EDA Capstone project - 1

Airbnb Bookings Analysis

Presented by

Ansh Srivastava


airbnb

Contents

Abstract

This study examined the relationship between various parameters of the AIRBNB dataset such as host id, hostname, neighbourhood group, neighbourhood, room type, price number of reviews, availability. An exploratory data analysis using field data points collected from the Airbnb listings in the metropolitan area of New York city reveals intriguing findings. The analysis helps us in understanding the most preferred hosts and neighbourhood groups by guests, the density of properties across the various neighbourhood, the number of room types belonging to each neighbourhood group, expensive neighbourhood groups, busiest hosts, preference of room types by guests, price of various room types. This analysis helps draw insights from the data and can be utilised for security, business decisions, understanding of customers and providers, behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

Introduction

What is Airbnb?Airbnb, as in “Air Bed and Breakfast” is a service that lets property owners rent out there spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world.

Today, Airbnb become one of a kind service that is used and recognized by the whole world.

Airbnb is based on a peer to peer business model. This makes it simple, easy to use, and tends to be more profitable for both parties. The model also gives you the opportunity to customize and personalize your guest’s experience the way you want

Problem Statement

For this project we have analyze Airbnb's New York City(NYC) data of 2019. NYC is not only the most famous city in the world but also top global destination for visitors drawn to its museums, entertainments, restaurants and commerce.

Our main objective to find out the key metrics that influence the listing of properties on the platform. For this , we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis(EDA) techniques.

Data analysis on thousands of listings provide through Airbnb is a crucial factor for the company.

We will be finding out the distribution of every Airbnb listing based on their location in NYC, including their price range, room type, listing name, and other related factors.

Dataset information:

Id- These gives us the listing id.

Name- Listing name

Host_id- host id

Host_name- host name

Neighbourhood-NYC neighbourhood

Neighbourhood_group-NYC brough

Latitude-listing latitude

Longitude- listing longitude

Room_type- type of rooms (Entire home/apt, Private room , Shared room)

Price- listing price

Minimum_nights- minimum nights people stay

Number_of_reviews- Total number of reviews

Last_review- Date of last reviews

Reviews_per_month- Average number of listings for this host

Calculated_host_list_count- Total number of listings for this host

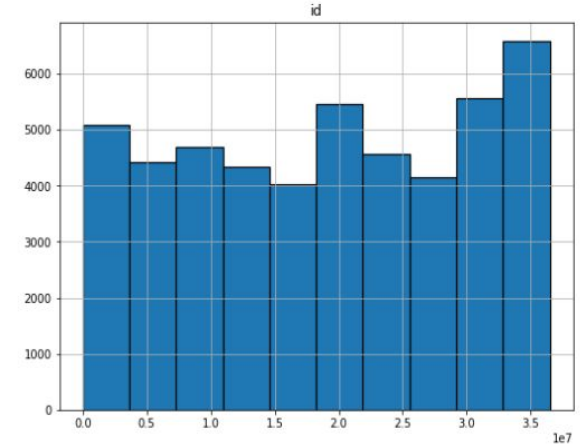
Availability_365- Number of days listing is available

