Storytelling Case Study: Airbnb, NYC

Presentation - I

BY,

AISHWARYA PRADEEP, MANISHA DASH & JAYANTI KUMARI

OBJECTIVE

People can utilize the web rental service Airbnb to market their vacant properties.

Airbnb suffered a significant revenue loss during the covid time.

Since people have started travelling once more, Airbnb is working to revive the industry and get ready to serve clients.

BACKGROUND

The revenue of Airbnb has significantly decreased during the last few months.

Airbnb wants to make sure that company is completely ready for this transformation now that the limits have started to loosen and people have started to travel more.

As a result, analysis has been done on a dataset made up of different New York Airbnb listings.

TOOLS USED

- Jupyter Notebook
- Tableau
- MS PowerPoint
- MS Word

DATA PREPARATION

- Understanding columns and datatypes

```
In [10]: # Check the Columns and datatypes
         airbnb_data.info()
```

```
 #   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
```

DATA PREPARATION

- Describing the dataset

In [11]:
```
# Describe the dataset
airbnb_data.describe()
```

Out[11]:

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_lis
count	4.889500e+04	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	38843.000000	4
mean	1.901714e+07	6.762001e+07	40.728949	-73.952170	152.720687	7.029962	23.274466	1.373221	
std	1.098311e+07	7.861097e+07	0.054530	0.046157	240.154170	20.510550	44.550582	1.680442	
min	2.539000e+03	2.438000e+03	40.499790	-74.244420	0.000000	1.000000	0.000000	0.010000	
25%	9.471945e+06	7.822033e+06	40.690100	-73.983070	69.000000	1.000000	1.000000	0.190000	
50%	1.967728e+07	3.079382e+07	40.723070	-73.955680	106.000000	3.000000	5.000000	0.720000	
75%	2.915218e+07	1.074344e+08	40.763115	-73.936275	175.000000	5.000000	24.000000	2.020000	
max	3.648724e+07	2.743213e+08	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.500000	

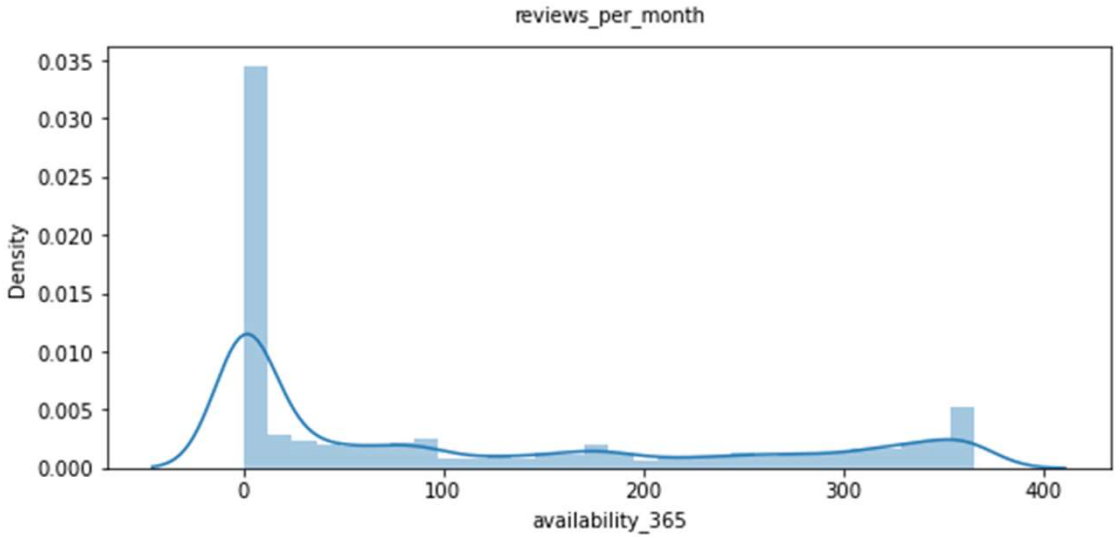
DATA PREPARATION

- Assessing the null values

```
In [7]: # Calculating the missing values in the dataset
        airbnb_data.isnull().sum()
```

```
Out[7]: 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
```

DISTRIBUTION OF ROOM AVAILABILITY



Listings showing 0-10 days availability should be verified and their hosts should be interviewed to find out the cause .

