

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

HRITIK SHARMA

Email- hritik.2.sharma@gmail.com

Contribution-

1. Data cleaning
 - ☐ Contributed in finding irregularities in data(Null values, erroneous values) ☐ Actively involved in removal of all data discrepancies.
2. Data validation
 - ☐ Successfully performed data validation of cleaned data and ensured that dataset is ready for analysis.
3. Strategizing Roles for the analysis
4. Contributed in strategizing the exploration process
5. Contributed in strategizing the visualization process
6. Presentation of the final report

Please paste the GitHub Repo link.

Github Link:- <https://github.com/hritiksharma11/capstone-project-air-bnb-eda-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)

This study aims to understand the customer's and host's behaviour and performance on the platform of Airbnb through an exploratory study that accounts for around 49,000 observations in New York, the United States of America, in neighbourhoods with a high number of accommodations listed on the Airbnb platform.

Our analysis can help in the understanding of different hosts and areas, which hosts are the busiest, which type of property listing is most preferred by customers, and which neighbourhoods have the highest traffic.

Our goal here is to perform an exploratory data analysis on the Airbnb NYC dataset, which could help in understanding the story the dataset entails.

Our approach was to strategically divide the whole analysis into six phases in order to divide the project into modules.

Ask-In this phase of the analysis, we asked ourselves what is the main problem statement we are trying to answer.

Prepare-In this phase, we made sure that the dataset we were given was good enough to answer our queries or that some additional data is required for our analysis. **Process**-In this phase, the dataset cleaning process was started.

Duplicate values removal-Firstly we checked for any duplicated observation through spreadsheet's remove duplicated functionality which was not found in our dataset.

Null values removal-Secondly, we checked for null values in the dataset and found that our dataset contained a large number of null values therefore for accurate results we dropped the columns containing nulls. In the review_per_month column, we assumed that NAN would imply zero reviews that month so we replaced it with zero.

Data validation-Thirdly, we validated our cleaned data based on data type, range, constraint, structure, and consistency.

Analyse-This phase was all about making sense of data by identifying trends and relationships within the data. This involved four steps

Organize data-Locate or access data for the problem

Format and adjust data-Filtering and sorting data

Get input from teammates-Getting valuable insights from teammates on possible trends. **Transform data**-Identifying trends and relationships in data.

Share-In this phase, we used visualizations to explain our findings mainly through matplotlib and seaborn libraries in python.

Act-Using our observations from above to solve the problem statement.

We dealt with missing data and outliers. That's a lot of work that Python helped us make easier. This analysis gave us great insight into which neighbourhood group has the highest and lowest average listing which can be correlated with traffic density in respective neighbourhood groups. Also, we learned which type of listing among the three(Apt./entire home, private room and shared room) are most preferred in each neighbourhood group.

