

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1. **Mohd Taufique:** taufiquemohd2@gmail.com
 - Data Exploration & Variables Identification.
 - Data Cleaning- Remove Duplicates and Handled Missing Values.
 - Numerical Variable Analysis: Price, Number of Reviews, availability_365.
 - Categorical Variable Analysis: Host name, neighbourhood_group & room type
 - Presentation, Technical Documentation, Project Summary
2. **Azhar Ali:** aliazhar007007@gmail.com
 - Data Understanding
 - Data Wrangling- Removing duplicates and tackled null values
 - Numerical Variable Analysis- calculated_host_listing_count, price
 - Categorical Variables Analysis- name, neighbourhood
 - Presentation, Technical Documentation
3. **Pushpam Raghuvanshi:** raghuvanshipushpam1991@gmail.com
 - Data Cleaning- imputing null values
 - Numerical variable Analysis- host id, price
 - Categorical Variable Analysis- room type, neighbourhood group
 - Presentation, Technical Documentation
4. **Aishwary Sharma:** aishwarys2@gmail.com
 - Data Cleaning- Handled missing values
 - Numerical Variable Analysis- minimum night
 - Categorical Variable Analysis- neighbourhood group
 - Presentation, Technical Documentation

Please paste the GitHub Repo link.

Mohd Taufique Github Link:- https://github.com/MOHD-TAUFIQUE/Airbnb-Bookings-Analysis_Capstone-Project

Azhar Ali Github Link:- <https://github.com/Azhar-ali7/Airbnb-NYC-EDA>

Aishwary Sharma Github Link:- <https://github.com/aishwrysharma/Airbnb-EDA-Capstone-Project>

Pushpam Raghuvanshi Github Link:- <https://github.com/pushpam-raghuvanshi/airbnb-bookings-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)

Airbnb (ABNB) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales. For the Exploratory Data Analysis, we are using Airbnb's New York City Booking data of 2019. This Dataset has around 49,000 observations in it with 16 columns and it is mix of categorical and numerical values. The listings at Airbnb generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and implementation of innovative additional services and much more.

The problem statement was to identify the room types and their count offered by the hosts, what are the popular & expensive neighborhood group, what are the availability and average prices of different room types across all neighborhood group, who are the Top hosts with listings, most frequent words used in the listings, proportion of listings and its distribution in different neighborhood group.

The first step in the analysis involved exploring the data, identifying the variables, and then performed data cleaning like removing the duplicates, anomaly correction, looking for any null values and tackling them.

The second step involved analyzing the different numerical and categorical features and show the analyzed result using different visualization charts like bar graph, clustered bar chart, box plot, pie chart, distplot etc. Finding the correlation between each variable and identify the important features that a guest looks at the time of booking.

The Final step involved is summing up the key observation & insights developed during the analysis. Some key takeaways were; Entire Home/apt is the most preferred room type, Manhattan is the most expensive city for booking all types of room, most guests prefer a cheap price, Queens and Manhattan are the favorite guest destination with highest number of reviews, the host Sonder (nyc) is the most popular host with highest listings and Dona is the busiest host with the most number of reviews, Williamsburg is the neighborhood with highest number of listings in the entire NYC etc.