

# Methodology Document

Step1:

Story Boarding

- Familiarised with Airbnb data dictionary, to gain an understanding of the meaning of each variable, the units of measurement, and the possible range of values.

Step2:

Data cleaning and preparation

- We begin with importing Airbnb dataset to Jupyter notebook to analyse the data using Python programming language and pandas library.

```
In [2]: df1=pd.read_csv('AB_NYC_2019.csv')
```

```
In [4]: df1.shape # check number of rows and columns present in dataset.
```

```
Out[4]: (48895, 16)
```

- The dataset has 16 columns and 48895 rows.
- Checked for duplicate values if present any.
- After checking for missing values, we found four columns with missing values i.e [name](#), [host\\_name](#), [last\\_review](#) and [reviews\\_per\\_month](#).
- For easier analysis we replaced the columns with missing values more than 15 percent i.e [last\\_review](#) and [reviews\\_per\\_month](#) with standardised (median or mode) value.

Below is snippet of code we used to replace missing values.

```
In [5]: # making a function to find percentage of missing values in each of the column.  
def find_null(df1):  
    return 100*df1.isnull().mean()
```

```
In [6]: find_null(df1) # Using above find_null function we found % of missing values in each column.
```

```
Out[6]: id                0.000000  
name                0.032723  
host_id             0.000000  
host_name           0.042949  
neighbourhood_group 0.000000  
neighbourhood       0.000000  
latitude            0.000000  
longitude           0.000000  
room_type           0.000000  
price               0.000000  
minimum_nights      0.000000  
number_of_reviews   0.000000  
last_review         20.558339  
reviews_per_month   20.558339  
calculated_host_listings_count 0.000000  
availability_365    0.000000  
dtype: float64
```

```

In [9]: # We have used mode method for filling missing values in column 'last_review' as it is a categorical column.
mod_df1=df1['last_review'].mode()[0]
df1['last_review']=df1['last_review'].fillna(mod_df1)

In [10]: # We have used median method for filling missing values in column 'reviews_per_month' as it is a continuous column.
med_df1=df1['reviews_per_month'].median()
df1['reviews_per_month']=df1['reviews_per_month'].fillna(med_df1)

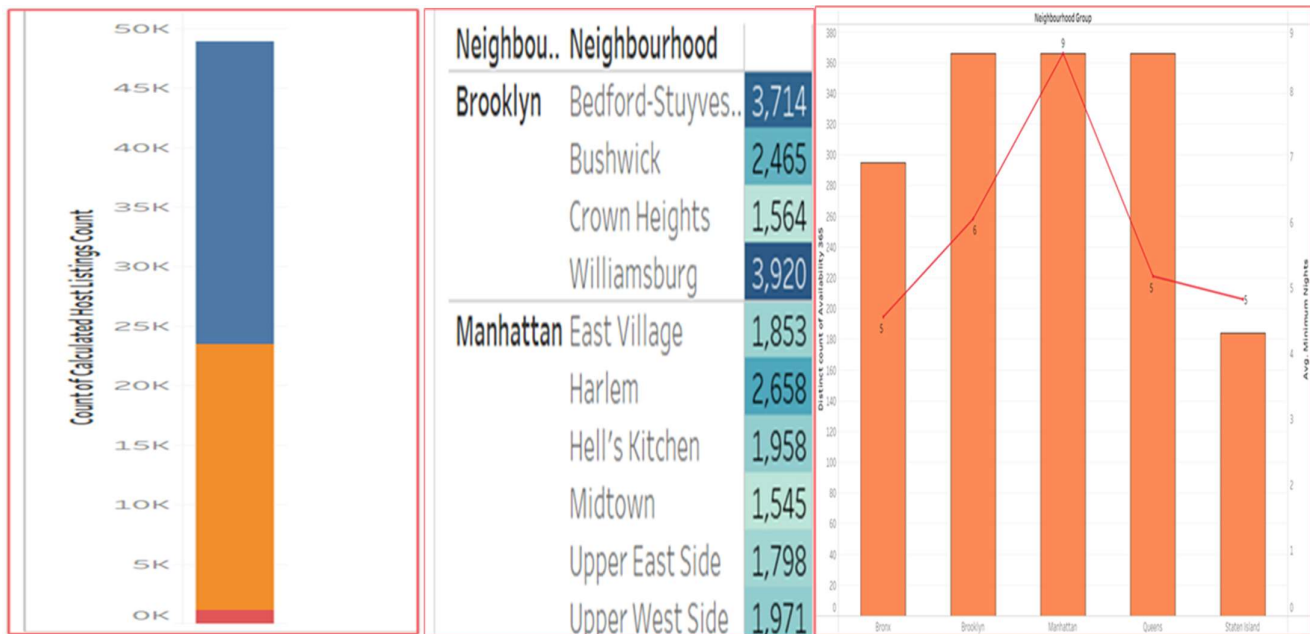
```

