forests

Article

Exploring National Park Visitors' Judgements from Social Media: The Case Study of Plitvice Lakes National Park

Carlotta Sergiacomi [1,*], Dijana Vuletić [2], Alessandro Paletto [3], Elena Barbierato [1] and Claudio Fagarazzi [1]

1 Department of Agriculture, Food, Environment and Forestry, University of Florence, p.le delle Cascine 18, 50144 Florence, Italy; elena.barbierato@unifi.it (E.B.); claudio.fagarazzi@unifi.it (C.F.)

2 Division for International Scientific Cooperation in Southeast Europe, Croatian Forest Research Institute, Cvjetno naselje 41, HR-10450 Jastrebarsko, Croatia; dijanav@sumins.hr

3 Research Center for Forestry and Wood, Council for Agricultural Research and Economics (CREA), p.zza Nicolini 6, 38123 Trento, Italy; alessandro.paletto@crea.gov.it

* Correspondence: carlotta.sergiacomi@unifi.it

Abstract: This study aims to conduct a survey of visitor reviews of the Plitvice Lakes National Park in Croatia to detect strengths and weaknesses of the park. In total, 15,673 reviews written in the period between 2007 and 2021 were scraped from the social media platform TripAdvisor. The research applies a comprehensive combination of multidimensional scaling, sentiment analysis, and natural language processing approaches to a sample area of international naturalistic interest. Analyzing the opinions of visitors, the authors identify: the main topics of interest related to the management of the park; and the strengths and weaknesses on the basis of definitely positive and decidedly negative reviews, respectively. The tested methodology is easily applicable for the analysis of different naturalistic contexts and protected areas, even in different countries, thanks to the use of translated reviews. The results obtained show that visitors to protected natural areas are not only interested in naturalistic and landscape aspects but also in issues such as accessibility and management of routes and visits.

1. Introduction

In the last decades, technological advances applied to the tourism sector have radically changed the way information is produced and consulted [1]. Tourists can access an increasing number of sources of knowledge and have many channels available to share their opinions on experiences and places. When the experiences are shared online, they help to define a concrete image of tourist destinations and to shape the decisions of future visitors [2,3]. In particular, social media platforms offer a space to freely share experiences and make judgements [4,5] through the so-called user-generated contents (UGC) [6–8]. For this reason, these platforms are becoming increasingly important both in the planning of destinations and in the definition of management priorities for places of tourist interest [9–12]. Social media can be considered as a rich source of news within which users create, circulate, and consult such information to mutually update each other on products, services, personages, and other objects of interest [13]. They are interactive platforms where individuals or larger communities share UGCs and include, among others, blogs, forums, or social networks [14]. Some social media are of general interest (e.g., Facebook or Twitter), while others are focused on more specific topics (e.g., professional networking on LinkedIn); some of them deal with media sharing (e.g., YouTube or Flickr), while others allow you to provide reviews on products and services (e.g., Google My Business or TripAdvisor).

In this study, TripAdvisor was chosen among the many available social media, because it is the largest travel website in the world, operating in 45 countries around the world [11]. It has more than 400 million visitors visiting every month [15] and more than 450 million reviews and opinions which concern more than seven million accommodations, restaurants, and attractions [16]. Besides, TripAdvisor is available in 28 languages [17]. TripAdvisor reviews are a source of information characterized by several positive aspects, including being free and easily accessible and covering a considerable number of years [3]. In addition to reviews, users can also publish other information, such as the country of provenance and the purpose of the trip. Therefore, user reviews on TripAdvisor combine textual comments (i.e., reviews) with concise ratings (i.e., bubbles). Although recent studies have shown that textual comments receive a lower priority than synthetic evaluations [18], it should be highlighted that users may have different priorities [19] that cannot be fully explained in choosing between one and five bubbles. Therefore, it becomes essential to develop tools which allow more information to be extrapolated from the textual component of the reviews.

The massive amounts of unstructured data that are continuously generated on the Internet necessarily require the use of automated procedures for this kind of data analysis [1,7,12]. Social media analytics is receiving increasing attention from companies in many sectors, because they try to analyze the large amount of data collected through different methods [6,20,21]. Content analysis (CA) is one of the available techniques for extrapolating and analyzing the text contents which is widely used in the tourism research field [11]. Sentiment analysis (SA) approach is part of the CA field, and it is a valid option to process this type of data automatically. SA uses computational linguistics and natural language processing (NLP) to analyze the text and identify the polarity of the judgements contained within it [1,8,16]. Another technique for analyzing unstructured textual data is that of multidimensional scaling (MDS), the main purpose of which is that of a better graphical visualization of the data in order to facilitate the understanding of the text structure [22]. In the international literature, the applications of MDS in tourism studies are numerous [23,24]. MDS is usually associated with cluster analysis, a particular application of which is text clustering [6].

Today, it is essential for the tourist community to identify destinations that provide them with meaningful experiences in natural contexts. In this way, protected forest areas and forested landscapes turn out to be popular destinations thanks to the multitude of natural values that take place within them [25]. In Croatia, this type of destination is well represented by national parks, which correspond to the second-highest level in the scale of protected areas (Law on Nature Protection, OG 88/13, 15/18, 14/19, 127/19). One of the most famous and visited national parks in Croatia is Plitvice Lakes National Park (PLNP). The choice of this well-known park was guided: on one side, by the need to validate a new methodology with a case study for which a great deal of information was already available on the activities and management problems with which to compare the final results; on the other side, by the fact that that social media data prove to be a better proxy of tourist visits in reference to the most popular parks [5].

To the best of our knowledge, no previous studies have focused on visitors' experiences for PLNP. The present study tried to fill this gap in the literature by conducting an in-depth analysis of TripAdvisor tourists' reviews on PLNP, by applying a comprehensive method of text mining and natural language processing techniques.

In particular, this study aims to answer the following research questions.

- RQ1. How to collect and investigate textual data by social media platform to investigate the preferences of users of protected areas?
- RQ2. How to extrapolate and analyze the management issues of greatest interest to visitors who choose protected areas as their destination?
- RQ3. How to identify the strengths and weaknesses of the management of protected areas from the point of view of visitors?

The management of protected forest areas as a potential tourist destination is particularly demanding. This complexity is due to the trade-off between the conservation of natural ecosystems and the promotion of tourist visits for economic reasons [26,27]. Therefore, it is particularly useful to define a flexible methodology for the analysis of the management of protected areas that considers the point of view of visitors. In the present study, the answers to the research questions will allow PLNP managers to monitor the satisfaction of local and international users and plan activities aimed at improving the quality of visits to the park.

The remainder of the paper is organized into the following five sections. Foremost, Section 2 provides a literature review on the analysis of nature-based tourism using MDS and NLP tools. After that, the methodology used is illustrated in Section 3. Section 4 shows the results, while Section 5 discusses the findings. Finally, Section 6 analyzes the limitations of the study and provides suggestions for useful application and future research directions.

2. Literature Review

2.1. Nature-Based Tourism

Nowadays, it is widely recognized that some segments of the tourism sector can be considered a "clean industry" and part of the Green Economy [28]. In particular, nature-based tourism is a growing key sector of this industry [26,29,30] which seeks to respond to a growing consumer demand for a return to nature [3,25]. This need is well explained by the fact that nature is capable of generating human well-being from a physical and psychological point of view. [20,25,31–34]. Moreover, natural areas are a place of refuge for biodiversity, in addition to providing restorative surroundings for people [26,31]. The establishment of protected areas created to conserve biodiversity and esthetic value of landscapes is one of the main pillars of nature-based tourism [29,30]. Thus, protected areas and nature-based tourism represent fundamental access for people to cultural ecosystem services [25,35,36]. Particularly, national parks are characterized by a high level of biodiversity protection among protected areas and, at the same time, provide tourism opportunities [5,26,27,37]. Thus, national parks play a very important role also in the tourism sector. For this reason, it is essential to analyze the factors that attract visitors and make visits to protected areas pleasant. Both internal components (e.g., expectations for places and activities) and external components related to tourism management (e.g., accessibility, means of transportations, etc.) strongly influence visitors' perception of the natural landscape [3]. Consequently, the management of nature-based tourism services must take into account the diversified opinions that visitors have towards nature in general and recreational activities in particular [38]. Therefore, it has become fundamental to evaluate how people perceive their recreational experiences in this type of protected area [8].

2.2. Content Analysis

Content analysis (CA) is a research tool to be adopted in order to identify some particular words or more general concepts within qualitative textual data [2,39] or to extrapolate homogeneous units of meaning from a complex text. Traditionally, CA involved human subjective interpretation by researchers, which has now been replaced by automated procedures and sophisticated software [4]. One of the possible approaches of CA is sentiment analysis (SA), which is also an important component of text mining. Text mining is an interdisciplinary field which draws on information retrieval, data mining, machine learning, statistics, and computational linguistics [40]. Valid overviews on SA were produced by Ma et al. and Alaei et al., to which reference should be made for further information [1,9]. In these contributions, the authors reconstruct the main historical stages that characterized the evolution of the SA and outline its most recent features and applications. Nonetheless, it can be synthetically said that the main purpose of SA is to distinguish between positive, negative, or neutral opinions [1,12,16]. Natural language processing (NLP) is one of the available tools for SA, but its application on UGC from social media in landscape design,

and planning research is still in a preliminary stage [21,41]. In the text analysis, MDS is a particularly valid automated computer algorithm. MDS is a data visualization technique based on the proximity of words and their spatial representation [23,24]. Another type of machine learning algorithm usually associated with MDS is that of cluster analysis, which is usually applied to transform unstructured word sets into structured clusters [21].

