A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Water Crisis and Tourism in Cape Town: Analysing the City's Destination Image Using Social Media Mining and Natural Language Processing

Costantino Scammacca del Murgo
55501

Work project carried out under the supervision of:

Leid Zejnilovic, Lénia Mestrinho

09/01/2024

Abstract

This study aims to develop a reliable system to collect, clean and classify social media data according to affective attributes of the destination image. Furthermore, it aims to analyze the evolution of the perception of Cape Town by the community of South African and international tourists writing on the TripAdvisor forum during the period of the water crisis. The results reveal significant changes in five of the eleven defined affective attributes, with a relevant decrease in perceptions of safety, cleanliness, attractiveness, comfort and enjoyment. Furthermore, it emerges that international tourists generally have more positive perceptions of the destination than domestic tourists.

Keywords: Destination Image, Water Crisis, Cape Town, Social Media Mining, Natural Language Processing

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Contents

Introduction

In an era where the global tourism industry faces unprecedented challenges and intensified competition, with an increasing percentage of tourists visiting a smaller number of destinations, it becomes crucial for these destinations to measure and monitor the perceptions of both domestic and international tourists. These perceptions can be significantly influenced by current events and crises. In this context, this thesis explores the impact of the severe water crisis experienced by Cape Town, South Africa, between 2015 and 2018, evaluating the image of this destination as perceived by the TripAdvisor community. This analysis focuses on the perspectives shared by local and international tourists on the Western Cape forum.

The concept of 'destination image' is a complex subject, but researchers agree that it refers to the combination of perceptions that individuals have of tangible and non-tangible features of a destination. These perceptions are crucial in the decision-making process of potential tourists, and many studies have found that it is mainly negative news that influences them the most. Destination image is a dynamic construct, constantly evolving in response to current events that gain media attention. It is subjective and depends on how individuals process this news based on their personal experiences and background, so it is natural that different perceptions develop among different groups of tourists. Precisely because of its dynamic nature, the image of a destination can be treated as a corporate brand, and resources should be invested in monitoring and improving it.

Over the past 15 years, with the rise of social media platforms and forums where individuals can freely express their opinions, vast pools of unstructured data have been generated. Many researchers have explored the possibility of using this data to conduct analysis in the tourism sector, validating this hypothesis and highlighting the importance of understanding the motivations behind individuals' willingness to express their opinions on platforms of various types. For instance, TripAdvisor, Expedia and, to some extent, Google Maps, guide users to review restaurants, hotels and tourist attractions, while Facebook or Twitter offer spaces for discussions on a wide variety of topics. However, even when the most suitable platform for the desired analysis is chosen, previous studies indicate that these data pools are often rich in 'noise' that needs to be filtered to identify the most useful information.

The main objective of this research is to develop a methodology to collect, process and categorize textual data from TripAdvisor based on various affective attributes of the destination image. Second, the thesis seeks to analyze the sentiment associated with these reviews. This analysis aims to highlight any changes in the perceptions of both local and international tourists towards the affective attributes of Cape Town's destination image during the water crisis. Additionally, it tests the hypothesis that the three-year-long drought had a negative impact on the perceptions of tourists visiting the city.

For this purpose, an in-depth analysis was conducted on approximately 120.000 tourist contributions published on TripAdvisor from 2012 to 2019. These contributions were extracted using popular web mining packages in Python. The classification and filtering of the reviews involved the use of Python keyword dictionaries and the GPT-3.5 Turbo model, accessible via the OpenAI API. For sentiment analysis, the BERT-base model was employed. T-Tests were performed to compare the average sentiments associated with each attribute and to identify significant changes over time and between different groups of tourists. The aim was to identify reviews that mention specific attributes of the destination image and to map the evolution of tourists' perceptions of these attributes the 4 years before and throughout the water crisis.

Literature Review

Destination Image Definition

In academic research on tourism, the concept of destination image refers to the different ideas and impressions that people hold about tourist destinations. Some experts believe it's a combination of overall opinions and assessments [1], while others think it's a collection of specific features described using descriptive words [2]. There is a marked distinction between considering the destination image as an evaluative judgement or as a descriptive set of qualities, yet these points of view are often merged into the same general term. Furthermore, there is debate as to whether this image is a cognitive (knowledge-based) or an affective (emotion-based) representation of the destination [3].

According to several academics [4][5][6][7], the destination image is a collection of all the concrete information, beliefs, ideas and emotions that individuals or groups associate with a specific location. This image blends observable features, such as scenery, attractions and accommodation, with less tangible elements, such as the perceived atmosphere, safety and hospitality of the destination. These perceptions are highly subjective and may be perceived differently by different individuals. Moreover, they significantly influence people's travel choices, travel behavior and even wider perceptions of a country's attractiveness for tourism, business, and international relations [8].

It is therefore crucial for tourism authorities of various destinations to craft an image that positively distinguishes them from their competitors, blending measurable attributes with holistic ones. To achieve this, tourist destinations often promote recognized functional and psychological characteristics of the destination, while others highlight the special features or experiences that make them unique.

Destination Image Relevance in the Tourism Industry

According to the World Tourism Organisation, 2018 saw an impressive 1.4 billion tourists visiting new destinations [9]. Despite the slowdown caused by the pandemic, the outlook for global tourism is very optimistic, with international arrivals in the first quarter of 2023 returning to 80% of their pre-pandemic levels. This recovery is particularly evident in the Middle East, Africa, and Europe, exceeding the global average [9]. The influx of tourists has consolidated the role of the tourism sector as a key economic driver and a significant mover in investments in new infrastructure [10]. Tourism is reported to account for more than 15% of the GDP of many European and world countries [11]. This sector is particularly crucial for smaller and developing nations. For example, in Africa, the GDP of Seychelles, Cape Verde and Mauritius owes 62%, 43% and 27% respectively to tourism [12]. The industry is also remarkable for its promotion of women and youth, with women making up about 65% of the workforce and half of the workers being 25 years old or younger. This is evident in Africa, where a higher percentage of women hold managerial positions in hotels and restaurants (31%) than in other sectors (21% overall) [12].

Simon Anholt compared the destination image of countries to the image of products and companies, suggesting that a strong, positive reputation, similar to corporate branding, can provide a competitive advantage and lead to greater success [13]. The tourism industry has grown considerably, with countless destinations competing for travellers' attention. Each year, the majority of international tourists - around 60 per cent - visit only a select group of 10 countries, leaving a vast number of other destinations to compete for the remaining 40 per cent of travellers. This reality underlines the intense competition and disparities within the tourism industry [14].

This competitive landscape has led local and national governments to invest in the tourism sector to promote their destination image. Several tourism destination promotion campaigns showed positive results. For instance, the ‘‘Visit Denmark” campaign reported a return of $16 in revenue for each dollar spent. Similarly, Tourism Ireland observed a 10 percent return from its television and online advertising efforts. Toronto's $4 million investment in rebranding led to a 26 per cent increase in international visitors. Dubai, on the other hand, has invested billions over the last 16 years, successfully establishing its image as a luxury destination [15].

With the growing importance of destination image in consumer decision-making, understanding how these images are formed becomes crucial. Such knowledge can assist destination marketers to shape an image that resonates with their target markets [16]. It is crucial for destination managers and marketers not only to project an attractive image, but also to ensure that they can maintain this image and exceed tourists' expectations to generate satisfaction and trust towards the destination [7]. Clearly, sustaining a distinctive destination brand and image requires significant resources.

Destination Image Characteristics and Components

Gallarza, Saura and Garcia explored the different elements that influence the destination image, describing it as complex, multiple, relative and dynamic [17].

The destination image is complex: although there is a general consensus that it constitutes the sum of the impressions a person holds about a location, the specific elements that form this impression are debated. Some studies emphasise cognitive factors (such as architecture, culture, natural environment, tourism infrastructure and the quality of services offered [18]), while others include both cognitive and affective elements (emotional reactions, such as the joy, excitement, or serenity) that a destination evokes [16]. Further research suggests that conative attributes, which are linked to the desire and intention to visit, also play a crucial role in shaping the destination image [17][16].

The destination image is multiple; it includes a number of specific attributes and components, such as physical characteristics, cultural offerings and available experiences. Nevertheless, it includes also the holistic perception, that each individual develops towards a destination, that cannot be fully described or understood by analysing only the individual elements that make it up. This holistic perception reflects the overall experience and emotional impact a destination has on an individual [17].

The destination image is dynamic, influenced by time and space. Time plays a crucial role, as individuals require time to form their own perceptions of a destination's image. Geography and space also significantly influence these perceptions. It has been observed that the destination images held by different groups, such as local and international tourists, are often inconsistent [19]. In this context, current events, whether positive, negative, or contradictory, can impact the perception of a destination image and, consequently, potential tourists' decision-making processes [4] [20]. Crises arising from geopolitical, natural, or financial events can rapidly and significantly damage a destination's image. This effect is often compounded by the media's tendency to emphasize unfavorable events, which can further discourage tourism [21]. These crises can be categorized into short-term, medium-term, and sustained crises, each affecting the destination image in different ways [22][23][24].

Finally, perceptions of a destination are relative and vary from one individual to another. In fact, each tourist's individual perception is influenced by their own experience as well as social and demographic factors. This dynamic and relative nature of the destination image is crucial for tourism marketing, as perceptions can be influenced and managed strategically across groups of individuals [17].

To effectively measure and understand both the attributes-based and holistic components of the destination image, past studies have employed both structured and unstructured research methods. These methods aim to capture the full range of tangible and intangible qualities that constitute a destination's image. Table 1 presents a list of cognitive and affective attributes that have been used to analyze the destination image.

Cognitive	Expenses/Price Levels ; Weather ; Tourist Attractions /Activities ; Nightlife and Amusements; Sports Facilities/ Events ; National Parks/Nature Adventures ; Regional Infrastructure/Transportation ; Beaches ; Shopping centers ; Accommodation Facilities ; Fairs, Exhibits, Festivals ; Centers for Information and Tours [4]
Affective	Crowdedness ; Cleanliness; Security, Economic Prosperity/Wealth ; Accessibility ; Urban Development; Extent of Commercialisation ; Political Stability ; Hospitality/Friendliness ; Diverse Traditions/Heritage ; Unique Gastronomy/Beverages ; Restful/Relaxing ; Atmosphere [4]

Table 1: examples of cognitive and affective attributes of the destination image

However, the existing literature has focused more on developing structured research methods which allows for a replicable though certainly not exhaustive statistical investigation. This study also utilises this methodology.

Cape Town Water Crisis

The episodes of extreme drought that Cape Town experienced between 2015 and 2018, which directly generated the water crisis, were part of an era of decreasing rainfall over the years. This coincided with one of the most extreme El Niño weather phenomena - associated with warmer and drier weather conditions in southern Africa - and was worsened by the increase in population density in Cape Town.

Examining rainfall in Cape Town over the past from 1993 to 2018, it appears that drought conditions have occurred in 10 of these years, half of which occurred in the last decade. Throughout this 25-year period, the average annual rainfall was approximately 572 millimeters in the first 10 years, decreasing to 423 millimeters in the last five years. Academics agree that this was the worst drought since 1904 and in recent years [25][26][27] and it led to severe domestic and commercial water shortages, disrupting the lives and safety of Cape Town's citizens and visitors [28].

