Airbnb Bookings Analysis

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 factor for the company. These millions of listings generate a lot of data - data that can be analysed and used for security, business decisions, understanding of customers' and providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more. This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

What is 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 and hospitality service for people to lease or rent short-term lodging including holiday cottages, apartments, homestays, hostel beds, or hotel rooms. Basically they provide the platform that is shared by hosts and visitors worldwide. Lets look at the dataset once.

Objectives

In this project, we will perform an exploratory data analysis(EDA) in order to investigate each of the variables and also come up with a conclusion for the relationship between variables. The main purpose is to identify which variables affect the price mostly. In addition to these, we will explore which neighborhood groups and room types are the most popular ones among the guests, and which hosts are the most preferred ones. The processes during the extraction can be listed as below:

1. Data Cleaning

2. Data Preprocessing

3. Data Manipulation

4. Data Visualization

5. Exploring the information

Questions that can be answered by this analysis

1. - Number of Neighborhood in each Neighborhood Groups

2. - Distribution of rooms according to latitude and longitude and room density in their regions

3. - Room types and their count in each neighborhood groups

4. - Relationship between room type and Price

5. - Price and Review Relationship

6. - Minimum, Maximum and Average Price according to the neighbourhood group

7. - The Most and Least Expensive Neighborhoods

Used Libraries

I have used several packages during the analysis of the historical data of Airbnb in NYC in order to make data manipulation and visualization. The list of packages used in this EDA can be seen below:

1. NumPy > To perform mathematical operations

2. pandas > For data frame

3. matplotlib > For Visualization of Data

4. seaborn > For Visualization of Data

5. plotly.express > For Visualization of Data

6. plotline > For Visualization of Data

7. warnings > For ignorance of any kind of unnecessary warnings

Lets now understand about every column in the data frame.

1. **'id'** => It is the unique id given for listings.

2. **'name'** => It is a column which contains the descriptions/names provided by the respective hosts for the reference of their customers.

3. **'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.

4. **'neighbourhood' and 'neighbourhood_group'** => These columns holds the information about the city and areas of the properties which are offered in airbnb New York.

5. **'Longitude' and 'Latitude'** => As the name suggests it just contains the longitude and latitude of the property location

6. **'Room_type'** => It displays the room_type of the property (either private room / entire home / shared room)

7. **'price'** => Its an important column which holds the price value of all those properties.

8. **'minimum_nights'** => It gives us information about the minimum number of nights that is offered by hosts for particular property.

9. **'number_of_reviews' and 'reviews_month'** => It contains information about the number of reviews and reviews per month for those properties and hosts hospitality.

10. **'availabilty_365'** => It gives information about the availability of the listings.

By using heatmap we can see that the last_review and reviews_per_month has maximum number of null values.

```
plt.figure(figsize=(10,7))
sns.heatmap(df.isnull(),yticklabels = False,cmap = 'plasma',cbar =False)
plt.title('Missing Columns')
plt.suptitle('2019 Airbnb Data')
```

```
[ ]  # percentage of null values
     str(round(df['reviews_per_month'].isna().sum()/len(df)*100,2))+'%'

     '20.56%'
```


2019 Airbnb Data

Missing Columns
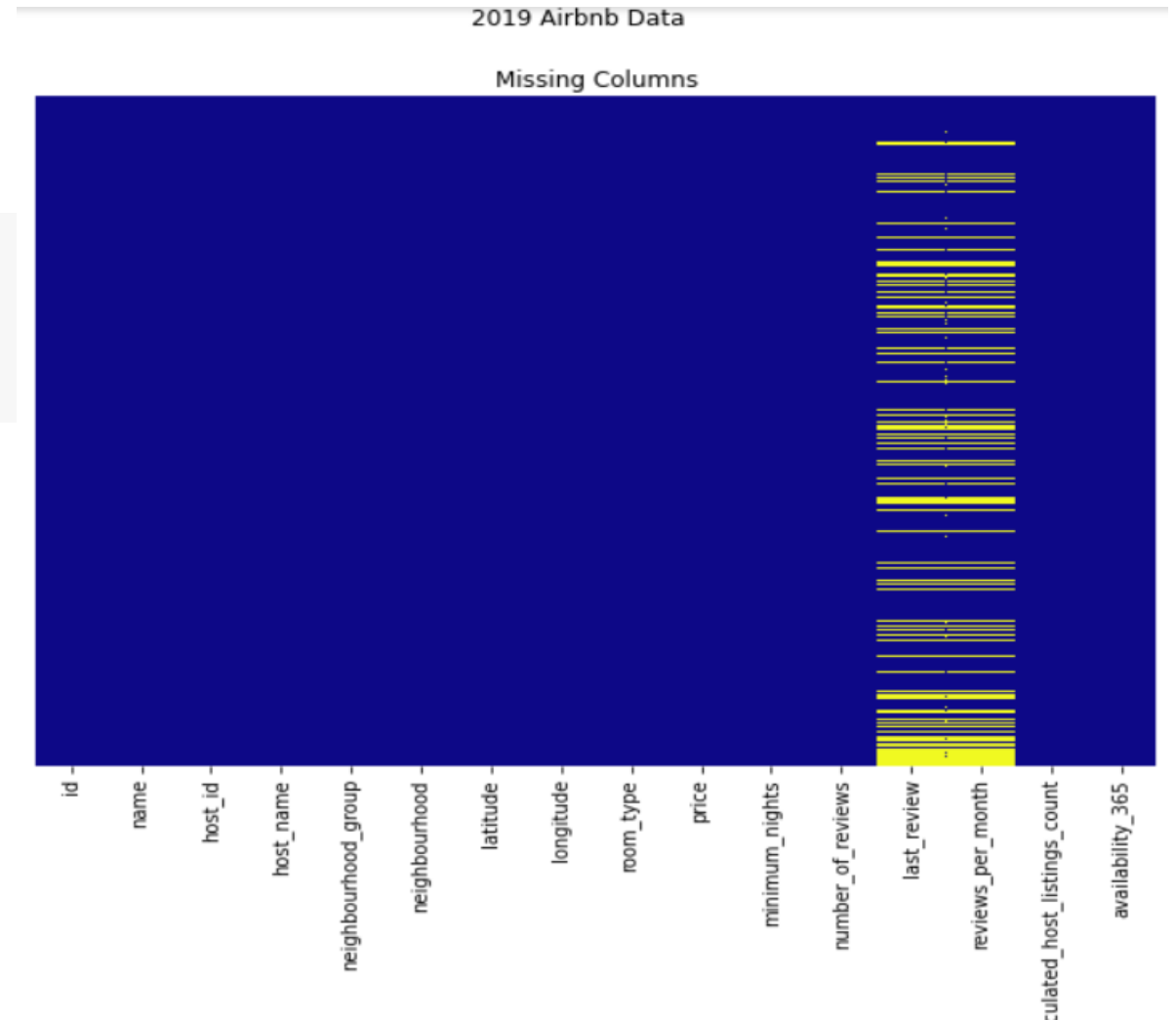
Calculating the number of neighborhood in each neighborhood group

```python
count_of_nbrhd_grp = pd.DataFrame(df.groupby(['neighbourhood_group'])
['neighbourhood'].count()).rename(columns={'neighbourhood':'Count_of_neighbourhood'}).reset_index()
count_of_nbrhd_grp
```

