




Article

Exploring User-Generated Content for Improving Destination Knowledge: The Case of Two World Heritage Cities

Nuno Antonio ^{1,2,*} , Marisol B. Correia ^{2,3,4,5}  and Filipa Perdigão Ribeiro ^{2,3} 

¹ Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

² Centre for Tourism Research, Development and Innovation (CiTUR), Campus of Penha, Portugal; mcorreia@ualg.pt (M.B.C.); fperdig@ualg.pt (F.P.R.)

³ School of Management, Hospitality and Tourism (ESGHT), Universidade do Algarve, Campus of Penha, 8005-139 Faro, Portugal

⁴ Research Centre for Tourism, Sustainability and Well-being (CinTurs), Campus of Gambelas, 8005-139 Faro, Portugal

⁵ CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

* Correspondence: nantonio@novaims.unl.pt

Received: 23 September 2020; Accepted: 12 November 2020; Published: 19 November 2020



Abstract: This study explores two World Heritage Sites (WHS) as tourism destinations by applying several uncommon techniques in these settings: Smart Tourism Analytics, namely Text mining, Sentiment Analysis, and Market Basket Analysis, to highlight patterns according to attraction, nationality, and repeated visits. Salamanca (Spain) and Coimbra (Portugal) are analyzed and compared based on 8,638 online travel reviews (OTR), from TripAdvisor (2017–2018). Findings show that WHS reputation does not seem to be relevant to visitors-reviewers. Additionally, keyword extraction reveals that the reviews do not differ from language to language or from city to city, and it was also possible to identify several keywords related to history and heritage; in particular, architectural styles, names of kings, and places. The study identifies topics that could be used by destination management organizations to promote these cities, highlights the advantages of applying a data science approach, and confirms the rich information value of OTRs as a tool to (re)position the destination according to smart tourism design tenets.

Keywords: user-generated content; data science; text mining; sentiment analysis; eWOM; UNESCO heritage sites

1. Introduction

From the beginning of the Web 2.0 platforms, we have witnessed an increase in travelers' and visitors' active participation when reviewing their travels or visits on social media. Studies point to the significant impacts on companies, organizations, and stakeholders of heritage and cultural tourism of the prevalence of participatory, interactive, and dynamic internet, centered on the consumer [1]. From the perspective of demand, these technologies allow tourists to access more information and more knowledge and control many aspects of their trip [2,3]. Consequently, platforms have radically transformed the way tourists perceive and engage with heritage [4]. One of the fundamental aspects of this involvement is the fact that tourists use new technologies to constantly share their experiences and impressions of the places they visit, thus competing with promotional materials and strongly influencing the decision-making process [5–7].

When we juxtapose cultural and heritage tourism and Web 2.0, there is still a lot of theoretically robust research to be developed [7]. In fact, there is an information gap between what visitors want and what they find during their trips. Additionally, from the perspective of management and marketing, it is necessary to understand who the UNESCO World Heritage tourists are in order to offer positive experiences without downplaying the much-needed support for place conservation [8]. Thus, one of the fundamental questions when carrying out tourist research on historical urban spaces is understanding what motivates individuals and families to visit heritage cities. This study attempts to answer this question through tourists' own reviews on online travel—research that, to date, does not seem to have been conducted, either in Portugal or elsewhere. On the other hand, comparative studies, which go beyond circumscribed local contexts, are essential to confront and unveil the similarities of what is often “thought to be historical-cultural particularisms and the uniqueness of local heritage” [9].

The general objective of this study is to emphasize the value of comprehensively exploring and understanding the main concerns and opinions the visitors' reviews portray about the sites they visit, based on the collection and exploitation of data available on TripAdvisor social media, so that those responsible for heritage management can arrive at informed decisions. As such, this study explores two World Heritage Sites (WHS) as tourism destinations—Salamanca in Spain and Coimbra in Portugal—by applying Data Science techniques such as Text Mining, Sentiment Analysis, and Market Basket Analysis to highlight patterns according to attraction, nationality, and repeated visits.

As a specific objective, we present a mixed exploratory methodology for the comparative analysis of an OTR (Online Travel Review) corpus of 8638 reviews in three languages: English, Spanish and Portuguese, published by visitors to UNESCO heritage cities—Coimbra, Portugal, and Salamanca, Spain, in 2017 and 2018. The two cities were chosen because they have common characteristics, as described below, and these three languages were chosen because they correspond to the languages of the relevant tourist markets in the two countries [10,11].

Our findings will provide yet another instrument for those responsible for heritage management to define objectives and strategies for the sites' sustainable promotion. Our research is based on the collection and exploitation of data available on TripAdvisor. It should be noted that in 2009, UNESCO and TripAdvisor launched a partnership to mobilize the support of visitors to preserve natural and cultural sites inscribed on the UNESCO World Heritage List [12].

From the perspective of research and analysis techniques and methods, there has been a growing interest in mixed methodological approaches, which encourage interdisciplinarity in studies in the area of Tourism and Hospitality [13,14]. However, concerning research with large amounts of textual data collected from social media and other forms of computerized media coverage, there seems to be an overall tendency for quantitative methods, which implies less informational wealth and a less holistic understanding of the phenomenon under analysis [15]. There are interesting exceptions, such as Rodrigues, Brochado, and Troilo [16], who focus on tourists' wellbeing by applying sentiment analysis to post-experience booking.com reviews. Thus, the specific aim of this study is to present how a mixed exploratory methodology for the analysis of an OTR corpus can be used to uncover useful insights which can be used by Destination Management Organizations (DMO) from UNESCO heritage sites to promote the sites or to improve the visitors' experience. Reviews published in 2017 and 2018, extracted from two WHS – Coimbra (Portugal) and Salamanca (Spain), were used to assess our methodological approach. The authors selected these cities due to their proximity, familiarity, native languages, markets' importance and place of residence.

The following research questions guided this study:

Q1—Is the UNESCO heritage classification a factor of attraction?

Q2—What are the perceptions of visitors?

Q3—How can Salamanca and Coimbra take advantage of the visitors' perceptions?

The two cities have common characteristics—both have been university cities since their foundation in the 13th century and both are prominent political and religious centers in each country's history.

Their tourist attractiveness depends mostly on these characteristics [17]. Also, both cities are classified as UNESCO heritage cities: Salamanca since 1998 and Coimbra since 2013.

Given Salamanca and Coimbra's geographical location, which are only 304 km apart (a car journey of just over three hours), and their similarity, both cities can be considered direct competitors in terms of tourism destination [17]. Thus, choosing these two cities seems to be appropriate to develop a comparative analysis, especially since some studies point to tourist flows between the two cities [17]. Not only do tourists visit these two cities, but they also visit multiple destinations in Europe in one trip. This is not a new phenomenon—the Grand tour and the Inter-rail are examples of this travel concept—and it is increasingly frequent and accessible to significant segments of the population, not only from Europe, but also from Asia and America. Other changes in travel patterns include a higher rate of individual or independently planned trips (as opposed to group planned), immersion in a specific region, and the growth of thematic trips [18]. These trends are particularly facilitated by online travel planning tools and naturally by OTRs.

OTRs are undoubtedly crucial in the decision-making process when choosing a destination. Consumers, more than ever, rely on choices through the internet. For tourist consumption, this medium is essential in purchasing, selecting destinations, and choosing what to visit at the destination itself [3]. Indeed, recent studies in marketing have been dedicated to exploring the influence of OTR on future travelers, from the perspective of those who will travel and how they interpret and assimilate the review contents [19]. Less frequent are studies focused on the reviews themselves or on reviews assessing visited places [20]. From the tourist product and evaluation perspective, most of the analyses have typically focused on OTRs related to hotels and published on Booking.com or TripAdvisor.

This paper follows the following structure: the next section summarizes the advantages of integrating the UNESCO List and briefly describes Salamanca and Coimbra as tourism destinations. This section is followed by a review of Web 2.0, eWOM (Electronic Word of Mouth), the TripAdvisor social media, and its importance for tourists and contemporary travelers. The following section presents the methodology, and the various methods proposed are illustrated with some results. This is then followed by two sections on the findings and conclusions.

2. UNESCO World Heritage List, Coimbra and Salamanca

Several studies show that the UNESCO listed heritage causes increased tourist flow and often brings more support from national and local governments for preservation or maintenance [21]. However, this support may vary according to different national contexts. In many cases, it may be a blessing in disguise [22]. One of UNESCO's objectives is to implement unique visiting experiences while simultaneously creating social and economic benefits for the places themselves [2,23]. Nonetheless, other impacts exist, such as changes in access to places or different uses of those places. Thus, it becomes necessary to systematically assess UNESCO heritage's effects to manage the sites, namely the construction of appropriate indicators, since UNESCO WHS's applications, especially since the 1990s, had as the primary objective the increase in visitor and tourism revenue [21,22]. However, studies that quantitatively measure economic impacts appear to be inconclusive [21,24].

