Google Play

Capstone Project

Play Store App Review Analysis

By

Sandip Dey

Data Science Trainee, AlmaBetter

WHY ANALYZE THE GOOGLE PLAY STORE?

- Mobile App Market is set to grow 20% by 2023.
- Android Apps comprise 90% of the Mobile App Market.
- What makes an App popular? Can we predict how popular it’s going to be?
- What are some interesting patterns in user behavior related to app usage & feedback

INTRODUCTION

- Android is the most popular operating system in the world, with over 3 billion active users spanning over 193 country.
- Google Play Store was launched on March 6, 2012, bringing together Android Market marking a shift in Google digital distribution strategy.
- There are more than 4 million apps found on Google Play Store.
- Android is the dominate mobile operating system today more than 85% of all mobile devices running Google OS. The Google Play Store is the largest and most popular Android app store.

- ## PROBLEM STATEMENT

- Two datasets are provided, one with **basic information** and the other with **user reviews** for the respective app.
- We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success

So, what factors influence an app's success?

- An app is said to be successful if it has:
- ❑ A high average user rating
- ❑ A good number of positive reviews
- ❑ A good number of monthly average users
- ❑ High revenue per customer and so on.

DATASET PREPARATION

- ✓ **Loading the data sets:** Two datasets, First Play store app dataset and User Reviews dataset.
- ✓ **Import Libraries:** NumPy, Pandas, Matplotlib and Seaborn.
- ✓ **Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data.
- ✓ **Data Imputation:** Filling the missing categorical values with mode and numerical values with median. Conversion of price, installs, reviews into numerical values.
- ✓ **Exploratory Data Analysis:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

DATASET DESCRIPTION

1. Attributes in Google Play store Data

I. App: The name of the app

II. Category: Category of the app belongs to,.

III. Rating: Rating has received from the users, range from 0.0 to 5.0

IV. Reviews: The number of reviews that the app received.

V. Size: The size of the app, The suffix "M" means Megabytes and the suffix "k" means Kilobytes

VI. Installs: Describes the number of install of the app

VII. Type: A label that indicates whether the app is free or paid

VIII. Price: The price value for the paid apps.

IX. Content Rating: It indicates the age group for user.

X. Genre: List of genres to which the app is belongs.

XI. Last Update: The date at which the app was last updated

XII. Current Version: Version of the app as specified by the developers.

XIII. Android Version: The Android OS the app is compatible with

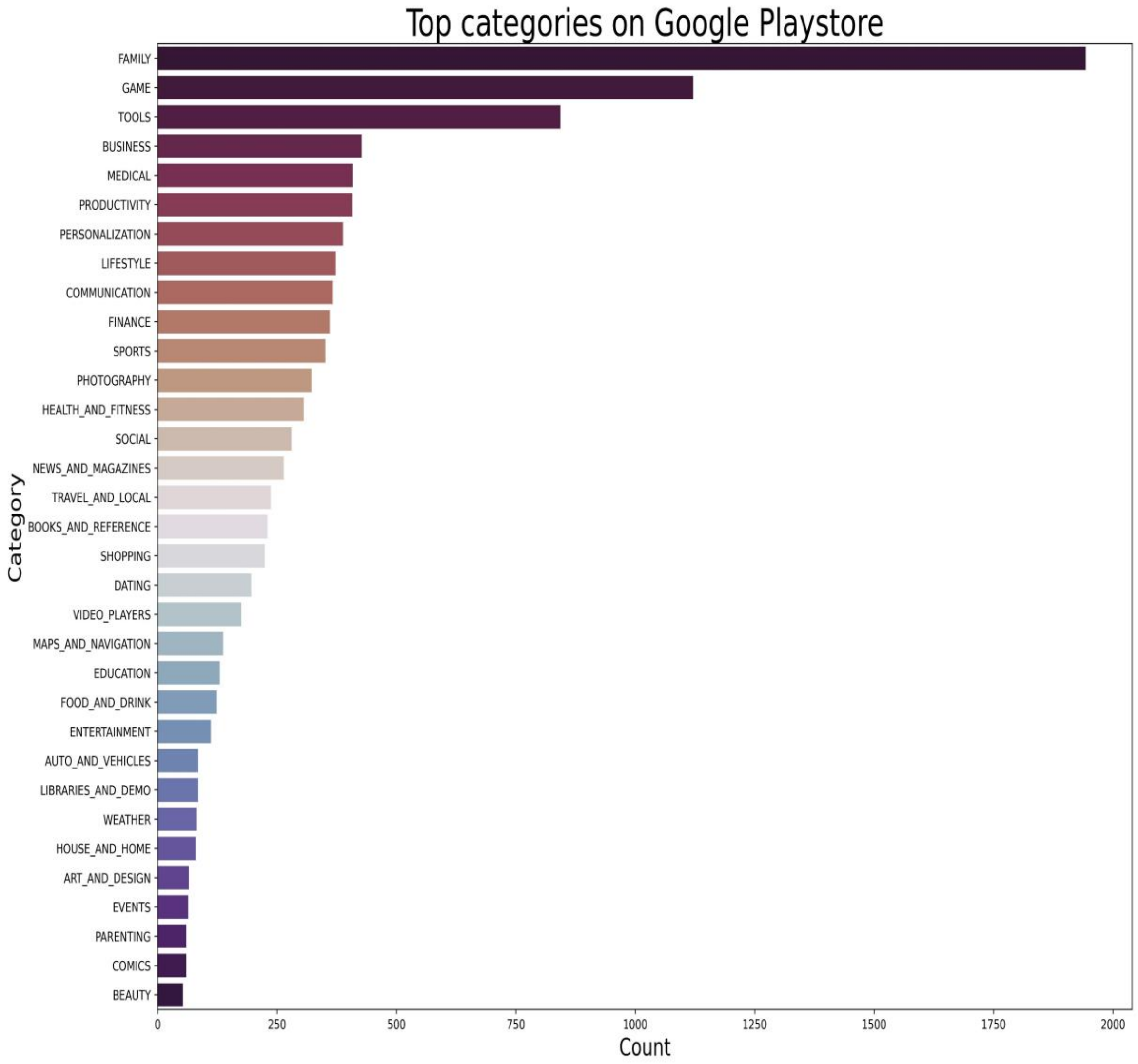
2. Attributes in User reviews

- App: The name of the app
- Translated Reviews: The review text in English
- Sentiment: Sentiment basically determines the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.
- Sentiment Polarity: Sentiment Polarity is float which lies in the range of (-1,1) where 1 means positive statement and -1 means a negative statement.
- Sentiment Subjectivity: Sentiment Subjectivity generally refer to personal opinion, emotion or judgment, which lies in the range of (0,1).

PROBLEM STATEMENT

1) Top categories on Google Play Store.
2) Which category apps have the greatest number of installs?
3) Which app category are paid or free with percentage?
4) Let’s have a look at the distribution of the ratings of the data frame
5) Which category of Apps from the ‘Content Rating’ column is found more on the play store?
6) Let’s have a look at the distribution of the Size of the data frame.
7) What are the top 10 installed apps in any category?
8) Which are the Apps having the highest numbers of reviews?
9) How does size impact on the number of installs of any application?
10) Which are the Genres that are getting installed the most in top 20 Genres?
11) Most Android Ver supported apps in play store.
12) App update details "By Year".
13) Histogram of subjectivity.
14) Is sentiment subjectivity proportional to sentiment polarity?
15) Sentiment_Polarity relation with paid and Free App.
16) Find the highest and the lowest rated Genres.

