Airbnb Case Study

Background

An American firm called Airbnb, Inc. runs an online marketplace for travel-related services and accommodations, especially homestays for holiday rentals. Airbnb gives hosts a platform to offer short-term housing and travel-related events for guests.

The most populous and diversified city in the US is New York City. Manhattan, Brooklyn, Queens, the Bronx, and Staten Island are among the five boroughs that make up the city; they were all grouped into one metropolis. It is widely acknowledged as the industry's global hub for financial services.

Problem Statement

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.

In the case study, we used Jupiter notebook to undertake preliminary data analysis and Tableau for data analysis and visualization.

Step 1

Importing Libraries

```
import warnings
warnings.filterwarnings("ignore")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

Data Conversion and Understanding

```
airbnb = pd.read_csv("AB_NYC_2019.csv")
airbnb.head(5)
```

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

No. of rows and columns of the dataset

```
airbnb.shape
```

```
(48895, 16)
```

- The dataset contains 48895 rows and 16 columns

Let's check if there are any missing values in the dataset

```
airbnb.isnull().sum()
```

```
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
```

After checking the missing values, lets drop column that are not required in the dataset

```
airbnb.drop(['id','name','last_review'], axis = 1, inplace = True)
```

```
airbnb.head(5)
```

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

Let's replace reviews per month containing missing values with a zero

```
airbnb.fillna({'reviews_per_month':0},inplace=True)
```

```
airbnb.reviews_per_month.isnull().sum()
```

```
0
```

- Now since there are no missing values in reviews per month column

Let's check the unique values in other columns

```
airbnb.room_type.unique()
```

```
array(['Private room', 'Entire home/apt', 'Shared room'], dtype=object)
```

```
len(airbnb.room_type.unique())
```

```
3
```

```
airbnb.neighbourhood_group.unique()
```

```
array(['Brooklyn', 'Manhattan', 'Queens', 'Staten Island', 'Bronx'],
      dtype=object)
```

```
len(airbnb.neighbourhood_group.unique())
```

```
5
```

```
len(airbnb.neighbourhood.unique())
```

```
221
```

Step 2: Data Wrangling

- No duplicate data was discovered when the Duplicate rows in our dataset were checked.
- Examined the dataset's null values.
- Name, host-name, last review, and review-per-month are examples of columns with null values.
- We removed the column name because there are fewer missing data and doing so won't significantly affect the analysis.
- Verified our dataset's formatting.
- Recognised and examined outliers.

Step 3: Data Analysis and visualization using Tableau

To visualise the data for the assignment, we utilised Tableau. The specific steps taken for each visualisation are listed below.

AirBnblisting distribution in NYC

The concentration of listings in each neighbourhood was visualised using a **BAR PLOT**. The variables Neighbourhood group and Id are used.

Room Type Analysis & Room Type Analysis w.r.t neighbourhoodgroup

To visualise the preference for a certain room type among the neighbourhood group, we built a **PIE CHART**.

Customer Booking w.r.t Minimum Nights

For Minimum nights, we designed the **BIN**.

The bins were used to represent the distribution of minimum nights based on the quantity of IDs reserved for each neighbourhood group.

Popular Neighbourhood

We organised neighbourhoods into rows, the total number of reviews into columns, and neighbourhood groups into colours.

We used a filter to display the top 20 neighbours according to the overall rating.

Price range preferred by customer

In order to construct a **BAR CHART**, we used pricing preference based on the quantity of reservations made in a given price range and the number of IDs. We have a $20 interval bin for the Price column.

Host with highest listing w.r.t Neighbourhood

This was examined in order to determine the maximum number of listings held by a single host and in which location. This would provide us with information on how the hosts are investing and expanding in a particular area.

To analyse the number of bookings, we used the Host ID in the x-axis and the Id in the y-axis. Because there were so many Host IDs, we narrowed it down to the top ten. The graph was color-coded according to the neighbourhood group.

Average price of Neighbourhood groups

Using Neighbourhood Groups as Columns and the Price column as Rows, we created a **BUBBLE CHART**. To highlight the various neighbourhood Groups in various colours, we included the Neighbourhood Groups to the Colour Marks card. Label the average price as well.

Neighbourhood Availability Analysis

We used a **BAR CHART** for availability (365 days) and a line chart for pricing (top 10 neighbourhood group, sorted by price) to build a dual axis graphic.

Tool Used

- Data cleaning and preparation: Jupyter notebook – Python
- Visualization and analysis: Tableau
- Data Storytelling: Microsoft PPT