Contents

- **Problem Statement**
- **Understanding of Dataset**
- **Features Description**
- **Data Exploration**
- **Data Cleaning**
- **Exploratory Data Analysis**
- **Conclusion**

Problem Statement

- Since 2008, Airbnb is expanding on traveling possibilities and more unique way of experiencing the world

- Today it is one of a kind service having millions of listings around the world

- These millions of listings generate a lot of data that can be analyzed for various reasons like security, business decisions etc.

- The given data includes all needed information to find out more about hosts, geographical availability, necessary metrics to draw conclusions.

- Here we will try to analyze various problems like:

  - Which room type is prefered in most popular neighbourhood?

  - Percentage share of price among various neighbourhood groups.

  - Top 10 expensive properties in New York City (NYC).

  - And many more....

Understanding of Dataset

- The given data is of Airbnb Inc. and it is an American Company that lets property owners rent out their spaces to travelers looking for a place to stay.

- The dataset that we would be analyzing the booking information of Airbnb in New York City from 2008 to 2019.

- It also includes around 49000 records of properties in different neighbourhood group of the city. Mainly, Manhattan, Brooklyn, Queens, Bronx and Staten Island.

Features Description

- The features in the dataset can be described as follows:

- id - This is the identity number of the property listed by a particular host.

- name - It represent the name of the property listed by the host.

- host_id - It represent the identity number of the hosts registered on Airbnb website.

- host_name - These are the names of the hosts who have listed their properties on Airbnb.

- neighbourhood_group - These are the names of the neighbourhood groups situated in the NYC.

- neighbourhood - These are the names of the neighbourhood present in the neighbourhood groups in NYC

- latitude – It is coordinates of latitude of the property listed.

Features Description

- longitude – It is coordinates of longitude of the property listed.

- room type - This represent the various types of room listed by host.

- price - This is the rent of the property listed in USD.

- minimum nights - This represent the minimum number of nights customer should rent the property.

- Number_of_reviews - This represent the number of customers reviewed the property.

- last_review - This represent the date when the property was last reviewed.

- reviews_per_month - It is the count of reviews per month received by the property.

- calculated_host_listings_count - It is the number of listings done by a particular host.

- Availability_365 - This represent the number of days the property is available among 365 days

Data Exploration

- Checking the first 5 rows of the dataset.

Data Exploration

- Checking for Categorical Columns.

```
# Checking for categorical columns
cat_var = [i for i in airbnb_data.columns if airbnb_data[i].dtype=='O']
cat_var
```

```
['name',
 'host_name',
 'neighbourhood_group',
 'neighbourhood',
 'room_type',
 'last_review']
```

- Checking for Numerical Columns.

```
# Checking for numerical columns
num_var = [i for i in airbnb_data.columns if airbnb_data[i].dtype!='O']
num_var
```

```
['id',
 'host_id',
 'latitude',
 'longitude',
 'price',
 'minimum_nights',
 'number_of_reviews',
 'reviews_per_month',
 'calculated_host_listings_count',
 'availability_365']
```

