Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Anukriti Shakyawar (shakyawaranukriti@gmail.com) Contribution- 1- Data Cleaning of Play Store Dataset a. Removing duplicate data from dataset. b. Removing null values c. Removing impurities from each columns of Play Store Dataset. 2- Data Visualization of User Review Dataset. a. Number of review type on the basis of categories b. Sentiment Subjectivity and Sentiment Polarity c. Sentiment polarity relation to the type of app d. Content Rating relation with Sentiment Polarity e. Categories Relation with Sentiment Subjectivity f. Word Cloud of Reviews Rajat Chaudhary (rajat.25.chaudhary@gmail.com) Contribution- 1- Data Cleaning of User Review Dataset a. Removing all duplicate data from the dataset b. Removing all null values from the dataset. 2- Data Visualization of Play Store Dataset a. Number of Apps per Category b. Number of apps per category differentiated by type c. Number of installs type-wise according to categories d. Number of installs for each category e. Impact of size on the number of installs f. Distribution of App Rating g. Rating on the basis of Size Raman Kumar (ramank445522@gmail.com) Contributions- 1- Distribution of Application Type 2- Distribution of Size of App 3- Installs per Rating 4- Reviews Senti ment 5- Distribution of Subjectivity 6- Content Rating based on Age Deepmala Srivastava (srvdeepmala@gmail.com) Contributions- 1- Presentation file
Please paste the GitHub Repo link.
Github Link:- https://github.com/Link/to/Repo
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

In this data set we were provided with two data set in which first one is Play Store data set which consists of our basic information like the name of app, categories and so on and the second data set was User Reviews in which we have customers reviews for apps, both data sets are connected with a key column which is App which has the names of all the apps present in data set. The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

There are some basic steps which we performed in our analyses. The first one is Integral Research, in this step we first uploaded the data and then perform basic operations such as info, shape and describe commands to know more about the data such as data type, columns, total number of entries and null values. This type of information gives us information to process to our next step. After that we performed data filtering, in this step we remove all the duplicate rows and the null values from our data set to provide more logical analyses and the third step was data cleaning in this step we cleared all the extraneous data (such as removing signs like $ and +), these steps are performed on the both the data sets. After the data filtering and cleaning we move to our next step which is data visualization for which we used a variety of packages such as Pandas, Numpy, Matplotlib, Seaborn, Word Cloud and Stopword

In this project we analyzed that most of the trending apps are from the categories like family, tools, and games. Most preferred Apps by users in point of size or weight are light size apps. Users also installed apps on the basis of their ratings. These ratings are defined on 2 points- Polarity and Subjectivity. **Basically any app made in the top 5 categories with a smaller size preferably free in type with higher number of review and so be in the positive polarity are more likely to be succeeded**. This analysis will help developers while preparing for their next apps.