CAPSTONE PROJECT SUBMISSION

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
V Rupesh Kumar Patro : (varnasipatro@gmail.com) Upload dataset to Google Colab · Analyze null values and filter them. · Data cleaning. · Correction of data types · Data wrangling · Data Visualizations · Technical Write up · PowerPoint presentation · Project summary **Shashank Maindola: (shashank.ddun@gmail.com)** · Upload dataset to Google Colab · Analyze null values and filter them. · Data cleaning. · Correction of data types · Data Visualizations · Technical Write up · PowerPoint presentation · Project summary
Please paste the GitHub Repo link.

Github Link:- https://github.com/rupeshpatro2001/Playstore-Data-Analysis

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 Google Play Store is a digital distribution service operated and developed by Google. It serves as the official app store for certified devices running on the Android operating system and its derivatives as well as ChromeOS, allowing the users to browse and download applications developed with the Android software development kit (SDK) and published through Google. Google Play store allows access to download a wide range of apps, books, movies, and television Content either free of charge or at a cost.

For every developer it is important to understand the needs of the user so that the developer can develop, update or modify the application according to the user needs. The application not only improves a user's experience but also provides an opportunity to monetize the app and earn money.

In this project we were provided with two datasets namely Playstore and User_review. The following are the details of the data sets

1. Playstore.csv contains all the details of the applications of Google Play. There are 13 features that describe a given app.

   'App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', 'Android Ver'

2. User_reviews.csv contains reviews of each app. The text data is converted using NLP technique to get three features:

   Sentiments (Positive, Negative, Neutral)

   Sentiment Polarity

   Sentiment Subjectivity

Firstly, we imported the library which were required to process our data, then we mounted the data from the drive link folder. Then we looked for the head and the tail of the data so we can get an insight of the data. After that Missing/Null values were looked for and we corrected using the methods. There were a lot of duplicate values present in our dataset which may have hampered our analysis so we removed them as well. There were visual impurities which needed to be filtered, hence we first filtered those so we can proceed to the next step
It was observed that there were columns within list which were numeric in nature but were considered as object in the dataset hence we converted them to numeric values as well. We merged the user_review and playstore dataset so we can visualize the data easily.
After our data was ready, we plotted different graphs for our observations using the libraries. Our observations were focused around the following:

1. The Top Apps
2. The User Rating
3. Type of App
4. The User Experience-I
5. The User Experience-II (Sentiments)

Based on the above observations we concluded the EDA and summarized our results with which android developer can provide a great user experience through their apps.