To address the drought and reduce water demand, the City of Cape Town implemented special measures. These included restrictions on water consumption and an increase in water costs, which ultimately led to the declaration of a local state of calamity in March 2017. It was during this period that the term 'Day Zero' was introduced into official discussions, indicating the imminent depletion of water reserves. The use of the term Day Zero was of great importance, as it sent a message of fear and panic to Cape Town residents and potential visitors. Fortunately, heavy rains in March 2018 allowed the 'Day Zero' alert to be postponed and eventually cancelled in June. This turnaround was due to the accumulation of water during six consecutive weeks of rain, coupled with effective resource management, enabling the city to secure sufficient reserves for a full year [29][30]. Appendix 1 provides a clear visualization of the water supply and demand in Cape Town between 2008 and 2017, illustrating the positive effects of the city's water resource management to avoid Day Zero.

Cape Town Water Crisis Impact on Tourism

The drought in Cape Town attracted the attention of local and international media. It coincided with one of the most severe El Niño-induced droughts in southern Africa and two historical events that could have contributed to a significant media attention: the 2015 Paris Agreement under COP 21 and the 2030 Agenda for Sustainable Development. In this context, politics also played its role. On one side the opposition party, the Democratic Alliance (DA), ruling the Western Cape province, and on the other side the national government, run by the African National Congress (ANC), did everything to gain political advantage from the drought [30].

Not unique to Cape Town, droughts have affected other major countries and destinations in the world such as Kenya in 2017, Morocco in 2018, Australia and Spain between 2008 and 2017. However, none of these countries adopted the communication strategy used in the Day Zero approach. This image of Cape Town as a city in crisis may have damaged the perception of the destination, leading tourists to postpone or cancel trips to what was one of the world's most popular destinations.

Although the absolute number of tourist arrivals in the Western Cape was not affected by the water crisis, as it continued to grow from 2014 with 1.279.000 arrivals until 2019 with 1.922.000 arrivals, the drought had an impact on tourist arrivals in Cape Town. In particular, at its peak between 2017 and 2018, it led to a decline in arrivals. This consequently affected potential tourism generated revenues for the city and the province. A study observed that the dry period was characterised by low growth rates in tourist numbers at the most iconic tourist attractions in the Cape Town area [31]. For April 2017-18, the tourism industry recorded a 12,6 per cent decline in arrivals from overseas, with further declines of 3,7 per cent and 1,3 per cent observed in May and June respectively [32]. This has led to a decrease in hotel occupancy and tourist arrivals at the city's most popular tourist sites such as the V&A Waterfront, Table Mountain cable car, Kirstenbosch Botanical Gardens, Cape Point and other destinations. Another interesting fact is that tourists were booking fewer nights than in previous years: in 2017, the average length of stay was 14,1 nights, a figure that dropped to 12,9 nights in 2018. Also the average spends per trip by international tourists decreased by 9% between 2017 and 2018.

Appendix 2 offers an in-depth analysis of the impact of the water crisis on local tourism-related businesses based on a report from the Tourism Council of South Africa and another report from Wesgro.

Methodology

This study aims to develop a comprehensive system for collecting, cleaning, and categorizing data from TripAdvisor. The focus of this categorization is on the affective attributes of the destination image, transforming semi-structured data into a structured format. Additionally, the study analyzes the sentiment of the reviews. This involves mapping the evolution of perceptions related to each affective attribute of the destination image among different groups of tourists, both domestic and international, who planned a trip to or visited Cape Town between 2012 and 2019.

The methodology of this study integrates social media mining and natural language processing techniques and consists of five phases: (1) Data Collection and Cleaning; (2) Exploratory Data Analysis; (3) Data Classification, leveraging keywords and GPT-3.5 Turbo; (4) Sentiment Analysis, employing the BERT model; and (5) Sentiment Mapping, with reference to the attributes of destination image, time, and the origin of tourists. Appendix 3 presents a flowchart detailing the main steps of the methodology.

Data Collection

TripAdvisor is a collaborative platform that allows users to gather information and share their travel experiences in forums. The platform was founded on the premise that tourists planning a trip highly value the opinions of other travelers, which significantly influence their decision-making process. A report by PhocusWire indicates that over 80% of potential travelers read between 6 and 12 online reviews before making accommodation choices, typically focusing on the most recent reviews [35]. Both positive and negative reviews have a significant impact on tourists' perspectives, with users tending to trust negative reviews more than positive ones [40]. Several studies have recognized the potential of using TripAdvisor for local tourism planning

[36][43]. However, many also point out the considerable amount of 'noise' in the data, highlighting the need for robust methods to classify and filter textual data [44].

Using a Python script based on the 'Selenium' and 'Beautiful Soup' libraries, reviews from the 'Cape Town' and 'Central Cape Town' forum sections on TripAdvisor were collected [37]. This process involved obtaining a list of all topic URLs, downloading the HTML code of each web page to retrieve general data and reviews. The data collected included the forum, the topic, the review title, the review date, and the review body. To protect the privacy of data subjects, the data were pseudonymized during the collection process. Direct identifiers, such as usernames and user pictures that could easily be traced back to individuals, were deliberately omitted. Additionally, all data cleaning steps applied to the textual data (referenced in the following paragraph) were performed during the data collection phase, before saving the final CSV file. Fig. 1 illustrates how some of these data are displayed on the TripAdvisor Forum.

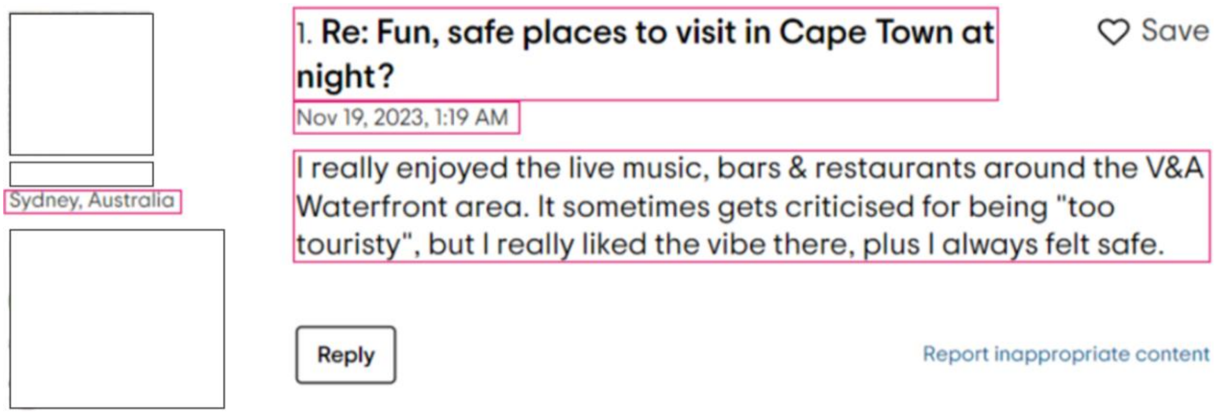


Fig. 1 Example of review on TripAdvisor Forum, highlighting the data extracted.

In total, 123.664 contributions were collected, covering the period from 2010 to 2023. These reviews offer insights from both domestic and international tourists who have visited Cape Town. All reviews were in English, as the data were collected from the South African forum for English-speaking countries. For the second part of the analysis, which examines the evolution of the affective attributes of the destination image across different groups of tourists (South African and non-South African), only contributions from users who stated their residency were included. This analysis takes into account information about residency as stated in the TripAdvisor users' profiles [61].

Appendix 4 provides a more detailed explanation regarding the quality of TripAdvisor data and the rationale for choosing it as the only data source.

Data Cleaning

In the initial phase of the analysis, the textual data were carefully prepared to ensure their quality. The preprocessing included several key steps to refine the dataset:

1. Initially, any duplicate entries were removed to maintain the uniqueness of the data.
2. To achieve uniformity, all reviews were converted to lower case, ensuring consistency throughout the dataset.
3. Stop words, commonly used words that offer information of minimal value for analytical purposes, were removed from the reviews.
4. Any numbers and spaces were removed to simplify the text.
5. All punctuation marks and hyperlinks were removed from the text.
6. To make the best use of the GPT-3.5 Turbo API, reviews with more than 750 tokens (about 600 words) were excluded.

These steps were essential to ensure that the textual data were properly prepared for analysis, and to minimize the risk of traceability for data subjects.

Data Classification Based on the Affective Attributes of Destination Image

Several methods for reviews classification have been developed. These techniques range from keyword-based methods to complex machine learning models. The methodology of this study benefits from both these approaches by implementing a two-phase classification process. The first phase is keyword-based, drawing inspiration from a study that analyzed the destination image of Athens during the Greek economic recession [45]. The second phase utilizes the GPT-3.5 Turbo API.

Keyword-based classification is straightforward and intuitive. It operates like a scanner, sifting through vast amounts of reviews to detect any of the predefined words. Whenever a text contains one or more of these predetermined words, the system flags it. For analyzing Cape Town's affective destination image, a list of affective attributes related to the city's Destination Image is compiled based on the existent literature [4][45]. These affective attributes include the Perceived Safety; Perceived Development; Perceived Cleanliness; Perceived Hospitability; Perceived Value; Perceived Attraction; Perceived Relax; Perceived Comfort; Perceived Authenticity; Perceived Excitement; and Perceived Enjoyment. Appendix 5 presents the full list of attributes and keywords used.

For each attribute, a set of keywords is chosen, and a list of synonyms and antonyms are defined for each keyword. All the keywords where sourced from Merriam-Webster, a reputable dictionary publisher, and Thesaurus, a popular website offering synonyms and antonyms. Based on the words defined, a Python dictionary is created and used to iterate through all the reviews under analysis and assign them an affective attribute based on the recurrence of the words in the dictionary in the text of the review. Reviews that feature more than one of the words defined in the dictionary fall into more categories of the affective attributes of the destination image, while those that do not present any word will be considered out of scope [45].

One of the strengths of dictionary-based approaches for text classification is their ability to achieve impressive precision, as they rely on co-occurrences of specific terms. However, a major limitation is their inability to detect intricate content correlations, which compromises the efficiency of categorization. The keyword-based classification labeled 11.488 reviews containing terms related to the attributes of the Destination Image.

However, a manual evaluation of a sample from the dataset revealed that some of the reviews considered relevant to destination image attributes were actually about private products or services in Cape Town. These reviews did not reflect the tourists' perceived image of the city and its hotspots.

To minimize noise in the dataset, an AI-based filter was implemented. This second phase involves using the OpenAI API to further categorize the reviews using advanced ML models, avoiding the need to develop one from scratch. OpenAI offers a range of specific models for different purposes. Some, such as the popular GPT 3.5 Turbo – chosen for this study - are capable of handling a variety of tasks. These models are accessible via an API key issued by OpenAI after creating an account and can be integrated using Python and other programming languages. To verify the performance of the classification based on GPT 3.5 Turbo, 200 random reviews were assessed manually. Reviews relevant to the attributes of the Cape Town Destination Image were marked with a 1, and those that mentioned the keywords but were not relevant were manually marked with a 0. A confusion matrix was created to assess the performance of the GPT-based classification, resulting in the following key indicators: the model demonstrated strong recall performance (0,9), which is critical to the main objective of this study: to minimize the misclassification of reviews relevant to the destination's image. Accuracy is slightly lower (0,75), but remains at a reasonable level, suggesting that the model accurately identifies relevant reviews in most cases. The overall accuracy (0,75) and F1 score (0,78) indicate a good level of performance of the model, although they show that there is space for improvement. The GPT-based classification allowed to filter out approximately 4.000 posts that did not reflect the tourists' perceived image of the city.

