Airbnb NYC Case Study – Methodology

- **ABHIRAMI SUBRAMANIAN**

Introduction:

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb provides a platform for hosts to accommodate guests with short-term lodging and tourism-related activities.[1]

New York City is the most diverse and populated city in the United States. The city is made up of 5 boroughs: 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's also the heartbeat of the American media, entertainment (along with California), telecommunications, law and advertising industries. [2]

Problem statement:

- In January 2020, Airbnb saw 50.2 million website visits.
- But with the onset of the pandemic, Airbnb was one of the many businesses in the travel industry to be hit hard.
- During the COVID-19 pandemic, Airbnb bookings dropped as much as 96% in some cities.[1] Due to this, Airbnb has seen a major decline in revenue.

Data:

- The data contains the bookings made through Airbnb in NYC area.
- It contains the following columns:
  - Id – Listing ID
  - Name – Name of Listing
  - Host_id – host ID
  - Host_name – Name of Host
  - Neighbourhood_group - Location
  - Neighbourhood - 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

Presentation – 1

Objective:

Now that the restrictions have started lifting and people have started to travel more, we have to make sure that it is fully prepared for this change.

The presentation will focus mainly on the following points:

- Get a better understanding about airbnb listings with respect to various parameters
- Understand the customer preferences
- Understand the customer booking trend

Exploratory Data Analysis:

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

- How are the airbnb listings spread out in NYC?
- What type of rooms do customers prefer?
- What could be the ideal number of minimum nights to increase customer bookings?
- Based on customer review:
  - Most preferred neighbourhood
  - Most preferred room type
- Who are the Hosts who have the highest listings w.r.t Neighbourhood?

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 –
  - Geography based bookings
  - Bookings based on room type
  - Number of reviews
  - Minimum number of nights
- Inferences have been made keeping in mind the above parameters

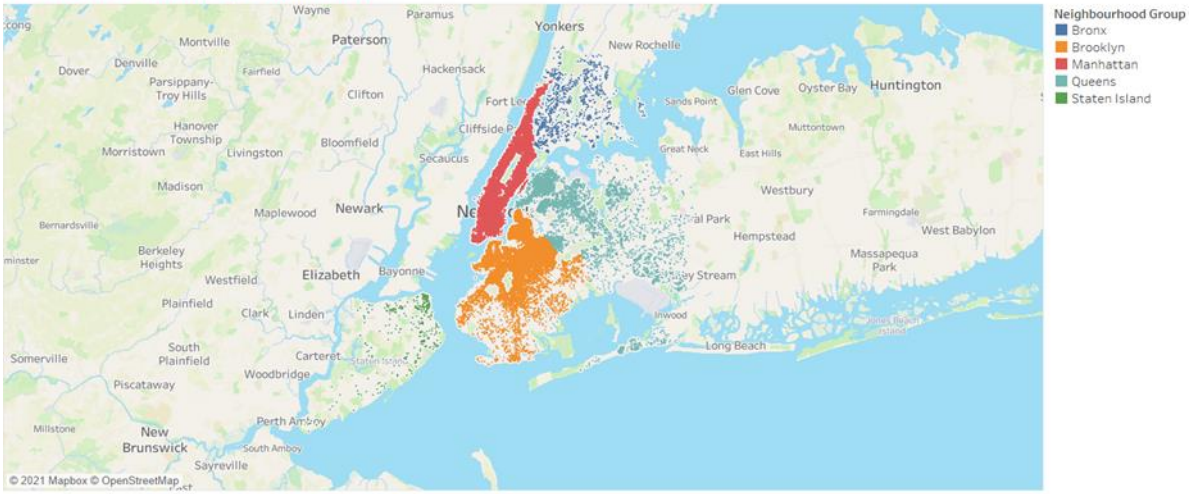
Explanation for EDA:

- How are the airbnb listings spread out in NYC?

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

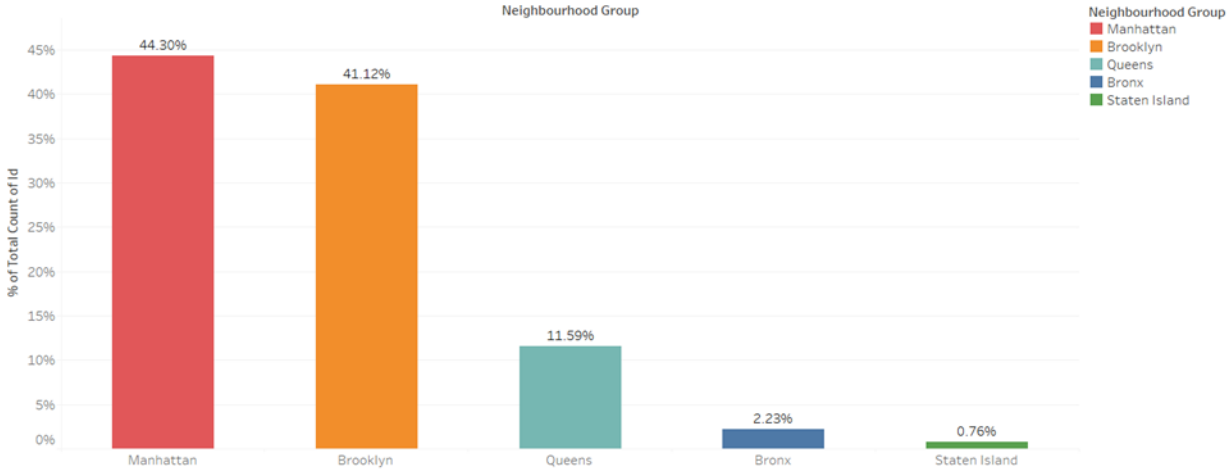
- Geographical plot: This was created using the parameters latitude, longitude, neighbourhood and neighbourhood group. This gave us an understanding on what area we were dealing with.
- Bar plot: This was used to understand the concentration of the listings in each neighbourhood. We use the parameters Neighbourhood group & CNT(Id).

Areas where AirBnB is present in NYC



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

Airbnb Listings concentration in NYC



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

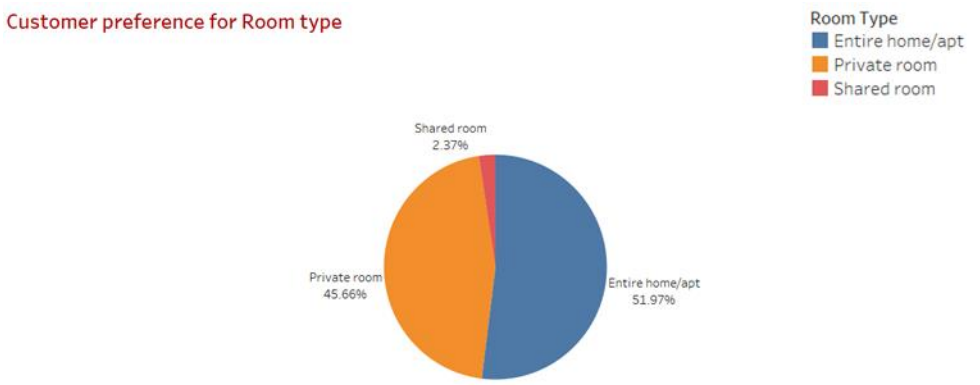
Inference:

- We see that, airbnb has good presence in Manhattan, Brooklyn & Queens.
- 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.