Social media analytics—in particular, SA—has been applied to social media in numerous tourism-related research fields [6,39]. The most investigated fields are food and wine tourism [19,39,41,42], hospitality [9,11,43,44], areas of interest or events in cities [4,16,45–47], and natural spaces with special regard to urban parks [20,21,31–33,48]. Conversely, national parks and nature reserves [3,6,8,25,27] are a field still not much investigated [8].

2.3. Nature-Based Tourism and Ccontent Analysis

According to the European Landscape Convention [49], landscape assessment processes should take into consideration public perception of places [50]. To evaluate visitors' perception towards natural destinations, traditional methods, such as in situ questionnaires, in-depth interviews, and focus groups, have long been employed. These techniques are usually time- and resource-consuming, in addition to not allowing the collection of results on a large scale or comparisons over time [3,6,8,27,32,50]. On the other hand, the development of modern tools for web analysis allows us to overcome all of these shortcomings. In the recent literature, numerous research contributions have used CA methods to analyze nature-based tourism destinations, but there are still few contributions that investigate the usability of the various social media platforms in relation to visits to protected areas [3].

Stoleriu et al. explores 226 online TripAdvisor reviews on Danube Delta through an automated CA in order to identify and quantify the main dimensions of visitors' experiences and memories [3]. Their results showed that managerial aspects linked to visit organisation (e.g., trip itinerary and visit duration) were more prominent themes in the tourists' reviews compared to the site characteristics. One of the main limitations of the study in relation to the use of TripAdvisor reviews is the lack of demographic and socioeconomic information of visitors. For this reason, it would be necessary to integrate this type of analysis with surveys that make it possible to evaluate the preferences of visitors based on their characteristics.

Two other recent studies [8,27] conducted SA in some national parks of South Africa. Hausmann et al. used SA and NLP techniques to analyze the content of image captions in 33,213 English posts published on Instagram relating to four national parks in South Africa [8]. The authors identified the main emotional components and the keywords formed by both a single word and a pair of adjacent words that recurred most in the posts. The results showed that the polarity of sentiment about national parks expressed by visitors on social media is generally positive, with a minor expression of negative feelings. This is significant to highlight the social role that national parks assume, favoring the development of positive interactions with nature and, therefore, well-being in visitors. Those authors found that visitors tend to idealize certain places or features of national parks and give them symbolic meaning. This meaning is what makes visiting experiences worth sharing and promoting. Among the problems identified by those authors in using this method, there are: on the one hand, the potential lack of representation of the sample of visitors who publish reviews; on the other hand, the use of an unconventional language (e.g., abbreviations, slang, emojis, etc.) which can make the use of automatic computational systems less effective. In almost the same area, Mangachena and Pickering conducted an analysis of 10,292 English tweets on Twitter about seven South African national parks [27]. Even in this case, they mostly found positive feelings and opinions related to the nature-based experience. Those authors identified a particular interest from visitors regarding specific events, such as commemorations related to the history of the park or discoveries of naturalistic interest. Furthermore, according to previous studies [8], some authors recognized that the use of concise texts, shortened words, and special characters (e.g., hashtags and emoticons), typical of social networks such as Instagram and Twitter, may also complicate text analysis of tourists' reviews [20].

Recently, Niezgoda and Nowacki investigated visitors' opinions towards one of the most visited protected areas in Poland, Tatra National Park [25]. Those authors elaborated a composite methodology made by text mining, NLP, and coding opinion procedures to process the data obtained from 624 English reviews published on TripAdvisor. The authors were interested in identifying the main reasons that led visitors to live experiences in the nature park and whether these were mainly related to the themes of ecological awareness and nature protection. The results of their study showed that the most active forms of entertainment (e.g., hiking, taking photos, mountain climbing) are the main motivation for visiting places in the open air. Those authors also highlight that in order to conduct this type of analysis it is necessary to assume that the reviews contain the elements considered most important by visitors, but it would be advisable to deepen the themes identified with more detailed surveys.

One of the latest applications of CA to national parks is that of Mirzaalian and Halpenny. In their study on Jasper National Park, they analyzed 17,224 English reviews on TripAdvisor [6]. In addition, that study analyzed destination loyalty statements using a keyword clustering approach. Among the main categories of visitor favorite destinations can be found waterfalls and lakes. Those authors acknowledge that one of the biggest limitations of this study is that the analysis did not concern some meaningful management aspects (e.g., transportation or outdoor activities).

3. Materials and Methods

The combination of several tools has made it possible to obtain different types of results that can be useful to the managers of the study area. On the one hand, the strengths and weaknesses of the PLNP from the visitor's point of view stemmed from the NLP technique (i.e., rapid automatic keyword extraction) based on SA scores. On the other hand, the MDS and cluster analyses were carried out to identify the topics most dealt with in the reviews released by PLNP visitors on TripAdvisor.

The different steps of the method used are summarized and described in a procedure flowchart (Figure 1).

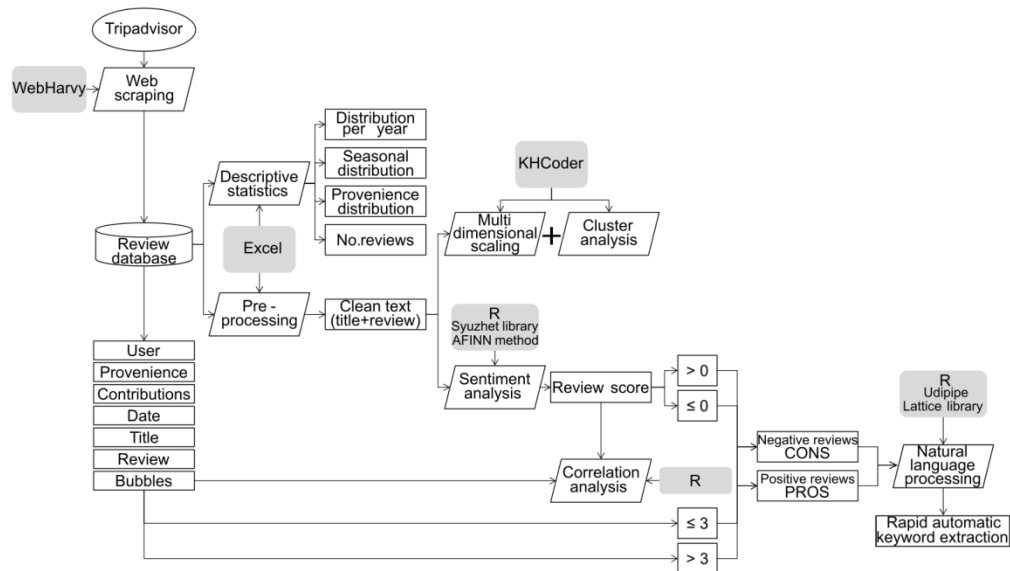**Figure 1.** Flowchart of the research procedure.

3.1. Study Area

Plitvice Lakes National Park (PLNP) is one of the most famous and visited national parks in Croatia. PLNP is located in the mountainous central part of the nation and is part of the Dinaric karst area. PLNP is the oldest protected area (designated 8 April 1949) and the biggest national park (29,685.15 ha) in Croatia. The park mainly consists of

forest areas, which represent about 81% of the total territory, with a complex system of lakes connected with waterfalls. The PLNP is well known for the rich biodiversity of its 296 square kilometers of forests. It is managed by the Plitvice Lakes National Park Public Institution (PLNPPI), founded by the Republic of Croatia and placed under the jurisdiction of the Ministry of the Environment and Energy (MEE). In addition, Plitvice is the only Croatian national park that is on the UNESCO World Heritage list (1979) as natural heritage and is entirely identified as a Natura 2000 site. Despite the large area of the park, only a small part of it represents the point of major tourist interest [37]. It is a lake system which includes 16 main lakes characteristic for their waterfalls, to which are added several other smaller lakes [51]. The park's finances derive entirely from the entrance tickets and hospitality services, including four hotels (380 accommodation units and 820 beds), two camping sites (2850 parking spaces for campers), seven restaurants, and eight other small park facilities (just under 3000 seats) [52]. The income of these activities is used for management and investments within the park area [37].

PLNP is one of the most visited natural sites in Central Europe and in the Mediterranean region [53]. The park's official statistics report a significant growth in the number of visitors per year, from 850,000 registered in 2007 to about 1.75 million in 2018. More than 80% of visitors visit the park in the period from May to September. The months of the greatest peak are July and August, when approximately 335,000 and 385,000 visitors were registered in 2017, with daily averages of about 10,800 and 12,400 visitors and reaching the maximum with over 16,000 visitors in a single day (August 2017). Consequently, the Park is often congested, causing both considerable discontent in the opinion of some visitors but above all putting safety procedures at risk and causing negative ecological impacts for the natural systems of the park [53].

3.2. Data Collection

Reviews relating to "Plitvice Lakes National Park" were scraped between October and November 2021 from the dedicated website on TripAdvisor (https://www.Tripadvisor.com/Attraction_Review-g303827-d554038-Reviews-Plitvice_Lakes_National_Park-Plitvice_Lakes_National_Park_Central_Croatia.html accessed on 26 September 2021).