In contrast, other studies highlight UNESCO's listing limited role and argue that the impact of being listed as world heritage depends on other contextual issues. For example, Prud'homme [25] concluded that there is an exaggerated perception of the impacts on local development. Results from the report by Research Consulting Ltd. & Trends Business Research Ltd. [26] confirm Prud'homme's findings, stating that 70% to 80% of the places appear to be doing little or nothing with the UNESCO listing to achieve significant economic impacts. For instance, Coimbra witnessed an immediate increase in visits to the University, but local businesspeople are not unanimous about the earned benefits [27]. Therefore, it would be relevant to explore if the "comparative advantage" based on Coimbra's and Salamanca's "outstanding cultural endowment included in the World Heritage List can transform their initial advantage into a competitive advantage" [28].

According to Poria, Butler, and Airey [29], the two most common reasons reported in the literature to visit a heritage site are education (i.e., the tourists' willingness to learn) and entertainment (i.e., the tourists' desire to be entertained). However, Poria et al. identified other reasons for visiting heritage sites, such as 'heritage experience', 'learning experience', and 'recreational experience', which are linked to the tourists' perception of the site in relation to their own heritage and their willingness to be exposed to an emotional experience. In a different study, Poria, Reichel, and Biran [30] identified five main motives for visiting a heritage site: learning, connecting with my heritage, leisure pursuit, bequeathing for children, and emotional involvement. This study suggested that heritage sites' managers and marketers should provide different tourists with different experiences to enhance the emotional involvement that visitors may feel at the visited site. Also, as tourist perceptions of a site may be associated with identifiable visitor characteristics (for instance, religion or nationality), this could help management identify those who perceive the site to be part of their own heritage.

The prestige associated with being listed, financial support, and access to specialized knowledge for preservation and conservation are essential benefits of the UNESCO WHS' brand. In addition, prestige acts as a catalyst for preserving heritage by governments and citizens alike [31]. The UNESCO World Heritage 'brand' value appears to be quite relevant in certain countries and communities. For example, in countries such as Portugal, the application and subsequent selection process is closely monitored by the national media and is of significant importance for enhancing cultural and national identity.

Coimbra, located in Portugal's center, on the Mondego riverbank, has about 134,156 inhabitants [32]. Data from 2017 indicate 430,076 overnight stays [33]. Its standard visitors are mostly of Portuguese origin (53.8%), followed by residents from Spain (13.6%), and Italy (6.2%) [9]. In 2017, the university registered more than 500,000 visitors, 20% of whom were French, 10% Brazilian, Spanish, Italian, and Portuguese [34].

Salamanca is a medium-sized Spanish city in the Iberian Peninsula interior, with 144,949 inhabitants [35]. In 1988 Salamanca was listed as UNESCO World Heritage and, in 2002, was branded as the European City of Culture. These events strongly contributed to the development of tourism, which was supported by the unique value of the city's unique architecture, with Roman, Medieval, Renaissance, and Baroque heritage, and of which the cathedral, the Plaza Mayor, and the University stand out [36]. In 2017, there was a total of 1,103,176 overnight stays [37].

3. User Generated Content (UGC) and Electronic Word of Mouth (eWOM)

Web 2.0 refers to the transition between the static HTML pages that emerged with Web 1.0 to the dynamic and interactive web pages. Web 2.0, also called Social Web or Social Era, consists of a participatory, interactive, and dynamic internet, focused on the consumer, whose contents are generated by the users (User Generated Content or UGC), that is, the users of the information are simultaneously consumers [38]. In prosumers' role, *consumer participation*, i.e., as active consumers and producers of information, results in UGC and opens new dimensions of communication [39]. According to [3], UGCs are considered more reliable than official sources because they are considered genuine and not focused on business.

OTRs are UGC forms that reflect evaluations and reviews about the visitor's own experience and destination. UGC can equate eWOM (Electronic Word of Mouth) and are crucial in shaping destinations' image and reputation [3]. When a destination exceeds visitors' expectations, they are likely to be motivated to share their positive experiences with others. On the other hand, disappointed visitors can use negative eWOM to spread their negative emotions. For this reason, tourism managers and stakeholders should take into consideration online reputation for understanding what visitors write about the products/services and brand traits. As Buhalis and Inversini ([39], p. 25) state, "online reputation should be understood as an asset for companies and organizations. Tourism managers should start to exploit all the information available online to boost tourists' experiences". Indeed, gathering and analyzing information from the tourist's perspective is imperative to understand the

tourist's experiences and a crucial source for designing new experiences and improving existing ones [40].

Interpersonal influence and eWOM are classified as the most relevant information sources when consumers are in the decision-making process of making a purchase. This influence can be especially significant in the tourism and hospitality sector, as its intangible products are difficult to assess before consumption [41]. Collaborative and interactive platforms, e.g., 2.0 platforms, offer an accessible medium—where people can express their opinion and suggestions on any subject and where they can become co-creators—to create a space for co-creation between entities and audiences [42]. Viral marketing is especially persuasive in the tourism and hotel industry, as users of 2.0 platforms are more willing to accept and trust information from people who are similar to them. Thus, viral marketing allows OTRs to reach a broader audience [19].

The widespread use of 2.0 platforms is currently causing radical changes in promoting tourist destinations through the clear strategy of incorporating UGC. Consequently, OTRs represent a precious source of information for both travel agencies and DMOs. By analyzing reviews and evaluations by visitors, DMOs adjust and adapt the promotion of tourist destinations to the new challenges posed by this new paradigm of interaction and communication [43].

UNESCO world heritage destinations are not different from other destinations and rely on eWOM to promote themselves. For example, Bergel and Brock [44] recognize that engaged customers can help promote heritage sites and influence customers' pay perception. Also, Mehmood et al. [45] acknowledge that good sustainable practices in heritage sites improve eWOM, positively influencing the site's image and tourists' travel intention to visit the site. A further example of Luli and Kawano [46] shows that employing data mining/analytics techniques to TripAdvisor data helps uncover the psychological impact of the technological intervention on the memorial heritage's survival.

With Web 2.0, tourists or visitors present increasingly documented OTRs, with an online story-telling process that seems to help them improve their experiences. An illustrative case is the study developed by Rahmani, Gnoth, and Mather [47], which uses Corpus Linguistics (in ICT, this technique is called Natural Language Processing or Text Mining, see below) to extract the "hidden" meanings in OTRs and identify their characteristics, issues explored below.

Introduced in 2000, TripAdvisor is one of the most used 2.0 platforms today, operating as an online travel guide that offers user-generated reviews on travel-related content. In 2018, it generated approximately 730 million reviews and user reviews, covering more than eight million ads for restaurants, hotels, vacation rentals, and attractions. It has 490 million unique visitors and sends approximately 80 million emails per week [48]. This platform allows users to search for hotels, activities, restaurants, flights, cruises, cars for rent, among other categories. Among the criteria applicable to would-be reviewers, TripAdvisor advises that reviews offer reliable advice from real travelers [49]. TripAdvisor advice is based on the fact that users' content must be based on "firsthand" experiences and provide a substantial contribution to the issue at hand.

4. Methodology

4.1. Mixed Methods Approach

The use of a plurality of methods from different disciplinary areas (ICT, Tourism, and Linguistics) adapts well to research topics in the social sciences that are subjective, complex, and multidimensional. These methods allow access to the OTR's textual richness to obtain different but complementary results [50], as is the present study's case. The complementarity of the methods used also offers a more practical approach and confirms the reliability of the results [14]. Since the perspective of both the quantitative and qualitative components of reviews provides a more comprehensive understanding of people's opinions [51], in the following sections, we present the various applied techniques employed to collect and analyze both components and the respective findings.

The use of a mixed methods approach from Data Science, especially Text Mining, Data Mining, and Machine learning, also makes it possible for less biased and more consistent analysis, uncovering unknown patterns and trends. While the algorithms employ the same criteria in all analyzed texts, a human analyst can hardly maintain the same standards and objectives over time. This human fragility is even more noticeable when the volume of texts requires multiple analysts.

4.2. Data Extraction and Description

To establish comparisons, we decided to select in Coimbra and Salamanca, respectively, the ten most recommended places for tourists by tourist information sites or digital newspapers with sections dedicated to tourism. For data description and analysis, the terms *visitor*, *tourist*, or *user* are used interchangeably.

The data set was extracted from TripAdvisor in April 2019 through a web robot, or simply bot, an application that performs tasks automatically [52]. With this bot, explicitly built for the C# programming language, 8,638 OTRs were extracted, relative to the reviews published in 2017 and 2018. The frequency and distribution of the OTRs extracted by city and language can be seen in Table 1.

Table 1. Frequency and distribution of OTR by city and language.

