

Capstone Project-1

Play Store App Review Analysis

Team Member

Pratik D. Jori

Pritesh Tambat

Someshwar Hydrabade

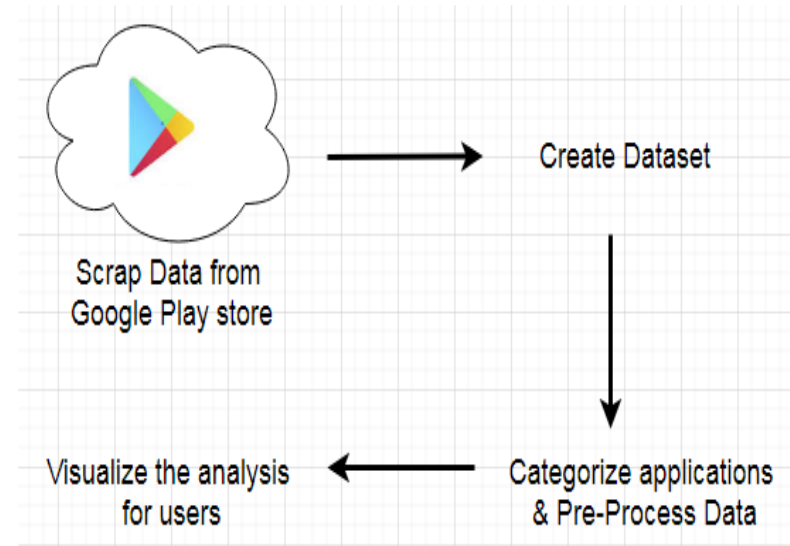
Sarvesha Kapse

Data Science Trainee, Almabetter

Introduction

Google play store is widely used platform for the download and install various mobile applications. It get flooded with thousands of new app everyday with many more thousand of developers are working independently or in team to make them succesfull.

In the exploratory data analysis we take data from play store which is in the form of dataset, then we clean and process the give data for the further visualization and analysis the user reviewes.



Problem Statement

- Android is expanding as an OS and Mobile app industry is increasing significantly and thus giving rise to more competitions to the one's that are creating applications.
- Due to the competition in the market and also expansion 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 we analyze and research relevant data.
- Two datasets are provided one with the information about the application and other with the user reviews for the applications
- We have examine and evaluate the both datasets to identify the important factor for the success of the apps.

Challenges Faced

- Reading the Datasets and comprehension problem statements.
- Handling the error, duplicates, NaN values in the datasets.
- Visualization to summarize the information of datasets and successfully communicate the trends and results to the reader.

Points for Discussion

- ☐ Data summary
- ☐ How many apps in each category
- ☐ How many ratings are getting for the apps give it's relation
- ☐ Which content rating is having higher number of apps
- ☐ Which are the top categories in Play Store
- ☐ What is percentage of the paid and free apps
- ☐ Which are the top 10 expensive apps
- ☐ What is relation between rating and it's affecting factors
- ☐ What are the Top 10 installed apps in any category
- ☐ Which App has higher number of reviews
- ☐ Which apps giving the highest earning
- ☐ What is average rating of the App in play store
- ☐ Conclusion

Data Summary

Google Play store Dataset

- **App** :- This Column contains the name of the app
- **Category** :-This contains the category to which the app belongs to the different categories.
- **Rating** :- This column contains the average value of the individual rating the app.
- **Reviews** :-This column contains the number of users have given their feedback for the app.
- **Size** :- This column contains the size of the app.

