Aston University

BUSINESS PROJECT

ANALYSIS OF AIRBNB IN THE CITY OF NEW YORK

BDM163

MSc Business Analytics

SEPTEMBER 2023

Student number : 220269920

Student Name : Ahmedmiya Kadri

Written under the supervision of Dr. Majid Azadi

Acknowledgement

I want to extend my sincere acknowledgment to Mazid Azadi, my project supervisor, for his crucial advising, supporting me consistently , and insights with knowledge during the course of this project. The examination of Airbnb in metropolitan of the New York City was considerably affected by their in-depth knowledge and skills in data analysis, their advice on data analysis methods, research methodology, and intelligent ideas improved the quality of this endeavour.

Throughout the duration of this education, their persistent support and commitment helped to keep me on track and motivated. I sincerely appreciate their mentoring, and I recognise the important role they had in the project's accomplishment.

Declaration

Declaration Include the following statement on the same page as the acknowledgments. I declare that I have personally prepared this report and that it has not in whole or in part been submitted for any other degree or qualification. Nor has it appeared in whole or in part in any textbook, journal or any other document previously published or produced for any purpose. The work described here is my/our own, carried out personally unless otherwise stated. All sources of information, including quotations, are acknowledged by means of reference, both in the final reference section and at the point where they occur in the text.

*

Abstract

The sharing economy has expanded quickly, revolutionising how people travel and locate places to stay. In this setting, Airbnb has become a dominating force, changing the hospitality sector, and offering travellers unusual housing alternatives. Due to its popularity as a travel destination, New York City has seen a considerable increase in the number of Airbnb listings, which has created both possibilities and difficulties for the local property market and the hotel sector.

The goal of this study is to conduct a thorough analysis of the New York City Airbnb industry and pinpoint the major variables affecting patron satisfaction. The sharing economy's emergence has completely transformed hospitality business, with Airbnb rising as an important player in the accommodation industry. In order to give a thorough examination of Airbnb's existence, influence, and user satisfaction levels, this research project looks into the varied Airbnb scene in New York City. The project develops along a number of related aspects. It begins by looking at how Airbnb listings are distributed geographically within New York City, highlighting the areas with the greatest number of listings. Understanding the platform's reach and use inside the city is based on this geographical study.

When it comes to the availability and distribution of Airbnb listings, socioeconomic characteristics are crucial. The study looks at the relationships between factors including population density, and neighbourhood demographics and the frequency of Airbnb lodgings in various locations. This study sheds light on the socioeconomic forces influencing how people use the platform. One of the objectives of this study is customer happiness. We analyse the elements that affect consumers' enjoyment with Airbnb in the setting of New York City using analysis of sentiment and feedback from customers. We also provide actionable analytics which will help to improve the user experience by identifying the key factors affecting happiness. By profiling various hosts and locations, the study reveals patterns and trends among hosts and their properties. Forecasting factors like pricing patterns and review scores using predictive modelling offers useful insight. The initiative also identifies the busiest hosts in the city and investigates the underlying causes of their high levels of activity. Understanding these host types may provide both new and seasoned Airbnb hosts with useful advice.

To conclude, this study offers a comprehensive examination of Airbnb in New York City, including information on its geographic spread, socioeconomic effects, customer happiness, impact on conventional hotels, financial ramifications, host and neighbourhood profiles, and predictive insights. This study adds to a fuller understanding of Airbnb's function in the city and the implications for the larger hospitality industry by exploring these aspects.

Contents

Chapter 1

1.1 Introduction

Since its launch, the Airbnb business has grown rapidly, upending the conventional hotel sector by matching guests with distinctive lodging provided by private hosts. Increased customer demand for personalised travel experiences, cost savings, and the desire for more genuine and local stays are some of the drivers driving the market's rise. In order to accommodate a wide range of travel preferences and budgets, the worldwide Airbnb business has grown to encompass thousands of cities across several nations.

With a wide variety of properties listed on the site, including flats, homes, villas, and unusual accommodation alternatives like treehouses and boats, the worldwide Airbnb business has experienced substantial development. The market size varies by area, with larger levels of Airbnb activity being observed in big cities and well-known tourist sites. The top three regions for both Airbnb listings and guest arrivals are North America, Europe, and Asia-Pacific.

The Airbnb market in New York has experienced substantial growth in recent years, fueled by the city's popularity as a major tourist destination and its vibrant neighbourhoods. Thousands of properties, including apartments, houses, lofts, and unique accommodations, are listed on the Airbnb platform, catering to diverse traveller preferences and budgets. The market provides an alternative to traditional hotel stays, offering visitors a chance to experience the city like a local.

With a broad selection of houses available for short-term rentals, New York City is home to one of the largest Airbnb marketplaces in the world. The scale of the Airbnb market varies by neighbourhood, with Manhattan, Brooklyn, and Queens being the most sought-after locations. There are a variety of property types available to suit different travel needs, from small studio flats in busy metropolitan locations to large luxury mansions in residential settings.

The varying demand and the city's varied neighbourhoods are reflected in the pricing dynamics of the New York Airbnb market. Based on elements including location, property size, amenities, and seasonal patterns, prices can vary widely. Major tourist destinations and commercial districts are typically near to one another, with costs in the latter being higher. Holidays, significant events, and seasonal fluctuations all affect price.

The regulatory environment in which the Airbnb market in New York City functions is complicated. To address issues with housing availability, tenant protections, and the effect of short-range leasing on the local housing souk, the city has put in place rules. These rules include restrictions on the length of rentals, the need for hosts to be present while guests are there, and prohibitions on particular kinds of homes. To guarantee compliance and prevent potential penalties, hosts must understand and follow these rules.

Both guests and hosts have access to special options thanks to New York's Airbnb market. Travellers have access to a broad range of lodging options in desirable areas, sometimes for less money than traditional hotels. They may have individualised stays and get a local's viewpoint on the city. In particular, if they take advantage of excellent visitor evaluations and provide outstanding hospitality experiences, hosts profit by monetizing their properties and generating additional cash.

The global Airbnb market has grown as a result of many causes. These consist of: Demand for travel is rising, as is the desire for unusual travel experiences. Economic aspects, including cost savings for visitors and increased revenue prospects for hosts. Technological developments, as well as the simplicity of reserving and communicating online via the Airbnb platform. Consumer attitudes and behaviour are shifting in favour of more individualised and regional travel experiences.

The Airbnb industry has brought forth a lot of advantages, but it also has issues and problems. These comprise regulatory difficulties and conflicts with regional laws in different regions. A negative effect on the conventional hospitality industry, which includes hotels and bed and breakfasts. Problems with security, safety, and quality assurance for all Airbnb rentals. There are worries about housing affordability and possible housing shortages in some areas as a result of lasting-time fee houses being turned into temporary Airbnb leasing.

But there are also opportunities in the Airbnb industry for other stakeholders: Travellers have access to a wide variety of lodging options that provide original and genuine experiences. Property owners may make money from their underused or unoccupied properties by selling them. Increased tourist and community expenditure can boost local economies. Governments can look for methods to control and tax the Airbnb business while boosting tourism and juggling the requirements of the local population.

The hospitality industry in New York City has changed as a result of the Airbnb market, which gives visitors a distinctive and different perspective on the city. The market's expansion reflects the shifting tastes of tourists looking for genuine, regional experiences. Although the market offers chances for hosts and guests, it also creates difficulties because of home availability, legal requirements, and neighbourhood issues. For the Airbnb market in New York City to expand sustainably and have a beneficial influence, it is essential to strike a balance between its advantages and these difficulties. To create legislation that encourage responsible hosting practises, safeguard affordable housing, and guarantee a peaceful coexistence between short-term rentals and the local community, policymakers, industry partners, and community members must cooperate cooperatively.

1.2 Problem of Research

1. How are Airbnb listings distributed geographically in New York City, and which areas have the largest concentration of listings?

2. What effects do socioeconomic variables have on the availability and distribution of Airbnb listings in various neighbourhoods?

3. What are the primary elements influencing New York City users' happiness with the Airbnb platform?

4. What effects does Airbnb's existence have on the conventional hotel sector, in terms of market shares, occupancy rates, and average daily prices?

5. What may the growth of Airbnb in New York City mean for the local economy?

6. What can we learn about different hosts and areas?

7. What do we learn from the predictions like prices, reviews etc.

8. Which hosts in the city are the busiest and what is the reason behind it?

9. Which room type is preferred in the most popular neighbourhood?

1.3 Objective of the Research

This project's main goal is to undertake a thorough examination of Airbnb's existence and effects in New York City. With its focus on the local economy, housing market, and customer happiness, this research intends to provide light on a variety of facets of Airbnb's operations. The initiative aims to by addressing certain research topics such as Examine the geographic distribution of Airbnb listings in New York City's neighbourhoods to find the regions with the greatest numbers of listings.

Examine Demographic Factors: To identify any potential discrepancies, look at how socioeconomic factors affect the availability and distribution of Airbnb listings in different neighbourhoods. Determine the main characteristics that lead to guest happiness with the Airbnb platform in New York City, and then identify the critical components that underpin positive encounters. Identify patterns and trends among Airbnb hosts and their homes to gain insight into the traits of various hosts and neighbourhoods. The busiest hosts in the city should be identified, and the causes of their high activity levels should be investigated. Analyse user preferences for different room kinds in the most well-liked areas to provide you information about lodging options.

To gain a better insight of user experiences and preferences, do sentiment analysis to determine user sentiment and satisfaction levels. This study intends to provide a thorough knowledge of Airbnb's influence on New York City's housing and hospitality scene as well as the elements influencing customer satisfaction and property demand in the Airbnb market by addressing these research goals. The results will provide important new information for both research and real-world applications in the travel and hospitality sectors.

1.4 Research contribution

The results of this study will add to the body of evidence already available about how the sharing economy, namely Airbnb, has affected the real estate market and hospitality sector in New York City. Policymakers, local governments, and stakeholders can use the findings to guide their decisions about rules, taxes, and policies for sustainable growth. Additionally, Airbnb hosts and platform operators may improve the visitor experience and further optimise their products with the help of the information collected through customer satisfaction study. In the end, this study seeks to promote a fair strategy that maximises the advantages of the Airbnb market while minimising any drawbacks in the larger context of the city's hotel industry. The following are some of the significant contributions that the proposed study on the New York City Airbnb business and factors influencing customer satisfaction will bring to the field's body of knowledge.

Research Objective 1:

Comprehensive Understanding of the Airbnb Market: The investigation will offer an in-depth knowledge of the Airbnb market dynamics in the city by conducting a thorough analysis of the distribution which is spatial of Airbnb listings in metropolitan of the New York City and looking at the socioeconomic factors influencing their presence. This will make it easier to pinpoint the places with the greatest concentration of listings and the elements influencing Airbnb's expansion in particular regions.

Research Objective 2:

Insights into satisfied customers elements: The study will determine the major elements that affect customer satisfaction on the Airbnb platform using sentiment analysis of customer reviews and comments. Recognising these aspects will help Airbnb hosts in New York City better their services, but it will also be helpful for the platform as a whole to increase customer experience.

Research Objective 3:

Strategy Recommendations: Based on the results of the research, local governments and policymakers may want to consider the following measures. It will be possible for policymakers to create suitable rules and taxation strategies to promote fair competition and sustainable growth in the hospitality sector if they have an understanding of the impacts of Airbnb's rise on the housing market and conventional hotels.

Research Objective 4:

The contribution to the Sharing Economics Literature: The study will expand on the little amount of empirical research that has already been done on Airbnb's effects in New York City. The study will offer a thorough picture of the impact of the sharing economy in the city's hospitality environment by studying both the favourable and unfavourable elements of Airbnb's existence.

Research Objective 5:

Generalizability: Although the research is centred on the metropolitan area of New York, the techniques used, and the conclusions drawn may be applied to other cities and areas dealing with the same possibilities and difficulties because of the growth of Airbnb and the economy of sharing.

The study contribution will, in general, offer insightful information to a range of stakeholders, including policymakers, local governments, the conventional hotel business, Airbnb hosts, and the larger academic community. The results of the study will aid in developing a fair and well-informed strategy for controlling the effects of Airbnb's expansion in New York City, which will eventually be advantageous to both the tourism sector and locals.

1.5 Summary table for the objectives

Objective 1	**Comprehensive Understanding of the Airbnb Market :** - By performing thorough analysis of spatial distribution and also considering the economic and social factors influencing the market.
Objective 2	**Gaining detailed insights on what factors satisfies a customer.** - By performing analysis of sentiment by analyzing customer reviews and comments.
Objective 3	**Recommending the strategies for customer retention and increasing the flow of customers:** - By analyzing the factors affecting the good and bad experiences of the customer we would successfully able to recognize the factors and suggest the key factors to the airbnb listers.
Objective 4	**To contribute the research for the hospitality sector and identifying the favorable and unfavorable factors:** - By running Logistic regression which is ordinal, we would be able to segregate the favorable and unfavorable factor's which can further would help with this sector.
Objective 5	**To make and increase the generalizability of this research for other locations.** - To use techniques which can be used for other locations and cities which would further increase the generalizability of the research.

1.6 Literature review

1. An examination of Airbnb's social and economic effects in New York.

Gabor Dudas analyzed the socio-economic condition of Air-bnb in New York city in order to map the presence of Airbnb's in the city. There were several factors which motivated his proposal like there were high number of Airbnb in the city, criticism and media attention, etc, the problem he tried to solve here is to understand how the spatial distribution is influenced by socio-economic factors and its underlying processes. The methodology followed here was The Inside Airbnb website, which offers geolocated Airbnb data for several places, including New York City, is where the researchers acquired comprehensive information on Airbnb properties.

The dataset contains facts on the location, the host, the kind of room, the price, the availability, and reviews. Additionally, the American Community Survey (ACS), which includes information on demographic, social, economic, and housing variables, was used to gather socio-economic data about the neighbourhoods of New York City. The researchers gathered housing and economic statistics as well as the relevant data at the Neighbourhood Tabulation Areas (NTA) level. To assess the attractiveness of various locations based on pertinent areas including eating and drinking, attractions, shopping, and sports and entertainment, points of interest (POI) data were collected from OpenStreetMap. The field of application of this research is concentrated on examining the geographic distribution of Airbnb, determining the key socioeconomic traits of regions having Airbnb listings, and comprehending the elements that draw Airbnb users. According to the data study, most of the postings (89%) are concentrated in a small number of Brooklyn and Manhattan neighbourhoods. Particularly, there are more postings in the southern and northern parts of Manhattan and Brooklyn. According to my research point of view, the possible impact of Airbnb on the neighbourhood housing market is not discussed in this research. According to research, the existence of Airbnb may have an impact on housing affordability and availability since some hosts may turn long-term rental properties into short-term lodging. The analysis makes no mention of the rules governing Airbnb in New York City. A more thorough examination of the Airbnb phenomena in the city may be provided by being aware of the regional laws, limitations, and enforcement initiatives that apply to short-term rentals. The study focuses on supply-side elements including the quantity of listings, varieties of rooms, and host traits. However, it would be advantageous to also take into account the viewpoints and experiences of Airbnb consumers. Although socioeconomic data are taken into account in the study, the paragraphs do not go into detail on the particular qualities of the areas where Airbnb listings are most prevalent. (Dudas, 2017)

2. The latest situation and evolution of Airbnb's lodging options in 166 different countries.

Czeslaw Adamiak analyzed the current state and development of Airbnb accommodation offer in different countries. The study's goal is to examine and comprehend the peculiarities of Airbnb postings in various nations. The authors want to look at what kinds of houses are being rented out, how many offers are hosted by specific platform users, and what factors affect how Airbnb listings are distributed. Also, The goal of the study is to comprehend how the supply of Airbnb relates to variables like economic growth, inbound tourism, metropolitan areas, and coastal regions. The study also intends to examine the expansion of Airbnb in various markets, the function of professional hosts as opposed to peer-to-peer hosts, and the regional differences in the advantages and effects of Airbnb activity. The problem here is to make Differentiating across nations' capacities, usage patterns, costs, and levels of satisfaction with Airbnb accommodations: The researchers want to improve the interpretation of the findings from regional studies on related subjects by looking at the worldwide variation of these properties of Airbnb listings. The study uses web-scraped data from 3,580,044 active Airbnb listings in 167 countries to achieve these objectives, making it the biggest web-scraped dataset ever utilised in an Airbnb study. According to type (single-room, single-home, multi-room, and multi-home listings) and the quantity of offers made by a single host, the listings are divided into categories. The empirical portion of the article analyses nations based on the variety of listings, dispersion in respect to tourist resources, expansion of the Airbnb rental stock, capacity, frequency of usage, pricing, and the happiness of guests. This study's field of application, which focuses on the size, organisation, distribution, dynamics, and usage of Airbnb rooms in 167 countries, describes Airbnb activity on a worldwide scale. The study's results include mapping the structure and distribution of Airbnb offers globally, determining the socioeconomic factors that influence Airbnb activity Analysis of variations in Airbnb apartments' capacity, use frequency, costs, and satisfaction.

Several factors might be studied or taken into account in the study, including the Impact on neighbourhood communities The possible effects of Airbnb on local communities are not specifically mentioned in the lines. Regulatory environment the regulatory environment around Airbnb in various nations is not discussed., Data and methodology restrictions: The usage of a huge dataset and the web-scraping technology are just briefly mentioned in the paragraphs. Comparative evaluation of the qualities of the destination Although the study examines nations based on their use of Airbnb, sustainability and long-term effects: The sustainability component of Airbnb is not mentioned stated in the paragraphs. (Adamiak, 2019)

3. Theory and real-world proof from Airbnb on shared housing and consumer choices.

The study by Yao cui the proposes to examine the important distinction between lodging sharing economy platforms like Airbnb and traditional hotels. The focus is on the fact that guests in the sharing economy may be sharing the property with a local host, which distinguishes it from the experience of occupying an entire space in traditional hotels. The paper further explores the theoretical ramifications of consumer preferences for shared living on both the platform and the social planner in light of the actual data that lends weight to this claim. The methodology described in this study involves combining three data sources to analyse the impact of an attack on Airbnb listings. Transactions Data: The Airdna-sold transactions dataset includes details on the results of each listing and trip date combination. The dataset divides the results into three categories: There are three possible outcomes: 1) the host disabled the listing, making it unavailable for rent; 2) the property was open for rent but unoccupied; and 3) the listing was rented by a visitor. The dataset for rental listings contains transaction costs and booking dates. The combined transaction data comes from two waves: the attack year, which includes the 24 weeks immediately after the attack date, and the control year, which includes the three months before and after the attack date. Listing characteristics: The insideairbnb.com website, which scrapes historical information from Airbnb listing webpages, is where you can find the listing characteristics dataset. Listings that were present throughout the duration of the research are identified using quarterly data from both years. The examination of Airbnb listings and their behaviour in reaction to an assault is the field of application in this context. The study focuses on transaction outcomes, price, booking dates, and listing qualities in order to comprehend the effect of the attack on the Airbnb market. There are a few potential study topics that might be taken into account or expanded upon: limits in the methodology The paragraph does not address any potential drawbacks or difficulties with the methods used, Analysis of the host's religious affiliation: Although it is mentioned in the text that one of the data sources contains host religious identification attributions, it is not explained how this data is used in the research. Comparison of the attack year and the control year: The paragraph indicates in passing that two waves of transaction data are gathered, spanning the attack year and the control year. External factor consideration: Although the study focuses on how the attack affected Airbnb listings, it may be useful to take into account additional external variables that might potentially affect the results (Cui, 2018).

