ISCTE Business School
Instituto Universitário de Lisboa

AIRBNB CUSTOMER SATISFACTION THROUGH ONLINE REVIEWS

Sara Raquel Pascoal Barbosa

Dissertation submitted as a partial requirement for the conferral of Master of Management of Services and Technology

Supervisor:

Prof. Doutor José Gonçalves Dias, Prof. Associado, ISCTE Business School, Departamento de Métodos Quantitativos para Gestão e Economia

September 2019

ISCTE Business School
Instituto Universitário de Lisboa

AIRBNB CUSTOMER SATISFACTION THROUGH ONLINE REVIEWS

Sara Raquel Pascoal Barbosa

Dissertation submitted as a partial requirement for the conferral of Master of Management of Services and Technology

Supervisor:

Prof. Doutor José Gonçalves Dias, Prof. Associado, ISCTE Business School, Departamento de Métodos Quantitativos para Gestão e Economia

September 2019

ISCTE Business School
Instituto Universitário de Lisboa

AIRBNB CUSTOMER SATISFACTION THROUGH ONLINE REVIEWS

Sara Raquel Pascoal Barbosa

"Your most unhappy customers are your greatest source of learning"

Bill Gates

Resumo

Com o desenvolvimento e maior acesso à Internet, dispositivos móveis e redes sociais, as pessoas começaram a publicar online as suas opiniões e avaliações de produtos e serviços. Estes comentários influenciam as decisões de compra de novos clientes e permitem às empresas obter um maior conhecimento sobre a experiência e satisfação dos seus clientes. Assim, tornou-se imprescindível para as estas, adotarem métodos capazes de analisar esta informação e extrair valor da mesma de modo a conseguirem atender de forma mais ajustada às necessidades dos seus clientes.

A área da hospitalidade foi uma das mais afetadas por esta tendência. Por esse motivo, este estudo vai ser focado nas reviews de uma plataforma online, o Airbnb, juntando assim também uma disrupção tecnológica desta mesma área. Este novo método de alojamento partilhado tem ganho cada mais seguidores pelas suas vantagens e diferenças em relação a hotéis mais comuns, mas também tem sido um assunto cada vez mais estudado por investigadores.

Os comentários estudados do Airbnb descrevem as experiências de cada hóspede relativamente ao alojamento onde permaneceram e são estudados através de Text Mining. Este consiste em vários métodos capazes de analisar grandes volumes de informação não estruturados como Big data para consequentemente compreender melhor a satisfação geral dos clientes, nomeadamente os fatores que a vão influenciar. Os resultados mostram que existem várias dimensões valorizadas e diferentes para as zonas estudadas em Sintra.

Palavras-chave: Comentários, Airbnb, text mining, satisfação.

Abstract

With the development and better access to the Internet, mobile devices and social media, people began to post online their opinions and reviews of products and services. These comments influence new customer buying decisions and qualify companies to gain superior insight into their customers' experience and satisfaction. Thus, it has become essential for companies to adopt methods capable of analyzing this information and extracting its value in order to better serve their customers' unmet needs.

The area of tourism and hospitality was one of the most affected by this trend. For this reason, this study will focus on the reviews of an online platform, Airbnb, so that it also studies the technological disruption in the mentioned industry. This new method of home-sharing has gained more and more followers for its advantages and differences compared to common hotels, which has triggered increasing researcher.

Airbnb's guest reviews describe each guest's experiences (the positive and negative aspects of their stay) and will be studied through Text Mining. This consists of several methods capable of analyzing large amounts of unstructured information such as Big Data, in order to better understand overall customer satisfaction, including the factors that will influence it. Results show that distinct dimensions are valued by guests and they are different in different areas of Sintra.

Key-words: Reviews, Airbnb, text mining, satisfaction.

Acknowledgments

I want to thank my supervisor, Prof. Dr. José Gonçalves Dias, for accepting this challenge and for the tireless support from the beginning to the end. All the support and availability, all the transmission of knowledge and guidance were essential. With this project, I not only completed my master's degree, but also went beyond my knowledge, learned a lot and would not have achieved it without the help and effort of my supervisor.

To all my family and boyfriend, especially my mother. The most wonderful and patient person who gave me the strength to never give up and made me believe that I was able to achieve my goals. Even far away, she has managed to be my pillar at all times and will always be my example to follow.

To all my colleagues and friends who supported this journey, directly or indirectly, by motivating me and believing me.

Table of Contents

List of Figures

List of Tables

Acronyms

BI	Business Intelligence
B2B	Business to business
B2C	Business to consumer
CEO	Chief Executive Officer
CRM	Customer Relationship Management
CTM	Correlated Topic Model
DTM	Document-Term Matrix
LARA	Latent Aspect Rating Analysis
LCR	Latent Classification regression
LDA	Latent Dirichlet Allocation
P2P	Peer to peer
TDM	Term-Document Matrix
TM	Topic Modeling
UCG	User-generated content
WOM	Word-of-mouth
EWOM	Electronic word-of-mouth

Chapter I – Introduction

1.1. Background

The sharing economy has revolutionized the business world. Thanks to the emergence of the smartphone ecosystem and widespread connectivity, information technologies have evolved and led to online sharing, collaboration, and content generation (Stephany, 2015). This concept has been portrayed as a new way of dealing with goods as a new way to transact them in the market since the focus is no longer on ownership but on access to resources. It is regarded as an inviting alternative for many consumers and is increasingly present in the basic activities of people's lives; thanks to the continued development of technology, which helps share products and services over the Internet.

There is increasing consumer awareness of promoting social and environmentally friendly activities, rapid growth of collaborative Internet communities and new activities in social sharing and commerce (Stephany, 2015; Hamari et al., 2016). Day by day, people are acting in favor of sustainability, economic gains, and better utilization of any product or service (Hamari et al., 2016), which is attractive to companies and researchers. Its scope reaches several business lines, including two, which are widely known to the Portuguese market: private passenger transport (Uber and Cabify) and tourist accommodation (Airbnb).

Without any doubt, the tourism and hospitality industry were the most affected industries by the sharing economy (Cheng, 2016). In fact, customers began to exchange traditional experiences for the unique experiences provided by sharing economy services. Their motivation is now to "live like a local", to meet new people, and to have a more realistic experience of the culture and customs of local communities (Bridges and Vásquez, 2016).

Airbnb is an online peer-to-peer platform that has become one of the most successful business models in the sharing economy, connecting a total number of 6 million unique accommodations in nearly 100000 cities and 191 countries (Airbnb, 2019a). It provides hosting-sharing services that allow apartment or room owners to rent them to guests for affordable prices. Moreover, studies have been carried out to prove that

Airbnb usually applies lower prices than home and hotel rates (Lee and Kim, 2018). Its growth has been significant in such a way that its popularity turned into a disruptive factor in the hotel system. This triggered the need for research on this unique operating system. Consequently, it is possible to find studies on the differences between staying at Airbnb accommodations and hotels (the difference of the experience that it provides) (for example, Lehr, 2015), characteristics of the peer-to-peer transactions (e.g. Botsman and Rogers, 2010), as well as legal and financial issues involving Airbnb (e.g. Ert et al., 2016).

Literature in tourism and hospitality shows the determinants of customer experience and the influence of different marketing strategies on customer perception and satisfaction (Kim and Kim, 2004). However, these are not appropriate in the case of sharing economy as a result of a different type of experience undergone by its customers. Therefore, there is a need to identify a new set of experience quality indicators for the sharing economy.

With the development of technology and the greater access to the Internet, the way goods and services have been acquired has changed. In fact, nowadays it is possible to share assessments on social media platforms, which are known as user-generated content. In the e-tourism era, many customers book hotels online and post reviews after their stay, in addition to rating them. These reviews describe customers' experiences, highlight product features and services, and provide customer insight in detail, thereby demonstrating overall customer satisfaction and generating an electronic word-of-mouth effect that will influence potential customers. They are more reliable than the information shared by the company in a way that it shapes the company's reputation and impact its sales (Öğüt and Onur Taş, 2012; Mudambi and Schuff, 2010).

More and more professionals are trying to manage customer' opinion, which consists of obtaining direct feedback from them. Service evaluation of the hosting experience can open doors to new ways of measuring service quality and customer satisfaction; and, consequently, adapt the strategies used by the company and ensure its success. At the level of accommodation sharing, there are still few studies that have focused on written client reviews as a way of assessing their experiences and identifying key attributes that influence those experiences and contribute to customer satisfaction or dissatisfaction.

1.2. Main objective

The main objective of this research is to understand consumer satisfaction concerning sharing economy service, Airbnb, through the analysis of online reviews. The study is focused on the Sintra area, one of the most touristic Portuguese spots.

The overall aim is to identify the dimensions valued by customers in an Airbnb experience (for example, the relation with the host) based on published online reviews so as the measure each dimension' weight in the evaluation and classification of the stay.

1.3. Specific objectives

Based on the literature review and Airbnb website data, we conducted research on the relationship that factors location, price, host, and commodities have on the various online reviews and satisfaction. In particular, we want to compare these factors concerning accommodations located in "not so tourist" locations with accommodations in locations near Sintra' center, where most of the tourist activities are concentrated.

This requires the checking of the added value, consistency and quantity of information. The information described in the reviews should discuss the main aspects of the house/room, its location, price, amenities, among others. After extracting data, it would be interesting to analyze some aspects between the chosen areas, such as the price average (per night, cleaning and service), the rating of the accommodation and the host (superhost or not), among others. Having a high volume of reviews with descriptive information (not a quantitative rating) is essential to ensure that there is enough data to measure customer satisfaction and for hosts to know in what aspects they can improve their customers' experience.

1.4. Research questions

The research questions put forward in this research are:

1. What are the most important factors in an Airbnb stay that will influence customer satisfaction? How are they ranked based on relative importance?
2. Can text mining techniques detect patterns in the text capable of providing value for decision-making?

1.5. Research methodology

In order to analyze customer satisfaction through textual reviews, it is necessary to obtain them through the website InsideAirbnb.com, where we can find the ratings and reviews written after users' stays.

After that, text mining techniques will be used through the software R, so as to carry out the first data cleansing and provide descriptive analysis, such as a WordCloud with the most usual words used in the reviews. Then, it will apply the Latent Dirichlet Allocation (LDA) to structure the topics. This analysis allows us to identify the most usual words used in the reviews and their frequency, which reflects the importance given by the guests to the described aspects.

1.6. Overall thesis structure

Following the objectives set, the presentation of the thesis structure is organized in five chapters:

Chapter I - Introduction introduces the topic as well as its analytic framework. Objectives and research questions are set. Finally, the methodology, scope and overall structure of the dissertation are discussed.

Chapter II - Review of literature presents the theoretical framework of the research topic, which serves as the basis for the development of the research process. It covers the main concepts related to the research topic, namely sharing economy and the Airbnb case, online reviews, and customer satisfaction. A relationship between these concepts is also established. It also presents the questions to be researched.

Chapter III - Methodology discusses the population and the sample of the study. It briefly discusses the methods of data collection and data analysis.

Chapter IV - Analysis of Results reports the results of the analysis of online reviews using R.

Chapter V - Conclusions sets out a summary of the findings in order to answer the objectives and research questions. Specific limitations of research are outlined and recommended ideas for future research are suggested.

Chapter II - Literature Review

2.1. Sharing economy

With the development of the Internet, the link between supply and demand became easier and generated collaborative consumption and sharing economy. This phenomenon quickly gained users, allowing new means of economic and social interaction (Belk, 2014). Therefore, people left the conventional model of acquiring products and services for a new one, where consumers have access to other consumers' assets, through business and digital market platforms (Dervojeda et al., 2013). Thus, markets that were once invisible have emerged: consumers began to share cars, clothes, dogs, boats, rooms, and houses (Geron, 2013). Furthermore, due to this phenomenon, consumers began to play an active role and became both consumer and produces, turning into "prosumers" (Stephany, 2015). People were captivated by this model of sharing and became micro-entrepreneurs of their own assets (Dervojeda et al., 2013). They are motivated by several factors: sustainability, in the sense of providing more environmentally friendly alternatives; better utilization of products and services; increased interpersonal interaction and economic gains (Hamari et al., 2016; Botsman and Rogers, 2010).

Sharing was also beneficial to companies that put emphasis on innovation and future orientation (Belk, 2014), by being present in industries, services and business models (Sundararajan, 2016). According to the European Commission's 2016 Report on the Presence and Size of Collaborative Economy in Europe, Europe has already more than 275 sharing economy platforms. In addition, this industry has become imperative for the economy and its operations have been researched within the areas of management, marketing and information technology (Karakostas et al., 2005). However, sharing economy is not an innovation. People have always lived based on sharing, according to collaborative consumption, without being aware of it. Whether it is in public transport, apartment blocks or public spaces (Sundararajan, 2016), the idea of saving the environment and associating with a community of people with whom we feel affinity (similarities) has always been a way of living (Rosenberg, 2013). With easy access to the Internet, this sharing has become more evident, promoting self-employment and the exchange of interactions among community members (Sundararajan, 2016).

2.1.1. Definition

The term "sharing economy" has been discussed by philosophers, economists, and entrepreneurs (Smolka and Hienerth, 2014). The first to name "sharing economy" as "collaborative consumption" were Botsman and Rogers (Botsman and Rogers, 2010). Since then, different authors have proposed other terms and definitions like "consumer-sharing system" (Lamberton and Rose, 2012) or "access-based consumption" (Bardhi and Eckhardt, 2012). There are also several other interchangeable terms, such as "peer-to-peer economics"; however, the most used one remains "sharing economy" (Stephany, 2013).

