A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

THE IMPACT OF COVID-19 ON AIRBNB: A NLP-BASED APPROACH OF ANALYZING SHIFTING RISK PERCEPTIONS AMONG TOURISTS DURING THE PANDEMIC

LORIAN ABAZI

Work project carried out under the supervision of:

Professor Qiwei Han

21-05-2021

Abstract

The COVID-19 pandemic has tremendously impacted our society, with massive economic shifts, including within the tourism industry. The aim of this research is to answer the question of how risk awareness has developed among tourists during the COVID-19 pandemic. Using Airbnb reviews, this research examines whether perception of cleanliness has changed since the outbreak and how this has affected indicators such as occupancy, price and monthly income. Based on Latent Dirichlet Allocation topic modelling algorithm, this paper shows that there has been a change in risk perception, which also has a positive impact on the performance indicators of cleanly-perceived properties.

Keywords: COVID-19 Impact, Hospitality Management, Tourism Industry, Sharing Economy, Machine Learning, Natural Language Processing, Sentiment Analysis

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22.

Table of Content

1. Introduction

1.1. Background and Problem Statement

"The impacts of COVID-19 on tourism are unprecedented. With borders closed, hotels shut down and planes on the ground, tourism has come to a total standstill in the last two months," This quote by Zoritsa Urosevic (UNWTO 2020), Director of Institutional Relations and Partnerships at UNWTO, describes the precarious situation which the tourism industry has been exposed to following the outbreak of the coronavirus. The outbreak of the COVID-19 pandemic has affected countries, societies and individuals across the globe and has resulted in a public health crisis - hospitals working at full capacity, cancelled operations, and perform triage (Gallina 2020). This dramatic development has severe consequences for the international and domestic tourism industry which has experienced a significant downturn. The UNWTO (2020) estimated that the pandemic will result in a loss of approximately $910 billion and a decline in international travel of 60-80% after the WHO declared COVID-19 a global pandemic.

Airbnb, as a sharing platform for housing particularly affected by the pandemic, reported a decline in revenues by 72%, leading to 1,800 job cuts as a result of the COVID-19 outbreak (Airbnb, 2020a). However, in the second half of the year 2020, Airbnb seemed to have recovered with returned profits and a successful IPO, where the company generated a market cap past the market cap of comparable industry giants such as Booking.com or Expedia (Abril, 2020). Despite this trend, many Airbnb hosts are still in a precarious situation. In the United States, Airbnb hosts have lost $4,036 since the beginning of the pandemic and have not seen booking increase since the lifting of certain COVID-19 measures (Lane 2020). The Forbes Magazine (2020) moreover revealed that 47% of Airbnb hosts did not feel safe renting out their properties, while over 70% of guests were afraid to stay in an Airbnb. This leads to the question of whether the risk perception of tourists have changed during the pandemic and how this impacts Airbnb's business model.

1.2. Objectives and Scientific Approach

Numerous studies have highlighted the importance of risk perception of tourists during and post health crises (Sharpley and Craven 2001; Baxter and Bowen, 2004; Rassy and Smith 2013). Currently, little is known about how COVID-19, which exceeds all previous tourism crises, has influenced the risk perception of travelers. Therefore, the central objective of this thesis is to assess the impact of COVID-19 on Airbnb and analyze whether the perception of risk has been changed due to the pandemic. The main variable that is considered in this regard is *perceived cleanliness.* At the beginning of the pandemic, it was unclear how exactly COVID-19 was transmitted. As the pandemic progressed, it became increasingly evident that the virus could be contracted by touching contaminated surfaces and breathing contaminated air. It can be thus hypothesized that guests made their choices of which listing to book based on the perceived cleanliness of other guests. Following this premise, this research paper focuses solely on properties in the city of London between January 2019 and January 2021. London is a suitable candidate for testing as the reviews here are largely written in English. Multilingual reviews would only lead to unnecessary complications in the topic modeling algorithm. Furthermore, January 2019 is an appropriate start date due to the consistency of the data and the fact that seasonal effects might be diminished.

This research relies on a topic modeling approach conducted with the Latent Dirichlet Allocation (LDA) algorithm, to assess how the perception of cleanliness has evolved over the course of the sample period. This model, combined with sentiment analysis, is used to implement a parameter that distinguishes between properties perceived as clean and otherwise. This distinguishing parameter will help to compare the development of price, monthly income and occupancy rate of properties in London with the goal to create a framework that can be used throughout other regions. The results will ultimately be used to drive recommendations for how Airbnb hosts can react towards COVID-19 to still be able to attract tourists as the pandemic continues.

2. Literature Review

In order to bring the research question into context, a few theoretical frameworks will be discussed. Since the research is designed to study the effects of COVID-19 on Airbnb, it is crucial to understand current trends in the sharing economy, of which Airbnb is one of the biggest players. Analyzing current trends in the sharing economy will also help to understand how resistant this industry is against disruptive events like a pandemic. Building upon this research, it will be explained how a pandemic can influence the hospitality and tourism industry in general. Airbnb is a player in the sharing economy, offering hospitality services with a client base of mostly, and therefore very suitable to study these effects. Finally, the topic of risk perceptions and the role this plays on the tourism industry will be addressed in order to lay the foundation of the generated hypotheses.

2.1. The Sharing Economy

Given the speed of which COVID-19 is spreading and considering the strict measures imposed on the society, the sharing economy is an ideal industry to study the impact of the pandemic. The sharing economy is defined as "people coordinating the acquisition and distribution of a resource for a fee or other compensation" (Belk, 2014, p. 1597). Many economic, social, and technological factors are driving the sharing economy in order to propel the industry into a successful phenomenon and stable economic sector (Mody et al., 2019). The sharing economy focuses on underutilized resources, and therefore, drives sustainability (Cheng et al., 2020; Geissinger et al., 2019). The sharing economy has had a positive impact on various industries, such as transportation (e.g., Uber), lodging (e.g., Airbnb) and household services (e.g., Care.com) and is thereby promoting a shift towards more sustainable production and consumption. To be more specific, it is increasingly linked with the discourse surrounding the circular economy, highlighted by the advantages of a transition from an ownership-based economy to an access-based economy (Botsman and Rogers, 2010; Curtis and Mont, 2020).

2.1.1. Short-term Impacts of COVID-19 on Sharing Platforms

COVID-19 has disrupted the sharing economy significantly. However, at this moment there is only a limited amount of research papers that discusses the impact of COVID-19 on this industrial sector. Most of the literature regarding the sharing economy and the COVID-19 pandemic thereby focuses on Uber and Airbnb, which correlates with the existing literature about the sharing economy generally (Muñoz and Cohen, 2018; Ritter and Schanz, 2018). Following this, it is necessary to distinguish the pandemic's short-term impacts of the pandemic between the macro- (e.g., societies, economies, governments), meso- (e.g., businesses, organizations, communities), and micro-levels (e.g., employees, individuals, consumers) (Baum et al. 2020). For example, COVID-19 measures from a macro perspective, like physical distancing, lockdowns, closed borders or increased hygiene standards have a significant impact on sharing platforms and its operations (Curtis and Mont 2020). Since these business models usually operate in a two-sided market, which requires a personal exchange in goods and services, restrictions on distances have severe consequences. With individuals being precautious in terms of avoiding contracting and spreading the virus, the demand for engaging with sharing platforms consequently seems to decrease in the same way (Mont et al. 2021).

However, a major advantage of sharing platforms is their ability to exploit information and communication technology (ICT) in order to potentially reduce the drawbacks experienced by other types of industries that are less digitalized. Exploiting the opportunities of smartphone applications, digital keys and virtual communication does not only reduce the need to meet in person but eventually keeps a certain quality of service. The digital nature of sharing platforms also increases the ability to adapt more quickly to new conditions and technologies, due to technical flexibility and higher agility, which results in low transition costs (Kamal 2020). Accordingly, the trend is that demand of sharing platforms is going to increase, especially those that manage the transition to contactless exchange of goods (Horgan et al. 2020).

2.1.2. Airbnb as a Major Participant in the Sharing Economy

Looking at the different types of sharing platforms, it becomes clear that they have been impacted in different ways and to a different extent. The type of sharing platform that has probably been most affected are hospitality platforms like Airbnb (Lee and Deale 2021; Boros et al. 2020; Hossain 2021). Airbnb's founding idea is based on a concept of living together with strangers for low costs and high host involvement (Zervas et al. 2017). Although the company developed a very successful business model, a significant number of consumers have also conveyed complaints, constraints and risk associated with staying in such accommodations even before the pandemic (Liang 2018; Phua 2019). According to this, perceived risk and lack of trust have been explored as obstacles for Airbnb (Jun 2020; Lee 2020). Previous studies thereby examined that distrust generates a negative attitude towards Airbnb properties and that Airbnb guests develop their risk perception mainly based on previous experiences and experiences of others. (So et al. 2018; Liang 2018).

During a pandemic, these obstacles become particularly observable through several key performance indicators. For instance, Airbnb reservations in many countries went down by a reported 90% linked directly to travel restrictions and lockdown measures. Consequently, COVID-19 perfectly illustrates the fragility of Airbnb towards external events and travel behavior of tourists that are closely related to risk perceptions. As a result, Airbnb's market valuation dropped from USD 31 billion in early 2017 to USD 18 billion in April 2020. Compared to 2019, revenues dropped more than 50% (Airbnb 2020a). While many experts expected a long-term decline in demand for Airbnb, in the US, Airbnb experienced more bookings in the summer of 2020 than during 2019. Mohamed (Mohamed 2020) examined that the reason for this might be that tourists perceive Airbnb properties as less risky for contracting the virus compared to hotels. Despite these first indications, yet only little is known about how the risk perception of tourists has changed due to the pandemic (Lee and Deale 2021).

