Play Store App Review Analysis EDA- Capstone

Mohammad Ammaar, Mohammad Shahzeb Khan, Nada Nasser, Gayatri Gupta, Mansur Shikalgar

Abstract:

Android is the dominant mobile operating system today more than 70% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store. We focus on analysing Google Play Store, the largest Android app store that provides features (ratings, reviews, type, ,install and number of downloads etc.). The overall objective of this analysis effort is to provide in-depth insight about play store review analysis in general.

1.Introduction

By this time, Google's ecosystem had expanded greatly compared to the humble beginnings of the Android Market. In fact, the Android Market was just one of the company's online markets. At the time, this was the only place that Google had to sell goods.

Analysing customer needs is one of the bizarre tasks in the business world today. Hence proposing analyse data to developer that what customer is likely to download, which category got the maximum downloads this all plays a crucial role in app development. Generally, customers download apps depending on number of downloads, positive reviews, negative reviews, ratings and comments. So, in this project we are going to help the users by categorizing positive, negative and neutral reviews and comments of the particular. we are going to help developer by analyzing the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization & developers. Also provide an idea about app that managed to get maximum and minimum number of downloads and predicting the category of apps that is most likely to be downloaded in the coming years. The dataset of google Play Store for analyzing is collected from dataset.

The purpose of our project is to gather and analyze detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.The Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.

2.Problem Denotation:

The Play Store apps data has enormous potential to drive app-making businesses to success. Android is expanding as an operating system and Mobile app industry is increasing in significantly and thus giving rise to more competitions to the one's that are creating applications .Due to the competition in the market and also expansion in order to help our developer understand what kinds of apps are likely to

attract more users and what is the motivating factor for the people to download an app we analyze and research relevant data. For the app development in dustry where they can analyse the downloads and demand off app download in the industry.

The Problem statements are :

1) Finding top apps in Playstore as per Category
2) Checking the proportion of Free and Paid Apps
3) Now checking the distribution of apps in terms of Size, Rating & Type
4) How importance is the rating of the application?
5) Which categories from the audience should the app be based on?
6) Which category has the most no. of installations?
7) How does the count of apps varies by Genres?
8) Let us examine the Free and Paid Apps available according to Category

3.Details of Columns on given Data set

There are two dataset:

Play Store Data Contains:

1. **App-** The app name
2. **Category-** Categorical label, which describes which broad category the app belongs to.
3. **Rating-** Continuous variable with a range from 0.0 to 5.0, which describes the average rating the app has received from the users.
4. **Reviews-** Continuous variable describing the number of reviews that the app received.
5. **Size-** The size of the app. The suffix M is used for megabytes, while the suffix K is used for kilobytes.
6. **Installs-** Categorical label that describes the number of installs.
7. **Type-** Label that indicates whether the app is free or paid.
8. **Price-** The price value for the paid apps.
9. **Content-** Rating Categorical rating that indicates the age group for which the app is suitable.
10. **Genre-** Semicolon separated list of genres to which the app belongs.
11. **Last Update-** The date the app was last updated.
12. **Current Version-** The current version of the app as specified by the developers.
13. **Android Version-** The Android operating system the app is compatible with

User Review Data(App, Sentiment ,Sentiment Polarity, Sentiment Subjectivity)

By diagnosing the data frame, we know that:

- There are 13 columns of properties with 10841 rows of data.

- Column 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object'

- Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'.

- Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'.

- Values of column 'Price' are strings representing price with symbol '$'

- Checking the duplicate data from Dataset and checking for NaN values across different columns.

.

4.Rating distributaion

From this distribution plotting, it implies that most of the apps in the Play Store are having rating higher than 4 or in the range of 4 to 4.6.
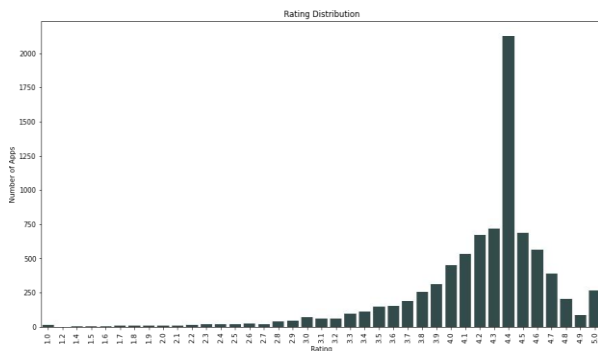

Fig 1: Pi chart of customer review .