From several attempts to define the "sharing economy", it is possible to identify specific characters. First, "sharing economy" is a "socioeconomic" model based on shared use (Botsman et al., 2012). It is guided by social interaction, through the exchange of assets allowing the appearance of new services. The consumers generate value by sharing and renting products on consumer-to-consumer platforms. Second, it promotes temporary access, rather than ownership of the good, excluding second-hand economy (Botsman, 2015). Products or services are not purchased, only access is facilitated. For example, a person who needs to go on a certain trip can simply go by car with another person who goes to the same destination instead of buying their own car. Third, the Internet provides collaborative consumption, as it provides creative and progressive ways for people to interact. In fact, anyone with Internet access can become a vendor, advertise their products and services, and create a trustworthy image in order to attract customers and connect with others (Matzler et al., 2015), who also consumes products and services from other consumers. Finally, "sharing economy" is an economy under demand, which means that there is an efficient use of fixed assets. The value added is determined in a negotiation between both parties (Botsman et al., 2012). Many companies choose to add the collaborative consumption model to the traditional supply and demand model. Thus, whenever necessary, "sharing economy" provides an available resource that is needed through exchange or rent, instead of the purchase of a product or service (Botsman et al., 2012).

According to Botsman (2013), "sharing economy" is categorized into four different activities: collaborative production (a company carries out the innovations of an online community), collaborative consumption (use of excess assets), collaborative lending (loans between people), and collaborative education (open access to person-to-person education). However, Belk (2014) does not agree with the inclusion of certain sharing arrangements (e.g. Couchsurfing, Zipcar) as being part of the “sharing economy”, since he considers the definition of Botsman and Rogers (2011) to be very vague and confusing. Moreover, the author argues that it should be more limited without including all types of exchange. Bearing this in mind, he redefined “sharing economy” as the acquisition and distribution of resources for a fee or compensation (Belk, 2014). Botsman (2013) argues that a fee or compensation is the equivalent of a “monetary or non-monetary benefit” that is, providers can both get money and something of similar value. In turn, Stephany (2015) divides the “sharing economy” into two models: business-to-consumer (B2C) and peer-to-peer (P2P). In the first, the company holds the assets and allows the consumer to access them through, for example company car rental. In the other model, consumers share products and services, like for example someone lending his own car so another person can enjoy it. Botsman (2013) also adds a third model – business-to-business (B2B) model – in which companies offer their underutilized assets to other companies.

2.1.2. Drivers

2.1.2.1. Technological driver

With the development of Web 2.0, innovative information technologies have emerged and have triggered a rapid growth of websites. This way, they encouraged sharing, collaboration, and user-generated online content (John, 2013). Unlike the previous generation that only looked for information on the Internet, it is now possible to create a peer-to-peer network business relationship. We entered the digital age (Denning, 2014), where people search for information, interact with others, create content or buy products and services through digital platforms (John, 2013). The act of sharing began essentially in social networks (Garbarino and Strahilevitz, 2004), making people acquire new habits of communication, perception, and even consumption.

From that moment on, an online marketplace emerged whereby sellers can reach an unlimited number of potential consumers and save on transaction costs (Botsman and Rogers, 2011; Stephany, 2014). By exposing their products in photographs and videos, they began to provide detailed information, showing their confidence. This is a key factor in sharing economy, because people must feel secure in sharing assets with strangers. Thus, most sites incorporate social networking features where personal information such as their interests, classifications, and tastes are made available to create a relationship of trust with the consumer, also reducing anonymity (Botsman and Rogers, 2011; Olson and Kemp, 2015).

New technologies began to adopt an infrastructure that would improve the entire sharing process. For the most recent generation, smartphones began to replace all other devices in such a way that they were used to accept and confirm reservations; to look for services or even to advertise the goods themselves more quickly and conveniently (Olson and Kemp, 2015; Stephany, 2014). Another factor that stimulated the growth of this economy was the emergence of online payment platforms such as PayPal (Owyang, 2013), increasing the degree of confidence in online transactions and reducing fraud cases.

2.1.2.2. Social driver

The major driver of the sharing economy at the social level is, of course, sharing itself (Smolka and Hienerth, 2014). Researchers consider three main reasons for sharing: volunteering to disseminate resources to the community (Botsman, 2013) extrinsic or intrinsic reasons (Franke and Shah, 2003) or even contribute to the public good (Lerner and Tirole, 2002).

According to Owyang (2013), the combination of a wide interaction between people and the trend for "helping others" has contributed to the growth of the sharing economy. The interaction itself generates greater word-of-mouth, where social networks lead to greater social engagement and people begin to trust more in other people (consumers) than in the information conveyed by companies or brands (Owyang, 2013; Stephany, 2015). It means that consumers become co-creators of value and many platforms emerge to motivate experience, opinion, and recommendation sharing through peer-reviewed systems (Schor, 2016). In fact, the intrinsic transparency of new platforms is a great source of confidence-building and social validation (Guttentag, 2015).

In addition, the sharing economy allows consumers to have more real and "home-made" experiences in order to become more familiar with other cultures and communities (Olson and Kemp, 2015).

2.1.2.3. Economic driver

Sharing economies offer benefits to both providers and users. In fact, much of the existing literature indicates that participation in the sharing economy is especially motivated by economic reasons (Schor, 2016).

From the perspective of the provider, sharing economy allows him/her to become a micro-entrepreneur using personal resources, that is, to receive an income from an idle resource that will be used by a consumer. This allows him/her to reduce property expenses, as what happens in Airbnb, whose hosts use the generated income to afford the monthly house fee or regular expenses (Airbnb, 2019b). Users find advantage in the increasing range of market offers at more accessible costs. However, it is essential to comply with a principle established by Botsman and Rogers (2011): products and services available must be able to satisfy the tastes and needs of consumers; otherwise the system is not self-sustainable. Owyang (2013) and Stephany (2015) also point out that it is important to consider population density and location in an urban area because these factors will influence the breadth of word-of-mouth, which will consequently grow sharing services.

2.1.2.4. Environmental driver

The consumer's environmental concern is one of the major factors in the growth of sharing economy. In these days, there is an increase in consumer awareness as consumers are looking for sustainable solutions in their day-to-day lives. Through sharing economy, the creation of new products that will later be discarded is discouraged while sharing and exchange are stimulated, trying to use fewer materials, reduce waste and provide greater accessibility to the available resources (Bardhi and Eckhardt, 2012; Botsman, 2015). For example, shared transportation savings avoid the expense of gasoline and maintenance, less environmental pollution and more convenient parking. Thus, this economy develops a more sustainable consumption and a lower environmental impact (Piscicell et al., 2015).

2.1.2.5. Political driver

With the rapid growth of sharing economy, legal issues come out because legislation is still not well defined in different areas (Guttentag, 2015). In the housing sector, it is known that there has been a proliferation of unlicensed rents, lack of permission and registration (Stephany, 2015). In the catering area, deficiencies in hygiene certificates and compliance with health standards have been reported, which is less frequent in restaurants. In the transport area, for instance, certain drivers do not pay compulsory fees. Sharing turned out to be an illegal, unauthorized business that harms traditional businesses, mainly the primary products and industry suppliers. Companies that use sharing platforms attempt to evade corporate responsibility that is applied to traditional models (Malhotra and Van Alstyne, 2014).

2.2. Airbnb

One of the leading industries in the sharing economy is home-sharing. In old times certain convents offered stay to people in need, until the first official platform appeared, Couchsurfing in 2003 (The Economist, 2012). However, the market has evolved, and many other online platforms have emerged, in which Airbnb stands out. It consists of a platform where it is possible for owners of underutilized assets (e.g., a whole house or a room) make it available for rent in order to earn extra money (Olson and Kemp, 2015).

It began in October 2007 in San Francisco, when two students – Brian Chesky and Joe Gebbia – were having difficulty in paying for their apartment rental. They took advantage of the fact that the hotels in the area were sold out due to a conference that was going to take place and they rented a space in their living room for guests who needed a place to stay overnight. With only three air mattresses in the room, they got the money to pay for the apartment rental and a business opportunity (Friedman, 2013; Olson and Kemp, 2015). With the help of a programmer, Nathan Blecharczyk, they created the site that connected various places available around the world with travelers looking for a place to stay (Sundararajan, 2016). This way, Airbnb has gained huge popularity among its users worldwide, disrupting the entire established hotel system and putting itself in front of most hospitality groups. Some analysts estimate that over the next five years, Airbnb will accumulate half a billion per day every year, managing to achieve one billion dollars annually by 2025 (Verhage, 2016).

In each Airbnb transaction, there are two parties: the host and the guest. The hosts enlist their properties, free of charge, which can be from private homes to private or shared rooms, or even another variety of properties such as caves, castles, treehouses, among others (Friedman, 2013). They must describe the space, put price, availability, and photos that are taken by photographers contracted by AirbnbGuests subscribe to the website so that they can access the various listings and search for the one that fits them best and be able to book them (Stephany, 2014). The search can be filtered by availability, price, geographic location, convenience, language of accommodation, and type of property. After the host's permission, both get in contact, payment is made and the stay is guaranteed (Airbnb, 2019b). Both hosts and guests have access to each other's Facebook profiles, e-mail addresses, phone numbers and reviews concerning what they are offering (Friedman, 2013). Airbnb charges 3% of the host's payment (Airbnb, 2019d) and usually less than 13% of the traveler. This can change depending on the way the subtotal changes (Airbnb, 2019d).

2.2.1. Accommodations

Most Airbnb rentals are for entire properties. There are also private rooms and a small percentage of shared rooms. In Lisbon, about 74% of the listings represent whole houses, 24% private rooms, and 2% shared rooms (Airbnb, 2019c).

A study carried out in 2013 concluded that the cost of renting a private room on Airbnb is always lower than the cost of renting a room in an average hotel, with an estimated cost savings of 21.2% (Metz, 2014). Another study conducted in one of the major US cities in 2013 showed that choosing a private room from Airbnb allows individuals to save 50% of the final payment and an entire apartment is about 20% cheaper (Priceonomics, 2013). This is backed by Guttentag (2015) that portrays Airbnb as a cheaper accommodation than the other traditional accommodation options, plus the advantages of staying in someone's home.

2.2.2. Why do travelers choose Airbnb?

There are several factors that lead travelers to use the Airbnb platform, the main motivation being the desire to "live as a local" (Airbnbcitizen, 2014).

Most Airbnb lodgings are located outside tourist areas (around 76% of hotels are not located in district centers) (Airbnbcitizen, 2014), offering a different and more authentic experience (Guttentag, 2015; Priceonomics, 2013). The fact that a person stays in someone else's home, also brings benefits, like for example having access to various facilities and appliances, such as kitchen, washing machine, among others, and also benefit from useful tips on local services given by hosts.

The most discussed factor is money-saving, whose reasons have already been mentioned. In Morgan Stanley's survey (2016) about shopping with conscience, price gains more importance than location and having an authentic experience. Travelers are beginning to prefer places farther from the center because they are cheaper.

Word-of-mouth is another important factor that leads people to try home-sharing experience. In fact, it gains more importance than other traditional marketing alternatives, as it includes sharing an experience, having positive and negative comments provided by real people and not by the company. People acquire knowledge about the sharing economy service and are more inclined to use it if they know the opinions of people who have actually tried it (Owyang et al., 2014).

Home sharing is definitely a form of sharing economy increasingly adopted by people. Those who have not yet used these services state that it is due to the lack of privacy (30%) and security (9%) (Ting, 2017). However, both the host and the guest, after the stay, can post a review evaluating the latter stay. This makes it possible for future guests to become familiar with hosts and vice versa, and also to build trust (Guttentag, 2015).

2.2.3. Airbnb versus hotel industry

After Airbnb was created and achieved success, researchers began to study the impact this sharing economy would have on the hotel industry. According to the research group report of the hotel industry STR, Airbnb has the same seasonal demand as hotels (Airbnbcitizen, 2016).

Many analysts say Airbnb has no significant impact on the hotel industry as its target is a concentrated audience of young travelers, whose goal is to save money. Mark Hoplamazian, CEO of Hyatt, in an interview with Yahoo! Finance in 2014 said that Airbnb is not considered a direct competitor because its product is different from branded hotels (Santoli, 2014). Thus, it does not satisfy many customers who value the traditional benefits such as meals in the room or gym in the hotel (Metz, 2014). According to Cathy Enz, a professor at the Cornell School of Hotel Management in Slate, Airbnb proves to be a better offer for the casual leisure traveler and groups. On the other hand, hotels are mostly satisfied only by business travelers and those who give preference to patterns of accommodation (Griswold, 2015).

Despite these opinions, most researchers believe Airbnb has a negative impact on hotels (Bridges and Vásquez, 2016). Effectively, the Meetings & Convention magazine reported that Airbnb could pose a great threat due to the lack of filling hotel rooms during major record conferences (Shapiro, 2014).

Another discussed cause is the advantage of the legal implications of the service. While hotels are required to comply with certain existing regulations to ensure security, it is suspected that Airbnb hosts will be able to provide lower prices to their customers for not incurring mandatory charges related to these regulations. According to Sundararajan (2016), Airbnb has created more demand in the overall accommodation sector, rather than diverting it from the hotel industry. This happens because Airbnb has a wide range of prices, locations and accommodation options (taking into consideration that Airbnb offers, for example, tents and treehouses).

2.3. Satisfaction

Customer satisfaction is a very subjective theme since it has no specific definition. It is a concept that is very much addressed by marketing and service management, as it leads to positive post-purchase behavior and is recognized as a critical success factor for several sectors, including hospitality (Peterson and Wilson, 1992).

The most common concept is based on the disconfirmation paradigm developed by Oliver (1980). It states that individuals anticipate the timing of the purchase transaction, create expectations, and form perceptions of service performance. Satisfaction comes from the disconfirmation of these expectations, that is, it is the result of the difference between expected and perceived performance. If perceived performance is higher than expected (positive disconfirmation), satisfaction occurs. Otherwise, expectations are not overcome (negative disconfirmation) and the client is dissatisfied (Oliver, 1980). This model was developed in the field of consumer research but has been encompassed in the satisfaction of e-commerce clients (Bhattacherjee, 2001).

Boone and Kurtz (1998) state that for an individual or a homogeneous group of customers to be satisfied with a specific product, its attributes should match their needs and exceed their expectations regarding the overall performance of the product. Attributes encompass their technical functions, but also the brand, after-sales service, warranty, vendor behavior, and other existing complimentary services.

Therefore, satisfaction depends on customer experience and will depend on the quality of product or service provided. Nevertheless, it may decrease if the created expectation is very high. After experiencing the service, the customer can evaluate the quality of the product or service in question.

