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 (2024) Efficacy of Automating the Analysis of TripAdvisor Data in Tourism and Climate Studies, Tourism Planning & Development, 21:6, 959-979, DOI: 10.1080/21568316.2024.2412060

To link to this article: https://doi.org/10.1080/21568316.2024.2412060

TOURISM PLANNING & DEVELOPMENT
2024, VOL. 21, NO. 6, 959–979
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

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,

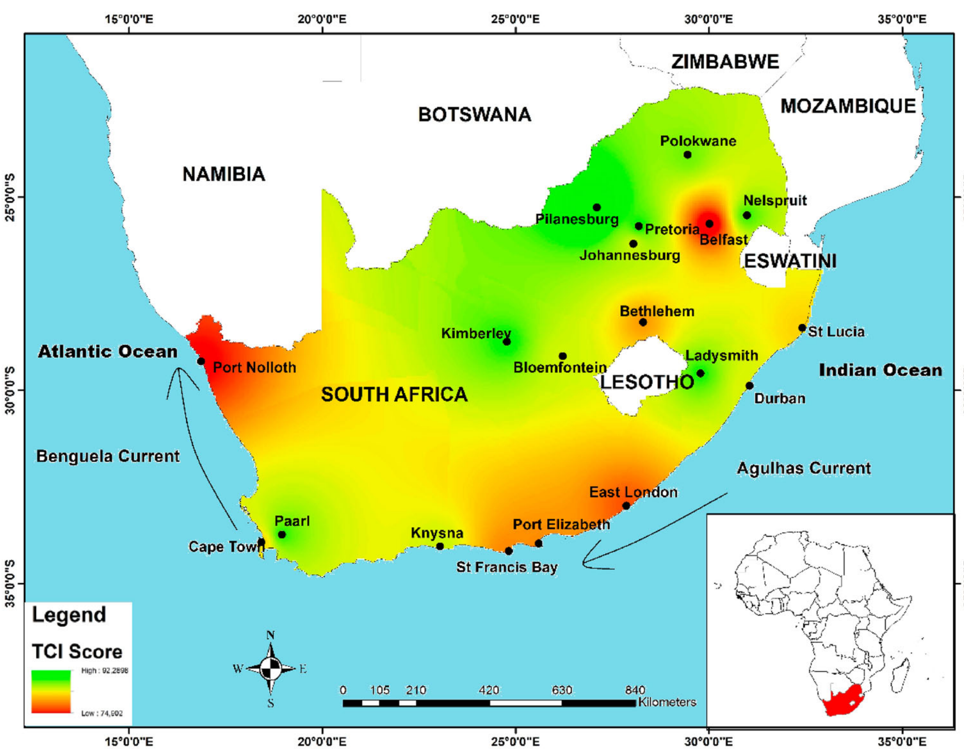

Figure 1. Map demonstrating the 19 different locations in South Africa where the study is based, with inverse distance weighting performed on TCI scores extracted from Fitchett et al. (2017).

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. $p = 0.94$.

