NOVA IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Text Mining of Airbnb Reviews

A holistic approach on reviewers' opinions and topics distribution

Ana Catarina Guinote Fernandes Alves Rodrigues

Dissertation presented as a partial requirement for the degree of Master of Information Management, Specialization in Marketing Intelligence

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

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

TEXT MINING OF AIRBNB REVIEWS

by

Ana Catarina Guinote Fernandes Alves Rodrigues

Dissertation presented as a partial requirement for the degree of Master of Information Management, Specialization in Marketing Intelligence

Supervisor: Diego Costa Pinto

June 2021

ABSTRACT

This thesis aims to perform a holistic investigation concerning how Airbnb accommodation features and hosts' attributes influence guest's reviews and how are the main topics distributed. A dataset containing almost 4 million reviews from major touristic cities in the world (Milan, Lisbon, Amsterdam, Toronto, San-Francisco, and Sydney) was used for the text mining analysis to uncover the reviews' social and market norms, as well as the guests' sentiments and topics distribution. This research uses both Mallet LDA (Latent Dirichlet Allocation) and Word2Vec methods to unveil the semantic structure and similarity between data in this study. This approach will allow hospitality providers to understand the impact of underlying factors on reviewers' opinions for further improvement of their services. Finally, this study develops a predictive unbiased model to forecast the review's scores, with an accuracy of 90.70%.

KEYWORDS

Airbnb; Text Mining; Online Reviews; Social Norms; Market Norms

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. BACKGROUND

As tourism's relevance increases, so do the offering of accommodations to satisfy the travelers' needs. Besides the hotels, new internet-based booking platforms that facilitate the spread of alternative accommodation offerings start to expand (Brauckmann, 2017). These collaborative platforms, labeled as "Sharing Economy", are challenging and redesigning traditional business models while ridding the tourism industry of monopolies and resource inefficiencies as they efficiently allocate assets and human resources (O'Regan & Choe, 2017). The research regarding Sharing Economy has verified an increase in the past years (Hossain, 2020).

In the last few years, these Sharing economy platforms have become particularly popular (Quattrone et al, 2016). This concept simplifies the relationship between suppliers and demanders, through a set of a peer-to-peer online marketplace, being the suppliers mostly individuals. In line with this increase, tourists have overcome the bias of "stranger-danger" (Suess et al., 2020), searching more and more for this type of experience. Recognized as the pioneer of the sharing economy is Airbnb, the marketplace for short-term rentals (Barron et al., 2018).

1.2. AIRBNB: THE COLLABORATIVE PLATFORM

Airbnb is an innovative collaborative platform for accommodation-sharing services that connects hosts and guests (Lu & Kandampully, 2016). Since its establishment in 2008, it has verified rapid growth and has connected more than 4 million hosts with above 800 million guests across more than 100,000 cities. It is a "community based on connection and belonging" where the hosts, as hospitality providers, "share their worlds to provide guests with the feeling of connection and being at home" (Airbnb, 2020).

The brand offers alternative accommodation for its users and challenges the models and practices of the conventional hotel industry (Bridges & Vásquez, 2016; Cheng, 2016; Zervas et al., 2017). On average, when compared to hotel rates, Airbnb rental offerings are valued 21.2% lower for houses and 49.5% lower for single rooms (Lee & Kim, 2018). Sainaghi and Baggio (2020) analyzed the possible substitution threat between these listings and hotels and have verified that there is a potential substitution threat, especially, during weekends and holidays, in which there is a partial synchronization in the daily occupancies. Through the delivery of lodging services, it generates and provides unique local experiences to its users (Luo, 2018).

1.3. ONLINE REVIEWS: THE MEASURE OF EXPERIENCES

In the platform, the experiences are measured through online reviews. Web 2.0 emerging technologies have played a major role in the development of several types of user-generated content on numerous websites, like booking platforms, in which the guests can discuss their experiences related to the services or products with other users (Plank, 2016). Furthermore, 81% of travelers are proved to consider these reviews important for their decisions (Statistic Brain, 2017). It is also verified that clients take other users' reviews increasingly into account to obtain information regarding accommodations, attractions, destinations, experiences and activities (Yoo & Gretzel, 2008; Park & Gretzel, 2007; Zhou et al., 2014). Thus, online reviews are a major information source that assists consumers and marketers in learning about the quality of the service (Chen & Xie, 2008) and, therefore, should be carefully and efficiently analyzed.

1.4. TEXT MINING: UNVEILING THE HIDDEN FACTORS

Text mining techniques are used to retrieve meaningful patterns and knowledge from unstructured text or raw data (Hung & Zhang, 2012). Sharda et al (2014) defined text mining as a semiautomatic process of extracting meaningful patterns from large volumes of unstructured text and transforming them into structured information. Regarding Airbnb, these techniques are employed to extract semantic characteristics from review texts. Ding et al (2020) used text mining techniques to extract “service quality attributes from online customer reviews”.

The steps to perform include “gathering, extracting, pre-processing, text transformation, feature extraction, pattern selection, and evaluation of results” (Liao et al., 2012). The statistical model involves word count analysis, probability model and frequency analysis (Chen et al., 2014). Additionally, the identification of data patterns to obtain high-quality information from text is the main objective of this approach (Aggarwal & Zhai, 2012).

Throughout the years, several studies were conducted applying a text mining analysis for several purposes. Text mining provides the tools for industries to understand and improve their products/services, enabling them to position themselves against their competition (Jain et al., 2013). Generally, online reviews, being text-based, encompass a lot more information to be analyzed, rather than online ratings, that are numerical (Ye et al., 2009). The analysis of these reviews is extremely valuable in the way that it offers a broader classification of the consumer experience, understanding the determinant factors of their satisfaction/unsatisfaction, and allows the possibility of analyzing their sentiments (Sparks and Browning, 2011). The impact of ratings on hotel websites has also been studied by other authors, such as Schuckert et al (2016) and Zhu & Zhang (2010). Moreover, using text mining methods, Zhang et al (2020) studied the relationships between the “host self-description, trust perception and purchase behavior” on Airbnb. Overall, text mining techniques allow the automatization of obtaining accurate and meaningful information to improve the decision-making process of companies (Fenn and LeHong, 2012).

1.5. PRIOR STUDIES AND RELEVANT FACTORS

Reviews’ analysis from previous studies revealed that factors such as price value, home atmosphere, sustainability and community are drivers of the choice of using Airbnb (Guttentag, 2015; Liang, 2015; Tussyadiah, 2015). On the other hand, unpredictability, lack of cost savings, lack of efficacy and distrust are viewed as restraints for using the platform (Liang, 2015; Tussyadiah, 2015; Tussyadiah & Pesonen, 2016a). Yang and Mao (2020) identified the “accessibility to points of interest, transport convenience, the surrounding environment, and market conditions” as location factors that contribute to lodging property performance.

Important Airbnb dimensions to consider are the “cleanliness” (Bridges & Vásquez, 2016), the “location” (Tussyadiah & Zach, 2016), the “economic benefits/cheaper price” (Guttentag & Smith, 2017) and the “household amenities” (Guttentag, 2015). Besides these features, Festila & Müller (2017) ensure that “authentic experience and host-guest interaction” are also a core dimension of the Airbnb experience. Additionally, Tussyadiah & Zach (2016) and Yannopoulou (2013) defend that the “time spent in local neighbourhoods” is another established measure to consider. The last researcher also argued that Airbnb involves a “meaningful life enrichment, human contact, access and authenticity”.

In this way, the interactions with the host are an essential criterion when evaluating the user experience (Tussyadiah & Pesonen, 2016a; Festila & Müller, 2017; Lampinen & Cheshire, 2016; Yannopoulou, 2013). Wu et al (2021) have verified that host-guest interaction increases guest's repeated reservations. Accordingly, Ostrom (2014) highlights social norms as necessary for a collective action to succeed, which corresponds to the social-moral relationship that determines demand alongside the price and market norms (Ariely et al., 2017).

Concerning the reviews itself, research has shown not only that negative reviews are more authentic and credible than positive reviews on Airbnb, but also the occurrence of social words is positively related to positive emotion, being, however, negatively linked to negative emotion in reviews (Zhang, 2019).

1.6. THEORETICAL GAP AND CONTRIBUTION

Prior studies are important to understand the Airbnb dimensions and features. Nevertheless, the previously presented researches lack an efficient and complete analysis of the reviews, where several hypotheses are considered to ensure full comprehension of the relationship between Airbnb’s components and user’s opinion.

Therefore, the focus of this master thesis is the detailed text mining analysis of Airbnb users’ reviews in different cities of the world, enabling the study of its evolution over time, with emphasis on the way that social norms (social-oriented) and market norms (business-oriented) influence the reviewer's opinion, considering a holistic approach to these norms together with users' sentiments, based on the identified topics to unveil the consumers' satisfaction.

Moreover, the predictive model that will allow the accurate prediction of the review's score rating complements this project, providing a solid contribution and creating a basis for future studies regarding other variables and subjects.

1.7. PAPER STRUCTURE

The paper is organized as follows. First, it presents the context of this study followed by the review of existing literature and hypothesis on social and market norms, as well as sentiments analysis and topic modelling. With the conceptual model and methodology carefully described, the following chapters regard the analysis of results and the review's ratings predictive model presentation. Finally, the results are discussed before detailing the main contributions, limitations and future research opportunities.

2. THEORETICAL BACKGROUND/LITERATURE REVIEW AND HYPOTHESIS

2.1. CONCEPTUAL DEVELOPMENT

The main research question of this master thesis can be defined as "How does the Airbnb market norms and social norms, along with guests' sentiments, influence the reviews? And how are the main topics distributed?" In this way, the conceptual model is as represented below:

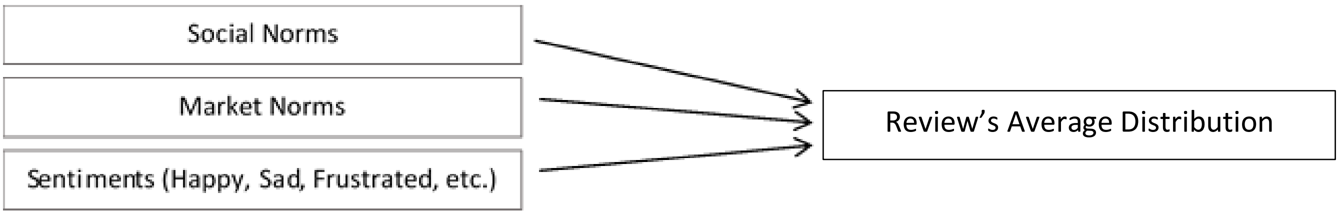

Figure 1 – Conceptual Model: Independent Variables and Dependent Variable

The dependent variable in this research is the review's distribution, whereas the independent variables encompass the social norms, the market norms and the specific sentiments (e.g. positive, neutral, negative) shown in reviews regarding the customer's opinions. These will be analysed considering the keywords/main topics present in the reviews.

2.1.1 SOCIAL NORMS IN HOSPITALITY

Social norms are considered a driver of behaviour in several social contexts (Krupka and Weber, 2013). These norms are connected with social and psychological concepts, such as factual convictions (Heiphetz et al., 2014), attitudes, which are directly related to an individual's preference (Petty and Brinol, 2010), or self and group efficacy, which means the belief on the capacity to accomplish an objective (Bandura et al., 1999). As mentioned before, social norms are highlighted by Ostrom (2014) as required for a collective action to succeed.

The Collaborative Consumption scenario is not different from the contexts abovementioned. From the literature presented, it is possible to understand that social norms are directly associated with the challenge that hosts in Airbnb have of matching the standards of the constantly evolving society.

More and more the society standards and customers' expectations regarding local accommodations are associated with authenticity and providing access to the "local experience". The interactions are considered authentic encounters that cannot be repeated in a conventional hotel setting (Tussyadiah, 2016) and tend to have a positive impact on the perceived authenticity (Liang et al., 2018). According to Mao & Lyu (2017), in Airbnb, the unique social nature tends to influence customers' emotional and behavioural responses, which will have an impact on the final review. Therefore, it is suggested that reviews associated with social norms influence the review overall score (H1).

The social nature of this peer-to-peer social and virtual interaction creates more chances of establishing social connectedness and possibly of producing stronger social ties (Lin et al. 2019; Perren and Kozinets 2018). Also, staying at peer-to-peer accommodation provides an opportunity to have closer connections, as it normally involves more human interactions between guests and hosts (Tussyadiah & Pesonen, 2016). Subsequently, the social interaction in collaborative consumption appears to avoid customers from posting negative reviews (Pera et al., 2019). In this way, social Interaction and Interpersonal contact with hosts are crucial parts of the sharing experience (Bucher et al., 2018).

Luo (2018) highlighted the word 'host' as great influencer over Airbnb users' recommendations and supported that a careful clarification of destination attractions aligned with a hosts' thoughtful service can contribute to customers' positive feelings. Also, Chen and Xie (2008) identified that helpfulness, flexibility and good communication play an important role in building up the initial trust, which could influence the Airbnb accreditation system. Furthermore, 66% of the text segments with the term host contained positive sentiments. Hence, it will also be tested whether host interaction has a positive impact on ratings (H1a).

Additionally, Ert et al (2016) performed controlled experiments that have indicated that the perceived trustworthiness of the photos posted by hosts impacts the guest's choice of booking and, therefore, the likelihood of Airbnb providers attaining bookings. Moreover, in the reviews, customers typically express sadness when their complaints are not solved by their hosts. Accordingly, customers seem to express anger when their listing facilities are not convenient (Luo, 2018). In order to avoid the unsatisfaction of customers, the host should make sure the descriptions associated with their listings truly match the customers' expectations upon arrival. Taking into consideration the abovementioned studies, it is possible to state that a significant gap in the relationship between guest's expectations and perceptions upon arriving at the accommodation can result in negative reviews (H1b).

H1: Social Norms impact the Overall Review.

H1a: Host Interaction has a positive impact on reviews.

H1b: The gap between the guest's expectations and perception has a negative impact on reviews.

2.1.2 MARKET NORMS IN HOSPITALITY

According to Ariely (2010), there are two different worlds. First, there is the one where social norms predominate, the second is where the market norms are in control. In this second one, the "exchanges are sharp-edged", meaning that wages, prices, rents, interest and costs-benefits are considered core drivers, being the market ruled by "the soulless exchange of capital for goods and services".

Prior studies have demonstrated that price value is one of the drivers of using Airbnb. On the opposite side, as mentioned in the introduction, unpredictability, lack of cost savings, lack of efficacy and distrust are viewed as restraints for using the platform (e.g. Tussyadiah, 2015). Moreover, "location" (city, beach, short, transport, nearby, shopping, bus), as well as a "good place for family”, and “nice home” (bed, water, bathroom) could lead to a recommendation, being the location and amenities features mentioned in most of the content of the online reviews of Airbnb users.

Additionally, characteristics, namely, “parks, bodies of water, airports and trains” are considered extremely valuable to travellers evaluating sharing accommodations. In reviews, customers showed concerns regarding appropriate transportation and location-related security issues (Luo, 2018). Also, “noise, floor, shower, parking, and door” have resulted in customers’ negative sentiments in reviews, which is directly linked to market norms perception of customers (Chen and Xie, 2008) and, subsequently, the cost-benefits of the accommodation and experience.

Guttentag & Smith (2017) and Guttentag (2015), verified that valued features are “economic benefits/cheaper price” and “household amenities”. In this way, there is enough literature to support the hypothesis that reviews associated with market norms influence the review overall score (H2), which will be tested in this study, focusing on understanding the impact of sentences regarding market/business norms in the reviews.