4. Aspects that impact the initial cost of short-term rentals: a strategy for controlling revenue:

The study by Diego de Jaureguizar Cervera suggests doing research on revenue management in the context of short-term flat rentals provided on websites like Airbnb. The goal is to provide a theoretical framework that explains how different aspects of a rental flat relate to the price at which it should be originally offered on the market. The parameters listed in the platform listing and the sociodemographic of the neighbourhood where the flat is situated are both taken into account in the analysis. The problem according to the study is the difficulty real estate professionals encounter when deciding on the best beginning price for residential real estate assets, particularly in the context of short-term rentals. The goal of the article is to give a useful tool that aids professionals in making educated pricing decisions by helping them comprehend the elements that influence the beginning price of short-term rental units. Big data analysis and the application of a technology transfer model make it possible to measure and estimate how various internal and external factors affect pricing. In order to investigate the link between flat qualities and rental price in the context of short-term rentals, the technique entails gathering data from various sources, using a technology transfer model based on the case study approach, and analysing the dataset. According to the sentence above, the area of application is income management in the context of short-term rental properties that are advertised on websites like Airbnb. The objective of the research is to create a technology transfer model that identifies the association between many elements (including both internal property qualities and sociodemographic indicators of the immediate vicinity) and the ideal short-term rental pricing. The study focuses on how these factors affect price choices in the real estate industry and their practical ramifications. The output of this study is the creation of a technology transfer model that identifies the link between several elements (interior property attributes and sociodemographic indicators) and the ideal short-term rental rate for flats. The technology transfer methodology for revenue management has been put into practise in Madrid, enabling revenue managers to calculate short-term rental flat market-adjusted pricing using data that is already available. Some potential areas that may be taken into account for additional investigation or study include: While the size/capacity of the apartments and the amenities supplied are mentioned as factors determining the rental price in the paragraph, there may be other factors that might also come into play, Dynamic pricing: The emphasis of the paragraph is on determining the first rental price at the time the property is first listed for sale. Competition in the market: Analysing the pricing of short-term rental flats in relation to other comparable properties might offer insightful information. Preferences and behaviour of customers: While the text indicates that elements influencing tourists' psychological behaviour are taken into account. (Cervera, 2022)

5. Airbnb's impacts on the property business: London statistics:

The research study by Amit Chaudhary was to apply a difference-in-differences approach to examine how Airbnb has affected house rental costs in the Greater London region while taking into account the Zoopla and Airbnb datasets. In this article, the endogeneity issues are addressed, and more accurate projections of Airbnb's effects on the rental market are given. To overcome potential endogeneity issues and demonstrate a causal link between Airbnb and home rental costs, the difference-in-differences approach is justified. The method aids in accounting for unseen variables, such as shocks in neighbourhood quality or changes in facilities drawing visitors, that may concurrently impact Airbnb supply and traditional market rental rates. The author seeks to reduce these endogeneity difficulties and give more reliable estimates of Airbnb's influence on rental prices by comparing rental price trends before and after the introduction of Airbnb and taking into account the variations between treatment and control groups. The methodology used is by quantifying the impact of Airbnb on house rental pricing using the difference-in-differences (DID) approach. Airbnb has a relatively minimal presence in that market category; hence the author uses homes with more than three bedrooms as the control group. The author intends to identify the causal relationship between Airbnb and rental prices by comparing the trends in rental prices between the treatment (properties impacted by Airbnb) and control groups (properties unaffected by Airbnb). The examination of Airbnb's effects on long-term rental costs in the Greater London region is the area of applicability in the previous two sentences. The author uses a difference-in-differences technique to study the relationship between the supply of available Airbnb rooms and the rental rates of different property types. The findings show that the presence of Airbnb significantly affects long-term rental pricing, particularly for smaller homes, and that the amount of the effect varies with the number of beds. The research would benefit from further discussion on constraints, other hypotheses, data concerns, heterogeneity, mechanisms, and policy implications. (Chaudhary, 2021)

6. Airbnb's financial gains and expenses:

In the study proposed by Josh Bivens, one of the well-known internet-based service companies (IBSFs) Airbnb's expansion is proposed to be evaluated for potential economic costs and advantages. The intent is to provide knowledge to the ongoing discussion regarding whether and how to regulate IBSFs like Airbnb. The paper seeks to explore hypotheses and offer actual data to add to the conversation on the financial effects of Airbnb's rental business. The proposed analysis aims to address a number of issues related to the growth of Airbnb, including its potential effects on local housing costs, residential neighbourhood quality of life, employment quality in the hospitality sector, and the capacity of local governments to enforce laws and collect necessary taxes. The paper seeks to offer a more knowledgeable view on the costs and advantages of Airbnb's presence in various areas by evaluating these problems. The problem here is that the discussion and controversy surrounding the growth of internet-based service providers (IBSFs) like Airbnb is the issue mentioned in the text. Contrasting opinions and ideas about the advantages and disadvantages of these platforms' economics frequently come up in arguments over whether and how to regulate them. IBSFs, according to its supporters, foster innovation, raise the standard of goods and services, and provide employment possibilities. On the other side, sceptics view them as efforts by capital owners and venture capitalists to get around rules and profit at the expense of established rivals. The methodology used by author is collecting and combining information from many sources, such as academic studies, surveys, economic indicators, and case studies. compare the benefits like Property owners can diversify into short-term rentals, Increased options and price competition for traveller's accommodations, Traveller's spending boosts the economic prospects of cities, to the potential costs like Housing expenses are growing for long-term tenants, Tax revenues for the local government decrease, externalities that affect nearby residents, Jobs might become scarcer and of worse quality. The application area covered in this study is the economic examination of Airbnb's effects on a variety of factors, such as housing costs, travel accommodations, local economies, taxation, zoning laws, and equity issues. The analysis' main goal is to comprehend the costs and advantages of Airbnb's growth as well as how it would affect various parties, including tenants, property owners, local governments, and city dwellers. The research findings are intended to help in policy discussions and decision-making surrounding the administration and regulation of online service providers like Airbnb and others in the sharing economy. The total result indicates that while Airbnb has certain advantages, such as a greater supply of vacation accommodations and transactional ease for property owners, it also poses considerable problems in terms of housing costs, taxes, (2020) (cui, 2020)zoning, and equity issues. In order to minimise any negative effects and achieve a balanced approach that answers the concerns of various stakeholders, the research emphasises the necessity for thorough regulation and policy decisions regarding the operation of Airbnb and comparable internet-based service providers. Following are some prospective topics that may be researched further or taken into account in the analysis: Effect on nearby businesses: The review primarily examines the costs and advantages of Airbnb from the viewpoints of visitors, property owners, and locals. Socioeconomic Disparities While the topic of the unequal advantages enjoyed by wealthy and white households is discussed, Framework for Regulation and Governance Although the research briefly addresses the necessity for regulation and policy decisions, more investigation of the particular regulatory difficulties is necessary. (Bevins, 2019)

7. Calculating Airbnb's effect on the hotel business:

In the analysis by Jessica Haywood, intends to evaluate how hotels and Airbnb do in various international areas. The objective is to evaluate both lodging choices' occupancy rates, market shares, average daily rates (ADR), and demand trends. In order to investigate the dynamics between the two industries and assess their competitive environment, the analysis uses data from STR (a data supplier for the hotel business) and Airbnb. Lack of transparency regarding Airbnb's effects on the hotel business and the type of demand it creates is one possible issue that may be deduced. The challenge in determining whether Airbnb's rise represents new demand or if it is replacing hotel stay. others who support Airbnb and others who oppose it disagree on this subject. By looking at occupancy rates, market shares, average daily rates, and demand trends for both hotels and Airbnb, the report aims to shed light on this issue. The methodology used here the analysis of data of market activity given to STR by Airbnb, Data for 13 markets, including Barcelona, Boston, London, Los Angeles, Mexico City, Miami, New Orleans, Paris, San Francisco, Seattle, Sydney, Tokyo, and Washington, D.C., is available from 1 December 2013 to 31 July 2016.Shared and private rooms were left out of the research in order to maintain data comparability with hotel data, and only full homes and flats were included. Additionally, since bigger parties are less likely to book hotel rooms, postings that could accommodate more than seven individuals were disregarded. The hotel business, and more especially the effect of Airbnb on the travel and lodging industry, is the subject of these paragraphs. The report looks at the growth rates of Airbnb's supply and demand, showing that both travellers and property owners are becoming more and more aware of it. Insights into the interactions between Airbnb and conventional hotels—particularly their effects on hotel demand—are the result of this investigation. The study recognises the complexity of the variables affecting the lodging sector, such as supply expansion, macroeconomic conditions, technology advancements, and the advent of alternative rent-by-owner platforms. (Haywood, 2016)

8. Research on Airbnb as a possible rival to the hotel business:

In the study proposed by Quynh Nguyen, The research project outlined in this article intends to add to the sparse body of empirical research on Airbnb as a possible rival to the hotel sector. Despite receive ng widespread publicity in TV news and periodicals, it shows that hospitality experts are not paying enough attention to the issue of Airbnb's effects on the hotel business. By conducting qualitative research to examine the interaction between Airbnb and the hotel business, the study seeks to close this gap. The issue with in this study is that few hospitality experts have focused on Airbnb as a possible rival to the hotel sector. Despite significant publicity in periodicals and on television, there is a dearth of empirical studies examining Airbnb's effects on the hotel business. This study gap is seen negatively because it prevents a thorough knowledge of the ramifications and interactions between traditional hotel owners and Airbnb. The research strategy and survey questions utilised in this study are described in the methodology. The competitor identification technique by Bergen and Peteraf served as the inspiration for the research. The primary goal of the study is to examine the demands travellers have when making hotel reservations, as well as whether and how effectively Airbnb satisfies those needs. This study applies to the hotel business and how it has reacted to the possibility of competition from Airbnb. Given the purpose of travel (business or pleasure) and market categories (budget, economy, midprice, upmarket, and luxury), the study seeks to ascertain if Airbnb competes in various market sectors within the hotel industry. The study's output is a collection of findings derived from a 10-question survey that was created using the Qualtrics platform. The survey data are examined and used to inform the competitor identification methodology developed by Peteraf and Bergen. The findings identify the degree to which Airbnb competes with various market niches in the hotel business. Several significant elements that could be overlooked from a research standpoint include: Statistical Analysis: No statistical analysis methods utilised to examine the survey data are mentioned in the paragraphs. Validation and Reliability: The paragraphs make no mention of the survey instrument's validity or dependability. Ethical Issues: No ethical issues pertaining to the poll, such as informed consent, data protection, or confidentiality. (Nguyen, 2014)

9.Study on the progress and development of Airbnb in the hotel business:

In the study by Daniel Guttentag on Airbnb is on the most significant recent developments in the travel industry. The purpose of his study is to evaluate the advancements in Airbnb-related research that have been made so far. The researchers looked through 132 peer-reviewed journal publications from diverse areas after searching several journal databases. Each paper's major characteristics were noted, and its substance was examined. there are places where additional study is needed as well as research gaps. These knowledge gaps suggest that certain facets of Airbnb have not been fully investigated or comprehended. As a result, the issue may be defined as the requirement for more study to fill in these knowledge gaps and offer a more thorough understanding of Airbnb's impact, ramifications, and dynamics on the tourist industry. A thorough evaluation of peer-reviewed journal publications on Airbnb is part of the technique outlined in the paragraph. The search was carried out utilising many online databases, including ScienceDirect Journals, EBSCOhost Hospitality and Tourism Complete. Only peer-reviewed journal articles were taken into consideration. The papers that cited reputable Airbnb articles using Google Scholar. The field of application of the study is to examine and summarise the current research on Airbnb from many disciplines and to offer insights into the body of knowledge and knowledge gaps in this field. Understanding the phenomena of Airbnb and how it impacts the travel and tourism industry is the study's main goal. The study's weaknesses are that it compares the effects of different scenarios on Airbnb; it would have been more useful to include an examination of the elements that have an impact on the Airbnb business model. (Guttentag, 2019)

10. How Airbnb impacts rent and housing costs:

Kyle Barron's suggested study found that The authors developed and conducted a study that focused on the reallocation of residences from the long-term rental market to the short-term rental market in order to evaluate the effects of house sharing. He wants to know whether house sharing affects rental rates, home prices, and the price-to-rent ratio favourably or unfavourably. It is challenging to forecast how house sharing may affect rental rates and property prices, but the authors are aware that there may be both positive and negative externalities. This analysis solely considers the short-term implications of Airbnb; it makes no attempt to account for changes in the overall housing stock or long-term consequences. The effect of the sharing of homes organisations like Airbnb on the housing sector is a worry raised by this study. The intent of the study is to evaluate how house sharing impacts these attributes and to offer support for the reallocation channel. Concerns have been made about the potential advantages of the sharing economy, particularly in the home-sharing sector, going mostly to investors and non-resident visitors, as well as about the possibility of rising living expenses for locals. According to rumours, home-sharing websites promote landlords to switch from long-term to short-term leases by rendering short-term rentals easier, which might drive up the cost of long-term rents. For the study's aims, data from the website for Airbnb will be collected and examined in order to comprehend the features and accessibility of Airbnb listings in the U.S. The investigation of Airbnb's impact on home prices and rents is the study's intended application area. The aim of the study is to ascertain how the dynamics of the housing market have been affected by the supply of Airbnb, as measured by the total number of listings. In the regression findings. The natural logarithms of the Zillow Rent Index (ZRI), Zillow Home Value Index (ZHVI), and Price-to-Rent Ratio (ZHVI/ZRI) are the variables that are dependent in this study. The research is carried out at the zip code stage, and other variables of control are incorporated into the regression models. These are only a few such variables that may be considered from a research standpoint. Conclusion as to cause: The study uses regression analysis to look at the relationship between the availability of Airbnb and the outcomes of the housing sector. Relations of space and time: The study investigates the impacts of Airbnb over a predetermined time frame and at the zipcode level. Contrast and compare Although it does not compare Airbnb to other websites that provide short-term rentals, the study is focused with how Airbnb influences the market for housing.

11. performance and Effectiveness of Airbnb : Factors and models for prediction:

In order to understand the Airbnb effects, difficulties, and user experiences, the research proposed by Efstathios Kirkos on Airbnb suggests investigations into a variety of platform aspects, including Disruptive Innovation Theory, Market Expansion and Platform Development, Taxation and Regulatory Framework, Impact on Property Value, Availability, and Urban Planning, Employment and Hotel Industry Performance, User Demographics, Pricing Dynamics, and Customer Characteristics and Satisfaction. The identification and study of elements that affect the performance of Airbnb listings is the issue or research topic described in this sentence. To comprehend the effects of many aspects on consumers' decisions, purchasing intents, and loyalty towards Airbnb lodgings, numerous research have been carried out. The paragraph outlines research projects that investigate how visitors' buying decisions and hotel room sales are influenced by variables such internet reviews, perceived value, perceived price, confidence in the host, cultural demands, room characteristics, host traits, and persuasive appeals. The approach for this research outlines the methods used for gathering and preparing the data for the investigation. The study's data originated from InsideAirbnb.com, a website that provides public use of data on listings on Airbnb in several locations. Thessaloniki is the subject of the study's particular dataset, which includes details about the accommodation, host, neighbourhood, availability, customer reviews, etc. This study's field of application is the performance evaluation of Airbnb listings. The determination of the studying is to comprehend the factors and predictive models that impact how well Thessaloniki's Airbnb listings function. The superhost badge, host attentiveness, supply of facilities, excellent overall guest ratings, and perceived value were among the relevant characteristics that the study discovered. According to the study, hosts may boost their income by increasing personal visibility, response rates, and amenities, as well as by earning the superhost label. Low-performing listings might use the prediction model used in the study to evaluate different tactics to improve their performance. Additionally, tax officials can utilise the model to spot probable tax evasion. The aspects which have been missed according to my research point of view are Generalizability: The study admits that there can be regional variations and concentrates on Airbnb listings in Thessaloniki, Greece., External Factors: The majority of the study's attention is focused on internal factors related to hosts, lodgings, and customer reviews. However, external factors like market competition, local legislation, and the state of the economy can also have an impact on how well Airbnb listings perform., Customer Preferences: Although the study looks at variables that affect listings' success, it does not specifically explore the demands and preferences of Airbnb users., Customer Preferences: Although the study looks at variables that affect how well listings function, it does not specifically go into the wants and preferences of customers. Sentiment analysis of reviews: The paper addresses the possibility of enhancing the input vector using sentiment analysis of Airbnb users' reviews. (Kirkos, 2021)

12. What factors influence Amsterdam Airbnb users' satisfaction? A probe of sentiments:

In the study proposed by Heyam Abdullah Bin Madhi, By identifying the elements that affect satisfaction, the study hopes to improve understanding of consumers' demands and increase customer satisfaction. The research tries to pinpoint the crucial elements that have an influence on consumers' happiness on the Airbnb platform by examining user comments and reviews. The problem here is that there haven't been many research done in the past that analyse consumer satisfaction aspects while also looking at listing features on the Airbnb site and guest online reviews. As a result, the study's goal is to pinpoint the elements that Amsterdam users of the Airbnb platform can depend on to feel satisfied. In order to prepare the datasets for research of consumer satisfaction on the Airbnb platform in Amsterdam, the process include collecting and cleaning online review comments and listing data. A total of 110,747 suitable comments were kept after the pre-processing and cleaning processes for additional analysis. The analysis of Airbnb listings and customer reviews to determine elements that affect customers' happiness is the study's field of application. In order to understand the polarity of customer evaluations, the study focuses on analysing the characteristics of Airbnb listings. The study's conclusion includes a list of the variables that influence how satisfied consumers are. The key variables that affect customers' pleasure, according to the authors, are costs, if the host is a super-host, and the style of accommodation. The topic clustering analysis carried out as part of the study showed that favourable comments were more likely to mention "Location" than negative ones to mention "Check-In." The survey also emphasised how Amsterdam as a location affected consumer happiness and its capacity to get favourable remarks. Several parts of this study should be further investigated, including Understanding the study's methodology might be improved by giving additional details about the exact sentiment analysis techniques, regression models, and visualisation approaches used. It would be helpful to look deeper into the practical consequences for both Airbnb hosts and platform operators with the intention of gaining a better understanding of how the Airbnb listings and customer reviews were gathered and whether any sample biases or limits existed. (Madhi, 2021)

1.7 Gap analysis:

A gap analysis based on the aforementioned 12 research publications may be done to determine the areas that have not been thoroughly studied or where further study is required. An examination of the research articles' gaps is given below:

Impact on Neighbouring Communities: Analysing thoroughly how Airbnb impacts surrounding communities would be a good way to get knowledge about the local social and economic dynamics. To create successful legislation, it is crucial to comprehend how short-term rentals affect local companies, citizens' quality of life, and community cohesiveness.

Data gathering and Methodology: Greater transparency and repeatability of the study would result from more details being provided about data gathering techniques and the methodology used. Other researchers may expand on the study's conclusions if the sources of the data, survey methods, and analytic procedures were made clear to them. Understanding how Airbnb competes with conventional hotel alternatives in various towns and regions would require doing a comparative study of destination attributes, taking into account elements like cultural attractions, amenities, and transit options. A detailed understanding of Airbnb's effects would be possible by looking at the variety of its effects across various market groups and geographical areas. Finding the variables that limit Airbnb's effects might help direct focused policy measures.

