DCS 52 : Airbnb NYC Storytelling Assignment - Leads

Prepared by :

1)Dharshak Chandra P

2)DigmaKumari Tarunkumar Patel

3) Divya Darshini

AGENDA

- Aim of Analysis
- Problem Statement
- Background
- Data preparation methods
- Key Analysis
- Appendix:
  - Data sources
  - Data methodology

Aim of Analysis

- 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 -
- Which type of hosts to acquire more and where?
- The categorisation of customers based on their preferences.
  - What are the neighbourhoods 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.
- What are the most popular localities and properties in New York currently?
- How to get unpopular properties more traction?

Problem Statement

- 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 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.

Data Preparation Methods

- Data cleaning to identify any missing values, duplicates or outliers
- Dropping irrelevant columns
- Generated seperate excel sheet for cleaned data using the data source provided
- Generating visual charts to better understand the data using bivariate analysis , univariate analysis , and much more using cleaned data file.

Before cleaning data: AB_NYC_2019.csv

```
n [39]: df = pd.read_csv("AB_NYC_2019.csv")
        df.head(5)
```

ut[39]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_
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	

```
In [2]: df = pd.read_csv("AB_NYC_2019.csv")

In [3]: df.info()

<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
```

After cleaning data: Airbnb_cleaned_dataset.csv

```
In [37]: df.to_csv(r'Airbnb_cleaned_dataset.csv',index=False, header=True)
         df.head(5)
```

Out[37]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_
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	

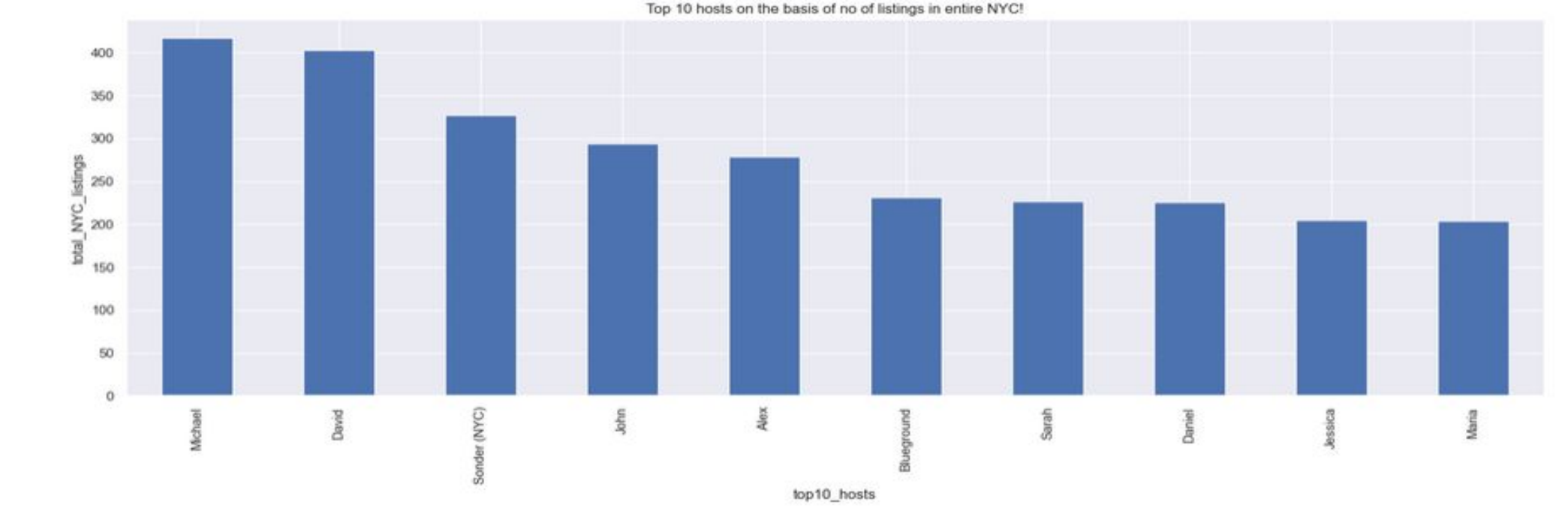
```
In [36]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48847 entries, 0 to 48846
Data columns (total 18 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              48847 non-null  object
 1   name                            48847 non-null  object
 2   host_id                         48847 non-null  object
 3   host_name                       48847 non-null  object
 4   neighbourhood_group             48847 non-null  object
 5   neighbourhood                   48847 non-null  object
 6   latitude                        48847 non-null  float64
 7   longitude                       48847 non-null  float64
 8   room_type                       48847 non-null  object
 9   price                           48847 non-null  int64
 10  minimum_nights                  48847 non-null  int64
 11  number_of_reviews               48847 non-null  int64
 12  last_review                     38811 non-null  datetime64[
 13  reviews_per_month               38811 non-null  float64
 14  calculated_host_listings_count  48847 non-null  int64
 15  availability_365                48847 non-null  int64
 16  price_range                     48847 non-null  category
 17  minimum_nights_range            48847 non-null  category
```

