AirBnB Exploratory Data Analysis

1. Aditya Bhople
2. Rupali Dawkore
3. Rachana Gupta
4. Piyush Sujeeth
5. Subhi Yadav

Abstract

In this project, we have analyze the data descriptively and statistically to determine how the variables are correlated to generate hypotheses useful for future decision-making. It's imperative to analyze the data carefully in order to obtain meaningful insights

that can assist in making better business decisions and understanding customer and host behavior. The dataset comprises millions of listings on the Airbnb platform. Airbnb offers short-term homestays and experiences on an online marketplace.

To work on data, we will be using different tools that are very common for performing simple and complex analyses, like *classifications of variables*, *histograms*, *textual mining*, and *measures of central tendency*.

Keywords: Exploratory data analysis, *histograms*, *textual mining*, and *measures of central tendency*

Problem Statement

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world.

Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate 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.

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

Explore and analyze the data to discover key understandings of the following important question to enhance bussiness of Airbnb

Introduction:

Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in that locale. It currently covers more than 81,000 cities and 220 countries worldwide.

For this project we are analyzing Airbnb's New York City (NYC) data of 2008. Right from choosing the quarters to offering support around the clock until we part, it brings all sorts of convenient measures right into our hands. Airbnb is one such wonderful platform that serves as an online marketplace to help people find accommodation during their trip as per their needs. Airbnb serves over 220 countries and regions across the world today.

Data analysis on thousands of listings provided through Airbnb is a crucial factor for the company. Our main objective is to find out the key metrics that influence the listing of properties on the platform. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis (EDA) techniques.

We have found out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, an other related factors. We have analysed this dataset from different angles and have come up with interesting insights. This can help in

making strategic data- driven decisions by the marketing team, finance team and technical team of Airbnb.

Analysis of Data Quality

We had to perform a few imputations and transformations on our dataset for us to create the desired visualizations. There were

no major inconsistencies or mismatches in the data, but most of the columns/features we were interested in did not contain data in the required format
and hence were manipulated in a way that their meanings are retained.

**Different types of room distribution**
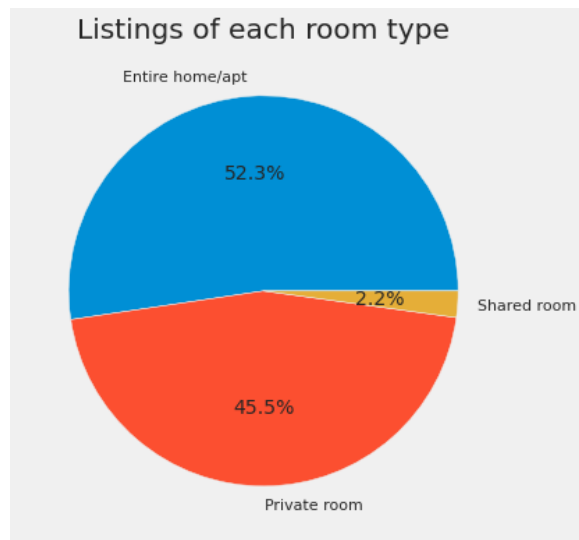
There is a very clear percentage division of the three different room types across the region, with 'Entire home/apt' accounting for 52.3% of listings, and shared rooms representing just 2.2%.

Top 10 neighbourhoods with maximum listings

Since we have more than two hundred neighbours listed out, so we will be taking a glimpse of only the top ten neighbours who has more listings.

	neighbourhood	total_prop
0	Williamsburg	3163
1	Bedford-Stuyvesant	3141
2	Harlem	2204
3	Bushwick	1942
4	Hell's Kitchen	1528
5	East Village	1489
6	Upper West Side	1482
7	Upper East Side	1405
8	Crown Heights	1265
9	Midtown	986
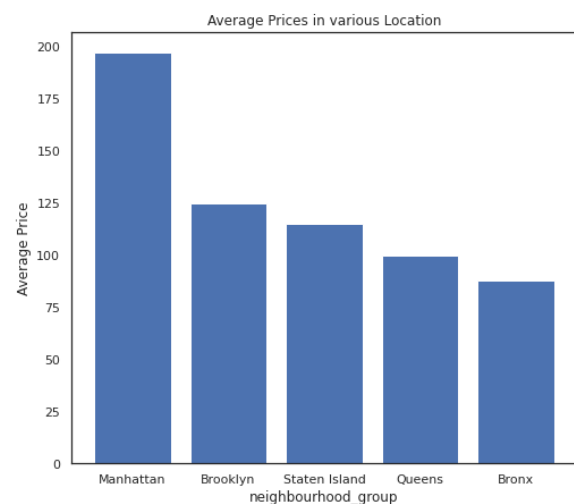
Conclusion

And At last after loading the dataset and performing Exploratory data analysis and analyzing the datasets properly we have come to some conclusions and insights which we have drawn from the given dataset.

This insights can be implemented and used at various departments of Airbnb such as marketing and business development to enhance their business and achieve exponential growth in their business.

The data of highly searched words and preferred location can be targeted by the marketing team to increase more conversions.

- Manhattan is the most focused place in New York for hosts to do their business
- Customers pay the highest amount in Brooklyn,Queens and Manhattan that is 10,000 and lowest amount is 10$.
- For the three types of room type (i.e. Entire home, Shared room, & Private room) average price of entire home is around $157, for shared room is around 60, and for private room is around 75.
- 'Entire home/apt' room type has the highest number of listing of 52% and 'Shared Room' is the least listed room type at only 2.4% in total.
- People stay for longer duration of time in Private rooms in Brooklyn and Manhattan.
- Words such as 'bedroom', 'cozy', 'private', 'apartment' and 'spacious' are used more frequently than words such as 'park', 'near', 'village' and 'heart'.
- Count of listing by top 10 hosts is almost 2.5%(1270 listings) of the whole dataset.
- Top three host base on their turnover are Sonder(nyc),Red awning, Henry and best host is Sonder(nyc)
- More customer preferred Manhattan location for night stay then Brooklyn
- 63.2% customer spend night in Entire home and 1.6% spend night in Shared room