Tourism & Management Studies, 19(2), 2023, 21-37 https://doi.org/10.18089/tms.2023.190202

Examining transaction-specific satisfaction and trust in Airbnb and hotels. An application of BERTopic and Zero-shot text classification

Examinando la satisfacción y confianza durante las estancias en Airbnb y hoteles: una aplicación de BERTopic y clasificación de texto Zero-shot

Manuel Rey-Moreno
Faculty of Tourism and Finance, University of Seville, Spain, mrmoreno@us.es

Manuel J. Sánchez-Franco
Faculty of Economics and Business Sciences, University of Seville, Spain, majesus@us.es

María de la Sierra Rey-Tienda
University Loyola of Andalusia, Spain, msreytienda@al.uloyola.es

Received: 24.07.2022; **Revisions required:** 14.10.2022; **Accepted:** 06.03.2023

Abstract

With a methodological approach, this article explores the application of data mining to the user-generated content of tourist accommodation on infomediation platforms and social networks. Its objective is to present an algorithm that allows the identification of service characteristics relevant to guest satisfaction and trust. Our study processes unstructured, natural-language data about Airbnb and hotel stays (the final dataset was 12,236 Airbnb sentences and 12,200 hotel sentences from 2018 until September 25 2021). Among the results is a computational algorithm that uses BERTopic to identify latent themes (or topics) in the narratives. Secondly, our analysis applies a Zero-shot classification approach for classifying guest reviews into labels related to guests' satisfaction and trust. Thirdly, we execute a Principal Component Analysis to investigate the sufficiency relationships between extracted topics, customer satisfaction, and trust-based labels. To sum up, and as practical implications, our study adds to the knowledge about the sharing economy by providing insights for developing marketing policies and a better understanding of hospitality services.

Keywords: Airbnb, hotels, satisfaction, Trust, BERT, Zero-Shot.

Resumen

El artículo analiza, desde una aproximación metodológica, la aplicación de la minería de datos al contenido generado por los usuarios en plataformas de infomediación y redes sociales de servicios de alojamiento turístico. El objetivo del paper es presentar un algoritmo que permita identificar los atributos más influyentes de este servicio en la satisfacción y confianza del huésped. Nuestro estudio procesa datos presentados en un lenguaje natural y desestructurado relativos a las estancias en hoteles y alojamientos Airbnb (la base de datos final fue de 12236 opiniones sobre servicios Airbnb y 12200 sobre hoteles, recogidas desde comienzos de 2018 hasta 25.09.2021). Entre los resultados obtenidos se encuentra un algoritmo computacional que utiliza BERTopic para identificar temas latentes en las narrativas. En segundo lugar, nuestro análisis aplica Zero-shot para clasificar las revisiones de los invitados en etiquetas relacionadas con su satisfacción y confianza. En tercer lugar, ejecutamos un Análisis de Componentes Principales para investigar las relaciones de suficiencia entre los tópicos extraídos y las etiquetas relacionadas con la satisfacción y confianza del cliente. Se añade al conocimiento sobre economía compartida nuevas perspectivas para el desarrollo de políticas de marketing y una mejor comprensión de los servicios de alojamiento.

Palabras clave: Airbnb, hoteles, satisfacción, confianza, BERT, Zero-shot.

1. Introduction

Tourism in Andalusia maintains significant economic relevance due to its high impact on both production and employment in this autonomous community. According to data from the latest Balance of the Tourism Year in Andalusia, published by the Ministry of Tourism and Sport of the Regional Government of Andalusia (2021), the income generated from tourism in this autonomous community reached €11 billion, accounting for 6.5% of its GDP. In 2021, the tourism sector employed an average of over 350,000 people, representing 11% of the total employment in Andalusia.

The Andalusian accommodation sector consists of approximately 100,000 establishments with nearly 1 million bed spaces. Among these, about 40% are dedicated to tourist apartments, while another 40% comprise hotels, guesthouses, hostels, and tourist flats. The remaining 20% consists of campsites, rural houses, and inns. As the capital of Andalusia, Seville holds a prominent position in terms of both supply and demand indicators in the tourism industry. It ranks as Spain's third most popular urban destination and the leading destination within Andalusia.

Tourist destinations in the era of the sharing economy adopt a model focused on Information and Communication Technology (ICT) and rooted in co-creation through value networks (Allee, 2003; Prahalad & Ramaswamy, 2004; Alqayed, Foroudi, Kooli, Foroudi, & Dennis, 2022). Specifically, peer-to-peer accommodation platforms (P2P accommodation) offer experiences that differ from traditional hotels. P2P accommodation not only provides social or physical interactions with local communities (Bresciani, Ferraris, Sanoro, Premazzi, Quaglia, Yahiaoui, & Viglia, 2021), but also fosters a sense of being at home at lower prices, targeting non-business guests (Zervas, Proserpio, & Byers, 2021). Consequently, P2P accommodation has emerged as a significant trend in the hospitality sector, prompting an examination of potential threats (Sainaghi & Baggio, 2020). However, there is no

conclusive evidence - only mixed findings - regarding guests' preferences and motivations when choosing sharing lodgings or evaluating hotels. As a result, the central question concerning Airbnb's competitive threat to hotels largely remains unanswered (Sainaghi & Baggio, 2020).

Numerous studies have investigated the perceptions and preferences of tourists staying in hotels in Andalusia and Seville for various purposes. Most of these studies draw their conclusions from surveys conducted using structured questionnaires. However, these techniques often introduce biases stemming from the formulation of the questions, potential response conditioning, or the utilisation of small sample sizes, among other factors. With the increasing use of ICT by tourists for consulting, booking, purchasing, and especially for leaving comments and evaluations about hotel services, it has become possible to explore these perceptions through alternative channels, such as Infomediation Platforms. Online reviews are essentially cognitive reconstructions of the customer's experiences at the hotel. Our study deviates from the conventional data collection approach via structured questionnaires and subsequent analysis using traditional statistical techniques. Instead, we employ data mining techniques to analyse the narratives generated and shared by hotel guests and Airbnb users, considering their genuine and up-to-date preferences. This approach is a foundation for designing policies that accommodation managers should develop.

Accordingly, our research aims to ascertain which conditions apply (equally or differently) to Airbnb and hotel accommodations based on user-generated content (UGC, used here interchangeably with electronic Word of Mouth-eWOM). Working with UGC lets us identify travellers' latent semantic structures in guests' usage motives associated with transaction-specific satisfaction and trust. UGC is also a more empathetic and reliable communication than unidimensional metrics to understand guests' needs and a global evaluation of relationship fulfilment by hospitality services. In this regard, BERTopic facilitates the organisation of free-form text (reviews) and the identification of the review's essential topics and associated sentences. In addition, our research applies a Zero-shot classification approach for classifying guest reviews into labels related to satisfaction and trust. A Principal Component Analysis (PCA) allows us to examine the relationship between topics and satisfaction/trust in terms of sufficiency.

Our work aims to address a specific objective related to filling a common methodological gap in the literature concerning tourist accommodation services. The novelty and originality of our study lie in the proposal of an algorithm that organises, classifies, and automatically evaluates natural language reviews provided by guests of tourist accommodations on infomediation platforms such as Airbnb and TripAdvisor. This algorithm enables the identification of behavioural patterns among users of each accommodation type studied, thereby facilitating the detection of potential differences between them.

After this introduction, the paper comprehensively reviews the relevant literature on the subject. Towards the end of this section, our paper formulates two research questions and outlines the methodological approach that will be employed to address them. The Research Method section provides specific details regarding data collection, data cleaning, and data pre-processing procedures. The subsequent section presents the research findings, categorised according to the data mining technique applied for each case, and distinguishes between their application to Airbnb accommodations and hotel stays. A Discussion section follows, where the main findings are analysed, and the research's theoretical and practical implications are identified. Finally, the article discusses the study's limitations and proposes potential future research directions.

2. Literature review

The number of tourists visiting urban areas has experienced exponential growth over the past decade, leading to significant transformations in both the hospitality sector and cities. Traditionally, the hospitality market comprised hotels and travel agencies that provided leisure travel. However, a new competitive landscape emerged in recent years, driven partly by the impact of ICT, enhanced connectivity, and the emergence of co-creation value networks. As a result, the internet revolution and the widespread adoption of the sharing economy have popularised short-term shared accommodation platforms (Deloitte, 2019).

New strategies and policies in the hospitality sector are essential for striking a balance between tourist activity and sustainable development, particularly in areas experiencing signs of over-tourism. According to Hall and Pennington (2016), the sharing economy is centred around online platforms that facilitate the sharing of underutilised assets or services between peers, either for free or for a fee. Peer-to-peer (P2P) accommodation platforms offer a combination of competitive pricing, commercial value, and social experiences (Sánchez-Franco & Rey-Moreno, 2021). The growth of P2P platforms in the vacation accommodation sector has been remarkable worldwide over the past five years. While the sector's revenues grew by over 10% between 2017 and 2018, it is estimated that the sector will continue to expand by up to 30% by the end of 2023 (Santos et al., 2021).

Airbnb is currently the world's largest P2P accommodation platform. Founded in 2008, it was created to provide hosts with spare rooms an opportunity to share their space with potential guests, and the disintermediation of traditional commercial channels has facilitated its growth. Today, Airbnb's database includes a vast collection of rooms, apartments, and homes in approximately 200 cities worldwide (Dogru et al., 2020). However, the increased supply of accommodations facilitated by Airbnb has caused concern among hotel managers in the traditional hospitality industry (Haywood et al., 2017).

Airbnb has emerged as a significant competitor to online travel agencies (OTAs), causing disruptions in the hospitality industry (Dogru, Mody, & Suess, 2019). It allows individuals to trade their underutilised properties, such as shared rooms, private rooms, or entire apartments, at competitive prices, emphasising the human connection as the primary shared asset (Dolnicar, 2018, 2020; Zach, Nicolau & Sharma, 2020; Quoquab & Mohammad, 2022). Additionally, it facilitates short-term rentals between hosts and guests as an alternative to traditional hotel stays. However, there is some controversy in the literature regarding the classification of Airbnb as a sharing economy service. Solano-Sánchez et al. (2021), for instance, argue that the original idea of individuals reserving rooms from private owners with unused spaces in their accommodations has diminished due to the rise of large companies engaged in purely commercial short-term rental activities. Therefore, according to their perspective, only a minority of the accommodations listed on Airbnb could be classified as true sharing economy services. Despite this debate, academics and managers are increasingly interested in peer-to-peer (P2P) accommodation (Dolnicar, 2020; Sánchez-Franco & Rey-Moreno, 2021). However, the motives that drive travellers to engage with short-term rental services remain unclear (Lalicic & Weismayer, 2018; Sainaghi & Baggio, 2020). On the one hand, Airbnb maintains a high level of transparency by sharing information with its hosting partners (Foroudi & Marvi, 2021), and guests perceive Airbnb as a cost-effective option (Guttentag et al., 2018; Liang, 2015). On the other hand, Airbnb creates significant competition and compels hosts and their reputational capital (Ikkala & Lampinen, 2014) to set high standards and adjust prices accordingly (Lalicic & Weismayer, 2018).

