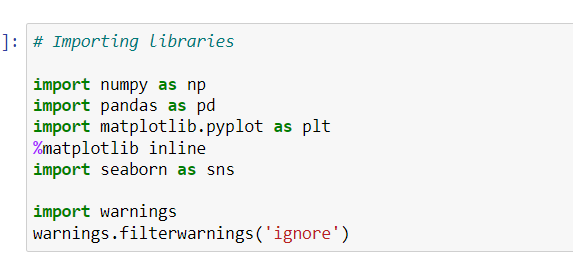
**Airbnb Case Study Methodology**

**Methodology Document:**

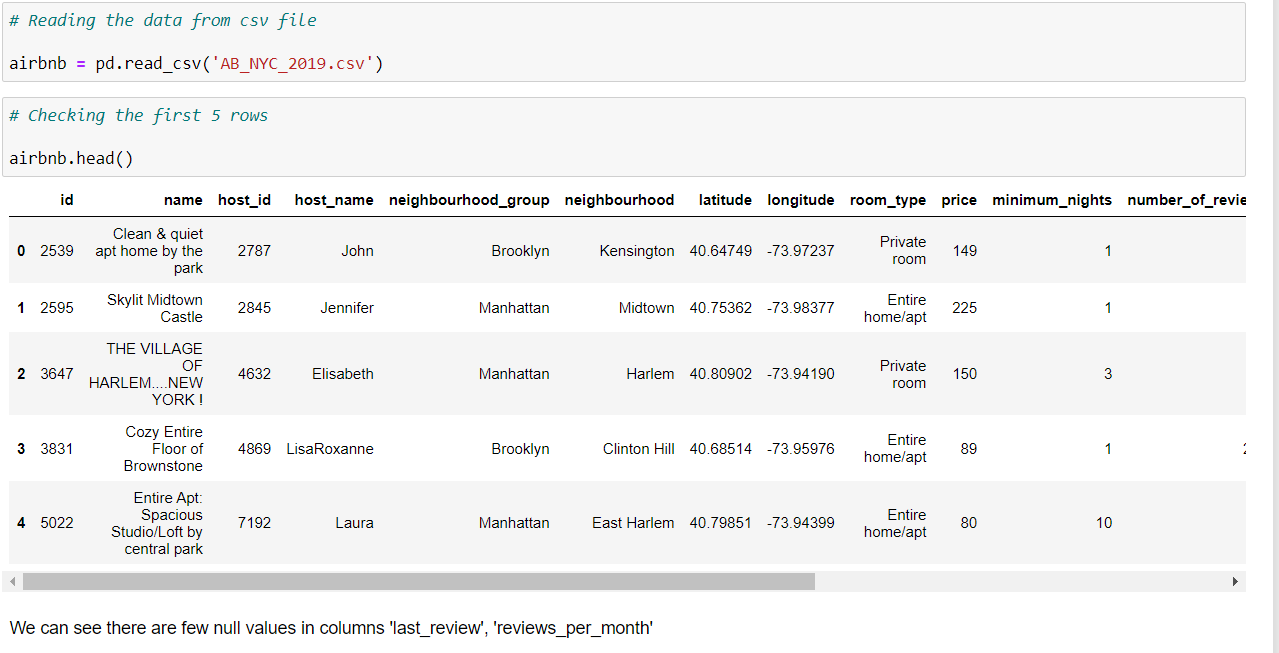
We have used Jupyter notebook to perform initial analysis of the dataset and Tableau for visualization.

