**Play Store Review Analysis**

**Bhoopalan M.S, Rajakumaran S, Gowtham R, Ranjith R**

**Abstract:**

Software application is vital because specific software is required in almost every industry, in every business, and for each function. In the Real-world android is the dominant mobile operating system today more than 85% of peoples using Android mobile. we aim on analysing Google play store that provides a particular app description and data such as reviews,

ratings, price and number of downloads and descriptions related to application functionality. we 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 overall objective of this analysis effort is to provide in-depth insight about play store review analysis in general.

**1. Introduction**

The Google Play Store started life as the “Android Market” in 2008. It launched alongside the very first Android devices, and its purpose was to distribute apps and games. The Android Market was extremely basic at the beginning. It didn’t support paid apps and games until 2009. However, as the Android platform grew, so did the Android Market. By 2012, it featured over 450,000 Android apps and games.

Many apps are being developed as apps are easy to create and its lucrative. But it’s important for developers to know which apps are loved by customers and are trending in market so that he develops only those apps and there is a high competition between app providers producing similar applications. 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 analysing 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 analysing is collected from dataset.

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

**2. Problem Definition:**

Our Analysis is divided

into four phases: data extraction, data cleaning, data visual-

ization, and applying different models, and it is depicted in

ﬁgure 8. First, we collect the data from the Kaggle website.

In the next step, we try to do data cleaning on the data set

to reduce the error percentage. After the data set is ready,

we try to analyze the data set using different plots and re-

move the stuff not needed from the data set. The last step

includes using different classiﬁcation algorithms on the data

set to see which one gives the highest percentage of accu-

racy. Finally, we narrate the analysis results to provide a

clear vision of the relationship among the areas of interest

Our Analysis is divided

into four phases: data extraction, data cleaning, data visual-

ization, and applying different models, and it is depicted in

ﬁgure 8. First, we collect the data from the Kaggle website.

In the next step, we try to do data cleaning on the data set

to reduce the error percentage. After the data set is ready,

we try to analyze the data set using different plots and re-

move the stuff not needed from the data set. The last step

includes using different classiﬁcation algorithms on the data

set to see which one gives the highest percentage of accu-

racy. Finally, we narrate the analysis results to provide a

clear vision of the relationship among the areas of interest

Our Analysis is divided

into four phases: data extraction, data cleaning, data visual-

ization, and applying different models, and it is depicted in

ﬁgure 8. First, we collect the data from the Kaggle website.

In the next step, we try to do data cleaning on the data set

to reduce the error percentage. After the data set is ready,

we try to analyze the data set using different plots and re-

move the stuff not needed from the data set. The last step

includes using different classiﬁcation algorithms on the data

set to see which one gives the highest percentage of accu-

racy. Finally, we narrate the analysis results to provide a

clear vision of the relationship among the areas of interest

Our Analysis is divided into four phases: data extraction, data cleaning, data visualization, and applying different models. First, we collect the data from the Kaggle website. In the next step, we try to do data cleaning on the data set to reduce the error percentage. After the data set is ready, we try to analyse the data set using different plots and remove the stuff not needed from the data set. The last step includes using different classiﬁcation algorithms on the dataset 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.

Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.

To Explore and analyse the data to discover key factors responsible for app engagement and success.

The Problem statements are

1. What are the top categories in Play Store?
2. Which category has the most no. of installations?
3. How are ratings Distributed in overall application?
4. Which Content ratings apps most present in Play store?
5. Are majority of the apps Paid or Free?
6. Which categories of apps have most paid and free apps?
7. How are ratings affected with corresponding app size and paid apps?
8. Find the most correlation using heatmap?
9. How peoples are reviews app using Sentiment analysis?
10. How peoples are reviews app with own Experience?

