Storytelling Case Study: Airbnb, NYC

By: Chinmay Kumar Sahu,
Sahil Panigrahi,
& Pranay Chide

AGENDA

- Introduction
- Business Understanding
- Problem Statement
- Objective
- EDA Approach
- Analysis & Insights
- Conclusion
- Credits
- Appendix

INTRODUCTION

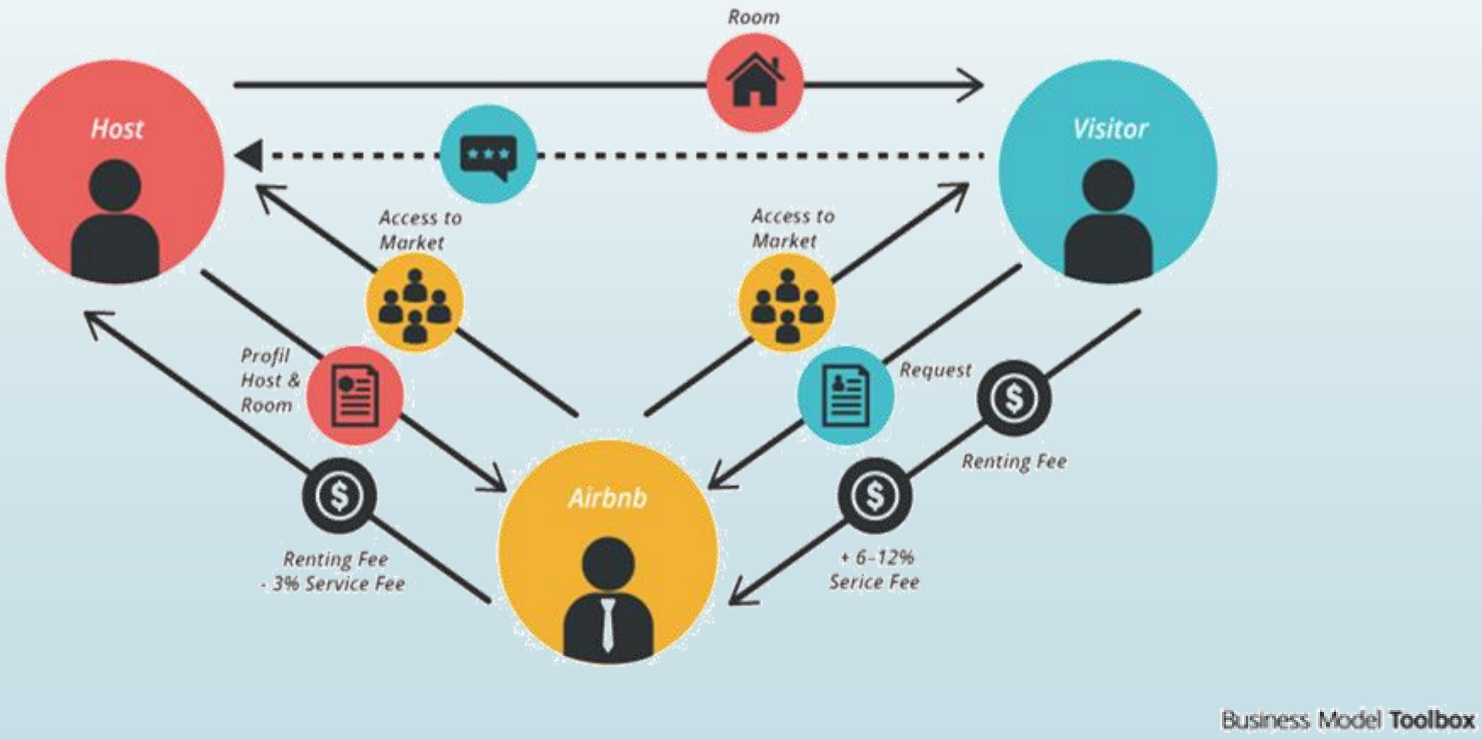
Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.

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.

BUISNESS UNDERSTANDING

The Airbnb business model is a two-sided marketplace that serves both property owners and guests. Property owners offer their homes or rental properties on the platform, while guests book these properties for a specified period.



PROBLEM STATEMENT

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.

OBJECTIVE

The presentation will focus mainly on the following points:

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

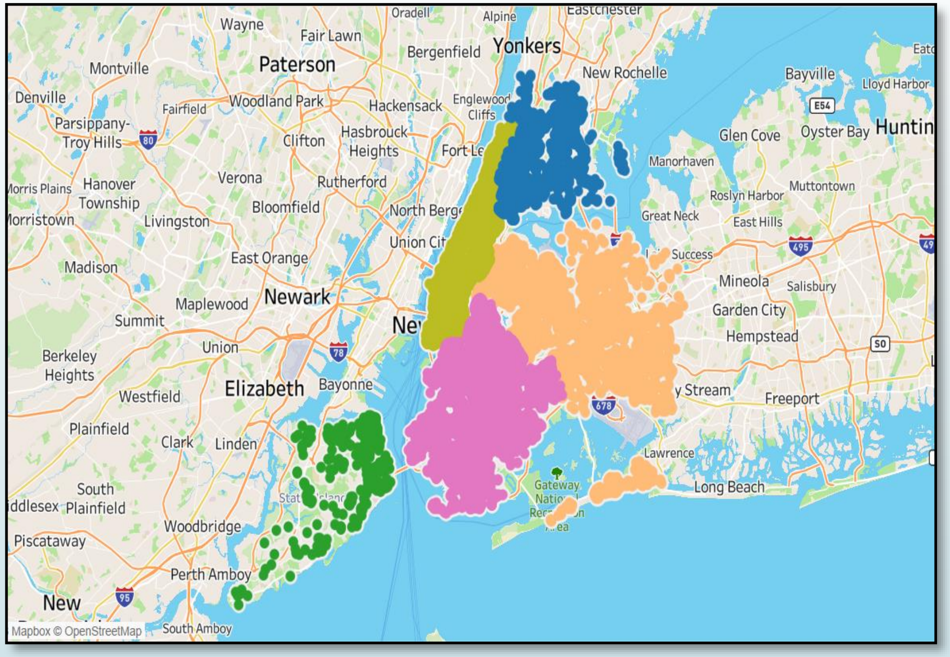
EDA APPROACH

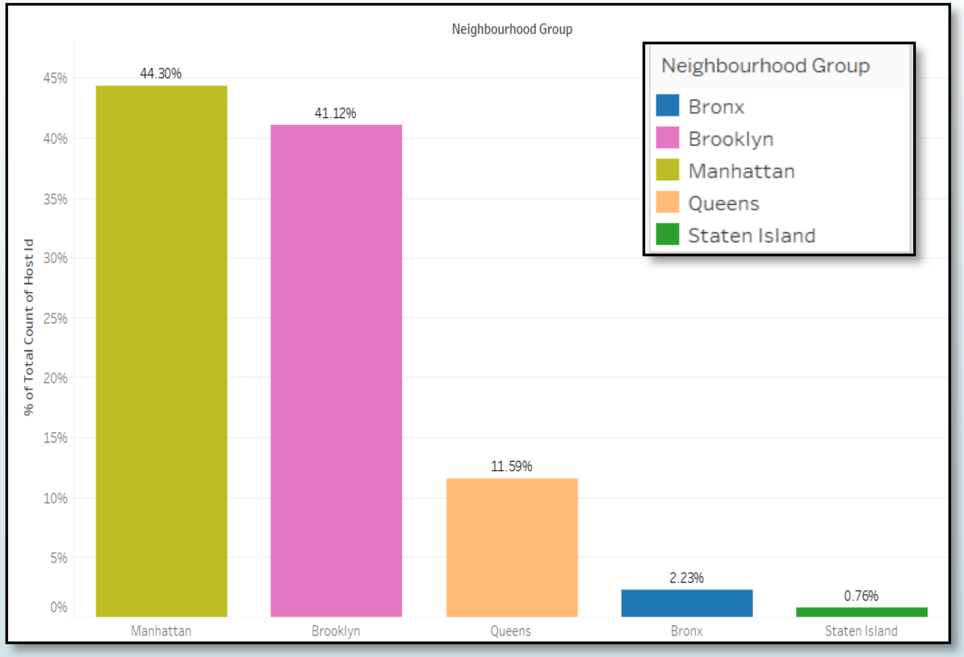
- *“AB_NYC_2019.csv”* file was loaded in Jupyter notebook to perform necessary steps to prepare & understand the data.
- After all steps were done and then file was downloaded under the name of *“EDA_AIRBNB_NYC.csv”.*
- This file was then used for data visualization in “*Tableau Public 2023.1”*.
- All the insights have been derived and enclosed in the upcoming slides.

*More on this is explained in the Methodology Document attached

ANALYSIS & INSIGHTS

Number of listings in NYC Neighborhood Group

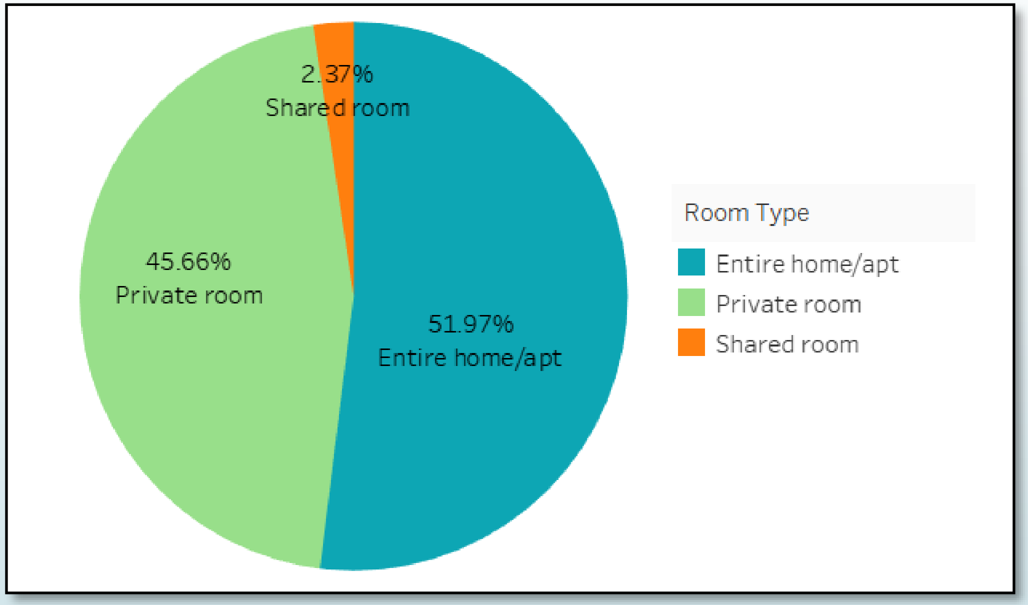




- The presence of Airbnb is substantially high in Manhattan, Brooklyn as compared to Queen, Bronx & Staten Island contribute in NYC
- Manhattan tops the listing with **44.30%**, followed by Brooklyn **41.12%** few reasons for could be high population density, tourism hub etc.
- Bronx **(2.23%)** & Staten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