	neighbourhood_group	Count_of_neighbourhood
0	Bronx	1091
1	Brooklyn	20104
2	Manhattan	21661
3	Queens	5666
4	Staten Island	373

```python
px.bar(count_of_nbrhd_grp , y = 'Count_of_neighbourhood' , x = 'neighbourhood_group', color = 'neighbourhood_group'
,height = 500 , width = 700, title ='Neighbourhoods in each Neighbourhood Groups<br><sup> 2019 Airbnb Data<sup>'
,text_auto = True).update_layout(paper_bgcolor = "rgba(0,0,0,0)").show()
```
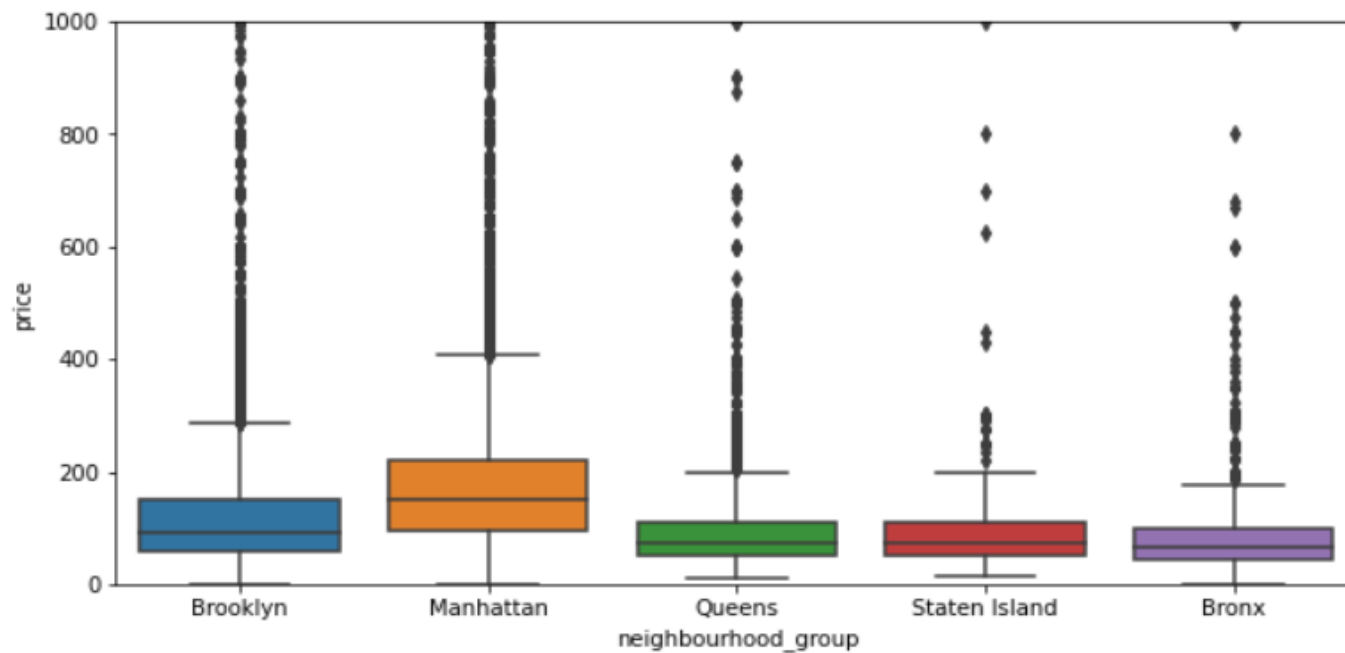

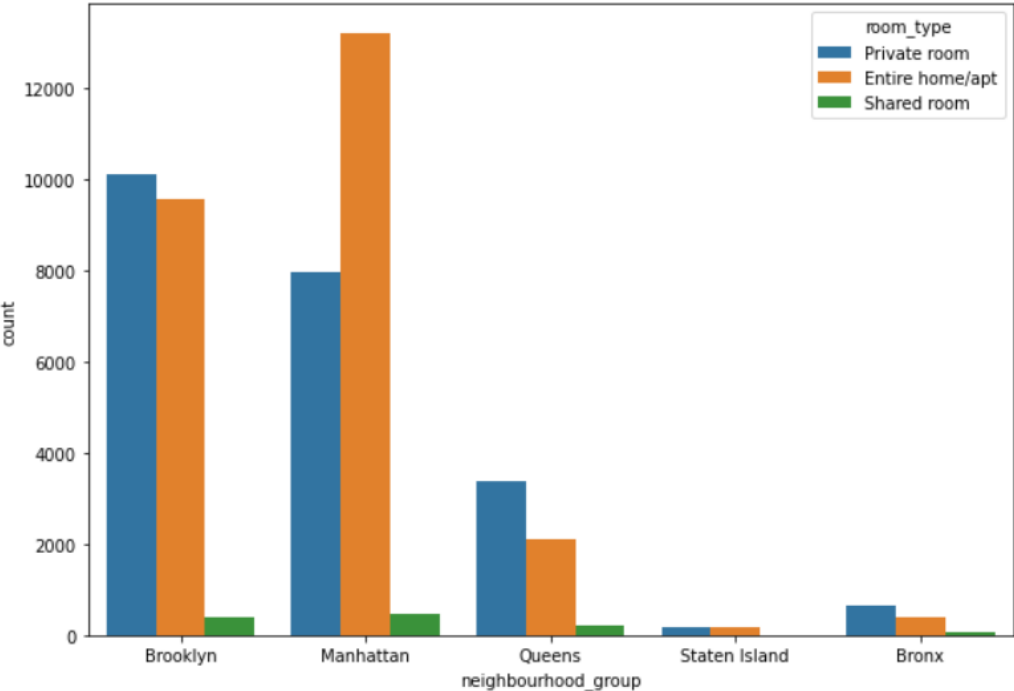Neighbourhoods in each Neighbourhood Groups
2019 Airbnb Data
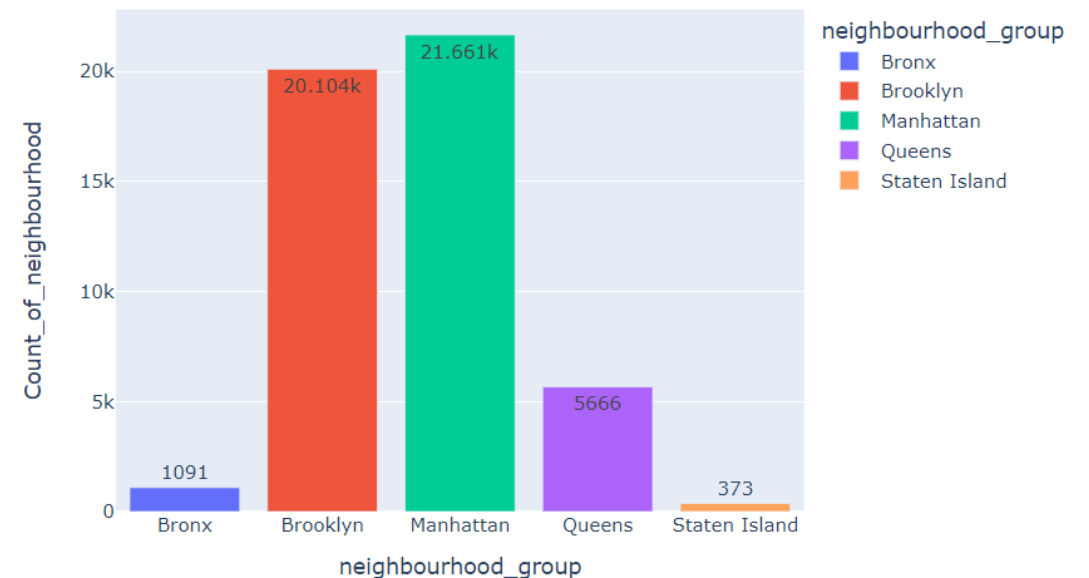
Range of Prices in Neighborhoods groups

```python
plt.figure(figsize=(10,5))
plt.ylim(0,1000)
sns.boxplot(x='neighbourhood_group', y='price',data=df)
plt.show()
```

