Capstone Project

Airbnb Booking Analysis

by
Ritik Vaidande
Data Science Practitioner
AlmaBetter, Bengaluru

INTRODUCTION

Airbnb is an American Company since 2007, operates an online marketplace focused on short-term homestays and experiences. The company acts as a broker and charges a commission from each booking.

What is Airbnb?

Airbnb, as in “Air Bed and Breakfast,” is a service that lets property owners rent out their spaces to travelers looking for a place to stay.

PROBLEM STATEMENT

Travel industries are having important reflection of the economy
from past few decades, and Airbnb housing price ranges are of great interest for both Hosts and Traveler.

Explore and Analyze the Data to Discover Key Understandings.

DATASET

Shape - (48895, 16)

Our dataset is feature rich containing, location with co-ordinates, prices, host name, room types, availability throughout season.

price
price of listings

name
representing an accommodation

latitude
coordinate

longitude
coordinate

host_id
unique id for host

neighbourhood_group
a group of area, 5 unique hoods

calculated_host_listing_count
no of registered listing under a host

id
unique id

neighbourhood
falls under group

room_type
3 unique room types

availability_365
number of days a host is available

minimum_nights
minimum nights stay required

number_of_reviews
total rating count of a listing

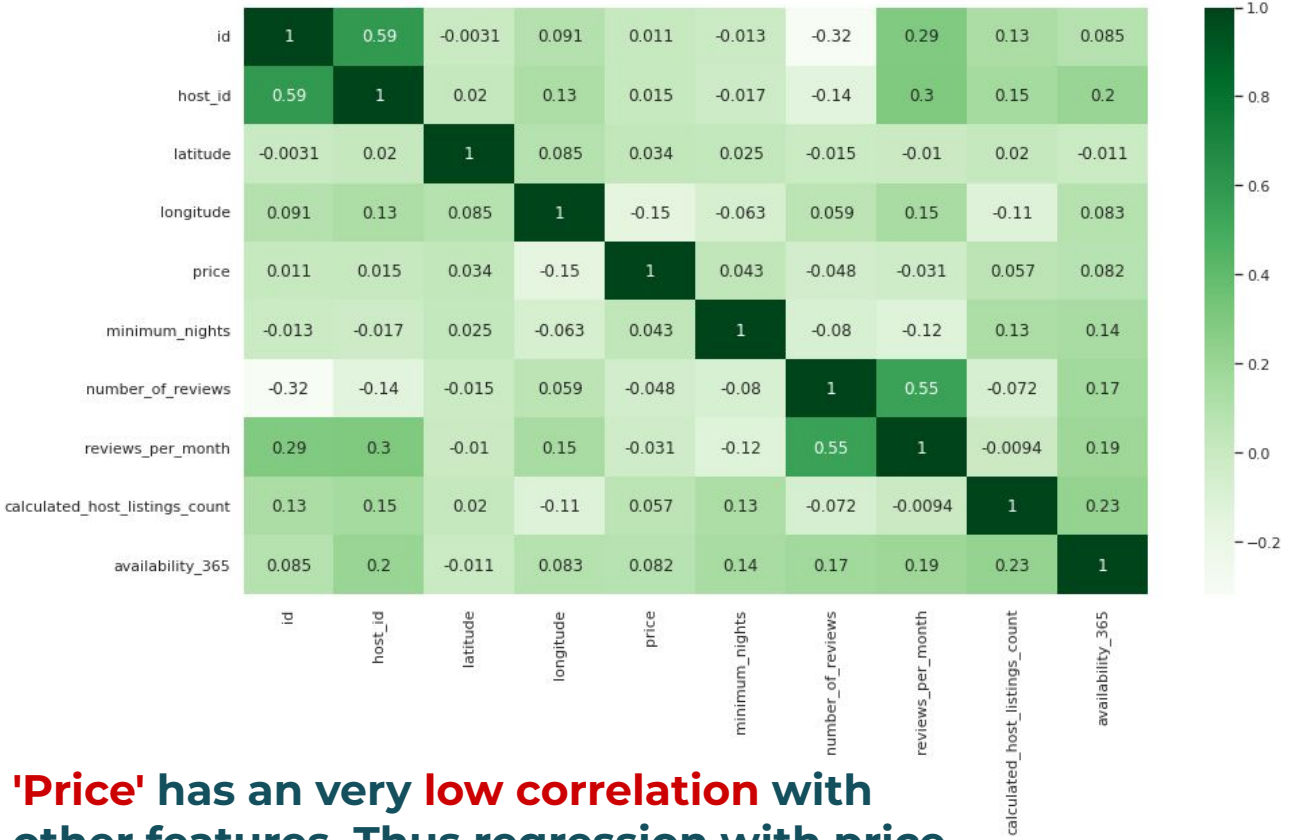
reviews_per_month
rate of reviews given per month

