Play Store App Review Analysis

Asif Ansari, Zubair Ahmad,
MD Ejaz Ansari
Data science trainees,
AlmaBetter, Bangalore

Abstract:

The google play store is one of the largest and most popular Android App stores. It has an ample amount of data that can be used to make an exquisite model. We have used raw data from diffrent websites. This data set contains many different features that can be used for predicting or analysing whether an app will be a successful one and not using any different features. This data is agonized from the Google Play Store. This almanac gives information about different classifier models that are used for prediction purposes and classifying which one has the excellent accuracy. This almanac also gives detailed information on feature extraction and the complete Data visualization done on this data set.

From this EDA project we have try to solve problem which is related to data science. By the extrect important information which is quite useful for us and there aim is convert it into visual form from the excel documents and to follow the step After cleaning the data we

have plotted various graphs using different python libraries such as matplotlib and seaborn that show relation between different attributes and we have successfully gained some good insights with it that can help the industries to capture the android market so that approach and analyzing data sets to summarize their main characteristics, often with visual methods.

Keywords: *ample, agonized, almanac, EDA, matplotlib, seaborn*

1.Introduction

Mobile applications are the fastest growing fragments of downloadable software application market. In this market Google Play Store is the most popular and fast growing platform. One of the major reason for its popularity is that it has about 82% applications which are totally free of cost. The Google Play Store had over 8.5 lakh applications and over 2.25 lakh android developers in mid of 2013.Now this platform have around 40 billion applications and over 500 million users around the world are installing apps from this platform. Android developers and users play a vital role in figuring the impaction that market interactions have on the upcoming future technology. Whereas, the lack of knowledge of working of applications and its vigorous of popular app markets are impacting both the developers and the users in this article, we will through the light on the vigorous of Google Play Store and how we can use different features from this data set for the future prediction purposes. We will provide a brief study of Google Play Store dataset which will give unique information. Our Analysis is divided into four phases mainly data extraction, data cleaning, data visualization and implementing different models.

Android is expanding as an Operating System.

It has taken around 70% of the world's market which is a genuine indicator of the ample amount of its popularity and the number of people using android in today's scenario. Our aim is to provide the accurate data to android developers to know them what exactly people wants in present time. It will also help them to find the factors that are affecting peoples to download an applications.
We analysed accoprding to categories like reviews, price, ratings, size and installs for this purpose and will find out how they are co-related or related to each other.
So 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 for example some people have motivation factor as

1.Low price,

2.Small size of apps,

3.High user reviews and rating

Analysis Methodology

Our Analysis is divided into following Phases:

Data extraction

Data cleaning

Data visualization

Interpret results

Conclusion

First, we collect the data from the given data set, in the form of a csv file of both the data set i.e play store data and review data. In the next step, we try to do data cleaning on the data set to reduce the error percentage in the data set which includes null values and duplicate values. After the data set is ready, we try to analyze the data set using different plots and remove the stuff not needed from the data set. The last step includes the use of different classification algorithms and visualization of models on the data set to see which one gives the highest percentage of accuracy. Finally, we narrate the analysis results to provide a clear vision of the relationship among the areas of interest. At the end we include a detailed discussion of the applicability and future research directions called Conclusion and future work.

Google Play Store Dataset

1. **App-** It tells us about the name of the application.

2. **Category -** It tells us about the category to which an application belongs.

3. **Rating-** It tells us about the ratings given by the users for a specific application.

4. **Reviews -** It tells us about the total number of users who have given a review for the application.

5. **Size-** It tells us about the size of the application on the mobile phone.

6. **Installs -** It tells us about the total number of installs/downloads for an application.

7. **Type -** It tells us whether the application is free or a paid one.

8. **Price-** It tells us about the price of the application.

8. **Content_Rating-** It tells us about the target audience for the application.

10. **Genres -** It tells us about the various other categories to which an application can belong.

11. **Last_Updated -** It tells us about when the application was updated.

12. **Current_Ver -** It tells us about the current version of the application.

13. **Android_Ver -** It tells us about the android version which can support the application on its platform.

Problem Definition

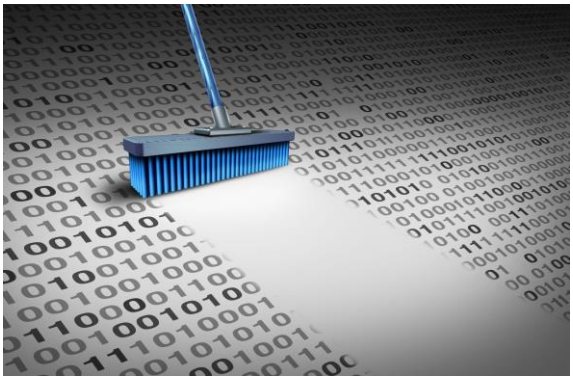
The Play Store apps data has high potential to drive the app-making businesses to make success. Android is expanding as an operating system exponentially and Mobile app industry is increasing significantly and thus giving rise to more competitions to the ones that are creating applications and which are new to this industry this is because of high competition in the market.