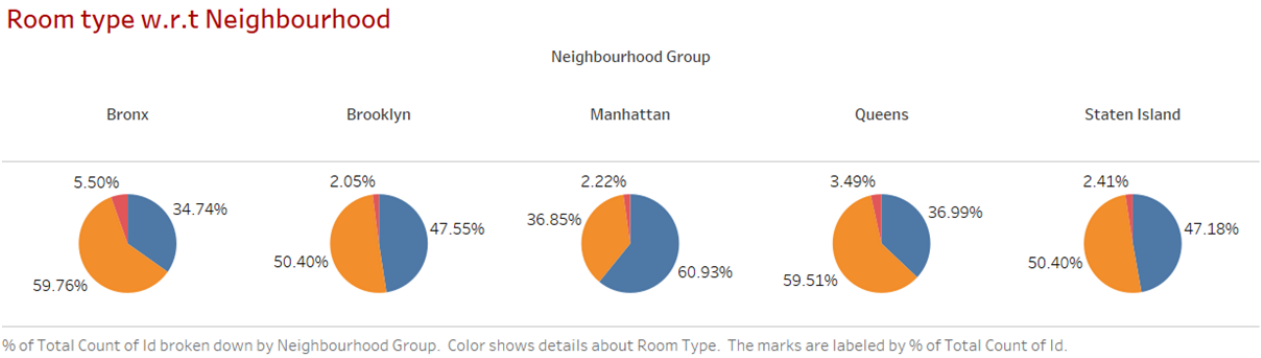
➢ What type of rooms do customers prefer?

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

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



Room Type and % of Total Count of Host Id. Color shows details about Room Type. Size shows count of Host Id. The marks are labeled by Room Type and % of Total Count of Host Id.



% of Total Count of Id broken down by Neighbourhood Group. Color shows details about Room Type. The marks are labeled by % of Total Count of Id.

Inference:

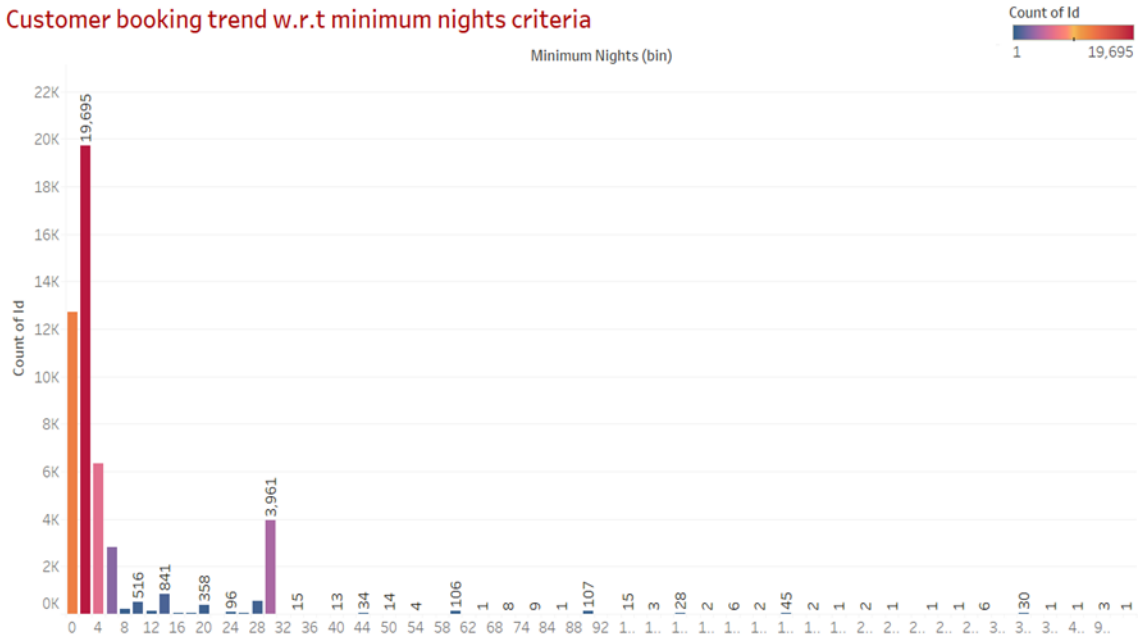
- There are three types of rooms - Entire home/Apartment, Private room & shared room.
- Overall, customers appear to prefer private rooms (45%) or entire homes (52%) in comparison to shared rooms (2.4%). Airbnb can concentrate on promoting shared rooms with discounts to increase bookings and also acquire more private listings.
- Queens & Bronx contribute 60% each to private rooms, more than the combined ratio of 45%. Whereas, Manhattan has a higher contribution in entire home (61%), compared to the combined ratio of 52%.

➢ What could be the ideal number of minimum nights to increase customer bookings?

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

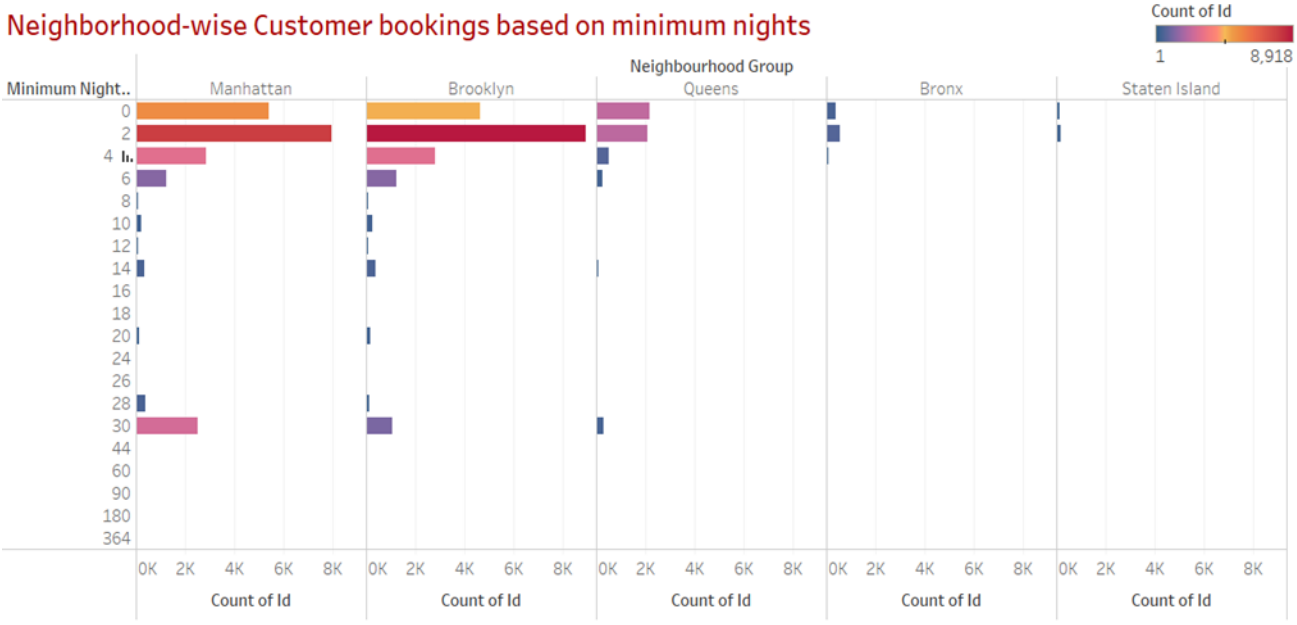
The parameters taken into account were: CNT(Id), Minimum Nights (This was binned, with a bin size of 2 for easier visualisation) & Neighbourhood Group.