ANALYSIS & INSIGHTS

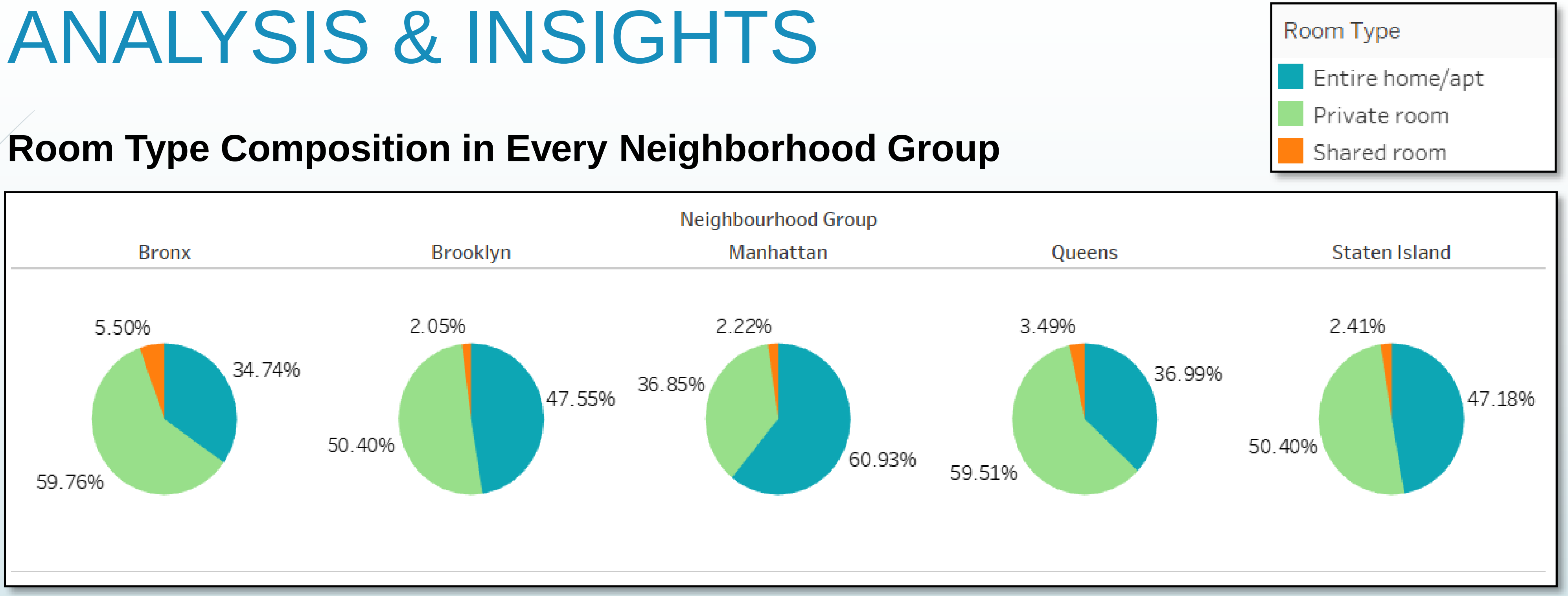
Room Type Categories in the Listings



- 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 seems to less suitable (2.37%)

ANALYSIS & INSIGHTS

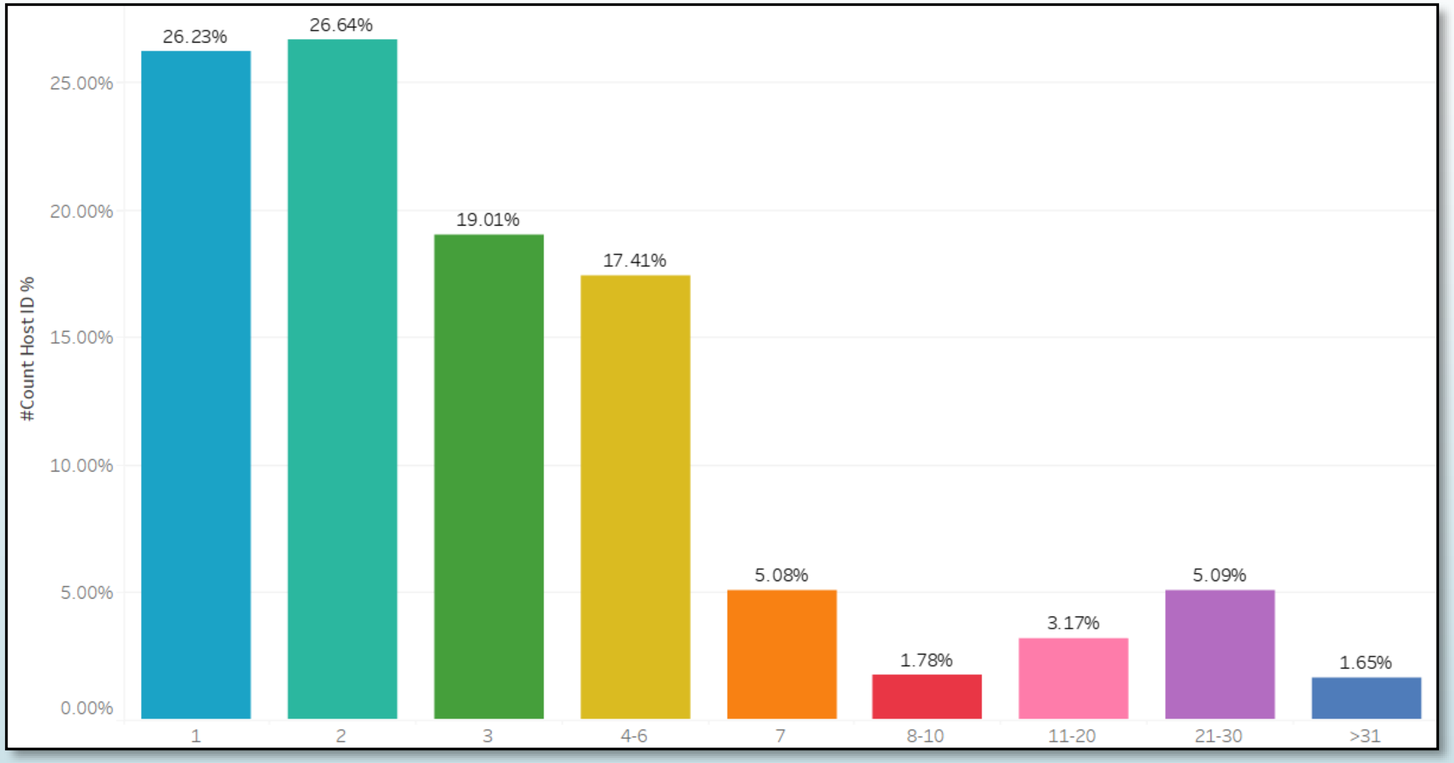
Room Type Composition in Every Neighborhood Group



- **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%.
- Very **less number of shared rooms** are available in each Neighborhood group.