H2: Market norms influence/have an impact on the overall review.

2.1.3 SENTIMENT ANALYSIS IN HOSPITALITY

Sentiment analysis corresponds to the process of extracting and categorizing opinions and emotions of users as positive, negative or neutral (Fernández-Gavilanes et al., 2016). An individual’s emotions analysis is attained through the identification of the text fragments that indicate a sentiment or opinion regarding a topic (Luo 2018; Nasukawa and Yi 2003).

The sentiment analysis is also referred to as “polarity analysis” (Liu, 2012), which can concern the dichotomization in positive or negative, considering a range of values, being]0,1] positive sentiment and [-1,0] negative sentiments (Cambria et al., 2013) or the trichotomization, which includes the neutral factor. Moreover, it can be classified into two categories, specifically, opinions (subjective) and facts (objective) (Schouten and Frasincar, 2016). While subjective statements are a representation of perspectives and judgements, objective statements express facts about a matter (Singh et al., 2014; Khan et al., 2014).

In this way, not only the sentiment classification, but also the subjectivity are necessary steps to perform accurate and efficient sentiment analysis (Pang and Lee, 2004). Furthermore, taking into consideration a specific text, it is important to link the information with a dictionary or lexicon, to assess the emotion strength (Mostafa, 2013).

This method is an excellent form of extracting observations from untreated data and converting it into valuable information to be further analysed by the interested parties. Previous studies denote that emotions analysis play a major role in unveiling a client's implicit feelings regarding the key subjects or features of accommodations (McAuley and Leskovec, 2013). The extraction of customer's opinions helps the brands management as well as its reputation (Pang and Lee, 2005), in the way that it provides the tools for customer relationship management analysis and strategy definition (Karakostas et al., 2005). In accordance, Guo et al. (2017) managed to understand guests' satisfaction and dissatisfaction top dimensions, regarding online reviews.

Prior research indicates that when customers perceive meetings as authentic and personal, they seem to experience more positive emotional responses (Hennig-Thurauet al., 2006), which is related to the previously presented theme of perceived authenticity and host interaction.

Bartel and Saavedra (2000) assert that interpersonal interaction with the host can impact customer emotions. In fact, these interactions induce positive emotions as a result of mutual relationship building (So et al., 2018). Also, Luo (2018) asserts that positive emotions are prompted by, for example, hosts' responses to customer questions and host resolution of customer problems.

On the other hand, witnessing other customers obtaining unreasonable treatment results in a negative evaluation of fairness which, as consequence, influences the other customer's individual evaluation (Mattila et al., 2014). Besides this, Cao et al. (2011) estimates of online user reviews indicated that some words have a positive impact, encouraging review votes, whilst others have a negative influence. Moreover, their findings indicate that the semantic characteristics have more impact than other features, regarding the number of helpful votes reviews obtain. Nevertheless, they also proved that extreme opinions collect more helpfulness votes than those with mixed or neutral considerations. In this way, the positive vs negative sentiment reviews impact on review's score will be tested (H3).

H3: Positive (vs Negative) sentiment reviews have an impact on Review's Score.

2.2. TOPIC AND PREDICTIVE MODELLING OF REVIEWS

2.2.1 TOPIC MODELLING

A precise and trustworthy sentiment analysis regards the analysis of the text segments, but also grouping them into topics. A topic concerns the clustering of words that frequently occur together. This modelling uncovers the key topics in a set of textual data, using a statistical model (Hong & Davison, 2010) and, also, allows understanding both the hidden semantic structures in a text (Aggarwal & Zhai, 2012) and the assessment and careful analysis of ambiguity in the words' connotation, regarding similar topics (Williamson et al., 2010). Therefore, topic models can search for patterns in the meaning of words and differentiate between uses of words with numerous connotations (McCallum, 2002), introducing in this way semantic meaning into the vocabulary.

In this way, it provides a starting point for an investigation of new forms of semantic representation (Griffiths et al., 2007). The author also revealed that the words that store high probability about the same topics, will tend to be greatly predictive of one another.

2.2.1.1. MALLET LATENT DIRICHLET ALLOCATION (LDA) METHOD

The Latent Dirichlet Allocation method is a three-level hierarchical Bayesian modelling process that clusters items into topics and the probabilities that describe each one (Blei, 2012). For this study, the Malled LDA will be used. This method consists of a topic model package that includes an "extremely fast and highly scalable implementation of Gibbs sampling, efficient methods for document-topic hyperparameter optimization, and tools for inferring topics for new documents given trained models". The *Gibbs Sampling* is a statistical technique created to promptly construct a sample distribution, to develop its topic models, being, therefore, normally used as a means of statistical inference (McCallum, 2002). This analysis enables the identification of the most frequent words used in reviews, pointing out the importance given by the guests to the described aspects.

Asuncion et al. (2010) presented how topic modelling increases software traceability. Chen et al. (2012) used LDA to discover relationships between software defects and software development, showing that LDA can easily scale to large documents. Tong and Zhang (2016) conducted two experiments using LDA, one regarding Twitter posts, in order to uncover what kind of topic the user talks more and is more interested in, and other concerning topic models on Wikipedia articles, understanding the series of article distribution over each topic.

Therefore, this method allows explaining the similarity between data, clarifying groups of observations. In this case, identifying the most verified topics in guest's Airbnb reviews.

2.2.1.2. WORD2VEC MODEL

Word2Vec is a technique to construct word embedding through vector representations of a certain word assessing the similarity metrics proposed by Mikolov et al (2013a; 2013b). It is proved to outperform traditional distributional methods (Baroni et al., 2014). Naili et al (2017) performed a study to assess various word embedding methods, in which it was concluded that Word2Vec presents the best word vector representations with a small dimensional semantic space. Moreover, it was proven that the quality of topic segmentation depends on the used language. In this way, using, for example, Arabic language decreases the abovementioned quality when compared to the English language, which is the chosen idiom for this study. It was also shown that this method provides a high quality of topic segmentation.

Jatnika et al (2019) performed a study using Word2Vec in which the similarity between words in English was measured, using word representation techniques to understand the correlation. For biomedical purposes, Minarro-Giménes et al (2014), has used word embedding to study the semantic and similarity association for information extraction. Zhang (2019) presented a two-stage text mining approach to classify construction accident causes, in which it was verified that the chosen approach outdid the other standard models considered in the analysis. This method greatest advantage is the fact that contextual similarity and semantic relationship between words can be inferred from the learned vectors (Khatua et al. 2019).

There are two methods of Word2Vec, namely the CBOW (Common Bag Of Words) and the Skip Gram. The first one forecasts the target word based on its neighbouring words, being more appropriate for large datasets, whereas the goal of the second method is the context prediction of a given word, being more suitable for smaller dimensional semantic spaces (Karani, 2018). Zhang et al (2015) and Alshari et al (2017) applied Word2Vec techniques in their studies, highlighting that it can reveal deep semantic features between words and, as already mentioned by other authors, it can be more effective than the baseline methods.

For all the specified reasons above, the sentiment analysis will be performed to test each topic's sentiment distribution for reviews.

2.2.2 PREDICTIVE MODELLING OF REVIEW'S RATINGS

With the evolution and increased number of online user reviews, natural language processing studies have started to focus on developing models that can predict the review's rating. Prior studies have shown that "user context information" is a significant source of data that should be taken into consideration (Tang et al., 2015) and that the extraction of other features like lexical patterns, semantic topics, words or syntactic structures can improve the performance of the model (Qu et al., 2010).

Additionally, some authors do not consider the review content as the only factor to be examined (Wang et al., 2010), since a user can comment positive words regarding a specific product, even if the rating he gave to the product was lower. This can happen when the person is tolerant or understanding, always depending on the situation. Pang and Lee (2005) studied class relationships for sentiment categorization regarding the rating scales, defined, usually, from 1 to 5.

After analyzing all the reviews extracted in the first phase, the present study will take into consideration the features previously obtained, namely from the sentiment and topic models, such as polarity, subjectivity score and the topics obtained, to explore the possible feature set for the model.

The reviews need to be labelled as categories and, for the training phase, these will need to be converted into a suitable format to use as input to the model. One-Hot encoding is the most common approach to perform this action. Binary encoding option, in which, the categorical data is transformed by "first assigning a numerical value to each category and then converting it to its binary representation" will be the applied technique (Seger, 2018).

Since ratings represent a certain order of classification, these problems are often tackled with regression models. Ning et al (2020) proposed a Convolutional Neural Network (CNN) model to predict movie ratings. In the present study, both regression and classification approaches will be tested to ensure the best decision of the chosen model.

Previous studies identified the K-folds cross-validation method as mostly leading to optimal model selection performance, since this method has a small variance (Syed, 2011). In this procedure, the dataset is divided in K number of "folds" and the model is trained on K – 1 data, being the remaining K used for the testing set. The process is repeated K number of times, ensuring each fold is only used for testing once. This will be the strategy to apply in this predictive model.

3. METHODOLOGY AND RESULTS

3.1. REVIEWS' ANALYSIS

3.1.1 DATA COLLECTION

To obtain a larger sample of data and ensure the reliability of the analysis, the dataset used in this study contains reviews information from six different cities of the world, namely Milan, Lisbon, Amsterdam, Toronto, San-Francisco and Sydney. Moreover, the numbers refer to two distant timelines, being randomly retrieved from both 2019 and 2016 (data from 2018 was used when no 2016 information was available). Therefore, enabling the study of its evolution over time. The data was retrieved considering the city distribution obtaining, in this way, an equitable proportion. This means that, if Lisbon contains 100 reviews and Milan only 10 in the database, then the training set will contain 10 reviews from Lisbon for each review from Milan. Then, the retrieved file would represent this proportional relationship. The figure below regards a visual representation of this result.

The datasets obtained contain details of thousands of accommodations and their customer reviews. Precisely, for this study, we are considering a total of 3.866.531 reviews.

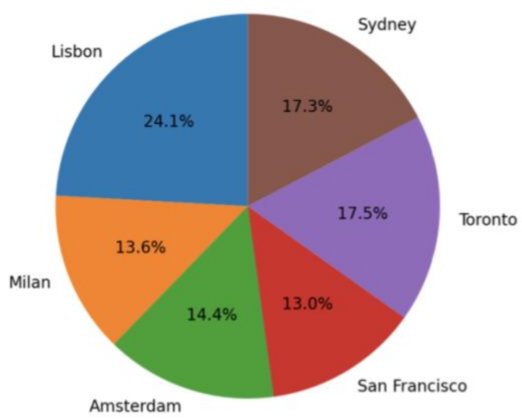

Figure 2 - Data Distribution Among the Cities

The listings dataset contains initial basic information in text format, containing the listings' ID, URL, name, summary, description, neighbourhood, transit, access, interaction, house rules and the latitude/longitude. It is also possible to verify the property type (apartment, house, loft, boat, room and others), room type, the number of bathrooms, bedrooms, included amenities, information regarding the price (total, weekly or monthly), associated fees (security deposit, cleaning, extra guests), minimum/maximum nights, cancellation policy and calendar availability. Moreover, it encloses the listings' total number of reviews, the dates of the first and last reviews, the scores rating and a detailed score from 1 to 10 concerning individually the accuracy, cleanliness, check-in, communication, location and value. Regarding the host, the dataset covers his/her name, date of initiation on Airbnb, location, response time and rate, whether he/she is a "super host", number of listings and if the identity is verified. The reviews dataset includes the listing, reviewer and review's IDs, the date, the name of reviewer and comments.

Thus, to collect the data, first, a dataset that exposes the present reviews scenario was created. Then, sampling was performed to reach a reduced dataset for this project, using the ETL process (Extract, Transform and Load), as explained in the following sub-chapters.

3.1.2 PRE-PROCESSING PHASE 1 - CLEANING

Python programming was the chosen language to perform the Natural Language Processing (NLP) of reviews. With NLP it will be possible to summarize and classify raw data into knowledge. As previously presented by Guzman and Maalej (2014), this automated method allows the extraction of features and sentiments in reviews, through a fine-grained analysis.

The first step was cleaning and preparation of the data for analysis. In the initial phase, the excel files were checked for impurity. In this way, took place a detailed removal/replacement of the impure or unrecognized characters, including useless punctuation that was not valuable for sentence tokenization.

Additionally, text lowercasing procedure was also applied to ensure the word's consistency. Besides, to guarantee that the core meaning of the sentence prevails and to assure that the analysed data has the highest information value possible, unnecessary numbers were also removed from these sentences. This last cleaning process increases the probability of pointing out only the relevant information, represented in words for the analysis. As an example, the sentence "We had 5 amazing meals", after this procedure, would indicate the core content, which is that this customer has had "amazing meals".

Moreover, we were able to reduce the inflection in words to their root forms and identify and remove the "Stop Words" that did not contain important significance to be used. Those words are removed from the analysis, as they are revealed as unnecessary, due to the fact that these are not measured as keywords in text mining. As an example, we can consider articles, prepositions or pronouns, among others. The technique applied was the classic method in which the Stop Words were tokenized and further compared with the NLTK stop-list (Kaur and Buttar, 2018).

Given that the comments in reviews were all in different idioms, only the ones in English were, in fact, considered for this analysis, to avoid possible translation problems, namely regarding the duplicate meaning of words and expressions.

Furthermore, reviews were also removed taking into consideration Pareto's Power Law 80/20 Distribution to clean the reviews that resulted as meaningless, mostly, containing zero words after this phase. As referred by Geerolf (2017), "in the social sciences, roughly 80% of the effects come from 20% of the causes". The final cleaned dataset contains 3.294.879 reviews.

3.1.3 PRE-PROCESSING PHASE 2 – TRAINING

All the previously preprocessed data was retrieved to a csv file ("reviews-csv"), containing 60.000 reviews that were taken into consideration for the development of this phase.

To test for confounding and interaction of the data, a stratified analysis was performed to decouple geographically and chronologically the reviews that, as abovementioned, were retrieved following their cities distribution.

Furthermore, according to the power law of their strength, some reviews were removed. In this way, and in order to keep the reviews with the closest length to the mean (defined by the mid quartile), both half of the remaining percentage from the shortest reviews and half from the longest were removed. This resulted in removing 10% of the reviews. So, by the end of this phase, the dataset contained 54.000 reviews.

The final step was to split each review into sentences, removing the non-sensical words and punctuation that became useless for this effect. In accordance, Stemming and Lemmatization of words were used to prepare the data for further processing. This process allowed to detect the derivation of words, considering each one is semantically linked. It is necessary to ensure that the semantically different words must be kept separate, as well as the fact that, for the same stem/lemma, morphological forms of a word should be taken into consideration (Mohan, 2015).

To perform the stemming process the Porters' algorithm was used. Proposed in 1980, this technique regards 5 steps within which rules are applied pending one passes the conditions. In that case, the rule is accepted, and the suffix is removed, moving forward to the following step. At the end of the fifth step, the final stem is obtained. As an example, the word "agreed" through a stemming process turns out to be "agree" (Jivani, 2011).

For the Lemmatization process, the SpaCy Phyton Library was used. This library has allowed to tokenize (break the document into words), recognize the name entities, detect nouns and, lastly, convert words in the second or third forms to their first form variants. The major difference concerning Stemming is the fact that Lemmatization ensures the roots obtained are actual words in the dictionary. The image below is a representation of this difference.

