Capstone Project- 1

EDA on Airbnb booking

by-
Team Achivers:
Abhishek Tiwari
Abhishek Luhera

❑ Overview of Air booking Analysis:

- ➢ **Introduction**
- ➢ **Abstract**
- ➢ **Problem Statements**
- ➢ **Steps Involved**
- ➢ **Exploratory Data Analysis**
- ➢ **Conclusion**

❑ Introduction:

➢ **Airbnb is an online marketplace for lodging, principally homestays for vacation rentals, and tourism activities based in the United States. Airbnb does not own any of the houses advertised; instead, it makes money from commissions on each booking.**

➢ **Airbnb is a database of accommodation bookings.**

➢ **We can grasp all of the features by exploring the data. The objective is to look into the data and then evaluate it with all of the necessary information.**

➢ **We have extracted crucial insights from the data through data exploration and data analysis.**

➢ **Airbnb is an abbreviated version of AirBedandBreakfast.com, which was the company's**

AI

❑ Abstract

➢ We have extensive experience with Airbnb Booking Analysis.

➢ We intend to concentrate solely on New York City.

➢ The investigation is mostly focused on the boroughs of Brooklyn, Manhattan, Queens, Staten Island, and the Bronx.

➢ This data includes information about the host, lodging prices, and reviews, among other things.

➢ With the use of exploratory data analysis, data wrangling, visualization, and other tools, we can extract a lot of information from the analysis.

❑ Problem Statement:

➢ **The customer can book a room through the AirBedandBreakfast.com (Airbnb website). To explore more, the customer is moving to the Queens, Bronx, Staten Island, Brooklyn, and Manhattan neighbourhoods.**

➢ **The customer requires assistance in identifying a number of bookings in each borough country and guiding that customer in selecting the best place to live at an affordable cost, analysing the room's price so that the customer can afford it, and selecting the appropriate room type based on their needs.**

➢ **The quantity of reviews, the number of reviews per month, and the availability of the rooms will all meet these criteria.**

➢ **All of these insights will be obtained through EDA and a thorough comprehension of the data**

❑ Steps Involved:

- ➢ **Data Wrangling**

- ➢ **Checked Correlation Between Each Features**

- ➢ **Treatment of Null Values**

- ➢ **Explored Visualization**

- ➢ **Transformation**

- ➢ **Conclusion**

❑ **Understanding the Data**

- There are 49,000 observations with various types of field in our dataset.
- List of field:

■ **Id**	■ **Price**
■ **Name**	■ **Minimum_nights**
■ **Host_id**	■ **Number_of_reviews**
■ **Host_name**	■ **Last_review**
■ **Neighbourhood_group**	■ **Reviews_per_month**
■ **Neighbourhood**	■**Calculated_host_listing_count**
■ **Latitude**	■ **availabilty_365**
■ **Longitude**	
■ **Room_type**	