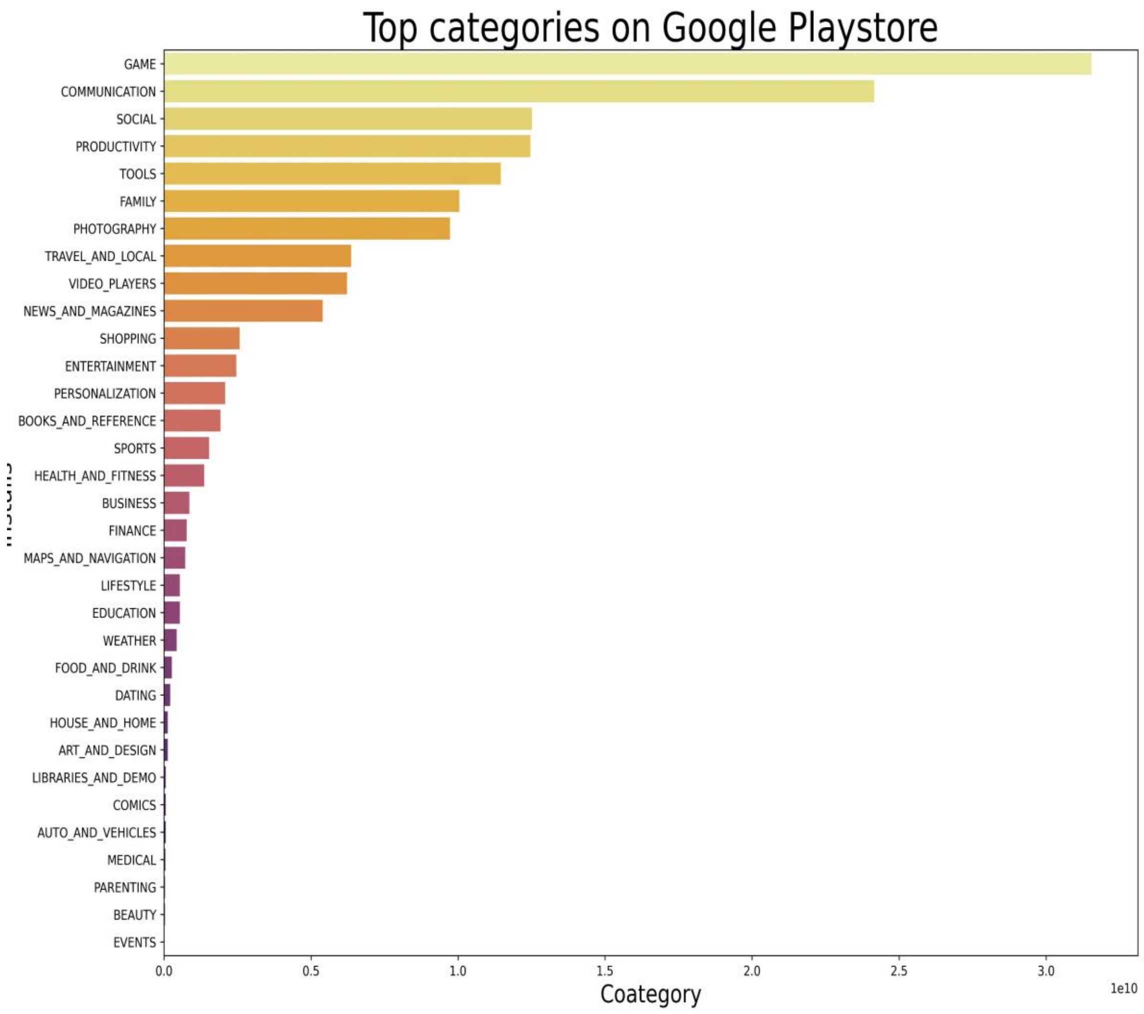
1. Top categories on Google Play Store



There are total 33 categories in the dataset From the above plot

and we can come to a conclusion that in playstore most of the apps are under Family & Game category and least are of Beauty & Comics Category.

2. Which category apps have the greatest number of installs?



From the above graph

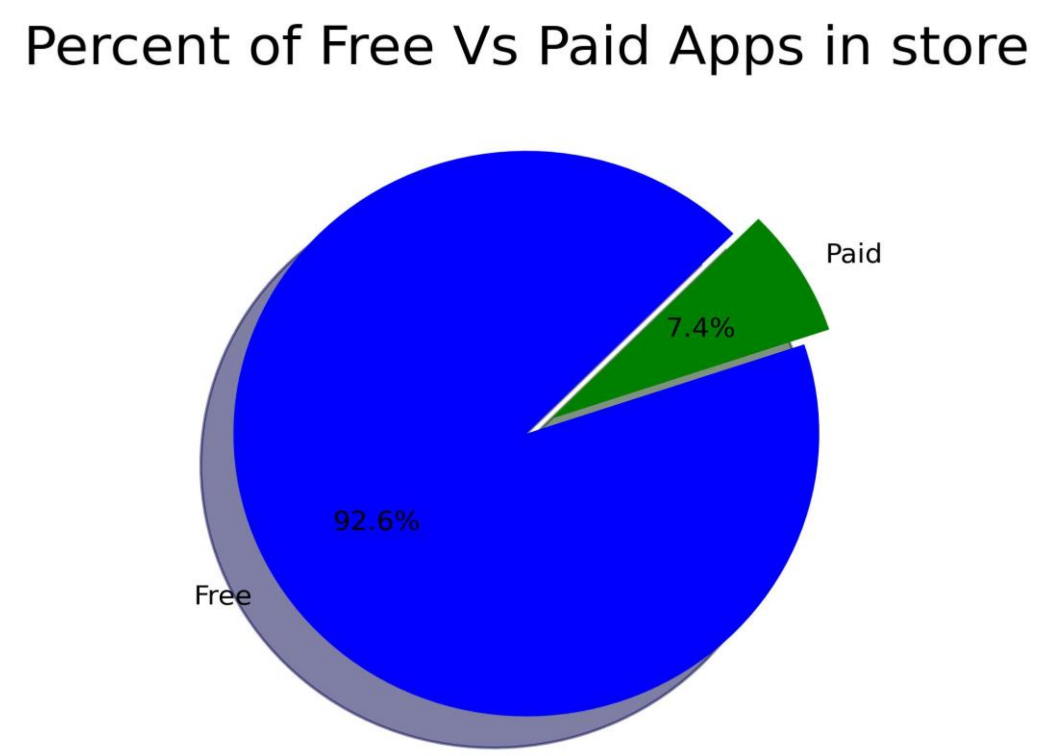
we can see that there are total of 33 categories in the dataset.

We can come to the conclusion that in the play store the top categories with the highest installs is-GAME category and least is EVENTS categories.

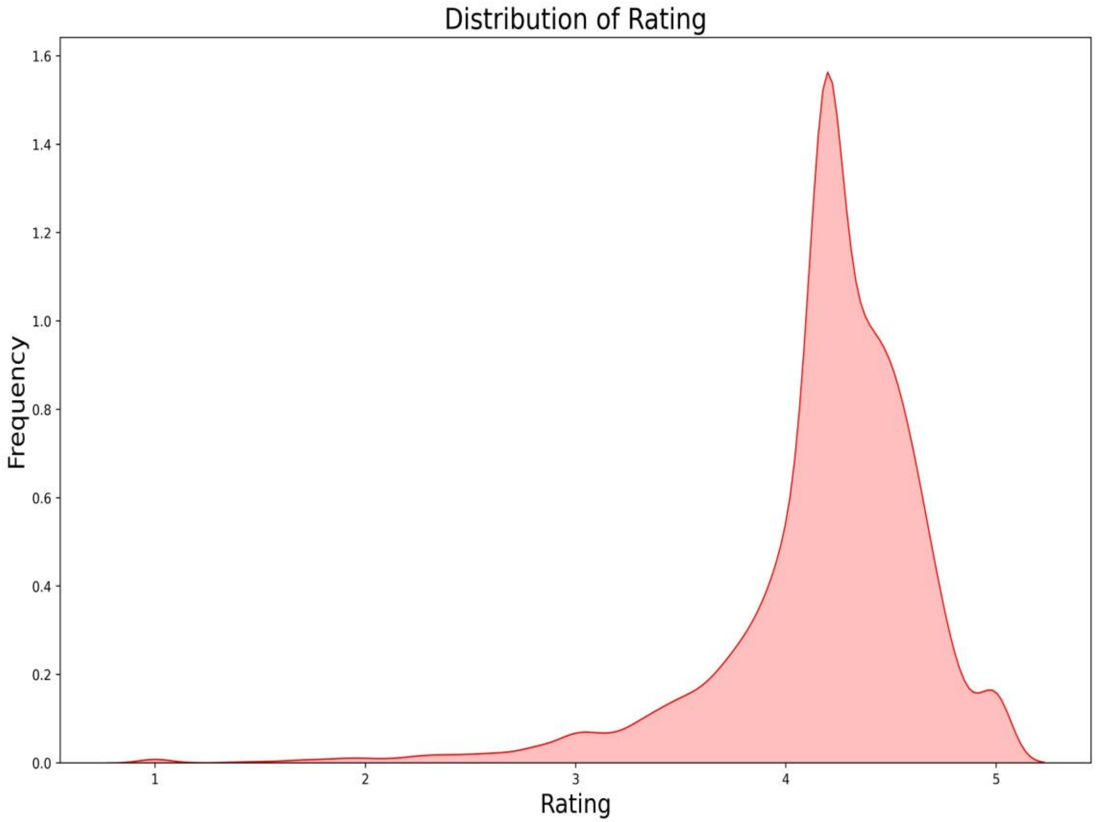
3. Which app category are paid or free with percentage?