		English	Spanish	Portuguese	TOTAL
Coimbra	Frequency	884	465	1643	2992
	Distribution (in relation to the total)	10.2%	5.3%	19.0%	34.6%
	Distribution (in relation to the city)	29.6%	15.5%	54.9%	100.0%
Salamanca	Frequency	1162	3897	587	5646
	Distribution (in relation to the total)	13.5%	45.1%	6.8%	65.4%
	Distribution (in relation to the city)	20.6%	69.0%	10.4%	100.0%
TOTAL	Frequency	2046	4362	2230	8638
	Distribution (in relation to the total)	23.7%	50.5%	25.8%	100.0%

The data set on an Excel file was organized according to the following labels:

City: name of the city (Coimbra or Salamanca);

FullText: text written by the user;

GlobalRating: quantitative rating for each location at the date of extraction, on a scale from 1 to 5, with 3 as the average value and 5 being excellent [53];

Language: language in which the review was written;

Location: user's place of residence (registered on TripAdvisor);

Name: user name. In order to post content and reviews, users have to create a profile on TripAdvisor. This name must be associated with a name/pseudonym;

PublishDate: date in which the review was published;

ReviewRating: quantitative rating attributed by the user;

SiteDesignation: name of the site.

4.3. Data Analysis

Although OTRs have two evaluation components, the quantitative ratings and the qualitative text written by the user [54], most research on OTRs focuses solely on the quantitative component, even though the qualitative component can provide a richer overview of OTRs [55]. For this reason, to analyze both components of the data, it was decided to use a diverse set of methods, such as Data Mining, Association Rules, Natural Language Processing, Text Mining, and Textual Analysis. This last analytical procedure is subsidiary to the others and supported the interpretation of the data.

Through database manipulation, machine learning, statistics, and data visualization, among other methods from data science and information technology, Data Mining allows you to find patterns in structured data and extract information, transforming information into knowledge [56,57]. Other methods such as Natural Language Processing, Text Mining, and Textual Analysis enable

information and knowledge extraction from unstructured data, as is the case with the qualitative component of OTRs. On the one hand, Text Mining comprises a set of techniques to characterize and transform the text using the words themselves as units of analysis (for example, frequencies, distribution, or presence/absence of specific terms). On the other hand, Natural Language Processing algorithms use syntactic and/or semantic processing based on statistical rules or methods to analyze, segment and extract information [58–60].

The data analysis included applying a set of sequential procedures to arrive at the extraction of information. This process, represented in Figure 1, involved applying several R packages [61].

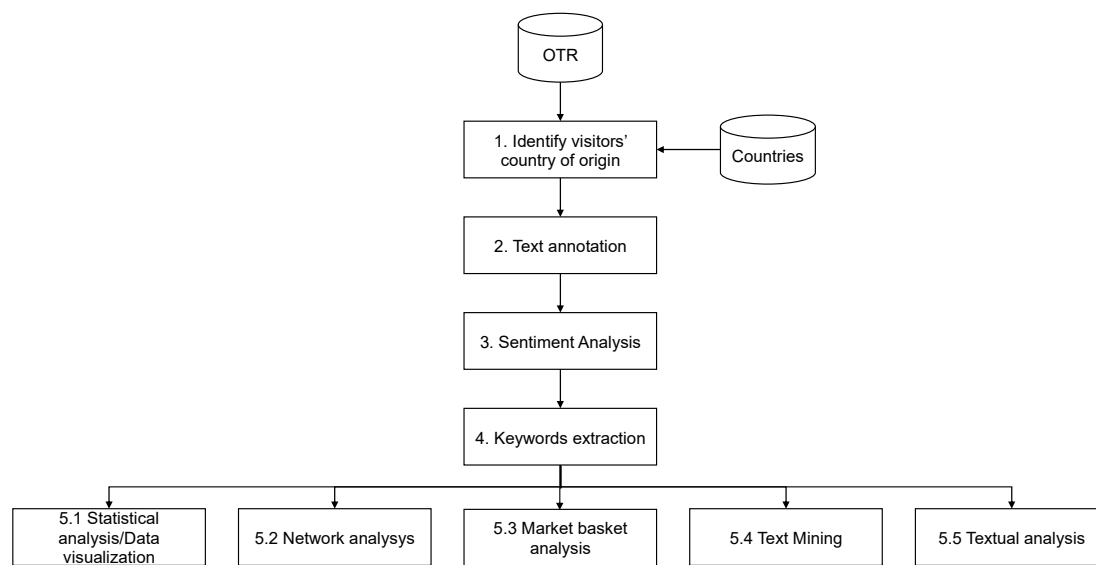


Figure 1. Diagram of data processing and analysis.

The first procedure was creating a category for each review with the respective user’s country of residence (procedure 1 in Figure 1). Although there is a field for Location, this is an open text box, which is not always filled in by users. Some just write the name of the city. Others complement the city name with the state (common in users from the USA and Canada). An added complexity to this procedure is that the language used (in our case, English, Portuguese, or Spanish) will change spelling (e.g., Spain, España, or Espanha). Therefore, to identify the country of residence, this first procedure involved applying Natural Language Processing techniques to divide the Location field text into separate words (tokenization). Then, using another set of data with the list ISO-3166 [62] with country designations in English, Portuguese, and Spanish, the user’s country was identified. This identification was accomplished based on the country’s designation and, when this was not possible, by comparing it against a list of cities and states created manually. This procedure allowed the country to be identified in 5675 reviews (approximately 66% of the total).

The following procedure (procedure 2 in Figure 1) involved a set of text annotation tasks such as counting word frequency (known as TF—Term Frequency) and calculating co-occurrences to explore the qualitative (textual) component of the reviews. With the application of the R package “udpipe” [63], it was possible to perform four common tasks in Natural Language Processing [59]: (1) “tokenization” (dividing the text into units—words and punctuation); (2) “part of speech tagging”—identifying the grammatical form of words based on the definition of the word and the context; (3) “Lemmatization”—reduction of the word to its lemma (i.e., the canonical form or the most common form of the word); (4) Dependency analysis—analysis of the grammatical structure of the sentences, identifying the “main” words and the relationships between “main” words and the other words.

The third procedure (number 3 in Figure 1) involved applying sentiment analysis to the texts. The analysis of feelings (or opinion analysis or opinion mining as it is also sometimes called) is the computational study of opinions about entities, individuals, events, topics, and attributes. Opinion mining allows quantifying opinions according to their polarity (positive, negative, or neutral) [64]. This access to users' opinions on hotels, restaurants, and tourist attractions is considered essential for developing customer service strategies and hence its increasing use in tourism management [65]. Sentiment analysis was performed with the R extension package, "Sentiment Analysis" [66]. For the English reviews, the package's base dictionary was used; for Spanish, the ElhPolar dictionary [67] and for Portuguese, the SentiLex-PT 02 dictionary [68]. The analysis was performed by sentences (taking advantage of the text annotation from the previous procedure), applying the "ruleSentimentPolarity" method, which assigns a value between -1 and 1 to each review, being -1 , very negative and 1 , very positive. The score for each review was calculated based on the average of the values of the respective sentences.

The fourth procedure (number 4 in Figure 1) involved applying another Natural Language Processing technique, which consists of keyword extraction—a sequence of one or more words that provide a compact representation of a document content [69]. The implementation of the RAKE algorithm [63] available in the R package "udpipe" [63] was used to extract the keywords. In simple terms, it can be said that this algorithm selects keywords by a three-step procedure: (1) selection of words not delimited by "stopwords" (non-semantic words which do not add meaning); (2) construction of a co-occurrence matrix of words; (3) calculation of a score according to co-occurrence.

After these four procedures, the analysis proceeded, using statistics, data visualization, network analysis, market basket analysis, text mining, and textual analysis. The results are presented in the following section.

5. Results

5.1. Data Mining

The 8638 reviews were written by 4695 users, which gives an average of 1.7 reviews per user, i.e., some single users posted more than one review. Even though the median was one review per user, the 3rd quartile consists of two reviews per user, with eleven users posting ten or more reviews. There was even one user who published fourteen reviews during the timespan and for the selected places under analysis. This user (a man from Spain), registered as someone between 50 and 64 years old and who had visited 100 cities so far, always wrote in Spanish, and had, at the time of data extraction, a total of 471 reviews published on TripAdvisor. Of the fourteen reviews he published on Coimbra and Salamanca, three reviews were published on Salamanca sites on 2/03/2018, 8/03/2018, and 27/07/2018. Finally, he published nine reviews about places in Coimbra on two consecutive days: 20 and 21/08/2018.

Data analysis by user and city showed that 3203 users published reviews from Salamanca, while 1863 users published reviews from Coimbra, revealing that 101 users posted reviews from the two cities.