2.3.1. Dissatisfaction

Some service failures correspond to a deficit in the attempt to match the customers' expectations, like for example some problem in the room's accommodation, resulting in low customer satisfaction and the resulting preference for other companies (McCollough et al., 2000).

According to the attribution theory, it is possible to highlight three factors about the cause of service failures that lead to customer dissatisfaction: causality, stability, and control (Weiner, 2000). Causality comes if the cause of the issue is the responsibility of the company or the customer. If the company is responsible, the customer may be dissatisfied and expect it to take some action to solve the problem (Iglesias, 2009). If there is no improvement, it is expected that the customer will not re-purchase the product or service. Stability refers to whether the problem is only fearful (unstable cause) or permanent (stable cause). When the cause is determined to be stable by the consumer, the customer expects the same result in the future. When the cause is unstable, future projection is already different and uncertain, causing less dissatisfaction than continuous service failures (Weiner, 2000). Control refers to whether the company has any control over the quality of service provided, according to the consumer's perspective. When a company has the habit of providing high-quality services, it is assumed that it has no control over a more recent failure (Hess et al., 2003). The experience and prior knowledge about the organization of a company's services are very important because if a brand has an excellent reputation, its failure can be regarded as a single element and not have a great impact on customer satisfaction (Laczniak et al., 2001).

It is also important to highlight that the attribution process has had an impact on customer assessment in online reviews and on consequent attitudes and behaviors (Weber and Sparks, 2010). The three factors related to the causes of dissatisfaction are linked to the motivation to post reviews of the products or services and are perceived by readers, who end up being influenced in their evaluation. The researchers concluded that negative electronic word-of-mouth (eWOM) leads to lower brand scores when message negativity is attributed to the brand and leads to positive assessments when message negativity is attributed to the communicator (Laczniak et al., 2001).

2.3.2. Satisfaction in peer-to-peer accommodation

Although there are many studies on hotel satisfaction (for example, Choi and Chu, 2001; Ren et al., 2016), the growth of sharing economy services has given rise to the need to identify the factors that influence the satisfaction of guests and their intentions when using the peer-to-peer hosting services in order to provide service quality. Since the services provided are different from the services of a traditional hotel, expectations and service evaluations are different as well. This is because they meet distinct needs, namely lower prices, more meaningful social experiences, and more sustainable stays.

Existing literature on sharing economy indicates that a major determinant of satisfaction and intention to participate in collaborative consumption is the reciprocity involved (Bellotti et al., 2015). That is, satisfaction comes from the reciprocal benefits (benefits received) obtained by the value gained from the exchange of social and material resources (Bellotti et al., 2015). Other researchers have studied the theory of self-determination, where the intention to participate in sharing economy depends on intrinsic and extrinsic motivations (e.g. Bellotti et al., 2015).

Some researchers (e.g. Botsman and Rogers, 2010) emphasize the sustainability factor, i.e. the desire to reduce waste and environmental impacts on consumption as one of the intrinsic motivations. This exchange of idle assets is termed as an increasingly sought-after social innovation (Sheth et al., 2011). It creates satisfaction as individuals feel more active and responsible citizens.

Regarding extrinsic motivations, the main one is the monetary benefit (Bellotti et al., 2015; Lamberton and Rose, 2012). Having access to idle and desired resources, with good quality, at a lower cost is very attractive and positively influences satisfaction and the intention to enjoy the advantages of sharing economy. Given the wide range of product and service options, there is an integrated online reputation system to reduce information asymmetry and make markets more competitive (Koopman et al., 2014). Another extrinsic motivation to emphasize is the desire for socialization. Creating new relationships, making new friends and meeting new people all contribute to the sense of belonging to a community both offline and online (Botsman and Rogers, 2010).

Other influencing factors of satisfaction in the case of peer-to-peer accommodation are location, which allows for a more local and less touristic experience, and access to a variety of domestic amenities (Guttentag, 2015; Kim et al., 2015). The location and amenities are attributes that most stand out, contributing on a large scale to guest satisfaction and the intention to acquire the services in the future.

Tussyadiah and Pesonen (2016) conducted a study in the United States to test the studied factors that influence customer satisfaction concerning peer-to-peer. You can conclude that it was essentially determined by monetary benefits, accommodation attributes (location, host reception by the host, and comfort of accommodation) and the fun that resulted from the whole experience.

2.4. User’s experience of Airbnb

Gentile et al. (2007: 397) states that “customer experience originates from a set of interactions between a customer and a product, a company or part of their organization, which causes a reaction”. The experience of an Airbnb customer is totally different from that of a hotel customer, but there is still much controversy as to the importance of each dimension of the Airbnb experience, which stands out for itself as a distinct Internet platform, calling itself “a trusted community market for people to list, discover and book unique accommodations around the world” and “connect people to unique travel experiences” (Airbnb, 2019c).

Some studies on Airbnb accommodations focus more on building relationships (Festtila and Müller, 2017; Tussyadiah and Pesonen, 2016; Yannopoulou, 2013), promotion of unique and real experiences (Tussyadiah and Pesonen, 2016), and significant social gatherings (Cheng, 2016); while traditional lodges focus more on functional aspects. For example, Yannopoulou (2013) argues that Airbnb essentially implies “significant enrichment in life, human contact, access and authenticity”. In this, it is worth mentioning in the reviews made by the clients, the hospitality of the hosts and the most advantageous location to know local restaurants and typical neighborhoods, supporting the studies stating that travelers have the will to feel home, even away from it, and to get in touch with local and authentic places (Airbnb, 2019c).

Guttentag (2015) concluded that the main attraction of Airbnb users is its low cost, portraying it as a hotel experience at a relatively lower cost (Festila and Müller, 2017). Other studies point out that Airbnb users valued practical attributes more than experiential ones, giving more importance to "enjoyment", "amenities", and "cost-saving" (by this order of importance) (Tussyadiah and Zach, 2016). The "location" factor did not reveal the expected importance. The reasons for these contradictory results are still unclear; but researchers suggest that this can be attributed to the lack of standard Airbnb accommodations (Tussyadiah and Zach, 2016).

Sthapit and Jiménez-Barreto (2018) found that most people booking accommodation on Airbnb were motivated by low price and location. The communication with the host was also highlighted, the experience of peer-to-peer accommodation being mostly valued for accessibility, coexistence, and economy.

The importance of the host in the experience of an Airbnb user is still unclear (Tussyadiah and Zach, 2016). There are people who value being received by the host and still appreciate tips offered about the place they are staying, while others do not have great opinion for lack of interaction with the host, since many of these hosts rent the entire apartment without ever being present (Cheng, 2016; Dredge and Gyimóthy, 2015). However, some studies portray this social interaction (guest-host) as one of the central dimensions of the experience (Festila and Müller, 2017; Tussyadiah and Zach, 2016; Yannopoulou, 2013).

Zhang et al. (2018) conducted a study where they demonstrated that perceived trust in relation to the host is very important and should be considered. To assess trust, there are three criteria: reputation, performance, and appearance. The main reputation indicator on Airbnb is the "superhost" emblem (Liang et al., 2017). Performance is relative to response speed and response rate (Malinen and Ojala, 2011). The quality of perceived information is also inherent (i.e., accuracy and integrity characteristics of the host that are provided) (Chen et al., 2014). As for the appearance, this incorporates the host profile and the way s/he receives the guests. A simple smile can reveal kindness, sociability, and honesty (Baudouin et al., 2000).

As for negative experiences, most are associated with hosts and their receptivity, as well as the amenities of the accommodations. Many of the complaints regard the deficit on the service quality expected, the amenities promised on the site compared to the ones actually made available to the guests, and also the hosts' attitude, that is whether they are friendly and able to solve any problem that arises during the stay (Sparks and Browning, 2010).

This raises the first research question: What are the most important factors in an Airbnb stay that will influence customer satisfaction? Given the lack of consensus in the literature, there is the need to understand whether all the dimensions studied are really important (location, host, price, and commodities) and how important they are.

2.5. Online reviews

Word-of-mouth (WOM) is defined as the exchange of information between consumers on the characteristics, use and properties of products and/or services (Kozinets, 2002). When disseminated by traditional means of communication, its strength was not significant. However, with the development of technologies and the Internet, along with the creation of the electronic word-of-mouth (eWOM), this has changed (Dellarocas, 2003).

Internet users began to express their opinions about consumer experiences by highlighting attributes of products and services they consume (Palese and Usai, 2018). This information shared on digital platforms is called "user-generated content" (UGC). It is estimated that about 91% of US Internet users create online content (Lenhart et al., 2004) and about 35% of users provide online reviews at least once a year (Max and Mace, 2008).

Online reviews are product reviews or in the form of numerical stellar ratings (1 to 5 or 1 to 10 usually) or through open comments with no defined structure. Numerical classification is a quantitative indication of overall performance showing generic evaluation. Meanwhile, a comment is much broader and can be easily used as an indicator of a reviewer's satisfaction since it can demonstrate the attitude and the feelings of the reviewer (Palese and Usai, 2018).

The content of online reviews includes a variety of aspects but usually is focused on two of them: the main service and the relational service. The main service consists of the company's objective and its fundamental competence in creating value for the customer. The relational service refers to consumer interactions with employees or customer services (Butcher et al., 2003). For example, in a hotel, the main service would then be comfort and room cleaning, while the relational service would be all the interaction established during the staying with the rest of the service team. Thus, much of the experience satisfaction is reported according to these two major components. However, despite the added value of relational elements from customer's perspective, a clean, comfortable room is considered much more important than a friendly welcome; thus, it has more impact on consumer overall assessment (Crosby and Stephens, 1987). That said it is known that more and more opinions and information are available online on a wide range of products and services in order to inform people about their alternatives (Mudambi and Schuff, 2010). EWOM has become a factor capable of shaping consumer behavior. There is a research that reveals a positive relationship created by user-generated positive and product sales (Dellarocas et al., 2004). This is essentially because people most readily accept and trust information given by others similar to themselves (Li, 2009). Online reviews have been proven to reach a large audience, attract more customer visits, increase the time spent on site, and create a sense of community among loyal customers (Dabholkar, 2006). Liu et al. (2017) state that customer analytics platforms are sources of reliable information to understand what drives customer satisfaction, which can help companies improve offerings to meet future demands and, therefore, obtain significant impact on sales practices and company pricing (Lu et al., 2014).

In this way, more and more companies have begun to induce customers to their recommendations platforms in order to recommend a product and be more successful. Some even hired people to generate this content and increase the rankings of certain products, which can mislead customers (Mudambi and Schuff, 2010).

2.5.1. Motivations

There are still few studies on the antecedents of online reviews made by customers. Henning-Thurau et al. (2004) demonstrated that there are three main reasons that lead people to generate evaluative content: altruism (worriedness with other customers); economic incentives (some people are paid to publish manipulative content) and homeostatic utility (attempt to restore imbalance after a bad experience). More detailed research by them showed that people evaluate products and services on online platforms primarily for reasons of concern with other consumers and community improvement. At the same time, there are also people who do it for more egocentric reasons such as emotional outbursts and economic incentives, altruism can play a more powerful role in people's decision to post online reviews.

Some people are encouraged to write reviews so as to belong to a given community status because some sites implement goal-driven online incentive hierarchies (Goes et al., 2016). Thus, the more reviews posted, the higher the status, possibly leading to the obtainment of some reward (usually a badge) (Liu et al., 2016).

Positive reviews are usually associated with a good quality product or service while negative ones refer to low-quality products (Herr et al., 1991; Lee et al., 2008). The more reviews there are, the more they will influence consumer attitudes (Lee et al., 2008); that is, if predominance is positive feedback, consumer ratings will also be more favorable. However, negative reviews have a greater influence on consumer decision-making in a negative way (Mizerski, 1982; Herr et al., 1991). This is associated with the strong willingness of customers to share their dissatisfaction when some experience did not live up to their expectations (Harrison-Walker, 2001).

According to Sparks and Browning (2011), a highly descriptive language is used in the negative reviews so that readers can feel they are living the experience. However, negative eWOM generally takes on an anti-normative connotation, diverging from prescriptive and descriptive rules. These negative reviews are attributed to people personality (they are seen as complainers and whiners) and not so much to negative customer experience (Laczniak et al., 2001).

2.5.2. Online reviews in the travel industry

From the beginning of the Internet in the ´90s, the hospitality companies start to introduce online analytical platforms operated by themselves or by outsourced organizations (Fang et al., 2016). In these sites, customers write reviews describing their opinions, ideas, and experiences about products, services, and brands (Kozinets, 2016). These reviews on digital sales platforms are called online travel reviews. They consist of product classifications and descriptions of other travelers' experiences and this is a subject that has been neglected by researchers, mainly on their impact on business performance (Dabholkar, 2006).

Travelers prefer to read reviews made by other experienced travelers than texts written by travel service platforms (Gretzel et al., 2007). These reviews play a very important role in consumer decision-making in travel and accommodation (Litvin et al., 2008). This is evidenced by studies such as Conrady (2007), where 61% of travelers claim to consult reviews before making any reservation and the study by Hock (2007) which showed that the reviews influence at least 75% of trips made. A later online survey about the use and impact of online travel reviews (Gretzel et al., 2007) showed that almost 98% of people planning to travel consulted online reviews during the planning and about 84% were even influenced in their travel choices (ComScore, 2007).

The most relevant sites for travel reviews are TripAdvisor.com, Booking.com, Virtualtourist.com, and LonelyPlanet.com (Chung and Buhalis, 2008; Gretzel et al 2007). On many of these sites, not only travelers can have access to other people's experiences, but also book flight reservations or stays (O'Mahony, 2008). TripAdvisor was considered the site that offers more content for travelers, having been part of travel planning for 45 million visitors in April 2011 (HotelMarketing, 2011). It is attractive because it has reviews on hotels, destinations, videos, photos so that people can evaluate various alternatives and compare them (Chung and Buhalis, 2008).

In peer-to-peer hosting, the importance of customer reviews is even bigger than for conventional hotels (Dredge and Gyimóthy, 2015). This happens because the hosts of peer-to-peer (e.g., Airbnb) accommodations are micro-entrepreneurs who do not promote their accommodations on television, radio or other media such as hotels. Thus, the only channel to their customers is the online platforms (Litvin et al., 2008).

