**AIRBNB NYC Case Study**

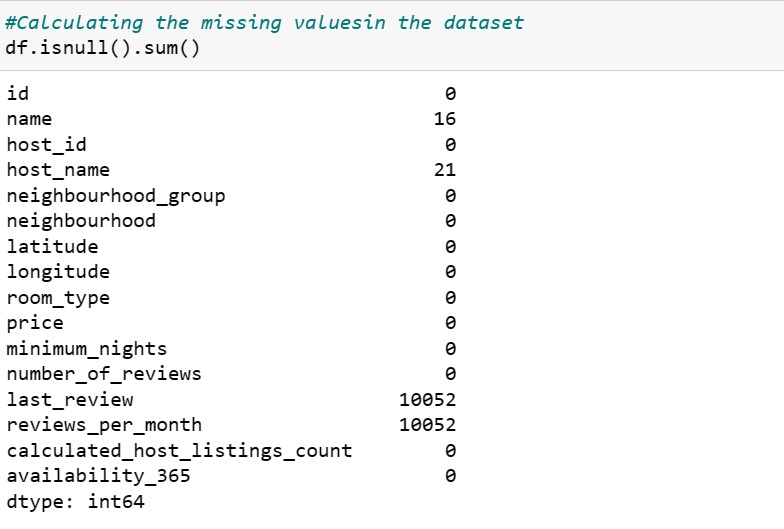
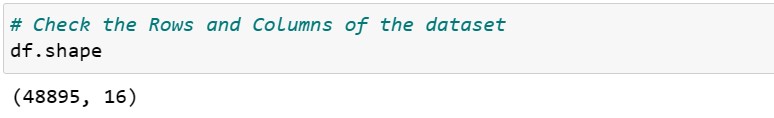
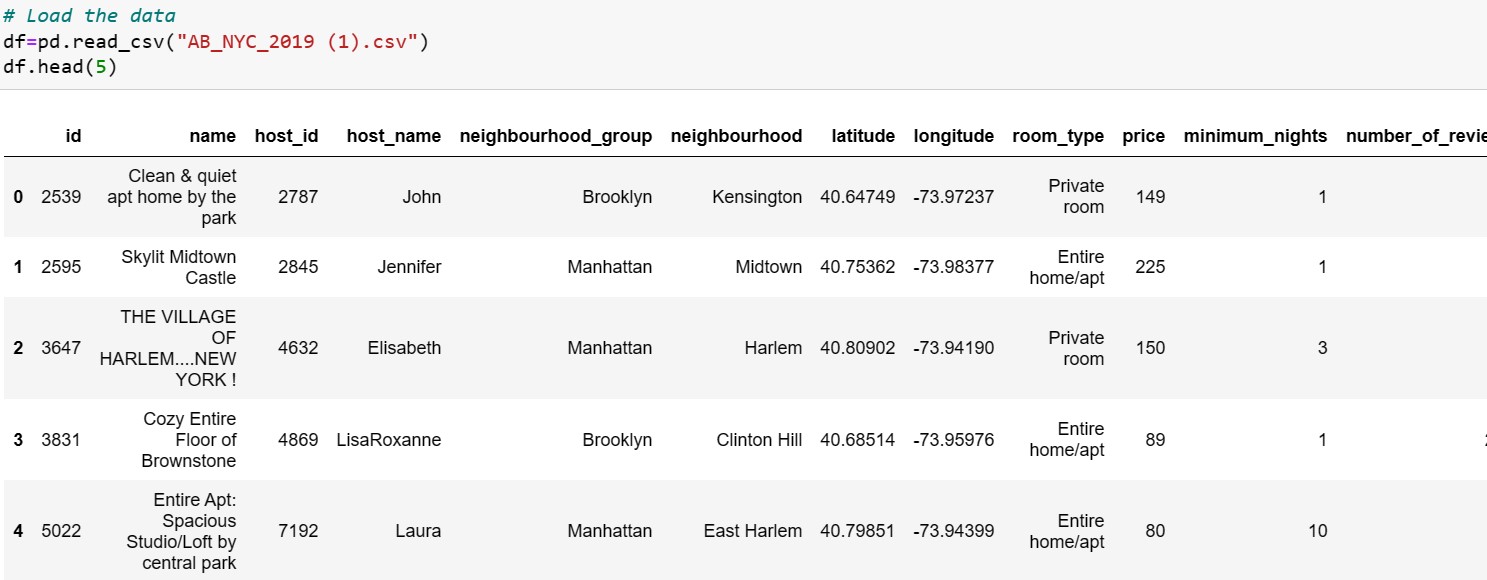
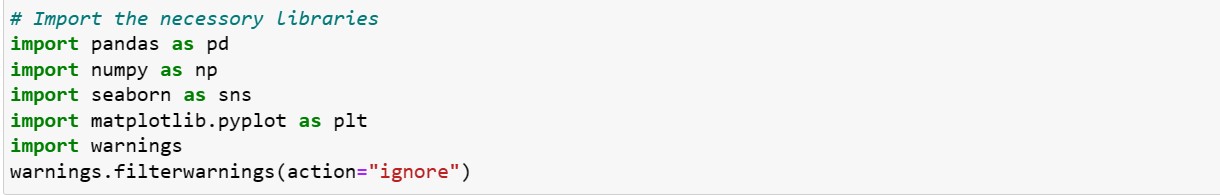
**Methodology Document PPT 1:**