Step 3:

Performing Data Analysis:

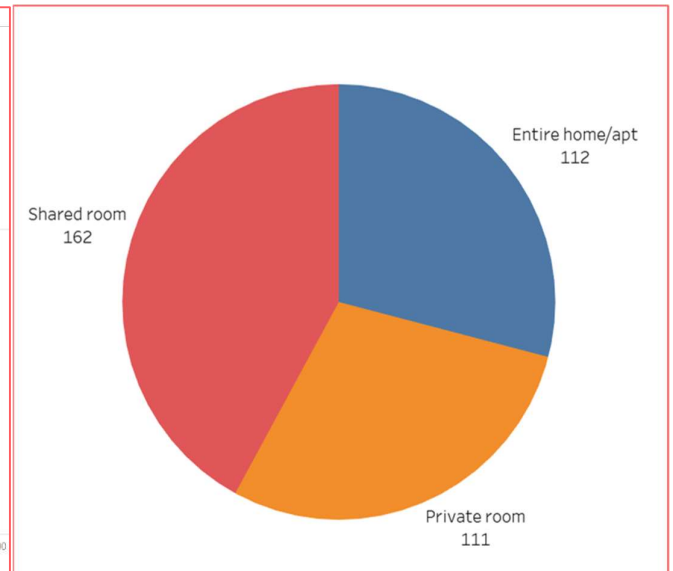
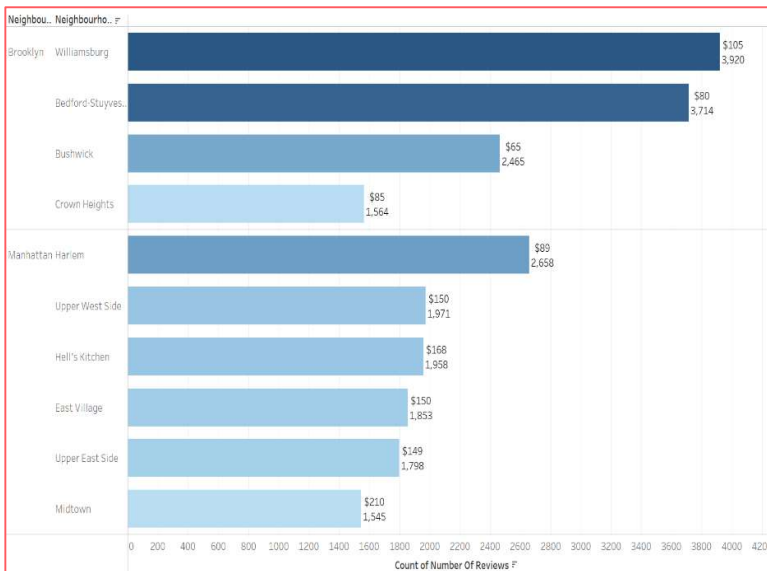
- After Data cleaning and preparation, we used Tableau to perform data analysis on the dataset for gaining insights on different aspects and attributes.

