Capstone Project

Project Title - Airbnb Bookings Analysis

- Aniket Satpute
- Kaiwalya Zankar
  (AlmaBetter Trainee)

Introduction

Airbnb, is an American company that operates an online marketplace for lodging, primarily homestays foe vacation rentals, and tourism activities.

Airbnb connects people who want to rent out their homes with people looking for accommodations in that locality .New York is the most populous city in the United States, and one of the most popular tourism and business places globally.

Airbnb offers people an easy, relatively stress-free way to earn some income from their property.

of the Project-

- **The objective of the project is obtain the meaningful analysis about the Airbnb Booking with respect to the factors associated with it.**
- **Mainly focused on terms as below-**

1. Data Prepping/Wrangling.
2. Checking the Null values for cleaning the Dataset for further analysis
3. Checking the unique values for Analyzing the Dataset for further analysis
4 Exploration of Neighbourhood group, Neighbourhood ,Roomtype, Price ,Reviews.
5. Correlation between variables.
6. Analysis of V hosts and key findings.

Problem Statement-

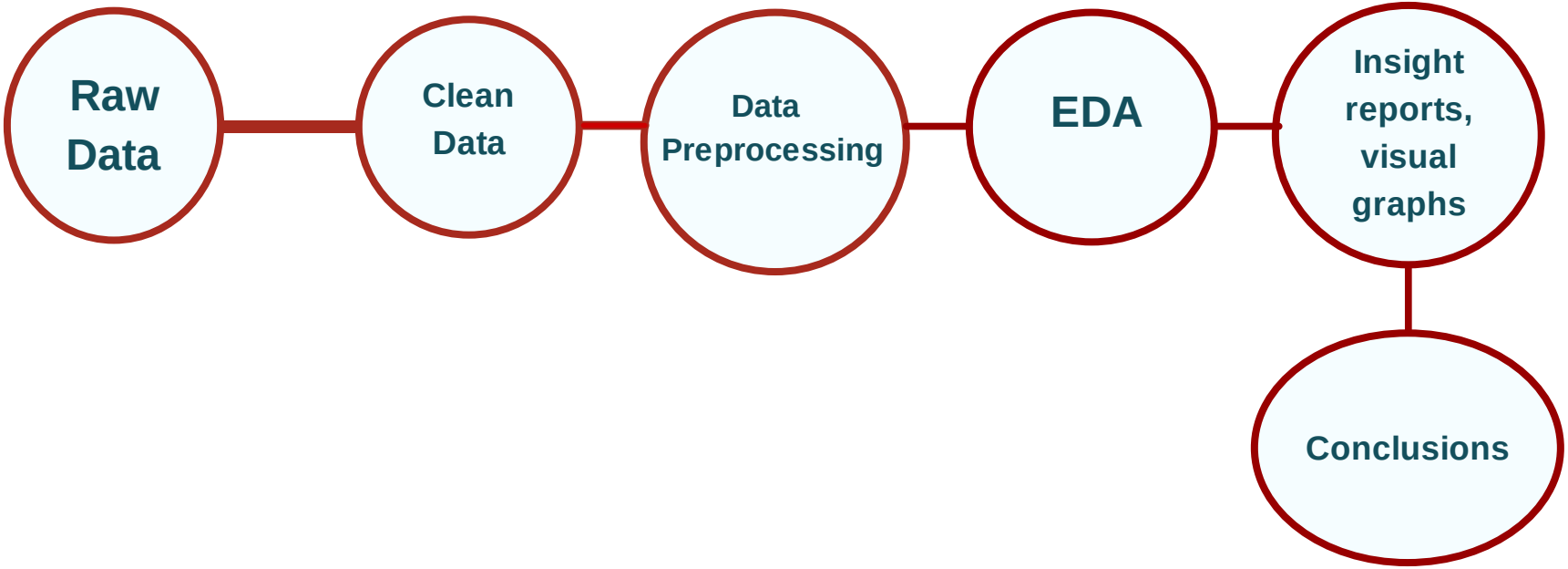
What can we learn about different hosts and areas?

What can we learn from predictions? (ex: locations, prices, reviews, etc)

Which hosts are the busiest and why?

Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Process Flow-



Data Pre-processing-

Clean Data:-

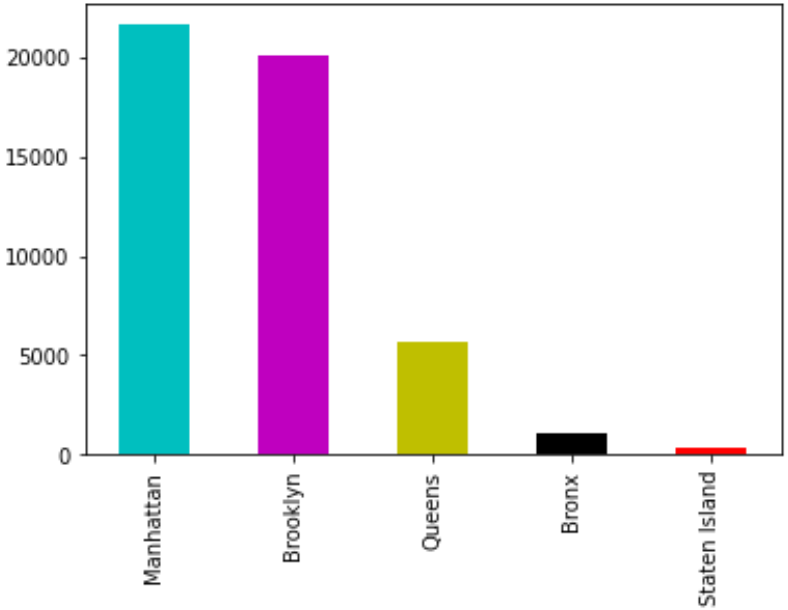
If Null values are present in the dataset then chances of errors are maximum.
To overcome this, we have to remove/replace the Null values in data set for features name as-
name, host name, last review, and reviews per month.
Also, last review, and reviews per month has huge Null values in data set.