Key Analysis

1. Top 10 hosts
2. Popular neighbourhood group
3. Customer booking w.r.t minimum nights
4. Airbnb availability in different Neighbourhood groups with respect to Minimum nights
5. Neighbourhood vs Availability w.r.t minimum nights
6. Median price of the different Neighbourhood Groups
7. Preferred room type w.r.t neighbourhood group

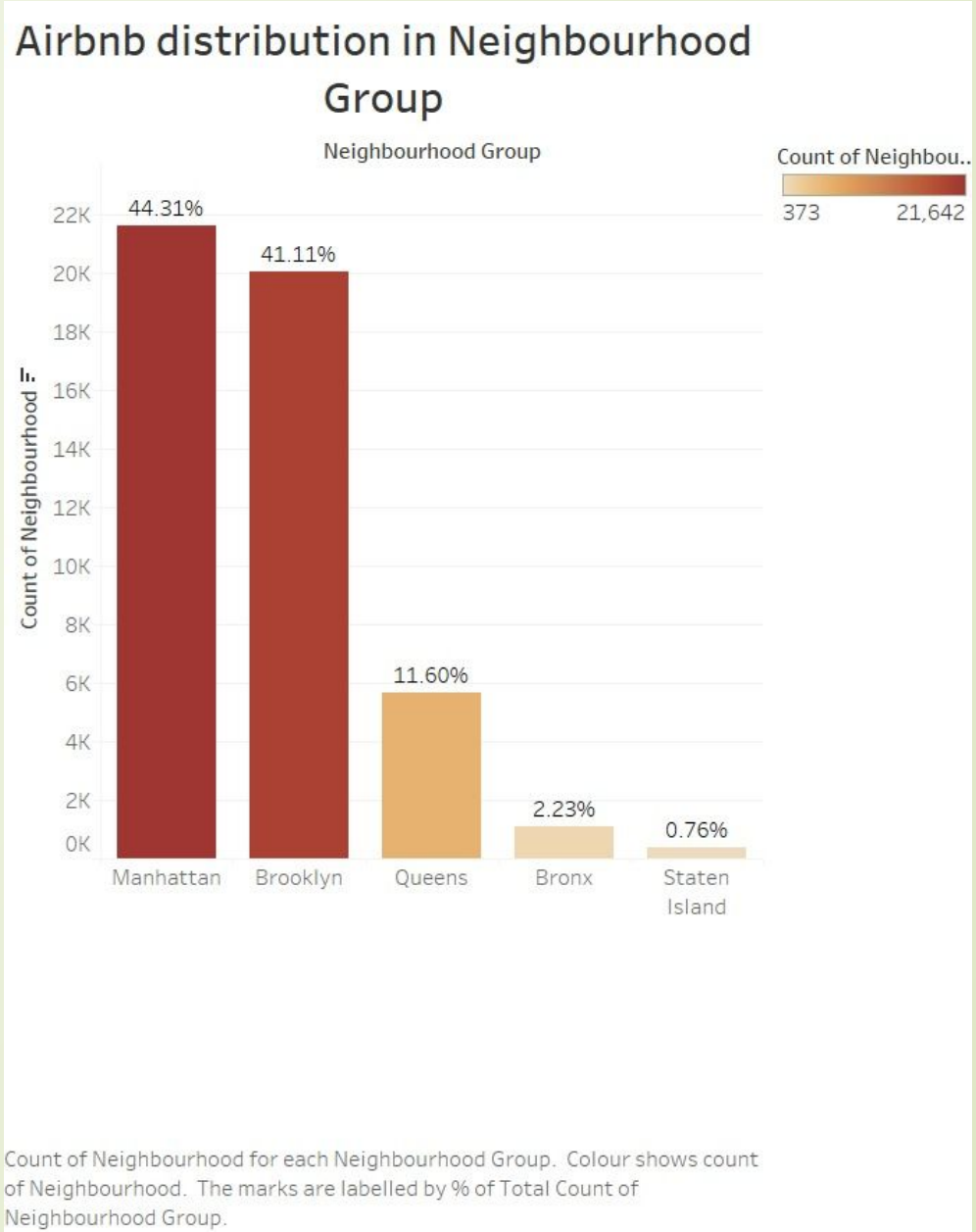
Top 10 hosts



- Among the 10 list of hosts for Airbnb seems Michael being the top most followed by David and to the end at 10 position its Maria.