Consequently, Airbnb has become a popular, cosy, and authentic opportunity. For example, Li, Hudson, and So (2019) summarise various studies on customer experience in hospitality services and confirm that the Airbnb customer experience encompasses the advantages of staying in a home, personalised service, social interactions, and authenticity. Unique local experiences (Tussyadiah & Pesonen, 2016), social encounters (Cheng, 2016), hospitality, and conversations with hosts (Tussyadiah & Zach, 2017; Belarmino, Whalen, Koh, & Bowen, 2017) are indeed the most valued attributes sought by guests in P2P accommodations.

Conversely, hotels are defined by the institutional contexts in which they operate. Unlike social distancing, hotels offer opportunities for contact with staff and facilitate indirect communication among guests (Osman, D'Acunto, & Johns, 2019). In addition to guest characteristics such as age, nationality, gender, or purpose of visit, guests prioritise factors such as location (proximity to major attractions, among others), room comfort and cleanliness, price, value for money, service quality (including staff friendliness and helpfulness), availability of amenities like parking or gym facilities, and security. From the perspective of hotels, Sthapit and Jiménez-Barreto (2018) emphasise the importance of room amenities and a location close to the destination's tourist attractions. UGC is widely recognised in social science as a reliable source for assessing the opinions and preferences of individuals in relation to a given topic. This content encompasses comments, messages, photographs, and videos found on websites, social networks, infomediation platforms, and various digital environments. In a recent study, Saura, Palacios-Marqués, and Ribeiro-Soriano (2023) present a summary table that compiles some of these studies. UGC holds particular significance in the tourism sector in general, and specifically in the accommodation sector, as it provides valuable insights into guests' experiences without any interference from researchers (Sánchez-Franco, Navarro-García & Rondán-Cataluña, 2016).

In this regard, UGC enables us to identify topics associated with guests' travel experiences, their motivations for using sharing environments like Airbnb, and the factors influencing transaction-specific satisfaction and trust compared to hotels. The marketing domain has extensively researched consumers' preferences in the decision-making process for accommodation services, examining the most influential attributes and features and the level of guest satisfaction or dissatisfaction (Ju et al., 2019; Mody, Suess, & Lehto, 2019). However, when it comes to Airbnb compared to hotels, the research has produced contradictory findings, suggesting either the emergence of new markets and consumption patterns or a clear substitution effect (where Airbnb attracts travellers with pricing options, a diverse range of spaces, and features associated with the location and local authenticity). While Airbnb and hotels offer distinct services and amenities, there is limited research focusing on differentiating preferences that impact the quality of guest relationships (satisfaction and trust), and, as a result, the conclusions remain contentious.

In a relational context, when comparing the outcomes of hospitality services to guests' expectations, Lovelock and Wirtz (2007) define satisfaction as the individual's feeling of pleasure or disappointment resulting from their stay. From a cognitive perspective, satisfaction is conceptualised as the affective response to the match or mismatch between the outcome and the standard of comparison (Oliver's disconfirmation of expectations model, 1997, 2010). Additionally, atmospheric theory (Baker, Levy, & Evans, 1992) identifies both tangible factors (such as accommodation amenities) and intangible factors (such as the ambience of the accommodation) that influence guests' pleasure, arousal, and willingness to return. Guest expectations, evaluations based on specific features, emotional evaluations using the theory of emotions, and sensory attributes all play a crucial role in generating satisfaction (Bagozzi, Gopinath, & Nyer, 1999; Baker, Levy, & Evans, 1992; Bigné, Andreu, & Gnoth, 2005; Lazos & Steenkamp, 2005; Mudie, Cottam, & Raeside, 2003; Oliver, 2010; Rodríguez & San Martín, 2008; Yu & Dean, 2001). Therefore, our research on UGC helps hosts and hotel managers identify guests' motives and provide enhanced services to

improve the guest experience, enhance the reputation of hospitality services, foster willingness to return, and even influence guests' willingness to accept a higher price.

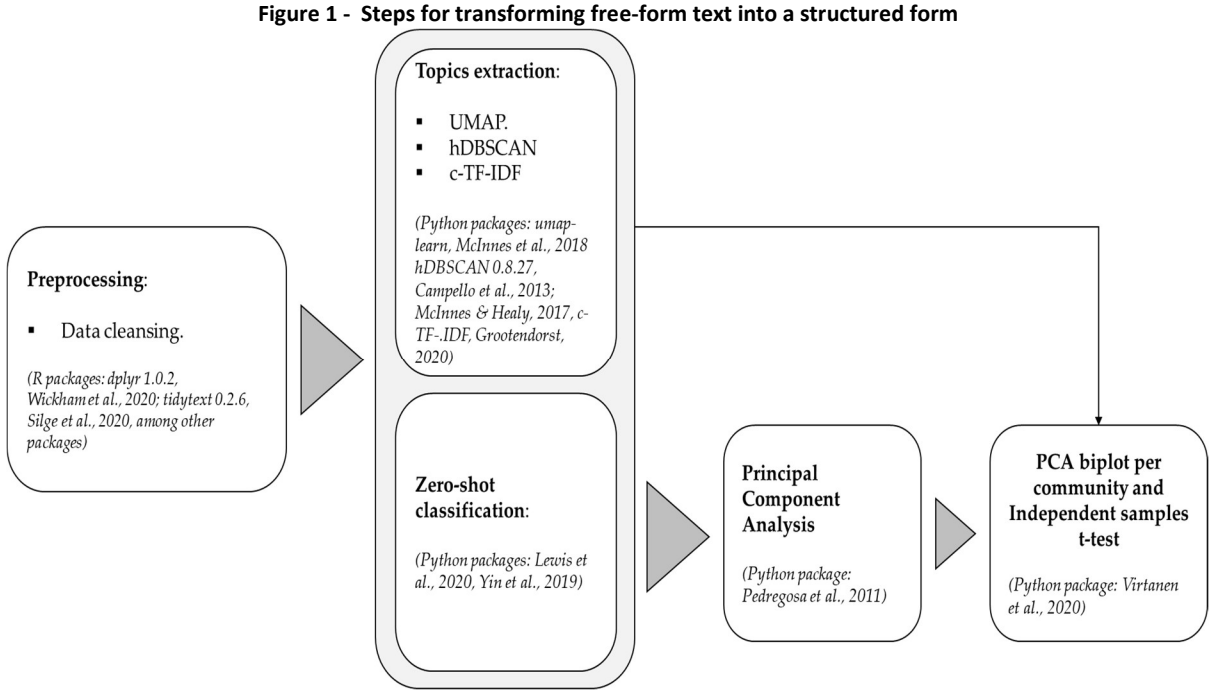
Moreover, enhancing the value-satisfaction-intention framework by incorporating additional critical factors, such as trust-related dimensions, allows for capturing the effects of the attribution process (Ye, Chen, & Paek, 2021). Trust is defined as "a subjective feeling that the trustee, whether it's a host or a hotel manager, will act in a certain way based on an implicit or explicit promise" (Ert, Fleischer, & Magen, 2016, p. 64). While "the most common reputation mechanism involves the presentation of online reviews of the seller by experienced users" (Ert, Fleischer, & Magen, 2016, p. 63), research based on user-generated content (UGC) focusing on Airbnb and hotels is limited to examining the overall effects of transaction-specific satisfaction and trust on behavioural intentions. Previous research indicates the lack of generalizability when comparing different features that guests consider when selecting between sharing models and more traditional options.

The aim of our paper is twofold. Firstly, we aim to present an automated identification and classification algorithm to manage large unstructured documentary archives effectively. This algorithm utilises the automatic generation of latent topics from user reviews and narratives, solving the challenge of handling such archives. Secondly, we aim to explore the significant features that impact guests' satisfaction and trust based on user-generated content (UGC) and social media, specifically online guest reviews.

Figure 1 provides an overview of the steps involved in transforming free-form text into a structured format and outlines the main approaches to address the following two research questions:

- What combination of data mining techniques is suitable for determining whether all the characteristics associated with hospitality services are sufficient to generate an outcome in the realm of hospitality services?
- Are there significant differences between the types of accommodations, namely Airbnb and hotels?

Figure 1 - Steps for transforming free-form text into a structured form



3. Method

Our research has meticulously curated a range of analytical tools, utilising a large dataset and employing techniques such as natural language processing (NLP), zero-shot classification, and principal component analysis (PCA). These approaches were carefully chosen due to their distinct advantages over alternative techniques and their alignment with the research problem under consideration.

Firstly, conducting big data analysis of reviews from information mediation platforms such as Booking.com offers several advantages over structured questionnaires. This approach provides access to extensive and diverse datasets comprising natural narratives, which offer a comprehensive view of customer feedback across various parameters, including service quality, location, cleanliness, and amenities. Additionally, since reviews are voluntarily posted, they provide a more authentic reflection of the customer experience, allowing for a broader and more genuine range of opinions. Moreover, big data analysis can uncover patterns and relationships that may not be readily apparent through traditional survey questions. Collecting data from information mediation platforms is also relatively cost and time-efficient, as the data is readily available, eliminating the need to conduct surveys, design questionnaires, and administer them to a sample of guests.

On the one hand, big data analysis does have certain limitations related to data generation, extract-transform-load (ETL) processes, and analysis and visualisation techniques (Biemer, 2014). Additionally, while big data analysis reveals drawbacks such as data quality issues, privacy concerns, potential biases, lack of control over variables, technical challenges, and difficulties in interpretation, it remains a valuable tool due to its accessibility, enabling more in-depth analysis and interpretation. Furthermore, it provides a detailed view of the customer experience, and the large sample sizes allow for greater statistical power and more robust analysis.

On the other hand, survey and market researchers are renowned for their in-depth understanding of research questions, particularly regarding qualitative and contextual aspects of social reality. They also have greater control over the variables they measure and can include multiple complementary indicators (Callegaro & Yang, 2018). However, traditional questionnaire-based research is a labour-intensive process that is infrequently updated and susceptible to response biases resulting from the wording of questions, which can distort information and create a disconnect from reality (Dolnicar, 2018; Zervas, Proserpio, & Byers, 2021).