NEW FIELDS CREATED

- Price range

Price Range

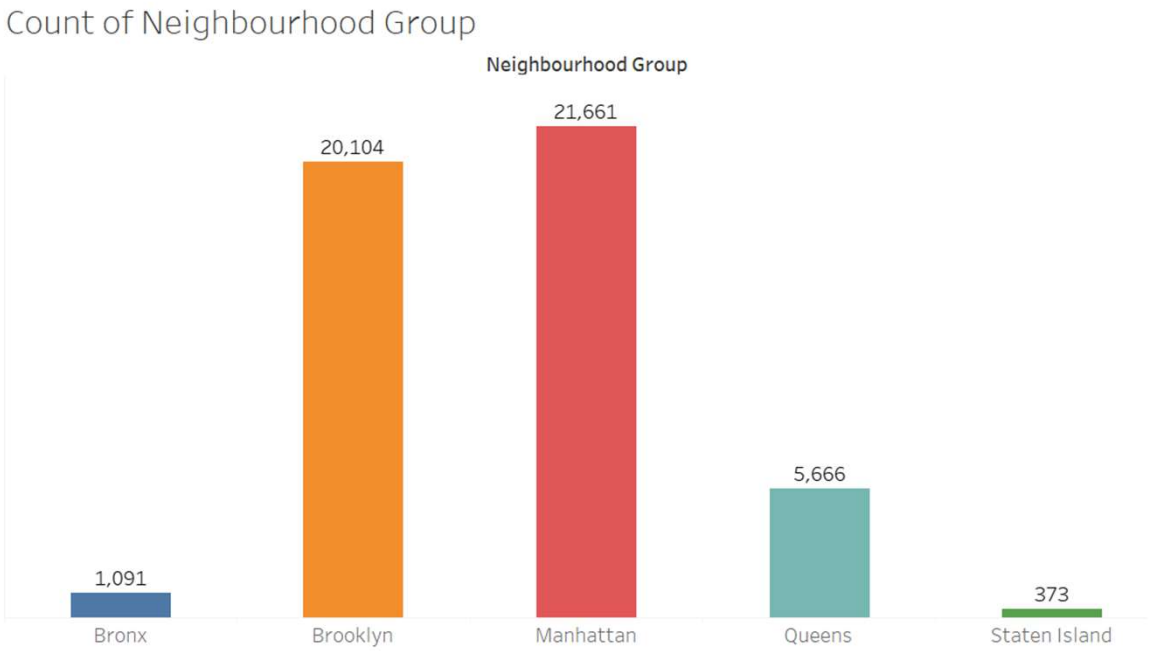
```
IF [Price] < 100 THEN "LOW"
ELSEIF [Price] >= 100 AND [Price] < 1000 THEN "MEDIUM"
ELSEIF [Price] >= 1000 AND [Price] < 5000 THEN "HIGH"
ELSE "VERY HIGH"
END
```