Additionally, the network analysis of users who posted reviews about more than one attraction, as displayed in the arc plot of Figure 2, clearly depicts visitors' preferences about which attractions to visit. In Figure 2, the size of the nodes represents the total number of reviews for the attraction. The arc width represents the number of visitors who posted reviews in both connecting attractions.

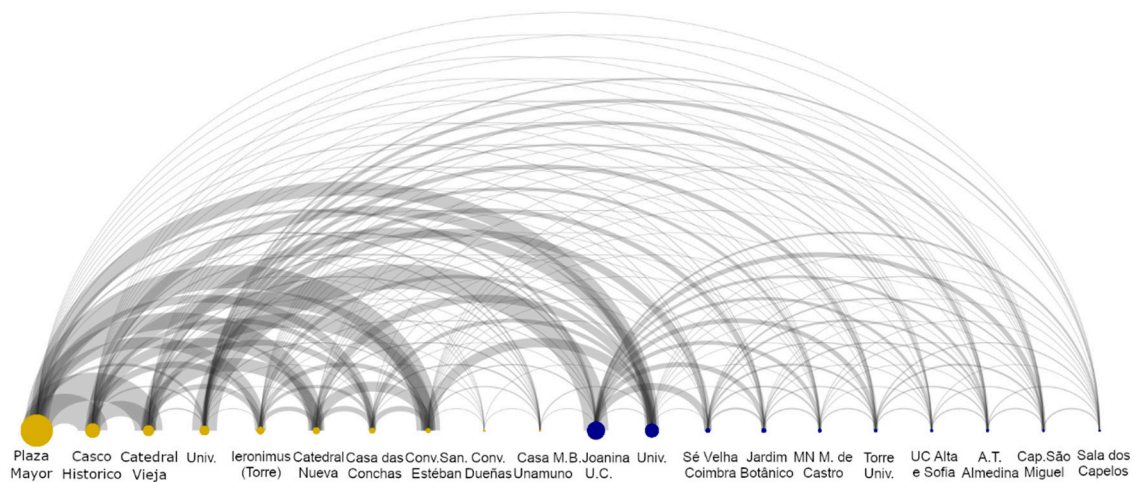


Figure 2. Visitors who posted reviews in multiple attractions.

We also employed the Apriori algorithm to discover association rules between visitors' reviews. The Apriori algorithm is typically employed in the retail industry to apply a "Market Basket Analysis", i.e., to uncover associations between items. The algorithm analyzes the items purchased together (itemset) in each transaction [70]. Also, the algorithm identifies rules, that is, associations of items that are bought together. For example, the rule "Product A + Product B => Product C" can be read as "Product A and product B are commonly purchased with product C". "Product A + Product B" is named the "Left-hand side" (LHS) of the rule, and "Product C" is called "Right-hand side" (RHS).

In this context of attractions visits, we considered each visitant's review as one transaction and applied the algorithm to understand the most common associations of reviews. Although the uncovered associations only reflect the patterns of visitors who posted multiple reviews on TripAdvisor and cannot be generalized to all visitors, it shows patterns of visitors' paths in the cities.

The rules uncovered by the algorithm are measured with three metrics:

Support: the proportion of all transactions with both LHS and RHS (e.g., the percentage of visitors who posted reviews about Catedral Vieja and Plaza Mayor).

$$\text{The formula is } \frac{\text{Number of transactions with both LHS and RHS}}{\text{Total number of transactions}} = P(\text{LHS} \cap \text{RHS}) \quad (1)$$

Lift: the factor by which the co-occurrence of LHS and RHS exceeds the expected probability of both co-occurring. Therefore, the higher the lift, the higher the probability of LHS and RHS occurring together. A lift value less than 1 means that LHS and RHS are probably mutually exclusive or even substitutes.

$$\text{The formula is : } \frac{\text{Support}}{\frac{\text{Number of transactions with LHS}}{\text{Total number of transactions}} \times \frac{\text{Number of transactions with RHS}}{\text{Total number of transactions}}} \quad (2)$$

Network analysis and association rules allow a quantitative analysis of the attractions where visitors stopover. Table 2 shows the top five association rules for each city, in which some prevalent association rules can be observed. For example, for Salamanca, the two top rules account for more than 10% of the visitors, showing that 10% who visited Plaza Mayor also visited the Catedral Vieja.

Table 2. Top five rules by support.

City	Lhs	Rhs	Support	Confidence	Lift
Coimbra	Cap. São Miguel	B. Joanina	0.039	0.890	1.527
	Cap. São Miguel, Univ. Coimbra	B. Joanina	0.027	0.943	1.617
	A. T. Almedina, Sé V. Coimbra	Univ. Coimbra	0.016	0.879	1.915
	Sé V. Coimbra, Torre UC	B. Joanina	0.014	0.813	1.394
	Cap. São Miguel, Sé V. Coimbra	B. Joanina	0.013	1.000	1.715
Salamanca	Casco historico, Catedral Vieja	Plaza Mayor	0.059	0.809	1.234
	Catedral Nueva, Catedral Vieja	Plaza Mayor	0.052	0.827	1.262
	Casas das Conchas, Catedral Vieja	Plaza Mayor	0.040	0.853	1.302
	Casco historico, Catedral Nueva	Plaza Mayor	0.038	0.847	1.293
	Casa das Conchas, Universidade	Plaza Mayor	0.037	0.815	1.244

As expected, when taking into account the users’ country of residence (see above), the vast majority of reviews are made by users from Portugal and Spain (Figure 3 shows all countries with more than 20 reviews). However, Table 3 also demonstrates that there is a different distribution in the two cities. While in Coimbra the number of reviews from foreigners is higher than the number of reviews from Portuguese visitors, and there are even more reviews from users residing in Spain than users living in Portugal, for Salamanca, the reviews from users living in Spain are clearly higher than the reviews of the users living in other countries.

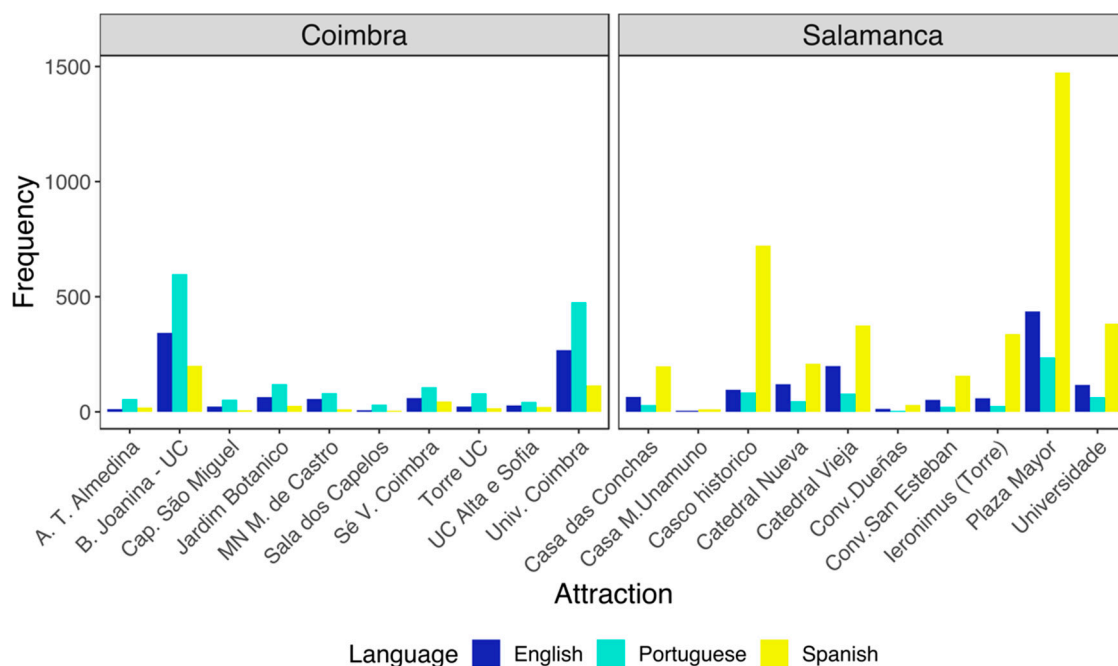


Figure 3. Frequency of reviews by location and language.

Table 3. Top frequency of reviews per country.

Country	Coimbra	Salamanca
Spain	572	2540
Portugal	488	195
United Kingdom	177	415
United States of America	158	170
Brazil	171	64
Argentina	17	90
Australia	42	58
Canada	77	21
Mexico	15	45
Belgium	27	24

The analysis of review frequencies by location and language (Table 1 and Figure 3) shows that the number of reviews in Spanish (4362) is almost twice the number of reviews in English and Portuguese, respectively 2046 and 2230. In turn, as Table 4 illustrates, there is also a high amplitude in the frequency of reviews by location. Salamanca's case stands out, where there is one place with only sixteen reviews, but the most visited has 2146. In Coimbra, only two places, *Biblioteca Joanina* and *Universidade* have more reviews than the city average (299.2), respectively, 1141 and 859. In Salamanca, only three places have more reviews than the city average (564.6). These are Plaza Mayor (2146), *Casco Histórico* (902), and *Catedral Vieja* (653).