Secondly, NLP is a powerful technique for extracting insights from textual data (here, guests' narratives). NLP can perform sentiment analysis, topic modelling, summarisation, and information extraction tasks. Furthermore, NLP handles unstructured or semi-structured data that traditional methods may not efficiently process. In particular, NLP has several advantages in automating tasks, improving decision-making, and analysing unstructured data. And it is applied to analyse large volumes of unstructured text data to extract insights and trends that can inform hospitality strategies.

Moreover, BERT, a pre-trained deep learning model, has several advantages over traditional NLP models. Its fine-tuning for various NLP tasks and its bidirectional nature improve its understanding of sentence context. (Liu et al., 2020; Petroni et al., 2019). BERT considers the context of both preceding and succeeding words, improving its ability to understand the meaning of a guest's narrative. And in particular, BERTopic is applied here as an unsupervised topic modelling technique supported by the transformer architecture as BERT. BERTopic generates sentence embeddings, which are clustered to identify related topics in a text corpus. BERTopic can be used to explore large volumes of unstructured text data and identify meaningful topics.

Nevertheless, NLP can encounter limitations regarding accuracy and performance when dealing with data that exhibit crucial characteristics. These characteristics include data in different languages, texts containing specific jargon or slang, texts with grammatical or spelling errors, texts with intricate metaphors or ironies, and texts lacking clear semantic information. Therefore, evaluating the suitability of our approach for the specific data types involved is vital.

Thirdly, zero-shot classification, a subfield of NLP, predicts a class that the model does not see during training. This method overcomes the limitations of supervised learning, which requires labelled data for each class. In our case, the zero-shot classification adapts to new classes of entities (related here to relationship quality) without retraining or fine-tuning the model (and saving time and resources). The zero-shot classification relies on a pre-trained language model that encodes semantic information from a text (here, guests' narratives) and compares it with class labels (Hugging Face, 2023; Xiang, Lampert, Schiele, & Akata, 2020; Yang, Ye, Zhang, & Huang, 2022). Likewise, the zero-shot classification highlights the ability to detect complex semantic relationships. However, it may have limitations regarding accuracy and performance when dealing with specific data types. For instance, it may have difficulty classifying classes with complex and subtle semantic representations, such as metaphors and ironies.

Fourthly, PCA is a dimensionality reduction technique that transforms a set of variables into a smaller set of uncorrelated components that capture most of the variance in the data and identify the most significant features. As a result, PCA can help reduce noise, improve visualisation, and enhance downstream models' performance. In particular, by reducing the number of input features using PCA, the downstream models can become less complex and more efficient, leading to better performance in terms of accuracy and speed.

PCA can also reveal hidden patterns and relationships among variables that may not be apparent otherwise (Jolliffe & Cadima, 2016). Compared to other dimensionality reduction techniques, such as Factor Analysis or Multidimensional Scaling, PCA (or adaptations of PCA) are relatively simple to implement and computationally efficient. PCA offers simplicity, flexibility, and scalability advantages, making it a more valuable tool for data reduction and visualisation. However, it is relevant to consider the limitations of PCA, such as its assumptions of linearity and sensitivity to outliers or its difficulty in interpreting the reduced-dimensional space.

3.1 Data collection

The infomediation platforms with which tourist accommodation guests usually interact have led to an evolution in the sector towards real-time UGC-based communication models. Online opinions or narratives tend to be more free, spontaneous, enlightening of guest experiences, and highly accessible from anywhere and at any time (Guo, Barnes & Jia, 2016).

Barbosa, Saura, and Bennett (2022) show in one of their studies the relevance of using Big Data Analytics techniques in the study of customer behaviour. Zhu, Lin, and Cheng (2020) point out that the user's own narratives are multifaceted and more reliable than other metrics. They refer to the capture of primary information based on questionnaires, which in addition to requiring a significant effort for data acquisition, usually accumulate various response biases, including those derived

from the wording of their questions, the excessive inconvenience generated by the respondent, as well as those of influence, retaliation or difficulty of generalisation due to the use of excessively small samples (Zervas, Proserpio & Byers, 2021; Dolnicar, 2018; Sun, Ma & Chan, 2018).

Given the described weaknesses of traditional methods of data capture and processing, our study proposes the alternative use of narratives generated by the hosts themselves, as well as data mining and machine learning techniques capable of extracting the essential elements of the original text (themes), thereby generating a simplified and understandable version of this in relation to the host profile.

This study analyses the Airbnb data obtained from the Inside Airbnb website http://insideairbnb.com/ (available on September 25 2021). In addition, for strictly academic purposes, our study extracts (from Tripadvisor) the reviews of hotels in Seville (Spain), with more than 1,000 reviews per hotel -available on September 25 2021. The reviews are publicly accessible. The final dataset was 12,236 Airbnb sentences and 12,200 hotel sentences (over 80 characters in each) from 2018 until September 25 2021.

Our study analyses only English-language reviews to maintain consistency between the texts analysed. Furthermore, our research preserves the amateur character of the host and selects only hosts with a single listing (here, entire apartments). Entire apartments represent 85 % of the Airbnb offer in Seville (AirDNA, 2022). In addition, assuming that pricing is a critical factor determining the sustainable success of the accommodation industry, "listings of entire apartments are strong substitutes for hotel rooms in all price segments" (Gyódi, 2017, p. 543). In addition, Airbnb listings are considered outliers when the price and the number of beds lie outside the interval formed by the 5 and 95 percentiles. In this regard, our study removes all Airbnb listings with over five beds. Finally, our research filters out listings with a price lower than 10 US dollars (not including cleaning fees or additional guests) and higher than 161 US dollars.

3.2. Data cleansing process

After data capture, it is necessary to transform the natural language in which these guest narratives are expressed into structured language. For this, the first step required is pre-processing and data cleaning. In our case, the process followed includes: (1) checking the spelling of sentences and removing duplicates, (2) discarding punctuation, digits and extra whitespaces, (3) removing a shortlist of common stop words to filter out overly common terms and a customised list of proper nouns, (4) fixings contractions, and compound terms, and (5) becoming text in ASCII and standardising it by lowercasing. Likewise, our study replaces terms related to the generic term accommodation (Airbnb, apartment, flat, lodging, condominium, property, and hotel, among others) with the [mask] token and the terms host and owner with the [staff] token.

4. Results

4.1 Data Mining

More and more work is being done on digital marketing and applying artificial techniques to improve the commercial policies that companies must develop (Saura, Ribeiro-Soriano, Palacios-Marqués, 2021). Our study displays how term usage differs between Airbnb and hotel guests (see Kessler, 2017) -using the Scattertext 0.1.6 package in Python.

Each point corresponds to the usage of a term, where the higher up the term is on the y-axis, the more it is used by Airbnb guests, and the further right, the more it is used by hotel guests. The top right of the plot is an area where the term frequency is high for both types. Points coloured blue (at the top of the figure) are associated with Airbnb, and points coloured red (at the bottom of the figure) are associated with hotels.

Our analysis displays terms frequently occurring in all sets of sentences but which are relatively infrequent compared to general term frequencies -using the Scattertext 0.1.6 package in Python. Corpus characteristicness is the difference in dense term ranks between the terms in all the sentences in our study and a general English-language frequency list.

Accordingly, the most frequent terms in both cases are related to staff, hospitality stay and walkability or night. Moreover, the most frequently cited by Airbnb guests are the kitchen and its equipment (for instance, a washing machine), feeling at home and helpful staff (communications, recommendations and tips on attractions and neighbourhood, among others). Likewise, the most frequent terms hotel guests cite are related to property services, breakfast buffet, hotel reception desk, and quality signalling drivers (star).

While Figures 2 and 3 allow us to identify discriminative terms for documents in the corpus, this approach reveals little on inter or intra-document statistical structure. Therefore, extracting and selecting distinct topics and their semantic communities are essential.