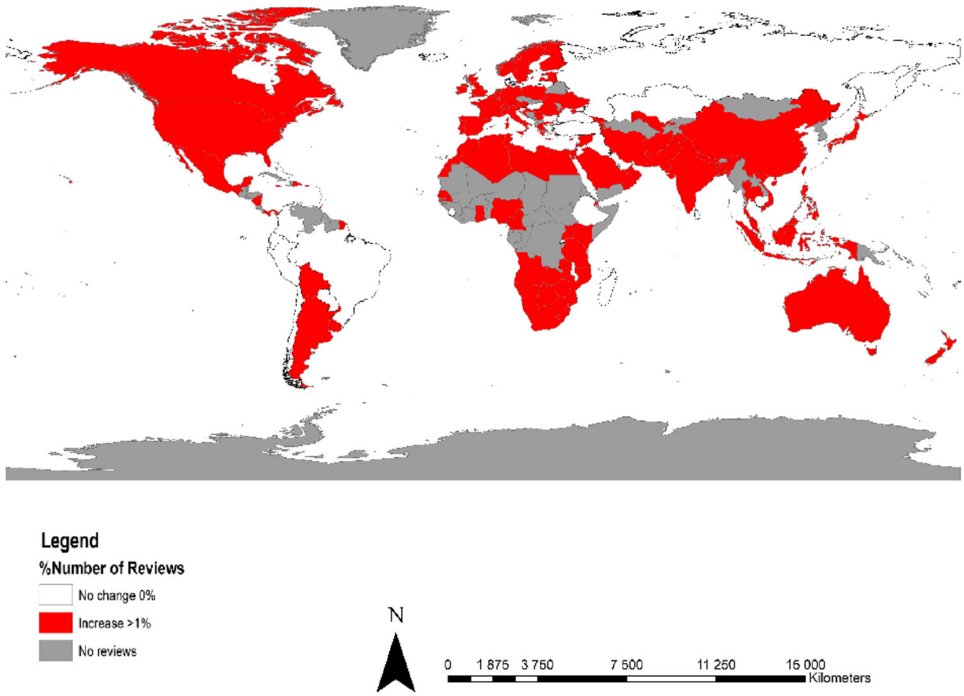

Figure 2. Map showing the percentage differences between the distribution of TripAdvisor reviews comparing the manual and web-scraping approaches.

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.

	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	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).

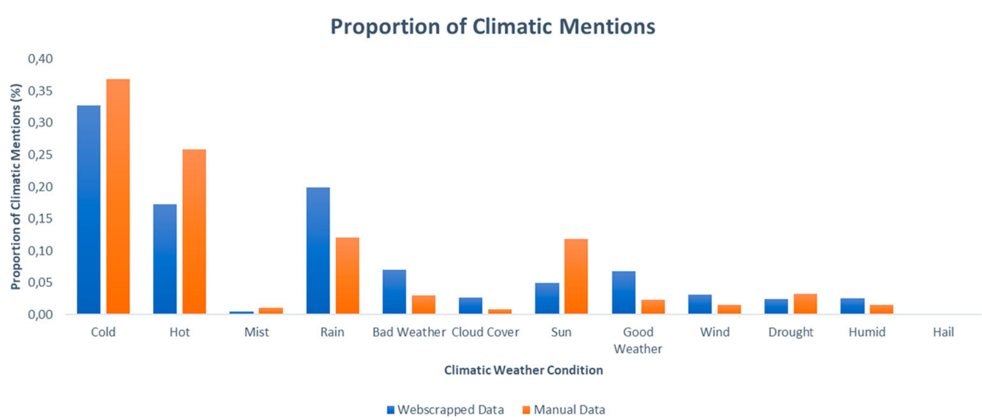

Figure 3. Proportion of climate references in the aggregated TripAdvisor reviews for 19 South African locations.

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.

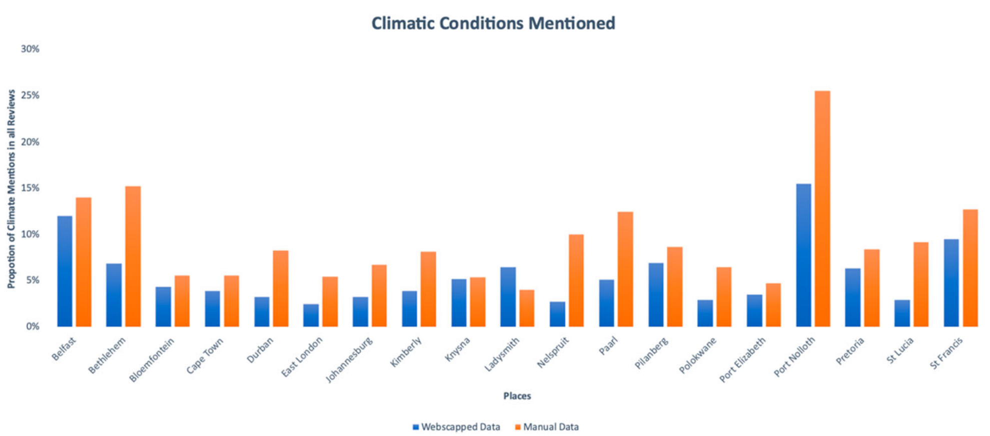

Figure 4. The proportion of TripAdvisor reviews that mention weather conditions, subdivided by location.

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. - Name of a hotel - Indicating a form of energy (wind turbines)	Over-counting in web-scraping	Integration with manually checking
Humid	- Humidifier - Wet/Damp	Over-counting in web-scraping	Integration with manually checking
Mist	- Wet or moist conditions - 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 - 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