Capstone Project
Airbnb data analysis and visualization

Yashwany Rawat

Ritish Das

Kshitij

Manoj

Rinit

Table of Content

- **Objective**

- **Data Summary**

- **Data loading and exploration**

- **Data Wrangling**

- **Correlation heatmap**

- **Airbnb analysis**

- **Some other questions**

- **Challenges**

- **Conclusion**

Objectives

We are going to analyse Airbnb dataset .We will be discussing following steps in upcoming slides.

- Data loading and exploration.

- Data Wrangling

- Data analysis and visualization.

- Conclusion.

Data Summary

- **id :** a unique id identifying an airbnb lisitng

- **name :** name representing the accomodation

- **host_id :** a unique id identifying an airbnb host

- **host_name:** name under whom host is registered

- **neighbourhood_group :** a group of area

- **neighbourhood :** area falls under neighbourhood_group

- **latitude :** coordinate of listing

- **longitude :** coordinate of listing

- **room_type :** type to categorize listing rooms

- **price :** price of listing

- **minimum_nights :** the minimum nights required to stay in a single visit

- **number_of_reviews :** total count of reviews given by visitors

- **last_review :** date of last review given

- **reviews_per_month :** rate of reviews given per month

- **calculated_host_listings_count :** total no of listing registered under the host

- **availability_365 :** the number of days for which a host is available in a year.

Data Wrangling and Exploration

Data Loading: Loading data from google drive and reading in a notebook

Data Exploration: Checking data in columns.

Numerical columns and categorical columns.

**Information and data types of columns**

Data Wrangling

Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis.

The steps below are included in data cleansing:-

1. Dealing with missing values.

2. For analysis, some columns may be added or removed.

3. Convert the column to the correct dtype format.

4. Duplicate values are removed

Correlation Heatmap
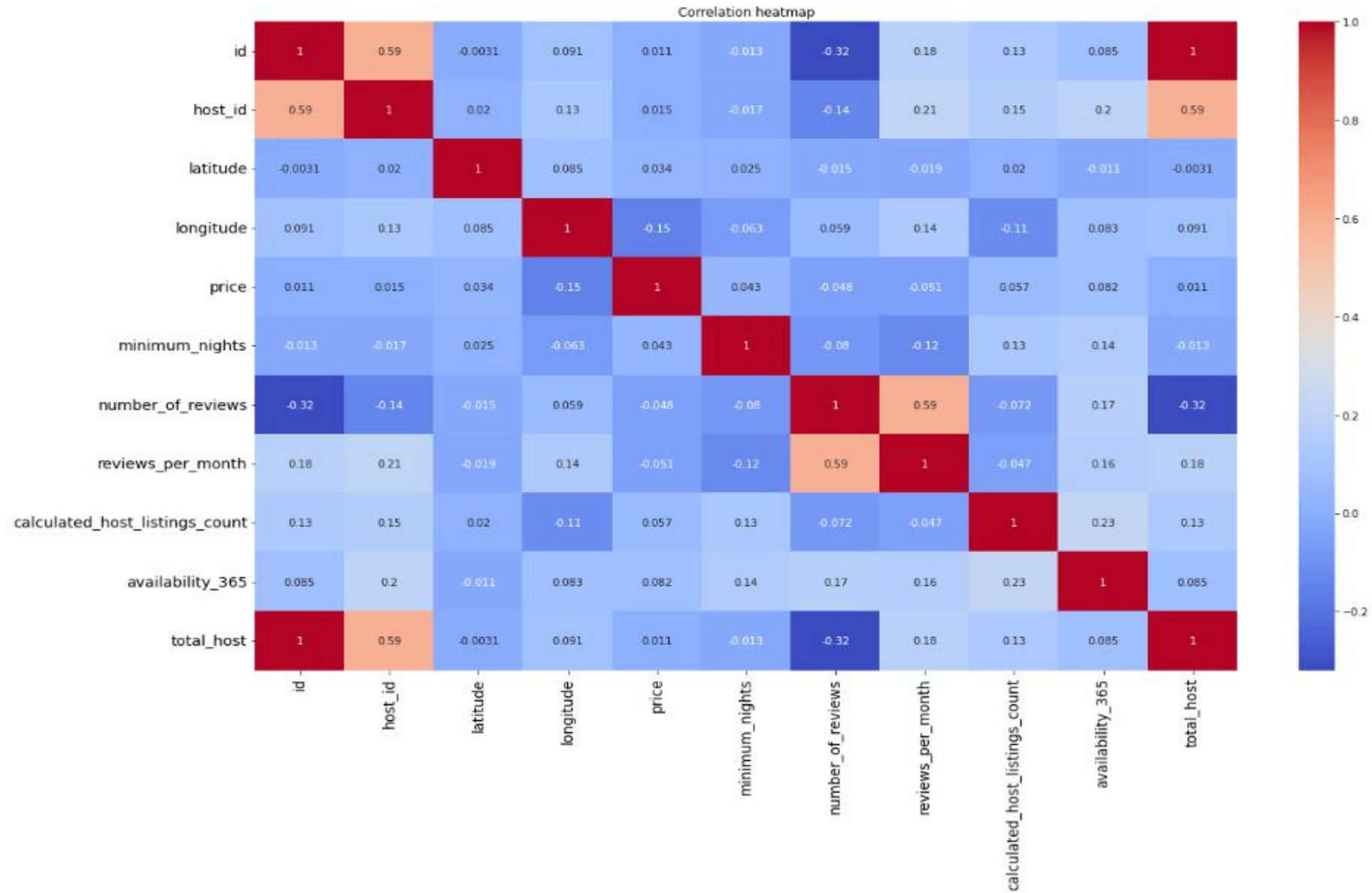
1. What can we learn about different hosts and areas?

As we can see most numbers of listings are from Manhattan-created Sonder (NYC), Blueground, Michael, David

	host_name	neighbourhood_group	calculated_host_listings_count
13221	Sonder (NYC)	Manhattan	327
1837	Blueground	Brooklyn	232
1838	Blueground	Manhattan	232
7279	Kara	Manhattan	121
7482	Kazuya	Brooklyn	103

# 2. What can we learn from predictions? (ex: locations, prices, reviews, etc)

	neighbourhood_group	number_of_reviews
0	Bronx	321
1	Brooklyn	488
2	Manhattan	607
3	Queens	629
4	Staten Island	333

