Capstone Project
Airbnb Booking Analysis

Content

➤ Team information

➤ Problem statement

➤ Objective

➤ Data summary

➤ Analysis based on:
  - ➤ Neighborhood group
  - ➤ Price
  - ➤ Room types
  - ➤ Number of reviews
  - ➤ Longitude and latitude

➤ Conclusion

AI

Team information :

Student Name	Student email ID
Swathi V Hebbar	swathihebbar21@gmail.com
Charishma Suddala	charishma123.suddala@gmail.com

Problem statement –

Airbnb Bookings Analysis:

- **EDA -** Exploratory data analysis (EDA) is used to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.(Wikipedia)

- Since 2008, guests and hosts have used Airbnb to expand on travelling possibilities and present a more unique, personalized way of experiencing the 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 - data that can be analyzed and used for security, business decision, guiding marketing initiatives, implementation of innovative additional services and much more.

Objective

- ✓ Finding the total number of listings available in New York

- ✓ To find the average price for available room types/neighborhoods

- ✓ To find the hosts (Who has the most number of listings and who has more number of reviews for their listings) through out New York

- ✓ To find how price is affecting the Maximum reviewed places.

- ✓ To see the price variation for different listings throughout New York

Data Summary

The **Airbnb** data set contains 48895 records with 16 variables.

Important columns :

▪ **'name'** :These are names of the properties given by specific hosts. Or i can say, 'name' is a column which has the descriptions provided by the respective hosts for the reference of their customers.

▪ **'host_id' and 'host_name'** : Many properties are being offered by many hosts. This 'host_id' and 'host_name' holds the records of all those hosts.

▪ **'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.
- **availabilty_365'** : Through this column, we get an idea about for how many days the properties are available.

