**Airbnb Booking Analysis**

**Mrutyunjaya Sahoo**

**Data Science Enthusiast**

**AlmaBetter, Bangalore**

**Abstract:**

Airbnb Bookings Analysis is based on understandings about location, property listing, property host, areas and their hostings.

You will gain information about factors affectingbooking like price, neighborhood area, etc.

The conclusions from this EDA can benefit who want to do business or who want to market their product. Important inferences have been provided throughout analysis in the collab notebook. This EDA will also help common people or customer to make choice decision which room to take according to their price, availability etc.

***Keywords:* Airbnb, Data Cleaning, Exploratory Data Analysis**

1. **Problem Statement**

Airbnb generates 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.

Explore and analyze the data to discover key understandings:

* Understanding different host has different reviews in their respective location.
* How different room type has different prices
* What is the price of different house
* Which host has different types of room and hosted at least some nights
* How many rooms available has the host with their reviews
* Which room has average prices that gives us how to book good rooms with their prices
* Understanding this airbnb data with heatmap

1. **Introduction**

Airbnb is an open online platform where people list their own housing for rent. Since 2008, it has grown in popularity and specially for those community which frequently use to travel. It is becoming a strong competitor to the hotel industry. It has millions of listing, which generate lots of data. I’m analyzing these data for making business decision, for looking best room type etc.

I will explore and visualize the dataset from Airbnb in New York using basic exploratory data analysis techniques. I will find out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors.

My object is to explore the data and extract useful information from the data and find out different relations between the columns.

1. **Airbnb Booking Dataset Insight**

This dataset has around 49,000 observations in it with 16 columns and it is a mix of categorical and numeric values It contains different hosts, the neighborhood group the properties are located in and the type of property customers most wish for. Exploring them will definitely help in understanding of the booking trends.

**Column Information**

* id : Unique Id.
* name :Name of listing.
* host\_id : unique id for each listed host.
* host\_name: Name of the host.
* neighbourhood\_group: Location
* neighborhood : Area
* latitude : Latitude coordinates
* longitude : Longitude coordinates
* room\_type: Type of listing
* price : Price of listing
* minimum\_nights: minimum nights to be paid for rooms
* number\_of\_reviews: No. of reviews
* last\_review: Content of the last review
* reviews\_per\_month: Number of checks per month
* calculated\_host\_listings\_count: Total count of hosting list
* availability\_365 :Availability around the year

1. **Steps involved**

* **Data Overview**

As a first step I overviewed the data, where I specially looked on understanding what each column means. So that I can be clear from what perspective I have to analyze our data. After understanding different column, I removed some columns because it has huge null values. Then we arrange the columns and did some basic visualization to see is there any correlation among columns.

* **Cleaning the Dataset**

Now I looked to clean the data. So first I found some null values and I replaced these null value according to their data type. After dealing with null values I stepped on to columns . So I removed some column. Then I replaced few null data which don’t make sense with other values. Finally, our data is ready for EDA.

* **Exploratory Data Analysis**

1. **host vs location analysis:**

The first focus is to see where there is different host has at different location that identifies that yes that person has that home ,also that helps people to choose as of their convenience for their booking.

2. **Price Analysis:** my focus is ‘Price’. Here first I looked in to average price of different room type across New York. From this I got to know that costly room type is ‘Entire home’. Then I looked at average price of room type according to different neighborhood group. Then I did few more on analysis on finding cheapest neighborhood and the cheapest listing throughout New York.



**3. Listing Analysis:** Here we understand different listing. In this I took a look at listing according to their neighborhood group. From the result we create inference that is someone want to do advertisement or marketing he should focus on Manhattan and Brooklyn. Then I saw data and looked different listing according to neighborhood group. We get to know that in Manhattan, entire room type is highly listed.

**4. Availability analysis:** Here you understand on availability of different room according to their neighborhood group and then the average availability of different room type. you get to know that private room has highest availability and entire home has least availability. The inference which come out from this result if host is having entire room then he will be making good money. But from customer point of view private room is the best as half of the year it is available.

**5. Conclusion**

I understand the data and have done the bookings analysis using this dataset.

* Understanding different host has different reviews in their respective location.
* How different room type has different prices
* What is the price of different house
* Which host has different types of room and hosted at least some nights
* How many rooms available has the host with their reviews
* Which room has average prices that gives us how to book good rooms with their prices
* Understanding this airbnb data with heatmap

**References**

1. Almabetter class notes
2. Github repository
3. Stack overflow