AI

❑ Cheking the count of null values:
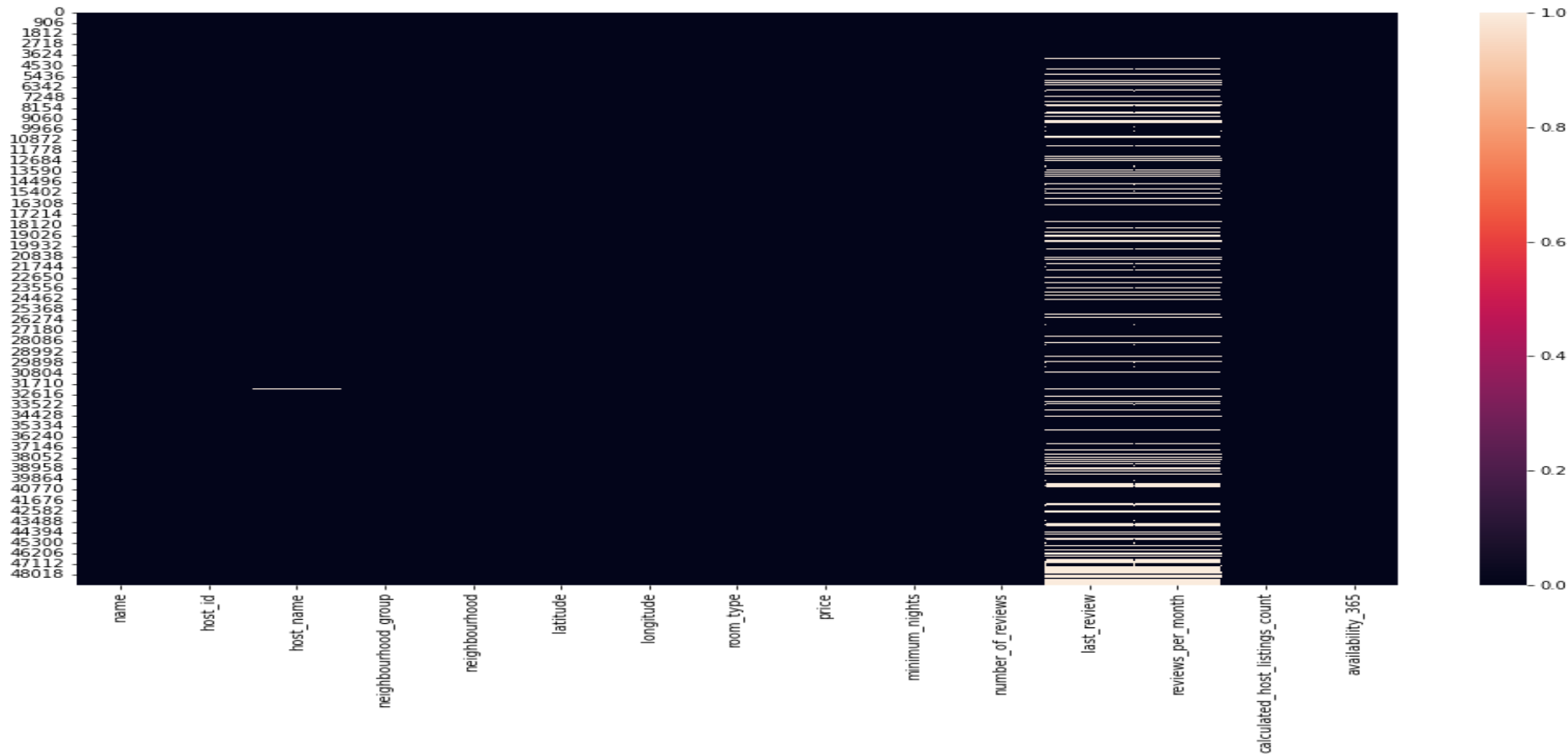
Correlation plot

- **Top 3 correlation** columns_

- *The number_of_reviews and reviews per_month columns has highest correlation of values is* **0.55**

- *The calculated_host_listings_count and availability_365 columns second highest correlation of values is* **0.23**

- *The reviews_per_month and availability_365 columns have correlation of values is* **0.19**

❑ **AGENDA**

➜ **We try to answer following questions for Airbnb:**

➢ **WHAT ARE THE ROOM TYPES AVAILABLE ?**

➢ **TOP 15 HIGHEST LISTING NEIGHBOURHOOD ?**

➢ **HOW MANY NUMBERS OF HOSTS WHICH HAS AVAILABILITY OF 365 DAYS I.E. 24X7 ?**

➢ **WHAT CAN WE LEARN FROM PREDICTIONS? (EX: LOCATIONS, PRICES, REVIEWS, ETC.)?**

➢ **WHAT IS THE AVERAGE PREFERRED PRICE BY CUSTOMERS ACCORDING TO THE LOCATION?**

❑ AGENDA (cont.…)

