Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Sl. No	Name	Email	Contribution
1.	Neha R	nehareddy.dr@gmail.com	1. Dataset: a) Importing dataset. 2. Data Clean up: a) Created the function to find out the Null Values. b) Addressed all the Null values and handling of the missing data were done. c) Filling all the missing values was done. 3. EDA Performing: ➔ EDA based on highest number of reviews: a) Analysed the top 5 neighbourhoods having the highest number of reviews per month. b) Analysed the top 5 neighbourhoods having the highest number of reviews. ➔ EDA based on top 10 neighbourhood: a) Analysed the maximum number of apartments for Airbnb. ➔ EDA based on the maximum number of prices in their respective neighbourhood group.

2.	Abhishek V L	abhisheklakshmana7@gmail.com	1. Dataset ➔ Understanding the data of the Airbnb. ➔ Checking the feature names present in the dataset. ➔ Finding the location of the Airbnb. 2. EDA on predictions: a) Analysed the locations. b) Analysed the prices. c) Analysed the reviews and so on…. ➔ Distribution of the room type and its distribution over the location. ➔ EDA performance: a) EDA based on the neighbourhood groups: ➔ Distributed Room type over a neighbourhood group. ➔ Ratios of respective room types are more or less the same over each neighbourhood group. b) EDA based on price: ➔ Price column is distributed over a room type. ➔ Surprising items in the price column.
3.	Swati R G	swatigajbhiye026@gmail.com	1. Data processing: ➢ Average preferred prices by customers according to the neighbourhood group for each category of room type. 2. EDA performing: ➢ Analysed the host having a highest number of apartments ➢ Analysed the average preferred price by customer according to neighbourhood group for each

			category of room type. a) EDA based on average price. ➢ Performed EDA on average price preferred for keeping good number_of_reviews according to neighbourhood_group.
4.	Nisarga C	nisargac5577@gmail.com	1. Data Analysis: ➔ Price in this range to get more reviews in specific room types and at a particular place. 2. EDA Performing: ➔ Analysed the top 10 neighbourhood having the number of apartments for Airbnb. ➔ Created a neighbourhood in each group having maximum prices in their respective neighbourhood group ➔ EDA on predictions : a) Locations b) Prices c) Reviews

Please paste the GitHub Repo link.

Github Link:- https://github.com/Nisargaavi/AIRBNB-BOOKING-ANALYSIS

Google Drive link:-
https://drive.google.com/drive/folders/1h74iXP-BPGJ8xb9qYPj6KJHokpCGvApv?usp=sharing

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

Summary:

We took this project because of mutual interest, so at least we had a basic knowledge of what all could have in the data set. And we got some of the data which I was expecting.
After downloading the data set, we just divided according to the columns and then data cleaning and handling of missing values was done, then we performed a few of the EDA's. Here we analysed the EDA's with the help of graphs with the help of matplotlib. pyplot and seaborn.

Problem Statement:

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values. Since 2008, guests and hosts have used Airbnb to expand on travelling possibilities and present a more unique, personalised way of experiencing the world. Today, Airbnb became one-of-a-kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analysed and used for security, business decisions, understanding of customers' and providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

Explore and analyse the data to discover key understandings (not limited to these) such as:

1. **How neighbourhood is related with reviews**
2. **Which are the top 5 hosts that have obtained the highest no. of reviews?**
3. **Which hosts have the highest number of apartments?**
4. **Which are the top 10 neighbourhoods? which has the maximum number of apartments for Airbnb?**
5. **What are the neighbourhoods in each group which are having maximum prices in their respective neighbourhood group?**
6. **What can we learn from predictions? (Ex: locations, prices, reviews, etc)**
7. **What is the distribution of the room type and its distribution over the location?**
8. **How does the Room type is distributed over Neighbourhood Group are the ratios of respective room types more or less same over each neighbourhood group**
9. **How is the price column distributed over room type and are there any surprising items in the price column?**
10. **What is the average preferred price by customers according to the neighbourhood group for each category of Room type?**
11. **What is the average price preferred for Keeping good number_of_reviews according to neighbourhood group?**
12. **Which hosts are busiest and why?**

Conclusion:

From the given dataset, after performing EDA based upon all of the above mentioned, we conclude the inferences which we found were:

1. **Top neighbourhood having the highest reviews per month is Theatre District with 58.50.**
2. **Top host with the highest number of reviews is Maya with 2273.**
3. **Maximum number of apartments for Airbnb is Williamsburg with 3920.**
4. **Neighbourhood in Manhattan having the maximum price is the Upper West Side with 10000.**
5. **Neighbourhood in Staten Island having the maximum price is Randall Manor with 5000.**
6. **The Neighbourhood in Bronx having maximum price is Riverdale with 2500.**
7. **Neighbourhood in Queens Which has the maximum price is Astoria with 10000.**
8. **The Neighbourhood in Brooklyn having the maximum price is Greenpoint with 10000.**
9. **Highest price of all ratings, minimumnights, availability_365, last review in order judge is Kathrine with review_per_month of 0.04 and last review of 2016-02-13 and availability_365 is 0 and price is 10000 and neighbourhood group is Queens.**
10. **Manhattan is most costly and Bronx is cheap for each room type.**

- **There was a lot of duplicate data.**
- **Choosing appropriate visualisation techniques to use was difficult.**
- **A lot of null values were there in the dataset.**