MASTERARBEIT

A look into Covid-19 spreading through the lodging sector

zur Erlangung des akademischen Grades

Master of Science

im Rahmen des Studiums

Cartography

eingereicht von

Estefania Ruiz Martinez

Matrikelnummer 11939551

ausgeführt am Institut für Geodäsie und Geoinformation der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien

Betreuung

Betreuer: M.Sc. Francisco Porras Bernárdez (TUW)

Vienna, 10.10.2021

Unterschrift (Verfasserin)

Unterschrift (Betreuer)

MASTER THESIS

A look into Covid-19 spreading through the lodging sector

For the Achievement of the Academic Title

Master of Science

Within the Degree Course

Cartography

Submitted By

Estefania Ruiz Martinez

Student ID 11939551

Completed at the Department of Geodesy and Geoinformation of the Faculty for Mathematics and Geoinformation at the Technical University Vienna

Supervision

Supervisor: M.Sc. Francisco Porras Bernárdez (TUW)

Vienna, 10.10.2021

Signature (Author)　　Signature (Supervisor)

Cartography M.Sc.

Master thesis

A look into Covid-19 spreading through the lodging sector

Estefania Ruiz Martinez

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

ITC UNIVERSITY OF TWENTE.

2021

A look into Covid-19 spreading through the lodging sector

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich

Author:	Estefania, Ruiz Martinez
Study course:	Cartography M.Sc.
Supervisor:	M.Sc. Francisco Porras Bernárdez (TUW)
Reviewer:	Ph.D. Mahdi Farnaghi (UT-ITC)
Cooperation:	if applicable

Chair of the Thesis
Assessment Board: Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner,TUW

Date of submission: 10.10.2021

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

A look into Covid-19 spreading through the lodging sector

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 10.10.2021 Estefania, Ruiz Martinez

To Hernando

Acknowledgment

First of all, I would like to thank M.Sc. Francisco Porras Bernárdez for his continuous guidance and his incredible amount and support during this work. I would like to thank Ph.D. Mahdi Farnaghi for the helpful feedback he provided.

I would like to thank all my Professors from TUM, TUW, TUD and UT-ITC for teaching me new concepts to get a proper understanding of the field of Cartography. Specially, I would like to thank Juliane Cron, for all the overall support.

Also, I would like to thank all my classmates for their encouragement during these years.

Finally, I would like to thank my family in Colombia. The distance was never a barrier for getting all the best motivation and love that were crucial for this achievement.

Abstract

This thesis integrates spatial analysis and text mining techniques to explore the experience of Airbnb users in the light of the health crisis of COVID-19. The main findings did not provide enough evidence to claim that after the outbreak of covid there was a significant change in the experience of Airbnb users in Rio de Janeiro and New York. On the one hand, the decrease of positive reviews was not significatively high, as the decrease was only 1% in Rio de Janeiro and 2% in New York. On the other hand, positive reviews containing covid terms accounted 87% of the covid reviews in Rio de Janeiro and 89% of the covid reviews in New York. Nevertheless, the analysis of reviews containing covid terms revealed topics related with the health crisis, such as the use of mask and hand sanitizer, cancellation of reservation and flight, and travel ban. Further analyses are required to prove if those situations had an influence in the experience of Airbnb users. The overall findings show how semantic and spatial analysis can contribute to the understanding of customer behaviour in the lodging sector, and furthermore during health crisis events.

Table of Contents

Acknowledgment i
Abstract ii
List of figures iv
List of tables v
List of maps vi
1 Introduction 1
1.1 Research objective and research question 2
1.2 Thesis structure 2
2 Scientific background and related work 3
2.1 Traditional hotel accommodations 3
2.2 Peer-to-peer accommodations 5
2.3 Customer behaviour during health crises 6
2.4 Concluding remarks 6
3 Methodology 7
3.1 Data and data collection 7
3.2 Pre-processing 7
3.3 Classification 8
3.3.1 Sentiment polarity 8
3.3.2 Covid terms 9
3.4 Keyword analysis 10
3.4.1 Frequent terms identification 10
3.4.2 Rare terms identification 10
3.4.3 Network diagrams 10
3.5 Spatial analysis 11
4 Results and discussion 11
4.1 Analysis of sentiment polarity 11
4.1.1 Analysis of frequent keywords 26
4.1.2 Analysis of rare keywords 31
4.1.3 Analysis of network diagrams 35
4.2 Analysis of covid-reviews 40
4.2.1 Analysis of frequent keywords 42
4.2.2 Analysis of rare keywords 43
4.2.3 Analysis of network diagrams 44
5 Conclusions 47
References 49

List of figures

Figure 1. Techniques used for sentiment analysis, taken from Medhat et al. (2014) 8
Figure 2. Network diagram of positive reviews pre-covid RJ. 36
Figure 3. Network diagram of positive reviews post-covid RJ. 36
Figure 4. Network diagram of positive reviews pre-covid NY. 37
Figure 5. Network diagram of positive reviews post-covid NY. 37
Figure 6. Network diagram of neutral reviews pre-covid NY. 38
Figure 7. Network diagram of neutral reviews post-covid NY. 38
Figure 8. Network diagrams of negative reviews pre-covid from RJ. 39
Figure 9. Network diagrams of negative reviews post-covid from RJ. 39
Figure 10. Network diagrams of negative reviews pre-covid from NY. 40
Figure 11. Network diagrams of negative reviews post-covid from NY. 40
Figure 12. Network diagram of positive covid-reviews from RJ. 45
Figure 13. Network diagram of negative covid-reviews from RJ. 45
Figure 14. Network diagram of positive covid-reviews from NY. 46
Figure 15. Network diagram of negative covid-reviews from NY. 46

List of tables

Table 1. Reviews per sentiment polarity from NY and RJ before and after the outbreak of covid. 12
Table 2. Count and percentage of properties per sentiment polarity from RJ and NY in both time periods. 13
Table 3. Count and percentage of properties from RJ and NY that shifted of polarity after the outbreak of covid. 13
Table 4. List with the 10 most frequent terms from reviews of properties in RJ before the outbreak of covid. 26
Table 5. List with the 10 most frequent terms from reviews of properties in RJ after the outbreak of covid. 27
Table 6. List with the 10 most frequent terms from reviews of properties in NY before the outbreak of covid. 28
Table 7. List with the 10 most frequent terms from reviews of properties in NY after the outbreak of covid. 29
Table 8. List with the 10 most rare terms from reviews of properties in RJ before the outbreak of covid. 31
Table 9. List with the 10 most rare terms from reviews of properties in RJ after the outbreak of covid. 33
Table 10. List with the 10 most rare terms from reviews of properties in NY before the outbreak of covid. 33
Table 11. List with the 10 most rare terms from reviews of properties in NY after the outbreak of covid. 34
Table 12. Frequency of covid terms. 41
Table 13. Number and percentage of covid reviews per sentiment polarity from properties in RJ and NY. 41
Table 14. List with 10 most frequent terms from covid reviews of properties in RJ. 42
Table 15. List with 10 most frequent terms from covid reviews of properties in NY. 43
Table 16. List with 10 most rare terms from covid reviews of properties in RJ. 44
Table 17. List with 10 most rare terms from covid reviews of properties in NY. 44

List of maps

Map 1. Density distribution of properties with positive polarity in RJ before the outbreak of covid. 14
Map 2. Density distribution of properties with positive polarity in RJ after the outbreak of covid. 15
Map 3. Change in the density of properties with positive polarity in RJ after the outbreak of covid. 15
Map 4. Density distribution of properties with neutral polarity in RJ before the outbreak of covid. 16
Map 5. Density distribution of properties with neutral polarity in RJ after the outbreak of covid. 17
Map 6. Change in the density of properties with neutral polarity in RJ after the outbreak of covid. 17
Map 7. Density distribution of properties with negative polarity in RJ before the outbreak of covid. 18
Map 8. Density distribution of properties with negative polarity in RJ after the outbreak of covid. 19
Map 9. Change in the density of properties with negative polarity in RJ after the outbreak of covid. 19
Map 10. Density distribution of properties with positive polarity in NY before the outbreak of covid. 20
Map 11. Density distribution of properties with positive polarity in NY after the outbreak of covid. 21
Map 12. Change in the density of properties with positive polarity in NY after the outbreak of covid. 21
Map 13. Density distribution of properties with neutral polarity in NY before the outbreak of covid. 22
Map 14. Density distribution of properties with neutral polarity in NY after the outbreak of covid. 23
Map 15. Change in the density of properties with neutral polarity in NY after the outbreak of covid. 23
Map 16. Density distribution of properties with negative polarity in NY before the outbreak of covid. 24
Map 17. Density distribution of properties with negative polarity in NY after the outbreak of covid. 25
Map 18. Change in the density of properties with negative polarity in NY after the outbreak of covid. 25
Map 19. Spatial distribution of properties in RJ and NY where covid terms where mentioned 42

1 Introduction

For almost 4000 years the hospitality industry has been providing products and services not only to travellers and tourists, but also to local, regional, and national populations. It accounts together with the tourism industry for no more 8% of employment and 9-10% of the gross domestic product (GDP) worldwide (C.S. Siu, 2019; Wood, 2015).

As customers are a key stakeholder in the hospitality industry, a critical matter for this industry has to do with customer satisfaction. Customer satisfaction is defined as the difference between customer expectations and the actual experience (Hwang & Seo, 2016).Therefore, by understanding customer experience, managers can improve customer satisfaction. In fact, it is well established how customer satisfaction and customer experience can determine customer loyalty, repeated purchase and hence, increase of profit (Xiang et al., 2015).

In the lodging industry, Airbnb is a business with growing popularity that offers peer-to-peer accommodation services in more than 65,000 cities worldwide. It is based on a different economic model, known as the sharing economy or collaborative consumption. In this model "people coordinate the distribution of idle resources for a fee or other compensation". Previous works indicate that Airbnb costumers may be expecting a different experience if compared with traditional hotel accommodations (Tussyadiah & Zach, 2017).

Recently, online customer reviews have gained popularity among scholars for the study of customer behaviour (Xiang et al., 2015). Online reviews are part of the umbrella term, user generated content (UGC), which describes all types of media content (e.g., text, video) that are publicly available, created outside of professional context and with some creative effort (Kaplan & Haenlein, 2010). As the voice of the customer, online customer reviews are relevant for the study of customer behaviour, as they can reveal aspects of customer experience and in the lodging sector, favoured hotel attributes (Mirzaalian & Halpenny, 2019; Xiang et al., 2015).

Online reviews are by nature unstructured text, and on the other hand, the advent of the web 2.0 plus economic, technological, and social drivers have facilitated their fast increase in number (Kaplan & Haenlein, 2010). In this view, researchers are employing text mining approaches for their analysis. Popular approaches are sentiment analysis and topic modelling. The later aims to extract the polarity (positive or negative) of opinions and the topic associated with them (Liu, 2015a), while the former seeks to identify the semantic topic of a document collection by using machine learning algorithms (Hodson, 2017).

The current crisis generated by the COVID-19 pandemic has posed a challenge to the hospitality industry. Due to different measurements to deal with the spread of the virus and hygiene concerns by the population, many hospitality businesses shutdown and others experimented a decreased of demand (Gursoy & Chi, 2020). On the other hand, customer behaviour may also be affected, for instance, Hall et al. (2020) report that even months after finishing quarantine, some individuals still try to avoid public spaces or crowded places.

While academics have already started to explore the behaviour of customers during the health crisis of COVID-19, none of the studies found have done it with customers of P2P accommodations. This thesis attempts to do it by exploring the behaviour of Airbnb customers in two geographically contrasting cities but as well located in countries highly affect by the health crisis of COVID-19. The main contribution is the integration of spatial analysis with text mining techniques to explore other aspects of the customer experience.

1.1 Research objective and research question

This thesis aims **to explore the experience of Airbnb users after the outbreak of COVID-19**, and therefore:

- To classify online reviews according to the sentiment polarity.
- To classify online reviews according to the presence/absence of covid-terms.
- To analyse keywords and their relationship.
- To analyse the spatial distribution of property listings according to the sentiment polarity and the presence of covid-terms in property' reviews.

To fulfil those sub-objectives, this thesis should answer the following research questions:

- How is the experience of Airbnb users according to the sentiment polarity?
- How is the experience of Airbnb users according to the presence/absence of covid-terms?
- Where do Airbnb users experience positive, neutral, and negative sentiments?
- Where did Airbnb users mention covid-terms?

1.2 Thesis structure

The thesis consists of the following chapters:

- **Introduction:** this chapter presents the main topics addressed by this thesis and their relevance. It also introduces the research problem and the corresponding research objectives and questions, and it finishes by explaining the structure of this work.
- **Scientific background and related research:** this section summarizes the main findings and limitations of previous works that employed text mining techniques to study the customer behaviour in the lodging industry. It also explains the contributions of this thesis on this subject matter.

- **Methodology:** this chapter contains the steps followed to answer the research questions. It also describes the data and how it was collected and furthermore, it explains the text mining approaches and spatial analysis techniques used.
- **Results and discussion:** this section presents the results and their interpretation according to the research questions.
- **Conclusions:** this section answers the research questions by summarizing the main findings of this work, it also explains the main constrains of the analysis and how they could be overcome.

2 Scientific background and related work

This section examines the contribution of text mining approaches to the study of customer behaviour in the lodging industry. It first introduces leading works in the traditional hotel industry, then it provides main research in the share economy, followed by studies in the context of health crises and it finishes with some concluding remarks.

2.1 Traditional hotel accommodations

A leading work in this domain is from Li et al. (2013). They applied text mining approaches and content analysis on online reviews with the aim of identifying key factors that contribute to customer satisfaction and determining whether they differ between luxury hotels and budget hotels in China. The results show that, first, customers of luxury hotels and budget hotels share most of the determinants of satisfaction and second, that customers consider important and are more satisfied with factors such as value for money, transportation convenience, food-and-beverage (F&B) or convenience of tourist destination. However, the findings of this study cannot be generalized because the data is only from one city in China, and it does not account for non-response bias.

Geetha et al.'s (2017) work also compares luxury hotels with budget hotels. They used sentiment analysis to evaluate whether there is a relationship between costumers' sentiment polarity and customer ratings and whether that relationship is consistent across hotel categories. Their findings indicate that consistency between customer sentiment polarity and customer ratings is present in both types of hotel and moreover, that customer sentiments can explain their ratings for both types of hotels too. The authors explained the following limitations, first, the analysis did not address other factors that may affect the ratings, second the algorithm used for sentiment analysis does not account for context specific words and sentiments and last, the reviews were collected from only one website.

Phillips et al.'s (2020)work addressed other factors that may influence the rating of customers, such as geographic and psychic distance. By integrating sentiment analysis and text mining techniques they found a negative relationship between distance and customer ratings. The authors recognized that their findings cannot be generalized as their work only considers one country (Portugal) and because of language-related limitations, for instance, the analysis did not consider the language of all customers and as consequence, the analysis did not capture the different degree of expressive power of costumer' language and their cultural background.

Ban et al. (2019) provided a general perspective about key attributes and their relationships for customers of hotels around the world. By using frequency analysis, semantic network analysis, factor analysis and regression analysis they identified five groups of key attributes: "Access", "F&B", "Purpose", "Tangibles", and "Empathy". Among them, the most influential was "Empathy". However, the authors recognized that the findings cannot be generalized as it is only related to the top 25 hotels in the world, and furthermore, as the analysis is frequency-based, it is not possible to extract additional meaning from the words, therefore the need of sentiment analysis.

Similarly, Galati & Galati (2019) also contributed to the identification of key hotel attributes but from a cross-cultural perspective (Italy, U.S and China). The results from the text link analysis indicates that there are significant differences in the perception and emphasis of hotel attributes among the different nationalities. Although this study considers reviews in different languages, it has several limitations, some derived from the nature of the algorithm and others from the methodology.

Hu et al. (2019) analysed the reviews of hotels in New York to identify topics that appear more often in negative reviews than in positive ones and furthermore, to evaluate whether the occurrence of topics vary with the grades of hotel. They employed structural topic model (STM) to this end and found that, first, factors such as gender, age, travel type do not significantly affect the dissatisfaction analysis, and second, that customers of high-end hotels mainly complain about service, while customers of low-end hotels mostly complain about facility-related problems. The limitation of this study has to do with generalization, since the data was collected from one platform and one city, and the time range is just up to the end of 2013. Furthermore, the STM model could include other covariates to get more insight from the analysis.

Park et al. (2020) used sentiment analysis to understand the revisiting behaviour of customers of hotels in Korea. They observed that while the reviews of one-time visitor contain more anxious and analytical words, the reviews of re-visitors contain more words in a sentence, and they express more positive or negative sentiments. This study has some limitations, for instance, it only addresses reviews written in English, and thus opinions and perceptions, which are written more often in native language, cannot be captured.

Moro et al. (2020) also compared high-end vs low-end hotels but in the context of airport hotel chains. They aimed to identify hotel attributes relevant to hotel guests in five European cities, namely, Amsterdam, Brussels, Frankfurt, London, Paris, and to that end, they used text mining approaches and topic modelling. As a result, the authors found that customers perceived in a similar way the service offered by the staff in both types of hotel, and that they consider the food as a relevant service for both types of hotel too. Furthermore, they highlight the relevance of cleanliness, punctuality, and transportation facilities. Limitations were also discussed, for instance, the analysis only considered two hotel chains, and did not consider the particularities of guests' choices.

The last paper examined in this literature review explored the customer experience of heritage hotels in India (Chittiprolu et al., 2021). In this work, the authors sought to identify the factors driving satisfaction and dissatisfaction on customers by comparing the topics present on positive- and negative-rated reviews using unsupervised text mining techniques. They found that there are different factors causing satisfaction and dissatisfaction on customers, however some factors causing satisfaction may also cause dissatisfaction when they are not properly handled. Furthermore, they observed that the reviews of satisfied customers mentioned tangible features of the hotel stay, while the reviews of unsatisfied customers are more about intangible aspects of the hotel service, such as the attitude of the staff, service failure, value for money, among others. This work is limited to only one country and data source, and it did not consider the cultural background and demographics of the reviewers.

2.2 Peer-to-peer accommodations

Tussyadiah & Zach (2017) used clustering analysis to identify main service attributes of P2P accommodations, they corroborated that costumers that choose this type of accommodations are driven by experiential and social motivations, however, since negative reviews were not identified, they could not make conclusions about aspects that needed improvement by hosts.

In response to Tussyadiah & Zach's (2017) work, Cheng & Jin (2019) employed topic modelling and sentiment analysis to identify key attributes that influence the experience of Airbnb's users. They found that attributes used to evaluate the experience in traditional hotel accommodations are the same used to evaluate P2P accommodations. Furthermore, contrary to what is commonly claimed in the literature, the results of the study do not provide evidence for an authentic tourist-host interaction.

At the same time, Zhang's (2019) work compares the experience of Airbnb's customers with customers of the traditional hotel industry in the USA. He used sentiment analysis and topic modelling to that end. His findings reveal specific topics on Airbnb, such us 'late check-in', "help from host' and 'food in kitchen', among some others.

Moreover, Serrano et al. (2020) used topic modelling and sentiment analysis to explore the preferences and attitudes of green Airbnb users, the results indicate that there is a positive bias in the reviews and that ‘sustainability’ and ‘host’ play a positive role in the experience of green Airbnb users.

2.3 Customer behaviour during health crises

Few attempts have been made to explore the impact of health crises on hotel customer behaviour. Indeed, most of the existing literature explores this issue in the context of the COVID-19 pandemic. In this view, F. Hu et al. (2021) used term frequency and sentiment analysis to study the expectations and preferences of hotel customers in China during the different stages of the COVID-19 situation. Their findings reveal changes in the customers’ evaluations that might be long lasting, and that hygienic requirements are not the only concern of customers. This research is limited to a single data source and country, as well as to five-star hotels, to a lower number of reviews if compared to the pre-COVID-19 period, and to a time range that cannot cover all the stage “After COVID-19”.

Similarly, Mehta et al. (2021) conducted a global study using sentiment analysis and topic modelling to evaluate customer satisfaction. They observed that the causes of dissatisfaction among customers are related with the staff, room, cleanliness, service, slow booking process, and the pandemic response by hotel. They also found that the effect of the pandemic on customer satisfaction differed according to the country and month. The limitation of this work has to do with the temporal coverage of the analysis and with the low number of reviews during specific months.

Likewise, Escandon-Barbosa & Salas-Paramo (2021) also used topic modelling and sentiment analysis to evaluate the behaviour of luxury hotels customers in Mexico. Their findings indicate that factors such as protocols, facilities, cleanliness, time restriction, food, changes in behaviour pattern and smart tourism had a significant influence on customer experience during the pandemic, however this influence differed according to tourist type (family, business, couple, alone and friends).