5.Different type of app on base of price(free/paid):

Here we see what is the percentage off free app and paid app in play store and from all app 92% app are free app and only 8% app are paid app.
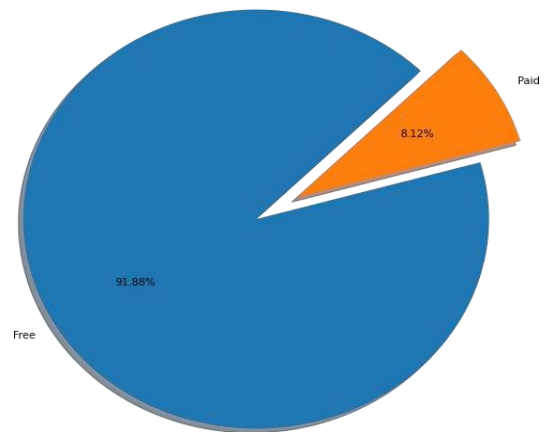

Fig 2: free vs paid app

6.Rating given by user from each category:

Here we see that Performance of all app categories is mostly Decent. We can say that most the apps with higher rating range of 4-4.7 are having high amount of reviews, size, and installs. In terms of price, it doesn't reflect a direct relationship with rating, as could see a fluctuation even at the range of high rating
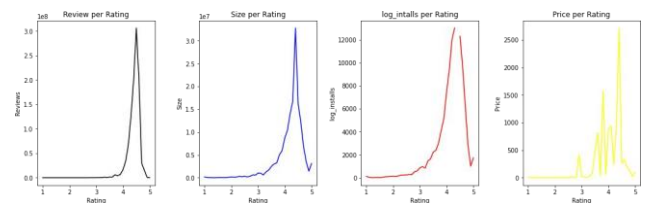

Fig 3: categories vs rating

7.Sentiment Analysis :

Here we conduct the graph which show type sentiment analysis from each category.

This graph help us to know which category get highest sentiment review as well as compare of +ve and -ve review .

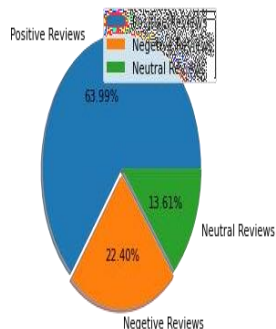A Pie Chart Representing Review Sentiments in Percentage

Fig 4: sentiments analysis pie chart

Distribution of Rating with Size:

Here we plot scatter plot, we can imply that majority of free apps are small in size and having high rating. While paid apps have quite equal distribution in terms of size and rating.

Fig 5: Scatter plot of Size with Rating

9.Conclusion :

The Google Play Store Apps report provides some useful details regarding the trending of the apps in the play store. As per the graphs visualizations shown above, most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps

from these categories are twice as much lesser than the category FAMILY but still used most. The trending of these apps are most probably due to their nature of being able to entertain or assist the user. Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps.

Other than that, the charts shown above actually implies that most of the apps having good ratings of above 4.0 are mostly confirmed to have high amount of reviews and user installs. The size and price shouldn't reflect that apps with high rating are mostly big in size and pricy as by looking at the graphs they are most probably are due to some minority. Furthermore, most of the apps that are having high amount of reviews are from the categories of SOCIAL, COMMUNICATION and GAME like Facebook, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free, Clash of Clans ,google apps etc.

Even though apps from the categories like GAME, SOCIAL, COMMUNICATION and TOOL of having the highest amount of installs, rating and reviews are reflecting the current trend of Android users, they are not even appearing as category in the top 5 most expensive apps in the store which are mostly from FINANCE and LIFESTYLE)

.As a conclusion, we learn that the current trend in the Android market are mostly from these categories which either assisting, communicating or entertaining apps.

Some important point:-

- Average rating of (active) apps on Google Play Store is 4.6.

- If we see individually app wise the communication app like Facebook and whats-up get highly reviewed .

- Medical and Family apps are the most expensive and even extend upto 80$.

- Users tend to download a given app more if it has been reviewed by a large number of people.

10. **References**:

1. Researchgate.net

2. GeeksForGeeks

3. jovian.com

4. Stackoverflow

5. github.com