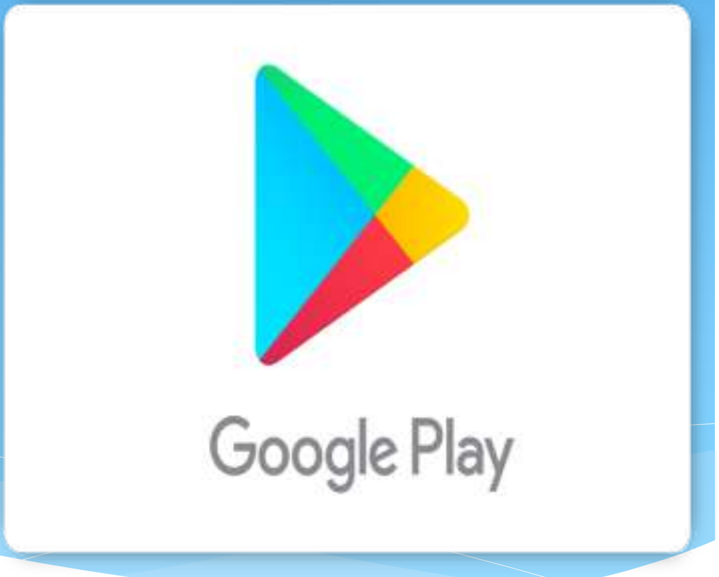
CAPSTONE PROJECT – 1

Google Play store App reviews Analysis



Presented By

SANJU KHANRA

Data Science Trainee, Alma Better

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

Contents

1.Introduction

2.Objective

3.Problem Statement

4.Description Of Data

5.Cleaning The Data

6.Data analysis and Visualization

7.Conclusion

Introduction

Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy.

Google Play store featured more than 5 million Android applications . Android is the most popular operating system in the world. With over 3bilion

Active user spanning over 190 countries. Applications are available through Google Play either for free or at a Paid.

Android applications, Games, Books, Movies and TV shows, Device updates

Lots of things are available.

Objectives

- Understand Consumer conduct and demand ,how they reacts to different Category and genres of Google Play store Apps
- Find the most popular and trending apps in recent times
- Find how small changes or update impact on app performances
- Analyze the Reviews ,Rating, Sentiments of People towards Various apps in play store
- Above all, help developers or client to recognize the gap, make the app better and meet customer expectations.

- Comparing different categories of applications based on the Android version.

Comparing the rates in different kinds of applications.

Problem statement

1. *Top categories on Google Play store?*
2. *Which category of App's have most number of installs?*
3. *How much percentage of apps are Free and Paid?*
4 . *What category of apps from the content Rating column are found more on play store?*
5. *Let's have a look at the Distribution of the ratings of the apps in data frame?*
6. What are the Top 10 installed apps in any category?
7. What are the Top 10 expensive Apps in play store?
8. What are the Apps with highest number of reviews?
9. What are the count of Top 20 Apps in different genres?
10. Which are the Genres that are getting installed the most in top 20 Genres?
11. Find the highest and the lowest rated Genres
12. App update details "By Year"
13. i) Before merging Which Category has highest number of average rating?
    ii) User reviews after merging Which Category has highest number of average rating?
14. Correlation Heat map
15. Pair Plot
16. Distribution of Sentiment subjectivity
17. Categories Relation with the Sentiment Subjectivity
18. Sentiment Polarity relation with paid and Free App
19. Which are the apps that have made the highest earning?
20. Let's have a look at the distribution of the Size of the data frame
21. Are Paid apps worth buying? (Analysis based on Average User Rating)

Description of Dataset

There are two dataset: Play store Data & User data

Play Store Data :-

App:- Name of the Application

Category :- Category of the Application

Rating :- Rating given to the Application

Reviews:- No of reviews given to the Application

Size:- Size of the Application

Installs:- No of downloads of the Application

Type:- Free or Paid Apps

Price:- Price of the Application if it is Paid

Content Rating:- It is Age appropriate or Not

Genres:- Type of Genres the Application belongs to

Last Update :- When the last time the Application is Updated

Current Version :- Current version of the Application

Android Version :- Minimum Android version required to run the Application

Description of Dataset

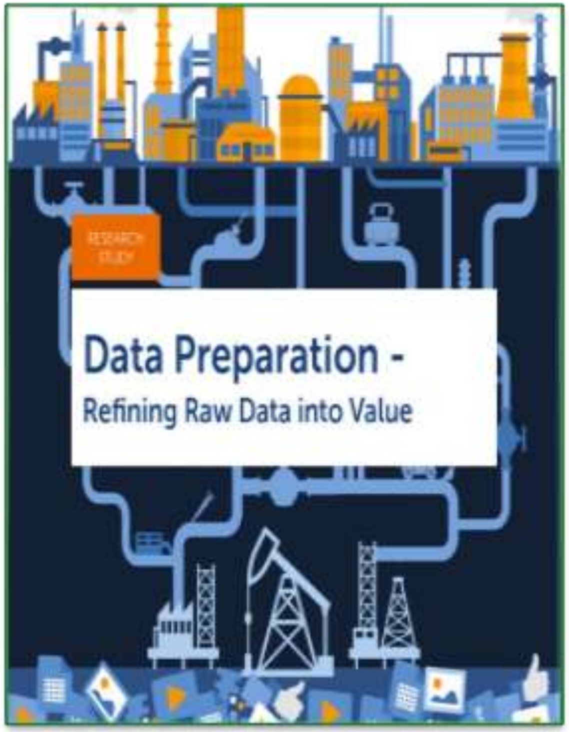
2. User Review Data :-

- **App –** An app name .
- **Translated Review –** Reviews being given by consumer.
- **Sentiment –** Sentiment given to an app by users(i.e. Positive, Neutral, Negative)
- **Sentiment Polarity –** The polarity of sentiment measures how negative or positive the context is. In the data we have, the polarity ranges from +1(Positive) to -1(Negative).
- **Sentiment Subjectivity –** The subjectivity of a sentiment is how likely that sentiment is to be based on data or factual information, versus personal opinions or public what is thinking exactly and public actually satisfied or not.

Dataset Preparation

Loading the data sets: Two datasets, First Play store app dataset and User Reviews dataset.

- **Import Libraries:** NumPy, Pandas, Seaborn and Matplotlib
- **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.



Data Cleaning

Data cleaning not just means removing the incorrect data or erroneous data. Many times we get the data which has all kinds of values some of them will cause problems during the analysis of the data and make our predictions incorrect.so we have to make sure our data has no erroneous values.

Data cleaning step : -

Removing Unwanted Values : Deleting of duplicate/incorrect or irrelevant or values use

Handling Missing Values : Handling missing values in our Dataset

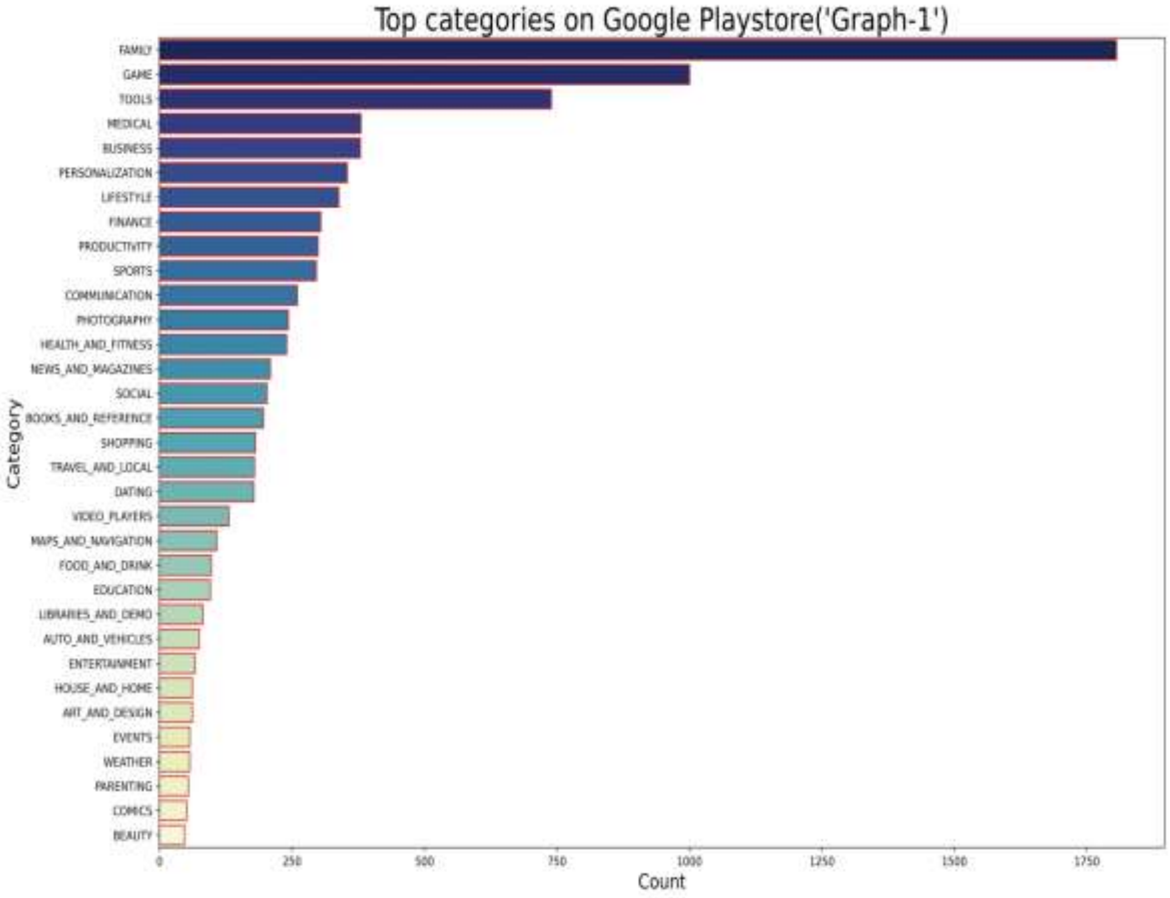
Handling Structural Errors : Fixing mislabeled categories or classes , Types , Strange , Name , conventions

Filtering Unwanted Outliers : Removing incorrect or unwanted outliers

Replacing Missing Values with mean and mode : Replacing missing values with mean is the most popular method of replacing missing values.