Figure 2 - Visualising differences based on term frequencies neighbourhood

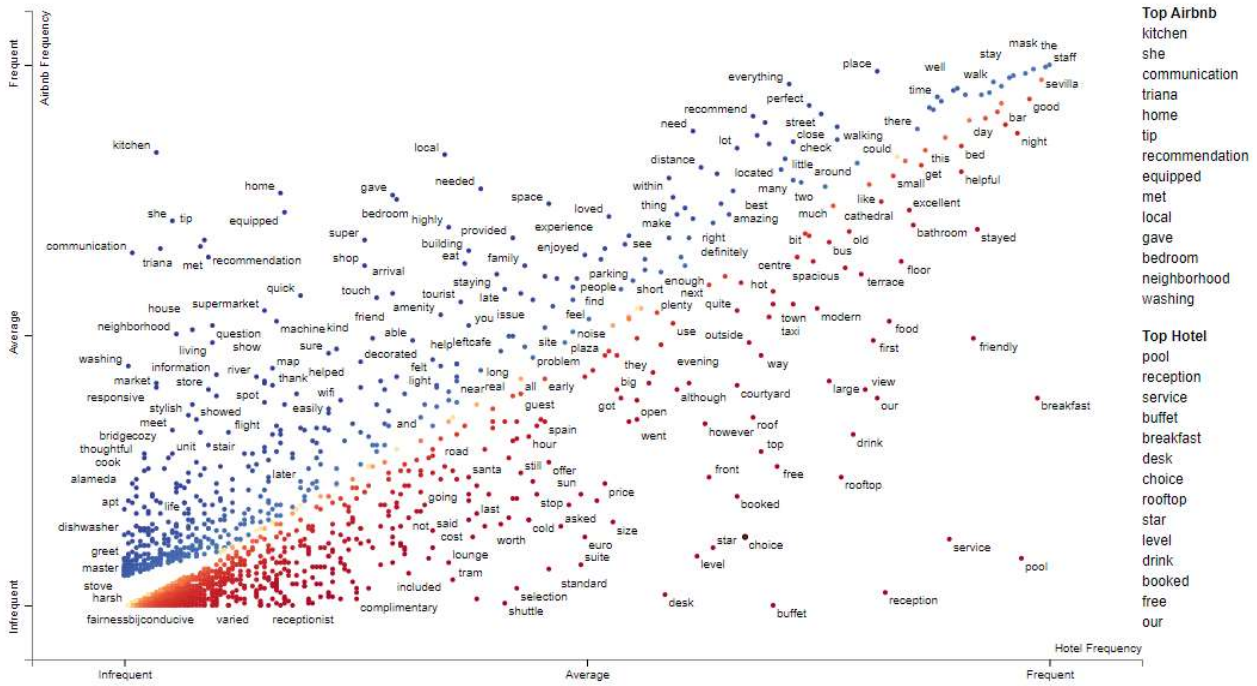
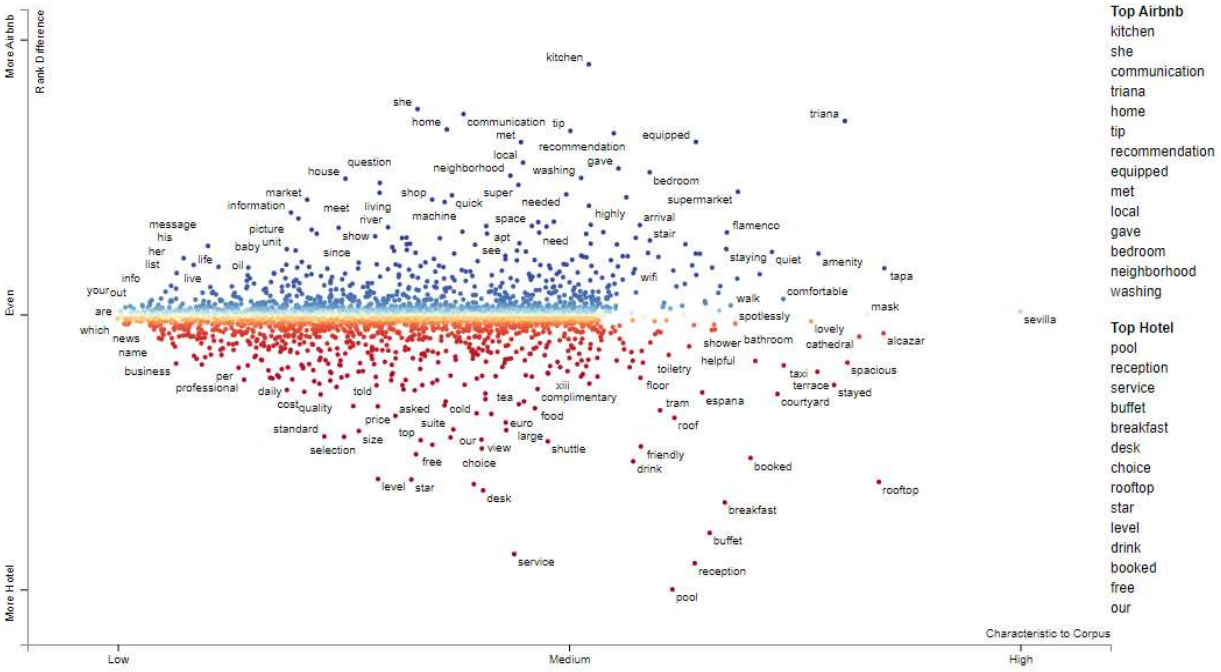


Figure 3 - Visualising terms of corpus characteristicness



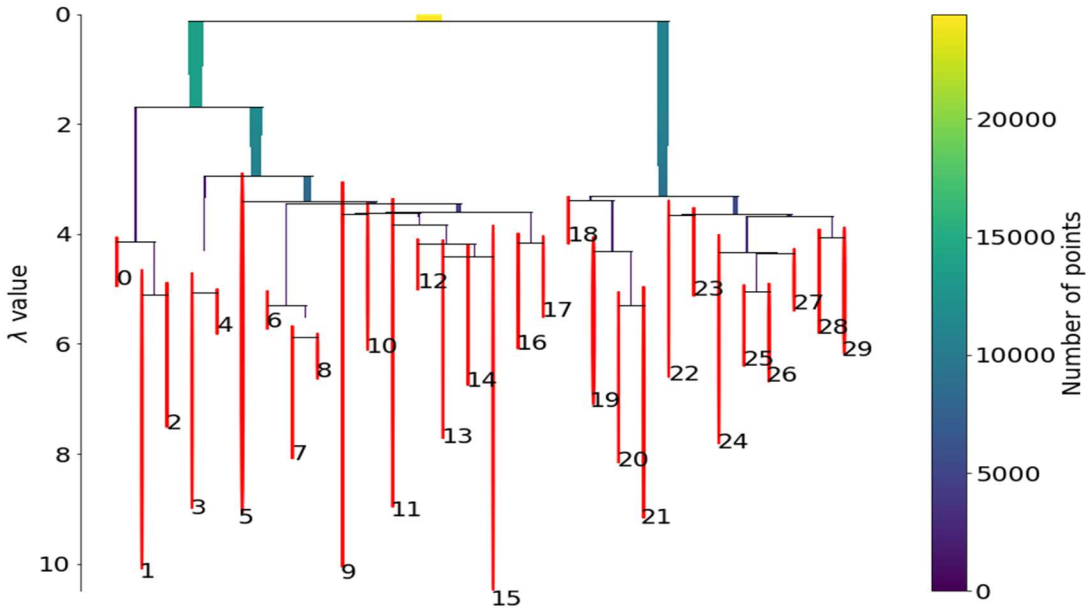
BERTopic's approach is a topic modelling technique that leverages transformers to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions (Grootendorst, 2020). It is based on Top2Vec (Angelov, 2020).

Our research firstly transforms our corpus into n-dimensional dense vector space by applying a SentenceTransformer model (here, all-MiniLM-L6-v2, the default model in BERTopic). Then, a Uniform Manifold Approximation and Projection for Dimension Reduction (from now on, UMAP) (McInnes, Healy, & Melville, 2018) is applied to our embeddings to create a lower-dimensional space of document vectors through the umap-learn 0.5.1 package in Python 3.8 (McInnes et al., 2018). Our proposal reduces the vectors to 5 dimensions (from now on, 5d-UMAP) and measures the distances between data points by cosine similarity. Experimentation and related literature here recommend the 15-nearest neighbours to emphasise the local structure, and the effective minimum distance between embedded points is set at 0.01.

Furthermore, to find such dense documents areas, the 5d-UMAP embedding is clustered with the Hierarchical Density-Based Spatial Clustering of Applications with Noise algorithm (from now on, hDBSCAN) (Campello, Moulavi, & Sander, 2013; McInnes & Healy, 2017). hDBSCAN extends DBSCAN and extracts stable clusters of varying densities (arbitrary shapes, sizes, and noisy points). The minimum size of clusters is set at 200. Likewise, the number of samples or density threshold is set at 25 (the minimum number of samples required before an area can be considered dense and a point be considered a core point).

Our analysis employs the hDBSCAN 0.8.27 package in Python 3.8. Figure 4 condenses the cluster tree. After obtaining the clusters, the following steps finally identify one topic vector per cluster. Our study employs a class-based variation TF-IDF approach. TF-IDF is a measure for representing the importance of a term to a sentence (or document) and combines the term frequency and the inverse document frequency. In this sense, c-TF-IDF allows us to compare the importance score of a term to a cluster (from now on, c-TF-IDF, with c being the identified cluster). The higher the c-TF-IDF score, the more representative it is of its topic.

Figure 4 - Visualising the condensed tree: hDBSCAN



Our study also creates a 2d-UMAP approximation where the continuous representation of topics can be easily visualised (see Figure 5).

Figure 5 - Visualising topics using the 2d-UMAP approximation

Next, network analysis by filtering the correlation matrix (> 0.20) provides an accurate set of methods and tools to produce structures (and substructures), showing a deeper understanding of the system. Our study employs the c-TF-IDF matrix for estimating the correlations between clusters.

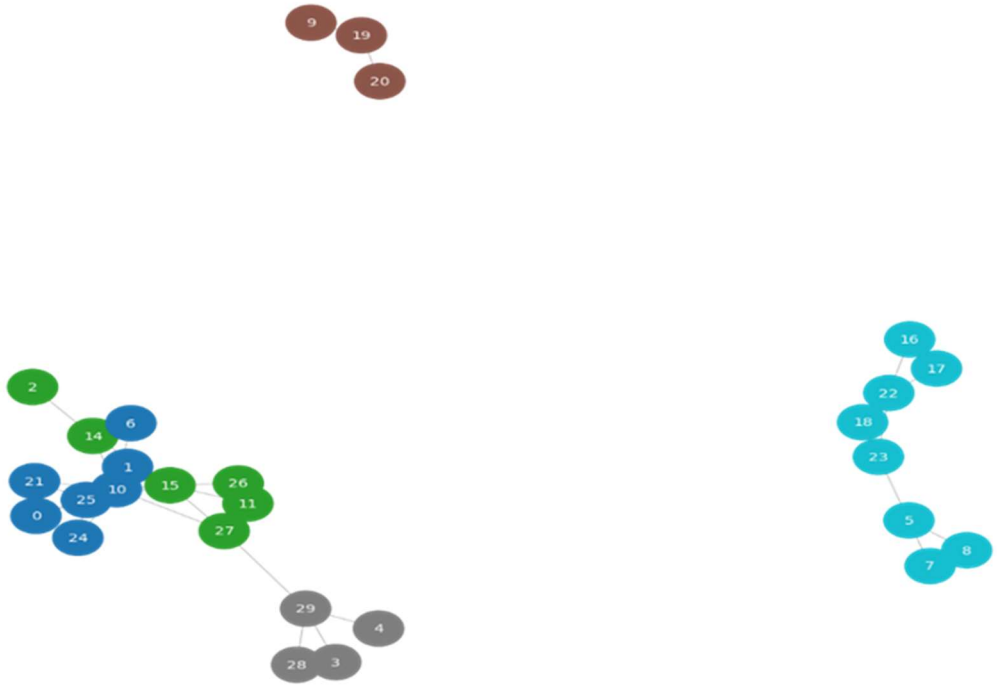
Community detection -based here on the Louvain algorithm- is used to examine the underlying semantic structure.

- The first community (green) identifies guests' satisfaction with their stay and behavioural intentions (in particular, recommendation-based topics: 2, 14, 11, 15, 26 and 27).
- The second community (dark blue) contains specific topics about walkability (topics: 0, 10 and 25), the transportation system (topic: 21) and neighbourhood amenities (topic: 24). Convenience and access are site-specific characteristics related to the location (topic 1), easy access to the hotel or other hotspots that attract many visitors. In addition, distance to major attractions and pedestrian-friendly infrastructure encouraging people to walk quietly and peacefully between the accommodation and hotspots are site-specific factors.
- The third community (grey), topics 3, 4, 28 and 29, is for staff. This community is known for caring, which may impact receiving ideal home-like feelings. Relationship drivers include being responsive to inquiries, keeping guests informed, and simply making guests feel welcome in the hospitality domain.