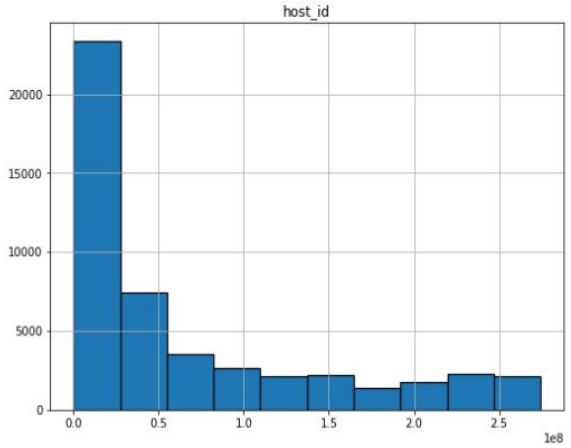
Number of Data set were nearly 49000

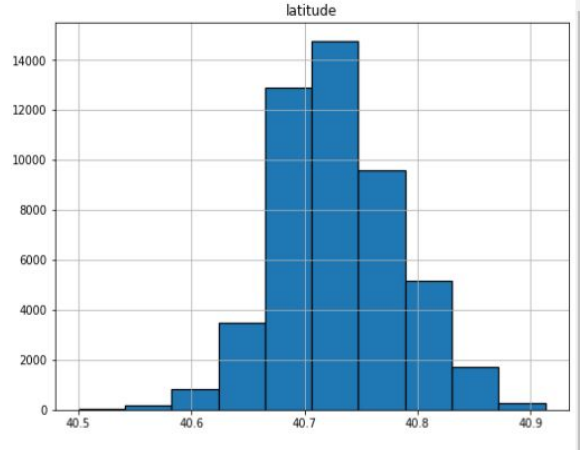
Agenda

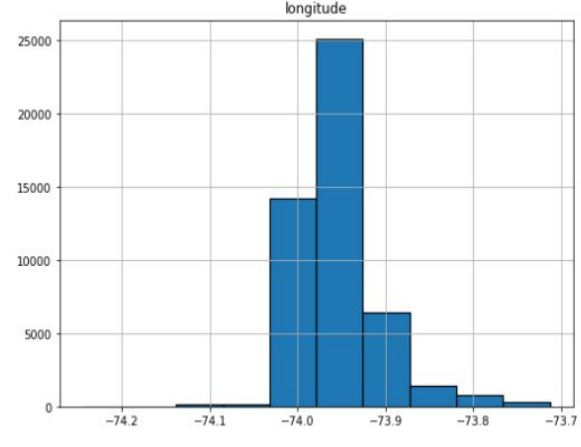
- Visualize the distribution for every feature
- Taking necessary columns only
- Most preferred room type
- Find the total count of each room type
- Find top 10 busiest hosts
- Find top hosts with most reviews
- Find minimum number of night stay
- Count of neighbourhood group
- Room type location per neighbourhood group
- Areas number of reviews
- Checking correlation for numerical data

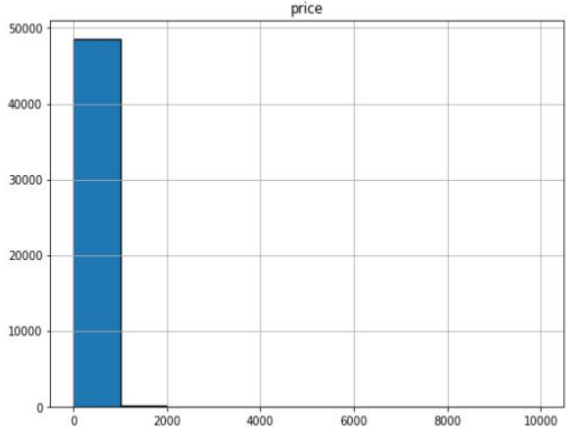
Visualize the distribution for every feature

