Airbnb NYC 2019
Case Study – Methodology

Authored by: Chinmay Kumar Sahu

Sahil Panigrahi

Pranay Chide

Introduction

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.

Airbnb was founded in 2008 by Brian Chesky, Joe Gebbia, and Nathan Blecharczyk. The idea for Airbnb came about when the three founders struggled to pay their rent in San Francisco and decided to rent out air mattresses in their apartment to attendees of a design conference in the city.

The idea proved to be successful, and the founders recognized the potential of a 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 of New York City Listings. 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.

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

Assumption:

Upon checking the data there was no direct relation found to customer satisfaction. Hence, we have taken number of reviews as the measure of liking of customer towards listings of properties.

Data Source:

Provided by Upgrad website.

Name: *AB_NYC_2019.csv*

Type: CSV

Column	Description
Id	Listing ID
Name	Name of Listing
Host_id	host ID
Host_name	Name of Host
Neighbourhood Group	Neighbourhood_group - Location
Neighborhood	Neighborhood - 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

Methodology:

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

Objective for Presentation -1:

The methodology was done while focus mainly on the following points:

1. The relation between various parameter given in the dataset
2. Performing Univariate & Bivariate analysis on the given parameters to get key insights.
3. Suggestion & Recommendation to improve the customer traction ultimately benefiting the Airbnb by increasing revenue.

Key Points for Presentation -1:

1. The data was analyzed through univariate and bivariate analysis.
2. Features were segmented in categorical, numerical & location variables.
3. Initially data set was prepared with the help of Python in Jupyter Notebook.
4. Then analysis and visualizations were done using Tableau 2023.1 Public considering various parameters.
5. The main parameters that have been taken into account for analysis are –
    - Geography based bookings
    - Neighbour Hood Groups
    - Bookings based on room type
    - Number of reviews
    - Number of nights
    - Price
    - Availability of listings
6. Inferences have been made keeping in mind the above parameters

Objective for Presentation -2:

The methodology was done while 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 traction.

Key Points for Presentation -:

- The analysis and visualizations were done using Tableau 2023.1 considering various parameters given in the data set.
- The key categorical features extensively sued are Neighborhood Group & Room type
- The emphasize the nature of analysis towards this particular group of reviewers following parameters were considered –
    1. Customer experience: Neighborhood, Room type & minimum nights offered
    2. Price variation: Price Distribution, Room type, Neighborhood Group, Number of reviews & Geography.

- The initial analysis done while keeping key parameters in mind with respect to Count of listings in those parameters.
- Furthermore, numerical features such as availability & minimum number of nights were subjected to univariate analysis with key categorical features
- Lastly the analyzing the customer reviews with the neighborhood group
- Recommendations have been made keeping in mind the above parameters.

STEP-1: Initial Analysis using Jupiter Notebook: Data Set Used: *AB_NYC_2019.csv*

Importing Necessary libraries for EDA

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing warning library

```
import warnings
warnings.filterwarnings(action="ignore")
```

```
df1 = pd.read_csv('AB_NYC_2019.csv')  # Import Dataset to notebook
```

```
pd.set_option("display.max_columns",200) #To make all columns visibile durign EDA
```

```
df1.info(verbose =True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   id                              48895 non-null  int64  
 1   name                            48879 non-null  object 
 2   host_id                         48895 non-null  int64  
 3   host_name                       48874 non-null  object 
 4   neighbourhood_group             48895 non-null  object 
 5   neighbourhood                   48895 non-null  object 
 6   latitude                        48895 non-null  float64
 7   longitude                       48895 non-null  float64
 8   room_type                       48895 non-null  object 
 9   price                           48895 non-null  int64  
 10  minimum_nights                  48895 non-null  int64  
 11  number_of_reviews               48895 non-null  int64  
 12  last_review                     38843 non-null  object 
 13  reviews_per_month               38843 non-null  float64
 14  calculated_host_listings_count  48895 non-null  int64  
 15  availability_365                48895 non-null  int64  
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

```
print(df1.columns.values)
```

```
['id' 'name' 'host_id' 'host_name' 'neighbourhood_group' 'neighbourhood'
 'latitude' 'longitude' 'room_type' 'price' 'minimum_nights'
 'number_of_reviews' 'last_review' 'reviews_per_month'
 'calculated_host_listings_count' 'availability_365']
```

DATA UNDERSTANDING

```
df1.shape
```

```
(48895, 16)
```

```
df1.head()
```

	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	

```
df1_null = 100*df1.isnull().sum()/len(df1)
df1_null
```

```
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                 0.000000
calculated_host_listings_count    0.000000
availability_365                  0.000000
dtype: float64
```

```
# Now there are certain columns that are not efficient to the dataset
df1.drop(['host_name','name','last_review'], axis = 1, inplace = True)
```

```
df1.info(verbose =True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 13 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   id                              48895 non-null  int64  
 1   host_id                         48895 non-null  int64  
 2   neighbourhood_group             48895 non-null  object 
 3   neighbourhood                   48895 non-null  object 
 4   latitude                        48895 non-null  float64
 5   longitude                       48895 non-null  float64
 6   room_type                       48895 non-null  object 
 7   price                           48895 non-null  int64  
 8   minimum_nights                  48895 non-null  int64  
 9   number_of_reviews               48895 non-null  int64  
 10  reviews_per_month               48895 non-null  float64
 11  calculated_host_listings_count  48895 non-null  int64  
 12  availability_365                48895 non-null  int64  
dtypes: float64(3), int64(7), object(3)
memory usage: 4.8+ MB
```