  In addition, these topics are related to customer service performance (communication about directions, tips or advice, accommodation rules, Wi-Fi instructions, or check-in/out, among others), and how staff deliver the service that contributes to a positive or negative hospitality experience.
- The fourth is related to a comfy place to sleep -noise and air conditioning- (dark brown), topics 9, 19 and 20.
- The fifth community (light blue) is concerned with in-room service, cleanliness and comfort.
  - Core services or tangible benefits sought by guests (for example, amenities and bath services or in-room-based topics: 16, 17 and 20) are associated with accommodation quality, including the size and type of accommodation, the functional space, the decoration and cleanliness of core services, and complaints resulting from out-of-accommodation activities.
  - In-room-based amenities (topic 13) are associated with kitchen facilities.
  - It is also concerned with the sensation of being "in a clean home" or experiencing hedonic comfort while staying in a hospitality accommodation (topics: 18 and 23).
  - Topics 5, 7 and 8 are about property features (swimming pool or views, among others) related to services, such as a terrace overlooking the city that enhances customer delight.

The average clustering coefficient equals 0.49 (modularity ≥ 0.3). Overall, the extracted communities (or metatopics) are easily interpretable and offer a coherent impression (see Figure 6):

Figure 6 - Visualising the network of topics



In addition, Table 1 displays the most informative terms on each topic.

Table 1 - Terms distribution among BERTopic topics sorted according to TF-IDF values

0: cathedral	0.34899	**5**: pool	0.16650	**10**: mask	0.07098	**15**: seville	0.22853	**20**: conditioning	0.12131	**25**: walk	0.17372
alcazar	0.23068	Courtyard	0.09214	walk	0.06212	mask	0.07438	mask	0.09107	walking	0.15352
walk	0.15862	Terrace	0.08469	location	0.05754	stay	0.04766	floor	0.07639	distance	0.12823
minutes	0.10988	Balcony	0.06772	city	0.04828	visit	0.03518	hot	0.05950	location	0.11048
location	0.08654	Mask	0.06447	located	0.04661	place	0.03362	stairs	0.05376	attractions	0.09942
cruz	0.07851	Rooftop	0.05382	breakfast	0.04365	recommend	0.03294	heat	0.04602	minutes	0.08837
restaurants	0.07322	Area	0.05328	restaurants	0.03813	located	0.03145	temperature	0.04474	city	0.07967
1: seville	0.34553	**6**: breakfast	0.24505	**11**: stay	0.31099	**16**: shower	0.24954	**21**: bus	0.20783	**26**: stayed	0.19638
stay	0.06542	Food	0.08559	stayed	0.20018	bathroom	0.19458	taxi	0.17988	stay	0.16704
place	0.04768	Good	0.07005	nights	0.17849	water	0.10492	airport	0.12196	nights	0.14124
city	0.04627	Coffee	0.05612	definitely	0.13634	towels	0.07390	shuttle	0.11729	mask	0.13310
visit	0.04358	Choice	0.04322	return	0.08511	hot	0.07000	euros	0.10637	enjoyed	0.07030
location	0.04353	Restaurant	0.04132	recommend	0.07681	bath	0.05653	taxis	0.09128	spent	0.06048
great	0.03484	Fresh	0.03921	staying	0.07096	toilet	0.05156	euro	0.08316	staying	0.05878
2: sevilla	0.41418	**7**: pool	0.46426	**12**: experience	0.16243	**17**: bed	0.33888	**22**: bed	0.15403	**27**: mask	0.15936
stay	0.07770	Bar	0.09837	really	0.08256	comfortable	0.23722	bathroom	0.14442	recommend	0.13951
place	0.06507	Rooftop	0.09731	thank	0.07701	bedstead	0.23328	comfortable	0.10886	highly	0.07964
spain	0.04930	Area	0.09413	perfect	0.07000	pillows	0.09922	bedstead	0.08646	definitely	0.05519
location	0.04400	Swimming	0.08218	welcome	0.06700	bedroom	0.08537	shower	0.08104	stay	0.03628
great	0.03904	Sun	0.06818	felt	0.06254	comfy	0.07780	mask	0.07369	booked	0.03411
beautiful	0.03501	Roof	0.06750	things	0.05891	couch	0.07715	clean	0.07291	return	0.03383
3: staff	0.45751	**8**: terrace	0.17502	**13**: kitchen	0.26501	**18**: clean	0.24828	**23**: clean	0.20539	**28**: staff	0.37163
caring	0.27706	Roof	0.11394	machine	0.15102	modern	0.09066	mask	0.09748	mask	0.11756
friendly	0.27179	Rooftop	0.10896	washing	0.14405	place	0.08844	comfortable	0.08807	valuable	0.11083
staffs	0.08408	Bar	0.09010	equipped	0.09053	decorated	0.08542	spacious	0.07346	amazing	0.10758
polite	0.05801	View	0.07629	washer	0.09050	space	0.07628	modern	0.05427	helpful	0.10587
extremely	0.04945	Sight	0.07532	cooking	0.07920	comfortable	0.06277	decorated	0.04413	welcoming	0.05413
attentive	0.04827	Courtyard	0.07301	laundry	0.06967	decor	0.06144	amenities	0.04297	great	0.05086
4: communication	0.15334	**9**: quiet	0.14172	**14**: sevilla	0.35581	**19**: noise	0.13826	**24**: restaurants	0.19280	**29**: mask	0.08568
questions	0.12576	Noise	0.14042	mask	0.06915	quiet	0.13787	supermarket	0.17022	check	0.07967
tips	0.08654	Night	0.11792	stay	0.05828	noisy	0.09237	bars	0.14727	met	0.07697
gave	0.08641	Street	0.09762	located	0.03538	street	0.08374	nearby	0.12125	arrived	0.06708
quick	0.06820	Noisy	0.07373	center	0.03535	hear	0.07392	grocery	0.11263	gave	0.06045
provided	0.05977	Hear	0.06690	city	0.03416	mask	0.07259	shops	0.10036	helpful	0.05725
helpful	0.05798	Sleep	0.06059	great	0.03395	night	0.06359	cafes	0.08785	quick	0.04946

4.2 Zero-shot classification

Our research applies a Zero-shot classification approach for classifying the corpus into labels related to relationship quality dimensions. Zero-shot text classification tasks "make use of the good performance that transformers have demonstrated in text entailment tasks" (Alcofarado et al., 2022, p. 126).

Yin, Hay, and Roth (2019) propose a pre-trained Natural Language Inference (NLI) approach as a ready-made Zero-shot sequence classifier. It is highly effective on larger pre-trained models like BART (a denoising autoencoder for pretraining sequence-to-sequence models; here, facebook/bart-large-mnli) (Lewis, Liu, Goyal, Ghazvininejad, Mohamed, Ley, Stoyanov, & Zettlemoyer, 2019). Our analysis does not previously need any labelled data. It operates by posing the sequence (here, sentences) to be classified as the NLI premise (developing a hypothesis from each candidate label).

Therefore, the task is to determine how confident the model is in our candidate label being relevant to the text and obtain a probability interpretation of the final result. Our study specifies the multiclass argument as false (the sum of the probability score is 1). For instance, a sentence about hospitality stays ("Excellent location, clean room and friendly staff, very comfortable beds. Tea and coffee facilities in the room first time I have known this in a hotel in Seville") could be predicted as caring (0.264 % confidence), competent (0.392 % confidence), communicative (0.262 % confidence) and delighted (0.082 % confidence).

Table 2 shows the main labels (and their synonyms according to hospitality services) related to the dimensions associated with satisfaction and trust-based dimensions, and based on Bakker, Voordt, Vink, and De Boon (2014); Dickinger (2011); Flavián, Guinaliu, and Gurrea (2006); Gefen and Straub (2004); Lu, Zhao, and Wang (2010); Mayer, Davis, and Schoorman (1995); Mody and Hanks (2020); Pappas (2019); Pijls, Groen, Galetzka, and Pruyn (2017); Sirdeshmukh, Singh, and Sabol (2002); and Tan and Sutherland (2004), among others.

Table 2 - Satisfaction and trust dimensions and candidate labels

Satisfaction	**Affective trust**		**Cognitive trust (reliability)**
	Integrity (and problem-solving orientation)	Benevolence	
delighted	committed	Caring	competent
fun	communicative	complacent	effective
happy	empathetic	friendly	helpful
loving	motivated	kind	reliable
pleasurable	purposeful	polite	skillful
welcoming			valuable

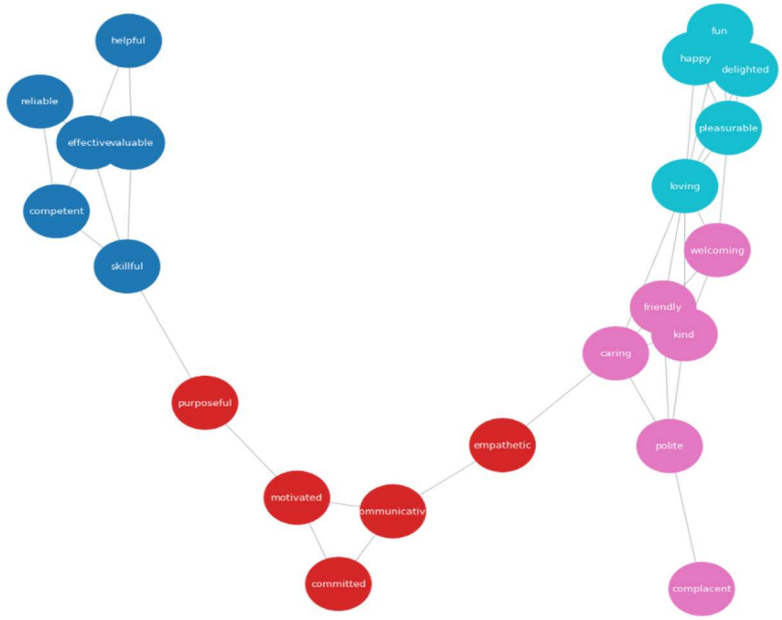
The content of each dimension is specified below:

- **Transaction-specific satisfaction**: Satisfaction is here conceptualised as a favourable response of guests about a discrete service encounter. It is empirically distinguished from overall satisfaction (Jones & Suh, 2000). Thus, it focuses on a service interaction that is rewarding, fulfilling and stimulating, for example, delighting, pleasurable, fun, happy or loving, among others.
- **Trust – integrity (and problem-solving orientation)**: Trust comprises three dimensions, reliability, integrity, and benevolence (Gefen & Straub, 2004; Geyskens, Steenkamp, Scheer, & Kumar, 1996; Tan & Sutherland, 2004, among others). Integrity has to do with the trustor's perception that the trustee adheres to principles that the trustor finds acceptable (see Flavian, Cuinaliu, & Gurrea, 2006; Mayer, Davis, & Schoorman, 1995). For instance, Airbnb or hotels act honestly when fulfilling their promises (credibility); hosts or staff are motivated to help guests if they need it (fostering communication between them). Thus, integrity is a key driver in building affective trust; guests expect Airbnb and hotels to adhere to their commitments and not act unfairly.