The study of Airbnb's influence would be reinforced by filling in these gaps and include these elements in further research, and its conclusions would have a wider range of relevance and applicability. Furthermore, a deeper comprehension of the intricate dynamics of the housing market in regard to home-sharing platforms like Airbnb would benefit from the exploration of practical ramifications, the examination of biases, and the exchange of precise methodological information.

We are going to cover these gaps by performing the following:

Understanding Industry Dynamics: Examining the elements influencing the availability, demand, and affordability in the New York City Airbnb business, Recognising the various sorts of properties that are offered (such as flats and separate spaces), as well as their appeal, Geospatial analysis: evaluating the density and demand centres of postings on Airbnb across various neighbourhoods, Analysing client feedback to determine the general attitude and degree of satisfaction with Airbnb services in New York City, Finding reoccurring Good and Bad factors: Identifying abilities and shortcomings by locating reoccurring both beneficial and detrimental factors expressed in customer evaluations. These gaps will be performed as a direct client perspective is absent from the existing approach, which is too focused on market trends and structure. It doesn't represent the opinions and encounters of actual consumers. Restricted Actual Time Insights: The research conducted at this time may be based on past data and might not correctly reflect current patterns and modifications in buyer habits and interests. The research develops stronger by completing in these gaps and putting the suggestions into practise, offering greater insights into the Airbnb industry while taking the perspectives and experiences of its customers into account.

Chapter: 2

2.1 Introduction

To guarantee a thorough and rigorous analysis, the approach for this study on the Airbnb industry and elements influencing consumer satisfaction in New York City should be properly planned. We will use a mixed-methods strategy to accomplish our study goals. In the beginning, we will use The Inside Airbnb website to gather thorough information about Airbnb homes in New York City, including their location, host traits, room kinds, and costs. Additionally, in order to comprehend the connection between neighbourhood attributes and the availability of Airbnb listings, data on socioeconomic status from the American Community Survey will be gathered. In order to pinpoint the key elements affecting consumer happiness on the Airbnb platform, we will also conduct sentiment analysis on customer feedback and suggestions. This will enable us to identify the factors that have the most effects on guests' experiences, such as price, host status (super-host), and room type. The impacting of Airbnb on the conventional hotel business will also be examined using a distinction in differences methodology in terms of occupancy rates, average daily prices, and shares of the market.

2.2 Materials and dataset.

This study attempts to analyse the New York City Airbnb industry and pinpoint the variables that affect price and client satisfaction. A thorough technique that includes data collection, analysis, and interpretation will be used to accomplish this aim. The technique combines both qualitative and quantitative approaches to offer a thorough insight of the city's Airbnb industry.

Collection of data: The Inside Airbnb dataset, which provides thorough details on Airbnb listings, hosts, reviews, and other pertinent factors for New York City, will serve as the main source of data for this study. We'll get this dataset and prepare it for more analysis. Secondary data from reliable sources including government databases, real estate firms, and tourism agencies will be taken into account to have a better knowledge of the New York City housing market and the effect of Airbnb.

2.3 Preprocessing of the data:

To eliminate any discrepancies, missing numbers, or unnecessary information, the obtained data will be carefully cleansed. The data will be filtered to highlight properties in New York City and remove items that were entered more than once or incorrectly. A thorough database for analysis will be built by combining various datasets. This will include comparing pertinent factors, such as property characteristics, host data, costs, and visitor ratings.

2.4 Analysis of the data:

The essential elements of the city of New York's Airbnb market, such as the quantity of listings, the average price, the rate of occupancy prices, and the distribution among neighbourhoods, will be summed up using descriptive statistics. We will compute dispersions and its measures which for example are standard deviation and range, to realize the variability of the data after computing measures of central tendency, such as average, middle value, and mode, for variables like listing prices, occupancy rates, and customer ratings. Following that, we will visualise the distribution of listing prices, customer reviews, and other important characteristics using density plots and histograms. Finally, look for any probable data outliers that might have a big impact on the statistics as a whole. Then, using geographical mapping tools to plot the distribution of Airbnb listings across various New York City neighbourhoods, we will perform geographic analysis. As a result, we will examine the concentration of listings in particular areas to identify popular and less popular regions. In order to understand how various groups contribute to the overall market, we will perform market segmentation, where we will segment the data based on various criteria, such as property types, host status (super-host or regular host), and customer demographics (e.g., solo travellers, families, business travellers).

The following phase is time series analysis, where we will examine temporal market patterns for Airbnb, such as variations in listing prices and occupancy rates over time. The analysis of customer satisfaction will then be performed by calculating the average customer satisfaction rating and identifying the most frequently mentioned positive and negative sentiments from customer reviews. Finally, based on sentiment scores and keyword analysis, reviews will be classified as positive, neutral, or negative. The examination of each host's performance will come next, during which we will determine parameters like each host's average occupancy rate, average nightly rate, and number of reviews. Based on visitor feedback and occupancy rates, we can also identify top-performing hosts. Last but not least, to show the results in an understandable and aesthetically pleasing way, we will employ a variety of data visualisation techniques including bar charts, pie charts, scatter plots, and heatmaps. An early insight of the distribution and features of the Airbnb market in New York City is provided via descriptive statistics analysis. It assists in locating trends, patterns, and areas of interest that may subsequently be thoroughly examined utilising sophisticated statistical approaches and machine learning technologies.

2.5 Findings and analysis:

- id - identification code for the property
- name – name of the property
- id_host- the host's unique identification number
- name_host – name of the host
- group_neighbourhood - the principal areas of the city
- neighbourhood_name - the city neighbourhoods
- lati – latitude of the property
- longi – longitude of the property
- type_room - types of rooms
- price - The cost of a single night
- min_nights - Minimum number of nights needed to reserve the space.
- no_of_reviews - total number of reviews received on the property.
- last_review - The most recent review date
- review_each_month - Monthly number of reviews
- calculate_host_listing_count - number of the host's properties that are listed on Airbnb.
- availability_yearly - The no. of days in one year that in which house is available.

2.6 Analysis of the data set in Python:

We will import the library of pandas which will help us to perform the manipulation and the analysis of the data, The Seaborn library, a data visualisation library built on Matplotlib, will then be imported. Seaborn will provide us with a high-level interface that will enable us to produce both aesthetically beautiful data visualisations and useful and appealing statistics representations. The Pyplot module from the Matplotlib package is then imported. One of the most popular libraries that will assist us in developing static, animated, or interactive visualisations in Python is Matplotlib. Lastly, we will import folium which is a popular framework for interactive maps. Python interactive map creation is

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43566 entries, 0 to 43565 Data
columns (total 16 columns):
 #   Column                        Non-Null Count  Dtype
```

---	------	43566	non-null	-----
0	id			float64
1	property_name	43566	non-null	object
2	id_host	43566	non-null	int64
3	name_host	43561	non-null	object
4	group_neighbourhood	43566	non-null	object
5	neighbourhood_name	43566	non-null	object
6	lati	43566	non-null	float64
7	longi	43566	non-null	float64
8	type_room	43566	non-null	object
9	price	43566	non-null	int64
10	min_nights	43566	non-null	int64
11	no_of_reviews	43566	non-null	int64
12	last_review	33070	non-null	object
13	review_each_month	33070	non-null	float64
14	calculate_host_listing_count	43566	non-null	int64
15	availability_yearly	43566	non-null	int64

dtypes: float64(4), int64(6), object(6) memory

usage: 5.3+ MB

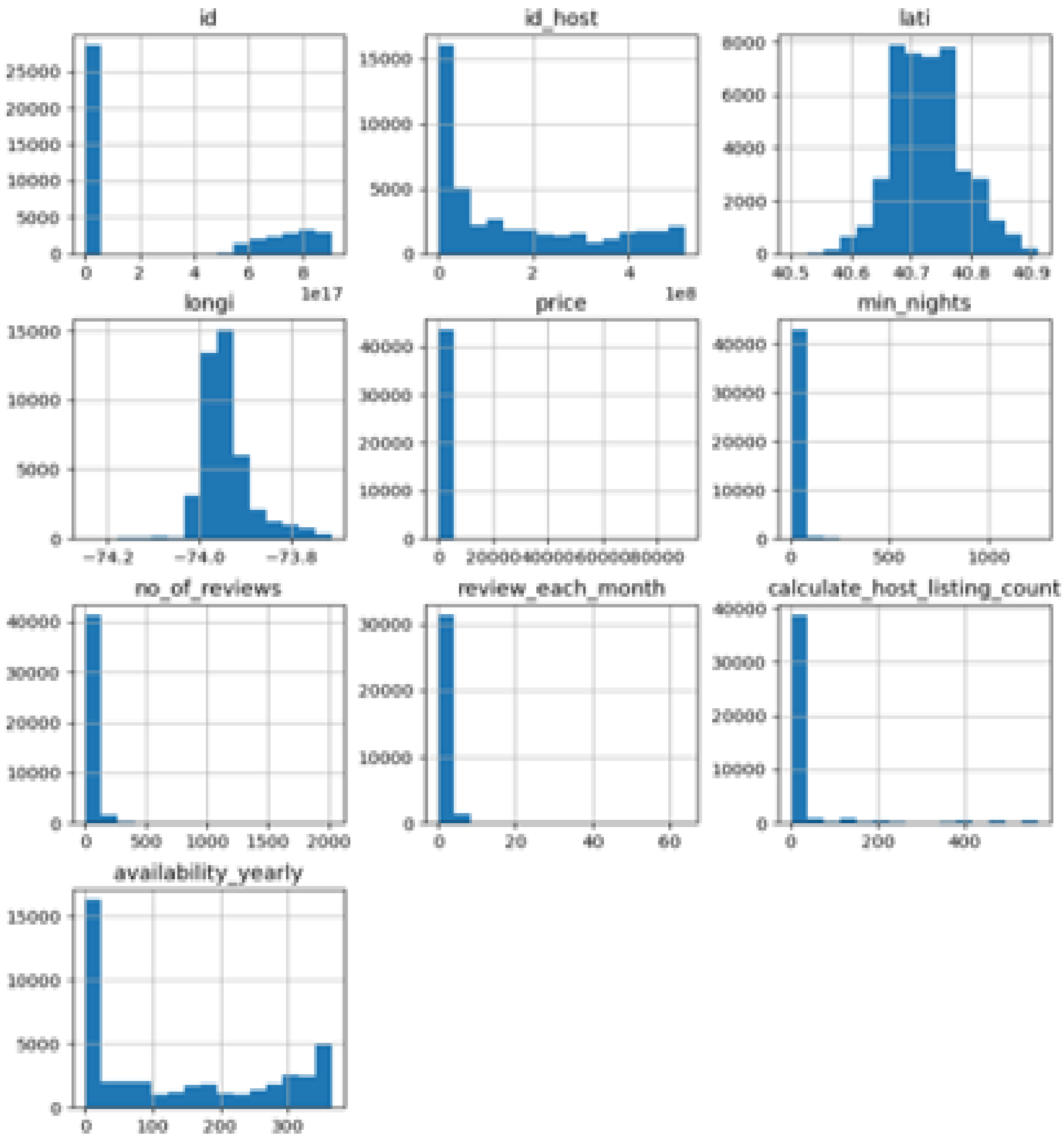
The size of the dataset and certain details about each column, such as the number of non-null entries and the kind of data it includes, may be seen using the info method. Now that we can see everything, there are 16 columns and 50,246 rows of data.Consider the fact that certain columns have null values. Usually, this is bad. To determine what to do with the null values, let's first determine whether they are meaningful. Only the columns last_review and reviews_per_month have a large percentage of null entries. This won't be an issue because neither of these columns will be the subject of our study. We'll get rid of them later on in this project. The host_name column also contains null values, but as we are not analysing these, this will not have an impact on our project. The quantity of null values is also unimportant.

2.7 Distribution of variables:

In order to see the dispersion of each variable , we will start to begin for looking the outliers, for that, we will create several histograms. It can also be used to determine whether the bulk of listings are for affordable, midrange, or upscale lodgings. along With the use of this information, hosts may establish reasonable min-night restrictions, and visitors can make travel plans.Also, With the use of this histograms, hosts may establish reasonable minimum-night restrictions, and visitors can make travel plans. adding that, It can suggest the popularity or activity levels of particular listings and can show the distribution of host interaction.It can say if a listing is available all year round, sometimes, or just during certain seasons.

In [ ]:
```
(round(newyrk.isnull().sum() / newyrk.shape[0] * 100, 2)).sort_values()
```

```
newyrk.hist(bins=15, figsize=(10,12));
```



- We can see from the histograms that certain significant variables, such as price and min_nights, are not evenly distributed. Let's use the describe function to view more dataset statistics in order to more clearly define these issues.

In [ ]:
```
newyrk[['price', 'min_nights', 'no_of_reviews', 'review_each_month',
    'calculate_host_listing_count', 'availability_yearly']].describe()
```

Out[ ]:

	price	min_nights	no_of_reviews	review_each_month	calculate_host_lis
count	43566.000000	43566.000000	43566.000000	33070.000000	43
mean	225.216109	18.700271	26.368544	1.219748	
std	818.738268	27.804388	57.740520	1.777084	
min	0.000000	1.000000	0.000000	0.010000	
25%	80.000000	2.000000	1.000000	0.130000	
50%	136.000000	15.000000	5.000000	0.550000	
75%	228.000000	30.000000	24.000000	1.800000	
max	90120.000000	1250.000000	2024.000000	63.950000	

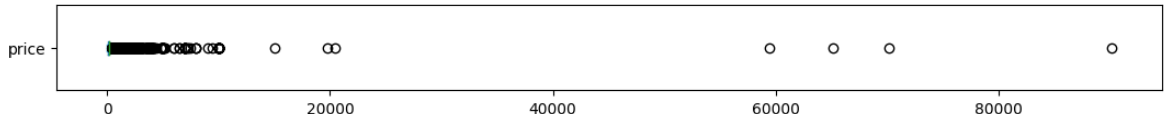
- It is evident that certain values are illogical. Consider the pricing column, for example. The average cost is 225.21, although the highest and lowest prices are 90120 and 0 respectively.
- The greatest number in the min_nights column is 1250, which is not very different. If a visitor needs to stay for at least a few years, how can someone expect to have their space reserved? It is very illogical!
- These kinds of values corrupt reality and whatever analysis we try to do. We have to start interacting with them.

2.7 Outlier Removing

- To examine the distribution of each of these columns in more detail, boxplots will be created for each of them. we will also check at how many and what proportions of prices fall between 0 and 500, as well as what proportions of minimum nights are longer than 30.

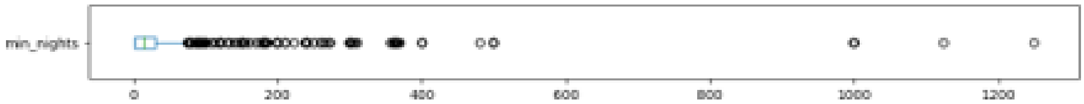
In [ ]:

```
# Using the box plot to display newyrk['price'].plot(kind='box', vert=False,
figsize=(13,1)) plt.show()

# Counting and calculating the proportion of prices that are equal to 500.

print('Values over $450.00: ')
print(len(newyrk[newyrk['price'] > 450]))
print('{:.4f}%'.format((len(newyrk[newyrk['price'] > 500]) / newyrk.shape[0]) *
print('\n')
print('Values equal to $0: ')
```




```
Values over $450.00:
        3375
        5.9909%


        Values equal to $0:
        16
        0.0367%
```

In [ ]:
```
# Using the box plot to display newyrk['min_nights'].plot(kind='box', vert=False,
figsize=(13,1)) plt.show()

#Counting the number of rooms and calculating the proportion of those having a l
print('29 nights above values: ')
print(len(newyrk[newyrk['min_nights'] > 30]))
print('{:.4f}%'.format((len(newyrk[newyrk['min_nights'] > 29]) / newyrk.shape[0]
```

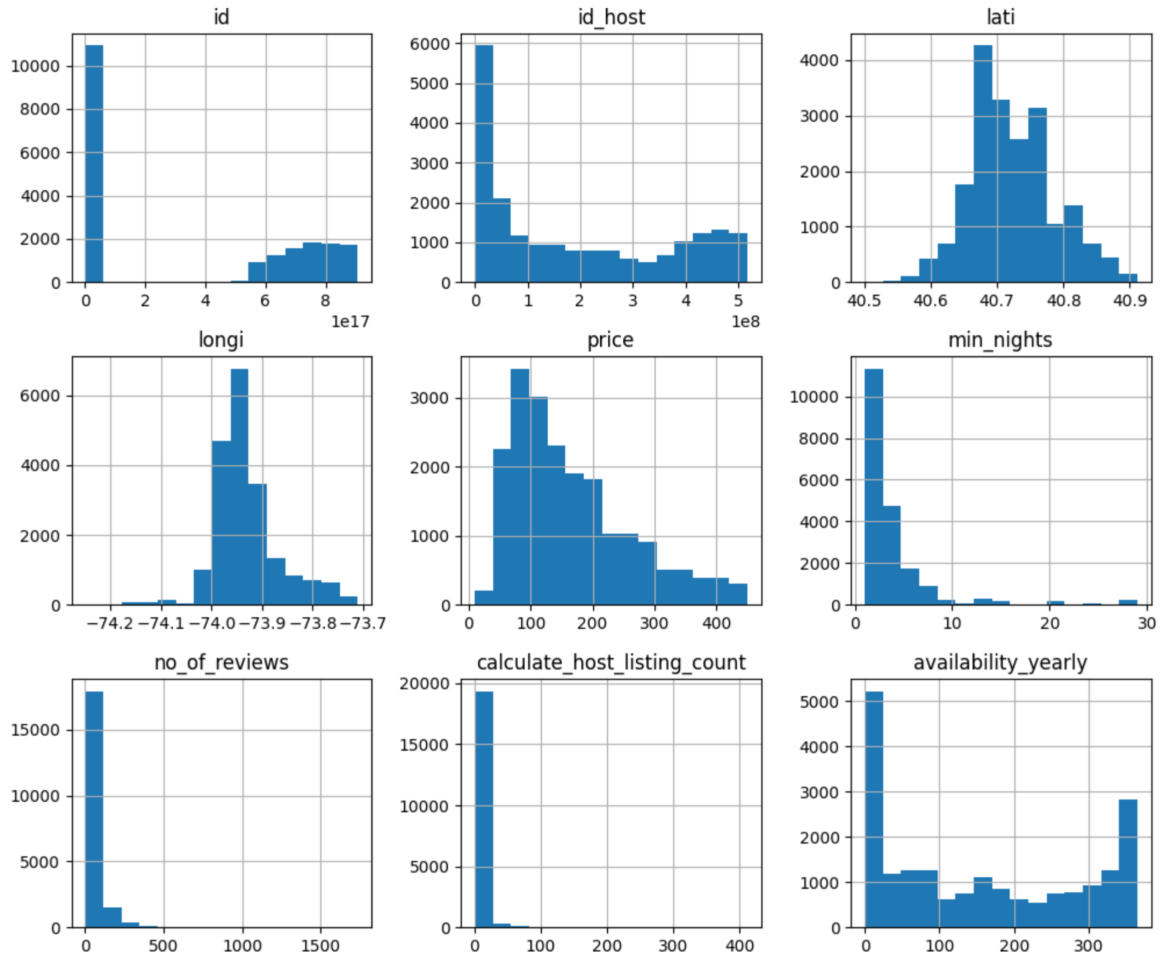



