AIRBNB NYC Case Study – Methodology

upGrad & IIITB
DS-C68 Batch
Authored by: **Syed Mohammad Bilal**

1

Introduction

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb provides platform for hosts to accommodate guests with short-term lodging and tourism-related activities.

New York City is the most diverse and populated city in the United States. The city consists of 5 borrows: Manhattan, Brooklyn, Queens, the Bronx and Staten Island, all of which were "grouped" together into a single city. It is widely recognized as the global center for the financial services industry. It is also the heartbeat of the American media, entertainment (along with California), telecommunications, and law and advertising industries.

Business Objective:

For the past few months, Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

Assumption:

As we are not aware about the nature of reviews, we have assumed that the properties, which received higher number of reviews, have a better customer liking.

Data Source

Provided with Airbnb New York City Listings Dataset till 2019 (48895 Rows * 16 Columns)

Column	Description
Id	Listing ID
Name	Name of Listing
Host_id	host ID
Host_name	Name of Host
Neighbourhood	Neighbourhood_group - Location
Neighborhood	Neighborhood - Area
Latitude & Longitude	Map co-ordinates
Room_type	Listing space type
Price	Price of listing
Minimum_nights	Amount of nights minimum
Number_of_reviews	number of reviews
Last_review	Lastest review
Reviews_per_month	number of reviews per month
Calculated_host_listings_count	no. of listings per host
Availability_365	no. of days when listing is available for booking

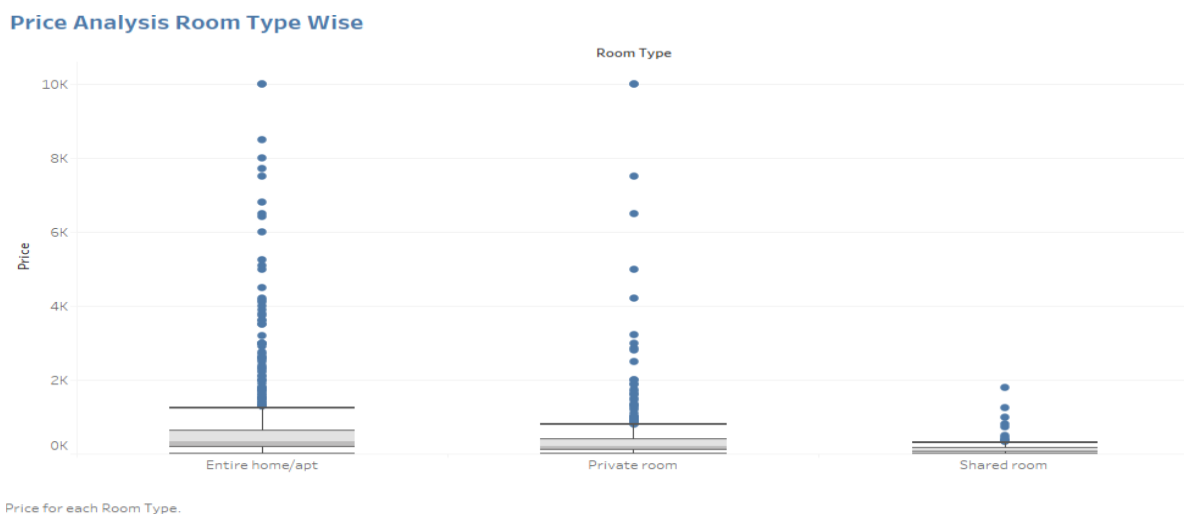
Column Name	No. of Rows and Column Datatype	
id	48895	int64
name	48879	object
host_id	48895	int64
host_name	48874	object
neighbourhood_group	48895	object
neighbourhood	48895	object
latitude	48895	float64
longitude	48895	float64
room_type	48895	object
price	48895	int64
minimum_nights	48895	int64
number_of_reviews	48895	int64
last_review	38843	object
reviews_per_month	38843	float64
calculated_host_listings_count	48895	int64
availability_365	48895	int64
dtypes: float64(3), int64(7), object(6)		

Data Wrangling:

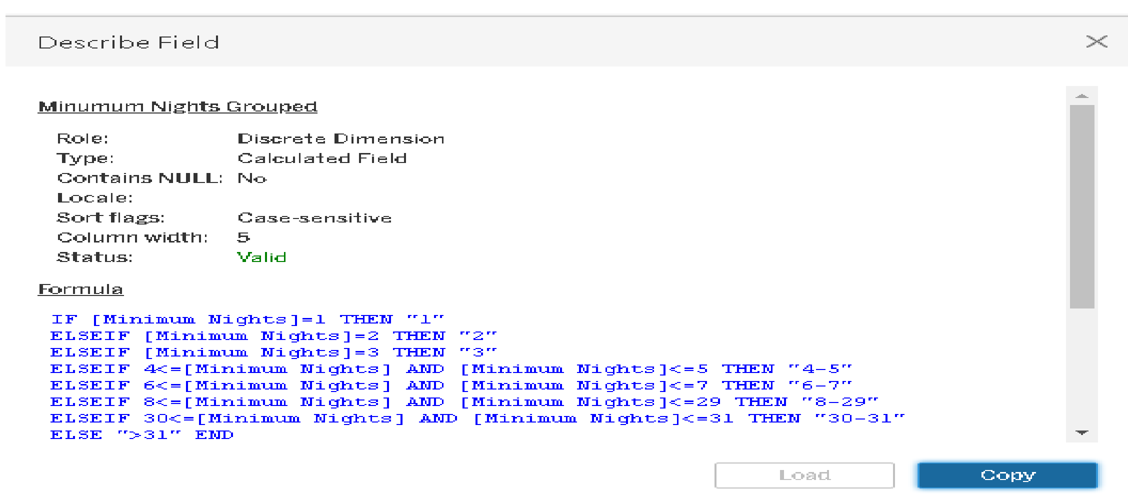
- Did univariate analysis using Tableau on the fields to see their distributions, the unique values in a field, the missing values and to check for outliers if any

- There was a small proportion of null values which would not affect my analysis so let them stay as it is

- Price was highly positively skewed so median was very close the lower quartile with some outliers as seen in the boxplot below



- Created a grouped field for Minimum Number of Days assuming null values belonged to the category.



Presentation – I:

Objective:

The presentation will focus mainly on the following points:

1. Get a better understanding about Airbnb listings with respect to various parameters
2. Understand the customer preferences
3. Understand the customer booking trend

Exploratory Data Analysis:

To understand some important insights we have explored the following questions:

1. How are the Airbnb listings spread out in NYC?
2. What type of rooms do customers prefer?
3. What could be the ideal number of minimum nights to increase customer bookings?

Based on customer review:

1. Most preferred neighborhood
2. Most preferred room type
3. Who are the Hosts who have the highest listings w.r.t Neighborhood?

Methodology

-> The data was analyzed through univariate and bivariate analysis.

-> The analysis and visualizations were done using Tableau considering various parameters.

-> The main parameters that have been taken into account for analysis are –

1. Geography based bookings
2. Bookings based on room type
3. Number of reviews
4. Minimum number of nights

-> Inferences have been made keeping in mind the above parameters

Explanation for EDA:

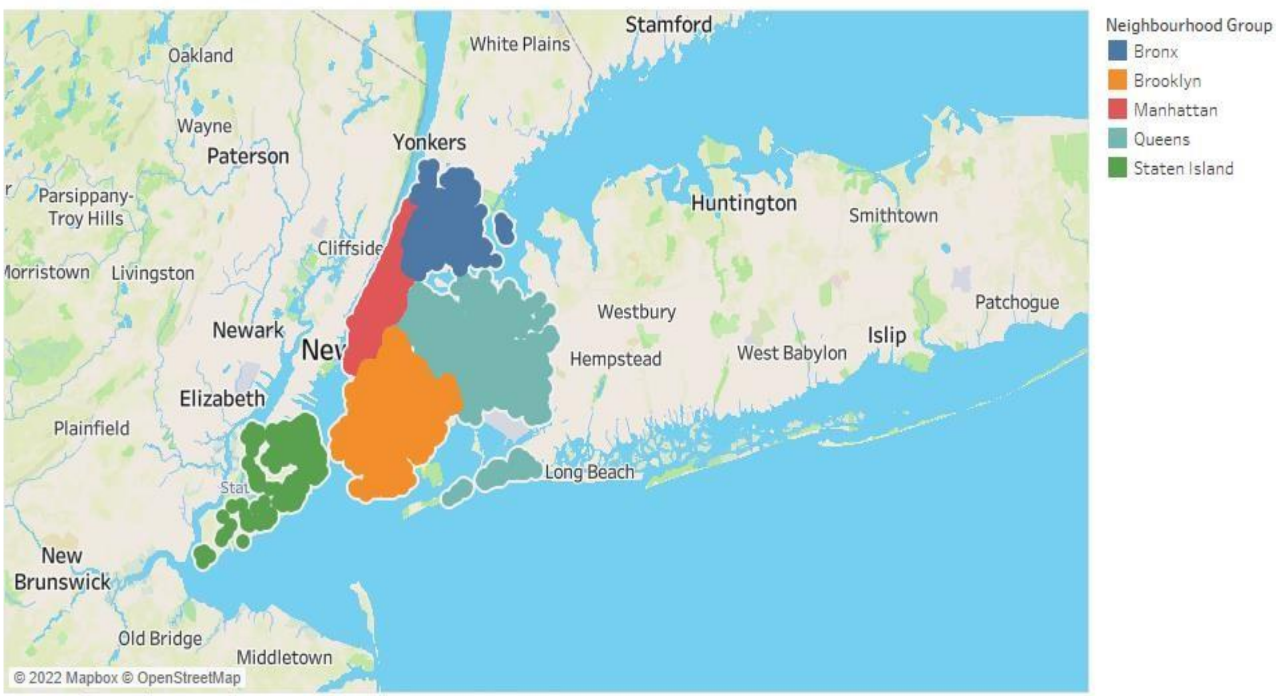
<u>How are the Airbnb listings spread out in NYC?</u>

We wanted to understand the spread of listings in the NYC areas and the concentration of listings in each neighborhood group. Two plots were used to explore this question:

- **Geographical plot**: This was created using the parameters latitude, longitude, neighborhoods, and neighborhood group. This gave us an understanding on what area we were dealing with.

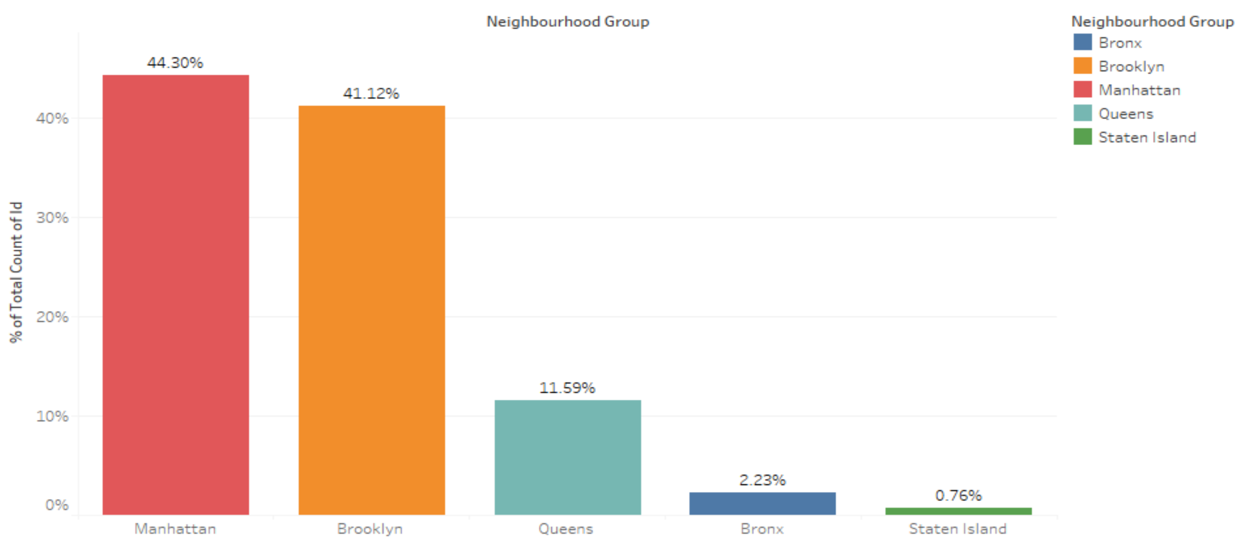
- **Bar plot**: This was used to understand the concentration of the listings in each neighborhood. We use the parameters Neighborhood group & CNT(Id).

Areas where AirBnb is present in NYC



Map based on Longitude and Latitude. Colour shows details about Neighbourhood Group.

AirBnb Listing distribution in NYC



% of Total Count of Id for each Neighbourhood Group. Colour shows details about Neighbourhood Group. The marks are labelled by % of Total Count of Id.

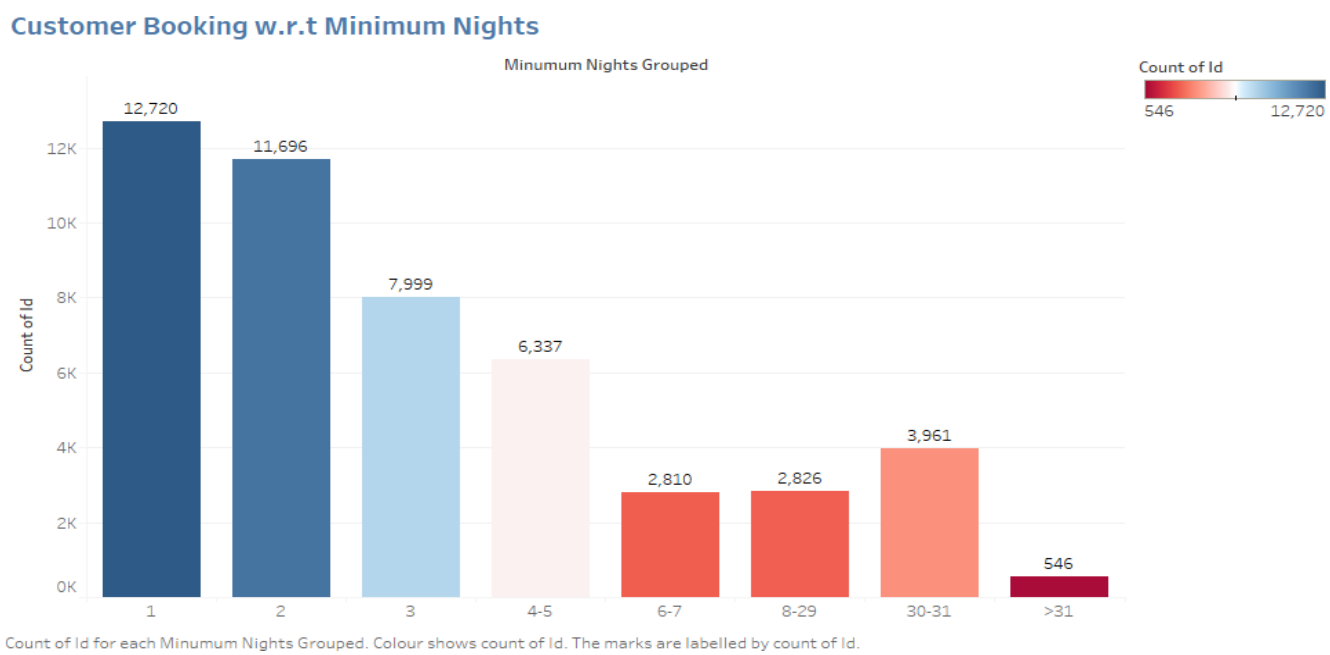
- **Inference:**

1. We see that, Airbnb has good presence in Manhattan, Brooklyn & Queens.
2. Listings are maximum in Manhattan (44%) & Brooklyn (41%) owing to the high population density and it being the financial and tourism hub of NYC. Staten Island (~1%) has the least number of listings, due to its low population density and very few tourism destinations.

