EDA Capstone Project 1

AIRBNB Bookings Analysis

Presented By-
Gyanendra kumar Porwal

Contents

1. Abstract
2. Introduction
3. Problem Statement
4. Dataset Information
5. Agenda
6. Challenges faced
7. Conclusion

Abstract

This study helps in examine the relationship between various parameters of the Airbnb NYC dataset such as host id ,neighbourhood group, neighbourhood, room type, price, number of reviews, availability etc. An exploratory data analysis using these data points collected from the Airbnb listings in the metropolitan area of New York city reveals various findings.This analysis helps us in finding the most preferred neighbourhood group by guest , most preferred room type, location on the basis of room available in various neighbourhood group and on the basis of room type, top unique neighbourhood, top unique host, variation of neighbourhood group with mean and median price, distribution of price with room type ,busiest host, availability etc. This analysis helps us to draw insights from the data that 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 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:

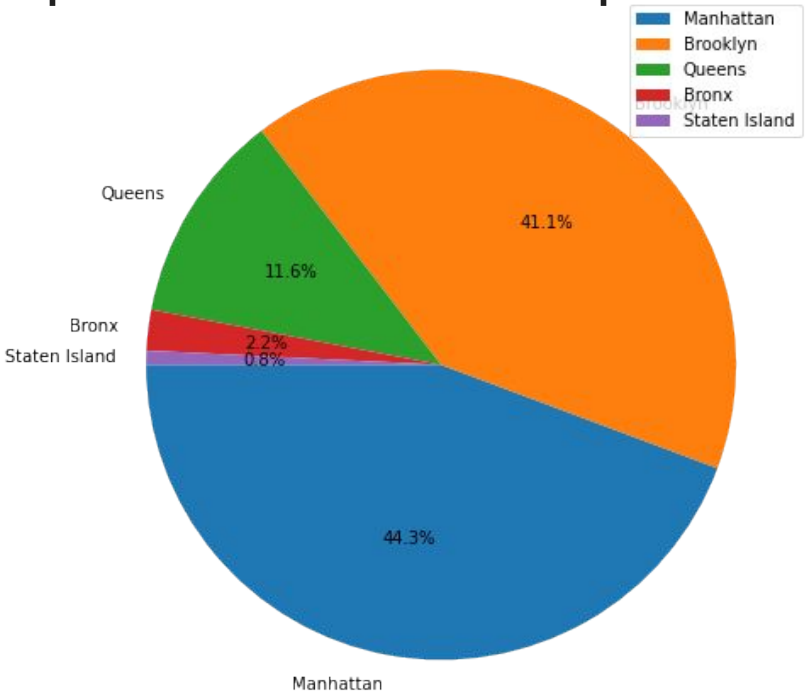
id: listing ID
name: name of the listing
host_id: host ID
host_name: name of the host
neighbourhood_group: location
neighbourhood: area
latitude: latitude coordinates
longitude: longitude coordinates
room_type: listing space type
price: price of room
minimum_nights: amount of nights minimum to stay
number_of_reviews: number of reviews
last_review: latest review
reviews_per_month: number of reviews per month
calculated_host_listings_count: amount of listing per host
avalability_365 : no of days rooms available in a 1 year(365days)

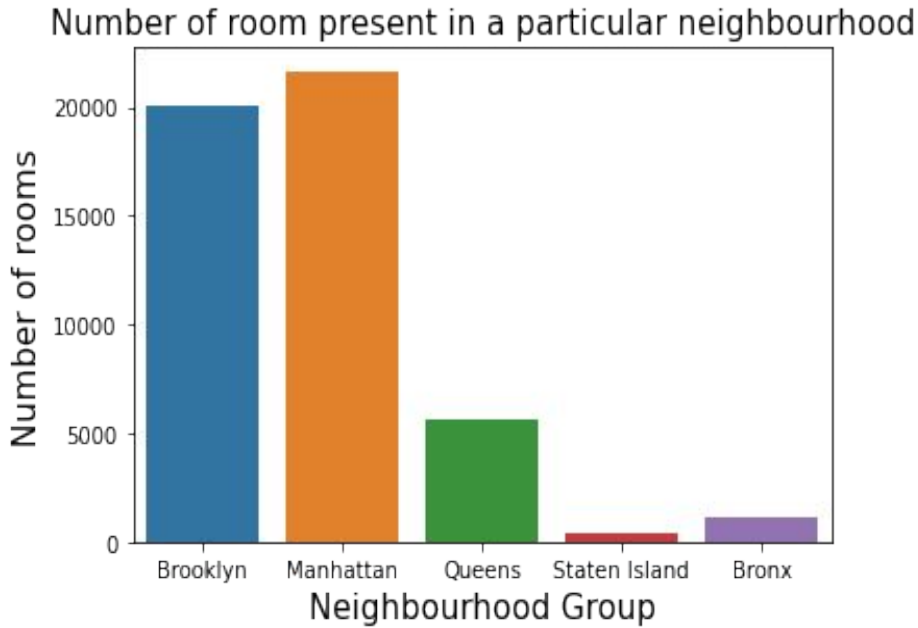
Agenda

1. Take the required column necessary for the analysis.
2. Checking correlation for numerical data.
3. Find out which neighbourhood_group has how much number of the room present in there respective region.
4. Find out the different types of room available to the person travelling and booking through airbnb site.
5. Find the different types of room available to the people in various neighbourhood group.
6. Plot on the basis on location(longitude , latitude) plot the room available in the given neighbourhood_group. Plot on the basis on location(longitude , latitude) plot various room types available.
7. Find out the unique neighbourhood and top 5 listing areas on the basis of neighbourhood also find out the unique host and top 5 listing areas on the basis of unique host_id .

