Capstone Project- 1

EDA on Airbnb booking

❖ Problem Statement

❖ **For this project we are analyzing Airbnb's New York City(NYC) data of 2019. NYC is not only the most famous city in the world but also top global destination for visitors drawn to its museums, entertainment, restaurants and commerce.**

❖ **Our main objective is to find out the key metrics that influence the listing of properties on the platform. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis (EDA) techniques.**

❖ **Data analysis on thousands of listings provided through Airbnb is a crucial factor for the company.**

❖ **We will be finding out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors.**

❖ **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	

❖ Agenda

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

- What is the average preferred price by customers according to the location?
- Number of active hosts per location (Where most of the hosts focused to own property?)
- Where the customer pays the highest and lowest rent according to location?
- Most popular/demanded host of Airbnb in New York
- Find the total count of each room type

Agenda(Cont...)

- Room types and their relation with availability in different neighbourhood groups?

- Which are the top 25 most used words in listing names?

- Find top 10 hosts with most listings

- Find the top three hosts based on their turnover

- Find total no. of nights spend per location

- Total no. of nights spends per room types

- Top 10 highest listing neighborhood

Map of New York City(NYC)

❖ 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 i.e $50 for Shared room and Private room.**
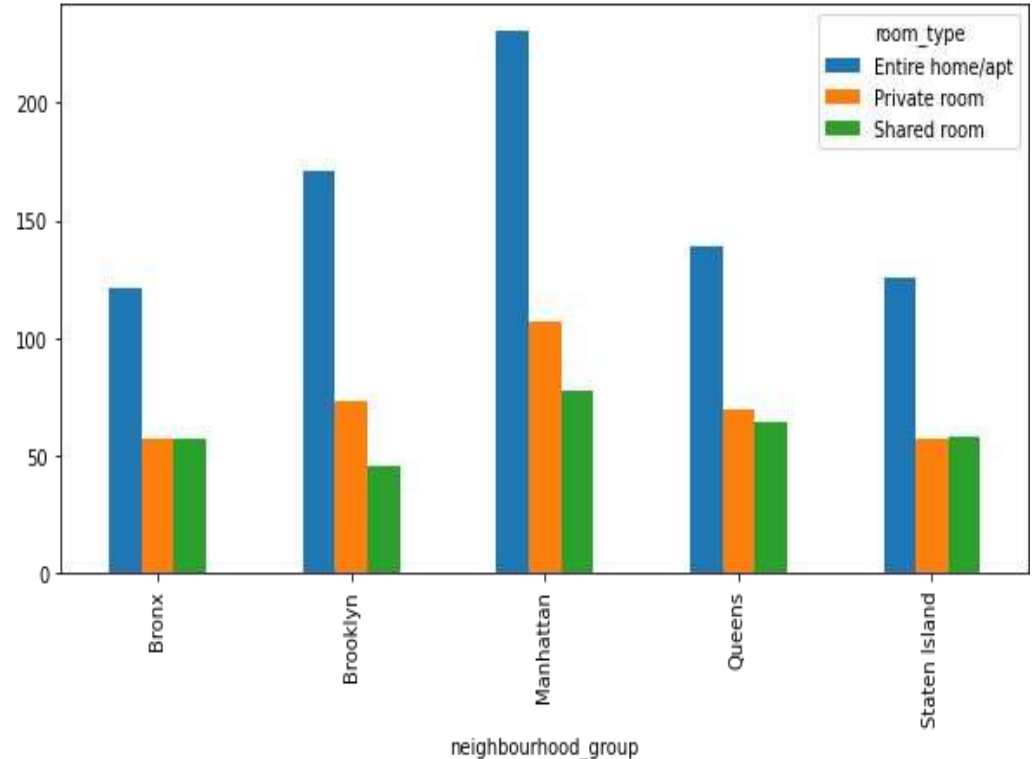
❖ **Number of active hosts per location (Where most of the hosts interested to own property?)**