The graph

is indicates that 92.6% apps are free to dowenload and rest 7.4% are paid apps.



4. Let’s have a look at the distribution of the ratings of the data frame.



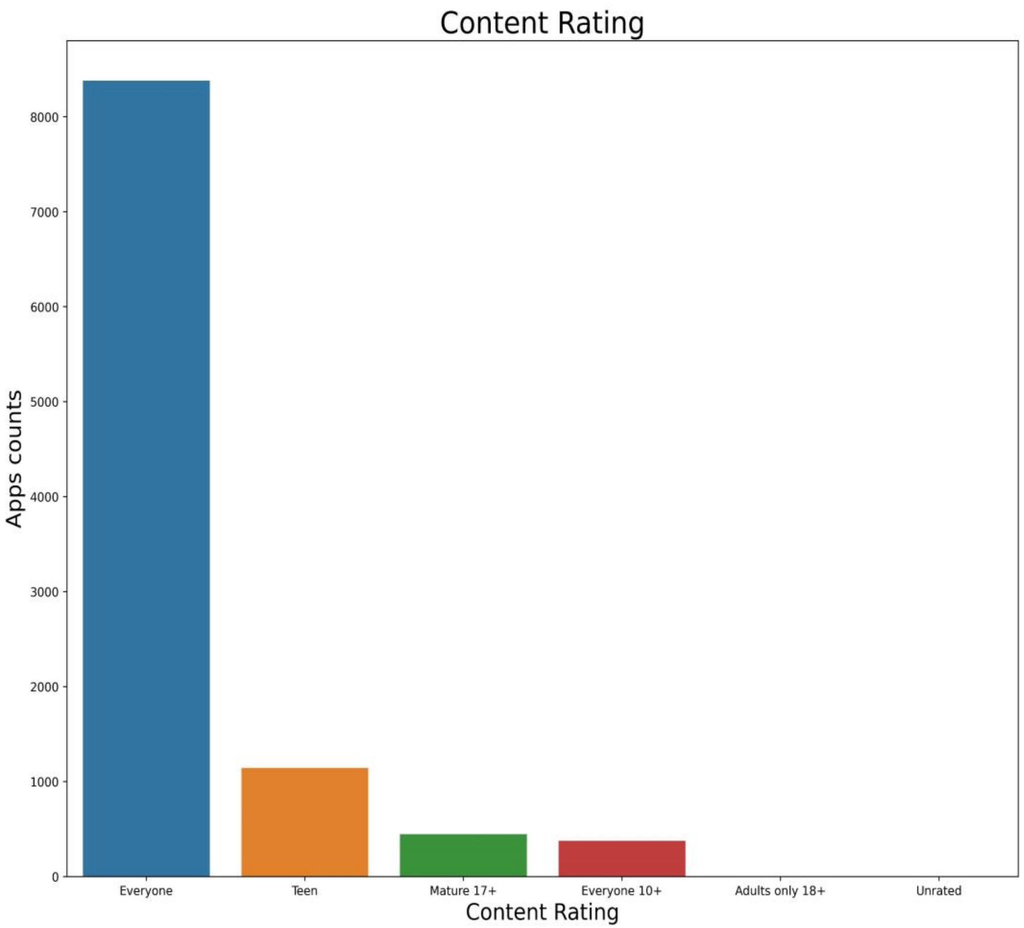
From the above graph,

we can come to the conclusion that most of the apps in the google play store are rated between 3.8 to 4.8.

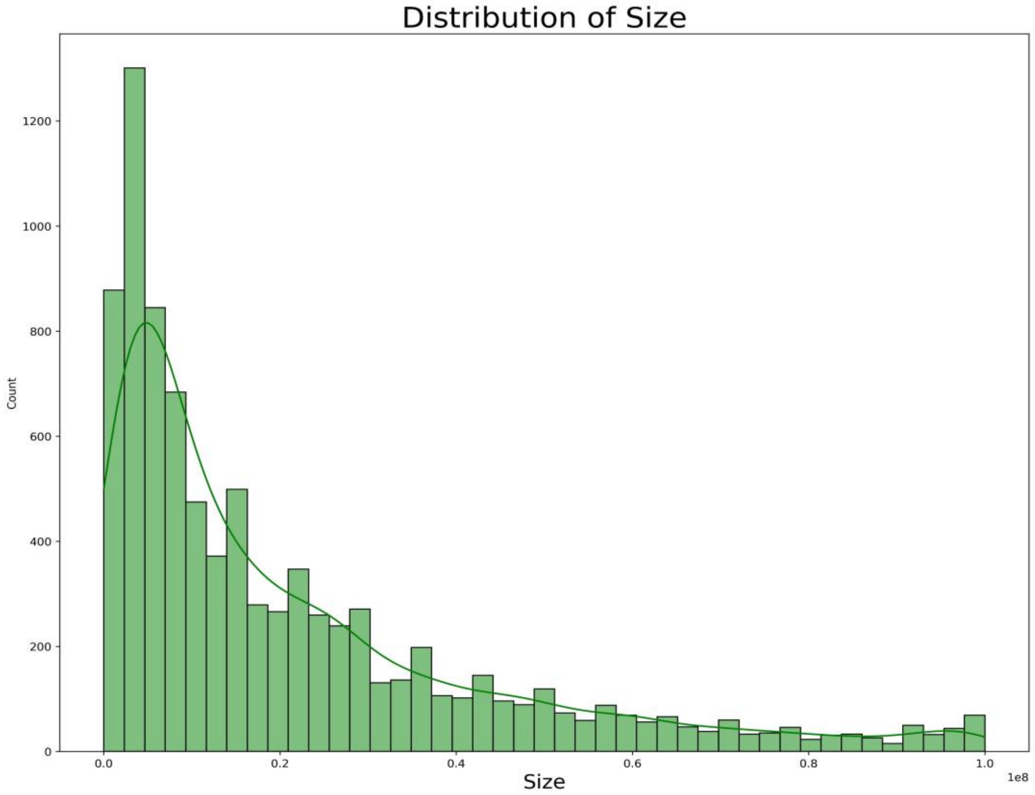
5. Which category of Apps from the 'Content Rating' column is found more on the play store?

From the above plot,

we can see that the Everyone category has the highest number of apps and unrated category has the lowest number of apps.