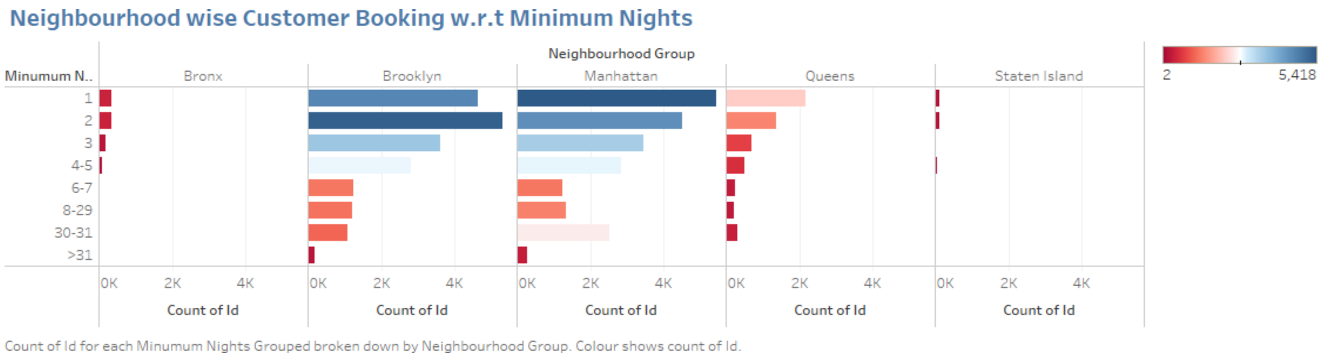
What type of rooms do customers prefer?

This question was addressed to understand the space needs of the customer and their preference. This has been explored using two pie charts.

- The first chart showed the overall preference of the customer across NYC.
- The second chart broke down the customer preference according to the neighborhood group.



- **Inference:**


Neighbourhood wise Customer Booking w.r.t Minimum Nights
Neighbourhood Group
Minumum N..
Bronx
Brooklyn
Manhattan
Queens
Staten Island
1
2
3
4-5
6-7
8-29
30-31
>31
0K
2K
4K
Count of Id
2
5,418
Count of Id for each Minumum Nights Grouped broken down by Neighbourhood Group. Colour shows count of Id.

- **Inference:**

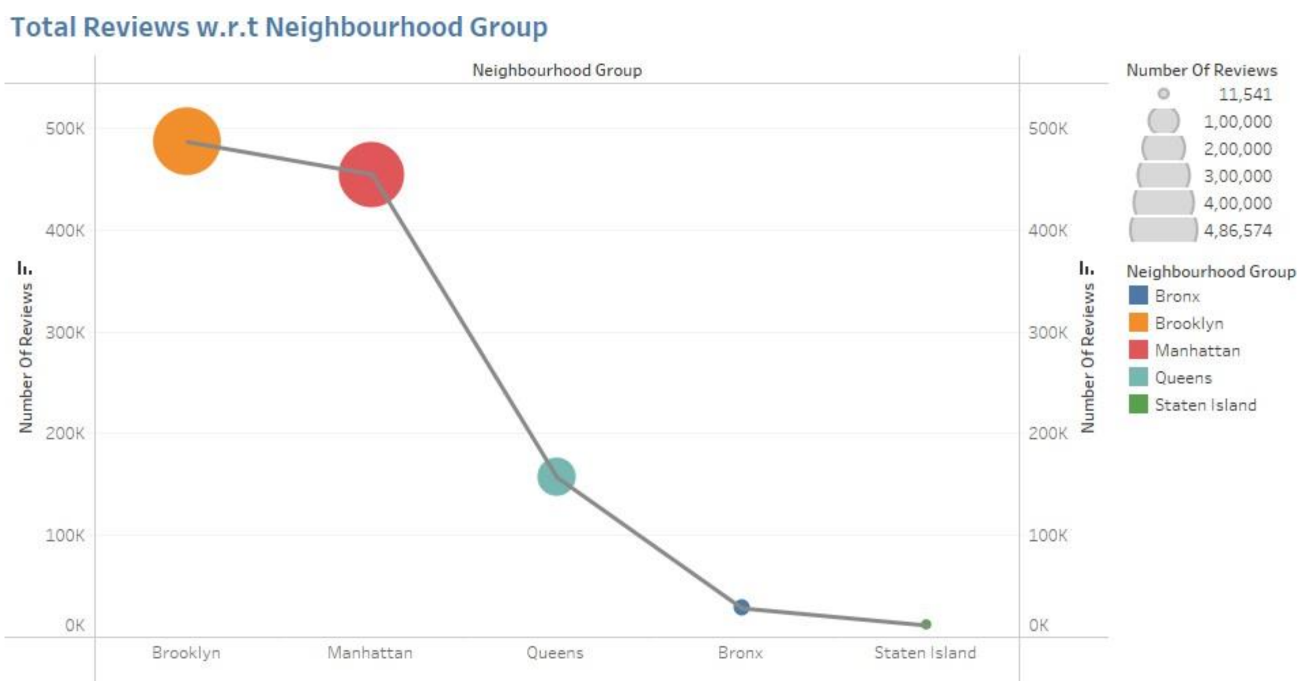
1. The listings with Minimum nights 1-6 have the most number of bookings.
2. We can see a prominent spike in 30 days; this would be because customers would rent out on a monthly basis. After 30 days, we can also see small spikes at 60 & 90 days, this can be explained by the monthly rent-taking trend.
3. Manhattan & Brooklyn have higher number of 30-day bookings compared to the others. The reason could be either tourists booking long stays or mid-level employees who opt for budget bookings due company visits.

Based on customer review:

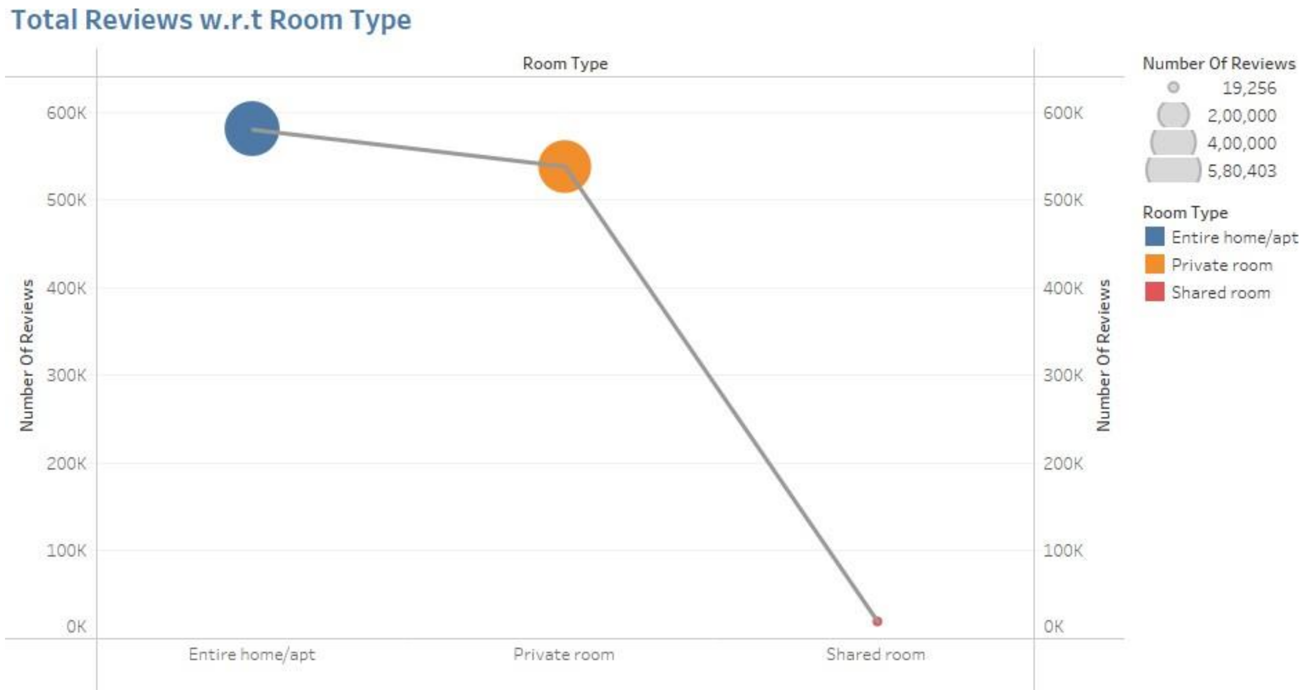
Most preferred neighborhood & Most preferred room type. The customer review parameter was chosen, as it is one of the most important factors to boost future bookings and listings. Here again, two different parameters that were taken for comparison: neighborhood & room type.

