

Capstone Project

EDA on Airbnb Dataset

- Akshat Mishra

Points for Discussion

- Introduction to Airbnb Business and Airbnb Dataset
- Data Summary
- Problem Statements and their findings
- Findings and conclusions

About Airbnb Business & Dataset

- **AirBnB** - Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales. Airbnb offers people an easy, relatively stress-free way to earn some income from their property.
- **AirBnB Dataset** - This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

Understanding the Dataset

To be specific,
this dataset has 48895
rows and 16 columns.

```
✓ [4] # Getting the basic information of dataset
0s airBnB_Data.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
```

Details on different columns in Dataset

- Id – ID is the unique ID given to each listing in the dataset
- Name – Name of listings
- host_id – IDs associated with hosts
- host_name – Name of the hosts who listed their properties on AirBnB
- neighbourhood_group – List of neighbouring cities from where people listed their properties.
- Neighbourhood – Neighbourhood is the sub division of neighbourhood_group
- latitude – latitude of the listing
- Longitude – longitude of listings

room_type – Different types of rooms which are listed on AirBnB

Price – Price of rooms per night

minimum_nights – Minimum number of nights people can stay.

number_of_reviews – Number of reviews associated to the listings.

last_review – Last review given by users

reviews_per_month – Number of reviews per month

calculated_host_listings_count – Number of a listings of hosts

availability_365 – Rooms availability

Checking for null values

- In this dataset, we have null values only in few columns which are “name”, “host_name”, “last_reviews”, “reviews_per_month”.
- We don’t have any need of host name and name because we have host_id which will work as unique identity. Names may have duplicates but host_ids are unique.
- For “last_reviews” and “reviews_per_month”, we will change the null value when we feel the need to do that.

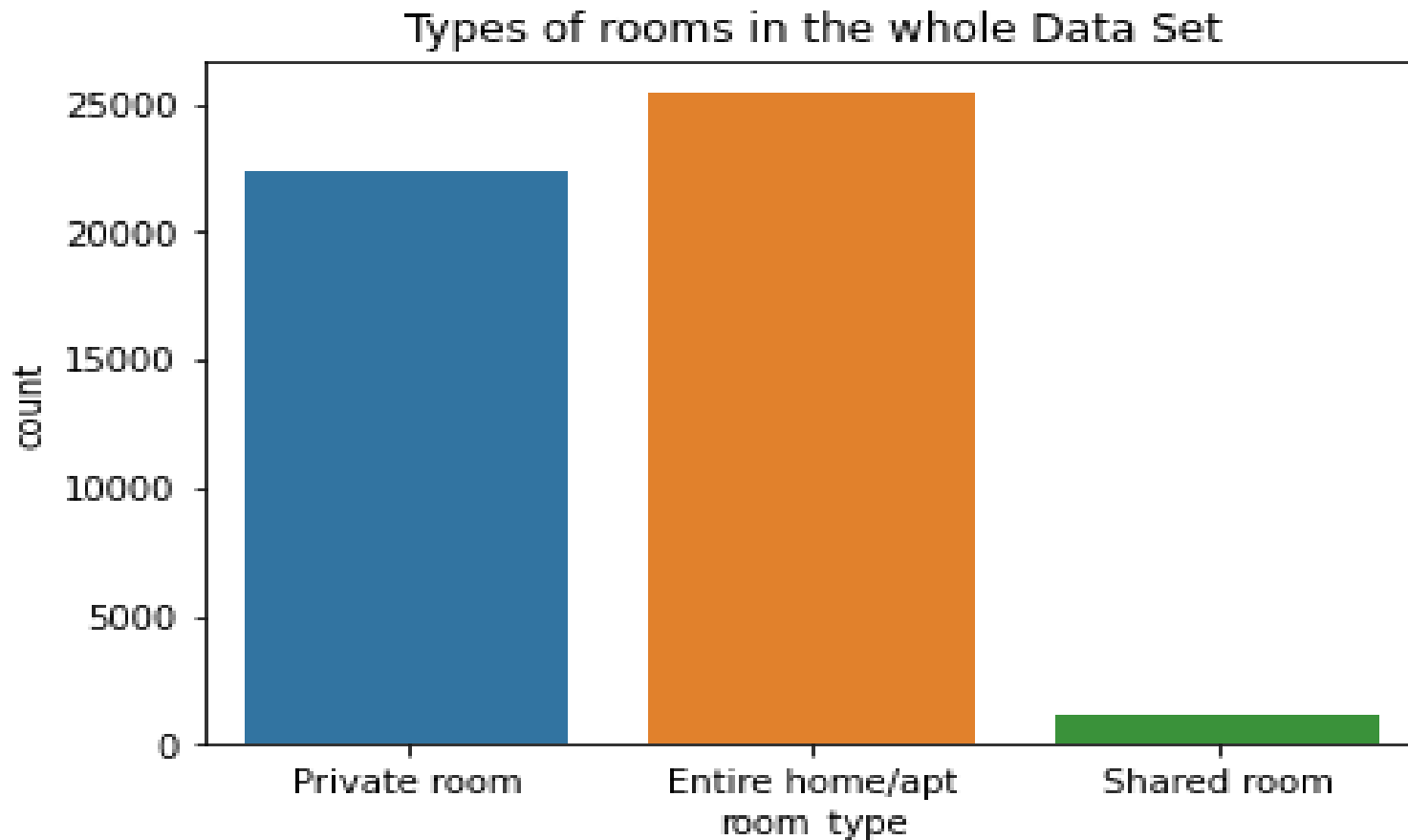
```
[5] # Calculating the total null values in the given dataset  
airBnB_Data.isna().sum()
```

```
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
```