Table 4. Review frequency statistics by location.

	Minimum	1. Quartile	Median	Average	3. Quartile	Maximum
Coimbra	43.0	87.5	133.5	299.2	211.5	1.1410
Salamanca	16.0	245.5	398.5	564.6	630.5	2.1460

5.2. Natural Language Processing and Text Mining

The application of Natural Language Processing and Text Mining methods to the reviews' qualitative component also revealed interesting results. For example, the frequency count showed that in the 2992 reviews written about Coimbra attractions, the word "Salamanca" was mentioned four times. In turn, in the 5646 reviews written about Salamanca's attractions, the word "Coimbra" was mentioned seven times. Below is an example of these references (in the original language) with gender, city of residence, age (when available), and date of review:

Extract 1—" (. . .) The University of Salamanca is a treasure of the world, not just Spain. For us, it proudly joins the family of Oxford, Cambridge, Sorbonne, Heidelberg, and Coimbra Universities, which we had visited in the past and greatly admired." (from Sudbury, Massachusetts, 02-07-2017) [our emphasis].

As described in OTR studies, Sentiment Analysis shows differences in the OTR's qualitative components depending on the language used (as in the quantitative component) [45]. Users who write reviews in Portuguese are those who, on average, produce a better quantitative assessment. However, when analyzing the sentiment, we notice that, for two tenths, this is not true for Salamanca. Here, users who write in Spanish present the higher average. Conversely, the English written reviews present the lowest average ratings, except for the quantitative assessment of Coimbra (Figure 4). However, these averages should be analyzed with some caution, as shown by [71], because Sentiment Analysis identifies discrepancies between quantitative and qualitative evaluations. See, for example, the review below in which the two evaluations show conflicting results:

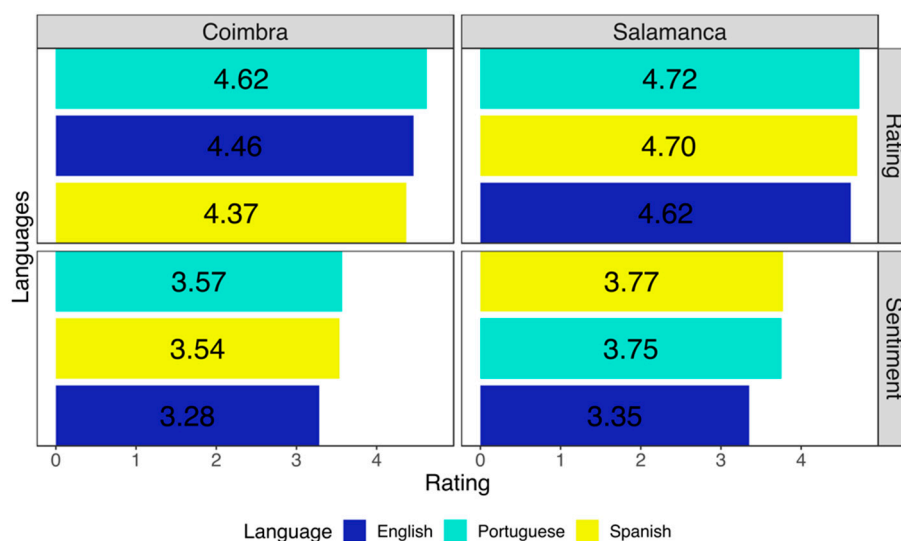


Figure 4. Quantitative assessment and sentiment. Note: The sentiment was normalized to use the same scale as the quantitative assessment (1 to 5).

Extract 2—“Tenia muchas ganas de fotografiar por la noche el puente romano y la catedral, cuando acudí comenzaba la iluminacion nocturna, una mezcla de luz increíble” (which translated to English means “I really wanted to photograph the Roman bridge and the cathedral at night, when I went there the night lighting began, an incredible mixture of light”) (resident in Provincia de Ciudad Real, Spain; Catedral Nueva, Salamanca, 15.05.2017. Rating: 1, Sentiment: 5).

The search for words and terms in the three languages related to the UNESCO heritage shows that the word UNESCO is mentioned more often in Coimbra than in Salamanca, but in both cases, unexpectedly in a very low proportion of the reviews (Table 5). Interestingly, words linked to heritage (e.g., “history”, “architecture”) are present in several reviews.

Table 5. Frequency and distribution of words and terms related to heritage (by language).

Word/Term	City	English	Portuguese	Spanish
UNESCO	Coimbra	17 (1.9%)	22 (1.3%)	1 (0.2%)
	Salamanca	2 (0.2%)	3 (0.5%)	5 (0.1%)
World heritage	Coimbra	10 (1.1%)	10 (0.6%)	3 (0.6%)
	Salamanca	1 (0.1%)	2 (0.3%)	29 (0.7%)
Ancestry	Coimbra	0 (0.0%)	2 (0.1%)	0 (0.0%)
	Salamanca	0 (0.0%)	0 (0.0%)	0 (0.0%)
Culture	Coimbra	10 (1.1%)	77 (4.7%)	5 (1.1%)
	Salamanca	8 (0.7%)	15 (2.6%)	53 (1.4%)
Architecture	Coimbra	53 (6.0%)	100 (6.1%)	7 (1.5%)
	Salamanca	143 (12.3%)	78 (13.3%)	160 (4.1%)
Tradition	Coimbra	8 (0.9%)	39 (2.4%)	2 (0.4%)
	Salamanca	5 (0.4%)	3 (0.5%)	10 (0.3%)
Legacy	Coimbra	1 (0.1%)	1 (0.1%)	0 (0.0%)
	Salamanca	0 (0.0%)	0 (0.0%)	2 (0.1%)
History	Coimbra	160 (18.1%)	367 (22.3%)	37 (8.0%)
	Salamanca	101 (8.7%)	78 (13.3%)	404 (10.4%)

Text annotation also made it possible to check that the adjectives (lemmas of the adjectives, i.e., words as used in dictionaries) in the reviews do not differ from language to language, nor by city (see Figure 5). In fact, for Coimbra, the two most frequent adjectives in the three languages are semantically equivalent.

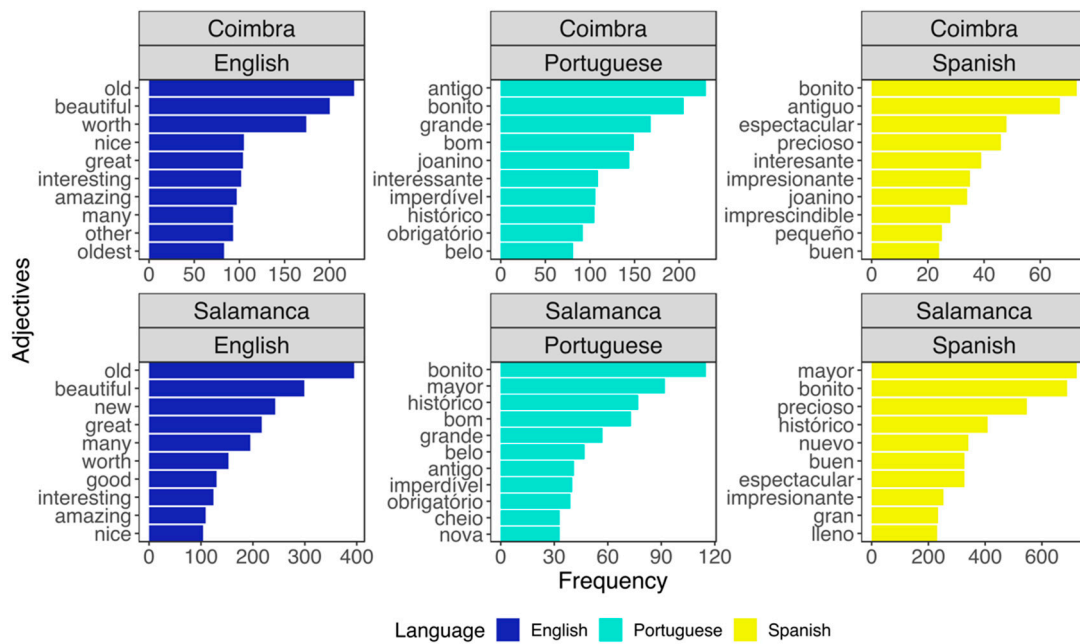


Figure 5. Top adjectives used in reviews.

The extraction of keywords through the RAKE method also provided some interesting findings. Figure 6 shows how the keywords differ by language and city and how many keywords are related to history and heritage, ranging from various references to architectural styles, names of kings, and places to references to specific centuries.