Area vs Number of reviews
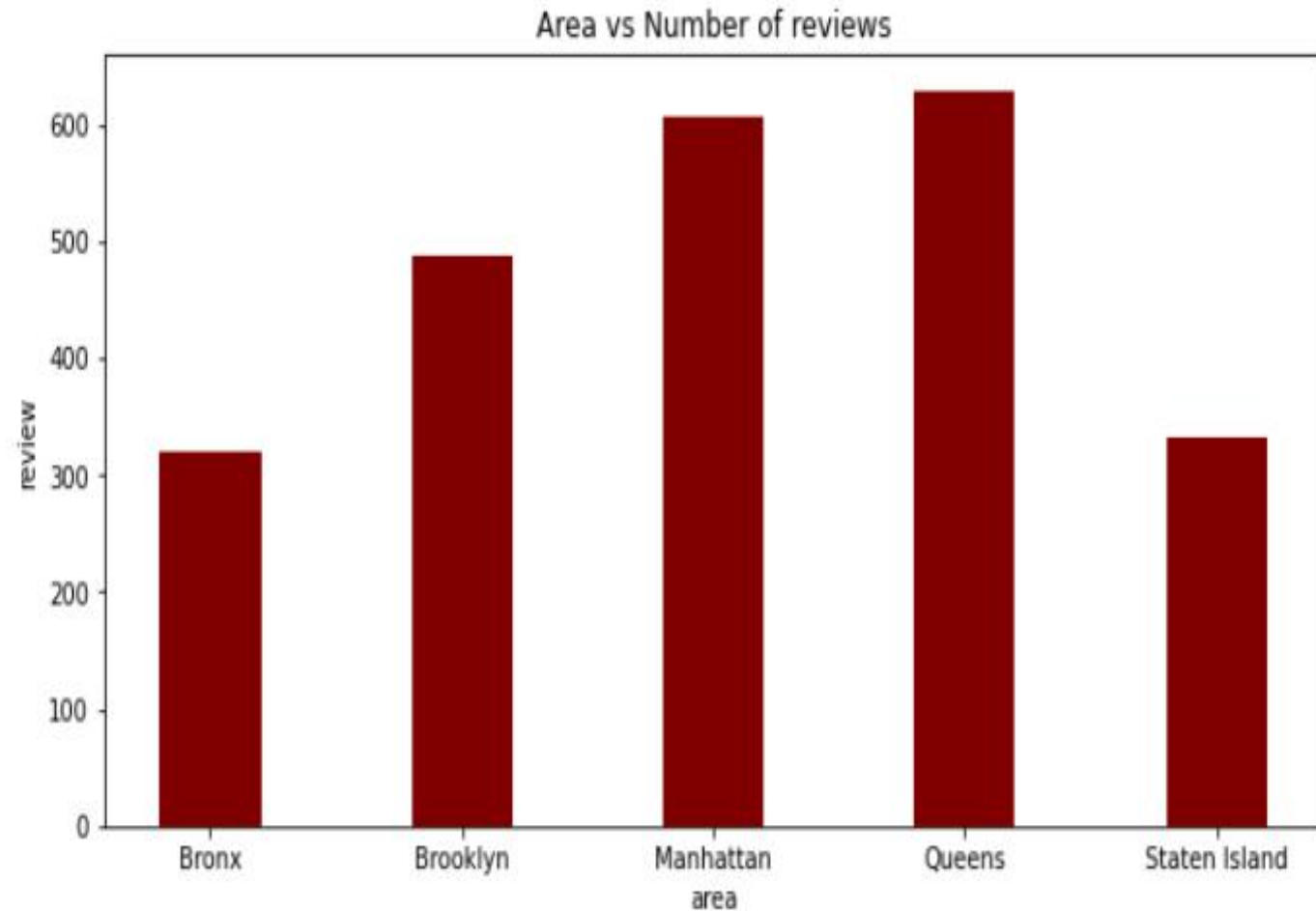
2. What can we learn from predictions? (ex: locations, prices, reviews, etc)

	price	number_of_reviews
0	0	95
1	10	93
2	11	113
3	12	8
4	13	9

From the above Analysis, we can say that most people prefer to stay in a place where the price is less.

Price vs Number of Reviews
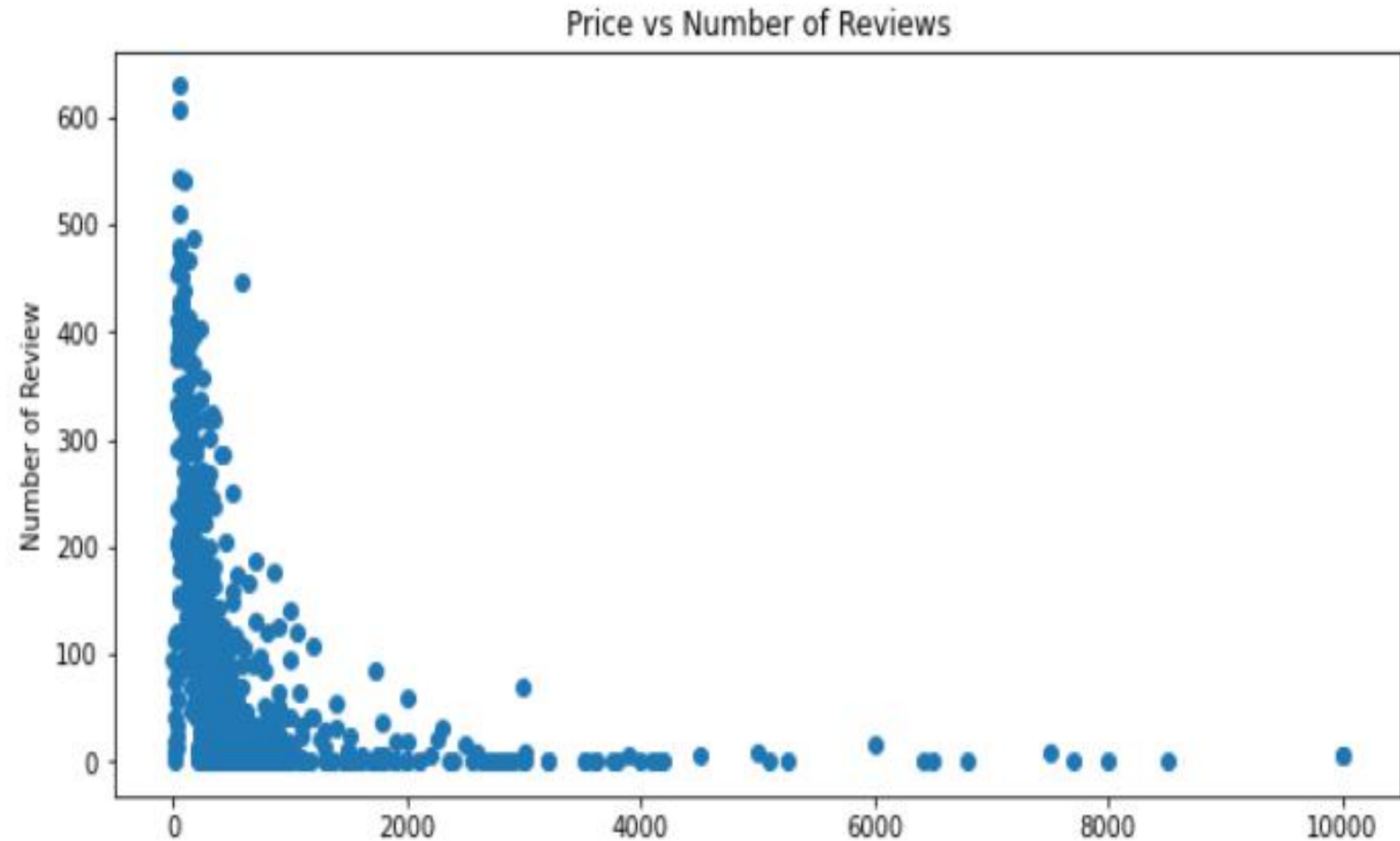
3. Which hosts are the busiest and why?