Treating NA

```
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
```

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 equated the null values of 'reviews_per_month' column as 0.

```
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
reviews_per_month       0
availability_365        0
dtype: int64
```
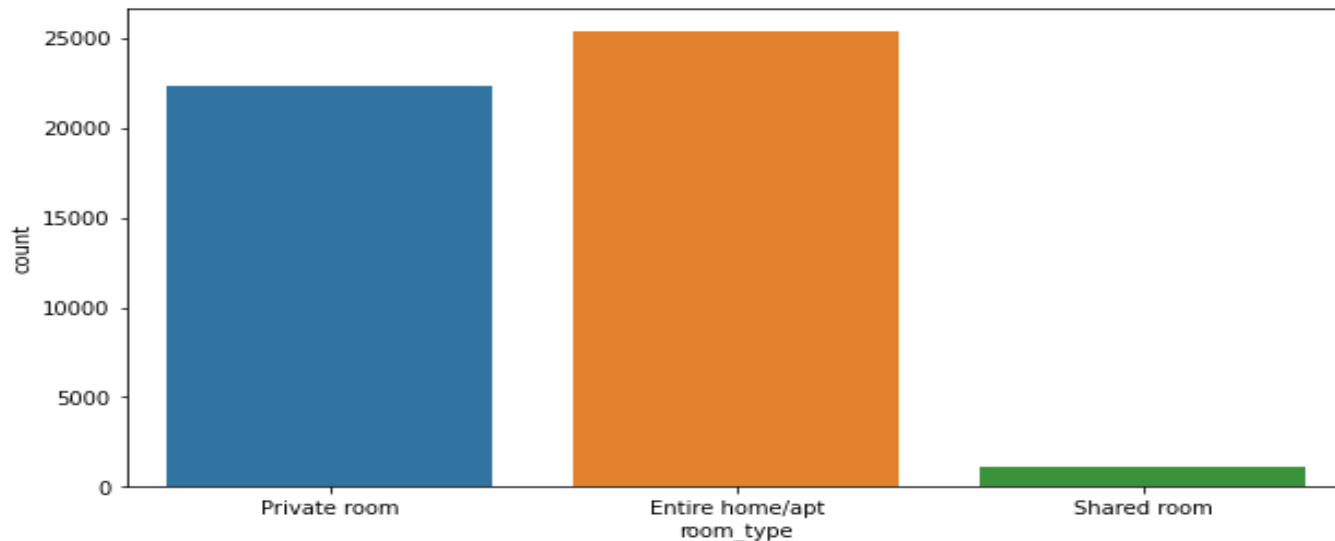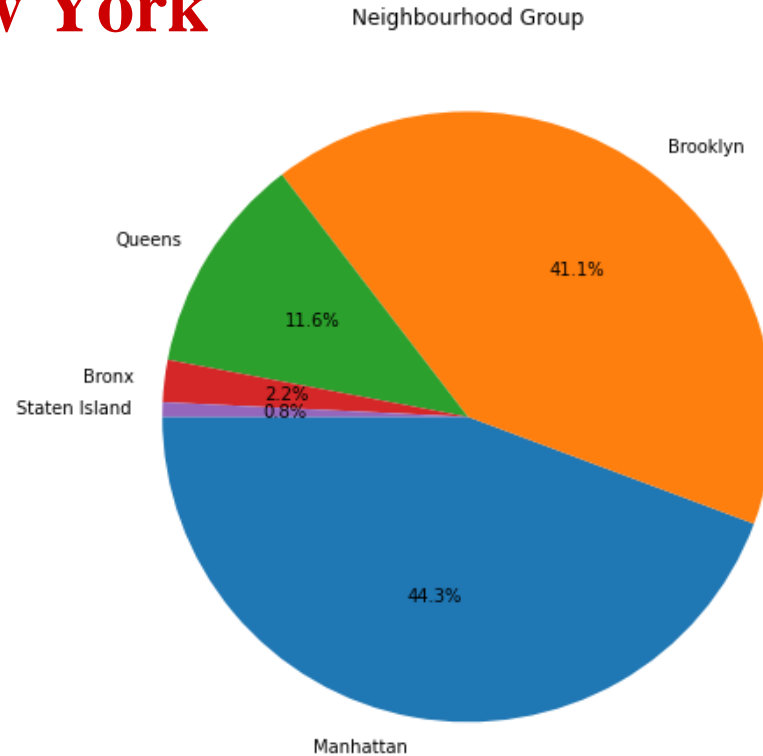
