Capstone Project

Airbnb Bookings Analysis

By Nayana Pradeep

PROBLEM STATEMENT

- Given is a rich dataset of the New York 2019 Airbnb Analysis which includes important details such as area, price, host name etc.

- Explore and analyze the data to have an understanding of how the Airbnb is being operated in NYC and the various factors responsible for the ups and downs of the business

OVERVIEW

Airbnb is an American company that operates an online marketing place for lodging, homestays for vacation and tourism activities.

The platform is accessible via website and mobile app. Airbnb does not own any of the listed properties. It profits by receiving commission from each booking.

During the last few years, the popularity of Airbnb rose so much so that it is now considered as a rival for all the hotel industries.

DATA SUMMARY

- airbnb_df has 48895 rows and 16 columns

Column Name	Description
id	Unique id for each listing
name	Name of the place of stay
host_id	Unique id for each host
host_name	Host name
Neighbourhood_group	Represents 5 main boroughs of New York . They are Manhattan, Brooklyn, Queens, Staten Island and Bronx
neighbourhood	Location of these listings ;There are 221 unique neighbourhoods.
latitude	Latitude coordinates
longitude	Longitude coordinates

DATA SUMMARY (Contd.)

Column Name	Description
room_type	There are 3 types of room :Shared room, Private room and an Entire Apartment
price	Price of each listing
minimum_nights	Minimum number of nights for which the booking is possible
number_of_reviews	Total number of reviews
last_review	Latest review
reviews_per_month	Number of reviews per month on an average.
calculated_host_listings_count	Number of listings per host
availability_365	Number of days for which the listing is available for booking

AIRBNB NUMERICAL STATISTICS

- The mean price of rooms is 157 dollars. The maximum price of a room is 10000 dollars.

- Surprisingly, there are listings available with minimum stay for more than an year .The maximum values for minimum_nights stay is 1250 which approximately is 3 and a half years

- Maximum number of reviews received by a listing is 629.

- There are certain listings with availability_365 as 0. This could either indicate that those listings are temporarily unavailable or bad data.

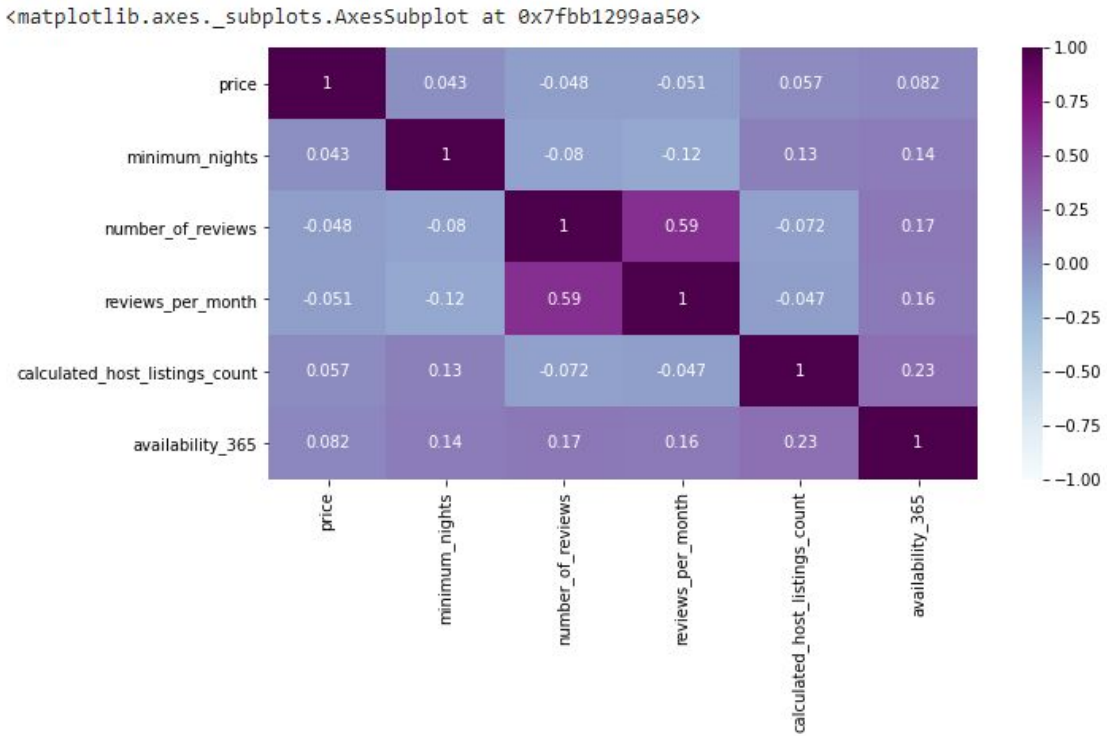
Handling Null Values

```
#Find the number of columns with null value for each column
airbnb_df.isnull().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
```

