CAPSTONE PROJECT

airbnb Exploratory Data Analysis

Individual Project By

Vivek Tripathi

email- Vivektrip93@gmail.com

Githublink- https://github.com/dhiruvivek/Alma-Better

Content

❑ **Problem Statement**

❑ **Objective**

❑ **Data Summary**

❑ **Analysis Based on:**
- **Neighbourhood Group**
- **Price**
- **Room Types**
- **Number of Reviews**
- **Longitude and Lattitude**

❑ **Conclusion**

AI

Problem Statement-

AirBnb Bookings Analysis:

- EDA-Exploratory Data Analysis (EDA) is an approach to analyze data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

- Since 2008, guests and hosts have used Airbnb to expand on traveling 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 for company.

- These millions of listings generate a lot of data - data that can be analyzed 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.

OBJECTIVE

- Finding the total number of listings available in New York.
- Find the average price .
- Find the owner/hosts who has the most number of listings throughout New York.
- Find out the reason how price is affecting the maximum reviewed place.
- To see the price variation for different listings throughout New York.

Data Summary

The AirBnb data set contains 48895 records with 16 Variables.

**Total Columns that Dataset Contains:**
- id
- Name
- host_id
- Host_name
- neighbourhood_group
- Neighbourhood
- latitude
- longitude

- room_type
- price
- minimum_nights
- number_of_reviews
- reviews_per_month
- Last_review
- calculated_host_listings_count
- availability_365

Important Columns:

- **'name'-** These are names of the properties given by specific hosts.
- **'host_id'-** Many properties re being offered by many hosts. This is the specific id given to host.
- **'neighbourhood_group' and 'neighbourhood'-** These columns holds the information about the city and areas of the properties which are offered in here.
- **'Longitude' and 'Latitude'-** As the name suggests it just contains the longitude and latitude of the property location.
- **'Room_type'** = It displays the room_type of the property (either private room / entire home / shared room)

- **'price'-** Its an important column which holds the price value of all those properties
- **'number_of_reviews' and 'reviews_per_month' -** It clearly tells about the number of reviews and reviews per month for those properties and hosts hospitality.
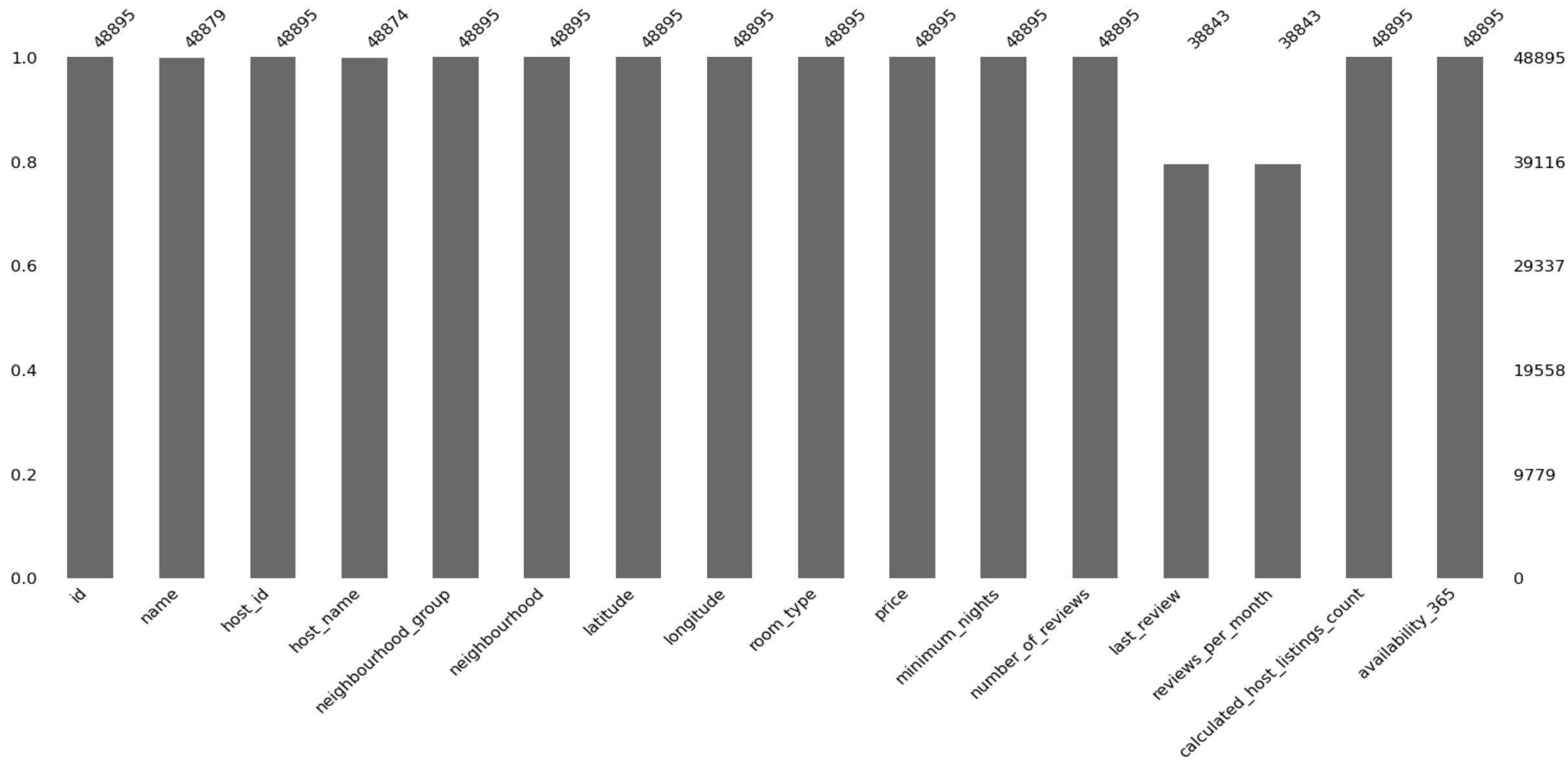
```
id                                  0
name                               16
host_id                             0
host_name                          21
neighbourhood_group                 0
neighbourhood                       0
latitude                            0
longitude                           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
```

There are 16 NA values in 'name' column, 21 NA values in 'host_name' column, 10052 NA values in 'last_review' & 'reviews_per_month' columns.

Treating NA

If we have a look at the number of null values throughout the data set, then we can observe that we found some of NA values in few columns.

- Last_review column contains 10052 records with null values.

  This column is not much of a help to analysis. So we drop that column.