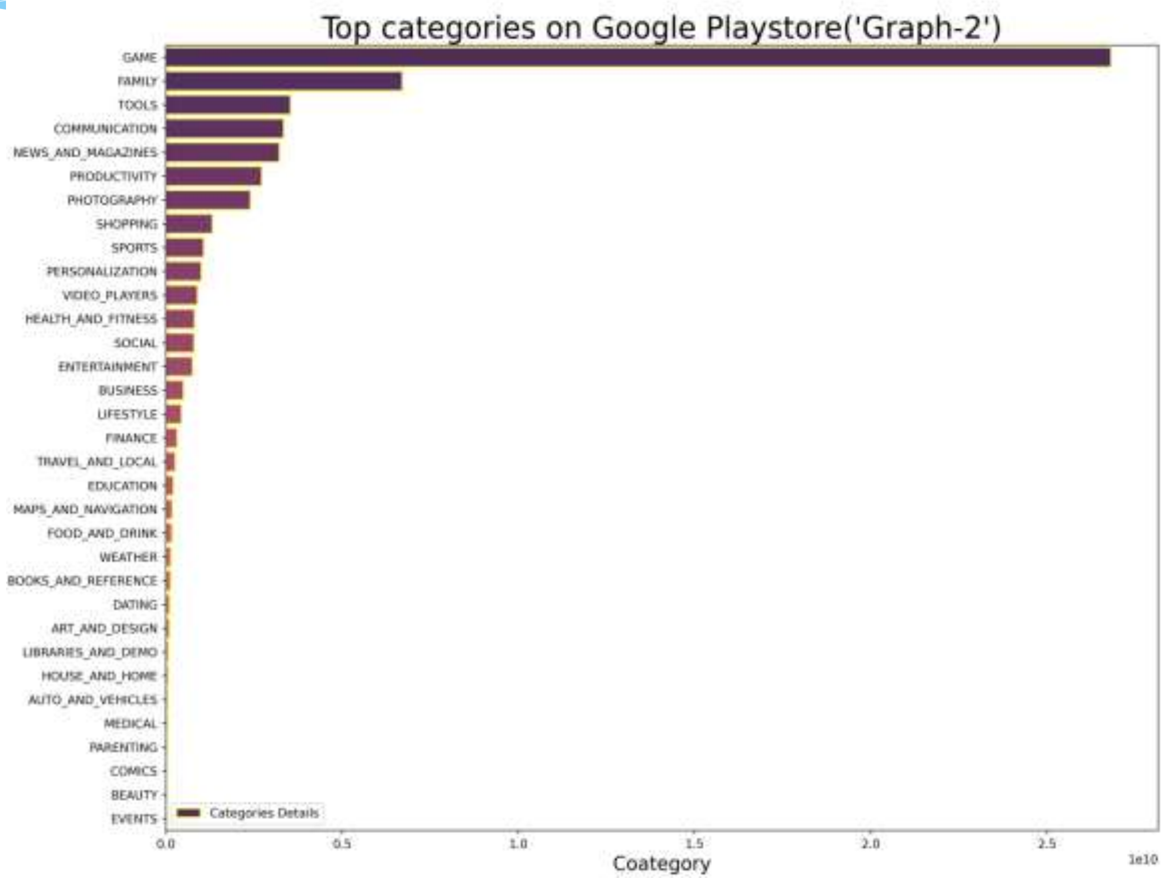
Data Analysis & Visualization

- **1. Top categories on Google Play store**
- So there are total 33 categories in the dataset From the above plot and we can come to a conclusion that in play store most of the apps are under Family & Game category and least are of Beauty & Comics Category.



Data Analysis & Visualization

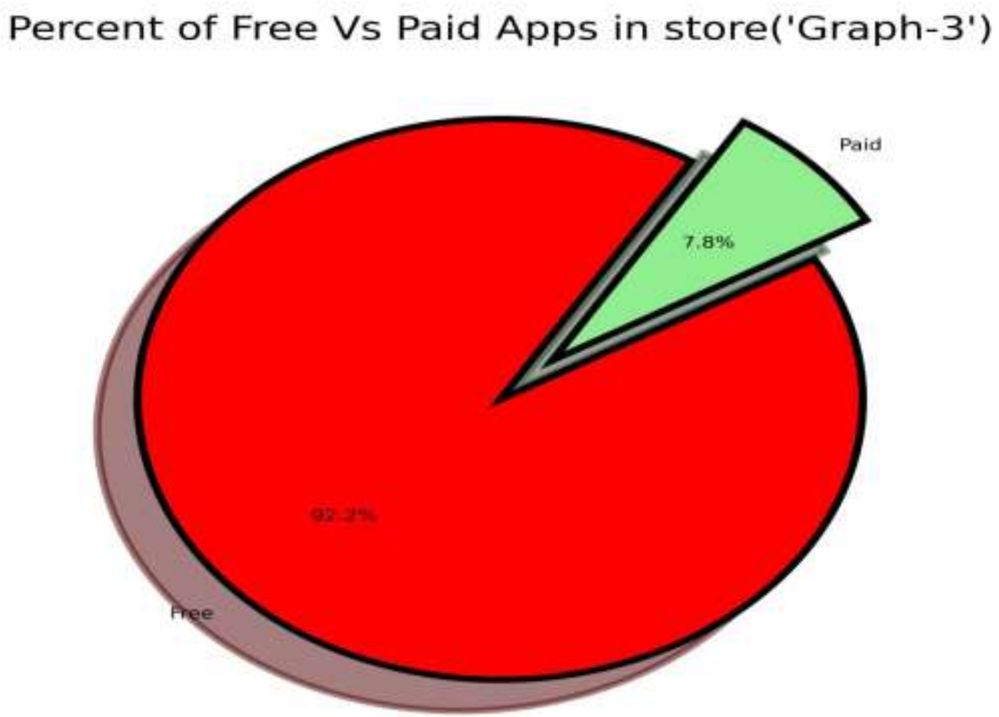
- **2. Which category apps have the most number of installs?**
- From the above graph we can see that there are total of 33 categories in the dataset. We can come to the conclusion that in the play store the top categories with the highest installs is- ”GAME” category and last is “EVENTS” categories.



Data Analysis & Visualization

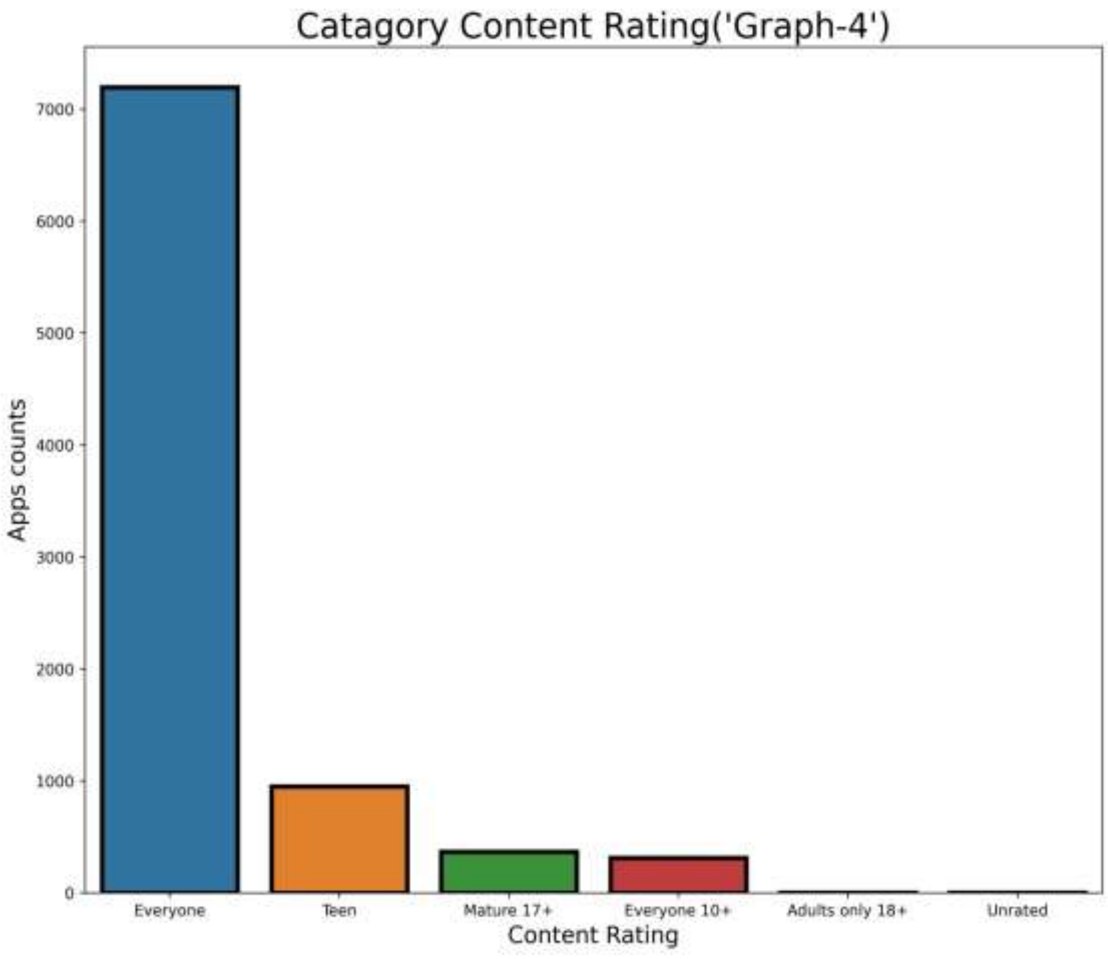
3. How much Types apps category percentage are paid or free ?

It is indicates most that 92.2% apps are free to download available and rest 7.8% are paid apps are buy Paid money.



Data Analysis & Visualization

4) Which category of Apps from the Content Rating column are found more on play store



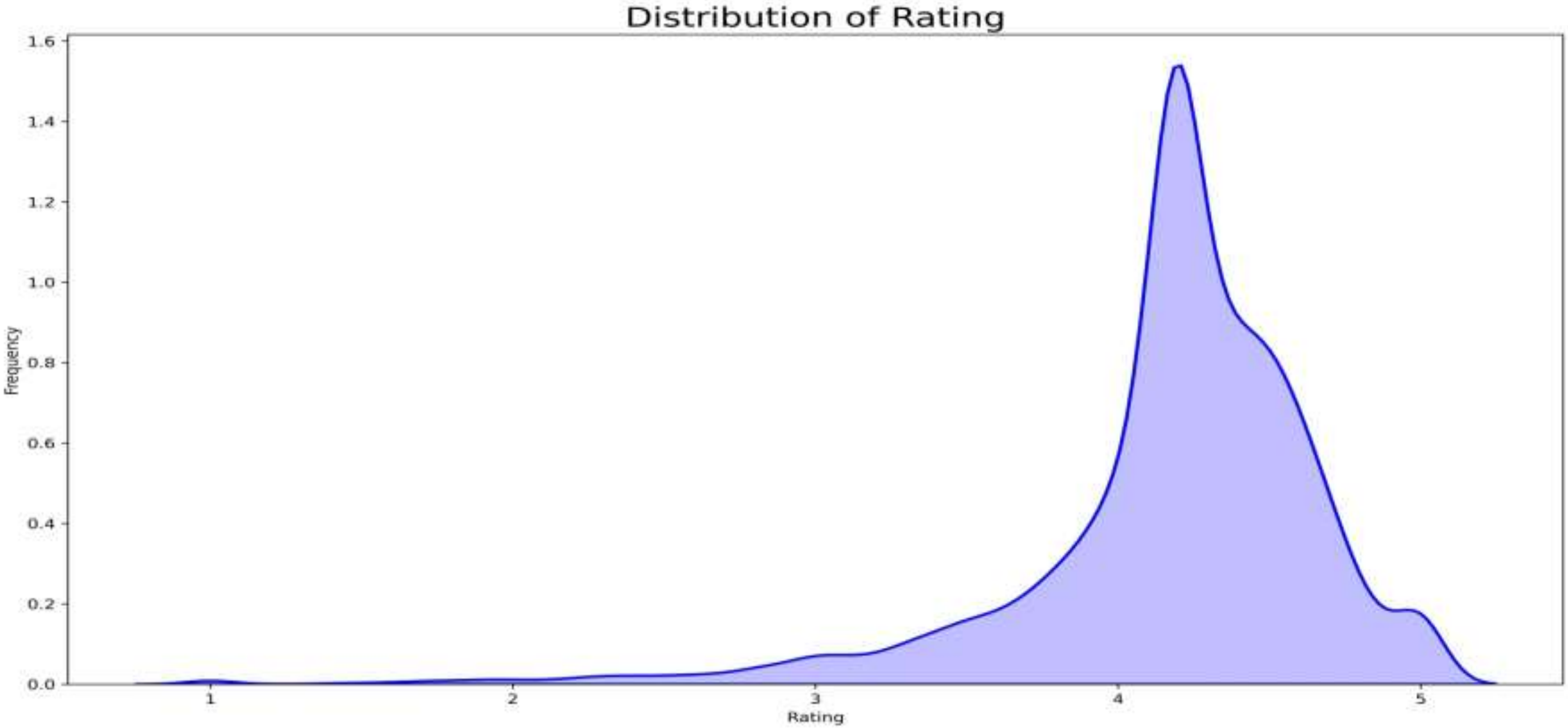
- we can see that the Everyone category has the highest number of apps and unrated category has the lowest number of apps.
- Everyone Content Rating Apps counts 7195 or unrated Apps counts is 2.