- **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
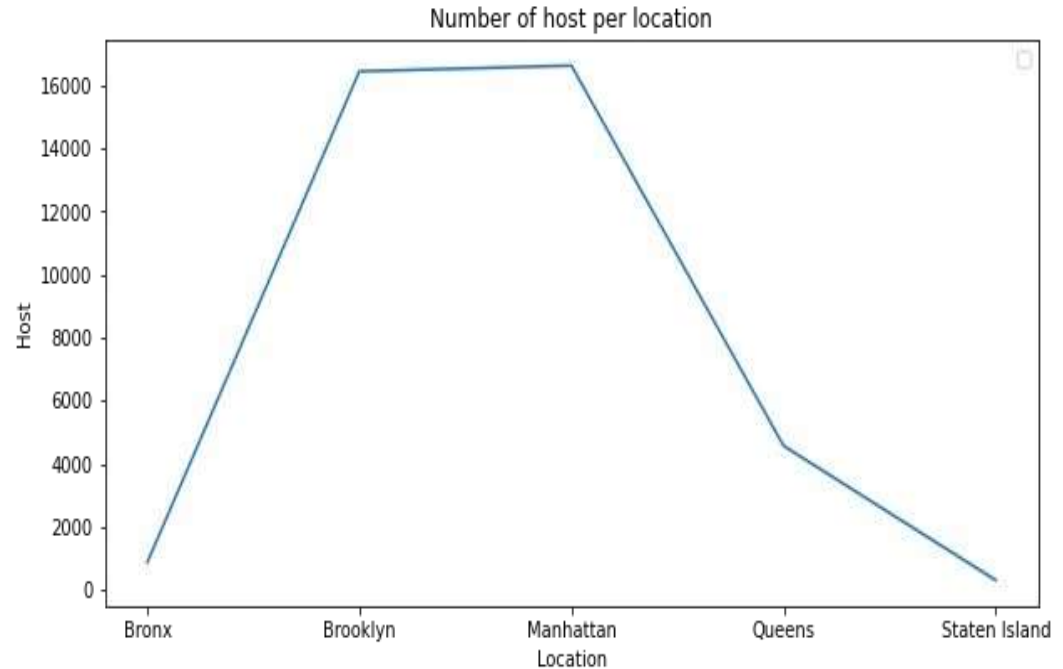
❖ **Where does the customers pay the highest and lowest rent according to location?**

- **Maximum price of rent in Brooklyn, Queens, Manhattan is almost same.**

- **Here we used log transformation to display minimum price with maximum price.**

- **Minimum price in Brooklyn, Queens, Manhattan is also same.**

Maximum and minimum price according to location
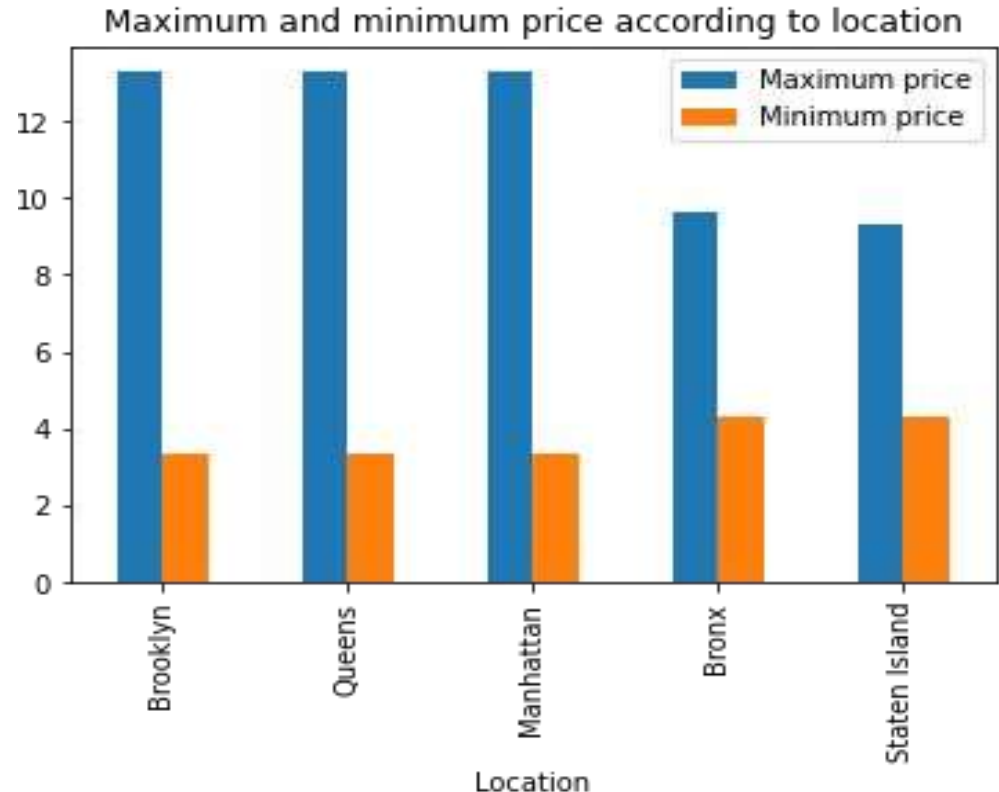
❖ Most popular/demanded host of Airbnb in New York City

- **Here we used Scatter plot for displaying the most demanded host with respect to Number of reviews and availability in a year.**

