

Capstone Project - EDA

Airbnb Bookings Analysis

TeamMembers

- Rajdeep Limbu
- Mohit Sakhare
- Mayur Chougule
- Aman Mishra
- Kiran Mishra



- **Overview**
- **Understanding of Dataset**
- **Data Wrangling**
- **Data Visualization & Exploratory Data Analysis (EDA)**
- **Conclusion**



Airbnb is an US based organization founded in 2007. It is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.

We have a Airbnb dataset which has all the booking details for different properties listed in NY. NY is amongst the most expensive places to live in the USA. We want to perform an in-depth analysis of one of the most densely populated cities in the world.

Understanding of Dataset

Column Type	Description
id	Listing id
name	Name of the Property
host_id	Unique id for each host
host name	Name of the Host
neighbourhood_group	Location
neighbourhood	Area
Latitudes	Latitude coordinates
Longitude	Longitude coordinates
room_type	Type of room
Price	Price in dollars
minimum_nights	Minimum nights stayed in the property
number_of_reviews	Number of reviews written for the property
last_review	Latest/Last review date for the property
reviews_per_month	Total reviews in a month
calculated_host_listing_counts	Total number of listings against the host id
availability_365	Number of days the property is available for booking

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

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 48895 entries, 0 to 48894
```

```
Data columns (total 16 columns):
```

#	Column	Non-Null Count	Dtype
0	id	48895 non-null	int64
1	name	48879 non-null	object
2	host_id	48895 non-null	int64
3	host_name	48874 non-null	object
4	neighbourhood_group	48895 non-null	object
5	neighbourhood	48895 non-null	object
6	latitude	48895 non-null	float64
7	longitude	48895 non-null	float64
8	room_type	48895 non-null	object
9	price	48895 non-null	int64
10	minimum_nights	48895 non-null	int64
11	number_of_reviews	48895 non-null	int64
12	last_review	38843 non-null	object
13	reviews_per_month	38843 non-null	float64
14	calculated_host_listings_count	48895 non-null	int64
15	availability_365	48895 non-null	int64

```
dtypes: float64(3), int64(7), object(6)
```

```
memory usage: 6.0+ MB
```

Fix "bad" data

```
df.isnull().sum()
```

```
id                0
name              16
host_id           0
host_name         21
neighbourhood_group 0
neighbourhood     0
latitude          0
longitude         0
room_type         0
price            0
minimum_nights    0
number_of_reviews 0
last_review       10052
reviews_per_month 10052
calculated_host_listings_count 0
availability_365   0
dtype: int64
```

- Total 4 columns contains null values.
- Columns named - last_review and reviews_per_month contains 10k+ null values.
- Either drop the above columns if they are not important for Analysis or put 0 in all the null values.
- Columns - 'names' and 'host_names' have some null values. So, replacing them with 'Not Applicable'.

```
df['last_review'].fillna(0, inplace = True)
df['reviews_per_month'].fillna(0, inplace = True)
df['name'].fillna("Not Applicable", inplace = True)
df['host_name'].fillna("Not Applicable", inplace = True)
```

Unique value insights

```
df.nunique()
```

id	48895
name	47905
host_id	37457
host_name	11452
neighbourhood_group	5
neighbourhood	221
latitude	19048
longitude	14718
room_type	3
price	674
minimum_nights	109
number_of_reviews	394
last_review	1764
reviews_per_month	937
calculated_host_listings_count	47
availability_365	366
dtype:	int64

- Columns such as `host_id` represents that there are 37457 hosts listed in the Dataset.
- We can see that `host_name < host_id`, from this we can conclude that the hosts might have more than one properties listed on Airbnb.
- NYC is divided among 5 groups.
- Room types are divided into 3 types.

Finding Neighbourhood Group and Room types

```
pd.DataFrame(df.neighbourhood_group.unique(), columns = ['neighbourhood_group'])
```

neighbourhood_group



0	Brooklyn
1	Manhattan
2	Queens
3	Staten Island
4	Bronx

