The impact of COVID-19 on sharing economy in Singapore: Role of customer's perceived risk

Kou Dan Ni Tara

Dulwich College (Singapore), Singapore, 71 Bukit Batok West Ave. 8, 658966, Singapore

kdnaijay@gmail.com

Abstract. The unprecedented COVID-19 pandemic has had a significant impact on the tourism industry. Previous research has studied the perspective of Airbnb hosts, but rarely mentioned the effect on consumer choice. As consumer choices may be driven by their risk perceptions, how COVID-19 may influence is not well-understood. This paper aims to investigate how COVID-19 may change consumer risk perception on different Airbnb listings, and such an impact on risk perception may affect their choices and overall business performance. The paper utilizes large-scale data about customer reviews and listing characteristics from Airbnb Singapore in the years 2019-2021, using NLP techniques to analyze, and extract the change in tourists' risk perception before and after the outbreak of COVID-19 in Singapore. Comparing the topical content from customer reviews, we find that customers tend to perceive their experience more significantly on cleanliness-related issues since the outbreak of COVID-19. Furthermore, sentiment Analysis is performed to classify customer reviews into positive and negative, respectively, such that all listings are classified according to their valence in the perceived cleanliness. We show that the number of bookings and revenues of listings perceived as clean and not clean by the customers has diverged since the pandemic, i.e., the listings that are perceived clean would receive more bookings and thus accrue more revenues than those that are perceived as not clean. Overall, the research results would provide insight for Airbnb hosts and platforms to manage listings in the post-pandemic era, especially on how to manage and communicate cleanliness-related risks.

Keywords: COVID-19 Impact, Sharing Economy, Risk Perception, Natural Language Processing.

1. Introduction

A sharing economy, by definition, is an IT-facilitated peer-to-peer model for commercial or non-commercial sharing of underutilized goods and service capacity through an intermediary without a transfer of ownership [1]. As a type of shared economy, Airbnb replaces the economy in the current residential needs. On the supply side, Airbnb is created to meet the extra demand that cannot be fulfilled by the number of hotels. On the demand side, the number of hotels is enough for consumers who are willing and able to pay the high price for hotels, but at the same time filters those who are not willing and able to pay the price. Therefore, Airbnb fulfills the needs of consumers willing to pay less for residential. In the past, Airbnb is a side occupation for people who have extra vacant houses to gain extra income; however, the platform becomes more professional and the main source of income for hosts. Airbnb steals the market share and put pressure on hotels, forming a disruption relation with the hotels.

Airbnb nowadays has certain characteristics: large-scale and platform. In terms of large-scale, Airbnb has the largest world hotel chain with data on Airbnb's website stating that about 2 million people stay in an Airbnb every night [2]. In terms of platform, Airbnb does not involve hotel management; instead, each host manages its own houses. Hosts put photos to reassure customers, lowering risk and uncertainty. Customer reviews have the same purpose, which is to serve as a proxy for room quality on Airbnb and reduce concerns, but usually fewer positive comments are sometimes communicated using more nuanced and subtle cues. Unlike hotels that belong to certain brands to guarantee their quality regardless of location, Airbnb has a higher risk in terms of quality. Thus, consumers have a greater need for extra information (reviews or photos) that help to lower risk.

In 2020, the outbreak of the COVID-19 pandemic affected countries, societies, and individuals across the globe and resulted in a public health crisis. This leads to industry a significant downturn for many industries, such as sharing economy. Thus, managing the severe consequences of COVID-19 for sharing economy would benefit from the crisis management framework that has been proven successful for managing diseases and natural disasters. Crisis management is a systematic attempt by organizational members with external stakeholders to avert crises or to effectively manage those that do occur [4]. The Crisis Management framework includes disasters and other forms of crises (e.g. epidemics, conflict, pollution) that can lead to a reduction in visitation to the affected area [5]. Previous studies have demonstrated the effect of epidemics and natural disasters on tourism. Epidemics have negative effects only for related areas, such as SARS affecting east Asia and the Ebola virus affecting Africa. Natural disasters (e.g. Tsunamis, floods, and volcanoes) constitute substantial negative motivators for prospective visitors [6]. However, COVID-19, compared to other crises, has a more significant effect in terms of width and depth. Most current research on COVID-19 focuses on hotels, but at this moment, there is only a limited amount of academic literature that describes and discusses the impact of COVID-19 on the Airbnb industry. Researchers have mixed evidence supporting both the positive and negative impacts of COVID-19 on Airbnb. The possible positive effect is that people have the need to travel during COVID-19, but Airbnb is favorable to hotels as it allows people to stay away from other tourists. On the other hand, COVID-19 reduces the overall number of tourists, which negatively impacts Airbnb.

Across the numerous studies, common ground can be found in that declines in sharing economy activities are majorly caused by concerns about tourists' health and security and by non-pharmaceutical interventions like lockdowns, surveillance, border control, and quarantine [6]. These observations are coherent with the concept of perceived risk theory [7]. According to this, the risk perception of tourists plays an important role in tourist behavior and travel decisions. However, it depends on the tourist's experience and knowledge in such critical situations. Therefore, determining the change in risk perception is essential to studying the impact of COVID-19 on tourists. The Risk perceptions have always existed: it is "safe" in the pre-covid period but shifts to "health" during COVID-19. Tourists increasingly look for individual social/physical distancing due to the fear of being infected and infecting others [8]. In theory, there are already existing studies showing the effect of COVID-19 on Airbnb hosts; however, only little is known about how COVID-19, which exceeds all previous tourism crises, has influenced the risk perception of travelers. The impact of the change of risk perception due to COVID-19 on business is unexplored, and thus this paper will fill this gap. The central objective of this paper is to assess the impact of COVID-19 on Airbnb in Singapore, a leading peer-to-peer accommodation platform, and analyze whether the perception of risk has changed due to pandemics. The main variable considered is perceived cleanliness.

Tourists' risk perception is highly subjective. Traditionally, researchers use surveys to study consumer risk perception, but customer review is used in this paper. Customer reviews that are from those who have stayed with Airbnb are used as they are believed to be a reliable source to measure customer perception, reducing uncertainty. To understand the change in risk perception, natural language process techniques are applied to analyze large-scale Airbnb customer reviews from 2019 to 2021, the year before and after the outbreak of COVID-19, so changes in risk perception can be observed. Then Topic modeling with the Latent Dirichlet Allocation (LDA) algorithm is used to extract topics from reviews, and the dynamics of risk factors such as perceived cleanness are analyzed across the whole period of pre- and post-COVID-19. Furthermore, sentiment analysis is applied to the customer reviews to examine discrepancies across price, monthly income, and occupancy rate of Airbnb listings to understand whether the perceived risks would impact customer choices. Overall, this research would provide empirical evidence by modeling and monitoring customer risk perception in the hospitality service industry before and after covid and how it changed. This research would also introduce methods to examine the impact of the change in customer risk perception on Airbnb's business performance.

The research is done by firstly collecting data on Airbnb customer reviews, then using topic modeling to identify topics related to cleanliness, tracking how the significance of these topics evolves throughout COVID-19. Next, sentiment analysis is used to distinguish positive and negative comments, and finally, the results of sentiment analysis are compared with Airbnb room availability data and revenue to discover whether people's risk perception would affect business performance. In conclusion, the risk perception after covid greatly shifted to cleanliness. The list perceived as negative receives fewer bookings and revenues than the list perceived as positive.

The paper has mainly three important contributions. First, the paper provides empirical evidence on how COVID-19 impacted Airbnb activity in Singapore, in which tourism plays a significant role in its economy. Second, the paper leverages NLP techniques to extract the topical content and sentiment that may characterize the tourists' perception changes of Airbnb after COVID-19. Third, the paper explains the discrepancy between listings through the mechanism of tourists' risk perception of cleanliness.

The rest of the paper is organized as follows. Section 2 reviews the related work on crisis management in the tourism industry. Section 3 introduces the Airbnb data in Singapore regarding the listing descriptions and customer reviews. Section 4 presents details of applying NLP techniques to extract the topics and sentiment from customer reviews and shows how customers' risk perception about the cleanliness of the Airbnb listings change over the course of COVID-19, which in turn affects listings bookings and revenues. Section 5 discusses the economic implications of risk perception changes for the Airbnb platform and hosts, which may shed light on hospitality management during the post-Covid era. Section 6 concludes with the limitation and future work.

2. Related Research

The crisis management framework consists of natural disasters and epidemics. To investigate the impact of covid on tourism, it's necessary to see how other crises affect tourism and apply it to the COVID-19 scenario. In 2019, Jaume Rosselló, Susanne Becken, and Maria Santana-Gallego published research on the effects of natural disasters on international tourism. Disasters and other forms of crises can lead to a reduction in visitation to the affected area, and may even spread to neighboring areas that are not impacted by the event. Several reasons explain the reduction of tourism. The decline in tourist arrivals may be due to people's risk perceptions and avoidance of regions that are deemed unsafe [9,10], or potential travelers may feel uncomfortable or have ethical concerns about traveling to a disaster region [5]. On the other hand, disasters may lead to an increase in visitation due to media coverage about a natural, or perhaps also man-made, phenomenon playing an informative role as a motivating factor to visit a region. Natural disasters and unexpected events may also cause the arrival of people from other countries for humanitarian reasons but also to visit friends and relatives who have been victims of those events. Additionally, the final reasons are death tourism, which involves traveling to places historically associated with death and tragedy [11]. This research explained the reason why some people may still travel during COVID-19 either because of media reports or visiting families and friends, but the decline is caused by the idea of risk perception, people perceive the place as unsafe. In the COVID-19 scenario, tourism is affected because people think traveling increases the likelihood of being infected or infecting others, so they feel unsafe or uncomfortable about traveling. This research provides insights for my research as it suggested that the essence of understanding the impact of COVID-19 on travel is to understand consumer risk perception.