- We then replaced the null values of 'reviews_per_month' column as 0.

```
name                              16
host_id                            0
neighbourhood_group                0
neighbourhood                      0
latitude                           0
longitude                          0
room_type                          0
price                              0
minimum_nights                     0
number_of_reviews                  0
reviews_per_month                  0
calculated_host_listings_count     0
availability_365                   0
```
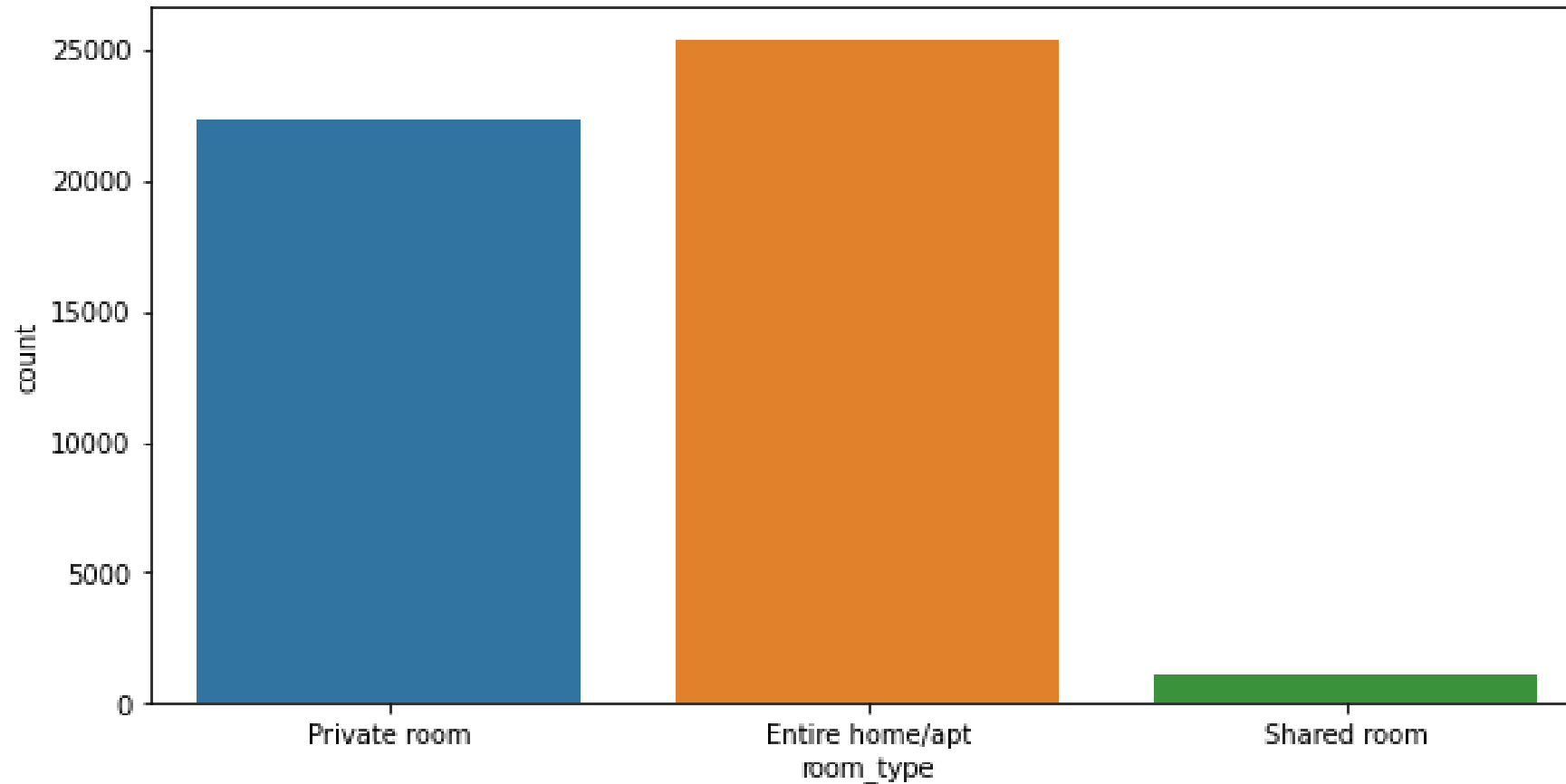
Outliers
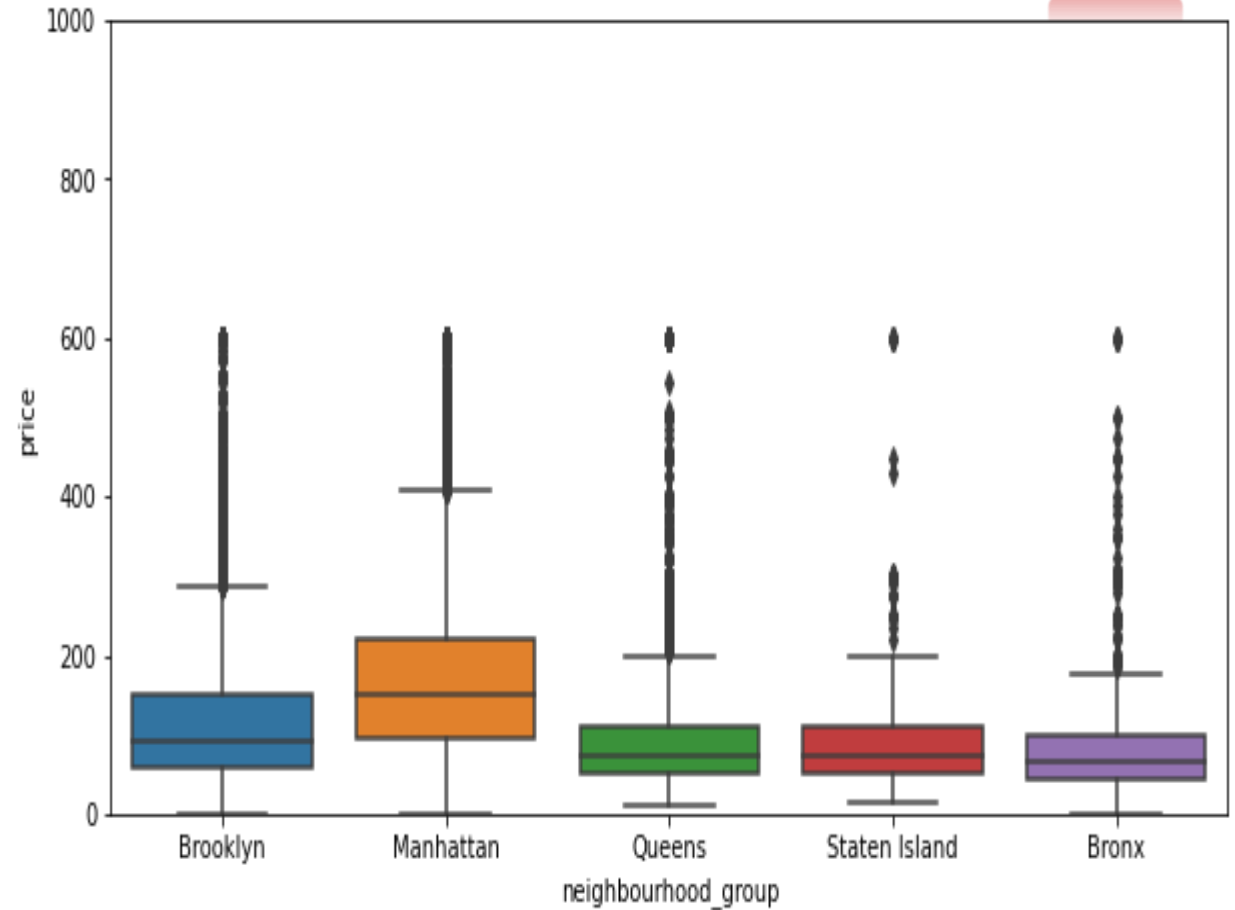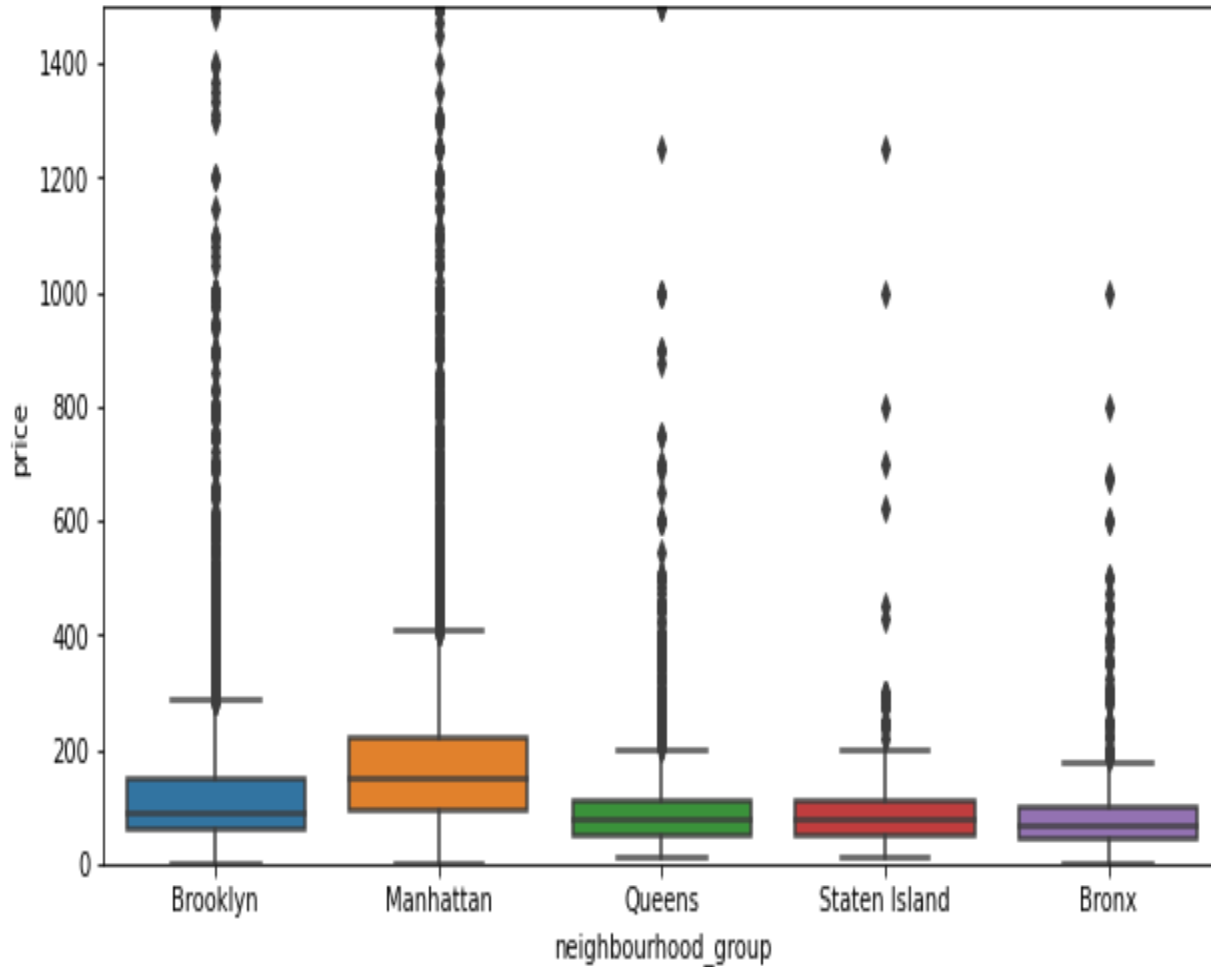
We observed there are only 1044 records which are having more than 500 as the price. So just taking the price of property which are more than 600 as 600, which will just minimize the number of outliers.
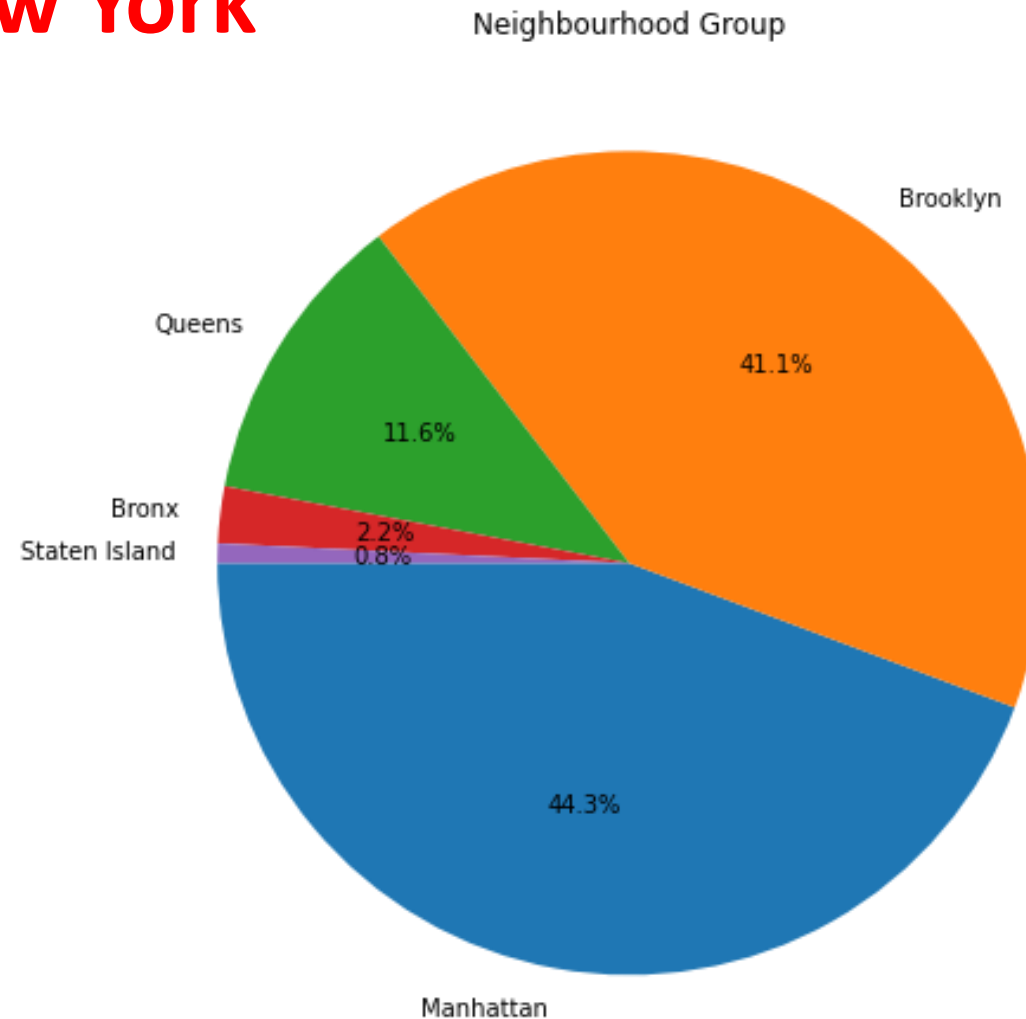
Neighbourhood Groups
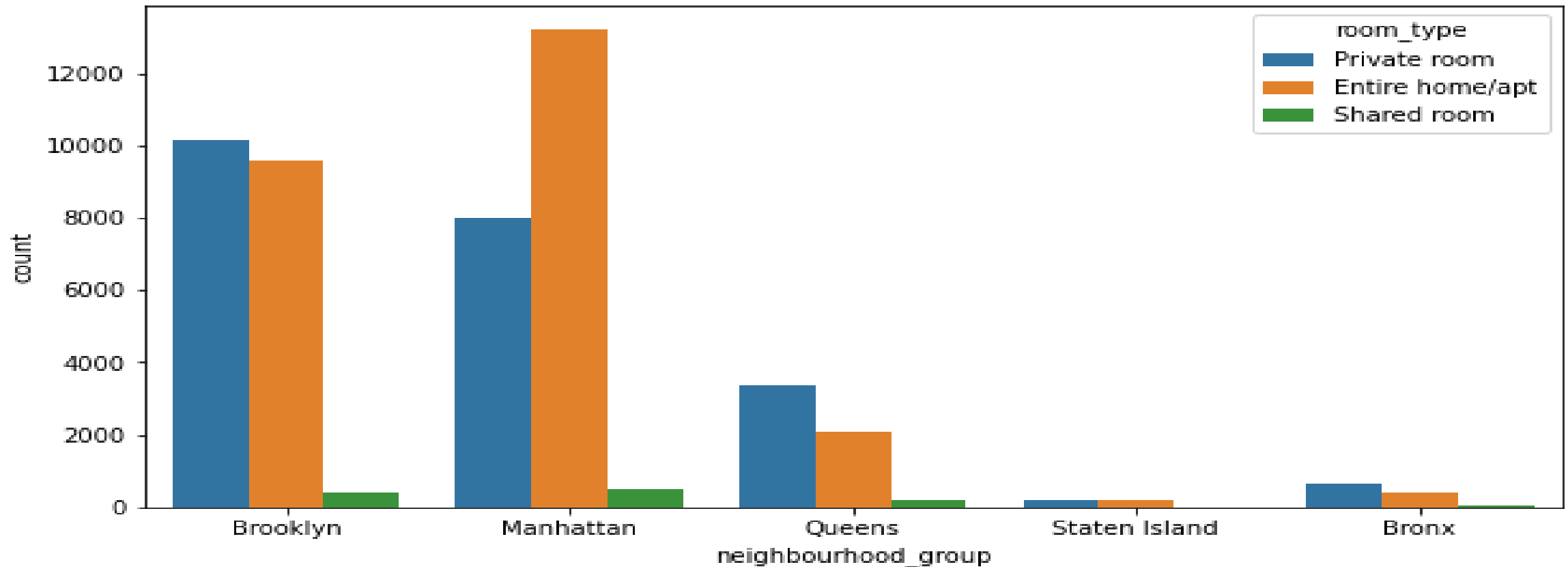
Number of properties available in different neighborhood groups is more in Manhattan(Home room type)
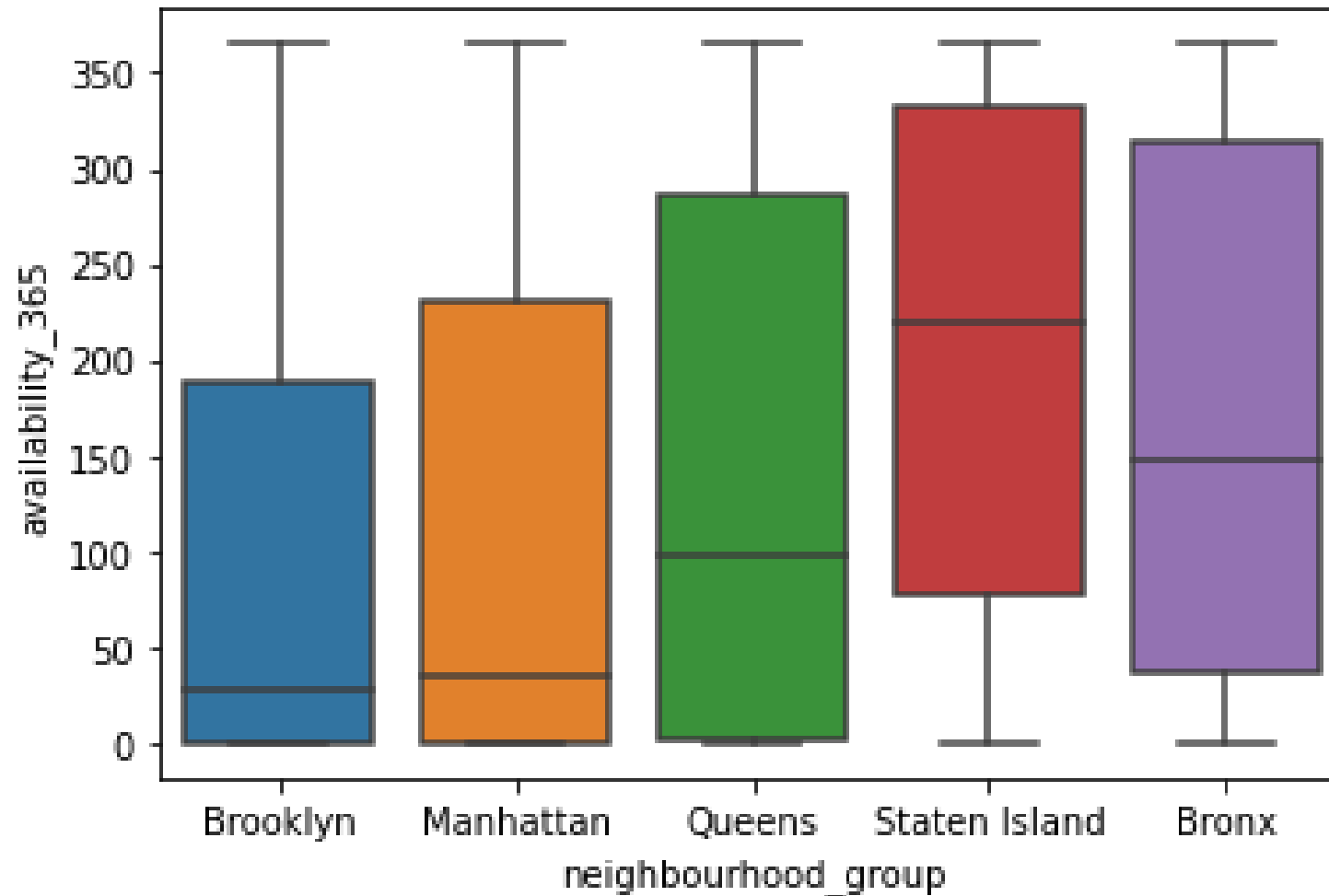
Room type throughout New York