|	host_name	host_id	room_type	number_of_reviews
10297	Dona	47621202	Private room	629
17726	Jj	4734398	Private room	607
25584	Maya	37312959	Private room	543
6253	Carol	2369681	Private room	540
8965	Danielle	26432133	Private room	510
3968	Asa	12949460	Entire home/apt	488
37766	Wanda	792159	Private room	480
22517	Linda	2680820	Private room	474
8643	Dani	42273	Entire home/apt	467
2960	Angela	23591164	Private room	466
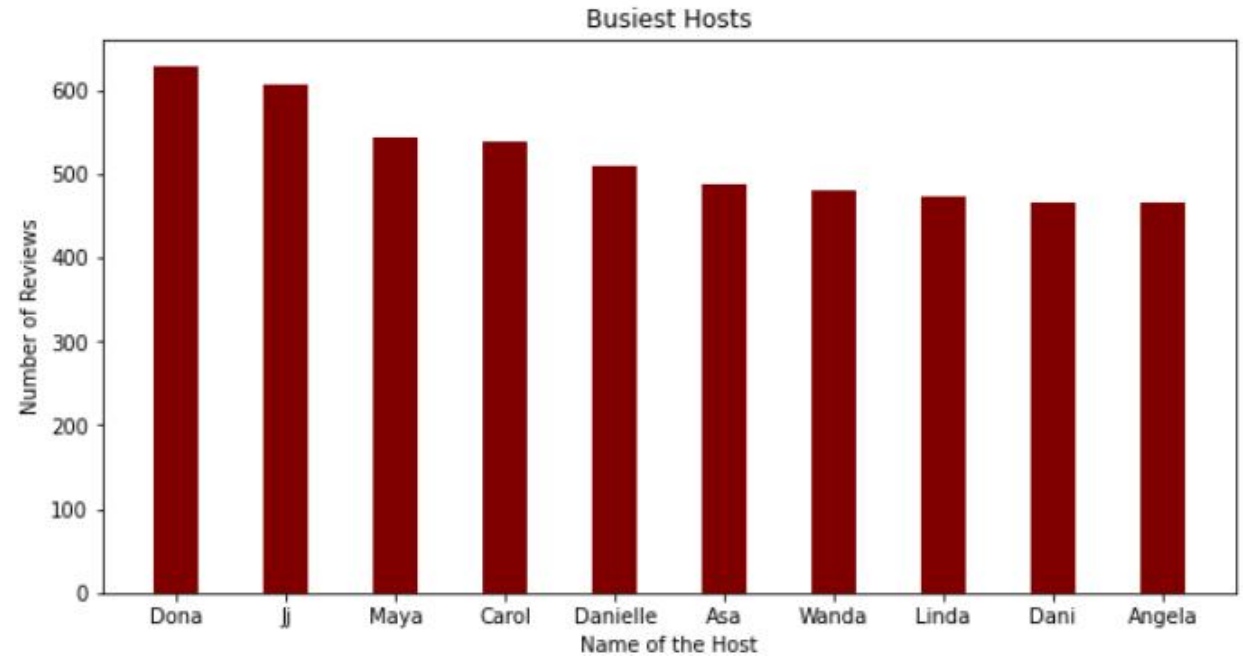
Busiest Hosts

Busiest hosts are:

- Dona
- Ji
- Maya
- Carol
- Danielle

Because these hosts listed room type as Entire home and Private room which is preferred by the most number of people.

4. Is there any noticeable difference of traffic among different areas and what could be the reason for it?
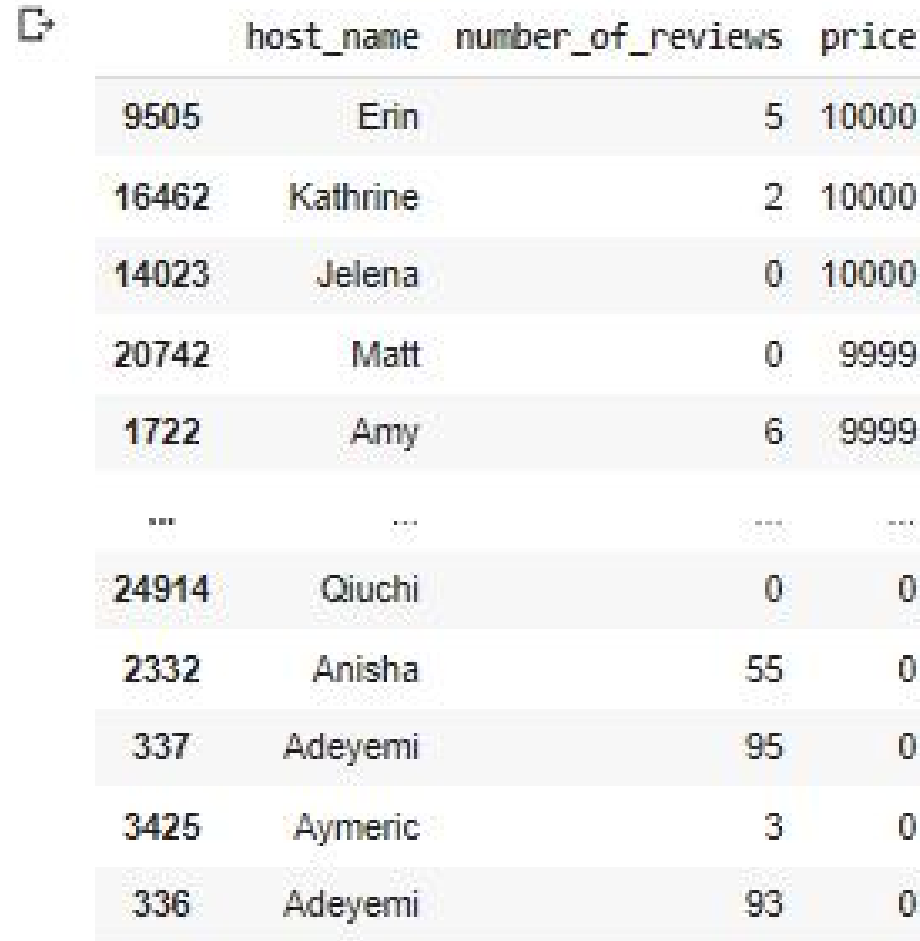
	neighbourhood_group	room_type	minimum_nights
6	Manhattan	Entire home/apt	13199
4	Brooklyn	Private room	10132
3	Brooklyn	Entire home/apt	9559
7	Manhattan	Private room	7982
10	Queens	Private room	3372
9	Queens	Entire home/apt	2096
1	Bronx	Private room	652
8	Manhattan	Shared room	480
5	Brooklyn	Shared room	413
0	Bronx	Entire home/apt	379
11	Queens	Shared room	198
13	Staten Island	Private room	188
12	Staten Island	Entire home/apt	176
2	Bronx	Shared room	60
14	Staten Island	Shared room	9