Coronavirus represents an economic super-shock [12]. An economic shock is "any change to fundamental macroeconomic variables or relationships that has a substantial effect on macroeconomic outcomes and measures of economic performance, such as unemployment, consumption, and inflation" [12]. The Covid-induced crisis is distinct in three ways: (1) the economic shock and the consequent travel decline are global; (2) the economic shock is more dramatic, with reductions to economic growth twice as big as those caused by regular shocks; (3) the shock has the potential to trigger structural changes in certain sectors of the industry [12]. The research mainly explained the impact of covid on Airbnb hosts. The first hypothesis stated that the proportion of investor-hosted Airbnb listings would drop since hosts who cannot cover the costs of the short-term market would enter the long-term rental house market. The second hypothesis claimed that Airbnb would recover from covid but would never get to the same level of pre-covid as no hosts would return to the market. They now recognize economic shock as a possible risk factor when making decisions.

Furthermore, the research by Marinko Škare, Domingo Riberio Soriano, and Małgorzata Porada-Rochoń in 2020 explained the impact of COVID-19 with the comparison of natural disasters and epidemics. The effect of COVID-19 on the travel and tourist industry will be unmatched by that of earlier pandemics [13]. The tourism business will be required to forgo innovation and product development, facilitate touring visitors and experience development, and organize efforts to deliver consumer experience by boosting traveler confidence and lowering perceived travel dangers. This again confirms that COVID-19 has a large impact on the tourism industry and businesses should focus on consumer experiences, which are determined by consumer risk perception. The paper focused on the perspective of Airbnb hosts, but the effect of COVID-19 on consumers remained unknown.

The research by Lajos Boros, Gábor Dudás, and Tamás Kovalcsik in 2020 explained the impact of COVID-19 on Airbnb, the most popular P2P platform. Due to the closing down of borders, fear of the virus, and the lockdown measures applied by local and national authorities [14]. The negative impact on Airbnb is mainly the large cancellation and rise in occupancy rate. Airbnb guests canceled their reservations or did not make new ones after the pandemic had started to spread [12]. Many authors concluded that the reason for declined bookings is due to the fact that tourists are concerned about their health and safety [15-18]. Another reason is that tourism may contribute to the spread of disease. The greater mobility of people and the accessibility and affordability of air travel contribute to the rapid spread of infectious diseases [19-21]. This research suggested that consumer risk perception had a great influence on their decisions. If travelers are concerned about their health and safety, it affects tourist flows negatively [22]. However, the authors also admit that the effects of diseases on Airbnb are rarely analyzed, due to the fact that since the emergence of this accommodation platform, there have been only a few disease outbreaks that affected tourism markets, and only with limited effect [22], so this research paper would help to fill a long-existing research gap.

Risk recovery is always a challenge since "the essence of risk management lies in maximizing the areas where we have some control over the outcome while minimizing the areas where we have no control over the outcome and the linkage between effect and cause is hidden from us" [23]. Therefore, understanding consumer risk perception and reacting to it is valuable and essential. The central concept of the paper is risk perception, which was also mentioned in the research by Pedro Moreira in his paper in 2008. Risk perception is an element of the general perceived image of products or services and was found to have a critical impact on organizational results [7]. The development and success of a tourist destination depend on the perceived image that tourists and residents have of the present reality and the future.

3. Data

3.1 Preprocessing

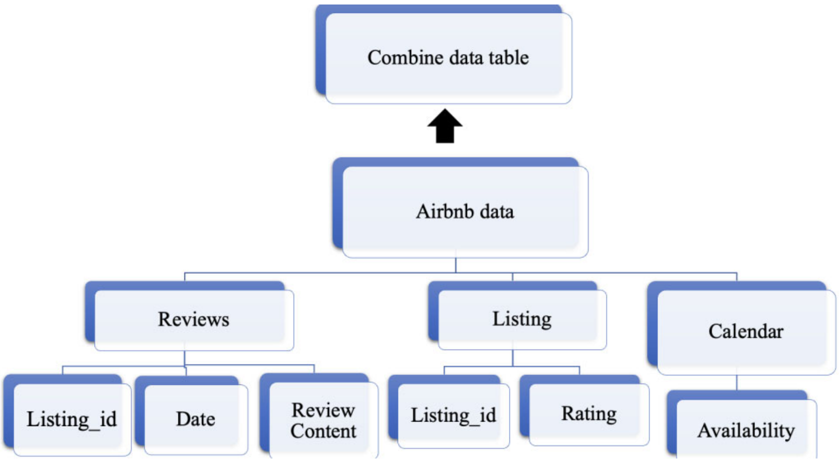

Figure 1. The Composition of the Final Airbnb Data

The initial customer review is imported from Airbnb between 2014 and 2022; therefore, reviews from 2014 to 2018 are deleted as they do not pertain to the covid period. The data used in this study are Airbnb customer reviews from 2019 (pre-covid) to 2021 (post-covid), so that the full length of time before and after covid can be determined. This research utilized three primary data sources, as depicted in the figure above. The entire data set is obtained by extracting three data types: reviews, listings, and calendars. The extracted fields from the review data included Listing id, date, and review content. The reviews contained 43,239 rows and 6 columns. Listing data consisted of 3,672 rows and 74 columns and extracted listing fields included Listing id and ratings. The calendar data contained 1339917 rows and 7 columns, with the extracted field being the room's availability. Below is a table containing the final Airbnb data, which consists of 85396 rows and 7 columns, based on the columns listed in the preceding section.

Table 1. Sample of the Final Airbnb data after emerging reviews, listing, and calendar

	listing_id	id	date	review_id	reviewer_name	comments	year
0	71609	496082140	2019-07-27	220147487	John	Attractive place to stay for 4pax when visiting S...	2019
1	71609	524932232	2019-09-07	2898879	Jeni	The location is very close to EXPO and Bus stop...	2019
2	7160	571338964	2019-12-01	287700916	Eigen John	Outstanding value for money. The place is near...	2019

Data preprocessing is required before the data can be used. As the customer reviews contain various languages, numbers, and symbols, the computers may not be able to identify them during the topic modeling procedure. Therefore, the deep translator language detection API module is employed to identify non-English languages and only English reviews are retained. Each individual has his or her writing style and punctuation, which complicates data analysis. Therefore, I remove numbers, short reviews, blank reviews, and punctuations, which in total account for 16% of the data, leaving 71,270 data. In addition, all reviews are converted to lowercase so that two identical words with different capitalizations will not be considered distinct. Although often overlooked, converting all text to lowercase is one of the simplest and most effective forms of text preprocessing. It applies to the majority of text mining problems and significantly improves output consistency. Automatic cancelation is removed as they are meaningless reviews produced by the platform itself. Stopwords such as “a” or “the” occur in abundance but are meaningless. They are eliminated from the text to place greater emphasis on the key points [24]. Lemmatization is used because each word has different forms, such as "run," "running," and "runs," which computers may interpret as three separate words. It is the process of reducing a word to its root form; consequently, these words are treated similarly, and the model learns that they can be utilized in similar contexts [25]. After preprocessing, there are 70,154 rows and 8 columns of data remaining. Among all 8 columns, the most useful columns are listing id, date, and review content.

3.2 Explorative data analysis

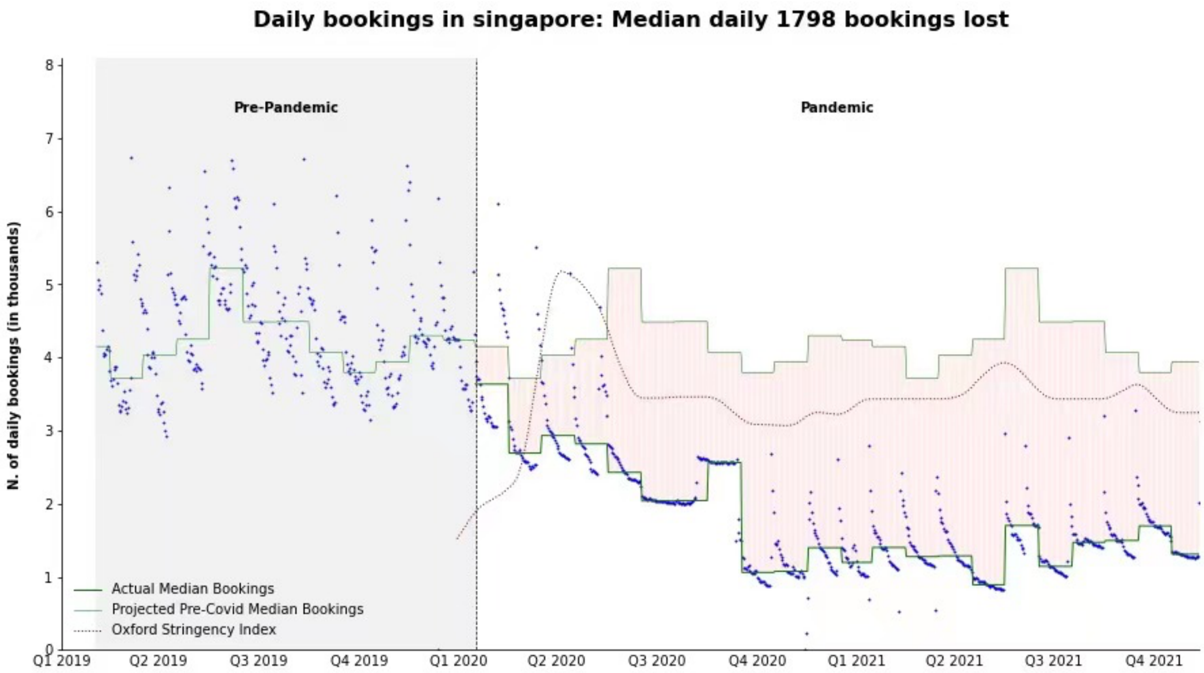

Figure 2. The comparison of daily bookings in Singapore during pre-pandemic and pandemic