```
29 nights above values:
2035
48.5057%
```

we can observe that only 6% of entries in the price column and only around 5% of the values in the min_nights column are bigger than 500 dollars. and 17 of our components are having a price of nill dollar as well. so, i think it is appropriate to eliminate around 5

% of the data for the purpose of making it more real to life like we found before that 70 percentage of this columns are below 230 dollars and for 30 nights respectively, we will now proceed with building a database which is clean named newyork_clean which will only contain the rows which have a price bigger than 500 and the lowest number of nights is not bigger then 30.Moreover, there mights be a probablity that few columns will meet both of this criterias, which directly means that the amount of data that we will loose is actually way less than 5%, also, we will eliminate the columns reviw_per_month and reviw_last after the creation of the new dataframe, as we discussed previously. now we will also confirm if the histograms can give a better result.

In [ ]:
```
# Picking the rows that meet the criteria that have been specified

newyrk_clean = newyrk[(newyrk['price'] <= 450) & (newyrk['price'] > 0) & (newyrk

#  desired column elimination

newyrk_clean.drop(['review_each_month', 'last_review'], axis=1, inplace=True)
newyrk_clean.reset_index(drop=True, inplace=True)

# Repeating the plotting of histograms.
```



- Even if the distribution is much more accurate currently, we can still observe, for instance, that the majority of prices are still below U$95.00.

2.8 Analysis of correlation

- Let's start searching for any correlations between the dataset's numerical variables.

- The map_heat function from Seaborn will be used to visualise the correlation matrix that we initially construct using the correlation technique.

In [ ]:

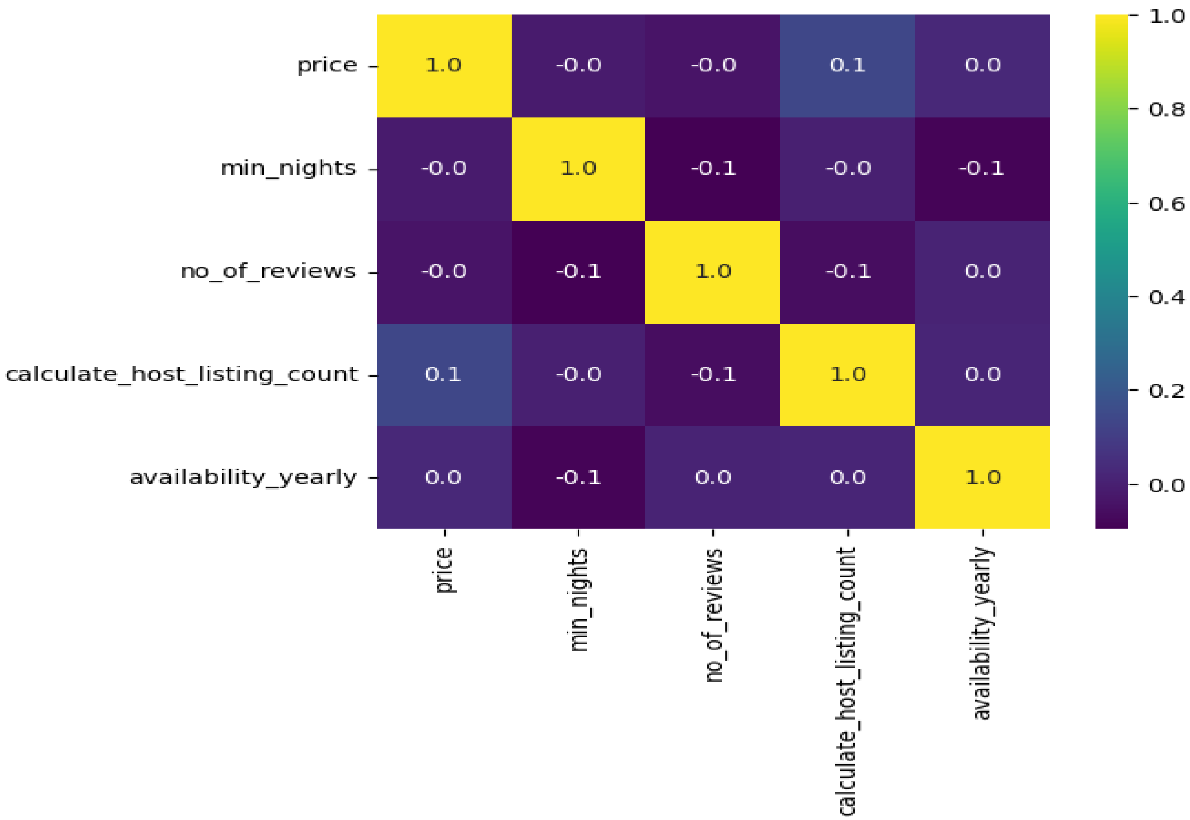
```
correlation = newyrk_clean[['price', 'min_nights', 'no_of_reviews', 'calculate_h
correlation
```

Out[ ]:

	price	min_nights	no_of_reviews	calculate_host_listing_(
price	1.000000	-0.019435	-0.040829	0.1
min_nights	-0.019435	1.000000	-0.097558	-0.0
no_of_reviews	-0.040829	-0.097558	1.000000	-0.0
calculate_host_listing_count	0.137110	-0.000773	-0.059154	1.0

availability_yearly	0.026306	-0.084865	0.009016	0.0

In [ ]:
```
sns.heatmap(correlation, cmap='viridis', fmt='.1f', square=True, linecolor='yell
```

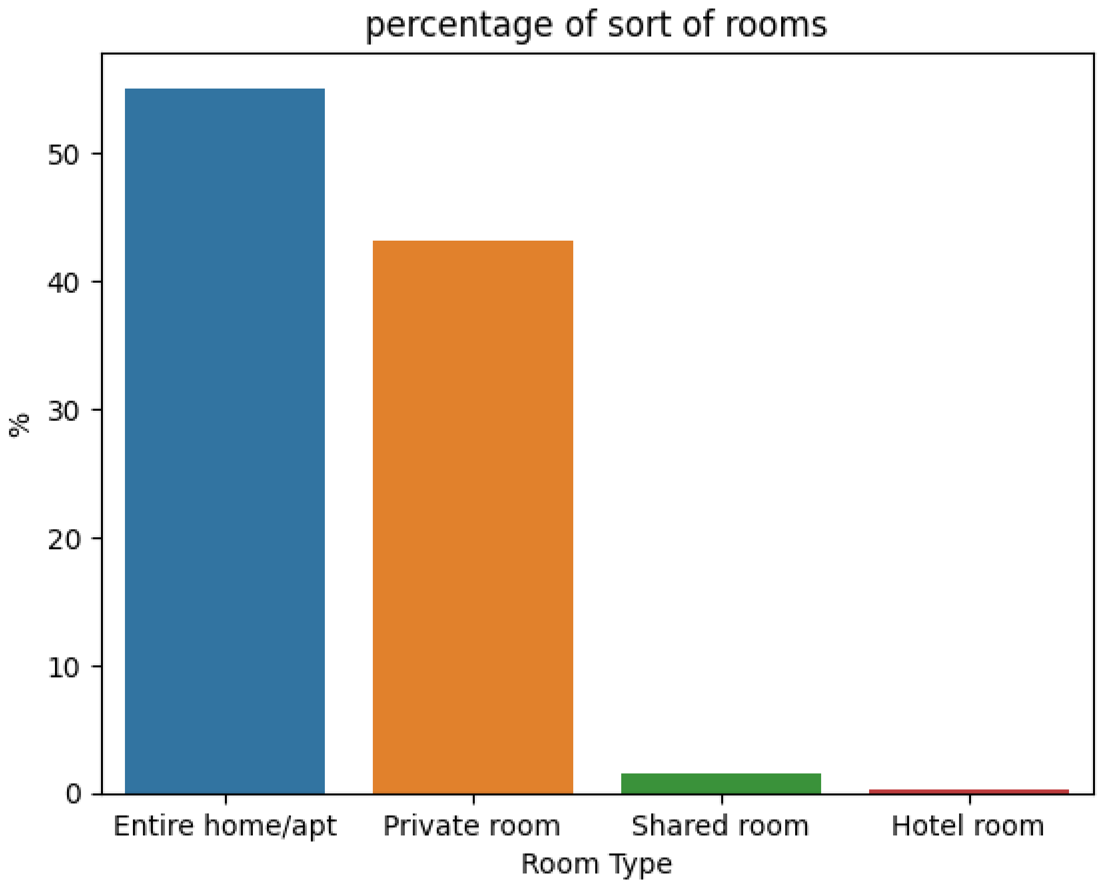


- The variables do not appear to have a strong correlation to one another, as we can observe.

analyzing further to help support the conclusions

Now that we have certain inquiries answered, we'll try to draw some conclusions from the data. We'll commonly utilise the by_group and counts_values methods for that. These two really effective methods let us quickly extract some helpful information from a vast volume of data. Let's check out which type of accommodation is most prevalent in the New York City Airbnb first.

In [ ]:
```
# Plotting the percentage of rooms for each room type plt.figure(figsize=(10, 6)) #
Adjust the figure size if needed ax = sns.barplot(x=type_room_pct.index,
y=type_room_pct.values) ax.set_title('Percentage of Room Types')

ax.set_xlabel('Room Type')
ax.set_ylabel('%')
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, horizontalalignment='right
plt.show()
```



We can clearly say from the above chart that full housing/flats had the highest amount of market in this short range leasing business, while the individual an private spaces closely followed at fourty percentage, on the other hand the shared and motel type of properties had around three and two percent of market respectively.

In [ ]:

```
# Estimating the quantity of each sort of room's rooms
type_room  =  newyrk_clean['type_room'].value_counts().sort_values(ascending=False

# Estimating the proportion of rooms for each sort of room.
type_room_pct  =  round(newyrk_clean['type_room'].value_counts(normalize=True)    *

# Presenting the outcomes
print(type_room) print('\n')
print(type_room_pct) print('\n')


# Graphing the proportion of each room  sorts.
rt = sns.barplot(x=type_room_pct.index, y=type_room_pct)
rt.set_title('percentage of sort of rooms')
rt.set_xlabel('Room Type')
rt.set_ylabel('%')
```

Entire home/apt 11014

Private room 8634

Shared room 297

Hotel room 54

Name: type room, dtype: int64

```
Entire home/apt     55.07
Private room        43.17
Shared room          1.49
Hotel room           0.27
Name: type_room, dtype: float64
```

In [ ]:

```
# Estimating the typical cost per sort of room
price_per_type = round(newyrk_clean.groupby('type_room').price.mean(), 2).sort_v


# Presenting the outcomes
print(price_per_type) print('\n')


# Graphing the mean cost for each kind of accommodation
plt.figure(figsize=(6, 4))
```

```
rt.set_ylabel('mean Price ($)') rt.set_xlabel('type_room')
rt.set_xticklabels(rt.get_xticklabels()) plt.show()
```

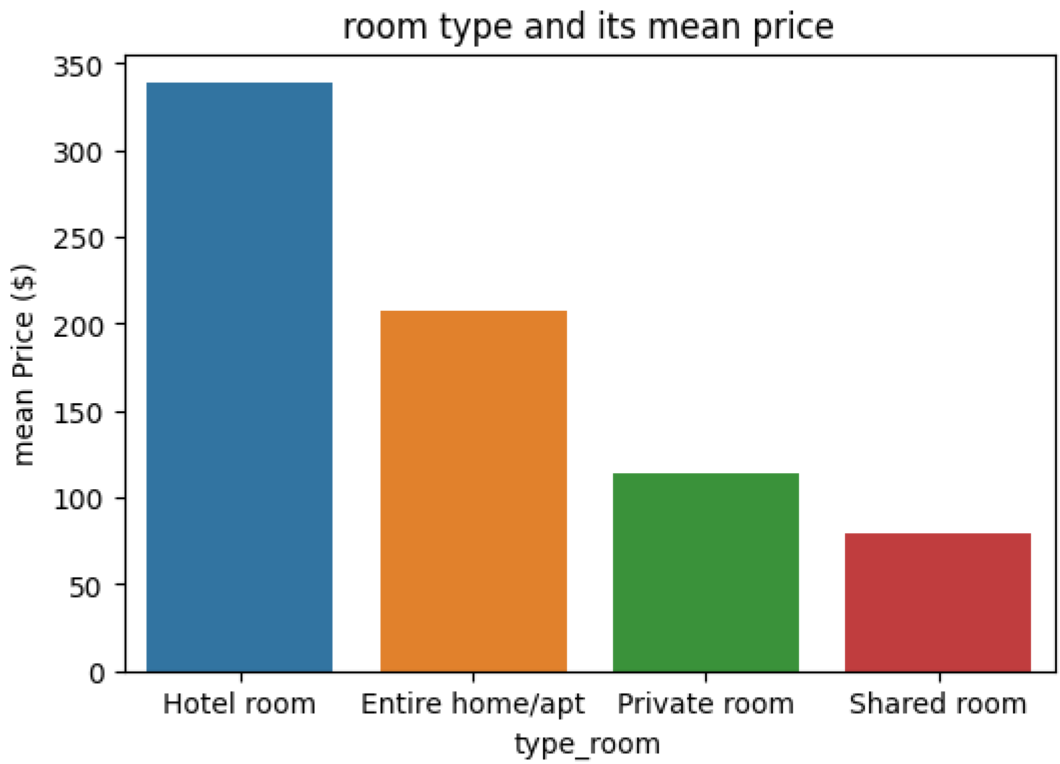
type_room

Hotel room 338.56

Entire home/apt 207.08

Private room 113.65

Shared room 78.98



On average, hotel rooms are the most expensive kind of accommodation. There are two potential causes for this:

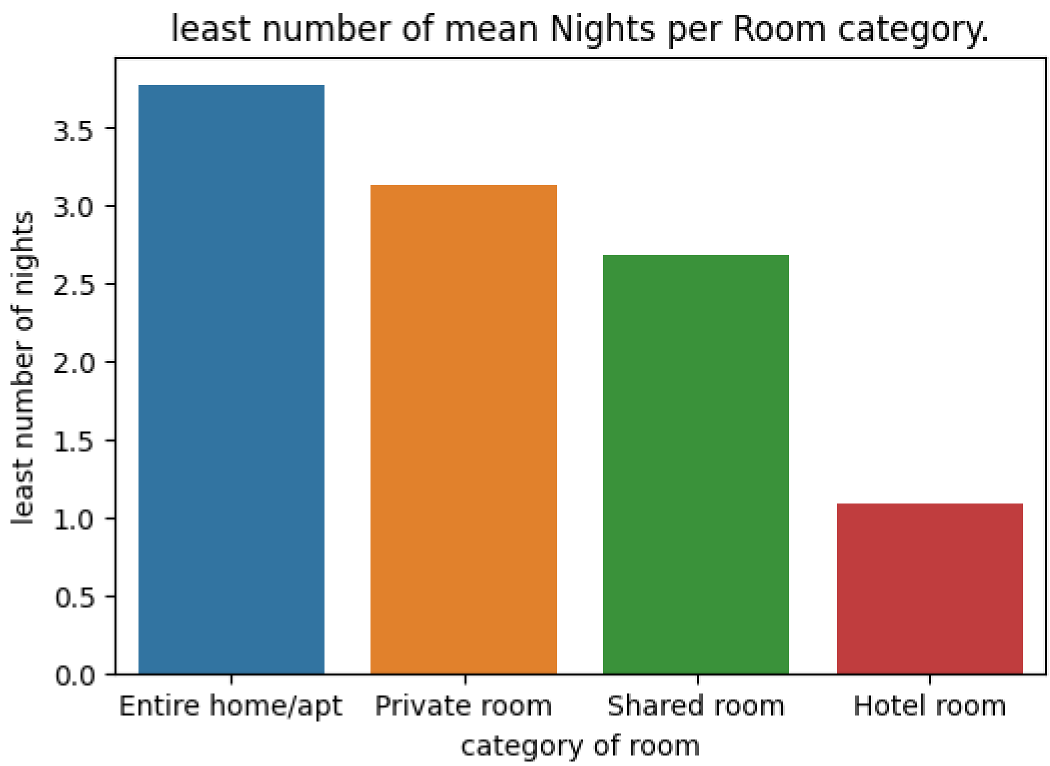
There are fewer hotels in the dataset, which affects the average. Hotels are inherently more expensive due to the number of staff and the services and alternatives accessible to the guest, such as room service and parking places. Apart from that, it seems sense to assume that private rooms are greater costlier compared to shared spaces and that full apartments are greater costlier when compared to shared rooms. Let's now examine the typical number of least nights for each type of lodging.

In [ ]:

```
# Estimating the typical least number of nights for each kind of room

avg_min_nights = round(newyrk_clean.groupby('type_room')['min_nights'].mean(), 2


# Presenting the outcomes
print(avg_min_nights) print('\n')


# Plotting the typical least number of nights for each kind of accommodation
```

```
rt.set_ylabel('least number of nights') rt.set_xlabel('category of
room') rt.set_xticklabels(rt.get_xticklabels()) plt.show()
```

```
type_room
Entire home/apt        3.77
Private room           3.13
Shared room            2.68
Hotel room             1.09
Name: min_nights,     dtype:   float64
```



Whole houses or flats give visitors an exclusive place, making them perfect for extended visits. All services and facilities are available to visitors, giving them a feeling of being at home. Convenience and Independence: Possessing their own home area, complete with a kitchen and several rooms, gives visitors a feeling of being at home while travelling. Quality for Longer Stays: The more comfortable typical nights suggest that visitors prefer to stay in complete houses or flats for extended periods of time, maybe for pleasure, employment, or vacation.

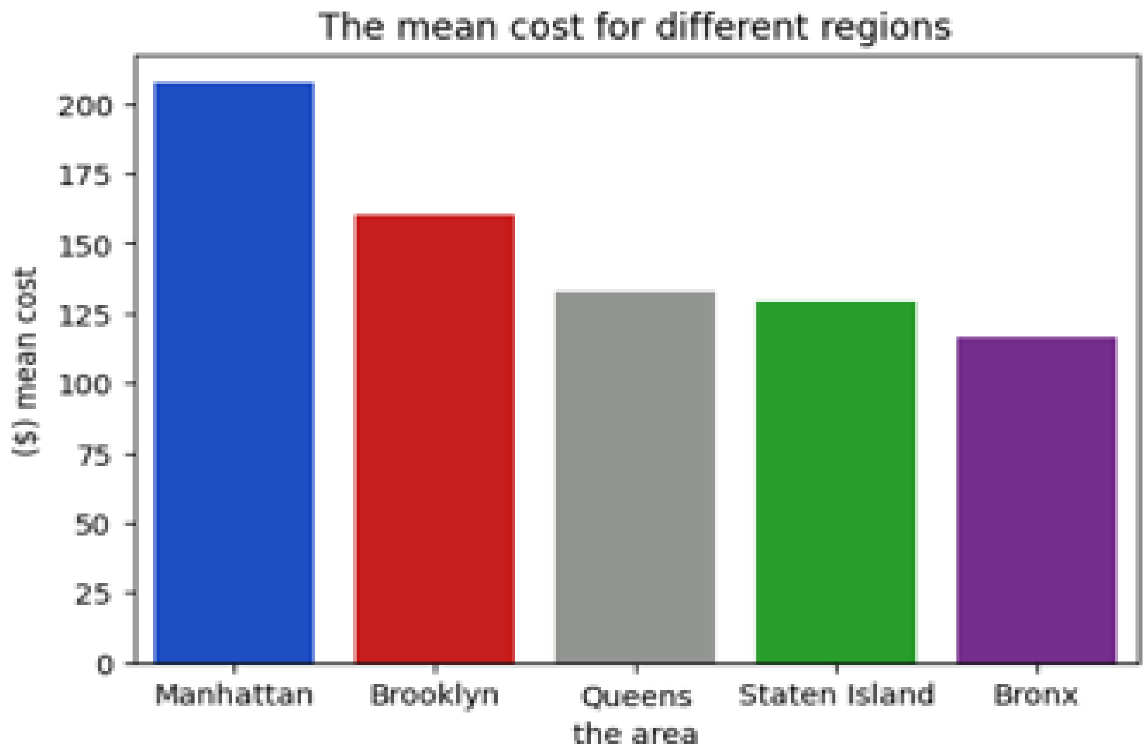
2.9 Exploring Locations

We'll proceed to inquire into costs in various parts of the city. Because it splits the city into five broad areas rather than numerous local neighbourhoods, we'll utilise the neighbourhood_group column for this instead of the neighbourhood column.

