# *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*

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib as plt
4 import seaborn as sns
5
```

```
1 df=pd.read_csv("AB_NYC_2019.csv")
2 df.head()
```

| | id | name | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | |

```
1 df.shape
```

```
(48895, 16)
```

```
1 df.dtypes
```

```
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
```

```
1 df.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
```

```
##Dropping colums that are insignificant for analysis-id,name and last_reviews
df.drop(["id","name","last_review"],axis=1,inplace=True)
```

```
df.head()
```

| | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_reviews | reviews_per_m |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2787 | John | Brooklyn | Kensington | 40.64749 | -73.97237 | Private room | 149 | 1 | 9 | |
| 1 | 2845 | Jennifer | Manhattan | Midtown | 40.75362 | -73.98377 | Entire home/apt | 225 | 1 | 45 | |
| 2 | 4632 | Elisabeth | Manhattan | Harlem | 40.80902 | -73.94190 | Private room | 150 | 3 | 0 | |
| 3 | 4869 | LisaRoxanne | Brooklyn | Clinton Hill | 40.68514 | -73.95976 | Entire home/apt | 89 | 1 | 270 | |
| 4 | 7192 | Laura | Manhattan | East Harlem | 40.79851 | -73.94399 | Entire home/apt | 80 | 10 | 9 | |

```
#replacing nan value of reviews_with_month by 0
df.fillna({"reviews_per_month":0},inplace=True)
```

```
df.isnull().sum()
```

```
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
```

```
#checking the unique values
df.neighbourhood_group.unique()
```

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

```
df.room_type.unique()
```

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

- ***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 review_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.***

## 1) Top 10 Host :

· We identified the top 10 Host Ids , Host Name with count of Host Ids using the tree map.

## 2) Preferred Room type w.r.t Neighbourhood group:

· We created a pie chart for understanding the percentage of room type preferred 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

## 3) For Variance of price with Neighbourhood Groups:

· We used a box and whisker's plot with Neighbourhood Groups in Columns and Price in Rows.

· We changed the Price from a Sum Measure to the median measure.

**4) Average price of Neighbourhood groups:**

• We created a bubble chart with Neighbourhood Groups in Columns and Price column in Rows.

•We added the Neighbourhood Groups to the colors Marks card to highlight the different neighbourhood Groups in different colors. Also Put Avg price in Label

# *Methodology Document PPT 2 :*

1**) Preferred Room type w.r.t Neighbourhood group**:

· We created a pie chart for understanding the percentage of room type preferred 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) **Customer Booking w r t minimum nights :**

· We created the bin for Minimum nights as shown below.

· The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.

**3) Neighbourhood vs Availability:**

· We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

## 4) Price range preferred by Customers:

· We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of $20.

## 5) Understanding Price variation w.r.t Room Type & Neighbourhood:

· We created Highlights Table chat by taking Room Type in rows & Neighbourhood Group in column.

· We took the average price in colour Marks card to highlight the different Room Type in different colors.

## 6) 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.

## 7) 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.

# Thankyou