Capstone Project Submission

Team Member's Name, Email and Contribution:

Saumya Dash	saumyadash9@gmail.com	Data Visualization, Data wrangling, Technical Documentation, Ppt.
Kumar Abhinav	Kumarabhinavthakur274@gmail.com	Data Cleaning, Data Manipulation, Problem Statements, Insights.

Please paste the GitHub Repo link.

Github Link:-
https://github.com/saumyadash9/Play-Store-App-Review-Analysis-Capstone-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)

Android users can download millions of applications from the Play Store for recreational purposes like gaming, watching movies, installing health apps, reading books, conducting business, etc.
We compared thousands of applications from numerous categories for this capstone project.
In order to help app developers work and successfully penetrate the Android market, we have studied the data to identify important elements responsible for app engagement and success.
Two dataset files—"Play store csv" and "User Reviews"—have been sent to us. One file has 13 dataframes, including "App," "Category," "Ratings," "Reviews," "Types," "Installs," "Genres," "Price," "Content Rating," "Last Updates," "Current Version," and "Android Version," and another file has three dataframes, including "App," "Translated Review," "Sentiment" , "Sentiment_Polarity" and"Sentiment_Subjectivity". We performed various data cleaning operation to avoid miscalculation and double counting of data during data analysis. Tasks we performed for this purpose are dropping duplicated instances from both the data frames,
dropping or substituting null values as per requirement,replacing numerical   columns given as 'obj ect' to 'float' values,Conversion of different Units given for a particular column to a common one, dr opor replace wrong data as per requirement .
Then we carried out Data Visualization. We formed problem statements which we used as a basis for our analysis
first we asked which category has maximum and minimum number of apps. For this we did group by operation and found out the relevant data frame and plotted a bar graph for the same. We realized that Family, Game and Tools category has maximum apps and Parenting, Comics and Beauty had minimum number of apps.
We did above analysis for Average downloads in each category as well, there we realized that Communication, Social and Video Players have maximum downloads and Beauty, Events and medical category has minimum downloads.
 Thirdly we asked if there is a gap in average app downloads and availability of apps, for this we drawn double variable bar graph with common dependent variable as Category of apps. Here, we can see that there is a disparity between the number of apps available in the market and their users in each category. For example, when compared to the quantity of apps in the market, the Medical

category has a high average installs.

In the next problem statement we saw that most of the categories have positive sentiments in public except for few categories such as Game.

In the 5th problem statement we tried to compare the Content Ratings of the apps in terms of availability of apps and average downloads in each rating, we realised that Everyone rated apps are widely available but Teen and Mature rated apps have sizeable user base.

Next we saw that most of apps in the play store are free except for family category where there are some paid apps. We also realised 60% of apps runs on android 4.0 and above, other 40% apps requires updating. One of the many categories that is well liked by both customers and service providers is gaming. However, it has been noted that some people have negative opinions about this category. Gaming companies must therefore enhance their offerings and give customers a hassle-free experience.

Some categories, including as communication, social media, video player, and gaming, have a huge market gap due to high user demand, yet there are very few apps to fill this gap. As a result, there is a lot of room for growth in these areas. Users prefer to download apps with teen and mature (17+ and more) ratings, even if the Play Store has the most apps with material that is rated for everyone. Consequently, there is room to build a

company around it. Because there aren't many paid apps in the app store, there is space to build a company using a monthly or annual subscription model. There is a lot of opportunity in this space for ed-tech, gaming, and entertainment.

Finally based on the insights that we draw from our analysis we gave some recommendations on our part which will definitely be helpful in capturing the android market and expanding the business.