Methodology

AIRBNB Case Study – Sonal Khot, Abhinandan Gupta, Amruta Patil

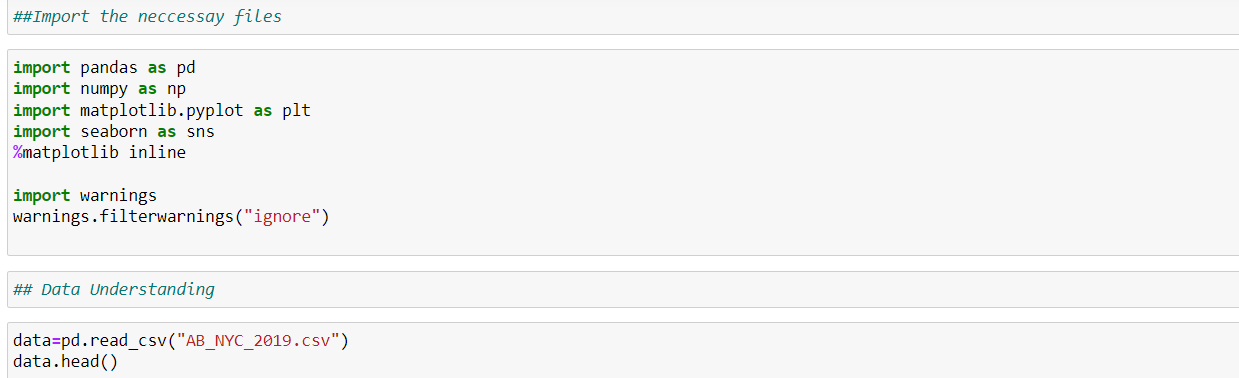
Methodology Document PPT 1:

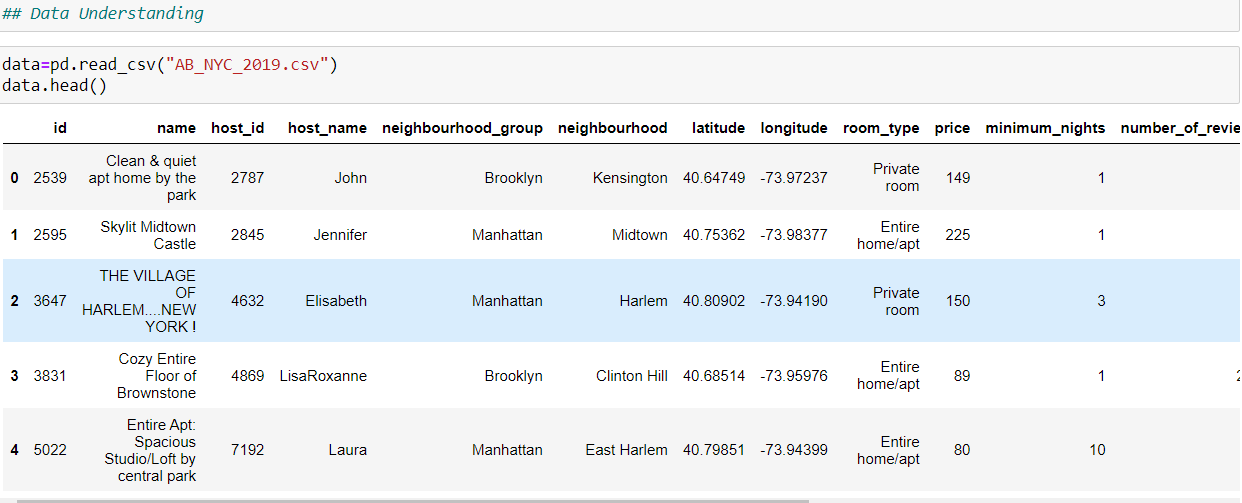
In this case study, we utilized Jupyter Notebook for the initial data analysis and preprocessing, and Tableau for in-depth data analysis and visualization to derive actionable insights.