The figure above indicates the change in daily bookings of Airbnb in Singapore before and after covid. COVID-19 started around the first quarter of 2020, highlighted in the figure with a black line, separating pre-covid and post-covid. The black-dot curve is the Oxford Stringency Index, which is a composite measure of the strictness of policy responses. The blue dots are the actual booking numbers of each Airbnb and the light green in the pre-pandemic section represents the median bookings before COVID-19. It is translated to the pandemic section, representing what the bookings are if no covid in the world. The dark green line in the pandemic section represented the actual median bookings after COVID-19. The pink area is the difference between the pre-covid bookings and the covid bookings. This area showed that the covid has significantly harmed Airbnb bookings, and there is a median daily 1,798 bookings lost. The figure helped to illustrate how covid negatively impacts Airbnb bookings.

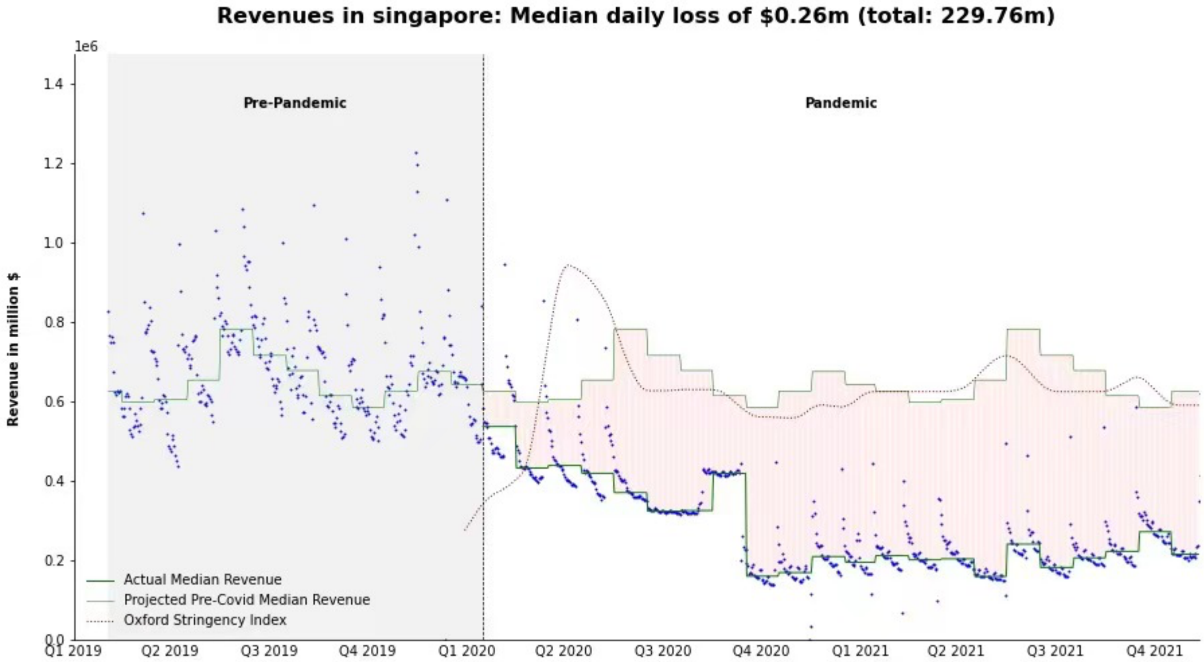

Figure 3. The comparison of Airbnb revenues in Singapore during pre-pandemic and pandemic

The figure represents the change in the revenues of Singapore Airbnb. The pink area between the light green line and the dark green line illustrated the difference between actual median revenue and projected pre-covid median revenue, showing the significant impact of COVID-19 on revenue. Due to covid, the median daily loss of revenue is $0.26 million and in total is $229.76 million.

These two figures confirmed that COVID-19 has a significantly negative impact on Airbnb bookings and revenues, which acted as the basis for analyzing consumer risk perception in the next section. It demonstrated that covid has an impact on Airbnb but did not illustrate how covid affects Airbnb. This is due to the change in consumer risk perception, which is explained in the next section.

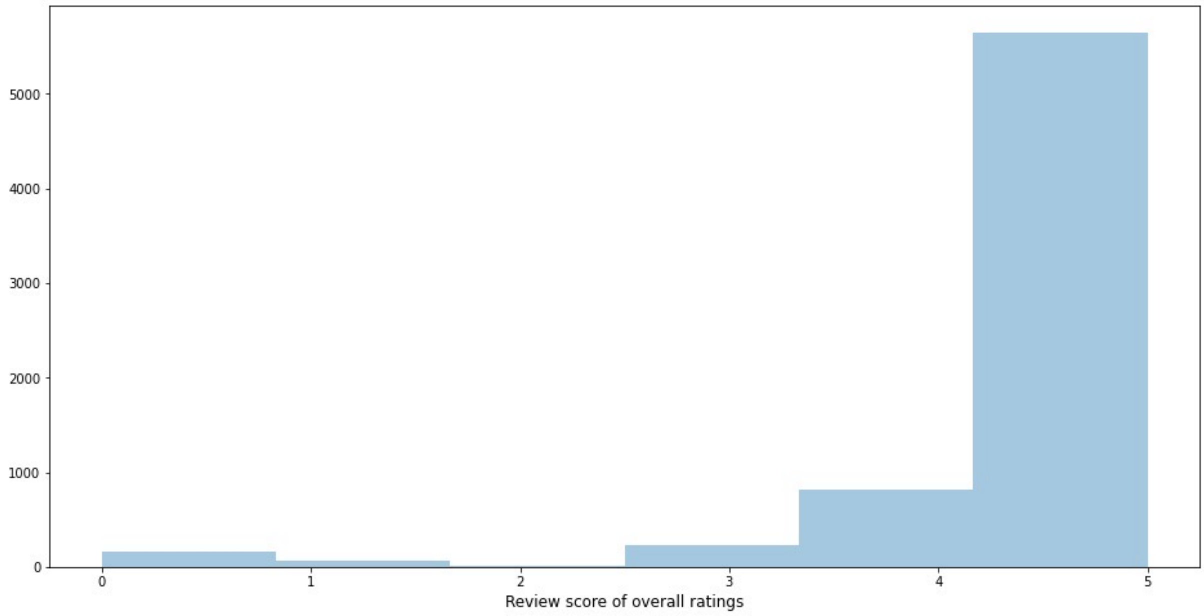

Figure 4. The distribution of the review scores of overall ratings

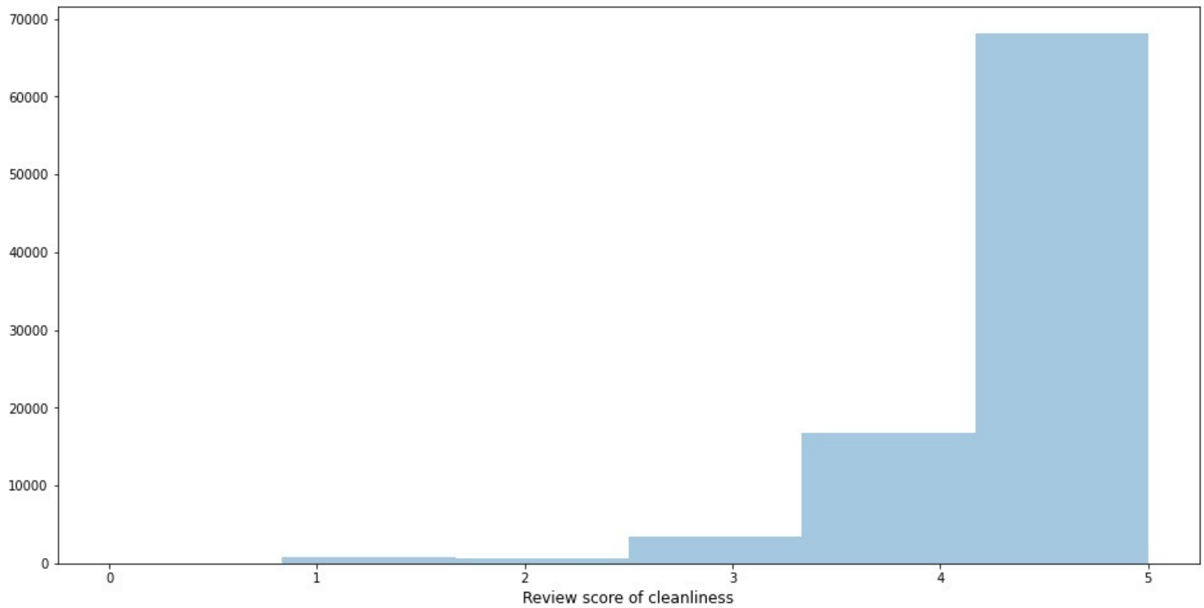

Figure 5. The distribution of the review score of the cleanliness rating

The two figures are the customer rating scores overall and the cleanliness rating, respectively. From the figure, ratings between 4 and 5 occupied more than 90% while the low ratings accounted for less than 10%, indicating that consumers are more likely to give high scores than low scores. Thus, the review content is increasingly important for analyzing the change in consumer risk perception.

