Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. **Kishor Kumar (kishor.aug10@gmail.com) :**

   I. Data Wrangling :

   Contributed with the exploration of data set 1 from the play store reviews. Locating and replacing Rating values that are out of range during wrangling.

   II. Analysis and Visualization :

   a). G1 : Category Vs Count Plot to See which category has the MAX number of Apps in the Playstore.
   b). G2 : Installs Per Category to Check which category has the Highest Installs.

   c). G3 : Rating Plot for the highest count of Ratings .

   III. Contributed in finalizing the Conclusion.
   IV. Contributed to the PPT by covering all the points.
   V. Contributed to the technical documentation's content, including the project's problem statement, purpose, and steps to be taken.

2. **Aishwarya K P (kpaishwariya@gmail.com) :**

   I. Data Wrangling :

Contributed in the exploration of data set 2 referred to as user reviews. Here, the sentiment subjectivity and sentiment polarity columns are verified for range and Null values are treated.

II. **Analysis and Visualization :**

a). G1 : Free vs Paid to compare the free vs Paid app

b). G2 : Price Range to compare the amount of free apps vs amount of paid apps.

c). G3 : Content Rating Distribution

III. **Contributed in finalizing the Conclusion.**

IV. **Contributed to the PPT by covering all the points.**

V. **Contributed to the technical documentation's content, including the project's problem statement, purpose, and steps to be taken.**

3. **Akshat Raj Kumawat (akssshat.raj@gmail.com):**

I. **Data Wrangling :**

Helped to check for null values in the "Rating" column and either replace them with another value using a method or remove them altogether. Changed the 'Review' column's data type.

II. **Analysis and Visualization :**

a). G1 : Rating Vs Category to check the highest rating per category.

b). G2 : Reviews Vs Category to check the highest reviews per Category.

c). G3 : Size distribution in Apps

III. **Contributed in finalizing the Conclusion.**

IV. **Contributed to the creation of the PPT.**

V. **Contributed to the technical documentation's content, including the project's problem statement, purpose, and steps to be taken.**

4. **Mourvika Shirode (mourvika95@gmail.com) :**

I. **Data Wrangling :**

Contributed to identifying the data type and null values in the "Size" and "Installation" columns, replaced the existing characters and modified the columns' data types.

II. **Analysis and Visualization :**

a). G1 : Rating vs App Size to check the relation of size of app and rating.

b). G2 : Android Version per No. of user

c). G3 : The top three Android Versions used today

III. **Contributed in finalizing the Conclusion.**
IV. **Contributed to the PPT by covering all the points.**
V. **Contributed to the technical documentation's content, including the project's problem statement, purpose, and steps to be taken.**

5. **Soumyadip Paul (paulsoumyadip99@gmail.com) :**

I. **Data Wrangling :**

Contributed to changing the data type and replacing characters in the "Installation" column. Checked for and dealt with duplicate values and combined the two sets of data.

II. **Analysis and Visualization :**

a). G1 : Sentiment Distribution of Apps

b). G2 : Sentiment Polarity per FREE/PAID Apps

c). G3 : Correlation graph of Merged Data set.

III. **Contributed in finalizing the Conclusion.**
IV. **Contributed to the PPT by covering all the points.**
V. **Contributed to the technical documentation's content, including the project's problem statement, purpose, and steps to be taken.**

Please paste the GitHub Repo link.

Github Link:- https://github.com/paulsoumyadip/eda_on_playstore_review

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.

Problem Statement:

The Google Play Store is the largest app market in the world. It generates more than double the downloads of the Apple App Store but makes only half the money as the Apple Store. Explore and analyze the data to discover the key factors responsible for app engagement and success.

Objective:

The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

Steps involved:

Loading the dataset:

We created a directorial path for the play store dataset, using the Pandas read function we read it. It has a shape (10841,13) which means it has 10841-row labels and 13 features or column labels.After reading it we found which are the dependent variables and which are the independent variables, there is one dependent variable which is the number of installs or just installs, others are independent variables.

Cleaning and Transforming Data:

Cleaning is the process of removing undesired features, values, or any suffix, prefix, or anything which can produce an exception. Transforming is completely a different process, transforming is required to ensure the consistent data type of features because inconsistent data type will generate an obstacle during the execution of the program. These two processes have specific subprocesses as follows.

Unwanted Data Removal:

In this step we ensured to make a data type of feature consistent by removing characteristics from the values of features, to make them usable.

Null values Treatment:

Two features of our data set have many null values, the first one is size and the second one is rating feature, so we filled null values with a mean, median, and mode because we can assume that the size of the application could be an average of sizes available for a particular category and talking about rating, we have taken the median of rating for each category, and then we filled Nan with a mode for categorical data.

Analysis using Visualization Tools:

Now we will be using the clean data sets and visualization tools to generate graphs and charts to gather some insight from the data set which may help us predict what are the key factors that might help in app engagement and success.