It is impossible to deny the increasing influence that the UGC has on travelers and the entire tourism industry (Papathanassis and Knolle, 2011). Thus, hospitality and tourism organizations must acquire tools capable of responding adequately to the impact that the UGC has held on the market, namely to this new behavior of travelers during the planning of their trips (Litvin et al., 2008). Business models must be modeled in order to develop or adapt their online presence (Ye et al., 2011). There is increasing knowledge about travel details and stays; access is increasingly broad, and sharing of travelers' perceptions about certain travel products (Litvin et al., 2008). Therefore, Starkov and Price (2007) recommend that this sharing of experiences should be integrated into the corporate marketing strategy, as the Sheraton Hotels and Resort group did by making the focus of their main online page the customer reviews of attracting the attention of other potential guests.

Travel services and accommodations must provide resources, such as new technologies, to gain some advantage in the travel industry (Litvin et al., 2008). The understanding how to use online travel reviews in their favor, by allowing customers to post and receive feedback, can provide more bookings and an improvement in services rendered (Ye et al., 2011).

2.5.3. Numerical rating

Another way to evaluate customer satisfaction, apart from online reviews, is numerical classification usually through stars (Gerdes et al., 2008). This is a quantitative alternative to evaluate the quality of service quickly and efficiently. According to the heuristic model of information processing, this option is designed for individuals seeking information without details, with lack of motivation, or without enough cognitive resources. The service can then be evaluated quickly, searching only for results appropriate to its needs through classifications (Fiske, 1992).

The given classification is usually by norm the generalization of the opinion reported by a review. Each classification has behind it several relevant (hidden) topics whose “average” of satisfaction has been translated into a number. Thus, this method of service evaluation becomes insufficient to determine customer satisfaction because it does not show the perceptions and attitudes of the customer (Luo and Tang, 2019). The hidden topics (e.g., price, location, cleaning, comfort) should be studied, essentially by peer-to-peer accommodations given that the relevant topics are heterogeneous. For example, an accommodation may have a rating of 4 for having a good location but an exaggerated price, while another may have the same rating for having a more affordable price but a more unfavorable location (McAuley and Leskovec, 2013). However, both methods of evaluation are important because they complement each other. According to Osborne (2019), on the Amazon website, some products of inferior value were appearing highly praised, making star rating effectively meaningless. Some items evaluated as poor products were exposed side by side with positive reviews regarding other items of better quality, influencing consumers to buy the worst items without being aware of it. Examples of this are the poorly translated versions of certain classic books, with notorious errors of writing and lack of coherence in sentences as well as the critically panned remakes of Hollywood films. These appeared classified as four\five stars when they actually could not even reach three stars according to the corresponding reviews.

2.6. Big data

Nowadays, people are spending more and more time and money online. In fact, what was once bought in person is now is purchased online. Thus, opinions on products and services are shared online, which can influence the opinion formatting of future consumers. Given the scale and growth of data on online purchases and usage, it is a challenge for companies to use such data in order to gain competitive advantage (Lau et al., 2005). Online data is becoming a popular way to understand the image and experience consumers have of a particular brand or product because they are not limited to quantitative information.

The set of these large-scale data available in various sources and media such as web pages, mobile transactions, and social media is called “big data”. It aims to generate new insights (usually in real-time) in order to complement other more conventional, generally static data sources (Xiang et al., 2015). Big data is usually characterized by three variables: volume, variety, and velocity. Volume is quantity of data and information, variety is the type of data, and velocity refers to the speed of data generation (Xiang et al., 2015).

The big data analysis is integrated into Business Intelligence (BI). It is used to better understand brand products, their competitors, customers, the entire market environment, technological impacts, and strategic plans (Xiang et al., 2015). These data are essential for the development of new knowledge, such as being able to create predictive models of customer relationships through the extraction of patterns and thus gain a better understanding for decision making (Chen et al., 2014).

Given that most data is available in an unstructured or semi-structured format, it is necessary to use data mining and econometric techniques to identify and extract useful information through large-scale textual documents and to be able to attribute causality and patterns (Rogers and Sexton, 2012). However, the experience and skills of many professionals on big data analysis, especially in the marketing area, is still little.

Researchers are very attracted to the media area since it offers a vast amount of data, which also appeals to the public that uses this information (Piryani et al., 2017). Tourism and hospitality industries are the frontrunners, with the number of postings on websites, forums, and referral systems, capable of generating insights that allow knowing and evaluating opinions, preferences and evaluations of guest experiences (Xiang et al., 2015). However, the methods and strategies to be adopted for Business Intelligence are still scarce, in relation to social media and referral systems (Xu et al., 2017). Park et al. (2012) have defined an inference structure driven by social networks to draw reliable profiles on their clients. He et al. (2013) analyzed Facebook and Twitter content about restaurant business through text mining. These examples are advances in technology to achieve commercial value from the user-generated content, in order to find advanced methods that allow exploring its potential in a systematic and effective way (Liu, 2012).

2.6.1. Text Mining

The rapid growth of big data has become a challenge for business intelligence, especially the analysis of unstructured texts (Chen et al., 2014).

Text mining, belonging to Text Analytics, is a semiautomatic process of extracting significant patterns from large amounts of unstructured text and transforming them into structured formats (Sharda et al., 2014). Basically, it summarizes a document into key numerical concepts, according to the similarities between them and so that they can later be analyzed to reveal relevant common properties. Typical assumptions of the text mining statistical model include word count analysis, probability model analysis, and concurrent frequency analysis (Chen et al., 2014).

Some researchers (e.g., Chau et al., 2007) considered text mining as an extension of data mining, since the purpose of both is the same but differ in the type of data to be processed. Data mining analyzes structured data and stored in databases, while text mining analyzes data in natural language texts and it is necessary to extract significant numerical indices (patterns identification) of unstructured texts in order to "transform text in numbers". Both are intended for document preprocessing and pattern discovery techniques in order to inform trends and significant features on specific topics (Nasukawa and Yi, 2003).

Text mining has the ability to organize large amounts of data that are available inside and outside organizations from consumer-generated content (including reports of their experiences with specific products and services). It can provide the knowledge needed to solve real-world problems (Godbole et al., 2010) and offers numerous benefits to organizations, especially those that deal with large amounts of textual data daily, such as catering and hospitality. Companies can understand consumers' opinions about the company itself and its products in order to improve the marketing strategy and to position itself against its competitors (Jain et al., 2013).

In the area of hotel and tourism, many earlier studies have focused on using user-generated content to understand the impact of ratings on hotel sites and customer feedback (Ye et al., 2011). However, in the past decade, more studies have been undertaken to examine consumer reviews based on big data sources (Chen et al., 2014; Sparks and Browning, 2011). The analysis of the online data can be much more valuable than those analyzed by traditional statistical and econometric methods because it provides a wider characterization of the consumer experience. It is an essential source for understanding both consumer satisfaction and determinants of consumer satisfaction (Sparks and Browning, 2011) since it is possible to determine the sentiment behind the reviews.

Many companies performed the analysis manually, but it has been proven to be a time-consuming and inconsistent process (Lau et al., 2005). They began using text mining techniques to make the process more automatic and to obtain more accurate information to enable them to make well-informed decisions (Fenn and LeHong, 2012). Although this has a cost of deploying software, it is the best method proposed for managing information embedded in hotel databases.

2.6.2. Sentiment Analysis

Sentiment refers to attitudes toward something negative, neutral, or positive. The term "sentiment analysis" emerged in 2003 (Nasukawa and Yi, 2003) and may also be termed emotional polarity analysis, review mining, subjectivity analysis, opinion mining, and appraisal extraction (Liu, 2012; Sharda et al., 2014).

Sentiment analysis is a process that automatically detects and studies the emotional content, i.e., opinions, feelings, evaluations, and emotions in relation to products, services, individuals, organizations and other topics (Liu, 2012). From this analysis, the polarity is determined, usually dichotomized in positive or negative, or by a range of values (]0,1] if positive and [-1,0] if negative) (Cambria et al., 2013). Nasukawa and Yi (2003) argue that the analysis of feeling is not only to classify a document as positive or negative but also to identify and understand the feeling attached to fragments of text. However, there is no definition of sentiment analysis that is accepted by most of the scientific community (Sharda et al., 2014).

Although, Pfister and Böhm (2008) indicate that it is dogmatic to map many emotional dimensions (e.g., confidence, surprise) into only one dichotomy or trichotomy (includes neutral position), each of the existing emotional dimensions has a distinct origin, meaning, and evaluation, which cannot be easily summarized. Therefore, Ghazi et al. (2010) suggest an approach to multidimensional emotions as an alternative to dichotomy or trichotomy in the sentiment analysis. Thus, Ekman (1992) proposes six fundamental aspects of emotion: anger, fear, sadness, disgust, joy, and surprise. Later Plutchik (1994) expanded this structure by adding two more dimensions: anticipation and trust. With eight sentimental dimensions, Plutchik (1994) describes them like a wheel, where negative emotions can balance with positive emotions: sadness-joy, fear-rage, repulsion-confidence, and startling-discovery.

The feeling has unique characteristics that allow it to be identified in the text. The process goes through two consecutive phases: analyzing the text segments and grouping them into topics and taxonomies; then classifying them according to two classes (positive and negative) and the range of polarities (Pang and Lee, 2004; Sharda et al., 2014). In addition, in order to calculate the feeling of the obtained text, it is necessary to compare it with a lexicon or a dictionary that determines the strength of feeling (Mostafa, 2013). It is important for companies to acquire professionals and tools to work on this analysis because every opinion posted online by an individual or company will have positive or negative connotations that will have repercussions on the reputation of the entity in question. Researchers are interested in people's feelings about specific topics because their understanding can be used to monitor customer experience in certain services or products in order to improve customer experiences (He et al., 2015). The existing literature on the polarity of feelings allows us to identify different applications: product reviews (Cambria et al., 2013), commentary on films (Mostafa, 2013), extraction of political orientations (Stieglitz and Dang-Xuan, 2013), forecasting of the stock market (Wong et al., 2008), product analysis (Nie et al., 2013) and track feelings trend (Gloor et al., 2009). Feelings can also be evaluated according to three different levels: in an entire document, in an individual sentence or aspects of an entity (Liu, 2012). The sentiment analysis at the aspect level are the most used, especially in the area of hospitality, because they approach users' opinions, experiences or concerns about the various dynamics of the business, being able to register behavioral patterns of the clients (Pan et al., 2007). For example, Guo et al. (2017) have been able to extract the dimensions that influence guest satisfaction and dissatisfaction according to online reviews.

Customers' textual feedback is one of the most important sources of information because they have more complete content, despite being unstructured, and allows for the identification of feelings and motivations regarding various hotel departments, so as to understand how their characteristics can influence overall customer satisfaction, helping in the delineation of strategies of Customer Relationship Management (CRM) (Ordenes et al., 2013; Belkahla and Triki, 2011). When extracting the text from written comments it is possible to classify them as positive or negative according to the polarity of the review (Cambria et al., 2013). Gan et al. (2016) report the sentiment analysis of online comments on a restaurant, which analyzed the feelings towards food and services provided along with prices. This analysis allowed restaurant managers to adjust prices according to the results of the analysis.

In the areas related to social network and online shopping, it helps in the development of marketing strategies by evaluating and predicting consumers' attitudes towards a brand (Medhat et al., 2014; Cambria et al., 2013). One example of its application is reported in Salehan and Kim (2016), which has been able to predict the number of users and the usefulness of a certain product through online customer reviews on Amazon.com. Twitter that is the most used social network for the analysis of feelings because it allows extracting and evaluating conversation patterns (Lipizzi et al., 2015).

In the area of hospitality and tourism, literature is still scarce. However, there is a growing concern from hotels and restaurants, as tourists use this information to purchase goods and services as well as select hotel units, which can in real-time monitor their performance and customer satisfaction (Lau et al., 2005; Grant-Braham, 2007). McAuley and Leskovec (2013) indicate that textual analyzes such as feelings analysis can be used to discover the implicit feelings of a client about the key topics or attributes of a hotel.

The classification of feelings can be divided into two categories: lexical-based approaches (through a pre-defined lexicon dictionary of feelings) or approach based on machine learning algorithms (such as neural networks and N-gram character-based model) (Lawrence, 2014; Medhat et al., 2014). These tools use data found in lexicographic resources to attribute feelings to many words. Together with a set of algorithms, they determine the polarity of a given document, classifying it as being positive, negative, or neutral. The first one is simpler and easier to use; however, the lexicon may not appear in the text of interest or may be used in a peculiar fashion. The second one can be built to a particular text, capturing the peculiarities of the language used in it (Paltoglou and Thelwall, 2010). Although the first is the least used, there have been some relevant studies where lexical-based web-service tools have been applied to online reviews. Xie et al. (2016) analyzed TripAdvisor's reviews with three different tools on four different hotels, by crossing an existing dictionary with the different characteristics of the hotels (e.g., room, staff, and location). Another study carried out by Pekar and Ou (2008), through 268 reviews from a hotel evaluation website, also allowed to analyze the feelings expressed regarding variants such as food, room service, facilities, and price. For this type of process, dictionary-based and auto-learning techniques were essential. Through lexicographic resources, certain tools are able to attribute feelings to a large number of words, determining their polarity and classifying them as positive, negative or neutral (Lawrence, 2014). There are also tools that provide a more detailed analysis and can identify emotional feeling such as happiness, sadness or anger; while others analyze the specific score of feeling.

When evaluating the positivity or negativity of a product or service, one quickly understands whether customers have a good opinion about it or not. However, it should be explicit that this view may be different according to gender. According to studies carried out on the difference that gender makes in feelings expressed in social networks (Volkova and Yarowsky, 2014), terms related to emotion (such as love and joy) are used in an unequal way by women and men. It can be concluded that the words used to express feelings about the same topics differ according to gender (Volkova et al., 2013). The evaluations of male authors are more direct and professional, with positives and negatives points in all reviews. This makes the analysis of feelings more difficult since these feelings are not so explicit. Also, more sarcasm is found, along with figurative language in order to disguise opinions in their views. Female author reviews are more extreme in feelings (very positive or very negative) using positive and more explicit language than men. A study by Thelwall (2016) on the reviews made on hotels by men and women was carried out. In positive reviews, women use simpler and more positive language, loaded with adjectives (such as adorable, delicious, incredible, beautiful) while men use factual words like beer, location, and construction. Both males and females value quality and low prices. In negative reviews, women presented simple and explanatory words, but always of an evaluative character, while men only indicated aspects to be avoided.

