Capstone Project

Airbnb Booking Analysis

(Shubham bhadouria)

WHAT IS AIRBNB

Airbnb began in 2008. It is an online platform that connects people all around the world for unique travel experiences. Initially, it is a platform for people to list their homes, but now it offers variety of tourism experiences within its platform. Home and room rentals are still the main business of Airbnb.

Dataset Description

The data set of Airbnb New York has 488950 rows and 16 columns

The columns are

1. 1. Id 2. Name 3. Host id 4. Host name 5. Neighbourhood group
   6. Neighbourhood 7. Latitude 8. Longitude 9. Room type
   10. Price 11. Minimum nights 12. Number of reviews 13. Last review
   14. reviews per month 15. calculated host listing count 16. Availability 365
2. 6

Objective of The Project

To find the top three hosts from each areas

To gather the maximum information of each neighbourhood group(ex: locations, prices, reviews, etc) from the dataset.

To find the top ten busiest hosts of Airbnb dataset.

To find the traffic among different areas of New York and reason for it?

Road Map for Data Analysis

Data Cleaning

- Dropping all the column which was not significant for doing data analysis. last_review and id columns are dropped as these are not significant for data analysis.
- The data has null values. last_review and reviews_per_month columns contain 10052 null values. Removing the null values from the column which is necessary for the observation.
- Data with zero null values gives accurate measures and help data analysts to give right suggestion to Airbnb

DATA TRANSFORMATION & ANALYSIS

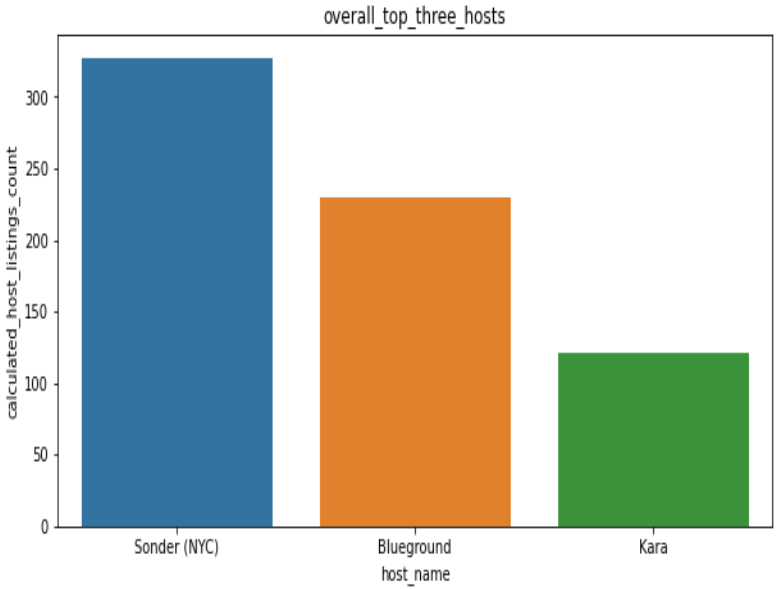
Reviews per month attribute has 10052 null values which are replaced with 0 assuming these hosts got zero reviews from the customers.

Price attribute has the zero values for many rows which is not possible and hence we replaced it with median value of price attribute.

All the columns are analysed deeply to gather information regarding the hosts and customers of New York.