- Minimum night range

Minimum Nights Rar

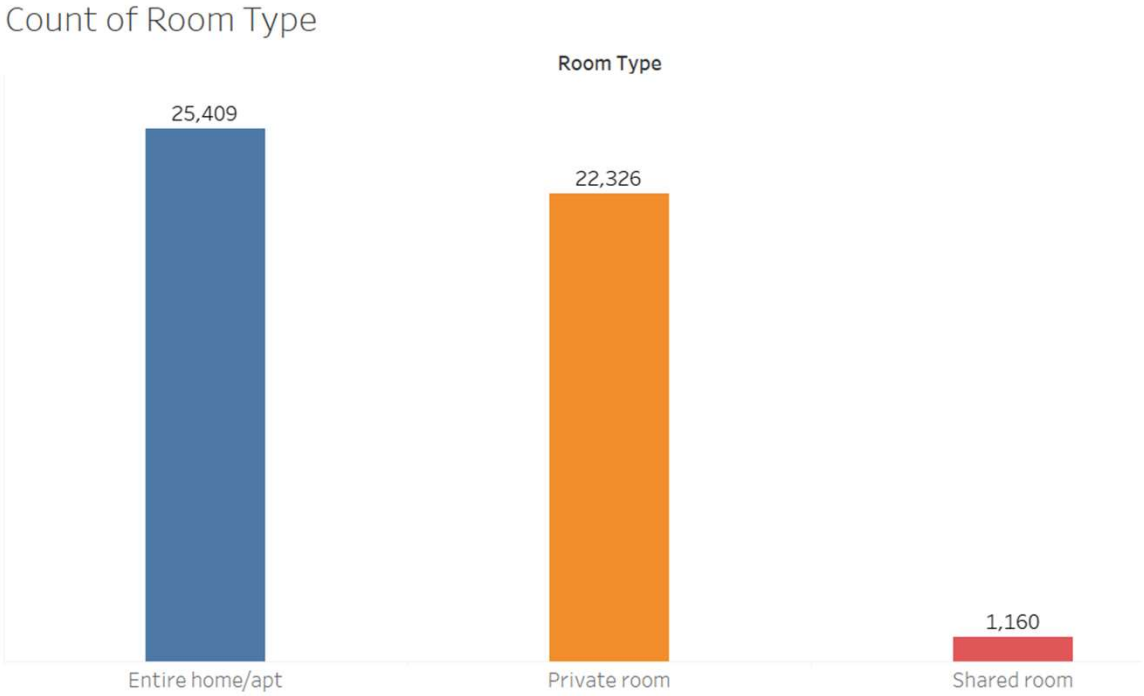
```
IF [Minimum Nights] < 8 THEN "Less than a week"
ELSEIF [Minimum Nights] >=8 AND [Minimum Nights] < 15 THEN "1-2 Weeks"
ELSEIF [Minimum Nights] >=15 AND [Minimum Nights] < 31 THEN "2-4 Weeks"
ELSEIF [Minimum Nights] >=31 AND [Minimum Nights] < 91 THEN "1-3 Months"
ELSEIF [Minimum Nights] >=91 AND [Minimum Nights] < 365 THEN "3 Months - 1 Year"
ELSE "More than a Year"
END
```

Most Booked Neighbourhood Groups



Most people prefer to stay in Manhattan (which constitutes almost 44% of the bookings) or Brooklyn (which constitutes almost 41% of the bookings) and Bronx and Staten Island are least preferred

