

https://dspace.org/

A discourse analysis of twitter posts on the perspectives of xenophobia in South Africa

Makhura, Busisiwe

2022
https://hdl.handle.net/10539/34010

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



A Discourse Analysis of Twitter Posts on the Perspectives of Xenophobia in South Africa

A Research Report submitted in partial fulfilment of the Degree of

Master of Commerce (Economics)

in the School of Economics and Finance,

University of the Witwatersrand

By

Busisiwe Makhura

Student No: 725692

Supervised by Dr. Sibulele Nkunzi

Word Count:

12771

Date:

30 January 2022

Acknowledgements

Writing this dissertation has been a long and at times, difficult journey but I have received a great deal of support, assistance and encouragement.

I would first like to thank my supervisor, Dr Sibulele Nkunzi who encouraged me to take the road less travelled and helped me every step of the way. There were times when I felt lost and overwhelmed, but you brought me through each time. Prof Manoel Bittencourt, who helped me get the ball rolling and inspired me to take a deep look into how the media influences our thoughts and behaviour.

To all my lecturers at Wits, this is a culmination of all your teachings, help and guidance over the years.

Mr Kyle Findlay, who generously shared his database of Tweets with me. At first, I thought I was crazy to think this could be done but you showed me that the magic of data science makes it all possible. Your website SUPERLINEAR has been a great source of inspiration and a helping hand, especially in those moments when I felt lost.

My Mother, Phuti and brother Karabo, thank you for your help, support, encouragement and patience. You've been with me from the very beginning and held me together when I felt like broken. I can't thank you enough.

Contents

Abstract 4
Introduction 4
Power, violence, and the scapegoat 6
Relative deprivation 11
Pride, Prejudice and Ethnic Tension 13
Resource mobilization 14
Methodology 16
Discourse Analysis 22
Sentiment Analysis 31
Results and Interpretation 32
Conclusion 36
Appendix 1: Additional data 38
Appendix 2: Licensing restrictions and limitations 41
References 43

Abstract

Strained relations between locals and immigrants are a long-standing reality in South Africa and often we have seen it become a bloody battle resulting in the loss of life. However, over time we have seen these occurrences make a move into a different arena from (seemingly) purely physical contestations to open and continuous dialogue on public platforms such as social media. In exploring the thought patterns and processes of the participating individuals we believe we may find an explanation of their resultant actions and behaviour feedback into the greater belief systems in society- particularly xenophobic and anti-migrant ideals. Discourse Analysis on Twitter posts related to the #PutSouthAfricaFirst hashtag at two points in time during 2020/2021 is used to get insight into the underlying causes of this ongoing conflict. Sentiment analysis concludes as an attempt to gauge political influence and user sentiment on various themes that appear to be recurrent in these ongoing conversations.

Keywords: Twitter, xenophobia, society, behaviour, dialogue

Introduction

The further we move in unpacking the topic of xenophobia, its causes and consequences the more we will come to terms with the weight of this task. The task we have undertaken calls on us to be humble in our works, remembering that whenever we deal with the human spirit we do so cautiously, aware of our limitations but still ambitious enough to attempt to contribute to efforts to understand, empower and improve the human experience. Going forward, many of the theories explored employ several assumptions which we accept and work with only to help to simplify the context in order to gain understanding but not to box in the human thought process or to put forward a final or exhaustive conclusions.

The use of labels such as riots, crowd, mobs, crazes or revolutions and refer to groups of people acting in one accord as a “collective” are avoided. Labels risk generalisation and assumptions regarding behaviour, this would not serve our purposes considering that collectives tend to differ in social

conditions, consequences and cultural make up (Weller & Quarantelli, 1973). This topic is heavy in voices, perceptions, pain and trauma that deepens with each occurrence of xenophobia is South Africa. The scars left behind are not collective, each is unique and deserves to be handled in a manner that aims to show kindness and preserves dignity. Introducing the perspective of the individual will be helpful although the issue at hand is a societal and national one.

South Africa as a country, its provinces and communities are arguably the most diverse in the world. A complicated and recent history combined with multiculturalism and overdue socio-economic teething problems makes for a potentially explosive balance. The aim of this paper to understand the human psyche, thought processes and resultant behaviour is vitally important before we can understand the role that social media, has to play in governing interactions between groups of people. ~~social~~Social media is used for this purpose because of the increased freedom in speech, expression, and interaction it affords individuals. Although it presents increased freedoms it is not without restrains – offense materials or those that conflict with the law will be removed by the platform. It allows us insights we could possibly only be afforded by listening in on private conversations. Although our main aim is to unpack these conversations, our chosen methodologies are in themselves an attempt at contributing to a potentially new form of research and methodology that employ social media data as a form of a listening tool into the activities of society to gain a better understanding of this new but increasingly important form of communication. It also allows us to expand on the social theories we explore as they help us understand the collective but the data we call on introduces the individual element.

We would be hard pressed to find a single theory that would suffice giving us the complete picture in understanding the workings of the human mind and behaviour that lead to collective action, particularly when these actions are violent and appear to be irrational or lacking explanations. We will spend some time calling on several schools, bringing together thoughts and theories from Economics, Sociology, Psychology, and Political sciences. Remaining cognisant of the limits of each and bringing them together with the hopes that they fill the gaps where the use of a single theory or thought would otherwise fail. In each of theories or school of thought we introduce we will explicitly state the specific theory or assumptions we employ. The first of the theories we investigate explores the dynamics

between violence, the state and power. The rise or formation of the state is important in understanding the distribution of power, the provision of resources as well as how that contributes to creation of hierarchies of social class and ultimately the scapegoat when these structures falter or even crumble. We, then follow with the important concepts of relative deprivation, resource mobilization and the contributing ideas to social identity and other theories that fan the flames of ethnic tension.

Having built the broader picture of the environment we aim to explore we conduct a Discourse Analysis on a collection of Twitter posts or tweets from the #PutSouthAfricaFirst movement. A movement whose purpose, direction and intentions are hotly contested, being labelled xenophobic, afrophobic and even pro- African. The theories we have explored create a base from which we begin our Discourse Analysis, isolating specific themes and keywords that we will look in our analysis and use for the coding of our material. These themes come up again in our sentiment analysis as we investigate the prevalence of these themes during ongoing conversations. We hope that this will shed some light on whether the theories we have formed around human interaction have a practical application or if we need to deepen our investigation and look elsewhere for a more suitable explanation. We investigate what is being said and how it is said. We take a step further by doing a sentiment analysis, going beyond just positive and negative to try and gauge the political element of these conversations.

Power, violence, and the scapegoat

"Institutions frame human behaviour and through these organisations people carry out their (complex) interactions [therefore in order to understand human behaviour we need to investigate the institutions they're subjected to]" (North,1991) – own emphasis.

Social development, its issues and challenges cannot be explained by looking at Economics or Political Studies in isolation. Perhaps the answers we seek regarding the development and (lack of) growth of our country and immediate surroundings can rather be found by delving into the grey and overlapping areas where we find the institutions that govern these environments. If we delve a little deeper into these complexities, we may be able to find what we need to better understand human behaviour and thinking

and the reasons why it seems to differ depending on the extent of development of the individual's surroundings.

More developed societies (or part within larger societies) have more sophisticated institutions in place to adequately provide good and services, carry out research and development and coordinate individual behaviour. The most powerful and centralised of these being the State. But what happens when the state fails or withdraws from these duties in some parts of the broader state? (North, et al., 2008, pp. 55-59)

Human behaviour, particularly violence needs to be prevented, contained and when it does occur it needs to be controlled, making a level of social order required. Throughout human history social order has prevailed, although the means by which it does so have differed. Primitive social order ruled among the hunter gathers but for the better part of the last 10 000 years, limited access order has lent structure to society. This default setting or "Natural State" social order is achieved through political and economic manipulation producing rents in order to paint a picture of a politically stable society. This may be the common setting however not for the more developed world. The resultant state may appear stable, politically as well as economically however it's built on a feeble foundation of exclusion, inequality, privilege, external influence and increased rent producing and seeking activities. This extreme but delicate balance tipped in the direction of the high powered and elite is not easily readjusted. (Roche, 1996).