Traffic Areas
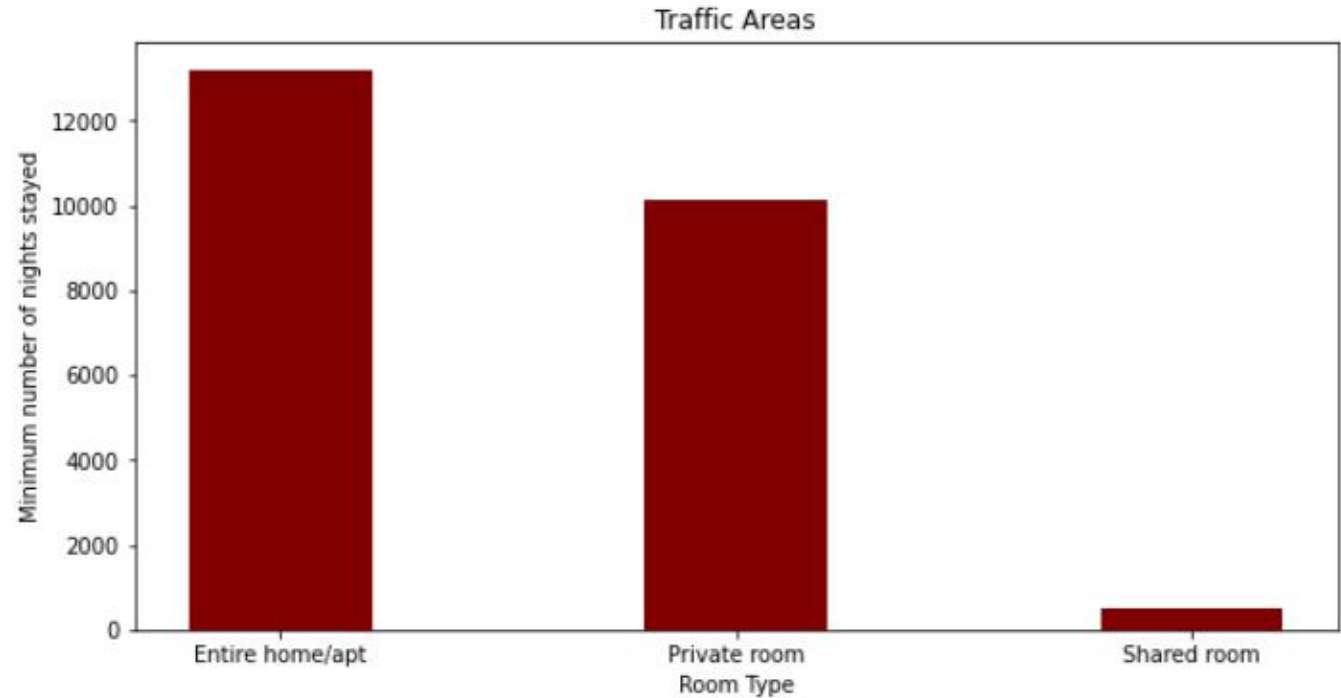
From the Above Analysis, We can Say that People are preferring Entire homes/apt or Private rooms which is present in Manhattan, Brooklyn, and Queens and people are preferring listings that are less in price

5. Who is the most expensive area's host name.

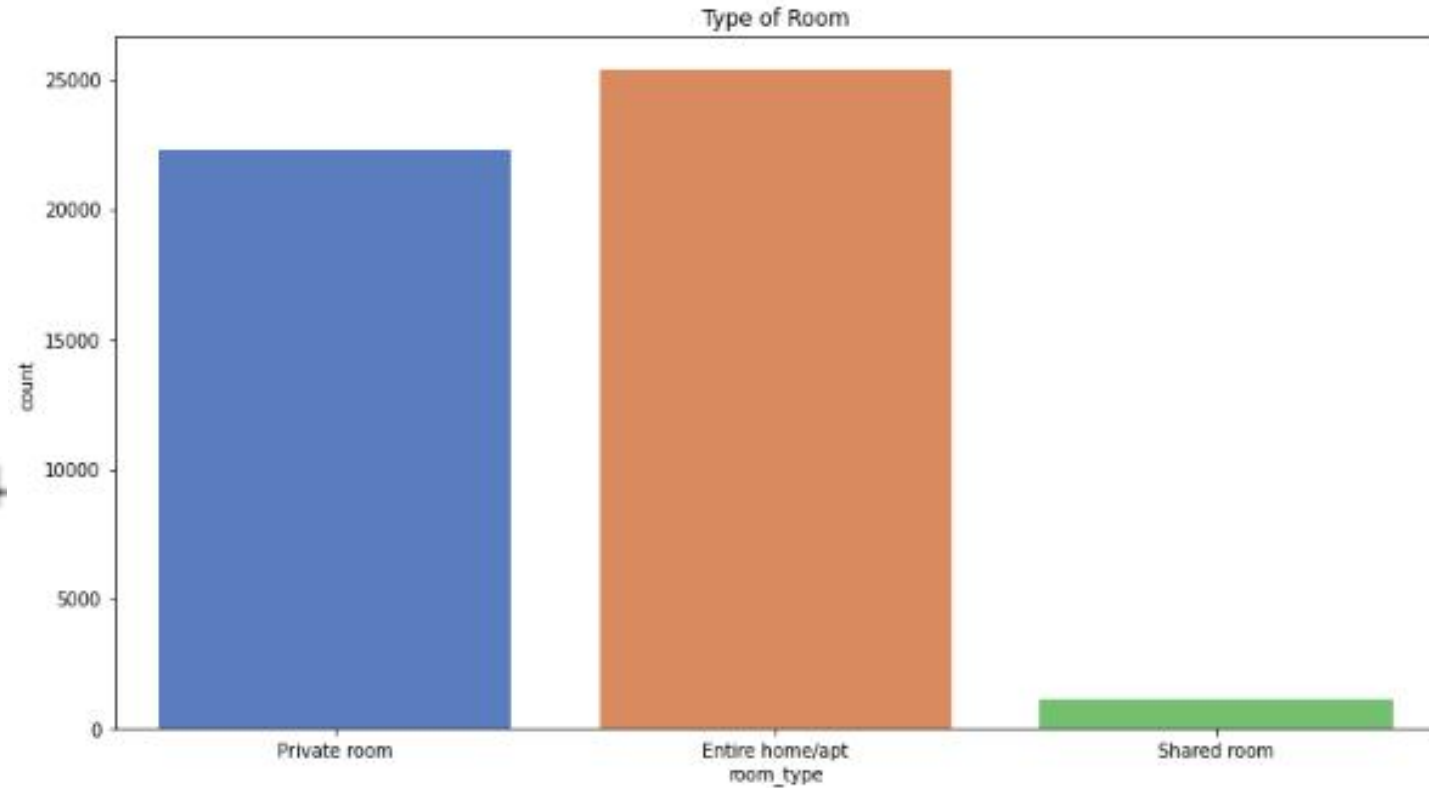**Jelena, Erin and Kathrine are the host of expensive area**

	host_name	number_of_reviews	price
9505	Erin	5	10000
16462	Kathrine	2	10000
14023	Jelena	0	10000
20742	Matt	0	9999
1722	Amy	6	9999
...	...	...	...
24914	Qiuchi	0	0
2332	Anisha	55	0
337	Adeyemi	95	0
3425	Aymeric	3	0
336	Adeyemi	93	0

32551 rows × 3 columns

6. Mostly preffered room type

Entire home/apt   25409
Private room   22326
Shared room   1160
Name: room_type, dtype: int64

