

https://dspace.org/

# Examining the Climate Sensitivity of Tourists in South Africa Using TripAdvisor: A Big Data Approach Through Web-Scraping

**Mokgehle, Dineo Revinwa**

**2025-02**

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/48707

Mokgehle, Dineo Revinwa. (2025). Examining the Climate Sensitivity of Tourists in South Africa Using TripAdvisor: A Big Data Approach Through Web-Scraping. [PhD thesis, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/48707

Downloaded from a DSpace repository. https://dspace.org/

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG



# Examining the Climate Sensitivity of Tourists in South Africa Using TripAdvisor:

## A Big Data Approach Through Web-Scraping

**Dineo Revinwa Mokgehle (1095685)**

**Supervisor: Prof Jennifer Fitchett**

**PhD Thesis**
**February 2025**

# Declaration

I declare that this thesis is my own, unaided work, except where otherwise acknowledged. It is being submitted for the Degree of Doctor of Philosophy in Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.



_____

Dineo Revinwa Mokgehle

_____4th_____day

of__February_____20_25_____at__Johannesburg Braamfontein Werf_____

# Abstract

The viability and competitiveness of tourism locations depends, in part, on weather and climate. Climate determines tourists' choice of locations, activities, and vacation lengths. Netnographic research utilising TripAdvisor reviews can objectively evaluate tourists' climatic sensitivity and determine the meteorological factors most important to them. Manually reading and coding hundreds of reviews is, however, time-consuming. Web-scraping, which is the process of obtaining large amounts of data from various websites, can automate this process to analyse hundreds of thousands of reviews nationwide and evaluate climatic influence more comprehensively. This thesis compares the outputs of manually collected and web-scraped TripAdvisor reviews, to evaluate the efficacy of a web-scraping approach. A comparative analysis of TripAdvisor reviews collected through manual and web-scraping methods was conducted for 19 South African tourist destinations. The objective was to assess the efficacy of web-scraping when exploring weather-related data. The two methodologies demonstrated a 3.5% difference in the total number of climatic occurrences. Although there were significant variations in the number of climate mentions in three locations, namely East London, Polokwane, and Knysna, no statistically significant differences were calculated between the two methodologies in terms of the total number of climate mentions. Web-scraping has been demonstrated to be an efficient method for the real-time collection of big data- which is large volumes of data sets that are generated at high speeds that cannot be handled effectively by traditional processing methodologies and has provided a comprehensive understanding of the sensitivity of tourists to climate. Sentiment analysis revealed that tourists expressed negative or extremely negative sentiments regarding the majority of weather scenarios, with the exception of good weather, mist, and hail. To comprehend the impact of climate on tourist perceptions, regional variations in the frequency and proportion of weather-related mentions were investigated. Cold, hot, and rainy weather were the most frequently mentioned weather conditions in South Africa, with significant regional variations. Cold conditions were most frequently mentioned in Gauteng, cloudy weather in the Western Cape, windy conditions in Mpumalanga, thunderstorms in the Eastern Cape, and hot weather in the Northern Cape. These results indicate that weather significantly influences tourist experiences, and web-scraping is a viable method for obtaining these insights. It has been shown that online scraping is an efficient method of acquiring large quantities of real-time data and providing a full understanding of visitors' sensitivity towards climate. Consequently, it may be considered and accepted as a reliable means of data collection, facilitating more comprehensive and widespread analyses of tourists' climatic sensitivity from TripAdvisor review.

# Acknowledgements

I would love to thank my Lord and Saviour Jesus Christ for His guidance, wisdom, love and being a pillar of strength in the various seasons of this PhD thesis. Glory and honour be to You.

I would like to acknowledge Professor Jennifer Fitchett for the love, patience, support and guidance offered throughout this research. I would like to honour the dedication and time invested in me by you. The helpful workshops you would organise for the awesome biometeorology lab to the encouragements you would send show your beautiful heart. You helped me gain confidence in myself again. It was through your guidance that I was able to collect data, interpret them and give valid recommendations successfully. Without your help throughout the year, it would be more difficult to reach this status of completion. I would also like to acknowledge your kindness for providing funding and opportunities for me to present my research at multiple conferences. You are a great supervisor.

I would like to acknowledge the University of the Witwatersrand Postgraduate Merit Award and the Council for Scientific and Industrial Research (CSIR) for funding me throughout the duration of this research.

I would like to also acknowledge the Makhoba family, Mr Pillay and Tshepiso for assisting in editing and modifying the code utilised for this research.

To my husband Bukhosi Nkala, I honour you. Thank you for your guidance, love and care throughout all those long nights and endless days conducting research. You are such a blessing to me. Lastly, I would also like to acknowledge my parents and siblings for their involvement, divine encouragements and motivation to a successful completion of this thesis.

# Table of contents

**DECLARATION** ..... I

**ABSTRACT** ..... I

**ACKNOWLEDGEMENTS** ..... II

**LIST OF FIGURES** ..... VI

**LIST OF TABLES** ..... VIII

**LIST OF ACRONYMS** ..... IX

**LIST OF SCIENTIFIC TERMS** ..... XI

**LIST OF EQUATIONS** ..... XII

**LIST OF WORKS** ..... XIII

**CHAPTER 1: INTRODUCTION** ..... 1

1.1 Background ..... 1
1.2 Rationale ..... 4
1.3 Aim and Objectives ..... 5
1.4 Structure of the Thesis ..... 6

**CHAPTER 2: LITERATURE REVIEW** ..... 8

2.1 Introduction ..... 8
2.2 Climate, Weather and Tourism ..... 9
*2.2.1 Tourist Destination Choices* ..... *9*
*2.2.2 Tourism Climate Indices* ..... *12*
*2.2.3 Tourists Weather Experiences, Perceptions and Preferences* ..... *16*
2.3 Climate Change ..... 18
2.4 Climate Change and Tourism ..... 23
*2.4.1 Global Perspective* ..... *23*
*2.4.2 African Perspective* ..... *26*
*2.4.3 South Africa* ..... *31*
2.5 Big Data and Tourism ..... 33
*2.5.1 The Application of Big Data in Tourism* ..... *33*
*2.5.2 Tourism and Social Media* ..... *35*
2.5.2.1 Online Review Platforms ..... 36
2.5.2.2 Social Media Analytics ..... 38
*2.5.3 Methods Used to Collect Data in Tourism Research* ..... *41*
2.6 Conclusion ..... 43

**CHAPTER 3: STUDY SITE** ..... 44

3.1 Study Site ..... 44
*3.1.1 Introduction* ..... *44*
3.2 Descriptive Geography of South Africa ..... 44
3.3 Climate and Weather ..... 45
3.4. Tourism in South Africa ..... 46

**CHAPTER 4: RESEARCH METHODOLOGY ..... 49**

4.1 INTRODUCTION ..... 49
*4.1.1 Data Collection ..... 49*
*4.1.2 Data Analysis ..... 53*
4.2 ETHICS CLEARANCE ..... 57

**CHAPTER 5: RESULTS ..... 59**

5.1 INTRODUCTION ..... 59
5.2 EFFICACY OF AUTOMATING THE ANALYSIS OF TRIPADVISOR DATA IN TOURISM AND CLIMATE STUDIES ..... 59
*5.2.1 Sample Size and Demographics ..... 59*
*5.2.2 Climatic sensitivity of tourists ..... 62*
5.3 SENTIMENT ANALYSIS ..... 71
*5.3.1 Sentiment Analysis ..... 71*
*5.3.2 Different Weather Themes and Sentiment Analysis ..... 73*
5.3.2.1 Cold ..... 73
5.3.2.2 Cloud ..... 74
5.3.2.3 Drought ..... 76
5.3.2.4 Bad Weather ..... 77
5.3.2.4 Good Weather ..... 79
5.3.2.5 Hot ..... 80
5.3.2.6 Humid ..... 82
5.3.2.7 Mist and Fog ..... 84
5.3.2.8 Rain and thunderstorm ..... 85
5.3.2.9 Sun ..... 88
5.3.2.10 Hail ..... 90
5.3.2.11 Wind ..... 91
5.4 REGIONAL EXPANSION OF ANALYSIS ..... 93
*5.4.1 Sample Size and Demographics ..... 93*
*5.4.2 Climatic mentions in reviews ..... 96*
5.4 CONCLUSION ..... 120

**CHAPTER 6: DISCUSSION ..... 122**

6.1 INTRODUCTION ..... 122
6.2 EFFICACY OF AUTOMATING THE ANALYSIS OF TRIPADVISOR DATA IN TOURISM AND CLIMATE STUDIES ..... 122
*6.2.1 Web-scraping ..... 122*
6.2.1.1 The Efficacy of Web-scraping ..... 122
*6.2.2 The Application of Web-scraping ..... 124*
6.3 THE EVALUATION OF TOURIST'S WEATHER EXPERIENCES FROM TRIPADVISOR REVIEWS ..... 127
*6.3.1 Factors Influencing Weather Experiences ..... 128*
*6.3.2 Analysis of Tourist Weather Experiences ..... 133*
6.3.2.1 Importance of Climatic Factors in Tourist Reviews ..... 133
6.3.2.2 The Relationship Between Tourists and the Weather They Experience ..... 134
6.3.2.3 Implications of Weather on Tourist Experiences ..... 136
*6.3.3 Implications for Tourism sector ..... 139*
6.4 LIMITATIONS ..... 141
*6.4.1 Data limitations ..... 141*
*6.4.2 Web-Scraping Limitations ..... 143*
*6.4.3 Sentiment Analysis Limitations ..... 144*

**CHAPTER 7: CONCLUSION ..... 146**

7.1 Introduction ..... 146
7.2 Synopsis and Significance of the Study Findings ..... 147
7.3 Achievement of Study Aim and Objectives ..... 148
7.4 Future Research Avenues ..... 152
**Reference List** ..... 154
**Appendix A: Ethics Letter** ..... 199
**Appendix B: Web scraping Code** ..... 200
**Appendix C: Publication** ..... 202

# List of Figures

**Figure 2. 1:** Confidence in anthropogenic contribution to climate change globally. "The evidence for elevated temperatures is mostly based on changes in metrics based on daily maximum temperatures; regional studies utilising additional indices (heatwave length, frequency, and severity) are also employed. Red hexagons denote areas where there is at least moderate confidence about a rise in hot extremes" (IPCC, 2021, p. 12). ..... 19
**Figure 2. 2:** "Global surface temperature changes recreated from paleoclimate archives (solid grey line, 1–2000) and direct data (solid black line, 1850–2020), both averaged and related to 1850–1900. b) Variations in global surface temperature during the last 170 years (black line) comparative to 1850–1900 and yearly aggregated" (IPCC, 2021, p. 7). ..... 20
**Figure 2. 3:** Map outlining different countries within Southern Africa where tourism and climate change research has been conducted. The darker the colour code the more research conducted. ..... 28
**Figure 3. 1:** Map displaying the various locations for which TripAdvisor reviews were retrieved with the TCI scores (Fitchett et al., 2017) interpolated using Inverse Distance Weight (IDW). ..... 45
**Figure 4. 1:** Data processing flowchart ..... 53

**Figure 5. 1:** Map illustrating the percentage differences in Tripadvisor review distributions gathered through manual and web scraping methods. ..... 61
**Figure 5. 2:** Proportion of climate references obtained through web scraping in the aggregated TripAdvisor reviews for 19 South African locations ..... 65
**Figure 5. 3:**The proportion of TripAdvisor reviews that mention weather conditions, subdivided by location. ..... 66
**Figure 5. 4:** Location-specific proportions of each climate component ..... 70
**Figure 5. 5:** Sentiment analysis of cold related climatic conditions ..... 74
**Figure 5. 6:** Sentiment analysis of cloud related climatic conditions ..... 76
**Figure 5. 7:** Sentiment analysis of drought-related climatic conditions ..... 77
**Figure 5. 8:** Sentiment analysis of bad weather-related climatic conditions ..... 79
**Figure 5. 9:** Sentiment analysis of good weather-related climatic conditions ..... 80
**Figure 5. 10:** Sentiment analysis of hot related climatic conditions ..... 82
**Figure 5. 11:** Sentiment analysis of humid related climatic conditions ..... 83
**Figure 5. 12:** Sentiment analysis of mist and fog related climatic conditions ..... 85
**Figure 5. 13:** Sentiment analysis of rain related climatic conditions ..... 88
**Figure 5. 14:** Sentiment analysis of sun related climatic conditions ..... 90
**Figure 5. 15:** Sentiment analysis of hail related climatic conditions ..... 91
**Figure 5. 16:** Sentiment analysis of wind related climatic condition ..... 93
**Figure 5. 17:** Map illustrating proportional distribution of TripAdvisor reviews from each country ..... 94
**Figure 5. 18:** Number of TripAdvisor reviews per year from the South African sample. ..... 95
**Figure 5. 19:** Number of Tripadvisor reviews per month from the South African sample. ..... 96
**Figure 5. 20:** Number of climatic mentions per year ..... 99
**Figure 5. 21:** Number of climatic mentions per month ..... 100
**Figure 5. 22:** Climatic mentions ..... 103
**Figure 5. 23:** Climatic conditions per province ..... 104

**Figure 5. 24:** Bad weather-related mentions in South Africa ..... 105
**Figure 5. 25:** Cold weather-related conditions in South Africa ..... 107
**Figure 5. 26:** Drought weather-related mentions in South Africa ..... 108
**Figure 5. 27:** Fog weather-related mentions in South Africa ..... 109
**Figure 5. 28:** Good weather-related mentions in South Africa ..... 110
**Figure 5. 29:** Hail weather-related mentions in South Africa ..... 111
**Figure 5. 30:** Heatwave weather-related mentions in South Africa ..... 112
**Figure 5. 31:** Hot weather-related mentions in South Africa ..... 113
**Figure 5. 32:** Humid weather-related mentions in South Africa ..... 114
**Figure 5. 33:** Mist weather-related mentions in South Africa ..... 115
Figure 5. 34**:** Rain weather-related mentions in South Africa ..... 116
**Figure 5. 35:** Sun-related weather mentions in South Africa ..... 117
**Figure 5. 36:** Thunderstorm weather-related mentions in South Africa ..... 118
**Figure 5. 37:** Wind-related weather mentions in South Africa ..... 119
**Figure 5. 38:** Cloud-related weather mentions in South Africa ..... 120

# List of Tables

**Table 2. 1:** Literature on Text Mining and Analysis of Online Reviews (After Taecharungroj and Mathayomchan, 2019) ..... 39

**Table 3. 1:** Tourist sites ..... 47

**Table 5. 1:** Results pertaining to the comparison between web scraped and manual reviews from various countries. ..... 62
**Table 5. 2:** Mentions of the weather in TripAdvisor reviews. ..... 62
**Table 5. 3:** Statistical analysis performed between the various climatic mentions between both datasets. ..... 63
**Table 5. 4:** Evaluations of weather experiences ..... 72
**Table 5. 5:** Proportion of climatic mentions ..... 98
**Table 5. 6:** Proportion of climatic mentions ..... 101

**Table 6. 1:** Misclassified words ..... 143

# List of Acronyms

APR: April

AR6: Assessment Report 6

BERT: Bi-directional Encoder Representation from Transformers

BnB: Bread and Breakfast

BMW: Bayerische Motoren Werke

CART: Classification and Regression Trees

CHGE: Centre for Health and the Global Environment

CNN: Convolutional Neural Network

CSIR: Council for Scientific Research

CSV: Comma Separated Values

CSS: Cascading Style Sheets

CT: Cape Town

CTI: Climate Tourism Index

CIT: Climate Index for Tourism

DSAT: Measurement of satisfaction of tourist destinations

DEC: December

ENNM: Ensemble Neural Network Model

eWOM: electronic Word of Mouth

FEB: February

FpK: Fanpage Karma

GDP: Gross Domestic Product

GHG: Greenhouse Gas

GPS: Global Positioning System

GUI: Graphical User Interface

HCI: Holiday Climate Index

HRECNM: Witwatersrand Human Non-Medical Research Ethics Committee Clearance Number

HTML: Hypertext Markup Language

HTTP: Hypertext Transfer Protocol

HOSVD: Higher-Order Singular Value Decomposition

IDE: Integrated Development Environment

IDW: Inverse Distance Weight

IPCC: Intergovernmental Panel for Climate Change

IOVO-SVM: Improved one-vs-one strategy-based Support Vector Machine

ITCZ: Intertropical Convergence Zone

JAN: January

JUL: July

JUN: June

LDA: Latent Dirichlet Allocation

NLP: Natural Language Processing

NOV: November

MAR: March

OCT: October

PMB: Pietermaritzburg

SA: South Africa

SDG: Sustainable Development Goals

SEPT: September

ST: Saint

SVM: Support Vector Machine

TCI: Tourism Climate Index

TF-IDF: Term Frequency-Inverse Document Frequency

UK: United Kingdom

USA: United States of America

UNEP: United Nations Environment Programme

UNFCCC: United Nations Framework Convention on Climate Change

UNWTO: United Nations World Tourism Organisation

WEF-TTCR: World Economic Forum - Travel & Tourism Competitiveness Report

WTTC: World Travel and Tourism Council

# List of Scientific Terms

°C: Degrees Celsius

COVID-19: Coronavirus Disease 2019

$CO_2$: Carbon Dioxide

PPM: Parts Per Million

PPB: Parts Per Billion

$N_2O$: Nitrous Oxide

NO: Nitric Oxide

$SO_2$: Sulphur Dioxide

$PM_{10}$: Particulate Matter With 10 Micrometres Diameter or Less

# List of Equations

Equation 1: Arithmetic Mean

# List of works

While this is a PhD by thesis, the following outputs have come from this work:

1. Mokgehle, D. and Fitchett, J.M. (2024). Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies. *Tourism Planning & Development*, 21(6), pp.1-21. DOI: 10.1080/21568316.2024.2412060

Status: The paper was published on the 11th of October 2024.

Conference presentations:

1. Mokgehle D and Fitchett JM. (2022). Examining the climate sensitivity of tourists in South Africa using TripAdvisor: A big data approach through web-scraping. African Biometeorology Colloquium, 8 June 2022.
2. Mokgehle D and Fitchett JM. (2022). Examining the challenges in developing a web-scraping tool in automating and analysing Tripadvisor reviews. Society of South African Geographers Biennial Conference, 12-14 September 2022
3. Mokgehle D and Fitchett JM (2023). Web-scraping vs Manual Data collection: A comparative analysis of climate related tourism reviews in South Africa. Wits Inter-faculty symposium on Climate, Sustainability and Inequality, 23- 24 August 2023.
4. Mokgehle D and Fitchett JM (2023). Web-scraping vs Manual Data collection: A comparative analysis of climate related tourism reviews in South Africa. From Recovery to Resilience in Tourism: Sustainable pathways for transforming tourism conference, 10-11 September 2023.

# Chapter 1: Introduction

## 1.1 Background

Climate is a key factor in shaping the sustainability of tourism sectors (Amelung *et al.,* 2007; Gössling *et al.,* 2012; Fitchett *et al.,* 2020). Climate significantly influences tourists' destination choices, length of stay, and types of holiday activities (Amelung *et al.,* 2007; Fitchett *et al.,* 2020). Additionally, climate has an effect on seasonality, and weather has an effect on the general contentment of tourists staying in a particular region (Becken, 2005; Fitchett *et al.,* 2018). This mainly has an impact on the likelihood of tourists recommending these attractions to others via word of mouth (Hübner and Gössling, 2012; Fitchett *et al.,* 2020). Tourism, as one of the fastest expanding sectors in the world, is highly susceptible to climate change due to its strong ties to the environment and climate (Fang *et al.,* 2018).

South Africa is anticipated to suffer abnormally large repercussions as a result of climate change, attributable to the increased intensity of modelled impacts and the reduced ability to adjust to changing conditions caused by more pressing developmental requirements (Reid and Vogel, 2006; Fitchett *et al.,* 2017). Thus, the viability of the tourism sector in South Africa is dependent on the type and degree of climate change, as well as the tourism sector and the adaptation capabilities of the country (Rogerson, 2016). South African tourism is also linked to local economic development strategies aimed at rescuing South Africans from the confines of poverty imposed by the apartheid past of the country (Rogerson, 2002; Fitchett *et al.,* 2017). Due to a disproportionate emphasis on outdoor attractions and a slow pace of climate change adaptation, the tourism sector in South

Africa is exposed to the full effects of climate change. These effects have the potential to have a significant negative impact (Fitchett *et al.,* 2016).

Internationally, South Africa is recognised to be among of the hardest afflicted nations by the consequences of climate change and is also ranked as a contributor to climate change (Scott *et al.,* 2019). The country now ranks as being one of the top 15 Greenhouse Gas (GHG) emitters and has a record of generating high emissions (Scott *et al.,* 2019). Conversely, tourism is one of the major sectors that fuels the economy of the country. It is highlighted that in Sub-Saharan Africa, South Africa is recognised as the country with the top tourism locations (WEF-TTCR, 2015). This is largely credited to the country drawing a huge number of tourists globally owing to its fascinating tourism assets which include a broad range of cultural heritage sites (Hall, 2016).

The South African tourism sector is placed as fourth internationally owing to its quick development due to its attractive tourism assets (Braimoh and Akindoabe, 2010; Amusan and Olutola, 2017). There are various tourist attractions within the tourist sector such as the Cape winelands, wildlife reserves, and diverse adventure activities all contribute to South Africa achieving the position of being the fourth largest and most rapid tourism sector globally (Braimoh and Akindoabe, 2010; Rogerson, 2016). Developing economies are more dependent on areas such as tourism as compared to established economies (Khan *et al.,* 2020). South Africa as a geopolitical force in sub-Saharan Africa needs to further advocate for work to be done to prevent climate change.

Assessing the reliance of the tourism sector on climate, how susceptible it is to climate change in the region, and the capacity for adaptation is essential for enabling successful containment of these impacts (Hambira and Saarinen, 2015; Fitchett *et al.,* 2018). The dependence on climate and the ability to adapt to climate change are essential for maintaining the tourism sector in a specific

location (Fitchett *et al.,* 2018). Accurate information regarding the current climate suitability of regions for tourism is essential for optimising the effectiveness of adaptation and mitigation actions (Hoogendoorn and Fitchett, 2016; Fitchett *et al.,* 2018). To then comprehend the comparative climatic advantages and drawbacks experienced by a given area and to evaluate its climatic suitability in comparison to alternative destination possibilities, an empirical and homogenous technique is needed (Fitchett *et al.,* 2020).

To do this, an impartial estimate of visitors' net climatic sensitivity throughout their holidays is required (Fitchett and Hoogendoorn, 2018; Fitchett *et al.,* 2020). This is challenging to authenticate; questioning tourists about climate can unwittingly induce them to contemplate climate even if it was not a topic of discussion during their vacation (Denstadli and Jacobsen, 2014; Fitchett *et al.,* 2020). Regional indices were created to measure the tourism suitability of a region by incorporating multiple climate data (Becken, 2010). However, due to a lack of climate data, these indices have been found to be biased in some cases and thus have rarely been tested in the Global South (Hoogendoorn and Fitchett, 2016). As a result, TripAdvisor provides a valuable forum for travellers to submit self-reported free-form reviews on their trip satisfaction (Mkono, 2012; Fitchett *et al.,* 2020).

Climate generally refers to the long-term average of weather patterns, whereas public perception often focuses on extreme weather events (Hansen *et al*., 2011; Keely and Syphard, 2017; IPCC, 2021). While weather and climate are inextricably linked, they should be examined separately. The word 'climate' refers to the predominant state of weather seen over a lower temporal resolution in a particular region (Hansen *et al*., 2011; Keely and Syphard, 2017; IPCC, 2021). In comparison, “weather is the expression of climate at a particular region and time” (Becken, 2010, p.2). In

alternative terms, while weather is concerned with variations and extremes, climate data takes these into account.

## 1.2 Rationale

The weather experienced during vacations is generally believed to influence the perception of tourist attractions and the overall vacation experience. Weather remains one of the few important factors that the tourism sector cannot affect. Tourist organisations and individual tourists must respond to weather conditions as they occur, yet there exists a deficit of empirical evidence regarding the relationship between weather and individual tourist perceptions, activities, and the reflection of these factors in tourist experiences, which is a fundamental postulation in the tourism sector. Consequently, an important knowledge deficit remains regarding a structural analysis of the impact of weather on particular tourists. TripAdvisor reviews potentially yield more unbiased results, as users are permitted to share any aspect of their vacation experiences. The frequency of climate or weather references provides a more precise indication of the importance of weather to travelers compared to surveys that directly inquire about weather conditions. The assessment of TripAdvisor reviews to ascertain tourists' climatic sensitivity has been successfully conducted across 19 destinations in South Africa (Fitchett and Hoogendoorn, 2018; 2019), offering insight into whether climate is significant in influencing tourists' positive and negative interactions in addition to as establishing the climatic conditions to which tourists exhibit the greatest sensitivity. However, all of these studies have used a manual approach in extracting data from the TripAdvisor site. This method can be labour intensive, prone to human errors and biases and cannot be rolled out on a larger scale. In light of the aforementioned issue, it is vital to investigate the South African tourist sensitivity through the use of TripAdvisor reviews through a platform that is not labour intensive, prone to human error or biases but is capable of collecting and integrating this

information. This information gateway will aggregate data from multiple sources and subsequently categorise it. The method employed for this purpose is referred to as web-scraping. This project aims to offer an affordable and straightforward method for gathering public tourism data from TripAdvisor, analysing it, and implementing it on a broader scale to investigate tourist sensitivity to weather and climate in South Africa, ultimately enhancing awareness of the importance of data collection in the tourism sector. The use of big data in tourism is of significance as it provides effective and efficient methodologies on extracting and processing data from various social media platforms, TripAdvisor for example, without directly interacting with individuals. Based on previous literature (Fitchett and Hoogendoorn, 2019) it is imperative that the following be answered: 1) What are the various climatic factors that tourists are most sensitive to? 2) What is the efficacy of web scraping in automating data in comparison to traditional methods of collating data in the field of biometeorology? 3) How can tourists' concerns and perspectives regarding climate and weather be analysed and better understood to develop effective adaptation strategies?

## 1.3 Aim and Objectives

**Aim:**

This thesis seeks to examine tourist's sensitivity to weather and climatic aspects while visiting various regions of South Africa through an analysis of big data from TripAdvisor.

**Objectives:**

1. Develop an effective web-scraping tool to run on all TripAdvisor reviews for a given region, and aggregate mentions of weather and climate

2. Test the effectiveness of the web-scraping tool against manual counts that have been published for south Africa

3. Expand the analysis to all accommodation establishments in South Africa through the combined use of web-scraping

4. Investigate the climatic suitability of tourists visiting South Africa through the use of sentiment analysis

## 1.4 Structure of the Thesis

Chapter two focuses on the literature review, addressing themes including climate, weather, climate change, tourism, and the utilisation of big data in tourism research. It synthesises the current literature, emphasising the scientific knowledge to which this study provides. The subsequent chapter examines the research site, detailing weather, economy, and tourism of South Africa. Chapter four explores the research approach utilised in the study, starting with the selection process of weather observatories, followed by an in-depth review of the methodological procedures, covering data cleaning, understanding the correlation between tourism and climate, developing the web-scraping tool, and conducting an analysis of data. Chapter five explores the results of the study, detailing the statistics obtained from multiple analyses. Chapter six analyses the findings, interpreting them with respect to the aims and objectives of the study, along with the pertinent source material. It assesses the efficacy of web-scraping and the relationship between

tourists and weather, discusses the discovered limitations, and emphasises the advantages of the study. The concluding chapter evaluates the attainment of the aims and objectives of the study, proposes avenues for future research, and examines the principal conclusions and their implications.

# Chapter 2: Literature Review

## 2.1 Introduction

Tourism refers to tourists (transnational tour operators and residents) visiting specific destinations due to a natural or human-made attractions (Dube and Nhamo, 2020). Tourism is among the most rapidly growing economic sectors (Nicolaides, 2020). The environment, tourism providers, local communities and tourists are all inextricably intertwined and dependent on one another, such that even minimal changes in one component can have a significant impact on the others (Moreno, 2017). For instance, the decision of a tour operator to add or exclude a location from their portfolio will significantly affect the environmental, social, and economic conditions of the destination locality (Moreno, 2017). Interdependence complicates tourism operations, as tourism flows are influenced to some extent by the operations of the sector and are impacted notably by social and environmental aspects that are not always predictable (Kádár and Gede, 2021). Climate and weather significantly influence tourists' decision-making processes, thereby playing an essential role in the tourism sector (Fitchett and Hoogendoorn, 2018). The relations between tourism, climate, and weather have been researched in various disciplines (Martin, 2005; Dogru *et al*., 2019). The variations that exist within climate spatially and temporally coupled with recreational activities give way to different techniques and methodologies needed to scrutinise environmental conditions for sustainable tourism to occur (Dogru *et al*., 2019).

In some economies, tourism plays a significant role; however, climate change can adversely affect this sector (Fitchett *et al.*, 2017; Leal *et al.*, 2024; Phuyal *et al.*, 2020; Tanrisever *et al.*, 2024; Zhou *et al.*, 2024). Climate change has the propensity to have widespread detrimental consequences on

the tourist sector on both global and local scales (Pandy and Rogerson, 2019). Additionally, United Nations World Tourism Organisation (UNWTO) and United Nations Environment Programme (UNEP; 2008) have indicated that climate change will influence tourism markets both indirectly and directly, with direct effects affecting tourists' key decision processes concerning travel destinations and timing, and indirect impacts resulting from environmental changes within a destination (Fitchett *et al.*, 2017). Various global entities have acknowledged the impact on tourism of climate change, while also highlighting the bidirectional link amid tourism and climate change, noting the influence of tourism on the environment. (Cavallaro *et al.*, 2021).

This chapter will review and analyse relevant literature based on the following themes: the impact, on tourism, of climate change and the application of big data in tourism as well as the complexities inherent in the interaction between tourism and climate change. The chapter will also aim to evaluate the use of big data and social media in the field of tourism at regional and global levels.

## 2.2 Climate, Weather and Tourism

### 2.2.1 Tourist Destination Choices

The climate of a tourists' home plays an important role in their sensitivity to climate, the value of climate in their choice of destination, and their level of acclimatisation (Koutroulis *et al.*, 2018). Unfavourable home weather conditions, whether in the year of trip or the preceding year, encourages tourists to go to destinations that have higher thermal comfort (Dube and Nhamo, 2020). Many locations across the world have tourist potential but are constrained by climate and weather (Gössling and Scott, 2024). For instance, if a destination lacks favourable climatic conditions, tourism managers may be less inclined to market it as it would be unprofitable, or

visitors may endure discomfort due to extreme temperatures if they visit certain destinations (Fitchett et al., 2020). The terms "push and pull factors" are frequently used when discussing travel behaviour (Chen *et al*., 2016). The selection of tourist destinations may be influenced internally by socio-economic variables or externally by destination qualities (Chen *et al*., 2016; Velea *et al.,* 2022). Weather and climate conditions might be considered one of such destination characteristics (Becken and Wilson, 2013; Velea *et al.,* 2022). Weather may operate as an attraction factor, luring travellers to a place with pleasant weather (Fitchett *et al*., 2020). Adverse climates or inclement weather in travellers' native countries might function as a catalyst for tourists to go to warmer or drier regions (Jeuring, 2017). The research carried out by Giddy *et al*, (2017) aimed at exploring American tourists' experiences with regards to weather and climatic factors highlighted the influence of climate and weather on destination choice. The findings demonstrated that, while respondents did not believe that weather conditions were a major role in their decision to explore the region, they did believe that weather that they had experienced had a noteworthy effect on the overall contentment of their vacation as the weather limited their ability to engage in outside activities (Giddy *et al*., 2017). A close connection appeared when integrating descriptions of unfavourable weather conditions with meteorological records, particularly for exceptionally high temperatures (Giddy *et al*., 2017). This demonstrates that the perceptions of American tourists accurately reflect climate ineptness for tourism, posing risks to the South African outdoor tourism business (Giddy *et al*., 2017). A comparable study done by Fitchett and Hoogendoorn (2019), demonstrated that 7.9% of tourists from TripAdvisor reviews examined, the climatic suitability of South Africa with regards to tourism. Tourist choices may be highly influenced by the climate in the destination of interest.

Another aspect that has also been defined as a challenge to the viability of tourism is seasonality (Gössling *et al.*, 2016). The seasonality of tourism influenced by various climatic factors at tourist locations and also through indirect or institutional influences such as school holidays or events (Martin *et al.*, 2014; Gössling *et al.*, 2016). The consistent need and urge to measure and evaluate the climatic suitability of tourist destination has long existed (Fitchett and Hoogendoorn, 2017). A study conducted by Fitchett *et al*., (2017) in South Africa aimed at measuring climatic suitability in 18 distinct locations proved that distinct seasonality with regards to Tourism Climate Index (TCI) scores is evident however, the contrast between summer and precipitation zones of South Africa ensures a net balance in climatic adaptability year-round. Climatic changes, such as precipitation and average temperature, may influence tourists' future decisions regarding travel frequency and duration, or may entice tourists to seek alternate holiday locations or go at a different period throughout the year (Hambira, 2011; Hambira and Mbaiwa, 2020). As a result, climate can have an effect on the seasonality of tourism, overall satisfaction, and the activities accessible to visitors during a holiday (Hoogendoorn and Fitchett, 2016).

Tourist satisfaction is a response or choice made by visitors in response to emotion or thought (Bigne *et al*., 2001). In general, satisfaction represents the emotional state and thinking that follows the experience of a prospect (Ranjanthran and Mohammed, 2010). Social and psychological factors like expression, behavior, and wants of tourists and environmental factors like weather and social group interactions effect satisfaction (Baker and Crompton, 2000). A customer's satisfaction with a product or service is called contentment (Ranjanthran and Mohammed, 2010). Reputation and service quality affect tourist satisfaction, according to multiple research (Bigne *et al*., 2001; Chen and Tsai, 2007; Ranjanthran and Mohammed, 2010; Ngoc and Trinh, 2015). Also, it appears that good contentment contributes to improved future behaviour, such as an enhanced motive to

return and a stronger inclination to suggest (Ranjanthran and Mohammed, 2010; Chen and Tsai, 2007). However, according to the findings of observational research conducted in two major cities located in Spain (Bigne *et al*. (2001), indicates that tourist satisfaction impacts the likelihood to recommend the destination to acquaintances, while it does not influence the intention to revisit. In contrast, factors including lodging quality, accessibility, scenic beauty, weather or climate, and cleanliness are recognised as the most important indicators of visitor satisfaction (Um et al., 2006; Som et al., 2012; Lai and Vinh, 2013).

Martin *et al*. (2019) published a study that added to the discussion on how to quantify tourist satisfaction. The objective was to assess the level of satisfaction among international visitors to South Africa using fuzzy numbers. Tourist destination satisfaction (DSAT) is evaluated through a model that considers factors like native language, age, and gender (Martin *et al.,* 2019). It was discovered that the empirical evidence for DSAT disparities between the segments studied. Females are more content than men; tourists between the ages of 56 and 65 are the most satisfied, while millennials under the age of 35 are the least satisfied; English native residents were more satisfied than other visitors.

### 2.2.2 Tourism Climate Indices

To enhance the sustainability of developmental measures, precise information concerning climate change in areas relevant to tourism is important (Hoogendoorn and Fitchett, 2016). Numerous indices have been established in the past few years to evaluate the impact of climate and climate change on tourism (Fergusson, 1964; Murray, 1972; Mieczkowski, 1985; Harlfinger, 1991; De Freitas *et al.*, 2008; Scott *et al.*, 2016; Fitchett *et al.*, 2017; Scott *et al.*, 2018). These indices have been used to quantify the climatic suitability for a range of tourist activities, namely water and

beach activities, wildlife viewing with respect to climate change (Kovács and Unger, 2014; Fitchett *et al.,* 2017). The Tourism Climate Index (TCI) was initially developed by Mieczowski (1985). Mieczkowski (1985) pioneered the utilisation of a TCI on a global level to assess climatic readiness for tourism for locations. The TCI is a composite index that incorporates numerous meteorological variables, including wind speed, daily sunshine hours, temperature, humidity, and precipitation, (Gössling *et al*., 2016; Fitchett *et al*., 2017).

Over the years, the TCI has been modified and applied to specific tourist settings and attractions such as beach tourism (De Freitas, 2003; Hasan *et al.,* 2020; Rutty et al., 2020), urban tourism (Scott et al., 2016), camping tourism (Ma *et al*., 2021), ski tourism (Demiroglu *et al.,* 2021). Due to the extensive use and relatively broad application of the TCI, it is particularly well-suited for preliminary investigations on climate appropriateness for tourism in the climatically diverse global South (Fitchett *et al.,* 2017). However, two critical criteria are noticeably omitted from these quantitative indicators of tourism climatic suitability. The first is concerned with the infrastructure of tourism lodging places and destinations and the extent to which it alleviates the effects of inadequate climatic conditions (Verbos and Brownlee 2017; Fitchett and Hoogendoorn, 2018). The second element is the tourist's place of origin, which influences volatility in part due to the climate they encounter most frequently (Gössling and Hall 2006; Fitchett and Hoogendoorn, 2018). These variations in tourists' preconceptions and criterion interactions of weather erode the predictive power of indices based solely on impartial climatic information in establishing the climatic suitability for tourism of a location (Gössling and Hall 2006; Fitchett and Hoogendoorn, 2018). The climatic components that appear to be important include radiation, windspeed effects and temperature (Becken, 2010; Hasan *et al.,* 2020). Tourists' perceptions of the weather cannot be attributed to a single weather or climatic component, since air conditions of the immediate

surrounding are reflections of a myriad of meteorological factors working simultaneously to create a particular weather condition (Becken, 2010; Hasan *et al.,* 2020). Thus, indices integrate various climatic data to provide a single indicator of climate appropriateness for a destination for certain tourism activities (Becken, 2010). Additionally, indices are frequently employed to reflect a tourist's acceptance of the ambient air conditions (Kovács *et al*., 2015).

Global tourism flow models have been developed that are predicated on a single environmental variable, including temperature, in order to predict the movement of tourists among nations (Hamilton *et al*., 2003; Priego *et al.,* 2015). De Freitas (2003) criticised a few of the early studies focusing on destination demand for being overly basic, as they depicted climate through using single parameters such as precipitation and/or temperature, rather than a multidimensional combination of varying factors. Hamilton and Lau (2004) discovered that temperature may be utilised as a primary determinant of travel demand, however temperature alone does not adequately capture visitors' destination climate concerns. De Freitas (2003) categorised the various aspects of tourist climate into three categories: thermal, physical, and aesthetic. The thermal aspect relates to the tourist's degree of satisfaction (De Freitas, 2003; Nasrollahi *et al.,* 2017). The physical component pertains to meteorological conditions such as rain and wind, that have an effect on tourism operations (de Freitas, 2003; Hoogendoorn and Fortune, 2018). The aesthetic element reflects a cognitive perspective, such as sunshine and cloud cover. De Freitas *et al*. (2008) proposed a new Climate Index for Tourism (CIT) that incorporates physical, thermal, and aesthetic factors affecting beach tourism. This index is akin to the TCI but differs in terms of it stemming from tourist reviews on their preferences with regards to weather and climate (Mieczkowski, 1985; Gössling *et al*., 2016). The CIT has indeed been employed recently (Kovács *et al*., 2015). It serves as a method for assessing the probability of occurrence of a variety of biometeorological and

meteorological circumstances (Kovács *et al*., 2015). The evaluations are derived from de Freitas's (2003) aspects of tourism and are subject to modification (Kovács *et al*., 2015). The preliminary CTI outlined by Zaninovic and Matzarakis (2009) was critiqued by Scott *et al*. (2012) because they were not derived from evaluations of what tourists prefer with regards to weather or climatic conditions that they would rather consider as ideal or unacceptable for diverse tourism activities or locations (Hoogendoorn and Fortune, 2018). Kovács *et al*. (2015) enhanced the CTI by adjusting the thermal aspects of the CTI and TCI utilising thermal experience and choice studies from Hungary. Scott *et al*. (2016) created the Holiday Climate Index (HCI), which is comparable to the CTI but is rather derived from published data on visitors' climatic interests and is utilised in a variety of metropolitan sites around Europe.

Other research has attempted to identify key criteria for tourism demand and thus to evaluate their relative significance to other variables (Gössling *et al*., 2016). Most research examines summer beach and coastal attractions (Lise and Tol, 2002; Köberl *et al*., 2016; Fitchett *et al.*, 2020) and winter season attractions (Damm *et al*., 2014). These studies seek to establish cut-off points for what constitutes favourable or unfavourable tourism weather conditions (Gössling *et al*., 2016). Furthermore, the perception of pleasant or unpleasant weather is arbitrary and is reliant upon the tourist's weather perceptions and preferences (Lohmann and Hübner, 2013). However, two critical criteria are noticeably omitted from these quantitative indicators of tourism climatic suitability. The first is concerned with the infrastructure of tourism lodging places and destinations and the extent to which it alleviates the effects of inadequate climatic conditions (Verbos and Brownlee 2017; Fitchett and Hoogendoorn, 2018). The second element is the tourist's place of origin, which influences volatility in part due to the climatic conditions they encounter most frequently (Gössling and Hall 2006; Fitchett and Hoogendoorn, 2018). These variations in visitors' preconceptions and

initial interactions of weather erode the predictive power of indices based solely on impartial climatic information in establishing the climatic appropriateness for tourism for a location (Gössling and Hall 2006; Fitchett and Hoogendoorn, 2018). While tourism climate indices are useful, their limitations however mean that it is crucial to find corroborating evidence of tourist experiences.

### 2.2.3 Tourists Weather Experiences, Perceptions and Preferences

The body of knowledge regarding tourist choices, preconceptions, perspectives, and destination engagements has broadened significantly (Lohmann and Hübner, 2013). Tourists' weather predilection needs to be apprehended to establish the acceptable scope of circumstances for major climatic variables and to enable the establishment of a framework of the profits and course of tourism (Dubois *et al*., 2016). The weather encountered by visitors at various tourist attractions is essential for various reasons including those outlined by Becken, (2012):

1. Weather has an influence on the tourist activities undertaken
2. Weather conditions are the key driver to a tourist's satisfaction or dissatisfaction with regards to a particular tourist activity undertaken
3. The safety of tourists is mainly reliant on weather conditions within the region

Given the complexity of the ambient environment, determining the precise preferences of tourists is challenging (Shambour et al., 2024). The issue is that human behaviour in response to weather systems is partially determined by their observations of weather (de Freitas, 2015). The perceptions of tourists may not always correlate with conventional weather parameters, for instance tourists' experiences with regards to temperature are not just impacted by ambient temperature but through humidity, wind, solar radiation as well as precipitation also (de Freitas,

2015; Hidayat *et al.,* 2024). Weather aspects can be seen as perfect, adequate, tolerable, or unsuitable depending on a variety of characteristics including the traveller's persona or destination type, career, culture, role and age (Tol *et al*., 2005; Rutty and Scott, 2010, 2014; Li *et al.,* 2018). Furthermore, Van Vuuren and Slabbert (2011) classified tourist travel reasons in South Africa as recreational activities, social experience, rest, seeking new experiences, personal worth and relaxation and self-improvement. Rutty and Scott (2014), for instance, used questionnaires to determine the thermal tolerance of coastal visitors. Temperatures below 34°C were reported to be tolerable, however thermal comfort varied by tourist age (elderly people tend to have a heat tolerance that is lower as compared to youth) and place of origin (Rutty and Scott, 2014). However, Dubois et al. 2016 then demonstrated that depending on the population surveyed and wording of questionnaire, these can result in quite different thresholds. Hewer *et al*. (2017) added to our understanding of how various factors such as trip duration, activity choices, age, and travel and expense management impact weather preferences and choices, revealing substantial variations, such as physically active campers placing a higher value on weather but being more considerate of adverse weather conditions and older campers placing a higher value on weather (Hewer *et al.,* 2017). Campers traveling shorter distances were more likely than those traveling greater distances to depart the campground due to inclement weather (Hewer *et al*. 2017).

Weather conditions are often not favourable, and travellers may often have varying reactions to adverse weather circumstances (de Freitas, 2003; de Freitas *et al*., 2015). Persistent floods, sudden snow or cold conditions, fog, wind, storms, extreme temperatures, and rain are all potential consequences of bad weather experiences (Gossling et al., 2016; Steiger et al., 2016; Jeauring, 2017; Giddy *et al*., 2017; Martinez-Ibarra *et al*., 2019). Temperature was observed to be more frequently mentioned by tourists in South Africa (Giddy *et al*., 2017).

## 2.3 Climate Change

Climatic change, as described by the United Nations Framework Convention on Climate Change (Hickman et al., 2021), are variations within global atmospheric conditions produced either directly or indirectly through anthropogenic activities and by natural climate variability over longer timeframes. Due to anthropogenic and natural processes constantly altering the atmospheric composition, these circumstances subsequently influence meteorological phenomena and conditions (Yanda, 2010; Gu *et al*., 2019). These shifts are significantly altering the temperature and precipitation cycles, as seen by seasonal flooding, increased temperatures, and droughts (Gu *et al*., 2019). Significant climate-change-induced meteorological fluctuations have been observed in South Africa since 1970, characterised by substantial intra and inter seasonal precipitation variability and rising drought (Rouault *et al*., 2013; Kruger and Nxumalo, 2015). Heatwaves and droughts, as well as associated hydro-meteorological extremes, are some of the most frequent natural extremes in South Africa (Kruger and Nxumalo, 2015). This predicament is exacerbated due to the reality that almost all human activities emit greenhouse gases (GHG), which are inextricably linked to economic development (Warburton *et al*., 2005; Reason, 2007; Stephan and Stephan, 2020; IPCC, 2021; Figure 2.1).

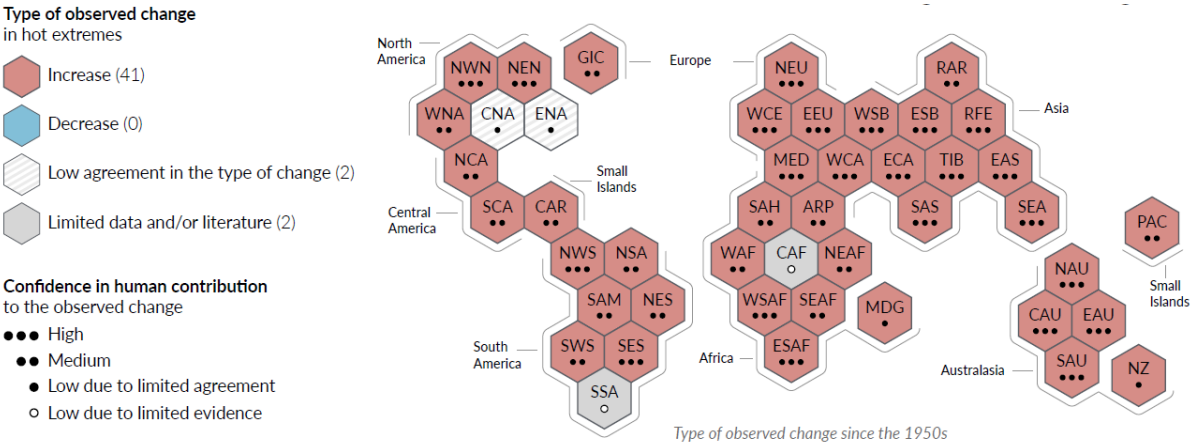



**Figure 2. 1:** Confidence in anthropogenic contribution to climate change globally. "The evidence for elevated temperatures is mostly based on changes in metrics based on daily maximum temperatures; regional studies utilising additional indices (heatwave length, frequency, and severity) are also employed. Red hexagons denote areas where there is at least moderate confidence about a rise in hot extremes" (IPCC, 2021, p. 12).

Anthropogenic induced increases in GHG emissions have been documented from roughly 1750 (IPCC, 2021). Concentrations in the atmosphere have steadily increased since 2011 (assessments published in Assessment Report 6, with seasonal averages of 332 parts per billion (ppb) for nitrous oxide ($N_2O$), 1866 ppb for methane ($CH_4$) and 410 parts per million (ppm) for carbon dioxide ($CO_2$) in 2019 (IPCC, 2021). Carbon dioxide as well as other GHGs are emitted into the atmosphere as a consequence of anthropogenic civilisation and related processes (Lu and Wang, 2018). The utilisation of fossil fuels in industrial and tourist operations releases dangerous carbon dioxide emissions into the atmosphere, causing heat to be trapped in the atmosphere of the earth (Lu and Wang, 2018). In the South African metropolitan centres of Johannesburg and Cape Town, domestic combustion and transportation emissions are significant causes of air pollution, owing mostly to their dense populations and primary economic activities (Lourens *et al.*, 2012). Nitrogen

oxide (NO), carbon monoxide (CO), sulphur dioxide ($SO_2$), nitrogen dioxide ($NO_2$), particulate matter with a diameter of 10 micrometres or less ($PM_{10}$), and different organic compounds are among the main pollutants generated by these operations (Lourens *et al.*, 2012).

A combination of proxies, archive data, and meteorological data has been utilised to construct complete climate assessments from 1850 to 2020 (Figure 2.2). These are also used to project future changes based on well-informed trajectories. Similar assessments have been conducted specific to South Africa, demonstrating regional changes in the climate (Sousa *et al.*, 2018; Dube *et al.*, 2018)

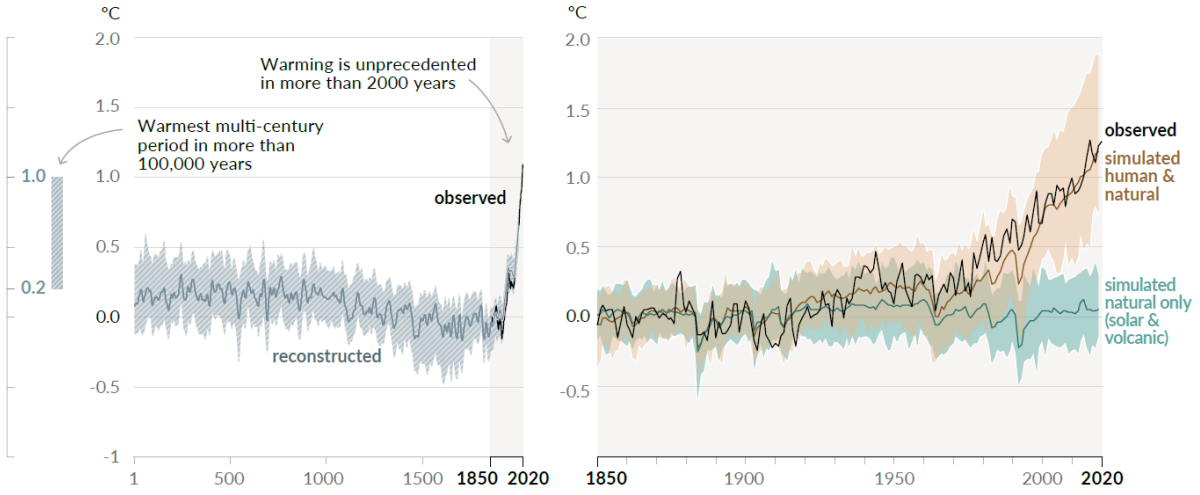



**Figure 2. 2:** "Global surface temperature changes recreated from paleoclimate archives (solid grey line, 1–2000) and direct data (solid black line, 1850–2020), both averaged and related to 1850–1900. b) Variations in global surface temperature during the last 170 years (black line) comparative to 1850–1900 and yearly aggregated" (IPCC, 2021, p. 7).

Global temperatures have increased, according to recent research (IPCC, 2021). The warmest periods ever documented happened subsequently; the hottest year was 2016, followed by 2015,

2017, and finally 2014 (IPCC, 2021). Future warming and GHG emission forecasts are on the rise. According to the IPCC (2021), GHG emissions are anticipated to decline in the future, and subsequently, the increase in global temperature may be kept to 1.5°C. Anthropogenic influences caused around 1.0°C of global warming, which varied from 0.8°C to 1.2°C by 2017 (IPCC, 2021). According to the IPCC, if the current pace of growth continues, global warming would likely reach 1.5°C between 2030 and 2052 (IPCC, 2021). The IPCC 2018 and 2021 special reports demonstrated that human-made forces in the climate system are explicitly influencing climate change (IPCC, 2018, 2021). More proof of global temperature rises is persistently mounting (Blunden and Arndt, 2017). As previously stated, global climates are constantly changing due to a variety of variables, including historical GHGs confined within the atmosphere (IPCC, 2018, 2021). In the future, climatic systems will create feedback mechanisms; the amplitude of these will be influenced by current human actions. Temperature rises are often accelerated by positive feedback, while they are slowed by negative feedback. In this situation, GHGs create positive feedback loops by absorbing heat and raising temperatures (Bellard et al., 2012; Mushewemhuka et al., 2021). These effects can be mitigated by reducing GHG emissions (Blunden and Arndt, 2017; IPCC, 2021). The inability to significantly and rapidly cut GHG emissions will lead to severe climate change, affecting all facets of life (IPCC, 2014, 2018, 2021).

As the world population increases prompting economic growth globally, so does the demand for natural resources increase, resulting in activities that emit GHGs including water vapor, carbon dioxide, methane, nitrous oxide, and ozone (Szulejko *et al*., 2016). Increased GHG emissions causes a rise in temperature across the globe, a phenomenon known as global warming, and it significantly contributes to climate change. Climate models project that worldwide climate change is occurring for the foreseeable future, with far-reaching consequences for sustaining life

(Szulejko *et al*., 2016). Since 1850, every one of the previous four decades have been warmer than the decade before it. In the initial twenty years of the twenty-first century (2001-2020), worldwide surface temperature was 0.99 [0.84-1.10] °C warmer than in 1850-1909 (IPCC, 2021). Climate researchers have determined a link between increasing global temperatures and rising sea levels, which has resulted in the flooding of certain remote island nations and coastal communities throughout the globe (Isobe, 2013). The detrimental effect of climate change on natural resources may be portrayed in a variety of other settings, including those of attractions of snow, where a key worry is the decline in tourism contentment and pleasant snow interactions (Cocolas *et al.,* 2015; Stockigt *et al.,* 2018).

Rising sea levels would complicate numerous infrastructure initiatives along coastal regions and have a detrimental influence on tourism (Patsch and Reineman, 2024). Rising sea levels will have a negative impact on coastal floods and anthropogenic disruptions, devastating low-lying coastal communities by inundation and flooding (Isobe, 2013). During 1950, global mean precipitation over land has most certainly grown, with a greater percentage increase since the 1980s (IPCC, 2021).

Human impact is very probably attributable to the trend of recorded precipitation variations ever since the mid-1900s, and it is quite probable that human impact has contributed to the sequence of documented variations in near-surface ocean salinity (Hsiang and Kopp, 2018). Since the 1980s, mid-latitude storm paths have most certainly migrated northward in the northern hemisphere, and southward in the southern hemisphere, with significant seasonal patterns (IPCC, 2021; Chemke, 2022). Human impact is extremely likely to be attributed to the poleward shift of the closely similar extratropical jet in austral summer in the Southern Hemisphere (Gerber and Son, 2014; Wood et al., 2020). Climate change will have a significant detrimental effect on water-based tourist

facilities, particularly those located around coastlines (Grenfell *et al*., 2016). Climate change impacts and will affect many geographic locations in different ways at the macro, 5th, and micro-scales. According to the Assessment Report 6 (AR6), existing information indicates that various portions of the globe would be affected through various ways (IPCC, 2021).

## 2.4 Climate Change and Tourism

### 2.4.1 Global Perspective

Tourism is a substantial worldwide economic industry that has increased substantially over the last 5 decades (UNWTO, 2011; 2018). The economic impact of the tourism sector globally has also increased, with the World Travel and Tourism Council (WTTC) assessing the input of the sector to the global economy in 2015 at US$7.2 trillion, representing 9.8% of global Gross Domestic Product (GDP), and 284 million employment opportunities, accounting for 9.1% of global jobs (WTTC, 2016a; Scott and Gössling, 2019). The tourist sector is of much significance in various destinations worldwide with more than 90 nations where tourism constitutes over 10% of the national GDP and a major share of jobs (WTTC, 2016b; Scott and Gössling, 2019). Given high anticipated increases in global tourist arrivals in many developing countries, international development agencies (WTTC, UNWTO) view tourism as a key component to the sustainable development goals (SDG) (Hall *et al.,* 2015; World Bank, 2017; Scott and Gössling, 2019).