Returning to the question posed above, what happens when what we call "the State" removes itself or fails to maintain a social order and to deliver basic goods and services to the people? When the state moves from being the "The first invisible Hand" that holds authority and provides provision to becoming the self-serving "Grasping hand"? (Olson, 1993)

In the Democratic republic of Congo (DRC) there remains a part of the country where the government does not exist or has seized to exist having withdrawn, leaving behind a vacuum of power. The dawn and the subsequent rise of the PlayStation brought with is a rise in the demand for Colton, a rarity used in the construction of the console which is found in the DRC. The demand and nature of the resource created an incentive for armed actors to monopolise violence in order to control access, distribution and

profits thereby forming a “government of their own”. This being one of many examples of the theory of monopolised violence as a theory for the formation of a state (Sanchez de la Sierra, 2013). Similarly occurs on a smaller scale within societies, much like within townships and rural areas in South Africa where the government and police forces are unwilling or even afraid to maintain authority also leaving behind a vacuum of power. Where definite leadership lacks and being faced with structures that appear to be failing, the gap is filled by communities that undertake to lead themselves. The age-old theory plays itself out in present day masked as vigilantism and criminality (Madhavan & Loren, 2011). South Africa’s history of segregation, discrimination and economic exclusion, makes service delivery even more important in the daily lives of individuals, for some more in more obvious and dire ways than for others- making it all the more crucial for the State to not only be strong but also efficient. Given this history and its direct impact on the environment and living conditions of the individual, the principle of “Equality without justice” leaves the underprivileged confused and conflicted, being faced with the realisation that the promises laid out in the Freedom Charter and their rights afforded by the constitution are undermined by the lack of delivery of the very materials needed to realise those rights. Delivery required from the authority by the social order (Landau, 2010). When the structures fail those, who are governed by them, environmental tensions simmer beneath the surface as competition ensues for the minimal resources that are available regardless of how efficient or suitable, they may be for the intended purpose. These cramped, poverty ridden spaces become a melting pot of language, culture, religion and nationality. The daily struggle becomes a battle on grounds of an individual and a nationality basis. Where Service delivery is simply the physical, material benefits of citizenship, slow (or a complete lack of) service delivery, poor housing and corruption opens the gap for self-governance to emerge in these desperate communities.

While the leaders’ appetites and consumption are not limited by their personal capacity but rather by their drive to operate within the greater power struggles in the political domain and they compete to these ends, the citizens do so for that which lacks- housings, medical care, land and infrastructure, informal trading and insufficient job opportunities and even a basic life source-access to water (Olson, 1993). Almost ironically, these poor societies are mostly located on the doorstep of affluence such as

Alexandra, a stone throw away from Sandton, the mecca of luxury. They're at higher risk of civil unrest as incubators of social alienation, poverty and intense competition for scarce resources where the black youth are the dominant group. Together with arguably the highest levels of inequality in the world, this makes for prime conditions for the breeding of crime. Coupled together with high volumes of desperate and poor immigrants whose physical features and unique linguistics, according to some South Africans, are different- xenophobic violence is imminent as desperation and frustration seek an outlet.

Contrary to opinions of state leadership, occurrences of xenophobia are neither spontaneous nor are they unexplained popular revolt nor is it irrational violent outbreak as they describe it. Further chastising those responsible for this blatant form of "criminality" and calling on the use of excessive, brutal policing does the very opposite of what is needed to address the issues and mend what is broken within our societies, instead it seems to serve as a catalyst further angering communities. A rather classic case of treating the symptoms and not the cause. Reiterating a point made earlier, reforming current social order requires confrontation, acknowledgment and action, not in a way that scrambles to find a scapegoat to silence the masses and pray that the 'craze' fades, but directed at the roots of the problem of failing structures and weak authority coupled with white privilege, distrust and an overall sad nation. (Madhavan & Loren, 2011). Recent years have seen South Africa's ranking in the happiness index has steadily declined. (Appendix 1).

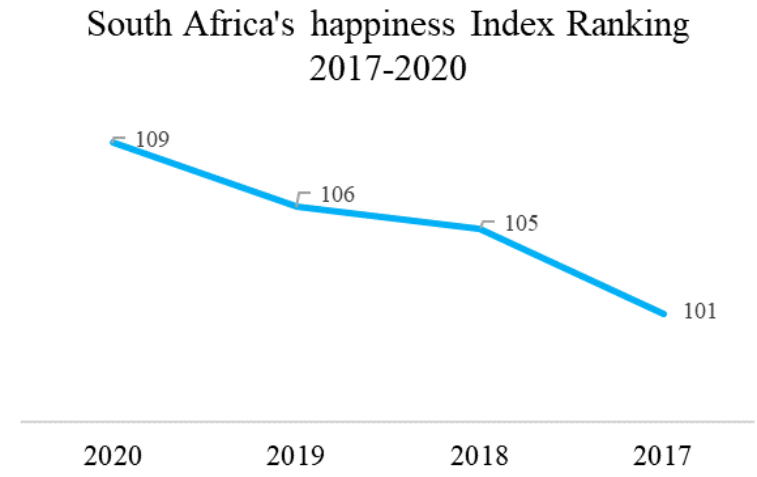


Figure 1: South Africa's happiness index over the last 5 years. Data in appendix 1

At this point it has become clear that xenophobia is far more complex than "criminality" or a "immigrant issue" (Yitzhaki, 1979). As raised before, vacuums of power are created when definitive, directive and legitimate leadership is lacking, there the opportunity arises for communities to captain themselves and for local leaders to divert attention and hostility towards immigrants as a means on "maintaining authority" where they've failed (Roche, 1996).

Moving forward we will work toward unpacking this phenomenon and in appreciating the complexities of the topic and in order to do it justice, drawing on several sources and theories to better understand the human element to this phenomenon, building towards the bigger picture. This is especially useful for the aims of this paper as together with the insights gained from social media conversations and interactions, helping to understand the cyclical relationship between the human psyche and resultant behaviour and its broader environment when combined with an intricate history and a tapestry of cultures like of South Africa. Furthermore, it will further help us understand why its occurrence are focused to certain races, living environments and income groups. The possible reasons, causes and theories are by no means exhaustive or conclusive- but include factors pertaining to the interactions and exposure of the groups to each other, cultural factors including issues of identity and nationalism and as well as material or economic factors. We aim to broaden our view and contribute to finding a treatment for this illness that appears to be spreading from the top down, from power platforms of government leadership to civilians (Vearney & Walker, 2020) . Xenophobia is in no way unique to South Africa but rather has become a global experience. The fast (unequal) globalisation of society may be one the attributing factors to this experience. With some societies and parts of the world getting left behind, this has prompted the mass movement of people across the globe seeking opportunity, prosperity or safety. What is unique in a sense is the stance of our governmental leaders relative to that of leaders in many other parts of the world faced with a similar immigration situation (Choane, et al., 2011). The USA and some parts of Europe have had their leaders take on a more negative stance. Uttering blatantly anti-immigration sentiments from their platforms of power. Although this may, at first glance be similar to the actions of South African leaders, their position is one of a straightforward fight back vs the local position of "scapegoat seeker" (Vearney & Walker,

2020). The story of a leadership unable to or afraid to take control of a deeply, unstable and angry nation (Madhavan & Loren, 2011).