2.2. Impact of Diseases on Tourism Industry

The example of Airbnb shows how fragile the tourism sector is and exposes its influenceability when faced with natural or human-made disasters or diseases outbreaks (Çakar 2018; Reddy et al. 2020). In the past 20 years, several major events have significantly disrupted the tourism industry such as terrorist attacks (e.g., New York City 2001, Bali 2002), the global financial crisis of 2008, the eruption of the volcano Eyjafjallajökull in 2010 or the 2004 Tsunami in Southeast Asia (Hall 2010; Lim and Won 2020). The most significant disease outbreaks affecting the tourism industry were the bovine spongiform encephalopathy ("mad cow disease") in 2002-2003, the severe acute respiratory syndrome (SARS-CoV) in 2003, the avian flu in 2004, the swine flu (H1N1) in 2009, Middle East Respiratory syndrome-related coronavirus in 2012 (MERS-CoV) and the Ebola outbreak in 2014 (Boros et al. 2020). All these events have emphasized the relevance of post-crisis management in the tourism industry and caused a series of academic research studying the effects of such outbreaks while highlighting the importance of precaution and preventive mechanisms (Sharpley and Craven 2001; Baxter and Bowen, 2004; Rassy and Smith 2013.; McAleer et al. 2010). However, as we see in the current COVID-19 crisis, governments worldwide have failed to introduce effective measures for disease-related tourism management and communication (Kuo et al. 2009; Lee et al. 2012; Joo et al. 2019).

Across the research papers, common ground can be found through the fact that declines in touristic activities are primarily caused by concerns of tourist's health and security as well as to non-pharmaceutical interventions like lockdowns, surveillance, border control and quarantine (Ryu et al. 2020; Ho et al. 2017; Lee et al. 2012). These observations are coherent with the concept of perceived risk theory. According to this framework, the consumer purchasing process is negatively influenced by unknown and uncertain factors which are evaluated by personal and subjective evaluations (Bauer 1980). Following this, risk perception can be explained as the subjective evaluation of risk of a threatening situation based on its features and severity

(Sjöberg et al. 2004; Moreira 2008). In this regard, it is worth mentioning that touristic activities are not only affected by diseases but also by the speed of its spread (Omrani and Shalhoub 2015; Findlater and Bogoch 2018). Due to an increasing mobility, accessibility and affordability of air travel, infectious diseases can spread more easily and rapidly across the globe.

According to this, risk perception of tourists plays an important role in tourist behavior and travel decisions, albeit depending on the tourist's experience and knowledge in such critical situations. Studies show that touristic activities decline significantly when tourists are concerned about their health and safety (Leggat et al. 2010; Pine and McKercher 2004; Yanni et al. 2010; Schroeder et al. 2013). Following this, negative effects in traveling appear quickly with long recovery times (Lean and Smyth 2009). Accordingly, the time for recovery seems to be connected to different cultural backgrounds as studies of the SARS outbreak in Taiwan reveal. Lim and Won (2020) examined that the risk is perceived differently and is affected by experiences regarding the virus and the trust in health organizations. In this regard, the recovery from tourist flows of Japan took much longer compared to the United States or Hong Kong.

2.3. Communication as a Key to Attract Customers in Times of Uncertainty

Many research papers connect travel decisions with the perception of risk where communication plays a crucial part in influencing tourists by driving fear, uncertainty and concerns (Baxter and Bowen 2004; Sparke and Anguelov 2012; Maphanga and Henema 2019; Jamal and Budke 2020). In general, infectious outbreaks always negatively affect consumer behavior – even if the disease is not harmful, such as in the case of the avian flu (Kim et al. 2020). Following this, it can be observed that the media can be responsible for setting an atmosphere of fear and uncertainty among tourists (Monterrubio 2010). Another aspect that plays an important role in this regard is the concept of word-of-mouth which refers to non-commercial flow of information regarding brands, products, companies and others (Harrison-Walker 2001). The reason why word-of-mouth is so influential is because humans usually rely more on what family, friends or reliable people say rather than commercial advertising. Therefore, word-of-mouth is often referred to as a consumer-driven communication tool which becomes even more significant in our modern society that is driven by social media and digital technology (Hennig-Thurau et al. 2002; Fong and Burton 2006). This puts an emphasize on communication as a mechanism for recovering in times of crisis. Scholars point out that that hotel facilities and destinations must convey a message of safety and appropriate health conditions in order to attract tourists which makes word of mouth as an influential communication method of particular interest in this regard (Yu et al. 2020).

3. Hypotheses

In order to investigate the impact of COVID-19 on Airbnb and analyze whether the perception of risk has shifted due to the pandemic, the following two central hypotheses are proposed which are extensively examined in the empirical part. First, the hypotheses relate to the general question of whether the perception of cleanliness as a risk factor has changed with the onset of COVID-19. In the next step, the hypotheses become more specific and determine how this potentially observed shift has affected the business. For this, the development occupancy rates, price per night and monthly generated income will be assessed over the course of the pandemic.

Hypothesis H1: Due to the outbreak of the coronavirus, the perception of cleanliness has changed significantly. It is expected that the awareness of hygiene as a risk factor has increased during the pandemic. This means that from March 2020, it is more likely that in an Airbnb review, the most relevant topic is cleanliness.

In this hypothesis, it is assumed that due to media coverage, general public fear and severity of the disease, the factor of hygiene moves into the focus of attention when people make reviews about Airbnb properties. Prior to the pandemic, other factors might have been more relevant than cleanliness, such as location and how well the property is connected to the public transportation system. The hypothesis is therefore confirmed if, on the one hand, the topic corresponding to hygiene has a higher weight in the reviews while other topics become increasingly irrelevant over the course of the pandemic.

Hypothesis H2: COVID-19 causes a significant difference in monthly occupancy, price per night and monthly generated income between Airbnb listings perceived as clean and properties not perceived as clean. Here, it is assumed that in times of uncertainty, travelers base the purchasing decision based on experiences of others according to the word-of-mouth framework. Therefore, tourists book units, where the perceived risk is as low as possible, at a higher rate. This would correspond with clean-perceived apartments which is why this hypothesis is confirmed if Airbnb listings classified as clean charge higher prices, are booked more frequently and consequently generate more income as compared to not perceived clean apartments.

4. Methodology

The empirical analysis and testing of the formulated hypothesis was carried out in a two-step approach. In the first step, reviews of Airbnb properties in London, United Kingdom were processed with the Latent Dirichlet Allocation (LDA) algorithm in the sample period of the 1st of January 2019 until the 31st of January 2021. This algorithm, an unsupervised machine learning model that is used for topic modeling, examines what Airbnb guests talk about in their reviews. The goal is to find a topic corresponding to hygiene before its relevance and dominance in the reviews will be analyzed over the course of the sample period.

Following this, each listing where cleanliness is the dominant topic will be labeled accordingly. At the end, a parameter will be induced to the dataset that will distinguish between properties where guests talk about cleanliness or not. Then, sentiment analysis is applied on each sentence per review that deals with cleanliness. For this, with the help of the Python library SpaCy, a library of synonyms related to cleanliness and dirtiness is created, which will be used to extract all the relevant sentences. If a property where cleanliness is a dominant topic contains a clean/dirty reference, the sentence will be extracted, and sentiment analysis will be applied on it. If the sentiment score is positive, the listing will be labeled as *perceived clean* and *not perceived clean* if otherwise.

4.1. Origin of the Data

This research uses Airbnb data which was provided by Inside Airbnb for bookings occurring between the 1st of January 2020 and 31st of January 2021 in London, United Kingdom. The selected time period and location allows deep insights into how COVID-19 has affected the short-term rental market in a metropole that is typically a main tourist and business traveler hub. According to Inside Airbnb, the data was verified, cleansed and aggregated and comes from public information on the Airbnb website (Inside Airbnb, 2021). For the purpose of this research, two files are extracted each month: listings and reviews and. The Airbnb listing data includes key property characteristics with timestamps that allow for the analysis of which factors are most affected by the COVID-19 pandemic. The variables include property location, property type, room type, Superhost status, allowed number of guests and Airbnb scores. The final dataset counts for 318,698 observations. The reviews' dataset includes a unique timestamp for each review. Furthermore, the dataset records a unique reviewer ID, the listing ID and the exact comment. The comments will later be used to determine whether a listing is *perceived clean* or not and for the topic modelling in order to determine the development of review topics around the term *clean*. Before presenting the results, the next section will explain how LDA topic modeling works in order to have a better understanding of the outcome of the analysis.

4.2. Topic Modelling with Latent Dirichlet Allocation

Topic modelling has been widely used in many different fields including scientific topic extraction, cryptocurrency, operation and risk extraction, communication research, marketing, investor attention modeling, computer vision and even bioinformatics (Blei and Lafferty 2007; Linton et al. 2017; Huang et al. 2017; Maier et al. 2018; Reisenbichler and Reutterer 2019; Liu et al., 2018). Generally speaking, topic modelling is a form of unsupervised machine learning that enables processing a large collection of data, but preserves statistical relationships in order to perform classification or summarization tasks (Bhat et al., 2019). Topic models are used for discovering latent variables which affect the composition of a document. These latent variables represent abstract topics. The main challenges in topic modeling are (1) handling the predominance of the most frequently appearing words in the estimated topics and (2) interpreting topics that are overlapping with common words. Thereby, topic modeling makes use of the word co-occurrence information to predict topics. The reason why co-occurrence is an appropriate method is due the way the human language is structured. Certain words appear more frequently than others which leads to a co-occurrence of these words where more words result in a predominance in estimated topics (Lei and Ying 2021).