➢ **WHICH HOSTS ARE THE BUSIEST AND WHY ?**

➢ **WHO IS THE TOP 5 NEIGHBOURHOOD ON THE BASES OF REVIEWS ?**

➢ **WHICH TYPE OF ROOM HAVE MAXIMUM BOOKINGS ?**

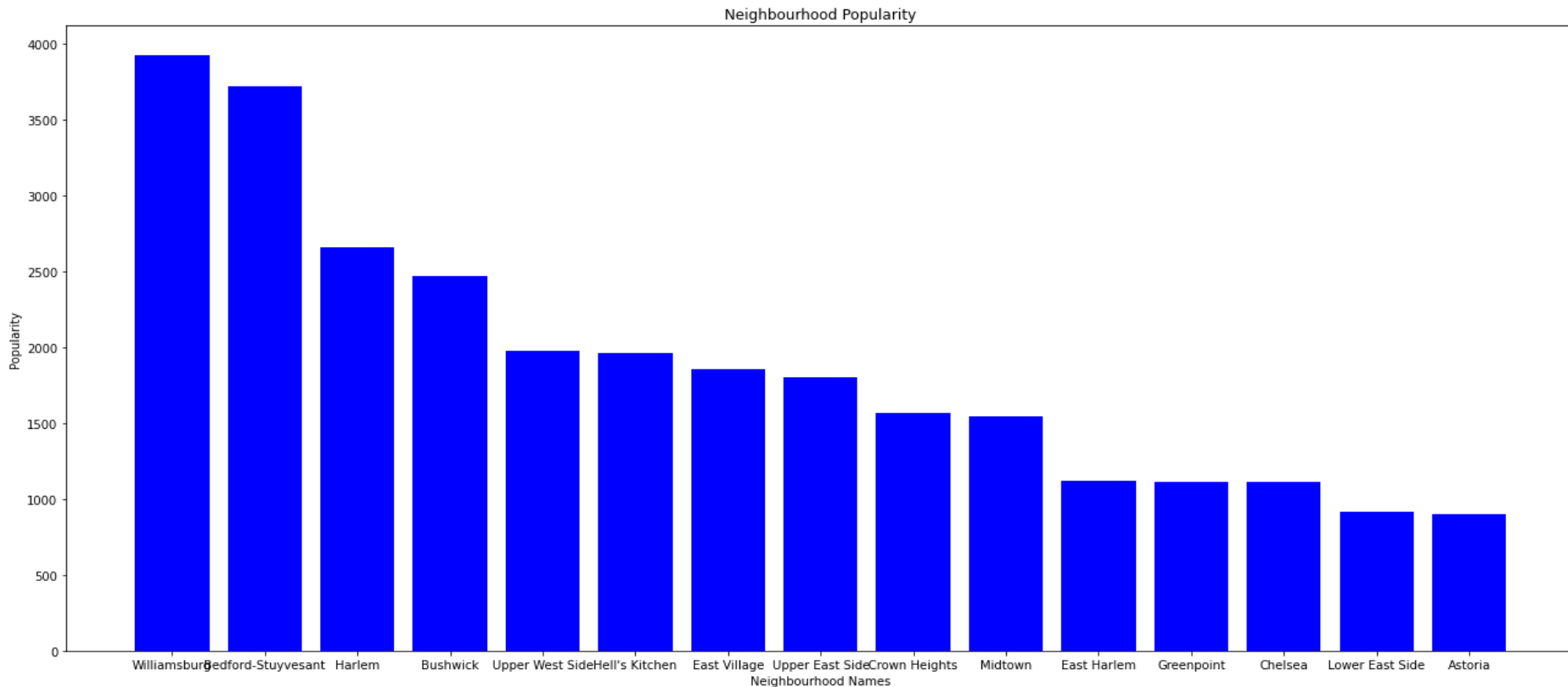➢ **WHICH TYPE OF ROOMS ARE MORE EXPENSIVES ?**

❑ Total count of each room type as per listing.

➢ **Based on the Analysis we found that:**

➔ **Entire home/apt** has the highest number of listing of 25409 among other room types.

➔ **Private room** has 22326 of listing among other room types.

➔ **Shared Room** is the least listed room type at only 1160 in total.
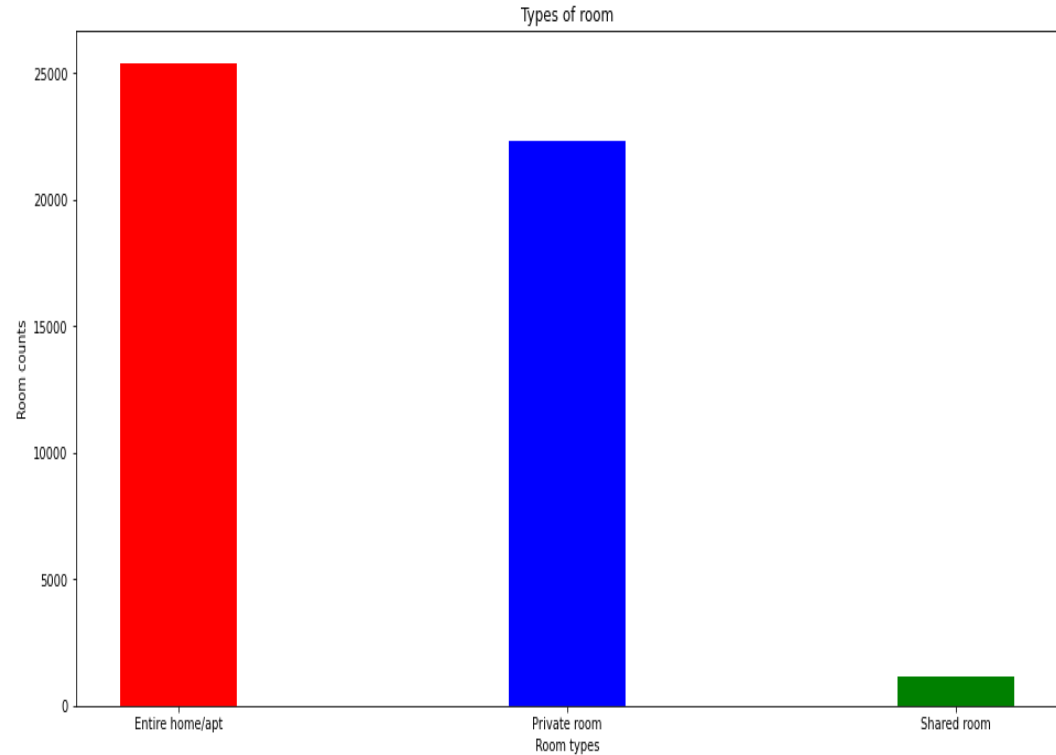
❑ TOP 15 MOST POPULAR LISTING NEIGHBOURHOOD ?