Data Visualization

Plotting For Host And count Of calculated host listing count

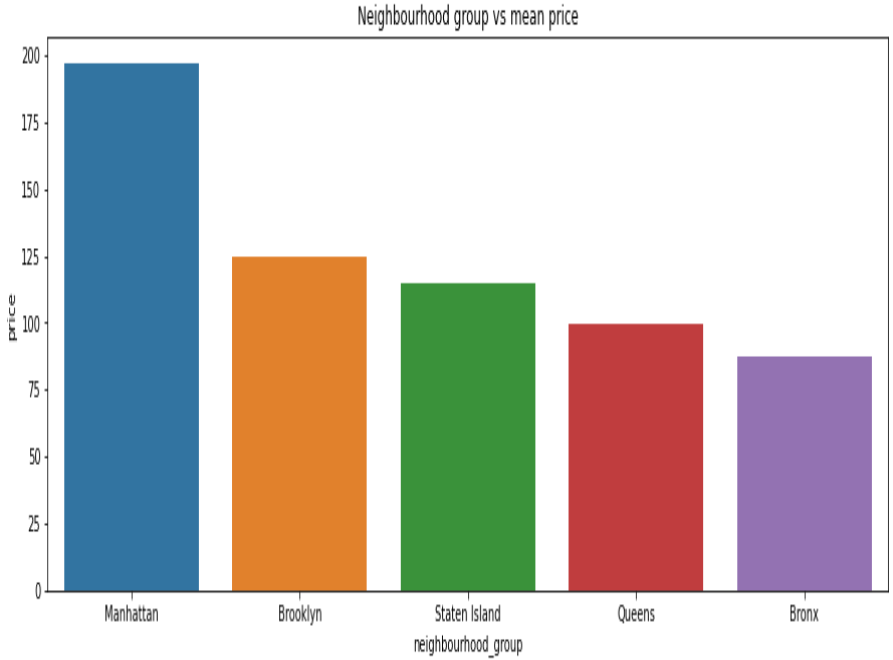


Conclusion

Sondar(NYC) is top host and has the maximum listing. The reason behind this may be the location of sondar hotels as these hotels lies in center of the city.

Data Visualization

Plot for finding the expensive Neighbourhood Group

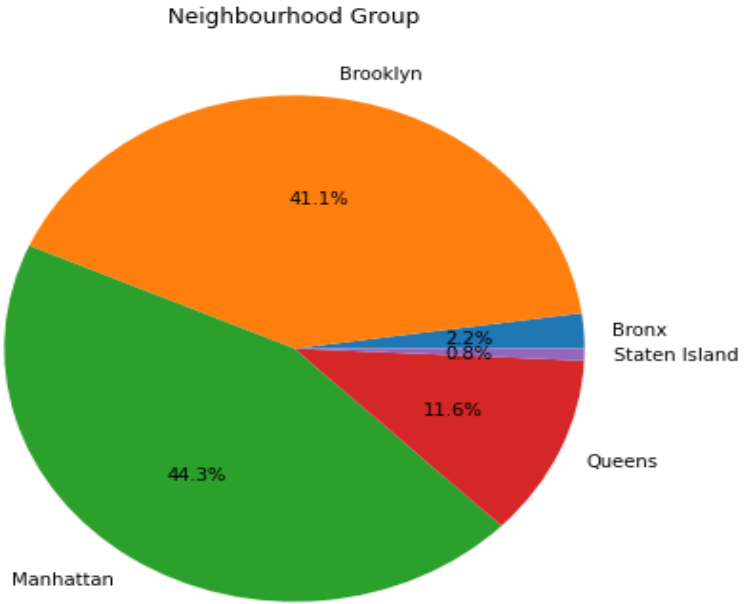


Conclusion

Manhattan is the most expensive neighbourhood group. The reason may be its location as it's situated at mid of the country so more denser and also have important financial status, so movement of financial act is happened here.

Data Visualization

Plot showing the Neighbourhood Group with the number of listing on Airbnb

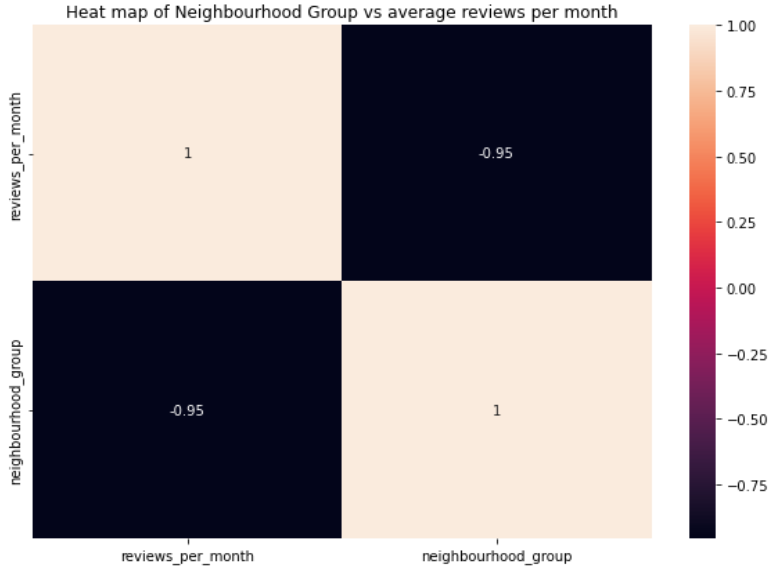


Conclusion

Plot shows that Manhattan has most share of property(44.3%) listing followed by Brooklyn(41.1%)

