NOVA IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Determinants of Airbnb guest Satisfaction: A comparative study between Europe and East Asia

Mónica Catarina Sousa Roseiro

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

BOOK SPINE

2019

Title:
Subtitle:

Student
full name

MEGI

2019

Title:
Subtitle:

Student
full name

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

DETERMINANTS OF AIRBNB GUEST SATISFACTION: A COMPARATIVE STUDY BETWEEN EUROPE AND EAST ASIA

by

Mónica Roseiro

Dissertation report presented as a partial requirement for obtaining the Master's degree in Information Management, with a specialization in Marketing Intelligence

Co-Advisor: Nuno António

Co-Advisor: Paulo Rita

November 2022

ABSTRACT

The rising sharing economy has impacted the lodging industry over the years because of Airbnb's popularity growth. Europe and East Asia are two culturally different regions, and considering that guest expectations might vary, this study analyzes and compares the determinants of Airbnb guest satisfaction in both regions, getting a new perspective on the topic. The analysis is based on a dataset containing English reviews between 2017 and 2021 from 6 different cities – Paris, Madrid, Rome (Europe), Tokyo, Beijing, and Hong Kong (East Asia). A topic modeling algorithm (Latent Dirichlet Allocation) is used to discover latent topics in Airbnb reviews. The findings reveal that most drivers of satisfaction are similar in both regions - location, accommodation, and amenities. In East Asia, guests value more the unique experience that Airbnb can provide, while in Europe, guests are more demanding and focus more on the negative attributes of the accommodation. Therefore, this study aimed to discover the factors that affect guest satisfaction in Airbnb, providing a new angle by comparing two geographic regions. From a practical perspective, it helps to understand how guest expectations vary and identify Airbnb's opportunities and challenges in the studied regions.

KEYWORDS

Sharing economy; Airbnb; Online Reviews; Text mining; LDA

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The popularity of the sharing economy has been increasing worldwide. It is a recent phenomenon that has changed industries with their new disruptive business models (CBRE, 2020). Sharing economy, also known as "collaborate consumption," is a term introduced by Botsman & Rogers (2010). They state that sharing economy describes the online mediated peer-to-peer services and goods exchange (Hamari et al., 2015). The development of new technologies and the growth of web 2.0 allowed the creation of new platforms that support collaboration. This evolution encouraged both parties to communicate and transact in real time without mediators (Belk, 2014).

The sharing economy has various advantages for the provider and the obtainer. The sharing economy platforms permit cheaper access to goods and services since it reduces transaction costs with fewer intermediaries. Moreover, it creates an additional revenue opportunity. Compared to traditional markets, the collaborative market is less bureaucratic. Instead, it is more dynamic and flexible. The sharing economy is expected to be worthen 335 billion dollars in 2025 (Perren & Grauerholz, 2015). Younger generations, between 25 and to 35-year-old, are also three times more likely to use a sharing platform rather than people aged 55 years or older (Lim, 2020). The tourism industry ecosystem has been changing over the last few years, with new players emerging from the sharing economy.

The sharing economy's most considerable growth and impact were in the tourism sector (Moreno-Gil & Coca-Stefaniak, 2020; Moro et al., 2019). Airbnb is a sharing economy platform that operates as an intermediary between hosts and guests looking for apartments, private rooms, villas, and other short-term accommodations for different prices. Airbnb is considered the pioneer of this shift and revolutionized the industry (Moon et al., 2019). Airbnb is present in over 100 000 cities, with over 6 million listings in 191 countries (Airbnb, n.d.). The value proposition offered by Airbnb – "live like a local" authenticity – attracts travelers, in addition to lower prices as primary reasons why guests choose the online platform over hotels.

The top three most influential factors when choosing a place are price, image/reputation, and reviews (Varma et al., 2016). The user-generated and reciprocal review system allows both hosts and guests to review each other, minimizing the risk and increasing trustworthiness.

The distributions of Airbnb reviews are strongly skewed towards higher rantings, and this outcome might be because the reviews are not anonymous, as most of the sharing economy platforms demand it to build confidence (Meijerink & Schoenmakers, 2020).

The increased internet usage allowed customers to share experiences, recommendations, and thoughts. This makes it easier to generate increasingly more data. Reviews are a reliable and important source of data for businesses and researchers. Electronic Word of Mouth (e-WOM) is the positive or negative exchange of information regarding products/services use or characteristics through digital and electronic means (Rodgers & Wang, 2011). The content that the guest produces while writing a review corresponds to e-WOM marketing (Liang et al., 2017).

The study of guest satisfaction has been researched over the last few years (Guttentag, 2019). Due to the easy access to large amounts of data, most research focus on extracting topics affecting customer satisfaction in one big city – like Los Angeles, New York, or Beijing. Besides, the fact that most studies focus on only one city has one significant disadvantage: the dimensions affecting customer satisfaction found in these studies are context-dependent since they include location-specific factors. Cheng and Jin (2019) cited as a limitation the importance of doing multicultural comparative studies to determine how subjects are influenced by context - as regions of stay.

For the analysis, the evolution of big data analytics, which text mining and NPL (natural processing language) are part of, allowed the extraction of patterns, associations, and valuable information from big amounts of unstructured data– text, video, image, and audio (Kumar et al., 2021).

Topic modeling through Latent Dirichlet Allocation (LDA) was applied to find similarities between the two regions' online reviews (Chen & Xie, 2020). LDA is a statistic model that identifies topics through similar words in distinct corpus documents. In addition, it discovers relations between the documents and topics (Karami et al., 2020).

The Asia-Pacific region has been growing and establishing as a destination trend in tourism over the last few years. Between 2010 and 2017, this region grew by 63%. The main reasons for tourism development in this region are the growth of air connectivity, infrastructure development, lower travel costs, and an easier online booking system (Akbulut & Ekin, 2019). It is the second most visited region, followed by Europe since 2002. Europe is the biggest travel destination, accounting for over 50% of international travelers and generating 37% of global tourism (Europe Statistics, n.d.). In addition, most Airbnb studies are focused on North America (40,2%) (Guttentag, 2019b). Europe and Asia are the regions with higher active listings in 2020, 2,475,055 and 1,490,069, respectively.

Also, the gross revenue (revenue made by the host and Airbnb) is much higher in Europe ($8,993,287,813) than in Asia Pacific ($3,855,016,273). The average price in Asia Pacific is 83$ while in Europe, it is more expensive – 90$ (Airbnb & Vacation Rental Statistics [2021], 2021). Since Asia is the largest continent and culturally broad, East Asia will be the region used for the study.

Three cities from each region were chosen to obtain a correct portrayal of factors that influence guest satisfaction. The criteria were based on the capital of the countries that had the most international tourist arrivals in 2019. The base year to decide the cities of the dataset was 2019 and not 2020 due to the strike of the new coronavirus. Compared to 2019, there was a global decrease of 73% in international tourist arrivals (International Tourism and Covid-19 | Tourism Dashboard, n.d.).

Therefore, the capitals of the countries with the largest international tourist arrivals in 2019 were chosen to represent the European region accurately. France, Spain, and Italy were the European countries with the most inbound arrivals (89.4 million, 89.4 million, and 64.5 million, respectively) (Statista, 2021). Hence, Paris, Madrid, and Rome are cities chosen to represent Europe. Regarding East Asia, in 2019, China, Japan, and Hong Kong had the most international tourist arrivals (65.7 million, 32.2 million, and 23.8 million, respectively) (UNWTO, 2020). Hence, Beijing, Tokyo, and Hong Kong are the cities chosen to represent East Asia. Therefore, a comparison between regions based on different regions with different characteristics and cultural backgrounds is lacking. Based on the identified gap, and based on the previous explanation, to verify that the factors influencing guest satisfaction do vary or do not vary across regions using Europe and East Asia, the following question was raised:

What are the influential consumer satisfaction factors in Europe versus East Asia?

This study aims to identify the main differences in Airbnb customer satisfaction between Europe and East Asia. Madrid, Paris, and Rome, representing Europe, and Beijing, Tokyo, and Hong Kong representing East Asia, will be used as input data to find the topics by region.

To address the research question previously stated, the study goals are:

- Describe the Airbnb property features and basic review characteristics of each city.
- Extract latent topics that compose the dimensions that influence guest satisfaction by region.
- Identify and compare the differences and similarities between both regions.

2. LITERATURE REVIEW

2.1 SHARING ECONOMY

The emergence of Internet-based technologies provided new opportunities, including developing new market alternatives such as the sharing economy. This economy creates disruptive innovations by facilitating the interaction of supply and demand (Hira & Reilly, 2017; Belk, 2014). The importance of the sharing economy has increased more than ever since the last decade (Sundararajan, 2016).

This concept created several new markets as consumers shared and monetized their own goods, such as houses or cars. In the last decade, sharing economy has been considered one of the major socio-economic developments (Cheng, 2016). Now, consumers have a new, two-sided role. Changed from having a passive to an active role in the market, they become “prosumers”. The prosumers are producers and consumers and, meanwhile, generate value. Consumer habits changed with the growth of the sharing economy (Botsman and Rogers, 2010). Companies must adapt to face new arrivals coming from this disruptive model (Davis, 2016). The best examples are Airbnb and Uber, which disrupted the mobility and travel sectors (Tussyadiah, 2016).

By creating new types of relationships, companies such as Airbnb, Uber, BlaBlaCar, and TaskRabbit have in common that they use web-based platforms as a mediator of relations and payments. Connecting buyers and sellers through these platforms reduces transaction costs and information uncertainty and improves the allocation of resources (Hira & Reilly, 2017; Cetim, 2020). Also, this business model creates economies of scale that allow for rapid growth and create new opportunities for new players looking to enter the market (PWC, 2020).

These changes made the market rethink the optimization of resources and how they should be offered and consumed (Lovins & Cohen, 2011). According to Belk (2014), a paradigm shift in society is happening from *“you are what you own”* to *“you are what you share”.* It is expected that the importance of sharing economy keeps growing and positively affect the economy since it catalyzes individual innovation and entrepreneurship (Sundararajan, 2014).

Sharing was always part of society; some examples are public transport, and community spaces according to the feeling of community. For this reason, it is widely studied in different fields, including economics and philosophy (Smolka and Hienerth, 2014). Sharing Economy is such a broad term that incorporates different markets, business models, and products. It is through short-term access that captures value (Daunoriené et al. 2015).

Although, only with internet development was it possible to do it on a bigger scale with the rise of e-commerce and online peer communities. Therefore, the term “sharing economy” as “collaborative consumption” was popularized by Botsman and Rogers (2010). After the initial concept introduced by Botsman and Rogers (2010), much research proposed new terms and definitions of the sharing economy. A sharing economy is “the peer-to-peer-based activity of obtaining, giving, or sharing access to goods and services, coordinated through community-based online services” (Hamari et al., 2015). Some sharing economy interchangeable terms are “collaborative economy, “peer-to-peer economy”, “gig economy”, and “on-demand economy” (Petropoulos, 2017).

Botsman and Rogers (2010) defined sharing economy as “Traditional sharing, bartering, lending, trading, renting, gifting, and swapping redefined through technology and peer communities - that is, remodeling business”. Belk (2014) considers this definition created by Botsman and Rogers (2011) ambiguous. He argues that not all types of exchange should be included in the definition of “sharing economy,” such as lending or giving. Therefore, he reformulated his definition by adding that some fee or compensation should be given when sharing or distributing resources. With this new element, Couchsurfing, according to Belk (2014), is not considered part of the sharing economy.

Based on the attempts to define the sharing economy, Frenken et al. (2015) identified common characteristics among the definitions. Firstly, it is an exchange made through consumer-to-consumer platforms, meaning that the suppliers can include non-professional individuals. This differs from the traditional two-sided market (Li, Moreno, & Zhang, 2015). Secondly, the products or services are facilitated instead of purchased - there is no transfer of ownership (Botsman, 2015). Finally, by being an on-demand economy, it efficiently uses physical assets, and the value is negotiated between both parties. It distances itself from the traditional supply and demand model since customers connect through digital platforms (Botsman & Rogers, 2010).

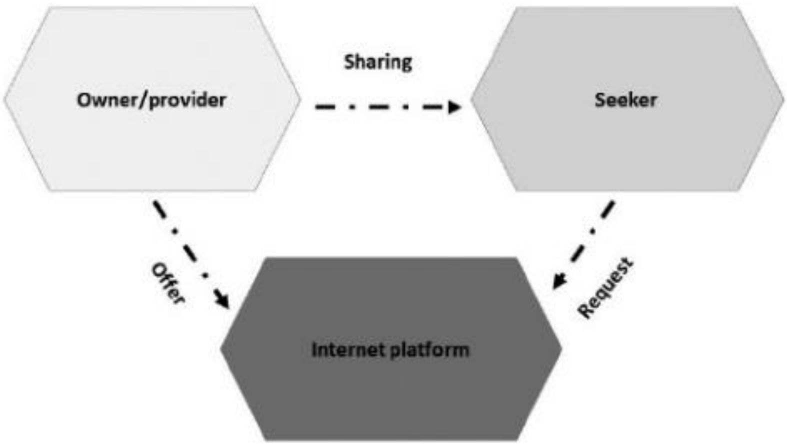

Figure 1- Sharing Economy Model. Source: (Grybaitė & Stankevičienė, 2016)

The sharing economy model is based on three main participants: the provider (owner of the asset), the buyer (who seeks the product), and the internet platform that creates an open market serving as an intermediate for the interaction (Grybaitė & Stankevičienė, 2016).

