Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. Vineeta Singh(singhvineeta0118@gmail.com):

Descriptive Analysis
- Dataframe description
- Dataframe shape

Data cleaning and Analysis:
- Drop unwanted columns['ID','last review']
- Box Plot of price

Data Wrangling
- Relation between 'Room_type' and 'minimum_nights'
- Relation between 'Room_type' and 'number of reviews'
- Relation between 'Neighbourhood group' and 'number of reviews'

Data Visualisation
- Relationship between Room_type' and 'minimum_nights'
- Relationship between Room_type' and 'Number of reviews'
- Relationship between Neighbourhood_group' and 'number of reviews'

2 Tushar Gupta (gupta.tushar1992@gmail.com):

Descriptive Analysis
- Dataframe description
- Dataframe shape

Data cleaning and Analysis
- Removing NaN values from columns like Price and Host Name
- Correlation Plot of all variables
- Box plot between price and neighborhood group

Data Wrangling
- Relation between 'Neighbourhood_group' and 'most number of host'
- Calculation of average price per neighborhood group
- Maximum benefitted host

Data Visualisations
- Relationship between 'Neighbourhood_group' and 'number of host'
- Relationship between Room_type' and 'Neighbourhood_group'

3. **Abhishek Mishra(abhishekmishra9559026@gmail.com):**

Descriptive Analysis
- Dataframe description
- Dataframe shape

Data cleaning and Analysis:
- Handling null values of 'reviews per month'
- Find zero price columns and remove them
- Removing null values from column name "name"

Data Wrangling
- Relation between 'Price' and 'reviews'
- Find top three busiest host

Data Visualisation
- Relationship between Room_type' and 'review' using scatter plot
- Plotting Graph of busiest host

4. **Mallesh (kurvamallesh36@gmail.com):**

DescriptiveAnalysis
- Dataframe description
- Dataframe shape

Data Wrangling
- Relationship between top three host and each neighbourhood group
- Room types preferred in Manahattan and Brooklyn

Data Visualisation
- Relationship between top busiest hosts in each neighbourhood group
- Geographical graph if neighbourhood

5. Arunesh Mishra(Arunesh12mishra@gmail.com):

Descriptive Analysis
Dataframe description
Dataframe shape

Data Wrangling
Room types preferred in Manahattan and Brooklyn

Data Visualisation
Geographical graph of neighbourhood
Relationship between busiest host in each neighbourhood group

Github Repository link:

Vineeta Singh- https://github.com/vineeta0118/Airbnb-EDA-capstone-project
Tushar Gupta-https://github.com/tushar2718/Airbnb-EDA-capstone-project
Abhishek Mishra- https://github.com/abhishekmishra-bareilly/EDA-capstone-project
AruneshMishra- https://github.com/kajuun/EDA-project
KurvaMallesh-https://github.com/kurvamallesh/EDA-project

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

Airbnb (AirBed & Breakfast, where Air was nothing but mattresses) comes with a small idea ofearning rent by Brian Chesky and Joe Gebbia in 2007 turned hospitality and travel industry on itshead and gained massive recognition now. In this Capstone Project, we were given dataset of 49000 observations distributed in 16columns.Columns are of two types categorical and numerical type.

Initially we started with analyzing the data. And after some initial analyzing , it was gathered that some columns are not useful and then we drop them. And some columns have NaN (null) values which also need removal. This whole process waster med as Data Cleaning and Analyses.

Next moving forward , started with Descriptive analysis over the raw data provided using box plot, correlation and other descriptive methods. The Project is now further broken into majorly four parts according to the question we need to answer.

At first, hosts and room_types were analysed. From that it was concluded that 'Entire home/apt' were most preferred type of room. On analysis, using data visualization techniques a clear relationship between room_type and minimum number nights spent was shown, which concluded that if the rooms type is 'Entirehome/apt' then customers spent more number of nights there.

Next after analyzing the neighbourhood group with respect to average price at each location, Manhattan was concluded as costliest state among others and possible reason could be the financialstateof entirecountry. People are not money conscious when spending money mainly in Manhattan and Brookyln as they want more privacy

And then analyzing was done based on 'number of reviews and host listing' which provided host were giving more reviews in most listed neighbourhood

Also we looked for the top three busiest host in US and also in top three busiest in each neighbourhood group

After exploring more about neighbourhood, we can conclude that shared room type were less likely preferred by US customers of Airbnb. And Bedford -studyvesent and Williams burgwere most populated from Manhattan state.

We can conclude that most of the Airbnb hosts are present in Manhattan and Brooklyn. After correlation between price and neighbourhood,