Customer booking trend w.r.t minimum nights criteria



Count of Id for each Minimum Nights (bin). Color shows count of Id. The marks are labeled by count of Id.

Neighborhood-wise Customer bookings based on minimum nights



Count of Id for each Minimum Nights (bin) broken down by Neighbourhood Group. Color shows count of Id. Details are shown for Neighbourhood Group. The view is filtered on Minimum Nights (bin), which keeps 20 of 81 members.

Inference:

- **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 rent out on a monthly basis. After 30 days, we can also see small spikes at 60 & 90 days, this can also be explained by the monthly rent taking trend.
- 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.

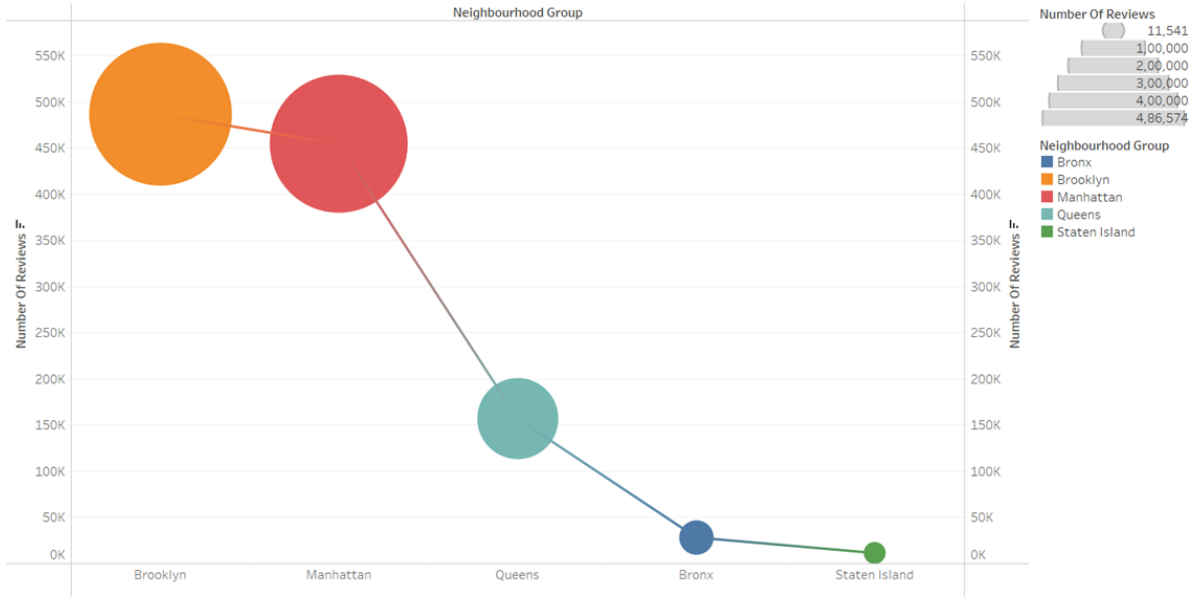
- Based on customer review: Most preferred neighbourhood & Most preferred room type

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 were taken for comparison: neighbourhood & room type.

We had earlier explore the same parameters with reference to volume of bookings under each heading. Here we analyse 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.

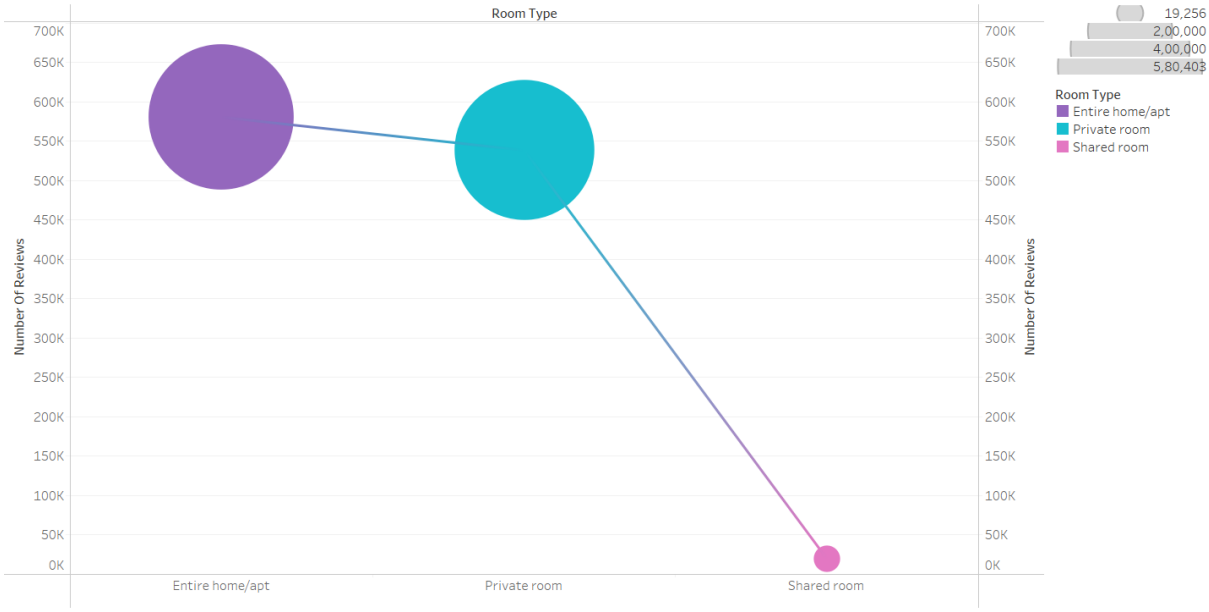
The parameters taken for analysis are: Room type; Neighbourhood group, SUM(Number of reviews)

