

# Air BNB Analysis Project

MUJTABA ALI & SAMRAT CHOWDHURY

Data science trainees,  
AlmaBetter, Bangalore

## Abstract:

Airbnb is an online marketplace connecting travelers with local hosts. On one side, the platform enables people to list their available space and earn extra income in the form of rent. On the other, Airbnb enables travelers to book unique homestays from local hosts, saving them money and giving them a chance to interact with locals. Catering to the on-demand travel industry, Airbnb is present in over 190 countries across the world.

## 1. Problem Statement:

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.

Explore and analyze the data to discover key understandings (not limited to these) such as

What can we learn about different hosts and areas?

What can we learn from predictions? (E.g. locations, prices, reviews, etc)

Which hosts are the busiest and why?

In this section, we will have the overview of the basic understanding of our dataset variables. What does particular features means and how it's distributed, what type of data is it. Airbnb dataset is having 16 columns in total. We can get this by basic inspection of our dataset. Some columns are not significant for our analysis which can also be kept off. Now let's look at some of the useful columns in our data set.

- **Id:** Unique listing ID
- **Name:** Name of the listing
- **Host\_id:** Unique host ID
- **Host\_name:** Name of the host
- **Neighbourhood\_group:** Location
- **Neighbourhood:** Area
- **Latitude:** Latitude coordinates

- **Longitude:** Longitude coordinates
- **Room type:** Listing space type
- **Price:** price in dollars
- **Minimum\_nights:** Amount of nights minimum
- **Number\_of\_reviews:** Number of reviews
- **Last\_review:** Latest review
- **Reviews\_per\_month:** Number of reviews per month
- **Calculated\_host\_listings\_count:** Amount of listing per host
- **Availability\_365:** Number of days when listing is available for booking

## 2. Introduction

**Airbnb** is an online marketplace connecting travelers with local hosts. Where you can list your space over their website.

In this Project We are analyzing the factors those are affecting the booking of host and many more. There are many columns like name of the list, longitude, latitude, room type, and price etc. Our goal to analyze all the columns by using python module like Pandas, matplotlib, and seaborn.

It helps us to book a good room, location and area. So that we can stay there safely.

## 3. Reasons for Analysis

The reasons for surge pricing are:

- Airbnb is the Vacation rental company. Where you can book room.

- 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,

## 4. Steps involved:

### ● Exploratory Data Analysis

After loading the dataset we performed this method by comparing our target variable that is Price with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

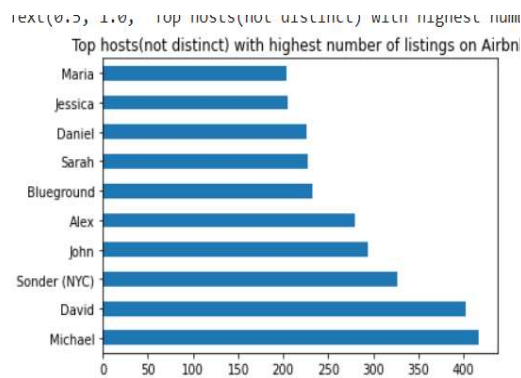
### ● Null values Treatment

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better

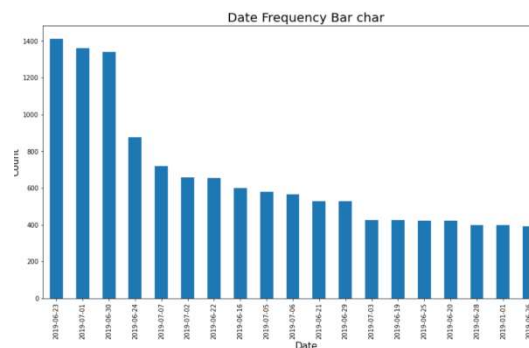
result. I applied mean to fill the missing values. Actually it depends on that kind of distribution we have in a particular column/feature. According to that I filled the missing values

## ● Visualization:

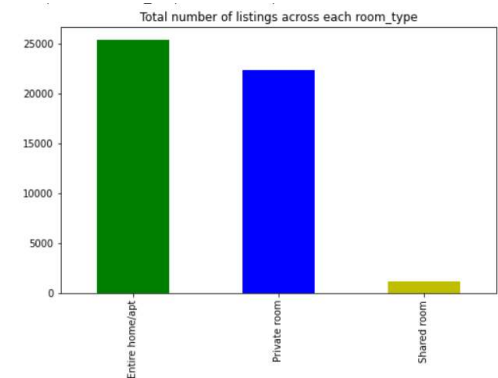
- ✓ Top host those has the highest listing on Airbnb.
- ✓ As you can see that Michel host has the highest no of listing on Airbnb.



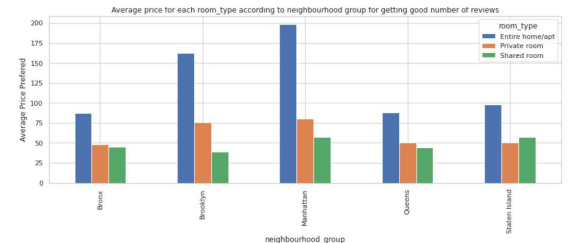
- ✓ This blow figure shows that in which date the highest no. of review has counted.



- ✓ Which kind of room has listed maximum no. of times and which one has minimum.
- ✓ As you can see that in below graph that Entire home/apartment has the maximum no. of listing. And shared room has minimum.



- ✓ This below graph shows that neighbourhood\_group vs. Average price Preferred.



## ● Encoding of categorical columns:

We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

## ● **Feature Selection**

In this project. We have not performed feature selection to select the features those are contributing more.

## ● **Standardization of features**

There is no need to apply Standardization on columns/features in This EDA project.

than the people who prefer to stay in entire home/apartment.

- ✓ Many rows are having values as 0 in price column, so this seems like an error which must be rectified by Airbnb.
- ✓ Island and Bronx neighbourhood\_group have very less numbers of listings.

## **References-**

1. <https://towardsdatascience.com/>
2. <https://www.geeksforgeeks.org/>

## 5. Conclusion:

- ✓ Sonder (NYC) host is having most number of listings on Airbnb in NYC.
- ✓ Williamsburg neighbourhood has most number of listings.
- ✓ Upper West Side, Astoria and Green point neighbourhoods have costliest listing in NYC.
- ✓ Bedford-Stuyvesant neighborhood has highest number of total reviews and Theater
- ✓ District neighborhood has highest number of reviews\_per\_month.
- ✓ Most of the listings on Airbnb in NYC are either Entire Home/Apartment or Private Room. The people who prefer to stay in entire home/apartment are likely going to stay longer, whereas people who prefer to stay in private\_room are likely to stay for a shorter period of time