

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name: Rudrajit Bhattacharyya
Email: rudrajitb24@gmail.com

Contribution:

1. Loading the Data
2. Data Cleaning
3. Data Exploration
 - a. Finding insights about the hosts in NYC
 - b. Finding insights about the neighborhood groups in NYC
 - c. Plotting locations, room types, price variation and availability over the map
 - d. Finding insights about the neighborhoods in NYC
 - e. Finding insights about the room types in NYC
4. Conclusion

Please paste the GitHub Repo link.

Github Link:- https://github.com/Rudrajit12/EDA_Capstone_Project

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

Airbnb, Inc. 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 website and mobile app. 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.

In this EDA project, a dataset of 49000 Airbnb listings in NYC is provided and my task is to explore and find out a few interesting insights which could be helpful in decision making. The main purpose of EDA is to detect any errors, outliers as well as to understand different patterns in the data.

My first task was to understand the problem statement and the variables, and then look out for any duplicate values, missing values or errors in the data. There were a few missing values which were dropped or imputed with a value depending upon the variable at hand. After dealing with missing values, I performed some univariate,

bivariate and multivariate analysis to uncover some insights about hosts, neighborhoods, room types etc.

Through EDA, I found out that there are many hosts with multiple properties listed in NYC, hosts who do not have the most number of reviews being occupied for the whole year, hosts who are the busiest and what could be the reason behind it. Exploring the neighborhoods gave me insights like which area is the most expensive, which area gets the most traffic, what type of rooms are available, what number of reviews each area gets, what is the availability of rooms in each area etc.

By the end of the analysis, I uncovered few very important insights such as:

- Host Maya is the busiest of all and there are multiple reasons for that.
- Brooklyn and Manhattan receive the most footfall due to multiple factors being in favor like many tourist attractions, number of properties, number of reviews etc.
- Manhattan is the most expensive area but receives a good amount of customers and most of them prefer to stay in Manhattan for a longer duration as compared to other neighborhoods.
- Entire Home/Apt or Private Rooms are the most preferred choice of room types in NYC as it receives the most number of reviews, huge numbers of these room types are listed and maintains privacy more than Shared Rooms.

Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Colab notebook, project presentation and video.

Drive Link:

<https://drive.google.com/drive/folders/1IYWk3S9kdMduH8IMw6nvWo2dlrs7oKwY?usp=sharing>