No. of reviews according to Neighbourhood



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

No. of reviews according to Room type



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

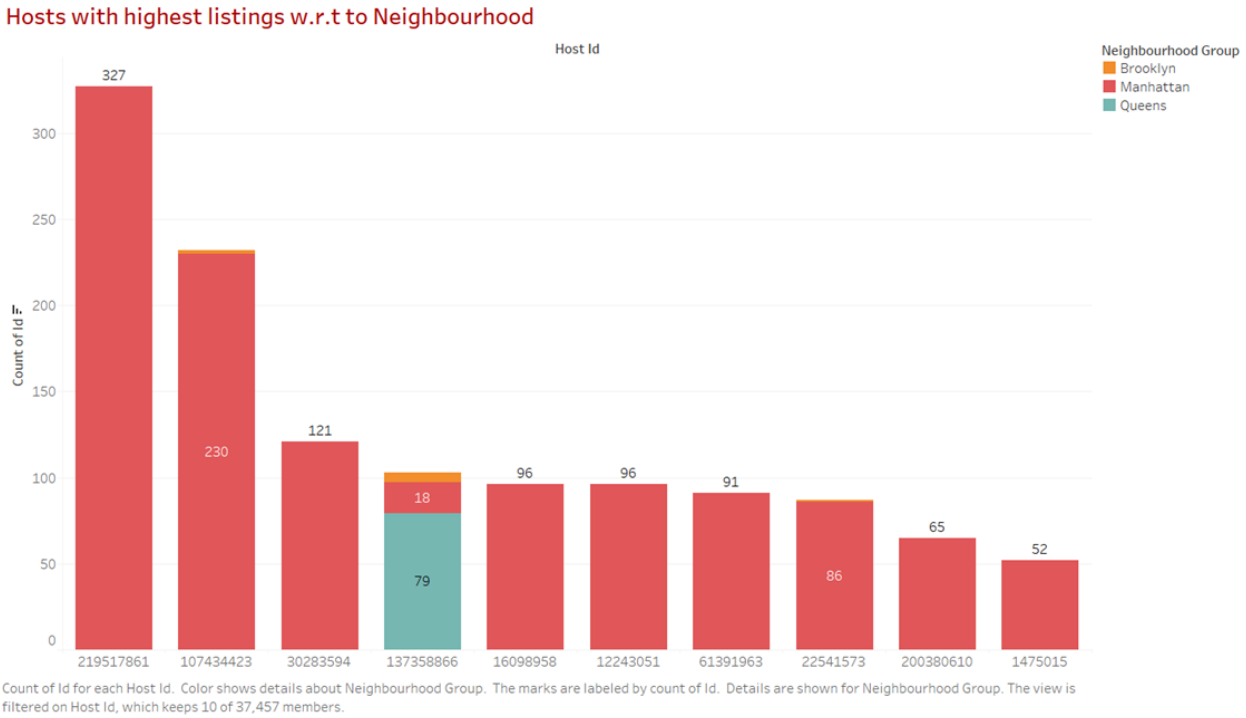
Inference:

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

➢ Who are the Hosts who have the highest listings w.r.t Neighbourhood?

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 neighbourhood group.



Inference:

- More experienced hosts know the market better.
- 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.
- This makes it more profitable for the host to acquire properties in the same area.

Presentation 2

Objective:

Now that the restrictions have started lifting and people have started to travel more, we have to make sure that Airbnb is fully prepared for this change.

The presentation will focus mainly on the following points:

- Understand customer preferences and customer experience in airbnb listings
- Understand the pricing relation to various parameters
- Recommendations to improve quality of new acquisitions and customer experience.

Exploratory Data Analysis:

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

- Customer preference for neighbourhood & room type
- Property demand based on minimum nights offered
- Price range preferred by customers
- Understanding Price variation w.r.t Room Type & Neighbourhood
- Understanding Price variation w.r.t Geography
- Top reviewed properties and their price range

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 important factors taken into consideration were customer booking volume and customer preference.
- 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 –
    - Customer experience : Neighbourhood, Room type & minimum nights offered
    - Price variation : Volume of customer booking, Room type, Neighbourhood, Number of reviews & Geography.
- Recommendations have been made keeping in mind the above parameters.

Explanation for EDA:

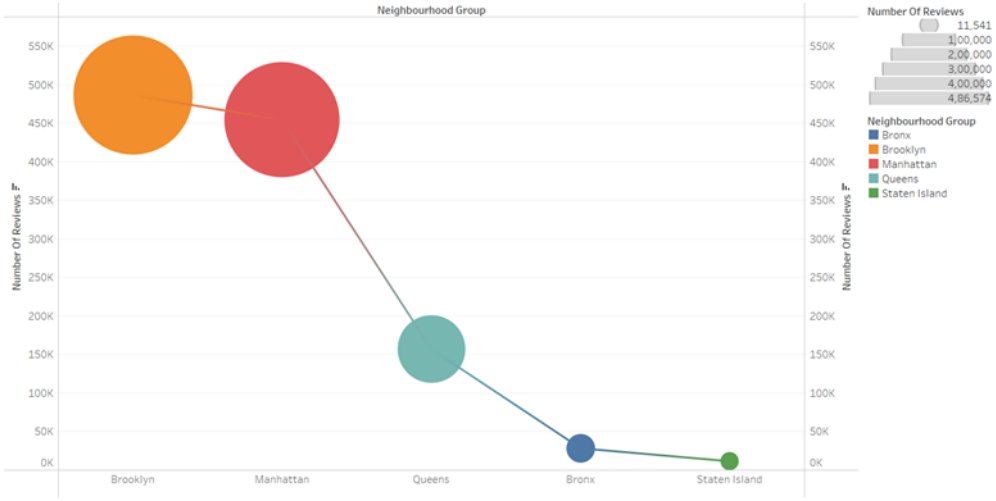
➢ Customer preference for neighbourhood & room type

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: neighbourhood & room type.

