Capstone Project

Airbnb Booking Analysis

By

Patan Ismail Alli Khan
(pathanismailalikhan@gmail.com)

What is Airbnb ?

Airbnb, as in “Air Bed and Breakfast,” is a service that lets property owners rent out their spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Airbnb is based on a peer-to-peer business model. This makes it simple, easy to use, and tends to be more profitable for both parties. The model also gives you the opportunity to customize and personalize your guests’ experience the way you want.

Problem Statement

Since 2008, guests and hosts have used Airbnb to expand on travelling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data that can be analysed and used for security, business decisions, understanding of customers' and providers (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around 49,000 observations in it with 16 columns and it is a mix of categorical and numeric values.

Explore and analyze the data to discover key understandings.

Workflow


Data Understanding
Data Cleaning And Manipulation
Exploratory Data Analysis

Data Understanding

id - Unique id of listing

name - Name of the listing

host_id - Unique id of host

host_name - Name of the host

neighbourhood_group - Location of the listing

neighbourhood - Area of the listing

latitude - Latitude of listing

longitude - Longitude of listing

room_type - Type of rooms

price - Price of listing

minimum_nights - Minumum number of nights to be paid for

number_of_reviews - Number of reviews given for the listing

last_review - Date of the last review given to the listing

reviews_per_month - No.of review given per month

calculated_host_listings_count - Total number of listings for host

availability_365
Number of days listing
is available

Data Cleaning and Manipulation

Our Dataset has 4 column's with Null Values

```
# Missing Values/Null Values Count
df.isnull().sum().sort_values(ascending=False)
```

```
last_review                    10052
reviews_per_month              10052
host_name                         21
name                              16
id                                 0
```

1. We can replace reviews_per_month with 0 as these Stays might not have been rated at all.

```
[9] # Handling Null values
    df.fillna({'reviews_per_month':0},inplace=True)
```

2. last_review is a date column. Replacing this with any random date might corrupt the dataset

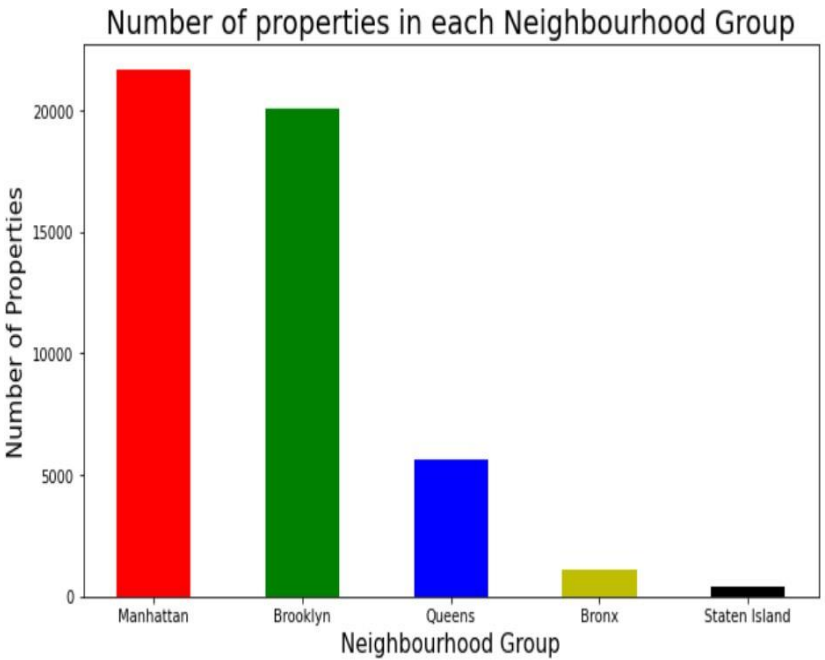
3. We will not be using host_name and name columns in our analysis.