4.2.1. The Foundation of Topic Modelling

Topic modelling with the LDA model is a relatively young method, dating its origin back to the early 1980s. Founders Salton and McGill (1983) thereby set the groundwork for modern topic modelling by developing a method to analyze text data. This method is called frequency-inverse document frequency (tf-idf) and calculates the statistical importance of a word within a document. Tf-idf vectorization thereby compares the number of occurrences of a word within a document with the number of occurrences of a word within the corpus, expressed by the formula below (Salton and McGill 1983).

$$w_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right)$$

However, this approach does not cover any statistical relationship between the terms in a text document which is why Deerwester et al. (1990) developed a different model (Latent Semantic Indexing (LSI)) that conducts a so-called low-rank approximation of the term-document matrix A. In this matrix, the tf-idf score is calculated for each entry according to the formula above in order to assign a weight to each term. Following this, a term has a high weight if it occurs frequently in a document but infrequently across the corpus. A is usually is a very noisy matrix due to its many dimensions, but through introducing the hyperparameter t number of topics, the dimensionalities are reduced with the so-called truncated singular value decomposition (SVD), which factorizes a matrix into the product of three independent matrices $M = U \times S \times V$ (Deerwester et al. 1990). This results in the following formula for the document-term matrix in LSI.

$$A \approx U_i \times S_t \times V_t$$

The problem Deerwester et al. (1990) could not solve was that the generated topics were very difficult to interpret and required a huge amount of input data. Hoffman (1999) therefore replaced the truncated SVD with a probabilistic method to reduce dimensions in the document-term matrix and called it the probabilistic Latent Semantic Indexing (pLSI). According to the following formula, pLSI adds probabilistic component to this approach so that one can say that topic z is present in document d with a probability of $P(z \mid d)$. Similarly, word w belongs to topic z with a probability of $P(w \mid z)$, as illustrated in appendix A1. The right side of the formula thereby describes the likelihood of seeing some document and then, based on the topic distribution in that document, the likelihood of finding a specific word within the sample document.

$$P(D,W) = P(D) \sum_{Z} P(Z \mid D) P(W \mid Z)$$

4.2.2. Latent Dirichlet Allocation (LDA)

This research performs topic modelling with the LDA algorithm. LDA thereby presents a Bayesian version of pLSI which uses Dirichlet priors α and β for the document-topic and word-topic distribution. It is an unsupervised machine learning algorithm developed by Blei et al. (2003) using a generative probabilistic topic modeling approach for collecting discrete data such as text corpora, genome sequences or collection of images (Lei and Ying 2021). In text collections, LDA explores all possible representations of documents as random mixtures over latent topics, where each topic is determined as a Dirichlet distribution over words. A synonym for Dirichlet can be distributions over distributions which allows for better generalization (Naushan 2020). As a result, Dirichlet distributions therefore gives an actual probability distribution likely to be observed given a certain distribution (Blei et al., 2003). Compared to pLSI, LDA works differently as illustrated in appendix A2.

First, a random sample representing the topic distribution is retrieved from a Dirichlet distribution Dir (α) which refers to $\theta_{(d,z)}$ and describes the per-document topic distribution. This probability is calculated over *d* documents of the corpus with the following formula.

$$P(Z_t = t \mid d) = \theta_{(d,z)}$$

From here, a topic *Z* is selected before a random sample, which stands for the word distribution, in topic Z is selected from a Dirichlet distribution Dir(β). This word distribution refers to ϕ_t from where a word *w* is chosen. The corresponding probability is approximated over vocabulary *V* of the corpus such that

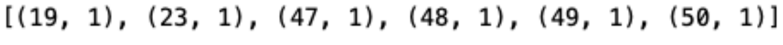
$$P(w_i \mid z_k = t) = \phi_{(zk,wi)}$$

The big advantage of LDA, when compared with pLSI and LSI thereby results in usually human interpretable topics, where each topic only contains words that are strongly associated with each other.

4.3. Implementation of LDA

Putting the LDA algorithm in practice and achieving good results requires some preprocessing work and hyperparameter fine-tuning. In this research, tuning the number of topics k was emphasized while for the rest of the parameters, default values were induced due to disproportionate computational tuning effort. Furthermore, the LDA model was tried out with different values for α and β, but large effects from adjusting these parameters could not be observed. However, before setting up the model, the text data must be preprocessed so that the algorithm can work with it (Naushan, 2020). For this, the text documents must be lowercased, punctuations must be removed, and all the terms must be tokenized and lemmatized.

Furthermore, so-called stopwords (e.g., *a, the, while*) and other frequently occurring words must be removed because they can potentially dilute the interpretability of the topics. Lemmatizing furthermore reduces inflectional forms of a word to a common base form (e.g., *am, are* to *be*) (Skorkovská, 2012). The lemmatized text will then be used to create the dictionary and corpus, where both will be both induced into the LDA algorithm. In the corpus creation process, a unique ID will be assigned to each term with the corresponding term frequency in the document. Bellow, an example of a corpus can be seen: For instance, (19, 1) implies that word ID 19 occurs one time in the respective document. Likewise, word ID 23 also occurs one time and so on.



```
[(19, 1), (23, 1), (47, 1), (48, 1), (49, 1), (50, 1)]
```

In order to determine the *k* number of topics where the LDA algorithm would show the best performance, the coherence value for each number of topics in a range from 1 to 30 will be calculated (Roeder et al. 2015). Appendix B shows the coherence scores on the Airbnb reviews data set. The LDA model appears to be optimal for either 22 or ten topics. However, after testing several *k* values, a smaller number of topics leads to better interpretability. This is why the LDA algorithm will be conducted with six topics (*coherence value = .5484*) which is still a good result (Roeder et al. 2015).

4.4. Descriptive Statistics

Before presenting the results of the LDA model and how the identified topics have evolved over time, the following section will outline some basic descriptive statistics of Airbnb properties in London in order to convey a sense of how the dataset is structured as well as how Airbnb is positioned in the city. The most popular neighborhoods are represented by Westminster, the Lower Hamlets and Camden. There are 60,864 unique listings that make up for 318,698 listing observations over the sample period. Appendix C provides the summary statistics of the listings in the underlying sample segmented by the pre and during COVID-19, as well as by the cleanliness classification. The two groups both pre and during COVID-19, show similar property and booking characteristics (e.g., number of accommodates, review scores and space type). However, it can also be seen that the share of booked houses has increased during the pandemic as well as the amount of Superhosts who furthermore have a larger share across clean listings. In addition to that, clean listings receive on average more reviews per month than not-clean listings. The number of listings started to decline in March 2020, coinciding with the number COVID-19 cases. The virus caused a significant decrease of 34.3% in the number of active listings from February 2020 to March 2020. From March to April 2020, due to lockdown measures and the burden on the health care system, there were 90% less Airbnb listings available.

The average property accommodates for approximately three people, about 30% of the listings are managed by Superhosts, which is an Airbnb classification for hosts with good response times and good ratings. The average price is €111 with standard deviation of €172. The resulting average generated income per month of €439, which is estimated by two nights per review received times the price per night. Furthermore, the average property gets 2.56 reviews per month which results in an occupancy rate of 4.5 days for each listing. Since the actual booking data is not available, occupancy rate is only estimated based on information about average booked nights and reviews per booking.

Using the LDA algorithm, combined with sentiment analysis on all the sentences that were related to cleanliness, each review was analyzed according to its dominant topic and sentiment score on the sentences related to *cleanliness*. Accordingly, a review was labeled as *perceived clean* if the most dominant topic was related to cleanliness and the corresponding sentiment score is positive. Consequently, all other reviews where the most dominant topic was not related to *cleanliness* and/or the sentiment score of a clean-related sentence was negative, the review is labeled as *not perceived clean.* It can be shown that prior to the pandemic, approximately 21% of the listings are *perceived clean* while only 14.8% of the observations are labeled *perceived clean* after the outbreak of COVID-19. This raises the question whether the tourists' hygiene standards have risen through the pandemic.

Delving into the statistics of the reviews, appendix E shows the distribution of document word counts. The graph shows that the reviews usually are short documents with a mean of 16 words and 88 words in the 99% percentile. Following this, appendix D displays the results of the LDA algorithm. The left panel thereby represents a general overview of the model and shows the prevalence of each topic as well as how they relate to each other. The topics are plotted as circles whose centers are defined by the computed distance between two topics. Thus, the closer to circles are, the closer the topics are related to each other. Following this, the prevalence of each topic is determined by the circle's area. The right part of the tool contains a bar chart representing the most frequent terms that are useful the topic interpretation. The two adjacent bars show the topic-related frequency of each term in red and the corpus-wide frequency in the color of gray. If no topic is selected, the right panel shows the top 30 most prominent terms of the dataset (Dossin 2018, Sievert and Shirley 2014).

The model identified six individual topics with the following relevant words: (1) *room, bathroom, bed, night, get, apartment;* (2) *station, walk, close, location, tube, minute;* (3) *home, make, feel, thank, host, stay;* (4) *check, easy, host, location, quick, clean;* (5) *location, clean, host, room, apartment, value, comfortable;* (6) *flat, host, locate, day, clean, location.* The LDA visualization shows that with topic 5, the algorithm successfully identified a topic with *clean* as a dominant term, and furthermore, generated other easily interpretable topics such as 2 (location and connection to public transport) or topic 3 (good relationship to host and convenient stay). Appendix F further displays the word counts and importance of topic key words which highlights the importance of the term *clean* within topic 5. The topic word clouds also confirm the relevance of this term in the corresponding topic (appendix H).

4.5. Hypothesis-Driven Evaluation of the Topic Modeling Results

After examining the datasets using descriptive methods, the next section evaluates the hypotheses formulated in *section 3* using the LDA algorithm and sentiment analysis with the Vader library. From here, practical implications can be generated that will result in recommended actions for both Airbnb and Airbnb hosts.

Hypothesis H1: As already indicated in the previous section, the LDA algorithm successfully identified a topic where the term *clean* has a significantly high relevance. An analysis of the development of topic relevance proves that with the onset of the COVID-19 outbreak, the topic of *cleanliness* is of increasing relevance, while other topics like location appear to have a decreasing importance. Figure 1 displays these results.

Empirically, figure 1 shows that the central range of normalized topic scores for the topic related to cleanliness has increased from approximately 22.5% to approximately 27.5% in the first wave of the pandemic over the course of three months. Between July and September 2020, notably the few months before the second COVID-19 wave, this rate remained at a constant level before it fell back to pre-COVID tines. Towards the end of the year 2020, the rate climbed back to the post outbreak level again. These results are of particular interest when comparing them to the topic weights other than *cleanliness*. Topic 4 which corresponds to the location of the property experienced a sharp decline with the onset of the coronavirus. Prior to the outbreak of COVID-19, its average topic weight was around 17.5% which decreased to less approximately 12.5% during the pandemic. This confirms the assumption that the risk factor hygiene is perceived as increasingly relevant while others lose importance. Finally, it can be shown that the perception of *cleanliness* has indeed changed due to COVID-19 and that Airbnb guests seem to be more aware of this factor.

