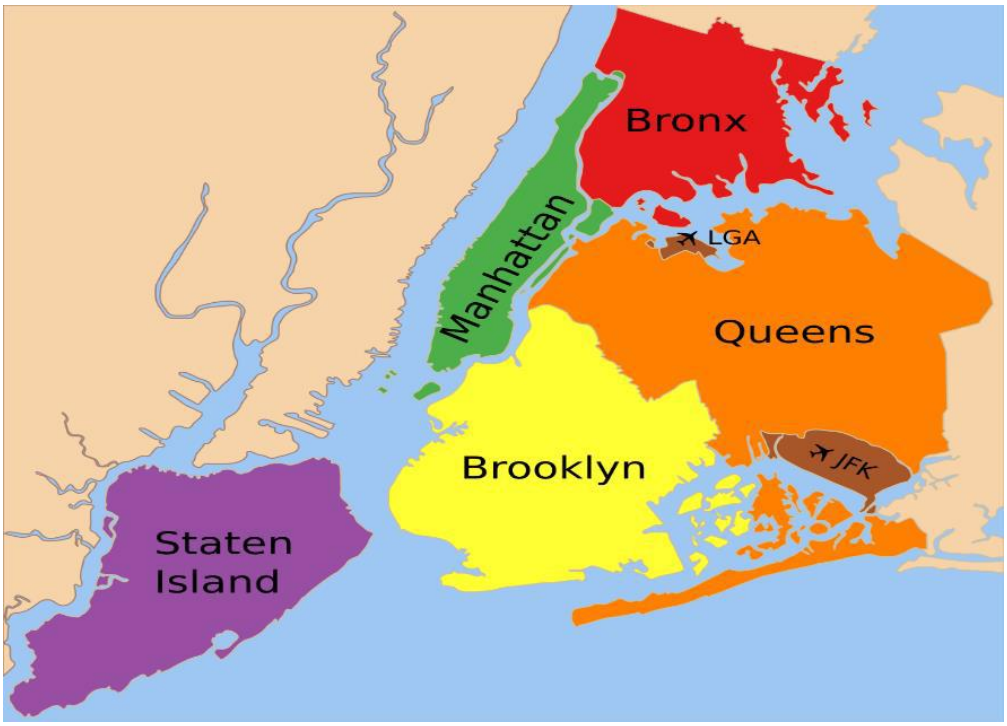
AIRBNB Case Study Methodology

Introduction

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb provides platform for hosts to accommodate guests with short-term lodging and tourism-related activities.

New York is the most diverse and populated 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 centre 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. 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.

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

In [1]:
```
#importing liabraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
#reading the data
bnb=pd.read_csv('AB_NYC_2019.csv')
```

In [3]:
```
bnb.head()
```

Out[3]:

	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	

In [4]:
```
len(bnb)
```

Out[4]:
```
48895
```

In [5]:
```
# columns datatypes in dataframe
bnb.dtypes
```

Out[5]:
```
id                                  int64
name                               object
host_id                             int64
host_name                          object
neighbourhood_group                object
neighbourhood                      object
latitude                          float64
longitude                         float64
room_type                          object
price                               int64
minimum_nights                      int64
number_of_reviews                   int64
last_review                        object
reviews_per_month                 float64
calculated_host_listings_count      int64
availability_365                    int64
dtype: object
```

```
In [8]: #checking the missing value percentage
        bnb.isna().mean()*100
```

```
Out[8]: 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
```

```
In [9]: # dropping last_review because of high missing values
        bnb.drop("last_review",axis=1,inplace=True)
```

- We removed the columns ‘Last Review’ which was not giving much information and replacing the column reviews_per_month null values with ‘0’ because of no reviews.

```
In [11]: # filling the reviews_per_month with 0 because of no reviews
         bnb.reviews_per_month=bnb.reviews_per_month.fillna(0)
```

```
In [12]: # checking for the null values after the imputation
         bnb.isna().sum()
```

```
Out[12]: 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
         reviews_per_month                 0
         calculated_host_listings_count    0
         availability_365                  0
         dtype: int64
```