WebHarvy software was used to scrape the reviews and obtain the following information:

- User data: name, origin, number of contributions (review number);
- Review data: date (month and year), travel purpose, number of bubbles (i.e., summary judgement), title and text of the review (i.e., extended judgement).

The software utilized is a visual web scraper that uses no script or code to scrape data. The program allows you to access the URL address of interest and to select the items that you want to collect. Thanks to the potential of the tool used, it was possible to carry out the immediate translation of the reviews and their respective titles by referring to the Google Translate plug-in. In this way, all of the reviews of all available languages were translated into English and used for subsequent analyses.

The study did not collect other types of socio-demographic information such as the age, occupation, and educational level of visitors. This is due to the fact that TripAdvisor profiles do not contain this kind of data [3]. The only personal information that TripAdvisor users commonly share is their country of origin. These data could be useful for analyzing the origin of visitor flows to the PLNP.

3.3. Multidimensional Scaling Method and Cluster Analysis

MDS and cluster analysis allow us to explore possible combinations or groups of words that share similar appearance patterns [22]. In particular, text clustering is a textual data mining method which converts the original sentences in a term-document-matrix using different feature extraction techniques [6,54]. In this way, it is possible to deduce the main elements perceived by the users (e.g., reviewers), which should be taken into consideration for an effective and rational management of the protected areas. The ease of analysis application and result interpretation are among the main advantages of the MDS [23,24].

The elaborations were carried out using KH Coder 3 software [25,39,54,55]. The KH Coder software combines two fundamental approaches of computer-based text analysis: the correlational approach, which consists in automatically extracting words from a text and analyzing them statistically; and the dictionary-based approach, which establishes coding rules for the different elements that form the text (e.g., sentences or groups of words) [55]. In order to identify the clusters of words, the Ward's minimum variance method or Ward's hierarchical clustering method was applied, as previously carried out by Barbierato et al. [39]. The Ward's method is a procedure that initially generates in clusters containing single objects. These clusters are gradually aggregated in such a way as to create clusters with the highest number of objects possible, but ensuring that the variance within each cluster is minimized [56]. The Ward's method was applied within the so-called Sammon space, which allows one to maintain a certain distance between words, preventing them from being excessively crowded and overlapping, giving more readable results [57]. Furthermore, among the options to define the distance, the cosine similarity coefficient was chosen, which is considered an efficient option in the presence of long documents (e.g., reviews) which contain, as in our case study, numerous words with an important frequency in each document [57]. A frequency threshold of 1500 terms was adopted on the basis of the term frequency–document frequency graph (i.e., TF–DF) (Figure 2a) in order to include exclusively the most representative terms that appear in several reviews. Based on the agglomeration graph (Figure 2b), it was chosen to generate seven clusters of 60 words each. For further information on the method, refer to the KH Coder software manual [57].
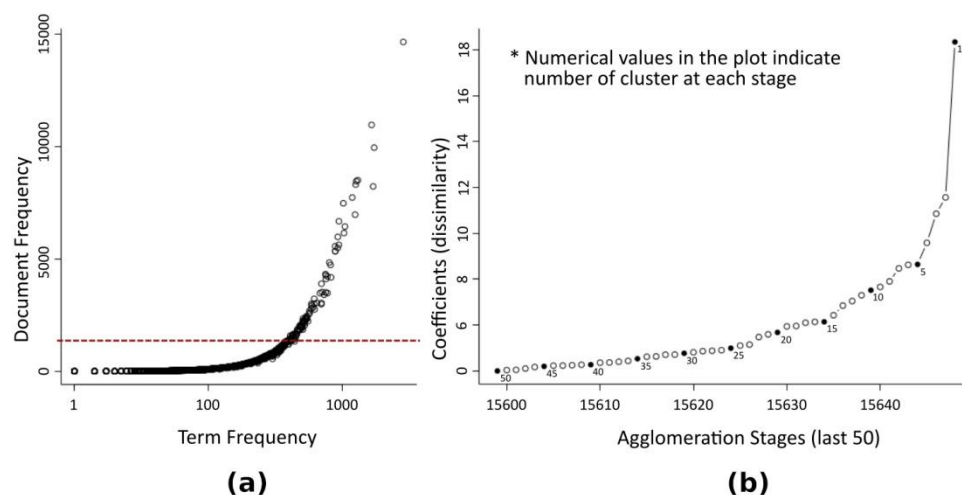
Figure 2. MDS model parameters for Plitvice Lakes National Park: TF–DF (**a**) and agglomeration graph (**b**).

3.4. Sentiment Analysis

Sentiment analysis (SA) research is driven by the importance of understanding consumer judgement [9]. In particular, SA can be used to understand consumer attitudes towards particular products, services, or places [16]. SA determines the positive or negative polarity of each relevant word in the text. Moreover, SA calculates a score based on a predefined lexicon contained within a library [39]. It should be specified that this score is not set on a reference scale between a predetermined minimum and maximum. The sentiment score varies both in reference to the text length and to the specific words contained therein. The only fixed references are the scores assigned to the individual words within the lexicon to be adopted. In the present study, the "syuzhet" library of R software was chosen, as it was applied in previous research that analyzed reviews on TripAdvisor [12,27,39]. The AFINN lexicon [58] was applied at the "syuzhet" library. Negative words and slang are commonly used in reviews on social networks (e.g., TripAdvisor). The AFINN lexicon is considered a valid option for evaluating this type of comment [59]. Furthermore, SA is widely applied to the analysis of quality perception through TripAdvisor reviews for

heritage sites and natural parks [45] and urban green areas [16]. For a more in-depth analysis of the procedure used by the software, please refer to Barbierato et al. [39].

3.5. Natural Language Processing

Natural language processing (NLP) is a technology that combines computer science and linguistics in order to interpret written texts [39]. In this study, the strengths and weaknesses of the PLNP were identified using a NLP procedure. The rapid automatic keyword extraction (RAKE) procedure is a method for extrapolating multi-word keywords from documents [60]. Candidate keywords are obtained by partitioning text through stop words (e.g., and, the, of, etc.) and phrase delimiters (e.g., ; and ,) and assigning a score to each candidate multiple keyword. Only double-word keyword candidates are searched in this study. Each of the two words that constitute the candidate keyword obtains a score that is given by the ratio between the number of times each single word co-occurs with the other word of the candidate keyword and the total frequency with which it appears by itself. The final RAKE score for the entire candidate keyword is the sum of the scores of each of the two words that form the candidate keyword [61]. The procedure was carried out through the "udpipe" library [61] of R software [62], considering only adjectives and nouns. Furthermore, only the first 20 keywords as a sequence of two adjacent words—defined as bi-grams—are considered, and a frequency threshold of 6 was adopted. In addition, the "lemma" option was chosen instead of "token". Through the lemmatization process, it is possible to group the different forms in which a word can be presented (e.g., singular and plural) in a single common voice. In this way, the various forms of the same reference word are counted as a single lemma, assuming a greater weight.

The analysis of definitely positive (bubbles > 3 and sentiment score > 0) and decidedly negative (bubbles ≤ 3 and sentiment score ≤ 0) reviews allowed us to identify strengths and weaknesses of the PLNP based on the visitor's judgement.

4. Results

4.1. Data Collection and Sample Description

Overall, 15,673 online reviews were automatically retrieved from the online review website TripAdvisor. The downloaded reviews date back to the period between 2007 and 2021.

Figure 3 shows the trend in the number of reviews registered on TripAdvisor for PLNP. This trend is considered to be related to the interest of visitors. The graph shows an important growth until 2015, followed by a slight decrease until 2019. In 2020, there is a significant drop (–88% compared to the previous year) due to the international and national restrictions on travel as a consequence of the COVID-19 pandemic.
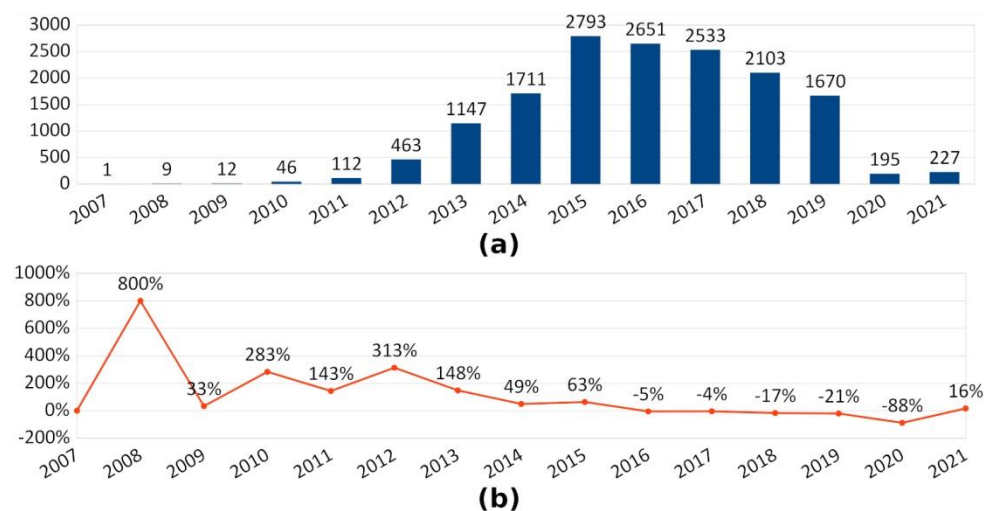
Figure 3. Frequency of reviews per year (**a**) and annual percentage growth rate of reviews (**b**).

