Storytelling Case Study: Airbnb, NYC

By : Sanchit Kumar Behera, Potnuru Koushik, Nikitha R.

Objective:

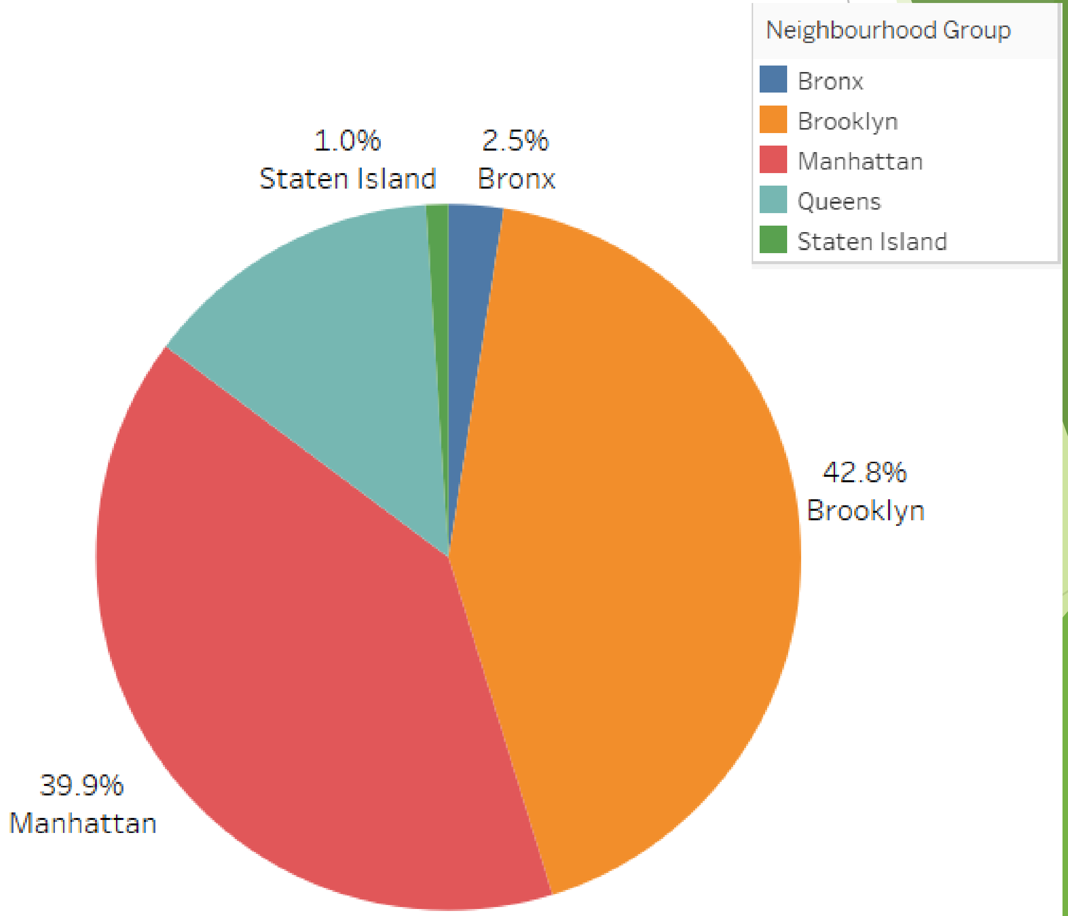
- Airbnb is an online platform using which people can rent their unused accommodations.
- During the covid time, Airbnb incurred a huge loss in revenue.
- People have now started travelling again and Airbnb is aiming to bring up the business again and e ready to provide services to customers.

Background

- 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.
- So, analysis has been done on a dataset consisting of various Airbnb listings in New York.

