

Airbnb Booking Analysis

RAUMAN SHARIFF, Data Science Enthusiast, Cohort

- LIMA

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

This project's primary goal is to better understand the dynamics of Airbnb rentals in New York City (NYC) by examining a range of variables, including cost, availability, neighborhood associations, types of rooms, and reviews from previous guests. We hope to learn more about pricing trends, neighborhood popularity, customer satisfaction, and market opportunities by looking at the average price, number of listings, availability, and customer reviews. Potential investors and hosts will find this analysis to be very helpful in making decisions and in understanding the main drivers of the popularity and profitability of

Airbnb rentals in New York City.

2. 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.

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

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

3. **Availability analysis:** Here we focused on availability of different room according to their neighborhood group and then the average availability of different room type. We 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.

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

5. 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 Bronx apart.
4. Manhattan have highest no. of listing.
5. In Manhattan entire home is mostly preferred but in Brooklyn ratio between entire home and private room is 50:50.
6. Private room has highest availability; Entire home has least availability.
7. Revenue generated by Entire home is highest irrespective of neighborhood group.