Outliers
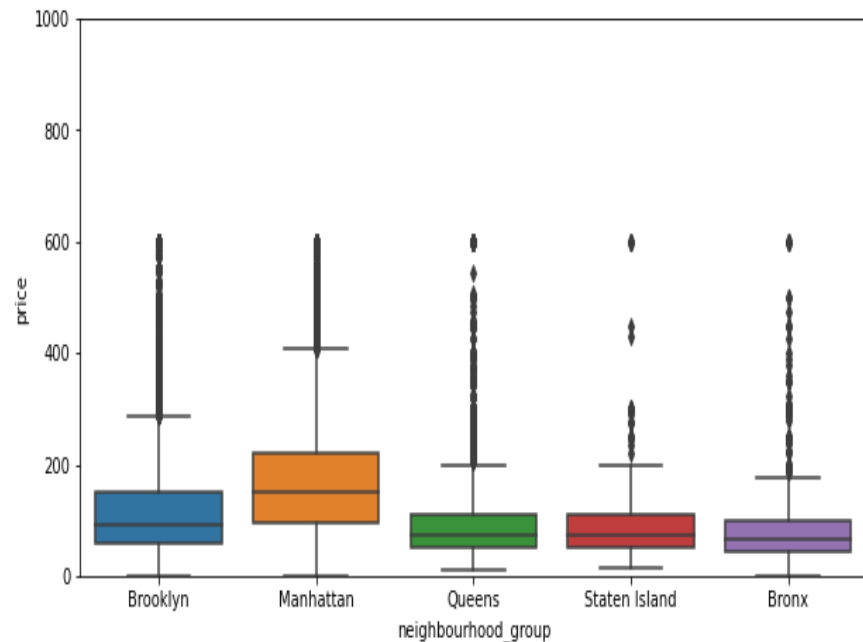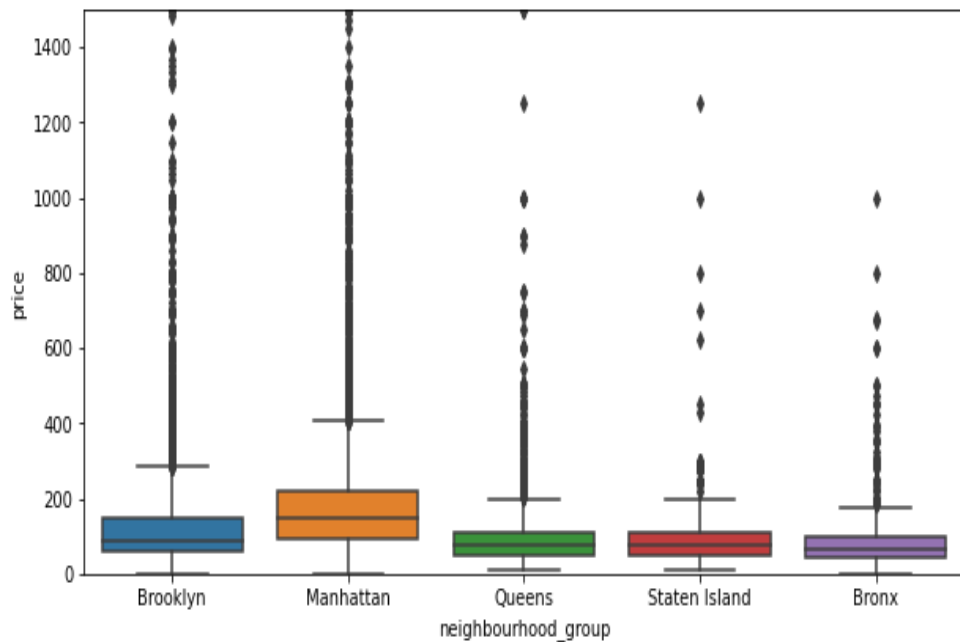
Neighborhood groups
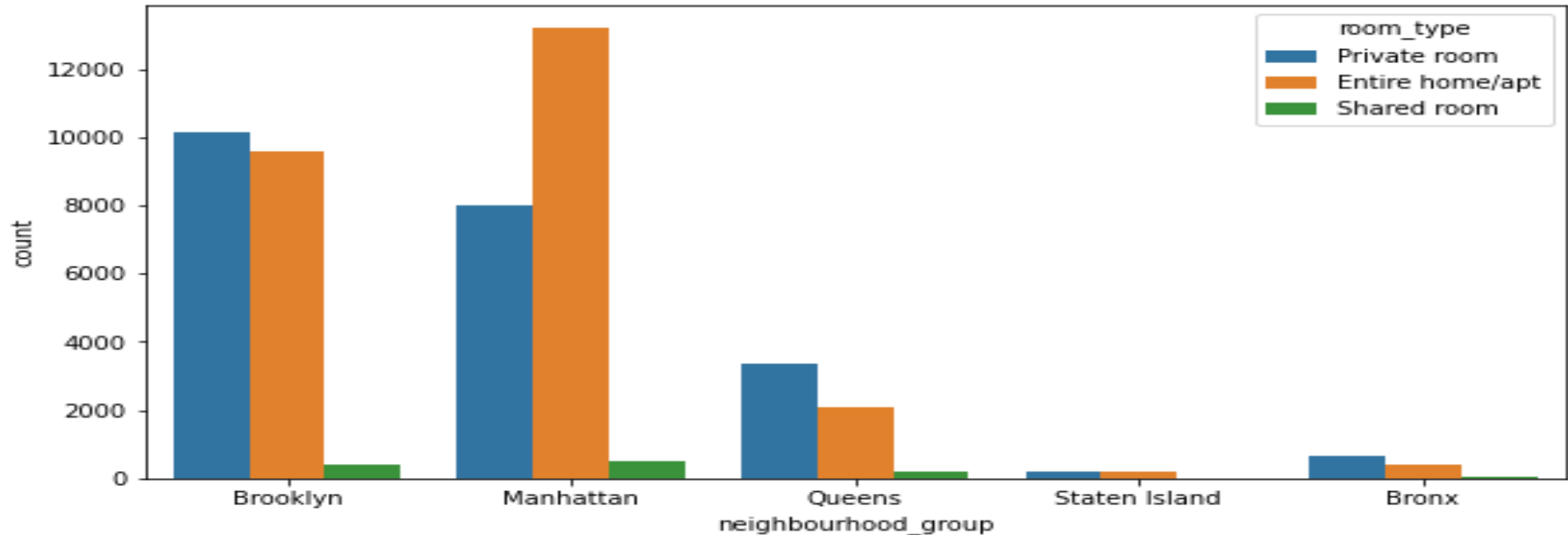
❑ Number of properties available in different neighborhood groups is more in Manhattan(Home room type)
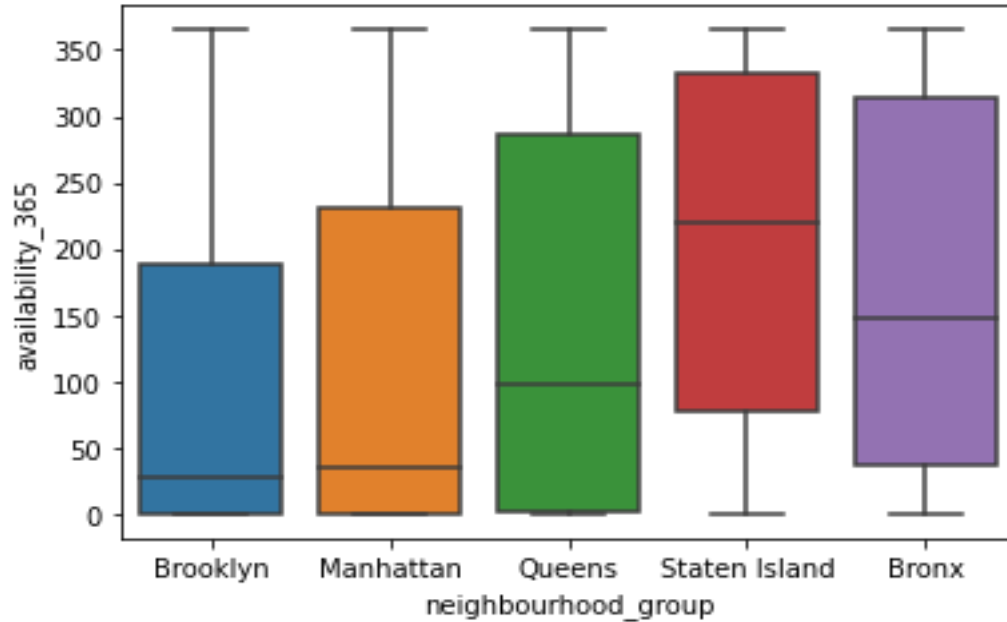
Room type throughout New York