last_review
date of last review given

CORRELATION-MATRIX

Highest Correlation With Price :
availability_365

Lowest Correlation With Price :
number_of_reviews

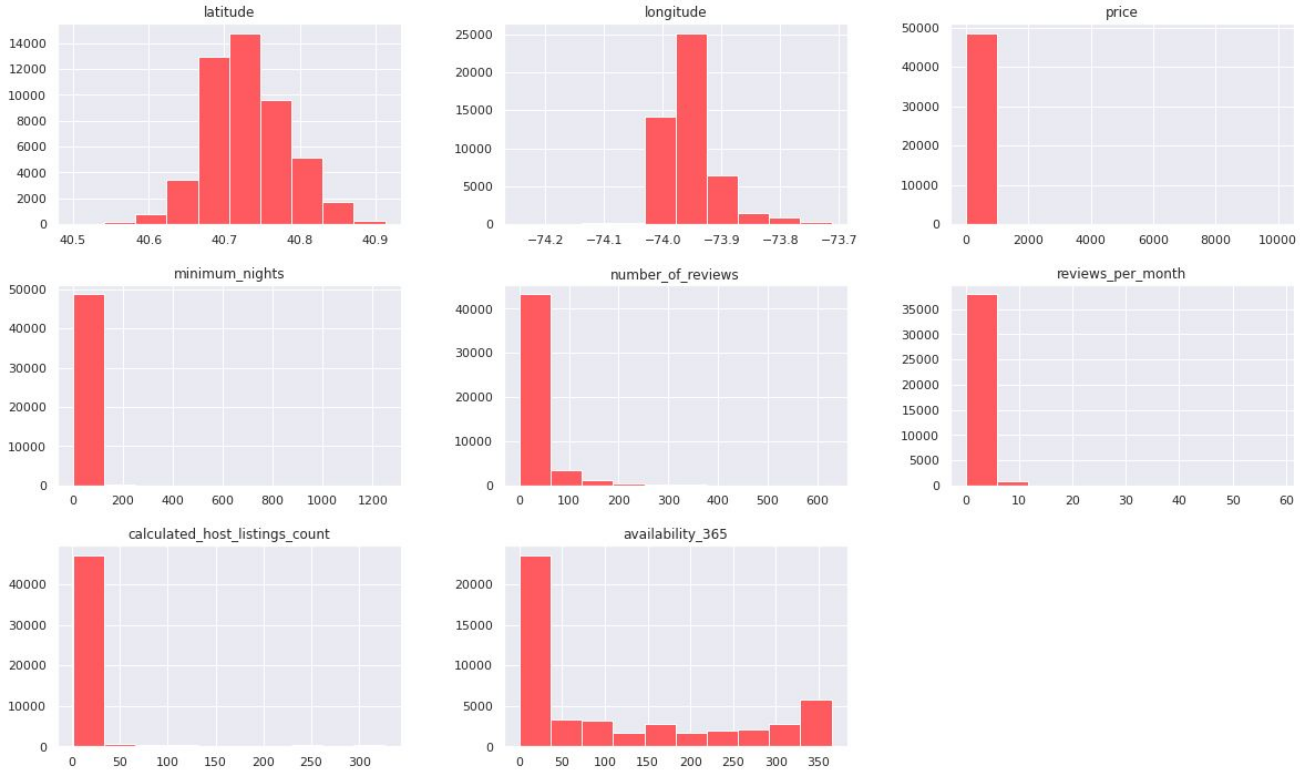


'Price' has an very low correlation with other features. Thus regression with price can be less accurate.

DISTRIBUTION OF DATA

Majority Data Distribution Type :

Positively Skewed



From the distribution plots of the numerical features, we can conclude that most of the data is positively skewed including price. Thus this biased distribution denotes there are many outliers in the data.

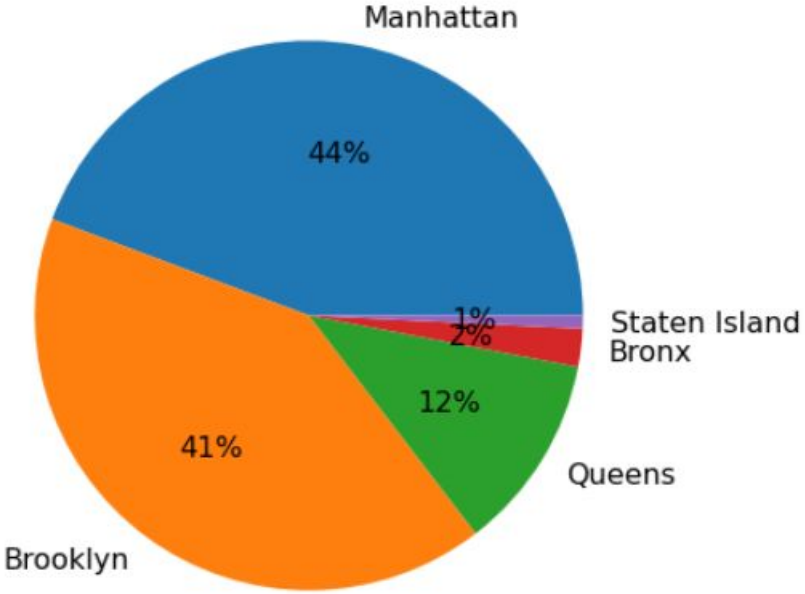
LISTINGS IN NEIGHBOUR-HOOD GROUPS

Highest Listings :