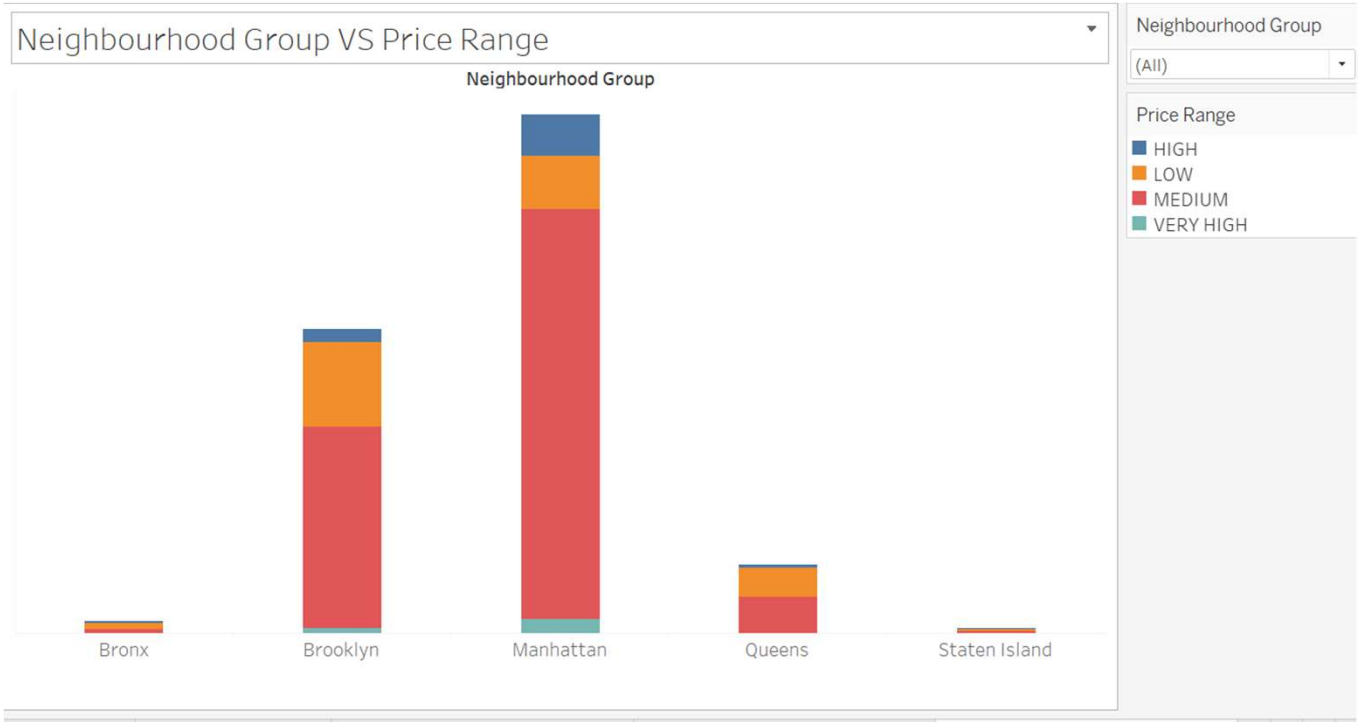
Most Booked Neighbourhood Groups



Almost 52% of the bookings were for an Entire home or apartment and 45% were for a private room.

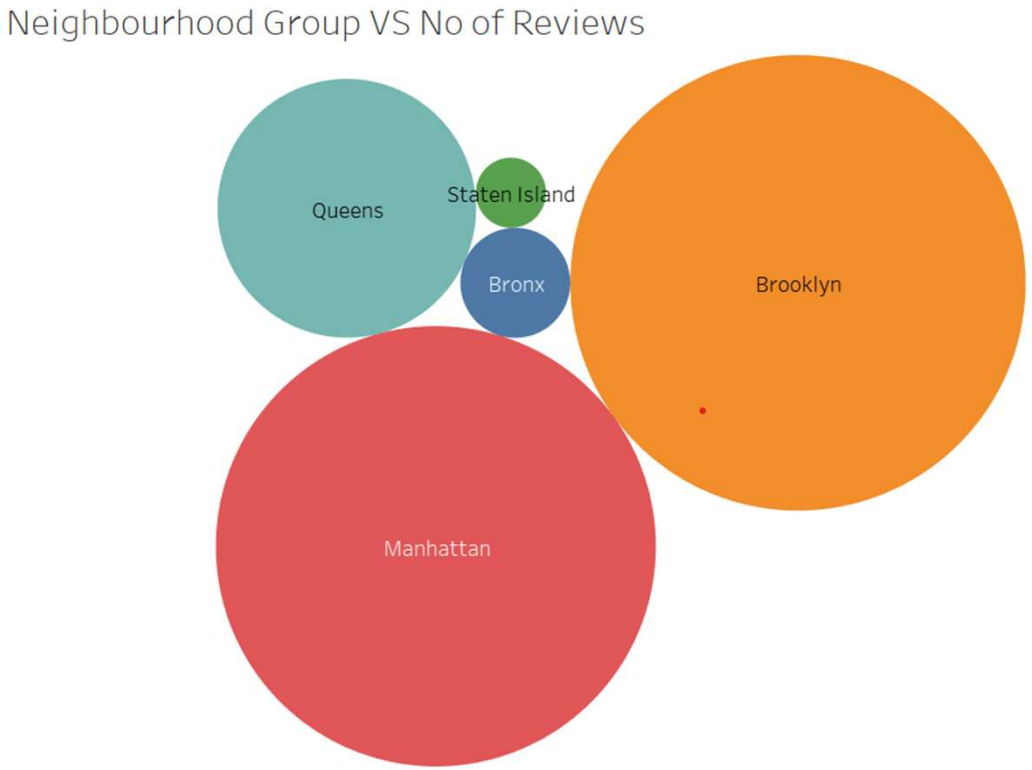
Very few people prefer to stay in Shared rooms.