The monthly and seasonal distribution of reviews (Figure 4) is consistent with the dynamics of visitor flows that have been analyzed in the current PLNP management plan [52]. The graph shows that in the summer—with special regard to August—the maximum peak is recorded. Instead, an intermediate influx of visitors is recorded on average in spring and autumn, even if the month of September still seems to be influenced by the importance of the summer flow. Winter is the season of least interest for visitors, as confirmed by the low number of revisions.
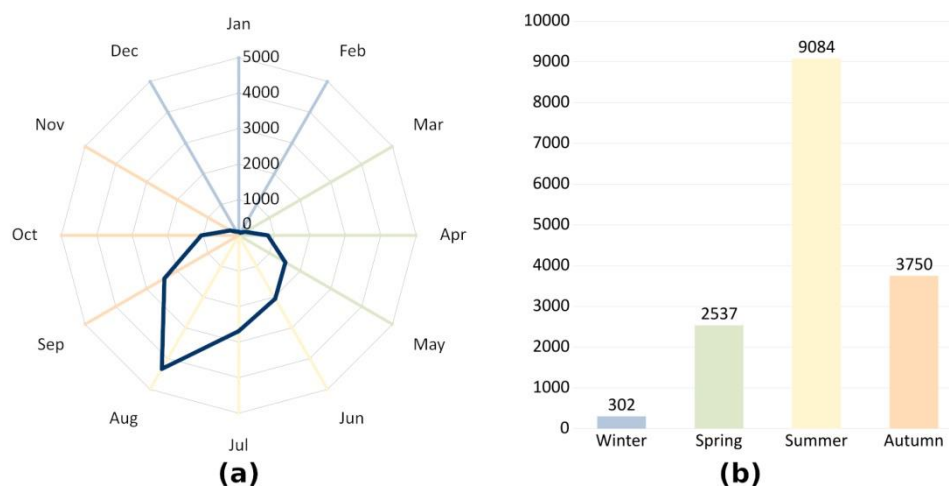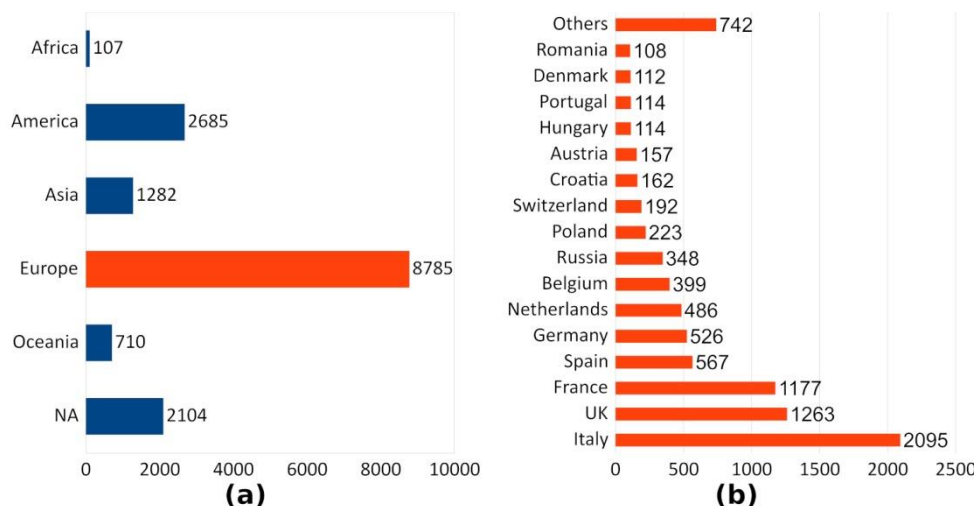

Figure 4. Monthly (**a**) and seasonal (**b**) distribution of reviews (average value for the period 2007–2021).

As regards the origin of PLNP visitors, Figure 5 shows that most of the visitors come from European countries. In particular, the largest flows are recorded from Italy, the United Kingdom, and France.

Figure 5. Provenance of the reviewers by continents (**a**) and from exclusively EU countries (**b**) (reference period 2007–2021).

4.2. Multidimensional Scaling Method and Cluster Analysis

The diagram derived from the MDS method shows seven clusters of words differentiated by color [54]. The results are in Figure 6. Cluster 1 (i.e., turquoise bubbles) concerns the principal elements that characterized PLNP landscape (i.e., "park", "lake", "waterfall") which are commonly associated with positive judgements ("beautiful"). Cluster 2 (i.e., yellow bubbles) is related to the theme of accessibility, including: the possible means of

transport to access and/or visit the park (i.e., "boat", "bus", "train", "car"); the organization into "route(s)" divided by length in terms of "hour(s)"; and the real entrance to the park, which concerns different activities, such as "parking" and the purchase of the "ticket". Cluster 3 (i.e., violet bubbles) is a hybrid set of aspects that characterize the park, emphasizing the beauty of the site on the one hand, using terms such as "nice" and "good", and the disadvantages related to overcrowding in the summer months of the high season, expressed by adjectives such as "many", "long", and "lot". Clusters 4 (i.e., red bubbles) and 6 (i.e., orange bubbles) contain the main favorable appreciations thus synthesizable: "great", "worth", "wonderful", "natural" connected to "nature", "beauty", and "experience" for Cluster 4; "stunning", "amazing", "clear", and "different" (in the positive sense of "different" landscapes and sceneries) relating in general to the "Croatia(n)" "national" park of "Plitvice" for Cluster 6. All of the positive adjectives of the Clusters 4 and 6 are also related to the nearest central terms of the Cluster 1. Cluster 5 (i.e., blue bubbles) contains the most negative elements, referring to the main problems related to the PLNP management: the presence of "crowd" and "queue(s)" in many different "point(s)", "path(s)", and "way(s)" of the area. Finally, Cluster 7 (green bubbles) represents a small deepening of the nearby Cluster 2 themes, recovering the theme of the fruition through the use of words such as "walk", "trip", and "tour". In this cluster, some information about the division in the "upper" and "lower" districts of the park are included.

Figure 6. Multidimensional scaling method and cluster analysis results for Plitvice Lakes National Park.

These results make it possible to identify the issues (i.e., the seven clusters) related to the PLNP management that are of greatest interest to visitors. The issues thus identified would be useful if applied to guide a participatory planning of the park in which samples of visitors were also involved.

4.3. Sentiment Analysis

The results of the SA are shown in Table 1. The reviews for PLNP are basically positive (mean value of 9.16) and the dispersion is relatively symmetrical (1st Qu. = 5; 3rd Qu. = 13). In fact, the mean value is shifted upwards, as the group of reviews designated with five bubbles represents over 78% of the total reviews (15,673). The SA results show that mean

and median values tend to increase with the increment in the number of bubbles (i.e., short judgement).

Table 1. Sentiment analysis scores for Plitvice Lakes National Park.

Bubbles	No. Reviews	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
◉○○○○	210	−27	−3	0	0.40	4	23
◉◉○○○	228	−19	−1	3	3.04	7	36
◉◉◉○○	641	−14	2	6	5.96	10	27
◉◉◉◉○	2317	−15	4	8	8.10	11	40
◉◉◉◉◉	12,277	−16	6	9	9.79	13	72
Total	15,673	−27	5	9	9.16	13	72

The non-normal distribution of the SA scores was visually verified through normal quantile plots, histograms, and box plots for each group related to the five review ratings (i.e., bubbles) (see Appendix A: Figures A1–A3). Furthermore, the Shapiro–Wilks test was performed for the groups of Bubbles 1, 2, 3, and 4 (in R, the Shapiro–Wilks test cannot be performed on sets of more than 5000 units). The p-value of all four groups (min $< 2.2 \times 10^{-16}$; max $= 0.002$) showed that the data do not follow a normal distribution. For this reason, the non-parametric Kruskal–Wallis test was applied to verify the correspondence between the SA scores and the bubbles assigned by the reviewers themselves.

The results confirmed the hypothesis of a statistically significant difference between the groups of bubbles in relation to the dependent variable of SA scores ($K = 848.91$; p-value $< 2.2 \times 10^{-16}$; $\alpha = 0.05$). In addition, a pairwise comparison using the non-parametric Mann–Whitney U test was conducted to highlight where the statistically significant differences between groups of bubbles are [34]. Although the differences within each pair of groups are statistically significant (Table 2), according to Barbierato et al. [39] the complete database was divided only into two sub-databases in order to simplify the data analysis: one definitely positive (bubbles > 3 and sentiment score > 0) and one decidedly negative (bubbles ≤ 3 and sentiment score ≤ 0), which were used separately in NLP analyses.

Table 2. Mann–Whitney U test ($\alpha = 0.05$) results for Plitvice Lakes National Park.