Data Visualization

Plot to fund The Correlation Between Neighbourhood group And Review Per Month

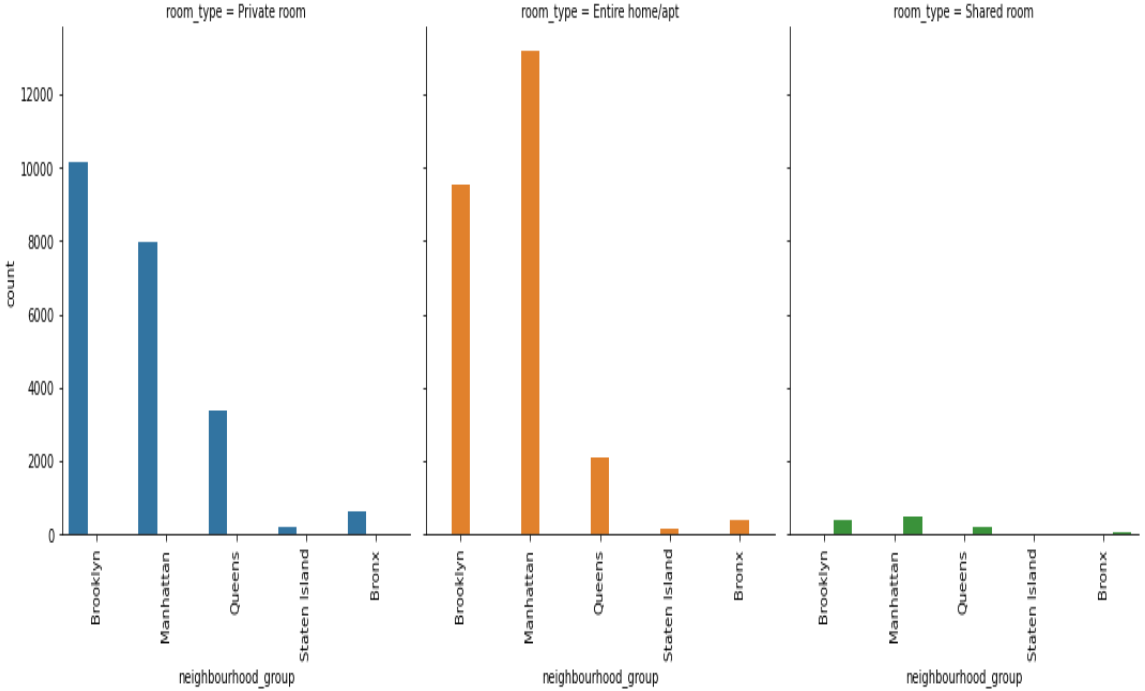


Conclusion

Plot shows that neighbourhood group and reviews per month are in a negative correlation with each other. That's why Manhattan have the lowest review even it's is more demanding

Data Visualization

Plot showing the count of different types of rooms listed in different neighbourhood groups

