Repositório ISCTE-IUL

What are Airbnb hosts advertising? A longitudinal essay in Lisbon

Structured Abstract:

Purpose- Considering the importance of the content created by the host for Airbnb consumers while making purchasing decisions, this study aims to analyze how the Airbnb hosts promote their properties by revealing the predominant attributes considered by hosts when advertising them.

Design/methodology/approach- The unstructured textual content of online Airbnb accommodations advertisements (property descriptions) is analyzed through a longitudinal text mining approach. We define a pipeline based on a Topic Modelling approach, that allows us not only to identity the most prevalent text attributes but also its distribution through time.

Findings- This research identifies and characterizes the attributes most advertised over time, on about 30,000 accommodations posted monthly over two years, between 2018 and 2020. Five main topics were identified in the data reflecting only pull motivations. Noteworthy is the slight changes in properties' descriptions topics along the two years, suggesting that "service" is increasingly being perceived by hosts as an important attribute of Airbnb guest experience.

Originality- Through a text analysis, this study provides an insight into peer-to-peer accommodation on the key attributes that hosts consider in the description of their properties to leverage the attractiveness of Airbnb. In the light of existing research, which has predominantly focused on the trustworthiness and attractiveness of the Airbnb advertisement, this research differentiates by analysing the main attributes in text over time. Given the Airbnb's changes since its inception, a longitudinal view is relevant to clarify how hosts advertise their properties and how it evolves in the light of these changes.

Keywords: Topic Modelling, Airbnb, hosts advertisements, push and pull factors, sharing economy

1. Introduction

The term "peer-to-peer sharing economy" has been frequently used to describe a business model that relies on digital technologies as a springboard to connect users/renters and owners/providers, allowing the efficient use and distribution of resources (goods or services) rather than private ownership (Parente *et al.*, 2018). The last decade has seen the emergence of sharing economy firms in many different industry sectors, empowered by the information and communication technology advances. Due to the nature of its service, the tourism and hospitality industry is the most affected by the peer-to-peer (P2P) sharing economy, with new business models in accommodation, transportation, restaurants, and travel planning being created (OECD, 2016).

A prominent example of the P2P sharing economy in the tourism industry is protagonized by Airbnb, an online platform through which individuals can offer their spaces/properties for short-term rentals. Airbnb presented an alternative value proposition based on cost-savings, household amenities and the potential for a more authentic local experience (Guttentag, 2019), becoming the leading provider of local accommodation in many tourist destinations around the world. Consequently, academic research about Airbnb has grown significantly in recent years. Scholars have been studying Airbnb from different angles, covering topics such as: advertising strategies (e.g., Liu and Mattila, 2017), guests motivations (e.g., Chiappa *et al.*, 2020; Mahadevan, 2018), value co-creation and co-destruction (e.g., Buhalis *et al.*, 2020), barriers to the use of Airbnb (e.g., Tussyadiah and Pesonen, 2018), the implications of its growth on the hotel industry (e.g., Zervas *et al.*, 2017) and the locals (e.g., Gant, 2016), among others.

Unlike traditional tourism accommodation such as hotels, where the brand and the star rating of a particular accommodation are indicators of the level of quality of the service and standards, the peer-to-peer market is characterized by heterogeneity (Einav *et al.*, 2016) and uncertainty

in its offer. On a peer-to-peer accommodation platform every property has its specific characteristics, and the hosts service is very variable dependent on host's level of professionalism. For that reason and since the P2P properties are not advertised on other promotional channels (e.g., T.V., radio, outdoors), guests use the information they can obtain on the platform, such as online reviews and host generated content in order to make purchasing decisions. To this end push and pull attributes (Crompton, 1979) of the lodging should be emphasized in host generated content, since those attributes are the enablers of guests decisions.

While online reviews have been examined in numerous studies, the host generated content, which provides information about the properties and the services, has been scarcely investigated. The advertisement of the property, with host created content, includes the name of the accommodation, photos, features of the space, amenities, prices, and cancellation policy, among other information useful for the potential guest. There is a description section with textual free-form fields dedicated to describing the space, the neighborhood, methods of getting around, interactions with guests, house rules and additional services. This description section provides meaningful information about what hosts consider important to leverage the attractiveness of their property.

Responding to Liang's et al. (2020) call on identifying the aspects or items in property descriptions written by hosts, as well as the call for longitudinal research (Correia and Kozak, 2021), this study explores the content of the textual description of properties on about 30,000 accommodations through a Topic Modelling approach over two years (between 2018 and 2020) that allows to ensure consistency in the findings.

We aim to identify the main topics chosen by Airbnb hosts to promote their properties and understand if they are aligned with guests motivations in the light of the push-pull approach

(Crompton, 1979). Therefore, we address a first research question: 1) "What are the topics considered by hosts when advertising their properties?"

Furthermore, under the ongoing debate on the increasing professionalization of Airbnb and whether or not Airbnb has already lost its sharing economy ethos (Demir and Emekli, 2021), it becomes relevant to analyze the dynamic nature of primary hosts concerns when advertising their properties. Accordingly, we address a second research question: 2) "How do these topics evolve over time?"

This research analyses the properties descriptions in Lisbon, the capital of Portugal, which has had a rapid and sustained growth over the last decades, evolving to an important tourist destination in Europe and in the world.

This article analyses the textual information publicized by the host and contributes to a better understanding, by practitioners and researchers, of how accommodations are advertised, a relatively under-studied problem in the literature (Liu and Mattila, 2017; Wisker *et al.*, 2019), and provides relevant input to decision making in this domain.

2. Literature review

This section starts by addressing the literature on Airbnb advertisements, the object of our study. Since understanding consumer motivations is an important competitive advantage to produce successful marketing strategies (Peter and Olson, 2009), we also address in this section the consumer motivations for choosing Airbnb that will serve as a basis for discussing the topics found in this study.

2.1. Airbnb advertisements

Given the nature of peer-to-peer accommodation, the effectiveness of the advertisement about a property relies both on the online reputation (e.g., guest rating scores and text reviews) acquired through time, and on the content produced by the host (Wisker *et al.*, 2019).

