**Exploratory Data Analysis Project of Airbnb Bookings**

Avinash Arora

Cohort Berlin Data Science, Alma Better

**Abstract:**

## This the technical document report of our project “Exploratory Data Analysis Project of Airbnb Bookings” as a part of our course at Alma Better. I started reading the data by analysing what the data set contains. What type of informations are there in the dataset.

## Contents:

I are here to explore a Airbnb dataset Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

**WORK FLOW:**

# Problem statement & collection of data: Identity the problem of analysis to be done and then collection the data regarding it.

# Data Cleaning & Manipulation: Raw data will have lot of null values and missing data. so it needs to be cleaned and manipulated as per our analysis

# Exploration Data Analysis: By using python libraries we write code to study and extract result.

# Data Visualization: The results of our analysis will be displayed in a visualization manner for better readability and understandability.

# Conclusion & Proposed Actions: Based on the EDA results conclusions will be drawn and necessary actions will be proposed to mitigate the risk.

## 

**Approach:**

1-First check dataset's and understand it.

2-Later I will check for any missing data in the data given.

3-I would check the type of data and divide it for our analysis.

4-I checked where there any outlier or unethical data in it if so, I would filer such

data for specific analysis.

5-Then do Data analysis by visualization techniques.

6-And then conclude with various outcomes from it.

**Dataset's**:

The Dataset contains

No. of Rows = 48895

No. of Columns = 16

**<class 'pandas.core.frame.DataFrame'>**

**RangeIndex: 48895 entries, 0 to 48894**

**Data columns (total 16 columns):**

**# Column Non-Null Count Dtype**

**--- ------ -------------- -----**

**0 id 48895 non-null int64**

**1 name 48879 non-null object**

**2 host\_id 48895 non-null int64**

**3 host\_name 48874 non-null object**

**4 neighbourhood\_group 48895 non-null object**

**5 neighbourhood 48895 non-null object**

**6 latitude 48895 non-null float64**

**7 longitude 48895 non-null float64**

**8 room\_type 48895 non-null object**

**9 price 48895 non-null int64**

**10 minimum\_nights 48895 non-null int64**

**11 number\_of\_reviews 48895 non-null int64**

**12 last\_review 38843 non-null object**

**13 reviews\_per\_month 38843 non-null float64**

**14 calculated\_host\_listings\_count 48895 non-null int64**

**15 availability\_365 48895 non-null int64**

**dtypes: float64(3), int64(7), object(6)**

**memory usage: 6.0+ MB**

From the above information I understand that few of the hosts are not listed at Airbnb as there are null values.

**All the columns present in the dataset**:

**Index(['id', 'name', 'host\_id', 'host\_name', 'neighbourhood\_group', 'neighbourhood', 'latitude', 'longitude', 'room\_type', 'price', 'minimum\_nights', 'number\_of\_reviews', 'last\_review', 'reviews\_per\_month', 'calculated\_host\_listings\_count', 'availability\_365']**

**Exploratory Data Analysis:**

Now we know that we are ready for an exploration of our data, we can make a rule that we are going to be working from left to right. The reason some may prefer to do this is due to its set approach – some datasets have a substantial number of attributes; plus, this way we will remember to explore each column individually to make sure we learn as much as we can about our dataset's.

Initial Observations :

#checked id are same or diffrent

#expermenting (one host have diffrent airbnb like john have 294 airbnd )

#some hotels have name repeated

#indiviual name have multiple airbnb

#221 loctions are avilable for airbnb

#some property is not listed yet or dont have host name

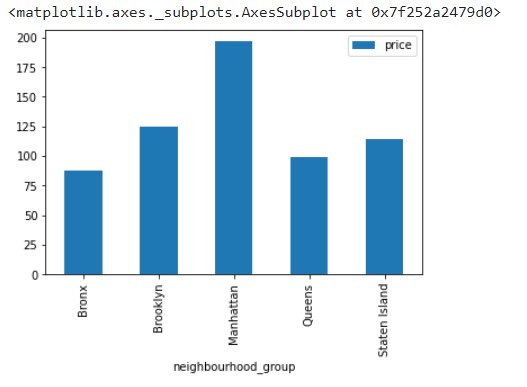
#47905  are listed yet some property are not listed or dont have host name

