AIRBNB NYC Case Study – Methodology

(Image courtesy: https://www.spinxdigital.com/)

Authored By: Ajinkya Survase

Jagruti Desai

Mohit Thakur

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 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

Price Analysis Room Type Wise

Price for each Room Type.
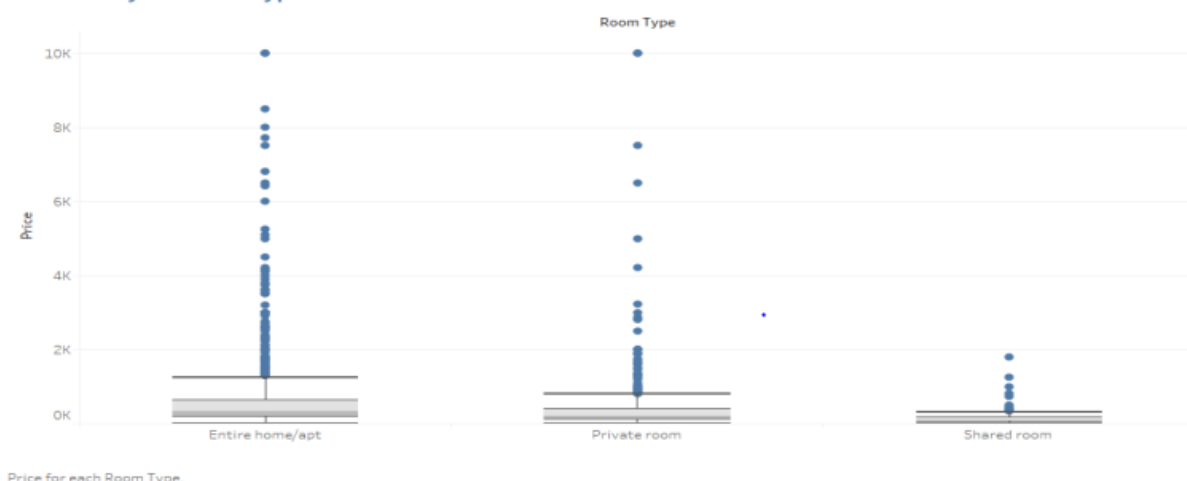
so let them stay as it is ¬ Price was highly positively skewed so median was very close the lower quartile with some outliers as seen in the boxplot below

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 neighbourhood

2. Most preferred room type

3. 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 –

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:

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. Bar plot were used to explore this question:

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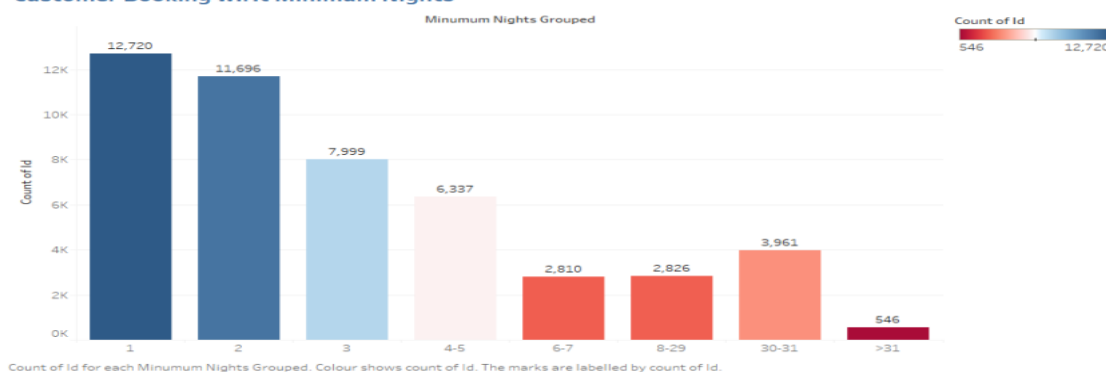