  Therefore, behaviours that are purposeful or committed - arising during service delivery- are critical incidents that provide insight into the character of the provider (Sirdeshmukh, Singh, & Sabol, 2002).
- **Trust – benevolence**: Benevolence is a fundamental component of hospitality relationships. It encompasses caring, complacent, friendly or kind behaviour (hosts care about guests). In this regard, it is defined as the extent to which a trustee (Airbnb or hotel) is believed to want to do good to the trustor (guests) with warmheartedness (willing to help the customer; Dickinger, 2011), aside from an egocentric profit motive (Flavian, Guinaliu, & Gurrea, 2006; Mayer, Davis, & Schoorman, 1995).
- **Trust – reliability**: Reliability in hospitality services may include service knowledge, effective treatment, and valuable customer service, among others, to help customers quickly, efficiently and effectively. Cognitive trust is thus the perception of a guest about the competence and knowledge of the (skilful) host or staff relevant to the intended behaviour (Lu, Zhao, & Wang, 2010).

Once the probability matrix has been identified, our analysis assesses how well scores on the instrument indicate the theoretical construct. It employs network analysis by filtering the correlation matrix (> 0.20) to research the bonds between labels. The average clustering coefficient equals 0.59 (modularity ≥ 0.30). Overall, based here on the Louvain algorithm, the extracted communities are easily interpretable and give an impression highly consistent with the typology shown in Figure 7.

Figure 7 - Visualising the network of candidate labels



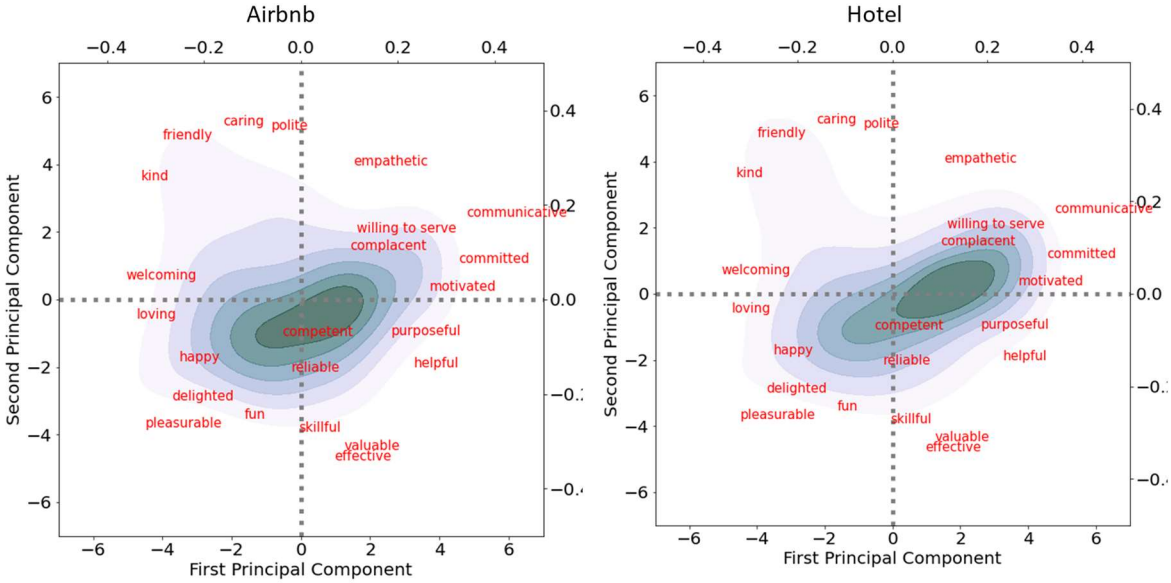
4.3 Principal Component Analysis

This research extends the previous knowledge by applying a Principal Component Analysis (PCA). PCA allows us to explain the variance in analysed data, extract the most informative candidate labels, create insightful biplots, and, in particular, test (per topic community and between both types of accommodation) the equality of means between both types of accommodation versus the alternative hypothesis of

differences between the means of each principal component. PCA is an exploratory data analysis method used to search a linear combination of the observed labels (here, Zero-shot probabilities matrix). Each principal component represents a percentage of the total variation captured from the dataset. In our analysis, principal component 1 holds 21.2 % of the information, while principal component 2 holds 14.6 % of the data (component 3 equals 11.3 %). Eight components explain 73 % of the variance. Our KMO value over 0.77 (> 0.7, good) and a significance level for Bartlett's test below 0.0001 suggest a substantial correlation in our probability matrix.

Our analysis in the biplot represents the labels' loading on the extracted components (PCA1 and PCA2 vectors) and the location of our data (their principal component scores) using a continuous probability density curve. As a result, sentences close to each other have similar probabilities of labels in our matrix. Vectors that point in the same direction can thus be interpreted as having similar meanings in the context set by our data. Although PCA could make characteristics less interpretable, the loadings here are readily understandable. In this sense, the more parallel to a principal component axis a vector is, the more it contributes only to that component (right angles represent a lack of correlation). For instance, labels such as welcoming and loving (left-hand side) or committed, motivated or purposeful (right-hand side) contribute to PCA1. On the other hand, polite or caring (top side) and valuable, skilful or effective (bottom side) contribute to PCA2. PCA1 could be conceptualised as a satisfaction/problem-solving orientation, while PCA2 represents benevolence/reliability (Figure 8).

Figure 8 - Principal component analysis biplot



The last step is to visualise the distribution of our data in this reduced space by communities of topics (and, consequently, narratives). Our analysis applies the Student's t-test for independent samples (and equality of variance) (see Table 3).

Table 3 - Independent samples t-test (Student Test)

Clusters (metatopics)	PCA 1 Airbnb	PCA 1 Hotel	PCA 2 Airbnb	PCA 2 Hotel	Airbnb vs Hotel: PCA1	Signif.	Airbnb vs Hotel: PCA2	Signif
	means							
Staff	-0.718	-1.815	1.422	2.489	10.772	*	-10.370	*
Noise	1.437	2.339	0.071	0.414	-9.236	*	-3.949	*
Walkability	0.050	0.560	-0.745	-0.353	-6.724	*	-7.284	*
In-room and bathroom	0.955	1.368	0.101	0.345	-3.327	*	-3.180	not
In-room amenities	0.662	1.291	-0.171	0.179	-1.466	not	-1.023	not
Clean & Comfort	-0.836	-0.090	-0.032	-0.019	-7.339	*	-0.179	not
Neighbour. amenities	0.610	0.745	-0.338	-0.198	-0.876	not	-1.681	not
Transportation	1.623	1.617	-0.462	-0.165	0.042	not	-2.210	not
Location	-0.544	0.277	-0.870	-0.420	-5.722	*	-3.774	*
Property views	-0.544	-0.489	-1.008	-0.580	-0.424	not	-4.818	*

* $p \leq 0.001$

5. Discussion

Our research describes a method of transforming free text (narratives) expressed by guests of tourist establishments into structured content as a basis for answering the research questions formulated at the beginning of the paper.

The combination of data mining techniques employed (BERTopic, Zero-shot Classification and Principal Component Analysis) allows us to obtain a valid procedure (algorithm) for the treatment of this unstructured data associated with the opinions of hotel establishment users (extracted from Tripadvisor) and Airbnb (obtained from its own infomediation platform).

The chosen data analysis methods are justified based on their suitability for the research problem, advantages over other alternatives, and availability in the existing literature and software tools. For example, using a database of thousands of guests' narratives enables comprehensive and diverse data collection, while NLP and zero-shot classification provide powerful techniques for textual data analysis. In addition, PCA helps to reduce noise and identify the most significant features.

In addition to the above, which, given the methodological nature of the paper, constitutes its main contribution, the application of these data mining techniques to the unstructured data captured allows us to answer one of the research questions initially formulated: "Are there significant distinctions between the most relevant characteristics (detected and) associated with hospitality services between types of lodging, Airbnb vs hotels?".

In relation to the latter, it should be noted that the main differences between the most influential attributes in visitor satisfaction and confidence with the accommodation service received, considering the Airbnb vs traditional hotels typologies, are as follows:

- Hotel guests (compared to Airbnb guests) tend to rate their experiences with staff performance (more specifically than Airbnb guests), with attributes (labels) associated with transaction-specific satisfaction (PCA1, left-hand side) and benevolence (PCA2, top-hand side). A hotel guest's transaction-specific satisfaction is thus more specifically influenced by the level of staff service that is in line with guests' expectations. The comments regarding staff are based on a guest's discrete assessment of a set of encounters occurring during the service process (tips on what to do at the destination).

  Hotel guests (compared to Airbnb guests) tend to rate their experiences related to noise and core services (in-room and bathroom facilities) with terms most highly associated specifically with problem-solving orientation labels (smooth communication, empathetic responsiveness or information provision from staff) (PCA1). In the case of noise, hotel guests also use terms more closely connected with the dimension of benevolence. Moreover, in-room and bathroom facilities tend to rate (for both types of guests equally) with words related to benevolence-based feelings (PCA2, top-hand side).

- Comments related to walkability (and the closeness of attractions and neighbourhood amenities) also tend to use terms associated with the problem-solving orientation dimension - particularly in the case of hotel stays. In the case of Airbnb stays, guests use words related to reliability-related tags to a greater extent than hotel guests.

- Among Airbnb guests, clean & comfort assessments are more specifically related to transaction-specific satisfaction labels than reviews published by hotel guests.

- For both guests, in-room and neighbourhood amenities and transportation systems load more problem-solving orientation in both hospitality services.

- Property views load higher on the satisfaction side for both guests. Likewise, among Airbnb guests, property views are more related explicitly to the reliability dimension.

- Location loads more on the problem-solving orientation labels side among hotel guests and more on satisfaction among Airbnb guests. Likewise, among Airbnb guests, location is more related explicitly to the reliability dimension.