6. Mostly preffered room type
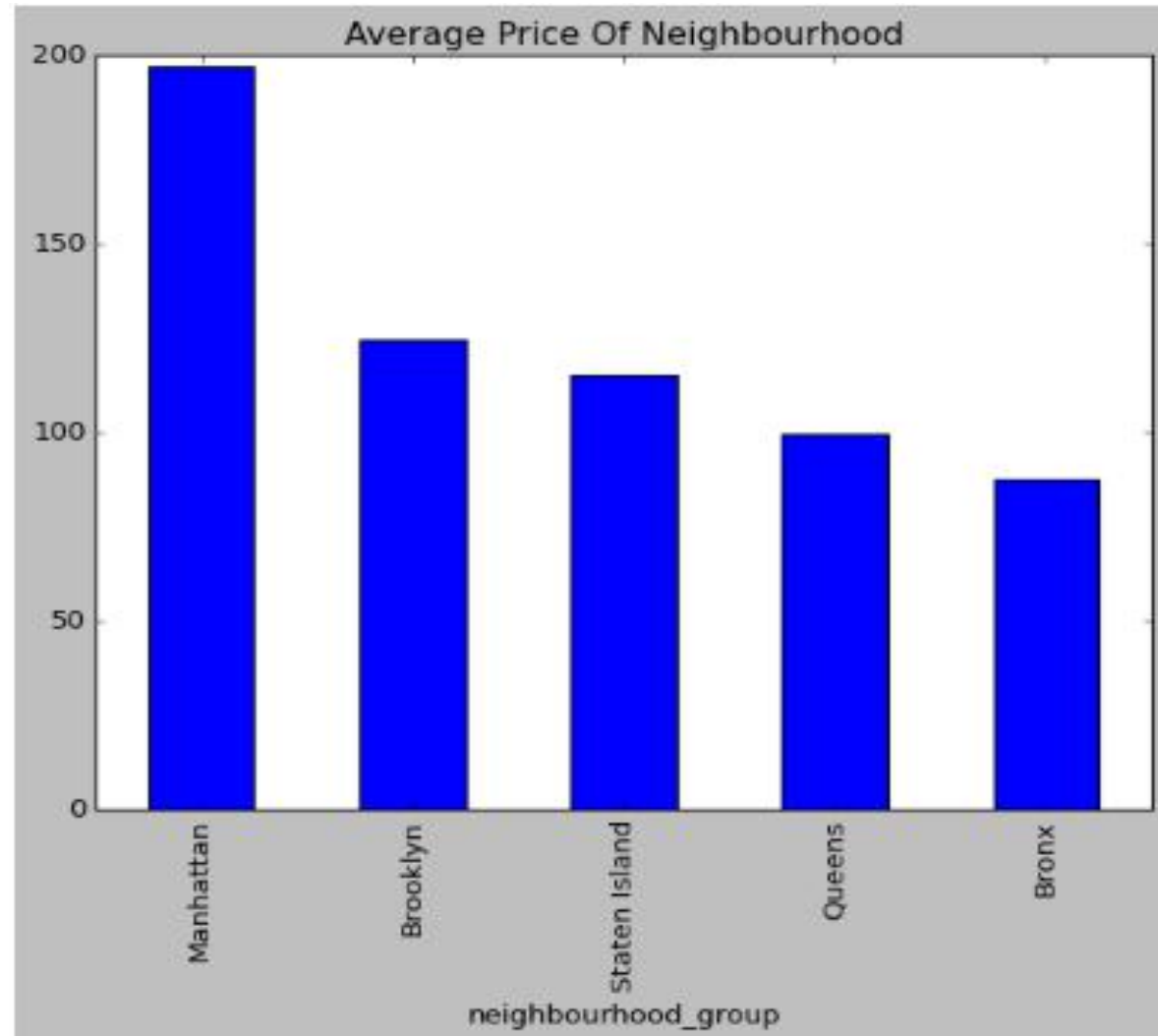

Room Type on Neighbourhood Group
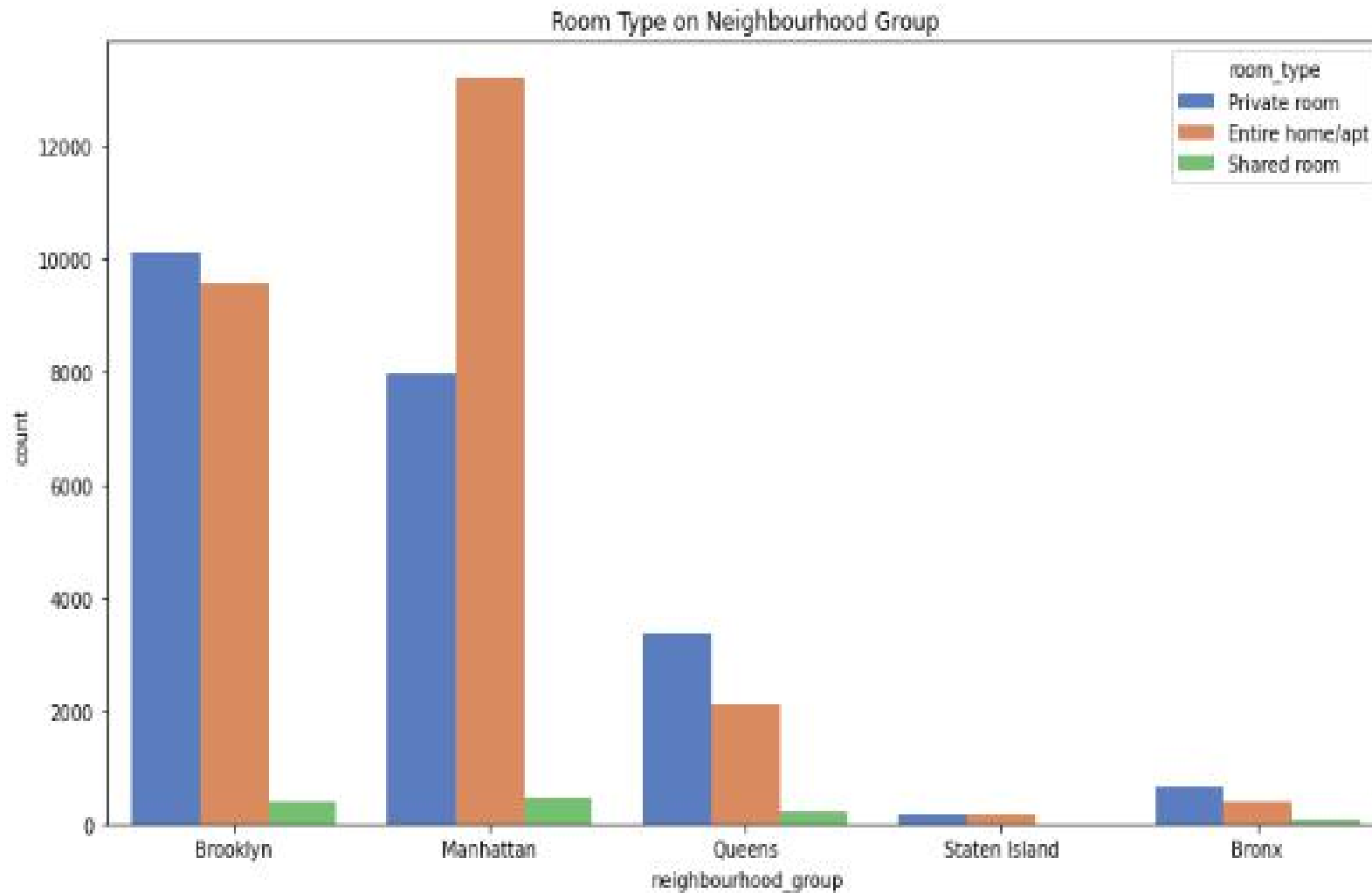
Entire home/apt prefered by people over private and shared room

7. Average Price of neighbourhood group

```
neighbourhood_group
Manhattan          196.875814
Brooklyn           124.383207
Staten Island      114.812332
Queens              99.517649
Bronx               87.496792
Name: price, dtype: float64
```