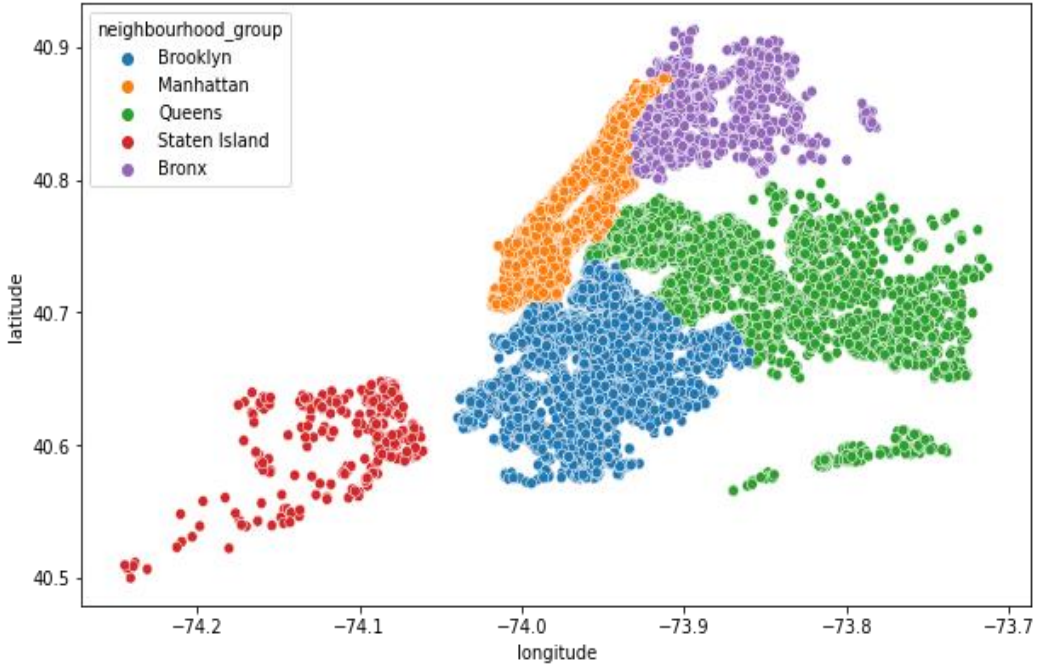
Libraries used-

- - numpy
- - pandas
- - matplotlib.pyplot
- - seaborn

Exploratory Data Analysis - Visualization and Descriptive Statistics

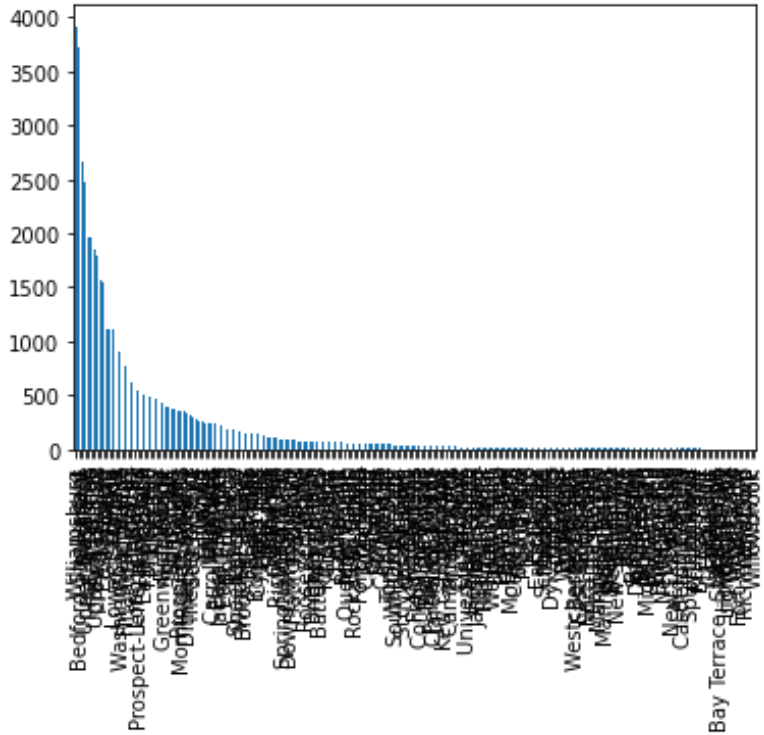
Exploration of Neighbourhood group





- From the above ,it looks like Manhattan and Brooklyn has more number of listing than the Queens, Bronx, and Staten Island.

