Capstone Project Submission

Team Member's Name, Email and Contribution:

1. Name: Shahrukh Ahmad
2. Email: shahrukhspn2016@gmail.com
3. Contribution:
   a. Contributed in the google colab notebook for the complete data cleaning and pre-processing part and in matplotlib plots used for data visualization.
   b. Contributed in the Technical Documentation for preparing the Abstract, Data Cleaning steps involved, Exploratory Data Analysis outcomes and the Conclusions.
   c. Contributed in the Presentation to find out the required contents, useful inferences, and the essential steps that are required for the proper flow of the presentation.

Please paste the GitHub Repo link.
GitHub Link: - https://github.com/Shahrukh2016/Play-store-review-analysis-EDA
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Google play store is one of the largest and most popular Android app stores. It contains huge amount of data that can be used to make an optimal model and can be used for identification of trends and future challenges. In this EDA project we have used two raw data sets of Google Play Store one is of play store attributes and other is of user reviews from Kaggle. The first dataset contains 13 different attributes and the second dataset contains the 5 other features that can be used for data manipulation and its analysis.

The very first and the most essential step for any EDA is data cleaning and for that purpose we have we have dropped all the duplicates the non-essential null values from different attributes. Since we have 1645 null values in our rating column and we could not drop that huge number of null values as it impacts our final results. So, we have replaced it by Mode of ratings.

After dropping the null and duplicated values we saw all columns are in different format which will create problem while visualizing. To overcome this problem, we have typecasted all the columns into required format. For ex- Replaced $ from Price, replaced "+" from Installs, replaced "," from Installs and converted it into required format.

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.

After doing some basic analysis and found some useful insights then we have merged both the data frame to find correlation between the columns of both datasets and we got very interesting results.

1. Doing in-depth analysis and after plotting proper visualization graphs we came to some good conclusions that reviews and installs has the positive correlation while price and rating share the negative correlation.
2. Category Art and Design has the maximum number of installs.
3. Developers and Managers should develop apps within **Family** and **Lifestyle** categories can be aimed for more profit i.e high revenue.
4. Almost 61% of people have positive sentiments while approx. 15% reacted negatively which is quite low in comparison (Rest are Neutral).
5. Compared with Free and paid apps, 92.12% apps are Free and 7.81% apps are paid.
6. As **Everyone** content rating contains all age group people, it has maximum i.e 81.80% apps.
7. Maximum number of apps belong to the **Family**, **Game** and **Tools** category.
8. The category **Game** is a potential unsaturated space for all developers, as it has a maximum number of installs.
9. People love to download apps from **Tools**, **Entertainment**, **Education**, **Business** and **medical** genres.
10. Average rating of apps on the play store is 4.17 which is quite good.
11. Users prefer to pay for apps that are light weighted.
12. Paid apps that are higher in size may not perform well in the market.
13. Users tend to download a given app more if it has been reviewed by a large number of people.
14. People tend to review harsher reviews for paid apps.
15. There is a positive correlation between **Installs** and **Rating**.
16. To develop an app which results with high rating needs to get updated with the latest version keeping it optimally sized.
17. It's good to develop a free app and have a content rating for everyone.

The dataset contains many possibilities to improve business values and have a good 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. Got to know many things and process about and how EDA is done.