```
# Exporting the cleaned file to the local coumputer for visualization through tableau.
df1.to_csv(r'C:\Users\ASUS\Desktop\Upgrad\Course 1 - Data Toolkit\AirBNB Case_Study\EDA_AIRBNB_NYC.csv',index=False, header=True)
```

Exported the prepared dataset under the name of *EDA_AIRBNB_NYC.csv* to local machine for further analysis in tableau public

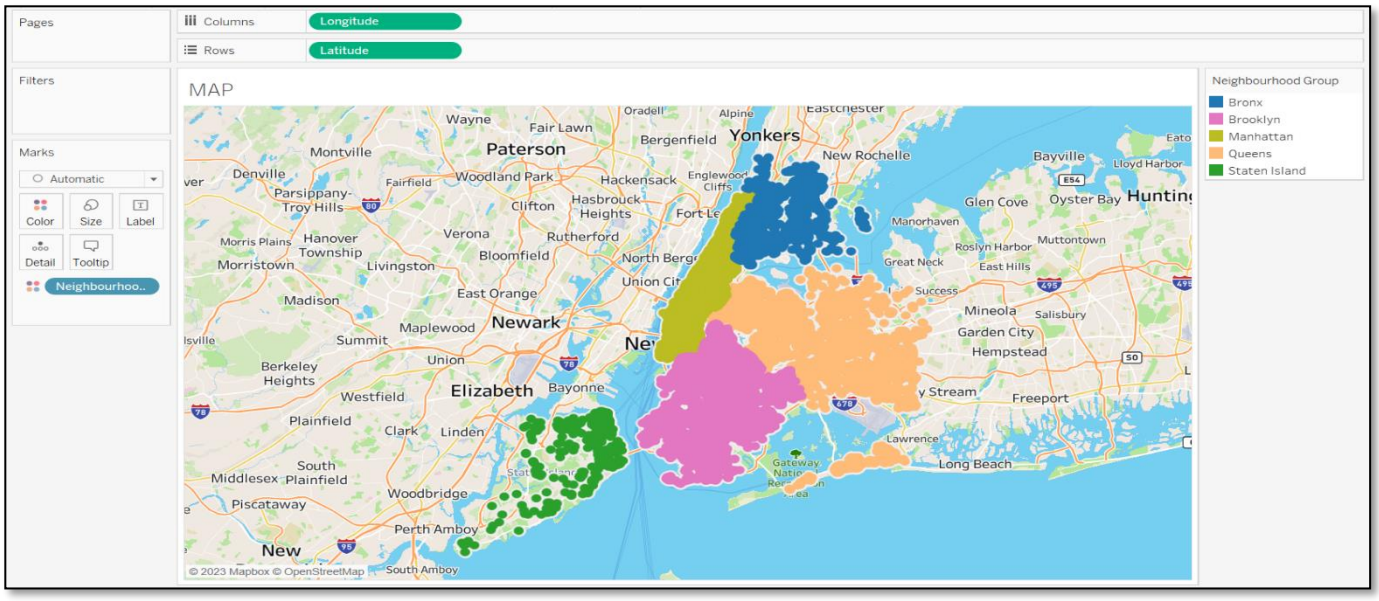
STEP-2: Imported the new dataset to Tableau Public 2023.1.

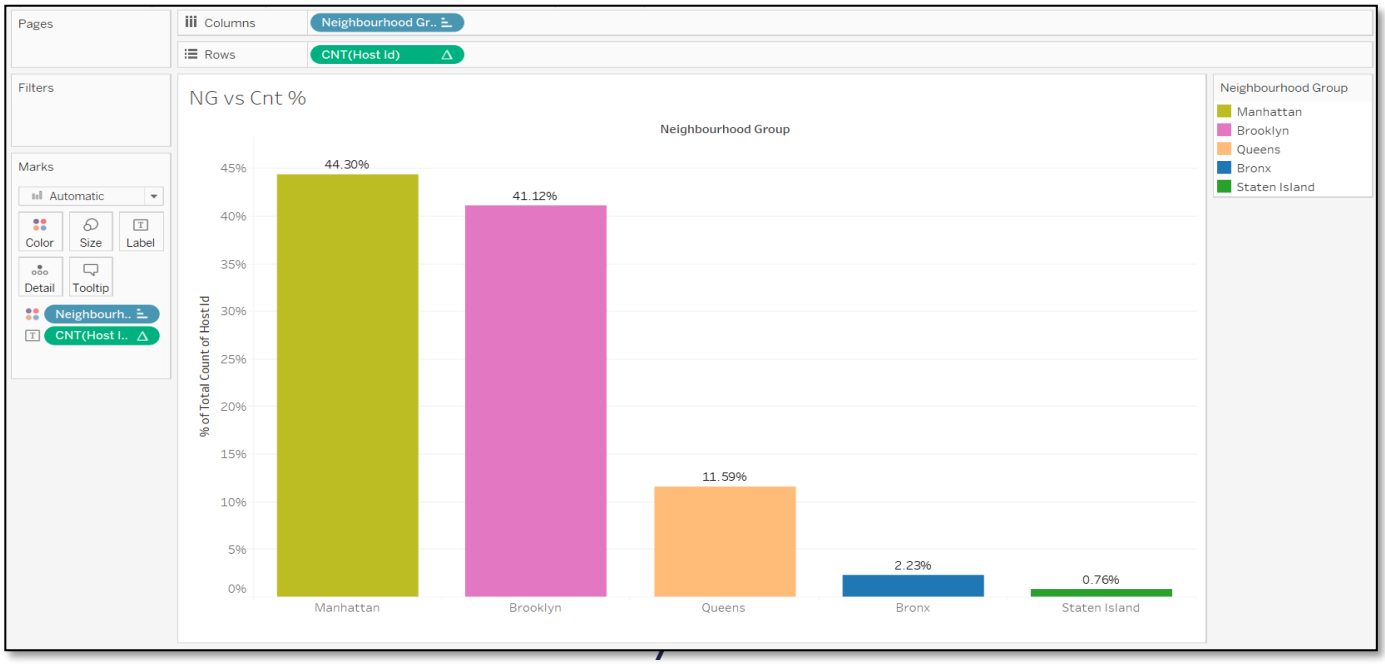
Explanation for EDA:

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 neighborhood group. Two plots were used to explore this question.
-**Geographical plot**: This was created using the parameters latitude, longitude, neighborhoods,and neighborhood 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 neighborhood. We use the parameters Neighborhood group & CNT(Host_Id).





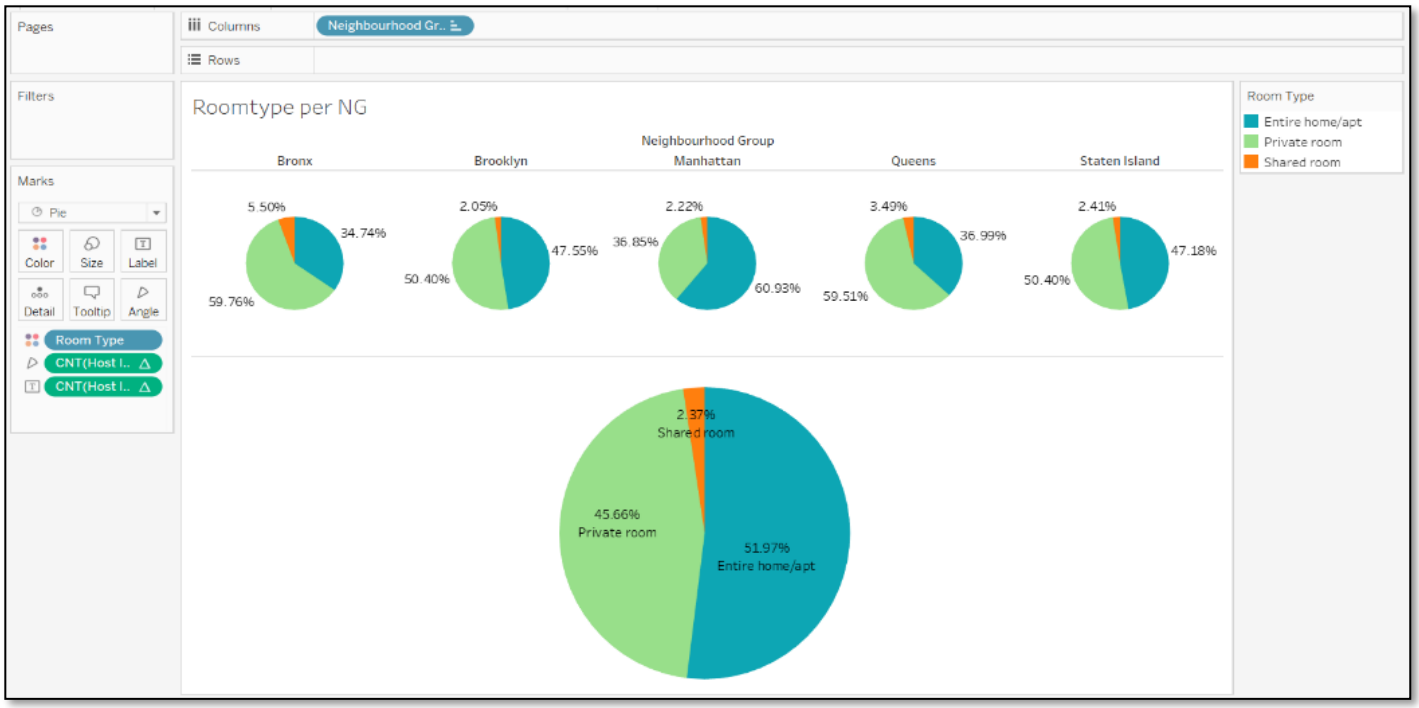
Inference:

- The presence of Airbnb is substantially high in Manhattan, Brooklyn as compared to Queen, Bronx & Staten Island contribute in NYC
- Manhattan tops the listing with 44.30%, followed by Brooklyn 41.12% few reasons for could be high population density, tourism hub etc.
- Bronx (2.23%) & Staten Island (~1%) has the least number of listings, due to its low population density and very few tourisms destination

What type of rooms do customers prefer?

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

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



Inferences:

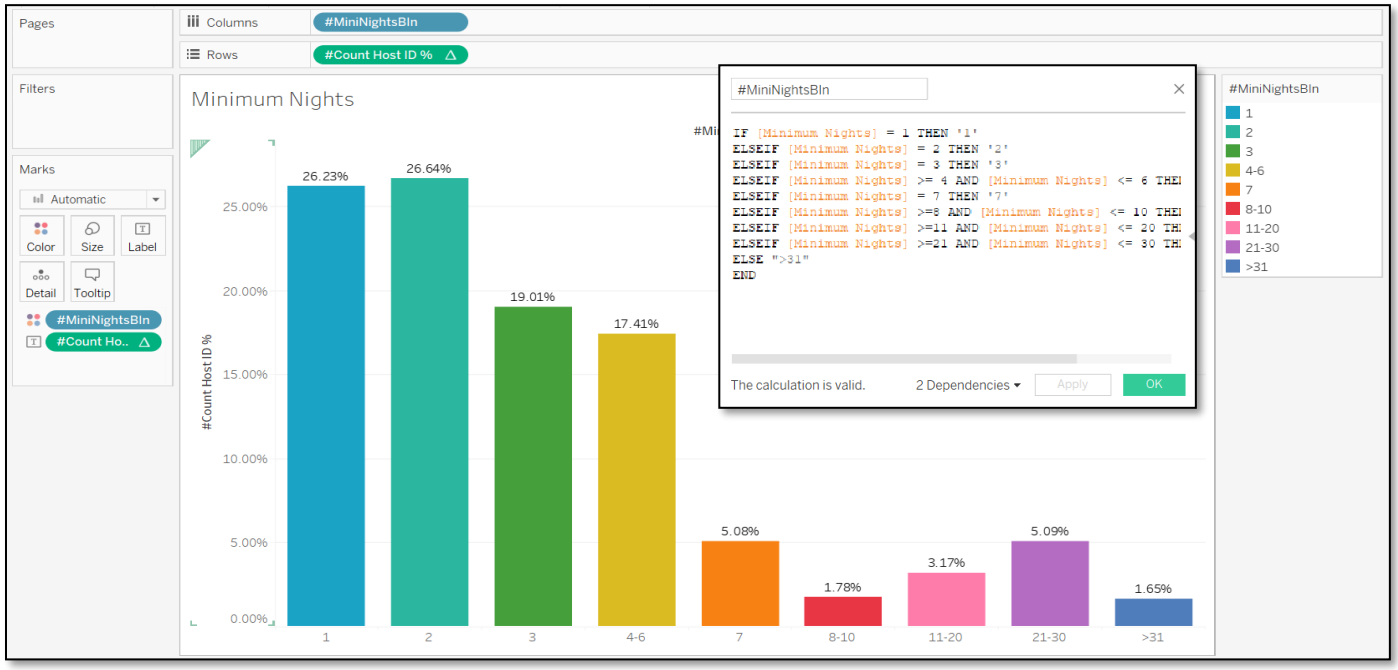
- There are namely three types of rooms in listings Entire Home/Apartment, Private Room & Shared Room.
- Overall, customers appear to prefer entire homes (51.97%) or private rooms (45.66%) while shared rooms seem to less suitable (2.37%).
- Private Rooms listings are dominant in every group except Manhattan where Entire home/apt is n majority.
- Private room shares are over 50% for Bronx, Brooklyn, Queen & Staten Island.
- Whereas, Manhattan has a higher contribution in entire home (61%), compared to the combined ratio of 52%.
- A smaller number of shared rooms are available in each Neighborhood group.

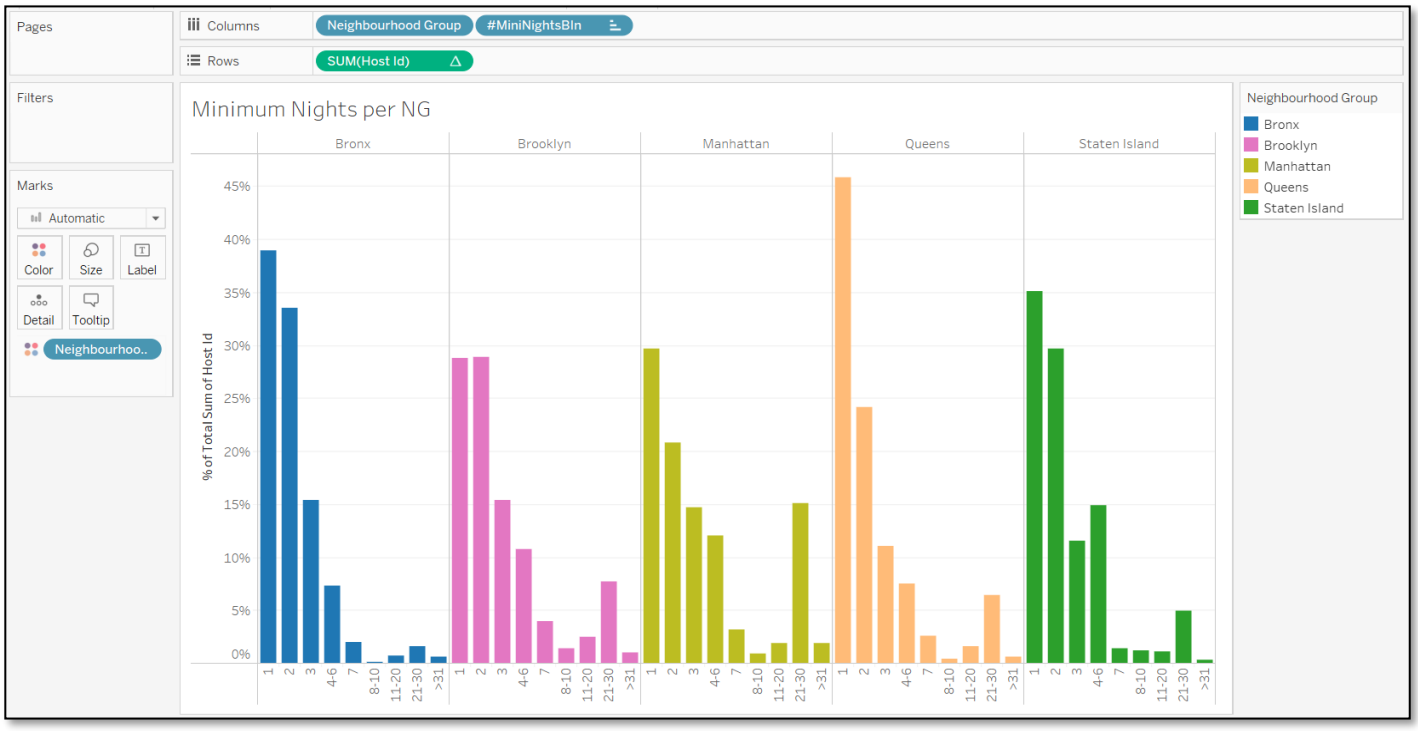
What is the preference in terms of the number of nights?

This question was addressed to understand the accommodation needs of the customer and their preference. Existing data has been segmented in bins to perform the below analysis

This has been explored using two bar graphs.

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





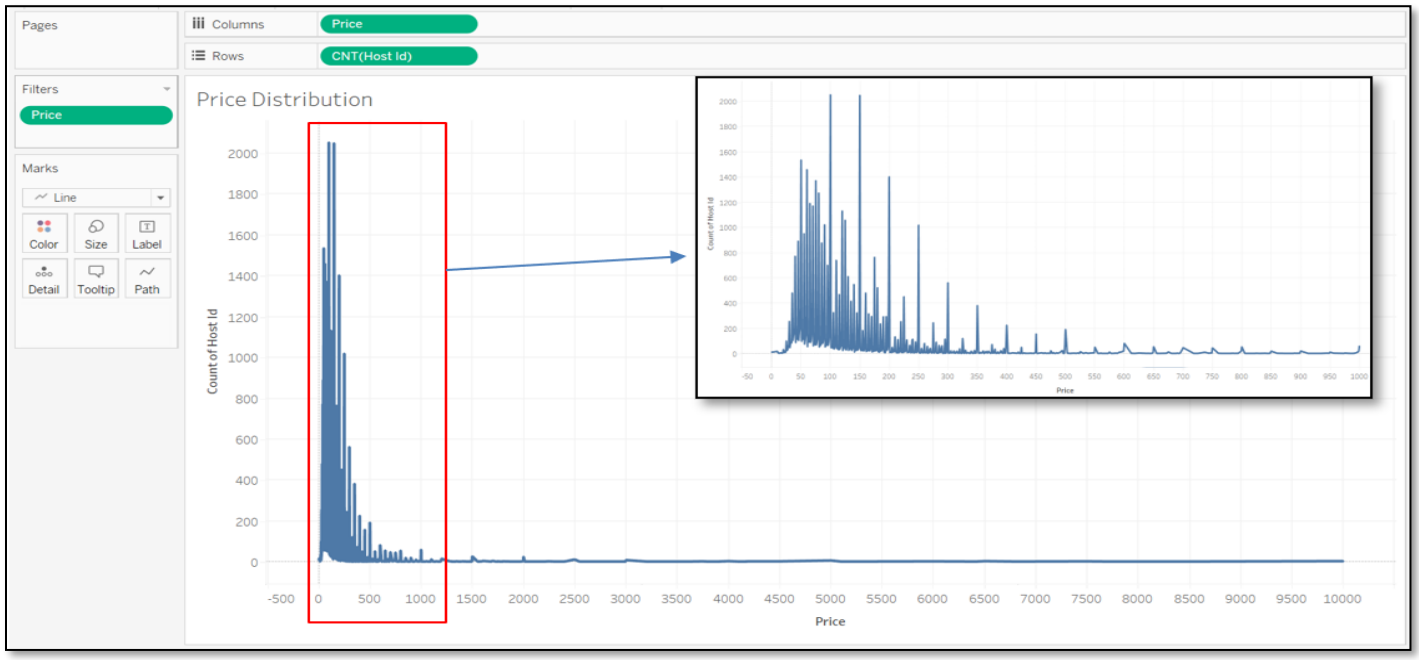
Inference:

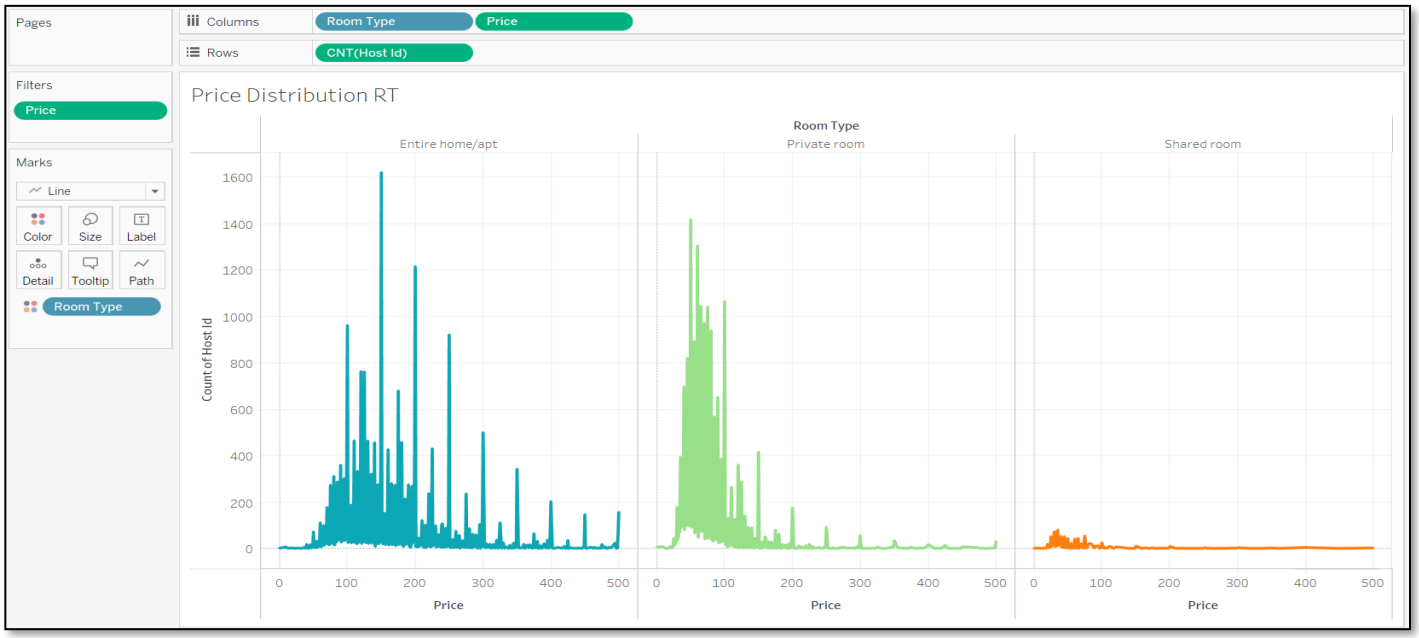
- Surprisingly the minmum_nights ranges from 1 to 1250.
- Customer prefers to book the for minimum of 2 nights (26.64%) followed by only 1 night (26.23) as per the data. These two constitutes more than 50% of preference of customers.
- We see substantially less number of bookings i.e.1.65% for more than 30 nights.
- 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, this can also be explained by the monthly rent taking trend.

