

**Airbnb Listing EDA (New York-2024) Data Analysis Project**



**Project Overview**

The project involves performing an Exploratory Data Analysis (EDA) on New York Airbnb data to derive insights into rental trends and patterns. Pandas, NumPy, Matplotlib, and Seaborn are the primary libraries used for cleaning, analysing, and visualizing the data.

**Objective**

The goal of this project is to:

1. Analyse **room types, prices, and availability** across different neighbourhoods.
2. Understand **host behaviour** and listing patterns.
3. Detect potential **outliers** in prices.
4. Provide recommendations for guests and hosts based on insights.

**Dataset**



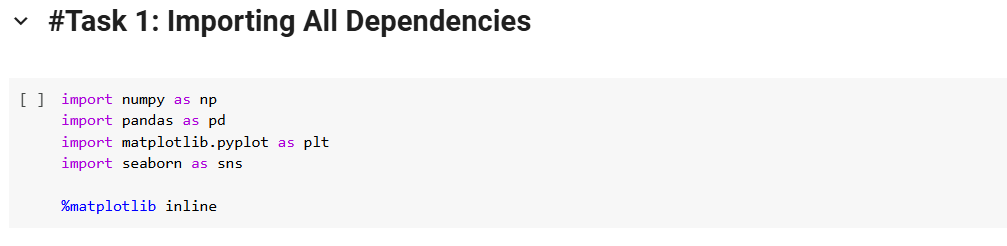
The dataset contains **20,765 entries and 22 features**, including:

* **id**: Unique identifier for each listing
* **name**: Title of the Airbnb listing
* **host\_name**: Name of the host
* **neighborhood\_group**: Group (borough) where the listing is located
* **latitude/longitude**: Geolocation of listings
* **price**: Nightly rental price
* **room\_type**: Type of accommodation (e.g., entire home, private room)
* **reviews\_per\_month**: Average monthly reviews for the listing
* **availability\_365**: Number of available days in the year