Appendix 6 offers an explanation of several aspects related to the use of GPT-3.5 Turbo. It covers the reasons for choosing this model, its associated costs, and the human-crafted instructions used to instruct the model. Additionally, it discusses the model's performance.

Sentiment Analysis

Sentiment Analysis is a field of Natural Language Processing (NLP) that aims to determine the sentiment or emotion expressed in a piece of text. It can be used to identify whether the sentiment of the content is positive, negative, neutral, or even to detect finer-grained sentiments like happiness, frustration, sadness.

Several NLP algorithms can be used for Sentiment Analysis. Appendix 7 features an intuitive matrix that displays the performance of BERT and other models on the GLUE (General Language Understanding Evaluation) tasks [46].

In this research, BERT-base, an open-source machine learning framework for NLP released in 2018 was chosen. BERT, short for Bidirectional Encoder Representations from Transformers, marked a revolution in the realm of Natural Language Processing (NLP). Unlike traditional models, that process words sequentially, BERT grasps a sentence in its entirety, capturing context from both preceding and following words. This bidirectionality offers a richer comprehension of language context, making it effective at various NLP tasks.

BERT base is used to process and extract the sentiment from the reviews that have been classified as in scope using the keywords defined above (I.e., the reviews that contain any word defined in the keywords dictionary). Each review is scored with a value from 0 (very negative sentiment) to 5 (very positive sentiment). Once the reviews have been categorized based on the Destination Image attributes, the reviews not referring to any of them are discarded.

Sentiment Mapping

For each of the attributes of Destination Image under analysis, the average sentiment extracted using BERT-base model is compared over different periods (pre crisis; during the crisis). The objective is to understand how the perception of each attribute has evolved over time and among different groups of tourists (South African, non-South African).

To compare these averages, a statistical method called the "two-sample paired t-test." is used. This method helps understand if the difference in average sentiment between the groups is real or just a coincidence. In statistics, p-values help determine if the results are significant. A low p-value (i.e. < 0,05) indicates that the observed difference is meaningful and not just random.

Results and Discussion

Exploratory Data Analysis

The dataset analyzed comprises 123.664 posts sourced from the Western Cape's TripAdvisor forum. These posts, spanning from 2010 to 2023, were contributed by both national and international tourists. Out of the total, 26% were from national visitors, 31% from international travelers, while 43% did not specify the author's country of origin (Appendix 8).

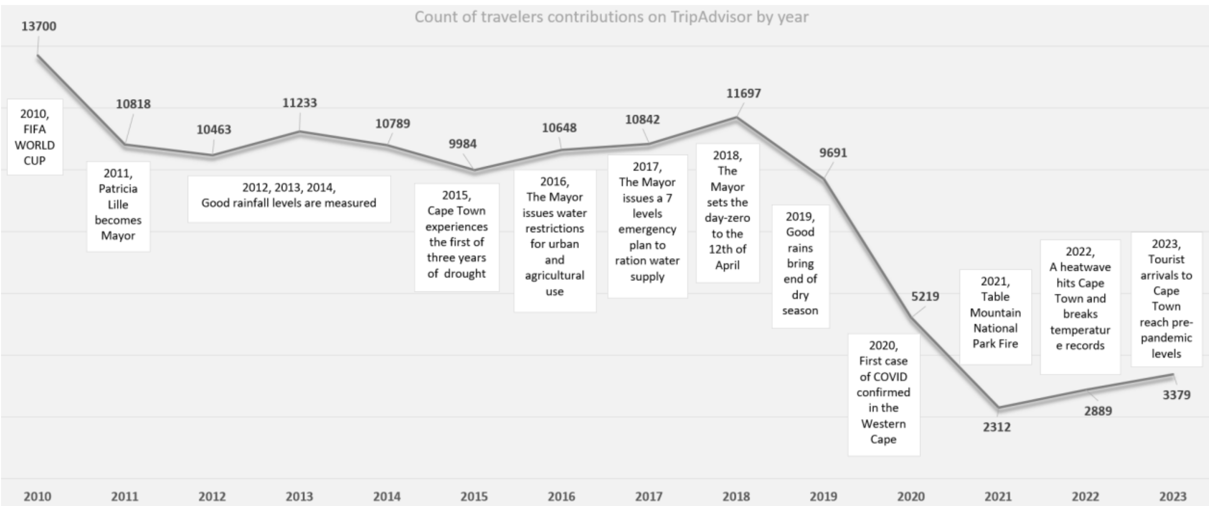


Fig. 2: Count of travelers contributions on TripAdvisor forum for Western Cape and main news by year, from 2010 to 2023

The 2010s were pivotal for Cape Town. The decade started on a positive note with South Africa proudly hosting the FIFA World Cup. It is well documented that sporting events can boost the tourism sector, produce substantial economic gains, and gain widespread media attention [47][48][49]. During the World Cup, Cape Town experienced an increase of 400.000 visitors [50]. This spike is evident in Fig. 2, which shows that 2010 saw the peak in the number of tourist contributions on TripAdvisor forum. This suggests that, at least for that year, there was a positive correlation between the increase in the number of tourists arrivals and the increase in contributions on the forum.

Between 2012 and 2015, Period 1 (P1), tourist arrivals in Cape Town remained constant at around 1.3 million visitors per year as indicated in Appendix 9. Similarly, the number of posts on TripAdvisor forum also remained stable at approximately 10.500 contributions per year (see Fig.2). Other than the election of a new mayor, Patrice de Lille, there were no events of major importance during these years [34]. From a climatic point of view, good rainfall levels were measured These four years were marked by political and climatic stability, with the water crisis not yet emerging as an issue. This is reflected in the absence of mentions of the crisis in the forum, as demonstrated in Appendix 10.1 (e.g., Word Cloud for P1). Therefore, Period 1 will serve as a baseline to assess any changes in the perceptions of the affective attributes of the destination image in the 4 subsequent years when the water crisis emerged.

From 2015 to 2019, Period 2 (P2), tourist arrivals in Cape Town saw a notable increase, peaking at 1.9M visitors in 2019—an impressive 33% growth from 2015, as detailed in Appendix 9. This upward trend is mirrored in the number of TripAdvisor comments over these years, except for 2019, which saw a dip in contributions (see Fig. 2). The surge in arrivals during these years is remarkable, especially when considering that Cape Town was experiencing its most severe drought in three centuries. Additionally, national and international media were broadcasting the municipality's 7-step plan designed to address the drought, rationing water for both locals and tourists [51]. In Period 2, the water crisis emerged as a major concern, which is evident in the increased frequency of related terms on the TripAdvisor forum, as highlighted in Appendix 10.2

(e.g., word cloud for P2). Therefore, this period will be analyzed to highlight any decline in the perceptions of Cape Town's destination attributes compared to Period 1

In March 2020, the first case of COVID-19 was detected, and as the pandemic rapidly spread globally, Cape Town's tourism industry experienced a sharp decline. By 2021, coinciding with a major fire in Table Mountain National Park, tourist arrivals in the region had dropped to a record low of only 293.000 as indicated in Appendix 11. This decline was reflected in the contributions to the Western Cape Forum, which also hit a minimum that same year. Then, in 2022, a new extreme heatwave hit the region and Cape Town. Only in 2023 the industry begun to show signs of recovery, with tourism arrivals returning to pre-pandemic levels [52]. This positive trend is also reflected in the number of contributions on the TripAdvisor forum that it's slowly increasing year on year. However, the main objective of this study is to examine changes in Cape Town's perceived destination image due to the water crisis, hence, the years from 2020 to 2023 are excluded from the analysis to avoid influencing the results with the impacts of COVID-19 and external climate-related factors.

The method defined in the methodology chapter was applied to all contributions posted in Periods 1 and 2 on the specific sub-forum dedicated to Cape Town Central and Cape Town Province. For P1: Out of 31.500 total contributions, 3.800 referenced attributes of the destination image. For P2: Out of 31.000 total reviews, 3.520 were about the destination image attributes.

Topics Analysis

On TripAdvisor each author can either create a new topic or comment an existing one contributing with their opinion on the subject. To illustrate the main topics discussed by tourists on TripAdvisor, the word clouds presented in Appendix 10 were created by analyzing the unique topics on the forum. This word cloud analysis focused on the top 300 words used in the forum's unique topics during Period 1 and Period 2.

To determine whether the topic of the water crisis gained relevance in period 2, a list of commonly used words related to the concept of the water crisis was defined. The frequency of each word in the topic titles was then annotated for each period, as detailed in Appendix 11. The term 'drought,' which was not mentioned at all in P1, recorded 19 mentions in Period 2. Similarly, the terms 'crisis,' 'shortage,' and 'restrictions' were not used in Period 1 but were mentioned 12, 10, and 7 times, respectively, in Period 2. This notable increase in the use of these terms suggests a rising awareness of water-related challenges during that period.

Overall, in Period 1, terms closely associated with the water crisis appeared a mere 5 times out of 4.042 unique topics discussed. In Period 2 instead, these terms were mentioned 112 times amongst 4.167 unique topics. This indicates that the subject of the water crisis increased from occupying 0,001% of the discussion in Period 1 to 2,7% of the discourse on the Cape Town Forum during Period 2. This significant increase suggests that while the water crisis began to attract attention, it did not emerge as one of the predominant topics of discussion.

In Appendix 10.2, which features the word cloud for Period 2, terms related to the water crisis are highlighted in red boxes, all of these words were absent in Appendix 10.1, which presents the word cloud for Period 1.

Overall, among the top 300 words for Period 2, 83 are related to tourism and destination image attributes, accounting for 27% of the total. In both periods, TripAdvisor users frequently discuss topics like beaches and the waterfront, wine and restaurants, as well as the sense of safety, relaxation, and the natural features and outdoor activities of the destination. These popular topics align with the findings of previous studies [53]. Notably, in both periods, there are even mentions of honeymoons, a topic that characterizes Cape Town's destination brand [55]. These findings support the hypothesis, as posited by several studies, that social media data can effectively capture individuals' perceptions of a destination.

Evolution of Affective Attributes Sentiment Across Time

Appendix 12 displays the aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's destination image over time. This includes the number of TripAdvisor contributions for each affective attribute (N), the total number of contributions for each defined period (n_1 and n_2), the percentage of total contributions (% N), the overall average sentiment (μ), and the average sentiment per attribute for each defined period (μ_1, μ_2). Additionally, it provides the p-values (p_1 and p_2), indicating any significant changes in the means for each attribute between Period 1 and Period 2.

