AIRBNB NYC ANALYSIS

VIVEK SONI,ASHISH PANDEY,RAVI
Data Science Trainees
AlmaBetter, Bangalore

Abstract

Airbnb, as in “Air Bed and Breakfast,” is a service that lets property owners rent out their spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

Problem Statement

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.Explore and analyze the data to discover key understandings (not limited to these) such as : 1.What can we learn about different hosts and areas?2. What can we learn from predictions? (ex: locations, prices, reviews, etc)3. Which hosts are the busiest and why?4. Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Goal

The aim of our project Airbnb is a leading online platform that offers its clients an ideal marketing place where they can shop and order for a wide range of hospitality services at the comfort of their home.

Airbnb has become a household name not only in America but also across the globe due to the strength of its mission and vision statements. The emphasis it places on excellence in everything it does has particularly placed it atop of Airbnb competitors other similar platforms.
The mission statement of a company presents the steps the management seeks to abide by while delivering its services to its target audience. In the case of Airbnb, the focus of its mission statement is on the experience it creates for its customers

In our case, missing data that is observed does not need too much special treatment. Looking into the nature of our dataset we can state further things: columns "name" and "host_name" are irrelevant and insignificant to our data analysis, columns "last_review" and "review_per_month" need very simple handling. To elaborate, "last_review" is date; if there were no reviews for the listing - date simply will not exist. In our case, this column is irrelevant and insignificant therefore appending those values is not needed. For "review_per_month" column we can simply append it with 0.0 for missing values; we can see that in "number_of_review" that column will have a 0, therefore following this logic with 0 total reviews there will be 0.0 rate of reviews per month. Therefore, let's proceed with removing columns that are not important and handling of missing data.

Introduction:

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

Id	Price
Name	Minimum_nights
Host_id	Number_of reviews
Host_name	Last_review
Neighbourhood_group	Review_per_month
Neighbourhood	Calculated_host_list_count
Latitude	Availbilty_365
Longitude	
Room_type	

Steps Involved:

- ## Raw Tables:

  It has information such as app name, category, rating, and more. And the other is a list of reviews for each app with the sentiment if that particular content of the review was positive, neutral, or negative. Unfortunately, we could not directly use these two files as they are not joined.

- ## Data Preparation and Cleaning:

  Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data, and the combining of data sets to enrich data. Data cleansing or data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate, or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

- ## Null Value Treatments:

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

- Removing Duplicate Entries:

  We don't want to count certain apps more than once when we analyze data, so we need to remove the duplicate entries and keep only one entry per app.

- Standardization of features:

  Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way. The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

- Exploratory Analysis and Data Visualization:

  In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images. This communication is achieved through the use of a systematic mapping between graphic marks and data values in the creation of the visualization.

Conclusion:

- THE PEOPLE WHO PREFER TO STAY IN ENTIRE HOME OR APT THEY ARE GOING TO STAY BIT LONGER IN THAT PARTICULAR NEIGHBOURHOOD ONLY.
- THE PEOPLE WHO PREFER TO STAY IN PRIVATE ROOM THEY WON'T STAY LONGER AS COMPARED TO HOME OR APT.
- MOST OF THE PEOPLE PREFER TO STAY IN LESS PRICE.
- IT THERE ARE MORE NUMBER OF REVIEWS FOR OARTICULAR NEIGHBOURHOOD GROUP THAT MEAN THAT PLACE IS TOURIST PLACE.

- IF THE PEOPLE ARE NOT STAYING MORE THAN ONE NIGHT MEANS THEY ARE LESS TRAVELLLERS.

References:

- Towards Data Science
- Science Direct
- Kaggle
- emarld