Number of room types available in New York which are of the type Private room, Entire home/Apartments, Shared room.

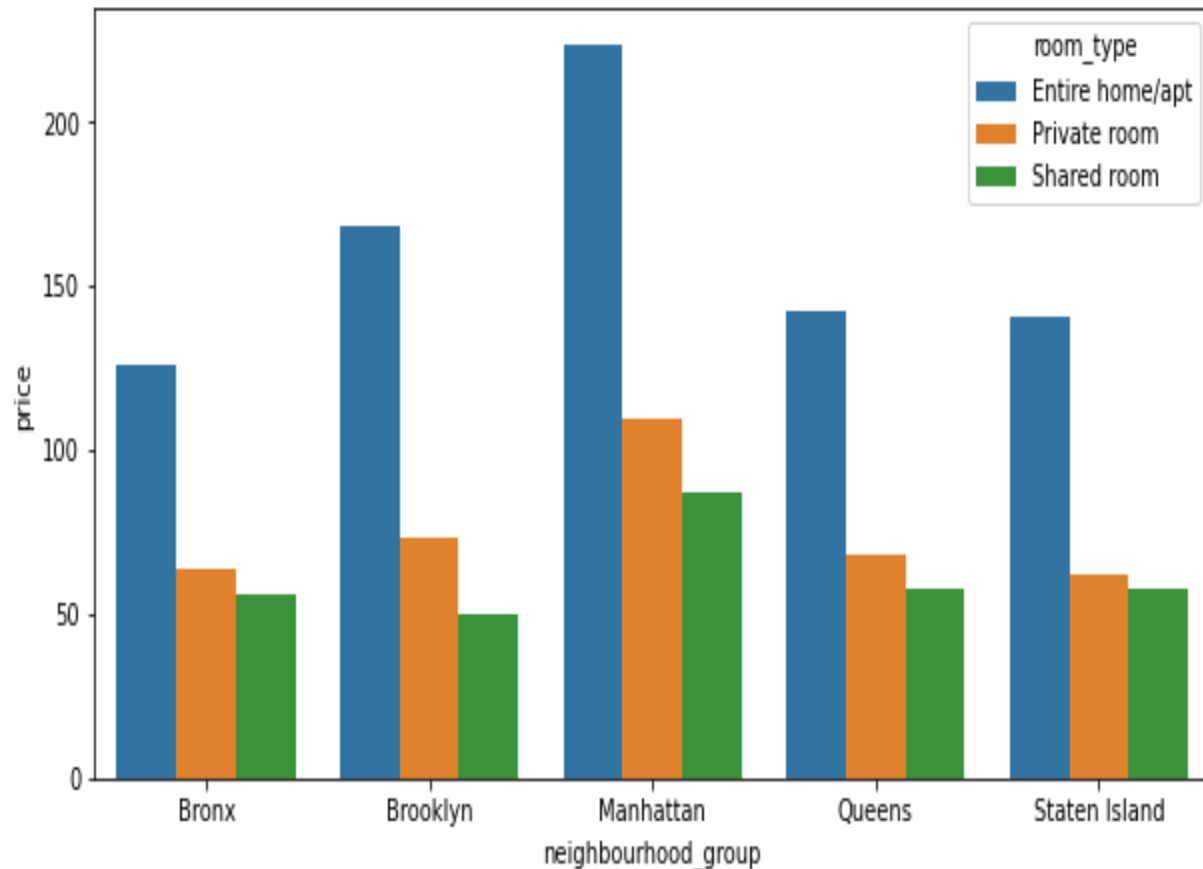# % of listings offered by each neighborhood groups in New York

Availability of the listings and neighborhood group

We can observe that the mean of listings count in Brooklyn, Manhattan and Queens are available only for 0 to 100 days a year.