5. Conclusions and implications

Our research contributes to the literature on hospitality services, offering insights for methodological research (theoretical implications) and allowing the design of customer service policies. In addition, it also allows for a more detailed understanding of the P2P phenomenon in tourist accommodation. Therefore, it could be said that our paper's contributions are structured in two directions. On the one hand, the study offers proposals to improve the research methodology in this field. On the other hand, it provides managerial contributions to designing customer satisfaction and trust policies and a more fine-grained understanding of P2P accommodation versus hotel services.

5.1 Theoretical implications

Firstly, the main implication of our research is methodological. The study applies novel techniques that improve the results obtained so far by others. A combination of BERTopic, Zero-shot classification and PCA allows us an appropriate analysis of the necessity and impact of hospitality attributes on relationship quality. Our research uses a text-mining technique to develop semantic structures to detect the subjects mentioned by visitors in their reviews. It enables gathering information in a trustworthy, authentic, and time-efficient manner. In addition, Zero-shot classification determines whether the hypothesis is true (entailment) or false (contradiction), given the premise, and concludes the confidence percentage score for each sentence and label. There is no need to provide training data, which is thus highly efficient. Thirdly, the PCA method displays

which topics (and their communities) are related to satisfaction and trust as requirements.

The combination of techniques detailed in the previous paragraph forms an automated identification and classification algorithm whose application makes it possible to solve the problem of managing large document archives by automatically generating latent topics in the users' own reviews or narratives, as well as their semantic grouping through an approach based on the discovery of patterns from machine learning models.

5.2 Practical implications

Regarding managerial implications, our findings provide relevant results for Airbnb and hotel managers, highlighting the attributes on which they should focus their services and activities from the perspective of guests' preferences and, consequently, from the degree of their satisfaction (and trust) with the service received. For example, according to our results, among Airbnb guests, labels associated with the location and the clean and comfortable hospitality services tend to load on transaction-specific satisfaction. On the other hand, among hotel guests (compared to Airbnb guests), labels highlighting guests' delight, ease of checking in/out, professionalism, and additional services, such as reservations, load highly on transaction-specific satisfaction and benevolence (caring, complacent, friendly or kind manners).

Moreover, hotel guests (compared to Airbnb guests) tend to rate their narratives with comments associated with a problem-solving orientation, such as location and walkability, in-room and bathrooms (and amenities), and noise. It is thus essential to anticipate and resolve problems that may arise before and during a service exchange. How hosts or staff deal with such issues is, therefore, crucial. For instance, a problem-solving orientation focused on a shower, a room temperature, or a sleep environment (noise) should be treated with empathy. In addition, it is a critical opportunity for the hospitality service provider to prove its commitment to the service (hearing the guests' problems, for example).

6. Limitations and future research

To sum up, terms related to hospitality features play an essential role in evidencing their contribution to guests' satisfaction or adequately influencing the perception of problem-solving orientation and reliability or benevolence. Nevertheless, future studies should delve into the positive review bias. In this regard, guests' narratives are influenced by their cultural scripts or demographic variables, such as age or income. Similarly, research should be conducted in other areas with various cultural scripts, tourist seasonality patterns, or tourist attractions to generalise the findings.

In addition to the above, future research could focus on several limitations of our paper. Among them: (1) To study more destinations with different tourism patterns or tourist attractions, (2) to compare our results with other ones obtained by different topic modelling approaches, (3) to analyse the influence of gender preferences (biological male versus female dichotomy) as a moderating factor in the relationship between service experiences with Airbnb and its pricing policies, and (4) to reflect critically about the consideration of Airbnb a being part of the sharing economy concept.

Credit author statement

All authors have contributed equally. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest: None

References

AirDND (2022). *MarketMinder*. Retrieved 10 January 2021 from https://www.airdna.co/

Alcofarado, A., Ferraz, T.P.., Gerber, R., Bustos, E., Oliveira, A.S., Veloso, B.M.,Siqueira, F.L.,& Costa, A.H.R. (2022). ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling. *Computational Processing of the Portuguese Language*. 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, ProceedingsMar 2022, 125–136.

Allee, V. (2003). *The Future of Knowledge. Increasing Prosperity Through Value Networks.* Burlington, MA: Elsevier.

Alqayed, Y. Foroudi, P., Kooli, K., Foroudi, M.M., & Dennis, C. (2022). Enhancing value co-creation behaviour in digital peer-to-peer platforms: An integrated approach. *International Journal of Hospitality Management*, 102, 103140. https://doi.org/10.1016/j.ijhm.2022.103140

Angelov, D. (2020). *Top2vec: Distributed representations of topics.* Available at: arXiv:2008.09470.

Bagozzi, RP, Gopinath, M.,& Nyer, P.U.(1999): The role of emotions in marketing. *Journal of the Academy of Marketing Science*, 27(2), 184-206

Baker, J., Levy, M., & Evans, J.R. (1992). An experimental approach to making retail store environmental decisions. *Journal of Retailing*, 68(4), 445-460.

Bakker, I., Voordt, Th., Vink, P., & De Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology*, 33(3), 405-421. https://doi.org/10.1007/s12144-014-9219-4

Barbosa, B., Saura, J.R., & Bennett, D. (2022). How do entrepreneurs perform digital marketing across the customer journey? A review and discussion of the main uses. *The Journal of Technology Transfer.* https://doi.org/10.1007/s10961-022-09978-2

Belarmino, A., Whalen, E., Koh, Y., & Bowen, J. T. (2017). Comparing guests' key attributes of peer-to-peer accommodations and hotels: mixed-methods approach. *Current Issues in Tourism*, 22(1), 1-7. https://doi.org/10.1080/13683500.2017.1293623

Bierner, P.P. (2010). Total survey error: Design, implementation and evaluacion. *Public Opinion Quarterly,* 74(5), 817-848. https://doi.org/10.1093/poq/nfq058

Bierner, P.P: (2014). *Dropping the "s" from TSE: Applying the paradigm to Big Data*. The 2014 International Total Survey Error Workshop (ITSEW 2014). Washington, DC: National Institute of Statistical Science

Bigné, J. E., Andreu, L., & Gnoth, J. (2005). The theme park experience: An analysis of pleasure, arousal and satisfaction. *Tourism Management*, 26(6), 833-844. https://doi.org/10.1016/j.tourman.2004.05.006

Bresciani S., Ferraris A., Santoro G., Premazzi, K., Quaglia, R. Yahiaoui, D., & Viglia, G. (2021). The seven lives of Airbnb. The role of accommodation types. *Annals of Tourism Research*, 88, 103170. https://doi.org/10.1016/j.annals.2021.103170

Callegaro, M., & Yang, Y. (2018). *The role of surveys in the era of Big Data.* In D.Vannette & J.Krosnick (eds.). The Palgrave Handbook of Survey Research. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-54395-6_23

Campello, R.J.G.B., Moulavi D., & Sander J. (2013). Density-based clustering based on hierarchical density estimates. In Pei J., Tseng V.S., Cao L., Motoda H., Xu G. (eds) *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science,* vol 7819. Berlin: Springer.

Cheng, M. (2016). Sharing economy: A review and agenda for future research. *International Journal of Hospitality Management*, 57, 60-70. https://doi.org/10.1016/j.ijhm.2016.06.003

Consejería de Turismo y Deporte Junta de Andalucía (2022): *Balance del año turístico en Andalucía (BATA)*, Empresa Pública para la Gestión del Turismo y el Deporte de Andalucía, Sevilla

Deloitte, (2019). *Prague Hospitality Report, Tourism, Hotels & P2P accommodation*. Deloitte Central Europe.

Dickinger, A. (2011). The trustworthiness of online channels for experience-and goal-directed search tasks. *Journal of Travel Research,* 50(4), 378-391. https://doi.org/10.1177/0047287510371

Dogru, T., Hanks, L., Mody, M., Suess, C., & Sirakaya-Turk, E. (2020). The effects of Airbnb on hotel performance: Evidence from cities beyond the United States. *Tourism Management*, 79, 104090. https://doi.org/10.1016/j.tourman.2020.104090

Dogru, T., Mody, M., & Suess, C. (2019). Adding evidence to the debate: Quantifying Airbnb's disruptive impact on ten key hotel markets. *Tourism Management*, 72, 27-38. https://doi.org/10.1016/j.tourman.2018.11.008

Dolnicar, S. (2018). *Peer-to-Peer Accommodation Networks: Pushing the Boundaries.* Oxford, United Kingdom: Goodfellow Publishers.

Dolnicar, S. (2020). Sharing economy and peer-to-peer accommodation – A perspective paper. *Tourism Review,* 76(1), 34-37. https://doi.org/10.1108/TR-05-2019-0197

Ert, E., Fleischer, A. & Magen, N. (2016). Trust and reputation in the sharing economy: the role of personal photos in Airbnb. *Tourism Management*, 55, 62-73. https://doi.org/10.1016/j.tourman.2016.01.013

Flavian, C., Guinaliu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43, 1–14. https://doi.org/10.1016/j.im.2005.01.002

Foroudi, P., & Marvi, R., 2021. Some like it hot: the role of identity, website, co-creation behavior on identification and love. *European Journal of International Management* (in press). https://doi.org/10.1504/EJIM.2023.10053692

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega*, 32(6), 407–424. https://doi.org/10.1016/j.omega.2004.01.006

Gentile, C., Spiller, N., & Noci, G. (2007). How to sustain the customer experience: An overview of experience components that co-create value with the customer. *European Management Journal*, 25(5), 395-410. https://doi.org/10.1016/j.emj.2007.08.005

Geyskens, I., Steenkamp, J-B.E.M., Scheer, L.K., & Kumar, N. (1996). The effects of trust and interdependence on relationship commitment: A trans-Atlantic study. *International Journal of Research in Marketing*, 13(4), 303-317. https://doi.org/10.1016/S0167-8116(96)00006-7

Grootendorst, M. (2020). *BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics.* Retrieved 15 December 2020 from https://doi.org/10.5281/zenodo.4430182.

Guo, Y., Barnes, S.J., & Jia, Q. (2016). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using Latent Dirichlet Allocation. *Tourism Managment*, 59, 467-483. https://doi.org/10.1016/j.tourman.2016.09.009

Guttentag, D., & Smith, S. L. (2017). Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *International Journal of Hospitality Management,* 64, 1-10. https://doi.org/10.1016/j.ijhm.2017.02.003

Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2018). Why tourists choose Airbnb: A motivation-based segmentation study. *Journal of Travel Research,* 57(3), 342–359. https://doi.org/10.1177/0047287517696980

Gyódi, K. (2017). Airbnb and the hotel industry in Warsaw: An example of the sharing economy? *Central European Economic Journal*, 2(49), 23-24. https://doi.org/10.1515/ceej-2017-0007

Hall, S. & Pennington, J. *How much is the sharing economy worth to GDP? (*2016). Retrieved July 6, from https://www.weforum.org/agenda/2016/10/what-s-the-sharing-economy-doing-to-gdp-numbers

Haywood, J., Mayock, P., Freitag, J., Owoo, K. A., & Fiorilla, B. (2017). Airbnb & hotel performance. *STR publication*. Hendersonville, USA.

