Are they sisters? Using text mining approach to understand perceptions about Madeira and Bermuda

Sandra Maria Correia Loureiro
Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL);
Lisbon, Portugal
Ricardo Godinho Bilro
Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL);
Tiago Ferro
INSTITUTO UNIVERSITARIO DE LISBOA (ISCTE-IUL) and Business Research
Unit (BRU/IUL)

Paper presented at the 48th Annual EMAC Conference, Hamburg, May 24-27, 2019.

Are they sisters? Using text mining approach to understand perceptions about Madeira and Bermuda

Abstract

The current study explores how tourists perceive two islands with identical culture, but different demographic characteristics. Two complementary methodologies are employed: *Netnography* and *Text Mining*. Eight encodings are used based on the literature: *Rating*, *Categories*, *Membership Level of reviewers*, *Language type*, *Tourism Experience Model*, *Content*, and *Symbology* and *Positive/Negative information*. With the first methodology, all encodings are analyzed, while in the *Text Mining*, only three: *TEM*, *ML* and *categories*. Customer satisfaction is slightly superior in Madeira. *Lifestyle and Leisure* and *Tourism, Travel and Commuting* are the main themes. Reviewers are engaged with the destinations and the re-purchase scenario is highly probable.

Keywords: Tourism Experience Model, Netnography, Text Mining, Island destination

Track: Tourism Marketing

1. Introduction

According to Schuckert *et al*. (2015, p.608) "71% of independent travel related bookings are done online, while 36% of all package tours are booked online". 65% of travellers looking for leisure activities or vacations will pursuit for information online before deciding to take the travel, and "69% of their plans are determined by online travel reviews" (Fang *et al*., 2016, p.498). Platforms as *Yelp* and *TripAdvisor* became multimillion dollars companies, due to this need for consumers to get information before taking any decision. Yelp, for instance, has more than 186 million users posting 150 million reviews online monthly, about the most varied subjects (ReviewTrackers, 2018). The *TripAdvisor* platform has 455 million users, which have written more than 661 million reviews since it's' foundation – it accounts with 7.5 million accommodations, restaurants and attractions. From those, over 1.2 million are hotels, 4.7 million are restaurants, and 975 thousand are attractions. The extension of these firms is worldwide, and the numbers show the importance of this platform to the online community (Smith, 2018).

The literature has studied online reviews and the tourism industry in multiple and different ways, since the effects of the first in the hotel booking intentions (Cheng *et al*., 2017), the effect of the user's membership level ranking on the consumer's decision making (Chen, 2015) or even trying to predict future trends based on past stats (Schuckert *et al*., 2015). A gap in the literature is found, that is, past studies when using online reviews to analyse the content, researchers tend to focus only in one destination, the current study explores how tourists perceive two islands with identical culture, but different demographic characteristics.

2.Lierature Review

The concept of *Customer Experience* has been considered in multiple and different ways (Hirschman & Holbrook, 1982; Gilmore & Pine, 2002; Thompson *et al*., 1989; Loureiro, 2014; Lemon & Verhoef, 2016). In tourism context, Otto and Ritchie (1996) claim that tourism experience may be defined as mental state defined by the participants feelings, while products and services are tangible and intangible – respectively, "experiences represent events that commit people in a particular manner and, as such, are memorable" (Mendes e*t al*., 2010, p. 112). Pine and Gilmore (1999) add that the tourism experience starts previously of the arrival to the endpoint and finishes with memories and intentions for visiting it in the future.

Tourism experience is a complex concept that involves a cetin number of players in the act (e.g.., Uriely, 2005; Bigné & Gnoth, 2005), consisting "of a continuous flux of related and integrated services which are acquired during a limited period of time, often in different geographical areas" (Mendes e*t al.*, 2010, p.112).

According to Deans and Gnoth (2012) and Gnoth and Matteucci (2014) the *Tourism Experience Model* (*TEM*) aggregates insights into experiences and experiencing in tourism, so as to be able to locate tourists in a commonly understood grid of references, meaning that it allows destination managers to understand what their position is, in the tourist's experiences. Positive memories are fundamental being linked to the destinations otherwise tourists may easily switch destinies for any other amusing experiences. The model has two axes, Consciousness (style of how tourists receive their experience of the destination) and Activity (recreational activities are all those that tourists have engaged in previously and as a matter of habit, training and repetition; exploratory activities, here tourists seek new insights, understandings, social and bodily feelings). The current study applies this model to the online tourism context.