**Dataset used**: AB\_NYC\_2019.csv

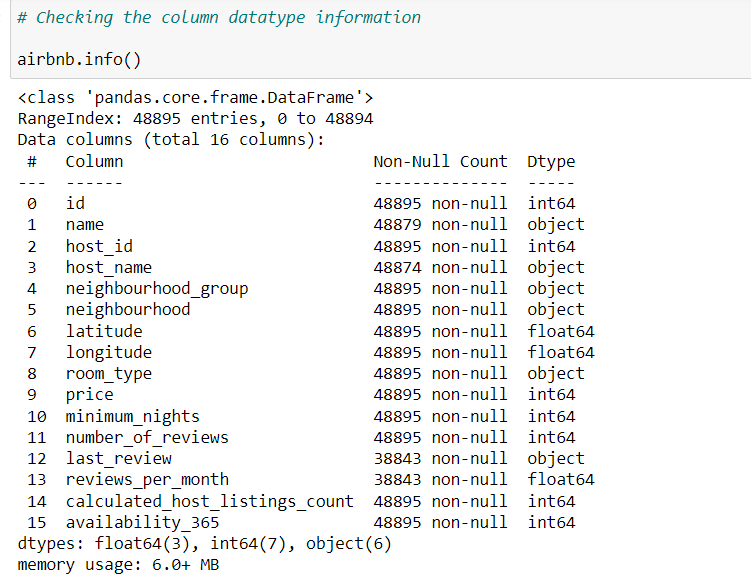
1. Importing libraries



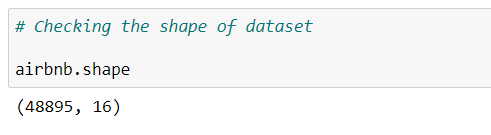
1. Loading dataset and reading the first 5 rows



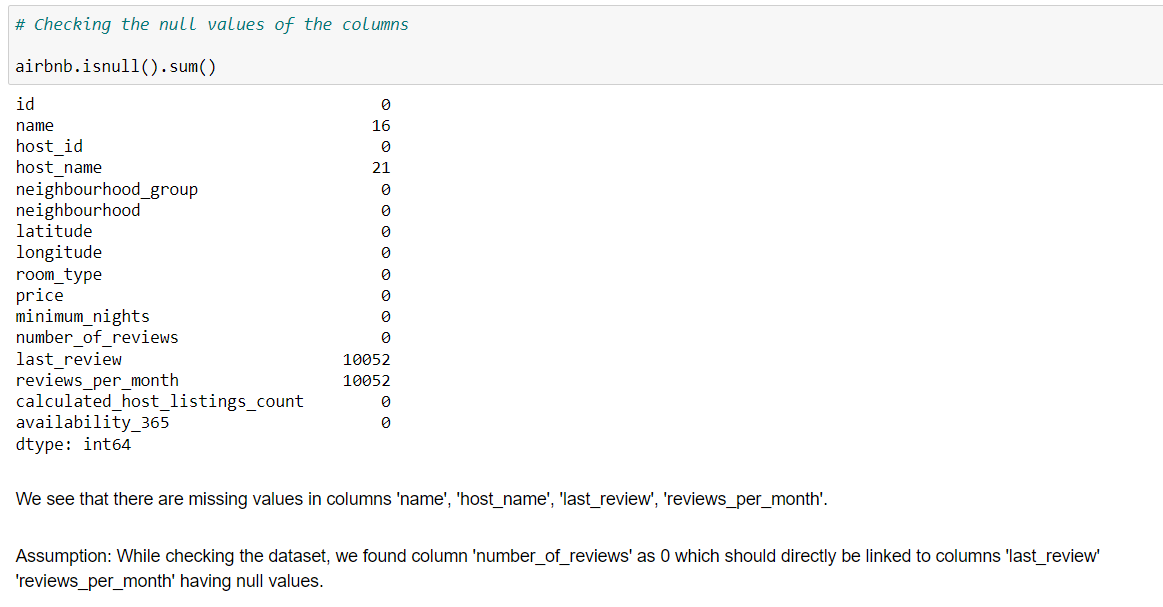
1. Checking the columns and the data types