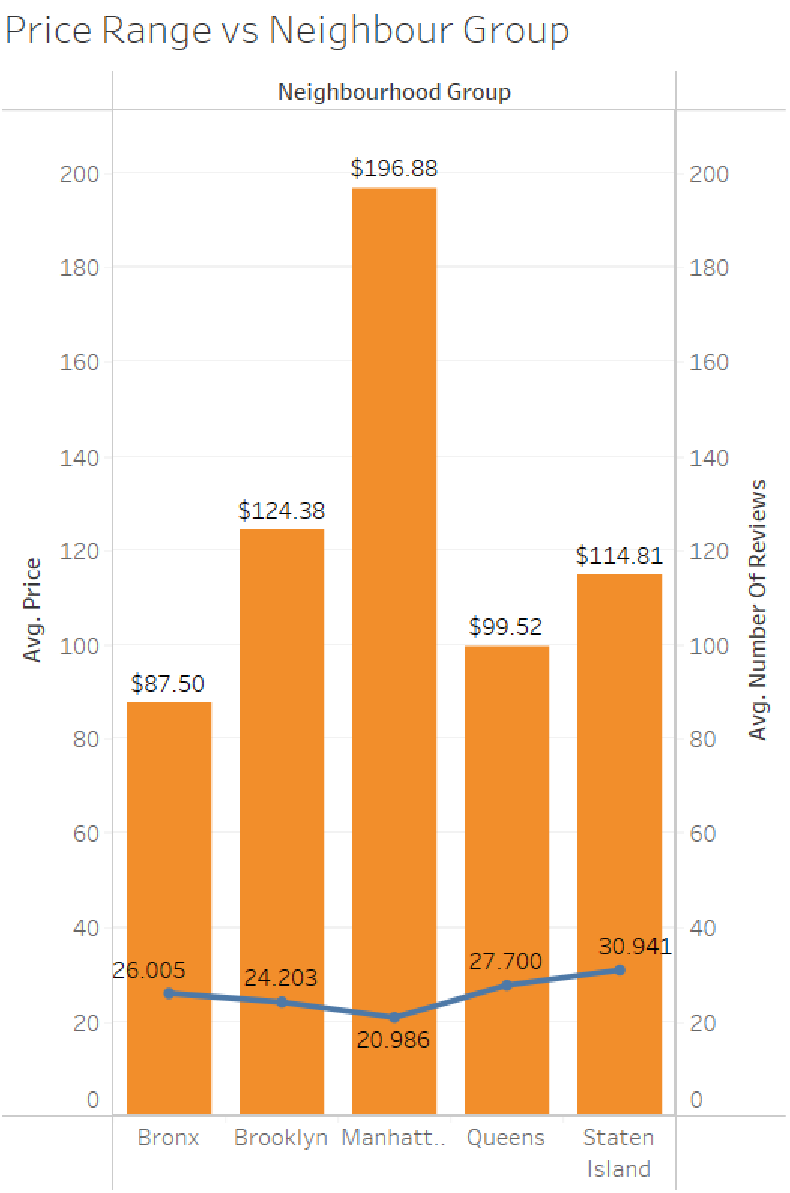
Neighborhood Group with respect to Booking

- Manhattan and Brooklyn constitute approx. 83% of all booking.
- Queens has the third highest share in booking approx. 14%.
- Least booking in Staten Island and Bronx