```
pd.DataFrame(df.room_type.unique(), columns = ['room_type'])
```

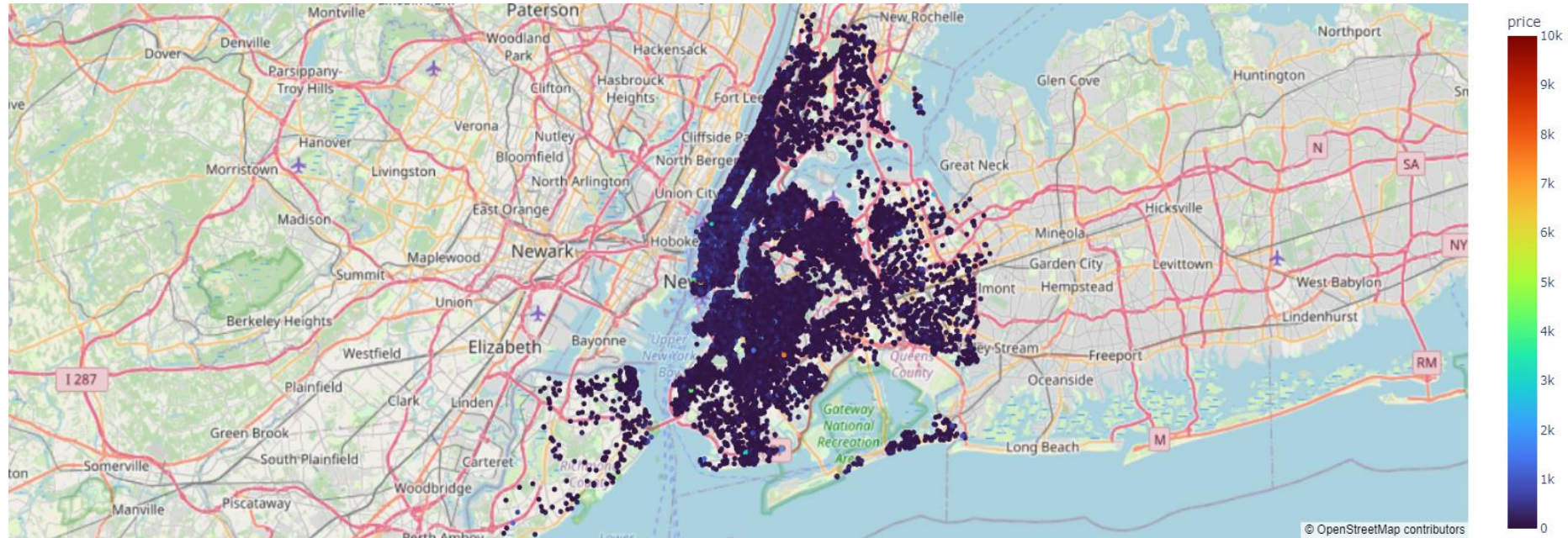
room_type



0	Private room
1	Entire home/apt
2	Shared room

Data Visualization

Plotting of Property co-ordinates w.r.t price on Map

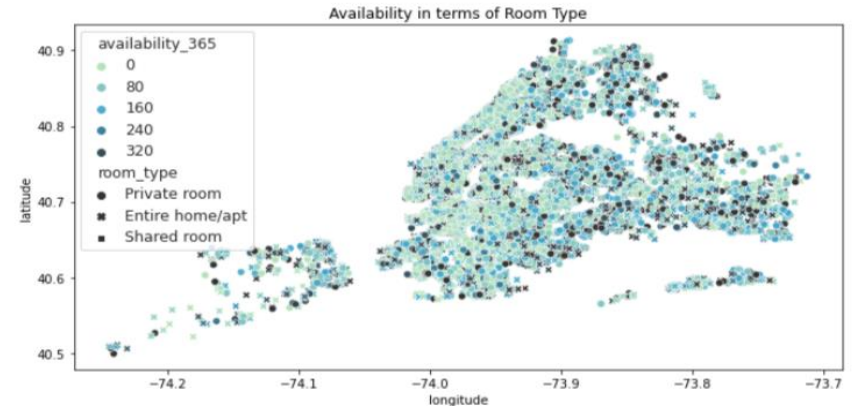
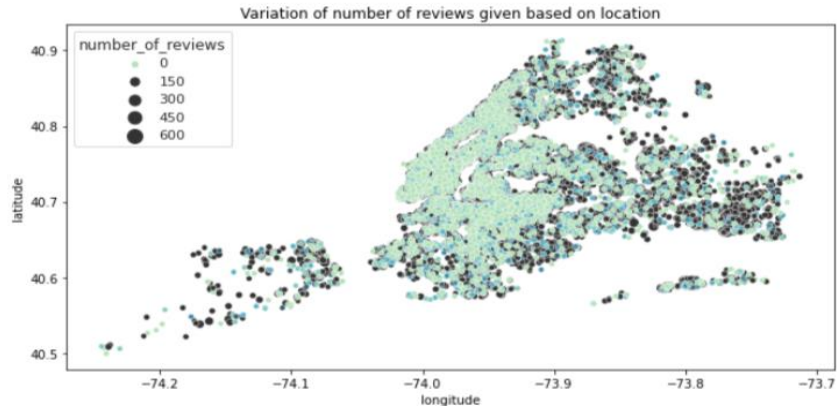
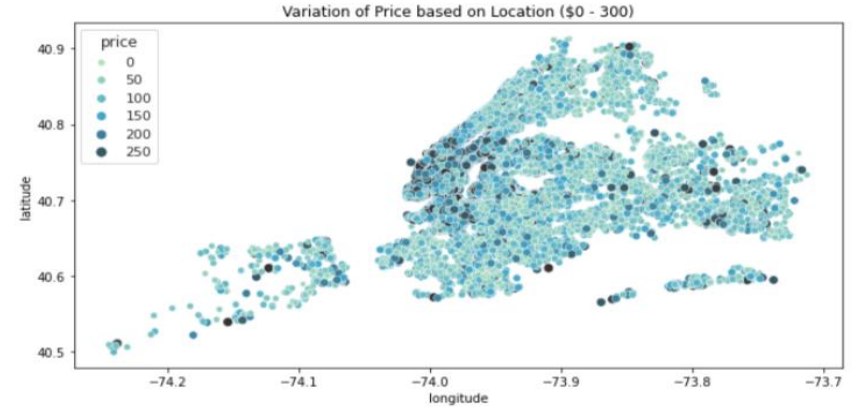
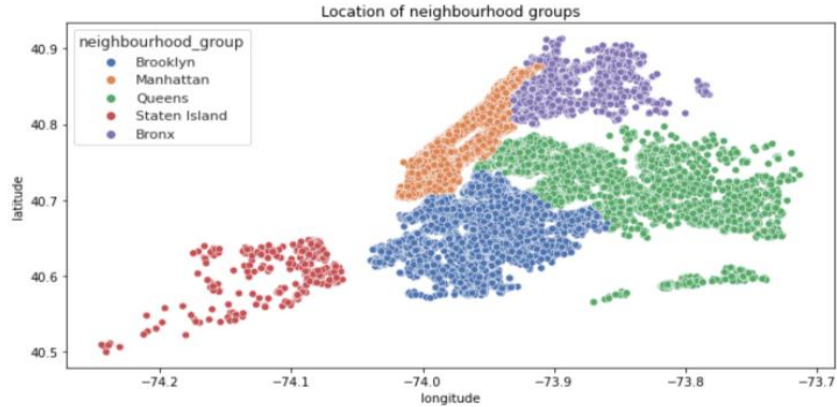


Insight :

From the map , it can be seen that majority of the property listings are populated in the major Industry focused and urban regions of Manhattan & Brooklyn.

Data Visualization

Plotting different graphs on the basis of Latitudes and Longitudes coordinates.



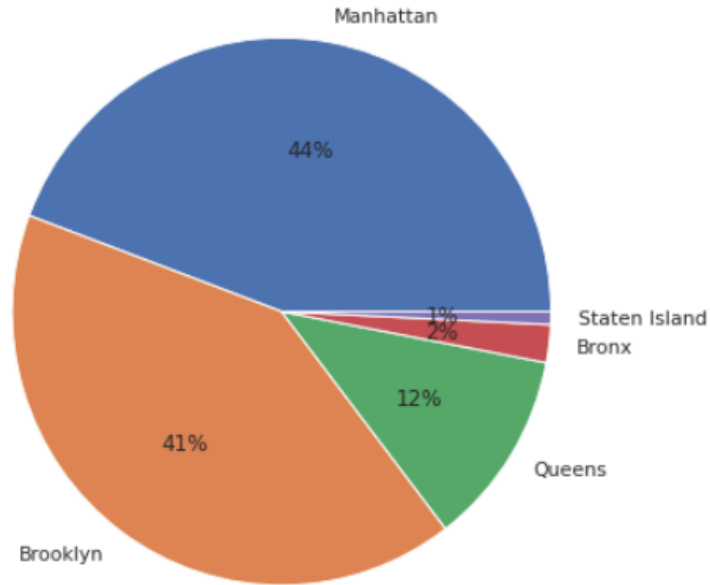
Plotting different graphs on the basis of Latitudes and Longitudes coordinates.

Insights :