Relative deprivation

The achievement of democracy proved to be a victory that brought the promise and expectation of not only a new national leadership but also a better life for the people and a brighter future for generations to come. The voices of leaders boomed into willing ears and further fuelled hopeful spirits with the Freedom Charter as their roadmap into a new world. The bright lights of the new South African shone bright, bright enough to be seen by our Northern neighbours, a beacon of hope for them too. Democracy not only exposed South Africa to the prospect of development and the opening of economic borders but with it led to an influx of migrants in search of sharing in the opportunities of the new "Rainbow Nation". All this happened at a time when the nation was still reeling from recent events and coming to terms with its newly inherited social and economic complexities. Proposed policies, heavily laden with promise together with a transition into democracy founded on negotiated settlements (with several stakeholders, many of which included neighbouring countries, during which promises were made) birthing plenty of ambiguity and confusion particularly in terms of the country's new foreign policies. The resultant socio-economic conditions were new territory for all, requiring decisive and immediate action lest the people become lost in their hopes and the leaders to their newfound ambition. Fast forward 27 years and the "young democracy" still battles weak institutions, riddled with ambiguity but now coupled with scapegoating, lack of responsibility and rife corruption. The Freedom Charter places service delivery at the forefront of citizenship, together with its material rights and privileges. We have previously touched on the point of the importance of service delivery particularly to a nation whose history has shaped the living conditions of its people and how "Equality without justice" has left a part of society not just confused and conflicted but now also disappointed and deprived.

The minds and actions we seek to understand live and operate in circumstances of poverty. Amartya Sen describes poverty as deprivation- being an environment that prevents one from reaching or living out their utmost potential. Consistently living with the lack of the most essential of resources,

inadequacies in levels of income, education and health have the ability to penetrate beneath the surface moving from being a physical condition to festering as insecurities, low self-confidence, powerlessness and the absences of rights- the most basic form of citizenry (Sen, 2011). As intricate as many of the issues we address in this paper, poverty, (relative) deprivation and life satisfaction are no different. It cannot be addressed by simply increasing income levels or curing homelessness hoping it will cure anti-immigrant sentiments held by the masses but a problem this complex calls for a transparent, all-encompassing solution that seeks to strengthen and empower, to ensure against risk and stabilise the foundations of weak institutions (Gordon, 2016).

In *Conflict, Power and Relative deprivation,* Korpi explores three forms of relative deprivation and their contribution to and the probability of conflict in society in relation to shifts in elements of power within that very society. For our purposes we will focus on only one of the discussed forms namely *Progressive deprivation* (Korpi, 1974). Progressive deprivation can be described as the steady improvement in the social and economic conditions of individuals which in turn leads to increased expectations for future circumstances but when there is a downturn or reversal in that trend of improvement, we have Progressive deprivation (Brush, 1996, p. 529) .This definition is a perfect fit for the South African case.

Post the transition into democracy expectations of a better future were possibly highest among the part of society that had suffered the brunt of segregation, extreme exclusion as well as both relative and absolute deprivation. Promises around housing, security, healthcare, education and employment raised expectations of a move out of poverty to live out the dream of 'an equal nation'. The state of the institutions of the country have had direct and dire consequences for the very people who had hoped for change. In the wake of deteriorating structures and weak and ambiguous policy we have seen a drastic reversal in the trend of (perceived) improvement, particularly in the vulnerable spaces within society. (Stark & Taylor, 1991)Progressive deprivation has become the daily reality of the South Africans. We have established the possible emotional and psychological the role that poverty has on the individual, this is one of the building blocks we will use to building an understanding of the reasons for violence within a specific income and ethnic group.

Pride, Prejudice and Ethnic Tension

We cannot discuss xenophobia or anti-immigrant sentiments without critically looking into issues of ethnic identification, social identity, nationalism and devaluation on an ethnic basis. This topic is particularly important for our purposes because it may help us shed some light or bring understanding to a phrase, I have heard one too many times but fail to fully comprehend. “Black people hate themselves”. This phrase never fails to rear its head when xenophobia occurs, ethnic tensions arise, are discussed or even suggested. In addition, this section is of particular importance as it brings together some points, we have discussed above under the umbrella of theories of social identity giving context and practical application to understanding not just the outward human behaviour as well as inward attitudes and cognition.

This section shifts focus from the physical violence aspect of the topic to the more discursive forms of conflict that occur among groups. This is important for our purposes, as we make use of the more expansive definition of xenophobia instead of boxing it in only in its forms of physical violence. In this case separating the causal factors that determine attitudes from those determining violent action becomes necessary. In addition, focusing solely on behavioural outcomes and violent outbreaks at the expense of analysis of xenophobic attitudes risks the underestimating of the spread and extent of xenophobia and takes away our power to attempt to predict future occurrences.

The building of a national identity was an integral part of the governmental transition in South Africa post-apartheid as with many other nations coming out of war, financial ruin and other forms of national disaster. It is done with the intention of rebuilding, recovering and forging ahead however it can have a potentially negative side where it is hierarchical. In the growing of national and cultural pride, we subsequently create the national “ingroup” and the immigrant “outgroup”. The outgroups are perceived as inferior where there is explicit comparison with other groups. Social Identity theory or SIT goes some way in making sense of the (perhaps unintended) adverse effects of nation or identity building.

The theory puts forward a hypothesis- the creation of an “ingroup identity” promotes competitive behaviour with the outgroup (Placeholder5). Ethnocentrism- the process of creation requires

differentiation or even the devaluation of the outgroup. In this case an extreme form of nationalism can lead to xenophobia or xenophobic sentiments (Herrera & Butkovich Kraus, 2016).

An alternative hypothesis speaks directly to another perspective we covered regarding deprivation, poverty, inequality and the reality of broken promises. Being economically vulnerable, deprived and feeling helplessness compounded with being in direct competition with the "inferior immigrant" breeds internal feelings of contempt (Herrera & Butkovich Kraus, 2016, p. 297) . This leads to a potential conclusion that xenophobic attitudes and resultant conflict have their roots in a materialist position that expresses itself through physical and verbal contestation.

Resource mobilization

Resource mobilization theory brings another angle to our investigation, that of how social movements organise and work to drawing in funds, media attention and alliance towards their cause, whilst highlighting grievances shared by a group, ideologies about social causes as well as realism and power. This theory becomes important in our case as we look at what has moved from being an single organised march to becoming a movement that has gained traction and attention. It is particularly useful as it embraces the differences between individuals and groups. It is important to emphasise that the theory we apply here is the *Entrepreneurial and organisational* form of the theory by McCartney and Zald. This form focuses on the role that economic factors and theories play in social movements as opposed to the theory put forward by Tilly and Adams, one whose focus is on the political process where the cause of collective action is attributable to sociological factors and political struggles- this doesn't assist us because actors are then assumed to be irrational and deviant, this kind of behaviour cannot be explained, whereas we are looking for an explanation for this behaviour. Zald and McCartney propose that feelings and grievances, although necessary, are not sufficient in explaining the rise of movements and their subsequent interaction with various groups. (McCarthy & Zald, 1977). The entrepreneurial theory points us to look at the access and control of resources as a crucial factor thereby creating linkages between the flow of resources, supply and demand as well as resultant individual action. This

theory sees actors as rational, interacting various types of resources including moral, ethical, cultural, socio-organisational material as well as human.

Together with the theory of progressive deprivation, a form of relative deprivation, allows us to bring in the view of the individual, without particularly overemphasising their role. There are acts that appear to occur between certain ethnic groups and aimed towards specifics (income) group- the individual becomes important understanding to the nature of the acts carried out (Jenkins, 1983). There's a risk of doing away with the importance of individual we have achieved this by including the theory of resource mobilization. However, this theory is particularly important because of its ability to express the intersections between social movements the media, authorities and other parties (McCarthy & Zald, 1977).

We bring in these theories in expanding our understanding because of the crucial role that they play in the birthing of ideologies (and general belief systems in societies), in this case ideologies that lend to the anti-immigrant sentiments, ripples that later become the rolling waves of the tension and violence that we witness.

We seem to have come across a pattern where violence is used as a means of interaction on an institutional level. We observe the monopolisation of violence as the building blocks to state formation in the presence of a power vacuum now in the light of social interaction violence again veers its ugly head only this time as a means of social control.

Up to this point we cannot clearly point to a single theory to be the cause or the root of xenophobic attitudes and acts of violence. Having looked at existing theories, what we have discovered is only but the tip of the iceberg. In order to expand to our investigation in order to fulfil the purpose of this paper, we need to investigate what exactly individuals are saying, in their private capacity. Unrestricted and unpoliced will perhaps shed further light. A critical analysis of the discourses and sentiment that surround the topic may answer the many questions we have.