Based on Price:

Most preferred neighborhood & most preferred room type. The price 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: across neighborhood & room type.

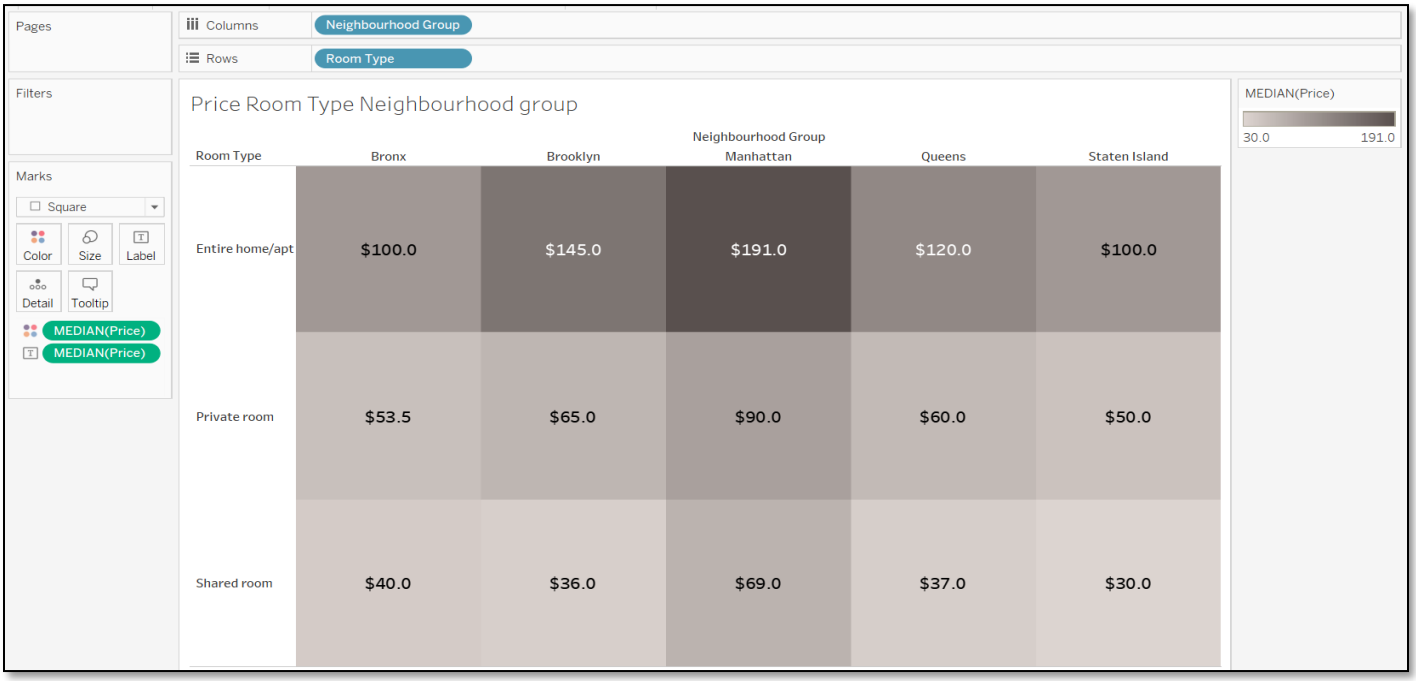
The parameters taken for analysis are: Room type, Neighborhood Group COUNT(Host_id), Price(Dimension), Median Price

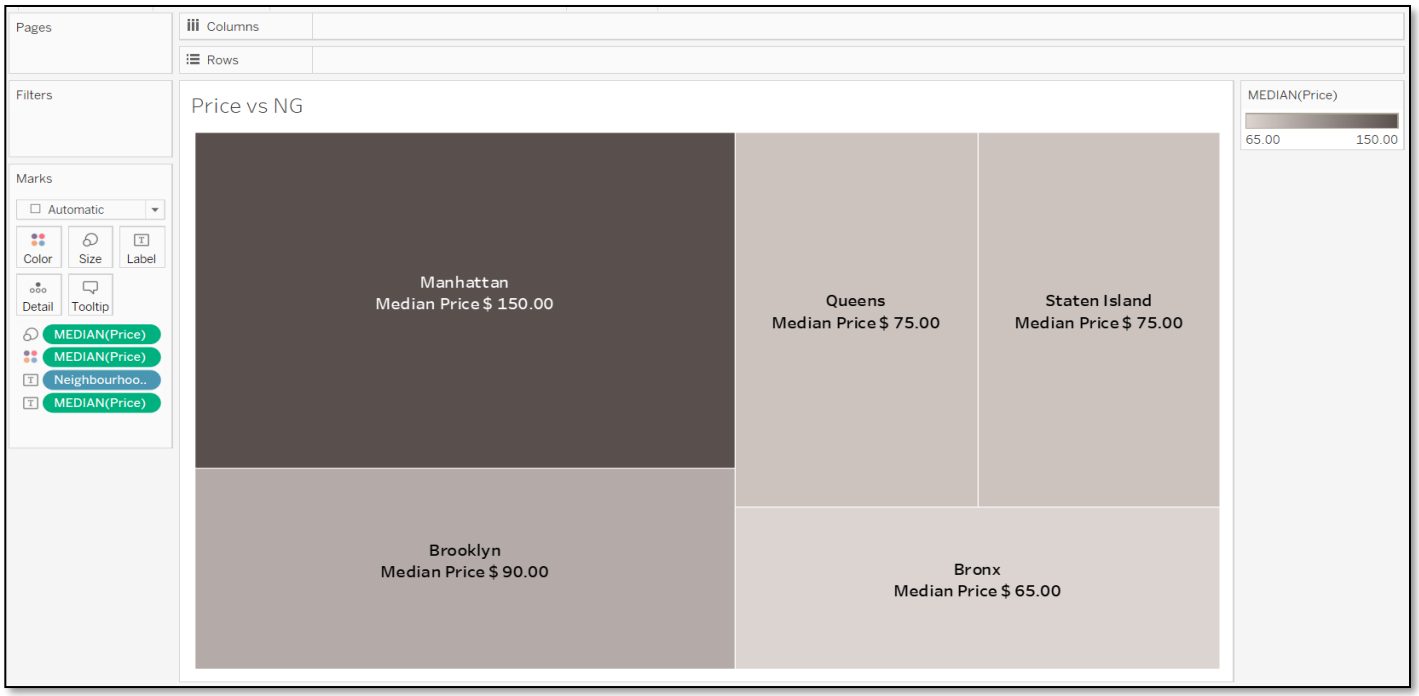




Inference:

- The distribution appears to be skewed towards left in a range of $1 to $10,000.
- Furthermore, on zooming in skewed region the majority distribution tends to be between $50 to $ 200.
- The prominent spikes have been observed at the prices $50, $100, $150, $200, $250, $300, $350 & $400 this be could be host preferring a whole number for pricing of their listing.

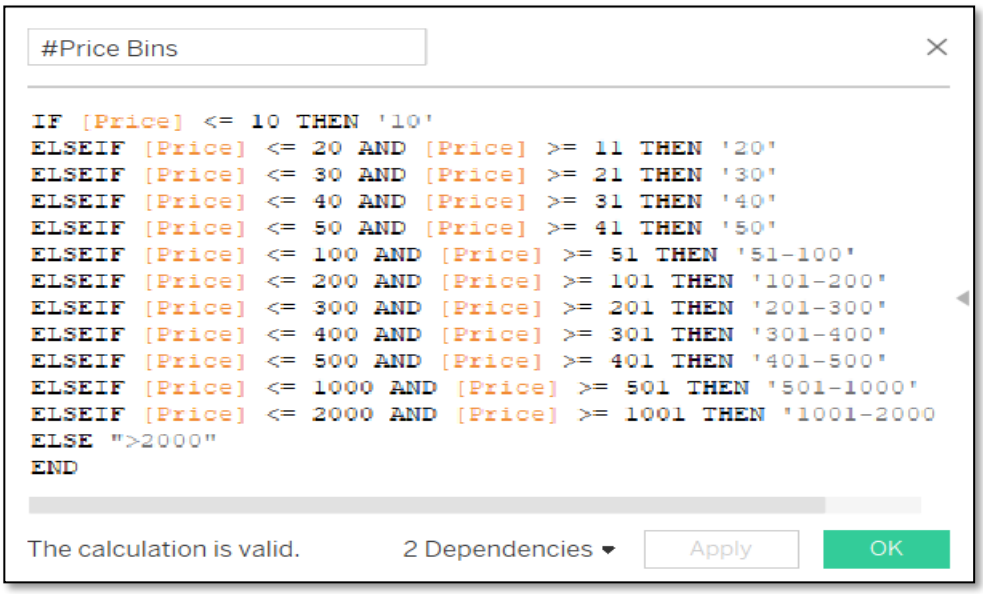




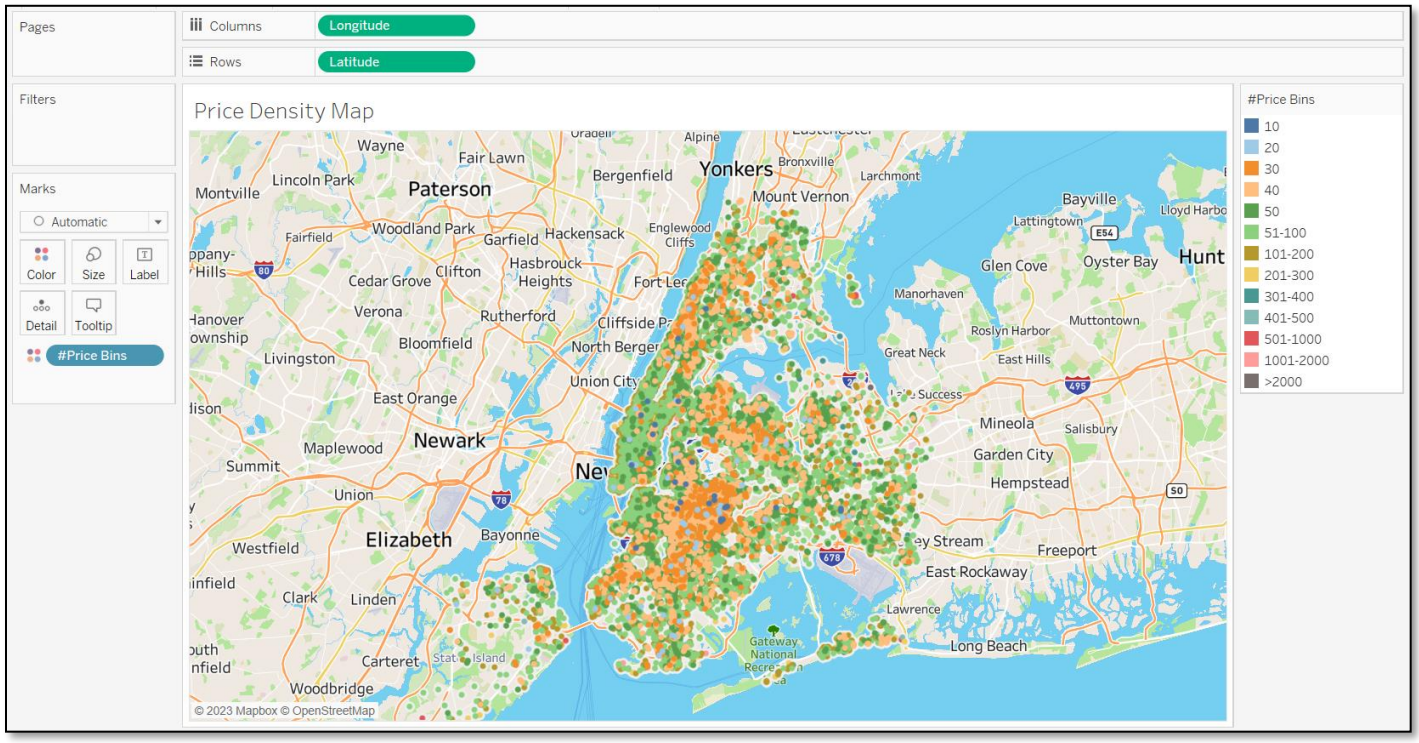
Inference:

- Manhattan appears to have the highest median price of $150.0. The 'Entire home/apt' room type in Manhattan is the most expensive at $191, much higher than the overall average
- 'Shared Room' type is the cheapest in Staten Island, Queens & Brooklyn.

For next step the Price Feature has been segregated in bins to further makes us understand the insights The steps followed to make the make is shown below



A price density chart has been plotted with Map coordinates to reveal the area of high & low prices.



Inferences:

- The map displays the price variation, which appears to be distributed uniformly in the in land areas.