- **In the table below we have displayed top 5 hosts of Airbnb in New York City.**

|	host_id	host_name	number_of_reviews	availability_365
12	22959695	Gurpreet Singh	1157	0.0
41	99392252	Michael	732	0.0
47	121391142	Deloris	693	0.0
123	792159	Wanda	480	0.0
125	37818581	Sofia	479	0.0
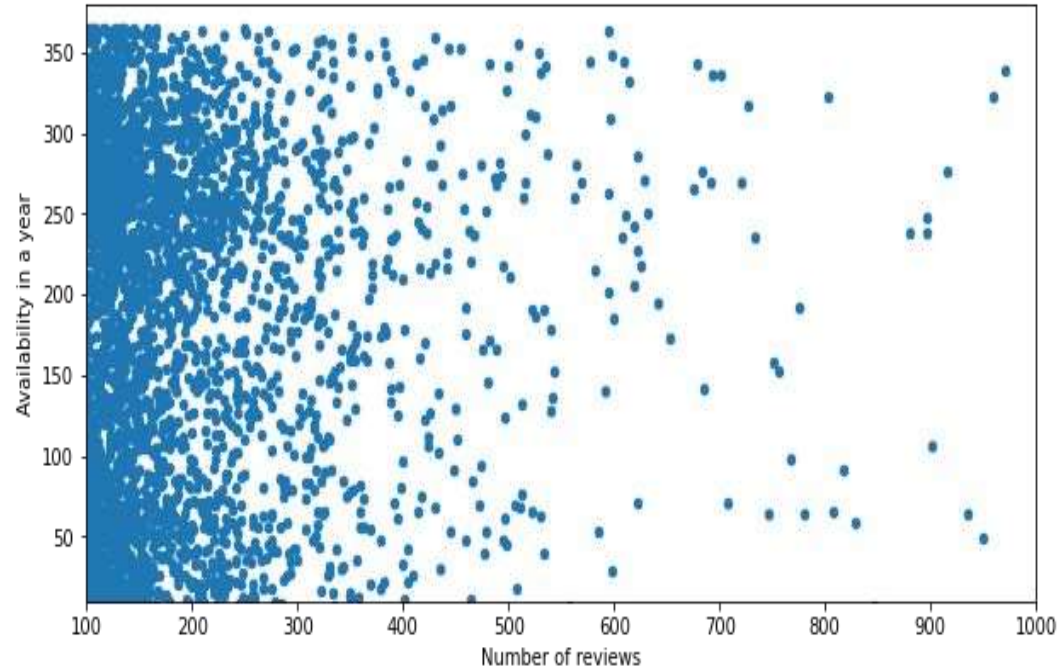
❖ **Total count of each room type as per listing.**

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

➜ **Entire home/apt** has the highest number of listing of 52% among other room types.

➜ **Private room** has 45.7% of listing among other room types.

➜ **Shared Room** is the least listed room type at only 2.4% in total.
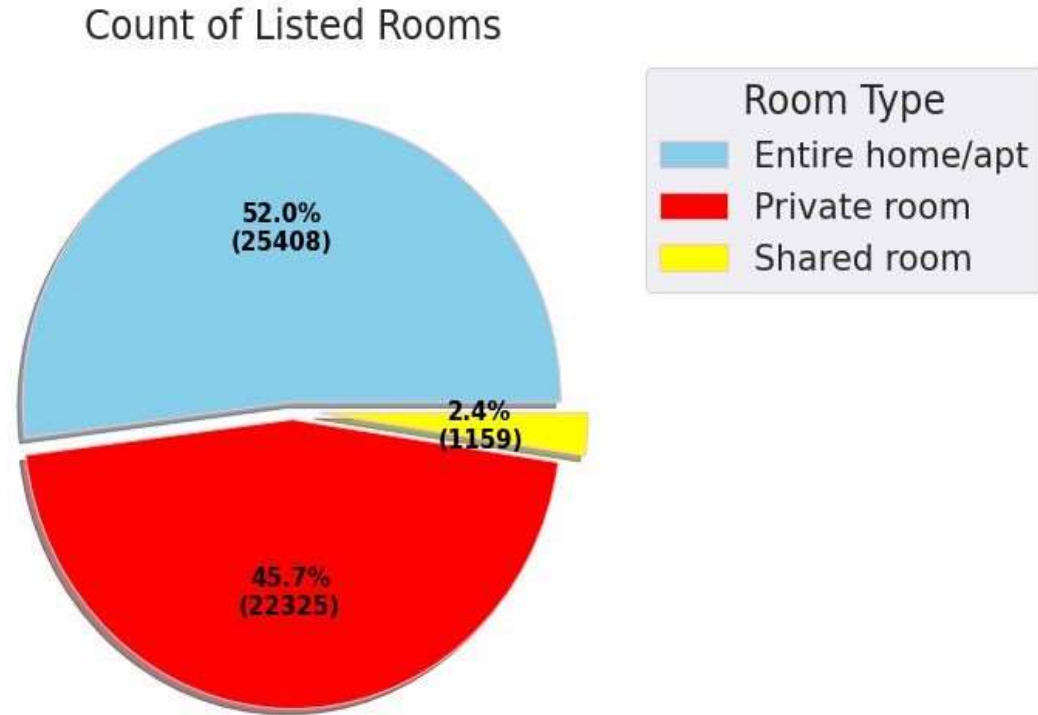
Count of Listed Rooms