We had earlier explore the same parameters with reference to volume of bookings under each heading. Here we analyze it with the number of reviews obtained. The number of reviews a customer gives for a particular listing directly implies the likability of the listing. Using this we would like to see if the findings match with our earlier observation.

The parameters taken for analysis are: Room type; Neighborhood group, SUM(Number of reviews).



The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Neighbourhood Group. For pane Sum of Number Of Reviews: Colour shows details about Neighbourhood Group. Size shows sum of Number Of Reviews.



The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Room Type. For pane Sum of Number Of Reviews: Colour shows details about Room Type. Size shows sum of Number Of Reviews.

- **<u>Inference:</u>**

1. In line with our earlier observation, we see the maximum reviews in listings for Manhattan & Brooklyn, implying that more bookings happen in these neighborhoods. The higher number of customer reviews imply higher satisfaction in these localities.

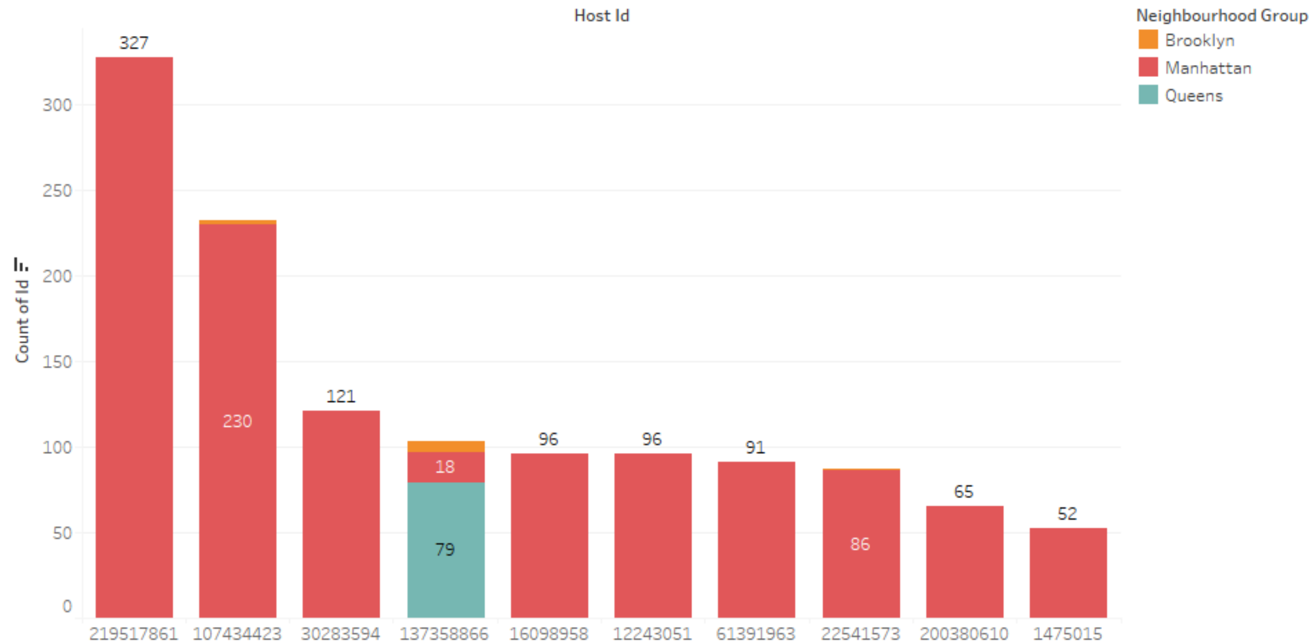
2. Also, we see the maximum reviews in room types ‘Entire home/apt’ & ‘Private rooms’. We can safely infer that, customers do not prefer ‘Shared rooms’.

<u>Who are the Hosts who have the highest listings w.r.t Neighborhood?</u>

This was explored to get an idea on the maximum listings held by a single host and in which area. This would give us an idea on how the hosts are investing and expanding in an area.

We have taken the Host ID in the x-axis with the CNT(Id) in the y-axis to understand the volume of booking. As there were huge number of Host ID, we have filtered it down to the top 10. The graph was color-coded based on neighborhood group.

Hosts having Highest Listing wrt Neighbourhood



Count of Id for each Host Id. Colour shows details about Neighbourhood Group. The marks are labelled by count of Id.The view is filtered on Host Id, which has multiple members selected.

- **Inference:**

1. More experienced hosts know the market better.
2. We observe a single host having multiple listings mainly in the Manhattan area. This is because Manhattan has the highest influx of tourists and financial enthusiasts visiting the city all year round.
3. This makes it more profitable for the host to acquire properties in the same area.

Presentation – II:

Objective:

The presentation will focus mainly on the following points:

1. Get a better understanding about Airbnb listings with respect to various parameters
2. Understand the pricing relation to various parameters
3. Recommendations to improve quality of new acquisitions and customer experience.

Exploratory Data Analysis:

To understand some important insights we have explored the following questions:

1. Customer preference for neighborhood & room type
2. Property demand based on minimum nights offered
3. Price range preferred by customers
4. Understanding Price variation w.r.t Room Type & Neighborhood
5. Understanding Price variation w.r.t Geography
6. Top reviewed properties

Methodology

-> The analysis and visualizations were done using Tableau considering various parameters.

-> The analysis was done keeping in mind the business side of the project.

-> The first half of the presentation focused on customer preference. The second half compared various parameters of customer preference with respect to price.

-> The following parameters were considered –

a. Customer experience: Neighborhood, Room type & minimum nights offered

b. Price variation: Volume of customer booking, Room type, Neighborhood, Number of reviews & Geography.

-> The first half of the presentation focused on customer preference.

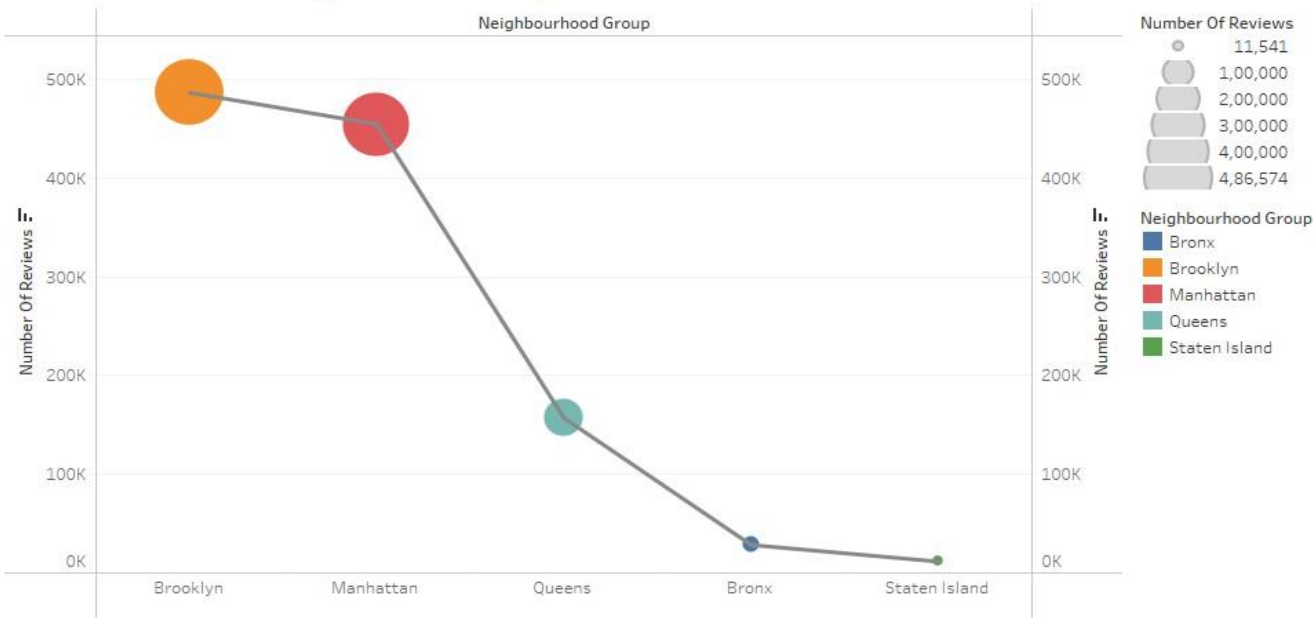
Explanation for EDA:

1. **<u>Customer preference for neighborhood & room type</u>**

We have explore the customer preference w.r.t volume and experience. The customer review parameter was chosen, as it is one of the most important factors to boost future bookings and listings. The number of reviews a customer gives for a particular listing directly implies the likability of the listing. The two different parameters were taken for comparison: neighborhood & room type.