2.4 Concluding remarks

This literature review shows how text mining approaches and NLP techniques has been used to leverage the large amount of online customer reviews to gain more insight about customer behaviour in the lodging industry. With this techniques scholars have been able to identify core attributes of customer experience, how those attributes might vary according to hotel type, how cultural background, distance, and customer profile influence the perception of hotel attributes, how ratings are related with customer sentiment polarity and moreover, how health crises affect customer behaviour. Furthermore, limitations were also presented, the common ones are related with language, cultural background, data sources and non-response bias. Though not a limitation, none of the works revised integrated spatial analysis with text mining techniques, and thus, this thesis attempts to do so, as an alternative to undercover other aspects of the customer behaviour.

3 Methodology

The analysis is divided in two parts, the first part dedicated to the analysis of sentiment polarity, the second one to the analysis of covid-related reviews. In the first part, reviews are classified according to the sentiment polarity and then, listings with reviews before and after the outbreak of covid are compared according to the sentiment polarity and topics. In the second part, reviews are further classified according to the absence/presence of covid terms and then the topics are extracted.

The focus of the analysis is on the experience of Airbnb users in Rio de Janeiro and New York, which are located in two of the most severely affected countries by COVID-19, namely, Brazil and U.S. Together with India, these countries are the top-3 countries with highest number of covid-cases reported. However, the selection of cities was according to data availability. The period of analysis covers one year after the outbreak of COVID-19, which the world health organization (WHO) declared on 31/12/2019 and one year before, in the case of the analysis of sentiment polarity. The following sections describe the data and data collection process, the process of pre-processing and the methods implemented for the analysis.

3.1 Data and data collection

Data was collected from the website Inside Airbnb, which provides information of the property listings, the reviews generated by users and the calendar availability for the next 365 days. Furthermore, the information of the property listings includes the geographic coordinates, which Airbnb anonymized by introducing an error between 0 and 150 meters. This information is of public domain dedication and readily available for more than 110 cities and up to several years (*Inside Airbnb*, 2021). After data was downloaded and processed, it was stored in a PostgreSQL database which was used further analysis.

3.2 Pre-processing

This step concerns about the preparation of data into a more structured and meaningful form appropriate for further analysis. In other words, raw text is turned into cleansed tokens. Mainly tasks in this process are unitization and tokenization, standardization and cleansing, stop word removal and stemming or lemmatization (Anandarajan et al., 2019; Sammut, 2017). All this tasks were implemented gradually, according to each of the steps of the analysis, and using the libraries available for text pre-processing in Python.

Before implementing the main tasks of pre-processing, automated postings (e.g., “This is an automated posting”), non-English, duplicated, empty reviews and reviews consisting only in two characters, numbers or NaN were discarded. Non-English reviews were identified using the python library Fasttext (Joulin et al., n.d., 2016), which categorizes texts according to the language.

Reviews were tokenized into single words and lowercased, stop words, emojis, proper nouns, special characters, numbers, punctuations, extra whitespaces, tabs, and newlines were removed. Furthermore, words were subjected to spelling correction and lemmatization, contractions were expanded and finally, words, adjectives and verbs were extracted.

3.3 Classification

Reviews were classified according to two criteria, sentiment polarity (positive, negative, neutral) and presence/absence of covid terms. The following subsections explain the classification process accordingly.

3.3.1 Sentiment polarity

Sentiment analysis (SA), also called opinion mining, is a field of study concerned about the extraction of positive or negative sentiments from natural language text and the targets of these sentiments (Liu, 2015a, 2015b). It can be also seen as a polarity classification problem, where a text can be classified as 'positive', 'negative' or 'neutral'. This classification problem can be taken to the document, sentence, or aspect level, where the aspect-level is the most meaningful one, as the classification problem is applied on the sentiment target (Alaei et al., 2019; Medhat et al., 2014).

There are two types of sentiment classification techniques, namely, machine learning and lexicon based, both with further subdivisions (see **Figure 1**). The machine learning approach considers the SA problem as a regular text classification problem and solves it using machine learning algorithms and/or linguistic features. On the other hand, lexicon-based approach classifies a text according to a collection of known and precompiled sentiment terms, also called sentiment lexicon (Medhat et al., 2014).

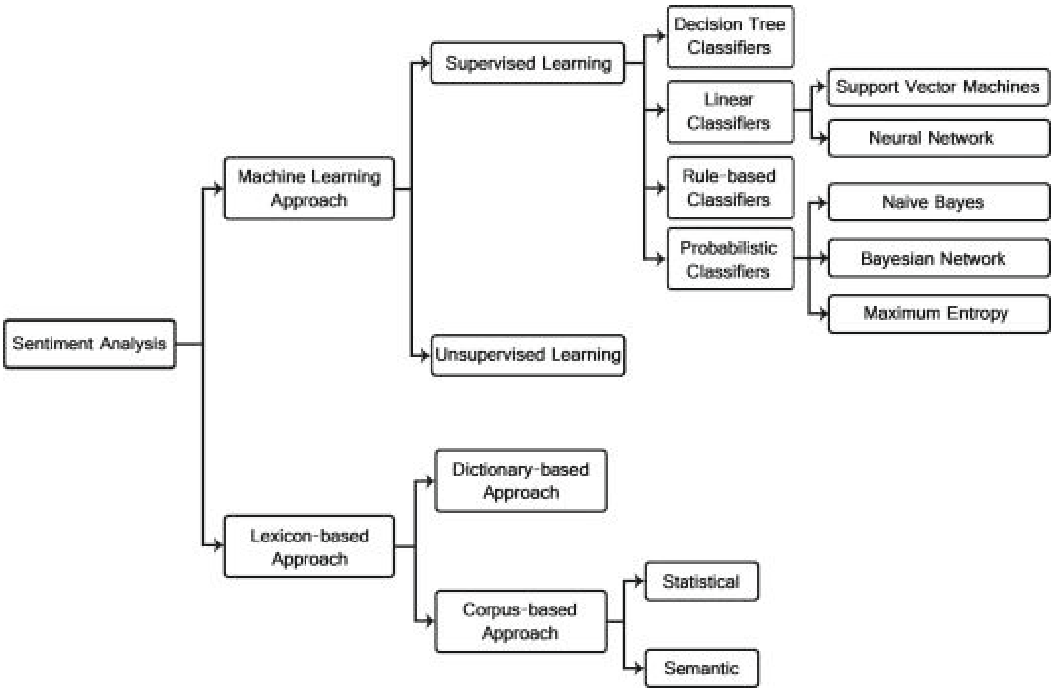


Figure 1. Techniques used for sentiment analysis, taken from Medhat et al. (2014)

This thesis deals with a lexicon-based and rule-based model, known as VADER, which stands for Valence Aware Dictionary for sEntiment Reasoning. This model is specially attuned to detect the polarity and intensity of sentiments in social media style text, yet it performs well across other domains if compared to other state-of-practice sentiment analysis tools. The sentiment lexicon used in this model combines well-established and human-validated sentiment lexicons (LIWC, ANEW, and GI) with lexical features used to express sentiments in social media text (e.g., emoticons, slangs, acronyms). Furthermore, it includes a series of heuristics that modify the intensity or the polarity of a sentence, such as, punctuations, degree modifiers, capitalization, negations, and conjunctions. Some of the advantages of this model are that it has been validated by humans and it does not require training data.

Python includes VADER as an independent library, which calculates a positive, negative, neutral, and compound score of a text. The compound score is the most frequently used and it results from adding the valence scores of each word and then normalizing it to be between -1 and +1. On the other hand, the positive, negative, or neutral scores represent the ratio of text that fall in each category (Hutto & Gilbert, 2014). This thesis uses the compound score, and for the classification the following thresholds:

- Positive review: compound score >= 0.001
- Neutral review: compound score > -0.001 and compound score < 0.001
- Negative review: compound score < = -0.001

3.3.2 Covid terms

Reviews were also classified according to the presence/absence of Covid-related terms. Covid-related terms can be of two types: context-specific, which means that those words appear very often in texts about the COVID-19 situation, such as, pandemic, mask, lockdown, quarantine. On the other hand, they can be words that people use to refer to the name of the virus or the disease and their abbreviations. Terms were collected mainly from scientific articles addressing sentiment analysis, text mining and linguistics to study aspects of the COVID-19 situation.

The most enriching source of covid-related terms was the study of Lillo (2020), which collected and categorized a total of 270 synonyms derived from a personally compiled corpus of tweets dating from late January to late May 2020. This collection contains synonyms of the standard terms for the coronavirus disease, COVID-19, and coronavirus, excluding figurative words and expressions, such as, the invisible enemy. These synonyms are often the result of slang, wordplay, verbal humour, bigotry or xenophobia, as examined by Lillo (2020).

Considering the scope of this work, the list of covid terms does not include open compound words and ambiguous words, however, hyphenated compound words were included but only the part before the first hyphen. Consequently, the final list of covid terms was reduced to 88 words. The classification of review employed a simple program coded in Python 3.8.8 to find words of the covid terms list in each of the reviews. This program generates a column indicating with 1 or 0 for the presence or absence of covid term and when a covid term is identify the frequency it appears in the review is store in a new column with the name of covid term.

3.4 Keyword analysis

According to Menner et al. (2016) topics can be identified as the most frequent terms in a user review, however, this analysis also includes rare terms, as they can reveal topics that are also relevant for the user experience. The following subsections explain the approaches to identify frequent and rare terms.

3.4.1 Frequent terms identification

Frequent terms were identified as the top 10 terms with the highest relative frequency (RTF) from the total of positive, neutral, and negative reviews. RTF is calculated as the ratio of the occurrences of each term to the maximal one and takes values from 0 to 1 (see the following equation).

$$RTF = \frac{TF}{\max(TF)}$$

3.4.2 Rare terms identification

Infrequent terms were identified as the top-10 terms with the highest TF-IDF score (term frequency - Inverse document frequency). With TF-IDF the frequency of terms gets offset by their occurrence across a set of documents (see equation below), thus frequent words that appear frequently among all documents get a lower score.

$$TF - IDF = \log\frac{N}{DF_i}(1 + \log TF_i)$$

3.4.3 Network diagrams

A network diagram or network graph represents how entities are connected among each other. Entities are nodes or vertices connected with links or edges. Entities can be any feature in the real world such as, individuals, organizations, nations among others. Links on the other hand, can be any attribute that tides them, such as contact, investment, friendship, co-occurrences and much more. In this analysis, nodes represent the most frequent terms found in reviews, while edges indicate the strength of the correlation among them.

Network diagrams are implement using the *igraph* package of the software R. An interesting feature of this package is that it offers the possibility to identify communities within networks using the cluster Louvain algorithm. This algorithm finds communities maximizing the modularity score. The modularity of the partition assesses the quality of the groupings within a network by measuring the density of edges inside communities in relation to edges between communities.

3.5 Spatial analysis

For the spatial analysis, Kernel Density estimation was used (KDE). Kernel density is a method that estimates a smooth density function of a variable using a kernel function (e.g., Gaussian, Epanechnikov, rectangular/uniform, triangular, biweight, cosine, optcosine) and a bandwidth radius. If applied to a two-dimensional space the kernel density function can estimate the intensity of points events in space. The kernel function gives more weight to points that are close to each other than farther away ones and the bandwidth determines the smoothness or roughness of the kernel histogram. A large bandwidth results in over-smoothed intensity values while a small bandwidth results in a spike around each point (Fortin, 2017; Sammut, 2017). According to this criteria bandwidths were selected for the analysis. KDE was implemented using ArcGIS Pro and the Spatial Analyst Tool, Kernel Density, which is built on the quartic kernel function introduced in (Silverman, 1986).

4 Results and discussion

A total of 26,262 reviews from 3,522 property listings in Rio de Janeiro, Brazil and 486,438 reviews from 18,751 property listings in New York, United States were analysed. There is a significantly reduction of reviews after the outbreak of COVID-19, 64% less reviews were posted from listings in U.S and 50% less reviews were posted from listings in Brazil.

4.1 Analysis of sentiment polarity

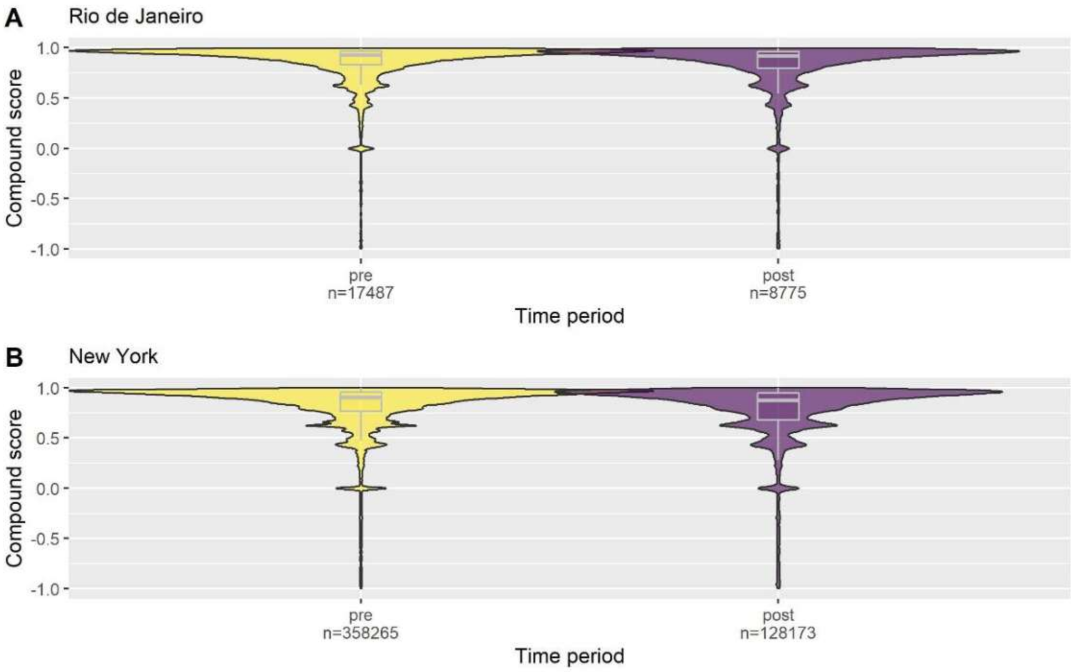
Analysis of reviews

The proportion of positive, neutral, and negative reviews from both Rio de Janeiro and New York were relatively the same before and after the outbreak of COVID-19 (See **Table 1**). After the outbreak of covid, positive reviews from Rio de Janeiro decreased 1%, neutral reviews remained the same amount and negative reviews increased 1%. Similarly, positive reviews from New York decreased 2% and negative and neutral reviews increased.

Table 1. Reviews per sentiment polarity from NY and RJ before and after the outbreak of covid.

		Pre		Post	
City	Polarity	Count	Percent (%)	Count	Percent (%)
Rio de Janeiro	positive	17064	98	8474	97
	neutral	197	1	113	1
	negative	226	1	188	2
	Total	17487	100	8775	100
New York	positive	346179	97	122170	95
	neutral	5192	1	2458	2
	negative	6894	2	3545	3
	Total	358265	100	128173	100

The long-left tail in the four plots of Figure x indicates that compound scores from reviews did not follow a normal distribution, and contrarily, there is a bias towards positive scores. This plot also shows that while the number of reviews is very low and relatively constant at negative scores, they start to increase a fluctuate in the neutral and positive part of the axis. It also shows that there are specific positive scores at which the number of reviews soars, having the highest peak near 1.



Analysis of property listings

The compound score of each review was averaged to obtain a summary score of each property listing. **Table 2** shows that properties with positive polarity declined 1.1% and 1.2% in Rio de Janeiro and New York respectively. Properties with neutral polarity increased 0.05% in Rio de Janeiro and 0.35% in New York. On the other hand, properties with negative polarity increased 1.1% in Rio de Janeiro and 0,8% in New York.

Table 2. Count and percentage of properties per sentiment polarity from RJ and NY in both time periods.

City	Polarity	Pre Count	Pre Percent (%)	Post Count	Post Percent (%)
Rio de Janeiro	positive	3478	98.75	3437	97.59
	neutral	21	0.60	23	0.65
	negative	23	0.65	62	1.76
	Total	3522	100	3522	100
New York	positive	18661	99.5	18436	98.32
	neutral	22	0.1	84	0.45
	negative	68	0.4	231	1.23
	Total	18751	100	18751	100

On the other hand, **Table 3** shows that number of listings that shifted from polarity was not significantly high, as it only occurred in 2.1% of the properties in New York and 3.6% of the properties in Rio de Janeiro. The shift displayed by most of the properties in both cities was from positive to negative, as it occurred in 1.7% of the properties in Rio de Janeiro and 1.2% of the properties in New York, followed by the shifted from positive to neutral and from negative to positive. However, the shift exhibited by less properties was, from neutral to negative, which occurred in 0.03% of the properties in Rio de Janeiro and from negative to neutral which occurred in 0.01% of the properties in New York.

Table 3. Count and percentage of properties from RJ and NY that shifted of polarity after the outbreak of covid.

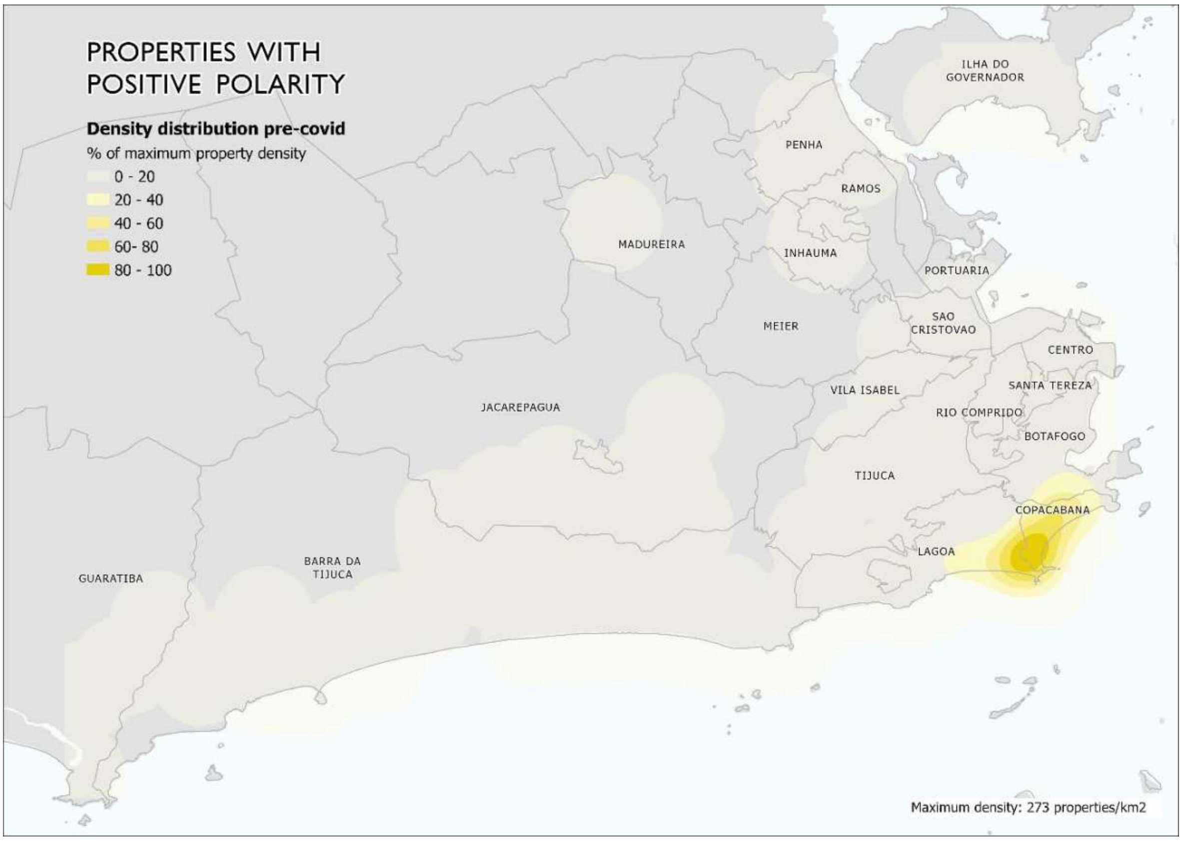
Polarity Pre	Polarity Post	Rio de Janeiro Count	Rio de Janeiro Percent (%)	New York Count	New York Percent (%)
neutral	positive	20	0.57	22	0.12
negative		22	0.62	60	0.32
Subtotal		42	1.19	82	0.44
positive	neutral	23	0.65	82	0.44
negative		0	0	2	0.01
Subtotal		23	0.65	84	0.45
positive	negative	60	1.70	225	1.20
neutral		1	0.03	0	0.00
Subtotal		61	1.73	225	1.20
positive	positive	3395	96.39	18354	97.88
neutral	neutral	0	0		0
negative	negative	1	0.03	6	0.03
Subtotal		3396	96.4	18360	97.91
Total		3522	100	18751	100