4. Method & Result

4.1 Topic modeling

As mentioned in previous sections, there are more than 70,000 reviews after preprocessing data, which is impossible to analyze one by one. Therefore, reviews could be divided into certain topics first using topic modeling, and then analyzing the topics, reducing the workload significantly. Topic modeling is a method for analyzing large volumes of unlabeled text data. It helps in discovering hidden topical patterns that are present across the collection, annotating documents according to these topics, and using these annotations to organize, search and summarize texts. In this paper, topic modeling is done using the Gensim library with an LDA model for topic modeling [26]. Genism is used because it can convert large-scale data at a fast speed. Latent Dirichlet Allocation (LDA) is a popular form of statistical topic modeling [27]. Words are converted into a bag of words first, then apply LDA, which looks at a document to determine a set of topics that are likely to have generated that collection of words. So, if a document uses certain words that are contained in a topic, LDA would say that the document is about that topic [31]. All reviews are assigned to four topics: cleanliness, communication, location, and value. Topics are divided into four in total because the Airbnb platform rating system also includes ratings from these four topics; that is all other sub-areas can fall into these four topics.

The table below shows the four topics and the correspondence topic number and keywords within the topics. The numbers in front indicate the probability of the words belonging to the topic.

Table 2. Words within the four topics

Topic number	Topic	Words within the topic
0	Communication	‘0.021*"u" + 0.018*"check" + 0.017*"host" + 0.015*"apartment" + 0.012*"airbnb" + 0.009*"stay" + 0.008*"well" + 0.008*"time" + 0.008*"day" + 0.007*"also"’

1	Location	'0.040*"great" + 0.040*"place" + 0.033*"location" + 0.023*"mrt" + 0.020*"clean" + 0.019*"stay" + 0.017*"good" + 0.016*"apartment" + 0.014*"close" + 0.014*"nice"'
2	Cleanliness	'0.057*"place" + 0.042*"stay" + 0.028*"host" + 0.025*"nice" + 0.024*"good" + 0.021*"great" + 0.020*"clean" + 0.015*"location" + 0.015*"really" + 0.013*"would"'
3	Value	'0.033*"room" + 0.014*"bathroom" + 0.012*"bed" + 0.010*"place" + 0.010*"small" + 0.009*"good" + 0.008*"one" + 0.007*"night" + 0.007*"bit" + 0.007*"toilet"'

None of the reviews fully belongs to a certain topic, and thus problem arises in identifying reviews for certain topics. For example, a review may have a 30% probability to belong to topic 0, 20% belong to topic 1, 10% belong to topic 2, and 40% belong to topic 3. In such cases, the reviews would assign to the topics with the highest probability. This would then give the results as shown in the table below. The index lists are the number of reviews, and the topic probability indicates how likely the review belongs to the topic with the maximum probability.

Table 3. The topic probability and topic assignment of reviews

	Index list	Topic number	Topic probability
0	0	2	0.890421
1	1	1	0.876521
2	2	1	0.531334
3	3	2	0.620590
…	…	…	…

However, the problem is not fully resolved. Some reviews are assigned to a certain topic with a lower than 50% or even 40% probability, so those reviews may not belong to the topic assigned due to the low probability. Therefore, to guarantee accuracy in assigning the topics, the threshold is imposed on the topic assignment. The threshold act as a limit on the probability, so the probability should be greater than the threshold in order to have a review belonging to certain topics. Therefore, some uncertain reviews whose maximum probability is smaller than the threshold will be considered as not belonging to any topics, and thus will be removed. As the threshold increases, the topic assignment becomes more accurate, but the number of reviews successfully assigned decreases since an increasing number of reviews are removed. The thresholds used in the paper are from 50 to 90, increased by a tenth. Different thresholds are set to reconfirm the results. The table below summarized the number of reviews assigned to each topic at different thresholds. When the threshold is 50, meaning that the topic probability should be greater than 50% so that the reviews could be assigned to the topic, there are 58,827 reviews left. Topic 1 had the most reviews, which is 23,856, and topic 0 had the least review 3519 reviews. When the threshold is 60,46,741 reviews are left. When the threshold is 70,35,059 reviews are left. When the threshold is 80, there are 23,713 reviews left. For all thresholds from 60 to 90, topic 2 have the most reviews and topic 0 have the least number of reviews. The number of reviews decrease from threshold 50 to 90, but the accuracy of the topic assignment increased.

Table 4. The Topic probability and total review numbers for the four topics

Threshold	Topic 0	Topic 1	Topic 2	Topic 3	Total reviews
50	3,519	23,856	23,278	8,174	58,827
60	2,022	19,429	19,906	5,384	46,741
70	1,075	15,177	15,539	3,268	35,059
80	566	10,474	10,946	1,727	23,713

90	251	4,352	3,219	209	8,031

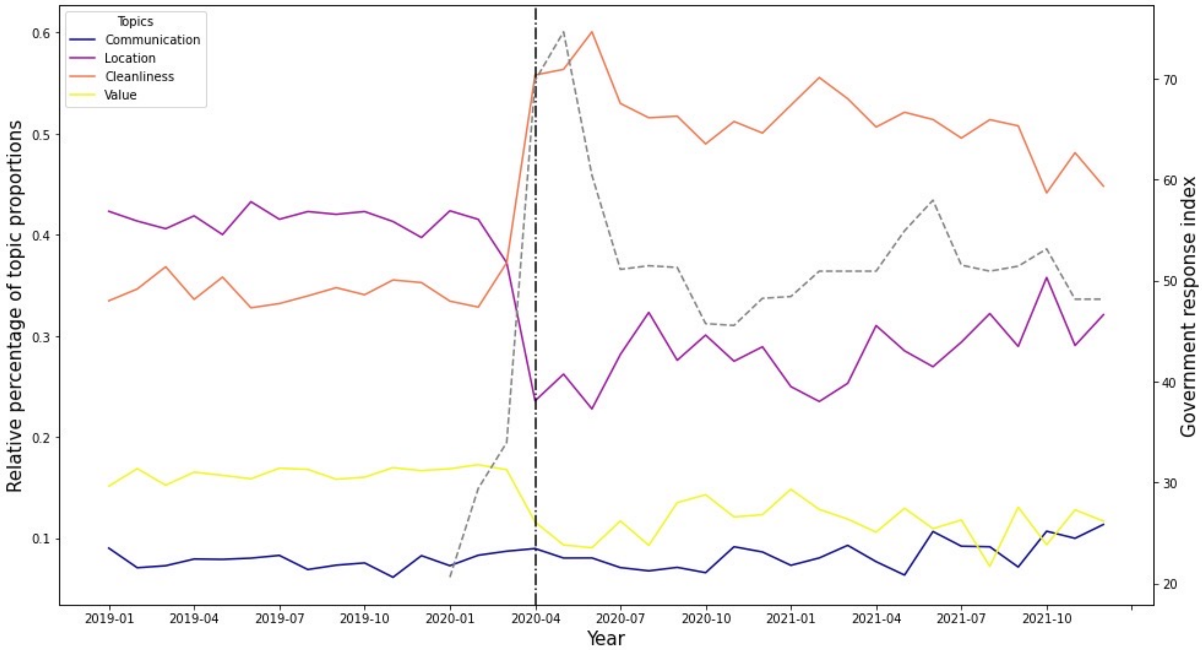

Figure 6. The topic distribution of four topics without threshold

The graph above has the topic probability of different topics on the y-axis and the date on the x-axis in order to compare whether the probability of the topics changed before and after covid. The black dot line indicates the time when covid starts, so it separates pre-covid and post-covid. The black dot curve is the Oxford Stringency Index representing the strictness of the government restriction. The graph above is without a threshold, which means that reviews are assigned to the topics with the maximum probability. The orange line represented topic 2, which is the cleanliness topic. It is the only topic that have a significant increase in probability and remained at the highest after covid existed, indicating that people discuss cleanliness more often, so they care more about this factor. In other words, consumers' risk perception of cleanliness increased significantly and remained at a high level after the pandemic. Additionally, the curve for topic 2 moves roughly in the same direction as the Oxford Stringency Index, which suggested that as the government responses to covid increase, the risk perception of cleanliness increases. Therefore, consumer risk perception of cleanliness is closely associated with the severity of covid.