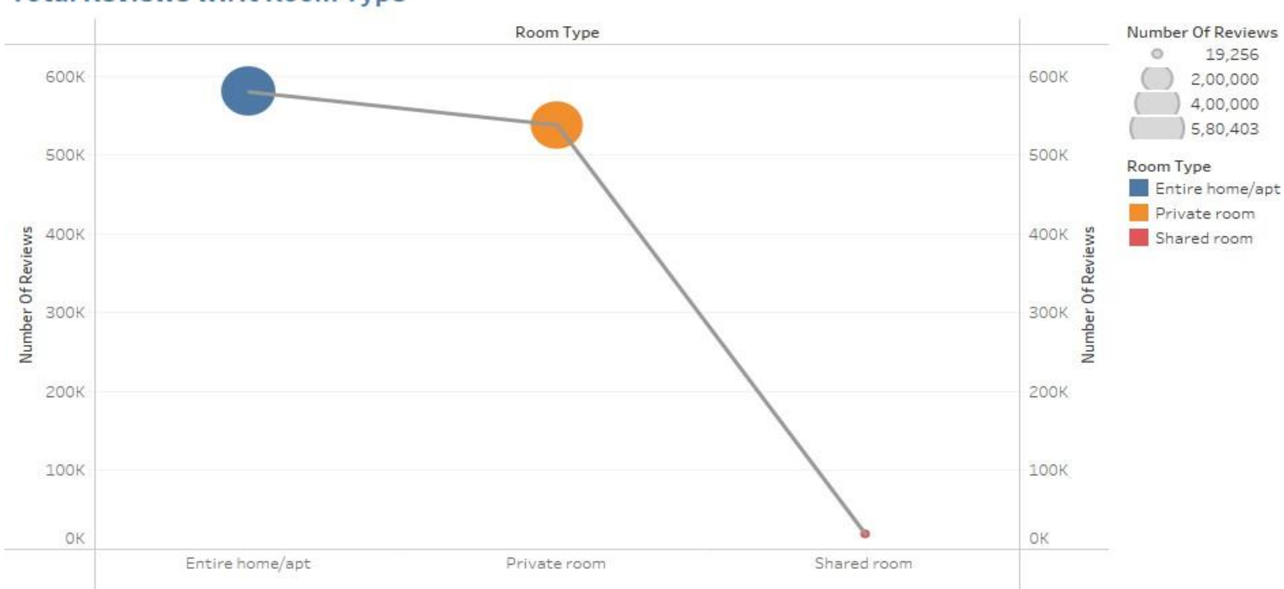
The parameters taken for analysis are: Room type; Neighborhood group, SUM (Number of reviews

Total Reviews w.r.t Neighbourhood Group

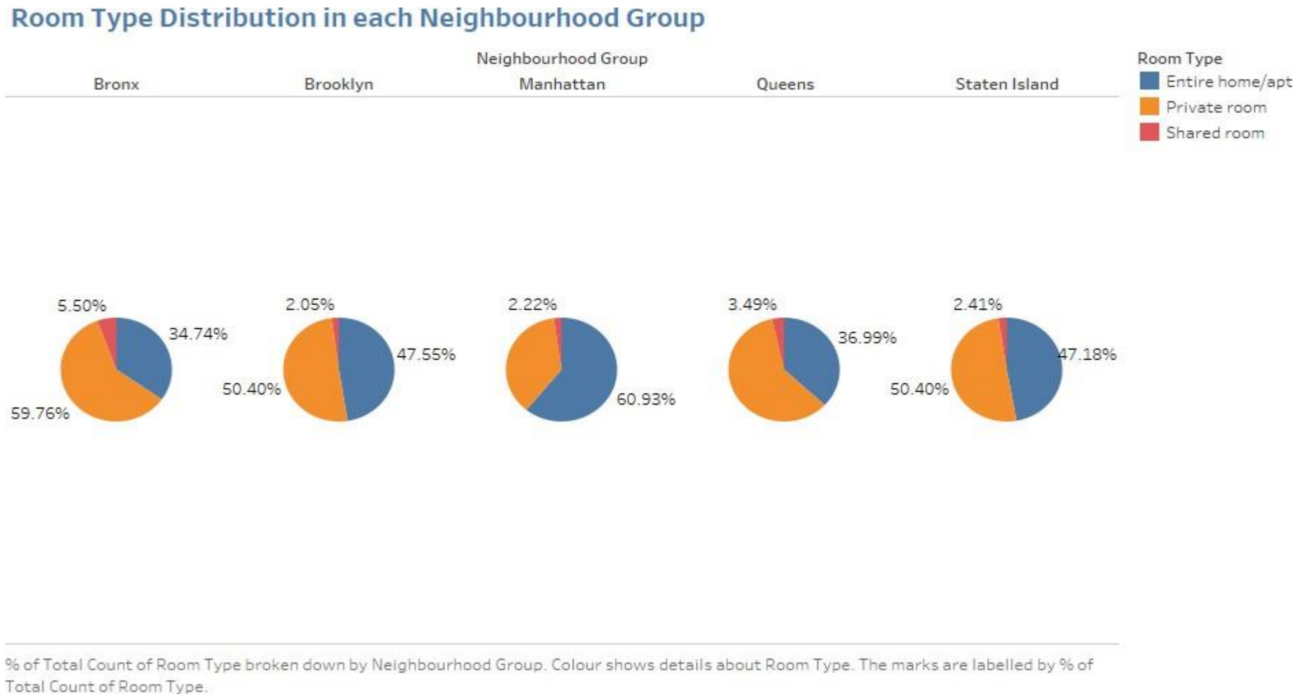


The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Neighbourhood Group. For pane Sum of Number Of Reviews: Colour shows details about Neighbourhood Group. Size shows sum of Number Of Reviews.

Total Reviews w.r.t Room Type



The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Room Type. For pane Sum of Number Of Reviews: Colour shows details about Room Type. Size shows sum of Number Of Reviews.



- **<u>Inference:</u>**

1. There are three types of rooms - Entire home/Apartment, Private room & shared room. Customers prefer private rooms or entire homes in comparison to shared rooms.

2. In addition, we can see maximum reviews in listings for Manhattan & Brooklyn, implying that more bookings happen in these neighborhoods. (The higher number of customer reviews imply higher satisfaction).