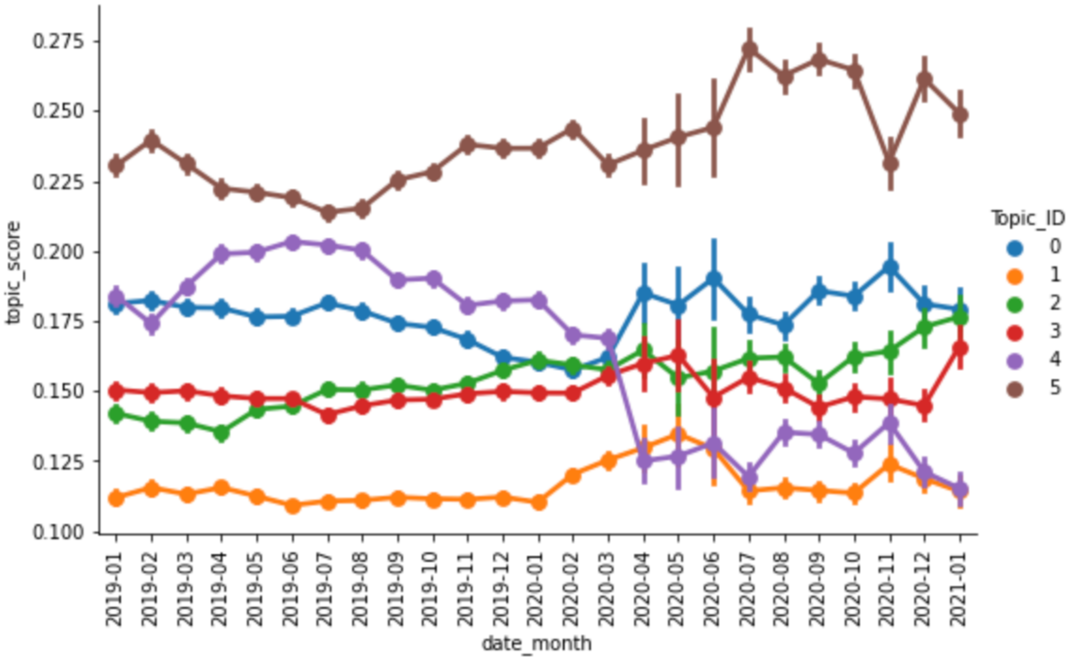


Figure 1: Development of topic weights

Hypothesis H2: The outbreak of COVID-19 impacted Airbnb listings whereby properties perceived as clean showed a better reaction towards the pandemic resulting in higher occupancy rates, increasing the price per night and consequently generated more monthly income compared to Airbnb listings not *perceived clean*.

Hypothesis H2 assumes that with the onset of COVID-19, the effects on occupancy rates, prices and monthly income of Airbnb listings can be observed, indicated by increasing attraction to *perceived clean* properties while the performance indicators for *not clean perceived* units show a declining trend. This assumption results from the fact that in times of pandemics, health is considered the biggest risk factor while travelling. The factor that corresponds to health issues in Airbnb properties is the perception of *cleanliness*, especially since at the beginning of the pandemic, there was a high uncertainty about how the virus was transmitted. It was clear that one way to contract COVID-19 was by touching surfaces that were covered with the virus or breathing in contaminated air. It is therefore assumed that *perceived clean* apartments are more in demand after the outbreak of the coronavirus. To test this hypothesis, the development of the previously mentioned performance indicators is analyzed over the course of the pandemic.

However, it is not examined whether the presented results are statistically significant which is why the statistical interpretation should be considered with reservation as biases could occur. Furthermore, only the parameter describing the perceived *cleanliness* is considered which should be extended to other variables in future studies.

Figure 2 represents the price development of Airbnb properties before and during COVID-19. The red line thereby marks the beginning of the global pandemic. It can be seen that prior to the pandemic there was a huge price gap between properties perceived clean and properties not perceived clean. Listings, where cleanliness was not actively been observed as positive, thereby, charge a higher price compared to properties that are *perceived clean*. With the onset of the pandemic, both lines appear to be closer compared to pre-COVID-times. During the second wave, the prices for *perceived clean* apartments begin to exceed those of *not perceived clean* units.

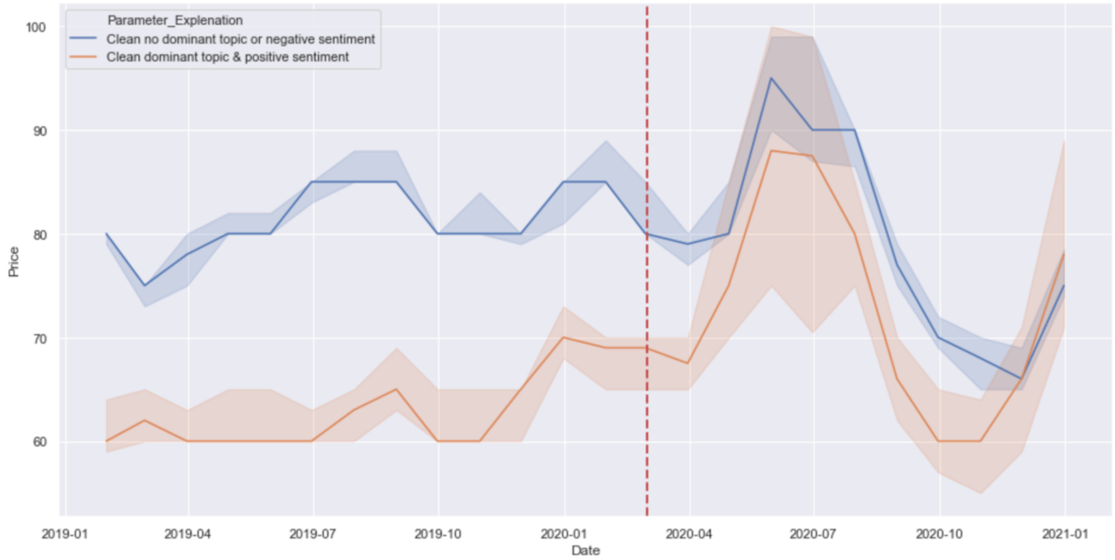


Figure 2: Price Development

Figure 3 shows the average monthly income of Airbnb properties that are perceived clean compared to units that are not perceived clean. Before COVID-19, both cohorts show a similar trend where *perceived clean* apartments have generated more income, which appears to be strongly affected by seasonal effects whereas *not clean perceived* apartments seem to be more resistant against this. However, both lines decrease sharply with the onset of the pandemic in March 2020, but properties perceived clean recover significantly quicker in the following months. The same development can be observed for average nights booked (Appendix G).

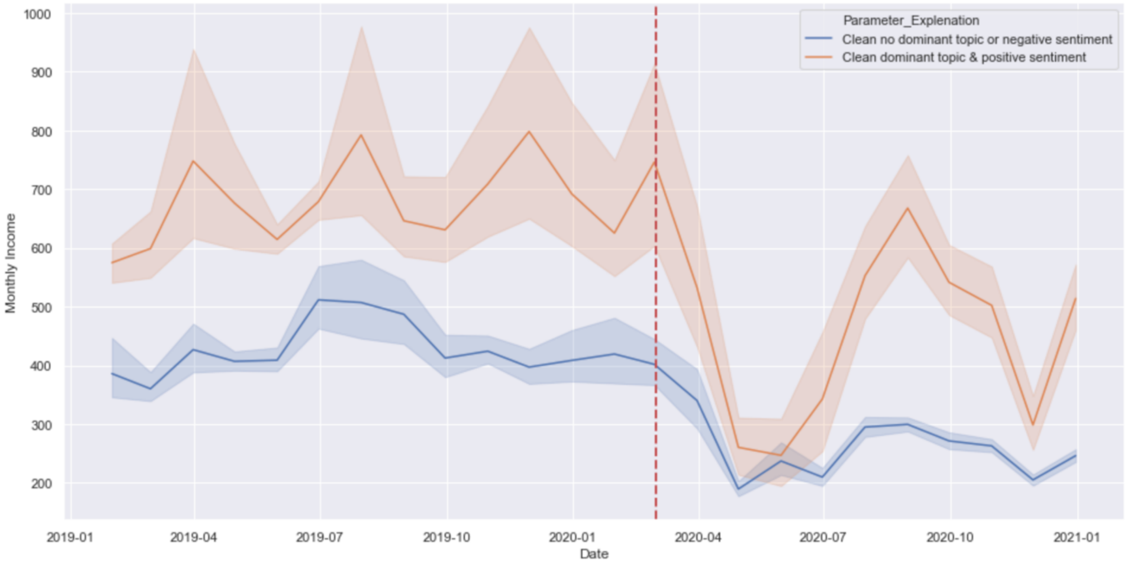


Figure 3: Development of Monthly Income

5. Discussion

The goal of this research was to determine whether the perception of *cleanliness* as a risk factor has changed due to the outbreak of COVID-19 and show the impact this has on key performance indicators such as occupancy rates, price per night and monthly generated income using the example of London during period between the 1st of January 2019 and the 31st of January 2021. The results from the LDA algorithm provide detailed knowledge about what Airbnb guests talk about in the reviews and how these topics evolve over time. In addition to these general insights, the research further indicated how this shift in risk awareness impacts the above-mentioned performance indicators. Since no research could have been found that combines topic modelling with sentiment analysis, the generated insights can provide significant added value to both Airbnb as an organization, as well as Airbnb hosts who seek to optimize their performance.

5.1. Summary and Interpretation of the Results

The results of the LDA model clearly indicate that the topic surrounding the *cleanliness* of Airbnb listings has gained relevance with the outbreak of COVID-19. Although hygiene had the strongest weight prior to COVID-19, due to the pandemic this share has experienced another incline. At the same time, other topics have lost relevance, such as the listing's location indicating that tourists put less importance in convenience and more about safety.

Considering the development of price, occupancy and monthly income over the sample period, underlines the implications of the LDA model. First, the price gap that existed between perceived clean units and not clean perceived ones has been diminished. Before COVID-19, *not clean perceived* listings charged higher prices per night on average. Figure 2 illustrates that one reason for this might that before the pandemic, price was mainly determined by factors like location and how well the property is connected to the public transport system. COVID-19 changed this picture. With the rise of the relevance of *cleanliness*, prices between *perceived clean* properties and *not perceived properties* almost equalize. Second, a similar trend can be observed with regards to monthly generated income and average nights booked. *Perceived clean* properties have always generated more income when compared with *not perceived clean* units - this has not changed with the outbreak of the pandemic. However, it can be observed that *perceived clean* properties clearly rebounded more significantly. A reason for this might be that due to the concept of word-of-mouth, people pay higher attention to references to cleanliness in the reviews, and thus, prefer properties where other guests have confirmed certain hygiene standards. This effect is even stronger during the outbreak of the second wave in November 2020. The combination of higher prices and higher occupancy rates of *clean perceived* properties ultimately leads to more income, and thus, results in a competitive advantage when compared to *not perceived clean* listings.