According to Botsman and Rogers (2010), for the sharing economy system to be self-sustainable, one principle must be accomplished: the products and services offered must meet the needs of consumers.

A survey showed which products people are more likely to share. It was reported that people are more willing to share or rent electronic devices (28%). Followed by intellectual property (26%), power tools (23%), bicycle (22%), clothing (22%), and household items (22%). (Nielsen, 2014).

Table 1 shows that the region that is more willing to share its own assets is Asia-Pacific. This region is above the average. On the other hand, Europe is the region which less intends to share its assets less. (Nielsen, 2014)

Asia-Pacific	Latin America	Middle East/Africa	North America	Europe	Global Average
78	70	68	52	54	68

Table 1 - Willingness to share own assets (%) (Nielsen, 2014)

2.1.1 Drivers

The sharing economy has been growing mainly since the last decade. According to an international study, there are many underlying motivations for the rise of its popularity (Comms, 2014). The research conducted by Havas Worldwide, partnered with Market Probe International, in 2014, found that the top three appealing factors for participating in the sharing economy are saving money, being active and useful, and reducing carbon footprint. In addition, several reasons indicate this growth. This growth can be divided into four main drivers: technological, social, economic, and environmental.

- **Economic Drivers**

Although technological drivers were the starting point of the big development of the sharing economy, economic reasons are the primary reason individuals participate in it. Collaborative consumption was seen as an opportunity for many during the financial crisis in 2008/2009 to get access to not affordable goods and prevent avoidable expenses (Rauch and Schleicher, 2015). Participating in the sharing economy as a provider utilizes unused assets to generate extra income and reduce ownership costs (Schor, 2016). At the same time, it increases the market range by providing a new alternative to consumers that is usually more economical and less expensive than the traditional options. With this, the money is distributed through the value chain towards the consumers instead of the intermediaries (Schor and Fitzmaurice, 2015).

- **Technological Drivers**

Technological development was the main reason why sharing economy became a global phenomenon since it allowed the sharing to be extended from a small group of people like family or friends to any person by creating a peer-to-peer network through the development of Web 2.0 (John, 2013). With web 2.0, creating content and communication among users was possible (Carroll & Roman, 2011). This way, technology facilitates interaction with trust by matching supply and demand (Hsu, Ju, Yen, & Chang, 2007).

Trust is a crucial element in the interaction in the sharing economy. Through User-Generated Content (UGC), one can share with another about their own experiences with the product or service, including giving ratings, reviews and sharing visual media such as photos or videos. Like this, they provide essential information, and users feel confident while sharing (Plank, 2016; Akehurst, 2009). Besides that, the increase in social media incorporation also helps to build trust. Social media increases credibility by decreasing anonymity (Teubner, 2014; Adam, Camacho, & Hassanein, 2014). This is important in the digital era where new habits were created: information search, communication among users, and buying products through digital platforms. (John, 2013).

Technology also allowed the introduction of online payment systems like PayPal, which diminishes fraud risk. Trust also increases with the intervention of intermediaries, such as Airbnb or Uber, between users and providers (Ranchordás, 2015; Owyang, 2013). The increase in smartphone penetration is also an important factor in the rise of the sharing economy since providers can take reservations, and users can search for it easier and faster (Olson & Kemp, 2015; Stephany, 2014). The low costs of developing a sharing platform due to technological developments were also key in the huge range of underutilized resources available to share (John, 2013).

- **Social Drivers**

During the last decades, there has been a paradigm shift in people's mentalities. The attitude towards product ownership and what they value changed: consumers now value more what they experience than they own. This drives the growth of sharing economy (Botsman and Rogers, 2010). Besides, feeling solidarity, community, and belongness to a group is also one of the reasons why sharing is attractive (Belk, 2010). According to Botsman & Rogers (2010), creating and developing connections is also a new opportunity. Airbnb is an excellent example since it allows travelers to connect with locals.

The interaction among individuals and the feeling of helping others generating word-of-mouth led to a higher trust in this organic information than the traditional information given by the brands. This helped the growth and popularity of the sharing economy (Owyang, 2013; Stephany, 2015). This way, through peer review systems, consumers become co-creators of value and social validation (Schor, 2016; Guttentag, 2015).

- **Environmental Drivers**

One of the main benefits of the sharing economy is the better use of idle resources, increasing its efficiency and the potential energy savings since it decreases waste, the production of new products, and raw materials while providing better accessibility to the resources available (Bardhi and Eckhardt, 2012). With the increasing awareness and concern about the environmental crisis, individual values and consumption choices are changing to pursue more sustainable options (Schor, 2010; Gansky, 2010). Therefore, the sharing economy is seen as a new alternative to overconsumption and a more sustainable behavior with a lower environmental impact (Botsman & Rogers, 2011; Piscicell et al., 2015).

2.1.2 Sharing Economy in the hospitality sector

The sharing economy is present in several industries, including travel, services, music, and car rental (PWC, 2015). The travel industry is considered one of the leading industries in the sharing economy. That is because it is one of the fastest-growing sectors in the world, with gross revenue of $1.6 trillion in 2017 (Quinby, 2017).

New players arrived a re-defined the tourism ecosystem and travel destinations (Koens et al., 2018). Some famous sharing economy companies in this market are Couchsurfing, founded in 2004, Airbnb, founded in 2008, and Windu, founded in 2011. These platforms act as intermediates between the owners, also known as hosts, and the guests. The integrated User Generated content significantly impacts the consumer decision-making process (Li and Wang, 2011). Hosts act as micro-entrepreneurs by providing travelers with accommodation for a cost and earning an extra income (Sundararajan, 2014; Jung et al., 2016).

The hospitality industry in the sharing economy is a success, with start-ups becoming global companies like Airbnb (Juul, 2015; Shuqair et al., 2021). Unlike traditional accommodation, peer-to-peer accommodation has no barriers to entering the market; it is not offered by professionals and has flexible listings (Guttentag, 2015; Tussyadiah and Zach, 2015). Some hosts have the flexibility to charge less since some fixed costs are covered (Guttentag, 2015). These platforms are an alternative to mainstream traveling, where travelers look for more individualized and authentic experiences. (Paulauskaite et al., 2017).

Even though there was a fast and massive development of the sharing economy in the tourism sector, it also had some negative consequences. Those are mainly due to legal issues, in which regulations are still not well defined (Guttentag, 2015). The increase in unlicensed rents and deficiency of permissions are among some legal discussions (Stephany, 2015).

Airbnb

Airbnb was created in 2008 in San Francisco by Brian Chesky, Joe Gebbia, and Nathan Blecharczyk. The idea initially developed by the three colleagues started when they had difficulties with rent payments and started renting their living room, taking advantage of the fact that the hotels were full because of the conference. Nathan Blecharczyk as a programmer joined the team by helping to create a website where travelers could look for places available (Friedman, 2013; Olson and Kemp, 2015; Sundararajan, 2016). Since then, Airbnb has become the biggest player in the sharing economy hospitality industry, disrupting the travel industry. In 2015, it only stayed behind the Hilton Group compared to the most valuable hotel chain (Hospitality, 2015). Airbnb revenues are generated by charging 13% of the guests and 3% from the host due to handling fees.

Airbnb's value proposition stands on the feeling of staying home, belonging to a community, and the uniqueness of the experience (Liu and Mattila, 2017). Besides having the same seasonal demand, research proved that Airbnb is not a direct competitor and a threat to traditional lodging options since the characteristics that satisfy travelers are different (Sundararajan, 2014). It was found that Airbnb is more used for causal and group traveling than hotels for business travel (Griswold, 2015). Sundararajan (2016) stated that the demand in the hospitality sector increased with Airbnb instead of taking it from traditional lodging options.

Airbnb is a peer-to-peer online marketplace where owners can list their idle resources, making it available for others to rent for a certain duration (Olson and Kemp, 2015). It contains more than 6 million available properties in 191 countries – more than the top five hotel groups combined (Airbnb, 2021).

Airbnb transactions include two parties: the host and the guest. Several property options include private and shared rooms, private houses, treehouses, and others. (Friedman, 2013). Hosts can list their proprieties without fee, stating space characteristics and description, price, number of guests allowed availability, and pictures (Zervas & Prosperio, 2016). The guests can access the listing and find the properties that fit their search according to the location, price availability, and type of property (Airbnb, 2019d).

Trust in the Airbnb website and as a company is noticed as one of the main reasons why the website stood out and became popular. The perception of the institution, including policies and infrastructures, and trust-building engines, affects consumer decision-making (Liang et al., 2018; Teubner, 2017).

Motivations to choose Airbnb

Airbnb, compared to hotels, still underperforms according to security and service quality, although many travelers still choose it. Different underlying motivations lead travelers to choose Airbnb. The primary motivation driver is utilitarian, and the top motivation is cost saving (Paulauskaite et al., 2017). Sthapit and Jiménez-Barreto (2018a) identified price and location as the main reasons travelers choose Airbnb. The main experimental motivation is local authenticity, including Airbnb's philosophy of "living like a local".

It was found that the practical benefits of Airbnb, including location, price, and household amenities, are the main motives, although other social factors as social interactions and closer contact with the culture, were mentioned (Paulauskaite et al., 2017; Guttentag et al.; 2018).

A study by Ting (2017) showed that 30% of interviewees never used Airbnb for privacy reasons and 9% for security reasons. This shows that the review system is even more important in peer-to-peer accommodation than in the traditional sector. The word-of-mouth generated through reviews is also a motive why people choose Airbnb because they buy first-hand opinions of people of experienced it (Owyang et al., 2014).

2.2 CUSTOMER SATISFACTION

Customer satisfaction is an abstract and ambiguous concept. It can be defined as the evaluation outcome of a former service or product experience (Fang et al., 2016). Oliver (1980) defined it as the fulfillment of response judgment. It involves expectations and performance since it can be considered how well a product performed according to what the customer expected. The customer is satisfied when the expectations meet or exceed the perceived performance. Otherwise, the customer is dissatisfied (Keller, 2022). According to Gaedeke et al. (1974), one is satisfied when a product attribute matches their needs, including technical functions and complimentary services as post-sales service.

Customer satisfaction is an individual response after a specific contact with a brand, product, or service that influences satisfaction (Meyer and Schwager, 2007). Different touchpoints can be divided into groups. The first one is direct contact which usually is initiated by the customer when searching, buying a product, or using a post-sales service. The second is usually unforeseen contact through advertisements, reviews, or word of mouth (Richardson, 2010).

Based on different definitions, Saxena (2017) identified four common characteristics. Firstly, it is related to a particular characteristic (product, service, among others). Secondly, it is an emotional or cognitive response. Thirdly, the response happens at a certain time (product search, after purchase). Lastly, the response comes after a certain judgment of expectations and evaluations.

2.2.1 Customer Satisfaction in the Hospitality Industry

It is crucial for different sectors, including hospitality, since positive satisfaction positively influences repurchase intentions, word of mouth, and profitability (Peterson and Wilson, 1992; Su, 2004).

Several studies analyzed the main factors influencing customer satisfaction in the lodging sector. Location, room cleanliness and comfort, service, security, and staff courtesy were revealed to be the most important factors determining customer satisfaction (Knutson, 1988; Akan, 1995). Value for money also has an important aspect alongside room and staff quality as the three most important aspects, according to Choi and Chu (2001).

At the peer-to-peer accommodation, the factors identified to influence customer satisfaction may differ since, as stated before, they fulfill distinct demands and expectations are different.

The determinants of customer satisfaction are highly related to the motivation to participate in the sharing economy. Satisfaction factors are the perceived reciprocal benefits of sharing (Bellotti et al., 2015). Sustainability and social innovation make the participants feel active and that they are contributing to reducing environmental impacts and responsible citizens, creating this inner satisfaction feeling (Sheth et al., 2011). Besides, the most influential factors in the collaborative consumption of the hostility industry include location and propriety facilities. According to Tussyadiah and Pesonen (2016), the most influential elements for satisfaction values for money, the accommodation characteristics (including host friendliness, comfort, and location), and the enjoyment of the experience.

2.3 ONLINE REVIEWS

Nowadays, online reviews have become a relevant factor in the customer decision-making process by helping to deduce the quality and performance of a service or product, decreasing insecurity (Liu, and Zhang, 2008). For businesses, it is an important source of information helping to understand more about their product and service quality (Chen & Xie, 2008). Online reviews are the product or service evaluations posted online (on the company or third-party websites) by the customers (Mudambi & Schuff, 2010).

The reviews can have a numerical rating, for example, 1 to 10, or be an open space for comments and media sharing. Open reviews are broader than numerical ones since one can understand the thoughts and feelings behind them, helping to understand better customer satisfaction. Ratings proved to be more used by consumers when searching for certain products or service information. Otherwise, when comparing products or services, open reviews, including looking at negative words, proved to be more helpful.

