**Airbnb Booking Analysis**

**Anjali Tidke, Shubham Dukare**

**Data Science Enthusiast**

**AlmaBetter, Bangalore**

**Abstract:**

The project is about the Analysis of Airbnb Bookings Data.

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.

This analysis helps business people to make business decisions. Airbnb analysis will describes us about different hosts, different areas, price listings based on areas, customer reviews, What can we learn from predictions 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:

* What can we learn about different hosts and areas?
* What can we learn from predictions (ex: locations, prices? reviews)
* Which hosts are the busiest and why??
* Is there any noticeable difference of traffic among different areas and what could be the reason for it??

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 listings, which generate lots of data. We are analyzing these data for making business decision, for looking best room type etc.

We will explore and visualize the dataset from Airbnb in New York using basic exploratory data analysis techniques. We 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.

The goal here is to explore the data and find useful insights 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 listing id.
* name: Name of the property.
* 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: Listing space types
* price: Price in dollars
* minimum\_nights: minimum nights required to stay
* number\_of\_reviews: No. of reviews written for the listing
* last\_review: Last reviewed date for the listing
* reviews\_per\_month: Total review per month for the listing
* calculated\_host\_listings\_count: Total no of listing against the host id
* availability\_365: Number of days when listing is available for booking.

1. **Steps involved**

* **Data Overview**

As a first step we take the overview of data, where we specially made our focus on understanding what each column means. So that we can be clear from what perspective we have to analyze our data. After understanding different column, we marked few important columns. These columns are neighborhood group, room type, price, minimum nights, reviews per month. Then we did some basic visualization to see is there any correlation among columns.

* **Cleaning the Dataset**

Now we started cleaning our data. So we first identified the null values and we replaced these null value according to their data type. After dealing with null values we moved on to those columns which we don’t need. Finally, our data is ready for EDA.

* **Exploratory Data Analysis**

1. What we can learn about different hosts and areas?

As we can see most number of listings are from Manhattan created Sonder (NYC), Blueground, kara, Kazuya.

1. What can we learn from predictions? (ex: locations, prices, reviews, etc.)

we can say that most people prefer to stay in place where price is less.

1. Which hosts are the busiest and why?

We can see that Dona host is busiest.

1. . Is there any noticeable difference of traffic among different areas and what could be the reason for it?

People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.

1. **Question & Answer**

Throughout the analysis, we tried to answer questions that help us understand the factors determining the data trends.

Q 1. In the dataset, price to be paid per night given, so how to find out total minimum expenditure for a stay?

Answer: Yes, we can estimate total minimum expenditure for a stay by multiplying price and number of minimum nights.

Q 2. Suddenly I’m planning for a trip but I’m not sure about where I can get room easily for a stay. can we analyze the ease of availability of room?

Answer: Yes, we can analyze the ease of availability of room in our data set. it can be done by grouping the neighborhood group, room type and then find mean of the availability365 of rooms. it gives the data of mean of availability of room type according to the neighborhood group.

Q 3. How to know about the prices of various room?

Answer: It can be done by finding out the mean price of various room types. we did the same in could reach to following conclusions;

1. Mean price of entire room is more than the mean price of private room.

2. Mean price of private room is more than the mean price of shared room.

Q 4. How to find out which neighborhood is costlier for a stay?

Answer. By finding out the mean price for each neighborhood group, anyone can compare neighborhood groups. In the given data set we found that Manhattan is costlier and Brooklyn is cheapest neighborhood group as per the mean price of various room type.

Q 5. How to know which room type is mostly available?

Answer: It can be done by grouping room type and find the mean of the availability365 according to the room type. In the data set we could find it that;

1. Private room has highest mean of availability.

2. Entire home has least mean of availability.

Q 6. How to find total listings by each neighborhood group.

Answer. It can be analyzed by grouping the neighborhood group and then count their listings. in our data we found that;

1. Manhattan & Brooklyn are having high no. of listing.

2. Staten Island and Bronx have low no. of listing.

**6. Conclusion**

We were able to answer some really important questions about the bookings analysis using this dataset.

1. The people who prefer to stay in Entire home or Apartment they are going to stay bit longer in that particular Neighbourhood only.

2. The people who prefer to stay in Private room they won't stay longer as compared to Home or Apartment.

3. Most people prefer to pay less price.

4. If there are more number of Reviews for particular Neighbourhood group that means that place is a tourist place.

5. If people are not staying more then one night means they are travellers.

**References**

1. Towards Data Science
2. Stack overflow
3. Medium
4. Git hub