❑ Number of room types available in New York which are of the type Private room, Entire home/Apartments, Shared room.
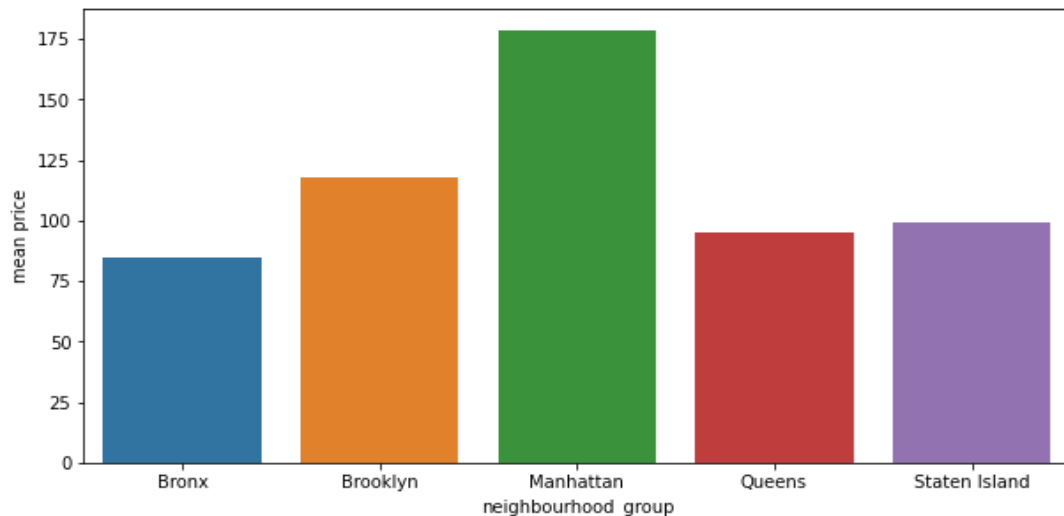
Percentage of listings offered by each neighborhood groups in New York