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

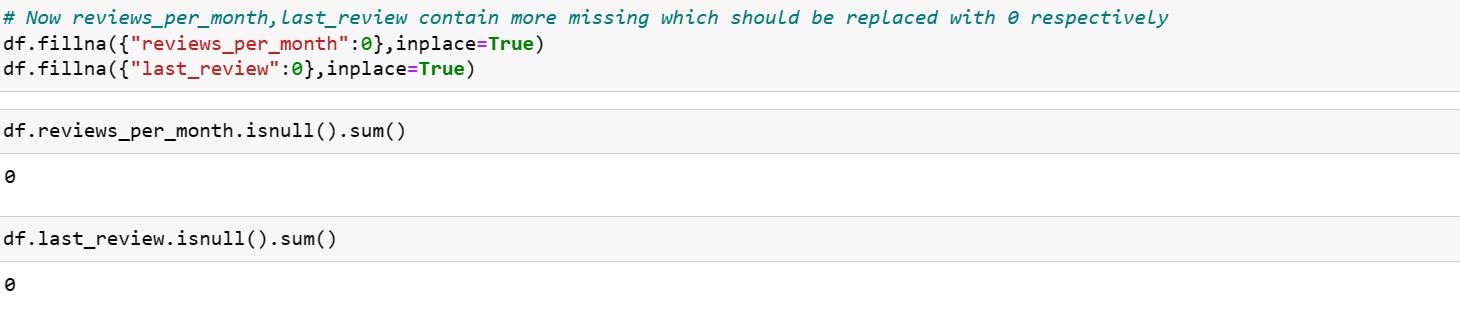
**Initial Analysis using Jupiter Notebook**: Data Set Used: AB\_NYC\_2019.csv

**Number of Rows**: 48895

**Number of Columns**: 16

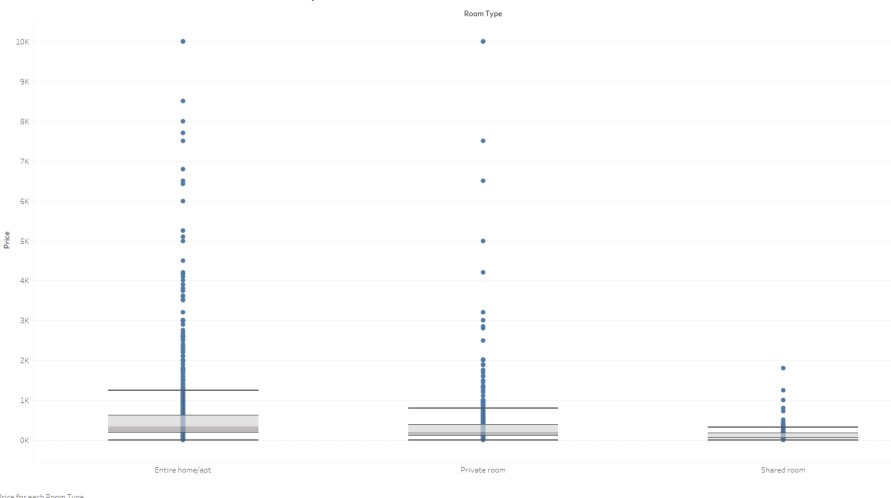


* We removed the columns Name which was not giving much information.