**Steps and Workflow**

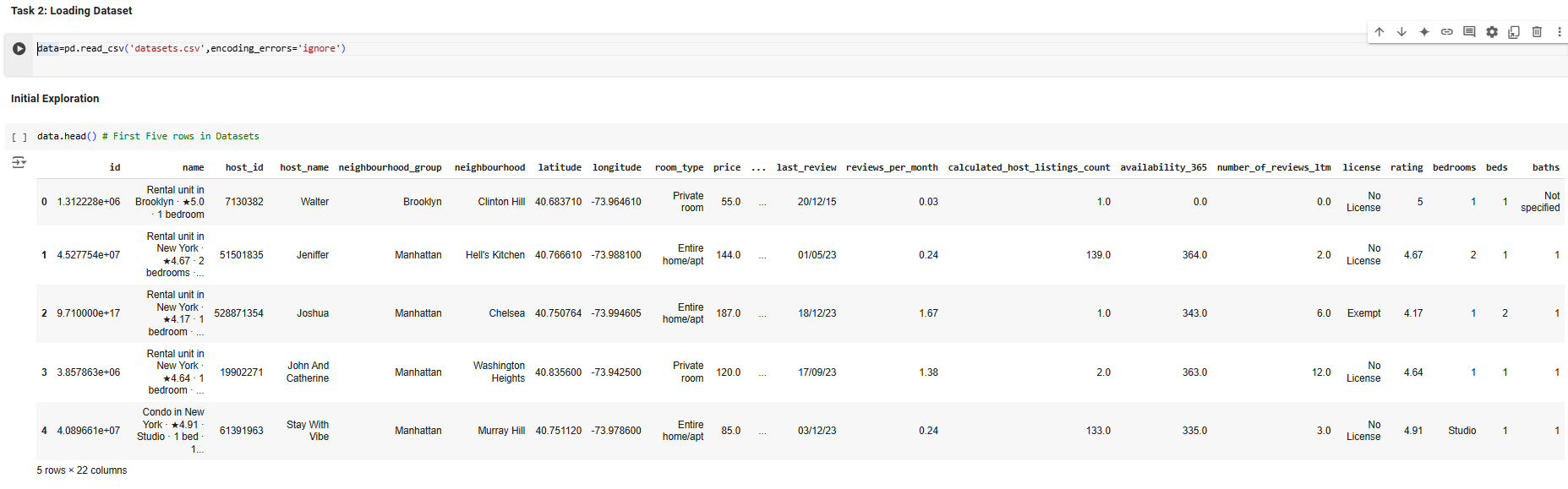
**Importing All Dependencies**

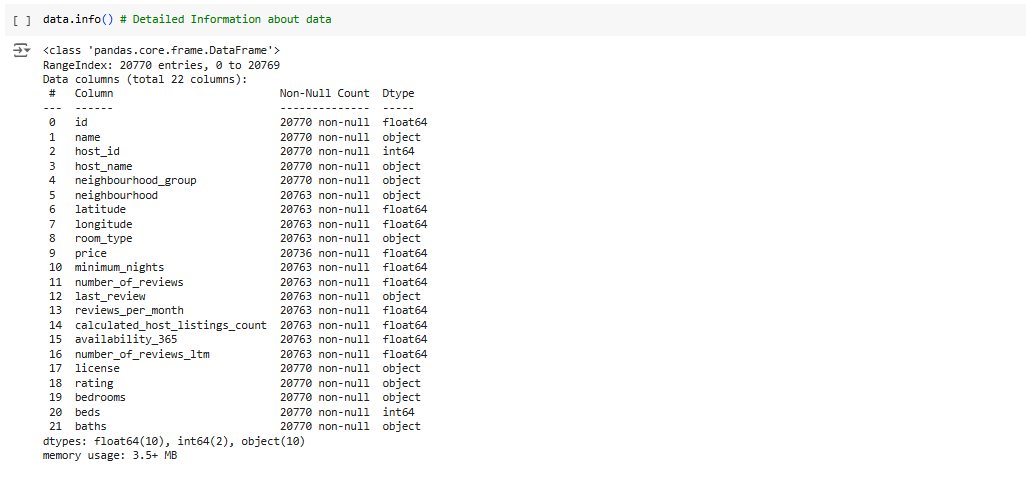
We begin by importing essential Python libraries such as NumPy, pandas, Matplotlib, and Seaborn. These libraries aid in tasks like analysing duplicate rows, handling missing data, changing data types, and uncovering various insights. Key insights include locality and review dependency, neighbourhood group patterns, availability trends, price distribution, geographical distribution, and review behavior.

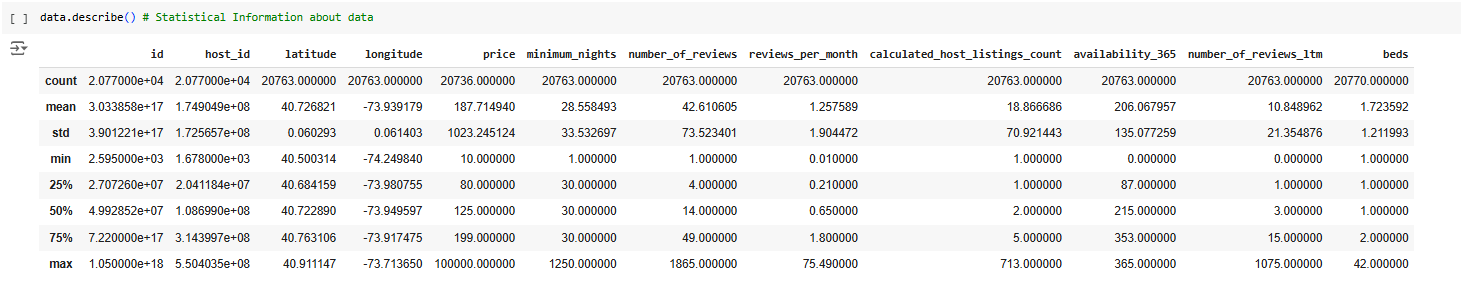


**Loading Dataset and Statistical Information about data**

We imported the dataset, a dataset.csv file obtained from the Kaggle website. The dataset was uploaded to Google Collaboratory, where we ran the commands shown in the screenshots below. We analysed the Airbnb dataset to extract overall details and described its statistical information.