The reviews mention two main types of aspects: main service and relational service. The first is product related, referring to the value created by the product/service for the individual. The second encompasses the other services associated with customer service or employee friendliness (Butcher et al., 2003). According to Crosby and Stephens (1987), according to customer evaluation, the leading service is more important than the relation service.

eWOM is electronic word of mouth in which customers share their experiences, views, and opinions on the internet through reviews, commentaries, and social media posts. (Burgess, et al., 2009). eWOM can influence purchase probability: positive comments increase the likelihood of purchases, whereas negative ones decrease (Hsieh et al., 2012). Studies showed that higher reviews and average ratings affect positivity sales (Duan, Gu, & Whinston, 2008). Negative reviews tend to impact customer decision-making more (Mizerski, 1982). This is common because the expectations do not match the product/service. Therefore, customers are more willing to share their dissatisfaction (Harrison-Walker, 2001).

Customers look for eWOM for mainly three reasons: looking for product or service information, the desire to belong to a community, and lastly, for entertainment (Schindler and Bickart, 2005).

There are different reasons behind the motivations for writing online reviews. Three factors have an important role when it comes to writing reviews. The first is altruism, caring, and helping other customers in decision-making. The second one involves financial reasons, getting something in return for publishing content. The homeostatic utility is the third reason to restore balance after a bad experience (Henning-Thurau, 2004). The feeling of belonging to a community is a common motivation to write online reviews (Goes et al., 2016).

2.3.1 Online Reviews in the Tourism Sector

The tourism sector is no exception, and travelers' product and services experience online opinion can impact business performance (Brochado et al., 2019; Dabholkar, 2006). Litvin (2008) references the importance of businesses adopting their models to answer to user-generated content's impact on the industry. Businesses should take advantage of online reviews' influence on customer decisions and use their resources to build online systems where customers can post their feedback (Ye et al., 2011).

It is proven that travelers prefer to read online reviews done by previous travelers than by professional travel platforms (Gretzel et al., 2007). There is evidence that around 98% of travelers checked online reviews while planning travel, and 84% considered that their choices were influenced by it.

Evidence shows that online reviews influence the consumer decision-making process in the lodging sector: a higher number of reviews positively influences hotel awareness and consideration (Vermeulen and Seegers, 2009).

In peer-to-peer accommodations, online reviews have a higher importance than traditional lodging options since micro-entrepreneurs manage accommodations there is no promotion of the properties on mainstream media (Dredge and Gyimóthy, 2015). Besides, it helps to reduce uncertainty and increase trust in choosing Airbnb (Zervas et al., 2015a).

Research has been done to find the underlying factors influencing Airbnb guest satisfaction. Using text mining techniques and a Latent Aspect Rating approach to take a deeper look at Airbnb reviews (2019) based in the city of Los Angeles, found that the five more common aspects referred when it comes to customer experience, including: “communication”, “experience”, “location”, “product/service”, and “value”.

Ding et al. (2021) employed LDA to find the factors of satisfaction and dissatisfaction. The study involved 12 cities from 11 different counties. As most studies overlook the positive experience, positive and negative reviews were separated, and the authors created two models. Accommodation facilities are the primary driver of Airbnb dissatisfaction. Airbnb hosts can also be an inflectional factor for both satisfaction and dissatisfaction.

A study in China found differences in the dimensions affecting consumer satisfaction between foreigners and domestic travelers through Airbnb reviews (Zhang & Fu, 2020). While foreign guests identified recommendation and booking flexibility as topics affecting satisfaction, domestic travelers identified cleanness as an important dimension (Zhang & Fu, 2020).

Sutherland and Kiatkawsin (2020) used the latent Dirichlet allocation method and New York City as a sample to find the determinants of Airbnb guest satisfaction which included “evaluation”, “location”, “accommodation unit”, and “management”. Sutherland, Kiatkawsin, and Kim (2020) also used LDA to do a comparative study between Singapore and Hong Kong to find differences in the reviews, which revealed that even though the number of topics extracted was different, the scope of the factors mentioned in the reviews was similar.

2.4 TEXT MINING

The increasing usage of the internet and online purchases generates large and unstructured data types. Big data can be defined as large datasets generated by various sources like online transactions, internet traffic, and social media (Xiang et al., 2015). Big data has three characteristics: volume, velocity, and variety. Volume refers to the quantity of available data, variety refers to data type, and velocity is the speed at which the data is generated. (Xiang et al., 2015).

With the growth of data available, including the increase of online reviews sharing, companies can have new insights through big data analytics and gain a competitive advantage on the image and opinion that consumers might have over the product or the brand itself rather than the traditional static sources of data (Lau et al., 2005; Davenport, Barth, & Bean, 2012). The insights can help to discover business problems and better understand competitors, customers, and the market (Xiang et al., 2015). In addition, it helps to improve decision-making, for example, by creating predictive customer models based on pattern extraction. (Chen et al., 2014)

The tourism sector is one of the leading regarding the amount of available data that has been able to find preferences and evaluate travelers’ experiences. Although, when it comes to social media data, the methods are still scarce (Xu et al., 2017).

Text mining includes techniques that extract meaningful information and patterns from unextracted text sets (Guerreiro & Rita, 2020; Hung & Zhang, 2012). It is a semi-automatic process of transforming unstructured data into structured forms (Sharda et al., 2014). It is also considered the data analysis of natural-language texts (Aggarwal and Zhai, 2012). The primary text mining process includes document clustering, quantitative text analysis, keyword extraction, sentiment analysis, and topic modeling.

The goal of text mining is the same as text mining but what differs is the type of processed data (Chau et al., 2007). While data mining analyses structured data, text mining seeks to analyze unstructured data - natural language texts - and transform them into numbers (Nasukawa and Yi, 2003).

Using text mining techniques, companies can have more valuable insights than using traditional methods into consumer satisfaction since it gives a broader representation of its experience by analyzing the sentiment behind (Sparks and Browning, 2011)

2.4.1 Topic Modelling

Topic modeling started in the early 80s as a generative probabilistic modeling branch (L. Liu et al., 2016). The methods were developed because of the need to analyze, classify, and summarize increasingly large amounts of data. The types of data using topic modeling include text data and genetic and biochemical sequences, images, videos, and geospatial data (Vayansky & Kumar, 2020).

Topic modeling is a text mining technique that includes different algorithms aiming to reveal, discover and annotate the thematic structure in a collection of documents, including finding hidden semantics and cluster themes as topics (Blei, 2012; Calheiros et al., 2017).

Several topic modeling methods include Latent semantic analysis (the most used), Non-Negative Matrix Factorization, Probabilistic Latent semantic analysis, and Latent Dirichlet Allocation. Different research fields use topic modeling, including bioinformatics, social network analysis, and software engineering (Blei, 2012).

3. METHODOLOGY

3.1 DATA COLLECTION AND SAMPLING

Guests usually share their online opinions and feelings after their stay at Airbnb. There are many ways to collect data from those reviews, including API´s and web scraping. In the present study, the data was collected from InsideAirbnb, a website that organizes and aggregates Airbnb data and provides a snapshot of statistics, ratings, reviews, and listings organized by city (Inside Airbnb: About, n.d.).

The data was extracted and collected in a CSV file that contained reviews since 2009. One dataset was extracted for each city: Paris, Madrid, Rome, Beijing, Hong Kong, and Tokyo, and afterward, grouped by regions.

Table 2 presents the variables contained in the dataset extracted from InsideAirbnb. It includes the Listing ID, Review Date, Reviewer ID, Name, and comments.

Variable	Description
Listing ID	Accommodation unit unique identifier
ID	Review unique identifier
Review Date	Date of review publication
Reviewer Id	Reviewer unique identifier
Name	Reviewer Name
Comments	Review

Table 2 – Explanation of the review´s dataset variables

Since the datasets include reviews not included in the analysis period, a filter was applied to keep only reviews between 2017 and 2021. Furthermore, the analysis will be based exclusively on English reviews. Therefore, the function *detect_language()* from the Python package *TextBlob* was used to detect the language of each review (Textblob, 2021). Afterward, all the reviews that were not written in English were deleted. Table 3 and Table 4 show the number of reviews before and after removing non-English reviews and the date filtering.

	Paris	Rome	Madrid	Europe
Number of Extracted Reviews	1,044,501	1,048,575	702,493	2,795,569
Number of English Reviews (2017-2021)	466,724	537,876	244,670	1,249,270

Table 3 – Total number of reviews in Europe

	Tokyo	Bangkok	Hong Kong	East Asia
Number of Extracted Reviews	262,591	274,012	88,484	312,991
Number of English Reviews (2017-2021)	128,630	137,857	46,536	313,023

Table 4– Total number of reviews in East Asia

For a general overview of the Airbnb listings in each city and region, a dataset including listing information from each city was also extracted. Again, it contained information from listings that were no longer active after 2017, so a filter was applied so only listings in which the last review was after 2017 would belong to the dataset. Table 5 describes the variables listed in the Airbnb accommodation listing dataset.

Variable	Description
Listing ID	Airbnb's unique identifier for the review
Host ID	Airbnb's unique identifier for the host
Neighborhood	The neighborhood is geocoded using the latitude and longitude against the neighborhood as defined by open or public digital shapefiles.
Room type	Airbnb home category: 3 Entire place 4 Private room 5 Shared room 6 Hotel room
Price	Average price, in local currency, per listing
Number of Reviews	Average number of reviews per listing
Last Review	Date of the last review per listing
Availability 365	The availability of the listing 365 days in the future as determined by the calendar. Note that a listing may not be available because it has been booked by a guest or blocked by the host.

Table 5 - Explanation of the listing´s dataset variables

3.2 Research Framework

Figure 2 illustrates the stages of the approach followed by this study to analyze and compare the reviews (unstructured text) related to the customer Airbnb experience. It focuses on the procedures and steps to analyze to identify unknown information and factors that influence customer satisfaction.

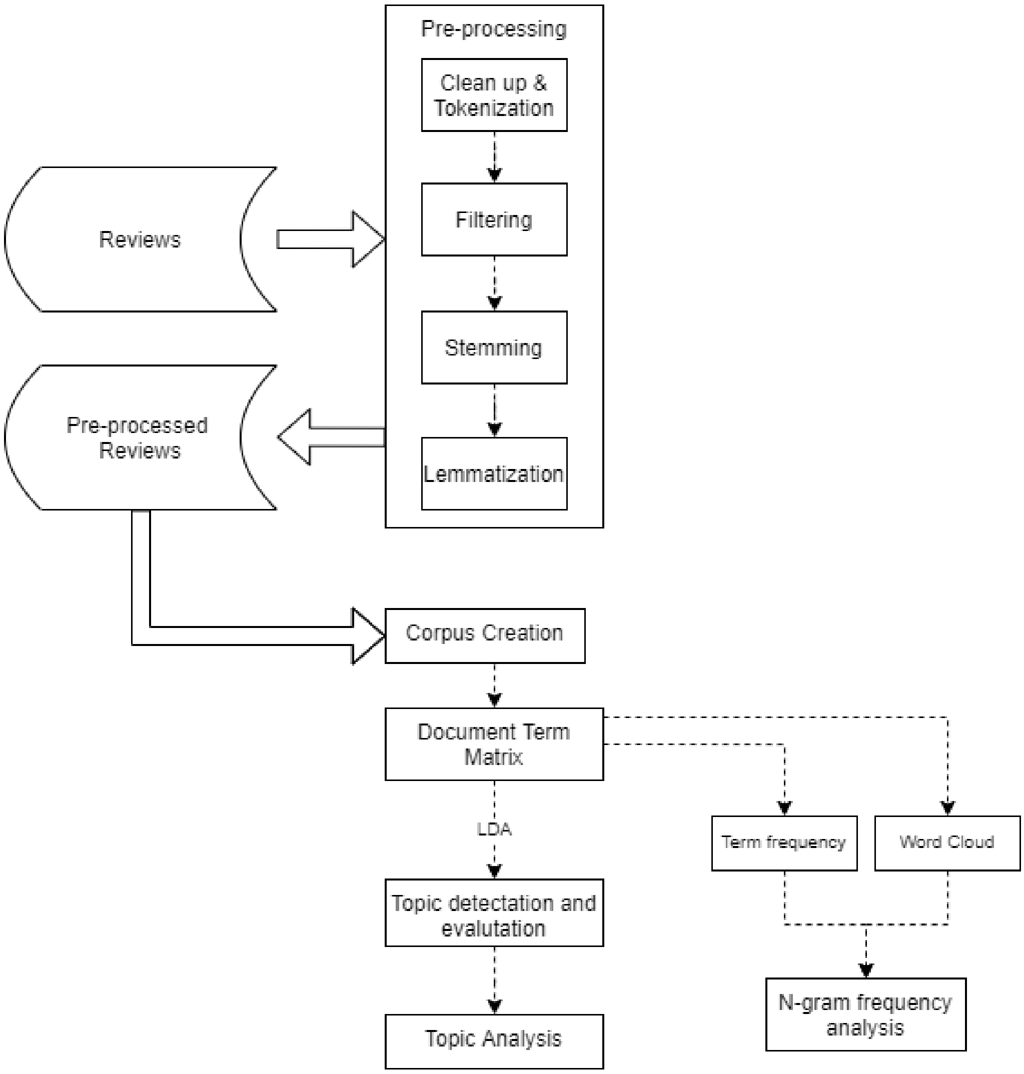

Figure 2 – Research Framework

Textual data is, by nature, unstructured; therefore, pre-processing and cleaning the data is extremely important to eliminate language-dependent factors. Pre-processing text data aim to reduce ambiguity and dimensionality of the data – remove complexities (Kwale, 2013). By removing high-dimension words (frequent words that produce unimportant high dimensions), the text becomes simpler and easier to structure, keeping relevant information (Patodkar & I.R, 2016). The Python package NLTK, among other functions, was used to perform the pre-processing (NLTK: Natural Language Toolkit, n.d.).