From the 9.497 (N) contributions that involved references to the attributes of the destination image of Cape Town, 31% contained a reference to the safety of the destination, followed by the perceived value with 15% and the perceived excitement 11%. Therefore, these three attributes are key in the perception of TripAdvisor's users towards Cape Town.

As detailed in the table in Appendix 12, the top three attributes with the highest average sentiment are perceived hospitality at 0,898, followed by perceived excitement at 0,740, and perceived comfort at 0,717. On the other hand, the three attributes with the lowest average sentiment are perceived authenticity at 0.464, perceived attraction at 0,456, and perceived enjoyment at 0,521. Given our sentiment scale ranging from -2 to 2, most attributes scored significantly above the neutral mark of 0. These findings indicate that despite the water crisis, the image of Cape Town has been positive from 2012 to 2019.

In line with existing literature [53], the results indicate that throughout the entire period analyzed, TripAdvisor users perceived Cape Town as a safe, valuable, attractive, exciting, authentic, hospitable, and relaxing destination. These perceptions align with reports from

TripAdvisor, which highlight that, on average, its users tend to post positive reviews about attractions listed on the website [41]. However, the average sentiment observed over these eight years does not necessarily indicate shifts in the perception of these attributes over time. Rather, it highlights the normative nature of destination images, suggesting that once established, these images can be particularly resistant to change even during crisis [54][45].

Across Periods 1 and 2 (P1 and P2), the findings indicate that among the eleven attributes of the destination image, nine registered a decrease in the average sentiment from P1 to P2. The attributes with the largest decreases were Perceived Enjoyment ($\Delta\mu$: -0,577), Perceived Cleanliness ($\Delta\mu$: -0,271), and Perceived Comfort ($\Delta\mu$: -0,258). The only two attributes that showed an increase were Perceived Hospitality ($\Delta\mu$: 0,035) and Perceived Relaxation ($\Delta\mu$: 0,009), both registering a minimum increase in the average sentiment between P1 and P2.

Overall, only five affective attributes of Cape Town's destination image recorded a significant change between period 1 and period 2. The perception of safety, as viewed by TripAdvisor visitors, decreased (μ1: 0,691; μ2: 0,611; p1-2: 0,031), suggesting that the city was perceived as less safe during the years of the water crisis. Similarly, the city was perceived as less clean, with a significant decrease in the Perceived Cleanliness attribute from P1 to P2 (μ1: 0,810; μ2: 0,539; p1-2: 0,019). Cape Town was viewed as less attractive during period 2 compared to the years before (μ1: 0,549; μ2: 0,353; p1-2: 0,002), and there was a highly significant decline in perceived comfort (μ1: 0,845; μ2: 0,586; p1-2: 0,000). Finally, the destination was perceived as significantly less enjoyable during the years of the water crisis (μ1: 0,778; μ2: 0,200; p1-2: 0,001).

Overall, the findings reveal a significant shift in the TripAdvisor community's perception between Period 1, characterized by relative political and natural stability, and Period 2, coinciding with the water crisis. During Period 2, the community perceived Cape Town as significantly riskier, less enjoyable, less clean, less attractive, and less comfortable. However, attributes such as Perceived Development, Perceived Value, Perceived Authenticity, and Perceived Excitement showed resilience to the water crisis's impact. Notably, the attributes of hospitality and relaxation even recorded an increase in average sentiment.

Determining whether the water crisis had only a short-term influence on the city's perceived attributes is challenging, especially considering the subsequent events that affected Cape Town in the last four years, including COVID-19, the Table Mountain National Park fire, and an extreme heatwave. The impact of these events is reflected in the reduced number of traveler contributions on the TripAdvisor forum in 2022, which were less than half of 2019.

Nevertheless, these results contribute to confirm that autonomous agents, such as non-tourism-related events and media coverage, can dynamically influence the destination image [55]. Furthermore, they reinforce the notion that social media data is a reliable source for analyzing and promoting the destination image [45][62][63][64]. Should the positive trend in tourist arrivals observed in 2023 persist, it will provide an opportunity to evaluate the resilience of Cape Town's destination image to the water crisis and other significant events that have hit the city [56].

Evolution of Affective Attributes Sentiment Across Groups of Tourists

Appendix 13 illustrates the evolution of Cape Town's affective destination image across different tourist groups: residents of South Africa, and non-residents. This includes the average sentiment per attribute for each group ($\mu_{SA\text{-}NSA}$), the average sentiment per period for each group ($\mu_{SA\text{-}NSA\ 1}$, $\mu_{SA\text{-}NSA\ 2}$), and the difference in average sentiment between the groups ($\Delta\mu_1$, $\Delta\mu_2$). For each affective attribute of the destination image, it provides the p-values that show whether there are any significant changes between the entire statistical population of each group (South African vs. non-South African) without reference to time ($p_{SA\text{-}NSA}$). It also includes the p-values indicating any significant changes in sentiment between the groups of tourists for Period 1 and Period 2 ($p_{SA\text{-}NSA\ 1}$ and $p_{SA\text{-}NSA\ 2}$). To visualize this, consider a jar of marbles as an analogy. Blue marbles represent the average sentiment of Group SA, and red marbles represent Group NSA. We are investigating whether there's a significant difference in the average size of blue marbles compared to red marbles. If there is a significant difference ($p < 0.05$), it suggests that the difference in sentiments between the two groups is not just due to random variation; rather, it represents a real, meaningful divergence.

The analysis of the evolution of destination image attributes across space reveals that, out of the 11 defined attributes, only two showed a significant difference in perception between local and international tourists. South African tourists perceived Cape Town as significantly less secure (μ_{SA}: 0,532; μ_{NSA}: 0,684; $p_{SA\text{-}NSA}$: 0,002) and less developed (μ_{SA}: 0,528; μ_{NSA}: 0,880; $p_{SA\text{-}NSA}$: 0,016) compared to international tourists. Additional t-tests were conducted to determine when these different perceptions were most pronounced. For the attribute of Safety, the difference was more marked in Period 1 ($\mu_{SA\ 1}$: 0,526; $\mu_{NSA\ 2}$: 0,753; $p_{SA\text{-}NSA\ 1}$: 0,001). Similarly, for the attribute of Perceived Development, the discrepancy was relevant in Period 1 ($\mu_{SA\ 1}$: 0,528; $\mu_{NSA\ 1}$: 0,880; $p_{SA\text{-}NSA\ 1}$: 0,030).

Considering that only two out of the eleven defined attributes showed a significantly different perception between groups of tourists, these results only partially align with the existing literature [57][45]. However, for nine of the eleven attributes, the average sentiment of non-South African (NSA) tourists was higher than that of South African (SA) tourists. This supports the hypothesis that international visitors generally have a more positive perception of Cape Town's affective destination image [58][45].

The lack of significant differences in perception between local and international tourists might be attributed to South Africa's status as the most economically unequal country in the world

[34]. This economic disparity led to the creation of extremely dangerous districts alongside safe and affluent areas within its cities, a reality that local tourists are likely more aware of compared to international tourists. This awareness could explain why locals perceive Cape Town as significantly less secure and developed. However, it is also possible that the poorer part of the population, who have fewer opportunities to travel is underrepresented on TripAdvsior. This could result in an online community dominated by the middle and upper classes of South Africa, whose social characteristics and travel preferences might be more aligned with those of non-South African tourists, thereby reducing the observed differences in perceptions.

Conclusion and Limitations

This study contributes to the field of ICT in tourism developing a methodology that leverages Python, the OpenAI API, and the T-Test to extract, classify, and analyze a large volume of comments on TripAdvisor. This methodology aims to study the evolution of the affective destination image of Cape Town as perceived by the community of tourists participating in the city's TripAdvisor forum.

The results confirm that Cape Town's overall image on TripAdvisor remained positive both before and during the water crisis. However, the community perceived the city as significantly riskier, less enjoyable, less clean, less attractive, and less comfortable. Additionally, an increase in the frequency of water crisis-related keywords from Period 1 to Period 2 indicates a growing interest in this topic among the TripAdvisor community.

Regarding tourist groups, South African and non-South African, the study reveals that only two out of eleven attributes, safety and development, were perceived significantly more positively by international tourists before the water crisis. During the second period, no significant shift was observed. Overall, nine attributes showed a decrease in sentiment from Period 1 to Period 2, suggesting that the water crisis had a negative impact on the TripAdvisor community's perception of the destination, with local tourists tending to have a more negative sentiment than international tourists.

This research contributes to understanding how the perception of affective attributes of the destination image evolves during medium-term crises, offering valuable insights, particularly for Cape Town. The primary stakeholders of this study are tourism authorities. The research reveals that a segment of the TripAdvisor community actively seeks up-to-date information on natural crises. Understanding tourists' perceptions in real-time would allow tourism bodies to promptly cluster frequently asked questions and promote positive aspects of the destination on social media. The importance of transparent communication by tourism entities regarding emergencies and other relevant information on platforms such as TripAdvisor is therefore emphasised. To this end, these entities could instruct their social media management departments to create official accounts on TripAdvisor and similar platforms to actively participate in forums and monitor new content. Furthermore, the study highlights the potential ease with which AI models and automation tools could be used to spread fake news on these online forums. Such misinformation could have a significant impact on the image of the destination, influencing tourists' perceptions and decisions. It is therefore crucial for tourism authorities to be vigilant and proactive in managing the online presence of their destinations.

For the academic community, the study recognises the potential of using social media data for tourism research and emphasises the usefulness of AI to filter large amounts of social media data to minimise noise, which is a major concern of researchers [44].

This study presents several limitations. Firstly, due to biases introduced by the COVID-19 outbreak and the heatwave that affected Cape Town from 2020 to 2022, it was not possible to determine whether the impact of the water crisis was short-term, as assessed in other studies analyzing the impact of long-term crises on the images of popular destinations worldwide.

Another limitation concerns the methodology used for data classification. While the combination of keyword and AI-based classification showed high precision and excellent recall, it is probable that some posts unrelated to the destination image of Cape Town were included in the analysis. Additionally, the analysis was primarily focused on specific affective attributes of the destination image, omitting the cognitive and conative attributes. This focus have limited the scope of our understanding of the destination image as a whole. Finally, the comparisons between groups of travellers were based on the residency of users as stated on TripAdvisor, rather than on a more comprehensive set of demographic data.

Appendices

Appendix 1: Visualization of the water supply and demand in Cape Town between 2008 and 2017.