When compared to the content generated by the host, guest reviews have received more academic attention in recent years. Although some studies question the authenticity of this content (Zhang, 2019), a number of empirical studies identify an influence of the review system on trust and purchase intentions (Chen and Chang, 2018), and so guest reviews are considered an important information source that other customers rely on to make purchase decisions. The content of text reviews, focusing on the extraction of attributes from text, has received increasing attention in recent years. Although the order of importance of each Airbnb attribute varies between studies, the content analysis based on guest reviews tends to focus on a similar collection of themes relating to the relationship with the host, the amenities of the property, and the convenience of the location (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017).

However, the review system (text reviews and ratings) alone is not enough for guests making purchasing decision. Moreover, reviews/ratings are uniformly high on websites like Airbnb (Bridges and Vásquez, 2018; Zhang, 2019), which makes it even more difficult for the guests to effectively differentiate properties (Ert *et al.*, 2016). Host generated content provides more complete and objective information about the stay, and studies have already confirmed its effect on trust and purchase decision in Airbnb (Ert et al., 2016; Liang et al., 2020; Tussyadiah & Park, 2018; Zhang et al., 2020).

Considering the importance of the seller's personal information in sharing economy, several studies have assessed the impact of the host self-description and personal photos on trust perception and booking intention. Tussyadiah and Park (2018) identify two self-presentation

patterns on the words hosts use to describe themselves online, through hierarchical cluster analyses (document clustering): hosts present themselves as well-travelled ("Travelers") or as having specific professions ("Workers"). Based on a questionnaire applied to consumers to rate the host (based on its description) in terms of trustworthiness and likelihood to book, the study demonstrates higher levels of perceived trustworthiness and booking intention to well-travelled hosts. Zhang (2020) applied a text analytics framework to identify the impact of linguistic and semantic features on trust perception. The topics hidden in self-descriptions were identified by Latent Dirichlet Allocation, and a deep-learning procedure was employed to automatically code trust perception. The results show that both readability and perspective taking of the self-description have a significant positive impact on trust perception, as well as the presence of topics related to family relationships, openness, service, and travel experience. Ert et al. (2016) analyzed Airbnb advertisements in Stockholm and conducted two experiments to investigate whether personal photos of hosts on Airbnb affect consumers' decision. They conclude that guests infer the host's trustworthiness from hosts photos, and that these photos can have a significant impact on guests' decision making.

Considering hypothetical advertising techniques, Liu and Mattila (2017) examine which advertising appeal is more effective: a belonging ("feeling at home") or uniqueness ("atypical places to stay") appeal. They find that "powerless" individuals exhibit higher levels of click-through intention and purchase intention to hypothetical Airbnb advertising focusing on belongingness, whereas "powerful" individuals respond more favorably to hypothetical Airbnb advertising focusing on uniqueness. They explain that these effects are mediated by the individual's connection to the brand.

Regarding exploring of real properties descriptions, Liang et al. (2020) have analyzed the effectiveness of the textual descriptions generated by hosts in influencing guests' booking and review-posting decisions. Specifically, the results show that both the width (number of

described items) and depth (level of detail) of properties descriptions are positively associated with their review volume.

There has been a research effort to understand the influence of guest/host generated content in trust perception and purchase intention/decision. However, few studies have focused on extracting attributes from the description of the properties, written by hosts, which reflects the host primary concerns and preferences when advertising the property. Lutz and Newlands (2018) search for cues in properties descriptions relating to demographic markers, travel modalities, and approaches to sociality, and compared it to the type of accommodation. They found that *entire home* hosts explicitly target business travelers, high-income guests, and couples looking for a romantic experience. In contrast, *shared room* hosts target younger guests that search for lower-priced experience and mention social interaction as an integral part of the experience.

The literature focusing on a global view of the main attributes referred by hosts to leverage the attractiveness of their property is still scarce. Given the Airbnb's changes since its inception, such as the proliferation of multi-unit hosts, and the professionalization of the platform (Dogru *et al.*, 2020), a longitudinal view is relevant to clarify how hosts advertise their properties and how it evolves in the light of these changes. To the best of our knowledge, this is the first study to perform a content analysis of properties description while employing a Topic Modelling approach, allowing, not only to identity the most prevalent factors considered by hosts when advertising the property, but also their pattern in time.

2.2. Travel motivations on Airbnb

Given the rapid emergence of Airbnb, numerous studies have been examining the factors that drive or deter from choosing Airbnb. Research shows that the practical/ utilitarian benefits (e.g., price, location, and household amenities) (Guttentag *et al.*, 2018; So *et al.*, 2018) and

experiential benefits (e.g., social interaction, authenticity, and novelty) (Guttentag *et al.*, 2018; Tussyadiah and Pesonen, 2016) are major factors motivating consumers to choose Airbnb.

Although there is no widely accepted theoretical or conceptual framework for understanding travel motivation, the push–pull framework initiated by Crompton (1979) provides a useful approach for examining travel motivations, that has been regularly used in tourism literature (Fluker and Turner, 2000; Kim *et al.*, 2003).

The concept behind this approach is that people travel because they are pushed by intrinsic desires and pulled by characteristics of a certain travel product. Pull factors include tangible resources, such as price, location, and physical aspects of the accommodation (Wang *et al.*, 2015). By contrast, the push factors are intangible motives which sustain the desire to travel, for example, uniqueness-seeking, interpersonal experience, and social networking (Wang *et al.*, 2015).

3. Dataset and variables

This study uses a dataset obtained from the Inside Airbnb website (http://insideairbnb.com/), which is sourced from public information available on the Airbnb website. The data presented in this website is a collection of monthly snapshots of listings available at a particular time for many cities around the world. The dataset consists of three main tables: ***listings***, which details accommodations data, showing 106 attributes for each one of the records; ***reviews***, which details the reviews given by the guests, and ***calendar***, which provides details about booking availability for 365 days in the future for each listing (active rental).