# No. of rental properties group by type and there count

```python
rental_properties= pd.DataFrame(df.groupby(['neighbourhood_group'],dropna=True)['room_type'].value_counts())
rental_properties
```



```python
plt.figure(figsize=(10,7))
sns.countplot(x='neighbourhood_group', hue='room_type', data=df)
plt.show()
```
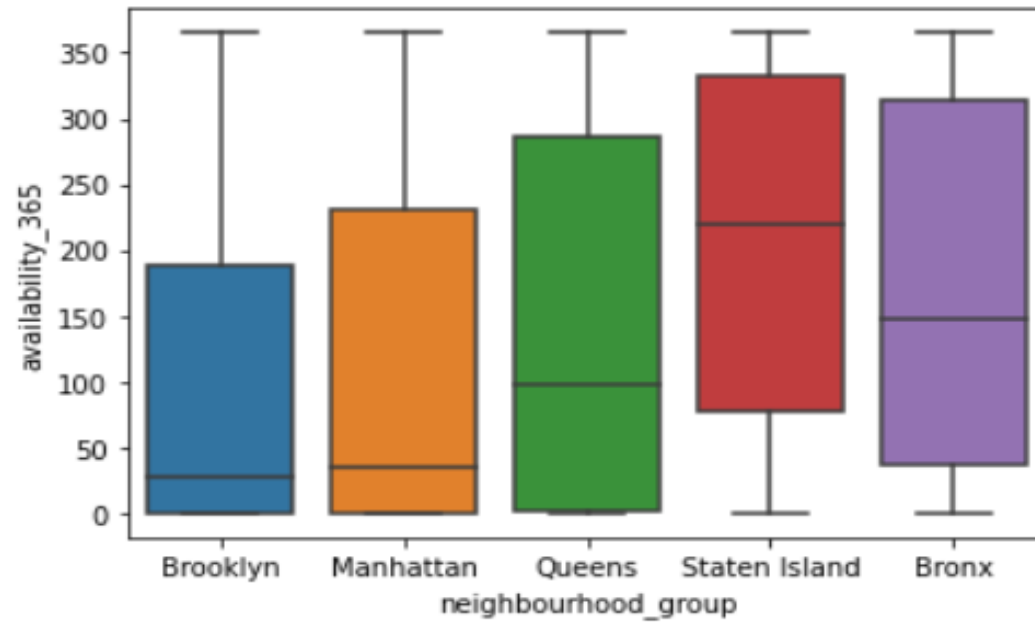
neighbourhood_group	room_type	
Bronx	Private room	652
	Entire home/apt	379
	Shared room	60
Brooklyn	Private room	10132
	Entire home/apt	9559
	Shared room	413
Manhattan	Entire home/apt	13199
	Private room	7982
	Shared room	480
Queens	Private room	3372
	Entire home/apt	2096
	Shared room	198
Staten Island	Private room	188
	Entire home/apt	176
	Shared room	9

# Total count of room types available in NYC
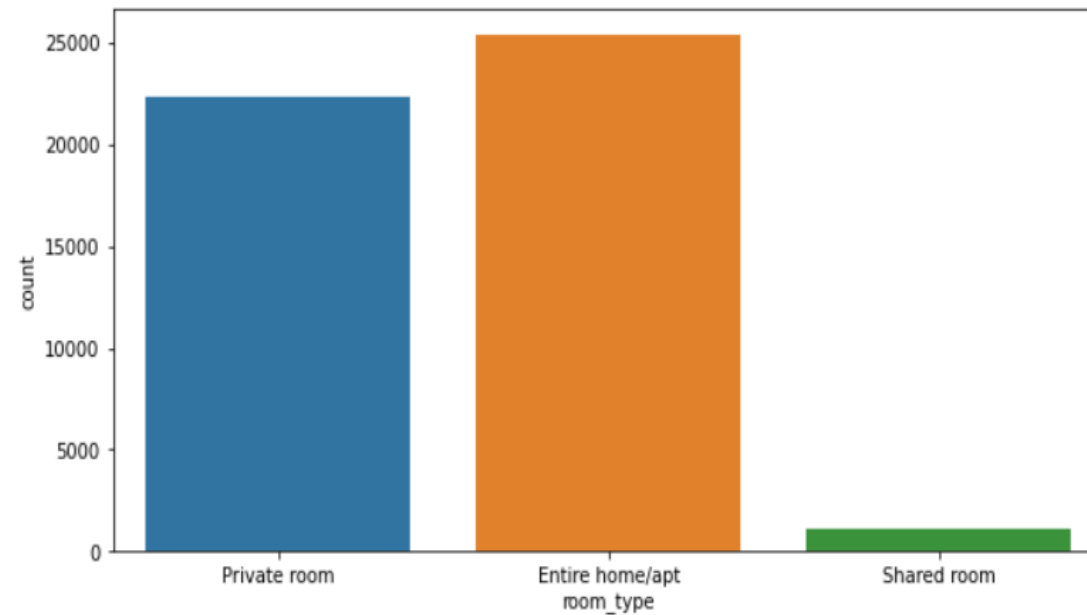

```python
plt.figure(figsize=(10,5))
sns.countplot(x='room_type', data= df)
plt.show()
```