- **The scatter plot graphs gives a focused segmentation of ' Neighbourhood Groups' in the entire NYC region.**
- **It also gives a clear understanding of distribution of prices , number of reviews and availability of different room-types for a year in the given region.**
- **The size of the dots and colour intensity represents the values for the respective graph type**

Neighbourhood wise listings

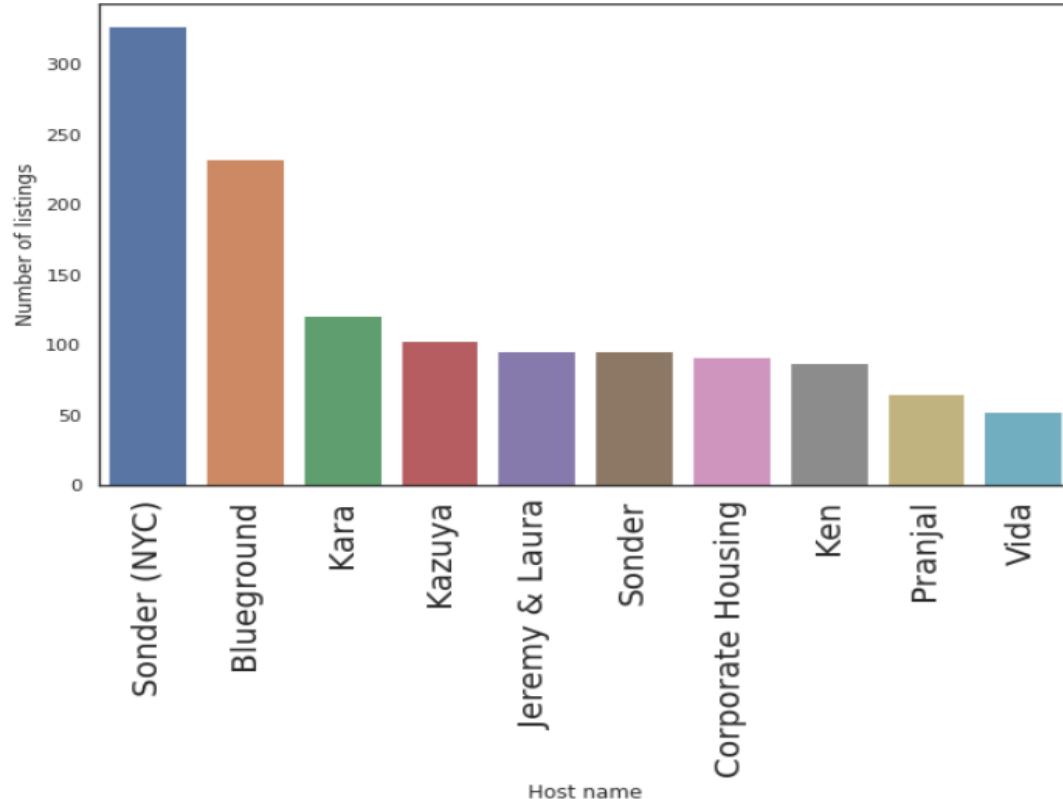
Pie chart showing percent of listings according to neighbourhood group



Insights :

- **Brooklyn and Manhattan has the maximum percentage of listings.**
- **Staten Island and Bronx contribute the lowest.**

Top Host having highest number of properties in NYC

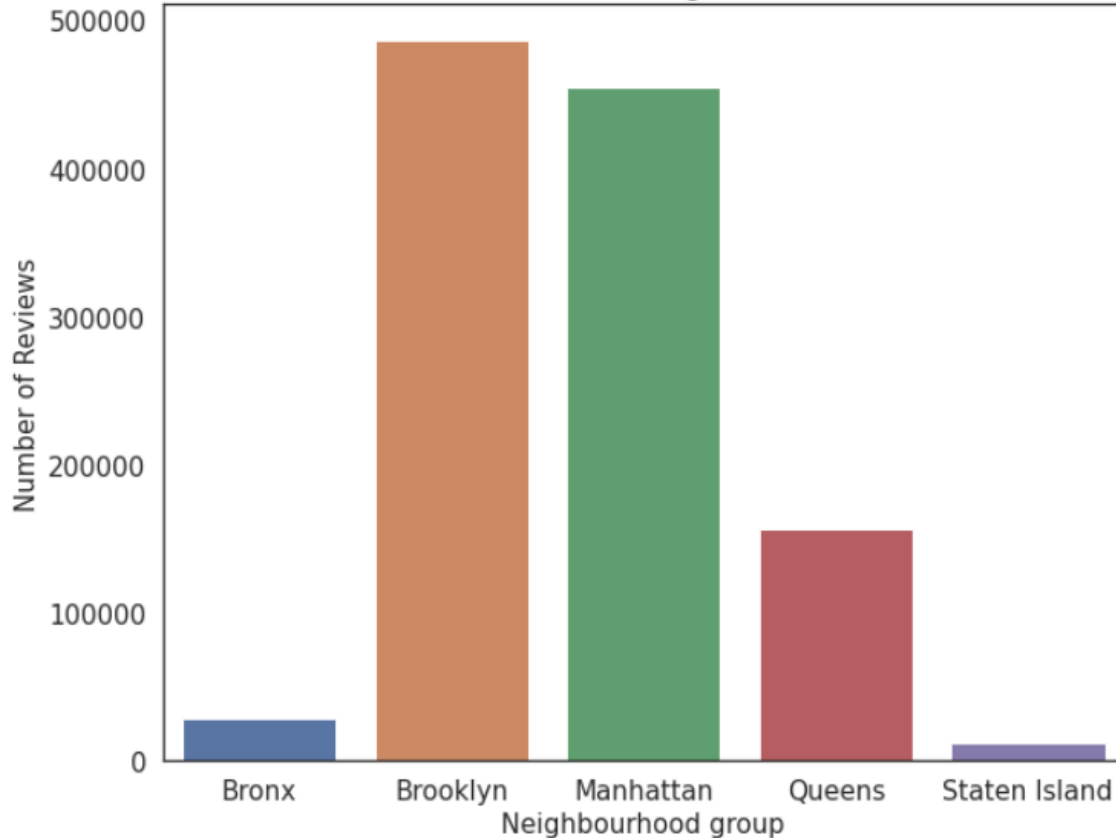


Insights :

- Host name - Sonder (NYC) has the most number of properties listed on Airbnb in the given region

