

Play Store App Review Analysis

Sameer Satpute
Data science trainee,
AlmaBetter, Bangalore

Abstract:

Mobile app distribution platform such as Google play store gets flooded with so many thousands of new apps everyday with many more thousands of developers working independently or in a team to make them successful. With immense competition from all over the globe, it is most important for a developer to know whether he is going in the right direction. Since most Play Store apps are free, the revenue model is quite unknown and unavailable as to how the in-app purchases, in-app adverts and subscriptions leads to the success of an app. Thus, an app's success is determined by the number of installs and the user ratings that it has received over its lifetime rather than the revenue it generated. In this project, on a smaller scale, i have tried to perform exploratory data analysis to go deeper into the Play Store data that i collected, discovering relationships with specific features such as how the number of words in an app name for instance, affect installs, in order to use them to find out which apps are more likely to succeed. Using that extracted features and the recent sentiment of users we have predicted the success of an app soon after it is launched into the Google Play Store.

Keywords: *Google-play store, mobile app*

1.Problem Statement

The expansion of smart phones is driving the fast development of mobile app stores. Currently, the two largest global platforms for app distribution are Apple's App Store and Google play store. I have picked Google play store and did a thorough analysis of its features that were available to me for predicting the success of a particular app. But the question arises that, is it necessary to do that? Well, the answer is yes, because an average of 5000+ mobile apps is released through the Google Play Store every day, according to the statistics shown by the Statist in their recent report where all the app in together compete for user attention.

The huge number of apps in the play store and the numbers of the app released every day make it competitive for the app developers, the companies who are building an app to come up with a unique idea that will definitely be bought by the end users. Because at the end of the day if the app does not perform well in the android market, then all the hard work behind building the app will go in vain. As the mobile industry is growing rapidly it is increasing the level of competition however, increased competition also leads to increased chances of failure. So, the developers need to do enough research as an enormous amount of time, effort and the money are invested into the process, so business cannot afford an app failure.

Success is a thing that does not come to one so easily and for that, I analysed the features of Google play store and came to a conclusion that will help developers to understand their app success rate using our proposed success parameters.

2. Introduction

It has been observed that the significant growth of the mobile application market has a great impact on digital technology. Having said that, with the ever-growing mobile app market there is also a notable rise of mobile app developers, eventually resulting in sky-highest revenue by the global mobile app industry. With immense competition from all over the globe, it is important for a developer to know that he is proceeding in the correct direction. To maintain this revenue and their place in the market the app developers have to find a way to stick into their current position. The Google Play Store is found the largest app market in the world. It has been observed that although it generates more than double the downloads than the Apple App Store but makes only half the money compared to the App Store. So, we scraped data from the Play Store to conduct our research on it.

For the EDA part, we analysed our dataset and created several Bar Charts to visualize the relationship between each attribute. Then we created a heat map of installs with other numeric features from there we could conclude that numeric features such as the number of ratings as well as subjectivity are all uncorrelated with installs. Additionally, we did an analysis on the features App Name, Content Rating, and Type and

reached some valid conclusion. From those Bar Charts, we observed the variance of the first 25 most frequently used word associated with each rating ranging from 1 through 5. From the EDA performed, we observed the importance of each feature and the correlations between each of them. Hence, from the analysis, it can be seen that determining the success rate of an app will play a very important role for developers and can bring certain changes that might affect the lifetime of their app.

2.1 Analysis

In today's scenario we can see that mobile apps playing an important role in any individual's life. It has been seen that the development of the mobile application advertise has an incredible effect on advanced innovation. Having said that, with the consistently developing versatile application showcase there is additionally an eminent ascent of portable application designers inevitably bringing about high as can be income by the worldwide portable application industry. may need to figure out how to stick into their present position. The Google Play Store is observed to be the biggest application platform. It has been seen that in spite of the fact that it creates more than two fold the downloads than the Apple App Store yet makes just a large portion of the cash contrasted with the App Store. In this way, I scratched information from the Play Store to direct our examination on it.