Cleaning the text includes removing numbers, spaces, and other unwanted characters like punctuation marks, HTML tags, special characters, and line breaks and normalizing to lowercase. Tokenization is the method that breaks the text into words, sentences, or other elements – tokens.

The most used approach, and the one used in this research, is to split the text into blank spaces – words (Kwale, 2013).

The goal of filtering it is to decrease the removing noise by reducing dictionary dimensionality by removing not important words. This step includes removing stop words and eliminating prepositions, interjections, conjunctions, and auxiliary verbs. These words are highly frequent, although they do not represent important information. This way, an analysis of the core words is performed (M.Kwale, 2013). Stop words were eliminated using the *stop_words* package (Stop-words, 2018). Although because this package does not contain all stop words, a new list containing additional words was created. Some of these words are 'even', 'etc.', 'also', and 'everything'.

Stemming is the process of reducing a word from its inflected form to its base form, maintaining a similar meaning, and simplifying the text. It includes eliminating plurals and other affixes (Kwale, 2013). The stemming was done using the *nltk.stem* package (NLTK :: Natural Language Toolkit, n.d.).

Lemmatization reduces the words to their morphological origin. This process reduces the number of unique words in the data, making the data less fragmented and easier to work with (Prabhakaran, 2018). The process was made with the help of the *nltk.stem* package (NLTK :: Natural Language Toolkit, n.d.).

After the process of cleaning the data, however, some mistakes were found. Therefore, new functions were created to eliminate those mistakes. After stemming, the word "us" was among the most frequent words. Hence, for analysis purposes, it was considered a stop-word and eliminated. In addition, after the data cleaning, the possessive case of the noun "*'s*" was presented as a word and among the most frequent words. Therefore, a function was created to eliminate these cases. Since "n't" was also one of the most frequent terms, it was investigated that it came from verb contraction. In this case, a replacement method was chosen. In this case, "isn’t" was replaced by "is not", "can’t" was replaced by "cannot", and "don’t" was replaced by "do not". Other relevant replacements were made, including "mtr" for "metro" and common mistakes found in the dataset as "minuteutes" for "minute".

In sum, data pre-processing can be split into three parts. The first one contained the first cleaning of the reviews, including capital letters, blank spaces, and punctuation. The second one ('Words Lemmatized') excluded stop words, tokenized, and lemmatized words for further analysis. The third one, used for analysis, did a steaming process, so the word is in the original form. Besides, all the extra procedures to correct review errors were made in this final step. Table 6 shows an example of the data-cleaning process.

Cleaned	Words Lemmatized	Final
it's not that hard to find the place but it's quite noisy and the room is narrow it's not polite but i have to say that i will not come back again	["'s", 'hard', 'find', 'place', "'s", 'quite', 'noisy', 'room', 'narrow', "'s", 'polite', 'say', 'come', 'back']	['hard', 'find', 'place', 'quit', 'noisi', 'room', 'narrow', 'polit','say', 'come', 'back']

Table 6 - Example of a processed review

The document term matrix is a mathematical matrix in which the cell's value is the frequency of terms that occur in a collection of documents. This matrix was used to aid and perform frequency analysis. A frequency table and a cloud of words were obtained to help to visualize the most occurrent terms giving a better understanding of the phenomena being studied (M.Kwale, 2013). In order to find the most common words in each dataset, the *nltk.probability* python was used. The frequency analysis was crucial in detecting errors made in the data pre-processing stage.

The word cloud was used for the visual representation of the most frequent words, in which the size of the word stands for the frequency – the bigger, the more frequent. The word cloud was obtained using the *wordcloud* package. Also, an n-gram analysis was performed to have a better context of the analysis, obtaining the top sequence of consecutive words.

3.3 Topic Analysis

In machine learning and natural language processing, topic modeling refers to methods that organize, understand, search, and summarize large amounts of data (Mo, n.d.).

For the topic modeling, Latent Dirichlet Allocation (LDA) was applied using the *Gensim* package. Latent Dirichlet Allocation is a generative probabilistic model based on the Bayesian probability principle applied to find hidden "topics" in the text, and group homogeneous topics (Blei, 2012).

The LDA assumes that each document comprises mixtures of latent topics, and each topic is described by a distribution over words (Blei et al., 2003).

The algorithm can be summarized in three steps:

1. Define parameters: the number of documents, the number of topics, and the number of iterations.
2. Assign each word to a particular topic randomly according to a Dirichlet distribution.
3. Repeat the steps for all words in the corpus.

As previously mentioned, one of the LDA model's downsides is that the number of topics (k) must be chosen *a priori*. Therefore, as the number of topics must be defined a priori, the coherence score is chosen to evaluate it. The goal is to have a high coherence score with the smaller number of topics possible, so the chances of the terms repeating themselves among the topics are lower. A higher coherence score means that the topic is more human interpretable and how similar the words are among each topic.

Figure 3 and figure 4 show the coherence score versus the number of topics. Analyzing Figure 3, the highest coherence score is approximately 0.6 when the number of topics equals four. This means that four topics will be chosen for the LDA model in Europe.

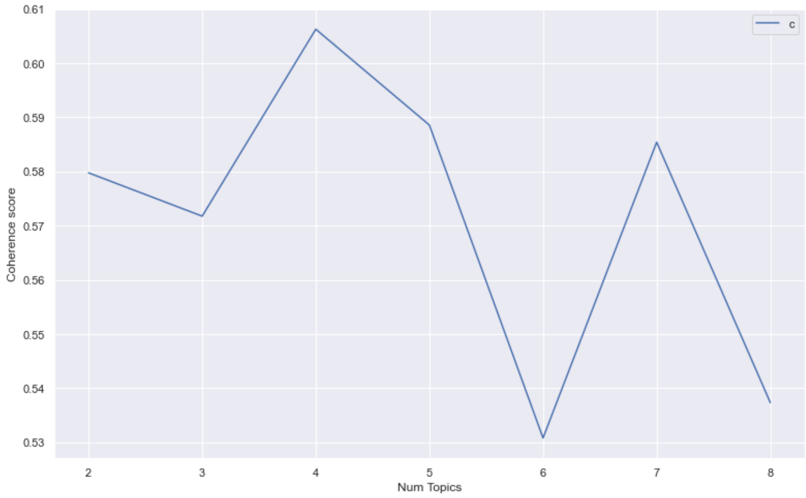

Figure 3 – Coherence Score (a) - Europe

Concerning East Asia, by analyzing figure 4, the highest coherence has an approximate value of 0.61, which happens when the number of topics is equal to four. Therefore, as in Europe, the East Asian LDA model will have four topics.

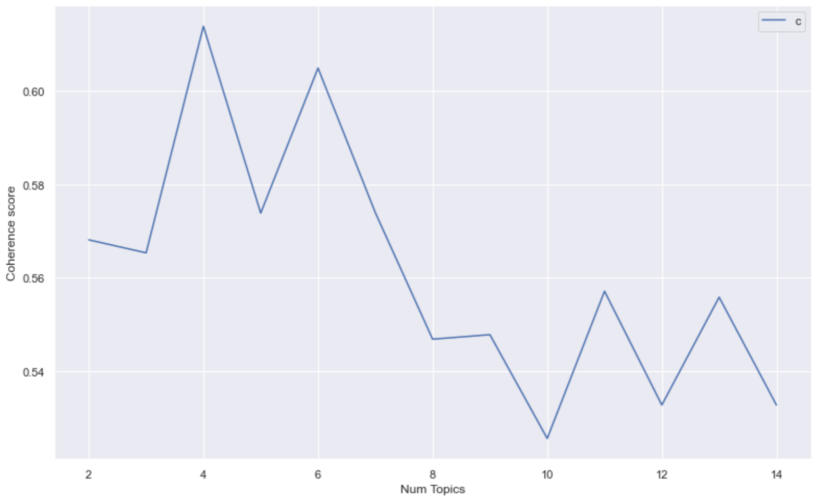

Figure 3 - Coherence Score (b) - East Asia

Therefore, each model will have four topics. The coherence score of the East Asian model is slightly superior to the European model, with 0.613 and 0.66, accordingly.

Following the LDA process, the guests' most valued topics of accommodation experience will be extracted.

4. RESULTS AND DISCUSSION

4.1. SAMPLE CHARACTERIZATION

The dataset containing guests' reviews is divided into two regions: Europe and East Asia, but firstly an analysis of Airbnb listings per city and region will be performed.

Table 7 describes the Airbnb listing by accommodation type, diving into cities and regions. In every city, an entire apartment is the most common form of short-time rental. The city that has the highest percentage of entire home rentals is Paris (84.63%). On the other side, Hong Kong has the lowest percentage of rentals from entire homes (46.12%). Comparing the regions, Europe has the highest percentage of listing with entire homes, 76.66%, compared to 64.87% from East Asia.

Regarding private rooms, Hong Kong has the highest percentage of private rooms on its listing (46,05%). On the other hand, Paris has the smallest percentage (12.71%). East Asia has a higher percentage of private rooms, 29.46%, against 20.41% in Europe. Considering hotel rooms, Hong Kong has the highest percentage of hotel rooms rented by Airbnb (5.23%). East Asia has a higher percentage of hotel rooms in their listing than Europe (3.12% and 2.28%, respectively). Regarding shared rooms, Tokyo has the highest percentage of shared rooms in their listing (3.10%), while Rome has the smallest (0.47%). East Asia has a greater percentage of shared rooms than Europe, 2.55% and 0.65%, respectively.

Type of Airbnb	Paris	Rome	Madrid	Beijing	Tokyo	Hong Kong	Europe	East Asia
Entire home/apt (%)	84.63	69.28	65.81	63.67	71.80	46.12	76.66	64.87
Private room (%)	12.71	26.55	32.29	35.07	21.61	46.05	20.41	29.46
Hotel room (%)	2.09	3.71	1.13	0.00	3.49	5.23	2.28	3.12
Shared room (%)	0.57	0.47	0.78	1.26	3.10	2.61	0.65	2.55

Table 7 – Airbnb type of accommodation per city

As presented in Table 8, all European cities in the sample have more listings than East Asian cities. Overall, Europe has approximately 4.8 times more listings than East Asia. On average, the city with the lowest rental prices is Madrid, followed by Rome and Hong Kong. On average, Europe has a lower price per night than East Asia, which is interesting since Europe has a higher percentage of entire homes than East Asia. Rome, followed by Madrid and Tokyo, have the highest average number of reviews per listing. Madrid, Rome, and Tokyo have the highest average monthly reviews per listing. Compared to East Asia, Europe has a higher number of average reviews per listing and per listing per month.

Regarding the availability for the next 365 days, on average, the city with the lowest availability is Paris – with approximately 111 out of 365 days available, followed by Tokyo, with approximately 115 out of 365 days available. On the other hand, Beijing and Rome have the highest availability for the next 365 days (approximately 224 and 210 days). Europe has more demand since it has fewer days available per listing than East Asia.

	Paris	**Rome**	**Madrid**	**Beijing**	**Tokyo**	**Hong Kong**	**Europe**	**East Asia**
Number of listings	35,887	19,090	13,391	3,102	8,231	2,832	68,368	14,165
Most Frequent Neighborhood	Buttes-Montmartre	I Centro Storico	Embajadores	怀柔区 / Huairou	Shinjuku Ku	Yau Tsim Mong	I Centro Storico	Shinjuku Ku
Average Price (in Euros)	124.68	109.95	107,89	244.023	129.12	117.22	117,28	151.95
Average Number of Reviews	29.73	55.20	47.96	7.98	47.80	37.62	40.41	29.07
Average Reviews per month	0.87	1.19	1.37	0.57	1.07	1.12	1.06	0.90
Availability	110.96	210.02	151.44	224.34	115.45	138.09	146.55	174.54

Table 8 – Listing´s summary statistics per city

4.2 N-GRAM ANALYSIS

Table 9 indicates the number of occurrences for each of the ten most used terms achieved because of the text mining procedure in Europe and Asia. In Europe, the top five most frequent terms include "apart", "great", "locat", "place", and "host". In Asia, the top five most frequent terms in Airbnb reviews are "place "stay", "great", "locat", and "clean".