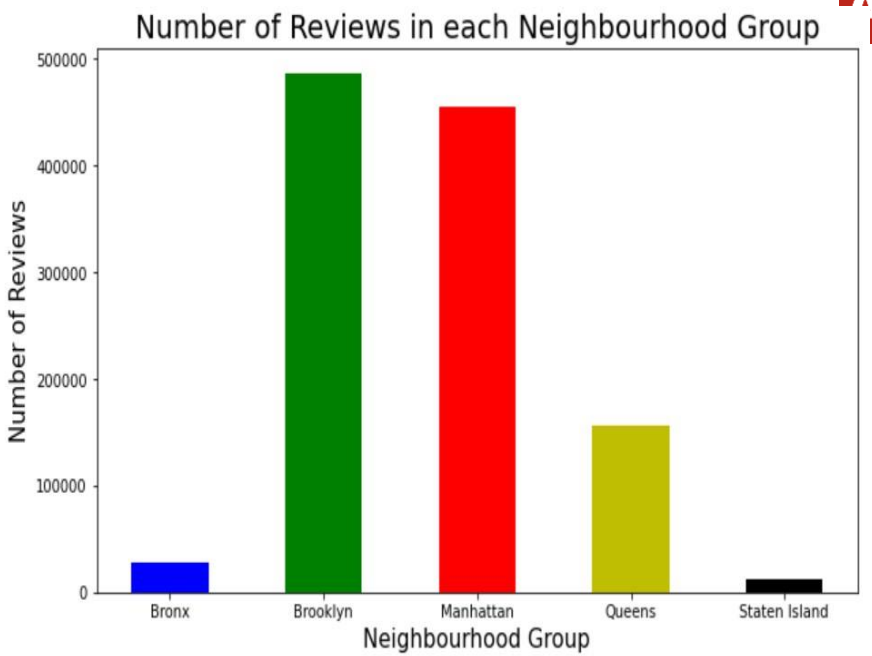
Exploratory Data Analysis

Numerical Features	Categorical Features
id	name
host_id	host_name
price	neighbourhood
minimum_nights	neighbourhood_group
number_of_reviews	room_type
calculated_host_listings_count	
availability_365	

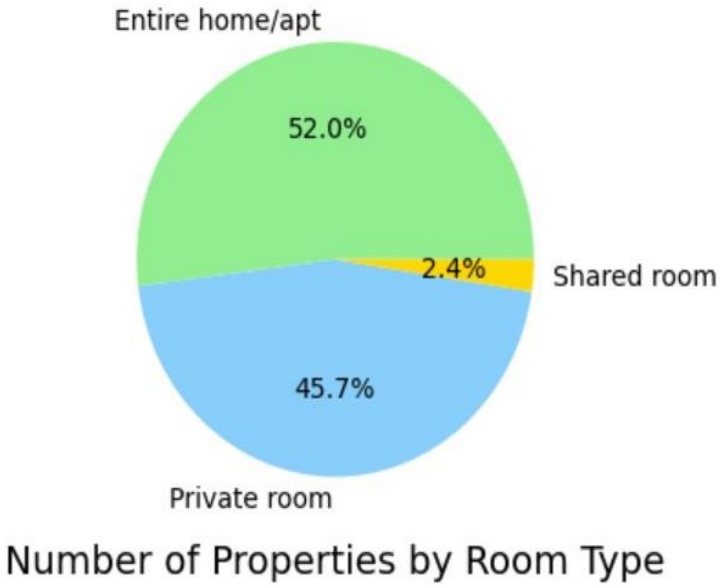
Correlation between features

From the above Heat Map we can see that there is no correlation between any of the features.

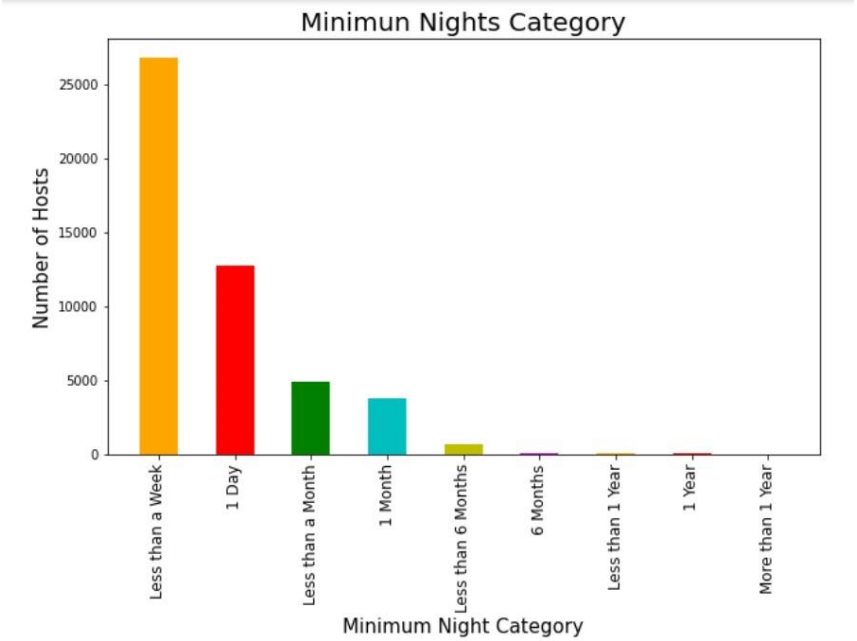




From the above bar graphs we can see that most number of properties are in Manhattan and Most number of reviews are for Brooklyn location properties.

