Capstone Project-1 Submission

Team Member's Name, Email, and Contribution:

1. Name: - Rajesh Mohanty
   Email ID:- Rmohanty385@gmail.com

   - Contributed to notebook and helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization
   - Contributed to the contents of PPT.
   - Contributed to Technical Documentation in the content of the problem statement goal of the project and steps involved.

2. Name: - Manishwar Gupta
   Email ID: - manishwargupta2109@gmail.com

   - Contributed to the notebook for data cleaning, data manipulation, and in EDA Visualization and finalizing the conclusion.
   - Contributed to the contents of PPT
   - Contributed to Technical Documentation in the content of the problem statement goal of the project and steps involved.

3. Name: - Manishankar Kumar Shaw
   Email ID:- Manishankarksaw15@gmail.com

   - Contributed to notebook and helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization
   - Contributed to the contents of PPT.
   - Contributed to Technical Documentation in the content of the problem statement goal of the project and steps involved.

4. Name: - Ashish Maharana
   Email ID:- ashishsid0467@gmail.com

   - Contributed to notebook and helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA

Visualization.
- Contributed in making of project summary

5. Name: - Sairindhri Jena
   Email ID:- rssairindhri.99@gmail.com

   - Contributed to notebook and helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization
   - Contributed to Technical Documentation in the content of the problem statement goal of the project and steps involved.

Please paste the GitHub Repo link.

Rajesh Mohanty Github Link: - https://github.com/Rmohanty385/Playstory-app-review-analysis

Manishwar Gupta Github Link: - https://github.com/manishwargupta/Play-Store-App-Review-Analysis

Manishankar Kumar Shaw Github Link: - https://github.com/manishankarksaw15/Play-store-App-and--user--review-analysis

Ashish Maharana Github Link: - https://github.com/ashishsid0467gmailcom/Play-store-app-user-review-analysis

Sairindhri Jena Github Link: - https://github.com/Sairindhrijena/Playstore-app-analysis/find/main

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 science can be summarized into five steps: capture, maintain process, analyze, and communicate. The analysis of Google Play Store application aided to build most reliable and more interactive applications. This would be very useful for app developers to build an application focused on certain discussed category in this analysis. This analysis will help in building the application with precise and accurate objectives.

In the initial phase, we focused more on the problem statements and data cleaning, in order to ensure that we give them the best results out of our analysis. Our major challenge was data cleaning, In Data Cleaning, we have performed few steps to ensure the data quality such as removing NAN values. During the Data Cleaning step we found that 13.60% of reviews were NaN values, and even after merging both the data frames, we could not infer much in order to fill them. Thus, we had to drop them.

The merged data frame of both play store and user reviews, had only 816 common apps. This is just 10% of the cleaned data, we could have given more valuable analysis if we had at least 70% - 80% of the data available in the merged data frames.

User Reviews had 42% of NaN values, which could have been used for developing an understanding of the category wise sentiments, which would help us to fill 13.60% NaN values of the Reviews column.

With the cleaned data, we have performed Exploratory Data Analysis to understand our dataset like number of installations for each category We explore the correlation between the size of the app and the version of Android on the number of installs and so on.

Our motive in whole project was to analyze the data and find out main components that affect users' decision to download app. After completion of analysis I concluded that user prefer more of free apps. Most of the apps present in play store are more or less of same size so size doesn't affect their decision much.

It was found that Most of the apps that are present on the google play store have rating in between 4 and 5.Also it was observed that Maximum number of applications present in the dataset are of small size.

We found most popular category of apps on two basis - Number of Installs and Number of reviews.

In the problem statement we are given with 2 datasets i.e. play store and User review data set in the user review dataset it was observed that User Reviews had 42% of NaN values, which could have been used for developing an understanding of the category wise sentiments, which would help us to fill 13.60% NaN values of the Reviews column.

Most of the reviews are of Positive Sentiment, while Negative and Neutral have low number of reviews. 8.Sentiment Polarity / Sentiment Subjectivity

Collection of reviews shows a wide range of subjectivity and most of the reviews fall in [-0.50,0.75] polarity scale implying that the extremely negative or positive sentiments are significantly low. Most of the reviews show a mid-range of negative and positive sentiments.

Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.

Sentiment Polarity is not highly correlated with Sentiment Subjectivity.

The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project. Many other interesting possibilities can be explored using this dataset.

From the results and process we have implemented, we can conclude that we have achieved this group project objective which is analyzing the Google Play Store apps and determine trends of the Google Play Store and both of our research questions.