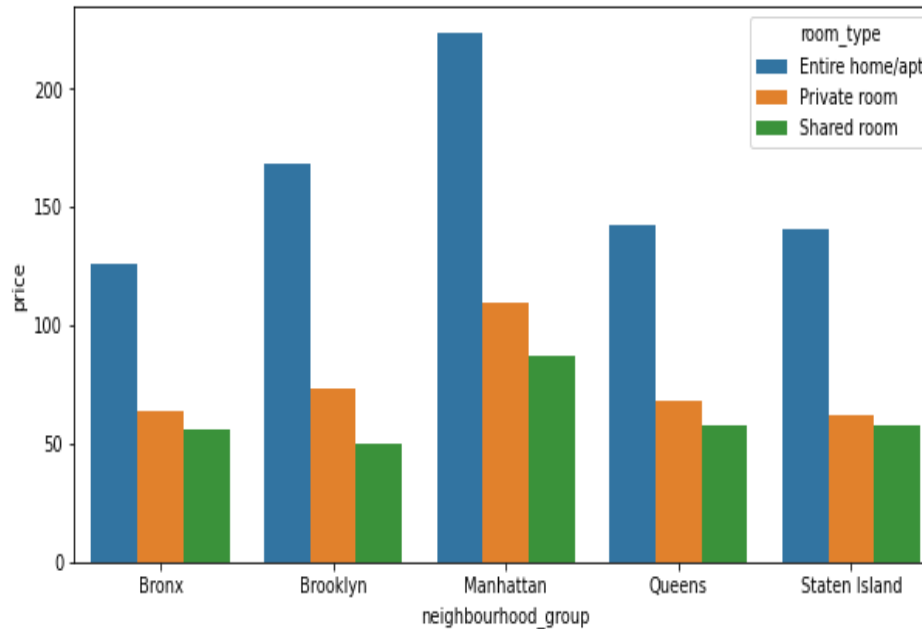# Availability of the listings and neighborhood group

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 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.
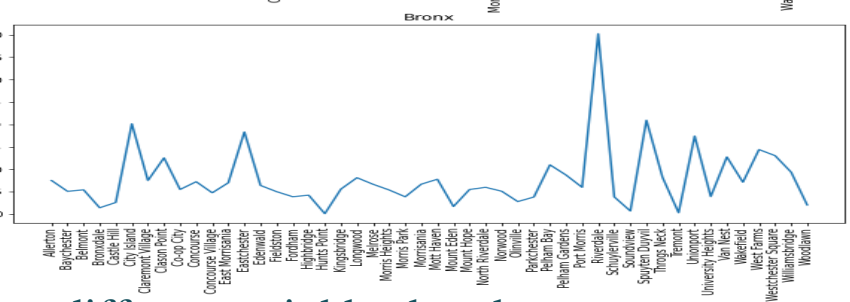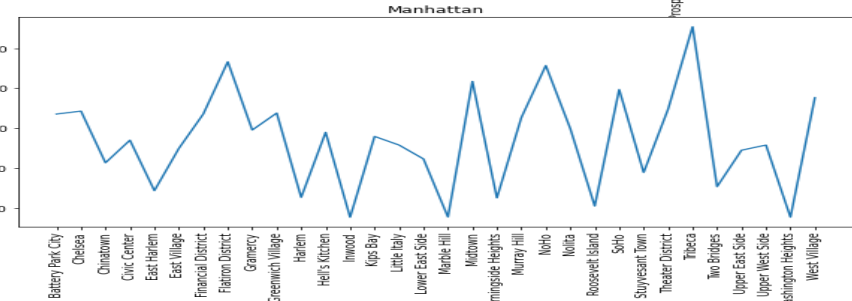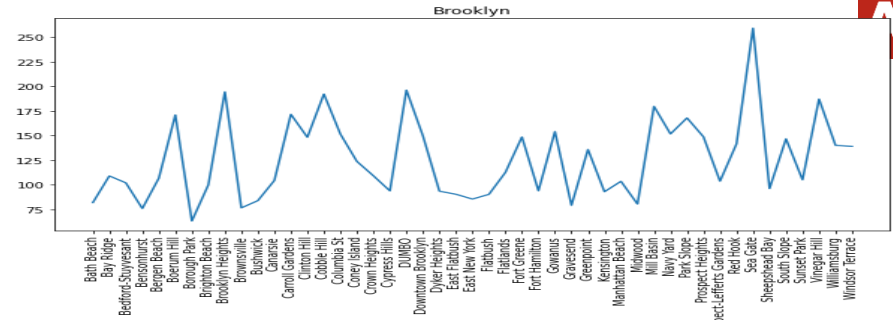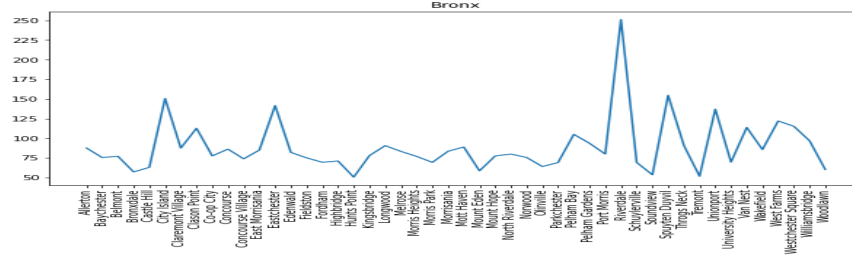
Average price for each room type
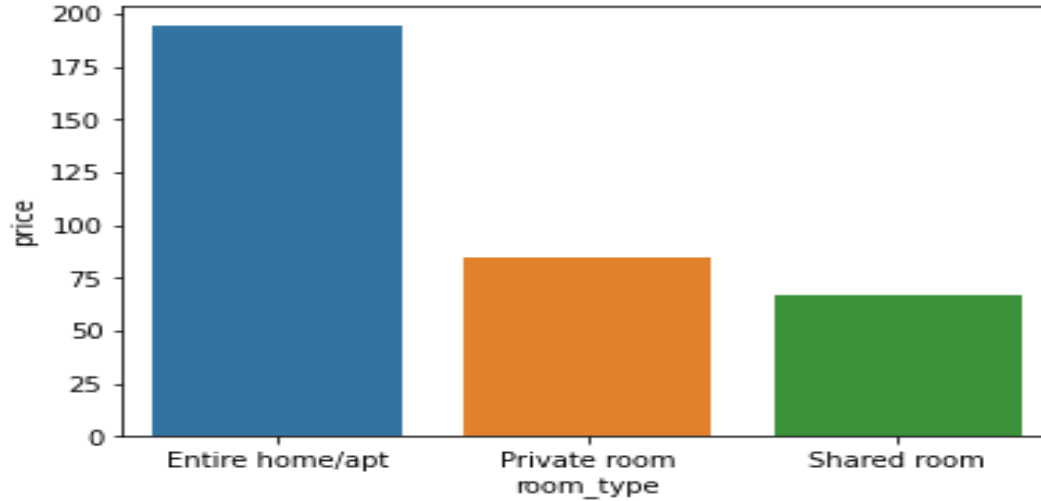
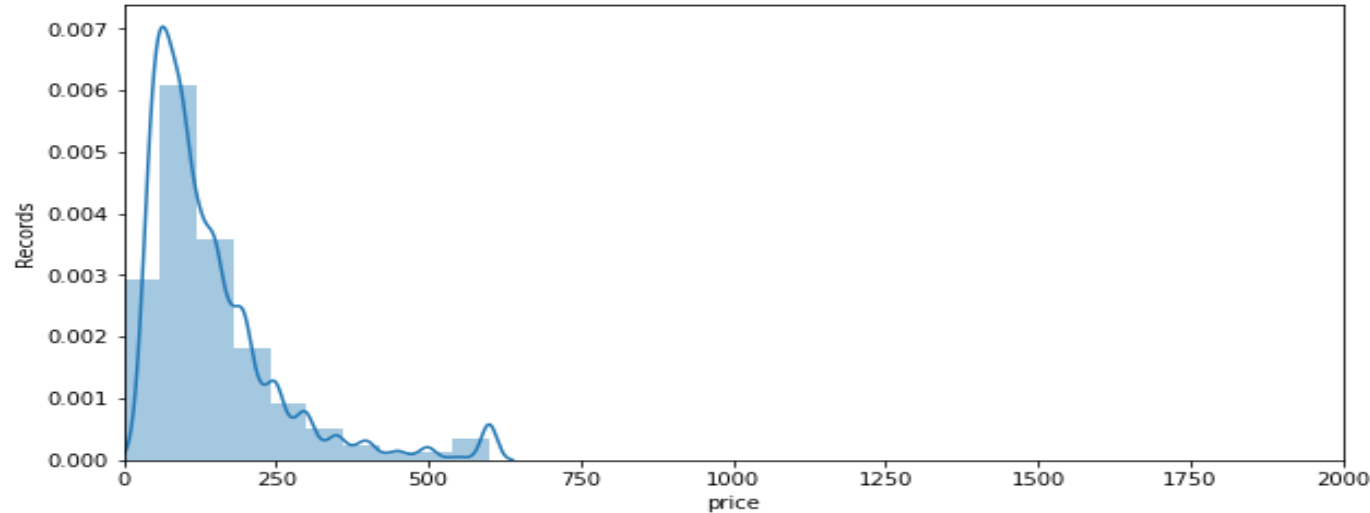|	room_type	price
0	Entire home/apt	194.211500
1	Private room	85.090881
2	Shared room	67.218103

