

Capstone Project

Airbnb Bookings Analysis

by
Ganesh Bayas
Data Science Trainee
AlmaBetter, Bengaluru

Introduction



Since 2007, Airbnb operates an online marketplace for lodging, homestays & tourism activities.

Dataset Information (48895, 16)

Null count:

name - 16

host_name - 21

last_review - 10052

reviews_per_month - 10052

id

unique id

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

price

price of property

neighbourhood

falls under group

availability_365

number of days a host is available

room_type

3 unique room types

minimum_nights

minimum nights stay required

number_of_reviews

total rating count of a listing

reviews_per_month

rate of review given per month

last_review

date of last review given

WordCloud

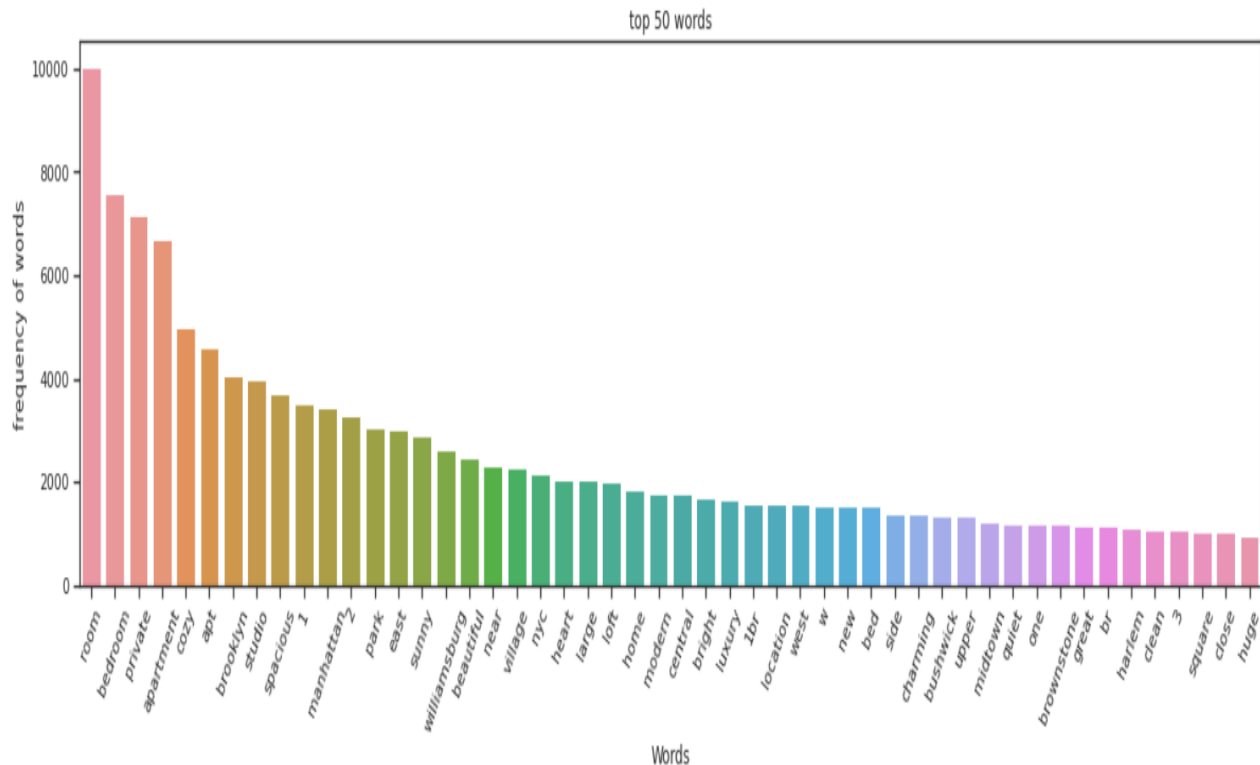


AI

Wordcloud Image

It shows all the frequently used words in dataset.

