# Airbnb Case Study Methodology

## Introduction

This document outlines the methodology used for the analysis of the Airbnb dataset. The analysis was performed using Jupyter Notebook, focusing on data wrangling, data analysis, and visualizations to extract insights from the dataset.

## Dataset

- Dataset Used: AB\_NYC\_2019.csv  
- Number of Rows: 48,895  
- Number of Columns: 16

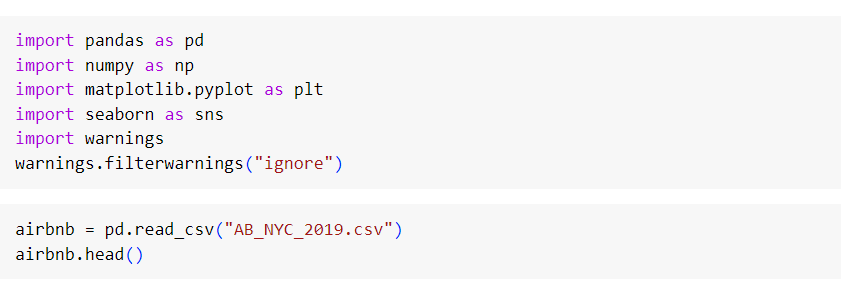
## Tools and Libraries

The following tools and libraries were used in the analysis:  
- Python Libraries: pandas, numpy, matplotlib, seaborn  
- Tool: Jupyter Notebook

## Step 1: Data Wrangling

Code:

#Importing and Reading Data

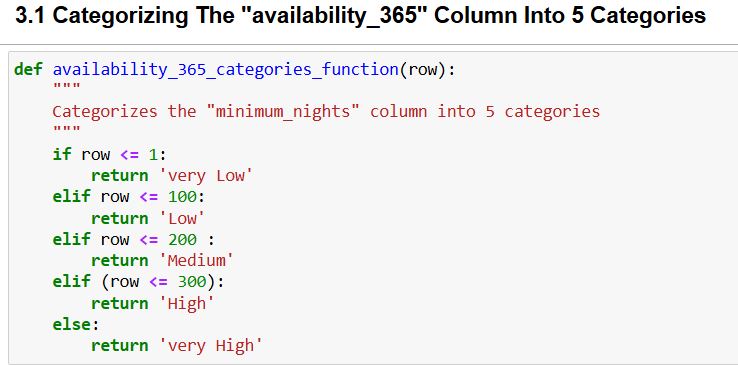


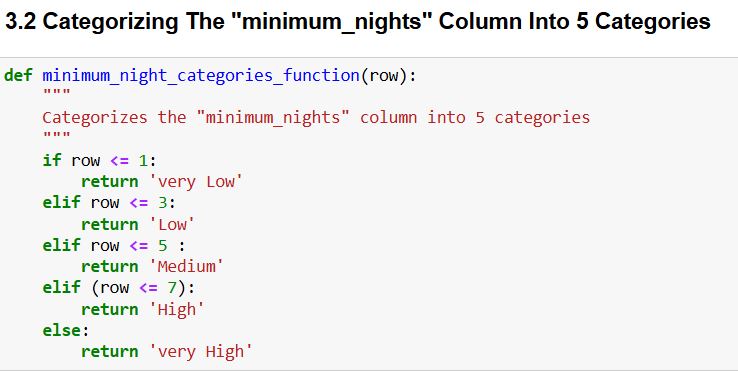
#Data Exploration and Variable Identification