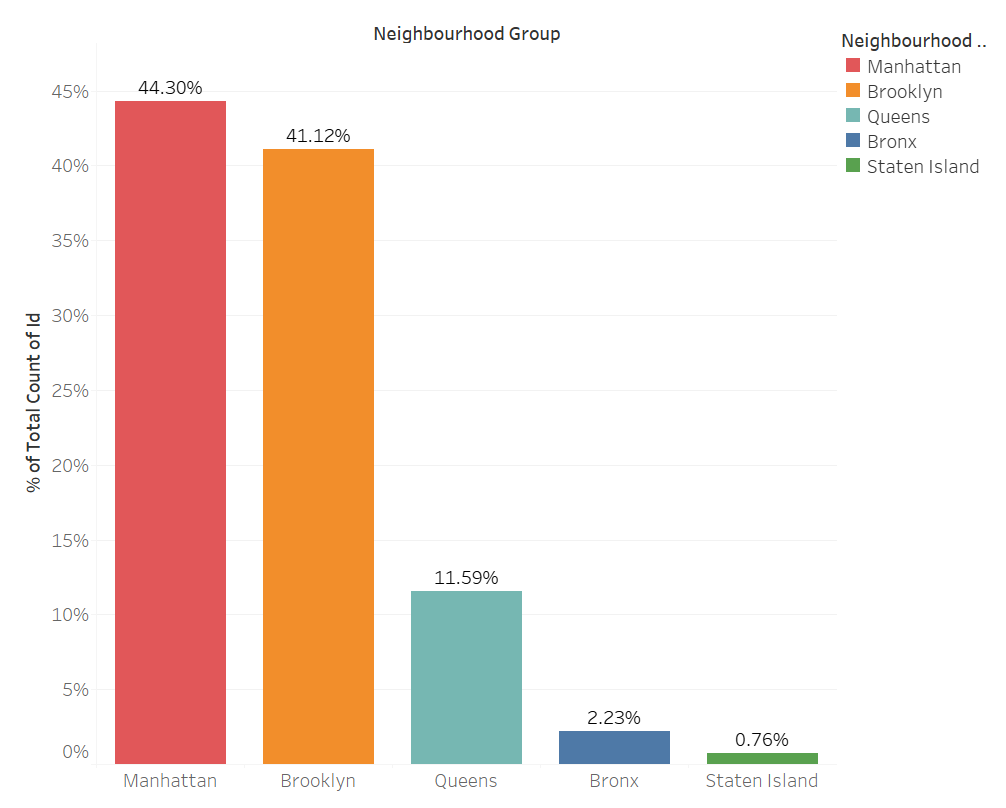
**Step 2: Data Wrangling:**

* Checked the Duplicate rows in our dataset and no duplicate data was found.  Checked the Null Values in our dataset. Columns like name, 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.



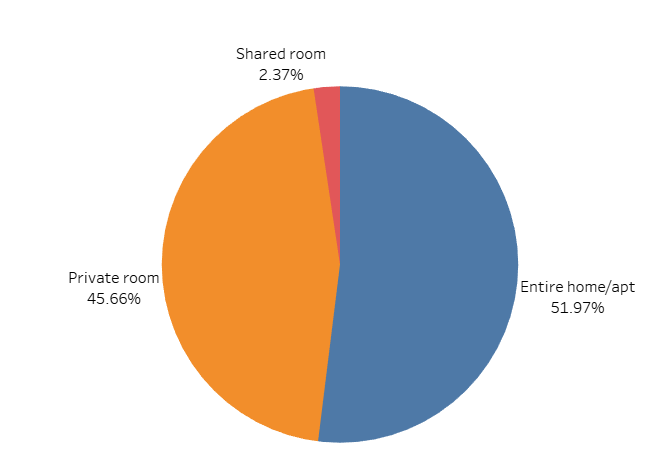
**Step 3: Data Analysis**

* Airbnb listings spread out in NYC



1. We see that, Airbnb has good presence in Manhattan, Brooklyn & Queens
2. Listings are maximum in Manhattan (44%) & Brooklyn (41%) owing to the high population density and it being the financial and tourism hub of NYC. Staten Island (~1%) has the least number of listings, due to its low population density and very few tourism destinations.

