AIRBNB Case Study

IIITB- Sep 2022 Batch

Vishwanath Rajasekaran

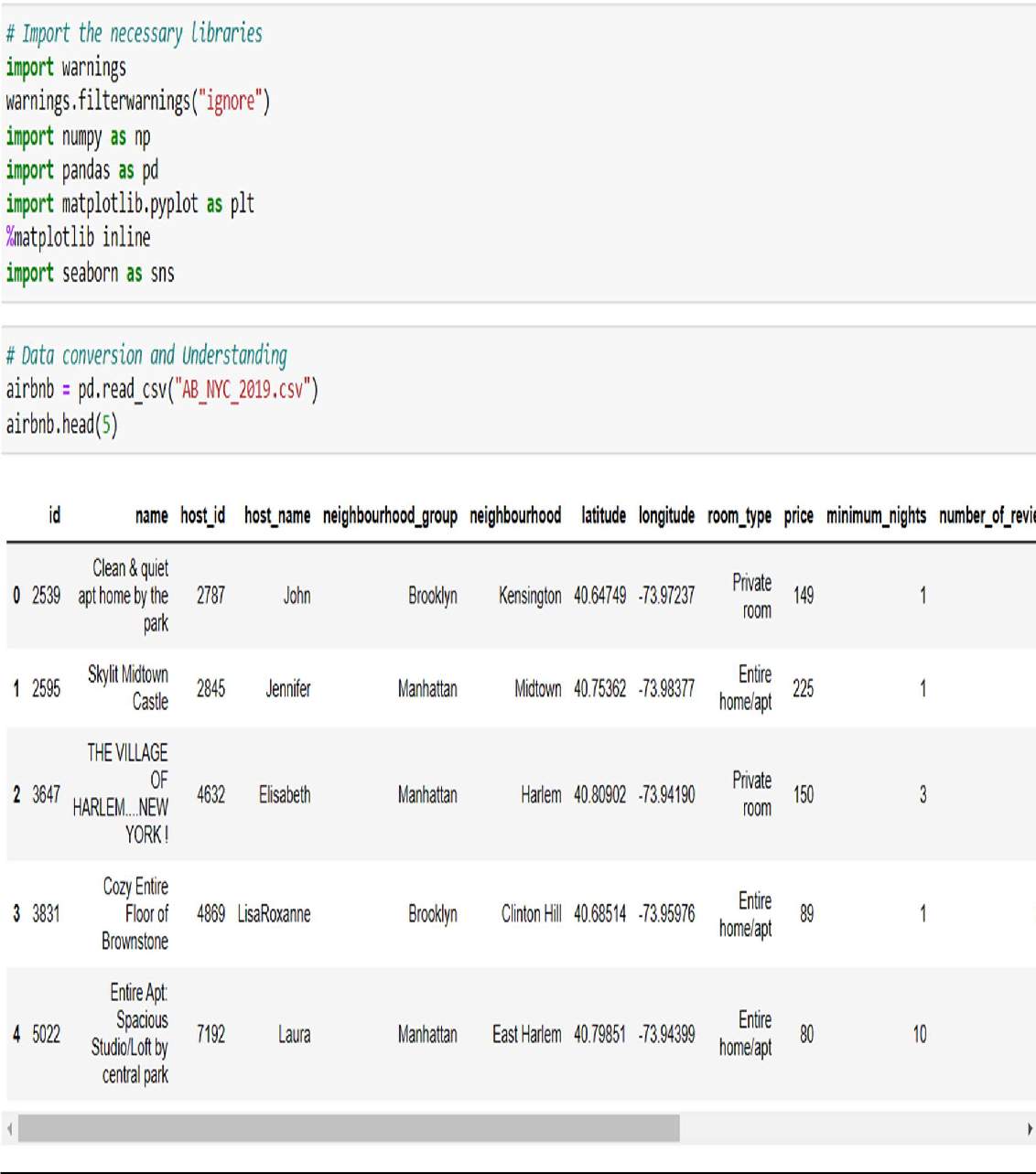
# Methodology Document PPT 1:

