Where You Sleep Tells What You Care About

Estela Marine-Roig(✉), Natalia Daries, Eva Martin-Fuentes, and Berta Ferrer-Rosell

University of Lleida, Lleida, Spain
{estela.marine,natalia.daries,eva.martin,berta.ferrer}@udl.cat

Abstract. Online reviews are valuable sources of information that can help other tourists gain an idea of the destination they seek. The aim of this research is to examine what types of content is reviewed by hotel guests and Airbnb guests in two of the main urban tourist cities in Spain (Barcelona and Madrid) to distinguish content between the destination where the accommodation is located and to observe variations on users of these platforms when they write reviews. Comparing more than one million reviews posted on TripAdvisor and Airbnb through compositional data analysis (CoDa), results show that there are different patterns between Barcelona and Madrid and between their accommodation mode according to the reviews. This study can help researchers and managers know what useful information can be extracted from online travel reviews.

Keywords: Peer-to-peer accommodation · User-generated content · Online travel reviews · Compositional data analysis · Spain

1 Introduction

Online travel reviews represent a prolific source of valuable information about consumers' preferences [1]. Their needs, wants, and demands and can guide businesses' proactive responses to cater to those preferences, especially by supplying satisfactory services and pre-purchase information [2]. In that way, reviews allow managers to analyze the virtues and defects of their businesses, motivate them to fix mistakes and solve problems, and, more broadly, help them to incrementally improve their companies [3]. By extension, by sharing their experiences in reviews that can shape the future of businesses, customers serve as co-creators of businesses, and their opinions and preferences can be used to generate more accurate user profiles and fine-tune customized recommendations. In parallel, competition between traditional and disruptive accommodations has spiked in the era of the collaborative economy, in which the success of business platforms depends on high levels of user engagement [4].

It is therefore valuable to know whether online reviews are different or perhaps even less valid for hotels versus P2P accommodations, which cannot be measured with star systems and have no established certificates of quality. In that case, readers can rely only on online reviews to form opinions about given services, which underscores the value of reviews as feedback that can be used to benefit consumers as well as the owners and managers of establishments [5].

This study aims to identify differential patterns in tourists' reviews according to the type of online travel platform used, focusing on content related to the destination's attraction factors: TripAdvisor (hotels) versus Airbnb (P2P accommodation) in the two most relevant tourism destination cities in Spain. Both TripAdvisor and Airbnb are useful to unveil users' destination image and preferences [6]. However, in addition to the different types of accommodation these platforms represent, some studies affirm that Airbnb reviews are overtly positive [7] and that the information they contain is shallow or presented in a specific way not really useful for customers. Besides, although TripAdvisor review content may be more extensive, the continuance in use of the platform is related to the credibility and usefulness of the review content [8]. Hence the importance of unveiling their potentially different patterns.

2 Literature Review

Perhaps the most notable provider of P2P products and services is Airbnb, whose outstanding marketing efforts have focused on the experiential side of its service—that is, offering opportunities to interact with hosts and local venues or to have authentic experiences at destinations [9]. Guests who choose Airbnb are frequent travellers who prefer the platform's accommodations that allow sharing experiences with friends, not families, and who have had good past experiences with and trust the brand [10]. Contrarily, guests who prefer hotels typically rate experiences with hotel websites highly, prioritise the comfort of dining in hotel restaurants, have been influenced by advertising, travel with their families, seek services offered by the establishment's infrastructure, and book accommodations via travel agencies, thus increasing their confidence in receiving refunds when necessary [11].

Few studies have focused on determining whether the same segment of consumers makes reservations in commercial hotels as in peer-to-peer accommodations such as Airbnb [12]. Some authors have argued that various micro-segments, including smokers, pet owners, long-term guests, groups of friends, and travellers who seek to share experiences with hosts, prefer to use P2P accommodations, whereas travellers looking for a more complete, professional, intimate service prefer conventional hotels [9, 13]. However, other researchers have affirmed that peer-to-peer accommodations and hotels share the same market in specific places and at specific times, particularly in countries or cities (e.g. Paris, London, and New York) where the available supply of accommodations is low, hotel beds are scarce, and the price of hotel rooms are excessive [12]. In those markets, P2P accommodations become competitive, especially when demand exceeds supply, during specific seasons or surrounding certain events (e.g., sporting events, concerts, and trade fairs), and where P2P accommodations make less use of dynamic pricing.