3. Methodology

The first study uses Netnography approach to answer to "How tourists perceive two island destinations with identical culture but different demographic characteristics, through *Social Networking Sites (SNS)*?", Netnography comprises *Internet* and *Ethnography* and its process results of a combination between the methodologies used in the study of cultural anthropology, applied to the online communities (Eastin *et al.*, 2011). According to Kozinets (2010), to use this approach researchers should follow six steps of ethnography: research planning, entrée, data collection, interpretation, ensuring ethical standards, and research representation. *TripAdvisor's* webpage is the SNS chosen because allow to get the information needed to achieve the aim of the study and has more than 600 million reviews and opinions, and more than 7.5 million different accommodations, airlines, attractions and restaurants. Moreover, it is said to have more than 455 million monthly average unique visitors (TripAdvisor, 2017).

Two island destinations are chosen: Madeira and Bermuda. For the first, Madeira, has a *GDP* of 5,244 million Euros (data for 2017), and its significance to the world is gradually being recognized - it was awarded as *Best Island Destination* by the *World Travel Awards* in 2017 and *Mankind Heritage* by *UNESCO*, in the same year. The second one is chosen based on the first, meaning that it had to be an island

destination with cultural similarities so that the purpose of this study made sense. After an extensive research, *Bermuda* is selected to be the second destination to be studied.

The second study employs text minding approach, a process that allow to deal with big flows of information (Big Data) (Kirilenko *et al*, 2017). One of the technique embedded in text mining is called *Sentiment Analysis (SA)*, which consists in a "computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes" (Liu & Zhang, 2012, p.:1), and it aims at studying the polarity of a certain text script (e.g. review, blog text or any user-generated content), being able to characterize if such excerpt is positive, neutral or negative.SA is quite useful in the tourism industry, because polarity may influence further travellers (Alaei *et al.*, 2017). SA will also, through the selected text, be able to analyse if the excerpt is based on subjective or objective premises. For instance, subjective reviews have as ground opinions and stereotype bias – reviewers let themselves be influenced by their feelings, beliefs and opinions given by others, while objective ones are accurate and based on concrete proofs and determinate interpretations (Feldman, 2013). The second study used the same SNS and islands, as in the first one.

4. Major Findings and Discussion

Regarding the *Madeira's* results of the first study, there are multiple data that shows how positive the reviewers have evaluated their stays in the island - average rating of 4.59, plus every category being classified over 4 and especially the fact that in 98.78% of the reviews it is possible to find positive information. Such aspects are believed to persuade readers to purchase the destination by increasing their booking intention (Chen & Tseng, 2011; Cheng *et al.*, 2017). Moreover, the fact that negative information is present in 33.45% of the reviews can be a positive factor in enhancing the reviewers credibility (Schuckert *et al.*, 2015; Zhong & Leung, 2013), as well as firms being able to identify possible services or products imperfections (Cheng & Tseng, 2011). The type of language mostly used by the reviewers is literal (vs figurative) – 79.27%, which is perceived by authors as being more useful for the readers (Fang *et al.*, 2016), thus increasing the purchase buying intention (Ashraf *et al.*, 2016). Also, the reviewers are mainly *Low Ranked* - 61.15%, which according to the literature indicts they are more willing to accept the information (Chen, 2015), but having *LRR* using figurative language – 22.51% of the reviews, may be a threat to the tourism of *Madeira*, once potential consumers tend to lower their booking intention against

reviews with such characteristics (Wu *et al.*, 2017). The results of the *Tourism Experience Model* indicate that it is a place where people tend to perform recreational activities in order to get closer to their existential being (Gnoth & Deans, 2012; Gnoth & Matteucci, 2014).