**3. EDA on given Data set**

There are two datasets:

1. **Play Store Data** (App, Category, Rating, Review, Size, Install, Type, current rating, genres, Last update, Current Var, Android Var)
2. **User Review Data** (App, Sentiment, Sentiment Polarity, Sentiment Subjectivity)

By basic analysis of the data frame, we understand from that:

* There are 10841 rows(apps) of data and 13 Columns (Properties of apps).
* Column Ratings in the type of ‘Float64’ and remaining all are in ‘Object’ type.
* From the sample data, we observe that some values in the columns like Installs, Price and Size have a few special characters (+, $, M, k). we need to convert into numeric format for easy to analyse.
* The special characters "," and "+" present in Installs column and "$" present in Price column need to be removed.
* Size Column, need to Convert into One format (kb to Mb).
* Check any Duplicated are there or not. if Duplicated apps found then remove it.

**3.1.** **Top Categories in Play Store**

From this Bar Chart plotting (Fig 1), most of the apps in the Play Store are in Family category and then Games, Tools...etc

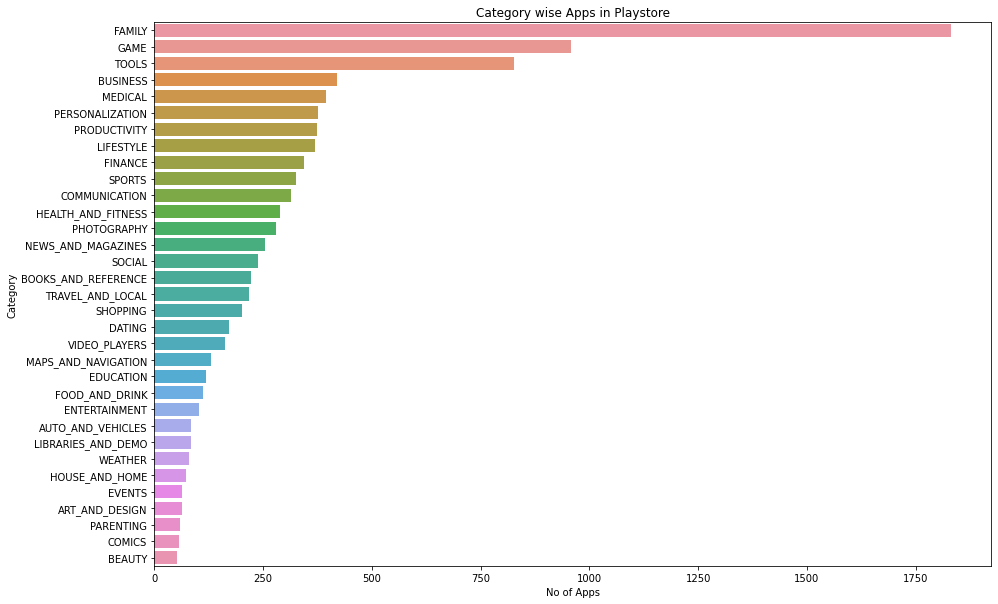


Fig 1: Bar chart of categories wise apps in Play Store.

**3.2. Categories vs Installation:**

From this Bar Chart plotting Categories wise, Communication have installed the greatest number of apps. Refer fig 2.

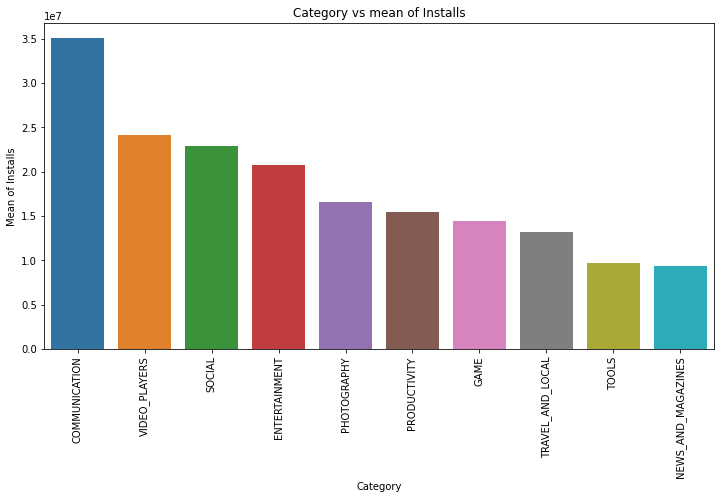


Fig 2: Categories vs Installation

## 3.3. Genres vs Installation

From this Bar Chart plotting Genre wise, Communication have installed the greatest number of apps. Refer fig 3.