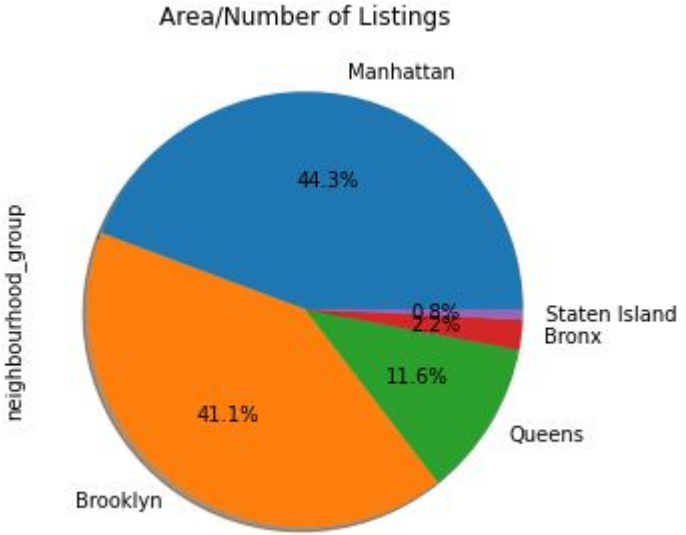
- Replaced the missing names of ‘name’ column with ‘NAME’
- Replaced the missing host names of ‘host_name' with ‘HOST NAME’
- Dropped the ‘last_review’ column
- Replaced the missing values of ‘reviews_per_month’ with 0