Problem statement (Points to explore)

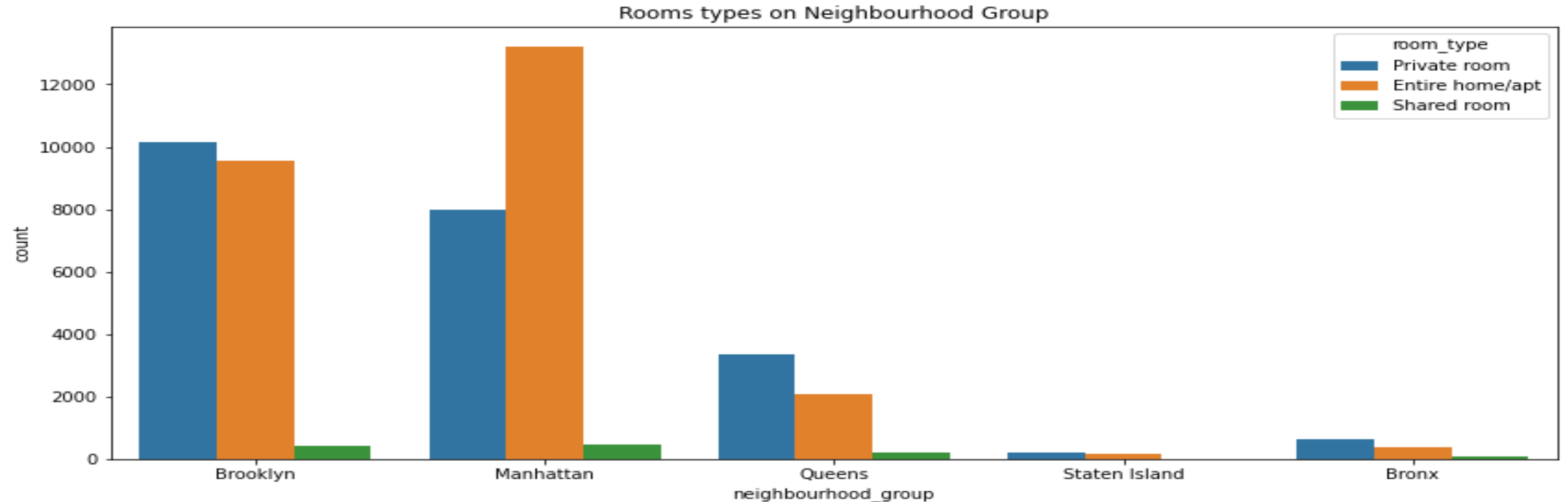
- What are the different room types in the whole dataset.?
- Different Room types on Neighbourhood Groups
- Top 10 neighbourhoods having maximum listings and their room types.
- Which room type is the most expensive.? And what is the average price of each room category.?
- What is the average price of rooms of all the neighbourhood groups.?
- Name of different neighbourhood groups and number of listings associated with them.?
- Hosts with maximum number of listings.
- Which are top neighbourhoods with respect to average price/day?