These are the graphs we plotted for analysing Host Acquisition:



Inferences from the graph:

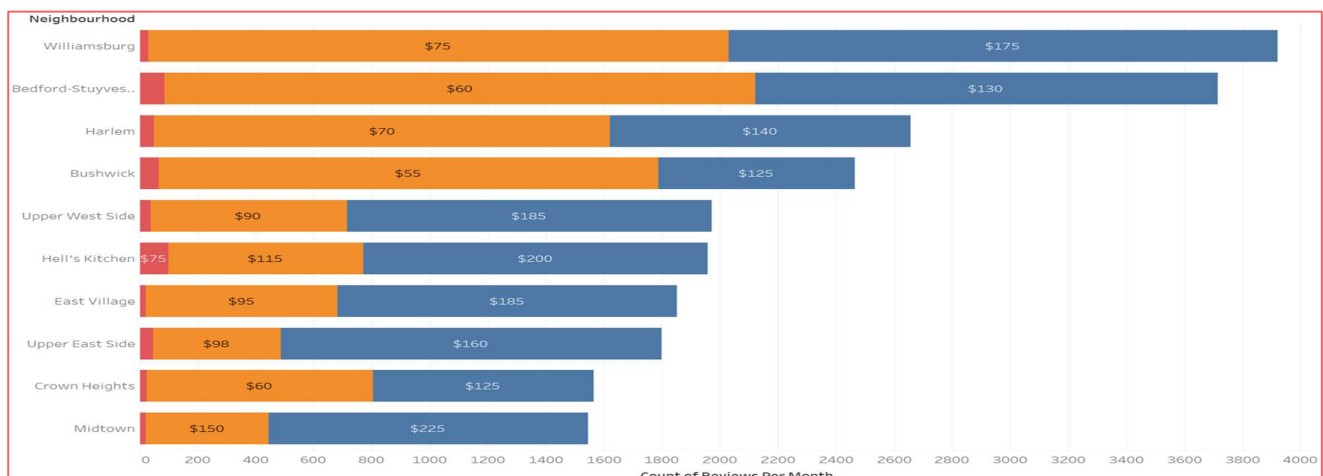
- Top ten neighbourhoods with most listings are mostly concentrated in Brooklyn and Manhattan with most popular room type being Entire home/apartment or Private room.
- We can also infer that Manhattan, Brooklyn and Queens have the highest list of hosts who are available for 365 days of the year with average minimum nights ranging from nine to five days, Manhattan being the highest.