Availability_365 and the neighborhood_group


```
sns.boxplot(data=df, x='neighbourhood_group',y='availability_365')
```

Overall contributions of each neighborhood in the count of listings throughout NYC

Neighbourhood Group

- Brooklyn — 41.1%
- Queens — 11.6%
- Bronx — 2.2%
- Staten Island — 0.8%
- Manhattan — 44.3%
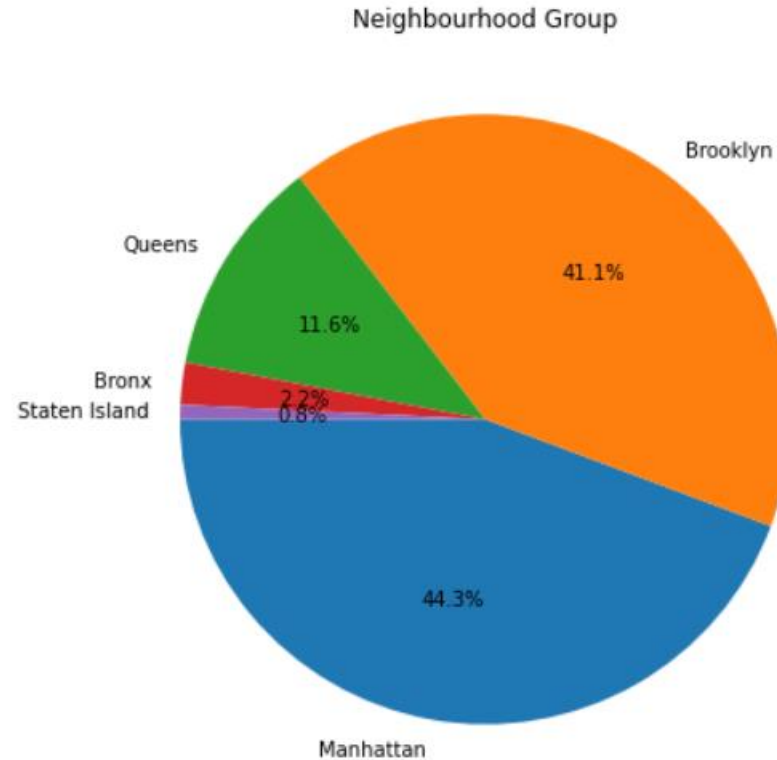
```python
plt.figure(figsize=(20,7))
plt.title("Neighbourhood Group")
plt.pie(df.neighbourhood_group.value_counts(), labels=df.neighbourhood_group.value_counts().index,autopct='%1.1f%%', startangle=180)
```

Finding the top best review of all time

```python
top_review_place=df.nlargest(10,'number_of_reviews')
top_review_place.head(5)
```

```python
plt.figure(figsize=(10,7))
sns.barplot(x='host_name',y='number_of_reviews', data=top_review_place)
plt.show()
```
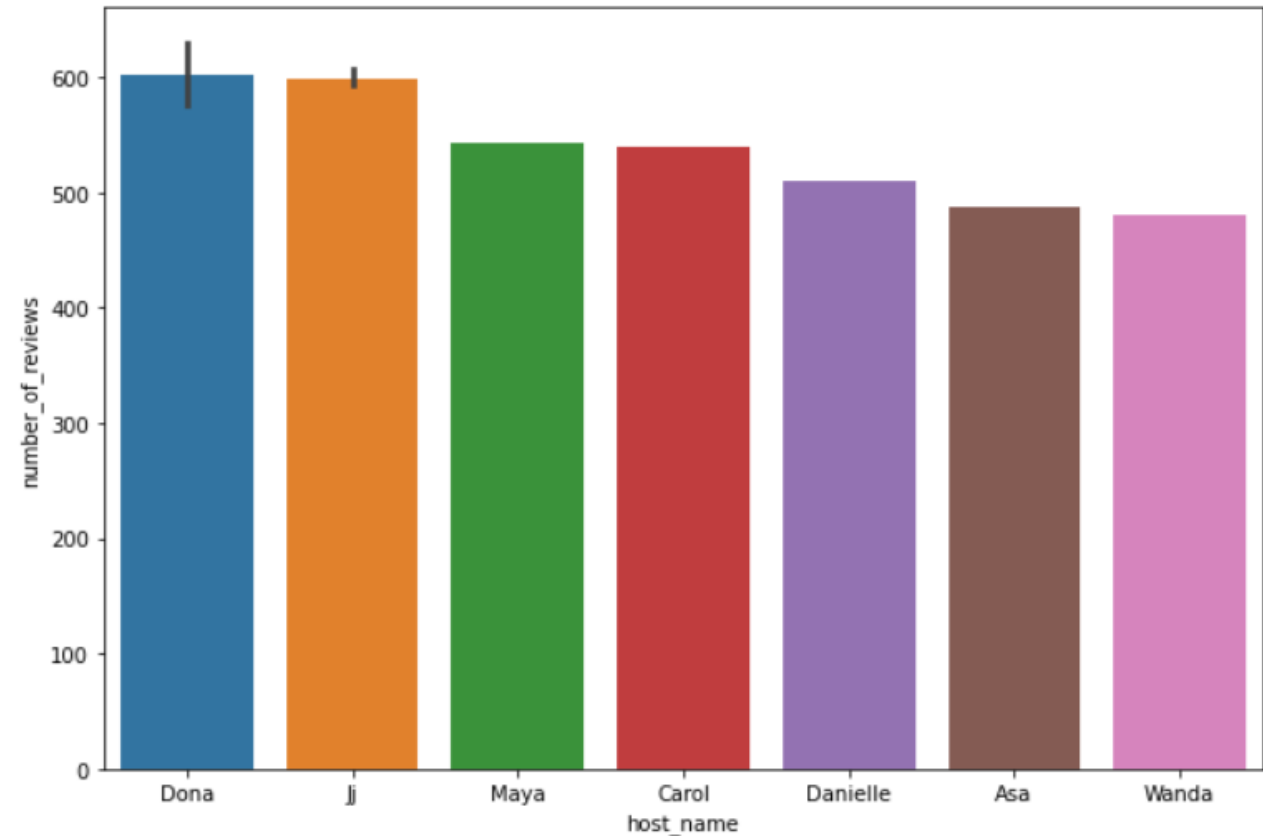
Number of reviews for each neighbourhood_group

```
nor_nbd_grp = df.groupby(['neighbourhood_group','room_type'])['number_of_reviews'].max().reset_index()
nor_nbd_grp
```

```
plt.figure(figsize=(10,7))
sns.barplot(x=nor_nbd_grp['neighbourhood_group'],hue=nor_nbd_grp['room_type'],y=nor_nbd_grp['number_of_reviews'])
```