Methodology

Having explored the various contributing factors to human behaviour and thought processes, gaining a deeper understanding on how society's belief system gives rise to its displayed behaviour which either alters or reinforces the existing belief system. The cyclical nature of the relationship further adds to its complexity.

We have chosen to use social media data in our endeavour, Twitter posts relating to #PutSouthAfricaFirst. We are drawn to this data form because it can take the form of social listening and audience intelligence, achieving this through observing the content and uncovering trends and insights that emerge from this data.

Social media is developing at a rapid pace, and it gives unique insights in analysing and understanding the ways in which people are using it to communicate as well as trends of behaviour. Behaviours and conversations on these channels are a mirror into the daily physical and mental behaviours of individuals. As social media continues to grow in both size and importance, an integral part of society and of daily life, it is vital to delve into this new area to apply research tools to understand it and form new methodologies to be able to leverage the insights that perhaps only these platforms can offer us. We have decided to take on the challenge and attempt to not only contribute to understand the social dynamics of the xenophobic plague but to also led a hand to forming these tools by embracing this new form of data along with the use machine learning in the data collection and analysis processes.

We had the option of using traditional data collection methods (i.e., Survey data) which would have been quicker, simpler and less stressful in many ways including being easier to administer, less time consuming and quite possibly most importantly, we would have been able to limit the data collection to particularly what we needed. Having a dataset that was created intentionally from the start would have removed much of the difficulty experienced in working with the raw Tweets, as Twitter and its posts are not created with the intention of being use as data. However social media data can give us a

quantitively and qualitatively rich data set, participants are open, willing and voluntary. This form of data also gives us the chance to address the shortcomings of the surveys.

These include respondent related issues such as:

- not feeling encouraged to provide accurate, honest answers
- Respondents may not feel comfortable providing answers that don't present themselves in a "politically correct" manner.
- Respondents may not be able to recall their answers or opinions due to memory or being disinterested.

Data: Collection and Challenges

Our analysis is based on two datasets across two specific periods following the same hashtag, #PutSouthAfricaFirst.

The first was generously made available to us by Kyle Findlay. It contained a total of 321 953 tweets, from 5th June to 14th August 2020.

The tweets were received in a comma delimited csv format, making it rather impossible to work with as is, it was abandoned for some time while we sought a solution for structuring the data. We later returned to the data; it was then put through an ETL process using Excel's Power query editor. Once the queries and connections were loaded, the resultant data was in the form of a table saved as a xlsx file.

The output was data sorted into columns including:

- created at (date),
- created at (time),
- language,
- mentions,
- user friends count,
- user. location,

- user.name,
- user. screenname,
- tweet type

and several other columns specifically related to tweets that are either retweets or replies to a tweet. We have not listed these as they were not used for reasons that will be explained a little further on.

The second database of tweets was extracted from directly Twitter using a Twitter API we have built and a Tweepy Python code that we custom scripted. The resultant database contained 18 000 tweets, from 30th December to 15th January. The number is significantly less than the first database we received as the filters we discuss below were used to narrow the data down to fit our requirements had already been hardcoded into the script and the extracted data was already filtered.

To be authorised to access Twitter and extract data from it, an application was made directly to Twitter for a Twitter developer account that would grant us access tokens and consumer keys that would be used for the authorisation. For the purpose we needed both read and write access to the data, therefore we required OAuth 1a Authentication. The keys and tokens were passed directly through the code in the OAuthHandler part of the script. The granting of the developer access to Twitter required agreement to a strict privacy agreement (the details of the license and privacy agreement conditions are further highlighted in the 'License and limitations' section of the appendix). To ensure that the privacy agreements are adhered to the Usernames and user descriptions associated with the individual tweets were removed from the database immediately after extraction, before any cleaning, structuring or analysis could begin.

Data: The ETL process

The cleaning of the data is applicable to both the Discourse Analysis and Sentiment Analysis parts of this paper.

The first dataset was cleaned, filtered and formatted as follows.

1. Having removed any elements that may be used in identifying the individual users, we filtered the data as follows:

a. Removing retweets and mentions (or replies) to isolate individual tweets rather than those that are in response to another. Each tweet was to be a stand-alone, representative of an idea, belief, or opinion.
b. It was not possible to use the available language filter to ensure we only had English tweets. Having done quality control checks we found that the coded language in the column was not completely reliable. Filtering by 'EN' did not return English tweets only. Taking into consideration that these tweets were mostly generated in South Africa, the mixing of two or more languages in a single tweet was a common occurrence.

   The language elements were catered for later during the data cleaning and structuring process.
c. The locations provided were also found to be unreliable however not owing to the Twitter app itself or the extraction process but on the part of the individual Twitter users:
   i. Not everyone grants the app "location permission" so that their actual location is shared with app or webpage on the device they're using.
   ii. Some users choose to manually input their locations. In these cases, we end up with locations such as,

In the Shadow of His wings,

Black Twitter Avenue,

Inside a cocoon,

Mo spacing,

Among several other creative locations/geographical descriptions. Having found this location was also omitted from the analysis.

Lastly the data was filtered for tweets with a Timestamp containing '2020/07' and '2020/08'. This period is of particular importance as it highlights the discourse in the weeks leading up the '29 August' march organised by the #PutSouthAfricaFirst movement.

The second dataset contains tweets for the period 15 December to 15 January. This period is also of interesting as it occurred at the same time as the trending of #Beitbridge following the influx of foreign nationals -mostly Zimbabwean nationals- towards the Beitbridge border post in an attempt to head home for the festive season and then to return to South Africa in January. The congestion was further aggravated by the Covid-19 protocols and the subsequent closure of the land borders on 11 January 2021 by President Cyril Ramaphosa. As mentioned above, the second dataset did not need to be filtered and structured as much as the previous as the requirements had already been included in the python script.

Lastly the columns in the two datasets were renamed to ensure consistency between the two datasets.

Having filtered and cleaned the data. Rapid Miner was used to create a representative sample as well as to aggregate, clean and structure the data in a way that it would be suited for our analysis purposes. Rapid Miner is a data science platform useful for preparing data, machine learning & predictive modelling.

Data: Creation of A Representative Sample

We created a representative sample of the data following the steps below:

- We generated unique IDs for each tweet- going forward this will be for identification purposes. This is will particularly useful in coding the tweets for the Discourse Analysis which will be addressed further below.
- A custom sample creation process was manually created in RapidMiner, it is for a stratified sample using the unique characteristics of each tweet. The same process was created and executed for both datasets. A stratified sample was chosen to allow the program group together tweets with similar attributes or characteristics in order to have a balance random sample. A stratified sample was chosen over a simple random sample as a stratified sample provides the

process with greater precision than a simple random sample of the same size would provide. Due to greater precision using a stratified sample allows us to make use of a smaller sample size reducing the required processing time.

- The extracted sample consists of 2066 for the July/Aug period and 2 628 For the Dec/Jan period (approximately 14% of the original dataset sets). The resultant number of tweets is the output of the process, although manually created.

Data: Cleaning and Structuring of The Data

The cleaning process below was carried out on all four sets. The two representative samples as well as the full datasets for the two time periods.

The steps to the cleaning process are as follows:

- All nominal values are converted to text to ensure that process is able to read and fully understand the data. This is done as nominal values tend to confuse the processing systems.
- The sets are tokenised by linguistic elements, where each element is turned into an individual item.
- Filtering and removal of any @ and &amp symbols. The @ is particularly to ensure to remove any mentions or usernames we may have missed in previous steps. This step acts as a final check. This is a step to ensure that no usernames remain in the set.
- Filtering and removal of any links to websites, articles, or media such as videos, pictures or audio. These forms of media will not form part of our analysis (we give further insights for this further below).
- Lastly, we filtered and removed content by length. We removed words fewer than 3 characters, which includes and, if, the, and other common articles. This is particularly useful for analysis later in the process, when looking into the use and prominence of words and/or word groups. If unremoved, theses words could come up as prominent which is not useful for our causes.
- The output is a cleaned dataset, however throughout the process the original text is maintained in a separate column comparison and readability purposes.

