AIRBNB Case Study IIIT B -Gaurav Chaudhary,Malaika Goveas,Divyanshi (DSC 59)

# Methodology Document PPT 1:

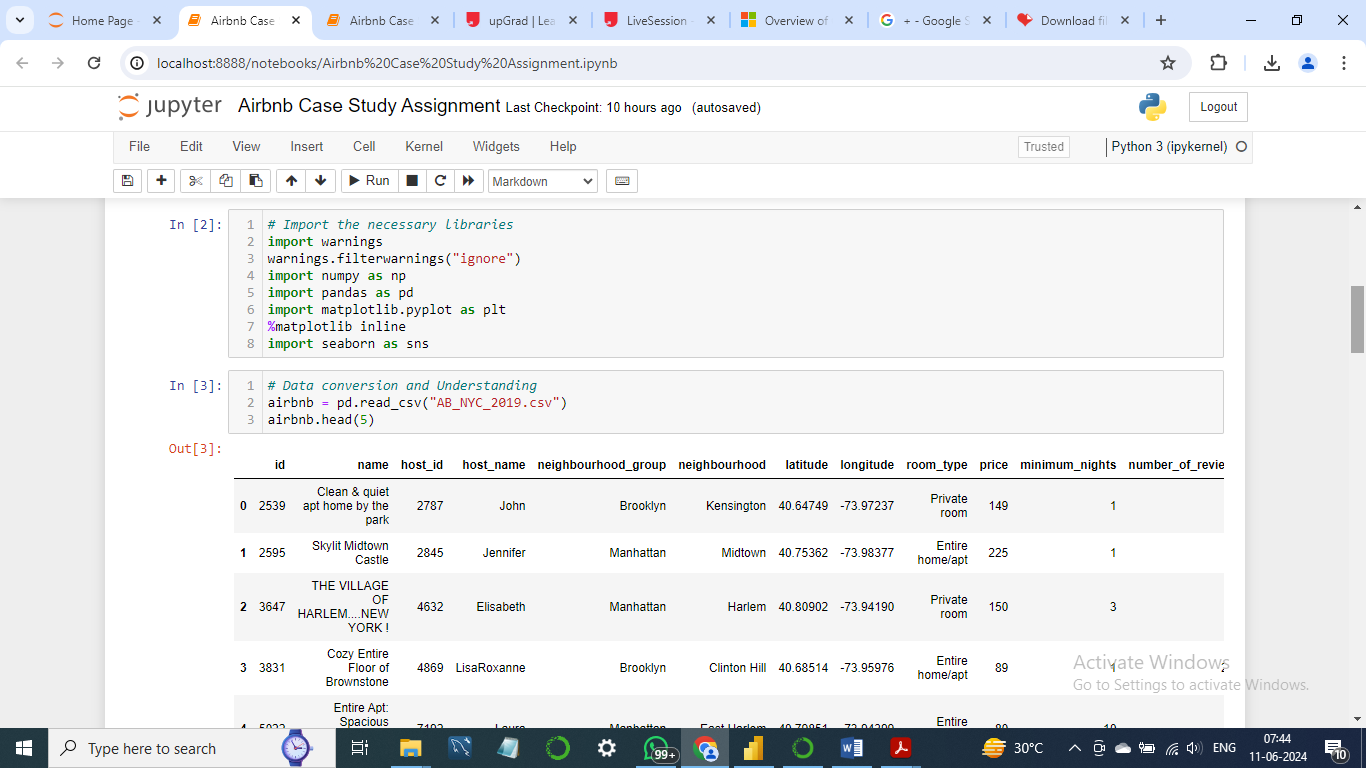
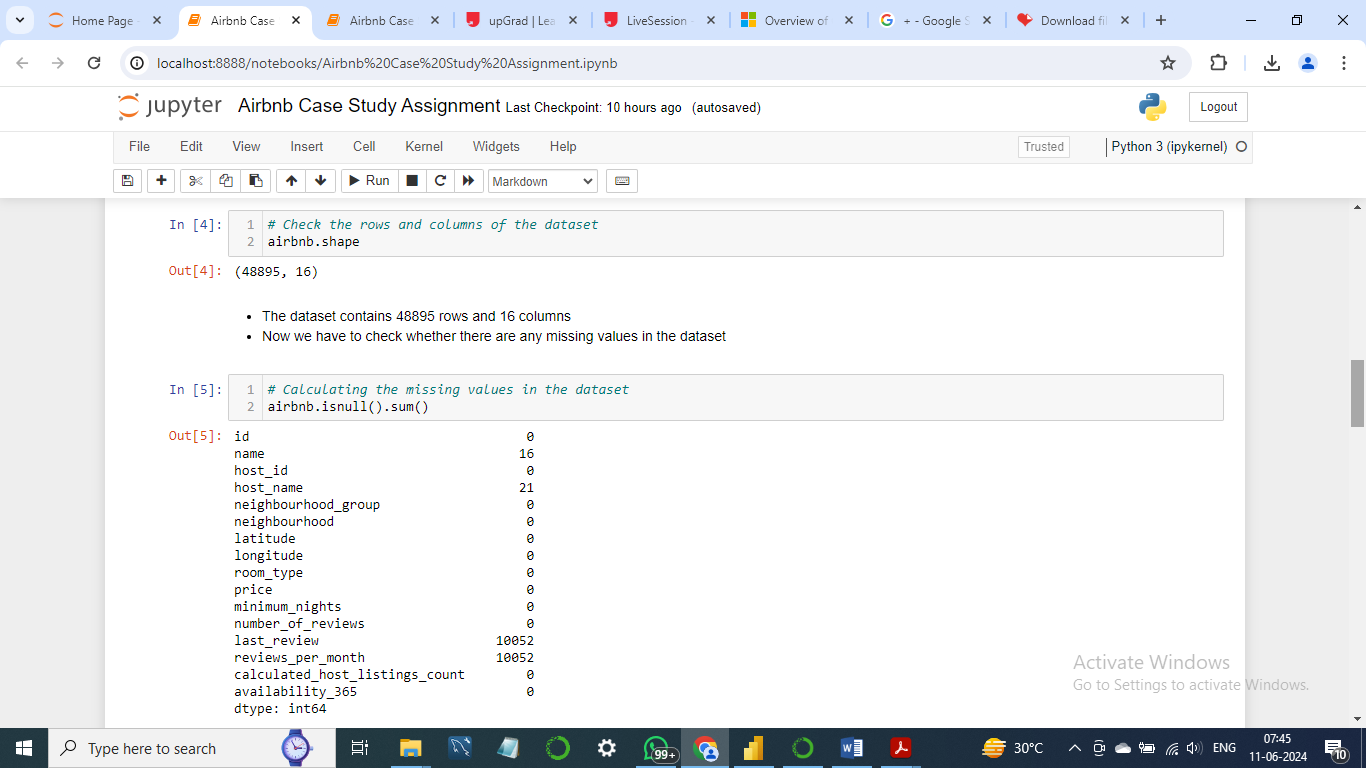
In the case study we have used **Jupiter notebook** to perform initial analysis of the data