ANALYSIS & INSIGHTS

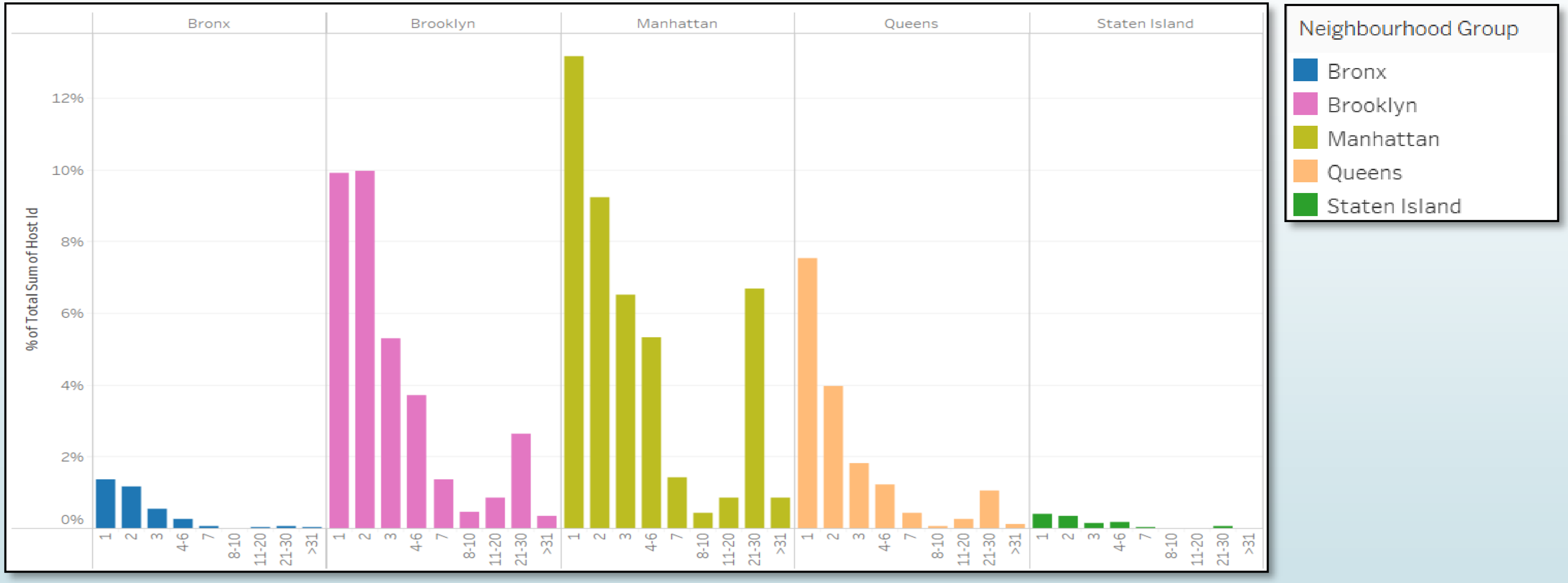
Minimum Number of Nights



- 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**.