8. Find the relationship of neighbourhood group with mean price & median price.
9. Find the variation of the price distribution with the different room type
10. Find out the Top 10 listing on the basis of the number of reviews.
11. Find out the top 10 neighbourhood & neighbourhood group with various room types according to the reviews per month.
12. Find the relationship between the price and availability of various room available in different neighbourhood?
13. Find the relationship between neighbourhood group and availability of room?

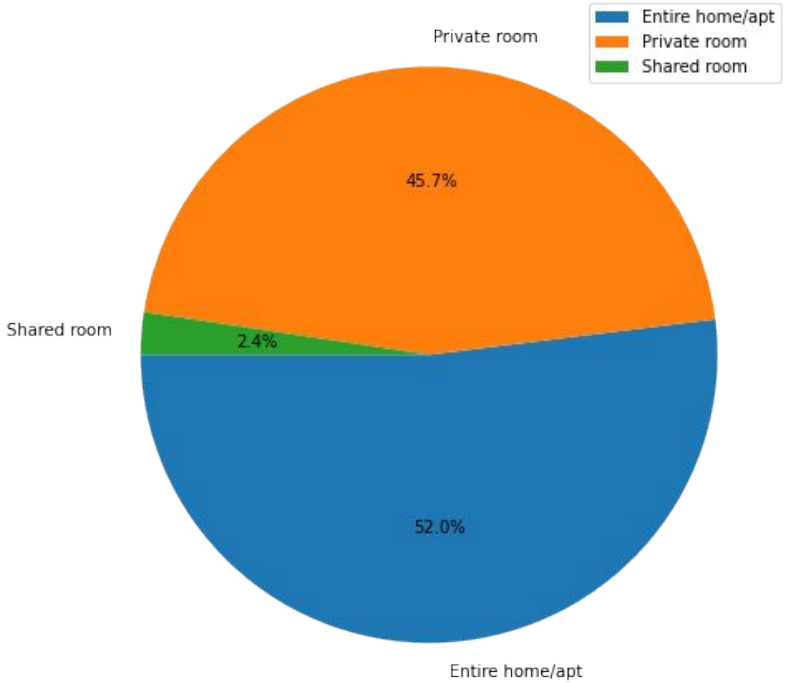
Neighbourhood_group has how much number of the room present in there respective region

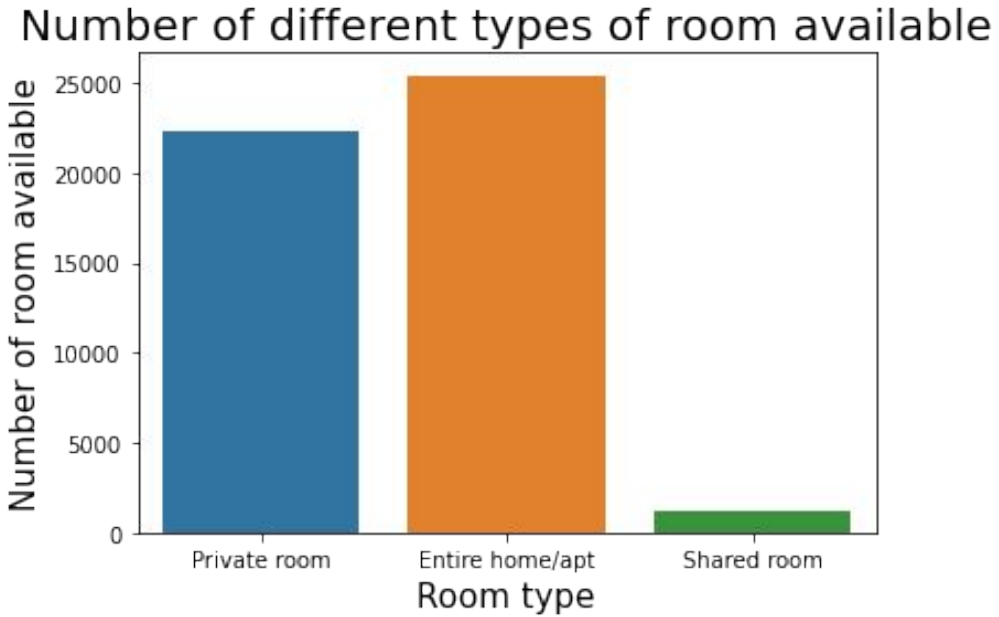




- We can see that that Manhatten **(44.3%)** is given the first position as it contains the most number of rooms followed by Brooklyn **(41.1%)** and Staten island has least rooms on airbnb site only **1.8%**.