Climate change and tourism have a symbiotic connection (Amusan and Olutola, 2017). Being the largest issue, that humanity is facing, climate change continues to affect the environment on a global scale (Nachmany *et al.,* 2014). The consequences of climate change may be seen in many industries, tourism included. As stated by the Centre for Health and the Global Environment

(CHGE; 2016), it is estimated that by the year 2050, climate change would have impacted one quarter of all species on land. As one of the fastest growing sectors in the world, tourism itself contributes to climate change (Becken et al., 2020; Pandy and Rogerson, 2019). The associated environmental effect related to tourist operations uncovers substantial issues to the globe, thereby considerably contributing to climate change (Loehr and Becken, 2021). Instances include GHG, deforestation, carbon emissions associated to the transportation engaged within tourist activities such as rail, aviation, water, and road all lead to the inefficient use of energy and waste generation (Amusan and Olutola, 2017). In the global south, summer-dependent coastal tourism, for instance, is substantially impacted by climate change in a variety of ways, such as elevated rainfall in some regions, a global increase in average temperatures, rising sea levels which result in damage to infrastructure, intrusion of salt water, as well as biodiversity loss and coral bleaching (Zeppel, 2012; Oyedotun *et al.*, 2018).

The threat that climate change presents to tourism expansion, has received limited attention (Becken, 2019; Hall, 2019; Janetos *et al.*, 2012; Saarinen and Rogerson, 2014; Scheyvens and Hughes, 2018). Within the context of developing nations, limited research has been published with regards to the existing relations between tourism and climate change (Hall *et al*., 2015; Scott *et al*., 2016). The tourist sector is sensitive to weather and climate and is also heavily impacted by a variety of variables such as the status of the natural environment, anthropogenic factors, and personal security perceptions (Scott *et al*., 2012a; Scott and Gössling, 2019). Climate change, according to the UNWTO *et al*. (2008, p. 38), is "...the biggest threat to the sustainability of tourism in the twenty-first century". The World Travel and Tourism Council (2018) endorsed the United Nations Framework Convention on Climate Change (UNFCCC) Climate Neutral Now program in 2018, promising to achieve climate neutrality by 2050 and collaborating on increasing industry

climate action (Gössling and Scott, 2018). As a result, the ramifications of climate change for numerous underdeveloped nations that intend to use tourism as a major prospective growth model must be carefully reviewed in policy development and planning, international aid programs, and global adaptive capacity negotiations (Gössling *et al.,* 2009; Scott *et al.,* 2012b). Given the extensive sectoral knowledge of the susceptibility of tourism to climate change (WTTC, 2015; Gössling and Scott, 2018), the divergent climate change effects encountered by the tourist sector at both local and global levels remain unknown. As indicated by tourist sector information in the geographic chapters of the Intergovernmental Panel on Climate Change (IPCC) Assessment Reports, possible implications in Africa, Asia, Central and South America, and Small Islands States remain inadequately recognised (Scott *et al*., 2016). Such regional gaps in knowledge are notable because they exist in various locations where tourism is expected to develop most rapidly within the next 30 years (Scott and Gössling, 2019). A further notable impediment to apprehending the local potential impacts of climate change for tourism economic viability and sustainability has been a dearth of analyses that take into account the broad variety of possible consequences and their interconnections at the destination scale (Scott *et al*., 2012b, 2016). The examination of climate change effects in seclusion is widespread in various research examined by Mora *et al*. (2018), which gives an inadequate and possibly misleading evaluation of the implications of climate change for a region or industry of concern. Due to the strong global interconnectedness of the tourist sector, the ramifications of climate change and affiliated adaptation and mitigation responses may also have crucial implications for destination countries, including the relative attractiveness of potential destinations (Hamilton *et al*., 2005; UNWTO *et al*., 2008; Scott *et al*., 2012a).

Research that has been conducted has in the past primarily focused on the effects of climate change on tourism demand, volatility, visitors' convenience, and destination selection (Amelung *et al*., 2007; Gössling *et al*., 2012; Serdeczny *et al*., 2017). Climate change scholars, in particular, focus on a wide variety of criteria such as the kind of destination, as well as region-specific, and worldwide studies (Hamilton *et al*., 2005; Amelung *et al*., 2007; Moreno, 2010; Gu *et al*., 2019). As a result, scholarly literature has concentrated primarily on the effects of climate change on the tourist sector, with little emphasis paid to the opposing side of the equation, namely the impact of the tourism sector on climate change (Gössling and Hall, 2006; Scott and Gössling, 2019). The tourism sector is experiencing considerable effects from climate change and simultaneously contributes significantly to its progression (Scott *et al*., 2012; Hambira and Saarinen, 2015). According to Gössling *et al*. (2012), tourism adds to climate change through a range of tourist activities, including increasing water usage, GHG emanations, and changes to local ecosystems.

### 2.4.2 African Perspective

The correlation between tourism and climate change may be viewed as bi-directional (Danish and Wang, 2018), as both entities influence each other. The daily operations, important investment decisions pertaining to the tourism sector may be affected by the climate and weather in the region (WTC, 2011; Damm *et al.,* 2020). Developing countries that depend significantly on land and resources and are vulnerable to climate change are generally found in the Sub - Saharan region (Kipkoech et al, 2015). These nations are mainly developing economies, making them more vulnerable to climate change and the hazards that it entails (Adger *et al*., 2003; Mushawemhuka *et al.,* 2021). This is mostly owing to insufficient financial resources and technological

advancements to address these changes through adaptation and mitigation (Kurukulasuriya *et al*., 2006; Hoogendoorn and Fitchett, 2018).

The susceptibility to climate change exacerbates the socioeconomic issues that African nations are experiencing (Hoogendoorn and Fitchett, 2018). Effects of climate change are apparent within activities surrounding African tourism. Different tourist destinations receive various tourist attention (Hoogendoorn and Fitchett, 2018). This is not only subject to weather, but the times of year, and if important events should be hosted adjacent to these locations (Hoogendoorn and Fitchett, 2018). Tourism and climate change research within Southern Africa is limited (Figure 2.3). It is thus necessary to have further research undertaken on the consequences that climate change will bring these tourism attractions and how it will affect the choices that tourists make with regards to destinations (Hoogendoorn and Fitchett, 2018).

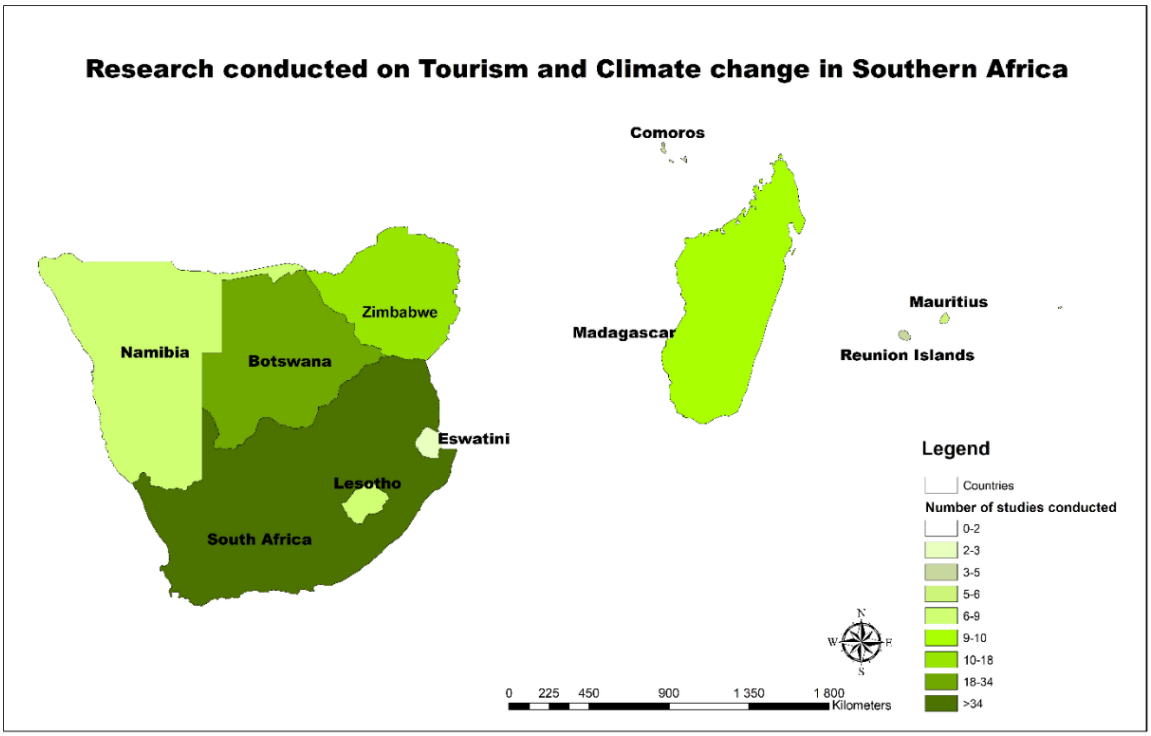



**Figure 2. 3:** Map outlining different countries within Southern Africa where tourism and climate change research has been conducted. The darker the colour code the more research conducted.

The existing research suggests that the effects of climate change within African countries is evident (Hambira and Saarinen, 2015). Long-term projections in climate variables including climate proxies, precipitation, temperature, and evapotranspiration have been examined (Chishakwe *et al*., 2012; van Wilgen, 2016), and these demonstrate climate change is occurring within Sub-Saharan Africa (Malisawa and Rautenbach, 2012: Mushawemhuka *et al.*, 2021).

Air pollution decreases the air quality in specific tourist destinations, thus affecting tourist attraction (Becken *et al.,* 2017). Sea level rises will occur due to the impact of climate change (Hoogendoorn and Fitchett, 2018; Hambira and Mbaiwa, 2020). This will be a challenge to most African developing nations since they already have a limited adaptive ability to climate change

and tourism is not a primary apprehension for climate change adaptation (Awuor *et al*., 2008; Snoussi *et al*., 2008; Hambira and Mbaiwa, 2020). According to research, the majority of tourists are more concerned about sea-level rises than they are about lodging or service quality (Lukoseviciute and Panagopoulos, 2021). The Egyptian government, for example, even had discussions surrounding their degree of readiness for sea level increases, as there are a few historical sites that would be impacted if the rise in sea levels continue (El-Masry *et al*., 2022).

The temperature increases linked with an increase in GHGs are the primary focus of climate change science and the major point of many predictions (Yong, 2021). Such temperature increases may adversely affect tourism, both directly and indirectly (Dube, 2022). Thermal discomfort of visitors occurs when temperatures surpass human amenity limits (Gössling and Schumacher, 2010; Hunt and Watkiss, 2011). Effects that interact, with varied degrees of severity and influence, incorporate the impact of thermal fluctuations on the natural environment components that compose the tourist attraction (Hambira *et al*., 2013; Proebstl-Haider *et al.,* 2021). Because of escalated severity of temperature rises with latitude, the allocation of characteristics affected by temperature, and the varied dependence of tourism on the natural environment, the severity of the danger of temperature increases to tourism must vary spatially (Rogerson and Sims, 2012; Hoogendoorn and Fitchett, 2018).

Climate change is expected to cause temperature variations in various African nations (Hoogendoorn and Fitchett, 2018). The consequence of these temperature rises, however, will make these locations less desirable (Gössling *et al*., 2006; Gössling and Schumacher, 2010). Implications are consistent storms that can result in flooding. Warm temperatures are not considered a concern for visitors in Zanzibar, Tanzania, despite typical maximum temperatures of 32°C (Gössling *et al*., 2006). Storms and humidity, conversely, are highlighted as climatic factors

of apprehension (Gössling *et al*., 2006). Temperature increase will, however, result in a direct reduction in tourist visits in certain nations (Seetanah and Fauzel, 2018). This may be shown in Botswana, where temperature increase may not have as many consequences as in Tanzania, but it will immediately lead to a drop in outdoor activities such as boat and canoe riding (Köberl *et al*., 2016).

Bird observation in Egypt might be hampered by algae growth around dams (Ahmed and Hefny, 2007; Hoogendoorn and Fitchett, 2018). Coral bleaching caused by increasing ocean temperatures is of great concern (Ahmed and Hefny, 2007; Marshall *et al*., 2011; Satyawan *et al.,* 2021). Coral bleaching is believed to be a major challenge to tourism linked with snorkelling and scuba diving in the Seychelles (Marshall *et al*., 2011; Hoogendoorn and Fitchett, 2018).

Rainfall will be enhanced in certain regions and decreases in precipitation in other areas will be experienced (Dube *et al.,* 2020) In either case, the consequence will be that outdoor tourist activities will be limited owing to excess water, or water-based tourism activities would be restricted due to low water levels (Preston-Whyte and Watson, 2005; Mbaiwa and Mmopelwa, 2009; Hoogendoorn and Fitchett, 2019; Dube *et al.,* 2020). Increased flooding activity in the Okavango Delta region might cause native crocodiles and hippopotamus to move deeper into tourism-occupied regions (Hambira and Saarinen, 2015; Hambira and Mbaiwa, 2020). These modifications will have a number of unintended consequences for tourism growth in the region (Mbaiwa and Mmopelwa, 2009; Hoogendoorn and Fitchett, 2019). Decreasing water levels in the Okavango Delta will have an impact on visitor options since water-based tourism activities would be considerably reduced (Hambira and Saarinen, 2015; Dube and Nhamo, 2020).

### 2.4.3 South Africa

The South African climate is also quite diverse, with summer and winter rainfall zones, sub-tropical to mid-latitude regions, and both humid and desert regions (Fitchett *et al.,* 2017). Additionally, South Africa is anticipated to suffer abnormally large reparations as a result of climate change, attributable to the increased intensity of modelled effects and the reduced ability to adjust to changing conditions caused by more pressing developmental requirements (Reid and Vogel, 2006; Fitchett *et al.,* 2017). Thus, the viability of the tourism sector in South African is dependent on the kind and degree of climate change, as well as the tourism sector and the adaptation capabilities of the country (Rogerson, 2016). Over 25% of the South African population is unemployed, but tourism has aided in the employment of 655,609 people in 2013 (Fitchett *et al.,* 2017). South African tourism is also linked to local economic development strategies aimed at rescuing South Africans from the confines of poverty imposed by the apartheid past of the country (Rogerson, 2002; Fitchett *et al.,* 2017). Due to a disproportionate emphasis on outdoor attractions and a slow pace of climate change adaptation, the South African tourism sector is vulnerable to the impacts of climate change (Tanrisever et al., 2024). These effects have the potential to have a significant negative impact (Fitchett *et al.,* 2016).

South Africa as a developing country is trapped in between two realities (Amusan and Olutola, 2017). Internationally South Africa is recognised to be among the hardest afflicted nations by the consequences of climate change and is also ranked as a contributor to climate change (Scott *et al.,* 2019). The country now ranks as being one of the top 15 GHG emitters and has a record of generating high emissions (Scott *et al.,* 2019). On the contrary, tourism is one of the major sectors that fuels the economy of the country. In Sub-Saharan Africa, South Africa is recognised as the country with the top tourism locations (World Economic Forum - Travel and Tourism

Competitiveness Report (WEF-TTCR), 2015). This is largely credited to the country drawing a huge number of tourists globally owing to its fascinating tourism assets which include a broad range of cultural heritage sites (Hall, 2016).

The South African tourism sector is placed as fourth internationally owing to its quick development due to its attractive tourism assets (Akinboade and Braimoh, 2010; Amusan and Olutola, 2017). There are various tourist attractions within the tourist sector such as the Cape Winelands, wildlife reserves, and diverse adventure activities all contribute to South Africa achieving the position of being the fourth largest and most rapid tourism sector globally (Braimoh and Akindoabe, 2010; Rogerson, 2016). Developing economies are more dependent on areas such as tourism as compared to established economies (Khan *et al.,* 2020). These economies demonstrated better resilience during the 2008-2009 global crisis, tourist sector within emerging countries in Sub-Saharan Africa showed stronger resilience (Amusan and Olutola, 2017). South Africa as a geopolitical force in sub-Saharan Africa needs to further advocate for work to be done to prevent climate change.

According to Perch-Nielsen (2010), beach tourist sector of South Africa is relatively sensitive to climate change. This suggests that South Africa is considered vulnerable to climate change, having a limited adaptation potential but a favourable susceptibility to climate change (Phillips *et al*., 2018). It implies that climate change will have an impact on the beach tourist sector of South Africa, but other regions across the world will be impacted far worse (Gössling *et al*., 2002; Hall, 2008; Gössling *et al*., 2012b; Scott *et al*., 2012a; Santos-Lacueva *et al*., 2017; Phillips et al., 2018).

Scarcity of water, as a result of climate change (Gössling *et al*., 2012b; Saarinen *et al*., 2012; Kaján and Saarinen, 2013), may have an impact on beach tourism in both the west and southern regions of South Africa (Cai *et al*., 2014). Serdeczny *et al*. (2017) believe that owing to the reduction in

precipitation, South Africa will be more vulnerable to droughts in the future. However, according to Fitchett *et al*. (2016a), under a conservative forecast, Cape Saint (St.) Francis would lose the whole beach by 2100, and any lodging enterprises erected near the coastline will be at significant danger of floods and storms. Climate change will have a detrimental environmental impact on Western Cape Province in South Africa (Dube *et al*., 2020). Wine farming will suffer as a result, as will tourism activity that is supported by this farming (Dube *et al*., 2020). The wine industry in South Africa is a major tourist attraction (Steyn, 2012).

Climate change functions in a fundamentally distinct manner from other conventional global issues but has major influence not just on one component but on various features of the world such as transportation, ecology, agriculture, forestry, and energy use (Hoogendoorn and Fitchett, 2018). Human endeavours such as tourism are among of the aspects most influenced by climate change. Sufficient research has concentrated mostly on the consequences of climate change on the environment with important conclusions by the IPCC (2021). However, the understanding concerning the interrelations that exist between tourism and climate change inside South Africa are still scarce (Pandy and Rogerson, 2021).

## 2.5 Big Data and Tourism

### 2.5.1 The Application of Big Data in Tourism

In the largely digitalised society of today, tourists produce and consume big data, through interacting, purchasing, and exchanging ideas, photos and experiences online (Li *et al*., 2020). Technological devices such as phones, sensors, tracking systems (Geographical Positioning System), and cameras that are integrated into human everyday life, become elements of data

generation and distribution through internet technology (Dimitrios *et al*., 2022). The implications of such data being easily accessible is that organisations now have access to millions of zettabytes of data pertaining to consumers, suppliers, company processes, and competitors (Dimitrios *et al*., 2022). The quantity of this data is created instantaneously, accumulates fast, without a defined structure, frequently without any homogeneity and is characterised in the phrase "Big Data" (De Mauro *et al*., 2015). The phrase “Big Data” alludes to ‘‘data sets whose volume is exceeding the capacity to interpret, collect, analyse and visualise traditional data, databases and software tools” (Padgavankar and Gupta, 2014, p.2219).

The Big Data paradigm is often represented in the three Vs which symbolise certain features and following the technological and scientific breakthroughs, are enhanced with additional aspects. The "core" of big data comprises of three essential parts (Chen *et al*, 2012; Buchholtz *et al*., 2014) which according to the extant research are:

• Velocity, which is connected to the methods used for data flow and their production pace.

• Volume, which refers to the vast quantity of this data, which needs cleaning, compaction, or circumstances for its data storage.

• Variety, which represents the variability of datasets and their layout, both in terms of the varieties in which they occur and in terms of their semantics.

Previously, the emphasis was on organised data that fit neatly in tables or relational databases, such as sales by product or by region (Li *et al*., 2020). Today, however, 80% of the data created is unstructured and thus cannot readily be arranged in tables (e.g., images, video files, or social media updates). Advancements in data generation and collecting have led to substantial innovations in data analysis and storage technology (Dimitrios *et al*., 2022). The usage of Big Data is not confined to business but also expands to other sectors such as health and governmental institutions

(Gandomi and Haider, 2015). Hence, it is acknowledged that Big Data may be applied on any forms of processes pertaining to various sectors, including the tourism sector (Dimitrios *et al*., 2021). The application of big data within the tourism sector is probably the simplest approach to better understand the interests and aspirations of the tourist (Li *et al*., 2018). In a sector that is dominated by an anthropocentric orientation with businesses concentrating, monitoring, and analysing Internet reviews left by the tourist before, during and after the end of the vacation experience may guide and enhance commercial activities (Dimitrios *et al*., 2020).

Big Data may also aid in making choices to design and supply better services, meet the demands of visitors, and even produce bigger profits (Chareyron *et al*., 2014). According to Li *et al*. (2018), Big Data in tourism seems to be in the initial stage, irrespective of the fact that tourism is one of the industries where Big Data might offer strategic advantage. While Mariani (2019) argues that the tourism sector is gradually embracing Big Data, yet there is a great deal of room for improvement in the application and comprehension of Big Data, particularly in Big Data analytics. Indeed, there is a rising usage of Big Data in the tourism business and some relevant research, the academics and practitioners are far from claiming that they have completely comprehended the application of Big Data in the tourist sector (Centobelli and Ndou, 2019). From what is discussed from previous researchers there is a need to better grasp the idea of the use of Big Data in tourism but also to better propose recommendations for future study because there is a research deficit on this problem.

### 2.5.2 Tourism and Social Media

Social media is a term that refers to a collection of various forms of online web 3.0 peer communications ranging from online content communities, social networking, consumer review

sites, and other forms of technological data and tools that allow for the exchange and creation of user generated content (Kaplan and Haenlem, 2010; Vassakis *et al*., 2019). Social networking is widely recognised as a crucial trend impacting the tourism sector (Xiang and Gretzel, 2010; Vassakis *et al*., 2019). Tourists all around the globe have embraced social media, allowing them to serve as media in different consumer-to-consumer interactions such as reviewing hotel room services, rating lodging amenities, and sharing thoughts on the quality of various tourist sites (Leung *et al*., 2013).

#### *2.5.2.1 Online Review Platforms*

Online reviews are becoming increasingly essential in tourism research (Höpken, et al., 2024). They represent tourists' satisfaction and dissatisfaction with their experiences (Banerjee and Chua, 2016). Online reviews, which are spontaneous and insightful comments supplied by tourists and are publicly available and accessible on review sites (Guo *et al*., 2017), are a legitimate blend of facts, perspectives, perceptions, and emotions (Wilson *et al.*, 2012; Q.; Ye *et al.*, 2014). Online reviews serve various purposes: they reflect the users' feelings, depict real-life experiences, offer suggestions, and give essential information (Ye et al., 2014). In the sense that users of the platform broadcast their material to other users. Online reviews are a valuable tool that may help a resort strengthen its competitive edge (Simeon *et al*., 2017).

Reviews may influence service provider performance (Ganzaroli *et al.*, 2017) as well as perceived value (Rodrguez-Daz and Espino-Rodrguez, 2018), and appearance (Barreda and Bilgihan, 2013; Wong and Qi, 2017). Users also influence tourist decisions, including their travel destination (Zhang *et al.*, 2010), possibility of reservation (Vermeulen and Seegers, 2009), loyalty (Gavilan *et al.*, 2018; Litvin *et al.*, 2008), and intention to recommend (Berezina *et al.*, 2016). Consumer-

generated media and online review platforms may be classified into two types: transaction-based online travel agencies and community-based websites (Xiang et al., 2017). Different characteristics of various tourist sites are documented on online review platforms such as LonelyPlanet, TripAdvisor, and Yelp based on a mix of user data, travel forums, and information tools entered on these platforms (Li *et al.,* 2018). Transaction-based systems, on the contrary, place a greater emphasis on electronic word-of-mouth exchanges through websites such as Booking.com and Expedia (Li *et al.,* 2018). This mode of communication is mostly concerned with the financial elements of tourism (Xiang *et al*., 2017). Variations in data sources for online reviews should be evaluated for the quality, accuracy, and representativeness of data in social media research (Xiang *et al.,* 2017). For example, collecting unstructured data from websites and applications such as X and Facebook are more impractical since it is unstructured, as opposed to structured data from online travel platforms such as TripAdvisor, which is more practical. User produced content refers to any type of data (e.g. videos, voice notes, text, and photographs) provided by various online users on various types of social media such as consumer review sites, discussion forums, and social networking sites (Kaplan and Haenlein, 2010; Li *et al.,* 2018). Social media networking services such as Facebook and X allow semi-private and public users to interact by sharing diverse personal interests (Bruns and Highfield, 2015). Most tourism firms, however, place a greater emphasis on internet discussion platforms such as TripAdvisor, where people with diverse shared interests express sentiment and personal experiences on tourist destination places (Schuckert *et al.,* 2015). Tourist reviews offer valuable insights into perceptions of weather conditions at specific destinations, which can be obtained through sentiment analysis to understand the impact of climate tourism (Shi *et al.,* 2024). By analysing the emotional tone, keywords, and recurring themes in reviews, researchers can identify patterns in how weather influences tourists' experiences and

overall satisfaction. Shi et al. (2024) demonstrated the potential of using social media data and sentiment analysis tools to study visitors' sentiments toward environmental factors, including weather, at urban river corridors. Such analysis can inform tourism operators about seasonal preferences, guide strategic planning for climate adaptation, and improve marketing efforts to match tourists' expectations regarding weather-related factors.

#### *2.5.2.2 Social Media Analytics*

Online reviews and electronic word-of-mouth (eWOM) have attracted attention from tourism researchers over the years (Dwivedi *et al*., 2020). The internet offers several intriguing options for gaining better knowledge of customer behaviour (Han and Anderson, 2021). TripAdvisor is widely regarded as the premier information and advising platform of the world for tourist decisions, constituting the biggest tourist network in the world (Law, 2006; Litvin and Dowling, 2018; Fitchett and Hoogendoorn, 2018). The number of evaluated reviews might vary but range from XX-YY. Statistical, and machine learning approaches such as Latent Dirichlet Allocation (LDA), naive Bayes classification, and neural network models have all been employed by the researchers.

**Table 2. 1:** Literature on Text Mining and Analysis of Online Reviews (After Taecharungroj and Mathayomchan, 2019)

| Authors | Purpose | Context | Platform | Methodology |
|---|---|---|---|---|
| **Fitchett and Hoogendoorn (2019a)** | Analysing tourist sensitivity to weather conditions | 18 locations in South Africa | 5,898 TripAdvisor reviews | Content analysis |
| **Bi *et al*. (2019)** | To conduct importance-performance analysis (IPA) | Two Five-Star Hotels, add country | 24,276 reviews on TripAdvisor | Latent Dirichlet allocation (LDA), the improved one-vs-one strategy based support vector machine (IOVO-SVM) and the ensemble neural network based model (ENNM) |
| **Cheng *et al.* (2019)** | To investigate the effect of online review contents on potential guests' trust perception | Lodge listings in New York City, US | 1,485 and 10,000 reviews on Airbnb | Manual content analysis and convolutional neural network (CNN) model |
| **Fitchett and Hoogendoorn (2019b)** | Aimed to investigate tourist climatic sensitivity and experiences related to climatic conditions and weather | Afriski Resort, Lesotho | 66 out of 259 TripAdvisor reviews | Content analysis |
| **Kirilenko *et al*., (2019)** | To identify attraction clusters | Attractions in Florida, US | 157,285 reviews on TripAdvisor | Network analysis using DIRT_LPAwb + algorithm |
| **Alrawadieh and Law (2019)** | Investigating the factors that determine hotel guest satisfaction | Best rated hotels in Istanbul, Turkey | 400 English TripAdvisor reviews | Text mining |
| **Taecharungroj (2019)** | To infer the possible place brand identities from user-generated content | 2 famous metropolitan areas in Bangkok, Thailand | 9,633 reviews on TripAdvisor and Google Maps | Content analysis using Leximancer |
| **Taecharungroj and Mathayomchan (2019)** | To develop a methodology that can analyse online reviews using machine learning techniques in such a way that practitioners in the fields of tourism and destination management can understand and apply the technique to improve their attractions | Tourist attractions, including beaches, islands, temples, a pedestrian street, and markets in Phuket, Thailand | 65,079 TripAdvisor online reviews | Machine learning techniques: latent Dirichlet allocation (LDA) and naïve Bayes modelling |
| **Fitchett *et al.* (2020)** | To explore tourist sensitivity and satisfaction with regards to weather and climate in 5 Tropical Indian Ocean Islands | Five Tropical Indian Ocean islands: Mauritius, Reunion, Seychelles, Maldives, and The Comoros (Hotels) | 13,618 TripAdvisor online reviews | Content Analysis |

| | | | | |
|---|---|---|---|---|
| **Sangkaew and Zhu (2020)** | Comprehending tourists' experiences within a local market | Phuket, Thailand | 2,934 TripAdvisor online reviews | Sentiment analysis |
| **Lee *et al* (2020)** | Investigating the mentions behind each review posted with regards to tourist attractions | Five Gatlinburg tourist attraction in Tennessee, United States of America (USA) | TripAdvisor online reviews (number not mentioned) | Polarity analysis |
| **Wang et al. (2020)** | Examining factors influencing tourist satisfaction within hotels | Hotels in China | 37,146 reviews from Ctrip.com | Natural Language Processing |
| **Nilashi *et al*. (2021)** | Investigates the effectiveness of a hybrid method using clustering, Higher-Order Singular Value Decomposition (HOSVD) and Classification and Regression Trees (CART) in analysing tourists' online reviews | Malaysian spa hotels | Reviews on TripAdvisor (Number of reviews not stated in article) | Higher-Order Singular Value Decomposition (HOSVD) and Classification and Regression Trees (CART) |

Various studies have applied the use of online reviews to study climate impact on tourism in southern Africa (Fitchett and Hoogendoorn, 2018, 2019; Stockigt *et al.*, 2018; Fitchett *et al.*, 2020; Dube and Nhamo, 2020). Any information presented on a travel website might be a source of data for tourism studies. Despite the allure of these prospects, the reality is that manually gathering a huge volume of internet data is tedious and nearly impossible to perform objectively and consistently (Han and Anderson, 2021). A conventional tourism research analysis may require data collection from a large number of accommodation establishments and restaurants across a variety of markets over an extended period of time (Han and Anderson, 2021). Manually gathering this data may be extremely time consuming and laborious (Han and Anderson, 2021). As a result, it is of importance for the process of collecting and storing data from travel websites to be automated.

### 2.5.3 Methods Used to Collect Data in Tourism Research

Previous studies directly collected data in a simulation lab or survey scenario where study participants pretended to complete internet activities (Min *et al*., 2015; Choi *et al*., 2017; Shin *et al*., 2019). A typical criticism levelled at this research is that there is a disparity between real consumer behaviour and the behaviour of the volunteers in these tests (Shin et al., 2019). Furthermore, while large-scale survey research is feasible, a variety of impediments make it a challenge (Wagner et al., 2020). In the interim, the range of online scraped data is often considerable (Han and Anderson, 2021). Obtaining currently accessible web data is less expensive than conducting a survey, even when the commercial service price is considered (Han and Anderson, 2021). The monetary and labour costs of conducting a survey rise continuously with each new sample, but web-scraping is a one-time price (Juszczak, 2023). Hotel costs fluctuate on a regular basis, and internet evaluations are updated even more often (Landers *et al.*, 2016). A single sampling may not be adequate in such a volatile market (Japec

et al., 2015). Multiple surveys must be conducted sequentially, which is costly and time-consuming, but gathering publicly available internet data allows researchers to extract as much data as they need in real time (Japec *et al.*, 2015).

Web-scraping is the process of developing computer software that will automatically collect, parse, and arrange data from the web (van den Broucke and Baesens, 2018). Web-scraping simply put aims to systematically capture data from various user websites and transfer it to a database, for storage purposes either in an unstructured or structured format (Mitchell, 2018). Generally, web-scraping may be carried out by building an automated application known as a virtual scraper (Mitchell, 2018). Various studies have applied the technique of web-scraping within the field of tourism to investigate tourists' reviews on hotel service, transportation or negative word mentions (Levy *et al*., 2013; Siew *et al.*, 2018; Ahmad *et al.*, 2020). Two studies (Terziyska and Damyanova, 2020; Barbierato *et al.*, 2021) focused on examining the key elements that are needed to make wine tourism a success in Italy utilised the application of web-scraping techniques such as text mining and sentiment analysis. The limitations faced however included issues with the data utilised, however the methods applied were more effective as compared to traditional surveys (Terziyska and Damyanova, 2020; Barbierato *et al*., 2021).

The application of web-scraping in the field of tourism biometeorology has nonetheless been on the rise of late (Saarinen *et al*., 2022). Some studies have employed the usage of this approach to automate data from online pages such as Twitter to examine climate change related research (Al-Saqaf and Berglez, 2019; Cecinati *et al*., 2019). However, the application of web-scraping within the field of tourism and climate change globally and within the southern part of Africa remains scarce (Saarinen *et al*., 2022).

## 2.6 Conclusion

This chapter demonstrates that tourism is an essential sector for developing economies all around the world. There is widespread accord among tourism researchers that tourists are sensitive to both climate and weather. The climate of a place is thought to impact a tourist's scheduling, destination choices and overall satisfaction of the trip. The climatic suitability of regions for tourism is frequently assessed using various climatic indices. However, the use of indices brings about disadvantages such as discounting the effects of tourists' native countries, expected weather conditions, and infrastructure in tourist accommodation businesses. This study aims to investigate tourists' experiences, perceptions, and sensitivity to weather and climatic conditions during vacations within 18 locations in South Africa through the application of big data through the analysis of TripAdvisor Reviews. Considering the significance of weather as a pull factor for tourism and in establishing the destination image, reports of adverse weather serve a dual purpose. This might enable for better-informed and hence more probable successful adaptation on the part of both tourists and the government. Furthermore, data might benefit in alerting future guests about the weather encountered by past visitors, especially those from the same region of origin, providing improved awareness among prospective tourists and hence a lower risk of disappointment.

# Chapter 3: Study Site

## 3.1 Study Site

### 3.1.1 Introduction

The study is mainly concentrated on the application of web-scraping to automate data concerning TripAdvisor reviews throughout 19 different locations. In the case of this aspect being achieved, the study will be expanded to the rest of the accommodation establishments throughout South Africa. This chapter presents an overview of the climate, weather, tourism and geographical features of the study region.

## 3.2 Descriptive Geography of South Africa

The 19 destinations include Polokwane, Pilanesberg National Park, Pretoria, Belfast, Nelspruit (Mbombela), Johannesburg, Kimberly, Bethlehem, Ladysmith, St Lucia, East London, St Francis Bay, Durban, Bloemfontein, Cape Town, Paarl, Knysna, Port Elizabeth (Gqeberha), and Port Nolloth to facilitate comparison with studies that have performed manual capturing of TripAdvisor reviews (Fitchett and Hoogendoorn, 2018; Figure 3.1). The surface area of South Africa is 1,219,602 $km^2$ and extends latitudinally from 22-35° South and longitudinally from 17-33° East (Figure 3.1). It is classed as semi-arid as two-thirds of the nation receives less than 500mm of rainfall each year (Van Heerden *et al.*, 1988; Tyson and Preston-Whyte, 2000). Directly influenced on the east by the Agulhas current and the warm Indian Ocean, and on the west by the Benguela current and the cold Atlantic Ocean, the mean annual rainfall of South Africa decreases from east to west (Jury, 2013). The South African climate varies greatly, with summer temperatures exceeding 32°C and winter temperatures several degrees below zero on the interior plateau (Jawtusch, 2014; Fitchett *et al.*, 2017). Notwithstanding these

differences, South Africa is deemed to have exceptionally favourable climatic conditions for tourism, and the climate of the country is a key appeal (Preston-Whyte and Watson, 2005; Fitchett *et al.*, 2017). Feasibly, the pleasant climate contributes significantly to the competitive position of South Africa as a tourism destination (Anholt, 2007; Fitchett *et al.*, 2017).

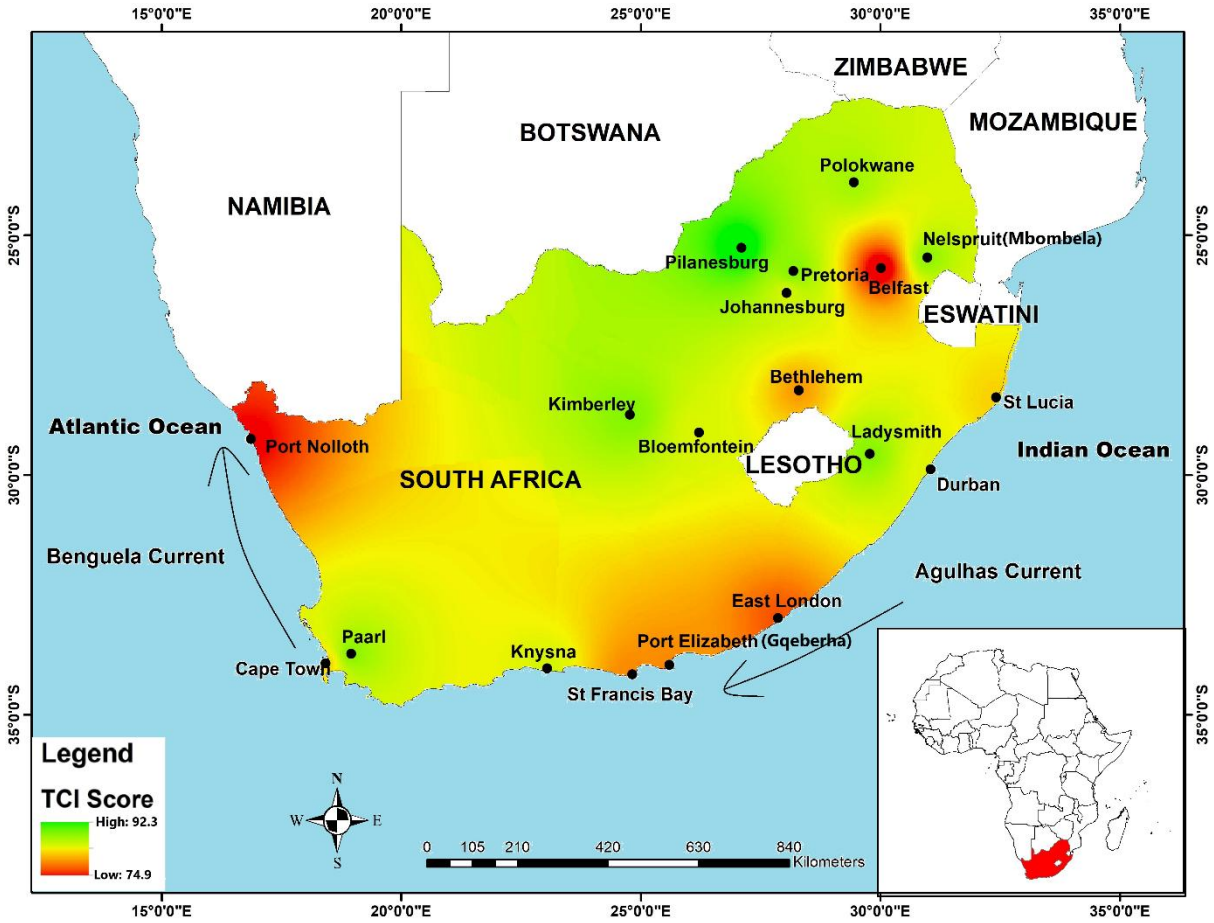



**Figure 3. 1:** Map displaying the various locations for which TripAdvisor reviews were retrieved with the TCI scores (Fitchett et al., 2017) interpolated using Inverse Distance Weight (IDW).

## 3.3 Climate and Weather

The climatic and meteorological attributes of South Africa are thoroughly documented (Taljaard, 1996; Tyson and Preston-Whyte, 2000; Landman *et al.*, 2017; Lennard, 2019). The

climate of South Africa can be broadly categorised as wet, humid, and warm in the east, to arid, hot, and dry in the west (Landman *et al.*, 2017; Lennard, 2019). Due to the latitudinal position of South Africa predominantly in the subtropics, the nation is affected by the seasonal movement of the Southern Hemisphere Westerlies and the Intertropical Convergence Zone (ITCZ), making seasonality a significant factor in the country (Tyson and Preston-Whyte, 2000; Chase and Meadows, 2007; Jury, 2013). Temperatures nationwide are typically elevated in summer and diminished in winter (Landman *et al.*, 2017). Furthermore, the majority of the country experiences rainfall in summer, but the southwestern Cape and western coast receive rainfall in winter, and the southern coast experiences rainfall year-round (Schulze and Maharaj, 2007; Lennard and Hegerl, 2015). The climate and weather of South Africa are both temporally and spatially variable across a variety of scales and are influenced by the latitudinal location, topography, and adjacent oceans (Taljaard, 1996; Nicholson, 2000; Tyson and Preston-Whyte, 2000; Landman *et al.*, 2017; Jury, 2018; Lennard, 2019).

## 3.4. Tourism in South Africa

To facilitate comparison of the previously published manually analysed TripAdvisor reviews with those from web-scraping, this study will concentrate on the same 19 locations in South Africa for which a manual TripAdvisor analysis of climate mentions has been conducted (Fitchett and Hoogendoorn, 2018, 2020; Figure 3.1). These locations span the nine provinces of South Africa, the three rainfall zones, and offer a range of different tourist attractions (Table 3.1). The South African tourist sector consists of various forms of tourism markets such as sports, wine, cultural, business, adventure, coastal and nature-based tourism. Each of these tourist destinations or hotels under study have unique climates and weather associated with them which make them ideal for evaluating the effect of weather and climate on tourism as a whole (Fitchett et al., 2017). The diversity and type of tourism offerings at each region is

specific and diverse. Cities such as Cape Town, Johannesburg and Durban serve as attraction sites for business tourism as well as provide various tourist activities such as museum tours and botanical gardens. Hence all these regions have tourist attractions that are sensitive to weather and climate. The tourist experiences vary in relation to weather, throughout each location. Climate and weather are of importance within tourist attractions, as there is an inter-dependence that exists between climate and tourism (Kozak et al., 2008; Moreno et al., 2008; Rutty and Scott, 2016).

**Table 3. 1:** Tourist sites

| **Tourist Segments** | **Tourist Locations** |
|---|---|
| Coastal | Port Elizabeth (Gqeberha)<br>Port Nolloth<br>St Francis Bay<br>Knysna<br>Durban<br>East London<br>Cape Town |
| Urban | Johannesburg<br>Pretoria<br>Cape Town<br>Durban<br>Kimberley<br>Bloemfontein |
| Rural and Countryside | Paarl<br>Polokwane<br>Ladysmith<br>East London<br>Knysna<br>St Francis Bay |
| Mountain | Bethlehem<br>Cape Town |
| Natural and Wildlife | Pilanesberg National Park<br>St Lucia<br>Knysna |
| Cultural and Historical | Nelspruit (Mbombela)<br>Johannesburg<br>Cape Town<br>Bethlehem |
| Adventure and Sports | Belfast<br>Knysna<br>Cape Town<br>Johannesburg |

South Africa offers a diverse range of tourist activities and attractions across these niche markets. The significance of tourism as a mechanism for economic and social growth in South

Africa necessitates the maintenance and preservation of tourist attractions to ensure their continued appeal to visitors. Additionally, South Africa should consistently seek to diversify its tourism market to attract a broader range of tourists. The South African climate is highly diverse, with summer and winter precipitation zones, subtropical and mid-latitude areas, as well as humid and arid zones (Nicholson, 2000; Fitchett *et al.*, 2017). South Africa is expected to suffer significant damages from climate change stemming from the high likelihood of projected impacts and the slow pace of adaptation as a result of other urgent specific interventions (Reid *et al.*, 2006; Fitchett *et al.*, 2017). Thus, the success of the tourism sector in South Africa is contingent on the characteristics and pace of climate change, as well as the extent of the tourism sector of the country to adapt in its entirety (Rogerson, 2016).

Tourism is recognised as an important factor to economic expansion in South Africa, consistently expanding at a rate that surpasses the national average (Henama, 2019). The sector generates foreign income and stimulates other economic sectors, resulting in the creation of jobs through direct, indirect, and induced influences (Akyürek et al., 2023). South African tourism is also linked to local economic development strategies aimed at rescuing South Africans from the confines of poverty imposed by the apartheid past of the country (Rogerson, 2002; Fitchett *et al.*, 2017). Due to a disproportionate emphasis on outdoor attractions and a slow pace of climate change adaptation, the tourism sector in South Africa is exposed to the full impacts of climate change. These effects have the potential to have a significant negative impact (Fitchett *et al.*, 2016b; Fitchett *et al.*, 2017).

# Chapter 4: Research Methodology

## 4.1 Introduction

This study applied Fitchett and Hoogendoorn's (2018, 2019) methodology for analysing visitor sensitivity to climate using TripAdvisor reviews to the same 19 tourist destinations in South Africa, using a web-scraping approach. In the case of this objective being achieved the study was then expanded to all accommodation establishments throughout South Africa. Sentiment analysis was then conducted to get the overall sentiments of tourists regarding the weather that they experienced at various tourist destinations. The main purpose of this research was to explore the possibilities of using web-scraping methods to enable a wider database of TripAdvisor reviews to be used in exploring the sensitivities of tourists to the weather. In doing this, all statistical tests conducted in the previous studies were replicated, which included the following: calculating and stating the overall number of reviews, the proportion of climatic/weather related mentions, exploring the relative climatic contribution of the climatic factors per location, validating it to the previously published manual results and lastly aggregating the various sentiments regarding weather mentions per location. In addition to this, a comparison of means test was conducted for each other components to determine if there is a significant difference between web-scraped and manually collated data.

### 4.1.1 Data Collection

To ensure consistency in the analysis and comparison of outcomes, this study followed the methodology provided by Fitchett and Hoogendoorn (2018; 2019) and Fitchett *et al.,* (2020) and implemented by Stockigt *et al.* (2018). Tripadvisor reviews may offer more objective responses, as reviewers may reflect on and discuss any aspects of their trip (Fitchett et al.,

2020). TripAdvisor provides an online forum for travellers to share feedback on their experiences, which has resulted in the spread of e-Word-of-Mouth recommendations to a broader audience (Choudhary and Sharma, 2022). For the purposes of this research, reviewer data such as total number of TripAdvisor reviews and TripAdvisor status of reviewers were not collected. Accommodation establishments were selected over TripAdvisor pages of attractions because they have the broadest and most consistent history of reviews on TripAdvisor. Secondly, an assessment of reviews of accommodation establishments enables direct comparisons to prior research undertaken in South Africa that focused on these establishments (Fitchett *et al.,* 2016). Accommodation establishments data from previous research which conducted manual counts on TripAdvisor reviews from 19 locations in South Africa ranging from backpackers and motels to five-star hotels was validated against the automated web-scraped data (Fitchett and Hoogendoorn, 2018; Stockigt *et al.,* 2018; Fitchett *et al.,* 2020). The five-star, four-star, three-star, two-star hotel, and Bed and Breakfast (BnB) or Inn (Amaral *et al.,* 2014; Fitchett and Hoogendoorn, 2018) are the classifications used by TripAdvisor for accommodation establishments. TripAdvisor reviews published in languages other than English were captured and evaluated independently from reviews posted in English. Dates of the review, the trip month (if dissimilar from the review date), and the country of origin of the individual who published the review were all recorded.

Reviews were captured for this research as well, working in reverse chronological order, across the three types of accommodation year by year from December 2022 (Amaral et al., 2014; Buzinde et al., 2010; Fitchett et al., 2020). This yielded a bigger and more detailed population sample. The basics of web-scraping were demonstrated using Python, a general-purpose programming language (Downey, 2014). This research does not introduce Python as a whole, but mainly the Python features essential to web-scraping. Web-scraping is a process or technique of extracting data from a website using computer software (Mahto and Singh, 2016;

Adhinugroho *et al.*, 2020). A web browser application or low-level Hypertext Transfer Protocol (HTTP) implementation is used to simulate data requests from the user's side (Mahto and Singh, 2016; Adhinugroho *et al.*, 2020). Web-scraping is the automated data collecting from many Web pages (Adhinugroho *et al.*, 2020). While web harvesting can be a convenient method for analysing large data sets, it has the potential to remove context from words or mislabel original data (Xiang *et al.*, 2018). To address this, the code was tested and revised to ensure that the website retrieves the correct data.

ChromeDriver is a Google Chrome extension that is downloaded. The Jupyter Notebook was downloaded and installed using the Anaconda prompt, along with the relevant Python libraries. Python was used for the purposes of this research, primarily owing of the large number of publicly accessible libraries that provide data extraction capability (Downey, 2014). The Integrated Development Environment (IDE) "Jupyter Notebook" was used throughout the development process because it features a Graphical User Interface (GUI) that facilitates coding and debugging. The primary libraries used to construct the code are NumPy, Pandas, PySimpleGUI, and Selenium. NumPy has a sizable library of high-level mathematical functions for manipulating big and multi-dimensional arrays (Mitchell, 2018). Pandas supports table manipulation data structures and operations (Mitchell, 2018). PySimpleGUI is a simple-to-use yet powerful tool for creating a straightforward graphical user interface (Mitchell, 2018). Selenium is used to implement the data scraping functionality (Mitchell, 2018). It is a robust framework for automating the browser, which is frequently used to test web applications. However, it is also effective for scraping data from websites due to its ability to more accurately imitate a user’s behaviour.

After successfully configuring the coding environment, the coding process began. The coding process involved the creation of lists which will aid in building the structure of the webpage. After that, a primary function was built based on the structure of the webpage and the logic

used to interact with it. In extracting the data, the code was specifically designed to extract the review date, title, location, rating, and content to determine the basic demographics pertaining to the sample. The data was then saved in a excel comma separated values (csv) format and was cleaned for further analysis.

The scraping of webpage content feature was then achieved. The tool "Inspect Elements" on Google Chrome was used to analyse the Hypertext Markup Language (HTML) code structure of the webpage and gain a thorough grasp of how to use the Cascading Style Sheets (CSS) selector to accurately and successfully scrape the intended data. After the coding phase was complete, testing was performed prior to delivery to ensure that there are no unexpected problems and that the proposed system is capable of performing the functions for which it was built. NumPy and Pandas was used to check for data duplication in the case of a vast volume of scraped data.

Once the data had been extracted from the TripAdvisor site through the use of the web-scrapper, the data was cleaned before analysis. The cleaning process consisted of the following:

- Removing dummy data
- Converting the data field to the correct sort of data
- Eradicating data redundancy by utilising the name, address, hotel id, and coordinates of the accommodation establishment
- Removal of inactive accommodation establishments

Sentiment Analysis was employed to evaluate the sentiment of the review, which included a variety of criteria for categorising the various lodging places as either very positive, positive, very negative, negative, or neutral. Upon achieving this objective, the same procedure was repeated in extracting more data throughout the whole of Southern Africa.

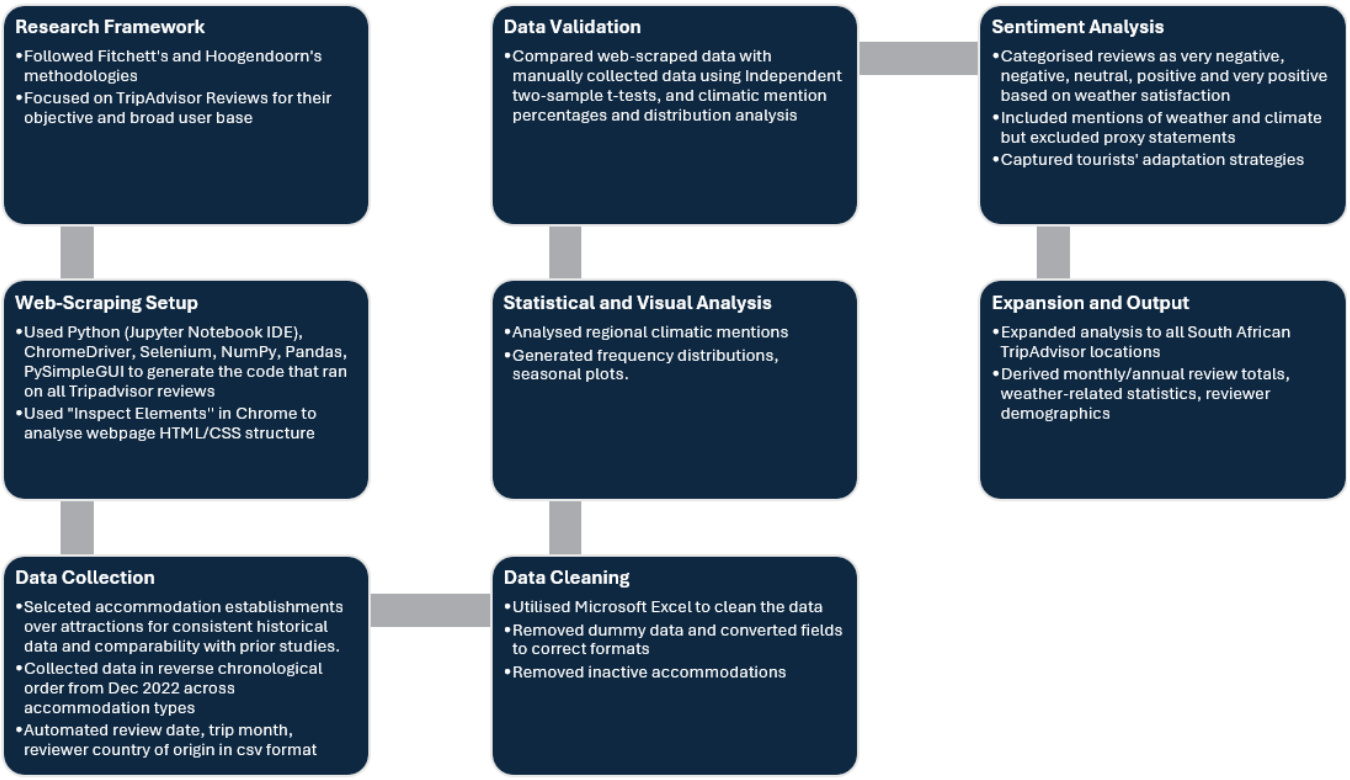



**Figure 4. 1:** Data processing flowchart

### 4.1.2 Data Analysis

A contextual method of content analysis was used to identify these climatic mentions. To be classified as a climate mention, material must include a comment about weather or climatic factors, including temperature, precipitation, wind, humidity, and cloudy/sunny circumstances (Hoogendoorn and Fitchett, 2018). Climate mentions did not include proxy statements such as the operation of air conditioning units, the lack of shade, or the sunset or sunrise due to the challenge of properly correlating these statements with a positive or negative weather experience encountered by the reviewer during their stay (O'Connor, 2010, Hoogendoorn and Fitchett, 2018).

In comparing the data, the time period used for the previous studies was applied even for the web-scraped data for accuracy purposes. To be considered a climate mention, text needed to contain information about climatic factors or weather, including temperature, rainfall, wind,

humidity, and cloudy or sunny circumstances. Climate references were categorised by the exact weather or climatic state, including heat, humidity, rain, wind, and storm, to determine the most relevant factors to visitors. The web-scraped data was analysed using the excel "if function" to identify reviews that contained mentions of climate or weather. Reviews were examined for recurring topics and predetermined terms using a content-based methodology such that if a review had a specified climate or weather mention the function would assign a value to it and then calculate the total amount of the mentions. The percentage of reviews mentioning climatic factors was determined, and the frequency distribution of climatic conditions mentioned was examined for each accommodation type and destination indicating the overall climatic sensitivity, the proportion of climatic mentions and the total climatic mentions per region. Finally, the statistics obtained were compared with the manually collected data presented in Fitchett and Hoogendoorn's (2019) TCI scores for each destination. The statistical tests included comparison of means test also known as an independent two-sample t-test. William Sealy Gosset introduced the t-test in 1908, and it is a widely used statistical method for comparing the means of two groups to determine whether they are significantly different (Jankowski et al., 2018). While employed as a statistician at the Guinness Brewery in Dublin, Ireland, Gosset devised the test. The t-test was particularly advantageous for small sample sizes, as it was more dependable than conventional methods such as the normal distribution (Mane, 2016).

The general formula for the t-statistic in a two-sample t-test is (Student, 1908):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{X}_1$ and $\bar{X}_2$ are the sample means of group 1 and 2.
- $s_1^2$ and $s_2^2$ are the sample variances of the two groups.
- $n_1$ and $n_2$ are the sample sizes of the two groups.

The t-test helps determine whether the difference between the means is statistically significant, based on the size of the samples and their variability (Mane, 2016). These tests were conducted based on the overall demographics and statistics of the reviewers, the proportion of climatic mentions, the overall climatic sensitivity, total climatic mentions per region to validate the results obtained from the web-scraped data relative to the results obtained from the manually collected data from Fitchett et al. (2019). Subsequent consideration was undertaken over the accuracy of manually acquired data in the classification of the climatic appropriateness of each place for tourism.