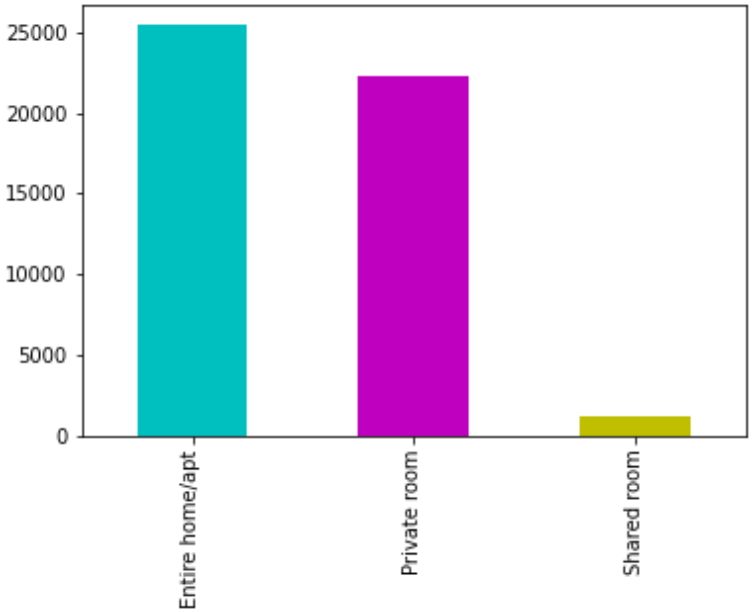
Exploration of Neighbourhood -

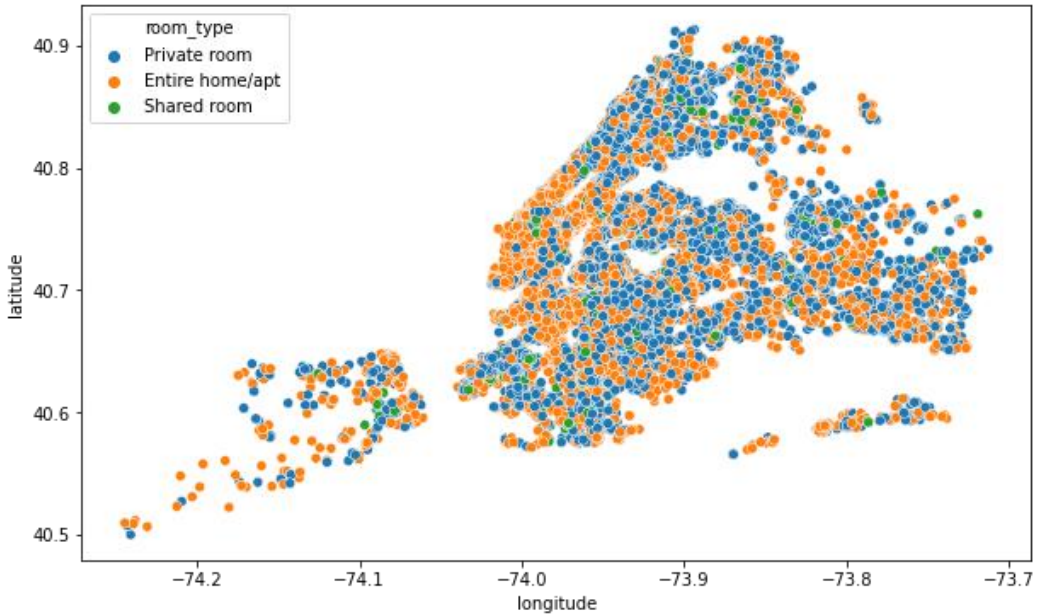


Williamsburg has the highest number of properties listed then the Brooklyn. And Willowbrook has the lowest number of properties listed .

.