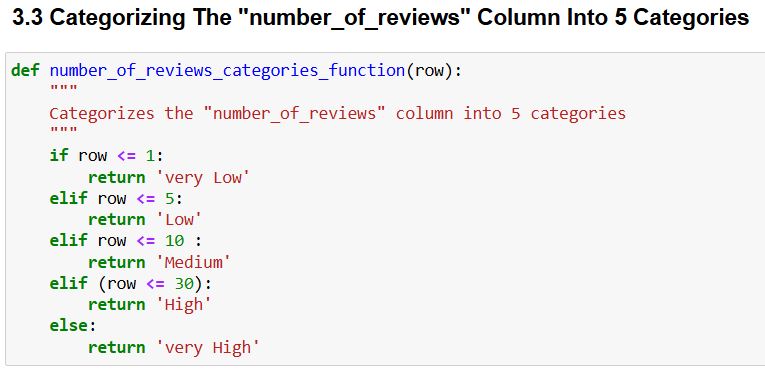


# Check for missing values  
airbnb.isnull().sum()

# Feature creation







#Fixing Columns

**Fix:** The **“reviews\_per\_month”** column is currently of object dtype ; changing it to **“datetime64”** would be more appropriate.

  
  
# Missing Value Analysis  
# Percentage of missing values

round((airbnb.isnull().sum()/len(airbnb))\*100,2)

**Detailed Analysis of Missing Values:**

1) Higher Pricing with Missing 'last\_review':

Observations indicate that listings without a 'last\_review' tend to have higher prices.

2)Lower Review Frequency for Shared Rooms:

Shared rooms are less likely to receive reviews compared to other room types.

3)High Prices Correlate with Fewer Reviews:

Listings with higher prices tend to have fewer reviews, suggesting that guests may be less inclined to leave feedback for more expensive stays.

**Descriptions:**

1. Loading and Understanding Dataset: The dataset is loaded into a pandas DataFrame.
2. Checking for Missing Values: The dataset is checked for missing values in each column.
3. Feature Creation: A new feature created for different numerical columns.
4. Fixing Columns: Fixing the column datatype to a more appropriate one.
5. Missing Value Analysis: The proportion of missing values in each column is calculated and analysis is done

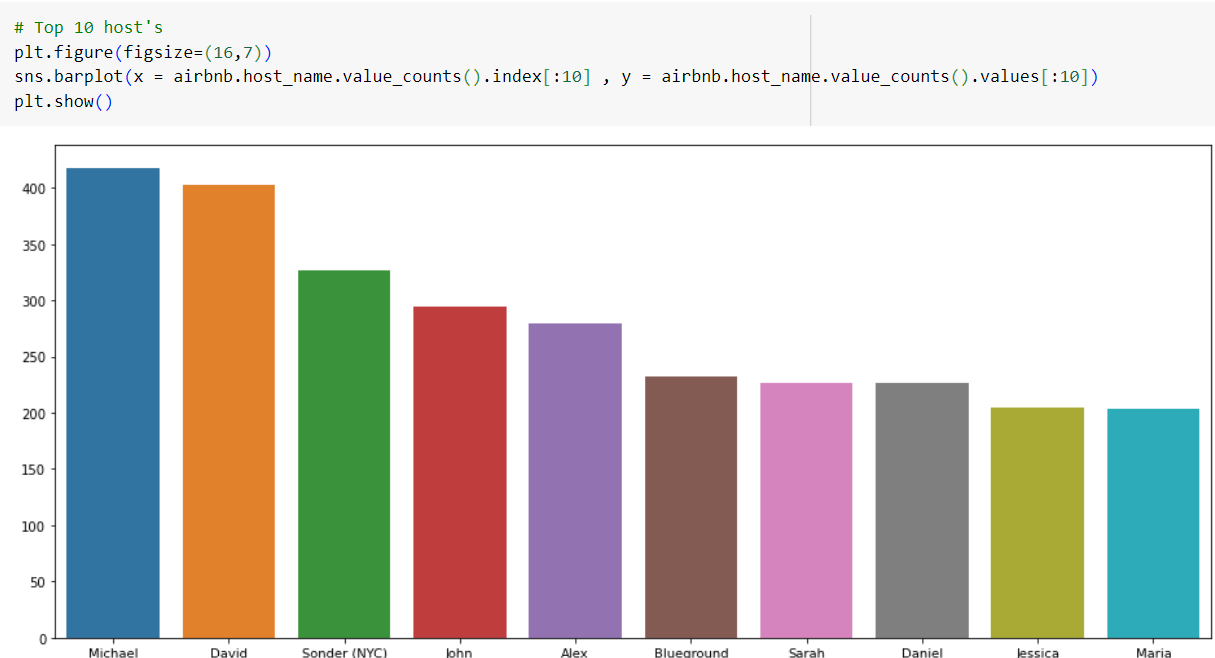
## Step 2: Data Analysis and Visualizations