5.2. Limitations

Although the methodology developed and executed in this research has not been found in any relevant literature, combining both LDA topic modeling and sentiment analysis provides some distinguishing advantages compared to existing studies. Firstly, LDA topic modelling assures that apartments are only labeled as *perceived clean* when cleanliness is specifically emphasized and not only mentioned as a side note. Secondly, including sentiment analysis further guarantees that properties are only perceived as clean if the respective statement has positive connotation. Otherwise, reviews indicating that the property is not clean might falsely be labeled as clean (e.g., "*not very clean",* etc.).

However, this method has various some limitations. For instance, the results do not control for any other variables except for the perceived *cleanliness.* Accordingly, future studies could build up on this framework by creating coefficients of perceived cleanliness and property type, host status, neighborhood, etc. Furthermore, the sample period only covers two years in one specific city and the occupancy rate could only be estimated. Additionally, the city of London was in strict lockdown throughout most of the pandemic leading to an imbalance between data pre and during the pandemic. It would thus be valuable to include a variety of cities from different cultural backgrounds in a longer sample period in order to observe how the perception of cleanliness has changed in different scenarios.

Furthermore, research observing tourists' risk awareness comparing the importance of location versus safety post the pandemic can prove to be valuable to Airbnb hosts promoting their properties. Nonetheless, the biggest limitation is from the LDA model itself. First, LDA fails to assess correlations between the topics which is why the algorithm could define two topics that are similar (as observed in Appendix E) and sometimes difficult to interpret. Additionally, due to its unsupervised nature, it is difficult to build a model that shows no signs of over- or underfitting. To solve this, supervised classification models could be induced to LDA which would be time consuming due to the huge amount of required training data (Naushan 2020).

5.3. Practical Implications

Looking back to the research question, it can be concluded that the risk perception of tourists has changed due to COVID-19. Not only has the awareness of hygiene risen, with COVID-19 the awareness of this risk factor stands outs against other parameters, such as location and convenience. Furthermore, the fact that a significant number of properties are *perceived clean* will help Airbnb to bounce back as the pandemic persists, increase traction on the platform and consequently increasing revenues. Although Airbnb has introduced a new cleaning protocol (Airbnb 2020b), which serves as a checklist for hosts to prepare the property according to scientific health standards, there are still listings where the enhanced hygiene standards are not mentioned in the reviews. Since the results show that listings where cleanliness is specifically mentioned, generate more income, Airbnb should also implement processes that incentivize guests to write comments related to the cleanliness of a listing that go beyond rating cleanliness on a scale from 1 to 10. This is an opportunity where hosts can be more involved and communicate with their guests that reviews should relate to cleanliness because of its strong impact on other tourists purchasing decision.

6. Concluding Remarks

The outbreak of the COVID-19 pandemic has been affecting society since March 2020, with repercussions that will follow for years to come. Lockdown measures put our planet at a standstill. International borders closed globally at an unprecedented rate and highlighted the deficiencies in countries' public health system. Disruptions have been felt far beyond the tourism industry, affecting every aspect of human life.

The objective of this research was to identify whether the risk perception of tourists has changed with the outbreak of COVID-19 and to demonstrate how this has affected Airbnb. Using

Airbnb's reviews as the dataset provided valuable insights into about what travelers pay attention to when renting Airbnb units. By combining both topic modelling with LDA and sentiment analysis, the results of this research paper not only show the increasing awareness of cleanliness as a risk factor over time, but also how the changing perception of cleanliness has influenced occupancy rates, prices per night and monthly generated income by Airbnb hosts. Additionally, this research contributes to the academic literature of health crises, travel risk perception and tourist behavior delivering empirical evidence that disruptive events, such as a pandemic, significantly influence risk awareness of tourists. From a practical perspective, this research provides valuable insights for Airbnb as an organization, alongside Airbnb hosts, by developing appropriate communication strategies for tourists in order to convey a message of safety and high hygiene standards. Airbnb should therefore make more use of the reviews section as it provides a relatively unbiased platform for travelers to express their perception of health risk related to the property and can aid other travelers in purchasing their future stays.

Several limitations must be considered when interpreting the results of this research. First, findings of this study exclusively consider the perception of cleanliness and do not consider any other variables. In addition, the sample period covers a time in the pandemic characterized by strict measures, which is why future studies should imperatively study the long-term change in hygiene perception, including throughout several countries. Findings are furthermore limited by the res itself. The LDA algorithm is an unsupervised machine learning model which is almost impossible to tune to perfection and consequently always leaves room for some error. However, based on the framework provided in this research, future studies should compare these results with other cultural and geographical regions in order to find differences and similarities in risk perception. As the pandemic subsides, it will be of particular interest to understand to what degree individuals' habits relating to safety remain over the long-term and to what degree risk perception reverts to pre-pandemic times, emphasizing pleasure and convenience.

List of References

Abril, D. 2020. "Airbnb's IPO Filing Reveals Huge COVID Impact." *Fortune*. https://fortune.com/2020/11/16/airbnb-ipo-initial-public-offering-coronavirus-impact/. Last access at 21.04.2021

Airbnb 2020. "A Message from Co-Founder and CEO Brian Chesky." https://news.airbnb.com/a-message-from-co-founder-and-ceo-brian-chesky/. Last access at 10.05.2021

Airbnb. 2020. "Die Erweiterte Reinigung Auf Airbnb." https://de.airbnb.com/d/enhancedclean?_set_bev_on_new_domain=1621436527_ODFjOGI4N2VmMzNj. Last access at 10.05.2021.

Bauer, R.A. 1980. "Consumer Behavior as Risk Taking." In *Dynamic Marketing for a Changing World*. Chicago, IL, USA: Hancock, R.S., Ed; American Marketing Association.

Baum, T., S. K. K. Mooney, R. N. S. Robinson, and D. Solnet. 2020. "COVID-19's Impact On the Hospitality Workforce - New Crisis or Amplification of the Norm." *Int. J. Contemp. Hosp. Manag., 32*, no. 32: 2813–29. DOI: https://doi.org/10.1108/IJCHM-04-2020-0314.

Baxter, E., and D. Bowen. 2004. "Anatomy of Tourism Crisis: Explaining the Effects on Tourism of the UK Foot and Mouth Disease Epidemics of 1967–68 and 2001 with Special Reference to Media Portrayal." *International Journal of Tourism Research* 6 (4): 263–73. DOI: https://doi.org/10.1002/jtr.487.

Belk, R. 2014. "You Are What You Can Access: Sharing and Collaborative Consumption Online." *J. Bus. Res*, 1595–1600.

Bhat, M.R., M.A. Kundroo, T.A. Tarray, and B. Agarwal. 2019. "Deep LDA: A New Way to Topic Model." DOI: https://doi.org/10.1080/02522667.2019.1616911.

Blei, D.M., and J.D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics*, 17–35.

Blei, D.M., A.Y. Ng, and M.I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, no. 3: 993–1002.

Boros, L., G. Dudas, and T. Kovalcsik. 2020. "The Effects of Covid-19 on Airbnb." *Hungarian Geographical Bulletin* 69: 363–81. DOI: 10.15201/hungeBoobruosl,l.L6.9e.4t.a3l.

Botsman, R., and R. Rogers. 2010. "What's Mine Is Yours: The Rise of Collaborative Consumption." *Harvard Business Review*, no. 88: 30–33.

Çakar, K. 2018. "Critical Success Factors for Tourist Destination Governance in Times of Crisis: A Case Study of Antalya, Turkey." *Journal of Travel & Tourism Marketing* 35 (6): 786–802. DOI: https://doi.org/10.1080/10548408.2017.1421495.

Cheng, M., G. Chen, T. Weidmann, M. Hadjikakou, L. Xu, and Y. Wang. 2020. “The Sharing Economy and Sustainability - Assessing Airbnb’s Direct, Indirect and Induced Carbon Footprint in Sydney.” *J. Sustain. Tourism* 28 (8): 1083–99.

Curtis, Steven Kane, and Oksana Mont. 2020. “Sharing Economy Business Models for Sustainability.” *J. Clean Prod.*, no. 266.

Deerwester, S., S. Dumais, T. Landenauer, G. Furnas, and R. Harshman. 1990. “Indexing by Latent Semantic Analysis.” *Journal of the American Society of Information Science* 41 (6): 391–407.

Dossin, L. 2018. “Experiments on Topic Modeling – PyLDAvis.” *Object-Oriented Subject.* https://www.objectorientedsubject.net/2018/08/experiments-on-topic-modeling-pyldavis/. Lass access at 05.05.2021.

Findlater, A., and I.I. Bogoch. 2018. “Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel.” *Trends in Parasitology* 34 (9): 772–83. DOI: https://doi.org/10.1016/j.pt.2018.07.004.

Gallina, P. 2020. " Covid-19 health crisis management in Europe: Decisive assessment is needed now". *International Journal of Infectious Diseases* 96: 416. DOI: https://doi.org /10.1016/j.ijid.2020.05.010

Geissinger, A., C. Laurell, C. Ödberg, and C. Sandström. 2019. “How Sustainable Is the Sharing Ecnonomy? On the Sustainability Connotations of Sharing Economy Platforms.” *J. Clean Prod.* 206: 419–29.

Hall, C.M. 2010. “Crisis Events in Tourism: Subjects of Crisis in Tourism.” *Current Issues in Tourism* 13 (5): 401–4017. DOI: https://doi.org/10.1080/13683500.2010.491900.

Harrison-Walker, L.J. 2001. “The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment as Potential Antecedents.” *Journal of Service Research* 4 (1).

Ho, L.-L., Y.H. Tsai, W.-P. Lee, S.-T. Liao, L.-G. Wu, and Y.C. Wu. 2017. “Taiwan’s Travel and Border Health Measures in Response to Zika.” *Health Security* 15 (2): 185–91. DOI: https://doi.org/10.1089/hs.2016.0106.

Horgan, D., J. Hackett, C.B. Westphalen, D. Kalra, E. Richter, M. Romao, A.L. Andreu, et al. 2020. “Digitalisation and COVID-19: The Perfect Storm.” *Biomed. Hub*, no. 5: 1–23. DOI: https://doi.org/10.1159/000511232.

Hossain, M. 2021. “The Effect of the Covid-19 on Sharing Economy Activites.” *Journal of Cleaner Production* 280 (1).

Huang, A.H., R. Lehavy, A.Y. Zang, and R. Zheng. 2017. “Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach.” *Management Science,* 64 (4): 2833–55.