Manhattan

Lowest Listings :

Staten Island

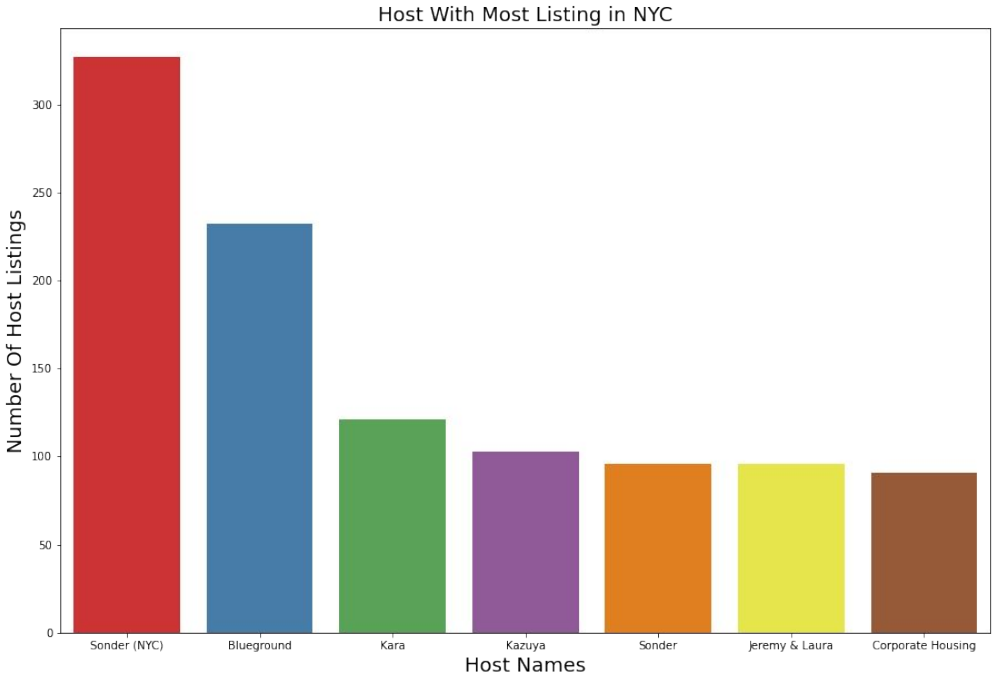


Around 85% of listings are held in neighbourhood groups Manhattan and Brooklyn

HOSTS WITH MOST LISTINGS

Top 7 hosts :

1. Sonder (NYC)
2. Blueground
3. Kara
4. Kazuya
5. Sonder
6. Jeremy & Laura
7. Corporate Housing



Sonder(NYC) with 327 listings has relatively high number of listings than other hosts

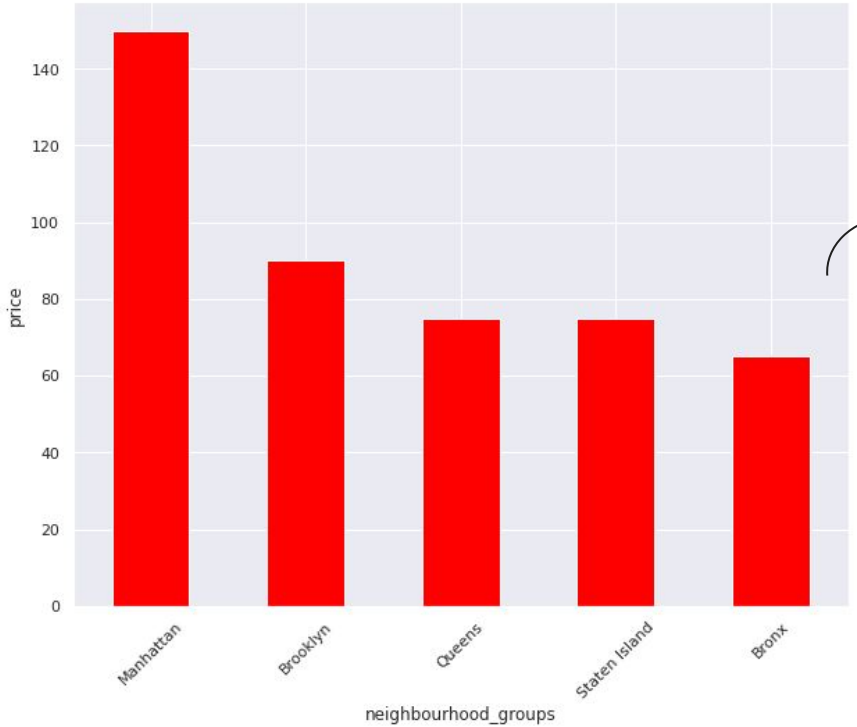
COSTLIEST NEIGHBOUR-HOOD GROUPS

Costliest Neighbourhood Group :