- We see spike in prices in coastal cities, owing to better view from stays and easy ferry reachability.
- we zoomed in, we also observed higher pricing near colleges or important monuments/landmarks.

Based on Availability:

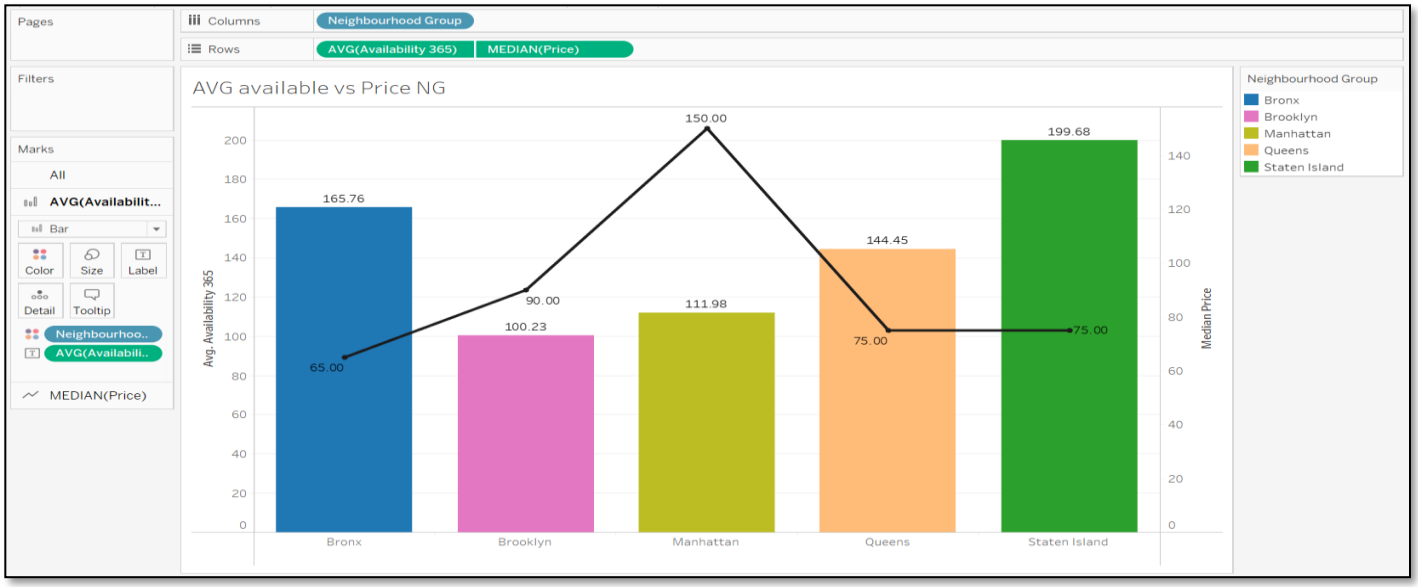
Most preferred neighborhood & most preferred room type. The availability parameter was chosen, to analyse the factor in bookings & preferences. Here again, two different parameters that were taken for comparison: across neighborhood & room type. To perform such analysis again the availability has been segmented in bins as shown below.

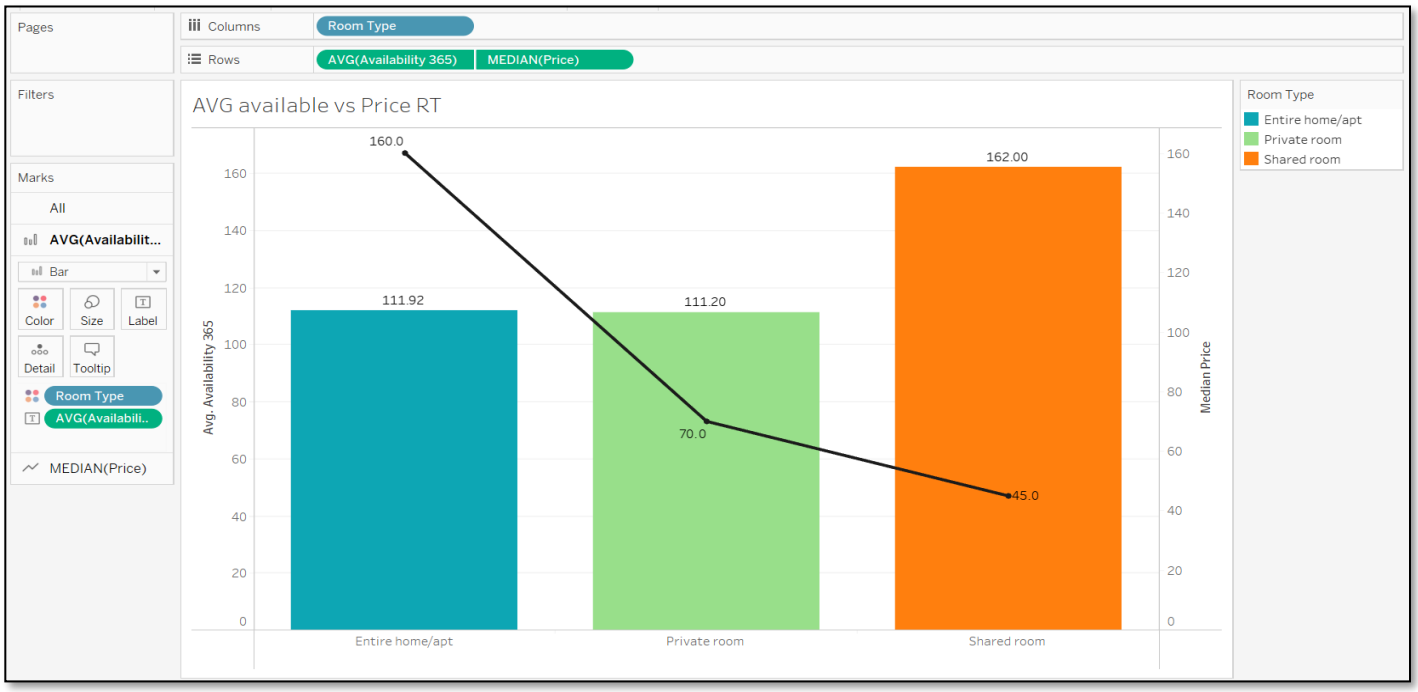
Neighbourhoo..	#Availbility Bins 15	16-30	31-60	61-90	91-120	121-150	151-180	181-210	211-240	241-270	271-300	301-330	331-365
Bronx	223	28	82	140	34	46	98	32	22	38	40	86	222
Brooklyn	9,333	842	1,273	1,444	573	547	942	496	515	699	705	878	1,857
Manhattan	9,715	782	1,225	1,251	565	625	946	712	663	807	745	1,090	2,535
Queens	1,664	163	393	563	170	258	435	140	142	202	225	370	941
Staten Island	49	8	23	45	5	13	28	14	12	19	28	35	94

Inference:

- Brooklyn & Manhattan has majority of there listings only available for less than 15 days of all the groups.
- 331-365 bin has comparatively high count from rest of the bins. This trend is observed in all the neighborhood group.
- The other bins expect first & last has uniform count regardless of the neighborhood group.

Next is the analysis involving median price for average availability of the listings across the neighborhood groups