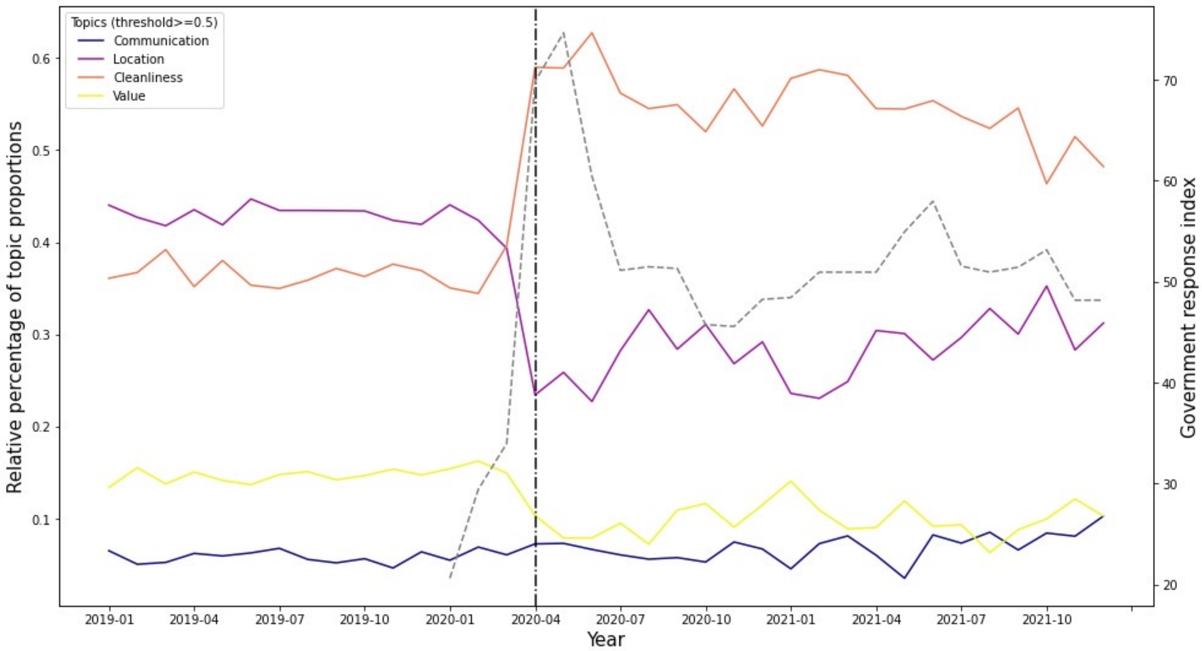

Figure 7. The topic probability of four topics when threshold is 0.5

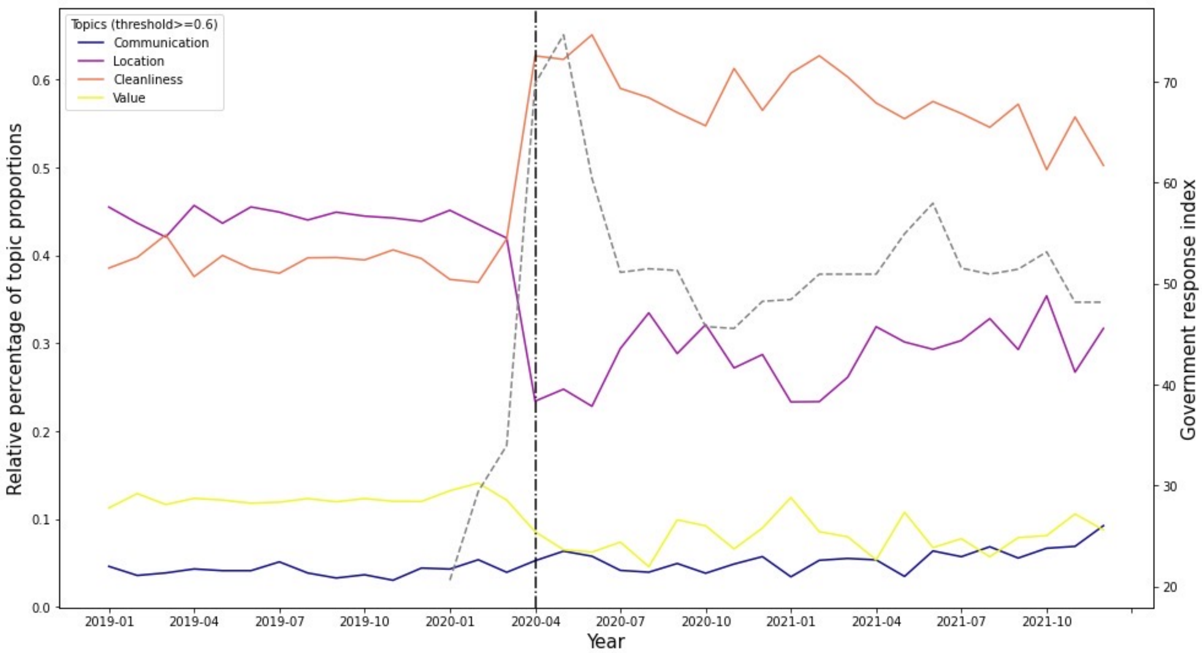

Figure 8. The topic probability of four topics when threshold is 0.6

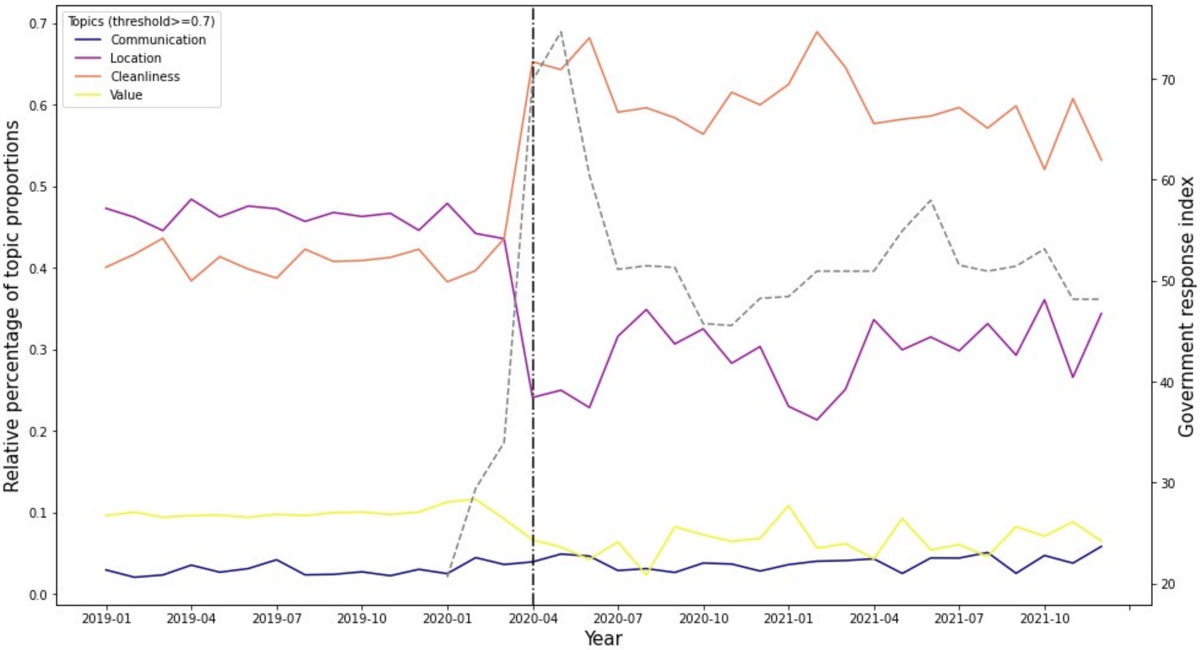

Figure 9. The topic probability of four topics when threshold is 0.7

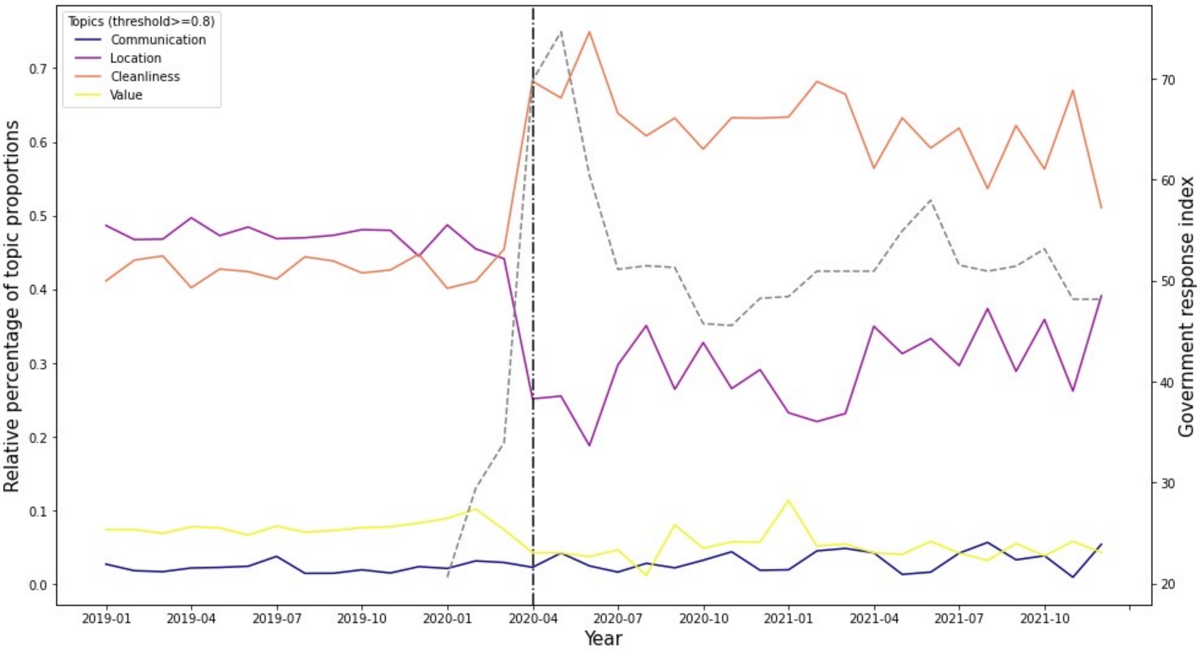

Figure 10. The topic probability of four topics when threshold is 0.8

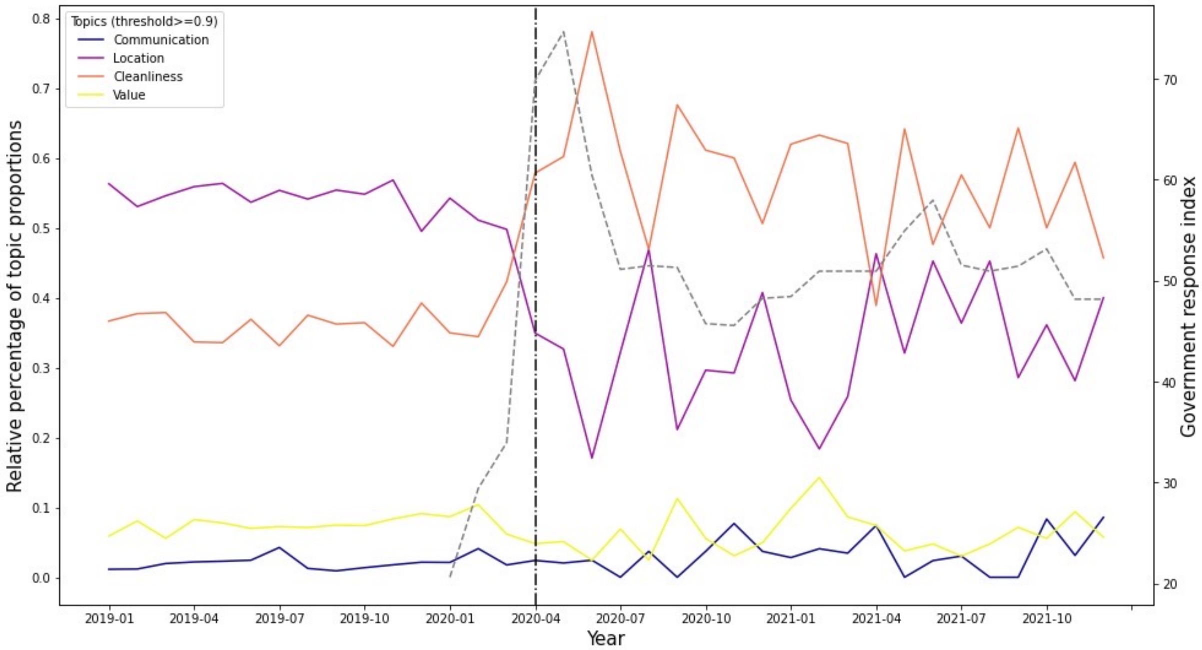

Figure 11. The topic probability of four topics when threshold is 0.9

The five figures above show the change of topic probability before and throughout the covid period with different thresholds from 50 to 90. The topic probability varies, but the trends are roughly the same. Topic 0 (communication) probability decreases significantly when the pandemic started, while topic 1 (location) and topic 3 (value) probabilities remained stable before and after COVID-19. However, the probability of topic 2 (cleanliness) rises greatly as COVID-19 existed in Singapore and remained at a high probability. The rise in topic probability indicated an increase in the number of discussions around cleanliness in consumer reviews, which suggested that consumer risk perception of cleanliness increased. Similar results that arose from different thresholds confirmed the fact that consumer risk perception towards cleanliness changed before and after the pandemic.

4.2 Sentiment Analysis

From the results in the last section, we could draw the conclusion that the probability of cleanliness topic increases after a pandemic occurred, but whether the reviews are positive or negative is unknown. Therefore, sentiment analysis is used to determine the positivity and negativity of the reviews under cleanliness topics and observed the change before and after COVID-19. Sentiment analysis is a process of extracting and understanding the sentiments defined in the text document [28]. It is an analytical technique that uses statistics, natural language processing, and machine learning to determine the emotional meaning of communications [29]. The NLP with artificial intelligence capability and text analytics are used to determine whether the sentiment of the opinion is positive, negative, or neutral [28]. This technique is used to determine whether the reviews are positive or negative, which revealed whether consumers are satisfied with cleanliness or unsatisfied with cleanliness in Airbnb. Sentiment analysis is done using VADAR Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media [30].

The VADER would provide a score as the result of the sentiment analysis. After emerged the results into the data, a table with a shape of 70,154 rows and 14 columns as shown below is obtained.

	listing_id	id	date	reviewer_id	reviewer_name	comments_x	year	clean_text	topic_number	topic_probability	month_year	comments_y	sentiment	sentiment_category
0	71609.0	496082140	2019-07-27	220147487.0	John	Attractive place to stay for pax when visit Si...	2019.0	attractive place stay pax visit singapore	2.0	0.890421	2019-07	Attractive place to stay for 4pax when visit S...	0.4404	positive
1	71609.0	524932232	2019-09-07	2898879.0	Jeni	The location is very close to EXPO and Bus sto...	2019.0	location close expo bus stop mrt station conve...	1.0	0.876521	2019-09	The location is very close to EXPO and Bus sto...	-0.1531	negative