Concerning the *Bermuda's* results for the first stydy, there are also numerous indicators that led us to claim the reviewers' experiences in the island was positive: an average rating of 4.43, all of the categories were evaluated with a rating above 4 and the fact of existing positive information in 96.70% of the reviews that were extracted. According to the literature studied these results point out that the *e-WOM* presented online in *TripAdvisor* is capable of persuading the readers in increasing the customer's booking intention (Chen & Tseng, 2011; Cheng *et al.*, 2017). Moreover, it was found negative information about *Bermuda* in 32.28%, which can be turned into a positive factor by increasing the reviewers' credibility (Schuckert *et al.*, 2015; Zhong & Leung, 2013), and allowing brands/firms to re-adjust their strategies based on the flaws presented (Cheng & Tseng, 2011). The percentage of reviewers using literal (vs figurative language) was massive – 80.31%, growing the usefulness of the review in the eyes of the readers (Fang *et al.*, 2016), consequently intensifying the customer's buying intention (Ashraf *et al.*, 2016). The vast majority of reviewers were classified in the *TripAdvisor* platform as *Low Ranked* – 56.69%, which according to the literature states that the reviewers are more agreeable in accepting other's information (Chen, 2015). A problem might arise due to having 21.11% of the reviews written by *LRR* with figurative language, as the literature points out that it can reduce the customer's booking intention (Wu *et al.*, 2017). The results of the *Tourism Experience Model* were identical to the ones already mentioned in the *Madeira's* case. According to the literature studied and the results obtained, *Bermuda* is perceived as a destiny where people, especially through recreational activities tend to get closer to their existential being (Gnoth & Deans, 2012; Gnoth & Matteucci, 2014).

In what concerns to the *Madeira's* results from the second study, it is clear that the sentiments demonstrated by the reviewers are extremely positive: an average polarity of 4.68; even the reviewers who manifested *disagreement* showed positive sentiments; every score above 3 in every coding and component (*Agreement*, *Irony* and *Subjectivity*) studied. Such extreme sentimentalism demonstrated, outcomes in reviews considered as more useful (Fang *et al.*, 2016), influencing positively the customer's decision making (Ashraf *et al.*, 2016). A higher polarity influences, in a higher level,

potential tourists of a destination (Alaei *et al.*, 2017), and having in mind that positive customer satisfaction is capable of transforming those potential customers in real ones (Chang *et al.*, 2009; Ribbink *et al.*, 2004), it can be infered that the reviews extracted are capable of converting readers in customers of a destination. In fact, the literature also shows that in order to increase the booking intention, destinations are dependent of positive emotions being demonstrated in *UCG* (Wang, 1999), as happens in the *Madeira's* case. The findings also prove that despite reviewers express *disagreement*, they continue to recommend the destination and informing potential customers, has pointed out in the literature (Bilro, 2017). Lastly, the topics with the higher positive connotation: *Hotel*, *Services* and *Gastronomy*; while the most negative ones were: *Measures*, *Tourism* and *Weather and Meteorology*. The results also show that the *Restaurant* category contains the topics being most positively evaluated – sentimentaly, while the *Cultural Activities* the worst in terms of positive sentiments. This is also linked with the experiences lived by the reviewers, as *Cultural Activities* are more involved with *Knowledge Seeker* it makes sense that it arises as the experience with the least positive sentiments involved, while *Pure Pleasure* was the one containing the best topics experiences. Regarding the last coding, the results show that the *Low Ranked* reviewers appear to be more satisfied with the topics that they mention online (3.56), but the difference is not huge in comparison with the *High Ranked* ones (3.49).

As for *Bermuda,* the results indicate that the reviewers felt positive sentiments: an average polarity of 4.10; the reviewers even manifesting *disagreement* in any part of their reviews showed sentiments above neutral; every score, generally, is above 4 except for *disagreement* (3.79). It is therefore possible to conclude that readers will find the reviews as more useful into making a decision (Fang *et al.*, 2016), which should be a positive one (Ashraf *et al.*, 2016). Also, has seen in the literature, the higher the average polarity score, more probable it is that the reviews will influence potential tourists of a destination, as well as the higher the probability of turning potential customers into real ones (Alaei *et al.*, 2017; Chang *et al.*, 2009; Ribbink *et al.*, 2004). Thus, the reviews extracted, through the results obtained, are capable of acting as a positive mediator into bringing new tourists into the island. A positive score in the reviews considered as in *disagreement* (3.79) also confirms that despite not being 100% in favor of everything, the positive customer satisfaction makes reviewers continue to recommend the destination and spread information to potential customers (Bilro, 2017).

Finally, the *Topics Sentiment Analysis* demonstrate that the topics being mentioned online showing the most positive sentiments are: *Sciences*, *Technology* and *Services*; while the most negative sentiments ones were: *Religion*, *Travel* and *Measures*. Such analysis allows the respective entities of the *Tourism of Bermuda* to take action into re-adapting their strategies and "repair" the topics causing less positive sentiments in the tourists (Chen & Tseng, 2011). The results of the *Category* coding states that *Hotel* is the category with the topics causing the most positive sentiments in the reviewers, while *Leisure Activities* the most negative ones. On the *Tourism Experience Model* level, the *Holist* experience contains the best topics while *Knowledge Seeker* the worst. Regarding the *Membership Level of Reviewers* the results were opposite to the Madeira, as *High Ranked* ones (3.52) felt highly satisfied than the *Low Ranked* ones (3.49).