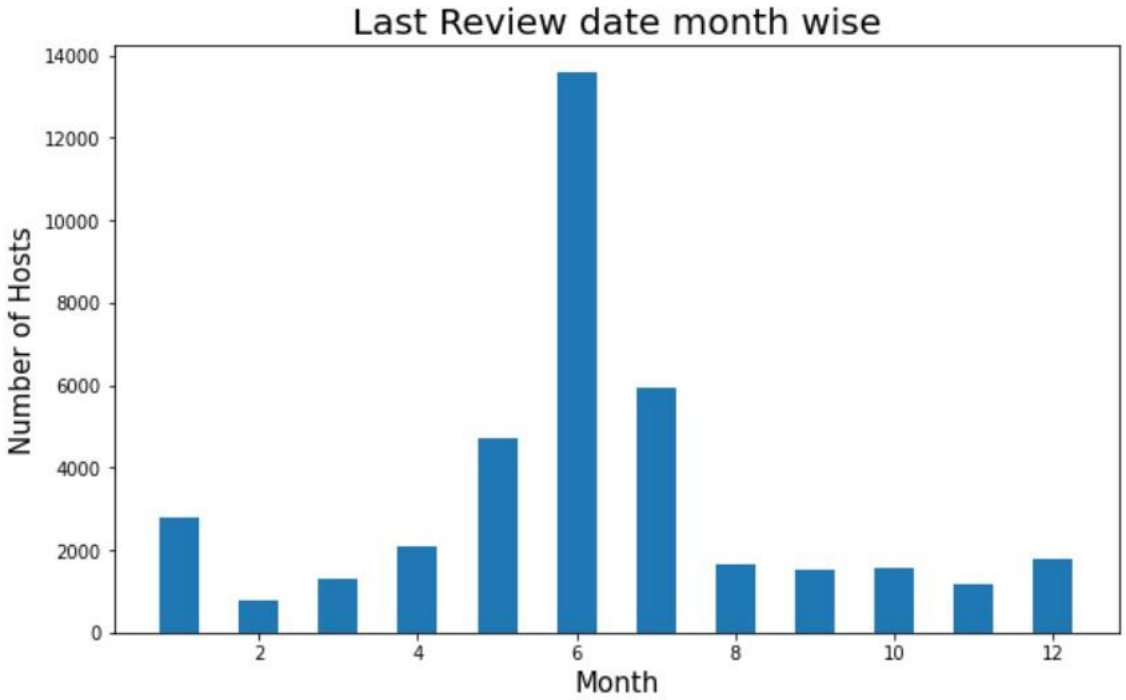


Number of Properties by Room Type

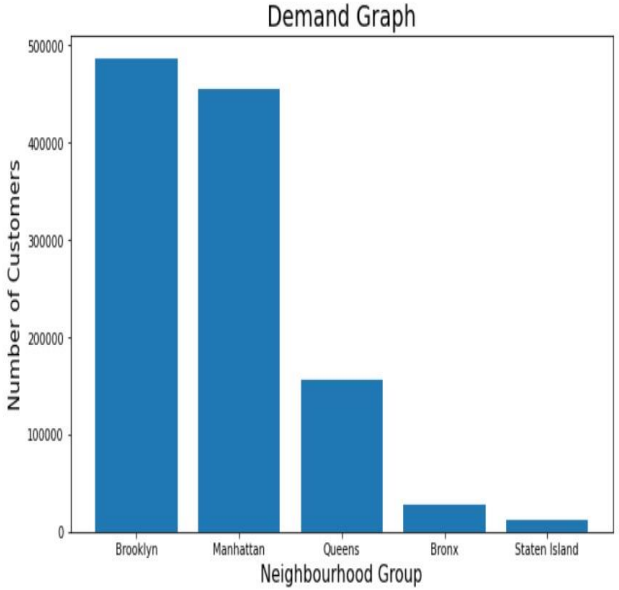


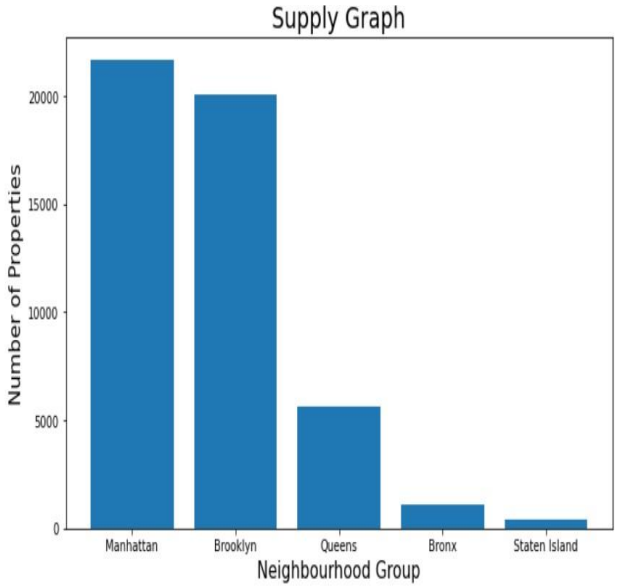
From the above Pie chart we can see that most properties offer Entire home or apartment room type.