Top 50 words

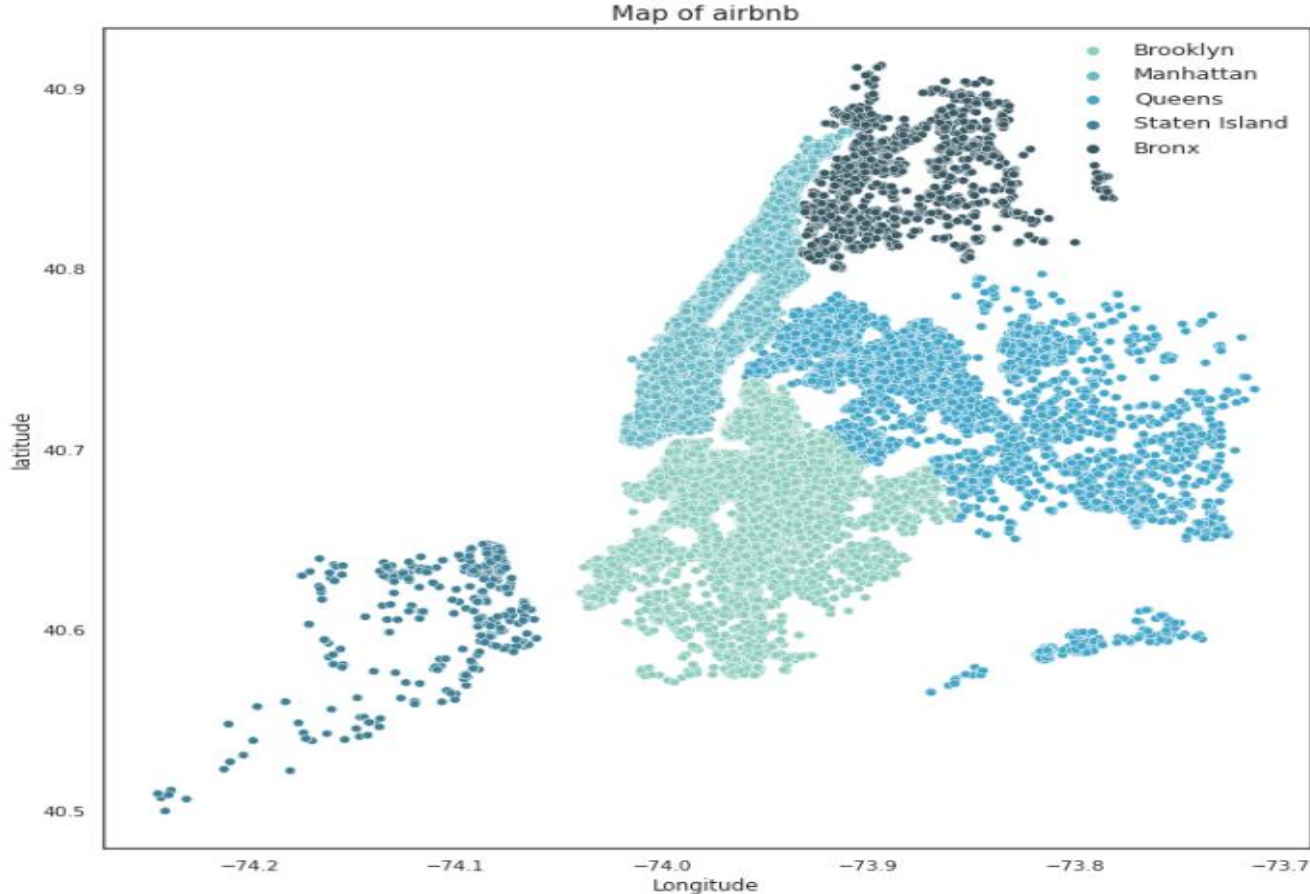


The above observation shows top 50 words with their frequency.

Location of Neighbourhood Groups

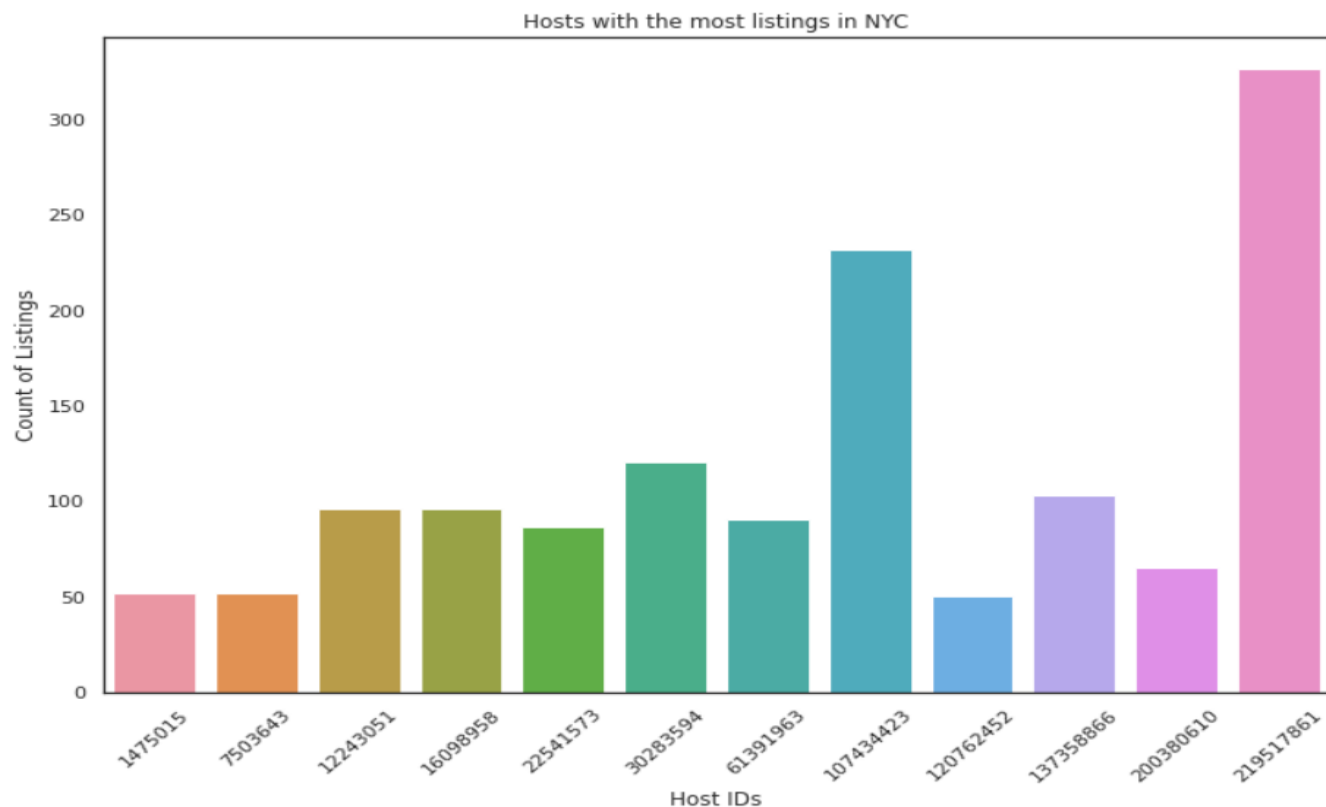
5 unique groups :

