



STORYTELLING CASE STUDY – AIRBNB NYC

CONTRIBUTORS:

SARAVANAN RAJU

ABHIRAM N

GAYATRI CHAUHAN

AGENDA

- Objective
- Background
- Key Findings
- Recommendations
- Appendix
 - Data Sources
 - Data Methodology
 - Data Model Assumptions

OBJECTIVE

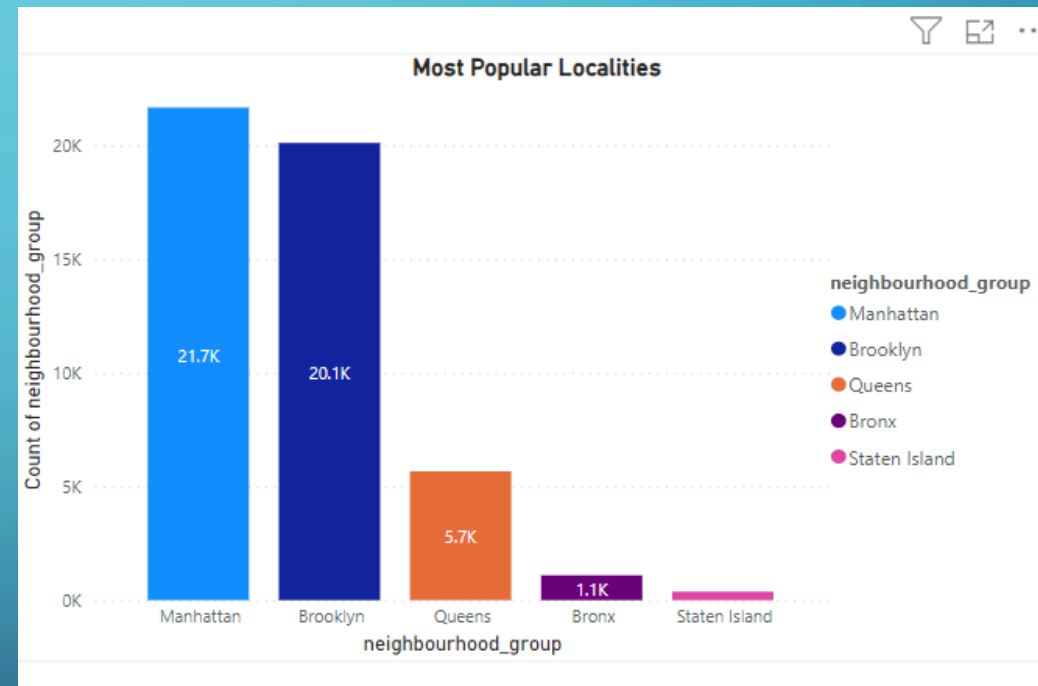
- Provide insights into the current market situation.
- Enhance the understanding of property and host acquisitions, operations and customer preferences.
- Provide upfront recommendations to the marketing and operations team.

BACKGROUND

- Airbnb, an online marketplace for short – and long-term homestays and experiences.
- For the past few months, Airbnb has seen a major decline in revenue.
- Now that the restrictions have been lifted and people have started to travel more.
- Airbnb wants to make sure that it is fully prepared for the change.

