**UNVEILING THE SECRETS OF AIRBNB IN NYC**
**METHEDOLOGY**

Dipjit Basak
Darsi Noel Jeffrey

BA || DS60
Upgrad IIIT Bangalore Executive PG Program in Data Science

Methodology Document

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

**Step 1: Importing library, Data Correction, Data Wrangling:**

- Importing different library for analysis
- 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.

**Step 2: 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 with respect to Neighborhood group:
   - We created a pie chart for understanding the percentage of room type preferred w.r.t. neighborhood group
3> Customer Booking w r t minimum nights:
   - We created the bin for Minimum night
4> Understanding Price variation w.r.t Room Type & neighborhood:
   - We created Highlights Table chat by taking Room Type in rows & neighborhood group in column.

- We took the average price in color Marks card to highlight the different Room Type in different colors.

5> Categorization w r t Geography:
- We used Geo location chart to plot neighborhood, neighborhood Group in map.

6> Popular Neighborhoods:
- We took neighborhood in rows and sum of reviews in column and took neighborhood groups in color.
- We used filter to show Top 20 neighbors as per the sum of reviews.

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