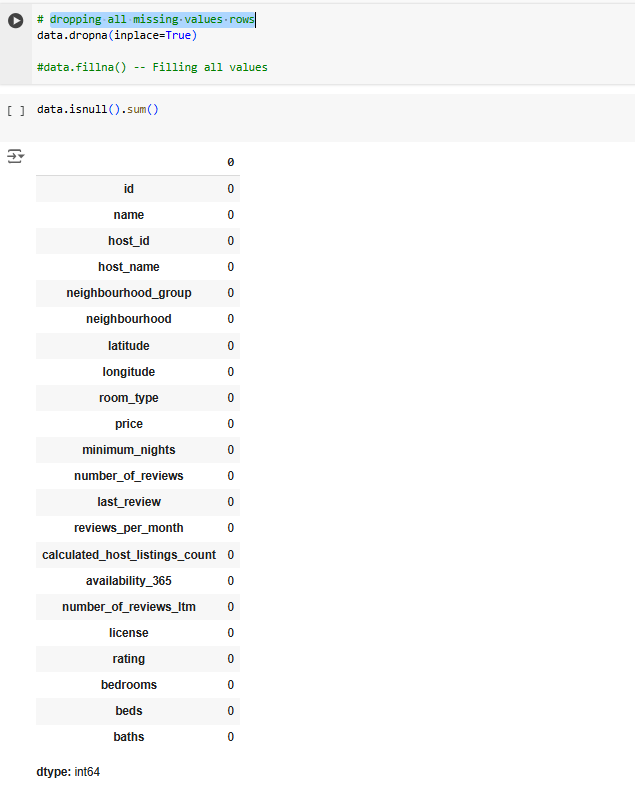


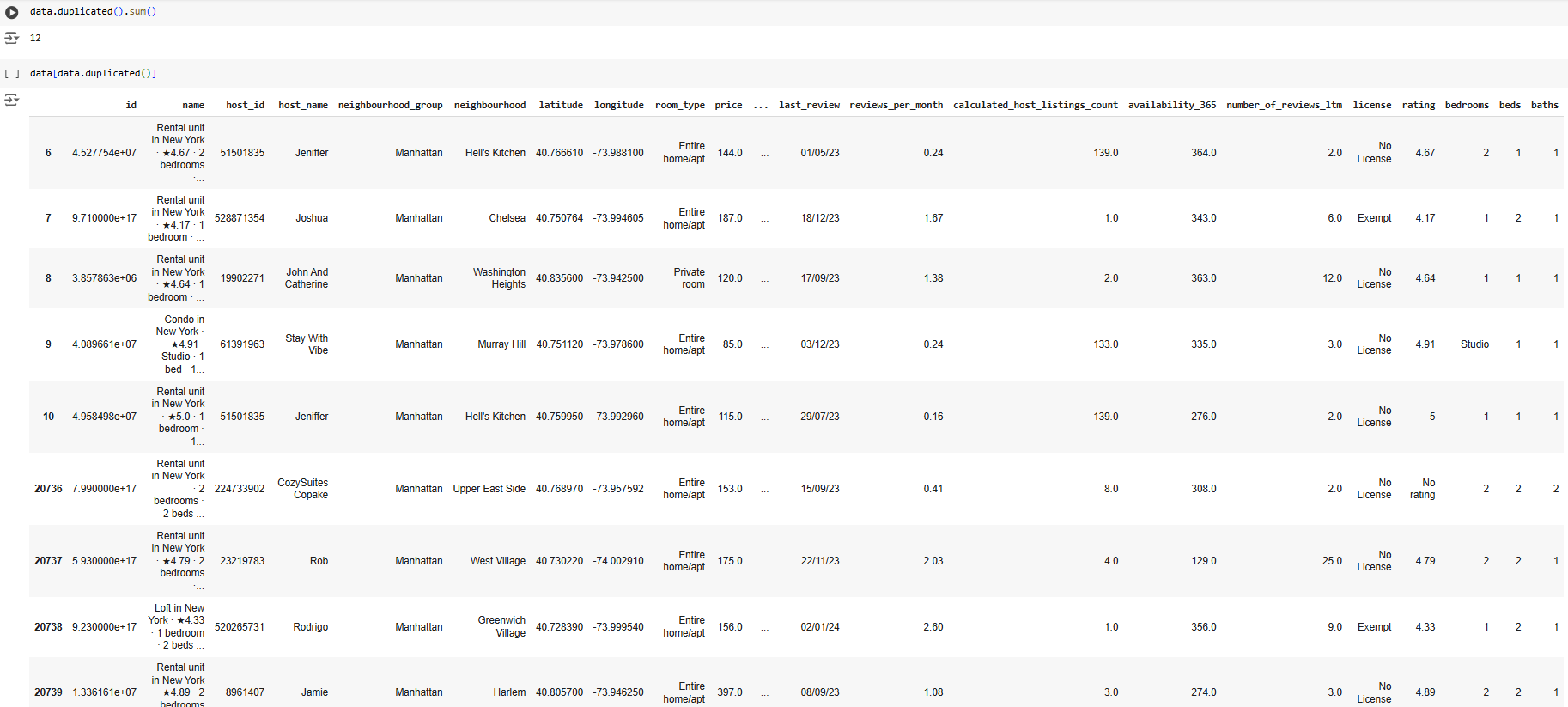


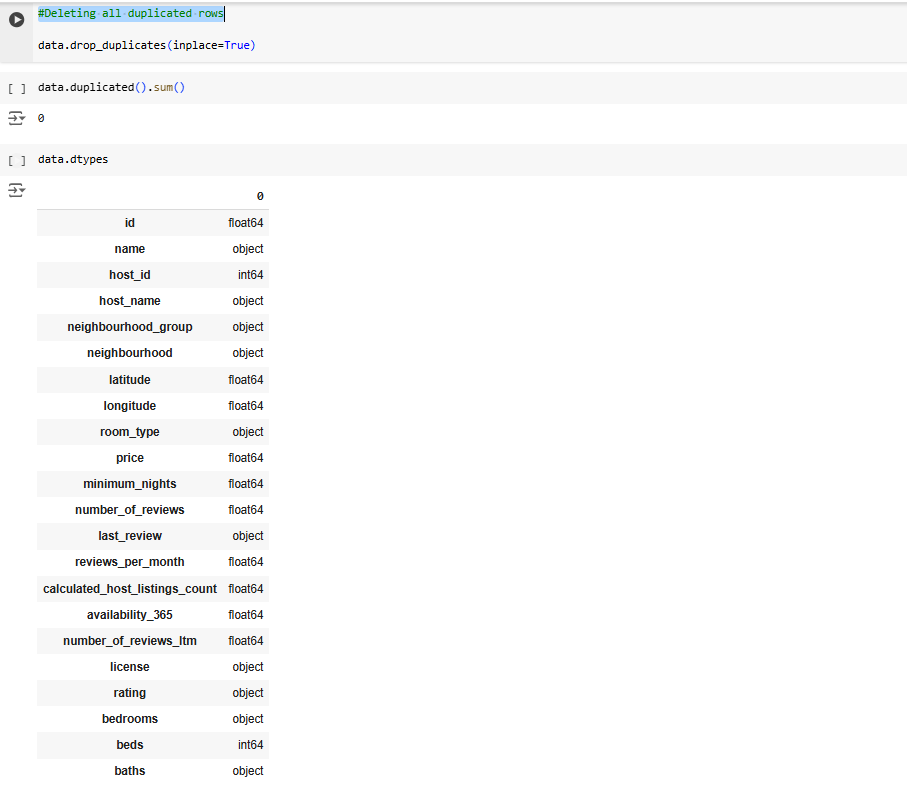
**Data Cleaning**

* **Handle missing data**: price, neighborhood, beds, latitude, longitude, room\_type, minimum\_nights, number\_of\_reviews, last\_review, reviews\_per\_month, calculated\_hosted\_listing\_count, availability\_365, numbers-of\_reviews\_itm columns had null values, dropping all missing values rows and Dealing Duplicate Rows.