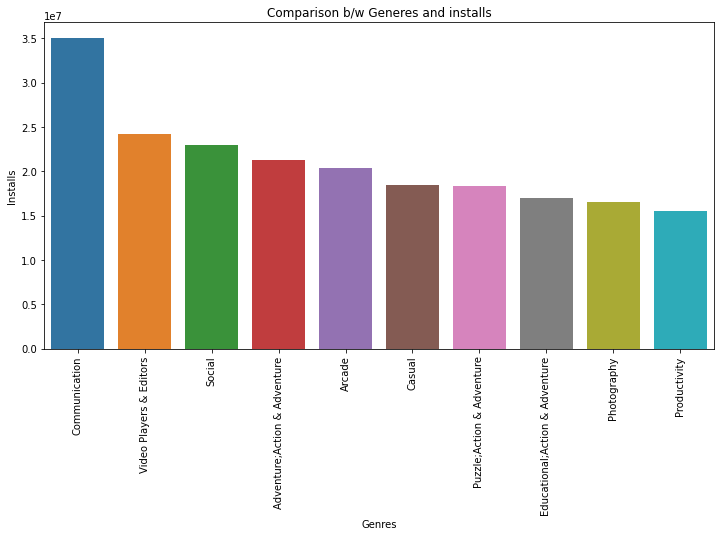


Fig 3: Bar chart of Genre vs No if Installs

## 3.4. Rating among apps

Here we see the most given rating among the apps in play store that is more than 2k apps got the rating of 4.3 Refer fig 4.

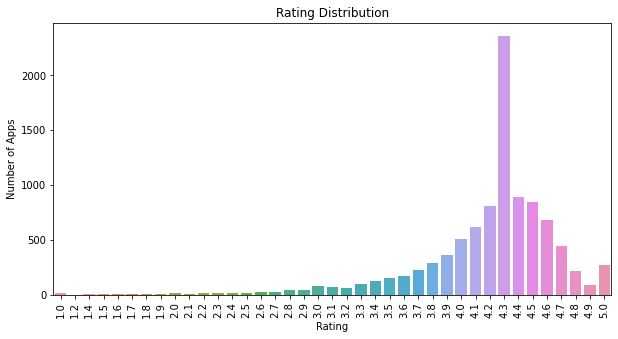


Fig 4: Most given rating among the apps in play store.

## 3.5. Content Ratings

Here we see the percentage of each Content of app and from this we conclude that everyone content rating is 81.86%. and then teen (10.73%). Refer fig 5.

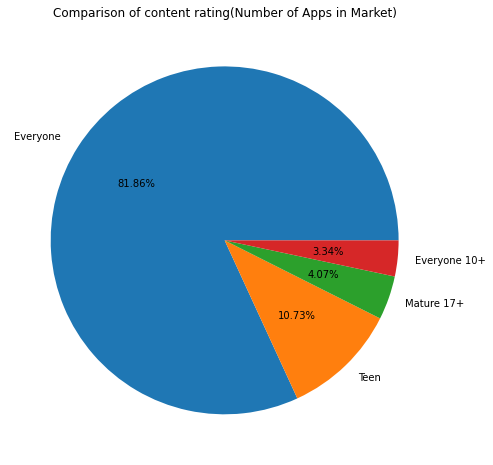


Fig 5: Content Rating.

**3.6. Category wise content rating**

Here we see the category wise content rating given by Everyone vs Teens.

Refer fig 6.