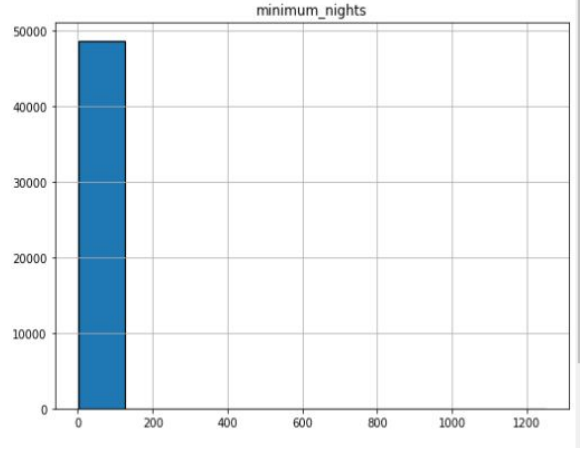


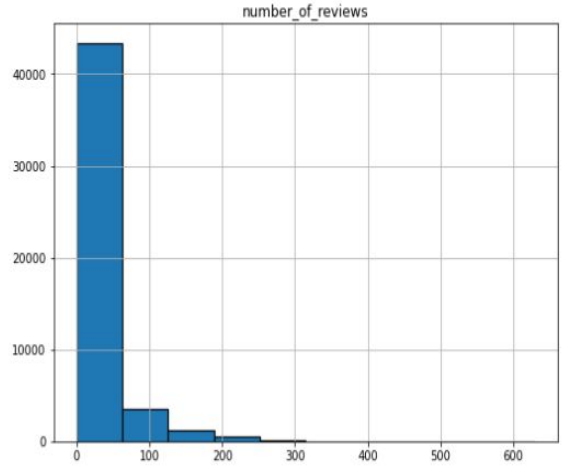


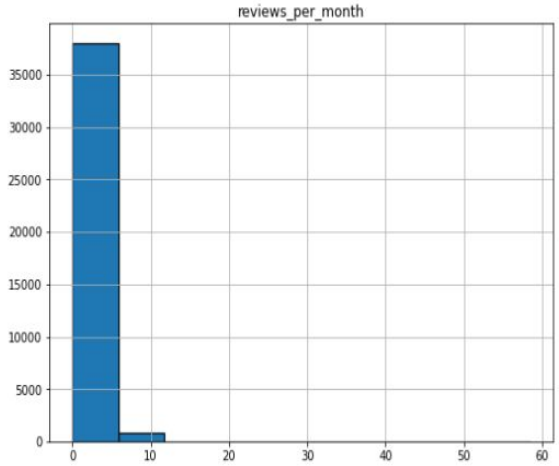


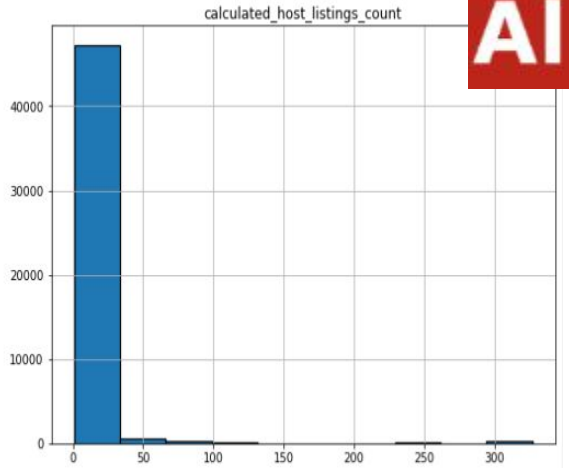


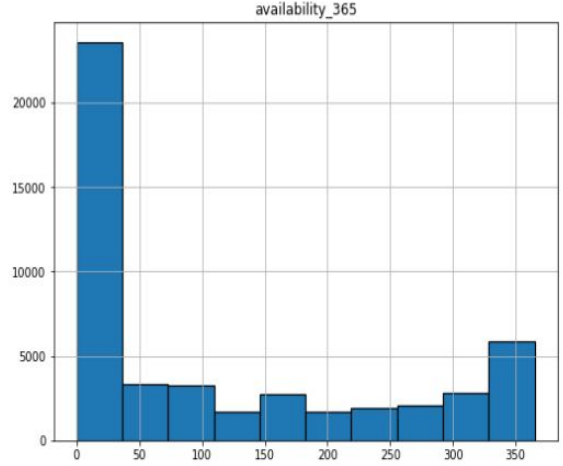





number_of_reviews
40000
30000
20000
10000
0
0
100
200
300
400
500
600

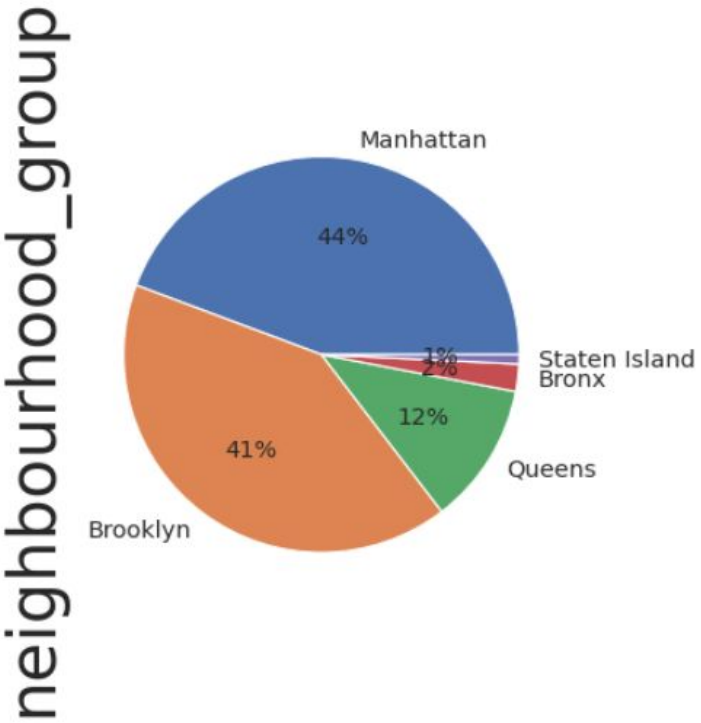

reviews_per_month
35000
30000
25000
20000
15000
10000
5000
0
0
10
20
30
40
50