Data Analysis & Visualization

5 .Let's have a look at the distribution of the ratings of the data frame

Note: From the above graph we can come to a conclusion that most of the apps in Google play store are Rating

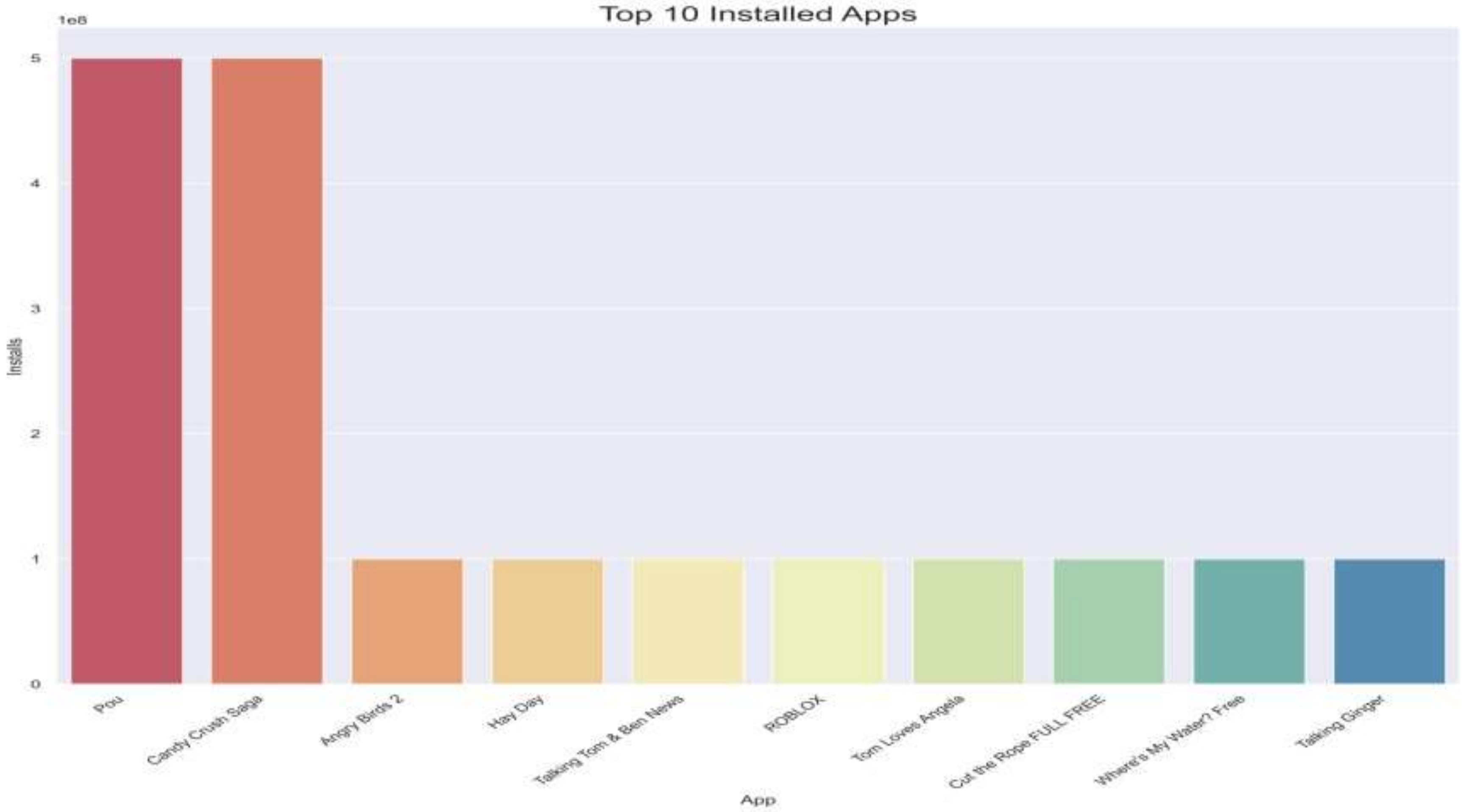
in between 3.5 to 4.8



Data Analysis & Visualization

6 : What are the Top 10 installed apps in any category?

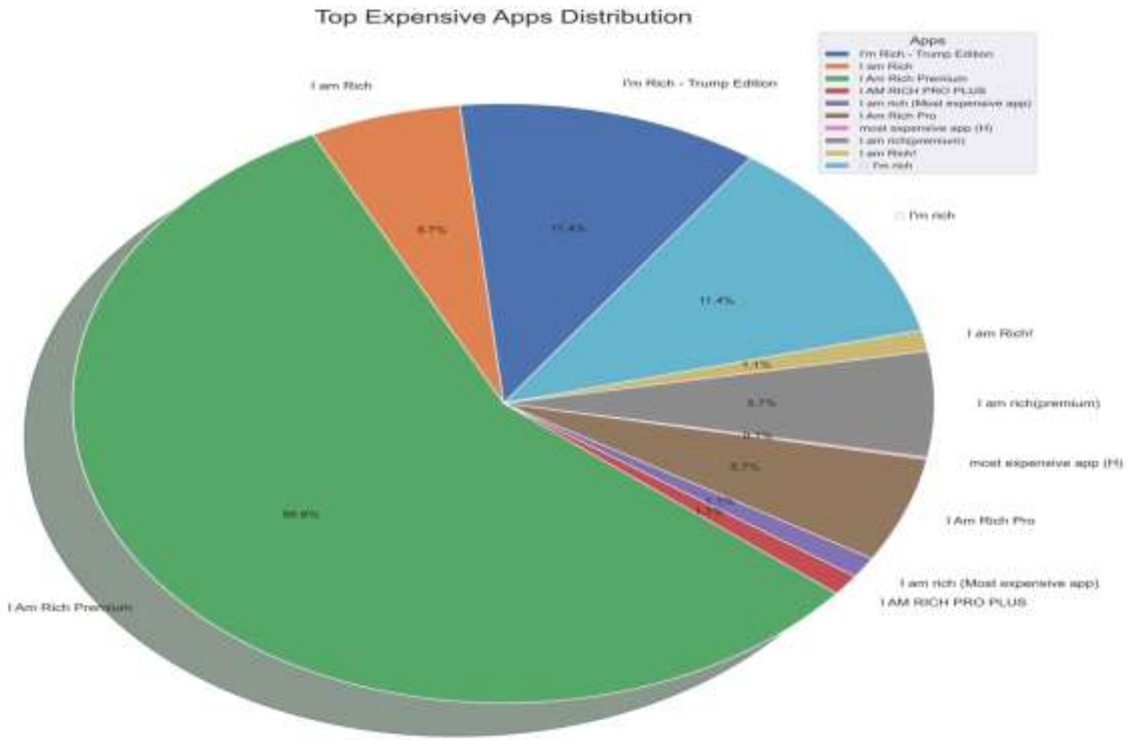
From the above graph we can see that in the FAMILY category POU, and CANDY CRUSH SAGA has the highest installs. In the same way we by passing different category names to the function, we can get the top 10 installed apps.



Data Analysis & Visualization

7 : Which are the top 10 expensive Apps in play store?

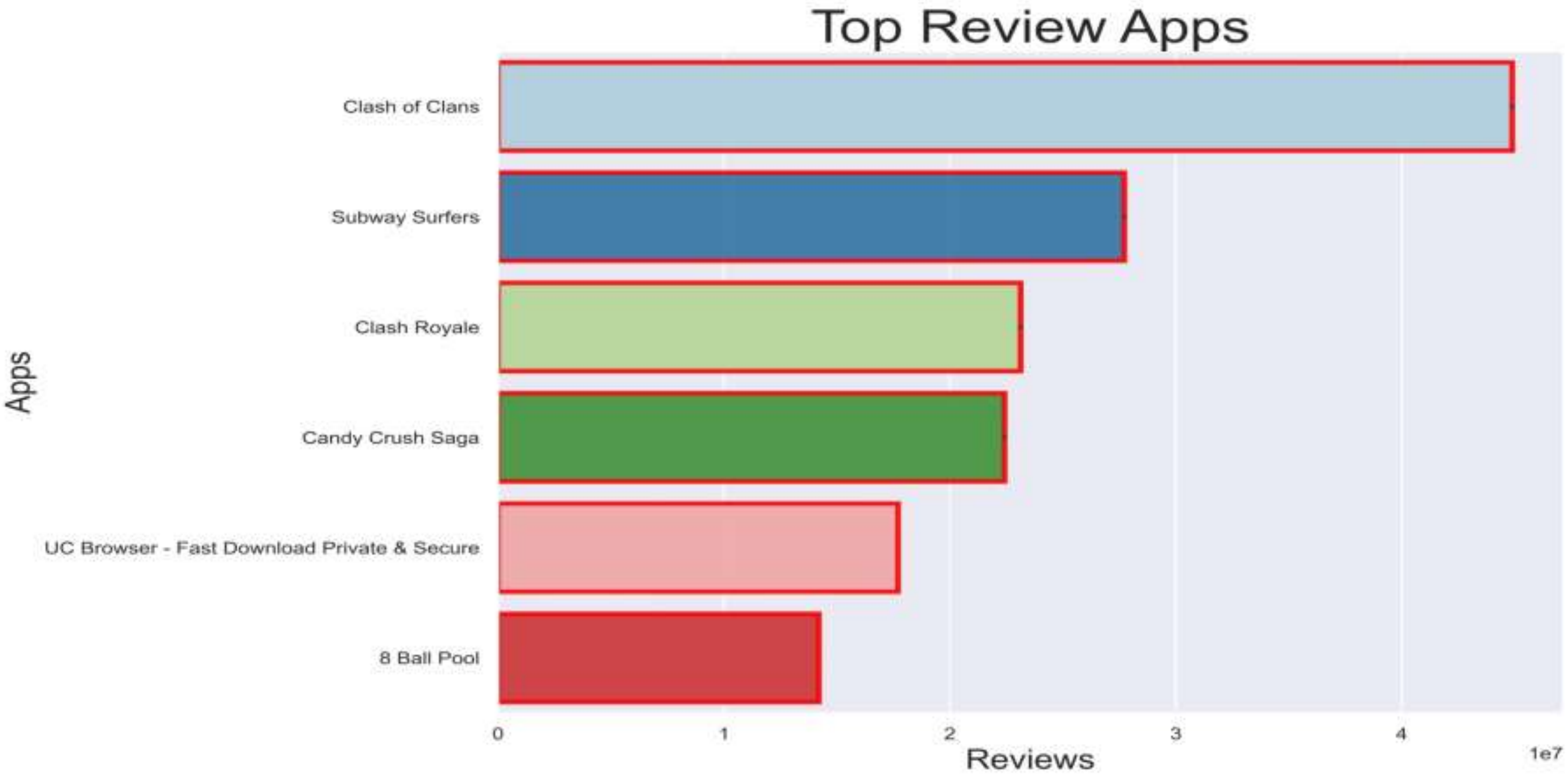
- From the above graph we can interpret that the App I Am Rich Premium is the most expensive app in the GOOGLE play store followed by I am Rich. I Am Rich Premium 56.8% app almost expensive. we also had to drop one row data for this visualization because the language of the app was Chinese and it was messing with the pie chart, In this data frame under 9934 row labels "I'm Rich/EU SOU Rico/أنا غني/我很有錢" in this app zero(0)times install that case this unwanted visualization



Data Analysis & Visualization

8 : Top the Apps with highest number of reviews?