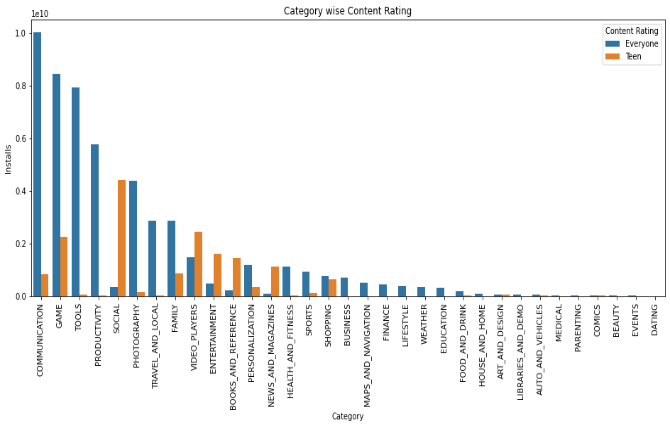


Fig 6: Free vs Paid Apps

**3.7. Content rating vs Installs**

Users from Teen category and 10+ Everyone category has installed a greater number of apps compared to others

Refer fig 7.

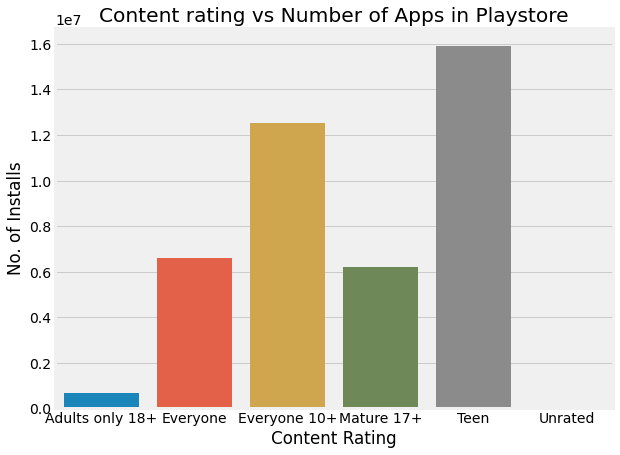


Fig 7: Category of users with more number of installs

**3.8.** **Paid vs Free Apps:**

Here we see that free apps present in play store is 92.17% and paid apps is 7.83%

Refer fig 8.

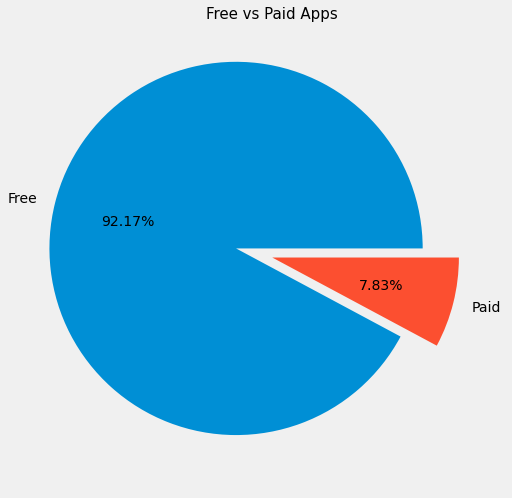


Fig 8: Free vs Paid Apps

**3.9. Category wise Free and paid apps count:**

Family category have installed more free and paid apps followed by game category compared to other categories

Refer fig 9.

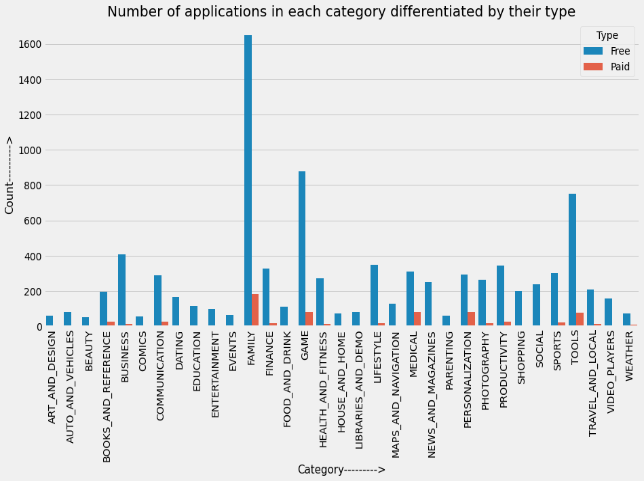


Fig 9: Category vs Free and Paid Apps

## 3.10 Ratings Affected by different Category basis

**3.10.1 Effect of size on Rating**  
From the above comparison between Size and Ratings discover that users like less size apps.