CORRELATION HEATMAP

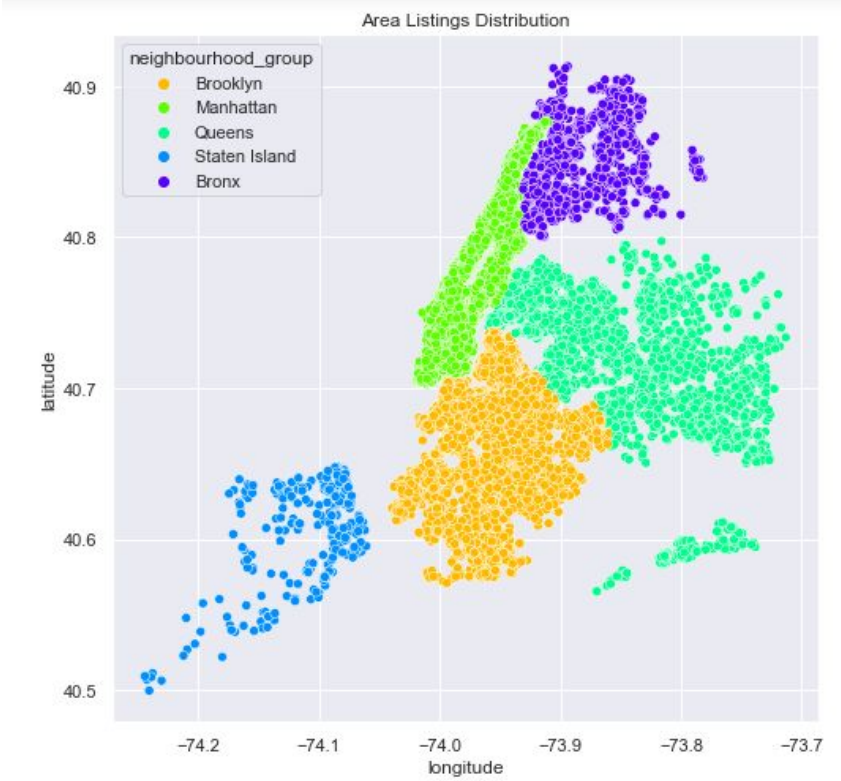


Neighbourhood Group

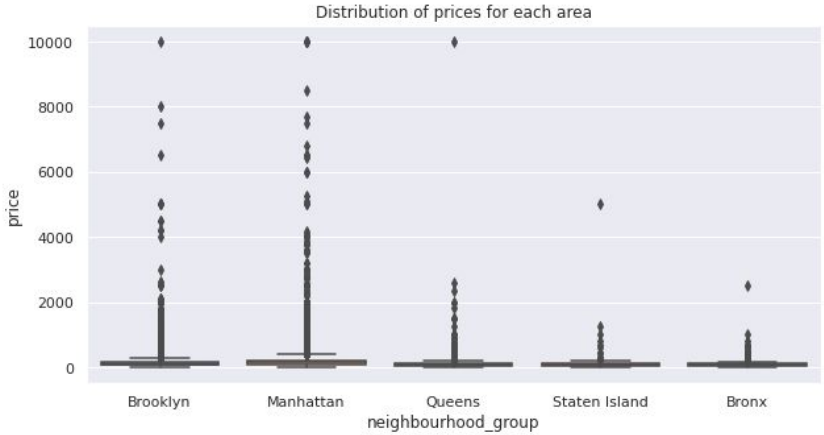
Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373

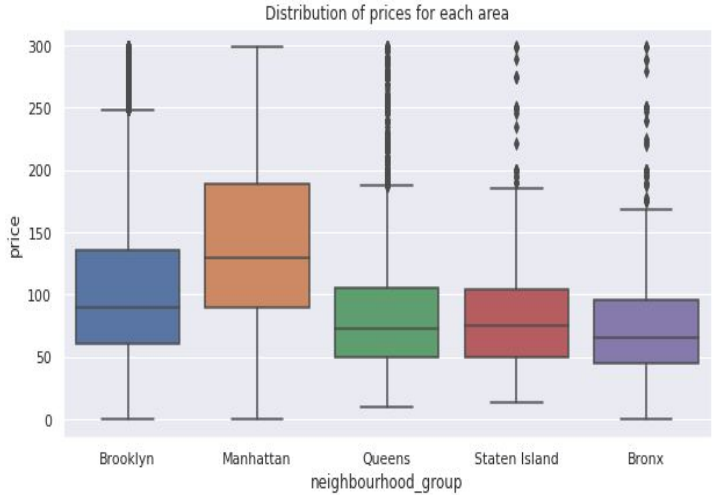


Neighbourhood Groups and Type of Rooms



REMOVING OUTLIERS





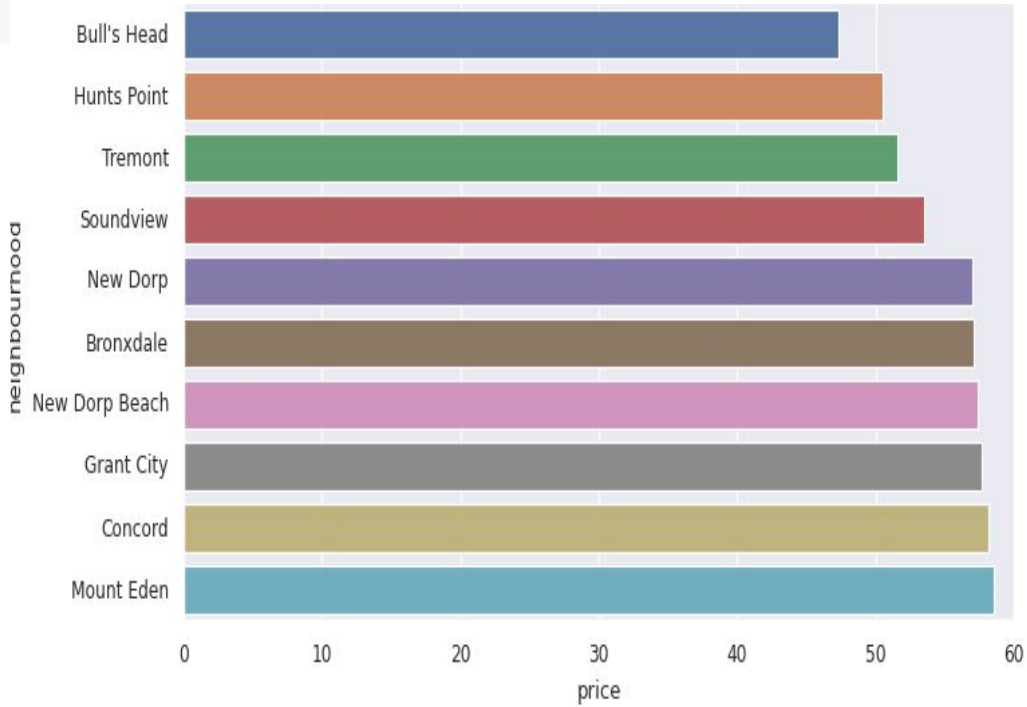
Above figures are visual representations of Price distribution among different areas before and after removing the outliers.
In this dataframe, the data with price range above 300 are outliers