❑ Top 15 most popular listing neighbourhood (cont.…)

➢ Here we can state that Williamsburg has the highest listing.

➢ And almost same in Bedford-Stuyvesant.

➢ It's a tourist attraction hence the number of listings could be higher

❖ **Number of active hosts per location (Where most of the hosts interested to own property?)**

- **Manhattan and Brooklyn is the most preferred place by hosts.**
- **Bronx and Staten Island is low traffic for host as compared to other location**
- **Queens has a average number of host.**

Number of host per location
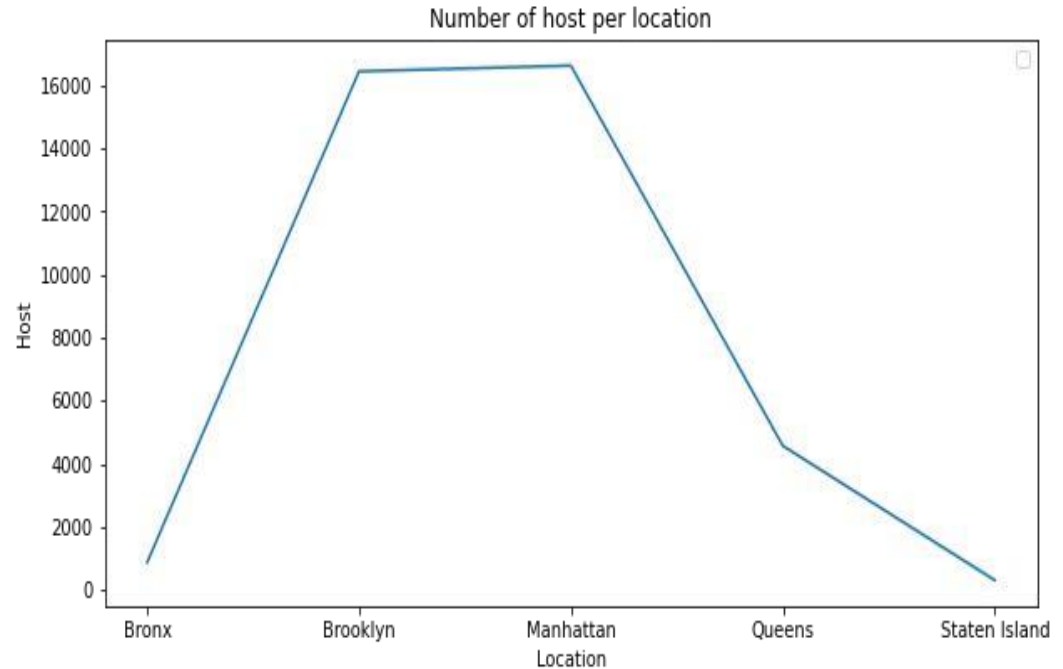
What can we learn from predictions? (ex: locations, p reviews, etc)?

