Methodology Document: Airbnb Case Study

- **Batch:** DS C57 June
- **Author**: Rajeev Punjabi

1. Introduction

The purpose of this methodology document is to outline the steps and approach used to analyse Airbnb data for the case study. The analysis aims to derive insights into various aspects of the Airbnb business to understand host behaviour, including geographic distribution, pricing trends, property types, market dynamics and key insights related to Airbnb listings in New York City (NYC).

2. Data Source

- The primary data source for this study is the "AB_NYC_2019.csv" dataset, containing information about Airbnb listings in NYC during the year 2019.
- The dataset includes attributes such as neighbourhood, room type, price, availability, and customer reviews.

3. Methodology Steps

3.1. Data Acquisition and Exploration

1. Loaded the dataset into a pandas DataFrame.

1. Importing libraries and reading the data

[1]:
```
# Let us import the necessary python libraries :

import pandas as pd
import numpy as np
import seaborn as sns
from scipy import stats
import os,sys
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
# Loading the Data :
df = pd.read_csv('Data File_AB_NYC_2019.csv')
```

2. Explore the structure of the dataset (number of rows, columns, data types).
3. Checked for missing values and decided on handling strategies (imputation or removal).
4. Identified relevant columns for analysis (e.g., neighbourhood, room type, price, customers' reviews).

3.2. Data Cleaning and Transformation

1. Removed unnecessary columns that do not contribute to the analysis.

2. Converted data types (e.g., date columns, categorical variables).
3. Handled missing values (impute or drop rows).

5. Handling missing values

```
[38]: # To see the number of missing values
      df.isnull().sum()
```

```
t[38]: id                                  0
       name                               16
       host_id                             0
       host_name                          21
       neighbourhood_group                 0
       neighbourhood                       0
       latitude                            0
       longitude                           0
       room_type                           0
       price                               0
       minimum_nights                      0
       number_of_reviews                   0
       last_review                     10052
       reviews_per_month                   0
       calculated_host_listings_count      0
       availability_365                    0
       availability_365_categories         0
       minimum_night_categories            0
       number_of_reviews_categories        0
       price_categories                    0
       dtype: int64
```

```
[39]: # Percentage of missing values
      round((df.isnull().sum()/len(df))*100,2)
```

4. Addressed outliers (if any) in price or other numerical attributes.
5. Created new features (e.g., average review score, price per night).