Manhattan

Cheapest Neighbourhood Group :

Bronx

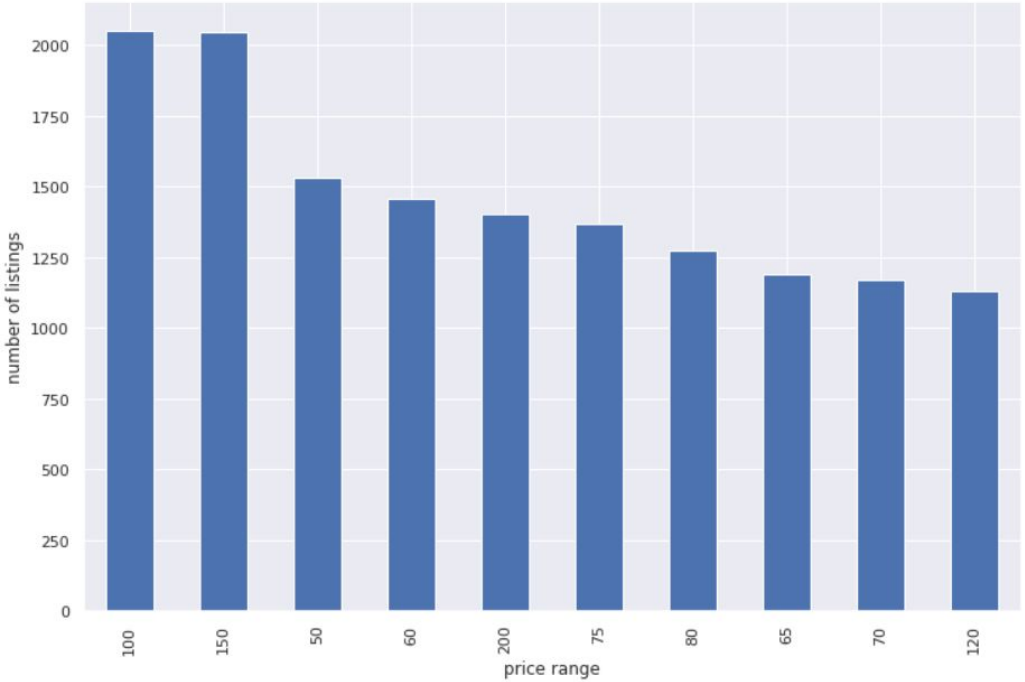


As we saw the Price distribution was positively skewed, therefore for the analysis, median values are chosen, because it is less sensitive to skewed data.

- **Manhattan is the costliest neighbourhood group in NYC**
- **Staten Island is not cheapest, though the number of listings were least.**

PRICE DESCRIPTION AND PRICE RANGE AGAINST LISTINGS

mean - 152$
std - 240$
min - 0$
25% - 69$
50% - 106$
75% - 175$
max - 10000$



- Almost 2k+ Airbnb's has price of 100$ and 150$.
- 1.5k Airbnb's have around 50$ price.