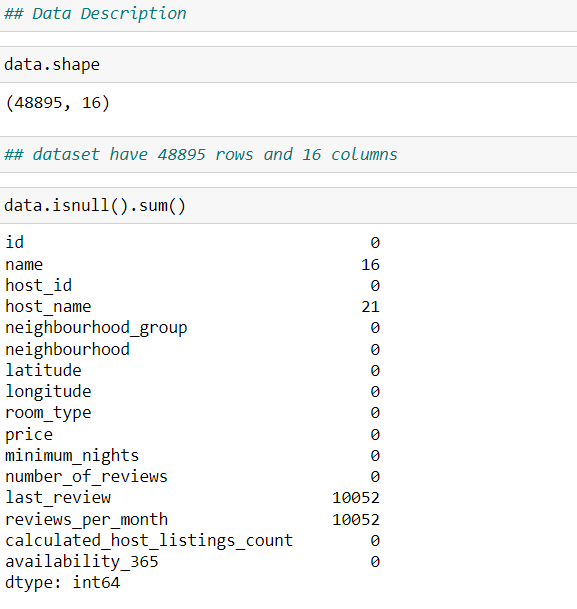
**Initial Analysis using Jupiter Notebook:** Data Set Used: AB\_NYC\_2019.csv

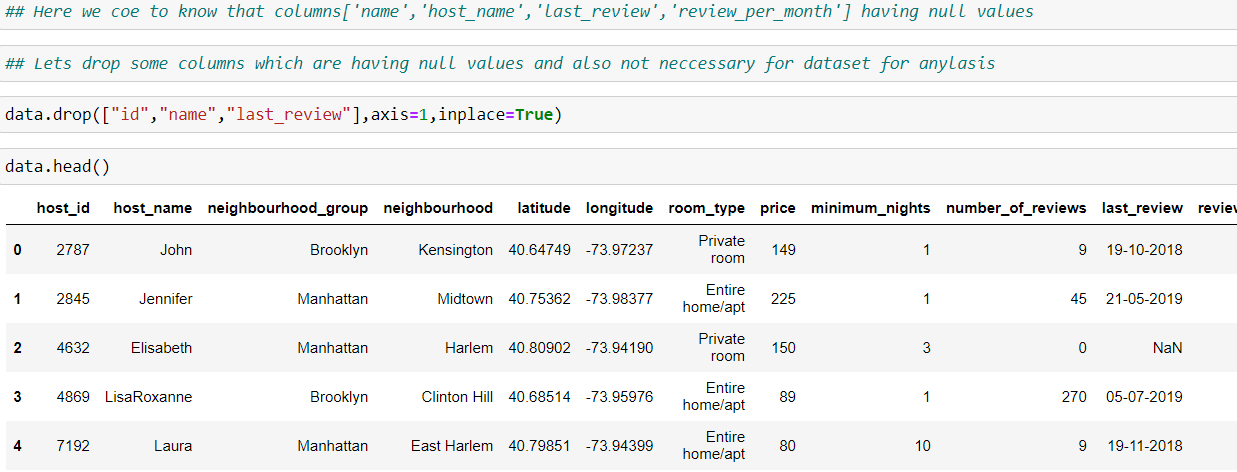
**Number of Rows:** 48895

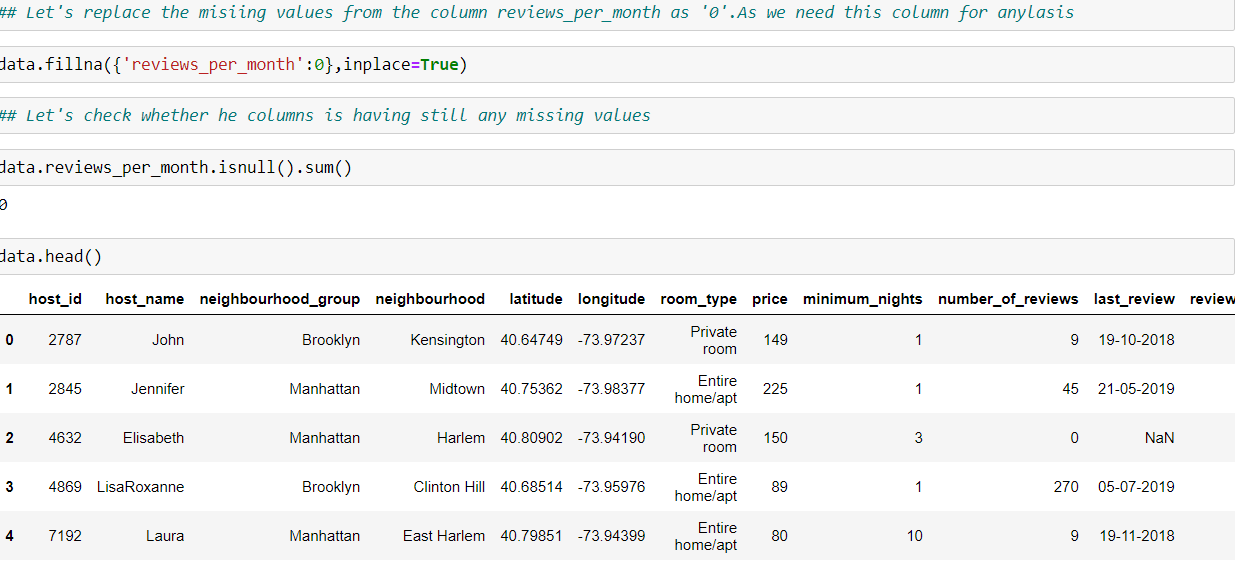
**Number of Columns:** 16













**Step 2: Data Wrangling:**

* **Duplicate Check:** Verified the dataset for duplicate rows, and none were found.
* **Null Value Check:** Identified null values in columns such as *name*, *host-name*, *last review*, and *review-per-month*.
* **Handling Missing Values:** Dropped the *name* column as its missing values were minimal and would not significantly affect the analysis.
* **Data Formatting:** Ensured consistent formatting across the dataset.
* **Outlier Identification:** Reviewed and addressed any outliers in the data to ensure accurate analysis.

**Data Analysis and Visualizations using Tableau:**

We used **Tableau and Excel** to perform data analysis and create visualizations for this assignment. Below are the detailed steps taken for each visualization.

PPT 1

1. Overview of NYC Airbnb Listing – created table by adding the Room type and Neighborhood in Rows and Columns while considered total count for value in table A in tableau and Average of the Prices in Table B for Value in Excel Worksheet
2. Room Type listing Added Room type in Rows and Neighbouhood in Columns with the Count of the listing as value and selected packed bubble for Visualization
3. Profit and Revenue insights – we used side by side bars with text table to compare neighborhoods as per room type to find the highest revenue generator
4. Pricing Analysis - 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.
5. Average Price of Neighborhood - We created a bubble chart with Neighborhood Groups in Columns and Price column in Rows.
6. Popularity of Localities and Properties – We created bar chart and sorted it to highlight the top 10 best performing neighbourhood and filter top 10 in given way



1. Minimum Night Insights – Created Groups for the Night by Right click to click option > Create > group then added the required numbers in a group and applied the same.
2. Neighborhood vs Availability - We created a dual-axis chart using bar chart for Availability 365 and a line chart for price for the top 10 neighborhood groups sorted by price.

Methodology Document PPT 2:

1. Host Acquisition Strategy - We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map



1. Property Insights by Neighbourhood – Created a Bubble chart and text chart using neighbourhood and the count
2. Room Type & Pricing Strategy – In this, we used stacked bars to show the Average price for each room type which is given on the bar as per each neighbourhood also showed the Room type Majority using the highlighting tables to highlight the highest neighbourhood in room types
3. Popularity of Properties - We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour. We used filter to show Top 20 neighbours as per the sum of reviews.
4. Optimizing Less Popular Properties – Used a side-by-side bars to analyse the Neighbourhoods and Room type as per the sum of reviews received per month

Tools used:

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