Exploration of Room Type

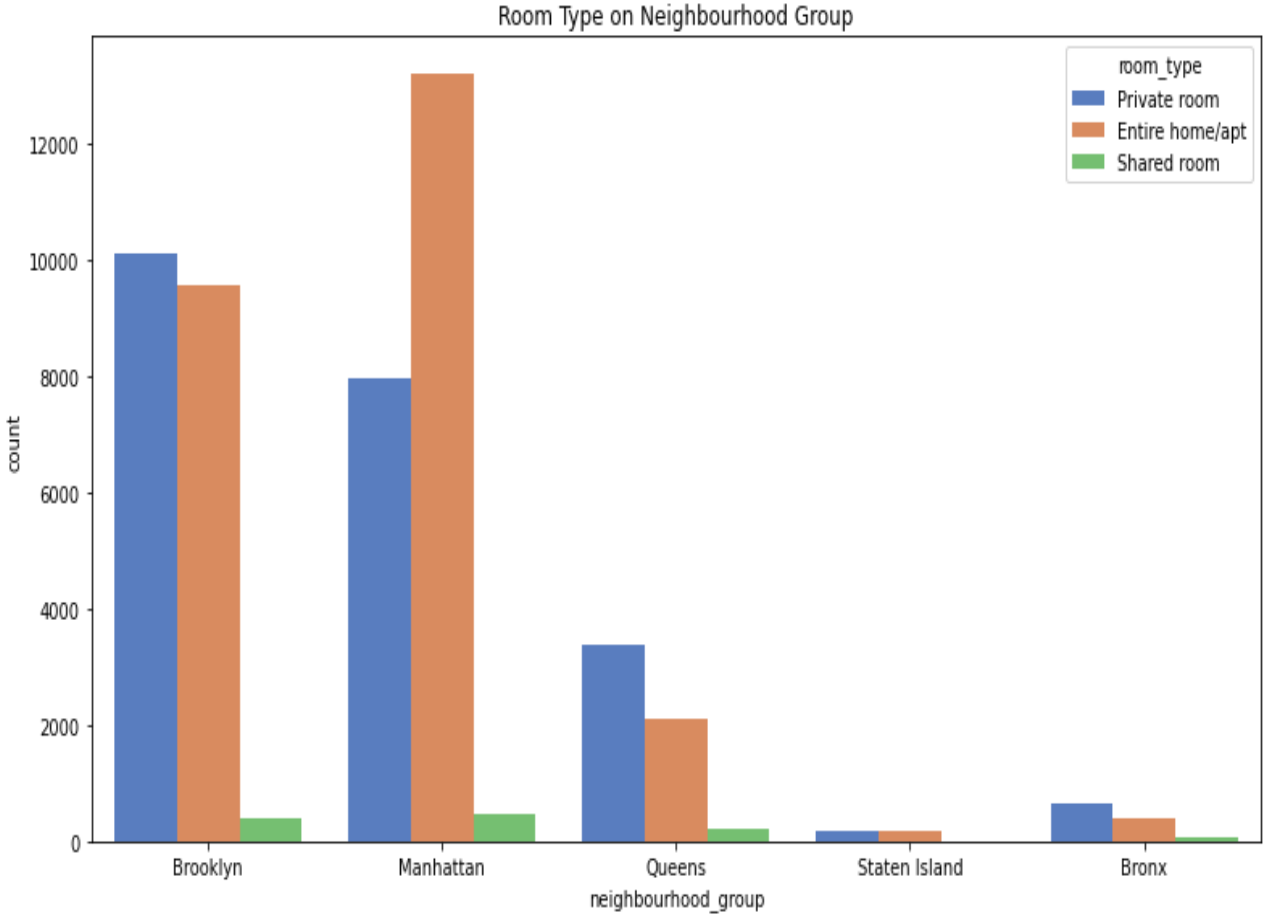




- There are 3 room type provided by the host**. Private ,Entire house , shared room**.
- Most of the rooms provided are private rooms and Entire home or apartments type

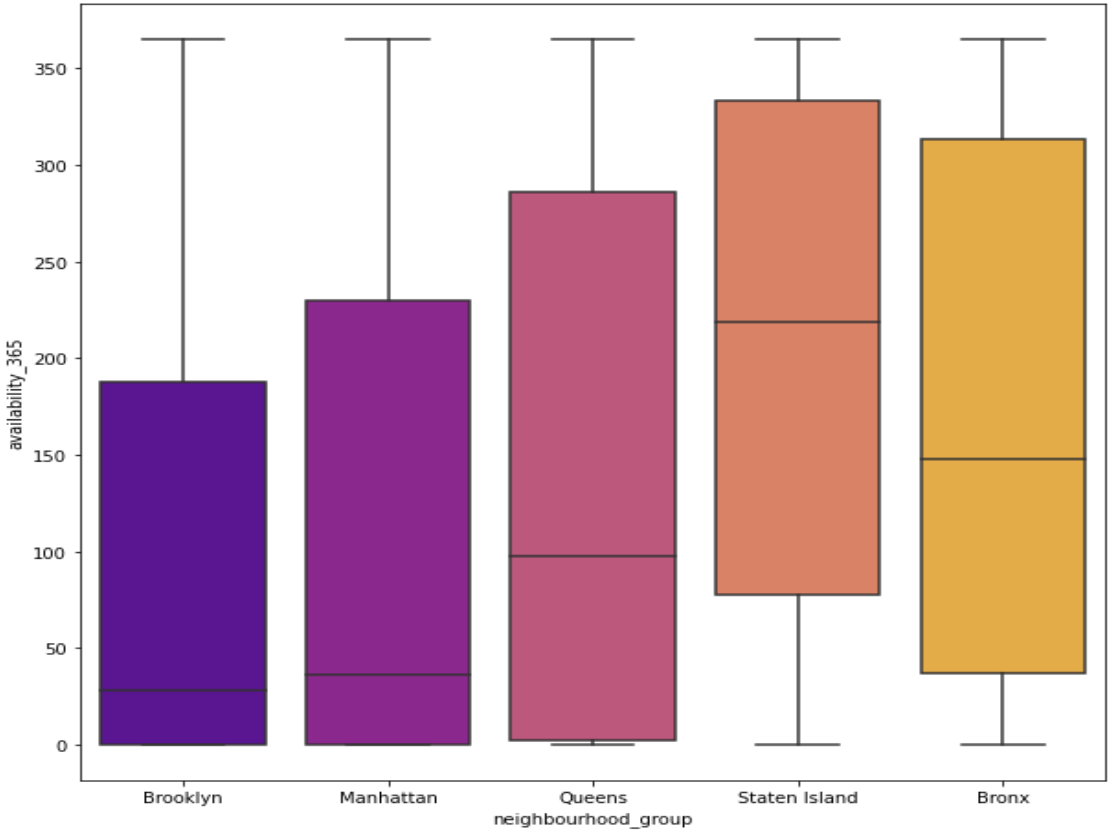
Room Types and Neighbourhood Group

From the Fig. It is clear that the Apartment and Private room data is more than that of Shared rooms. In general, Shared room cost less and can be very useful for Travellers who moves from city to city quite frequently.