What are the different room types in the whole dataset.?



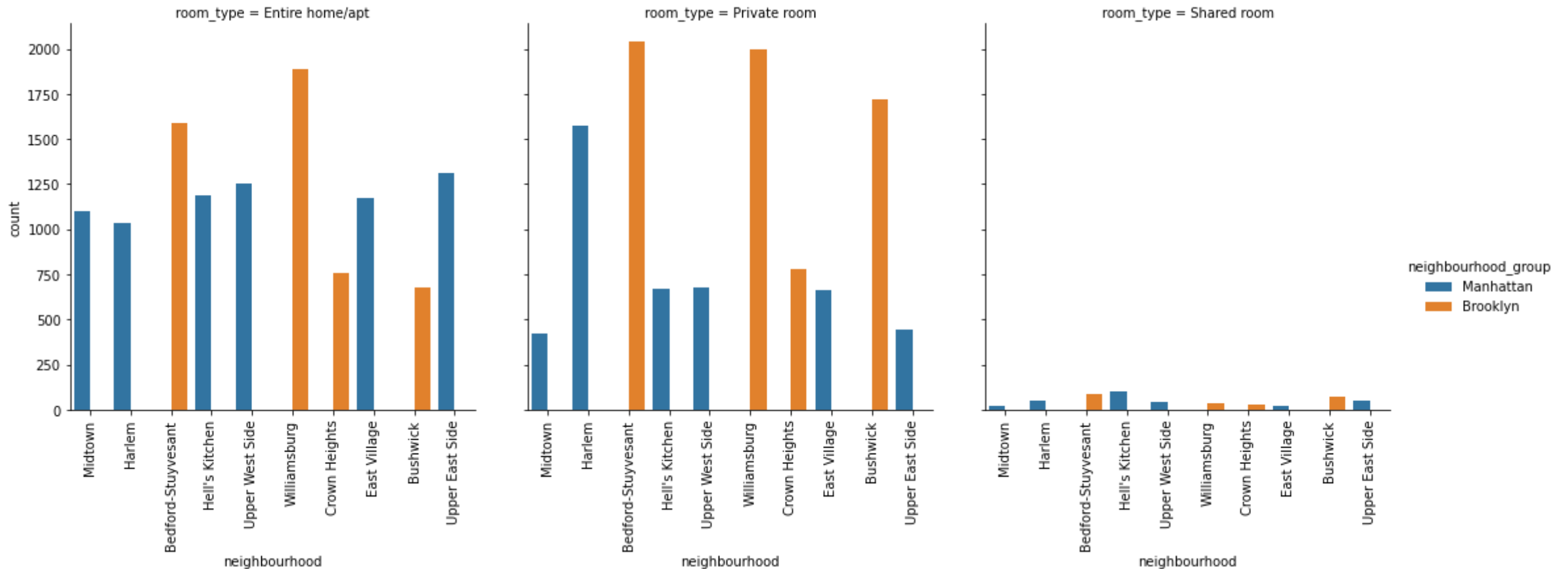
Comments - This shows that the entire home / apt category has the highest number of rooms followed by private rooms. Shared room is the least preferred category.