Spatial distribution of sentiment polarity

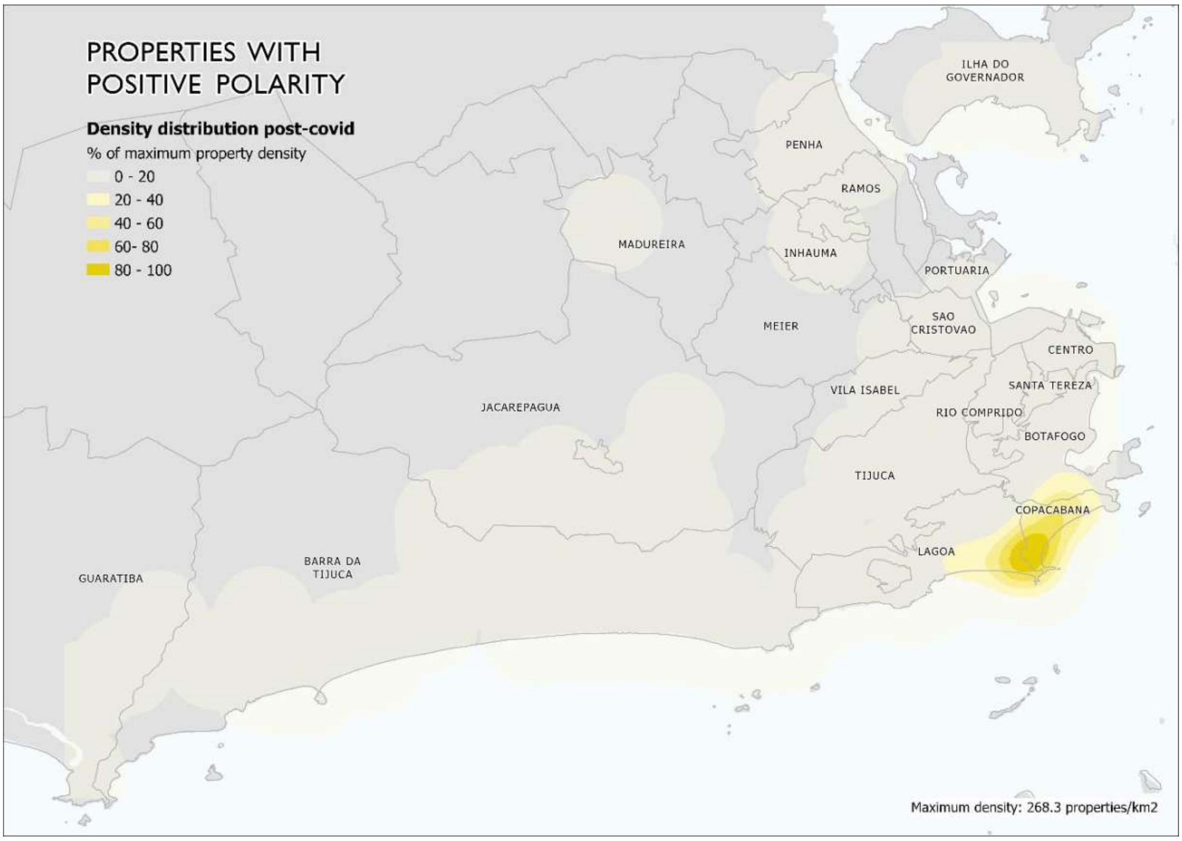
Rio de Janeiro, Brazil

Properties with positive polarity

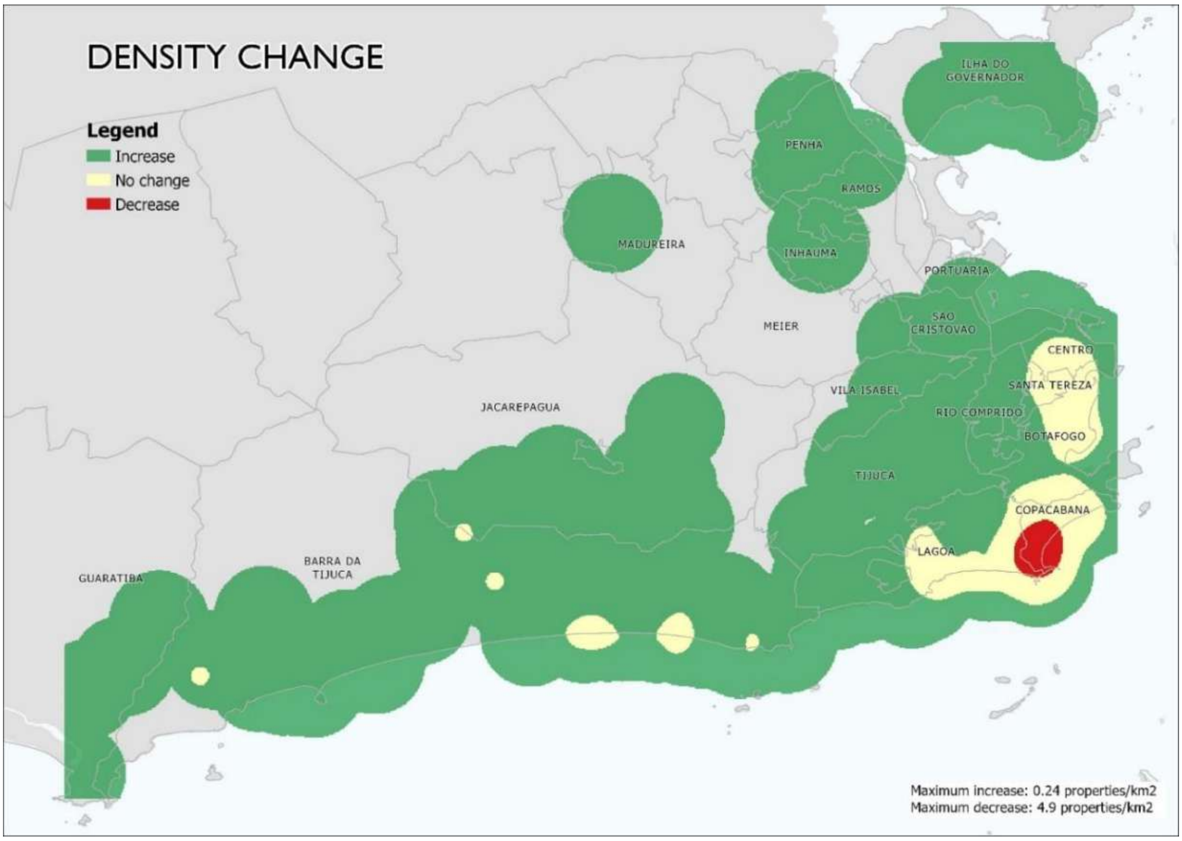
Before and after the outbreak of covid properties with positive polarity in Rio de Janeiro were distributed over twenty out of the thirty-three districts of the city. **Map 1** and **Map *2*** show that the highest concentration of these properties occurred, in both time periods, over the border of the districts Copacabana and Lagoa, as this area held between 80% and 100% of the maximum density of these properties. In these districts an also in Botafogo, also occurred between 20% and 40% of the property density. After the outbreak of covid the number of properties with positive polarity declined 1.1%, and although no noticeable in **Map *2***, this decrease is mainly reflected in a decrease of the concentration of properties with positive polarity in the same area with the highest concentration of these properties during both time periods (see **Map 3**).



Map 1. Density distribution of properties with positive polarity in RJ before the outbreak of covid.



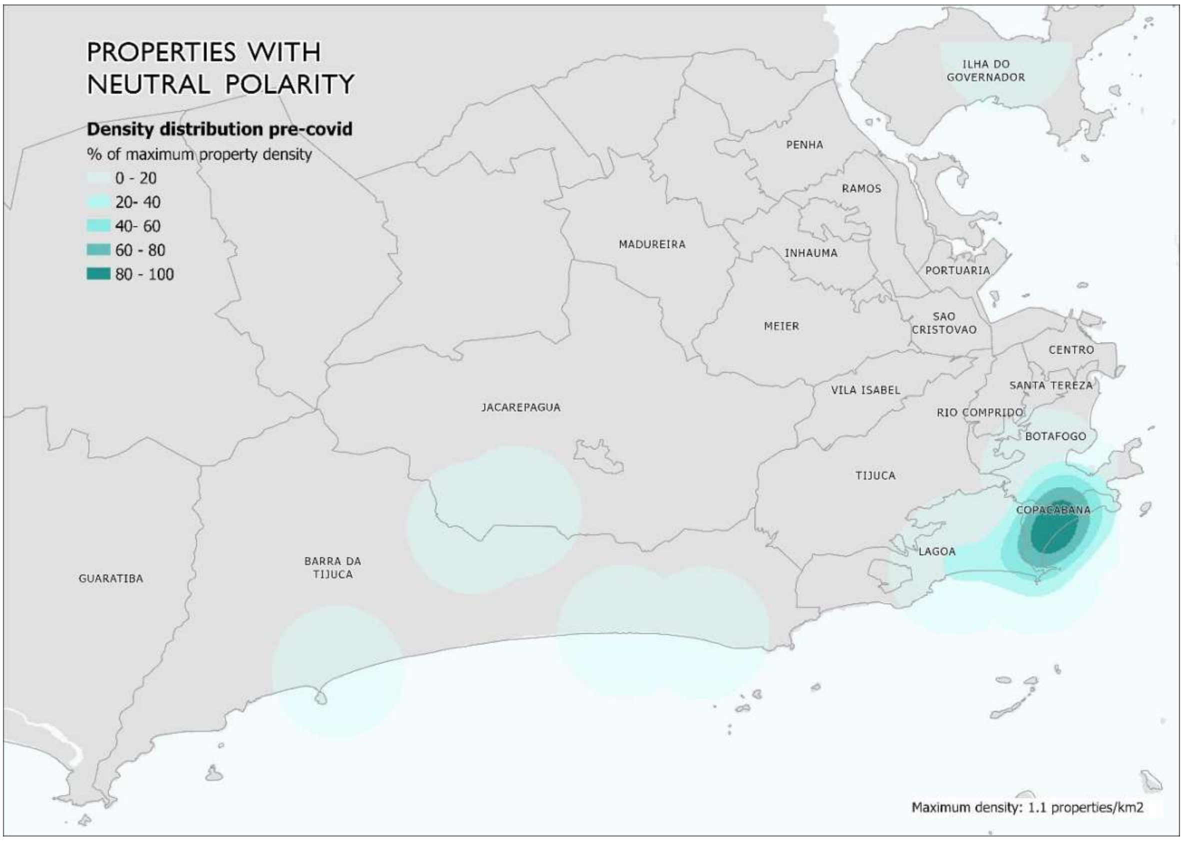
Map 2. Density distribution of properties with positive polarity in RJ after the outbreak of covid.



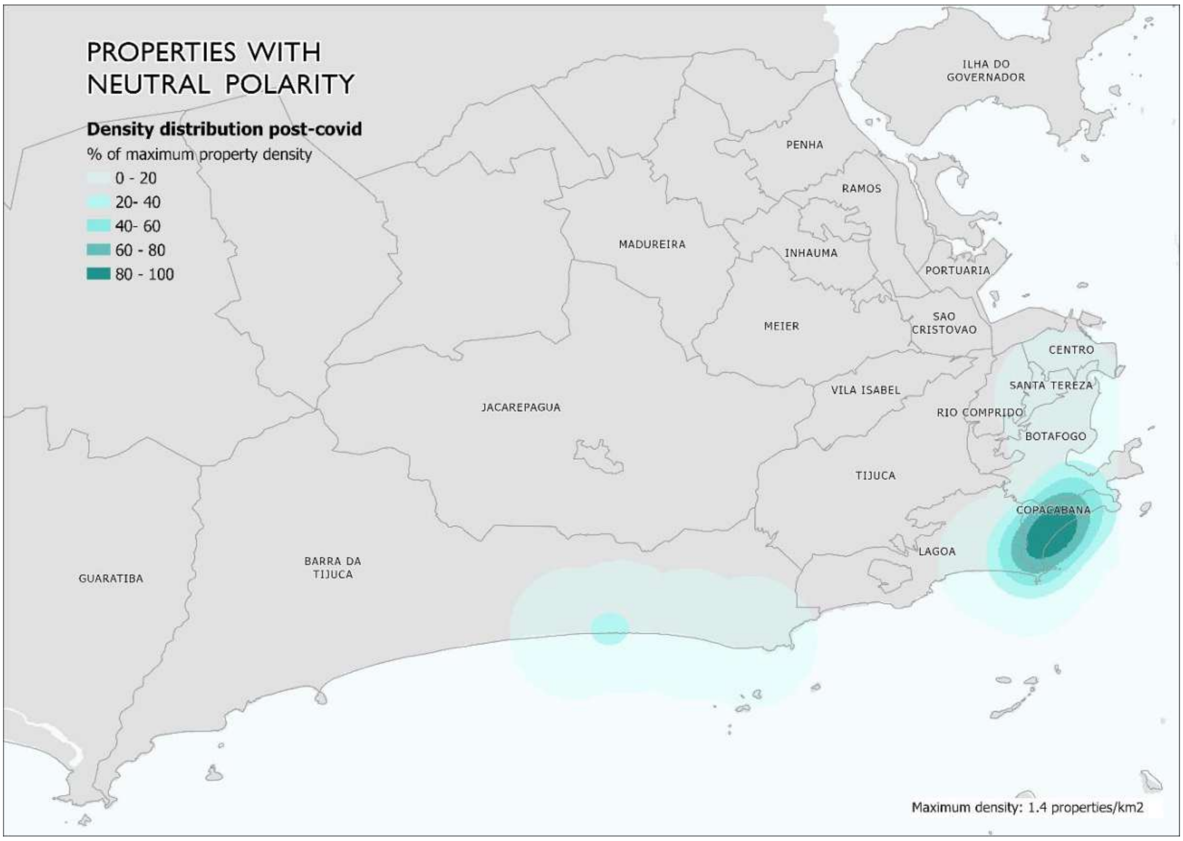
Map 3. Change in the density of properties with positive polarity in RJ after the outbreak of covid.

Properties with neutral polarity

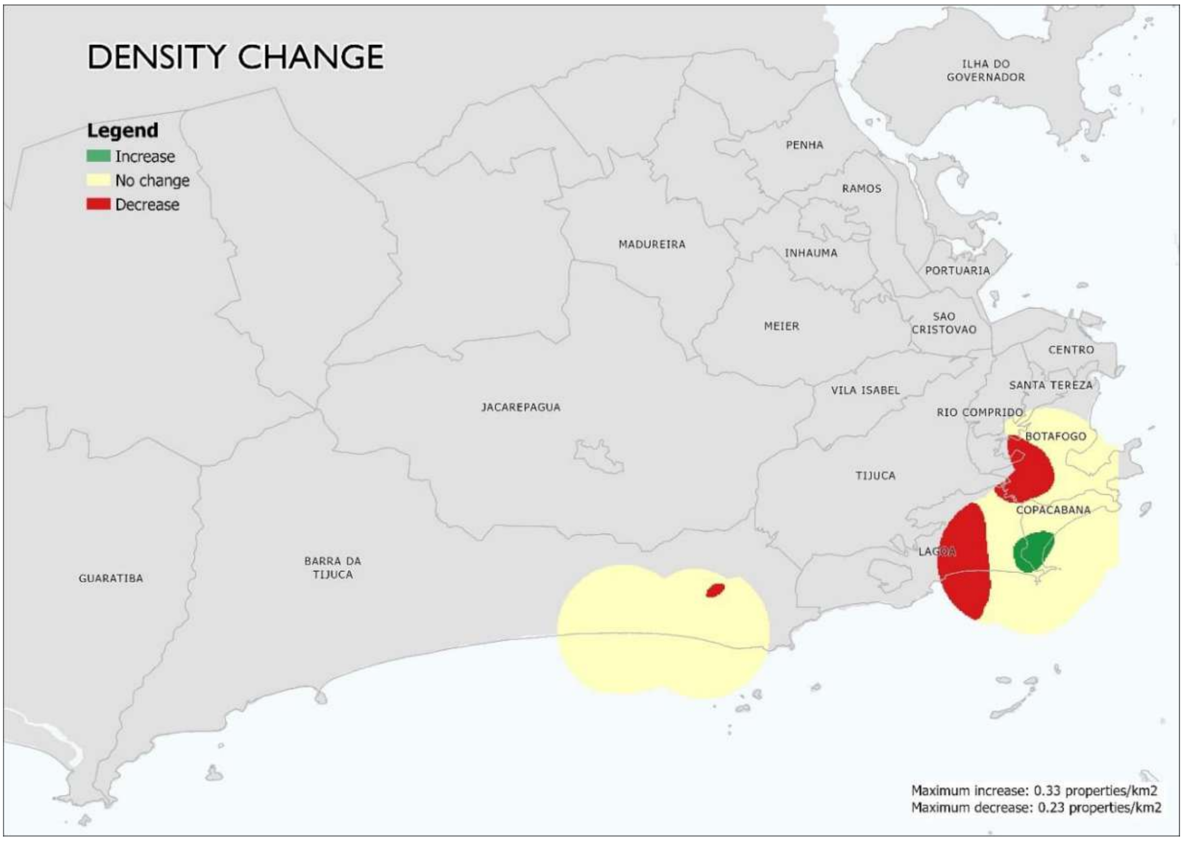
Before the outbreak of covid there were 21 properties with neutral polarity distributed over six districts, like properties with positive polarity, the range between 20% and 40% of the maximum density of these properties took place in the districts of Copacabana, Lagoa and Botafogo, however, the maximum concentration of these properties was in Copacabana (see **Map 4**). After the outbreak of covid, these properties were distributed over five districts, and Copacabana still held the maximum concentration, however, there was an increase of these properties of 9.5% which is reflected in a new are with the range between 20% and 40% of the property density in the district Barra da Tijuca and in a reduction of this range over the district Lagoa (see **Map 5**). **Map 6** also shows that the increase in neutral properties was mainly over the area with the highest concentration of these properties.



Map 4. Density distribution of properties with neutral polarity in RJ before the outbreak of covid.



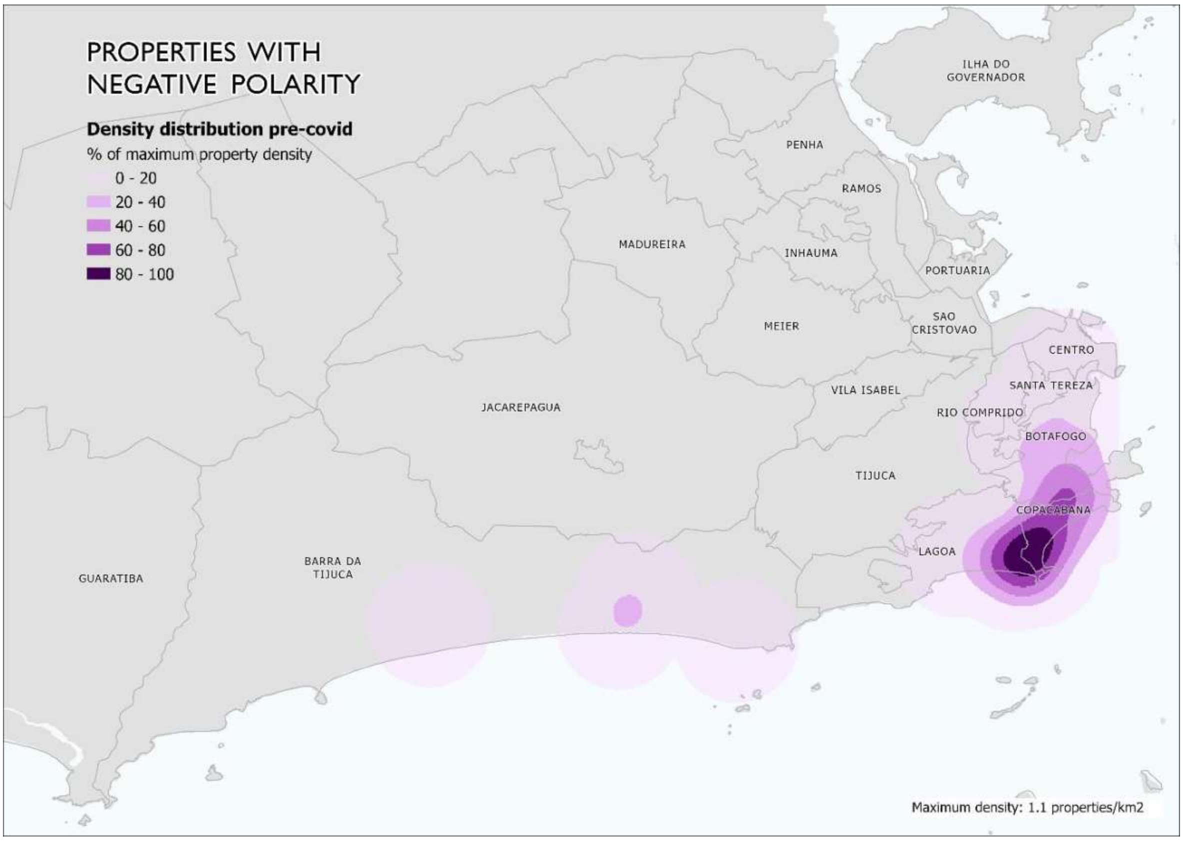
Map 5. Density distribution of properties with neutral polarity in RJ after the outbreak of covid.