The frequency distributions of all climate mention and of each distinct category of climatic circumstances was studied, allowing for quantitative comparisons of overall count data within each group and qualitative analyses of inter-class heterogeneity (Fitchett *et al.*, 2018). With a particular emphasis on rainfall, the primary risk factor for beach tourism, the data was further studied through theme analysis and identification of important quotes. Statistical and visual comparisons, such as the t-test to check for correlations was done with previously manually captured data from TripAdvisor from previous studies specifically from the locations mentioned in this study. All climate references were classified according to the exact weather or climatic state, including heat, humidity, rain, wind, storm, and so on, to identify which climatic factor is predominantly relevant to visitors. The overall number of climatic mentions in each category was computed annually, monthly, and per accommodation establishment.

Once this has been validated, the study was expanded to all of the TripAdvisor locations in South Africa, mainly focusing on accommodation establishments to further comprehend the sensitivity of tourists to climate and weather. This involved scraping data up to the date of analysis which was the end of 2021. The same procedure used prior to fulfilling this objective was also utilised here. Once this was done, means were determined for all the TripAdvisor data captured. This data includes the aggregate total of reviews (per month and per year). Frequency distributions were utilised to provide a visual depiction of the outcomes. In the case of certain reviewers providing only a province or town that they reside in; this was changed to the country rather. Lists were generated which included the number of reviews for each location in South Africa. Other details included the mean rainfall, temperature, wind and humidity as well as other climate and weather-related mentions. From this data, monthly seasonal plots as well as choropleth maps to demonstrate the proportional representations of TripAdvisor reviews from each region, were then generated through the use of ArcGIS software.

Last, sentiment analysis was conducted. Sentiment analysis of tourist opinions is essential for the ability to interpret complex feedback, ensure accuracy, and generate context-rich insights. Manual sentiment analysis, in contrast to automated methods, enables researchers to understand and interpret the complex structure of human language, particularly in industries such as tourism, where emotions, expectations, and subjective experiences are notable factors. Subjective statements are frequently present in tourist feedback, which automated tools may misinterpret. Researchers can capture complexities, such as the distinction between a tourist's overall experience and specific instances associated with weather or service, through manual sentiment analysis. Human analysts are more proficient at identifying sarcasm, humour, or elaborate emotions that algorithms may find difficult to accurately classify (Lugmayr et al., 2017). Contextual understanding remains a challenge for Natural Language Processing (NLP)

tools, and this is particularly true in industries such as tourism, where experiences can be personal.

Upon the web scraped data being further analysed for reviews containing weather related mentions, each review that had climatic mentions were further analysed to evaluate the sentiment of the reviewer in relation to the interaction or experience regarding the weather. The reviews were further classified into five different categories namely (very negative, negative, neutral, positive and very positive), the reviews were classified as very negative if the weather experienced at the particular location caused extreme distress or dissatisfaction to the point of cancellation or recommending that people should not visit that particular location, negative classifications were assigned to dissatisfaction, such as discomfort caused by weather, neutral classifications were based on either no outright negative or positive sentiment was attributed to the weather experienced, however an experience was classified as positive if a reviewer expressed satisfaction, such as appreciating hot weather. Frequency distribution graphs and tables were generated further visualizing the results obtained. Furthermore, additional relevant data was recorded, including the adaptation strategies employed by tourists or accommodations, such as whether a hotel provided alternative activities during bad weather. This is crucial because it has the potential to influence the level of satisfaction that tourists experience with the weather. When tourists are incapable of adjusting to unfavorable weather factors, they are more likely to experience dissatisfaction (de Freitas et al., 2015; Gössling et al., 2016).

## 4.2 Ethics Clearance

Ethics clearance was obtained from the Witwatersrand Human Non-Medical Research Ethics Committee, clearance number HRECNM22-01-077. In automating the data from TripAdvisor,

screen names were not captured and were not reported on, however screen names can be pseudonyms.

# Chapter 5: Results

## 5.1 Introduction

The following chapter explores the results obtained from this research. The chapter is structured into four primary parts, whereby each section provides an exposition of the outcomes derived from the used methodological techniques. This research provides the findings derived from the development and implementation of a web-scraping tool specifically intended to automate the process of extracting reviews from the popular travel website, TripAdvisor, for the purpose of the analysis of climate sensitivities among tourists. The tool was developed with the aim of optimising the procedure of gathering user-generated reviews for the purpose of analysis. This study aims to assess the effectiveness of automating the analysis of TripAdvisor data in the fields of tourism and climate research. Specifically, the investigation focuses on comparing two distinct methodologies while examining reviews referenced in a sequential manner, adhering to chronological order, progressing retrospectively through successive years from 2019. In the third section, the sentiment analysis findings will be provided to provide a deeper understanding of the impact of weather and climate on visitors' perceptions. Subsequently, the research will further explore the expansion of the analysis on various hotel establishments throughout South Africa.

## 5.2 Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies

### 5.2.1 Sample Size and Demographics

The utilisation of web scraping techniques yielded a total of 26,687 TripAdvisor reviews pertaining to 168 tourist accommodation establishments located across 19 specified

destinations within South Africa. This quantity surpassed the manually collected reviews previously documented by Fitchett and Hoogendoorn (2018), which amounted to 5898, representing an increase of 352.48% (Table 5.1). Notably, South Africa accounted for the highest proportion of reviews obtained through web scraping, constituting 12,226 (45.81%, n=26,687) of the total reviewers identified. Subsequently, the United Kingdom and the United States of America followed, contributing 7.6% (n=26,687) and 7.4% (n=26,687) of the total reviews, respectively, out of the 26,687 reviews collected. Reviews sourced from Australia comprised 3.45% of the entire dataset obtained through web scraping. Furthermore, contributions from an additional 96 countries were documented.

Among the total reviews analysed, 2604 (44.15%) originated from individuals residing in South Africa from the manually collected data (Fitchett and Hoogendoorn, 2018). A minor percentage of reviewers (14.54%, n=5898) did not disclose their residential location in their TripAdvisor profiles (Fitchett and Hoogendoorn, 2018). The residual reviews encompass contributions from 96 different countries of origin (Fitchett and Hoogendoorn, 2018), with the predominant sources being residents of the United Kingdom (9.82%, n=5898), the United States of America (3.54, n=5898), and Australia (2.04%, n=5898).

A 1.64% increase, which was statistically insignificant, was noted in the total number of South African tourists visiting tourist destinations from the manual approach (Figure 5.1), whereas a 2.22% insignificant difference was also observed in the number of tourists coming from the United Kingdom between the web scraped and manual approaches (Figure 5.1). A statistically insignificant 1.41% increase was also denoted between both approaches in the reviewers coming Australia (Figure 5.1). More countries such as Argentina, Canada, France, Germany, Spain, China, and India have increased by more than 1% from the manual results. A total of 11.28% of the countries had no reviews which was consistent with the manual approach.

Countries that showed no change in the number of reviews obtained resulted in 26.15% (Figure 5.1).

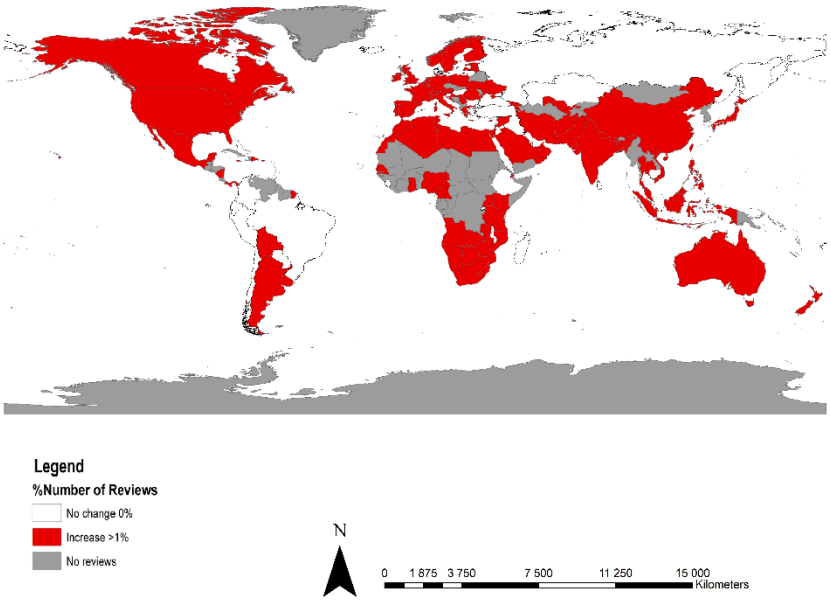



**Figure 5. 1:** Map illustrating the percentage differences in Tripadvisor review distributions gathered through manual and web scraping methods.

The highest percentage of reviews recorded was from the United Kingdom (7.6%, n=26,687), which has a statistically insignificant 2.22% difference from the manually recorded data (Table 5.1). United States followed after (7.4%, n=26,687), which yielded a 3.86% difference between both approaches (Table 5.1). The remainder of the assessments (displayed in Figure 5.1a) included Australia (3.45%, n=26,687) which showed a 1.41% insignificant difference between both approaches. South African reviewers contributed most of the reviews within the web scraped data, with over 45.81%, n=26,687 written by locals (Table 5.1), particularly from Gauteng (21.22%, n=12,226). These statistics indicate that there is a statistically insignificant difference (p=0.94) with those from the manually captured data which yielded 44.17%, n=5898

of South African reviewers. All these results indicate a statistically insignificant difference between both the approaches (Table 5.1).

**Table 5. 1:** Results pertaining to the comparison between web scraped and manual reviews from various countries.

| Countries of Origin | Total Number | | Percentage (%) | | Percentage difference (%) |
|---|---|---|---|---|---|
| | W | M | W | M | |
| **South Africa** | 12226 | 2604 | 45.81 | 44.17 | 1.64 |
| **United Kingdom** | 2015 | 579 | 7.6 | 9.82 | 2.22 |
| **United States of America** | 1966 | 209 | 7.4 | 3.54 | 3.86 |
| **Australia** | 921 | 120 | 3.45 | 2.04 | 1.41 |
| **T-test for the comparison of country percentages** | | | | | |
| ***p= 0.94*** | | | | | |

## 5.2.2 Climatic sensitivity of tourists

Among the 26, 687 TripAdvisor reviews, 1,163 mentioned the weather, equating to an average climatically sensitivity of 4.4% (Table 5.2). This is statistically different (p=0.014) from the manually collected data yielding an overall sensitivity of 7.9% (Table 5.3). Of the 1,163 mentions of weather in the web scraped TripAdvisor reviews, cold conditions were mentioned the most (35.68%, n=1,163), followed by rain (14,70%, n=1,163), which differed to the findings obtained by manually gathering the data where the cold (38.8%, n=494) was mentioned the most followed by hot (28.88%, n=494; Table 3).

**Table 5. 2:** Mentions of the weather in TripAdvisor reviews.

| Web scrapped and Manual Data | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tourist Destination** | **Number of Reviews Consulted** | | **Number of Mentions of Climate** | | **Percentage of Climate Mentions** | | **Most Frequently Mentioned Climate Factor** | | **Proportion within Climate Mentions** | | **P-value (Proportion of climate mentions)** |
| | **W** | **M** | **W** | **M** | **W** | **M** | **W** | **M** | **W** | **M** | |
| **Belfast** | 326 | 172 | 36 | 24 | 11% | 14% | Cold | Cold | 0.67 | 0.83 | *0.55* |
| **Bethlehem** | 1021 | 171 | 73 | 26 | 7% | 15% | Cold | Cold | 0.42 | 0.85 | *0.63* |
| **Bloemfontein** | 1624 | 470 | 71 | 26 | 4% | 6% | Cold | Cold | 0.45 | 0.62 | *0.36* |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cape Town** | 1448 | 604 | 58 | 33 | 4% | 6% | Hot | Hot | 0.19 | 0.36 | *0.55* |
| **Durban** | 1260 | 304 | 39 | 25 | 3% | 8% | Cold | Sun | 0.54 | 0.24 | *0.44* |
| **East London** | 3777 | 462 | 92 | 25 | 2% | 5% | Cold | Hot | 0.25 | 0.36 | *0.02* |
| **Johannesburg** | 1731 | 345 | 52 | 23 | 3% | 7% | Hot | Hot | 0.31 | 0.35 | *0.898* |
| **Kimberly** | 808 | 360 | 31 | 29 | 4% | 8% | Cold | Hot | 0.45 | 0.59 | *0.93* |
| **Knysna** | 2853 | 603 | 141 | 32 | 5% | 5% | Cold | Cold | 0.35 | 0.41 | *0.01* |
| **Ladysmith** | 883 | 179 | 57 | 25 | 6% | 4% | Cold | Cold | 0.47 | 0.56 | *0.37* |
| **Nelspruit (Mbombela)** | 1817 | 251 | 43 | 25 | 3% | 10% | Cold | Cold | 0.51 | 0.40 | *0.1* |
| **Paarl** | 1285 | 202 | 65 | 25 | 5% | 12% | Cold | Hot | 0.26 | 0.36 | *0.16* |
| **Pilanesberg** | 1749 | 327 | 118 | 28 | 7% | 9% | Cold | Cold | 0.42 | 0.43 | *0.10* |
| **Polokwane** | 733 | 249 | 22 | 16 | 3% | 6% | Hot | Hot | 0.41 | 0.63 | *0.02* |
| **Port Elizabeth (Gqeberha)** | 1279 | 423 | 46 | 20 | 4% | 5% | Cold | Cold | 0.43 | 0.55 | *0.304* |
| **Port Nolloth** | 149 | 47 | 21 | 12 | 14% | 26% | Hot | Cold | 0.48 | 0.50 | *0.56* |
| **Pretoria** | 830 | 299 | 56 | 25 | 7% | 8% | Cold | Hot | 0.43 | 0.52 | *0.76* |
| **St Lucia** | 2353 | 273 | 70 | 25 | 3% | 9% | Rain | Hot, rain | 0.26 | 0.32 | *0.53* |
| **St Francis** | 761 | 157 | 67 | 20 | 9% | 13% | Cold | Cold | 0.33 | 0.50 | *0.17* |
| **Total** | **26678** | 5898 | **1163** | 464 | **4.4%** | 7.9% | **Cold** | Cold | **0.37** | 0.39 | |

There is a statistically insignificant difference (p=0.72) between the mentions of cold mentions in the web scraped and manual data, whereas hot climatic mentions yielded a p-value of 0.05, indicating statistical insignificance between the two approaches.

**Table 5. 3:** Statistical analysis performed between the various climatic mentions between both datasets.

| Climatic Condition | Manual (%) | Web-scrapped (%) | Percentage difference (%) | output-value |
|---|---|---|---|---|
| **Hot** | 26.60 | 18.34 | 8.26 | *0.05* |
| **Cold** | 39.70 | 37.54 | 2.16 | *0.72* |
| **Rain** | 12.43 | 13.14 | 0.71 | *0.68* |
| **Bad weather** | 2.87 | 5.89 | 3.02 | *0.05* |
| **Good weather** | 2.37 | 4.92 | 2.55 | *0.05* |
| **Mist** | 1.28 | 0.16 | 1.12 | *0.78* |
| **Sun** | 11.57 | 10.89 | 0.68 | *0.45* |
| **Humid** | 2.06 | 1.17 | 0.89 | 0.76 |
| **Drought** | 2.9 | 2.47 | 0.43 | *0.80* |
| **Cloud** | 0.8 | 2.93 | 2.13 | *0.06* |
| **Wind** | 2.35 | 2.77 | 0.42 | *0.78* |

| | | | | |
|---|---|---|---|---|
| **Overall climatic sensitivity** | 7.9 | 4.4 | 3.5 | *0.014* |

Based on the observations made in Figure 5.2, it can be inferred that the indicators of thermal comfort remain the most referenced climatic factors in the review. This suggests an increased sensitivity to thermal comfort. After conducting a comparison of means test on the datasets as shown in Table 5.3, it was found that the total climatic sensitivity exhibited a statistically significant percentage difference of 3.56 (p=0.033). The observed difference of 2.16% in the proportion of mentions related to 'Cold' between the data obtained by online scraping and manual collection did not demonstrate statistical significance (p=0.72, Table 5.3). The hot climatic references observed a drop of 8.26% based on the manual data, which was determined to be statistically insignificant at a significance level of p=0.05. In comparing sun climatic references, it was seen that there was a minimal difference of 0.68% between the web scraped and manual procedures. This difference was determined to be statistically insignificant, with a p-value of 0.45 (see Table 5.3). There was no evidence of statistical significance (p=0.68) noticed in the comparison of rain-related mentions between the manual and web scraped procedures, with just a marginal difference of 0.71% observed (see Table 5.3 and Figure 5.2). Both good and bad weather conditions exhibited p-values of 0.05, indicating statistical insignificance. Furthermore, there was a notable fluctuation of 3.5% between the two techniques (Table 5.3). The changes in drought, wind, and humid climatic conditions were all found to be below 1%, rendering them statistically insignificant (p=0.80, p=0.78, p=0.76). There was a lack of statistical significance (p=0.06) seen in the comparison of cloud-related mentions between the manual and web scraped methodologies (see Table 5.3). A marginal variation of 1.12%, which was determined to be statistically insignificant (p=0.78), was noticed in the frequency of references associated to mist (Table 5.3; Figure 5.2).

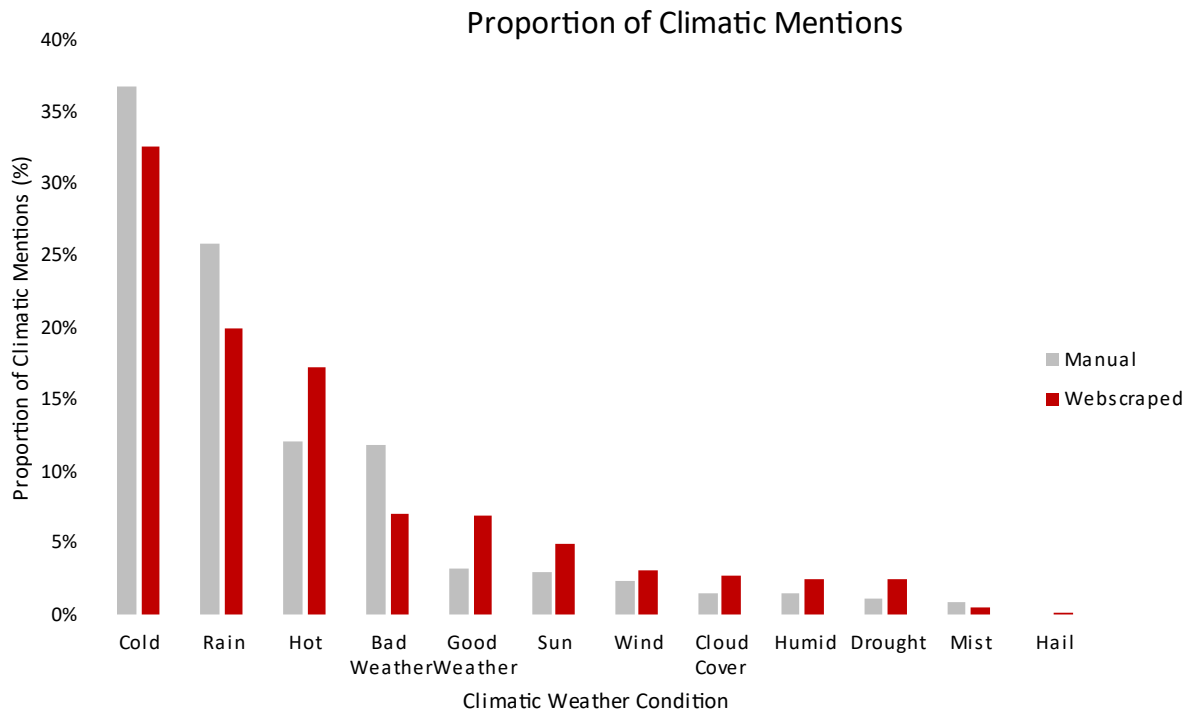



**Figure 5. 2:** Proportion of climate references obtained through web scraping in the aggregated TripAdvisor reviews for 19 South African locations

Statistically significant variations were noted in East London (p=0.02), Knysna (p=0.03) and Polokwane (p=0.02) between both manual and web scraped data. No statistically significant variations with regards to the other nine climatic references obtained through web scraping in the 17 destinations were observed between regions in both web-scraped and manually collected data, as evidenced by Figure 5.3 and Table 5.3. The web scraped data reveals that Port Nolloth is the region most frequently associated with climatic references, which is also related to the findings obtained from the manually collected data. Port Nolloth appears to have the highest number of recorded climatic references obtained through web scraping, which supports the findings obtained with the web scraped data. Belfast has the second highest recorded climatic mentions for both methods (Figure 5.2). Figure 5.3 and Table 5.3 both indicate that Port Nolloth received fewer reviews overall for both datasets, but a significant percentage of these reviews

included climatic references. Knysna had the most TripAdvisor reviews for both datasets but had the lowest recorded climatic mentions for both datasets (Figure 5.3 and Table 5.3). Knysna had a 1% difference resulting from both web scraping and manual approaches as depicted in Figure 5.3 and Table 5.3.

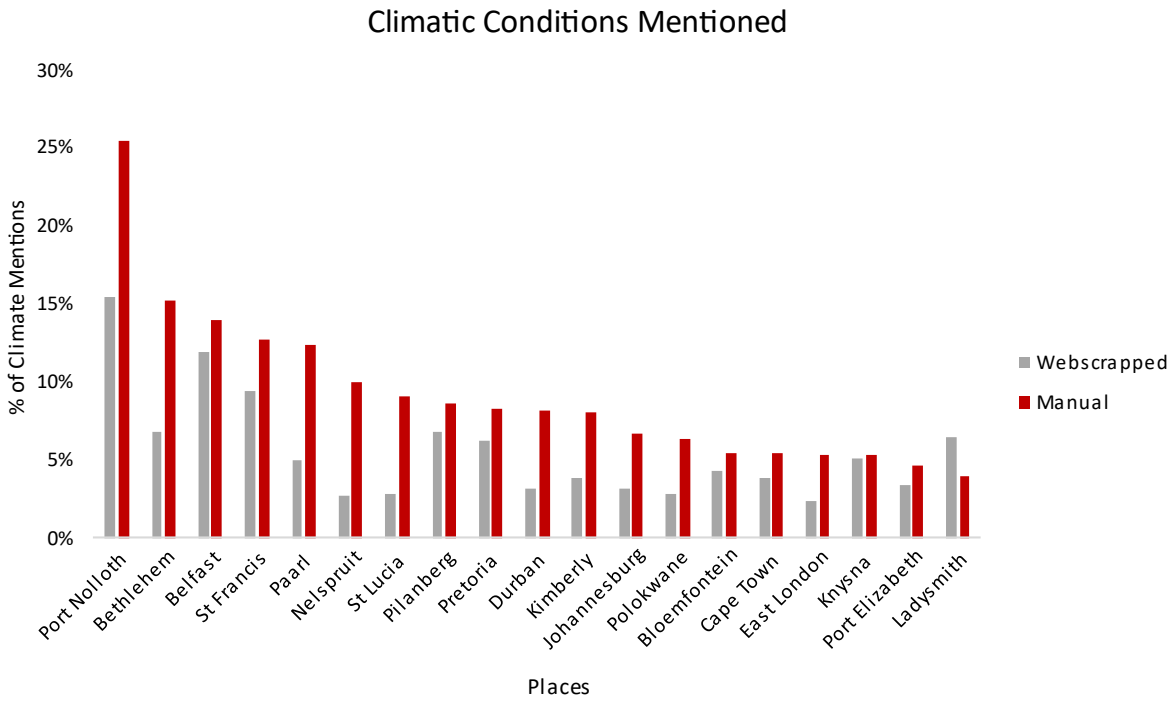
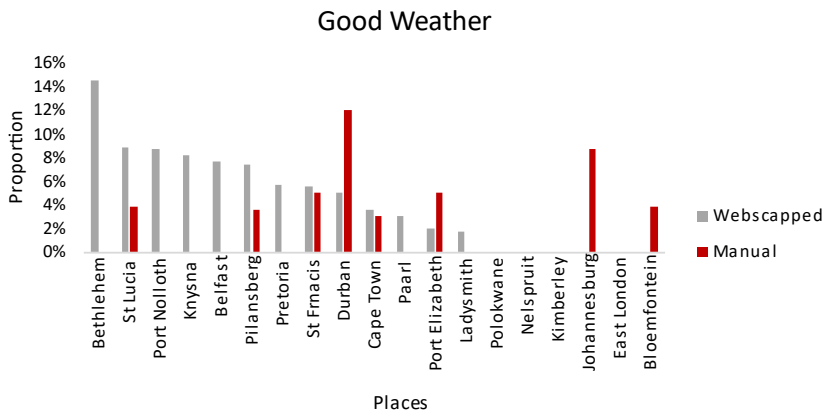


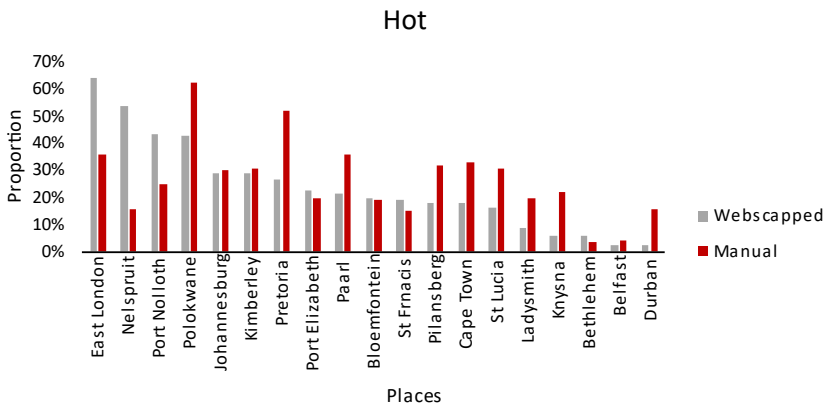

**Figure 5. 3:**The proportion of TripAdvisor reviews that mention weather conditions, subdivided by location.

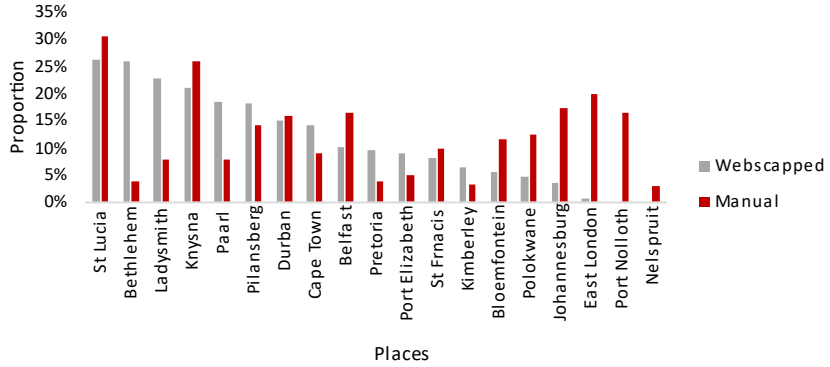
Figure 5.4 demonstrates a reduced degree of statistical variability across the datasets across all different climatic references obtained through web scraping, in a comprehensive manner. As stated before, there was no statistically significant difference seen in climate references when comparing the results obtained from web scraping and manual data collection methods. According to the data shown in Figure 5.4, 16 tourist locations exhibited climate variables that did not demonstrate statistically significant differences. The weather conditions in St Francis

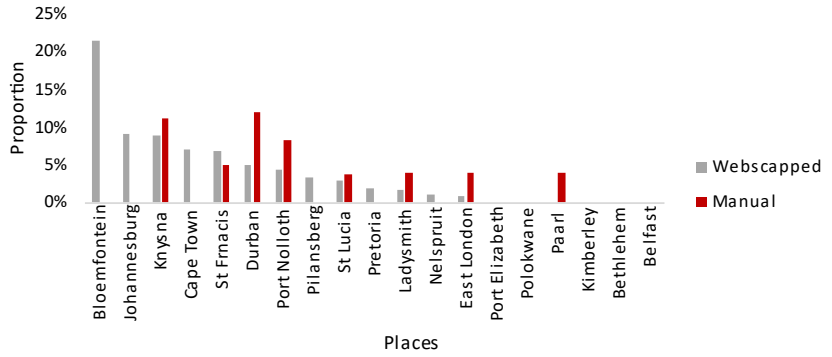
exhibited slight variations of 0.56%, whereas Bethlehem experienced large changes of 14.49% and Johannesburg saw a reduction of 8.70% in the web scraped data compared to the manually obtained data. However, it is important to note that these differences were also not statistically significant. The analysis of meteorological data in Port Nolloth revealed a drop of 32.42% when comparing web scraped data to manual data. Conversely, in Nelspruit (Mbombela), a marginal rise of 1.43% was seen in the web scraped data. It is important to note that these changes were not statistically significant. The web-based scraped data revealed an 8.89% increase in bad weather occurrences in Nelspruit (Mbombela) compared to the manual data, as well as a 6.25% reduction relative to the manual data. All these changes remained statistically insignificant. The statistical analysis reveals that the percentage variances in drought occurrence vary from 17.86% in Bethlehem to -19.23% in St Lucia.

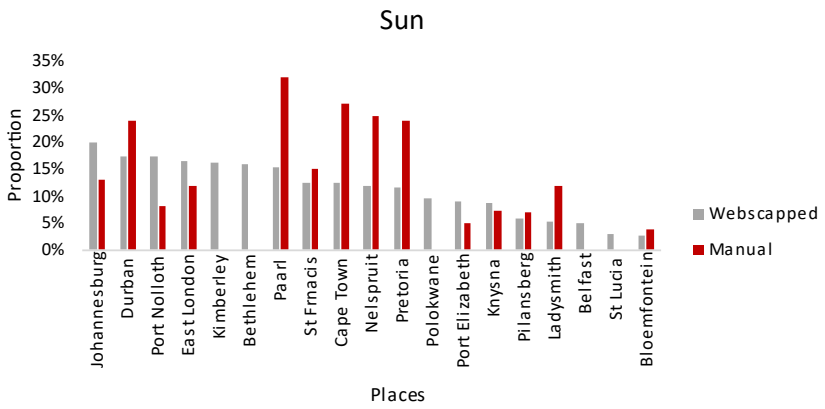
Statistically significant differences in climatic references were observed in East London, Knysna, and Polokwane utilising both methodologies. The percentage variations in references to hot weather ranged from an increase of 20.38% to a percentage difference of -19.64%. In Knysna, the variation in favourable weather references ranged from a decrease of 16.06% to an increase of 8.22%. East London experienced a significant decrease of 19.4% in rainfall-related mentions and a significant increase of 28.13% in references to hot weather. These areas exhibited considerable statistical variances overall.

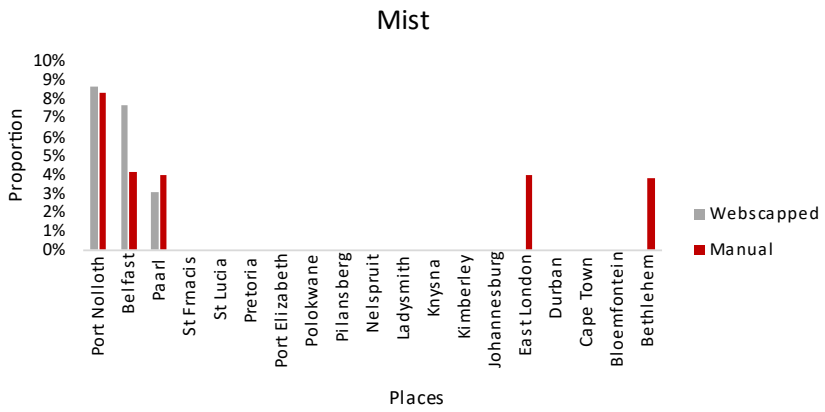

Good Weather
16%
14%
12%
10%
8%
6%
4%
2%
0%
Proportion
Bethlehem
St Lucia
Port Nolloth
Knysna
Belfast
Pilansberg
Pretoria
St Frnacis
Durban
Cape Town
Paarl
Port Elizabeth
Ladysmith
Polokwane
Nelspruit
Kimberley
Johannesburg
East London
Bloemfontein
Places
Webscapped
Manual

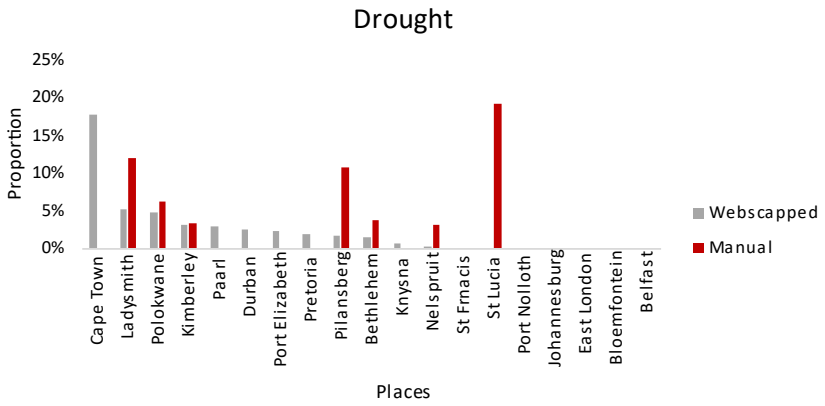

Hot
70%
60%
50%
40%
30%
20%
10%
0%
Proportion
East London
Nelspruit
Port Nolloth
Polokwane
Johannesburg
Kimberley
Pretoria
Port Elizabeth
Paarl
Bloemfontein
St Frnacis
Pilansberg
Cape Town
St Lucia
Ladysmith
Knysna
Bethlehem
Belfast
Durban
Places
Webscapped
Manual

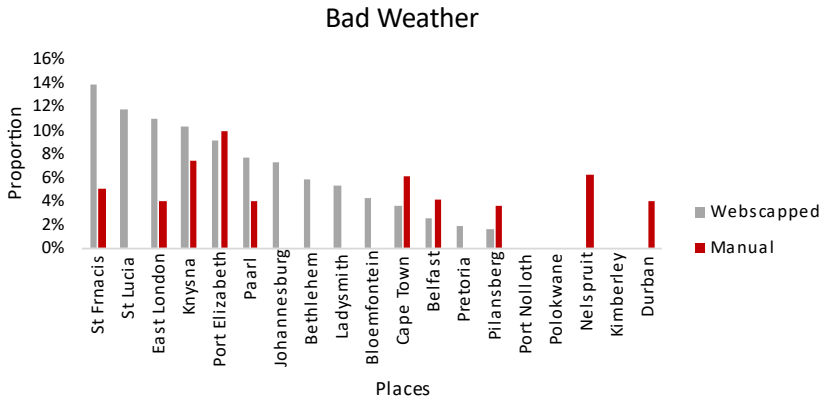

Rain
35%
30%
25%
20%
15%
10%
5%
0%
Proportion
St Lucia
Bethlehem
Ladysmith
Knysna
Paarl
Pilansberg
Durban
Cape Town
Belfast
Pretoria
Port Elizabeth
St Frnacis
Kimberley
Bloemfontein
Polokwane
Johannesburg
East London
Port Nolloth
Nelspruit
Places
Webscapped
Manual


Wind
25%
20%
15%
10%
5%
0%
Proportion
Bloemfontein
Johannesburg
Knysna
Cape Town
St Frnacis
Durban
Port Nolloth
Pilansberg
St Lucia
Pretoria
Ladysmith
Nelspruit
East London
Port Elizabeth
Polokwane
Paarl
Kimberley
Bethlehem
Belfast
Places
Webscapped
Manual


Sun
35%
30%
25%
20%
15%
10%
5%
0%
Proportion
Johannesburg
Durban
Port Nolloth
East London
Kimberley
Bethlehem
Paarl
St Fmacis
Cape Town
Nelspruit
Pretoria
Polokwane
Port Elizabeth
Knysna
Pilansberg
Ladysmith
Belfast
St Lucia
Bloemfontein
Places
Webscapped
Manual


Mist
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
Proportion
Port Nolloth
Belfast
Paarl
St Fmacis
St Lucia
Pretoria
Port Elizabeth
Polokwane
Pilansberg
Nelspruit
Ladysmith
Knysna
Kimberley
Johannesburg
East London
Durban
Cape Town
Bloemfontein
Bethlehem
Places
Webscapped
Manual


Drought
25%
20%
15%
10%
5%
0%
Proportion
Cape Town
Ladysmith
Polokwane
Kimberley
Paarl
Durban
Port Elizabeth
Pretoria
Pilansberg
Bethlehem
Knysna
Nelspruit
St Fmacis
St Lucia
Port Nolloth
Johannesburg
East London
Bloemfontein
Belfast
Places
Webscapped
Manual


Bad Weather
16%
14%
12%
10%
8%
6%
4%
2%
0%
Proportion
St Fmacis
St Lucia
East London
Knysna
Port Elizabeth
Paarl
Johannesburg
Bethlehem
Ladysmith
Bloemfontein
Cape Town
Belfast
Pretoria
Pilansberg
Port Nolloth
Polokwane
Nelspruit
Kimberley
Durban
Places
Webscapped
Manual

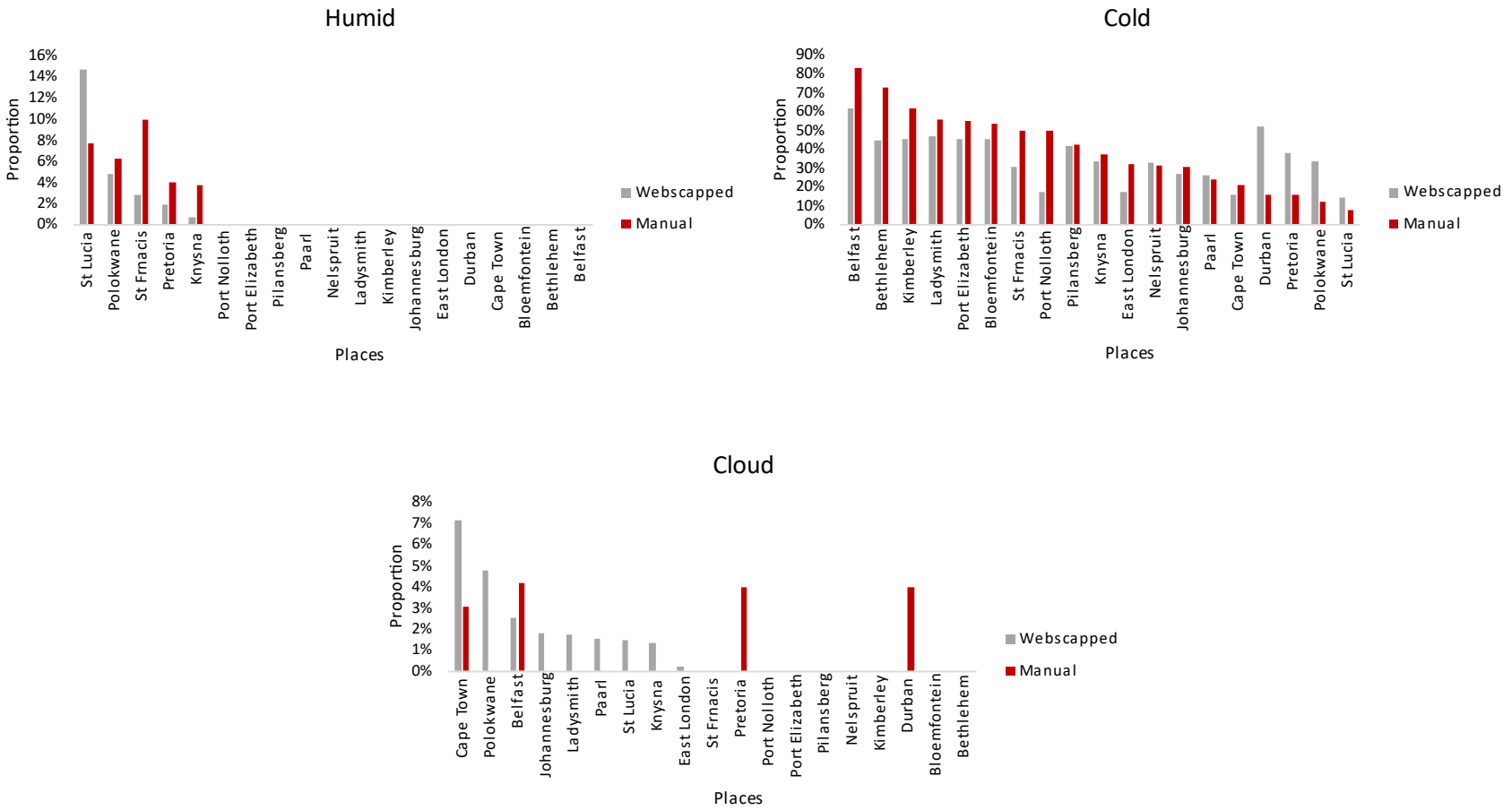



**Figure 5. 4:** Location-specific proportions of each climate component

## 5.3 Sentiment Analysis

The assessment of weather conditions by tourists constitutes a crucial focal point within the scope of this research. The relevance of evaluating an experience, particularly one in which weather plays a key role, was deemed to be of significance due to its influence on the whole vacation experience. The subsequent section will address the reviewers' contentment with the various climatic factors within the reviews obtained from the web-scraped data. The remarks pertaining to climate have been categorised into five distinct sentiment analysis categories: very negative, negative, neutral, positive, and very positive. The presence of very positive and positive references suggests a high level of contentment with the weather conditions, whilst the presence of very negative and negative references suggests a significant degree of discontent with the climatic conditions. This section will provide a comprehensive analysis of each climatic condition, delving into their specific characteristics and illustrating them by means of a range of climate conditions.

### 5.3.1 Sentiment Analysis

The assessment of weather conditions by tourists is of importance, as it can significantly influence the likelihood of their return to a particular tourist destination or their inclination to recommend it to others. The significance of appraising an experience based on weather or climatic conditions was underscored, given its impact on the overall holiday experience. Within tourist reviews, a spectrum ranging from very negative to very positive evaluations were discernible. Specifically, positive evaluations constituted 54.9% of the total, with 30.4% categorised as very positive. Neutral assessments comprised 7.3%, while negative sentiments accounted for 5.2%, and very negative evaluations constituted 2.2% (Table 5.4).

**Table 5. 4:** Evaluations of weather experiences

| | | Total Sentiment Analysis | | | | | |
|---|---|---|---|---|---|---|---|
| **Climatic Conditions** | | **Very Negative** | **Negative** | **Neutral** | **Positive** | **Very Positive** | **Totals** |
| **1** | Cold | 2.31% | 5.75% | 8.42% | 31.83% | 51.68% | 3371 |
| **2** | Cloud | 3.23% | 3.87% | 3.87% | 16.13% | 72.90% | 155 |
| **3** | Drought | 0.61% | 2.45% | 4.91% | 20.25% | 71.78% | 163 |
| **4** | Bad Weather | 0.00% | 15.15% | 3.03% | 15.15% | 66.67% | 33 |
| **5** | Good Weather | 0.00% | 0.00% | 3.57% | 14.29% | 82.14% | 28 |
| **6** | Hot | 1.39% | 5.31% | 5.18% | 36.28% | 51.83% | 791 |
| **7** | Humid | 4.71% | 20.00% | 17.65% | 25.88% | 31.76% | 85 |
| **8** | Mist | 0.00% | 0.00% | 2.33% | 25.58% | 72.09% | 43 |
| **9** | Rain | 2.70% | 4.98% | 7.82% | 24.89% | 59.60% | 703 |
| **10** | Sun | 2.49% | 1.97% | 5.77% | 31.63% | 58.14% | 762 |
| **11** | Hail | 0.00% | 0.00% | 0.00% | 66.67% | 33.33% | 9 |
| **12** | Wind | 1.58% | 5.26% | 4.74% | 22.63% | 65.79% | 190 |
| | **Total** | **2.21%** | **5.18%** | **7.34%** | **30.40%** | **54.87%** | **100.00%** |

Table 5.4 outlines the sentiment analysis associated with various climatic conditions obtained through web-scraping. Cold, hot and sun related conditions have a significant proportion of very positive sentiments, indicating that many people view these weather conditions favourably. Good weather unsurprisingly has a high proportion of positive and very positive sentiments. Rain also has a considerable number of positive sentiments, suggesting that rain might be observed as beneficial or pleasant by various tourists. Drought has a high percentage of very positive sentiments, possibly indicating relief or appreciation when drought conditions are alleviated. Cloud, mist, and wind generally have a mix of both positive and negative evaluations, but with a notable presence of positive sentiments. Hail is perceived positively by most of the tourists, which might be surprising given its potential for damage. Humid has relatively balanced evaluations. Bad weather has the same number of positive and negative evaluations but no very negative evaluations, which is counterintuitive but could indicate resilience or appreciation for challenges faced during such conditions.

### 5.3.2 Different Weather Themes and Sentiment Analysis

Cold conditions are categorised as part of thermal comfort. Mentions related to cold weather accounted for 53.2% (n= 6,333) of all climatic references obtained through web-scraping. Most of these mentions, totalling 51.7% (Table 5.4), expressed very positive sentiments. Conversely, 8.06% of mentions conveyed negative or very negative sentiments. Cold conditions can negatively affect a traveller's thermal comfort, either by discouraging outdoor activities, or by affecting their overall satisfaction with the stay, as exemplified by the following reviews.

#### *5.3.2.1 Cold*

> "Great location, nice friendly staff, good food and the animals were terrific. Had a couples massage at the spa, very relaxing and quite enjoyable. Weather was a bit cold, so eating was a little uncomfortable due to no heat in the dining area." October (Oct) 2019, Colorado.

> "Overall, I am satisfied. Though the hotel is the best within the hospitality industry, I called maintenance to come fix my heater or assist in its performance. It was freezing cold in the evening and in the morning." July (Jul) 2017, Pretoria.

The reviews (Figure 5.5) demonstrate the influence of cold weather on multiple aspects of a guest's experience, including outdoor activities, comfort in hotels, and eating locations. Having appropriate amenities like heating systems, fireplaces, and warm clothes may reduce the impact of cold weather, leading to increased visitor pleasure.

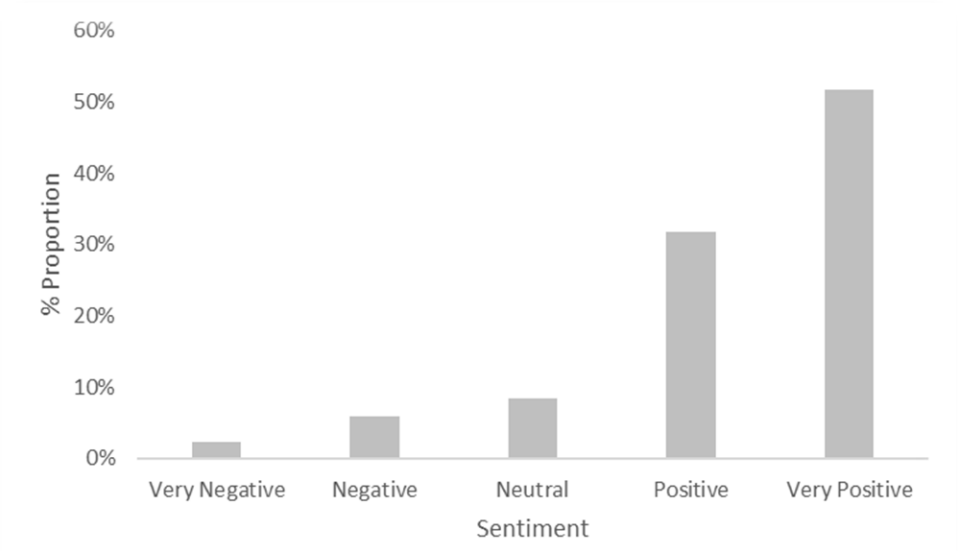



**Figure 5. 5:** Sentiment analysis of cold related climatic conditions

*5.3.2.2 Cloud*

Clouds constituted of 2.4% (n= 6,333) of the overall tourist climatic reviews. Clouds may additionally correlate with tangible climatic phenomena, for instance, rainfall or storms; however, it is improbable that activities would be cancelled, or tourists would depart prematurely solely due to overcast or cloudy conditions. Of the 155 cloud-related comments, 72.9% were very positive (Table 5.4; Figure 5.6). The following examples show how the travellers are affected by cloudy weather conditions:

> "Restaurant scenery, there are more features, huts, night stars, clouds I will never forget."

> "We visited and stayed in Cabin 3 (Chalet) for 2 nights, at the end of May 2017. PROS: We loved the location near the end of the road (so there was not a lot of car traffic or noise), and the sunset views from the lawn outlook were beautiful, even when it was overcast. Fortunately for us the weather was decent, some fog and overcast, but not uncomfortably humid or balmy."

In certain cases, however, clouds can negatively influence the overall experience of a tourists visiting a particular destination altogether. The negative evaluations pertaining to clouds amounted to 3.9%, with 3.2% being very negative. Some examples of such sentiments include:

> "We were only for one night and unfortunately it was raining. We did not see the Drakensberg because of the clouds. It is a really beautiful place, maybe we will come again when sun is shining. It is on the top of a hill with a naturally surrounding. The dinner was excellent, and the wine cellar is a gem." 2021 Cape Town, South Africa.

> "A lovely, simple place, downside: bathrooms were a bit old and water trouble in rooms, and the weather wasn't to good when we went, was cloudy and rainy, Positives lovely rooms clean and neat helpful friendly staff, lovely meals taste and variety (I am vegetarian and compared to most place they offered a variety of options, close to the ocean beautiful scenery, good kids entertainment, they have a small spa but the treatments were good". Sep 2017, Canada, Mississauga.

Overall, the reviews indicate that weather, particularly cloud cover and rain, can influence the visitor experience, especially when it affects outdoor activities or visibility of scenic landscapes. However, positive aspects of the stay, such as accommodations, dining experiences, and helpful staff, can mitigate the impact of unfavourable weather conditions.

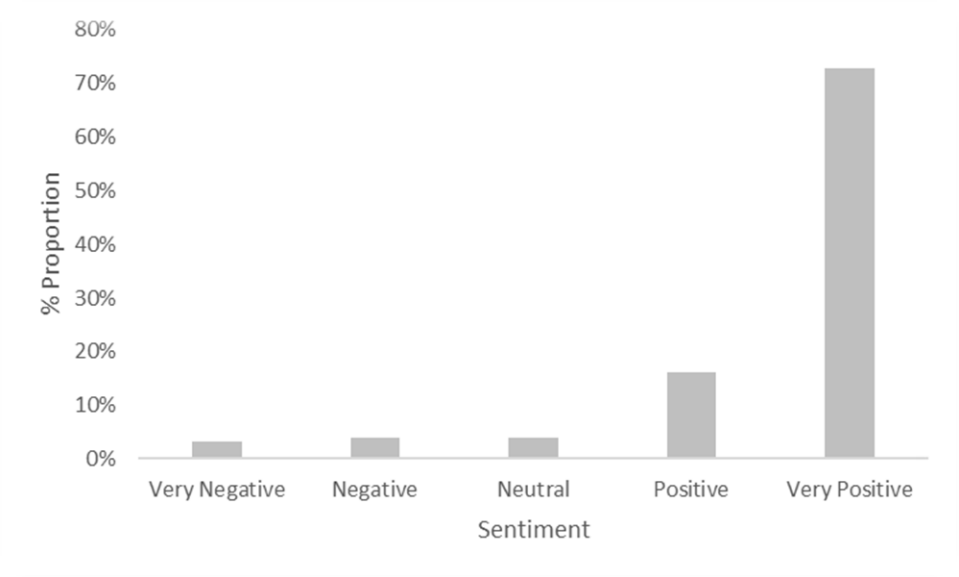



**Figure 5. 6:** Sentiment analysis of cloud related climatic conditions

*5.3.2.3 Drought*

Drought accounted for 2.6% (n= 6,333) of the climatic mentions in reviews, with 0.6% of these instances categorised as very negative (Figure 5.7). Both negative and positive mentions had major variations with negative mentions constituting 2.5% and positive mentions constituting 20.3% of the total, with a majority (71.8%) falling under the predominantly very positive sentiment category (Table 5.4). Some reviews highlighted the effects of drought conditions on wildlife, the environment, water availability, conservation efforts, and the importance of raising awareness about the impacts of drought as in the following examples:

> "Staff are very nice great to have texts with gate code morning off stay and warm welcome. Tents are as described very nice relaxing place. The drought has taken its toll with dam near empty, animals are around but sightings are scarce. Pool area is nice but too small for 10 tents in a 36-degree day."
>
> February (Feb) 2019.

"The Cape is suffering under a severe drought and the city of Cape Town has imposed water restrictions. Blaauw Heim, however, has a borehole and thus a lot of water. The rest of our visit was exactly as we knew it would be. Great hospitality, the best beds in the Western Cape and Blaauw Heim's signature breakfasts on the patio! Thank you, Johan, Jo-Anne, Richard, Rafiki, Kat en Voël! We will be back." January (Jan) 2018, Pretoria, South Africa (SA).

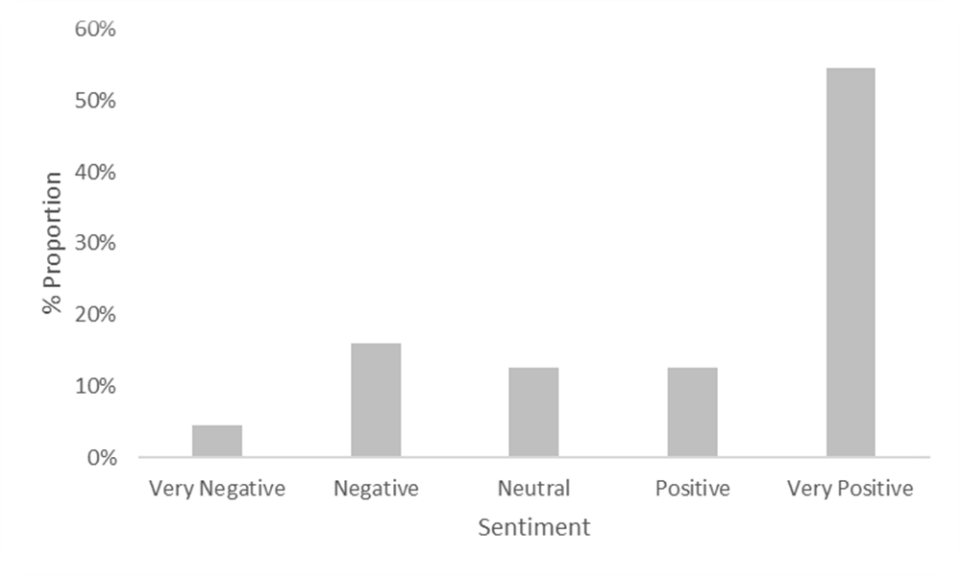



**Figure 5. 7:** Sentiment analysis of drought-related climatic conditions

*5.3.2.4 Bad Weather*

Bad weather is typically linked to cold, wet, and windy conditions. Bad weather constituted 0.5% (n= 6.333) of the weather conditions with majority of the mentions being classified as very positive (66.7%; Table 5.4; Figure 5.8). Most of the bad weather-related reviews focused mainly on the need for seeking shelter and comfort, and having to cancel outdoor activities in the case that adverse weather conditions were to be experienced, such as in these examples:

> “This was truly the highlight of our stay in the Port Elizabeth (Gqeberha) area. Unfortunately, we did not get to swim or do any of the hiking routes due to bad weather, but staying in a cabin in the woods with that awesome view was completely worth it”. Jan 2015, Location not found.

> “We stayed for one night only. The weather was dire, so we were indoors most of the time. This is a very interesting building, and the Indian theme is carried out tastefully. We had dinner in the casual restaurant and the food was good with prompt service despite being very busy. Oceanview room small but clean and well appointed. Staff were very helpful and considerate. The bad weather meant that the bar was jam packed, so we had only one drink, but it looked like a great location”. May 2016, United Kingdom (UK).

> “We enjoyed two nights at the Westlodge BandB. We had a lovely room with an open bathroom. Rezne is a great host and offers excellent cooking and good talk. Unfortunately, we had bad weather during our stay, so we could not fully benefit from the nature highlights in the surrounding of Graskop. The town offers not enough to overcome bad weather periods”. March (Mar) 2015, France.

In general, these reviews illustrate the substantial impact unfavourable weather conditions can have on guests' experiences, ranging from constraining outdoor activities and wildlife observation opportunities to restricting guests indoors and impeding exploration of the nearby area. Nonetheless, establishments that prioritise comfort and hospitality during adverse weather conditions can alleviate the adverse effects and offer a welcoming sanctuary for guests.