60

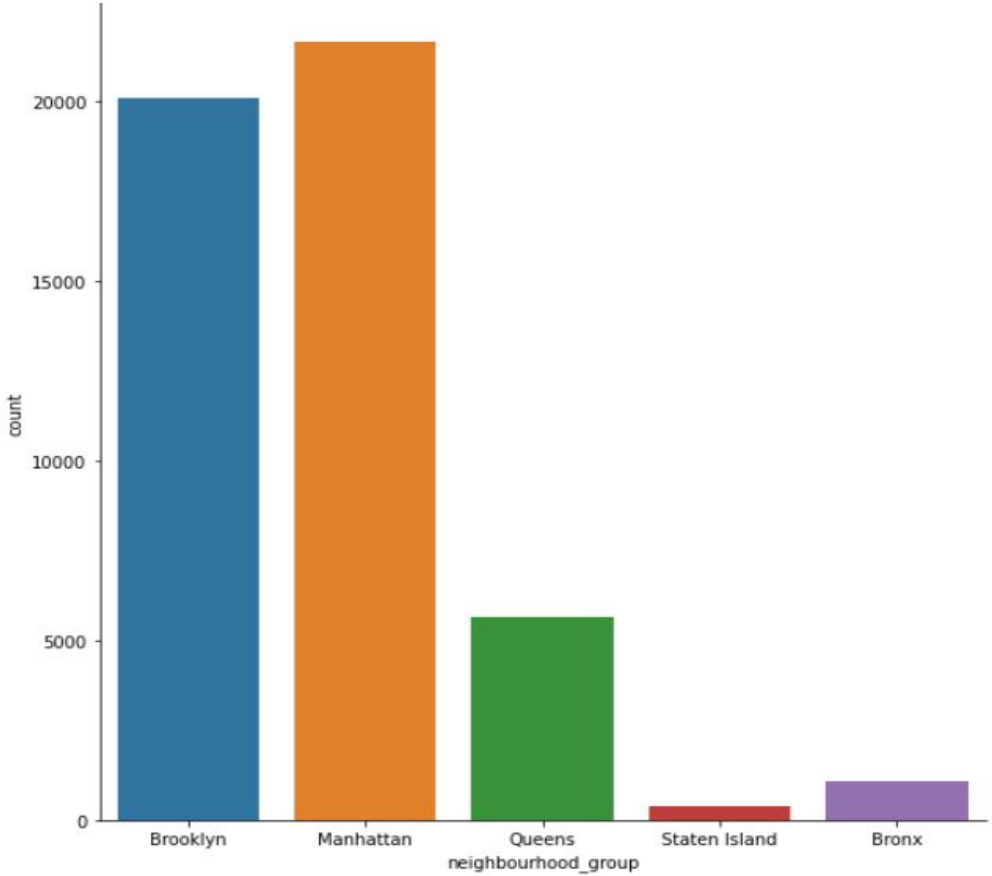

calculated_host_listings_count
40000
30000
20000
10000
0
0
50
100
150
200
250
300


availability_365
20000
15000
10000
5000
0
0
50
100
150
200
250
300
350

Neighbourhood groups

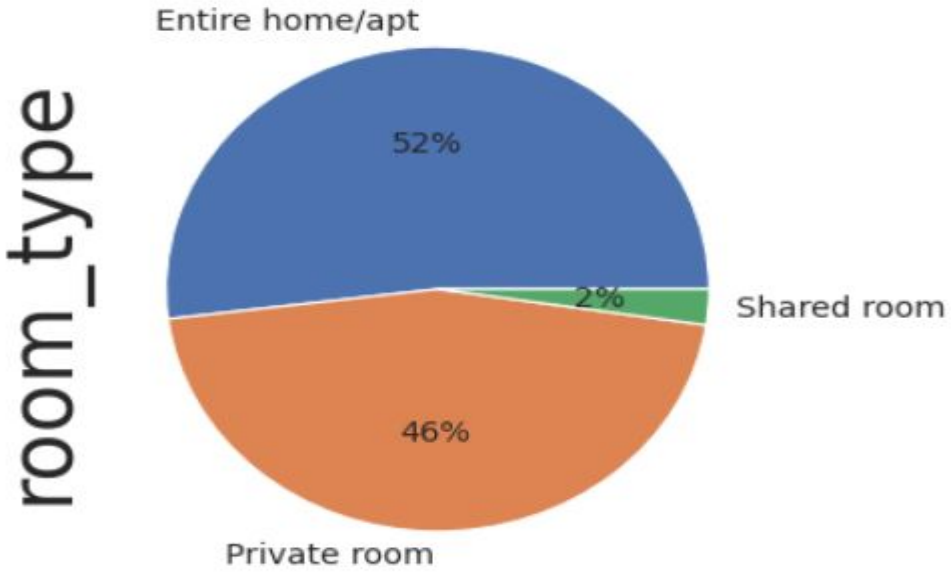
- Manhattan has the highest number of listing of about 44% followed by brooklyn of 41%.
- Staten Island stands the list number of listing 1%.