The prepared datasets are used for both the Discourse and Sentiment Analyses.

Discourse Analysis

South Africa has a complex history, combined with even more colourful interactions with its neighbours, immediate as well as a little further afield. Xenophobia in South Africa is almost as old as democracy itself- if not older. Most recently we have seen it become a widespread, continuous dialogue on public platforms, a stark difference from the previous violence, bloodshed and displacement of the "other". In considering the posts we will not label the conversation or a side of it as xenophobic or not but rather we examine the comments and contributions as being either of a positive, negative of neutral tone. In this section we use the language element of the tweets to attempt to extract pointers to the underlying knowledge base and gain a better understanding of how physical socio-economic conditions of communities' feedback back into the belief systems and resultant speech and actions. Xenophobic violence is the proverbial axe hanging over our heads. One that calls for permanent solutions. This leads us to the question.

"How do the things people say and do affect society at large? And how does society in turn influence its people?"

Discourse Analysis to help us navigate towards the answers to these questions. Owing to the nature of the conflicting systems of philosophy that sociologists, political thinkers, and other schools have developed around the term, it becomes important to fully and clearly lay out each of the individual elements of discourse theory that we will be using and embracing throughout our study.

Discourse can be broadly defined as a conversation or discussion. However, due to the nature of human interaction, discourse is better described as all forms and means of communication. Critical Discourse Analysis and political Discourse Analysis are both text-only based approaches, we will borrow elements from both approaches as we work only with written texts. Discourse is unavoidably and heavily linked to a form of written and spoken language, highlighting the explicit link between expression and human knowledge. However, it is important to reiterate that people communicate through speech, text, sound, visuals, taste, and smells, all of which are a type of discourse. This type of multimodal approach would also examine pictures, movies, games, food etc. in a bid to reveal a myriad of social truths. This study

is associated with semiotics- the study of how signs stand for different objects. Although Twitter is a multimedia platform, we will be look only at the written Twitter posts- excluding all other media forms- therefore allowing us to follow the text based approached mentioned above.

People use languages to speak and produce written texts drawing from their individual knowledge, which itself is drawn from the pool of generally accepted knowledge and the belief system held by the broader society. While individuals produce elements that feed into the knowledge system, these very same discourses feed back into society either reinforcing the existing or igniting new social ideals. What a society holds to be true shifts and changes over time. It is highly dependent on the exchange and expression of ideas within communities as well as the means and speed at which these exchanges occur. Leading to questions of how these people move to shape the "flows of knowledge". Exposure, income, education, background -among many others- determine the opinions that one holds. Everyone will hold onto their truth differently depending on their personal intersectionality of the above and several other elements. This shapes their position in their larger community where the "other" maybe excluded. Discourse theory as we follow it concerns itself with questions of power and institutional hierarchies that in turn lead to questions of dominance and resistance- this will follow closely on our previous discussion above- *Power, Violence, and the Scapegoat*.

Establishing the context

The first step in the process of Discourse Analysis is the establishing of context. There exists a very close relationship between discourse and its context- which helps interpret the meaning of utterances. Context is foundational-essentially without context, there is no discourse.

Looking at the tweets we need to take into consideration more than the content of but also the context in which they occur- in which the discourse has taken place. A difference in context may have a significant effect on the content, interpretation and effects of the discourse. As with many of the points we explore going forward. '*Context'* has a few definitions, for the purpose of clarity and consistency we will be explicitly clear in the definition that we will choose to use going forward.

H. G. Widdowson explained context to mean, "those aspects of the circumstance of actual language use which are taken as relevant to meaning". Whereas during his study of the existing relationships between

discourse and literature, defined context in both a broad and narrow sense. In the broader sense, he took it refers to being knowledgeable of external factors as well as other parts of the text under consideration. Something he'd refer to as the "co-text". On the other hand, the narrow sense would refer only to the external factors affecting the text under consideration (H.G. Widdowson, 2000, p.126). On the other hand, we have George Yule who also considered the meaning of "context". The definition he provides is the one we will follow throughout our analysis. His definition is as follows:

"Context is the physical environment in which a word is used." (George Yule, 2000, 128)

Although the definitions above are shaped from differing perspectives to serve different purposes, they centre around a common point: the undeniable link between context and the environment in which discourse takes place.

To extract more meaning from the text we are working with we will break the context down further into three segments- the linguistic, cultural and situational and apply them directly to our analysis.

Linguistic

The linguistic elements tie in with the concept of othering that we discussed in previous sections. The in-group and out-group context comes through strongly using words such as us, we, them, they, these/those people. The use of these deictic expressions- to be more specific, these personal expression- assign roles and positions within the discourse.

We have a clear distinction between the in group, being the locals or citizens and the out group or the 'others", being immigrants or foreigners. The distinction is made implicitly during one's utterances referring to another.

Cultural

Cultural context includes the culture, customs and background or history of the communities (in this case it may even be the country of South Africa as a whole) in which the speakers, in this case Twitter users. The language used or how participants choose to frame their arguments largely dependent on the social structure and belief/value system of the societies the individuals come from as well as their daily experiences and interactions and how they perceive them. As such we conclude that language use is

strongly influenced by several factors such as social role, social status, sex and age, and the unique disparities in the socio-economic statuses of citizens owing mainly to their income group.
Social roles are culture-specific functions, institutionalised in a society and recognised by its members.

We remain constantly aware of the uniqueness of the individuals experience, dependency on the state and the proximity of their interactions with people from other nations and the extent to which they must share their already limited resources. These have a strong influence on the language used in expressing opinions and thoughts, which collectively shape the structures and value systems of communities.

Situational

The situational context is a breakdown of elements such as the environment, time and place within which the discourse occurs. To make this clearer we will address this also under three headings, namely, field, tenor, and mode.
The field describes the ongoing activity around the discourse. Tenor refers to the greater social relationship enacted by the discourse- highlighting how the use of language not only describes existing interactions and relationships but further influencing the topic or subject of the discussion. Lastly, Mode investigates the medium or channel used to transmit the greater message and whether the transmission is delayed, or the contact is immediate and direct.

In the context of our analysis the communication takes place on the Twitter platform that allows for immediate, unrestricted communication or transmission of the intended message. The #PutSouthAfricaFirst subject of the discussion is largely centred around the presence of illegal immigrants and the perceived roll-on effects of their close presence in other areas of life including living conditions, employment, and the availability of resources. Ness The period we have chosen to look plays out on a field with the background of the Covid-19 pandemic with other pertinent issues in the foreground including allegations made against foreigners (particularly Nigerians) ranging from criminal acts to human trafficking during the June-August period to the failing processes and systems at the country's largest land border post- the Beitbridge Border post.

Exploring the production process

Twitter is a microblogging and social networking platform on which interact through messages known as "tweets". Tweets are posted, liked, and retweeted using the website service or user app. Twitter Users can group posts together by topic to forma type of ‘conversation” by use of hashtags, where words or phrases prefixed with a # sign. Each tweet is restricted to 280 characters, audio and video tweets are limited to 140 seconds. The Twitter platforms allow users to tweet and interact with minimal restrictions, the platform affords users increased freedom of speech- giving us privileged insights into real time, -mostly-unfiltered personal opinions. This is well suited to our intended purpose as it allows us a view into the perceived truths and belief systems held by individuals regarding the presence of immigrants and the effect it has on their daily lives.

We will be using #PutSouthAfricaFirst in two distinct time periods. The hashtag has grown in popularity in recent months, having been birthed in early 2020, resulting in several planned marches and calls to action. Uniquely, this hashtag has not faded away like many that trend but has become a continuous and open conversation among Twitter users and occasionally transitioning into mainstream media forms. For the purposes of our investigation, we will limit the time period of the first dataset to July and August 2020. This period is particularly interesting as it looks at the interactions before and immediately after the 29 August march, organised by the #PutSouthAfricaFirst movement. It also captures the hive of activity of a call to action by residents of Sunninghill, Pretoria and Hillbrow, Johannesburg to ‘clean up’ the streets following a public cry by a young lady detailing her ordeal in which she and a friend were lured into a Hillbrow apartment, allegedly with the intention of trafficking them. The apartment was said to be owned and occupied by Nigerian nationals.