AirBnb Listing distribution in NYC
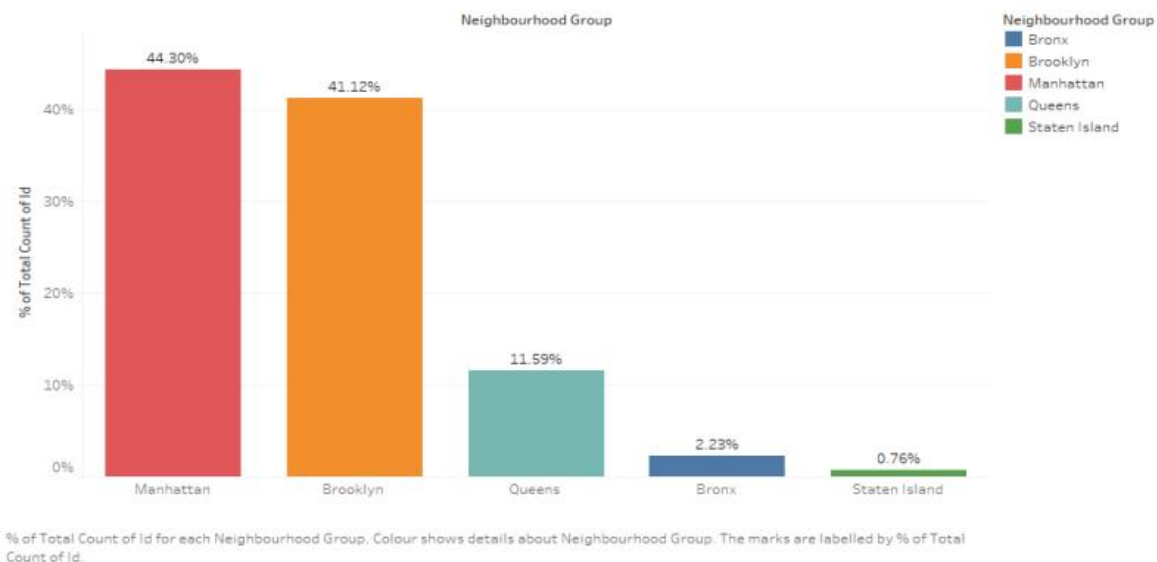
• Inference: 1. We see that, Airbnb has good presence in Manhattan, Brooklyn & Queens. 2. Listings are maximum in Manhattan (44%) & Brooklyn (41%) owing to the high population density and it being the financial and tourism hub of NYC. Staten Island (~1%) has the least number of listings, due to its low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

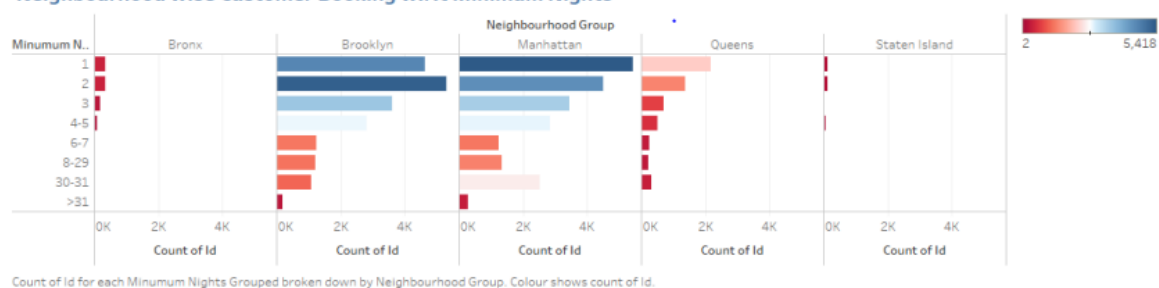**Property demand based on minimum nights offered**

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

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 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.

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

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 that 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 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; Neighbourhood group, SUM(Number of reviews)