Exploratory Data Analysis

Most reviewed Neighbourhood

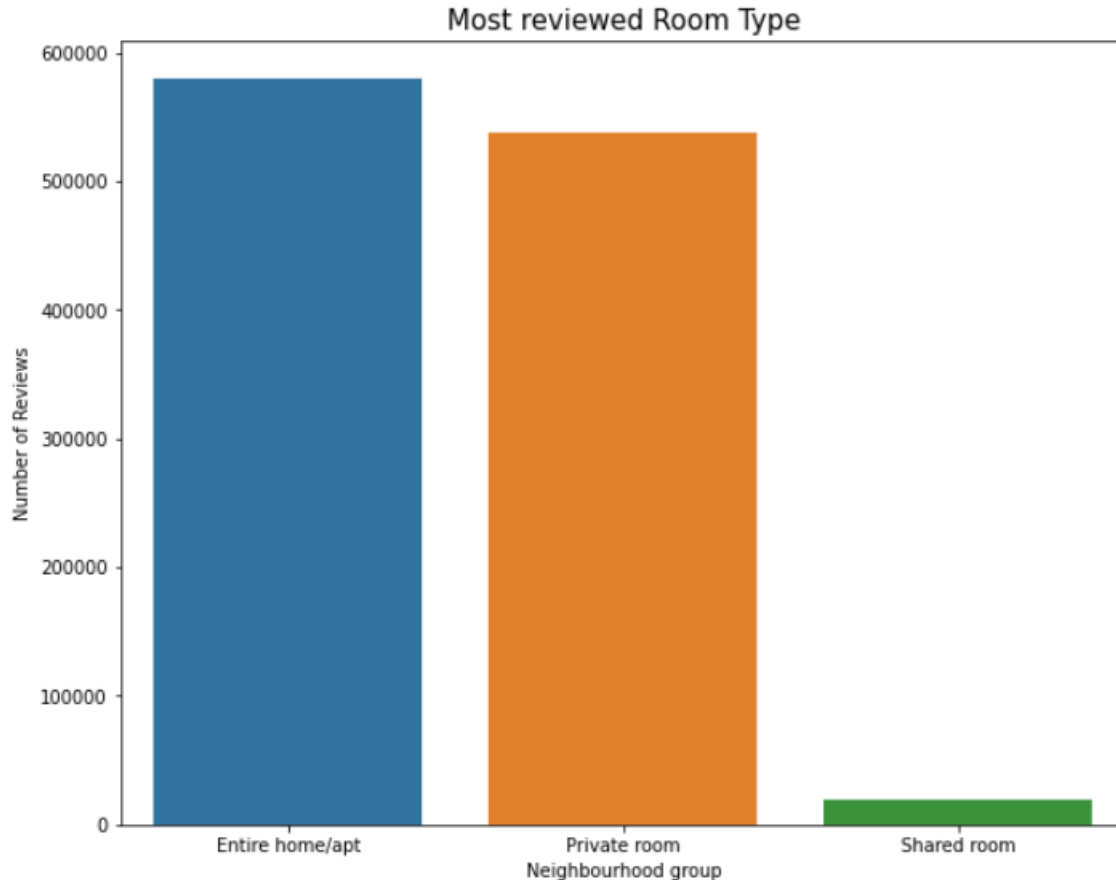


Insights :

- The Brooklyn region received the maximum number of reviews per month followed by Manhattan , Queens , Bronx and Staten Island.
- The graph also signifies that Brooklyn region may have the most number of listed properties preferred by the customers .

Exploratory Data Analysis

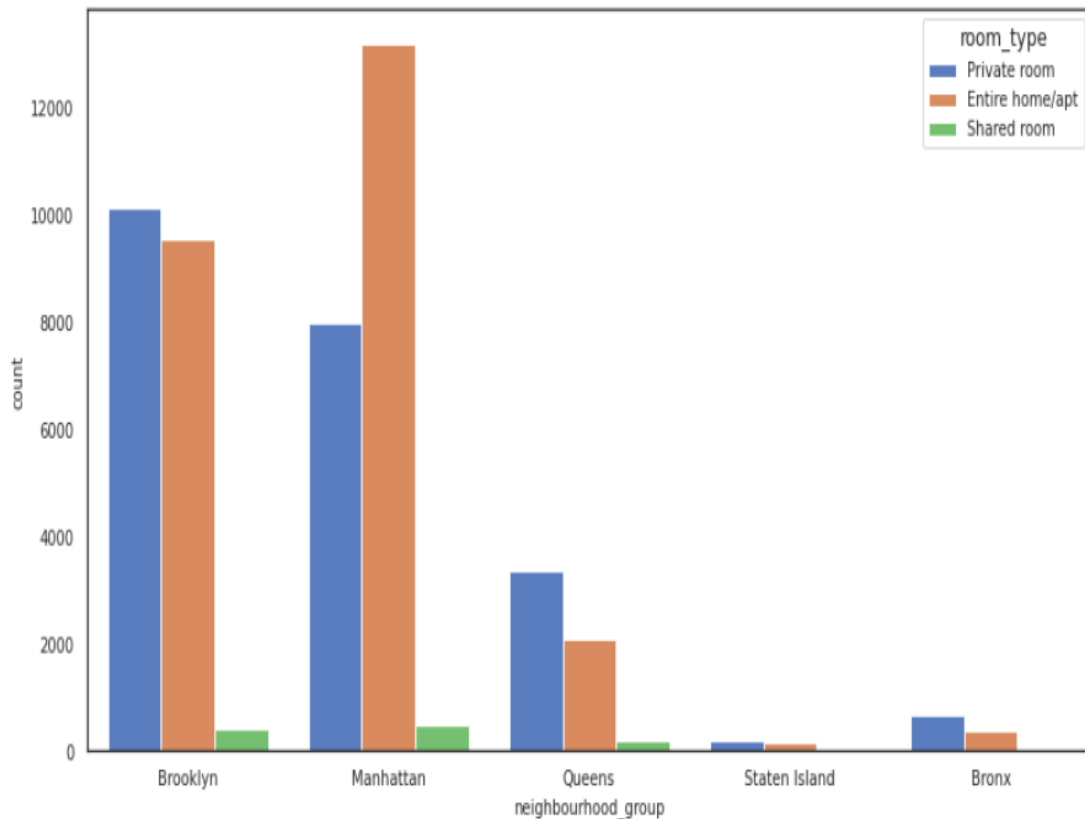
Most reviewed Room Types



Insights :

- Entire home/apartment is the most reviewed room-type followed by private & shared room-type.

Room-Types in the Neighbourhood Group

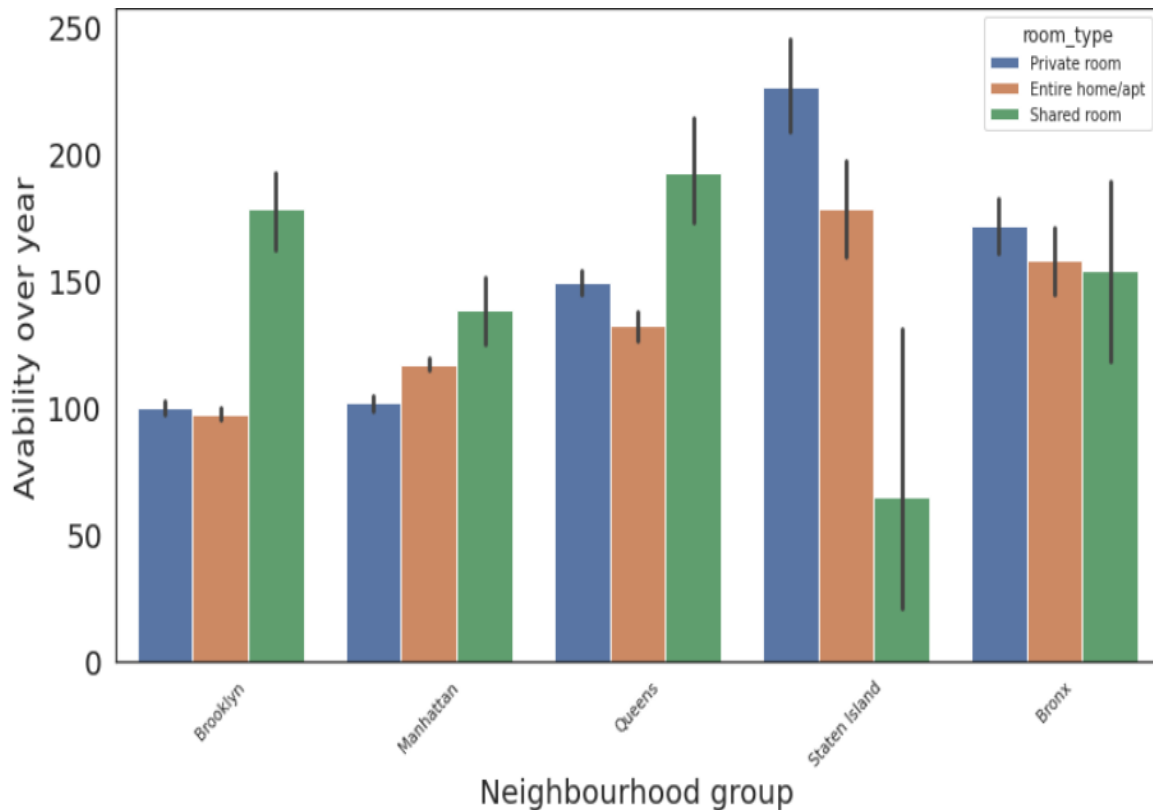


Insights :

- In Brooklyn & Queens region, private room-type is the most preferred type , followed by entire-house/apartment.
- In Manhattan , Entire house/ apartment is the most preferred type followed by private rooms and shared rooms