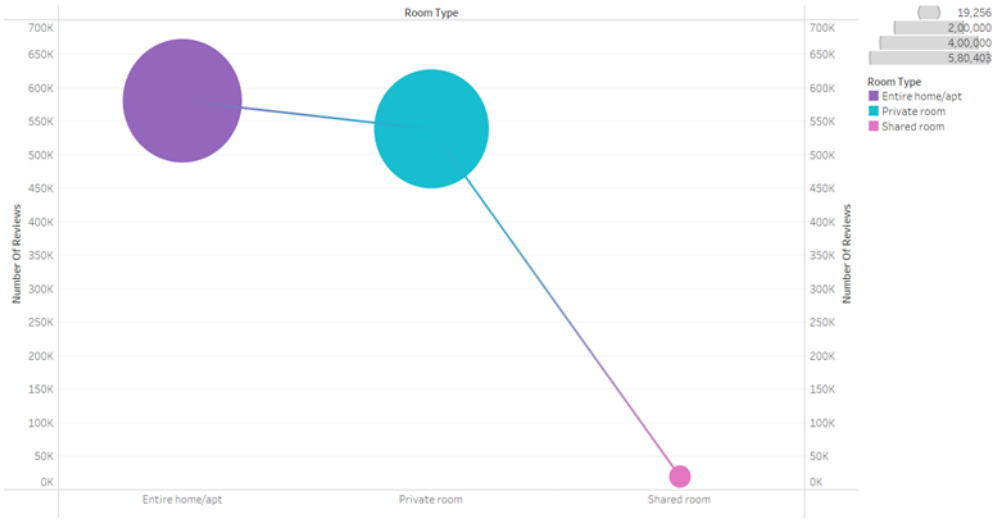
The parameters taken for analysis are: Room type; Neighbourhood group, SUM(Number of reviews)

No. of reviews according to Neighbourhood



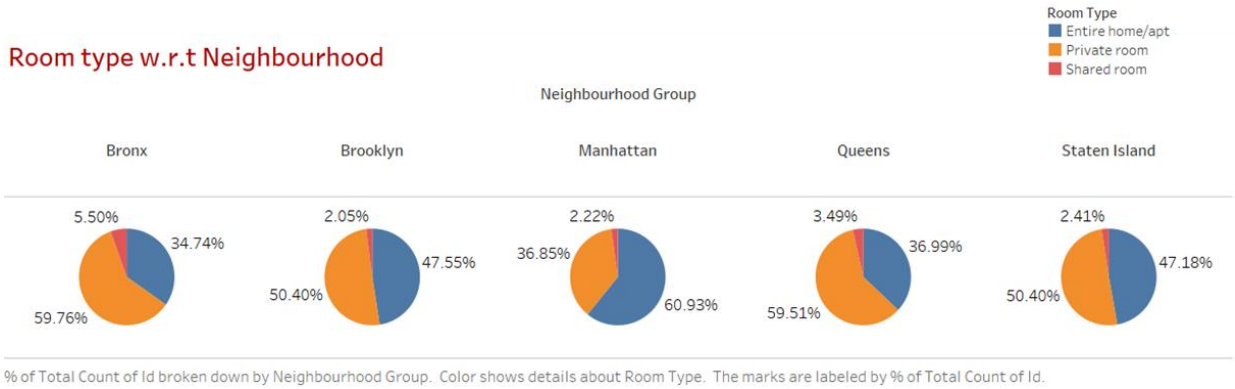
The trends of sum of Number Of Reviews and sum of Number Of Reviews for Neighbourhood Group. Color shows details about Neighbourhood Group. For pane Sum of Number Of Reviews: Size shows sum of Number Of Reviews.

No. of reviews according to Room type



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

Room type w.r.t Neighbourhood



% of Total Count of Id broken down by Neighbourhood Group. Color shows details about Room Type. The marks are labeled by % of Total Count of Id.

Observation:

- 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.

- Also, we can see maximum reviews in listings for Manhattan & Brooklyn, implying that more bookings happen in these neighbourhoods. *(The higher number of customer reviews imply higher satisfaction)*

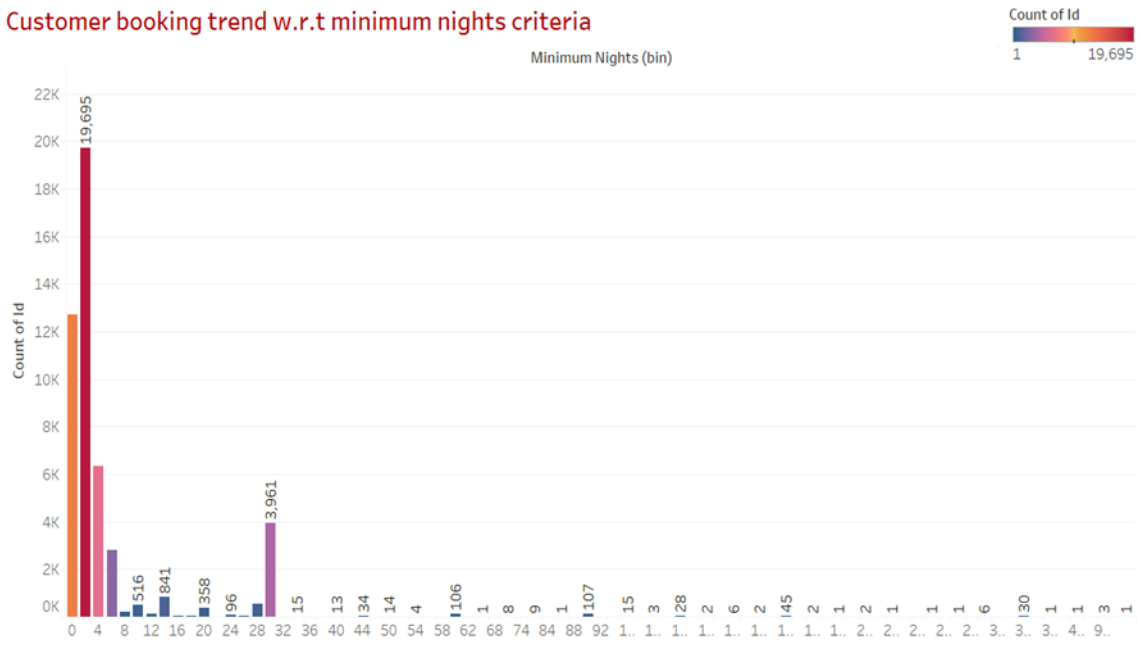
Recommendation:

- Airbnb can concentrate on promoting shared rooms with targeted discounts to increase bookings.
- New acquisitions can be explored to acquire 'private rooms' in Manhattan and Brooklyn and 'entire homes' in Bronx and Queens.