Area vs Number of reviews
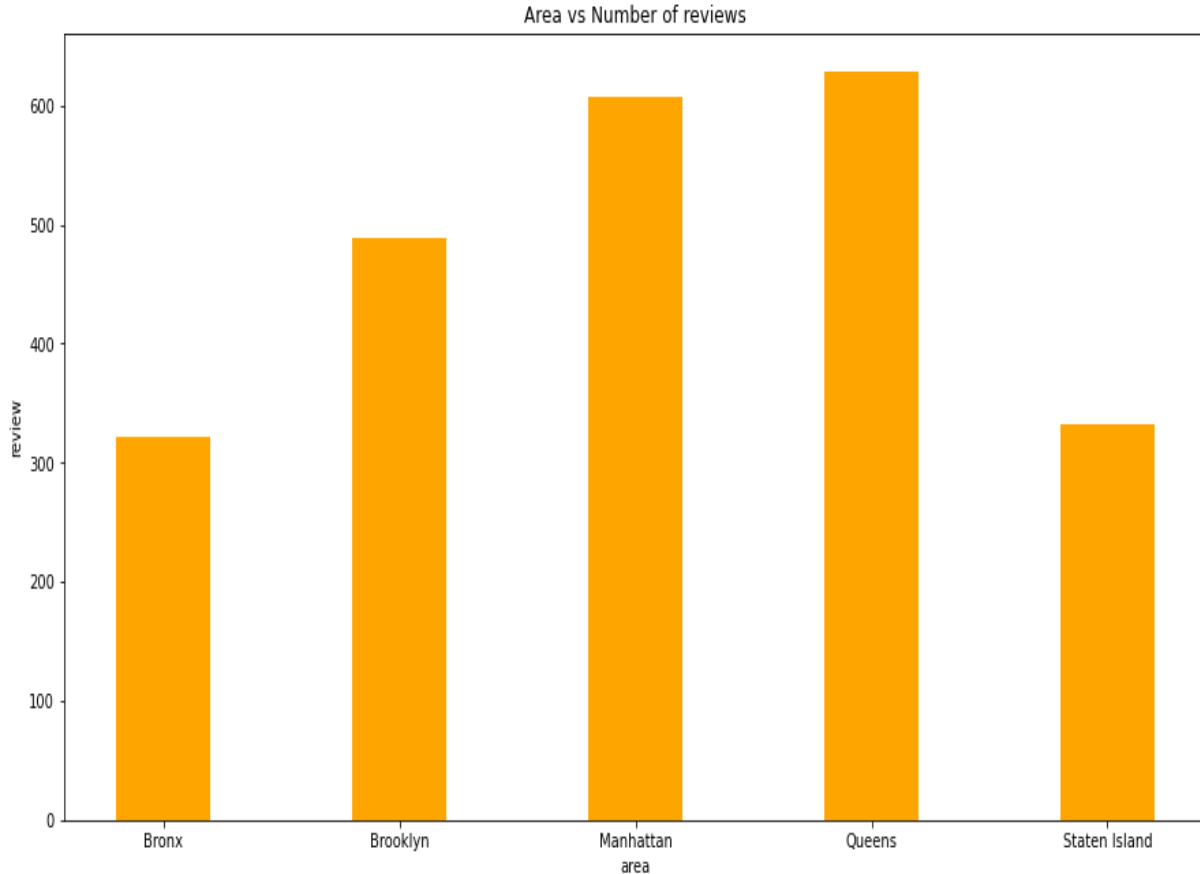
❖ **It is clear with the Area vs Numbers of Reviews graph that queen has the highest reviews according area wise with 629 counts**

❖ **Manhattan has the 2nd most numbers of reviews with 607 counts.**

❖ **What is the average preferred price by customers according to the location?**

- **As we can see in Manhattan average price of Entire home/apt is highest, compared to other location.**

- **Average price of shared room is lowest in brooklyn.**

- **Queens, Staten Island and Bronx share almost the same price for Shared room and Private room.**
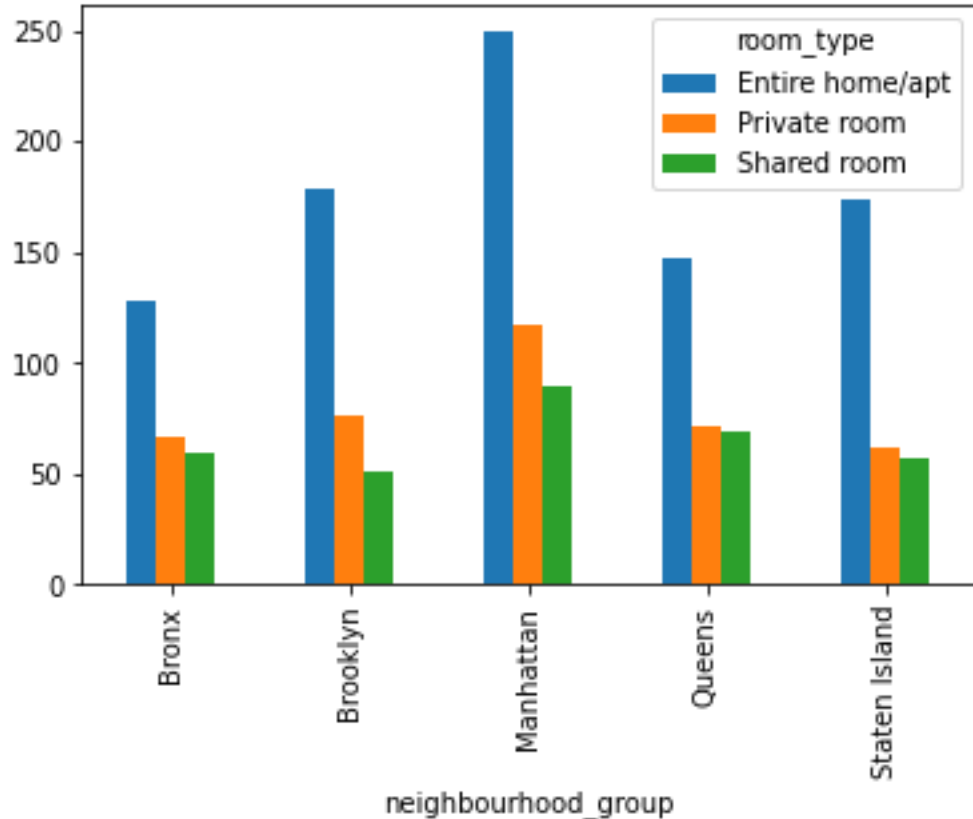
❖ **Which host are the busiest?**