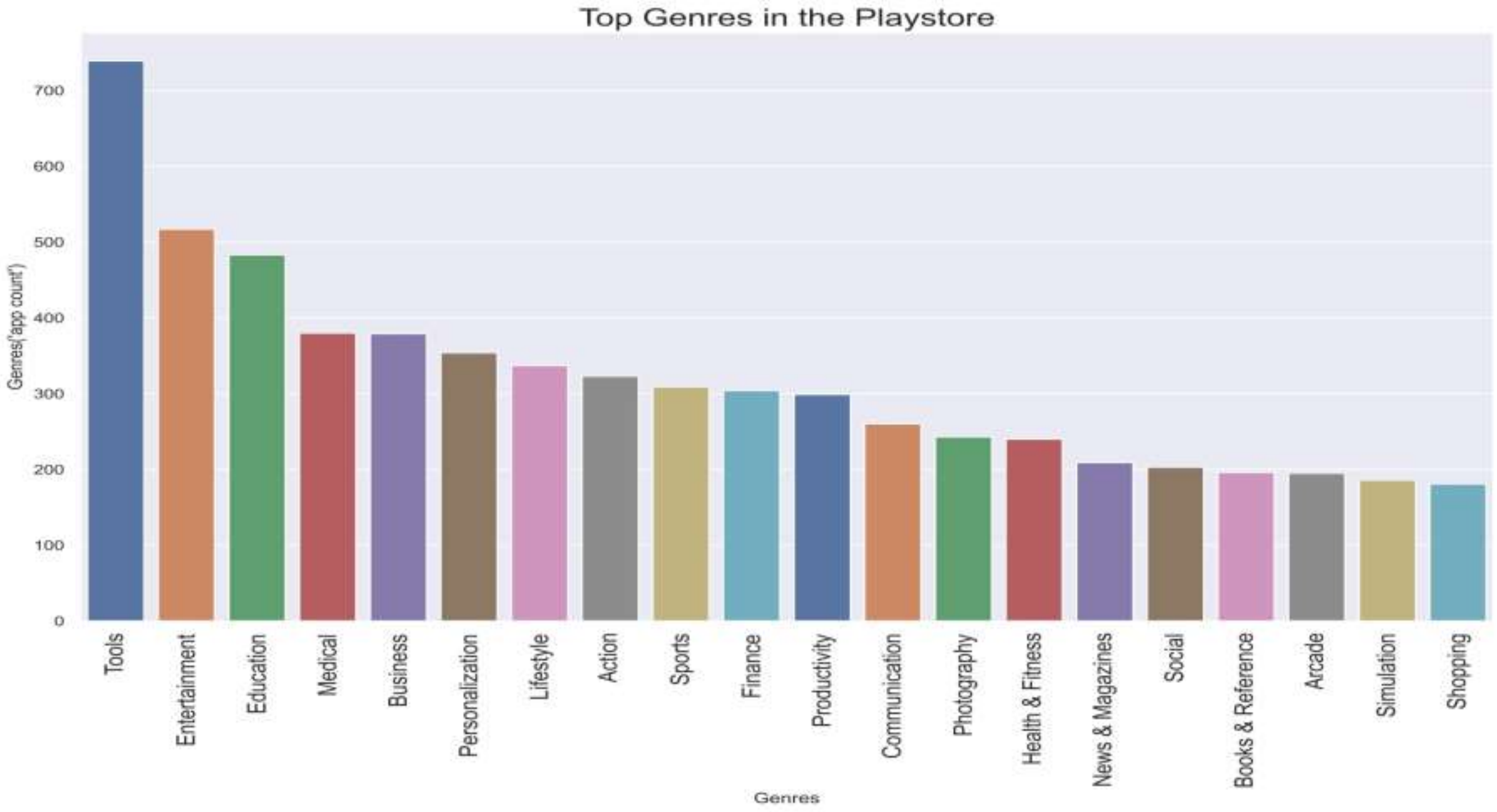
From the above data frame we can interpret, and come to conclusion that the Apps like Clash of Clans, Subway Surfers, Clash Royale, and Candy Crush Saga , UC Browser - Fast Download Private & Secure,8 Ball Pool, has the highest number of reviews on GOOGLE play store. This top apps under ,Clash of Clans, Subway Surfers are most reviews app. Clash of Clans 44893888 numbers of reviews apps and Subway Surfers 44891723 numbers of reviews app.



Data Analysis & Visualization

9 : What are the count of Top20 Apps in different genres?

After visualization we can see that the Highest Number of Apps found in the Tools and Entertainment genres followed by Education, Medical and many more . Tools Genres app count 739, Entertainment Genres app count 517, and Education Genres app count 483.

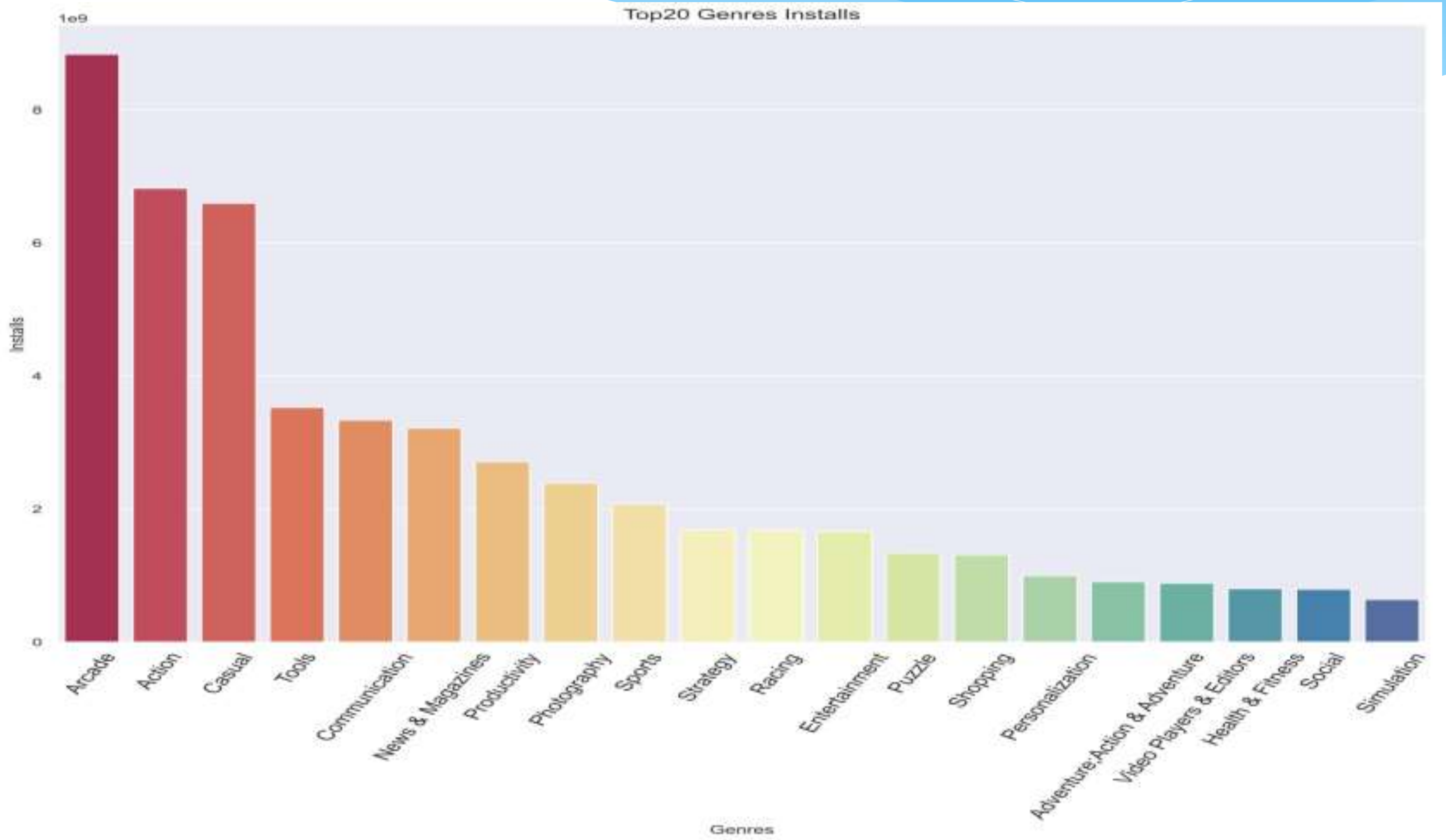


Data Analysis & Visualization

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

After visualization we can come to the conclusion that maximum app install comes under Arcade Genres and followed by Action, Casual and Tools Genres, Arcade Genres install 8836079153 times, and

Action Genres most install times 6818939040.



Data Analysis & Visualization

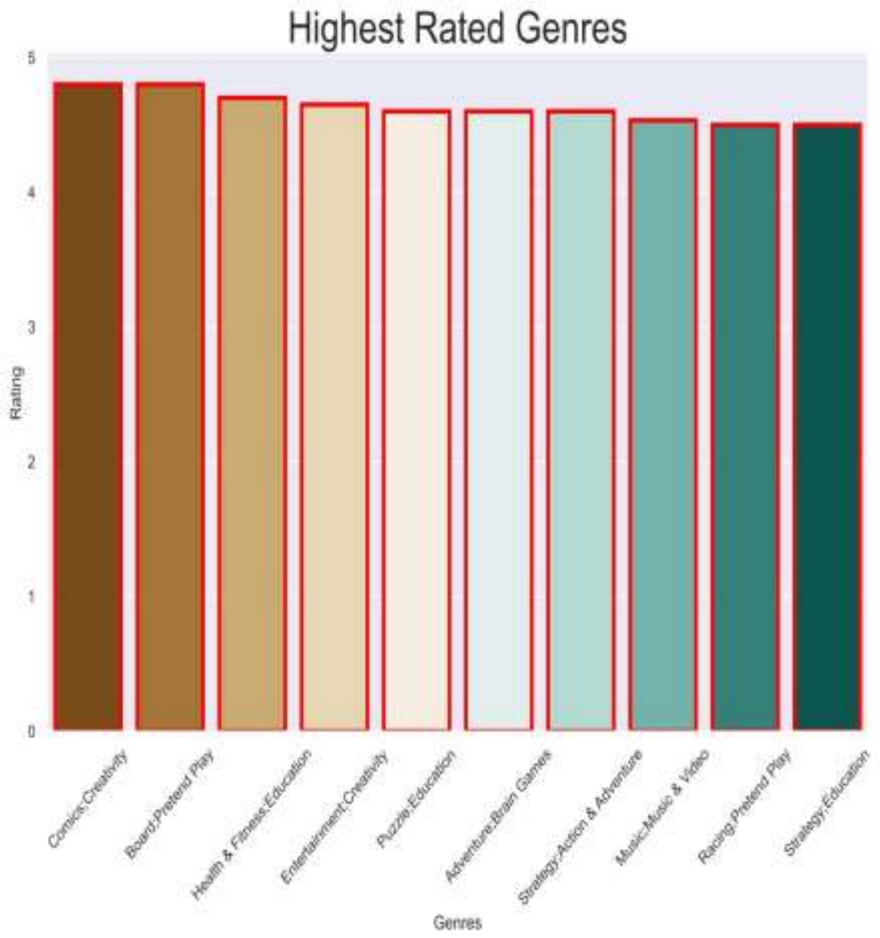
11A.Find the highest Genres

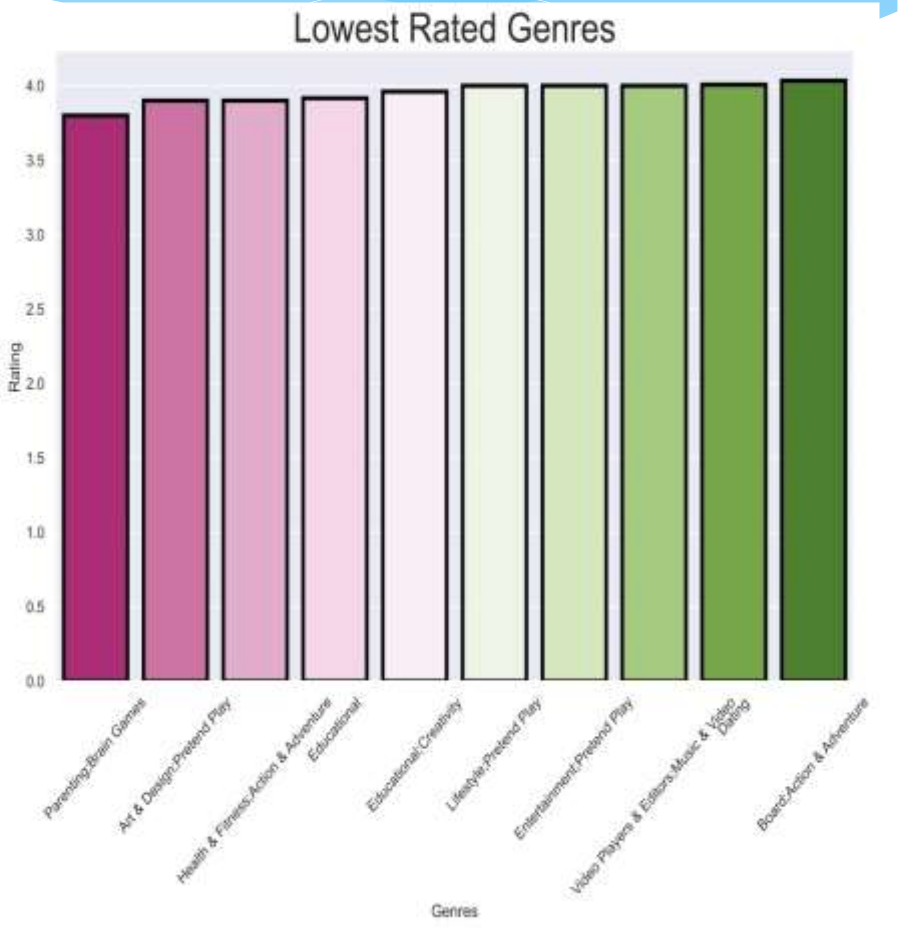
- From the above graph we can see that Comics : Creativity and

  Board : Pretend Play are the highest rated genres. Comics : Creativity Genres rating is 4.8000, Board : Pretend Play rating is 4.8000.