### Inferences from the graph:

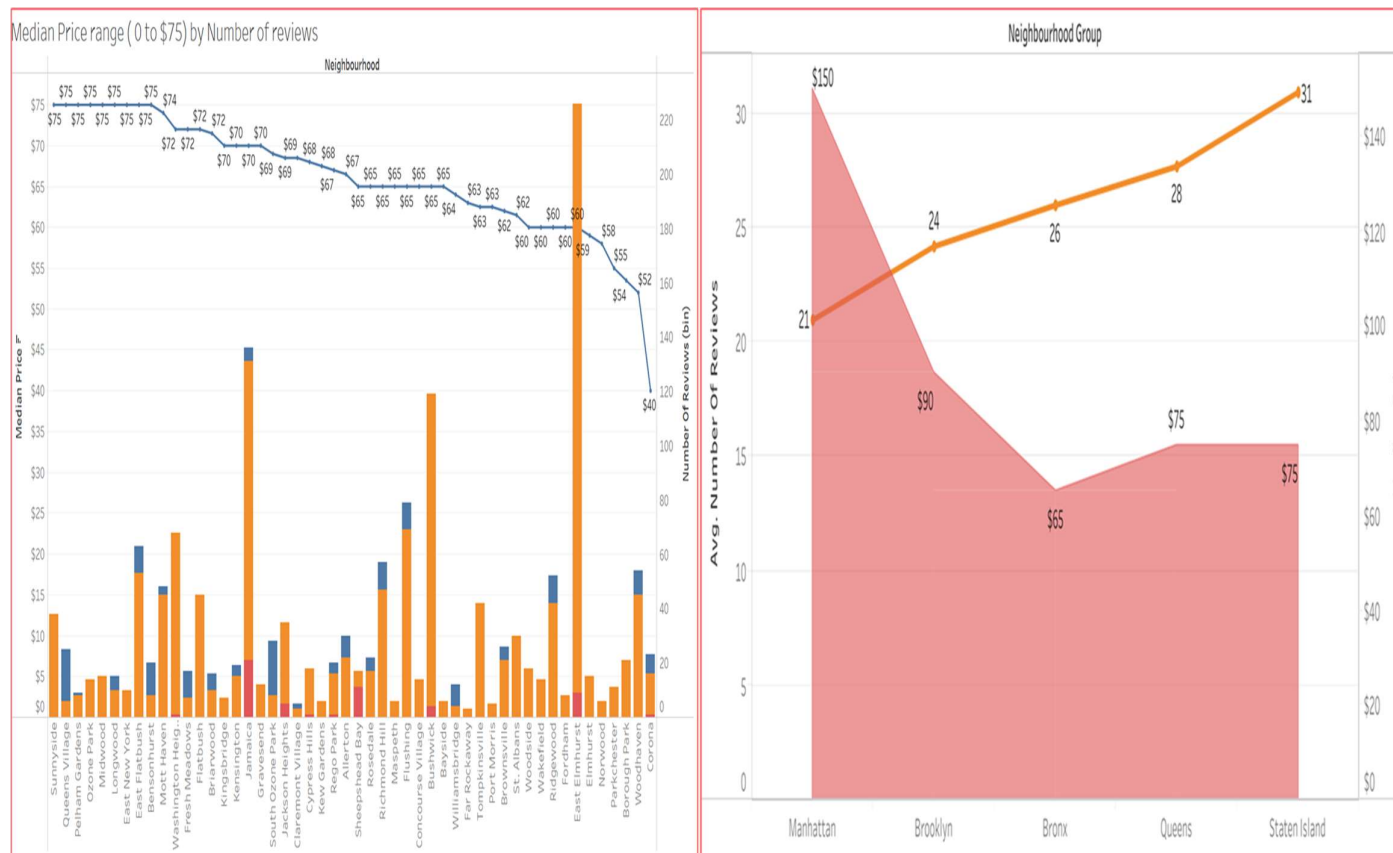
- Top ten neighbourhoods with most number of reviews and number of listings for each neighbourhood, Williamsburg being the highest in Brooklyn and Harlem in Manhattan, with the median price ranging from \$105 to \$85 in Brooklyn and \$210 to \$89 in Manhattan.
- There are fluctuations in median price ranges in the neighbourhoods, in Brooklyn price seems to be decreasing w.r.t rank, but in Manhattan it is the opposite, as the neighbourhood with a smaller number of reviews has high price compared to the neighbourhood with high number of reviews.
- Pie chart depicts room type which is available most days of the year.

These are the graphs, we plotted to analyse customer preference:



## Inferences from the graph:

- From the above graphs we can infer that Top ten neighbourhoods with highest reviews per month are Williamsburg, Bedford-Stuyvesant, Harlem, Bushwick, Upper West Side, Hell's Kitchen, East Village, Upper East Side, Crown Heights and Midtown Where Entire home/Apartment is being preferred by 51.97%, Private Room by 45.66% and Shared Room by 2.37%.