The representation of the average price of the neighbourhood group

Average Price Of Neighbourhood
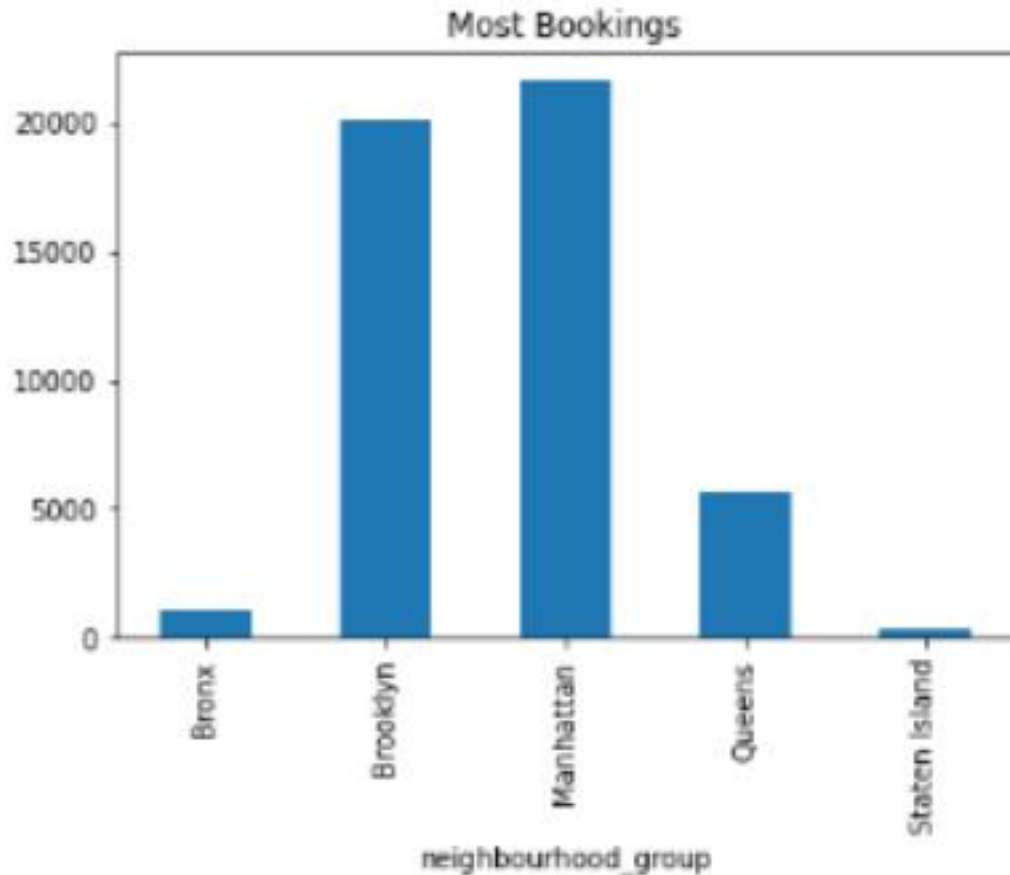
8. Which Neighbourhood Group had Most Bookings

```
neighbourhood_group
Bronx               1091
Brooklyn           20104
Manhattan          21661
Queens              5666
Staten Island        373
Name: room_type, dtype: int64
```
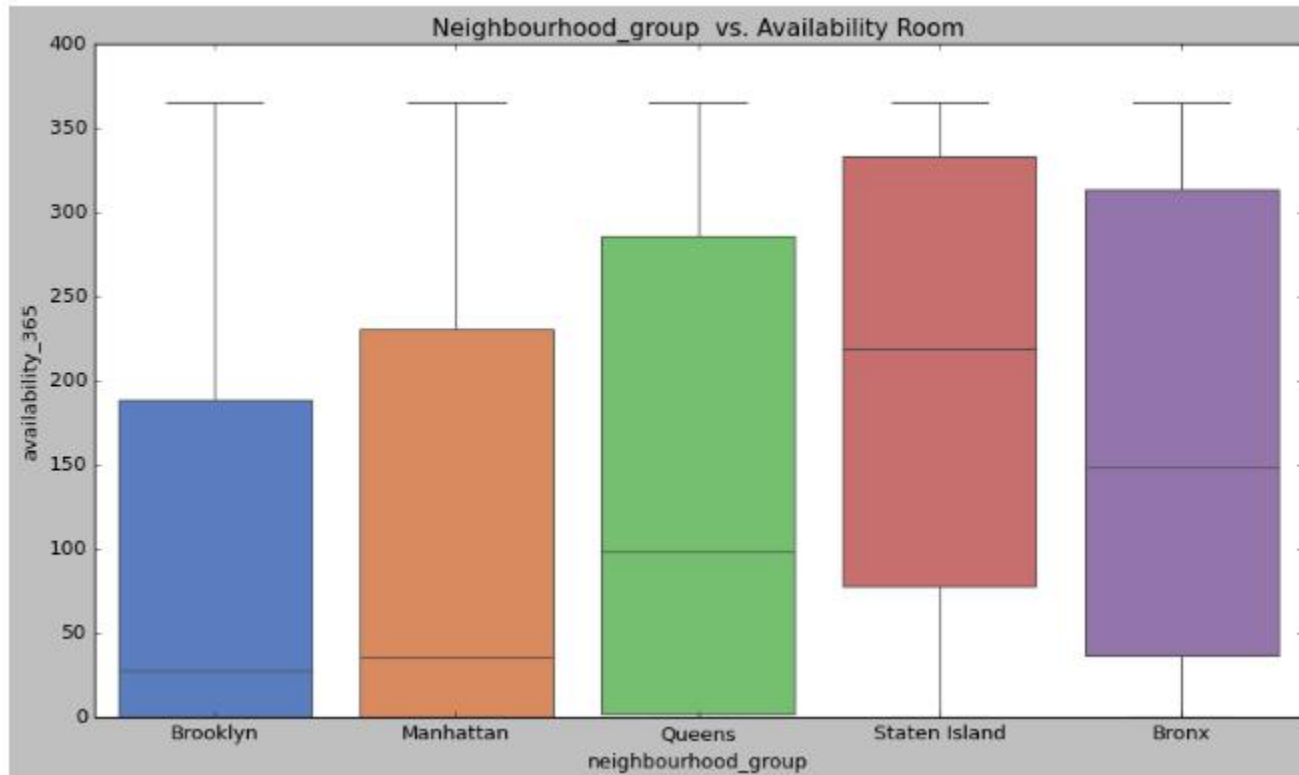