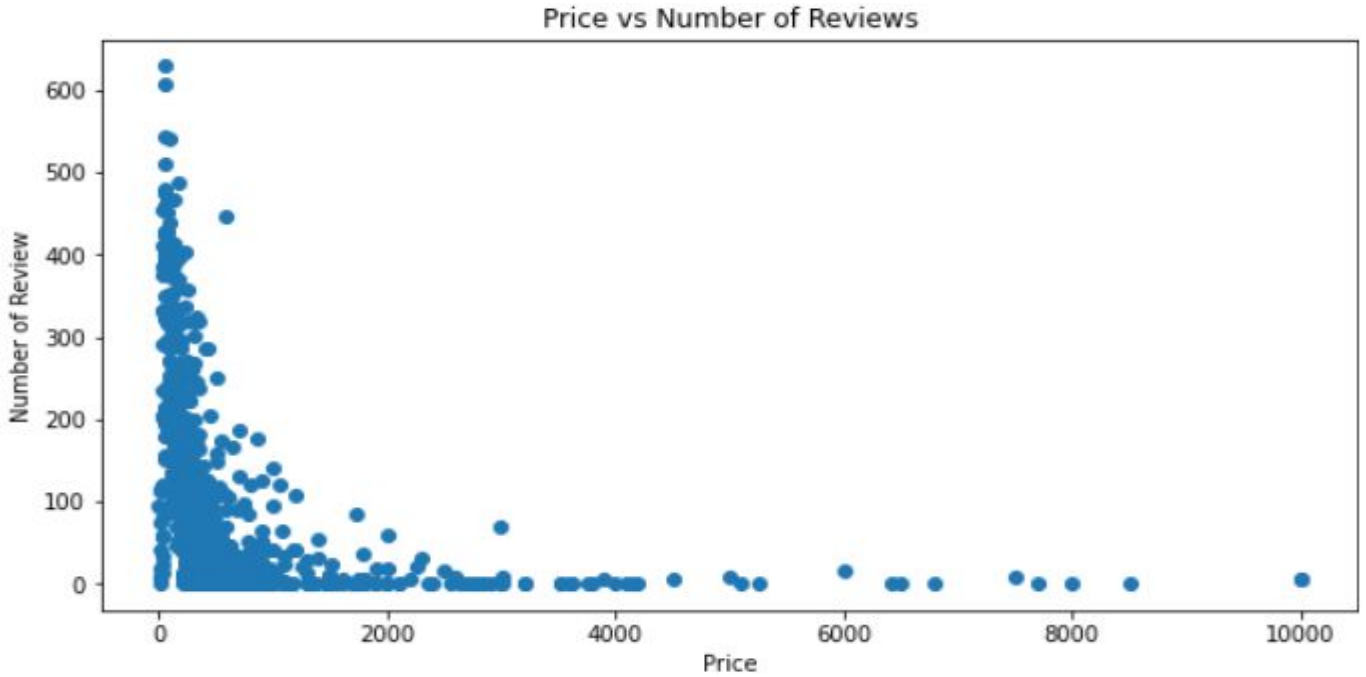


Various room type

- The entire home/apt room are the highest in NYC followed by private rooms.
- Shared room holds are only 2% in AIrbnb NYC.