Every company main objective is to provide free app for getting more Installs but at same

time inside app some features are paid so they get there revenue from that. Apps like Medical, finance, lifestyle are most of paid app.

2.2 Google Play store Dataset

The dataset consists of Google play store application and is taken from Kaggle, which is the world's largest community for data scientists to explore, analyse and share data.

This dataset is information of play store application to analyse the market of android. Here it is a downloaded dataset which a user can use to see the Android market of different use of classifications music, camera etc. With this, client can predict see whether any given application will get lower or higher rating level. This dataset can be more over used for future references for the proposal of any application. Additionally, the disconnected dataset is picked so as to choose the estimate exactly as online data gets revived all around a great part of the time. With the assistance of this dataset I will examine various qualities like rating, free or paid and so for utilizing Hive and after that I will likewise do forecast of various traits like client surveys, rating etc.

2.3 Python

Most of the info scientist use python due to the good built-in library functions and therefore the decent community. Python now has 70,000 libraries. Python is simplest programing language to select up compared to other language. That's the most reason data scientists use python more often, for machine learning and data processing data analyst want to use some language which is straightforward to use. That's one among the most reasons to use python. Specifically, for data scientist the foremost popular data inbuilt open source library is named panda. As we've seen earlier in our previous assignment once we got to plot scatterplot, heat maps, graphs, 3-dimensional data python built-in library comes very helpful

choose to wait a few minutes to see if the rates go back down.

3. Steps involved:

- **Exploratory Data Analysis**

After loading the dataset I performed this method by comparing our target variable that is Pricing_Type with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

- **Null values Treatment**

In dataset contains a large number of null values also I see that is

unnecessary and change my output accuracy hence I moved with dropping such things in the beginning of project to obtain better result.

- **Establish basic observation**

From dataset wrangling I found some basic observation such top category with highest ratings, which app has maximum reviews, total paid and free apps.

- **Data mining**

Data mining is process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. From dataset I derive some meaningful insights such as content rating of app, price category, bulky vs light app, category count of app so that developer easily get idea what type app we need to make for getting popular on play store.

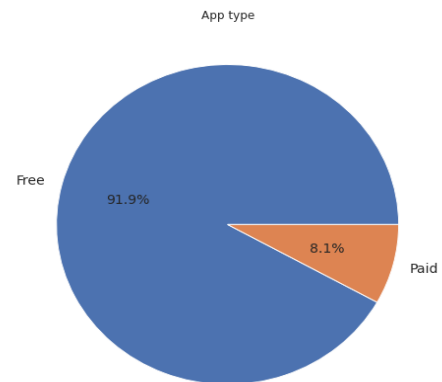
- **Sentiment analysis**

Sentiment analysis is a technique through which you can analyse a piece of text to determine the sentiment behind it. It combines machine learning and natural language processing (NLP) to achieve this. Using basic Sentiment analysis, a program can understand whether the sentiment behind a piece of text is positive, negative, or neutral.

Using review dataset I analyse free app review and paid app review that is negative, positive and neutral. i also used wordcloud to know which word mostly use in reviews.