For both regions, the importance of the house experience ("apart, "place", "clean") and location ("locat") seem to be the most mentioned among reviews, confirming what previous studies showed. The word "host", one of the Airbnb most unique points, is also on most frequent terms showing the importance the relation with the host has a significant value during the stay. Besides, guests mentioned only positive terms, including adjectives such as "great", "nice," and "perfect", demonstrating positive feelings towards most reviews.

Europe		East Asia	
Term	**Frequency**	**Term**	**Frequency**
apart	755,487	place	199,056
great	723,917	stay	165,535
locat	717,634	great	143,884
place	650,996	locat	130,303
host	386,144	clean	108,535
clean	366,596	host	89,988
recommend	353,162	nice	85,595
nice	324,601	good	83,886
walk	321,268	apart	79,408
good	249,42	room	78,822

Table 9 – Top 10 most frequent terms

The word cloud, figure 4, highlights the most frequent words used by guests in the reviews. It gives an overall and visual point of view on the most mentioned topics in the Airbnb reviews.

Figure 4 shows the most frequently used words in Europe and East Asia. Besides the ones mentioned previously, guests mainly mention positive adjectives in Europe. It includes adjectives regarding the apartment, “comfort”, “clean”, and “spacious”. Besides, surrounding services are also mentioned, such as “bar”, “restaurant”, and “supermarket”. Guests highly mentioned parts of the house and house facilities in their reviews, such as “bed” and “room”. Lastly, location is also mentioned, including the name of the cities used for the analysis, “Paris”, “Rome”, and “Madrid”, but also characterizes public transportation - “metro”, “central”, and “close”. Regarding the neighborhood, guests use words such as “area”, “street”, and “quiet”.

The most frequent words used in Asia, figure 4 (b), exhibited mostly positive adjectives as well, including “perfect”, “spacious”, “wonder” and “beauti”. Surrounding services are also mentioned, such as “mall”, “restaurant”, and. House facilities are also pointed out including “shower”, “bed”, and “toilet”. Last of all, location is highly pointed out, the city where the guest stayed like “tokyo”, “bangkok”, and “hongkong”, the neighbourhood using words such as “street” and “quiet”, accessibility of public transportation, such as “metro”, “train”, “station”.

Figure 4– Word Cloud –(a) Europe; (b) East Asia

Regarding the top 2-gram terms, observed in Figure 5, it is clear the emphases that the Airbnb clients put on location since “great location” is the most frequent term for Europe and Asia. Besides, location is again clearly a factor of importance when it comes to guest satisfaction since it is mentioned in four out of the ten most frequent 2-gram words in Europe (“great location”, “walking distance”, “minute walk”, “metro station”).

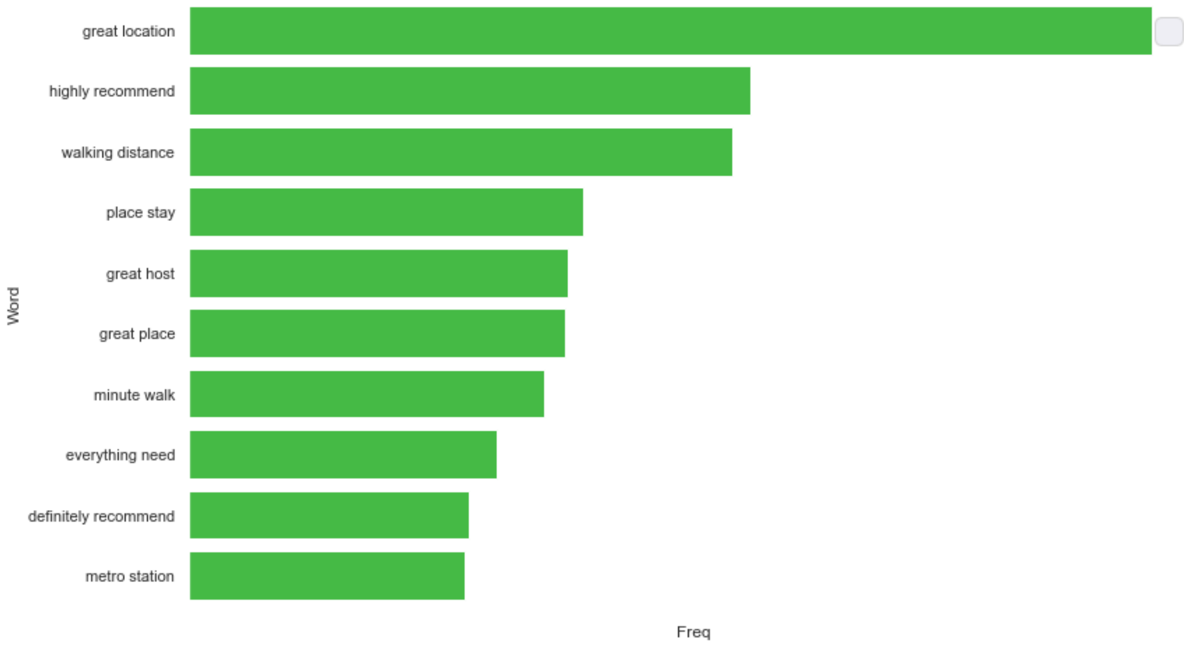

Figure 5 - Bi-gram Term Frequency (a) - Europe

The same applies to East Asia as six out of the ten most mentioned 2-gram term, including topics such as “great location”, “good location”, “train station”, “minute walk”, “walking distance”, showing once again the importance of the apartment location on the guest reviews.

In both figures, the recommendation of the place of visit is also directly mentioned in the reviews. Examples include “highly recommend” and “definitely recommend”. This point is more mentioned in the Europe reviews than the Asia guest reviews since it is not only mentioned on two out of ten topics but also because the second most mentioned topic, while in Asia, it is the fourth. On the other hand, Asia mentions cleaning in the reviews showing the importance of that feature to the guests staying at East Asian places. Once more, only positive terms are among the most frequent terms, including the facilities of the house provided ("everything need", "great place", "great place").

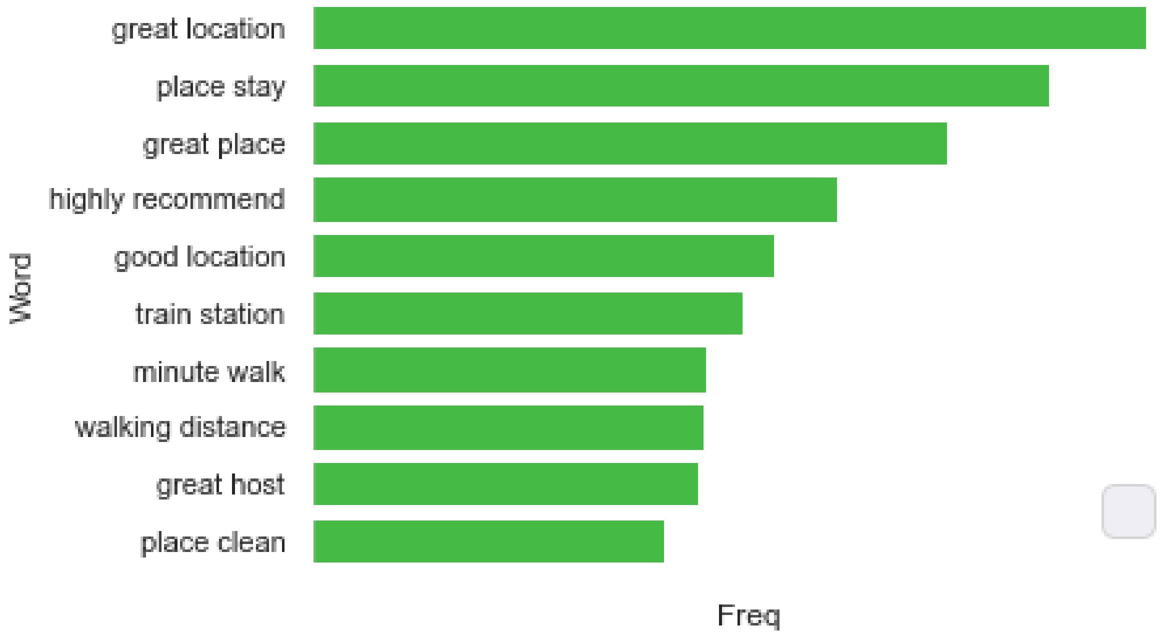

Figure 5 – Bi-gram Term Frequency (b) - East Asia

By analyzing the tri-gram frequency, figure 6, it is possible to conclude that, once again, the most relevant topics are regarding the location of the Airbnb, and the overall opinion of the stay, not only for Europe but also for East Asia.

In Europe, the primary focus is the walking distance and the apartment location ("within walking distance", "apartment great location", "place great location"). Besides, expressions mentioned come from an automated posting that happens when the host cancels the place. The original review is "The host canceled this reservation the day before arrival. This is an automated posting". This is the only negative point mentioned in the 3-gram analysis of the European Airbnb reviews.

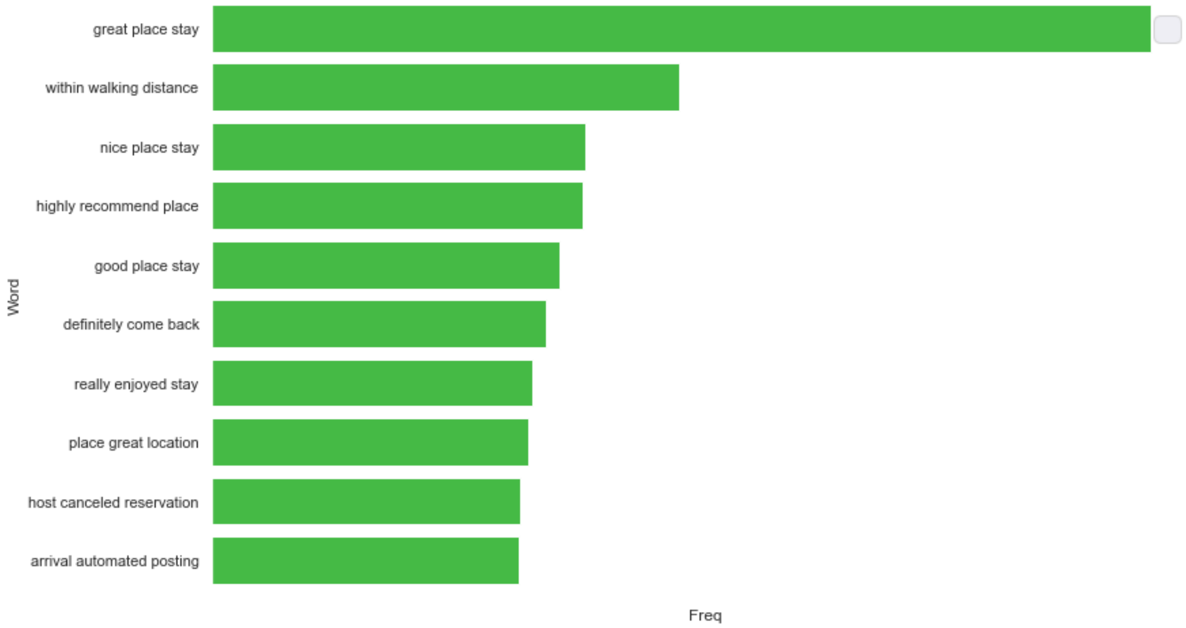

Figure 6 - Tri-gram Term Frequency (a) - Europe

Regarding East Asia, the reviews also mention the satisfaction of the guest regarding their experience at the Airbnb (“great place stay”, “nice place stay”, “good place stay”, “nice place stay”). Once more, the location quality is an important point taken into consideration by the guests (“within walking distance”, “place great location”). In Figure 6 (b), two points are not mentioned in the tri-gram term frequency in Europe. The first is the guest's direct recommendation to future visitors (“highly recommend place”). The second is the intention to re-visit the Airbnb accommodation, indicating great satisfaction with the service provided (“definitely come back”).

