Methodology

1. Research Problem

● For the past few months, Airbnb has seen a major decline in revenue due to the lockdown imposed during the pandemic.

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

2. Business Understanding

Airbnb is an American company based in San Francisco, California. It operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. The platform is accessible via the website and mobile app.

3. Whom are we presenting?

● **Data Analysis Managers:** These people manage the data analysts directly for processes and their technical expertise is basic.

● **Lead Data Analyst:** The lead data analyst looks after the entire team of data and business analysts and is technically sound.

● **Head of Acquisitions and Operations, NYC:** This head looks after all the property and hosts acquisitions and operations. Acquisition of the best properties, price negotiation, and negotiating the services the properties offer falls under the purview of this role.

● **Head of User Experience, NYC**: The head of the user experience looks after the customer preferences and handles the properties listed on the website and the Airbnb app. Basically, the head of the user experience tries to optimize the order of property listing in certain neighborhoods and cities in order to get every property the optimal amount of traction.

4. Method of Analysis along with code:

1. Data Understanding and Preparation

Before we start the basic understanding of the data in hand, we imported relevant libraries available in Python. Below are the libraries that we imported

 # Importing Libraries

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

import warnings

warnings.filterwarnings('ignore')
```

We started with **Understanding the Data** in hand provided by running basic functions to load and interpret the variables, data types of the variables, dimensions, and size of the data frame. Below is the code used for the same.

 # Load the dataset

airbnbs=pd.read_csv('AB_NYC_2019.csv')

airbnbs.head()

 # Dimensions

airbnbs.shape

 # Data-Types

airbnbs.info

#Basic Data Interpretation:

● There are 16 columns and 48895 rows in the data frame.

● There are 3 floats, 7 integers, and 6 objects data type values in the data frame.

● There seem to be many columns with missing values.

2. Variables in the data frame:

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

The above understandings lead us to perform basic **Numeric and Categorical analysis** in depth by using the following function in Python

airbnbs.describe()

#Analyzing categorical values

airbnbs.select_dtypes(include=[‘object’]).describe()

3. Handling Missing Values and Outliers:

● Then we moved to handle missing values and outliers in the data frame. Starting with the missing values, we identified two columns having an equal percentage of missing values which were last_review and reviews_per_month of around 20.56%. And also, the other two columns had quite minimal missing values which were host_name of 0.4% and name of the place of 0.3%.Imputed the missing values of reviews_per_month with a 0.

 # Checking missing values percentages

round(100*(airbnbs.isnull().sum()/len(airbnbs.index)),2).sort_values(ascending = False)

 # Extracting Numeric and categorical columns:

Obj_types = [i for i in airbnbs.select_dtypes(include=np.object).column

['name',
 'host_name',
 'neighbourhood_group',
 'neighbourhood',
 'room_type',
 'last_review']

Num_dtypes =[i for i in airbnbs.select_dtypes(include =np.number)]

['id',
 'host_id',
 'latitude',
 'longitude',
 'price',
 'minimum_nights',
 'number_of_reviews',
 'reviews_per_month',
 'calculated_host_listings_count', 'availability_365']

Plotting the spread of outliers:
We preferred capping (statistical)method for treating outliers .Following code s are given below

outlier treatment for price:

Q1 = airbnbs.price.quantile(0.10)

Q3 = airbnbs.price.quantile(0.90)

IQR = Q3 - Q1

airbnbs = airbnbs[(airbnb.price >= Q1–1.5*IQR) & (airbnbs.price <= Q3 + 1.5*IQR)]

outlier treatment for minimum_nights:

Q1 = airbnbs.minimum_nights.quantile(0.10)

Q3 = airbnbs.minimum_nights.quantile(0.90)

IQR = Q3 - Q1