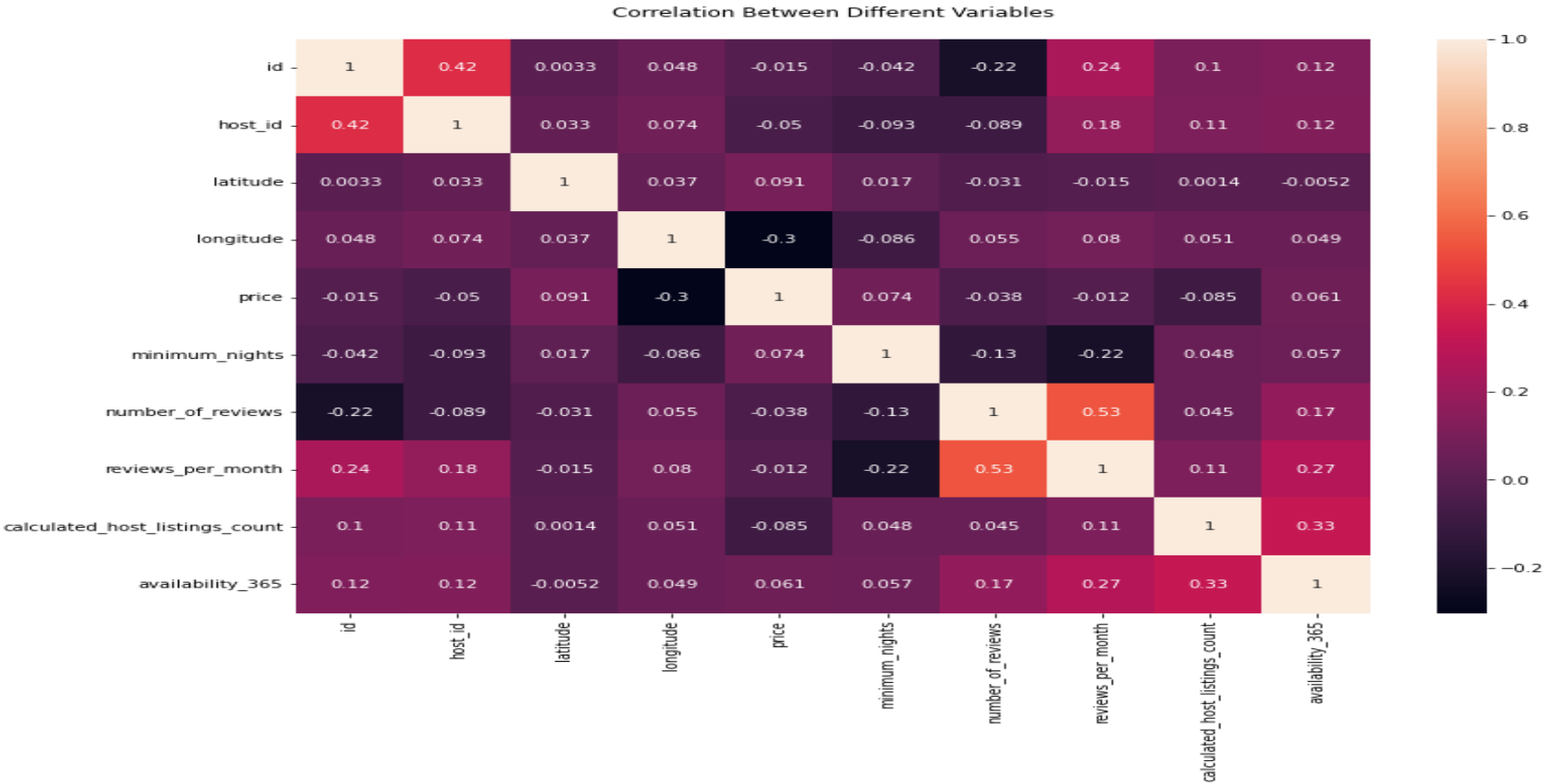


Relation between Neighbour group and Availability of Room

From the fig. we can say that Staten Island have higher number of rooms available for long days, followed by Bronx, Queens, Manhattan, but Brooklyn have very less rooms available as it is the busiest neighborhood.



Heat Map to visualize the correlated variables



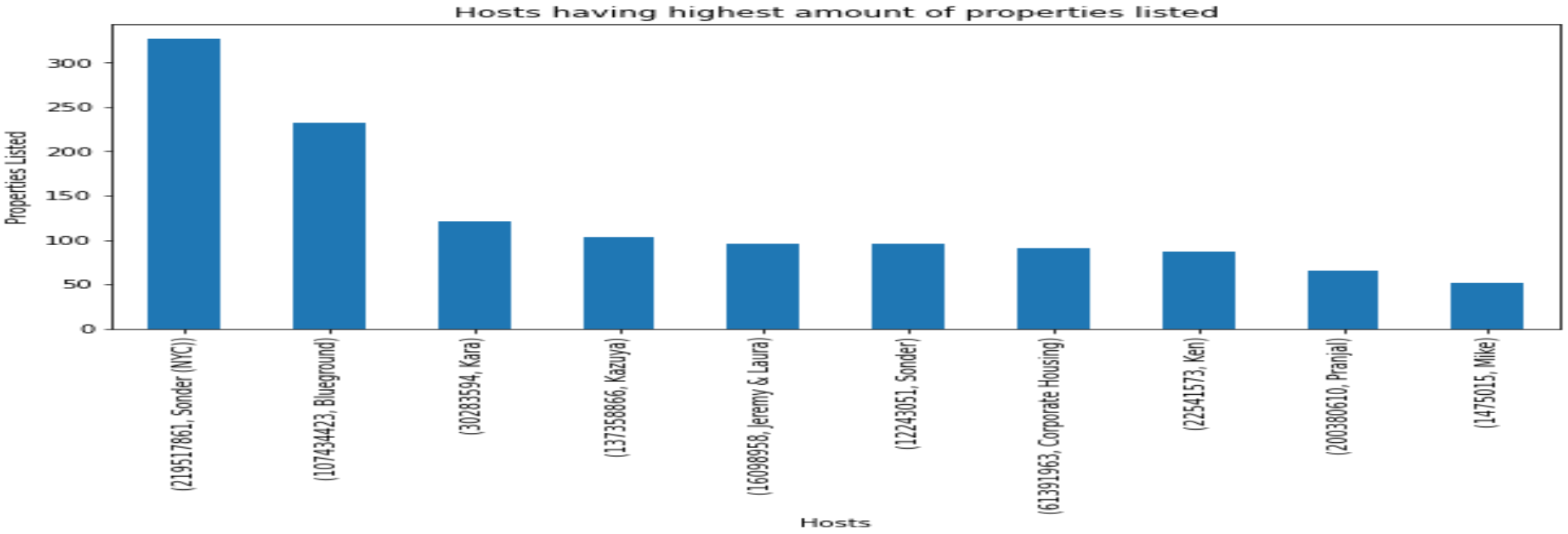
From the graph above, we know that there is not a strong correlation except review_per_month and number_of_review.