Total Reviews w.r.t Neighbourhood Group

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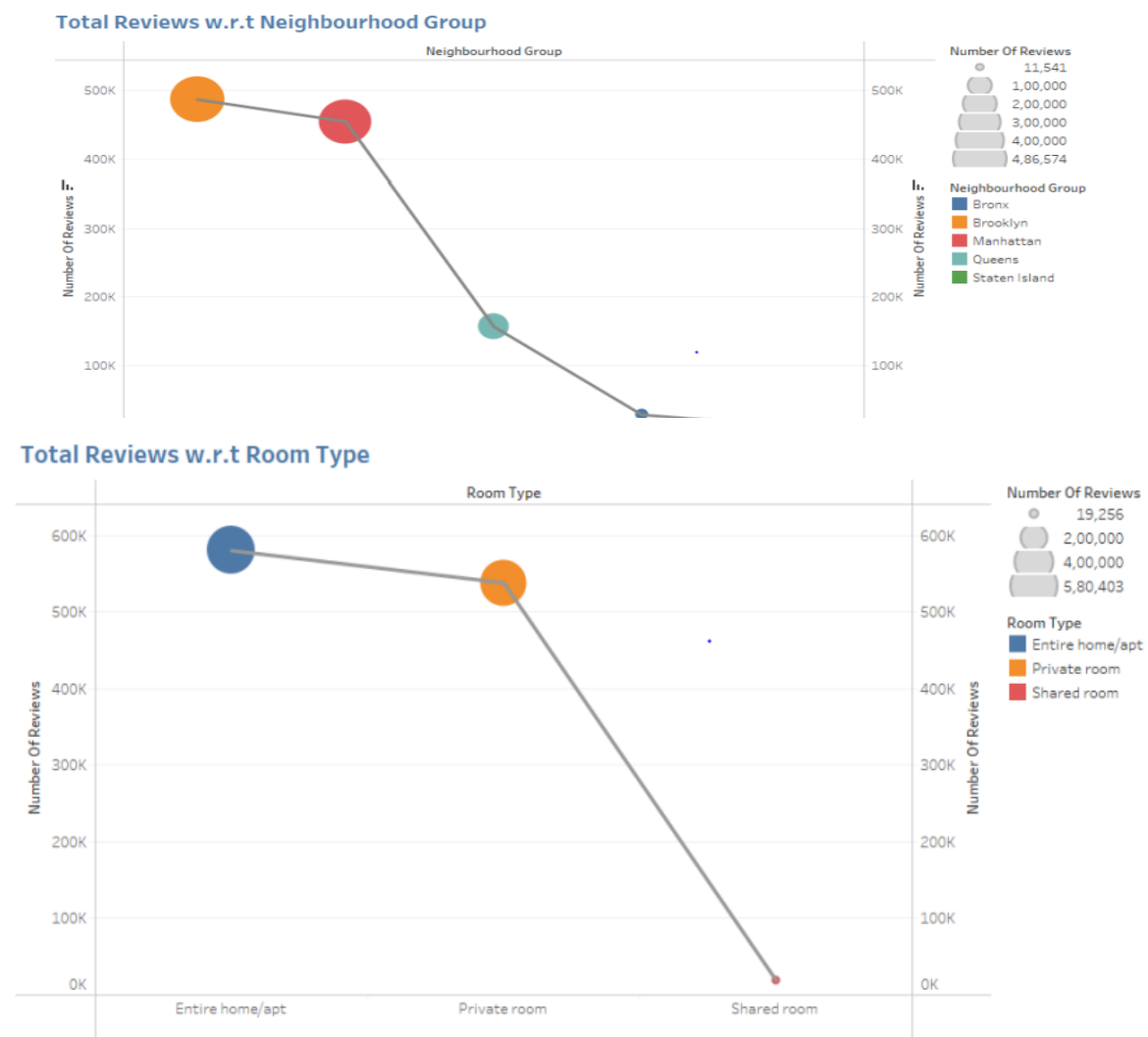
The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Room Type. For pane Sum of Number Of Reviews: Colour shows details about Room Type. Size shows sum of Number Of Reviews.
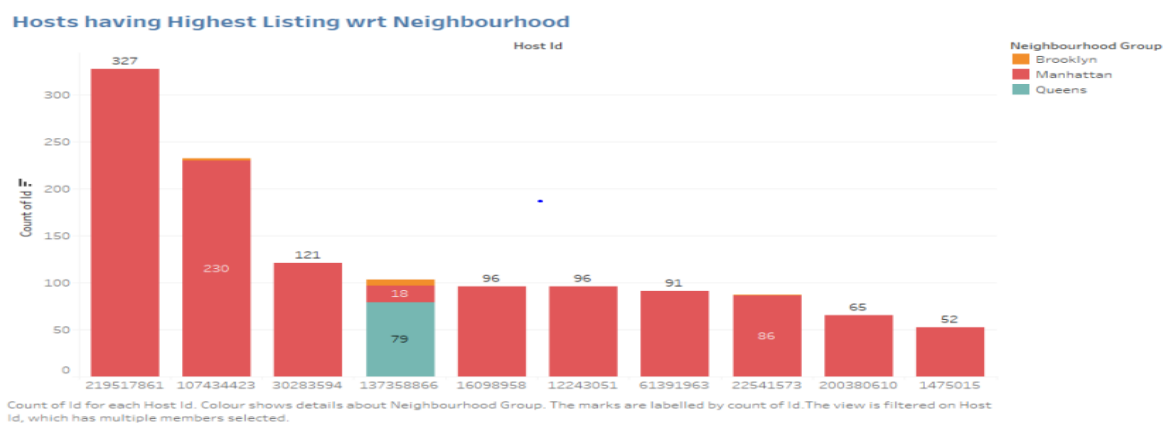
Inference:

1. 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.

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.

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.

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 & Neighbourhood

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: Neighbourhood, Room type & minimum nights offered

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

Geography.