Room Type
- Entire home/apt
- Private room
- Shared room

52.0%
(25408)

2.4%
(1159)

45.7%
(22325)

❖ **Room types and their relation with availability in different neighbourhood groups?**
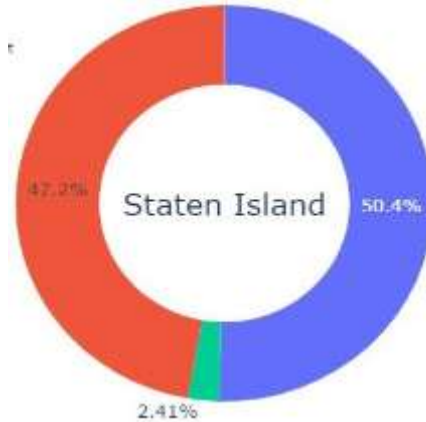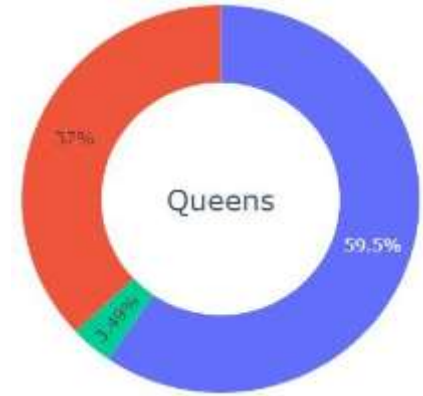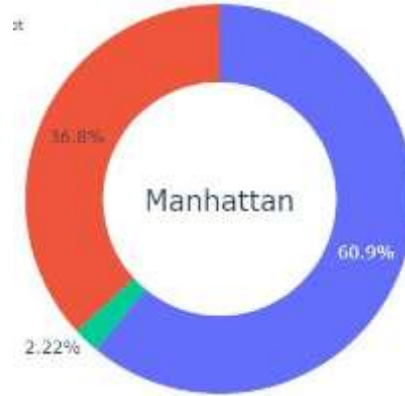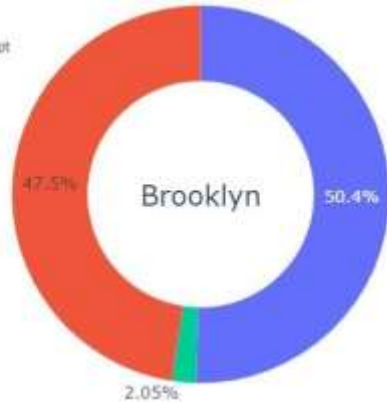
❖ **As we can see every neighbourhood group is dominated by the private room type at around 55%.**

Room types and their relation with availability in different neighbourhood groups?(contd.)

❏ **Brooklyn and Manhattan have the least availability of rooms overall, as low as 0 days.**

❏ **Staten Island and Bronx has the highest availability rate overall at around 300 days.**

❏ **Form this analysis we can say that people stay for longer duration of time in Private rooms in Brooklyn and Manhattan.**