Average price for each room type in neighborhood

	neighbourhood_group	room_type	price
0	Bronx	Entire home/apt	126.266491
1	Bronx	Private room	63.751534
2	Bronx	Shared room	56.466667
3	Brooklyn	Entire home/apt	168.437389
4	Brooklyn	Private room	73.328563
5	Brooklyn	Shared room	50.225182
6	Manhattan	Entire home/apt	223.726797
7	Manhattan	Private room	109.401027
8	Manhattan	Shared room	87.310417
9	Queens	Entire home/apt	142.643130
10	Queens	Private room	68.285291
11	Queens	Shared room	57.656566
12	Staten Island	Entire home/apt	141.034091
13	Staten Island	Private room	62.292553
14	Staten Island	Shared room	57.444444

If we compare the average price per stay in different neighborhood groups,
we can observe that its high in Manhattan for Entire home/ Apartment.

Average price for each neighborhood in NY
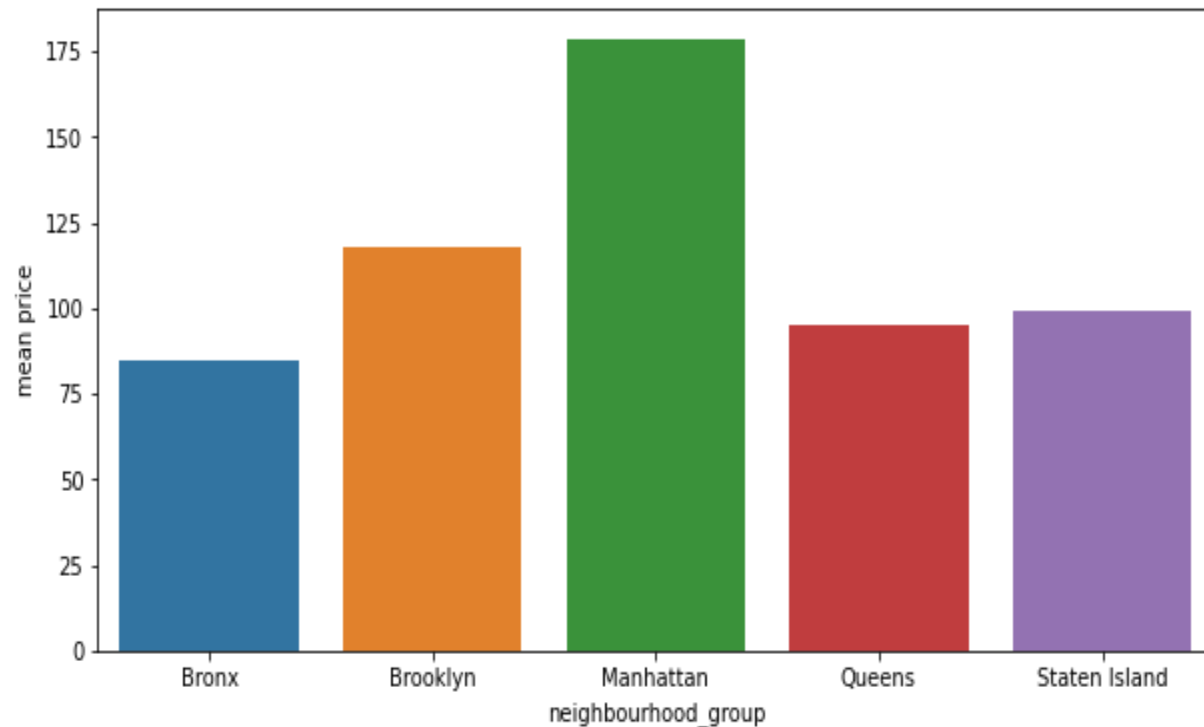
	neighbourhood_group	price
0	Bronx	85.067828
1	Brooklyn	118.076055
2	Manhattan	178.575227
3	Queens	95.420755
4	Staten Island	99.329759

Now we are comparing the overall average price per stay throughout New York for different neighborhood groups. Its clear that its high in Manhattan
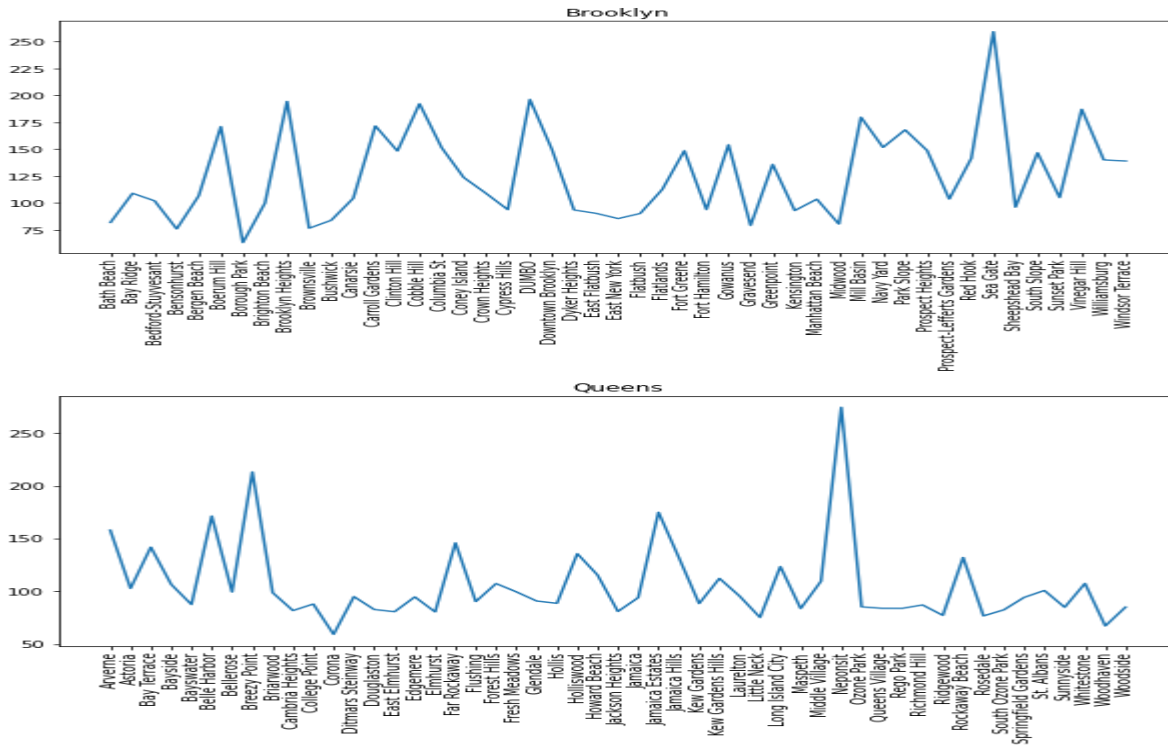
Average price for each room type
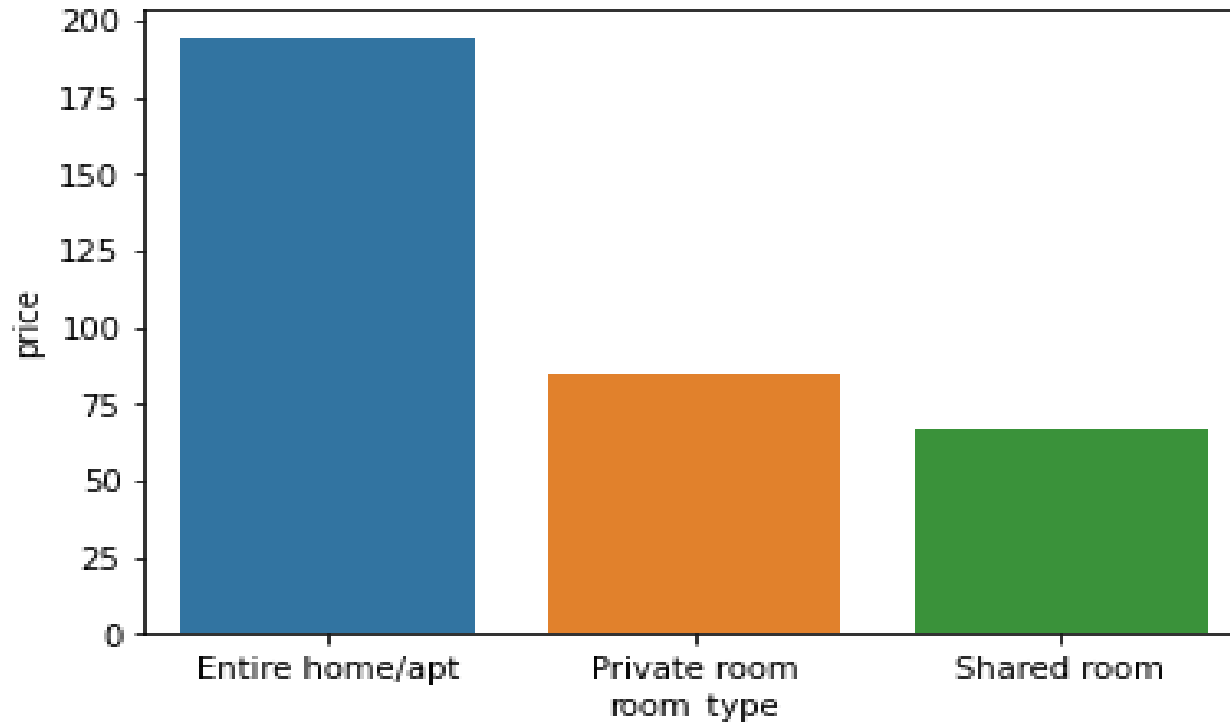
	room_type	price
0	Entire home/apt	194.211500
1	Private room	85.090881
2	Shared room	67.218103