Inside Airbnb. (2021). Get the data. http://insideairbnb.com/get-the-data.html. Last access 20.04.2021.

Jamal, T., and C. Budke. 2020. “Tourism in a World with Pandemics: Local-Global Responsibility and Action.” *Journal of Tourism Futures*. DOI: https://doi.org/Doi: 10.1108/JTF-02-2020-0014.

Joo, H., B.A. Maskery, L.D. Rotz, Y.-K. Lee, and C.M. Brown. 2019. “Economic Impact of the 2015 MERS Outbreak on the Republic of Korea’s Tourism-Related Industries.” *Health Security* 17 (2): 100–108. DOI: https://doi.org/10.1089/hs.2018.0115.

Jun, S. 2020. “The Effects of Perceived Risk, Brand Credibility and Past Experience on Purchase Intention in the Airbnb Context.” *Sustainability* 12 (12): 5212–29.

Kamal, M. M. 2020. “The Triple-Edged Sword of COVID-19: Understanding the Use of Digital Technologies and the Impact of Productive, Disruptive and Destructuve Nature of the Pandemic.” *Inf. Syst. Manag.*, no. 37: 310–17. DOI: https://doi.org/10.1080/10580530.2020.1820634.

Kim, J., J. Kim, and L.R. Tang. 2020. “Effects of Epidemic Disease Outbreaks on Financial Performance of Restaurants: Event Study Method Approach.” *Journal of Hospitality and Tourism Management*, no. 43. DOI: https://doi.org/10.1016/j. jhtm.2020.01.015.

Kuo, H.-I., C.L. Chang, B.-W. Huang, C.-C. Chen, and M. McAleer. 2009. “Estimating the Impact of Avian Flu on International Tourism Demand Using Panel Data.” *Tourism Economics* 15 (3): 501–11. DOI: https://doi.org/10.5367/000000009789036611.

Lane, L. 2020. “How Bad Are Covid-19 Pandemic Effects On Airbnb Guests, Hosts?” *Forbes Magazine*, 2020. https://www.forbes.com/sites/lealane/2020/06/09/how-bad-are-covid-19-pandemic-effects-on-airbnb-guests-hosts/?sh=61cbcb2e7432.

Lean, H., and R. Smyth. 2009. “Asian Financial Crisis, Avian Flu and Terrorist Threats: Are Shocks to Malaysian Tourist Arrivals Permanent or Transitory?” *Asia Pacific Journal of Tourism Research* 14 (3): 301–21.

Lee, C.-K., H.-J. Song, L.J. Bendle, M.-J. Kim, and H. Han. 2012. “The Impact of Non-Pharmaceutical Interventions for 2009 H1N1 Influenza on Travel Intentions: A Model of Goal-Directed Behavior.” *Tourism Management* 33 (1): 88–89. DOI: https://doi.org/10.1016/j. tourman.2011.02.006.

Lee, S.H. 2020. “New Measuring Stick on Sharing Accommodatgion: Guest-Perceived Benefits and Risks.” *International Journal of Hospitality Management* 87.

Lee, S.H., and C. Deale. 2021. “Consumers’ Perception of Risks Associated with the Use of Airbnb before and during the COVID-19 Pandemic.” *Emerald Insight*.

Leggat, P.A., L.H. Brown, P. Aitken, and R. Speare. 2010. “Level of Concern and Precaution Taking among Australians Regarding Travel during Pandemic (H1N1) 2009: Results from the 2009 Queensland Social Survey.” *Journal of Travel Medicine,* 17 (5): 291–95. DOI: https://doi.org/10.1111/j.1708-8305.2010.00445.x.

Lei, H., and Chen Ying. 2021. “Exclusive Topic Modeling.” *Department of Statistics and Applied Probability, National University of Singapore*, 1–5.

Liang, L. 2018. “Understanding Repurchase Intention of Airbnb Consumers: Perceived Authencity, EWoM and Price Sensitivity.” *International Journal of Hospitality Management* 38 (1): 73–89.

Lim, J., and D. Won. 2020. “How Las Vegas’ Tourism Could Survive an Economic Crisis?” *Cities* 100. DOI: https://doi.org/10.1016/j.cities.2020-102643.

Linton, M., E.G.S. Teo, E. Bommes, C. Chen, and W.K. Härdle. 2017. “Dynamic Topic Modelling for Cryptocurrency Community Forums.” *Applied Quantitative Finance*, 355–72.

Liu, L., L. Tang, W. Dong, S. Yao, and W. Zhou. 2018. “An Overview of Topic Modeling and Its Current Applications in Bioinformatics.” *SpringerPlus* 5 (1).

Maier, D., A. Waldherr, P. Miltner, A. Wiedemann, A. Niekler, A. Kleinert, B. Pfetsch, G. Heyer, U. Reber, and T. Häussler. 2018. “Applying Lda Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology.” *Communication Methods and Measures* 12 (2–3): 93–118.

Maphanga, P.M., and U.S. Henema. 2019. “The Tourism Impact of Ebola in Africa: Lessons on Crisis Man- Agement.” *African Journal of Hospitality, Tourism and Leisure* 8 (3): 1–13.

McAleer, M., B.-W. Huang, H.-I. Kuo, C.-C. Chen, and M. Cheng. 2010. “An Econometric Analysis of SARS and Avian Flu on International Tourist Arrivals to Asia.” *Environmental Modelling & Software* 25 (1): 100–106.

Mody, M, C. Suess, and X. Lehto. 2019. “Using Segmentation to Compete in the Age of the Sharing Economy: Testing a Coreperiphery Framework.” *Int. J. Hospit. Manag.* 78: 119–213.

Mohamed, T. 2020. “Airbnb IPO Could Be the ‘steal of the Century’ If People Keep Switching from Hotels to Homes, Jim Cramer Says.” *Business Insider*. https://news.yahoo.com/airbnb-ipo-could-steal-century-105012112.html?guccounter=1. Last access at 10.05.2021

Mont, Oksana, Steven Kane Curtis, and Yuliya Voytenko Palgan. 2021. “Organisational Response Strategies to COVID-19 in the Sharing Economy.” *Sustainable Production and Consumption*, no. 28: 52–70. DOI: https://doi.org/10.1016/j.spc.2021.03.025.

Monterrubio, J.C. 2010. “Short-Term Economic Impacts of Influenza A (H1N1) and Government Reaction on the Mexican Tourism Industry: An Analysis of the Media.” *International Journal of Tourism Policy* 3 (1). DOI: https://doi.org/10.1504/IJTP.2010.031599.

Moreira, P. 2008. “Stealth Risks and Catastrophic Risks: On Risk Perception and Crisis Recovery Strategies.” *Ournal of Travel & Tourism Marketing* 23 (2–4): 15–27. DOI: https://doi.org/10.1300/J073v23n02_02.

Muñoz, P., and B. Cohen. 2018. “A Compass for Navigating Sharing Economy Business Models.” *Calif. Manage. Rev.*, no. 61: 114–47.

Naushan, H. 2020. “Topic Modeling with Latent Dirichlet Allocation.” *Towards Data Science*. https://towardsdatascience.com/topic-modeling-with-latent-dirichlet-allocation-e7ff75290f8#_=_. Last access at 10.05.2021.

Omrani, A.S., and S. Shalhoub. 2015. “Middle East Respiratory Syndrome Coronavirus (MERS- CoV): What Lessons Can We Learn?” *Journal of Hospital Infection* 91 (3): 188–96. DOI: https://doi.org/10.1016/j. jhin.2015.08.002.

Phua, V.C. 2019. “Perceiving Airbnb as Sharing Economy: The Issue of Trust in Using Airbnb.” *Current Issues in Tourism* 22 (17): 2051–55.

Pine, R., and B. McKercher. 2004. “The Impact of SARS on Hong Kong’s Tourism Industry.” *International Journal of Contemporary Hospitality Management* 16 (2): 139–43. DOI: https://doi.org/10.1108/09596110410520034.

Rassy, D., and R.D. Smith. n.d. “The Economic Impact of H1N1 on Mexico’s Tourist and Pork Sectors.” *Health Economics* 22 (7): 824–34. DOI: https://doi.org/10.1002/ hec.2862.

Reddy, M.V., S.W. Boyd, and M. Nica. 2013. “Towards a Post-Conflict Tourism Recovery Framework.” *Annals of Tourism Research* 84. DOI: https://doi.org/10.1016/j.annals.2020.102940.

Reisenbichler, M., and T. Reutterer. 2019. “Topic Modeling in Marketing: Recent Advances and Research Opportunities.” *Journal of Business Economics* 89 (3): 327–56.

Ritter, M., and H. Schanz. 2018. “The Sharing Economy: A Comprehensive Business Model Framework.” *J. Clean Prod.*

Roeder, M., A. Both, and A. Hinneburg. 2015. “Exploring the Space of Topic Coherence Measures.” *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. DOI: https://doi.org/10.1145/2684822.2685324.

Ryu, S., H. Gao, J.Y. Wong, E.Y.C. Shiu, J. Xiao, M.W. Fong, and B.J. Cowling. 2020. “Non-Pharmaceutical Measures for Pandemic Influenza in Non-Healthcare Settings –International Travel-Related Measures.” *Emerging Infectious Diseases* 26 (5): 961–66. DOI: https://doi.org/10.3201/eid2605.190993.

Salton, G., and M. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw Hill.

Schroeder, A., L. Pennington-Grey, K. Kaplanidou, and F. Zhan. 2013. “Destination Risk Perceptions among U.S. Residents for London as the Host City of the 2012 Summer Olympic Games.” *Tourism Management* 38: 107–19.

Sharpley, R., and B. Craven. 2001. “The 2001 Foot and Mouth Crisis – Rural Economy and Tourism Policy Implications: A Comment.” *Current Issues in Tourism* 4 (6): 527–37. DOI: https://doi.org/10.1080/13683500108667901.

Sievert, C. and Shirley, K.E. 2014. " LDAvis: A method for visualizing and interpreting topics". DOI: 10.13140/2.1.1394.3043.

Sjöberg, L., B.E. Moen, and T. Rundmo. 2004. “Explaining Risk Perception. An Evaluation of the Psychometric Paradigm in Risk Perception Research.” *Rotunde* 84.