6. Let’s have a look at the distribution of the Size of the data frame.



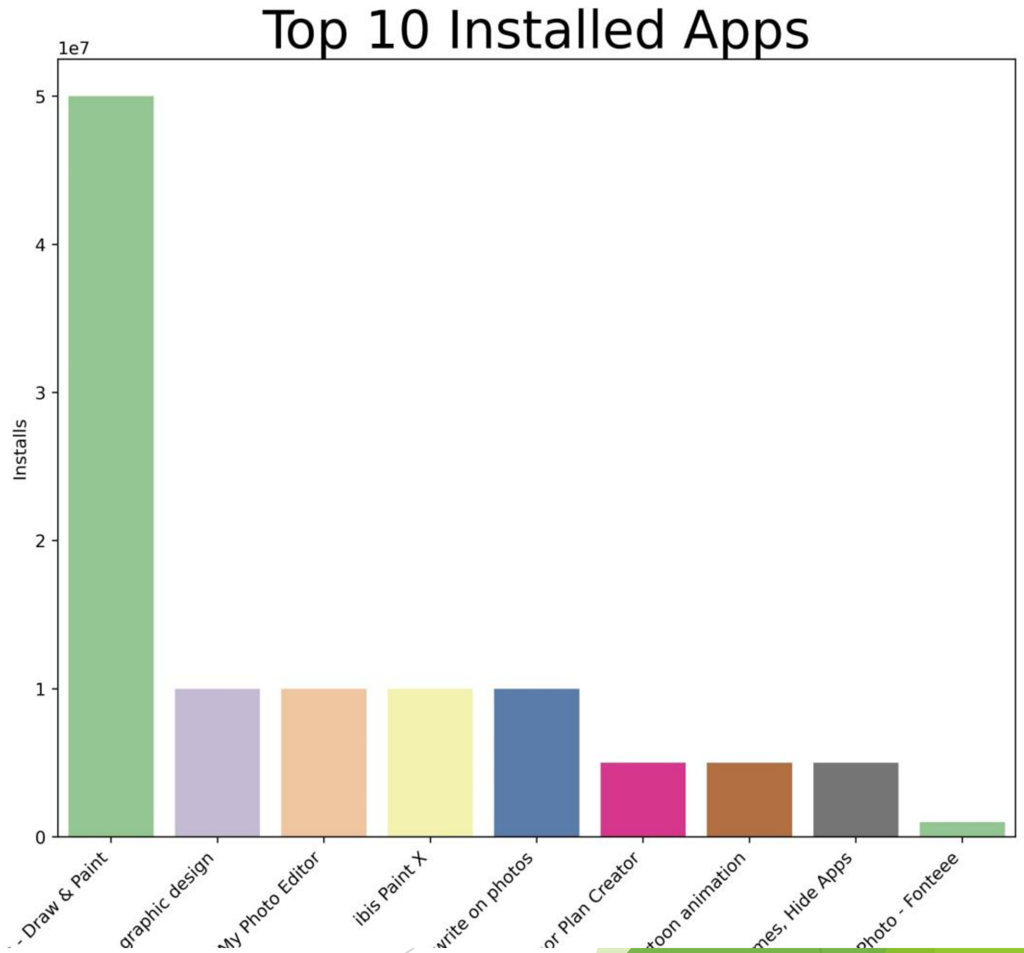
From the above histogram graph,

we can come to the conclusion that maximum number of applications present in the dataset are of small size.

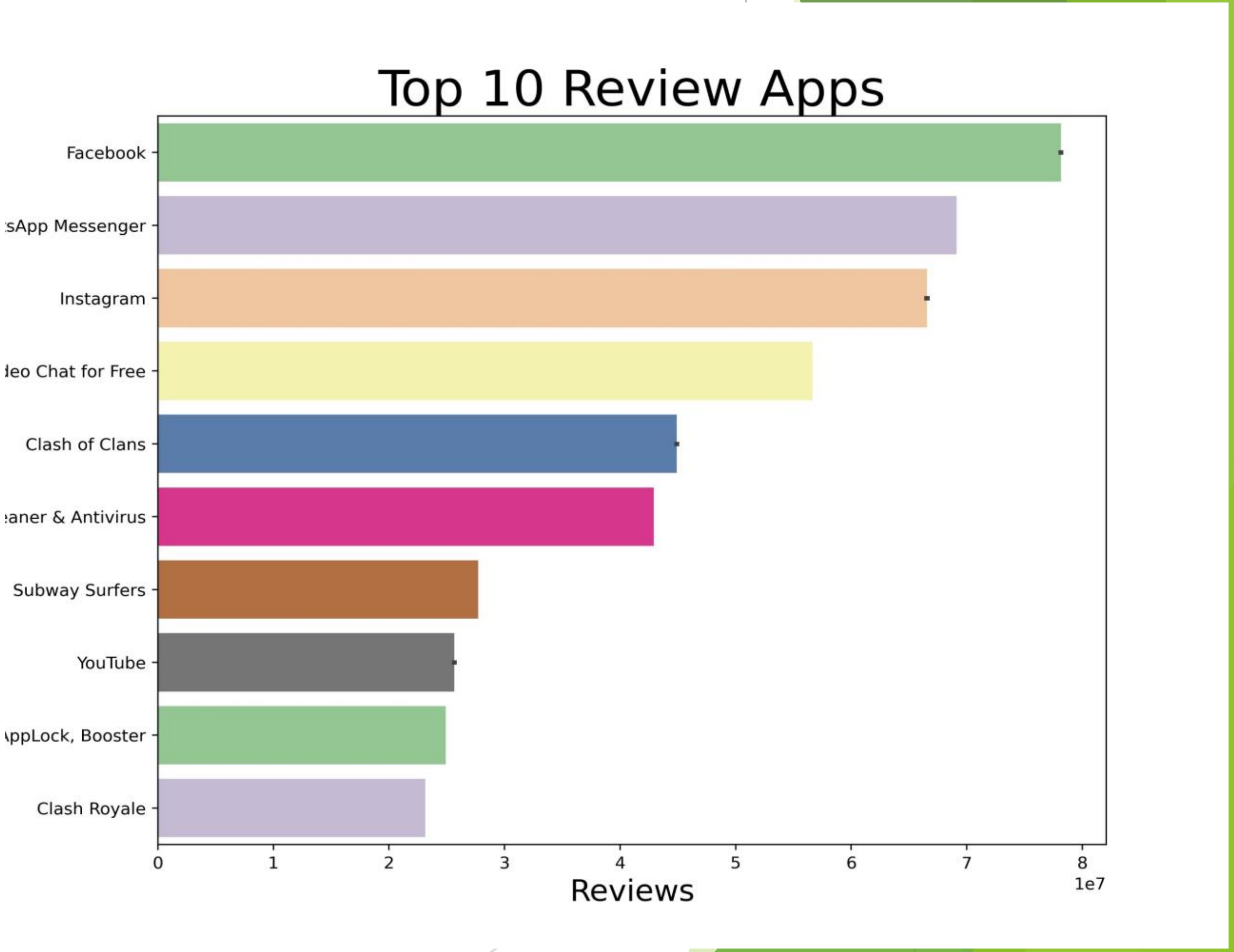
7. What are the top 10 installed apps in any category?

From the above graph

we can see that in the "ART_AND_DESIGN" category's "Sketch - Draw & Paint" has the highest install and Text on Photo-Fonteee has lowest install.