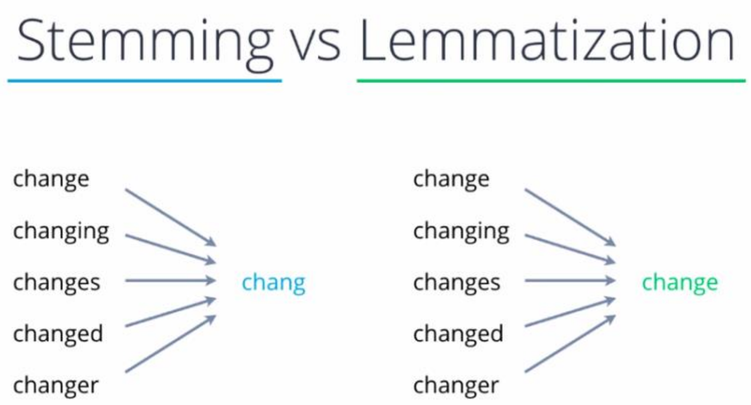

Figure 3 - Stemming vs Lemmatization. Image downloaded from https://medium.com/swlh/introduction-to-stemming-vs-lemmatization-nlp-8c69eb43ecfe in September 2020

3.1.4 SENTIMENT MODEL

The sentiment model was defined and prepared to be used in the analysis. The reviews were evaluated, being identified the text parts that match with specific sentiments or opinions.

For each sentence a polarity score was assigned from -1 to +1, meaning, respectively, very negative and fully positive. Also, a subjectivity score was provided from 0 to +1, correspondingly factual and very subjective sentence. Additionally, a capped polarity score was given considering thresholds, matching negative (-1), neutral (0) and positive (+1) sentences.

To complete this model the thresholds key values defined were 0.0, being lower or equal values considered as negative, and +0.3, being higher or equal values considered as positive sentiments. The chosen values for the polarity score went through a thorough analysis process that started with -0.33 and +0.33 obtaining a non-satisfactory outcome. Then, -0.5 and +0.5 was inspected, but as the prior testing, the results were not pleasant. Therefore, the final decision relied on unequal threshold values for negative and positive score interval. A possible reason can be the fact that the neutral sentiments mostly regard a score of approximately 0, being this interval shorter than the prior sentiments mentioned.

Further testing was performed to increase the quality and efficiency of classification. This sentiment model was created based on a convolutional and a recurrent neural network (CNN and RNN). Training, validation and testing sets were created based on, respectively, 80 / 10 /10 % split of data.

The first stage focused on training the tokenizer and Word2Vec model. In this way, tokenization and padding of data need to be ensured, in the first place, to transform the sentences into a numerical representation of the words. Furthermore, embeddings were also taken into consideration, to improve the performance and ensure the representation of similar words.

For the CNN, local characteristics were retrieved, understanding patterns to obtain the sentence embedded "opinion" features. On the other hand, with the RNN it was possible to find the overall context of the whole sentence (Wang et al., 2016). For this purpose, the LSTM (Long-Short Term Memory) unit was used to output the sentence knowledge. This type has been proved to surpass the traditional RNN Tanh unit (Chung et al., 2014). The training resulted in a 98% accuracy score.

3.1.5 TOPIC MODEL

A topic model was developed, with the dictionary obtained in 3 layers. First, there are two main topics: business and social. The second layer was created taking into consideration the MalletLDA, which, as explained before, is a statistical inference method created to construct a sample distribution to obtain the most frequent words used.

In this phase, it was necessary to identify the best number of topics to use in this study. In this way, the coherence score was plotted versus the number of topics. After analysing the following graph, the decision relied on 10 topics as the optimal number, since this is the value from which the coherence score (quality of the clusters) begins to stabilize, meaning that adding more topics would not result in significant improvements in this model, as it was verified up until that point of the graph. Since LDA in an unsupervised model, within the interval of 2 and 20 topics, several training processes took place before achieving the result of the presented topic.

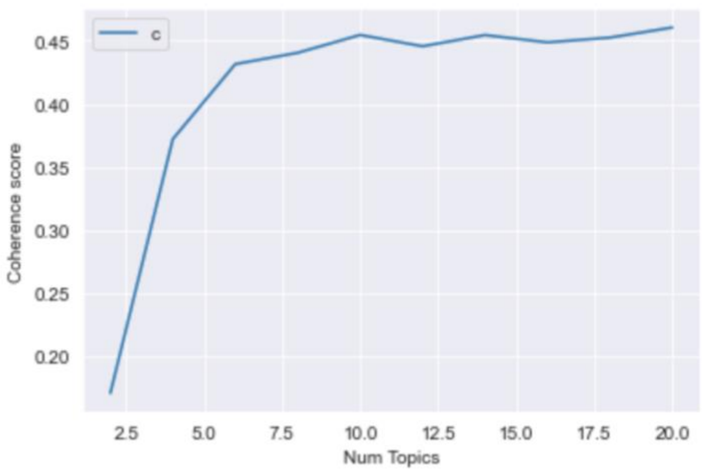

Figure 4 - Elbow Graphic

Furthermore, to ensure the correct number of topics, the topic's keywords were evaluated, as well as the interaction between these topics. In this way, the goal is to minimize the intra-clustering distance and maximize the inter-clustering distance. Therefore, the similarity within a cluster must be verified, taking into consideration the keywords defined for each topic. Moreover, each cluster must regard different topics, so, there should not be high similarities between clusters.

The figure below was obtained using the LDAvis tool, that plots the topics as "circles in the two-dimensional plane". The midpoints are, then, obtained by computing the intra-topic distance. Furthermore, multidimensional scaling is used to obtain the inter-topic distance, in a two dimensions graphic (Sievert & Shirley, 2014).

By analysing the graphic, it is possible to understand that on the upper right side there are topics similar between them. In this way, there was the need to merge clusters of topics to achieve a more accurate number of topics for this second-layer. As a result, the following 7 topics were defined as the final clusters:

- Location – corresponds to the city
- Neighbourhood– regards the neighbourhood

- Stay – described the customer experience
- Apartment – related to the accommodation architectural aspects
- Interiors – regards the decoration and features present inside of the listing
- Good Host – relates to the host interaction with the guest
- Logistic – corresponds to the organization matters

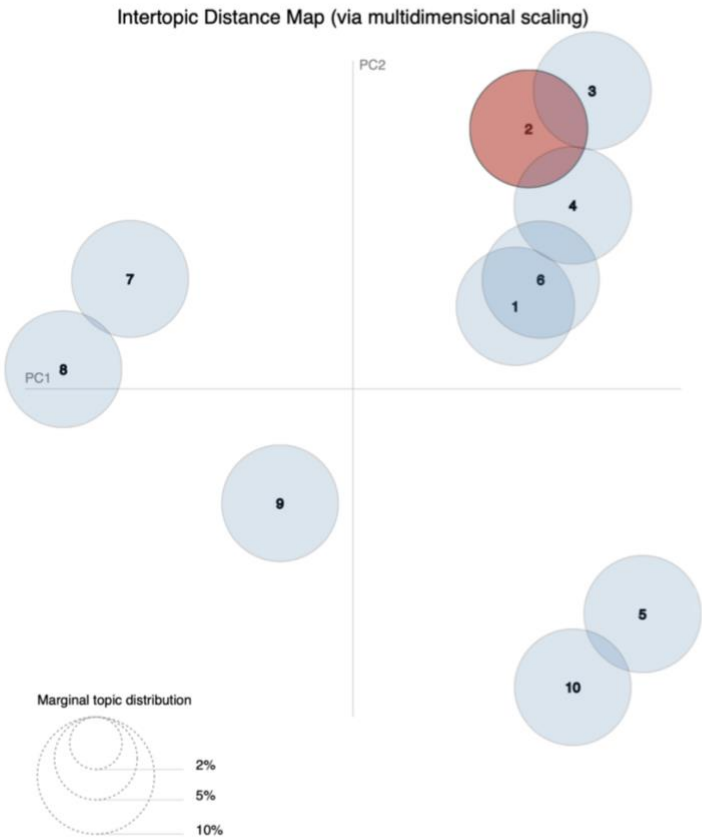

Figure 5 - Inter-Cluster Distance

Subsequently, to form the third layer, the information was refined by clustering the defined top words for each topic with the k-means algorithm, obtaining groups of homogenous clusters that are different from each other. Moreover, results were enhanced with similar words from a "word2vec" model.

The CBOW (Common Bag of Words) approach was chosen for this analysis. As identified above, this is the most appropriate method for large datasets, forecasting the target word based on its neighbouring words. The vectorial space used was 100, being each word represented by 100 real values. By understanding the closest vectors, similar words were identified, concerning the similarity score between -1 and 1. For each set of keywords, the enrichment occurred in the top 30 related words with a similarity score above 0.7. Cleaning was performed to ensure all the words that appear in more than one set were deleted from the sets, with an exception for the set in which they were more frequent. Finally, the similarity matrix was completed.

The following step was to decide on the number of sub-topic clusters to use. According to the first graphic below, and similar to the analysis performed before, the elbow method shows an optimal cluster number of 6. However, this result might not be clear to point out when analysing the elbow. In this way, the Silhouette coefficient was also taken into consideration to assess the intra-cluster consistency and cohesion. Analysing the second graphic below, it is clear that the ideal number of sub-topics to include is 6, where the consistency is closest to 1.

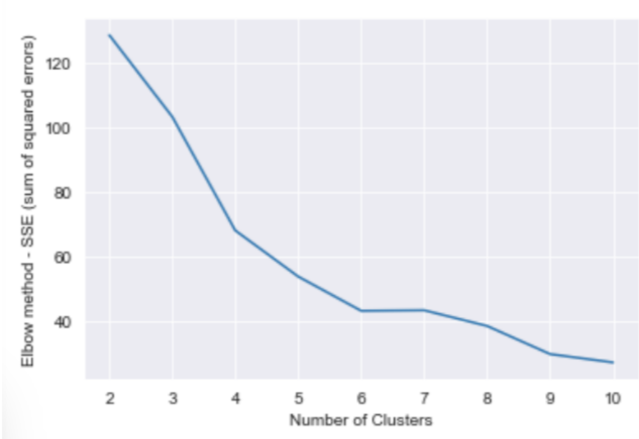

Figure 6 - Elbow method

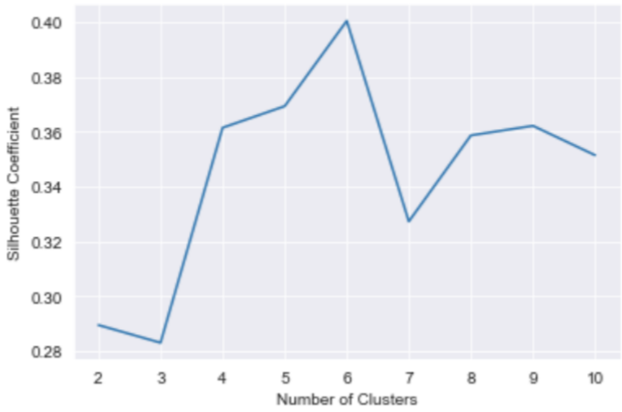

Figure 7 - Silhouette coefficient

Therefore, for each topic obtained in the second layer abovementioned, the sub-topics were taken into consideration. It is very important to point out that the goal of the third layer is to ensure the final chosen topics are clearly representative of the reviews' most frequent themes. So, for that reason, some second layer topics were not divided into sub-topics, as they were representative and clear enough.

Finally, with a fine-grained dictionary prepared, the topic model phase was concluded with 15 topics: advice, apartment, booking, communication, experience, good host, interiors, location, meal, neighbourhood, noise, shops, surrounding, transport, and trip.

3.1.6 ANALYSIS OF RESULTS

The following goal is to proceed to the text mining of reviews, analyzing each one according to the sentiments expressed and if either social or market norms are present. This segmentation results in a more organized dataset in which the sentiments and types of norms that influence the reviews are clearly specified and scored.

With the trained dataset and both sentiment and topic model prepared, a set of approximately 150.000 reviews were retrieved from the database to be processed and analyzed. Regarding these reviews, first, per review, a matrix was obtained assigning values for each review sentence, taking into consideration both topic, sentiment, subjectivity and the capped sentiment value, as previously modelled. With the processed reviews, the focus was the sentences that were, then, split and paired with the classifications.

With the organized dataset, the second part consisted of assessing the relationships between variables. The goal will be to verify the proven relationship analysis from previous studies and various hypotheses formerly presented.

From table 1 and analysing figure 8 below, it is possible to verify that, on average, each review has 54 words, with an average of 49 words of standard deviation from the mean, which regards a high deviation and, consequently, high length variability. Moreover, we can expect reviews with only 1 word or 604 words, which was the longest review retrieved.

	Mean	Std	Min	25%	50%	75%	Max
Words	53.82	48.59	1.0	21.0	41.0	72.0	604.0

Table 1 - Word Count Statistics

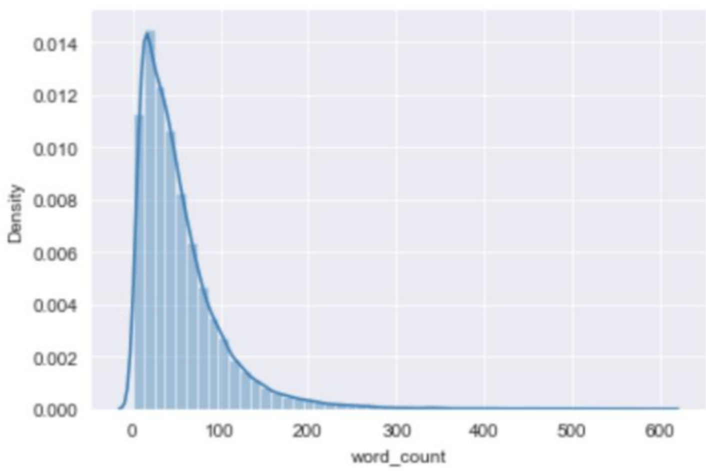

Figure 8 -Word Count Graphic Distribution

From table 2 and figure 9 below it is possible to verify that each review has around 4 sentences, with an average of 2.6 sentences of standard deviation from the mean. Besides, we can expect reviews with only 1 sentence or a maximum of 35 sentences, being this the highest verified value.

	Mean	Std	Min	25%	50%	75%	Max
Words	3.90	2.61	1.0	2.0	3.0	5.0	35.0

Table 2 - Sentence Count Statistics

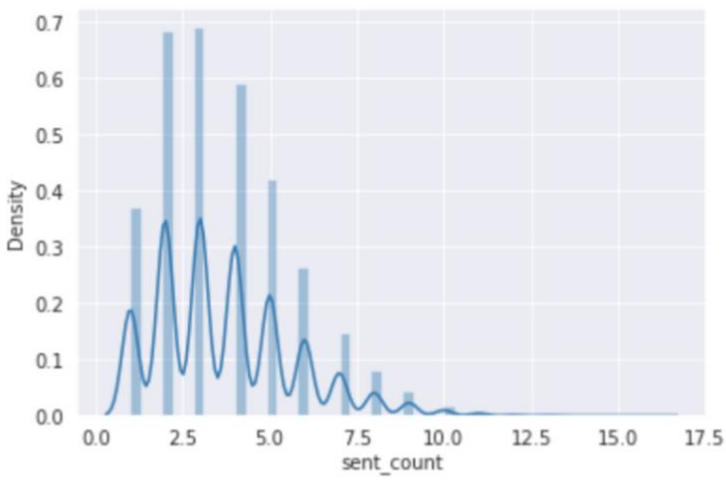

Figure 9 - Sentence Count Graphic Distribution

Regarding the first layer of topics, these reviews concern in its majority 52% social topics and approximately 48% regard business themes. Despite having very similar percentages, the social-oriented vision is identified with more frequency in this study reviews.