what can we learn about different hosts and areas

	host_name	neighbourhood_group	calculated_host_listings_count
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	230
9742	Michael	Manhattan	212
3250	David	Manhattan	202
9741	Michael	Brooklyn	159
6808	John	Manhattan	151
3249	David	Brooklyn	142
7275	Kara	Manhattan	135
432	Alex	Manhattan	134
9856	Mike	Manhattan	134

As we can see most number of listing are from Manhattan created Sonder(NYC),Blueground, Michael, David, John, Kara, Alex and Mike.

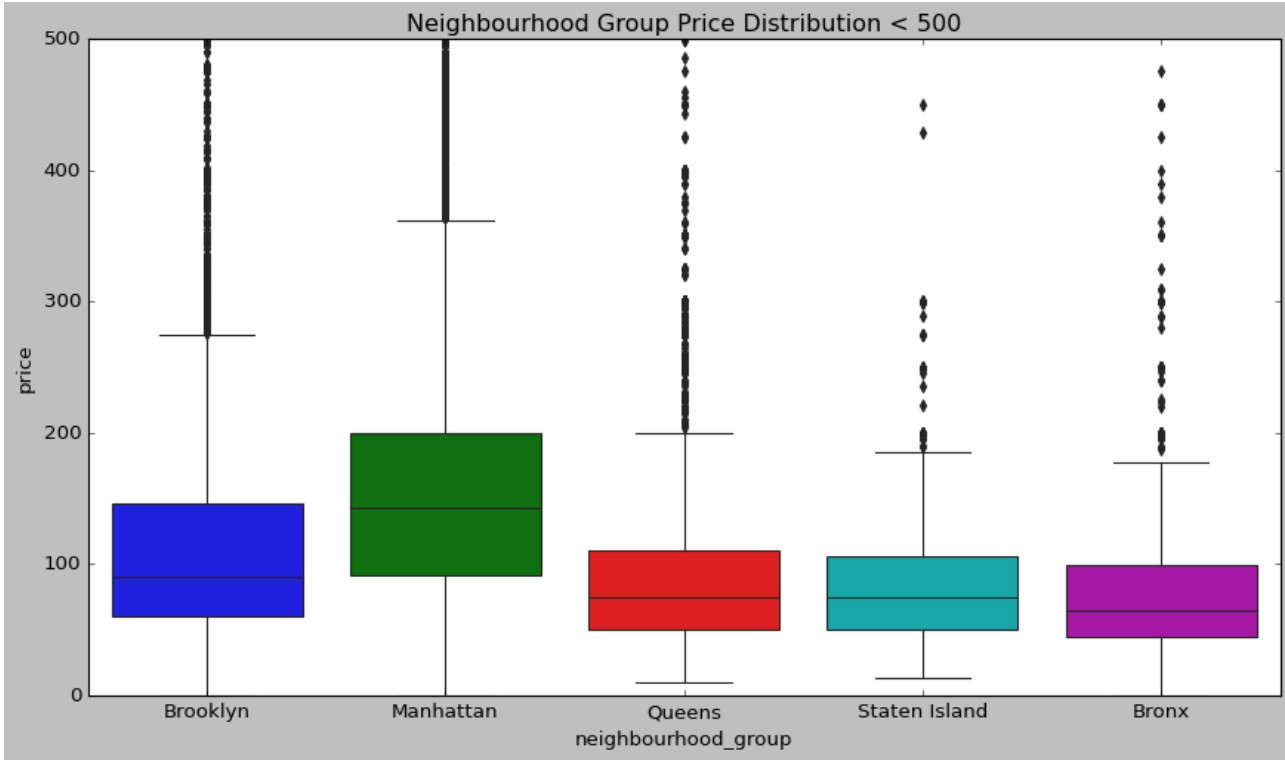
Hosts having highest amount of properties listed