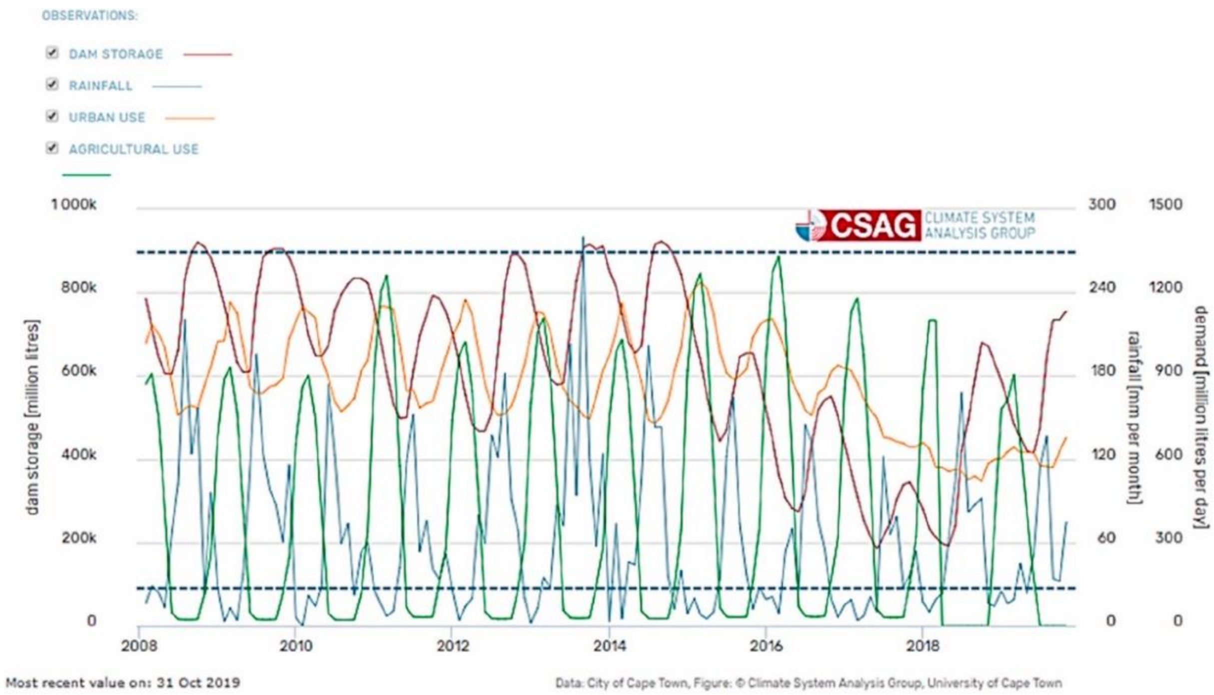


Fig. 3 Water supply and demand in Cape Town between 2008 and 2017 [26]

Appendix 2: In-depth analysis of the impact of the water crisis on local tourism-related businesses.

The impact on tourism businesses has been disproportionate, with small and lower category hotels being the most affected. Four- and five-star accommodation establishments were not as affected [30]. This could be attributed to the perceived ability to provide water from alternative sources to ensure a comfortable stay for guests. A study found that some high-end hotels had invested in water harvesting technologies that allowed them to operate without relying on the city's water supply.

The Tourism Council of South Africa conducted a data collection on a two-year basis, involving various players in tourism, such as car rental companies, travel agents, event organisers and tour operators, to assess the impact of the drought on tourism businesses. The goal was to understand perceptions and experiences regarding the impact of the so-called 'Day Zero' on the operations of these businesses, both in Cape Town and the Western Cape province.

The results show that during the peak of the drought, especially in the first and second half of 2018, tourism businesses felt the effects of the water crisis in different ways. About 40% of the businesses reported a negative impact on their activities, while 35% perceived no significant impact. In addition, 20% of the businesses indicated a considerable impact and the remaining 5% noted some form of negative impact. These data confirm previous observations that the drought had a negative impact on the region's tourism sector. A few months later, in the first half of 2019, less than half of survey participants reported that drought had a negative impact on their business operations.

During the Cape Town water crisis, tourism businesses faced several challenges. In addition to water supply problems, the industry had to cope with an increase in water bills, which particularly affected small and medium-sized businesses. This increase was one of the strategies adopted by the municipality to reduce water demand. The increase in operating costs had a negative impact on employment in the tourism sector. A report by Wesgro found that the drought caused a loss to the Western Cape tourism industry of between 1.707 and 4.024 jobs per year, with a reduction in revenue between 723 million and 1.7 billion rand per year [31]. Revenue losses were the result of a combination of factors, including a decline in hotel employment, a decrease in consumer spending, and a general decline in tourist arrivals in the province.

Appendix 3: Flowchart detailing the main steps of the methodology.

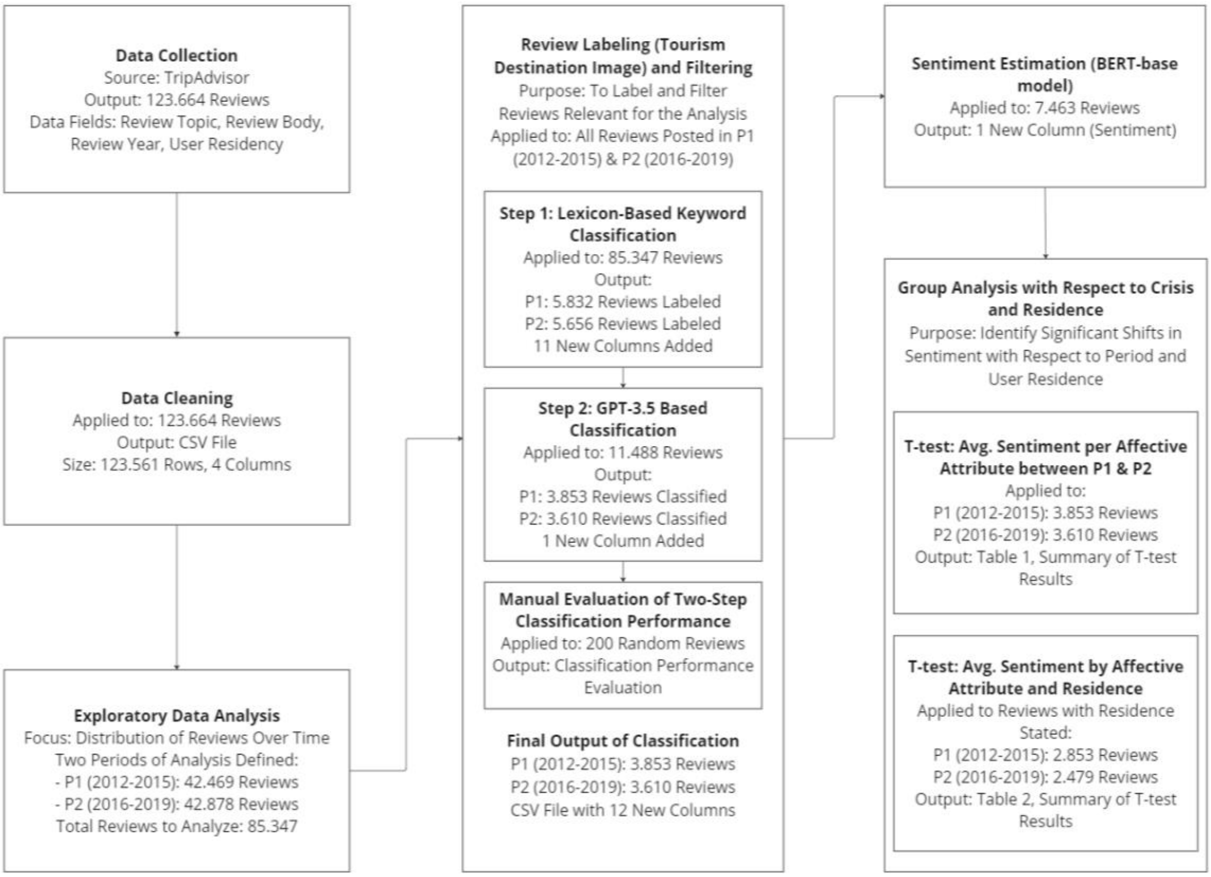


Fig. 4 Flowchart detailing the main steps of the methodology.

Appendix 4: Detailed explanation regarding the data quality of TripAdvisor and the rationale for choosing it as the only data source.

TripAdvisor, founded on 14 February 2000 by Stephan Kaufer [34], has become a key player in Web 2.0 [33], particularly in the tourism sector. TripAdvisor serves as a collaborative platform, allowing users to gather information, share their feedback on travel experiences and participate in interactive travel forums. Over the course of two decades, TripAdvisor has grown to become the world's largest travel site with 463 million monthly users, operating in 49 global markets in 28 languages, accumulating about one billion reviews and ratings [38][39].

TripAdvisor was created on the assumption that tourists planning a trip highly value the opinions of other travellers that influence their decision-making.
Organisations and researchers have invested considerable resources to validate this assumption and to understand the needs of travellers. According to a report by PhocusWire, over 80% of potential travellers read between 6 and 12 online reviews before finalising their accommodation choices, typically focusing on the most recent reviews [35]. Both positive and negative reviews significantly influence tourists' perspectives, with users tending to trust negative reviews more than positive ones [40].
In academic research, TripAdvisor is a prominent data source: over 17.000 articles on Google Scholar and 2.110 on Science Direct include 'TripAdvisor' and 'Data Mining' as keywords. Several studies have detailed their methodologies for extracting data, typically relying on open-source programming languages popular for their efficiency and cost-effectiveness [41]. However, some researches have used paid web scraping services, and few have conducted manual data collection.
It is widely recognised that tourists' decision-making is influenced by reviews on platforms such as Google Maps and TripAdvisor. Another matter is the assessment of the quality of reviews on TripAdvisor, and the legitimacy and authenticity of reviews and shared experiences. On this topic, TripAdvisor publishes the annual Content Moderation Transparency Report, which analyses a year's worth of travellers' contributions. In 2020, more than 26 million reviews were submitted to the site, with more than two million submissions (8,6 per cent of all submissions) rejected or removed for violating community standards or considered false. Of note, only 3,6 per cent of all review submissions, about 1 million reviews, were identified as fake [42].
Despite TripAdvisor's content moderation report, assessing the quality and authenticity of all contributions remains a challenge. However, several studies confirm the hypothesis on the

possibility of using TripAdvisor for local tourism planning needs [36]. An interesting study cross-validated the reliability and validity of TripAdvisor visit patterns against independent data sources such as mobile tracking data and official visitor surveys. The research found a strong correlation between travel patterns derived from the social media and those derived from mobile tracking and surveys, establishing the reliability and representativeness of TripAdvisor visit data [43]. Other research, however, underlined the considerable amount of noise in the data, highlighting the need to develop robust methods for classifying and filtering textual data. The noise does not affect the authenticity of reviews, but rather the quality of the analysis conducted by researchers, who have to take it into account [44].

Appendix 5: Full list of affective attributes used for reviews classification.

The list of affective attributes and keywords used for reviews classification is compiled based on the existent literature [4][45]. These affective attributes include the Perceived Safety; Perceived Development; Perceived Cleanliness; Perceived Hospitability; Perceived Value; Perceived Attraction; Perceived Relax; Perceived Comfort; Perceived Authenticity; Perceived Excitement; and Perceived Enjoyment.