The second period we look at trended very closely with #BeitBridge. We have chosen to examine this topical hashtag over the December 2020- 15 January 2021 period, during which it trended continuously for several days following the influx of Zimbabweans towards the Beitbridge border post in an attempt to head home for the festive season. Commonly known to be busier at that time of the year, the situation

was further aggravated by the Covid-19 protocols initiated by officials before the subsequent closure of the land borders.

Preparing the material for analysis

Considering the volume of the data we are working with (a total of 4694 tweets across the two samples). We combined the two representative samples from the two time periods and then created a third representative sample. We want to understand the cyclical relationship between discourse and action as well as the reverse. We will keep two sets of the example. One where the tweets will be kept in their original state so that we can read them clearly, these will be used in the coding process. The second set will undergo the same cleaning and structuring process that we used in the sentiment analysis process to conduct quantitative analysis on it.

Coding the material

Having explored the background of the topic and its potential causes of this phenomenon at the start of this paper we were able to identify and extract the list below of themes and keywords that we expect to find upon examining the data:

- Power and violence
- Institutions and hierarchies
- The state, state formation and its legitimacy
- Social development
- Lack of resources, deprivation, and poverty
- Cultural and identity theories
- Citizenship, its associated rights, and privileges

The volume of tweets in the representative samples was too vast to manually code the categories and discourse strands to then extract the themes and keywords. We combined the two sample datasets from the two time periods and then reused the sampling process used previously to create a stratified representative sample of both time periods to get to explore the broader conversation over an extended

time. Combining the two sets allows us to capture the conversations and any resultants changes that may have occurred in the period in between (i.e., September to November). The created sample contained 657 tweets.

These 657 tweets were individually read and analysed to extract the theme being addressed as well as its linguistic, cultural, situational and context. The process employed; we will refer to as *evolutionary coding* where our expected themes have evolved from theoretical considerations to a fully operational list based on our empirical data. (The full extracted list is shared in Appendix 1).

We found that the operational list did not deviate from the initial list instead included a few new insights which we have narrowed down into broader topics that we can explore deeper in the rest of our analysis.



Figure 2: Word cloud for #PutSouthAfricaFirst during the July-Aug period

Examining the structure of the text

Having collapsed the individual topics into broader themes, we are able to move towards establishing a link between the empirical themes and our earlier theoretical investigations. At this point in the Discourse Analysis, we examine the structure of the text for insights into the strands of discourse to establish whether they exist in silo formation or if there is a significant overlap between the themes which represent elements of the broader belief system. The themes we have extracted and grouped have pointed to the themes we have explored and discussed at length above.

Discourse and cultural statements:

The collection of tweets leads to two main ideas, two main discursive strands. Namely the Pro-South African group of tweets and the more anti-immigrant (also labelled "xenophobic" sentiment). The first, holds the main idea that a country, its State, and its people should be held as a priority and be provided for before individuals from other countries (particularly from the poorer neighbouring country) can be catered for, maintaining that citizen rights should be placed and respected above all else. These rights must be fulfilled by the State, failing which full responsibility should be assumed by the same State. The second strand holds a similar sentiment but differing in the direction of the focus and blame. The second point on which they hold a similar belief but then differ on the point of who should assume responsibility is on the volume of immigrants within South Africa's borders. Although the commonality in their arguments lies in that the volume of foreign nationals is just too great and coupled with the few and dwindling resources, our systems are failing to cope. The two groups appear to move in the same direction here but then diverge on who should assume responsibility for these failings and what should be done about them. One moves towards blaming the State for porous borders, corrupt border officials, failing institutions including Home Affairs and International relations as well as poor legacy systems. The other points to the immigrants themselves questioning their reasons, motivation, and efforts in seeking refuge within South Africa's borders, echoing the sentiment that each country is riddled with its own problems and failings and should not be burdened with cries from their neighbours and the continent at large. This could be very important in explaining the reason behind the change in

interactions between locals and foreigners over the years. Initially foreign nationals were seen to be the source or at least the aggravating factor of the poverty, lack and scarcity that is the daily experience of millions of South Africans however now the focus it seems to have and continues to shift to the State and its failure to deliver and help its people. Another indicator of this is the *Perception of corruption* measure in the Happiness Index, which is significant high in South Africa (the numbers are included in Appendix 1). These positions are strengthened and made clearer by and through some of the related hashtags, some which are illustrated below:



#WeWantOurCountryBack

#LandIsOurDignity

#TogetherUnitedSAns

#Xenophobia

#ForeignersGoHome

#BeitBridge

Figure 4: Top 6 hashtags that trended alongside #PutSouthAfricaFirst

Identification of linguistic and rhetorical mechanisms:

During analysis of the tweets, in both periods, we found there to be a kind of golden thread that runs through the statements made. The challenge is in "how' the sentiment is brought across using language, the overall message however is very clear. South Africans are grossly unhappy with the state of their country, in terms of service delivery, their living conditions and the overall degradation of state institutions pointing to corrupt and irresponsible politicians and officials who seem to be unbothered by the rapidly deteriorating state of affairs.

The position on which individuals differ is the direction in which the blame should be placed, and it is on this point we can divide our Twitter users into two sub-groups, reiterating the ideas we discussed above of the Pro-South African and anti-immigrant sentiment (which we explore further in the sentiment analysis below).

In a case where one sides argument is dominant it would be much easier to assess the argument as the grammar features (an extension of the linguistic context we discussed above)- the use of words like *us*,

them, we and they – would point in one direction but in our case, we must make a distinction. In our investigation we investigated the use of these words used by locals (the in group) to refer to immigrants (the out group)

Sentiment Analysis

The sentiment analysis process was carried out using RapidMiner. The platform has an inbuilt sentiment analysis function to extract sentiment from text however we have chosen not to use it but to manually build a custom process from scratch and individually hard code the sentiment of all the tweets in the stratified sample we extracted. The sample will work as a learning tool for the built process. The building process is outlined below as well as our reasons for designing the process as opposed making use of the automated one. The same process was carried out for both time periods.

Using the sample data, we manually hard coded the tweets individually as positive, negative or neutral (which included the unrelated posts as well) in a column called "Sentiment". The Sentiment attribute is then set to the role of attributing labels to the Tweets. We then imported the data into the process.

The first step is conducting the sentiment analysis on the training model (i.e., the sample data) that we will use for the larger data.

We wanted to extend our analysis beyond just the standard positive, negative, neutral sentiment to extract sentiments of the key themes we had isolated during the Discourse Analysis. In the sentiment process we included steps to generate and extract the following attributes (which are given in detail in Appendix 1):

1. Security

2. Identity

3. The State

4. Ownership

5. Jobs and employment

6. Political

The Naives Bayes model is then applied to the data for the sentiment analysis. Naives Bayes was chosen as it does not require a great volume of training data, the created representative sample is sufficient for training the model. In addition, it is able to handle both continuous and discrete data and greatly reduces the processing required.

The steps as applied to the smaller dataset, the training model, are passed down into the main dataset set through a set the multiplier. The performance of the model, or its accuracy lies in the coded training model. Spot checks are conducted on the output and any incorrectly assigned instances are manually added to the training model and the process is rerun, to improve the performance of the model.

Challenges:

We chose to hard code the sentiment of the sample because text we are working with is especially complex due to language use on the Twitter platform. Twitter users tend to use shorthand text language as well as use a combination of languages within a single tweet. We have taken steps to cater for this however there are instance where a non-English or uniquely South African words or phrases are used. The topic we are working with makes it a little more complicated particularly determining what is negative or positive. We also want to have more control over the meaning of what is positive, negative and neutral as opposed to relying on RapidMiner's interpretation, to achieve this we built the process outlined above. It is important to mention that we have taken positive to include "pro-South African" stances or sentiments.

We encountered several tweets that were completely unrelated to the topic. Examples such as individuals using a popular hashtag to advertise something they are selling or in a bid to attract to more followers for their personal profiles. We had taken measures to filter these out in the initial stages however owing to the size of the data, there are a few tweets that remained. Many of these were captured as neutral.