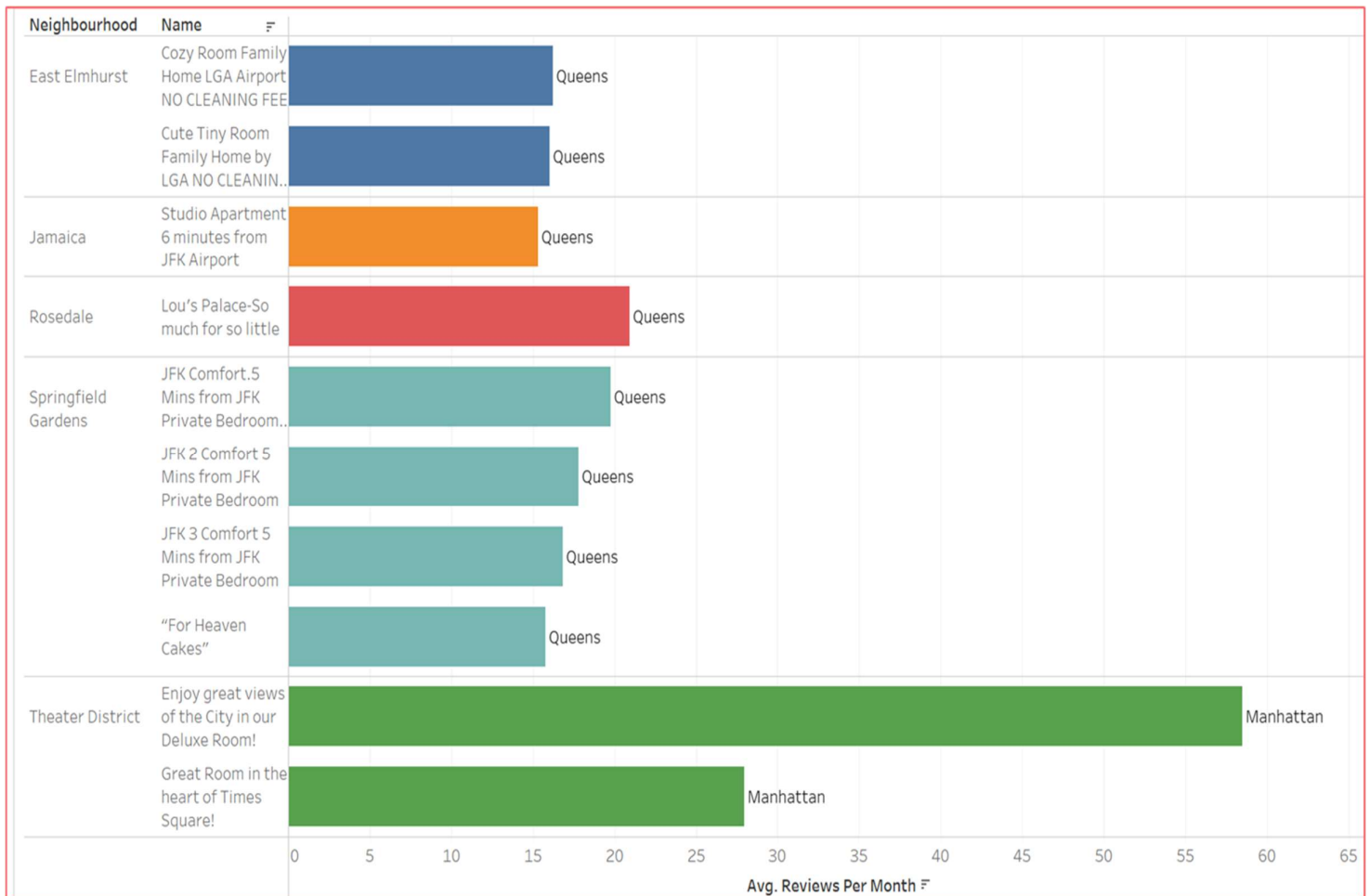


## Inferences from the graph:

- Median Price below \$75 has highest bookings for private room. The price ranges from \$40 to \$60 has range of 100 to 180 number of reviews, where median price \$60 to \$75 has 180 to 220 number of reviews. East Elmhurst having the highest number of reviews with price of \$60. Jamaica and Bushwick being the next highest with median price of \$70 and \$65.
- Staten island has the highest average number of reviews and Manhattan with least average number of reviews.

- Manhattan has the highest price of \$150 compared to any other neighbourhood groups, Bronx being the least with price range starting from \$65.