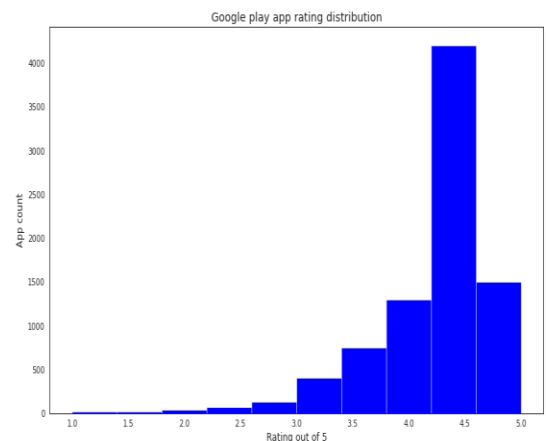
3.1 EDA

3.1.1 Free vs Paid



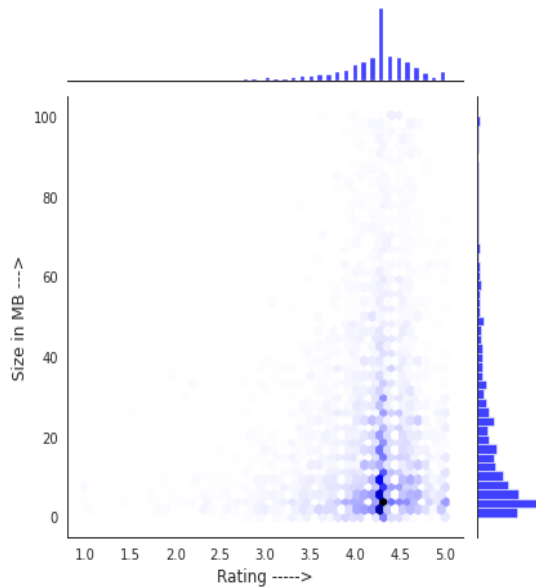
Here we can see that 91.9% apps are free and 8.1% apps are paid on Google Play Store, so we can say that Most of the apps are free on Google Play Store.

3.1.2 Ratings of apps



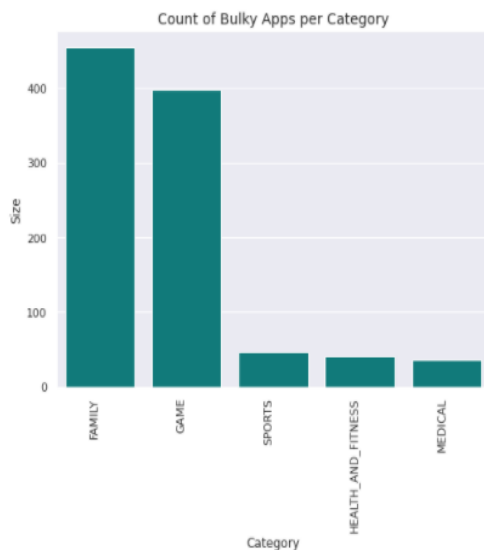
Rating of apps from above graph clearly see that distributed mostly above 4.1.

3.1.3 Light vs Bulky



From above graph we can conclude that most of app rating above 4.0 and has size around 0.1 MB to 40 MB.

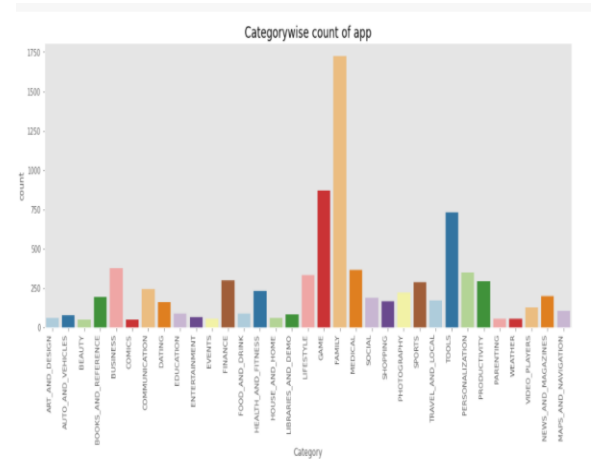
3.1.4 Bulky Apps distribution



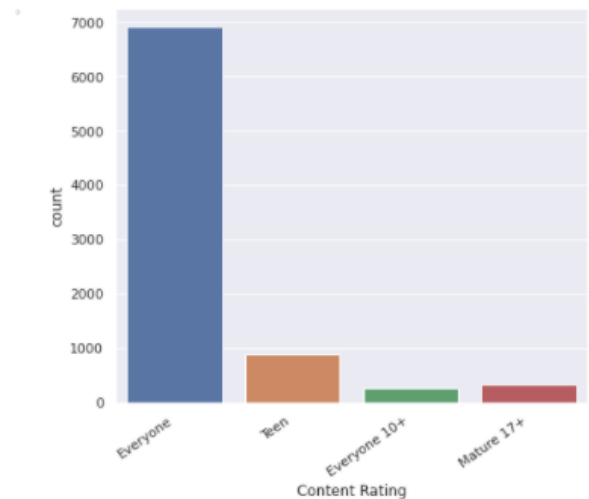
We can clearly see in above graph that family category apps are more bulky.