airbnbs = airbnbs[(airbnbs.minimum_nights >= Q1–1.5*IQR) & (airbnbs.minimum_nights <= Q3 + 1.5*IQR)]

outlier treatment for minimum_nights:

Q1 = airbnbs.number_of_reviews.quantile(0.10)

Q3 = airbnbs.number_of_reviews.quantile(0.90)

IQR = Q3 - Q1

airbnbs = airbnbs[(airbnbs.number_of_reviews >= Q1–1.5*IQR) & (airbnbs.number_of_reviews <= Q3 + 1.5*IQR)]

outlier treatment for reviews_per_month:

Q1 = airbnbs.reviews_per_month.quantile(0.10)

Q3 = airbnbs.reviews_per_month.quantile(0.90)

IQR = Q3 - Q1

airbnbs = airbnbs[(airbnbs.reviews_per_month >= Q1–1.5*IQR) & (airbnbs.reviews_per_month <= Q3 + 1.5*IQR)]

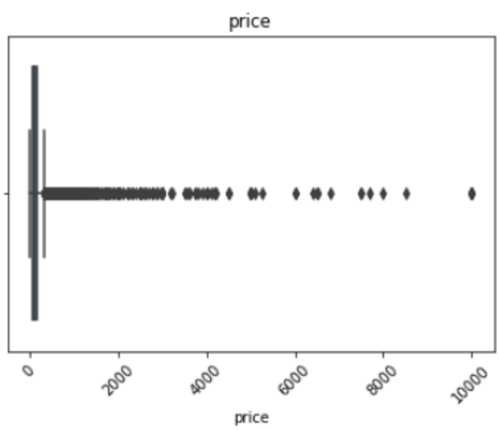
 # outlier treatment for calculated_host_listings_count:

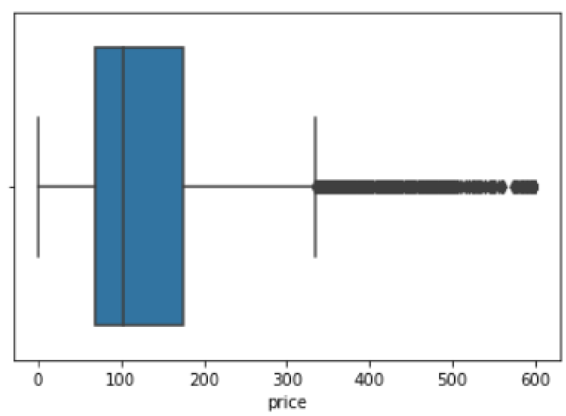
Q1 = airbnbs.calculated_host_listings_count.quantile(0.10)

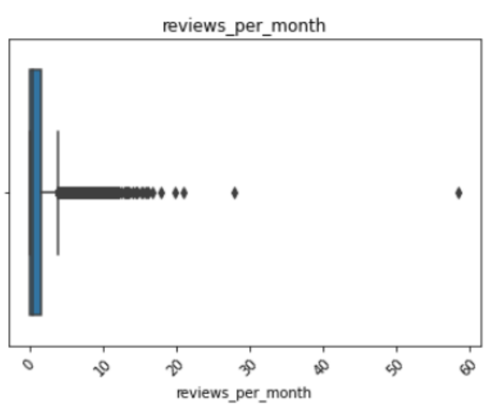
Q3 = airbnbs.calculated_host_listings_count.quantile(0.90)

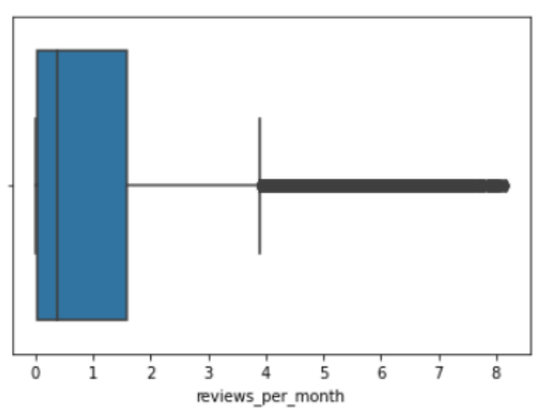
IQR = Q3 - Q1

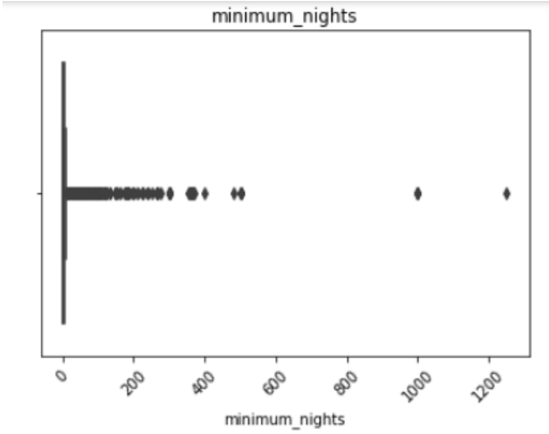
airbnbs = airbnbs[(airbnbs.calculated_host_listings_count >= Q1–1.5*IQR) & (airbnbs.calculated_host_listings_count <= Q3 + 1.5*IQR)]

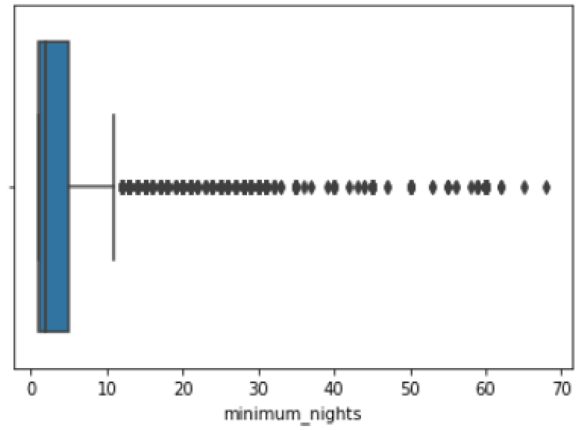


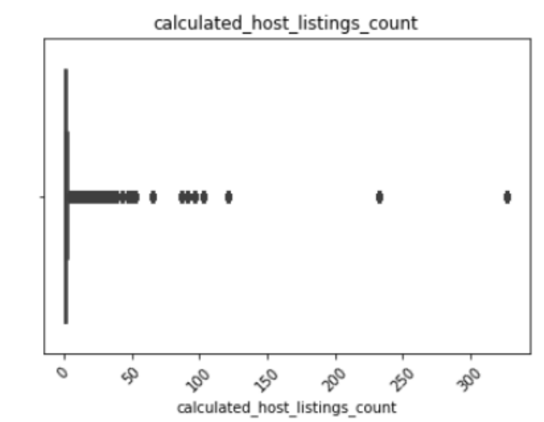


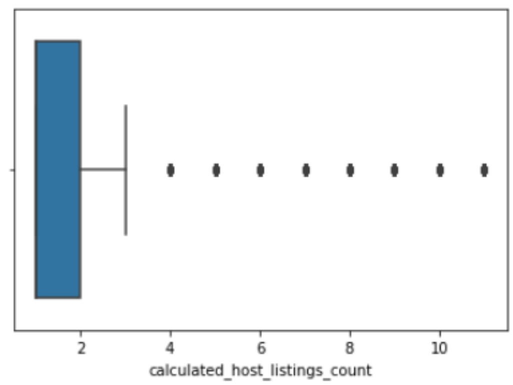




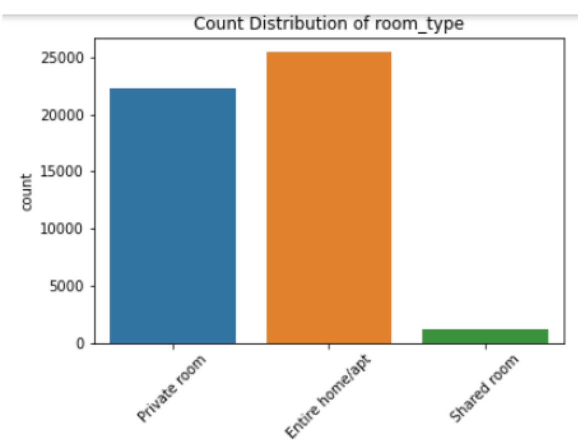


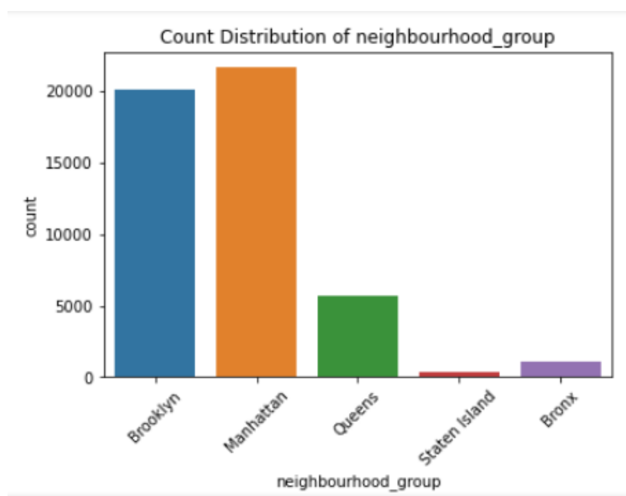






Categorical Univariate Analysis:





Parallel to we created all other Bivariate and Multivariate plots using **Tableau.**

5. Evaluating Questions

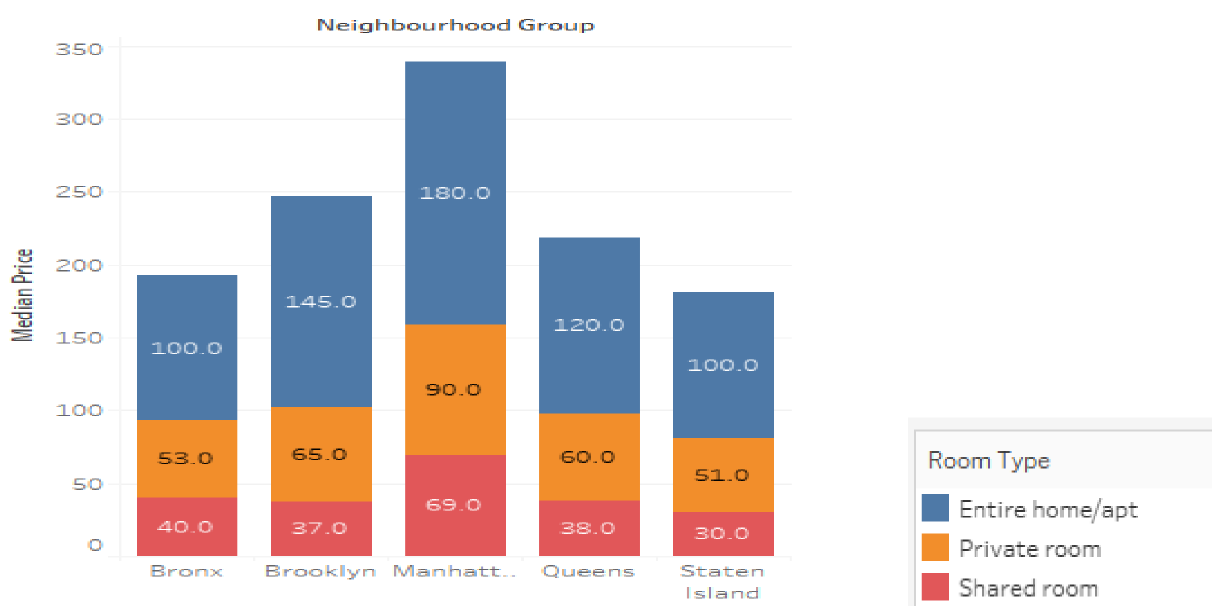
The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue such as -

1. Which type of hosts to acquire more and where?
2. The categorization of customers based on their preferences.
    - What are the neighborhoods they need to target?
    - What is the pricing ranges preferred by customers?
    - The various kinds of properties that exist w.r.t. customer preferences.
    - Adjustments in the existing properties to make it more customer-oriented.
3. What are the most popular localities and properties in New York currently?
4. How to get unpopular properties more traction?

Bivariate and Multivariate plots using Tableau.

Created a calculated field for Minimum night grouped

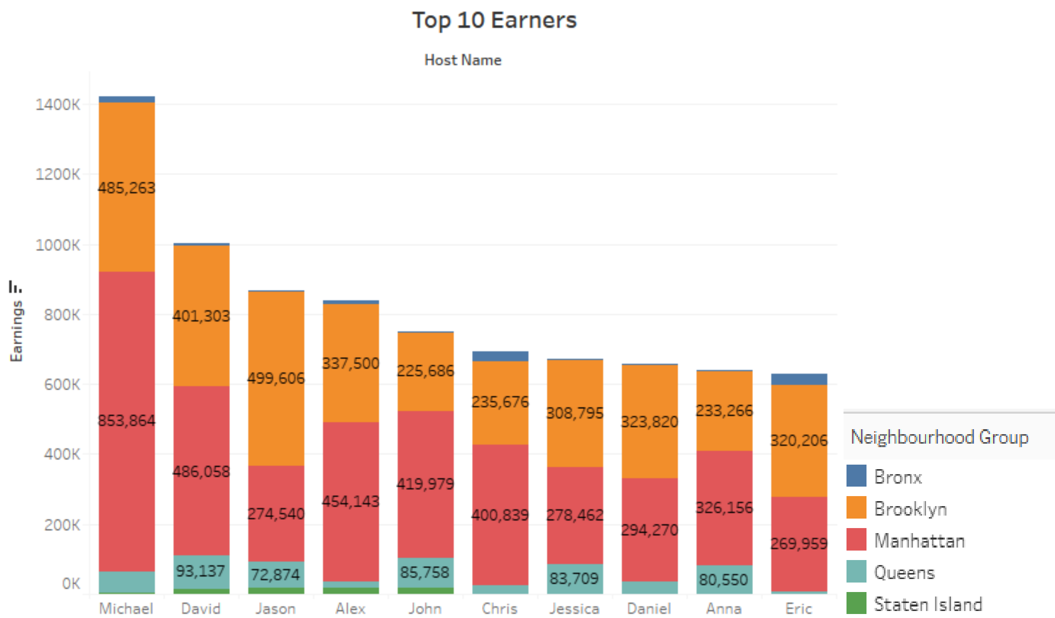
Created a calculated field for Earnings



- Entire Home apartments are popular in Manhatten

- Shared room is not much popular in neighborhood group accept Manhatten.
- Its clearly depicted from the graph Manhatten is costlier among all the neighborhood groups.

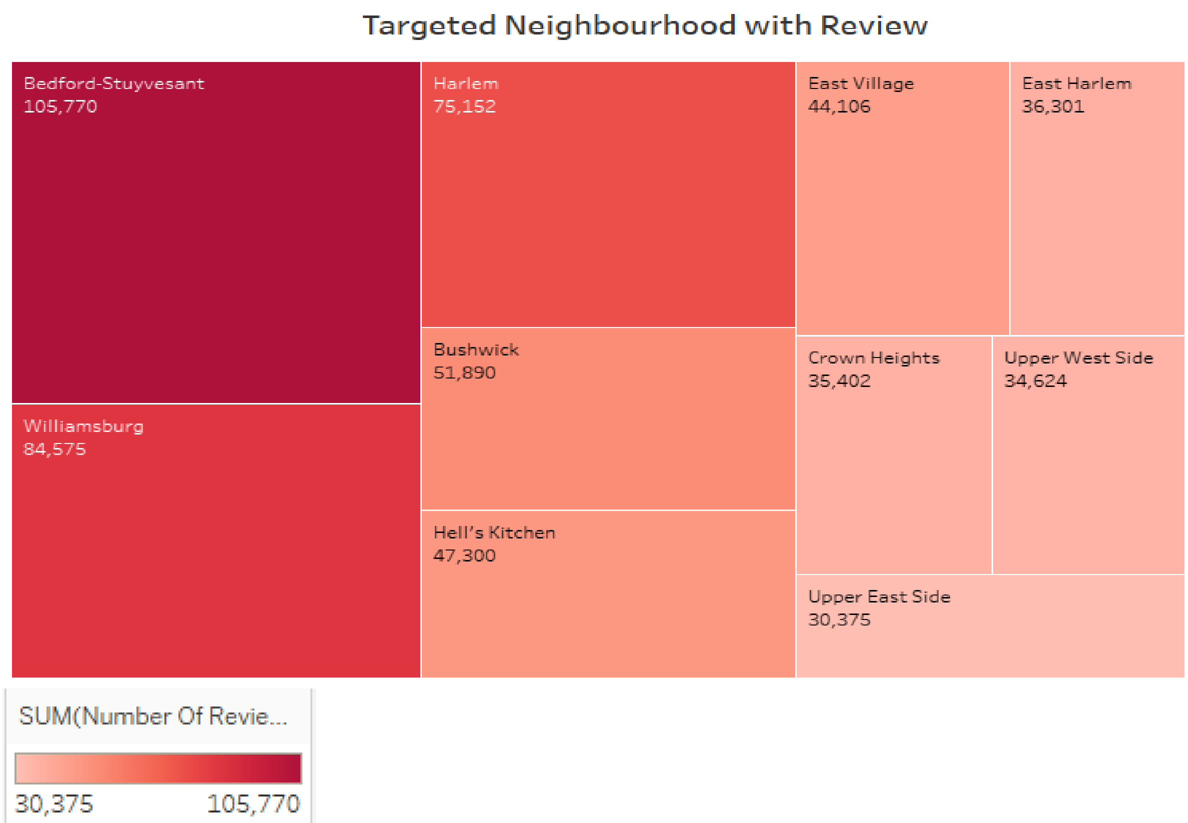
List of top 10 host to Acquire



- The graph depicts the top 10 host who are earnings more.
- Michael is top earner who is earning more and he belongs to Manhattan.

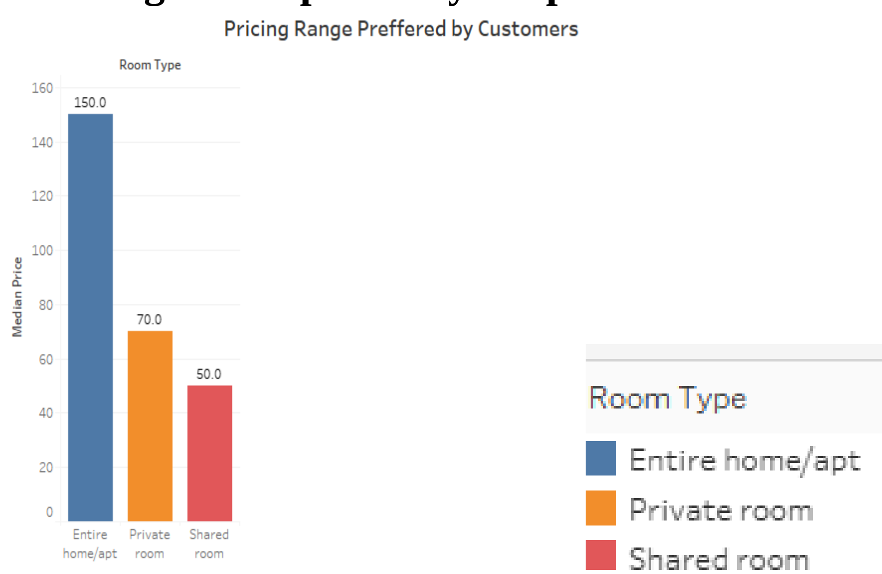
Targeted Neighborhood

