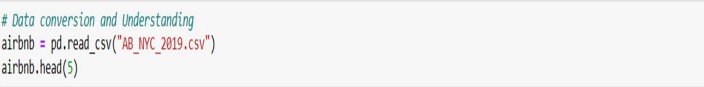
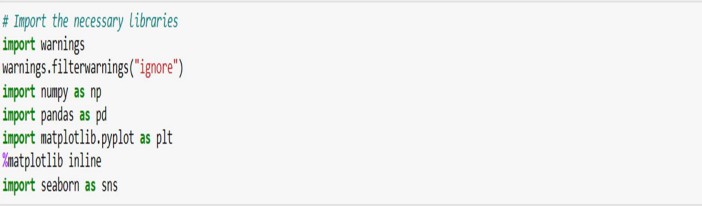
# Methodology Document PPT 1:

Python Jupyter notebook has been used in the case study to perform initial analysis of the data and Tableau for data analysis and visualization.

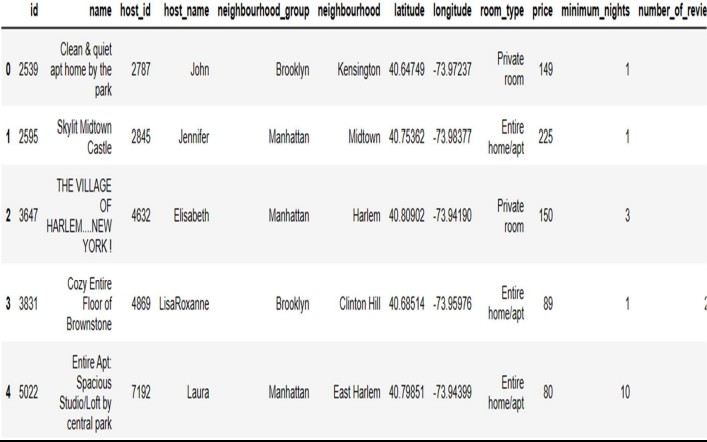
**Jupiter Notebook Initial Analysis :** Data Set Used: AB\_NYC\_2019.csv

**Number of Rows:** 48895

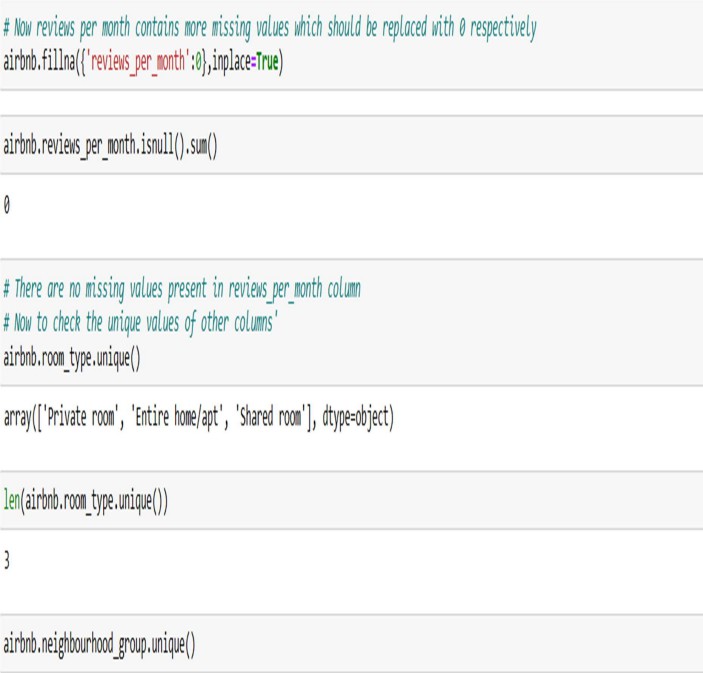
# Number of Columns: 16

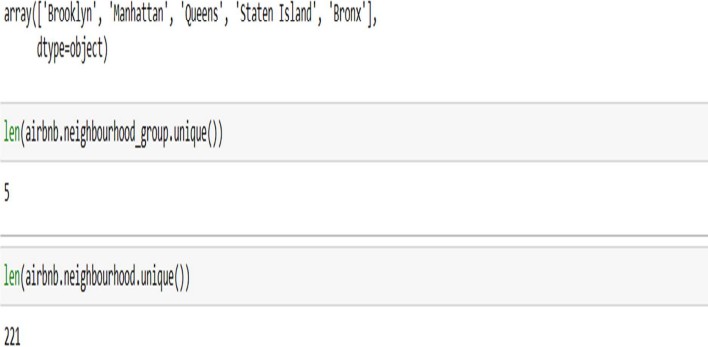






Columns like Id, Name, Last Review have been removed as they were not giving much information.





# Step 2: Data Wrangling:

* Duplicate rows were checked in our dataset and no duplicate data was found. \
* Null Values were checked in our dataset. Columns like name, host-name, last review and review-per-month have null values.
* Dropped the column name as missing values are less and dropping it won’t have significant impact on analysis.
* Formatting was checked in our dataset.
* Outliers were identified and reviewed.

# Data Analysis and Visualizations using Tableau:

Tableau was used to visualize the data for the assignment. Below are the detailed steps used for each visualization.

## Top 10 Host:

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



## Preferred Room type with respect to Neighbourhood group:

* + A Pie Chart was created for understanding the percentage of room type preferred with respect to 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.

## Neighbourhood Groups Price variance:

* + A box and whisker’s plot was used with Neighbourhood Groups in Columns and Price in Rows.
  + The Price was changed from a Sum Measure to the median measure.

## Neighbourhood groups Average price :

* + A Bubble chart was created 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. We have also Put Avg price in Label.

1. 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:

* + Dual axis chart was created using Bar Chart for availability of 365 and a line chart for price of top 10 neighbourhood group sorted by price.

**Methodology Document PPT 2:**

## Room type with respect to Neighbourhood group:

* + A Pie Chart was created to understand the percentage of room type preferred with respect to the neighbourhood group
  + Room Type was added 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.

## Availability vs Neighbourhood :

* + Dual axis chart was created using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

## Price Range preferred by Customers:

* + Pricing preference was taken 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.

1. Understanding Price variation with respect to Room Type & Neighbourhood:
   * Highlights Table chat was created 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.
2. Price variation w r t Geography:
   * Geo location chart has been used to plot neighbourhood, neighbourhood Group in map to show case the variation of prices across.
3. Popular Neighborhoods:
   * We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour.
   * Filter has been used to show Top 20 neighbours as per the sum of reviews.
4. Tools used:
   * Data cleaning and preparation: Jupyter notebook – Python
   * Visualization and analysis: Tableau
   * Data Storytelling: Microsoft PPT