Capstone Project Submission

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

Team Member's Name, Email and Contribution:

1)**Sourav Karmakar**
Email: karmakarsourav2024@gmail.com
- Data Cleaning, Sorting, Manipulation.
- Framework establishment and Error Debugging
- Analysis of Best Average Rating of categories and Genres
- Last Update impact, Rating and Average Reviews Analysis
- Overall people's sentiment analysis according to subjectivity, reviews, genres
- Bar plot, reg plot, Histogram, Line plot, PI-plot
- Data melting
- Project Summary

2)**Meenu Benny**
Email: meenubenny555@gmail.com
- Data Visualization
- Bar plot
- Finding Free and Paid Category App
- Data Melting
- Sentiment Analysis according to Rating
- Approach towards Plan
- Technical Documentation

3)**T. Ompriya Subudhi**
Email: ompriyasubudhi2000@gmail.com
- Seaborn
- Matplotlib
- Most Reviews, Rating analysis
- Variation of Reviews in particular month, year analysis
- Bar plot
- pi-plot
- Data Visualization

4) **Thasniya Mansoora**
Email: thasniyamansoora91@gmail.com
- Analyze factors affecting Installation
- Finding Top Content Rating
- Correlation establishment
- Matplotlib
- Seaborn, bar plot
- PPT Presentation

Please paste the GitHub Repo link.

Github Link:- https://github.com/souravkarma/playstore-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)

Data has been provided by AlmaBetter .
From these analyses, clients can understand the needs of their customers in different aspects of product development and it will help them to solve the problems with the market superiors.

 The main objective of this project is to deal with the data provided by the company and to analyze the data in every possible aspect. It will help companies to understand consumer behaviors, their buying patterns and give more clarity to developers regarding future important decisions.

Our Approaches

  At first We are following four steps for cleaning our data that is:-
1)Treatment of null values
2)Treatment of Duplicate Rows
3) Conversion of necessary column values from string to required datatype
4)Replacing the missing values with mean, median or mode.

After cleaning the data we simply analyze our first data frame playstore data csv and solve the problems of top average rated category,genres,most reviewed, installed apps etc.

For sentiment analysis we merge two dataframe play store data and user review data. It helps us to find sentiment subjectivity, Top review sentiments, Genre's sentiment, correlation between various columns and many more things.

Conclusion

After doing this project we came to many interesting conclusions. First, we have done the removal of null values from rows and columns and the same goes with the removal of duplicates from the datasets. Then we did the formatting for each of the required columns in each dataset.

Apps with the category Family and the genre tools are in large numbers. But interestingly Category EDUCATION and Genres 'Board;Pretend Play' and 'Comics;Creativity' both get the best average rating in the playstore.

App Rating is directly proportional with the recent update. From this we can be sure that with the latest update the reviewers are giving better ratings.

We also found that Ratings are directly proportional with Average Reviews till 4.5. But after that we observe some deviation and eventually reviews decrease after 4.5 ratings. 4.5 Rating apps have a maximum 1.1 million reviews.

We found that most of the opinion on sentiment subjectivity lies high in the range 0.4 to 0.7. It means people give reviews after some experience.

We also found people's sentiment are more towards paid app less towards free app

GAME category has maximum no of translated reviews but, most positive percentage acquired by HEALTH_AND_FITNESS category and GAME has most negative reviews among top 5 categories. So, it concludes that most reviews do not suggest the most positive sentiment percentage.

Action Genres got the most strong positive sentiment among all genres. It suggests that people prefer Action genres apps far more than any other genres.

google play store has more free apps than paid. and Family category has most free and paid apps in playstore. MEDICAL category has highest number of paid apps considering quantity.

Social,entertainment,event,education,comic* These popular category apps are always free.

We found that in terms of both Top Total and Average Apps Reviews outcomes Facebook,Instagram,whatsapp messenger,Clash of Clans these apps are common. so we can conclude that these 4 apps are most popular and trending among masses.

Subway Surfers app has the highest rating along with highest reviews followed by Instagram,clash of clans,Facebook,whatsapp messenger,clean master.

We got the highest reviewed apps are those which were last updated in january,2018 and got lowest review in February and March update. It concludes that Most of the companies are majorly updating their apps in January month, so in those months they get most reviews.

We also found that 4.4 rated apps are the most reviewed apps in google play store and from 2.5 to 3.1 there are almost equal numbers of positive, negative, neutral reviews. It states that neither top rating nor less rating makes an app good or bad.

We found that GAME,COMMUNICATION,SOCIAL these category apps got the most installation(above 1 crore).

Most heavy apps got 10cr installations, among top 10 heavy apps 4 apps got above 1 crore installations so, it concludes that customers still preferred heavy apps.

Most Installed, Rated, reviewed positive sentiment apps are always in range between 4 and 4.5 rating.

Most of the apps which are used by the users have a content rating of 'Everyone'.

From install,rating,reviews correlation graph we also found install and reviews are positively correlated means when Installation is more reviews are more.