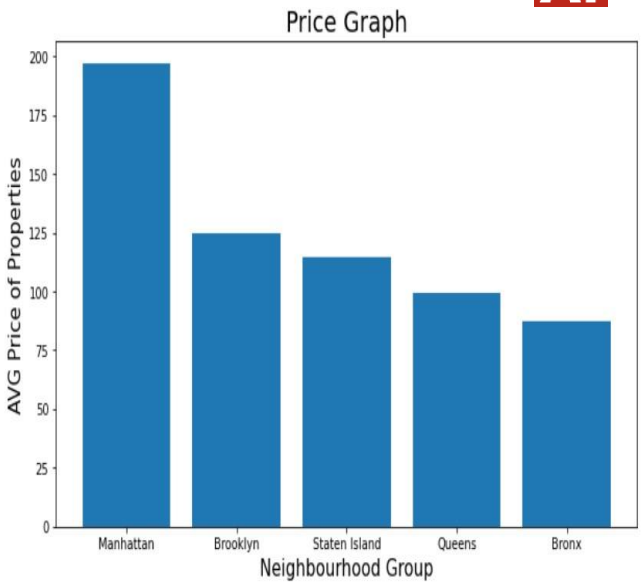
The Bar graph shows that many properties prefer minimum 1 day or less that a week nights to be paid for



From the above bar graphs we can observe that Most number of people have either vacated / changed there property in the month of June







From the above bar graphs we can observe that

1.Brooklyn has high demand for properties

2.More Properties are available in Manhattan

3.Manhattan has the highest average price

Exploring into Neighborhoods

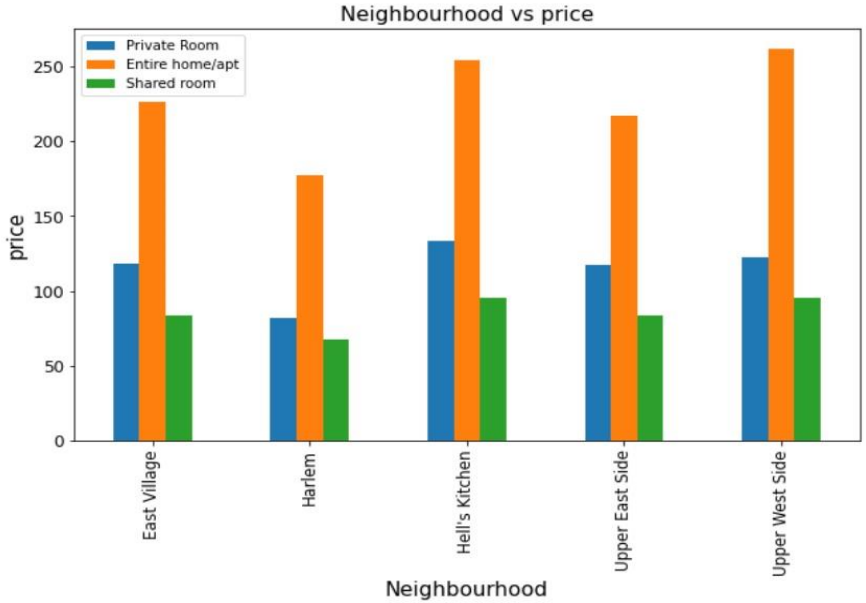
```
[25] # This function will give subset of inputed "Neighbourhood Groups" all "Neighbourhoods"
     def get_neighbourhoods(n_grp):
       ng_df=df[df['neighbourhood_group']==n_grp]
       neighbourhoods_count=ng_df['neighbourhood'].value_counts().reset_index()
       nh1=ng_df[ng_df['neighbourhood']==neighbourhoods_count['index'][0]]
       nh2=ng_df[ng_df['neighbourhood']==neighbourhoods_count['index'][1]]
       nh3=ng_df[ng_df['neighbourhood']==neighbourhoods_count['index'][2]]
       nh4=ng_df[ng_df['neighbourhood']==neighbourhoods_count['index'][3]]
       nh5=ng_df[ng_df['neighbourhood']==neighbourhoods_count['index'][4]]
       final_df=pd.concat([nh1,nh2,nh3,nh4,nh5])
       return final_df
```