Relation between Neighbourhood group & Availability of rooms
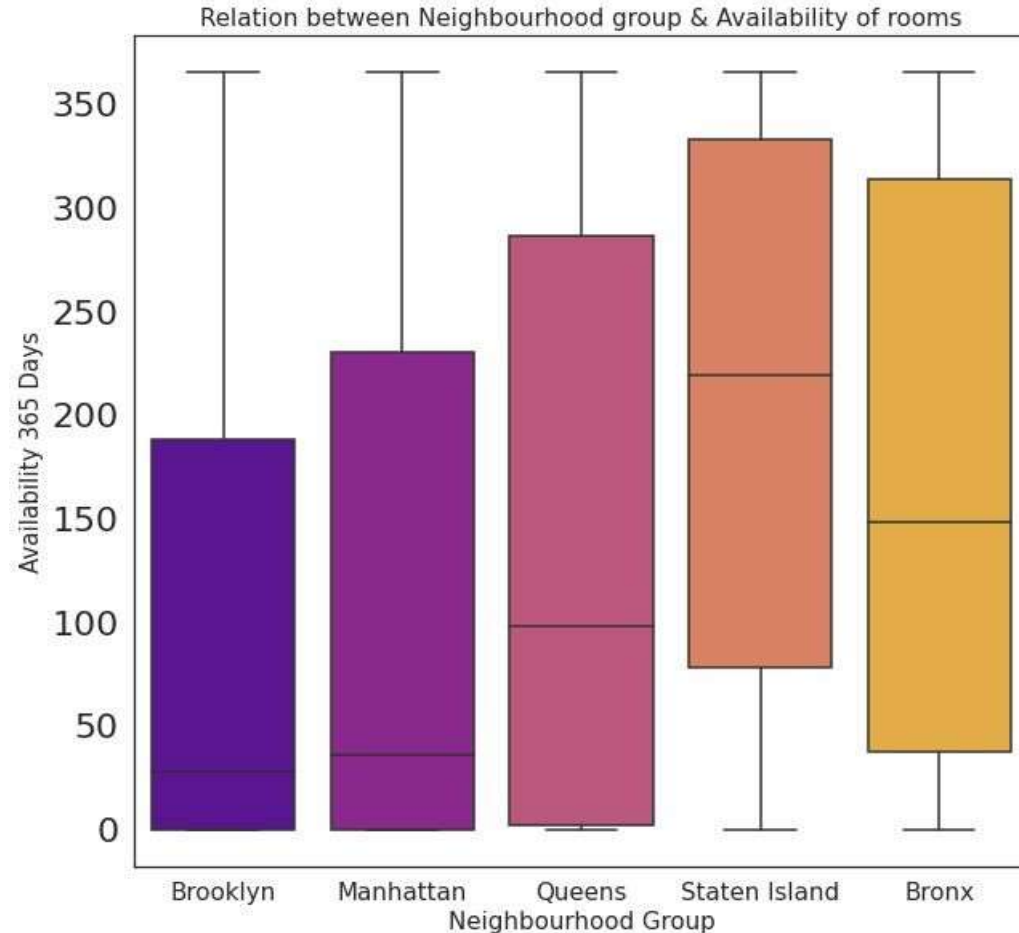
❖ **Top 25 most common words used in listing names**
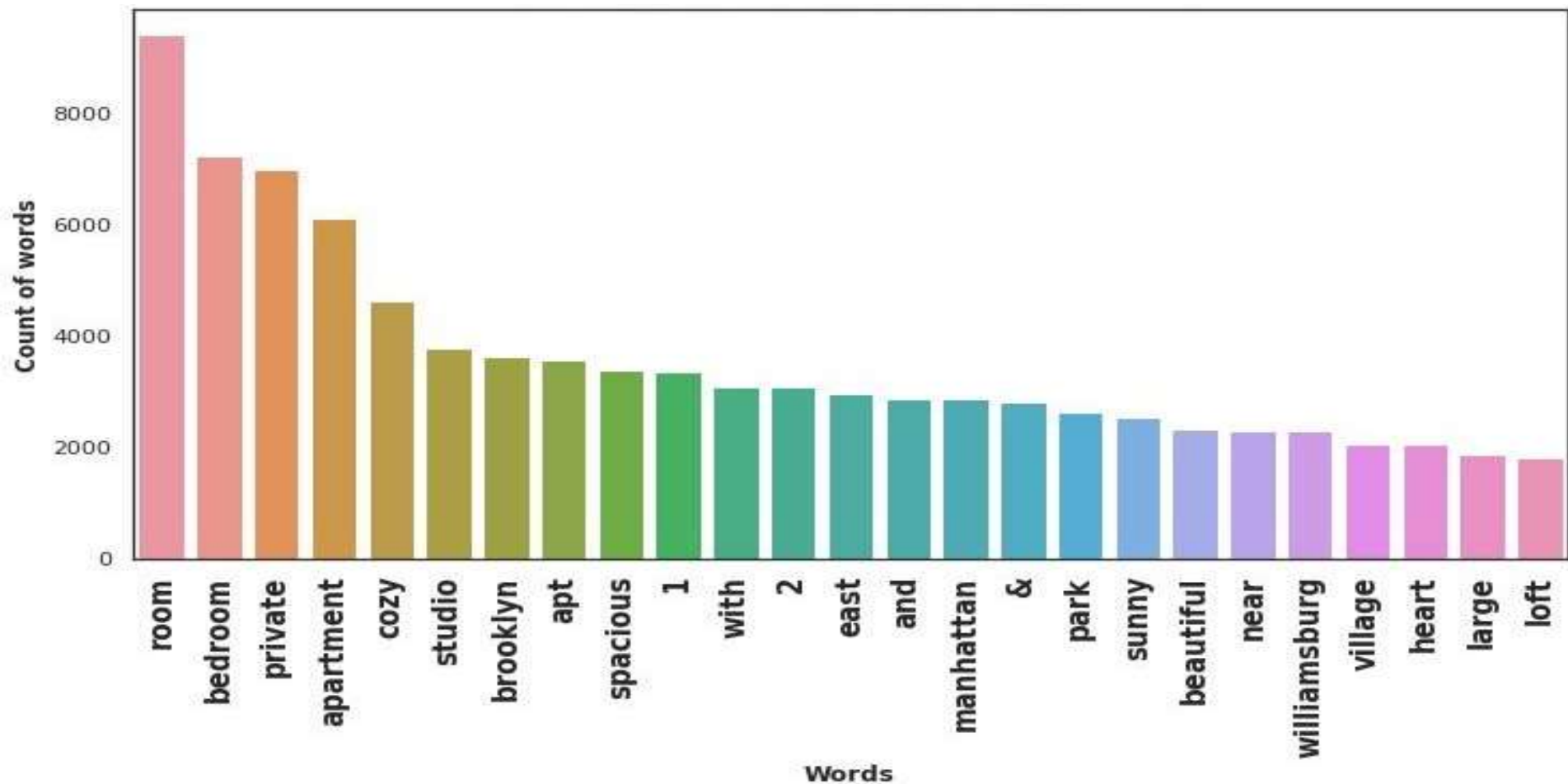
Top 25 most common words used in listing names (contd.)