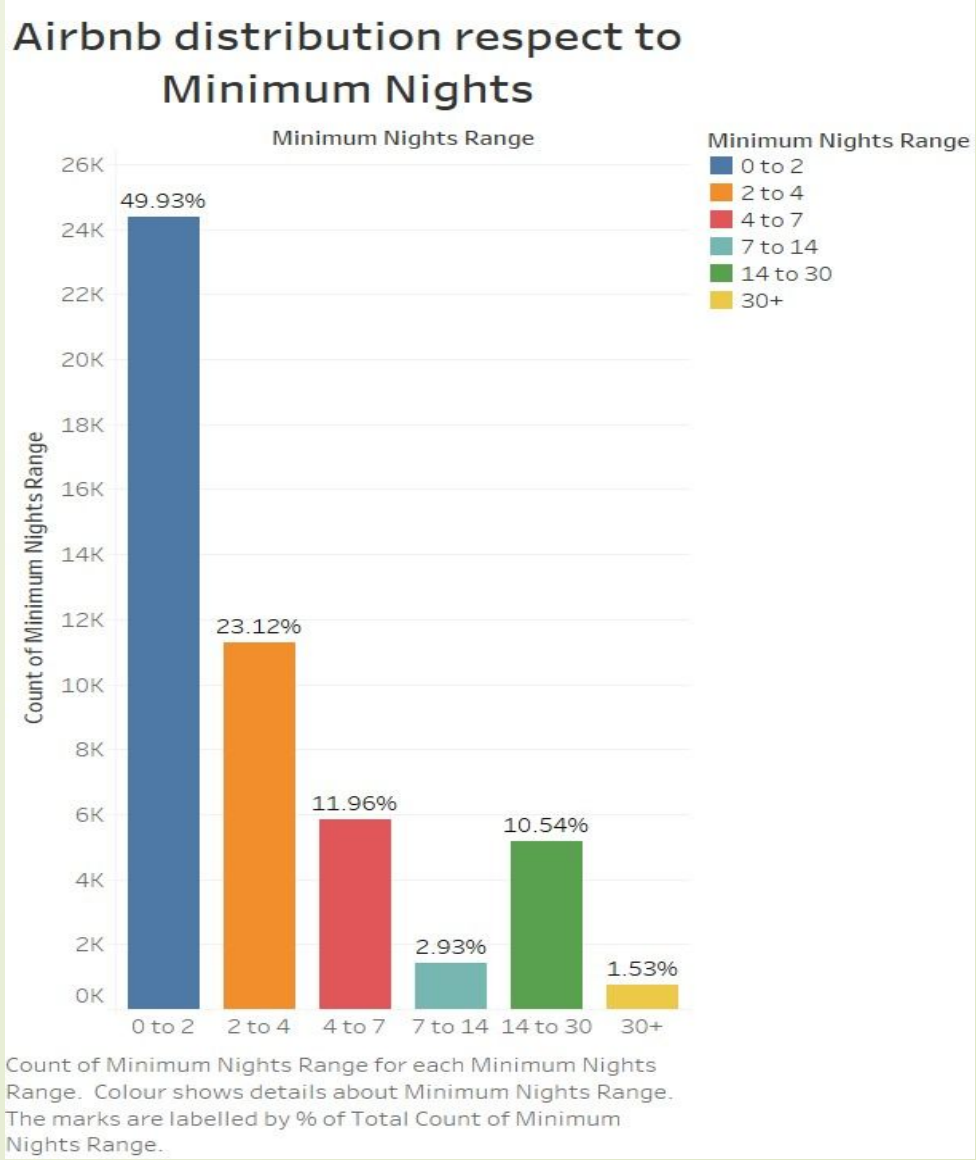
Popular Neighbourhood Group

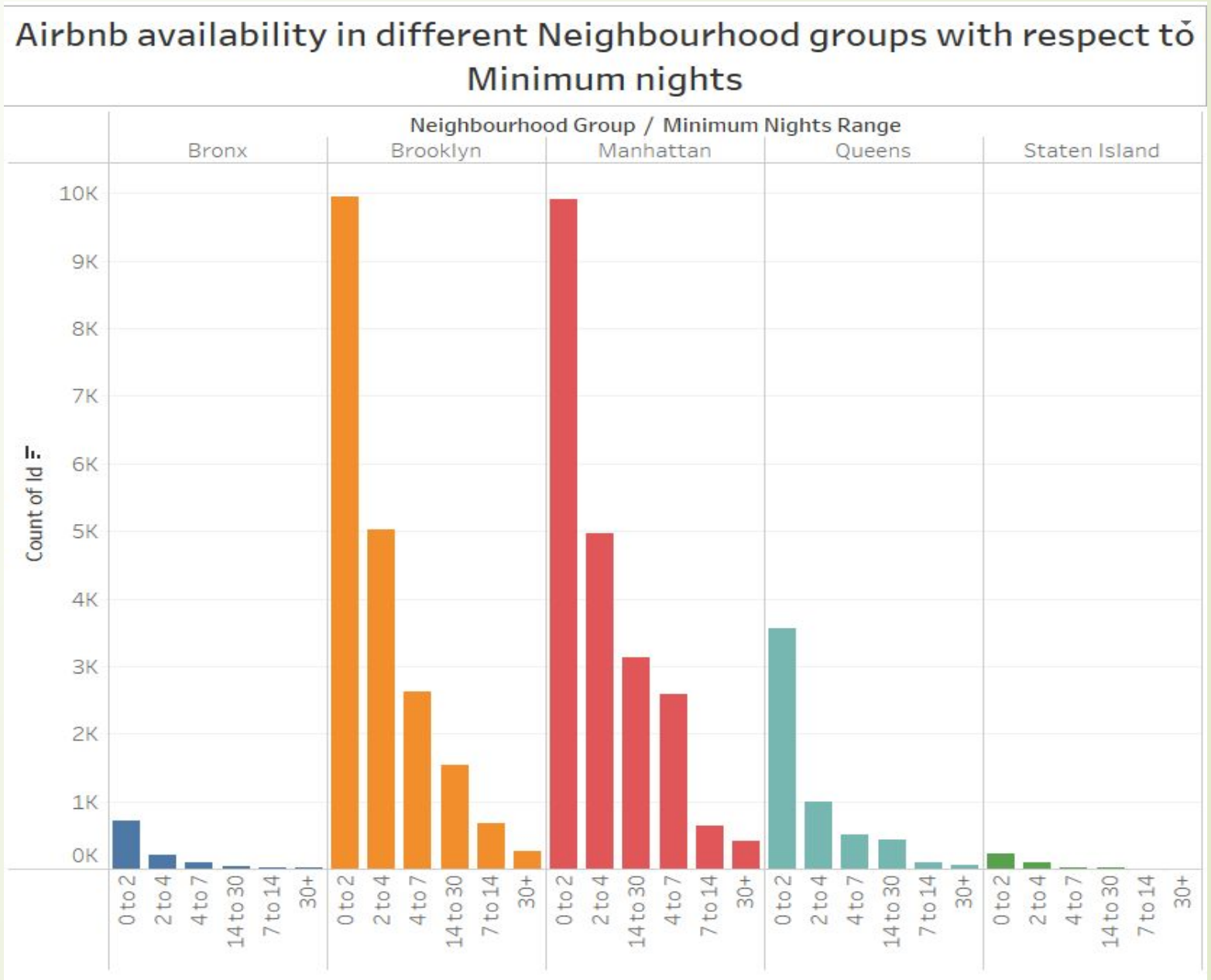
- Manhattan & Brooklyn have the highest share of Airbnb listings in the New York.
- Staten island has the least number of listings.
- Queens is third preferred airbnb spot followed with one third being preferred in Bronx.



Customer booking w.r.t minimum nights

- A very huge hike rate could be seen for Airbnb booking for 0-2 days listing out 2-4 days at just 50 % of previous one.
- 7 -14 & 20+ has a very ower bar followed with 4-7 & 14 - 20.