- We can clearly comprehend that this location are most targeted location among all neighborhood, most of the people would prefer to go this neighbor location/area only.
- Review wise Bedford-Stuyvesant is best among all location/area.



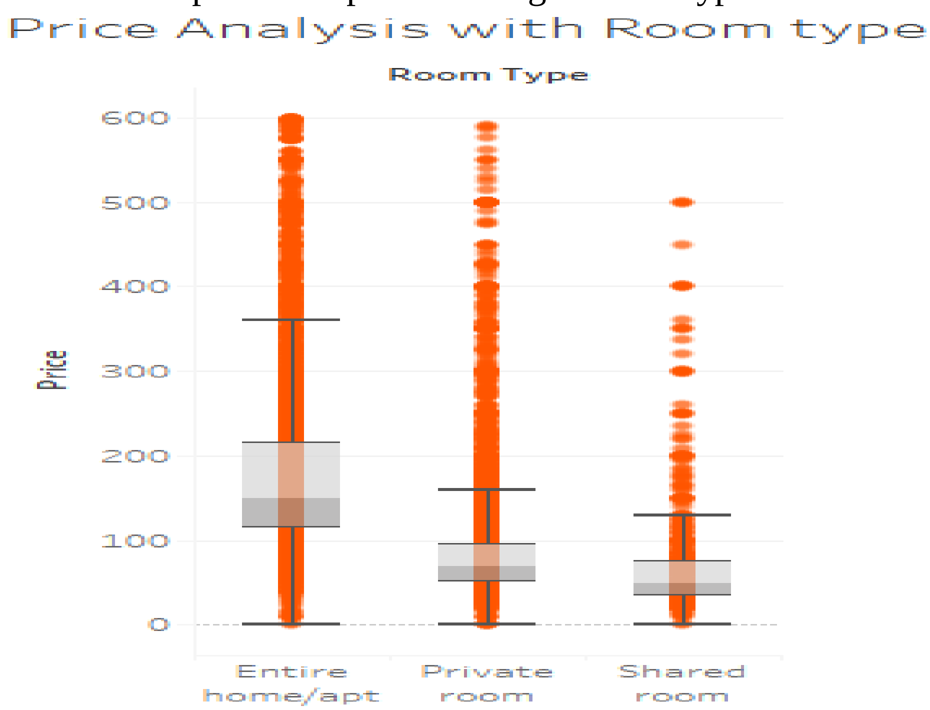
- We can clearly comprehend that this location are most targeted location among all neighborhood, most of the people would prefer to go this neighbor location/area only.
- Review wise Bedford-Stuyvesant is best among all location/area.

Average Price prefer by People

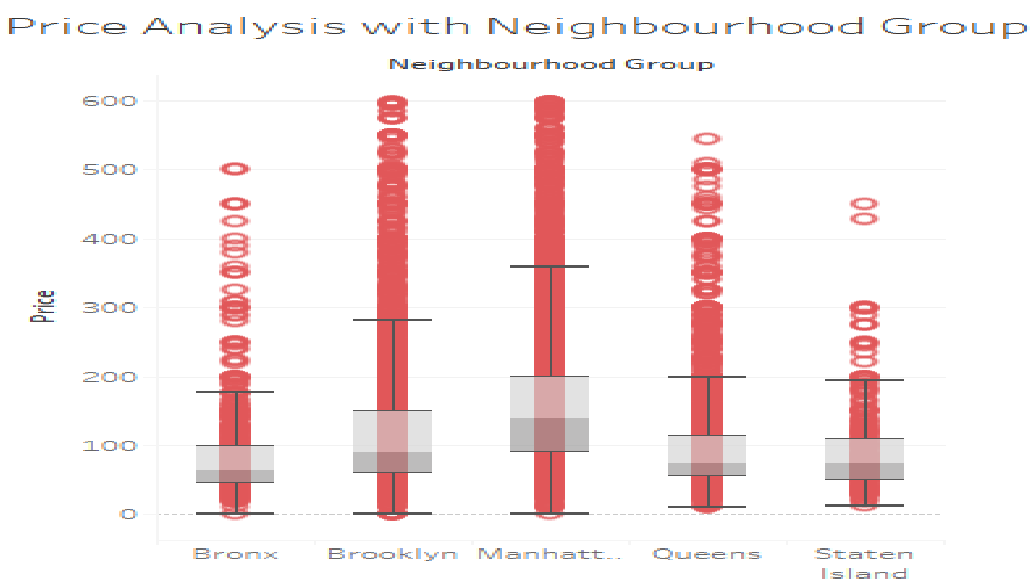


Clearly depicts from the graph on the basis of room type the average price preferred by customer for Entire Room is 160.

- For Private Room is 70
- Shared Room is 45
- Entire home/apt is most prefer among all three types.

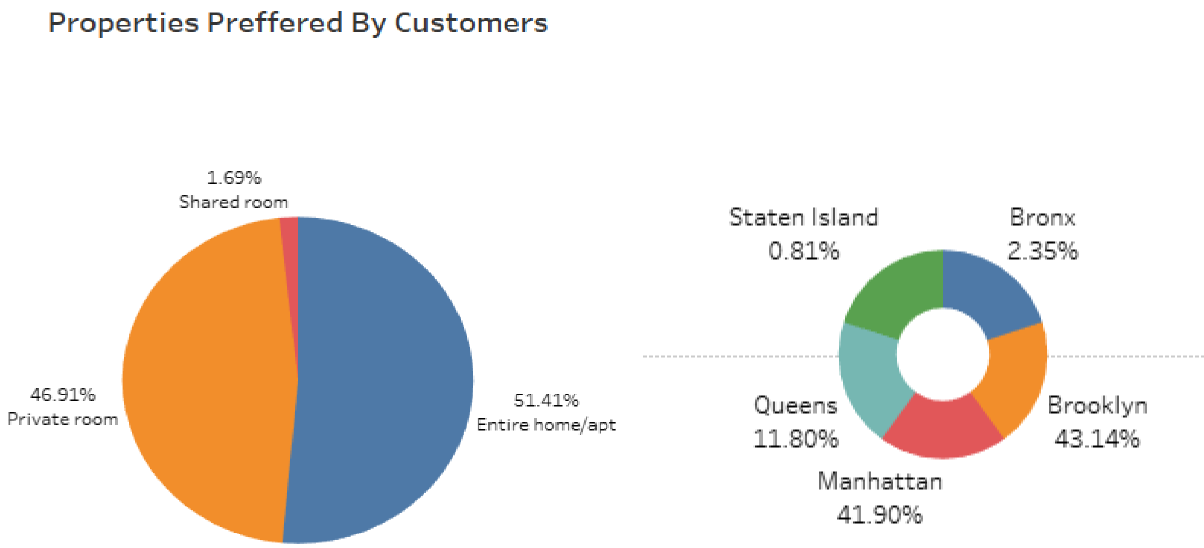


-



Graph depict that Manhattan is costlier then Brooklyn among all neighborhood groups

#Types of properties by customers preferences and total neighborhoods by neighborhood groups



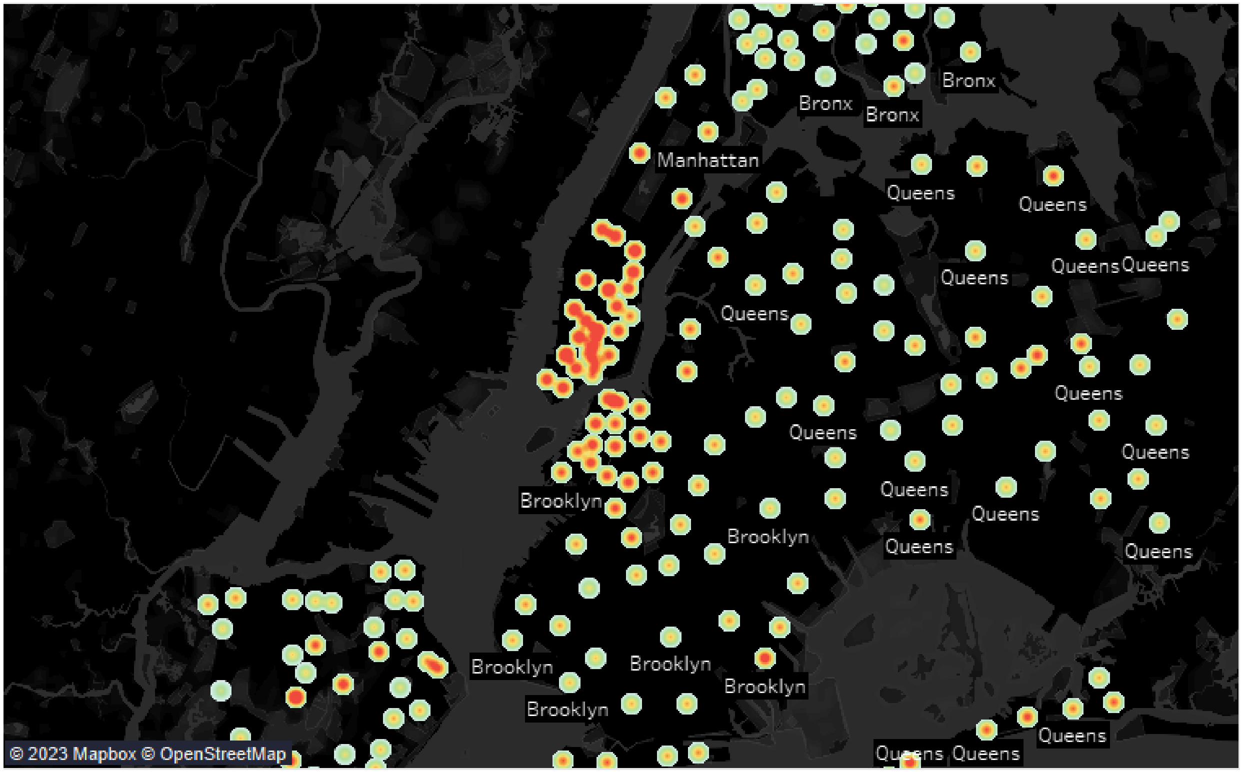
There are three types of rooms – Entire Home/Apartment, Private Room & Shared Room

• Overall customers appear to prefer Entire Home (51-41%) & Private Room (46.91%) in comparison to the shared room (1.69%).

• Airbnb can focus on promoting shared rooms with discount offers to increase booking of a shared room with discounts.