Results and Interpretation

We have explored numerous theories in this paper, many of which have been thought to offer at least a part in explaining xenophobic or anti-immigrant belief systems that exist in our societies. We explored the interactions and conversations on the Twitter platform to investigate if there is any truth to these ideals. Hearing what participating individuals have to say is the perhaps our best attempt at getting insight into peoples' thought process and the belief system that informs their thinking.

We started off with an extracted a list of what we expected to be prevalent themes as we coded the material and set out the context. The themes and context we extracted at the end did not deviate from the expected instead it expanded the scope.

The voices that dominate this conversation seems to those faced with a daily struggle for basic resources and find themselves in direct competition with our foreign neighbours seeking relief form the pains of their homelands. Previously we saw frustrations result in bloody battles in the streets where the foreigner was labelled the cause for the deprivation that plagued the people but now, we see the conversations play out in the *Twitter Streets* where we can hear them voice their complaints. This time also directed at the government who they fault for being greedy, corrupt, and incompetent. Perhaps 27 years is too long a time to expect people to wait for a "new democracy's" teething problems to be resolved. The people also seem to have changed their perception of the foreigner, although the outgroup still exists, the State is implored to prioritise its own before it can look to assist its ailing neighbours while the gap between the haves and have nots stretches ever wider. Not only has social media given a platform for freedom of expression and to be heard but it has also made the world smaller, creating a global community. We see, share, and experience each other's worlds highlighting the vast differences between nations, neighbours and friends. Relative deprivation has never been more obvious.

Our initially investigation into power, violence, institutions, resources, deprivation, and social identity was informed by previous literature as well as personal thoughts. Although they seemed to make logical sense, we had no way of making sure, no way of "hearing" the thoughts of participating individuals to point us to the right track. That was until now. We initially set out to expand on what we know about

xenophobia and anti-immigrant notions but also to attempt to make a unique contribution to 'new school' methodologies. Making use of Twitter data gave us the listening device we needed to possibly delve into people's thought processes and really get into their heads and see where these sentiments are coming from.

Sentiment analysis results and interpretation:

Jul/Aug Sentiment		
	Count	%
Positive	3201	21,71
Neutral/unrelated	1769	11,99
Negative	9778	66,3
Total	14748	100

The results of the sentiment analysis for the July-August dataset are shown in the tables above. The positive tweets amounted to a 21,71%, we had included the 'pro-South African' sentiment into this category. These were also the earlier days of the movement, a lot of the discussions around the hashtag were about how it seemed to a movement that perpetuated xenophobia and how freedom of speech was being used as an excuse to abuse foreign nationals on a platform that affords individuals anonymity.

Dec/Jan Sentiment		
	Count	%
Positive	7020	39
Neutral/unrelated	1746	9,7
Negative	9234	51,3
Total	1800	100

The results of the sentiment analysis for the December-January dataset are shown in the tables above. The positive tweets amounted to 39%. During the time of the *Beitbridge saga* many of the positive tweets pointed towards better border controls and documentation systems as well as increased accountability on the part of officials particularly moving towards the roll out of the Covid-19 vaccine.

The negative tweets accounted for just over half of the total number of tweets (51,3%). These shifted the focus in a different direction from the officials, towards the immigrants themselves raising issues

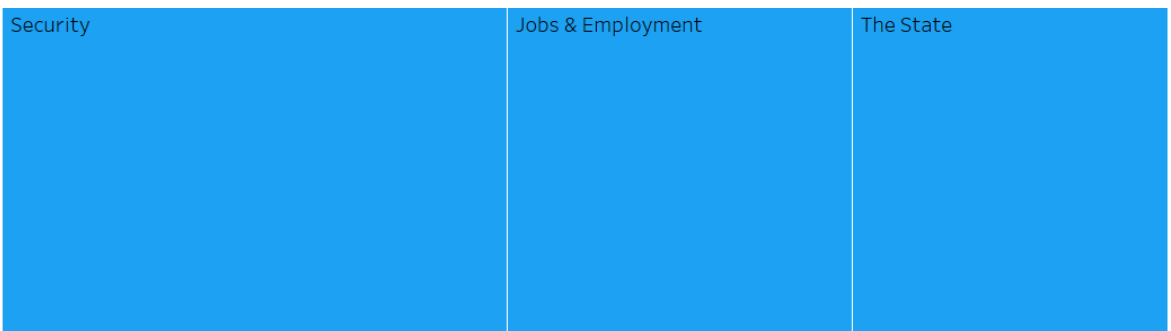
such shock at the vast number of individuals at the border post as well as the presenting of the fake Covid test results.

In both sets we also had a small number of neutral or unrelated posts. This was expected as owing to the nature of the topic as well as the platform, it is unlikely that an individual would take to Twitter without having a particular standpoint on the topic. The unrelated tweets are mostly made up by individuals riding the wave of a popular hashtag to get as many people as possible to see their post either for advertising purposes or for "clout"- to request likes, retweets or follows.

The other attributes were created to extract the prevalence of the themes we have explored thought the paper. The results below show that we are possibly on the right track in explaining the underlying causes of ethnic tensions, xenophobia and anti-immigrant beliefs in our societies.

Themes	No of tweets	%
Security	329	50,08
Identity	132	20,09
The State	257	39,12
Ownership	174	26,48
Jobs & Employment	263	40,03
Poltical	293	44,6

The results are based on the sample set of 657 tweets. The number of tweets and percentages illustrate the number of tweets that addresses or alluded to each of the themes (outlined in the Appendix). Each of the themes in heavily represented throughout the dataset showing how prevalent these issues are and how important they are to the participating groups. When we started out, we were not sure if these topics would contribute to the discussion, but we see that they do and significantly so. The literature addresses these points, and we now see that individuals are doing the same in their conversations and interactions.



Conclusion

The task we undertook saw us delve deeper than we anticipated. We covered a number of theories, each rich, that contributed significantly to gaining a better understanding of the individuals participated in the broader conversation, their thought process and belief system. We called on social media, Twitter in particular, to help us comb through the conversations and hear what is being said.

Theories relating to the dynamics between violence, the state and power, the rise or formation of the state and the provision of resources, shaped our process in analysing the discourse around the #PutSouthAfricanFirst movement to help us move closer to the ideas fuelling the movement aided by discussions regarding of relative deprivation, resource mobilization and the ideas relating to social identity. These points came across strongly in the interactions among individuals and helped us to understand the reason behind the shift from physical violence on ongoing dialogue, as the focus itself shifted from placing blame on the individuals that have crossed borders seeking safety, security, and prosperity to placing on the shoulder of those elected to prioritise, protect and serve those who have placed them in those very seats of power.

In the Sentiment Analysis we moved beyond just the positive, negative, and neutral ideals but looked deeper movement whose purpose, direction and intentions are hotly contested and often unclear. We

looked at the occurrence of the main themes in the discussions and found that every one of the themes were explored had a place in these conversations. The literature calls on these themes to offer an explanation for these occurrences and now we know that individuals are discussing the same topics in their conversations and interactions. These inform their thought processes and resultant behaviour and owning to the cyclical relationship we had alluded to earlier, also feedback into the general belief system of our societies. The nature of the discourse is continuously and rapidly changing, more so than ever due to the rate at which these notions are shared, reshared and contested on social media platforms.

Using social media data was in itself an ambitious undertaking. As much as we sought to unpack the dynamics between South African and foreign nations and their interactions, we also made an attempt to contribute to the body of knowledge as a whole through exploring new methodologies and data sources. Although we admit that the processes are not yet perfect, they have shown us what is possible and through experiencing the many challenges and pitfalls shown us what needs to be improved upon and refined. However, we now have a means of hearing the true perceptions and opinions of individuals and groups of people in a setting closer to ‘living room conversations’ than ever before.