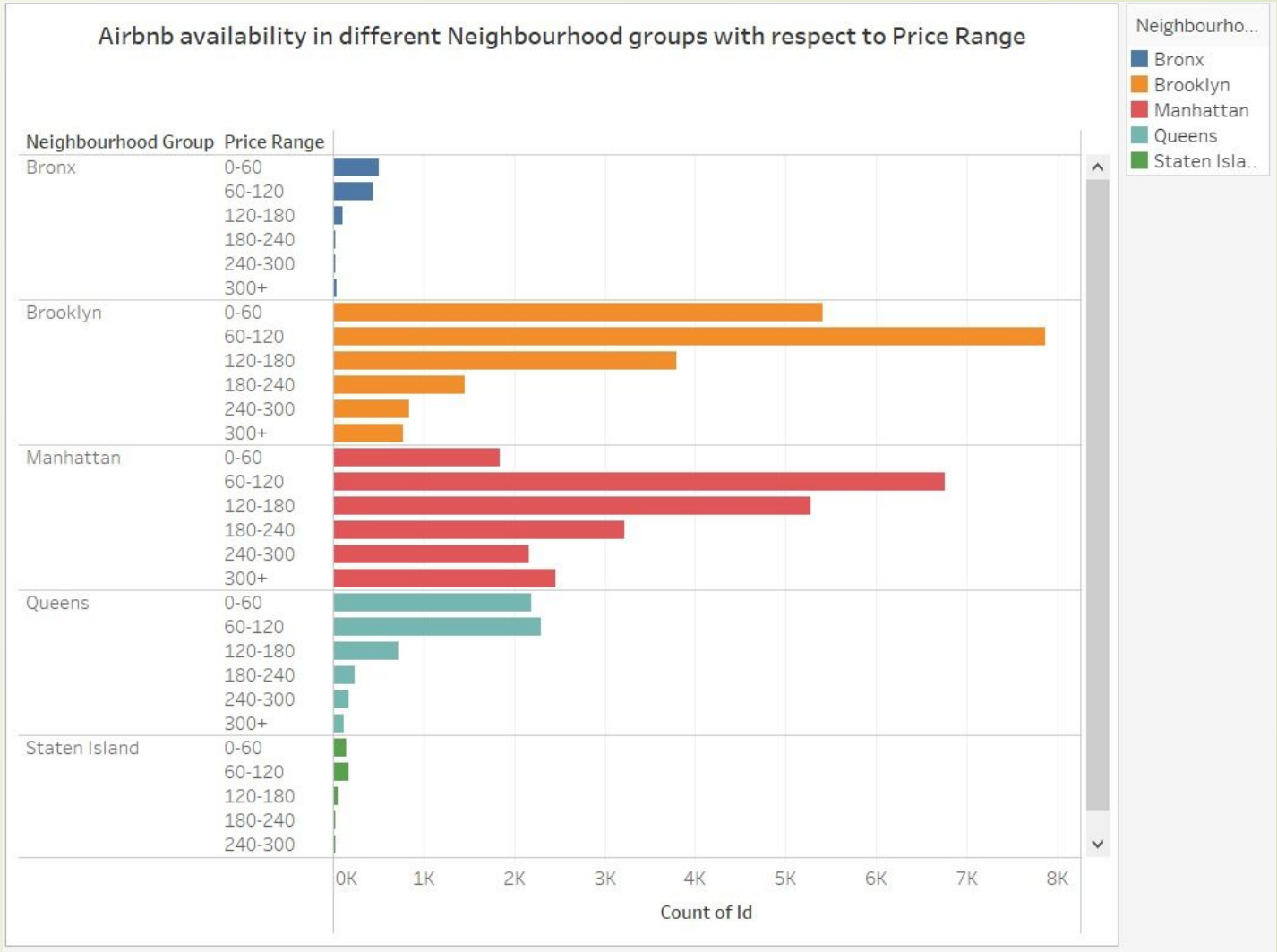




- Most of the listings in all the different Neighbourhood groups have the Minimum nights requirement to be between 0 to 4 days.