Step 2: Data Wrangling:

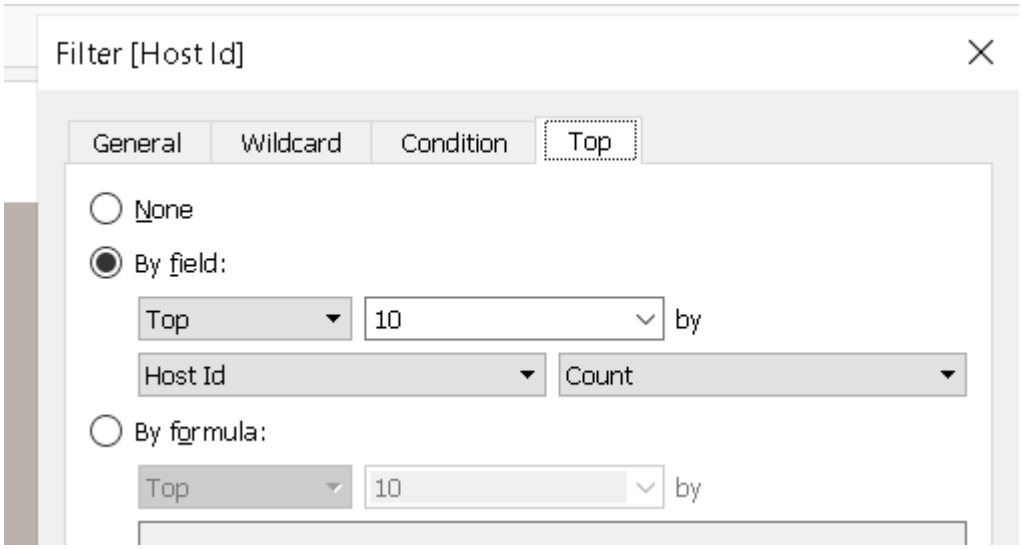
- Checked the Duplicate rows in our dataset and no duplicate data was found.
- Checked the Null Values in our dataset. Columns like name, host_name, last review and reviews_per_month have null values.
- We've dropped the column name as missing values are less and dropping it won't have significant impact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

Methodology Document PPT 1:

1) **Top 10 Host :**
   - We identified the top 10 Host Name with neighbourhood group and count of Host names using the bar chart.



2) **Preferred Room type w.r.t count of host listings:**
   - We created a pie chart for understanding the percentage of room type preferred wrt count host listing.
   - We added Room Type to the colors Marks card to highlight the different Room Type in different colors and count of Host Id to the size

3) **Most valued customers:**
   - we have created a bar chart for neighbourhood and host name with host listing count.

4) **Average price of room type per neighbourhood groups:**
   • We created a stacked bar chart with Neighbourhood Groups in Columns and Price column in Rows.
   •We added the room type to the colors Marks card to highlight the different room types in different colors. Also put Avg price in Label.

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

6) **Customer preference per night stay:**
   We have created bar chart, most of the customer visit either for 1 day or 2 days. But there is strange trend for 30 days also.

Methodology Document PPT 2 :

1) **Average Price w.r.t Neighbourhood group:**
   - We created a bubble chart for understanding the average price w r t neighbourhood group.
   - We added Room Type to the colors Marks card to highlight the different room Type in different colors and count of Host Id to the size.

2) **Neighbourhood vs Availability:**
- We created a bar chart for availability 365 for different neighbourhood.

3) **Distribution of properties wrt different areas.**
   The density of properties is high in Manhattan compare to others areas.

4) **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.

5) Popular Neighborhoods:

- We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in color.
- We used filter to show Top 20 neighbours as per the sum of reviews.

6) <u>Highest reviewed properties.</u>

- Bar chart has been created between properties with count of reviews.
- Neighbourhood group has been added for colors marks card.