This study uses the subset of the ***listings*** table corresponding to the Lisbon territory, restricted to accommodations for which an English-language description is available, and considering the snapshots between April 2018 and April 2020. It is noteworthy that the hosts

advertisements analyzed do not reflect the Covid-19 pandemic effect, since the pandemic situation started to freeze tourism activity in Lisbon by the end of March 2020. From the ***listings*** table, the following attributes are selected: *Id, latitude, longitude, neighbourhood_group_cleansed* (with information about the municipality in which the property is located, e.g., Cascais), and eight textual attributes related to the property description (*summary, space, neighborhood_overview, notes, transit, access, interaction, house_rules*) that were concatenated into a single attribute *description_complete*. Since all these eight attributes contain textual data, this aggregation is made to provide more meaningful context to the adopted Topic Modelling approach. The average length of the *description_complete* is 328 words (terms), varying between 1 and 1,232 words.

The number of accommodations is not fixed during the two years' time span. Our dataset has an average of 18,139 listings (accommodations) per month, varying between 15,533 accommodations (April 2018) and 18,728 accommodations (April 2020).

4. Methodology

Topic Modelling is a natural language processing and text mining method that can be applied to a set of text documents for discovering the latent abstract "topics" occurring in those documents in an unsupervised way. This technique has increasingly been used in research because it is an automated process that enables the analysis of much larger text collections than would be feasible to process manually. In the tourism and hospitality industry, Topic Modelling has been widely used to extract important attributes of the product/service from online reviews of hotels (Guo *et al.*, 2017), tourist attractions (Taecharungroj & Mathayomchan, 2019) or local accommodation (Cheng & Jin, 2018). The most common method for Topic Modelling is the Latent Dirichlet Allocation (LDA) for its better probability statistical foundation that overcomes the overfitting problem of other methods (Blei *et al.*, 2003). The LDA approach

requires that the number of topics to be discovered in the text is chosen by the researcher. LDA generates "topics" as lists of words based on the likelihood of words to occur within documents and infers mixtures of these topics in each document.

Figure 1 illustrates the adopted research method. The general framework consists of creating the LDA model with the accommodation's descriptions (listings) posted between April 2018 and April 2020 and then applying the model on aggregated documents.

[Figure 1 near here]

As usually done in text mining, due to the large number of non-informative words and characters in the online texts, a text pre-processing stage is required to improve the effectiveness and efficiency of extracting the important attributes.

The pre-processing of the accommodations description texts encompasses the following steps: (1) remove non-alphabetic characters; (2) convert all text into lowercase letters; (3) filter the accommodations with English text (77% of total accommodations); (4) remove stop-words (common words with little meaning, such as articles and auxiliary verbs) and low frequency words (terms that are not present in at least 15% of the documents).

After pre-processing the data, we have created the LDA model by considering the description of each accommodation, obtained in a particular snapshot (time point of the month), as a separate document. The number of topics was fixed based on the semantic coherence of different models, ranging from two to eight topics. Topic coherence measures have been proposed in the literature to evaluate the interpretability of topics learned from a statistical topic model (Röder et al., 2015). Such measures, based on word similarity of the top words of a topic, are grounded on a reference corpus. In our experiments, the *Cv* measure, proposed for LDA by Röder et al. (2015), achieved its maximum values for 4, 5, and 6 topics. The final 5-

topic model solution was adopted based on a qualitative manual evaluation that considered its clearness and interpretability, and also based on its proximity to other tourism literature.

In order to observe the evolution of topics over time, all the accommodations descriptions from the same snapshot were aggregated in one document (time-level aggregation), leading to a dataset composed by 24 documents used for inference.

In addition, the discussion of the results is made considering the difference between professional and non-professional hosts. One of the criteria considered in the literature for distinguishing professional from non-professional accommodations is the number of properties offered by a single host (Adamiak, 2019). In this study, professional accommodations are considered the accommodations managed by providers who host more than three properties.

The implementation of the mentioned methodology was done using Python. The data pre-processing was performed by Python Natural Language Processing libraries, such as: *re* (Van Rossum, 2020) for regular expressions operations; *langdetect* for language detection; *Natural Language Toolkit* (Bird *et al.*, 2009) for removing stop-words and *scikit-learn* (Pedregosa *et al.*, 2011) for converting the collection of text documents to a matrix of token counts. The LDA method was implemented with *gensim* (Rehurek and Sojka, 2011), a Python library designed specifically for Topic Modelling. For data manipulation, such as filtering or data aggregation, we have used the *pandas* library (McKinney and others, 2010), a high-level data manipulation tool.

5. Results and discussion

Taking into account the method illustrated in Figure 1, this section is organized as follows: first we describe the topics identified in accommodation's descriptions by applying the LDA model; then, we present the results of inferring the topics through time.

5.1. Topics identified in accommodation's descriptions

Topic Modelling analysis produced five topics: Topic 1: "Location (Lisbon city center)", Topic 2: "Location (Lisbon Outskirts)", Topic 3: "Service", Topic 4: "Amenities" and Topic 5: "Accessibility". Figure 2 represents the topics in two dimensions, corresponding to the first two Principal Components (PC), and was adapted from the result produced by pyLDAvis tool (Sievert and Shirley, 2015). Each circle in the image represents one topic. The area of the circle represents the frequency of each topic over the entire corpus, while the distance between circles indicates the similarity between topics. The smaller the distance, the more common words the topics have. These representations are complemented by the lists of the most frequent terms for each topic along with their overall frequency.

[Figure 2 near here]

The first topic is mainly related to geographic places that surround the city center, including terms exclusive of the historic center of Lisbon such "alfama", "bairro alto", and "chiado" (neighborhoods of the historic center). The second topic, complementary to the previous one, is associated with locations at Lisbon outskirts like "cascais" and "sintra". Both topics refer to the description of the locations and characteristics of the neighborhoods where the accommodations are placed. The first topic emphasizes the central location of the accommodations in the traditional neighborhoods in the heart of Lisbon and describes the proximity to points of interest such as historic monuments (e.g., São Jorge Castle), typical Portuguese restaurants or the river. Topic 2 emphasizes the privacy and quietness of the neighborhoods, characteristic of Lisbon outskirts, as well as the proximity to the beach. As previous research has shown based on guest reviews, a nice view from the apartment, a central location, and the vitality and safety of the neighborhoods are important attributes in guest satisfaction (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017). On the

other hand, neighborhood with bad ambiance and distant from points of interest is perceived negatively and may discourage users from choosing a particular accommodation.