11B.Find the lowest rated Genres

- graph we can see that Parenting : Brain Games is the lowest rated genres. Brain Games is the lowest rated genres this rating is 3.800

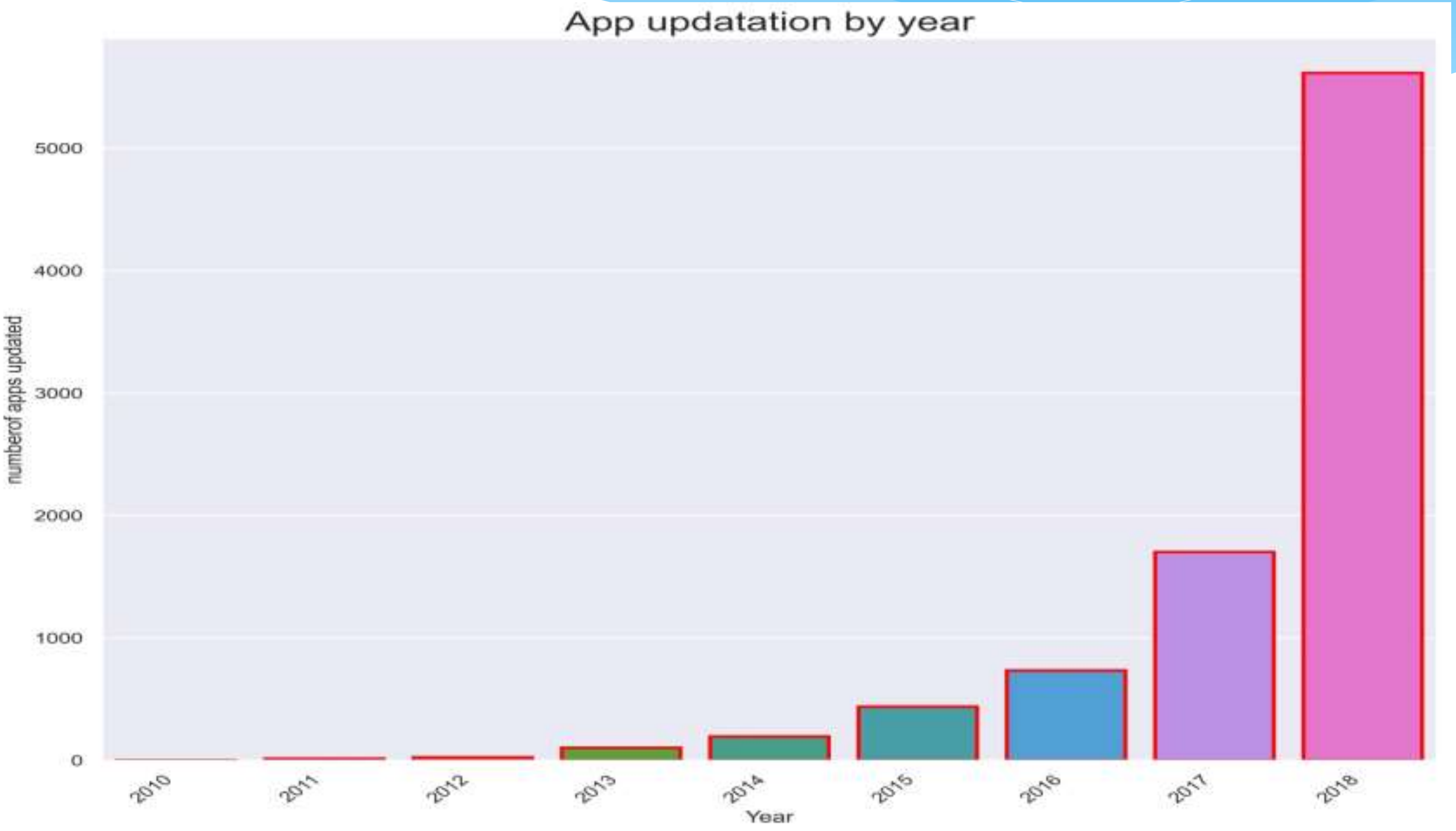




Data Analysis & Visualization

12. What year more Apps update details "By Year”

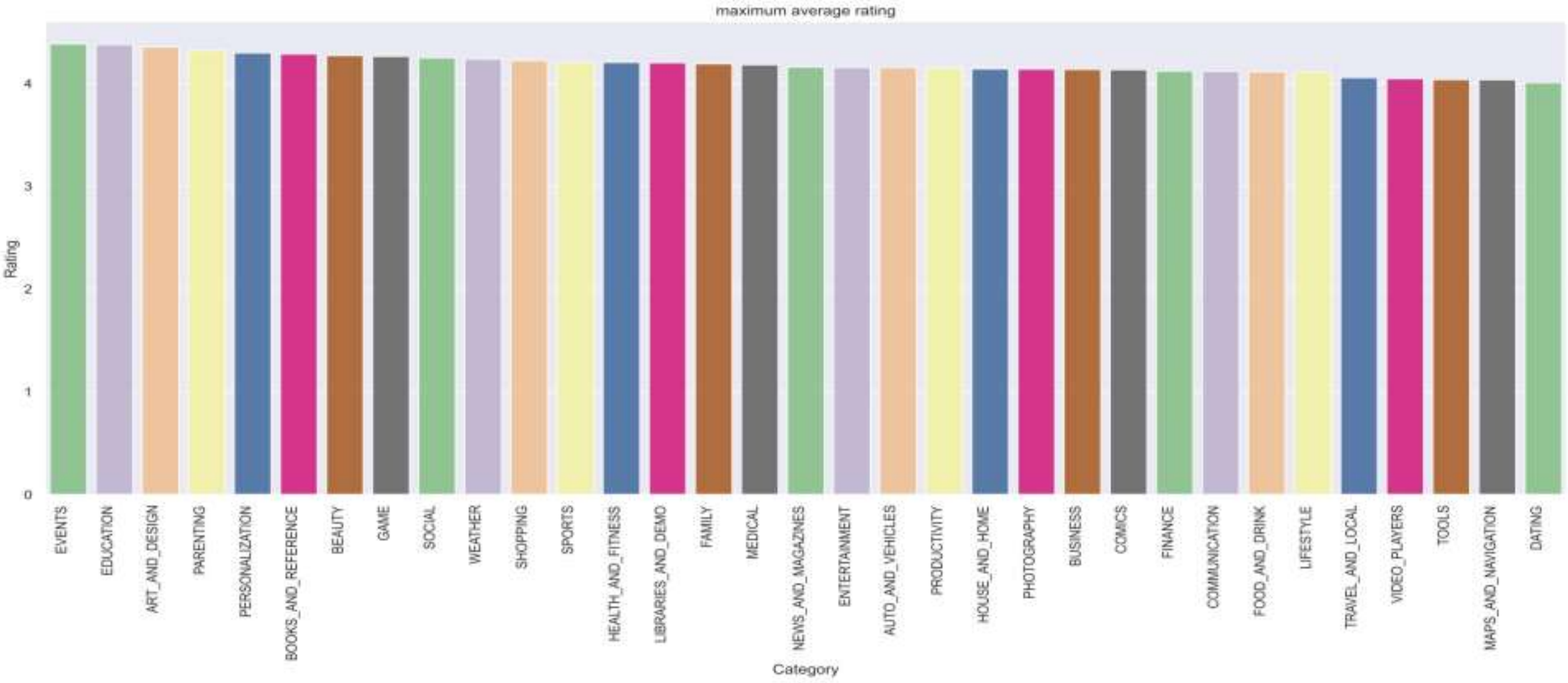
- From this plot we can see that a very wide range of app updated in play store during 2017-2018, Actually in 2017-2018 under Almost More of the apps will update new version and in this 2018 year 5615 number of apps will be updated, and in this 2017 year 1720 numbers of app will be updated.



Data Analysis & Visualization

13 (i) . Before Merging Which Category has highest number of average rating?

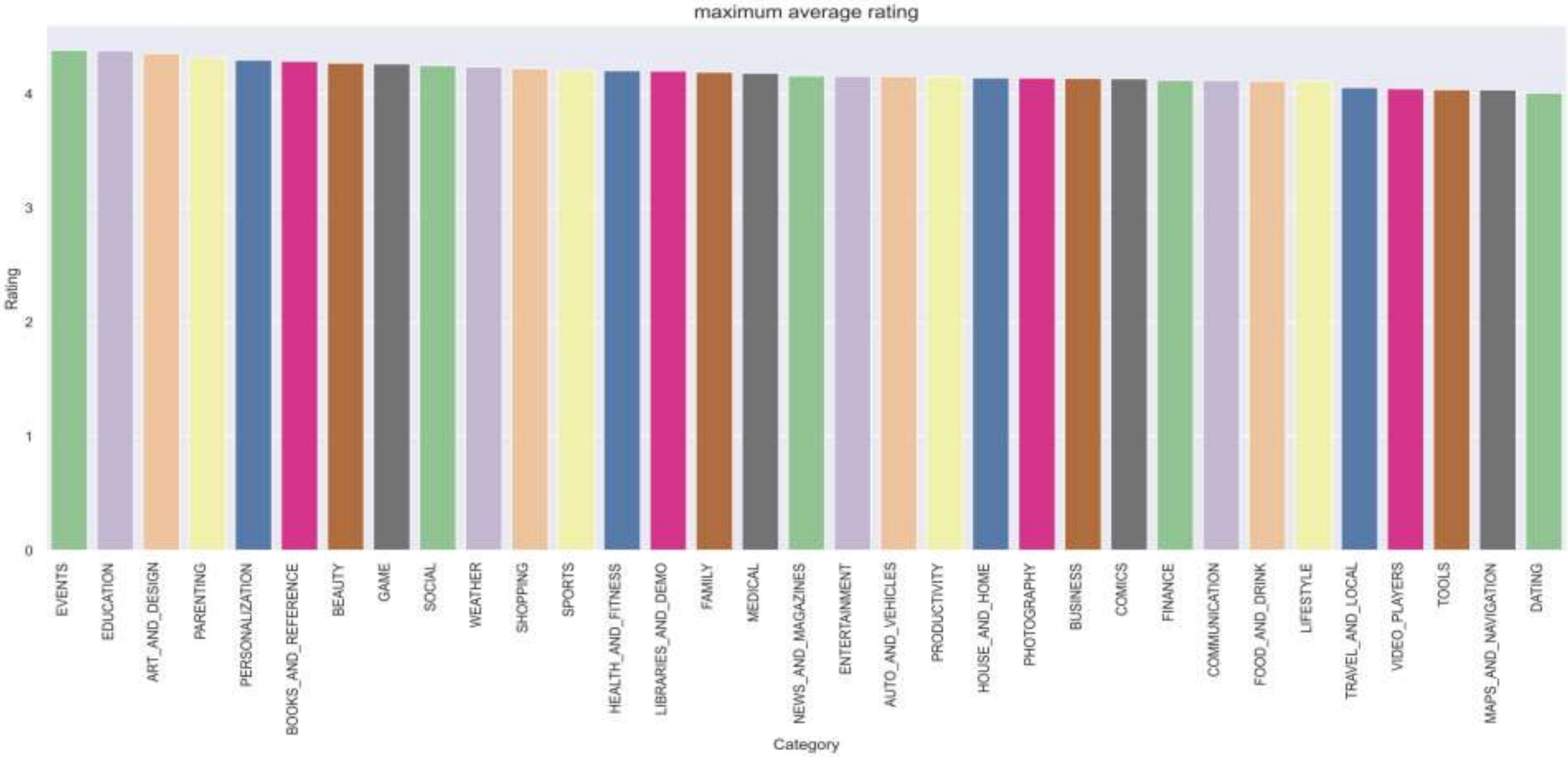
- From the above graph we can see that in the "EVENTS" category has the highest average rating. Category of EVENTS Rating 4.381924, and Category of EDUCATION Rating 4.377999. before merging play store category rating all 33Categories average rating is 4.183



Data Analysis & Visualization

13 (ii).User reviews after merging Which Category has highest number of average rating?