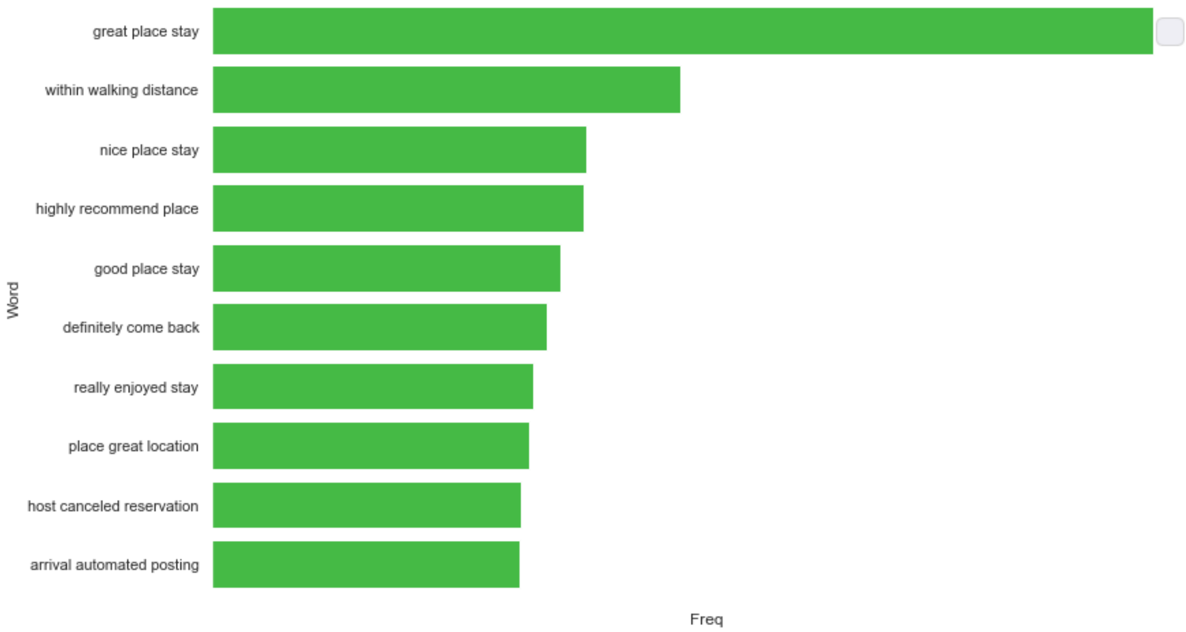

Figure 6 - Tri-gram Term Frequency (b) – East Asia

4.3 TOPIC ANALYSIS

An LDA analysis was performed to find the determinants of Airbnb guest satisfaction. This algorithm, as explained before, is a technique that allows classifying the documents into themes. Four topics were used for the analysis – as it was the best number according to the coherence score. The table below shows the terms that are most common within each topic. Therefore, each topic was named to represent a category that can influence guest satisfaction. Table 10 shows the topics identified that influence guest satisfaction in Europe and East Asia.

In Europe, the first topic was named "accommodation" because it contains general information regarding the apartment where the guest stayed. The most common words in this topic include “apart”, “great”, “locat”, and “place” since they have the highest betas. Besides the judgment of the apartment and the location, the host and cleanness of the apartment are also dimensions that affect guest satisfaction, represented by this topic. Only positive adjectives are allocated to this topic, including great and nice.

The second topic is called “location" since it incorporates aspects of guest satisfaction regarding distances, surrounding services, public transportation, and the neighbourhood. The word that is more represented in this topic – with higher betas – include “walking”, “restauran”, and “minut”. It includes the importance of the position of the apartment's location regarding public transportation, for example, the metro, and the importance of walking distance. It includes services available nearby as restaurants and shops. Ultimately, the fact that the word minute is part of this topic emphasizes the importance of all these previously mentioned factors being within a short distance of the apartment.

The third topic, referred to as the “negative aspects,” consists of subjects that had complications since it includes words like issue and problem. The words with the highest betas are “arrive”, “day”, and “work”. Considering the words with the highest betas and the previous analysis, especially the n-gram analysis, it is possible to conclude that some problems were identified to occur at the check-in. Other words related to check-in are “late” and “wait,” which one can assume that one complaint that guests make is related to check-in complications regarding waiting times. Besides, other issues can be found in the connection services – Wi-Fi. The increasingly flexible working, working from home, and digital nomads reflect the importance of having a stable connection so Airbnb can be seen as a reliable option for those seeking those services. Further, the word reservation shows that guests also complain about the reservation process.

The fourth topic is “amenities” since it describes different features of Airbnb or equipment used or provided for a specific purpose. This shows the importance of the facilities on customer satisfaction, including the rooms and the resources and furniture provided by Airbnb to the guests. The words with the highest betas are “room”, “night”, and “bed”. The parts of the house included in the topic include rooms, bathroom, and kitchen. Besides, it includes the word bed which can be analyzed as a determinant factor taken into consideration when it comes to satisfaction since it affects the guest’s sleeping quality. The adjective small is also part of the most relevant words on this topic, which can have two interpretations depending on what the guest expects. Small can be considered cozy, convenient, and adequate for the guests' needs. On the other hand, it can be seen as the expectations do not match what the guest is confronted with, and the space is not enough for the total guests and does not match their needs.

Regarding East Asia, as previously stated, the coherence score showed that the best result happens when the number of topics is four. Therefore, the consumer satisfaction factors in East Asia will be divided into four categories.

The first topic is designated as “location” because it contains keywords regarding surrounding services, public transportation, time, distance, and quality provided. The words with the highest betas are “good”, “apart”, and “conveni”. This topic includes location factors that guests in East Asia consider relevant. It includes the importance of the walking distance of the apartment to a certain destination. Plus, the convenience of the apartment location is also a key factor and can be related to other words mentioned in the topic, such as a restaurant (services nearby) and metro (public transportation). Other words like easy, minute, and close emphasize the importance of a good apartment location, impacting guest satisfaction.

The second topic describes different adjectives regarding the overall stay at an Airbnb. Therefore, this topic was called “accommodation”. This topic contains themes about the apartment (“place”) but also the quality of the visit. It is visible that it contains positive overviews for the apartment since it contains adjectives and verbs that describe Airbnb positively - “great”, “nice”, “love”, and “recommend”. It mentions the host, and from that group of words, “help” is related to the host. Therefore, it shows the importance of the host and how they can be helpful in different areas during a guest's stay. It also includes clean as one of the words presented in this topic, showing the importance of it when it comes to qualities that one describes while visiting an Airbnb.

The third topic was named “experience” because it includes words that make Airbnb a unique and distinct lodging option over the traditional offers. It includes words that go along with the Airbnb slogan “feels like home”. These words include “home”, “airbnb”, “experience”. Besides this topic, it also includes activities one does during a holiday, such as “beach” or “go”.

The fourth topic - “amenities” – includes categories that include parts of an apartment and facilities that one can use while staying at an Airbnb place. The words with a higher beta are “room”, “bed”, and “kitchen”. It includes room, kitchen, and bathroom as part of the house. Besides, this topic also includes facilities provided by the host as a bed or utensils to cook. In addition, the word used describes the tools provided by Airbnb for the guests. As stated in topic 4, in Europe, the word " small " can have two meanings, a positive and a negative. The positive one occurred when one expected the place to be small and it is convenient according to the guest's needs. The negative one happens when the guest expects a bigger place, and the limited space can perturb their needs.

Europe				East Asia			
Accommodation	**Location**	**Negative Aspects**	**Amenities**	**Location**	**Accommodation**	**Experience**	**Amenities**
apart	walk	arriv	room	apart	stay	hous	bed
great	restaur	day	night	conveni	great	beach	kitchen
locat	minut	work	bed	walk	locat	back	small
stay	madrid	late	small	easi	clean	experi	use
place	area	wifi	bathroom	locat	nice	come	water
host	metro	issu	kitchen	close	host	home	work
clean	distanc	problem	use	restaur	recommend	day	bathroom
madrid	shop	reserve	peopl	metro	love	go	cook
recommend	home	wait	live	minut	help	airbnb	overal

Table 10 – Topics affecting guest satisfaction – Europe and East Asia

5. CONCLUSIONS

The sharing economy platform emerged as a disruptive model that has revolutionized markets by changing tourist consumption patterns, including the tourist lodging sector. Airbnb is the biggest player in the peer-to-peer tourism sector. The reviews written by the guests are important for the accommodation's reliability, helping other travelers' decision-making. In addition, it is a valuable source of information regarding customer satisfaction and preferences.

To extract hidden dimensions influencing satisfaction, firstly, the data was pre-processed and then analyzed - frequency analysis, n-gram analysis, and topic modeling were the methods used. The data was extracted from Inside Airbnb, and each region dataset contains aggregated English reviews between 2017 and 2021 from 3 different cities. The cities chosen to represent each region were based number of international tourist arrivals in 2019.

Frequency analysis showed that the reviews tend to be skewed positively and contain mostly positive aspects. The fact that only positive adjectives were among the most frequent terms supports what has been studied in the literature. Location, the apartment unit, and the host were the three main subjects mentioned in the topmost frequent words. Once more, 2-gram and 3-gram analyses showed the importance of location and having public transportation nearby since it is a point that the guests emphasize. There was found that the host, the wish to re-visit the apartment, and the recommendation to the next guests are also among the most frequent topics that the customer writes about in the reviews. In the East Asian reviews was observed that cleaning was a point mentioned in the n-gram analysis that was not in the European reviews.

This thesis had three goals: i) describe Airbnb features property of each city; ii) extract the topics affecting guest satisfaction; and iii) compare the determinants of satisfaction between regions.

The first goal consisted of analyzing the European and East Asian Airbnb listings to obtain information about the accommodations and their differences which helps to have a deeper knowledge of the regions, including the type of accommodation offered, price ranges, and the number of properties available. Regarding the type of offer, Airbnb in Europe and East Asia have a similar distribution. The Airbnb entire property is the most popular among both regions, with a slightly superior percentage in Europe. On the other hand, the less common in both regions is the shared room. Moreover, the prices, on average, are higher in Europe compared to East Asia. Europe also has a higher number of properties available and a higher Airbnb offer. This is interesting information since the European cities considered in the dataset have a considerably lower population than the East Asian cities, showing that Airbnb is more developed in the European market and that there is an opportunity in the East Asian market. Although, this high popularity in Europe has affected the housing market, including the increase in rental prices. Paris, for example, in 2017 created a law that limited the short-time rental to 120 nights per year (Coffey, 2017). Therefore, the property owners would have to analyze better their entry into the short-time rental market.

The second goal of this study aims to understand the dimensions affected by the guest regarding satisfaction by region. For each region, four topics were extracted. Previous studies mentioned in the literature review that likewise used text mining techniques show that "communication", "experience", "location", "product/service", "value", "cleanliness", and "accommodation unit" are the main determinants of satisfaction. The extracted factors influencing guest satisfaction found in European cities were "accommodation", "location", "negative aspects", and "amenities". The location of the accommodation, neighborhood, and surrounding services, including public transportation, was proven by this study as a crucial factor in consumer satisfaction. The place of stay “accommodation” topic, which includes the accommodation, and its condition and qualities, confirms the previous studies showing that the service and the accommodation unit are principal factors influencing guest satisfaction. The amenities can also be proven by previous studies considering that it includes factors of the accommodation unit and service since it is the house extras like a bed or cooking utensils. Although not accordingly to previous studies, one new factor was found – negative comments. This factor includes points the guest did not enjoy during their stay at the Airbnb.

Interestingly, one factor was dedicated to this topic since Airbnb reviews tend to be mostly positive, therefore also the lack of negative topics. Regarding the East Asian factors - “location”, “accommodation”, “experience”, and “amenities”. Location, as stated, is once again proven to be a crucial factor in defining customer satisfaction. Amenities are also proved by previous research to be one of the influential matters impacting consumer satisfaction. In this study, the factor of experience was shown to be also significant regarding customer satisfaction. This topic includes the Airbnb branding strategy of giving the guest a feeling of belonging and the experience of living like a local, as well as other experiences regarding peer-to-peer accommodation. It is the differential factor from traditional lodging options.

Regarding the accommodation topic, this factor incorporates different qualities from different features of the Airbnb overall service/experience. It sustains what previous studies have shown. Price or value for money is considered one key factor when researching the underlying factors of guest satisfaction. Although, in this study, there is no evidence that the price is one of the decisive components of consumer satisfaction as there is no reference to it during the analysis performed.

The third goal of this study is to compare the determinants influencing guest satisfaction in Europe and East Asia. Comparing the regions, three out of four topics were similar – overall opinion, location, and amenities. This indicates that there are no significant differences in the factors influencing customer satisfaction. The location topics include, in both regions, the importance of walking distance from the accommodation. In addition, both topics incorporate the importance of having surrounding service services and public transportation next to the apartment. Although, in Asia, there is a bigger emphasis on convenience than the type of services around the apartment. The amenities were the second topic considered valuable for the guest staying in Europe and East Asia. As stated, this topic includes furniture and accommodation characteristics. Both topics contain the importance of the different parts of the house, including the room, the bathroom, and the kitchen.

Regarding furniture, the bed was found to be a key factor in customer satisfaction, implying that the sleeping quality is very important for the guests in both regions. The third similar topic, accommodation, includes important factors influencing guest satisfaction. Besides the apartment, the host, location, and cleanliness are also mentioned. Since the location per se has its own topic, host assistance is important for the guest. Besides, the cleanliness of the apartment is a point considered in both regions.

Regarding the differences, Europe has a topic dedicated to the negative aspects of the guest stay, while East Asia has one topic dedicated to the unique experience that Airbnb can give - “living like a local”. As stated, the topic extracted from the European reviews, negative aspects, includes issues felt by the guests during their stay. It includes problems with the checking, including waiting times. Another problem was the wi-fi connection, which can be a deal breaker for guests, especially remotely working. The growth of digital nomads is proof of that. Evidence shows that digital nomads rely on Airbnb because most of their accommodation is booked through that platform (Hannonen, 2020d).