2	71609.0	571338964	2019-12-01	287700916.0	Eigen John	Outstanding value for money The place is near...	2019.0	outstanding value money place near changi airp...	1.0	0.531334	2019-12	Outstanding value for money. The place is near...	0.8146	positive
3	71609.0	573744808	2019-12-07	312920122.0	Jean-Philippe	Really hepful with everything Answered really...	2019.0	really hepful everything answered really gast ...	2.0	0.620590	2019-12	Really hepful with everything. Answered really...	0.3612	positive
4	71609.0	581140325	2019-12-24	96914254.0	Jade	Great location great hospitality	2019.0	great location great hospitality	1.0	0.576262	2019-12	Great location, great hospitality	0.8481	positive

Figure 12. Sample of the data with results of sentiment category and sentiment analysis score

The sentiment in the second right column is the compound score detected by VADER and the right column is the sentiment status given based on the score. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive) [30]. The threshold for dividing reviews into different categories is 0.05. Reviews with a compound score greater or equal to 0.05 are considered positive, while those that are less or equal to -0.05 are negative reviews. Neutral sentiment reviews are those between -0.05 and 0.05. The table below shows the number of positive and negative reviews on each topic.

Table 5. The number of reviews for each sentiment category in each topic

Topic number	Sentiment Category	Number of reviews
0	Negative	722
	Neutral	381
	Positive	3,922
1	Negative	175
	Neutral	956
	Positive	25,212
2	Negative	165
	Neutral	660
	Positive	24,098
3	Negative	1,367
	Neutral	559
	Positive	8,115

Since the Airbnb rating from the listing data is also an indication of sentiment analysis but from a numerical perspective. I decided to compare the results of the sentiment analysis and the rating for each topic and see if it corresponds with the sentiment analysis. Theoretically, the positive listing of the topic should have a higher score and the negative listing should have a lower score. The results of the average reviews from positive listings and negative listings and the difference between the average ratings of the two listings are shown in the table below.

Table 6. The average rating for positive and negative listing and its difference for each topic

Topic	Average rating for positive listing	Average rating for negative listing	Difference in rating
communication	4.876	4.755	0.121

location	4.798	4.652	0.146
cleanliness	4.733	4.501	0.232
value	4.671	4.511	0.160

From the results, the positive listing of all four topics, which is Airbnb that consumers perceived as good in those aspects, received higher average ratings than the negative listing which consumers perceived did poorly in the four aspects. Therefore, the results of the score correspond with the sentiment analysis for customer reviews. The two indications of sentiment analysis displayed the same pattern. Additionally, the cleanliness topic has the greatest difference in average ratings between positive listing and negative listing compared with the other three topics, meaning that consumers care about cleanliness more than other topics. So, if an Airbnb host have done poorly in cleanliness, consumers are more likely to give low ratings to the hosts than if they did poorly in communication, location, or value. This again confirmed the fact that consumers have a risk perception towards cleanliness, so they would react stronger on issues related to cleanliness.