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

Explanation for EDA:

1. 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.
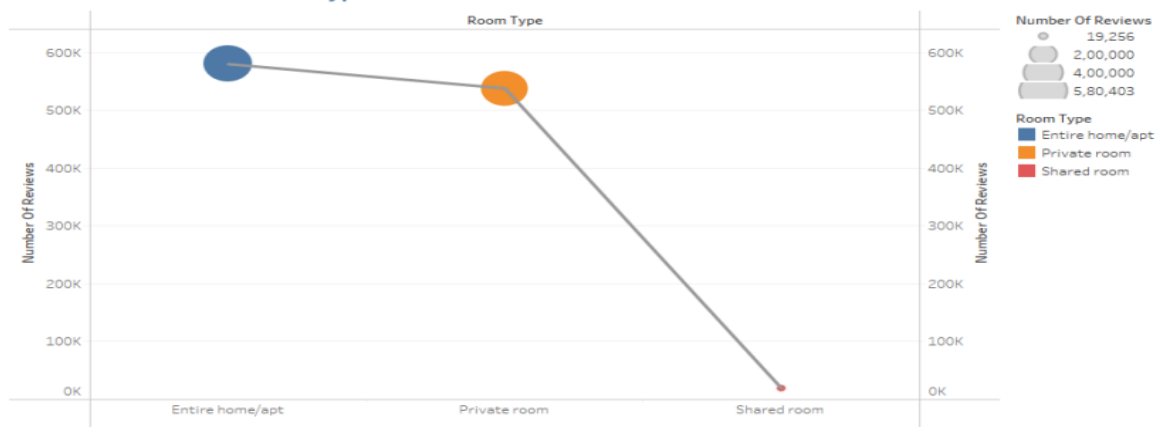
12 The parameters taken for analysis are: Room type; Neighbourhood group, SUM(Number of reviews

Total Reviews w.r.t Neighbourhood Group
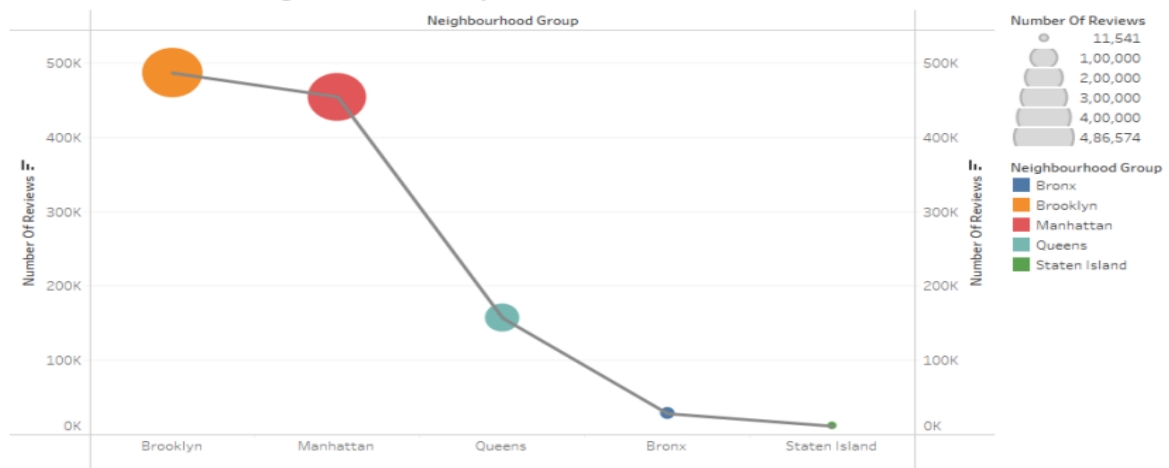

The trends of sum of Number Of Reviewsandsum of Number Of Reviews for Neighbourhood Group. For pane Sum of Number Of Reviews: Colour shows details about Neighbourhood Group. Size shows sum of Number Of Reviews.

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.

Room Type Distribution in each Neighbourhood Group
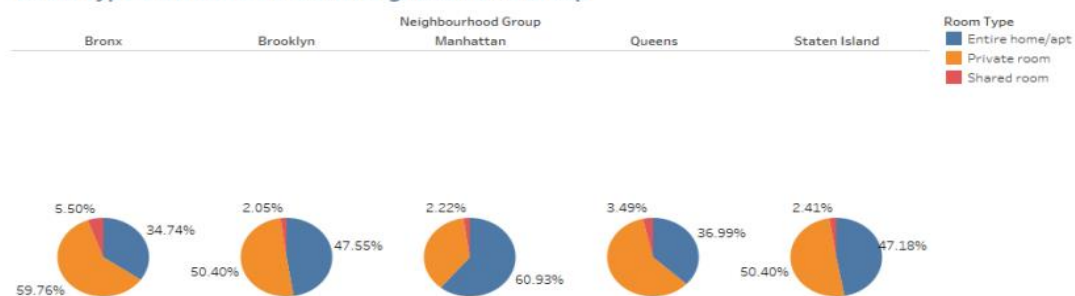
% of Total Count of Room Type broken down by Neighbourhood Group. Colour shows details about Room Type. The marks are labelled by % of Total Count of Room Type.