The analysis of feelings contains innumerable applications. Effectively, this tool is an excellent method for extracting views from unstructured documents and obtaining valuable data through business intelligence instruments. In addition, by tracking public opinion, it contributes to brand reputation and management (e.g., Pang and Lee, 2005) in order to define the best strategy for retaining and managing customer relationships (CRM) (Karakostas et al., 2005).

2.7. Previous studies

Cheng and Jin (2019) used text mining to investigate which aspects influenced the experiences of Airbnb users embedded in online reviews (big data). To analyze the data, they started to carry out text mining to identify the semantic and relational insights of the reviews. Then they identified themes and concepts previously studied in the literature. This was performed with the software Leximancer. The last stage was the sentiment analysis, so as to identify guests' positive and negative feelings related to the attributes identified earlier.

They concluded that the evaluation described in the reviews was made in comparisons with experiences customers had in hotels and four key aspects were highlighted: location, amenities, host, and recommendation (in this order of importance). The price was not identified as a factor influencing satisfaction, because the preference was for a location not only close to transport, shopping areas, and city, but also a good place for the family; a friendly and helpful host and the basic amenities that made a house enjoyable. The negative reviews highlighted noise coming from outside, shower, parking, and not being welcomed by the host.

Luo and Tang (2019) carried a study to identify the dominant aspects and the emotions incorporated in the textual reviews and to understand their influence on the general (numerical) classification. The reviews were from Airbnb in Los Angeles. They used the modified Latent Aspect Rating Analysis (LARA) approach to assess the relative importance of each aspect of an entity (i.e. an Airbnb listing) contributing to the overall rating and to identify the latent sentiment rating of each aspect, which contributes to the overall rating. Regarding the feelings, they used the NRC Emotion Lexicon to test the polarities of feeling and extensions of eight emotion dimensions defined by Plutchik (1994). Latent Classification Regression (LCR) was used to generate the overall classification of each aspect for each Airbnb listing. They managed to obtain five main aspects that stand out: communication, value, product\service, location, and experience. Communication is relative to the interactions between host and guest throughout the process (from reservation to the end of stay). Value refers to price and recommendation (price-quality ratio). Product\service matches the amenities available. Location includes distance to the city and access to transports. Experience covers service, environment, and neighborhood. Location was the most significant aspect and the value the least. As for feelings, these were mostly of joy and surprise. Joy was associated with product/service, value, and location and surprise with communication and experience.

After reaching a better understanding of these business intelligence tools and their applications, the second research question is raised: Can text mining techniques and sentiment analysis detect patterns and polarities in the text, capable of providing value for decision-making?

Chapter III: Methodology

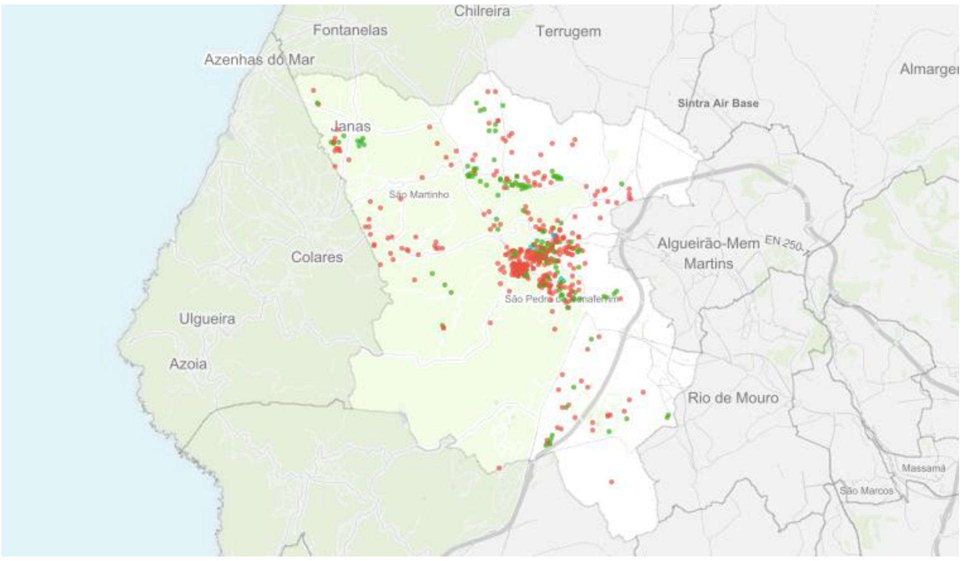
3.1. Population and sample

The search is based on guests' reviews for accommodation on Airbnb in Sintra County, comprising 3 main areas: Santa Maria e São Miguel, São Martinho, and São Pedro de Penaferrim. This choice is due to the fact that this area is very attractive to locals and tourists, therefore allowing different perspectives to be studied. Santa Maria e São Miguel represents the touristic area (where downtown Sintra attracts thousands of tourists). São Pedro de Penaferrim represents a more local area, linked to other parishes in more remote areas (most traditional areas, far away from crowded areas, full of tourists). São Martinho is near the mountains and by the sea; in a calmer and not so touristic area (it contains well-known beaches and natural beauty). This division will allow us to study the differences between the three areas, and what is most valued by Airbnb guests, according to the location they choose, then analyzing whether Airbnb guests prefer more remote and traditional areas to have an experience similar to the locals.

3.2. Data collection

To extract online reviews and all the data of the various lists of Sintra's accommodations, we used the Airbnb Inside website. It shows that Sintra's county currently contains 531 listings, 68.9% of which are whole houses or apartments with an average price of 109€ per night, 29.4% private rooms and the rest are shared rooms. However, not all of these accommodations contain guest reviews, because there has been no registration made by Airbnb in these accommodations, and the total number also varies, as some accommodations will pop up while others are flagged as unavailable.

Figure 1 shows the location of the various accommodations along the desired area and differs according to the type of housing (red corresponds to houses, green corresponds to private rooms, and blue to shared room). It is possible to observe that most accommodations are located in São Pedro de Penaferrim and Santa Maria e São Miguel (encompassing touristic and local areas, essential for the present study).



Source: Airbnb (2019c)

Figure 1. Map of Sintra's Airbnb accommodations

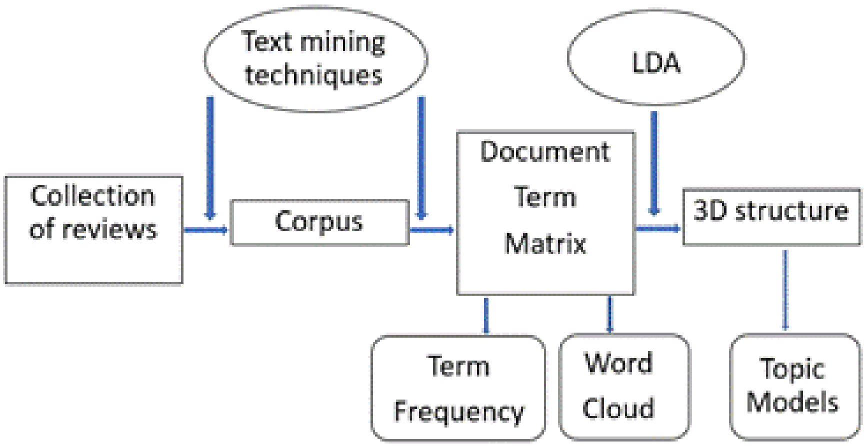
In addition, according to the site, it is estimated that the period of stay in this area is 92 nights per year, with an occupancy rate of 25.2%. Most hosts (about 70%) are owners of several accommodations and not just one, and some even have eight properties in their name. The number of total accommodation reviews is 15 902, this being a growing number.

Various indicators were collected concerning each accommodation: type, price (a simulation of the reservation was made to obtain the price per night[1], cleaning price and service price), amenities, and general classification of the accommodation. The name of the host was also withdrawn and whether s/he is a superhost, since according to the literature studied, this is a factor that influences the confidence and can impact guest satisfaction (Zhang et al., 2018). Only comments in English were extracted to facilitate the analysis. The accommodations that were inactive did not have English reviews and did not have all the information needed (e.g., number rating, description of the amenities, or guest nationality) or shared rooms were excluded.

[1] It is important to note that for some households the price varies according to the height of the year. In that case, the value is at time of data collection (low season).

3.3. Data analysis

Figure 2 shows the steps of the approach to be followed in order to analyze the unstructured text related to customer evaluations. It starts by gathering the information into an Excel file. Then, text mining techniques are applied to clean the text of irrelevant words. After obtaining a corpus, text mining techniques are reapplied to analyze it and obtain a frequency table of terms and a word cloud. In the last step, Latent Dirichlet Allocation is applied to obtain a 3D structure capable of providing the most relevant topics.



Source: Own creation based on Feinerer (2018).

Figure 2. Methodologic strategy

3.3.1. Text Mining

The Excel file is later transformed into a .csv (Comma-separated values) file so that one can analyze the unstructured texts in the R statistic tool. To convert the text into a corpus (collection of text documents) and be able to perform text mining procedures, the "tm" package in R was adopted (Feinerer, 2018). This package has several functions, which allow the conversion of unstructured into structured data by reducing dimensionality of data and still keeping relevant information. Quantitative and qualitative data are analyzed through several phases (Tomar, 2017):

- Removal of punctuation;

- Removal of whitespaces;

- Removal of stopwords;

- Text conversion to lower case, so that words like “host” and “Host” are considered the same word for analysis;

- Removal of numbers;

- Application of the stemming process;

- Removal of words that are irrelevant (e.g., hostnames).

The removal of stop words is based on the elimination of prepositions, interjections, auxiliary verbs and other terms irrelevant to the analysis. This prevents those words from appearing as the most frequently used words listed; thus, it allows obtaining a correct analysis of the core words used in the text. Some common words were not removed this way; thus, they had to be removed by a specific manual code. The stemming phase allows reducing the words to a single root word (including different verb tenses with the same semantic meaning), to avoid being classified as different (for example, “booking” is reduced to “book”). To perform all these transformations, the function applied to all elements of the corpus was the tm_map() function (Feinerer, 2018).

After obtaining a cleaned up corpus and representative of the core set of relevant words, the next step is to create a Term-Document Matrix (TDM) or a Document-Term Matrix (DTM). These mathematical matrices describe the frequency of terms that occur in a collection of documents and may differ according to whether terms are rows and documents are columns (TDM) or backward (DTM) (Feinerer, 2018). They require two extensions: reviews (usually text mining is performed on the documents) and each of the terms considered; each cell contains information about how often each term occurs in each of the reviews.

To reduce the size of the term-document matrix and remove the terms that occur in very few documents (sparse terms), the function used is the inspect(removeSparseTerms()). This allows reducing the data set without losing significant relations inherent to the matrix.

For the analysis and interpretation of the term-document matrix, a frequency table was obtained (which describes the number of occurrences of each term) and a cloud of words (a way to visualize the text corpus and understand the frequently used words) (Tomar, 2017).

3.3.2. Latent Dirichlet Allocation

In order to generate the most relevant topics, we adopted the topicmodels package, which that uses the document term matrix served as an input for the basic infrastructure for the assembly of thematic models (Feinerer, 2018). This package implements the latent Dirichlet allocation (LDA). It tries to discover the hidden semantic structures represented by abstract topics within a collection of documents, based on Bayesian analysis (Bernhard Learns, 2017).

The LDA is a three-level hierarchical Bayesian modeling process that groups items into topics (words or terms that have been identified) and the probabilities that characterize each topic (Blei, 2012). This occurs through the β distribution, which analyzes the relationship between the theme and the term identified. All values of β are positive and close to zero. These values reveal the strength of the relation between the term and the corresponding topic. The LDA algorithm is usually calculated with three dimensions (terms, reviews, and topics) but can be calculated with only the desired number of topics and the document term matrix created for text mining. It was also used an upgraded version of LDA, the Correlated Topic Model (CTM) to construct the topic models (Bernhard Learns, 2017).

Chapter IV – Results

4.1. Sample characterization

The data set, consisting of 2444 comments and 123 accommodations, was extracted from the Inside Airbnb platform to an Excel file. Regarding the gender, 1401 comments are from females, 991 are from males and 52 are from couples (female and male). Accommodations are classified into 72 apartments and 31 private rooms. Regarding the place, there are 1060 comments from São Pedro de Penaferrim, 820 from Santa Maria e São Miguel, and 564 from São Martinho. Guests are mostly from the United States of America (526 comments), the United Kingdom (373), Canada (210), Germany (184), the Netherlands (151) France (138), Australia (92), Spain (74), Portugal (73), and Belgium (62).

4.2. Text mining

Figure 3 shows the number of occurrences for each of the twenty most used terms, obtained through the text mining procedure. It can be seen that the top 5 of the most frequent words is composed of the following words: “place”, “Sintra”, “stay”, “great”, and “house”. It is immediately visible from the beginning, the importance of the location, either by the word “Sintra” or even by “locat” and “walk”, which confirms previous studies. Another aspect to consider is the fact that there are several positive adjectives, present in the top five (“great”) and within the 20 most frequent words (“love”, “beauti”, “clean”, “perfect”). This shows the presence of positive feelings in most of the comments. Furthermore, other very common words regarding the type of accommodation and its amenities are: “place”, “house”, “apart”, and “room”.