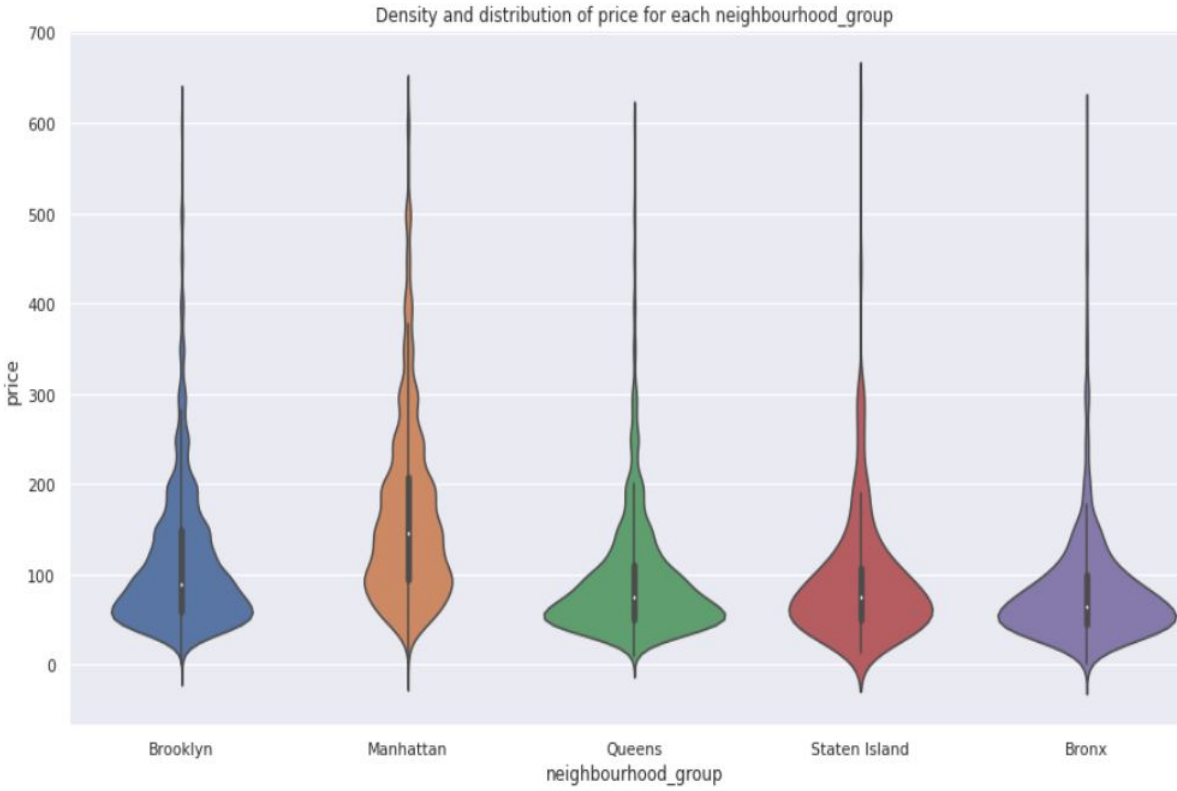
Price To review Analysis



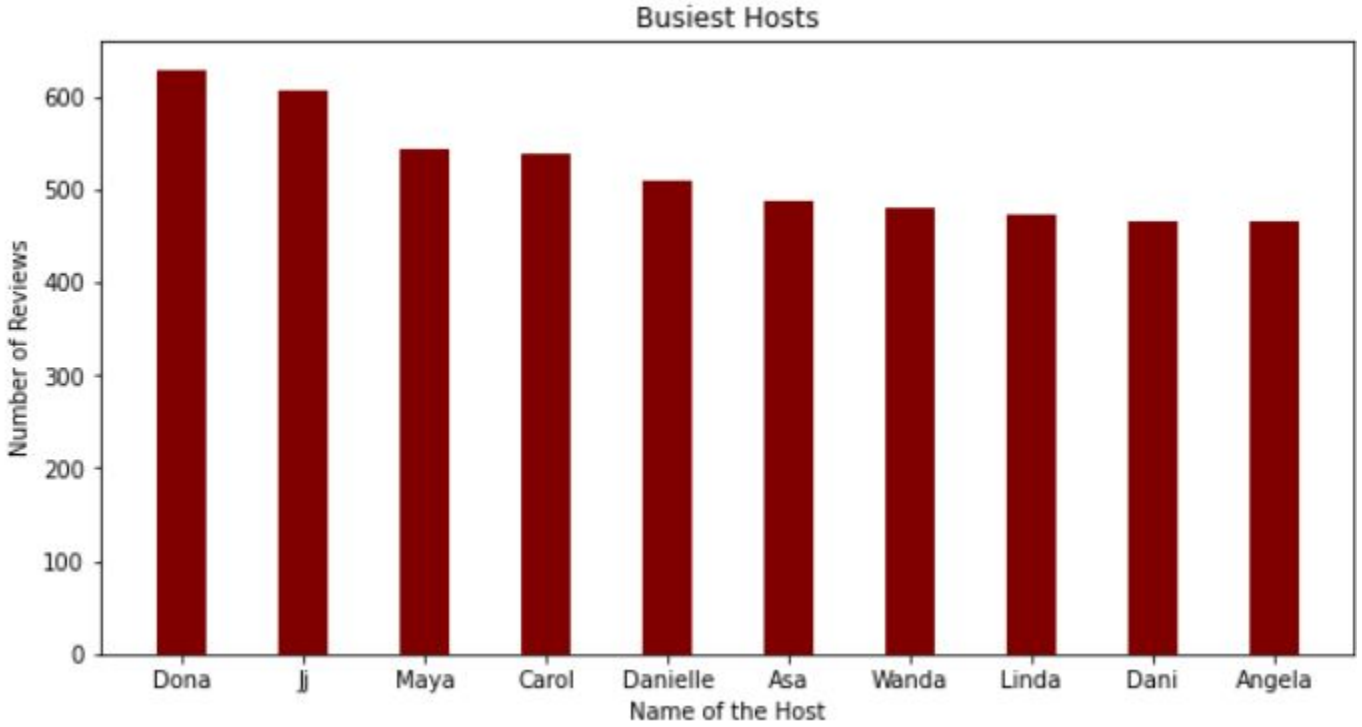
Price on X axis and Reviews on Y axis.
It say us that people like more to stay where price is less.

Analyzing the price across neighbourhoods

- As usual Manhattan being the costliest place to live in NYC, having average price more than 145 USD followed by Brooklyn with around 90 USD on an average for the listing.
- Queens and Staten Island has nearly the same of 75 USD.
- The highest price range could go just above 360