Neighbourhood Group
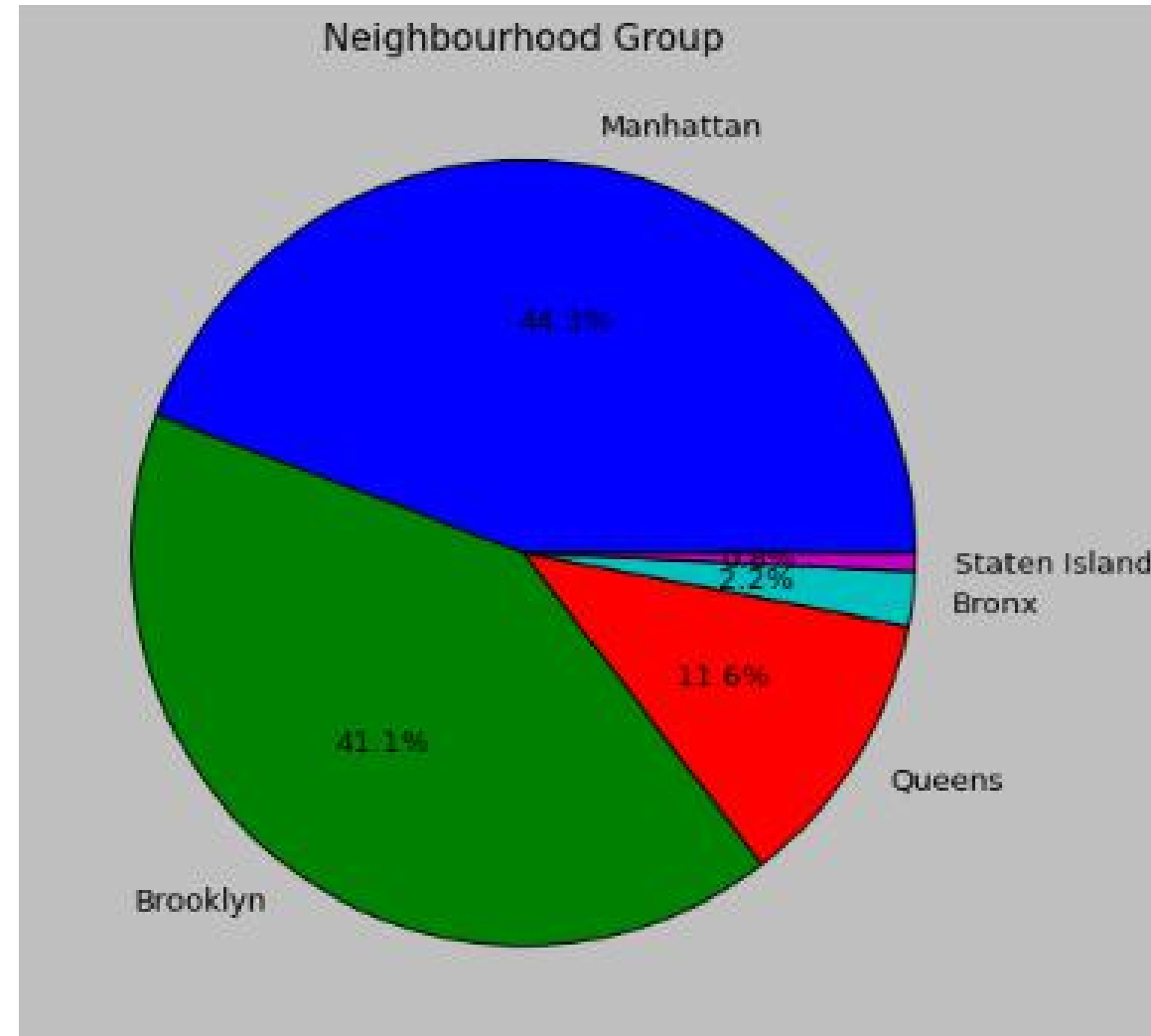
8. Which Neighbourhood Group had Most Bookings


```
last_review
2019-06-23    1413
2019-07-01    1359
2019-06-30    1341
2019-06-24     875
2019-07-07     718
              ...
2015-01-07       1
2014-12-30       1
2014-12-22       1
2014-12-19       1
2011-03-28       1
Name: last_review, Length: 1764, dtype: int64
```
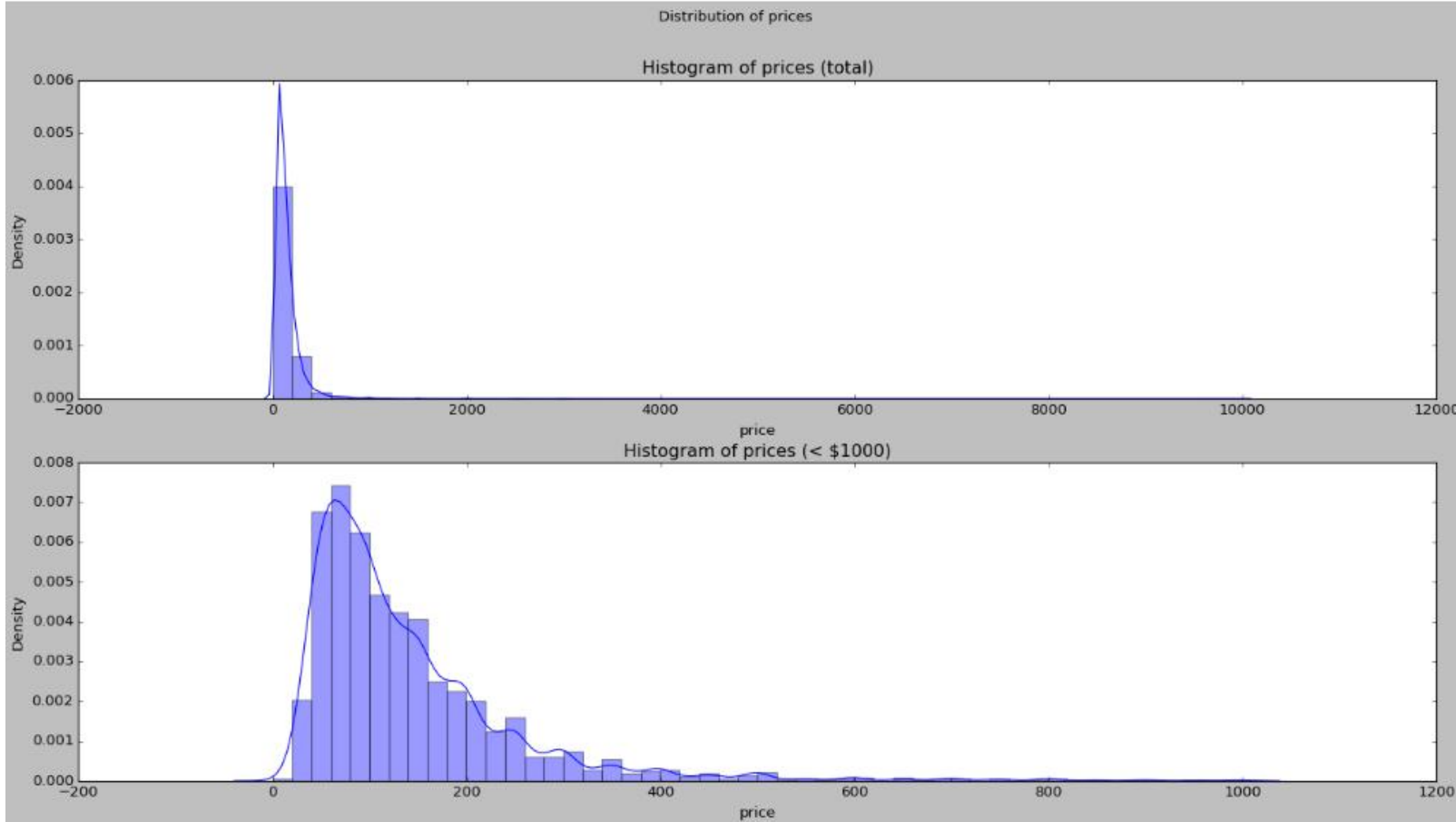
AIRBNB in Manhattan had the most bookings than any other Neighbourhood group and the least bookings are from staten Island