3.1.5 Category count of app

From the below chart we can find that most of the apps which are on Google Play Store belong to Family, Gaming and Tools.



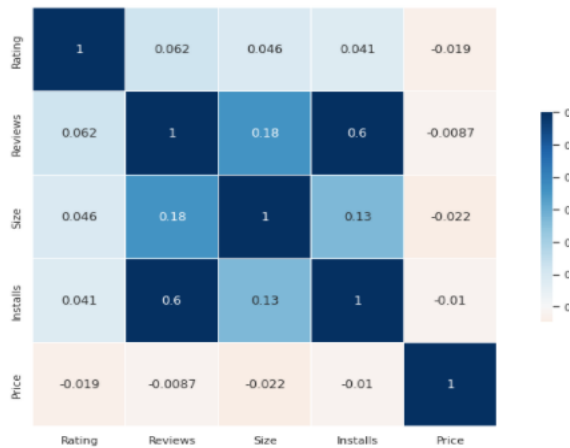
3.1.6 Content rating(count)



This shows almost all apps present on play store is made for everyone means there is no restriction to use based on age but some apps made based on category as above shows.

3.1.6 Correlation

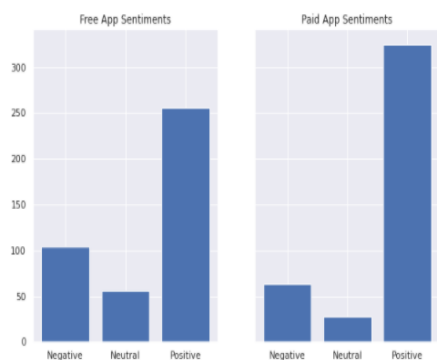
Positive correlation of 0.6 exists between the number of reviews and number of downloads



3.2 Sentiment analysis

I have plotted the positive, negative and neutral reviews for each category and observed that the Health and Fitness apps perform the best with more than 85% positive reviews. On the other hand, Game and Social apps have a higher fraction of negative reviews. We compared the reviews between free and paid apps and found that people are harsher towards free apps whereas users are more tolerant when they are paying for it.

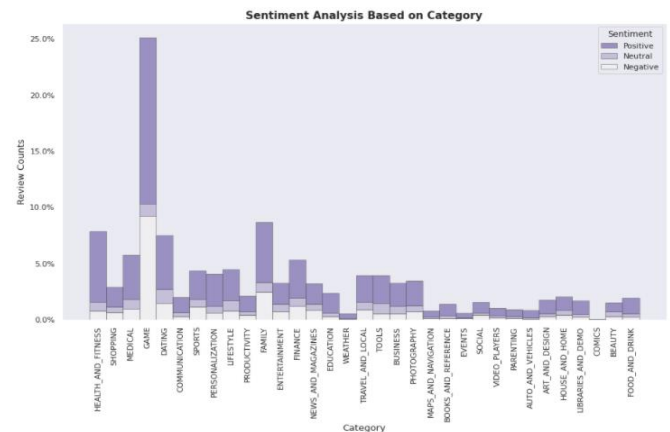
3.2.1 Free vs Paid



In Free apps as we can see negative as well as neutral review so high variance. paid app has more positive review very less neutral

