Capstone Project-1 Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. Name: - Manjeet Srivastava
   Email-ID: srivastavamanjeet01@gmail.com
   - Contributed in data preparation, cleaning and visualization.
   - Contributed in editing and making graphs for the PPT.
   - Contributed in finalizing raw collab notebook.

2. Name: - Vivek Kumar
   Email ID: viku35@gmail.com
   - Contributed in notebook for data preparation and manipulation.
   - Contributed in analysis part of the PowerPoint presentation.
   - Contributed in analysis of graphs of Technical Documentation.

3. Name: - Rishabh Srivastav
   Email ID: rishabhsri999@gmail.com
   - Contributed in notebook for data preparation, manipulation and EDA visualization.
   - Contributed in preparation of Project Summary.
   - Contributed in Technical Documentation in creation of problem statement, goal of project and steps involved.

4. Name: - Shubham Bharti
   Email ID: shubhambharti731@gmail.com
   - Contributed in notebook for data cleaning and preparation.
   - Contributed in preparation of PowerPoint presentation.
   - Contributed in conclusion of Technical documentation.

5. Name: - Durga Shankar

Email ID: dspathak50@gmail.com
- Contributed in notebook for data cleaning and manipulation.
- Contributed in finalizing PowerPoint presentation.
- Contributed in preparation of Project Summary.

Please paste the GitHub Repo link.

Manjeet's GitHub Link: - https://github.com/manjeetsrivastava/play-store-app-reveiw-analysis

Vivek's GitHub Link: - https://github.com/endlessstory35/google-play-store

Rishabh's GitHub Link: -https://github.com/RishabhSrivastav-1994/play-store-app-review-analysis

Shubham's GitHub Link: - https://github.com/Shubh-am1999/play_store_app_analysis.git

Durga's GitHub Link: - https://github.com/bigdeepak/play-store-app-reveiw-analysis

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 is the most in demand industry subject currently. It's quite apt to say, "Data is the new oil." In this capstone project, we have done and exploratory data analysis on Google play store based on existing user data and reviews. This analysis would help the app and gaming industry to focus more positively on industry demands and requirements. Moreover, this would aide app developers to build applications and games focused on certain discussed category in this analysis.

In the initial phase, we focused more on creation of the problem statements and data preparation, in order to proceed with our analysis. One of the challenges, we faced initially was data cleaning. In Data Cleaning, we performed few steps to remove unnecessary data and as such, 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.

The goal of the entire project was to evaluate the data and identify the key factors that influence users' decisions to download applications. We came to the conclusion that users favor free apps after finishing our analysis. The majority of apps available in the Play Store are roughly the same size, therefore size has no bearing on their choice.

Most of the apps on the Google Play market have ratings between 4 and 5, it was discovered. Additionally, it was noted that the majority of apps in the dataset are of a small size.

We found most popular category of apps on two basis - Number of Installs and Number of reviews. Personalization wins in former criteria whereas Sports wins in later criteria.

In the problem statement, two datasets are provided: the Play Store dataset and the User reviews dataset. In the User reviews dataset, it was found that User Reviews had 42% of NaN values, which could have been used to develop an understanding of the category-wise sentiments and fill in the Reviews column's 13.60% NaN values.

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

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

Sentiment Polarity is not highly correlated with Sentiment Subjectivity.

The dataset offers vast opportunities to raise business values and make a difference. It is not restricted to the issue considered for this project. Using this dataset, many further intriguing options might be investigated.

We may infer from the outcomes and the methodology we used that we were successful in achieving the goal of our group project, which was to examine Google Play Store apps and identify market trends related to both of our research topics.