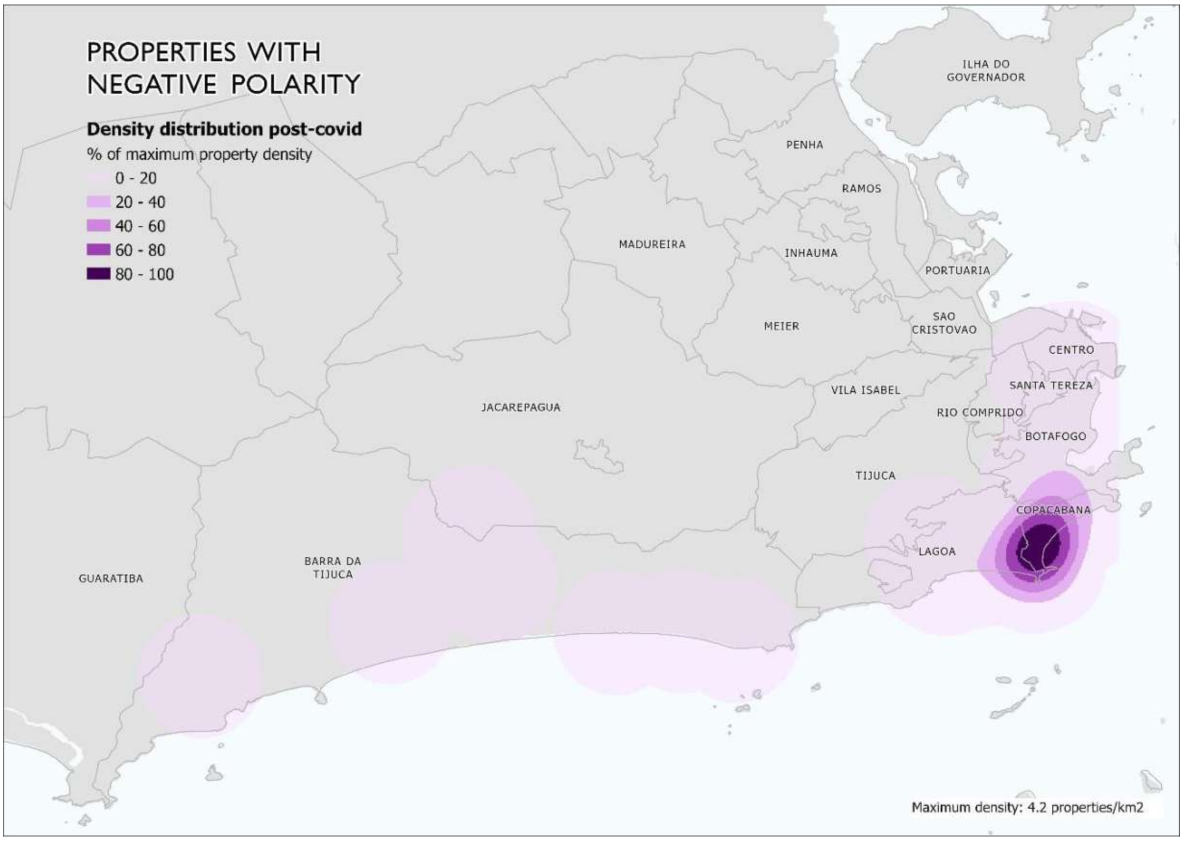
Map 6. Change in the density of properties with neutral polarity in RJ after the outbreak of covid.

Properties with negative polarity

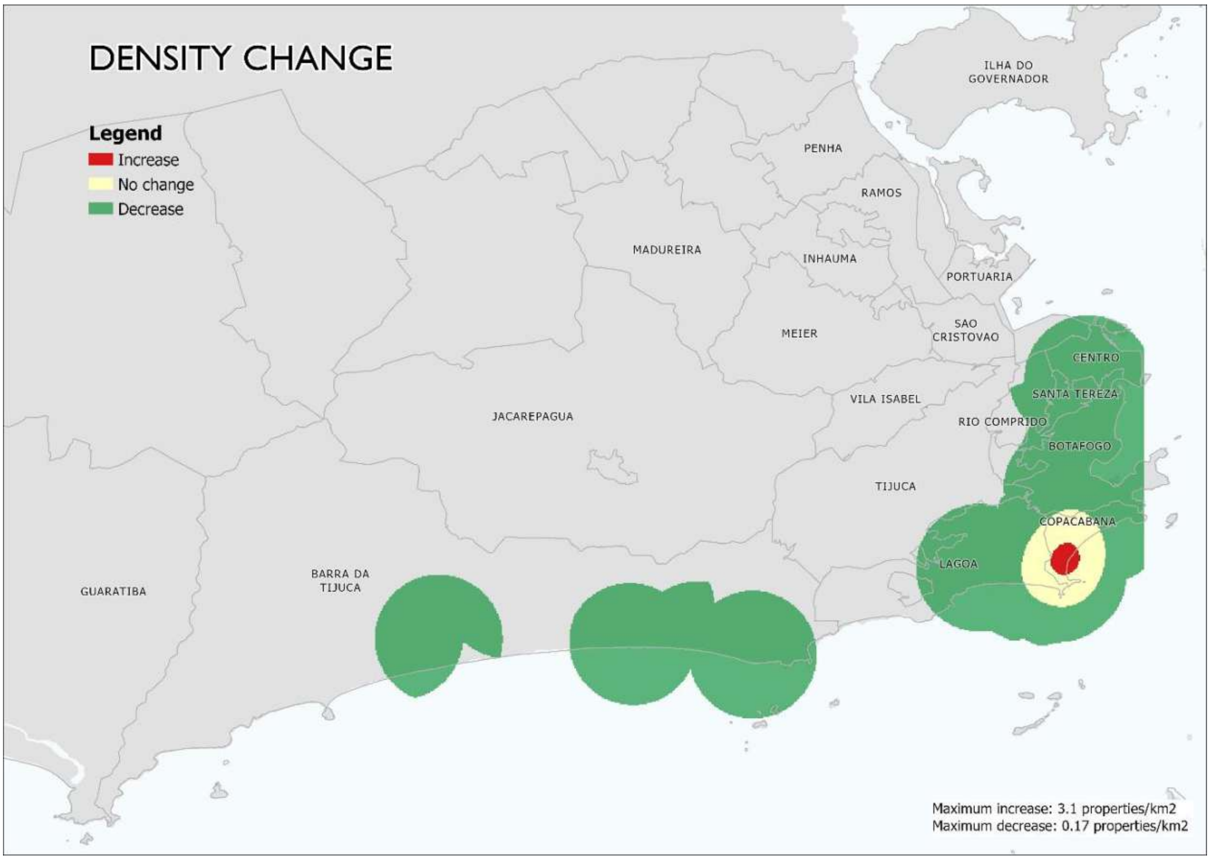
Before the outbreak of covid there were 23 properties with negative polarity distributed over five districts, like properties with positive polarity, the highest concentration of these properties was over the border between Copacabana and Lagoa, however the range between 20% and 40% of the maximum density of these properties covered more area of Botafogo and there was a small area with this range in the district Barra da Tijuca (see **Map 7**). After the outbreak of covid, the number of properties with negative polarity increased 170% and they distributed over two more districts, the area with the highest concentration of these properties was still in the same location, however, the area of the range between 20% and 40% got shrink, covering less area in Botafogo and no area in Barra da Tijuca (see **Map 8**). **Map 9** also shows that the increase in properties with negative polarity mainly occurred in the area with the highest concentration of these properties, as there was and increase in density of these properties.



Map 7. Density distribution of properties with negative polarity in RJ before the outbreak of covid.



Map 8. Density distribution of properties with negative polarity in RJ after the outbreak of covid.

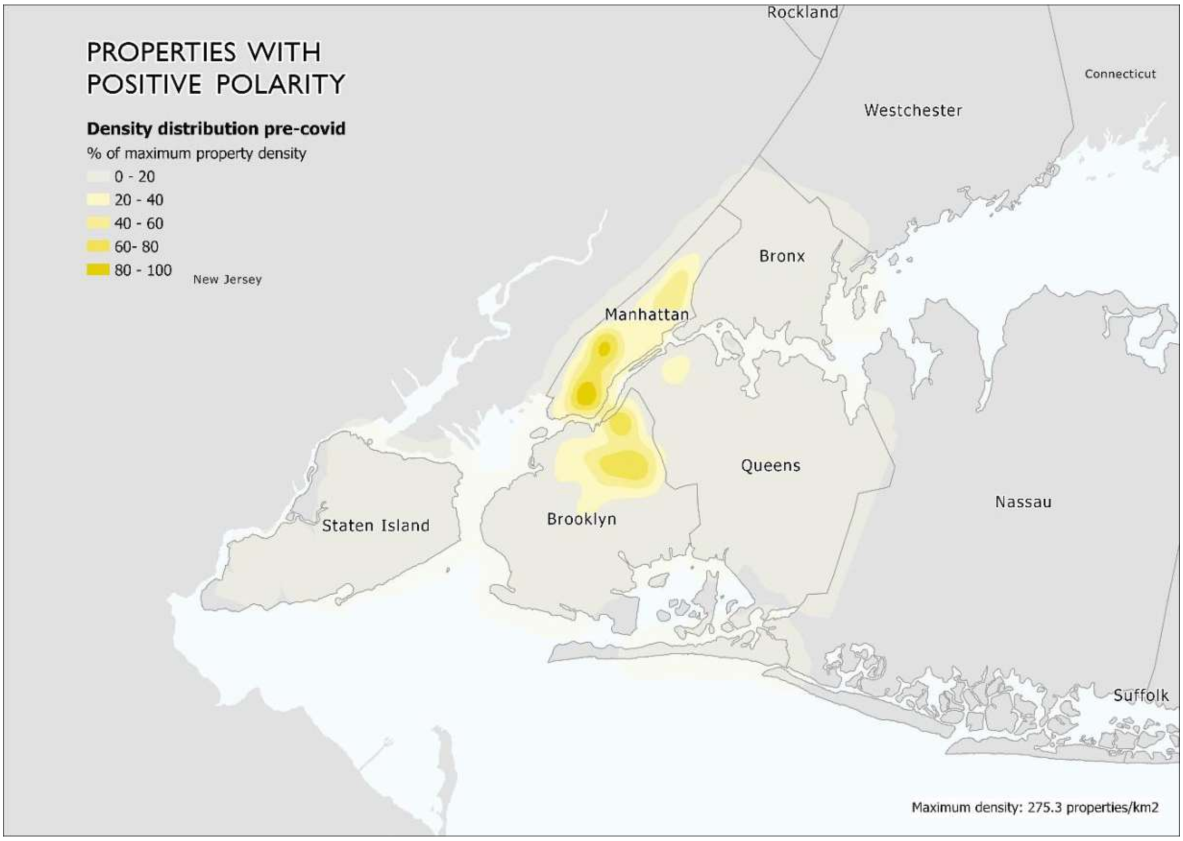


Map 9. Change in the density of properties with negative polarity in RJ after the outbreak of covid.

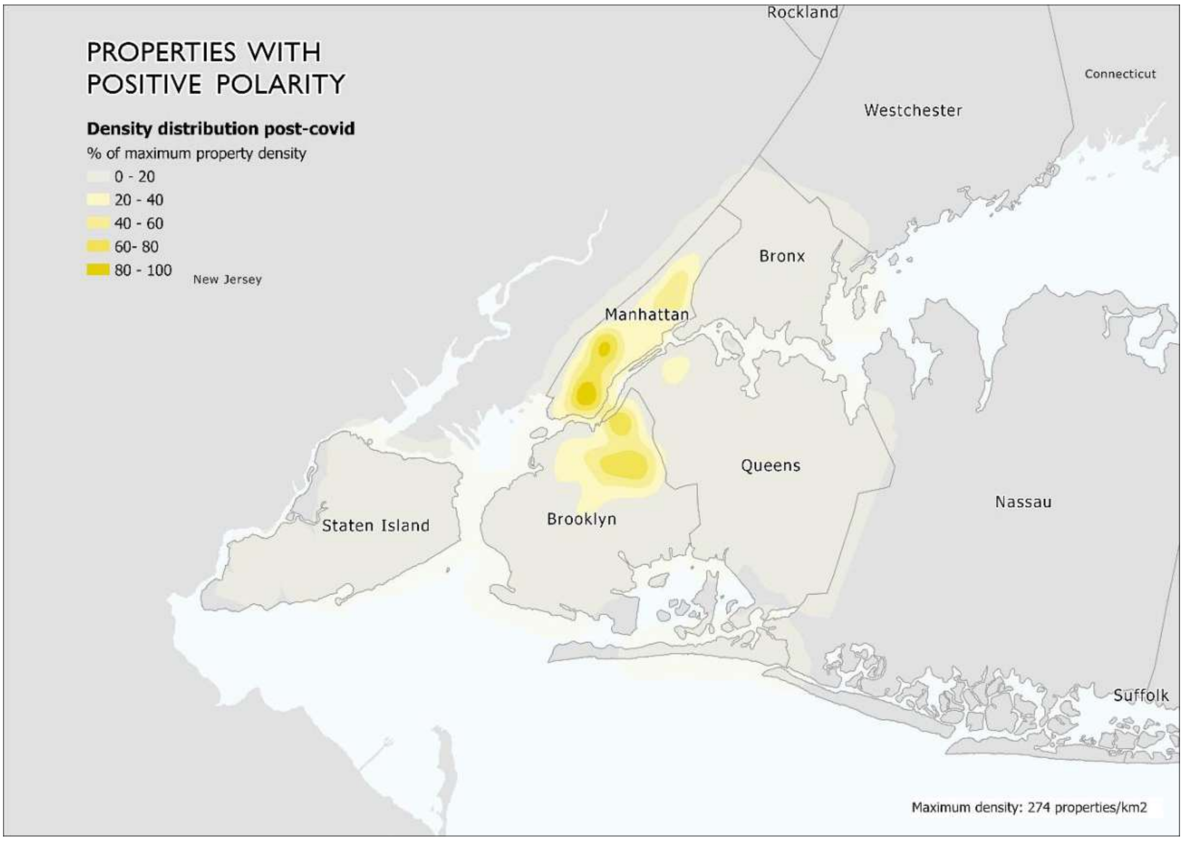
New York, United States

Properties with positive polarity

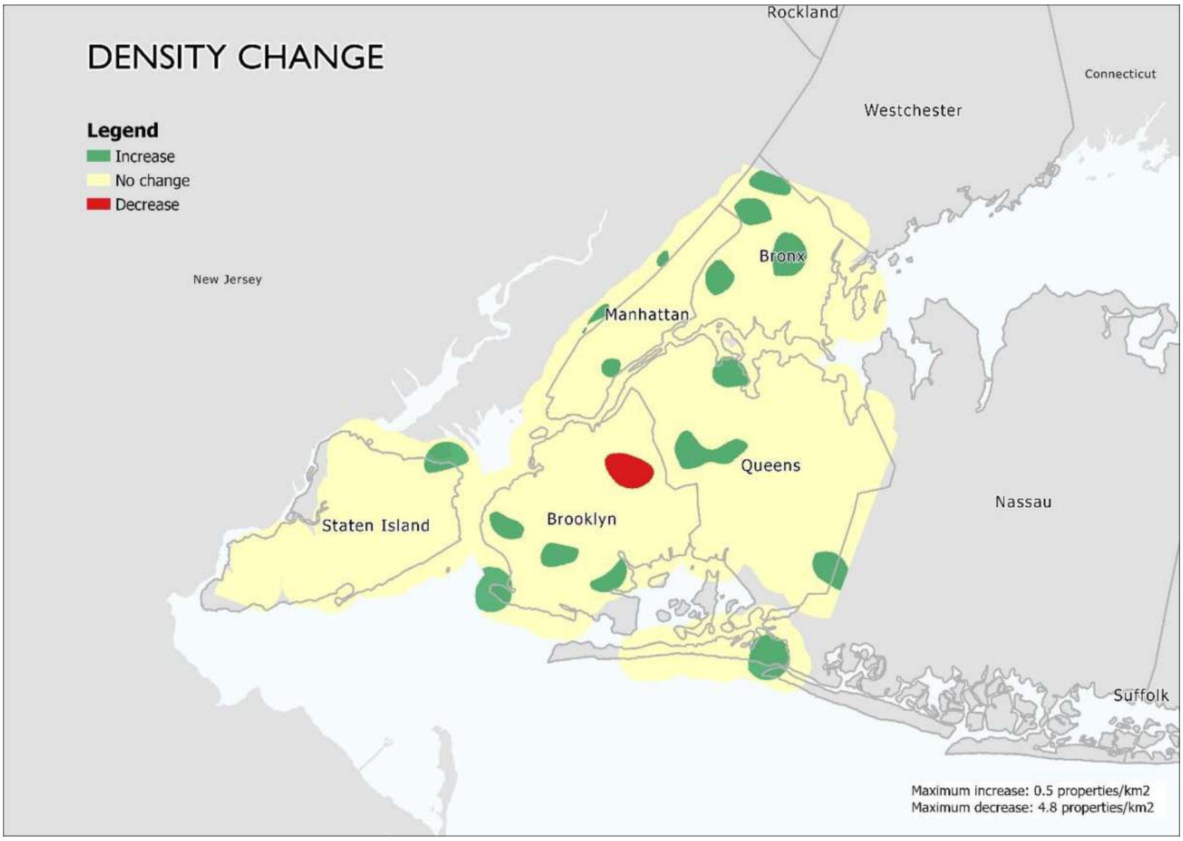
Before and after the outbreak of covid properties with positive polarity were distributed over the five boroughs of the city and two counties. **Map 10** and **Map 11** show that the highest concentration of these properties occurred, in both time periods, in two areas of Manhattan, as they held between 80% and 100% of the maximum density of these properties. In this district as well as in Brooklyn and in small area of Queens took place the range between 20% and 80% of the property density. After the outbreak of covid the number of properties with positive polarity declined 1.1%, and although no noticeable in **Map 11**, this decrease is mainly reflected in a decrease of the concentration of properties with positive polarity in the area with range between 60% and 80% of the property density in Brooklyn (see **Map 12**).



Map 10. Density distribution of properties with positive polarity in NY before the outbreak of covid.



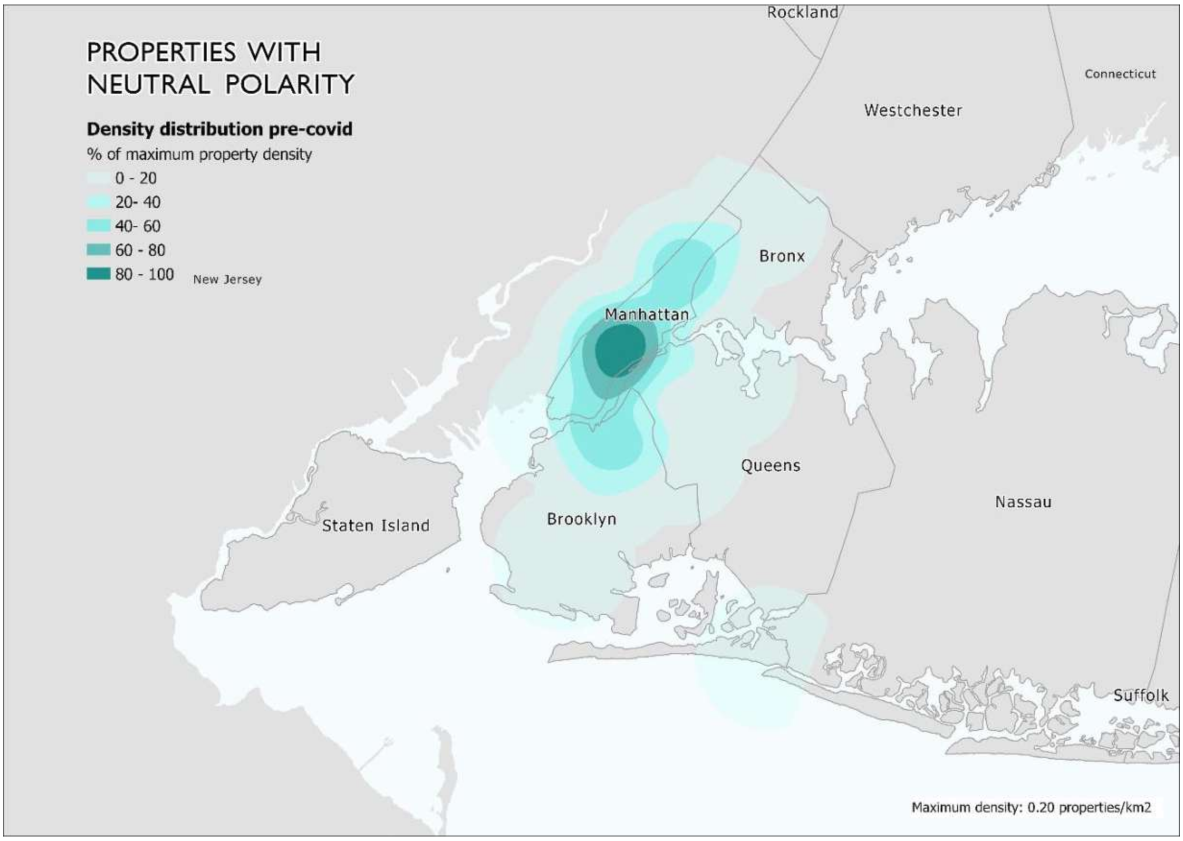
Map 11. Density distribution of properties with positive polarity in NY after the outbreak of covid.