5. Conclusions and Implications

5.1 Theoretical and managerial implications

Regarding theoretical implications, this study has given a unique perspective on studying together, both the islands of *Madeira* and *Bermuda* on a reviewer's perspective. Combining the *Kozinets* netnographic process – with a set of unique codings as: *rating*, *category*, *membership level of reviewers*, *language type*, *tourism experience model*, *content*, *symbology* and *positive/negative information*; with a text mining process (namely, *Text Classification*, *Topics Extraction* and *Sentiment Analysis*) was also a new approach to the existing literature. Secondly, the *Tourism Experience Model* -created by Deans and Gnoth (2012)- is used in a unique way, as this study is able to provide an innovative approach in characterizing the two island destinations according to the different experiences proposed by the model. Thirdly, the methodologial struture of this study can be used for further research using any set of destinations, as well as it is a flexible methodology, which allows, for instance, to change the codings according to the objective of the project.

Concerning managerial implications, the current study allows to recommend the promotion of *Low Ranked Reviewers* into *High Ranked* ones, due to the fact that the language effect is attenuated when reviews are written by the second (Wu *et al.*, 2017) – this strategy might overcome lowering the customer's booking intention caused by *LRR* using figurative language. It could be done through a partnership between the refered entities of each island and the firms responsible for the *Social Networking Sites*.

Secondly, based on the results of the *Text Classification* and *Topics Extraction,* it is important that the communication/promotion strategies of the destinations (and the firms and brands that operate in the tourism industry) are based on *Lifestyle and Leisure* and *Tourism, Travel and Commuting*. Other themes might also play an important secondary role in these campaigns, such as *Art and Culture* and *Environment, Weather and Energy*. Thirdly, one consideration when promoting the islands should include the type of experiences that the island is known and sought by. The fact that both islands are mainly experienced through their *Recreational Activities*, makes it a key element when promoting it to potential tourists. Finally, this study permits firms to recognize the negative information circulating through the *TripAdvisor* online community regarding a destination, which allows to re-define strategies into turning that negative information into positive. On the other hand, the already positive information being mentioned, can also be analysed to reinforce the strengths of the destinations.

5.2 Limitations and future research

In a strategic point of view it is possible to claim that the limitations are avenues for future investigation. The following enlisted topics present the current limitations of this study, as well as possible future researches associated. First, this study is limited to a certain amount of reviews presented in *August of 2017* in the *TripAdvisor* platform, which provide a limited amount of data (1783 reviews). In future research the sample obtained, for Madeira and Bermuda, may be enlarged to provide more accurate results;

Second, the conducted study aimed merely two island destinations, in order to obtain more complete and rigorous results, the amount of island destinations should be increased;

Third, this study uses only one online community – *TripAdvisor*, which may or may not have skewed the results. In future research, more than one *Social Networking Site* should be used, to provide a comparison between the results obtained in each one of them;

Fourth, in future research, the methodology used can be applied to any destinations that are intended to be studied – which means that it is possible to study any destinations presented in the *TripAdvisor* platform using this methodology;

Lastly, the methodology of this study was limited to two different techniques: *Netnography* and *Text Mining*. In future research further techniques may be used to confront the results – e.g: questionnaries, interviews, etc., as well as create a new set of techniques to study these islands, or other destinations.

References

Alaei, A. R., Becken, S. & Stantic, B. (2017). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, 1-17. https://doi.org/10.1177/0047287517747753

Anderson, R. E. & Srinivasan, S. S. (2003). E-Satisfaction and E-Loyalty: A Contingency. *Psychology & Marketing*, 20(2),123–138.

Ashraf, A. R., Thongpapanl, N. & Spyropoulou, S. (2016). The connection and disconnection between e-commer businesses and their customers: Exploring the role of engagement, perceived usefulness and perceive ease-of-use. *Eletronic Commerce Research and Applications*, 20,69-86.

Bigné, L. & Gnoth, J. (2005). The Theme Park Experience: An Analysis of Pleasure, Arousal and Satisfaction. *Tourism Management*, 26, 833-844.

Bilro, R. G., 2017. **The role of consumer-brand engagement in a digital marketing era**. Unpublished doctoral dissertation, ISCTE - Instituto Universitário de Lisboa.