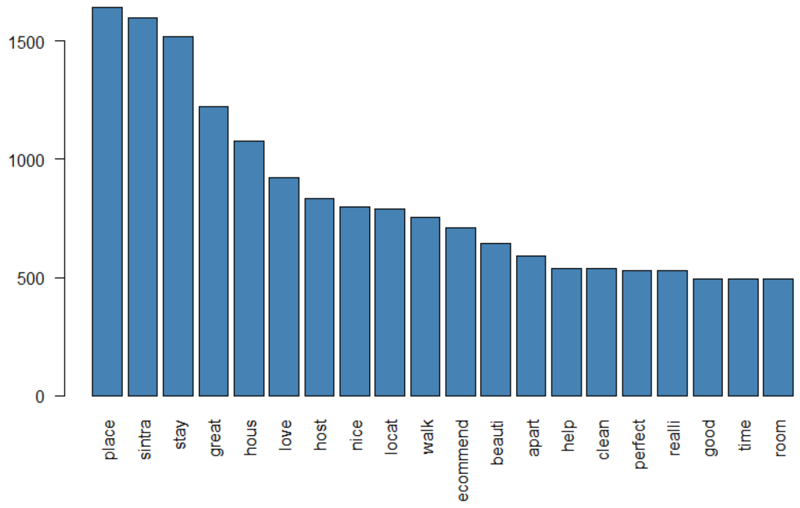


Figure 3. Term Frequency

Four clouds of words were obtained, one covering all the comments and three for each of Sintra regions (São Pedro de Penaferrim, Santa Maria e São Miguel, and São Martinho). The goal is to be able to analyze the attributes that are most important to guests by their chosen location.

Figure 4. Sintra World Cloud

Figure 4 is a cloud of words, presenting in a more visual way the most frequent words throughout all the comments. The highlighted words are the same as in Figure 3; however, there is a broader range of terms. Thus, in general, the attribute that most guests value during their stay is the "host" and it is the most used word in reviews followed by "location".

Regarding location, other connected words are "restaur", "park", "car", "visit", and "close". This expresses a certain concern about the environment around the accommodation, and the need for transport to move around; however, there is no specific word that refers to tourist attractions. Another term that gained more prominence was the host and other words like "feel", "welcom", "friend", "thank", "experi", and "help", all related to her/him. This result reinforces that the second attribute that most influence the satisfaction of an Airbnb customer, location. It is again observed that there are several adjectives related to the experience itself ("wonder", "perfect", "enjoy", "great", "well", "love", "amaz", "recommend", and "good") and some comments related to the environment itself and surroundings (e.g. "comfort", "clean", "area", "quiet", "beauti", and "view"). The word "home" appears in the wordcloud for the first time and brings us back to the Airbnb mission, which is to provide local and traditional experience, making guests feel as if they were in their own home, one of the great differentiating factors from a hotel room.

Figure 5. São Pedro Penaferrim Word Cloud

Figure 5 illustrates the wordcloud for <u>São Pedro Penaferrim</u>. This is the most traditional area, not far from the attractions, where it is possible to get a local experience. In general, there are many adjectives and words that appeal to the positivism of the experience ("enjoy", "beauti", "good", "apart", "wonder", "need") and others that appeal to the host ("welcome", "thank", "help", "friend"). It should be highlighted a concern for location and neighborhood ("park", "restaur", "area", "quiet", "farm"). What differs from others is the existence of words that refer to access to attractions (e.g., "visit", "walk", “close”, “car”, “get”), which reveals that even choosing this type of stay, there is no total indifference to the touristic area.

Figure 6. Santa Maria e São Miguel Word Cloud

Figure 6 shows the wordcloud for Santa Maria e São Miguel, the center of Sintra and its attractions. In this cloud, new words with great distinction appear, such as “locat”, “host”, and “walk”. In addition, there are words related to the peripheral and tourism, such as “centr”, “castl”, “town”, “restaur”, and others (e.g. “close”, “minute”, “train”, and “station”) that also facilitate the visit to the capital, Lisbon (an aspect that has proved to be important in some reviews). Apart from this, adjectives and words of recommendation continue to be noticeable.

Figure 7. São Martinho Word Cloud

Figure 7 shows the wordcloud is related to the mountainous neighborhood of São Martinho. Some words are highlighted such as “farm”, “natur”, “live”, “interest”, and "work". The host is again mentioned as is the word "people". There is still a concern about location and access ("car", "area", "walk"), although noticeably less. This area provides an experience that differs from the others; however, it is possible to detect positive feelings by the prominence of the adjectives ("beauti", "nice", and "experi").

In all clouds of words, it is possible to observe the unanimity of the most used words ("stay", "place", "great", "Sintra"), which refers to the positive feelings regarding accommodation and the Sintra area.

4.3. Latent Dirichlet Allocation

It was used three topics for the topic analysis (construction of a topicmodel) [2]. Figure 8 shows the terms that are most common within each topic, where β is the probability of each term per topic.

The first topic essentially relates to feelings and the host. Most of the feelings are positive and related to the whole experience. Sintra proves to be a perfect place for accommodation, regardless of the chosen area due to the most common words in this topic: "stay", "great", "place", and "Sintra" (they have the highest beta). The host is definitely one of the most valued aspects, essentially for his/her availability, help, and welcome.

[2] In the first try, 10 main terms were shown per topic from a model with 9 topics. However, it was difficult to name the dimensions due to the lack of correlation between the mains terms. In the second attempt, six topics were chosen, but the problem remained. The solution was to keep only three dimensions with 15 main terms, where it was visible the relation between the main terms and it confirms the previous results obtained. Then proceed to the labeling or assignment of a distinct topic-by-topic, giving more relevance to terms with higher beta. At this stage, it is also possible to analyze which words are much more probable in one topic compared to another. In the end, the topic analysis used three topics.

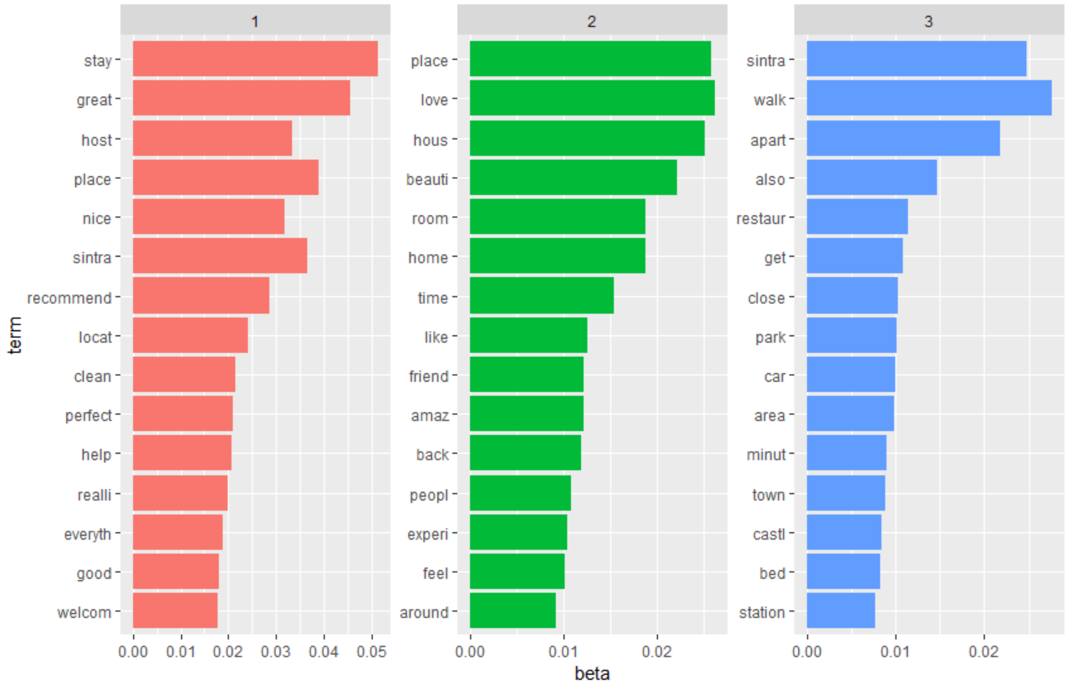


Figure 8. LDA results: Topic model

The second topic is related to accommodation and its characteristics. The words with the highest beta and therefore most common in this topic are: “place”, “love”, “hous”, “beauti”, “room”, and “home”. This last word is very common in every area, revealing the closeness and sense of belonging that one normally feels when being in another person’s house.

The third topic refers to location and neighborhood. Nearby touristic attractions have a lot of influence, as well as other outdoor activities (e.g. farm and restaurant). Access to them, including the means of available transport (including car or car rental) and the time each requires, is very important as well.

4.4. Characterization of Airbnb places

Table 1 provides key summary statistics of the variables available in the data set at the apartment/room level. LDA can also model each document as a mixture of topics. We can examine the probability of each topic per document, which is given by γ (“gamma”). The mean of topics 1, 2, and 3 refer to the gammas of LDA, it means that approximately more than 30% (gamma = 0.3) of the words in the document were generated from each topic since the three of them have similar gammas. In the reviews of São Pedro de Penaferrim and Santa Maria e São Miguel, it is possible to observe more words coming from location (topic 3 with higher gamma), while in the reviews of São Martinho most of the words are related to accommodation (topic 2). In all areas, there is an almost excellent classification, even though there is a great difference of average price practiced and in São Martinho there are not so many hosts considered superhost compared to the other two areas. This is justified by the fact that most of the accommodations in this area are whole houses (and not rooms) with more amenities compared to other areas. The cheapest area is the central area, followed by the more traditional zone and finally the mountainous zone. The same happens with what concerns the average price of the service and differs in the cleaning level, where São Martinho practices lower prices than São Pedro de Penaferrim. The fact that São Martinho is farther from the center is compensated by the fact that it offers, on average, more amenities.

Table 1. Descriptive statistics of the Airbnb places

	São Pedro de Penaferrim		**Santa Maria São Miguel**		**São Martinho**		**Total**	
n	49		22		13		84	
%	58.3		26.2		15.5		100.0	
	Mean/prop.	s.d.	Mean/prop.	s.d.	Mean/prop.	s.d.	Mean/prop.	s.d.
Topic1	0.333	0.024	0.337	0.020	0.323	0.028	0.332	0.024
Topic2	0.317	0.028	0.321	0.030	0.365	0.043	0.326	0.035
Topic3	0.350	0.032	0.342	0.025	0.312	0.036	0.342	0.033
Type	0.673		0.636		0.692		0.667	
Superhost	0.469		0.591		0.077		0.440	
Classification	4.673		4.705		4.500		4.655	
No. of comments	21.837		37.273		43.231		29.190	
Price	99.306		55.909		152.769		96.214	
Cleaning price	38.469		12.727		26.308		29.845	
Service price	55.980		28.500		170.077		66.440	
no. of amenities	18.735		17.136		20.923		18.655	

Chapter V – Conclusion

This study aimed to understand the unstructured content that is published in digital media (in this case, reviews), through text mining techniques, in order to achieve better decision making. The dimensions that most affected Airbnb customer satisfaction were identified through the R software, beginning with the first phase of cleaning and processing the data and the second phase extract the frequent terms and comparing through different functions.

In order to obtain a faithfully representative matrix document, in the first phase, it was necessary to remove specific words through due to a large influx of unnecessary terms for analysis, such as host names. It was also considered replacing these names with the word "host" but it was noticed that even without this, the host was already a very prominent factor.

In the LDA part, given the large sample of comments, it became difficult to group terms into similar topics, thus showing a good correlation. The terms always seemed much dispersed from each other, so it was necessary to analyze with a smaller number of topics but easier to relate them to each other.

After analyzing the results, it is possible to compare them with the studied literature and answer the first question concerning which factors influence customer satisfaction and its order of importance. The most outstanding is the location and the host.

The location was, without any doubt, considered the most valued aspect by customers. They described Sintra, the area where they were staying, and other specifics worries such as the most convenient transportation to reach tourist attractions, the nearest supermarkets, and restaurants, as well as location-related safety issues. This way, we can distinguish three types of people in relation to location: (1) those who really preferred to be near tourist spots and were not concerned about having a traditional experience; (2) those who valued a more local place and away from the confusion, yet shared information on the means of transport (and prices) for both the Sintra tourist area and Lisbon and (3) the others who were generally families with children that rented whole houses with many amenities, so they can relax and enjoy their stay quietly, without much interest in the sights.

The host was the second most important factor mentioned in almost every comment. Confidence and friendliness are influencing customer satisfaction and consequently boosting positive reviews. The guests praised the communication that had taken place if they had been well received, if the host had given tips on places to visit and if they had been accessible to any existing questions or problems. It is important to note that replacing hosts names in comments by the word “host” could have caused impact on the ranking of the importance of the factors.

The third factor highlights the accommodation itself and the amenities. There was a great concern on the part of the guests that the expected level of service was the same as the one delivered, i.e. that the amenities described on the site did match the ones available. For example, one of the apartment claimed to have outdoor pool on the platform, however when guests arrived at the apartment, the pool was not functional. In addition, the accommodation had to be clean, organized and presentable (tidy and not old).

The fourth factor is related to the traditional experience: to “live as a local”. The great essence of Airbnb was described in the comments when they named the accommodation as “home” and it is a big gap between this platform and hotels. However, this factor is dependent of the previous ones because guests can only have the whole local experience if the location should not be close to tourist areas but also not too far away; hosts are empathetic and receptive and amenities are the same as promised.

Price is the last factor, as opposed to the literature studied. Although the suited prices were very different from each other with the intent to compare them, they were rarely mentioned in the reviews. These results may result from the fact that comments are positive. Some negative comments, whose customers were dissatisfied, emphasized that the price paid did not match the service provided. However, many transport prices were mentioned regarding taxi and bus, mainly in the comments from the less centered accommodations.

The analysis of feelings could not be performed because the results obtained were inconclusive, there was not enough sample to support them. Only 19 comments out of 2444 were negative and even the accommodation’s ratings were high (at 122 properties, 116 were greater than or equal to 4), with the lowest score being 3. However, individual reviews were read to understand the reasons that caused the most dissatisfaction: some guests in farther accommodations emphasized that it was impossible to move without a car; others near the touristic spots complained about the noise from outside and traffic (difficult to park); lacks of soundproof of rooms and house as there were annoying noises from neighbors (eg barking dog); poor communication from the host and wrong location at the website and consequently guests got lost. Some minor issues and/or easy resolution were soon solved by the hosts and that avoid dissatisfaction and negative comments. This is in line with what has been studied in the literature.