Refer fig 10.1

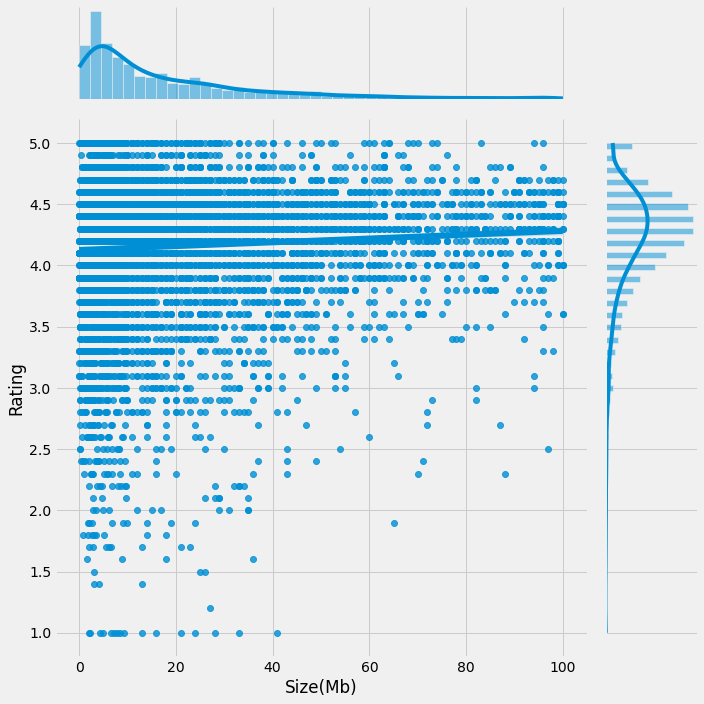
****

Fig 10.1: Effect of size on ratings

**3.10.2 Effect of price on Rating**  
From the above comparison between Size and Ratings discover that users like less price apps.

Refer fig 10.2

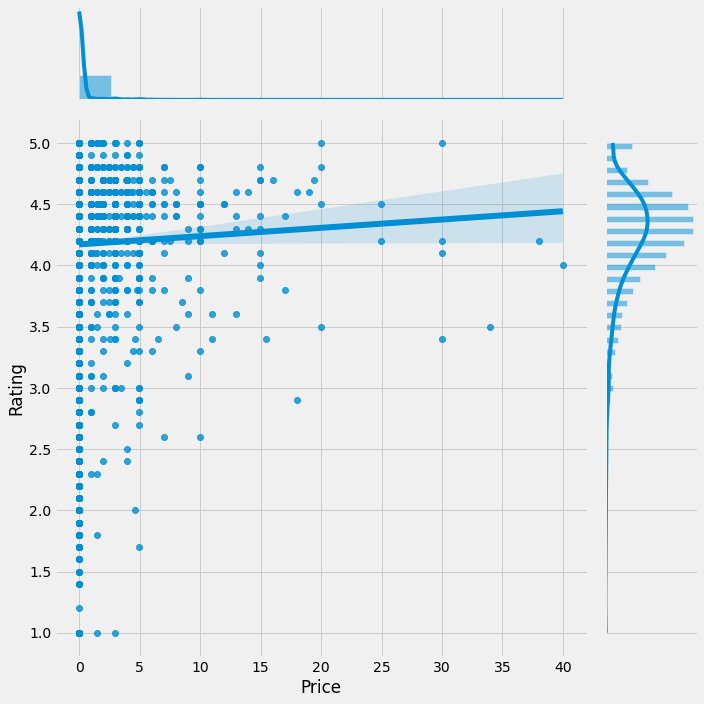


Fig 10.2: Effect of price on ratings

**3.11. Top 5 most reviewed apps**

Here we see that Facebook has most reviews and has overall rating of 4.1 whereas clash of clans and messenger has the 5th most reviewed app and has overall rating of 4.6

Refer fig 11.