By comparing average price through out New York in different available room types, we can observe that, Home/ Apartment is the category which has maximum average price.
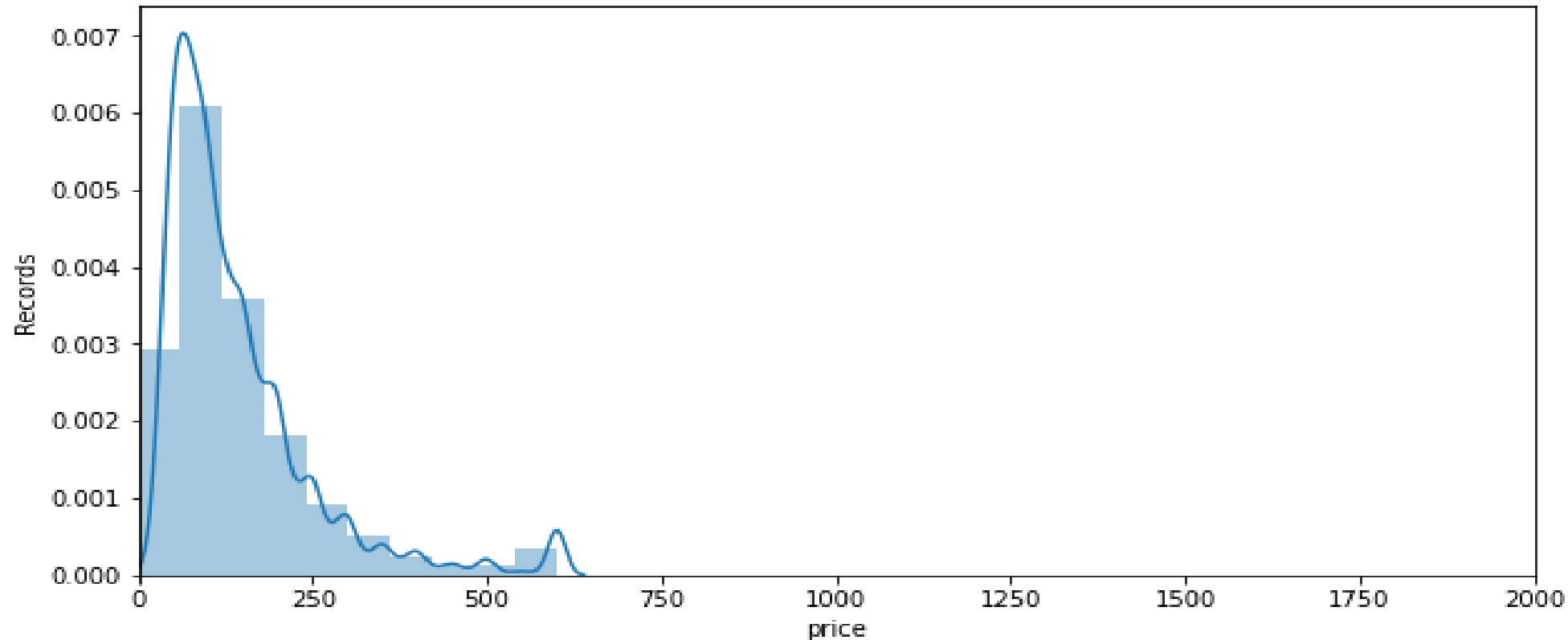
Price variation for each neighborhood group

We can see the price fluctuations through out different neighborhood groups. It differs from place to place, heavily in Manhattan.
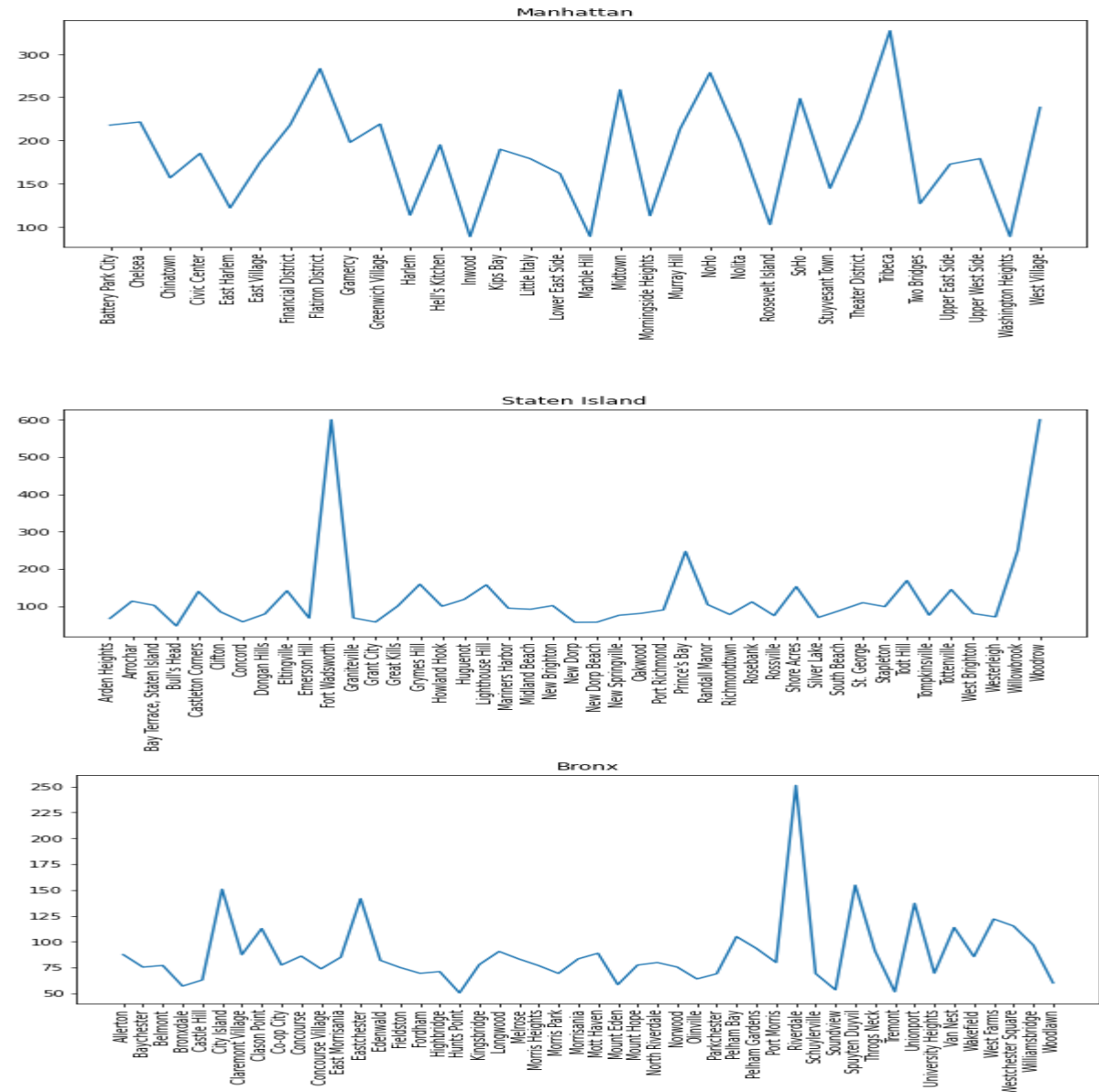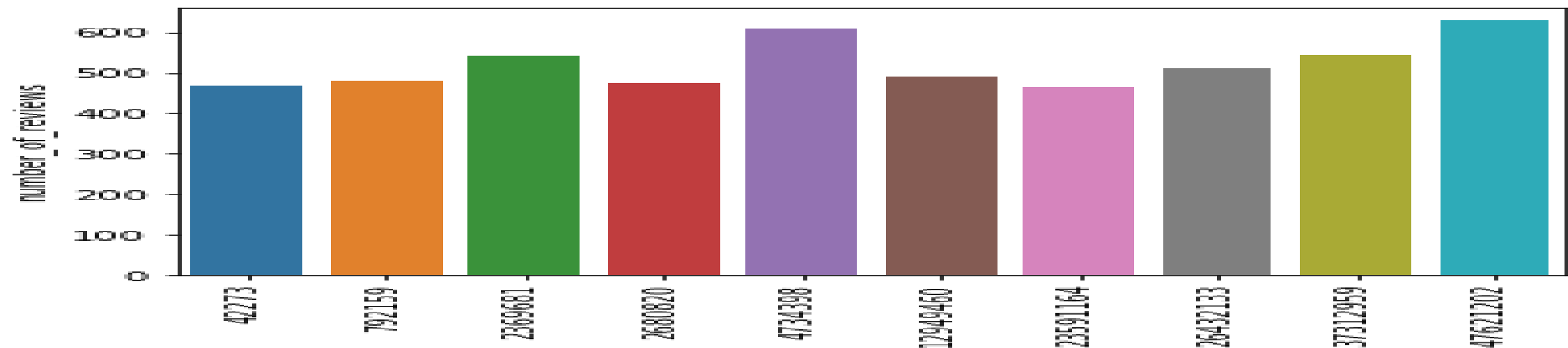
Frequency distribution of price throughout New york