Data Cleaning

- we find that the data in the column last_review and column reviews_per_month has 20.55% data missing. So we can simply drop those columns as we don't need those columns in our analysis.


```
airbnb_data.isnull().sum()
```

```
id                                0
name                              0
host_id                           0
host_name                         0
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
```

```
clean_df.isnull().sum()
```

```
id                                0
name                              0
host_id                           0
host_name                         0
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```
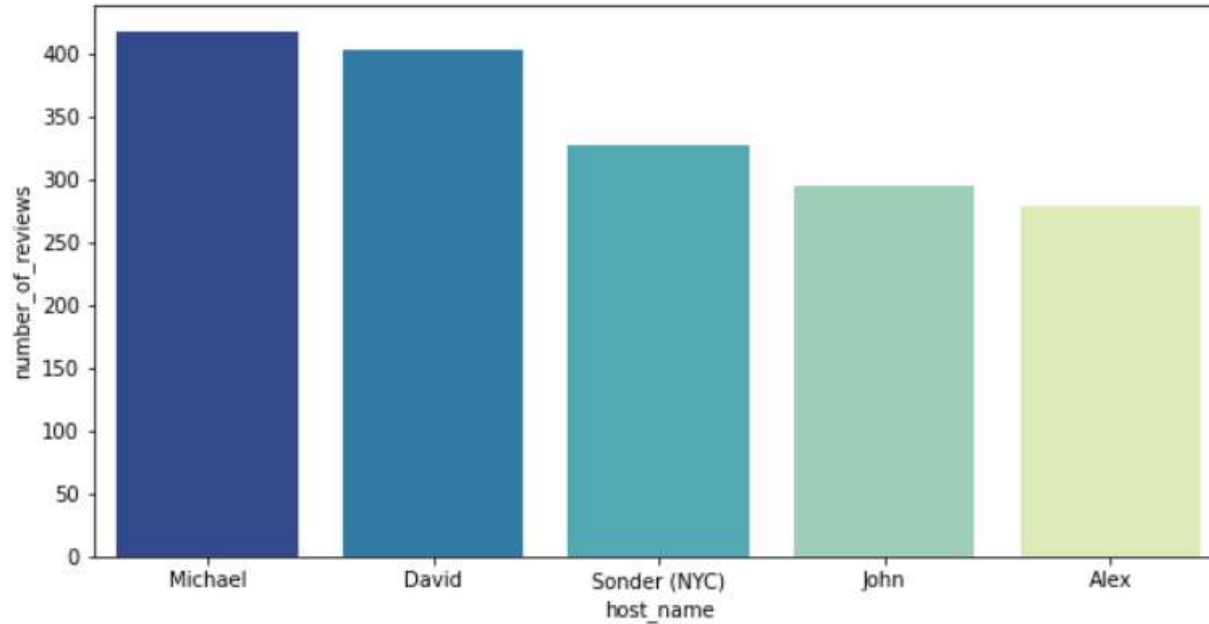
Top 10 hosts which has most number of listings.
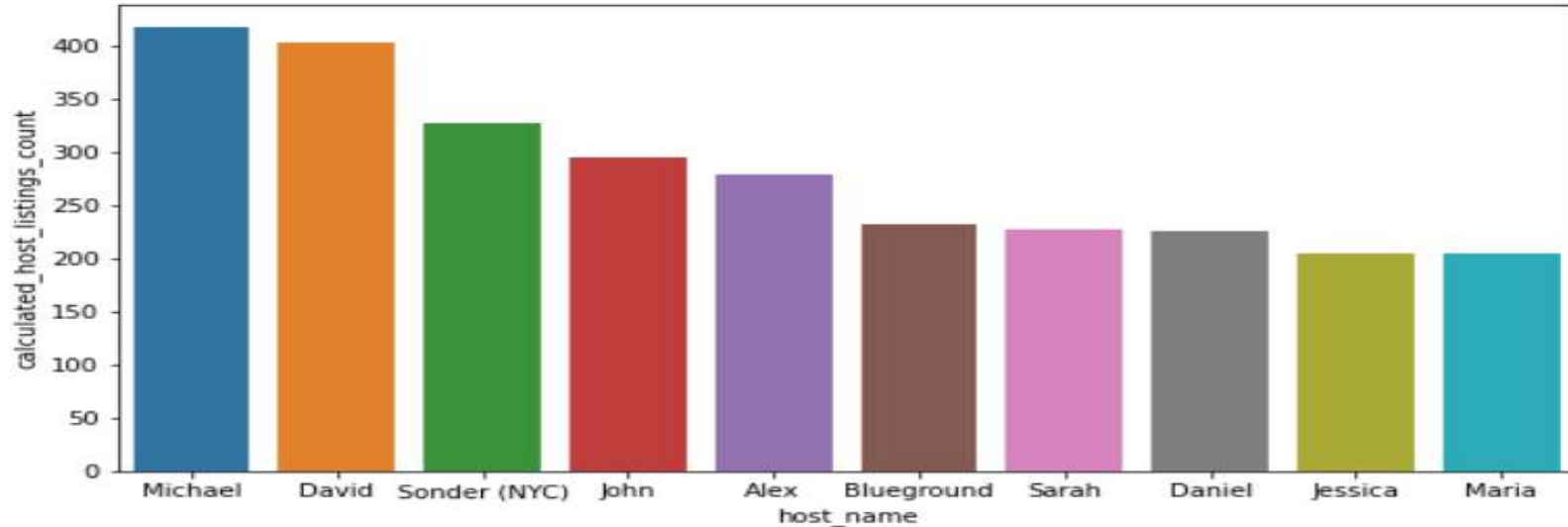