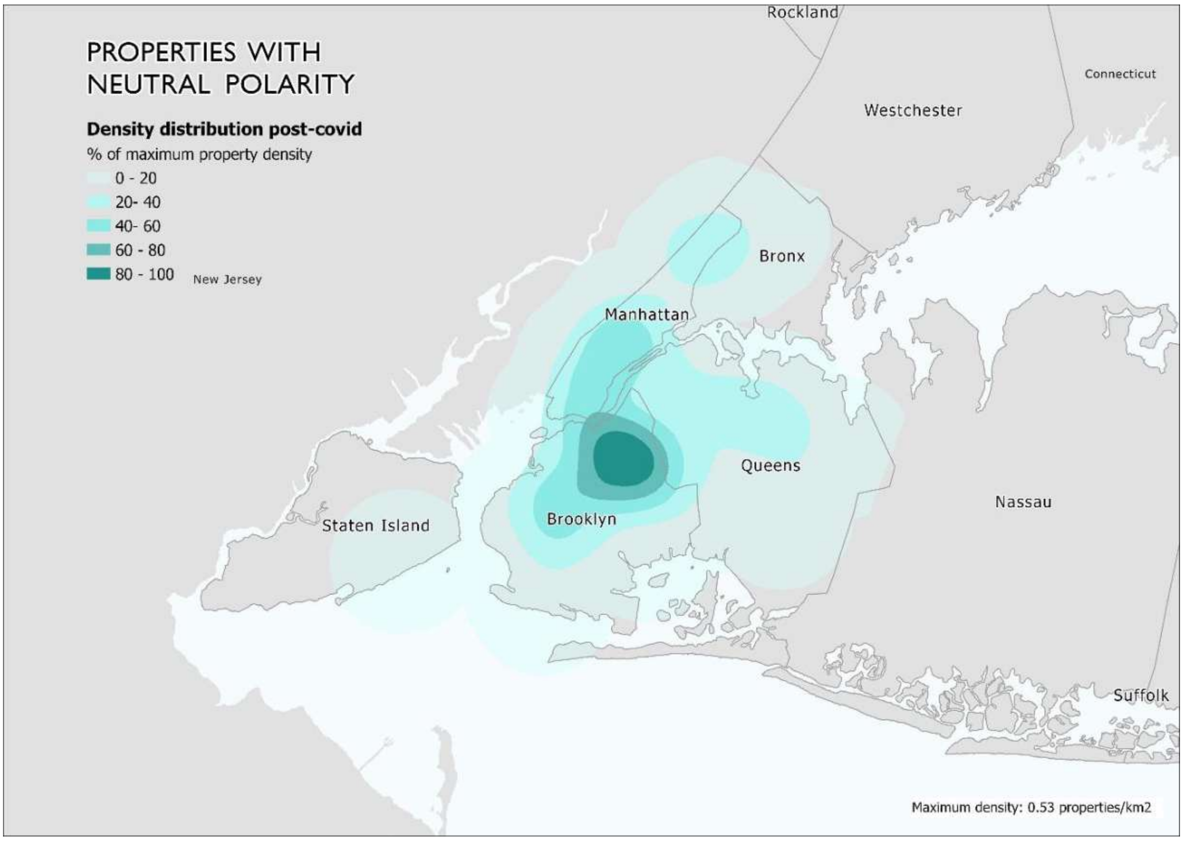
Map 12. Change in the density of properties with positive polarity in NY after the outbreak of covid.

Properties with neutral polarity

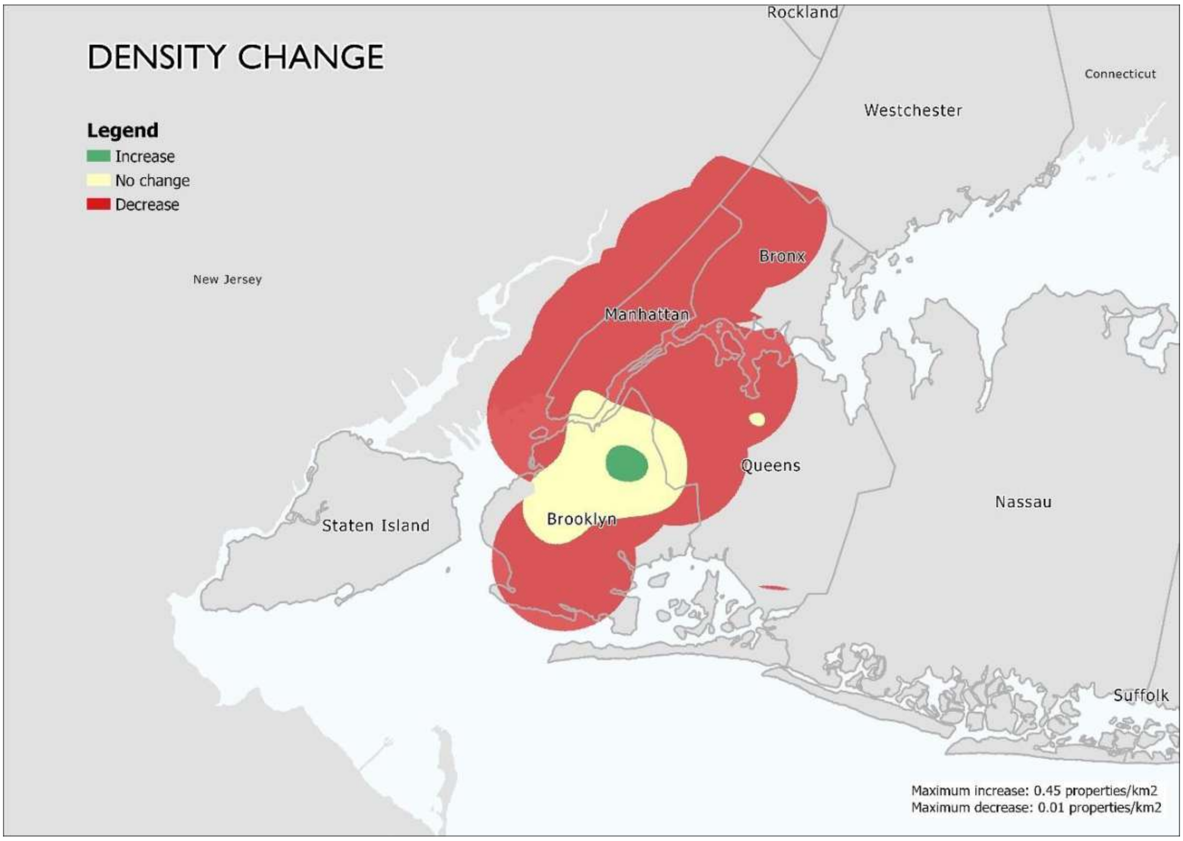
There were 22 properties with neutral polarity before the outbreak of covid, which were distributed over four of the boroughs. Like properties with positive polarity, the highest concentration of these properties occurred in Manhattan. I this borough and as well in Brooklyn occurred the range between 20% and 40% (see **Map 13**). However, after the outbreak of covid, properties with neutral polarity increased 74%, and this increase was mainly reflected in an increase of the concentration of properties with this polarity in Brooklyn, and an expansion of the range between 20% and 40% of the property density to Queens (**Map 14**). **Map 15** summarizes the changes previously presented.



Map 13. Density distribution of properties with neutral polarity in NY before the outbreak of covid.



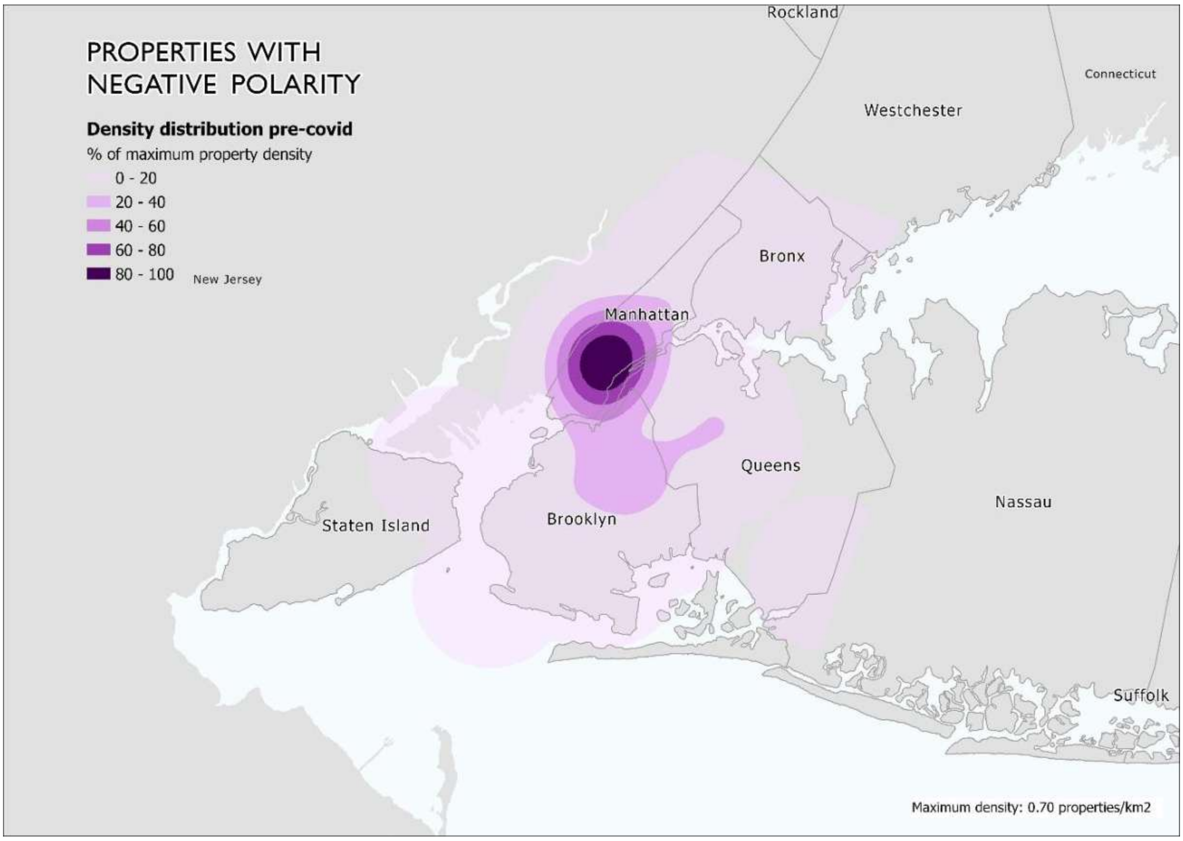
Map 14. Density distribution of properties with neutral polarity in NY after the outbreak of covid.



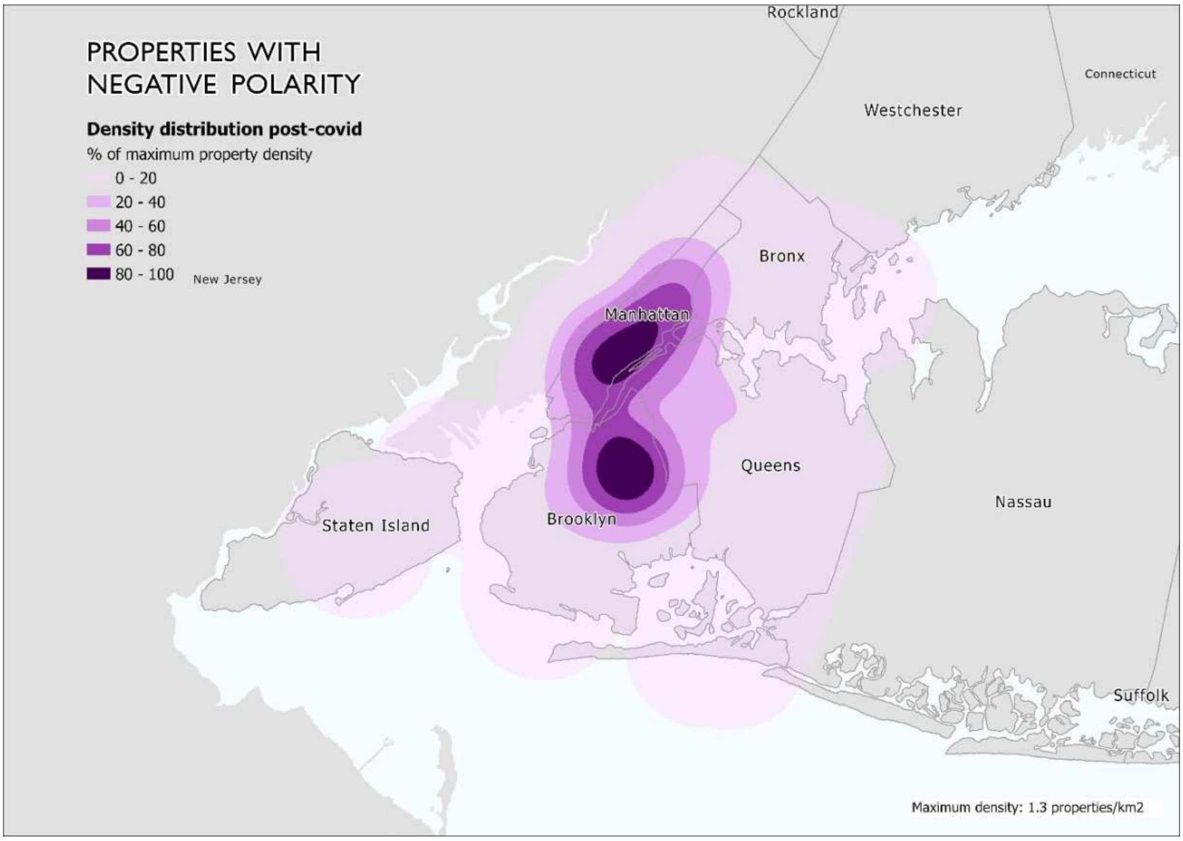
Map 15. Change in the density of properties with neutral polarity in NY after the outbreak of covid.

Properties with negative polarity

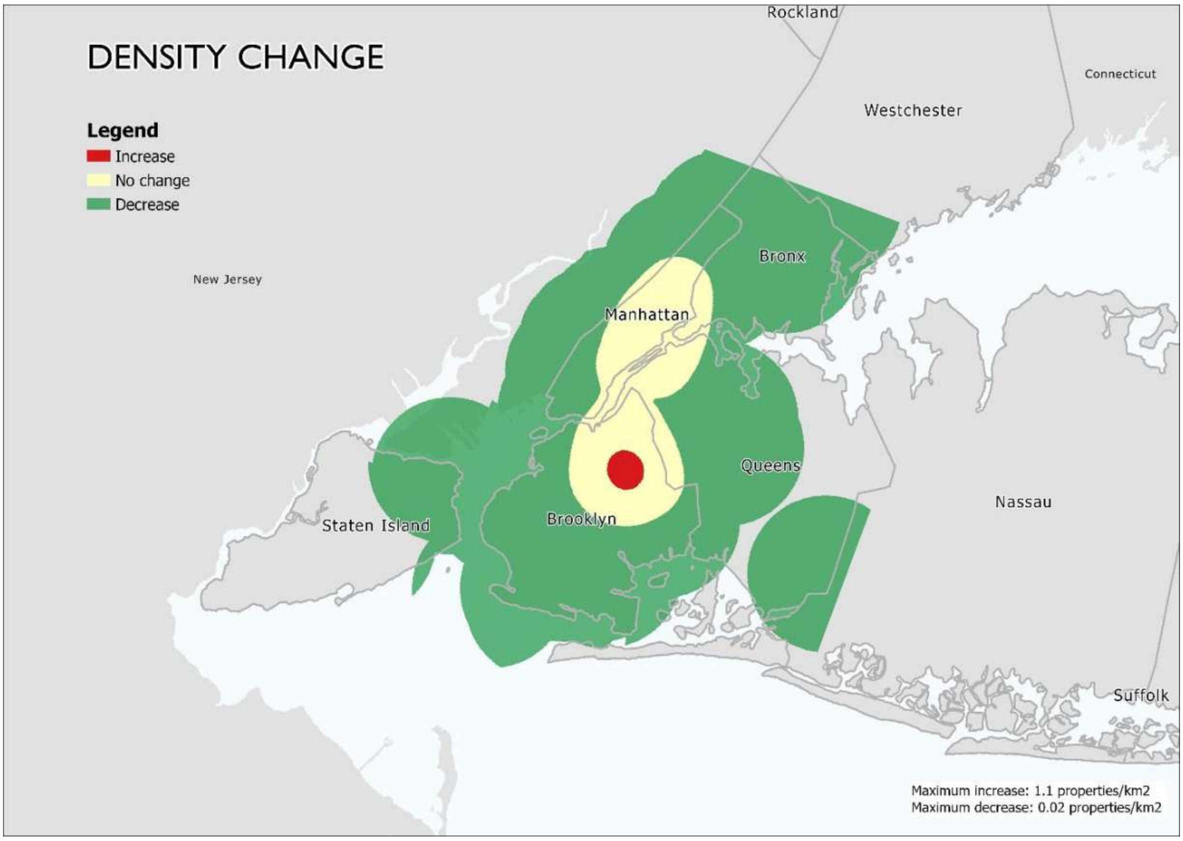
There were 68 properties with negative polarity before the outbreak of covid and distributed over the five boroughs of the city. Like properties with positive and neutral polarity, Manhattan had the highest concentration of properties with negative polarity, however the area that concentrated between 40% and 80% of the maximum density of properties with this polarity still covered Manhattan (see **Map 16**). After the outbreak of covid, the number of properties with negative polarity increased 275% and this change was reflected as a new area with the range between 80% and 100% of the property density of in Brooklyn and an extension of the area with the range between 40% and 80% of the property density to Brooklyn as well (see **Map 17**). **Map 18** summarizes the changes.



Map 16. Density distribution of properties with negative polarity in NY before the outbreak of covid.



Map 17. Density distribution of properties with negative polarity in NY after the outbreak of covid.



Map 18. Change in the density of properties with negative polarity in NY after the outbreak of covid.

Although in both cities the number of properties with positive polarity declined and the number of properties with neutral and negative polarity increased, depending on the polarity they displayed different changes in the density of properties. In both cities the distribution of density to the maximum density of properties with positive polarity over the city did not change after the outbreak of covid, however, in Rio de Janeiro the decline of properties with this polarity mainly occurred in the same area of the highest concentration of properties, while in New York it mainly occurred over an area with a range of density between 60% and 80% of the maximum property density. After the outbreak of covid, in Rio de Janeiro, properties with neutral and negative polarity tended to get more concentrated towards the area of the highest density of properties, while in New York, the area with the highest concentration of properties with neutral polarity moved to a different borough and a new area with the highest concentration of properties with negative polarity emerged.

4.1.1 Analysis of frequent keywords

Rio de Janeiro, Brazil

Pre-covid reviews

Fifteen different terms represented the 10 most frequent keywords of positive, neutral, and negative reviews pre-covid from property listings in Rio de Janeiro (**Table 4**). The top of positive, neutral and of negative reviews had in common 40% of these terms (“place”, “apartment”, “location”, “host”, “stay” and “room”), while the top of positive and neutral reviews shared 30% of them (“time”, “beach” and “restaurant”). There were not major differences in the frequency in which those terms were mentioned in positive, neutral, and negative reviews. The top one and second was always occupied by the keywords **place** and **apartment**, however the keyword **beach** went from being in the sixth place in the top of positive reviews to occupy the third place in the top of neutral ones. The term **host** was more relevant in the negative reviews, as it went from being in the fourth position in the top of positive reviews and sixth position in the top of neutral ones, to occupy the third place in the top of negative reviews. Just one keyword was unique to the top of positive (“view”) and neutral (“price”) reviews, while four keyword were to the top of negative ones (“day”, “night”, “issue”, “work”).

Table 4. List with the 10 most frequent terms from reviews of properties in RJ before the outbreak of covid.