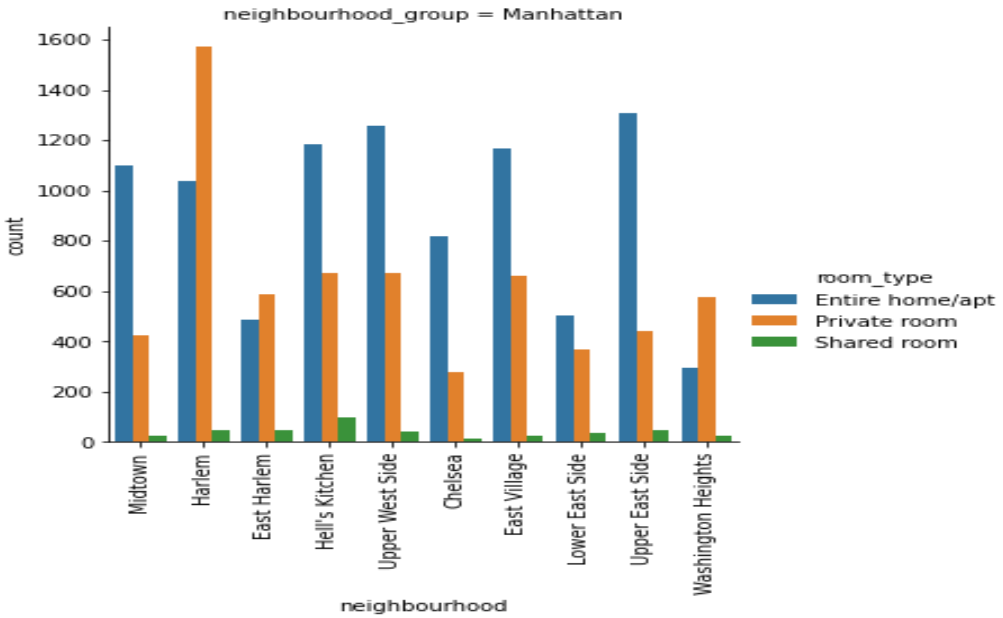


Conclusion

In Manhattan, people prefer entire room/apartment. In Brooklyn, people prefer either private room or entire home. People neglect the shared rooms. People are more concerned about privacy.

Data Visualization

Plot For Finding The top neighbourhood inside Manhattan



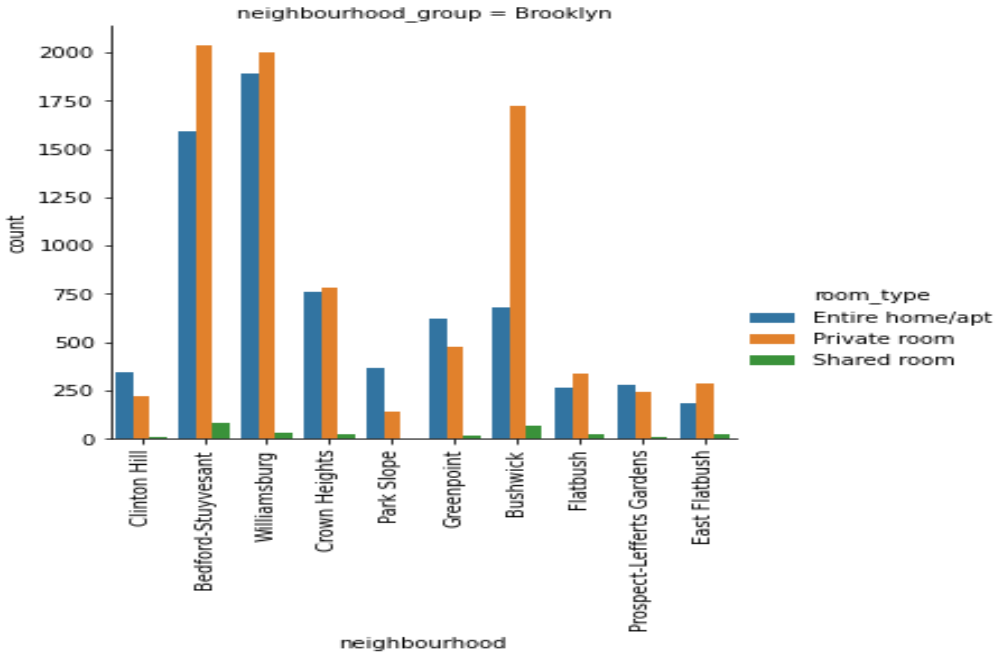
Conclusion

The bar plot between different neighbourhood of Manhattan based on the total count of room listing.