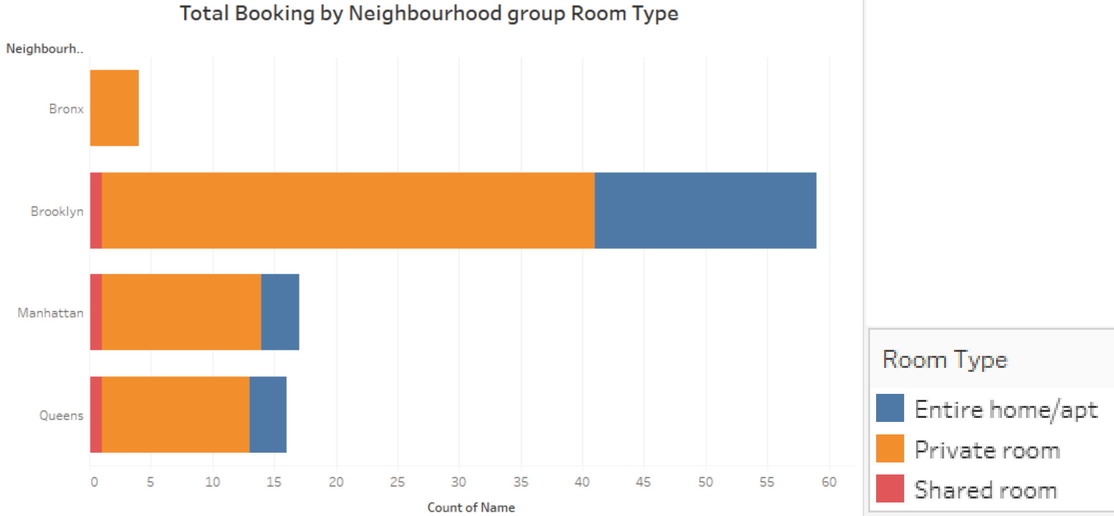
Most popular localities and properties in Newyork

Most Popular Properties Average Price in neighbourhood Room Type



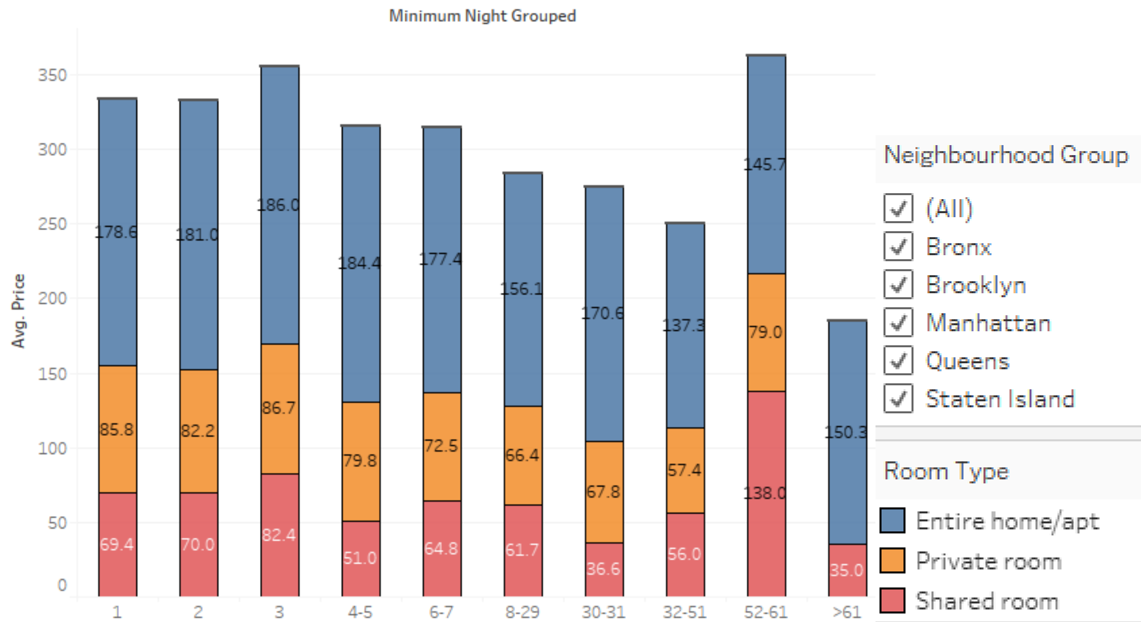
- According to this map more the darker side represents the most popular localities and the lighter side represents the least popular.
- We can conclude that Manhattan, Brooklyn & Queens are much popular than Bronx and Staten Island.

Total Booking by Neighbourhood group Room Type



Clearly depicts from the graph most popular neighborhood group for booking wise is Brooklyn.

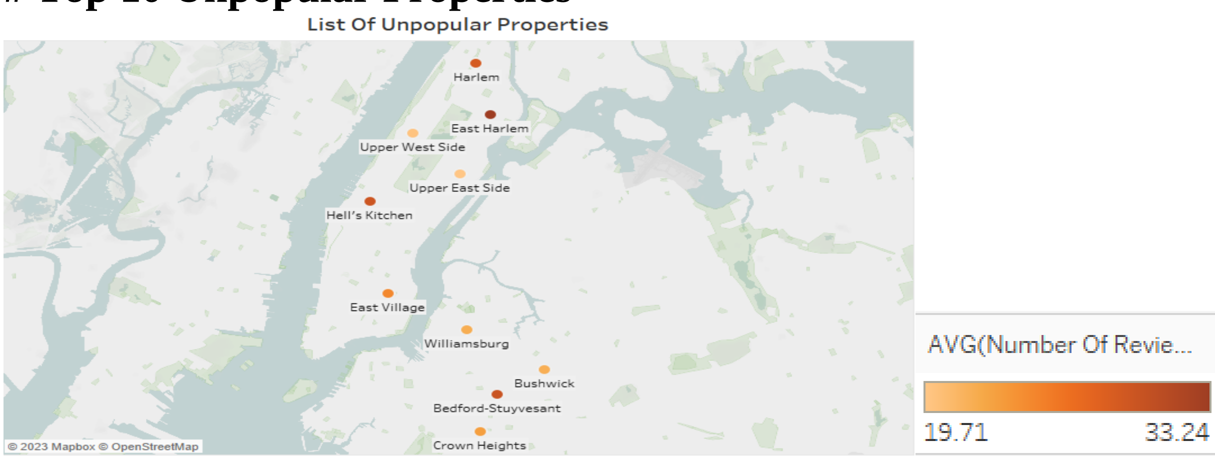
Here customers prefer private room among all three types.



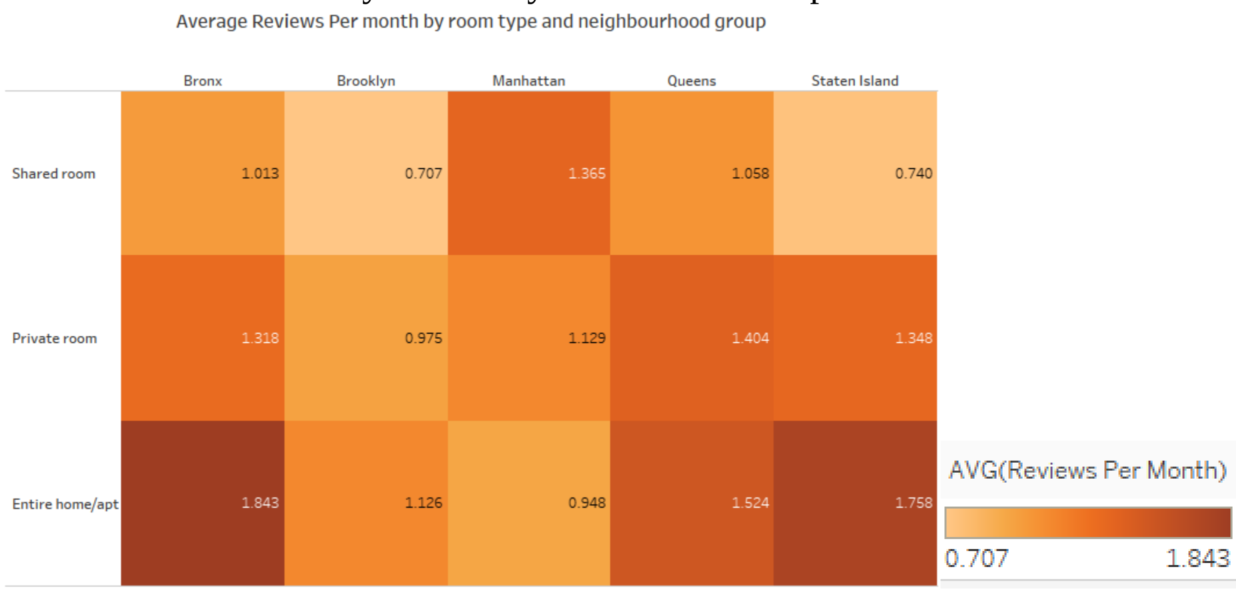
Graph clearly depicts that minimum stay night of 32-51 is comparatively cheaper cost in Entire home apartment among all the neighborhood group the others two.

Minimum 3 night is costlier in entire home apartment.

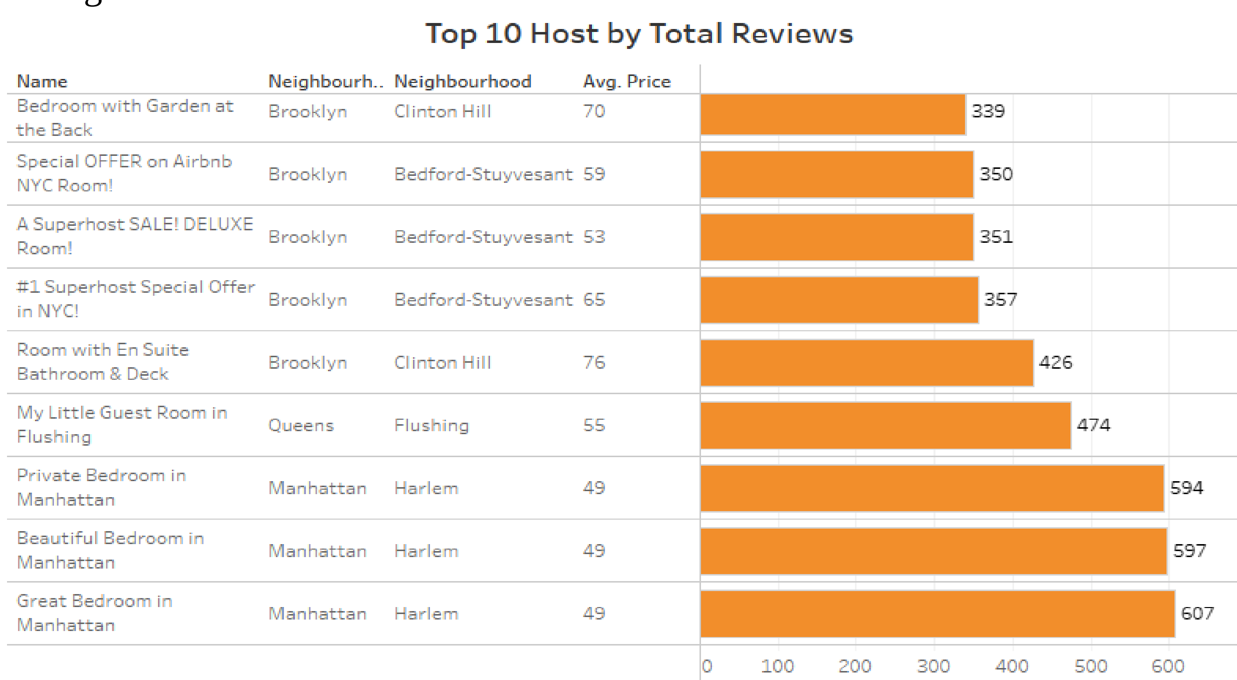
Top 10 Unpopular Properties



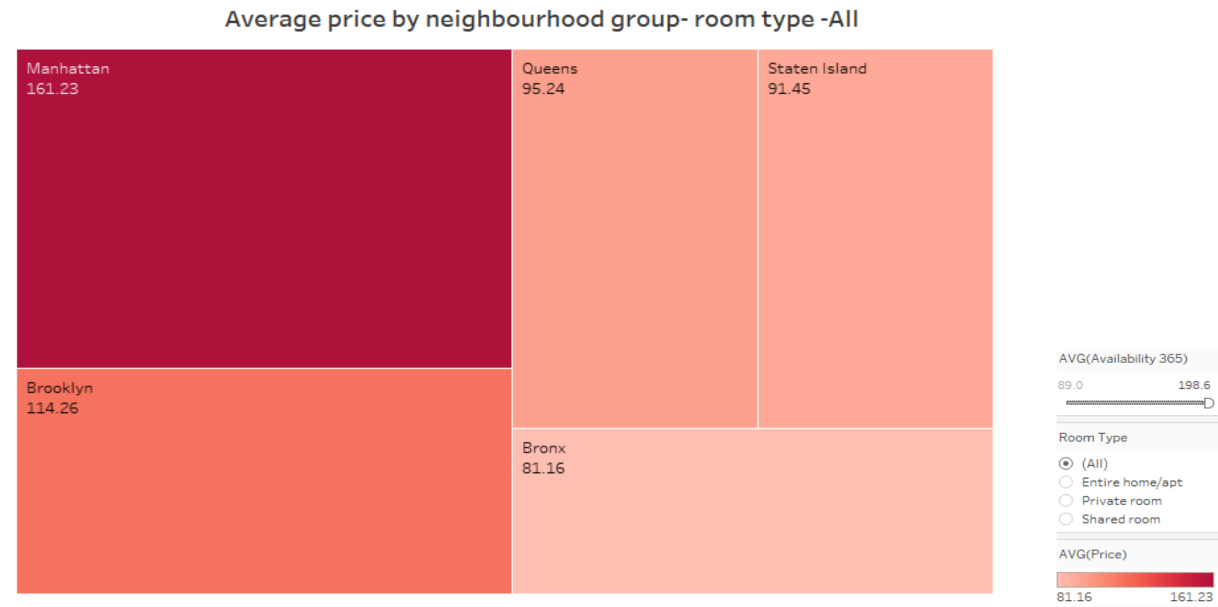
Graph clearly comprehend that light colour of city Upper Westside,Bushwick,Williamsburg etc are least popular city where people do not wish to visit or there may not be any tourist attraction point.

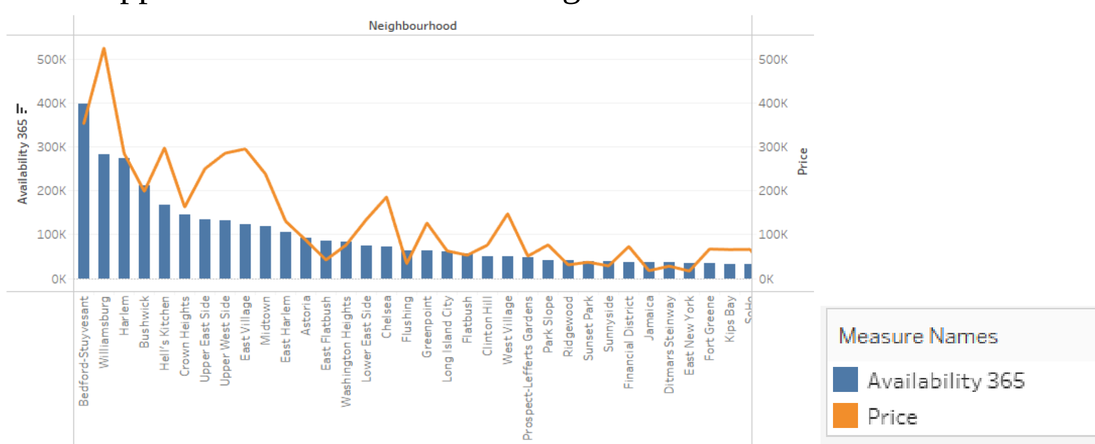


Graph depicts that Bronx has highest review per month for entire home apartment among all others