Affective Attribute of the Destination Image	Keywords
Safe Environment	
Perceived Safety	“Risky”, “Scary”, “Dangerous”, “Uncertain”, “Safe”, “Secure”, “Stable”, “Certain”
Perceived Development	“Undeveloped”, “Decline”, “Deteriorating”, “Developed”,”Evolving”, “Progressing”, “Stagnant”,”Modern”
Hospitable Environment	
Perceived Cleanliness	“Dirty”,”Unclean”,”Soiled”,”Pollution”, “Clean”,”Unpolluted”, “Bright”,”Neat”,
Perceived Hospitability	“Unhospitable”, “Inhospitable”, “Hostile”, “Unwelcoming”, “Unfriendly”, “Hospitable”, “Friendly”, “Courtesy”, “Warmth”, “Generosity”

General Mood and Atmosphere	
Perceived Value	“Insignificant”,‘‘Worthless”,”Unremarkable”,”Disappointing”,”Dispensable”, “Valuable”, “Remarcable”,”Beneficial”,”Useful”, “Worth”
Perceived Attraction	“Attractive”, “Charming”,”Stimulating”, “Scenic”,”Fascinating”, “Unattractive”, “Unappealing”,”Awkward”,”Bad”, “Ordinary”
Relaxing Effect	
Perceived Relax	“Stressful”,”Upsetting”,”Chaotic”, “Frenetic”, “Relax”, “Peaceful”,” Restful”, “Recreational”
Perceived Comfort	“Terrifying”,”Frightening”,”Alarming”,”Shocking”,”Comfortable”,”Convenient”,”Cozy”, “Relieving”
Authenticity of the Experience	
Perceived Authenticity	“Artificial”,”Unreal”,”Fake”,”False”,”Unnatural”, “Authentic”, “Convincing”,”Original”,”Pure”, “True”
Perceived Excitement	“Boring”,”Uninteresting”, “Tiresome”,”Dull”,”Mundane”, “Thrilling”, “Exciting”,”Interesting”,”Fun”, “Eventful”
Perceived Enjoyment	“Unpleasure”, “Nasty”, “Undesirable”, “Unenjoyable”, “Pleasure”,”Amusing”, “Joy”,” Satisfaction”

Table 2: Full list of affective attributes used for text classification.

Appendix 6: Detailed explanation of several aspects related to the use of GPT-3.5 Turbo.

For this study, the model chosen is GPT 3.5 Turbo. OpenAI describes it as the most capable and cost-effective in the GPT-3.5 series, optimised for chat and standard completion tasks. Its price is $0,0010 per 1K token for input and $0,0020 per 1K token for output.1K tokens correspond approximately to 750 words. Thus, a 500-word review with a 250-word instruction for the model would cost $0,0010 for processing. Analysing 10.000 reviews of this length would cost about $10. To make a comparison, GPT 4 costs 10 times as much. In order to make efficient use of resources, it is essential to carefully choose the model according to one's needs, as well as to study and test its use in order to minimise errors and save money.

GPT 3.5. Turbo functions like a chat, in which the user provides an instruction (input) and receives a response (output) from the AI model. The following instruction was meticulously crafted to further classify the remaining reviews from the keyword-based classification and minimise noise in the dataset:

“Analyze TripAdvisor reviews to determine their relevance to specific attributes related to tourist destinations, excluding mentions of businesses or products.

Focus Attributes: Safety, development level, hospitality, cleanliness, value for money, attractiveness, relaxation, authenticity of the environment, emotional impact, overall pleasantness.

Scoring System:

Score 1: If the review discusses any of the listed attributes in the context of the destination (not a business or product). This applies to both positive and negative mentions. Look for explicit or implicit references to these attributes as they relate to tourists' experiences or perceptions of the destination.

Score 0: If the review does not mention any of these attributes, focuses solely on businesses/products, or is unrelated to tourists' perceptions or experiences of the destination.

Guidelines: Ensure attributes mentioned are in the context of the destination experience, not just business or product reviews. Be alert to indirect references where attributes are inferred but not directly stated. Prioritize accuracy; when uncertain, reassess the context for correct classification. Apply these criteria uniformly across all reviews.

Output: Provide a final output of 1 or 0 for each review. A single relevant reference to any of the attributes results in a score of 1; otherwise, it's 0. Do not provide any text in the output, just the score.”

Appendix 7: Performance of BERT and other models on the GLUE (General Language Understanding Evaluation) tasks.

The GLUE (General Language Understanding Evaluation) benchmark provides resources to train, evaluate, and compare language models. It includes nine challenging tasks to assess how well an NLP model understands language. Here is an overview of how BERT and other models performed on these tasks [46]:

BERT's Performance on GLUE:

Task	Average	Grammatical	Sentiment Analysis	Similarity	Paraphrase	Question Similarity	Contradiction	Answerable	Entail
BERTLARGE	**82.1**	**60.5**	**94.9**	**86.5**	**89.3**	**72.1**	**86.7/85.9**	**92.7**	**70.1**
BERTBASE	79.6	52.1	93.5	85.8	88.9	71.2	84.6/83.4	90.5	66.4
OpenAI GPT	75.1	45.4	91.3	80.0	82.3	70.3	82.1/81.4	87.4	56.0
Pre-OpenAI SOTA	74.0	35.0	93.2	81.0	86.0	66.1	80.6/80.1	82.3	61.7
BiLSTM+ELMo+Attn	71.0	36.0	90.4	73.3	84.9	64.8	76.4/76.1	79.8	56.8

Fig 5. Performance of BERT and other models on the GLUE tasks.

Appendix 8: Picart with the breakdown of the main nationalities of the authors in the dataset.

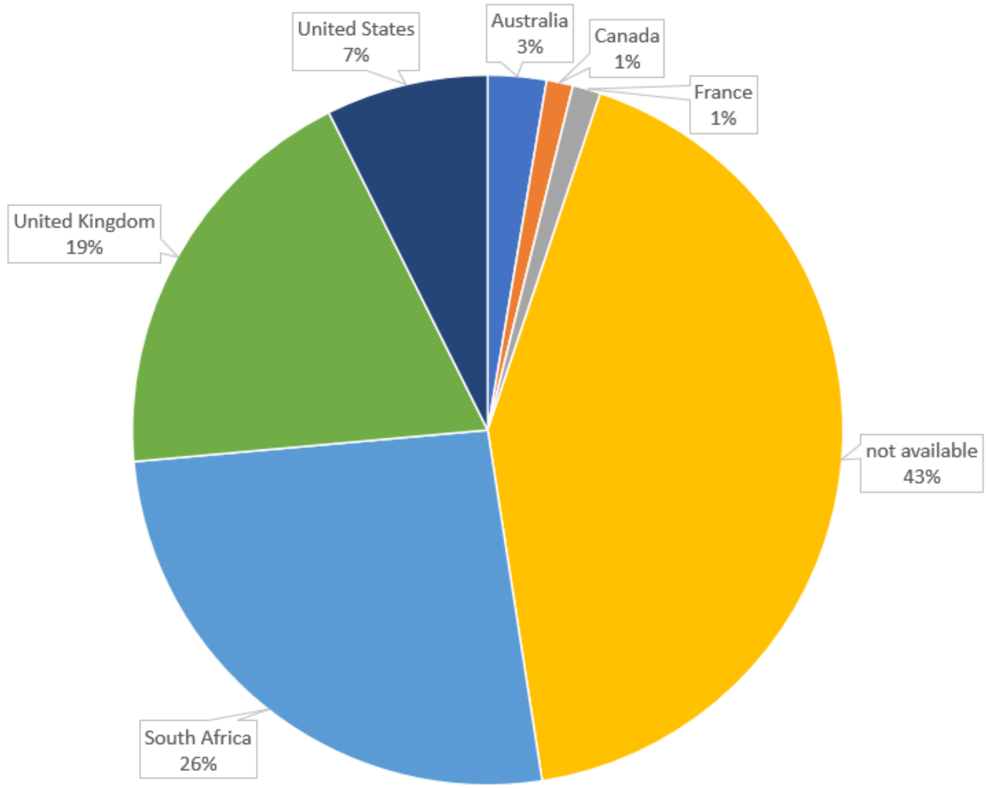
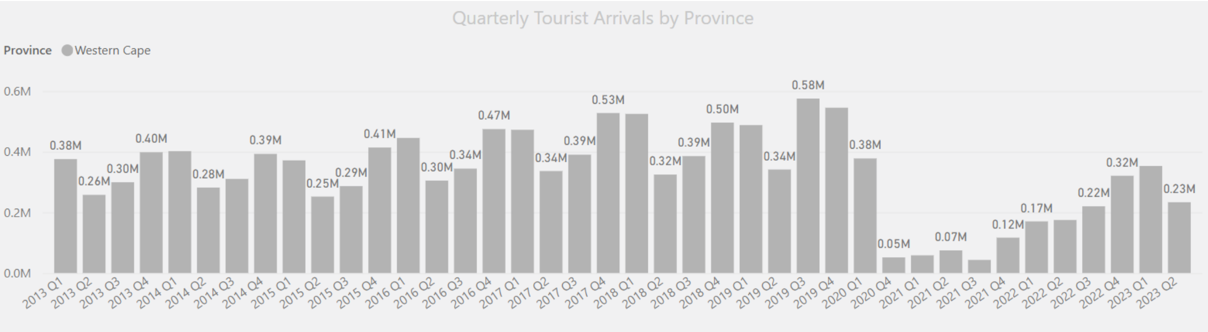


Fig. 6 breakdown of the main nationalities of the authors in the dataset.

Appendix 9: count of tourist arrivals in Western Cape by quarter (from 2013 to 2022).

Fig. 7: count of tourist arrivals in Western Cape by quarter (from 2013 to 2022).



Appendix 10.1: Words Cloud for period 1

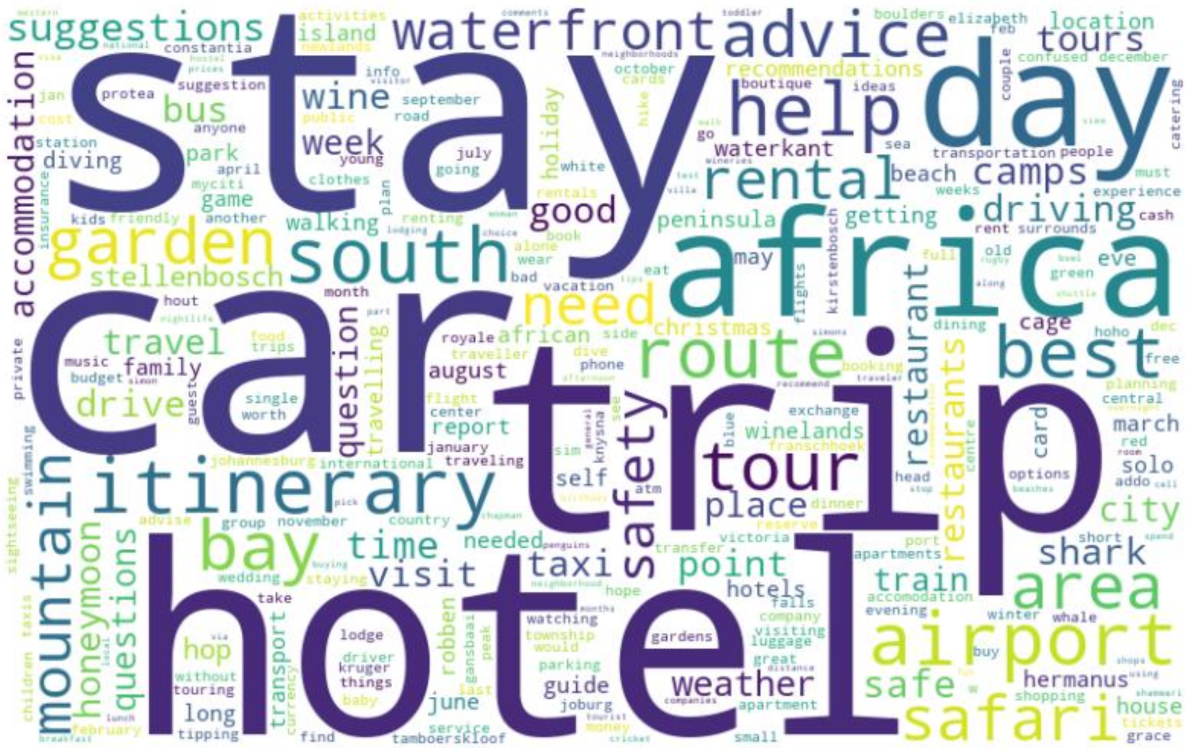


Fig. 8: top 300 words by frequency in topics for period 1 (2012-2015).