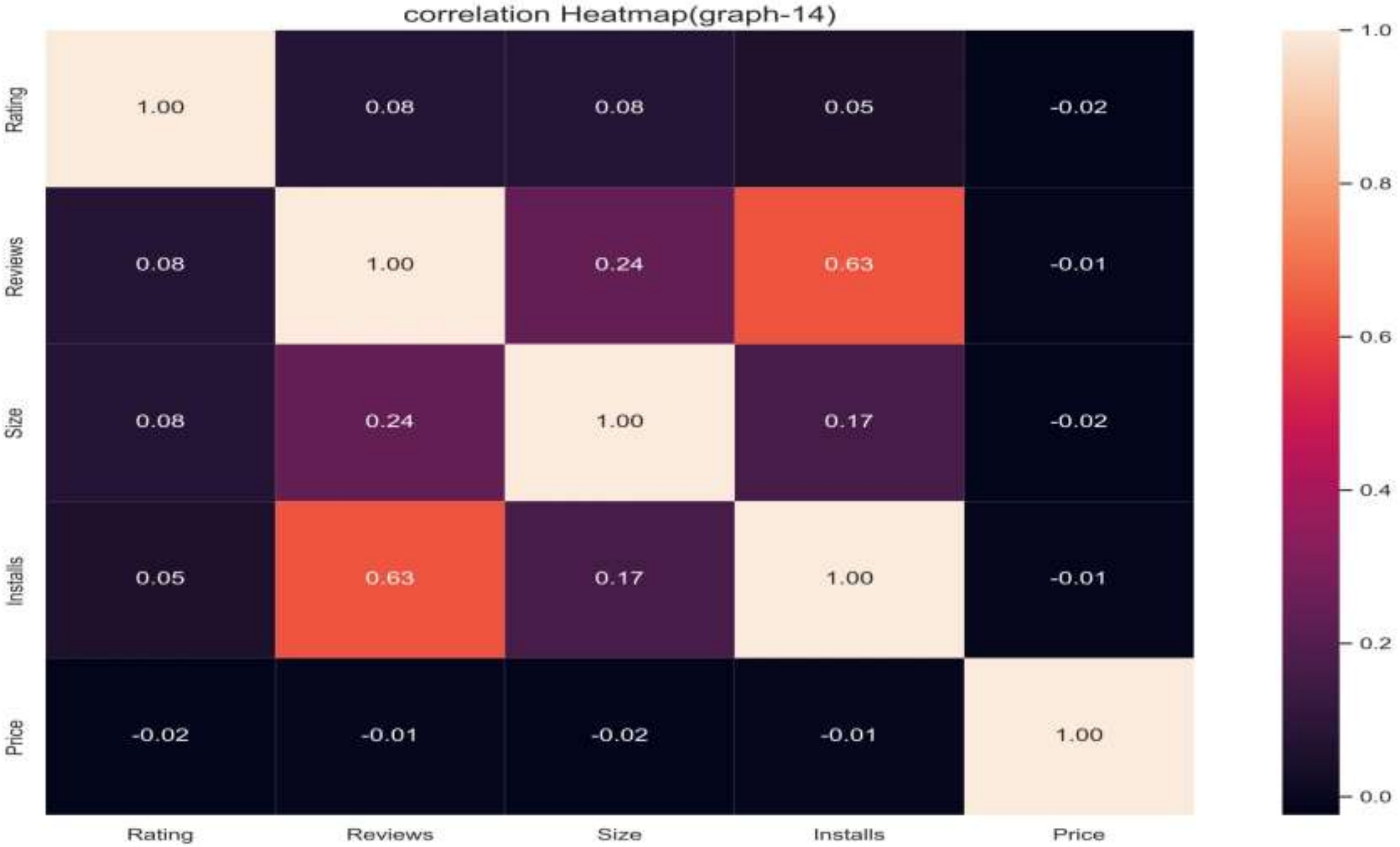
After merging User reviews than graph we can see that in the "COMICS" category has the highest average rating. Highest COMICS Category rating is 4.7000, and lowest ENTERTAIMENT category rating is 4.05642. User reviews After merging all 33Categories Average rating 4.274.



Data Analysis & Visualization

14. Correlation Heat map ?

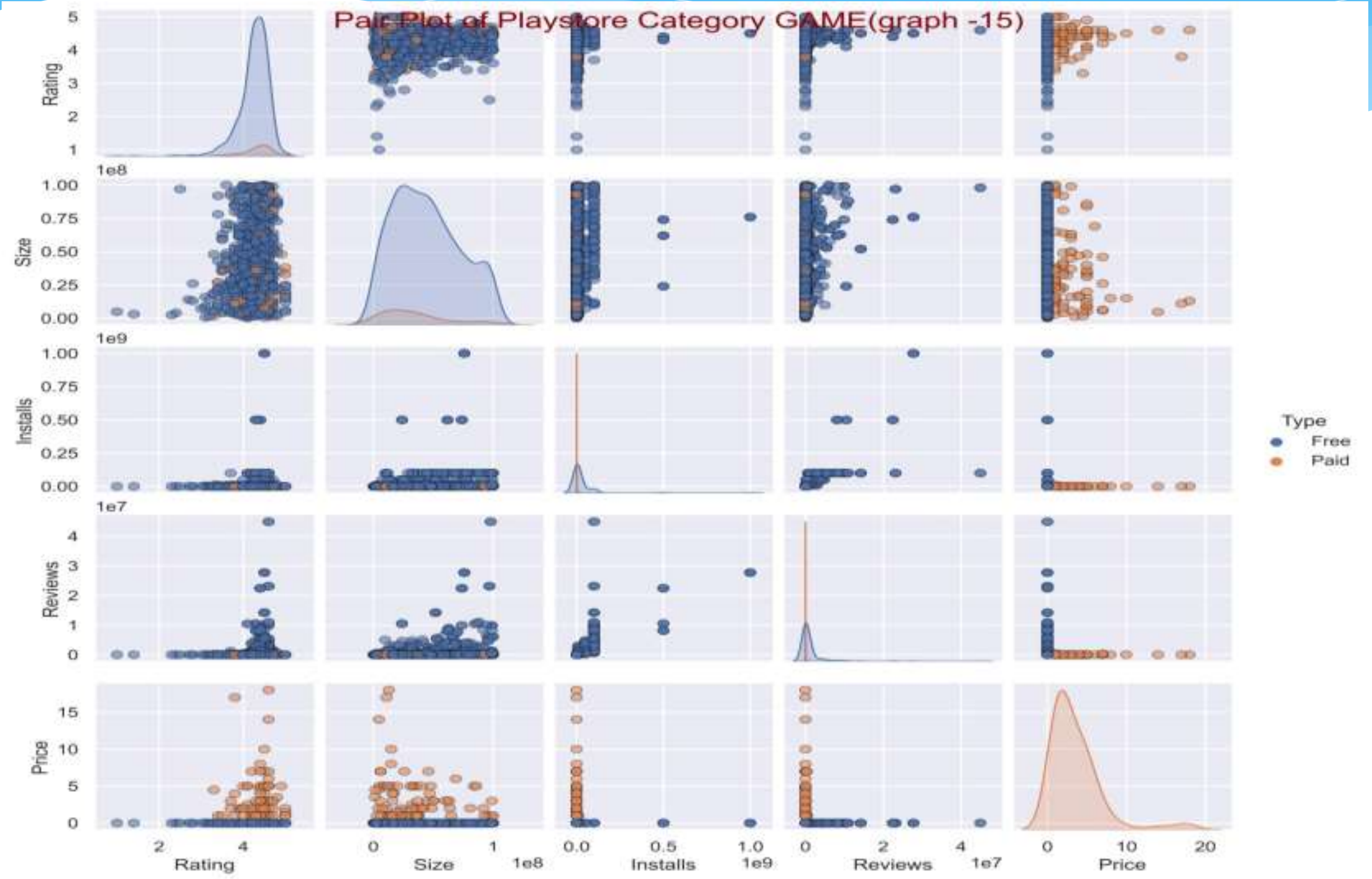
There is a strong positive correlation between the Reviews and Installs. The Price is slightly negatively correlated with the Rating, Reviews, and Installs. The Rating is slightly positively correlated with the Installs and Reviews.



Data Analysis & Visualization

15 . Pair Plot

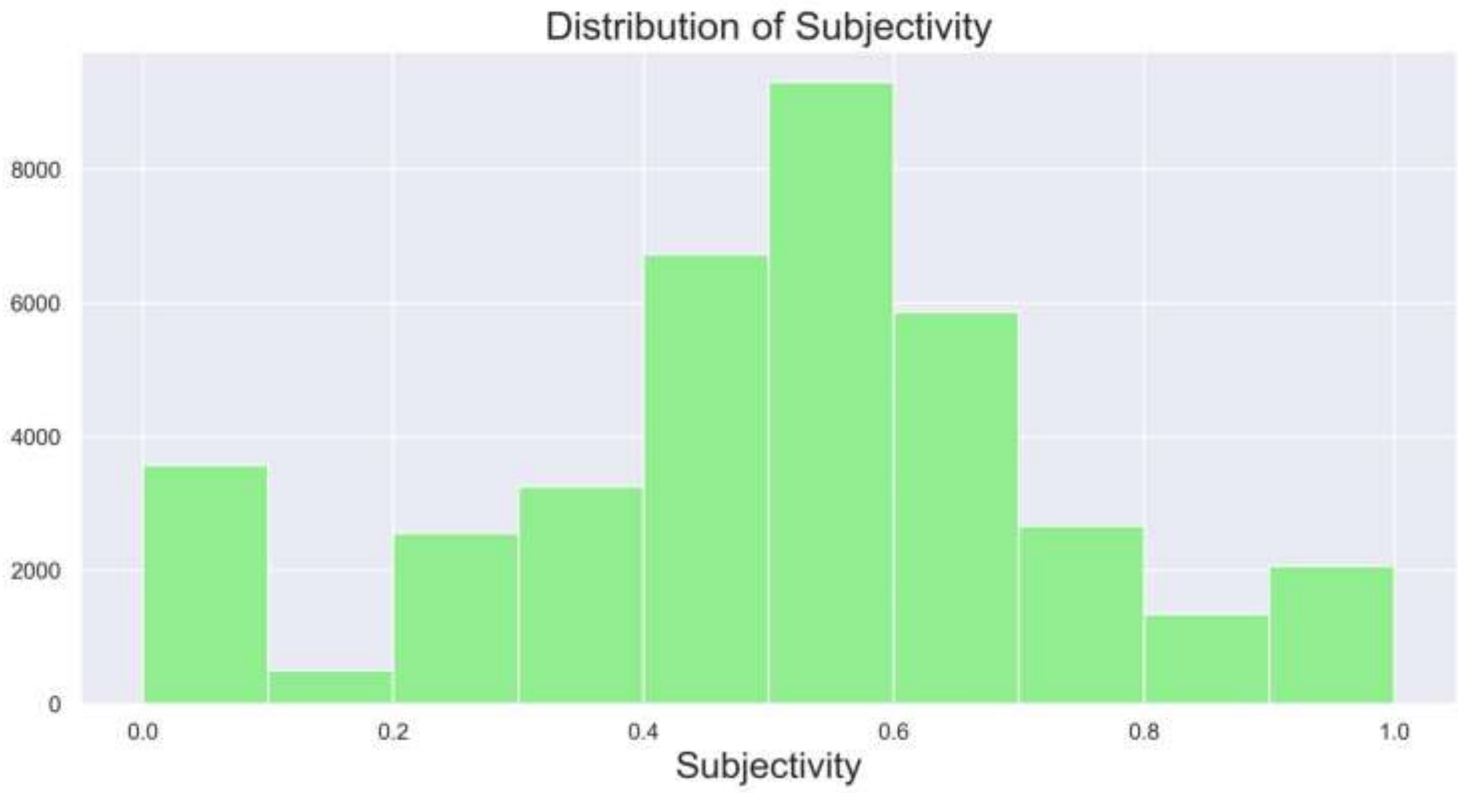
Game Category Rating, Reviews, Size, Installs and Price Check relationship To Type Free and Paid.