- **As we see in the analyses Dona has the highest number of reviews among the top 10 with 629 counts.**
- **J was at 2ⁿᵈ position with 607 counts.**
- **Maya and Carol at 3ʳᵈ and 4ᵗʰ position with almost same number of counts 543 and 540.**

Busiest hosts
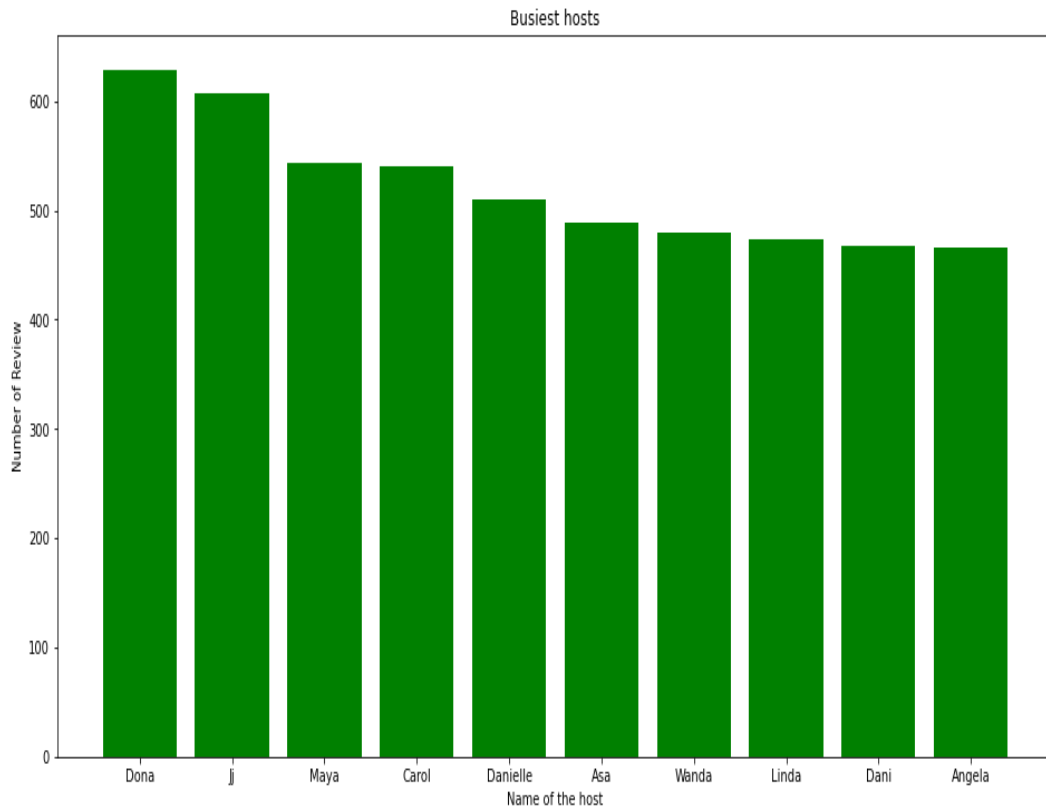
❖ Who is the top 5 neighborhood on the bases of reviews?