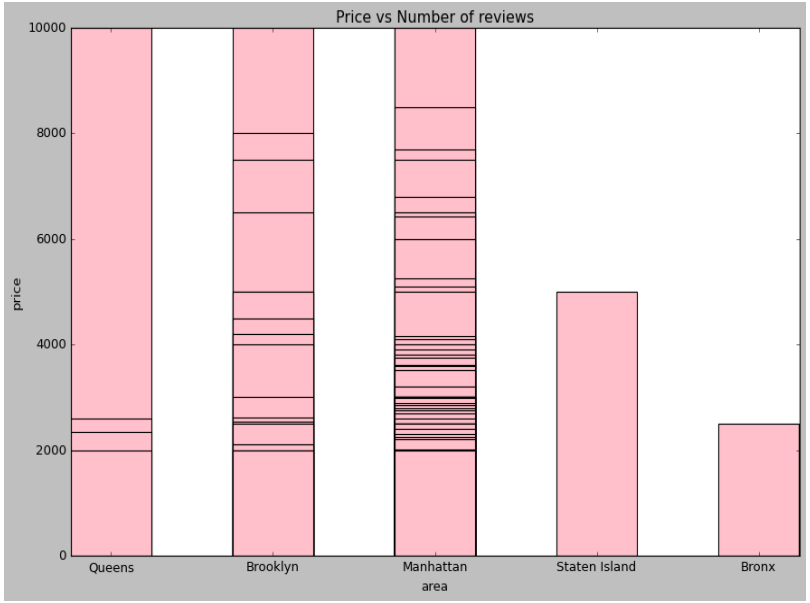
We can see that Sonder(NYC) has the highest number of properties that are listed but his property was not in the top 5 highest reviews table we saw earlier.

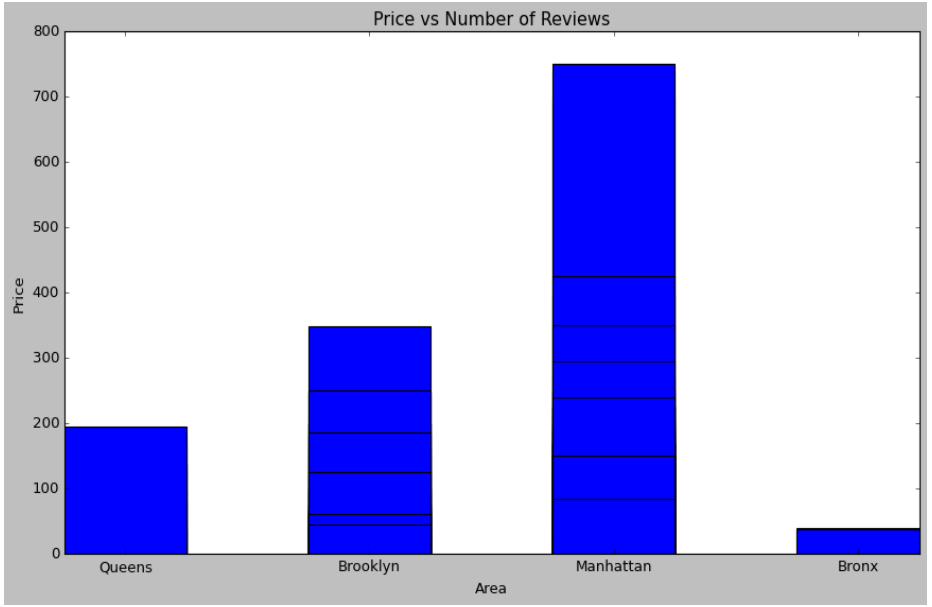
Neighbourhood group with Price

The information we got from the graph above dots are the rooms with a higher price.
Also, we can see that the Manhattan region has more expensive room price. From the boxplot we can observe a couple of things about the distribution of prices for Airbnb booking analysis we can state that Manhattan has the highest range price for the listing with about $ as an average price, followed by Brooklyn with $ 90 per night, Queens and Staten Island seems to have a very similar distribution.
The Bronx is the cheapest.



Number of Reviews





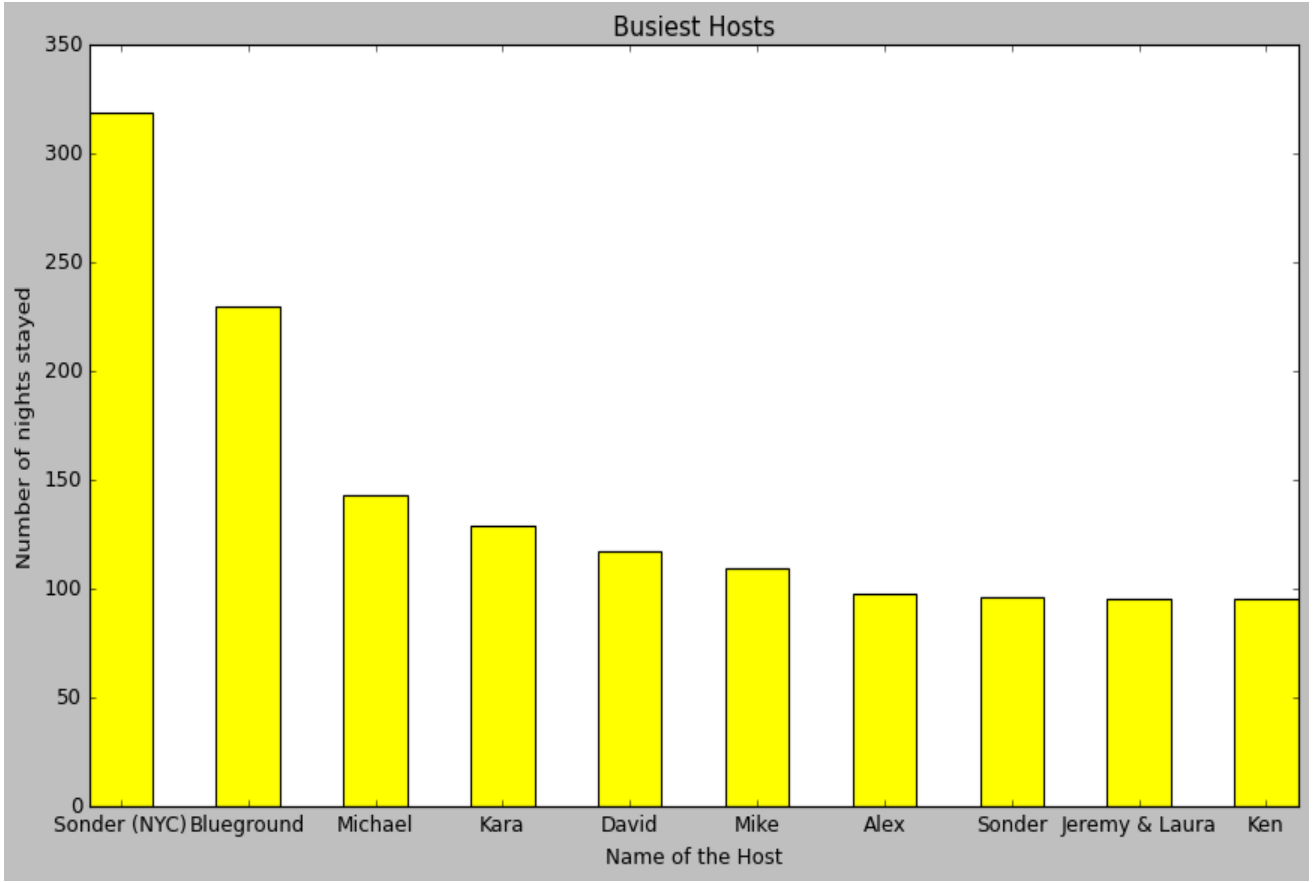
From the fig, we can say that Manhattan has having the highest price for the service but the number of reviews for Manhattan is also high as the Manhattan is the busiest neighborhood in our dataset, followed by the Brooklyn and Queens