2. Creating categories from existing features

2.1 Categorizing the "availability_365" column into 5 categories

```
In [11]: df.head()
```

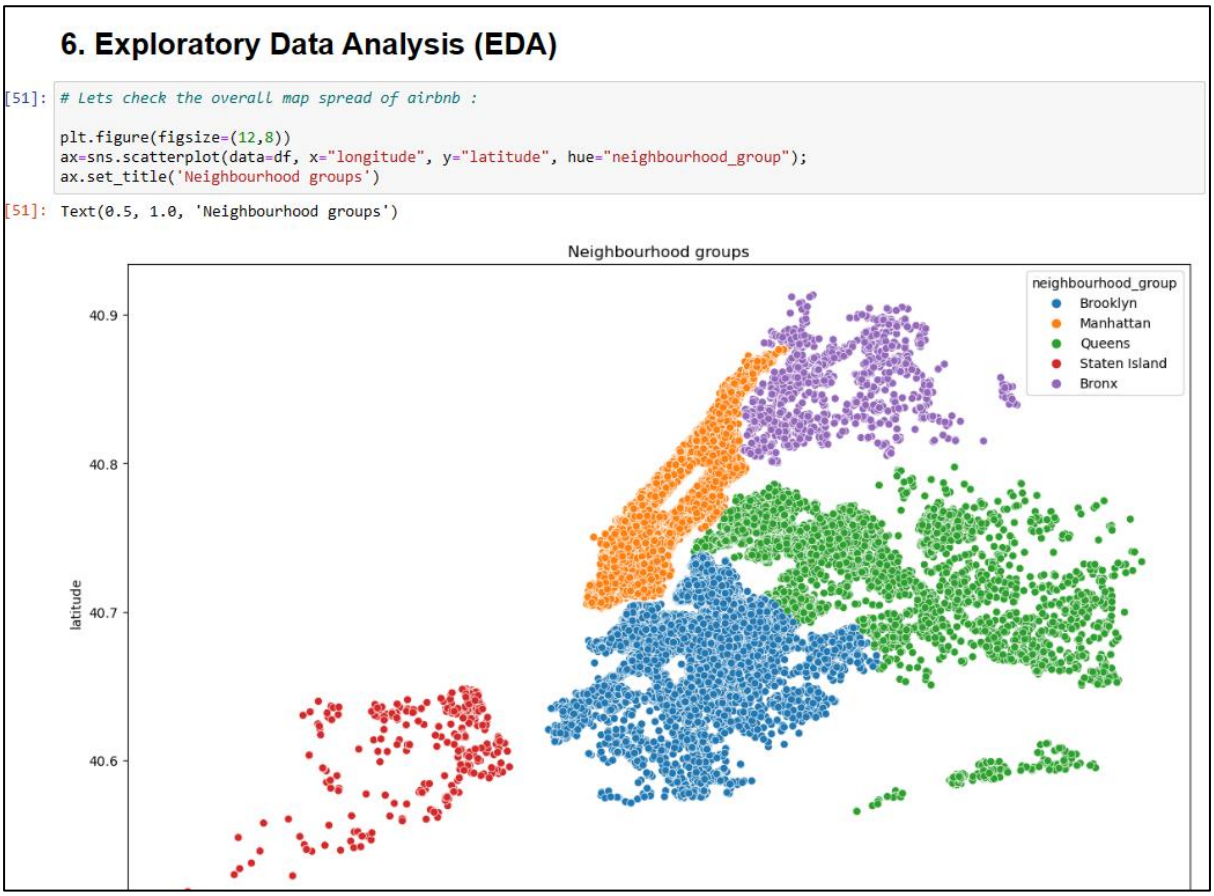
Out[11]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revi
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

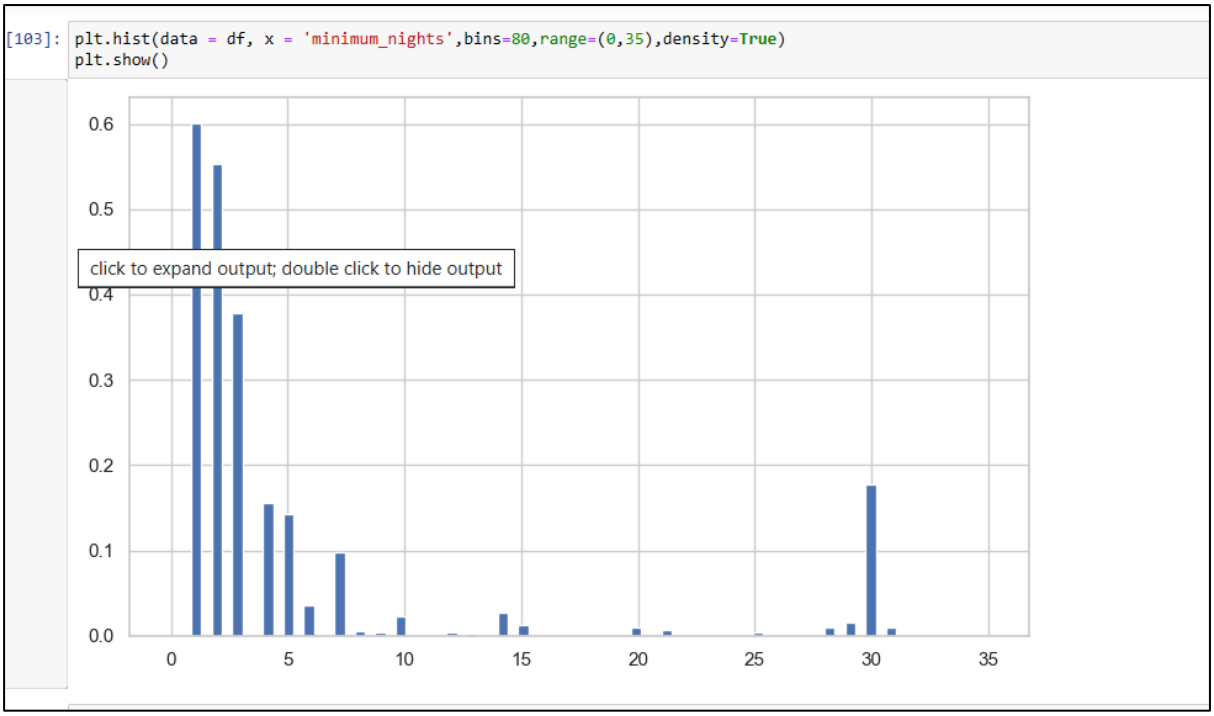
```
In [12]: def availability_365_categories_function(row):
             """
             Categorizes the "minimum_nights" column into 5 categories
             """
             if row <= 1:
                 return 'Very Low'
             elif row <= 100:
```