Skorkovská, L. 2012. “Application of Lemmatization and Summarization Methods in Topic Identification Module for Large Scale Language Modeling Data Filtering.” *7499*. DOI: https://doi.org/10.1007/978-3-642-32790-2_23.

So, K.K.F., H. Oh, and M. Somang. 2018. “Motivations and Constraints of Airbnb Consumers: Findings from a Mixed-Methods Approach.” *Tourism Management* 67 (1): 224–36.

Sparke, M., and D. Anguelov. 2012. “H1N1, Glo- Balization and the Epidemiology of Inequality.” *Health & Place* 18 (4): 726–36. DOI: https://doi.org/10.1016/j. healthplace.2011.09.001.

Yanni, E.A., N. Merano, and P. Han. 2010. “Knowledge, Attitudes, and Practices of US Travelers to Asia Regarding Seasonal Influenza and H5N1 Avian Influenza Prevention Measures.” *Journal of Travel Medicine* 17 (6): 374–81. DOI: https://doi.org/10.1111/ j.1708-8305.2010.00458.x.

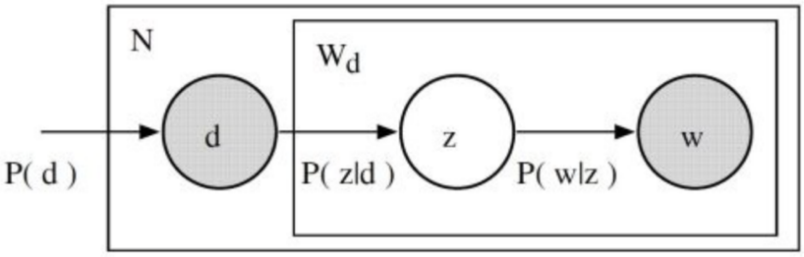
Yu, J., Seo, J., Hyun, S.S. 2021. "Perceived hygiene attributes in the hotel industry: customer retention amid the COVID-19 crisis." International Journal of Hospitality Management 93. DOI: https://doi.org/10.1016/j.ijhm.2020.102768.

Zervas, G., D. Proserpio, and J.W. Byers. 20167. “The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry.” *Journal of Marketing Research* 54 (5): 687–705.

List of Appendices

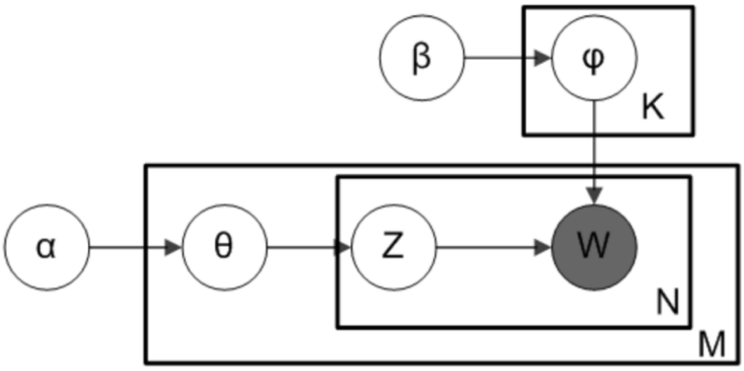
Appendix A: Topic Model Descriptions

Appendix A1: pLSI Model Description



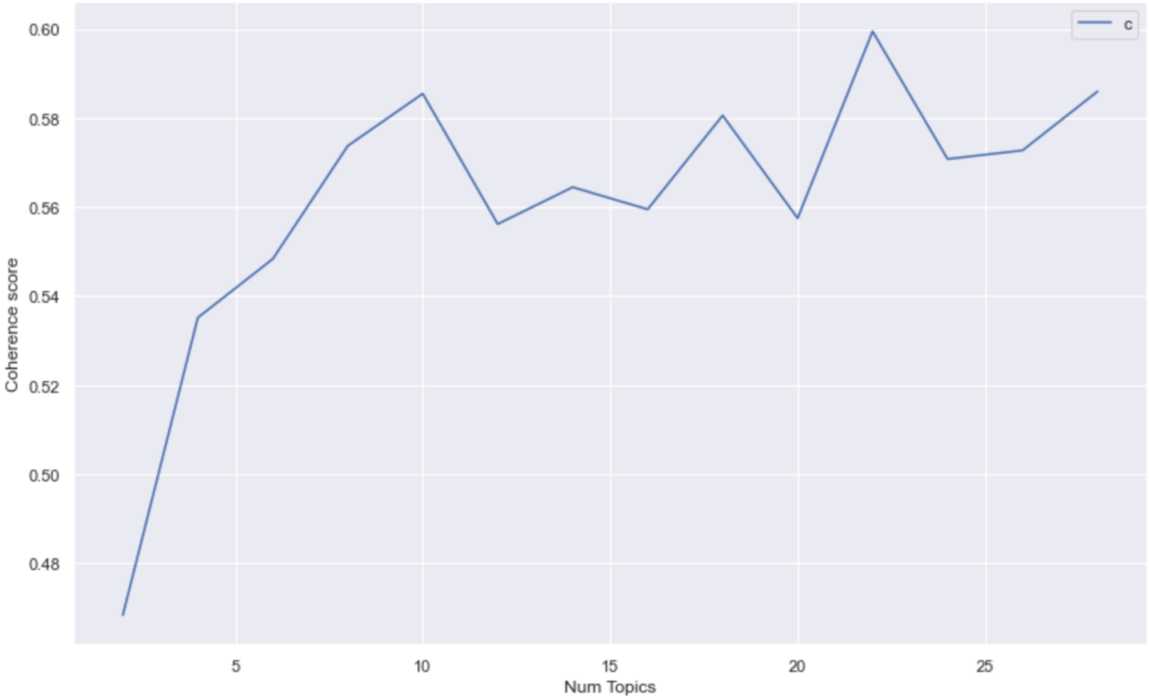
This plate notation displays the concept of the pLSI developed by Hofmann (1999). The grey circles correspond to the observable variables which are the documents and words in this regard. Thereby topic *z* is present in document *d* with a probability of $P(z \mid d)$. Similarly, word *w* belongs to topic *z* with a probability of $P(w \mid z)$.

Appendix A2: LDA Model Description



Compared to the pLSI model, the LDA model contains two hyperparameters α and β which correspond to the document-topic density and topic-word density respectively.

Appendix B: Scores of Coherence Value Analysis



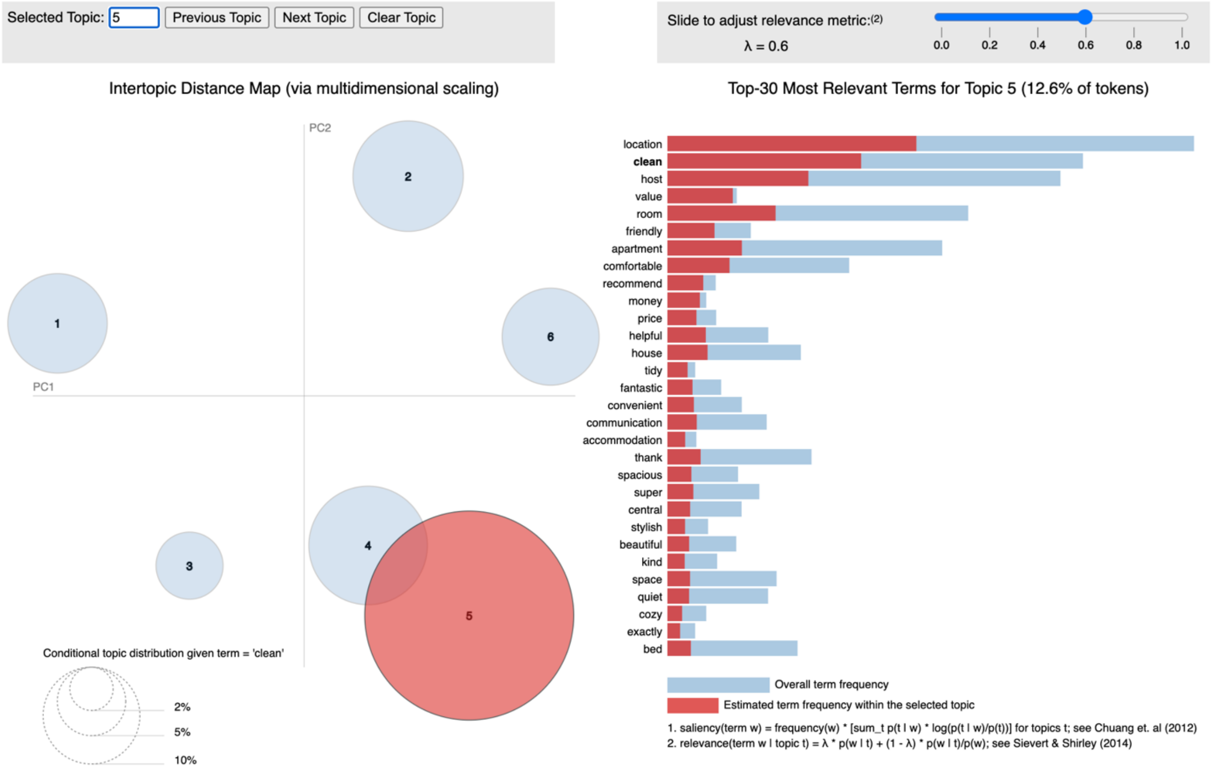
This graph represents the results of the coherence value analysis for the LDA algorithm. The model works best for 22 topics. However, the model uses six topics due to better interpretability.