Topic 3, labelled as "Service" refers to the Airbnb service performance, highlighting the availability of the host, the house rules, extra services and guest arrival and departure (check-in and check-out). The topic "Service" reveals some host service characteristics, such as helpfulness (e.g., "*we are here to help with whatever you need and give you the same tips we give our foreign friends*"), flexibility (e.g., "*Check-in is at 15:00, but can be scheduled at a different time, according to both needs.*"), responsiveness and prompt communication (e.g., "*During your stay I will be available by phone/email*") and how guest is welcomed (e.g., "*one of us will meet you personally and happily help you plan out a memorable route*"); characteristics that are consistently proposed as an important attribute of guest evaluation (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017). However, this topic essentially captures the functional/logistic value of service, not focusing on the host-guest interaction and experience.

Topic 4, labelled as "Amenities", focus on physical aspects of the accommodation, including rooms descriptions (e.g., "kitchen", "bathroom", "bedroom"), the availability of equipment/furnishings (e.g., "fully equipped", "tv", "machine", "sofa", "bed") or other supporting goods (e.g., "wifi", "towels"). This thematic is also found recurrently in the previous studies on Airbnb guests reviews (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017). These studies reveal that guests value the quality of the space (cleanliness, comfort, well-equipped) and its privacy, as well as facilities for daily use. The provision of amenities that are not expected is appreciated by Airbnb guests, while negative feedback is given when the facilities were not accurate to what was described.

Finally, Topic 5, labelled as "Accessibility" highlights a pedestrian scale of short/walking distances and access to public transportation (e.g., train station, airport, subway, bus). As well as apartments view and neighborhoods ambiance, convenience of transportation (proximity to public transportation, parking, and walking distances) has been consistently reported in the literature as an attribute valued by guests, since it defines how the tourist will explore the destination (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017).

Although these topics are in accordance with the attributes identified in the literature on guest reviews; our macro analysis shows that property descriptions do not emphasize the concepts of *authenticity*, *novelty* and *social interaction* of the accommodation experience, pillars of Airbnb's original value proposition (Guttentag, 2019). In order to attract guests to choose their accommodation, hosts are focusing on the pull motivations such as the location, accessibility, and the functional attributes of the stay (amenities and procedures).

Property descriptions include a number of implicit sentences that enable push motivations: such as authenticity-seeking ("*you'll be able to experience authentic life*"), uniqueness-seeking ("*plan out a memorable route*"), or a call to dream about ("*The simplest country house that one can imagine!*"). However, the appeal to the intrinsic drivers (push motivations) is done punctually, not being expressed at a macro view.

Figure 3 shows an excerpt of descriptions that illustrate the presence of each one of the topics. In these examples, we have highlighted the most prominent words in the text, according to the corresponding topic, to better illustrate the presence of the topic in the text.

[Figure 3 near here]

5.2.   *Inference of topics through time*

A longitudinal view of the topics enables us to track the shift in hosts primary concerns. Figure 4 shows the monthly evolution of topics, revealing a constant increase of the proportion of the

topic "Service" over time (3 percentual points in two years), compensated with a decrease of the proportion of the topics related to location.

[Figure 4 near here]

The topic "Service" is strongly related to the logistics of the service, such as house rules, extra services, and guest arrival and departure (check-in and check-out), while also captures some host service characteristics, such as flexibility and availability.

The results of the longitudinal analysis show that hosts are increasingly aware of the importance of their own service, being a growing concern when promoting their properties. However, with regard to their role in the accommodation experience, hosts are mainly concerned with showing themselves as facilitators (Cheng and Jin, 2019), offering solutions to the problems during the stay, and showing flexibility in accommodating guests requests (particularly with check-in and check-out times), rather than guest-host social relationship builders (Tussyadiah and Pesonen, 2016).

In parallel with the evolution of topics, Figure 4 also shows the number of accommodations hosted by professional and non-professional providers (considering the criteria mentioned in Section 4). The results show that the offer of professional hosts is growing more quickly than non-professional hosts, as verified in other countries in the study of Adamiak (2019). The increasing professionalization of Airbnb can also help to explain the increase focus of hosts on service. However, further research is required to explain this pattern in the primary concerns of hosts.

6. Conclusions and implications

Airbnb accommodation's descriptions act as an advertisement for the property and so reflect what the hosts believe will leverage the attractiveness of their property. This research applied

a Topic Modelling approach to the accommodation's descriptions, posted by hosts, and identified five main topics: "Location (Lisbon city center)", "Location (Lisbon outskirts)", "Service", "Amenities", and "Accessibility".

These topics are in accordance with the main attributes found in the guest reviews (home benefits, neighborhoods ambiance, convenience of transportation, central location and service) in previous studies (Celata *et al.*, 2020; Cheng and Jin, 2019; Tussyadiah and Zach, 2017),suggesting that hosts are aligned with guests' expectations. However, the aspects related to the emotional-based experiential drivers such as host-guest interaction and authentic experience, are not being captured by our macro view. The topics found in our study reflect attributes of the travel product (pull factors), providing evidence that the internal drives (push factors) that inspire someone to travel are not being valued by hosts as a promotion strategy.

Besides providing a global picture of the primary concerns of hosts, this study analyzes how they evolve over two years. Our longitudinal analysis shows a steady increase in the proportion of the "Service" topic over time, offset by a decrease in the proportion of location-related topics, revealing that hosts are increasingly concerned with promoting their own service.

In the light of existing research, which has predominantly focused on the trustworthiness and attractiveness of the Airbnb advertisement, this work differentiates by analysing the hosts perspective of an attractive Airbnb experience in the text, exploring its main attributes in view of push-pull motivations to choose Airbnb. Methodologically, this paper contributes to recent research efforts on extracting meaningful information from large, unstructured human-authored text data, and by defining a pipeline based on a Topic Modelling approach, that allows us not only to identity the most prevalent text attributes but also its distribution through time.