- **<u>Recommendation:</u>**

1. Airbnb can concentrate on promoting shared rooms with targeted discounts to increase bookings.

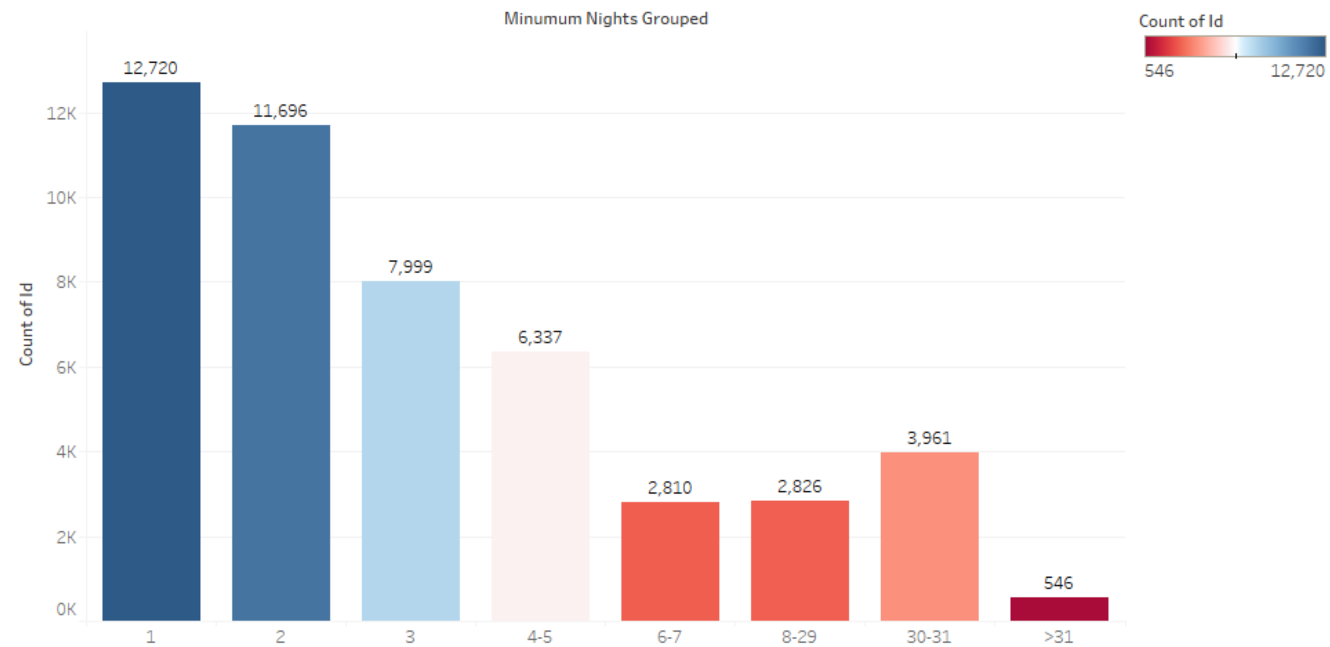
2. New acquisitions can be explored to acquire 'private rooms' in Manhattan and Brooklyn and 'entire homes' in Bronx and Queens.

2. Property demand based on minimum nights offered

We wanted to observe the customer booking pattern and demand of property based on the minimum number of stay nights. This was chosen to understand for what type of stay customers use Airbnb; short-stay or long-stay. Here, we took into account the volume of booking and the neighborhood- wise volume of booking.

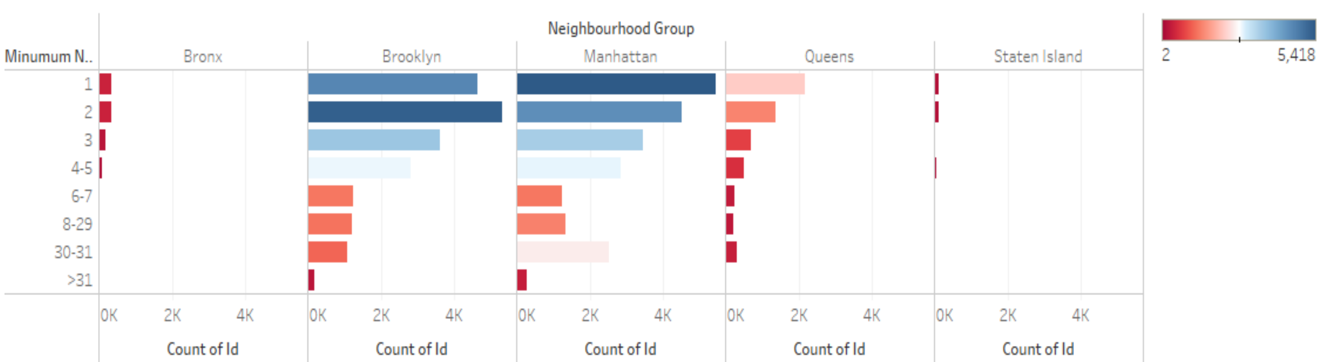
The parameters taken into account were: CNT(Id), Minimum Nights (This was binned, with a bin size of 2 for easier visualization) & Neighborhood Group.

Customer Booking w.r.t Minimum Nights



Count of Id for each Minumum Nights Grouped. Colour shows count of Id. The marks are labelled by count of Id.

Neighbourhood wise Customer Booking w.r.t Minimum Nights



Count of Id for each Minumum Nights Grouped broken down by Neighbourhood Group. Colour shows count of Id.

- **Inference:**

1. The listings with Minimum nights 1-6 have the most number of bookings. We can see a prominent spike in 30 days. This would be because customers would prefer renting out on a monthly basis. After 30 days, we can also see small spikes at 60 & 90 days, this can be explained by the monthly rent-taking trend.

- **Recommendation:**

1. More number of hosts & listings with monthly rental duration (30-60-90) can be acquired. We see a good potential in the 30-day rental window. Manhattan & Brooklyn have higher number of 30 day bookings compared to the others, these areas can be further targeted.

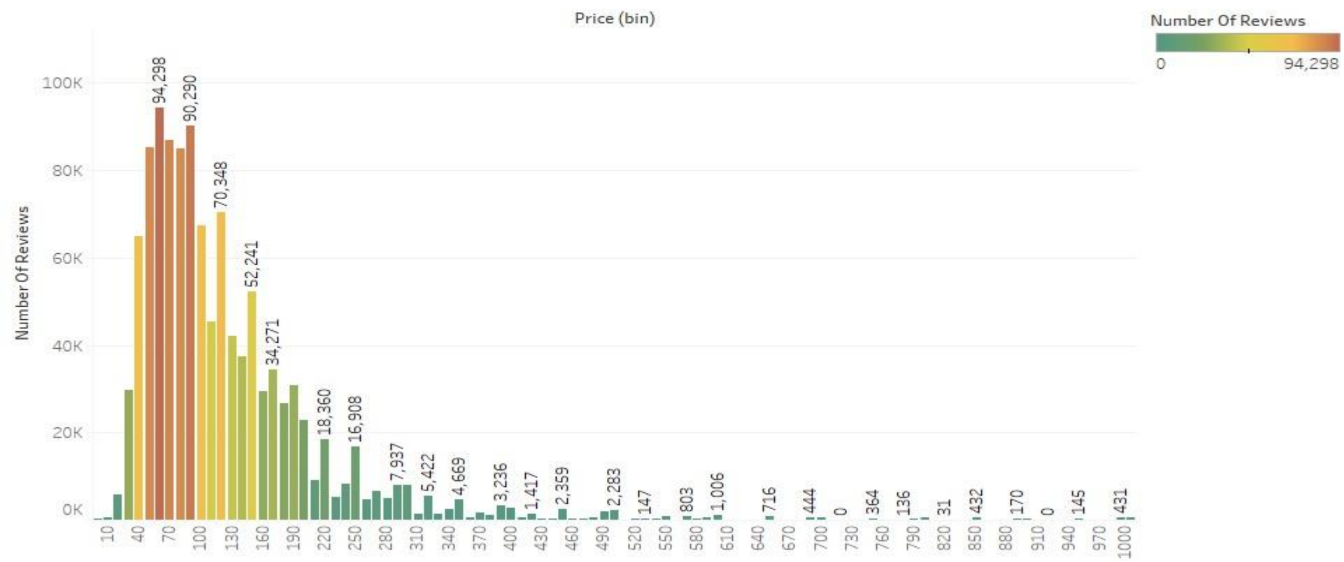
2. Also, weekly or bi-weekly rentals can also be acquired as these can be used customers stranded in NYC for quarantine purposes.

3. Price range preferred by customers

For any business to operate it has to have a fair understanding of the customer-buying pattern. So we have tried to understand the most preferred price range for customers. Using this we can try to improve the listings in the price range preferred by the customer.