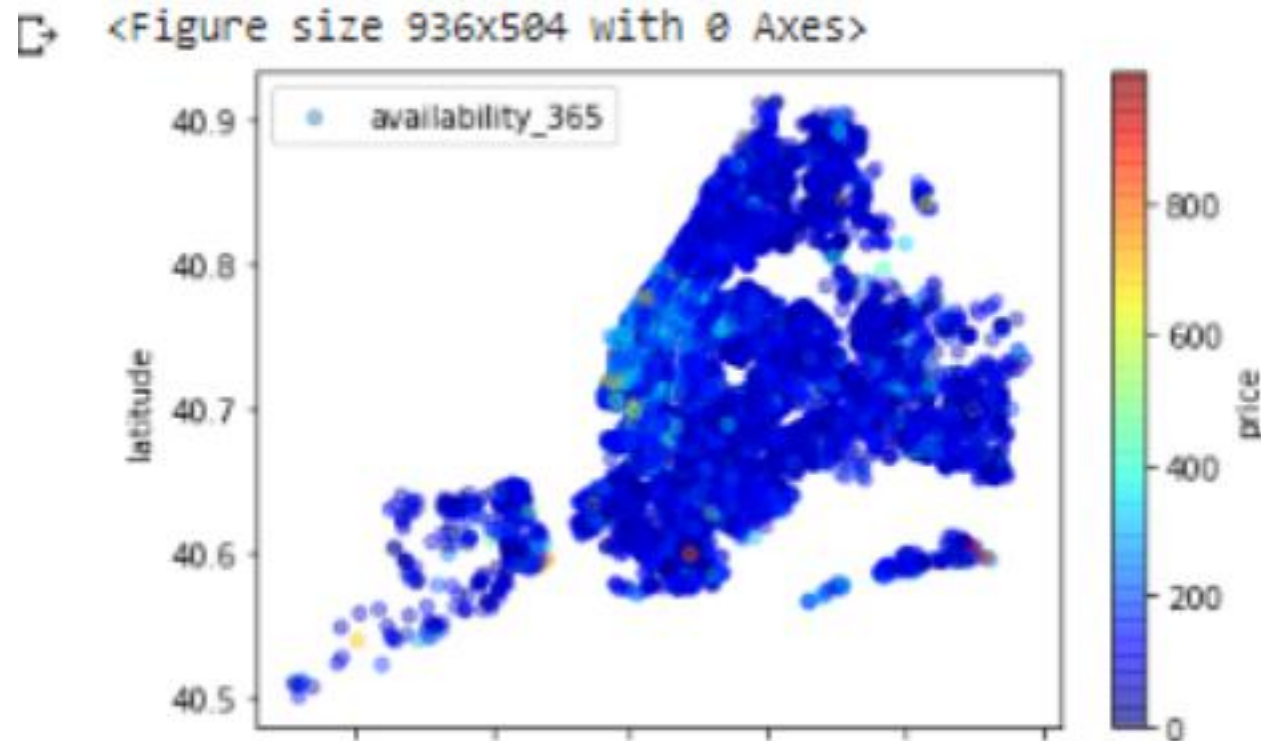# 9. Neighbourhood group vs. Availability Room

As we can see the relationship between the neighbourhood and availability

10. Which Neighbourhood group has most number of EntireHome/APT Room_Type ?

Here is histogram of price and price less than 1000

11. What is the price distribution

`<Figure size 936x504 with 0 Axes>`

The dot in red colors has more price than other room type and Manhattan is expensive

12. Which host available at the time of night

	host_name	minimum_nights
3590	Genevieve	1250
8281	Peter	1000
7018	Meg	999
3751	Glenn H.	999
675	Angie	999
4989	John	500
5923	Laura	500
472	Amanda	500
395	Aliya	500
9579	Shining	480

The name of the host name is Genevieve under which people does more night stay

13. Is there a relationship between price and reviews?
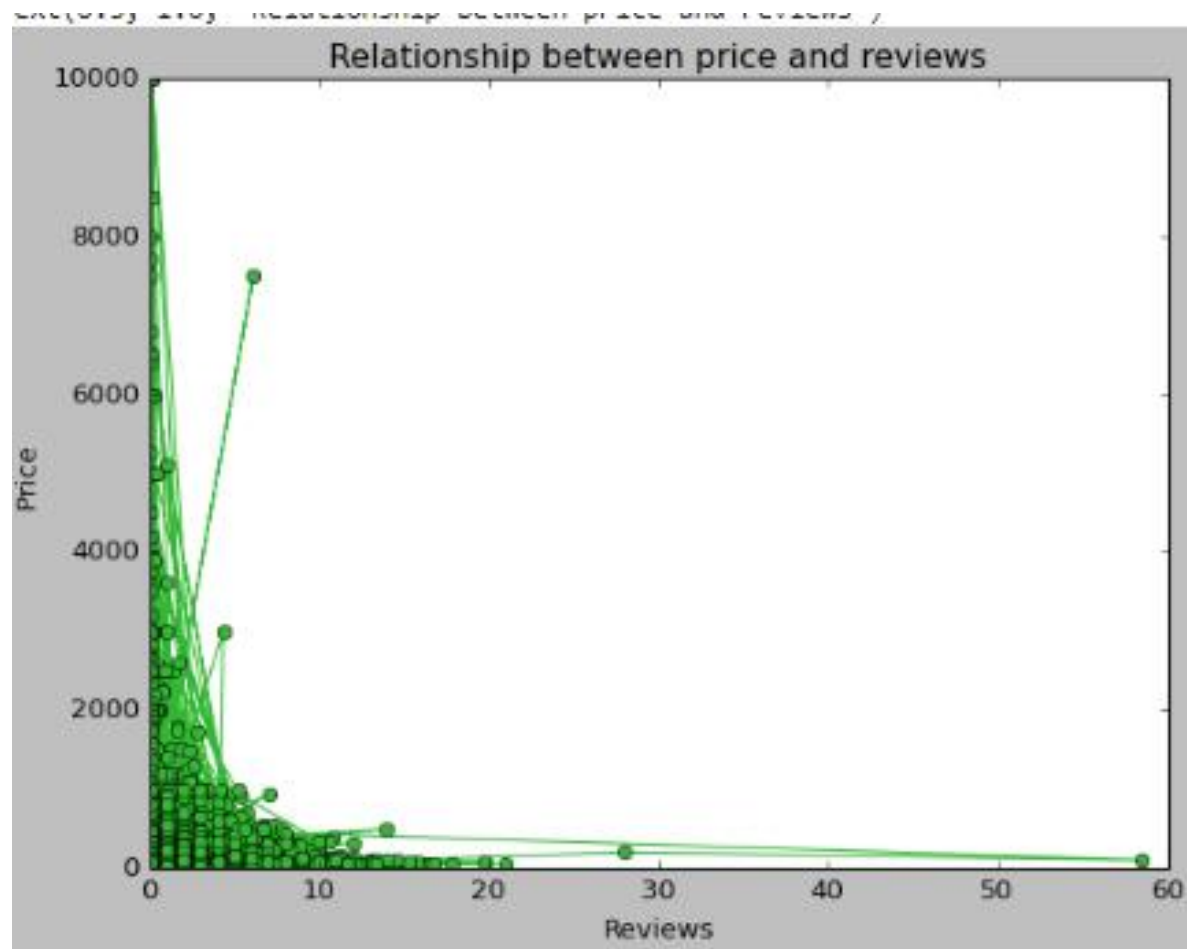
There is a negative trend relationship between price and reviews.
If the no. of reviews increses, price will decreases.

Challenges

- Choosing visualization for different analysis.

Conclusion

1. Those who want to live in an entire house or apartment will stay in that neighbourhood for a little while longer.

2. Compared to a home or apartment, people who choose to stay in a private room won't remain as long.

3. The majority of individuals prefer to pay less.

4. A location is a tourist destination if there are more reviews for a particular neighbourhood group.

5. People are considered visitors if they do not remain more than one night.

6. People prefer entire apartment over private rooms and shared rooms

7. The highest average price of the neighbourhood is 196.875814

Thank You