❏ As we can see most of the listing names include words related to property type such as 'bedroom', 'cozy', 'private', 'apartment' and 'spacious'.

❏ It is interesting to see that words related to proximity or connection to public places such as 'park', 'near', 'village' and 'heart' rank lower in chart.

Hosts with most listings in New York
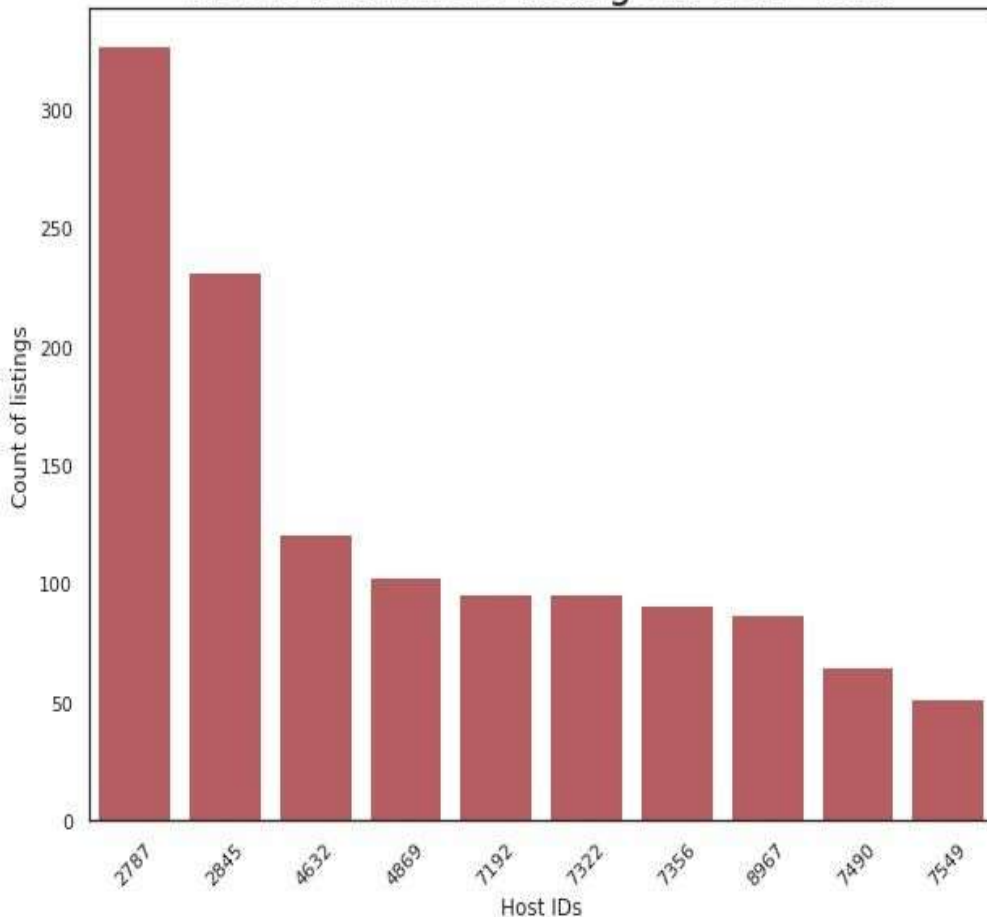
❖ Top 10 hosts with most number of listings.

❏ **From the chart, we can see that count of listing by top 10 hosts is almost 2.5%(1270 listings) of the whole dataset.**

❏ **Even one of the hosts has more than 300 listings!**

❖ **Top three hosts based on their turnover**

top 3 host based on their turnover

- **Here we are trying to find top three host based on their turnover and compare between those three hosts.**

- **We can state that Sonder(NYC) is the top host based on his high turnover.Red Awning and Henry secured 2nd and 3rd position respectively.**