- In NYC, Michael is a host with most number of properties listed i.e. 417. Followed by David and Sonder(NYC) with 403 and 327 properties respectively.
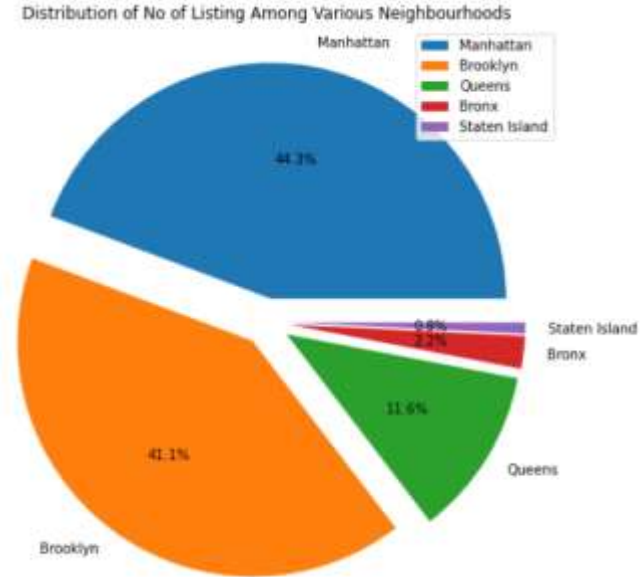
Top 5 host has maximum no of reviews.

- Host with most no. of reviews is Michael followed by David and Sonder (NYC)

Neighbourhood group with most no of listing.

- The Manhattan has highest percentage of listings i.e 44.3%. Among all five neighbourhood groups Manhattan and Brooklyn has dominance.
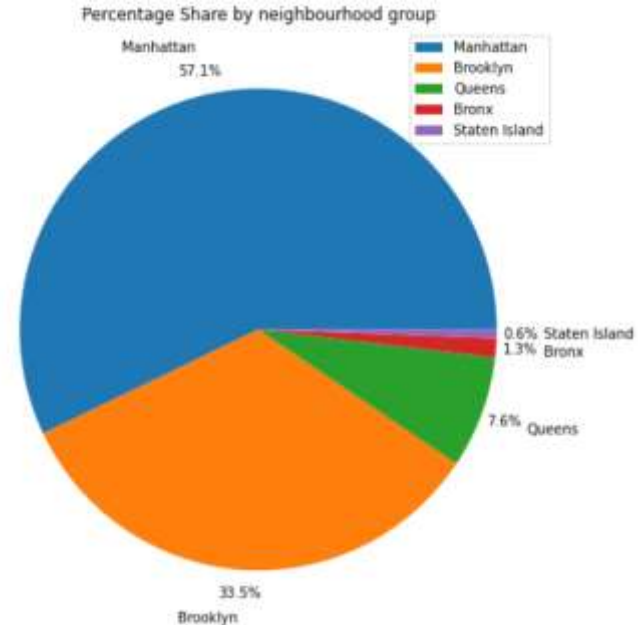

Distribution of No of Listing Among Various Neighbourhoods

Room type prefered in most popular neighbourhood.

- Private Room type is prefered by most of the people for stays.


Prefered Room Type in Popular Neighbourhood
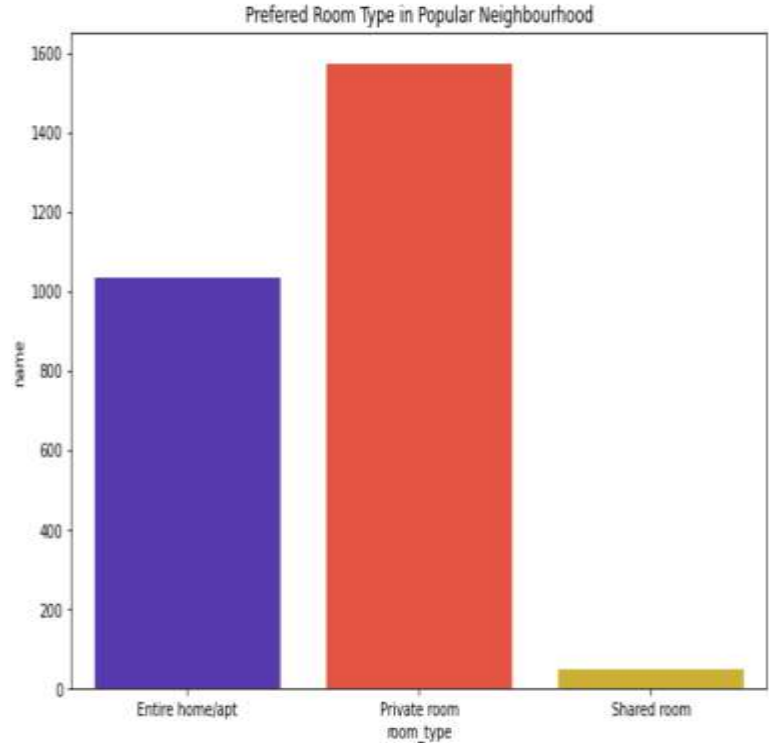
Percentage share of price among various neighbourhood groups.