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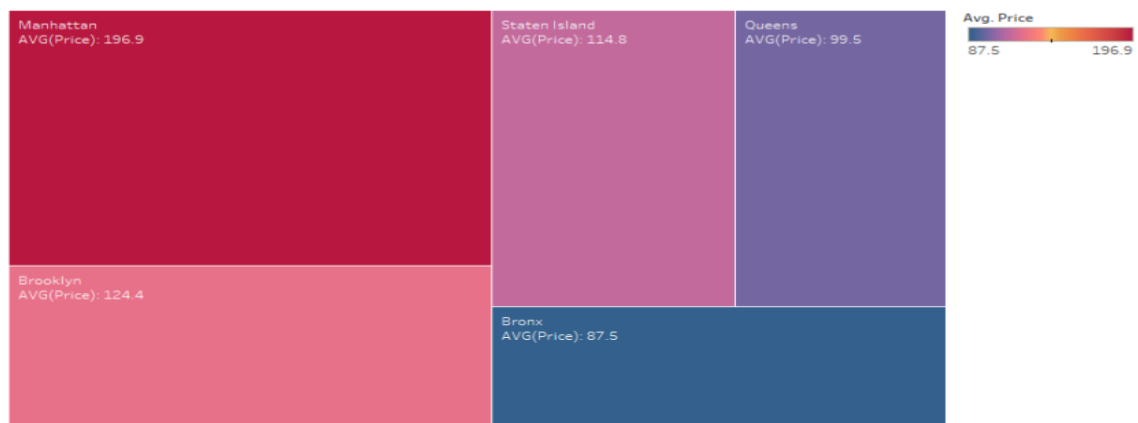
Neighbourhood Groupandaverage of Price. Colour shows average of Price. Size shows average of Price. The marks are labelled by Neighbourhood Groupandaverage of Price


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

Top reviewed properties:

We have gotten various insights in the above questions regarding price range or neighbourhood. 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.
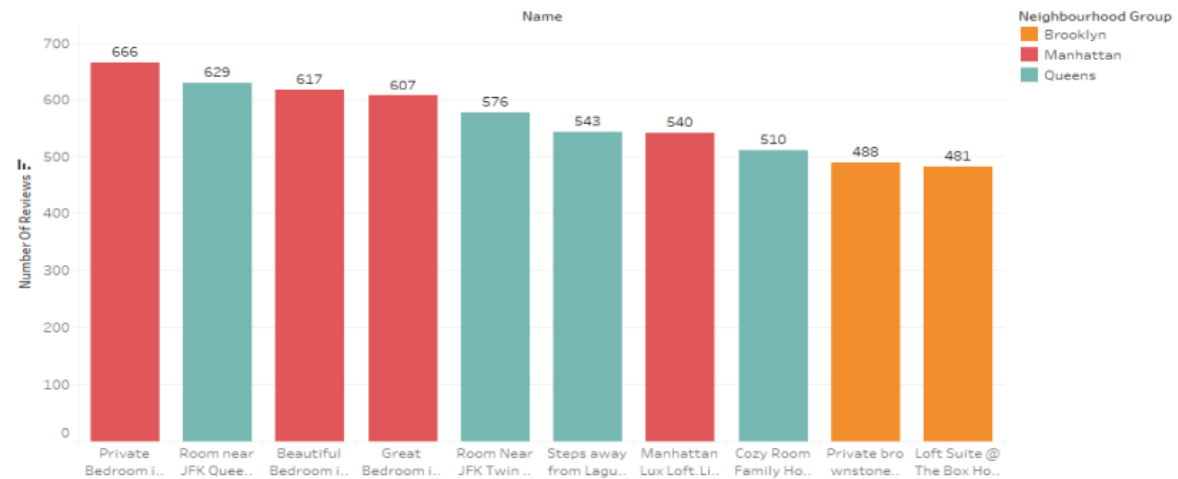
Sum of Number Of Reviews for each Name. Colour shows details about Neighbourhood Group. The marks are labelled by sum of Number Of Reviews. The view is filtered on Name, which has multiple members selected.