Data Analysis & Visualization

16. Distribution of Sentiment subjectivity

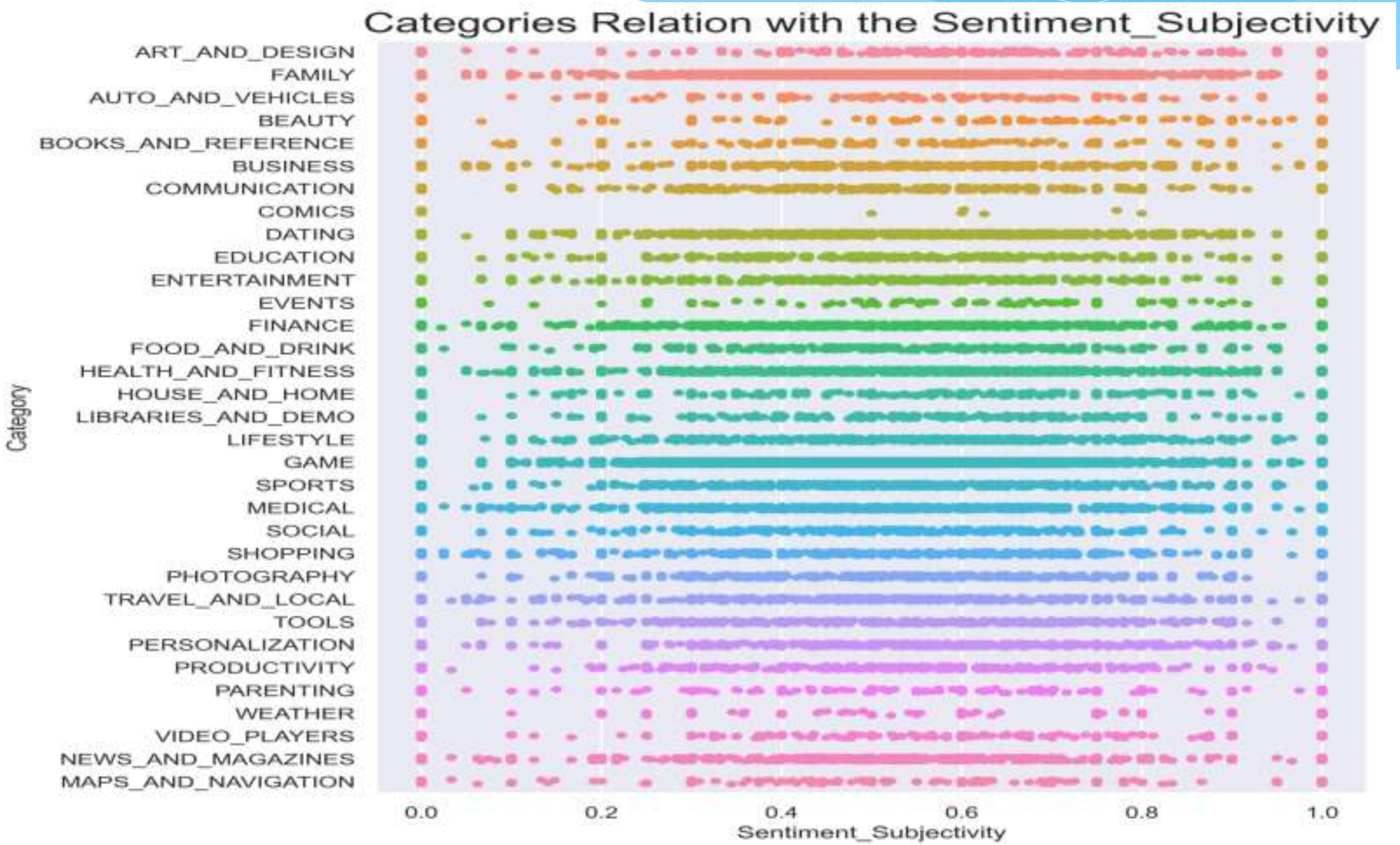
- this graph it can be seen that maximum number of sentiment subjectivity lies between 0.4 to 0.7.
- So we can conclude that maximum number of users give reviews to the applications, according to their experience



Data Analysis & Visualization

17 . Categories Relation with the Sentiment Subjectivity

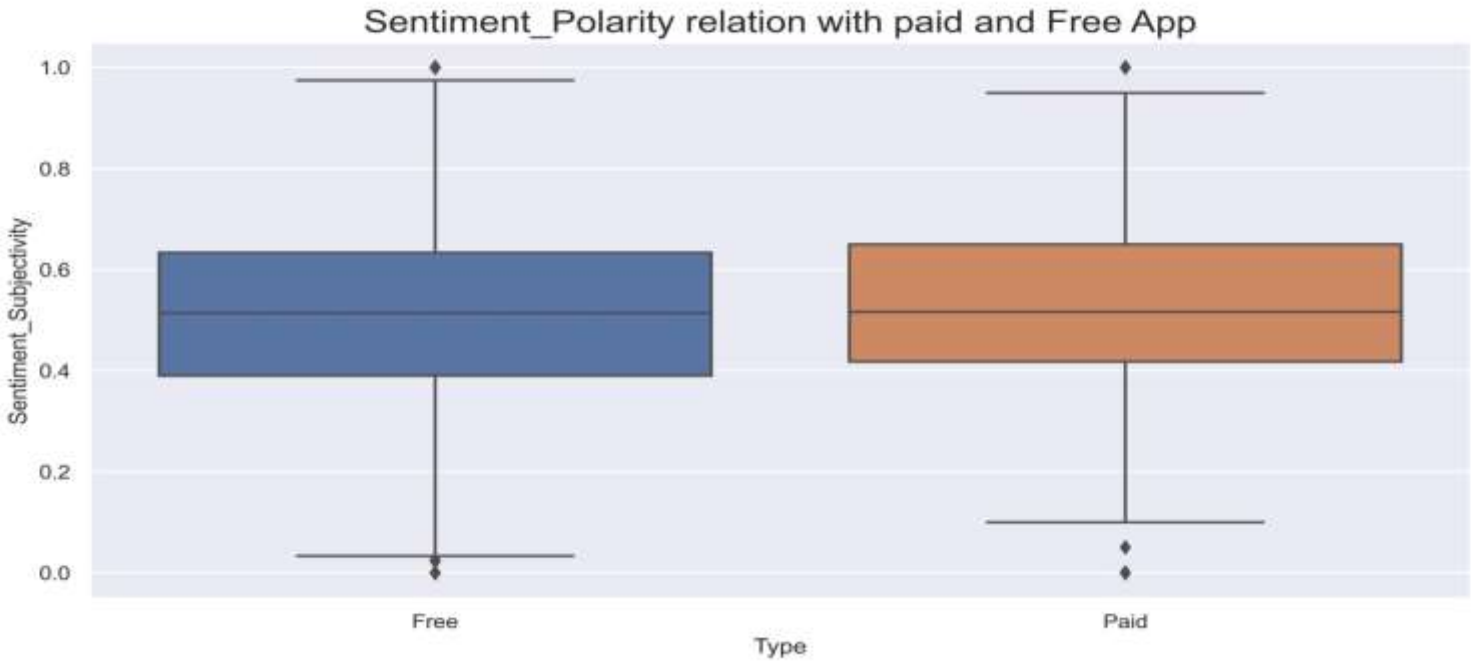
User sentiment subjectivity ART-AND-DESIGN ,FAMILY , Play store App category Relation will be check here.



Data Analysis & Visualization

18. Sentiment _ Polarity relation with paid and Free App

- In this graph we can come to the conclusion that , Paid type Apps more than Free type Apps Sentiment Subjectivity



Data Analysis & Visualization

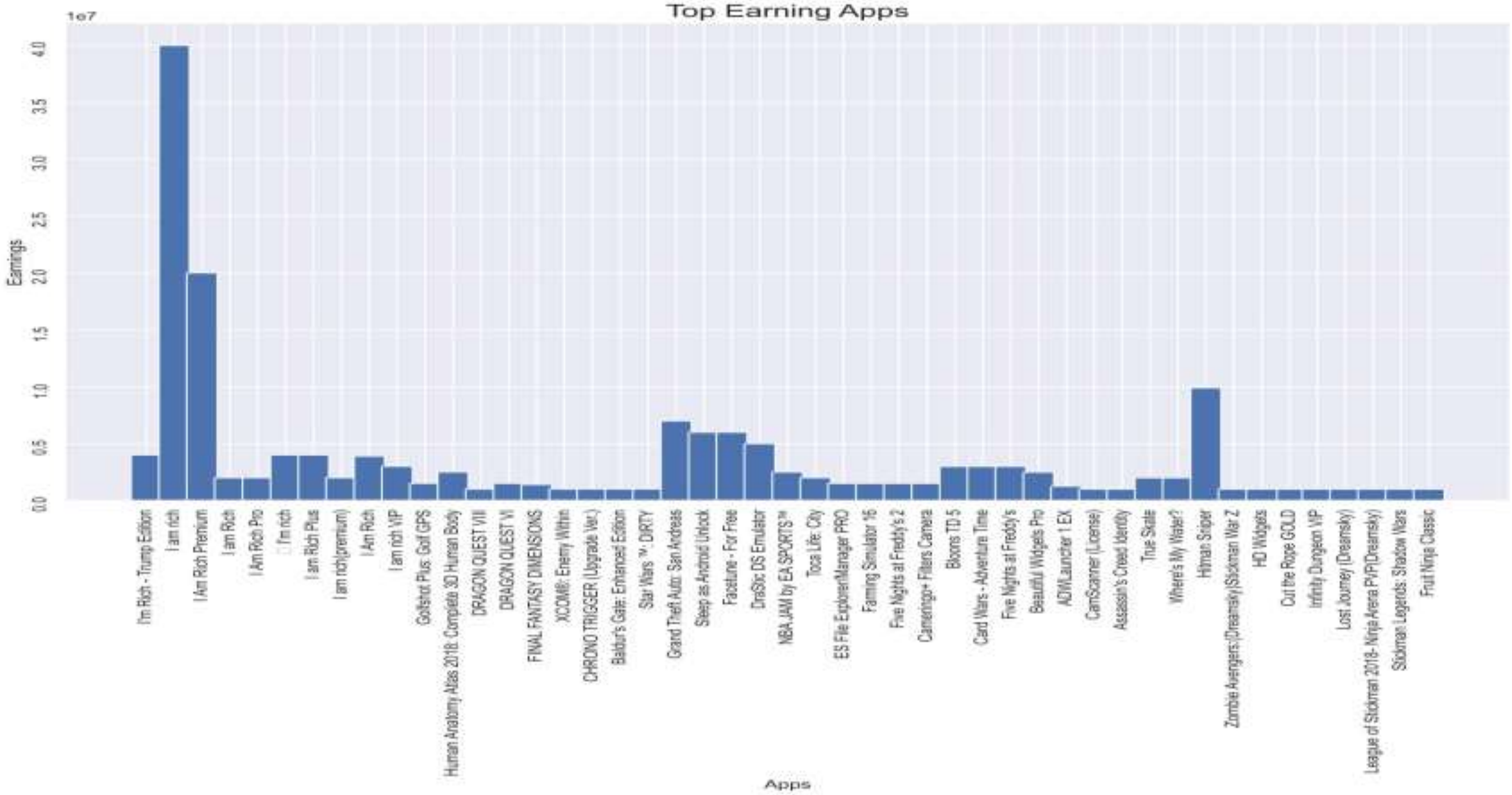
19 : Which are the apps that have made the highest earning