Pair of Groups of Bubbles		W	p-Value
◉○○○○	◉◉○○○	17,998	6.934×10^{-6}
◉○○○○	◉◉◉○○	33,963	$<2.2 \times 10^{-16}$
◉○○○○	◉◉◉◉○	86,522	$<2.2 \times 10^{-16}$
◉○○○○	◉◉◉◉◉	348,787	$<2.2 \times 10^{-16}$
◉◉○○○	◉◉◉○○	52,868	5.059×10^{-10}
◉◉○○○	◉◉◉◉○	141,873	$<2.2 \times 10^{-16}$
◉◉○○○	◉◉◉◉◉	584,911	$<2.2 \times 10^{-16}$
◉◉◉○○	◉◉◉◉○	589,860	1.306×10^{-15}
◉◉◉○○	◉◉◉◉◉	2,551,728	$<2.2 \times 10^{-16}$
◉◉◉◉○	◉◉◉◉◉	11,975,881	$<2.2 \times 10^{-16}$

4.4. Natural Language Processing: The RAKE Analysis

The RAKE analysis was applied to the two sub-databases obtained dividing positive from negative reviews considering the SA scores. The double-word keywords most frequently encountered in TripAdvisor reviews for PLNP were identified by the RAKE analysis (Figure 7). The most cited characteristics can be identified both in the definitely

positive reviews, to be interpreted as the main strengths, and in the decidedly negative reviews, to be read as the most critical weaknesses. Definitely positive RAKE analysis results (Figure 7a)—deriving from the sub-database containing the reviews with bubbles > 3 and sentiment score > 0—show that the natural heritage and landscape elements are the most appreciated aspects of the PLNP. In particular, the "UNESCO" designation is considered as an extremely positive characteristic, as highlighted by three keywords: "UNESCO heritage", "UNESCO site", and "UNESCO list". The negative results—deriving from the sub-database containing the reviews with bubbles $\leq$ 3 and sentiment score $\leq$ 0—show that the main weaknesses are represented by the phenomenon of crowding ("many people"), because the presence of a "mass tourism" during the "high season" is the cause of complex management problems, such as "traffic jam" and "endless queue" (Figure 7b). In addition to "long (waiting) time", there are also complaints about the organization of "parking lot" and the "high price" of the entrance ticket.
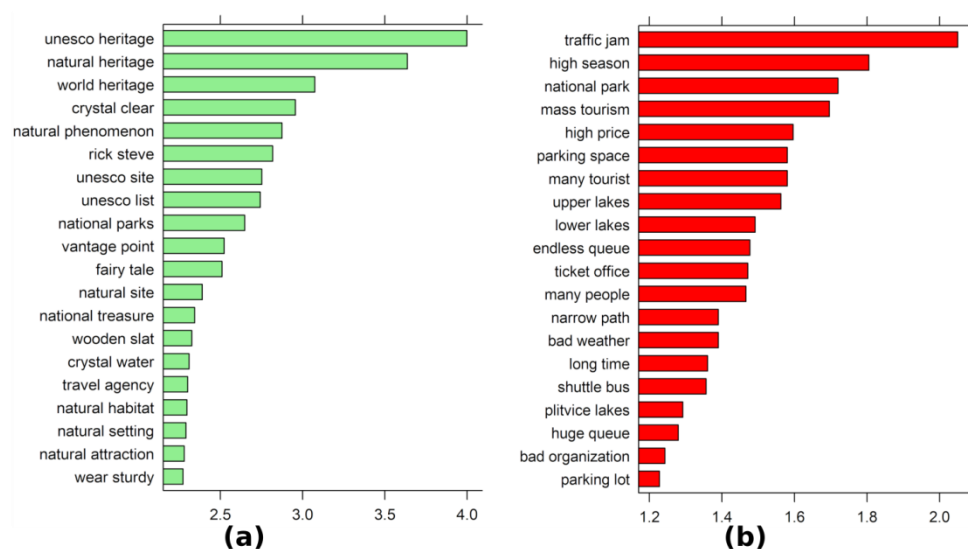
Figure 7. RAKE analysis for positive (**a**) and negative (**b**) reviews for Plitvice Lakes National Park.

5. Discussion

5.1. Answers to Research Questions

The importance of the PLNP at national and international levels is now recognized (Figures 3 and 5). The descriptive statistics highlighted the recurring seasonal trend of visits (Figure 4). This trend has made it essential to implement strategies to redistribute tourist pressure acting on the protected area in a more balanced way.

Regarding the first research question (RQ1), the research has shown that efficient tools exist as an alternative to manual coding (e.g., the software WebHarvy) to collect extensive data relating to lengthy textual reviews (e.g., TripAdvisor online platform). Moreover, the combination of CA with MDS method and cluster analysis turned out to be exhaustive to analyze visitors' preferences and perception for areas of naturalistic interest. First of all, these techniques make it possible to identify the most important symbols and attributes that characterize national parks in accordance with the visitors' opinions. The SA results (Table 1) confirm that national parks and, in general, nature-based experiences arouse positive sentiments in visitors, as already found in other studies [6,8].

MDS methods and cluster analysis are valid instruments to investigate the principal management issues from visitors' point of view (RQ2). The seven clusters identified by this study can help guide a participatory discussion on the issues that visitors consider most important for the reality of PLNP. As stated by Hausmann et al., visitors to national parks tend to idealize some particular places in their destinations, assigning them meanings that make those places worth visiting [8]. In fact, some of the naturalistic and landscape aspects of the PLNP (Cluster 1, 4, and 6, Figure 6) assume a symbolic meaning that almost

exclusively attracts the interest of visitors. The most recurring element is the complex aquatic ecosystem of lakes and waterfalls. Also Mirzaalian and Halpenny have identified this type of water elements as one of the main categories of destinations preferred by visitors and a recurring element in the reviews of naturalistic sites [6]. On the one hand, the water system represents the most important naturalistic attraction of the PLNP, but it is also the place where visitors flock the most, representing the fulcrum of tourist organizational problems. In this way, interest in high landscape and environmental or historical values of other areas of the park is excluded a priori. The most evident example is that of the large forest area which is not mentioned at all in any clusters. Other relevant aspects identified are those of accessibility and management of paths and visitors (Clusters 2, 5, and 7, Figure 6). The results obtained show that visitors are aware of and interested in discussing and expressing opinions on organizational issues related to the fruition of places, as already found by Stoleriu et al. [3]. In particular, words like "route" (Cluster 2), "experience" (Cluster 4), "path" (Cluster 5), and "walk" (Cluster 7) emphasize the attention of visitors towards active experiences (e.g., hiking or nature photography). Other studies have also identified these activities as being of great interest in the outdoor visits [25]. In addition, the organizational capacity and the entertainment activities promoted by a tourist destination is an indispensable experiential factor for all those who do not have naturalness as their primary interest [25]. In any case, the most relevant management aspect identified is the management of visitor flows and the problem of overcrowding (Cluster 3 and 5, Figure 6), which was also found by the RAKE analysis.

About the third research question (RQ3), NLP techniques proved to be fundamental to highlight strengths and weaknesses that characterize the image of PLNP. These techniques are of greater interest to identify the negative aspects to be solved and improved rather than the positive aspects to maintain and enhance. The problem of overcrowding is already widely recognized by the Plitvice Lakes National Park Management Plan 2019–2028 [50], which talks about the dissatisfaction of visitors (e.g., due to numerous encounters on the trails or impossibility of taking good photos of pristine landscapes) and the countless organizational problems (e.g., the overcoming of the physical capability of means of transport such as buses and boats or the inability to find parking) detected in the high season [53]. Visitor congestion caused by the crowds of visitors and the consequent recreational conflicts are recurring themes also in other studies focused on the use of protected areas of international interest [21,25,63]. Only a small part of the PLNP's surface represents the main focal point [37], with the "upper lake(s)" and "lower lake(s)" zones (see Figures 6 and 7), where the majority of visits are concentrated [51]. This means that an organizational and promotional effort could be conducted to make the other parts of the park more attractive with activities and guided tours. In fact, the organization of specific events, preferably connected to naturalistic aspects, are of particular interest and attract a large number of visitors as found by Mangachena and Pickering [27].

The automated text analysis processes on social media can provide park managers useful information relating to environment and organizational perception of visitors [27] with a view to collaborative and participatory planning.

5.2. Theoretical Implications

This study makes significant theoretical contributions in the management of areas of naturalistic interest. Firstly, the research demonstrates the flexibility and effectiveness in using an automated approach to obtain information from a large amount of content generated by visitors. From a methodological point of view, the web scraper software applied, WebHarvy, proved to be a valid alternative to manual coding tools. One of the most important innovations of this study is the use of reviews in different languages. In fact, the automatic translation procedure made it possible to use a large number of reviews compared to previous studies that only used reviews written in English [6,8,11,16,25,27,33,39]. Secondly, this study answers a series of research questions regarding the users' judgement on the management of areas of naturalistic interest. In fact, it was possible to identify

the topics most cited in visitor reviews, give an order of importance to their discussion, and summarize those that are considered the most important strengths and weaknesses. The study made it possible to extend the use of text mining and NLP techniques already widely applied in other research topics related to tourism in general [9,19,39,44,45] but less explored [8] in nature-based tourism [6,25,27].

Finally, the use of this innovative technique for a well-known study area of international interest (i.e., Plitvice Lakes National Park) allowed to validate the effectiveness of the tool, finding results in accordance with previous knowledge. This step will permit extending the use of the method to other less investigated areas of naturalistic interest, being able to contribute substantially to the identification of key management factors.