```
rt.set_xticklabels(rt.get_xticklabels()) plt.show()
```

group_neighbourhood

Manhattan	207.30
Brooklyn	160.19
Queens	132.94
Staten Island	128.89
Bronx	116.70
Name: price,	dtype: float64

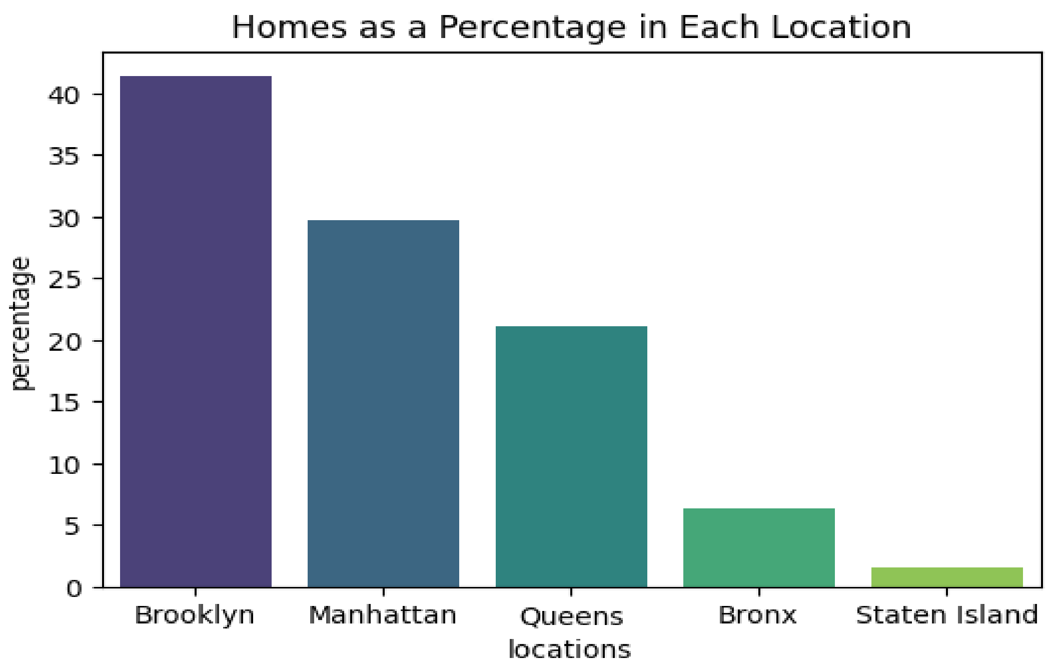


Given that it is the wealthiest and most populated area of New York, Manhattan is significantly more costly than the rest of the city.

Let's investigate how many rooms are available in each region before we wrap up.

```
bar.set_xticklabels(bar.get_xticklabels()) plt.show()
```

```
Brooklyn         8263
Manhattan        5941
Queens           4221
Bronx            1266
Staten Island     308
Name: group_neighbourhood, dtype: int64

Brooklyn         41.32
Manhattan        29.71
Queens           21.11
Bronx             6.33
Staten Island     1.54
Name: group_neighbourhood, dtype: float64
```



The Manhattan and Brooklyn regions have the most available rooms on Airbnb, while having the two highest average pricing. Thus, it follows that they are the most costly and in-demand places.

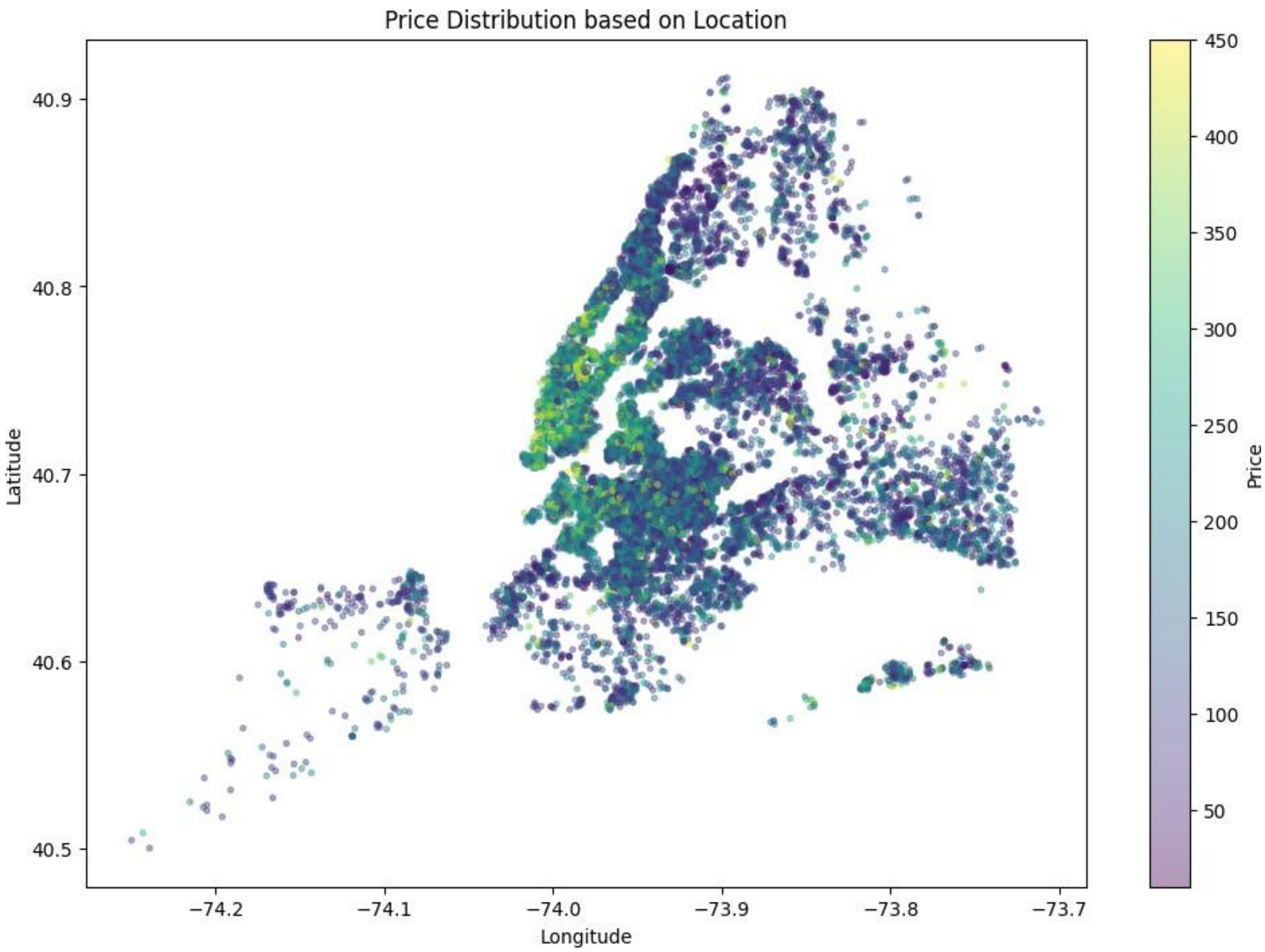
2.10 Exploring the Geographical Data

Given that we have some knowledge of the price distribution for rooms in New York City, let's map this information geographically and try to identify the areas of the city where average Airbnb costs are greater and lower. Our dataset's latitude and longitude columns will be used to make a scatter plot first.

In [ ]

```
import matplotlib.pyplot as plt


# plotting a scatter for the given information.
```

```
scatter = plt.scatter(x=newyrk_clean['longi'], y=newyrk_clean['lati'], alpha=0.4 plt.xlabel('Longitude')

plt.ylabel('Latitude')

plt.title('Price Distribution based on Location') plt.colorbar(scatter, label='Price')

plt.show()
```



However, maps like this make it difficult to visualise the city. We can observe certain trends in this particular instance because New York is one of the most well-known cities in the world. For instance, we can recognise Manhattan Island and Central Park, the white square in the island's centre.

With this in mind, it is simpler to notice that Lower Manhattan has more expensive pricing than the majority of the other areas shown on the map. There are some greater values in the area near the Brooklyn Bridge, and we can also see some very high prices on the east side of Queens.

The fact that we are familiar with the layout of the city enables us to have a general understanding of what is happening. It would have been more difficult to draw any conclusions from this plot if that had not been the case. we'll work with folium to improve this visualisation and produce a genuine map.

2.11 Generating visualization using Folium

With the help of the package Folium, it is simple to view data that has been altered in Python on an interactive leaflet map.

However, the dataset we are utilising does not contain a ZIP column, and Folium requires ZIP codes rather than latitude and longitude. To do this, we'll utilise a different datasetfrom Airbnb. Additionally accessible on Inside Airbnb is this dataset.

2.12 Data Preparation

This dataset is substantially larger than the one we are using, with 94 columns and a great deal more data. Read it into a dataframe now.

[ ]:

```
In [ ]: newyrkk.shape
Out[ ]: (24790, 80)
In [ ]: newyrkk.head()
```

Out[ ]:

	id	Reviews	listing_url	scrape_id	last_scr
0	2595.0	Notre séjour de trois nuits.\r Nous avons ...	https://www.airbnb.com/rooms/2595	2.020000e+13	05/06/
1	14991.0	My stay at Bianca's was lovely! The flat is co...	https://www.airbnb.com/rooms/14991	2.020000e+13	05/06/
2	5136.0	My family had a wonderful stay at Rebecca and ...	https://www.airbnb.com/rooms/5136	2.020000e+13	05/06/
3	15341.0	First off, it's an super spacious apartment. ...	https://www.airbnb.com/rooms/15341	2.020000e+13	05/06/
4	59709.0	Fun, airy, sunny, great apartment on the edge ...	https://www.airbnb.com/rooms/59709	2.020000e+13	05/06/

5 rows × 80 columns

Its id column has the same ids as those in the ny dataframe, as you can see. With just the two columns we require—the id and zipcode columns—we'll generate the zip dataframe using this updated dataset.

In [ ]:
```
zip_code = newyrkk[['id','zipcode']]
zip_code.shape
```

Out[ ]:
```
(24790, 2)
```

Let's then combine it with our other dataframe. We will only utilise the combined dataframe moving forward. Let's look at some details regarding it as well.

In [ ]:
```
data_merge = newyrk_clean.merge(right=zip_code, on='id',how='left')
data_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'> Int64Index:
20014 entries, 0 to 20013 Data columns (total 15
columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----

 0   id                            20014 non-null  float64
 1   property_name                 20014 non-null  object
 2   id_host                       20014 non-null  int64
 3   name_host                     20014 non-null  object
 4   group_neighbourhood           20014 non-null  object
 5   neighbourhood_name            20014 non-null  object
 6   lati                          20014 non-null  float64
 7   longi                         20014 non-null  float64
 8   type_room                     20014 non-null  object
 9   price                         20014 non-null  int64
 10  min_nights                    20014 non-null  int64
 11  no_of_reviews                 20014 non-null  int64
 12  calculate_host_listing_count  20014 non-null  int64
 13  availability_yearly           20014 non-null  int64
 14  zipcode                       10262 non-null  float64
```

dtypes: float64(4), int64(6), object(5)

memory usage: 2.4+ MB

Our consolidated dataframe has 41654 rows altogether. The zip column has several null values, which is terrible news. Let's check to determine whether that is a sizable sum.

In [ ]:

```
data_merge['zipcode'].isnull().sum() / len(data_merge) * 100
```

48.72589187568702

we have no null value in this data set as it is evident from the above result.

```
data_merge['zipcode'].unique()
```

array([5.90230000e+04, 7.37800000e+03, 1.86084000e+05, ...,
       4.43335070e+08, 2.36596830e+07, 1.08550734e+08])

```
data_merge = data_merge[data_merge['zipcode'] != '']
data_merge.reset_index(drop=True, inplace=True)
data_merge.shape
```

(20014, 15)

We have 41654 rows in the dataset after eliminating the null values.In our entirely clean dataset, there are still 41654 rows

2.13 Map Plotting Using Folium

- The map will then be generated. We'll take these five actions to do this: Create a figure, a map, and a layer of choropleths. Make some marks on the map that point to some well-known New York sights; Show the map.

```
# Building the figure

fig = folium.Figure(width=500, height=500)
```

The Choropleth's creation

airbnb_geo = folium.Choropleth(

geo_data='https://raw.githubusercontent.com/fedhere/PUI2015_EC/master/mam161

name = ‘newyork’

```
In [ ]: data_merge['zipcode'].isnull().sum() / len(data_merge) * 100
```

```
Out[ ]: 48.72589187568702
```

we have no null value in this data set as it is evident from the above result.

```
In [ ]: data_merge['zipcode'].unique()
```

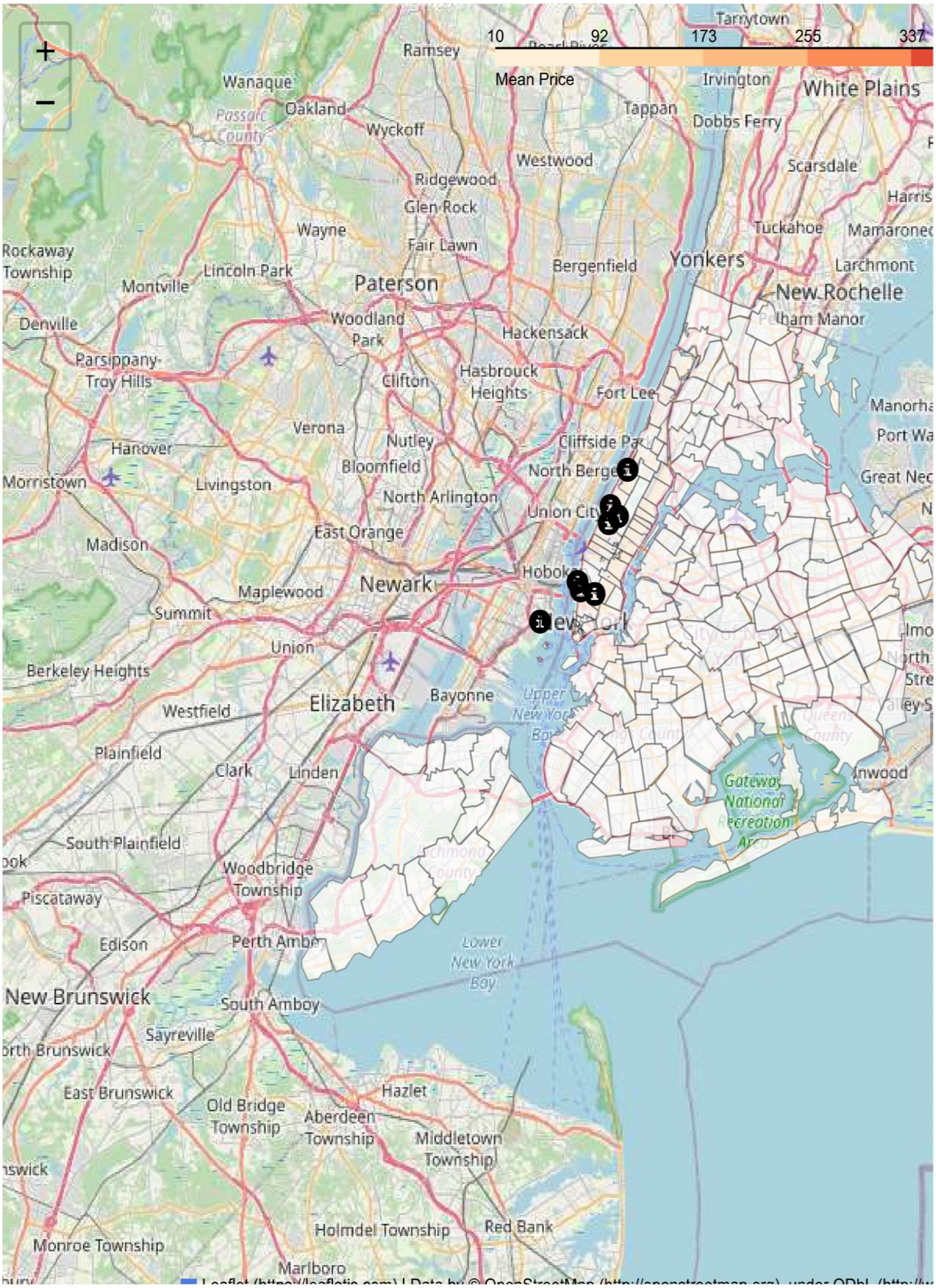
```
Out[ ]: array([5.90230000e+04, 7.37800000e+03, 1.86084000e+05, ...,
               4.43335070e+08, 2.36596830e+07, 1.08550734e+08])
```

```
In [ ]: data_merge = data_merge[data_merge['zipcode'] != '']
        data_merge.reset_index(drop=True, inplace=True)
        data_merge.shape
```