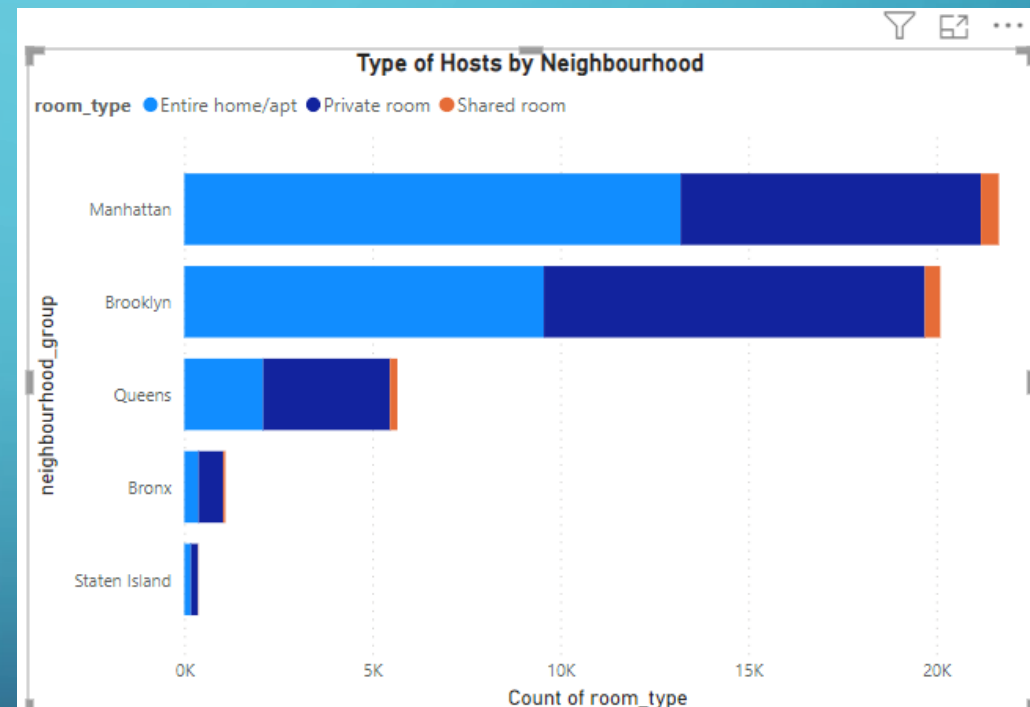
NEIGHBOURHOOD'S TO BE TARGETED

- Manhattan and Brooklyn are top neighbourhood groups and mostly people prefer to book the entire home or private room.
- Manhattan has highest number of home/apt properties of total listed properties.
- Very less number of shared rooms are available in each Neighbourhood group.



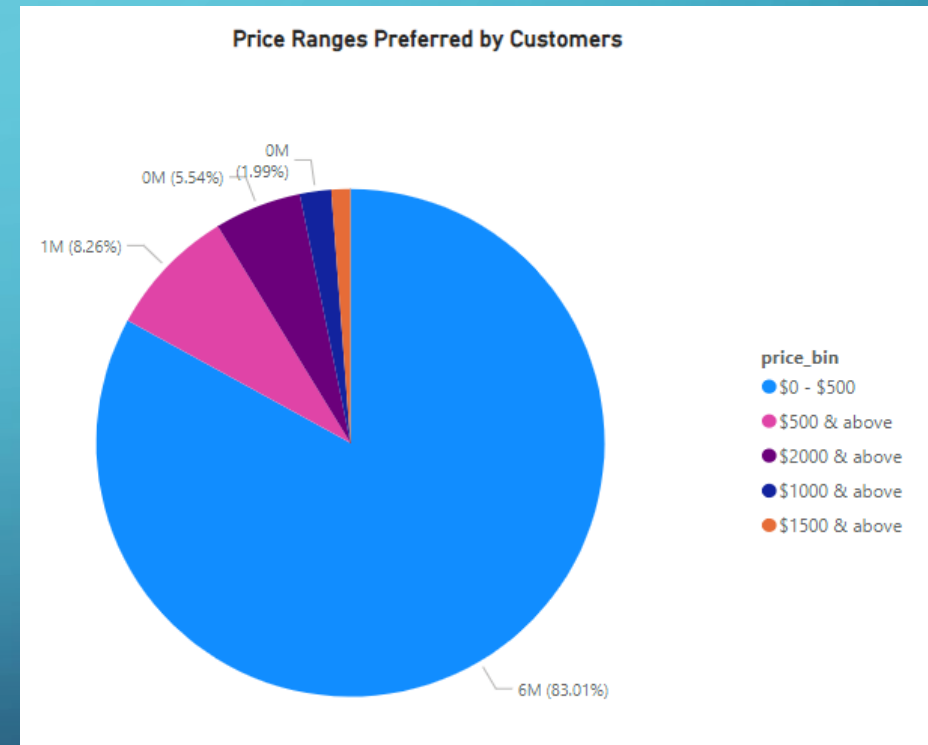
TYPE OF HOSTS AND THEIR LOCATION PREFERENCES

- Entire home/Apartments are the top choice of most people.
- Shared accommodations has the least preferences.
- Manhattan and Brooklyn again tops in the location.