From the managerial standpoint, the findings of this study help Airbnb practitioners to reflect about the value proposition they promote through text. Currently, hosts are promoting the

practical/utilitarian benefits (pull factors) of their accommodations and giving less attention to the value of experiential benefits (e.g., social interaction, authenticity, and novelty).

Since the main tangible attributes (price, location, and household benefits) can be promoted through other visual content of Airbnb website, such as the map, photos, or pre-defined lists of amenities, the text description section should appeal to the intrinsic desires of guests (push factors). Furthermore, Airbnb hosts benefit from using the brand name "Airbnb", however, they are not entirely promoting the original value proposition of Airbnb, supported on authentic local experiences and social interaction. In that sense, the text description written by hosts should focus more on the genuine sense of the experience that guests could have (e.g., "Here you'll be able to experience authentic life in a historic neighborhood in the very heart of Lisbon.", "You can feel the typical rituals of a Lisboner"), and on the positive impact that hosts can have in their experience (e.g., "One of us will meet you personally and happily help you plan out a memorable route", "If i cannot be there for some reason, my family also lives near by.") .

This research brings to light the urgency of defining and promoting what distinguishes Airbnb from conventional forms of accommodation. While hotels seem to be adopting Airbnb's strategies to be closer to customers, Airbnb seems to lose some of its initial characteristics that have justified its growth, which can end up in a non-differentiated strategy.

7. Limitations

Despite the significance of this study, it has still some limitations that should be addressed in further research. First, this study applies LDA, a popular unsupervised Topic Modelling approach to efficiently discover the topics and weights from a large corpus of text. However, standard LDA does not allow the use of supervised labels to incorporate independent expert knowledge into the learning process. Future research could apply a semi-supervised approach,

where a sample of documents (accommodation's descriptions) is manually read and tagged, with the topics observed by independent expert labelers, to guide the model learning. Such semi-supervised approach could improve document classification and the alignment of the resulting model with human expectations for Topic Modelling and extraction. Second, this research presents a macro view of the main topics present in the text written by hosts. Future research could focus on micro differences between the attributes in guest reviews and the attributes in the respective description written by the host. A finer analysis could clarify about the alignment between guests and hosts perspective of the Airbnb experience. Third, we focused our research in Lisbon, a comparison with other cities may provide additional and more generalizable evidence. Fourth, the current study is restricted to descriptions written in English, but it would be interesting to incorporate data in a bilingual or multilingual context for cross-cultural analysis. Finally, this paper explores data before the COVID-19 pandemic in Lisbon, future comparative studies between the advertisements written before and after the beginning of the pandemic would be a useful next step forward that could demonstrate a greater focus on hygiene and safety.

References

Adamiak, C. (2019), "Current state and development of Airbnb accommodation offer in 167 countries", *Current Issues in Tourism*, Taylor & Francis, Vol. 0 No. 0, pp. 1–19. doi:10.1080/13683500.2019.1696758.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, " O'Reilly Media, Inc."

Blei, D.M., Ng, A.Y. and Jordan, M.T. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3 No. January 2001, pp. 993–1022.

Bridges, J. and Vásquez, C. (2018), "If nearly all Airbnb reviews are positive, does that make them meaningless?", *Current Issues in Tourism*, Taylor & Francis, Vol. 21 No. 18, pp. 2065–2083. doi:10.1080/13683500.2016.1267113.

Buhalis, D., Andreu, L. and Gnoth, J. (2020), "The dark side of the sharing economy: Balancing value co-creation and value co-destruction", *Psychology and Marketing*, John Wiley & Sons, Ltd, Vol. 37 No. 5, pp. 689–704. doi:10.1002/mar.21344.

Celata, F., Capineri, C. and Romano, A. (2020), "A room with a (re)view. Short-term rentals, digital reputation and the uneven spatiality of platform-mediated tourism", *Geoforum*, Elsevier, Vol. 112 No. August 2019, pp. 129–138. doi:10.1016/j.geoforum.2020.04.007.

Chen, C.C. and Chang, Y.C. (2018), "What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness", *Telematics and Informatics*, Elsevier, Vol. 35 No. 5, pp. 1512–1523. doi:10.1016/j.tele.2018.03.019.

Cheng, M. and Jin, X. (2019), "What do Airbnb users care about? An analysis of online review comments", *International Journal of Hospitality Management*, Elsevier, Vol. 76 No. September 2017, pp. 58–70. doi:10.1016/j.ijhm.2018.04.004.

Chiappa, G. Del, Sini, L. and Atzeni, M. (2020), "A motivation-based segmentation of italian airbnb users: An exploratory mixed method approach", *European Journal of Tourism Research*, Vol. 25 No. 2020, pp. 1–20. doi:10.54055/ejtr.v25i.420.

Correia, A. and Kozak, M. (2021), "Past, present and future: trends in tourism research", *Current Issues in Tourism*, Taylor & Francis, Vol. 0 No. 0, pp. 1–16. doi:10.1080/13683500.2021.1918069.

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

Demir, E. and Emekli, G. (2021), "Is Airbnb no longer a sharing economy platform? Evidence from Europe's top 10 Airbnb destinations", *Anatolia*, Routledge, Vol. 00 No. 00, pp. 1–19. doi:10.1080/13032917.2021.1890626.

Dogru, T., Mody, M., Suess, C., Line, N. and Bonn, M. (2020), "Airbnb 2.0: Is it a sharing economy platform or a lodging corporation?", *Tourism Management*, Vol. 78 No. November, pp. 0–17. doi:10.1016/j.tourman.2019.104049.

Einav, L., Farronato, C. and Levin, J. (2016), "Peer-to-Peer Markets", *Annual Review of Economics*, Vol. 8, pp. 615–635. doi:10.1146/annurev-economics-080315-015334.

Ert, E., Fleischer, A. and Magen, N. (2016), "Trust and reputation in the sharing economy: The role of personal photos in Airbnb", *Tourism Management*, Elsevier Ltd, Vol. 55, pp. 62–73. doi:10.1016/j.tourman.2016.01.013.

Fluker, M.R. and Turner, L.W. (2000), "Needs, motivations, and expectations of a commercial whitewater rafting experience", *Journal of Travel Research*, Vol. 38 No. 4, pp. 380–389. doi:10.1177/004728750003800406.