After doing sentiment analysis for the customer reviews, we find out that the positive listing tends to have a higher rating as well. However, in order to discover whether consumer risk perception has an impact on business performance, we should compare sentiment analysis results with Airbnb booking numbers and revenue. If the consumer risk perception has an impact on business performance, then Airbnb bookings and revenue are expected to change accordingly. Specifically, when consumers perceived Airbnb as clean, then Airbnb is supposed to have more bookings and generate more revenue; while if Airbnb is perceived as not clean, then it should have fewer bookings and less revenue. This trend is expected to increase during the COVID-19 period in order to illustrate the fact that COVID-19 changed consumer risk perception of cleanliness. Therefore, the columns of booking number and revenue are extracted from the calendar data and emerged into the Airbnb data with sentiment analysis results on it.

Airbnb should be divided into positive listings and negative listings based on their perceived cleanliness, which is Airbnb that is perceived as clean and Airbnb that is not. The threshold of dividing is 50%. The top 50% of hosts with high sentiment analysis scores become the positive listing while the other 50% of hosts are in the negative listing. The statistical results of the sentiment analysis compound score are found. The mean score of the whole list is 0.690. The maximum score and minimum score are 0.999 and -0.995, respectively. The 50% line lies at the score of 0.781. Therefore, hosts with a compound score higher than 0.78 are in the positive listing and are those that consumers perceived as clean rooms. Conversely, hosts with compound scores lower than 0.78 are in the negative listing, as they are perceived as not clean by consumers. Then, a graph is made combining the revenue and booking results of positive and negative listing. As shown in the figure below, the x-axis is the date and the y-axis represents the number of bookings. The red curve is the positive listing, and the black curve is the negative listing. The figure revealed the relationship between consumer risk perception and the number of bookings for Airbnb. As mentioned in the earlier section, COVID-19 started in January 2020, so before that time is pre-pandemic, and after is pandemic. Broadly speaking from the figure, the number of bookings for both positive listings and the negative listing has a significant decrease from pre-pandemic to the pandemic, showing a negative impact of COVID-19 on Airbnb booking numbers as a whole. Before the pandemic, the positive listing and the negative listing have around the same number of bookings with nearly no difference. This indicated that consumers do not have a risk perception towards cleanliness as their choices are not influenced by cleanliness. However, after the pandemic, the difference between positive and negative listing increased greatly, especially from December 2020 to May 2021, when the difference reached its maximum. Positive listings have a lot more bookings than negative listings, which demonstrated that Airbnbs that are perceived as cleaner by consumers would have more bookings. Consumers' risk perception towards cleanliness would make them more likely to choose Airbnb which is said to be clean by other customers in the customer reviews, reducing the risk of infection. The only exception is from August 2020 to November 2020 when the negative listing has more bookings than the positive listing. This phenomenon could be due to the that consumers have higher requirements for cleanliness, so the rooms that are originally perceived as clean are no longer clean anymore to new customers. Therefore, they are moved to the negative listing, but the booking numbers are not reduced at the moment.

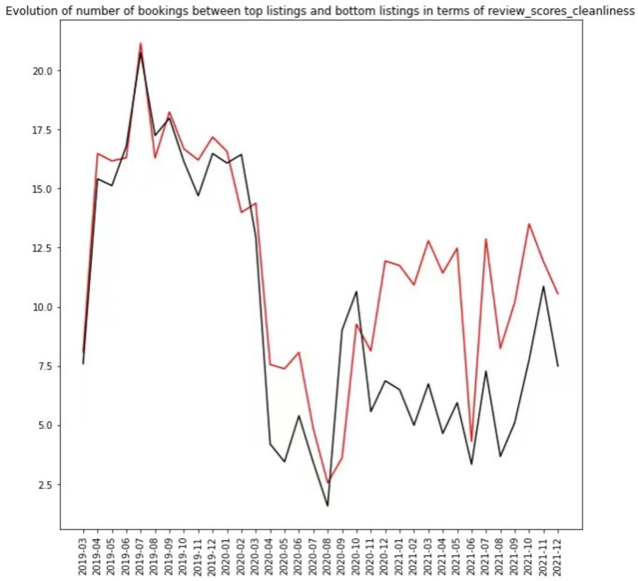

Figure 13. The evolution number of bookings between top and bottom listings of cleanliness

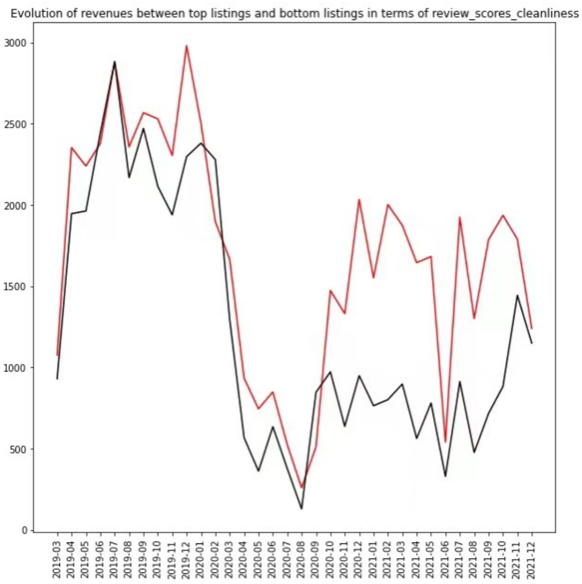

Figure 14. The evolution of revenues between top and bottom listings of cleanliness

The figures above show the difference in revenue between the positive listing and the negative listing. The x-axis represents the date while the y-axis represents the revenue. The red curve represents the 50% of Airbnb rooms customers viewed as clean while the black curve represents the 50% of Airbnb rooms that customers viewed as not clean. Following the same trend with the numbers of booking figure, before the pandemic (before January 2020), the difference between the red curve and black curve is small, suggesting that the clean Airbnb and the not clean Airbnb makes the same amount of revenue. Therefore, it indicated that cleanliness is not the main factor for customers when choosing which Airbnb to live in; in other words, consumer risk perception is not towards cleanliness before the pandemic. However, after the pandemic occurred, the overall revenue for Airbnb fell. The difference between positive listing and negative listing widens as the duration of the pandemic increased, reaching its maximum at the end of 2020 and the start of 2021. Positive listings nearly all the time made more profits than negative listings. The difference demonstrated that consumers are more likely to choose Airbnb which is perceived as clean over the others, so the hosts in positive listings make more revenue. Thus, the pandemic changed consumer risk perception to cleanliness, which would play a great role in consumer decisions. Due to the shift in consumer risk perception, cleanliness is essential for hospitality business performance as hosts with a clean room make more profits than hosts with no clean room.

5. Implication

The results from section three showed that the topic probability of topic 2 (cleanliness) increased significantly after COVID-19 occurred at all different thresholds, while other topics decreased, suggesting that risk perception changed greatly in COVID-19. COVID-19 shifted consumer risk perception towards cleanliness. People cared more about cleanliness due to COVID-19 since they are afraid of being infected or infecting others. The comparison between positive listing and negative listing suggested that hosts in the positive listing, consumers perceived as clean, have more bookings, and gained more revenue than hosts in the negative listing, which consumers perceived as not clean. Airbnb hosts with high cleanliness risks make fewer profits. Consumer risk perception of cleanliness is the main factor influencing consumer decision and thus would affect business performance. Therefore, noticing the change in consumer risk perception and reacting to it would help Airbnb hosts make more profits.

According to the results, the hosts should make a greater effort on maintaining the cleanliness of the room as consumer risk perception shifts toward cleanliness. In terms of cleaning, every host should sweep, vacuum, dust, or mop areas before sanitizing to make sure there's no dust on furniture or the floor so that consumers are more likely to perceive it as clean. Targeting COVID-19 prevention, hosts should wash dishes and laundry at the highest heat setting possible so that any bacteria left from the previous customers are completely dead. Hosts should wipe down all surfaces with disinfectants. In order to inform consumers about the cleanliness of the room, they should state on their website their cleaning process and the results of it, guaranteeing to consumers that the rooms are clean and safe. Also, the hosts should create a non-contact check-in process by putting the key in certain locations and avoiding a large crowd staying in the same place, which would allow consumers to feel safe and reduce the possibility of infection. They should emphasize advertising cleanliness more so that more customers can be attracted.

As for the Airbnb platform, they should make a strict guideline or handbook for every Airbnb host on the cleaning process, including how to clean, what to use, and what to achieve, so that customers know that Airbnb has a unified and strict requirement on cleanliness, lowering the uncertainty and risk. They could provide a list of disinfectants for Airbnb hosts to use against COVID-19 that are approved by the local regulatory agencies. Airbnb platform could make a statement on the front page of the website stating that all hosts have followed the cleaning handbook and used effective disinfectants. Additionally, the platform could also use certain symbols or identification to help consumers differentiate between hosts that are clean and are not clean. They could also add a more-detailed rating system for customers to refer to and make their decisions, providing more information to customers on cleanliness. Overall, cleanliness should be essential in advertising.