	Positive			Neutral			Negative		
No.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
1	**place**	9178	1	**place**	24	1	**apartment**	160	1
2	**apartment**	8113	0,88	**apartment**	24	1	**place**	120	0,75
3	**location**	8065	0,88	**beach**	24	1	**host**	115	0,72
4	**host**	5222	0,57	**location**	15	0,63	**location**	79	0,49
5	**stay**	4562	0,50	**stay**	11	0,46	**day**	68	0,43
6	**beach**	4348	0,47	**host**	9	0,38	**stay**	62	0,39
7	**time**	2587	0,28	**room**	8	0,33	**room**	62	0,39
8	view	2524	0,28	**time**	6	0,25	night	54	0,34
9	**restaurant**	2480	0,27	**restaurant**	6	0,25	issue	37	0,23
10	**room**	1701	0,19	price	4	0,17	work	35	0,22

Post-covid reviews

On the other hand, the 10 most frequent keywords of positive, neutral, and negative reviews post-covid were represented by sixteen different terms (**Table 5**), 25% of them were shared by the top of positive, neutral and negative reviews ("place", "location", "apartment" and "host"), another 25% ("stay", "beach", "view" and "restaurant") by the top of positive and the top of neutral reviews and 12,5% ("time" and "room") by the top of positive and negative reviews. As in the top of reviews pre-covid, there were not major differences in the frequency of those terms among positive, neutral, and negative reviews. The top one was always occupied by the term **place** and **apartment**; however, the second place was occupied by the keyword **location** in the top of positive reviews and by the keyword **beach** in the top of neutral reviews. Unlike the top of reviews pre-covid, the term **host** became less frequently mentioned in the negative reviews, as it was 11% more times mentioned in neutral reviews and 7% more times mentioned in positive reviews than in negative ones. There were not unique keywords to the top of positive reviews, while two ("need" and "street") were unique to the top of neutral reviews and four to the top of negative ones ("day", "night", "water" and "people").

Table 5. List with the 10 most frequent terms from reviews of properties in RJ after the outbreak of covid.

No.	Positive			Neutral			Negative		
1	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
2	**place**	4525	1	**place**	17	1	**apartment**	158	1
3	**location**	3874	0,86	**beach**	16	0,94	**place**	114	0,72
4	**apartment**	3839	0,85	**location**	12	0,71	**host**	76	0,48
5	**host**	2505	0,55	**host**	10	0,59	**location**	67	0,42
6	**stay**	2312	0,51	**stay**	9	0,53	day	61	0,39
7	**beach**	1967	0,43	**apartment**	9	0,53	**time**	60	0,38
8	**view**	1299	0,29	need	4	0,24	**room**	53	0,34
9	**time**	1263	0,28	**view**	3	0,18	night	49	0,31
10	**restaurant**	1127	0,25	street	3	0,18	water	38	0,24
No.	**room**	829	0,18	**restaurant**	3	0,18	people	36	0,23

Comparison

When comparing all the top keywords from reviews pre- and post- covid we could see that four keywords were constant in all of them, namely, **place**, **apartment**, **location**, and **host**. The top of positive reviews pre-covid shared all the frequent keywords with the top of positive reviews post-covid. The top of neutral reviews pre- and post- covid differ by three keywords ("price", "need" and "street") and the top of negative ones by four ("issue", "work", "water" and "people). The term **beach** was relatively more frequently in the top of neutral reviews pre-covid than post-covid, and the keyword **host** went to be 24% less frequently mentioned in the negative reviews post-covid than pre-covid.

New York, United States

Pre-covid reviews

Fifteen different terms represented the 10 most frequent keywords of positive, negative, and neutral reviews pre-covid from property listings in New York (**Table 6**), 46% of these terms ("place", "apartment", "location", "stay", "host", "room", "time") were shared by the top of positive, neutral, and negative reviews and only one term, **space**, by the top of positive and neutral reviews. The top one keyword was always the term **place**, and the top second was always **location**, except in the top of negative reviews, where the second place was occupied by the term **room**, which was 56% more frequent in negative reviews than in positive and neutral ones. The term **host** was also more frequently mentioned in negative reviews as it is 41% more frequent there than in neutral reviews and 16% more frequent there than in positive ones. The terms **home** and **area** were unique terms to the top of positive reviews, the terms **station** and **subway** to the top of neutral reviews and the terms **day**, **night,** and **bathroom** to the top of negative ones.

Table 6. List with the 10 most frequent terms from reviews of properties in NY before the outbreak of covid.

	Positive			Neutral			Negative		
No.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
1	**place**	206130	1	**place**	1160	1	**place**	3828	1
2	**location**	115551	0,56	**location**	547	0,47	**room**	3304	0,86
3	**stay**	102795	0,50	**stay**	536	0,46	**host**	2339	0,61
4	**host**	93211	0,45	**room**	345	0,30	**apartment**	2316	0,61
5	**apartment**	89161	0,43	**apartment**	265	0,23	night	1962	0,51
6	**room**	62843	0,30	subway	240	0,21	**time**	1553	0,41
7	**space**	46827	0,23	**host**	228	0,20	**stay**	1523	0,40
8	**time**	44286	0,21	station	216	0,19	bathroom	1514	0,40
9	home	31288	0,15	**time**	185	0,16	day	1448	0,38
10	area	30974	0,15	**space**	184	0,16	**location**	1431	0,37

Post-covid reviews

Likewise, the 10 most frequent keywords of positive, neutral, and negative reviews post-covid of property listings in New York were represented by fifteen different terms, 40% of them were shared by the top of positive, neutral, and negative reviews ("apartment", "place", "stay", "host", "room", "time"), and 20% by the top of positive and negative reviews ("location", "space" and "home"). The term **room** was also more frequently mentioned in negative reviews, as it was 55% more frequent there than in neutral reviews and 65% more frequent there than in positive ones. Similarly, the term **host** was relatively more frequent in negative reviews as it was mentioned 48% more times there than in neutral reviews and 26% more times there than in positive ones.

Like in the top of reviews pre-covid, here the top one term of positive, neutral, and negative reviews was also **place** and the top second from negative reviews was **room**, however the top second was replaced by the term **stay** in the top of positive and neutral reviews.

There was just one unique term to the top of positive reviews ("area"), as well as to the top of neutral ones ("book") and four to the top of negative reviews ("day", "night", "bathroom" and "door").

Table 7. List with the 10 most frequent terms from reviews of properties in NY after the outbreak of covid.

	Positive			Neutral			Negative		
No.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
1	**place**	70122	1	**place**	486	1	**place**	1930	1
2	**stay**	37576	0,54	**stay**	302	0,62	**room**	1755	0,91
3	**location**	34874	0,50	**location**	220	0,45	**host**	1329	0,69
4	**host**	29807	0,43	**room**	175	0,36	**apartment**	1047	0,54
5	**apartment**	24587	0,35	**space**	112	0,23	**night**	911	0,47
6	**room**	17956	0,26	**host**	102	0,21	**stay**	870	0,45
7	**space**	17153	0,24	**time**	90	0,19	**time**	863	0,45
8	**time**	13749	0,20	**apartment**	86	0,18	**day**	861	0,45
9	**home**	11197	0,16	book	75	0,15	**bathroom**	632	0,33
10	area	8943	0,13	**home**	61	0,13	**door**	515	0,27

Comparison

When comparing all the 10 most frequent keywords from reviews pre- and post- covid, we could see that all of them have in common seven keywords ("apartment", "place", "stay", "host", "room", "time", "location"). The top of positive reviews pre-covid shared all the keywords with top of positive reviews post-covid. The top of neutral reviews pre- and post- covid differed for three keywords ("subway", "station" and "book") and the top of negative reviews pre- and post- covid differed for just one keyword, **door**. The keywords **room** and **host** were relatively more frequent in the negative reviews post-covid than pre-covid.

Comparison between Rio de Janeiro and New York

All the tops of frequent keywords from properties in Rio de Janeiro, as well as from properties in New York had in common four terms, namely, **apartment**, **place**, **location,** and **host**, and all of them differed by twelve keywords, **price**, **need**, **street**, which were unique to the tops neutral of Rio de Janeiro, **subway**, **station,** and **book,** which were unique to the tops neutral of New York, **issue**, **work**, **water** and **people**, which were unique to the tops negative of Rio de Janeiro and **bathroom** and **door**, which were unique to the tops negative of New York

All the tops of positive reviews had in common seven keywords, namely, **apartment**, **place**, **location, host, room, time** and **stay,** and they differed by the terms **home** and **area**, which were unique to the tops of New York, and the terms **beach**, **restaurant** and **view** which were unique to tops of Rio de Janeiro.

All the tops of neutral reviews have in common five keywords, namely, **apartment**, **place**, **location**, **host** and **stay**, and all of them differed by six the keywords, namely, **price**, **need**, **street**, which were unique to the tops of Rio de Janeiro, and **subway**, **station**, and **book,** which were unique to the tops of New York. The top of neutral pre-covid of Rio de Janeiro shared the keyword **room** with the all the tops of neutral from New York.

Like all the tops of positive reviews from Rio de Janeiro and New York, all the tops of negative ones shared seven keywords, two of them are unique to all these top 10, namely, **day** and **night**. The top of negative reviews pre- and post- of Rio de Janeiro have in common the keyword **location** with the top of negative reviews pre-covid of New York, and the contrary occurs with the keyword **stay** which is shared by top of negative reviews pre-covid from Rio de Janeiro and the top pre- and post- of negative reviews from New York. All the tops of negative reviews differ by five keywords, namely, **issue**, **work**, **water**, **people,** and **door**. The keyword **bathroom** which is unique to the top of negative reviews from New York.

All the top ones were descriptors of property, namely, **place** and **apartment**, being **apartment** most often mentioned in the negative reviews of Rio de Janeiro. Other descriptors of property were **space**, **home**, and they were only found in the top of New York. The keyword **space** was unique to the top of positive and neutral reviews pre- and post- covid, the keyword **home** was unique to the top of neutral and positive reviews post-covid, as well as to the top of positive reviews pre-covid.

Different to the second place of all the top of Rio de Janeiro, which was occupied by descriptors of property, the second place of all the top of New York was occupied by three different keywords, namely, **room**, which appeared in top of negative reviews pre- and post- covid; **location**, which appeared in the top of positive and neutral reviews pre-covid and **stay** which replaced location after the outbreak of covid.

On the other hand, the keyword **host** was more often mentioned in the negative reviews pre-covid of Rio de Janeiro than in the negative ones pre-covid of New York. On the other hand, it was more often mentioned in the reviews post-covid of New York than in the reviews post-covid of Rio de Janeiro. The keyword **room** was more often mentioned in the negative reviews of New York than Rio de Janeiro.

While the keyword **day** was relatively more often found in the negative reviews of Rio de Janeiro than the keyword **night**, the opposite occurred in the negative reviews of New York, where the keyword **night** was relatively more often found than the keyword **day**. The keyword **location** was more often found in positive than in neutral reviews and in neutral reviews than in negative ones.

4.1.2 Analysis of rare keywords

Rio de Janeiro, Brazil

Pre-covid reviews

Twenty-eight different terms represented the 10 rarest terms from positive, neutral, and negative reviews pre-covid from property listings in Rio de Janeiro (

Table 8). The top of positive reviews shared one keyword with the top of neutral reviews ("year"), as well as one keyword with the top of negative reviews ("view"). The keyword **year** was as rare in positive reviews as in neutral ones, however, the keyword **view** was 13% less rare in negative reviews than in positive ones. 80% of keywords were unique to the top of positive reviews and 90% to the top of neutral reviews as well as to the top of negative ones.

Table 8. List with the 10 most rare terms from reviews of properties in RJ before the outbreak of covid.

	Positive		Neutral		Negative	
No.	Term	Score	Term	Score	Term	Score
1	**year**	1	**year**	1	sheet	1
2	word	1	xbox	1	review	1
3	wish	1	**time**	1	renter	1
4	wifi	1	thumb	1	pay	1
5	wife	1	stay	1	block	1
6	welcome	1	star	1	shop	0,93
7	wait	1	saucer	1	spot	0,92
8	**view**	1	rockstar	1	**view**	0,87
9	value	1	right	1	experience	0,84
10	treat	1	reply	1	clean	0,83

Post-covid reviews

Likewise, twenty-eight different terms represented the rarest keywords from positive, neutral, and negative reviews post-covid (

Table *9*). The keywords **view** and **time** were shared by the top of positive reviews and the top of neutral ones, and they were as rarely mentioned in the latter as in the former. 80% of the keywords were unique to the top of positive reviews as well as to the top of neutral ones, while 100% of the keywords were unique to the top of negative reviews.

Table 9. List with the 10 most rare terms from reviews of properties in RJ after the outbreak of covid.

	Positive		Neutral		Negative	
No.	Term	Score	Term	Score	Term	Score
1	wow	1	**view**	1	reservation	1
2	worth	1	**time**	1	problem	1
3	work	1	supermarket	1	comment	1
4	wait	1	street	1	cold	0,93
5	visit	1	stay	1	doubt	0,82
6	**view**	1	star	1	mosquito	0,81
7	value	1	sim	1	guest	0,80
8	**time**	1	show	1	frill	0,78
9	think	1	renovation	1	change	0,77
10	thanks	1	place	1	truth	0,77

Comparison

When comparing all the top rarest keywords pre- and post- covid, we can see that there are six keywords found in common in both time-periods. The keyword **time** was shared by the top of neutral pre-covid with the top of neutral and positive post-covid, and it was equally rare in all of them. The keyword **view** was shared by the top of positive pre-covid and the top of positive and neutral post-covid, and it was also equally rare in all of them. The keywords **wait** and **value** that were unique to the top of positive and the keywords **stay** and **star** that were unique to the top of neutral, they were as rare in the top pre-covid as in the top post-covid.

New York, United States

Pre-covid reviews

The 10 rarest keywords from positive, neutral, and negative reviews pre-covid from property listings in New York were represented also by twenty-eight terms (**Table 10**). From these terms, **visit** is shared by the top of positive and neutral, and the term **stay** by the top of neutral and negative. Both terms are as rarely found in neutral reviews as in positive and negative reviews, respectively.

Table 10. List with the 10 most rare terms from reviews of properties in NY before the outbreak of covid.

	Positive		Neutral		Negative	
No.	Term	Score	Term	Score	Term	Score
1	zone	1	way	1	train	1
2	worth	1	**visit**	1	thanks	1
3	worry	1	think	1	television	1
4	woman	1	term	1	**stay**	1
5	welcome	1	subway	1	pay	1
6	want	1	**stay**	1	neighborhood	0,93
7	walk	1	star	1	need	0,92
8	**visit**	1	spot	1	host	0,87
9	**value**	1	room	1	complaint	0,84
10	**trip**	1	review	1	comment	0,83

Post-covid reviews

On the other hand, the 10 rarest keywords from positive, neutral, and negative reviews post-covid were represented by twenty-six terms (**Table 11**). There was one term in common between the top of positive and neutral ("visit"), and with the top of negative ("worth"). There were two terms in common between the top of neutral and top of negative, **stay**, and **room**. The term **worth** was as rarely found in positive reviews as in negatives, the term **visit** was also as rare in positive reviews as in neutrals, and the term **stay** was as well, as rare in neutral reviews as in negative, however, the term **room** was slightly less rare in negative reviews than in neutral ones.

Table 11. List with the 10 most rare terms from reviews of properties in NY after the outbreak of covid.

	Positive		Neutral		Negative	
No.	Term	Score	Term	Score	Term	Score
1	year	1	wifi	1	**worth**	1
2	**worth**	1	**visit**	1	thanks	1
3	window	1	time	1	**stay**	1
4	welcome	1	**stay**	1	**room**	0,93
5	**visit**	1	star	1	refund	0,82
6	view	1	spot	1	place	0,81
7	vibe	1	smoking	1	money	0,80
8	value	1	service	1	host	0,78
9	use	1	**room**	1	condition	0,77
10	trip	1	review	1	comment	0,77

Comparison

When comparing all the top rarest keywords pre- and post- covid, we can see that twelve keywords were found in both time-periods. The keyword **visit** was shared by the top pre- and post- covid of positive and neutral reviews and was as rare in the top of pre-covid as in the top of post-covid. The keyword **stay** was shared by the top pre- and post- covid of neutral and negative reviews, and it was also as rare in the top pre-covid as in the top post-covid. The keyword **worth** was shared by the top of positive pre- and post- covid and the top of negative post-covid and was equally rare in all of them. The keywords **welcome**, **value** and **trip** that were unique to the top of positive, the keywords **star**, **room** and **review** that were unique to the top of neutral and the keyword **thanks** that was unique to the top of negative, were as rare in the top pre-covid as in the top post-covid, however, the keywords **host** and **comment** that were unique to the top of negative were slightly less rare in the top post-covid.

Comparison between Rio de Janeiro and New York

There were no keywords in common among all the tops from both cities. However, some of the keywords are shared among specific top 10. From the keywords that were unique to the top of positive reviews, the keyword **value** was shared by all the tops positive, the keyword **welcome** was shared by the top pre-covid from Rio de Janeiro and the top pre- and post- covid from New York, and the keyword **year** was shared between the top pre-covid from Rio de Janeiro and the top post-covid from New York, all these keywords were equally rare in all the mentioned top 10.

From the keywords unique to the top of neutral reviews, only the keyword **star** was shared by all of them and was also equally rare in all of them. On the other hand, from the unique terms to the top of negative reviews, there were only two keywords shared between cities, the keyword **pay** which was shared between the tops pre-covid from Rio de Janeiro and from New York and that was equally rare in both cities, and the keyword **comment**, shared by the tops post-covid of both cities but slightly less rare in the top of New York.

Other keywords that were shared between cities and equally rare among different top 10 were:

- **Wifi**: shared between the top of positive pre-covid of Rio de Janeiro and the top of neutral post-covid of New York.
- **Think**: shared between the top of positive post-covid of Rio de Janeiro and the top of neutral pre-covid of New York.
- **Thanks**: shared between the top of positive post-covid of Rio de Janeiro and the top of negative pre- and post- covid of New York.
- **Time**: shared by the top of neutral pre-covid and the top of post-covid positive and neutral of Rio de Janeiro with the top of neutral post-covid of New York.
- **Stay**: shared by the top of neutral pre- and post- covid of Rio de Janeiro and the top of neutral and negative pre- and post- covid of New York
- **Review**: shared by the top of negative pre-covid of Rio de Janeiro with the top of neutral pre- and post- covid of New York.
- **Worth**: shared by the top of positive post-covid of Rio de Janeiro with the top of positive pre- and post- covid and the top of negative post-covid of New York.
- **Visit**: shared by the top of positive post-covid with the top of positive and neutral pre- and post- covid of New York.

The keyword **view** was shared by the top of positive and negative pre-covid and the top of positive and neutral post-covid of Rio de Janeiro with the top of positive post-covid of New York. This keyword was only less rare in the top of negative pre-covid from Rio de Janeiro. On the other hand, the keyword **spot** was shared by the top of negative pre-covid of Rio de Janeiro with the top of neutral pre- and post- covid of New York.

4.1.3 Analysis of network diagrams

Network diagrams show the connection among keywords. The strength of the connection is represented by the thickness of the edges and according to the Spearman correlation score. The size of the nodes indicates number of connections of each keyword and the colours represent communities.

Positive reviews, Rio de Janeiro

Three communities represent the reviews pre- and post- covid (**Figure 2**, **Figure 3**). These communities have three topics in common, one about attributes of the property (e.g., bathroom, room, kitchen, bedroom), another one about the description of the experience (e.g., day, time, night) and another about the facilities in the neighbourhood (e.g., restaurant, bar, supermarket).

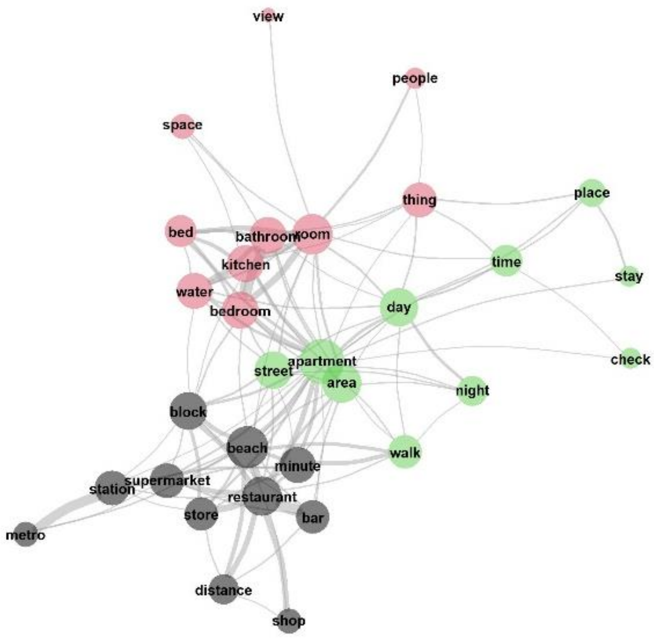


Figure 2. Network diagram of positive reviews pre-covid RJ.

