Capstone Project Submission

Team Member's Name, Email & Contribution:

Name: Saniya Bubere
Email:saniyabubere06@gmail.com

Contribution:

1. **Exploring the data**
2. **Null Values Treatment**
3. **Visualizations for the following:**
    - Top 10 Highest Reviews based on App and Top 10 Lowest Review App
    - Top Categories on Google Playstore using Barplot
    - Print Maximum & Minimum Size of Apps & Print Maximum And Minimum Price of Ap
    - Wrote a function in which we get to know the rating of every app by giving Rating As Input
    - Top 10 Paid App & Percent of Free Vs Paid Apps instore using pie chart
    - Which age group in Content Rating has more Contribution use chart to show output using bar chart
    - Which Sentiments have Greater Portion using pie chart & Calculate Average of Sentiment Polarity & Sentiment Subjectivity
    - Sentiments based on each Numeric Data & Installation with each and every SENTIMENTS & Plot a Bar Graph for Same

Name: Ashi Saxena
Email: ashisaxena710@gmail.com

Contribution:

1. **Analysed the given data sets**
2. **Dealt with the null values, removed unnecessary columns**
3. **Explored and visualized the data based on the following objectives:**
    - Objective-1: Obtain the top 15 costly apps and their category.
    - Objective 2: Observing the number of apps in each category.
    - Objective-3: Observe the sentiments of the free and the paid type app.
    - Objective-4: Observing the installs of each category.
    - Objective-5: Draw comparsion among the top Apps in the communication and game category.

Name: Rahul Jadhav
Email: rahultj333@gmail.com@gmail.com

1. **Exploring the data**
2. **Null Values Treatment**
3. **Visualizations for the following:**
    - comparisons in % of Paid and free app
    - find the contest rating analysis with apps
    - No of apps according to category
    - no of installs - according to category
    - average rating in each category
    - revenue generated
    - % of sentiment
    - common phrase - positive review
    - common phrases - negative review

Name: Rishabh Rohil

Email: rishabh.r@iitgn.ac.in

1. **Exploring the data**
2. **Null Values Treatment**
3. **Visualizations for the following:**
    - box plots for size and rating
    - find the contest rating analysis with apps
    - Based on Content Rating
    - Based on Category
    - based on rating
    - Based on Size Range
    - Based on Android Version
    - Based On Last Update
    - Based On revenue Generation

Please paste the GitHub Repo link.

Github Link:-
https://github.com/SaniyaBubere/EDA-GOOGLE-PLAY-REVIEW-ANALYSIS-PROJECT

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

Data Set: The Play Store App Review Analysis This dataset consists of two CSV files:googleplaystore.csv&googleplaystore_user_reviews.csv

googleplaystore.csv contains the following columns such as [Appname, Category, Rating, Size, Installs, Price, etc] and googleplaystore_user_reviews.csv. contains the following columns such as [Appname, User_Review, Sentiment, etc]

The objective of the project is to visualize the dataset provided to explore meaningful patterns and statistics.

Problem Statement: Data is taken from the Google play store dataset. Every row contains various entries regarding a certain app. We will be doing Exploratory data analysis on this data set, which is a very important step in data science cycle, as it not only helps in taking very initial business decisions .Our objective will be to structure the data, clean it and present certain trends that we observe that can help us draw very preliminary conclusions about the probability of success of a newly launched app. Also our objective is to Explore and analyze the data to discover key factors responsible for app engagement and success.

Approach: After loading the dataset and importing the libraries like Numpy, Pandas, Matplotlib, Seaborn, Plotly, etc., our first step was to clean the data. After cleaning the dataset, it got reduced from 135 columns to 19. Then we tried to establish meaningful relationship between these variables and presented the insights we got using visualization charts like pie chart, scatter plot, correlation plot, bar plot, etc.

Conclusion: After completion of the Project we have learned data cleaning and how to filter the data. We learned to differentiate between which data is essential and which is not from the given dataset and how to visualize it using various visualization techniques.

The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.After performing the EDA on the Dataset we made various observations and found that The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product. Dataset can also be used to look whether the original ratings of the app matches the predicted rating to know whether the app is performing better or worse compared to other apps on the Play Store.