Different types of room available to the guest booking through airbnb site

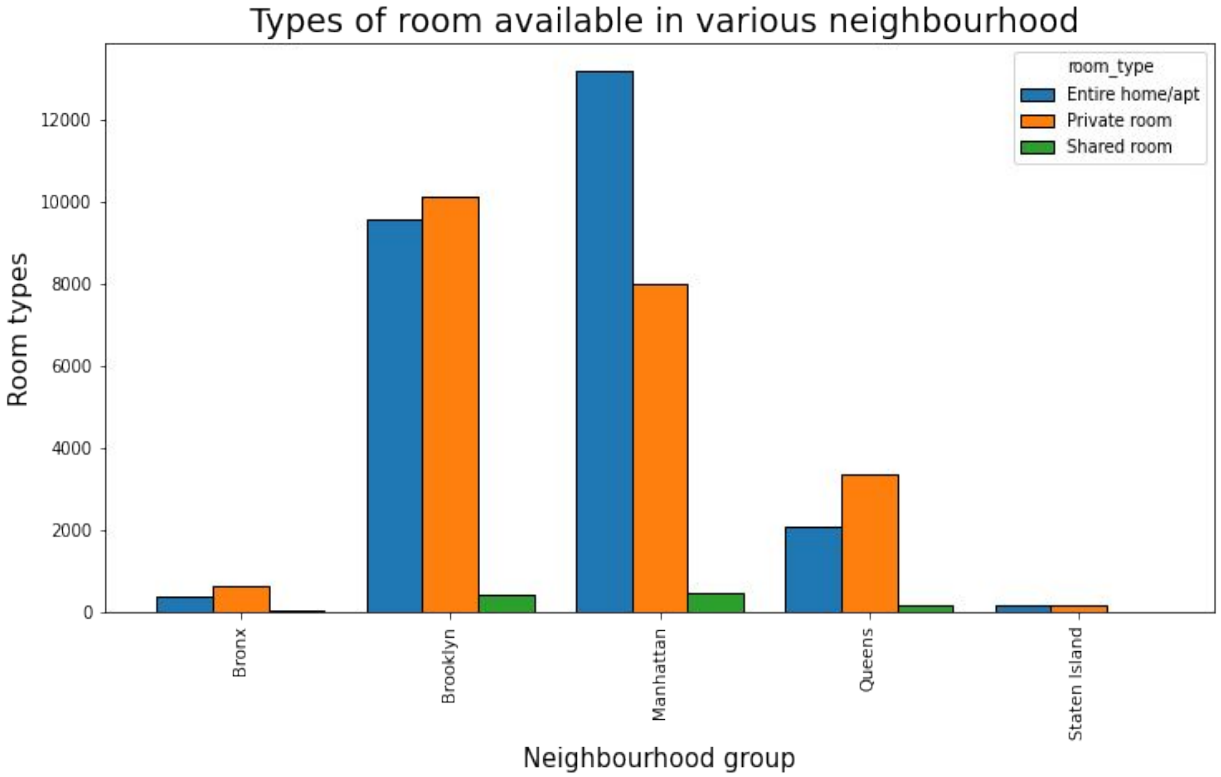




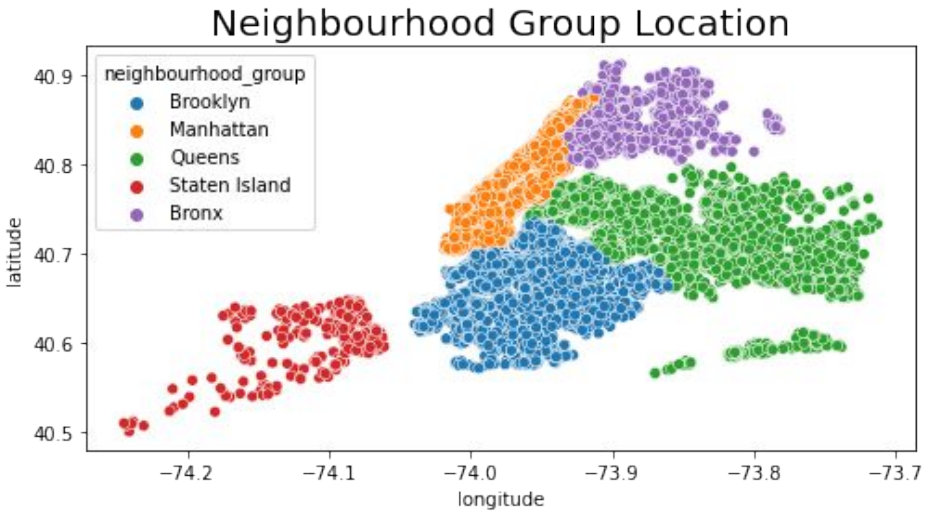
- The room type which is available highest is Entire home/apt**(52.8%),** private room**(45.7%)** & shared room**(2.4%)** is very low in comparison to the other two room type

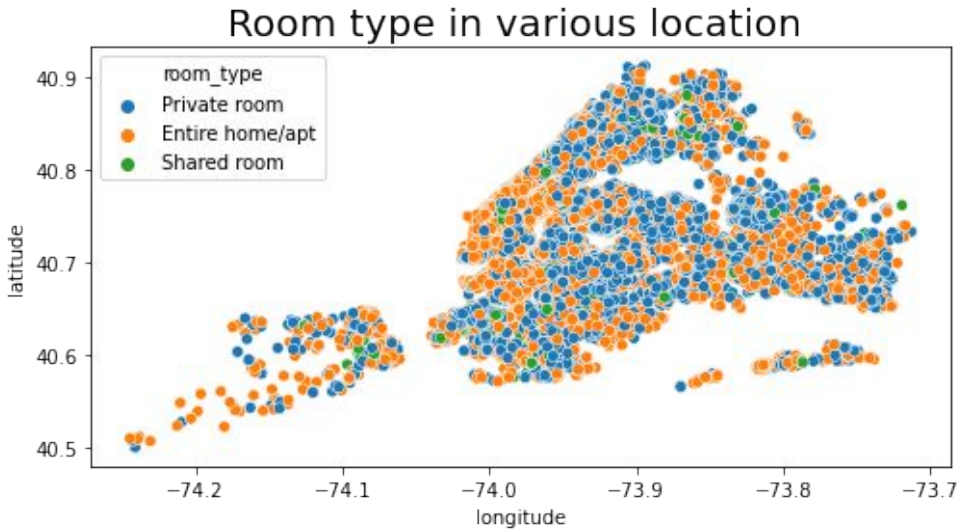
Different types of room available to the people in various neighbourhood group