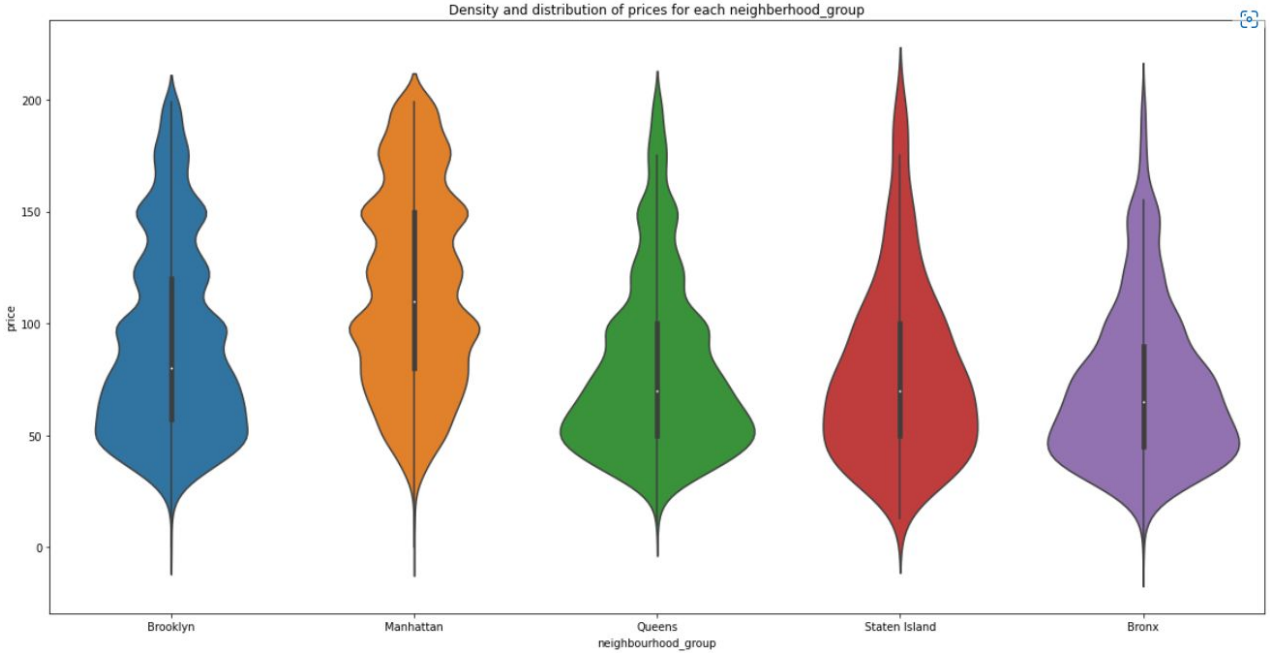
PRICE DISTRIBUTION OVER NYC

Highest Number of Expensive Rooms :

Manhattan

Highest Number of Cheap Rooms :

Bronx



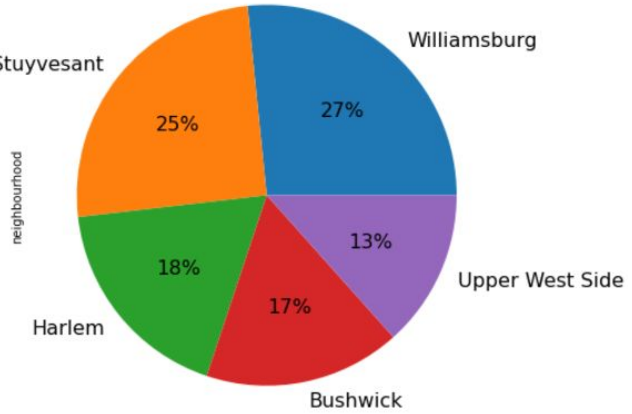
- **Manhattan** Airbnb's has the highest average price.
- **Bronx** provides most number of budget friendly rooms.
- Queens and Staten Island have similar IQR (Inter-Quartile Range) and Median.

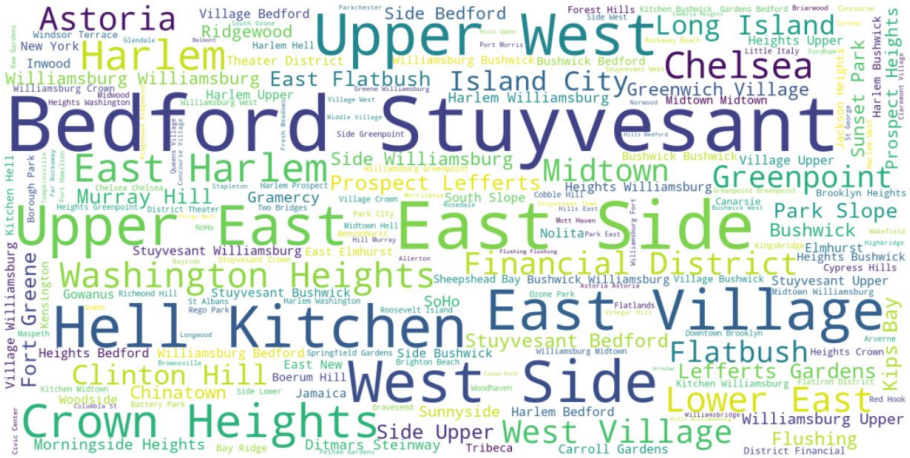
HOTTEST TRANSACTION AREAS

Top 5 hottest transaction areas :

1. Williamsburg
2. Bedford-Stuyvesant
3. Harlem
4. Bushwick
5. Upper West Side

Top 5 hottest areas by percentage. Williamsburg is the hottest area of transaction followed by Bedford-Stuyvesant.





This wordcloud shows most prominent neighbourhood areas respect to listings.