In terms of the Singapore government, they should reinforce the COVID-19 recovery strategy as the number of bookings and revenues dropped after COVID-19, meaning that most customers still avoid traveling in the presence of COVID-19. Therefore, an effective recovery strategy would enhance the COVID-19 situation in Singapore, reducing the risk of COVID-19, and more people are willing to travel. This will generate more revenue for the hospitality industry, and thus national GDP. Singapore government could also reinforce Airbnb inspection by setting stricter requirements and increasing the frequency of inspection so that the sanitary condition of Airbnb can be guaranteed. This way, customers would perceive less risk and go to Airbnb more often.

6. Conclusion

In conclusion, COVID-19 negatively impacted the hospitality industry's performance and shifted consumer risk perception toward cleanliness, which is shown by the increase in the probability of cleanliness in customer reviews before and after COVID-19. Additionally, sentiment analysis showed that consumer risk perception would affect the business performance of Airbnb. Combining sentiment analysis with calendar data, the results suggested that the listings that are perceived as clean would have more bookings and revenues than the listings that are perceived as not clean. The difference between the two listings increases after COVID-19, indicating cleanliness is the essential factor influencing customer risk perception. Customers are more willing to choose clean places over not clean hosts.

With these results, the Airbnb hosts and platform can reinforce the requirements on cleanliness and put greater emphasis on the advertisement of cleanliness, creating a clean environment for customers. The platform can also set unified guidelines for all hosts to follow to maintain similar qualities. In addition, the Singapore government could work on a recovery strategy, lowering the threats of COVID-19 to the tourism industry, as well as enhancing Airbnb speculation to supervise and monitor the cleanliness of Airbnb, generating more revenues and thus GDP for the country.

This paper uses empirical analysis to provide the mechanism that explains the change in customer choices in selecting Airbnb hosts after the pandemic using the risk perception framework.

However, the researchers have several limitations. Firstly, the data is only for Airbnb customer reviews, and thus cannot be generalized to all hospitality services. Since Airbnb and hotels have great differences in terms of price, management style, and qualities. Therefore, the applicable range for the results has limitations. In addition, the data only target Singapore Airbnb, so it cannot represent globally as a whole. Since different countries have different markets for Airbnb with different competitors and regulations, the strategy and environment for Singapore Airbnb may differ from other countries, so the results cannot be fully generalized.

In the future, the research could extend to the whole hospitality industry, analyzing not only Airbnb customer risk perception but also hotel guests' risk perception. Additionally, the research could be done across other places, using data from western countries to generalize the results globally.

References

[1] Schlagwein, Daniel, et al. "Consolidated, Systemic Conceptualization, and Definition of the 'Sharing Economy.'" Jasist, vol. 71, no. 7, pp. 817-38, asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24300. Accessed 4 Aug.2022.

[2] Deane, Steve. "2022 Airbnb Statistics: Usage, Demographics, and Revenue Growth." Stratos, 4 Jan. 2022, www.stratosjets.com/blog/airbnb-statistics/. Accessed 5 Aug. 2022.

[3] Pearson, Christine M., and Judith A. Clair. "Reframing Crisis Management." The Academy of Management Review, vol. 23, no. 1, 1998, pp. 59-76. JSTOR, https://doi.org/10.2307/259099. Accessed 22 Aug. 2022.

[4] Bhati, A., Upadhayaya, A., & Sharma, A. (2016). National disaster management in the ASEAN-5: An analysis of tourism resilience. Tourism Review, 71, 148–164.

[5] Rosselló, Jaume, et al. "The Effects of Natural Disasters on International Tourism: A Global Analysis." Tourism Management.

[6] Moreira, Pedro. "Stealth Risks and Catastrophic Risks." Journal of Travel and Tourism Marketing, vol. 23, nos. 2-4, Jan. 2008, pp. 15-27, https://doi.org/10.1300/J073v23n02_02. Accessed 23 Aug. 2022.

[7] Bresciani, Stefano, et al. "The Seven Lives of Airbnb. The Role of Accommodation Types." Annals of Tourism Research, vol. 88, May 2021, www.sciencedirect.com/science/article/abs/pii/S0160738321000323. Accessed 10 Aug. 2022.

[8] Kozak, M., Crotts, J. C., & Law, R. (2007). The impact of the perception of risk on international travellers. International Journal of Tourism Research, 9, 233–242.

[9] Sönmez, S., Apostolopoulos, Y., & Tarlow, P. (1999). Tourism in crisis: Managing the effects of terrorism. Journal of Travel Research, 38, 13–18.

[10] Foley, M., & Lennon, J. (1996). JFK and dark tourism: Heart of darkness. Journal of International Heritage Studies, 2(2), 195–197.

[11] Dolnicar, Sara, and Samira Zare. "COVID19 and Airbnb – Disrupting the Disruptor." Annals of Tourism Research, vol. 83, July 2020, p. 102961, https://doi.org/10.1016/j.annals.2020.102961. Accessed 26 Aug. 2022.

[12] Škare, Marinko, et al. "Impact of COVID-19 on the Travel and Tourism Industry."Technological Forecasting and Social Change, vol. 163, Feb. 2021, p.120469, https://doi.org/10.1016/j.techfore.2020.120469. Accessed 23 Aug.2022.

[13] Ren, X. 2020. Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States. Eurasian Geography and Economics 1–12. Doi: 10.1080/15387216.2020.1762103

[14] Kuo, Hsiao-i, et al. "Estimating the Impact of Avian Flu on International Tourism Demand Using Panel Data." Tourism Economics, vol. 15, no. 3, Sept. 2009, pp. 501-11, https://doi.org/10.5367/000000009789036611. Accessed 26 Aug. 2022.

[15] Mao, C.-K., Ding, C.G. and Lee, H.-Y. 2010. Post-SARS tourist arrival recovery patterns: An analysis based on a catastrophe theory. Tourism Management 31. (6): 855–861. Doi: 10.1016/j.tourman.2009.09.003

[16] Lee, C.-K., Song, H.-J., Bendle, L.J., Kim, M.-J. and Han, H. 2012. The impact of non-pharmaceutical interventions for 2009 H1N1 influenza on travel intentions: A model of goal-directed behavior. Tourism Management 33. (1): 89–99. Doi: 10.1016/j. tourman.2011.02.006

[17] Joo, Heesoo, et al. "Economic Impact of the 2015 MERS Outbreak on the Republic of Korea's Tourism-Related Industries." Health Security, vol. 17, no. 2, Apr. 2019, pp. 100-08, https://doi.org/10.1089/hs.2018.0115. Accessed 26 Aug. 2022.

[18] Davis, Xiaohong M., et al. "International Travelers as Sentinels for Sustained Influenza Transmission during the 2009 Influenza A (H1N1) pdm09 Pandemic." Journal of Travel Medicine, vol. 20, no. 3, 1 May 2013, pp. 177-84, https://doi.org/10.1111/jtm.12025. Accessed 26 Aug. 2022.

[19] Omrani, A.S. and Shalhoub, S. 2015. Middle East respiratory syndrome coronavirus (MERS- CoV): what lessons can we learn? Journal of Hospital Infection 91. (3): 188–196. Doi: 10.1016/j. jhin.2015.08.002

[20] Findlater, Aidan, and Isaac I. Bogoch. "Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel." Trends in Parasitology, vol. 34, no. 9, Sept. 2018, pp. 772-83, https://doi.org/10.1016/j.pt.2018.07.004. Accessed 26 Aug. 2022.

[21] Boros, Lajos, et al. "The Effects of COVID-19 on Airbnb." Hungarian Geographical Bulletin, vol. 69, no. 4, 22 Dec. 2020, pp. 363-81, https://doi.org/10.15201/hungeobull.69.4.3. Accessed 26 Aug. 2022.

[22] Bernstein, P. (1996). Against the gods: The remarkable story of risk. New York: John Wiley.

[23] Khanna, Chetna. "Text Pre-processing: Stop Words Removal Using Different Libraries." Towards Data Science, 10 Feb. 2021, medium.com/towards-data-science/ text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a.Accessed 17 Aug. 2022.

[24] Lang, Niklas. "Stemming Vs. Lemmatization in NLP." Towards Data Science, Feb.2019, towardsdatascience.com/stemming-vs-lemmatization-in-nlp-dea008600a0.Accessed 17 Aug. 2022.

[25] Řehůřek, Radim & Sojka, Petr. (2010). Software Framework for Topic Modelling with Large Corpora. 45-50. 10.13140/2.1.2393.1847.

[26] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3. Jan (2003): 993-1022.

[27] Shivaprasad, T. K., and Jyothi Shetty. "Sentiment Analysis of Product Reviews: A Review." 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Piscataway, IEEE, 2017, pp. 298-301, https://doi.org/10.1109/ICICCT.2017.7975207. Accessed 18 Aug. 2022.

[28] Korolov, Maria. "What Is Sentiment Analysis? Using NLP and ML to Extract Meaning." CIO, 9 Sept. 2021, www.cio.com/article/189218/ what-is-sentiment-analysis-using-nlp-and-ml-to-extract-meaning.html. Accessed 18 Aug. 2022.

[29] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[30] BRIGL, Tobias, 2019, Extracting Reliable Topics using Ensemble Latent Dirichlet Allocation [Bachelor Thesis]. Technische Hochschule Ingolstadt. Munich: Data Reply GmbH. Available from: https://www.sezanzeb.de/machine_learning/ensemble_LDA/