- The Manhattan has the maximum no. of room type in the category of Entire home/apt & Brooklyn has highest no. in the category of Private room.



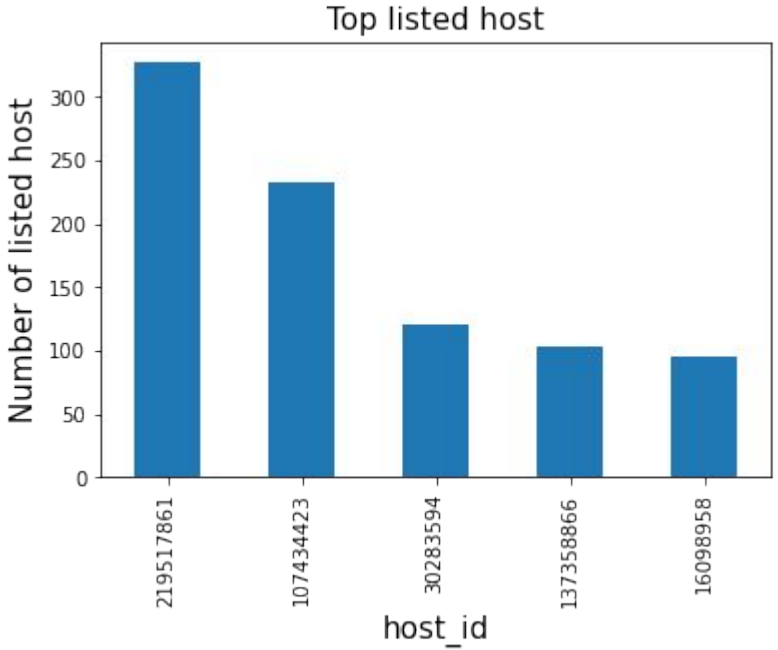
1. Plot on the basis on location(longitude,latitude) plot the room available in the given neighbourhood_group.
2. Plot on the basis on location(longitude,latitude) plot various room types available.

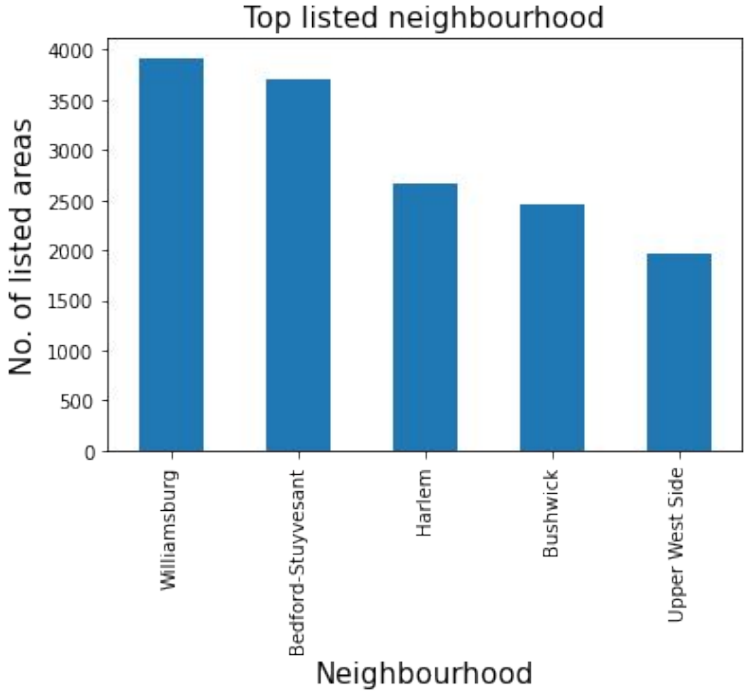




1. Brooklyn contains all rooms in closed packed space, manhattan in less width and long space, queens are wider and longer, bronx in smaller closed packed space than brooklyn and staten island have rooms available seperately in a unclosed or openly manner.
2. We can see the private room and entire home/apt largely in various location as they are present in larger amount in comparison to shared room.