	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
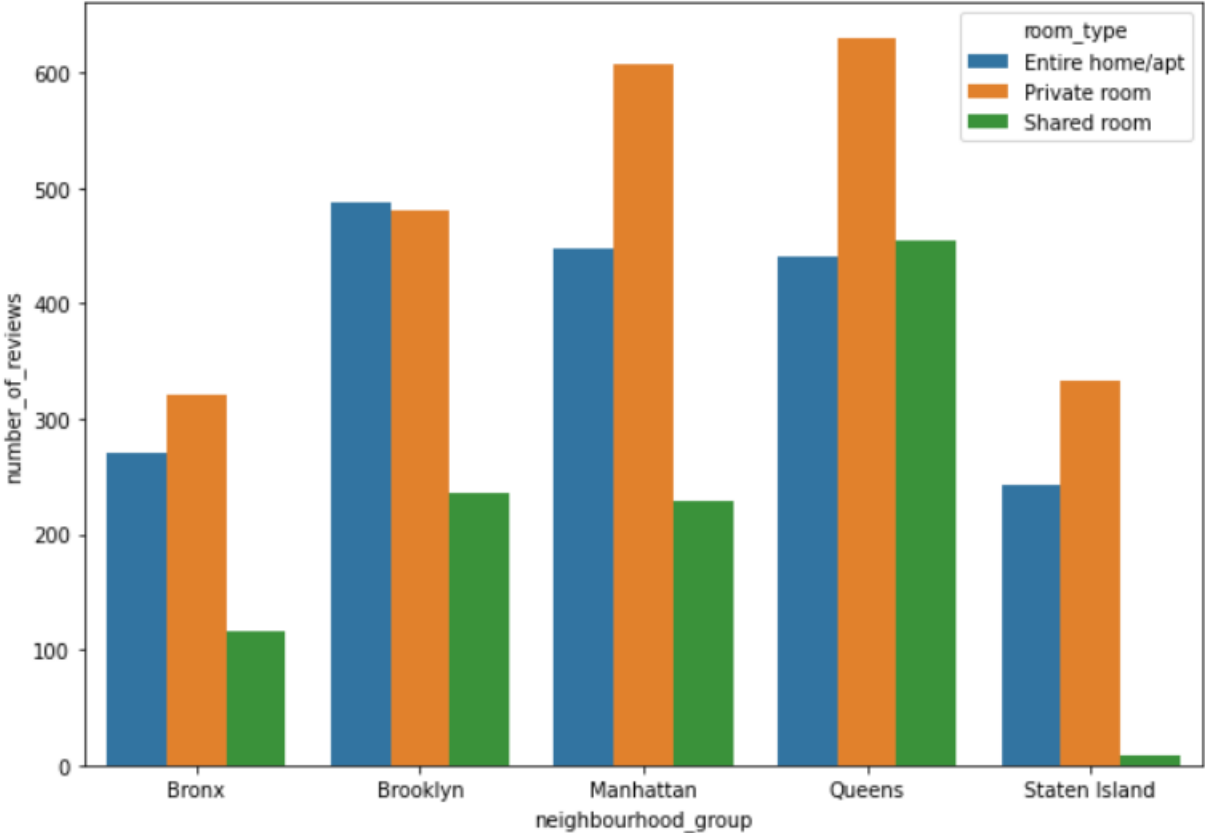
Most available room of all time

```python
most_available_place=df.nlargest(10,'availability_365')
most_available_place.head(5)
```

```python
[ ]  # Average price of most available place
     price_avrg=most_available_place.price.mean()
     print('Average price of most avialble place: {}'.format(price_avrg))
```

```
Average price of most avialble place: 150.3
```

Average price based on minimum_nights of stay in descending order

```
[ ] min_nights_stay = df.groupby(['minimum_nights','neighbourhood_group','neighbourhood'],dropna=True)['price'].mean().reset_index()
```

```
min_nights_stay.sort_values('minimum_nights',ascending=False,inplace=True)
min_nights_stay
```

	minimum_nights	neighbourhood_group	neighbourhood	price
2331	1250	Manhattan	Greenwich Village	180.000000
2330	1000	Manhattan	Battery Park City	400.000000
2329	999	Manhattan	Harlem	99.000000
2328	999	Manhattan	Greenwich Village	110.000000
2327	999	Brooklyn	Williamsburg	79.000000
...	...	...	...	...
134	1	Queens	Corona	48.891892
133	1	Queens	College Point	67.571429
132	1	Queens	Cambria Heights	76.416667
131	1	Queens	Briarwood	78.566667
0	1	Bronx	Allerton	85.363636

2332 rows × 4 columns

```
plt.figure(figsize=(11,7))
plt.scatter(min_nights_stay['minimum_nights'],min_nights_stay['price'])
plt.ylim(0, 600)
plt.xlabel('Minimum_nights')
plt.ylabel('Price')
```

Text(0, 0.5, 'Price')
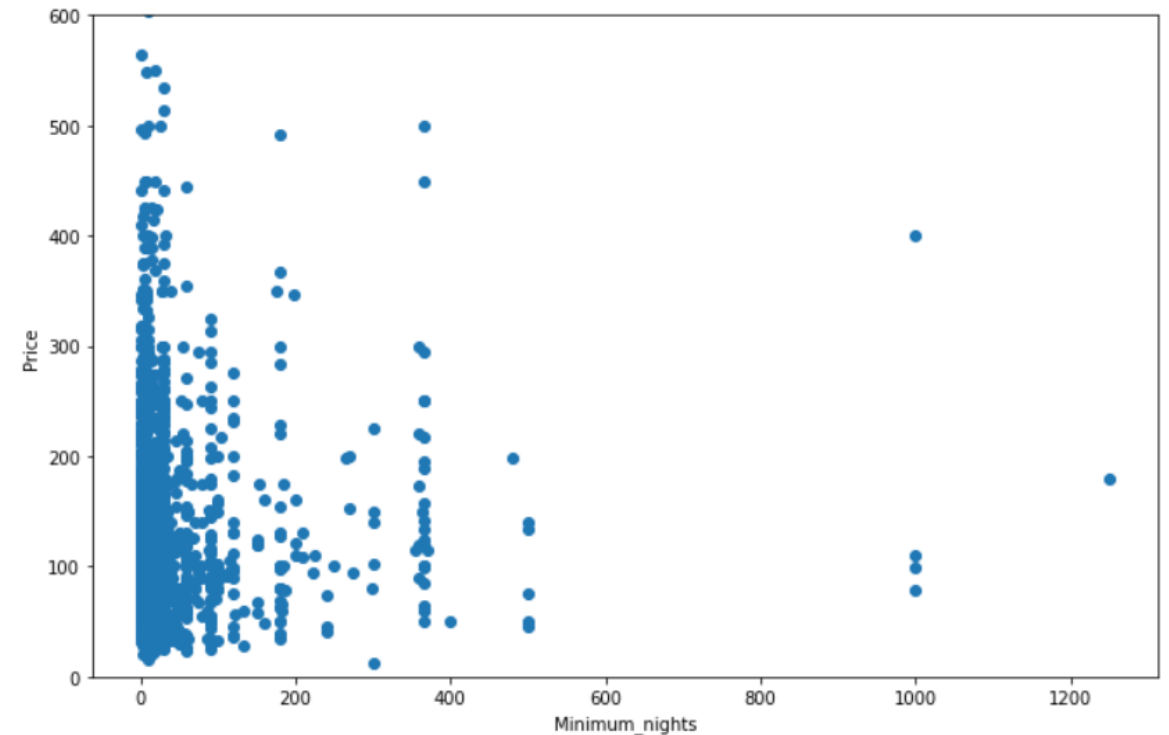
Average price per night for each neighbourhood along with room_type

```python
avg_price_per_night_nbd = df.groupby(['neighbourhood_group','room_type'])['price'].mean().reset_index()
avg_price_per_night_nbd
```