Harlem has the most room listing count and Washington heights has least room listing count

Data Visualization

- **Plot For Finding The top neighbourhood inside Brooklyn**

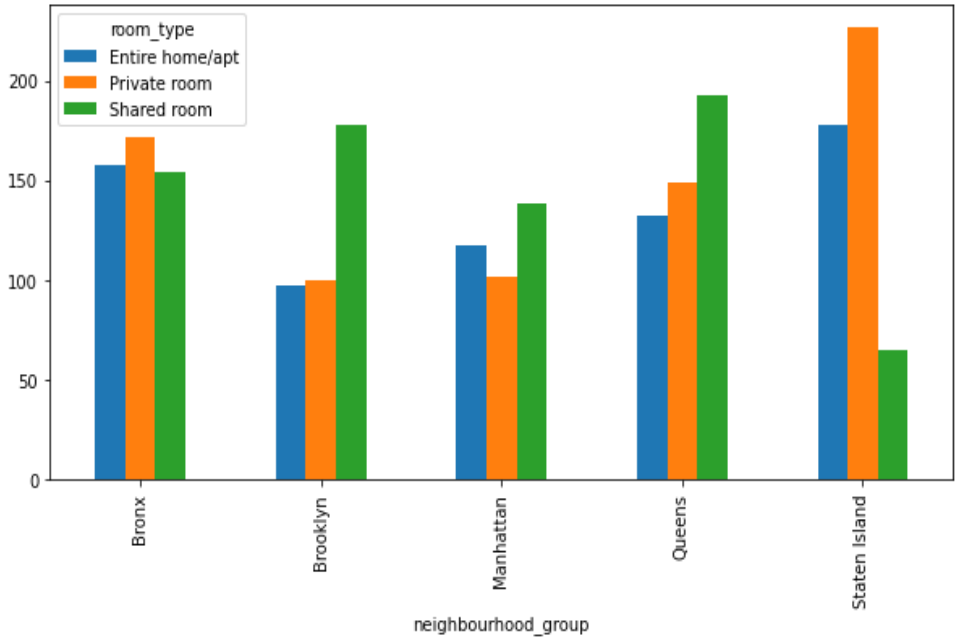


Conclusion

The bar plot between different neighbourhood of Manhattan based on the total count of room listing. Bedford Stuyvesant and Williamsburg has the most room listing count and East Flatbush has least room listing count

Data Visualization

- **Plot showing availability of different types of rooms among different neighbourhood groups**



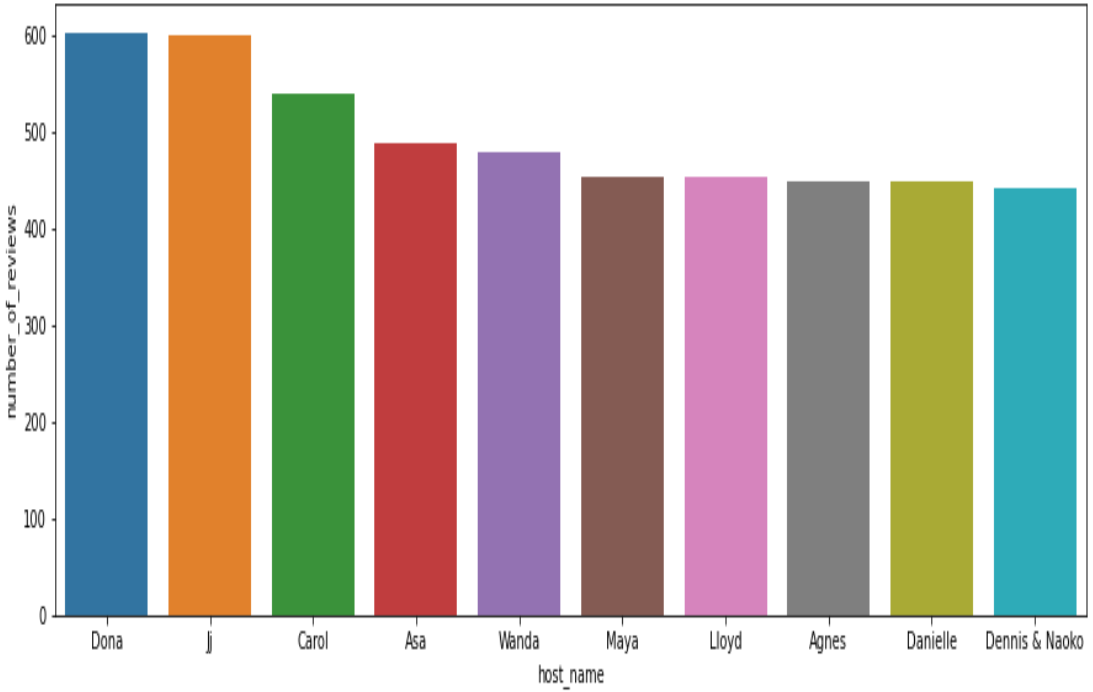
Conclusion

The bar plot between different neighbourhood group based on the total count of availability through out the year.