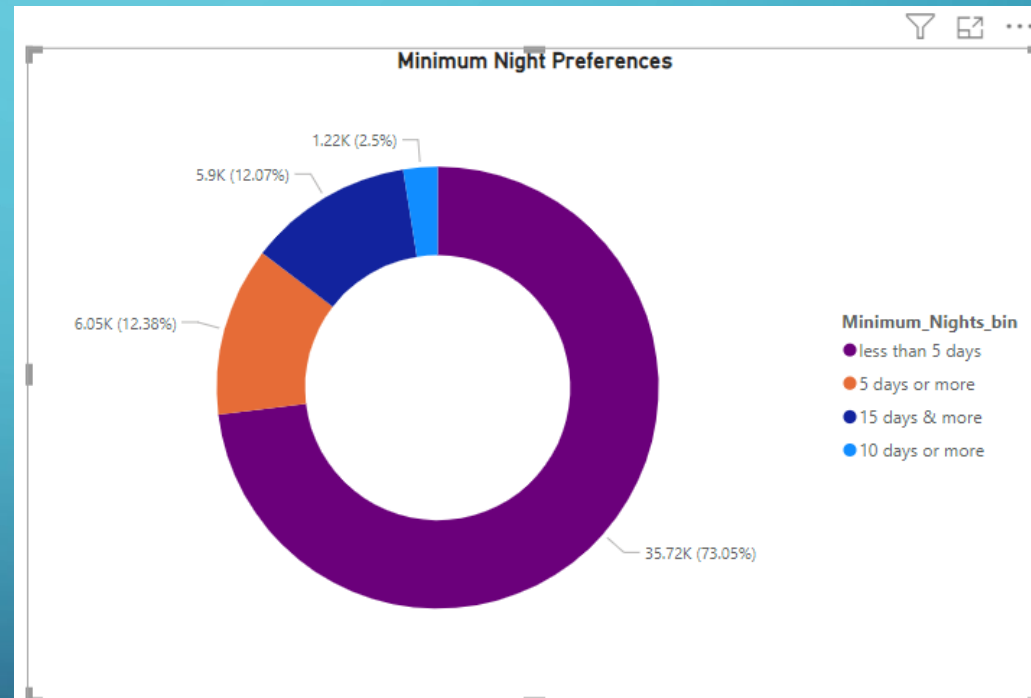
PRICE RANGES PREFERENCES BY CUSTOMERS

- Most preferred price ranges fall under \$500.
- Over 83% of customers fall in this price range.



HOW LONG CUSTOMERS STAY?

- Most customers(73%) stay for less than 5 days.
- Close to 25% customers stay more than 5 to 15 days.



KEY FINDINGS AND RECOMMENDATIONS

- Shared accommodations has the least preferences. These need to be inspected and customized to private rooms to meet customer demand.
- More than 80% of the listings are in Manhattan and Brooklyn neighbourhood.
- Threshold of minimum nights should be less than 10 nights to make property more customer-oriented.

A decorative graphic on the left side of the slide, consisting of white and light blue lines that resemble a circuit board or a stylized tree. The lines are vertical and horizontal, with small circles at the ends, creating a complex, branching pattern.

APPENDIX

DATA SOURCES

DATA METHODOLOGY

DATA ASSUMPTIONS

DATA SOURCES

The columns in the dataset are self-explanatory. You can refer to the diagram given below to get a better idea of what each column signifies.

Note: The price column contains the price/night.