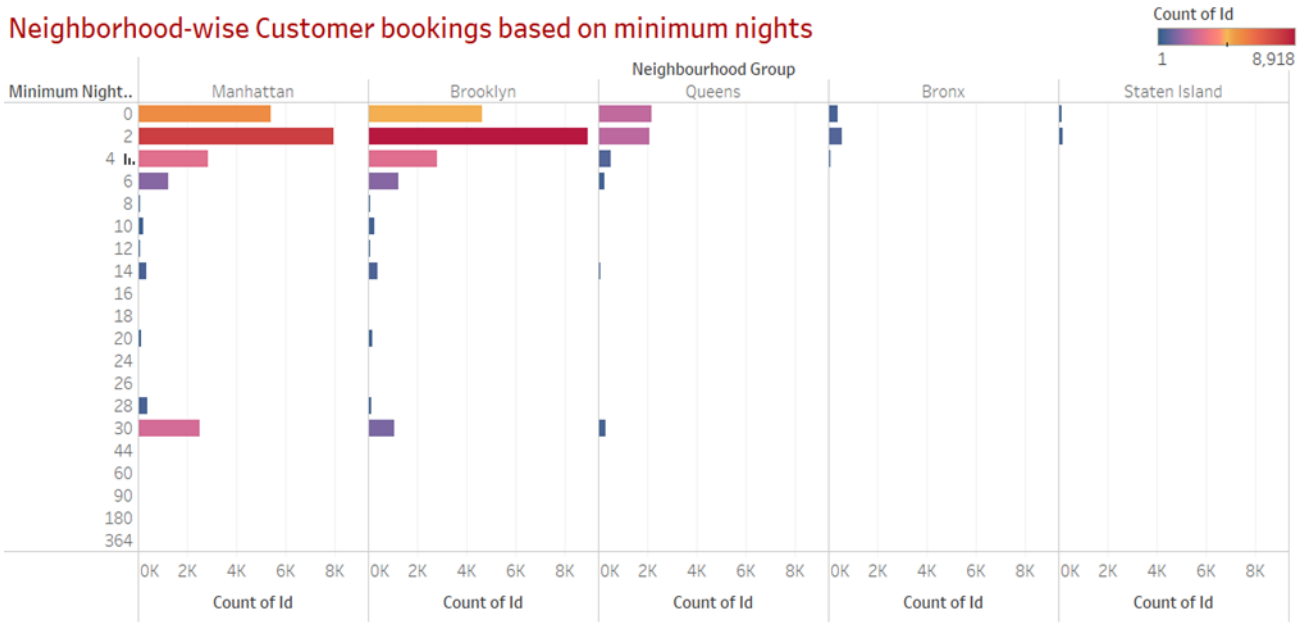
➢ Property demand based on minimum nights offered

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 do customers use airbnb; short-stay or long-stay. Here, we took into account the volume of booking and the neighbourhood-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 visualisation) & Neighbourhood Group.



Count of Id for each Minimum Nights (bin). Color shows count of Id. The marks are labeled by count of Id.



Count of Id for each Minimum Nights (bin) broken down by Neighbourhood Group. Color shows count of Id. Details are shown for Neighbourhood Group. The view is filtered on Minimum Nights (bin), which keeps 20 of 81 members.

Observation:

- **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 also be explained by the monthly rent taking trend.

Recommendation:

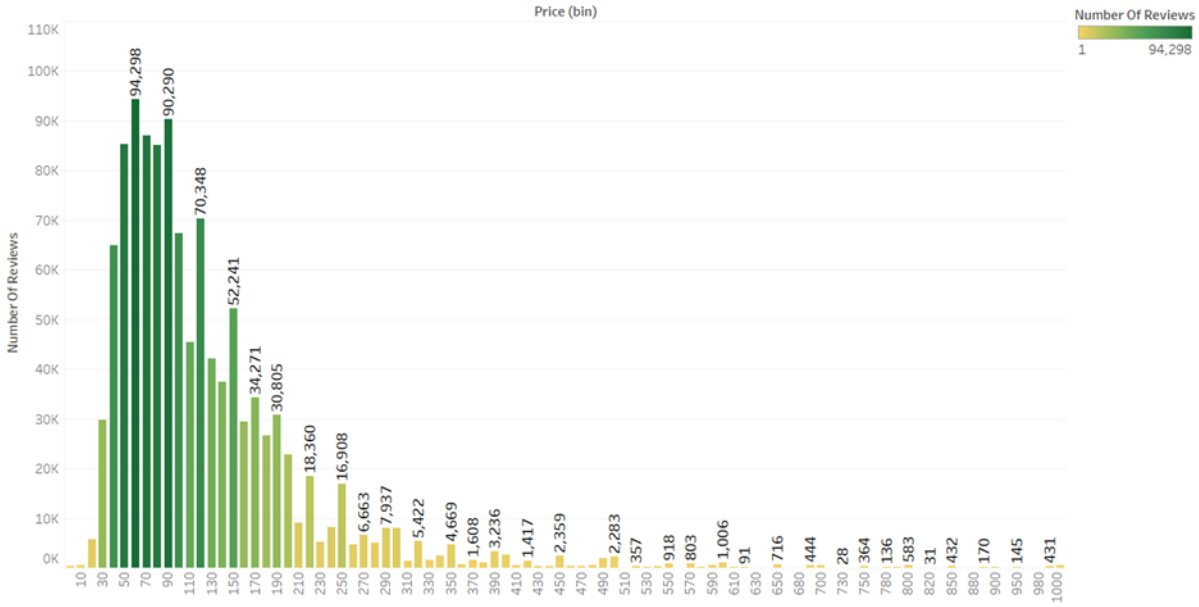
- More number of hosts & listings with monthly rental duration (30-60-90) can be acquired. We see a good potential in the 30-day rental window. Manhattan & Brooklyn have higher number of 30 day bookings compared to the others, these areas can be further targeted.
- Also, weekly or bi-weekly rentals can also be acquired as these can be used customers stranded in NYC for quarantine purposes.

➢ 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.

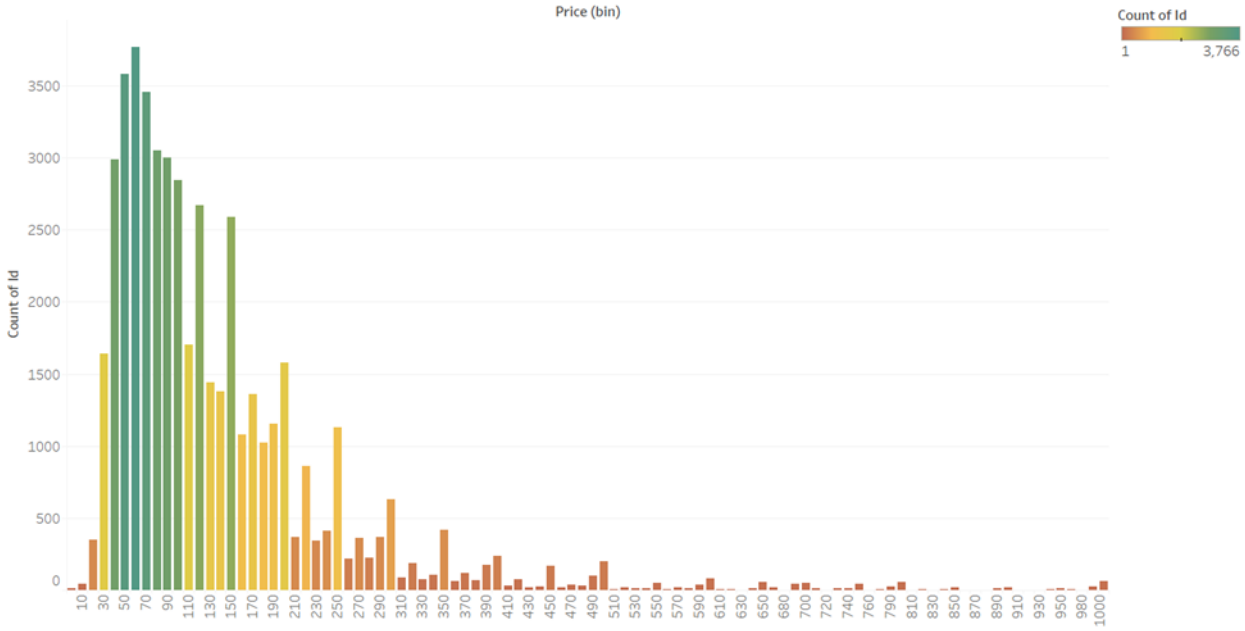
We have considered the volume of booking and number of reviews in a particular price range. For easy visualisation we have binned the Price with a bin size of 10. Also owing to the enormous value range we have observed the variation till $1000. As there was very little data beyond this, we decided to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