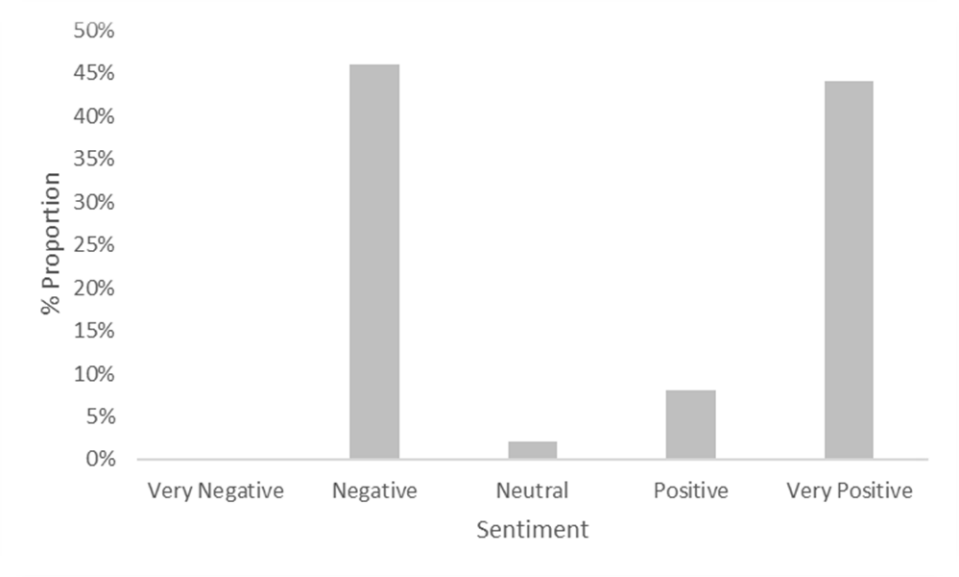



**Figure 5. 8:** Sentiment analysis of bad weather-related climatic conditions

*5.3.2.4 Good Weather*

Good weather is typically linked to sunny, less windy, and clear skies weather conditions. Good weather constituted of 0% very negative and negative weather mentions, with majority of the mentions being classified as very positive (82.1%; Table 5.4; Figure 5.9). Neutral sentiments amounted to 3.6% with positive sentiments being at 14.3% (Table 5.4). Most of the good tourists enjoying scenic views and pleasant atmosphere, as well as enhanced safari experiences, such as in these examples:

> "The service was amazing, owner is such an overall great human being, very helpful and chilled. The room as specious, clean and neat with a beautiful patio that gives incredible sunrise view in the morning and beautiful sunset in the evenings. the place is in a perfect location close to all amenities. The weather was beautiful, and it was an overall amazing experience." Jan 2019, location not found.

> "An amazing experience. This really was the whole reason for me going to Africa to do a safari and this did not disappoint in any way the accommodation was lovely and unique. The food was delicious, and the service was perfect. Special thanks to Noni our guide who was amazing and so passionate about the

> animals and Casper the man who kept us all in order. He was so bossy but so funny and lovely. On the one afternoon that we had good weather we were in the plunge pool when two baby giraffes walked past how magical is that? The stars were so bright there. Unfortunately, we did not get to see the free lions only the old ones in the enclosure, but I did manage to get to Shamwari and see the Born Free Foundation encloser which is very sad on one hand but as the guide said it's a…" Jan 2014, UK.

The observations show how positive weather conditions may greatly influence tourists' experiences throughout their vacations. Good weather is essential for creating memorable and satisfying visitor experiences, whether it involves enjoying beautiful views and pleasant atmospheres or engaging in outdoor activities like safaris.

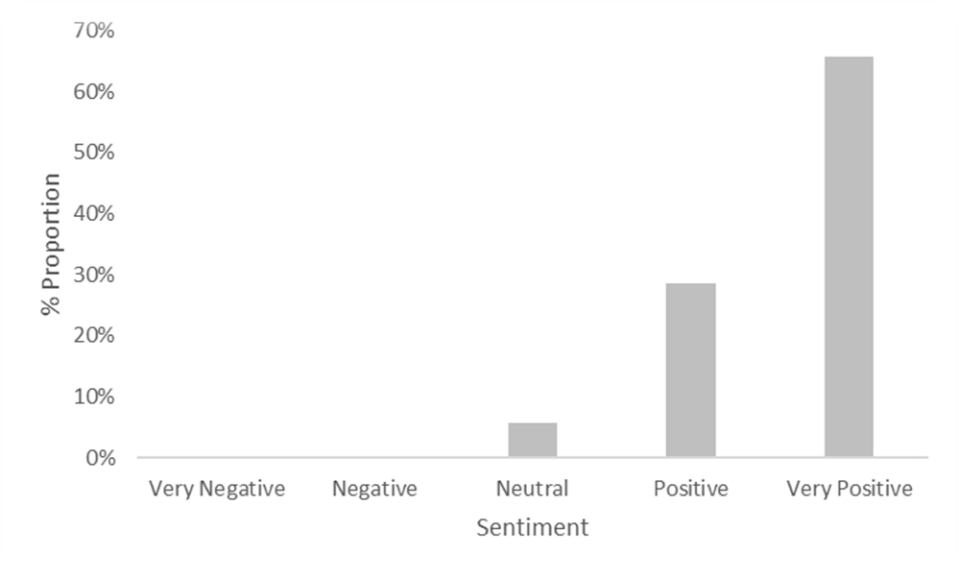



**Figure 5. 9:** Sentiment analysis of good weather-related climatic conditions

*5.3.2.5 Hot*

Hot weather mentions accounted for 6.5% (n= 6,333) of the weather-related mentions. Very negative sentiments amounted to 1.4% of the weather mentions with negative sentiments being

5.3%, positive sentiments amounting to 36.3% (Table 5.4; Figure 5.10) and very positive sentiments amounting to 51.83%. Hot weather directly influences tourists' thermal comfort, which can influence their contentment or dissatisfaction when staying at a particular hotel destination, as exemplified in these reviews:

> "Our hosts were attentive; food was excellent, and we had a lovely overnight stay. it was extremely hot, and the pool and sundowners were welcome relief from the heat. We opted to have the braai prepared by the hosts and the meal was generous and delicious." November (Nov) 2020, Pretoria, SA.

> "Thank you to the staff for always delivering exceptional service. We will always return to this beautiful, secluded place. We highly recommend Return Africa Pafuri - just maybe not during a heat wave!" Nov 2021, Location not found.

> "Rangers did not try to show us the big five, I saw a lack of interest. Place was very nice, food was ok, staff was very helpful, noting to say. I will recommend your Lodge. When we visited your location weather was very hot, more than 40-degrees, just because of this we did not sleep well but the place was so nice it was not a big issue". Mar 2016, London.

> "From the welcoming cold towels on a very hot day to the beautifully thought-out huts and the amazing guided-game drives, the whole experience couldn't have been better. Customer service at its best. Stay as long as you can." April 2016, London.

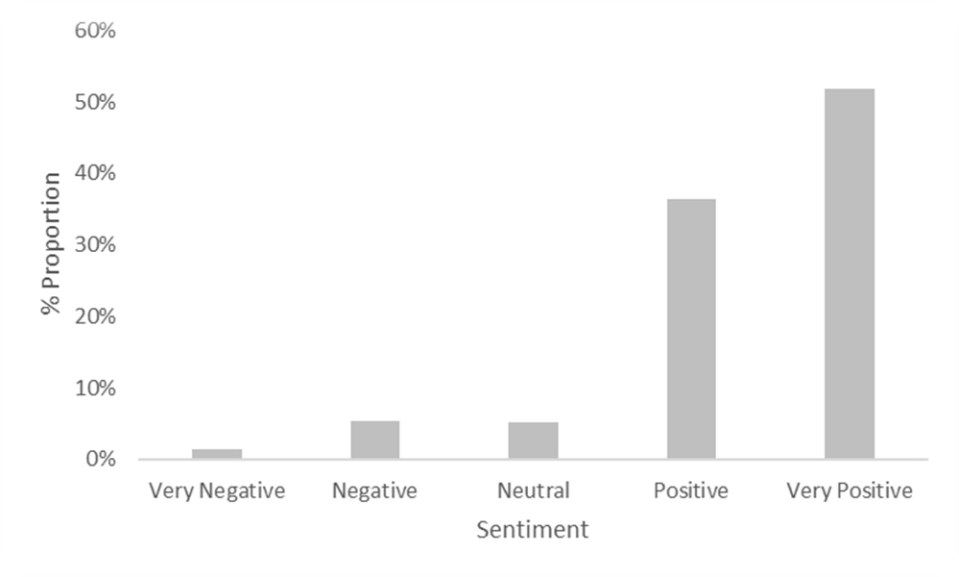



**Figure 5. 10:** Sentiment analysis of hot related climatic conditions

*5.3.2.6 Humid*

Humid conditions constituted only 1.3% (n= 6,333) of total weather-related mentions. Regarding review sentiments, very negative expressions accounted for 4.7%, while 25.9% were positive mentions and 31.8% were very positive sentiments concerning humidity (Table 5.4; Figure 5.11). Several reviews addressing humid conditions primarily highlighted discomfort due to inadequate air-conditioning, the adverse effects of humidity on sleep quality, and appreciation for implemented cooling measures, as in these examples:

> "We stayed at the African Rock hotel for 2 days upon arrival in Johannesburg and again for 1 night before departing South Africa. We were greeted at the airport by the hotel driver. He was friendly and was already waiting for us when we arrived. The drive to the hotel was in a new BAYERISCHE MOTOREN WERKE (BMW) which was very comfortable. He had cold water for us which was much appreciated after a long flight and being greeted with humid weather. The ride to the hotel was less than 20 minutes. Upon arrival at the hotel, we were greeted warmly and provided with a small fruit cocktail while filling out the required details (not much paperwork and only took 2 minutes). We were then given

a quick tour and shown to our room. We appreciated that the hotel is a small boutique style with only 9 rooms. …". Feb 2017, Ontario, Canada.

"Staff at the hotel is exceptional. Breakfast and buffet restaurant are very good. Aircon in our bedroom was not working and it was very hot and humid during the night; our sleep was affected because of this. We had the opportunity to report it only in the next morning." Mar 2017, London.

These assessments underline the considerable influence of humid weather on tourists' experiences throughout their visits. Inadequate cooling facilities, like as air conditioning, may lead to discomfort and disrupted sleep, eventually reducing the overall quality of the stay. However, visitors may still appreciate other parts of the rooms and services, highlighting the importance of resolving climate-related concerns to enhance guest contentment.

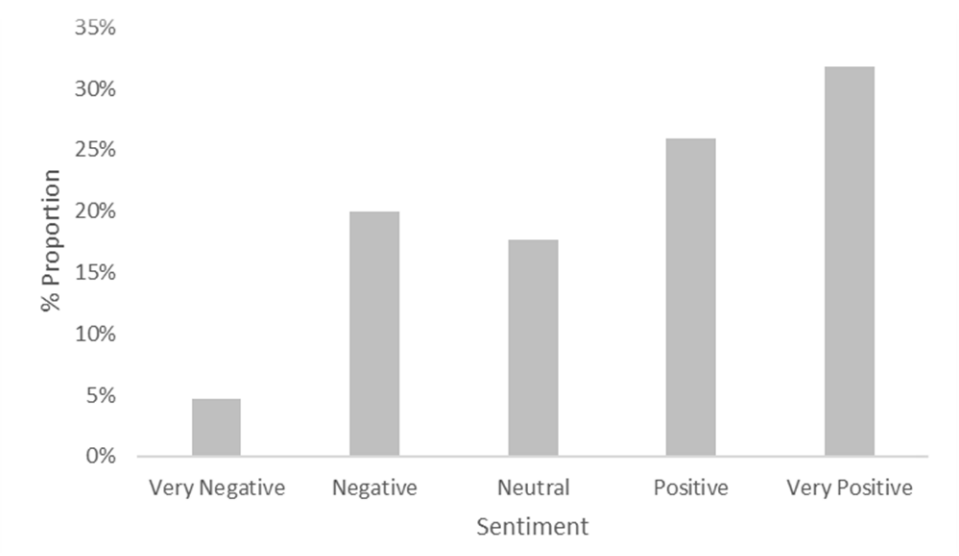



**Figure 5. 11:** Sentiment analysis of humid related climatic conditions

*5.3.2.7 Mist and Fog*

Mist and fog related weather conditions tend to be mentioned interchangeably within reviews. Mist related weather conditions constituted of 0.7% (n= 6,333) of the overall climatic mentions (Figure 5.12). A total of 72.1% of the reviews resulted in very positive sentiments from reviewers with 0% of the reviews being negatively associated with reviewers (Table 5.4). Some of the reviews pertaining to both mist and fog related mentions covered various themes such as the preference of fog-related weather, scenery, the lack of visibility and disappointment thereof as well as precautions and recommendations. The following reviews exemplify this:

> "We were so unlucky with the weather for the 3 nights we were in Joburg - being mid-June a cold front passed over and the temperature never rose above 5C. It was also very foggy, and we could not appreciate the gardens at the Lucky Bean, which was disappointing. We had 2 rooms, and both were spacious - which meant they took a lot of heating - as central heating is not usually required in South Africa, we were given radiators and gas stoves (the latter being effective but a little alarming to light and impossible to leave unattended). The beds had sufficient blankets. Wi-Fi excellent. What was warm was the welcome, and the general helpfulness of staff both before and during our stay. Breakfasts are jolly as people tend to gather round the large Table and chat. Help yourself cereal…." Jul 2016, Edinburgh, UK.

> "Very beautiful place to stay...although cold and rainy, the rooms are warm with heaters and electric blankets provided. Roads to Estatean was quite bad due to rain and mist so a high ride vehicle is preferable." April (Apr) 2022, Johannesburg, SA.

> "Owners were extremely welcoming and helpful, with advice on local sites, etc. Good sized clean, tidy, well-equipped room. The only downside was that the Wi-Fi did not work in our room, we had to go to the dining room to connect. Dinner (pre-ordered) was nice, but quite expensive. Breakfast was well prepared and there was a wide selection. Unfortunately, while we were there was low cloud and mist in

the mountains, but the owners were able to offer suggestions on where to go where we could see something interesting". September (Sept) 2014, Belize.

"A very nice house. Cosy and friendly atmosphere, both in rooms, lounge, kitchen and especially from the hosts and their staff. The home-cooked dinner was very good, with good wines. Very nice breakfast with many alternative dishes. Only negative thing was the rainy and foggy weather but that leads to even higher chance to visit Westlodge again to benefit from all service." Sweden, 2016

The reviews highlight the significant impact of fog and mist on travellers' experiences throughout their visits.

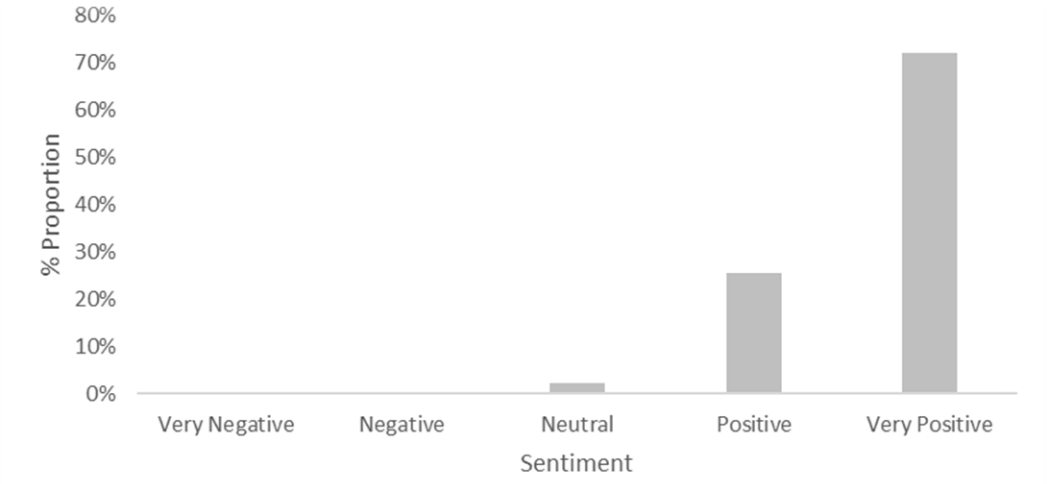



**Figure 5. 12:** Sentiment analysis of mist and fog related climatic conditions

*5.3.2.8 Rain and thunderstorm*

Rain and thunderstorms were frequently discussed together in the reviews. Rain was mentioned in 11.1% (n= 6,333) of all reviews (Figure 5.13). Negative sentiments accounted for 5.0% of the reviews, whereas positive sentiments constituted 24.9%. Overall, most sentiments (59.6%) were

very positive, with 2.7% of reviews expressing very negative sentiments. Most reviews detailed experiences such as adventurous encounters, missed opportunities, and cancelled activities due to the rain and thunderstorms encountered, for example:

> "Tops was hit by a power outage for several hours, but all staff pitched in to ensure our comfort and apart from the darkness, it was business as usual although the afternoon drive had to be cancelled because the storm was raging - but the offer to go out was made. Torches and umbrellas were handed out to help folks get to their rooms. Be aware that some rooms were badly affected by the driving rain (not rooms 1 and 5) and some guests had soaked bedding and clothes - but again…." Jan 2015, Knysna, South Africa.

> "What a disappointment. We found ourselves stranded in East London when our South Africa coastal trip was cancelled by a horrendous storm and relied on a guidebook recommendation to quickly find a respite from the rain. It boasts its 4-star rating at the front desk, but any basis for that rating is nowhere to be found in either the accommodations or the staff. It has a picturesque setting overlooking the river, but all of the one-bedroom rooms have a view of only a garden. The rooms are worn out—lots of water spots on the ceiling from inadequate roofing and water runoff, empty refrigerators that look dingy in their cabinet, no safe in the rooms, and generally the feel of a tired hotel. A sign at the front desk says that renovation is underway, but it will not cure the biggest…". April 2013, New York.

> "Normally you expect that you get what you pay for but that is not always true. Spending 24 hours at Kariega Game Reserve makes you dependable on good weather and we were not fortunate. We did get a cold afternoon/evening trip around the area and saw amazing elephants. But our wake-up call came an hour earlier than expected and we were informed that the rain was pouring down. Everything was cancelled in the upcoming hours, and we had our stay shorted. Well, you cannot always know about the weather. The lodging was ok, the food was great, but I recommend you stay at least 2 days to avoid bad weather experiences". Feb 2018, Sweden.

> "I can only add to what everyone else says about this lodge. I cannot wait to plan another trip to Bukela. The drives were thrilling, we experienced all sorts of weather - it was even fun when the rain poured

down. The encounters with the large mammals were, of course, exciting (and frequent) but even more so were the unexpected sightings of tiny frogs, bugs, birds, butterflies etc, all of which our ranger Stuart patiently taught us about. He was calm and reassuring, knowledgeable and entertaining. We discovered so much more than we expected to and enjoyed every single minute. Arriving back at the lodge after a drive we were greeted with damp cloths and drinks appropriate to the weather and a roaring fire when a chill wind blew. The staff were all charming and took great…." December (Dec) 2010, London.

"We had 3 lovely nights at Bushmans Kloof - a place we have wanted to visit for ages. The weather did not quite play ball as it was quite cold and wet which meant that the flowers didn't open up until the last day when the sun re-appeared. We had a really nice guide who took us out in the rain and did not seem to mind that he got drenched driving the open Land Rover. We could not reach the bushman paintings on the first morning because the level of the river had risen so much due to the heavy rain, but we had a lovely walk in the drizzle and mist, nevertheless. We got to see some of the paintings on the day before we left, and our guide gave a really informative talk on the history of the region and its earliest inhabitants. We met some interesting people staying at the lodge and…." Sept 2012, Pietermaritzburg (PMB), KwaZulu-Natal.

Guests' satisfaction with their lodgings and entire experience may be affected by how well facilities manage and react to adverse weather conditions that can lead to interrupted activities, comfort difficulties, lost opportunities, and necessary changes.

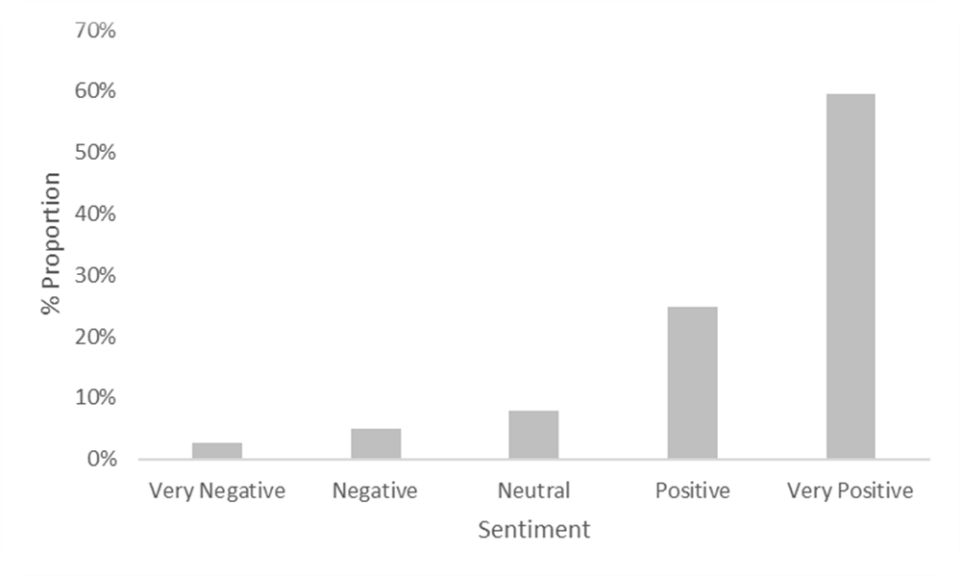



**Figure 5. 13:** Sentiment analysis of rain related climatic conditions

*5.3.2.9 Sun*

Sun related conditions constituted 12.0% (n= 6,333) of all climatic mentions obtained through web-scraping, with a majority of mentions being very positive (58.1%) and only 2.5% categorised as very negative (Figure 5.14). Sunny conditions are widely regarded as crucial within most reviews as they significantly influence the vacation experience. Some of the reviews pertaining to sunny conditions, as in the following examples, highlighted delightful safari experiences, breathtaking scenic views, and provided suggestions for enhancing future tourists' experiences:

> "We had no idea what to expect when we booked safari at Pumba as part of our honeymoon trip but wow this place is amazing. We loved every second. The game drives were great fun, as was our driver. The sunset and sunrise stops were a lovely touch. Our villa was AMAZING and probably the best accommodation we will ever stay in. Sunbathing while watching elephants, monkeys, zebras, and any other safari life will be hard beaten. The food was exceptional and the staff a real credit to Pumba. A fantastic time."

> “Everything from beginning to end of our stay here was amazing - we stayed for 3 nights and had 6 game drives. Our ranger, Matthew, was so enthusiastic and knowledgeable and went out of his way to make sure we saw everything we hoped to see, and more! Coffee breaks and Sunset stops provided fabulous views. Our room was huge and had an amazing view over the waterhole. All the staff were friendly and helpful and the food in the restaurant was lovely. The whole place had a lovely atmosphere, and it was truly a fantastic experience. Would love to do it again!!”

> “What an amazing experience! We stayed at the River Camp which was small and very exclusive (but not as expensive as many other game reserves) and were treated really well. As it is so small (only 6 tents I think) there is a real personal service with plenty of excellent food and drink (all included). Its very child friendly and great fun. Our two children are still buzzing about their time in the tents and eating around the camp fire. We had a look at the Forest Camp, and this looked fantastic as well, although a little bigger. Arriving by boat up the river is a great way to start and the early morning boat ride to the game drive along a misty, silent river watching the sun rise can only be described as 'magical'. The tents are really comfortable and the 'living/eating' area well…”. May 2007, UK.

> “The service was amazing, owner is such an overall great human being, very helpful and chilled. The room as specious, clean and neat with a beautiful patio that gives incredible sunrise view in the morning and beautiful sunset in the evenings. the place is in a perfect location close to all amenities. The weather was beautiful, and it was an overall amazing experience.” Jan 2019, location not found.

The reviews mentioned illustrate that sunny weather enhances tourists' satisfaction throughout their visits. Sunny weather significantly enhances visitors' satisfaction and joy throughout their stay by providing opportunities for outdoor activities, beautiful scenery, a good state of mind, and unforgettable experiences.

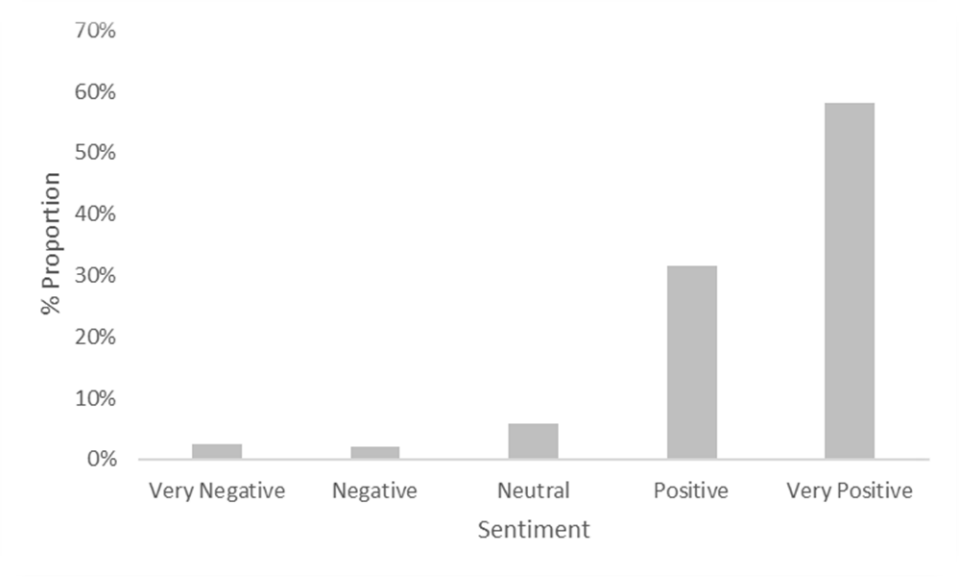



**Figure 5. 14:** Sentiment analysis of sun related climatic conditions

*5.3.2.10 Hail*

Hail conditions made a minor contribution to the overall reviews, with only nine mentions reflecting reviewers' sentiments. All mentions regarding hail were positive, with 66.7% classified as positive and 33.3% as very positive (Figure 5.15). Most reviews concerning hail conditions offered recommendations and cautioned future visitors, for example:

> "Facilities and Staff personnel met my expectations and beyond. I hereby suggest that the hotel should have a covered packing in case of hail storms as well as the heavy sun heating the cars. It is really a pleasure to have stayed at your hotel." Cape Town (CT) 2019, Pretoria, South Africa.

Overall, while hail weather conditions may present challenges for guests, such as limiting outdoor activities, it also provides opportunities for establishments to prioritise indoor comfort amenities and implement measures to enhance guest safety and satisfaction, such as covered parking.

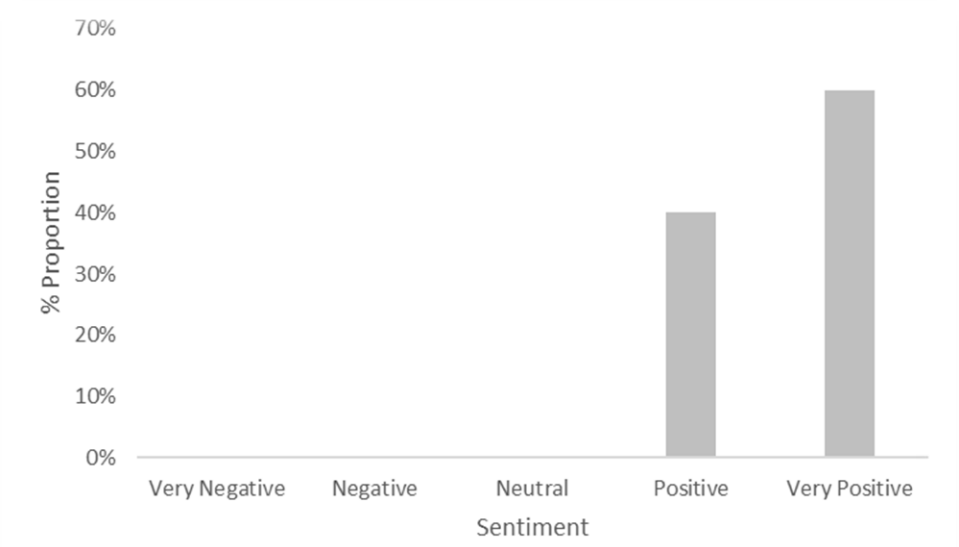



**Figure 5. 15:** Sentiment analysis of hail related climatic conditions

*5.3.2.11 Wind*

Wind accounted for 3.0% (n= 6,333) of the total climatic mentions, with 1.6% categorised as very negative, 5.3% as negative, and 65.8% as very positive (refer to Table 5.4; Figure 5.16). The majority of reviews addressed topics such as disturbed sleep and comfort issues, as well as precautionary measures and adaptations, as exemplified by these reviews:

> "Very scenic and beautiful location. Very friendly staff and helpful with any problems that might arise. Wind was unpleasant, everything that could squeak in the luxury suite did squeak on windy nights which lead to several poor night's sleep!" Jul 2014, South Africa.

> "My review about 4 years ago opened the doors of the Inn for Indian tourists. I have been away for four years. We finished our business in Oudtshoorn and decided we would have lunch at Eight Bells on the way home. We got to the hotel at 12.20 p.m. and sat at the Tables under the trees to enjoy the beautiful

breeze and the fantastic views of the mountains and the valleys. After 15 minutes I got up to go and call a waitress. Our order was taken for food and drinks. The drinks came. 20 minutes later my wife finished her drink and wanted to order a refill, but no one came. Wind picked up, so we got up and moved to the Tables on the patio under umbrellas. After waiting for some time my wife called our waitress who heard her but went past us on to our first Table to collect the empty…". Nov 2019, South Africa.

"This is our very last night of our 3 day stay here, as I am writing this, we had a really incredible stay, from the moment we entered Collen ( the owner) sat us down gave us a welcome drink, explained everything to us in supper great detail, when we wanted a boat tour he phone for us and booked it for us, he gave us insight on what to do every day and all the little things he did went such a long way for example 1) he offered every day if we wanted to take a cooler box with ice ( free of charge) so that we can pack our cold water bottles in it, or a picnic 2) free masks and snorkelling gear was available to us to take to cape Vidal ( although we didn't as it was windy) 3) every night you have yoghurt and fruits you can take in the main living room 4) free maps ( both for…".Canada, June (Jun) 2015.

"Fabulous guest house! Very clean, big rooms with a/c and excellent breakfast. The two pool areas were lovely, and Adel and Shaun were very helpful both before and during our stay having booked a last-minute hippo and croc tour for us because our whale watching had been cancelled due to high winds. Parking was secure. Would thoroughly recommend." Oct 2019, UK.

In summary, these observations indicate the significant impact of windy weather on tourists' experiences throughout their visits. Responding well to windy circumstances is crucial for guaranteeing visitor satisfaction and ease, whether it involves interruptions in sleep or outside eating.

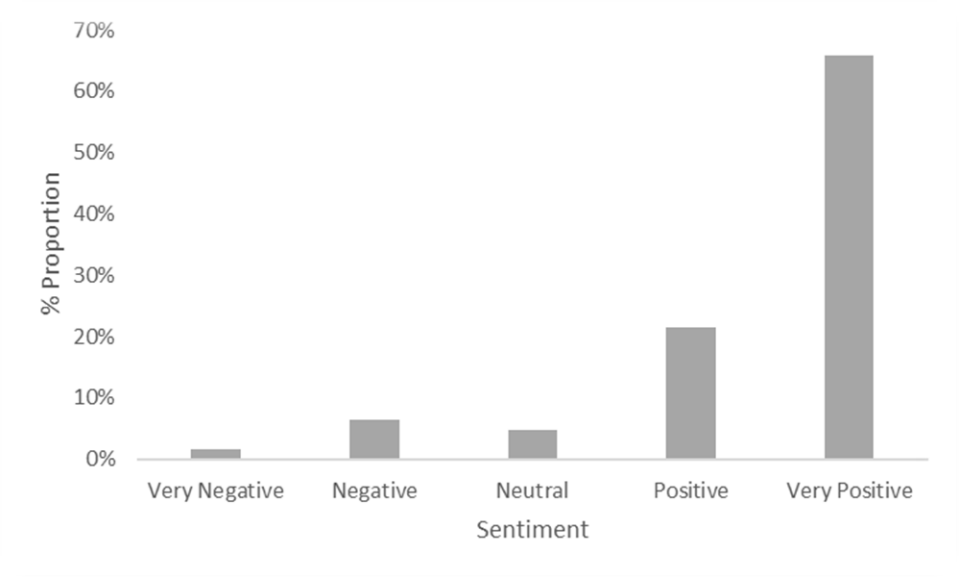



**Figure 5. 16:** Sentiment analysis of wind related climatic condition

## 5.4 Regional Expansion of Analysis

Based on the demonstrated efficacy of web-scraping as a valuable tool for analysing the climate sensitivity of tourists in South Africa, the scope of the research was extended to encompass the entirety of South Africa. A comprehensive dataset comprising 161,280 TripAdvisor reviews was collected, through the application of web-scraping, from 433 tourist accommodation establishments located across various tourist destinations spanning nine provinces in South Africa.

### 5.4.1 Sample Size and Demographics

Of the sample of 161,280 TripAdvisor reviews, 31.1% (n=161,280) were written by South African residents, with 11.4% (n=161,280) originating from residents of Gauteng, indicating a substantial presence of local residents in the tourism sector and a notable influx of tourists from the economic centre of the country. A further 16.8% (n=161,280) reviewers opted not to disclose their location.

The remaining reviews originated from 96 different countries, with the largest proportions authored by residents of the United Kingdom (21.1%, n=161,280), the United States of America (8.0%, n=161,280), and Australia (2.9%, n=161,280; Figure 5.17).

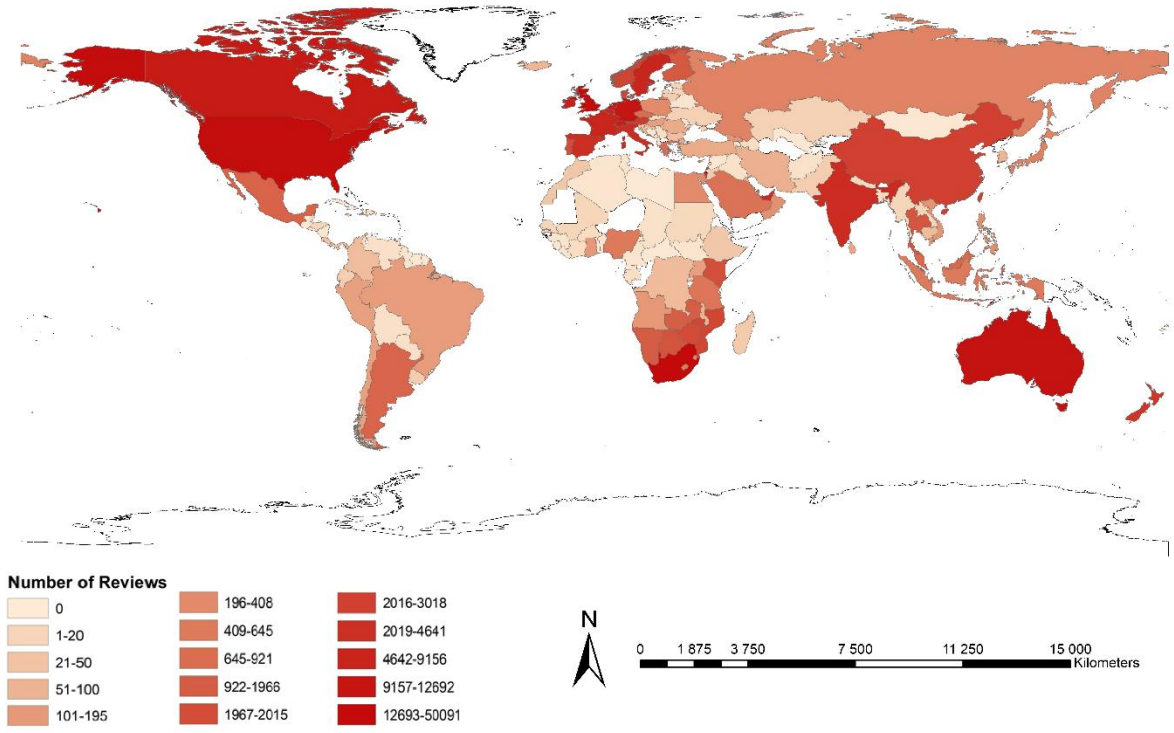



**Figure 5. 17:** Map illustrating proportional distribution of TripAdvisor reviews from each country.

The number of TripAdvisor reviews pertaining to hotel establishments within South Africa increased from 2008 to 2016, peaking at 16,948 reviews in 2016 (Figure 5.18). This coincided with the growth of the platform. Subsequently, a gradual decline in review frequency was observed from 2016 to 2019. Notably, the year 2008 registered the lowest number of documented reviews,

totalling 862. This could have been as a result of the global financial crisis that occurred between 2007 to 2008. Early concerns of the H1N1 (Influenza) pandemic were reported in 2008 which also could have impacted tourist travel. The period between 2015 and 2017 witnessed the highest volume of reviews (Figure 5.18). The most substantial increase occurred between 2015 and 2016, with a total of 4,971 reviews added. However, in 2020, the global tourism sector was significantly impacted by the COVID-19 pandemic, leading to widespread lockdown measures in many countries, including South Africa. Consequently, there was a sharp decline of 10,123 reviews observed from 2019 to 2020 (Figure 5.18). Nevertheless, in the years leading up to 2020, a modest increase in review activity was evident.

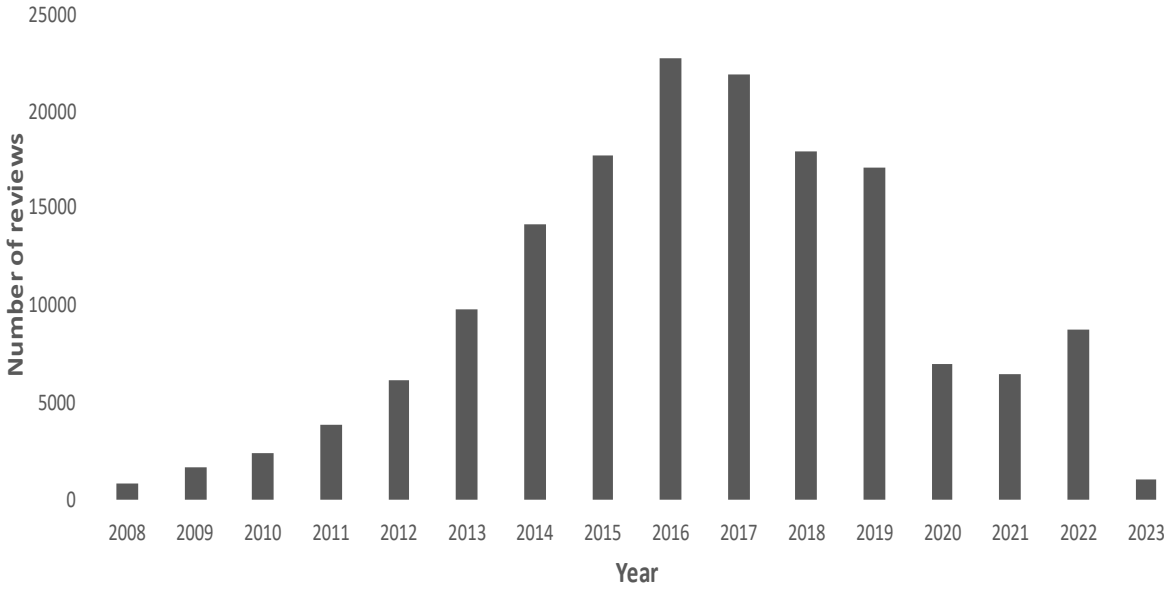



**Figure 5. 18:** Number of TripAdvisor reviews per year from the South African sample.

The bulk of reviews of South African accommodation establishments are authored during January (10.9%, n= 161,280), March (9.3%, n=161,280), and October (9.6%, n=161,280), whereas the

lowest volume of reviews is documented in June (5.9%, n=161,280) and July (6.9%, n=161,280; Figure 5.19).

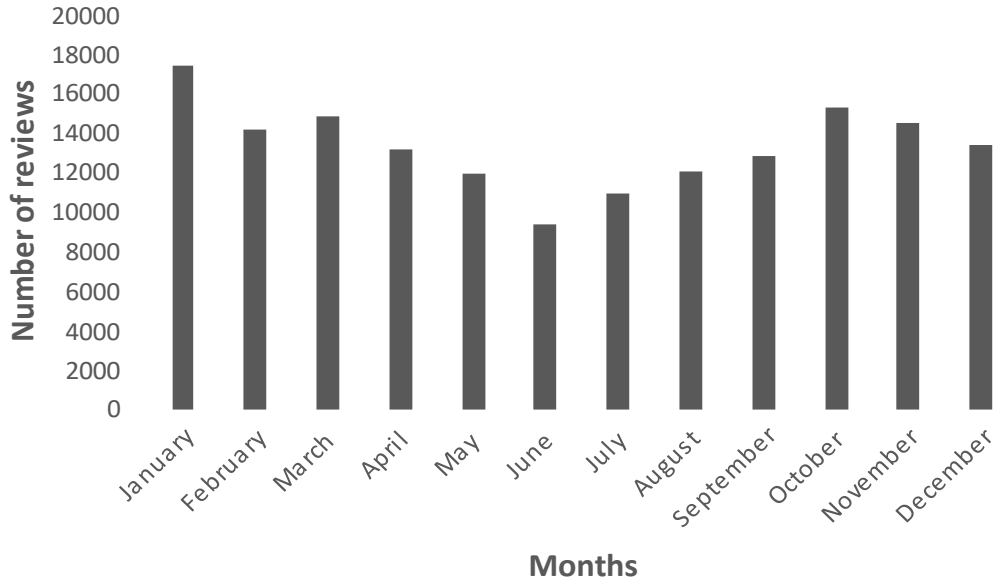



**Figure 5. 19:** Number of Tripadvisor reviews per month from the South African sample.

## 5.4.2 Climatic mentions in reviews

A comprehensive overview of the distribution of climatic or weather-related mentions across various provinces in South Africa is displayed (see Table 5.5). The data is derived from a total of 161,280 reviews web-scraped, out of which 5.6% mentioned climate. Eastern Cape has the second highest proportion within climate mentions at 43.2%, with cold being the most frequently mentioned climate factor (Table 5.5). Despite having a lower number of reviews consulted compared to Western Cape, it has a higher percentage of climate mentions at 6.1%. Free State and North West have similar patterns, with cold being the predominant climate factor mentioned and proportions within climate mentions at 11.5% and 31.7%, respectively.

Gauteng, despite having a considerable number of reviews consulted, only has a small percentage (3.2%) that mention the climate, predominantly cold conditions, contributing to only 9.8% proportion within climate mentions. In KwaZulu-Natal, cold is also the most frequently mentioned factor with an 8.2% proportion within climate mentions. Limpopo and Mpumalanga both have cold as the dominant climatic mention, but Limpopo has a higher proportion within climatic mentions at 33.2%. The Northern Cape stands out as it is the only province where hot is the most frequently mentioned climatic factor. It also has one of the highest percentages of climatic mentions (11.1%) and contributes to more than half (50.4%) of all such discussions in this context.

Lastly, Western Cape, despite having by far the largest number of reviews consulted, only contributes to 9.9% proportion within climatic mentions, with cold conditions being predominantly mentioned. The prominence of "cold" as a frequently mentioned climatic factor across almost all provinces except Northern Cape indicates that reviewers are more inclined to discuss colder weather conditions. This could be attributed to various factors including but not limited to tourism activities associated with colder climates or specific weather events during periods when reviews were written. The deviation from this trend for the Northern Cape could be due to its semi-arid climate making hot weather more noteworthy for reviewers. The high percentage (50.0%) in proportion within climatic mentions suggests that hot weather significantly impacts visitors' experiences in this province. Provinces like Gauteng having low percentages in both categories suggest that either their climates are moderate or other non-climactic factors play more crucial roles in shaping visitors' experiences and feedback.

**Table 5. 5:** Proportion of climatic mentions

| Province | Number of Reviews Consulted | Number of Mentions of Climate | Percentage of Climate Mentions | Most Frequently Mentioned Climate Factor | Proportion within Climate Mentions |
|---|---|---|---|---|---|
| **Eastern Cape** | 24114 | 1461 | 6.06% | cold | 43.19% |
| **Free State** | 4132 | 235 | 5.69% | Hot | 11.49% |
| **Gauteng** | 9746 | 315 | 3.23% | Hot | 9.84% |
| **KwaZulu-Natal** | 18869 | 951 | 5.04% | Hot | 8.20% |
| **Limpopo** | 18930 | 1033 | 5.46% | cold | 33.20% |
| **Mpumalanga** | 8409 | 633 | 7.53% | cold | 32.23% |
| **North West** | 7054 | 625 | 8.86% | cold | 31.68% |
| **Northen Cape** | 4138 | 458 | 11.07% | Hot | 50.44% |
| **Western Cape** | 65888 | 3312 | 5.03% | Hot | 9.87% |
| **Total** | **161280** | **9023** | **5.59%** | | |

Figure 5.20 illustrates the annual distribution of climatic or weather-related mentions in South Africa from 2008 to 2023. A careful analysis reveals a fluctuating yet significant increase in the number of mentions over the years, indicating a growing attention towards climate and weather issues. From 2010 to 2013, there was a gradual increase in climatic mentions per year (Figure 5.20). This period marked an era of increased awareness but had not yet reached its peak. The mentions were relatively low, with less than 400 per year.

A noticeable spike occurred in 2013, where the mentions almost doubled from the previous year. This could be attributed to specific climatic events or policy changes that drew public attention (Figure 5.20). A gradual increase is recorded in 2014 and 2015, which could suggest reduced climatic activities or perhaps a shift in public focus. However, this was short-lived as there was an unprecedented decline in mentions from 2016 to 2018. From 2019 and onwards, there is a sharp decline to levels even lower than those recorded before the peak years (pre-2013).

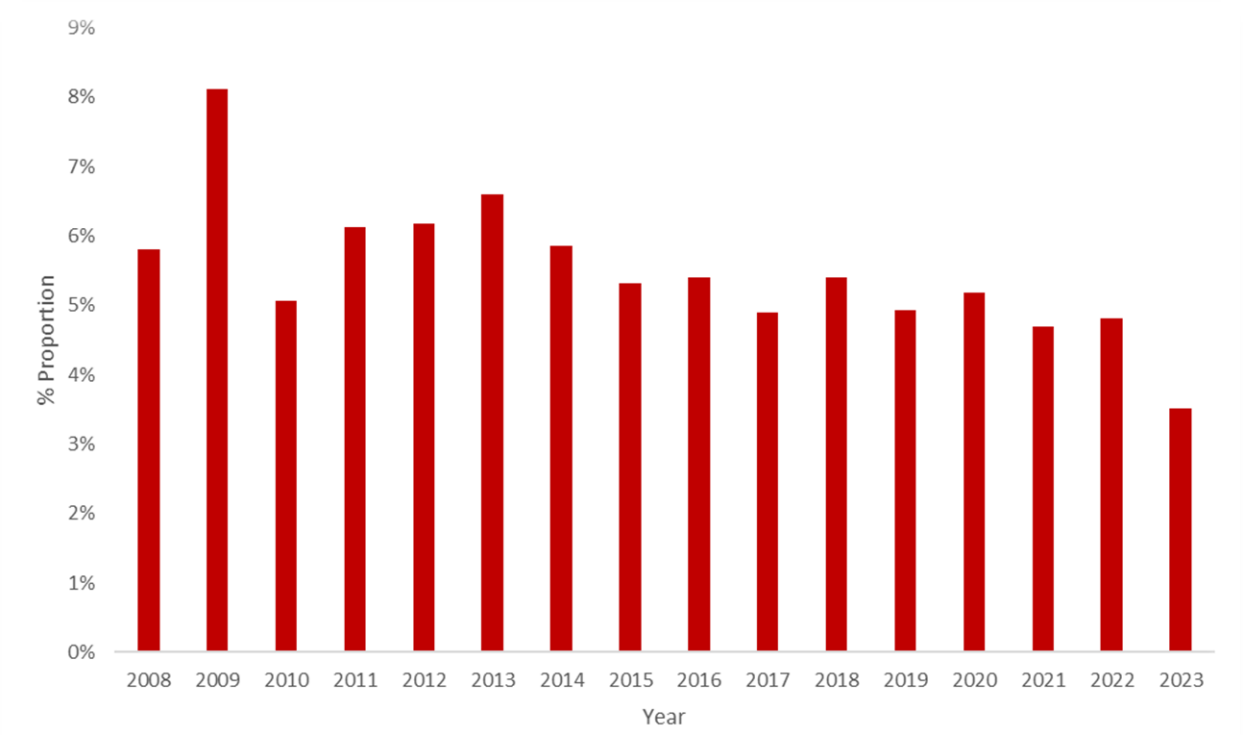



**Figure 5. 20:** Number of climatic mentions per year

There is a percentage proportion of 5.3% of climate related mentions in January. This is followed by a significant drop in February and March, where mentions are reduced. From April to June, a gradual increase is observed. Another peak occurs in July. Although not as pronounced as the peak for January (Figure 5.21).

The trend declines sharply again in August before rising slightly in September. From September, the trend maintains consistency through to December. This pattern suggests that while there are periods of heightened attention to climatic events, there are also periods of relative stability.

The pronounced peaks during January and July may correspond with extreme weather conditions associated with summer and winter seasons, respectively. These could be due to fluctuations in seasonal temperatures, resulting in warmer-than-average summers and cooler-than-average

winters, particularly in the southern hemisphere. Conversely, the reduced mentions during intermediary months could indicate periods of moderate weather conditions (Figure 5.21).

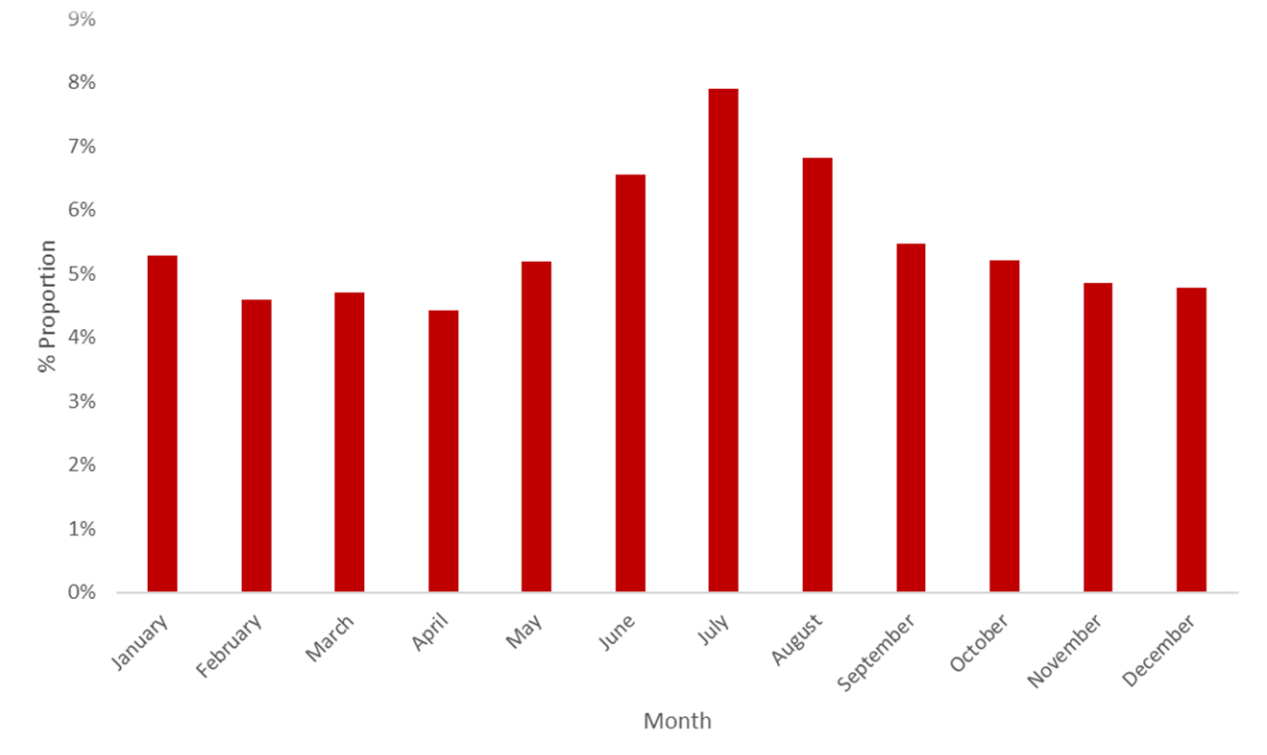



**Figure 5. 21:** Proportion of climatic mentions per month

The total number of climatic mentions is 13,540, distributed among various weather conditions. Cold weather is the most frequently mentioned condition, accounting for a significant 25.6% of all mentions. This prominence suggests that cold weather might have a substantial impact on tourists, warranting frequent discussions or reports. Rain follows with a 13.5% proportion of climatic mentions. Given that rain can influence various aspects of tourist's overall vacation experience often leading to cancelled trips, shortened vacation durations, transportation, and outdoor activities, it is important to acknowledge it.

Other conditions like hot weather (8.6%), sun (5.0%), and wind (3.0%) also contribute to the discussion but are less prominent than cold and rain (Table 5.6). These conditions might be regular occurrences but may not significantly impact daily life as much as colder temperatures or rainfall. Drought, humid conditions, bad weather, good weather, and mist each constitute 1.3%, 0.7%, 0.6%, 0.5% and 0.8% of the total mentions respectively. These low percentages indicate that while these conditions occur, they might not be as prevalent or impactful to warrant frequent tourist discussions. Weather conditions such as heatwave, thunderstorm, hail have less mentions as well ranging from 0-2%, indicating their infrequent occurrence or lesser impact on tourists' daily life in South Africa (Table 5.6).

In summary, the analysis reveals a distinct emphasis on cold and rainy conditions. These weather types are predominantly featured in conversations, indicating a notable interest or concern among tourists. Hot weather, while also a topic of discussion, attracts attention to a lesser extent. This suggests that while hot weather conditions are noteworthy, they may not incite as much conversation or concern as their colder counterparts. Interestingly, other extreme weather conditions such as heatwaves and thunderstorms are discussed less frequently. This could imply that these weather phenomena, despite their severity, may not be as common or impactful on the daily lives of the reviewers, leading to their reduced presence in conversations. This distribution could be reflective of the prevailing climatic conditions in South Africa or public perception and reaction towards these diverse types of weather conditions.

**Table 5. 6:** Proportion of climatic mentions

| Climatic Conditions | Climatic Mentions | Proportion of Climatic Mention |
|---|---|---|
| **cold** | 3472 | 26% |
| **cloud** | 159 | 1% |

| | | |
|---|---|---|
| **drought** | 177 | 1% |
| **fog** | 47 | 0% |
| **Heatwave** | 22 | 0% |
| **Hot** | 1162 | 9% |
| **Humid** | 91 | 1% |
| **Bad Weather** | 87 | 1% |
| **Good Weather** | 67 | 0% |
| **Mist** | 101 | 1% |
| **Rain** | 1821 | 13% |
| **Sun** | 742 | 5% |
| **Thunder** | 53 | 0% |
| **Thunderstorm** | 230 | 2% |
| **Hail** | 22 | 0% |
| **Wind** | 405 | 3% |
| **Total** | 13540 | 100% |

Cold weather is prominently mentioned, with over 25% of the total mentions attributed to it (Figure 5.22). This could indicate a significant impact of cold weather on tourists, as tourists are sensitive to thermal comfort, and as they might be anticipating warmer conditions in what is marketed as 'sunny South Africa'. Rain and Sun also have notable mentions, though significantly less than Cold, at around 13% and 5%, respectively (Figure 5.22). These Figures might suggest that rain and sunny conditions are common but not as impactful or noteworthy as cold weather in public discourse.

Other climatic conditions like cloud, drought, fog, heatwave, humid conditions, bad and good weather, mist, thunder and thunderstorm, hail, and wind have fewer mentions, at below 5%. This low frequency could be due to their infrequent occurrence or lesser impact on daily life.