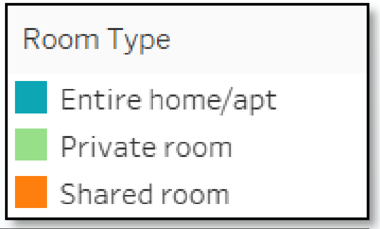
ANALYSIS & INSIGHTS

Minimum Number of Nights per Neighborhood Group

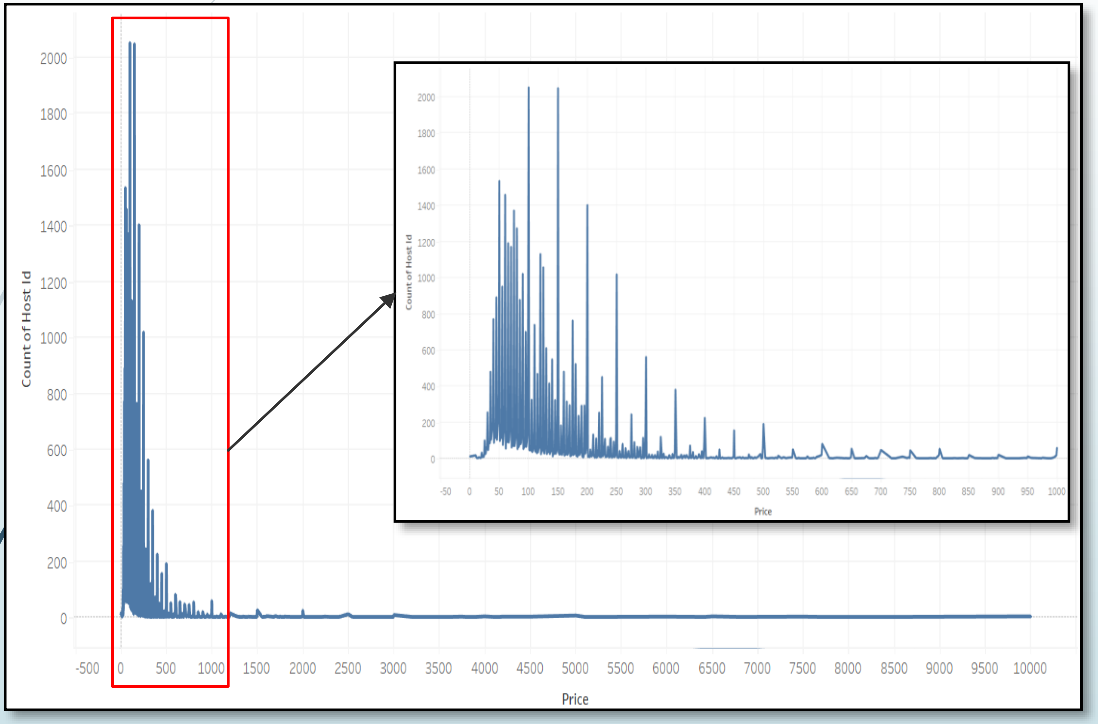


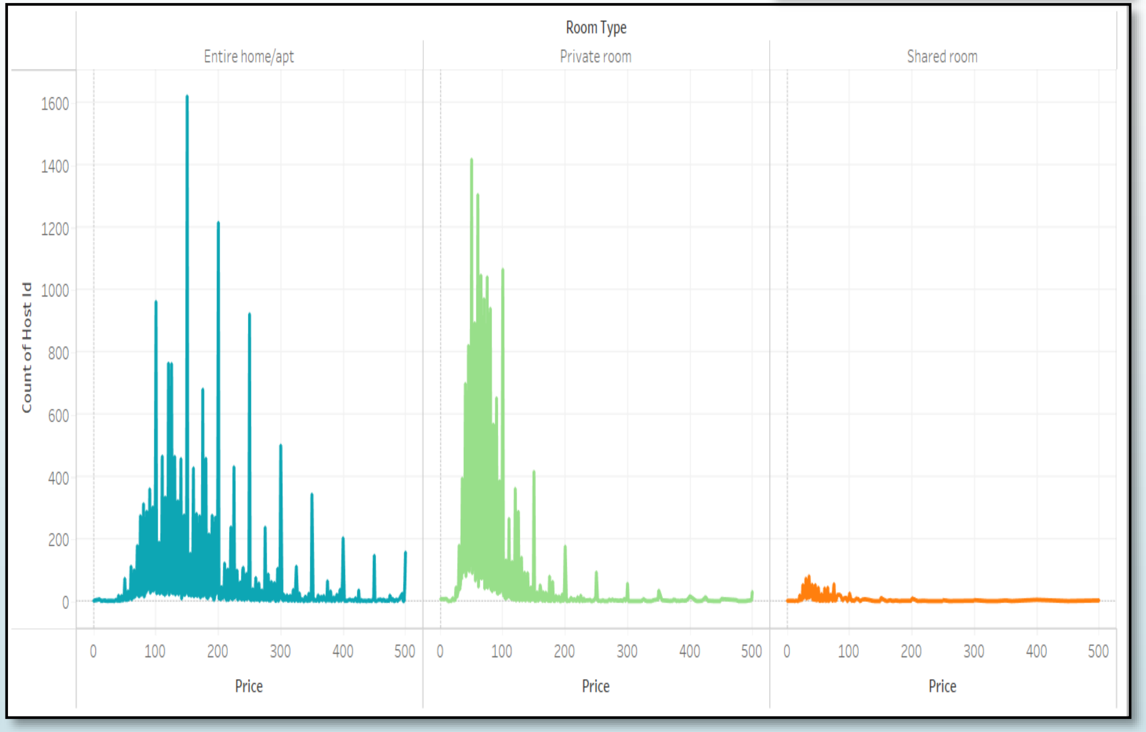
- The listings with **Minimum nights 1-6** have the most number of bookings. We can see a prominent spike in **30 days**, this would be because customers would rent out on a monthly basis.
- After 30 days, we can also see small spikes, this can also be explained by the monthly rent taking trend.
- **Manhattan & 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 bookings due company visits.

ANALYSIS & INSIGHTS



Price Distributions of the listings



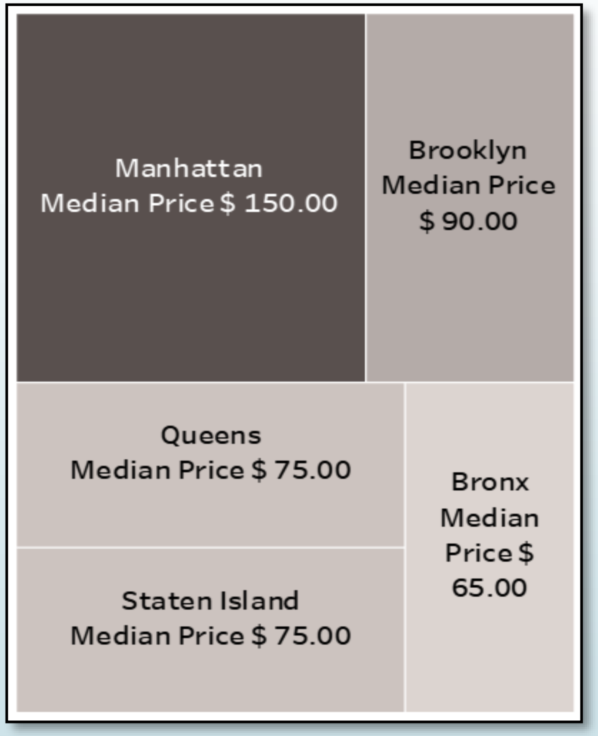


- The distribution appears to be **skewed towards left** in a range of $1 to $10,000.
- Furthermore on zooming in skewed region the majority distribution tends to be between $50 to $ 200.
- The prominent spikes has been observed at the prices $50, $100, $150, $200, $250, $300, $350 & $400 this be could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

ANALYSIS & INSIGHTS

Median Price, Room Type & Neighborhood group

Room Type	Bronx	Brooklyn	Manhattan	Queens	Staten Island
	Neighbourhood Group				
Entire home/apt	$100.0	$145.0	$191.0	$120.0	$100.0
Private room	$53.5	$65.0	$90.0	$60.0	$50.0
Shared room	$40.0	$36.0	$69.0	$37.0	$30.0