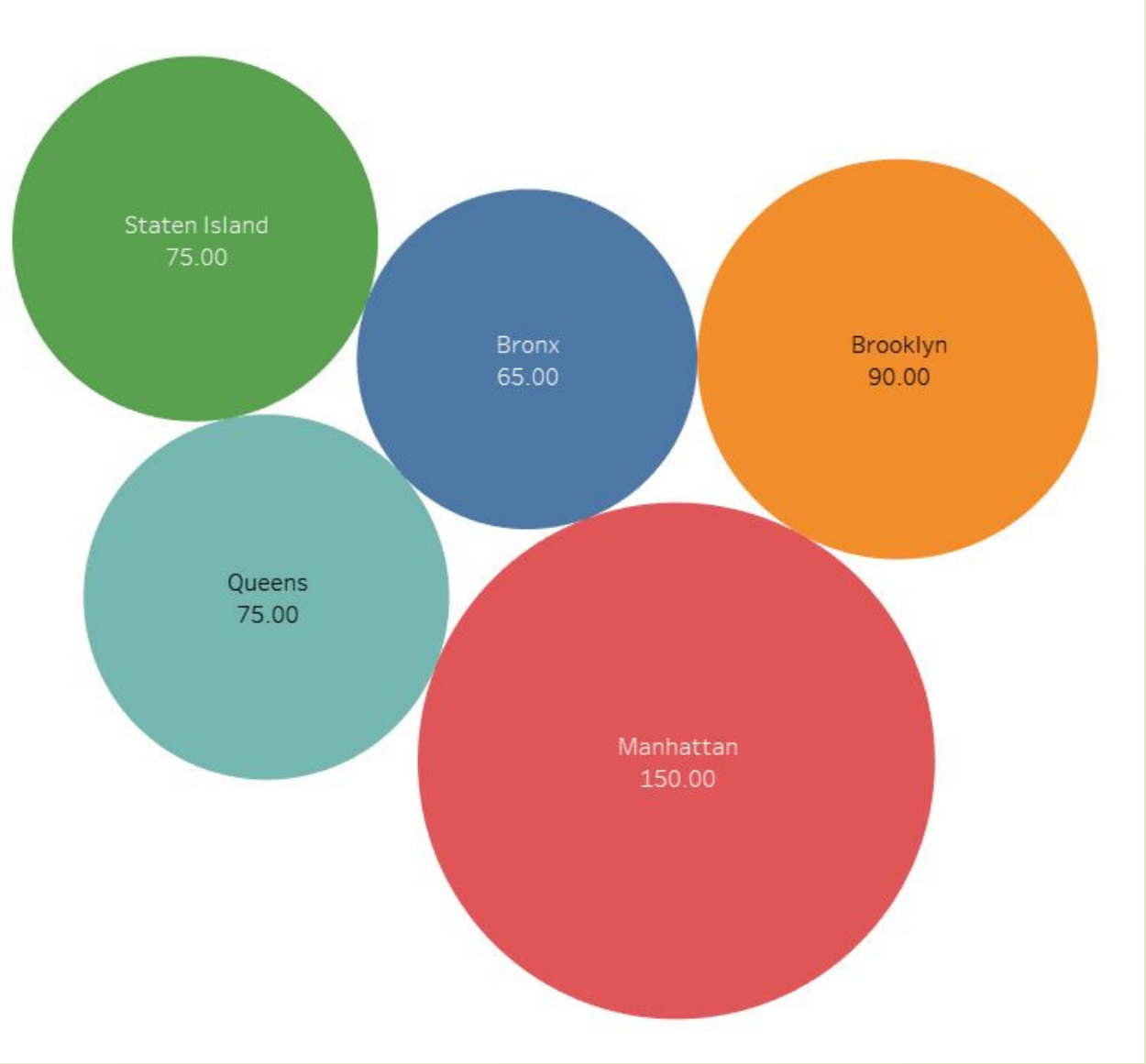
Neighbourhood vs Availability w.r.t minimum nights

- From graph , maximum availability for rooms is in Brooklyn & Manhattan. Manhattan & Queens 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 friendly booking.
- However in both the neighbourhood groups maximum price range preferred is 60 - 180.