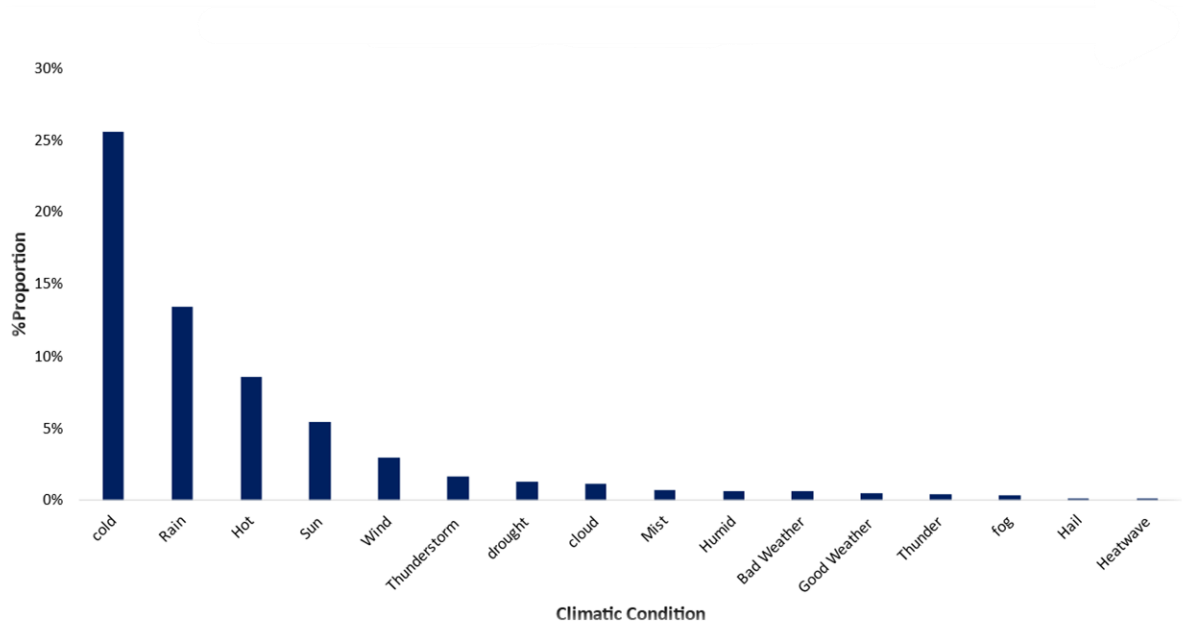



**Figure 5. 22:** Climatic mentions

There is considerable variation in climatic conditions mentioned in reviews across the nine provinces (Figure 5.23). The highest proportion of climatic mentions is calculated for the Eastern Cape at over 40%, suggesting that this region experiences diverse and pronounced climatic phenomena. This could potentially impact tourist activities, necessitating adaptive measures to mitigate adverse effects.

In contrast, Free State (11.5%), Gauteng (9.8%) and KwaZulu-Natal (8.2%), have relatively lower mentions below 20% (Figure 5.23). This indicates milder climatic conditions which might offer a more comfortable environment for tourists.

Mpumalanga (32.2%), North West (31.7%), Northern Cape (50.4%) exhibit an intermediate level of climatic mentions ranging from 20% to 40% (Figure 5.23). These regions may experience moderate climatic variations requiring consideration in tourist planning and activities.

The Eastern Cape region had a high frequency in climatic mentions nearing 60%, indicating pronounced climatic conditions similar to the Northern Cape (Figure 5.23). Tourists visiting these areas may need to be particularly sensitive to weather patterns and adapt their plans accordingly.

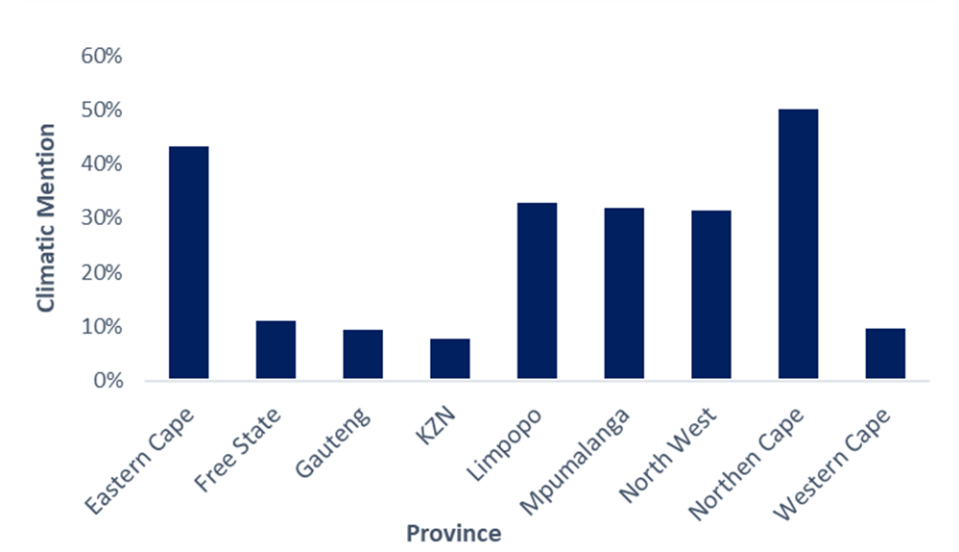



**Figure 5. 23:** Climatic conditions per province

Figure 5.24 provides an overview of the distribution of bad weather-related tourist mentions across various provinces, with each province contributing differently to the total number of mentions. The Free State (1.7%) had the highest contribution among the other provinces. This is followed by the Eastern Cape and Gauteng, which also have a significant 1.5% and 1.6% proportion of climatic mentions respectively (Figure 5.24). These two provinces appear to be the most frequently mentioned in discussions related to bad weather. KwaZulu-Natal and Western Cape, each contributing 1.0% and 1.1% respectively, indicate a moderate level of bad weather-related tourist

mentions in these regions (Figure 5.24). This suggests that while these provinces do experience bad weather, they may not be as severely affected as regions such as the Eastern Cape and Gauteng, or perhaps they have effective measures in place to manage these conditions.

Limpopo, Mpumalanga, and the North West regions, with contributions below 1%, have fewer mentions. This could indicate that these provinces experience less severe weather conditions, or perhaps they are less popular tourist destinations. Finally, the Northern Cape region may have negligible contributions. This could suggest that this province either experiences fewer severe weather conditions or are less frequently visited by tourists.

This distribution could be attributed to a range of factors including the geographical location, climate conditions, and tourist attractions available in each province. The Eastern Cape and Gauteng having higher percentages might be experiencing more frequent or severe bad weather conditions affecting tourists' experiences or these provinces could potentially have more tourists reporting about the weather due to their popularity.

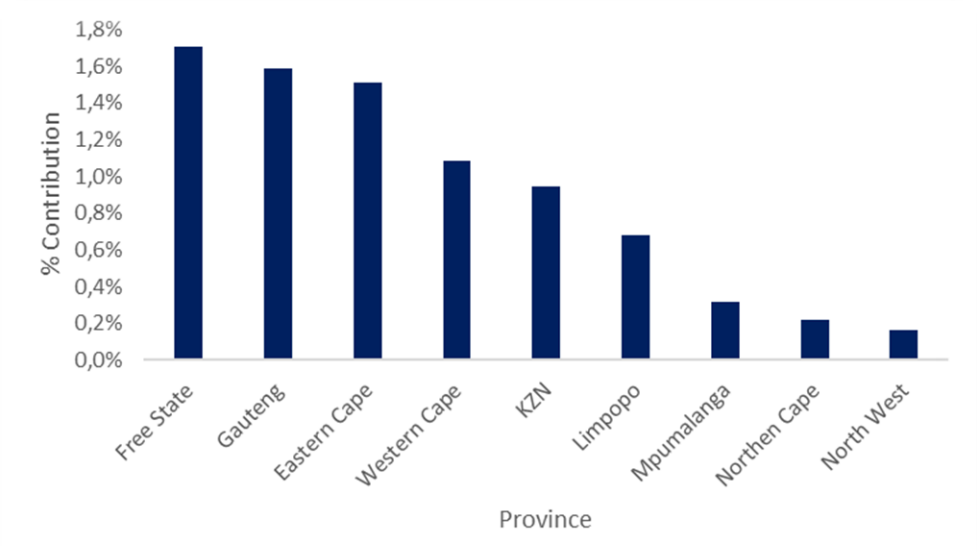



**Figure 5. 24:** Bad weather-related mentions in South Africa

The Eastern Cape has the second-highest percentage of cold-related weather mentions, at around 43.2% (Figure 5.25). This could be indicative of frequent cold weather conditions or that tourists are more vocal about such weather conditions. Free State (45.5%) also has a high percentage of cold-related weather mentions (Figure 5.25). This could indicate that the Free State region experiences significant cold weather conditions.

KwaZulu-Natal has a moderate percentage of cold-related weather mentions, approximately 30%. Gauteng, on the contrary, has the highest percentage of cold-related weather mentions, above 50% (Figure 5.25). This suggests that this region might experience the coldest weather conditions, given that the majority of cold weather mentions originate from reviewers who have visited hotel establishments in that area.

Limpopo (33.2%), Mpumalanga (32.2%) and the North West (31.7%) also have moderate percentages of cold-related weather mentions. Meanwhile, the Northen Cape and KwaZulu-Natal have the lowest percentage of cold-related weather mentions. This could be due to less frequent cold weather or fewer mentions within tourist reviews from these areas (Figure 5.25)

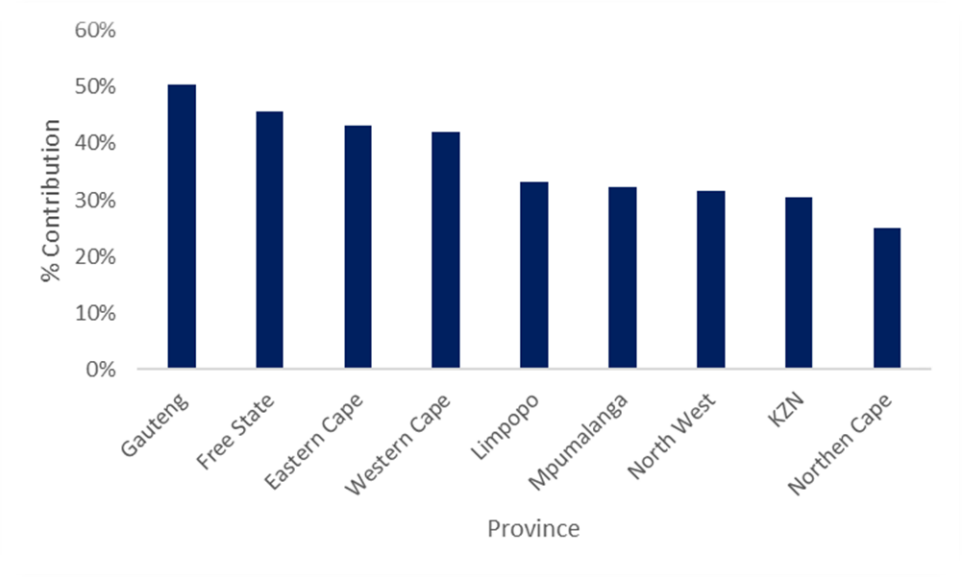



**Figure 5. 25:** Cold weather-related conditions in South Africa

KwaZulu-Natal, Limpopo, and Western Cape are the most frequently mentioned in reviews related to drought conditions, each contributing to 2.6%, 3.1% and 2.5% of the mentions respectively (Figure 5.26). This high level of mentions could be indicative of these provinces experiencing more severe or frequent drought conditions.

By contrast, moderate mentions of drought related climatic conditions are evident within the Gauteng (1%), and North West (1.4%) regions, respectively. This suggests that while drought conditions may be present in these provinces, they are not as severe or frequent as in KwaZulu-Natal, Limpopo, and the Western Cape (Figure 5.26), or the impacts of drought are less well detected by the tourists visiting these destinations. Conversely, Eastern Cape (1%), Free State (0.9%), Mpumalanga (0.6%), and Northern Cape (0.9%) have the least mentions.

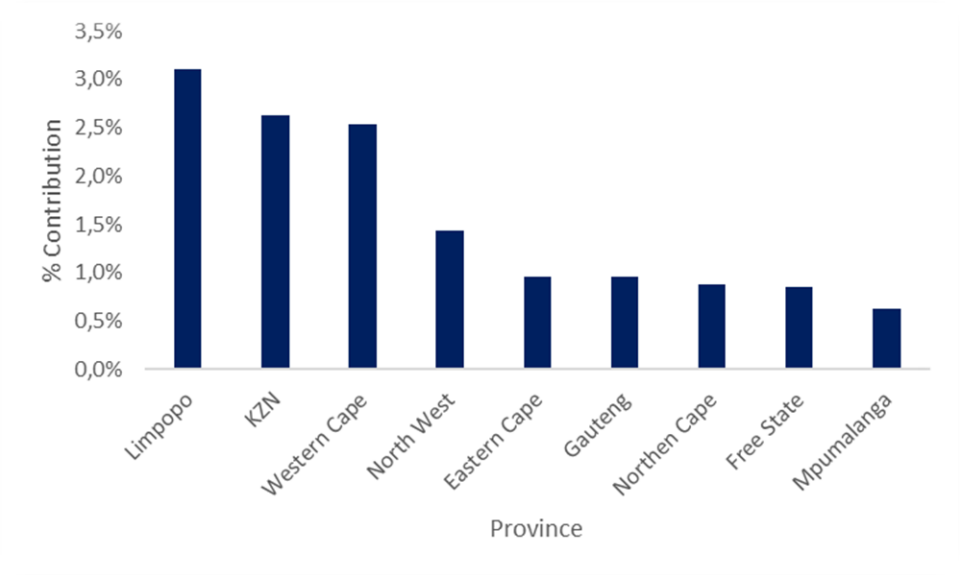



**Figure 5. 26:** Drought weather-related mentions in South Africa

Gauteng (1.0%) and Mpumalanga (1.0%) are most frequently mentioned in reviews related to fog, indicating that this weather condition is significant in these areas. This could be due to geographical factors or climate conditions that favour the formation of fog. This is followed by the Western Cape (0.7%) and Eastern Cape (0.5%) regions (Figure 5.27).

By contrast, the Free State (0%) and North West (0%) have the lowest proportion of mentions, suggesting that fog is less common or less reported in these provinces. Moderate levels of fog-related mentions are identified for KwaZulu-Natal (0.4%), Limpopo (0.1%) and Northern Cape (0.4%; Figure 5.27). This suggests that while fog is not the most dominant weather condition in these provinces, it does occur with some regularity. This variation in mentions could reflect the varying frequency and intensity of fog conditions across different provinces (Figure 5.27).

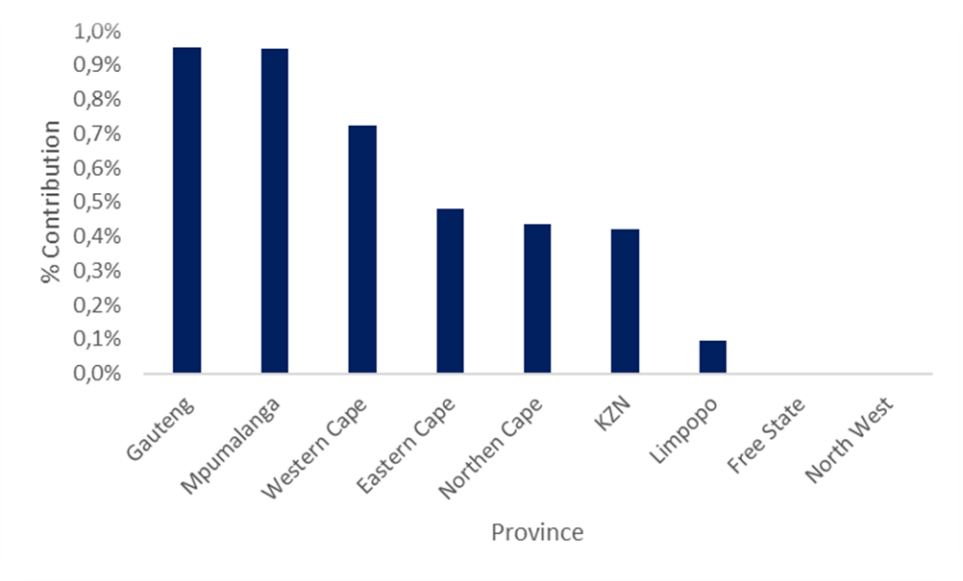



**Figure 5. 27:** Fog weather-related mentions in South Africa

In the Eastern Cape (0.7%), there is a moderate percentage of good weather mentions (Figure 5.28). This could indicate a moderate level of satisfaction with the weather among visitors. By contrast, Limpopo has the lowest percentage (0.2%) among all provinces, almost negligible (Figure 5.28). This could suggest that good weather is not a common topic of discussion in this region, possibly due to less favourable weather conditions. Gauteng (0.6%) has a slightly higher percentage than the Free State but is still quite low, under 1%. This suggests that while good weather might be slightly more common than in Gauteng, it is still not a major topic of discussion within reviews. KwaZulu-Natal (0.8%) being the third highest, experiences a significant increase in good weather mentions (Figure 5.28).

Mpumalanga (0.6%) also has low good weather mentions but is slightly higher than Limpopo. The Northern Cape (0.4%) and Free State (0.4%), however, have lower mentions of good weather (Figure 5.28). This could be due to less favourable weather conditions leading to fewer discussions about the weather. Finally, the Western Cape (1.0%) and North West (1.0%) regions have the

highest mentions of good weather within the regions, surpassing all other provinces. This suggests that the Western Cape and North West have the most favourable weather conditions.

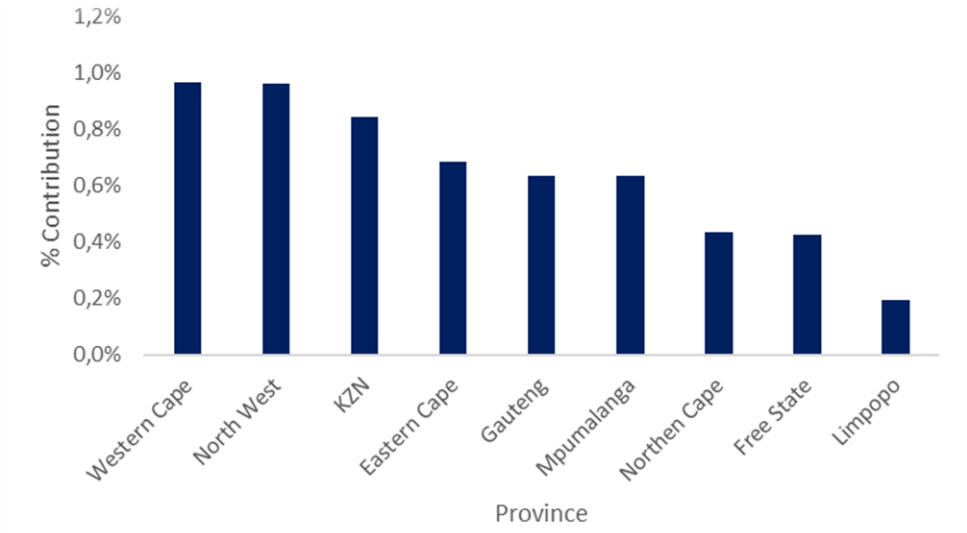



**Figure 5. 28:** Good weather-related mentions in South Africa

The Northern Cape (0.4%) and Western Cape (0.4%) regions, are the highest in hail weather related mentions, indicating that hail is a significant weather concern in this region (Figure 5.29). The Western Cape has a moderate percentage of hail-related weather mentions, suggesting that hail is a common weather phenomenon here, but not as prevalent as in the Northern Cape. Similarly, the Free State (0.3%) and Gauteng (0.3%) regions also have moderate percentages of hail-related weather mentions, indicating that hail is a common weather phenomenon in these regions as well (Figure 5.29). KwaZulu-Natal (0.0%), Limpopo (0.0%) and Mpumalanga (0.0%) alternatively, have relatively low to zero percentages of hail-related weather mentions. This indicates that hail is less of a concern amongst tourists in these regions compared to the Western Cape and Northern Cape.

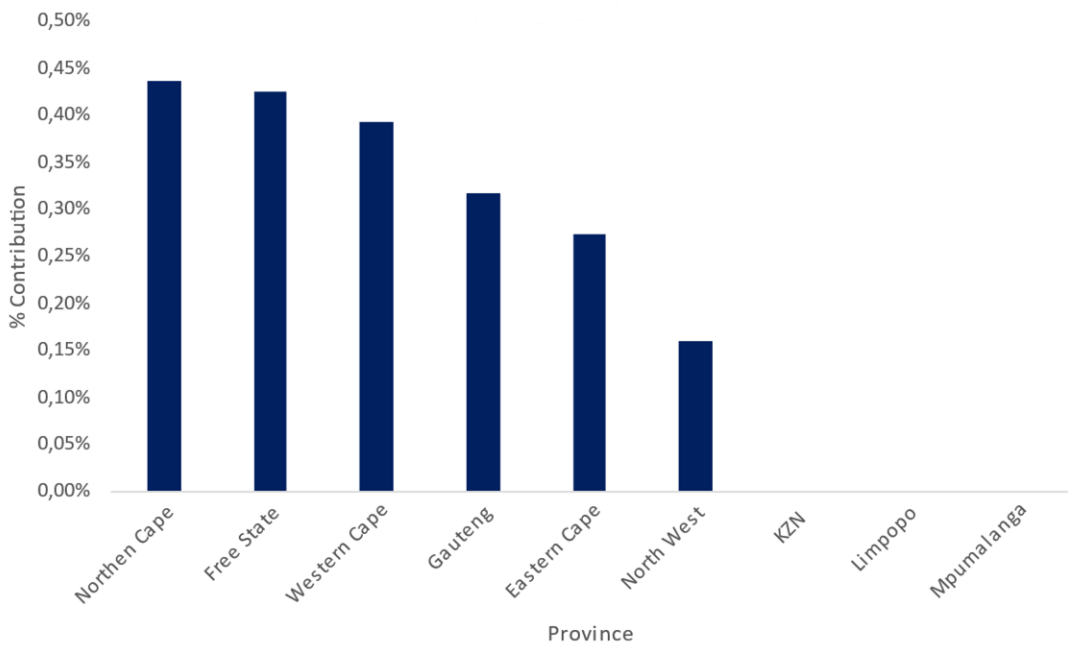



**Figure 5. 29:** Hail weather-related mentions in South Africa

North West (0.6%) and the Northern Cape (0.7%) almost have similar contributions to heatwave-related mentions both below 1% (Figure 5.30). This suggests that these regions have relatively the highest instance of heatwave mentions. This is then followed by the Mpumalanga region with a contribution of 0.5% of heatwave-related conditions. The Free State (0.4%), Gauteng (0.3%) and Western Cape (0.3%) regions all have similar contributions to heatwave related mentions (Figure 5.30).

Limpopo (0.1%) has less contributions to heatwave-related mentions. The KwaZulu-Natal and Eastern Cape regions, conversely, have negligible or zero contributions, indicating that heatwave-related weather mentions are almost non-existent within tourist reviews (Figure 5.30).

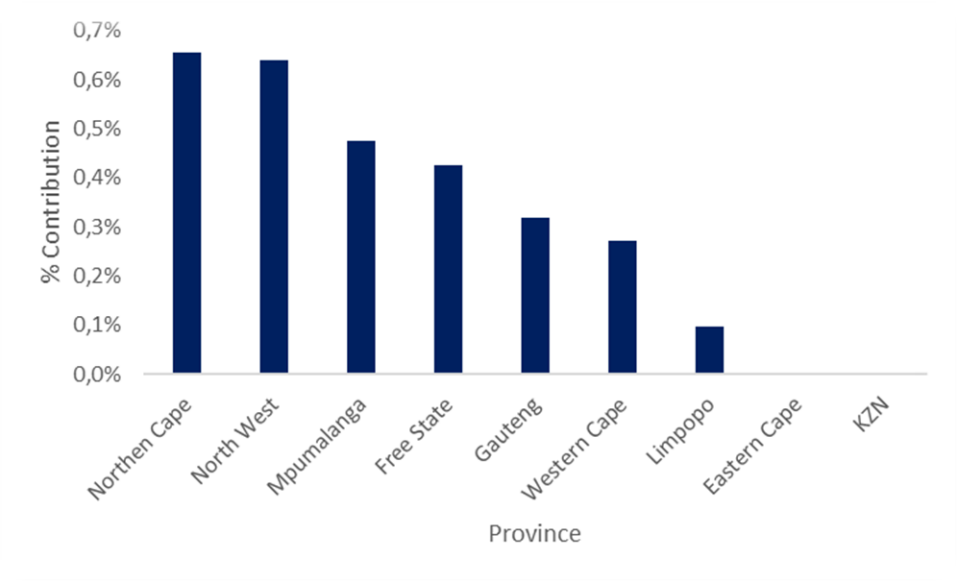



**Figure 5. 30:** Heatwave weather-related mentions in South Africa

The Northern Cape in South Africa is particularly notable for its weather, as it accounts for over 50.4% of all mentions of hot weather (Figure 5.31). This is likely attributable to the semi-arid to arid climate of the region, which is known for its hot, dry summers. By contrast, the other provinces, namely, Eastern Cape (11.5%), Free State 11.5%), Gauteng (9.8%), KwaZulu-Natal (8.2%), Limpopo (16.1%), Mpumalanga (12.0%) and North West (7.0%) each account for less than 20% of hot weather mentions (Figure 5.31). This diversity in climate could be the reason for the lower percentage of hot weather mentions in these regions.

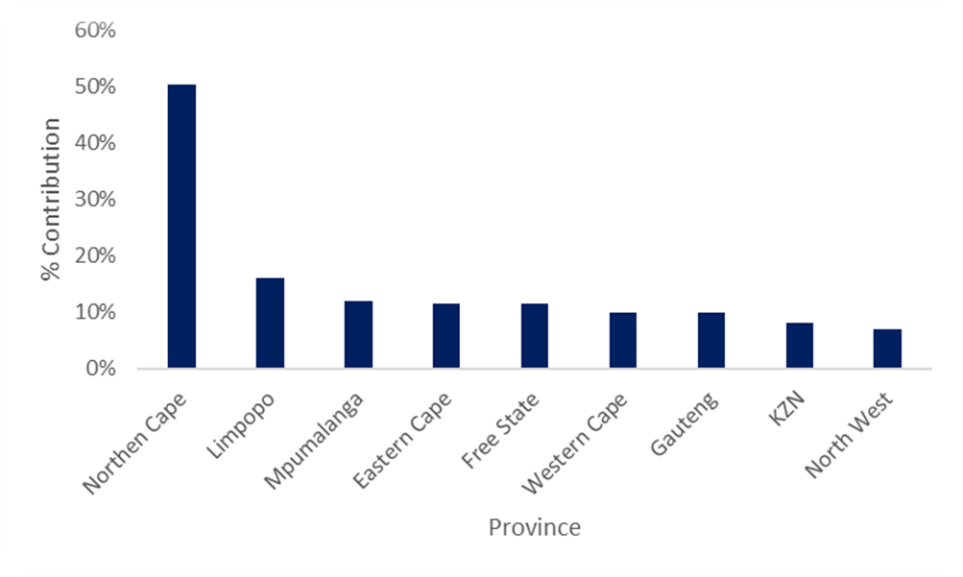



**Figure 5. 31:** Hot weather-related mentions in South Africa

KwaZulu-Natal is distinct among the South African provinces for its high percentage of humid weather mentions, which is approximately 5.2% (Figure 5.32). This could be interpreted as an indication that most tourists visiting accommodation establishments in KwaZulu-Natal experience more humid weather conditions, or alternatively, it could suggest that there is a heightened level of reporting about humidity within tourist reviews within this province. Either of these are possible as KwaZulu-Natal does experience high levels of relative humidity year-round.

Conversely, Eastern Cape (0.9%), Free State (0.9%), Gauteng (1.3%), Limpopo (0.7%), Mpumalanga (0.3%), North West (0.0%), Northern Cape (0.0%), and Western Cape (0.4%) all have significantly lower mentions of humid weather, all falling below 2% (Figure 5.32). This could imply that these provinces either experience less humid weather conditions, or there are less reviews that mentions humidity weather conditions in these regions.

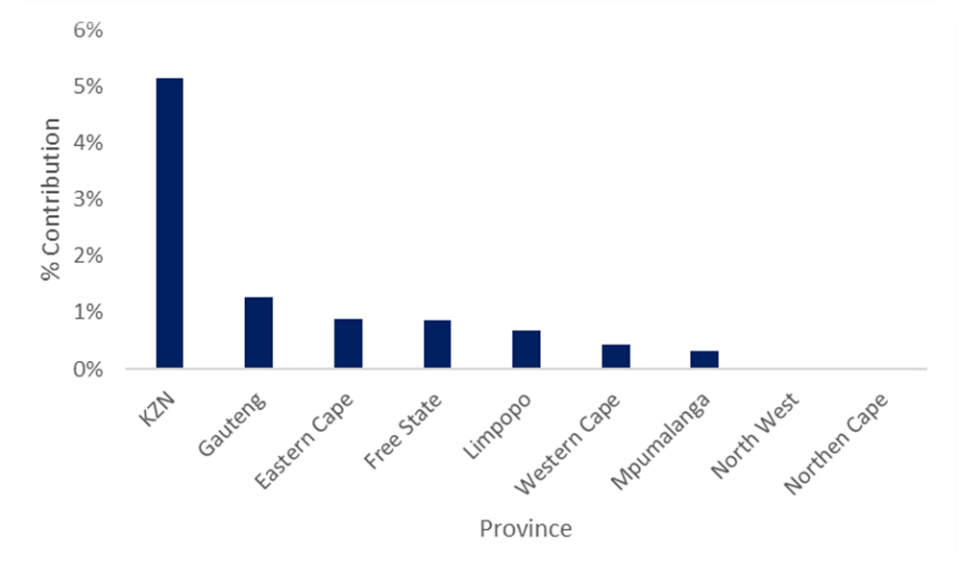



**Figure 5. 32:** Humid weather-related mentions in South Africa

Mpumalanga is characterised as having the highest percentage of mist-related weather mentions among all provinces, contributing to 2.8% (Figure 5.33). This could suggest a higher prevalence or awareness of misty weather conditions in this region. Following Mpumalanga, the Western Cape contributes 1.7% to the mist-related weather mentions. This implies that misty conditions are also relatively common or frequently reported within tourist reviews in this province.

KwaZulu-Natal ranks third, with 1.1% of the mist-related weather mentions. This indicates a lower, but still notable, occurrence or reporting of misty weather in KwaZulu-Natal (Figure 5.33). The remaining provinces, namely, Eastern Cape, Free State, Gauteng, Limpopo, Mpumalanga, and Northern Cape, have very low percentages of mist-related weather mentions. This could suggest that misty conditions are less common, less noticed, or less documented within tourist reviews in these provinces.

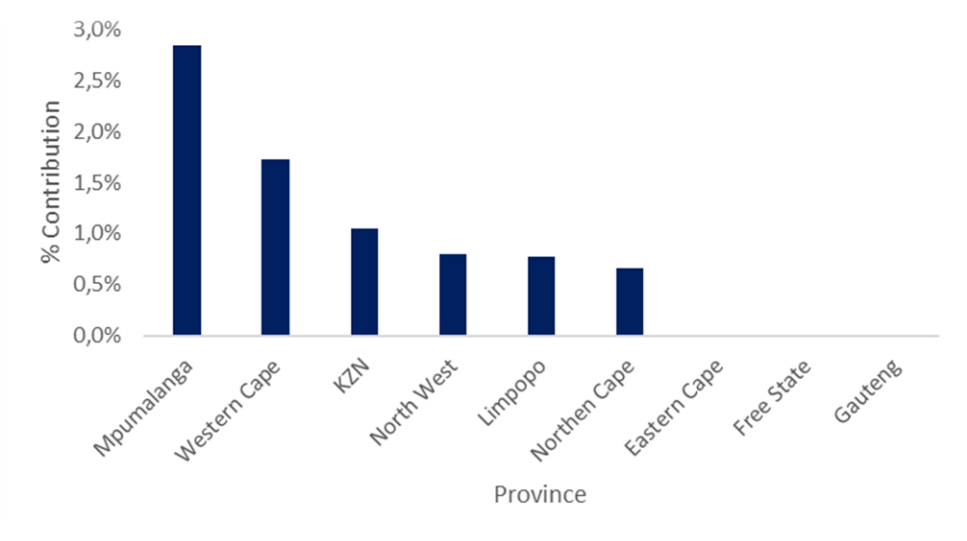



**Figure 5. 33:** Mist weather-related mentions in South Africa

The Eastern Cape is at the forefront with approximately 28.7% of rainfall mentions (Figure 5.34). This could be an indication that it experiences more rainfall, or that discussions related to rain are more prevalent in this province. Following closely is Limpopo, with 28.5% (Figure 5.34). This suggests a significant amount of rain-related tourist reviews documented in this area.

KwaZulu-Natal and Western Cape contribute similarly to the discussion, each above 20%. Mpumalanga, however, contributes slightly less than 15%. North West, with just above 20%, indicates a slightly higher frequency of rain-related mentions in contrast to Gauteng. The Northern Cape has the least mentions, below 6%, suggesting that rain is a less common occurrence or topic of discussion within reviews in this area (Figure 5.34).

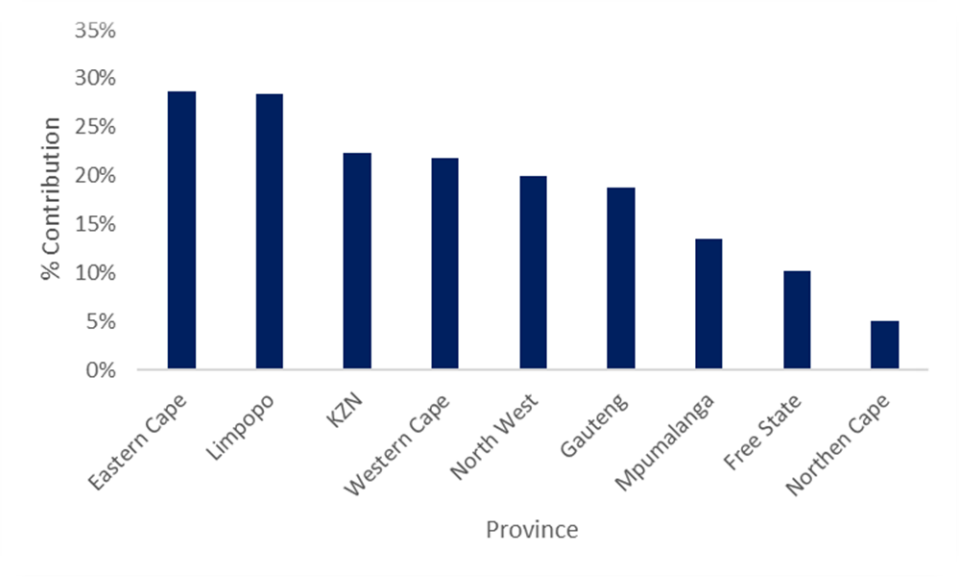



Figure 5. 34**:** Rain weather-related mentions in South Africa

The North West region has the highest percentage of sun-related mentions, with 29.3% (Figure 5.35). This data suggests that the North West is most frequently associated with sunny weather. In comparison, Mpumalanga (20.1%), and KwaZulu-Natal (18.0%) also have significant sun-related mentions, with their percentages surpassing 15% (Figure 5.35). This indicates that sunny weather is a common characteristic associated with these areas.

On the contrary, the Eastern Cape (3.2%) and Western Cape (4.7%) have the lowest percentages of sun-related mentions, falling under 5%. This suggests that sunny weather is less commonly associated or reported in these areas. Lastly, the provinces of Free State (19.2%), Northern Cape (9.4%), and Limpopo (11.2%) have moderate sun-related mentions (Figure 5.35).

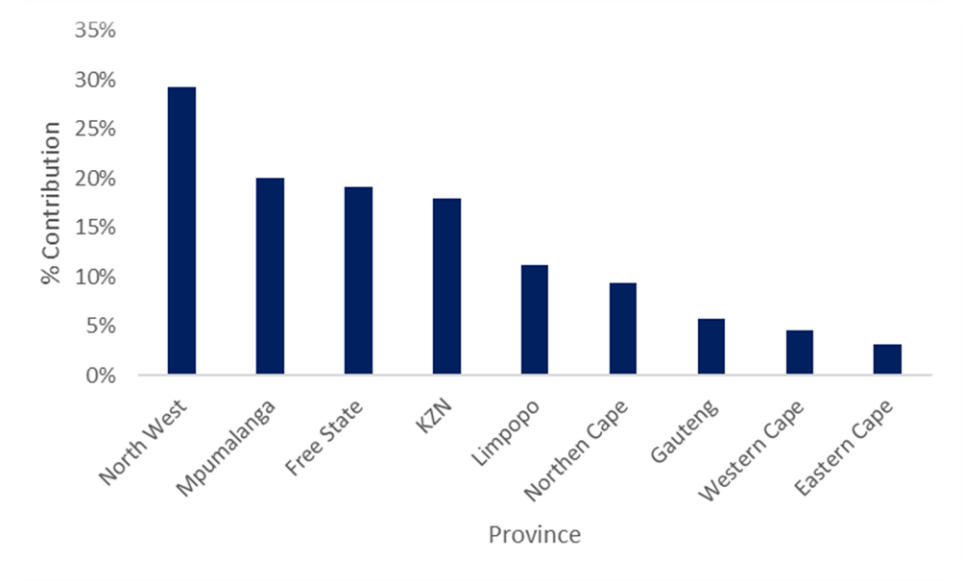



**Figure 5. 35:** Sun-related weather mentions in South Africa

The Eastern Cape leads with the highest percentage of thunderstorm mentions, contributing 6.8% (Figure 5.36). This could indicate a higher frequency of thunderstorms or more discussions about this weather condition in this region. KwaZulu-Natal (5.4%), Gauteng (4.4%) and the Western Cape (3.7%), indicate a moderate level of thunderstorm activity as well as tourist reviews mentioning thunderstorm phenomena (Figure 5.36).

Free State, Limpopo, Mpumalanga, and North West have moderate contributions, ranging between 2% to 4%. This suggests a relatively lower level of thunderstorm mentions compared to the Eastern Cape, indicating less frequent thunderstorm activity. Lastly, the Northern Cape region has the least mentions, contributing less than 2% (Figure 5.36). This could imply fewer thunderstorms or less discussion about this weather condition in these regions, suggesting a lower frequency of thunderstorm-related weather conditions.

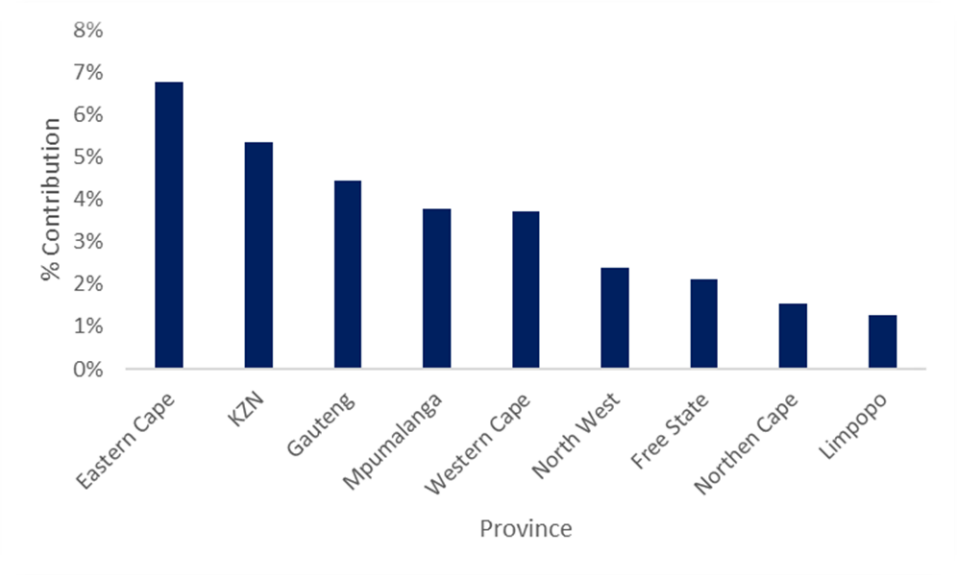



**Figure 5. 36:** Thunderstorm weather-related mentions in South Africa

Mpumalanga (8.9%) has a higher frequency of wind-related weather mentions (Figure 5.37). This suggests that wind might be a significant weather concern in this area. The Western Cape (6.5%) and Free State (6.4%) regions also have significant mentions of wind-related weather, as indicated by their respective contributions (Figure 5.37). This implies that wind is a notable weather factor in these regions as well.

Conversely, Gauteng (1.6%) and Eastern Cape (0.6%) have exceptionally low contributions, under 2%, indicating minimal mentions of wind-related weather (Figure 5.37). This suggests that wind is not a significant weather concern in these areas. Lastly, KwaZulu-Natal (3.2%), Limpopo (3.1%), and Northern Cape (4.4%) have moderate mentions of wind-related weather. This indicates that wind might be a periodic or seasonal concern in these areas.

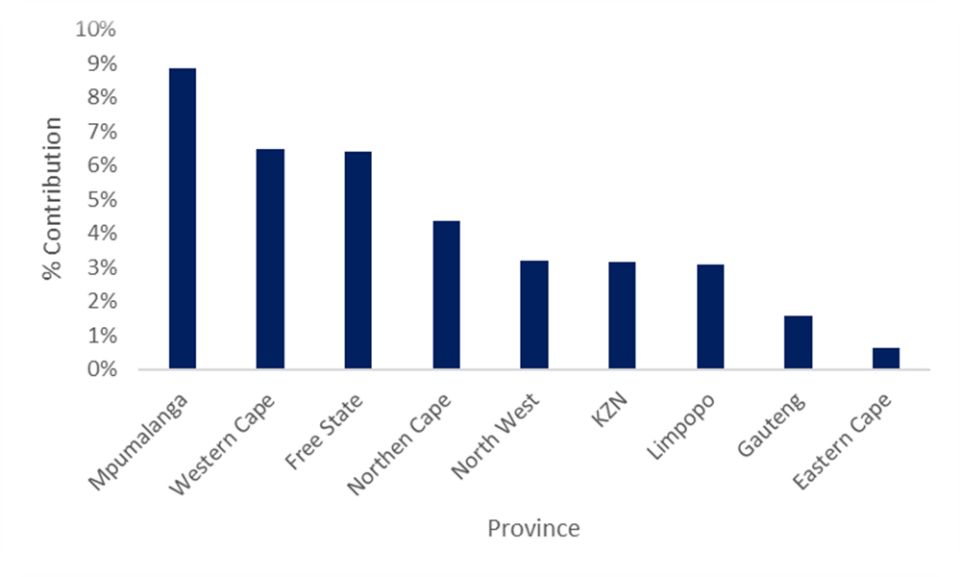



**Figure 5. 37:** Wind-related weather mentions in South Africa

Western Cape demonstrates the highest frequency of cloud weather related mentions with an overall contribution of 3.0% (Figure 5.38). The high frequency of cloud mentions indicates the type of climate tourists are susceptible to when visiting various hotel establishments throughout the western cape. This is then followed by KwaZulu-Natal (1.4%) and Mpumalanga (1.3%) with contributions less than 3% further indicating moderate cloud mentions throughout those regions (Figure 5.38). The Northern Cape had the lowest mentions with less than 1% of cloud related weather mentions within tourist reviews.

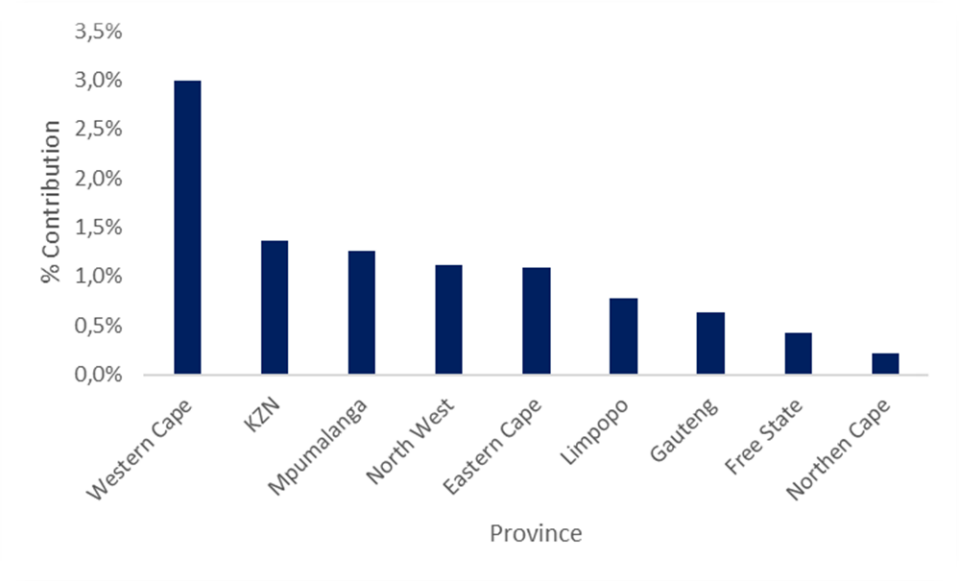



**Figure 5. 38:** Cloud-related weather mentions in South Africa

## 5.4 Conclusion

This research aimed to conduct a comparative analysis of TripAdvisor reviews obtained by manual collection and web-scraping methods. No statistically significant differences were noted between both approaches for the overall climatic mentions. It was determined that web-scraping is effective in being less tedious, obtaining vast amounts of data in real-time and therefore can be accepted a good source of obtaining data.

The sentiment analysis undertaken in this research provided insights into the attitudes of diverse guests towards various weather conditions. It was observed that apart from instances of good weather, mist and hail, all other weather scenarios elicited predominantly negative to very negative sentiments. Regional variations in the distribution and proportion of different weather mentions were analysed to gain a deeper understanding of how weather and climatic conditions influence

tourist perceptions based on their evaluations. Among the prevailing weather conditions in South Africa, cold, hot, and rainy weather predominated, with notably higher proportions observed.

# Chapter 6: Discussion

## 6.1 Introduction

This research had the main objective of exploring the possibilities of using web-scraping methods to enable a wider database of TripAdvisor reviews to be used in exploring the sensitives of tourists to the weather. The application of methods in big data collation and analysis from tourist reviews was conducted. This study aimed to test the efficacy of the web-scraping approach in comparison to the manual approach, in assessing the climatic sensitivity of tourists in South Africa through TripAdvisor reviews. In this section, the efficacy of web-scraping in scientific data collection techniques, the different scientific fields the technique has been used and whether or not it was successful will be discussed. The comparison of both the manual and automated approaches, and finally, the limitations of web-scraping will be thoroughly discussed. The study was then expanded to accommodate all accommodation establishments throughout South Africa. The utilisation of sentiment analysis was employed to further understand how weather influences the overall satisfaction or the general contentment of tourists regarding weather experiences.

## 6.2 Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies

### 6.2.1 Web-scraping

#### *6.2.1.1 The Efficacy of Web-scraping*

The study employed the utilisation of web-scraping in automating the collection of data from TripAdvisor. Big data is a term that originated in order to describe volume of data that is available

in the world, particularly online (Khasan et al., 2023). The data is in most cases unstructured; however, it is a valuable resource for data analysis and visualisation (Faccia et al., 2022). It necessitates a notable quantity of effort to organise (Chen, 2019). Web-scraping is a method that has been developed to collect and organise this huge volume of data. Web-scraping is the process of developing computer software that will automatically collect, parse, and organise data from the web (vanden Broucke and Baesens, 2018). Web-scraping, in essence, is the process of systematically extracting data from a variety of user websites and transferring it to a database for storage in either an unstructured or structured format (Mitchell, 2018). In general, web-scraping can be accomplished by developing an automated application known as a virtual scraper (Mitchell, 2018). The first objective of the study was to develop a web-scraping tool that could automate big data from TripAdvisor, to gain more knowledge on the sensitivity of tourist to the climate in South Africa. The overall process of developing the tool was successful as results were generated that validate the technique of web-scraping against manual collection. However, there were challenges that limited the scale of data collection. These challenges included limitations in scraping reviews due to restrictions from the website of TripAdvisor, the complexity of filtering relevant weather-related content, and the variability in how users express their experiences with weather. As a result, although the automation allowed for more efficient data collection, the number of relevant reviews was not as high as initially anticipated. This indicates that while web-scraping is effective, the scope and accuracy of results are dependent on the quality of available data and the limitations of the platform. Nevertheless, through this technique a broader understanding of how tourists perceive weather and are affected by it was analysed due to the success of automating TripAdvisor reviews through web-scraping. In order to investigate travellers' opinions regarding hotel service (Levy et al., 2013), transportation (Ahmed et al., 2020),

or overall sentiment (Siew et al., 2018), various studies have implemented web-scraping techniques. The implementation of web-scraping techniques, including sentiment analysis and text mining, was employed in two studies (Terziyska and Damyanova, 2020; Barbierato et al., 2021) that examined the critical components required for the success of wine tourism in Italy. The data utilised presented challenges; processing unstructured data, which comprises 80% of the data available online, can be challenging. Assessing valuable insights from this unstructured format requires sophisticated methods and tools, and this was a major problem in the development of the web-scraping tool that was utilised in this thesis. The paper employed the Fanpage Karma (FpK) web-scraping tool. While recognising that various tools entail inherent limitations, the choice of a tool must correspond to the specific goals of the scraping task, potentially constraining flexibility. For instance, the code was word specific in this thesis, which brought about the integration of manual work when analysing the data. While tools like FpK facilitate scraping for individuals requiring coding skills, more sophisticated customisations, such as integrating scraping tools with other platforms, necessitate programming expertise. The study addressed legal issues, highlighting the necessity of ensuring that data scraping complies with privacy laws and does not obtain information without consent. However, the techniques implemented were more advantageous than conventional surveys (Barbierato et al., 2021).

### 6.2.2 The Application of Web-scraping

Web-scraping is particularly suitable for exploratory research. It allows researchers to understand general patterns and gather preliminary evidence that can inform more detailed studies (Brenning and Henn, 2023). This flexibility makes it an invaluable tool for initial hypothesis generation and trend analysis. Based on this information the technique has been applied in various fields such as

business intelligence (Bhatt et al., 2023), tourism (Adhinugroho et al., 2020; Pereira et al., 2023; Barbera et al., 2023), customer behaviours and hospitality (Han and Anderson, 2021) and e-commerce (Valova et al., 2023).

The application of web-scraping in the field of biometeorology has been on the rise of late (Saarinen *et al*., 2022). This approach has been used in some studies to automate data from online pages such as X to examine climate change related research (Al-Saqaf and Berglez, 2019; Cecinati *et al*., 2019). Other studies applied the utilisation of web-scraping to mainly extract and analyse tourists' reviews from various online platforms (Chu et al., 2022). This study applied web-scraping as a means of collecting big data in order to validate the technique against manual collection and to further analyse the relations that exists between weather and tourists in South Africa as a whole. Through exploring the percentage differences in climatic mentions overall and per location, slight variations were evident ranging from cold climatic mentions being referenced the most (2.2%, p=0.72), to the least mentioned climatic reference being mist (1.1%, p=0.78), however, these variations were small and in most instances were statistically insignificant. It was found that all the percentage differences between the climatic references in both approaches did not show any statistical significance when comparing the results obtained (Table 5.6). Overall, this means it would be reasonable to use web-scraping in the place of a manual approach. By employing an aspect-based sentiment analysis model grounded in the Bidirectional Encoder Representation from Transformers (BERT), researchers (Chu et al., 2022) effectively demonstrate the potential of web-scraping in sentiment analysis The methodology of the study, which includes iterative translation for dataset enhancement and classification using the Term Frequency-Inverse Document Frequency (TF-IDF) method, illustrates the comprehensive approach necessary for extracting nuanced insights from unstructured data (Chu et al., 2022). A study conducted by Minor et al.

(2023) investigates the relationship between meteorological conditions and social media activity, applying web-scraping to collect a large dataset of over three and a half billion posts from Facebook and Twitter. By applying climate econometric methods, the researchers establish a clear link between adverse weather conditions and increased social media use, providing insights into how external environmental factors influence online behaviour. However, the application of web-scraping within the field of tourism and climate change globally and within the southern part of Africa remains scarce (Saarinen *et al*., 2022). Barbera et al. (2023) used web-scraping techniques to enhance the data acquisition process from TripAdvisor. Through the use of web-scraping techniques, the researchers successfully amassed data that could aid them in interpreting changes in tourism trends post-pandemic from the social sites of TripAdvisor. This data included multiple elements, such as the performance of a competitor, and consumer preferences (Barbera et al., 2023). This enables individuals to acquire information and perspectives that could have remained overlooked or proven challenging to gather using traditional manual data collecting techniques (Mustapha et al., 2024; Tjaden, 2023). This research emphasises the use of web-scraping as a multifaceted instrument for gathering large-scale data from social media sites such as TripAdvisor.

It was noted that ‘cold’ climatic references were mentioned the most throughout both approaches. Port Nolloth had the highest climatic mentions for both approaches. Knysna had the most TripAdvisor reviews but had the lowest climatic references for both approaches. This further indicates that web-scraping can be trusted as a good data collection source as the results indicate similarities between both manual and web-scraping methods. It was found that though the use of artificial intelligence and machine learning, the automation of data from TripAdvisor is possible, more efficient, and less tedious, as the analysis of a large proportion of reviews manually obtained (often over 5000) can be prone to human error (Landers et al., 2016; Karthikeyan et al., 2019).

These tools have been applied to climate change research, particularly on Twitter, and developments have been made in sentiment analysis using code to classify emotions (Al-Saqaf and Berglez, 2019; Cecinati et al., 2019; Sailunaz and Alhajj, 2019). According to Karthikeyan et al. (2019) web-scraping can be classified as a very fast and effective data collection method and moreover can cover a large number of websites within a short time period. With web-scraping, researchers can extract data in a structured format, making it easy to analyse and compare (Karthikeyan et al., 2019).

## 6.3 The Evaluation of Tourist's Weather Experiences from Tripadvisor Reviews

The findings obtained from the research also supplement existing literature that focus on the utilisation of tourist reviews to effectively analyse the effect of weather on tourist's sentiments through a qualitative approach (de Freitas, 1990; Becken and Wilson, 2010; Jeuring, 2013). The various ways in which weather and climate affects tourists' general contentment with regards to various tourism destinations provides valuable insight into the complex relationship that exists between tourism, climate and weather. Moreover, this research provides and contributes to the expanding corpus of literature that delves into how the utilisation of tourist reviews in examining tourist overall vacation experiences and satisfaction. As observed in previous literature (Hong, 2020; Baleiro, 2023; Sangkaew and Zhu, 2023), the use of tourist reviews as demonstrated in TripAdvisor has been of significance in terms of being a valuable data source. It is anticipated that the increasing utilisation of online blog content for scientific research purposes would significantly expand in the future.

### 6.3.1 Factors Influencing Weather Experiences

The South African tourist sector consists of various forms of tourism markets such as sports, wine, business and cultural, adventure, coastal and nature-based tourism (Rogerson and Rogerson, 2021). Each of the tourist destinations and hotels studied here have unique climates and weather associated with them which make them ideal for evaluating the effect of weather and climate on tourism as a whole (Fitchett et al., 2017). The diversity and type of tourism offerings at each region is specific and diverse. Cities such as Cape Town, Johannesburg and Durban serve as attraction sites for business tourism as well as providing various tourist activities such as museum tours and botanical gardens (Fitchett et al., 2017). Hence all these regions have tourist attractions that are sensitive to weather and climate. The tourist experiences vary in relation to weather, throughout each location. Climate and weather are of importance within tourist attractions, as there is an inter dependence that exists between climate and tourism (Kozak et al., 2008; Moreno et al., 2008; Rutty and Scott, 2016). The results obtained in this study confirm that weather is of importance according to tourist reviews as the data indicated a total of 161,280 reviews web-scraped, out of which 5.59% had climatic references. The presence of these climatic references indicates the effect that weather has on tourism and give reference to the climatic suitability of a region (Fitchett and Hoogendoorn, 2018). Various climatic aspects were mentioned by reviewers. As according to De Freitas (2003), tourism climate aspects can be categorised into three facets, namely: thermal, aesthetic and physical. The physical facet includes mentions relating to mist, rain, thunderstorms and sun while the aesthetic aspect includes climatic mentions in reference to sun, mist, clouds and clear skies. The thermal facet involves mentions of heat, uncomfortable temperature, comfortable temperatures, moderate temperatures as well as humidity and coolness (de Frietas, 2003). Other mentions used in this study included bad and good weather.