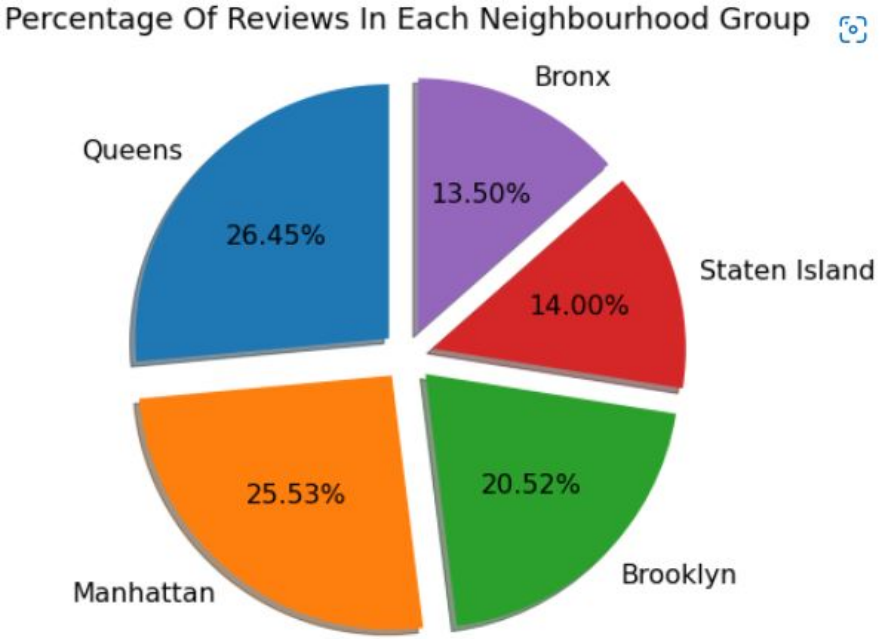
REVIEWS AGAINST NEIGHBOUR-HOOD GROUP

Highest Reviews :

Queens

Lowest Reviews :

Bronx

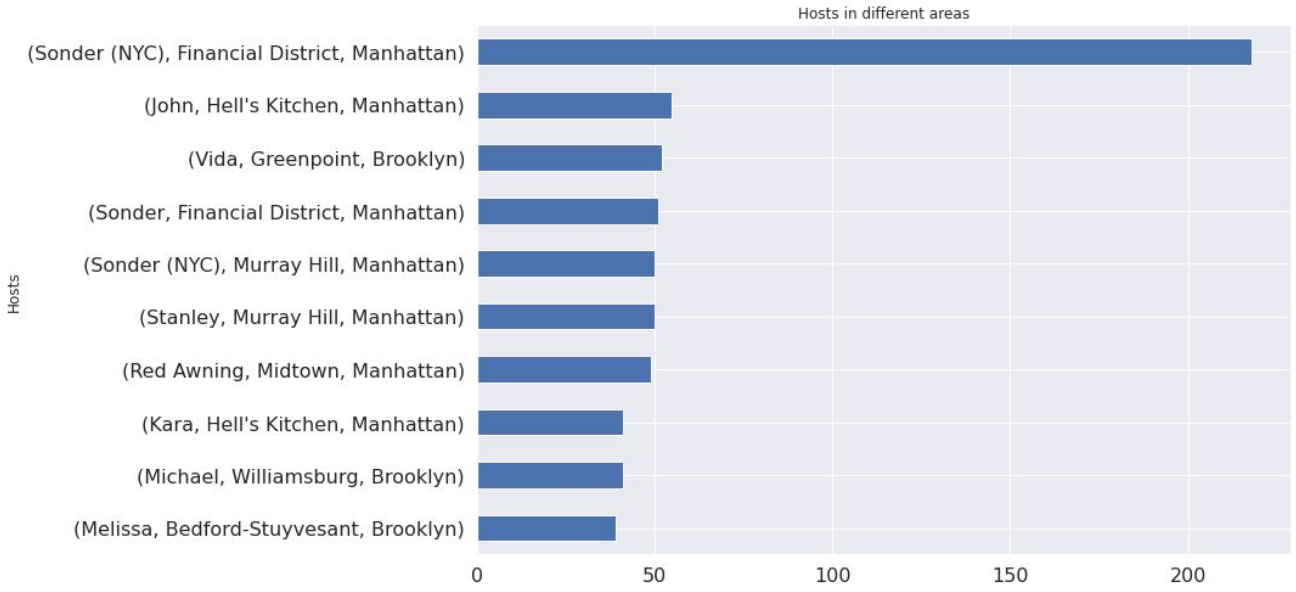


More than 50% of reviews are from Queens and Manhattan. Even though we saw the number of listings of Queens were less then half compared to Manhattan or Brooklyn, Queens has the most reviews than any another neighbourhood group. This shows that Queens is most likely to preferred by people.

BUSIEST HOSTS AND THEIR NEIGHBOUR-HOODS

Most Busiest Host :
Sonder (NYC)

Most Busiest Neighbourhood:
Financial District, Manhattan



These are the Top 10 busiest host among their respective neighbourhoods. Financial District is the busiest among all neighbourhood. Most busiest hosts are from Manhattan region.