Hugging Face (2023). *What is Zero-Shot classification?* Retrieved 10 March 2021 from https://huggingface.co/tasks/zero-shot-classification

Ikkala, T., & Lampinen, A. (2014), Defining the price of hospitality: networked hospitality exchange via Airbnb. *In Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 173-176. Baltimore: ACM, MD

Jolliffe, I.T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 374 (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Jones, M. A., & Suh, J. (2000). Transaction-specific satisfaction and overall satisfaction: An empirical analysis. *Journal of Services Marketing*, 14, 147-159. https://doi.org/10.1108/08876040010371555

Ju, Y., Back, K. J., Choi, Y., & Lee, J. S. (2019). Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *International Journal of Hospitality Management*, 77, 342-352. https://doi.org/10.1016/j.ijhm.2018.07.014

Kessler, J.S. (2017). *Scattertext: a Browser-Based Tool for Visualising how Corpora Differ.* Retrieved 10 March 2021 from arXiv: 1703.00565v3

Lalicic, L., & Weismayer, C. (2018). A model of tourists' loyalty: the case of Airbnb. *Journal of Hospitality and Tourism Technology,* 9(1), 80–93. https://doi.org/10.1108/JHTT-02-2017-0020

Lazos, F.J.M., & Steenkamp, J.B.E.M. (2005). Emotions in consumer behaviour. A hierarchical approach. *Journal of Business Research*, 58(10), 1437–1445. https://doi.org/10.1016/j.jbusres.2003.09.013

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* 7871–7880.

Li, J., Hudson, S., & So, K. K. F. (2019). Exploring the customer experience with Airbnb. *International Journal of Culture, Tourism and Hospitality Research*, 13(4), 410-429. https://doi.org/10.1108/IJCTHR-10-2018-0148

Liang, L.J. (2015). *Understanding repurchase intention of Airbnb consumers: perceived authenticity, EWoM and price sensitivity.*Unpublished Master's Thesis, University of Guelph, Canada

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2020). *RoBERTa: A robustly optimised BERT pretraining approach.* arXiv preprint arXiv: 1907.11692

Lovelock, Ch., & Wirtz, J. (2007). *Services marketing: People, technology, strategy.* 6th ed. New Jersey: Prentice Hall.

Lu, Y., Zhao, L., & Wang, B. (2010). From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers' purchase intention. *Electronic Commerce Research and Applications,* 9(4), 346–360. https://doi.org/10.1016/j.elerap.2009.07.003

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organisational trust. *Academy of Management Review*, 20(3), 709-734. https://doi.org/10.2307/258792

McInnes L., & Healy J. (2017). Accelerated Hierarchical Density-Based Clustering. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 33-42.

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861. https://doi.org/10.48550/arXiv.1802.03426

Meyer, C., & Schwager, A. (2007). Understanding customer experience. *Harvard Business Review*, 85(2), 116.

Mody, M., & Hanks L. (2020). Consumption authenticity in the accommodations industry: The keys to brand love and brand loyalty for hotels and Airbnb. *Journal of Travel Research*, 59(1), 173-189. https://doi.org/10.1177/0047287519826233

Mody, M., Suess, C., & Lehto, X. (2019). Going back to its roots: can hospitableness provide hotels competitive advantage over the sharing economy. *International Journal of Hospitality Management*, *76*, 286-298. https://doi.org/10.1016/j.ijhm.2018.05.017

Mudie, P., Cottam, A., & Raeside, R. (2003). An exploratory study of consumption emotion in services. *The Service Industries Journal*, 23(5), 84–116. https://doi.org/10.1080/02642060308565625

Oliver, R. (1997). *Satisfaction: a behavioural perspective on the customer*. New York: McGraw-Hill.

Oliver, R. (2010). *Customer Satisfaction.* Wiley International Encyclopedia of Marketing

Osman, H., D'Acunto, D., & Johns, N. (2019). Home and away: Why do consumers shy away from reporting negative experiences in the peer-to-peer realms? *Psychology & Marketing*, 36(12), 1162–1175.

Pappas, N. (2019). The complexity of consumer experience formulation in the sharing economy. *International Journal of Hospitality Management*, 77, 415-424. https://doi.org/10.1016/j.ijhm.2018.08.005

Petroni, F., Rocktaschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., & Miller, A. (2019). Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (2463-2473). Association for Computational Linguistics. https://dolorg/10.18653/V1/d19-1248

Pijls, R., Groen, B.H., Galetzka, M., & Pruyn, A.T.H. (2017). Measuring the experience of hospitality: Scale development and validation. *International Journal of Hospitality Management*, 67, 125-133. https://doi.org/10.1016/j.ijhm.2017.07.008

Prahalad, C.K., & Ramaswamy, V. (2004). *The Future of Competition: Co-Creating Unique Value with Customers*. Boston, MA: Harvard Business Press.

Quoquab, F, & Mohammad, J. (2022). The salient role of media richness, host-guest relationship, and guest satisfaction in fostering airbnb guests' repurchase intention. *Journal of Electronic Commerce Research*, 23(2), 59-76.

Rodríguez, I., & San Martín, H. (2008). Tourist satisfaction. A cognitive – affective model. *Annals of Tourism Research*, 35(2), 551–573. https://doi.org/10.1016/j.annals.2008.02.006

Sainaghi, R., & Baggio, R. (2020). Substitution threat between Airbnb and hotels: Myth or reality? *Annals of Tourism Research,* 83, 102959. https://doi.org/10.1016/j.annals.2020.102959

Sánchez-Franco, M.J., Navarro-García, A. & Rondán-Cataluña, F.J. (2016). Online customer service reviews in urban hotels: A data mining approach. *Psychology & Marketing*, 33(12), 1174-1186. https://doi.org/10.1002/mar.20955

Sánchez-Franco, M.J., & Rey-Moreno, M. (2021). Do travelers' reviews depend on the destination? An analysis In coastal and urban peer-to-peer lodgings. *Psychology & Marketing*, 39(2), 441–459. https://doi.org/10.1002/mar.21608

Santos, J.A.C., Fernández-Gámez, M.A., Solano-Sánchez, M.A., Rey-Carmona, F.J., & Caridad y López del Río, L. (2021). Valuation models for holiday rentals'daily rates: Price composition based on Booking.com. *Sustainability*, 13(1), 292, https://doi.org/10.3390/su13010292

Saura, J.R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Exploring the boundaries of opern innovation: evidence from social media mining. *Technovation*, 119, 102447. https://doi.org/10.1016/j.technovation.2021.102447

Saura, J.R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. *Industrial Marketing Management,* 98, 161-178. https://doi.org/10.1016/j.indmarman.2021.08.006

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, 66(1), 15–37. https://doi.org/10.1509/jmkg.66.1.15.18449

Solano-Sánchez, M.A., Santos, J.A.C., Santos, M.C., & Fernández-Gámez, M.A. (2021). Holiday rentals in cultural tourism destinations: a comparison of Booking.com-based daily rate estimation for Seville and Porto. *Economies*, 9(4), 157, https://doi.org/10.3390/economies9040157

Sthapit, E., & Jimenez-Barreto, J. (2018). Exploring tourists' memorable hospitality experiences: An Airbnb perspective. *Tourism Management Perspectives*, 28, 83-92. https://doi.org/10.1016/j.tmp.2018.08.006

Sun, Y., Ma, H., & Chan, E.H.W. (2018). A model to measure tourist preference towards scenic spots based on social media data: A case of Dapeng in China. *Sustainability,* 10(1), 43. https://doi.org/10.3390/su10010043

Tan, F.B., & Sutherland, P. (2004). Online consumer trust: A multi-dimensional model. *Journal of Electronic Commerce in Organizations,* 2(3), 40-58. https://doi.org/10.4018/jeco.2004070103

Tussyadiah, I. P., & Pesonen, J. (2016). Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research*, 55(8), 1022-1040. https://doi.org/10.1177/0047287515608505

Tussyadiah, I. P., & Zach, F. (2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing*, 34(5), 636-652. https://doi.org/10.1080/10548408.2016.1209153

Virtanen, P, Gomers, R., Oliphant, T.E.,Haberland, M., Reddy, T.,...& Van Mulbregt (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261-272. https://doi.org/10.1038/s41592-019-0686-2

Xian, Y., Lampert, C.H., Schiele, B., & Akata, Z. (2020). *Zero-Shot learning: A comprehensive evaluation of the good, the bad and the ugly*. arXiv preprint arXiv: 1707.00600

Yang, G., Ye, Z., Zhang, R., & Huang, K. (2022). A comprehensive survey of Zero-Shot image classification: methods, implementation and fair evaluation. *Applied Computing and Intelligence*, 2 (1), 1-31. https://doi.org/10.3934/aci.2022001

Ye, S., Chen, S., & Paek, S. (2023). Moderating effect of trust on customer return intention formation in peer-to-peer sharing accommodation. *Journal of Hospitality & Tourism Research,* 47(2), 328-353. https://doi.org/10.1177/10963480211014249

Yin, W., Hay, J., & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 3914–3923, Hong Kong, China, November 3–7, 2019.

Yu, Y., & Dean, A. (2001). The contribution of emotional satisfaction to consumer loyalty. *International Journal of Service Industry Management*, 12(3), 234–250. https://doi.org/10.1108/09564230110393239

Zach F.J., Nicolau J.L., & Sharma A. (2020). Disruptive innovation, innovation adoption and incumbent market value: The case of Airbnb. *Annals of Tourism Research*, 80, 102818. https://doi.org/10.1016/j.annals.2019.102818

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American psychologist*, 35(2). https://doi.org/10.1037/0003-066X.35.2.151

Zervas G., Proserpio D., & Byers J.W. (2021). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters*, 32, 1–16. https://doi.org/10.1007/s11002-020-09546-4

Zhu, L., Lin, Y., & Cheng, M. (2020). Sentiment and guest satisfaction with peer-to-peer accommodation: when are online ratings more trustworthy? *International Journal of Hospitality Management*, 86 (3688), 102369. https://doi.org/10.1016/j.ijhm.2019.102369