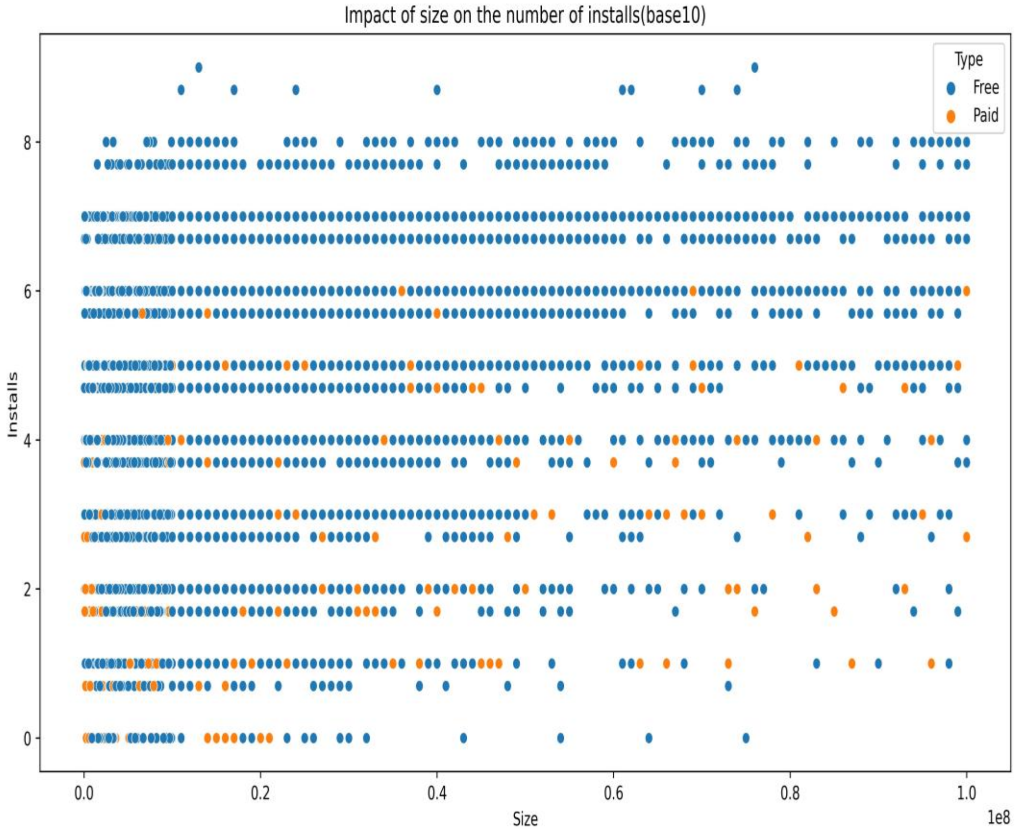
8. Which are the Apps having the highest numbers of reviews?



From the above plot

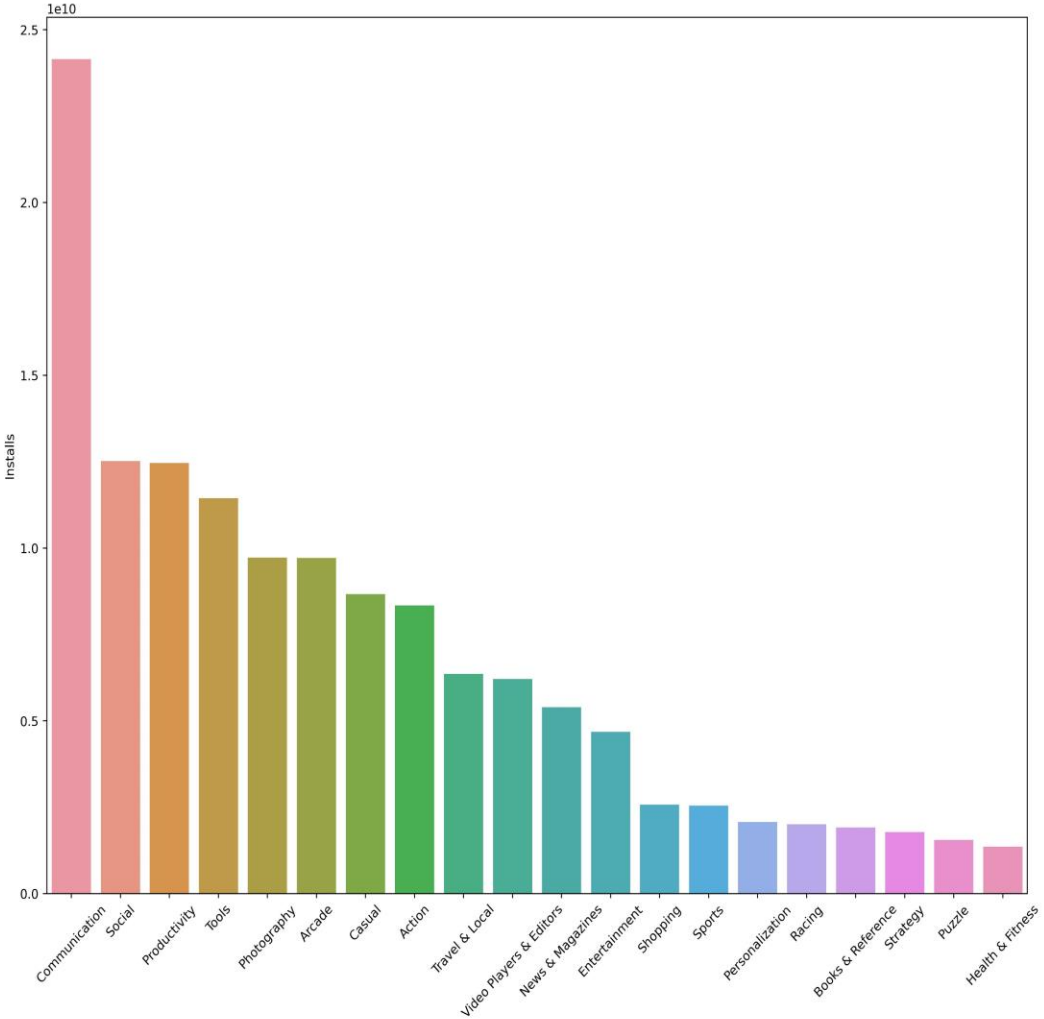
we can see that Facebook has the highest number of reviews and Clash Royale has lowest number of reviews.

9. How does size impact on the number of installs of any application?



It is clear from the above mentioned plot that size may impact the number of installations. Bulky applications are less installed by the user.

10. Which are the Genres that are getting installed the most in top 20 Genres?

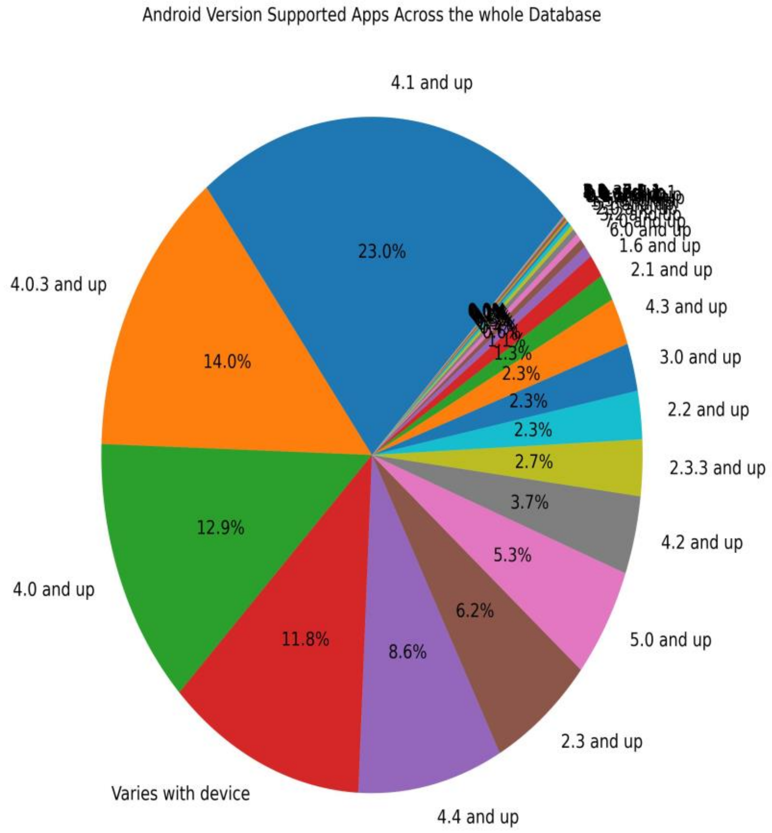


From the above plot we can come to the conclusion that maximum app install comes under Comunication Genres and followed by Social, Productivity and Tools Genres