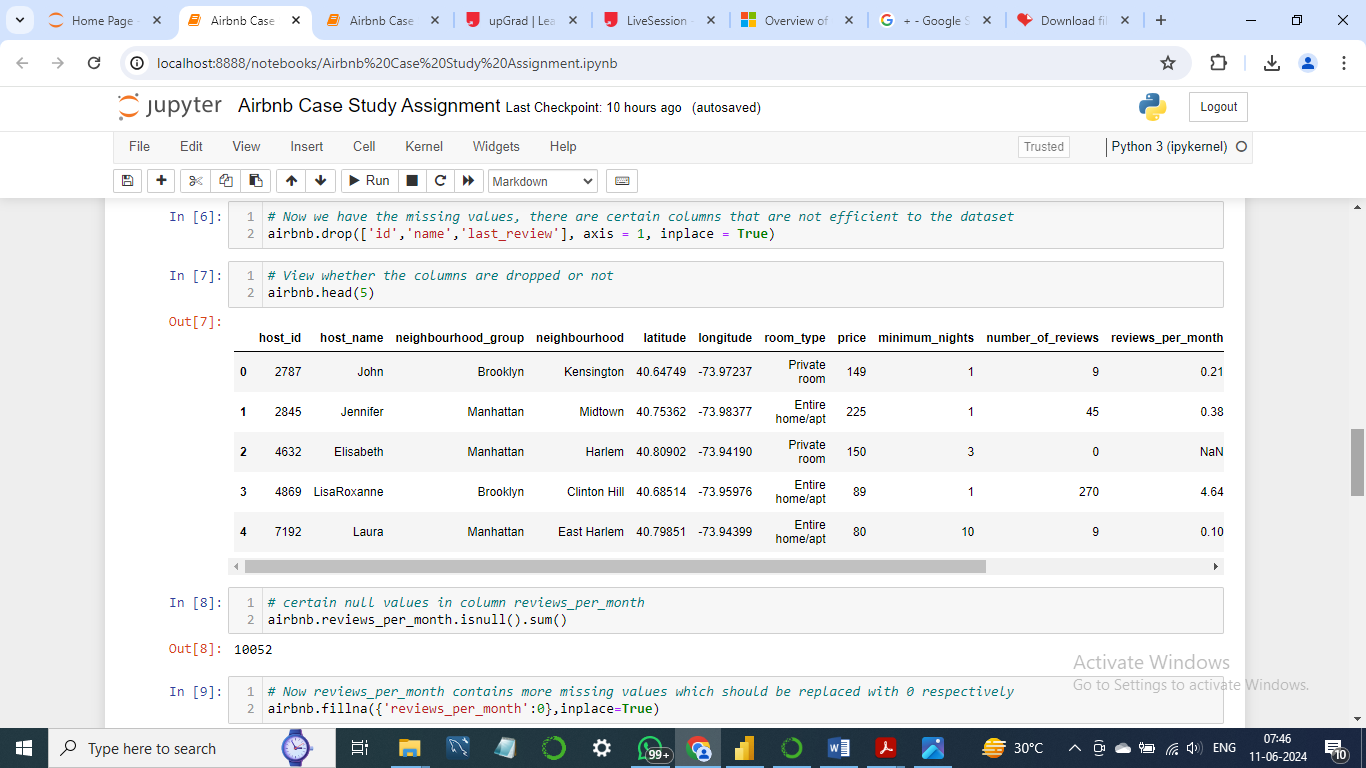
and **Power Bi for data analysis and visualization**.