Gant, A.C. (2016), "Holiday rentals: The new gentrification battlefront", *Sociological Research Online*, Vol. 21 No. 3, pp. 112–120. doi:10.5153/sro.4071.

Guo, Y., Barnes, S.J. and Jia, Q. (2017), "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation", *Tourism Management*, Elsevier Ltd, Vol. 59, pp. 467–483. doi:10.1016/j.tourman.2016.09.009.

Guttentag, D. (2019), "Progress on Airbnb: a literature review", *Journal of Hospitality and Tourism Technology*, Vol. 10 No. 3, pp. 233–263. doi:10.1108/JHTT-08-2018-0075.

Guttentag, D., Smith, S., Potwarka, L. and Havitz, M. (2018), "Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study", *Journal of Travel Research*, Vol. 57 No. 3, pp. 342–359. doi:10.1177/0047287517696980.

Kim, S.S., Lee, C.K. and Klenosky, D.B. (2003), "The influence of push and pull factors at Korean national parks", *Tourism Management*, Vol. 24 No. 2, pp. 169–180. doi:10.1016/S0261-5177(02)00059-6.

Liang, S., Schuckert, M., Law, R. and Chen, C.C. (2020), "The importance of marketer-generated content to peer-to-peer property rental platforms: Evidence from Airbnb", *International Journal of Hospitality Management*, Elsevier, Vol. 84 No. February 2018, p. 102329. doi:10.1016/j.ijhm.2019.102329.

Liu, S.Q. and Mattila, A.S. (2017), "Airbnb: Online targeted advertising, sense of power, and consumer decisions", *International Journal of Hospitality Management*, Elsevier Ltd, Vol. 60, pp. 33–41. doi:10.1016/j.ijhm.2016.09.012.

Lutz, C. and Newlands, G. (2018), "Consumer segmentation within the sharing economy: The case of Airbnb", *Journal of Business Research*, Elsevier, Vol. 88 No. December 2017, pp. 187–196. doi:10.1016/j.jbusres.2018.03.019.

Mahadevan, R. (2018), "Examination of motivations and attitudes of peer-to-peer users in the accommodation sharing economy", *Journal of Hospitality Marketing and Management*, Routledge, Vol. 27 No. 6, pp. 679–692. doi:10.1080/19368623.2018.1431994.

McKinney, W. and others. (2010), "Data structures for statistical computing in python", *Proceedings of the 9th Python in Science Conference*, Vol. 445, pp. 51–56.

OECD. (2016), *OECD Tourism Trends and Policies 2016 PART I Policies for the Tourism Sharing Economy*, OECD Publishing, available at: www.Blablacar.com,.

Parente, R.C., Geleilate, J.M.G. and Rong, K. (2018), "The Sharing Economy Globalization Phenomenon: A Research Agenda", *Journal of International Management*, Elsevier, Vol. 24 No. 1, pp. 52–64. doi:10.1016/j.intman.2017.10.001.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., *et al.* (2011), "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

Peter, J.P. and Olson, J.C. (2009), *Consumer Behavior & Marketing Strategy*, *Dana*.

Rehurek, R. and Sojka, P. (2011), "Gensim--python framework for vector space modelling", *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, Vol. 3 No. 2.

Van Rossum, G. (2020), *The Python Library Reference, Release 3.8.2*, Python Software Foundation.

Sievert, C. and Shirley, K. (2015), "LDAvis: A method for visualizing and interpreting topics", pp. 63–70. doi:10.3115/v1/w14-3110.

So, K.K.F., Oh, H. and Min, S. (2018), "Motivations and constraints of Airbnb consumers: Findings from a mixed-methods approach", *Tourism Management*, Elsevier Ltd, Vol. 67, pp. 224–236. doi:10.1016/j.tourman.2018.01.009.

Tussyadiah, I.P. and Park, S. (2018), "When guests trust hosts for their words: Host description and trust in sharing economy", *Tourism Management*, Elsevier Ltd, Vol. 67, pp. 261–272. doi:10.1016/j.tourman.2018.02.002.

Tussyadiah, I.P. and Pesonen, J. (2016), "Impacts of Peer-to-Peer Accommodation Use on Travel Patterns", *Journal of Travel Research*, Vol. 55 No. 8, pp. 1022–1040. doi:10.1177/0047287515608505.

Tussyadiah, I.P. and Pesonen, J. (2018), "Drivers and barriers of peer-to-peer accommodation stay– an exploratory study with American and Finnish travellers", *Current Issues in Tourism*, Vol. 21 No. 6, pp. 703–720. doi:10.1080/13683500.2016.1141180.

Tussyadiah, I.P. and Zach, F. (2017), "Identifying salient attributes of peer-to-peer accommodation experience", *Journal of Travel and Tourism Marketing*, Routledge, Vol. 34 No. 5, pp. 636–652. doi:10.1080/10548408.2016.1209153.

Wang, T., Luo, Y. and Rebecca, L. (2015), "The investigation of consumer motivations to patronize boutique hotels using push-pull theory : a case study in Xiamen, China", *International Journal*

of Tourism Cities, Vol. 1 No. 4, pp. 317–328. doi:10.1108/IJTC-08-2014-0010.

Wisker, Z.L., Kadirov, D. and Bone, C. (2019), "Modelling P2P Airbnb online host advertising effectiveness: the role of emotional appeal information completeness creativity and social responsibility", *International Journal of Culture, Tourism, and Hospitality Research*, Vol. 13 No. 4, pp. 505–523. doi:10.1108/IJCTHR-03-2019-0045.

Zervas, G., Proserpio, D. and Byers, J.W. (2017), "The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry", *Journal of Marketing Research*, Vol. 54 No. 5, pp. 687–705. doi:10.1509/jmr.15.0204.

Zhang, J. (2019), "What's yours is mine: exploring customer voice on Airbnb using text-mining approaches", *Journal of Consumer Marketing*, Vol. 36 No. 5, pp. 655–665. doi:10.1108/JCM-02-2018-2581.