11. Most Android Ver supported apps in play store.

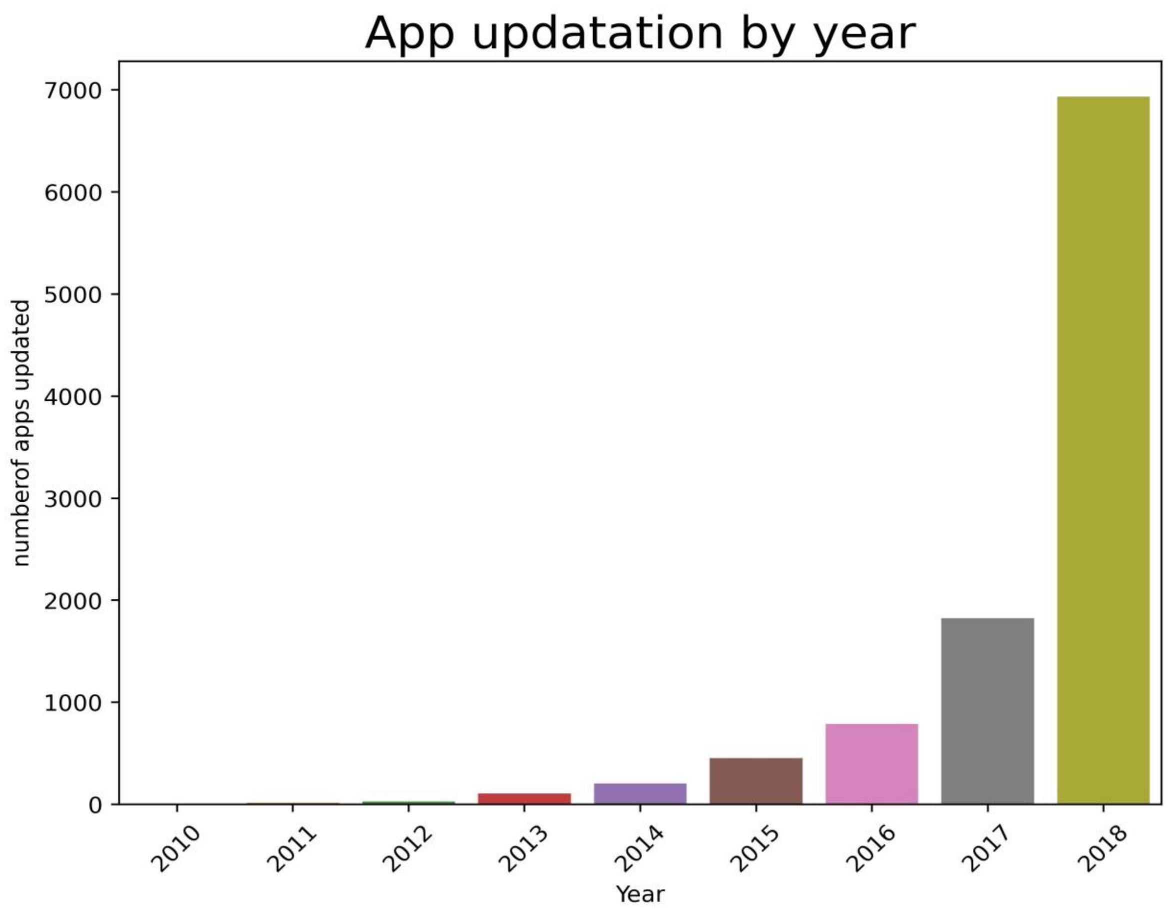
After identify total distribution percentage on data, given details of more app supported Android OS version Basically android 4.0

and above version supported app ratio is very higher and more then 60% apps support only on android 4.0 and above.



12. App update details "By Year".

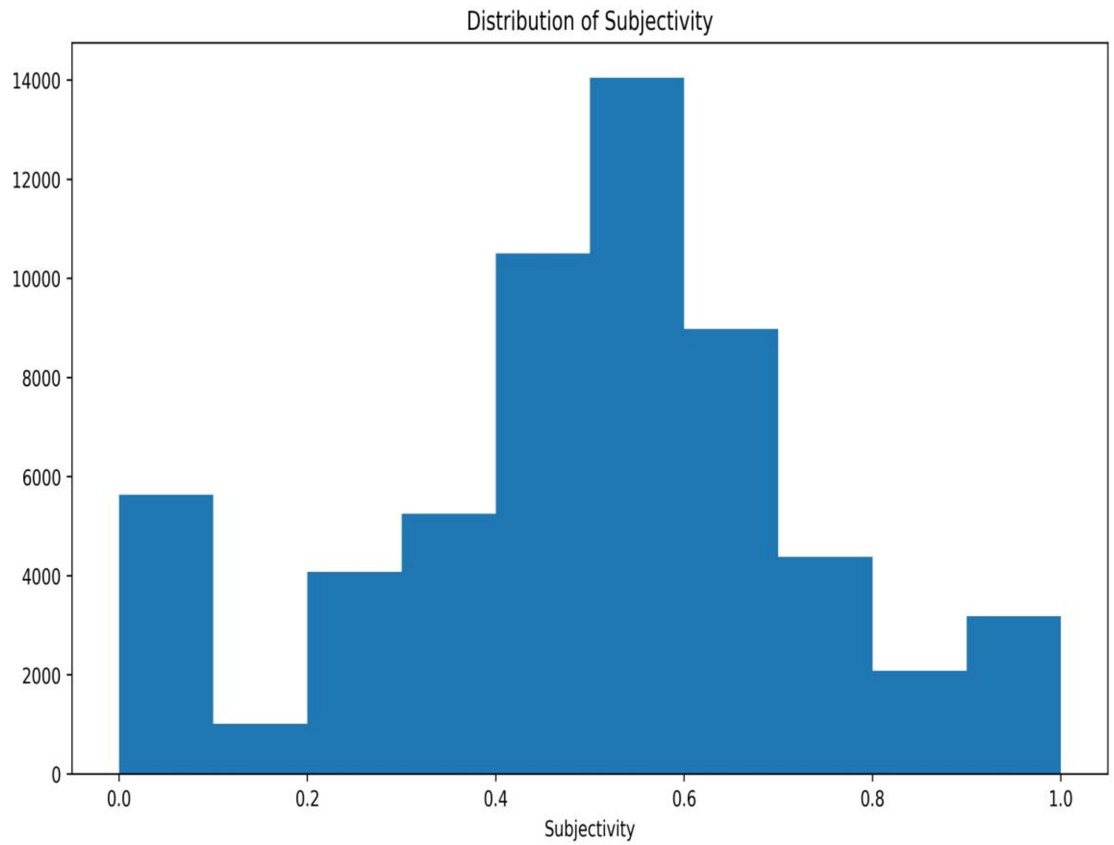
From this barplot we can see that a very wide range of app updated in play store during 2017-2018