We can observe that the price range between 0 to 250 is the price which most listings in New York offers for their places.
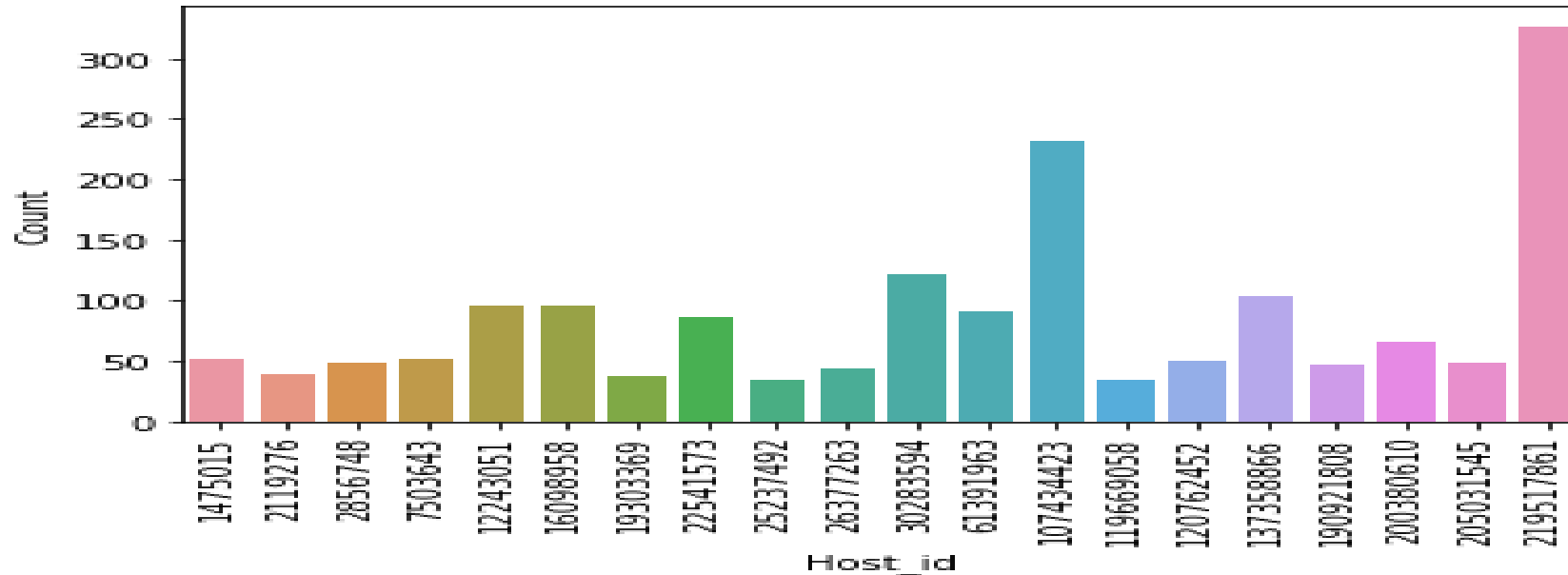
Hosts who got more number of reviews

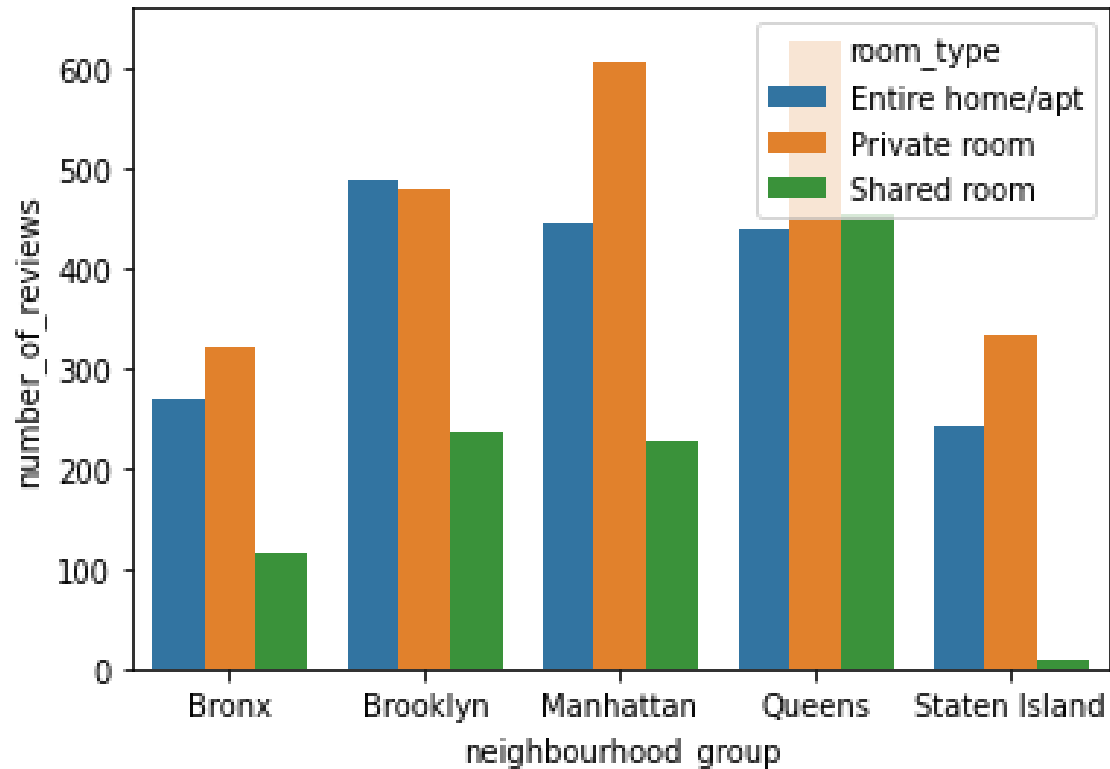|	host_id	number_of_reviews
23499	47621202	629
7365	4734398	607
21316	37312959	543

Host with the host_id 47621202 has the highest number of reviews throughout New York.

Famous Host/ host with more properties

We can see the list of hosts id who possess more number of properties in New York. The maximum listings are offered by host id 219517861

Places where number of reviews were max

	neighbourhood_group	room_type	number_of_reviews
0	Bronx	Entire home/apt	271
1	Bronx	Private room	321
2	Bronx	Shared room	116
3	Brooklyn	Entire home/apt	488
4	Brooklyn	Private room	480
5	Brooklyn	Shared room	236
6	Manhattan	Entire home/apt	447
7	Manhattan	Private room	607
8	Manhattan	Shared room	229
9	Queens	Entire home/apt	441
10	Queens	Private room	629
11	Queens	Shared room	454
12	Staten Island	Entire home/apt	242
13	Staten Island	Private room	333
14	Staten Island	Shared room	9

We trying to plot a maximum reviews obtained for listings in particular neighborhood groups with selected room types. Queens clearly has the maximum reviews.
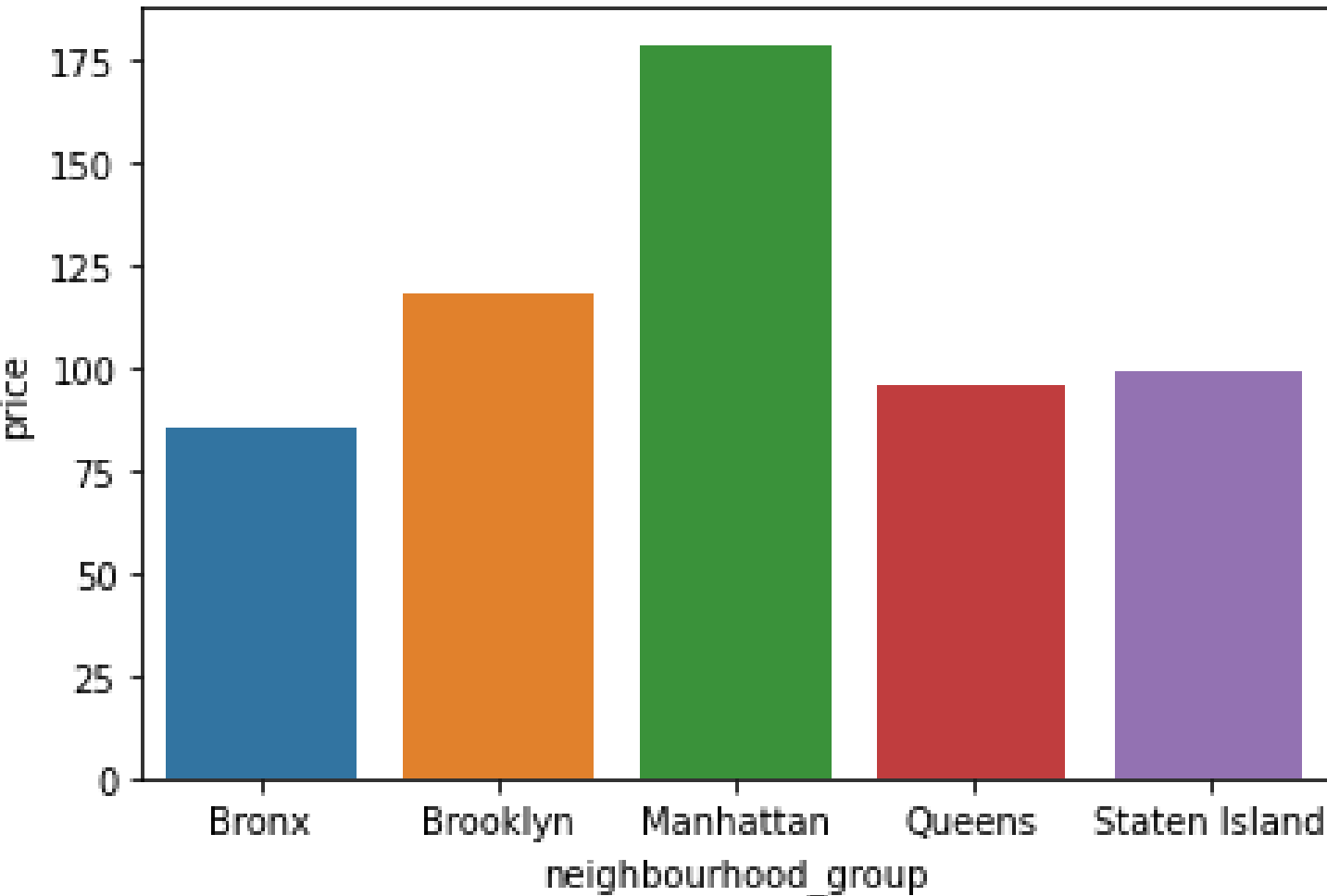
Average price per night
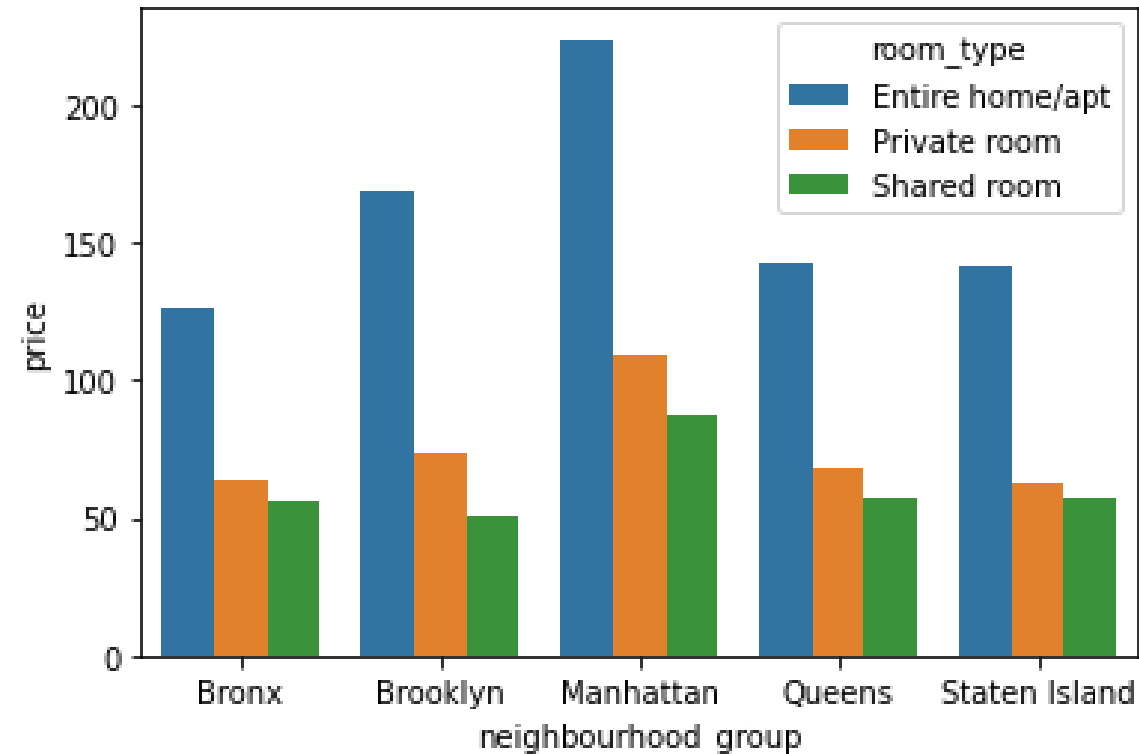
Manhattan has the highest price for a single night in all room types.