```
Out[ ]: (20014, 15)
```

```
In [ ]:     airbnb_geo = folium.Choropleth(

            geo_data='https://raw.githubusercontent.com/fedhere/PUI2015_EC/master/mam161
            name='newyork_airbnb',
            data=data_merge,
            columns=['zipcode', 'price'],
            key_on='feature.properties.postalCode',
            fill_color='BuPu',
            highlight=True,
            bins=8,
            nan_fill_color='gray',
            fill_opacity=0.8,
            line_opacity=0.5,
            legend_name='Average Price',
            reset=True
```

Out[72]:



10
92
173
255
337
Mean Price
Tarrytown
Ramsey
Wanaque
Oakland
Passaic County
Wyckoff
Irvington
White Plains
Tappan
Dobbs Ferry
Westwood
Scarsdale
Ridgewood
Glen Rock
Wayne
Fair Lawn
Tuckahoe
Rockaway Township
Montville
Lincoln Park
Paterson
Bergenfield
Yonkers
Larchmont
New Rochelle
Pelham Manor
Denville
Woodland Park
Hackensack
Parsippany-Troy Hills
Clifton
Hasbrouck Heights
Fort Lee
Verona
Nutley
Cliffside Park
Hanover
Bloomfield
North Bergen
Morristown
Livingston
North Arlington
Union City
Madison
East Orange
Hoboken
Maplewood
Newark
Summit
Union
Berkeley Heights
Elizabeth
Bayonne
Westfield
Upper New York Bay
Plainfield
Clark
Linden
Gateway National Recreation Area
Inwood
South Plainfield
Richmond County
Woodbridge Township
Piscataway
Edison
Perth Amboy
Lower New York Bay
New Brunswick
South Amboy
Sayreville
East Brunswick
Old Bridge Township
Hazlet
Aberdeen Township
Middletown Township
Holmdel Township
Red Bank
Monroe Township
Marlboro

We can now see how much an Airbnb in New York City costs. In the above map we can see clearly the expensive location of airbnb in New York city, also it was expected that in the Location Which is Manhattan and specifically in its Lower part, it is a highly costly area of the city. Additionally, it is home to the most well-known attractions, which undoubtedly contributes to the high rates by raising demand for lodging. Staten Island and the Bronx have affordable housing. They all have a more costly location, but as we observed earlier in the project, this may be due to the small number of rooms in these areas, which can skew the average price.

The cost of living varies greatly between Brooklyn and Queens. These places contain a large proportion of affordable neighbourhoods, but they also have a sizable number of wealthy neighbourhoods. For example, locations near Manhattan are often more costly.

As we indicated while discussing the scatter map, several neighbourhoods in the south of Brooklyn as well as the east side of Queens have some higher pricing.

We undertook significant data manipulation processes, including data exploration, cleansing, analysis, and visualisation, to achieve these objectives.

Having stated so, the findings are as follows:The two most prevalent types of rooms are private rooms and full apartments;Private and shared rooms are often more costly than hotel rooms and full residences; Manhattan and Brooklyn,which are also the costliest areas, are home to more than 80% of the rooms. Yes, you will probably have to pay extra if you want to stay close to the city's top attractions.

Some conclusions that we can make from the above analysis are : The two most prevalent types of rooms are individual spaces along with complete houses:

According to the research, individual rooms and full houses are the most common lodging options on Airbnb in city of New York Because of the solitude and convenience they offer, visitors frequently look for these choices. While complete houses are popular with individuals hoping for a experience like thier own home during their stay, separate spaces are popular with singles and couples seeking an equilibrium between security and expense.

Independent and shared lodgings are often pricier than hotel accommodations and full housings:

The report emphasises that compared to individual and communal lodging,

hotel lodgings and complete residences are frequently pricier. This is understandable given that motels frequently offer a variety of services and facilities while fully furnished housing offer an entire living experience. These choices are frequently chosen by tourists eager to spend extra for a deluxe or special stay. Manhattan and Brooklyn, which are also the most costly regions, are home to more than 75% of the accommodations:

The data clearly shows that Manhattan and Brooklyn are home to the greatest number of Airbnb properties, or about 75% of them. These areas are highly sought-after by tourists due to their popular sites, enterprises, and historical sites. Nevertheless, it's important to know that because these are desirable sites, lodging costs are frequently more here than they are in other parts of the metropolitan area.As we indicated while discussing the scatter map, several neighbourhoods in the south of Brooklyn as well as the east side of Queens have some higher pricing.

2.14 Conclusion For the socio_economic analysis

We undertook significant data manipulation processes, including data exploration, cleansing, analysis, and visualisation, to achieve these objectives.

Having stated so, the findings are as follows:The two most prevalent types of rooms are private rooms and full apartments;Private and shared rooms are often more costly than hotel rooms and full residences; Manhattan and Brooklyn,which are also the most costly areas, are home to more than 80% of the rooms. Yes, you will probably have to pay extra if you want to stay close to the city's top attractions.

Some conclusions that we can make from the above analysis are : The two most prevalent types of rooms are individual spaces along with complete houses:

According to the research, individual rooms and full houses are the most common lodging options on Airbnb in city of New York Because of the solitude and convenience they offer, visitors frequently look for these choices. While complete houses are popular with individuals hoping for a experience like thier own home during their stay, separate spaces are popular with singles and couples seeking an equilibrium between security and expense.

The cost of lodging near popular city attractions is frequently higher:

A further significant conclusion from the data is that visitors who want to stay close to New York City's most popular sights should be ready to devote a bigger amount of their money to lodging. There is frequently a cost associated with the comfort and accessibility to well-known destinations, leisure centres, and commercial areas. This discovery is essential for tourists trying to strike an equilibrium between price and area during their visit to the metropolis.

In a nutshell, the research provides useful information on the patterns and valuing dynamics of Airbnb lodging in New York City, enabling potential guests to make wise decisions on their personal tastes and financial constraints.

Chapter 3

Airbnb Customer satisfaction analysis in New York city

3.1 Introduction

A new socioeconomic system called the "sharing economy" enables people to briefly rent out their own possessions, such their cars or spare rooms in their homes. The purpose of this study is to look at the factors that affect consumers' happiness when utilising websites for private rentals in the sharing economy. Several approaches that we are going to be including analysis of sentiment, clustering of various words and regression which is ordinal logistic, and its visualisation, we are going to examine large data set of Airbnb's online reviews and listings in the New York city. Results show that the pricing, value, sanitation, rate, host interaction, ease of check-in, precision of the property description, and whether the host is a wonderful host or not all have a substantial influence on the polarity of Airbnb guests' evaluations in the New York city. Surprisingly, in New York, there was no evidence that the surrounding area of the property affected customer satisfaction. (L., 2019) (Back, 2019)

When utilising the Airbnb platform in New York, what factors contribute to consumer satisfaction?

Using analysis of sentiments, grouping of topic, regression, and visualisation approaches, the New York Airbnb online reviews dataset will be examined concurrently with the posting's information dataset.

3.2 Client happiness analysis.

cognitive assertion that includes the fulfilment of a need or want and the pleasure experienced as a result of such success. Clients' decisions, views, opinions, feelings, or emotional reactions in regard to their entire interactions with an item or service may be referred to as their level of fulfilment. Yet, increasing customer satisfaction comes at a cost, so firms must be careful not to overdo it and spend extravagantly without reaping the rewards. To achieve the maximum level of customer satisfaction at the lowest cost, businesses must understand their customers and what makes them happy. It may be challenging for many firms to satisfy clients. (DDDA, 2019)
The initial phase in increasing client fulfilment is to ascertain current levels and the elements that led to them. Assessment and documentation of the client's experience are excellent ways to do this. Customers' reviews and remarks, which they left after obtaining an experience, might be examined to discover more about how satisfied they were. Opinions posted by Airbnb customers will be examined in order to comprehend customer contentment and the factors that affect it better. (Lau, 2016)The perceptions and satisfaction of clients of Airbnb have been studied using a variety of academic approaches, some focusing on user experience and technology and others on

administration, travel, and construction. A broad spectrum of subjects are covered, such as review prejudice, the hospitality markets, community ranking, cost and community projections,
rental distribution, user feels and requests, picture extraction, request mining, assessing schema, integrating schema, and evaluation of population. (Sari R, 2019)

3.3 **Methodology , Data Collection and Preparation**

For academics, company owners, and other consumers, analysing internet review comments has become increasingly significant and valuable over the years, according to a large body of literature. Since user comments depict the opinions and emotions they have when utilising a service or an item, they may be utilised to gauge and analyse users' levels of satisfaction as well as pinpoint crucial variables that contributed to these levels of happiness. (Fensel, 2017)
Customers publish evaluations about their experiences and thoughts on the accommodations they stay at on the Airbnb website. This gives information that may be used to analyse the variables that influence customer satisfaction. The present study included specific listing information obtained from the Inside Airbnb platform as well as reviews left by visitors to Airbnb on their stays in New York. (cuquet m, 2017)

The two datasets were easily obtainable due to the fact that Airbnb data is widely available in the sites which are normal in common people. The review dataset has six traits and more than 6500 count of rows, while the listing dataset has more than 80 attributes that characterise the host, house characteristics, and scored ratings. Then, prior to the analysis phase, data sets were first processed and cleaned because doing this step frequently leads in findings that are usable. The cleaning method for the listings dataset consisted of removing characteristics with faults, 0 vales (blanks), and unnecessary vales. (Fuentes, 2017) After then, other vertical line data were created or deduced from already-present properties, such as guest involvement in month deduced from the host's beginning time. (Hidayaanto an, 2018)Columns like id comnt, id review, and username that were unwanted were removed. In order to reduce the number of ratings and make the data comprehensible for analysis using home devices with constrained capabilities, a filter was further applied using the day of the remark, picking only responses from a certain period of time. Dummy contributions made of characters or numbers were also eliminated. (osborrne, 2020)

3.4. Data Analysis

Sentiments analysis: Sentiment analysis was the initial step in the analytical process. Excel's Text Analytics module was used to do the analysis of the sentiment. Two separate documents are generated by this analysis of satisfaction. The initial one is for sorting remarks by subject, including areas classifications and topics kinds (this were subsequently used in the mining of the words), and the other one that is the one is for sorting comments by polarity, such as good, bad, or balanced , and a trust score.

Regression :analysis of satisfaction of customers was the initial step in the analytical process. Excel's Text Analytics module was used to do the analysis of the sentiment. Finding the variables influencing consumers' happiness was the second phase in the investigation, which was regression analysis. The two datasets were combined to launch this stage. By identifying the vales which are of the independent values (IV) has a statistic_important validity on our dependent values (DV), that is polarity, the logistic ordinal (OLR) regression test was adopted to address the study topic. The variable which are depend is variable which is ordinal and also categorisedxgfre32, with a value indicated by a scale of three values which are (bad (n)=0, neutral (NEU)=1, and good(P)=2), and the values which are independent are a mixture of both qualitative and categorised information, which is why the OL regression test was used for this study. Only two values which are (accommodation and number of beds) had a multicollinearity coefficient value greater than 0.8, indicating they were firmly linked. As a result, the beds variable was eliminated from the test. In the pursuit of interpretation , we will take a look for contributing variable of the values that are dependent would be difficult without performing the test to determine multicollinearity. The final set of variables contained 20 attributes after the highly correlated variables were removed. And finally, the visualizations were created.

3.5 Analysis of sentiment and clustering of the topics.

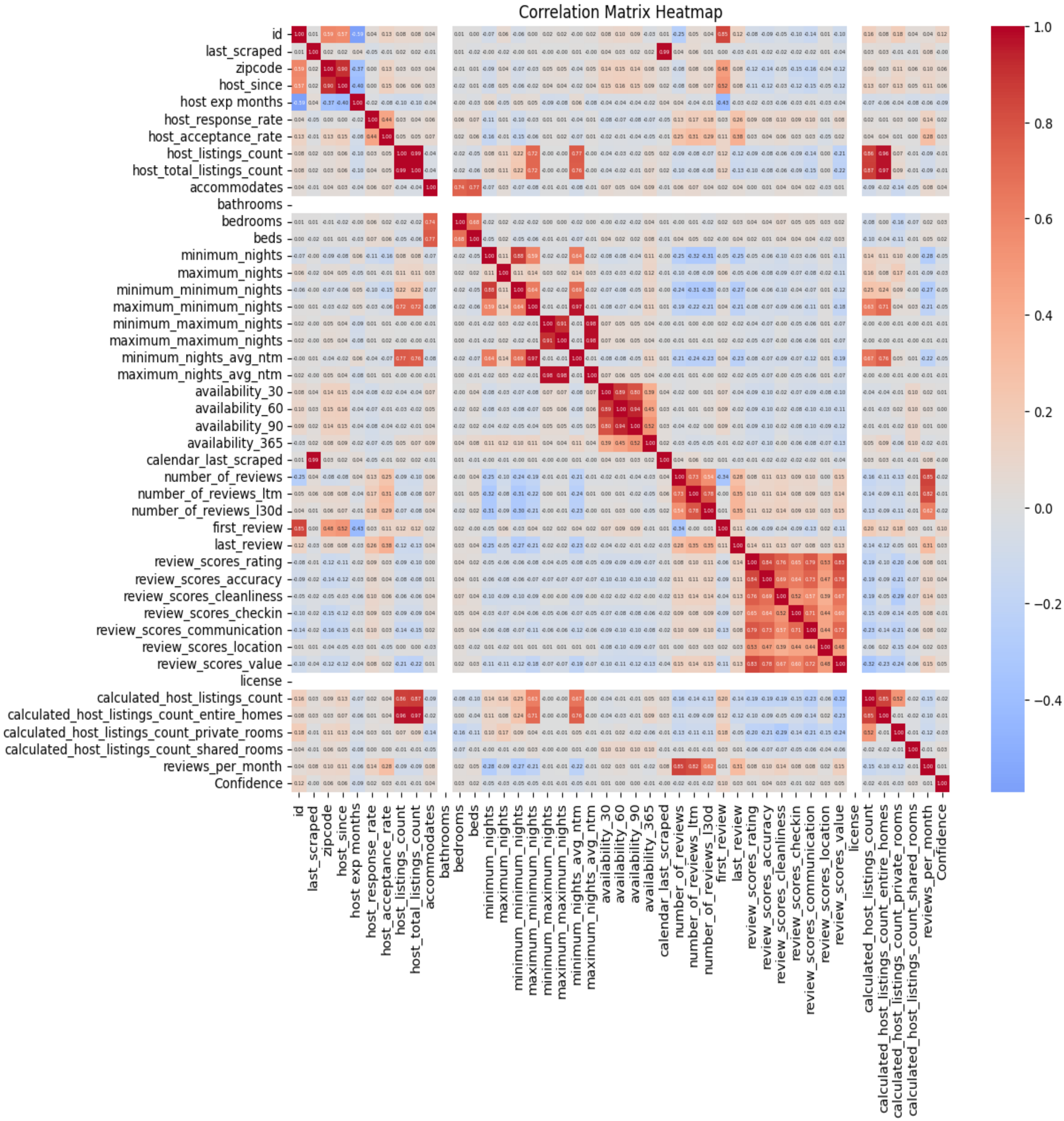
According to the sentiment analysis's findings, New York's Airbnb users were generally having a good experience. Users expressed satisfaction with the location and host of their encounter, according to the topic clustering results. In this case, a location's probability score of 51% indicates that 51% of the favourable evaluations contain the word "location."

Since consumers worry about the closeness of various amenities to their homes, such as public transit, restaurants, and stores, location is a key predictor of their happiness. The phrase "host" is the second most prevalent theme in the good results, accounting for 30.89% of the favourable results, showing that many users were content with the way the people who hosted them were handling them. Check-in, on the other hand, was the subject that regularly generated an unfavourable response. The majority of the unfavourable attitude is represented by the phrase "checkin," which has a probability score of 88.67%. This outcome serves as a signal to hosts to enhance their checkin capabilities in order to make the procedure quick, effective, and seamless. You may accomplish this by according to the official Airbnb checkin procedure instructions. The word clusters in the table below reflect the previous results. Though the majority of reviews on Airbnb are good, looking more closely at the areas where there have been complaints enables the people living in the air-bnb to address these issues by correcting any issues or establishing objectives for the future (for example, upgrading checkin procedures). A further indication that New York's overall atmosphere can play a crucial role in determining the favourable impressions of those staying on Airbnb is the sturdy presence of subjects associated to the city's surroundings among favourable reviews as opposed to their frail occurrence amongst negative ones (such as place of residence, services, conveyance, bistros, estates and factories).

Good			Bad		
Term	Chances	Score	Term	Chances	Score
locality	51.67%	5168	Check in	88.67%	8861
host	30.89%	3081	Host	2.21%	226
facility	3.37%	335	Location	1.87%	188
Clean	1.89%	185	Room	0.95%	98
Room	1.75%	176	Facility	0.53%	57
Walk	1.55%	157	Bed	0.23%	28
Transportation	1.31%	135	Transportation	0.18%	19
Bed	1.22%	126			
Great	0.89%	85			

3.6 Generating the correlation Matrix

```
n [23]: import pandas as pd

    import numpy as np import
    statsmodels.api as sm from
    scipy.stats import spearmanr

In [25]: data = pd.read_excel('C://Users/assel/OneDrive/Desktop/airbnb project/excel python

In [26]: categorical_columns = ['source', 'host_location', 'host_response_time', 'host_verif
 'neighbourhood', 'neighbourhood_cleansed', 'neighbourhood_gr', 'room_type', 'bathrooms_text',
 'has_availability', 'instant_