Different Room types on Neighborhood Groups



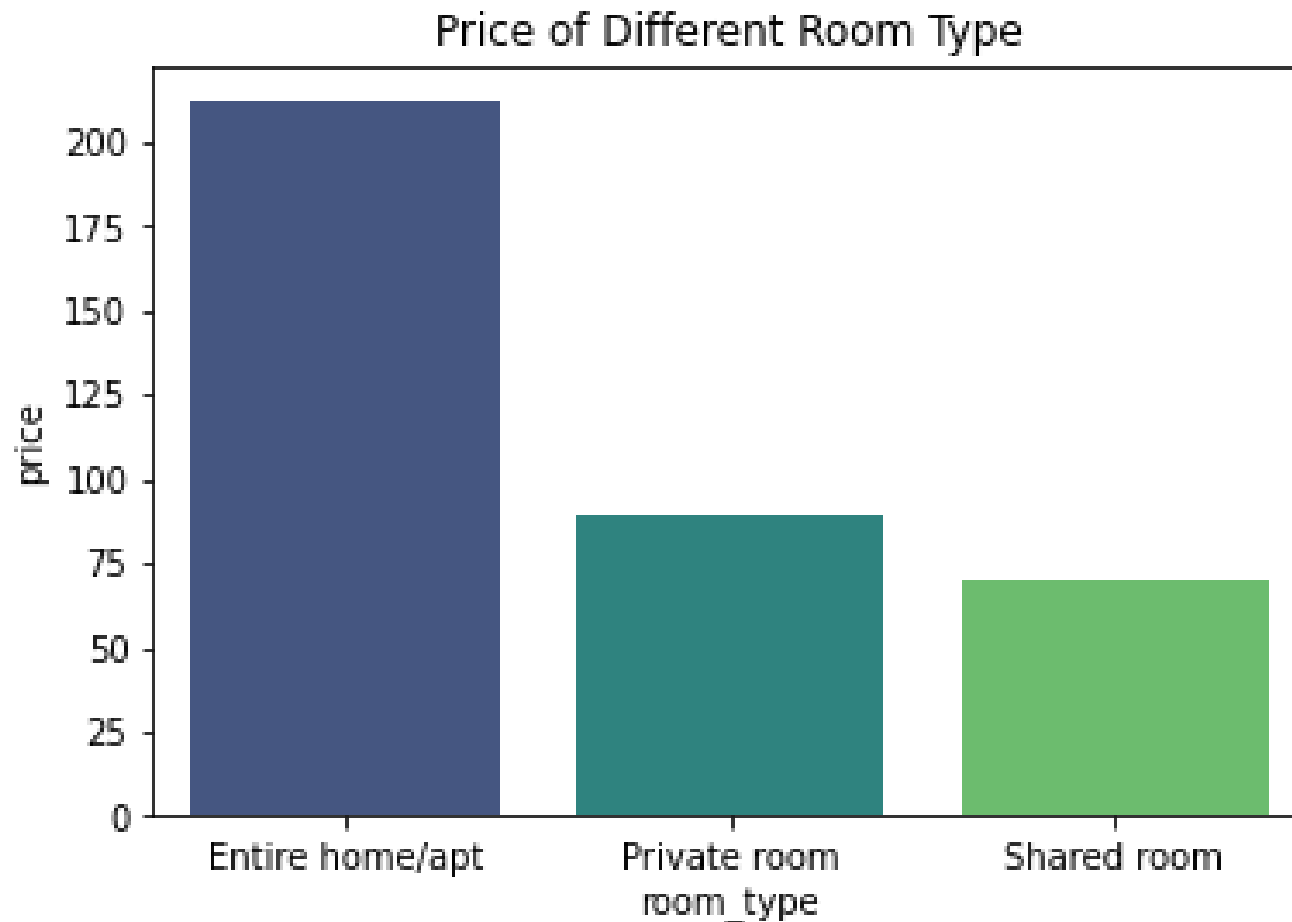
Comments - In Manhattan(neighborhood group), we have highest "Entire home / apt" category whereas Brooklyn has the highest number of private rooms. Staten Island is having the least number of listings in all three types of rooms category.

Top 10 neighborhoods having maximum listings and their room types.