and negative reviews as compared to positive reviews

3.2.2 Based on category

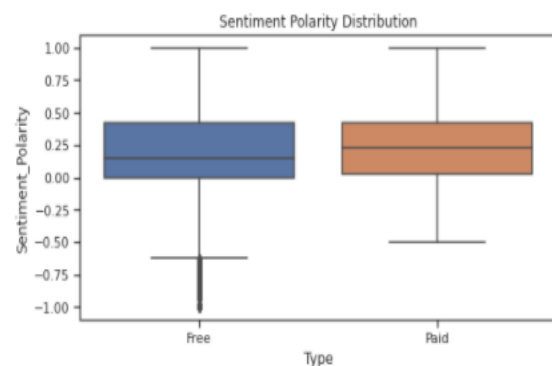


Family, Sports and Health & Fitness apps perform the best, Having more positive reviews

On the contrary, many Game and Social apps perform decent leading to 50% positive and 50% negative.

3.2.3 Sentiment polarity

Sentiment polarity for an element defines the **orientation of the expressed sentiment**, it determines if the text expresses the positive, negative or neutral sentiment of the user about the entity in consideration.



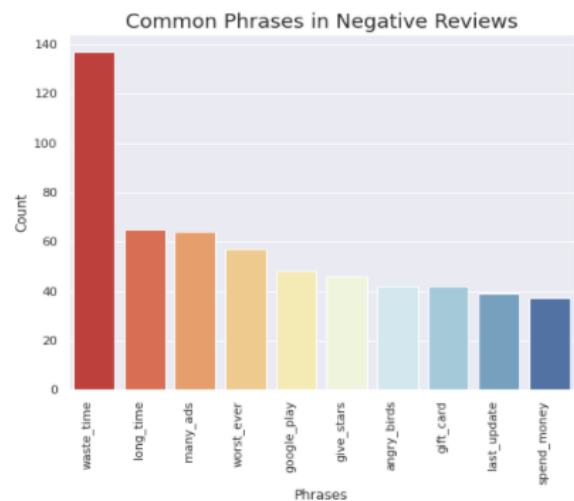
Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analysing data from social network websites.

2) Wordcloud for paid app



Phrase	Count
many_ads	245
free_version	108
much_better	106
best_ever	104
user_friendly	96
works_great	95
highly_recommend	93
first_time	88
wish_could	88
way_many	86

2) Negative Review



For Negative review mostly use phrases are waste time, long time, worst ever.

4. Data Processing

The dataset collected from the Play store is semi structured or unstructured and contains significant superfluous data (defined as not contributing significant meaning).Some data type needs to change in required format as int, float, date. Sizing of apps needs to convert in one measurement KB or MB. Pre-processing includes various tasks including stemming, lowercase conversion, Units, punctuation, and excluding terms.

4.1 Software

I Use Google Collab Notebook for Python coding along with matplotlib and seaborn libraries for visualization.

5. Conclusion

- **Average rating** we see of apps on Google Play Store is **4.18**
- Most of user like light app and they pay for it also. Thus, paid app that is bulky may not do well in the market that's reason for getting good installs.
- Most of the **top rated apps are optimally sized between ~0.1MB to ~40MB - neither too light nor too heavy.**

- **Medical and Family** apps are the most expensive and even extend upto 80\$.
- Users download a given app more if it has been reviewed by a more number of people.
- Paid apps have a slightly higher number of favourable reviews than free apps.
- Free apps get more negative and neutral feedback, suggesting a wider range of opinions.
- **Clash of Clans** app has most number of reviews. While **Subway Surfers** is most number of install app
- When it comes to free apps, users are more pessimistic and harsh than when it comes to paid apps.
- More than half users rate **Family, Sports and Health & Fitness apps** positively. Apps for games and social media get mixed reviews, with 50 percent positive and 50 percent negative responses.
- Many ads ,free version such phrases most commonly see in reviews.
- Worst ever, long time waste time negative review that factors taken into consideration for improving or develop new app

References-

1. Stackoverflow
2. GeeksforGeeks
3. Jovian
4. Research paper based on play store analysis.