- In Staten Island & Bronx region private room-type is most , whereas shared room-type is the least preferred.

Exploratory Data Analysis

Availability of room-type by different neighbourhood groups



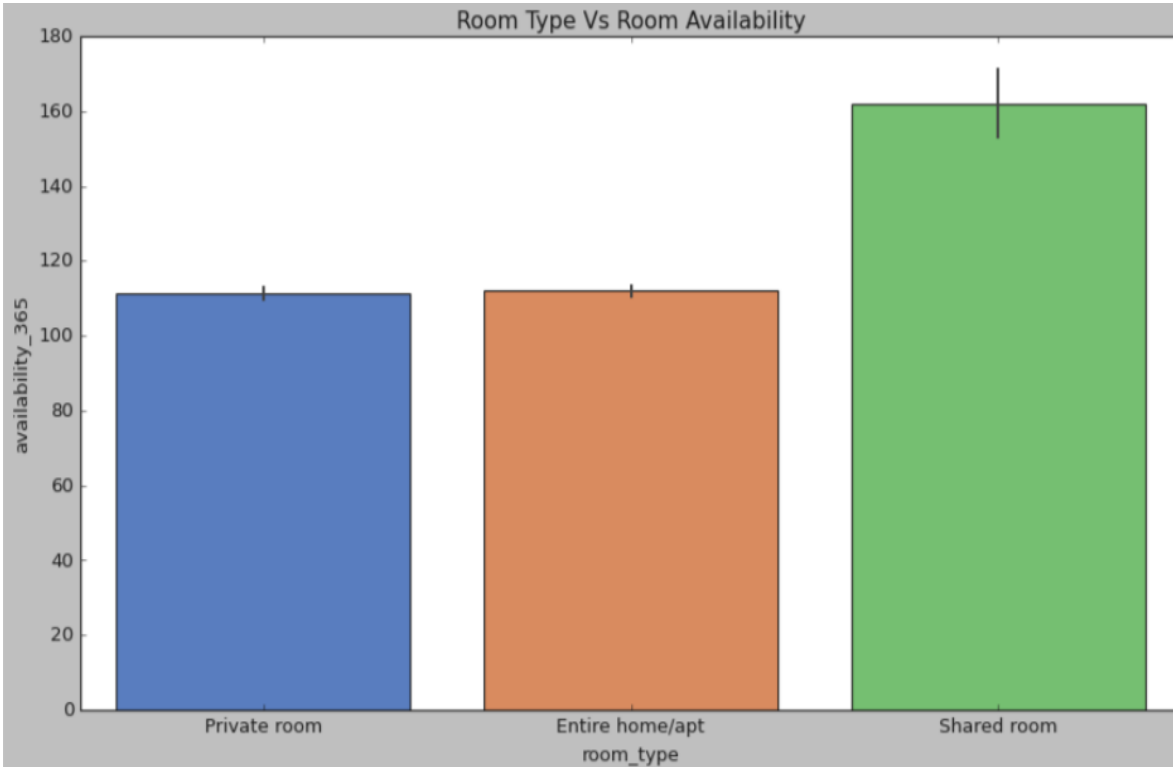
Insights :

- Private room-types have the most number of availability in the Staten Island region.
- Entire house/ Apartment room-type have the most number of availability in the Staten Island region.
- Shared room-type have the most number of availability in the Queens region.

Exploratory Data Analysis



Room-type V/s Availability

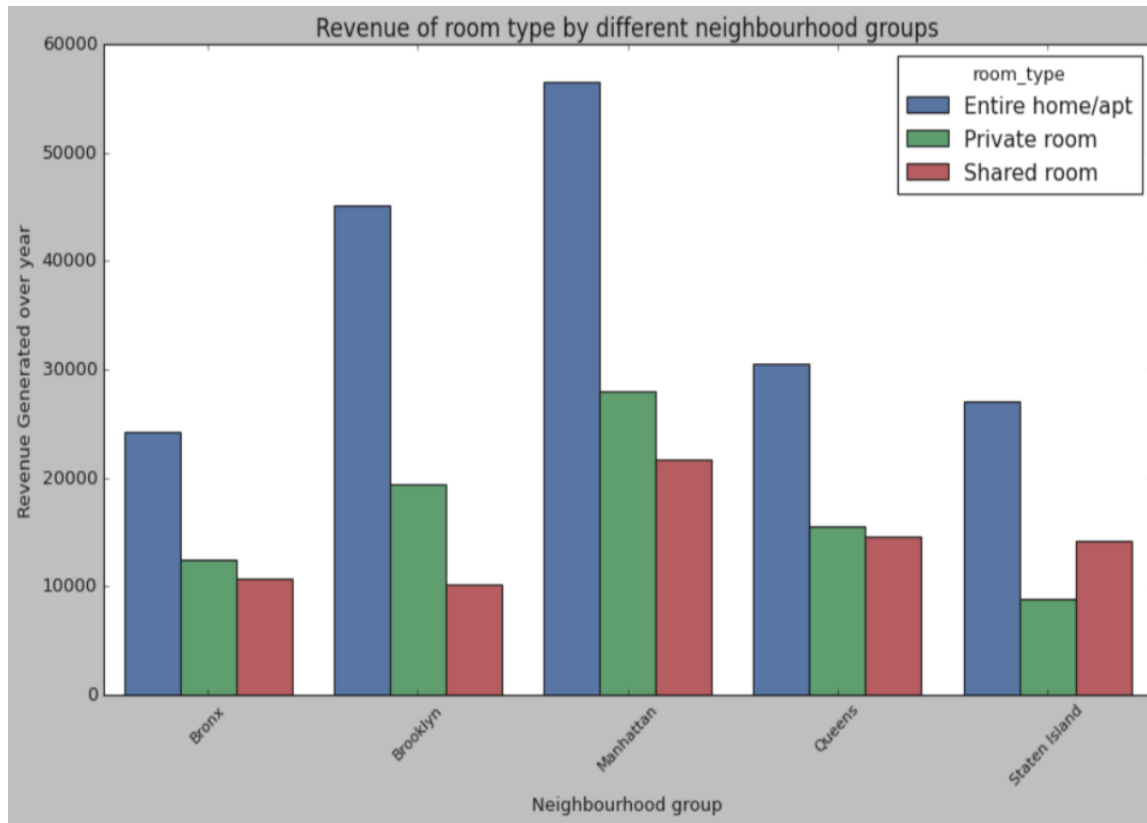


Insights :

- Shared room-type have the most number of availability throughout the year followed by Entire house/ apartment and Private rooms.

Revenue of room-type by different neighbourhood groups