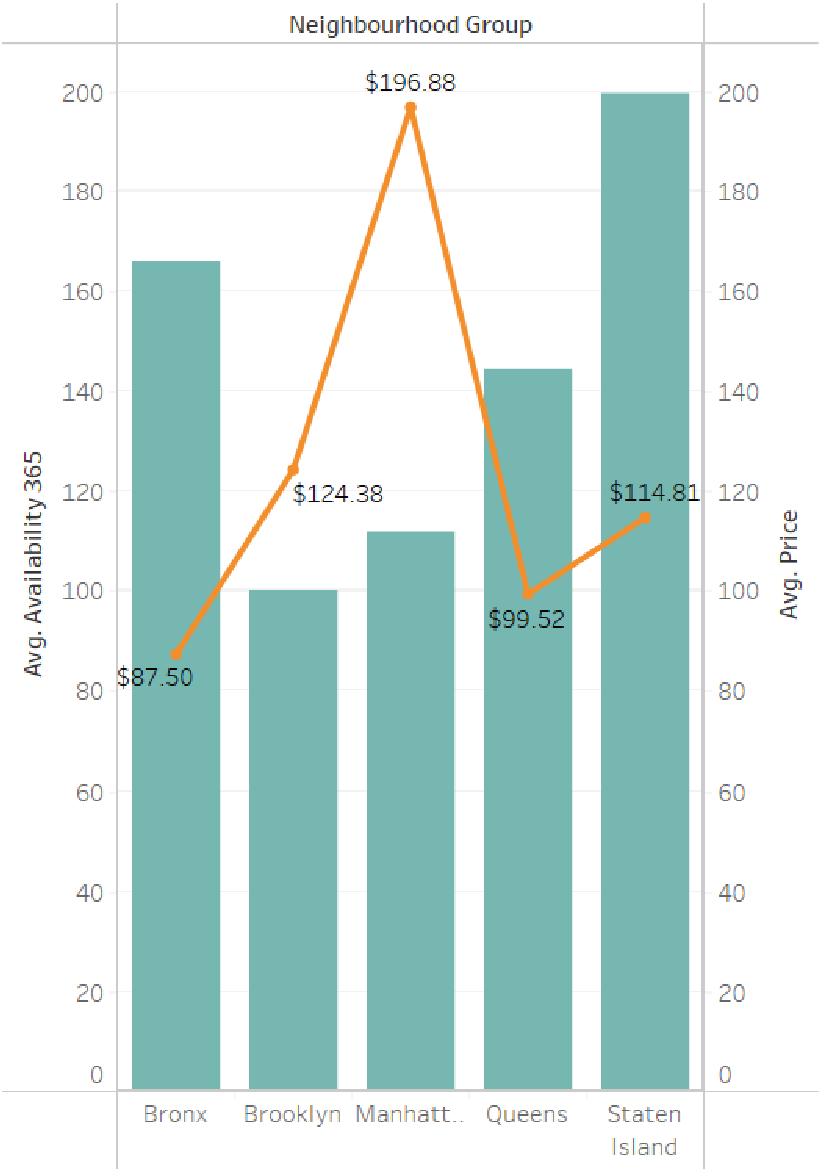
Price Range with Neighborhood Group

- Brooklyn and Manhattan has property of higher price.
- The average booking price stays around 100k – 120k with least in Bronx.



