**Airbnb Booking Analysis**

**Ritik Vaidande**

**Data Science Practitioner**

**AlmaBetter, Bangalore**

**Abstract:**

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

We will gain information about factors affecting booking 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:

* 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 listing, 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**

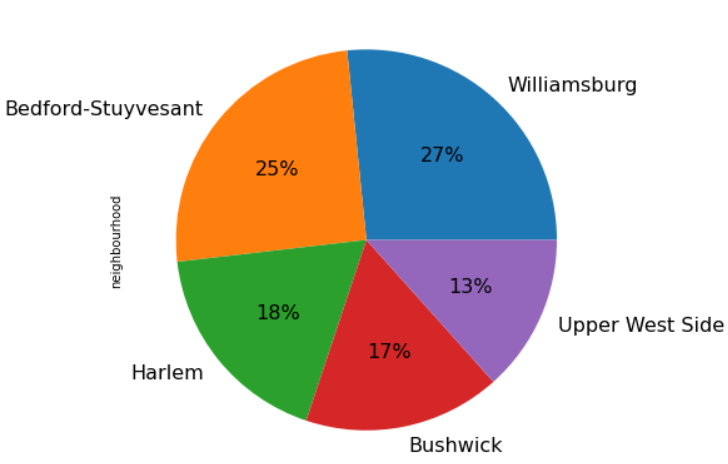
As a first step we take the overview of data, where we specially made our focus on understanding what each column means. So that we can be clear from what perspective we have to analyze our data. After understanding different column, we marked few important columns. These columns are neighborhood group, 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. So we removed last review column. Then we replaced few data which don’t make sense with other values. Here we replaced zero price of property with mean price according to their room type and neighborhood type. Finally, our data is ready for EDA..

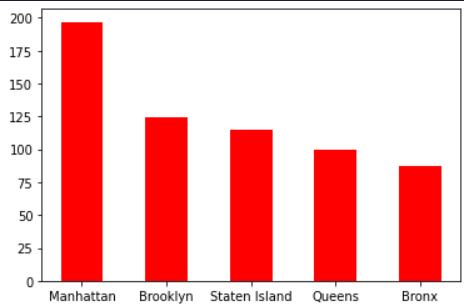
* **Exploratory Data Analysis**

1. **Area Analysis:** The first focus we put on is ‘Area’. Here first we looked in to most listings in the neighbourhood. From this we get to know that Sonder (NYC) neighbourhood group has the most number of listings. Then we looked at neighborhood group which has the most number of listings, which came out to be Manhattan. From this analysis we made inference that if Manhattan and Brooklyn are both combined, almost 85% listings are done every day just in this regions. Where Staten Island and Bronx are very few listed areas thus this specifies that these are unpopular and not densely populated areas.



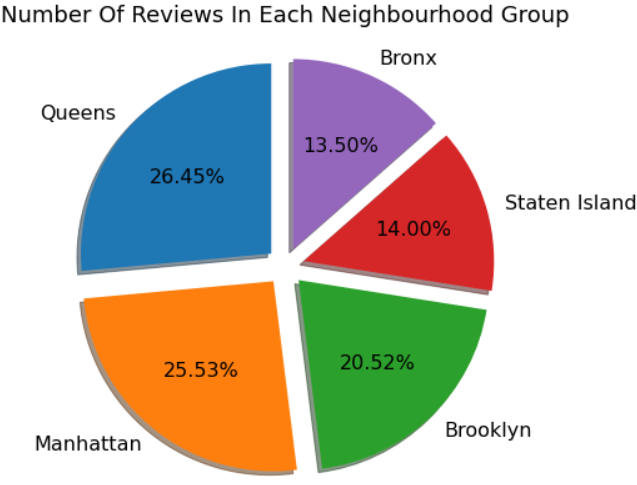
Top 5 Transaction Areas

**2. Price Analysis:** Here we first looked into the average price of listings. From this analysis we made inference that Manhattan is the costliest neighbourhood group although it was on top on the number of listings and Staten Island which was having the least number of listings is not the cheapest neighbourhood group among all. Manhattan airbnb's has the highest average price. Bronx provides most number of budget friendly rooms. Airbnb’s with price range of 150 to 200 dollars are above 4k+ listings. Also if a person is low on budget, he may prefer to book an Airbnb in Bronx.



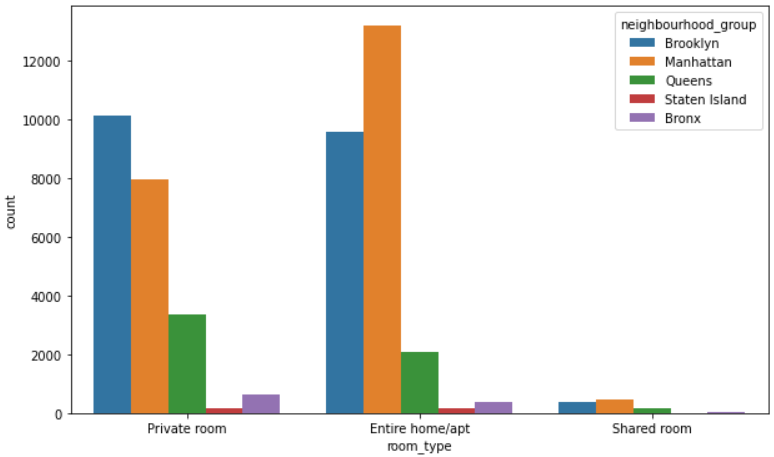
Costliest Neighbourhoods

**3. Review analysis:** We get to know that Airbnb’s having 4 reviews per month are around 130 and Airbnb’s having 1 review per month are around 178. The inference which come out from this result that if people are rating the hosts with poor reviews more than good reviews. This shows that people will not prefer to come again to the rooms they were not comfortable.



Percentage of Reviews as per neighbourhood group

1. **Preferability analysis:** To confirm the above inference we analysed the type of rooms people prefer to stay. We found out that ‘Entire Home’ is room type people prefers the most almost in every neighbourhood. The inference which come out from this result if host is having entire room then he will be making good money. We were astonished that irrespective of any neighborhood group, entire home is making way ahead revenue then other room types. In Brooklyn, Private rooms are more preferred instead of Entire Home.



1. **Question & Answer**

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

Q 1. 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 2. 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 3. 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.\

Q 4. How to know which host are the busiest?

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

1. Financial District from Manhattan neighbourhood is the busiest host amongst all hosts.

**6. Conclusion**

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

1. Entire home/apt is highly expensive.
2. Manhattan living cost is highest, Bronx living cost is lowest.
3. Cheapest neighborhood is Bulls head.
4. Cheapest listing is Bronx apart.
5. Manhattan have highest no. of listing.
6. In Manhattan entire home is mostly preferred but in Brooklyn ratio between entire home and private room is 50:50.
7. Private room has highest availability; Entire home has least availability.
8. Sonder have maximum property in New York.

**References**

1. Towards Data Science
2. StackOverflow
3. Medium
4. GitHub