   In [22]: plt.fig(figurize=(11, 10)) sns.heat_map(corelation_matrix,)
, annot_kws={"size": 4}, ccmap='colwam', plt.title('Corelation_Matrix Heat_map') plt.show()
```


Correlation Matrix Heatmap
id
last_scraped
zipcode
host_since
host exp months
host_response_rate
host_acceptance_rate
host_listings_count
host_total_listings_count
accommodates
bathrooms
bedrooms
beds
minimum_nights
maximum_nights
minimum_minimum_nights
maximum_minimum_nights
minimum_maximum_nights
maximum_maximum_nights
minimum_nights_avg_ntm
maximum_nights_avg_ntm
availability_30
availability_60
availability_90
availability_365
calendar_last_scraped
number_of_reviews
number_of_reviews_ltm
number_of_reviews_l30d
first_review
last_review
review_scores_rating
review_scores_accuracy
review_scores_cleanliness
review_scores_checkin
review_scores_communication
review_scores_location
review_scores_value
license
calculated_host_listings_count
calculated_host_listings_count_entire_homes
calculated_host_listings_count_private_rooms
calculated_host_listings_count_shared_rooms
reviews_per_month
Confidence
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

3.7 Correlation matrix and the analysis of sentiment dataset:

The correlation coefficients between variables in a dataset are shown visually in a correlation heatmap matrix. The intensity and direction of the correlations between the pairs of variables are represented by colours. connection coefficients typically fall between positive 1 and 1, with negative 1 denoting a perfect adverse connection, 1 denoting a perfect relationship that is positive, and 0 denoting no connection at all.Correlation values are displayed on the heatmap using an assigned colour scale. Some common colour palettes are:

Correlations that are favourable are represented by red shades, with shades of red indicating higher associations. Darker coloration of blue represents unfavourable associations, and lighter hues represent weaker correlations that are adverse. For 0% correlation (no association), use white or an unaltered shade.

The chart's horizontal and vertical values match the variables that are being associated. The x and y axes of each variable are labelled. The association across the two parameters is represented by the connection of a row and a column. In order to uncover trends and make knowledgeable judgements in data analysis and modelling, this correlation heat chart grid will provide us a visual depiction of how variables in a dataset are linked to one another.

The last collection of parameters, after removing the most strongly associated factors included 28 attributes, including the following data points: experience_host_months, time_response_host, rate_response_host, host_super, total_count_list, pic_host_profile, identity_verified_host, cleaned_neighbourhood, type_property, type_room, bath_accmodates, bed_rooms, type_beds, cost and review_communication.

3.8 Performing a logistic regression which is ordinal on the dataset :

Descrip.	Attribut	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	*Upper Bound*
Where neutral is zero that sentiment	[Polarity=0]	-33.729	9049.911	.000	1	.997	-17771.229	17703.772
Time for long since this property has been noted in months	**Host_in_month_Exp**	**.029**	**.001**	**836.171**	**1**	**<0.001**	**0.027**	**0.031**
The interaction rate of the owner of the house with the guests	Host_rate_Respo	0.314	0.321	0.957	1	0.328	-0.315	.943
Total lising count of property listed by the host	Host_count_list_Tot	.000	6.604e-5	5.120	1	0.024	0.000	-2.000e-5
Review ranges from 1 to 5 where 1 is the lowest rate and 5 is the highest rate	Review_scor_clean	0.142	0.118	1.456	1	0.228	-0.089	0.373
	Review_score_check	0.107	0.17	0.396	1	0.529	-0.227	0.441
	Review_scor_rate	-0.594	0.212	7.82	1	0.005	-1.01	-0.178
	Review_scor_accurate	-0.172	0.171	1.019	1	0.313	-0.507	0.162
	Review_score_interact	0.495	0.168	8.677	1	0.003	0.166	0.824
	Review_scor_area	-0.346	0.12	8.269	1	0.004	-0.582	-0.11
	Review_score_value	0.266	0.149	3.201	1	0.074	-0.025	0.557
	Accommodation of 1	-0.222	1.298	0.029	1	0.864	-2.766	2.321
	Accommodation of 2	-0.126	1.294	0.009	1	0.922	-2.663	2.411
	Accommodation of 3	-0.236	1.296	0.033	1	0.855	-2.776	2.304

The number of guests accommodated by a given listing.	Accommodation of 4	-0.315	1.294	0.059	1	0.808	-2.852	2.222
	Accommodation of 5	-0.221	1.297	0.029	1	0.865	-2.763	2.321
	Accommodation of 6	-0.263	1.294	0.041	1	0.839	-2.799	2.274
	Accommodation of 7	-0.152	1.309	0.013	1	0.908	-2.718	2.414
	Accommodation of 8	-0.151	1.301	0.014	1	0.907	-2.702	2.399
	Accommodation of 9	-0.45	1.362	0.109	1	0.741	-3.119	2.219
	Accommodation of 10	-0.871	1.327	0.431	1	0.511	-3.472	1.729
	Accommodation of 11	0.511	1.403	0.132	1	0.716	-2.24	3.261
	Accommodation of **12**	-0.45	1.376	0.107	1	0.744	-3.147	2.247
	Accommodation of **13**	-1.003	1.736	0.334	1	0.563	-4.405	2.398
	Accommodation of **14**	0.908	1.655	0.301	1	0.583	-2.337	4.152
	Accommodation of **15**	-0.82	1.64	0.25	1	0.617	-4.035	2.395
	Accommodation of **16**	**-.45**	**1.265**	**0.250**	**1**	**0.619**	**-0.2348**	**1.945**
	No. of bathrooms = 1	-18.852	0.815	534.384	1	<.001	-20.45	-17.253
	No. of bathrooms =1.5	-18.794	0.823	521.139	1	<.001	-20.407	-17.18
	No. of bathrooms =2	-19.061	0.82	540.281	1	<.001	-20.669	-17.454
	No. of bathrooms =2.5	-18.894	0.848	495.886	1	<.001	-20.557	-17.231
	No. of bathrooms =3	-19.02	0.863	486.232	1	<.001	-20.71	-17.329
	No. of bathrooms =3.5	-18.243	0.93	384.691	1	<.001	-20.066	-16.42

Number of bathrooms	No. of bathrooms =4	-18.645	1.054	313.036	1	<.001	-20.71	-16.58
Bookable instantly yes = 1, bookable instantly no = 0	Bookable instantly = yes	-.585	.779	.617	1	.399	-2.245	.891
	Bookable instantly = no	0a	.	.	0	.	.	.
Number of bedrooms	No. of Bedrooms = 0	2.414	2.486	0.943	1	0.332	-2.459	7.287
	No. of Bedrooms = 1	2.569	2.485	1.069	1	0.301	-2.301	7.44
	No. of Bedrooms = 2	2.52	2.485	1.029	1	0.31	-2.35	7.39
	No. of Bedrooms = 3	2.17	2.488	0.761	1	0.383	-2.707	7.047
	No. of Bedrooms = 4	2.157	2.503	0.743	1	0.389	-2.748	7.062
	No. of Bedrooms = 5	2.866	2.532	1.282	1	0.258	-2.096	7.828
	No. of Bedrooms = 6	2.295	2.732	0.706	1	0.401	-3.06	7.651
	No. of Bedrooms = 7	2.19	3.542	0.382	1	0.536	-4.751	9.132
Listing price in dollar	Price	0	0	0.64	1	0.424	-0.001	0
Interaction in an hour = 0, a few hours = 1, a day = 2, a few days or more = 3	[host_respond_time=0]	0.802	0.31	6.676	1	0.01	0.194	1.41
	[host_ respond _time=1]	0.408	0.094	18.795	1	<.001	0.223	0.592
	[host_ respond _time=2]	0.289	0.065	19.995	1	<.001	0.162	0.415
	[host_ respond _time=3]	1.154	.893	1.671	1	.196	-.596	2.904
Does host have profile picture? No = 0, Yes = 1	[host_profile_image=0]	-0.299	0.66	0.205	1	0.651	-1.592	0.995
	[host_profile_image =1]	0[a]	.	.	0	.	.	.
	[host_himself_verification=0]	0.035	0.086	0.168	1	0.682	-0.133	0.203

Is host's identity verified? No = 0, Yes = 1	[host_himself_verification 1]	0[a]	.	.	0	.	.	.
1 = entire rental unit, 2 = Entire condo, 3 = private room rental unit, 4 = private room townhouse, 5 = private room condo, 6 = private room in loft, 7 = private room in home, 8 = Entire loft, 9 = Entire home , 10 = Entire townhouse, 11 = Entire guest suite, 12 = Entire guesthouse, 13 = private room in guest suite, 14 = private room in houseboat, 15 = Private room, 16 =Entire place, 17 = Shared room in rental unit, 18= Private room in guesthouse, 19 = Entire serviced apartment, 20 =Boat, 21 =	[type_house = 1]	3.328	2.495	1.779	1	0.182	-1.562	8.219
	[type_house =2]	3.17	2.499	1.609	1	0.205	-1.728	8.068
	[type_house =3]	1.837	3.576	0.264	1	0.607	-5.172	8.846
	[type_house =4]	1.715	3.577	0.23	1	0.632	-5.297	8.726
	[type_house =5]	1.871	3.58	0.273	1	0.601	-5.147	8.889
	[type_house =6]	1.601	3.583	0.2	1	0.655	-5.422	8.623
	[type_house =7]	1.89	3.577	0.279	1	0.597	-5.12	8.9
	[type_house =8]	3.31	2.501	1.752	1	0.186	-1.591	8.212
	[type_house= 9]	3.422	2.498	1.876	1	0.171	-1.474	8.318
	[type_house=10]	3.441	2.499	1.895	1	0.169	-1.458	8.339
	[type_house =11]	2.89	2.501	1.336	1	0.248	-2.011	7.792
	[type_house =12]	3.501	2.536	1.906	1	0.167	-1.47	8.471
	[type_house =13]	1.164	3.593	0.105	1	0.746	-5.879	8.207
	[type_house =14]	-1.359	3.98	0.117	1	0.733	-9.16	6.442
	[type_house =15]	2.551	3.676	0.482	1	0.488	-4.653	9.756
	[type_house =16]	2.417	2.529	0.913	1	0.339	-2.54	7.375
	[type_house =17]	1.927	3.588	0.288	1	0.591	-5.106	8.959
	[type_house =18]	1.882	3.665	0.264	1	0.608	-5.301	9.066
	[type_house =19]	3.539	2.512	1.985	1	0.159	-1.384	8.462
	[type_house =20]	21.049	6229.117	0	1	0.997	-12187.796	12229.893
	[type_house =21]	2.031	3.626	0.314	1	0.575	-5.076	9.137
	[type_house =22]	2.498	2.69	0.862	1	0.353	-2.775	7.771
	[type_house 23]	1.986	3.63	0.299	1	0.584	-5.129	9.102
	[type_house =24]	1.716	3.695	0.216	1	0.642	-5.527	8.958
	[type_house =25]	2.147	3.692	0.338	1	0.561	-5.089	9.383
	[type_house =26]	0.725	3.714	0.038	1	0.845	-6.555	8.004
	[type_house =27]	3.221	2.727	1.395	1	0.238	-2.125	8.566
	[type_house e=28]	3.002	2.596	1.337	1	0.247	-2.086	8.089
	[type_house =29]	1.788	3.726	0.23	1	0.631	-5.516	9.091
	[type_house =30]	1.843	3.611	0.26	1	0.61	-5.234	8.92
	[type_house =31]	2.954	3.703	0.636	1	0.425	-4.303	10.211

Shared room in home, 22 = Entire cottage, 23 = Private room in bed and breakfast, 24 = Shared room in loft, 25 = Private room in serviced apartment, 26 = Room in serviced apartment, 27 = Tiny home, 28 = Entire bungalow, 29 = Shared room in condo, 30 = Room in hotel, 31 = Shared room in townhouse, 32 =Floor, 33 = Private room in hostel, 34 = Houseboat, 35 = Room in boutique hotel, 36 = Room in aparthotel, 37 = Private room in resort, 38 = Entire villa, 39 = Shared room in guesthouse, 40 = Private room in tiny home. 41 = Private room in earthen home, 42 = Private room in	[type_house =32]	-17.97	5797.919	0	1	0.998	-11381.682	11345.742
	[type_house =33]	20.272	9052.056	0	1	0.998	-17721.431	17761.975
	[type_house e=34]	22.213	9052.055	0	1	0.998	-17719.489	17763.915
	[type_house =35]	1.391	3.588	0.15	1	0.698	-5.64	8.423
	[type_house e=36]	1.553	2.569	0.366	1	0.545	-3.481	6.588
	[type_house =37]	-19.121	5797.92	0	1	0.997	-11382.835	11344.592
	[type_house 38]	3.52	3.549	0.984	1	0.321	-3.436	10.477
	[type_house =39]	-0.782	4.357	0.032	1	0.858	-9.322	7.759
	[type_house =40]	1.055	3.969	0.071	1	0.79	-6.723	8.834
	[type_house e=41]	1.791	3.982	0.202	1	0.653	-6.014	9.597
	[type_house e=42]	0.638	4.355	0.021	1	0.884	-7.897	9.173
	[type_house =43]	0.404	3.538	0.013	1	0.909	-6.529	7.337

bungalow, 43 = Barn								
Shared room = 1, private room = 2 and entire home = 3.	[type_ room=1]	-1.586	2.558	0.385	1	0.535	-6.599	3.427
	[type_ room =2]	-0.092	0.462	0.039	1	0.843	-0.998	0.814
	[type_ room =3]	0a	.	.	0	.	.	.

The above table we have obtained from SPSS after performing the regression which is logistic ordinal, we have obtained the following columns :

Description : This heading gives an explanation of the characteristic or predictor variable that is being studied.

Attribute : Describes the particular subcategory or level of the predictor variable under study. Typically, in this type of regression which is ordinal_logistic, they are the values of the ordinal responder variable.

Estimate : It is a representation of the estimated log-odds (logit) of the outcome variable for a change of 1 unit in the variable that serves as the predictor. The variances connected to each predictor level are represented by these estimations.

Std_error : gives a rough approximation of the average deviation of the estimate's distribution of the sample. It aids in comprehending the accuracy of the calculated coefficients.

Wald : Estimate ^2 divided by matching standard error is the Wald statistic. It is used to test if the parameter estimate is equal to zero (no impact), which is the null assumption.

Sig : represents the Wald statistic's related p-value. If the null assumption (no impact) is true, it represents the likelihood of seeing the calculated coefficient.

Interval_confidence 95% and its lower and bound upper : This gives the 95% trust interval's lower and upper boundaries for the calculated parameter. It displays the scope in which the actual population value falls with a level of trust of 95%.

We will concentrate on the interpretation of the estimates log_odds and their corresponding levels of significance while discussing the table of ordinal logistic regression. Based on the figures, we will determine if a variable that can be predicted has a substantial impact on the ordinal response variable and the trajectory of that impact.

The estimate for the predictor factor 'Attribute' suggests that, on a median, the odd of logs of being in the top group of the reply statistic rise by 'Estimate' for a 1-unit increase in 'Attribute. The ordinal answer to this estimate is statistically significant with a value_p of "Sig," indicating that "Attribute" has a major effect on it.

In the above table the regression on the attribute which is host experience in month that is, the number of months a host has registered his airbnb on the market in months. If we look at its est_estimate value , its 0.29 in the odd_logs which means that in a span of a month , the raise is 0.29 in host_Experience. Next is the std_error which erro_standard which is less than 0.01, this lower score of error_Standard Implies an accurate estimate. This is followed by Wald value which for this factor is 837.6 which is good Since a high Wald score often denotes a more substantial effect. degrees of freedom for the Wald statistic in this instance, which is one. For hypotheses to be tested, this is helpful. As we are aware that the Wald statistic's related p_value. The p_value in this instance is less than 0.001, demonstrating the great statistical significance of the influence of host experience in months.

Interval of 95% Trust (CI): Lower_Bound: The bottom limit of the estimate's 95% accuracy interval. For the sake of this research, it is around 0.0027. Upper_Bound: The highest point inside the estimate's 95% accuracy range. For the purpose of the present inquiry, it is around 0.031. The range of values between which we may be 95% certain that the true parameter for the population (impact_of_host_experience_in _months) resides is known as the confidence_interval with a level of 95%. So in summary for this factor, we can say that A value of 0.29 shows that for every one month more that the host experience, the log_odds of being in the top group of the response to the variable rise. This impact appears to be precise and statistically significant based on the low p_value (0.001) and the small trust interval.

Some other values which were found to be statistically significant and having value_p of less than 0.05 were review_rating_score : As greater review scores make the place to stay and the host appear more favourable to prospective visitors. Due to the trust, it inspires and the guarantee of a pleasant visit, visitors are more inclined to select a location with high evaluations, the other factors were score_review_Communication, time_response_hour, rate_response : as the amount of time a host takes in order to reach out to the customer is ofcourse considered as a good attitude towards customer and the quality of communication the host maintains with the guests is also vital. Other significant factors were 11_people_accomodation, cleanliness_rate : as the more reviews on cleanliness and hygiene, the more customer are attracted, in todays world with the rise of certain diseases it is mandatory for every accommodation business to maintain certain level of cleanliness and hygiene and other factors was property able to accommodate 11 people which is quite surprising as people are likely to arrange a gathering of more then 10 people in the property.

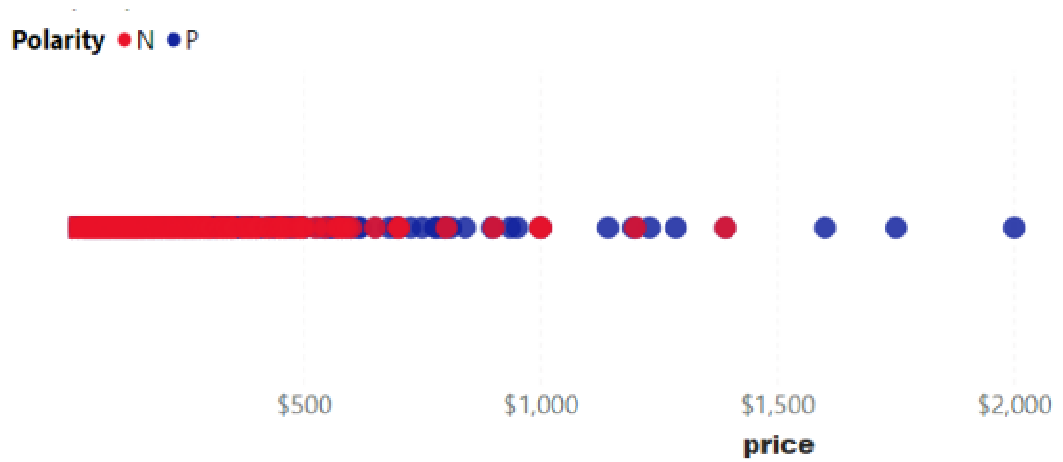
The Est_Estimate values for rate_response, rate_cleanliness, verified_identity_host, experience_of_host_months was also high with values greater than 0.10 this tells us that in order to be good host and to attract more customers, the host should be timely responsive and have good communication with the guests. Also, if the experience of the host with the property is more in time implies to customer trusting the host much more and the chances of property going on rent increases.

4 Conclusion for sentiment_analysis

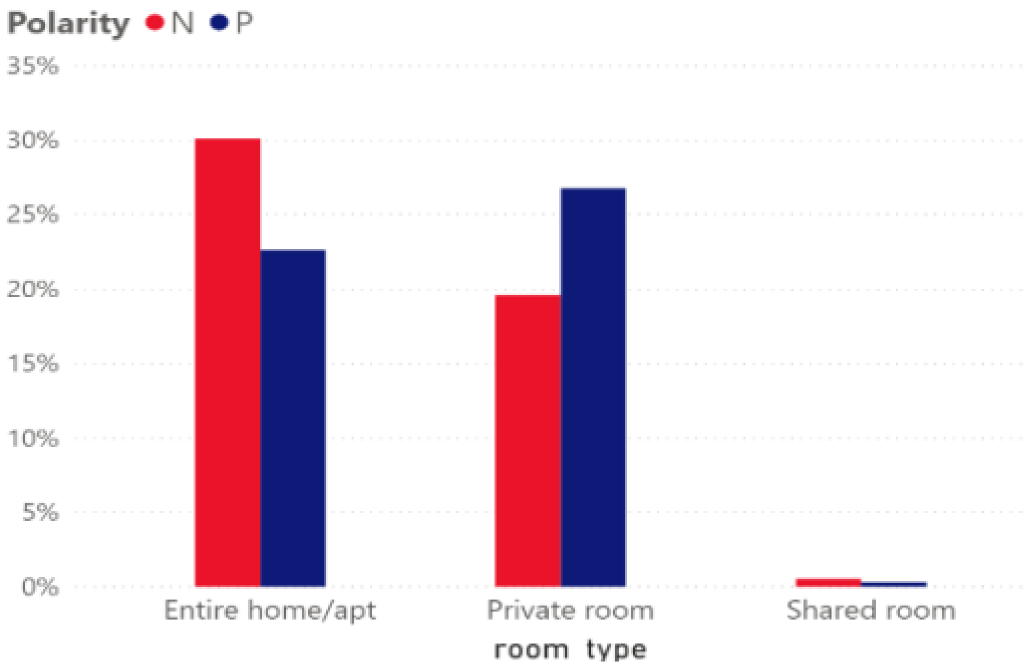
4.1 Polarity reviews and the influencing factors.

Price, room type (entire residence or apartment), the fact that the host is not a super, review_score_assessment, review_score_accuracy, cleanliness, review_ score checkin, reviewscore communication, and review score value are the factors that significantly affect guests' satisfaction after using an Ordinal logistic test.

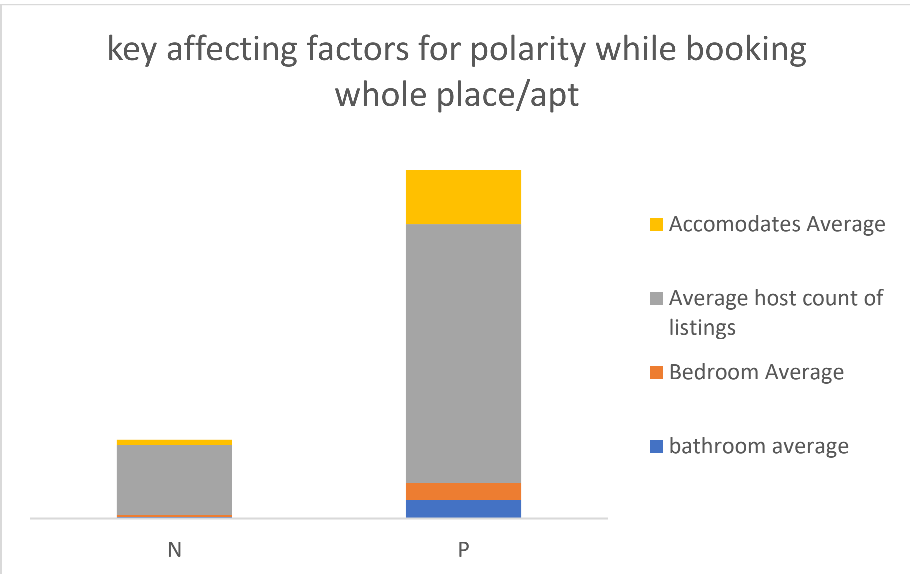
1) Price_ : Cost is the first element that affects customers' happiness, with a value of p less than 0.05. There is a correlation between real estate cost on a daily basis and the mood of reviews. The graph demonstrates that the number of bad reviews decreases as the cost rises. In other words, high-priced residences are more frequently connected with favourable reviews than bad ones, and vice versa. Additionally, residences with certain characteristics which are more luxurious are typically more expensive than similar properties. Therefore, it is anticipated that the advantages of having opulent features would encourage clients to spend more. (Larpin, 2019)



2) Room category: The 2nd variable to consider is the kind of space, which might be a whole house or flat, a private room, or a room that is shared. The graph below displays, for each of the 4 feature kinds, the proportion of good and bad comments based on identical polarity. unfavourable remarks for complete homes or apartments, for illustration, account for 30% of the 4 groups' combined unfavourable feedback. Reviews of this kind of complete home or flat are more likely to be unfavourable than favourable. Several aspects that influence the complete home or flat polarity after evaluating the mean of many parameters depending on user emotion (good versus bad). The outcome suggests that having less standard square feet has an impact on feeling bad. (Chattopadhyay, 2019)



Additionally, as seen the below figure, the likelihood of receiving a disappointed visitor will rise as host postings expand. Consequently, it is proof that there is a poor association between host total postings and customer happiness. This can be the result of renters being too busy managing all of their properties at once to provide excellent client service. Over time, hosts will fall short of providing their visitors with the standard of care. (Teuber, 2016)



host_super: The absence of a superhost is the 3rd influencing element. The graph below indicates that whether an individual is a superhost or not, it will not alter the favourable rating of an impression by displaying a comparable distribution of superhost within the favourable remarks. On the other hand, the fact that the host isn't a super-host significantly affects the unfavourable feelings. This is consistent with results that show that needy services have a bigger influence on compensation levels than valuable amenities.

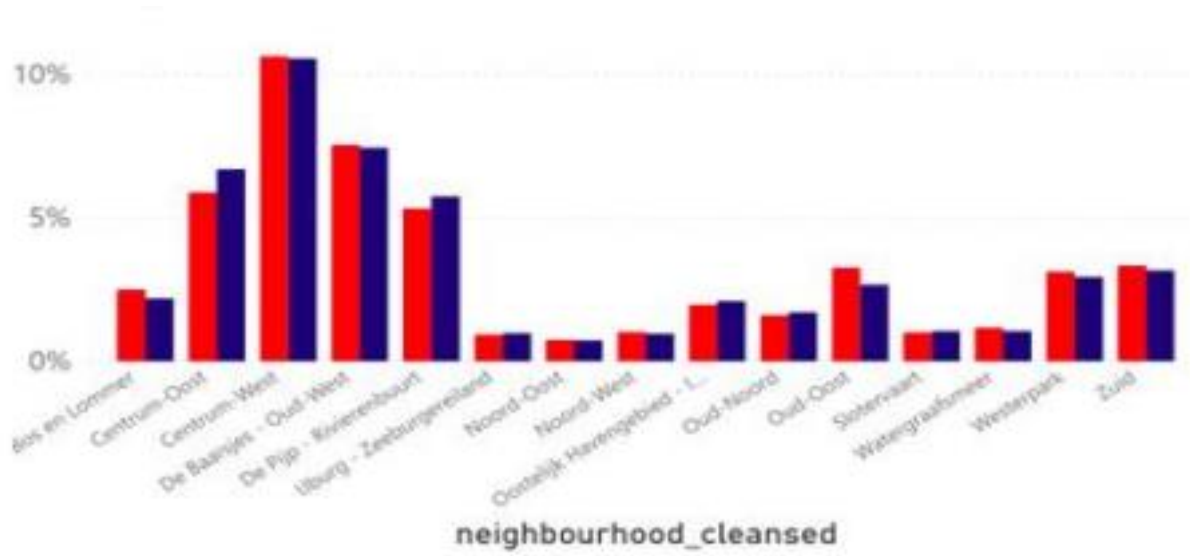
Individuals frequently have particular standards for quality. While missing this level will have a major bad impact on how they feel, achieving it will not have a good impact on it. (Line, 2014)

4.2 Various listing locations and the distribution of polarity.

Polarity ● N ● P

Neighbouring Area: Surprisingly, the ordinal logistic test showed that neighbouring area wasn't impacting the polarity as shown in the above table, despite the fact that the area plays a significant role as an effecting element on topic clustering findings.After carefully examining the findings, it appears that the precise location inside each neighbouring area has the most influence, not the neighbouring area as a whole. The difference in site within the same neighbouring area zone in terms of resources, tranquilly, transit, etc. is what has led to this predicament.

The association between polarity and neighbouring area was depicted in the below figure, using geographies and a chart with bars to support the Ordinal logistic finding. The bar chart, which shows that each neighbouring area essentially has an identical dispersion rate of good and bad responses, supports the finding that the mappings show a comparable pattern of both of the distribution of responses.



A thriving example of the collaborative system in overall, and specifically in the tourism business, is Air-bnb. Although there is still a lot of discussion about Air-bnb in the industry, there has only been limited research on the subject. The analysis of Air-bnb listings' attributes as influences on feedback polarity is being done. Costs, whether it is a super_host which is host, and the style of housing are the key determinants of clients' enjoyment, according to the writers. Additionally, in the topic grouping test, the phrase location appeared most frequently in good remarks, whereas checkin did so in poor ones.

nevertheless, it has been demonstrated that New York as a city has had an effect on Airbnb's clients' contentment and may serve as a justification for favourable reviews. By examining variables as a proportion of each component polarity out of the same in general polarity, this study aimed to prevent the biased sample effect. There are certain restrictions on this research with regard to this endeavour. As there is a restriction with a non-random sample, the usage of a single sample, like New York, might have an impact on the outcome on other places. (Falk, 2019)

4.3 Overall Conclusion from socio_economic and sentiment analysis for city of New York

An in-depth examination of Airbnb's operations in New York City has uncovered an immense amount of information on how both socioeconomic factors and customer perceptions have affected the company's operations. We have developed a comprehensive grasp of Airbnb's impact on the regional economy, its adherence to socio_economic considerations, and the essential components that fuel consumer pleasure by exploring these areas.

The socioeconomic research provided light on popular accommodation kinds and significant lodging choices, providing interesting insights into the nature of Airbnb listings around the city of New York. Individual rooms and whole condos appeared as the most common choices among the wide range of accommodation possibilities, with guests usually choosing these options for their personal space and convenience. Solo apartments are preferred by individuals and couples looking for a compromise between privacy and cost, while whole residences are particularly popular with people wishing an experience similar to their own home. The research also revealed an important trend in lodgings costs. Compared to lodging and whole homes, independent and shared lodgings are frequently more expensive. This difference might be ascribed to lodges' more extensive range of amenities and services, which makes them a more costly choice. Well-furnished accommodations, on the other hand, give visitors a full living experience, making them a desirable option for those prepared to spend more funds on an opulent or special stay.

The data also highlighted the concentration of Airbnb lodgings in particular neighbourhoods, namely Manhattan and Brooklyn. Due to their well-liked attractions, various businesses, and historical landmarks, these regions accounted for more than seventy-five percent of Airbnb properties, making them highly sought-after by travellers. It's crucial to keep in mind that the popularity of certain regions frequently results in increased hotel expenses, forcing potential visitors to choose between locality and costs.

This comprehensive study of Airbnb in the Metropolis of New York City has shed light on the healthy link between it and the socioeconomic structure of the city. These observations provide a solid basis for strategic planning as well as well-informed choice-making in the dynamic short-term rental market.

Insightful information about the elements affecting customer happiness and influencing booking choices in the Airbnb ecosystem has been gleaned from the analysis of sentiment of customer feedback. The likelihood of a visitor being in the top group are greatly impacted by parameters like the host's_experience_in_months, according to a study of the est_estimate values. With p-values under 0.05, a number of other variables were also shown to be significant in statistics. Review_communication_score, review _rating_score, and response_time in hours become important variables affecting customer impressions.

The implications of this complex consumer sentiment analysis provide hosts and those who use Airbnb with useful information. Hosting practises may be strategically improved by placing an emphasis on quick responses, straightforward communication, and upholding strict standards for sanitation. Hosts may greatly raise visitor happiness, get rave ratings, and draw in more customers by matching their practises to these client expectations. Understanding the key elements influencing customer satisfaction is crucial for Airbnb in order to improve platform standards that encourage hosts to give priority to crucial elements like interactions, sanitation, and appropriateness of accommodations. Enhancing these features can help guests have enhanced overall experiences, encourage guest loyalty, and boost Airbnb's standing as a dependable business.

In summary, our in-depth investigation on Airbnb in the Metropolis of New York City has shed light on the company's symbiotic relationship with the views of customers. The platform's managers and hosts may use these data as a springboard for strategic decision-making and tactical tweaks that will improve visitor satisfaction and help them succeed in the competitive temporary rental market.

4.4 Recommendation and Future Works

Future research and evaluation may be built on a strong basis provided by the Metropolis of New York. Here are a few probable directions for future prospects:

Analysis according to the change on customer behaviours : To track the changes in attitudes and socio_economic patterns through period of time, conduct periodic research. In order to spot emerging trends, patterns, and movements in customer habits and the dynamics of markets, this may necessitate analysing data extending a number of years.

Expanding the geographics: In order to assess how attitude and socio_economic characteristics differ in other areas, expand the investigation to additional cities or regions. Researching geographical variations might give hosts and the Airbnb platform useful knowledge to customise their approach.

Extenuating circumstance impacts : Investigating the effects of important occurrences on the socio_economic environment and the attitudes of Airbnb clients. Examples include slumps in the economy, severe healthcare incidents, and modifications to legislation. It is essential to comprehend how outside influences affect the short-term rental industry in order to modify plans and regulations.

Implementing techniques such as machine learning and forecasting: To forecast trends in sentiment and socioeconomic aspects, use machine learning models. In order to make proactive modifications, advanced modelling techniques which are statistical can assist hosts and the owners of Airbnb in foreseeing changes in client tastes and market conditions.

Analysis of government policies and its impacts: Examine how governmental laws, such as those governing rentals for brief periods, affect user perception and the dispersion of Airbnb properties. The host company and their services can benefit from knowing how policies affect the market as they overcome regulatory barriers.

Analysis of the segmentation of the users: Creating user groups based on various factors (such as geographies, different demographics and travel intentions) and examine how each segment views the property_listings, costs, and support services provided by Airbnb. The Airbnb journey could possibly be improved by adjusting techniques based on various user categories.

To comprehend the host and guest encounters as they actually occur inside the Airbnb environment, do an ethnographic examination. This subjective methodology may offer insightful information on user objectives, behaviour, and challenges. We can improve user experiences, make more educated choices, and create a more sustainable temporary housing market by exploring these future research paths. By doing so, we will get a deeper knowledge of the functioning of Airbnb as a whole.

References

repossessed from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3136138, (2020)

Adamiak, C. (2019, october 2). Retrieved from https://www.tandf_online.com/doi/full/10.1080/13683500.2019.1696758?scroll_=top&needAccess=true&role=tab

Back, J. (2019). *An investigation examines the uneven effects of Airbnb's service quality factors on customer satisfaction. publication for the reason of hospitality in journal and international*

repossessed from https://_/10.1016_/j_ijhm.2018.

Bevins, J. (2019, March 26). repossessed from https://www_epi_org/publication/ the pros and costs of using AirBnB economically There is no justification for local decision-makers to let Airbnb to evade taxes or controls.

Cervera, D. d. (2022, November 16). Repossesed from https://www_frontiersin_org/articles/10.3389/fpsyg.2022.994910/fu_ll

Chattopadhyay. (2019). *The goal of the study was to determine if various Airbnb listing criteria consistently affect price. In the International Journal of Hospitality Management, the study was published.* Repossesed from https://doi_org/10.1016/j_ijhm_2019.03.008.

Chaudhary, A. (2021, October 19). Repossesed from https://deliverypdf_ssrn_com/delivery_php?ID=600090118106119105070094071101122069101037019087067074070003086094099126010083072124050026032061006034030081073121097079068112029029009051050013073015127020018079022046085012008118027120095112096078079015115091

Cui, Y. (2018, MAR 10). Repossesed from https://papers_ssrn_com/sol3/papers_cfm?abstract_id=3136138

cui, Y. (2020, 7 7). Repossesed from https://papers_ssrn_com/sol3/papers.cfm?abstract_id=3136138

cuquet m, f. a. (2017). Retrieved from https://doi_org/10.1109/GIOTS_2017_8016274.

DDDA, N. (2019). Retrieved from https://doi_org/10.1109/ICIMTech_2019_8843779.

Dudas, G. (2017, october). Repossesed from https://www.research_gate_net/publication/320443736.A.socio-economic_.analysis_of_metropolitan_of _New_York_City_for_airbnb

Falk. (2019). *studied how different attributes affected the cost of postings on Airbnb in both urban and rural areas. Their research, which highlighted the importance of particular characteristics in influencing the costs of these listings, was published in the Interna*. Repossesed from https://doi_org/10_1016/j_ijhm_2019

Fensel, B. (2017). Retrieved from https://doi.org/10.1109/GIOTS.2017.8016274.

Fuentes, M. (2017). *Fernandez, C. Mateu, E. Martin-Fuentes, and E. Mariine_Roiig_/ Modelling an evaluation system for peer_from_peer housing: Airbnb stars.* Retrieved from https://doi_org/10_1016/j_ijhm_2017_10_016

Guttentag, D. (2019, June). Retrieved from https://www_ressearchgate_net/publicaton/333880155_review_litratre_of_upcoming_of_air-bnb

Haywood, J. (2016, july). Repossesed from https://str.com/sites/default/files/2019-07/Airbnb-and-Hotel-Performance.pdf

Hidayaanto an, w. a. (2018). Retrieved from https://doi_org/10_1109/EECSI_2018_8752912.

Kirkos, E. (2021, 12 10). Retrieved from https://_ejtr_vumk_eu/indx_php/about/artcle/download/2142/518

L., t. (2019, 07).

Larpin, S. (2019). *examining how many factors affect the cost of Airbnb rentals in both urban and rural areas.* Repossesed from 10.1016/j_ijhm_2019.04.023.

Lau, F. (2016). Spending fewer bucks to discover more: Investigating guest interaction in the Airbnb environment. Repossesed from https://doi_org/10_1016/j_ijhm_2019.04.007

Line, H. (2014). *Airbnb guests' and hosts' perspectives on peer-to-peer interactions. Published in Volume 77, Pages 405–414 published in the global journal of the Management of care industry.* Repossesed from https://doi_org/10_1016/j_ijhm.2018_08_004.

Madhi, H. A. (2021). Retrieved from https://pdfs.semanticscholar.org/ada4/60675a5989f997576514bec57ea90b39644f.pdf?_gl=1*1tbqyby*_ga*MTU5NTY3NTg0OC4xNjg5MzQxNzE0*_ga_H7P4ZT52H5*MTY4OTM0MTcxMy4xLjAuMTY4OTM0MTg1MS4yMi4wLjA.

Nguyen, Q. (2014, 1 12). Retrieved from Several significant elements that could be overlooked from a research standpoint include:

osborrne, J. (2020). Repossesed from https://doi_org/10_4135/9781071802724_n6.

Sari R, L. Y. (2019). Repossesed from https://doi_org/10_1109/ICIIMTech_2019.8843746.

Teuber, H. (2016). Peer from peer leasing operations in the cooperative society *facilitators and inhibitors. Annual Proceedings of the Hawaii World Conference on Network Technologies.* Repossesed from https://doi_org/10_1109/HICSS_2016_593.

Appendix

Data analysis code for analysis of social and economics.

In [ ]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import folium
```