- We clearly seen that Manhattan and Brooklyn has dominates the price share percentage and contribute to 90.6% of Total listing price.

Percentage Share by neighbourhood group

Manhattan 57.1%

Manhattan
Brooklyn
Queens
Bronx
Staten Island

0.6% Staten Island
1.3% Bronx
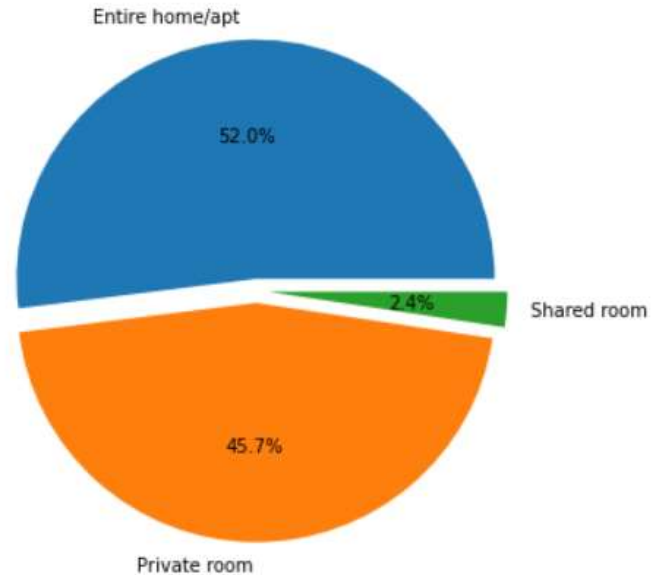7.6% Queens
33.5% Brooklyn

Averagely minimum_nights people stayed in each room_type.

- If someone is booking Entire home/apt , They tend to stay for longer duration on an average 8 to 9 days

Relationship Between Room Type and Average Night Stay
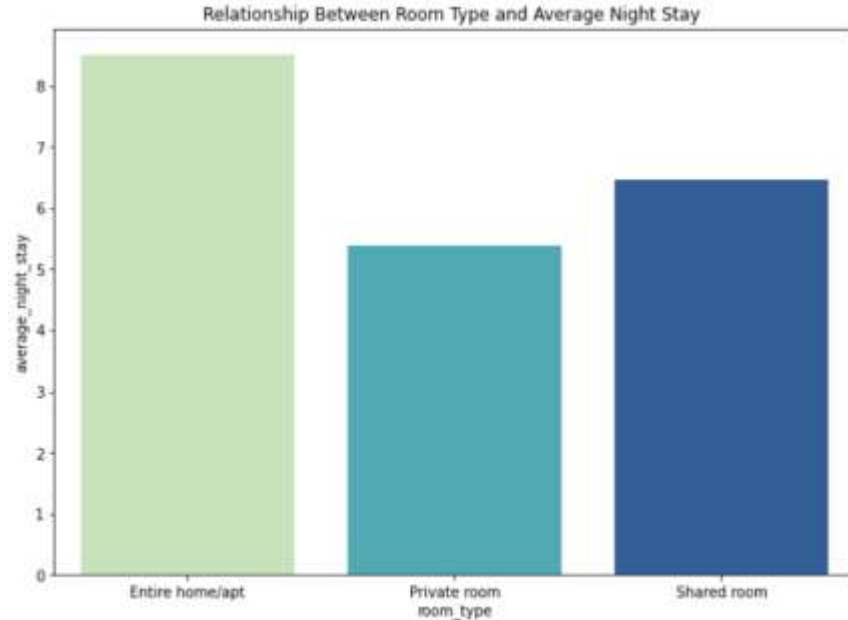
# Percentage Share of No. of properties of Room types across Airbnb(NYC).

- Entire Home/Apt is the highest contributor of Room type in terms of number of properties stands at 52.0% whereas Shared rooms contributing very little at 2.4%.

Percentage Share Properties Hold by Room Types

List down the Neighbourhood Groups in descending order of average price