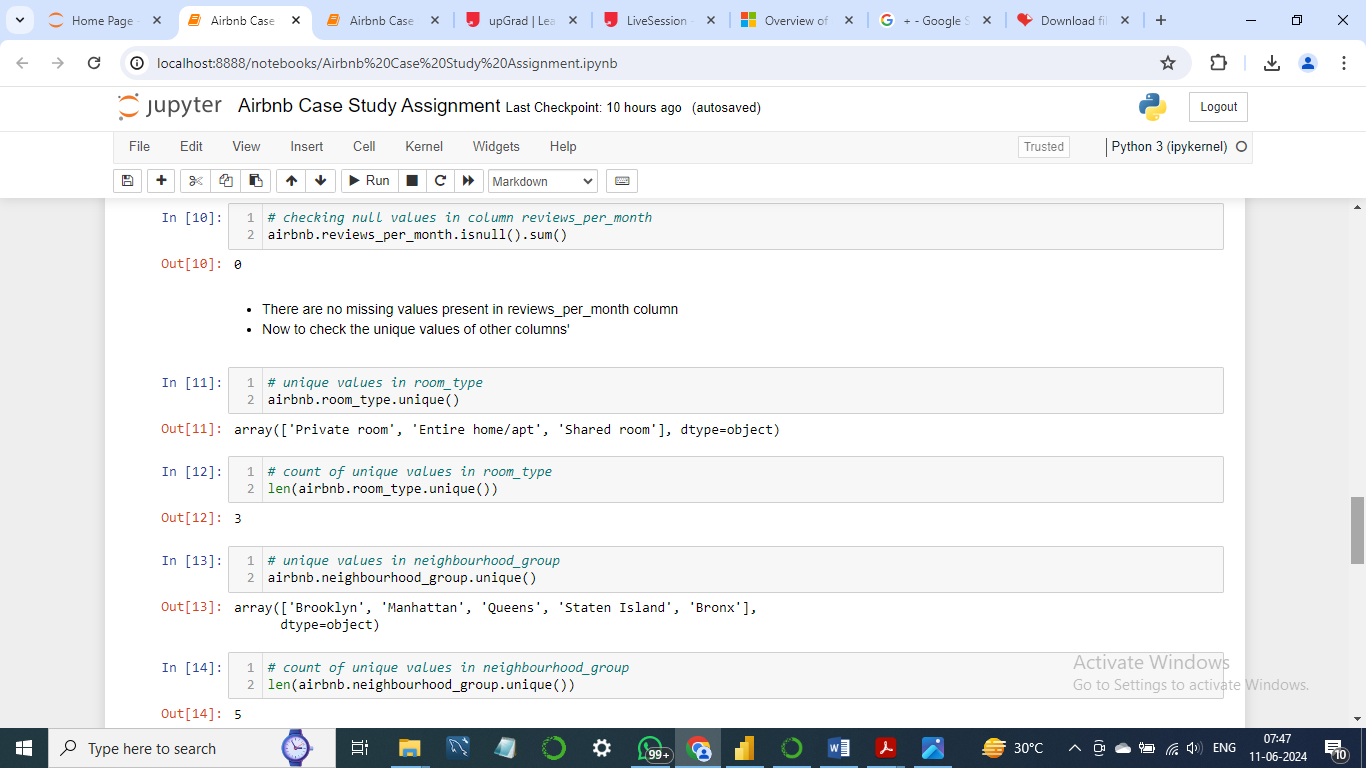
**Initial Analysis using Jupiter notebook**: Data Set Used: AB\_NYC\_2019.csv