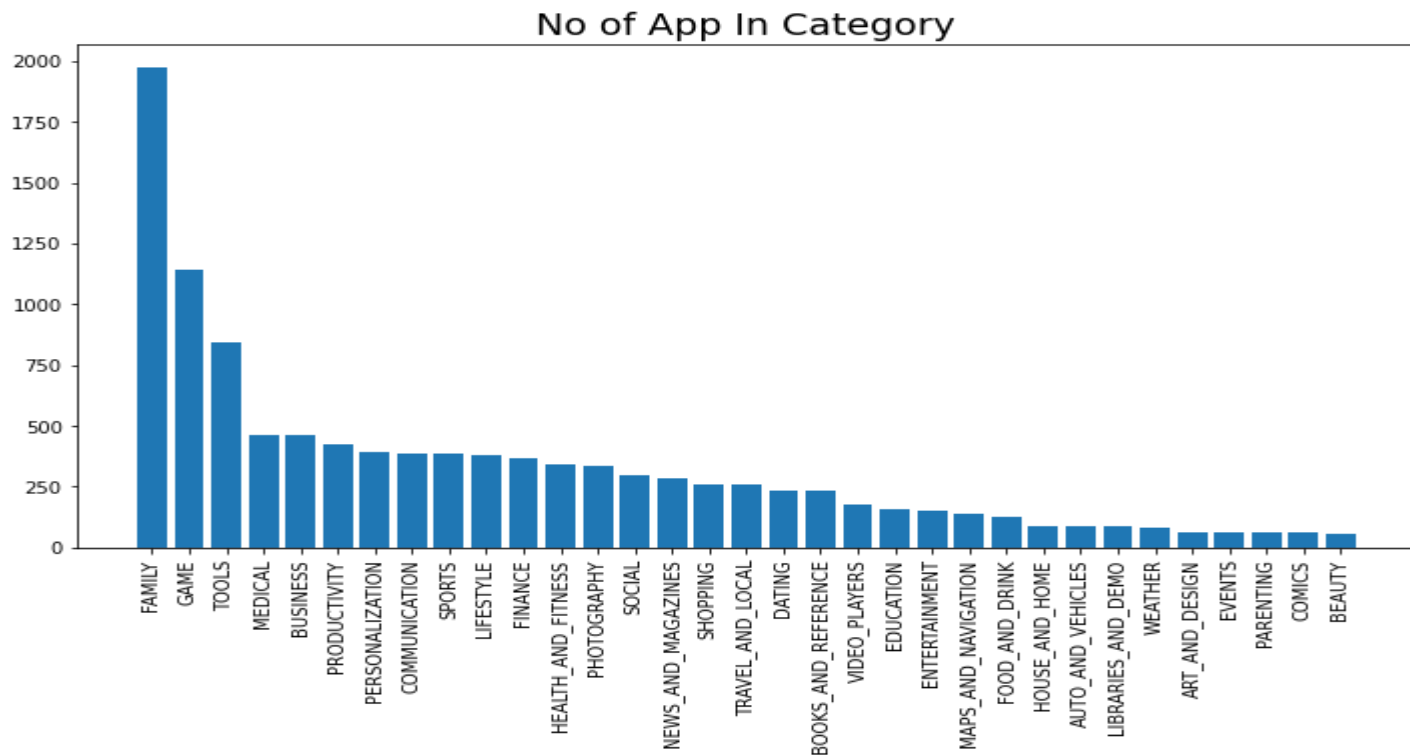
- **Installs** :- This column indicates the number of time that the app has been downloaded from the play store.
- **Type** :- This column contains only two values- free and paid.
- **Price** :- For paid apps this column contains the price of the app, for free apps it contains the value 0.
- **Content Rating** :- It indicates the users of the app and their age group.
- **Genre** :- This column contains to which genre the app belongs to.
- **Last Updated** :- This column contains the info about the date on which the last update for the app was launched.

- **Current Version :-** Contains information about the current version of the app available on the play store.
- **Android Version :-** Contains information about the version of the android OS on which the app can be installed.