Neighbourhood and Prices

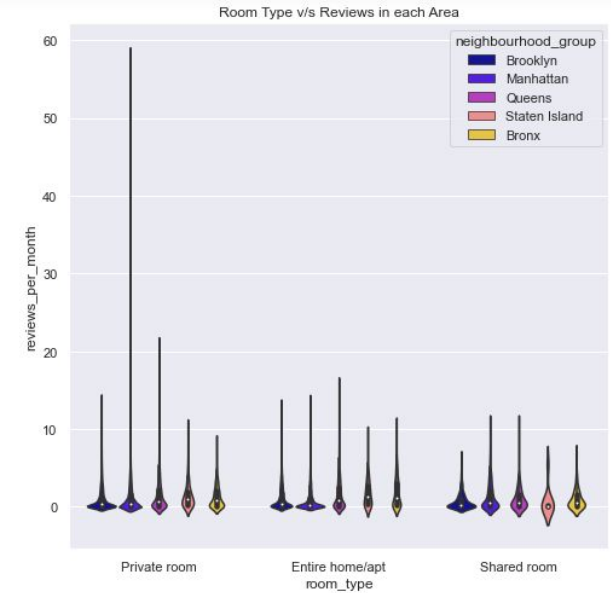
```
mean_prices_neighbourhood
```

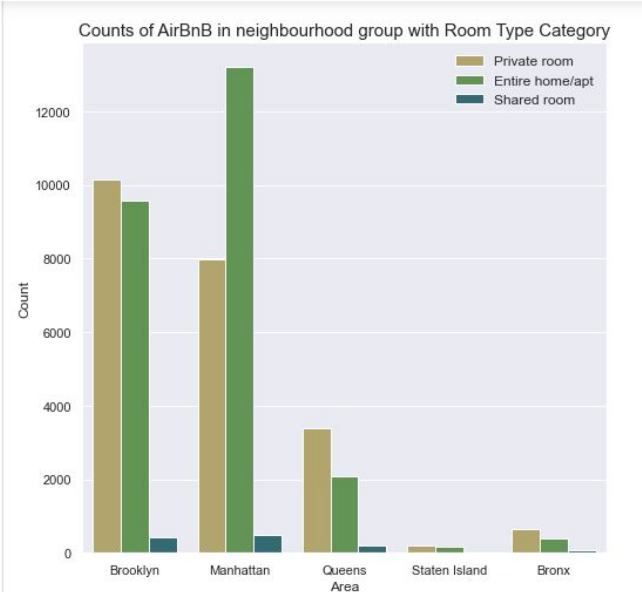
	neighbourhood	price
0	Bull's Head	47.333333
1	Hunts Point	50.500000
2	Tremont	51.545455
3	Soundview	53.466667
4	New Dorp	57.000000
...	...	...
216	Riverdale	442.090909
217	Sea Gate	487.857143
218	Tribeca	490.638418
219	Woodrow	700.000000
220	Fort Wadsworth	800.000000