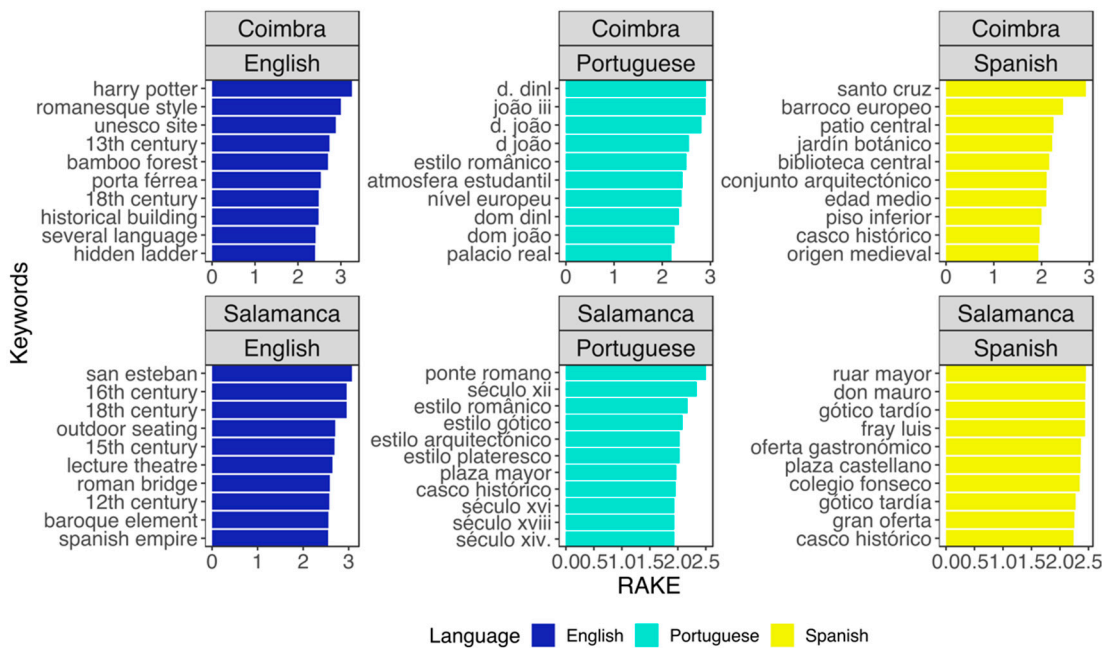


Figure 6. Top 10 keywords identified by the RAKE method.

However, the main surprise was the keyword with the highest RAKE index, for Coimbra, in English: “Harry Potter”. In fact, the subsequent textual analysis of the reviews show that the visiting users observe the *Biblioteca Joanina*, the streets of Coimbra, or the University space from the perspective of the “Harry Potter” universe (which is a fictional main character in a series of seven children’s and

youth novels written in English by the British J.K Rowling (1998-2007). The novels were also adapted for cinema):

Extract 3—“Enjoyed our walk through the narrow streets of historical Coimbra to visit the university. (. . .) An inspiration for J.K. Rowlings and the Harry Potter novels.” (resident in Jerusalem, Israel; Jardim Botânico, Coimbra, 01.08.2017).

6. Discussion of Results

The analysis carried out using both quantitative and qualitative methods shows that some locations collected more reviews than others. In Coimbra, the number of reviews from foreigners compared to Portuguese nationals is higher, while in Salamanca, the number of reviews of users residing in Spain and in Spanish are higher than other users. These results indicate that Coimbra receives proportionally more foreign visitors than Salamanca. Thus, two explanatory hypotheses can be put forward: either Coimbra is promoting itself adequately in the foreign market or is not promoting itself well within the national market. The hypotheses could also be true inversely for Salamanca: the city promotes itself adequately among the national public or promotes itself poorly among international visitors. Finally, it appears that most visitors come from European or Latin American countries.

The study of associations can enable DMO to understand visiting patterns better and implement measures to improve the tourists' experiences, such as (1) introducing the visitors of one city to the other city; since we found common visitors to both cities, some of which who actually compare them, one could hypothesize that both cities cater for similar tourist “targets”, therefore cross-marketing could be employed to present the other city to visitors of one city; (2) when visiting the more popular attractions, visitors could be encouraged to check the less popular ones. This DMO could mitigate the effects of overtourism in the most popular attractions and help visitors get a more wholesome experience.

One of our research questions asked if becoming listed as UNESCO heritage was an attraction factor for visitors. We were able to ascertain, using the RAKE method, that the word “UNESCO” or the equivalent phrase, “world heritage”, seldomly occurred; however, the words “architecture” and “history” appeared more frequently. These results seem to indicate that among the four WHS reasons for applying to the UNESCO list shown by Research Consulting Ltd. & Trends Business Research Ltd. [26]: celebration, SOS designation, brand for marketing/logo, and catalyst for building space. Clearly, the brand category is not being worked on either in Coimbra or Salamanca, and, based on our corpus, we could say that it does not seem to be an attraction factor for visits. This result helps to clarify the answer to the question formulated by Remoaldo, Vareiro, Ribeiro, and Marques [72] “Do the cities declared by UNESCO as World Heritage Sites have an outstanding tourist competitive advantage over the ones not benefiting from such a label?”. In conclusion, the UNESCO reputation or brand does not seem to be appropriately explored.

As mentioned above, the analysis revealed the transposition of the Harry Potter fiction universe to the ancient space of Coimbra or, in other words, the evocation of the Harry Potter universe from the immersion in the Coimbra environment, at least for visitors who wrote reviews in English. This phenomenon, common in some works of fiction could be used by those responsible for promoting the city of Coimbra to sponsor it among tourists.

7. Conclusions

It is not possible to understand modern heritage tourism without considering the people who ‘consume’ heritage [73]. Following that logic, this study explored the way people ‘consume’ the heritage of two UNESCO listed cities using a corpus collected from what people themselves write and post about that same ‘consumption’. From a methodological perspective, this study contributes to the tourism literature on heritage studies by demonstrating how a data science approach that uses natural language processing, machine learning, and other methods can be used to explore OTRs in order to gain a clearer picture and understanding of what kind of things visitors value and want to find in their experience. Even though our study is not unique in some of its methodological choices

(see, for instance, how Rodrigues et al. rely on UGC produced by tourists to identify thermal and spa's attractiveness [16]), we believe it may be used as a methodological framework to understand the applicability of various methods to analyze OTRs of UNESCO World Heritage sites. The study showed how the various techniques could complement each other and the ability to establish standardized information on the expectations, perceptions, and appraisals of visitors, who are simultaneously users of TripAdvisor. Also, we sought to highlight the advantages of applying various methods of analysis to a comparative study and demonstrate the informational wealth of OTRs as an instrument to (re) position the destination.

Theoretically, we believe this study uncovers some weaknesses in terms of the World Heritage List recognition by tourists or visitors to Salamanca and Coimbra, even though econometric studies show that becoming included among the sites on the World Heritage List could have the stimulating effect of promoting tourism (namely in China) [74]. Nonetheless, other studies seem to question the World Heritage List's direct impact on tourism attraction [75]. Indeed, our findings showing unexpected few occurrences of the UNESCO acronym (or synonyms) in the OTRs demand further studies to explore this link better. This finding could mean that the brand is not living up to its promise [76] and, therefore, destination marketing organizations should focus on enhancing the UNESCO 'brand'. Empirically, it is possible to state that analyzing these particular OTRs about Coimbra and Salamanca revealed the semantic similarity of adjectives across the three different languages and the two cities. However, when applying the RAKE method, the keywords were no longer that similar, and we found that many are related to history and heritage, ranging from various references to architectural styles, names of kings, and places. Additionally, the keyword with the highest RAKE index, for Coimbra, in English was "Harry Potter", which was rather unexpected. This presents interesting challenges in terms of investigating further the language and discourse of OTRs, as well as provides useful insights to be used by DMOs in Coimbra. This study is not without limitations. First, it only focuses on two cities. It could be useful to replicate it with other listed cities from different geographic regions. This replication would expand our understanding of what tourists/users want and value, particularly relevant to promote cities, leverage economic development, and contribute to strength place identity. Second, UNESCO World Heritage tourists are male and female in equal numbers, are highly schooled, are employed, and travel in small groups (two to five people) [8]. Although not tested in our analysis, these demographics could help explain the precise and relatively sophisticated textual construction of the collected OTRs and the frequent analogies and cultural references to other places, issues that will be explored in the future. We believe we have proven our study's relevance and its potential for future research, namely for exploring differences in OTR discursive and rhetorical patterns. In particular, the relevance of multilingual studies, which have been underexplored, despite the growing importance of intercultural issues in computer-mediated communication [77].