Price variation w.r.t Number of reviews



Sum of Number Of Reviews for each Price (bin). The marks are labeled by sum of Number Of Reviews. The data is filtered on Reviews Per Month, which keeps non-Null values only.

Pricing range preferred by customers



Count of Id for each Price (bin). Color shows count of Id. The view is filtered on Price (bin), which keeps 98 of 188 members.

Observation:

- 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 - $160. This is the price range most preferred by most customers.

Recommendation:

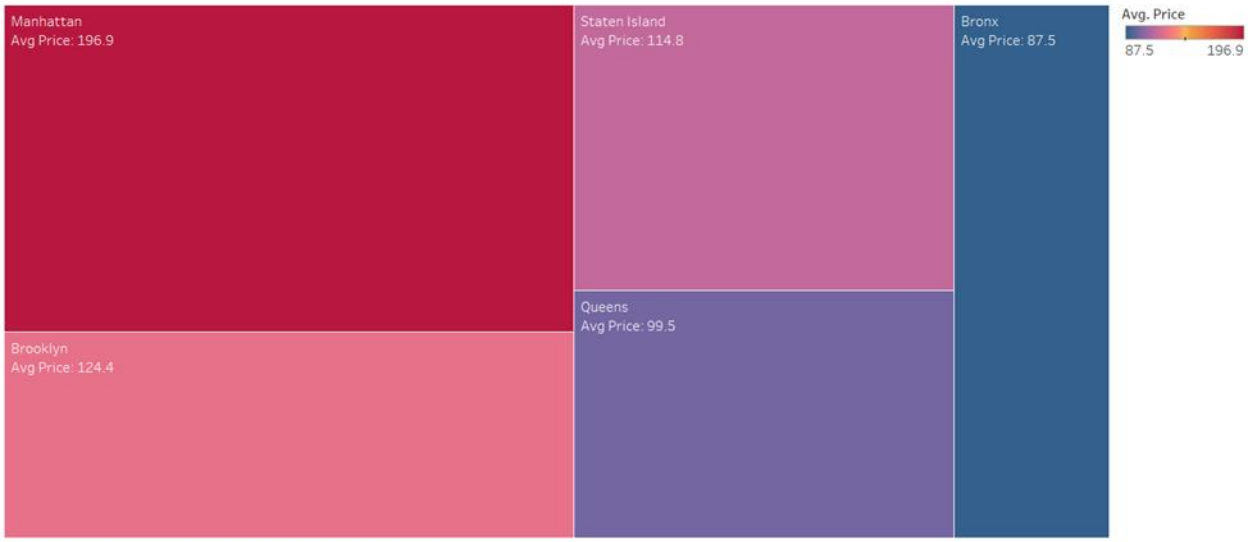
- New acquisitions and expansion can be done in the price range of $40 - $160 as it satisfies both parameters of; volume of customer traffic and customer satisfaction.

➢ Understanding Price variation w.r.t Room Type & Neighbourhood

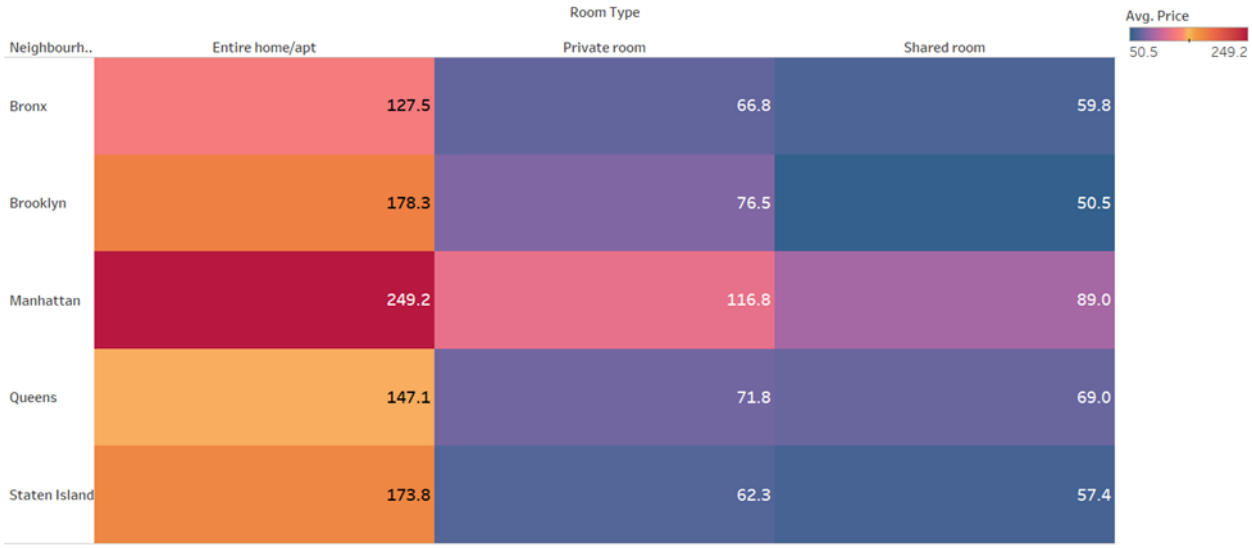
Now that we have obtained the optimum price range for listings, let us explore which neighbourhoods and room types fit in this category.

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 neighbourhood mainly consisted of numbers we decided to go ahead with highlight table to display the highest and lowest values.



Neighbourhood Group. Color shows average of Price. Size shows average of Price. The marks are labeled by Neighbourhood Group.



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

Observation:

- Manhanttan 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.
- 'Shared Room' type is the cheapest in Brooklyn.