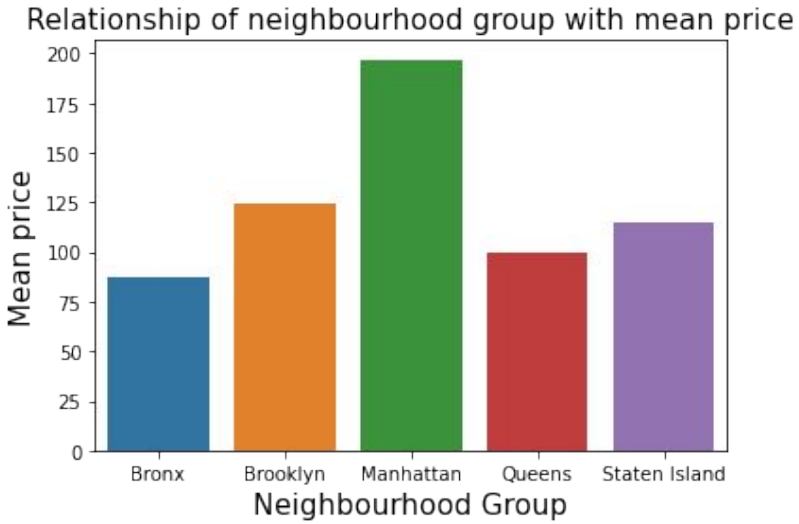
The unique neighbourhood and top 5 listing areas on the basis of neighbourhood also find out the unique host and top 5 listing areas on the basis of unique host_id .

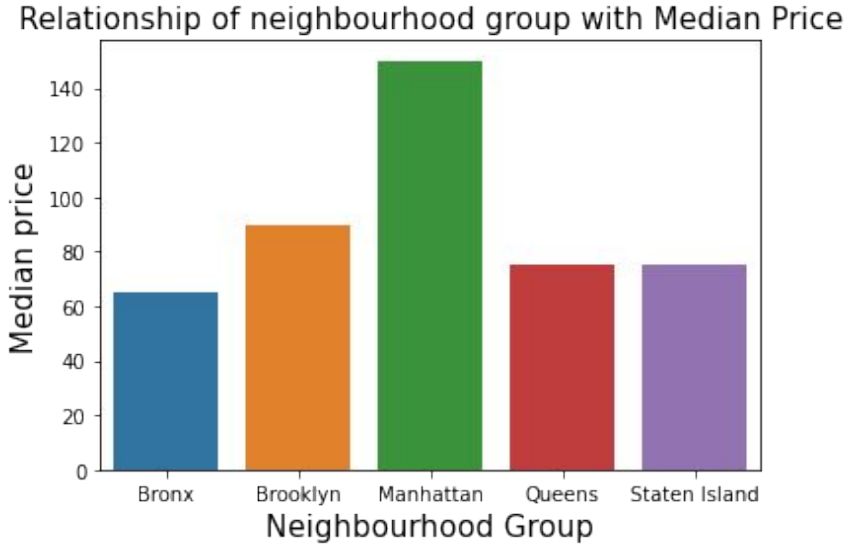




- The host with host_id 219517861 has highest listing of 327.
- Williamsburg neighbourhood has the most listings followed by bedford-stuyvesant

The relationship of neighbourhood group with mean price & median price.