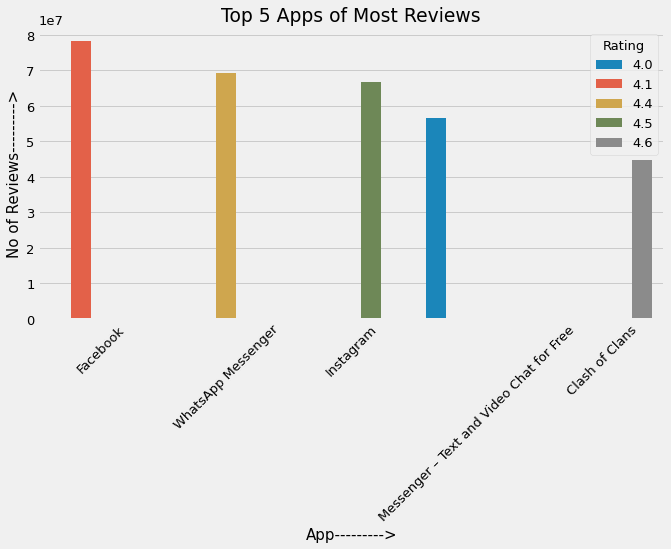


Fig 11: Free vs Paid Apps

**3.12. Correlation between columns**

From the Heat map Correlation map, we can get ‘Installs’ and ‘Reviews’ are more Correlated. Hence the most installed apps are having more reviews and less installed apps are having lesser reviews.

Refer fig 12.

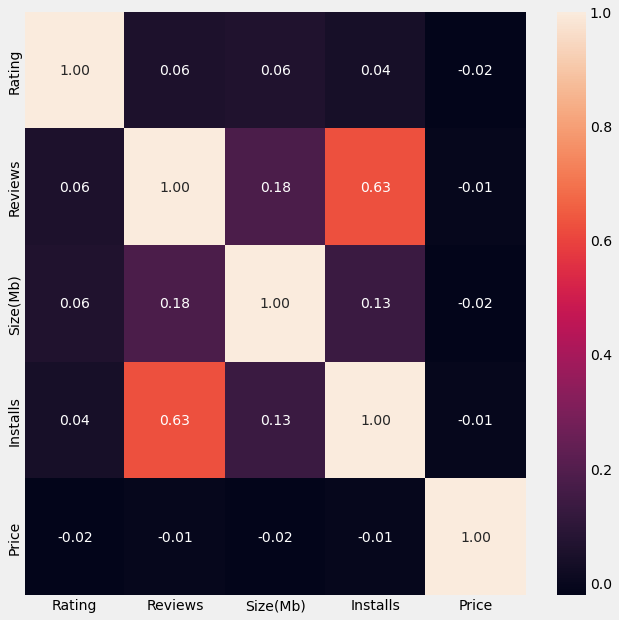


Fig 12: Correlation Heat Map

## 3.13. Sentiment Analysis on different Category basis.

Here we conduct the graph which show type sentiment analysis from each category. This graph helps us to know which category get highest sentiment review as well as compare of +ve and -ve review.

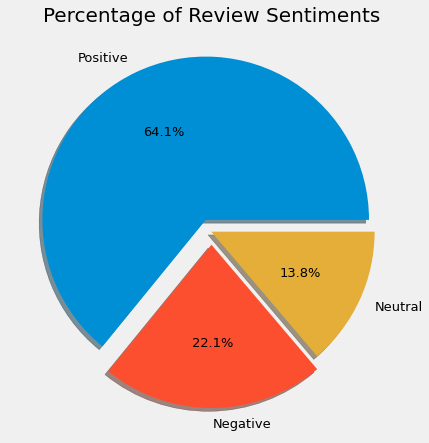


Fig 3.13: Percentage of Review Statement

From above the pie chart represent that most of the reviews are positive. Refer Fig 3.13