Revenue Calculation per room-type: `df["revenue_generated"] = (365-df["availability_365"]) * df["price"]`



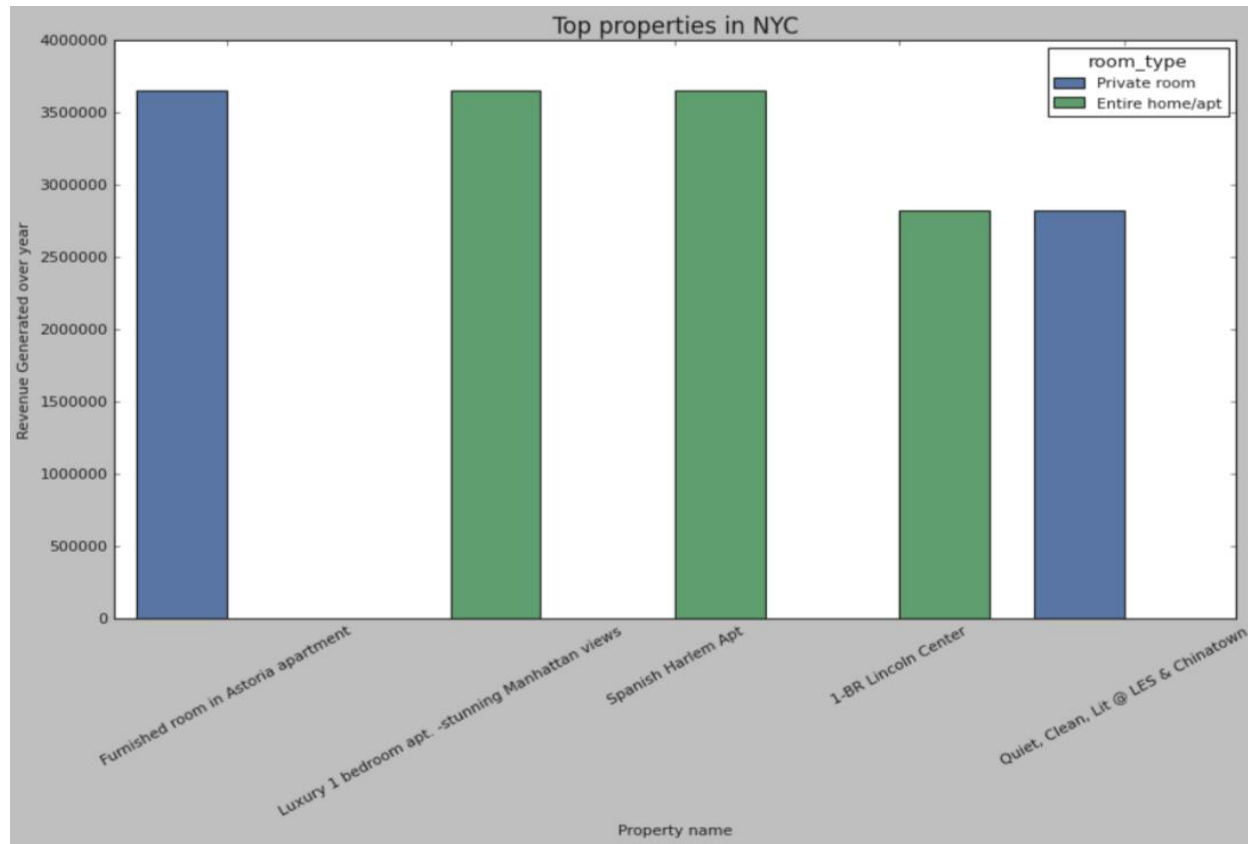
Insights :

- All the three room-type found to be the top-most revenue generating segment for Manhattan region.
- Hence, Manhattan region generates the maximum revenue across all the regions in NYC, followed by Brooklyn , Queens , Bronx & Staten Island.

Exploratory Data Analysis



Top properties in NYC



Insights :

- The 'Furnished room in Astoria apartment' and 'Luxury 1 bedroom apt. – stunning Manhattan views' are the top-most revenue generating properties across different regions for private and entire home/apartment room-type respectively.

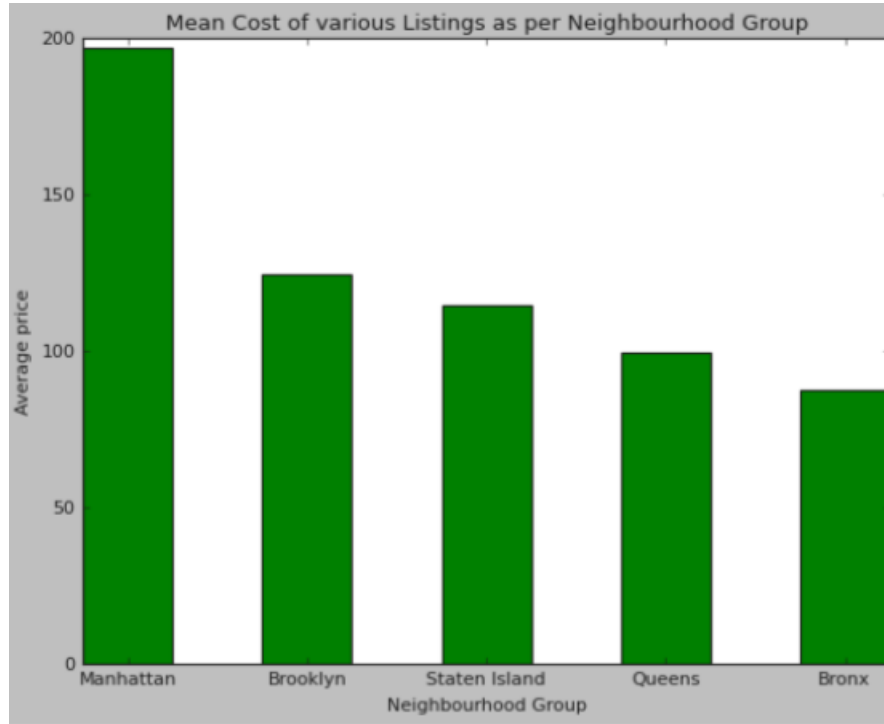
Mean Price of various Property Types



Insights :

