Data Analysis Project on New York City Airbnb Open Data-set

About Airbnb

- Airbnb, Inc. is San Francisco, California based company which act as a broker and provides an online marketplace for short-term homestays.
- The company charges a commission from each bookings.
- Aibnb providing it premium service and experience to customers since 2008. Today Airbnb has millions of listings. These listings generates lost of data.
- Analyzing this data become crucial factor for the company. This data can be use for business decisions, marketing, implementations of initiatives, additional services and much more.
- Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

Problem Statement

- Given dataset describes the listing activity and metrics in NYC, NY for 2019. It include all information about host, listed properties, geographical location, prices reviews and all other required metrics.

- Analyse the given dataset make different predictions and draw meaningful conclusion in order to grow the business.
-
- Also state what can we learn from different predictions.

Business Benefits of the analysis

- By analyzing the given data set customer can make several decision about their journey and the location.
- Customer could take idea about expenses for the accommodation and which room to prefer in the particular area during journey.
- Finding the perfect location for night stay and the most preferred airbnb property according to previous customer reviews will be easy.
- This report can attract the customer who wanted to plan a trip but not visited that place before by checking about the location and number of option available for the home stays.
- This report may increase reputation and company revenue growth along with the other businesses by increasing tourism.

About Dataset

- This dataset has round 48895 listings and 16 Columns.
- Given dataset contains null values as well which we have to consider while doing analysis. Last_review and reviews_per_month has more null values.
- There are 5 neighbourhood group in which all listings located.
- Nearly 80-85% of listings located in Manhattan and Brooklyn.
- In Manhattan booking price is bit higher as compared to other neighbourhood groups.
- There are 3 kind of room type (i.e Shared Room, Private Room, Entire home/Apt).
- Out of which Shared room are least preferred by the customer even after having less price for booking.

Inspiration

- ▫What can we learn about different hosts and areas?
- ▫What can we learn from predictions? (ex: locations, prices, reviews, etc)
- ▫Which hosts are the busiest and why?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Questions

- 1.In which Neighbourhood group there is maximum number of properties listed ?
- 2.Which host has maximum number of properties listed ?
- 3.Which host has maximum properties listed in neighbourhood groups having maximum properties listed ?
- 4.What is the average price in different properties listed ?
- 5.What may be the reason of having high price in that neighbourhood groups ?
- 6.What is the most prefered room type in the every neighbourhood groups ?
- 7.Total availability of properties having different room type?
- 8.Which one is the busiest host ?
- 9.Which property has maximum number of reviews ?

Analysis questions

* Count of reviews per month

* Show total room types

* Find the total number of shared rooms, private rooms, entire home/apt

* Create a slicer for dates to show last reviewed information

* The prices for each neighborhood group

* Create a table for host to check the count of properties been listed for each neighborhood group.