❑ 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
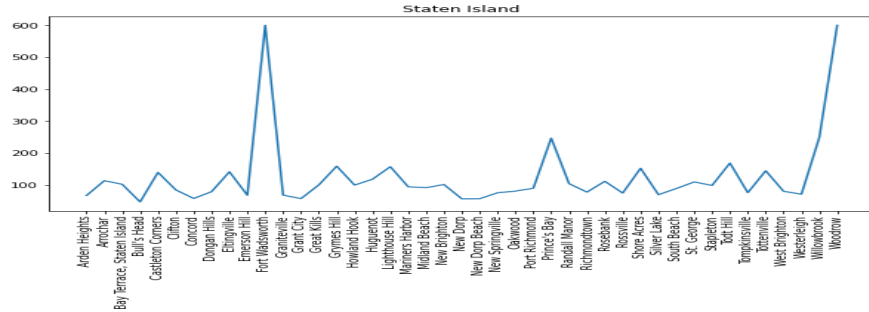
❑ We can see the price fluctuations through out different neighborhood groups. It differs from place to place heavily in Manhattan.

Frequency distribution of prices 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.
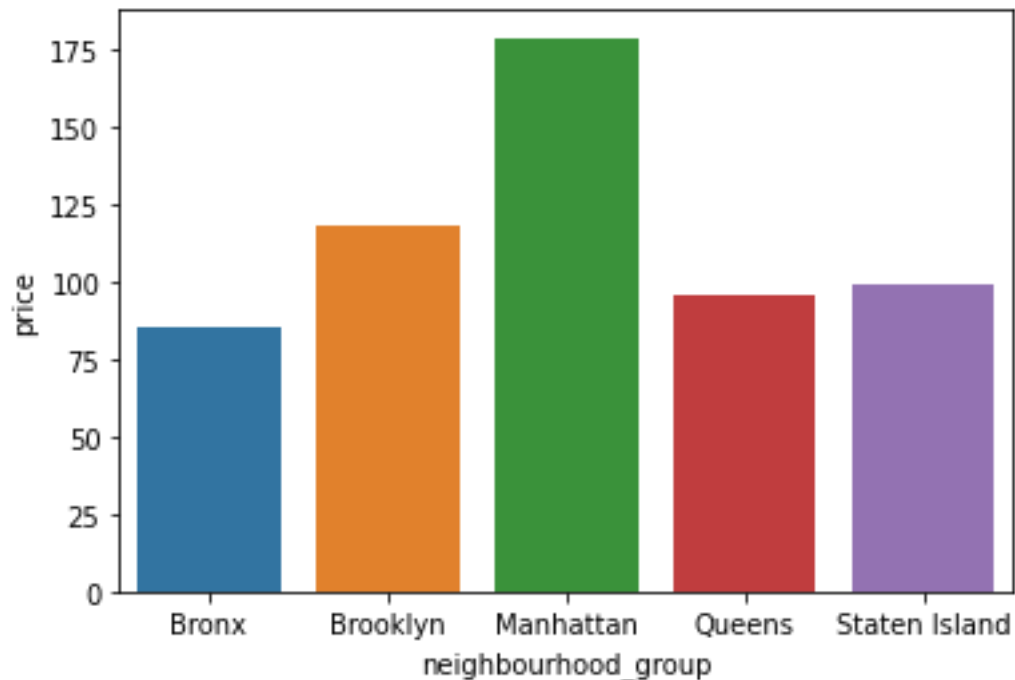
Average price per night
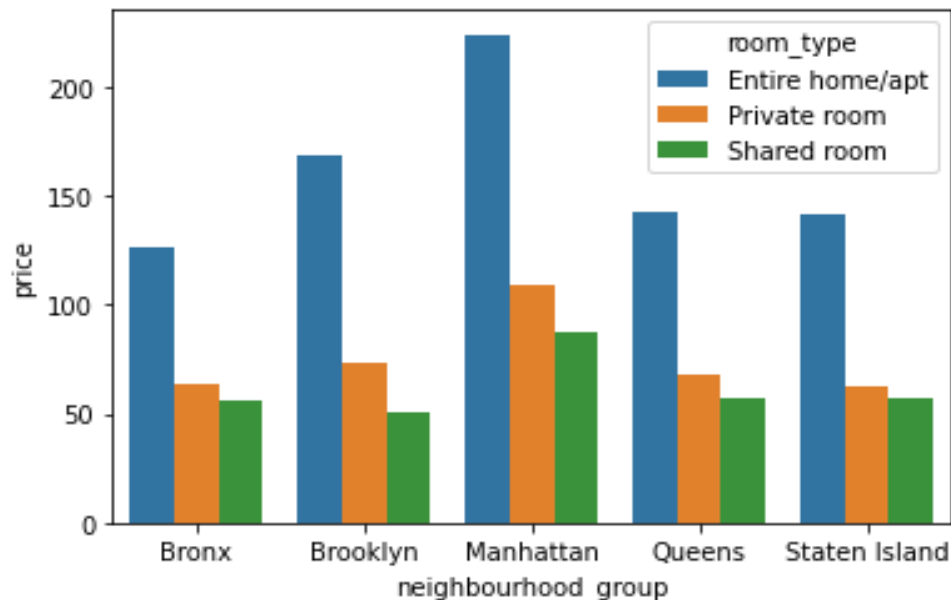
❑ Manhattan has the highest price for a single night in all room types.

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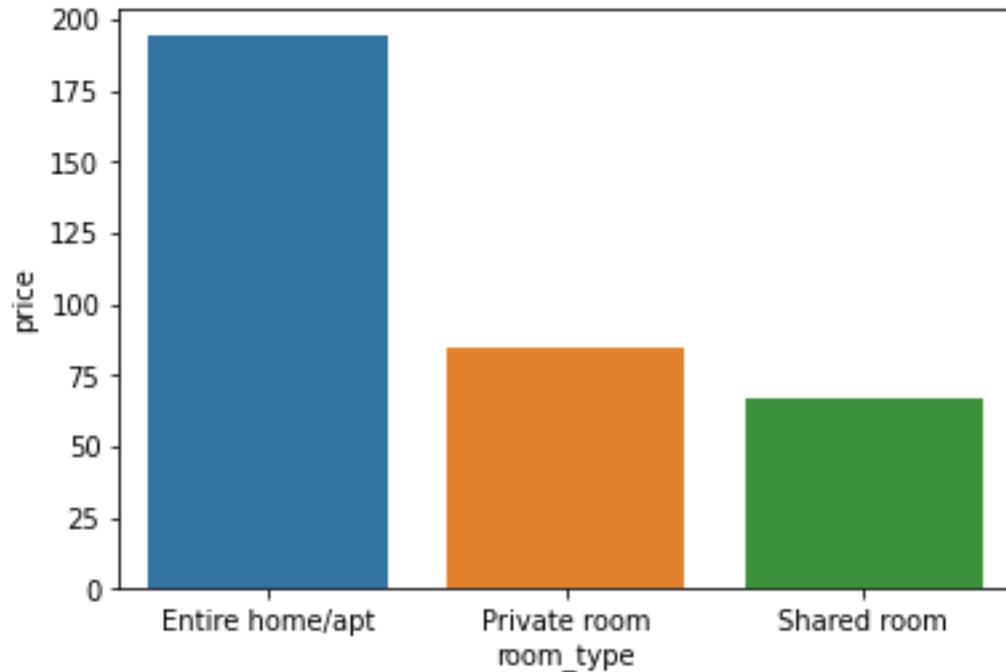
	room_type	price
0	Entire home/apt	194.211500
1	Private room	85.090881
2	Shared room	67.218103

❑ Entire home/Apartments is type Which has the highest average Price for night in all over New York