Norms	Value
Business	48.071053
Social	51.928947

Table 3 - Norms Distribution

Figure 10 - Business vs Social Distribution

Analyzing the polarity of sentences, it was possible to obtain an overall of these review sentiments classified in negative (-1), neutral (0) and positive (+1) sentences. Approximately 56% regard positive reviews, which means that this study's reviews are mostly identified with a positive sentiment, whereas negative and neutral reviews are almost equally distributed by around 20% each, as it is possible to verify through the table and figure below.

Given these broad results, more exhaustive sentiment analysis was performed to achieve more detailed scores.

Polarity	Value
-1	22.904651
0	21.593962
1	55.501388

Table 4 - Sentiment Polarity

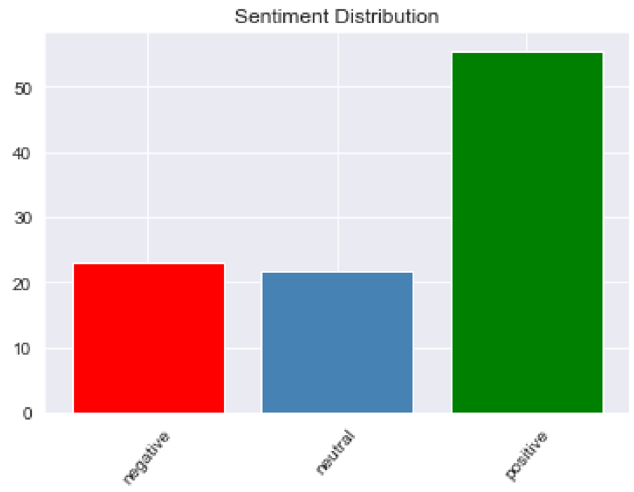

Figure 11 - Sentiment Distribution

Connecting the two major norms of this study, namely, business and social, with the sentiments expressed by the reviewers, it was possible to realize that business reviews regard mostly positive opinions (above 25% of reviews). Whereas neutral and negative reviews are similarly distributed, meaning that these last two appear with less frequency than positive sentiments, having the neutral sentiment (with around 11% of reviews) a higher impact than the negative (with approximately 9% of reviews), in this Business scenario.

On the same hand, but with a notorious difference, these results are corroborated in social reviews also. However, in this case, the negative sentiments have a higher expression when compared to neutral sentiments. The graphic below can suggest that, even if it is a minor percentage difference (around 4%), when it comes to social-oriented reviews, the guests can tend to have stronger sentiments, namely regarding negative feelings, which represent near 14% of these reviews. On the other hand, neutral sentiments represent 10% of social-oriented reviews. However, this can be only data oscillations.

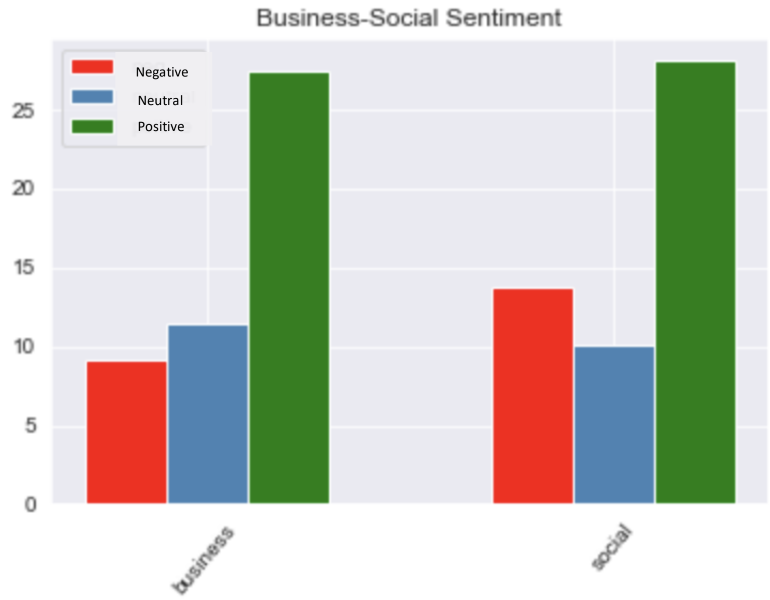

Figure 12 - Sentiment Distribution per Business and Social Reviews

Taking into consideration the **subjectivity** score that, as abovementioned, results in values from 0 to 1, regarding namely factual and very subjective sentences, the graphics below show the verified subjectivity values in this study's reviews, respectively overall and, secondly, business and social-oriented results.

Analyzing the following graphic, it is possible to retrieve that more than half of these reviews are classified with a subjectivity score equal to or higher than 0.4. Specifically, around 72% of the reviews are strongly subjective, which leaves less than 30% of reviews with low subjectivity and, therefore, more precise opinions.

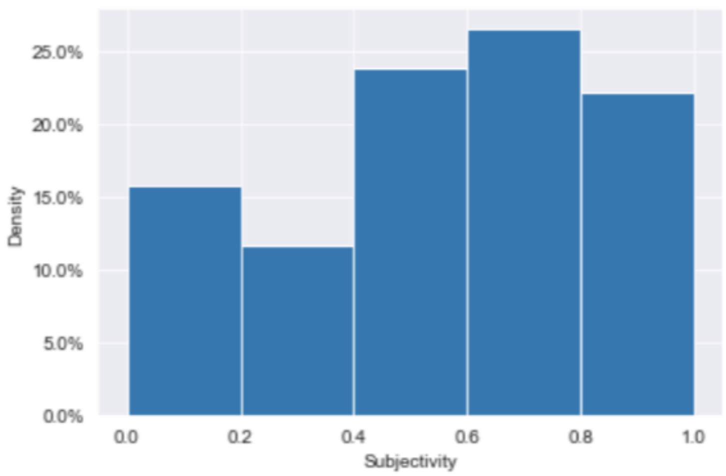

Figure 13 - Overall review's subjectivity score

It was also possible to obtain a norm's segmented vision. Taking this into consideration, the graphic below shows reviews related to business norms follow the subjectivity distribution previously described. When compared with social-oriented reviews, in this case, it is possible to state that reviews that reveal opinions related to social interaction are more subjective than reviews that regard business topics, namely, the apartment, interiors, among others. Particularly, more than 75% of social reviews represent medium or high subjectivity scores, between 0.4 and 1, whereas is the business case, these concern around 70%, as explained above.

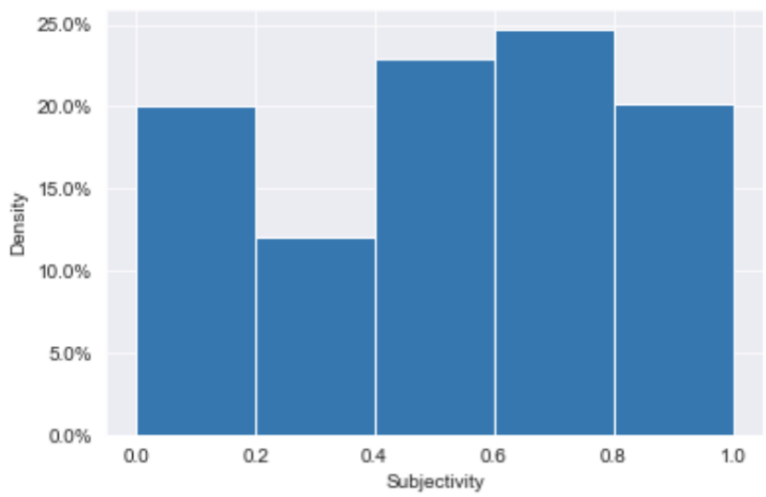

Figure 14 - Subjectivity score of business-oriented reviews

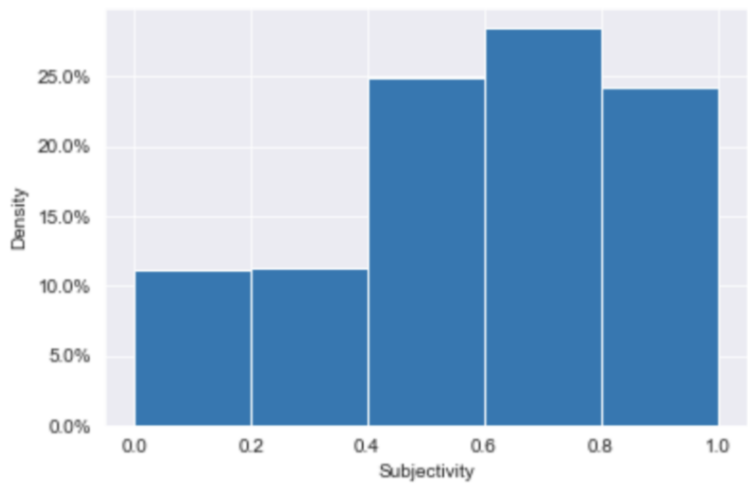

Figure 15 - Subjectivity score of social-oriented reviews

Regarding the analysis of the **topics** mentioned in the customer reviews, all the defined topics have a positive mean, however, through the analysis of the standard deviation values present in the table below, it is possible to understand that some topics can have negative sentences. As an example, we can consider the “Advice” or “Interiors”, “Location”, “Neighborhood”, “Noise”, among others. These results show that there is a lot of variability between topics’ values.

Topic	Mean	Standard Deviation
Advice	0.24	0.32
Apartment	0.42	0.28
Booking	0.32	0.29
Communication	0.47	0.35
Experience	0.66	0.23
Good Host	0.40	0.27
Interiors	0.19	0.29
Location	0.14	0.17
Meal	0.34	0.33
Neighbourhood	0.23	0.26
Noise	0.13	0.31
Shops	0.24	0.29
Surrounding	0.29	0.23
Transport	0.22	0.26
Trip	0.32	0.33

Table 5 - Sentiment's Real Value Distribution per Topic

Taking into consideration both business and social norms, the graphic below shows evidence that business topic “Apartment”, which is represented in red, demonstrates a higher weight in reviewer’s sentences, being, therefore, the main keyword more present in these reviews, with more than 30% of frequency.

When weighting social topics, the “Advice” is the most verified topic (more than 20%), followed by the customer “Experience” that is the second social-related topic more present in Airbnb reviewers' opinions (around 10%). Additionally, but with less frequency, guests also refer to the "Trip" details in their reviews (about 7%).

It is also possible to understand that topics such as "location", "surrounding", "communication", "noise", "meal" or "shops" are not frequently reviewed (between 1-2%), when compared to other topics. This can mean that these themes might not be as important for reviewers as the others.

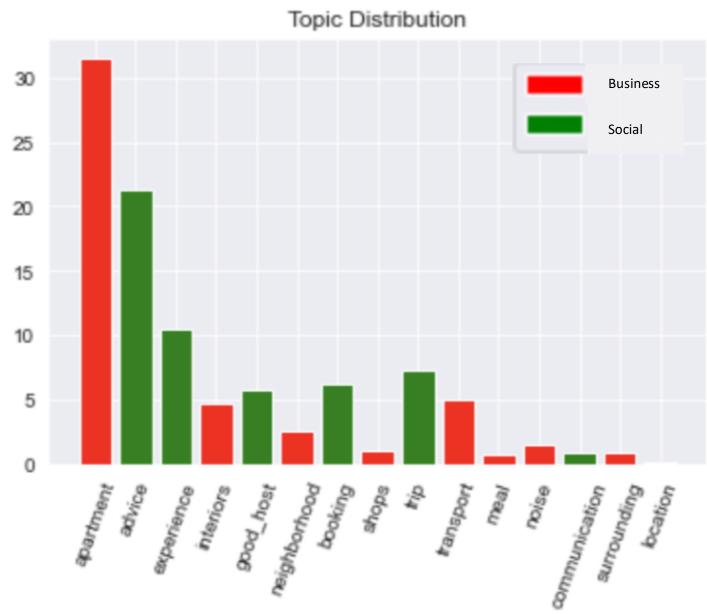

Figure 16 - Topic Distribution

As mentioned above, it was verified that the general sentiment distribution regards 56% of positive feelings and evenly around 20/20 for neutral and negative sentiments. Nevertheless, when normalizing the topics and sentiments distribution by data, which means the percentage values were used instead of the absolute values, the identified conclusions are different across the mentioned topics. In this way, the following paragraphs describe the most common topic's distribution per reviewers' sentences, stating the conclusions reached through the analysis of the graphic below.

The apartment topic is more often referenced positively in sentences, with over 20% of positive frequency. Additionally, but on much less scale, this topic does also appear in reviews considered neutral and, lastly, negative (less than 5%).

The advice received is a topic normally mentioned in negative sentences. However, the difference from positive sentences is not that high (roughly 2% difference), meaning that advice can usually be considered negative or positive by reviewers. The ones whose advice received was negative are slightly more frequent (approximately 9%).

On the other hand, the overall experience is considered by reviewers to have been, in its majority, positive (nearly 10%). Some have also measured their experience as neutral (1%), but not negative.

Regarding the host being a “good host”, the reviewer’s opinion around this matter is mostly positive or neutral (between 2-4%). A negative connotation, in this case, is verified, but on small scale.

When expressing themselves about the “booking”, overall, the reviewer’s present a positive meaning (around 4%). Other situations might reveal the booking as having negative or neutral importance for the review (almost 2% each).

Taking into consideration the Trip itself, Airbnb customer’s reviews reveal a mostly positive sentiment about the trip (about 4%). However, some consider it as negative or neutral, being the percentage of negative reviews close to the positive ones, differing only around 1% under. Topics such as interiors, neighbourhood or shops have a smaller expression, when compared to the abovementioned.

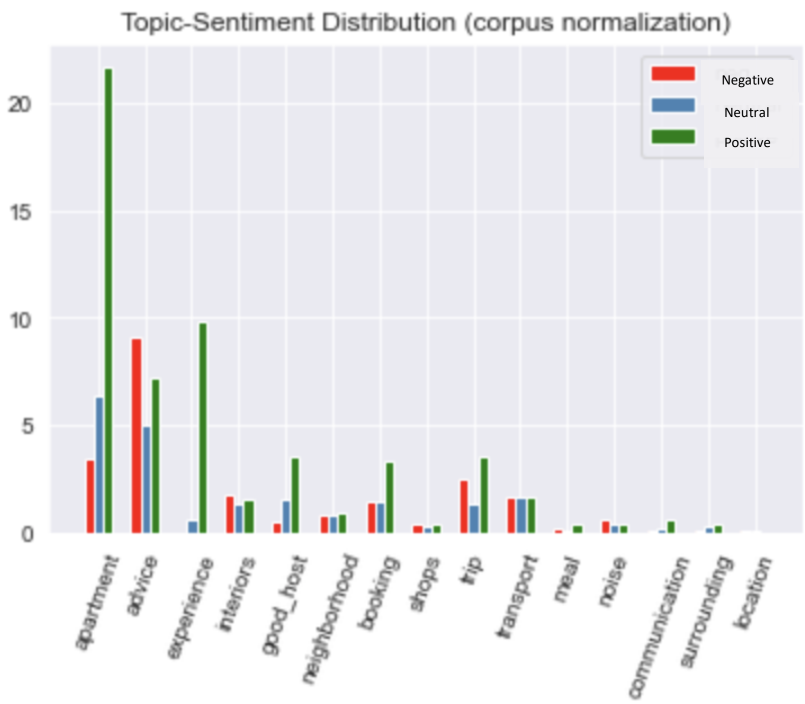

Figure 17 – Sentiment/Topic Distribution Normalized by Data

In another approach and to support the analysis performed above, if the data is normalized by topic (graphic below), which means, in this case, the percentage to be analyzed is calculated with the number of data per topic, almost every subject is, on average, more present in positive sentences. However, when compared to the positive sentiment, the negative sentiment represents a strong presence in most of the topics with a clear exception for when the reviewers are talking about the apartment, their experience and the good host.