- **Total turnover of Sounder(NYC) is more than 50000$. Red Awning and Henry is quite behind from the 1st place.**
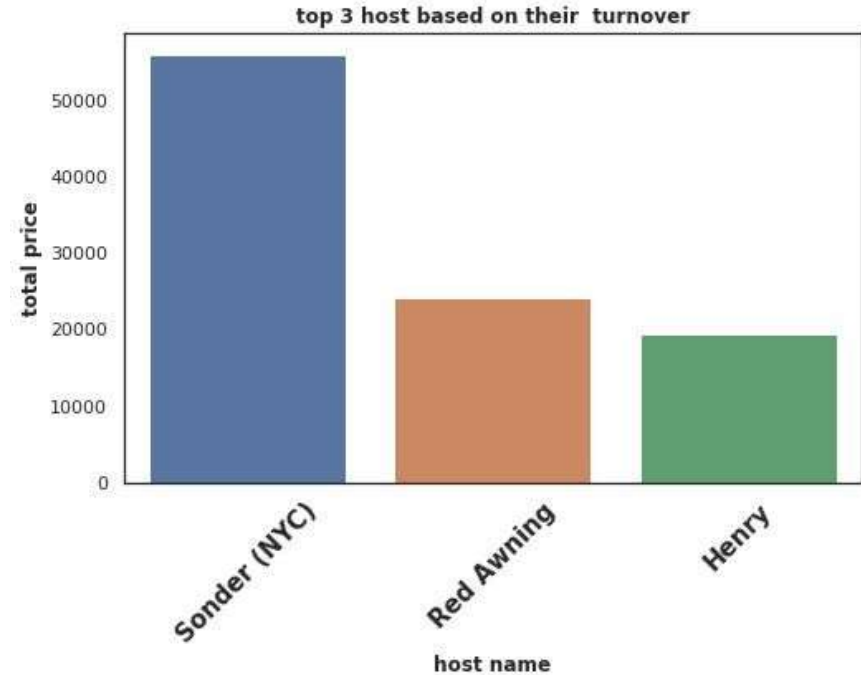
❖ **Find total no. nights spend as per location**

- **The locations are categorized based on its total number of minimum nights spend by customer.**

- **From here we can get an overall idea of which location the customer prefers more.**

- **We can state that more customers prefer Manhattan and Brooklyn for night stay as compared to other locations.**