Furthermore, East Asia has an exclusive topic regarding the Airbnb exclusive experience, the feeling of "home away from home" as one of their slogans indicates. Therefore, there is a significant difference between these two topics. One discusses the negative aspects of being in an Airbnb, bringing up disadvantages regarding traditional lodging offers. While the topic of East Asia brings the value added by Airbnb and its unique experience compared to traditional competitors.

Considering information from the sampling analysis, it is known that Airbnb in Europe has higher prices and more offer compared to the East Asian accommodation units. Consequently, it is expected that the customers will also be more demanding. On the other hand, the East Asian Airbnb market has fewer offerings compared to the European. Travelers looking for peer-to-peer accommodation belong more to a niche market, therefore valuing the experience of Airbnb and its advantages compared to the traditional lodging market.

This information shows that East Asia has the potential to grow, not only in terms of the number of units that Airbnb can have but also in market share. Besides, a new tendency has been observed in the Airbnb business market in Europe. There has been a shift in the business model from consumer-to-consumer to business-to-consumer with the professionalization and commercialization of the Airbnb market (Demir & Emekli, 2021i). This supports the study's evidence that differentiator factors influence guest satisfaction between Europe and East Asia. As previously stated, satisfaction depends on the expected performance of the product/service (Oliver, 1980), and since there is a higher professionalization of Airbnb in the European market, guests expect the accommodation to perform there. The other side of the professionalization of Airbnb in Europe is the loss of Airbnb, the feeling of home away from home, or the feeling of living like a local and being welcomed by the host.

Theoretically, this study helped advance the state-of-the-art by finding attributes that contribute to satisfaction are influenced by micro-level (house, location, host friendliness) and macro-level (activities around, region, and cultural area), besides breaking the sample into regions with climatic and cultural differences. The insights provided can be useful to aid traditional players and hosts to understand what the customer value in their competitors to learn and differentiate by leveraging their strengths against the factors influencing the customer experience. Airbnb, as a company, should analyze and strategy its policies regionally since guest expectations and demands differ. For the hosts, this study could help to identify priorities to improve their accommodation to increase satisfaction, including facilities that guests value the most, including bed and wi-fi connection. Besides, hosts can use the insight gained from this research to better target their listings to customers who might be most satisfied and boost the host's online reputation. Competing players can use the information to strategize how to leverage the customer topics of interest, improving their strengths and providing better customer service.

As a limitation of this study, it can be pointed the fact that only English reviews were analyzed. Therefore, there is a loss of information, especially from domestic guests, and that is important to evaluate the topics influencing guest satisfaction. Besides, only three cities were used as a sample of each region, Europe and East Asia; therefore, there might be a lack of generalization of results.

There are opportunities to extend this research by including more cities in the sample to reach further generalized results. Moreover, future research can include other regions and non-English reviews to compare guest satisfaction drivers and their variations according to the region of stay and domestic vs. international guests. Since most of the studies are also done using InsideAirbnb data, which contains data from big cities, there is a lack of information regarding satisfaction in smaller cities. Thus, future studies could analyze and compare which guest satisfaction drivers differ between big and small cities. Regarding the period of analysis, this study included reviews between 2017 and 2021, therefore, the results are a picture of the overall period. Although the perception and expectations of Airbnb guests have been changing since there has been an increase in the professionalization of businesses, the future analysis could break the reviews into periods of time to discover if the drivers of guest satisfaction have changed over time. A new algorithm approach could also be used since LDA is known for working better with longer documents (Mazarura et al., 2020).

To conclude, based on this study, comparing Europe and East Asia, there is evidence of a slight difference in the determinants of guest satisfaction according to the region of stay. Some factors do not depend on the region of stay as for location and accommodation quality. Nevertheless, since the Airbnb market differs, guest expectation also varies; therefore, the factors influencing satisfaction can vary according to the region.

REFERENCES

Aggarwal, C. C., & Zhai, C. (2012). A Survey of Text Classification Algorithms. *Mining Text Data*, 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6

Airbnb Statistics: Airbnb Stats For Your Market [2022]. (2021). All the Rooms. https://www.alltherooms.com/analytics/airbnb-statistics/

Airbnb Trust & Safety - Your safety is our priority. (n.d.). Airbnb. https://www.airbnb.com/trust?locale=en

Akan, P. (1995). Dimensions of service quality: a study in Istanbul. *Managing Service Quality: An International Journal*, *5*(6), 39–43. https://doi.org/10.1108/09604529510796575

Akbulut, O., & Ekin, Y. (2019). An Assessment of Culture Tourism in Europe Destination. *International Journal of Contemporary Tourism Research*, 43–49. https://doi.org/10.30625/ijctr.549471

Akehurst, G. (2008). User generated content: the use of blogs for tourism organisations and tourism consumers. *Service Business*, *3*(1), 51–61. https://doi.org/10.1007/s11628-008-0054-2

Altimeter Research Theme: Digital Economies. (2013, June). *The Collaborative Economy: Products, services, and market relationships have changed as sharing startups impact business models. To avoid disruption, companies must adopt the Collaborative Economy Value Chain*. http://www.collaboriamo.org/media/2014/04/collabecon-draft16-130531132802-phpapp02-2.pdf

Bardhi, F., & Eckhardt, G. M. (2012). Access-Based Consumption: The Case of Car Sharing: Table 1. *Journal of Consumer Research*, *39*(4), 881–898. https://doi.org/10.1086/666376

Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, *67*(8), 1595–1600. https://doi.org/10.1016/j.jbusres.2013.10.001

Bellotti, V., Ambard, A., Turner, D., Gossmann, C., Demkova, K., & Carroll, J. M. (2015). A Muddle of Models of Motivation for Using Peer-to-Peer Economy Systems. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2702123.2702272

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. -.

Botsman, R. (2015, May 27). *Defining The Sharing Economy: What Is Collaborative Consumption–And What Isn't?* Fast Company. https://www.fastcompany.com/3046119/defining-the-sharing-economy-what-is-collaborative-consumption-and-what-isnt

Botsman, R., & Rogers, R. (2011). *What's Mine is Yours*. Collins.

bs4. (2016, March 3). PyPI. https://pypi.org/project/bs4/

Brochado, A., Rita, P., & Moro, S. (2019). Discovering patterns in online reviews of Beijing and Lisbon hostels. *Journal of China Tourism Research*, 15(2), 172-191.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675-693.

CBRE. (2020). *Short time rentals*. https://www.hospitalitynet.org/file/152008780.pdf

Cetim, R. (2020). SHARE OR NOT SHARE AT ALL ECONOMY: EFFICIENCIES AND PROBLEMS OF PLATFORMS. -.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chen, X., & Xie, H. (2020). A Structural Topic Modeling-Based Bibliometric Study of Sentiment Analysis Literature. *Cognitive Computation*, *12*(6), 1097–1129. https://doi.org/10.1007/s12559-020-09745-1

Chen, Y., & Xie, J. (2008). Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science*, *54*(3), 477–491. https://doi.org/10.1287/mnsc.1070.0810

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

Choi, T. Y., & Chu, R. (2001). Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *International Journal of Hospitality Management*, *20*(3), 277–297. https://doi.org/10.1016/s0278-4319(01)00006-8

Coffey, H. (2017, November 15). Strict new rules imposed on Airbnb rentals in Paris. *The Independent*. https://www.independent.co.uk/travel/news-and-advice/airbnb-paris-rules-new-rental-limit-bookings-stay-nights-latest-a8054771.html

Crosby, L. A., & Stephens, N. (1987). Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry. *Journal of Marketing Research*, *24*(4), 404. https://doi.org/10.2307/3151388

Dabholkar, P. A. (2006). Factors Influencing Consumer Choice of a "Rating Web Site": An Experimental Investigation of an Online Interactive Decision Aid. *Journal of Marketing Theory and Practice*, *14*(4), 259–273. https://doi.org/10.2753/mtp1069-6679140401

Daunorienė, A., Drakšaitė, A., Snieška, V., & Valodkienė, G. (2015). Evaluating Sustainability of Sharing Economy Business Models. *Procedia - Social and Behavioral Sciences*, *213*, 836–841. https://doi.org/10.1016/j.sbspro.2015.11.486

Davis, P. (2016). How Do Sharing Economy Companies Grow? A Comparison of Internal and External Growth Patterns of Airbnb and Uber. *Chancellor's Honors Program Projects*. https://trace.tennessee.edu/cgi/viewcontent.cgi?article=2925&context=utk_chanhonoproj

Demir, E., & Emekli, G. (2021). Is Airbnb no longer a sharing economy platform? Evidence from Europe's top 10 Airbnb destinations. *Anatolia*, *32*(3), 470–488. https://doi.org/10.1080/13032917.2021.1890626

Ding, K., Choo, W. C., Ng, K. Y., Ng, S. I., & Song, P. (2021). Exploring Sources of Satisfaction and Dissatisfaction in Airbnb Accommodation Using Unsupervised and Supervised Topic Modeling. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.659481

Dredge, D., & Gyimóthy, S. (2015). The collaborative economy and tourism: Critical perspectives, questionable claims and silenced voices. *Tourism Recreation Research*, *40*(3), 286–302. https://doi.org/10.1080/02508281.2015.1086076

Duan, W., Gu, B., & Whinston, A. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing*, *84*(2), 233–242. https://doi.org/10.1016/j.jretai.2008.04.005

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506. https://doi.org/10.1016/j.tourman.2015.07.018

Friedman, T. L. (2013, July 20). *Opinion | Welcome to the 'Sharing Economy.'* The New York Times. https://www.nytimes.com/2013/07/21/opinion/sunday/friedman-welcome-to-the-sharing-economy.html

Gaedeke, R. M., Boone, L. E., & Kurtz, D. L. (1974). Contemporary Marketing. *Journal of Marketing*, *38*(4), 119. https://doi.org/10.2307/1250415

Goes, P., Guo, C., & Lin, M. (2014). Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2413052

Gretzel, U. (2017). *Role & Impact of Online Travel Reviews*. https://www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf

Griswold, A. (2015, July 6). *Airbnb Is Thriving. Hotels Are Thriving.* Slate Magazine. https://slate.com/business/2015/07/airbnb-disrupting-hotels-it-hasnt-happened-yet-and-both-are-thriving-what.html

Grybaitė, V., & Stankevičienė, J. (2016). Motives for participation in the sharing economy – evidence from Lithuania. *Ekonomia i Zarzadzanie*, *8*(4), 7–17. https://doi.org/10.1515/emj-2016-0028

Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*, 43, 269-272.

Guttentag, D. (2013). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192–1217. https://doi.org/10.1080/13683500.2013.827159

Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2017). Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study. *Journal of Travel Research*, *57*(3), 342–359. https://doi.org/10.1177/0047287517696980

Guttentag, D. (2019). Progress on Airbnb: a literature review. Journal of Hospitality and Tourism Technology, 10(4), 814–844. https://doi.org/10.1108/jhtt-08-2018-0075

Hamari, J., Sjöklint, M., & Ukkonen, A. (2015). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*, *67*(9), 2047–2059. https://doi.org/10.1002/asi.23552

Hannonen, O. (2020d). In search of a digital nomad: defining the phenomenon. *Information Technology &Amp; Tourism*, *22*(3), 335–353. https://doi.org/10.1007/s40558-020-00177-z

Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, *4*(1), 60–75. https://doi.org/10.1177/109467050141006

Havas. (2014). *THE NEW CONSUMER AND THE SHARING ECONOMY*. https://dare.havas.com/news-archive/the-new-consumer-and-the-sharing-economy/

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, *18*(1), 38–52. https://doi.org/10.1002/dir.10073

Hira, A., & Reilly, K. (2017). The Emergence of the Sharing Economy: Implications for Development. *Journal of Developing Societies*, *33*(2), 175–190. https://doi.org/10.1177/0169796x17710071

Hsu, M. H., Ju, T. L., Yen, C. H., & Chang, C. M. (2007). Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International Journal of Human-Computer Studies*, *65*(2), 153–169. https://doi.org/10.1016/j.ijhcs.2006.09.003

https://www.condorferries.co.uk/tourism-in-europe-statistics. (n.d.). Condor Ferries. https://www.condorferries.co.uk/tourism-in-europe-statistics

Hung, J. L., & Zhang, K. (2011). Examining mobile learning trends 2003–2008: a categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher Education*, *24*(1), 1–17. https://doi.org/10.1007/s12528-011-9044-9

Inside Airbnb: Adding data to the debate. (n.d.). Inside Airbnb. http://insideairbnb.com/behind.html

International Tourism and Covid-19 | Tourism Dashboard. (n.d.). Retrieved October 1, 2022, from https://www.unwto.org/tourism-data/international-tourism-and-covid-19

John, N. A. (2013). The social logics of sharing. *The Communication Review*.

Jung, J., Yoon, S., Kim, S., Park, S., Lee, K. P., & Lee, U. (2016). Social or Financial Goals? *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. https://doi.org/10.1145/2851581.2892328

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review Through Text Mining. *IEEE Access*, *8*, 67698–67717. https://doi.org/10.1109/access.2020.2983656

Keller, L. K. K. P. T. (2022). *Framework for Marketing Management (6th Edition) by Philip T. Kotler Kevin Lane Keller(2015–01-26)*. Pearson.