When analyzing the neutral sentiments, these types of opinions are, on average, equally distributed amongst topics, meaning that these are mainly distributed between 20% and 40% of occurrence per topic. The neutral sentiment is less verified in the experience topic, where the frequency regards only approximately 1%. The highest percentage of occurrence concerns the location topic. Therefore, some reviewers mention the location as a neutral opinion, in this way, neither being satisfied or dissatisfied with it.

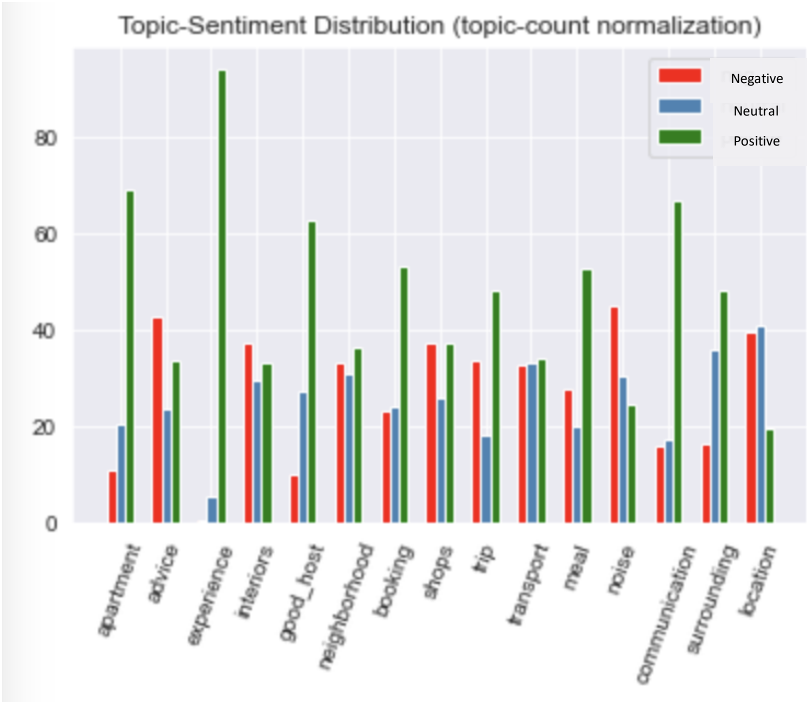

Figure 18 - Sentiment/Topic Distribution Normalized by Topic-Count

To better understand each Topic's distribution, the following table represents the sentiment's mean by topic. It is possible to identify similarities between groups, such as:

- Communication, Apartment and Good Host;
- Meal, Trip and Booking;
- Shops and Advice, Neighborhood and Transport;
- Location and Noise.

Topic	Mean	Standard Deviation
Experience	0.66	0.23
Communication	0.47	0.35
Apartment	0.42	0.28
Good Host	0.40	0.27
Meal	0.34	0.33
Trip	0.32	0.33
Booking	0.32	0.29
Surrounding	0.27	0.23

Shops	0.24	0.29
Advice	0.24	0.32
Neighbourhood	0.23	0.26
Transport	0.22	0.26
Interiors	0.19	0.29
Location	0.14	0.17
Noise	0.13	0.31

Table 6 - Singular Topic Sentiment Distribution

The pie charts below represent a visual graphic of the detailed distribution of each sentiment across all topics, to better understand the mean values identified above. First of all, the *experience* topic mean score is the highest (0.66), which reveals strong opinions, mainly positive (94%) and almost none negative. On the other hand, the topics *communication, apartment* and *good host* demonstrate a similar distribution, being all between almost 63% to 68.8% of positive opinions expressed regarding them.

From the figures below, it is possible to understand that communication's sentiment distribution is affected almost evenly by the weight of neutral (17.4%) and negative sentences (16%), resulting in a lower positive mean value of 0.47. Whereas apartment and good host's positive mean distribution (0.42 and 0.40) is mostly affected by Neutral opinions, respectively 20.3% and 27.2%.

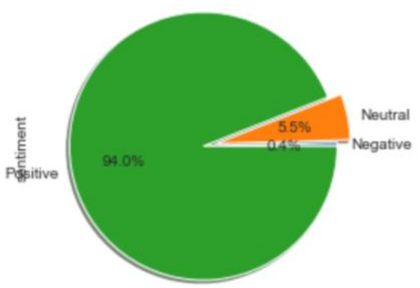

Figure 19 - Experience Sentiment Mean

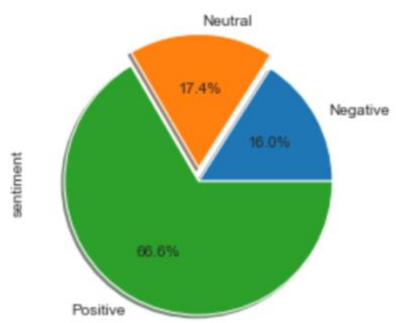

Figure 20 - Communication Sentiment Mean

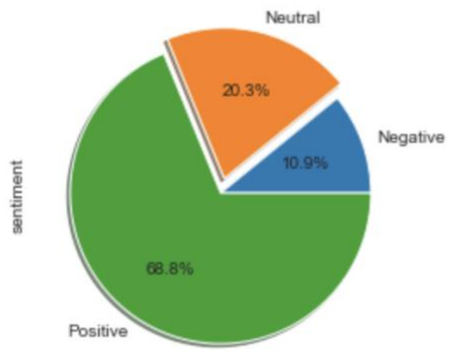

Figure 22 - Apartment Sentiment Mean

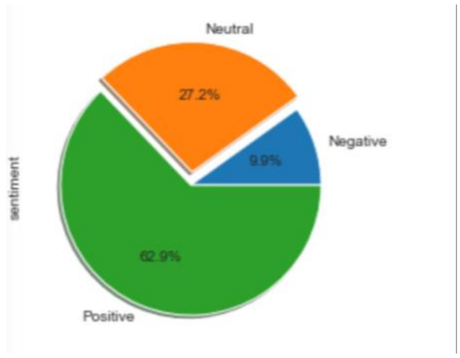

Figure 21 - Good Host Sentiment Mean

When analyzing *meal* and *trip* topics, it is notable that these topics' mean distribution (0.34 and 0.32) is visibly caused by the weight of the negative sentences showed in color blue in the pie chart below. In the *meal* topic, positive sentences regard around 52% of reviews, whereas in 48% is the percentage of positive reviews verified concerning *trip*.

Booking, on the other hand, has the highest positive sentences value of this group (53%) and a lower percentage of negative sentences (23%) than the prior two topics. However, verifies a lower mean (0.32) than, for example, *meal* (0.34). In this case, what can happen is this topic being more affected by negative and neutral sentences together than the *meal* topic, resulting in a lower mean distribution than this last topic.

Surrounding's registered score, 0.27, represents approximately 48% of positive and 35.6% of neutral sentences. Similarly to the case mentioned above, this topic has a positive sentences value (48.2%) parallel to the *trip* topic's value (48.1%), however, has a significantly lower negative sentences percentage, with 17.4% less than *trip*, and, yet, this last topic has a higher mean score (0.32). This can mean that either the negative sentences have high strength, or the neutral sentences have a strong impact, in such a way that it impacts significantly, decreasing the *Surrounding* topic mean.

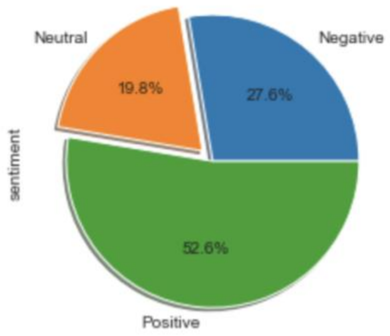

Figure 24 - Meal Sentiment Mean

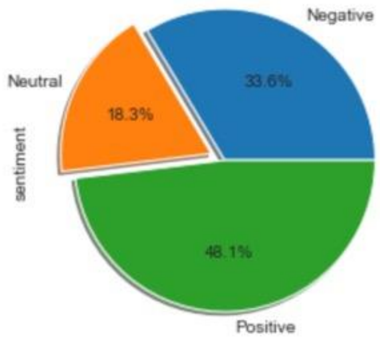

Figure 23 - Trip Sentiment Mean

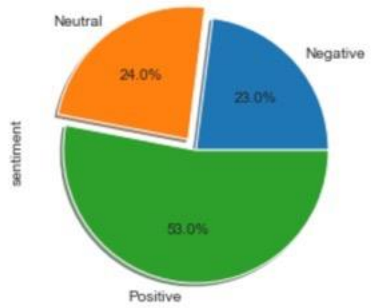

Figure 25 - Booking Sentiment Mean

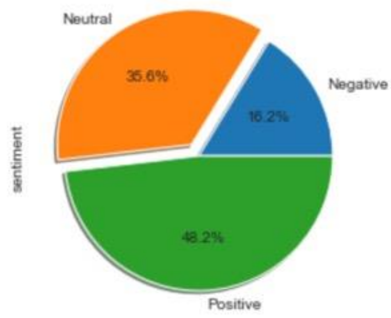

Figure 26 - Surrounding Sentiment Mean

Shops and *advice* have a mean of approximately 0.24, which is a lower value, when compared to the topics above. These topics distribution is impacted by the high weight of the negative sentiments, respectively 37.1% and 42.8%. Nevertheless, its connotation is only positive due to both positive and neutral sentences, that together contribute to the positive meaning of “Shops” and “Advice” in the reviews. The same conclusion is applied to *neighbourhood* and *transport*.

An aspect to point out is the fact that *advice* has the highest negative value in this group and a higher mean than *neighbourhood* and *transport* topics that, respectively, weight 0.23 and 0.22. In fact, the *neighbourhood* contains more positive (36.1%) and fewer negatives (33%), which suggests that the weight of the negative sentences in the reviews regarding *advice* can have less strength than in the reviews regarding the *neighbourhood* or *transports*.

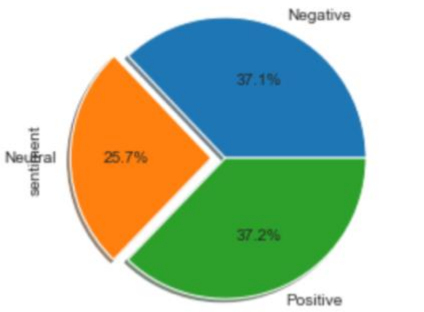

Figure 27 - Shops Sentiment Mean

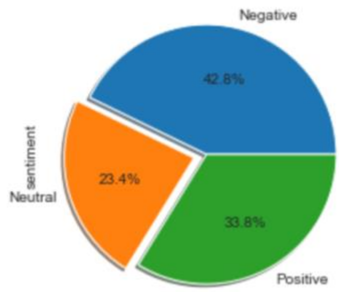

Figure 28 - Advice Sentiment Mean

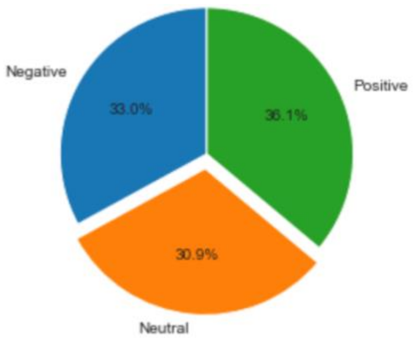

Figure 30 - Neighbourhood Sentiment Mean

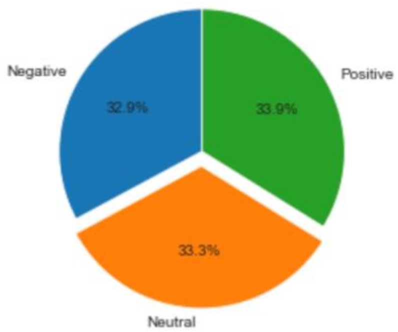

Figure 29 - Transport Sentiment Mean

The last three topics contain the lowest mean of the studied topics. *Noise* has the highest negative value, almost 45%, being, therefore, the lowest mean of all the topics, with 0.13. Regarding *interiors* and *location*, this last one has 39.7% of negative sentences, whereas the first one contains 37.3%. The *location* has a higher percentage of neutral sentences (40.7%) when compared to *interiors* (29.5%). For positive sentences, *interiors'* reviews are classified as positive in a frequency of 33.2%, while *location* reviews correspond to 19.7% of positive sentences. Following these values, *location* and *interiors* score, respectively, a mean of 0.14 and 0.19.

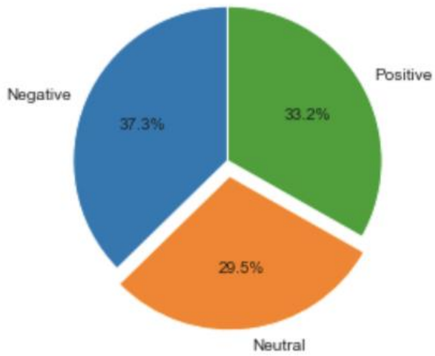

Figure 31 - Interiors Sentiment Mean

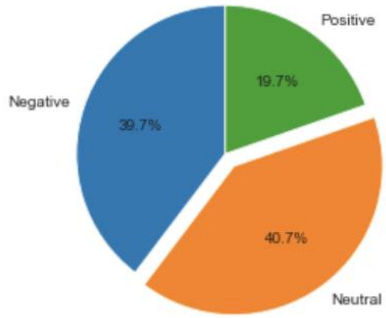

Figure 32 - Location Sentiment Mean

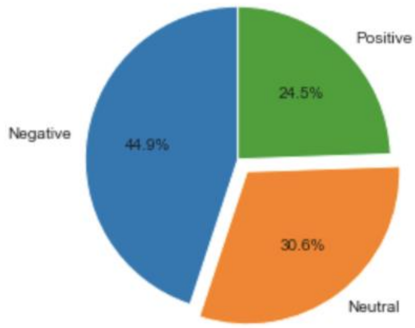

Figure 33 - Noise Sentiment Mean

Analyzing, now, the perspective from the type of sentences sentiments, regarding the reviewer's negative sentences, it is possible to understand that both the *advice* received, the *apartment* they have stayed in and the *trip* itself are the most mentioned topics in negative sentences, corresponding, respectively, to almost 40%, 15% and 10% of negative sentences. *Experience* and *location* are the less mentioned with less than 1% of occurrence.

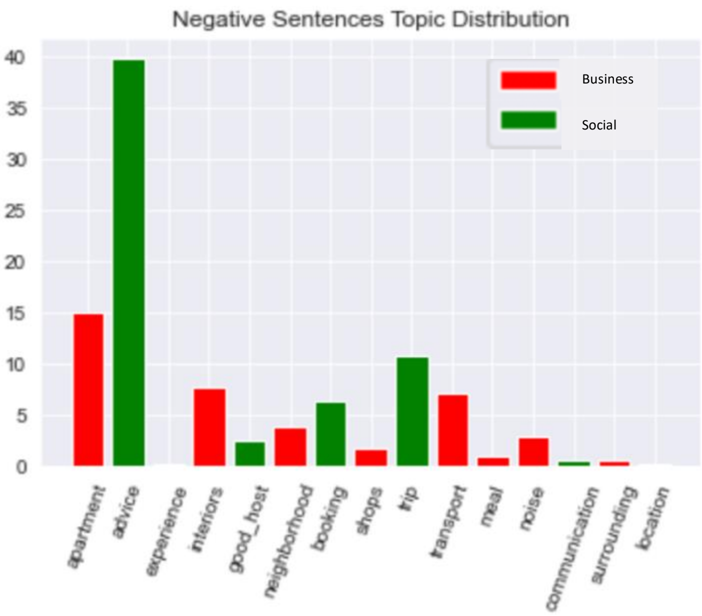

Figure 34 - Negative Sentences Topic Distribution

Taking into account the neutral sentences, the most mentioned topics overall are the *apartment* (29.6%), as well as the *advice* received (23.12%). Customer's neutral sentences do not refer as frequently to the topics such as *location*, *meal and communication* (less than 1% of occurrence).