Find Earning generated is given by the formula:

- **Earning = Installs * Price**

The top four apps with highest earnings found on GOOGLE Play store are:-

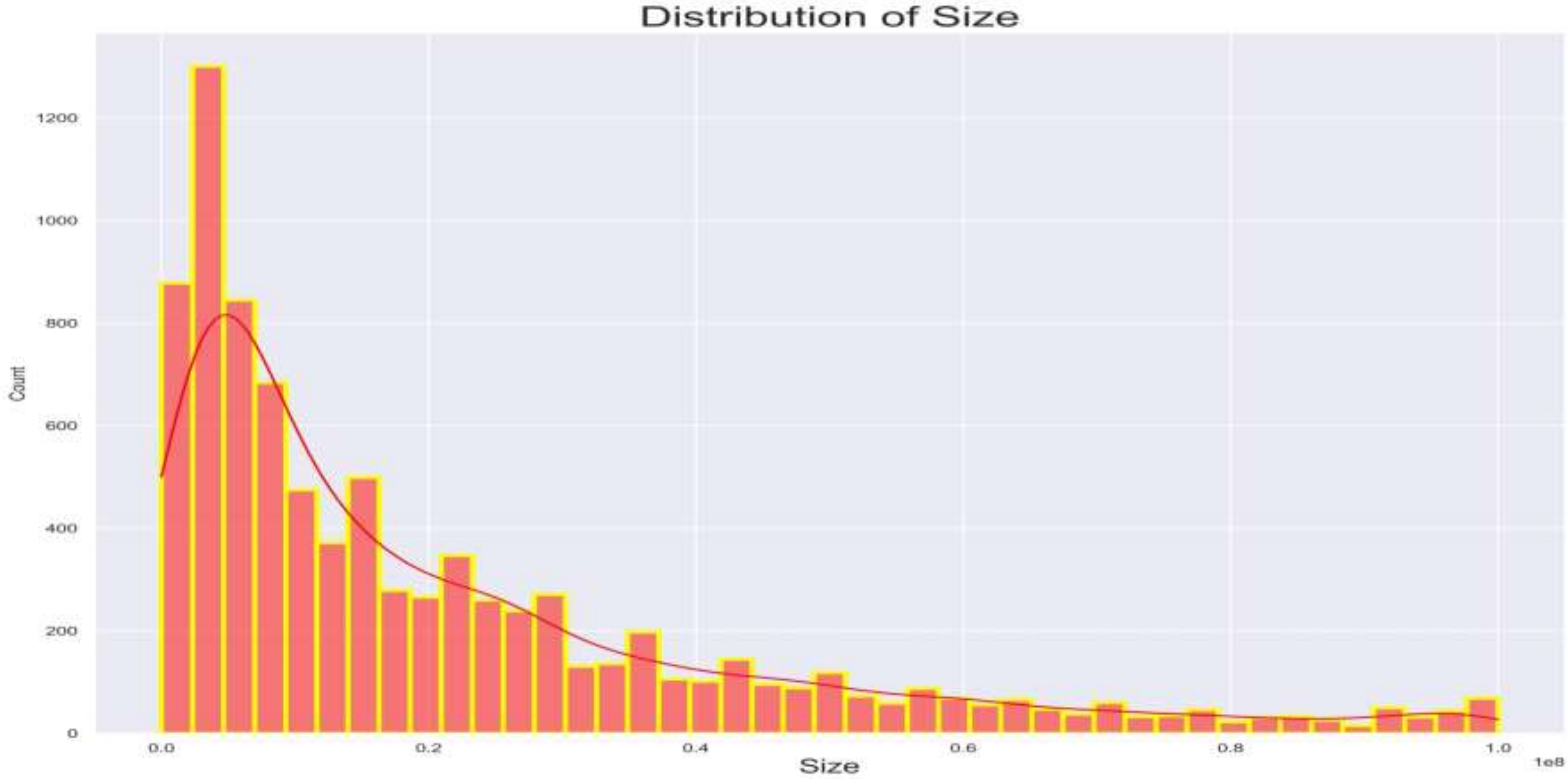
- I am Rich
- I am Rich Premium
- Hit man Sniper
- Grand Theft Auto: San Andreas



Data Analysis & Visualization

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

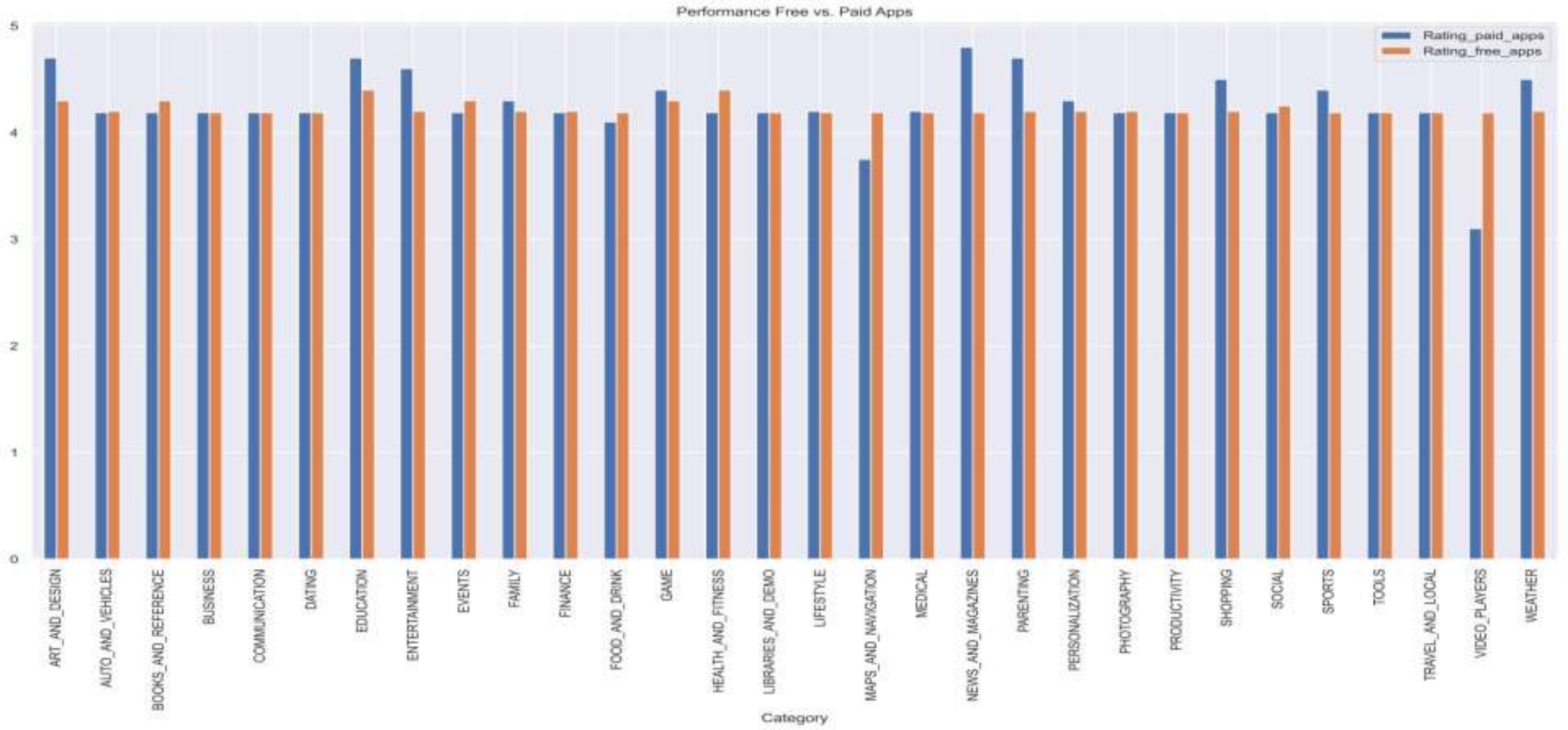
- we can come to the conclusion that maximum number of applications present in the dataset are of small size



Data Analysis & Visualization

21. Are Paid apps worth buying? (Analysis based on Average User Rating)

Looks like paid apps perform marginally better than the free apps, check category under high rating paid apps and free apps and check category under low rated paid apps and free apps high rated paid Category apps NEWS_AND_MAGAZINES Rating 4.800000 and high rating free apps EDUCATION Ratings 4.400000 and low rated paid Category apps VIDEO_PLAYERS Rating 3.100000 and low rated free category apps HOUSE_AND_HOME Rating 4.187877.



CONCLUSION

1) In play store present most of the apps are under Family & Game category and least are of Beauty & Comics Category.

2)Category with the highest number of installs is Game and least number of installs is EVENTS categories.

3) 92.2% apps are Free and 7.8% apps are paid in type.

4) Most of the apps in the Google play store are rated between 3.8 to 4.8.

5) play store maximum number of apps are available for Everyone and then for Teen Everyone content Rating 7195 and Teen Content Rating 915.

6) FAMILY category POU, and CANDY CRUSH SAGA has the highest installs 5 Billions
times install in two apps.

7) In Play store most expensive app is I'm Rich - Trump Edition price actually 400.00 this install only 10000 times but I Am Rich Premium is price 399.99 but this maximum time install 50000 times.

8) In This Play store Game category belongs Clash of Clans game is most reviews
estimate 44893888 reviews

9) Top20 Apps in different genres highest genres count Tools 739

10) The Genres that are getting installed the most in top 20 Genres Arcade Genres 8836079153 times installs this the highest install in Play store.

11) In Genres Highest rating genres is Comics Creativity is Rating 4.800000
and Lowest rating genres is Parenting Brain Games Rating 3.800000

12) App updated in play store during 2017-2018 in 2017 update 1702 apps and in 2018 5615 apps will be update.

13) The "EVENTS" category has the highest average rating. "EVENTS" category rating is 4.381924, before merging play store category rating all 33Categories average rating is 4.183

After merging User reviews than the "COMICS" category has the highest average rating. Highest COMICS Category rating is 4.7000, and lowest ENTERTAIMENT category rating is 4.05642. user reviews After merging all 33Categories Average rating 4.274.

14) Correlation heat map The Price is slightly negatively correlated with the Rating, Reviews, and Installs. The Rating is slightly positively correlated with the Installs and Reviews.

15) Pair plot Game Category Rating, Reviews, Size, Installs and Price Check relationship To Type Free and Paid.

16) 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.

17) In this chart show that Sentiment Subjectivity in all category GAME and FAMILY is underlying distribution is maximum

18) Free type app compare to Paid type app high Sentiment Subjectivity.

19)n this Google Play store I am Rich

I am Rich Premium

Hit man Sniper

Grand Theft Auto: San Andreas are top four highest expensive app

20)The conclusion that maximum number of applications present in the dataset are of small size.

21)Looks like paid apps perform marginally better than the free apps, check category under high rating paid apps and free apps and check category under low rated paid apps and free apps high rated paid Category apps NEWS_AND_MAGAZINES Rating 4.800000 and high rating free apps EDUCATION Ratings 4.400000 and low rated paid Category apps VIDEO_PLAYERS Rating 3.100000 and low rated free category apps HOUSE_AND_HOME Rating 4.187877.

THANK YOU!