Manhattan
Brooklyn
Queens
Staten Island
Bronx



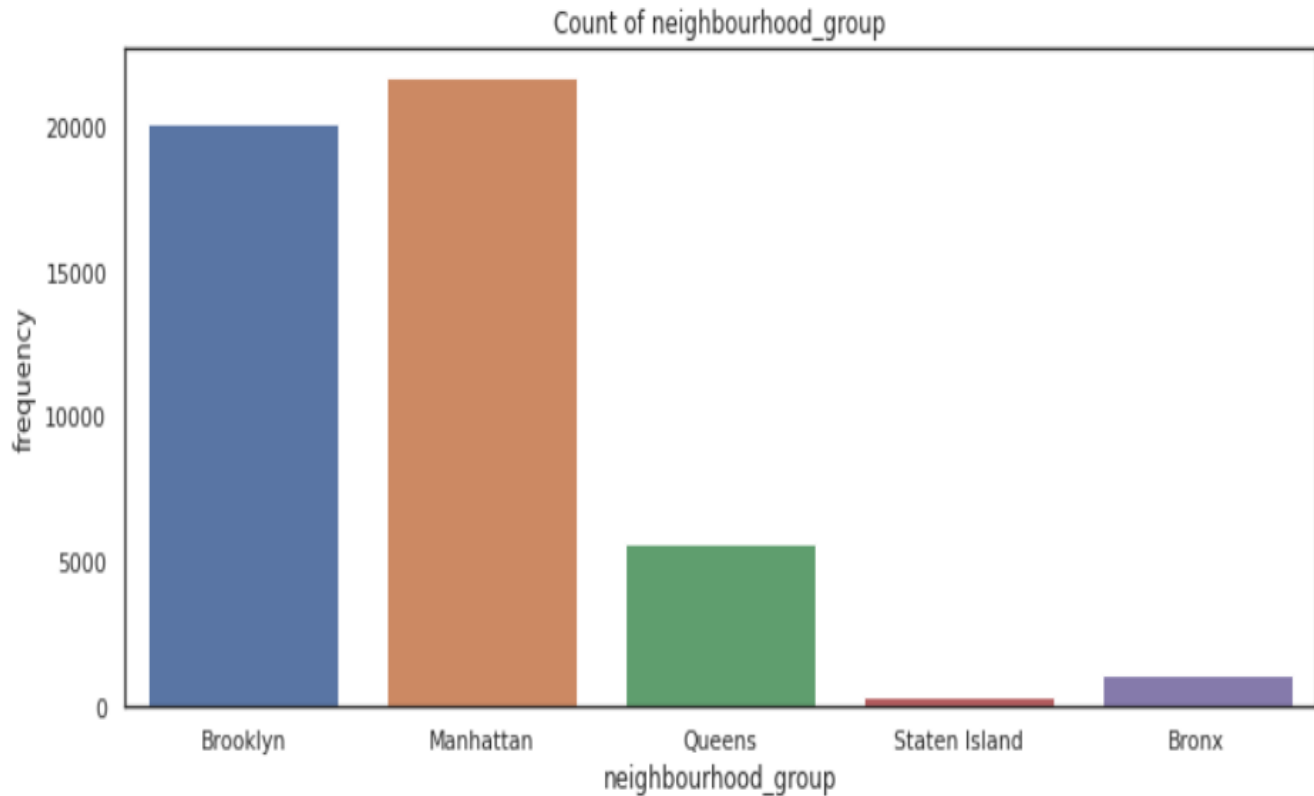
**There are 211 unique neighbourhoods
falls under 5 groups**

Host with most listings in NYC



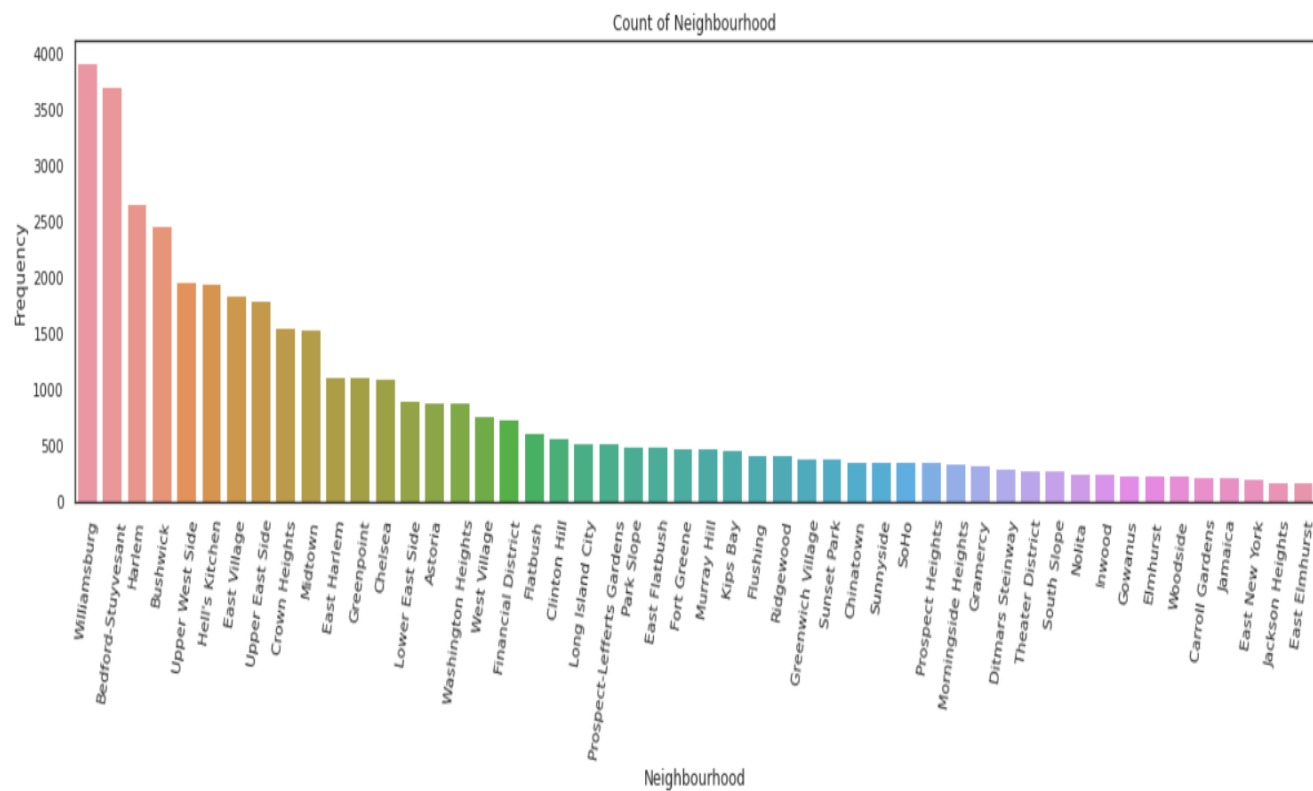
The above bar chart shows Hosts with the most listings in NYC it also shows distinct host ids with number of listings.

