Capstone Project Submission

Team Member's Name, Email and Contribution:

1) Akshay Shinde
Email : akshayshinde7276@gmail.com

Contribution:
a) Contributed in operations using Numpy & Pandas library and Data visualisation.
i) Used all Pandas and numpy library functions.
ii) Used all methods of python operations.
iii) Done Technical Documentation.

2) Pratiksha More

Email : morepratiksha7447@gmail.com

Contribution:
a) Done data visualization and perform python functions
i) With the help of groupby.
ii) Used all methods of python operations.
iii) Completed PPT.

2) Somya Gupta

Email :guptasomya84@gmail.com

Contribution :
a) Contributed in operations usings Pandas library and data visualisation.
i) Used all Pandas library functions.
ii) Used all methods of python operations.
iii) Done Summary.

3) Prashant Gour

Email: gourpc05@gmail.com

Contribution:
a) Contributed in operations using Matplotlib , Seaborn and visualization.
i) Used all Matplotlib and Seaborn library functions.
ii) Used all methods of python operations.
iii) Completed Technical Documentation.

/Please paste the GitHub Repo link.

https://github.com/pratikshamor/EDA1_Capston-Project_Play-Store-App-Review-Analysis

Summary of Capstone project and components:

Data science can be summarized into five steps: capture, maintain process, analyses, and communicate. The analysis of Google Play Store applications aided to build more reliable and more interactive applications. This would be very useful for app developers to build an application focused on certain discussed categories in this analysis. This analysis will help in building the application with precise and accurate objectives.
In the initial phase, we focused more on the problem statements and data cleaning, in order to ensure that we give them the best results out of our analysis. Our major challenge was data cleaning, In Data Cleaning, we have performed a few steps to ensure the data quality such as removing NAN values. During the Data Cleaning step we found that 13.60% of reviews were NaN values, and even after merging both the data frames, we could not infer much in order to fill them. Thus, we had to drop them.

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.Our motive in the whole project was to analyse the data and find out main components that affect users' decision to download apps. After completion of analysis I concluded that users prefer more free apps. Most of the apps present in play store are more or less of the same size so size doesn’t affect their decision much.

It was found that Most of the apps that are present on the google play store have ratings in between 4 and 5.Also it was observed that Maximum number of applications present in the dataset are of small size.

In this EDA project, we are provided with different datasets related to play store data :

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

So our agenda is too:

1) Loading the data into data frame.

2) Finding out the basic information about the data.

3) Cleaning the data (Check for null or missing values in the dataset).

4) Extracting statistics from the dataset.

5) Exploratory analysis and visualizations.

- As the first step, we found the basic information about the data , we founded that our first data set playstore.csv contains 10841 rows and 13 columns then we performed data filtering in which we checked for null values in data set and deal with them. We found the types of different datasets as int64, float64 etc. Further, we divided the project into five main parts

  i.e. different category wise and comparison wise.

- Then we counted the no. of applications in each category differentiated by their type and placed a pie chart on it.

- There are columns like reviews, rating, price, size etc. so we gone through them and found overall comparison of category & install (category and count wise) and the count of rating in each category differentiated by customer ratings.

- We implemented one addition in this previous data from play store we provided some benefits to customers rating wise**.** We distributed ten coins to each customer who rated the apps so total no. of coins used are 390803 coins.

- 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

Conclusion and future work:

From the given data we have concluded lots of outputs which make us easy to understand and get the actual calculations of data comparing with category. We have also understand the actual data analysis and how to take insides from the data and prepare understandable work for present.

This data set contains a large amount of data that can be used for various purposes. Currently, the data wrangling and visualizing this data set can be used for future developers and Google play store team to glance at the google play store market and what categories of the apps should be made to keep google play store popular in the future. It can be used to improve business values and google play store in general. It is not just limited to the problem we solved. Using this data set, we applied various libraries. Using this data set the future work includes the prediction of other parameters such as the number of reviews and installs based on the regression model, identifying the categories and statistics of the most installed apps, exploring the size of the app and its version of Android, etc. on the number of installs.