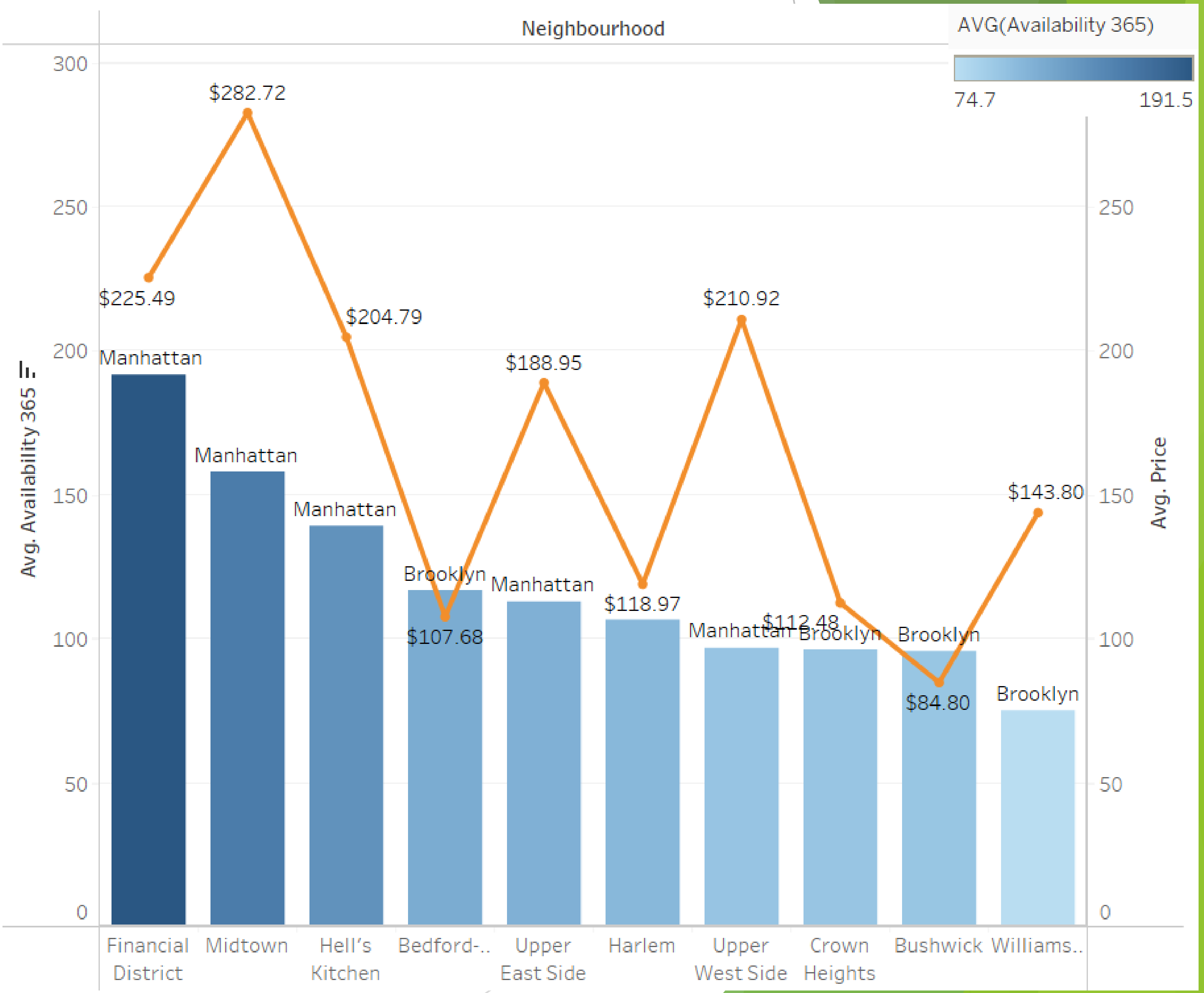
Neighborhood Group and Availability

- In terms of availability Staten Island and Bronx in an average has a higher value.
- Brooklyn and Manhattan being booked for longer duration show higher popularity.
- The comparison with respect to Price and Availability 365 shows even with lower prices in Bronx and Staten Island it remains available for a longer period.