**Univariate Analysis**

**Top 10 Hosts**

Description:

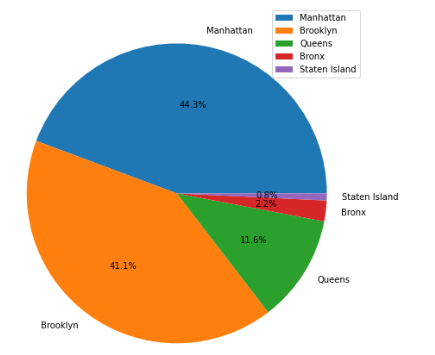
- Top 10 Hosts: The top 10 hosts by the number of listings are identified and plotted.



**Preferred Neighbourhood Group**

Description:

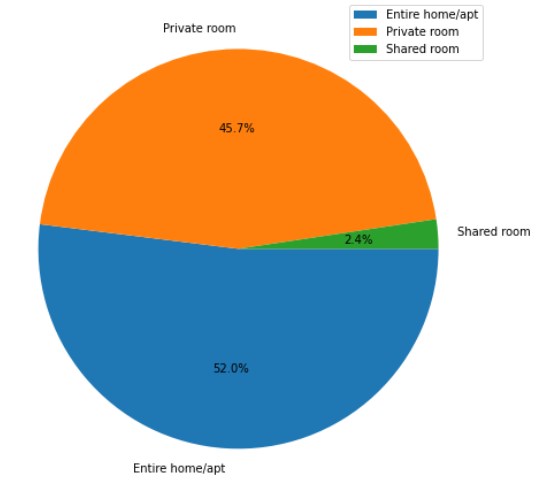
- Neighbourhood Group: The distribution of listings among different neighbourhood groups is visualized using a pie chart.



**Preferred Room Type**

Description:

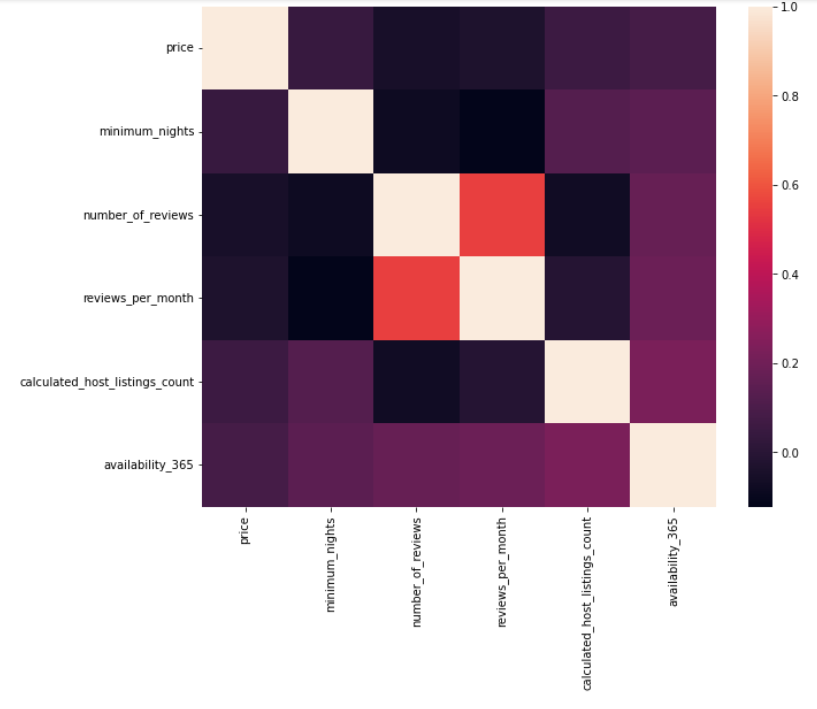
- Room Type: The distribution of room types preferred by hosts is visualized using a pie chart.



