

Capstone Project Submission

Team Member's Name, Email and Contribution:

- **Name - Deepak Karki**
Email – dishunwab0@gmail.com
 - Data Understanding
 - Feature Analysis
 - Null value check, Fill null values
 - Data Visualization
 - Bar charts,
 - Research Analytics
 - Technical documentation
- **Name – Prakshita Agrawal**
Email – prakshita21@gmail.com
 - Data Understanding
 - Feature Analysis
 - Remove Null values Columns
 - Data Visualization
 - Box plot, Pie Chart, Heat map
 - Multivariate Analysis
 - Correlation matrix
- **Name – Bharat Soni**
Email - bharatsoni0047@gmail.com
 - Data Understanding
 - Data Conduct
 - Collect information and idea
- **Name – Bhatu Sonawane**
Email – sonawanebhatu406@gmail.com
 - Data Understanding
 - Data Conduct
 - Collect information and idea
- **Name – Sopan Wadekar**
Email – sopanwadekar777@gmail.com

- Data Understanding
- Data Conduct
- Collect information and idea

Please paste the GitHub Repo link.

Github Link:-

[https://github.com/praxita21/Airbnb-Analysis/blob/main/Airbnb_Bookings_Analysis_Prakshita_Agrawal_Capstone_Project.i
pynb](https://github.com/praxita21/Airbnb-Analysis/blob/main/Airbnb_Bookings_Analysis_Prakshita_Agrawal_Capstone_Project.ipynb)

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

Problem statement :-

Finding out the reviews and see the busiest and non-busiest area steps by providing offers based on the stay and review factors. Using the data provided, this paper aims to analyze the data to determine which variables are correlated, with customer. An analysis will also be done to identify the people wants for the rooms.

APPROACH:

Exploratory Data Analysis (EDA) is an approach to analyze data. The first and foremost task is that the data analysis to view the data and tries to make some sense out of it. Later we figure out what question we want to ask and how to use the available data to get the insights and all the answers that we need from the data set. EDA helps us to

- Delve into the dataset
- Examine the relationships among the variables
- Identify the interesting observation
- Develop an initial idea of possible associations among the predictors and the target variable. Our airbnb data set consists of 48895 columns and 16 rows records. out of which 4 objective, integer, and float and Boolean data type.

Conclusion -:

Starting with the dataset loading, we have performed data pre-processing, EDA

Manhattan and Brooklyn are the two top most popular neighborhood groups in terms of hosts count, number of reviews, number of listing, maximum number of nights spends in these areas. So it might also be reason of traffic and high prices.

For other neighborhood groups namely Queens, Bronx and Staten island there aren't as popular as these two, especially on Staten Island.

The dataset can be further used for price prediction by building a linear model. The data needs to be treated of outliers and skewness for a linear regression as well as other models.

This Airbnb(NYC 2019) Dataset For The Year 2019 Appeared to be very rich Dataset with a variety of columns that allowed us to do deep data exploration.

In the column name and host_name which have 16 and 21 null value only. Null values are present in last_review and reviews_per_month which can be dropped both have most null values is 10052.

From the dist. plots it can be observed that latitude and longitude data seem to be normally distributed and most of the numeric_features are positively skewed.

People stay for longer duration of time in private rooms in Brooklyn and Manhattan.

More customers preferred Manhattan location for night stay then Brooklyn.

Entire home/apt room type has the highest number of listing of 52% and shared room is the least listed room type at only 2.4% in total.

63.2% costumer spend night in entire home and 1.6% spend night in shared room.