Hosts who got more number of reviews
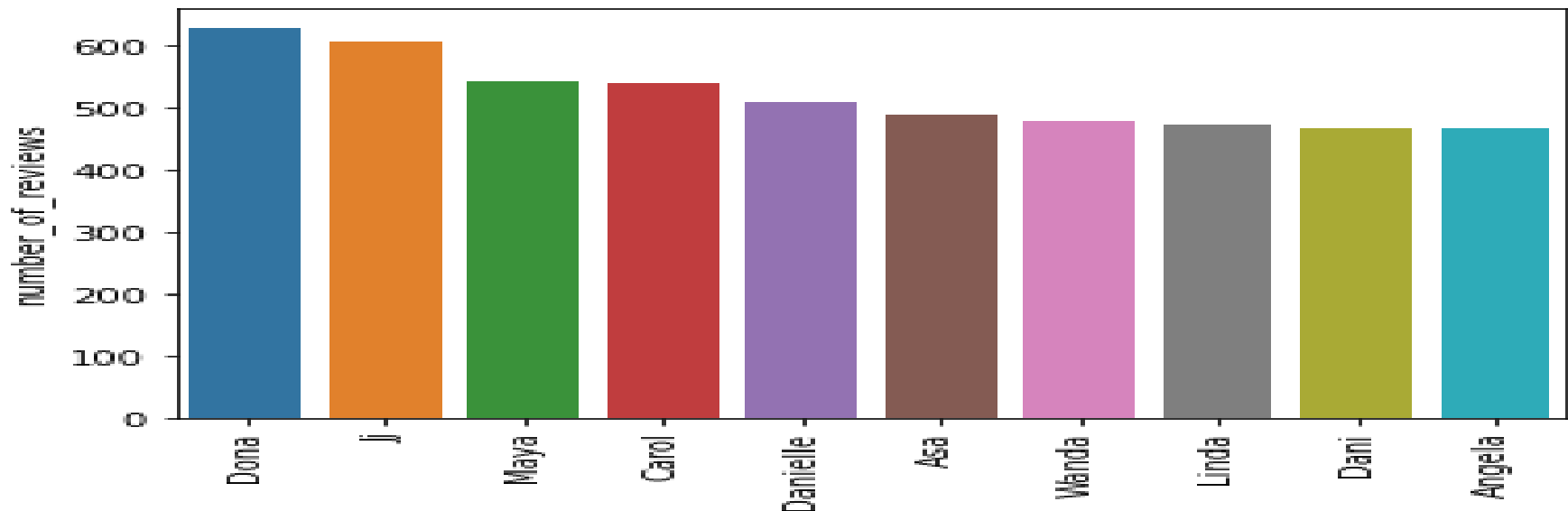
	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	availability_365
11759	9145202	Room near JFK Queen Bed	47621202	Dona	Queens	Jamaica	40.66730	-73.76831	Private room	47	1	629	14.58	333
2031	903972	Great Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82085	-73.94025	Private room	49	1	607	7.75	293
2030	903947	Beautiful Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82124	-73.93838	Private room	49	1	597	7.72	342

❑ Dona is the host who got more number of reviews throughout New York.

Famous host / host with more properties

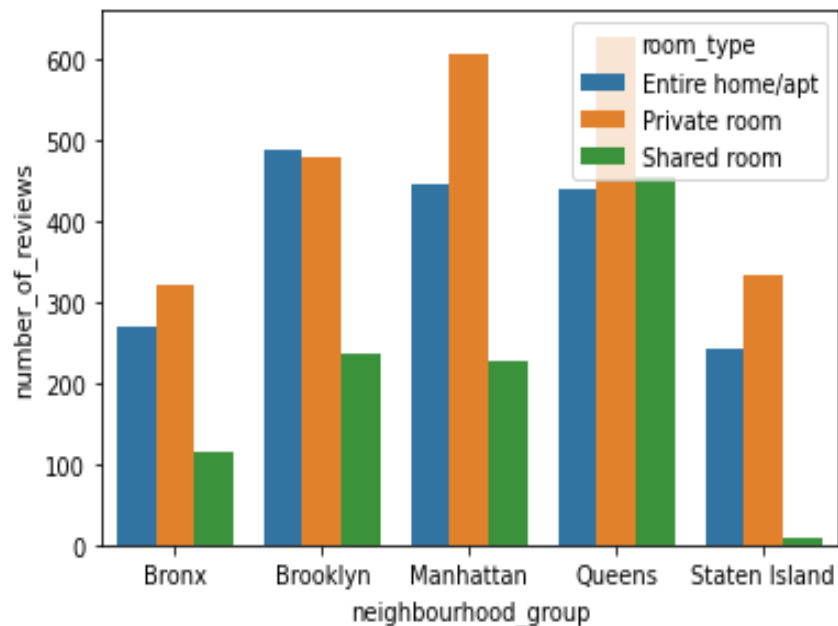|	host_name	neighbourhood_group
38293	Sonder (NYC)	Manhattan
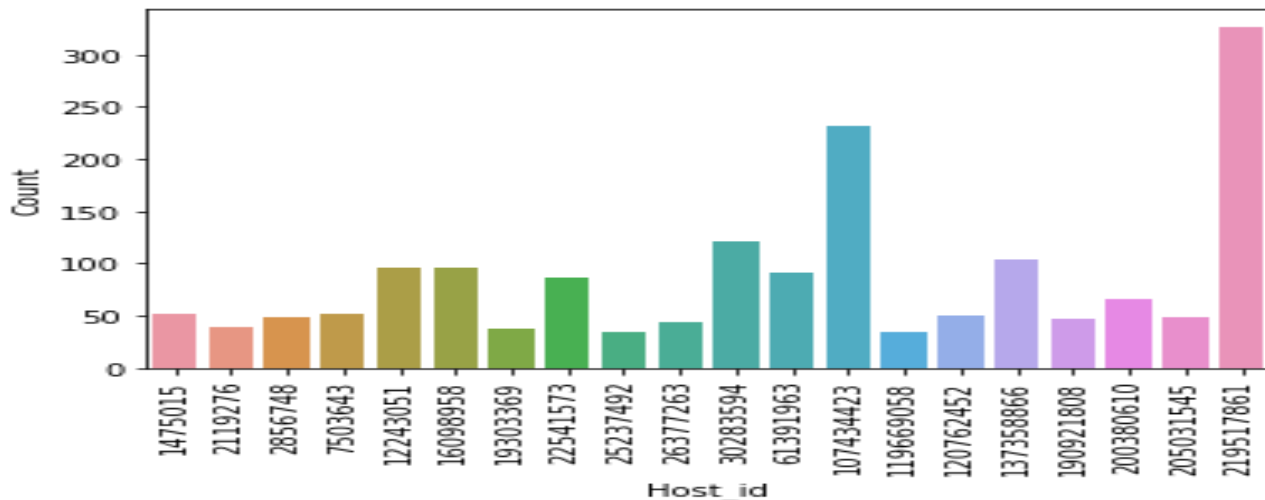
❑ We can see the list of hosts id who possess more number of properties in New York. The maximum listings are offered by host id 219517861.

❑ We can see that the host who has more listings in New York is **Sonder** and he is from Manhattan.