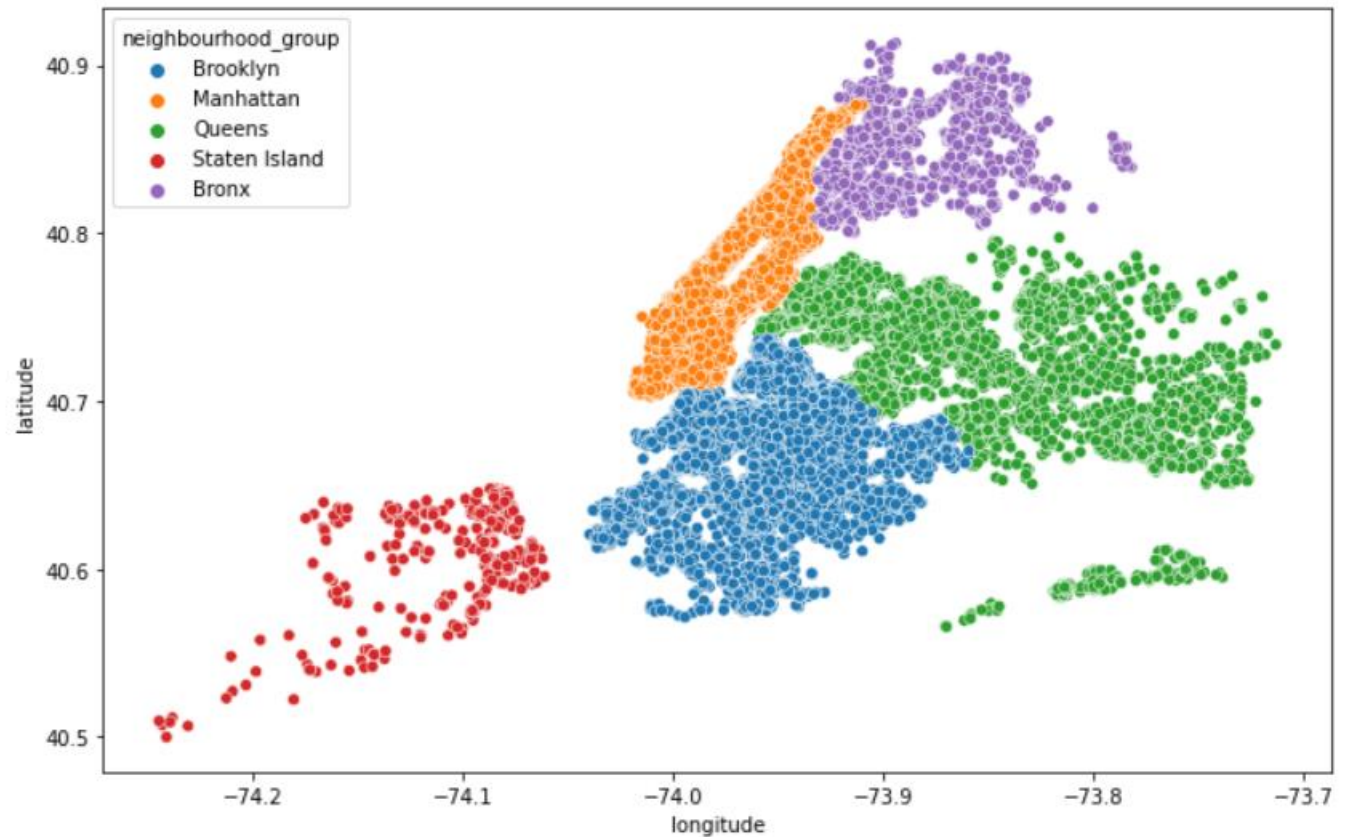|	neighbourhood_group	room_type	price
0	Bronx	Entire home/apt	127.506596
1	Bronx	Private room	66.788344
2	Bronx	Shared room	59.800000
3	Brooklyn	Entire home/apt	178.327545
4	Brooklyn	Private room	76.500099
5	Brooklyn	Shared room	50.527845
6	Manhattan	Entire home/apt	249.239109
7	Manhattan	Private room	116.776622
8	Manhattan	Shared room	88.977083
9	Queens	Entire home/apt	147.050573
10	Queens	Private room	71.762456
11	Queens	Shared room	69.020202
12	Staten Island	Entire home/apt	173.846591
13	Staten Island	Private room	62.292553
14	Staten Island	Shared room	57.444444