Neighborhood Group analysis



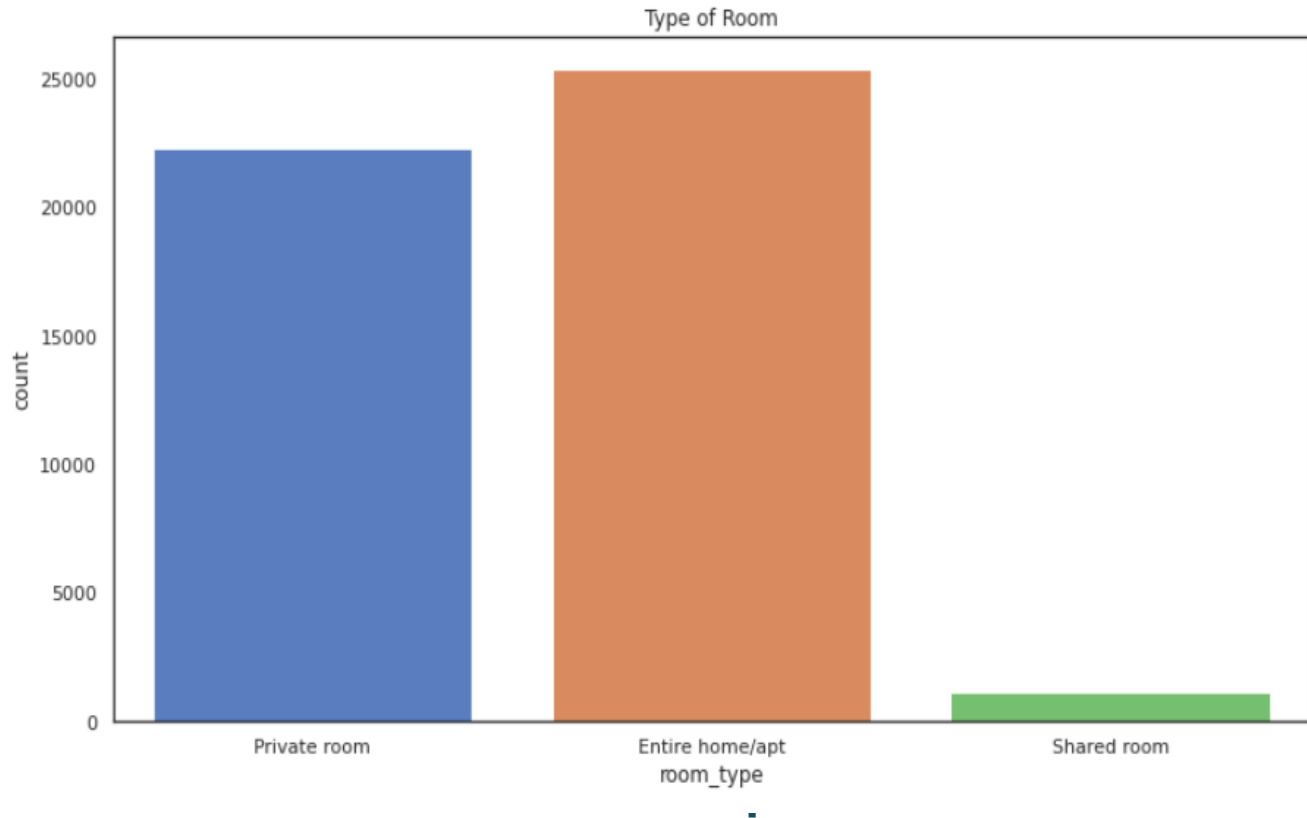
The above chart shows neighbourhood groups booking frequency the highest bookings are in Brooklyn and manhattan.

Neighbourhood demand chart



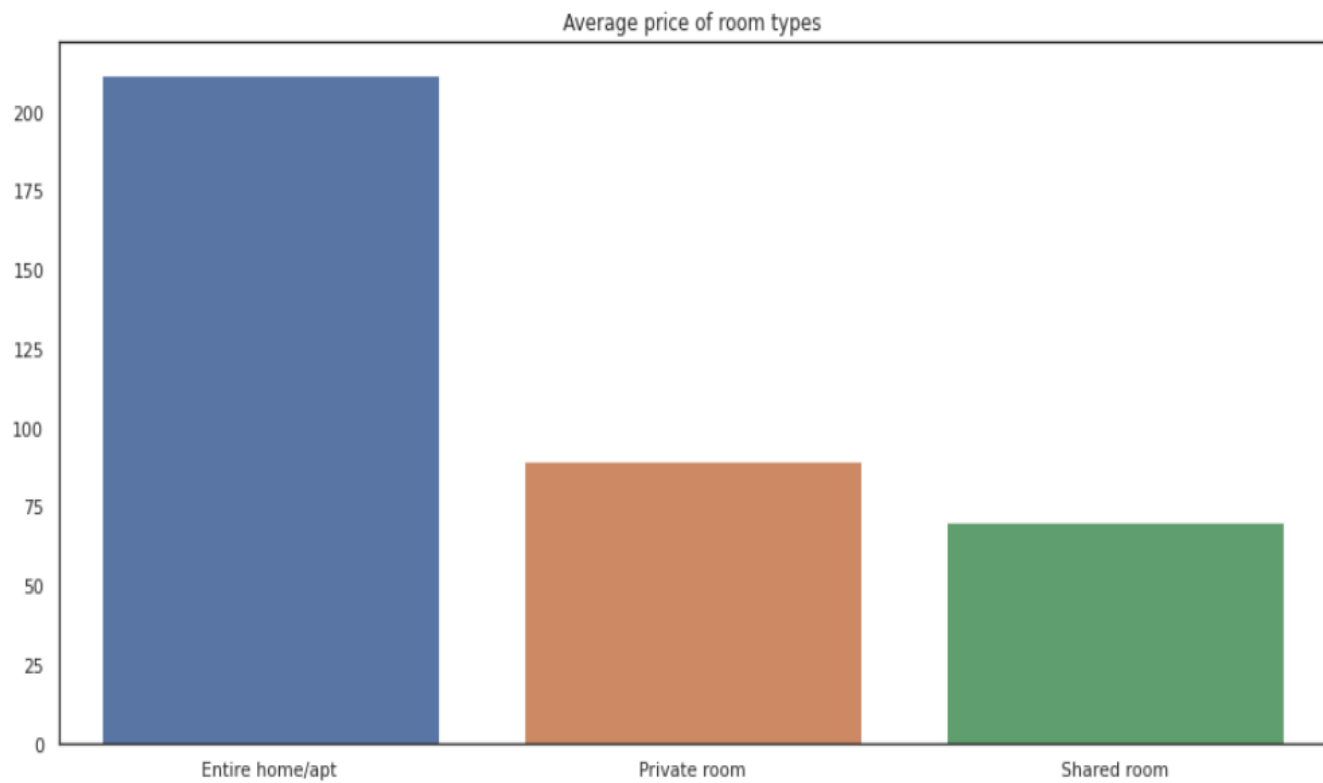
The above chart shows us neighborhood towns with the highest and lowest demand from customers across the platform.