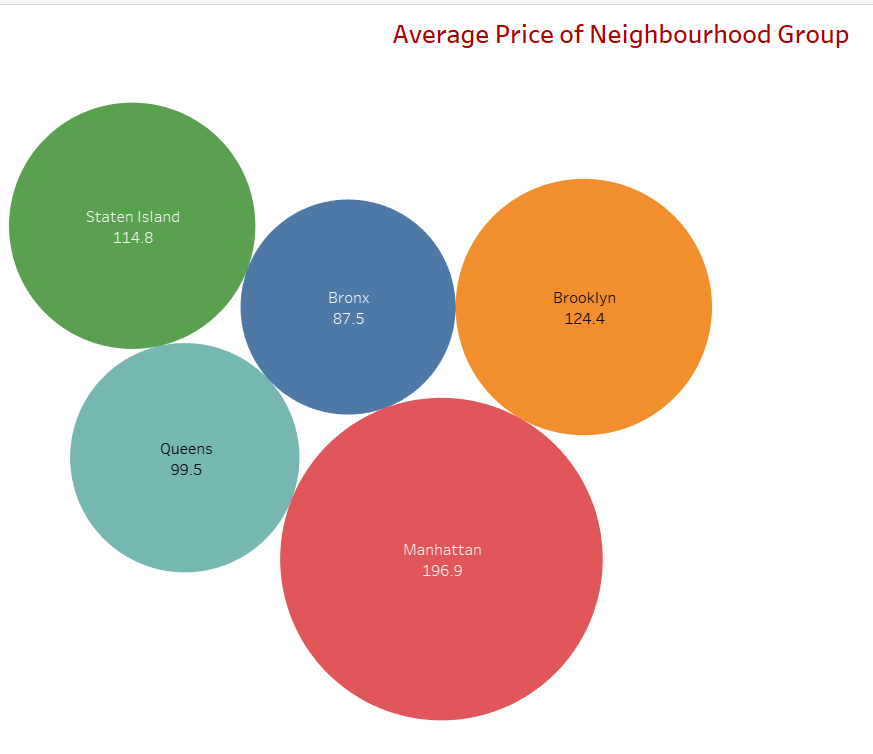
* Room Types preferred more by Customers



1.There are three types of rooms - Entire home/Apartment, Private room & shared room.