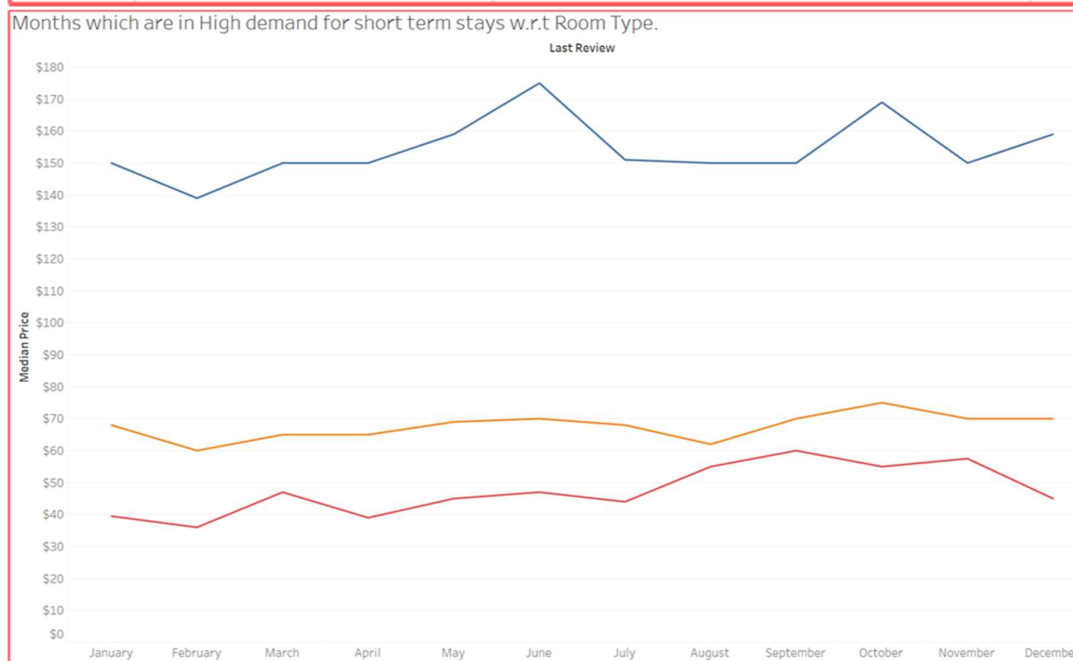
### Analysis on Popular localities and properties



### Inferences from the graph:

- East Elmhurst, Jamaica, Rosedale, Springfield Gardens and Theater District are the neighbourhoods which have the most popular properties in New York, where Two most popular properties are situated in Manhattan and Queens having the remaining eight most popular properties based on Average Reviews Per Month.

These are the Graphs we plotted to analyse short-term stays w.r.t Customer Preferences:

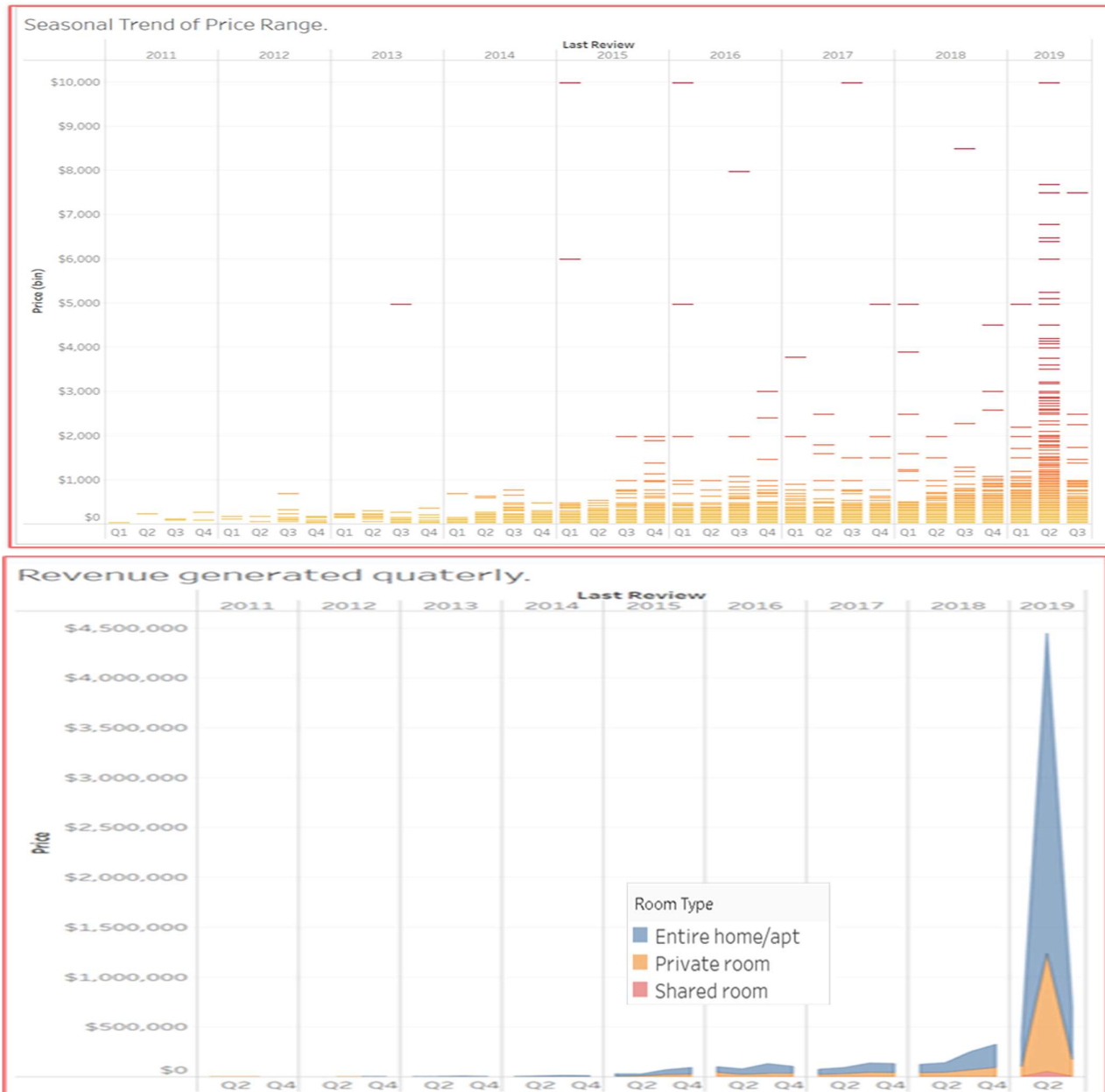


Inferences from the graph:

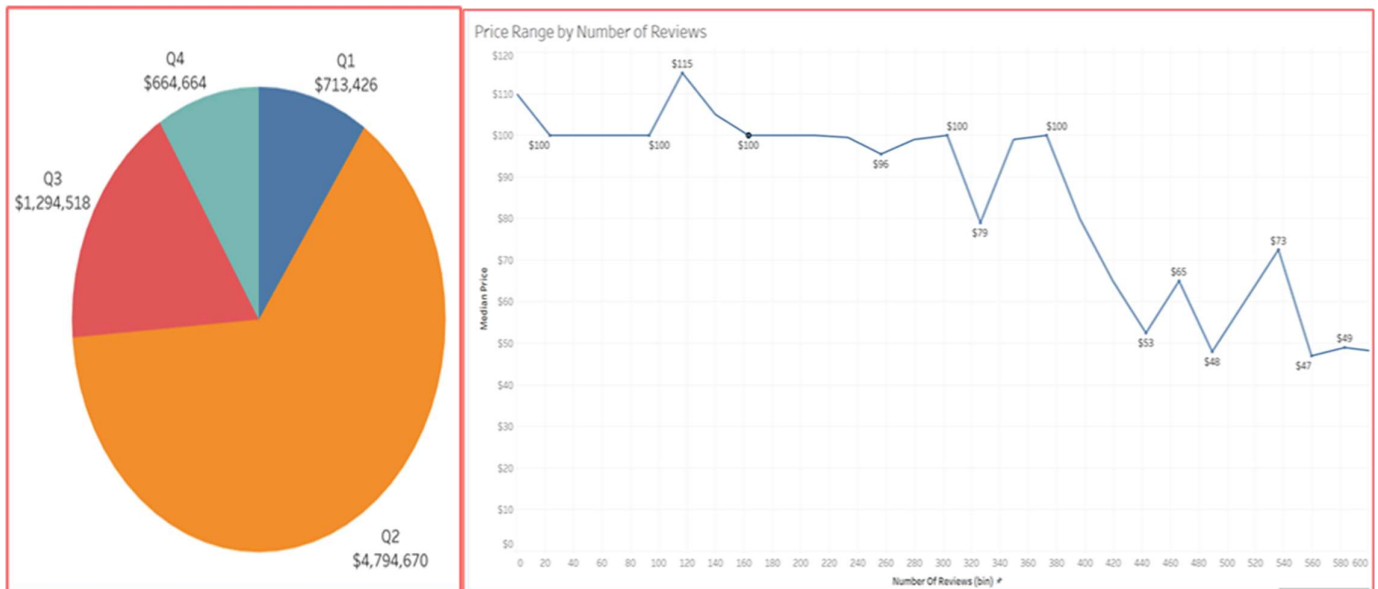
- Upper west Side, Upper East Side, Midtown, Hell's Kitchen, Harlem and East Village have most number of reviews for short term stays between 30 days to 60 days in Manhattan. Bedford-Stuyvesant, Williamsburg, Bushwick and Crown Heights have the most number of reviews for short term stays between 30 days to 60 days.