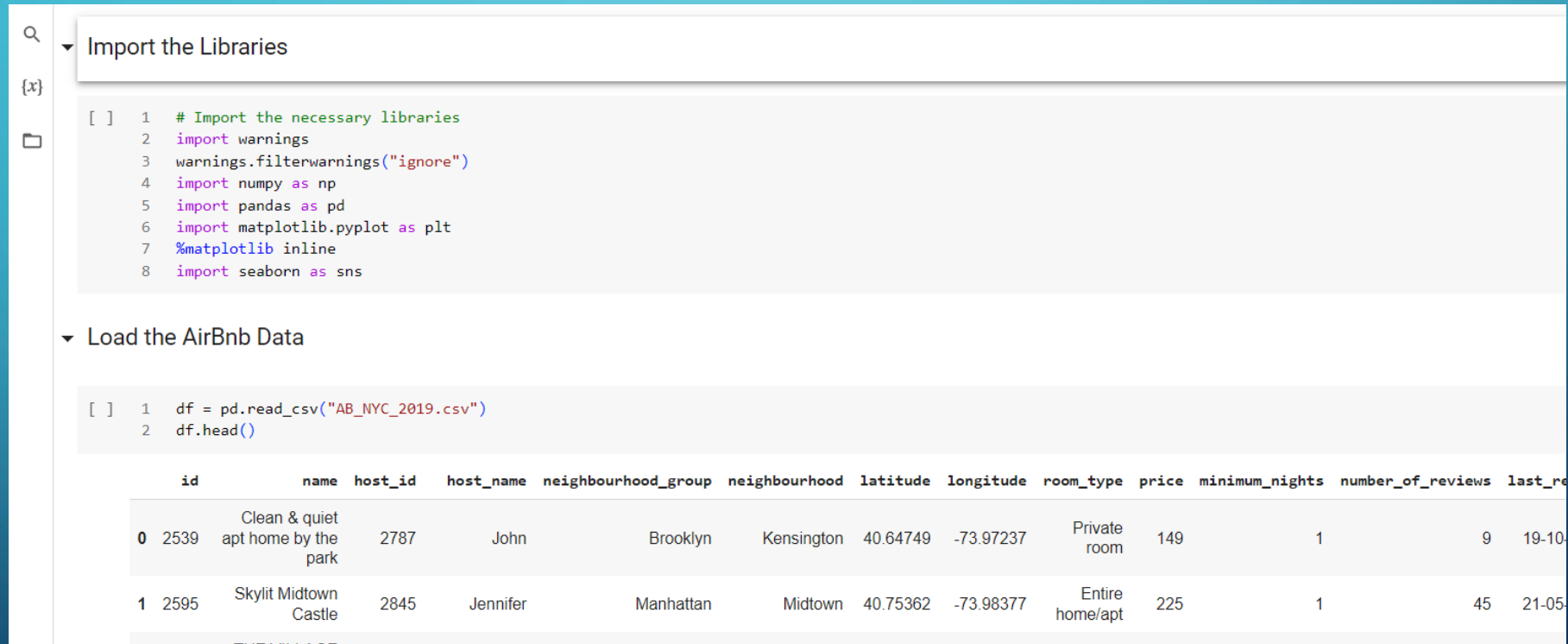
| Column | Description |
|--------------------------------|--|
| id | listing ID |
| name | name of the listing |
| host_id | host ID |
| host_name | name of the host |
| neighbourhood_group | location |
| neighbourhood | area |
| latitude | latitude coordinates |
| longitude | longitude coordinates |
| room_type | listing space type |
| price | |
| minimum_nights | amount of nights minimum |
| number_of_reviews | number of reviews |
| last_review | latest review |
| reviews_per_month | number of reviews per month |
| calculated_host_listings_count | amount of listing per host |
| availability_365 | number of days when listing is available for booking |

Dataset Description

DATA METHODOLOGY

- Conducted Data Analysis on Airbnb, NYC dataset.
- Data-Cleaning, Preparation & adding features were done through Python.
- Created charts and Visualization through Power-BI

DATA METHODOLOGY – DATA CLEANING/PREPARATION



The screenshot shows a Jupyter Notebook interface with two sections: "Import the Libraries" and "Load the Airbnb Data".