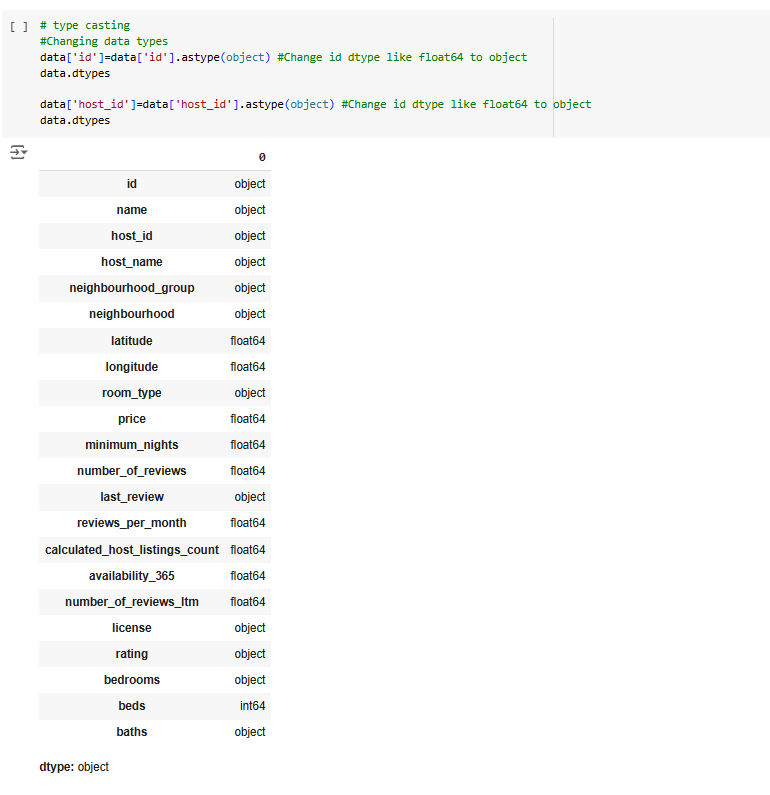




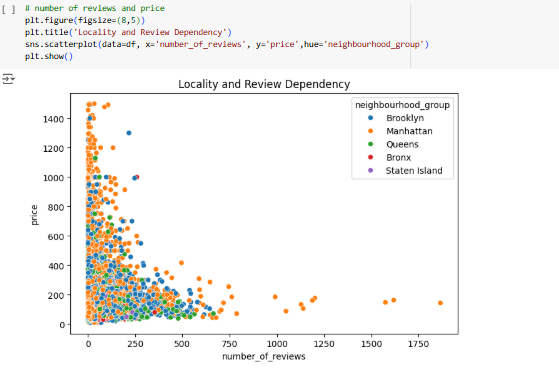




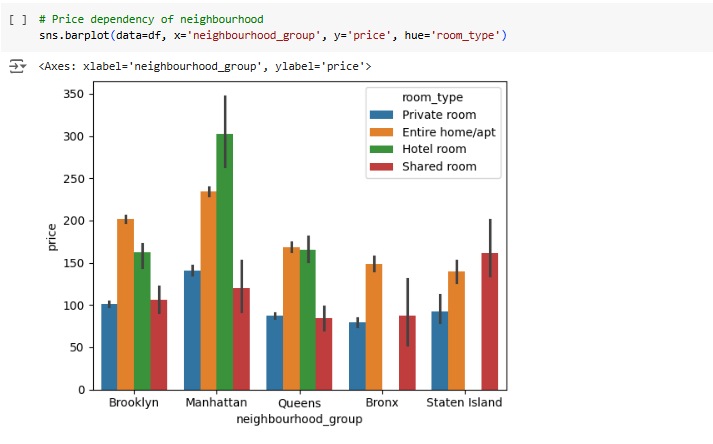
* **Fix data types**: Converted datatypes like id and host\_id to a **float64** object.



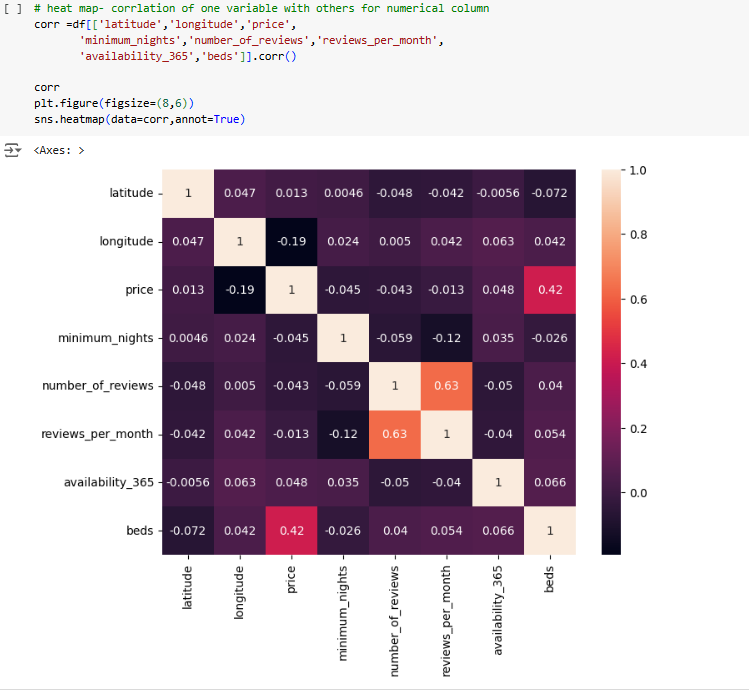
* **Locality and Review Dependency**:
  + Analyse no\_of\_reviews, price and neighbourhood\_group for showing locality and Review Dependency using Scatterplot.