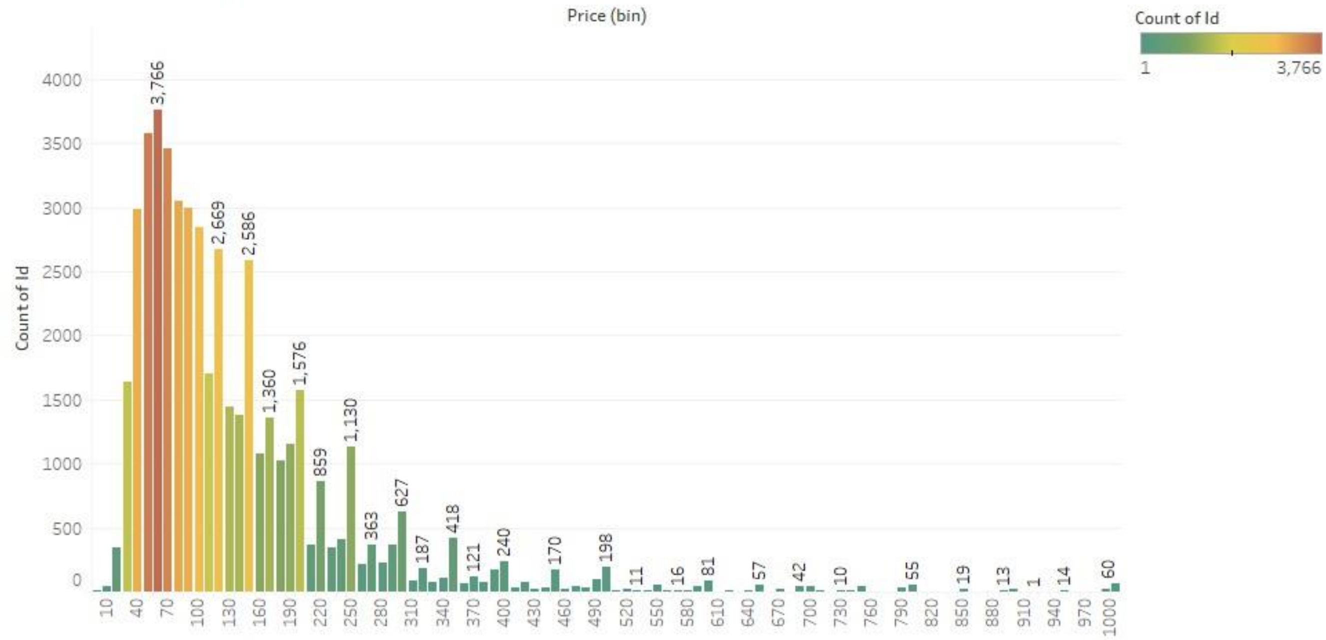
We have considered the volume of booking and number of reviews in a particular price range. For easy visualization, we have binned the Price with a bin size of 10. Also owing to the enormous value range, we have observed the variation until $1000. As there was very little data beyond this, we decided to filter it.

Price Variation wrt Reviews



Sum of Number Of Reviews for each Price (bin). Colour shows sum of Number Of Reviews. The marks are labelled by sum of Number Of Reviews.The view is filtered on Price (bin), which has multiple members selected.

Preferred Price By customers



Count of Id for each Price (bin). Colour shows count of Id. The marks are labelled by count of Id. The view is filtered on Price (bin), which has multiple members selected.

- **<u>Inference:</u>**

1. We have taken pricing preference based on two parameters – volume of bookings done in a price range and number of reviews in a price range. From both the graphs, the favorable price range is $40 - $190. This is the price range most preferred by most customers.

- **<u>Recommendation:</u>**

New acquisitions and expansion can be done in the price range of $40 - $190 as it satisfies both parameters of volume of customer traffic and customer satisfaction.

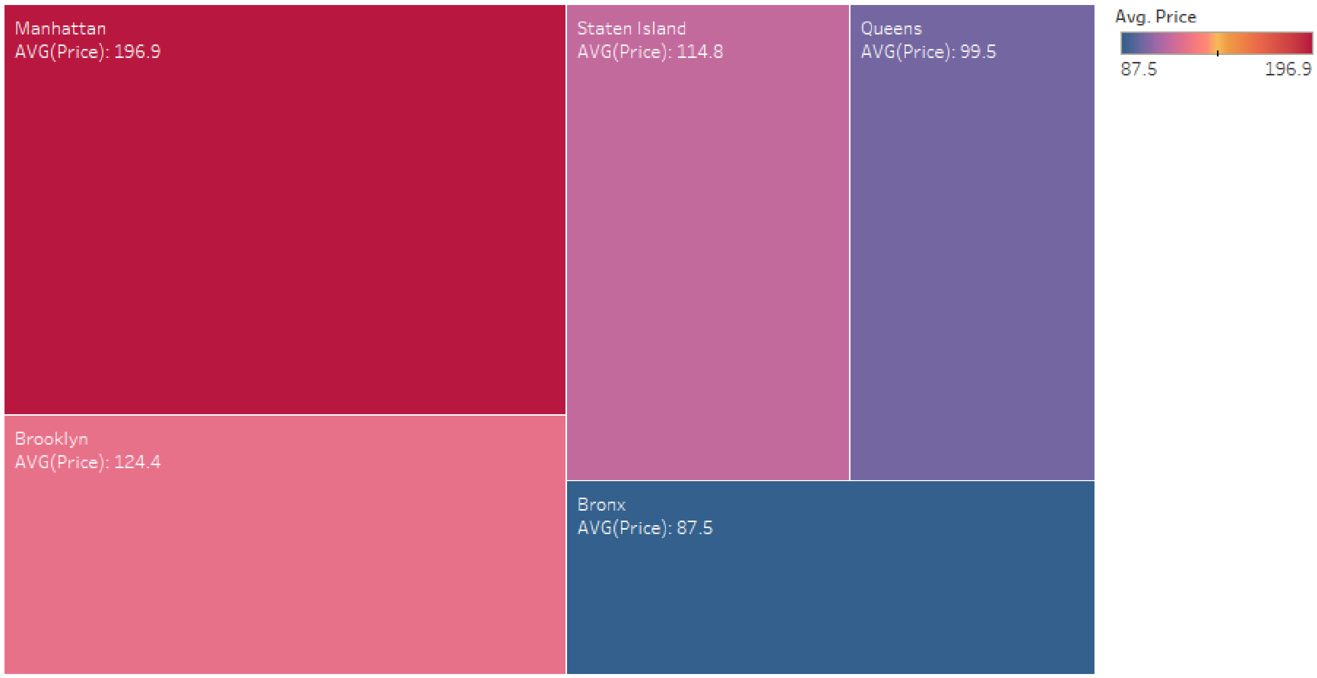
4. **Understanding Price variation w.r.t Room Type & Neighborhood**

Now that we have obtained the optimum price range for listings, let us explore which neighborhoods and room types fit in this category.

We have created two graphs to explore this question:

- **Tree map**: We wanted to understand the average price distribution in the 5 boroughs of NYC. The tree map was created with Avg(Price) for ‘size’ and ‘color’.
- **Highlight table**: As the comparison table containing the room type and neighborhood mainly consisted of numbers we decided to go ahead with highlight table to display the highest and lowest values.

Avg. Price in each Neighbourhood



Neighbourhood Groupandaverage of Price. Colour shows average of Price. Size shows average of Price. The marks are labelled by Neighbourhood Groupandaverage of Price.

Avg. Price in each Neighbourhood wrt Room Type

Neighbourhood Group	Room Type: Entire home/apt	Room Type: Private room	Room Type: Shared room
Bronx	127.5	66.8	59.8
Brooklyn	178.3	76.5	50.5
Manhattan	249.2	116.8	89.0
Queens	147.1	71.8	69.0
Staten Island	173.8	62.3	57.4

Avg. Price: 50.5 – 249.2

Average of Price broken down by Room Type vs. Neighbourhood Group. Colour shows average of Price. The marks are labelled by average of Price.

- **Inference:**

1. Manhattan appears to have the highest average price of $196.9. The 'Entire home/apt' room type in Manhattan is the most expensive at $250, much higher than the overall average.
2. 'Shared Room' type is the cheapest in Brooklyn.

- **Recommendation:**

1. In line with the earlier recommendation, we observe that ‘private rooms’ of Manhattan & Brooklyn and ‘entire homes’ in Bronx and Queens Fall in the favorable price range ($40-$190).
2. Brooklyn has an average price of $124. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.