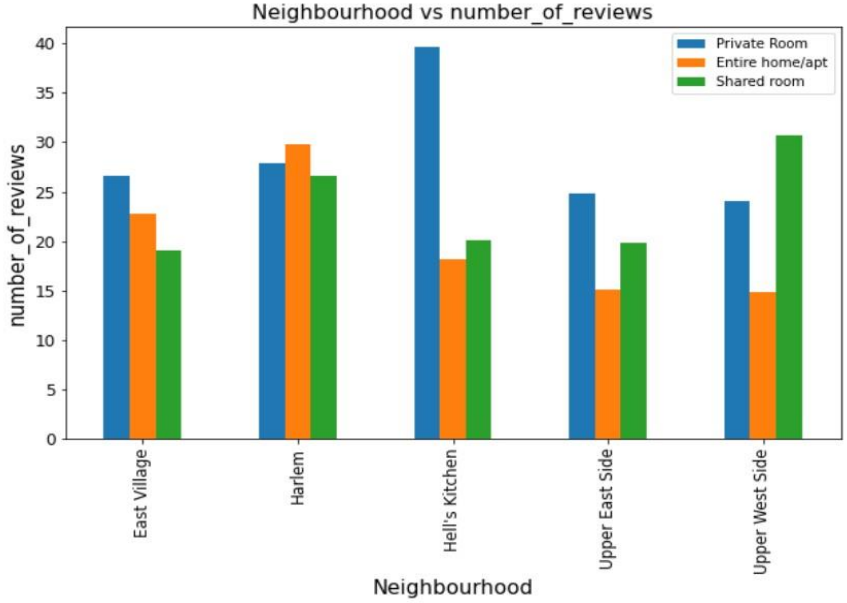
The above function will take Neighborhood group as input and return the subset of all Neighborhoods in that group.

```
#This function will further subset the inputed "room_type" from all "Neighbourhoods"
def get_roomtype(room_type,neighbourhood_group):
  ng_subset=get_neighbourhoods(neighbourhood_group)
  result=ng_subset[ng_subset['room_type']==room_type]
  return result
```

The above function will take room type and Neighborhood group as input and further subset for each room type.

Manhattan



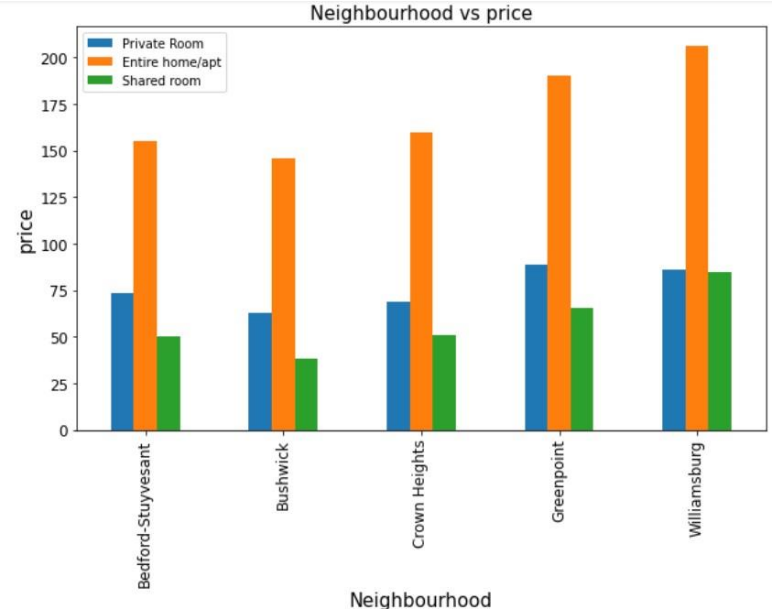


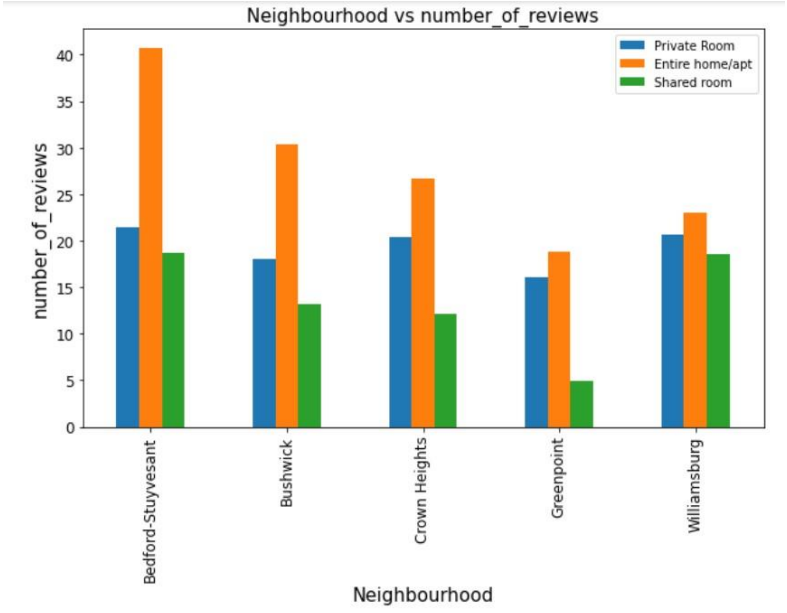
From the above bar graphs we can observe that