Problem Statements

- Find out the relationship between paid app install and free app install and count the number
- Find out the top 10 application and their view.
- Find out the list of top category of application which contain the highest number of application
- Find the average rating of application
- Which category of application have most number of installation.
- Find out top 10 installed application in any category.
- What are the count of apps in different genres?
- What percentage of free and paid application?
- Find out the percentage of positive negative and neutral rating.

2. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. This is because of when we combine multiple data sources of data, there is opportunities for data to duplicate or mislabeled.



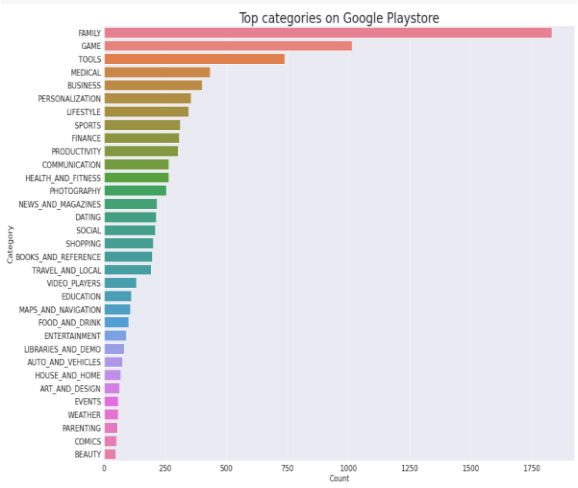
3. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations. It is the veryfirst and most important step in every Data Science project. After cleaning the raw data EDA involves the process of discovering the patterns which may give us good insights and that can help in future challenges. EDA is the process of analyzing the dataset to discover trends, and outliers, and form hypotheses based on our understanding of the dataset.

It involves creating statistics for numerical data in the dataset, finding correlation between numerical values and creating various graphical representations, plots by using visual methods to understand the data in a short and better way. In this article, we will understand EDA with the help of a Google play store dataset. We will use python and its essential libraries and will try to generate some good insights with it.

3.1 Top Category Application

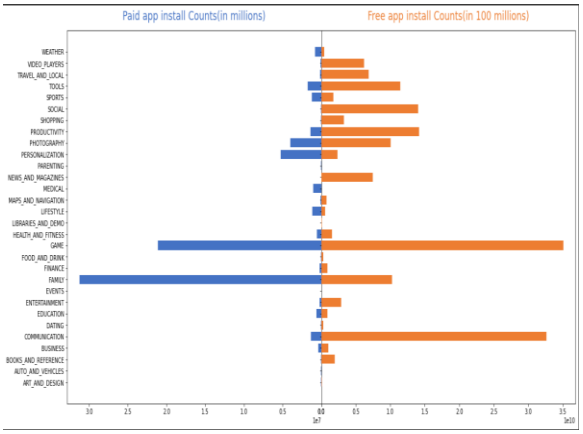
As we know all the application are belongs to different category on the play store on the basis of which we found that Family, Game , Tools are the top category have highest number of application.



3.2 List of free and paid app Install count

From the following graph we can see the list of free and paid application install count which shows that the which category of application had

installed count while comparing free and paid.



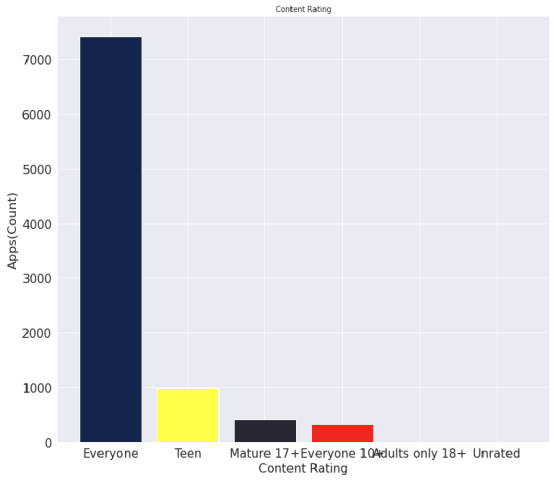
So we can see that Family, Game have highest paid user and Business, Game have highest free application user.