Room type demand analysis



The above chart shows there are 3 types of rooms in which entire home/apt having highest demand followed by private room then shared room.

Price analysis of Room types

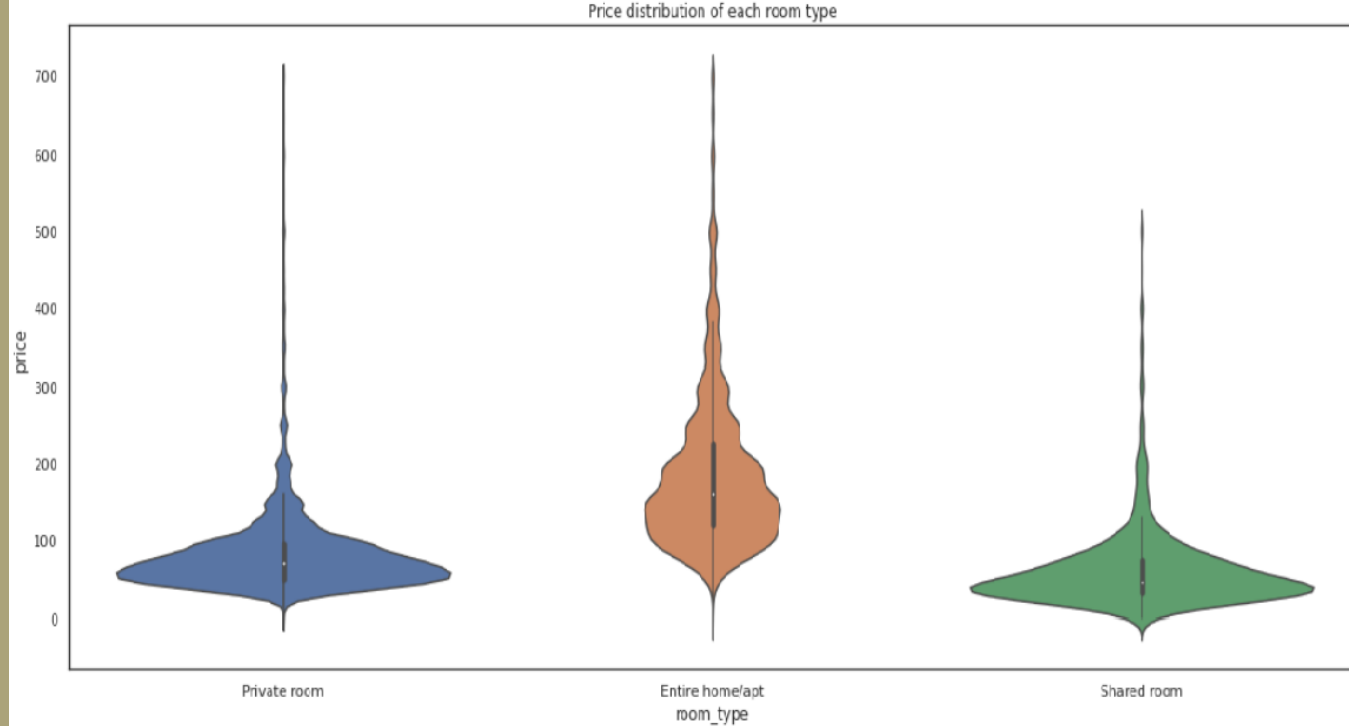


In above chart Entire home/apt has more than 50% avg cost private room and shared room avg cost is not having price difference morethan 20%.

Price distribution analysis

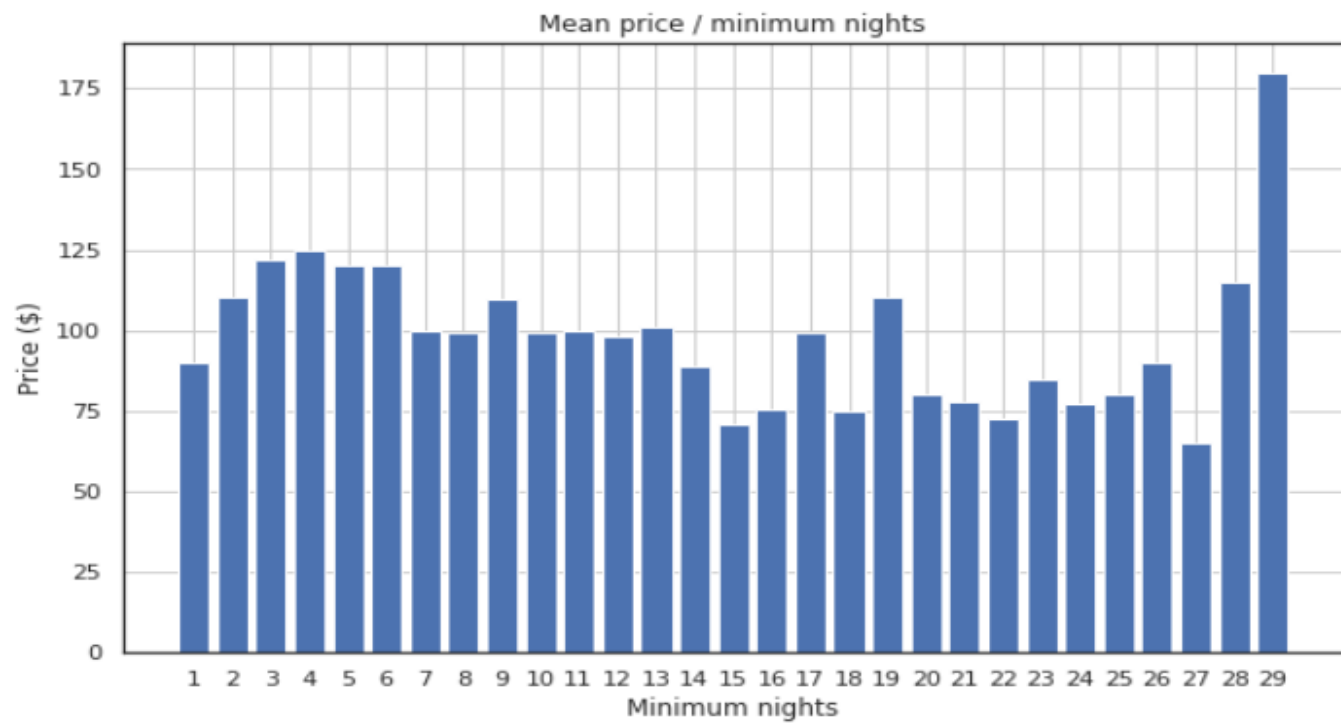
we have considered to divide the whole price range into three categories