1. Hell's Kitchen and Upper West side in Manhattan have Highest Average Price

2. Hell's Kitchen has the highest number of reviews for Private room Type of Properties.

Brooklyn



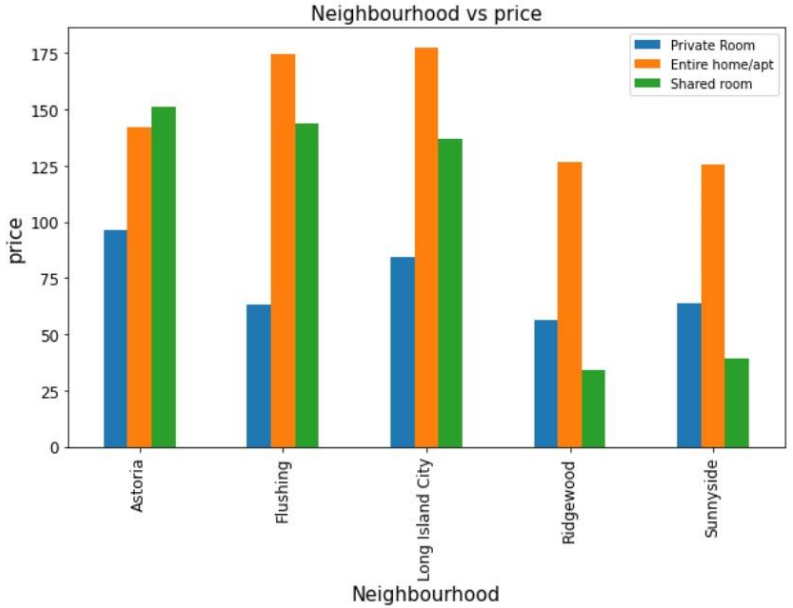


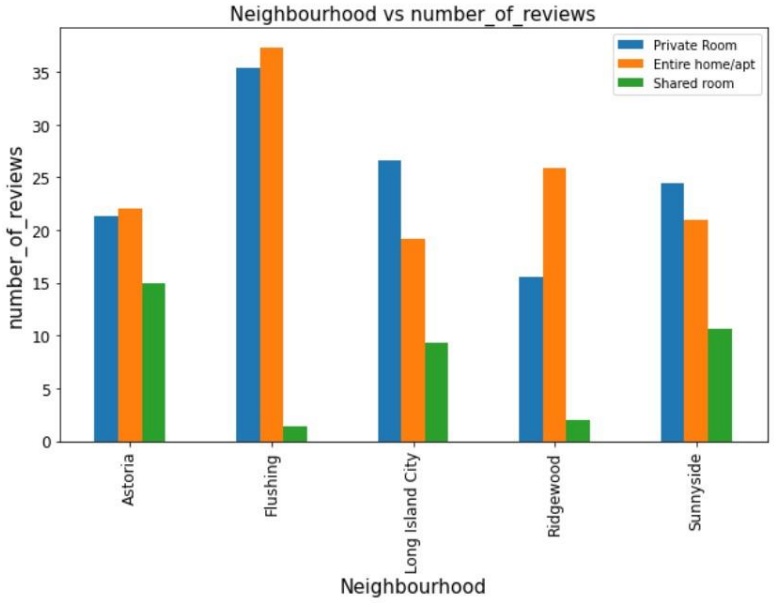
From the above bar graphs we can observe that

1. Williamsburg in Brooklyn has the Highest Average Price for all room types.

2. Most reviews are for the Entire home/apt type in Brooklyn and Williamsburg has the least reviews for Entire home/apt type.