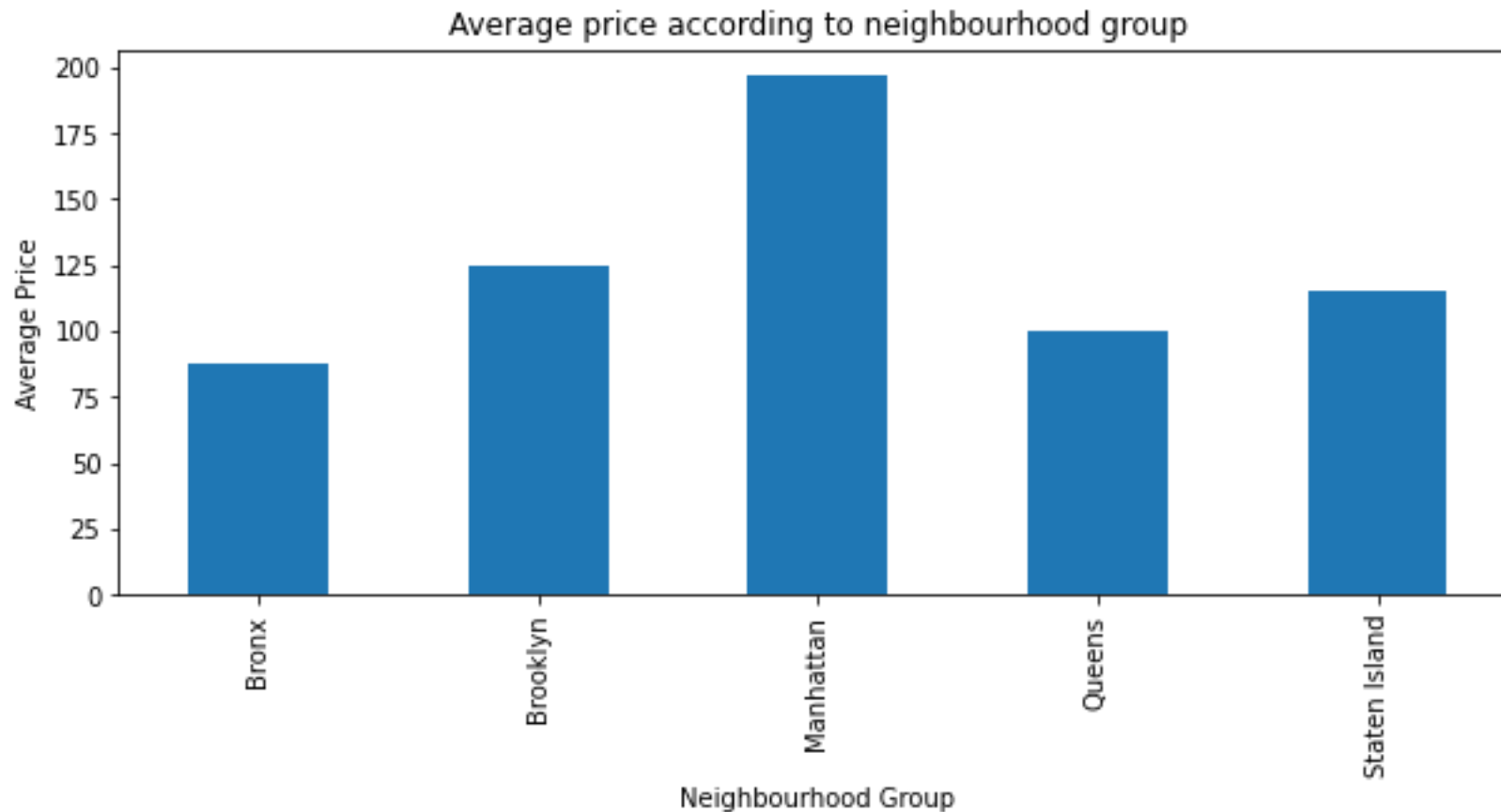
Comments - After calculating we got to know that top 10 neighbourhoods belongs to only 2 neighbourhood_groups which are "Manhattan" and "Brooklyn". And their associated room types can also be seen in the plot clearly.

Which room type is the most expensive.? And what is the average price of each room category.?



