Airbnb booking analysis

Pratik Chaudhari
Data science Trainee
AlmaBetter

Abstract:

Airbnb is an open online platform where people list their housing for rent. Since 2008, it has grown in popularity, especially for communities that frequently use travel. It is becoming a strong competitor in the hotel industry. It has millions of listings, which generate lots of data. We are analyzing these data for making business decisions, for looking the best room type, 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 the analysis in the collab notebook. This EDA will also help common people or customers to make a choice decision which room to take according to their price, availability, etc.

Keywords: Airbnb, EDA (exploratory data analysis)

- 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 in traffic among different areas and what could be the reason for it?

- Which properties are most preferable?

- how many properties are available for more than 200 days?
- which neighborhood group has the highest number of Airbnb?

1. Problem Statements:

There are a few problems that we discussed in this EDA .problems are mentioned below:

2. Introduction:

Airbnb 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 the website and mobile app. Airbnb does not own any of the listed properties instead, it profits by receiving a commission from each booking. Airbnb is a shortened version of its original name,AirBedandBreakfast.com

The goal here is to explore the data, find useful insights, and find out the different relations between the columns.
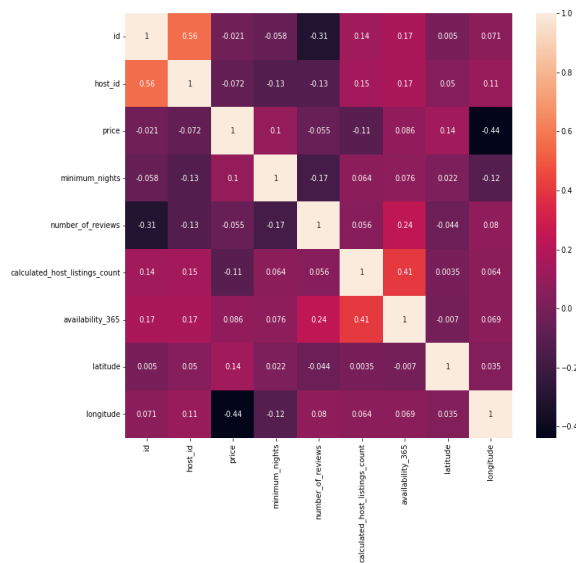
iii.   Host_id: unique id for each listed host.
iv.   Host_name: Name of the host.
v.   Neighbourhood_group: Location
vi.   Neighborhood: Area
vii.   Latitude: Latitude coordinates
viii.   Longitude: Longitude coordinates
ix.   Room_type: Listing space types
x.   Price: Price in dollars
xi.   Minimum_nights: minimum nights people wants to stay on property
xii.   Number_of_reviews: No. of reviews written for the listing
xiii.   Last_review: Last reviewed date for the listing
xiv.   Reviews_per_month: Total review per month for the listing
xv.   Calculated_host_listings_count: Total no of listing against the host id
xvi.   Availability_365: Number of days when the listing is available for booking.

4.Steps involved:
  - **Data study**

In the beginning, we need to understand the dataset to make any analysis of the dataset. So our first step has to be understanding each column what it depicts and what we can get through that information.
We came to know that there are a few columns that we don't need in any of our analyses, so we can keep only those columns in the dataset that we need for our analysis.

3.Data set :

observations in it with 16 columns and it is a This dataset has around 49000 mixes of categorical and numerical values. now we can discuss the column information those are given below:

i.   Id: Unique listing id.
ii.   Name: Name of the property.

After that, we also came to know that there are few null values in our dataset also.
In the price column, there are a few 0 values of properties and that can not be possible so we have to take some action on that too.

• Data cleaning

So, first of all, we only keep those columns which we need for our analysis those columns are 'id', 'name', 'host_id', 'host_name', 'neighbourhood_grop, 'neighborhood', 'room_type', 'price', 'minimum_nights', 'number_of_reviews', 'calculated_host_listing_counts' and 'availability_365'.
Now in the second step, we have a few null values in columns 'name' and 'host_name', so we fill those null places with 'not known and 'not present.
We also replace 0 values in the price column with the median of the rest of the non-zero values.

• EDA (Exploratory Data Analysis)

After we clean our dataset. we are ready to analyze our data set.
We did various data analyses based on prices of properties, availabilities, Neighborhood groups etc. most of our EDA can be done with the help of a group.
We also develop data visualization for different questions for better understanding. through our analysis, we found different trends and relationships among various columns. The goal here

is to explore the data, find useful insights, and find out the different relations between the columns.

5. Conclusions:

So after EDA we came with a few conclusions on our dataset and that can be very useful for both sides .so all the conclusions which we get are given below:

- Most of the listings are from Manhattan and Brooklyn.
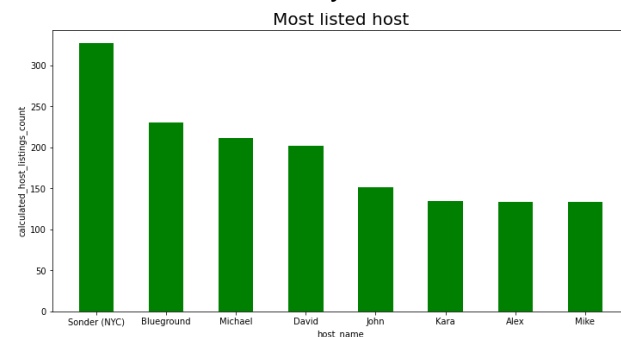
- Sonder(NYC) has the most number of listings and after her Blueground hold second place in listings.
- Entire rooms and privaterooms have greater mean prices than other room types because demand for entire rooms and private rooms are on higher sides.
- Airbnb spread across all the neighbourhood_group.from our analysis we found out that most numbers of the private rooms are located in Brooklyn and most numbers of the entire homes are present in

Manhattan. The Bronx has the least number of all kinds of room_type.

- Busiest hosts are Sonder(NYC), Michael, and Blueground, and the reason cause they are providing the entire apt or private rooms which are preferred by most of the visitors.

- we can see Private Bedroom in Manhattan has almost no reviews of 666 so here we can say that Private Bedroom in Manhattan is the most preferable property of the whole data set.
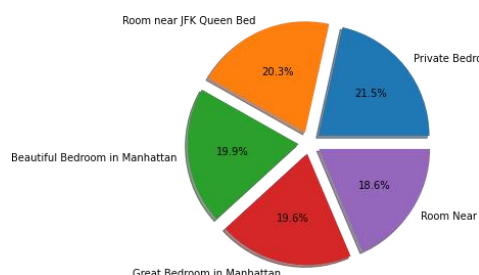
- Only 27.62% of properties are available for more than 200 days which means we can not book throughout the years few properties.

- Manhattan has the most number of properties and after Manhattan there is Brooklyn, and the Bronx and Staten Island have the least number of Airbnb.