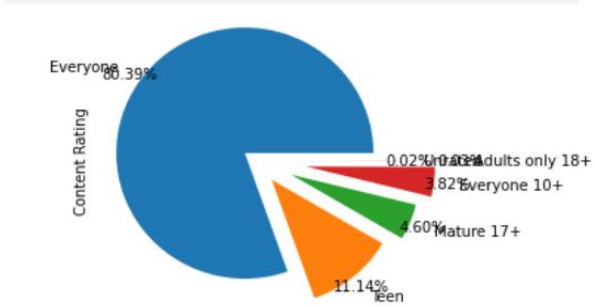
3.3 Content rating Vs Number of application

Content ratings in play store are used to describe the minimum maturity level of content in apps. While manipulating and visualizing our data we have observed that:

1. Mostly 90% of total apps are targeting audiences in every age group and hence open for everyone.

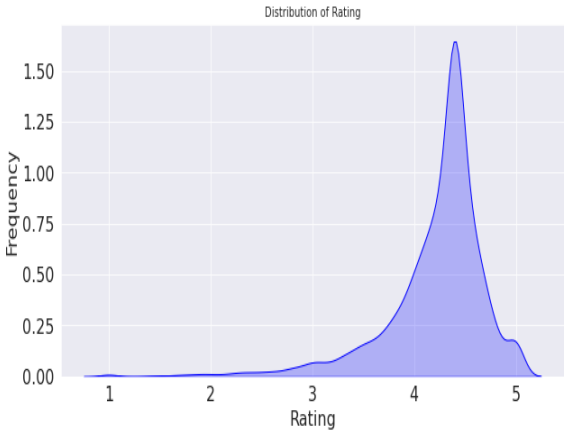
2. Very few(less than 500 apps) are catering to only the adult population i.e Mature 17+.

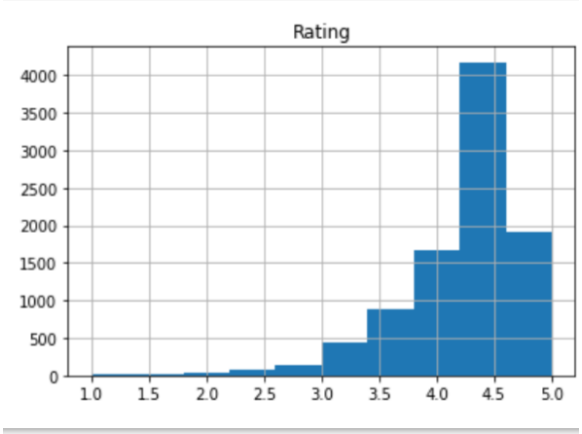




3.4 Rating Analysis

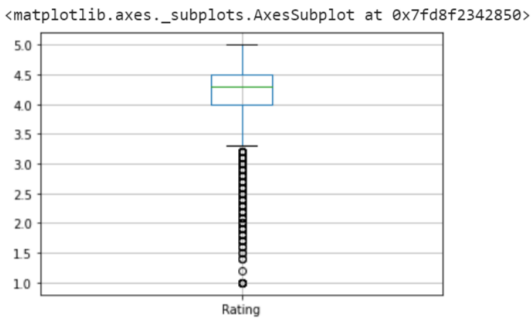
This graph is basically oriented for the marketing prospective. which company should have to invest for the good rating and which company have highest number of average rating and more number of installation.





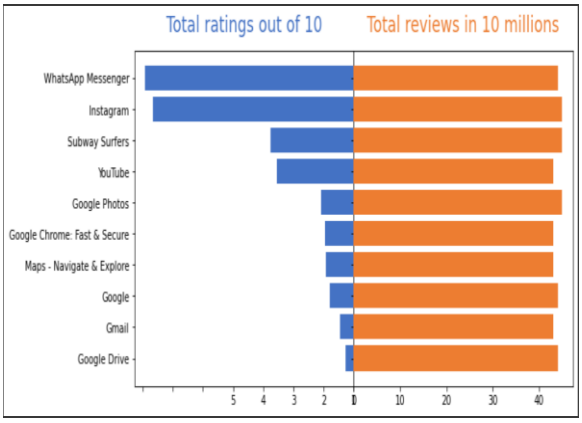
From this graph we can also say that

- Average rating of apps is 4 to 4.5
- Maximum rating is 4.3



3.5 Top 10 highest rating apps

These are the list of top 10 application have Highest rating and maximum review by the user.

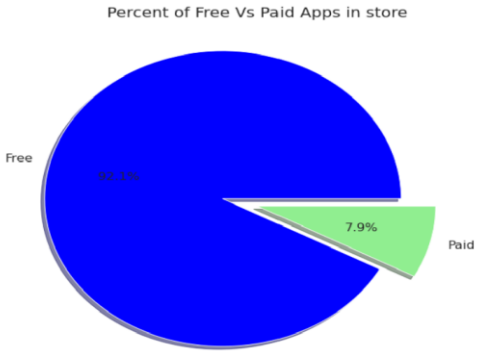


- Whatsapp, Instagram, Subway surface, Youtube, Google are in the list of top 5 which have maximum installed

- Google Chrome, Map- Navigate & Explore, Google, Gmail, Google Drive are in the list of top 5 to 10

3.6 Paid Vs free user

As we know that maximum number of audience are prefer to use free apps so we can see from the pie chart.