# here some host name and property name is same

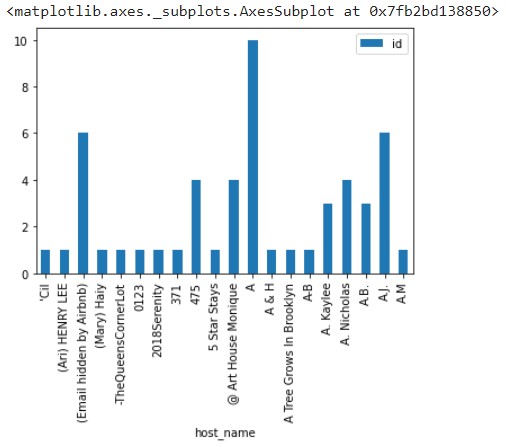
#same property  with different room\_types & varied price ranges

#Same hosts have different locations.

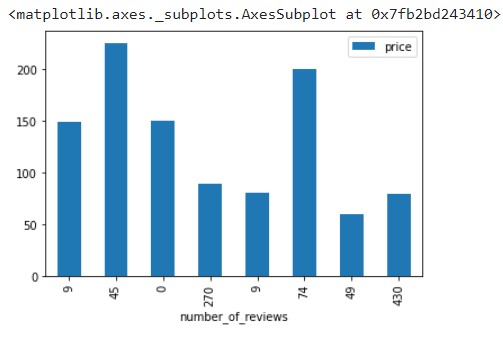
**Observing the dataset by plotting the different types of graph**