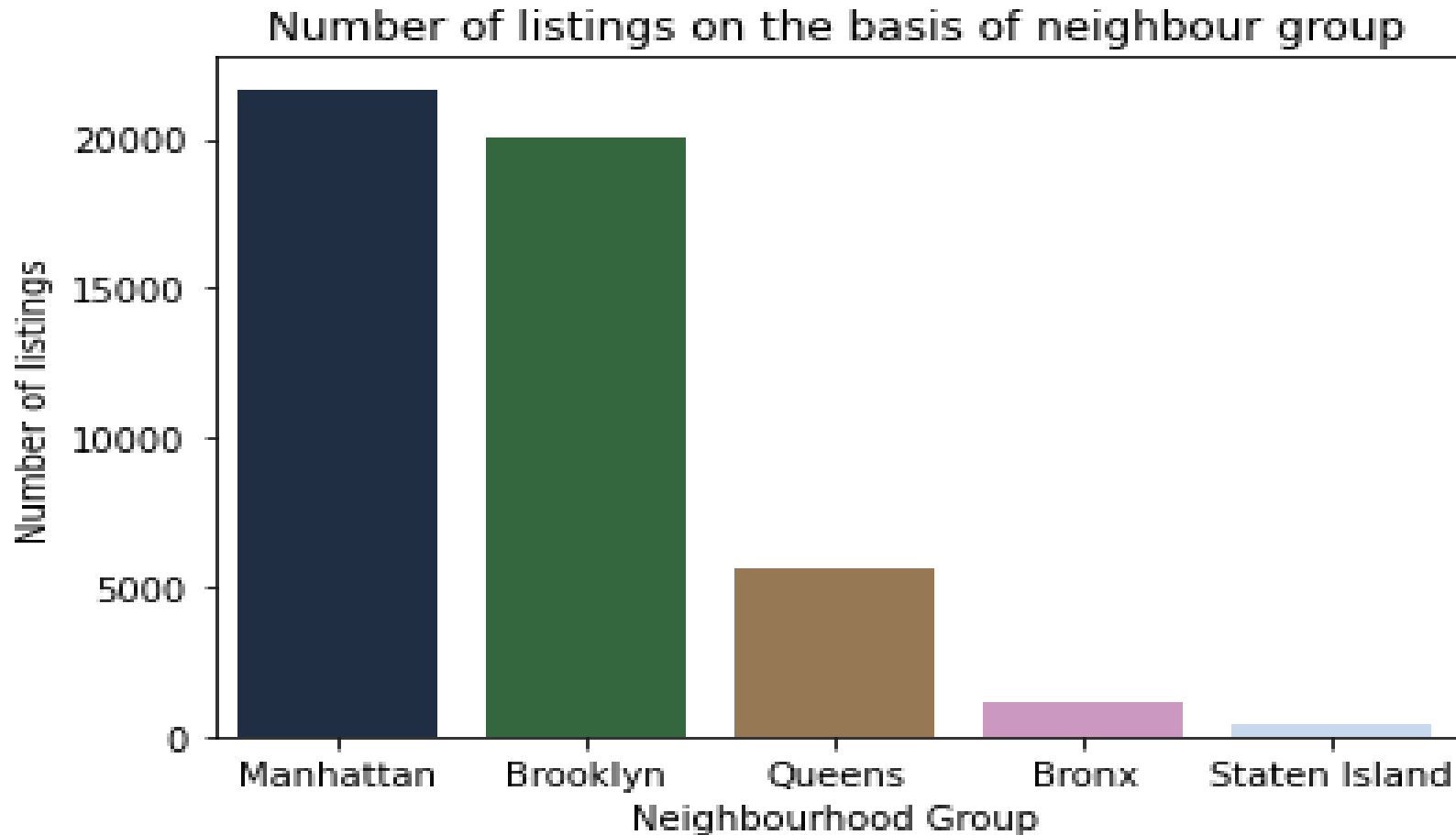
Comments - The most expensive category of rooms is the "Entire home/apt" followed by "Private room" and "shared room" is the least expensive comparing to others.