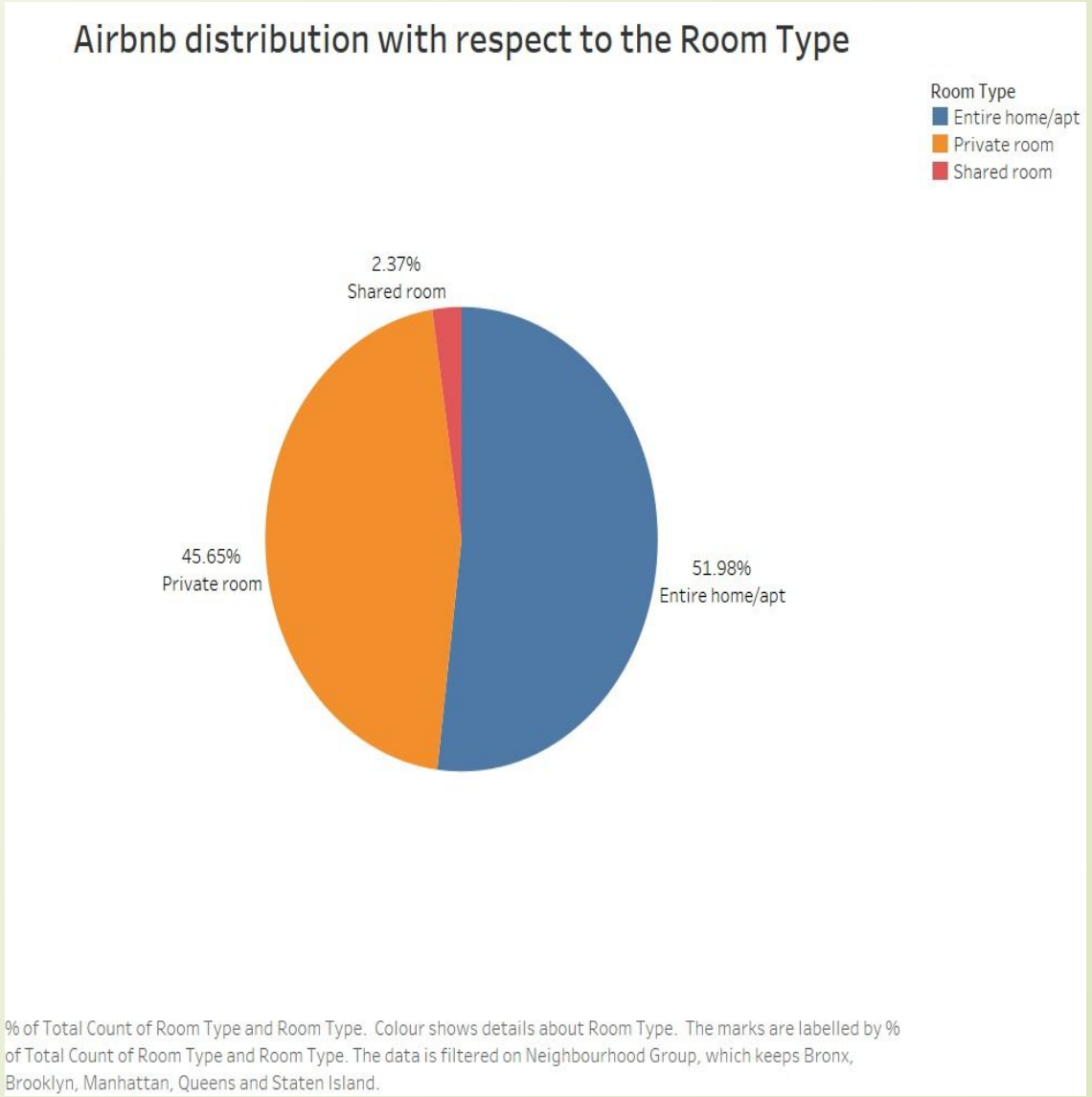
Median price of the different Neighbourhood Groups

- From the data we see that Manhattan has higher median price than the other Neighbourhood Groups.
- Brooklyn has the 2nd highest median price.
- Bronx is the least.



Preferred room type w.r.t neighbourhood group

- From chart seems the private and Entire home occupies is more preferred by customer than share room i.e 2.22%.



Appendix- DATA SOURCES:

- Dataset Overview: The dataset contains Airbnb listings' details, including hosts, locations, prices (per night), and various attributes.
- Column Explanations: Columns are self-explanatory; reference the provided diagram for specific meanings.

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

Appendix - Data Methodology

Conducted a thorough analysis on the given Airbnb dataset. The process included:

- Cleaning the dataset using Python with the help of Pandas library.
- Created new columns to convert numerical columns to categorical columns which helps in providing more precise information.
- Created the visualization charts using Tableau.