AIR BNB Case Study – SANDEEP SUMAN PRADHAN

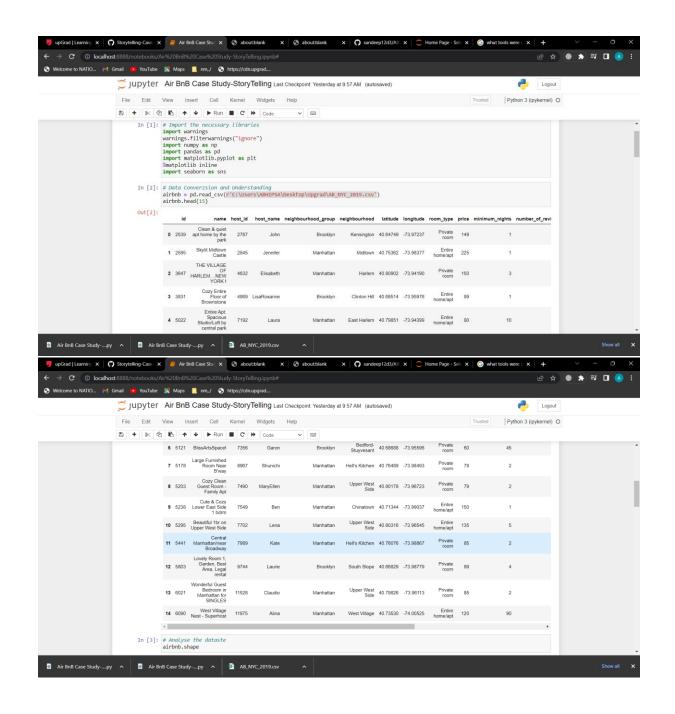
Methodology Document PPT-1:

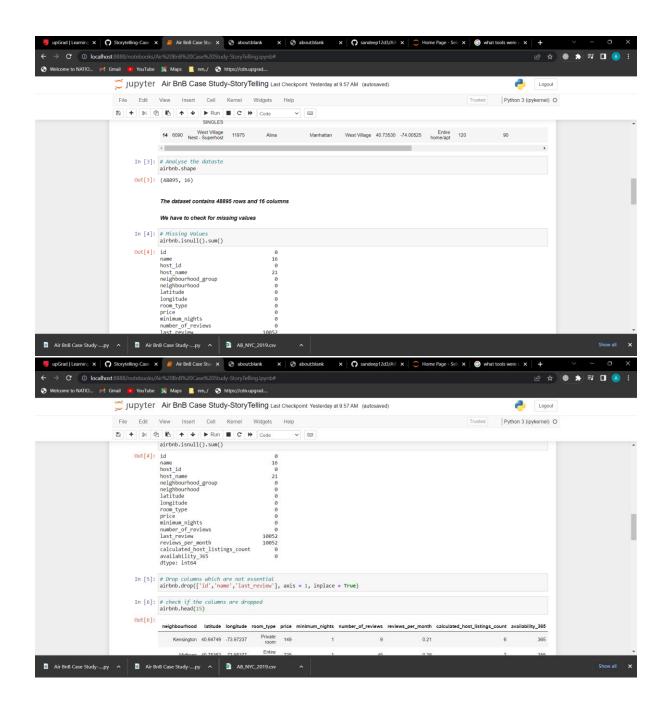
In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

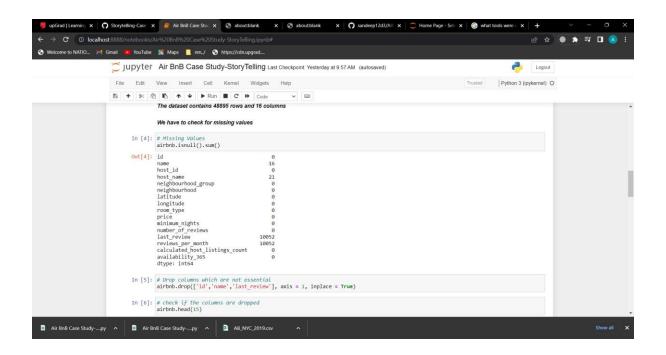
Analysis using Jupiter Notebook: Data Set Used: AB_NYC_2019.csv

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 for Duplicate rows in the dataset
- Checked for Null Values in the dataset. Columns like name, host-name, last review and review-permonth had null values.
- We dropped the columns where missing values were less and which would not impact our analysis in the dataset
- Identified and reviewed 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.

- 1) Top 10 Host:
 - We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map.
- 2) Preferred 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
- 3) 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.
- 4) 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.
- 5) 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.
- 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.

Methodology Document PPT 2:

- 1) 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
- 2) 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.
- 3) 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)
- 4) 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.
- 5) 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.
 - 6) 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.

7) Tools Used:

Data cleaning and preparation: Jupyter Notebook-Python

Visualization: Tableau

Data story Telling and Presentation : Microsoft Powerpoint