Appendix C: Descriptive Statistics of Airbnb Listings

	Before Covid				During Covid			
	Not Clean		Clean		Not Clean		Clean	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Unit Capacity (Accommodates)	3,239	1,926	3,145	1,865	3,239	1,926	3,145	1,865
Indicator if Entire Home	0,577	0,475	0,503	0,491	0,619	0,484	0,577	0,494
Indicator if Private Room	0,367	0,467	0,430	0,482	0,370	0,482	0,410	0,491
Indicator if Shared Room	0,005	0,070	0,005	0,071	0,004	0,064	0,004	0,064
Price	118,352	173,491	107,177	179,019	106,302	157,855	99,594	208,766
Monthly Income	398,120	1724,429	685,196	3302,831	291,749	1558,856	454,739	1285,852
Nights Booked	3,903	2,324	7,200	4,120	2,923	1,640	5,484	4,446
Comments per Month	2,178	1,582	4,378	2,783	1,509	0,994	3,126	3,421
Review Scores (Accuracy)	9,481	0,684	9,572	0,617	9,521	0,794	9,612	0,660
Review Scores (Check-In)	9,558	0,639	9,640	0,595	9,627	0,722	9,689	0,625
Review Scores (Communication)	9,586	0,623	9,679	0,574	9,646	0,723	9,714	0,621
Review Scores (Location)	9,488	0,586	9,563	0,569	9,600	0,647	9,657	0,578
Review Scores (Value)	9,280	0,717	9,363	0,672	9,305	0,823	9,390	0,709
Indicator if Superhost	0,207	0,359	0,273	0,416	0,284	0,442	0,334	0.466
Number of Observations	50.798		23.001		24.325		5.379	

Note that nights booked is an estimate calculated by multiplying the number of reviews by 2. Monthly Income is calculated by multiplying the price with nights booked. Furthermore, clean refers to listings with *cleanliness* as a dominant topic and a positive sentiment score while not clean properties either have a negative sentiment score and/or a dominant topic not related to *cleanliness*.

Appendix D: Visualization of the LDA Results



Selected Topic: 5

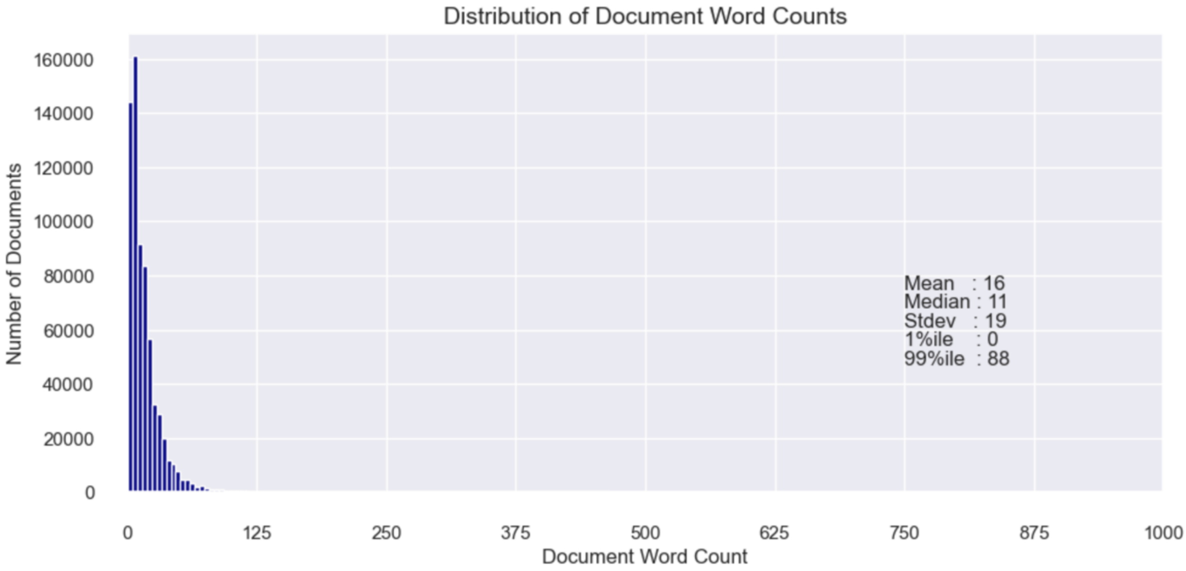
Slide to adjust relevance metric:(2)

λ = 0.6

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

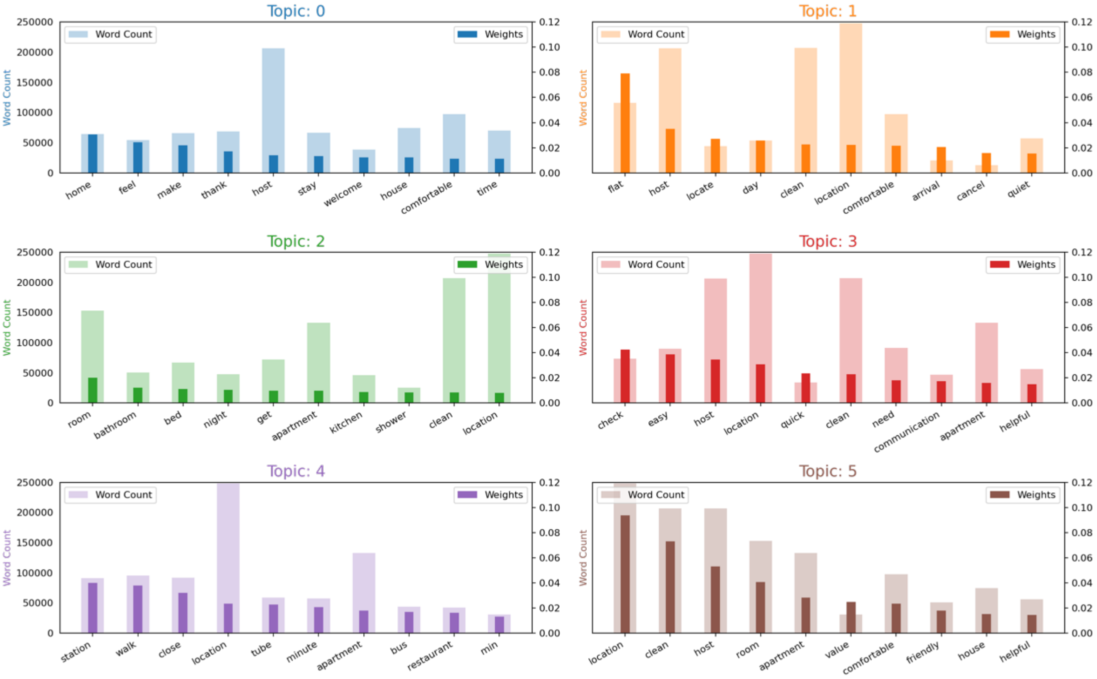
Here, the results of the LDA model are visualized. Accordingly, the topic where cleanliness has the highest relevance is assigned to bubble 5. The closer the bubbles are located to each other, the more related the topics are.

Appendix E: Distribution of Document Word Counts



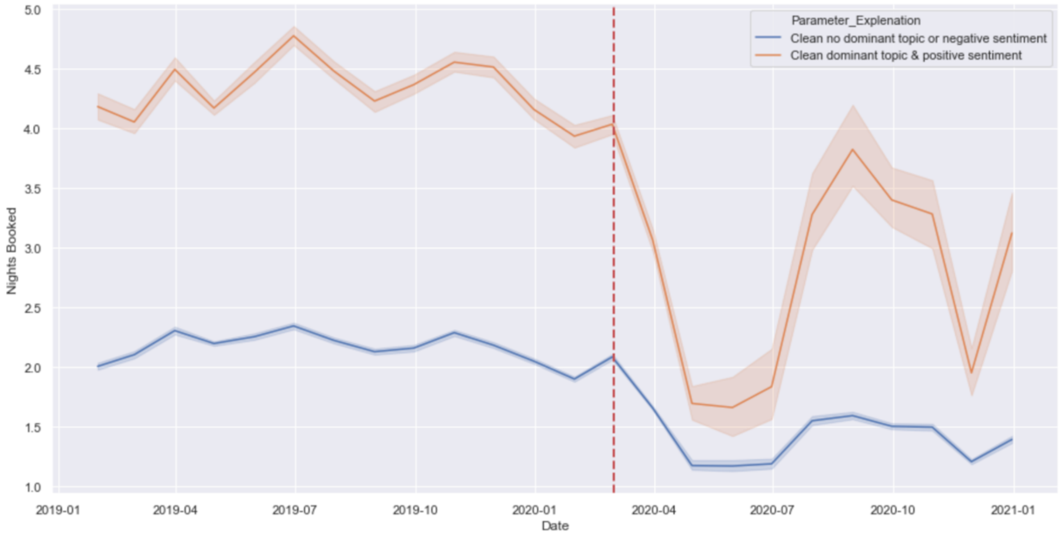
The graph shows that most of the documents are relatively short with an average of 16 words.

Appendix F: Word Count and Importance of Topic Keywords



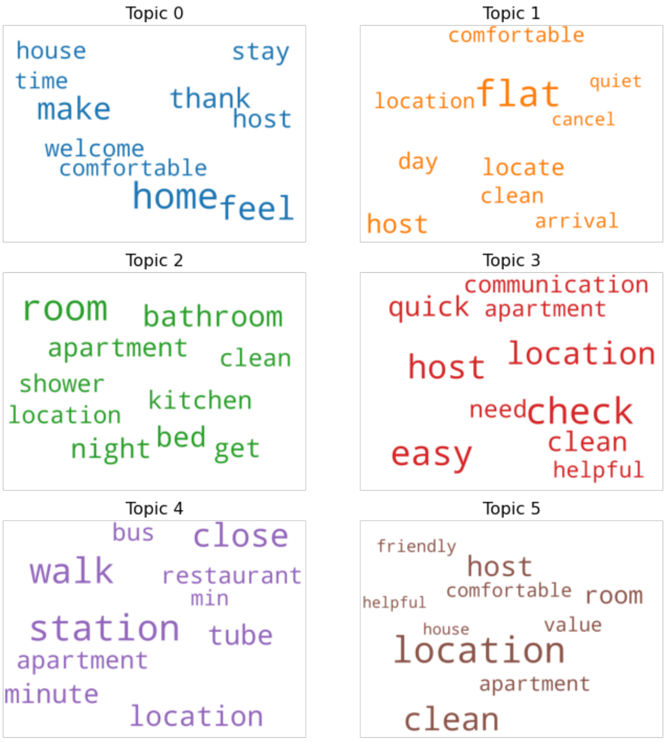
Topic 5 corresponds to the topic where cleanliness is very relevant. It can be seen that it is counted approximately 200,000 times and has a weight of approximately 8%.

Appendix G: Development of Monthly Nights Booked



The average occupancy rate shows a similar trend compared to monthly income. The orange line corresponding to *perceived clean* apartments rebounds more significantly in the COVID-19 pandemic, compared to *not clean perceived* apartments

Appendix H: Word Clouds of Topics



These word clouds illustrate the most relevant words in each corresponding topic where the size of the word relates to its term-frequency. It can be seen that in topic 5, the term clean is one of the biggest words highlighting its dominance and relevance in the respective topic.