Zhang, Lz., Yan, Q. and Zhang, L. (2020), "A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb", *Decision Support Systems*, Elsevier, Vol. 133 No. September 2019, p. 113288. doi:10.1016/j.dss.2020.113288.
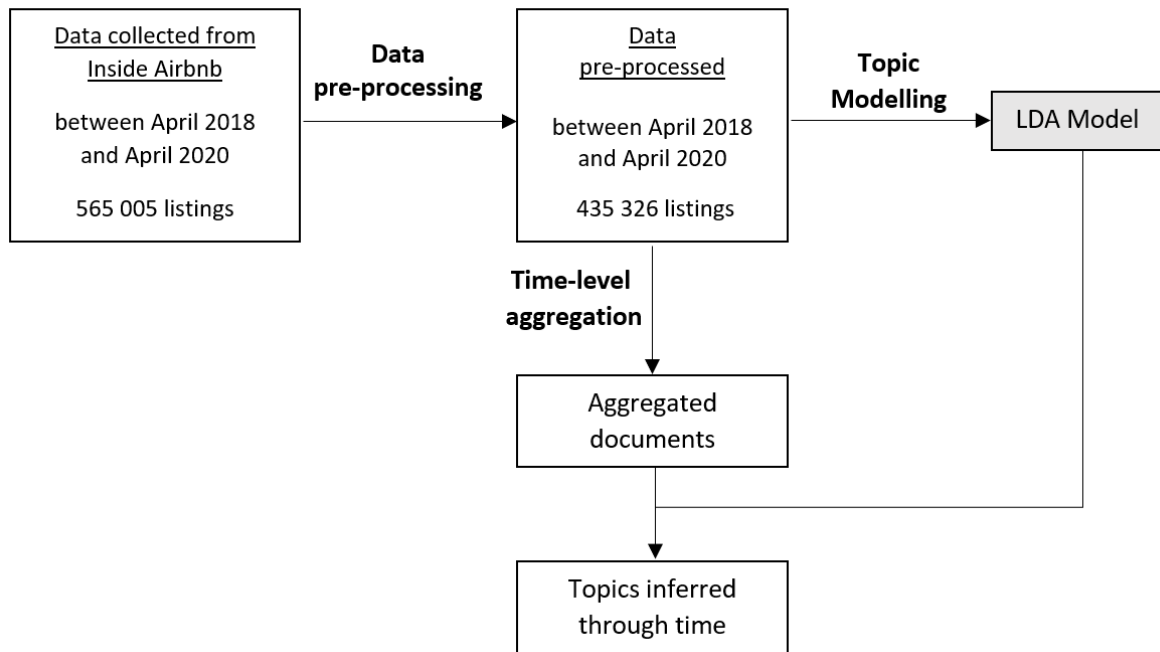
Figures

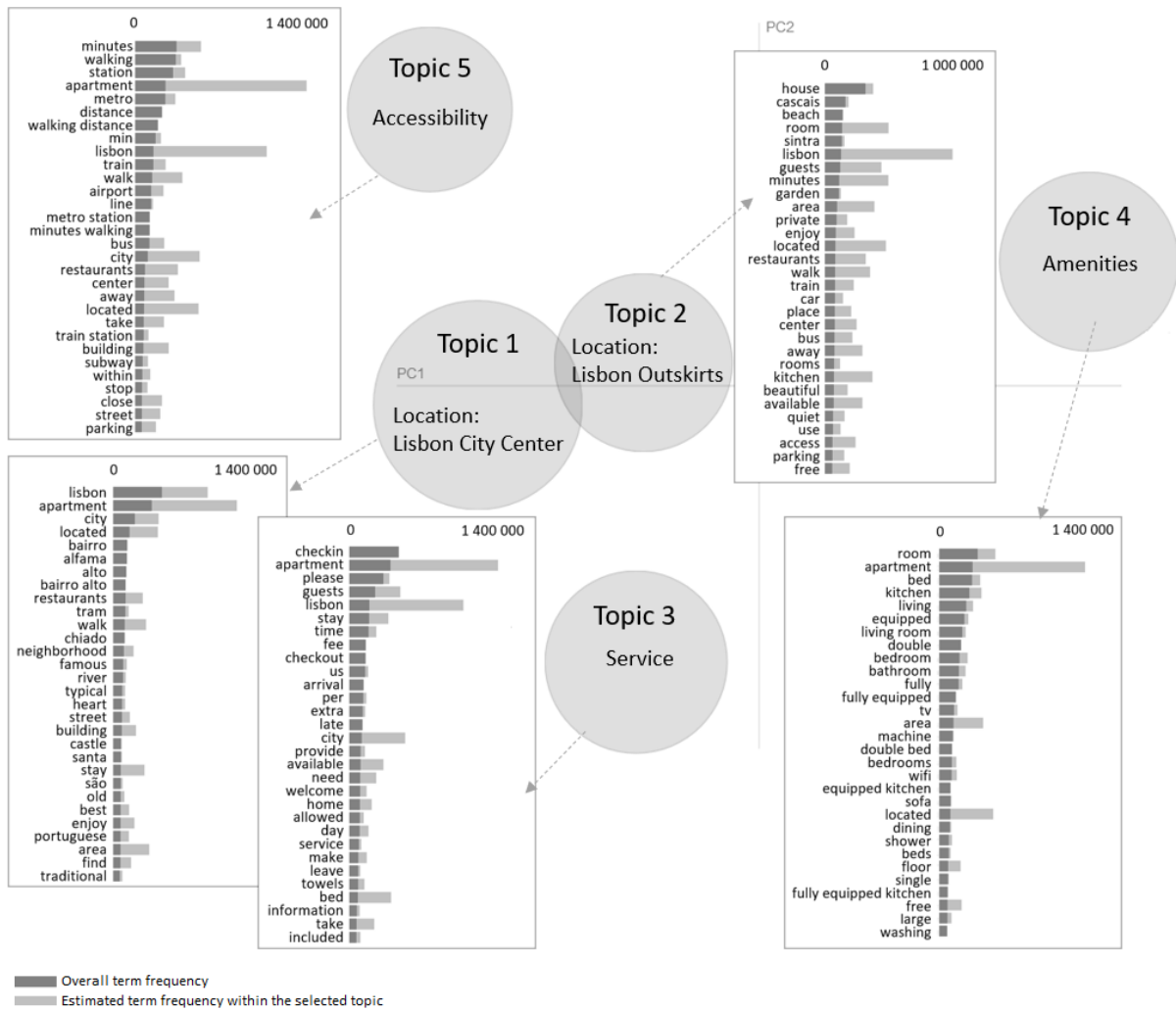Figure 1- Illustration of the Proposed Method.