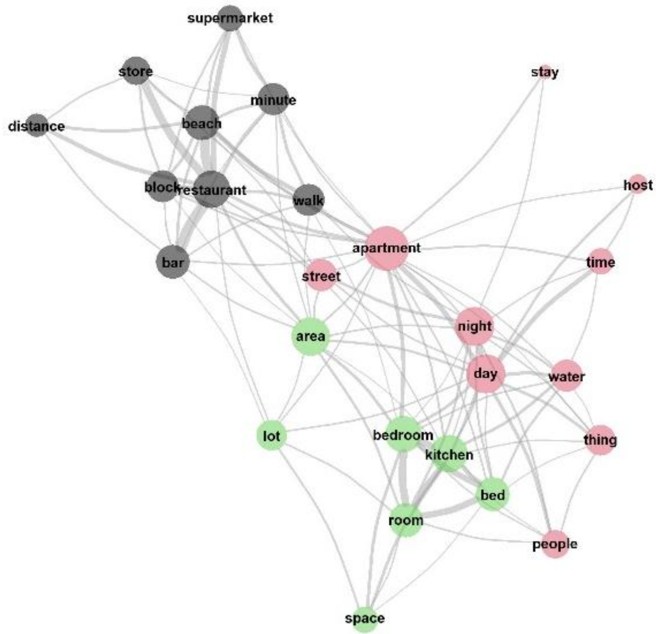


Figure 3. Network diagram of positive reviews post-covid RJ.

Positive reviews, New York

Three communities represented the reviews pre-covid, while two represented the reviews post-covid (**Figure 4**). However, the communities from reviews pre- and post-covid had the same topics in common (**Figure 5**). These topics were the same found in the positive reviews from Rio de Janeiro, namely, description of the experience, facilities in the neighbourhood and attributes of the property.

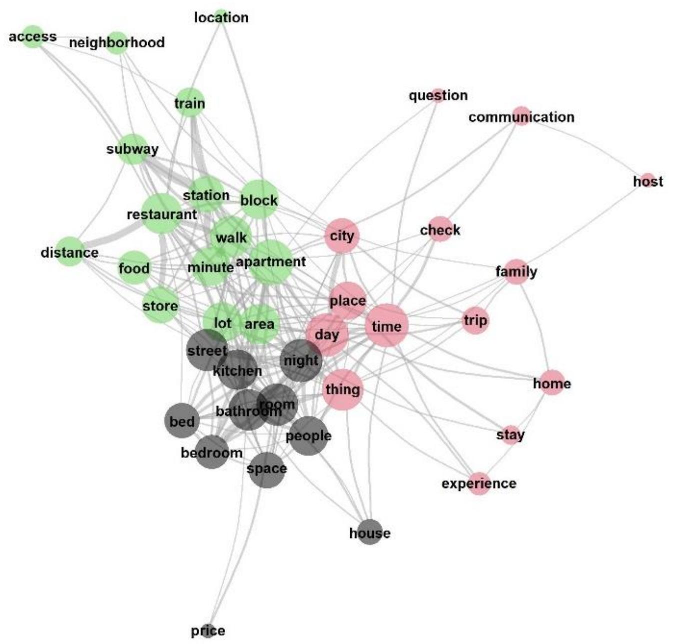


Figure 4. Network diagram of positive reviews pre-covid NY.

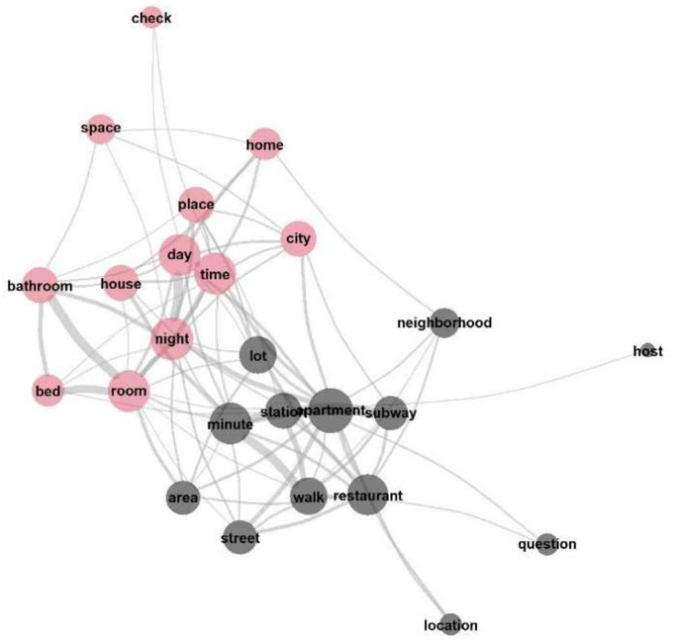


Figure 5. Network diagram of positive reviews post-covid NY.

Neutral reviews, New York

Four communities represented the reviews pre-covid and post-covid (**Figure 6**, **Figure 7**). The communities from reviews pre-covid contained descriptors of the experience, facilities in the neighbourhood, as well as positive descriptors of the location (“minute”, “walk” and “location”, “beat) and negative ones also (“noise”, “street”). On the other hand, communities from reviews post-covid contained descriptors of the host’ service, experience as well as positive descriptors of the location.

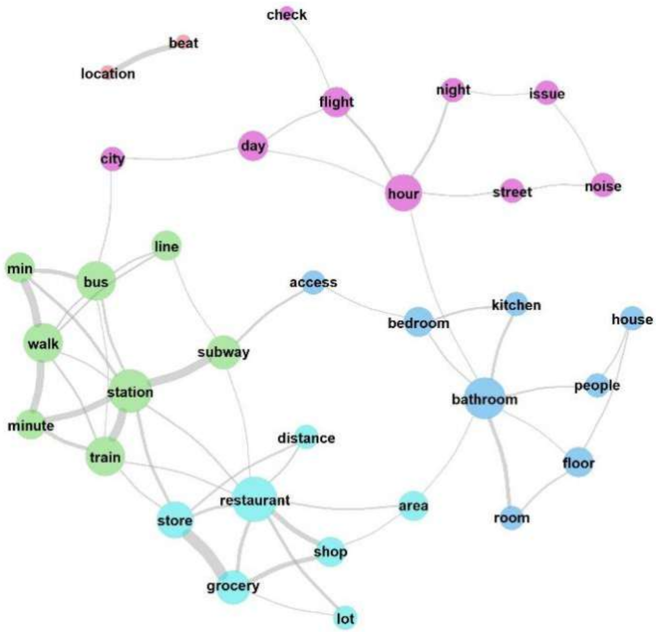


Figure 6. Network diagram of neutral reviews pre-covid NY.

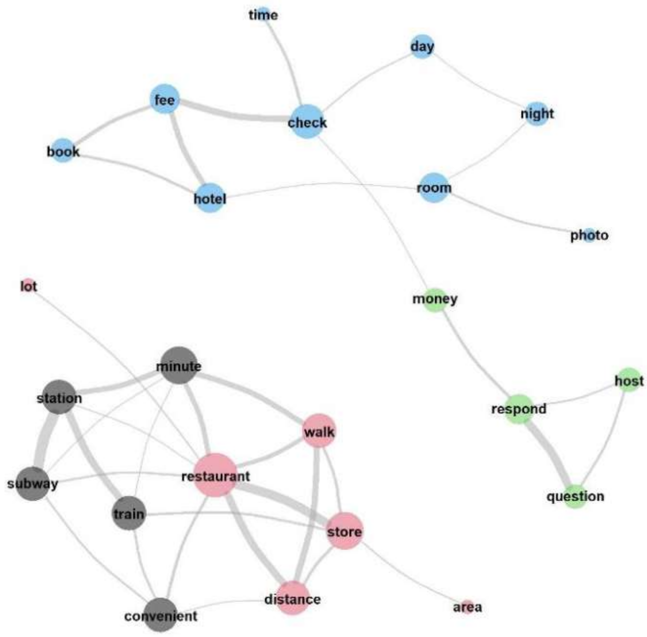


Figure 7. Network diagram of neutral reviews post-covid NY.

Negative reviews, Rio de Janeiro

Five communities represented the reviews pre-covid while three represented the reviews post-covid (**Figure 8**, **Figure 9**). While the communities from the reviews pre-covid explicitly mentioned issues with host, water, service and cleanliness, the communities from reviews post-covid do not explicitly describe issues but the experience, facilities in the neighbourhood and attributes of the property.

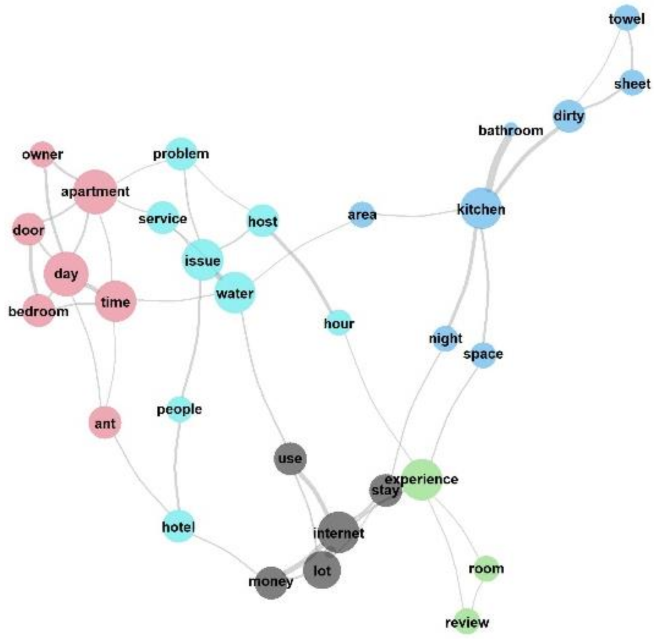


Figure 8. Network diagrams of negative reviews pre-covid from RJ.

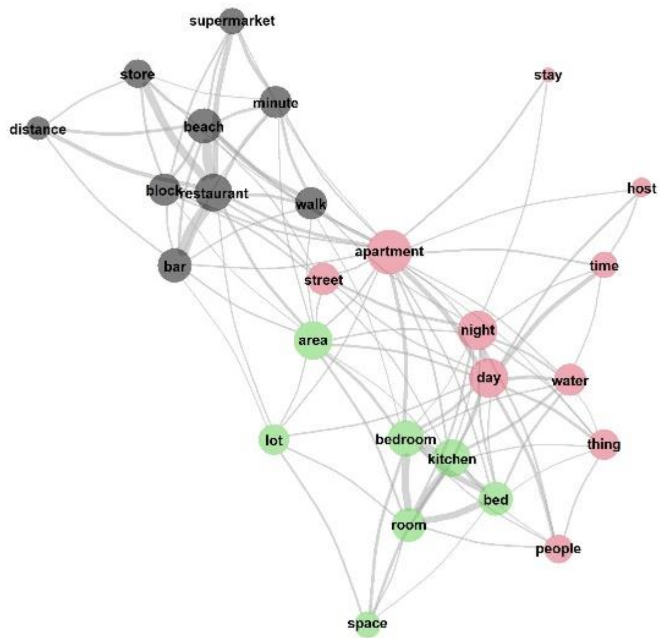


Figure 9. Network diagrams of negative reviews post-covid from RJ.

Negative reviews, New York

Five communities represented the reviews pre-covid and seven the reviews post-covid (**Figure 10**, **Figure 11**). Communities from reviews pre-covid contained descriptors of the experience (“day”, “time”, “stay”, “night”), host’s service (“key”, “check”, “response”, “message”) attributes of the property as well as description of problems with water, fees, cleanliness. On the other hand, the topics from reviews post-covid are less explicit, but still indicate issues with the smell, radiator, and cleanliness.

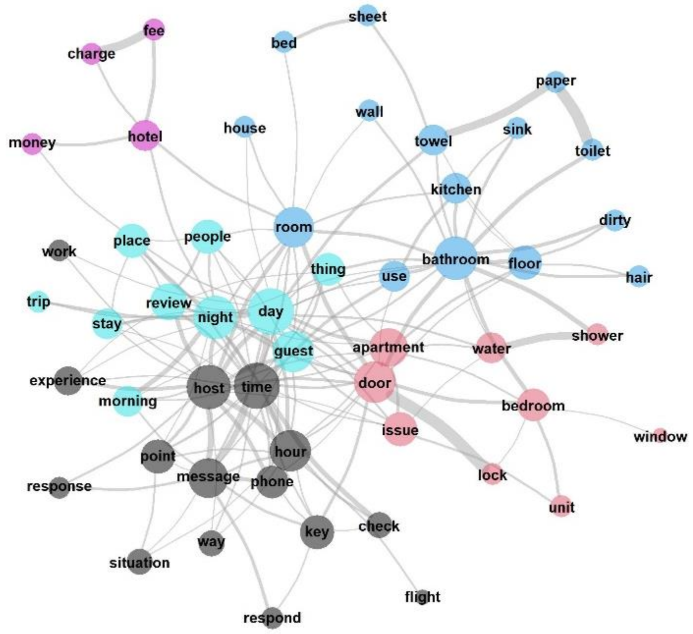


Figure 10. Network diagrams of negative reviews pre-covid from NY.

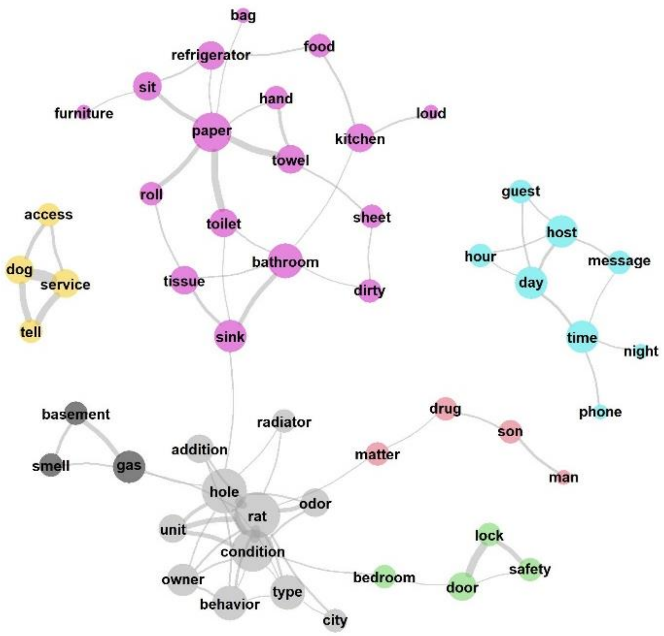


Figure 11. Network diagrams of negative reviews post-covid from NY.

4.2 Analysis of covid-reviews

A total of 1222 reviews from New York and a total of 85 reviews from Rio de Janeiro contained covid-terms. This represents a proportion of 0.32% of the total of reviews from Rio de Janeiro and 0.25% of the total from New York. A total of seven covid-terms were identified in the reviews, and the covid-terms found in Rio de Janeiro are the same found in New York. The most frequent covid-term found in reviews from both cities is "Pandemic", however, the second most frequent covid-term found in reviews from Rio de Janeiro was "quarantine" followed by "corona", and in reviews from New York was "covid", followed by "mask" and "quarantine (see **Table 12**).

Table 12. Frequency of covid terms

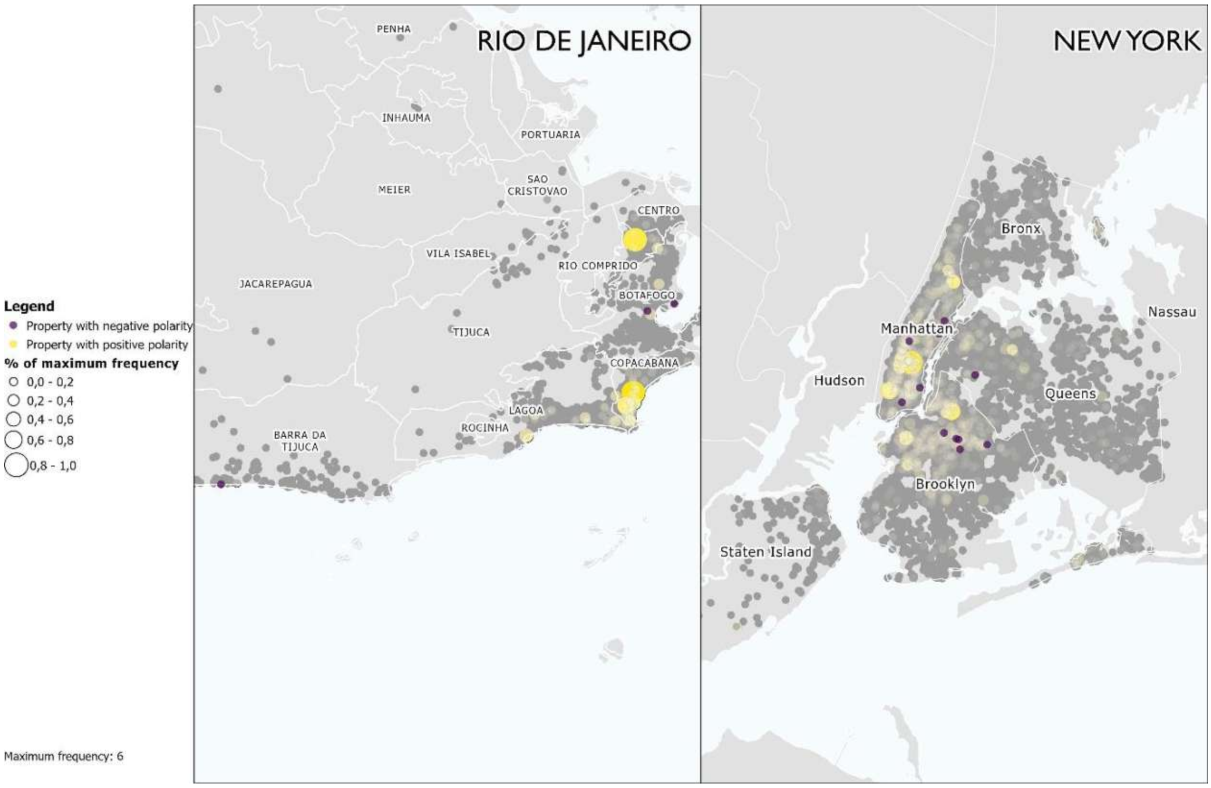
	Rio de Janeiro		New York	
Covid-term	Count	Percent (%)	Count	Percent (%)
corona	16	15	58	4
coronavirus	13	12	99	7
covid	14	13	319	23
lockdown	4	4	59	4
mask	9	8	132	10
pandemic	34	31	579	42
quarantine	20	18	136	10
Total	110	100	1382	100

Most of the covid-reviews were positive, as they accounted for 87% and 89% of the total covid-reviews from Rio de Janeiro and New York respectively. As for negative reviews, they represent 13% and 10,3% of the covid-reviews from Rio de Janeiro and New York respectively, and neutral ones only occurred in covid-reviews from New York, representing 0.4% (**Table 13**).

Table 13. Number and percentage of covid reviews per sentiment polarity from properties in RJ and NY.

	Rio de Janeiro		New York	
Sentiment	Count	Percent (%)	Count	Percent (%)
Positive	74	87	1098	89,3
Neutral			5	0,4
Negative	11	13	126	10,3
Total	85	100	1229	100

Covid-reviews were posted from 2,3% of the properties in Rio de Janeiro and 5,7% of the properties in New York. Map x shows the distribution of properties where covid-reviews were posted (covid-properties) coloured according to the average polarity of the property, here properties with neutral polarity were not considered as they only occurred in New York. The size of the circles indicates the number of times a covid term was mentioned in the reviews of the property but normalized by the maximum frequency in which they were mentioned in both cities, which allows to compare between cities. We can see that covid properties with positive polarity followed more a cluster distribution, while covid properties with negative polarity followed more random distribution.



Map 19. Spatial distribution of properties in RJ and NY where covid terms where mentioned

4.2.1 Analysis of frequent keywords

Rio de Janeiro, Brazil

Sixteen different terms represent the 10 most frequent keywords of positive and negative covid-reviews from property listings in Rio de Janeiro (**Table 14**). The top of positive and negative covid-reviews share 25% of these terms ("place", "apartment", "host" and "time"), however the term "host" is 35% more frequent in the negative covid-reviews than in the positive ones.

Table 14. List with 10 most frequent terms from covid reviews of properties in RJ.