Appendix 1: Additional data

South Africa's Happiness Index:

Year	Ranking	Score	GDP per Capita	Social Support	Health-life expectancy	Freedom to make life choices	Generosity	Perception of Corruption
2020	109	4.814	0.902	1.259	0.407	0.435	0.126	0.060
2019	106	4.722	0.960	1.351	0.469	0.389	0.130	0.055
2018	105	4.724	0.940	1.410	0.330	0.516	0.103	0.056
2017	101	4.829	1.055	1.385	0.187	0.479	0.139	0.073

South Africa's happiness Index numbers over the last 5 Years as sourced from the Happiness Index reports for the years 2017-2020. The numbers were extracted from the individual reports and recorded in the table above

Additional Details on the dataset contributed by Mr Findlay:

The dataset received from Kyle Findlay, after we had it through the ETL process consisted of the following attributes:

- created at (date),
- created at (time),
- language,
- mentions,
- user friends count,

- user. location,
- user.name,
- user. screenname,
- tweet type

and a number of other columns specifically related to tweets that are either retweets or replies to a tweet.

These are not important to us as we only made use of the isolated tweets, not mentions or retweets.

Collapsing keywords and themes into an Operational list for Discourse Analysis:

Having looked at each of the individual tweets to identify the topics addressed by the user, we complied the list below of themes and keyword:

- Identity
- Citizenship
- Property
- Possession
- Language
- Offence
- Jobs And Employment
- Borders
- Safety
- Opinion
- Position
- Crime
- Politics
- Land
- Illegal Immigration
- Power And Institutions
- Influence
- Citizen Rights
- History

The keywords were collapsed into broader groups as follows:

1. Security

1. Crime
2. Border control
3. Safety

2. Identity
    a. Citizenship
    b. Language
    c. Citizen rights
    d. History
3. The State
    a. Influence
    b. Power
    c. Institutions
4. Ownership
    a. Land
    b. Possession
    c. Property
5. Jobs
    a. Employment
    b. and social position
    c. Job
6. Political.
    a. Malema
    b. Mashaba
    c. Ramaphosa
    d. Julius
    e. EFF
    f. ANC

Appendix 2: Licensing restrictions and limitations

DATABASE OF TWEETS RECEIVED FROM KYLE FINDLAY:

A database of 321 953 raw tweets related to #PutSouthAfricansFirst was received from Kyle Findlay following the signing of a Memorandum of Understanding (MoU) in which the following was agreed to:

- In final thesis, individuals are not to be personally identified. The identities of non-public individuals are to be protected.
- The raw dataset is not to be shared with a third party.
- Raw outputs will not be shared with a third party.

DEVELOPER TWITTER:

Requested access tokens and consumer tokens for creating the authentication within the script for the Twitter API to create the crawler to extract tweets directly from Twitter.

As per the agreements:

- Display the Twitter mark to indicate the source of the data.
- To not publish any individual tweets showing the username of an individual
- Pass on the extract content to a third party for use or storage purposes.

- To maintain the security of the Twitter API and will not make available to a third party, any tokens, keys, passwords or other login credentials to the Twitter API or Twitter Developer account
- Explicitly agreed that none of the content accessed would be used for commercial purposes.
    - In the case where a tweet has been displayed in the paper, it is the real, unmodified Tweet from the real account.
- In order to comply with the accepted Terms and policies in order to obtain the license. The data has been used as an aggregate analysis of Twitter content. The usernames, user descriptions and other identifiers we removed from the data before any processing began. These were not stored. For the purposes of coding the data in the Discourse Analysis process, unique in order to create hierarchies and extract required words and phrases. No personal data associated with the content of the Tweets has been stored.

LIMITATIONS OF USING TWEEPY:

Using the access tokens and keys granted through the license, Tweepy (the Python library) was used to access and extract the tweets.

- Following the Developer agreements and having been granted a license for the standard API. There are limits to the content that can be accessed in a given time period. Allowing the bot retrieve tweets up to 7 days ago and is limited to scraping 18,000 tweets per a 15-minute window.

RAPID MINER EDUCATION LICENSE:

A RapidMiner education license has been granted. The only license request being that RapidMiner is to credit for the steps and processes for which it has been used.

TRENDSMAP PRO TRIAL:

The trial was applied for and granted for educational/research purposes. It did not carry any additional requirements.

TABLEAU:

A student licence was requested and granted. It did not carry any additional requirements.

References

Brush, S., 1996. Dynamics of Theory Change in the Social Sciences: Relative deprivation and collective violence. *Journal of Conflict Resolution*, [online] 40(4), pp.523-545. Available at: <https://journals.sagepub.com/doi/abs/10.1177/0022002796040004001> [Accessed 23 July 2020].

Choane, M., Shulika, L. and Mthombeni, M., 2011. *An Analysis of The Causes, Effects and Ramifications of Xenophobia in South Africa*. [eBook] Africa Insight. Available at: <https://www.researchgate.net/publication/282878743_An_Analysis_of_the_Causes_Effects_and_Ramifications_of_Xenophobia_in_South_Africa>

Douglass C. North, John Joseph Wallis and Barry R. Weingast, 2008. Violence and the Rise of Open-Access Orders. *Journal of Democracy*, 20(1), pp.55-68.

Gordon, S., 2016. Understanding Evaluations of Foreigners in Modern South Africa: The Relationship Between Subjective Wellbeing and Xenophobia. *Journal of Happiness Studies*, [online] Available at: <https://link.springer.com/article/10.1007/s10902-016-9838-6> [Accessed 8 August 2020].

Herrera, Y. and Butkovich Kraus, N., 2016. Pride Versus Prejudice: Ethnicity, National Identity, and Xenophobia in Russia. *Comparative Politics*, 48(3), pp.293-315.

Jenkins, J., 1983. Resource Mobilization Theory and the Study of Social Movements. *Annual Review of Sociology*, [online] 9(1), pp.527-553. Available at: <https://www.jstor.org/stable/2946077>.

Korpi, W., 1974. Conflict, Power and Relative Deprivation. *American Political Science Review*, [online] 68(4), pp.1569-1578. Available at: <https://www.jstor.org/stable/1959942?seq=1> [Accessed 23 July 2020].

Landau, L., 2020. *Opinion | What's Behind the Deadly Violence in South Africa?* [online] Nytimes.com. Available at: <https://www.nytimes.com/2019/09/16/opinion/south-africa-xenophobia-attacks.html> [Accessed 11 July 2020].

McCarthy, J. and Zald, M., 1977. Resource Mobilization and Social Movements: A Partial Theory. *American Journal of Sociology*, [online] 82(6), pp.1212-1241. Available at: <https://www.jstor.org/stable/2777934>.

Roche, R., 1996. Collective violence as social control. *Sociological Forum*, 11(1), pp.97-128.

Sanchez de la Sierra, R., 2013. On the Origin of States: Stationary Bandits and Taxation in Eastern Congo. *SSRN Electronic Journal*.

Sen, A.K., 1987. Poverty and entitlements. In *Food policy* (pp. 198-204). Johns Hopkins University Press Baltimore.

Stark, O. and Taylor, J., 1991. Migration Incentives, Migration Types: The Role of Relative Deprivation. *The Economic Journal*, 101(408), p.1163.

Vearey, J. and Walker, V., 2020. *OP-ED: In Sickness and In Health: Why the Minister of Home Affairs Has a Duty of Care to Tackle Endemic Anti-Foreigner Sentiment*. [online] Daily Maverick. Available at: <https://www.dailymaverick.co.za/article/2019-09-26-in-sickness-and-in-health-why-the-minister-of-home-affairs-has-a-duty-of-care-to-tackle-endemic-anti-foreigner-sentiment/?#gsc.tab=0> [Accessed 12 July 2020]

Weller, J. and Quarantelli, E., 1973. Neglected Characteristics of Collective Behaviour. *American Journal of Sociology*, [online] 79(3), pp.665-685. Available at: <http://ttps://www.jstor.org/stable/2776263>.

Yitzhaki, S., 1979. Relative Deprivation and the Gini Coefficient. *The Quarterly Journal of Economics*, [online] 93(2), p.321. Available at: <https://www.jstor.org/stable/1883197> [Accessed 22 July 2020].