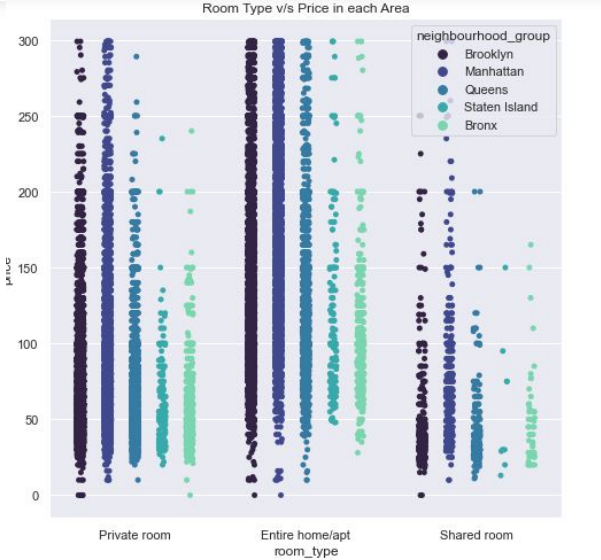
221 rows × 2 columns



Room Types

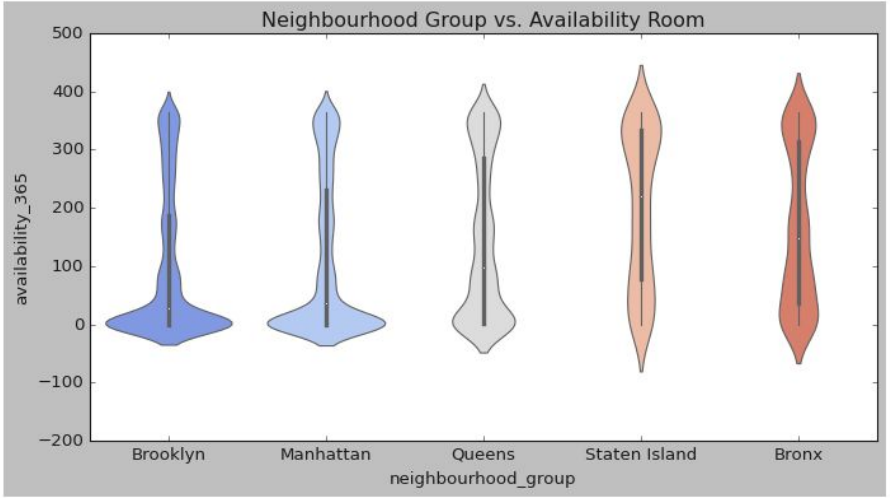


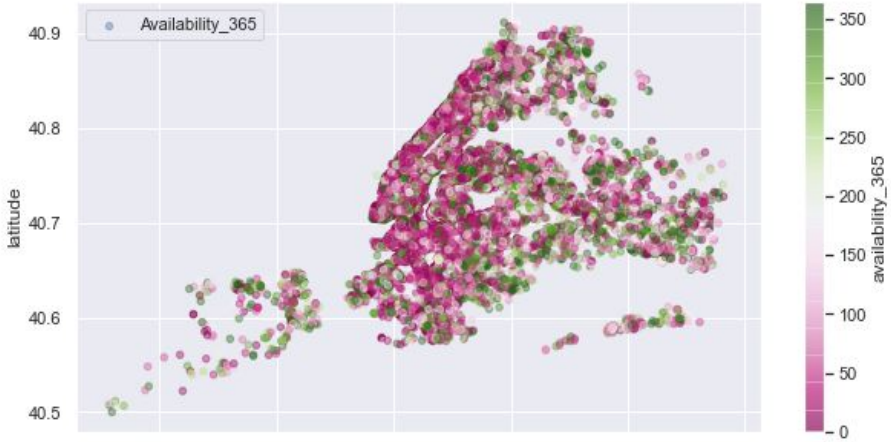




Irrespective of the area, 'entire home/apartment' stay is costlier than the other two. Most number of listings are for an entire apartment followed by private rooms.