- Bushwick has the least price of \$50 while Midtown has the highest price of \$200.
- June and November are high in demand for short term stays between 30 to 60 days where most sorted listing is Entire home/Apartment.

Graphs we plotted to do Analysis on Prices:







Inferences from the graph:

- The price has increased as the years progressed, it can indicate that the demand for rental properties has increased over time, leading to an increase in prices of rental. It can also suggest that the supply of short-term rental properties has not kept up with the demand, leading to an increase in prices.
- The second quarter of 2019 was a highly profitable period for Airbnb, and the most popular types of listings during that time were Entire Home/apartment and private rooms.
- The median price range between \$75 and \$50 has highest number of reviews.



## Step 4: Summary

After Data Analysis we arrived at conclusions mentioned below:

- Prioritize acquiring hosts in Brooklyn and Manhattan, focusing on "Entire home/apartment" or "Private room" listings, as they are the most popular room types. Consider Manhattan, Brooklyn, and Queens for longer stays as they have the most year-round available hosts.
- Offer discounts in lower median price neighbourhoods like Bushwick in Brooklyn or Washington Heights in Manhattan, and promote highly reviewed areas such as Williamsburg in Brooklyn and Harlem in Manhattan.
- Encourage hosts to offer more private rooms with median prices ranging from \$40 to \$75, and incentivize competitive pricing in Manhattan. Promote Staten Island's high average reviews to encourage bookings in the area.
- Promoting shared rooms with attractive price and incentives for short term stays will be profitable as the number of listings for shared rooms is limited but their average price is placed less and availability is high.
- Encourage hosts in Upper West Side, Upper East Side, Midtown, Hell's Kitchen, Harlem, and East Village to offer short-term stays between 30 to 60 days, as these neighbourhoods have the most number of reviews for this length of stay. Encourage hosts in Bedford-Stuyvesant, Williamsburg, Bushwick, and Crown Heights to offer short-term stays between 30 to 90 days, as these neighbourhoods have the most number of reviews for this length of stay.
- Consider offering promotions or discounts for bookings made in June and November, which are high-demand months for short-term stays between 30 to 60 days. This could encourage more bookings during these periods and help increase revenue.

- Investing in the development of new short-term rental properties to help meet the growing demand and prevent further price increases.
- Encourage hosts to price their listings competitively within the median price range of \$75 to \$50, as this is where the highest number of reviews are found. Consider offering special promotions or discounts during off-peak periods to encourage more bookings and maximize revenue throughout the year.

#### Step 5: Presentation

- Made a presentation to present insights and recommendations gained after data analysis.