3 Methodology

3.1 Data Collection

In March 2020, more than one million reviews written from 2010 to 2019 were downloaded from Inside Airbnb (www.insideairbnb.com) and TripAdvisor.com, which corresponded to accommodations of the two most populated Spanish urban tourist destination

cities: Madrid and Barcelona. For this research, only English reviews (442,701 for Airbnb and 895,285 for TripAdvisor) were analyzed. The English reviews from InsideAirbnb were selected automatically from the dataset using OpenRefine, an open source application for handling big data. The filtering was done using Google Language detection tools with a Python scriplet from OpenRefine. After adding a blacklist of non-significant words, and a list of composite terms of interest, we proceeded to content analyze reviews through key-term counts (counting the number of times a certain key-term appears in reviews) with the KHCoder software. In the case of Airbnb reviews, this resulted in a total analysis of 31,543,003 words/terms for Barcelona, of which 81,504 were unique; and of 22,872,293 words in the case of Madrid, of which 62,823 were unique. In the case of TripAdvisor reviews, the process resulted in an analysis of 41,329,101 total words, of which 102,326 were unique; and of a total of 22,915,536 words in the case of Madrid, of which 74,474 were unique. Then, these keywords were classified through intercoder reliability technique into eight predetermined content categories on destination attraction factors [6], gaining percentages of classified words in each category. CoDa was used to deal with those percentages.

3.2 Data Analysis

When analyzing percentages, it is necessary to bear in mind their proportionality, otherwise results may be misinterpreted [14, 15]. In other words, Euclidean Distance consider that the pairs of percentages 1% to 2% and 11% to 12% are mutually distant (1% of difference), but in the first pair the proportional increase is of 100%, while in the second pair, it is of less than 10%. Consequences of not considering the characteristics of data carrying relative information can be found in Pawlowsky-Glahn et al. [16].

The most common approach to deal with data carrying relative information is to transform the data into logarithms of ratios [14]. They constitute a natural way of distilling the information about the relative size and tend to meet the distributional assumptions of classical statistical models. The so-called centered log-ratio (clr) transformation computes the log-ratios of each component (in this study, the content categories) over the geometric mean of all of the components (content categories), including its own. Once we have the clrs computed, compositional squared distance between two compositions x and y (platform per city x and platform per city y) assumes log-ratios carry all of the needed information about relative differences [17] and see the differences between cities and platforms for each clr summing the squares. Computations of compositional distance reveal which of the content categories contributes the most and the least to differentiating the platforms per cities. Thus, it is possible to measure which content categories contribute mainly to differentiate the pairs of cities and platforms.

4 Results

Barcelona, in relative terms has double the number of words referring to Sports than Madrid, as well as words referring to Sun, Sea, and Sand. Reviews from Madrid, however, contain more words referring to Food and Wine. It seems that there are differences between both cities, but proportionality of the content's appearance is respected in reviews of both platforms (Table 1).

Regarding compositional distance, Madrid presents the greater gap between platforms (0.822). The content categories contributing the most to differentiating platforms in Madrid are Sports (0.293) and Urban environment (0.2886).

Table 1. % of content categories per city and per platform, considering the keywords classified into the categories (first 4 columns) and Compositional distances and contributions to compositional distances between platforms (per cities) within content categories (last 2 columns).

Content Categories	Bcn TA	Bcn AB	Mad TA	Mad AB	Bcn TA-AB	Mad TA-AB
Food and Wine	26.98	20.94	29.59	22.22	0.025	0.012
Intangible Heritage	0.34	0.25	0.49	0.35	0.051	0.030
Leisure & recreational activities	16.15	11.48	16.30	12.61	0.060	0.006
Nature & active tourism	1.48	1.66	1.72	1.61	0.045	0.012
Sports	1.49	1.01	1.03	0.50	0.085	0.293
Sun, Sea, Sand	7.22	7.72	3.92	3.88	0.026	0.028
Tangible Heritage	13.69	15.19	15.83	14.43	0.040	0.007
Urban Environment	32.65	41.76	31.10	44.40	0.117	0.286
Compositional distance					0.671	0.822