The Bronx has the lowest reviews.

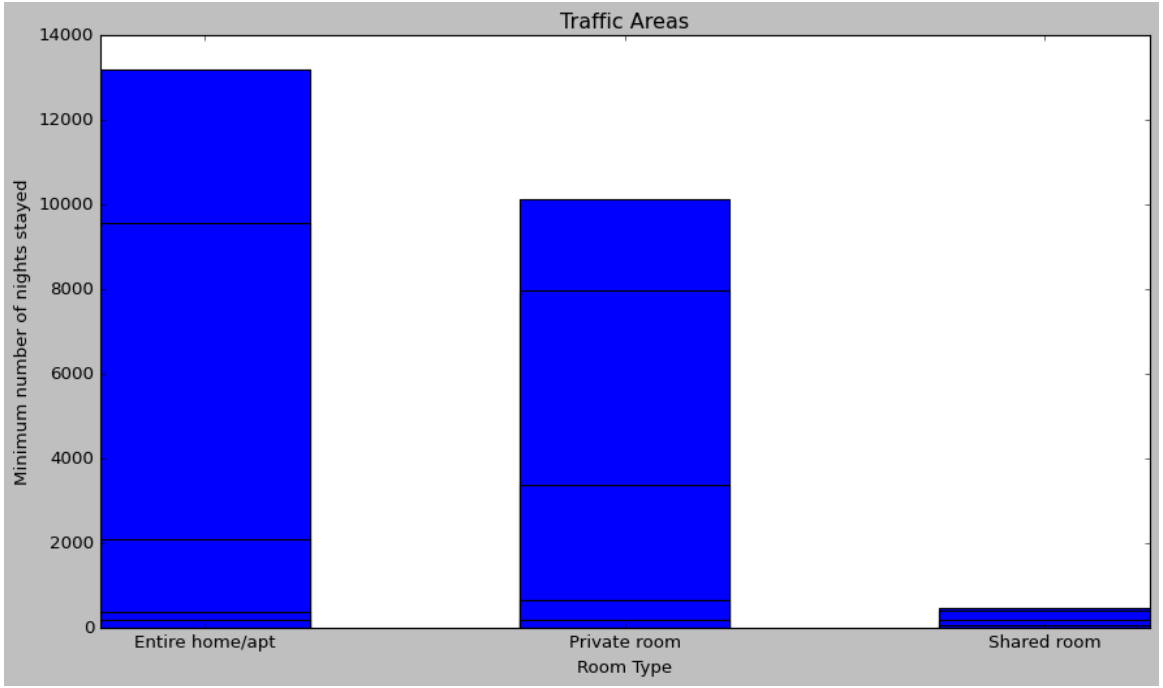
Which hosts are the busiest and why?

Busiest Hosts are
1. Sonder
2. Blueground
3. Michael
4. Kara
5. David

Because these hosts are listing there places in Manhattan which is the most popular place according to our previous analysis and the room type Entire home/apartments which is the most number of people



Is there any noticeable difference of traffic among dif ferent areas and what could be the reason for it



From the analysis we can stay that people are preferring entire home/apartments or private room which are pre sent in Manhattan ,Brooklyn, Queens and people are preferring listings which are less in price.

Conclusion

- From the given Dataset,there are 3 unique room_type and 5 neighbourhood_ group ranking followed by entire apartment > private room > shared room and for neighbourhood_group ranking followed manhattan at top and Staten Islan d at bottom.
- Neighbourhood- Williamsburg has the highest property in brooklyn and also brooklyn has mor number of private rooms unlike manhattan which has entire apartments
- The price of manhattan is more expensive unlike shared rooms are least expensive one.