- 1.cheap (price range below or equal to 80\$)
- 2.Affordable(for price range 80 to 500\$)
- 3.Expensive(for price range more then 500\$)



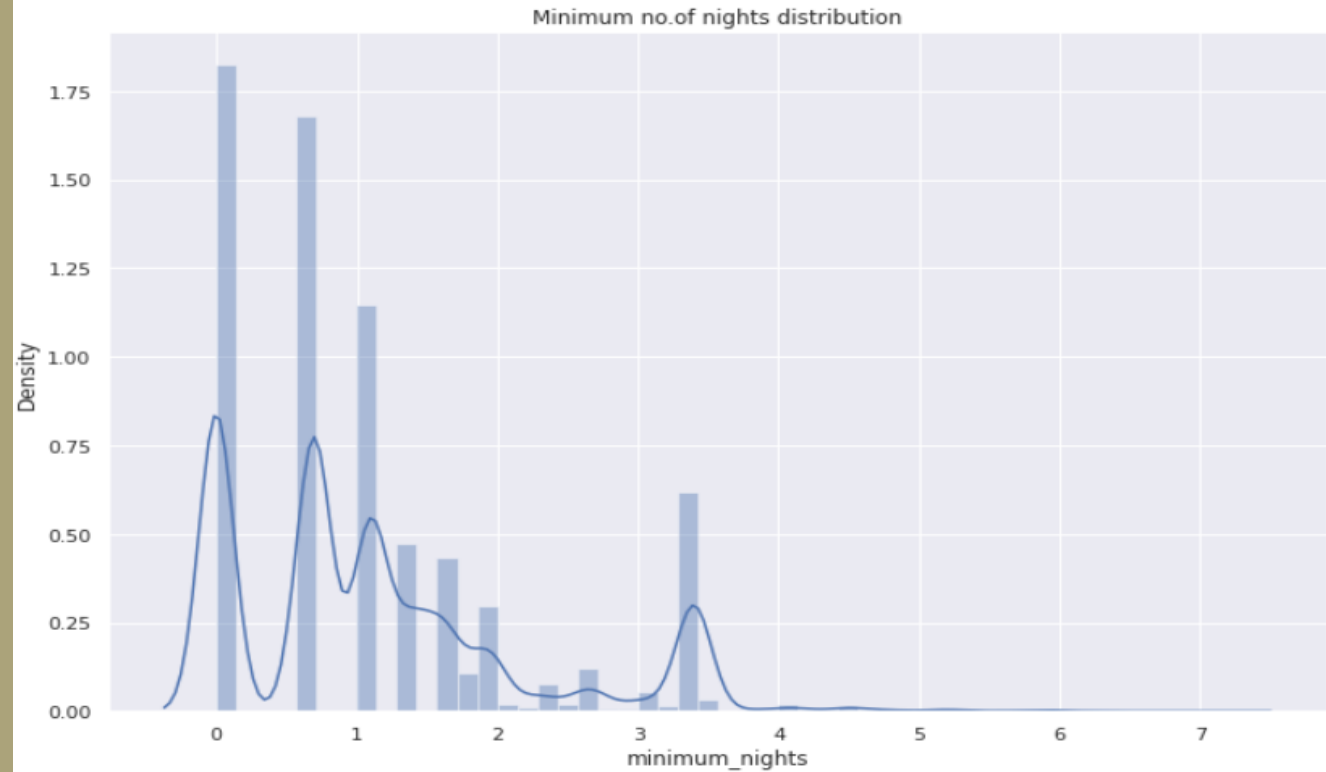
so, it look like people have more intrest in having "affordable" rooms/apartments rather than having cheap and expensive rooms

Mean price per night analysis



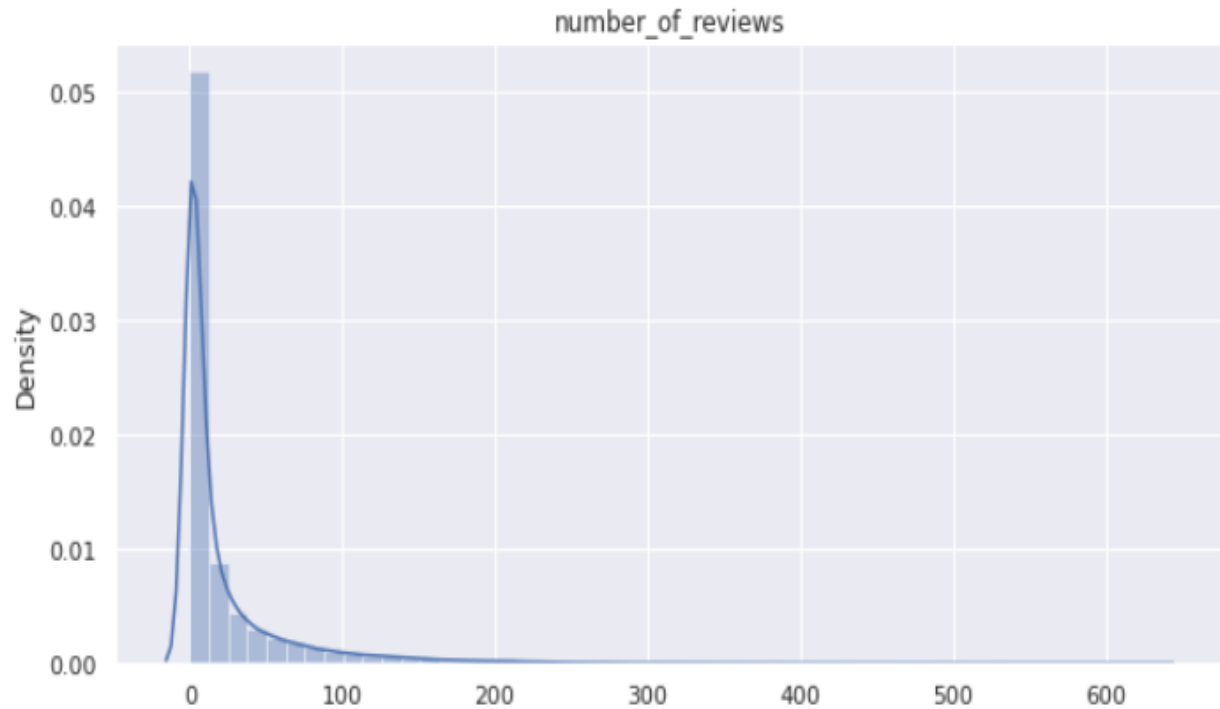
**It's generally cheaper to stay in rooms
between 14 and 28 nights.
Usually, the minimum required nights to stay
in a room is around 2.**