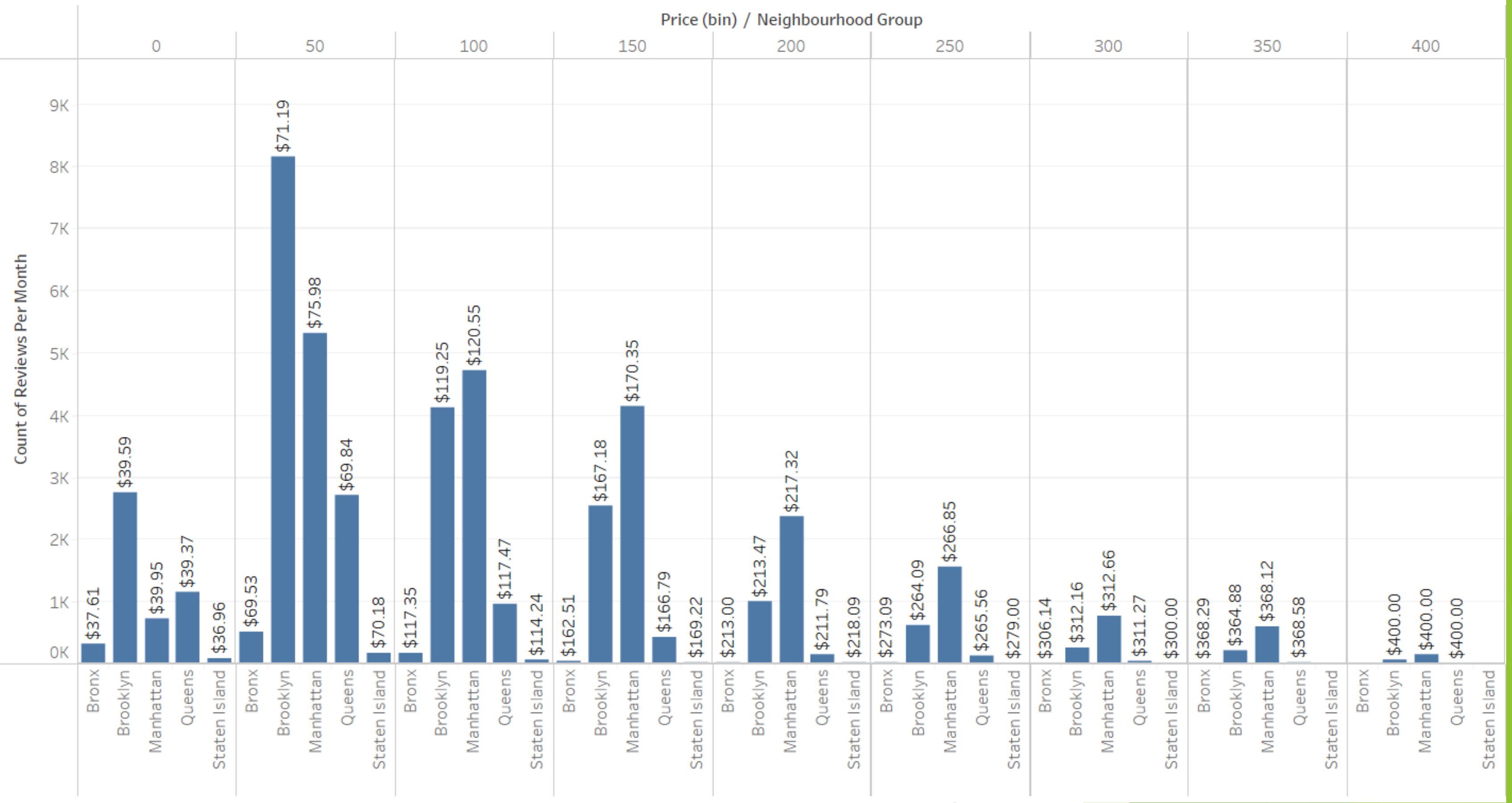
Neighborhood with respect to Availability and Price

- With respect to Neighborhood Financial District is most available and Williamsburg at the least.
- Manhattan is among the top area where in an average rooms are available year round.
- Midtown’s availability is lower but is on the pricier side.