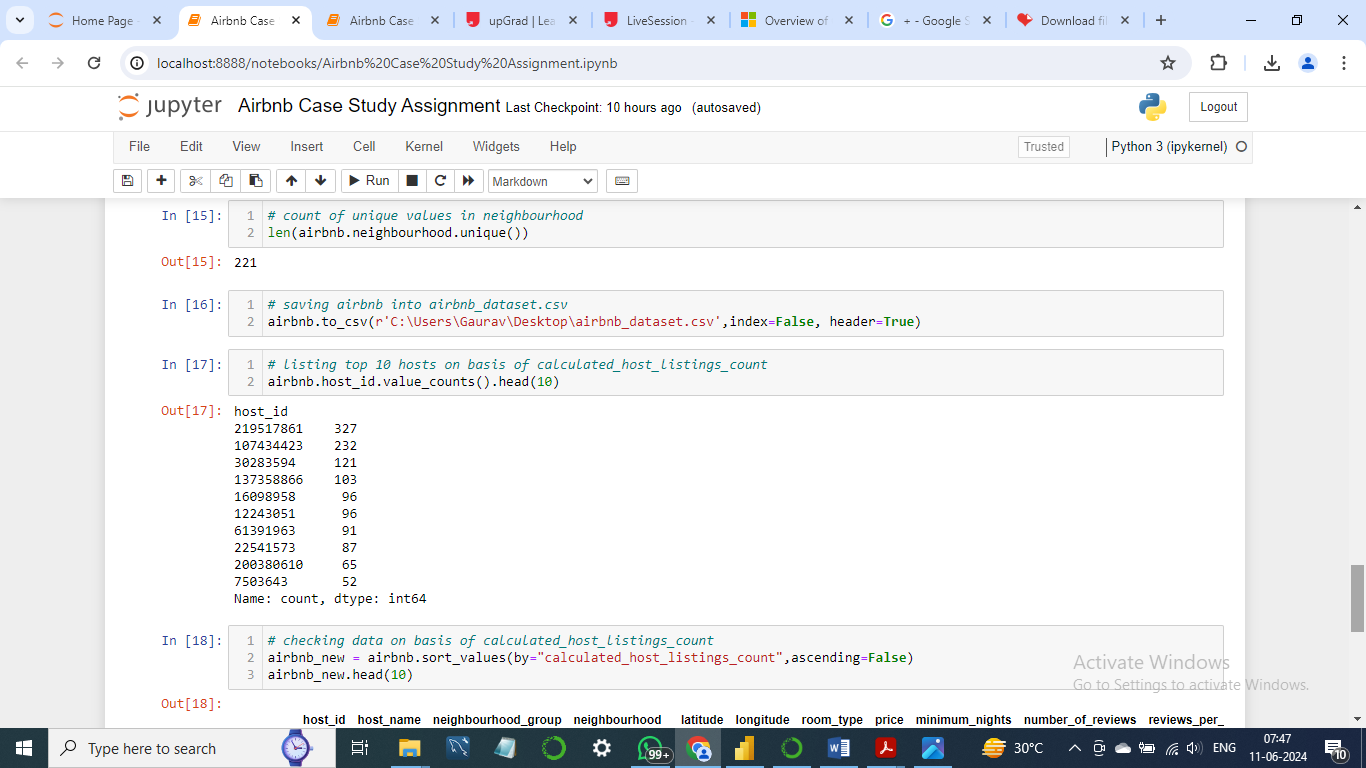
Number of Rows: 48895

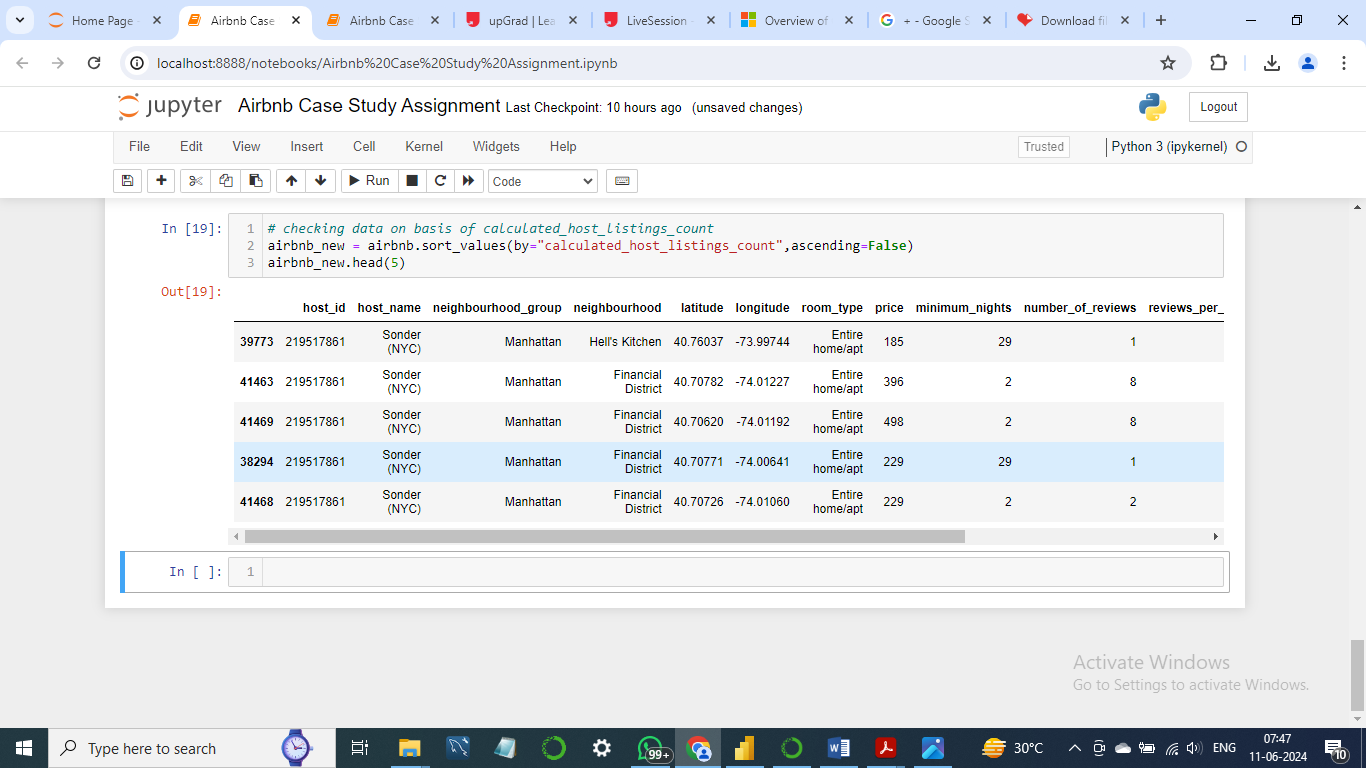
Number of Columns:16











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.

## Data Analysis and Visualizations using Power BI:

We have used Power BI to visualize the data for the assignment. Below are the detailed steps

used for each visualization.

**1) Top 10 Host:**

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

**2) Popular Neighbourhood:**

* 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.

**3) Average price of Neighbourhood 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 neighbourhood Groups in different colors. Also Put Avg price in Label.

**4) For Variance of price with Neighbourhood Groups:**

* We used a box and whisker’s plot with Neighbourhood Groups in Columns and Price

in Rows.

* We changed the Price from a Sum Measure to the median measure.

**5). Preferred Room type with respect to Neighbourhood Groups:**

* 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

**6) 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.

**7) 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.

Methodology Document PPT 2:

**1) Popular Neighbourhood:**

* 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.

**2) 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.

**3) 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

**4) Customer Booking with respect to minimum nights:**

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

**5) 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.

**6) 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.

**7) 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.

**8) Tools used:**

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