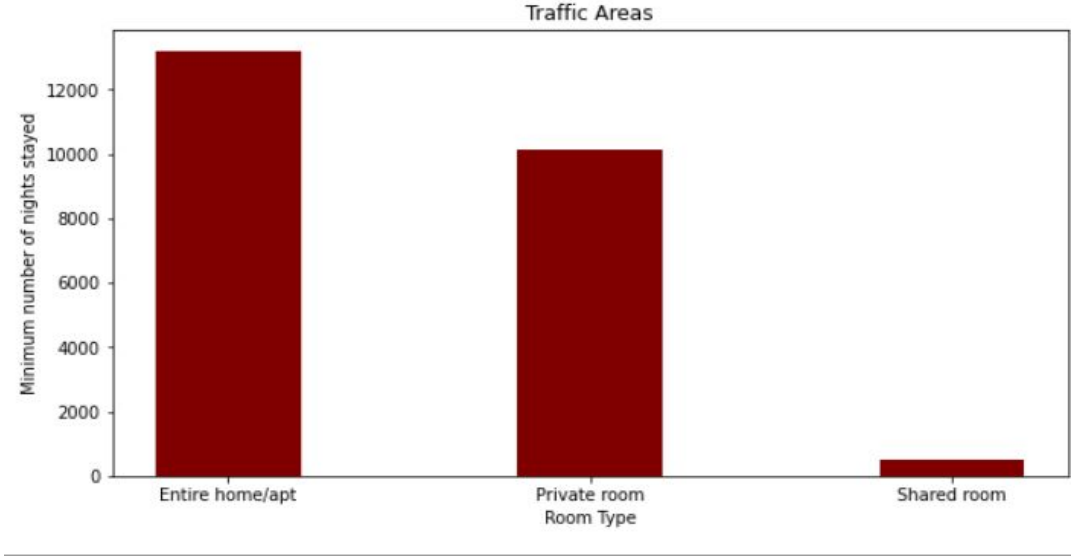


Top 10 busiest hosts with most review



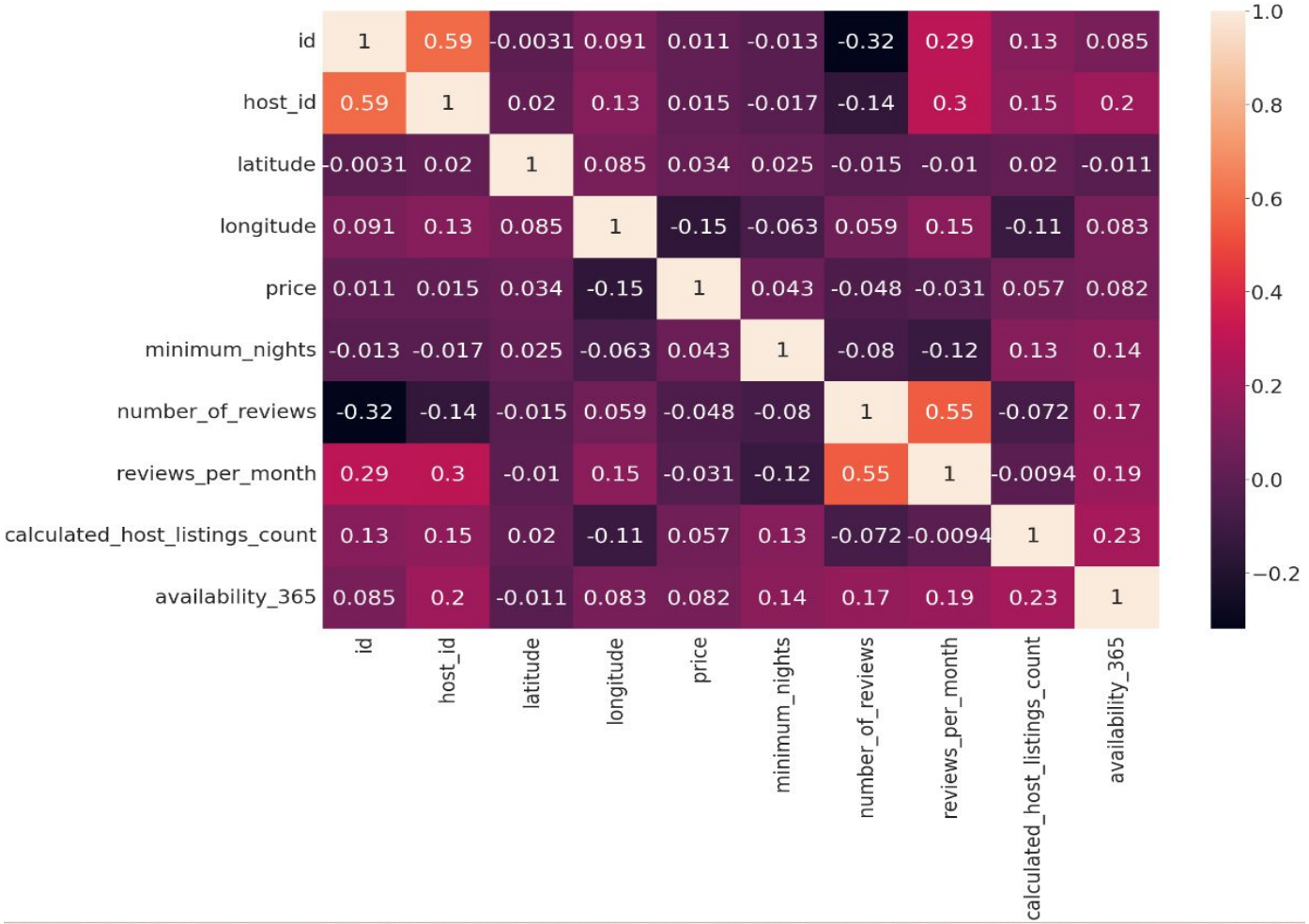
- Dona, Ji, Maya, Carol, Danielle has the maximum Number of reviews.
- **Because these hosts listed room type as Entire room** and Private room which is preferred by most number of people.