Chang, H. H., Wang, Y.-H. & Yang, W.-Y. (2009). The impact of e-service quality, customer satisfaction and loyalty on e-marketing: Moderating effect of perceived value. *Total Quality Management*, 20(4), 423 –443.

Chen, C. C. & Tseng, Y.-D. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50,755-768.

Cheng, I. C. C., Lam, L. W., Chow, C. W. & Fong, L. H. N. (2017). The effect of online reviews on hotel booking intention: The role of reader-reviewer similarity. *Journal of Hospitality Management*, 66, 54-65.

Chen, J. (2015). Examining the influence of membership levels on consumers' decision making. *Procedia Computer Science*, 55, 1258-1264.

Eastin, M. S., Daugherty, T. & Burns, N. M. (2011). *Handbook of Research on Digital Media and Advertising: User Generated Content Consumption.* New York: Information Science Reference.

Fang, B., Ye, Q., Kucukusta, D. & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52,98-506.

Fan, W., Wallace, L., Rich, S. & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49,77-82.

Feldman, R. (2013). Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 56(4),82-89.

Gilmore, J. H. & Pine, B. J. (2002). *Differentiating hospitality operations via experiences: Why selling services is not enough*, s.l.: Cornell Hotel and Restaurant Administration Quarterly.

Gnoth, J. & Deans, K. (2012). **Developing the tourism experience model**. Otago, Otago University.

Gnoth, J. & Matteucci, X. (2014). A phenomenological view of the behavioural tourism research literature. *International Journal of Culture, Tourism and Hospitality Research*, 8(1), 3-21.

Hirschman, E. C. & Holbrook, M. B. (1982). Hedonic Consumption: Emerging Concepts, Methods and Propositions. *Journal of Marketing*, 46, 92-101.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H. & Li, X. (2017). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*, 1-14. doi: https://doi.org/10.1177/0047287517729757

Kozinets, Robert V. (2010). **Netnography: Doing ethnography research online**. London: Sage Publications Ltd.

Lemon, K. N. & C.Verhoef, P. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing: AMA/MSI Special Issue*, 80:69-96.

Liu, B. & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In: C. C. Aggarwal & C. Zhai, eds. *Mining text data*. Charu C. Aggarwal • ChengXiang Zhai ed. New York: Springer: Springer, pp. 415-463.

Loureiro, S.M.C. (2014). The role of the rural tourism experience economy in place attachment and behavioral intentions. *International Journal of Hospitality Management*, 40(5), 1-9.

Mendes, J. d. C., Valle, P. O. d., Guerreiro, M. M. & Silva, J. A. (2010). The tourist experience: Exploring the relationship between tourist satisfaction and destination loyalty. *Original scientific paper*, 58(2), 111-126.

Otto, J. E. & Ritchie, J. R. B. (1996). The service experience in tourism. *Tourism Management*, 17(3),165-174.

Pine II, B.J. & Gilmore J.H. (1998). Welcome to the experience economy. *Harvard Business Review*, 76(4), 97-105.

ReviewTrackers, 2018. *reviewtrackers.com.* [Online] Available at: https://www.reviewtrackers.com/online-reviews-survey/Accessed in: 16 September 2018.

Ribbink, D., Riel, A. C. v., Liljander, V. & Streukens, S. (2004). Comfort your online customer: quality, trust and loyalty on the internet. *Managing Service Quality*, 14(6), 446-456.

Schuckert, M., Liu, X. & Law, R. (2015). Hospitality and tourism online reviews: recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32, 608–621.

Thompson, C. J., Locander, W. B. & Pollio, H. R. (1989). Putting Consumer Experience Back into Consumer Research: The Philosophy and Method of Existential-Phenomenology. *Journal of Consumer Research*, 16(2), 133-146.

Uriely, N. (2005). The Tourist Experience: Conceptual Developments. *Annals of Tourism Research*, 32(1),199–216.

Wang, N. (1999). Rethinking authenticity in tourism experience. *Annals of Tourism Research*, 26(2), 349-370.

Wu, L., Shen, H., Fan, A. & Mattila, A. S. (2017). The impact of language style on consumers0 reactions to online reviews. *Tourism Management*, 59, 590-596.

Zhong, L. & Daniel Leung, R. L. &. B. W. (2013). eTourism in China: A Review of the Chinese Language Literature. *Asia Pacific Journal of Tourism Research*, 18(5), 464-482.