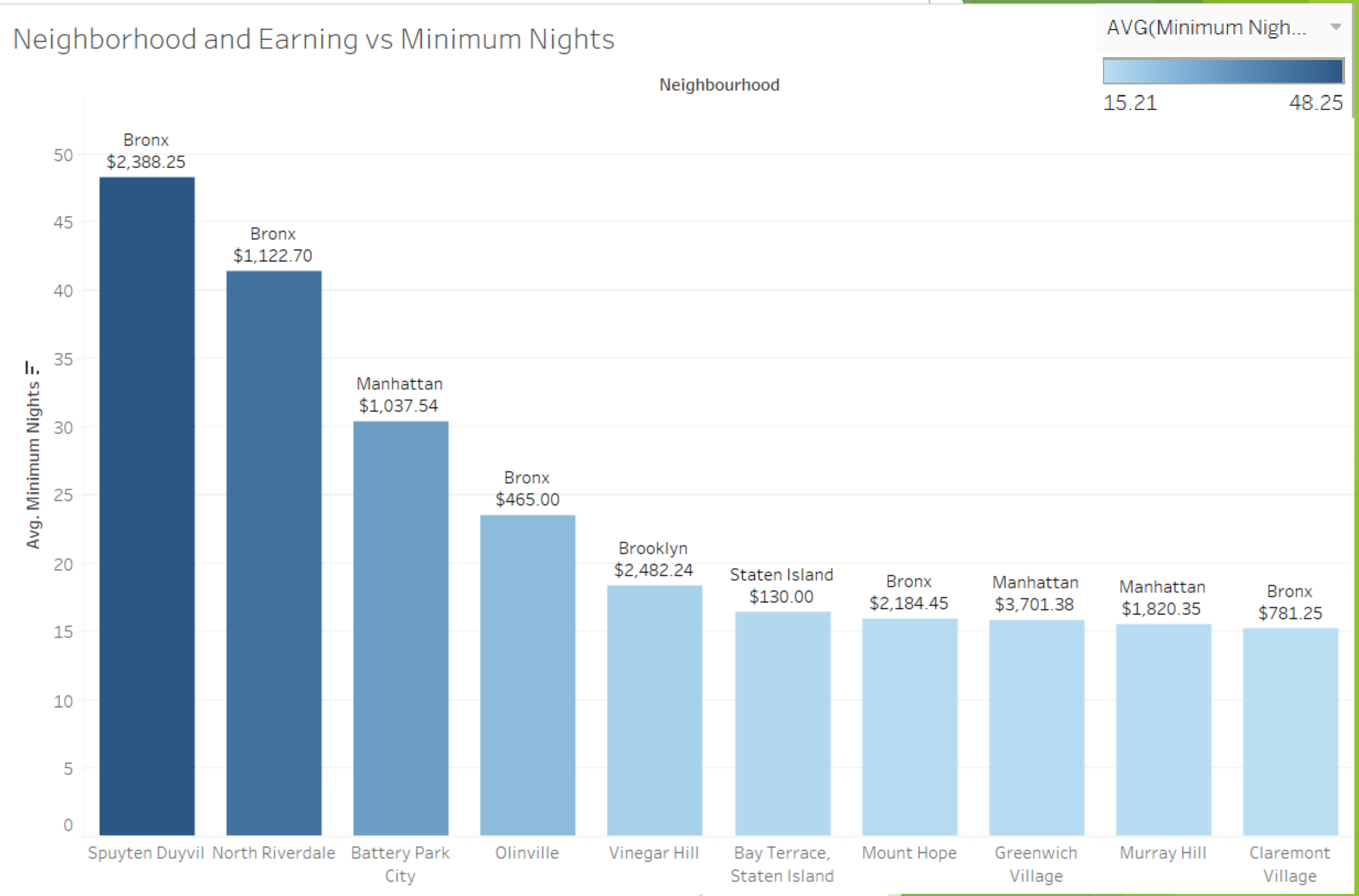
Price Binning with Neighborhood Group

- Most of the booking are done from the range 30$ to 170$ in average.
- In the lower Bin range 0-100 Brooklyn has higher average price range.
- Whereas in the higher Bin range of 100$ - 200$, Manhattan has higher average pricing range.