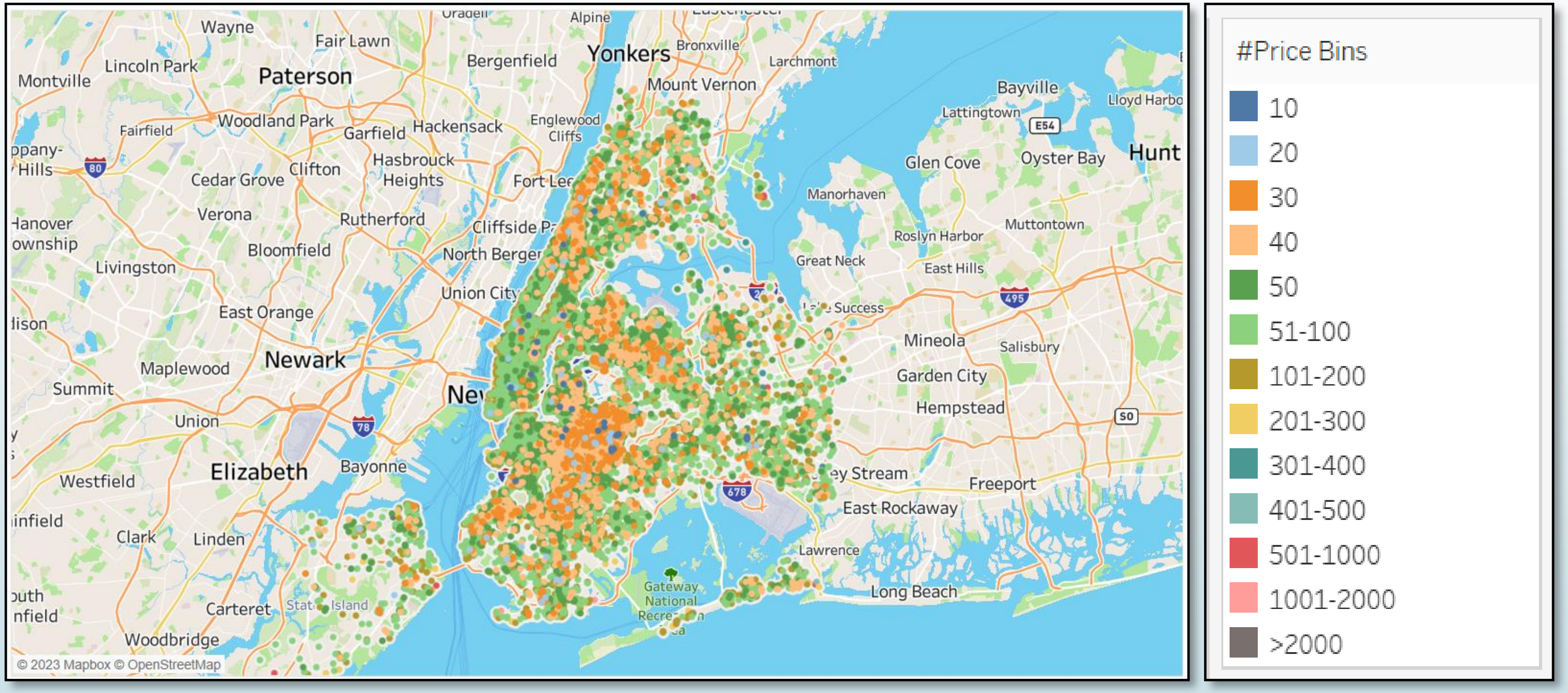


- Manhattan appears to have the highest median price of $150.0. The 'Entire home/apt' room type in Manhattan is the most expensive at $191, much higher than the overall average
- 'Shared Room' type is the cheapest in Staten Island, Queens & Brooklyn.

***NOTE**: Median price is considered in place of Average price as the price distribution is skewed towards left (shown in the pervious slide).*

ANALYSIS & INSIGHTS

Price Density across Neighborhood Group



- 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.

ANALYSIS & INSIGHTS

Availability of listings in Neighborhood Group

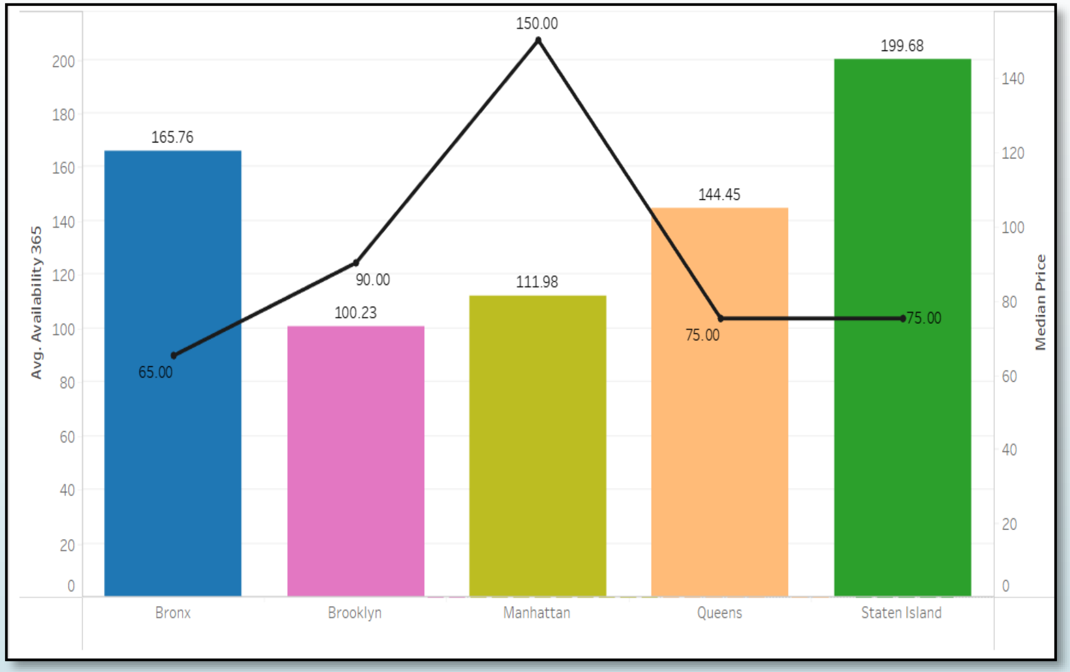
	#Availbility BIns												
Neighbourhoo..	15	16-30	31-60	61-90	91-120	121-150	151-180	181-210	211-240	241-270	271-300	301-330	331-365
Bronx	223	28	82	140	34	46	98	32	22	38	40	86	222
Brooklyn	9,333	842	1,273	1,444	573	547	942	496	515	699	705	878	1,857
Manhattan	9,715	782	1,225	1,251	565	625	946	712	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