2. Overall, customers appear to prefer private rooms (45%) or entire homes (52%) in comparison to shared rooms (2.4%).

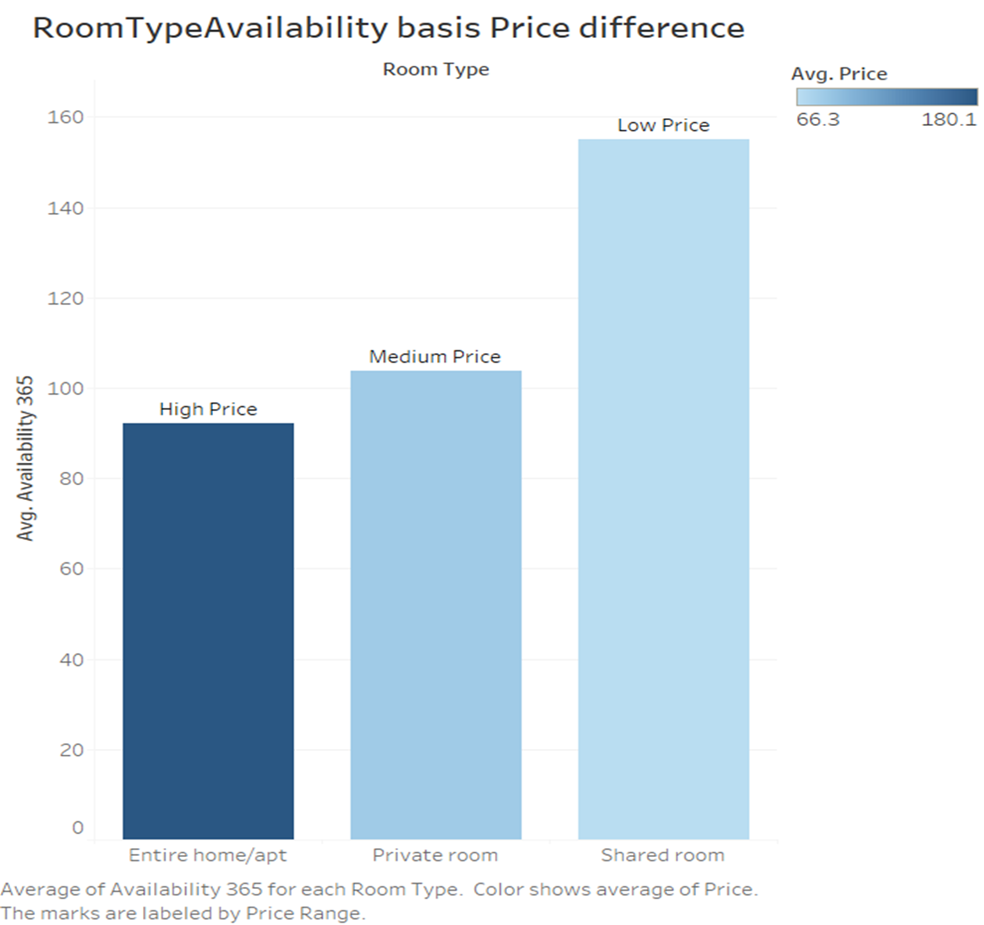
* Average price of Neighbourhood groups



1.The average price of listed properties in Manhattan is around 196.9, which is highest among all neighbourhoods.

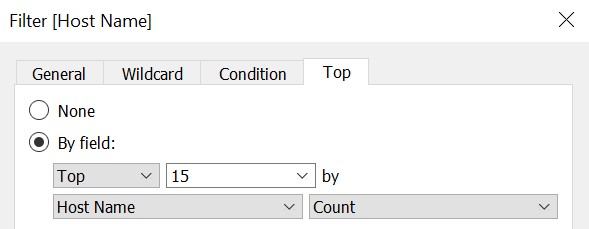
2. Average price for Brooklyn is second highest i.e. 124.4.

* Room Type availability basis Price difference

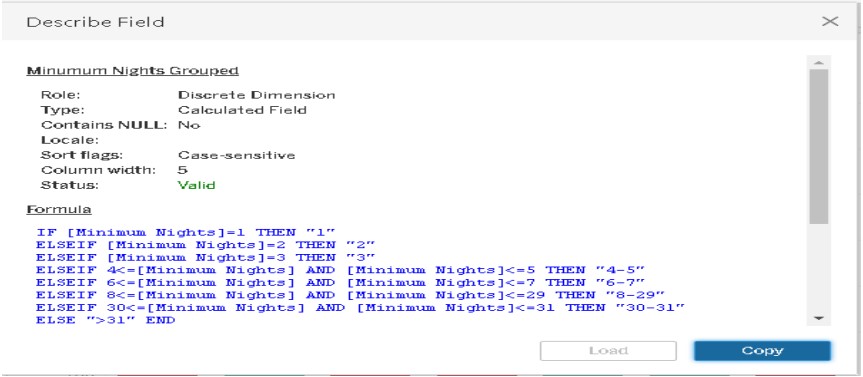


 **TOP 15 HOST HAVE RECEIVED HIGHEST NUMBER OF REVIEWS?**

* We identified the top 15 Host Name with Number of reviews using the Scatter Plot.



* **How are the Airbnb listings spread out in NYC?**
  + We wanted to understand the spread of listings in the NYC areas and the concentration of listings in each neighbourhood group. –
  + **Bar plot:** This was used to understand the concentration of the listings in each neighbourhood. We use the parameters Neighbourhood group & CNT(Id).
* **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
* **Customer Booking w r t minimum nights:** 
  + We created the bin for Minimum nights as shown below.