Minimum no of nights distribution



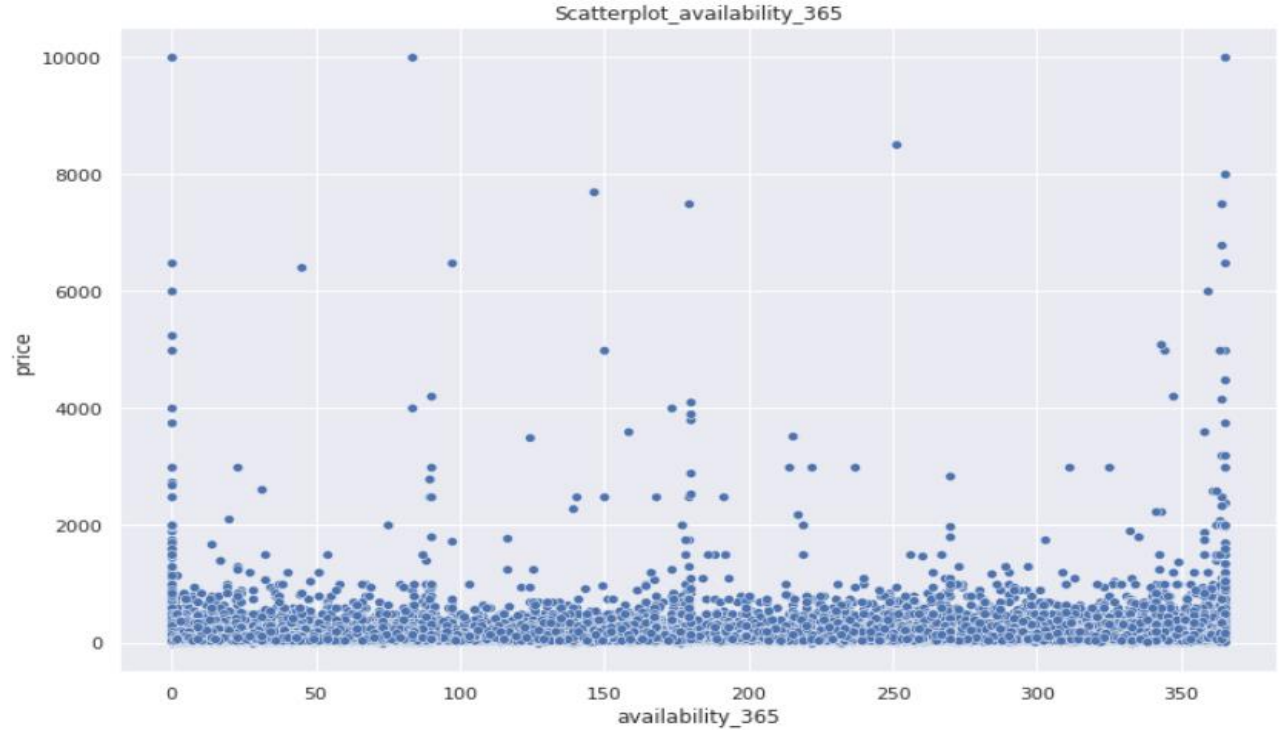
This plots shows that majority of room booking is one for 1 to 4 days. Box-Cox transformed plot strictly shows that the majority of booking lies between 0 to 3 days.