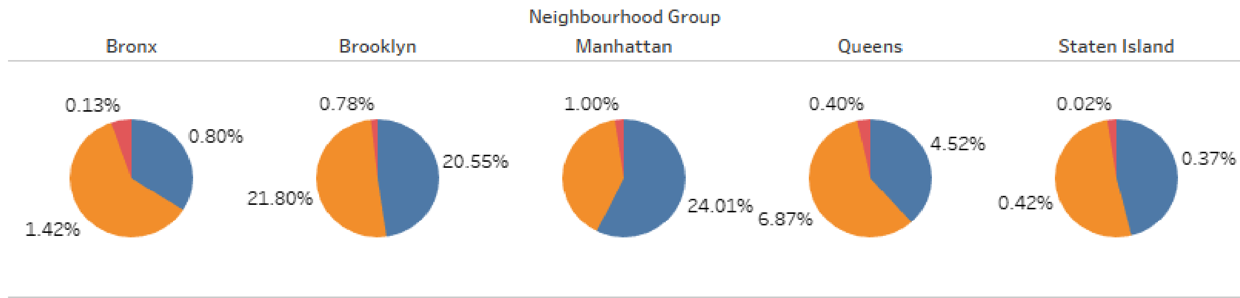


Graph clearly depict that Great Bed Host in Mahanttan having highest rating among Others.

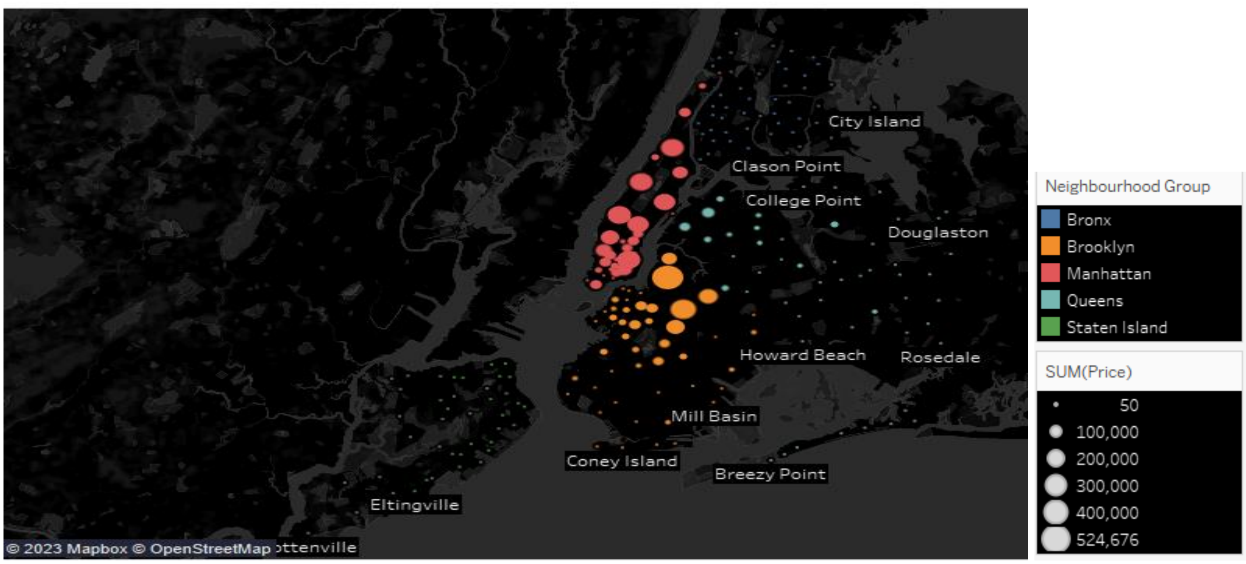


- The average price of listed properties in Manhatten is 161.23 which is highest among all neighborhoods.
- Average price for Brookklyn is second highest i.e 114.26.
- Bronx appears to be an affordable neighborhood.
- 

- Availability on Bedford is highest and its price is on lower side.it is a good choice for customers
- After Bedfords, Harlem follow the same trend.
- On the other hand, William's price is high and has average availability.



- Manhattan and Brooklyn are top neighbourhood groups and mostly people prefer to book the entire home or private room.
- Manhattan has highest number of home/apt properties.



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

Adjustments in the existing properties to make it more customer-oriented

• With the exception of Manhattan and Brooklyn, every other city needs to alter its marketing plan to boost sales.

• Most customers prefer to invest their money in the $40 to $160 range. Try a fresh marketing tactic to draw customers, such as offering deals and reductions.

• Every unpopular locality needs to alter their current plan in order to increase revenue, such as by creating a tourism draw.

• Increase the customer's purchasing ability, etc.