Airbnb Data Analysis

BY : - Ankur Nain

ABSTRACT:

Purpose –The purpose of this project is to review the extant literature on Airbnb is one of the most significant recent innovations in the tourism sector to assess the research progress that has been accomplished to date.

INTRODUCTION：Airbnb is an online marketplace and hospitality service, aiding users to rent or lease accommodation not limited to bed and breakfasts, hostels, homestays, apartments, rooms or hotels. Airbnb does not own any of the properties but collects a brokerage fee and service fee percentages from both the host and the guest per booking. Airbnb was established in August of 2008 and founded in San Francisco, California. Accommodation located all over the world can be booked online using a tablet, mobile phone, PC or Mac (“About Us – Airbnb”, 2017).

Business Understanding:

As you know Airbnb runs on the aggregator-based model, its main income comes from the commission on every booking made. Here the property owner can list their properties for free.

On the Airbnb platform, the company carries out booking and monetary transactions. The company mainly earns from the transactions and the ways are as below.

The **business model of Airbnb** is based on the aggregator business model. This aggregator business model is a network model.

Here the company will collect the data about a specific good or the service providers. They will make the providers their partners and sell their services under its brand.

The idea behind Airbnb is simple: Find a way for local people to make some extra money renting out their spare home or room to people visiting the area. Hosts using this platform get to advertise their rentals to millions of people worldwide, with the reassurance that a big company will handle payments and offer support when needed. And for guests, Airbnb can offer a homey place to stay that has more character, perhaps even with a kitchen to avoid dining out, often at a lower price than what hotels charge.

Data Description:

For this project we have analyze Airbnb ‘s New York City(NYC) data of 2019.NYC is not only the most famous City in the world but also top global destination for visitors drawn to its museums,entertainment,restaurants and commerce.

Our main objective is to find out the key metrics that influence the listing of properties on the plateform. for this ,we will explore and visualization the dataset from Airbnb in NYC using basic explorating Data .

The idea behind Airbnb is simple to Find a way for local people to make some extra money renting out their spare home or room to people visiting the area.

Hosts using this platform get to advertise their rentals to millions of people worldwide, with the reassurance that a big company will handle payments and offer support when needed. And for guests, Airbnb can offer a homey place to stay that has more character, perhaps even with a kitchen to avoid dining out, often at a lower price than what hotels charge.

DATA DESCRIPTION:

Features Name	Type
Id	int64
name	object
host_id	int64
host_name	object
neighbourhood_group	object
neighbourhood	object
latitude	float64
longitude	float64
room_type	object
price	int64
minimum_nights	int64
number_of_review	int64
last_review	object
review_per_month	float64
calculated_host_listing_count	int64

avilablity_367 int64

Data type :

float64(3)
int64(7)
object(6)

Features Breakdown:

Id – these gives us the listing id.
name – Listing name.
Host_id/Name – host id/Name
neighbourhood_group - NYC borough
neighbourhood – NYC Neighbourhood
Latitude- listing latitude
Longitude – listing longitude
Room_type-type of rooms (Entire home/apt,Private room ,shared room)
Price – listing price
Number_of_review & Last_review - Total number of reviews & Date of last review
Calculated_host_listing_count – total number of listing for this host
Availablity_365 – number of days listing is available

Number of data set were nearly 49000.

Exploratory data analysis:

exploratory data analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization 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 and thereby contrasts traditional hypothesis testing.

Exploratory data analysis tools: Specific statistical functions and techniques you can perform with EDA tools include:

- Clustering and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.
- Univariate visualization of each field in the raw dataset, with summary statistics.
- Bivariate visualizations and summary statistics that allow you to assess the relationship between each variable in the dataset and the target variable you're looking at.
- Multivariate visualizations, for mapping and understanding interactions between different fields in the data.
- K-means Clustering is a clustering method in unsupervised learning where data points are assigned into K groups, i.e. the number of clusters, based on the distance from each group's centroid. The data points closest to a particular centroid will be clustered under the same category. K-means Clustering is commonly used in market segmentation, pattern recognition, and image compression.
- Predictive models, such as linear regression, use statistics and data to predict outcomes.

Types of exploratory data analysis: There are four primary types of EDA:

- Univariate non-graphical : This is simplest form of data analysis, where the data being analyzed consists of just one variable. Since it's a single variable, it doesn't deal with causes or relationships. The main purpose of univariate analysis is to describe the data and find patterns that exist within it.
- Univariate graphical : Non-graphical methods don't provide a full picture of the data. Graphical methods are therefore required. Common types of univariate graphics include:
    - Stem-and-leaf plots, which show all data values and the shape of the distribution.

Histograms, a bar plot in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values.
Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

Multivariate nongraphical : Multivariate data arises from more than one variable. Multivariate non-graphical EDA techniques generally show the relationship between two or more variables of the data through cross-tabulation or statistics.
Multivariate graphical : Multivariate data uses graphics to display relationships between two or more sets of data. The most used graphic is a grouped bar plot or bar chart with each group representing one level of one of the variables and each bar within a group representing the levels of the other variable.

Other common types of multivariate graphics include:

Scatter plot, which is used to plot data points on a horizontal and a vertical axis to show how much one variable is affected by another.
Multivariate chart, which is a graphical representation of the relationships between factors and a response.
Run chart, which is a line graph of data plotted over time.
Bubble chart, which is a data visualization that displays multiple circles (bubbles) in a two-dimensional plot.
Heat map, which is a graphical representation of data where values are depicted by color.

Analysis Summary:

We defined some points which can help airbnb in their business:

➢ As people loved to stay when in need for maximum number of nights in entire apartment and Brooklyn stand second highest focused placed by people.

➢ Also, as the average cost in brooklyn is 80 USD which is less then Manhattan Airbnb can increase number of entire apartment in Brooklyn.

➢ Top most place in nighbourhood is williamsburg and Bedford.

➢ Top rated host as per number of reviews are Dona and jj.

➢ The availability of rooms in year is highest in queens island and lowest in Brooklyn.

➢ So it is a great opportunity to increase some entire apartment in brooklyn.