5.3. Practical Implications

The results show that social media analysis can be very validly applied to the nature-based tourism field [8]. In particular, these techniques can help decision makers and managers to interpret the online image of national parks constructed by visitors [3,8]. CA—with special regard to SA—effectively identifies negative trends in online reviews, making the tourism operators of national parks capable of being proactive and developing targeted strategies [9]. On the one hand, the method adopted makes it possible to monitor the perception of visitors' recreational experiences in order to plan attractive and well-organized tourist activities. On the other hand, the need to create protected areas and implement conservation and enhancement strategies within them would be supported by similar results [8,53]. In fact, the results of this study demonstrate the high interest and involvement that visitors have towards these very popular tourist destinations. Furthermore, starting from the results obtained, social media could be used by tourism actors (e.g., park managers, tour operators, etc.) to communicate their strategies and marketing proposals to consumers [6]. In particular, for the PLNP both the topics of greatest interest treated by visitors in their reviews and the less contemplated elements are identified, thanks to the use of the methodology adopted. Particularly, the forest ecosystem is not taken into consideration by the visitor reviews, while it would represent the largest percentage of the park area. In line with what has been identified in the current Management Plan [52], it becomes essential to enrich the program of visits with activities that encourage the exploration of all areas of the park. For example, experiences of great interest [25], such as group excursions or guided naturalistic visits, could generate greater appreciation for the complexity of the park's natural systems other than the aquatic ones already widely known. Given the importance attached by visitors to events and special occasions, a further solution to improve the management of the PLNP could be to organize theme-days, highly appreciated by visitors to national parks [27], in order to attract tourists even in less crowded periods, for example, during the winter season, and, therefore, reduce the pressure of the summer season. The PLNP managers could monitor the effectiveness in the proposal of the new visiting programs and events by repeating in the future an analysis of the TripAdvisor reviews with the method adopted in this study in order to search for the presence or absence of the "forests" theme among the interests of visitors.

Thus, in general, from a managerial point of view, these findings can help PLNP managers to better understand visitors' preferences. Furthermore, in this way, managers can more consciously decide which aspects to devote more attention to and how to best redistribute investments to ensure visitor satisfaction.

5.4. Limitations and Future Research

Through the use of social media, it is possible to involve visitors in a first level of participation for protected natural resource management, that of information gathering. In fact, it is extremely complex to include visitors in the subsequent steps of the process, first of all, because it would be necessary to involve very large samples to be representative for the entire population and, secondly, because it is difficult to find simple and adequate channels to contact and interview so many people. Conversely, one of the most relevant

advantages is due to the opportunity to carry out investigations on very large samples at extremely low costs. It is also true that other social media (e.g., Instagram and Twitter) allow analysis on a larger scale [8,27], even if they reported some difficulties in processing much shorter texts with a definitely lower amount of information [27].

In the present study, in order to obtain a consistent sample (15,673 online reviews) it was decided to use TripAdvisor reviews on the PLNP issued over a long period (2007–2021). Future research could investigate shorter periods of time to analyze the evolutionary dynamics of the park as well as the effectiveness of the different management strategies used over the years. Furthermore, it must be said that the analysis was restricted to a single Croatian National Park, even if it is the best known (i.e., PLNP). A further study could be, for example, that of a broader analysis of the overall network of national parks that would make it possible to systematize the monitoring and management of protected areas based on a shared investigation effort. It should also be noted that the study presents some biases related to the habits of people in the use of social media. In fact, it has been demonstrated that social media are mostly used among younger people [8,32], which highlights the fact that the analyzed sample is not representative of some categories of people (i.e., children and elderly). The absence of socio-demographic information from TripAdvisor users does not allow for more extensive surveys on the characteristics of the sample [3], while it would be advisable to analyze the preferences of visitors based on their personal characteristics through subsequent in-depth surveys. In fact, it has not been forgotten that the combination of current and traditional survey methods certainly allows the carrying out of very extensive investigations but also allows one to deepen some aspects of the issue in detail [3]. Likewise, it is assumed that all reviews analyzed come from honest opinions of visitors. However, this assumption may not be true, as fake reviews are not uncommon, and it is likely that some of them were included in the sample used in this as well as other sector studies [19]. Since that of natural areas, and in particular of national parks, is a topic not yet particularly deepened in the CA field [3], it could be useful to develop a recreational dictionary specific for national parks that can improve the accuracy of the analysis of the text thanks to the reference to specific terms for the description of the perception of natural environments [8]. Finally, future research could exploit the information available relating to the country of provenance in order to investigate the different preferences expressed by visitors from diverse geographic clusters [27], which have not been investigated in this study.

Despite the above-mentioned limitations, it is believed that the research conducted can be a reliable and useful starting point in the context of tourism analyses to deepen the opinions of the users of the areas of naturalistic interest and extrapolate from their reviews important information for better planning of management activities.

6. Conclusions

The present study investigated the strengths and weaknesses of the PLNP through a large sample of visitor reviews. The results demonstrated the flexibility and effectiveness of applying the developed method to unstructured textual data of online reviews. The present study contributes to fill a research gap in visitor perception analysis for natural areas. The management of the forest area of the PLNP is complex, as it must combine the conservation of natural ecosystems and the tourist destination promotion. In other words, the management must consider the trade-off between the tourism-recreation function and other ecosystem services. The combined use of different and complementary techniques allowed us to develop two research branches in parallel. In the first, the sentiment analysis scores were used to implement a natural language processing technique (i.e., RAKE analysis) from which the strengths and weaknesses of the PLNP have been extrapolated from the visitors' point of view. In the second, the multidimensional scaling method and cluster analysis were used to identify the key topics covered in visitors' reviews. In accordance with the latter result, it might be appropriate to involve visitors in a more in-depth investigation so as to collect visitors' opinions on the priorities defined by the park managers. Despite

the limitations encountered, the social media data analysis turns out to be an exhaustive investigation method capable of providing useful information. On the one hand, theoretical advantages can be achieved, contributing in the field of research to the definition of increasingly in-depth and efficient survey tools, and, on the other hand, it is possible to obtain practical information to be provided to the figures who deal with the management and planning related to protected natural areas.