5. **Understanding Price variation w.r.t Geography**

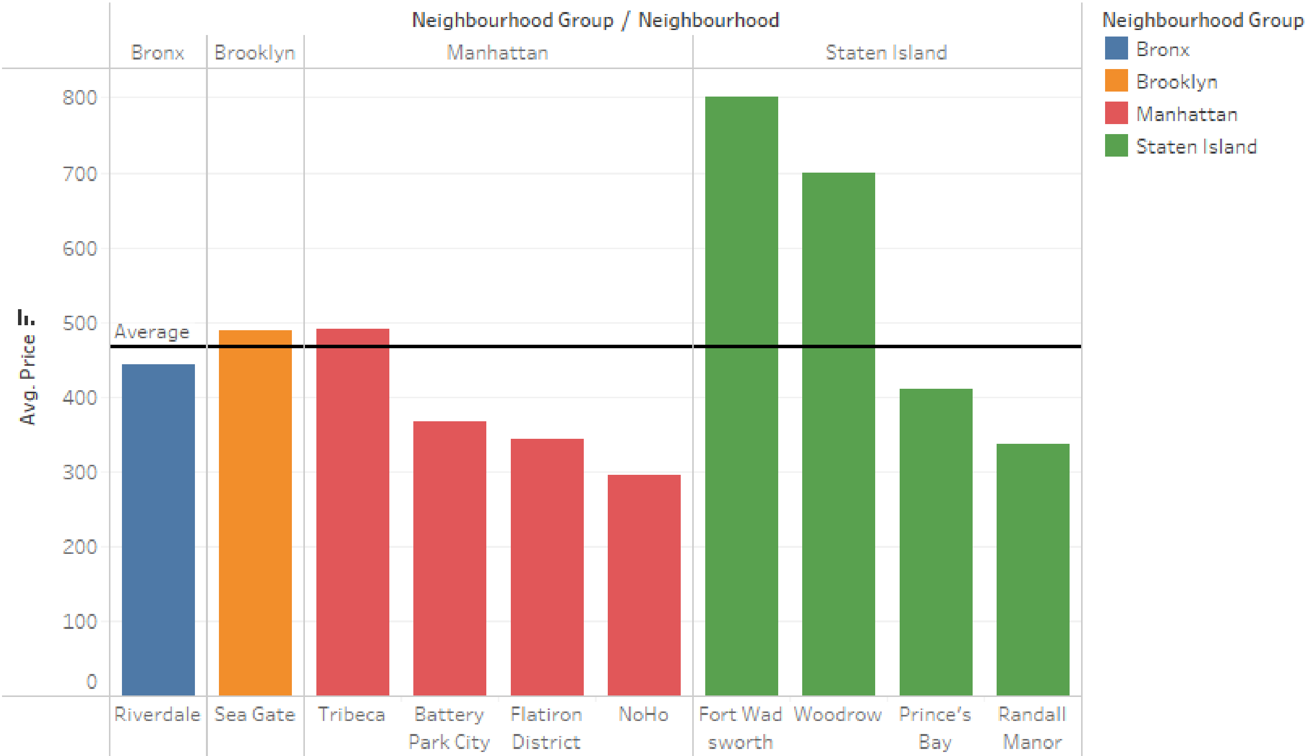
We had earlier explore the price variation with respect to location. We now deep dive to understand how it varies across difference areas/geographies.

- We wanted to understand if the geography played a part in rising prices. For this, we plotted a geographical map to understand the price density and variation
- To further correlate our finding, we took the top 10 neighborhood with maximum average price. We used the findings in this to confirm our observation obtained from the geographical map.

Price Density and Variation

Map based on Longitude and Latitude. Colour shows average of Price. Size shows average of Price. The marks are labelled by Neighbourhood Group.

Top 10 Neighbourhod Based on Avg. Price



Average of Price for each Neighbourhood broken down by Neighbourhood Group. Colour shows details about Neighbourhood Group.The view is filtered on Neighbourhood, which has multiple members selected.

- **Inference:**
  1. The map displays the price variation, which appears to be distributed uniformly in the inland areas. We see spike in prices in coastal cities, owing to better view from stays and easy ferry reachability. When we zoomed in, we also observed higher pricing near colleges or important monuments/landmarks.
  2. The bar graph confirms our inference, as we observe that the top 10 neighborhoods according to price are those that are situated near the sea or are next to important institutions/companies/landmarks.

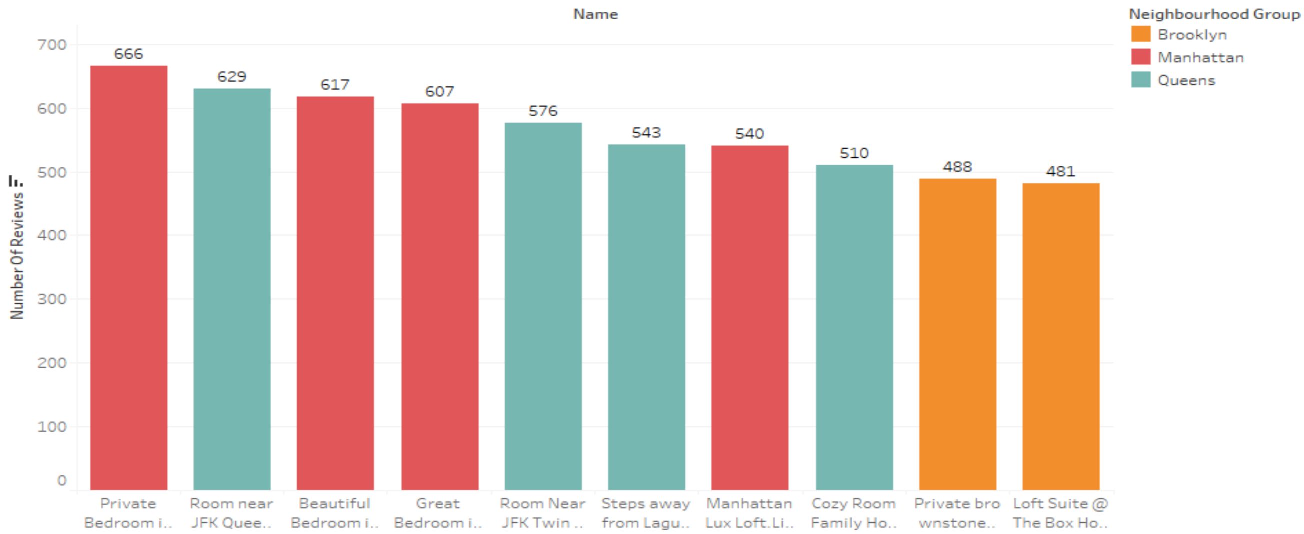
- **Recommendation:**

1. Increasing acquisitions and new properties in coastal regions can increase customer bookings.

6. Top reviewed properties

We have gotten various insights in the above questions regarding price range or neighborhood. To confirm and correlate our observations, we have visualized the Top 10 most reviewed properties. This would give us an overall idea of whether our analysis agrees with the customer preference. We have taken the “name” of the listings and calculated how many reviews each listing received.

Top 10 Property as per Reviews



Sum of Number Of Reviews for each Name. Colour shows details about Neighbourhood Group. The marks are labelled by sum of Number Of Reviews. The view is filtered on Name, which has multiple members selected.

- **Inference:**

  1. Manhattan, Brooklyn and Queens have the most liked properties (most reviewed).
  2. the most reviewed property "Private Bedroom in Manhattan", though it appears to be steeply priced still has managed to get the maximum number of reviews making it the most favorable property in NYC.

Recommendations Consolidated:

1. Promotion of shared rooms with targeted discounts to increase bookings.
2. More number of hosts & listings with monthly rental duration (30-60-90) can be acquired. We see a good potential in the 30-day rental window. Manhattan & Brooklyn have higher number of 30-day bookings compared to the others; these areas can be further targeted.
3. Weekly or bi-weekly rentals can also be acquired, as these can be used customers stranded in NYC for quarantine purposes.
4. New acquisitions and expansion can be done in the price range of $40 - $190 as it satisfies both parameters of volume of customer traffic and customer satisfaction.
5. New acquisitions can be explored to acquire 'private rooms' in Manhattan and Brooklyn and 'entire homes' in Bronx and Queens.
6. Brooklyn has an average price of $124. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.
7. Increasing acquisitions and new properties in coastal regions can increase customer bookings.