3.3. Exploratory Data Analysis (EDA)

1. Descriptive Statistics: Compute basic descriptive statistics such as mean, median, and standard deviation for key variables like price, number of reviews, etc.



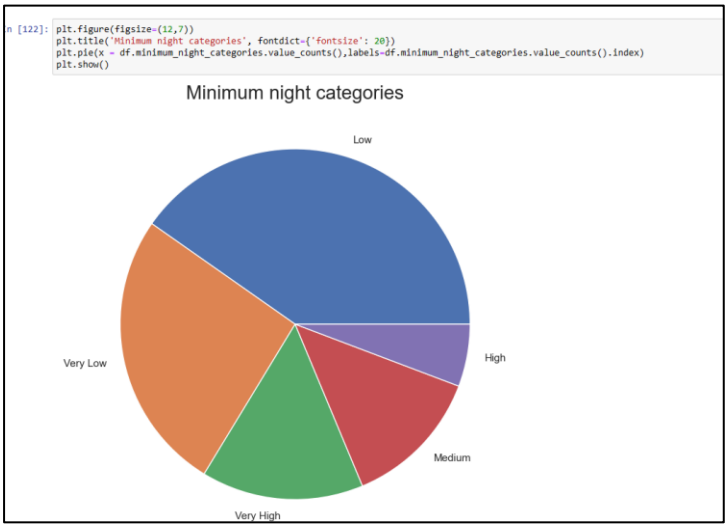
2. Visualizations: Create visualizations (e.g., histograms, scatter plots, bar charts, heatmap) to explore the distribution and relationships between different variables like for example, Room types across different neighbourhood groups. neighbourhood groups (Boroughs). This helps in identifying patterns and trends in the data.



3. Correlation Analysis: Perform correlation analysis to identify relationships between variables. This helps in understanding which factors influence pricing and other metrics.
4. Analysed customer booking patterns based on minimum nights required.
5. Explored listing availability in different neighbourhoods.
6. Investigated preferred price ranges among customers.
7. Compared price variation by room type and neighbourhood.

3.4. Key Findings and Insights

1. Identified top neighbourhoods based on customer bookings.
2. Highlighted neighbourhoods with high review counts (popularity indicators).
3. Investigated correlations between minimum nights required and customer bookings.



4. Suggested opportunities for expansion or optimization based on listing availability.

3.5. Recommendations

1. Provided actionable recommendations for Airbnb hosts:
    - Optimal pricing strategies based on neighbourhood and room type.
    - Insights into customer preferences (e.g., minimum nights, price ranges).
    - Neighbourhoods with growth potential.
2. Suggested areas for further research (e.g., impact of reviews on bookings).

4. Conclusion

In conclusion, this methodology document outlines the steps taken to analyse Airbnb data for the NYC case study. The insights gained from this analysis will inform strategic decisions for hosts and contribute to a deeper understanding of the Airbnb market in NYC as mentioned in PPT1 and PPT2. (refer attachments)
