Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Member Name: • Shaik Ahmad Basha ahmadshaik982basha@gmail.com Contribution: • Exploring Data • Data Wrangling • Data Cleaning • Checking for Null Values • Checking for Duplicated Values • Analyze Categories with More Apps • Analyze Apps Based On their Updated Years • Analyze Ratings, Installs and Reviews for Different Categories • Analyze Types of Apps • Analyze Average prices and Revenue Generated for Paid Apps • Analyze Sentiments of User Reviews • Analyze Top 10 Genres with most Apps • Performed EDA on Datasets
Please paste the GitHub Repo link.
GitHub Link: - https://github.com/ahmedshaik982/Play_Store_App_Review_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)

Problem Statement:

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. We have two data sets. One is play store data which contains each app (row) . And the other dataset is user reviews.

Our main objective is to explore and analyze these datasets to discover key factors responsible for app engagement and success.

Approaches:

The first step imported all the necessary libraries like NumPy, Pandas etc. and then collected the data. I started with understanding the data like what are the columns and their meanings and data types.

After that, the second step is data preprocessing. Data preprocessing is a process where raw data is converted into clean data. Play Store App data has 10841 rows and 13 columns. User Reviews data has 37427 rows and 6 columns. The two datasets have null values and duplicated values. I removed null values for particular features and created some new features for Exploratory data analysis

And then I moved on to Exploratory data analysis where I analyzed various features to get useful insights from the data. For Visualization, I used bar charts, heatmaps, scatter plots, distribution plots and pie charts.

Conclusions:

Most apps are from FAMILY category followed by GAME category and the least is BEAUTY.
EVENTS category has highest ratings. COMMUNICATION category has highest installs.
GAME category has highest reviews
Most of the apps (66.97%) are updated in 2018. 2018 updated apps have higher ratings, installs and reviews
Most of the apps are free (92.63%) and only 7.37% of apps are paid apps. Free Apps has high installs (99.41%) than paid apps (0.59%).
Most number of ratings are in range of 4 to 5 and most prices of paid apps are in the range of 0 to 50 USD and a very few are in the range of 50 to 450 USD
A very few paid apps have highest revenue. Most of the paid apps have some decent amount of revenue. LIFESTYLE category has highest average revenue.
FINANCE, LIFESTYLE category has high average prices and the least is SOCIAL.
The top 5 revenue apps are Minecraft, I am rich, I Am Rich Premium, Hitman Sniper, Grand Theft Auto: San Andreas.
There is total 115 genres exist in the data. The genre with most number of apps is Tools.
Facebook app has higher reviews. Most of the reviews are positive i.e., 64.04%, 21.29% reviews are Negative, and 14.67% reviews are Neutral.
The most popular app (more positive reviews) is ColorNote Notepad Notes.