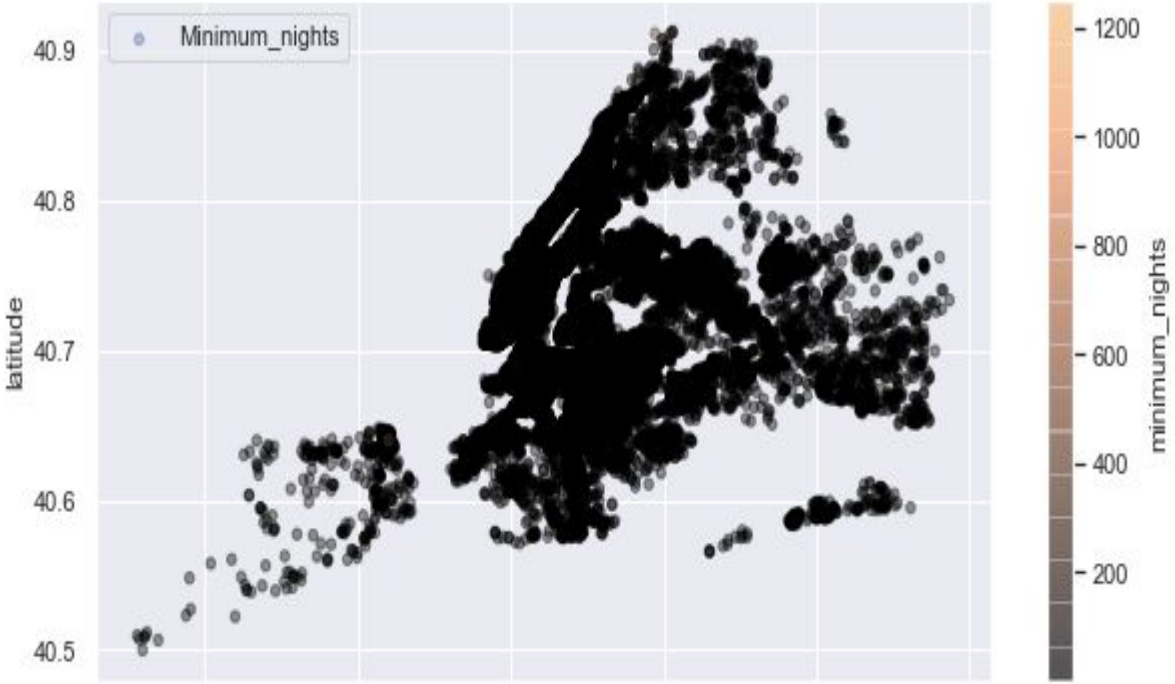
AVAILABILITY_365





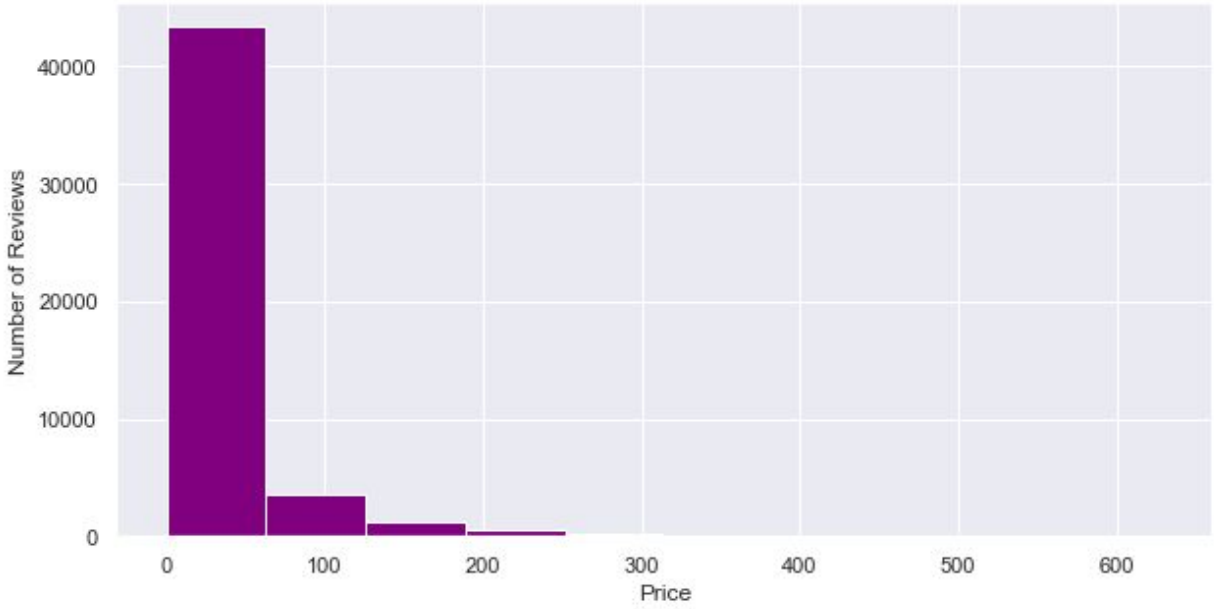
Staten Island has the least density when it comes to the availability of listings through the year. The fact that Staten Island has the least number of listings could also be contributing to this.

MINIMUM_NIGHTS



Almost 90 percentage of the listings have their minimum number of nights below 150. This could mean that the airbnb listings look more attractive to short term visitors of the place.

Reviews



The most number of reviews are for the listings with less price range from(0 to 50 $)

SUMMARY

- Manhattan and Brooklyn have the highest share of hotels.We deduced this using the pie charts and scatterplots.
- Manhattan is the busiest location in New York despite the relative higher cost of living. Availability is the main reason Manhattan is thriving in the airbnb business. Manhattan has the maximum number of listings and is mostly available throughout the year
- Manhattan has expensive and Bronx has low priced rooms. But we can see there are more price outliers. We have categorised the rooms into three categories based on price.0-100,100-300 and 300+ as cheap, medium and Expensive rooms. Price range of 0 to 300 is more appealing to the majority of the customers
- There is a notable difference in the number of people willing to stay in shared room with respect to a private room or an apartment.Most of the listings are for the stay in an entire apartment
- Listings in the less price range seem to have more reviews
- calculated_host_listings_count is a very helpful and accurate column of the given dataset.

THANK YOU