Inference:

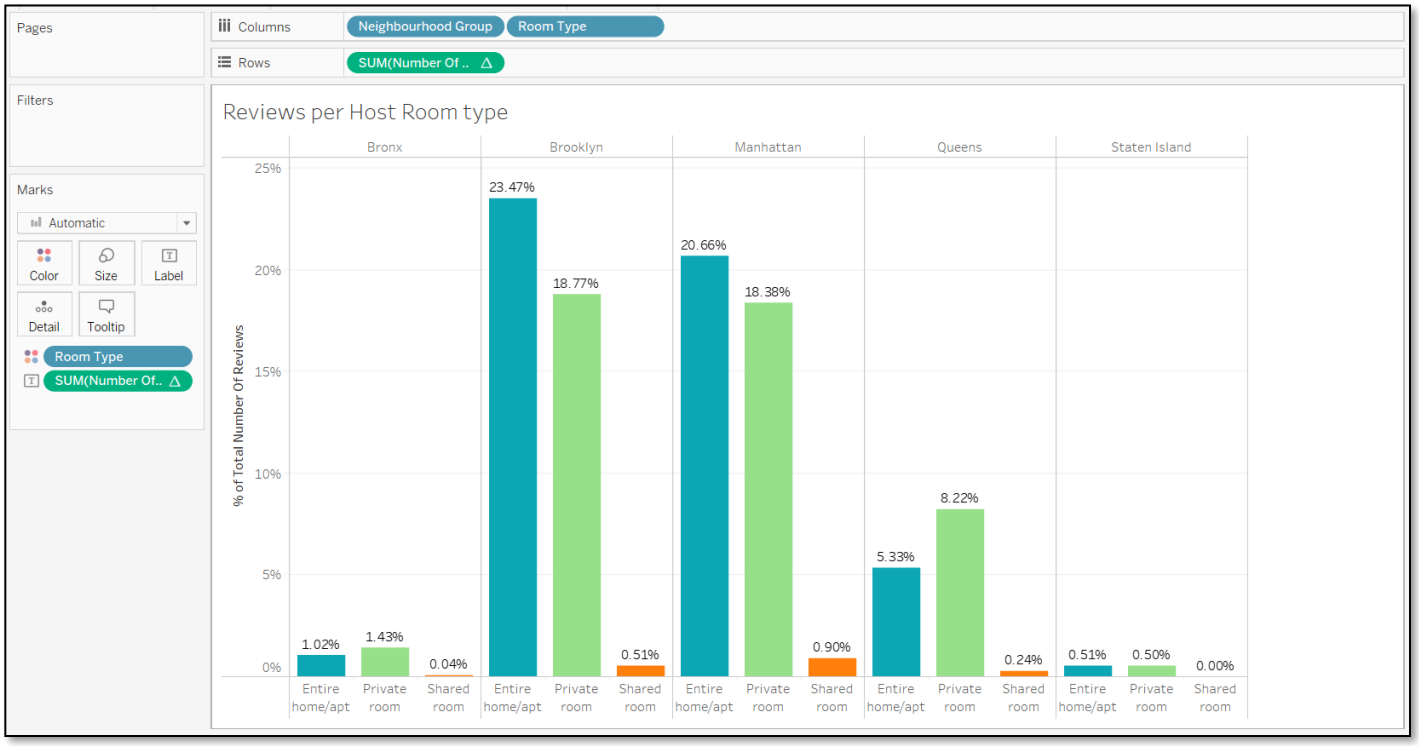
- Average availability of listings in Manhattan is comparatively less than other groups with higher pricing.
- Shared rooms are frequently available with having less price for accommodation.

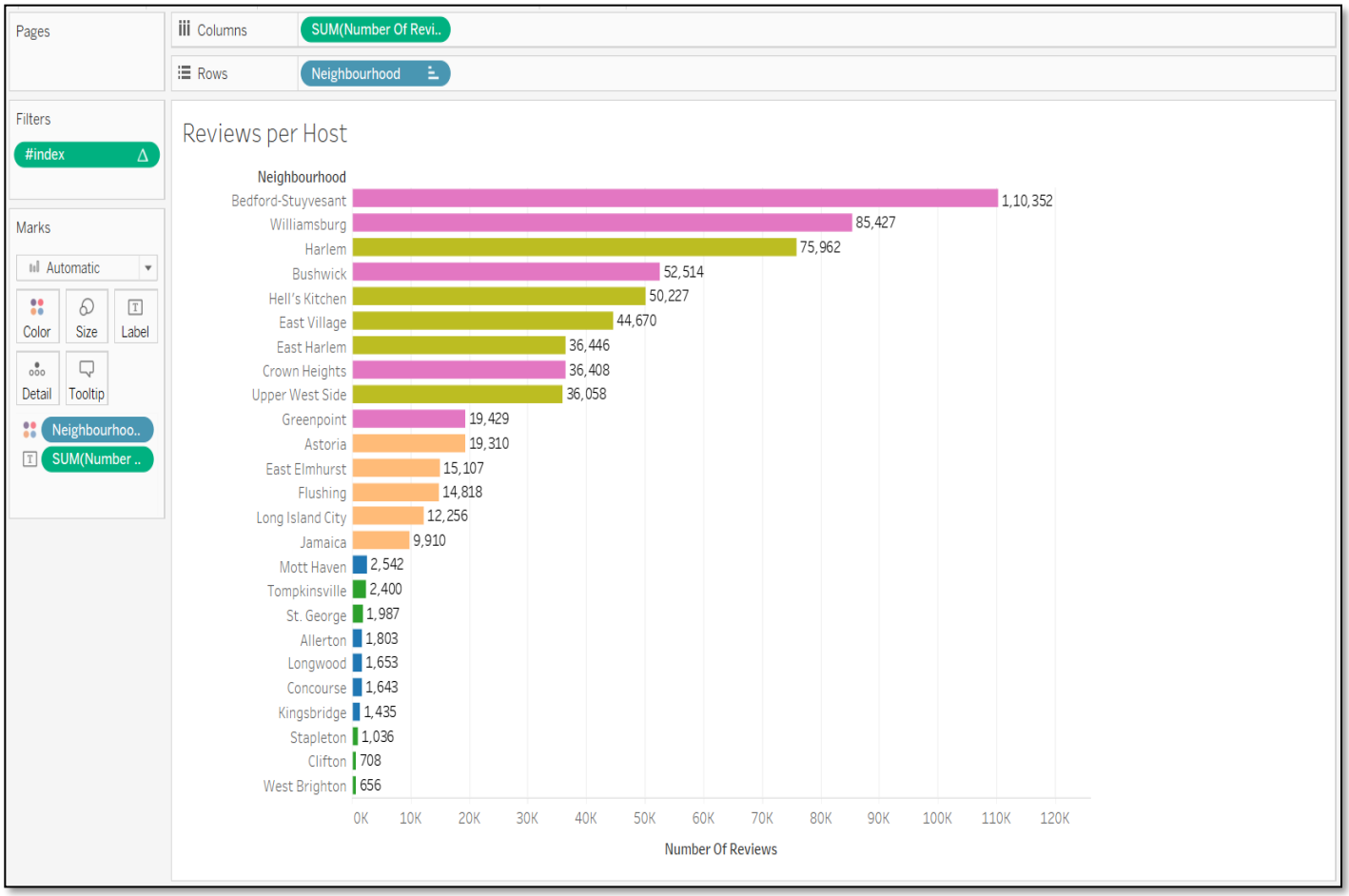
Based on customer review:

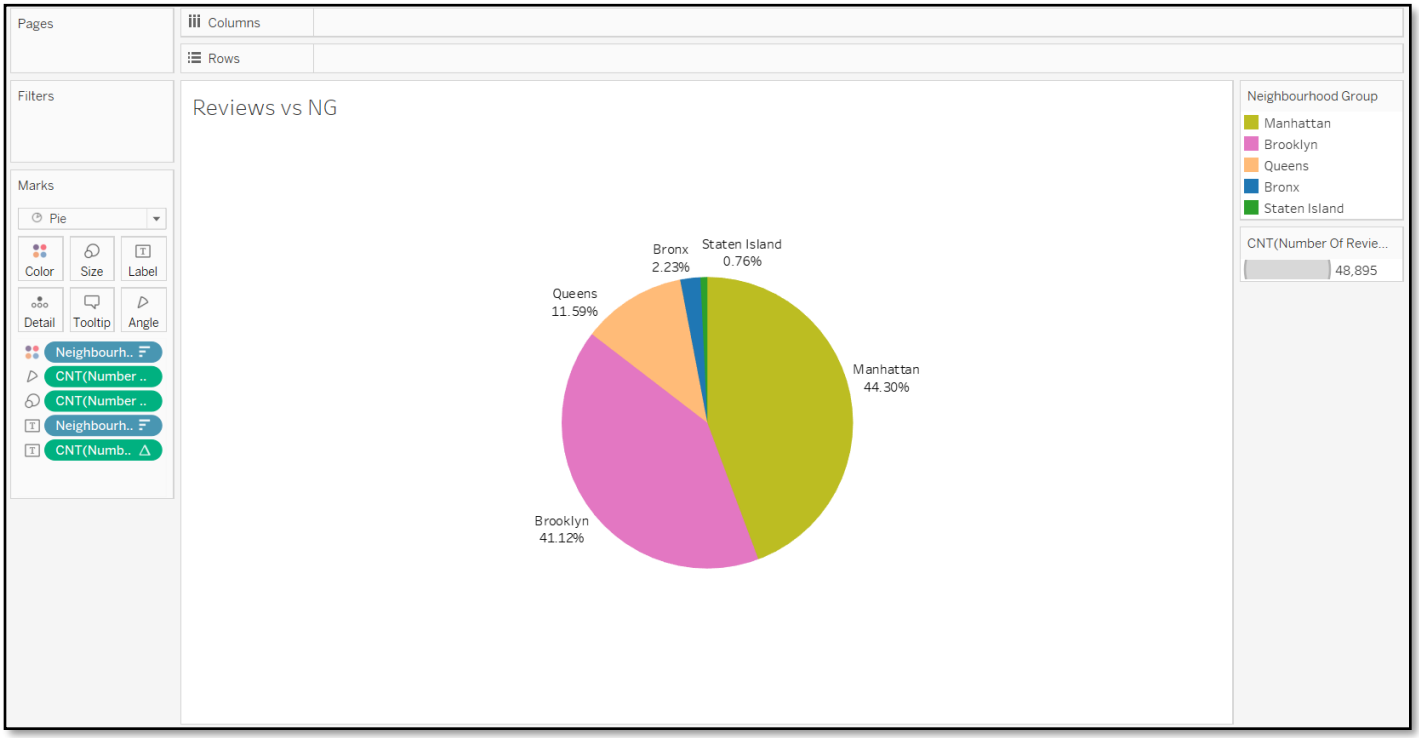
Most preferred neighborhood Most preferred room type The customer review parameter was chosen, as it is one of the most important factors to boost future bookings and listings Here again, two different parameters that were taken for comparison neighborhood room type.

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 Neighborhood group, SUM(Number of reviews)




Pages
Columns
SUM(Number Of Revi..
Rows
Neighbourhood
Filters
#index
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Neighbourhoo..
SUM(Number ..
Reviews per Host
Neighbourhood
Bedford-Stuyvesant 1,10,352
Williamsburg 85,427
Harlem 75,962
Bushwick 52,514
Hell's Kitchen 50,227
East Village 44,670
East Harlem 36,446
Crown Heights 36,408
Upper West Side 36,058
Greenpoint 19,429
Astoria 19,310
East Elmhurst 15,107
Flushing 14,818
Long Island City 12,256
Jamaica 9,910
Mott Haven 2,542
Tompkinsville 2,400
St. George 1,987
Allerton 1,803
Longwood 1,653
Concourse 1,643
Kingsbridge 1,435
Stapleton 1,036
Clifton 708
West Brighton 656
0K 10K 20K 30K 40K 50K 60K 70K 80K 90K 100K 110K 120K
Number Of Reviews


Pages
Columns
Rows
Filters
Marks
Pie
Color
Size
Label