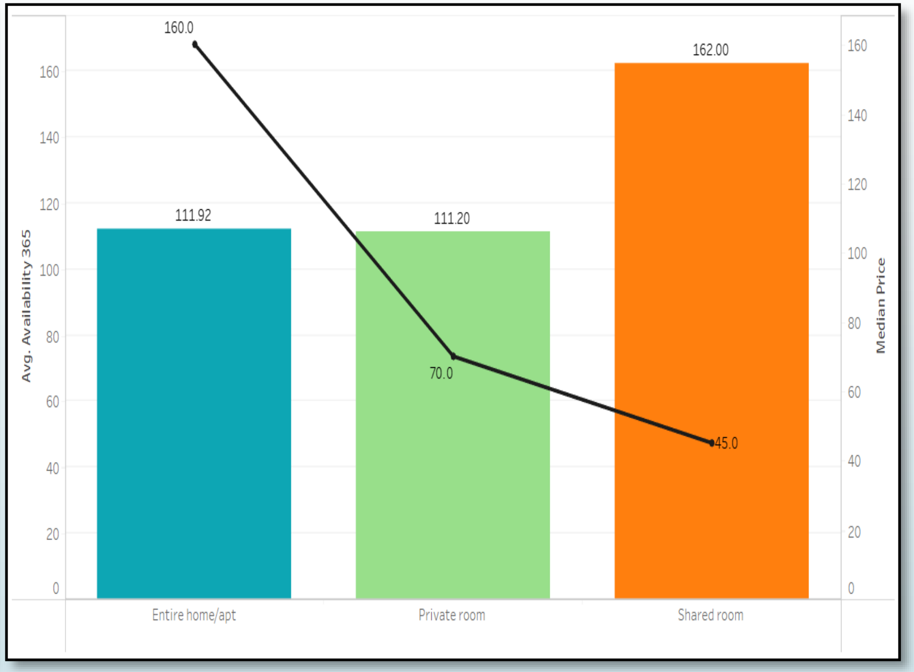
Room Type
- Entire home/apt
- Private room
- Shared room

- **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.

ANALYSIS & INSIGHTS

Availability of listings & Median Price in Neighborhood Group & Room type

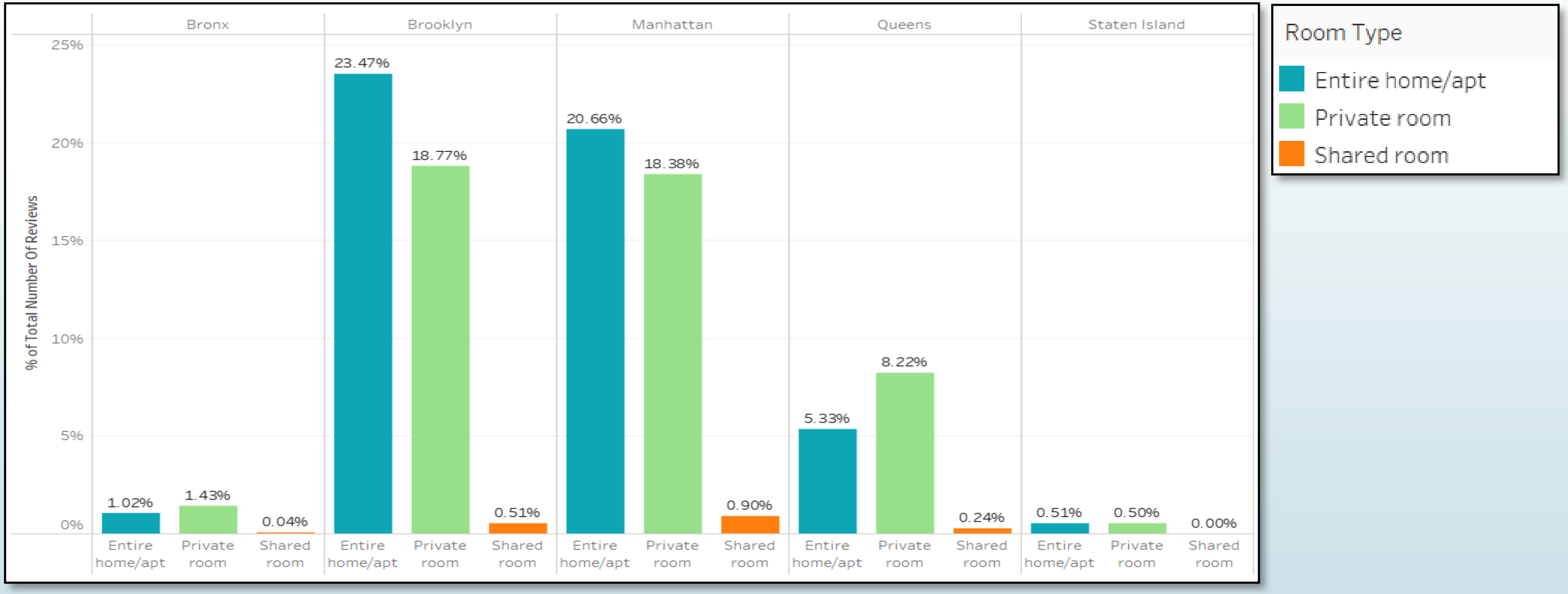




- Average availability of listings in Manhattan is comparatively less than other groups with higher pricing.
- Shared rooms are frequently available with having less price for accommodation.