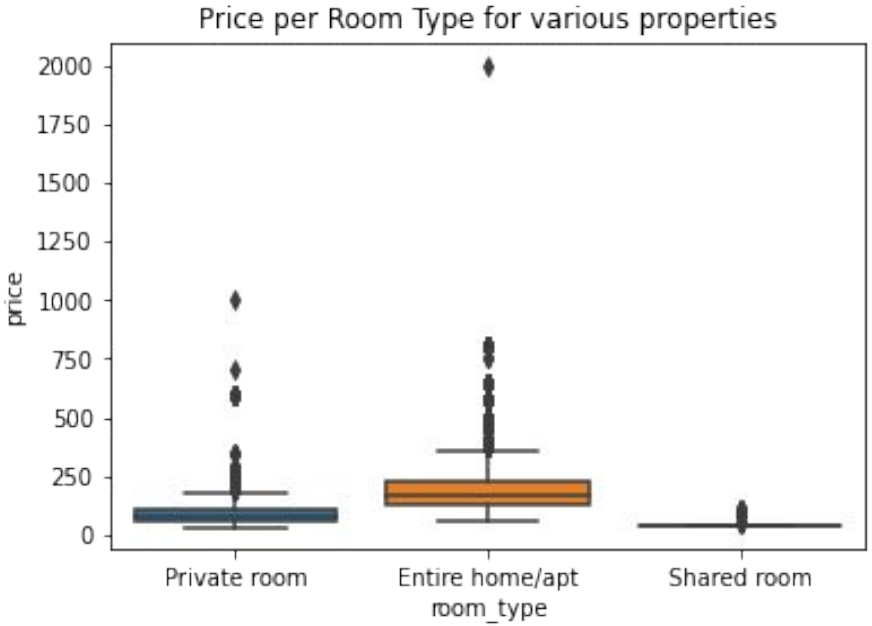


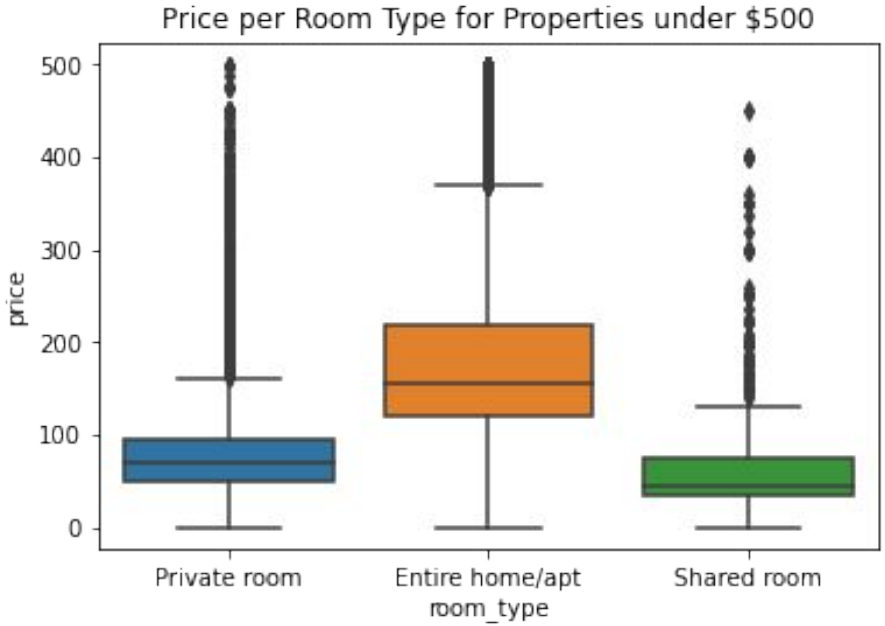


- We can conclude that according to the average price the manhattan is the most expensive place in the given airbnb dataset with $197 followed by brooklyn $124 then comes staten island $115, queens with $99 and at last is bronx with $87.

- We can see that from median price perspective the manhattan has the highest price $150 followed by brooklyn $90. Queens and staten island have equal median price $75 and bronx is the lowest with $65.

The price distribution with the different room type

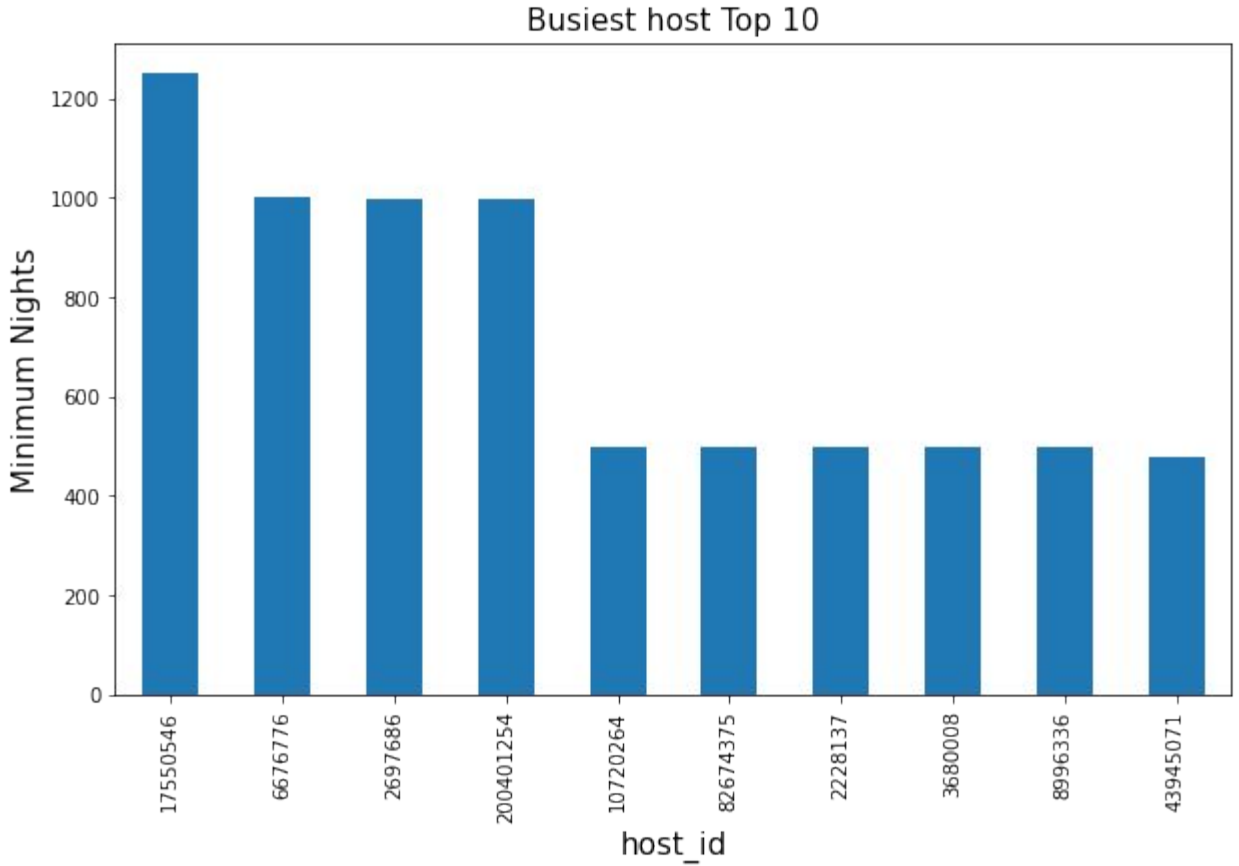




- The variation in the price of various room type can be seen and to clearly see the price range since they contains high price values we filter out the data with price under 500.
- The Entire home/apt have the largest range with median(50 percentile) value $160 ,private room on second with median value $70 and shared room with lowest range and median value $45.

The busiest host in the given data with reference to the host_id and minimum nights columns.