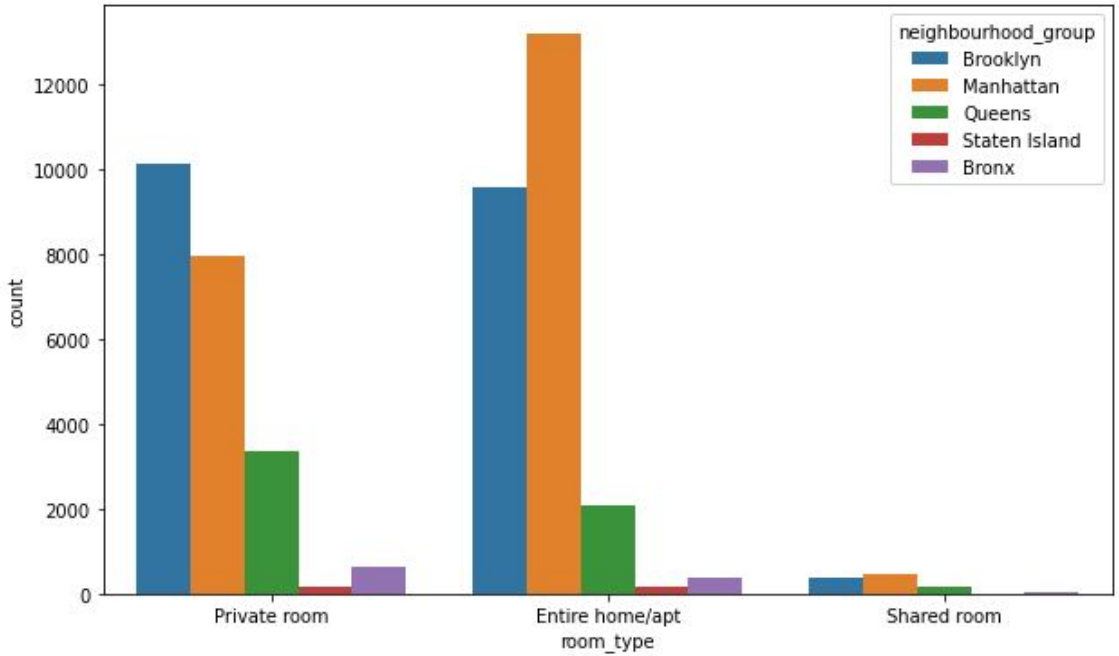
ROOM TYPE PREFERABILITY

Most Preferred Room Type :

Entire Home

Least Preferred Room Type :

Shared Room



- **90% of the listings on Airbnb are either Private Room or Entire/Home. Guests mostly have opted for Private Room/Entire Room as it's much easier for guests to stay in and rent on Airbnb.**
- **Entire Home is also the highest service used across NYC.**
- **In Queens, Private rooms were more used.**

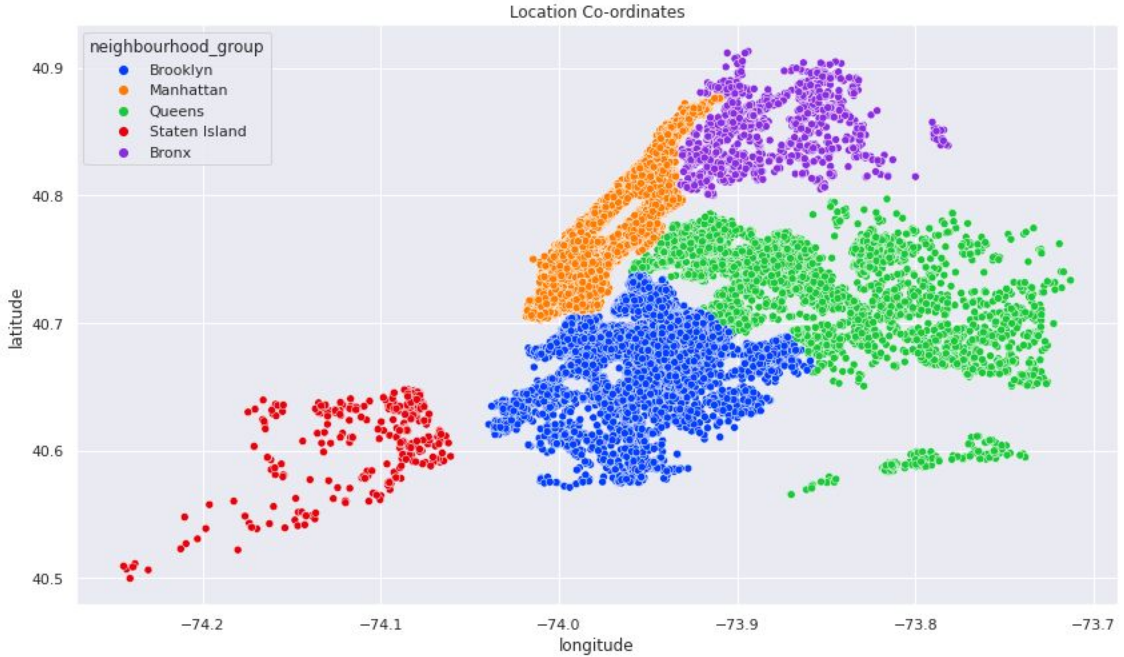
LOCATION CO-ORDINATES OF NEIGHBOUR-HOOD GROUP

