**AirBnB NYC DataSet:**

AirBnB offers a platform for rental properties. Let it be Apartments, Houses or single rooms AirBnB got it. The prices of AirBnB varies upon location, amenities provided by the hosts and etc. There are multiple factors that effect the price on AirBnB. For the brief understanding about the price predictions and factors affecting the prices I have gone through article published by ‘Gustavo Santos’.

Here in this project the author used python on exploring ‘NYC Airbnb dataset’.

The initial learning from using basic .describe() function were:

-Most popular host with highest number of appearances

-In overall NYC, most of the properties were located in the ‘Manhattan’ island(44%).

-Most of the bookings were for homes or apartments but not single rooms.

-average price in NYC is $152 and the average bookings were for weekly.

**Price Variations:**

People seek stay for a week or further longer than 7 nights in NYC. By further exploration using graphs for analysis, it concluded that the popular hosts focus on renting their properties for 30 days for people who are on business purpose rather than tourist. And the avg expense would be around 120$ to 150$. And according to 2019 data in similar locations, the prices were better than hotels.

Chart, histogram

Description automatically generated

Almost 85% of the listing were from Manhattan and Brooklyn.

Map

Description automatically generated

Heat Map created using Folium Map package to demonstrate where are the real estates and the prices(Red means Expensive)

**What does the Top 10 hosts offer different than others?**

First and foremost, thing is the location. All their properties were located in the heart of NYC.

Focusing on business rentals over tourism.

They charge over the average prices; As the location act as a key factor.

**DATA CLEANING:**

For analysis;

In the Airbnb dataset, some columns obviously have improper formats which interrupts the analysis process. For example, variables (price, acceptance rate, response rate, date etc.) with string datatype. And few characters like ‘$’, ‘%’ or any currency symbols need to be removed.

For the room types the data will be saved in the same column. For example, 1.5bathroom, 2bathroom etc. It is better that these columns should be split into two different columns.

The Boolean columns should be converted to binary.

**Based on grouping some of the main features we can maintain the integrity of analysis. Few of them are;**

**Hosts;** Based on the response time of the host the prices vary. So, Host response time column should be sorted.

**Availability** ; The prices of rooms vary on their availability. We should be able to see the availability of rooms for next year also.

**Size;**  Room prices vary based on the length and width also. There will price difference because of the size even though the room type is same.

**Location;**  The closer we are to the significant place the higher the prices.

This way we need to nurture the code. So, that the visualization of the data will be appropriate to the customer needs as well as we can satisfy the host also.