

# **Play Store App Review Analysis**

**Divya Prakash Kedia  
Sakshi Ravindra Ghugare  
Alma better Trainee,  
Alma better Bangalore.**

## **Abstract:**

Software application is vital because specific software is required in almost every industry, in every business, and for each function. It becomes more important as time goes on. Mobile app distribution platform such as Google play store gets flooded with millions of new applications uploaded by developers every day. This project helps to analyze the database which can help to find the key factors which are responsible for the app engagement, popularity and success.

The objective of this is to analyze the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization & developers

## **1. Introduction:**

In current time, the Google Play Store is the largest and most popular android app store. It is Developed and Operated by Google, launched on 6th March, 2012. Approximately 3.48 million apps are in the Play store and it provides wide collection of data on features like ratings, price and number of downloads and apps description. It is important for developers to know which

apps are loved by customers and are trending in market so that he develop only those apps and also there is a high competition between app providers producing similar applications. The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

## **2. Problem Statement:**

The Play Store apps data has enormous potential to drive app-making businesses to success. Android is expanding as an operating system. It has captured around 74% of the total market which is a true indicator of the huge amount of population using android. Mobile app industry is increasing significantly and thus giving rise to more competitions to the one's that are creating applications.

The main objective is to build an analysis, which could help the ordinary people to find the best apps based on the previous data.

## **3. Exploring the database:**

We have provided with two databases

- Play store database

1. Shape of this database is (10841, 13).
  2. Out of this thirteen columns we have numeric values.
- User reviews database
1. Shape of this database is (64295, 5).
  2. Here there are only two numeric values found. Sentiment Subjectivity, Sentiment polarity.

#### **4. Features selection**

The columns are also known as features, one or more different features are grouped together for different analyses to form a data frame

- top\_genres\_df: It contains the rows of top 'Genres' based on the app counts.
- genres\_paidapp\_count: It holds the rows of top 'Genres' and corresponding total paid app installs count.
- genres\_freeapp\_count: It holds the rows of top 'Genres' and corresponding total free app installs count.
- app\_max\_review : it hold the rows of top 'apps' which content maximum review.
- genres\_ratings\_df: It takes the rows of top 'Genres' and their corresponding mean Rating.
- top\_30\_genres\_df: This data frame contains the rows of features top Genres, Free app installs count, Paid app installs count and Rating.
- cat\_type\_installs\_df: It contains the rows of Category wise free and paid apps installed.
- top\_5\_downloaded\_app: it contains the rows of top 5 downloaded apps.
- cat\_mean\_rating\_df: It contains the rows of Category wise mean rating of all the apps.
- content\_rating\_df: It holds the rows of different aged people wise ratings count.
- no\_null\_user\_reviews\_df: It holds the different users' non null Sentiment type, Sentiment Subjectivity distribution and Sentiment Polarity.
- sentiment\_subjectivity\_df: It holds the different users' sentiment subjectivity distribution.

#### **5. Data visualization and analysis**

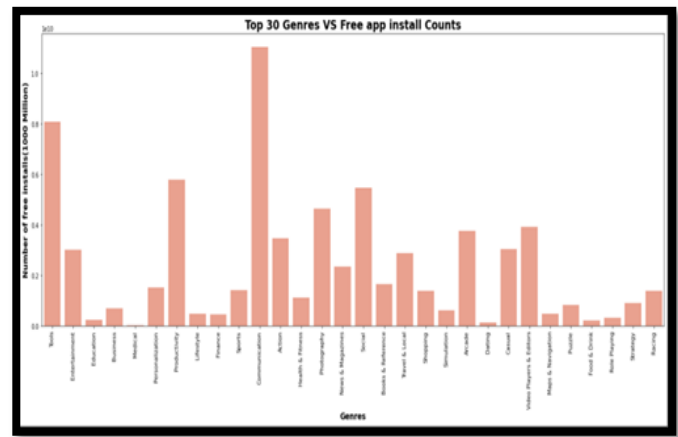
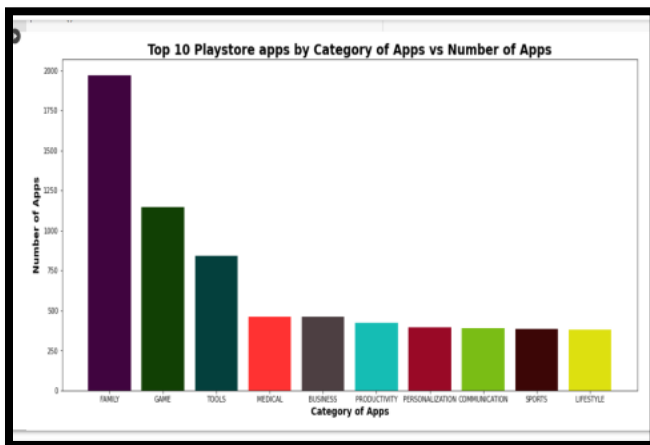
In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.