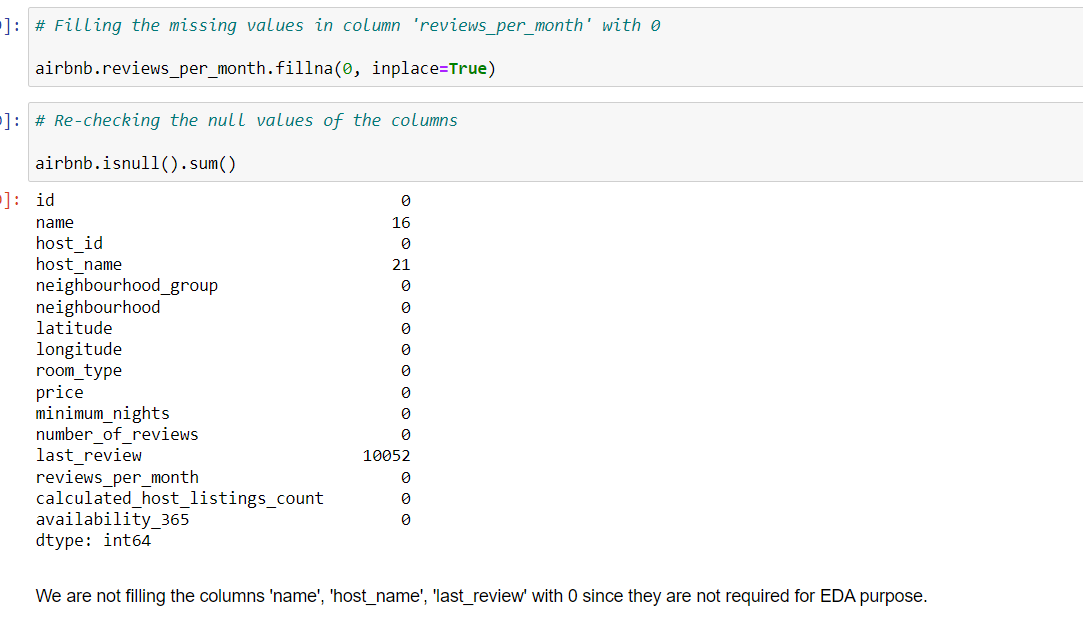
1. Finding dataset rows and columns



1. Checking null values in dataset



1. Filling null values and checking columns



**Data Wrangling:**

**Imputation and Handling Missing Values**:

* Imputed missing values for reviews\_per\_month using mean/median imputation.
* We filled the column reviews\_per\_month and left the other columns as it was present.
* Checked the data types of the columns
* Performed few visualizations to find relationships

**Feature Engineering**:

* Created new features such as review\_rate\_per\_year from reviews\_per\_month to capture annual review rate

**Data Transformation:**

* Extracted date features (year, month, day) from last\_review to analyze seasonal trends

**Data Visualization using Tableau:**

We have used Tableau to visualize the data for the case study. Below are the detailed steps used for each visualization.

***Univariate Analysis:***

1. **Distribution of Availability** – We have used histogram to find the distribution of room availability
2. **Distribution of Minimum nights** – We have used histogram to find the distribution of minimum nights offered by hosts
3. **Distribution of Reviews** – We have used histogram to find the distribution of number of reviews
4. **Distribution of Neighbourhood groups** – We have used histogram to find the distribution of neighbourhood groups
5. **Distribution of Room Types** – We have used histogram to find the distribution of different types of rooms
6. **Distribution of Price** – We have used histogram to find the distribution of room price

***Bivariate Analysis:***

1. **Hosts vs Reviews** – We have created column chart with host name in rows and average number of reviews in columns, sorted by top 10 highest to lowest
2. **Reviews/month vs Neighbourhood groups** – We have created a map showing the top 10 hosts having highest reviews per month across neighbourhood groups
3. **Avg. price vs Neighbourhood groups** - We have created a bubble chart to find the comparison of neighbourhood groups with respect to average price
4. **Avg. price vs Neighbourhoods** - We have created a bubble chart to find the comparison of neighbourhood with respect to average price, sorted by top 10 highest to lowest
5. **Neighbourhood groups vs Minimum nights** – We have created column chart with neighbourhood groups in rows and average minimum nights in columns, sorted from highest to lowest
6. **Neighbourhood group vs Reviews** - We have created a Gantt bar chart to find the comparison of neighbourhood groups with respect to average number of reviews, sorted from highest to lowest
7. **Neighbourhood vs Reviews** - We have created a bar chart to find the comparison of neighbourhoods with respect to average number of reviews, sorted from highest to lowest to find the top 10

***Multi-variate Analysis:***

1. **Room types vs Neighbourhood groups** - We have created stacked column chart with neighbourhood groups in rows and count of room types in columns, room type designated with colours
2. **Neighbourhood vs Availability** – We have created a pareto chart (combination of line and bar chart) to compare neighbourhoods with respect to average price and availability
3. **Hosts having highest reviews across Neighbourhood groups** - We have created a map to compare neighbourhood groups with respect to hosts having highest sum of reviews, sorted by top 10 from highest to lowest
4. **Availability of Room types across Neighbourhood groups** - We have created a pareto chart (combination of line and bar chart) to compare neighbourhoods with respect to average price and availability

**Conclusion:**

* Summarized key insights derived from the data analysis and visualizations
* Provided actionable recommendations based on the findings, such as pricing strategies for hosts, identifying high-demand neighbourhoods, and improving listing descriptions to attract more reviews