5   Conclusions

The initial work concludes that reviews describe the characteristics that define the cities: Barcelona, as the Olympic host city and located on the coast; and Madrid, the imperial city with a wide gastronomy. TripAdvisor users comment very similar things (specific pattern), and Airbnb users also talk about very similar things among them (specific pattern) regardless of the city. This study seeks to provide researchers and tourism destinations with a straightforward method to compare user reviews from different platforms, as well as to obtain relevant information on user profiles and preferences, to improve communication strategies and contributes to understand the new consumer dynamics in the accommodation sector. Destination managers should consider differences in tourist preferences to develop new services, activities and experiences that meet the needs of all visitors.

In the future work of this study, we aim to analyse other urban destinations and to include the valence analysis (positive or negative) of reviews, to confirm the results and strengthen the theoretical contribution, as well as to focus on patterns about the accommodation features in more detail.

References

1. Marine-Roig, E.: Content analysis of online travel reviews. In: Fuchs, M., Gretzel, U., Höpken, W., Xiang, Z., (eds.) Handbook of e-Tourism, pp. 557–582. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-05324-6_31-1

2. Hsu, C.-L., Yu, L.-C., Chang, K.-C.: Exploring the effects of online customer reviews, regulatory focus, and product type on purchase intention: Perceived justice as a moderator. Comput. Human Behav. **69**, 335–346 (2017)

3. Clemons, E.K., Gao, G.G., Hitt, L.M.: When online reviews meet hyperdifferentiation: a study of the craft beer industry. J. Manag. Inf. Syst. **23**, 149–171 (2006)

4. Kesgin, M., Murthy, R.S.: Consumer engagement: the role of social currency in online reviews. Serv. Ind. J. **39**, 609–636 (2019)

5. Martin-Fuentes, E., Fernandez, C., Mateu, C., Marine-Roig, E.: Modelling a grading scheme for peer-to-peer accommodation: stars for Airbnb. Int. J. Hosp. Manag. **69**, 75–83 (2018)

6. Lalicic, L., Marine-Roig, E., Ferrer-Rosell, B., Martin-Fuentes, E.: Destination image analytics for tourism design: an approach through Airbnb reviews. Ann. Tour. Res. 86, 103100 (2021)

7. Zervas, G., Proserpio, D., Byers, J.W.: A first look at online reputation on Airbnb, where every stay is above average. Mark. Lett. **32**(1), 1–16 (2020). https://doi.org/10.1007/s11002-020-09546-4

8. Filieri, R., Acikgoz, F., Ndou, V., Dwivedi, Y.: Is TripAdvisor still relevant? The influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. Int. J. Contemp. Hosp. Manag. **33**, 199–223 (2021)

9. Tussyadiah, I.P.: An exploratory study on drivers and deterrents of collaborative consumption in travel. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, pp. 817–830. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14343-9_59

10. Zhang, G., Cui, R., Cheng, M., Zhang, Q., Li, Z.: A comparison of key attributes between peer-to-peer accommodations and hotels using online reviews. Curr. Issues Tour. **23**, 530–537 (2020)

11. Varma, A., Jukic, N., Pestek, A., Shultz, C.J., Nestorov, S.: Airbnb: Exciting innovation or passing fad? Tour. Manag. Perspect **20**, 228–237 (2016)

12. Haywood, J., Mayock, P., Freitag, J., Owoo, K.A., Fiorilla, B.: Airbnb & hotel performance. STR Publ (2017)

13. Cheng, M., Jin, X.: What do Airbnb users care about? An analysis of online review comments. Int. J. Hosp. Manag. **76**, 58–70 (2019)

14. Coenders, G., Ferrer-Rosell, B.: Compositional data analysis in tourism: review and future directions. Tour Anal **25**, 153–168 (2020)

15. Egozcue, J.J., Pawlowsky-Glahn, V.: Compositional data: the sample space and its structure. TEST **28**(3), 599–638 (2019). https://doi.org/10.1007/s11749-019-00670-6

16. Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R.: Modeling and analysis of compositional data. Wiley, Chichester (2015)

17. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., Pawlowsky-Glahn, V.: Logratio analysis and compositional distance. Math. Geol. **32**, 271–275 (2000)