In [ ]:

```
!pip install folium
```

In [ ]

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
%matplotlib inline
```

- For data manipulation we have used pandas.
- for data visualization we have used seaborn and matplotlib.
- for dealing with geographical data we have used folium.

In [ ]:

```
newyrk = pd.read_csv('C:/Users/assel/OneDrive/Desktop/airbnb project/listings.cs
```

In [ ]:

```
newyrk.head()
```

Out[ ]:

	id	property_name	id_host	name_host	group_neighbourhood	neighbourhood
0	5121.0	Rental unit in Brooklyn · ★4.52 · 1 bedroom · ...	7356	Garon	Brooklyn	Bedford-Stuy
1	2595.0	Rental unit in New York · ★4.68 · Studio · 1 b...	2845	Jennifer	Manhattan	Mi
2	14991.0	Rental unit in New York · ★4.93 · 1 bedroom · ...	59023	Bianca	Manhattan	Lower Ea
3	5136.0	Rental unit in Brooklyn · ★5.0 · 2 bedrooms · ...	7378	Rebecca	Brooklyn	Suns
4	59709.0	Rental unit in New York · ★4.77 · 2 bedrooms · ...	186084	Ricardo & Ashlie	Manhattan	Chin

In [ ]:

```
newyrk.info()
```

```
Out[ ]:  last_review                          24.09
         review_each_month                    24.09
         name_host                             0.01
         id                                    0.00
         property_name                         0.00
         id_host                               0.00
         group_neighbourhood                   0.00
         neighbourhood_name                    0.00
         lati                                  0.00
         longi                                 0.00
         type_room                             0.00
         price                                 0.00
         min_nights                            0.00
         no_of_reviews                         0.00
         calculate_host_listing_count          0.00
         availability_yearly                   0.00
         dtype: float6
```

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 43566 entries, 0 to 43565 Data

columns (total 16 columns):

\# Column Non-Null Count Dtype

<u>Location exploration code :</u>

```
# Estimating the regional average cost
region_price = round(newyrk_clean.groupby('group_neighbourhood').price.mean().so

# Presenting the outcomes
print(region_price) print('\n')

# Visualising the regional mean cost

plt.figure(figsize=(6, 4))

colors = ["blue", "red", "grey", "green", "purple"]

rt = sns.barplot(x=region_price.index, y=region_price.values, palette=sns.xkcd_p rt.set_title('The mean cost for different regions')

rt.set_ylabel('($) mean cost') rt.set_xlabel('the area')
```

Home as percentage in each location code:

```
# Estimating the number of rooms in each area

region_no_rooms = newyrk_clean['group_neighbourhood'].value_counts()


# Estimating the proportion of rooms in each area

region_no_rooms_pct = round(newyrk_clean['group_neighbourhood'].value_counts(nor


# Presenting the outcomes
print(region_no_rooms) print('\n')
print(region_no_rooms_pct) print('\n')


# Visualising the number of rooms by area.

plt.figure(figsize=(6, 4))

bar = sns.barplot(x=region_no_rooms_pct.index, y=region_no_rooms_pct.values, pal bar.set_title('Homes as a
Percentage in Each Location') bar.set_ylabel('percentage')

bar.set_xlabel('locations')
```