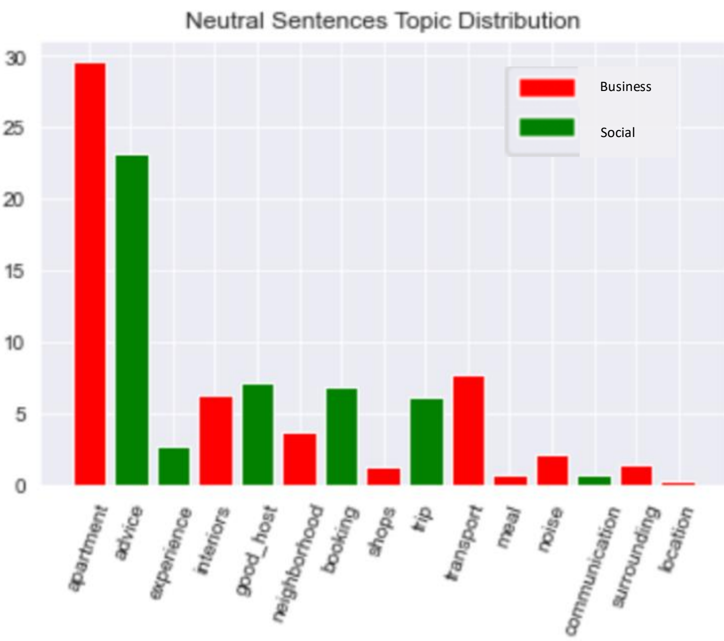

Figure 35 - Neutral Sentences Topic Distribution

When customers' reviews are considered positive, the most frequent topic referred to is the *apartment* (39%). *Experience* and *advice* are also the topics of positive reviews, respectively, 17.7% and 12.9%. *Location*, as it has been noticed, is a topic that is almost not reviewed across neutral, positive or negative sentiments (<1%).

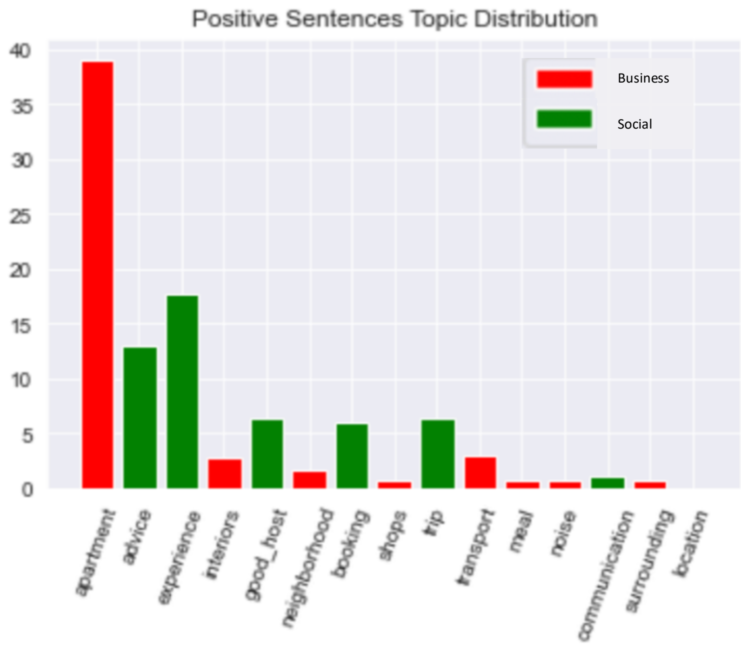

Figure 36 - Positive Sentences Topic Distribution

To complete the analysis, the **top 10 words** of each topic were also identified. The most verified words for each topic ordered by frequency are specified below. It is possible to state that the words are not exclusively used in each topic, in this way, the word "great" or "stay" are referenced in more than one topic, being the same word used in reviews regarding different themes.

- **Apartment** - place, apartment, clean, nice, great, stay, comfortable, location, really, recommend;
- **Advice** - recommend, highly, location, thank, thanks, arrival, perfect, good, definitely, quick;
- **Experience** - great, location, host, stay, lovely, amazing, wonderful, fantastic, hosts, experience;

- **Interiors** - room, kitchen, bathroom, bed, small, big, shower, living, bedroom, private;
- **Good Host** - host, helpful, friendly, kind, accommodating, responsive, great, welcoming, stay, really;
- **Neighbourhood** - quiet, area, neighbourhood, distance, walking, short, walk, location, restaurants, minute;
- **Booking** - easy, time, check, great, access, stay, location, left, return, definitely;
- **Shops** - nearby, local, food, restaurants, corner, eat, grocery, places, shopping, good;
- **Trip** - stay, home, house, definitely, feel, experience, recommend, enjoyed, make, really;
- **Transport** - close, walk, station, location, train, central, walking, minutes, city, convenient;
- **Meal** - coffee, breakfast, tea, delicious, morning, bread, milk, provided, fruit, fresh;
- **Noise** - night, bit, people, noisy, little, loud, street, stay, noise, location;
- **Communication** - communication, great, good, host, flexible, excellent, check, contact, stay, process;
- **Surrounding** - lot, near, restaurants, thanks, location, plenty, good, interesting, city, close;
- **Location** - public, transport, transportation, city, location, good, restaurants, near, transit, center;

3.2. SCORE RATING MODEL

Given a dataset extracted from the data analysed in the previous stage, the predictive model will learn how to label reviews based on the feature sets and behaviour of data defined to teach this model.

The structure of the predictive model needs to be well defined right from the beginning of this phase. In this way, the different possibilities of variables to select as features for the predictive modelling were evaluated.

From the sentiment and topic modelling performed earlier, it is possible to obtain as key features the polarity score, subjectivity score and different topics reached. The difference between them is that this last one is categorical. Thus, encoding was used to ensure the topics are suitable for processing. One-Hot Encoding (OHE) was used to convert the categories into numerical representations. This encoding considers all the categories a topic relates to and applies a binary variable for each value, as a dummy variable.

Moreover, the length of reviews is also an important feature, however, the reviews do not always have the same length. In this way, truncation and padding, as already mentioned before, were used to ensure longer reviews are shortened (truncated) and shorter reviews are extended with the padding values.

To define the possible feature sets, several combinations of variables can be used. In this way, all the possibilities are defined in the table below. Set of variables 1, 2 and 3 use the average values of the feature sets described above. Whereas, feature sets 4 to 9 regard merged variables. For this purpose, topics and sentiment are merged in one vector, for example, in the case of feature set 4, the following assumption is made: topic score = polarity. Additionally sets 5 and 6 are extensions of this assumption. On the other hand, set 7 corresponds to the polarity score mitigated by subjectivity, which, as verified before can impact the model and bias the results. This formula for this scenario is *topic score = polarity ∗ (1 − subjectivity)*. Further testing will be performed to assess the best set of features to guarantee the best model performance.

Feature Set	Content
1	Topic OHE + average polarity + average subjectivity
2	FS1 + review length
3	FS1, normalized by review length
4	Topics vector as a counter for polarity score
5	FS4 + review length
6	FS4, normalized by review length
7	Topics vector as a counter for polarity score, mitigated by subjectivity score
8	FS7 + review length
9	FS7, normalized by review length

Figure 37 - Predictive Model Possible Feature Sets

3.2.1. LABELLING THE DATA

To start the labelling process, a sample of the data was extracted and categorized by two coders to ensure the different perspectives are taken into consideration. The process of labelling the data included using 5 rating classes, namely from 1 to 5, where 1 means very dissatisfied and 5 means very satisfied.

With the data categorized, the accuracy of the classifications needed to be guaranteed. In this way, the ratings were validated using the following agreement strategy. First, if there was a difference in classification given and it was higher than 2, those samples would be removed from the dataset. For example, if one coder classified as 2 and the other as 5. On the other hand, if the difference was exactly 2 classes, then, the ratings' mean would be applied. So, in the case were one coder categorized as 3 and the other as 5, the final rating would be 4. Furthermore, when the difference was 1 the lower rating score was the one used. In this way, with a review labelled as 3 by one coder and 4 by the other, the final rating would be 3.

The following step encompassed evaluating all the results and understanding if the data obtained, after the validation process abovementioned, was truthful and suitable for the purpose. During the process, it was verified that it would be necessary to label more data, in order to guarantee that the data to create the predictive model is representative enough for the model to learn from the most accurate examples possible.

Taking this into consideration, a positive bias issue was detected in the data. Precisely 85% of the sample was categorized as 5 stars, which would lead to an unbalanced dataset that is exactly what cannot be used for training the predictive model, since it does not correspond to the whole scenario we need the model to learn. As explained above, more data extraction and more labelling took place.

The final dataset resulted, then, in 6000 rated reviews. As it is shown in the following pie chart, 31.1% of reviews are 5 stars, being 26.3% labelled as 4 stars. Whereas, 16.1% are considered 3 stars and 2 and 1 stars were given to, respectively, 13.7% and 12.8% of reviews. Moreover, the 1st star class is the class with less data, around 770 samples. In this way, there was the need to balance the dataset using this number of reviews for each of the labelling classes. The final result was 3850 samples used for training the model. Further ahead, 385 were used for the testing phase.

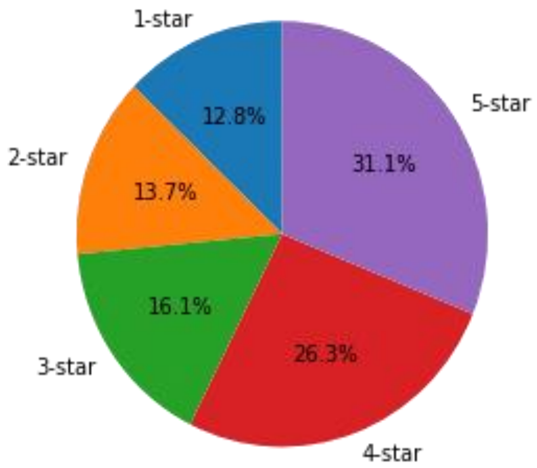

Figure 38 - Class Distribution

3.2.2. BUILDING THE PREDICTIVE MODEL

To start creating the model it is necessary to decide on which approach to use. First, a regression can be the solution if the goal is to ensure the model performs the labelling process as the coder would. In this case, the model learns the stars and categorizes the data into those ratings quantitatively in integers. On the other hand, if the goal is to teach the model to learn the categories itself and not the numbers, teaching the model to classify and knowing in which category it should be included, then the classification should be used.

Both regression and classification will be tested to understand the best solution. To train and test the model, the method K-folds cross-validation (CV) was used. In this way, the process was divided into a training (90% of data) and, further, testing set (10% of data).

For the purpose, the data was split into 5 folds. All the possible combinations of folds were created, and the training was performed with 4 folds (3465 data), leaving the remaining to assess the model performance. With the estimates of performance, the 95% confidence intervals and mean model performance were measured. With the model defined, the testing occurs with 385 data.

3.2.2.1. FEATURE SET TESTING

To test the model's performance with all the feature sets and understand the best set to include in the final model. The goal is to find the set with the best R-squared value, which means the reviews are highly explained by the chosen variables. This choice needs to be aligned with the accuracy, whose value must be the highest possible, since the bigger it is, the higher is the number of correct predictions the model has performed.

From the analysis of the table below, it is possible to understand that the accuracy is higher in the feature sets that consider the polarity mitigated by subjectivity score, obtaining, not only, more accurate results, as well as higher R-squared values, which means the model has a high explanatory capacity. For this reason, the chosen feature set is the 7th, as it is the point from where the explanatory capacity will start to decrease, as well as the accuracy. This set represents the optimum combination of values.

Feature Set	R^2	Accuracy
1	0.606	0.661
2	0.563	0.662
3	0.566	0.661
4	0.708	0.761
5	0.712	0.760
6	0.713	0.746
7	**0.875**	**0.883**
8	0.853	0.878
9	0.846	0.860

Figure 39 - Feature Set R-Squared and Accuracy Values

3.2.2.2. REGRESSION

The goal of the regression model is to understand the relationship between the reviews and the defined feature set, namely, the topic's vector as a counter for polarity score, mitigated by subjectivity score. The final goal is to predict the quantitative rating (1-5) of the reviews, taking into consideration all these variables.

Several regression models were tested. The table below shows the achieved results, obtained using K-folds, as explained earlier. The decision of the best model relies on the evaluation of the error of the model through the Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), as well as, finding the best R-squared value, which ensures the determination coefficient (R^2).

The model that proves to have a high explanatory capacity and, therefore, being a good predictor is the XGB, in which the R^2 equals 0.899, the highest R-squared value. Moreover, the errors of this model are very low when compared to others.

Thus, it is verified that the linear regression models are not suitable in this case, where the problem seems to be correctly tackled with a non-linear model, in this case, decision-tree-based XGBoost.

Model	MAE	MSE	RMSE	R^2
Linear	0.884	1.287	1.134	0.357
Ridge	0.887	1.286	1.134	0.357
Lasso	0.886	1.288	1.135	0.356
Decision Tree	0.162	0.272	0.521	0.864
Gradient Boosting (GB)	0.418	0.300	0.548	0.850
eXtreme-Gradient Boosting (XGB)	**0.237**	**0.202**	**0.449**	**0.899**
SVR	0.528	0.630	0.794	0.685

Figure 40 - Performance of Regression Models

3.2.2.3. CLASSIFICATION

Given the labels defined, the classification model will predict the categories in which the set of reviews fit. The evaluation of the most accurate model will identify the percentage of correctly classified examples from all predictions performed.

Additionally, precision needs to be taken into consideration to measure the result relevancy, in order words, to understand how much of the predicted reviews were correctly assigned to the specific category. Recall measures how much of the actual reviews were correctly rated.

Furthermore, the F1 score can be evaluated to see the weighted average between precision and recall, when there is an uneven class distribution.

Several classification models were tested. The table below shows that the overall results are better than when using regression models. For classification, the best model is the XGB with 90.7% of accuracy, approximately 0.91 in a total of 1 in both precision, recall and F1 score, being these the highest values of all the classification models.

Model	Accuracy	Precision	Recall	F1
Logistic	65.12	0.663	0.651	0.646
KNN	81.36	0.814	0.814	0.812
Decision Tree	78.47	0.818	0.785	0.790
Extra Trees	75.64	0.766	0.756	0.756
Random Forest	87.66	0.879	0.877	0.877
GB	88.16	0.883	0.882	0.882
XGB	**90.70**	**0.909**	**0.907**	**0.907**
Lin. SVC	63.36	0.642	0.634	0.630
SVC	82.10	0.822	0.821	0.820

Figure 41 - Performance of Classification Models

4. DISCUSSION AND CONCLUSIONS

This thesis proposed a holistic approach regarding how Airbnb accommodation features and hosts' attributes influence guest's review score sentiment and how are the main topics distributed. After thoroughly cleaning and training the data, a sentiment model was defined, and the topic model was developed. Therefore, obtaining a fine-grained dictionary prepared for processing the reviews. A text-mining technique was applied in an unsupervised approach to analyse the predominant topics identified by Airbnb guests when expressing their opinions regarding their rental experiences.