Queens



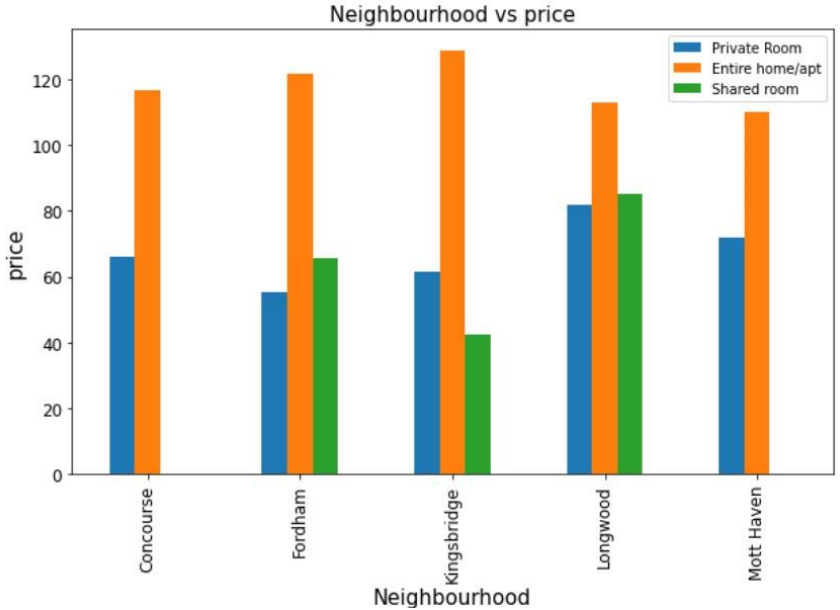


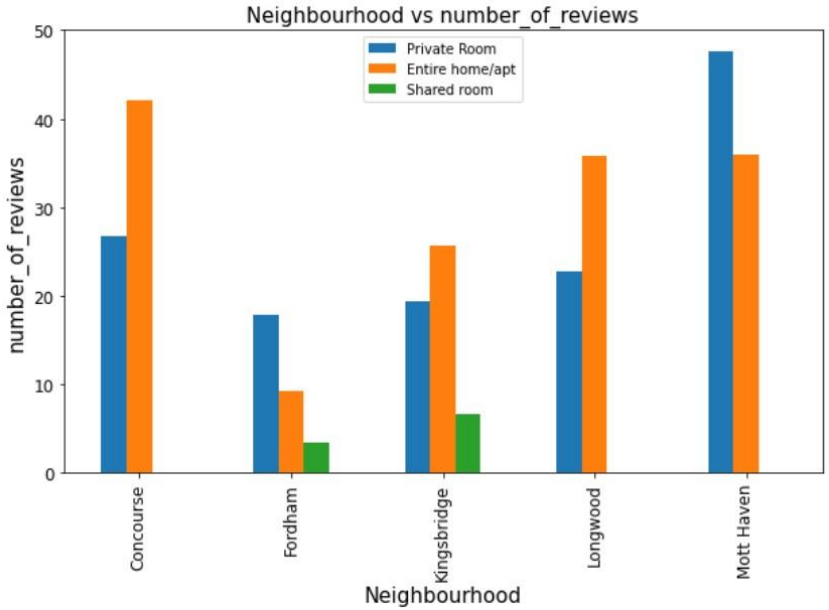
From the above bar graphs we can observe that