	Positive			Negative		
No.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
1	**place**	72	1	**apartment**	14	1
2	**apartment**	55	0,76	**host**	13	0,93
3	stay	31	0,43	owner	11	0,79
4	location	27	0,38	**place**	10	0,71
5	**host**	26	0,36	house	9	0,64
6	beach	23	0,32	air	9	0,64
7	day	17	0,24	people	8	0,57
8	**time**	16	0,22	rule	7	0,50
9	month	15	0,21	refund	7	0,50
10	view	15	0,21	**time**	7	0,50

New York, U.S

In New York, on the other hand, twenty-one different terms represent the 10 most frequent keywords of positive, neutral, and negative covid-reviews (**Table 15**), 9,5% of these terms are shared by the top of positive, neutral, and negative covid-reviews ("room", "day"), whereas 25% are shared by the top of positive and negative covid-reviews ("place", "apartment", "stay", "host" and "time"). Like Rio de Janeiro, the term **host** is more frequently mentioned in negative covid-reviews than in positive ones. From being in the fourth position in the top of positive covid-reviews, without appearing in the top of neutral ones, it became number one in the top of negative covid-reviews. The term **room** is also found more frequently in the negative and neutral covid-reviews than in the positive ones. From being in the seventh position in top of positive covid-reviews, it occupied the second position in the top of neutral as well as negative covid-reviews. The term **day** was also more frequently mentioned in neutral as well as negative covid-reviews. From being in the last position in the top of positive covid-reviews it became number four in the top of neutral as well as negative covid reviews. Differently from the 10 most frequent words of positive and negative covid-reviews from Rio de Janeiro, that did not have any covid-term, **covid** occupied the third position of the top of neutral covid-reviews.

Table 15. List with 10 most frequent terms from covid reviews of properties in NY.

	Positive			Neutral			Negative		
No.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.	Term	Freq.	R. Freq.
1	**place**	1038	1	bed	3	1	**host**	104	1
2	**apartment**	582	0,56	**room**	3	1	**room**	96	0,92
3	**stay**	515	0,50	***covid***	3	1	**place**	96	0,92
4	**host**	478	0,46	**day**	2	0,67	**day**	90	0,87
5	**time**	473	0,46	king	2	0,67	**time**	76	0,73
6	location	368	0,35	cabin	1	0,33	**apartment**	76	0,73
7	**room**	353	0,34	ceiling	1	0,33	**stay**	60	0,58
8	space	341	0,33	hostel	1	0,33	night	53	0,51
9	home	280	0,27	dorm	1	0,33	people	46	0,44
10	**day**	263	0,25	plug	1	0,33	issue	37	0,36

4.2.2 Analysis of rare keywords

Rio de Janeiro, Brazil

The 10 most rare keywords of positive covid-reviews from property listings in Rio de Janeiro have no terms in common with the 10 most rare keywords of negative ones (**Table 16**). While no covid-term was found in the top of positive covid-reviews, **covid** and **quarantine** occupied the first and second position of the top of negative ones.

Table 16. List with 10 most rare terms from covid reviews of properties in RJ.

	Positive		Negative	
No.	Term	Score	Term	Score
1	wait	0,75	***covid***	0,476
2	condo	0,73	***quarantine***	0,456
3	serve	0,72	home	0,456
4	house	0,69	friend	0,456
5	charm	0,67	apartment	0,451
6	begin	0,65	check	0,418
7	help	0,64	travel	0,407
8	deck	0,58	value	0,391
9	window	0,57	iron	0,391
10	studio	0,57	ice	0,391

New York, U.S

Almost like Rio de Janeiro, the 10 most rare keywords of positive, neutral, and negative covid-reviews from New York have no terms in common except for **covid** (**Table 17**), which went to be the third last rare term in the top of positive covid-reviews to occupy the fourth position in the top of neutral covid-reviews and the first position in the top of the negative ones. Apart from **covid**, another covid-term, **coronavirus**, occupied the sixth position in the top of positive covid-reviews.

Table 17. List with 10 most rare terms from covid reviews of properties in NY.

	Positive		Neutral		Negative	
No.	Term	Score	Term	Score	Term	Score
1	home	1	stay	0,84	***covid***	1
2	eats	1	day	0,65	property	0,79
3	birthday	1	king	0,57	cancellation	0,78
4	cleaner	0,96	***covid***	0,54	scruple	0,76
5	value	0,94	room	0,45	crew	0,74
6	***coronavirus***	0,93	bed	0,45	conference	0,70
7	guideline	0,92	sleep	0,35	unit	0,69
8	**covid**	0,90	plug	0,35	people	0,67
9	thanks	0,89	hostel	0,35	rental	0,65
10	umbrella	0,86	eye	0,35	host	0,64

4.2.3 Analysis of network diagrams

Rio de Janeiro, Brazil

Positive covid-reviews were represented by six different communities of keywords (see **Figure 12**), while negative covid-reviews were represented by five (see **Figure 13**). The communities from positive covid-reviews contain descriptors of the experience, host' service, facilities in the neighbourhood and attributes of the property, but rather than been grouped in each community they are mixed together and some of them are related with the pandemic situation ("crisis", "travel" and "covid", "owner").

On the other hand, communities from negative covid-reviews contained more descriptors of the experience and some of them were also related with the pandemic (“travel”, “ban”, “covid”, “state”, “cancel”, “flight”, “crisis”, “country”).

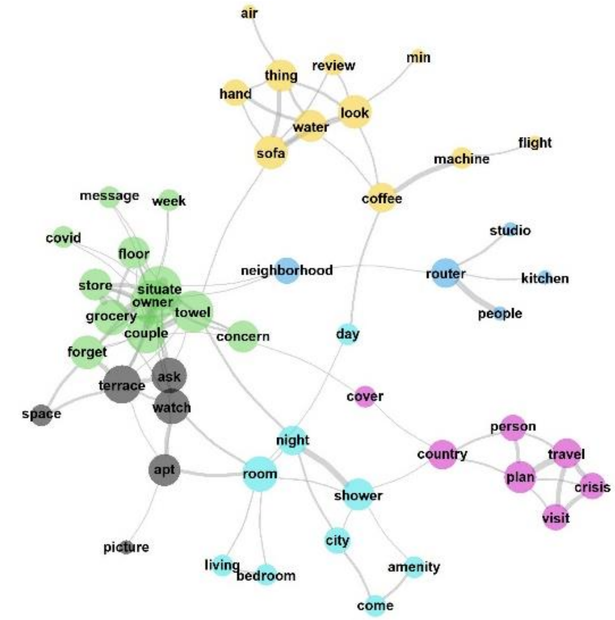


Figure 12. Network diagram of positive covid-reviews from RJ.

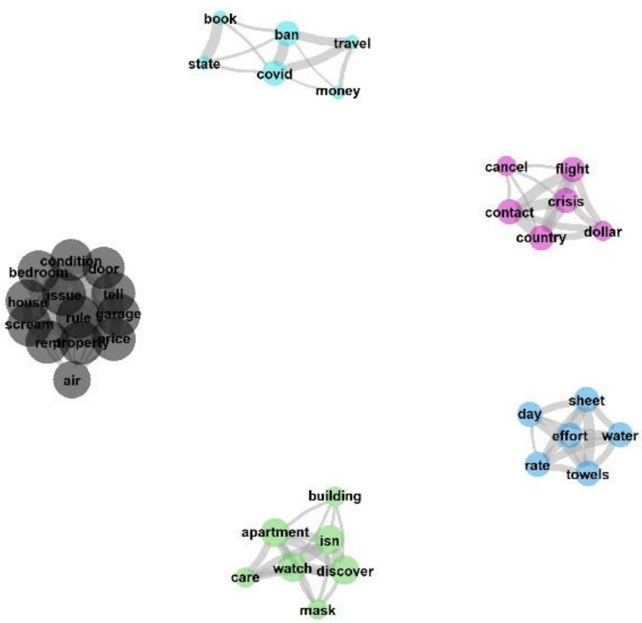


Figure 13. Network diagram of negative covid-reviews from RJ.

New York, U.S

Eight communities of keywords represented positive covid-reviews (see **Figure 14**), while seven communities represented negative ones (see **Figure 15**). Communities from positive covid-reviews contained mostly descriptors of the experience and some of them were related with the pandemic situation (“hand”, “sanitizer”, “mask”).

They also described specific issues with mattress, cleanliness, and cancellation of reservation. On the other hand, communities from negative reviews described also issues related with infestation and cleanliness, indicated hygiene and safety concern, and contained descriptors of the experience.

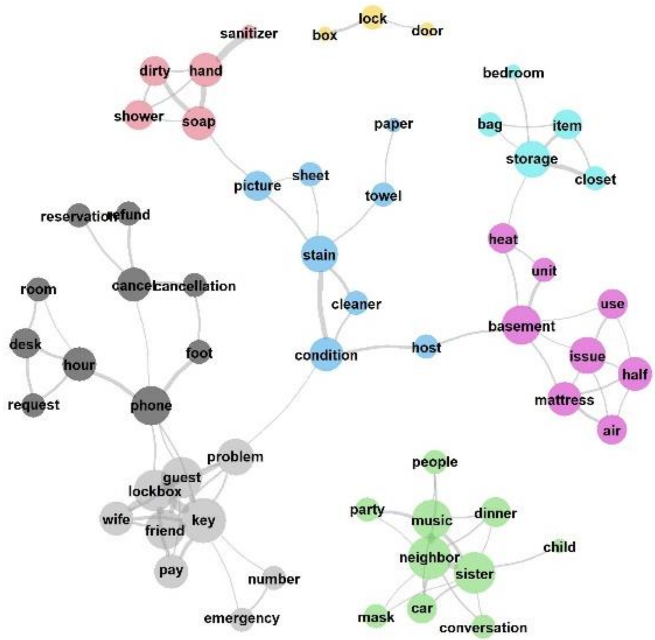


Figure 14. Network diagram of positive covid-reviews from NY.

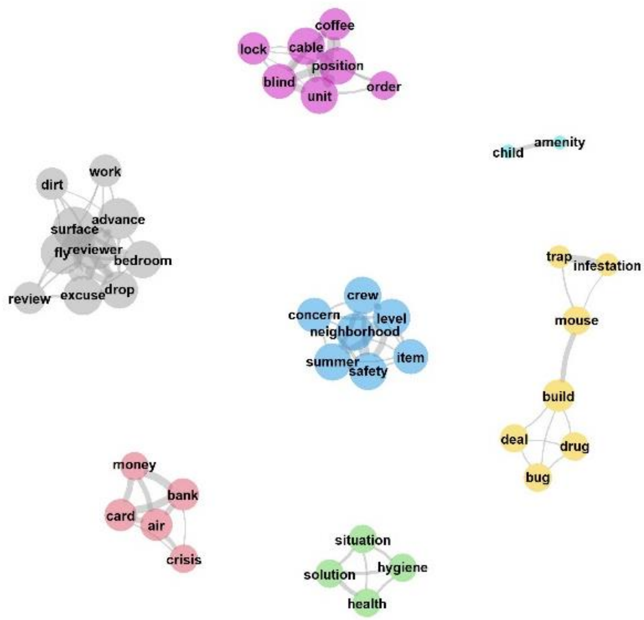


Figure 15. Network diagram of negative covid-reviews from NY.

5 Conclusions

After one year of the outbreak of COVID-19, the analysis of reviews according to the sentiment polarity did not provide enough evidence to identify a significant change in the experience of Airbnb users in Rio de Janeiro as well as in New York. Nevertheless, the analysis of reviews that contained of covid-terms, indicates that users in both cities experienced situations related with the pandemic, such as the use of mask and hand sanitizer, cancellation of reservation and flight, and travel ban. However, this situations appeared in the network diagram of positive and negative reviews, and furthermore, positive covid-reviews accounted for 87% of the total covid-reviews from Rio de Janeiro and 89% of the covid-reviews from New York.

The sentiment analysis indicates that after the outbreak of covid, relatively less reviews from Rio de Janeiro and New York were positive, and more were negative. Positive reviews from Rio de Janeiro declined 1% and from New York 2%, whereas negative reviews from both cities increased 1%. Furthermore, that analysis shows that Airbnb users in both cities and in both time-periods had a positive bias in their assessment of the experience, since positive reviews were never less than 94% of the total reviews.

According to the sentiment polarity, the analysis of frequent keywords shows that **location**, **host**, **room**, **day**, and **night** were key elements of the experience that had in common users in Rio de Janeiro and in New York, not only before the outbreak of covid but also after it. Furthermore, the analysis reveals that **restaurant**, **beach**, and **view** were only key elements in the experience of users in Rio de Janeiro, before and after the outbreak of covid, since they only appeared in the list of the top frequent keywords of Rio de Janeiro, while **subway** and **station** were only key elements in the experience of users in New York, but only before the outbreak of covid, as they as well, only appeared in the list of the top frequent keywords of New York.

The analysis also shows that there are elements of the user experience that only played a key role in the negative assessment depending on the city and time period. For instance, **bathroom** was only a key element in the experience of users in New York before the outbreak of covid, while, after the outbreak, **water** and **people** were only key elements in the experience of users in Rio de Janeiro and **door** in the experience of users in New York. Further analyses are required to verify whether these elements could be linked to issues derived from the pandemic.

The analysis of rare keywords, on the other hand, revealed that the **Wi-Fi** played a key role for some users in Rio de Janeiro only before the outbreak of covid, and for some users in New York only after it. This analysis also indicates that the **view** also played a key role in the experience of some users in New York, however only after the outbreak of covid. Other elements played a key role in the experience of some users depending on the city, time-period, and polarity.

The analysis of network diagram supports the findings of the keyword analysis and uncovered other key elements of the user experience such as the service and the convenience of the location. Furthermore, it reveals that dirtiness is a constant issue encountered by users in Rio de Janeiro and in New York.

As for the findings from the analysis of frequent keywords from covid-reviews, they did not highly differ from the findings of this analysis when applied to all the reviews. Positive reviews from both cities were also mainly dominated by a description of the property (e.g., place, apartment, space, home), host, and its location, and negative reviews also provided other key elements of the experience. Negative reviews from Rio de Janeiro very frequently considered **refund** and **rule**, and from New York, the topic **people**. At the same time, the analysis of rare keywords showed that **quarantine** was a frequent topic of negative reviews from some users in Rio de Janeiro, as well as **cancellation** was of negative reviews from some users in New York.

As for the spatial analysis of the sentiment polarity, it indicates that before the outbreak of covid, in Rio de Janeiro, average positive, neutral, and negative experiences were highly concentrated in the same area, whereas in New York, they were in the same borough but positive ones over two areas, and negative and neutral ones over one and shared the same area. After the outbreak of covid, the decline of properties with positive polarity and the increase of properties with neutral and negative polarity, was displayed differently by each city. In Rio de Janeiro it was reflected in an increase of the concentration of average neutral and negative experiences in the same area with the highest density of positive, neutral, and negative experiences, on the other hand, in New York it was reflect as a movement of the area with highest concentration of average neutral experiences towards Brooklyn, and the emergence of a new area with high concentration of average negative experiences.

The location of properties where Airbnb users mentioned covid-terms differed according to the average sentiment polarity of the property in both cities. When it comes to properties with positive polarity, they followed a cluster distribution, however properties with negative polarity followed a random distribution.

These findings show how semantic and spatial analysis can contribute to the understanding of customer behaviour in the lodging sector, and furthermore during health crisis events. However, this study is not without limitations. The results cannot be generalized as they only considered the experience of Airbnb users in two cities and only expressed in English language. Furthermore, the analysis of sentiments is limited to the document level, which overlooks part of reviews that might have contrasting polarities, and which does not allow to identify the target of the sentiment, nevertheless a future research can implement a sentiment analysis at the aspect level to overcome this. Even though the analysis of keywords and network diagrams allowed to identify some topics, it was not very conclusive, since it is required a proper interpretation of the context in which keywords were mentioned, nevertheless it could be overcome by employ trained models for topic modelling.

References

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, *58*(2), 175–191. https://doi.org/10.1177/0047287517747753

Anandarajan, M., Hill, C., & Nolan, T. (2019). *Practical Text Analytics* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-95663-3_7

Ban, H. J., Choi, H., Choi, E. K., Lee, S., & Kim, H. S. (2019). Investigating key attributes in experience and satisfaction of hotel customer using online review data. *Sustainability (Switzerland)*, *11*(23). https://doi.org/10.3390/su11236570

C.S. Siu, R. (2019). *Economic Principles for the Hospitality Industry* (1st Editio). Routledge. https://doi.org/https://doi.org/10.4324/9781315108520

Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, *76*(September 2017), 58–70. https://doi.org/10.1016/j.ijhm.2018.04.004

Chittiprolu, V., Samala, N., & Bellamkonda, R. S. (2021). Heritage hotels and customer experience: a text mining analysis of online reviews. *International Journal of Culture, Tourism, and Hospitality Research*, *15*(2), 131–156. https://doi.org/10.1108/IJCTHR-02-2020-0050

Escandon-Barbosa, D., & Salas-Paramo, J. (2021). Tourism Amidst COVID-19: consumer experience in luxury hotels booked through digital platforms. *Tourism Recreation Research*, *0*(0), 1–6. https://doi.org/10.1080/02508281.2021.1940708

Fortin, M.-J. (2017). Spatial Analysis of Ecological Data. In *Encyclopedia of GIS*. https://doi.org/10.1007/978-3-319-17885-1

Galati, F., & Galati, R. (2019). Cross-country analysis of perception and emphasis of hotel attributes. *Tourism Management*, *74*(November 2018), 24–42. https://doi.org/10.1016/j.tourman.2019.02.011

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. *Tourism Management*, *61*, 43–54. https://doi.org/10.1016/j.tourman.2016.12.022

Gursoy, D., & Chi, C. G. (2020). Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda. *Journal of Hospitality Marketing and Management*, *29*(5), 527–529. https://doi.org/10.1080/19368623.2020.1788231

Hall, C. M., Scott, D., & Gössling, S. (2020). Pandemics, transformations and tourism: be careful what you wish for. *Tourism Geographies*, *22*(3), 577–598. https://doi.org/10.1080/14616688.2020.1759131

Hodson, J. (2017). Encyclopedia of Machine Learning and Data Mining. In *Encyclopedia of Machine Learning and Data Mining*. https://doi.org/10.1007/978-1-4899-7687-1

Hu, F., Teichert, T., Deng, S., Liu, Y., & Zhou, G. (2021). Dealing with pandemics: An investigation of the effects of COVID-19 on customers' evaluations of hospitality services. *Tourism Management*, *85*(July 2020), 104320. https://doi.org/10.1016/j.tourman.2021.104320

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, *72*(January), 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, *May*, 216–225.

Hwang, J., & Seo, S. (2016). A critical review of research on customer experience management: Theoretical, methodological and cultural perspectives. *International Journal of Contemporary Hospitality Management*, *28*(10), 2218–2246. https://doi.org/10.1108/IJCHM-04-2015-0192

Inside Airbnb. (2021). About. http://insideairbnb.com/about.html

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (n.d.). *FASTTEXT.ZIP: COMPRESSING TEXT CLASSIFICATION MODELS*. Retrieved October 12, 2021, from https://github.com/facebookresearch/fastText

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). *Bag of Tricks for Efficient Text Classification*. https://github.com/facebookresearch/fastText

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Li, H., Ye, Q., & Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research*, *18*(7), 784–802. https://doi.org/10.1080/10941665.2012.708351

Lillo, A. (2020). COVID-19, the beer flu; Or, the disease of many names. *Lebende Sprachen*, *49*(5), 411–438. https://doi.org/10.1515/les-2020-0021

Liu, B. (2015a). Introduction. In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (pp. 1–15). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9781139084789

Liu, B. (2015b). Preface. In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (pp. 162–174). https://doi.org/10.4018/978-1-5225-2031-3.ch009

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Mehta, M. P., Kumar, G., & Ramkumar, M. (2021). Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis. *Tourism Recreation Research*, *0*(0), 1–18. https://doi.org/10.1080/02508281.2021.1894692

Moro, S., Lopes, R. J., Esmerado, J., & Botelho, M. (2020). Service quality in airport hotel chains through the lens of online reviewers. *Journal of Retailing and Consumer Services*, *56*(July). https://doi.org/10.1016/j.jretconser.2020.102193

Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Current Issues in Tourism*, *23*(5), 605–611. https://doi.org/10.1080/13683500.2018.1549025

Phillips, P., Antonio, N., de Almeida, A., & Nunes, L. (2020). The Influence of Geographic and Psychic Distance on Online Hotel Ratings. *Journal of Travel Research*, *59*(4), 722–741. https://doi.org/10.1177/0047287519858400

Sammut, C. (2017). Density Estimation. In *Encyclopedia of Machine Learning and Data Mining*. https://doi.org/10.1007/978-1-4899-7687-1

Serrano, L., Ariza-Montes, A., Nader, M., Sianes, A., & Law, R. (2020). Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: a text mining approach. *Journal of Sustainable Tourism*, *0*(0), 1–19. https://doi.org/10.1080/09669582.2020.1838529

Silverman, B. . (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Tussyadiah, I. P., & Zach, F. (2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel and Tourism Marketing*, *34*(5), 636–652. https://doi.org/10.1080/10548408.2016.1209153

Wood, R. C. (2015). What Kind Of Industry ? Structure , Strategy And The Nature Of The Hospitality Industry. In *Hospitality Management : A Brief Introduction* (pp. 20–35). SAGE Publications Ltd.

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013

Zhang, J. (2019). What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, *36*(5), 655–665. https://doi.org/10.1108/JCM-02-2018-2581