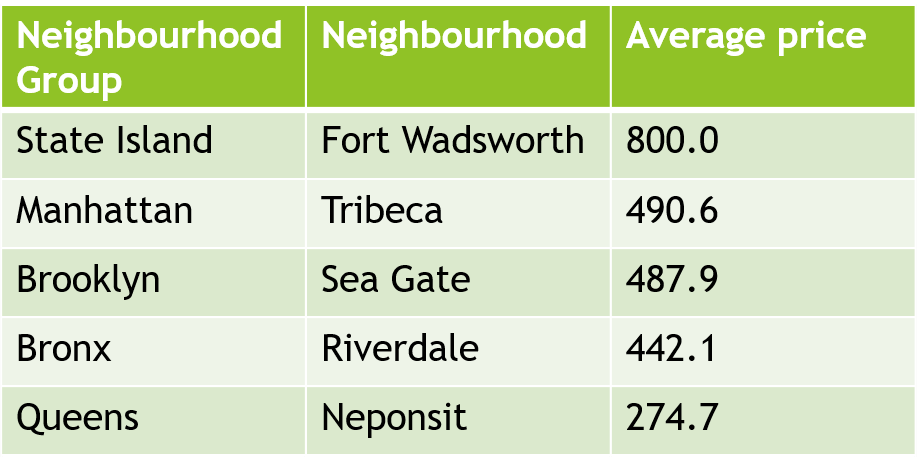
* + The bins were used to display the distribution of minimum nights based on the number of ids booked for each neighbourhood group.
* **Neighbourhood vs Availability:**
* We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

 **Popular Neighbourhoods:**

* We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour.
* We used filter to show Top 10 neighbours as per the sum of reviews
* We see that Bedford-Stuyvesant from Brooklyn is the highest popular with 1,10,352 no of reviews in total followed by Williamsburg**.**
* Harlem from Manhattan got the highest no of reviews followed by Hell’s kitchen.

 **Average Price According To Neighbourhood And Neighbourhood Group:**

* State Island with Fort Wadsworth neighbourhood is highest average price among all group



**Methodology Document PPT 2:**

 **What are the pricing range preferred by Customers w r t to Review?**

* We have considered the volume of booking and number of reviews in a particular price range. For easy visualization, we have binned the Price with a bin size of 10. Also owing to the enormous value range, we have observed the variation until $1000. As there was very little data beyond this, we decided to filter it
* We have taken pricing preference based on two parameters – volume of bookings done in a price range and number of reviews in a price range. From both the graphs, the favourable price range is $60- $190. This is the price range most preferred by most customers.

 **Which Room Types available for more minimum night stay contradicting their price range?**

 On average Entire home/apt typesare preferred more by thecustomers followed by Private rooms and then Shared Rooms. Mostly because they are also available for a higher number of minimum night’s stay window booking as compared to Private and Shared rooms.

 **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.

 **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
* **Types of Rooms available for N number of days in a year based on their price range.?**

Average of availability 365 for each Room Type. Colour showsof average of Price.

The marks are labelled by Price Range.

* **Popular Neighbourhoods:**
* We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour.
* We see that Bedford-Stuyvesant from Brooklyn is the highest popular with 1,10,352 no of reviews in total followed by Williamsburg**.**
* Harlem from Manhattan got the highest no of reviews followed by Hell’s kitchen.

 **Which Locations that have received the highest number of the Reviews?**

* Staten Island — Silver Lake, Staten Island — Richmond town, Staten Island — Eltingville, Staten Island — Huguenot, and Brooklyn — Manhattan Beach are the Top 5 locations with Low Price ranges that have received the highest number of reviews on average being the lowest in the Price range. On the contrary, Queens — Neposit, Manhattan — NoHo, Manhattan — Tribeca, Staten Island — Willow brook, and Manhattan — Flatiron District is the highest in the Price range and have received a low number of reviews.

* **Tools used:**

• Data cleaning and preparation: Jupyter notebook – Python

• Visualization and analysis: Tableau

• Data Storytelling: Microsoft PPT