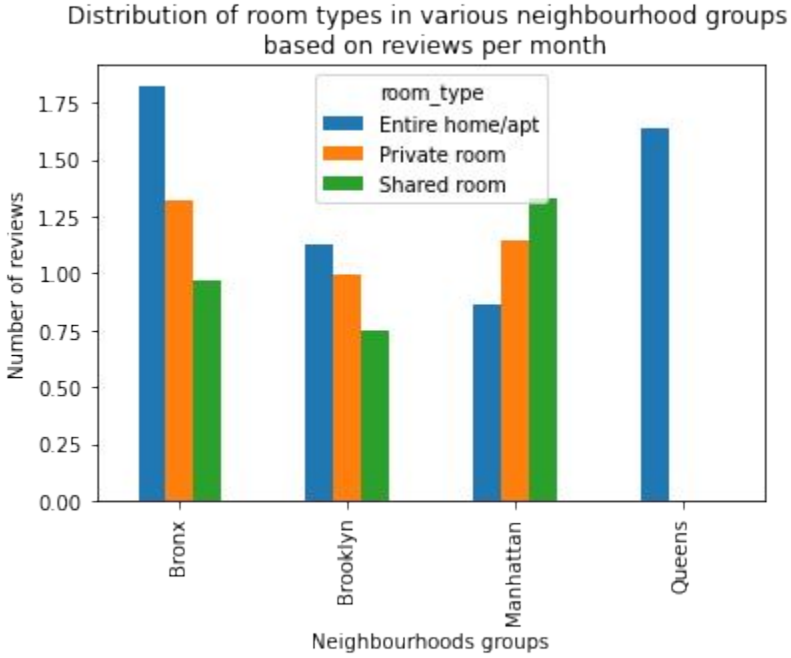
- We can say that host with the host_id 17550546 is the busiest host as the number of minimum nights spend is the highest at this particular host listing.

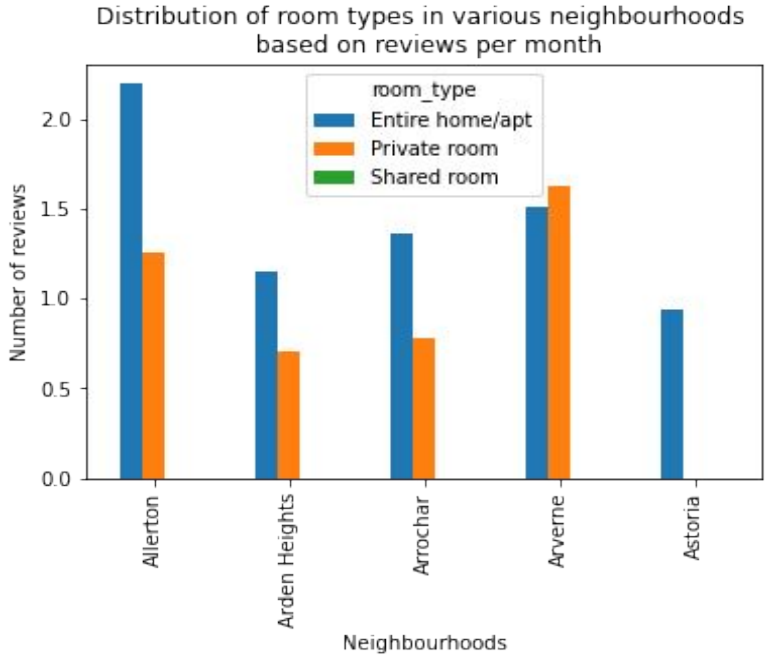


Top 10 listing on the basis of the number of reviews.

- There are 4 manhattan,4 queens and two brooklyn neighbourhood type in top 10 reviewed listing.
- Total of 9 room type are private room and one is entire home/apt and most of them with price range $46-$49.
- The highest no. of reviews given is 629.

The top 10 neighbourhood & neighbourhood group with various room types according to the reviews per month.

