Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Nandeesh Umesha - umeshanandeesh@gmail.com
Contributions:

1. Introduction
2. Brainstorming of questions
3. Data Cleaning:
    a. Box plots
    b. Missing values
    c. Outliers
    d. Distributions
4. Data Dictionary
5. Are free apps of poor quality?
6. Are free apps more popular?
7. What type of apps are users ready to pay for?
8. Does the size of the application matter?
9. Correlation map
10. Market Research of android apps
11. Powerpoint presentation
12. Technical Documentation
13. Conclusion

Vilas Sonawane - vilassonawane01@gmail.com :
Contributions:

1. Brainstorming of questions
2. Data Cleaning:
    a. Missing values
    b. Outliers
3. Correlation map
4. Market Research of android apps
5. Sentiment polarity, sentiment subjectivity and popularity of an app
6. Are apps with good ratings more popular?
7. Powerpoint presentation
8. Technical Documentation
9. Conclusion

Vikas Singh - vikassingh.3201@gmail.com :

1. Brainstorming of questions
2. Introduction - Revenue sources

Asad Aslam - aslamasad55@gmail.com :

1. Brainstorming of questions

Please paste the GitHub Repo link.

Github Link:- https://github.com/Nandeesh-U/Playstore_App_Review_Analysis

Drive link:-

https://drive.google.com/drive/folders/1XOrFSFiYAxMMytL7_p9y4pTtgZDL8TWO?usp=sharing

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Google Play Store is a digital distribution service operated and developed by Google. It is an official platform for registered android OS mobile users to download the apps published by various developers. In 2019 alone, 116 Billion apps were downloaded across the world. Consumer spending on Android apps and games increased by 80% in 2019-20. Developers around the world(excluding China) have earned more than $80 Billion with Google Play.

The Play Store apps data has enormous potential to drive app-making businesses to success. In this project, we explore the Playstore apps data set and identify key insights and factors that a developer can utilize to make data-driven decisions.

The dataset included two CSV files. The first one contains information like rating, reviews, installs, price and category for 9660 apps. The other one consists of 64k written user reviews for 1074 apps. This file also included additional data for sentiment, sentiment polarity and sentiment subjectivity of the review calculated from sentiment analysis.

We divided the entire project into three parts - Data cleaning, Exploration and Conclusion. In the data cleaning, part, we started by preparing a detailed data dictionary listing the data type, count or range of the data and definition of the variable. Missing values were identified and handled appropriately. For each numerical variable, box plots and histograms were used to identify the outliers and treat them accordingly. Some variables like installs, size, price were listed in a string format. Appropriate string manipulation techniques were used to convert them into numerical data types.

In the data exploration, we divided the task into sub-questions in a group discussion. For example:
Are free apps more popular than paid apps?
Does the size of the app negatively affect its popularity?
Do paid apps have more customer satisfaction?
What type of apps are users ready to pay for?

We also performed market research of Google Playstore and identified key insights. We used various data visualization techniques like scatter plots, box plots, histograms, treemaps, pie charts, heatmaps and bar charts wherever necessary.

In the conclusion part, we assimilated the answers we found. We identified some counterintuitive insights like “more popular apps have higher sizes on average”. We also proved some commonsensical points like “good customer rating increases an app’s popularity”. The dominance of big tech players in the market was observed in some categories of apps.