Distribution of reviews in the top neighbourhoods
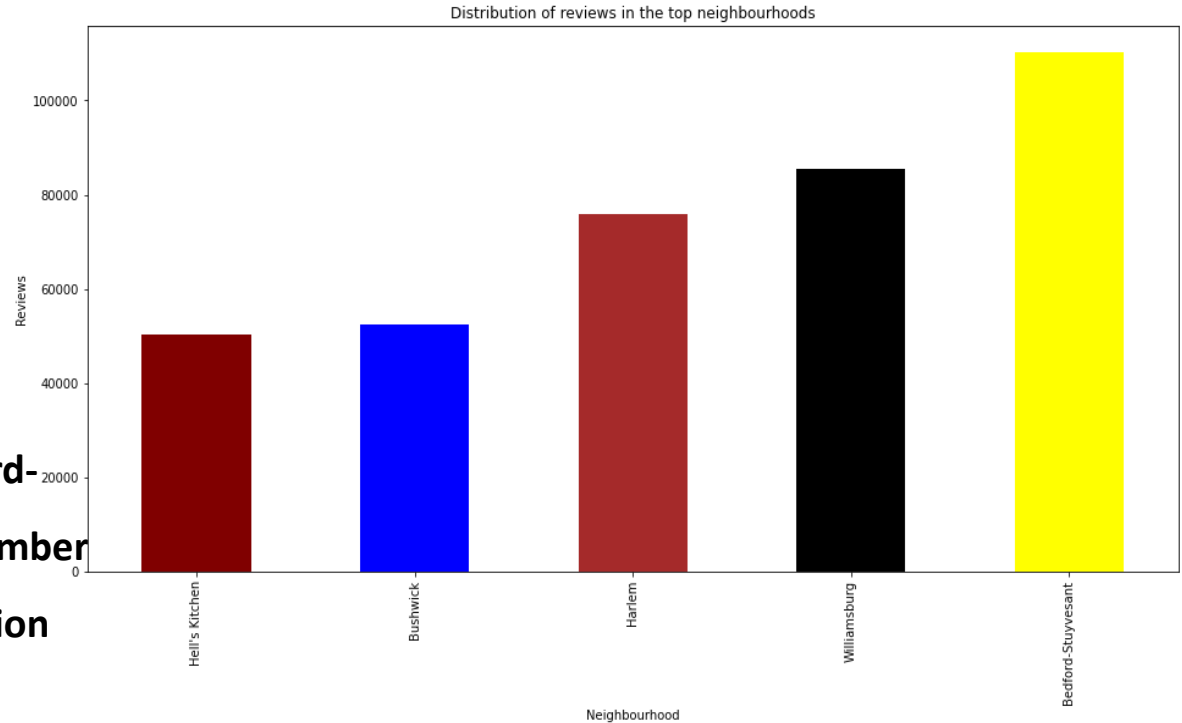
❖ **According to the graph Bedford-stuyvesant has the highest number of reviews and at the 1st position among other neighborhood.**

❑ **Which type of rooms have maximum bookings or more popular?**

➢ **Entire home/apt is the first choice of tourist with maximum bookings.**

➢ **Private room is the second choice.**

➢ **Shared room is least preferred by tourist.**

Popular rooms type

Entire home/apt

Number of entries

Shared room

Private room
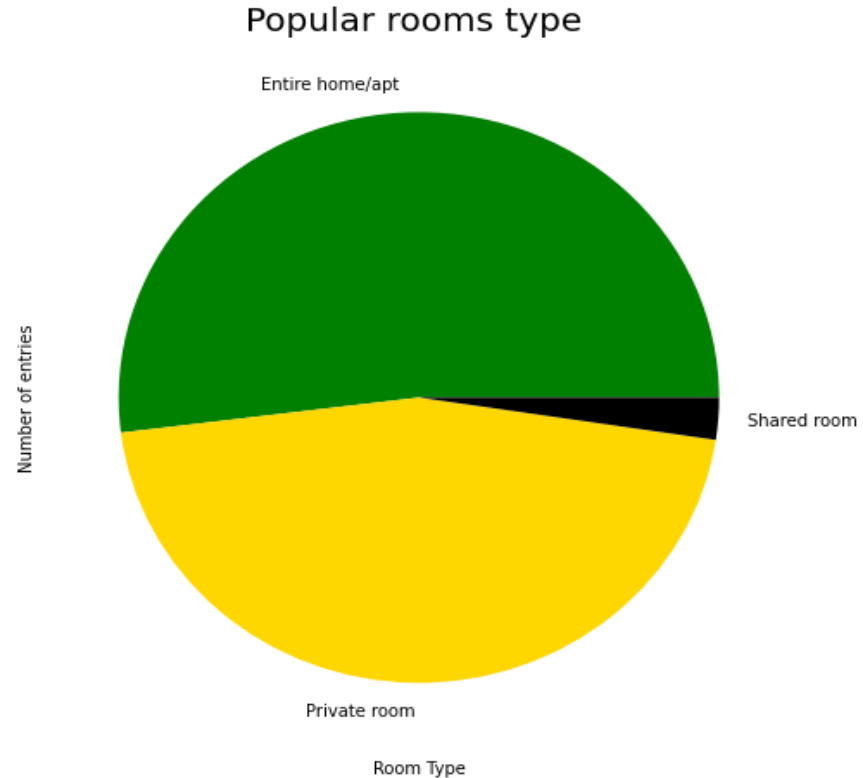
Room Type

❑ Which type of rooms are most expensive?