- Manhattan is the costlier neighbourhood group to stay in with average price of 197 followed by Brooklyn , Staten Island , Queens and Bronx.
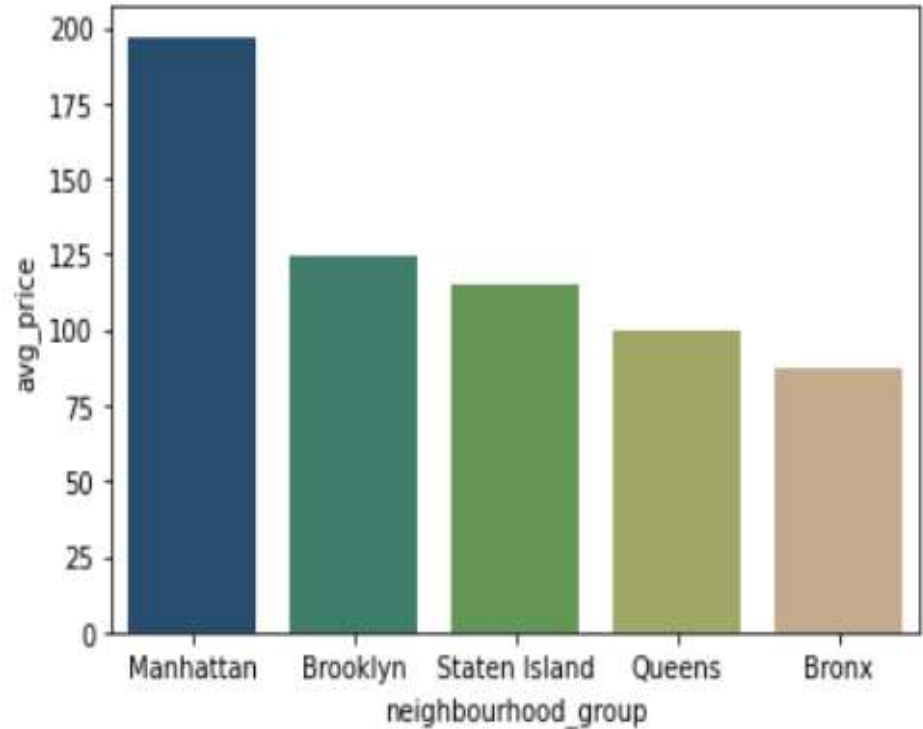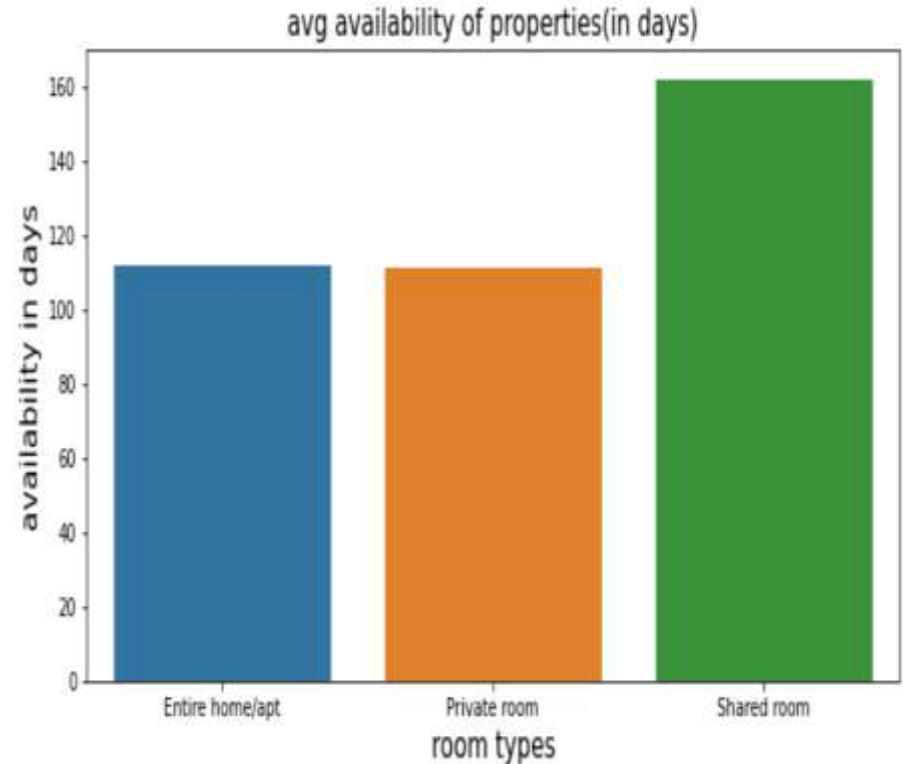
Top 10 expensive properties in New York City(NYC)

- There are three properties that are most expensive in NYC of price 10000 per night.

Top 10 Expensive Properties
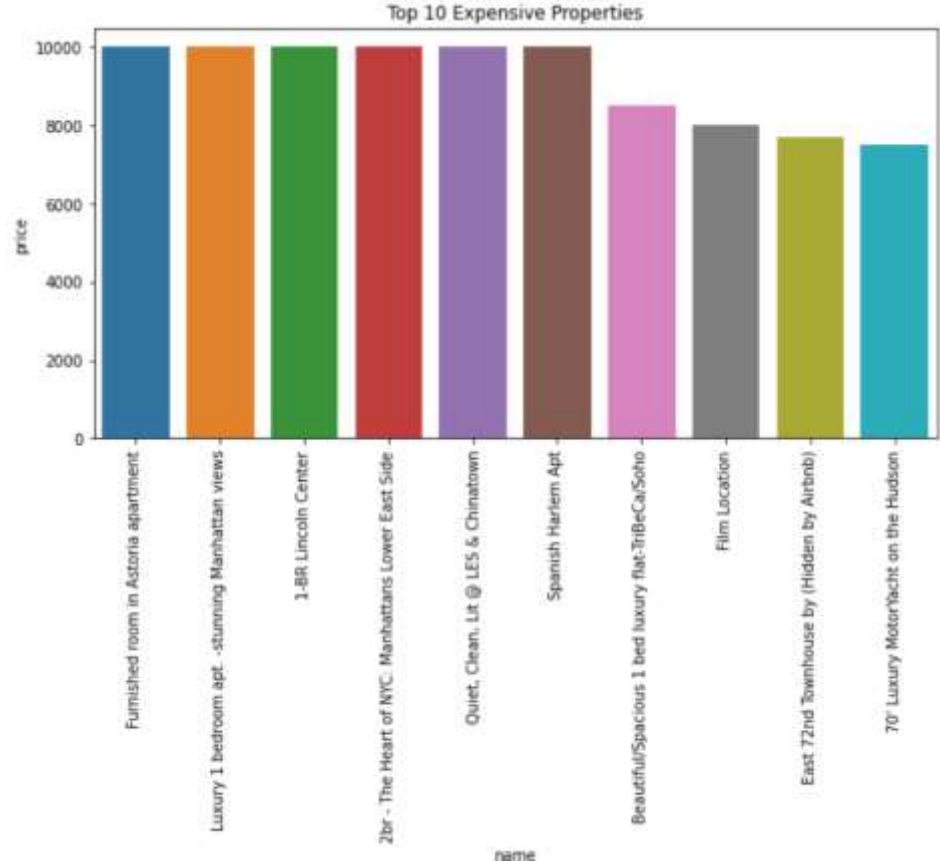
Average availability of properties by room types.

- Average availability of shared rooms is higher than others i.e 162 days out of 365.

avg availability of properties(in days)

Conclusion

- It is found that Manhattan has highest number of listings and probably the most active neighborhood group in terms of hotels - contributing to 44.3% of total; followed by Brooklyn with 41.1%. Thus the price in those two area is also high having contribution of about 90.6%.

- In Manhattan the most popular neighbourhood is Harlem with 2658 listings and private rooms being most attractive room type among the travellers.

- Michael is a host with 417 listings having highest number of reviews. So we can conclude that Michael is probably host with best and wide variety of properties across NYC.

Conclusion

- As Airbnb eases our struggle of finding best and affordable stays on trips, we found that more listings are from type entire home/appartment which is around 52% of total listings in NYC, followed by private rooms (45.7% of total listings.)

- For other neighbourhood groups namely Queens, Bronx and Staten island there are not as many listing options available as Manhattan and Brooklyn. Probably because those areas are not as mainstream as others

Thank You !!