A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images analysis.

##### **5.1 Top 'Category' Analysis**

- The analyze is done between the top 10 Categories which was having more number of downloads.
- Here we can see the number of applications in the Categories like Family, Game, Tools and etc.
- Categories like Sports and Lifestyle are the least.



## 5.2 Top ‘Genres’ Analysis

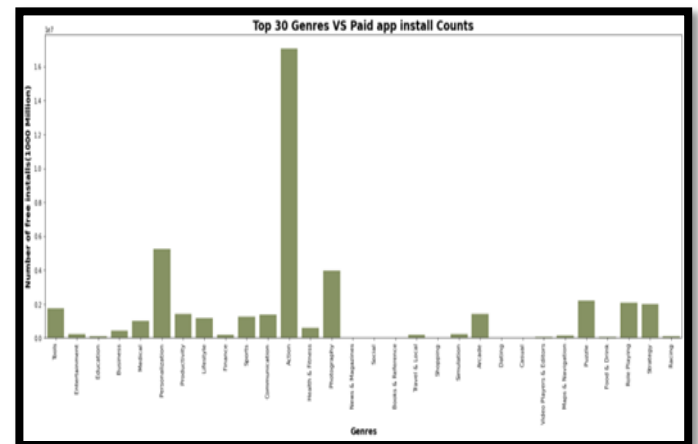
- The analyze is done between the top 10 Genres which was having more number of downloads
- Here we can see the number of applications in the genres like Tools, Entertainment, Education and etc.
- The developers are mostly focusing on these genres because of the people's daily basis requirements.
- Genres like Educational, Parenting, Music are having comparatively less.

## Top Free apps

- When compared, the both plots it was observed that, people are showing more interest on free apps like communication, Tools etc.
- As seen earlier the Educational, Parenting and Music are the genres in the least Top free apps.

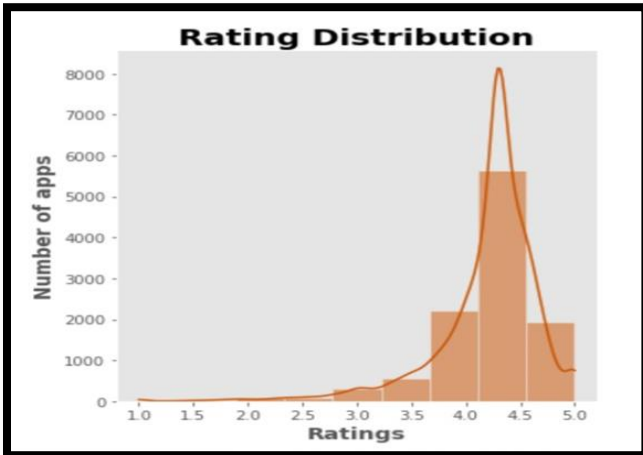
## Top Paid apps

- When it's coming, commercial people are preferring apps like Games, Photography, and Personalization.
- People are preferring less on Educational, Event, and Art & Design.



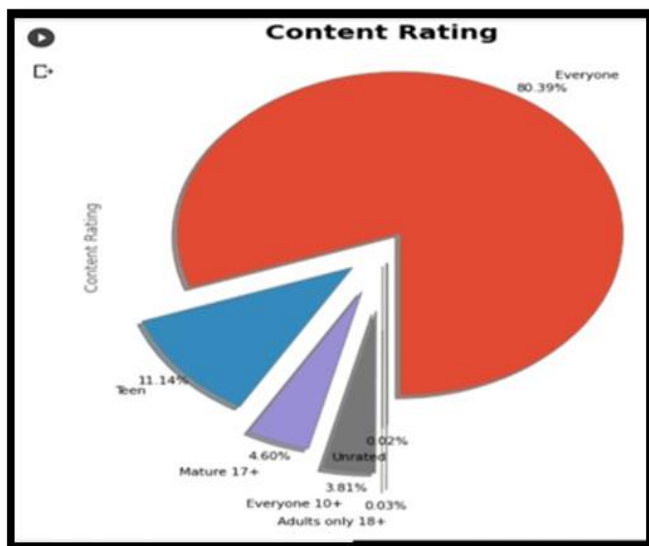
## 6. Rating-Distribution:

- In general Rating is the main scale factor, which measures how much people are satisfied with the product.
- Here, most of the people rated the apps between the ranges of 4 to 5, which can be considered as good.
- Moderate rating lies between 2.5 to 4.
- And below average or poor apps are in the range of 0 to 2.5 which are less in count.



### 6.1 Impact of 'Rating' :

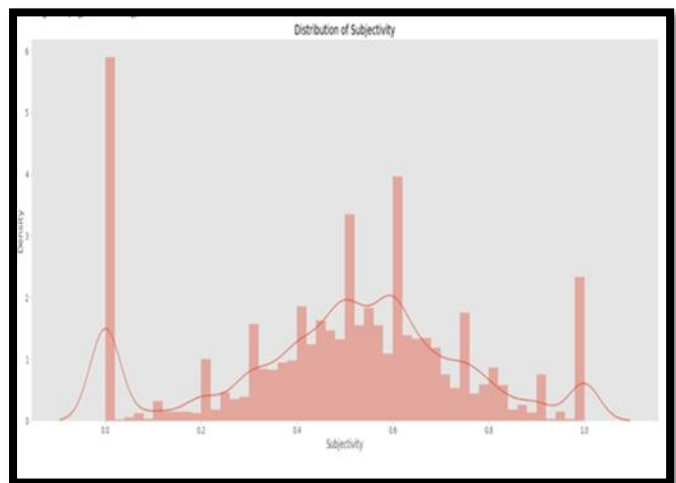
- The content rating shows the results for general contents as high.
- The content rating type 'Everyone' has the most percentage value of 80.39%.
- 'Teen' contents are second in the order with the percentage of 11.14%.
- Adult's only and unrated contents are least in this plot, 0.03% and 0.02% respectively.



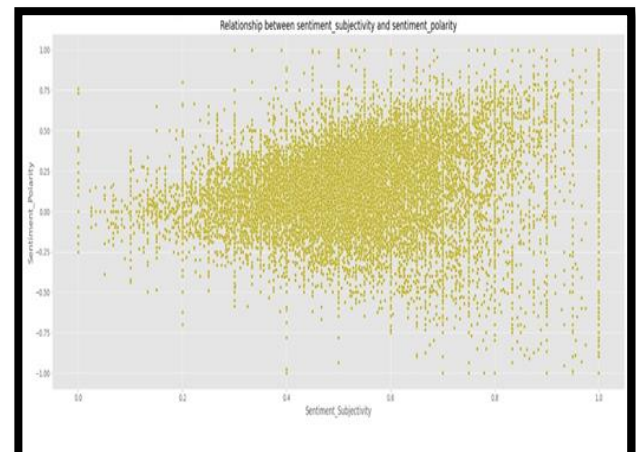
## 7. User sentiments analysis:

### Distribution of Subjectivity:

- Subjectivity lies mostly between 0.5 and 0.65.
- It shows that the average content and apps reviews subjectivity are mostly relevant.
- Subjectivity of 100% has slightly occurred frequently.
- The nearly 0 subjectivity has a considerable amount of frequency.



### 7.1 Relationship between Subjectivity and polarity:



From the above scatter plot it can be observed that sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.

- Whether or not does the size of app matters.

## **8. Conclusion:**

Thus the app development companies could decide what application should be developed and they can also see the prediction of their developed application. In this they also get to see the categorized reviews of all the application in one interface which will help them decide which app is liked by the users and which apps need to be developed more. The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project.

After analyzing the dataset we have got answers to some of the interesting facts which any of the android users would love to know.

- Top 10 apps available per category in Play store.
- No. of Installs per Category.
- Top 7 apps in any category.
- Top 30 Genres for free Installation.
- Top 30 Genres for paid Installation.
- Top apps in play store based in Reviews.
- Mean Rating for all Categories
- Rating distribution for apps.
- Types of Content rating for the apps.