Most Booked Neighbourhood Groups



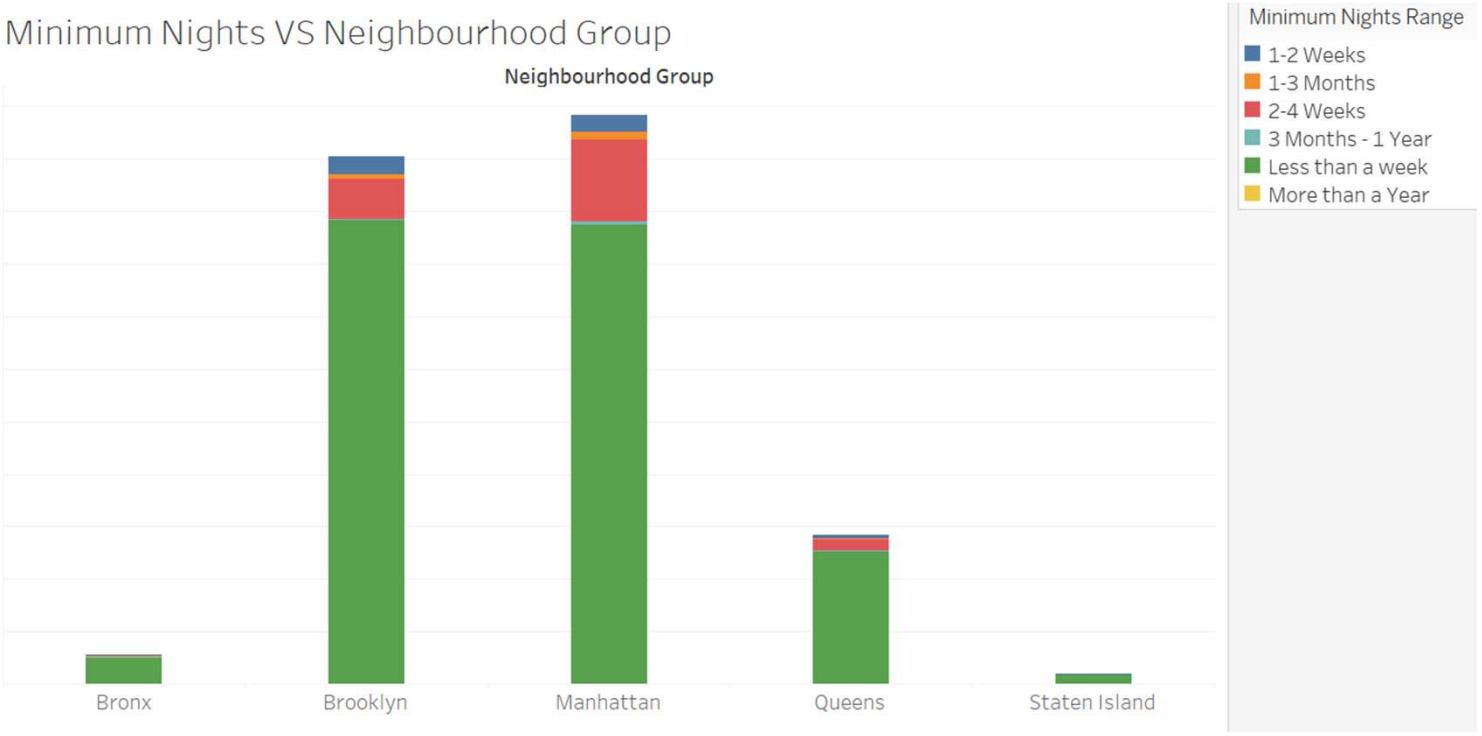
By a very high margin, people prefer to go by medium priced places. This applies to all the neighborhood groups. Very high priced places are rarely booked.