For topic modelling the MalletLDA method retrieved *advice, apartment, booking, communication, experience, good host, interiors, location, meal, neighbourhood, noise, shops, surrounding, transport* and *trip* as the most frequent topics used. Regarding the distribution of the main topics, *apartment, interiors, neighbourhood, shops, transport, meal, noise and surrounding* are considered business-oriented, referring to the Airbnb accommodation features. The other topics are social-oriented, with an exception for the location that is verified in such low scale to be distinguished.

Based on the extracted results, reviews expressing positive sentiments were very frequent on the business-oriented topic *apartment*. However, taking into consideration the first layer of topics (business vs social), the positive sentiments were more common for social-oriented reviews, with topics such as the *experience, advice* or *good host.* In this way, since these social-oriented topics present a positive connotation in guest's reviews, H1a is supported. In detail, from this analysis, it is possible to confirm that host interaction related words seem to be used in reviews to express positive feelings overall. Therefore, contributes to positive sentiments.

As identified in prior studies and mentioned before, the interaction with the host is represented by the "helpfulness, flexibility and good communication", alongside with a "careful clarification of destination attractions aligned with a host's thoughtful service" and "authenticity and providing access to the local experience". Kmeans method and further enhancement with word2vec model allowed to obtain the groups of words that are the most frequent per topic. In this way, it is possible to prove that this concept of host and guest interaction is exactly what these three social-oriented topics refer to, namely, regarding the *advice* topic, the keywords were found to be “recommend” and “highly”. Furthermore, within the *experience* topic, the top words were “great” and “location”, whereas, for the *good host*, the main words to point out are “hosts” and “helpful”.

Regarding the guest’s sentiments overall, in this analysis, positive sentiments are more verified than negative and neutral combined. In this way, the reviews tend to be positive either regarding business or social norms. However, when it comes to negative feelings, these are more frequent in social-oriented reviews, which means that more negative opinions can be detected when guests give their view on the *advice* received rather than the *apartment*. The advice seems to be both referenced in negative and positive reviews, with a higher weight in the negative case. This can occur since the advice in this study is perceived as, not only the host’s advice, but also the advice the guests are giving to other guests in the reviews and that can be negative if their Airbnb experience was not pleasant.

The main question that this paper proposed to answer is “How does the Airbnb market norms and social norms, along with guests’ sentiments, influence the reviews? And how are the main topics distributed?”. Accordingly, analyzing the perspective of the main two layers of this study, the most important business-oriented topic mentioned in reviews is the *apartment*, while the most present social-oriented topic is the *advice*. Both H1 and H2 were supported in this study, taking into consideration the fact that social topics regard 52% of reviews, whilst business norms represent approximately 48% of the guest’s reviews. Therefore, social-oriented and business-oriented norms influence the guest’s reviews.

Concerning hypothesis H3, that aims to verify if the “Positive (vs Negative) sentiment reviews have an impact on Review’s Score”, in this study, precisely the sample labelling performed for the predictive model, it is possible to say that reviews regarding positive sentiments tend to have higher ratings. However, this cannot be proven with 100% accuracy, since the actual review score of the extracted reviews was not made available by Airbnb.

Additionally, H1b refers to the gap between guest’s expectations and actual perceptions of the accommodation having a negative impact on reviews. Regarding this hypothesis, the *advice* topic in this study could indicate that the suggestions or interaction with the host did not meet the expectations and negative reviews can come from those actions. However, as advice also regards the guest's opinions and not only the host's advice, this cannot be ensured in this analysis.

To complete the study, a predictive model was developed to predict the overall ratings in customer reviews. The data was labelled with the help of two coders and was then analysed and evaluated following a defined labelling agreement. With the dataset ready, training and testing sets were created, in which both classification and regression models with different feature sets were tested to assess the best predictive model for the Airbnb reviews.

This model rates the reviews in a score from 1 to 5, where 1 means very dissatisfied and 5 means very satisfied. The best predictive model chosen provides an accuracy of 90.70%, an F1 score of 0.907 and addresses the problem as classification process, in this way, ensuring the reviews are correctly classified in the classes defined without bias. The model was achieved using XGBoost Classifier, a decision-tree-based model, in which all the possibilities are taken into consideration.

To conclude, this study provided, not only a thorough analysis of the Airbnb reviews in different cities, understanding the impact of social-oriented and business-oriented opinions alongside the most frequent topics and their associated keywords, combined with all the encompassed sentiments expressed. The predictive model adds value to this study in a way that no bias is verified with the high accuracy obtained and, therefore, a more accurate and genuine interpretation of the accommodation, interaction and overall experience can be obtained with high trustworthiness.

4.1. THEORETICAL AND METHODOLOGICAL CONTRIBUTIONS OF THIS RESEARCH

This research will be determinant to understand the underlying factors of the Airbnb review score rating attribution, not only being essential to comprehend the behaviour of the guests and how their sentiments influence the ratings but also crucial to clarify how the social and market norms impact this matter. The comprehension and evaluation of these factors are critical for hosts, the Airbnb and competitors, such as Hotels. Both hotels and Airbnb hosts should project the customer experience to deliver their hospitality products/services always focusing on the customer perspective (Bharwani & Jauhari, 2013). With this information, hotels can understand which factors they need to study to deliver better service to their customers. For example, given the fact that guests value the local experience, hotels can improve by delivering services that satisfy this consumer need.

The novelty of this contribution lies in the holistic and differentiated approach related to the variables measured, being the first, to the best of the writer's knowledge, performed for these cities with these detailed factors under consideration and combining a predictive model. In addition to prior researches, previous projects have focused on discussing Airbnb accommodations' uniqueness and travel experiences compared to hotels (Lehr, 2015), studying Airbnb's legal and financial matters (Ert et al., 2016), analysing the profile and role of helpful reviewers in online social travel networks (Lee et al., 2011) or investigating factors impacting the choice/refusal to use Airbnb (Stollery & Jun, 2017; Guttentag, 2015; Tussyadiah, 2015).

Additionally, other authors have examined the reviews evaluating experiences of sharing economy-based accommodations (Cheng & Jin,2019; Zhang, 2019; Ding, 2020; Luo, 2018). Many studies, such as Ding (2020), mainly focus on extracting and analysing the review's attributes. Whereas Luo (2018) performed an investigation of the Airbnb lodging aspects, including the prediction of aspect level weights. However, no previous studies were found encompassing directly social and business-oriented norms with Airbnb reviews, aligned with sentiment analysis, topic modelling and prediction of review's ratings with the specific set of cities under study.

4.2. PRACTICAL IMPLICATIONS OF THIS RESEARCH

Considering the practical implications, these findings show evidence that both market and social norms must be taken into consideration by either new or existing hospitality providers. To start providing accommodation or to consolidate the current listings, making sure the guest's expectations are met or exceeded is extremely important. Our findings suggest that a comfortable and clean apartment is one of the most important factors evaluated by the customers. This is one way that hosts can obtain positive sentiments from their guests, which will be reflected in the reviews.

Furthermore, the hospitality providers must bear in mind that failing to deliver the guest's expectations can result in negative reviews. The gap between what customers expect and what they perceive when they arrive at the accommodation has a high impact on guest's opinions, which reflects in the reviews they write about either the accommodation itself or the service provided. Hospitality suppliers must not allow the existence of this gap by any chance. On the other hand, if this gap is verified, the focus must be on trying to mitigate the guest's disappointment as much as possible during their stay. This can be achieved through the host's willingness to help, showing the guest that he/she is a responsive and friendly host, aiming to ensure the guest has the best stay possible. In fact, a host's helpfulness can have a great impact on the guest's final opinion on the stay. In this way, a host that failed to deliver the guest's expectations, but has been available, communicating and helping in every aspect to mitigate the problems caused is more valued than a host that failed to deliver the expected and did not help in any way to improve the guest experience.

In accordance, another way to obtain good reviews is to ensure these customers have the best experience possible. Customers appreciate a "wonderful", "amazing" and "fantastic" booking experience. The experience level can be related to the host, as it is most of the times, since good host interaction does result in reviewer's positive feelings. Hence, when the experience is associated with a "helpful", "friendly", "responsive" and "welcoming" host that has given "perfect" advice and recommendations during the stay, among other factors, this can certainly lead to a satisfied guest. These positive sentiments can result in high review rating, as a consequence of all the effort providers put into the services delivered. These reviews show not only how good the experience was, but also qualify the host in the Airbnb Community, where all existing and new guests search for their accommodation options. In this way, a high review score rating exposes the delivery of service and accommodation that meets or surpasses the customer expectations, which can lead to the recognition of the accommodation and/or host in the community and, consequently, result in more guests or even repeated guests to their listings.

5. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

As a limitation of this study, it can be pointed out the fact that, for the analysis part, only 150k reviews were used. Ding et al. (2020) analysed service quality attributes in 242.020 Airbnb reviews from Malaysia listings. For future analysis, more reviews could be analysed to reach further generalized conclusions. If possible, the closest amount to the total number of reviews retrieved.

Moreover, the algorithm used for the analysis is unsupervised (LDA), which means the result can be different each time the algorithm is re-trained and so will the topics that the algorithm might determine. Future research should use different models, preferably with only supervised algorithms, in order to ensure the research can be generalized. Other models, such as the LARA model, that was developed by Wang et al. (2010), analyses the hidden aspects and their weights in the overall rating. This model regards both supervised and unsupervised approaches, including latent rating regression (LRR) algorithm. This approach was employed by authors such as Luo (2018). However, this still regards unsupervised algorithms, which for further studies should be avoided. On the other hand, Abinaya et al. (2019) presented two methods to detect online review categories, being one unsupervised and the other a supervised learning probabilistic activation method to retrieve the grammatical relationship between reviews. Future studies must analyse several options available to improve the study.

Furthermore, future studies regarding the Airbnb or other Collaborative platforms can include other variables and more factors to study the reviews and retrieve even more detailed results, such as the geographical location and understand how can that influence the review and if there is a pattern or tendencies in reviews from different geographical locations. For instance, Luo (2018) showed that location is one of the five lodging aspects considered by guests and summarizes the listings locations in his study, presenting a map with the listings grouped by cheap and expensive locations. Additionally, Cheng & Jin (2019), also identified the location as one of the key attributes that influence Airbnb user's experiences. In their investigation for the International Journal of Hospitality Management, these authors have presented the geo-location of each listing studied in Sydney.

Besides location, other topics that had a low expression in the current study, such as meal topic, can be further addressed to comprehend and study in detail their impact in the reviews, for example by extracting more reviews with those topics. Additionally, other rating scales and more features can be studied to address the best way to classify the reviews. Different improvements can be performed with this thesis as a basis, the goal will be to always ensure the reliability of the extracted data and accuracy of the results.

6. REFERENCES

Abinaya, V., & Krishnakumari, K. (2019). A Two-Level Aspect Categorization For Sentiment Classification Using Association Rule Mining And Deep Hyper Graphs. SSRG International Journal of Computer Science and Engineering, Special Issue ICMR Mar 2019,41-48. http://www.internationaljournalssrg.org/uploads/specialissuepdf/ICMR-2019/2019/CSE/10.%20KSK1129.pdf

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.

Alshari, E., Azman A., Doraisamy, S., Mustapha, N., & Alkeshr, M. (2017). Improvement of Sentiment Analysis Based on Clustering of Word2Vec Features. In: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA), Lyon. p. 123–126. doi:10.1109/ DEXA.2017.41.

Airbnb. (2020). *About us.* Airbnb News. Retrieved on January 24th, 2020 from https://news.airbnb.com/about-us/

Ariely, D. (2010). Predictably irrational: the hidden forces that shape our decisions. New York: Harper Perennial.

Ariely D., Gneezy U., & Haruvy E. (2017). Social Norms and the Price of Zero. Journal of Consumer Psychology, Issue Marketplace Morality, Volume 28, 181-191

Asuncion, H.U., Asuncion, A.U., & Taylor, R.N. (2010). Software traceability with topic modeling. in Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1. ACM.

Bandura, A., Freeman, W., & Lightsey, R. (1999). Self-efficacy: The exercise of control. Journal of Cognitive Psychotherapy, 13, 158–166.

Bartel, C., & Saavedra, R. (2000). The Collective Construction of Work Group Moods. Administrative Science Quarterly, 45(2), 197-231. doi:10.2307/2667070

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In ACL; 1:238-247.

Barron, K., Kung, E., & Proserpio, D. (2018). The Sharing Economy and Housing Affordability: Evidence from Airbnb. 5-5. 10.1145/3219166.3219180.

Bharwani, S., & Jauhari, V. (2013). An exploratory study of competencies required to cocreate memorable customer experiences in the hospitality industry. International Journal of Contemporary Hospitality Management, 25(6), 823-843

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

Brauckmann S. (2017). "City tourism and the sharing economy – potential effects of online peer-to-peer marketplaces on urban property markets", Journal of Tourism Futures, Vol. 3 Issue: 2, pp.114-126

Bridges, J., & Vásquez, C. (2016). If nearly all Airbnb reviews are positive, does that make them meaningless? Curr. Issues Tour. 1–19.

Bucher, E., Fieseler, C., Fleck, M. & Lutz, C. (2018). Authenticity and the Sharing Economy. 4. 294-313. 10.5465/amd.2016.0161. https://doi.org/10.5465/amd.2016.0161

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

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. Decision Support Systems, 50(2), 511–521. https://doi.org/10.1016/j.dss.2010.11.009

Chen, TH., Thomas, S.W., Nagappan, M., & Hassan, A. (2012). Explaining software defects using topic models. IEEE International Working Conference on Mining Software Repositories. 189-198. 10.1109/MSR.2012.6224280.

Chen, X., Huang, Q., Davison, R., & 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.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3), 477-491.

Cheng, M. (2016). Sharing economy: a review and agenda for future research. Int. J. Hosp. Manag. 57, 60–70.

Cheng M., & Jin X., (2019). What do Airbnb users care about? An analysis of online review comments. International Journal of Hospitality Management, Volume 76, Part A, Pages 58-70

Chung, J., Gülçehre, Ç., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. ArXiv, abs/1412.3555.

Ding, K., Choo, W., Ng, K. & Ng, S. (2020). Employing structural topic modeling to explore perceived service quality attributes in Airbnb accommodation. International Journal of Hospitality Management. 91. 102676. https://doi.org/10.1016/j.ijhm.2020.102676

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

Fenn, J., & LeHong, H. (2012). Hype Cycle for Emerging Technologies 2012, Gartner. Retrieved from http://www.gartner.com/DisplayDocument?doc_cd=233931.

Fernández-Gavilanes, M.F., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F.J. (2016). Unsupervised method for sentiment analysis in online texts" Expert Systems with Applications 58, 57-75.

Festila, M., 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, 54-63

Geerolf, F. (2017). A Theory of Pareto Distributions. UCLA

Griffiths, T.L., Steyvers, M., & Tenenbaum, J.B.T. (2007). Topics in Semantic Representation. *Psychological Review, 114*(2), 211-244.

Guo, Y., Barnes, S.J., & 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.

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

Guzman, E. & Maalej, W. (2014). How do users like this feature? a fine grained sentiment analysis of app reviews. In Requirements Engineering Conference (RE), 2014 IEEE 22nd International, pages 153–162.

Heiphetz, L., Spelke, E. S., Harris, P. L., Banaji, M. R. (2014). What do different beliefs tell us? An examination of factual, opinion-based, and religious beliefs. Cognitive Development, 30, 15–29.

Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are All Smiles Created Equal? How Emotional Contagion and Emotional Labor Affect Service Relationships. Journal of Marketing, 70(3), 58–73. https://doi.org/10.1509/jmkg.70.3.58

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in twitter. In Proceedings of the first workshop on social media analytics (pp. 80-88). ACM.