	neighbourhood_group	room_type	price
0	Bronx	Entire home/apt	126.266491
1	Bronx	Private room	63.751534
2	Bronx	Shared room	56.466667
3	Brooklyn	Entire home/apt	168.437389
4	Brooklyn	Private room	73.328563
5	Brooklyn	Shared room	50.225182
6	Manhattan	Entire home/apt	223.726797
7	Manhattan	Private room	109.401027
8	Manhattan	Shared room	87.310417
9	Queens	Entire home/apt	142.643130
10	Queens	Private room	68.285291
11	Queens	Shared room	57.656566
12	Staten Island	Entire home/apt	141.034091
13	Staten Island	Private room	62.292553
14	Staten Island	Shared room	57.444444

Average price per night in neighborhood groups

	neighbourhood_group	price
0	Bronx	85.067828
1	Brooklyn	118.076055
2	Manhattan	178.575227
3	Queens	95.420755
4	Staten Island	99.329759

We can observe that even if Consider overall Manhattan listings, The average price for a night is High in Manhattan

Average price for night in all room types in New York
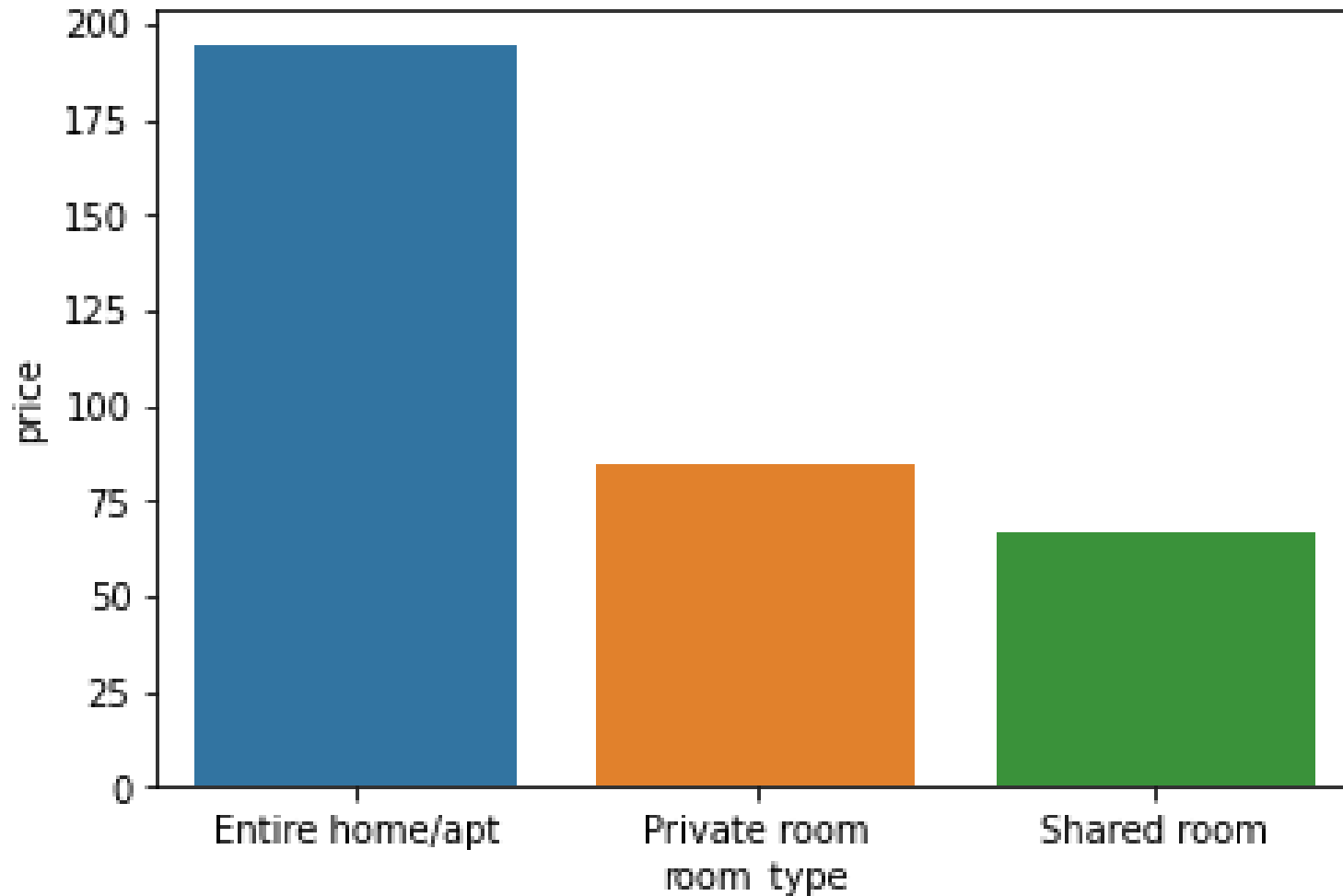
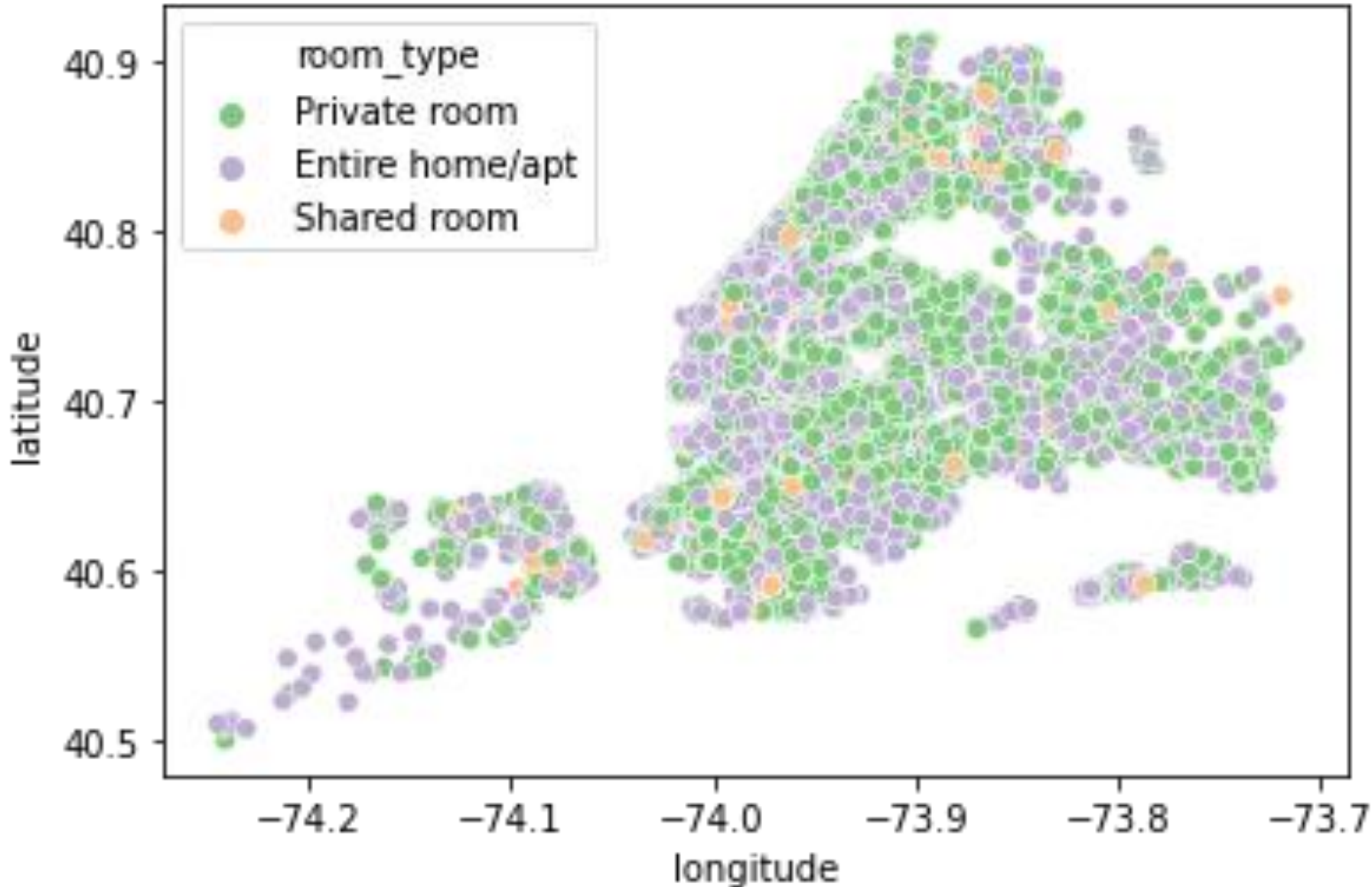|	room_type	price
0	Entire home/apt	194.211500
1	Private room	85.090881
2	Shared room	67.218103

