EDA CAPSTONE PROJECT
ON AIRBNB

Nishank Rana (nr9556@gmail.com)

- Upload the dataset into google colab.
- Import some libraries and then mount the drive in google colab.
- Analyze null values and filter them.
- Data Visualization.
- Data wrangling.
- Observe and understand the dataset.
- Find out the Conclusion.

Github Link:- https://github.com/NishankRana/Airbnb-Data-Analysis

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Project Content:

Since 2008, guests and host have used Airbnb to expand on travelling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in New York City, for 2019. This data file includes all needed information to find out more about hosts and necessary metrics to make predictions and find out conclusions. In this project I tried to provide answers to the following states of problems.

Problem Statements:

- 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?

Approach:

In this project, I had performed a descriptive and exploratory analysis of the data, to understand how the phenomena of each variable behave individually and transversely, in addition to generate hypotheses useful for future decision-making. To complete the project, I had divided the task in several categories as in below.

Steps Involved:

1. Importing Modules and Loading Data

In the first I import some of the useful libraries like Panda, NumPy, Seaborn and Matplotlib and just load the data and checked that what kind of data I'm going to work on.

2. Understand More About The Data

When I successfully filtered out the data and now I have to find out what type of information in our dataset has given.

3. Exploring and Visualizing Data

I have explored the data using various techniques and found some useful information from it.

4. Calculating the null or missing values

Our dataset contained few numbers of null values for which I checked and analyze null values and treated it for cleaning and filtering.

5. Correlation Check

The correlation matrix to understand how are the features interrelated with each other and get an idea which are the most important features in this data.

6. Conclusion

- Manhattan neighbourhood has the most listed Airbnb as compared to others.
- Queens has most number of visitors which means people visited more here.
- In the month of july people vitisted more in manhattan.
- Brooklyn and Manhattan has maximum price which costs more than 5000 because of metropolitan city.
- Bronx neighbourhood has their maximum price Airbnb which costs 1000 per night.
- The top 5 host has listed their airbnb which has room type as private.
- The major traffic areas are where room type as entire home/apartment and private room.
- Peoples doesn't like to stay in shared room airbnb.
- The demand of Airbnb is too high in Manhattan and Queens neighbourhoods.
- In Staten Island the demand of Airbnb is too low as compared to other neighbourhoods.
- In Staten Island neighbourhood the pricing of the Airbnb is quit high. Therefor people doesn't book Airbnb there.
- In staten Island neighbourhood the lowest price tagged Airbnb 13$ is little bit higher than other neighbourhoods and maximum price for single night is 5000.
- It would be better to record the reviews on the scale of 0 to 5. It will help to identify the hosts whose services are better.