Barplot of average price

```
sns.barplot(x=avg_price_per_night_nbd['neighbourhood_group'],hue=avg_price_per_night_nbd['room_type'],y=avg_price_per_night_nbd['price'])
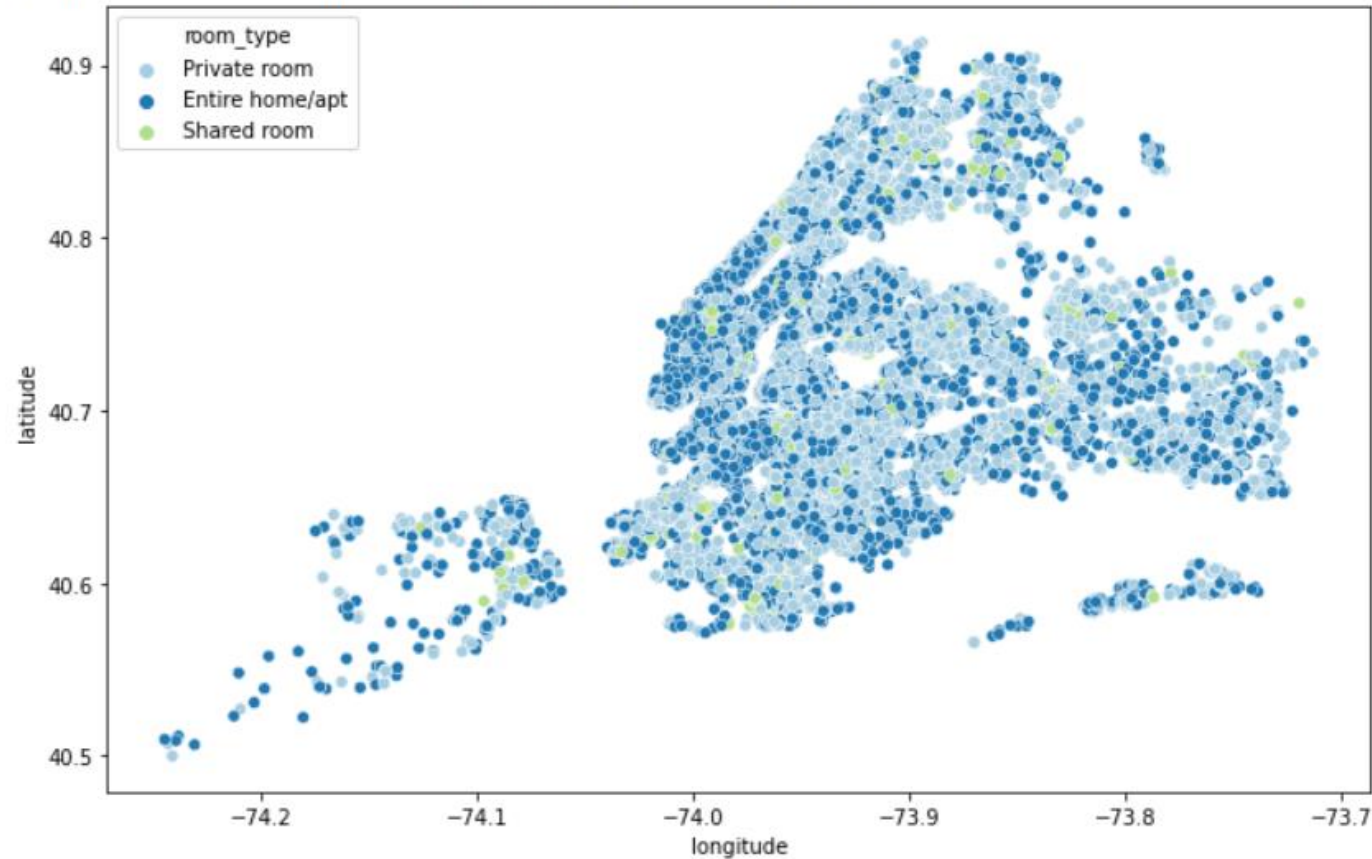```
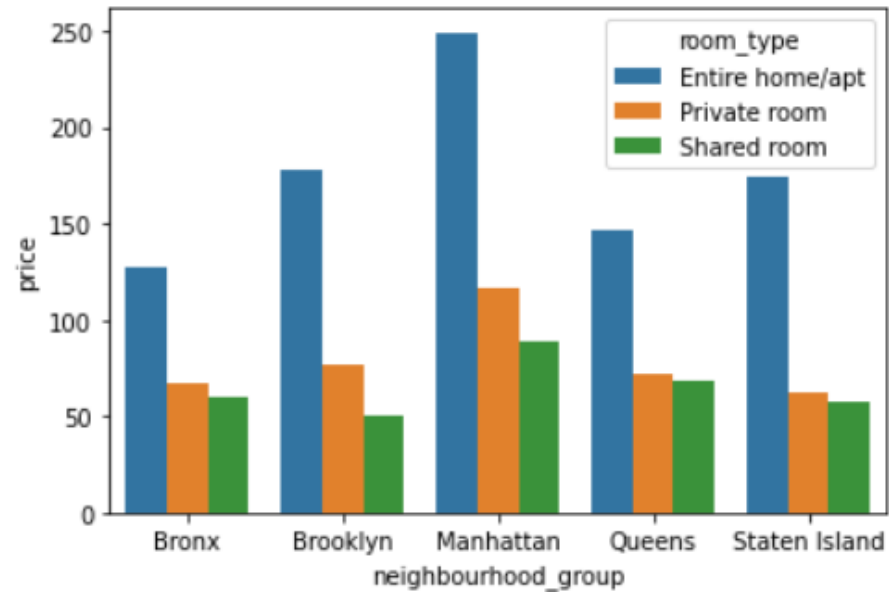
# Map based on properties of Neighbourhood_group

```
#viasualizing the newtork map from the goven data
plt.figure(figsize=(11,7))
sns.scatterplot(df.longitude,df.latitude,hue=df.neighbourhood_group)
```

Maps with respect to room_types available throughout NY

```python
plt.figure(figsize=(11,7))
sns.scatterplot(df.longitude,df.latitude,hue=df.room_type,palette='Paired')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fd1d874c730>

