Play Store App Reviews Analysis

Team Members
Saurabh Funde
Kalyani Motkari
Sameer Ambekar

Abstract:

Play Store and formerly Android Market, is a digital distribution service operated by Google. Play store is a marketplace (App) for downloading Android applications for smart android devices. Google Play Store has published 2.6 million apps till Dec 2021. Smartphones sells soaring across world. It's a growing market for App developers to build applications for smart devices. By performing EDA on given dataset's, it gives important information for App developers for developing Android applications for users.

1. Introduction

The mobile app industry has been active for over a decade in worldwide. Google Play's largest markets are India, South-east Asia and South America and many other countries. Annual Google Play app revenue for year 2021 was $47.9 billion dollars. Games remain the primary form of revenue generation for Google Play store. Gaming category contributes large part of revenue for Google. Our goal is to perform EDA on given two dataset's and analyze the data for app developments. Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

2. Problem Statement

Smart devices sell soaring across world such as Smartphones, Smart wearables and etc. It's a big market for app developers. It's important to find out what type of apps people downloading before developing an App for users and list that App on Play Store for downloading. It's important to explore and analyze the Play store data to discover key factors responsible for app engagement and success like app size, type and other various factors

Following are the two datasets for EDA.

1. Play Store App Data: This dataset contains total 13 columns. The names of columns are App, Size, Price, Reviews, Category, Rating, Content Ratings etc.
2. User Reviews: This dataset contains 5 columns. The names are App, Translated Reviews, Sentiment, Sentiment Polarity and Sentiment Subjectivity.

3. EDA Methodology

Our Exploratory data analysis approach is divided into four phases: datasets loading, data exploration, data wrangling and data visualization's.

4. EDA Approach and Challenges

Using Python libraries both datasets loaded for EDA. After initial data exploration of both datasets, we found that there is a need of some data cleaning like changing data types of some columns, finding out NaN values in each column, duplicated values in columns and special characters etc. We wrangled the required columns for EDA.

Play store app dataset has 13 columns and 10841 rows. Rating is the only column which has maximum NaN values. App column has 9659 unique App names, 33 unique categories. There were 1181 duplicated app entries in App column. All duplicated App names removed and only latest entry allowed in dataset. Rating columns missing values filled with median of same column.

User reviews dataset has 5 columns. This dataset contains translated reviews, sentiment of apps etc. There are three unique sentiment entries in Sentiment column Positive, Negative and Neutral.

Play store data and User reviews datasets merged on App column. This dataset contains 816 unique app names in app column. By doing EDA on this merged dataset, we got to know about sentiment by category wise and if there is an any correlation between app Size, Type and Reviews on sentiment of app.

5. Data Visualizations

After doing data wrangling, we plotted various plots using Python viz. libraries such as Seaborn and Matplotlib for the purpose of data interpretation. Few of them plots are added in this documentation.

The below pie chart shows about the Type of apps from Play Store Dataset. 92.19% Apps are Free type and only 7.81% are Paid in type for downloading.
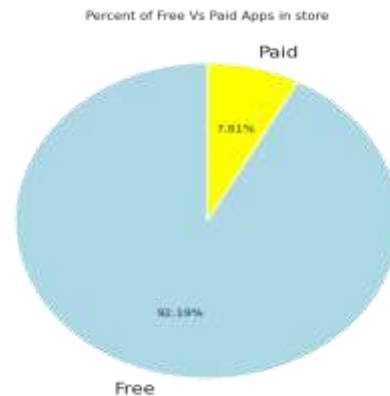
Fig: 5.1 Type of Apps

The category wise app count plot shows that the Family, Game, Tools, Business and Medical are top categories for developing an app. Lifestyle, Finance, Sports and various other are emerging categories for an app success on Play store.
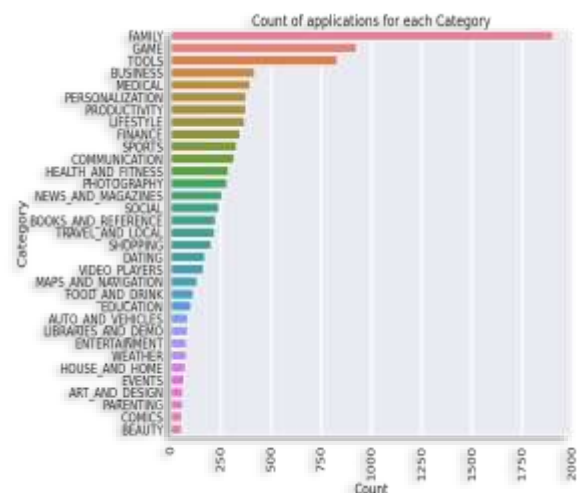
Fig. 5.2 Category wise Count of Apps

This following scatterplot shows that most of apps are having Rating more than 3.5 and size of Apps less than 50 MB. The mean rating of Free and Paid type apps is 4.18 and 4.26 respectively.

Fig. 5.3 Size Vs Rating Viz.

This is correlation matrix of a Merged dataset. We found no significant relation between Rating, Reviews, Size, Installs and Price with respect to the sentiment polarity and sentiment subjectivity. There is some relation between two columns Reviews and Installs.
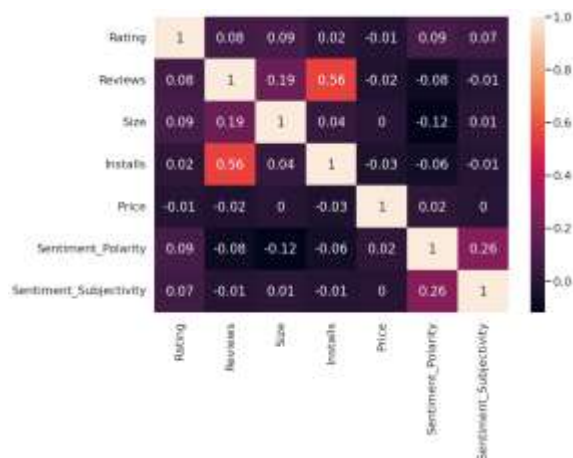

Fig. 5.4 Correlation Matrix of Merged Dataset

6. Conclusion

The following are the conclusion we can draw from this EDA.

- 92.19% apps are Free in type. It's important factor for a success an app.

- Size of app should be minimum. In Play store dataset we found that 8700+ apps are having size less than 50Mb.

- 81% Apps are having content rating as Everyone. It's also important factor for a success of an app.

- There are many other emerging categories for App development.

7. References

www.towardsdatascience.com
www.medium.com