Entire home/Apartments is type Which has the highest average Price for night in all over New York
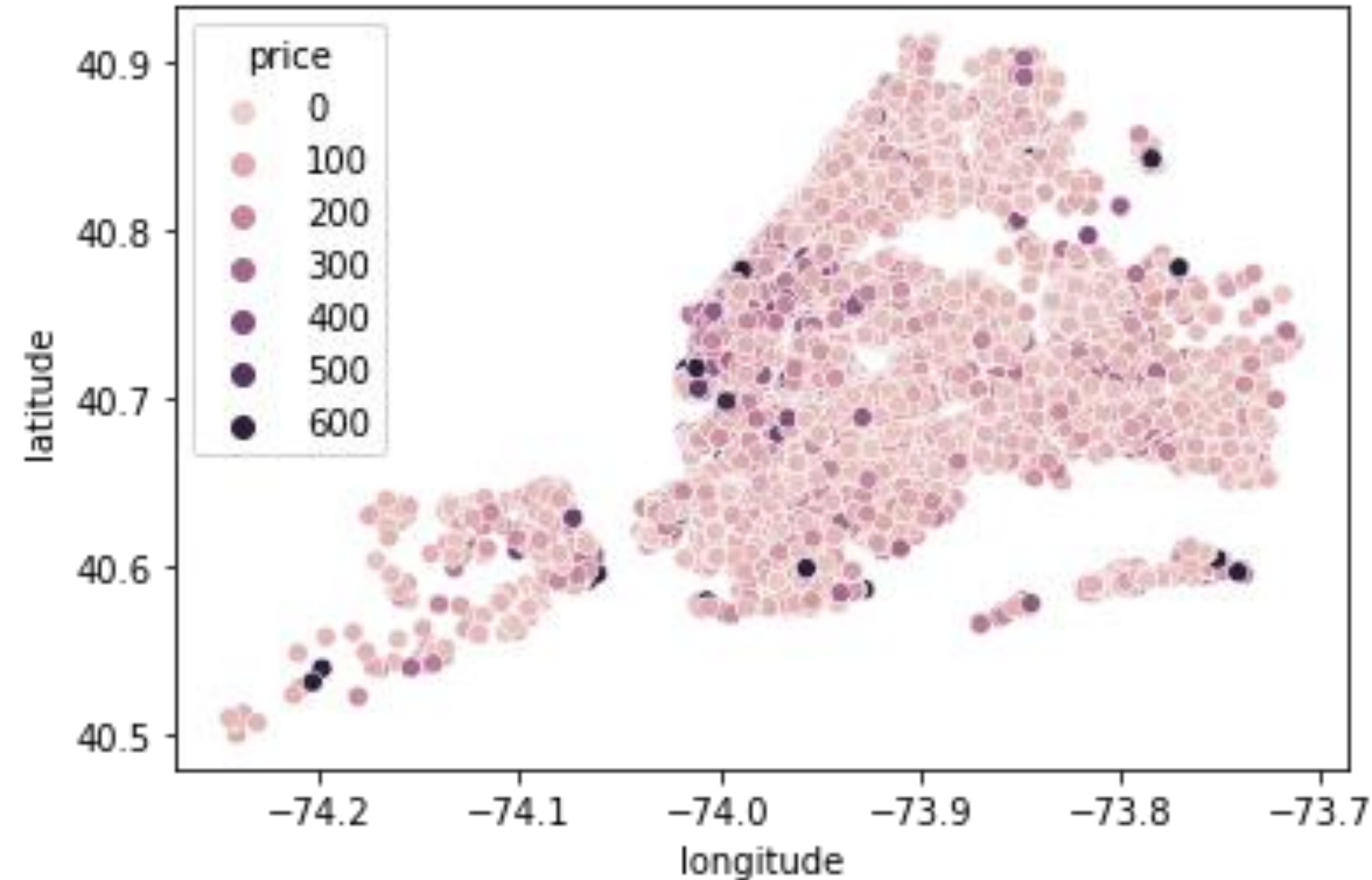
Longitude and latitude
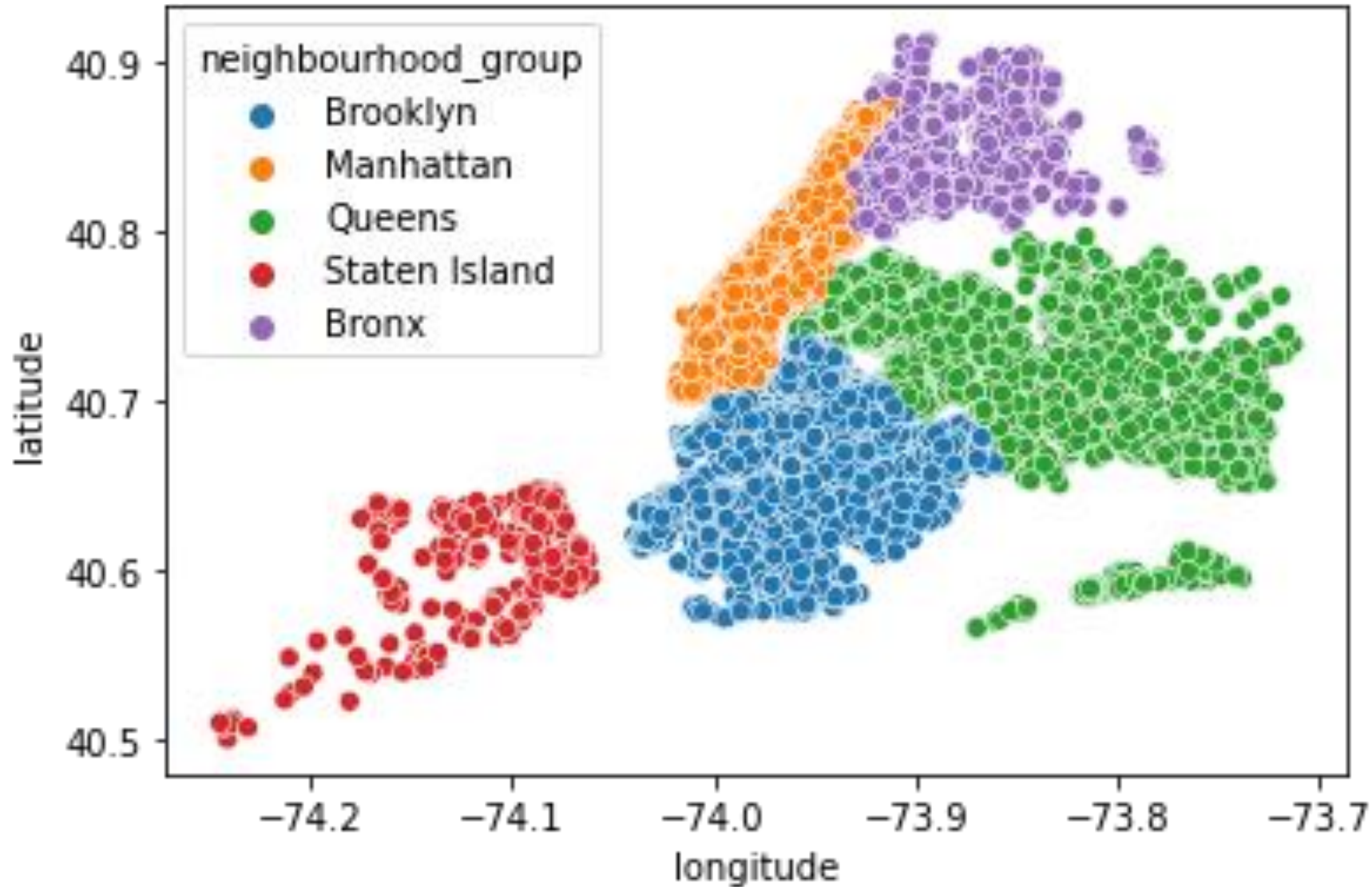
We can see the geo locations of all neighborhood groups throughout New York

Room types available in respective location

We can see the different room types available throughout New York

Price variations for different location

We can clearly see the places which is offering listings with maximum price throughout New York

Conclusion

- Price depends on Neighbourhood group. Its high in Manhattan. Manhattan is the city which has a lot of price fluctuations. We can even find the less prices for some listings in particular neighbourhood. We can even find the costliest listings in the same neighbourhood.

- We can get the famous host from Manhattan and the number of properties he is offering. He owns many properties throughout Manhattan, New York, which in turn is the busiest city in New York.

- We can observe one host from Queens has the property with more number of reviews. Queens is the place where there are listings available with reasonable price when compared to New York. We can conclude that people tend to go to the place with reasonable price.

- Throughout NY, Entire home/apt is the room type which is mostly in demand.

- Manhattan is the place which is famous and can be a good option for the companies to invest on properties of Entire home/Apt.

- We can increase the number of room type of places whichever got more number of reviews.

Thank You