➤ **Entire home/apt was the most expensive room type with 211.8 mean price among the other listing.**

Cost per room type
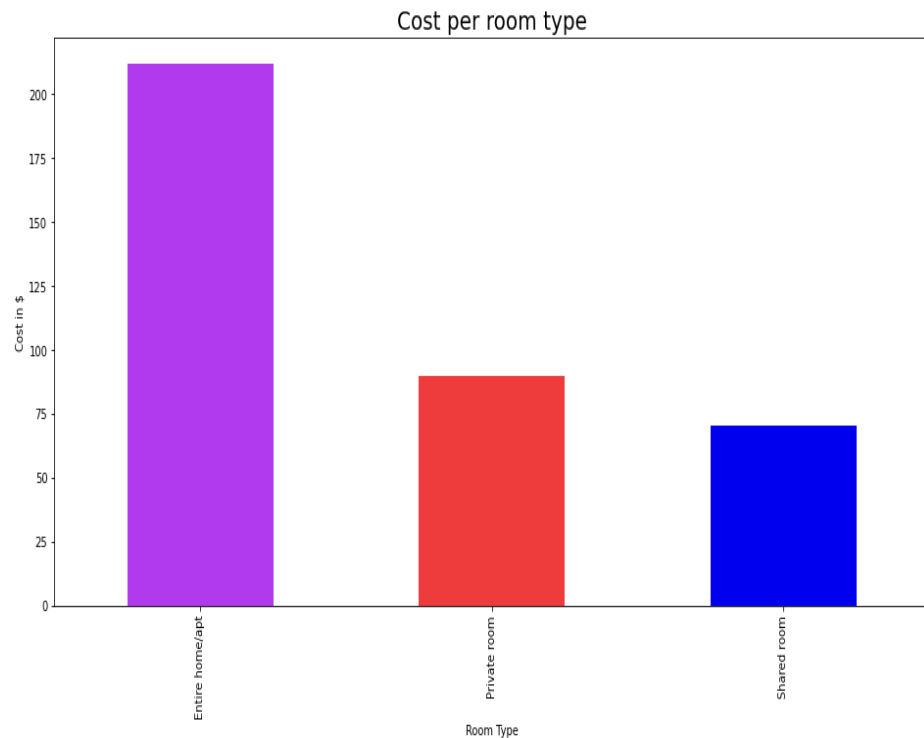
The most expensive room type is "Entire home/apt" with 211.8 mean price

Challenges Faced

- **Reading the dataset and understanding the meaning of some columns.**
- **For answering some of the questions we had to understand the business model of airbnb that how they work.**
- **Handling NaN values,null values and duplicates.**
- **Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.**

❑ **Conclusion:**

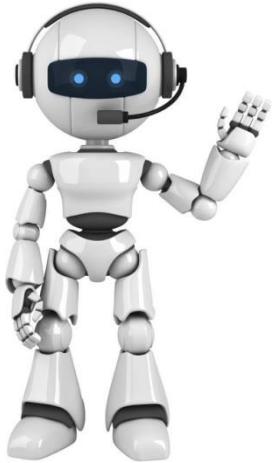➜ **We defined some points which can help airbnb in their business:**

- **Manhattan is the most focused place in New York for hosts to do their business**
- **Customers pay highest amount in Brooklyn,Queens and Manhattan that is $10,000 and lowest amount is $10.**
- **For the three types of room type (i.e. Entire home, Shared room, & Private room) average price of entire home is around $157, for Shared room is around $60, and for private room is around $75.**
- **Top three host base on their turnover are Sonder(nyc),Red awning, Henry and best host is Sonder(nyc)**

❏ Conclusion (contd.):

- **'Entire home/apt' room type has the highest number of listing of 52% and 'Shared Room' is the least listed room type at only 2.4% in total.**

- **People stay for longer duration of time in Private rooms in Brooklyn and Manhattan.**

- **There are totally 894 hosts which has availability of 365 days**

- **Count of listing by top 10 hosts is almost 2.5% (1270 listings) of the whole dataset.**

- **More customer preferred Manhattan location for night stay than Brooklyn**

- **63.2% customer spend night in Entire home and 1.6% spend night in Shared room**