

Capstone Project Submission

Instructions:

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

Team Member's Name, Email, and Contribution:
Aishwarye Gangwar Contributions: 1-Studying the Data set 2-Data cleaning: 1. taking important columns 2. Replacing null/NAN values 3-Information about hosts and areas 4-Predictions 5-Busiest host 6-Traffic across the area and reasons 7-Preferable properties 8-Number of Airbnb 9-Correlation between Data set 10-Conclusion
Please paste the GitHub Repo link.
GitHub Link:- https://github.com/laishwarye/first-eda-capstone-project Colab link :- https://colab.research.google.com/drive/1Ni7QyOnKb8Lo_I-38fEJPqrIlzDaB30c?usp=sharing
Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Airbnb 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 the website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving a commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

As a first step, we take the overview of data, where we specially made our focus on understanding what each column means. So that we can be clear from what perspective we have to analyze our data. After understanding different columns, we marked a few important columns and kept them for the further analysis process.

For problems, we perform different analyses. Most of the analysis can easily be done with the group by and after analysis we also perform data visualization .in data visualization we make 'Bar', 'Histogram', 'Boxplot', 'Pie chart' and 'Heatmap'. We found out about the availabilities of properties also.

We also found out about the most preferable properties across the areas. We also established a correlation graph with our data frame to find out how each column is correlated to the other.

There are various points we can take out from our analysis :

- Most numbers of listings are from Manhattan
- Host Sonder(NYC) have the most number of listing.
- Heavy requirement of entire or private rooms.
- Cause of the heavy requirement of entire or private rooms, prices of them are on the higher side.
- Different room type densities in different neighborhood groups.
- Busiest hosts are Sonder, Michael, and Blueground.
- Most of the traffic happening in Manhattan, Brooklyn
- Private bedroom in Manhattan is the most preferable property.
- Properties availabilities.
- Number of Airbnb across the different neighborhoods.