Appendix A

The non-normal distribution of the sentiment analysis scores was visually verified in the following graphs.
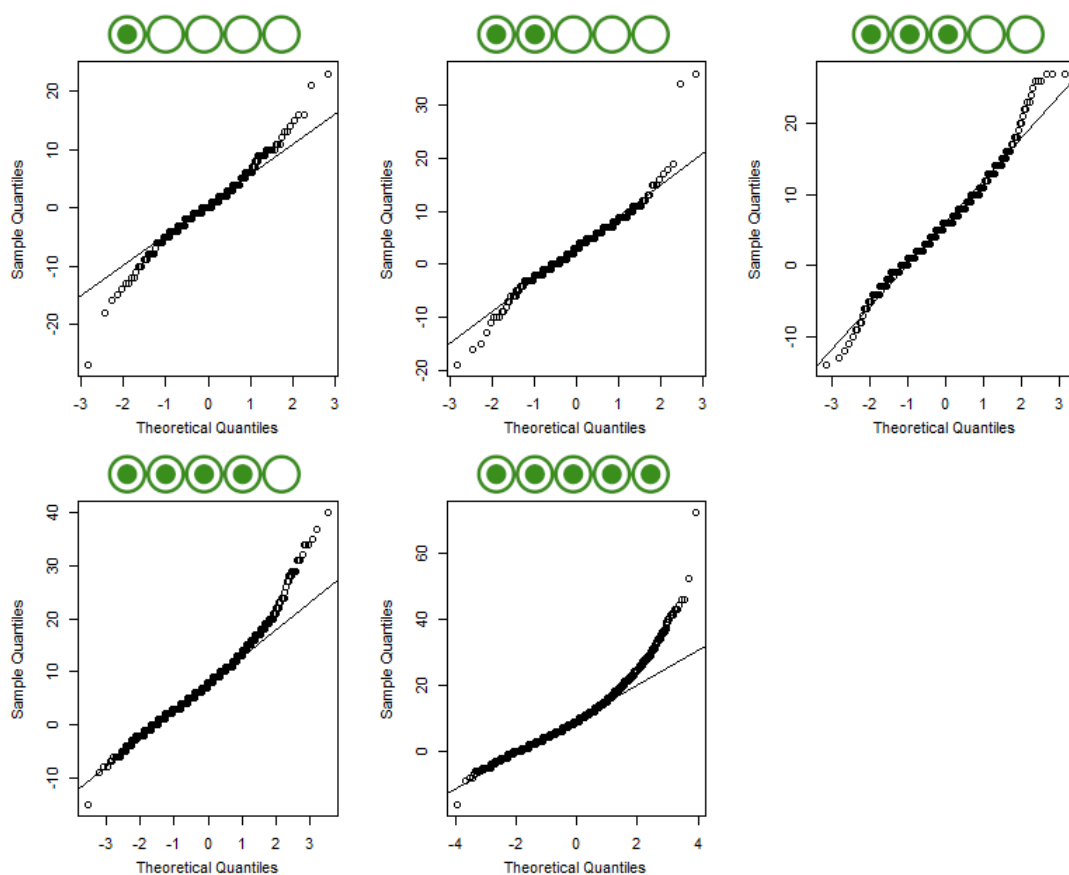
Figure A1. Quantile-quantile plots of the variable "score" for the five groups of bubbles.
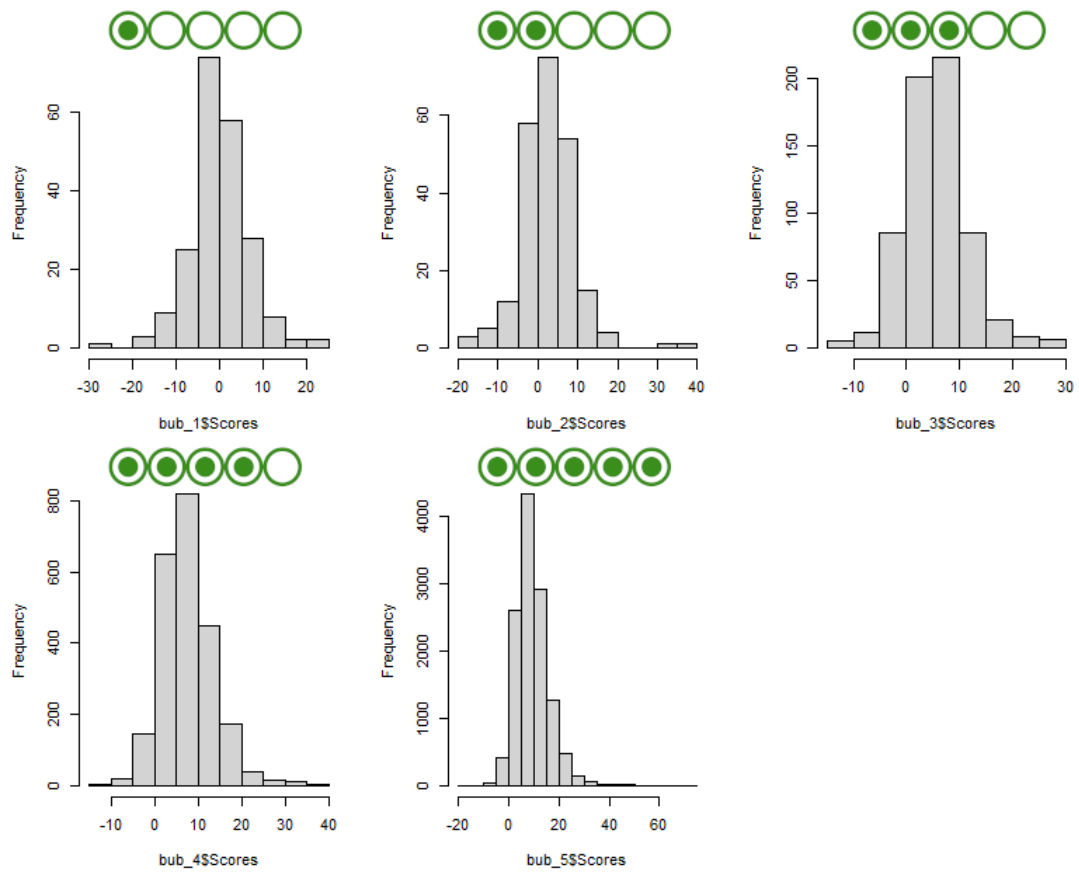
Figure A2. Histograms of the variable "score" for the five groups of bubbles.
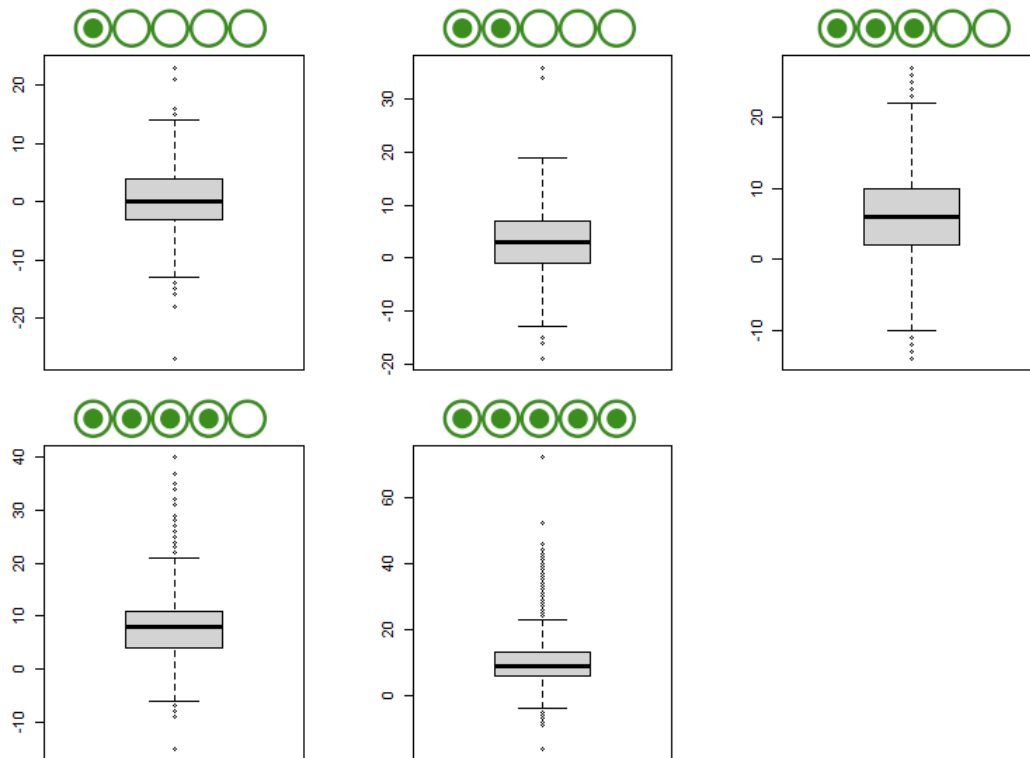

Figure A3. Box plots of the variable "score" for the five groups of bubbles.

References

1. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [CrossRef]
2. Zhang, Y.; Cole, S.T. Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts. *Tour. Manag.* **2016**, *53*, 13–27. [CrossRef]
3. Stoleriu, O.M.; Brochado, A.; Rusu, A.; Lupu, C. Analyses of Visitors' Experiences in a Natural World Heritage Site Based on TripAdvisor Reviews. *Visit. Stud.* **2019**, *22*, 192–212. [CrossRef]
4. Lai, L.S.L.; To, W.M. Content analysis of social media: A grounded theory approach. *J. Electron. Commer. Res.* **2015**, *16*, 138–152.
5. Tenkanen, H.; Di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci. Rep.* **2017**, *7*, 1–11. [CrossRef]
6. Mirzaalian, F.; Halpenny, E. Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *J. Destin. Mark. Manag.* **2021**, *20*, 100598. [CrossRef]
7. Yang, M.; Han, C. Revealing industry challenge and business response to Covid-19: A text mining approach. *Int. J. Contemp. Hosp. Manag.* **2020**, *33*, 1230–1248. [CrossRef]
8. Hausmann, A.; Toivonen, T.; Fink, C.; Heikinheimo, V.; Kulkarni, R.; Tenkanen, H.; Di Minin, E. Understanding sentiment of national park visitors from social media data. *People Nat.* **2020**, *2*, 750–760. [CrossRef]
9. Ma, E.; Cheng, M.; Hsiao, A. Sentiment analysis—A review and agenda for future research in hospitality contexts. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3287–3308. [CrossRef]
10. Xiang, Z.; Du, Q.; Ma, Y.; Fan, W. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tour. Manag.* **2017**, *58*, 51–65. [CrossRef]
11. Yu, Y.; Li, X.; Jai, T.M. (Catherine) The impact of green experience on customer satisfaction: Evidence from TripAdvisor. *Int. J. Contemp. Hosp. Manag.* **2017**, *29*, 1340–1361. [CrossRef]
12. Valdivia, A.; Luzón, M.V.; Herrera, F. Sentiment Analysis in TripAdvisor. *IEEE Intell. Syst.* **2017**, *32*, 72–77. [CrossRef]
13. Blackshaw, P.; Nazzaro, M. *Consumer Generated Media (CGM) 101: Word-of-Mouth in the Age of Web-Fortified Consumer*; Nielsen, A., Ed.; BuzzMetrics White Pap: Diepoldsau, Switzerland, 2006.
14. Khan, M.F.; Jan, A. Social Media and Social Media Marketing: A Literature Review. *IOSR J. Bus. Manag.* **2015**, *17*, 12–15. [CrossRef]
15. Filieri, R.; Acikgoz, F.; Ndou, V.; Dwivedi, Y. Is TripAdvisor still relevant? The influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. *Int. J. Contemp. Hosp. Manag.* **2021**, *33*, 199–223. [CrossRef]
16. Ghahramani, M.; Galle, N.J.; Duarte, F.; Ratti, C.; Pilla, F. Leveraging artificial intelligence to analyze citizens' opinions on urban green space. *City Environ. Interact.* **2021**, *10*, 100058. [CrossRef]
17. Tripadvisor Tripadvisor Investor Relations. Available online: https://ir.tripadvisor.com/ (accessed on 11 January 2022).
18. Aicher, J.; Asiimwe, F.; Batchuluun, B.; Hauschild, M.; Martina, Z.; Egger, R. *Information and Communication Technologies in Tourism 2016*; Springer: Cham, Switzerland, 2016; pp. 369–382. [CrossRef]
19. Pezenka, I.; Weismayer, C. Which factors influence locals' and visitors' overall restaurant evaluations? *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 2793–2812. [CrossRef]
20. Roberts, H.; Resch, B.; Sadler, J.; Chapman, L.; Petutschnig, A.; Zimmer, S. Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Plan.* **2018**, *3*, 21–33. [CrossRef]
21. Huai, S.; Van de Voorde, T. Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. *Landsc. Urban Plan.* **2022**, *218*, 104307. [CrossRef]
22. Borg, I.; Groenen, P.J.F.; Mair, P. *Applied Multidimensional Scaling and Unfolding*, 2nd ed.; Springer: Cham, Switzerland, 2018; ISBN 978-3-319-73470-5.
23. Marcussen, C. Multidimensional scaling in tourism literature. *Tour. Manag. Perspect.* **2014**, *12*, 31–40. [CrossRef]
24. Chhetri, P.; Arrowsmith, C.; Jackson, M. Determining hiking experiences in nature-based tourist destinations. *Tour. Manag.* **2004**, *25*, 31–43. [CrossRef]
25. Niezgoda, A.; Nowacki, M. Experiencing nature: Physical activity, beauty and tension in Tatra National Park-analysis of tripadvisor reviews. *Sustainability* **2020**, *12*, 601. [CrossRef]
26. Kaffashi, S.; Radam, A.; Shamsudin, M.N.; Yacob, M.R.; Nordin, N.H. Ecological conservation, ecotourism, and sustainable management: The case of Penang National Park. *Forests* **2015**, *6*, 2345–2370. [CrossRef]
27. Mangachena, J.R.; Pickering, C.M. Implications of social media discourse for managing national parks in South Africa. *J. Environ. Manag.* **2021**, *285*, 112159. [CrossRef] [PubMed]
28. Duffy, R. Nature-Based Tourism and Neoliberalism: Concealing Contradictions. *Tour. Geogr.* **2015**, *17*, 529–543. [CrossRef]
29. Balmford, A.; Beresford, J.; Green, J.; Naidoo, R.; Walpole, M.; Manica, A. A global perspective on trends in nature-based tourism. *PLoS Biol.* **2009**, *7*, e1000144. [CrossRef] [PubMed]
30. Dedeke, A.N. Creating sustainable tourism ventures in protected areas: An actor-network theory analysis. *Tour. Manag.* **2017**, *61*, 161–172. [CrossRef]
31. Schwartz, A.J.; Dodds, P.S.; O'Neil-Dunne, J.P.M.; Danforth, C.M.; Ricketts, T.H. Visitors to urban greenspace have higher sentiment and lower negativity on Twitter. *People Nat.* **2019**, *1*, 476–485. [CrossRef]