ANALYSIS & INSIGHTS

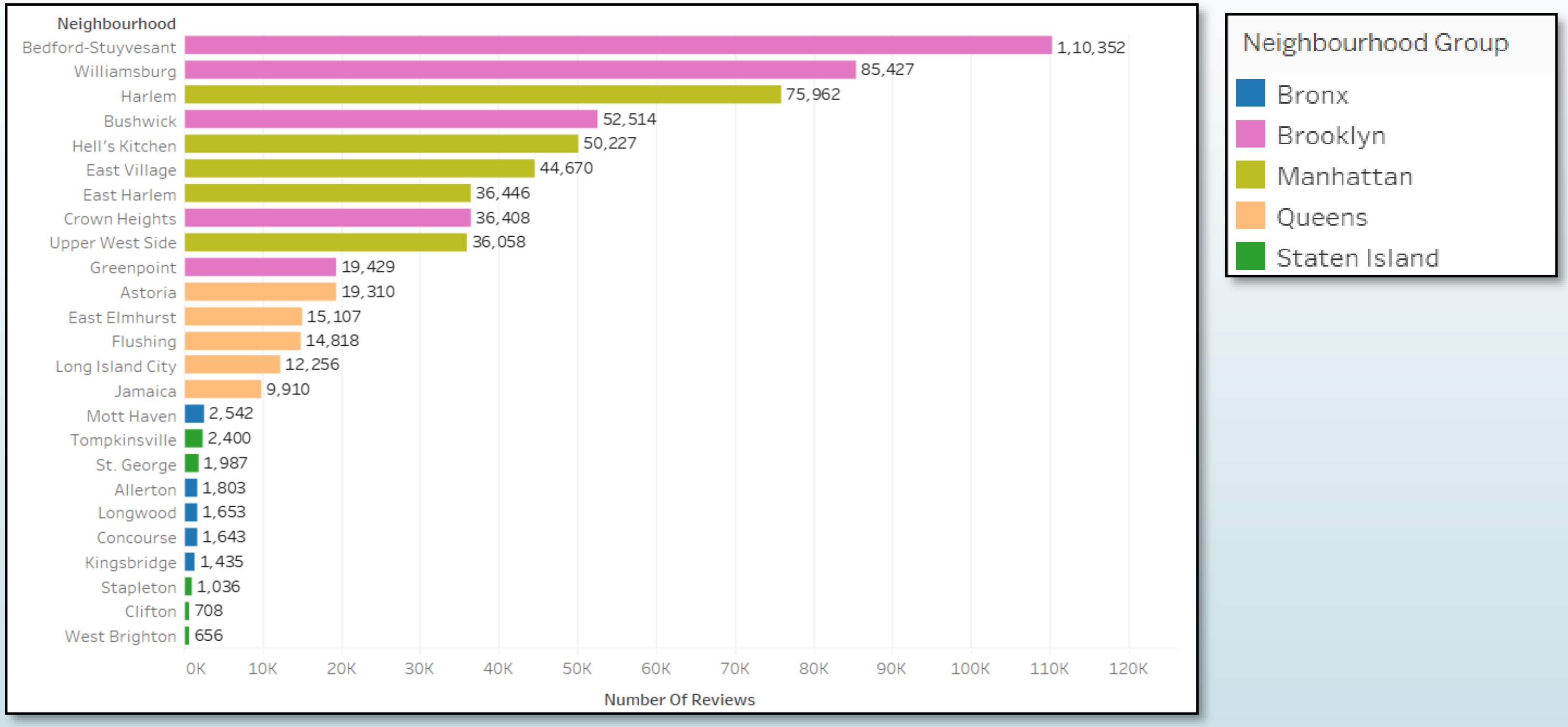
Customer Reviews of Room Types in every Neighborhood Group



- 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 tourism 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.

ANALYSIS & INSIGHTS

Top5 Highest Customer Reviews in every Neighborhood Group



- We see that **Bedford-Stuyvesant from Brooklyn** is the highest popular with 1,10,352 no 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.

CONCLUSION

- Bronx & Staten Island has **costal region** which could be leveraged for more traction of customers
- As proportion of shared rooms is substantially low these could be targeted with discounts to increase bookings.
- 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.
- **Weekly or bi-weekly rentals** can also be acquired, as these can be used customers stranded in NYC for quarantine purposes.
- **The price range of $50 -$200** as it satisfies both parameters of volume of customer traffic and customer satisfaction.
- 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 roomto 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.

APPENDIX

Data Dictionary

Column	Description
Id	Listing ID
Name	Name of Listing
Host_id	host ID
Host_name	Name of Host
Neighborhood_Group	Neighbourhood_group -Location
Neighborhood	Neighborhood -Area
Latitude	Map co-ordinates
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

APPENDIX

Data Understanding

Provided with Airbnb New York City Listings Dataset till 2019 (48895 Rows * 16 Columns)

```
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
```

```
# calulating the percentage of sum of missing values for each columns
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                 20.558339
calculated_host_listings_count     0.000000
availability_365                   0.000000
dtype: float64
```

The following features were dropped as they had null values & cannot be imputed based on the their category 'host_name', 'name' & 'last_review'.

APPENDIX

Methodology :

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

Assumptions :

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.

CREDITS

- [UPGRAD]
- [https://docs.python.org/3/]
- [TABLEAU]
- [AIRBNB]
- [Airbnb Business understanding]
- [Tableau Viz]

THANK YOU