Count the number of rooms in every neighbourhood group

```python
count_of_room = df.groupby(['neighbourhood_group']).room_type.count().sort_values().reset_index().rename(columns={'room_type':'Count of room'})
count_of_room
```
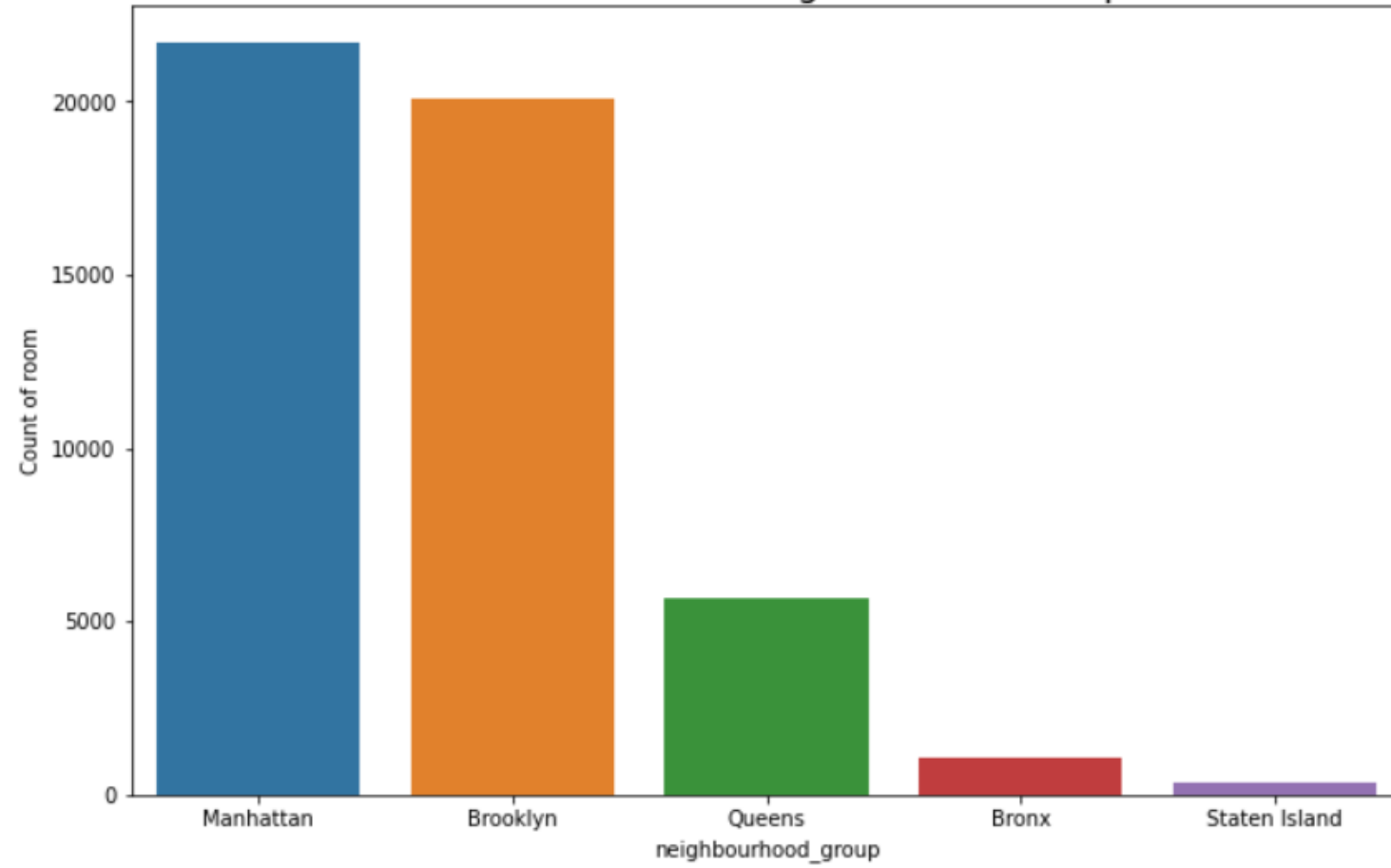
	neighbourhood_group	Count of room
0	Staten Island	373
1	Bronx	1091
2	Queens	5666
3	Brooklyn	20104
4	Manhattan	21661

```python
plt.figure(figsize=(11,7))
ax = sns.barplot(data = count_of_room.sort_values(by ='Count of room',ascending = False), x ='neighbourhood_group',y='Count of room')
plt.title('Count of rooms vs Neighbourhood Group',fontsize=17)
plt.suptitle('2019 NYC Airbnb Data',fontsize=14)
```

2019 NYC Airbnb Data

Count of rooms vs Neighbourhood Group

Now, let's check for the distribution of room types across all neighbourhood groups of NYC!

```
#scatter plot(geom_jitter) by using plotnine library to check room density distribution in each neighborhood group
(ggplot(df, aes(x = 'neighbourhood_group', y = 'room_type', color = 'room_type')) + geom_jitter() + theme_minimal() +
theme(legend_position="bottom", plot_title = element_text(vjust = 0.5)) +
labs(title = "Distribution of Room according to Neighbourhood group", x = "Neighbourhood Groups", y = "Room Types", color = " "))
```


Distribution of Room according to Neighbourhood group
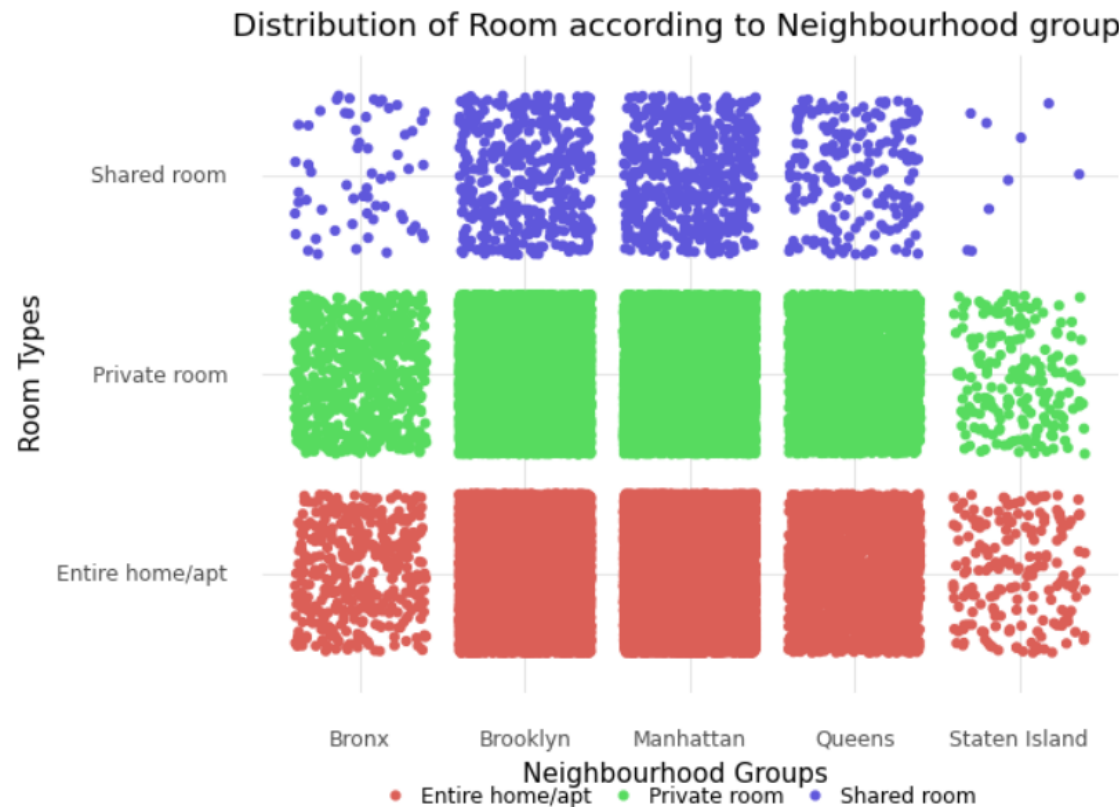
Final Conclusion

In this study, we address the explanatory analysis of the airbnb data with several key features such as price, neighbourhood, neighbourhood group, room type, number of reviews, etc. By using these data

- We obtain price and neighbourhood relationship, i.e., Manhattan is the most expensive Airbnb region when we compare the other neighbourhood groups. On the other hand, the least expensive region is Bronx.

- Another analysis is conducted by using room type. The results show that the entire home/apt type is more preferable and the others are private room and shared room, respectively.

- In terms of listings Manhattan is on the top.

- In terms of reviews Queens is on the top.

- In terms of availability Manhattan has most has most availability of room all over the year

- Number of reviews are also investigated to find which neighbourhoods take the most review according to the neighborhood group.