Airbnb Booking Analysis

Nileshkumar , Abhishek , Animesh & Prajwal
Data Science Enthusiast, Cohort - Antwerp
AlmaBetter, Bangalore

Abstract:

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

We will gain information about factors affecting booking like price, neighbourhood area etc.

The conclusions from this EDA can benefit those 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 customers to make a decision which room to take according to their price, availability etc.

Keywords: Airbnb, Data Cleaning, Exploratory Data Analysis

1. Problem Statement

The main objective of this exploratory data analysis (EDA) project is to understand the dynamics of Airbnb rentals in New York City (NYC) by analysing various factors such as pricing, availability, neighbourhood groups, room types, and customer feedback. By examining the average price, number of listings, availability, and customer reviews, we aim to gain insights into the pricing patterns, popularity of neighbourhoods, customer satisfaction, and market opportunities. This analysis will provide valuable information for potential investors or hosts to make informed decisions and understand the key factors influencing the success and profitability of Airbnb rentals in NYC.

- 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??

2. Introduction

Airbnb is an open online platform where people list their own housing for rent. Since 2008, it has grown in popularity and especially for those communities which frequently travel. It is becoming a strong competitor to the hotel industry. It has millions of listings, which generate lots of data. We are analysing this data for making business decisions, for looking for the best room type etc.

We will explore and visualise 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.

3. Airbnb Booking Dataset Insight

This dataset has around 48,895 observations in it with 16 columns and it is a mix of categorical and numeric values It contains different hosts, the neighbourhood 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
- neighbourhood : 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.

4. 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 analyse our data. After understanding different columns, we marked a few important columns. These columns are neighbourhood group, room type, price, minimum nights, reviews per month. Then we did some basic visualisation to see if there is 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 values according to their data type. After dealing with null values we moved on to those columns which we don't need. So we removed the last review column. Then we replaced a 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 neighbourhood type. Finally, our data is ready for EDA..

● Exploratory Data Analysis

1. **Price Analysis:** The first focus we put on is 'Price'. Here first we looked into the average price of different room types across New York. From this we get to know that the costly room type is 'Entire home'. Then we looked at the average price of the room type according to different neighbourhood groups. From this analysis we made inference that if a salaried employee wants to increase his savings then he will prefer to work in the Bronx. Then we did a few more analyses on finding the cheapest neighbourhood and the cheapest listing throughout New York.

2. Listing Analysis: Here we focused on different listings. In this we take a look at listing according to their neighbourhood group. From the result we made inference that if someone wants to do advertising or marketing he should focus on Manhattan and Brooklyn. Then we deep dive into the data and looked at different listings according to the neighbourhood group. We get to know that in Manhattan, the entire room type is highly listed.

3. Availability analysis: Here we focused on availability of different rooms according to their neighbourhood group and then the average availability of different room types. We get to know that the private room has the highest availability and the entire home has least availability. The inference which comes out from this result is that if the host has an entire room then he will be making good money. But from a customer point of view, a private room is the best as half of the year it is available.

5. **Profitability analysis:** To confirm the above inference we created some new columns which talks about revenue generated by different hosts and their property. First we did an analysis on revenue generation of different room types by their neighbourhood group. Seeing the result, we were astonished that irrespective of any neighbourhood group, the entire home is making way ahead in revenue then other room types. Then we deep dived and looked into which host is having highest no. of property, which neighbourhood group has highest review etc.

6. 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 easily get room for a stay. Can we analyse the ease of availability of rooms?
Answer: Yes, we can analyse the ease of availability of rooms in our data set. It can be done by grouping the neighbourhood group, room type and then finding the mean of the availability365 of rooms. it gives the

data of mean availability of room type according to the neighbourhood group.

Q 3. How to know about the prices of various rooms?
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 an entire room is more than the mean price of a private room.
2. Mean price of a private room is more than the mean price of a shared room.

Q 4. How to find out which neighbourhood is costlier for a stay?
Answer. By finding out the mean price for each neighbourhood group, anyone can compare neighbourhood groups. In the given data set we found that Manhattan is costlier and Brooklyn is the cheapest neighbourhood group as per the mean price of various room types.

Q 5. How to know which room type is mostly available?
Answer: It can be done by grouping room type and finding the mean of the availability365 according to the room type. In the data set we could find it that;

1. Private room has the highest mean of availability.
2. Entire home has the least means of availability.

Q 6. How to find total listings by each neighbourhood group.
Answer. It can be analysed by grouping the neighbourhood group and then count their listings. in our data we found that;

1. Manhattan & Brooklyn are having high no. of listings.
2. Staten island and Bronx have low no. of listings.

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 listing is the Bronx.
4. Manhattan has the highest no. of listings.
5. In Manhattan the entire home is mostly preferred but in Brooklyn the ratio between the entire home and private room is 50:50.
6. Revenue generated by Entire home is highest irrespective of neighbourhood group.

References

1. Towards Data Science
2. Stack overflow
3. Medium
4. Github