Even with the obstacles faced, the answer to the study's second question regarding the ability of text mining to provide value for decision making is affirmative. By having a more depth knowledge of the reviews' content posted on the platform, hosts can understand what their customers (in this case, guests) are looking for and try to meet their expectations and needs more easily. Thus, strategies and decisions can be outlined with more precision and effectiveness

Bibliography

Airbnb (2019a). About us. Retrieved from https://press.airbnb.com/about-us/ Acessed: on 3rd february 2019.

Airbnb (2019b). Trust and security. Retrivered from https://www.airbnb.pt/trust Acessed on 3rd february 2019.

Airbnb (2019c): Inside Airbnb Lisbon. Retrivered from http://insideairbnb.com/lisbon/ Acessed: on 5th january 2019.

Airbnb (2019d). What are service fee. Retrivered from https://www.airbnb.pt/help/article/1857/what-is-the-airbnb-service-fee Acessed on 3rd february 2019.

Airbnbcitizen (2014). Airbnb economic impact around the world. Retrieved from https://www.airbnbcitizen.com/airbnb-economic-impact-around-the-world/. Accessed on 11th November 2018.

Airbnbcitizen (2016). New research on Airbnb and hotels. Retrieved from https://www.airbnbcitizen.com/new-research-on-airbnb-and-hotels/. Accessed on 11th November 2018.

Bardhi, F., and Eckhardt, G. M. (2012). Access-based consumption: The case of car sharing. *Journal of Consumer Research*, 39 (December), 881–898.

Baudouin, J. Y., Gilibert, D., Sansone, S., and Tiberghien, G. (2000), When the smile is a cue to familiarity, *Memory*, 8(5) 285–292.

Bernhard Learns (2017). Topic Models (LDA and CTM) with R and topicmodels. Retrivered from: http://www.bernhardlearns.com/2017/05/topic-models-lda-and-ctm-in-r-with.html Accessed on 20 th May 2019.

Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), 1595–1600.

Bellotti, V., Ambard, A., Turner, D., Gossmann, C., Demkova, K., and Carroll, J.M. (2015). A muddle of models of motivation for using peer-to-peer economy systems. In: Proceedings of CHI 2015 (pp. 1085-1094) Conference on Human Factors in Computing Systems, Crossings, South Korea.

Belkahla, W. and Triki, A. (2011). Customer knowledge enabled innovation capability: Proposing a measurement scale, *Journal of Knowledge Management,* 15(4), 648–674.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model, *MIS Quarterly*, 25(3), 351-370.

Blei, D.M. (2012). Probalistic topic models. *Communications of the ACM*, 55(4): 77-84.

Boone, L. E.; Kurtz, D. L. (1998). Marketing contemporâneo. 8. ed. Rio de Janeiro: LTC.

Botsman, R. (2013). The sharing economy lacks a shared definition, *Fast Company* 21: 2013. Retrivered from: https://www.fastcompany.com/3022028/the-sharing-economy-lacks-a-shared-definition Accessed on 20 th november 2018.

Botsman, R. (2015). Defining the sharing economy: What is collaborative consumption – And what isn't?, *Fast Company*. Retrieved from http://www.fastcoexist.com/3046119/defining-the-sharing-economy-what-iscollaborativeconsumption-and-what-isnt. Accessed on 25 th May 2019.

Botsman, R., and Rogers, R. (2010). Beyond zipcar: Collaborative consumption. *Harvard Business Review*, 88(10), 30.

Botsman, R. and Rogers, R. (2011). *What's Mine is Yours. How collaborative consumption is changing the way we live*, Harper Collins Publishers. London, UK.

Bridges, J., and Vásquez, C., (2016). If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism*, 21(18), 1–19.

Butcher, K., Sparks, B., and O'Callaghan, F. (2003). Beyond core service, *Psychology and Marketing*, 20(3), 187-208.

Cambria, E., Schuller, B., Xia, Y., and Havasi, C. (2013). New avenues in opinion mining and sentiment analysis, *IEEE Intelligent Systems*, 2, 15-21.

Chau, M., Shiu, B., Chan, I., and Chen, H. (2007). Redips: Backlink search and analysis on the Web for business intelligence analysis, *Journal of the Association for Information Science and Technology*, 58(3), 351-365

Chen, X., Huang, Q., Davison, R., and Hua, Z. (2014). The moderating effects of contextual facts on a buyer's trust in e-commerce platforms and sellers, *Pacific Asia Conference on Information Systems*, 1-18.

Cheng, M. (2016). Sharing economy: a review and agenda for future research, *International Journal Hospital Management*, 57, 60-70.

Cheng,M. and Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76 (19), 58-70.

Choi, T.Y., and Chu, R., (2001). Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry, *International Journal of Hospitality Management*, 20, 277-297.

Chung, J. Y., and Buhalis, D. (2008). Web 2.0: A study of online travel community. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 70-81), Innsbruck: Springer.

ComScore (2007). *Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior*. Retrieved from http://www.comscore.com/Press_Events/Press_Releases/2007/11/Online_Consumer_Reviews_Impact_Offline_Purchasing_Behavior Accessed on 15 th December 2018.

Conrady, R. (2007). Travel technology in the era of Web 2.0. In R. Conrady & M. Buck (Eds.), *Trends and Issues in Global Tourism 2007* (pp. 165-184). New York: Springer.

Crosby, L. A., and Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry, *Journal of Marketing Research*, 404-411.

Dabholkar, P. (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.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms, *Management Science*, 49(10), 1407-1024.

Dellarocas, C., Awad, N., and Zhang, X. (2004). Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning. *Proceedings of the 24th International Conference on Information Systems*, Washington D.C.

Denning, S. (2014). An economy of access is opening for business: five strategies for success, *Strategy & Leadership*, 42(4), 14–21.

Dervojeda, K., Verzijl, D., Nagtegaal, F., Lengton, M., Rouwmaat, E., Monfardini, E. and Frideres, L. (2013). The Sharing Economy. Accesseibility Based Business Models for Peer-to-Peer Markets. European Union. Retrieved from http://ec.europa.eu/DocsRoom/documents/13413/attachments/2/translations/en/renditions/native. Accessed: 20 January 2019. Accessed on 26 th November 2019.

Dredge, D., and Gyimóthy, S. (2015). The collaborative economy and tourism: critical perspectives, questionable claims and silenced voices, *Tourism Recreation Research*, 40(3), 286–302.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200.

Ert, E., Fleischer, A., and Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb, *Tourism Management*, 55, 62-73.

European Commission (2016). The use of collaborative platforms. URL: http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/FLASH/surveyKy/2112. Accessed: 11 November 2018.

Fang, B., Ye, Q., Kucukusta, D., and Law, R. (2016), Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management,* 52, 498–506.

Feinerer, I. (2018). Introduction to the tm Package Text Mining in R. *Unplublished article.* Retrivered from https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf. Accessed on 20 th May 2019.

Fenn, J., and LeHong, H. (2012). Hype Cycle for Emerging Technologies 2012, Gartner. Retrieved from http://www.gartner.com/DisplayDocument?doc_cd=233931. Accessed on 15 th February 2019.

Festila, M., and Müller, S. (2017). The impact of technology-Mediated consumption on identity: the case of airbnb. Paper Presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from Daguerreotype to laserphoto, *Journal of Personality and Social Psychology*, 63(6), 877-889.

Franke, N., and Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users, *Research Policy*, 32(1), 157–178.

Friedman, T. (2013). Welcome to the “Sharing Economy”. Retrieved from http://www.nytimes.com/2013/07/21/opinion/sunday/friedman-welcome-to-the-sharing-economy.html. Accessed on 16 th November 2018.

Gan, Q., Ferns, B. H., Yu, Y., and Jin, L. (2016). A text mining and multidimensional sentiment analysis of online restaurant reviews, *Journal of Quality Assurance in Hospitality & Tourism*, 18(4), 465-492.

Garbarino, E., and Strahilevitz, M. (2004). Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation, *Journal of Business Research*, 57(7), 768–775.

Gentile, C., Spiller, N., and Noci, G. (2007). How to sustain the customer experience: An overview of experience components that co-create value with the customer, *European Management Journal*, 25, 395-410.

Gerdes, J., Stringam, B. B., and Brookshire, R. G. (2008). An integrative approach to assess qualitative and quantitative consumer feedback, *Electronic Commerce Research*, 8(4), 217-234.

Geron, T. (2013). Airbnb and the unstoppable rise of the share economy. Forbes. Retrieved from http://www.forbes.com/sites /tomiogeron/2013/01/23/airbnb-and-the-unstop-pable-rise-of-the-share-economy/. Accessed on 30 th October 2018.

Ghazi, D., Inkpen, D., Szpakowicz, S. (2010), Hierarchical approach to emotion recognition and classification in texts, in A. Farzindar, V. Keselj (eds.), Proc. 23rd Canadian Conference on Artificial Intelligence, Ottawa, On *Lecture Notes in Artificial Intelligence 6085*, Springer Verlag, 40–50

Gloor, P., Krauss, J., Nann, S., Fischbach, K., and Schoder, D. (2009). Web science 2.0: Identifying trends through semantic social network analysis. *Computational Science and Engineering, 2009.CSE'09. International Conference*, vol. 4: 215-222.IEEE

Godbole, S., Bhattacharya, I., Gupta, A., and Verma, A. (2010). Building re-usable dictionary repositories for real-world text mining. *Em Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1189–1198). New York, NY, USA: ACM.

Goes, P., Guo, C., and Lin, M. (2016). Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange, *Information Systems Research*, 27(3), 497e516.

Grant-Braham, B. (2007). The social media and travel chatter, *Hospitality in Focus*. Retrivered from http://eprints.bournemouth.ac.uk/13634/2/Croner_-_Hospitality_in_Focus_-_14th_September_2007.pdf. Accessed on 6th December 2018.

Gretzel, U., Yoo, K. H., and Purifoy, M. (2007). *Online travel review study: Role and impact of online travel reviews*: Laboratory for Inteligent Systems in Tourism, Texas A&M University. Retrieved from http://www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf Accessed on 10th December 2018.

Griswold, A. (2015). Airbnb is thriving. So are hotels. How is that possible? Slate. Retrieved from http://www.slate.com/articles/business/moneybox/2015/07/airbnb_disrupting_hotels_it_hasn_t_happened_yet_and_both_are_thriving_what.html. Accessed on 20th February 2019.

Guo, Y., Barnes, S.J., and Jia, Q. (2017). Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation, *Tourism Management*, 59, 467–483.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current issues in Tourism*, *18*(12), 1192-1217.

Hamari, J., Sjoklint, M., and Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption, *Journal of the Association for Information Science and Technology*, 67(9), 2047-2059.

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, 60-75.

He, W., Zha, S., and Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry? *International Journal of Information Management*, 33(3), 464-472.

He, W., Wu, H., Yan, G., Akula, V., and Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks, *Information and Management*, 52(7), 801-812.

Henning-Thurau, T., Gwinner, K., Walsh, G., and Gremler, 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.

Herr, P.M., Kardes, F.R., and Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: Accessibility-diagnosticity perspective. *Journal of Consumer Research*, 17, 454-462.

Hess, R. L., Ganesan, S., and Klein, N. M. (2003). Service failure and recovery: the impact of relationship factors on customer satisfaction. *Journal of the Academy of Marketing Science*, 31(2), 127-145.

Hock, R. (2007). *The traveler's web: An extreme searcher guide to travel resources on the Internet*. Medford, NJ: Information Today Inc.

HotelMarketing (2011). *TripAdvisor grows and grows and grows*. Retrieved 16 Apr 2011, from http://www.hotelmarketing.com/index.php/content/article/tripadvisor_grows_and_grows_and_grows Hotel News Now. 2016. New Airbnb data sheds light on 13 global markets. Retrieved from http://hotelnewsnow.com/Articles/77191/New-Airbnb-data-sheds-light-on-13-global-markets. Accessed on 27 th November 2018.

Jain, M., Kumar, M., and Aggarwal, N. (2013). Web usage mining: An analysis, *Journal of Emerging Technologies in Web Intelligence*, 5(3), 240–246.

Iglesias, V. (2009). The attribution of service failures: effects on consumer satisfaction, *The Service Industries Journal*, 29(2), 127-141.

John, N. A. (2013). The social logics of sharing, *The Communication Review*, 16(3), 113–131.

Karakostas, B., Kardaras, D., and Papathanassiou, E. (2005). The state of CRM adoption by the financial services in the UK: an empirical investigation, *Information & Management*, 42(6), 853-863.

Kim, W. G., and Kim, D. J. (2004). Factors affecting online hotel reservation intention between online and non-online customers, *International Journal of Hospitality Management*, 23(4), 381-395.

Kim, J., Yoon, Y., Zo, H. (2015). Why people participate in the sharing economy: A social exchange perspective. In: Proceedings of Pacific Asia Conference on

Information Systems (PACIS) 2015. Retrivered from http://aisel.aisnet.org/pacis2015/76/ Accessed on 2nd November 2018.

Koopman, C., Mitchell, M., and Thierer, A. (2014). The sharing economy and consumer protection regulation: The case for policy change. Mercatus Working Paper, George Mason University, Arlington, VA. Retrivered from http://mercatus.org/publication/sharing-economy-and-consumer-protection-regulation-case-policy-change. Accessed on 2nd November 2018

Kozinets, R.V. (2002). The field behind the screen: Using netnography for marketing research in online communities, *Journal of Marketing Research*, 39, 61-72.

Kozinets, R.V. (2016). Amazonian forests and trees: multiplicity and objectivity in studies of online consumer-generated ratings and reviews, a commentary on de Langhe, Fernbach, and Lichtenstein, *Journal of Consumer Research*, 42(6), 834–839.

Laczniak, R. N., DeCarlo, T. E., and Ramaswami. S. N. (2001). Consumers' responses to negative word-of-mouth communication: An attribution theory perspective. *Journal of Consumer Psychology*, 11(1), 57-73.

Lamberton, C. P., and Rose, R. L. (2012). When is ours better than mine? A framework for understanding and altering participation in commercial sharing systems, *Journal of Marketing*, 76(4), 109-125.