Author Contributions: All authors contributed equally to this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Funds provided by FCT—Foundation for Science and Technology through project UIDB/04470/2020.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Tscheu, F.; Buhalis, D. Augmented reality at cultural heritage sites. In *Information and Communication Technologies in Tourism*; Inversini, A., Schegg, R., Eds.; Springer International: Cham, Switzerland, 2016; pp. 607–619.
2. Ndou, V. New approaches for managing tourism complexity: Implications and insights. In *Digital Culture and e-Tourism: Technologies, Applications and Management Approaches*; Hershey: New York, NY, USA, 2011; pp. 123–138.
3. De Ascaniis, S.; Cantoni, L. Online visit opinions about attractions of the religious heritage: An argumentative approach. *Church Commun. Cult.* **2017**, *2*, 179–202. [[CrossRef](#)]

4. Europeana. Moving Targets: Engaging Cultural Tourists with Collections and Listings Content Online. 2015. Available online: http://pro.europeana.eu/files/Europeana_Professional/Publications/Moving%20Targets%20%20Colour.pdf (accessed on 6 May 2019).
5. du Cros, H.; McKercher, B. *Cultural Tourism: The Partnership between Tourism and Cultural Heritage Management*, 2nd ed.; Routledge: New York, NY, USA, 2015.
6. Xiang, Z.; Fesenmaier, D. Big Data analytics. Tourism design and smart tourism. In *Analytics in Smart Tourism Design: Concepts and Methods*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer: Cham, Switzerland, 2017; pp. 299–307.
7. Hausmann, A.; Weuster, L. Possible marketing tools for heritage tourism: The potential of implementing information and communication technology. *J. Herit. Tour.* **2018**, *13*, 273–284. [[CrossRef](#)]
8. Adie, B.A.; Hall, C.M. Who visits world heritage? A comparative analysis of three cultural sites. *J. Herit. Tour.* **2016**, *12*, 67–80. [[CrossRef](#)]
9. Fortuna, C.; Gomes, C.S. Turismo, cidade e universidade: O caso de Coimbra. In *Turismo e Cultura: Destinos e Competitividade*; Cravidão, F., Santos, N., Eds.; Imprensa da Universidade de Coimbra: Coimbra, Portugal, 2014.
10. Instituto Nacional de Estatística. Estatísticas do Turismo. 2019. Available online: <https://www.ine.pt/xurl/pub/133574> (accessed on 23 October 2020).
11. Estadística de Movimientos Turísticos en Fronteras (FRONTUR). 2019. Available online: <https://www.ine.es/daco/daco42/frontur/frontur1219.pdf> (accessed on 23 October 2020).
12. United Nations Educational, Scientific and Cultural Organization (UNESCO). UNESCO and Online Travel Giant Tripadvisor Launch World Heritage Partnership. 2009. Available online: <https://whc.unesco.org/en/news/561> (accessed on 3 May 2019).
13. Walker, C.; Baxter, J. Method sequence and dominance in mixed methods research: A case study of the social acceptance of wind energy literature. *Int. J. Qual. Meth.* **2019**, *18*, 1–14. [[CrossRef](#)]
14. Molina, J.; Font, X. Mixed methods in sustainable tourism research: An analysis of prevalence, designs and application in JOST (2005–2014). *J. Sustain. Tour.* **2016**, *24*, 549–573. [[CrossRef](#)]
15. Mason, J. Mixing methods in a qualitative driven way. *Qual. Res.* **2006**, *6*, 9–25. [[CrossRef](#)]
16. Rodrigues, H.; Brochado, A.; Troilo, M. Listening to the murmur of water: Essential satisfaction and dissatisfaction attributes of thermal and mineral SPAs. *J. Travel Tour. Mark.* **2019**. [[CrossRef](#)]
17. Gomes, C.S. Novas imagens para velhas cidades? Coimbra, Salamanca e o turismo nas cidades históricas. *Sociologia* **2012**, *23*, 37–49.
18. Larsson, T.; Reibring, A. Tracking multi-destination travel in Europe from long-haul source markets—US, China, Japan and India. In *European Travel Commission Handbook, Market Intelligence Report*; Europe Translation Agency (ETC): Brussels, Belgium, 2018; Available online: http://www.etc-corporate.org/?page=report&report_id=122 (accessed on 11 May 2019).
19. Sánchez-Jiménez, M.A.; Correia, M.B.; Matos, N. Análisis de las redes sociales como instrumento de comunicación en los destinos turísticos del Algarve. *Revista GEINTEC-Gestão Inovação Tecnologias* **2018**, *8*, 4476–4494. [[CrossRef](#)]
20. Ribeiro, F.P.; Antonio, N.; Correia, M.B. Uma abordagem metodológica para a análise comparativa de comentários de viagens online de duas cidades património da UNESCO. In *Turismo & História: Perspetivas Sobre o Património da Humanidade no Espaço Ibero-Americano*; Henriques, C., Herédia, M., Eds.; Editora da Universidade de Caxias do Sul: Caxias do Sul, Brazil, 2020; pp. 304–331.
21. VanBlarcom, B.; Kayahan, C. Assessing the economic impact of a UNESCO World Heritage designation. *J. Herit. Tour.* **2011**, *6*, 143–164. [[CrossRef](#)]
22. Caust, J.; Vecco, M. Is UNESCO World Heritage recognition a blessing or a burden? Evidence from developing Asian countries. *J. Cult. Herit.* **2017**, *27*, 1–9. [[CrossRef](#)]
23. United Nations Educational, Scientific and Cultural Organization (UNESCO). Report on the World Heritage Thematic Programmes. 2016. Available online: <http://whc.unesco.org/archive/2016/whc16-40com-5D-en.pdf> (accessed on 8 May 2019).
24. Kayahan, B.; VanBlarcom, B. Cost benefit analysis of UNESCO world heritage site designation in Nova Scotia. *Rev. Econ. Anal.* **2012**, *4*, 247–273.

25. Prud'homme, R. The Socio-Economic Impacts of Inclusion of a Site on the World Heritage List: Three Studies. 2008. Available online: <http://www.prudhomme.com/resources/2008+Impact+Liste+Patrimoine.pdf> (accessed on 2 May 2019).
26. Research Consulting Ltd. & Trends Business Research Ltd. World Heritage Status: Is There Opportunity for Economic Gain? 2009. Available online: <http://www.lakesWH.co.uk/documents/WHTheEconomicGainFinalReport.pdf> (accessed on 9 May 2019).
27. Soldado, C. Há um "Efeito Unesco" em Coimbra? Público. 2015. Available online: <https://www.publico.pt/2015/06/22/p3/noticia/ha-um-efeito-unesco-em-coimbra-1823585> (accessed on 2 December 2019).
28. Cuccia, T.; Guccio, C.; Rizzo, I. The effects of UNESCO World Heritage List inscription on tourism destinations performance in Italian regions. *Econ. Model.* **2016**, *53*, 494–508. [CrossRef]
29. Poria, Y.; Butler, R.; Airey, D. Links between tourists, heritage, and reasons for visiting heritage sites. *J. Travel Res.* **2004**, *43*, 19–28. [CrossRef]
30. Poria, Y.; Reichel, A.; Biran, A. Heritage site perceptions and motivations to visit. *J. Travel Res.* **2006**, *44*, 318–326. [CrossRef]
31. United Nations Educational, Scientific and Cultural Organization (UNESCO). World Heritage Information Kit: Benefits of Ratification. 2000. Available online: <http://www.unesco.org/whc/nwhc/pages/doc/main.htm> (accessed on 5 April 2019).
32. Pordata. População Residente, Estimativas a 31/12. 2018. Available online: <https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente++estimativas+a+31+de+Dezembro-120> (accessed on 6 May 2019).
33. Pordata. Hóspedes Nos Alojamentos Turísticos: Total e por Tipo de Alojamento. 2018. Available online: <https://www.pordata.pt/Municipios/H%C3%B3spedes+nos+alojamentos+tur%C3%ADsticos+total+e+por+tipo+de+alojamento-750> (accessed on 8 May 2019).
34. Notícias de Coimbra. 500 mil Pessoas Visitaram a Universidade de Coimbra em 2017. 2018. Available online: <https://www.noticiasdecoimbra.pt/500-mil-pessoas-visitaram-universidade-coimbra-2017/> (accessed on 7 May 2019).
35. Instituto Nacional de Estadística (INE). Cifras Oficiales de Población de los Municipios Españoles. 2017. Available online: <https://www.ine.es/jaxiT3/Datos.htm?t=2911> (accessed on 11 May 2019).
36. Berzosa-Pareja, R. Salamanca: Planificación y gestión turística de un destino patrimonial. *Polígonos Revista de Geografía* **2018**, *30*, 171–194. [CrossRef]
37. Instituto Nacional de Estadística (INE). Encuesta de Ocupación Hotelera. Número de Viajeros por Puntos Turísticos-Salamanca (Datos de 2010 a 2017). 2017. Available online: <https://www.ine.es/jaxiT3/Tabla.htm?t=2078&L=0> (accessed on 11 May 2019).
38. O'Reilly, T. What is Web 2.0: Design patterns and business models for the next generation of software. *Int. J. Digit. Econ.* **2007**, *65*, 17–37.
39. Buhalis, D.; Inversini, A. Tourism branding, identity, reputation co-creation, and word-of-mouth in the age of social media. In *Tourism Management, Marketing, and Development: Volume I: The Importance of Networks and ICTs*; Mariani, M.M., Baggio, R., Buhalis, D., Longhi, C., Eds.; Palgrave Macmillan: New York, NY, USA, 2014; pp. 15–40.
40. Tussyadiah, I.P. Toward a theoretical foundation for experience design in tourism. *J. Travel Res.* **2014**, *53*, 543–564. [CrossRef]
41. Dias, J.A.; Ribeiro, F.P.; Correia, A. Online reviews of short-term visits: Exploring sense of place. *Int. J. Cult. Tour. Hosp. Res.* **2013**, *7*, 364–374. [CrossRef]
42. Iglesias-Sánchez, P.P.; Correia, M.B.; Jambrino-Maldonado, C. Challenges of open innovation in the tourism sector. *Tour. Plan. Dev.* **2019**, *16*, 22–42. [CrossRef]
43. Roque, V.; Fernandes, G.; Raposo, R. Identificação dos Media Sociais utilizados pelas organizações de gestão de destinos: O caso de estudo do destino turístico Serra da Estrela. *Revista Turística Desenvolvimento* **2012**, *17/18*, 225–234.
44. Bergel, M.; Brock, C. Visitors' loyalty and price perceptions: The role of customer engagement. *Serv. Ind. J.* **2019**, *39*, 575–589. [CrossRef]
45. Mehmood, S.; Liang, C.; Gu, D. Heritage image and attitudes toward a heritage site: Do they really mediate the relationship between user-generated content and travel intentions toward a heritage site? *Sustainability* **2018**, *10*, 4403. [CrossRef]

