Reservation Analysis for Airbnb

Problem Statement

AIRBNB's main function is to offer homestays for vacation rentals and tourism-related activities; the company does not own any of the houses on its list. Simply put, it's an internet marketplace with all the properties posted.

House Types and Various Neighborhood Groups

There are three different types of Houses listed in Airbnb.

- Entire room/Apt
- Shared Room
- Private Room

There are five neighborhood groups.

- Brooklyn
- Manhattan
- Queens
- Staten Island
- Bronx

Observations

Price Variation

While prices vary depending on the style of property, on average, the costs of an entire home or apartment are more costly, whereas a shared room costs less. When it comes to neighborhood pricing, Manhattan has pricey homes of all kinds.

Housing distribution on the map

About 49000 houses are on the list. where Staten Island has 373 fewer listings of residences while Manhattan has 21661 more listings.

Availability of Hosts

The dataset availability_365 column talks about the host availability for 365 days, if it is 0 then the host is busy.

Dataset

The dataset comprises 16 columns and roughly 49000 rows.
The term "host_name" refers to the name of the host providing services to guests.
neighborhood_group: represents the city
neighborhood: represents areas of the city.
latitude and longitude represent the house's location.
room_type: represents the type of room (private, shared, or apartment)
price: represents the cost of the houses.
minimum_nights: the number of nights spent by customers.

last_review: the date of the last customer review
reviews_per_month: the number of reviews per month
calculated_host_listings_count: the number of hosts listing.
availability_365: the number of hosts available annually

Actions Required

Investigative Data Analysis

We loaded the dataset, then used a variety of charting approaches to do EDA. A few graphs based on the dataset's features have been presented visually. The dataset yielded some insights for us.

Null values Management

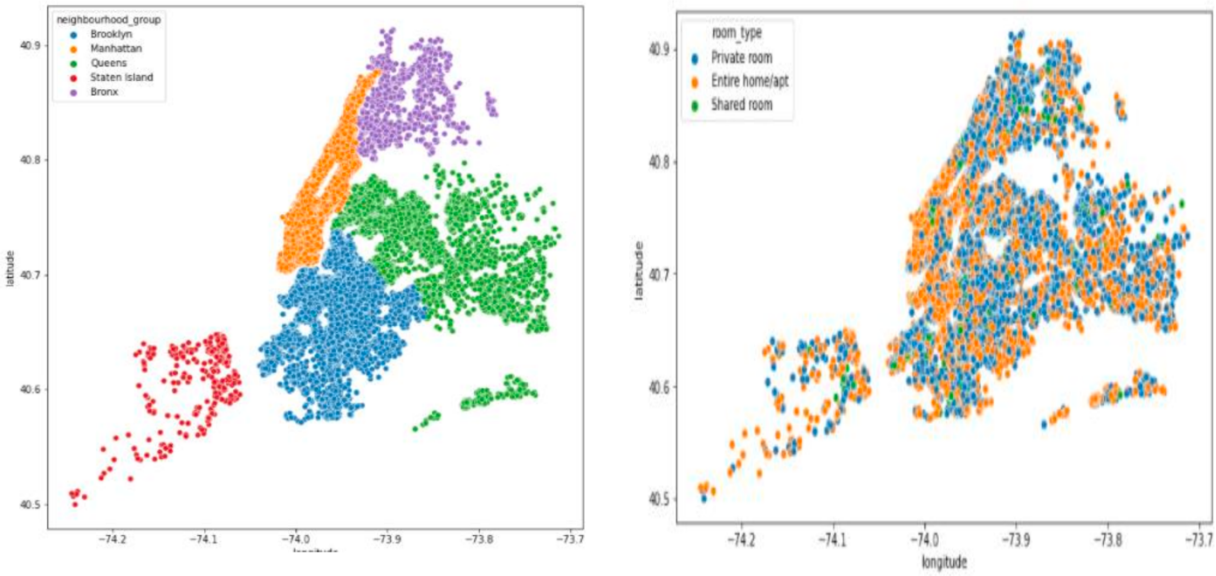
Numerous null values can be found in our dataset, which could have an impact on our findings. We have added zeros to the reviews column in place of null entries. We have substituted a median for the null entries in the price column. For the analysis, unnecessary columns were eliminated.