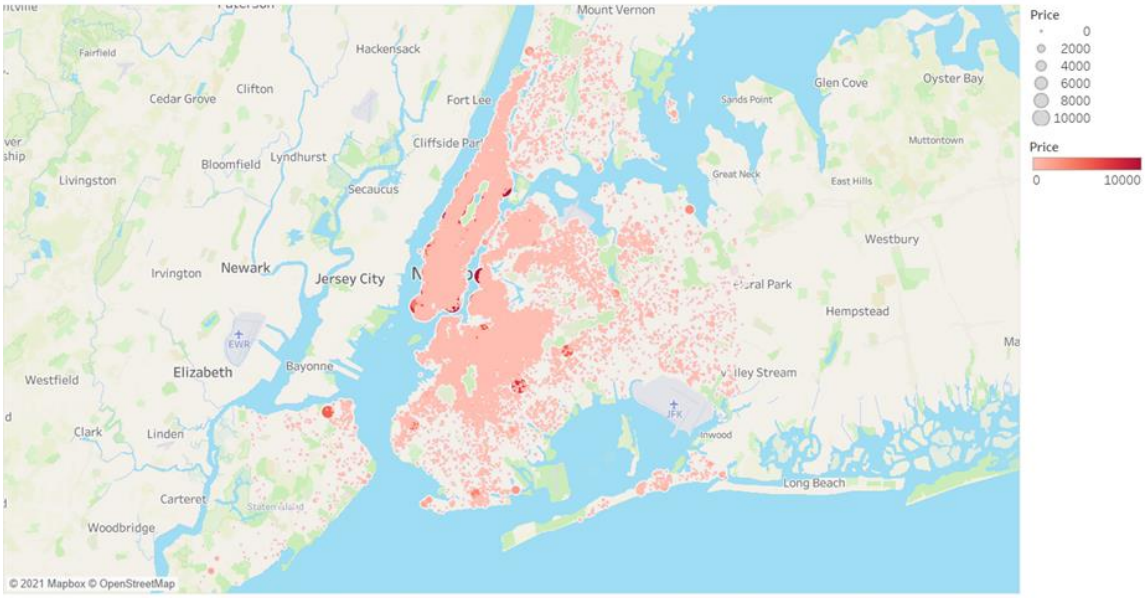
Recommendation:

- In line with the earlier recommendation, we observe that 'private rooms' of Manhattan & Brooklyn and and 'entire homes' in Bronx and Queens fall in the favorable price range ($40-$160).
- Brooklyn has an average price of $124. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.

➢ 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 neighbourhood with maximum average price. We used the findings in this to confirm our observation obtained from the geographical map.



Map based on Longitude and Latitude. Color shows details about Price. Size shows details about Price. Details are shown for Neighbourhood and Neighbourhood Group.

Average of Price for each Neighbourhood broken down by Neighbourhood Group. Color shows details about Neighbourhood Group. The view is filtered on Neighbourhood, which keeps 10 of 221 members.

Observation:

- 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.
- The bar graph confirms our inference, as we observe that the top 10 neighbourhoods according to price are those that are situated near the sea or are next to important institutions/companies/landmarks.

Recommendation:

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

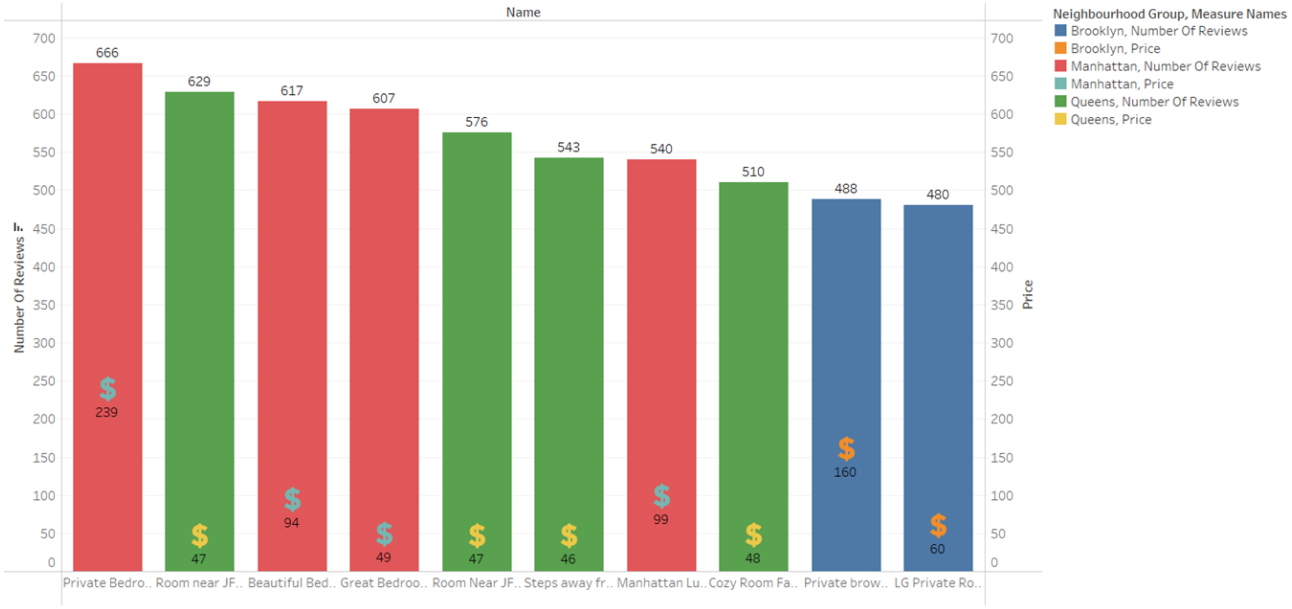
➢ Top reviewed properties and their price range

We have gotten various insights in the above questions regarding price range or neighbourhood.

To confirm and correlate our observations, we have visualised 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. We have also simultaneously seen the pricing range for the listing.

Top 10 Listings based on number of reviews along with their pricing



Observation:

- Manhattan, Brooklyn and Queens have the most liked properties (most reviewed). Most of the properties also fall in our estimated favourable range of $40-$160.
- 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:

- Promotion of shared rooms with targeted discounts to increase bookings.

- 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.
- Weekly or bi-weekly rentals can also be acquired as these can be used customers stranded in NYC for quarantine purposes.
- New acquisitions and expansion can be done in the price range of $40 - $160 as it satisfies both parameters of; volume of customer traffic and customer satisfaction.
- New acquisitions can be explored to acquire 'private rooms' in Manhattan and Brooklyn and 'entire homes' in Bronx and Queens as they fall in the favorable price range ($40-$160).
- Brooklyn has an average price of $124. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.
- Increasing acquisitions and new properties in coastal regions can increase customer bookings.

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.