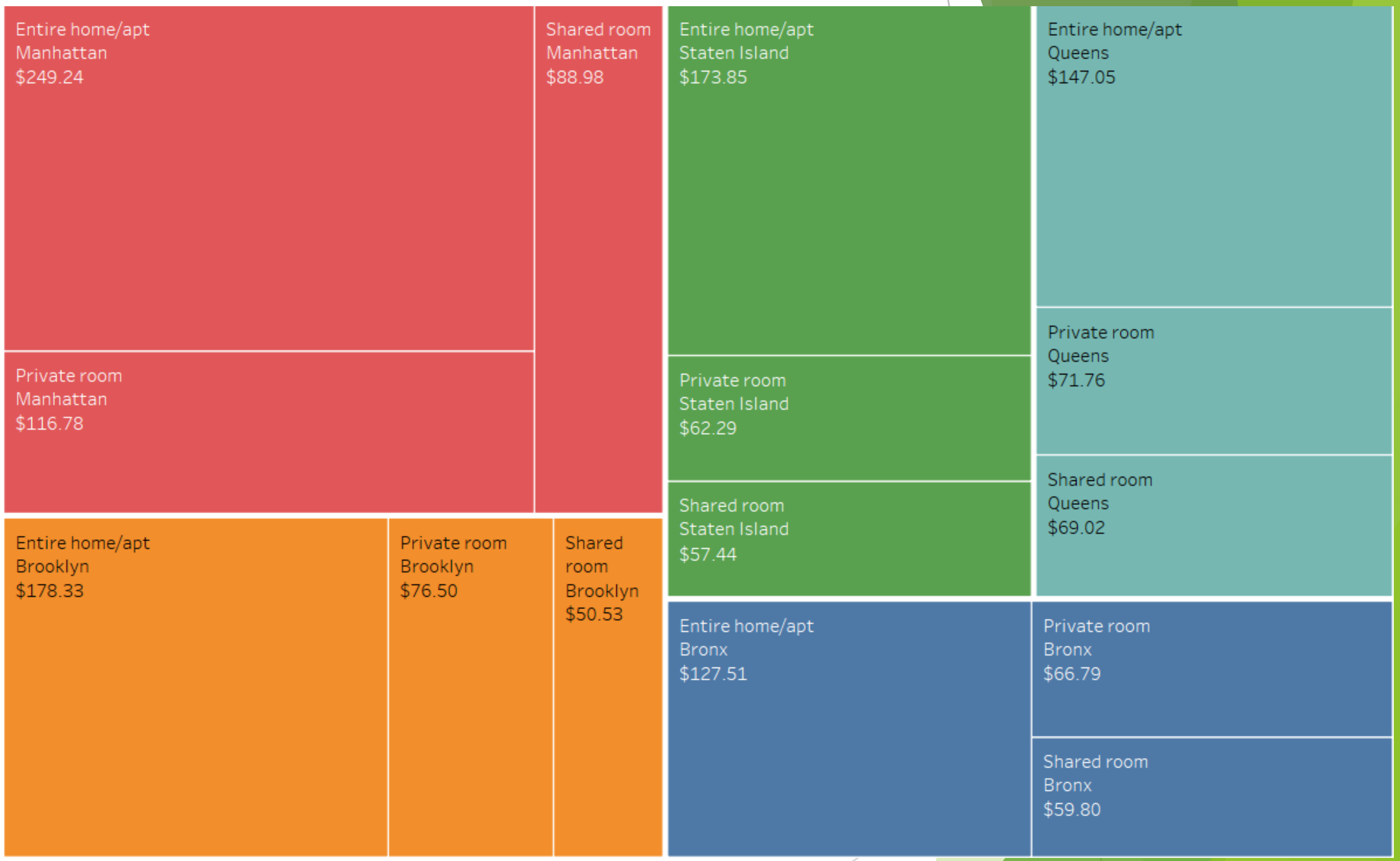
Minimum Nights vs Neighborhood and Earning

- Properties in Spuyten Duyvil has been occupied for a longer period of around 50 days
- Whereas in Claremont Village in an average a property is booked only for 15-16 days.
- The top 3 property booked for longer period are Sputyen Duyvil, North Riverdale and Battery Park City.



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

- The 'Entire home/apt' room type in Manhattan is the most expensive at $250, much higher than the overall average.
- ‘private rooms’ of Manhattan & Brooklyn has the highest average.
- ‘Shared Room' type is the cheapest in Brooklyn with $50.5.