Khan, M., Wu, X., Xu, X., & Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. *2017 IEEE International Conference on Communications (ICC)*. https://doi.org/10.1109/icc.2017.7996801

Kiatkawsin, K., Sutherland, I., & Kim, J. Y. (2020). A Comparative Automated Text Analysis of Airbnb Reviews in Hong Kong and Singapore Using Latent Dirichlet Allocation. *Sustainability*, *12*(16), 6673. https://doi.org/10.3390/su12166673

Knutson, B. J. (1988). Ten Laws of Customer Satisfaction. *Cornell Hotel and Restaurant Administration Quarterly*, *29*(3), 14–17. https://doi.org/10.1177/001088048802900309

Koens, K., Postma, A., & Papp, B. (2018). Is Overtourism Overused? Understanding the Impact of Tourism in a City Context. *Sustainability*, *10*(12), 4384. https://doi.org/10.3390/su10124384

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(1), 100008. https://doi.org/10.1016/j.jjimei.2021.100008

Lau, K. N., Lee, K. H., & Ho, Y. (2005). Text Mining for the Hotel Industry. *Cornell Hotel and Restaurant Administration Quarterly*, *46*(3), 344–362. https://doi.org/10.1177/0010880405275966

Lees-Haley, P. (2004). Commentary on Butcher, Arbisi, Atlis, and McNulty (2003) on the Fake Bad Scale. *Archives of Clinical Neuropsychology*, *19*(3), 333–336. https://doi.org/10.1016/j.acn.2003.12.004

Li, J., Moreno, A., & Zhang, D. J. (2015). Agent Behavior in the Sharing Economy: Evidence from Airbnb. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2708279

Liang, L. J., Choi, H. C., & Joppe, M. (2017). Understanding repurchase intention of Airbnb consumers: perceived authenticity, electronic word-of-mouth, and price sensitivity. *Journal of Travel & Tourism Marketing*, *35*(1), 73–89. https://doi.org/10.1080/10548408.2016.1224750

Liang, L. J., Choi, H. C., & Joppe, M. (2018). Exploring the relationship between satisfaction, trust and switching intention, repurchase intention in the context of Airbnb. *International Journal of Hospitality Management*, *69*, 41–48. https://doi.org/10.1016/j.ijhm.2017.10.015

Lim, W. M. (2020). The Sharing Economy: A Marketing Perspective. *Australasian Marketing Journal*, *28*(3), 4–13. https://doi.org/10.1016/j.ausmj.2020.06.007

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1). https://doi.org/10.1186/s40064-016-3252-8

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

Lovins, H. L., & Cohen, B. (2011). *Climate Capitalism: Capitalism in the Age of Climate Change* (First Edition). Hill and Wang.

Luo, Y., & Tang, R. L. (2019). Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *International Journal of Hospitality Management*, *80*, 144–154. https://doi.org/10.1016/j.ijhm.2019.02.008

Mayer, C., & Schwager, A. (2007). Understanding customer experience. *Harvard Business Review*.

matplotlib. (2022, October 8). PyPI. https://pypi.org/project/matplotlib/

Mazarura, J., de Waal, A., & de Villiers, P. (2020). A Gamma-Poisson Mixture Topic Model for Short Text. *Mathematical Problems in Engineering*, *2020*, 1–17. https://doi.org/10.1155/2020/4728095

Mizerski, R. W. (1982). An Attribution Explanation of the Disproportionate Influence of Unfavorable Information. *Journal of Consumer Research*, *9*(3), 301. https://doi.org/10.1086/208925

M.Kwale, F. (2013). An Efficient Text Clustering Framework. *International Journal of Computer Applications*, *79*(8), 30–38. https://doi.org/10.5120/13763-1607

Mo, Y. (n.d.). *Topic modelling for supporting systematic reviews* [MA Thesis]. University of Manchester.

Moon, H., Miao, L., Hanks, L., & Line, N. D. (2019). Peer-to-peer interactions: Perspectives of Airbnb guests and hosts. *International Journal of Hospitality Management*, *77*, 405–414. https://doi.org/10.1016/j.ijhm.2018.08.004

Moreno-Gil, S., & Coca-Stefaniak, J. A. (2020). Guest editorial. *International Journal of Tourism Cities*, *6*(1), 1–7. https://doi.org/10.1108/ijtc-03-2020-174

Moro, S., Rita, P., Esmerado, J. and Oliveira, C. (2019), "Unfolding the drivers for sentiments generated by Airbnb Experiences", *International Journal of Culture, Tourism and Hospitality Research*, Vol. 13 No. 4, pp. 430-442. https://doi.org/10.1108/IJCTHR-06-2018-0085

Mudambi, & Schuff. (2010). Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185. https://doi.org/10.2307/20721420

Nasukawa, T., & Yi, J. (2003). Sentiment analysis. *Proceedings of the International Conference on Knowledge Capture - K-CAP '03*. https://doi.org/10.1145/945645.945658

Nielson. (2014). *Is Sharing the New Buying?* https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/global-share-community-report-may-2014.pdf

NLTK :: Natural Language Toolkit. (n.d.). Retrieved October 30, 2022, from https://www.nltk.org/

numpy. (2022, October 12). PyPI. https://pypi.org/project/numpy/

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.1177/002224378001700405

Olson, M., & Kemp, S. (2015). *Sharing economy. An In-Depth Look At Its Evolution & Trajectory Across Industries*. http://collaborativeeconomy.com/wp/wp-content/uploads/2015/04/SharingEconomy-An-In-Depth-Look-At-Its-Evolution-and-Trajectory-Across-Industries

On, H. (2018, May 30). *Record valuation for AirBnB*. Hospitality ON. https://hospitality-on.com/en/airbnb/record-valuation-airbnb

pandas. (2022, October 19). PyPI. https://pypi.org/project/pandas/

Patodkar, V. N., & I.R, S. (2016). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *IJARCCE*, *5*(12), 320–322. https://doi.org/10.17148/ijarcce.2016.51274

Paulauskaite, D., Powell, R., Coca-Stefaniak, J. A., & Morrison, A. M. (2017). Living like a local: Authentic tourism experiences and the sharing economy. *International Journal of Tourism Research*, *19*(6), 619–628. https://doi.org/10.1002/jtr.2134

Perren, R., & Grauerholz, L. (2015). Collaborative Consumption. *International Encyclopedia of the Social & Behavioral Sciences*, 139–144. https://doi.org/10.1016/b978-0-08-097086-8.64143-0

Peterson, R. A., & Wilson, W. R. (1992). Measuring customer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, *20*(1), 61–71. https://doi.org/10.1007/bf02723476

Petropoulos, G. (2017). *An economic review of the collaborative economy*. Bruegel.

Piscicelli, L., Cooper, T., & Fisher, T. (2015). The role of values in collaborative consumption: insights from a product-service system for lending and borrowing in the UK. *Journal of Cleaner Production*, *97*, 21–29. https://doi.org/10.1016/j.jclepro.2014.07.032

Plank, A. (2016). The hidden risk in user-generated content: An investigation of ski tourers' revealed risk-taking behavior on an online outdoor sports platform. *Tourism Management*, *55*, 289–296. https://doi.org/10.1016/j.tourman.2016.02.013

PWC. (2020). *Sharing or paring? Growth of the sharing economy*. https://www.pwc.com/hu/en/kiadvanyok/assets/pdf/sharing-economy-en.pdf

Quimby, D. (2017, May). *Hotels vs. the (OTA) World*. https://www.phocuswright.com/Travel-Research/Research-Updates/2017/Hotels-vs-the-OTA-World

Ranchordas, S. (2015). Does Sharing Mean Caring? Regulating Innovation in the Sharing Economy. *J.L. Sci. & Tech*.

Rauch, D. E., & Schleicher, D. (2015). Like Uber, But for Local Governmental Policy: The Future of Local Regulation of the "Sharing Economy." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2549919

Reporter, G. S. (2017, May 9). *Smarter regulation for the sharing economy*. The Guardian. https://www.theguardian.com/science/political-science/2015/may/20/smarter-regulation-for-the-sharing-economy

Rodgers, S., & Wang, Y. (2011). Electronic Word of Mouth and Consumer Generated Content. *Handbook of Research on Digital Media and Advertising*, 212–231. https://doi.org/10.4018/978-1-60566-792-8.ch011

Rosenberg, T. (2013, June 5). *It's Not Just Nice to Share, It's the Future.* New York Times. https://opinionator.blogs.nytimes.com/2013/06/05/its-not-just-nice-toshare-its-the-future/?_r=0

Sazema, A. (2017). Literature Review on Customer Satisfaction. *Sage*.

scikit-learn. (2022, October 26). PyPI. https://pypi.org/project/scikit-learn/

Schindler, R., & Bickart, B. (2005). Published Word of Mouth: Referable, Consumer-Generated Information on the Internet. *Online Consumer Psychology: Understanding and Influencing Behavior in the Virtual World*.

Schor, J. (2016). Debating the Sharing Economy. *Journal of Self-Governance & Management Economics*.

Schor, J., & Fitzmaurice, C. (2015). Collaborating and Connecting: The Emergence of the Sharing Economy. *Handbook of Research on Sustainable Consumption*.

Sharda, R., Delen, D., & Turban, E. (2014). *Business Intelligence*. Pearson.

Sheth, J. N., Sethia, N. K., & Srinivas, S. (2010). Mindful consumption: a customer-centric approach to sustainability. *Journal of the Academy of Marketing Science*, *39*(1), 21–39. https://doi.org/10.1007/s11747-010-0216-3

Shuqair, S., Pinto, D. C., So, K. K. F., Rita, P., & Mattila, A. S. (2021). A pathway to consumer forgiveness in the sharing economy: The role of relationship norms. *International Journal of Hospitality Management*, 98, 103041.

Smolka, C., & C, H. (2014). The Best of Both Worlds: Conceptualizing Tradeoffs between Openness and Closedness for Sharing Economy Model. *12th International Open and User Innovation Workshop*.

spacy. (2022, October 20). PyPI. https://pypi.org/project/spacy/

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, *32*(6), 1310–1323. https://doi.org/10.1016/j.tourman.2010.12.011

Statista. (2021, March 22). *European countries with the highest number of international tourist arrivals in 2019*. https://www.statista.com/statistics/261729/countries-in-europe-ranked-by-international-tourist-arrivals/

Stephany, A. (2015). *The Business of Sharing: Making it in the New Sharing Economy* (2015th ed.). Springer.

Sthapit, E., & Jiménez-Barreto, J. (2018). Sharing in the host–guest relationship: perspectives on the Airbnb hospitality experience. *Anatolia*, *29*(2), 282–284. https://doi.org/10.1080/13032917.2018.1448876

stop-words. (2018, July 23). PyPI. https://pypi.org/project/stop-words/

Su, A. Y. L. (2004). Customer satisfaction measurement practice in Taiwan hotels. *International Journal of Hospitality Management*, *23*(4), 397–408. https://doi.org/10.1016/j.ijhm.2004.02.002

Sundararajan, A. (2016). *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*. MIT Press.

Sutherland, I., & Kiatkawsin, K. (2020). Determinants of Guest Experience in Airbnb: A Topic Modeling Approach Using LDA. *Sustainability*, *12*(8), 3402. https://doi.org/10.3390/su12083402

Teubner, T. (2014). *THOUGHTS ON THE SHARING ECONOMY*. Proceedings of the International Conference on e-Commerce.

Teubner, T., Adam, M., Camacho, S., & Hassanein, K. (2014). *Understanding resource sharing in C2C platforms: The role of picture humanization*. Australasian Conference on Information Systems, Auckland, New Zeland. https://doi.org/10.13140/2.1.3578.0167

textblob. (2021, October 22). PyPI. https://pypi.org/project/textblob/

Touchpoints Bring the Customer Experience to Life. (2014, July 23). Harvard Business Review. https://hbr.org/2010/12/touchpoints-bring-the-customer

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. (2016). 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

UNWTO. (2020, December). *World tourism barometer and Statistical Annex*. https://www.e-unwto.org/doi/abs/10.18111/wtobarometereng.2020.18.1.7

Varma, A., Jukic, N., Pestek, A., Shultz, C. J., & Nestorov, S. (2016). Airbnb: Exciting innovation or passing fad? *Tourism Management Perspectives*, *20*, 228–237. https://doi.org/10.1016/j.tmp.2016.09.002

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*(1), 123–127. https://doi.org/10.1016/j.tourman.2008.04.008

Why are online reviews in the sharing economy skewed toward positive ratings? Linking customer perceptions of service quality to leaving a review of an Airbnb stay. (2021). *Journal of Tourism Futures*.

wordcloud. (2022, June 27). PyPI. https://pypi.org/project/wordcloud/

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, *27*(2), 634–639. https://doi.org/10.1016/j.chb.2010.04.014

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, *54*(5), 687–705. https://doi.org/10.1509/jmr.15.0204

Zhang, Z., & Fu, R. J. C. (2020). Accommodation Experience in the Sharing Economy: A Comparative Study of Airbnb Online Reviews. *Sustainability*, *12*(24), 10500. https://doi.org/10.3390/su122410500