What is the average price of rooms of all the neighborhood groups.?



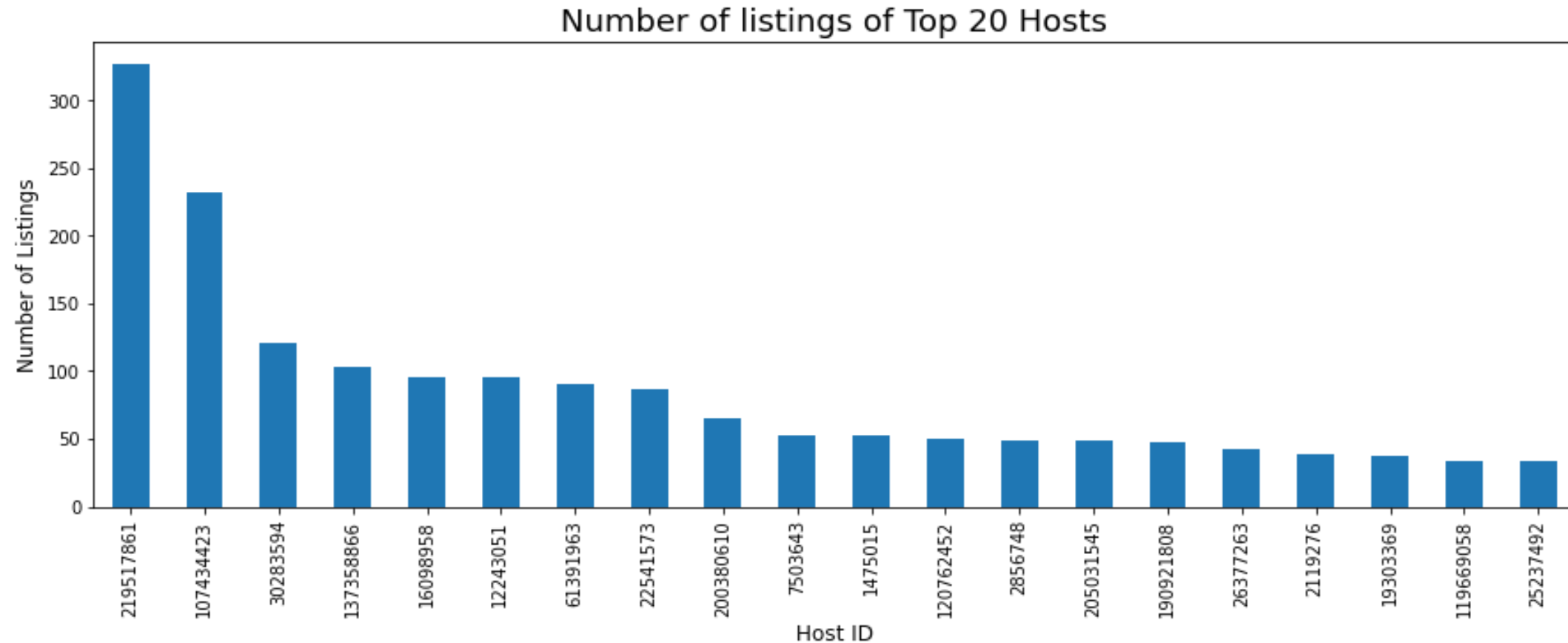
Comments - Highest room price is in "Manhattan" and the lowest is in "Bronx".

Name of different neighborhood groups and number of listings associated with them.?