Appendix 10.2 Words Cloud for period 2

Fig. 9: top 300 words by frequency in topics for period 2 (2016-2019).

Appendix 11: Frequency of water crisis related terms in period 1 and period 2.

Word	Frequency in P1	Frequency in P2
water	4	60
drought	0	19
crisis	0	12
shortage	0	10
restrictions	0	7
zero	1	4
TOT	5	112

Table 3: Frequency of water crisis related terms in period 1 and period 2.

Appendix 12: Aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's destination image over time.

Appendix 12 displays the aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's destination image over time. This includes the number of TripAdvisor contributions for each affective attribute (N), the total number of contributions for each defined period (n_1 and n_2), the percentage of total contributions (% N), the overall average sentiment (μ), and the average sentiment per attribute for each defined period (μ_1, μ_2). Additionally, it provides the p-values (p_1 and p_2), indicating any significant changes in the means for each attribute between Period 1 and P2.

Attribute	**N**	**n_1**	**n_2**	**% N**	**μ**	**$\Delta\mu$**	**μ_1**	**μ_2**	**$p_{1\text{-}2}$**
Safe Environment									
Perceived Safety	2921	1411	1510	0,31%	0,649	-0,087	0,691	0,611	0,031
Perceived Development	201	118	83	0,02%	0,706	-0,033	0,720	0,687	0,411
Hospitable environment									
Perceived Cleanliness	494	148	152	0,05%	0,673	-0,271	0,810	0,539	0,019
Perceived Hospitability	460	265	195	0,05%	0,898	0,035	0,883	0,918	0,366
General mood and atmosphere									
Perceived Value	1430	817	613	0,15%	0,586	-0,020	0,595	0,574	0,348
Perceived Attraction	1020	536	484	0,11%	0,456	-0,195	0,549	0,353	0,002
Relaxing Effect									
Perceived Relax	474	267	207	0,05%	0,700	0,009	0,697	0,705	0,463
Perceived Comfort	609	309	302	0,06%	0,717	-0,258	0,845	0,586	0,000
Authenticity of the experience									
Perceived Authenticity	679	363	316	0,07%	0,464	-0,104	0,512	0,408	0,122
Perceived Excitement	1063	569	496	0,11%	0,740	-0,079	0,777	0,698	0,085
Perceived Enjoyment	146	81	65	0,02%	0,521	-0,577	0,778	0,200	0,001

Table 4. Aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's destination image over time.

Appendix 13: Aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's between groups of tourists.Appendix 13 illustrates the evolution of Cape Town's affective destination image across different tourist groups: residents of South Africa, and non-residents. This includes the average sentiment per attribute for each group ($\mu_{SA\text{-}NSA}$), the average sentiment per period for each group ($\mu_{SA\text{-}NSA\ 1}$, $\mu_{SA\text{-}NSA\ 2}$), and the difference in average sentiment between the groups ($\Delta\mu_1$, $\Delta\mu_2$). For each affective attribute of the destination image, it provides the p-values that show whether there are any significant changes between the entire statistical population of each group (South African vs. non-South African) without reference to time ($p_{SA\text{-}NSA}$). It also includes the p-values indicating any significant changes in sentiment between the groups of tourists for Period 1 and Period 2 ($p_{SA\text{-}NSA\ 1}$ and $p_{SA\text{-}NSA\ 2}$).

Attribute	**Residency**	$\mu_{SA\text{-}NSA}$	$p_{SA\text{-}NSA}$	μ_1	$\Delta\mu_1$	$p_{SA\text{-}NSA\ 1}$	μ_2	$\Delta\mu_2$	$p_{SA\text{-}NSA\ 2}$
Safe Environment									
Perceived Safety	SA	0,532	0,002	0,526	0,227	0,001	0,540	0,082	0,153
	NSA	0,684		0,753			0,621		
Perceived Development	SA	0,528	0,016	0,467	0,396	0,030	0,630	0,277	0,151
	NSA	0,880		0,863			0,906		
Hospitable environment									
Perceived Cleanliness	SA	0,766	0,852	0,896	-0,063	0,372	0,552	0,093	0,357
	NSA	0,738		0,833			0,645		
Perceived Hospitability	SA	0,719	0,102	0,747	0,131	0,212	0,667	0,221	0,147
	NSA	0,881		0,877			0,888		
General mood and atmosphere									
Perceived Value	SA	0,502	0,083	0,519	0,079	0,170	0,471	0,110	0,144
	NSA	0,590		0,598			0,581		
Perceived Attraction	SA	0,419	0,226	0,513	0,085	0,192	0,270	0,083	0,241
	NSA	0,475		0,598			0,353		
Relaxing Effect									
Perceived Relax	SA	0,594	0,143	0,610	0,140	0,160	0,571	0,079	0,324
	NSA	0,710		0,750			0,650		
Perceived Comfort	SA	0,677	0,474	0,782	0,045	0,362	0,532	0,026	0,426
	NSA	0,683		0,826			0,558		
Authenticity of the experience									
Perceived Authenticity	SA	0,373	0,149	0,470	0,097	0,239	0,224	0,173	0,125
	NSA	0,478		0,567			0,396		
	SA	0,699	0,283	0,779	0,000	0,498	0,573	0,127	0,108

Perceived Excitement	NSA	0,738		0,779			0,700		
Perceived Enjoyment	SA	0,490	0,410	0,741	- 0,026	0,462	0,182	0,152	0,323
	NSA	0.538		0.714			0.333		

Table 5 aggregated outcomes of sentiment analysis, mapping the evolution of Cape Town's between groups of tourists.

References

[1] Kock, F., Josiassen, A. and Assaf, A. (2016), "Advancing destination image: The destination content model", Annals of Tourism Research, Vol. 61, pp. 28–44, doi: 10.1016/j.annals.2016.07.003.

[2] Prayag, G. and Ryan, C. (2012), "Antecedents of Tourists' Loyalty to Mauritius The Role and Influence of Destination Image, Place Attachment, Personal Involvement, and Satisfaction", Journal of Travel Research, Vol. 51, pp. 342–356, doi: 10.1177/0047287511410321.

[3] Baloglu, S. and Brinberg, D. (1997), "Affective images of tourism destinations", Journal of Travel Research, SAGE Publications Ltd, Vol. 35 No. 4, pp. 11–15, doi: 10.1177/004728759703500402.

[4] Echtner, C.M. and Ritchie, J.R.B. (1993), "The Measurement of Destination Image: An Empirical Assessment", Journal of Travel Research, SAGE Publications, Vol. 31 No. 4, pp. 3–13, doi: 10.1177/004728759303100402

[5] Crompton, J.L. (1979), "Motivations for pleasure vacation", Annals of Tourism Research, Vol. 6 No. 4, pp. 408–424, doi: https://doi.org/10.1016/0160-7383(79)90004-5.

[6] Baloglu, S. and McCleary, K.W. (1999), "A model of destination image formation", Annals of Tourism Research, Elsevier Ltd, Vol. 26 No. 4, pp. 868–897, doi: 10.1016/S0160-7383(99)00030-4

[7] Del Bosque, I.R. and San Martín, H. (2008), "Tourist satisfaction a cognitive-affective model", Annals of Tourism Research, Pergamon, Vol. 35 No. 2, pp. 551–573, doi: 10.1016/J.ANNALS.2008.02.006

[8] Onatski, Alexei, Marcelo J. Moreira, and Marc Hallin. 2014. "Signal Detection in High Dimension: The Multispiked Case.". https://www.researchgate.net/publication/232503167_Signal_Detection_in_High_Dimension_The_Multispiked_Case

[9] UNWTO. 2023. https://www.unwto.org.

[10] Athanasopoulou, Anna. 2013. "Tourism as a Driver of Economic Growth and Development in the EU-27 and ASEAN Regions." Associate Fellow of the European Union Centre in Singapore, December 2013. https://aei.pitt.edu/102635/

[11] Camera dei deputati. 2020. https://www.camera.it.

[12] UNCTAD. 2017. https://unctad.org.

[13] Anholt, Simon. 2009. "Handbook on Tourism Destinations Branding." World Tourism Organization (WTO). https://issuu.com/bibliotecaperiodismoturistico/docs/handbook_on_tourism_-_destination_b

[14] World Population Review. (2024), "Most Visited Countries 2023", available at: https://worldpopulationreview.com/country-rankings/most-visited-countries (accessed 7 January 2024).

[15] Souiden, Nizar, Riadh Ladhari, and Nizar El-Chiadmi. 2017. "Destination Personality and Destination Image." Journal of Hospitality and Tourism Management. https://www.sciencedirect.com/science/article/abs/pii/S1447677016301802

[16] Gartner, W.C. (1994), "Image Formation Process", Journal of Travel & Tourism Marketing, Taylor & Francis Group, Vol. 2 No. 2–3, pp. 191–216, doi: 10.1300/J073V02N02_12.

[17] Gallarza, M.G., Saura, I.G. and García, H.C. (2002), "Destination image: Towards a Conceptual Framework", Annals of Tourism Research, Pergamon, Vol. 29 No. 1, pp. 56– 78, doi: 10.1016/S0160-7383(01)00031-7.

[18] Stylos, Nikolaos, and Andreas Andronikidis. 2013. "Exploring the Cognitive Image of a Tourism Destination." TOURISMOS: An International Multidisciplinary Journal of Tourism. https://www.researchgate.net/profile/Nikolaos-Stylos/publication/259345076_Exploring_the_cognitive_image_of_a_tourism_destination/links/0c96052b202d0eab78000000/Exploring-the-cognitive-image-of-a-tourism-destination.pdf

[19] Stylidis, D., Belhassen, Y. and Shani, A. (2015), "Destination image, on-site experience and behavioural intentions: path analytic validation of a marketing model on domestic tourists", Current Issues in Tourism, Vol. 20, pp. 1–18, doi: 10.1080/13683500.2015.1051011.

[20] Jenkins, O.H. (1999), "Understanding and measuring tourist destination images", International Journal of Tourism Research, John Wiley & Sons, Ltd, Vol. 1 No. 1, pp. 1– 15, doi: https://doi.org/10.1002/(SICI)1522-1970(199901/02)1:13.0.CO;2-L.