Lau, N., Lee, H., and Ho, Y. (2005), Text Mining for the Hotel Industry, *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 344–362.

Lawrence, L. (2014). Reliability of Sentiment Mining Tools: A comparison of Semantria and Social Mention. Retrieved from http://essay.utwente.nl/65302/. Accessed on 27th March 2019.

Lee, J., Park, D. H., and Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view, *Electronic Commerce Research and Applications*, 7(3), 341-352.

Lee, S., and Kim, D. Y. (2018). Brand personality of Airbnb: Application of user involvement and gender differences, *Journal of Travel & Tourism Marketing*, *35*(1), 32-45.

Lehr, D. D. (2015). *An analysis of the changing competitive landscape in the hotel industry regarding Airbnb*, Master thesis, Dominican University of California, San Rafael, CA.

Lenhart, A., Horrigan, J., and Fallows, D. (2004). Content Creation Online. Washington, D.C., Pew Internet & American Life Project. Retrieved from: http://www.pewinternet.org/reports Accessed on 17th January 2019

Lerner, J., and Tirole, J. (2002). Some simple economics of open source, *The Journal of Industrial Economics*, 50(2), 197–234.

Li, C. (2009). Steering word-of-mouth online, *PRWeek*, 12(8), 13.

Li, N., and Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast, *Decision Support Systems*, 48(2), 354–368.

Liang, S., Schuckert, M., Law, R., and Chen, C. C. (2017). Be a "superhost": the importance of badge systems for peer-to-peer rental accommodations, *Tourism Management*, 60, 454–465.

Lipizzi, C., Iandoli, L., and Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams, *International Journal of Information Management*, 35(4), 490–503.

Litvin, S. W., Goldsmith, R. E., and Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management, *Tourism Management,* 29(3), 458-468.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

Liu, X., Schuckert, M., and Law, R. (2016). Online incentive hierarchies, review extremity, and review quality: Empirical evidence from the hotel sector, *Journal of Travel & Tourism Marketing*, 33(3), 279e292.

Liu, Y., Teichert, T., Rossi, M., Li, H., and Hu, F. (2017). Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews, *Tourism Management,* 59, 554–563.

Luo, Y., and Tang, R. (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.

Malhotra, A., and Van Alstyne, M. (2014). The dark side of the sharing economy and how to lighten it, *Communications of the ACM*, 57(11), 24–27.

Malinen, S., and Ojala, J. (2011). Perceptions of trust between online auction consumers, *International Journal of Web Portals*, 3(4), 15–26.

Matzler, K., Veider, V., and Kathan, W. (2015). Adapting to the sharing economy, *MIT Sloan Management Review*, 56(2), 71-81.

Max, H., and Mace, M. (2008). Online communities and their impact on business: Ignore at your peril. Rubicon Consulting Inc. Retrieved from: http://rubiconconsulting.com/insight/whitepapers/2008/10/.

McAuley, J., and Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. October. Proceedings of the 7th ACM Conference on Recommender Systems. ACM, New York, NY, pp. 165–172.

McCollough, M. A., Berry, L. L., and Yadav, M. S. (2000). An empirical investigation of customer satisfaction after service failure and recovery, *Journal of Service Research*, 3, 121-137.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 5(4), 1093–1113.

Metz, D. (2014). *Should hoteliers be more concerned about this company?* Retrieved from Professional Convention Management Association. Retrieved from http://www.pcma.org/be-in-theknow/pcma-central/convention-and-meetings-news/news-landing/2014/04/07/shouldhoteliers-be-more-concerned-about-this-company-#.U-6hk_nxq2k/ Accessed on 10th January 2019

Mizerski, R. W. (1982). An attribution explanation of the disproportionate influence of unfavorable information, *Journal of Consumer Research,* 9, 301-310.

Morgan Stanley (2016). Morgan Stanley Reports Fourth Quarter and Full Year 2016. Retrieved from https://www.morganstanley.com/about-us-ir/shareholder/4q2016.pdf Accessed on 1st February 2019

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments, *Expert Systems with Applications,* 40, 4241-4251.

Mudambi, S. M., and Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com, *MIS Quarterly,* 34(1), 185-200.

Nasukawa, T., and Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77), ACM.

Nie, X., Liu, L., Wang, H., Song, W., and Lu, J. (2013). The opinion mining based on fuzzy domain sentiment ontology tree for product reviews, *Journal of Software*, 8(11), 2682-2687.

Öğüt, H., and Onur Taş, B. K. (2012). The influence of Internet customer reviews on the online sales and prices in hotel industry, *The Service Industries Journal*, 32(2), 197-214.

Oliver, R. L. (1980). A cognitive model of the antecedentes and consequences of satisfaction decisions, *Journal of Marketing Research*, 17(4), 460-469.

Olson, M. and Kemp, S. (2015). Sharing economy. An In-Depth Look At Its Evolution & Trajectory Across Industries. Retrieved from http://collaborativeeconomy.com/wp/wp-content/uploads/2015/04/Sharing-Economy-An-In-Depth-Look-At-Its-Evolution-and-Trajectory-Across-Industries-.pdf. Accessed on 23th November 2018

O'Mahony, D. (2008). *Unlocking the power of your hotel website*. Retrieved from www.eyefortravel.com/node/13912. Accessed: 25 February 2019

Ordenes, F., Burton, J., Theodoulidis, B., Gruber, T., and Zaki, M. (2013). Analyzing customer experience feedback using text mining: A linguistics-based approach, *Journal of Service Research*, 17(3), 278-295.

Osborne, H. (2019). Amazon shoppers misled by 'bundled' star-ratings and reviews. *The Guardian*. Retrievered from: https://www.theguardian.com/technology/2019/apr/05/amazon-shoppers-misled-by-bundled-product-reviews Accessed on 10th April 2019

Owyang, J. (2013). A Market Definition Report of 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. Retrieved from http://www.collaboriamo.org/media/2014/04/collabecon-draft16-130531132802-phpapp02-2.pdf. Accessed on 10th December 2018

Owyang, J., Samuel, A., and Grenville, A. (2014). *Sharing is the new buying: How to win in the collaborative economy*. VisionCritical, Crowd Companies.

Palese, B., and Usai, A. (2018). The relative importance of service quality dimensions in E-commerce- experiences, *International Journal of Information Management*, 40, 132–140.

Paltoglou, G., and Thelwall, M. (2010). A study of information retrieval weighting schemes for sentiment analysis. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1386–1395, Uppsala, Sweden

Pan, B., MacLaurin, T., and Crotts, J.C. (2007). Travel blogs and the implications for destination marketing, *Journal Travel Research*, 46(1), 35–45.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, 271

Pang, B., and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*: 115-124. Association for Computational Linguistics.

Papathanassis, A., and Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach, *Tourism Management*, 32(2), 215-224.

Park, S., Huh, S., Oh, W., and Han, S. P. (2012). A social network-based inference model for validating customer profile data, *MIS Quarterly*, 36(4), 1217-1237.

Peterson, R.A., and Wilson, W.R., (1992). Measuring customer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, 20(1), 61-71.

Pekar, V., and Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews, *Journal of Vacation Marketing*, 14, 145–155.

Pfister, H. and Böhm, G. (2008). The multiplicity of emotions: A framework of emotional functions in decision making, *Judgment and Decision Making*, 3(1), 5-17.

Piryani, R., Madhavi, D., and Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015, *Information Processing and Management*, 53(1), 122–150.

Piscicelli, L., Cooper, T., and 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.

Plutchik, R. (1994). *The Psychology and Biology of Emotion*. New York, US: HarperCollins College Publishers.

Priceonomics (2013). Airbnb vs Hotels: A Price Comparison. Retrieved from: https://priceonomics.com/hotels/. Accessed on 20 November 2018.

Ren, L., Qiu, H., Wang, P., and Lin, P. M. C. (2016). Exploring customer experience with budget hotels: Dimensionality and satisfaction, *International Journal of Hospitality Management*, 52, 13-23.

Rogers, D., and Sexton, D. (2012). Marketing ROI in the era of big data: The 2012 BRITE\NYAMA Marketing in Transition Study. Technical report, Columbia Business School. Retrieved from https://www.iab.com/wp-content/uploads/2015/05/2012-BRITE-NYAMA-Marketing-ROI-Study.pdf Accessed on 27th March 2019

Rosenberg, T. (2013). It's Not Just Nice to Share, It's the Future. *The New York Times*. Retrieved from http://opinionator.blogs.nytimes.com/2013/06/05/its-not-just-nice-to-share-its-the-future/?_r=0. Accessed on 8th April 2019

Salehan, M., and Kim, D. (2016). Predicting the performance of online consumer reviews, *Journal Decision Support Systems*, 81(C), 30-40.

Santoli, M. (2014). *Hyatt CEO sees no serious threat from fast-growing Airbnb*. Retrieved from Yahoo! Finance. Retrieved from http://finance.yahoo.com/blogs/daily-ticker/hyatt-ceosees-no-serious-threat-from-fast-growing-airbnb-200252874.html/ Accessed on 2nd March 2019

Sheth, J.N., Sethia, N.K., and Srinivas, S., (2011). Mindful consumption: A customer-centric approach to sustainability, *Journal of the Academy of Marketing Science*, 39(1), 21–39.

Schor, J. (2016). Debating the sharing economy, *Journal of Self-Governance & Management Economics*, 4(3), 1-15.

Shapiro, M. (2014). Coming to terms with Airbnb: Like it or not, alternative lodging is a game-changer for meetings, *Meetings & Conventions*, 29-36.

Sharda, R., Delen, D., and Turban, E., (2014). *Business Intelligence: A Managerial Perspective on Analytics* (3rd ed.). Harlow: Pearson.

Smolka, C., and Hienerth, C. (2014). *The Best of Both Worlds: Conceptualizing Trade-offs between Openness and Closedness for Sharing Economy Models*. Paper presented at *12th International Open and User Innovation Workshop, Harvard Busines School, Harvard University, Boston, MA, USA.*

Sparks, B. A., and Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310-1323.

Starkov, M., and Price, J. (2007). *Are search engines becoming obsolete in the Web 2.0 frenzy?* Retrieved from www.4hoteliers.com/4hots_fshw.php?mwi=2389 Accessed on 14th May 2019

Stephany, A. (2014). *Beyond Uber and Airbnb: The future of the 'sharing economy'*. Retrieved from: www.latimes.com/opinion/opinion-la/la-olsharing-economy-20140519-story.html/ Accessed on 2nd March 2019

Stephany, A. (2015). *The Business of Sharing. Making it in the New Sharing Economy*. Hampshire, UK: Palgrave Macmillan.

Sthapit, E., and Jiménez-Barreto, J. (2018) Sharing in the host–guest relationship: perspectives on the Airbnb hospitality experience, *Anatolia*, 29(2), 282-284.

Stieglitz, S., and Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277-1291.

Sundararajan, A. (2016). The sharing economy. The end of employment and the rise of crowd-based capitalism. Toppan Best-set Premedia Limited. USA.

Ting, D. (2017). Airbnb is becoming an even bigger threat to hotels says a new report. Skift. Retrieved from https://skift.com/2017/01/04/airbnb-is-becoming-an-even-bigger-threat-to-hotels-says-a-new-report/. Accessed on 19th January 2019

The Economist. (2012). Online house-sharing Mi casa, su cash. Retrieved from http://www.economist.com/blogs/babbage/2012/07/online-house-sharing.

Thelwall, M. (2016). Book genre and author gender: Romance> Paranormal- Romance to Autobiography> Memoir. *Journal of the Association for Information Science and Technology*, 68(5), 1212-1223.

Tomar, S. S (2017). Text Mining in R: a Tutorial. *Springboard blog.* February 2018. Retrieved from https://www.springboard.com/blog/text-mining-in-r/ Accessed on 20th May 2019

Tussyadiah, L., and Pesonen, J. (2016). Impacts of peer-to-Peer accommodation use on travel patterns, *Journal of Travel Research*, 55(8), 1022–1040.

Tussyadiah, L., and Zach, F., (2016). Identifying salient attributes of peer-to-peer accommodation experience, *Journal Travel Tourism Marketing*, 34 (5), 636-652.

Verhage, J. (2016). One Wall Street Firm Expects Airbnb to Book a Billion Nights a Year Within a Decade. Retrieved from https://www.bloomberg.com/news/articles/2016-04-11/one-wall-street-firm-expects-airbnb-to-book-a-billion-nights-a-year-within-adecade. Accessed on 14th May 2019

Volkova, S., Wilson, T., and Yarowsky, D. (2013). Exploring demographic language variations to improve multilingual sentiment analysis in social media. In EMNLP (pp. 1815-1827).

Volkova, S. and Yarowsky, D. (2014). Improving gender prediction of social media users via weighted annotator rationales. In NIPS 2014 Workshop on Personalization. Retrieved from http://www.cs.jhu.edu/~svitlana/papers/VYNIPSPersonalization14.pdf Accessed on 17th May 2019

Weber, K., & Sparks, B. (2010). Service failure and recovery in a strategic airline alliance context: Interplay of locus of service failure and social identity, *Journal of Travel & Tourism Marketing*, 27(6), 547-564.

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective, *Educational Psychology Review*, 12, 1-14.

Wong, K. F., Xia, Y., Xu, R., Wu, M., and Li, W. (2008). Pattern-based opinion mining for stock market trend prediction, *International Journal of Computer Processing of Languages*, 21(4), 347-361.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., and Uysal, M. (2015). What can big data and texto analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.

Xie, K. L., Chen, C., and Wu, S. (2016), Online consumer review factors affecting offline hotel popularity: Evidence from Tripadvisor, *Journal of Travel & Tourism Marketing*, 33(2), 211-223.

Xu, X., Wang, X., Li, Y., and Haghidhi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors, *International Journal of Information Management*, 37, 673–683.

Yannopoulou, N. (2013). User-generated brands and social media: couchsurfing and Airbnb, *Contemporal Management Research*, 9(1), 85-90.

Ye, Q., Law, R., Gu, B., and 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.

Zhang, L., Yan, Q., and Zhang, L. (2018). A computational framework for understanding antecedents of guests' perceived trust towards hosts on Airbnb, *Decision Support Systems*, 115, 105-116.