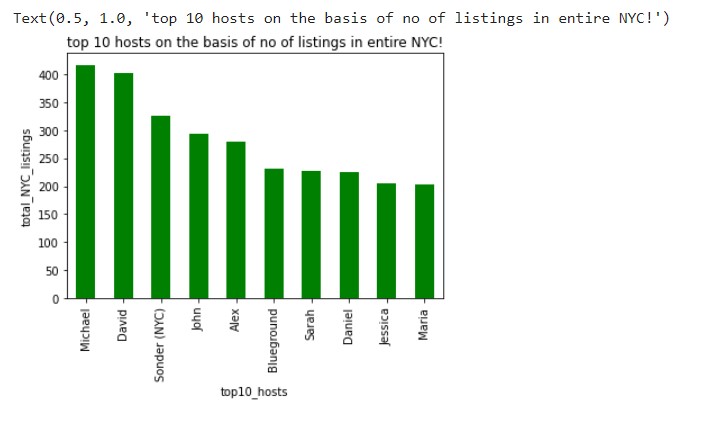
#manhattan have  the most costly airbnb



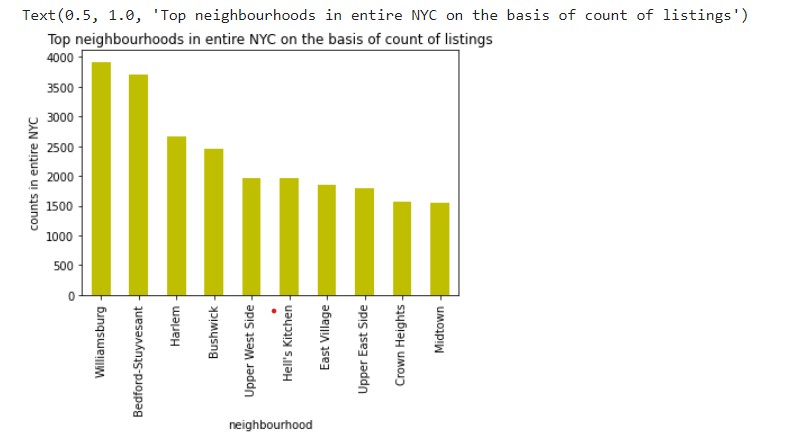
#Same host\_name and count of IDs



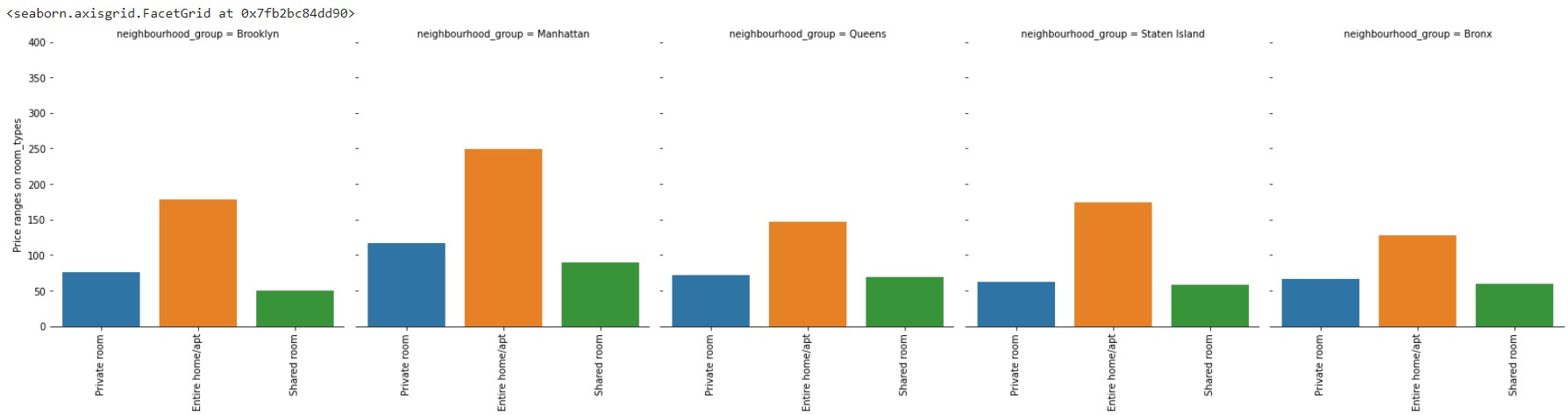
#relation between number\_of\_reviews and pricing



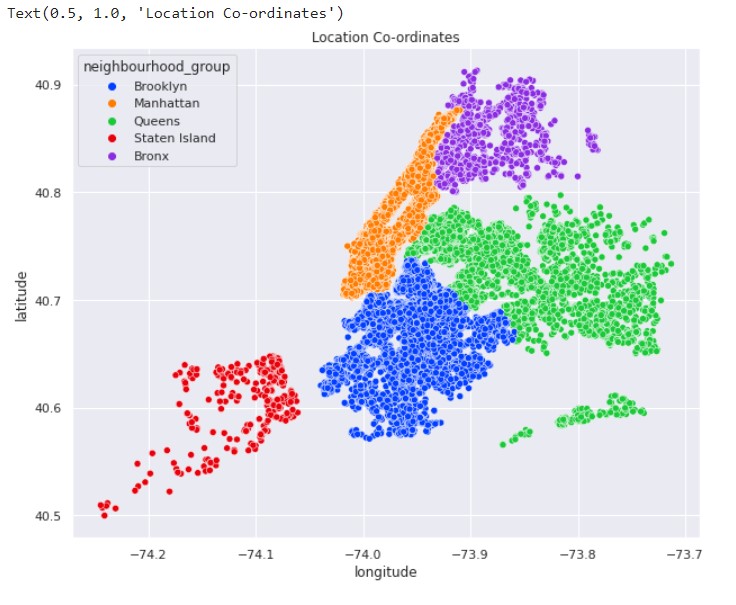
#top 10 hosts on the basis of no of listings in entire NYC!



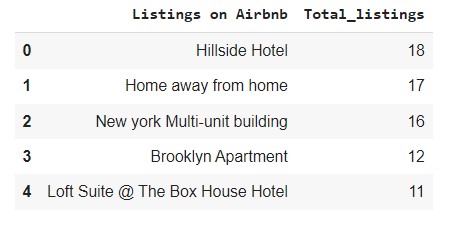
#checking top 10 neighbourhoods on the basis of no of listings in entire NYC!