**Correlation Among Different Columns**

Description:

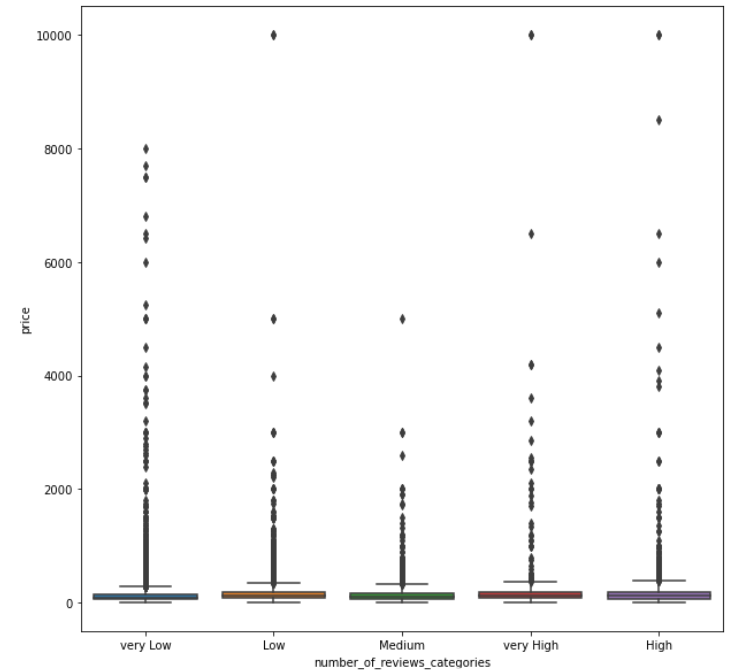
- Correlation Analysis: A heatmap is created to visualize the correlation among different numerical columns in the dataset.



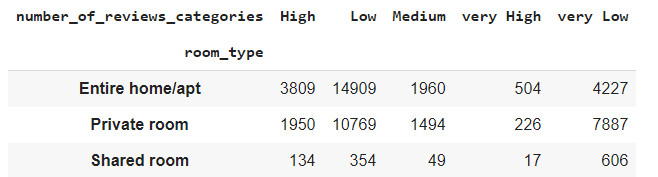
**Review Category vs. Price**

Description:

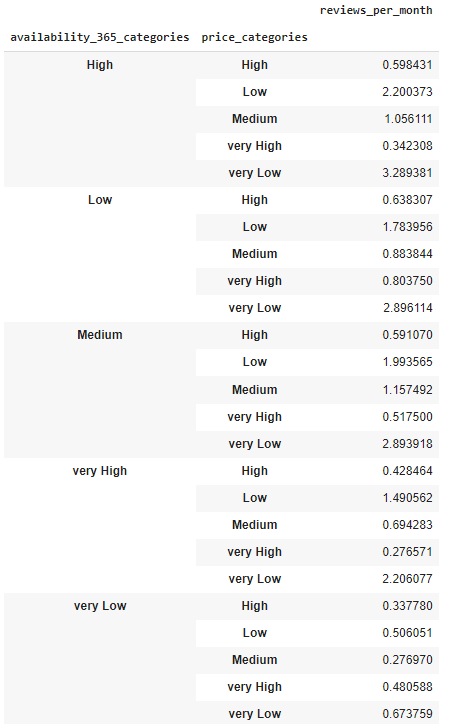
- Reviews vs. Price: A box plot is created to analyze the distribution of reviews across different price categories.



**Room Type and Reviews Per Month**



**Availability 365, Price Category, and Reviews Per Month**



## Conclusion

This methodology document provides a detailed overview of the steps and code used in the analysis of the Airbnb dataset. The analysis includes data wrangling, visualization, and correlation analysis to extract meaningful insights.