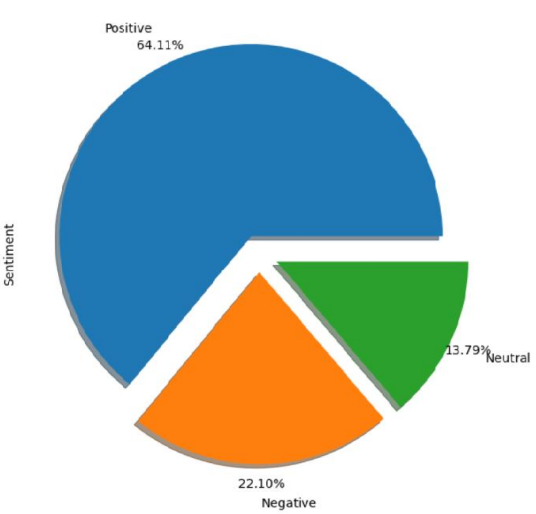
- 92.1% of them are the free user
- 7.9% of them are paid user

3.7 Sentiment Percentage of Review

Sentiment play important role in identifying weather people like this apps or not.
We have also determine the overall sentiment of review.

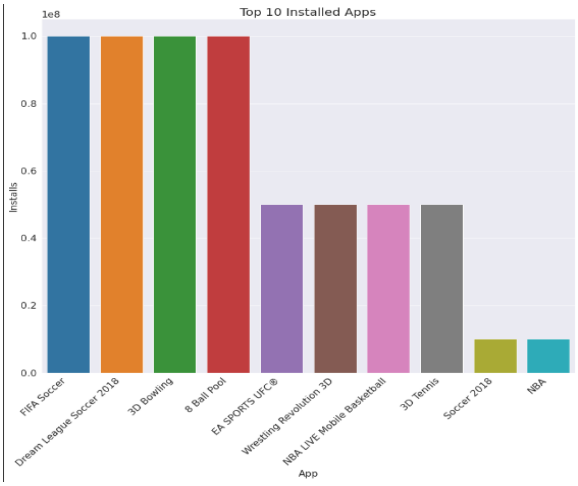
From the following pie chart we can say that say that mostof the apps on google play store have received positive reviews 64% by the

user, while some of the apps have received negative reviews as well 15% approx and rest of them are neutral.



3.8 Top 10 Game apps installed

It is playing the crucial role for the marketing prespective that which kind of application having maximum number of user in the gaming category

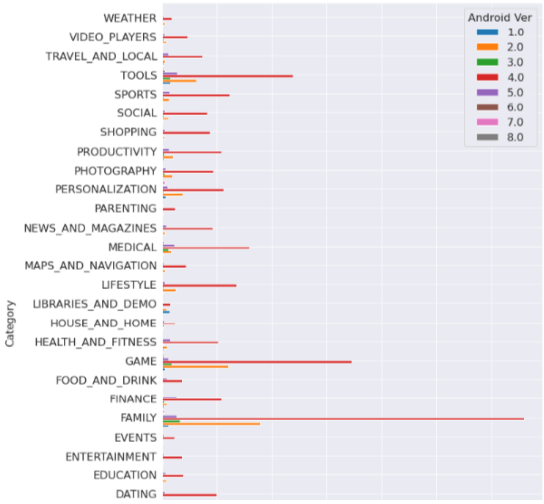


The above graph show

- FIFA Scorer, Dream League Soccer, 3D Bowing, 8 Ball pool are the maximum number of installed.
- EA SPORTS UFA, Wrestling Revolution 3D, NBA LIVE Mobile Basketball, 3D Tennis are in the list of top
- Soccer 2018 and NBA are in the list of top 8 to 10

3.8 Android Version of apps

Following graph shows android version which is used by the categories of application



- version 2.0 is mostly used by application
- Second maximum mostly used version is 4.0

4. Conclution

This data set provided contains a large amount of data that can be used for various purposes and we can extrect many possible data which is quite useful to understand or useful for the business purpose like all the question come out from the dataset which we have seen above.

4. References

1. Almabetter Deskboard and class material
2. Pandas and Numpy libraries
3. GeeksforGeeks
4. Pandas and Numpy libraries
5. Stack overflow
6. Youtube
7. Researchgate.net
8. W3school.com