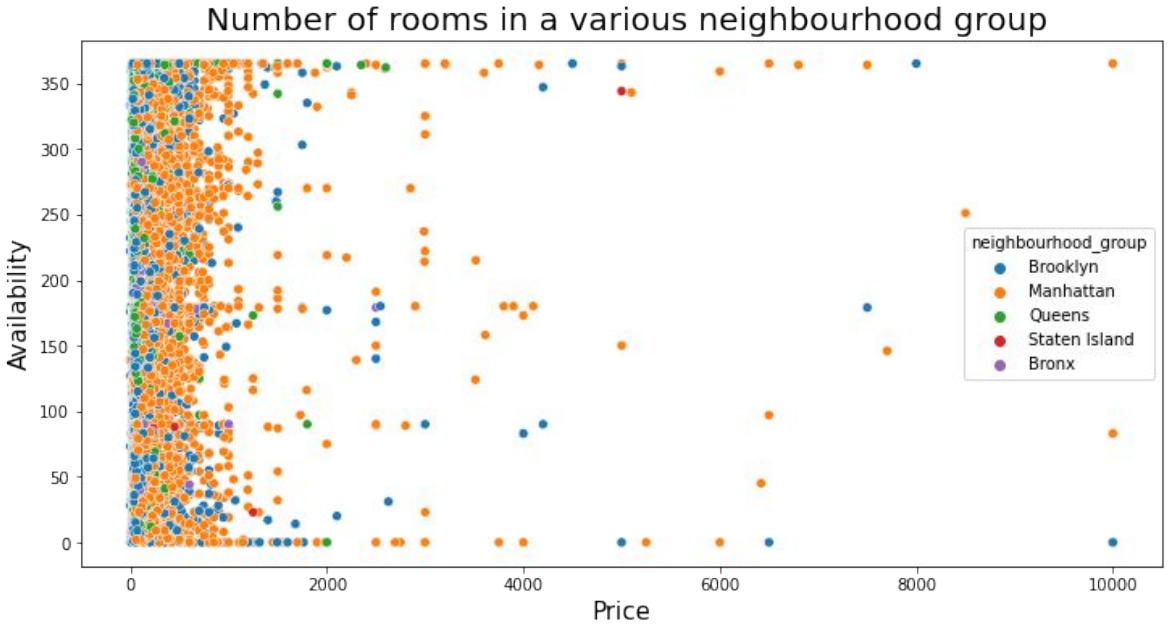




- Except for Manhattan, in all the neighbourhood groups, 'Entire home/apt' is the room type that gets most traffic. We can say that with respect to neighbourhood that none of top 10 locations have reviews for shared room but both the parameters give Allerton neighbourhood in Bronx with 'Entire home/apt' room type as the highest therefore Allerton, Bronx has the highest traffic

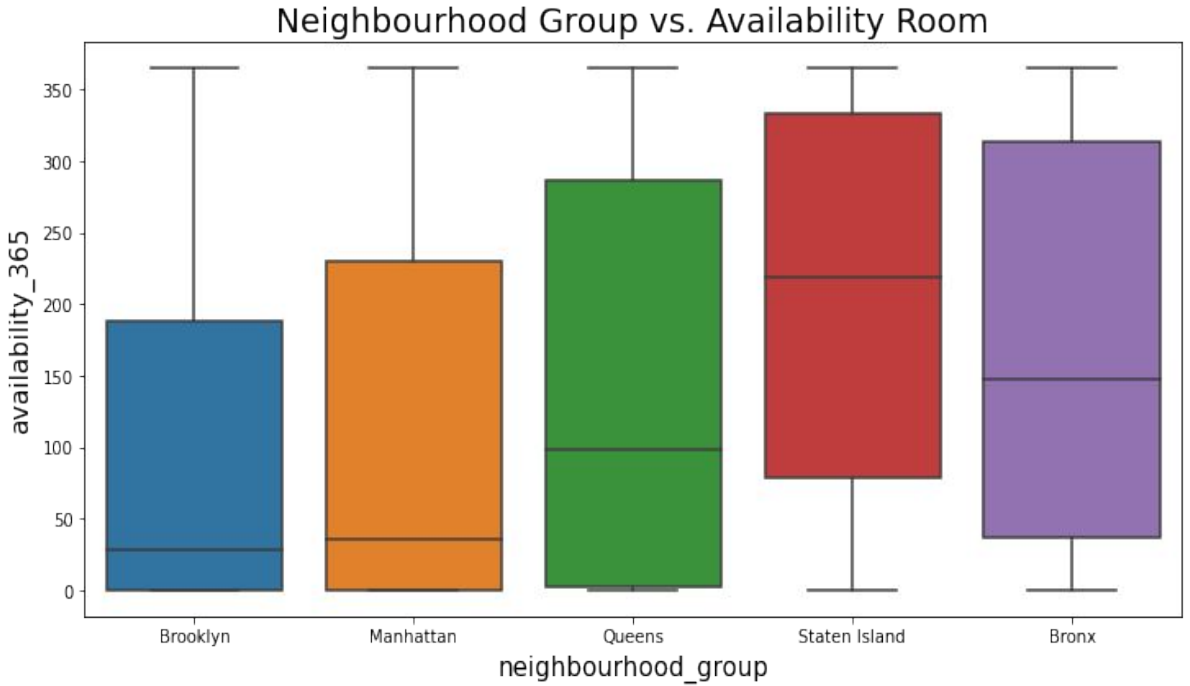
The relationship between the price and availability of various room available in different neighbourhood?

- From the scatter plot we can see that room with price range less than $1000 are available with all the values higher or lower it depends on other factors.
- After price $2000 less room are available and become less further $4000.They can have high availability but they are less in number.
- At last we can see only 3 room available which most expensive with price range of $10,000.



The relationship between neighbourhood group and availability of room

- Staten island has the highest median value on the basis of availability followed by bronx , queens, manhattan and least is brooklyn.
- Queens has the highest
- range of availability in a particular neighbourhood group.



Challenges faced

- Reading the data and understanding the columns.
- For answering some question we had to understand the business model of airbnb and how they work.
- Handling the Nan values , null values and duplicates.
- Creating multiple visualization 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.

Conclusion

The given data appear to be very rich dataset with a variety of columns that allowed us to do the exploration on each significant columns present in the dataset.

- Most of the data is not highly correlated
- Manhatten has the highest no of room present in a neighbourhood group.
- Entire home/apt is the maximum room type preferred by guest.
- Manhatten is maximum in entire home/apt category followed by brooklyn in private room.
- Brooklyn is most dense according to the location and shared rooms are much less in comparison to the other two.
- The host with host_id 219517861 has highest listing of 327.
- Williamburg neighbourhood has the most listings followed by bedford-stuyvesant.
- Manhatten has the highest price range according to both mean and median prices.
- The Entire home/apt have the largest range of price value with median value of $160.
- The host with host_id 17550546 is the busiest host as the number of minimum nights spend is the highest.

- The highest no. of reviews given is 629
- Allerton, Bronx with 'Entire home/apt' room type has the highest traffic according to the reviews per month.
- Staten island has the highest median value on the basis of availability

Thank you