Room availability in Staten Island is highest among all the groups because of its distance from the rest of the city.

Data Visualization

- **Plot For Finding the busiest host of Airbnb dataset**

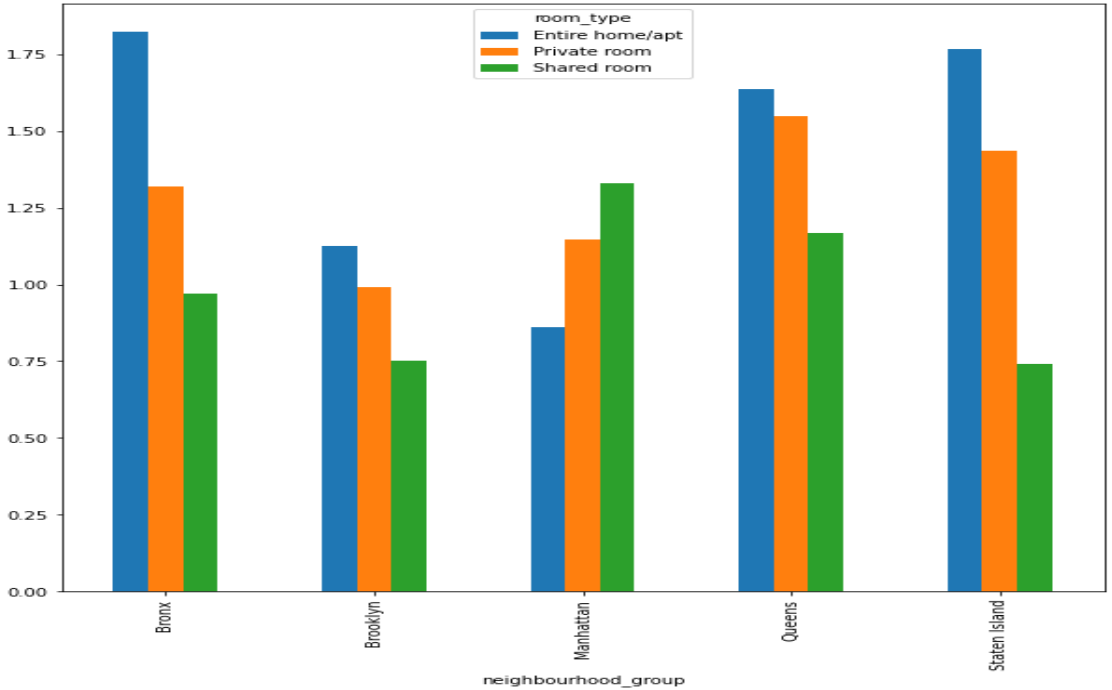


Conclusion

The bar plot between different hosts irrespective of the neighbourhood group based on the average reviews per month.
Among the top ten busiest hosts, most of the hosts are from Queens because Queens has a significantly lower cost of living.

Data Visualization

- **Plot For Finding Traffic among all the neighbourhood groups**



Conclusion

The bar plot between different neighbourhood groups and average reviews per month.

Room type: Entire home/apt of Bronx has the highest traffic among all.

Room type: Private rooms of Queens has the highest traffic among all the groups.

Room type: Shared rooms of Manhattan has the highest traffic among all groups

Challenges Faced

- **Since the data was huge so it was very challenging to completely understand the data.**
- **Dealing the missing data and outliners values is a very tricky task.**
- **It took a lot of efforts to find out the insignificant data and eliminates them so that a relevant dataset can be created for analysis.**

Future Scope of Work

There is always a better solution to the problems in future than the current one.

- **Future analyst can find the relation between the minimum number of night stayed by the customer in different room types of every neighbourhood group.**
- **Future analyst can find top ten neighbourhood of each neighbourhood group demanding on average price of room.**

.

Thank You