According to the results obtained from the study, rain was classified as the third most mentioned climatic factor. Rain accounted for 13.5% of all references to climate. This could be attributed to the significant impact that rain can have on the range of activities available for tourists to engage in (Hübner and Gössling, 2012). Acknowledging the impact of rain on various aspects of tourists' vacation experience, such as trip cancellations, shorter durations, transportation issues, and limitations on outdoor activities, is crucial. This finding aligns with other prior case studies. Chen and Lin (2017) observed that, in Taiwan, wet days had a negative influence on whale watching. Hübner and Gössling (2012) identified that a significant number of tourists exhibited unwillingness to revisit Martinique following their encounter with intense rainfall. The significance of weather lies in its ability to influence the perception of a location and impact decisions on future travel arrangements (Gössling et al., 2013; de Freitas et al., 2015; Gössling et al., 2016; Park et al., 2018).

In South Africa, cold, hot, and rainy weather conditions predominated, with significantly higher proportions noted. Furthermore, various provinces demonstrated heightened frequencies of particular meteorological conditions: Gauteng experienced the highest frequency of cold conditions, the Western Cape recorded the most references to cloudy weather, Mpumalanga exhibited the windiest conditions, the Eastern Cape had the greatest number of thunderstorm-related mentions, and the Northern Cape noted the highest proportion of hot-related mentions. Cold weather generally leads to lower TCI values, signifying less favourable conditions for tourism. Winter months typically produce the lowest TCI scores owing to decreased tourist activity (Efe et al., 2022). Precipitation and cloudy conditions adversely impact TCI, as they correlate with discomfort and fewer outdoor activities. Research indicates that heightened precipitation is associated with lower TCI ratings, especially in coastal areas during the summer (Efe et al., 2022).

Thunderstorms may significantly reduce tourism attractiveness, resulting in significant decreases in TCI. The unpredictability of such weather conditions can dissuade tourists, affecting overall visitation rates (Dubois et al., 2016). Although adverse weather conditions typically lower TCI values, certain tourists may pursue particular experiences linked to these conditions, such as winter sports or storm watching, suggesting a complex relationship between weather and tourism preferences.

Over the past years, there has been a growing interest in the interactions between tourism, weather, and climate, with emphasis on the effect that climate change has on tourism (Fitchett and Hoogendoorn, 2018). Tourists are vastly sensitive to multiple weather or climate related conditions especially in the case of choosing a specified tourist destination (Fitchett et al., 2017). Climate and weather both have the potential to either positively or negatively affect the tourism sector, as tourists always seek for places that are associated with favourable weather conditions (Fitchett et al., 2019). South Africa is a country that is highly diverse in terms of climate and weather with temperature variations throughout the land (Fitchett et al., 2019). The most common meteorological variable that tourists in South Africa are most sensitive to is temperature. Temperature has a more immediate and noticeable physiological and psychological effect on individuals compared to other weather variables like wind or humidity (Tian et al., 2021). High temperatures can lead to discomfort, irritability, and fatigue, while cooler temperatures can lead to discomfort due to insufficient clothing or preparation (Tian et al., 2021). Tourists may be more inclined to comment on or complain about these physical sensations, making temperature one of the most discussed weather variables in reviews and feedback (Zeiss et al., 2022). To illustrate, coastal regions in South Africa such as Durban are associated with warm temperatures which most tourists would classify as favourable weather conditions, indicating their level of thermal comfort.

However, colder conditions reduce the level of thermal comfort of tourists in cooler regions, predominantly in winter. Favourable climate is a critical prerequisite for beach tourism (Wilkins et al., 2018). Stahl et al. (2020) conducted a study that corroborated these sentiments, revealing that over 70% of tourists consider climate to be a critical factor in their decision-making process when selecting tourist destinations. The research demonstrates that outdoor thermal comfort, which is determined by the date, temperature, solar radiation, and wind speed, is an important factor in the selection of a destination (Stahl *et al.,* 2020).

Some reviews from the web-scraped data, predominantly had negative sentiments with regards to cold weather conditions especially in the case of outdoor located tourist destinations. The most frequently cited meteorological conditions in this research were cold, hot, rain and sun, which links to previous research that was conducted through the manual approach (Fitchett et al., 2018; Fitchett et al., 2019). Extensive web-scraping and manual data reviews were conducted for the 19 destinations based on these weather conditions. The graphs depicting various weather locations showed diverse perspectives on weather regarding both approaches. With regards to "Hot" climate references, it was found that web-scraped data had more climatic references than manual data in 16 of the 19 destinations which emphasises the sensitivity of tourists to thermal comfort. In investigating the weather preferences of various French tourists Dubois et al. (2016) found that tourists were most sensitive to rainfall, which is seen as the most inconvenient factor, followed by cold weather and radiation fluxes. The research identified a disparity between exposure to the sun and heat, contrasted with precipitation and cooler temperatures. The high tolerance for elevated temperatures and the lack of worry regarding heatwaves stand in contrast to a pronounced aversion to rainfall, which significantly influences decision-making processes (Dubois et al., 2016). This indicates that tourists are indeed sensitive to specific weather conditions mainly pertaining to heat,

cold and sunshine and this impacts the likelihood of them returning. In another study based on investigating the impact of weather conditions on the behaviour of tourists at a beach destination, it was found that the most important weather condition for tourists when deciding to visit the beach is that it is sunny (43%) with the absence of rain (37.3%; Toubes et al., 2020). When it comes to rainfall, tourists generally prefer regions with less rainfall as it has the potential to hinder most tourist activities such as sightseeing and outdoor associated activities (Ngxongo, 2021). Additionally, windy conditions also influence tourists' decisions when selecting a destination as windy conditions can affect a range of activities such as surfing, wildlife safaris and beach visits (Scott and Lemieux, 2010). Humidity and sunshine duration can also affect tourists' daily activities, as regions associated with high humidity levels can result in tourists not engaging in some activities (Adiguzel et al., 2022).

Assessing tourists' self-reported satisfaction with a destination and their dependence on weather can help accommodation providers address specific concerns effectively. Analysing weather reports across destinations in South Africa and examining relevant climatic factors can help identify areas that are particularly susceptible to climate change. Understanding tourists' perceptions related to climate is crucial for developing effective adaptation strategies, and TripAdvisor reviews can provide valuable insights in this regard. The use of web-scraping has increased in biometeorology recently, with studies using this approach to extract data from online sources for climate change research. However, web-scraping has not been widely applied to the research of tourism and climate change, both globally and in southern Africa. The use of this data acquisition technique is recommended for evaluating tourists' climate sensitivity in different areas.

## 6.3.2 Analysis of Tourist Weather Experiences

### *6.3.2.1 Importance of Climatic Factors in Tourist Reviews*

Weather does not exist in isolation as a phenomenon (Kim et al.,2017; Fitchett et al., 2020). The impact that weather and climate have on tourist experiences is of significance as it can influence the overall satisfaction regarding the vacation (Jeuring, 2013). The sentiment analysis results reveal that climatic factors strongly impact the shaping of tourists' experiences and overall satisfaction. Positive evaluations constituted 54.9% of the total reviews, with 30.40% being very positive. Neutral evaluations contributed to 7.3%, while negative sentiments accounted for 5.2%, and very negative evaluations comprised of 2.2% of the overall sentiments (Table 5.6). These figures indicate that while a majority of tourists view climatic conditions favourably, there is still a notable proportion that experiences adverse effects due to unfavourable weather. Tourists' loyalty may be adversely affected by perceived unfavourable weather, particularly if they anticipate similar conditions in the future (Zeng et al., 2023). Alomar and Almhameed (2024) have found that tourist activity reaches its peak during moderate weather conditions and decreases during extreme weather. Consequently, extreme temperatures may discourage tourism.

Climatic factors such as cold, heat, and sun-related conditions garnered a significant number of very positive sentiments, suggesting that many tourists enjoy these weather conditions. Rain, often perceived negatively in many contexts, surprisingly received a considerable amount of positive feedback, indicating that some tourists find rain beneficial or pleasant. Interestingly, drought conditions also had a high percentage of very positive sentiments, possibly reflecting relief or appreciation when drought conditions are alleviated, which deviates from previous research whereby drought conditions were often negatively associated in relation to tourist experiences.

*6.3.2.2 The Relationship Between Tourists and the Weather They Experience*

Tourists' experiences with weather conditions show a complex relationship, as evidenced by the varied responses to different climatic variables. The reviews indicate the complex ways in which tourists perceive weather (Jeuring and Peters, 2013). For instance, upon considering the various aspects of tourism climate, clouds are categorised under the aesthetic aspect, which primarily means that they have aesthetic worth and do not disrupt tourist activities as such. However, in this study there were reviews that had references relating to clouds having an effect on tourist activities. In one instance, a tourist experienced rain and clouds, which prevented them from seeing the Drakensberg mountains. This implies a significant negative impact, as the view of the Drakensberg was likely a key attraction for their visit. This then indicates that clouds do have the potential of affecting the tourist sentiment regarding the destination or vacation experience. Tourists in the Arctic demonstrate a profound aversion to weather conditions that decrease visibility, such as overcast skies (Gawor and Dolnicki, 2018). This aversion is particularly evident in regions such as Spitsbergen, where the overall tourist experience may be reduced by the increased precipitation resulting from climate change, which may result in more cloudy days (Denstadli and Jacobsen, 2014). Tourist satisfaction is notably influenced by the quality of the service landscape in protected areas, which includes the visibility of natural attractions. Guests' dissatisfaction may result from poor visibility, which decreases the perceived value of the site (Oviedo-Garcia et al., 2019). The satisfaction levels of visitors in urban settings such as Valparaiso are influenced by the mode of transportation, as they perceive the city differently depending on their mode of transportation. These experiences can be further complicated by limited visibility caused by weather (Martin et al., 2021).

Although clouds can obscure views and impact satisfaction, certain tourists may still value the overall experience, valuing other aspects such as cultural engagement or distinct environments, indicating a complex relationship between visibility and satisfaction.

According to previous research (Lee et al., 2014; Gössling et al., 2016), hot and sunny conditions are often perceived as good and acceptable weather whereas rain and cold associated weather is observed as bad weather. However, in this study, rain and cold conditions had some evaluations that were predominantly positive to very positive that could have been attributed to tourists' country of origin as well as their preferences. In some instances, the cold or rainy conditions did not negatively influence the tourist's general contentment regarding the weather. In contrast, in other cases, it was discovered that both cold and rainy weather-related conditions were associated with negative sentiments (Gössling et al., 2016). Some reviewers noted that the cold weather in specific regions had notable effects on their comfort, while rain-related conditions obscured scenic views and frequently resulted in the cancellation of beach visits, which detracted tourists' overall satisfaction with their vacation. In a study conducted by Jeuring (2013), rain was generally regarded as a negative factor, as some tourists reported that the continuous rainfall had an adverse impact on their overall experience. In some cases, however, it was observed as positive as it added to the aesthetic value of the region. This indicates that the weather evaluations regarding tourist' experiences are rather varied. Of the small proportion of negative reviews, cold weather negatively impacted dining experiences and overall comfort due to inadequate heating, detracting from overall satisfaction. Conversely, clear weather conditions at night associated with the cold, allowing for stargazing, significantly enhanced the ambiance and overall experience, underscoring the importance of favourable weather in tourist satisfaction.

Weather often has an impact on the type of holiday activities available for tourist participation (Beckon and Wilson, 2010; Giddy et al., 2017; Fitchett et al., 2020). In this study for instance, rain, fog and cloudy condition often obscured the scenic views that tourists aimed to see in certain regions. In previous research conducted, it was observed that great visibility underwater was of significance in activities such as snorkelling and diving and as such rain related conditions could have an impact on the turbidity of the water (Ziegler et al., 2012; Jarvis et al., 2016). Similarly, bad weather, including cold fronts and fog in Johannesburg, limited outdoor activities and visibility, yet tourists still found aspects of their stay enjoyable.

The results indicate that windy conditions often had both negative and positive evaluations. In certain instances, a light breeze was often perceived as relief from extremely hot conditions. This aligns with particular research that found a light breeze to be optimal for beach tourists, whereas intense wind is considered unacceptable (Rutty and Scott, 2013). In other instances, windy conditions contributed to the lack of sleep as the conditions were unpleasant at night due to wind. On the contrary, severe weather conditions, including heat waves and storms had more pronounced negative effects. High temperatures exceeding 40°C impacted tourists' sleep quality and overall comfort, while a horrendous storm in East London led to the cancellation of a coastal trip, consequentially reducing overall satisfaction.

*6.3.2.3 Implications of Weather on Tourist Experiences*

The sentiment analysis illustrates that while weather conditions do affect tourist satisfaction, the extent of this impact varies. Weather plays a critical role in regions where outdoor activities are the main attractions, influencing the overall experience more significantly (Fitchett et al., 2017). For instance, the inability to hike or swim due to bad weather led to a notable decrease in

satisfaction. Conversely, tourists in regions where indoor activities or facilities mitigate the impact of weather, such as the provision of cold towels on hot days, reported higher satisfaction despite adverse weather conditions such as observed in this study. Research has shown that the general contentment of tourists and their likelihood to return to a tourist destination are directly influenced by their perception of their experience with the weather in the region, particularly in adverse weather conditions (Kim et al., 2017). In high altitude tourist destinations, tourists demonstrated greater loyalty when they anticipated favourable weather, suggesting that positive expectations may alleviate dissatisfaction (Zeng et al., 2023). Furthermore, providing comfort measures could influence tourists' perceptions of outdoor thermal conditions, as demonstrated in coastal tourism, where beachgoers preferred warmer temperatures (Rutty and Scott, 2015).

The diverse responses to weather conditions also highlight the importance of adaptive measures by tourism operators. Providing amenities that mitigate the effects of extreme weather, such as efficient heating during cold weather or cold towels during hot weather, can enhance overall satisfaction. Research on outdoor personal heating and cooling highlights the development of a Janus textile that can provide localised heating and cooling, improving personal thermal comfort in outdoor settings (Luo et al., 2020). Tourists are more likely to seek environments that offer thermal comfort, as evidenced by increased retail sales during extreme temperatures, suggesting that amenities catering to comfort can drive consumer behaviour (Yoo et al., 2024). The tourism sector is increasingly responding to climate change by implementing energy-efficient solutions, which can enhance the overall experience for visitors (Wiratama et al., 2024). Research shows that tourists are willing to adapt their travel choices based on climate amenities, indicating that effective management of extreme weather can lead to higher satisfaction and longer stays (Barrios and Ibañez, 2014). Additionally, clear communication about weather expectations and potential

impacts on planned activities can help manage tourist expectations and improve their overall experience.

In summary, the sentiment analysis confirms that weather significantly influences tourist satisfaction, particularly in regions where outdoor activities are prominent. While some tourists can overlook adverse weather due to other positive aspects of their stay, extreme conditions often lead to discomfort and dissatisfaction. Therefore, tourism operators should consider climatic factors and implement adaptive measures to ensure a positive experience for all tourists.

The research was qualitative and was aimed at deepening the understanding of tourists' experiences related to weather. The objective was to investigate the insufficient empirical evidence regarding tourists' sensitivity to weather conditions (Denstadli et al., 2011) and to gather data from observation that assesses the responses, views, requirements, responses, and desires of tourists using web-scraping techniques (de Freitas, 2003). Various weather themes were identified throughout the evaluations of tourist reviews regarding weather. This suggests that the impact of weather and climate on tourism is more complex than previously thought, as illustrated by the various experiences of tourists (Becken and Wilson, 2010; Denstadli et al., 2011; Jeuring, 2013). Narratives frequently incorporated subjective evaluations, characterising weather as an experiential occurrence. Tourists evaluate a variety of weather factors, determining whether they are favourable or unfavourable in accordance with their plans and expectations. This evaluative process has been implicitly presumed in previous research studies, but it has not been considered in the context of weather experiences. Hence, the present comprehension of how tourists think about and perceive weather is enhanced by the weather themes identified in this study. This study provides insight into the intricate and complex relations that exist between tourism and weather

(de Freitas, 2003; Matzarakis et al., 2007), which further contributes to the existing knowledge in biometeorology.

### 6.3.3 Implications for Tourism sector

The study highlights the complex relations between weather, climate and tourism. It provides evidence that climate and weather have an effect on the general contentment of tourists. South Africa is anticipated to suffer abnormally large reparations as a result of climate change, attributable to the increased intensity of modelled impacts and the reduced ability to adjust to changing conditions caused by more pressing developmental requirements (Reid and Vogel, 2006; Fitchett *et al.*, 2017).

Internationally South Africa is recognised to be among of the hardest afflicted nations by the consequences of climate change and is also ranked as a contributor to climate change (Scott *et al.*, 2019). The country now ranks as being one of the top 15 GHG emitters and has a record of generating high emissions (Scott *et al.*, 2019).

The beach tourism sector in South Africa is relatively sensitive to climate change (Friedrich et al., 2020). This indicates that South Africa is perceived as susceptible to climate change, having a limited adaptation potential but a favourable susceptibility to climate change (Phillips *et al*., 2018). It implies that the South African beach tourist sector will be affected by climate change, but other regions across the world will be impacted far worse (Gössling *et al*., 2002, 2012b; Hall, 2008; Scott *et al*., 2012a; Santos-Lacueva *et al*., 2017; Phillips et al., 2018).

Scarcity of water, as a result of climate change (Gössling *et al*., 2012b; Saarinen *et al*., 2012; Kaján and Saarinen, 2013), may have an impact on beach tourism in both the west and southern regions of

South Africa (Cai *et al*., 2014). Serdeczny *et al*. (2017) believe that owing to the reduction in precipitation, South Africa will be more vulnerable to droughts in the future. Climate change will have a detrimental environmental impact on the Western Cape in South Africa (Dube *et al*., 2020). Wine farming will suffer as a result, as will tourism activity that is supported by this farming (Dube *et al*., 2020). The South African wine industry is a major tourist attraction (Steyn, 2012).

Climate change does not operate like any other conventional global concern but has major influence not just on one component but on various features of the world such as transportation, ecology, agriculture, forestry, and energy use (Hoogendoorn and Fitchett, 2018). Human endeavours such as tourism are among of the aspects most influenced by climate change. Sufficient studies have concentrated mostly on the consequences of climate change on the environment with important conclusions by the IPCC (IPCC, 2021). However, the understanding concerning the interrelations that exist between tourism and climate change inside South Africa are still scarce.

This research illustrates the weather and climate that tourists experience and how that shapes their overall satisfaction. By acknowledging tourists' personal connections to weather, tourism businesses can better cater to their needs and preferences. For example, providing amenities that offer alternatives for outdoor activities during inclement weather can enhance tourist experiences and satisfaction. Through the utilisation and application of web-scraping technologies, a deeper comprehension of tourists' weather-related concerns and inclination can be gained (Jeuring, 2017). Therefore, destination managers can develop more effective adaptation strategies to mitigate the impact of weather variability on tourist satisfaction. Understanding the relationship between weather and tourist behaviour can help tourism stakeholders anticipate and respond to changing weather patterns and climate conditions, thereby enhancing the resilience of tourism destinations. A realistic depiction of probable weather conditions at destinations is crucial for aligning tourists'

expectations and improving their overall satisfaction (Jeuring, 2017). Understanding tourists' weather-related concerns and preferences can inform destination branding campaigns and marketing efforts.

## 6.4 Limitations

This research, akin to all research, encountered limitations. This section of the discussion chapter addresses specific challenges that arose during the research process.

The parts that follow highlight and analyse the main constraints that impeded the comprehensive objectives and goals of the thesis. The main difficulties were the quantity of data collected, the development of the web-scraping tool employed, and the statistical analyses conducted on the data. It is crucial to recognise these limitations, as they have the potential to affect the accuracy of the findings and inform future research, as outlined in Chapter 7 (Conclusion). This section defines what steps were implemented to mitigate these limitations throughout each stage of the research.

### 6.4.1 Data limitations

Both manual data collection and web-scraping techniques possess inherent limitations. The process of manually collecting data can be susceptible to human error and can also be a time-consuming task (Birnbaum, 2005; Landers et al., 2016). Conversely, web-scraping techniques may not possess an exhaustive understanding of the context and can be influenced by technical limitations or inaccuracies during the scraping process (Dallmeier, 2021). This phenomenon was further shown in the research, as some phrases were interpreted as climatic references while lacking such connotations. When data is collected manually, there is a risk of introducing biases

or variations in the identification, recording, or categorisation of mentions due to the subjectivity of the reader (Birnbaum, 2005; Landers et al., 2016). Diverse readers may construe identical references in varying ways, which could result in incongruities in the compiled data. Conversely, web-scraping can offer a methodical and uniform methodology for acquiring references, albeit it may be deficient in comprehending the subtle connotations of the reference within its contextual framework. Web-scraping algorithms commonly depend on predetermined patterns or keywords, which can lead to the inclusion of irrelevant mentions or the omission of mentions that necessitate contextual comprehension (Egger et al., 2022).

Three out 19 destinations that had climatic references had insignificant differences between both approaches, namely, East London ($p=0.02$), Knysna ($p=0.01$) and Polokwane ($p=0.02$). This could be attributable to differences in interpreting what constitutes "good" or "bad" weather in manual data as opposed to web-scraped data, which automates words as they are. In contrast to web-scraping, which generally entails the extraction of references through predetermined patterns or keywords, manual data collection affords the reader the ability to grasp the significance of the reference within its contextual framework (Dallmeier, 2021). The comprehension of a text may be subject to the reader's personal interpretation, experience, and level of knowledge (Saarinen et al., 2022); for example, in this study, the word ‘humidifier’ utilised by a reviewer complaining that the room needed one, could mean that the reviewer was indirectly stating that the conditions in the room were arid. In web-scraping however, that is not the case, as the code is designed to automate data according to what is documented in the code. Another example included the misclassification of the words such as ‘hotel’ for ‘hot’ and ‘Sunday’ for ‘sun’ within the web-scraped data (table 6.1). Whereas within the manually collected data, such misclassifications did not exist.

**Table 6. 1:** Misclassified words

| Misclassified words: | Misclassification example | Reason attributed to the misclassification | Possible solutions to this |
|---|---|---|---|
| **Cold** | -Cool temperatures or an unkind welcome | Over-counting in web-scraping | Integration with manually checking |
| **Warm** | -Warm temperature, or friendly welcome. | over-recording in web-scraped data | Coupled words that could be utilised to signify warm conditions. |
| **Sun** | -Sunlight or warmth (temperature) or even form part of a name of a hotel | Over-counting in web-scraping | Coupled words that could be utilised to signify sun or sunshine. |
| **Wind** | -To rest or calm down.<br>-Name of a hotel<br>-Indicating a form of energy (wind turbines) | Over-counting in web-scraping | Integration with manually checking |
| **Humid** | -Humidifier<br>-Wet/Damp | Over-counting in web-scraping | Integration with manually checking |
| **Mist** | -Wet or moist conditions<br>-Name of a hotel e.g. MistyRiver | Over-counting in web-scraping | Integration with manually checking |
| **Hot** | -Hot temperatures or a '**hot**el' | Over-counting in web-scraping | Integration with manually checking |
| **Rain** | -Rain as in precipitation<br>-Bathroom shower head classification | Over-counting in web-scraping | Integration with manually checking |

### 6.4.2 Web-Scraping Limitations

One important limitation encountered during the development of the web-scraping tool was the occurrence of errors in the code design. When websites undergo structural changes, or when the structure of a website differs from that of similar websites, the process of data extraction may fail or produce inaccurate results (Wang and Kirilenko, 2021). To mitigate this, frequent manual checks of the scraped data were performed, and updates to the code were made as required. While this approach helped minimise inaccuracies, it could not fully eliminate the possibility of errors occurring due to unforeseen changes in website architecture. As a result, there may be some impact on the overall accuracy of the results, especially when such errors went undetected for a time.

Another limitation was the technical barrier involved in writing and managing the scraping code. The development process requires programming skills, which can be a notable obstacle for non-technical users (Singrodia et al., 2019). This limitation was addressed by leveraging existing web-scraping libraries and tools that simplify some aspects of the coding process. However, this reliance on external tools still demands a basic understanding of programming, which could limit accessibility for broader research use. The accuracy of the data extraction was not significantly compromised by this limitation, but the usability of the tool by non-experts remains restricted.

Additionally, the tool required access to a server to automate data collection requests, which presented a logistical limitation. Setting up and maintaining a server can be costly and technically demanding. To address this, the tool was run on local machines for smaller datasets or rented cloud servers for larger data pulls. While this strategy reduced costs, it did introduce occasional delays in data collection, potentially impacting the timeliness of the data but not its accuracy.

### 6.4.3 Sentiment Analysis Limitations

With regard to the evaluation of weather, the categorisation of tourist sentiment was in five aspects namely, very negative, negative, neutral, positive and very positive. The categories may be viewed as a limitation due to the oversimplification of the complexity of emotions expressed in the reviews evaluated (Fitchett et al., 2020). This method of analysis may not fully convey the entire variety of nuanced sentiments or perceptions that tourists may have regarding the weather and climate that they experience during vacations (Fitchett et al., 2020). Towards addressing this, the study incorporated contextual analysis alongside sentiment scoring, which helped add some depth to the interpretation of results. Secondly the study only focused on English reviews, which may miss

reviews that could enhance the comprehension of tourist’s perceptions of climate and weather. This language restriction could lead to an incomplete understanding of tourists' perceptions of weather and climate, especially in destinations with diverse international visitors. While translation tools could have been used, time and resource constraints prevented their inclusion in this study. Consequently, the findings may not be fully representative of all tourists' experiences, particularly those from non-English-speaking backgrounds, which could marginally affect the comprehensiveness of the conclusions.

# Chapter 7: Conclusion

## 7.1 Introduction

Tourist organisations and individuals must contend with the weather as it occurs, yet there is limited empirical evidence on how weather influences tourist perceptions and experiences. This research gap highlights the need for a structured approach to understanding the effects of weather on tourists. TripAdvisor reviews offer more unbiased insights, as they reflect tourists' experiences without prompting specific inquiries about weather. Exploring tourist sensitivity across South Africa using TripAdvisor reviews through a non-labour-intensive, unbiased platform is vital. In addressing these issues, the first part of the research focused on developing a web-scraping tool that could automate the data needed and testing the effectiveness of the tool against manual counts that have been published for south Africa. In understanding the relationships between tourists and how weather affects them, the subsequent segment of the study concentrated mainly on expanding the research to all accommodation establishments throughout the whole of South Africa and investigating the climatic suitability of tourists through the use of sentiment analysis. The findings and conclusions of the work are found in chapter 5 (Results) and 6 (discussion). This chapter focuses on mainly providing a synthesis of the work done from the earlier chapters and reflects on how and the degree to which the aims and overall objectives of this research were achieved. Lastly, avenues for future research are then discussed.

## 7.2 Synopsis and Significance of the Study Findings

The introduction of this research outlines the main gaps in knowledge and issues in the literature, aiming to contribute to new knowledge. This section briefly describes how the thesis has contributed to the field of biometeorology in South Africa. The thesis makes important contributions, particularly by developing a standard method to assess tourists' sensitivity to weather and climate in South Africa through the analysis of TripAdvisor big data.

One of the key contributions is the development of an effective web-scraping tool that aggregates mentions of weather and climate from TripAdvisor reviews across 19 locations in South Africa, and later, additional regions. This addresses the need for a tool to obtain data on the relationship between weather and tourist perspectives and activities, which is crucial for the tourism business. Analysing large datasets to understand how weather affects tourist behaviour helps operators tailor their services to varying weather conditions. Accurate climate suitability information is essential for optimising adaptation and mitigation actions, ensuring the sustainability of the tourism sector (Hoogendoorn and Fitchett, 2016; Fitchett et al., 2018). Web-scraping automates and integrates this information, providing a comprehensive view. Analysing these reviews can aid in comprehending the effects of climate change has on tourism, revealing which climatic conditions are most relevant to visitors. This data assists in evaluating climate change effects, understanding tourist concerns for effective adaptation, and increasing industry awareness of the importance of data collection.

In this study, the web-scraping demonstrated the effectiveness of automating data from TripAdvisor to further analyse the climate sensitivity of tourists and validate the findings with the manual approach. The limitations of this, however, included, the time employed in the development of the code to effectively automate the data from the website as well as the

misclassification of weather-related themes within the analyses. It is however depicted that web-scraping can be accepted method of data collection just as the manual approach. Additionally, changes to the TripAdvisor website layout or algorithms can impact the accuracy of the extracted data (Egger et al., 2022). However, when coming to analysing the perception of tourists regarding weather, both methods have their strengths and weaknesses. Tourist assessments, for example, might offer information on the influence of heatwaves or extreme cold spells on people's comfort levels, capacity to engage in outdoor activities, and overall trip satisfaction (Liu et al., 2021). Furthermore, studying tourist assessments might aid in the identification of situations or aspects that may lead to favourable or poor experiences under various weather and climatic conditions (Ngxongo, 2021). This information may guide urban and regional planners in designing and creating initiatives aimed at creating suitable infrastructure and providing good services in improving the overall well-being and satisfaction of tourists under various meteorological conditions (Abundabar and Pongpong, 2022). Web-scraping may also be used to investigates the economic effects of weather and climate on the tourism sector. Evaluating visitor assessments allows researchers to gain understanding of the influence of weather and climate on tourists' opinions and preferences (Fitchett et al., 2019).

## 7.3 Achievement of Study Aim and Objectives

The Introduction (Chapter 1, Section 1.2) defines the objectives and goal of the research. The primary objective of this research was to examine tourist’s sensitivity to weather and climatic aspects while visiting various regions of South Africa through an analysis of big data from TripAdvisor. The objectives of the study were aimed at tackling the primary objective of the

research by contributing to the knowledge gaps that were addressed in the introduction and literature review chapters. The work that was utilised to address the study objectives was the automated data from the TripAdvisor website working from 2023 going back in time. In dealing with the objectives of the study, the limitations discussed in Chapter 6 encountered while conducting the research have not undermined the overall accomplishment of the aim and study objectives. The rationale and methodology behind the primary objectives of the research, as well as the extent to which they were achieved, are presented below from chapter 1 (Introduction).

1. *Develop an effective web-scraping tool to run on all TripAdvisor reviews for a given region, and aggregate mentions of weather and climate*

This objective was achieved through a series of stages. The scope and key parameters were determined, with a particular emphasis on weather-related terms such as "temperature" and "rain." The tool was created by utilising web-scraping frameworks such as Scrapy or Beautiful Soup, and scripts were written to extract review content. In order to address issues such as blocked requests, automation and error-handling features were implemented. The extracted data was cleaned and pre-processed to remove irrelevant information and filter reviews by language. The sentiment associated with these mentions was assessed using sentiment analysis, and statistical techniques in Excel were employed to identify weather-related keywords. The data was subsequently aggregated by region and time and presented in the form of frequency distributions and maps. The tool was subjected to testing and validation to guarantee the precision and thoroughness of the data collected. The main goal of these analyses was to develop a web-scraping tool that could extract data from TripAdvisor, thereby offering a clearer comprehension of the correlation between weather and tourism. The efficacy of the tool was determined to be consistent with that of other

published works. The web-scraping tool functioned efficiently and was capable of automating data extraction from the website. The data was subsequently utilised to conduct analyses.

*2. Test the effectiveness of the web-scraping tool against manual counts that have been published for south Africa*

This objective was achieved through conducting a comparative analysis, analysing the outputs of manually collected and web-scraped TripAdvisor reviews, to evaluate the efficacy of a web-scraping approach. Utilising a sample of 26,687 web-scraped TripAdvisor reviews, this study examined the comparability of results obtained from web-scraping relative to manually collected data for 19 South African destinations. All statistical tests that were conducted in the previous studies were replicated during this process. These included the following: calculating and stating the total number of reviews, determining the proportion of climatic or weather-related mentions, exploring the relative climatic contribution of the climatic factors per location to investigate the variability of the two datasets. In addition, a comparison of means test was conducted for each component to determine whether there was significant variation between the data that was manually collated and that was scraped from the web. This objective addressed the research gap presented in the rationale in the first chapter detailing web-scraping as a potential method that could upscale manual analyses to hundreds of thousands of reviews, and to improve accuracy and reduce bias. A percentage difference of 3.5 was noted between both methodologies for the total amount of climatic mentions captured. A significant difference was evident in three of the 19 locations per climate mention (Knysna, Polokwane and East London), however no statistically significant differences were noted between both approaches for the overall climatic mentions. It is worth noting that web-scraping, being driven by code, may lack the nuanced understanding of text that a human reader possesses. Nevertheless, it was determined that web-scraping is effective in

being less tedious, obtaining vast amounts of data in real-time and in offering a comprehensive perspective on the sensitivity of tourists towards climate and therefore can be accepted a good source of obtaining data.

*3. Investigate the climatic suitability of tourists visiting South Africa through the use of sentiment analysis*

To enhance comprehension of the relationship that exists between tourists and weather, sentiment analysis was conducted. This objective aimed to answer the research gap concerning the lack of empirical data concerning the relationship between weather and individual tourist perceptions, activities, and how this is reflected in tourist experiences, a concept that is central to the tourism business. This objective was achieved through conducting both a coded statistical and manual approach such as classifying the various sentiments that tourists felt regarding the weather that they experienced at a particular region, into five main categories and aggregating the various sentiments. It was concluded that the sentiment analysis undertaken in this research provided insights into the attitudes of diverse guests towards various weather conditions. It was observed that apart from instances of good weather, mist and hail, all other weather scenarios elicited predominantly negative to very negative sentiments.

*4. Expand the analysis to all accommodation establishments in South Africa through the combined use of web-scraping*

Due to the success of achieving the first and the second objectives which involved the development and testing of the web-scraping tool, the study was then expanded to all accommodation establishments throughout South Africa. Due to trouble with the code, not all accommodation establishments were reached which was a limitation. However, the data that was obtained did give

an overview of the validity and efficacy of the web-scraping tool to collect and parse data for analysis. The tool was able to automate sufficient data to perform the necessary statistics to get a general outline of the relationship between weather and tourists throughout the country. The objective was achieved through utilising the same code that was generated in the first objective to collect and parse data from the TripAdvisor website. The same techniques in order to analyse the data were utilised in this regard as well.

## 7.4 Future Research Avenues

Integrating manual data collection with web-scraping techniques can prove advantageous in addressing the constraints of both methodologies. The process of manually collecting data can yield a detailed and context-specific analysis of references, whereas web-scraping can effectively amass a substantial amount of data (Khder, 2021). Through the integration of these methodologies and the utilisation of suitable techniques for data validation and verification, a dataset that is more comprehensive and dependable can be acquired (Lauer et al., 2018). On the contrary, the integration of tourist data that has climatic or weather-related information sourced from hotel reviews through the application of web scraping, with thermal comfort indices can provide an enhanced apprehension of the effect of climatic factors on the comfort experiences of tourists. This approach employs an integrated methodology to evaluate climate sensitivity, thereby offering important implications for enhancing comfort levels in hotel settings. Based on the above-mentioned information, manual comparisons can be done for accuracy purposes in order to identify the various language components. Validation and verification of data retrieved through web-scraping is another area of prospective research in biometeorology. This is essential for ensuring

the precision and reliability of the obtained data. In addition, future biometeorology research should concentrate on standardising web-scraped data. This involves defining formats and methods for organising and analysing obtained data in order to ensure reliability and comparability throughout research. Precisely determining the dependency of the tourism sector on climate, vulnerability to climate change, and ability to adapt is crucial for effectively mitigating its impacts. Ensuring that the tourism sector within a region is sustainable requires an understanding of its dependence on climate and its capability to adapt to climate change. To achieve this, it is vital to have precise information on a regions' current climate suitability for tourism, which can be evaluated using an empirical and consistent approach to compare it to other destinations in which web-scraping data may be of use.

# Reference List


Abundabar, A. P., and Pongpong, M. A. (2022). Assessing the Local Tourism Infrastructure and Services towards Tourism Development of San Jose. *International Journal of Applied Research in Tourism and Hospitality*, 1(2): 97-116.

Abu Khasan, R., Khomonenko, A.D. and Kuranova, O.N. (2023). Big Data Analytics: Integration, Import Substitution, Operativeness, Reliability and Availability. *T-Comm*, 17(10): 59–65.

Adger, W.N., Huq, S., Brown, K., Conway, D., Hulme, M. (2003). Adaptation to climate change in the developing world. *Progress in Development Studies*, 3(3): 179-195.

Adiguzel, F., Bozdogan Sert, E., Dinc, Y., Cetin, M., Gungor, S., Yuka, P., Sertkaya Dogan, O., Kaya, E., Karakaya, K. and Vural, E. (2022). Determining the relationships between climatic elements and thermal comfort and tourism activities using the tourism climate index for urban planning: a case study of Izmir Province: Tourism climate index for urban planning. *Theoretical and Applied Climatology*, 147(3): 1105-1120.

Adhinugroho, Y., Putra, A.P., Luqman, M., Ermawan, G.Y., Mariyah, S., Pramana, S. (2020). Development of online travel Web scraping for tourism statistics in Indonesia. *Information Research: an International Electronic Journal*, 25(4): 880 https://doi.org/10.47989/irpaper885

Agüero-Torales, M. M., Cobo, M. J., Herrera-Viedma, E., and López-Herrera, A. G. (2019). A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. *Procedia Computer Science*, 162: 392-399.

Ahmed, A.Z., Rodríguez-Díaz, M. (2020). Significant labels in sentiment analysis of online customer reviews of airlines. *Sustainability*, 12(20): 1–18.

Ahmed, M.T., Hefny, M. (2007). *Climate change and tourism: An Egyptian perspective*. Cairo: Egyptian Biodiversity Clearing House Mechanism.

Akinboade, O.A., Braimoh, L.A. (2010). International tourism and economic development in South Africa: A Granger causality test. *International Journal of Tourism Research*, 12(2): 149-163.

Akyürek, S., Allen, R.D.W. and Bahcelerli, N.M. (2023). Examining the impact of tourism on economic growth in the Republic of South Africa. *Revista de Gestão e Secretariado*, 14(10): 17254-17271.

Alrawadieh, Z., Law, R. (2019). Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers. *International Journal of Culture, Tourism and Hospitality Research*, 13(1): 84-97.

Al-Saqaf, W., and Berglez, P. (2019). How do social media users link different types of extreme events to climate change? A study of Twitter during 2008–2017. *Journal of Extreme Events*, 6(02): 1950002.

Amaral, F., Tiago, T. Tiago, F. (2014). User-generated content: tourists' profiles on TripAdvisor. *International Journal of Strategic Innovative Marketing*, 1(3): 137-145.

Amelung, B., Nicholls, S. and Viner, D. (2007). Implications of global climate change for tourism flows and seasonality. *Journal of Travel Research*, 45(3): 285-296.

Amusan, L., Olutola, O. (2017). Climate change and sustainable tourism: South Africa caught in-between. *African Journal of Hospitality, Tourism and Leisure,*6(4): 1-15

Anholt, S. (2007). Competitive identity, policy and practice. *A Development Education Review*, 4: 3-13.

Awuor, C. B., Orindi, V. A., Adwera, A. O. (2008). Climate change and coastal cities: The case of Mombasa, Kenya. *Environment and Urbanisation,* 20(1): 231–242

Ayeh, J.K., Au, N. and Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, 52(4): 437-452.

Baker, D.A., Crompton, J.L. (2000). Quality, satisfaction, and behavioural intentions. *Annals of Tourism Research*, 27(3): 785-804.

Baleiro, R. (2023). Understanding visitors' experiences at Portuguese literary museums: An analysis of TripAdvisor reviews. *European Journal of Tourism Research*, 33: 3305-3305.

Banerjee, S.B. (2002). Corporate environmentalism: The construct and its measurement. *Journal of Business Research*, 55(3): 177-191.

Barbera, G., Araujo, L., and Fernandes, S. (2023). The Value of Web Data Scraping: An Application to TripAdvisor. *Big Data and Cognitive Computing*, 7(3): 121.

Barbierato, E., Bernetti, I., Capecchi, I. (2021). Analysing TripAdvisor reviews of wine tours: an approach based on text mining and sentiment analysis. *International Journal of Wine Business Research*, 34(2): 212-236.

Barimah, P. T., Kwadwo, S. O., and David, O. (2015). Assessment of people's knowledge and perception on climate change: A case study of Asunafo North District, Ghana. *International Journal of Innovative Research in Science, Engineering and Technology*, 4(1): 18417-18424.

Barreda, A., Bilgihan, A. (2013). An analysis of user-generated content for hotel experiences. *Journal of Hospitality and Tourism Technology*, 4(3):263-280.

Barrios, S. and Rivas, J. I. (2014). Climate amenities and adaptation to climate change: a hedonic-travel cost approach for Europe. Climate Change and Sustainable Development, Working Paper No. 20.2014. https://doi.org/10.2139/ssrn. 2403728

Baum, T., and Lundtorp, S. (2001). *Seasonality in Tourism*, 1-4. Oxford: Elsevier.

Becken, S. (2005). Harmonizing climate change adaptation and mitigation: The case of tourist resorts in Fiji. *Global Environmental Change*, 15(4): 381-393.

Becken, S., and Wilson, J. (2013). The impacts of weather on tourist travel. *Tourism Geographies*, 15(4): 620-639.

Becken, S. (2010). *The importance of climate and weather for tourism: literature review*. LEAP, New Zealand. https://doi.org/10182/2920

Becken, S. (2019). Decarbonising tourism: mission impossible? *Tourism Recreation Research*, 44(4): 419-433.

Becken, S., Whittlesea, E., Loehr, J., Scott, D. (2020). Tourism and climate change: evaluating the extent of policy integration. *Journal of Sustainable Tourism*, 28(10):1603–1624.

Becken, S., Wilson, J. (2013). The impacts of weather on tourist travel. *Tourism Geographies*, 15(4): 620–639.

Becken, S., Wilson, J., & Reisinger, A. (2010) *Weather, Climate and Tourism: A New Zealand Perspective*. Land Environment and People Research Report No. 20. Available at http://www.lincoln.ac.nz/Research-Centres/LEaP/Publications/#ResearchReports(open in a new window) (accessed 2 August 2012)

Becker, S. (1998). Beach comfort index–a new approach to evaluate the thermal conditions of beach holiday resorts using a South African example. *GeoJournal*, 44(4): 297-307.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. and Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology letters*, 15(4): 365-377.

Berrittella, M., Bigano, A., Roson, R., Tol, R.S. (2006). A general equilibrium analysis of climate change impacts on tourism. *Tourism management*, 27(5): 913-924.

Bhatt, C., Bisht, A., Chauhan, R., Vishvakarma, A., Kumar, M. and Sharma, S. (2023). Web Scraping Techniques and Its Applications: A Review. *2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT)*, 1-8.

Bi, J.W., Liu, Y., Fan, Z.P., Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70: 460-478.

Bigano, A., Hamilton, J. M., Maddison, D. J., and Tol, R. S. (2006). Predicting tourism flows under climate change: An editorial comment on Gössling and Hall. *Climatic Change*, 79: 175-180.

Bigne, J.E., Sanchez, M.I., Sánchez, J. (2001). Tourism image, evaluation variables and after purchase behaviour: inter-relationship. *Tourism Management*, 22(6): 607-616.

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review Psychology,* 55: 803-832.

Brenning, A. and Henn, S. (2023). Web scraping: a promising tool for geographic data acquisition. *arXiv (Cornell University),* 2305-19893.

Briedenhann, J. and Wickens, E. (2004). Tourism routes as a tool for the economic development of rural areas—vibrant hope or impossible dream? *Tourism management*, 25(1): 71-79.

Bruns, A., Highfield, T. (2015). Social media and the public sphere. *The Routledge companion to social media and politics*, 56-74.

Buchholtz, S., Bukowski, M., Sniegocki, A. (2014). Big and open data in Europe: A growth engine or a missed opportunity. *Warsaw Institute for Economic Studies Report Commissioned by demosEUROPA*, 10: 116.

Bujisic, M., Bogicevic, V., Parsa, H.G., Jovanovic, V., Sukhu, A. (2019). It's raining complaints! How weather factors drive consumer comments and word-of-mouth. *Journal of Hospitality and Tourism Research*, 43(5): 656-681.

Buzinde, C.N., Manuel-Navarrete, D., Kerstetter, D., Redclift, M. (2010). Representations and adaptation to climate change. *Annals of Tourism Research,* 37(3): 581-603.

Cai, W., Borlace, S., Lengaigne, M., Van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., 2014. 290 McPhaden, MJ; Wu, L. (2014). Increasing frequency of extreme El Nino events due to greenhouse warming. *Climate Change*, 4: 111-116.

Caldeira, A. M., and Kastenholz, E. (2018). It's so hot: predicting climate change effects on urban tourists' time–space experience. *Journal of Sustainable Tourism*, 26(9): 1516-1542.

Cecinati, F., Matthews, T., Natarajan, S., McCullen, N., and Coley, D. (2019). Mining social media to identify heat waves. *International Journal of Environmental Research and Public Health*, 16(5): 762.

Cecinati, F., Matthews, T., Natarajan, S., McCullen, N. and Coley, D. (2019). Mining social media to identify heat waves. *International Journal of Environmental Research and Public Health*, *16*(5): 762.

Centobelli, P. and Ndou, V. (2019). Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism*, 22(15): 1862-1882.

Chareyron, G., Da-Rugna, J. and Raimbault, T. (2014). Big data: A new challenge for tourism. In *International conference on big data (big data)*, 5-7

Chase, B.M. and Meadows, M.E. (2007). Late Quaternary dynamics of southern Africa's winter rainfall zone. *Earth-Science Reviews*, 84(3-4): 103-138.

Chen, C.C., Huang, W.J. and Petrick, J.F. (2016). Holiday recovery experiences, tourism satisfaction and life satisfaction–Is there a relationship? *Tourism Management*, 53: 140-147.

Chen, C.F. and Tsai, D. (2007). How destination image and evaluative factors affect behavioural intentions? *Tourism management*, 28(4): 1115-1122.

Chen, H., Chiang, R.H. and Storey, V.C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*: 1165-1188.

Chen, J., Lu, W. and Zavalina, O. (2019). Organizing data, information, and knowledge in big data environments. In *Proceedings of the 18th Joint Conference on Digital Libraries*, 459-460.

Cheng, X., Fu, S., Sun, J., Bilgihan, A. and Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management*, 71: 366-377.

CHGE (Centre for Health and the Global Environment). (2016). Climate Change and Biodiversity Loss. Retrieved February 2022 from http://www.chgeharvard.org/topic/climatechange-and-biodiversity-loss

Chishakwe, N., Murray, L. and Chambwera, M. (2012). *Building climate change adaptation on community experiences: Lessons from community-based natural resource management in southern Africa*. London: International Institute for Environment and Development. Retrieved from http://pubs. iied.org/pdfs/10030IIED.pdf

Choi, S., Mattila, A.S., Van Hoof, H.B. and Quadri-Felitti, D. (2017). The role of power and incentives in inducing fake reviews in the tourism industry. *Journal of Travel Research*, 56(8): 975-987.

Choudhary, F. S., and Sharma, A. (2022). Exploring antecedents of electronic word-of-mouth in tourism: a case of Tripadvisor. In *Handbook on Tourism and Social Media*, 322-336.

Chua, A. Y., and Banerjee, S. (2013). Reliability of reviews on the Internet: The case of TripAdvisor. In *World Congress on Engineering and Computer Science: International Conference on Internet and Multimedia Technologies*, 453-457

Chu, M., Chen, Y., Yang, L., and Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. *Frontiers in Psychology*, 13: 1029945.

Cocolas, N., Walters, G., Ruhanen, L. (2016). Behavioral adaptation to climate change among winter alpine tourists: an analysis of tourist motivations and leisure substitutability. *Journal of Sustainable Tourism*, 24(6): 846-865.

Conradie, D. C. U., and Kumirai, T. (2012). The creation of a south African climate map for the quantification of appropriate passive design responses. In *4th CIB International Conference on Smart and Sustainable Built Environments*, Sao Paulo, Brazil (1–8). International Council for Research and Innovation.

Cremers, H. R., Wager, T. D., and Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PloS One*, 12(11): 0184923.

Dallmeier, E. C. (2021). Computer vision-based web scraping for internet forums. In *2021 7th International Conference on Optimisation and Applications (ICOA)*, 1-5

Damm, A., Köberl, J. and Prettenthaler, F. (2014). Does artificial snow production pay under future climate conditions? –A case study for a vulnerable ski area in Austria. *Tourism Management*, 43: 8-21.

Davis, N.E. (1968). An optimum summer weather index. *Weather*, 23(8): 305-317.

De Freitas, C. R. (2015). Weather and place-based human behavior: recreational preferences and sensitivity. *International Journal of Biometeorology*, 59: 55-63.

De Freitas, C.R. (2003). Tourism climatology: evaluating environmental information for decision making and business planning in the recreation and tourism sector. *International Journal of Biometeorology*, 48(1): 45-54.

De Freitas, C.R., Scott, D. and McBoyle, G. (2008). A second-generation climate index for tourism (CIT): specification and verification. *International Journal of Biometeorology*, 52(5): 399-407.

De Mauro, A., Greco, M. and Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In *AIP Conference Proceedings*, 1644 (1): 97-104. American Institute of Physics.

de Sandoval, J.L.X. (2024). Rise and fall of TripAdvisor: The Lack of Participation and Its Causes. In *International Conference on Tourism and Information and Communication Technologies* (pp. 93-100). Springer, Cham.

Demiroglu, O. C., Turp, M. T., Kurnaz, M. L., and Abegg, B. (2021). The Ski Climate Index (SCI): fuzzification and a regional climate modeling application for Turkey. *International Journal of Biometeorology*, 65: 763-777.

Denstadli, J.M., Jacobsen, J.K.S., Lohmann, M. (2011). Tourist perceptions of summer weather in Scandinavia. *Annals of Tourism Research*, 38(3): 920-940.

Denstadli, J.M. and Jacobsen, J.K.S. (2014). More clouds on the horizon? Polar tourists' weather tolerances in the context of climate change. *Scandinavian Journal of Hospitality and Tourism*, 14(1): 80-99.

Dogru, T., Marchio, E.A., Bulut, U. and Suess, C. (2019). Climate change: Vulnerability and resilience of tourism and the entire economy. *Tourism Management*, 72: 292-305.

Dong, Y., Hu, S., and Zhu, J. (2018). From source credibility to risk perception: How and when climate information matters to action. *Resources, Conservation and Recycling*, 136: 410-417.

Doran, A., and Schofield, P. (2023). Tourism and the Seasons. Dubois, G., Ceron, J. P., Gössling, S., and Hall, C. M. (2016). Weather preferences of French tourists: lessons for climate change impact assessment. *Climatic Change*, 136: 339-351.

Downey, A. 2014. *Think Python: How to Think Like a Computer Scientist*. Green Tea Press.

Dube, K. and Nhamo, G. (2020). Tourist perceptions and attitudes regarding the impacts of climate change on Victoria Falls. *Bulletin of Geography. Socio-economic Series*, 47(47): 27-44.

Dube, K. and Nhamo, G. (2020). Vulnerability of nature-based tourism to climate variability and change: Case of Kariba resort town, Zimbabwe. *Journal of Outdoor Recreation and Tourism*, 29: 100-281.

Dube, K. and Nhamo, G. (2018). Climate variability, change and potential impacts on tourism: Evidence from the Zambian side of the Victoria Falls. *Environmental Science and Policy*, 84: 113-123.

Dube, K. (2022). Nature-based tourism resources and climate change in Southern Africa. In *Protected Areas and Tourism in Southern Africa: Conservation Goals and Community Livelihoods*: 161-174. Routledge

Dube, K., Nhamo, G. and Chikodzi, D. (2020). Climate change-induced droughts and tourism: Impacts and responses of Western Cape province, South Africa. *Journal of Outdoor Recreation and Tourism*, 39:100319. https://doi.org/10.1016/j.jort.2020.100319

Dubois, G., Ceron, J.P., Dubois, C., Frias, M.D. and Herrera, S. (2016). Reliability and usability of tourism climate indices. *Earth Perspectives*, 3: 1-8.

Dubois, G., Ceron, J.P., Gössling, S. and Hall, C.M. (2016). Weather preferences of French tourists: lessons for climate change impact assessment. *Climatic Change*. 136(2): 339–351.

Efe, B., Gözet, E., Özgür, E., Lupo, A.R. and Deniz, A. (2022). Spatiotemporal Variation of Tourism Climate Index for Türkiye during 1981–2020. *Climate*, 10(10): 151.

Egger, R., Kroner, M., and Stöckl, A. (2022). Web Scraping: Collecting and Retrieving Data from the Web. In *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (67-82). Cham: Springer International Publishing.

El-Masry, E.A., El-Sayed, M.K., Awad, M.A., El-Sammak, A.A. and Sabarouti, M.A.E. (2022). Vulnerability of tourism to climate change on the Mediterranean coastal area of El Hammam–EL Alamein, Egypt. *Environment, Development and Sustainability*. 24(1): 1145–1165.

Faber, J., and Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental press journal of orthodontics*, 19: 27-29.

Faccia, A., Cavaliere, L.P.L., Petratos, P. and Mosteanu, N.R. (2022). Unstructured over structured, big data analytics and applications in accounting and management. In *Proceedings of the 2022 6th International Conference on Cloud and Big Data Computing* (37-41).

Falk, M. (2013). Impact of long-term weather on domestic and foreign winter tourism demand. *International Journal of Tourism Research*, 15(1): 1-17.

Fang, Y., Yin, J., Wu, B. (2018). Climate change and tourism: A scientometric analysis using CiteSpace. *Journal of Sustainable Tourism*, 26(1): 108-126.

Fergusson, P. (1964). Summer weather at the English seaside. *Weather*, 19(5): 144-146.

Feyrer, J., Sacerdote, B. (2009). Colonialism and modern income: Islands as natural experiments. *The Review of Economics and Statistics*, 91(2): 245-262.

Filieri, R., Acikgoz, F., Ndou, V. and Dwivedi, Y. (2020). Is TripAdvisor still relevant? The influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. *International Journal of Contemporary Hospitality Management*, 33(1): 199-223.

Filieri, R., Alguezaui, S., and McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism management*, 51: 174-185.

Fitchett, J. M., and Hoogendoorn, G. (2018). An analysis of factors affecting tourists' accounts of weather in South Africa. *International Journal of Biometeorology*, 62: 2161-2172.

Fitchett, J. M., and Hoogendoorn, G. (2019). Exploring the climate sensitivity of tourists to South Africa through TripAdvisor reviews. *South African Geographical Journal*, 101(1): 91-109.

Fitchett, J. M., Fortune, S. M., and Hoogendoorn, G. (2020). Tourists' reviews of weather in five Indian Ocean islands. *Singapore Journal of Tropical Geography*, 41(2): 171-189.

Fitchett, J. M., Robinson, D., and Hoogendoorn, G. (2017). Climate suitability for tourism in South Africa. *Journal of Sustainable Tourism*, 25(6): 851-867.

Friedrich, J., Stahl, J., Hoogendoorn, G. and Fitchett, J.M. (2020). Exploring climate change threats to beach tourism destinations: application of the hazard–activity pairs methodology to South Africa. *Weather, Climate, and Society*, 12(3): 529-544.

Gandomi, A. and Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2): 137-144.

Ganzaroli, A., De Noni, I. and van Baalen, P. (2017). Vicious advice: Analysing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. *Tourism Management*, 61: 501-510.

Gavilan, D., Avello, M. and Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66: 53-61.

Gawor, Ł. and Dolnicki, P. (2018). Arctic tourism development with regard to legal regulations and environmental protection. *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 32(2): 289-298.

Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., Hontou, V., Gakis, N., and Lalas, D. P. (2019). Climatic preferences for beach tourism: An empirical study on Greek islands. *Theoretical and Applied Climatology*, 137(1): 667-691.

Giddy, J.K., Fitchett, J.M. and Hoogendoorn, G. (2017). Insight into American tourists' experiences with weather in South Africa. *Bulletin of Geography*. 38(38): 57–71.

Gössling, S. and Hall, C.M. (2006). Uncertainties in predicting tourist flows under scenarios of climate change. *Climatic Change*, 79(3): 163-173.

Gössling, S. and Scott, D. (2018). The decarbonisation impasse: Global tourism leaders' views on climate change mitigation. *Journal of Sustainable Tourism*, 26(12): 2071-2086.

Gössling, S., and Schumacher, K. P. (2010). Implementing carbon neutral destination policies: Issues from the Seychelles. *Journal of Sustainable Tourism*, 18(3): 377–391.