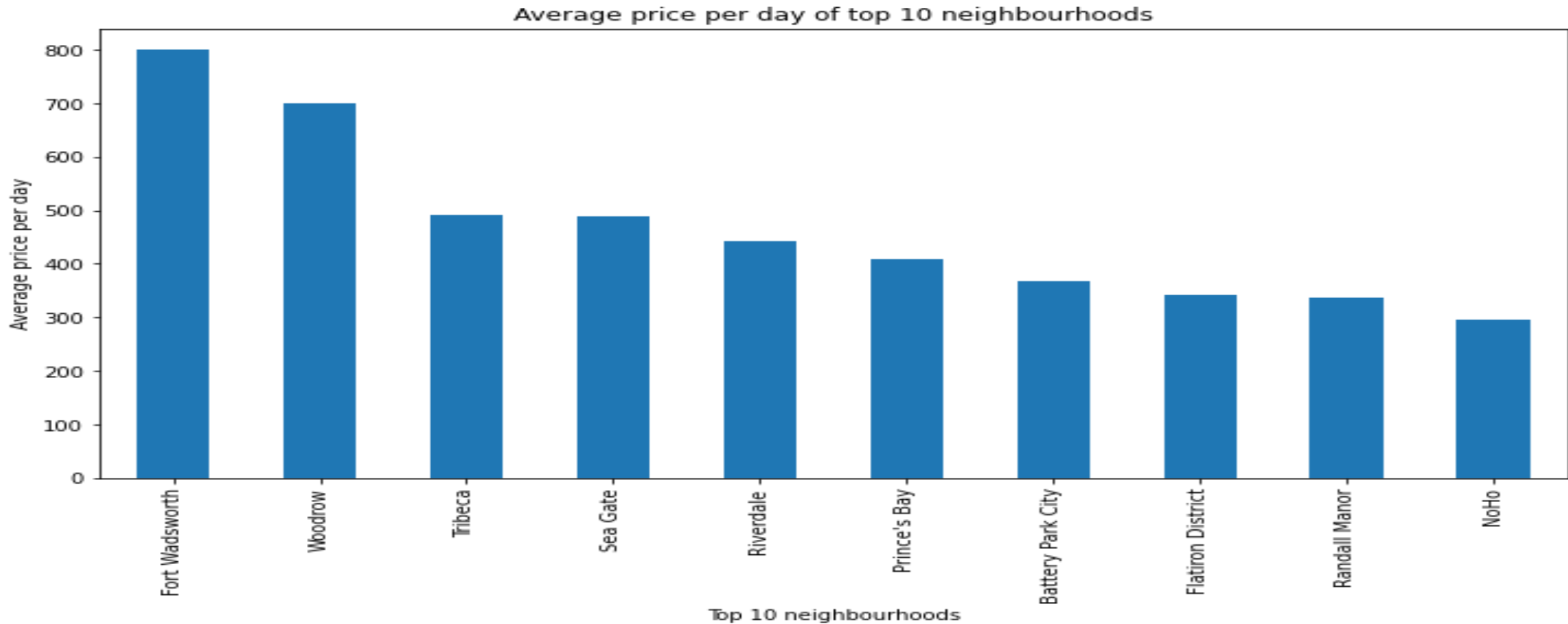
Comments - Manhattan, Brooklyn and Queens are the top 3 neighborhood groups in terms of total number of listings associated with them.

Hosts with maximum number of listings.



Representing top 20 hosts according to their number of listings in decreasing order.

Which are top neighborhoods with respect to average price/day?



Comments - Among all the neighborhoods under neighborhood groups "Fort Wadsworth" is the most expensive neighborhood where average per day price of rooms is highest in the whole data set.

Few more points can be drawn from the exploration -

On the basis of above exploration and analysis, following points can be considered -

- **"Neighborhood group - Manhattan"** : Most preferred location is "Manhattan".

Why ?

- In the whole Airbnb data set, most preferred room type is "Entire home/ apt". And the highest number of this room category is found in Manhattan. Which means people have more option for their desired room category in Manhattan that's why they prefer Manhattan.
- Average room price on the basis of neighborhood group, is highest in Manhattan. But if look at the bigger picture and try to figure out the most expensive rooms on the basis of neighborhood, we found that the most expensive rooms are in "Fort-Wadsworth" which is in Staten Island. And that is least preferred place.

- **"Neighborhood Group - Staten Island"** : Least preferred place is "Staten Island"

Why?

- Most preferred room category is "Entire home/apt" and Staten Island has the least number of rooms of this category.
- The average per day price on the basis of neighborhoods, in Staten Island there are very expensive neighborhoods.

Thank You