[21] Avraham, E. (2004), "Media Strategies for Improving an Unfavorable City Image", Cities, Vol. 21, pp. 471–479, doi: 10.1016/j.cities.2004.08.005

[22] Parsons, W. (1996), "Crisis management", Career Development International, MCB UP Ltd, Vol. 1 No. 5, pp. 26–28, doi: 10.1108/13620439610130614.

[23] Ritchie, B.W. (2004), "Chaos, crises and disasters: a strategic approach to crisis management in the tourism industry", Tourism Management, Vol. 25 No. 6, pp. 669–683, doi: https://doi.org/10.1016/j.tourman.2003.09.004.

[24] Avraham, Eli, and Eran Ketter. 2012. "Media Strategies for Marketing Places in Crisis: Improving the Image of Cities, Countries and Tourist Destinations." https://www.researchgate.net/publication/286009170_Media_strategies_for_marketing_places_in_crisis_Improving_the_image_of_cities_countries_and_tourist_destinations

[25] Booysen, M.J., Wijesiri, B., Ripunda, C. and Goonetilleke, A. (2019), "Fees and governance: Towards sustainability in water resources management at schools in post-apartheid South Africa", Sustainable Cities and Society, Vol. 51, p. 101694, doi: https://doi.org/10.1016/j.scs.2019.101694.

[26] LaVanchy, G.T., Kerwin, M. and Adamson, J. (2019), "Beyond 'Day Zero': insights and lessons from Cape Town (South Africa)", Hydrogeology Journal, Vol. 27, doi: 10.1007/s10040-019-01979-0.

[27] Sousa, P.M., Blamey, R.C., Reason, C.J.C., Ramos, A.M. and Trigo, R.M. (2018), "The 'Day Zero' Cape Town drought and the poleward migration of moisture corridors",

Environmental Research Letters, IOP Publishing, Vol. 13 No. 12, p. 124025, doi: 10.1088/1748-9326/aaebc7

[28] Nhamo, G. and Agyepong, A.O. (2019), "Climate change adaptation and local government: Institutional complexities surrounding Cape Town's Day Zero", Jamba: Journal of Disaster Risk Studies, AOSIS (pty) Ltd, Vol. 11 No. 3, doi: 10.4102/jamba.v11i3.717.

[29] Time. 2018. "What It's Like to Live Through Cape Town's Massive Water Crisis." https://time.com/cape-town-water-crisis/.

[30] Dube, K., Nhamo, G. and Chikodzi, D. (2022), "Climate change-induced droughts and tourism: Impacts and responses of Western Cape province, South Africa", Journal of Outdoor Recreation and Tourism, Vol. 39, p. 100319, doi: https://doi.org/10.1016/j.jort.2020.100319.

[31] WESGRO. (2018), Western Cape Destination Performance Report: Annual 2018.

[32] OECD. (2021), Water Governance in Cape Town, South Africa.

[33] Valdivia, Ana, Luzón, M. Victoria, and Herrera, Francisco. "Sentiment Analysis in TripAdvisor." In Affective Computing and Sentiment Analysis, edited by Erik Cambria, Nanyang Technological University. https://sentic.net/sentiment-analysis-in-tripadvisor.pdf.

[34] Wikipedia. 2023. "Mediterranean Climate." https://en.wikipedia.org/wiki/Mediterranean_climate; Tripadvisor Foundation. "Water Crisis Timeline."; Cape Town News. "Cape Town Timeline."

[35] Prabu, Karthick. 2014. "Vast Majority of TripAdvisor Users Read at Least 6-12 Reviews Before Choosing Hotel." Phocuswire. https://www.phocuswire.com/Vast-majority-of-TripAdvisor-users-read-at-least-6-12-reviews-before-choosing-hotel.

[36] Minkwitz, A. (2018), "Tripadvisor as a Source of Data in the Planning Process of Tourism Development on a Local Scale", Turyzm, Vol. 28, pp. 49–55, doi: 10.2478/tour-2018- 0014.

[37] TripAdvisor. "Western Cape Forum." https://www.tripadvisor.com/ShowForum-g312653-i9207-o10420-Western_Cape.html.

[38] Statista Research Department. 2023. "Tripadvisor - Statistics & Facts." https://www.statista.com/topics/3443/tripadvisor/#topicOverview.

[39] Tripadvisor. "Media Centre." https://tripadvisor.mediaroom.com/IN-about-us.

[40] Gavilan, D., Avello, M. and Martinez, G. (2017), "The influence of online ratings and reviews on hotel booking consideration", Tourism Management, Vol. 66, pp. 53–61, doi: 10.1016/j.tourman.2017.10.018.

[41] Ali, Twil, Marc, Bidan, Bencharef, Omar, Kaloun, Soulaimane, and Larbi, Safaa. 2021. "Exploring Destination's Negative E-reputation Using Aspect Based Sentiment Analysis Approach: Case of Marrakech Destination on TripAdvisor." Tourism Management Perspectives 40. DOI: 10.1016/j.tmp.2021.100892. https://www.sciencedirect.com/science/article/pii/S2211973621001057.

[42] Tripadvisor. 2021. “2021 Review Transparency Report.” PRNewswire, Oct. 27. https://tripadvisor.mediaroom.com/2021-10-27-Tripadvisor-Content-Moderation-Transparency-Report-Reveals-New-Data-In-Fight-Against-Fake-Reviews.

[43] Ma, Shihan, Kirilenko, Andrei. 2021. “How Reliable Is Social Media Data? Validation of TripAdvisor Tourism Visitations Using Independent Data Sources.” In Information and Communication Technologies in Tourism 2021. ISBN: 978-3-030-65784-0. https://link.springer.com/chapter/10.1007/978-3-030-65785-7_26#:~:text=The%20validity%20was%20explored%20in,to%20be%20reliable%20and%20representative.

[44] Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2018), “Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews”, Information Technology and Tourism, Springer Berlin Heidelberg, Vol. 18 No. 1–4, pp. 43–59, doi: 10.1007/s40558-017-0098- z.

[45] Gkritzali, A., Gritzalis, D. and Stavrou, V. (2017), “Is Xenios Zeus Still Alive? Destination Image of Athens in the Years of Recession”, Journal of Travel Research, SAGE Publications Inc, Vol. 57 No. 4, pp. 540–554, doi: 10.1177/0047287517705225.

[46] Muller, Britney. 2022. “BERT 101: State Of The Art NLP Model Explained.” https://huggingface.co/blog/bert-101.

[47] Zagnoli, Patrizia, and Elena Radicchi. 2009. “Do Major Sports Events Enhance Tourism Destinations?” DOI: 10.2478/v10141-009-0031-z. https://intapi.sciendo.com/pdf/10.2478/v10141-009-0031-z#:~:text=Major%20sports%20events%20are%20seen,of%20stay%20during%20the%20event.

[48] Decker, Evan. 2023. “Sports Tourism: A Game-Changing Phenomenon in the World of Travel.” Sports Tourism Manager at Visit Billings, June 14. https://brand.mt.gov/Programs/Office-Of-Tourism/News/news-articles/sports-tourism

[49] United Nations World Tourism Organization. “Sports Tourism.” https://www.unwto.org/sport-tourism.

[50] Szymanski, S. (2010), “World Cup Tourism Totals”, available at: https://www.csmonitor.com/Business/The-Sports-Economist/2010/1001/World-Cuptourism-totalsc (accessed 7 January 2024).

[51] Thompson Barton. (2018), “Running Out of Water: Cape Town, the U.S., and Drought”, 6 February, available at: https://law.stanford.edu/2018/02/06/running-out-of-water-capetown-the-u-s-and-drought/ (accessed 7 January 2024)

[52] South African Tourism. 2023. https://www.southafrica.net/gl/en/corporate/category/media-centre.

[53] Maumbe, K. and Van Wyk, L. (2016), “Destination Cape & Western Cape: Analysis of destination brand awareness, identity and image in the domestic market ”, Travel and Tourism Research Association: Advancing Tourism Research Globally, Vol. 24.

[54] Anholt, Simon. 2010. ‘‘Definitions of Place Branding – Working Towards a Resolution.” Place Branding and Public Diplomacy 6 (1). DOI: 10.1057/pb.2010.3. https://doi.org/10.1057/pb.2010.3.

[55] Gabbioneta, Claudia, and De Carlo, Manuela. 2019. ‘‘The Role of News Articles, Prior Destination Experience, and News Involvement in Destination Image Formation.” DOI: 10.1002/jtr.2251. https://onlinelibrary.wiley.com/doi/abs/10.1002/jtr.2251.

[56] Thwala, Konanani, and Dube, Kaitano. 2023. ‘‘Examining COVID-19 Recovery and Resilience of Cape Town as a Tourist Destination.” DOI: 10.1007/978-3-031-28340-6_13. https://www.researchgate.net/publication/370734105_Examining_COVID-19_Recovery_and_Resilience_of_Cape_Town_as_a_Tourist_Destination/citation/download.

[57] Bonn, M., Joseph-Mathews, S. and Dai, M. (2005), “International versus Domestic Visitors: An Examination of Destination Image Perceptions”, Journal of Travel Research, Vol. 43, pp. 294–301, doi: 10.1177/0047287504272033.

[58] Wang, D., Hu, S., Feng, L. and Lu, Y. (2022), “Tourism Destination Image Perception Model Based on Clustering and PCA from the Perspective of New Media and Wireless Communication Network: A Case Study of Leshan”, edited by Ning, X.Wireless Communications and Mobile Computing, Hindawi, Vol. 2022, p. 8630927, doi: 10.1155/2022/8630927.

[59] Baker, Aryn. 2019. ‘‘What South Africa Can Teach Us as Worldwide Inequality Grows.” Time, May 2. https://time.com/longform/south-africa-unequal-country/.

[60] Sieff, Kevin. 2018. ‘‘As Cape Town’s Water Runs Out, The Rich Drill Wells; The Poor Worry About Eating.” The Washington Post. https://www.washingtonpost.com/news/world/wp/2018/02/23/feature/as-cape-towns-water-runs-out-the-rich-drill-wells-the-poor-worry-about-eating/

[61] https://www.tripadvisor.com.sg/ShowTopic-g1-i12105-k12168234-Country_of_Origin-Tripadvisor_Support.html

[62] 20th International Scientific Conference Economics and Management - 2015 (ICEM-2015) The relation between destination image and social media user engagement – theoretical approach Krzysztof Stepaniuk. https://www.researchgate.net/publication/286444546_The_Relation_between_Destination_Image_and_Social_Media_user_Engagement_-_theoretical_Approach

[63] Lin, M.S., Liang, Y., Xue, J.X., Pan, B. and Schroeder, A. (2021), “Destination image through social media analytics and survey method”, International Journal of Contemporary Hospitality Management, Emerald Group Holdings Ltd., Vol. 33 No. 6, pp. 2219–2238, doi: 10.1108/IJCHM-08-2020-0861/FULL/PDF.

[64] Sultan, Mohammad Tipu, Farzana Sharmin, Alina Badulescu, Darie Gavrilut, and Ke Xue. 2021. ‘‘Social Media-Based Content towards Image Formation: A New Approach to the Selection of Sustainable Destinations” Sustainability 13, no. 8: 4241. https://doi.org/10.3390/su13084241 https://www.mdpi.com/2071-1050/13/8/4241