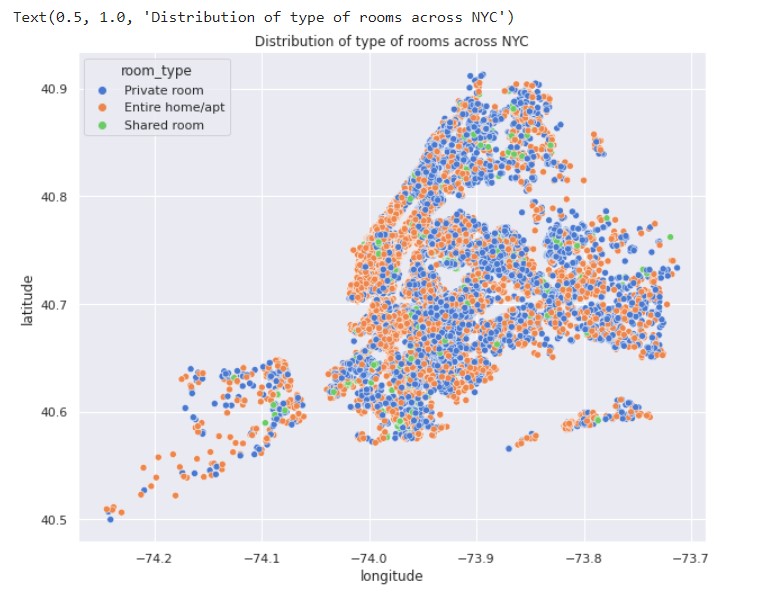
#checking types of rooms according to prices



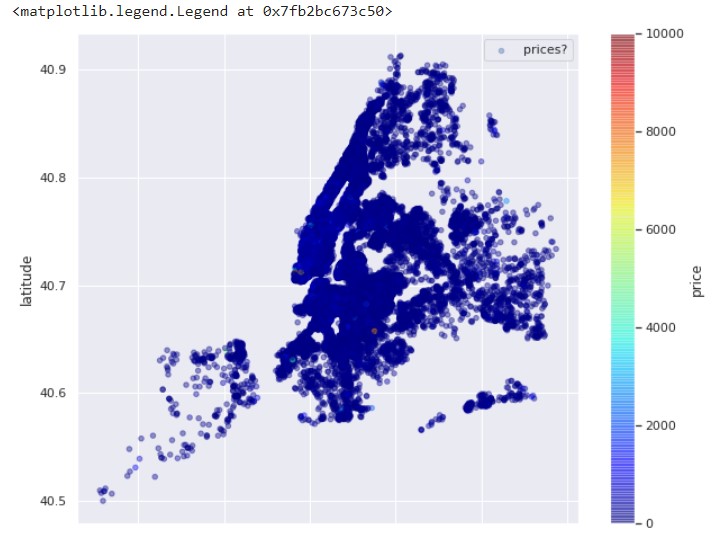
#trying to find where the coordinates belong from the latitude  and longitude



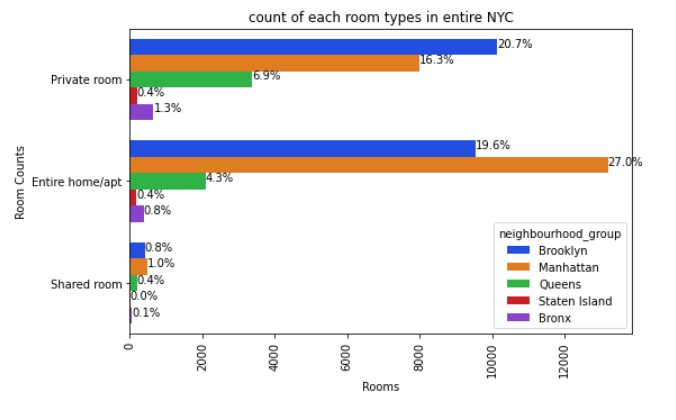
#hillside hotel is busiest



#Distribution of type of rooms across NYC



#Prices across area



#count of each room types in entire NYC

**Conclusion**

1. **Manhattan is the most costly Airbnb as More people want to live here than the city can hold**, **and that has driven up prices for the available apartments and houses.**
2. **Hillside hotels are the busiest by containing total no. of 18 listings.**
3. **Michael is the top host in the number of listings in entire NYC as it contains 417 hotels.**
4. **Manhattan has the highest number of listing in entire NYC 20000+ listings.**
5. **Williamsburg is the top neighborhood in the entire NYC on the basis of count of listings.**
6. **Manhattan is the most reviewed listing in NYC containing an average of 58.50 reviews per month.**
7. **Prices are different for the same types of room for different locations.**

**Important points to keep in mind while doing the analysis of the dataset.**