Hossain, M. (2020). Sharing Economy: A Comprehensive Literature Review. 87. 10.1016/j.ijhm.2020.102470

Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003‑2008: A categorical meta-trend analysis using text mining techniques. Journal of Computing in Higher education, 24(1), 1-17.

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

Jatnika, D., Bijaksana, M., & Ardiyanti, A. (2019). Word2Vec Model Analysis for Semantic Similarities in English Words. Procedia Computer Science. 157. 160-167. 10.1016/j.procs.2019.08.153.

Jivani, A. (2011). A Comparative Study of Stemming Algorithms. Int. J. Comp. Tech. Appl. 2. 1930-1938.

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

Karani, D. (2018). Introduction to Word Embedding and Word2Vec. Retrieved from https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa

Kaur, J., & Buttar, P. (2018). STOPWORDS REMOVAL AND ITS ALGORITHMS BASED ON DIFFERENT METHODS. International Journal of Advanced Research in Computer Science, 9(5), 81-88. https://doi.org/10.26483/ijarcs.v9i5.6301

Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. Journal of King Saud University – Computer and Information Sciences, 26(3), 258-275. https://doi.org/10.1016/j.jksuci.2014.03.009

Khatua A., Khatua A., & Cambria E. (2019). A tale of two epidemics: contextual Word2Vec for classifying twitter streams during outbreaks. Inf Process Manag. 56(1): 247–257.

Krupka, E. L. & Weber, R., (2013), Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?, Journal of the European Economic Association, 11, issue 3, p. 495-524 https://doi.org/10.1111/jeea.12006

Lampinen, A., Cheshire, C., (2016). Hosting via Airbnb: motivations and financial assurances in monetized network hospitality. Paper Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San José, California DOI:10.1145/2858036.2858092

Lee, H. L., Law, R., & Murphy, J. (2011): Helpful Reviewers in TripAdvisor, an Online Travel Community, Journal of Travel & Tourism Marketing, 28:7, 675-688 https://doi.org/10.1080/10548408.2011.611739

Lee, S., & 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. 188 (Retrieved from https://scholar.dominican.edu/masters-theses/188)

Liang, L. J. (2015). Understanding repurchase intention of Airbnb consumers: perceived authenticity, EWoM and price sensitivity. Master Thesis, University of Guelph (Retrieved from http://hdl.handle.net/10214/8813)

Liang L. J., Choi H. C. & Joppe M. (2018) Understanding repurchase intention of Airbnb consumers: perceived authenticity, electronic word-of-mouth, and price sensitivity, Journal of Travel & Tourism Marketing, 35:1, 73-89, DOI: 10.1080/10548408.2016.1224750

Lin, M., Miao L., Wei W., & Moon H. (2019), “Peer Engagement Behaviors: Conceptualization and Research Directions,” Journal of Service Research, 22(July) 388–403.

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

Lu, C., & Kandampully, J. (2016). What drives customers to use access-based sharing options in the hospitality industry? Research in Hospitality Management, 6(2), 119-126.

Luo, Y. (2018). “What Airbnb Reviews Can Tell Us? An Advanced Latent Aspect Rating Analysis Approach.” *Graduate Theses and Dissertations*. 16403.

Mao, Z., & Lyu, J. (2017). Why travelers use Airbnb again? International Journal of Contemporary Hospitality Management, 29(9), 2464-2482.

Mattila, A., Hanks, L., & Wang, C. (2014). Others service experiences: emotions, perceived justice, and behavior. European Journal of Marketing. 48. 10.1108/EJM-04-2012-0201.

McAuley, J., & 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.

McCallum, A. K. (2002) "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. Available at arXiv preprint arXiv:1301.3781.

Mikolov, T., Yih, W. T., & Zweig, G. (2013b). Linguistic regularities in continuous space word representations. *Proceedings of HLT-NAACl* (pp. 746–751).

Minarro-Giménez, J. A., Marin-Alonso, O., & Samwald, M. (2014). Exploring the application of deep learning techniques on medical text corpora. *Studies in Health Technology and Informatics, 205*, 584–588.

Mohan, V. (2015). Preprocessing Techniques for Text Mining - An Overview.

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

Naili, M., Chaibi, A. H., & Ben Ghezala, H. H. (2017). Comparative study of word embedding methods in topic segmentation. In Procedia Computer Science (Vol. 112, pp. 340–349). Elsevier B.V. https://doi.org/10.1016/j.procs.2017.08.009

Nasukawa, T., & Yi, J. (2003). "Sentiment Analysis: Capturing Favorability Using Natural Language Processing." In Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003

Ning, X., Yao, L., Wang, X., Benatallah, B., Dong, M., & Zhang, S. (2020) Rating prediction via generative convolutional neural networks based regression, Pattern Recognition Letters, Volume 132, Pages 12-20, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2018.07.028.

O'Regan M., & Choe J., (2017) Airbnb and cultural capitalism: enclosure and control within the sharing economy. Anatolia. 28. 1-10. 10.1080/13032917.2017.1283634.

Schouten, K., & Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering, 28(3), 813-830.

Ostrom, E. (2014). Collective action and the evolution of social norms. Journal of Natural Resources Policy Research, 6, 235–252.

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., & 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.

Park, Y. A., & Gretzel, U. (2007). Success factors for destination marketing web sites: A qualitative meta-analysis. Journal of travel research, 46(1), 46-63.

Pera, R., Viglia, G., Grazzini, L., & Dalli, D. (2019). When empathy prevents negative reviewing behavior. Annals of Tourism Research, 75, 265-278. https://doi.org/10.1016/j.annals.2019.01.005

Perren, R., & Kozinets, R. V. (2018). “Lateral Exchange Markets: How Social Platforms Operate in A Networked Economy”, Journal of Marketing, 82 (January), 20–36.

Petty, R. E., & Brinol, P. (2010). Attitude change. In Baumeister, R. F., Finkel, E. J. (Eds.), Advanced social psychology: The state of the science (pp. 217–259). New York, NY: Oxford University Press

Plank, A. (2016). The hidden risk in user-generated content: An investigation of ski tourers' revealed risk-taking behaviour on an online outdoor sports platform. Tourism Management, 55, 289-296.

Qu, L., Ifrim, G., & Weikum, G. (2010). The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns. COLING.

Quattrone, G., Proserpio, D., Quercia, D., Capra, L., & Musolesi, M. (2016). Who Benefits from the "Sharing" Economy of Airbnb?. Proceedings of the 25th International Conference on World Wide Web.

Liao, S., Chu, P., & Hsiao, P. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. Expert Syst. Appl., 39, 11303-11311.

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

Schuckert, M., Liu, X., & Law, R. (2016). Insights into suspicious online ratings: direct evidence from TripAdvisor. Asia Pacific Journal of Tourism Research, 21(3), 259- 272.

Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-237426

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

Sievert, C. & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. 10.13140/2.1.1394.3043.

Singh, S. K., Paul, S. & Kumar, D. (2014). Sentiment Analysis Approaches on Different Data Set Domain: Survey. International Journal Of Database Theory and Application, 7(5), 39-50.

So, K. K. F. & Oh, H. & Min, S. (2018). Motivations and constraints of Airbnb consumers: Findings from a mixed-methods approach. Tourism Management. 67. 10.1016/j.tourman.2018.01.009.

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.

Statistic Brain (2017). Internet Travel & Hotel Booking Statistics. Retrieved from https://www.statisticbrain.com/internet-travel-hotel-booking-statistics/

Stollery, A., & Jun, S. H., (2017) "The antecedents of perceived value in the Airbnb context", Asia Pacific Journal of Innovation and Entrepreneurship, Vol. 11 Issue: 3, pp.391-404

Suess, C., Woosnam, K. M., & Erul, E. (2020). Stranger-danger? Understanding the moderating effects of children in the household on non-hosting residents' emotional solidarity with Airbnb visitors, feeling safe, and support for Airbnb. Tourism Management. 77. 103952. https://doi.org/10.1016/j.tourman.2019.103952

Syed, A. (2011). A Review of Cross Validation and Adaptive Model Selection.

Tang, D., Qin, B., Liu, T., & Yang, Y. (2015). User Modeling with Neural Network for Review Rating Prediction. IJCAI.

Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. Computer Science & Information Technology. 6. 201-210. 10.5121/csit.2016.60616.

Tussyadiah, I. P. (2015). An exploratory study on drivers and deterrents of collaborative consumption in travel Information and communication technologies in tourism 2015 (pp. 817-830): Springer.

Tussyadiah, I. P. (2016). Factors of satisfaction and intention to use peer-to-peer accommodation. International Journal of Hospitality Management. 55. 70-80. 10.1016/j.ijhm.2016.03.005.

Tussyadiah, I. P., & Pesonen, J. (2016). Impacts of peer-to-peer accommodation use on travel patterns, Journal of Travel Research, 55 (8) pp. 1022-1040

Tussyadiah, I. P., & Pesonen, J. (2016a). Drivers and barriers of peer-to-peer accommodation stay–an exploratory study with American and Finnish travellers. Current Issues in Tourism, 1-18

Tussyadiah, L., Zach, F. (2016). Identifying salient attributes of peer-to-peer accommodation experience. J. Travel Tour. Mark. 1–17.

Wang, H., Lu, Y., & Zhai, Chengxiang. (2010). Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 783-792. 10.1145/1835804.1835903.

Wang, X., Jiang, W., & Luo, Z. (2016). Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers, pages 2428–2437

Williamson, S., Wang, C., Heller, K., & Blei, D. (2010). The IBP compound Dirichlet process and its application to focused topic modeling.

Wu, J., Cai, J., Luo, X., & Benitez, J. (2021). How to increase customer repeated bookings in the short-term room rental market? A large-scale granular data investigation. Decision Support Systems. 113495. https://doi.org/10.1016/j.dss.2021.113495

Yang, Y. & Mao, Z. (2020). Location advantages of lodging properties: A comparison between hotels and Airbnb units in an urban environment. Annals of Tourism Research. 81. 102861. https://doi.org/10.1016/j.annals.2020.102861

Yannopoulou, N., (2013). User-generated brands and social media: couchsurfing and airbnb. Contemp. Manag. Res. 9 (1), 85–90

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert systems with applications, 36(3), 6527-6535.

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? Information Technology & Tourism, 10(4), 283-295.

Zervas, G., Proserpio, D., Byers, J.W., (2017). The rise of the sharing economy: estimating the impact of Airbnb on the hotel industry. J. Mark. Res. LIV 687–705.

Zhang, D., Xu, H., Su, Z., & Xu, Y. (2015). Chinese comments sentiment classification based on word2vec and SVM perf. *Expert Systems with Applications*, *42*(4), 1857–1863. doi:10.1016/j.eswa.2014.09.011

Zhang, F. (2019). A hybrid structured deep neural network with Word2Vec for construction accident causes classification. International Journal of Construction Management. 1-21. 10.1080/15623599.2019.1683692.

Zhang J., (2019) "What's yours is mine: exploring customer voice on Airbnb using text-mining approaches", Journal of Consumer Marketing

Zhang, L., Yan, Q. & Zhang, L. (2020). A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb. Decision Support Systems. 133. 113288. https://doi.org/10.1016/j.dss.2020.113288

Zhou, L., Ye, S., Pearce, P. L., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. International Journal of Hospitality Management, 38(2014), 1-10.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of marketing, 74(2), 133-148.

7. ANNEXES

Topic	Top Word	Top Word Count
Apartment	Place	58511
	Apartment	56481
	Clean	41243
	Nice	35974
	Great	28548
	Stay	22117
	Comfortable	20414
	Location	18633
	Really	12510
	Recommend	11218
Advice	Recommend	10924
	Highly	7100
	Location	6345
	Thank	6313
	Thanks	5851
	Arrival	5663
	Perfect	5367
	Good	5116
	Definitely	4588
	Quick	4171
Experience	Great	41937
	Location	14769
	Host	9614
	Stay	8640
	Lovely	7801
	Amazing	7605
	Wonderful	6812
	Fantastic	4564
	Hosts	3110
	Experience	2635
Interiors	Room	10429
	Kitchen	5314
	Bathroom	4940
	Bed	4416
	Small	3476
	Big	2618
	Shower	2490
	Living	2404
	Bedroom	2371
	Private	1891
Good_Host	Host	14350
	Helpful	11195
	Friendly	8747

	Kind	4474
	Accommodating	3152
	Responsive	2979
	Great	2393
	Welcoming	2139
	Stay	2058
	Really	1990
Neighbourhood	Quiet	5301
	Area	4667
	Neighbourhood	3582
	Distance	3002
	Walking	2771
	Short	2136
	Walk	1946
	Location	1829
	Restaurants	1679
	Minute	1431
Booking	Easy	12219
	Time	12088
	Check	8013
	Great	4926
	Access	3136
	Stay	3016
	Location	2235
	Left	2104
	Return	1829
	Definitely	1817
Shops	Nearby	1828
	Local	1410
	Food	1220
	Restaurants	1153
	Corner	874
	Eat	786
	Grocery	742
	Places	724
	Shopping	701
	Good	676
Trip	Stay	25922
	Home	8517
	House	5505
	Definitely	5357
	Feel	4126
	Experience	4038
	Recommend	3302
	Enjoyed	3143
	Make	2919
	Really	2819
Transport	Close	11353
	Walk	8526
	Station	8409

	Location	6383
	Train	4193
	Central	3942
	Walking	3555
	Minutes	3252
	City	3235
	Convenient	3023
Meal	Coffee	2221
	Breakfast	1790
	Tea	790
	Delicious	614
	Morning	551
	Bread	449
	Milk	419
	Provided	388
	Fruit	373
	Fresh	354
Noise	Night	3416
	Bit	2227
	People	1765
	Noisy	1142
	Little	713
	Loud	648
	Street	622
	Stay	622
	Noise	542
	Location	507
Communication	Communication	3111
	Great	1340
	Good	756
	Host	719
	Flexible	683
	Excellent	548
	Check	546
	Contact	541
	Stay	405
	Process	318
Surrounding	Lot	1686
	Near	1541
	Restaurants	649
	Thanks	628
	Location	508
	Plenty	358
	Good	345
	Interesting	344
	City	322
	Close	263
	Public	792
	Transport	382
	Transportation	252

Location	City	114
	Location	97
	Good	87
	Restaurants	67
	Near	66
	Transit	60
	Center	56

Table 7 - Topic's top words ordered by frequency

Topic	Percentage of Occurrence
Advice	39.82
Apartment	14.94
Trip	10.75
Interiors	7.59
Transport	7.15
Booking	6.24
Neighbourhood	3.75
Noise	2.90
Good Host	2.46
Shops	1.81
Meal	0.92
Communication	0.64
Surrounding	0.59
Location	0.24
Experience	0.20

Table 8 - Negative Topic's Mean

Topic	Percentage of Occurrence
Apartment	29.63
Advice	23.12
Transport	7.68
Good Host	7.19
Booking	6.91
Interiors	6.36

Trip	6.19
Neighbourhood	3.73
Experience	2.68
Noise	2.09
Surrounding	1.37
Shops	1.33
Communication	0.74
Meal	0.70
Location	0.26

Table 9 - Neutral Topic's Mean

Topic	Percentage of Occurrence
Apartment	39.06
Experience	17.70
Advice	12.98
Good Host	6.46
Trip	6.35
Booking	5.94
Transport	3.04
Interiors	2.79
Neighbourhood	1.69
Communication	1.10
Shops	0.75
Surrounding	0.72
Meal	0.72
Noise	0.65
Location	0.05

Table 10 - Positive Topic's Mean