Import the Libraries

```
[ ] 1 # Import the necessary libraries
2 import warnings
3 warnings.filterwarnings("ignore")
4 import numpy as np
5 import pandas as pd
6 import matplotlib.pyplot as plt
7 %matplotlib inline
8 import seaborn as sns
```

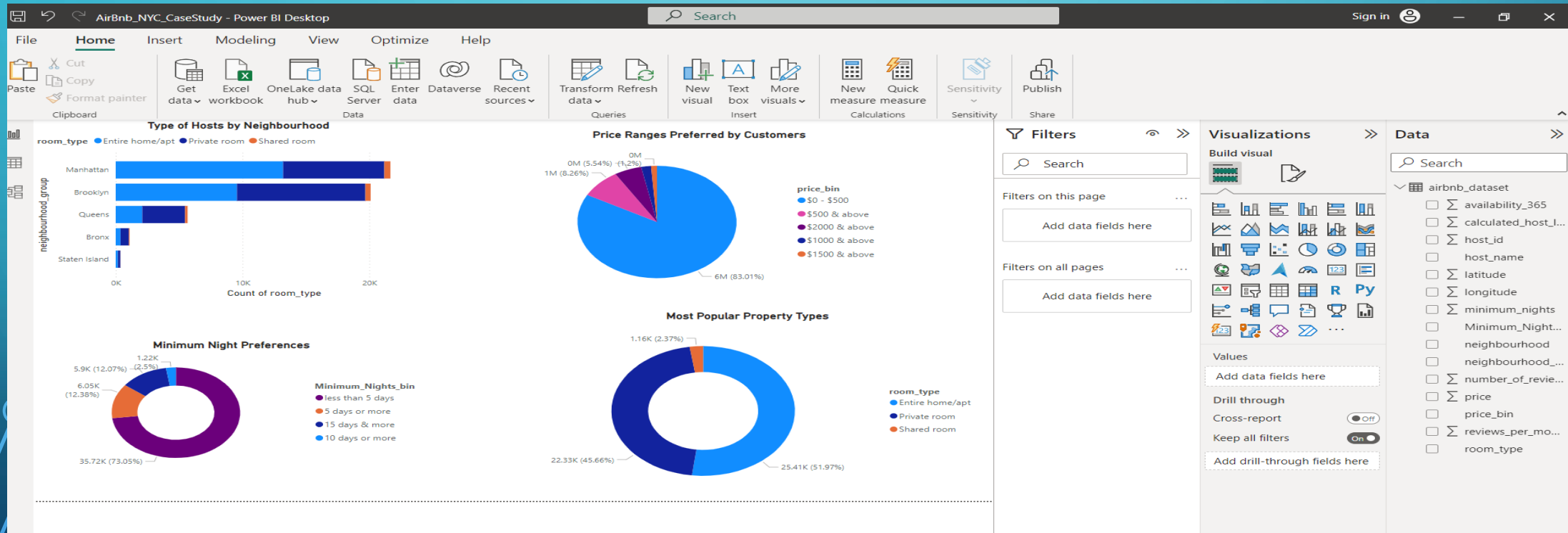
Load the Airbnb Data

```
[ ] 1 df = pd.read_csv("AB_NYC_2019.csv")
2 df.head()
```

The output of the second cell is a DataFrame with the following columns: id, name, host_id, host_name, neighbourhood_group, neighbourhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review. The first two rows of data are shown below:

| | id | name | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_reviews | last_review |
|---|------|------------------------------------|---------|-----------|---------------------|---------------|----------|-----------|-----------------|-------|----------------|-------------------|-------------|
| 0 | 2539 | Clean & quiet apt home by the park | 2787 | John | Brooklyn | Kensington | 40.64749 | -73.97237 | Private room | 149 | 1 | 9 | 19-10- |
| 1 | 2595 | Skylit Midtown Castle | 2845 | Jennifer | Manhattan | Midtown | 40.75362 | -73.98377 | Entire home/apt | 225 | 1 | 45 | 21-05- |

DATA METHODOLOGY – VISUALIZATION THRU POWER BI



DATA ASSUMPTIONS

Categorical Variables:

- room_type
- neighbourhood_group
- neighbourhood

Continous Variables(Numerical):

- Price
- minimum_nights
- number_of_reviews
- reviews_per_month
- calculated_host_listings_count
- availability_365
- Continous Variables could be binned in to groups too

Location Variables:

- latitude
- longitude

Time Varibale:

- last_review

Variable Categories