- The Mean Price for Entire home/ apartment across all regions in NYC is maximum in the given room-type group.

Mean cost of various listings as per Neighbourhood Group



Insights :

- The mean cost for various listings is the maximum in Manhattan region .
- Hence, Manhattan region is the most expensive across all regions in NYC, followed by Brooklyn & Staten Island.

Exploratory Data Analysis

Most Affordable listings in NYC

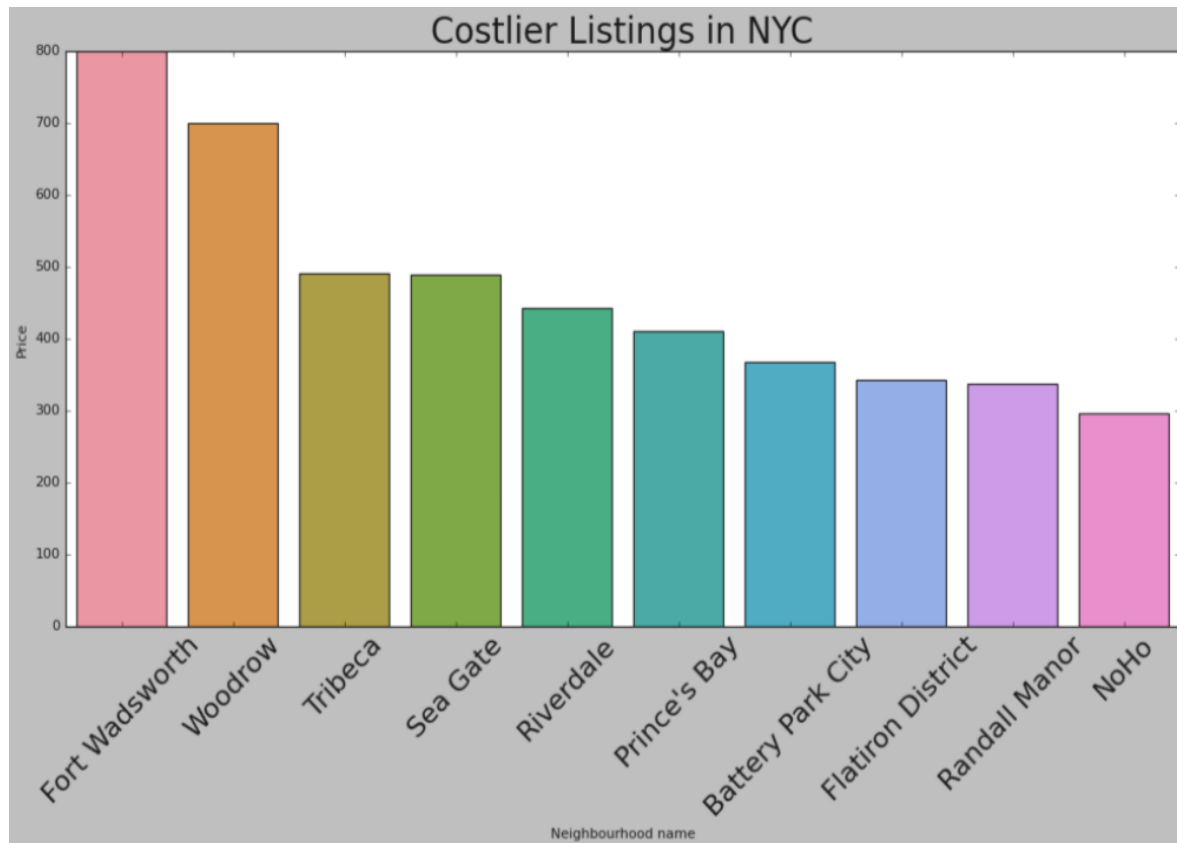


Insights :

- The property 'Bull's Head' is the most affordable across all the properties in NYC .

Exploratory Data Analysis

Most Costliest listings in NYC



Insights :

- The property 'Ford Wadsworth' is the most expensive across all the properties in NYC .

The given Airbnb dataset gives a great insight into prices , reviews and availability throughout the year .
The conclusions from the dataset are as follows:

1. Most people prefer either private rooms or entire house/apartment over shared rooms owing to greater privacy, comfort and better reviews.
2. Focus should be given on improving the bookings of shared-rooms , by providing better offers and implementing new marketing strategy.
3. Manhattan region generates the maximum revenue , even though it is the most expensive of all the regions in NYC.
4. More focus should be given towards improving the sale in Staten Island & Bronx region , by focusing on increasing the number of listings , providing better convenience and increasing the counts of shared room-type in these two regions as it appears to be the profitable segments.
5. Entire house/ apartment generates maximum revenue , due to good pricing with better facilities as compared to other room-types.

**THANK
YOU**