Conclusion

- With respect to neighborhood group there is business is scalable in the Manhattan, Brooklyn and Queens area.
- Staten Island and Bronx have more availability annually, so different campaign can be run to increase it’s booking. Staten Island even has higher price in an average than Bronx and Queens.
- The top 3 property booked for longer period are Sputyen Duyvil, North Riverdale and Battery Park City.
- With respect to neighborhood Bronx had a higher average earning as compared to others.

Appendix : Methodology

AIRBNB Case Study

Methodology Document PPT 1:

In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

Initial Analysis using Jupiter Notebook: Data Set Used: AB_NYC_2019.csv

Number of Rows: 48895

Number of Columns: 16

```
# Import the necessary libraries
import warnings
warnings.filterwarnings("ignore")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
# Data conversion and Understanding
airbnb = pd.read_csv("AB_NYC_2019.csv")
airbnb.head(5)
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revi
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

```
# Check the rows and columns of the dataset
airbnb.shape
```

```
(48895, 16)
```

- The dataset contains 48895 rows and 16 columns
- Now we have to check whether there are any missing values in the dataset

```
# Calculating the missing values in the dataset
airbnb.isnull().sum()
```

```
id                                    0
name                                 16
host_id                               0
host_name                            21
neighbourhood_group                   0
neighbourhood                         0
latitude                              0
longitude                             0
room_type                             0
price                                 0
minimum_nights                        0
number_of_reviews                     0
last_review                       10052
reviews_per_month                 10052
calculated_host_listings_count        0
availability_365                      0
dtype: int64
```

```
# Now we have the missing values, there are certain columns that are not efficient to the dataset
airbnb.drop(['id','name','last_review'], axis = 1, inplace = True)
```

```
# View whether the columns are dropped
airbnb.head(5)
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revie
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

Methodology- Contd.

Step 2: Data Wrangling:

- Checked the Duplicate rows in our dataset and no duplicate data was found.
- Checked the Null Values in our dataset. Columns like name, host-name, last review and review-per-month have null values.
- We've dropped the column name as missing values are less and dropping it won't have significant impact on analysis.
- Checked the formatting in our dataset.
- Identified and review outliers.

Data Analysis and Visualizations using Tableau:

We have used tableau to visualize the data for the assignment. Below are the detailed steps used for each visualization.

1) **Top 10 Host:**
   - We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map.

2) **Preferred Room type with respect to Neighbourhood group:**
   - We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
   - We added Room Type to the colours Marks card to highlight the different Room Type in different colours and count of Host Id to the size.

3) **For Variance of price with Neighbourhood Groups:**
   - We used a box and whisker's plot with Neighbourhood Groups in Columns and Price in Rows.
   - We changed the Price from a Sum Measure to the median measure.

4) **Average price of Neighbourhood groups:**
   • We created a bubble chart with Neighbourhood Groups in Columns and Price column in Rows.
   •We added the Neighbourhood Groups to the colors Marks card to highlight the different neighbourhood Groups in different colors. Also Put Avg price in Label.

5) **Customer Booking w r t minimum nights:**
   - The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.

6) **Popular Neighborhoods:**
   - We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour.
   - We used filter to show Top 20 neighbours as per the sum of reviews.

7) **Neighbourhood vs Availability:**
   - We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

8) **Price range preferred by Customers:**
   - We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of $20.

9) **Understanding Price variation w.r.t Room Type & Neighbourhood:**
   - We created Highlights Table chat by taking Room Type in rows & Neighbourhood Group in column.
   - We took the average price in colour Marks card to highlight the different Room Type in different colours.

10) **Price variation w r t Geography:**
   - We used Geo location chart to plot neighbourhood, neighbourhood Group in map to show case the variation of prices across.

11) **Popular Neighborhoods:**
   - We took neighborhood in rows and sum of reviews in column and took neighborhood groups in color.
   - We used filter to show Top 20 neighbors as per the sum of reviews.

12) **Tools used:**
   - Data cleaning and preparation: Jupyter notebook – Python
   - Visualization and analysis: Tableau
   - Data Storytelling: Microsoft PPT