Minimum number of night stayed in traffic areas



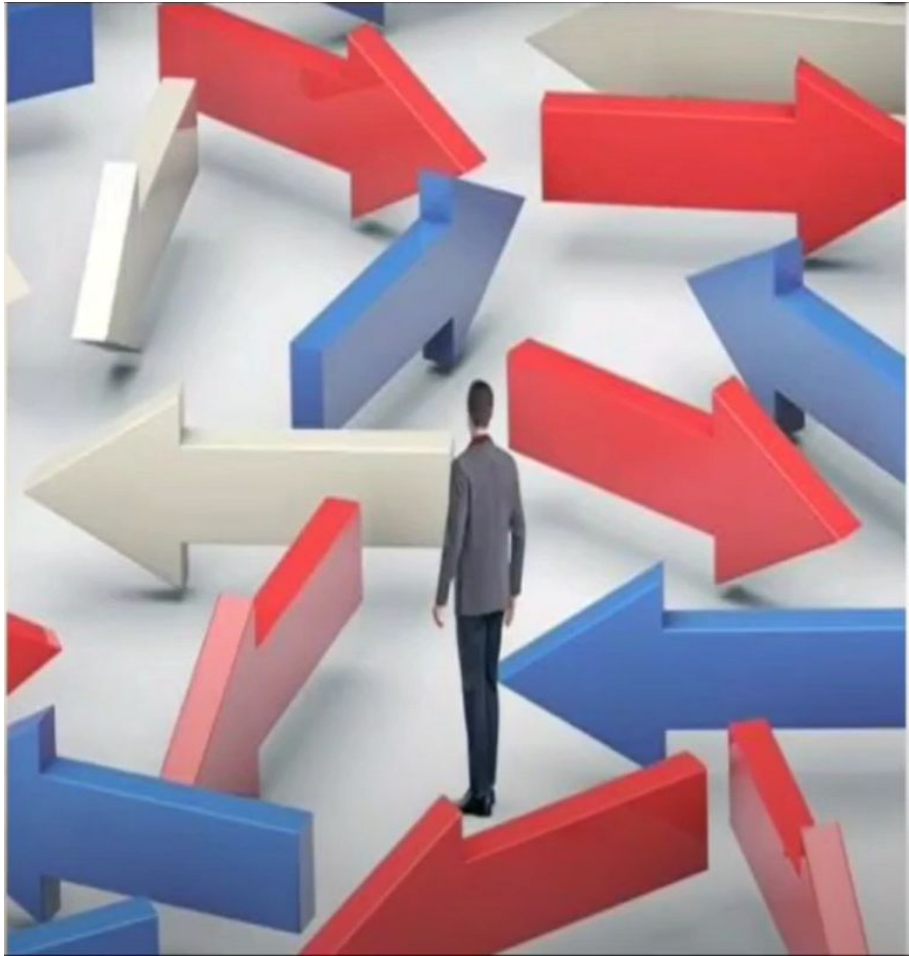
- From the Above Analysis We can Say that People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.

Correlation Matrix



Challenges faced

- Reading the dataset and understanding of columns.
- For answering some of the questions we had to understand the business model of Airbnb how they work.
- Handling NAN values, null values and duplicates.
- Designing multiple visualizations to summarize the Information in the dataset and successfully communicate the results and trends to the reader.
- Removing the outliers for some data set. Finding and sorting few impossible dataset.



Analysis summary

- **Manhattan is the most focused place in New york for hosts to do their business.**
- As people loved to stay when in need for maximum number of nights in Entire Apartment and Brooklyn stands second highest focused place by people. Also, as the average cost in Brooklyn is 90 USD which is less than Manhattan Airbnb can increase number of entire apartment in Brooklyn.

Conclusion

- Evident that Manhattan and Brooklyn are the neighbourhood groups with most number of Airbnb rooms.
- Room types listed mostly is either Entire home or Private rooms.
- Reviews across Manhattan and Brooklyn further says they are most popular.
- Price ranges options based on nights can help plan his expenses for longer stay.
- Helpful for a new host who wants to enter into business with Airbnb.
- The people who prefer to stay in Private room they won't stay longer as compared to Home or apartment.
- If there are more number of reviews for particular Neighbourhood group that means that place is a tourist place.
- Most people prefer to pay less price.
- If people are not staying more then one night means they are travellers.


airbnb
airbnb
AI

Thank You!!