Gössling, S., Abegg, B. and Steiger, R. (2016). "It was raining all the time!": Ex post tourist weather perceptions. *Atmosphere*. 7(1): 1-10

Gössling, S., Hall, C.M. and Scott, D. (2009). The challenges of tourism as a development strategy in an era of global climate change. In E. Paluso (ed.), *Rethinking development in a carbon-constrained world. Development cooperation and climate change*. Ministry of Foreign Affairs Helsinki: 100-119. Available at http://gci.org.uk/Documents/Paluso_Finland.pdf (accessed 6 February 2025)

Gössling, S., Scott, D., Hall, C.M., Ceron, J.P., Dubois, G. (2012). Consumer behaviour and demand response of tourists to climate change. *Annals of Tourism Research*, 39(1): 36-58.

Gounden, D. (2021). *An assessment of visitor profiles, consumption patterns and perceptions as well as the state of coastal and marine tourism (specifically beach) sites in KwaZulu-Natal Province, South Africa* (Doctoral dissertation).

Grenfell, S.E., Callaway, R.M., Grenfell, M.C., Bertelli, C.M., Mendzil, A.F. and Tew, I. (2016). Will a rising sea sink some estuarine wetland ecosystems? *Science of the Total Environment*, 554: 276-292.

Guo, Y., Barnes, S.J. and Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59: 467-483.

Hall, C.M. (2015). Islands of sustainability or analogues of the challenge of sustainable development? *Routledge International Handbook of Sustainable Development*: 55-73.

Hall, C.M. (2019). Constructing sustainable tourism development: The 2030 agenda and the managerial ecology of sustainable tourism. *Journal of Sustainable Tourism*, 27(7): 1044-1060.

Hambira, W.L. and Mbaiwa, J.E. (2020). Tourism and climate change in Africa. In: *Routledge Handbook of Tourism in Africa*. Routledge: 98–116.

Hambira, W.L., Saarinen, J. (2015). Policymakers' perceptions of the tourism–climate change nexus: Policy needs and constraints in Botswana. *Development Southern Africa*, 32(3): 350-362.

Hambira, W.L., Saarinen, J., Manwa, H. and Atlhopheng, J.R. (2013). Climate change adaptation practices in nature-based tourism in Maun in the Okavango Delta area, Botswana: How prepared are the tourism businesses? *Tourism Review International*, 17(1): 19-29.

Hamilton, J.M. and Lau, M.A. (2006). The role of climate information in tourist destination choice decision making. *Tourism and Global Environmental Change*, 243-264

Hamilton, J.M., Maddison, D.J. and Tol, R.S. (2005). Climate change and international tourism: a simulation study. *Global environmental change*, 15(3): 253-266.

Hamilton, L.C., Rohall, D.E., Brown, B.C., Hayward, G.F. and Keim, B.D. (2003). Warming winters and New Hampshire's lost ski areas: an integrated case study. *International Journal of Sociology and Social Policy*, 23(10): 52-73.

Han, S., Anderson, C.K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. *Cornell Hospitality Quarterly*, 62(1): 89-104.

Hansen, J., Sato, M., Kharecha, P. and Von Schuckmann, K. (2011). Earth's energy imbalance and implications. *Atmospheric Chemistry and Physics*, 11(24): 13421-13449.

Harlfinger, O. (1991). Holiday bioclimatology: a study of Palma de Majorca, Spain. *GeoJournal*, 25(4): 377-381.

Hasan, K., Abdullah, S.K., Islam, F. and Neela, N.M. (2020). An integrated model for examining tourists' revisit intention to beach tourism destinations. *Journal of Quality Assurance in Hospitality and Tourism*, 21(6): 716-737.

Henama, U.S. (2019). Tourism as a neoliberal economic messiah: the case of South Africa. In *Neoliberalism in the Tourism and Hospitality Sector* (29-51). IGI Global.

Hewer, M.J., Scott, D.J. and Gough, W.A. (2017). Differences in the importance of weather and weather-based decisions among campers in Ontario parks (Canada). *International Journal of Biometeorology*. 61(10): 1805–1818.

Hickmann, T., Widerberg, O., Lederer, M. and Pattberg, P. (2021). The United Nations Framework Convention on Climate Change Secretariat as an orchestrator in global climate policymaking. *International Review of Administrative Sciences*, 87(1): 21-38.

Hidayat, N.M., Hidayati, R., Turyanti, A. and Al Maula, S.F. (2024). Modification of The Thermal Comfort Index Based on Perceptions for Urban Tourism Around Jakarta. *Jurnal Meteorologi dan Geofisika*, 25(1): 1-15.

Hong, J. E. (2020). Qualitative Analysis of Users' Negative Reviews on TripAdvisor: International Tourists' Reviews on Gyeongbokgung Palace in Seoul, Korea. *International Journal of Geospatial and Environmental Research*, 7(1): 5.

Hoogendoorn, G. and Fitchett, J.M. (2018). Tourism and climate change: A review of threats and adaptation strategies for Africa. *Current Issues in Tourism*, 21(7): 742-759.

Hsiang, S. and Kopp, R.E. (2018). An economist's guide to climate change science. *Journal of Economic Perspectives*, 32(4): 3-32.

Hübner, A., Gössling, S. (2012). Tourist perceptions of extreme weather events in Martinique. *Journal of Destination Marketing and Management*, 1(1-2): 47-55.

Huda, C., Heryadi, Y. and Budiharto, W. (2024). A tourism dataset from historical transaction for recommender systems. *Data in Brief*, 52:109990.

Hunt, A. and Watkiss, P. (2011). Climate change impacts and adaptation in cities: a review of the literature. *Climatic change*, 104(1): 13-49.

Intergovernmental Panel on Climate Change (IPCC). (2018). Summary for Policymakers. In: Masson-Delmotte V, Zhai P, Pörtner HO, et al. (eds.), Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emissions pathways, in the context of strengthen the global response to the threat of climate change, sustainable development and efforts to eradicate poverty. Switzerland: World Meteorological Organisation.

IPCC (2023). Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental

Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. *IPCC,* 1–34. doi: https://doi.org/10.59327/ipcc/ar6-9789291691647.001.

International Finance Corporation. (2017). *Twenty Reasons Sustainable Tourism Counts for Development*. World Bank.

Isobe, M. (2013). Impact of global warming on coastal structures in shallow water. *Ocean Engineering*, 71: 51-57.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C. and Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, 79(4): 839-880.

Jankowski, K.R., Flannelly, K.J. and Flannelly, L.T. (2018). The t-test: An influential inferential tool in chaplaincy and other healthcare research. *Journal of health care chaplaincy*, 24(1): 30-39.

Jarvis, D., Stoeckl, N. and Liu, H.B. (2016). The impact of economic, social and environmental factors on trip satisfaction and the likelihood of visitors returning. *Tourism Management*, 52: 1-18.

Jawtusch, J. (2014). Climate change and disaster risk guide: Which climate and disaster risks affect South African and how do they affect South Africa? Retrieved January 16, 2015, from. http:// www.brotfueralle.ch/fileadmin/deutsch/2_Entwicklungpolitik_allgemein/

Jeacle, I., and Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organisations and Society*, 36(4-5): 293-309.

Jeuring, J.H.G. (2017). Weather perceptions, holiday satisfaction and perceived attractiveness of domestic vacationing in The Netherlands. *Tourism Management*. 61: 70–81.

Johnson, P. A., Sieber, R. E., Magnien, N., and Ariwi, J. (2012). Automated web harvesting to collect and analyse user-generated content for tourism. *Current Issues in Tourism*, 15(3): 293-299.

Jury, M.R. (2013). Climate trends in southern Africa. *South African Journal of Science*, 109: 1-11.

Kádár, B. and Gede, M. (2021). Tourism flows in large-scale destination systems. *Annals of Tourism Research*, 87: 103-113.

Kaján, E., Saarinen, J. (2013). Tourism, climate change and adaptation: A review. *Current Issues in Tourism*, 16(2): 167-195.

Kämpfer, S. and Mutz, M. (2013). On the sunny side of life: Sunshine effects on life satisfaction. *Social Indicators Research*, 110: 579-595.

Kaplan, A.M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1): 59-68.

Kareem okab Alomar, A. and khayoun Almhameed, A. (2024). Temperatures and Their Impact on Tourism Activity in Dhi Qar Governorate: An Applied Climate Study. *International Journal of Religion*, 5(10): 606-617.

Keeley, J.E. and Syphard, A.D. (2016). Climate change and future fire regimes: examples from California. *Geosciences*, 6(3): 37.

Khan, A., Chenggang, Y., Hussain, J., Bano, S. and Nawaz, A. (2020). Natural resources, tourism development, and energy-growth-CO2 emission nexus: A simultaneity modelling analysis of BRI countries. *Resources Policy*, 68: 101751.

Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and Its Applications*, 13(3): 145-168.

Kim, S., Park, J.H., Lee, D.K., Son, Y.H., Yoon, H., Kim, S. and Yun, H.J. (2017). The impacts of weather on tourist satisfaction and revisit intention: a study of South Korean domestic tourism. *Asia Pacific Journal of Tourism Research*, 22(9): 895-908.

Kipkoech Anderson, K., Emmanuel, T. and Bangali Solomon, F.A.R.A., 2015. State of Knowledge on CSA in Africa. In *Case Studies from Ethiopia, Kenya and Uganda Forum for Agricultural Research in Africa, Accra, Ghana ISBN*, 978-9988.

Kirilenko, A.P., Stepchenkova, S.O. and Hernandez, J.M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management*, 72: 400-410.

Köberl, J., Prettenthaler, F. and Bird, D.N. (2016). Modelling climate change impacts on tourism demand: A comparative study from Sardinia (Italy) and Cap Bon (Tunisia). *Science of the Total Environment*, 543: 1039-1053.

Koutroulis, A.G., Grillakis, M.G., Tsanis, I.K. and Jacob, D. (2018). Mapping the vulnerability of European summer tourism under 2 C global warming. *Climatic Change*, 151(2): 157-171.

Kovács, A. and Unger, J. (2014). Modification of the Tourism Climatic Index to Central European climatic conditions–examples. *Quarterly Journal of The Hungarian Meteorological Service*, 118(2): 147-166.

Kovács, A., Unger, J., Gál, C.V. and Kántor, N. (2016). Adjustment of the thermal component of two tourism climatological assessment tools using thermal perception and preference surveys from Hungary. *Theoretical and Applied Climatology*, 125(1): 113-130.

Kozak, M., Gokovali, U. and Bahar, O. (2008). Estimating the determinants of tourist spending: a comparison of four models. *Tourism Analysis*, 13(2): 143-155.

Kruger, A.C. and Nxumalo, M.P. (2017). Historical rainfall trends in South Africa: 1921–2015. *Water SA*, 43(2): 285-297.

Kurukulasuriya, P., Mendelsohn, R., Hassan, R., Benhin, J., Deressa, T., Diop, M., Eid, H.M., Fosu, K.Y., Gbetibouo, G., Jain, S. and Mahamadou, A. (2006). Will African agriculture survive climate change? *The World Bank Economic Review*, 20(3): 367-388.

Kyriakidis, A., Felton, J. (2008). Too hot to handle? The hospitality industry faces up to climate change. *The Travel and Tourism Competitiveness Report. Geneva: World Economic Forum*.

Lai, W.H. and Vinh, N.Q. (2013). Online promotion and its influence on destination awareness and loyalty in the tourism industry. *Advances in Management and Applied Economics*, 3(3): 15.

Landman, W.A., Malherbe, J. and Engelbrecht, F. (2017). South Africa's present-day climate. *Understanding the Social and Environmental Implications of Global Change*: 7-12.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., and Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, 21(4): 475.

Lauer, C., Brumberger, E., and Beveridge, A. (2018). Hand collecting and coding versus data-driven methods in technical and professional communication research. *IEEE Transactions on Professional Communication*, 61(4): 389-408.

Law, R. (2006). Internet and tourism—part XXI: TripAdvisor. *Journal of Travel and Tourism Marketing*, 20(1): 75-77.

Leal Filho, W., Dinis, M.A.P., Nagy, G.J., Fracassi, U. and Aina, Y.A. (2024). A ticket to where? Dwindling snow cover impacts the winter tourism sector as a consequence of climate change. *Journal of Environmental Management*, 356: 120554.

Lee, J.J., Gino, F. and Staats, B.R. (2014). Rainmakers: Why bad weather means good productivity. *Journal of Applied Psychology*, 99(3): 504.

Lee, J., Benjamin, S. and Childs, M. (2020). Unpacking the emotions behind TripAdvisor travel reviews: The case study of Gatlinburg, Tennessee. *International Journal of Hospitality and Tourism Administration*, 23(2): 347-364.

Lennard, C. and Hegerl, G. (2015). Relating changes in synoptic circulation to the surface rainfall response using self-organising maps. *Climate Dynamics*, 44: 861-879.

Lennard, C. (2019). Multi-scale drivers of the South African weather and climate. *The Geography of South Africa: Contemporary Changes and New Directions*: 81-89.

Leung, D., Law, R., Van Hoof, H. and Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel and Tourism Marketing*, 30(1-2): 3-22.

Levy, S.E., Duan, W. and Boo, S. (2013). An analysis of one-star online reviews and responses in the Washington, DC, lodging market. *Cornell Hospitality Quarterly*, 54(1): 49-63.

Li, J., Chen, D., Li, X. and Godding, L. (2018). Investigating the association between travelers' individual characteristics and their attitudes toward weather information. *Travel Behaviour and Society*, 10: 53-59.

Li, J., Xu, L., Tang, L., Wang, S. and Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68: 301-323.

Lise, W. and Tol, R.S. (2002). Impact of climate on tourist demand. *Climatic Change*, 55(4): 429-449.

Litvin, S. W., and Dowling, K. M. (2018). TripAdvisor and hotel consumer brand loyalty. *Current Issues in Tourism*, 21(8): 842-846.

Litvin, S.W., Goldsmith, R.E. and Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3): 458-468.

Liu, B. (2012). Sentiment analysis: A fascinating problem. *Sentiment Analysis and Opinion Mining* (1-8). Cham: Springer International Publishing.

Liu, W. Y., Yu, H. W., and Hsieh, C. M. (2021). Evaluating forest visitors' place attachment, recreational activities, and travel intentions under different climate scenarios. *Forests*, 12(2): 171.

Loehr, J. and Becken, S. (2021). The tourism climate change knowledge system. *Annals of Tourism Research*, 86: 103073.

Lohmann, M. and Hübner, A.C. (2013). Tourist behavior and weather. Understanding the role of preferences, expectations and in-situ adaptation. *Mondes du Tourisme*, (8): 44-59.

Lourens, A.S.M. (2012). *Air quality in the Johannesburg–Pretoria megacity: its regional influence and identification of parameters that could mitigate pollution.* (Doctoral thesis submitted to North-West University).

Low, P. (2024). Computerised Content Analysis of TripAdvisor Reviews: A Case Study of a Street Food Restaurant with One Michelin Star. *TEM Journal*, 13(1): 326-333.

Lu, J.L. and Wang, C.Y. (2018). Investigating the impacts of air travellers' environmental knowledge on attitudes toward carbon offsetting and willingness to mitigate the environmental impacts of aviation. *Transportation Research Part D: Transport and Environment*, 59: 96-107.

Lukoseviciute, G. and Panagopoulos, T. (2021). Management priorities from tourists' perspectives and beach quality assessment as tools to support sustainable coastal tourism. *Ocean and Coastal Management*, 208: 105646.

Luo, H., Zhu, Y., Li, Q., Xu, Z., Hong, Y., Ghosh, P., Kaur, S., Wu, M., Yang, C. and Qiu, M., (2020). Outdoor personal heating and cooling by a Janus textile. https://doi.org/10.21203/rs.3.rs-118029/v1

Ma, S., Craig, C.A. and Feng, S. (2021). Camping climate resources: The camping climate index in the United States. *Current Issues in Tourism*, 24(18): 2523-2531.

Mahto, D.K., Singh, L. (2016). A dive into Web Scraper world. *International Conference on Computing for Sustainable Global Development (INDIACom)*, (3): 689-693.

Malisawa, M.S. and Rautenbach, C.D.W. (2012). Evaluating water scarcity in the Southern African Development Community (SADC) region by using a climate moisture index (CMI) indicator. *Water Science and Technology: Water Supply*, 12(1): 45-55.

Mane, A.B. (2016). From a brewer to the faraday of statistics: William Sealy Gosset. *Muller Journal of Medical Sciences and Research*, 7(2): 147-149.

Marchi, V., Apicerni, V., Marasco, A. (2021). Assessing Online Sustainability Communication of Italian Cultural Destinations–A Web Content Mining Approach. *Information and Communication Technologies in Tourism 2021*: 58-69. Springer, Cham.

Mardani, M., Mardani, H., De Simone, L., Varas, S., Kita, N., Saito, T. (2019). Integration of machine learning and open access geospatial data for land cover mapping. *Remote Sensing*, 11(16): 1907.

Mariani, M. (2019). Big data and analytics in tourism and hospitality: A perspective article. *Tourism Review*, 75(1): 299-303

Marine-Roig, E. and Clavé, S.A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing and Management*, 4(3): 162-172.

Marshall, N.A., Marshall, P.A., Abdulla, A., Rouphael, T. and Ali, A. (2011). Preparing for climate change: recognising its early impacts through the perceptions of dive tourists and dive operators in the Egyptian Red Sea. *Current Issues in Tourism*, 14(6): 507-518.

Martín, J.C., Saayman, M. and du Plessis, E. (2019). Determining satisfaction of international tourist: A different approach. *Journal of Hospitality and Tourism Management*, 40: 1-10.

Martín, J.M.M., Aguilera, J.D.D.J. and Moreno, V.M. (2014). Impacts of seasonality on environmental sustainability in the tourism sector based on destination type: An application to Spain's Andalusia region. *Tourism Economics*, 20(1): 123-142.

Martín, M.B.G. (2005). Weather, climate and tourism a geographical perspective. *Annals of tourism research*, 32(3): 571-591.

Martín, J.C., Saayman, M. and du Plessis, E. (2019). Determining satisfaction of international tourist: A different approach. *Journal of Hospitality and Tourism Management*, *40*: 1-10.

Martin, J.C., Román, C., Moreira, P., Moreno, R. and Oyarce, F. (2021). Does the access transport mode affect visitors' satisfaction in a World Heritage City? The case of Valparaiso, Chile. *Journal of Transport Geography*, 91: 102969.

Matthews, L., Scott, D., Andrey, J. (2021). Development of a data-driven weather index for beach parks tourism. *International Journal of Biometeorology*, 65(5): 749-762.

Matzarakis, A. (2001). Climate and bioclimate information for tourism in Greece. In *Proceedings of the first international workshop on climate, tourism and recreation, ed. by A. Matzarakis and CR de Freitas. International Society of Biometeorology, Commission on Climate Tourism and Recreation*: 171-183

Mbaiwa, J.E. and Mmopelwa, G. (2014). Chapter forty-one perceived effects of climate change on The Tourism business in the Okavango Delta, Botswana. In *Meeting the challenges of*

*Climate Change to Tourism: Case studies of best practice*, 417. Cambridge Scholars Publishing.

McKercher, B., Prideaux, B., and Thompson, M. (2023). The impact of changing seasons on in-destination tourist behaviour. *Tourism Review*, 79(6): 1197-1209

Mieczkowski, Z. (1985). The tourism climatic index: a method of evaluating world climates for tourism. *Canadian Geographer/Le Géographe Canadien*, 29(3): 220-233.

Min, H., Lim, Y. and Magnini, V.P. (2015). Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed. *Cornell Hospitality Quarterly*, 56(2): 223-231.

Minor, K., Jensen, M.L., Hamilton, L., Bendixen, M., Lassen, D.D. and Rosing, M.T. (2023). Experience exceeds awareness of anthropogenic climate change in Greenland. *Nature Climate Change*, 13(7): 661-670.

Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. " O'Reilly Media, Inc.".

Mkono, M. (2012). A netnographic examination of constructive authenticity in Victoria Falls tourist (restaurant) experiences. *International Journal of Hospitality Management*, 31(2): 387-394.

Mokgehle, D. and Fitchett, J.M. (2024). Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies. *Tourism Planning & Development*, 21(6): 1–21. https://doi.org/10.1080/21568316.2024.2412060.

Mora, C., Spirandelli, D., Franklin, E.C., Lynham, J., Kantar, M.B., Miles, W., Smith, C.Z., Freel, K., Moy, J., Louis, L.V. and Barba, E.W. (2018). Broad threat to humanity from cumulative

climate hazards intensified by greenhouse gas emissions. *Nature Climate Change*, 8(12): 1062-1071.

Morabito, M., Crisci, A., Barcaioli, G. and Maracchi, G. (2004). Climate change: the impact on tourism comfort at three Italian tourist sites. *Advances in Tourism Climatology. Berichte des Meteorologischen Institutes der Universtät Freiburg*, 12: 56-64.

Moreno, A. 2017. *Climate change impacts: the vulnerability of tourism in coastal Europe*. Routledge: 30-47

Morgan, R., Gatell, E., Junyent, R., Micallef, A., Özhan, E., and Williams, A. T. (2000). An improved user-based beach climate index. *Journal of Coastal Conservation*, 6: 41-50.

Munro, A. (2020). Using experimental manipulation of questionnaire design and a Kenyan panel to test for the reliability of reported perceptions of climate change and adaptation. *Climatic Change*, 162: 1081-1105.

Murray, R. (1972). A simple summer index with an illustration for summer 1971. *Weather*, 27(4): 161-169.

Mushawemhuka, W.J., Fitchett, J.M. and Hoogendoorn, G. (2021). Towards quantifying climate suitability for Zimbabwean nature-based tourism. *South African Geographical Journal*, 103(4): 443-463.

Mustapha, S., Man, M., Bakar, W.A.W.A., Yusof, M.K. and Sabri, I.A.A. (2024). A Demystified Overview of Data Scraping. *International Journal of Data Science and Advanced Analytics*, 6(6): 290-296.

Mutz, M. and Kaempfer, S. (2011). ... and Now the Weather: Does Weather Influence the Assessment of Political and Economic Issues? *Zeitschrift Fur Soziologie*, 40(4): 208-226.

Nachmany, M., Fankhauser, S., Townshend, T., Collins, M., Landesman, T., Matthews, A., Pavese, C., Rietig, K., Schleifer, P. and Setzer, J. (2014). *The Globe climate legislation study: a review of climate change legislation in 66 countries.* Global Legislators' Organisation (GLOBE).

Nasrollahi, N., Hatami, Z. and Taleghani, M. (2017). Development of outdoor thermal comfort model for tourists in urban historical areas; A case study in Isfahan. *Building and Environment*, 125: 356-372.

Ngoc, K.M. and Trinh, N.T. (2015). Factors affecting tourists' return intention towards Vung Tau City, Vietnam-A mediation analysis of destination satisfaction. *Journal of Advanced Management Science*, 3(4): 292-298.

Ngxongo, N. A. (2021). The impact of climate change on visitor destination selection: A case study of the Central Drakensberg Region in KwaZulu-Natal. *Jàmbá: Journal of Disaster Risk Studies*, 13(1): 1161. https://doi.org/10.4102/jamba.v13i1.1161

Nicholson, S.E. (2000). The nature of rainfall variability over Africa on time scales of decades to millenia. *Global and Planetary Change*, 26(1-3): 137-158.

Nicolaides, A. (2020). Sustainable ethical tourism (SET) and rural community involvement. *African Journal of Hospitality, Tourism and Leisure*, 9(1): 1-16.

Nilashi, M., Samad, S., Ahani, A., Ahmadi, H., Alsolami, E., Mahmoud, M., Majeed, H.D. and Alarood, A.A. (2021). Travellers' decision making through preferences learning: A case on Malaysian spa hotels in TripAdvisor. *Computers and Industrial Engineering*, 158: 107-348.

Noome, K. and Fitchett, J.M. (2019). An assessment of the climatic suitability of Afriski Mountain Resort for outdoor tourism using the Tourism Climate Index (TCI). *Journal of Mountain Science*, 16(11): 2453-2469.

O'connor, P. (2008). User-generated content and travel: A case study on TripAdvisor. com. *Enter*, 47-58.

Oviedo-García, M.Á., Vega-Vázquez, M., Castellanos-Verdugo, M. and Orgaz-Agüera, F. (2019). Tourism in protected areas and the impact of servicescape on tourist satisfaction, key in sustainability. *Journal of Destination Marketing and Management*, 12: 74-83.

Oyedotun, T.D.T., Ruiz-Luna, A. and Navarro-Hernández, A.G. (2018). Contemporary shoreline changes and consequences at a tropical coastal domain. *Geology, Ecology, and Landscapes*, 2(2): 104-114.

Padgavankar, M.H. and Gupta, S.R. (2014). Big data storage and challenges. *International Journal of Computer Science and Information Technologies*, 5(2): 2218-2223.

Pandy, W.R. and Rogerson, C.M. (2019). Urban tourism and climate change: Risk perceptions of business tourism stakeholders in Johannesburg, South Africa. *Urbani Izziv*, 30: 225-243.

Park, J.Y. and Jang, S. (2018). The impact of sold-out information on tourist choice decisions. *Journal of Travel and Tourism Marketing*, 35(5): 622-632.

Perch-Nielsen, S.L., Amelung, B. and Knutti, R. (2010). Future climate resources for tourism in Europe based on the daily Tourism Climatic Index. *Climatic change*, 103(3): 363-381.

Pereira, S., Araújo, S. and Lopes, S. (2023). Web Scraping Method for Extracting Search Results' Data: Implementation for Airbnb Experiences. *Advances in Tourism, Technology and Systems: Selected Papers from ICOTTS 2022*, 1:(467-479).

Phillips, M.R., Jones, A.L. and Thomas, T. (2018). Climate change, coastal management and acceptable risk: Consequences for tourism. *Journal of Coastal Research*, 85(10085): 1411-1415.

Phuyal, R.K., Bhuju, D., Devkota, T., Acharya, S., Bishwokarma, D. and Devkota, N. (2020). Impact of Climate Change on Tourism Sector in Nepal. *Journal of Development and Administrative Studies*, *28*(1-2): 19-34.

Preston-Whyte, R.A. and Watson, H.K. (2005). Nature tourism and climatic change in Southern Africa. *Tourism, Recreation and Climate Change*: 130-142.

Priego, F.J., Rosselló, J. and Santana-Gallego, M. (2015). The impact of climate change on domestic tourism: a gravity model for Spain. *Regional Environmental Change*, 15(2): 291-300.

Pröbstl-Haider, U., Mostegl, N. and Damm, A. (2021). Tourism and climate change–A discussion of suitable strategies for Austria. *Journal of Outdoor Recreation and Tourism*, 34: 100394.

R.-Toubes, D., Araújo-Vila, N., and Fraiz-Brea, J. A. (2020). Influence of Weather on the Behaviour of Tourists in a Beach Destination. *Atmosphere*, 11(1): 121.

Ranjanthran, M. and Mohammed, B. (2010). Domestic Tourism: Perception of domestic tourist on tourism products in Penang Island. *Asian Journal of Management Research*, 1(2): 795-816.

Reason, C.J.C. (2007). Tropical cyclone Dera, the unusual 2000/01 tropical cyclone season in the Southwest Indian Ocean and associated rainfall anomalies over Southern Africa. *Meteorology and Atmospheric Physics*, 97(1): 181-188.

Reid, P., Vogel, C. (2006). Living and responding to multiple stressors in South Africa—Glimpses from KwaZulu-Natal. *Global Environmental Change*, 16(2): 195-206.

Rhemtulla, M., and Bottesini, J. G. (2022). On the reliability of reliability estimates in infancy research. *Infant and Child Development*, 31(5): e2327.

Richins, H. and Scarinci, J. (2009). Climate change and sustainable practices: A case study of the resort industry in Florida. *Tourismos*, 4(2): 107-128.

Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P., and Martens, P. (2014). Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: Evidence from Aruba. *Tourism Management*, 41: 245-256.

Rodríguez-Díaz, M. and Espino-Rodríguez, T.F. (2018). A methodology for a comparative analysis of the lodging offer of tourism destinations based on online customer reviews. *Journal of Destination Marketing and Management*, 8: 147-160.

Rogerson, C.M. (2016). Climate change, tourism and local economic development in South Africa. *Local Economy*, 31(1-2): 322-331.

Rogerson, C.M. (2002). Tourism-led local economic development: The South African experience. *Urban Forum*, 13(1): 95-119

Rogerson, C.M. and Rogerson, J.M. (2021). Niche tourism research and policy: International and South African debates. *African Journal of Hospitality, Tourism and Leisure*, 10(4): 1131-1151.

Rogerson, J.M. and Sims, S.R. (2012). The greening of urban hotels in South Africa: Evidence from Gauteng. In *Urban forum*. Springer Netherlands, 23(3): 391-407

Rony, M.K.K. and Alamgir, H.M. (2023). High temperatures on mental health: Recognizing the association and the need for proactive strategies—A perspective. *Health Science Reports*, 6(12): e1729.

Rosselló-Nadal, J. (2014). How to evaluate the effects of climate change on tourism. *Tourism Management*, 42: 334-340.

Rouault, M., Roy, S.S. and Balling Jr, R.C. (2013). The diurnal cycle of rainfall in South Africa in the austral summer. *International Journal of Climatology*, 33(3): 770-777.

Rutty, M. and Scott, D. (2010). Will the Mediterranean become "too hot" for tourism? A reassessment. *Tourism and Hospitality Planning and Development*, 7(3): 267-281.

Rutty, M. and Scott, D. (2014). Thermal range of coastal tourism resort microclimates. *Tourism Geographies*, 16(3): 346-363.

Rutty, M. and Scott, D. (2015). Bioclimatic comfort and the thermal perceptions and preferences of beach tourists. *International Journal of Biometeorology*, 59: 37-45.

Rutty, M., Scott, D., Matthews, L., Burrowes, R., Trotman, A., Mahon, R., and Charles, A. (2020). An inter-comparison of the Holiday Climate Index (HCI: Beach) and the Tourism Climate Index (TCI) to explain Canadian tourism arrivals to the Caribbean. *Atmosphere*, 11(4): 412.

Saarinen, J. and Rogerson, C.M. (2014). Tourism and the millennium development goals: Perspectives beyond 2015. *Tourism Geographies*, 16(1): 23-30.

Saarinen, J., Fitchett, J. and Hoogendoorn, G. (2022). *Climate Change and Tourism in Southern Africa*. Routledge.

Sailunaz, K., and Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36: 101003.

Sangkaew, N. and Zhu, H. (2020). Understanding tourists' experiences at local markets in phuket: An analysis of TripAdvisor reviews. *Journal of Quality Assurance in Hospitality and Tourism*. 23(1): 89-114.

Sangkaew, N., and Zhu, H. (2022). Understanding tourists' experiences at local markets in Phuket: An analysis of TripAdvisor reviews. *Journal of Quality Assurance in Hospitality and Tourism*, 23(1): 89-114.

Santos-Lacueva, R., Clavé, S.A. and Saladié, Ò. (2017). The vulnerability of coastal tourism destinations to climate change: The usefulness of policy analysis. *Sustainability*, 9(11): 2062.

Satyawan, I.A., Wibisono, B.I. and Binangun, P.S. (2021). The impact of climate change in tourism sector in Java Island: a literature review. In *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, 724(1): 012111

Saurkar, A. V., Pathare, K. G., and Gode, S. A. (2018). An overview on web scraping techniques and tools. *International Journal on Future Revolution in Computer Science and Communication Engineering*, 4(4): 363-367.

Scalabrini, E., Ferreira, J., Zekan, S.B. and Fernandes, P.O. (2023). Destination Image Through TripAdvisor´ s Reviews Analysis. *Communication Design and Branding: A Multidisciplinary Approach*, 205-221.

Scheyvens, R. and Hughes, E., 2021. Can tourism help to "end poverty in all its forms everywhere"? The challenge of tourism addressing SDG1. In *Activating critical thinking to advance the sustainable development goals in tourism systems*, 215-233.

Schuckert, M., Liu, X. and Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel and Tourism Marketing*, 32(5): 608-621.

Scott, D. and Gössling, S. (2018). *Tourism and Climate Change Mitigation: Embracing the Paris Agreement: Pathways to Decarbonisation*. European Travel Commission: Brussels, Belgium. Available at https://etc-corporate.org/uploads/2018/03/ETC-Climate-Change-Report_FINAL.pdf(open in a new window) (accessed June 2022)

Scott, D., and Lemieux, C. (2010). Weather and climate information for tourism. *Procedia Environmental Sciences*, 1: 146-183.

Scott, D., Amelung, B., Ceron, J.P., Dubois, G., Gössling, S., Peeters, P. and Simpson, M.C. (2008). *Climate change and tourism: Responding to global challenges*, 230: 1-38

Scott, D., Gössling, S. and Hall, C.M. (2012). International tourism and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3): 213-232.

Scott, D., Hall, C.M. and Gössling, S. (2019). Global tourism vulnerability to climate change. *Annals of Tourism Research*, 77: 49-61.

Scott, D., Hall, C.M. and Gössling, S. (2016). A review of the IPCC Fifth Assessment and implications for tourism sector climate resilience and decarbonisation. *Journal of Sustainable Tourism*, 24(1): 8-30.

Scott, D., Hall, C.M. and Stefan, G. (2012). *Tourism and climate change: Impacts, adaptation and mitigation*. Routledge.

Scott, D., Rutty, M. and Peister, C. (2018). Climate variability and water use on golf courses: optimisation opportunities for a warmer future. *Journal of Sustainable Tourism*, 26(8): 1453-1467.

Scott, D., Rutty, M., Amelung, B., and Tang, M. (2016). An inter-comparison of the holiday climate index (HCI) and the tourism climate index (TCI) in Europe. *Atmosphere*, 7(6): 80.

Schulze, R.E. and Maharaj, M. (2007). Rainfall seasonality. *South African Atlas of Climatology and Agrohydrology Water Research Commission, Pretoria, RSA, WRC Report*, 1489(1): 06.

Serdeczny, O., Adams, S., Baarsch, F., Coumou, D., Robinson, A., Hare, W., Schaeffer, M., Perrette, M. and Reinhardt, J. (2017). Climate change impacts in Sub-Saharan Africa: from physical changes to their social repercussions. *Regional Environmental Change*, 17(6): 1585-1600.

Shi, L., Maruthaveeran, S., Yusof, M.J.M. and Dai, C. (2024). Exploring Factors Influencing Recreational Experiences of Urban River Corridors Based on Social Media Data. *Applied Sciences*, 14(10): 4086.

Shin, H., Perdue, R.R. and Pandelaere, M. (2020). Managing customer reviews for value co-creation: An empowerment theory perspective. *Journal of Travel Research*, 59(5): 792-810.

Shambour, Q., Abualhaj, M., Abu-Shareha, A. and Kharma, Q. (2024). Personalized Tourism Recommendations: Leveraging User Preferences and Trust Network. *Interdisciplinary Journal of Information, Knowledge, and Management*, 19: 017.

Siew, A., Lee, H., Lee, A.S.H., Yusoff, Z. and Zainol, Z. (2018). Know your Hotels Well! an Online Review Analysis using Text Analytics. *Article in International Journal of Engineering and Technology*. 7:341–347.

Simeon, M.I., Buonincontri, P., Cinquegrani, F. and Martone, A. (2017). Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*, 8(2): 220-238

Som, A.P.M., Marzuki, A. and Yousefi, M. (2012). Factors influencing visitors' revisit behavioral intentions: A case study of Sabah, Malaysia. *International Journal of Marketing Studies*, 4(4): 39.

Sousa, P.M., Blamey, R.C., Reason, C.J., Ramos, A.M. and Trigo, R.M. (2018). The 'Day Zero'Cape Town drought and the poleward migration of moisture corridors. *Environmental Research Letters*, 13(12): 124025.

Stephan, A. and Stephan, L. (2020). Achieving net zero life cycle primary energy and greenhouse gas emissions apartment buildings in a Mediterranean climate. *Applied Energy*, 280: 115932.

Steyn, J.N. (2012). Managing climate change impacts on tourism: mitigating and adaptive strategies with special reference to the Western Cape Province of South Africa: recreation and tourism. *African Journal for Physical Health Education, Recreation and Dance*, 18(3): 552-564.

Stockigt, L., Hoogendoorn, G., Fitchett, J.M. and Saarinen, J. (2018). Climatic sensitivity and snow-based tourism in Africa: an investigation of TripAdvisor reviews on Afriski, Lesotho. *Submitted for the Proceedings of the Society of South African Geographers Biennial Conference*.

Student. (1908). The probable error of a mean. *Biometrika*, 1-25.

Szulejko, J.E., Kumar, P., Deep, A. and Kim, K.H. (2017). Global warming projections to 2100 using simple CO2 greenhouse gas modeling and comments on CO2 climate sensitivity factor. *Atmospheric Pollution Research*, 8(1): 136-140.

Taecharungroj, V. and Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*. 75: 550–568.

Taecharungroj, V. (2019). User-generated place brand identity: Harnessing the power of content on social media platforms. *Journal of Place Management and Development*, 12(1): 39-70.

Taljaard, J.J. (1994). Atmospheric circulation systems, synoptic climatology and weather phenomena of South Africa. Part 1: Controls of the weather and climate of South Africa. *Africa Weather Service Technology Paper*, *27*.

Tanrisever, C., Pamukçu, H., and Baydeniz, E. (2024). Climate Change in Tourism: understanding the impacts and opportunities for sustainability. In *Future Tourism Trends Volume 1: Tourism in the Changing World*. Emerald Publishing Limited, 33-45

Terziyska, I. and Damyanova, R. (2020). Winescape through the lens of organised travel–a netnography study. *International Journal of Wine Business Research*, 32(4): 477-492.

Tian, X., Zhang, L. and Wei, W. (2016). The Design and Implementation of Automatic Grabbing Tool in Tripadvisor. In *2016 4th International Conference on Enterprise Systems (ES)*. IEEE, 76-80

Terziyska, I. and Damyanova, R. (2020). Winescape through the lens of organised travel–a netnography study. *International Journal of Wine Business Research*, 32(4): 477-492

Tian, X., Deng, Y., Wargocki, P. and Liu, W. (2021). Effects of increased activity level on physiological and subjective responses at different high temperatures. *Building and Environment*, 201: 108011.

Tjaden, J. (2023). Web Scraping for Migration, Mobility, and Migrant Integration Studies: Introduction, Application, and Potential Use Cases. *International Migration Review*, 01979183231208428.

Tol, R.S., O'Neill, B. and van Vuuren, D.P. (2005). A critical assessment of the IPCC SRES scenarios. *Research Unit, Sustainability and Global Change, Hamburg University.*

Tung, V. W. S., and McKercher, B. (2017). Negotiating the rapidly changing research, publishing, and career landscape. *Tourism Management*, 60: 322-331.

Tyson, P.D., Preston-Whyte, R.A. (2000). The Weather and Climate of Southern Africa, Oxford.

Um, S., Chon, K. and Ro, Y., 2006. Antecedents of revisit intention. *Annals of tourism research*, 33(4): 1141-1158.

United Nations World Tourism Organisation (UNWTO), United Nations Environmental Programme (UNEP), World Meteorological Organisation (WMO). (2008). Climate Change and Tourism: Responding to Global Challenges. UNWTO: Madrid–UNEP: Paris–WMO: Geneva.

UNWTO (United Nations World Tourism Organisation). (2019). Tourism highlights. World Tourism Organisation, 2011. *Tourism towards 2030: Global Overview*. UNWTO.

Valova, I., Mladenova, T., Kanev, G. and Halacheva, T. (2023). Web Scraping-State of Art, Techniques and Approaches. In *2023 31st National Conference with International Participation (TELECOM)*, 1-4.

van Heerden, J., Terblanche, D. E., Schulze, G. C. (1988). The southern oscillation and South African summer rainfall. *Journal of Climatology*, 8: 577-597.

Van Vuuren, C. and Slabbert, E. (2011). Tourism-Travel behaviour of tourists to a South African holiday resort. *African Journal for Physical Health Education Recreation and Dance*, 17(4): 694.

vanden Broucke, S., Baesens, B. (2018). From Web Scraping to Web Crawling. In *Practical Web Scraping for Data Science:* 155-172.

Vassakis, K., Petrakis, E., Kopanakis, I., Makridis, J. and Mastorakis, G. (2019). Location-based social network data for tourism destinations. In *Big Data and Innovation in Tourism, Travel, and Hospitality*. Springer, Singapore: 105-114

Velea, L., Gallo, A., Udristioiu, M. T., and Puiu, S. (2022). Rating Tourists' Interest in Tourism-Tailored Climate and Environmental Products. *Journal of Tourism and Services*, 13(25): 230-247.

Verbos, R.I. and Brownlee, M.T. (2017). The Weather Dependency Framework (WDF): A tool for assessing the weather dependency of outdoor recreation activities. *Journal of Outdoor Recreation and Tourism*, 18: 88-99.

Vermeulen, I.E. and Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1): 123-127.

Vu, H. Q., Li, G., Law, R., and Zhang, Y. (2018). Tourist activity analysis by leveraging mobile social media data. *Journal of Travel Research*, 57(7): 883-898.

Vu, H. Q., Li, G., Law, R., and Zhang, Y. (2019). Exploring tourist dining preferences based on restaurant reviews. *Journal of Travel Research*, 58(1): 149-167.

Wagner, S., Mendez, D., Felderer, M., Graziotin, D. and Kalinowski, M. (2020). Challenges in survey research. *Contemporary Empirical Methods in Software Engineering*, 93-125.

Wang, L., Wang, X.K., Peng, J.J. and Wang, J.Q. (2020). The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded rationality behavioural decision support model. *Tourism Management*, 76: 103961.

Warburton, M.L. and Schulze, R.E. (2005). Detection of climate change: a review of literature on changes in temperature, rainfall and streamflow, on detection methods and data problems. *Water Research Commission, Pretoria, RSA*, 257-274.

Wilson, A., Murphy, H. and Fierro, J.C. (2012). Hospitality and travel: The nature and implications of user-generated content. *Cornell Hospitality Quarterly*, 53(3): 220-228.

Wiratama, I., Wijaya, I. and Kenedy, F.V. (2024). Climate Mitigation and Waste Management in The Tourism Industry for A Sustainable Ecosystem. *International Journal of Applied Science and Sustainable Development (IJASSD)*, 6(1): 12-22.

Wong, C.U.I. and Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews-the case of Macau. *Tourism Management Perspectives*, 23: 19-29.

Xiang, Z., Schwartz, Z., Gerdes Jr, J.H. and Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44: 120-130.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2018). Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews. *Information Technology and Tourism*, 18(1): 43-59.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58: 51-65.

Yanda, P.Z. (2010). Climate change impacts, vulnerability and adaptation in Southern Africa. *Sarua Leadership Dialogue Series*, 2(4): 11-30.

Ye, Q., Li, H., Wang, Z. and Law, R. (2014). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. *Journal of Hospitality and Tourism Research*, 38(1): 23-39.

Yong, E.L. (2021). Understanding the economic impacts of sea-level rise on tourism prosperity: Conceptualisation and panel data evidence. *Advances in Climate Change Research*, 12(2): 240-253.

Yoo, J., Eom, J. and Zhou, Y. (2024). Thermal comfort and retail sales: A big data analysis of extreme temperature's impact on brick-and-mortar stores. *Journal of Retailing and Consumer Services*, 77: 103699.

Zaninović, K. and Matzarakis, A. (2009). The bioclimatological leaflet as a means conveying climatological information to tourists and the tourism industry. *International Journal of Biometeorology*, 53(4): 369-374.

Zare, S., Hasheminezhad, N., Sarebanzadeh, K., Zolala, F., Hemmatjo, R., and Hassanvand, D. (2018). Assessing thermal comfort in tourist attractions through objective and subjective procedures based on ISO 7730 standard: A field study. *Urban Climate*, 26: 1-9.

Zeiss, H., Graw, K., and Matzarakis, A. (2022). Impact of the Destination Weather Conditions on Decision and Complaint Behavior of Guests Who Booked Vacation Rentals. *Atmosphere*, 13(12): 1998.

Zeng, Y., Filimonau, V., Wang, L.E. and Zhong, L. (2023). The impact of perceived unfavorable weather on tourist loyalty in high-altitude destinations: The case of the Qinghai-Tibet plateau, China. *Journal of Outdoor Recreation and Tourism*, 43: 100658.

Zeppel, H. (2012). Climate change and tourism in the Great Barrier Reef Marine Park. *Current Issues in Tourism*, 15(3): 287-292.

Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. and Huang, M. (2021). IPCC 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Zhang, Z., Ye, Q., Law, R. and Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4): 694-700.

Zhang, Y., Tan, H., Jiao, Q., Lin, Z., Fan, Z., Xu, D., Xiang, Z., Law, R. and Zheng, T. (2024). A Predictive Model Based on TripAdvisor Textual Reviews: Early Destination Recommendations for Travel Planning. *SAGE Open*, 14(2): 21582440241246434.

Zhou, W., Faturay, F., Driml, S. and Sun, Y.Y. (2024). Meta-analysis of the climate change-tourism demand relationship. *Journal of Sustainable Tourism*, 32: 1762–1783

Ziegler, J., Dearden, P. and Rollins, R. (2012). But are tourists satisfied? Importance-performance analysis of the whale shark tourism industry on Isla Holbox, Mexico. *Tourism Management*, 33(3): 692-701.

# Appendix A: Ethics Letter

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

**HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)**

Registration number: REC-101114-044

11 May 2022

Re: Ms. Dineo Mokgehle (1095685)

Waiver letter number: HRECNMW22/05/05

To whom it may concern,

Ms. Mokgehle is a registered PhD student at the School of Geography, Archaeology and Environmental Studies at the University of the Witwatersrand, Johannesburg. This letter is to confirm that, at the time of writing, Ms. Mokgehle does not need ethical clearance for her PhD study entitled *'Examining the climate sensitivity of tourists in South Africa using TripAdvisor: A big data approach through web scraping'*. This decision has been reached based upon a description of the project supplied by Ms. Mokgehle to the University Human Research Ethics Committee (Non-Medical), which has been evaluated by the Chairs and Deputy Chairs. If, however, if Ms. Mokgehle changes the methods of data collection and analysis for this study, this decision may no longer be valid. If such changes take place, this should be communicated to the University Human Research Ethics Committee (Non-Medical) as soon as possible. This waiver letter is valid until 10 May 2025.

Please feel free to contact me should you require any further information.

Thank you.

Yours sincerely,
S Schoeman

**Shaun Schoeman** (Administrative Officer)
Solomon Mahlangu House, 10th Floor, Room 10004, Jorissen Street, Braamfontein, Johannesburg
Private Bag 3, Wits 2050
T + 27(0)11 717 1408 | E Shaun.Schoeman@wits.ac.za | hrec-medical.researchoffice@wits.ac.za
**www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/**

# Appendix B: Web scraping Code

Python

```
***code generated by Dineo Mokgehle, 2022/06/18***
import csv
from csv import reader
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
import time
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.service import Service as ChromeService
i = 0
with open('LINKS.csv', 'r', encoding="utf-8-sig") as read_obj:
    csv_reader = reader(read_obj)
    for row in csv_reader:
        path_to_file                                                    =
     f"C:/Users/dineo/Desktop/TripAvisorReviews2023/Old/{i}.csv"   #  Modify
     the directory path here
        url = row[0]
driver                                                                  =
     webdriver.Chrome(service=ChromeService(ChromeDriverManager().install()
     ))
driver.get(url)
try:
    WebDriverWait(driver, 10).until(
        EC.presence_of_element_located((By.XPATH,   ".//div[contains(@data-
     test-target, 'expand-review')]"))
    )
except:
    print("Expand review element not found or took too long to load.")
csvFile = open(path_to_file, 'a', encoding="utf-8-sig")
csvWriter = csv.writer(csvFile)

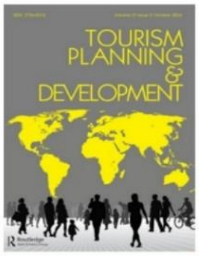
while True:
    time.sleep(2)
    driver.find_element("xpath",        ".//div[contains(@data-test-target,
     'expand-review')]").click()

    container = driver.find_elements("xpath", "//div[@data-reviewid]")
    dates = driver.find_elements("xpath", ".//div[@class='cRVSd']")
    Loc = driver.find_elements("xpath", ".//span[@class='RdTWF']")

    for element in container:
        rating  =  element.find_element("xpath",  ".//span[contains(@class,
     'ui_bubble_rating bubble_')]") \
                .get_attribute("class").split("_")[3]
        title  =  element.find_element("xpath", ".//div[contains(@data-test-
     target, 'review-title')]").text
        review   =   element.find_element("xpath",   ".//q[@class='QewHA   H4
     _a']").text.replace("\n", "  ")
        date  =  "  ".join(dates[container.index(element)].text.split("  ")[-
     2:])
        location = ''
        try:
            reviewer_element = Loc[container.index(element)].text
            location = reviewer_element
```

```
        except Exception as e:
            location = "Location not found"
        csvWriter.writerow([date, rating, location, title, review])
if "ui_button nav next primary disabled" in driver.page_source:
    break
else:
    driver.find_element("xpath", './/a[@class="ui_button nav next primary
     "]').click()
driver.quit()
csvFile.close()
i = i + 1
```

# Appendix C: Publication




Routledge
Taylor & Francis Group

Tourism Planning & Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rthp21

## Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies

Dineo Mokgehle & Jennifer M. Fitchett

**To cite this article:** Dineo Mokgehle & Jennifer M. Fitchett (11 Oct 2024): Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies, Tourism Planning & Development, DOI: 10.1080/21568316.2024.2412060

**To link to this article:** https://doi.org/10.1080/21568316.2024.2412060

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 11 Oct 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rthp21

TOURISM PLANNING & DEVELOPMENT
https://doi.org/10.1080/21568316.2024.2412060

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

# Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies

Dineo Mokgehle and Jennifer M. Fitchett

School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, Johannesburg, South Africa

**ABSTRACT**

The viability and competitiveness of tourism locations depends, in part, on climate. Netnographic research utilising TripAdvisor reviews can objectively evaluate tourists' climatic sensitivity and the meteorological factors most important to them. Manually reading and coding hundreds of reviews is time-consuming. Web-scraping can automate this process and analyse hundreds of thousands of reviews nationwide to evaluate climatic influence more comprehensively. This paper compares the outputs of manually collected and web-scraped TripAdvisor reviews, evaluating the efficacy of a web-scraping approach. The total number of climatic mentions captured varied by a statistically insignificant 3.56%, demonstrating a reasonable efficacy of web-scraping. Out of 26,687 TripAdvisor reviews, 1,163 mentioned climate. This 4.4% mean sensitivity to climatic variables is statistically significantly different to the 7.9% yielded from manual data collection. The two methodologies had minor statistically insignificant differences in climatic references at 19 destinations, ranging from 0.42% to 8.26%. A webscraping approach is thus possible but should be used with caution.

**ARTICLE HISTORY**
Received 6 October 2023
Accepted 24 September 2024

**KEYWORDS**
Climate; weather; TripAdvisor; web-scraping; reviews

## 1. Introduction

Some of the most important elements in a tourist deciding their choice of destination, timing, and length of vacation are the destination's weather and climate (Zeiss et al., 2022). This in turn will have an impact on whether they are likely to visit the location again or suggest it to other tourists (Fitchett et al., 2020). For most tourist destinations, favourable weather results in an increase in the number of tourists (Toubes et al., 2020). Tourist climatic preferences may vary depending on the type of tourist activities taking place, the tourist destination, and the tourists' place of origin (Velea et al., 2022). Research that was conducted in Southern Europe (Georgopoulou et al., 2019) and the Caribbean Islands (Rutty & Scott, 2013) indicated that tourists that were mainly from tropical regions favoured moderate to hot temperatures, whereas those from temperate

**CONTACT** Jennifer M. Fitchett jennifer.fitchett@wits.ac.za

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

regions preferred cooler conditions. Climate plays a significant role in determining tourism seasonality as a whole while also affecting tourism patterns globally (McKercher et al., 2024). This creates both push and pull factors (Doran & Schofield, 2023). The impact of weather events on tourists may result in dissatisfaction in the tourists' overall vacation experience (Caldeira & Kastenholz, 2018). Seasonal patterns define the spatial and temporal distribution of tourists year-round, demonstrating the sensitivity of the tourism industry to climatic fluctuations (Doran & Schofield, 2023), as these weather-related events directly influence tourists' destination decisions. The impact of climatic conditions varies considerably based on the different types of tourist activities conducted upon visiting the destination; outdoor activities are often more susceptible to climatic fluctuations than those indoors (Tanrisever et al., 2024). The perception of the credibility of sources of climate information, whether by the public or media entities, also influences the decision of tourists and travellers (Dong et al., 2018).

There are various challenges to measuring tourists' sensitivity to climate. One of these is that results from studies on temperature thresholds for tourism that have been identified for a specific cohort of tourists at a destination for a given period of time cannot be easily applied to other destinations, populations, and environments (Zare et al., 2018). However, internationally applicable indices and thresholds that use common methodologies and representative samples across different places and people are needed to achieve generalisable findings and to compare destinations (Dubois et al., 2016). Directly enquiring about climate through questionnaires and interviews can heighten people's awareness of climate, artificially enhancing measures of climatic sensitivity (Barimah et al., 2015; Fitchett & Hoogendoorn, 2019; Munro, 2020). It is also essential to consider multiple factors, such as temperature, humidity, precipitation, and wind, to accurately evaluate climate preferences (Ngxongo, 2021). These indices include indices, such as the Tourism Climate Indices and new-generation indices derived, such as the Beach Climate Index (BCI; Morgan et al., 2000), Holiday Climate Index (HCI) (Scott et al., 2016), Holiday Climate Index Beach (HCI:Beach) (Rutty et al., 2020), Holiday Climate Index Urban (HCI:Urban) (Scott et al., 2016), Climate index for tourism (De Freitas et al., 2008), are helpful, especially if they are customised for various forms of tourism (Demiroglu et al., 2021).

Tourists are more likely to contemplate the impact of climate on their travel experiences when prompted to consider it. A post-facto analysis of open-ended, unsolicited reviews posted on TripAdvisor can circumvent this, and provide a more robust measure of the climatic sensitivity of tourists (Fitchett et al., 2020; Fitchett & Hoogendoorn, 2018, 2019; Stockigt et al., 2018). TripAdvisor reviews may offer more objective responses, as reviewers may reflect on and discuss any aspects of their trip (Fitchett et al., 2020). TripAdvisor provides an online forum for travellers to share feedback on their experiences, which has resulted in the spread of e-Word-of-Mouth recommendations to a broader audience (Choudhary & Sharma, 2022). Although there has been a recent increase in researchers studying TripAdvisor reviews to gain insights into tourists' perspectives on various topics (Baleiro, 2023; Hong, 2020; Sangkaew & Zhu, 2022), there is a lack of research on reviewers' experiences with weather during their trips (Fitchett et al., 2020). Studies of the sensitivity of tourists to weather and climate have utilised TripAdvisor in various ways (Fitchett et al., 2020; Fitchett & Hoogendoorn, 2018, 2019; Stockigt et al., 2018). This is beneficial in the formulation of targeted adaptation plans

that primarily address tourist concerns. All of the aforementioned studies, however, obtained TripAdvisor data manually, which is labour intensive, resulting in comparatively low total datasets, and may be susceptible to errors and human biases. web-scraping offers the potential to upscale these types of analyses to hundreds of thousands of reviews, and to improve accuracy and reduce bias (Mitchell, 2024; Tjaden, 2023). This study seeks to compare the results achieved through both methodologies mentioned in analysing TripAdvisor tourist reviews over 19 distinct tourist destinations in South Africa.