Most Booked Neighbourhood Groups



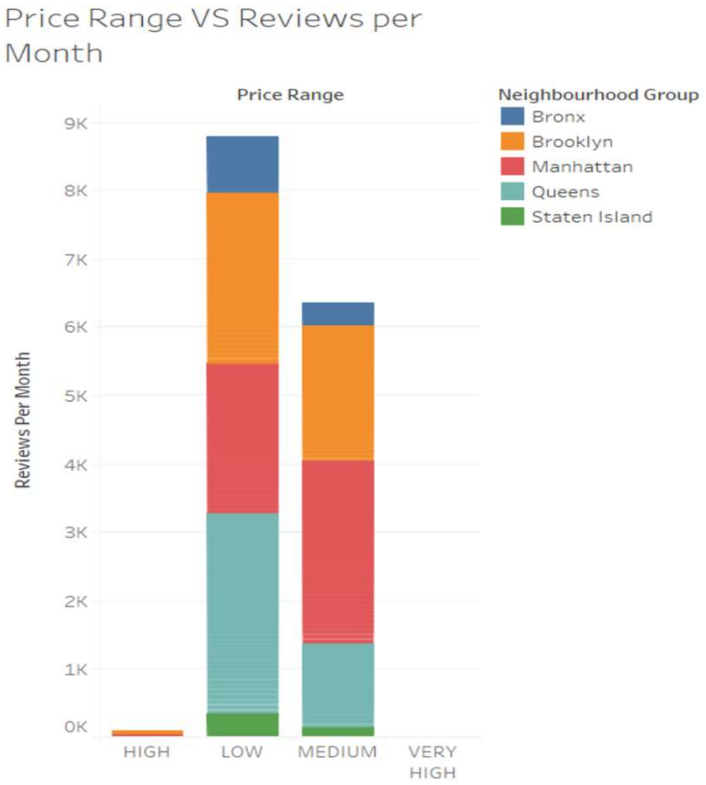
Most of the reviews are for bookings in Manhattan and Brooklyn. But there are more than 10000 records without any reviews.

Most Booked Neighbourhood Groups



Bookings are usually for less than a week duration. Brooklyn and Manhattan have quite few bookings for 2-4 weeks. Customers rarely book for more than a year.

Price Range VS Reviews Per Month



Low price range has received more reviews.

CONCLUSION

- The dataset is not in the best condition as there are 10000+ values missing for ‘Last_review’ and ‘reviews_per_month’.
- The columns like Room_type and Neighborhood_group are highly imbalanced.
- Manhattan and Brooklyn are the most popular neighbourhoods. So they should be targeted more.
- Private room or homes are preferred more by the customers with a medium price range. So such places should have more room availability.
- Most bookings are for a short period but some bookings are for 2-4 weeks. So customers should be provided with good services for such long stays.

TABLEAU DASHBOARD

Tableau dashboard for the case study

https://public.tableau.com/app/profile/aishwarya.pradeep/viz/AirBnBdashboard_16834628803670/Dashboard1?publish=yes