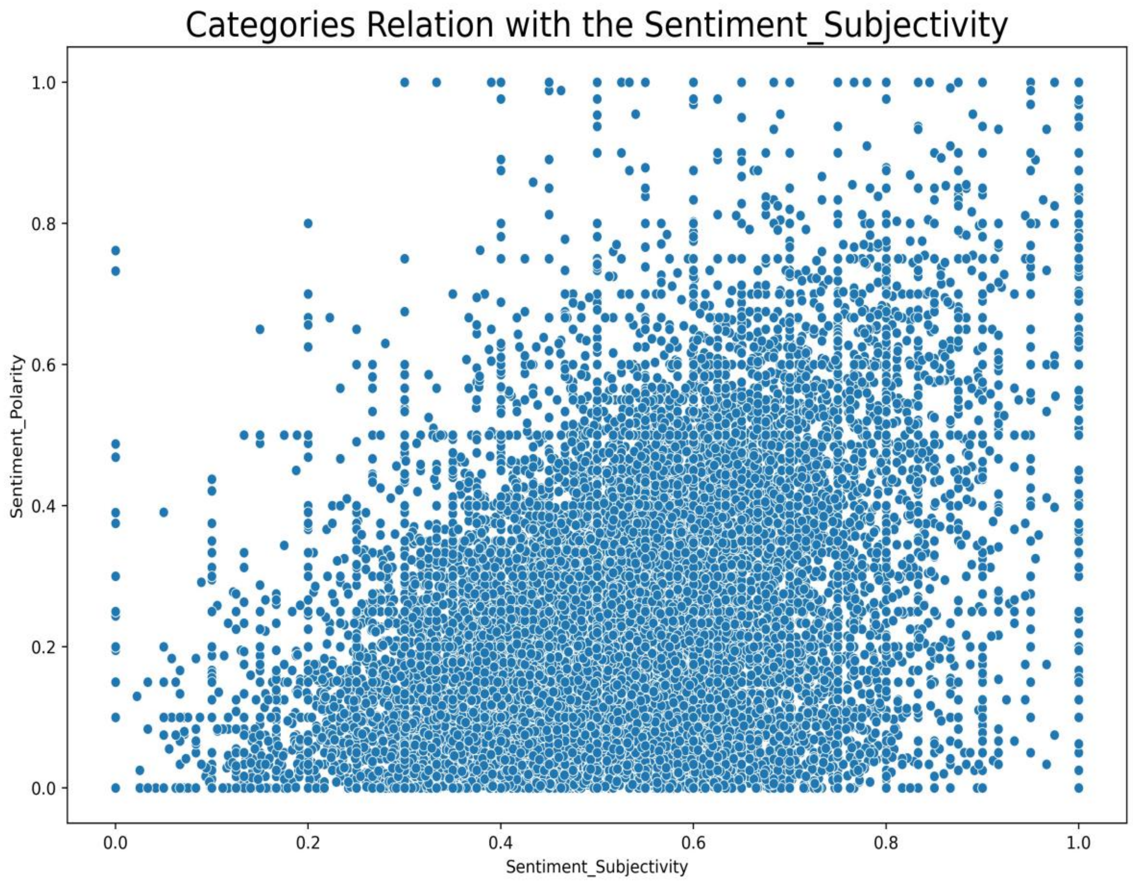
13. Histogram of subjectivity.

From this graph it can be seen that maximum number of sentiment subjectivity lies between 0.4 to 0.7.

So we can conclude that maximum number of users give reviews to the applications, according to their experience.



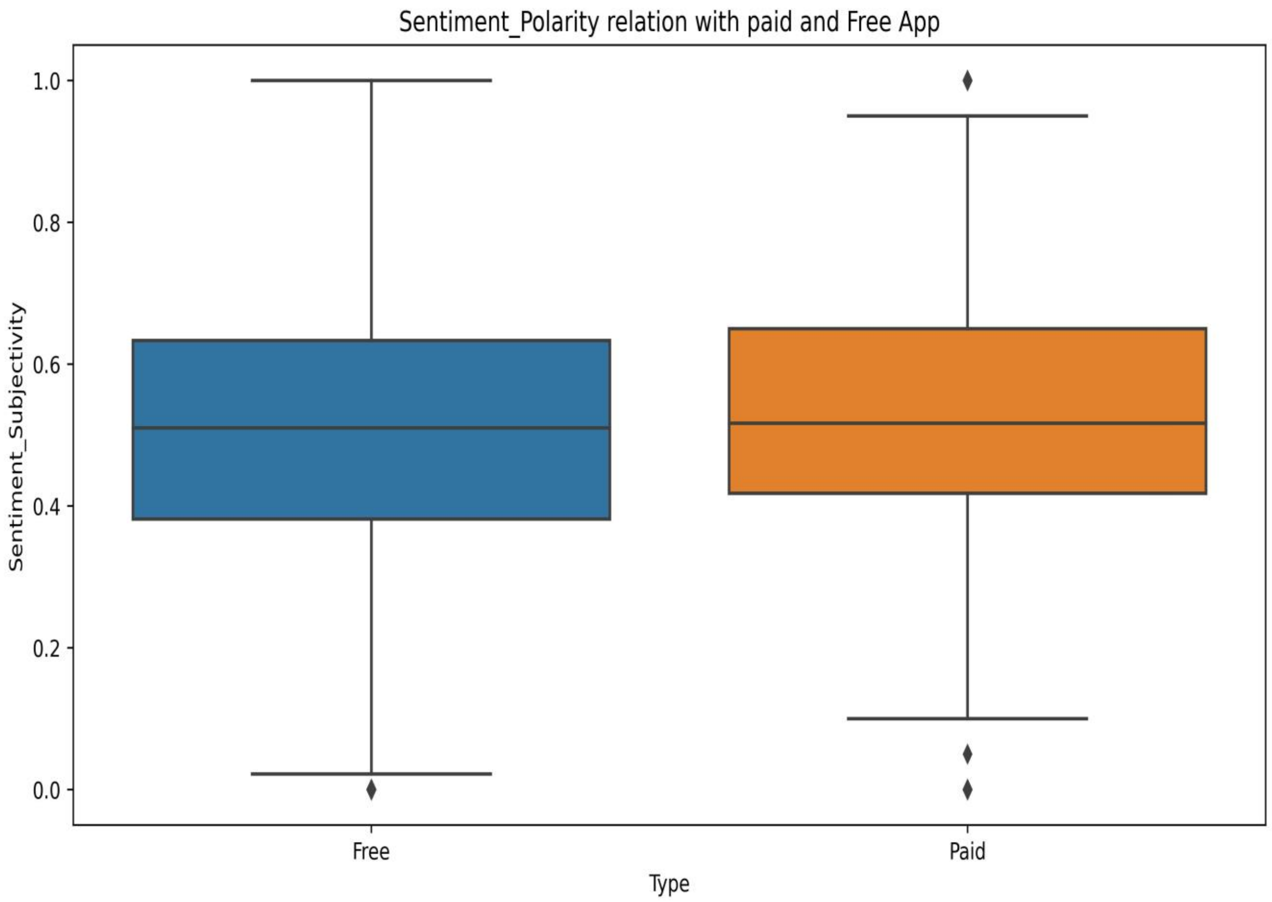
14. Is sentiment subjectivity proportional to sentiment polarity?



From the above scatter plot

it can be concluded 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.