32. Kovacs-Györi, A.; Ristea, A.; Kolcsar, R.; Resch, B.; Crivellari, A.; Blaschke, T. Beyond spatial proximity-classifying parks and their visitors in london based on spatiotemporal and sentiment analysis of twitter data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 378. [CrossRef]

33. Plunz, R.A.; Zhou, Y.; Carrasco Vintimilla, M.I.; Mckeown, K.; Yu, T.; Uguccioni, L.; Sutto, M.P. Twitter sentiment in New York City parks as measure of well-being. *Landsc. Urban Plan.* **2019**, *189*, 235–246. [CrossRef]

34. Zhang, T.; Zhang, W.; Meng, H.; Zhang, Z. Analyzing visitors' preferences and evaluation of satisfaction based on different attributes, with forest trails in the Akasawa National Recreational Forest, Central Japan. *Forests* **2019**, *10*, 431. [CrossRef]

35. MEA Millenium Ecosystem Assessment (MA). *Ecosystems and Human Well-Being: Synthesis*; Island Press: Washington, DC, USA, 2005.

36. Hausmann, A.; Slotow, R.; Burns, J.K.; Di Minin, E. The ecosystem service of sense of place: Benefits for human well-being and biodiversity conservation. *Environ. Conserv.* **2016**, *43*, 117–127. [CrossRef]

37. Vurnek, M.; Brozinčević, A.; Čulinović, K.; Novosel, A. Challenges in the Management of Plitvice Lakes National Park, Republic of Croatia. In *National Parks—Management and Conservation*; Suratman, M.N., Ed.; IntechOpen: London, UK, 2018; pp. 55–72.

38. Winter, P.L.; Selin, S.; Cerveny, L.; Bricker, K. Outdoor recreation, nature-based tourism, and sustainability. *Sustainability* **2020**, *12*, 81. [CrossRef]

39. Barbierato, E.; Bernetti, I.; Capecchi, I. Analyzing TripAdvisor reviews of wine tours: An approach based on text mining and sentiment analysis. *Int. J. Wine Bus. Res.* **2021**, *1751*, 1062. [CrossRef]

40. Gupta, V.; Lehal, G.S. A Survey of Text Mining Techniques and Applications. *J. Emerg. Technol. Web Intell.* **2009**, *1*, 60–76. [CrossRef]

41. Gan, Q.; Ferns, B.H.; Yu, Y.; Jin, L. A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews. *J. Qual. Assur. Hosp. Tour.* **2017**, *18*, 465–492. [CrossRef]

42. Zuheros, C.; Martínez-Cámara, E.; Herrera-Viedma, E.; Herrera, F. Sentiment Analysis based Multi-Person Multi-criteria Decision Making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews. *Inf. Fusion* **2021**, *68*, 22–36. [CrossRef]

43. Li, H.; Liu, Y.; Tan, C.W.; Hu, F. Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 1713–1735. [CrossRef]

44. Serrano, L.; Ariza-Montes, A.; Nader, M.; Sianes, A.; Law, R. Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: A text mining approach. *J. Sustain. Tour.* **2021**, *29*, 1134–1152. [CrossRef]

45. Lee, J.; Benjamin, S.; Childs, M. Unpacking the Emotions behind TripAdvisor Travel Reviews: The Case Study of Gatlinburg, Tennessee. *Int. J. Hosp. Tour. Adm.* **2020**, *23*, 347–364. [CrossRef]

46. Grandi, R.; Neri, F. Sentiment Analysis and City Branding. In *New Trends in Databases and Information Systems*; Springer: Cham, Switzerland, 2014; Volume 241, pp. 339–349. ISBN 9783319018621.

47. Pearce, P.L.; Wu, M.Y. Entertaining International Tourists: An Empirical Study of an Iconic Site in China. *J. Hosp. Tour. Res.* **2015**, *42*, 772–792. [CrossRef]

48. Song, Y.; Fernandez, J.; Wang, T. Understanding perceived site qualities and experiences of urban public spaces: A case study of social media reviews in Bryant Park, New York city. *Sustainability* **2020**, *12*, 36. [CrossRef]

49. Council of Europe. *European Landscape Convention*; Council of Europe: Florence, Italy, 2000.

50. Koblet, O.; Purves, R.S. From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally. *Landsc. Urban Plan.* **2020**, *197*, 103757. [CrossRef]

51. Mandić, A. Protected area management effectiveness and COVID-19: The case of Plitvice Lakes National Park, Croatia. *J. Outdoor Recreat. Tour.* **2021**, 100397. [CrossRef]

52. *Plitvice Lakes National Park Management Plan 2019–2028*; Plitvice Lakes National Park Public Institution: Zagreb, Croatia, 2019; ISBN 9789534850350.

53. McCool, S.F.; Eagles, P.F.J.; Skunca, O.; Vukadin, V.; Besancon, C.; Novosel, A. Integrating Marketing and Management Planning for Outstanding Visitor Experiences in a Turbulent Era: The Case of Plitvice Lakes National Park. In *Mediterranean Protected Areas in the Era of Overtourism*; Springer: Cham, Switzerland, 2021; pp. 221–240. [CrossRef]

54. Nattuthurai, P. Content Analysis of Dark Net Academic Journals from 2010–2017 Using KH Coder. *ACET J. Comput. Educ. Res.* **2017**, *12*, 1–10.

55. Hirahara, S. Evaluation of a structure providing cultural ecosystem services in forest recreation: Quantitative text analysis of essays by participants. *Forests* **2021**, *12*, 1546. [CrossRef]

56. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]

57. Higuchi, K. *KH Coder 3 Reference Manual*; Ritsumeikan University: Kyoto, Japan, 2016; p. 99.

58. Nielsen, F.Å. AFINN Word Database an Affective Lexicon. Available online: https://www2.imm.dtu.dk/pubdb/pubs/6010-full.html (accessed on 17 December 2021).

59. Naldi, M. A review of sentiment computation methods with R packages. *arXiv* **2019**, arXiv:1901.08319.

60. Rose, S.; Engel, D.; Cramer, N.; Cowley, W. Automatic keyword extraction from individual document. In *Text Mining: Applications and Theory*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2010; pp. 1–277. ISBN 9780470749821.

61. Wijffels, J. *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the "UDPipe" "NLP" Toolkit*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

62. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

63. Mirzaalian, F.; Halpenny, E. Multi-Dimensional Explorations into Visitors' Experience Sharing through TripAdvisor Using Social Media Analytics: An Investigation on Jasper National Park. In Proceedings of the TTRA Canada 2018 Conference, Halifax, NS, Canada, 25–28 September 2018.