

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

- | | |
|--------------------|--|
| 1. Rohan Solanke | rohansolanke1702@gmail.com |
| 2. Meghana RS | meghanars70@gmail.com |
| 3. Rishekh Dubey | 91221rishekhdbey@gmail.com |
| 4. Prabhujeet Kaur | prabhuilu09@gmail.com |
| 5. Sunil R M | sunilrm1999@gmail.com |

Contribution Role:

1. Rohan Solanke

- a. neighbourhood_group df
 - i. neighbourhood_group vs price
- b. longitude_df
- c. latitude_df
- d. Room type per neighborhood group
 - i. Brooklyn
 - ii. Manhattan
 - iii. Staten Island
 - iv. Bronx
- e. Locations on which different types of rooms are present
 - i. Private
 - ii. Entire home/apartment
 - iii. Shared Room

2. Meghana R S.

- a. Airbnb Bookings NYC dataframe
 - i. 1.1 Cleaning and transforming features in df
 - ii. 1.2 Null value treatment
- b. Different host names with maximum number of entries in the provided Airbnb Bookings dataset.
- c. Calculated host listings count vs host name.
- d. Different types of rooms available vs the number of rooms.
- e. Average of review per month vs different room types
- f. Correlation between different features in the dataset.
- g. Handling outliers in the price column using IQR method.
 - i. 7.1 Comparing the nature and distribution of prices allotted for rooms before and after trimming the outliers from the price column.
 - ii. 7.2 Comparing the price variation(without outliers) vs different types of rooms.
- h. Finding the average price of top 10 most reviewed listings in NYC.

3. Rishekh Dubey

- a. Finding the Hotels that have the same neighborhood.
- b. Finding the Hotels that provide only private rooms as room type.
- c. Minimum nights that are available in the hotels.
- d. Hotels that have a minimum of 2 or more nights.
- e. Rooms which cost less than or equal to 100.

4. Prabhujeet Kaur

- Finding the average price of NYC based on the number of reviews.
- Finding the neighborhood group vs number of reviews.
- Finding the average reviews of NYC in different neighborhoods.
- Comparing the price in different latitude and longitudes.
- Comparing the number of reviews in different latitude and longitudes.

5. Sunil R M

- Extract the Day, Month and Year from the last_review column.
- Create new columns like Day, Month and Year.
- Year wise types of rooms getting a last review and plot Bar graph.
- Find how many Neighborhood groups are available 365 days.
- How many last reviews of every month (Count last_reviews per month).

Please paste the GitHub Repo link:

<https://github.com/Prabhujeeet-K/Airbnb-Booking-Analysis>

Drive link:- https://drive.google.com/drive/folders/1T7Snq109ecZ_IR59nFzoF9OvAp2T5nRW

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. Explore and analyze the data to discover the key factors responsible for bookings.

As a team, we were tasked with locating a dataset of our choosing, and performing cleaning, and EDA. We choose an AIRBNB dataset because of the varied and interesting features.

Our primary objectives were to deliver insights to understand customer demands better and thus understand the demands for bookings considering various factors available in the data.

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

First of all we find the null values, then removing the null values and then we start the EDA finding the relation between different features by different visualization tools and then concluded the analysis.

Conclusion:

- Sonder is the busiest host on Airbnb after then the Michael comes.
- Shared rooms are less available compared to other room types but when we check the numbers of reviews we find out those private and shared rooms are preferred more compared to the entire room. So it means we have to increase the availability of room type other than entire must be increased to gain profit.
- Some of features positively correlated with each other.
- The price of private and shared room types is less compared to the entire room type.
- The Topmost reviewed room is the private type which has a price below 50\$. It means users prefer cheap rooms.
- Manhattan has the highest number of hotels which have availability 365 then Brooklyn comes, as Manhattan is famous for museums, stores, parks and theatres that's why more hotels are available.
- If we see closely the private room type is highest in Brooklyn then Manhattan comes. So as we discuss before private room type is high in demand and also this place near to Manhattan so visitors can visit both in cheap prices.
- Maximum hotels consider one minimum night which is good for visitors to stay.
- Most of the last reviews came in June month.
- Staten Island has the maximum number of reviews than Queens that's why most of the rooms are not available all 365 days compared to Brooklyn and Manhattan.
- The silver lake in Staten Island has more reviews than other places in NYC. Most of the prices are less in Staten Island and Queens because the number of reviews was more in these places than in

other boroughs.

However, features related to accommodation, room type, and neighborhoods in Manhattan, Queens, and Brooklyn play an important role in determining future price of the listings. These features would be crucial for AirBnb to predict revenue.