46. Luli, V.D.D.; Kawano, N. Online tourist reviews and accidental conveyors of memories of the atomic bomb. *J. Tour. Cult. Chang.* **2020**, *18*, 514–531. [CrossRef]
47. Rahmani, K.; Gnoth, J.; Mather, D. Tourists' participation on Web 2.0: A corpus linguistic analysis of experiences. *J. Travel Res.* **2018**, *57*, 1108–1120. [CrossRef]
48. TripAdvisor. 35 Amazing TripAdvisor Statistics and Facts (2019) By the Numbers. DMR—Business Statistics. 2018. Available online: <https://expandedramblings.com/index.php/tripadvisor-statistics/> (accessed on 17 April 2019).
49. TripAdvisor. About TripAdvisor. 2019. Available online: http://www.tripadvisor.com/pages/about_us.html (accessed on 17 April 2019).
50. Olsen, W. Triangulation in social research: Qualitative and quantitative methods can really be mixed. In *Developments in Sociology*; Holborn, M., Ed.; Causeway Press: Ormskirk, UK, 2004.
51. António, N.; De Almeida, A.; Nunes, L.; Batista, F.; Ribeiro, R. Hotel online reviews: Different languages, different opinions. *Inform. Technol. Tour.* **2018**, *18*, 157–185. [CrossRef]
52. Dunham, K.; Melnick, J. *Malicious Bots: An Inside Look into the Cyber-criminal Underground of the Internet*; CRC Press: Boca Raton, FL, USA, 2009.
53. TripAdvisor. All about Your TripAdvisor Bubble Rating. 2018. Available online: <https://www.tripadvisor.com/TripAdvisorInsights/w810> (accessed on 20 April 2019).
54. European Commission. Study on Online Consumer Reviews in the Hotel Sector. Final Report. 2014. Available online: http://rpaltd.co.uk/uploads/report_files/hotel-reviews.pdf (accessed on 6 May 2019).
55. Duan, W.; Yu, Y.; Cao, Q.; Levy, S. Exploring the impact of social media on hotel service performance: A sentimental analysis approach. *Cornell Hosp. Q.* **2016**, *57*, 282–296. [CrossRef]
56. Adedoyin-Olowe, M.; Gaber, M.M.; Stahl, F. A Survey of Data Mining Techniques for Social Media Analysis. 2013. Available online: <http://arxiv.org/abs/1312.4617> (accessed on 1 May 2019).
57. Liao, S.H.; Chu, P.H.; Hsiao, P.Y. Data mining techniques and applications—A decade review from 2000 to 2011. *Expert Syst. Appl.* **2012**, *39*, 11303–11311. [CrossRef]
58. Feldman, R.; Sanger, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*; Cambridge University Press: New York, NY, USA, 2007.
59. Manning, C.D.; Manning, H.S. *Foundations of Statistical Natural Language Processing*, 2nd ed.; The MIT Press: London, UK, 1999.
60. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural language processing: An introduction. *J. Am. Med. Dir. Assoc. JAMIA* **2011**, *18*, 544–551. [CrossRef]
61. R Core Team. R: A Language and Environment for Statistical Computing. 2016. Available online: <https://www.R-project.org/> (accessed on 10 April 2019).
62. International Organization for Standardization. ISO Country Codes. 2017. Available online: <https://www.iso.org/obp/ui/#search> (accessed on 13 February 2019).
63. Wijffels, J. UDpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the “UDpipe” “NLP” Toolkit (Version R package version 0.8). 2018. Available online: <https://CRAN.R-project.org/package=udpipe> (accessed on 6 April 2019).
64. Liu, B.; Zhang, L. A survey of opinion mining and sentiment analysis. In *Mining Text Data*; Aggarwal, C.C., Zhai, C.X., Eds.; Springer: Boston, MA, USA, 2012; pp. 415–463.
65. Ge, J.; Vazquez, M.; Gretzel, U. Sentiment analysis: A review. In *Advances in Social Media for Travel, Tourism and Hospitality: New Perspectives, Practice and Cases*; Sigala, M., Gretzel, U., Eds.; Routledge: New York, NY, USA, 2018; pp. 243–261.
66. Feurriegel, S.; Proelochs, N. Sentiment Analysis: Dictionary-Based Sentiment Analysis (Version R Package Version 1.3-2). 2018. Available online: <https://CRAN.R-project.org/package=SentimentAnalysis> (accessed on 20 April 2019).
67. Saralegi, X.; San Vicente, I. Elhuyar at TASS 2013. In *Proceedings of the XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural*; Esteban, A.D., Loinaz, I.A., Román, J.V., Eds.; El Congreso Español de Informática: Madrid, Spain, 2013; pp. 143–150.
68. Silva, M.J.; Carvalho, P.; Sarmiento, L. Building a sentiment lexicon for social judgement mining. In *Computational Processing of the Portuguese Language*; Caseli, H., Villavicencio, A., Teixeira, A., Perdigão, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 218–228.

69. Rose, S.; Engel, D.; Cramer, N.; Cowley, W. Automatic keyword extraction from individual documents. In *Text Mining*; Berry, M.W., Kogan, J., Eds.; John Wiley & Sons Ltd.: Chichester, UK, 2010; pp. 1–20.
70. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1994; pp. 487–499.
71. António, N.; de Almeida, A.; Nunes, L.; Batista, F.; Ribeiro, R. Hotel online reviews: Creating a multi-source aggregated index. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3574–3591. [[CrossRef](#)]
72. Remoaldo, P.; Vareiro, L.; Ribeiro, J.C.; Marques, V. Tourists' motivation toward visiting a World Heritage Site: The case of Guimarães. In *Tourism & History. World Heritage—Case Studies of Ibero-American Space*; Henriques, M.C., Ed.; Editora da Universidade de Caxias do Sul: Caxias do Sul, Brazil, 2017; pp. 99–121.
73. Light, D.; Prentice, R.C. Who consumes the heritage product? Implications for European heritage tourism. In *Building a New Heritage: Tourism, Culture and Identity in the New Europe*; Ashworth, G.J., Larkham, P.J., Eds.; Routledge: London, UK, 1994; pp. 90–116.
74. Yang, C.-H.; Lin, H.-L. Is UNESCO recognition effective in fostering tourism? A comment on Yang, Lin and Han: Reply. *Tour. Manag.* **2011**, *32*, 455–456. [[CrossRef](#)]
75. Cellini, R. Is UNESCO recognition effective in fostering tourism? A comment on Yang, Lin and Han. *Tour. Manag.* **2011**, *32*, 458–460.
76. Armstrong, G.; Kotler, P. *Principles of Marketing*, 15th ed.; Pearson Prentice Hall: New York, NY, USA, 2014.
77. Tian, Y. Engagement in online hotel reviews: A comparative study. *Discourse Context Media* **2013**, *2*, 184–191. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).