Detail
Tooltip
Angle
Neighbourh..
CNT(Number ..
CNT(Number ..
Neighbourh..
CNT(Numb..
Reviews vs NG
Neighbourhood Group
Manhattan
Brooklyn
Queens
Bronx
Staten Island
CNT(Number Of Revie...
48,895
Bronx 2.23%
Staten Island 0.76%
Queens 11.59%
Manhattan 44.30%
Brooklyn 41.12%

Inference:

- The listings of the Entire Home/apt has maximum reviews in Brooklyn & Manhattan. While Queens has more reviews on Private rooms.
- The traction towards shared room has been very low for every group hence the lowest number of reviews.
- The reason more reviews for Manhattan & Brooklyn group is they are a hub of financial sector & has many tourisms centered places
- As we have seen Manhattan & Brooklyn have higher number of 1-6 & 30-day bookings compared to the others. This could be the reason for higher percentage of booking of Entire home/apt.
- We see that Bedford-Stuyvesant from Brooklyn is the highest popular with 1,10,352 no's of reviews in total followed by Williamsburg 85,427.
- Harlem from Manhattan got the highest no of reviews followed by Hell's kitchen.
- The higher number of customer reviews hints the frequency of renting is more & hence higher satisfaction in these localities.

Recommendations Consolidated:

- Bronx & Staten Island has **costal region** which could be leveraged for more traction of customers
- To generate **more revenue Entire Home/apt** *type* should be focused as it has more traction as well good pricing. More offers & services would attract customers.
- Brooklyn has an median price of $90. As there are already many listings available in Manhattan, Brooklyn can be considered for expansion.
- Average availability of listings in Manhattan Brooklyn is comparatively less than other groups with higher pricing. These areas could be targeted to acquire more properties.
- Shared rooms are frequently available with having less price for accommodation. Listing of shared room to be decreased as the revenue generated would be less
- Also upon values missing *in last_review* and *reviews_per_month* carrying *NaN* values on purpose, meaning they are not missing at random as these hosted sites/places have not received any reviews from the customers. Hence, these places would be least preferred by the future customers and would also be facing bad business from our side.
- As proportion of shared rooms is substantially low these could be targeted with discounts to increase bookings or can be looked into to be converted to other types.
- **Weekly or bi-weekly rentals** can also be acquired, as these can be used customers stranded in NYC for quarantine purposes.
- 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.
- **New acquisitions and expansion** can be done **in the price range of $50 -$200** 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.
- Increasing acquisitions and new properties in coastal regions can increase customer bookings.

Link for Tableau Visualizations:

https://public.tableau.com/app/profile/chinmay.kumar.sahu2740/viz/ABNDraft-1/ReviewsvsNG?publish=yes