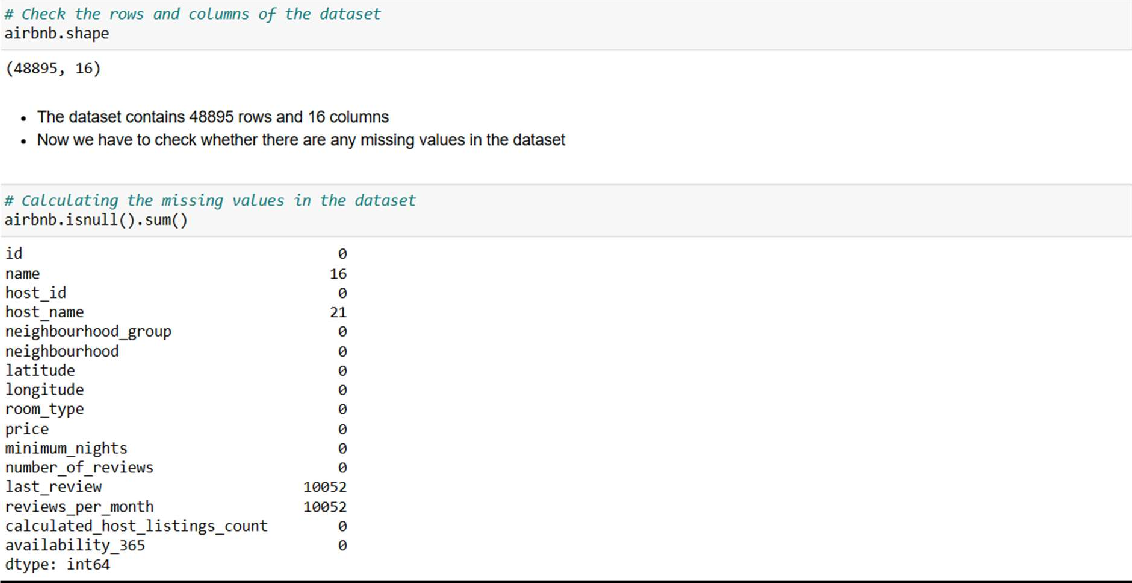
In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

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

**Number of Rows:** 48895

## Number of Columns: 16







* We removed the columns like Id, Name, Last Review which was not giving much information.



## Step 2: Data Wrangling:

* Checked the Duplicate rows in our dataset and no duplicate data was found.
* Checked the Null Values in our dataset. Columns like name, host-name, last review and review-per-month have null values.
* We’ve dropped the column name as missing values are less and dropping it won’t have significant impact on analysis.
* Checked the formatting in our dataset.
* Identified and review outliers.

## Data Analysis and Visualizations using Tableau:

We have used tableau to visualize the data for the assignment. Below are the detailed steps used for each visualization.

## Top 10 Host:

* + We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map.



## Preferred Room type with respect to Neighbourhoodgroup:

* + We created a pie chart for understanding the percentage of room type preferred w r t neighbourhoodgroup.
  + We added Room Type to the colors Marks card to highlight the different Room Type in different colors and count of Host Id to the size.

## For Variance of price with NeighbourhoodGroups:

* We used a box and whisker’s plot with NeighbourhoodGroups in Columns and Price in Rows.
* We changed the Price from a Sum Measure to the median measure.

## Average price of Neighborhood groups:

* We created a bubble chart with Neighbourhood Groups in Columns and Price column in Rows.
* We added the Neighbourhood Groups to the colors Marks card to highlight the different neighbourhoodGroups in different colors. Also Put Avg price in Label.

## Customer Booking w r t minimum nights:

* We created the bin for Minimum nights as shown below.



* The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.

## Popular Neighborhoods:

* 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.

## Neighbourhood vs Availability:

* We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

# Methodology Document PPT 2:

## Room type with respect to Neighbourhood group:

* + We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
  + We added Room Type to the colours Marks card to highlight the different Room Type in different colours and count of Host Id to the size

## Customer Booking with respect to minimum nights:

* We created the bin for Minimum nights as shown below.



* The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.

## Neighbourhood vs Availability:

* We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

## Price range preferred by Customers:

* We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of $20.

## Understanding Price variation w.r.t Room Type & Neighbourhood:

* We created Highlights Table chat by taking Room Type in rows & Neighbourhood Group in column.
* We took the average price in colour Marks card to highlight the different Room Type in different colours.

## Price variation w r t Geography:

* We used Geo location chart to plot neighbourhood, neighbourhood Group in map to show case the variation of prices across.

## Popular Neighborhoods:

* 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.

## Tools used:

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