Displaying information on a Leaflet map

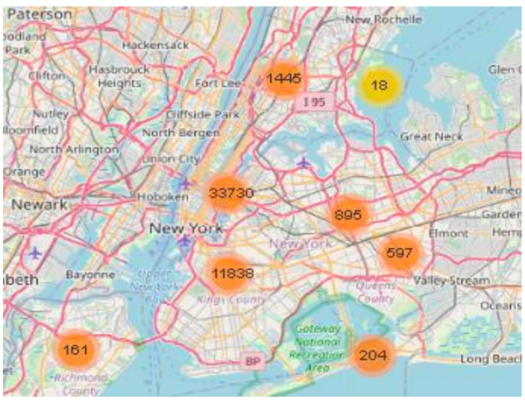
We received each house's latitude and longitude information. Plotly libraries and folium have been used to create a map-based visual representation of the data. Using scatter plots and other visual aids, we visually portrayed neighborhood groups and each home.

Visualization

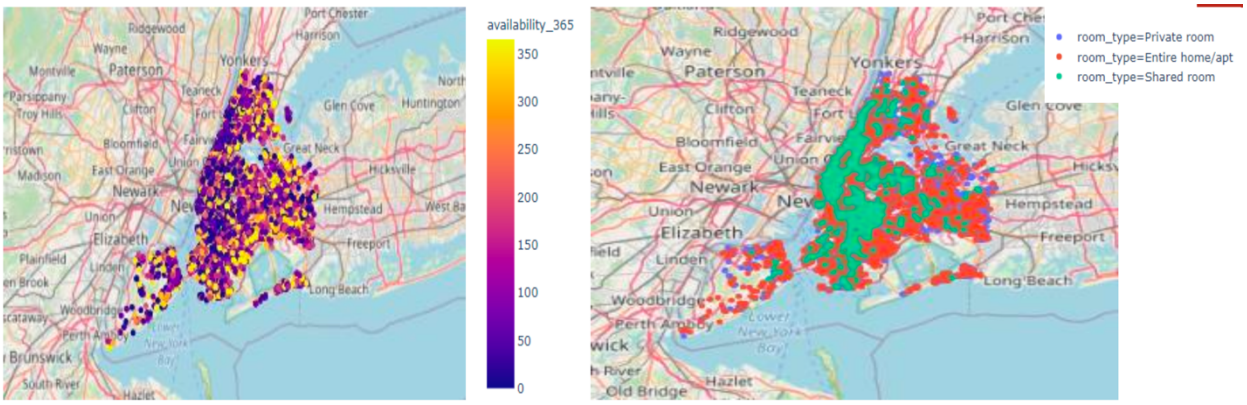
Using latitude and longitude data, we have created two scatter plots above that show the shared room types and neighborhood groupings throughout the New York City map.



The two plots above display the total number of hotels as clusters. One image is zoomed out, while the other is zoomed in, allowing us to observe the hotels' distribution over the map.



The hotels that are open for business are displayed in the first of the two plots above, while the second plot displays different kinds of rooms dispersed throughout the NYC area.



Conclusion

One of the most popular platforms for vacation rentals and lodging is Airbnb, which specializes in homestays. These days, one of the most popular brands for providing hosts and guests with a positive experience is Airbnb. To make traveling simple and convenient, the data is used to improve knowledge of every little detail. The information is used to demonstrate the necessary conclusions.16 columns and 49000 observations with null and missing values comprise the data that we have been provided. There was inaccurate information present. Gaining understanding of the data and drawing conclusions from it was crucial. Finding the columns with the highest number of missing values is the first step. Replace the zero values to address the missing values. Some of the values in the price column are missing, thus the zero values and missing values are replaced by the column's median (in this example, the median is used instead of the column's mean because the data contains too many outliers).

In the reviews_per_month column, NA values had been replaced with zeros. After performing operations on the data, the summary statistic was examined by outlining the methodology and concluding that no insignificant values were present. It is now possible to create visualizations using the data. A range of findings are concluded with the help of plotly express and folium libraries.

This dataset contains AIRBNB data from 2008 onwards. This dataset contains 16 columns, around 49000 observations, and a combination of categorical and numerical values.
The detailed Insights from the analysis:
The Neighborhood of the New York city has 5 groups:

- Brooklyn
- Manhattan
- Queens
- Staten Island
- The Bronx

Properties from Manhattan are a bit pricey followed by Brooklyn and Staten Island
Top 3 Hosts from the dataset are:

- Sonder(NYC)
- Blue ground
- Kazuya

Since, Reviews are important here are the Top 3 Hosts who hold the most reviews:

- Dona
- Asa
- Dennis & Nauko

There are 3 room types Entire home/apt. (~25000), Private room (~23000), Shared room(>500)
The visualization that plotted on maps also shows where all the properties are located along with their price, availability, and the type of the room.