Most Scattered Neighbourhood :

Queens

Most Dense Neighbourhood :

Manhattan



- Manhattan and Brooklyn co-ordinates are more dense, as we saw around 85% listings are held by both neighbourhoods.
- As you can see that Staten Island and some areas of Bronx are slightly far from other neighbourhoods, this could be a reason they are much cheaper than other neighbourhoods.

ROOM TYPES AGAINST NEIGHBOUR-HOOD GROUPS

Most Owned Room Type :

Entire Home

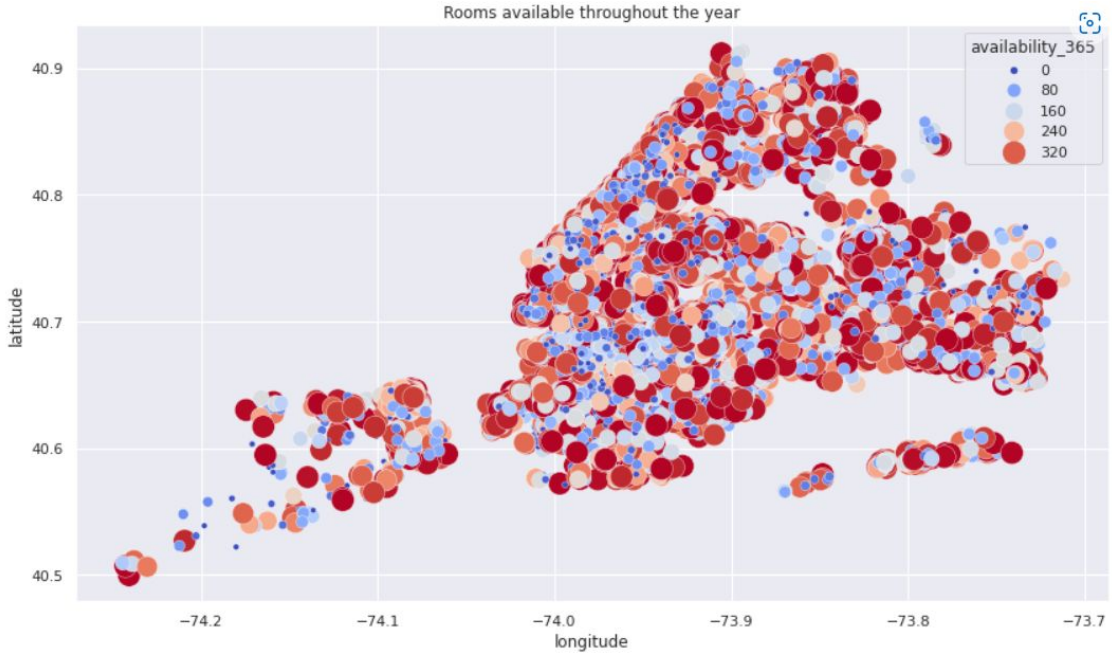
Least Owned Room Type :

Shared Room

- By the above scatterplot of latitude vs longitude we can conclude that there's Shared Room listings throughout NYC are very less as compared to Private and Entire Home. Entire Home is also the highest service used across New York City.
- 90% of the listings on Airbnb are either Private Room or Entire/Home. Guests mostly have opted for Private Room/Entire Room as it's much easier for guests to stay in and rent on Airbnb.

ROOM AVAILABILITY THROUGHT THE YEAR

- Most rooms are available throughout the year
- Every neighbour-hood as the almost same proportion of 365 days room availability.



- As we discussed in previous analysis that Bronx and Staten Island listings are way cheaper than compared to others like Manhattan and Queen.
- As you can see that Staten Island and Bronx listings are available throughout the year, now this might be case that they are way cheaper than other neighbourhood groups.

SCOPE OF IMPROVEMENT OF DATASET

As dataset has few qualifying attributes to value a property, more features can be added like bedroom, bathroom, property age (it might be one of the most important one), applicable tax rate, distance to nearest airport, hospital or schools.

In presence of ratings, hosts can be classified and ranked, special discount or offer can be
given to highest rated hosts following marketing strategy

CONCLUSION

Manhattan and Brooklyn are the two distinguished, expensive & posh areas of NYC.

Though location of property has high relation on deciding its price, but a property in popular location doesn’t mean it will stay occupied in most of the time.

Most visitors don’t prefer shared rooms, they tend to visit private room or entire home.

Thank You

This presentation is an part of the EDA Capstone Project by Ritik Vaidande, AlmaBetter.