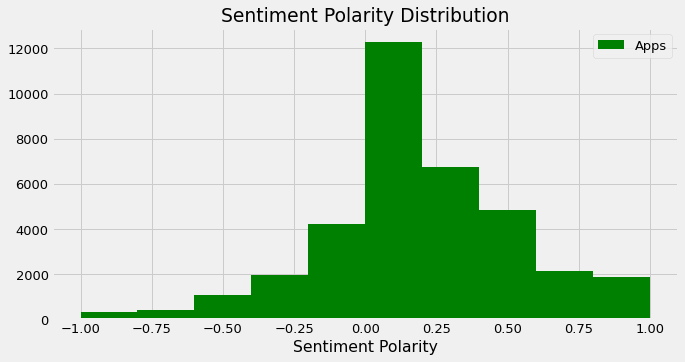


Fig 3.13.1: Sentiment Polarity

From above the Distribution chart represent that most of the reviews are positive and polarity value are positive. Refer Fig 3.13.1

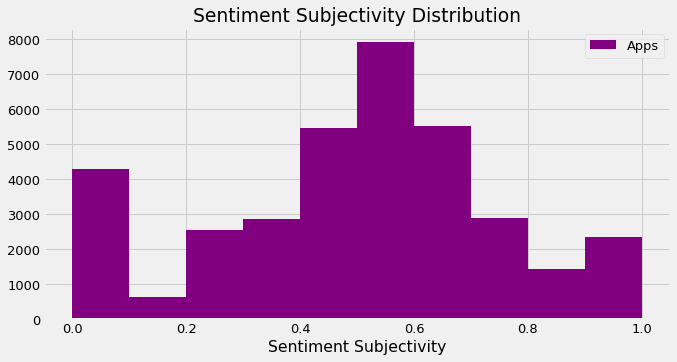


Fig 3.13.2: Sentiment Subjectivity

From above the Distribution chart represent that maximum number of sentiment subjectivity of Apps between 0.4 to 0.7. From this, we can conclude that the maximum number of the users give reviews to the applications, according to their own experience. Refer Fig 3.13.2

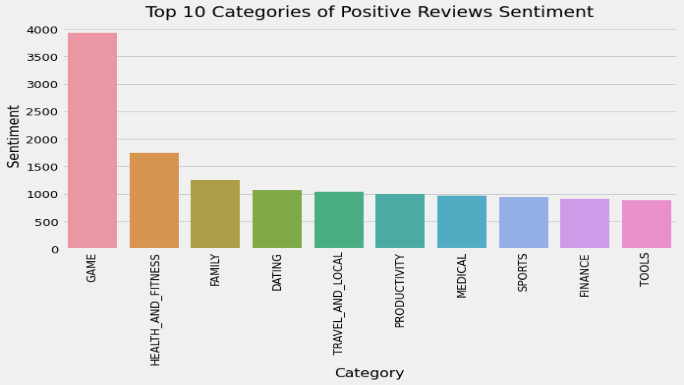


Fig 3.13.3: Category vs sentiment

From the above chart we can conclude that game category has the most positive review sentiments followed by health and fitness compared to other category.

**4. Conclusion**

The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product. As per the graphs visualizations shown above,

1. The highest amount of application based on categories is family, followed by game and tools.

2. Communication Category has highest number of installs by the user followed by Video Players, Social category etc.,

3. The highest number of reviews in total is achieved by game category, but the highest amount of reviews/app is achieved by social. It means user is likely to review social applications compare to other categories.

4. when we are comparing Genre with installs of application, Communication Category have highest installation followed by Video Players, Social, Action & Adventure.

5. In this distribution ratings majority of the apps had a rating of 4.3, and most apps lie between 4 and 5.

6. Users from Teen category and 10+ Everyone category has installed a greater number of apps compared to others

7. In play store, a greater number of free apps are there compared to paid apps. Maximum users prefer free apps compared to paid apps.

8. Most of the users likes lesser size apps compared to bigger size apps as per the Size vs Installs chart.

9. The most installed apps are having more reviews and less installed apps are having lesser reviews. Hence Number of installs is directly proportional to that app review

10. Most of the apps got positive reviews. 'Game' categories had highest positive review followed by 'Health and Fitness', 'Family', and dating.

11. The maximum number of sentiment subjectivity is between 0.4 to 0.7. From this, we can conclude that the maximum number of the Users give reviews to the applications, according to their own experience.

**5. References**

1. [www.geeksforgeeks.org](http://www.geeksforgeeks.org)
2. [www.kaggle.com](http://www.kaggle.com)