## 2. Materials and methods

### 2.1. Study site

The Republic of South Africa has a total land area of 1,219,602 km$^2$, and spans 22–35°S, 17–33°E (Figure 1). South Africa is influenced by the cold Atlantic Ocean and Benguela current to the west and the warm Indian Ocean and Agulhas current to the east, with the average annual rainfall decreasing from east to west (Jury, 2013). A significant proportion of the country receives less than 500 mm of precipitation per year, which classifies it as semi-arid (Tyson & Preston-Whyte, 2000). The South African climate is varied, with summer temperatures frequently surpassing 32°C and winter temperatures dropping below freezing on the interior plateau (Fitchett et al., 2017; Jawtusch, 2014). Nevertheless,

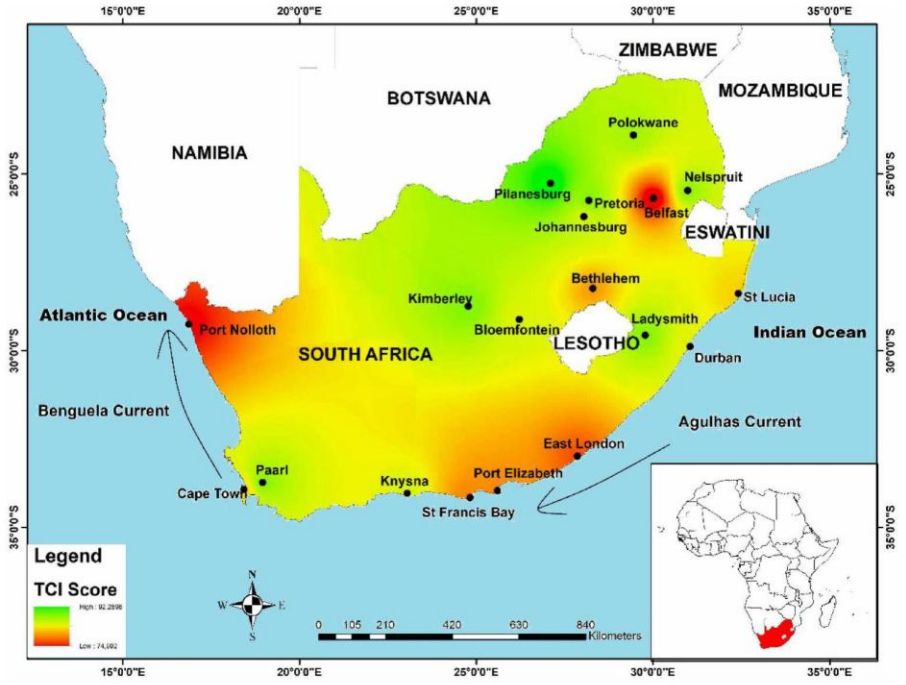



**Figure 1.** Map demonstrating the 19 different locations in South Africa where the study is based, with inverse distance weighting performed on TCI scores extracted from Fitchett et al. (2017).

the climate is generally considered ideal for tourism and a significant attraction (Fitchett et al., 2017; Preston-Whyte & Watson, 2005).

To facilitate comparison of the previously published manually analysed TripAdvisor reviews with those from web-scraping, this study will concentrate on the same 19 locations in South Africa for which a manual TripAdvisor analysis of climate mentions has been conducted, namely: Polokwane, Pilanesberg National Park, Pretoria, Belfast, Nelspruit, Johannesburg, Kimberly, Bethlehem, Ladysmith, St. Lucia, St. Francis Bay, Durban, Bloemfontein, Paarl, Cape Town, Knysna, Port Elizabeth, East London, and Port Nolloth (Fitchett & Hoogendoorn, 2018, 2019; Figure 1). These locations span the nine provinces of South Africa, the three rainfall zones, and offer a range of different tourist attractions (Table 1). The South African tourist sector consists of various forms of tourism markets such as sports, wine, cultural, business, adventure, coastal and nature-based tourism. Each of these tourist destinations or hotels under study have unique climates and weather associated with them which make them ideal for evaluating the effect of weather and climate on tourism as a whole (Fitchett et al., 2017). The diversity and type of tourism offerings at each region is specific and diverse. Cities such as Cape Town, Johannesburg and Durban serve as attraction sites for business tourism as well as providing various tourist activities such as museum tours and botanical gardens. Hence all these regions have tourist attractions that are sensitive to weather and

**Table 1.** Tourist segments and the sites of this study.

| Tourist segments | Tourist locations |
|---|---|
| Coastal | Port Elizabeth |
| | Port Nolloth |
| | St Francis Bay |
| | Knysna |
| | Durban |
| | East London |
| | Cape Town |
| Urban | Johannesburg |
| | Pretoria |
| | Cape Town |
| | Durban |
| | Kimberley |
| | Bloemfontein |
| Rural and Countryside | Paarl |
| | Polokwane |
| | Ladysmith |
| | East London |
| | Knysna |
| | St Francis Bay |
| Mountain | Bethlehem |
| | Cape Town |
| Natural and Wildlife | Pilanesburg National Park |
| | St Lucia |
| | Knysna |
| Cultural and Historical | Nelspruit |
| | Johannesburg |
| | Cape Town |
| | Bethlehem |
| Adventure and Sports | Belfast |
| | Knysna |
| | Cape Town |
| | Johannesburg |

climate. The tourist experiences vary in relation to weather, throughout each location. Climate and weather are of importance within tourist attractions, as there is an inter dependence that exists between climate and tourism (Kozak et al., 2008; Moreno & Amelung, 2009; Rutty & Scott, 2016).

### 2.2. Research methodology

This study applies Fitchett and Hoogendoorn's (2018, 2019) methodology for analysing visitor sensitivity to climate using TripAdvisor reviews to the same 19 tourist destinations in South Africa, using a web-scraping approach. The main purpose of this paper is to provide a comparison between the manually captured data and web scrapped derived data. In doing this, all statistical tests conducted in the previous studies will be replicated which will include the following: calculating and stating the overall number of reviews, the proportion of climatic/weather-related mentions, exploring the relative climatic contribution of the climatic factors per location and then utilising principal component analysis to explore the co-variability of the two datasets. In addition to this, a comparison of means test will be conducted for each other components to determine if there is a significant difference between web scraped and manually collated data.

#### 2.2.1. Data collection

For this study, the same criteria pertaining to the type and amount of accommodation establishments utilised in the (Fitchett & Hoogendoorn, 2018, 2019) papers were also used. Reviews from a minimum of three accommodation establishments per location were consulted, including a four/five-star hotel, a guesthouse or BnB, and a one-star motel-type resort in accordance with the methodology provided in the Fitchett and Hoogendoorn's (2018, 2019) research. Reviews were captured for this research as well, working backwards in time, year by year across the three types of accommodation from December 2022 (Amaral et al., 2014; Buzinde et al., 2010; Fitchett et al., 2020). This yielded a bigger and more detailed population sample. The data was obtained through web-scraping, which involved developing a unique code to automate data extraction from the TripAdvisor website. To accomplish this, a Python algorithm was employed to automatically extract and manage vast amounts of data from the website. The Python scripting language has a web-scraping library known as "Selenium" that can extract particular details, such as structured data from the HTML/XML content. In extracting the data, the code was specifically designed to extract the review date, title, location, rating, and content to determine the basic demographics pertaining to the sample. The data was then saved in a excel csv format and was cleaned for further analysis.

#### 2.2.2. Data analysis

To be considered a climate mention, text needed to contain information about weather or climatic conditions, such as temperature, rainfall, wind, humidity, and cloudy/sunny circumstances. Climate references were categorised by the exact weather or climatic state, such as heat, humidity, rain, wind, and storm, to determine the most relevant factors to visitors. The web-scraped data was analysed using the excel "if function" to identify reviews that contained mentions of climate or weather. Reviews were

analysed for recurring themes and predetermined terms using a content-based methodology such that if a review had a specified climate/weather mention the function would assign a value to it and then calculate the total amount of the mentions. The proportion of reviews mentioning climatic factors was calculated, and the frequency distribution of climatic conditions mentioned was examined for each accommodation type and destination indicating the overall climatic sensitivity, the proportion of climatic mentions and the total climatic mentions per region. Principal Component Analysis (PCA) was then used to explore the spatial clustering of destinations based on climate remarks.

Finally, the statistics obtained were compared with the manually collected data presented in Fitchett and Hoogendoorn's (2019) research for each destination. The statistical tests included comparison of means tests and t-tests. These tests were conducted based on the overall demographics and statistics of the reviewers, the proportion of climatic mentions, the overall climatic sensitivity, total climatic mentions per region to validate the results obtained from the web scraped relative to the results obtained from the manually collected data from the Fitchett and Hoogendoorn (2019) paper. Subsequent consideration was undertaken over the accuracy of manually acquired data in the classification of the climatic appropriateness of each place for tourism.

## 3. Results

### *3.1. Sample size and demographics*

web-scraping extracted 26,687 TripAdvisor reviews of 168 tourist accommodations in the 19 destinations in South Africa, which was 352.48% more than the 5898 manually collected reviews (Table 2). The highest amount of reviews recorded was from the United Kingdom (7.6%) which resulted in a 2.22% difference from the manually recorded data (Table 2). This was followed by the United States which yielded a 3.86% difference between both approaches (Table 2). The remainder of the assessments as (displayed in Figure 2(a)) included Australia which showed a 1.41% difference between both approaches. South African reviewers contributed the majority of the reviews, with over 45.81% written by locals (Figure 2(a)), particularly from Gauteng (21.22%). These statistics indicate that there is a statistically insignificant difference ($p = 0.94$) with those from the manually captured data which yielded 44.17% of South African reviewers. However, the location of 20.37% of the reviewers was not disclosed which was 5.9% more than the manually collected data. All these results indicate a statistically insignificant difference between both the approaches (Table 2).

**Table 2.** Results pertaining to the comparison between web scraped and manual reviews from various countries.

| Countries of Origin | Total Number | | Percentage (%) | | Percentage difference (%) |
|---|---|---|---|---|---|
| | W | M | W | M | |
| South Africa | 12,226 | 2,604 | 45.81 | 44.17 | 1.64 |
| United Kingdom | 2,015 | 579 | 7.6 | 9.82 | 2.22 |
| United States of America | 1,966 | 209 | 7.4 | 3.54 | 3.86 |
| Australia | 921 | 120 | 3.45 | 2.04 | 1.41 |

Notes: *T*-test for the comparison of country percentages. $p = 0.94$.

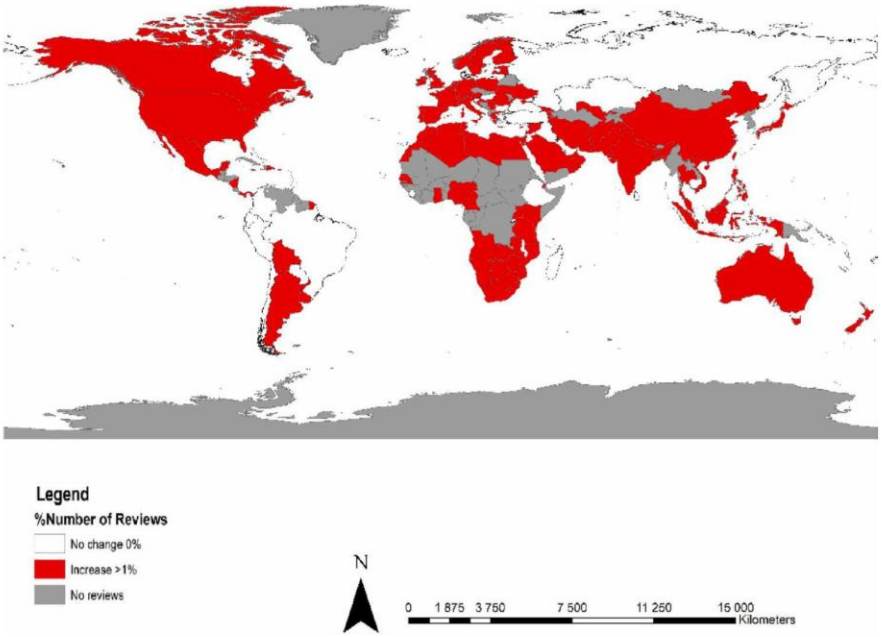



**Figure 2.** Map showing the percentage differences between the distribution of TripAdvisor reviews comparing the manual and web-scraping approaches.

The reviewers' countries of origin were documented and visualised (Figure 2). A 1.64% increase, which was statistically insignificant, was noted in the total number of South African tourists visiting tourist destinations from the manual approach (Figure 2), whereas a 2.22% insignificant difference was also observed in the number of tourists coming from the United Kingdom between the web scraped and manual approaches (Figure 2). A statistically insignificant 1.41% increase was also denoted between both approaches in the reviewers coming Australia (Figure 2). More countries such as Argentina, Canada, France, Germany, Spain, China and India have increased by more than 1% from the manual results. A total of 11.28% of the countries had no reviews which was constant from the manual approach. Countries that showed no change in the number of reviews obtained resulted in 26.15% (Figure 2).

### *3.2. Climatic sensitivity of tourists*

Among the 26, 687 TripAdvisor reviews, 1,163 included climate-related references, equating to an average sensitivity to the weather of 4.4% (Table 2). This is statistically different ($p = 0.014$) from the manually collected data yielding an overall sensitivity of 7.9% (Table 3). Of the 1,163 mentions of weather in the web scraped TripAdvisor reviews, cold conditions were mentioned the most (35.68%, $n = 1,163$), followed by rain (14.70%, $n = 1,163$), which differed to the findings obtained by manually gathering the data where the cold (38.8%, $n = 494$) was mentioned the most followed by hot (28.88%, $n = 494$; Table 3).

**Table 3.** Mentions of the weather in TripAdvisor reviews.

| Tourist Destination | Web scrapped and manual data | | | | | | | | | | *P*-value (Proportion of climate mentions) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of Reviews Consulted | | Number of Mentions of Climate | | Percentage of Climate Mentions | | Most Frequently Mentioned Climate Factor | | Proportion within Climate Mentions | | |
| | W | M | W | M | W | M | W | M | W | M | |
| Belfast | 326 | 172 | 36 | 24 | 11% | 14% | Cold | Cold | 0.67 | 0.83 | *0.55* |
| Bethlehem | 1,021 | 171 | 73 | 26 | 7% | 15% | Cold | Cold | 0.42 | 0.85 | *0.63* |
| Bloemfontein | 1,624 | 470 | 71 | 26 | 4% | 6% | Cold | Cold | 0.45 | 0.62 | *0.36* |
| Cape Town | 1,448 | 604 | 58 | 33 | 4% | 6% | Hot | Hot | 0.19 | 0.36 | *0.55* |
| Durban | 1,260 | 304 | 39 | 25 | 3% | 8% | Cold | Sun | 0.54 | 0.24 | *0.44* |
| East London | 3,777 | 462 | 92 | 25 | 2% | 5% | Cold | Hot | 0.25 | 0.36 | *0.02* |
| Johannesburg | 1,731 | 345 | 52 | 23 | 3% | 7% | Hot | Hot | 0.31 | 0.35 | *0.898* |
| Kimberly | 808 | 360 | 31 | 29 | 4% | 8% | Cold | Hot | 0.45 | 0.59 | *0.93* |
| Knysna | 2,853 | 603 | 141 | 32 | 5% | 5% | Cold | Cold | 0.35 | 0.41 | *0.01* |
| Ladysmith | 883 | 179 | 57 | 25 | 6% | 4% | Cold | Cold | 0.47 | 0.56 | *0.37* |
| Nelspruit | 1,817 | 251 | 43 | 25 | 3% | 10% | Cold | Cold | 0.51 | 0.40 | *0.1* |
| Paarl | 1,285 | 202 | 65 | 25 | 5% | 12% | Cold | Hot | 0.26 | 0.36 | *0.16* |
| Pilanesberg | 1,749 | 327 | 118 | 28 | 7% | 9% | Cold | Cold | 0.42 | 0.43 | *0.10* |
| Polokwane | 733 | 249 | 22 | 16 | 3% | 6% | Hot | Hot | 0.41 | 0.63 | *0.02* |
| Port Elizabeth | 1,279 | 423 | 46 | 20 | 4% | 5% | Cold | Cold | 0.43 | 0.55 | *0.304* |
| Port Nolloth | 149 | 47 | 21 | 12 | 14% | 26% | Hot | Cold | 0.48 | 0.50 | *0.56* |
| Pretoria | 830 | 299 | 56 | 25 | 7% | 8% | Cold | Hot | 0.43 | 0.52 | *0.76* |
| St Lucia | 2,353 | 273 | 70 | 25 | 3% | 9% | Rain | Hot, rain | 0.26 | 0.32 | *0.53* |
| St Francis | 761 | 157 | 67 | 20 | 9% | 13% | Cold | Cold | 0.33 | 0.50 | *0.17* |
| Total | **26,678** | 5,898 | **1,163** | 464 | **4.4%** | 7.9% | **Cold** | Cold | **0.37** | 0.39 | |

Note: Shading represents statistical significance at $p < 0.05$.

On comparison, cold climatic mentions between both methods yielded a *p*-value of 0.72 which indicated statistical insignificance, whereas hot climatic mentions yielded a *p*-value of 0.05 indicating statistical insignificance between the two approaches. This indicates that visitors are proportionately more sensitive to thermal comfort (Figure 3).

Based on the observations made in Figure 3, it can be inferred that despite the statistically insignificant variations in the frequency of mentions of hot between the manual and web scraped methods, the indicators of thermal comfort remain the most referenced climatic factors in the review. This suggests an increased sensitivity to thermal comfort. After conducting a comparison of means test on the datasets as shown in Table 4, it was found that the total climatic sensitivity exhibited a statistically significant percentage difference of 3.56 ($p = 0.033$). The observed difference of 2.16% in the proportion of mentions related to "Cold" between the data obtained by online scraping and manual collection did not demonstrate statistical significance ($p = 0.72$, Table 4). The hot climatic references observed a drop of 8.26% based on the manual data, which was determined to be statistically insignificant at a significance level of $p = 0.05$. In comparing sun climatic references, it was seen that there was a minimal difference of 0.68% between the web scraped and manual procedures. This difference was determined to be statistically insignificant, with a *p*-value of 0.45 (see Table 4). There was no evidence of statistical significance ($p = 0.68$) noticed in the comparison of rain-related mentions between the manual and web scraped procedures, with just a marginal difference of 0.71% observed (see Table 4 and Figure 3). Both good and bad weather conditions exhibited *p*-values of 0.05, indicating statistical insignificance. Furthermore, there was a notable fluctuation of over 2% between the two techniques. The changes in drought, wind, and humid climatic conditions were all found to be below 1%, rendering them statistically inconsequential ($p = 0.80$, $p = 0.78$, $p = 0.76$). There was a lack of statistical significance ($p = 0.06$) seen in the comparison of cloud-related mentions between the manual and web scraped methodologies (see Table 4). A marginal variation of 1.12%, which was determined to be statistically insignificant ($p = 0.78$), was noticed in the frequency of references associated to mist (Table 4; Figure 3).

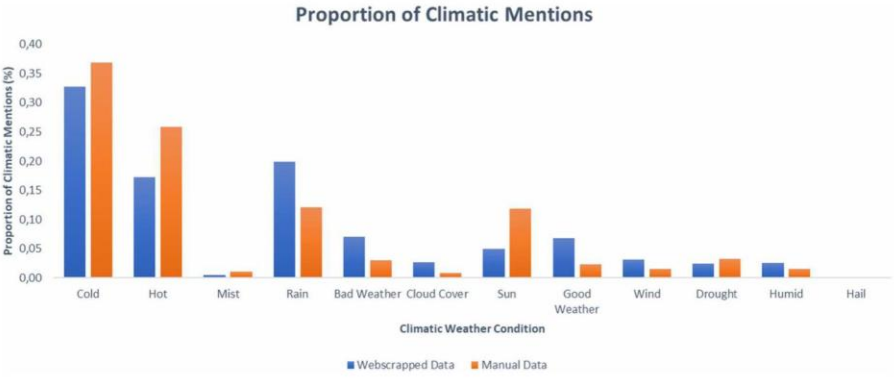



**Figure 3.** Proportion of climate references in the aggregated TripAdvisor reviews for 19 South African locations.

**Table 4.** Statistical analysis performed between the various climatic mentions between both datasets.

| | Manual % | Web scrapped. % | Percentage difference (%) | output-value |
|---|---|---|---|---|
| Overall climatic sensitivity | 7.9 | 4.4 | 3.5 | *0.014* |
| **Climatic condition** | | | | |
| Hot | 26.60 | 18.34 | 8.26 | *0.05* |
| Cold | 39.70 | 37.54 | 2.16 | *0.72* |
| Rain | 12.43 | 13.14 | 0.71 | *0.68* |
| Bad weather | 2.87 | 5.89 | 3.02 | *0.05* |
| Good weather | 2.37 | 4.92 | 2.55 | *0.05* |
| Mist | 1.28 | 0.16 | 1.12 | *0.78* |
| Sun | 11.57 | 10.89 | 0.68 | *0.45* |
| Humid | 2.06 | 1.17 | 0.89 | 0.76 |
| Drought | 2.9 | 2.47 | 0.43 | *0.80* |
| Cloud | 0.8 | 2.93 | 2.13 | *0.06* |
| Wind | 2.35 | 2.77 | 0.42 | *0.78* |

Note: Shading represents statistical significance at $p < 0.05$.

Statistically significant variations were noted in East London ($p = 0.02$), Knysna ($p = 0.03$) and Polokwane ($p = 0.02$) between both manual and web scraped data. No statistical variations with regards to the other nine climatic references in the 17 destinations were observed between regions in both web-scraped and manually collected data, as evidenced by Figure 4 and Table 3. The web scraped data reveals that Port Nolloth is the region most frequently associated with climatic references, which is also directly related to the findings obtained from the manually collected data. Port Nolloth appears to have the highest number of recorded climatic references, which supports the findings obtained with the web scraped data. Belfast has the second highest recorded climatic mentions for both methods (Figure 3). Figure 4 and Table 3 both indicate that Port Nolloth received fewer reviews overall for both datasets, but a significant proportion of these reviews included climatic references. Knysna had the most TripAdvisor reviews for both datasets but had the lowest recorded climatic mentions for both datasets (Figure 4 and Table 3). Knysna had a 1% difference resulting from both approaches as depicted in Figure 4 and Table 3.

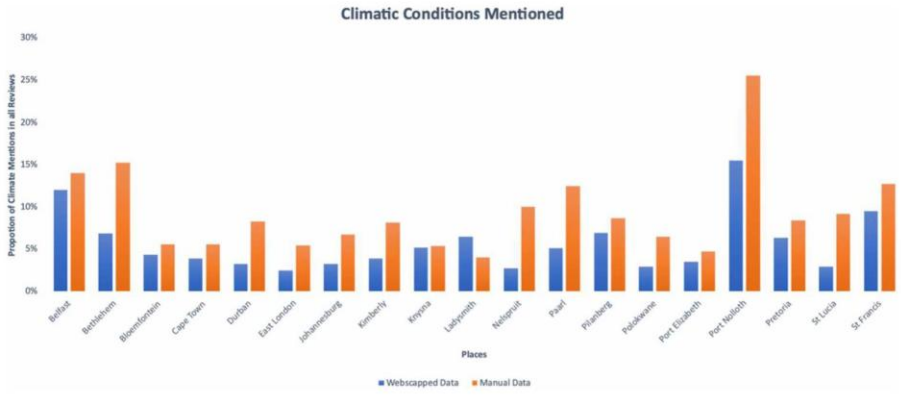



**Figure 4.** The proportion of TripAdvisor reviews that mention weather conditions, subdivided by location.

Figure 5 demonstrates a reduced degree of statistical variability across the datasets across all different climatic references, in a comprehensive manner. As stated before, there was no statistically significant difference seen in climate references when comparing the results obtained from web-scraping and manual data collection methods. However, according to the data shown in Figure 5, it can be seen that 16 tourist locations exhibited climate variables that did not demonstrate statistically significant differences. The weather

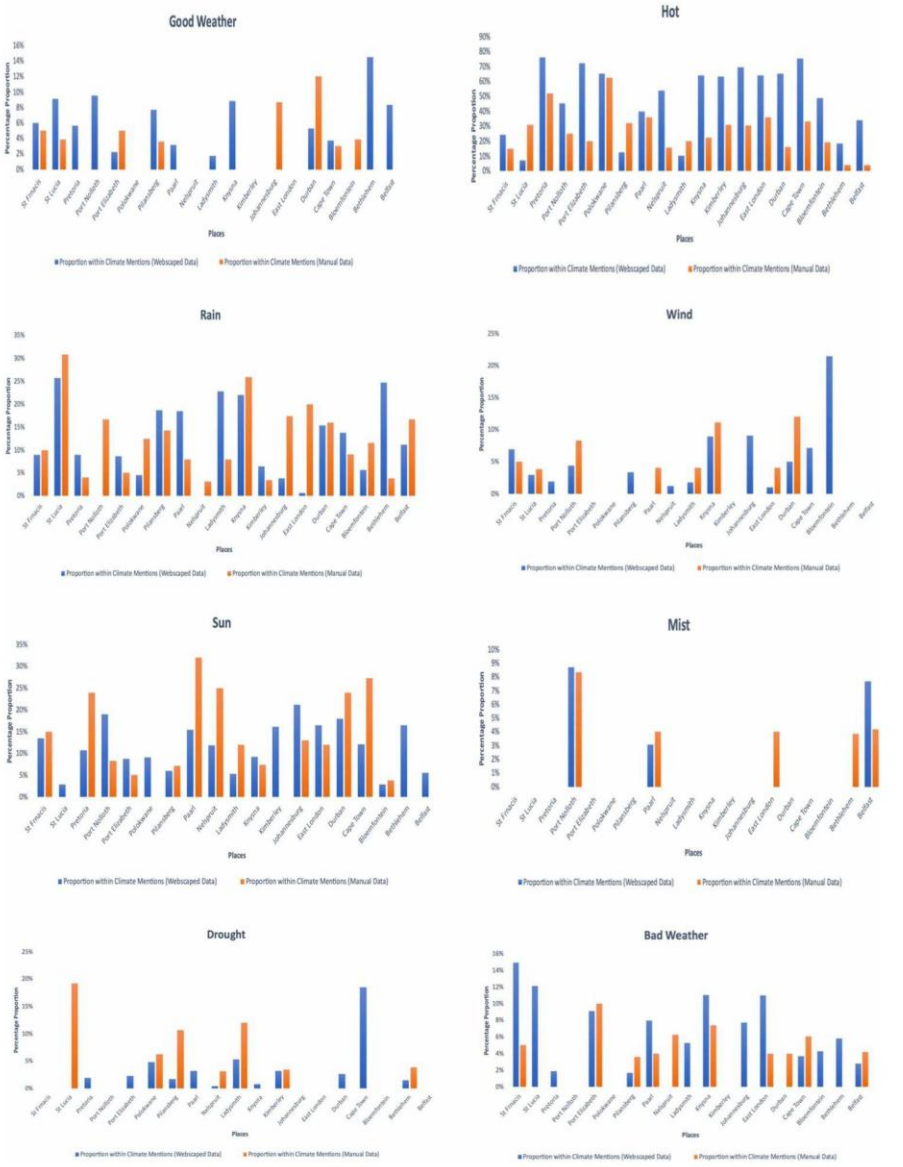



**Figure 5.** Location-specific proportions of each climate component.

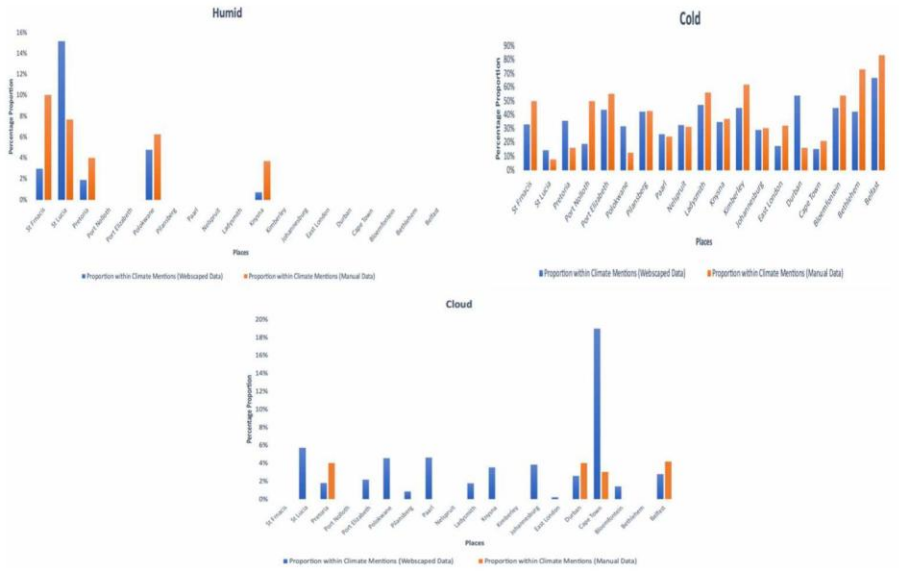



**Figure 5** *Continued*

conditions in St Francis exhibited slight variations of 0.56%, whereas Bethlehem experienced large changes of 14.49% and Johannesburg saw a reduction of 8.70% in the web scraped data compared to the manually obtained data. However, it is important to note that these differences were not statistically significant. The analysis of meteorological data in Port Nolloth revealed a substantial drop of 32.42% when comparing web scraped data to manual data. Conversely, in Nelspruit, a marginal rise of 1.43% was seen in the web scraped data. However, it is important to note that these changes were not statistically significant. The web-based scraped data revealed an 8.89% increase in bad weather occurrences in Nelspruit compared to the manual data, as well as a 6.25% reduction relative to the manual data. All of these changes remained statistically insignificant. The statistical analysis reveals that the percentage variances in drought occurrence vary from 17.86% in Bethlehem to −19.23% in St Lucia, indicating that these variations are statistically inconsequential. The climatic conditions in Nelspruit exhibited the most pronounced statistically significant disparity, with a percentage difference of 38.31%. Conversely, Pretoria had a percentage drop of 25.08%. The frequency of mist climatic references was found to be the lowest in both approaches, exhibiting a difference of 3.52% in Belfast and with a slight variation of 0.37% in Port Nolloth. The region of Bethlehem had the highest percentage difference at 3.85%; yet it is important to note that all of these variances were deemed statistically insignificant. The wind-related conditions in Bloemfontein exhibited statistically insignificant variations of 21.43%, whereas in Paarl, there was a percentage drop of 4%. Statistically insignificant percentage declines were seen in humid circumstances, particularly at St Lucia (7.02%) and St Francis (−7.22%), using both web scraped and manual procedures. The city of Bethlehem had a marginal and statistically insignificant rise of 22.24% in the frequency of rain-related references. Conversely, the town of Port Nolloth witnessed a marginal and statistically insignificant decline of 16.67% in web scraped data.

Statistical significance in climatic mentions was seen in three areas, namely East London, Knysna, and Polokwane, across both methods. The observed percentage differences for hot climatic remarks varied from a positive value of 20.38% to a negative value of −19.64%. The region of Knysna exhibited variations in good weather conditions, with a range spanning from a decrease of 16.06% to an increase of 8.22%. The region of East London had a significant reduction of 19.40% in the frequency of comments connected to rainfall, while there was a notable rise of 28.13% in references to hot weather. Overall, these locations exhibited statistically significant disparities.

## 4. Discussion

The process of web-scraping enables the collection of vast quantities of data in a short amount of time (Simeon et al., 2017). The use of automated methods for collecting and analysing an extensive amount of tourist reviews is advantageous, particularly in instances in which manual collection and analysis of each review would be time-consuming and inefficient (Vu et al., 2019). Furthermore, web-scraping offers a more comprehensive and objective perspective on user ratings. In addition, the use of web-scraping enables a more comprehensive examination of data (Khder, 2021). Conventional techniques for data collection, such as surveys and questionnaires, are frequently constrained by the limited number of participants who can be accessed, in addition to the time and resource constraints associated with data collection and analysis (Vu et al., 2018). web-scraping, on the contrary, enables the simultaneous accumulation of data from multiple Internet platforms, thereby facilitating a more comprehensive examination of tourist ratings (Mitchell, 2024).

Barbera et al. (2023) highlight the potential of web data scraping tools to capture information from TripAdvisor's social pages, enabling businesses to compare the performance of competitors and obtain new insights into customer preferences. Tian et al. (2016) propose a web-scraping tool that is particularly designed to capture data from TripAdvisor, with a concentration on actual customer reviews and hotel sentiment. Johnson et al. (2012) discuss the use of automated web harvesting to extract review data from TripAdvisor, shedding light on the advantages and disadvantages of utilising user-generated content for tourism research. Research conducted by Saurkar et al. (2018) provides an overview of web-scraping techniques and tools, emphasising their significance in the collection of structured data from unstructured web content. The above-mentioned case studies all demonstrate the utility of web-scraping in acquiring TripAdvisor data and its potential to improve tourism industry decision-making and analysis.

Through the utilisation of web-scraping, the study obtained 352.48% more reviews than the manually collected data. TripAdvisor has documented 1,059,165 million hotel reviews all over the world (TripAdvisor, 2013). According to TripAdvisor's fact sheet, as of 2013, there were more than 125 million reviews and opinions available on the site (Chua & Banerjee, 2013; TripAdvisor, 2013). Through exploring the percentage differences in climatic mentions overall and per location, slight variations were evident ranging from cold climatic mentions being referenced the most (2.16%, $p = 0.72$) to the least mentioned climatic reference being mist (1.12%, $p = 0.78$). However, these variations were small and were mostly statistically insignificant. It was found that all the percentage differences

between the climatic references in both approaches, did not show any statistical significance when comparing the results obtained. Overall, this means it would be reasonable to use web-scraping in the place of a manual approach. Three out 19 destinations that had climatic references had insignificant differences between both approaches, namely, East London ($p = 0.02$), Knysna ($p = 0.01$) and Polokwane ($p = 0.02$). This could be attributable to differences in interpreting what constitutes "good" or "bad" weather in manual data as opposed to web scraped data, which automates words as they are. In contrast to web-scraping, which generally entails the extraction of references through predetermined patterns or keywords, manual data collection affords the reader the ability to grasp the significance of the reference within its contextual framework (Dallmeier, 2021). The comprehension of a text may be subject to the reader's personal interpretation, experience, and level of knowledge (Saarinen et al., 2022) for example in this study the word humidifier utilised by a reviewer complaining that the room needed one, could mean that the reviewer was indirectly stating that the conditions in the room were humid. In web-scraping however, that is not the case, as the code is designed to automate data according to what is documented in the code. Another example included the misclassification of the words such as "hotel" for "hot" and "Sunday" for "sun" within the web scraped data (Table 5). Whereas within the manually collected data, such misclassifications didn't exist.

There were no significant differences in the comparison of climatic mentions per region between the datasets, indicating the effectiveness of web-scraping in this regard (Table 5). It was noted that "cold" climatic references were mentioned the most throughout both approaches. Port Nolloth had the highest climatic mentions for both approaches. Knysna had the most TripAdvisor reviews but had the lowest climatic references for both approaches. This further indicates that web-scraping can be trusted as a good data collection source as the results indicate similarities between

**Table 5.** Misclassified words in web-scraping TripAdvisor reviews.

| Misclassified word | Example | Reason attributed to the misclassification | Possible solutions to this |
|---|---|---|---|
| Cold | - Cool temperatures or an unkind welcome | Over-counting in web-scraping | Integration with manually checking |
| Warm | - Warm temperature, or friendly welcome. | over-recording in web scraped data | Coupled words that could be utilised to signify warm conditions. |
| Sun | - Sunlight or warmth (temperature) or even form part of a name of a hotel | Over-counting in web-scraping | Coupled words that could be utilised to signify sun or sunshine. |
| Wind | - To rest or calm down.<br>- Name of a hotel<br>- Indicating a form of energy (wind turbines) | Over-counting in web-scraping | Integration with manually checking |
| Humid | - Humidifier<br>- Wet/Damp | Over-counting in web-scraping | Integration with manually checking |
| Mist | - Wet or moist conditions<br>- Name of a hotel e.g. Misty River | Over-counting in web-scraping | Integration with manually checking |
| Hot | - Hot temperatures or a "**hot**el" | Over-counting in web-scraping | Integration with manually checking |
| Rain | - Rain as in precipitation<br>- Bathroom shower head classification | Over-counting in web-scraping | Integration with manually checking |

bot manual and web scraped methods. In a study that was conducted by Karthikeyan et al. (2022), it was found that through the use of artificial intelligence and machine learning, the automation of data from TripAdvisor is possible, more efficient, and less tedious, as the analysis of a large proportion of reviews manually obtained (often over 5000) can be prone to human error (Landers et al., 2016). These tools have been applied to climate change research, particularly on Twitter, and exciting developments have been made in sentiment analysis using code to classify emotions (Al-Saqaf & Berglez, 2019; Cecinati et al., 2019; Sailunaz & Alhajj, 2019). A proportion of climatic mentions didn't yield any results among both approaches as well. According to Karthikeyan et al. (2019) web-scraping can be classified as a very fast and effective data collection and moreover can cover a large number of websites within a short time period. With web-scraping, researchers can extract data in a structured format, making it easy to analyse and compare (Karthikeyan et al., 2022).

It is noteworthy that manual data collection and web-scraping techniques possess inherent limitations. The process of manually collecting data can be susceptible to human error and can also be a time-consuming task (Birnbaum, 2004; Landers et al., 2016). On the other hand, web-scraping techniques may not possess a comprehensive understanding of the context and can be influenced by technical limitations or inaccuracies during the scraping process (Dallmeier, 2021). This phenomenon was further shown in the research, as some phrases were interpreted as climatic references while lacking such connotations. When data is collected manually, there is a risk of introducing biases or variations in the identification, recording, or categorisation of mentions due to the subjectivity of the reader (Birnbaum, 2004; Landers et al., 2016). Diverse readers may construe identical references in varying ways, which could result in incongruities in the compiled data. Conversely, web-scraping can offer a methodical and uniform methodology for acquiring references, albeit it may be deficient in comprehending the subtle connotations of the reference within its contextual framework. web-scraping algorithms commonly depend on predetermined patterns or keywords, which can lead to the inclusion of irrelevant mentions or the omission of mentions that necessitate contextual comprehension (Egger et al., 2022).

Integrating manual data collection with web-scraping techniques can prove advantageous in addressing the constraints of both methodologies. The process of manually collecting data can yield a detailed and context-specific analysis of references, whereas web-scraping can effectively a mass a substantial amount of data (Khder, 2021). Through the integration of these methodologies and the utilisation of suitable techniques for data validation and verification, a dataset that is more comprehensive and dependable can be acquired (Lauer et al., 2018). On the other hand, the integration of climate data sourced from hotel reviews with thermal comfort indices can provide an enhanced understanding of the impact of climatic conditions on the comfort experiences of tourists. This approach employs an integrated methodology that incorporates both subjective and objective data to evaluate climate sensitivity, thereby offering significant implications for enhancing comfort levels in hotel settings. Based on the above-mentioned information, manual comparisons can be done for accuracy purposes in order to identify the various language components. As this was already conducted throughout South Africa, the study can therefore be expanded throughout the whole of South Africa through a web scraped approach.

In this study, the web-scraping demonstrated the effectiveness of automating data from TripAdvisor to further analyse the climate sensitivity of tourists and validate the findings with the manual approach. The limitations of this however included, the time employed in the development of the code to effectively automate the data from the website as well as the misclassification of weather-related themes within the analyses. It is however depicted that web-scraping can be accepted method of data collection just as the manual approach. Additionally, changes to the TripAdvisor website layout or algorithms can impact the accuracy of the extracted data (Egger et al., 2022). However, when coming to analysing the perception of tourists regarding weather, both methods have their strengths and weaknesses. Tourist assessments, for example, might offer information on the influence of heatwaves or extreme cold spells on people's comfort levels, capacity to engage in outdoor activities, and overall trip satisfaction (Liu et al., 2021). Furthermore, studying tourist assessments might aid in the identification of situations or aspects that may lead to favourable or poor experiences under various weather and climatic conditions (Ngxongo, 2021). This information may guide urban and regional planners in designing and creating initiatives aimed at creating suitable infrastructure and providing good services in improving the overall well-being and satisfaction of tourists under various meteorological conditions (Abundabar & Pongpong, 2022). web-scraping may also be used to investigate the economic effects of weather and climate on the tourism industry. By evaluating visitor assessments, researchers may get insight into the influence of weather and climate on tourists' opinions and preferences (Fitchett & Hoogendoorn, 2019).

Tourism operations and individuals must contend with the weather as it occurs, yet there is limited empirical evidence on how weather influences tourist perceptions and experiences. This research gap highlights the need for a structured approach to understanding weather's effects on tourists. TripAdvisor reviews offer more unbiased insights, as they reflect tourists' experiences without prompting specific inquiries about weather. Exploring tourist sensitivity across South Africa using TripAdvisor reviews through a non-labour-intensive, unbiased platform is vital.

South Africa is a country that is highly diverse in terms of climate and weather with temperature variations throughout the region (Fitchett & Hoogendoorn, 2019). The most frequently cited meteorological conditions in this research were cold, hot, rain and sun, which links to previous research that was conducted (Fitchett & Hoogendoorn, 2018; Fitchett & Hoogendoorn, 2019). The most common meteorological variable that tourists in South Africa are most sensitive to is temperature (Fitchett & Hoogendoorn, 2018). To illustrate, coastal regions in South Africa such as Durban are associated with warm temperatures which most tourists would classify as favourable weather conditions, indicating their level of thermal comfort. On the other hand, colder conditions reduce the level of thermal comfort of tourists in cooler regions, predominantly in winter. Some reviews from the web scraped data predominantly had negative sentiments with regards to cold weather conditions, especially in the case of outdoor located tourist destinations.

With regards to "Hot" climate references, it was found that web-scraped data had more climatic references than manual data in 16 of the 19 destinations which emphasises the sensitivity of tourists to thermal comfort. According to a study conducted by Dubois et al. (2016) in investigating the weather preferences of various French tourists, it was found that tourists were most sensitive to rainfall, which is seen as the most inconvenient factor, followed by cold weather and radiation fluxes. The study found an asymmetry

between sun and heat on one hand, and rain and cold on the other. The high tolerance to high temperature and the absence of concern over heatwaves contrast with a deep aversion to rain, with also strong effects on decisions (Dubois et al., 2016).

## 5. Conclusion

This study sought to conduct a comparative analysis of TripAdvisor reviews obtained by manual collection and web-scraping methods, with the aim of assessing the effectiveness of the web-scraping strategy in 19 different destinations throughout South Africa. A percentage difference of 3.5 was noted between both methodologies for the total amount of climatic mentions captured. A significant difference was evident in three of the 19 locations per climate mention (Knysna, Polokwane and East London), however no statistically significant differences were noted between both approaches for the overall climatic mentions. It is worth noting that web-scraping, being driven by code, may lack the nuanced understanding of text that a human reader possesses. Nevertheless, it was determined that web-scraping is effective in being less tedious, obtaining vast amounts of data in real-time and in offering a comprehensive perspective on the sensitivity of tourists towards climate and therefore can be accepted a good source of obtaining data.

Validation and verification of data retrieved through web-scraping is another area of prospective research in biometeorology. This is crucial for ensuring the precision and reliability of the obtained data. In addition, future biometeorology research should concentrate on standardising web scraped data. This involves defining formats and methods for organising and analysing obtained data in order to ensure reliability and comparability throughout research. Precisely determining a tourism industry's reliance on climate, susceptibility to climate change, and ability to adapt is crucial for effectively mitigating its impacts. Ensuring that the tourism sector within a region is sustainable requires an understanding of its dependence on climate and its capacity to adapt to climate change. To achieve this, it is vital to have precise information on a regions' current climate suitability for tourism, which can be evaluated using an empirical and consistent approach to compare it to other destinations in which web-scraping data may be of use.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

DRM has received funding for her doctoral studies from and the University of the Witwatersrand Postgraduate Merit Award and the Council for Scientific and Industrial Research (CSIR). JMF receives funding from the University of the Witwatersrand Friedel Sellschop Grant, and from GENUS.

## Research ethics

Ethics clearance was obtained from the Witwatersrand Human Non-Medical Research Ethics Committee, clearance number HRECNM22-01-077.

## References


Abundabar, A. P., & Pongpong, M. A. (2022). Assessing the local tourism infrastructure and services towards tourism development of San Jose. *International Journal of Applied Research in Tourism and Hospitality*, *1*(2), 97–116. https://doi.org/10.52352/ijarthy.v1i2.805

Al-Saqaf, W., & Berglez, P. (2019). How do social media users link different types of extreme events to climate change? A study of Twitter during 2008–2017. *Journal of Extreme Events*, *6*(2), 1950002. https://doi.org/10.1142/S2345737619500027

Amaral, F., TiAgo, T., & TiAgo, F. (2014). User-generated content: Tourists' profiles on TripAdvisor. *International Journal of Strategic Innovative Marketing*, *1*(3), 137–145.

Baleiro, R. (2023). Understanding visitors' experiences at Portuguese literary museums: An analysis of TripAdvisor reviews. *European Journal of Tourism Research*, *33*, 3305–3305. https://doi.org/10.54055/ejtr.v33i.2839

Barbera, G., Araujo, L., & Fernandes, S. (2023). The value of web data scraping: An application to TripAdvisor. *Big Data and Cognitive Computing*, *7*(3), 121. https://doi.org/10.3390/bdcc7030121

Barimah, P. T., Kwadwo, S. O., & David, O. (2015). Assessment of people's knowledge and perception on climate change: A case study of Asunafo North District, Ghana. *International Journal of Innovative Research in Science, Engineering and Technology*, *4*(1), 18417–18424. https://doi.org/10.15680/IJIRSET.2015.0401003

Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, *55*(1), 803–832. https://doi.org/10.1146/annurev.psych.55.090902.141601

Buzinde, C. N., Manuel-Navarrete, D., Kerstetter, D., & Redclift, M. (2010). Representations and adaptation to climate change. *Annals of Tourism Research*, *37*(3), 581–603. https://doi.org/10.1016/j.annals.2009.10.018

Caldeira, A. M., & Kastenholz, E. (2018). It's so hot: Predicting climate change effects on urban tourists' time–space experience. *Journal of Sustainable Tourism*, *26*(9), 1516–1542. https://doi.org/10.1080/09669582.2018.1478840

Cecinati, F., Matthews, T., Natarajan, S., McCullen, N., & Coley, D. (2019). Mining social media to identify heat waves. *International Journal of Environmental Research and Public Health*, *16*(5), 762. https://doi.org/10.3390/ijerph16050762

Choudhary, F. S., & Sharma, A. (2022). Exploring antecedents of electronic word-of-mouth in tourism: A case of TripAdvisor. In *Handbook on tourism and social media* (pp. 322–336). Edward Elgar Publishing.

Chua, A. Y., & Banerjee, S. (2013). Reliability of reviews on the internet: The case of TripAdvisor. In *World Congress on Engineering & Computer Science: International Conference on Internet and Multimedia Technologies* (pp. 453–457).

Dallmeier, E. C. (2021, May). Computer vision-based web scraping for internet forums. In *2021 7th International Conference on Optimization and Applications (ICOA)* (pp. 1–5). IEEE.

De Freitas, C. R., Scott, D., & McBoyle, G. (2008). A second generation climate index for tourism (CIT): Specification and verification. *International Journal of Biometeorology*, *52*(5), 399–407. https://doi.org/10.1007/s00484-007-0134-3

Demiroglu, O. C., Turp, M. T., Kurnaz, M. L., & Abegg, B. (2021). The Ski Climate Index (SCI): Fuzzification and a regional climate modeling application for Turkey. *International Journal of Biometeorology*, *65*(5), 763–777. https://doi.org/10.1007/s00484-020-01991-0

Dong, Y., Hu, S., & Zhu, J. (2018). From source credibility to risk perception: How and when climate information matters to action. *Resources, Conservation and Recycling*, *136*, 410–417. https://doi.org/10.1016/j.resconrec.2018.05.012

Doran, A., & Schofield, P. (2023). Tourism and the seasons. In P. Goulding (Ed.), *Tourism: A temporal analysis* (pp. 29–42). Goodfellow.

Dubois, G., Ceron, J. P., Gössling, S., & Hall, C. M. (2016). Weather preferences of French tourists: Lessons for climate change impact assessment. *Climatic Change*, *136*(2), 339–351. https://doi.org/10.1007/s10584-016-1620-6

Egger, R., Kroner, M., & Stöckl, A. (2022). Web scraping: Collecting and retrieving data from the web. In *Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications* (pp. 67–82). Springer International Publishing.

Fitchett, J. M., Fortune, S. M., & Hoogendoorn, G. (2020). Tourists' reviews of weather in five Indian Ocean islands. *Singapore Journal of Tropical Geography*, *41*(2), 171–189. https://doi.org/10.1111/sjtg.12318

Fitchett, J. M., & Hoogendoorn, G. (2018). An analysis of factors affecting tourists' accounts of weather in South Africa. *International Journal of Biometeorology*, *62*(12), 2161–2172. https://doi.org/10.1007/s00484-018-1617-0

Fitchett, J. M., & Hoogendoorn, G. (2019). Exploring the climate sensitivity of tourists to South Africa through TripAdvisor reviews. *South African Geographical Journal = Suid-Afrikaanse Geografiese Tydskrif*, *101*(1), 91–109. https://doi.org/10.1080/03736245.2018.1541022

Fitchett, J. M., Robinson, D., & Hoogendoorn, G. (2017). Climate suitability for tourism in South Africa. *Journal of Sustainable Tourism*, *25*(6), 851–867. https://doi.org/10.1080/09669582.2016.1251933

Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., Hontou, V., Gakis, N., & Lalas, D. P. (2019). Climatic preferences for beach tourism: An empirical study on Greek islands. *Theoretical and Applied Climatology*, *137*(1), 667–691. https://doi.org/10.1007/s00704-018-2612-4

Hong, J. E. (2020). Qualitative analysis of users' negative reviews on TripAdvisor: International tourists' reviews on Gyeongbokgung Palace in Seoul, Korea. *International Journal of Geospatial and Environmental Research*, *7*(1), 5.

Jawtusch, J. (2014). Climate change and disaster risk guide: Which climate and disaster risks affect South African and how do they affect South Africa.

Johnson, P. A., Sieber, R. E., Magnien, N., & Ariwi, J. (2012). Automated web harvesting to collect and analyse user-generated content for tourism. *Current Issues in Tourism*, *15*(3), 293–299. https://doi.org/10.1080/13683500.2011.555528

Jury, M. R. (2013). Climate trends in Southern Africa. *South African Journal of Science*, *109*(1), 1–11. https://doi.org/10.1590/sajs.2013/980E

Karthikeyan, P., Aishwariya Rani, V., Jeyavarshini, B., & Muthupriyaadharshini, M. N. (2022). Sentimental analysis and classification of restaurant reviews. In *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2022* (pp. 247–262). Springer Nature Singapore.

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, *13*(3), 145–168. https://doi.org/10.15849/IJASCA.211128.11

Kozak, M., Gokovali, U., & Bahar, O. (2008). Estimating the determinants of tourist spending: A comparison of four models. *Tourism Analysis*, *13*(2), 143–155.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research. *Psychological Methods*, *21*(4), 475. https://doi.org/10.1037/met0000081

Lauer, C., Brumberger, E., & Beveridge, A. (2018). Hand collecting and coding versus data-driven methods in technical and professional communication research. *IEEE Transactions on Professional Communication*, *61*(4), 389–408. https://doi.org/10.1109/TPC.2018.2870632

Liu, W. Y., Yu, H. W., & Hsieh, C. M. (2021). Evaluating forest visitors' place attachment, recreational activities, and travel intentions under different climate scenarios. *Forests*, *12*(2), 171. https://doi.org/10.3390/f12020171

McKercher, B., Prideaux, B., & Thompson, M. (2024). The impact of changing seasons on in-destination tourist behaviour. *Tourism Review*, *79*(6), 1197–1209. https://doi.org/10.1108/TR-06-2023-0420

Mitchell, R. (2024). *Web scraping with python*. Farnham: O'Reilly Media, Inc.

Moreno, A., & Amelung, B. (2009). Climate change and tourist comfort on Europe's beaches in summer: A reassessment. *Coastal Management*, *37*(6), 550–568.

Morgan, R., Gatell, E., Junyent, R., Micallef, A., Özhan, E., & Williams, A. T. (2000). An improved user-based beach climate index. *Journal of Coastal Conservation*, *6*(1), 41–50. https://doi.org/10.1007/BF02730466

Munro, A. (2020). Using experimental manipulation of questionnaire design and a Kenyan panel to test for the reliability of reported perceptions of climate change and adaptation. *Climatic Change*, *162*(3), 1081–1105. https://doi.org/10.1007/s10584-020-02709-2

Ngxongo, N. A. (2021). The impact of climate change on visitor destination selection: A case study of the Central Drakensberg Region in KwaZulu-Natal. *Jàmbá: Journal of Disaster Risk Studies*, *13*(1), 1161. https://doi.org/10.4102/jamba.v13i1.1161

Preston-Whyte, R. A., & Watson, H. K. (2005). Nature tourism and climatic change in Southern Africa. In C. M. Hall & J. Higham (Eds.), *Tourism, recreation and climate change* (pp. 130–142). Channel View Publications.

Rutty, M., & Scott, D. (2013). Differential climate preferences of international beach tourists. *Climate Research*, *57*(3), 259–269.

Rutty, M., & Scott, D. (2016). Comparison of climate preferences for domestic and international beach holidays: A case study of Canadian travelers. *Atmosphere*, *7*(2), 30.

Rutty, M., Scott, D., Matthews, L., Burrowes, R., Trotman, A., Mahon, R., & Charles, A. (2020). An inter-comparison of the Holiday Climate Index (HCI: beach) and the Tourism Climate Index (TCI) to explain Canadian tourism arrivals to the Caribbean. *Atmosphere*, *11*(4), 412. https://doi.org/10.3390/atmos11040412

Saarinen, J., Fitchett, J., & Hoogendoorn, G. (2022). *Climate change and tourism in Southern Africa*. Routledge.

Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, *36*, Article 101003. https://doi.org/10.1016/j.jocs.2019.05.009

Sangkaew, N., & Zhu, H. (2022). Understanding tourists' experiences at local markets in Phuket: An analysis of TripAdvisor reviews. *Journal of Quality Assurance in Hospitality & Tourism*, *23*(1), 89–114. https://doi.org/10.1080/1528008X.2020.1848747

Saurkar, A. V., Pathare, K. G., & Gode, S. A. (2018). An overview on web scraping techniques and tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 363–367.

Scott, D., Rutty, M., Amelung, B., & Tang, M. (2016). An inter-comparison of the Holiday Climate Index (HCI) and the Tourism Climate Index (TCI) in Europe. *Atmosphere*, *7*(6), 80. https://doi.org/10.3390/atmos7060080

Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017). Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*, *8*(2), 220–238. https://doi.org/10.1108/JHTT-10-2016-0067

Stockigt, L., Hoogendoorn, G., Fitchett, J. M., & Saarinen, J. (2018, October 1–5). Climatic sensitivity and snow-based tourism in Africa: An investigation of TripAdvisor reviews on Afriski, Lesotho. In *Proceedings of the Biennial Conference of the Society of South African Geographers*. University of the Free State.

Tanrisever, C., Pamukçu, H., & Baydeniz, E. (2024). Climate change in tourism: Understanding the impacts and opportunities for sustainability. In *Future tourism trends volume 1: Tourism in the changing world* (pp. 33–45). Emerald Publishing Limited.

Tian, X., Zhang, L., & Wei, W. (2016, November). The design and implementation of automatic grabbing tool in TripAdvisor. In *2016 4th International Conference on Enterprise Systems (ES)* (pp. 76–80). IEEE.

Tjaden, J. (2023). Web scraping for migration, mobility, and migrant integration studies: Introduction, application, and potential use cases. *International Migration Review*, 01979183231208428. https://doi.org/10.1177/01979183231208428

Toubes, R. D., Araújo-Vila, N., & Fraiz-Brea, J. A. (2020). Influence of weather on the behaviour of tourists in a beach destination. *Atmosphere*, *11*(1), 121. https://doi.org/10.3390/atmos11010121

TripAdvisor. (2013). Fact sheet. http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html

Tyson, P. D., & Preston-Whyte, R. A. (2000). *The weather and climate of Southern Africa*. Oxford University Press.

Velea, L., Gallo, A., Udristioiu, M. T., & Puiu, S. (2022). Rating tourists' interest in tourism-tailored climate and environmental products. *Journal of Tourism and Services*, *13*(25), 230–247. https://doi.org/10.29036/jots.v13i25.438

Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2018). Tourist activity analysis by leveraging mobile social media data. *Journal of Travel Research*, *57*(7), 883–898. https://doi.org/10.1177/0047287517722232

Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2019). Exploring tourist dining preferences based on restaurant reviews. *Journal of Travel Research*, *58*(1), 149–167. https://doi.org/10.1177/0047287517744672

Zare, S., Hasheminezhad, N., Sarebanzadeh, K., Zolala, F., Hemmatjo, R., & Hassanvand, D. (2018). Assessing thermal comfort in tourist attractions through objective and subjective procedures based on ISO 7730 standard: A field study. *Urban Climate*, *26*, 1–9. https://doi.org/10.1016/j.uclim.2018.08.001

Zeiss, H., Graw, K., & Matzarakis, A. (2022). Impact of the destination weather conditions on decision and complaint behavior of guests who booked vacation rentals. *Atmosphere*, *13*(12), 1998. https://doi.org/10.3390/atmos13121998