Distribution of number of reviews



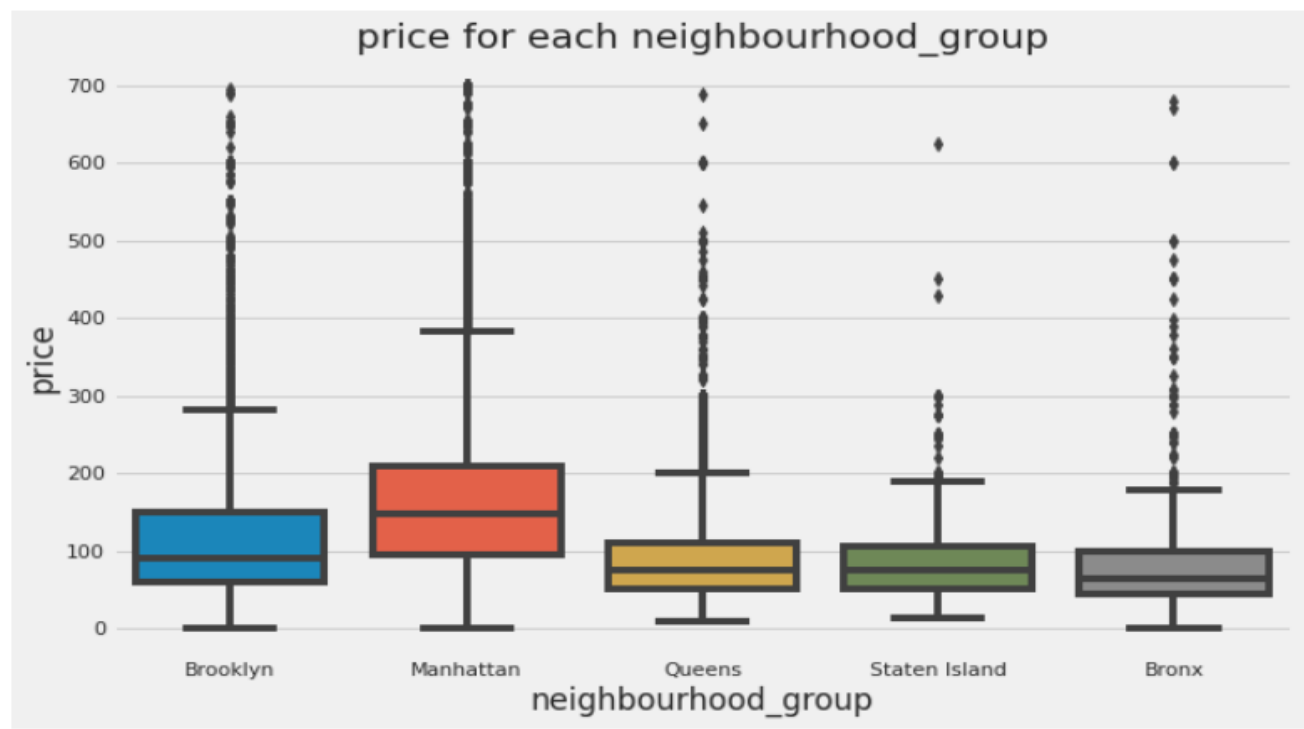
Number of reviews are highly dense from 0 to 100 reviews. We can say that most of the rooms are not rated and those which are frequently occupied only those are rated.

365 days availability analysis



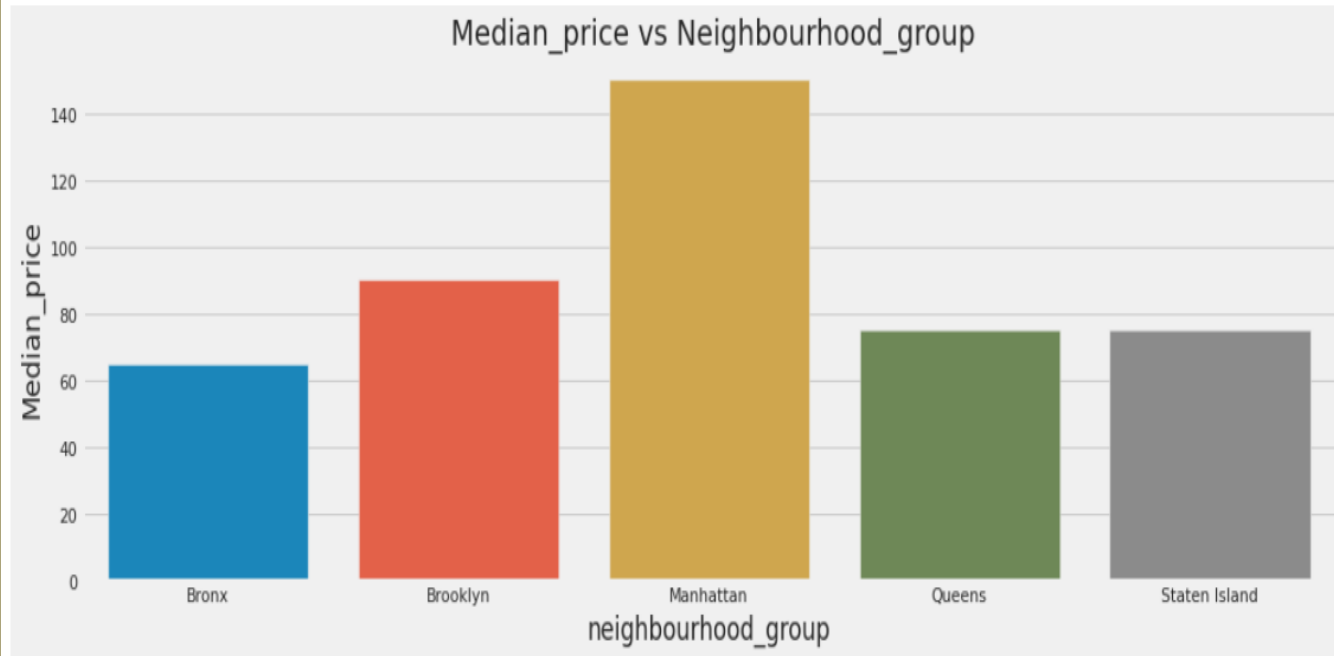
From above plot we can see that most of the available rooms are in the range of 0 to 2000. Very few are available for price above 2000\$,this is quite obvious that there are very few peoples who prefer to have expensive Rooms

Neighbourhood group wise price analysis



We can see that Manhattan is the most expensive destination immediately followed by Brooklyn. Queens, Staten island and Bronx, are having price range less as compared to other two. we know no of reviews is directly proportional to no of guests.

Neighbourhood group mean price analysis



Above chart shows neighborhood wise median prices as we can see Manhattan is the most expensive neighborhood group with \$145 followed by Brooklyn then Queens, staten islands, Bronx.

Limitation

Dataset features in terms of modern world, are of very poor quality in deciding the valuation of a property

User ratings of hosts aren't available, it would've been better to rank our hosts based on user satisfaction and ratings. Normally a low rated property tends to lower their price

In order to have a better analysis regarding the quality of the properties, it would be interesting if we had an analysis of sentiments with property valuations.

The exact number of guests count also missing

Scope of Improvement

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

Time series analysis can be done to make prediction on occupancy rate based on tourist season

Conclusion

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

Comparing the price/popularity variables suggests that people who travel and use Airbnb tend to prefer listings that are cheaper.

Have a room "close" to things that affect popularity (it might be a good idea to include those words in the room name).

NYC shared rooms tend to be clustered in the city center, perhaps because there are more travelers who want to visit the most famous cities.

Thank You

As part of EDA Capstone Project by