Figure 2- Representation of the topics as an intertopic distance map of two dimensions, and histograms with the 30 most relevant terms in each topic (adapted from the result produced by pyLDAvis tool).

Topic 1

LOVING ALFAMA APARTMENT IV it's a cozy and charming place, with the best location you can have on the historic city center – in the heart of Alfama – and near Lisbon's most beautifully locations such as Praça do Comércio, Graça, Castelo, Bairro Alto and Chiado. Alfama contains many important historical attractions, with many Fado bars and restaurants. The great 1755 Lisbon Earthquake did not destroy Alfama which has remained a picturesque labyrinth of narrow streets and small squares. Lately the neighborhood has been invigorated with the renovation of the old houses and new restaurants where Fado— Portuguese typical melancholy music—can be enjoyed. Alfama is a neighborhood with a strong local life.

Topic 2

The simplest country house that one can imagine!
For a vacation nature-oriented at a country village near Sintra and the beaches (Praia Grande).
Small country house with a large country yard, of around 2 thousand square meters.
Very private area

Topic 3

Jorge and Filipe want to show you just how great Lisbon is! Check-in is at 15:00, but can be scheduled at a different time according to both needs. One of us will meet you personally and happily help you plan out a memorable route. Checkout is until 11:00, but it can also be adjusted. We normally like to send you off in person, but you can also just leave the keys and close the door behind you. We hope to offer you all the comfort during your stay so that you can enjoy Lisbon's light as much as possible. Hello and welcome to our apartment. Here you"ll be able to experience authentic life in a historic neighborhood in the very heart of Lisbon

Topic 4

Very well decorated, high quality three bedroom, large room with sofas for five people and table for eight people, television in suite, living room and kitchen, CO2 detectors and flood detector, central air conditioning, washing machine, Washing machine, dishwasher, Wi-Fi, CD and DVD player, refrigerator and freezer drawers, espresso machine, large balcony with table, two elevators to access the apartment without any step, high security entrance door. It has double glazing for heat, cold and sound insulation. It has central air conditioning throughout the apartment. Close to two Supermarkets. Free Parking.

Topic 5

Stylish newly renovated apartment in the center of the neighborhood voted as the "best to live in Lisbon". Only 100 meters from the metro and bus stations. The quiet yet vibrant neighborhood offers various restaurants, cafes and local shops all in a walking distance. Very close to the airport - 5 minutes by car and only 10 minutes by metro. A 10-minute ride by metro to the city center but if you want to go further you can catch the train to Sintra for example, one metro stop away.
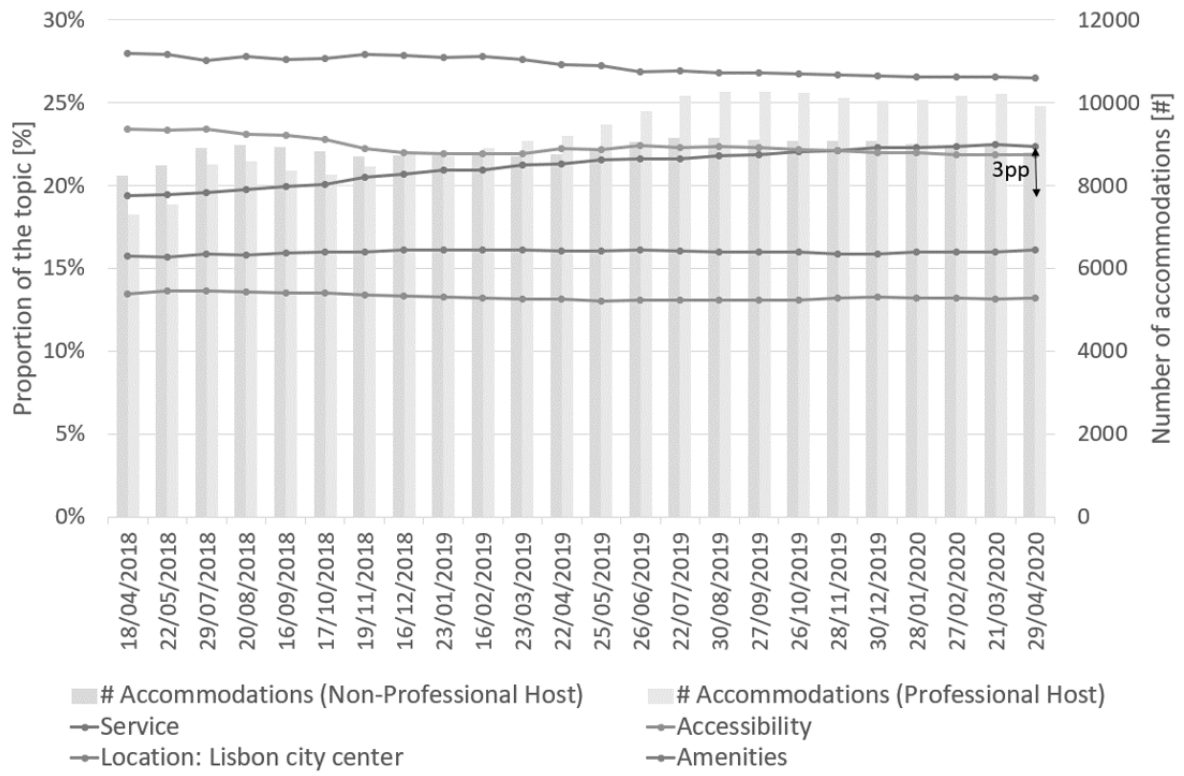
Figure 3- Excerpt of descriptions illustrating each topic

Figure 4- Topics monthly evolution.