Places where number of reviews were at 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.
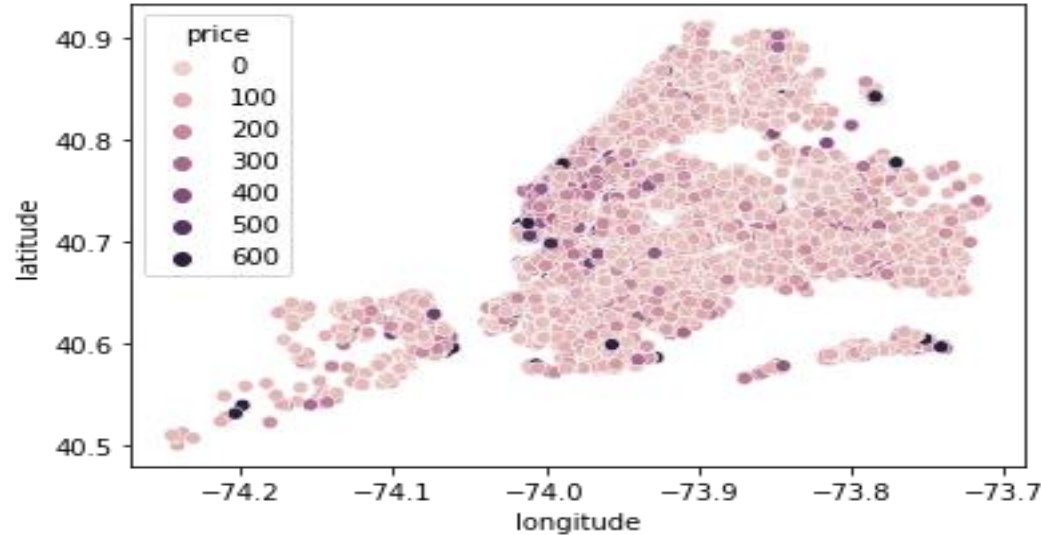
Longitude and Latitude
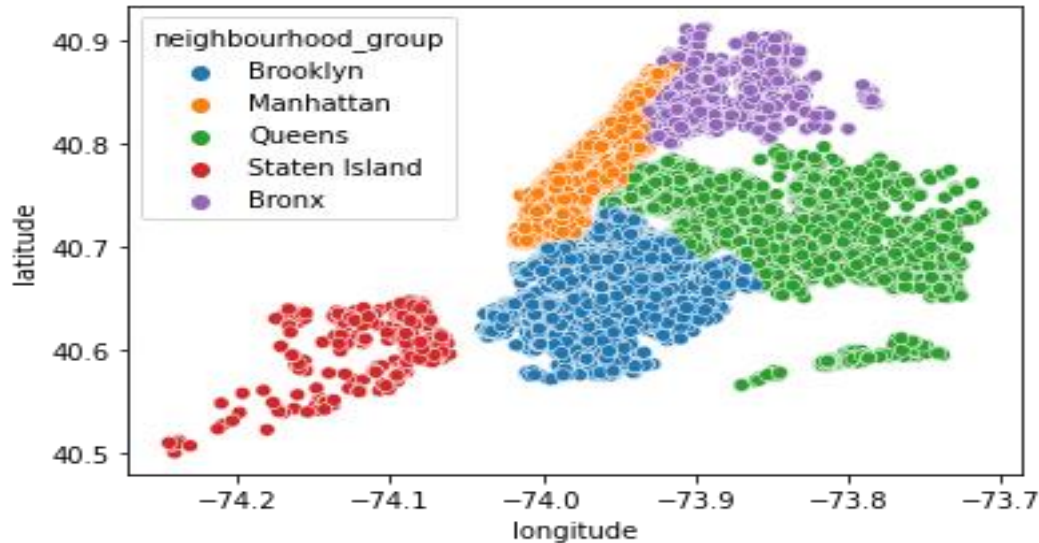
❏ We can see the geo locations of all neighborhood groups throughout New York.

Room types available in respective location
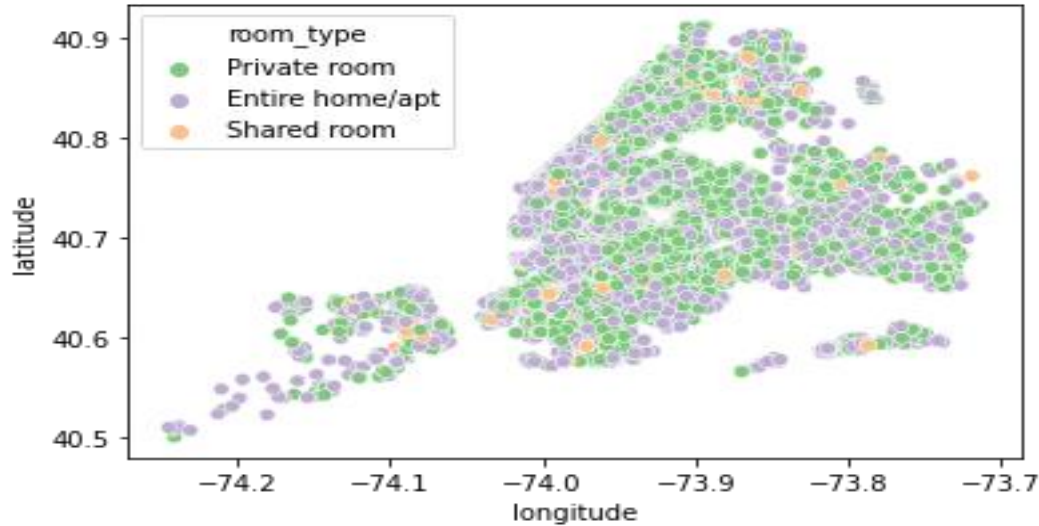
❑ We can see the different room types available throughout New York

Price variations for different location

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

Conclusion

(1) 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.

(2) We can get the famous host (Sonder from Manhattan) from 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.

(3) We can observe host Dona 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 manhattan. We can conclude that people tend to go to the place with reasonable price.

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

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

(6) We can increase the number of room type of places whichever got more number of reviews.

Thank You