**Exploratory Data Analysis of Airbnb NYC Bookings**

**Rubina Fathima**

**Data science trainee,**

**AlmaBetter, Hyderabad**

**Abstract:**

The Airbnb Data set of New York City is a collection of customers who make online hotel bookings.Neighbourhood in NYC rent their houses and apartments to make extra income.Tourists and people who are in need to stay in a home like environment opt for such bookings.

My analysis can help to understand what could be the reason for less reviews,high bookings.Price of the different room types and reviews of them.

***Keywords: Data analysis, availability,reviews,room type,price.***

**Problem Statement**

Data provided has customer information who make bookings of hotels.It has data spread in the form of rows and columns.

Rows are the various hosts and columns represent their features like calculated number of days stayed,reviews per month,name of the host and their span of stay.

**Dataset**

Airbnb dataset consists of

48895 rows,

16 columns.

Features of the dataset are explained as follows

Id: column which is unique to every observation

Name: gives name of the hotel of residency place for booking

Host\_id: Unique id given to the host Host\_name:Name of host

Neighbourhood\_group:There are places which are grouped together into neighborhood group

Neighborhood: Column has city names Latitude and Longitude: These describe the location of the hotel.

Room\_type:This again has three categorizations like single room, shared and entire home for booking

Price: Charges for every room type Minimum\_nights:Number of days stayed. Number\_of\_reviews:The reviews given by hosts after the stay.

Last\_review:This gives date when the review was recorded

Reviews\_per\_month:This gives the average of reviews for every host.

Calculated\_host\_listings\_count:column has count of hosts in the dataset.

Availability\_365:Hotels or room types availability is given here.

Every column has a datatype to describe the content inside fields.Datatype of the columns is explained in the figure below.

**Exploratory Data Analysis**

Airbnb Dataset consists of hotel

booking observations of hosts in

NYC.

I have done some exploratory data analysis using python to gain

insights of the dataset.

Null Values: Dataset has few columns whose fields are empty, filling such fields to perform analysis.

Room types of the hotel bookings

The dataset has a Private room ,Shared room and Entire home/apartment.

Observation-1:

Number of room types is given on the top of the bar graph to have a clear view of type and count of rooms in hotels.

Analysis shows shared rooms are

the least of the bookings.

Entire room bookings of the dataset are very close to the Private room bookings.Hence the Shared room types are rendered with less bookings.

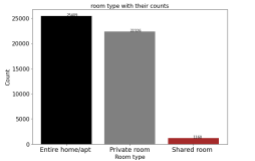


Figure:1

Reviews of each room type are very crucial to the online booking platforms.New users will start using hotels with high ratings from past customers.

Observation-2:

The hotels with high bookings sometimes face less average ratings.

This could be due to low

maintenance of the hotels.

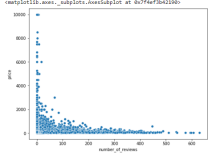
Figure:2

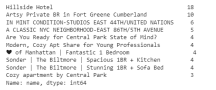
Top 5 Hotel Names who have high

bookings but below average

reviews.Among these

Hillside Hotel is the top and others are given in the figure.

Figure:3



Observation-3:

Scatterplot is used to show the distribution of reviews per month and price.

Here I have tried to find out the relation between price and reviews.

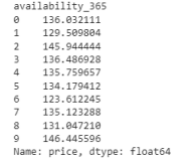
Figure shows that most of the reviews are awarded to the room prices which are low.

Average price of the room types is given above.

Observation-5:

Analysis over availability 365 column is Below observations states that

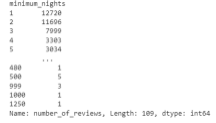
availability “0” has an average price of 136 As the availability increases the price increases of the room type.

Observations-5:

Here I have explored the columns minimum nights and count of the nights the host has stayed.

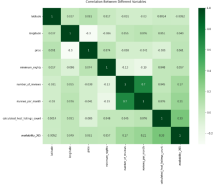
This observation gives a glance that single day bookings are at high

compared to more days bookings.

Observation-6:

Here I have tried to show correlation among features.

Reviews column and longitude have high correlation compared to other features.

Figure-4

Observation-7:

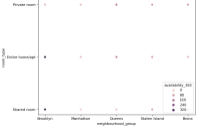
Here I have shown the relationship between availability and price column. High availability is shown in the neighborhood group brooklyn.

It conveys that the reason behind

Brooklyn's availability could be that the place is less among tourist attractions. Low number of visitors to brooklyn could also be the reason.

Other neighborhood groups like

Manhattan, Queensland, States Island and Bronx.

Figure-5

**Conclusion:**

After all the observations I have

collected points to explain

**●** Airbnb is an online booking platform

where the locals earn through giving

their houses on rent.

● Reviews from customers give a lot of

information on the room types and their

price.

● Availability of the room type is more

important because it has a direct impact

on price.

● Reviews from single bookings and

multiple bookings are different.

● Some neighborhood groups like

Brooklyn show high availability of

room types like entire homes and

shared rooms

**References:**

1. Analytics Vidhya

2. Medium platform