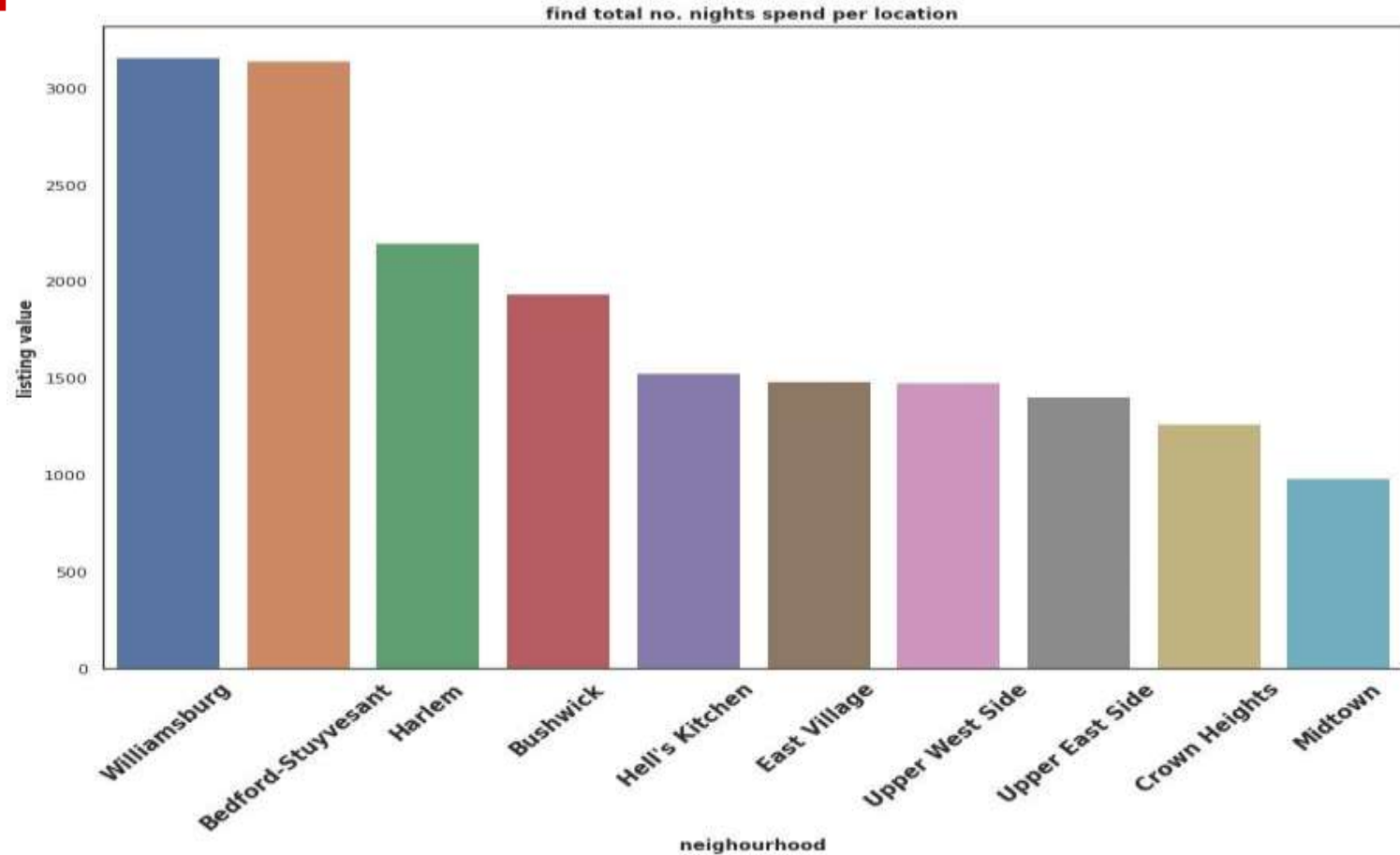find total no. nights spend per location
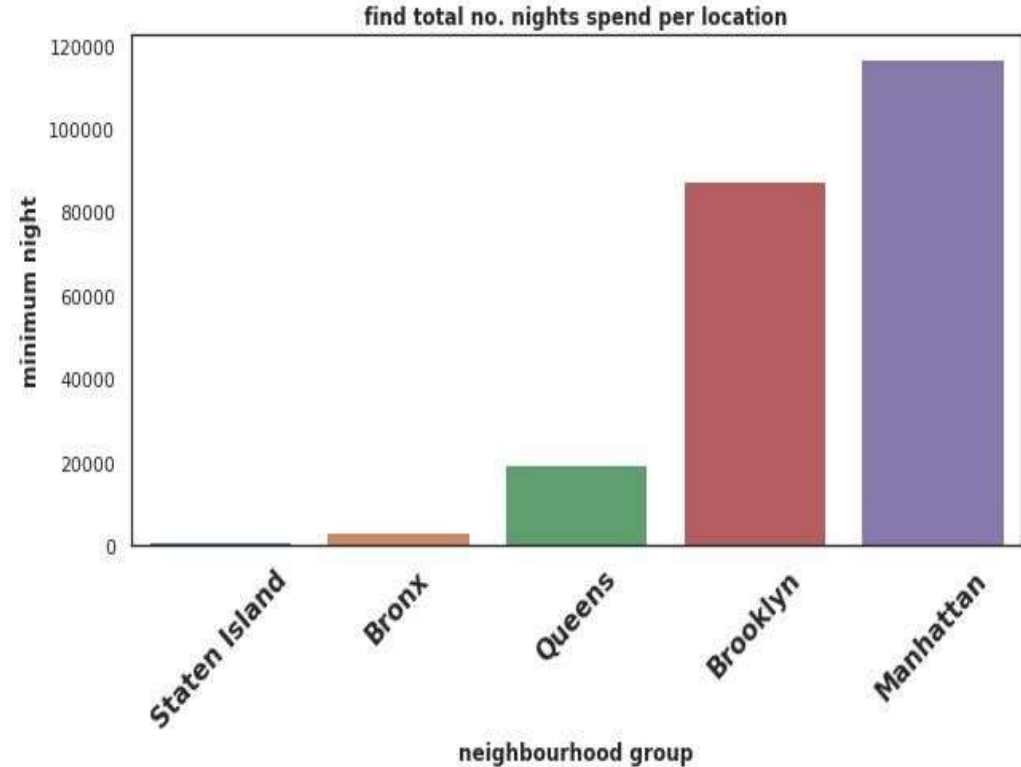
❖ **Total no. of nights spend per room types**

- **Here we can state that in which room type customers prefer for night stay.**

- **From pie chart we can conclude that 63.2% customers spend night in entire home/apt.**

- **Only 1.6% customers spend night in shared room.**

total no. of nights spend per room types

find total no. nights spend per location
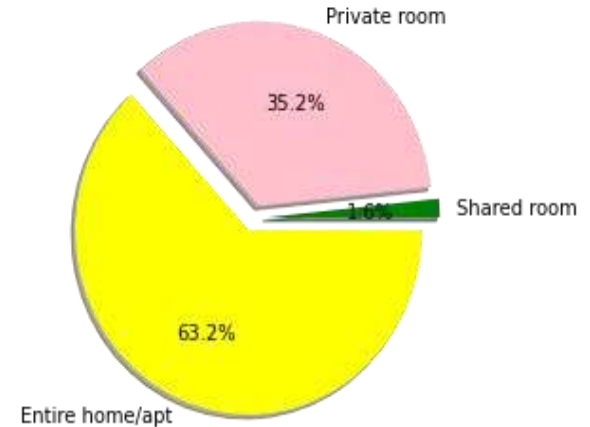
Top ten highest listing neighbourhood(contd.)

- **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.**

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.**

Analysis Summary:

➔ **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)**

Analysis Summary (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.**

- **Words such as 'bedroom', 'cozy', 'private', 'apartment' and 'spacious' are used more frequently than words such as 'park', 'near', 'village' and 'heart'.**

- **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**

Thank you