15. Sentiment_Polarity relation with paid and Free App.

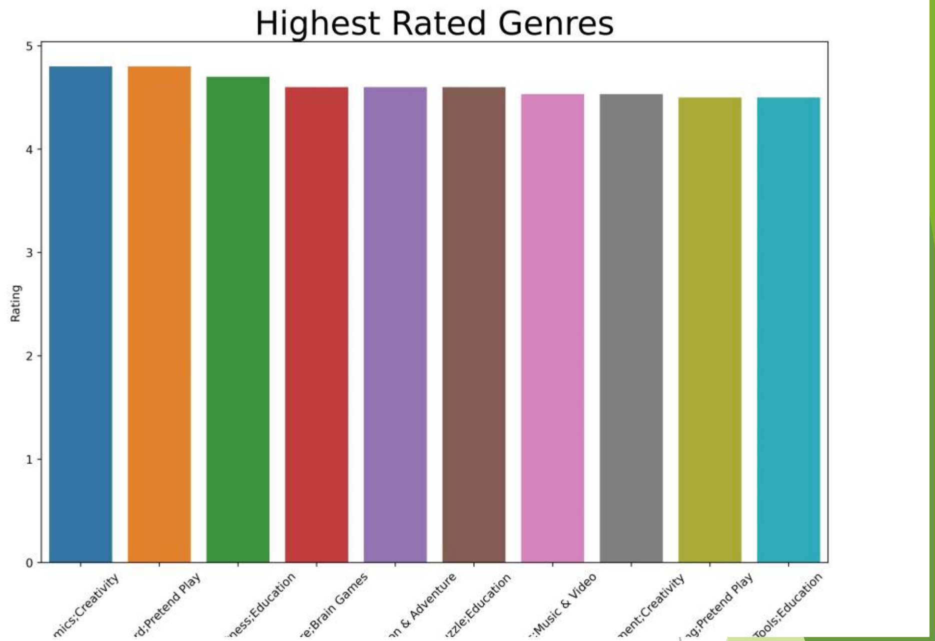


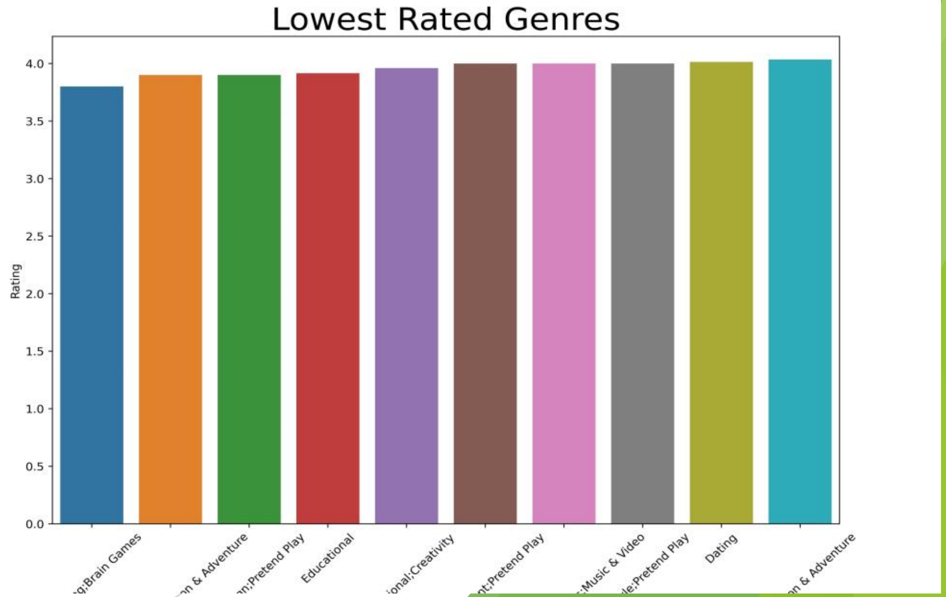
16. Find the highest and the lowest rated Genres.

From the above graph

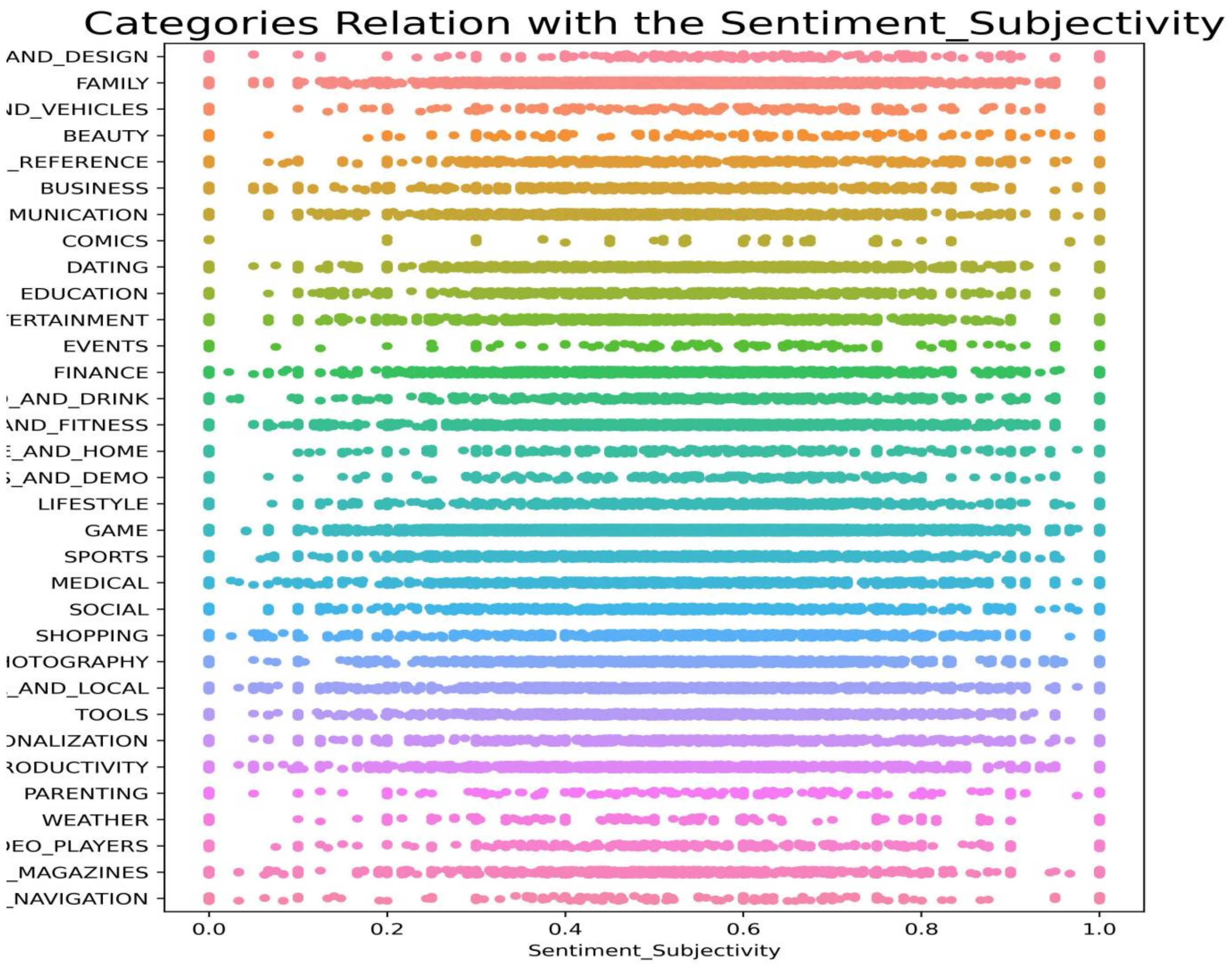
we can see that Comics;Creativity and Board;Pretend Play are the highest rated genres

and Parenting:Brain Games is the lowest rated genres





17. Categories Relation with the Sentiment_Subjectivity



CONCLUSION

1. In play store most of the apps are under Family & Game category and least are of Beauty & Comics Category.
2. 92.6% apps are Free and 7.4% apps are paid in type.
3. Category with the highest number of installs is Game and least number of installs is EVENTS categories.
4. Most of the apps in the google play store are rated between 3.8 to 4.8.
5. Education category has a highest mean rating of 4.37 and Dating category has lowest 4.01 rating.
6. Family, Game and Tools are top three categories having 1943, 1121 and 843 app count.
7. Everyone category has the highest number of apps and Unrated category has the lowest number of apps.
8. maximum number of applications present in the dataset are of small size.
9. Facebook has the highest number of reviews.
10. Bulky applications are less installed by the user.
11. Comunication Social, Productivity and Tools are top Genres.
12. Overall sentiment count of merged dataset in which Positive sentiment count is 64%, Negative 22% and Neutral 14%.
13. Maximum number of sentiment subjectivity lies between 0.4 to 0.7.
14. It's good to develop a Free type app and having a content rating for Everyone.
15. Comics;Creativity and Board;Pretend Play are the highest rated genres and Parenting:Brain Games is the lowest rated genres

THANK
YOU