AIRBNB Case Study IIIT-B   
Submitted by   
Snehil Tiwari, Sushma Mahagaonkar, and Rahul Patra

**Methodology Document for Presentation 1:**

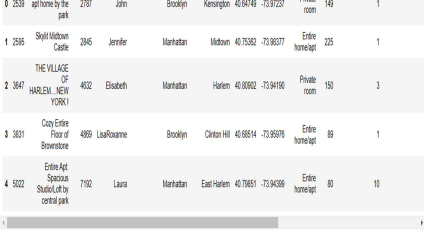
For the case study, we employed Jupiter Notebook to conduct the initial data analysis, and Tableau for both data analysis and visualization purposes.

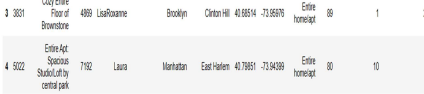
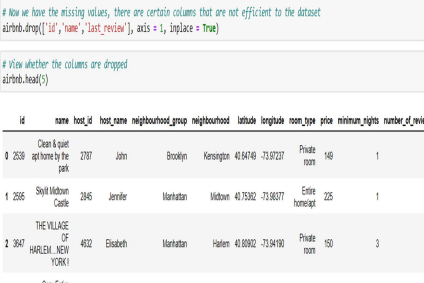
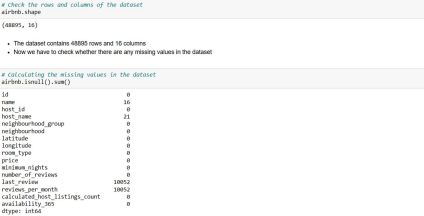
Initial Analysis using Jupiter Notebook:

- Dataset Used: AB\_NYC\_2019.csv

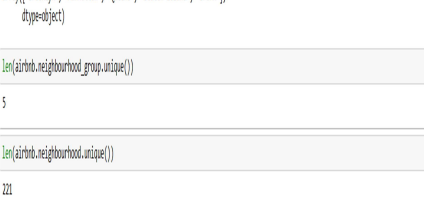
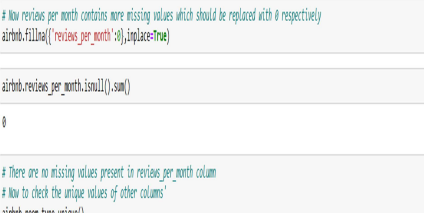
- Number of Rows: 48,895

- Number of Columns: 16





* We eliminated columns such as Id, Name, and Last Review, as they provided limited information and were deemed unnecessary for our analysis.

  
  
**Step 2: Data Wrangling**

1. We examined the dataset for duplicate rows and confirmed that no duplicate data was present.
2. Null values were assessed in our dataset, specifically in columns such as name, hostname, last review, and review-per-month.
3. After considering the impact on the analysis, we made the decision to drop the "name" column due to the relatively low number of missing values, which would not significantly affect the analysis.
4. The dataset underwent a comprehensive check to ensure proper formatting and consistency.
5. We identified and addressed outliers in the review data.

**Data Analysis and Visualizations using Tableau:**

For the assignment, we utilized Tableau to create visualizations of the data. Here are

the step-by-step details for each visualization:

1) Top 10 Hosts:

* Using a Tree map visualization, we determined the top 10 Host IDs and Host names based on the count of host IDs.



2) Preferred Room Types by NeighbourhoodGroup:

* Pie chart visualization was used to analyze the percentage distribution of

preferred room types in relation to each Neighbourhood group.

* The "Room Type" field was assigned to the Color Marks card to differentiate room types using distinct colors and count of Host Id to the size.

3) Variance of price with Neighbourhood Groups:

* Utilized a box and whisker plot with NeighbourhoodGroups displayed on the Columns axis and Prices on the Rows axis.
* Converted the Price measure from a Sum Measure to a median measure to capture the central tendency more effectively.

4) Average price of Neighbourhood groups:

* Generated a bubble chart featuring NeighbourhoodGroups on the Columns axis and Price columns on the Rows axis.
* Assigned distinct colors to the NeighbourhoodGroups by adding them to the colors Marks card.
* The Avg price was also displayed on the Label.

5) Customer Booking with respect to minimum nights:

* Bins were created to categorize the minimum nights as shown below.
* The distribution of minimum nights was visualized using these bins, showcasing the number of IDs booked for each Neighbourhood group.



6) Popular Neighborhoods:

* Neighborhoods were arranged in rows, and the sum of reviews was displayed in columns, while the Neighbourhood groups were distinguished by color.
* A filter was applied to present the Top 20 neighborhoods based on the sum of reviews.

7) Neighbourhood vs Availability:

* A dual-axis chart was generated, combining a bar chart for Availability 365 and a line chart for price.
* The top 10 Neighbourhood groups, sorted by price, were selected for this visualization.

**Methodology Document for Presentation 2**:

1) Room type with respect to Neighbourhood group:

* A pie chart was generated to analyze the percentage distribution of preferred room types in relation to each Neighbourhood group.
* Room Type was assigned to the color Marks card, enabling distinct colors for different room types, while the count of Host Ids determined the size of each section.

2) Customer Booking with respect to minimum nights:

* Bins were created to categorize the minimum nights, as illustrated below.
* The distribution of minimum nights, based on the number of IDs booked for each Neighbourhood group, was visualized using these bins.



3) Neighbourhood vs Availability:

* A dual-axis chart was created, featuring a bar chart for Availability 365 and a line chart for price.
* This visualization focused on the top 10 Neighbourhood groups, sorted by price.

4) Price range preferred by Customers:

* A bar chart was generated to examine the pricing preferences based on the volume of bookings in different price ranges.
* The Price column was binned with a $20 interval to facilitate the visualization.

5) Understanding Price variation with respect to Room Type & Neighbourhood:

* A Highlights Table chart was created, with Room Type in rows and Neighbourhood Group in columns.
* The average price was used as the color Marks card to differentiate room types using distinct colors.

6) Price variation with respect to Geography:

* A Geo Location chart was utilized to plot the Neighbourhood and Neighbourhood Groups on a map, showcasing the variation in prices across different locations.

7) Popular Neighborhoods:

* Neighbourhood were arranged in rows, and the sum of reviews was displayed in columns, while the Neighbourhood Groups were distinguished by color.
* A filter was applied to highlight the top 20 Neighbourhood based on the sum of reviews.

8) Tools used:

* For data cleaning and preparation: Jupyter Notebook with Python.
* For visualization and analysis: Tableau.
* For data storytelling: Microsoft PowerPoint.