1. The Maximum price in Queens is very less compared to Manhattan and Brooklyn.

2. Entire home/apt and Private Room Type in the Flushing Area has highest number of reviews compared to other Areas.

Bronx



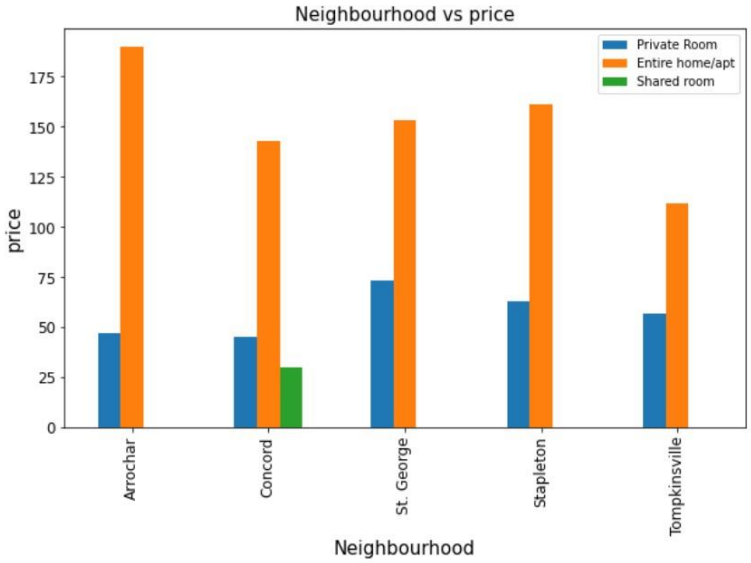


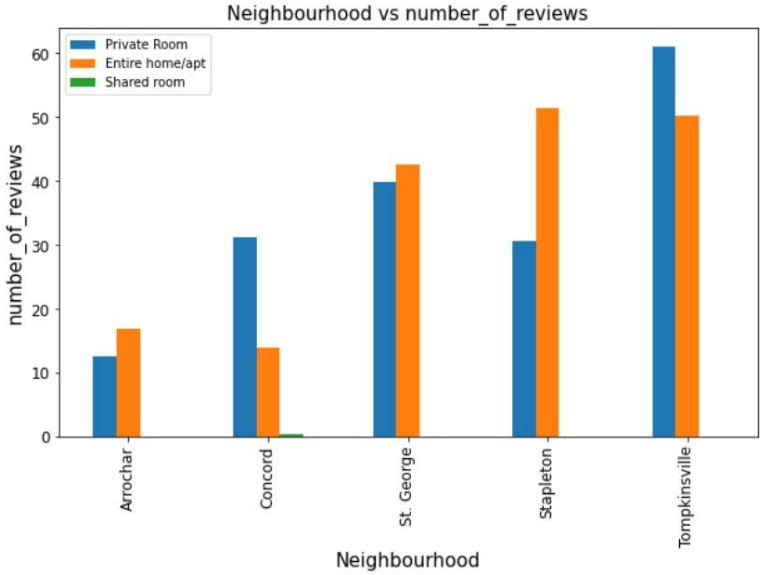
From the above bar graphs we can observe that

1.Entire home/apt Type Prices are high in almost all Areas in Bronx.

2. Most of the reviews are for Entire home/apt or Private Room type.

Staten Island

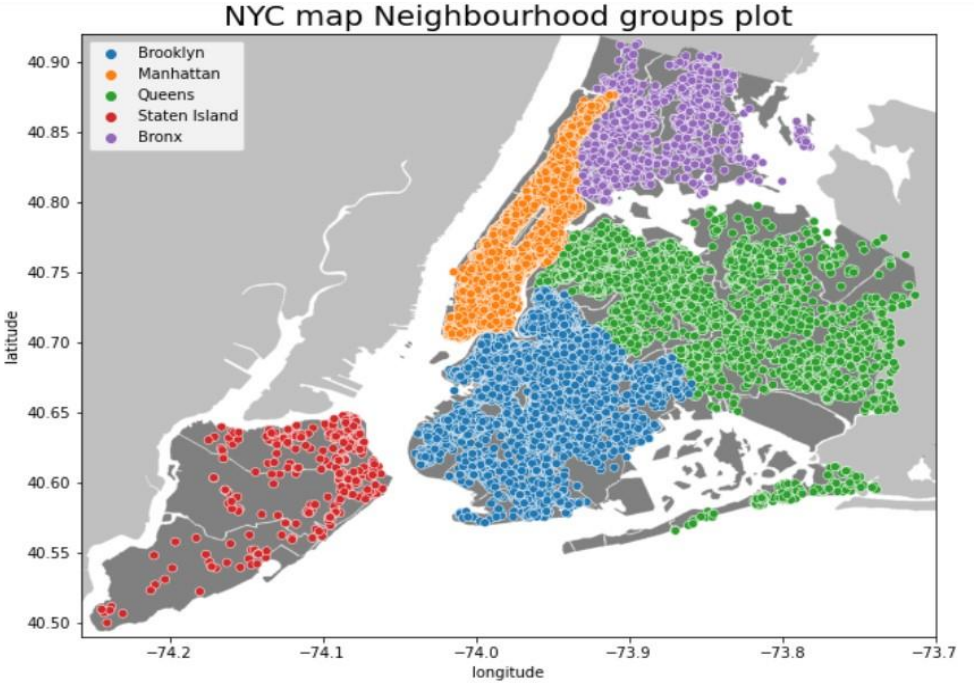




From the above bar graphs we can observe that

1.Only Concord Area is having shared room type in Staten Island

2. Most reviews are for Tompkinsville Area.



From the above graph we can see an overview of the properties listed in the New York City differentiated by locations.

Conclusion

1.Most number of Properties are in Manhattan and Brooklyn followed by Queens.

2.Most people prefer Brooklyn and Manhattan followed by Queens

3.Most people prefer Entire home/apt type followed by private rooms and only few people prefer Shared room type.

4.Most people have either vacated or changed the property in the month of June.

5.Airbnb can increase number of properties in Brooklyn as there is large demand and get higher profits

6. In Manhattan, Hell's Kitchen and Upper West side have Highest Average Price

7. in Brooklyn , Williamsburg has the Highest Average Price for all room types.

8. The Maximum price in Queens is very less compared to Manhattan and Brooklyn and its even lower in Bronx.

9.Only Concord area in Staten Island has shared room.

THANK YOU