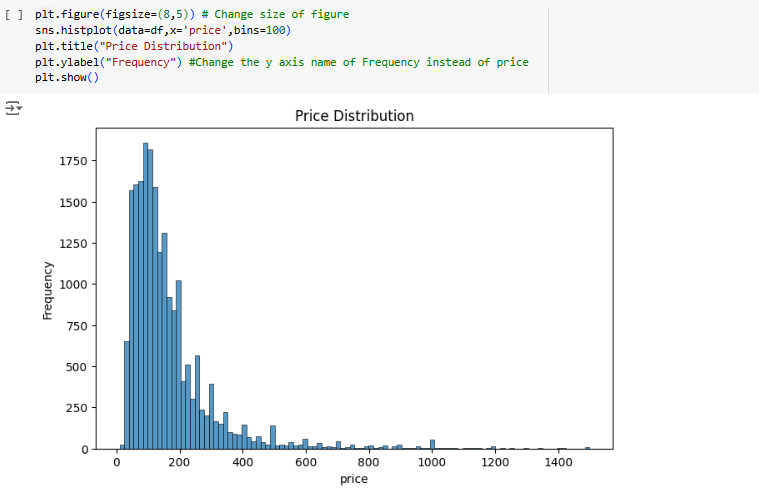
* **Neighborhood group insights**:
  + Analyzed **price variations by boroughs**.
  + Manhattan had the **highest average prices** using bar plots.



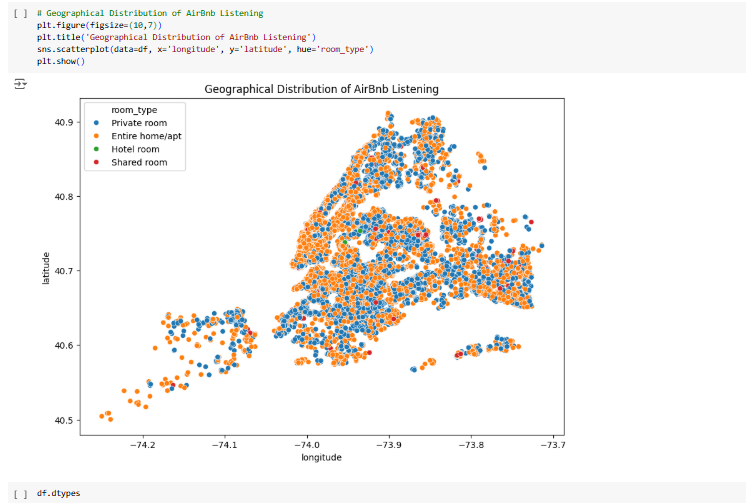
* **Availability trends**:
  + Used **heatmaps** to show correlations among price, availability\_365, number\_of\_reviews, beds, latitude ,longitude, minimum\_nights and reviews\_per\_month.



* **Price distribution**:
  + Used **histograms** to show the distribution of prices.
  + Majority of the listings were priced between **$0 - $1500**.



* **Geographical Distribution of Airbnb Listing:**
  + Analysed distribution with multiple listings using **scatterplot** to identify listing like Private Room, Entire Home/Apt, Hotel Room, Shared Room.



* **Review or Host behavior**:
  + Used **pair plots** to show relationships between number of reviews, price, and availability.



**3. Data Visualization**

* **Pairplot**: To see correlations among price, availability, and number of reviews.
* **Heatmap**: Showing correlations among numerical features.
* **Histograms and Boxplots**: To detect outliers in price.
* **Bar Charts**: Displaying room types and neighborhood group distributions.

**Key Findings and Insights**

1. **Price Trends**:
   * **Manhattan** has the most expensive listings, followed by Brooklyn.
   * **Entire homes/apartments** cost significantly more than private or shared rooms.
2. **Room Type Distribution**:
   * **Entire homes/apartments** are the most common, but **private rooms** offer budget-friendly options.
3. **Outliers in Price**:
   * Few listings priced greater than **1500** were detected, indicating the need to filter such extreme values.
4. **Availability Patterns**:
   * Listings with **high availability** tend to have lower prices and more reviews, likely due to better guest experience.
5. **Host Behaviour**:
   * Some hosts manage **multiple listings**, indicating a trend toward professional hosting.

**Recommendations**

* **For Guests**:
  + Look for listings with high availability and good reviews for a better experience.
  + **Private rooms** in Brooklyn offer affordable stays compared to Manhattan.
* **For Hosts**:
  + Improve **availability** and **review response rates** to attract more bookings.
  + Manage pricing effectively to compete within the borough's market.

## Conclusion

This project offers valuable insights into the New York Airbnb market, helping both guests and hosts make informed decisions. By using **EDA techniques**, we identified key trends and developed actionable recommendations. Future improvements can involve advanced analytics and predictive modelling to further enhance the findings.