Capstone Project Submission

Team Member's Name, Email and Contribution:

Name : Sanjay Jaiswal
Email id : sanjayjaiswal1996@gmail.com

Contribution:

- Introduction to Data
- Data cleaning
- Correlation
- Average review across each category
- Top app category in play store
- Top app genres in play store
- Distribution of app rating
- Percentage of free app Vs paid app in play store
- Percentage of review sentiments
- Polarity Vs Subjectivity
- Which are the apps with highest number of reviews
- Rating Vs Type

Name : Sachin Dubey
Email id : dubey.sachin086@gmail.com

Contribution:

- Data cleaning
- App installed according to each category
- Content rating
- Merging of dataframes on apps
- Distribution of app size
- Top expensive and earning app in play store
- Distribution of sentiment subjectivity
- How many apps are present in each category according to their version(free/paid)

Name : Jayalaxmi Mekap
Email id : youmailjaya.inbox@gmail.com

Contribution:

- Data cleaning
- App installed according to each genres
- Distribution of app price
- What are the top five installed app in any category
- Average of installation of app across the year
- Distribution of sentiment polarity

Please paste the GitHub Repo link.

Github Link :
https://github.com/victorious07/EDA_Capstone_Project

Drive Link :
https://drive.google.com/drive/folders/1GzAFd3EGBaFiwQ65WFFD2iZPNBzVJE7U?usp=sharing

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions.

(200-400 words)

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.

Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.

Explore and analyze the data to discover key factors responsible for app engagement and success.

- Apart from the given columns we added one more column i.e ‘year’ basically we want to calculate the installations of apps across the year.
- In this project we have two csv files. The first one is play store data.csv and second one is user reviews.csv.
- As the first step we perform data wrangling and data cleaning operation after that we remove the duplicate elements from the dataset.
- First we are dealing with a play store data file in which we calculated the average reviews across each category and we also calculated top category and top genres of apps in the given dataset.
- Further we also calculated apps installed according to their category and genres. We observe the maximum number of apps present in google play store comes under Tools, Entertainment and Education Genres but as per the installation and requirement in the market plot, scenario is not the same. Maximum installed apps come under Communication,Productivity and Social Genres.
- We also observe that the percentage of free apps is more than 90% in the given dataset.
- After that we draw a histogram of app rating and app size and observe that most of the user downloaded apps which have rating lies between 4.2 to 4.5 and if we are talking about histogram of app size scenario is the same as most of the apps installed which are small in size.
- Second we are dealing with a user review file in which we draw the histogram of sentiment subjectivity and observe the maximum number of sentiment subjectivity lies between 0.4 to 0.7. From this we can conclude that the maximum number of users give reviews to the applications, according to their experience.
- We also draw the pie chart of review sentiments and observe that the percentage of positive sentiments is near about 64%.