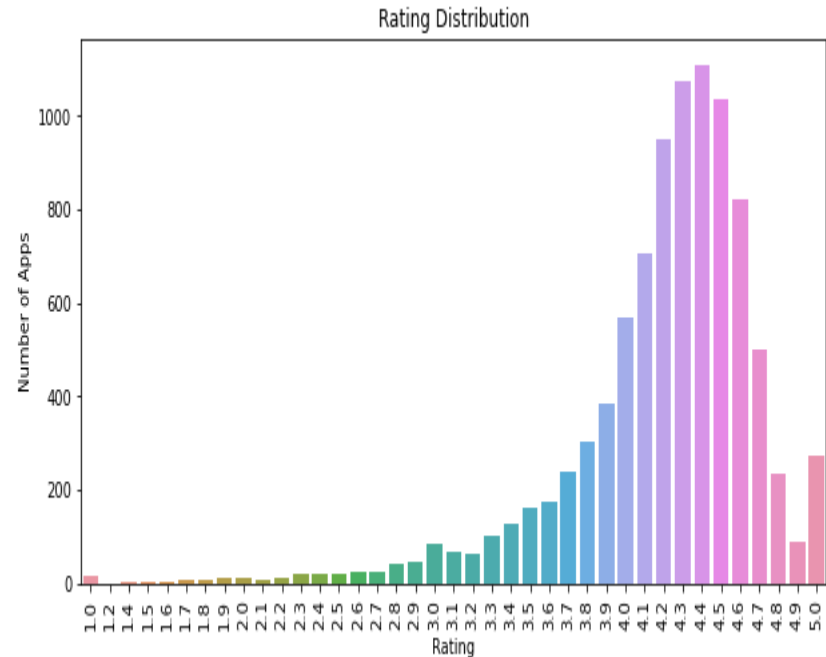
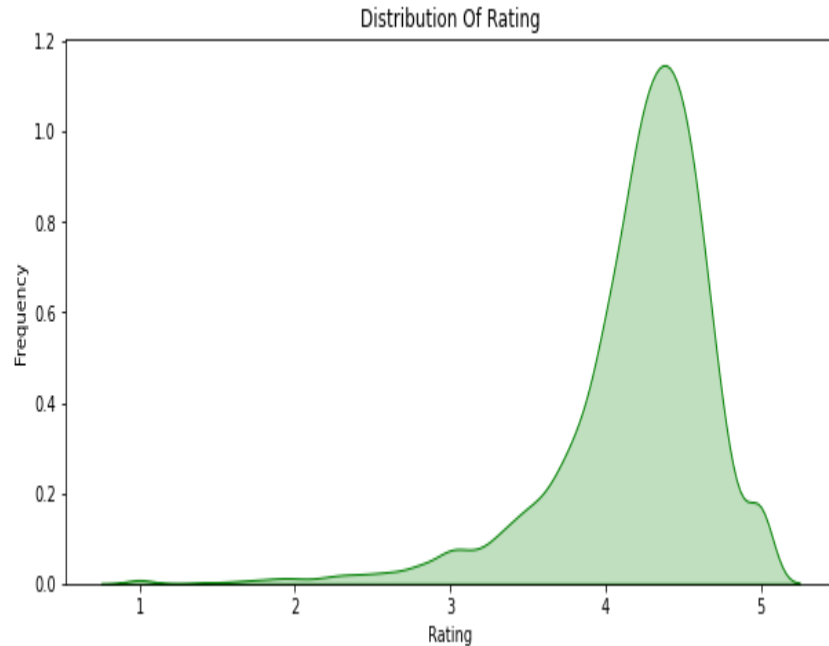
User Review Dataset

- **App :-** Contains the name of the app with a short description.
- **Translated Review :-** It contains the English translation of the review given by the user of the app.
- **Sentiment :-** It gives the emotion of the review writer. It can be Positive, Negative, or Neutral.
- **Sentiment Polarity :-** It represents the polarity of the review. Its ranges $[-1,1]$, where 1 represents 'Positive statement' and -1 represents a 'Negative statement'.
- **Sentiment Subjectivity :-** This value gives how close a reviewer's opinion is to the opinion of the users.

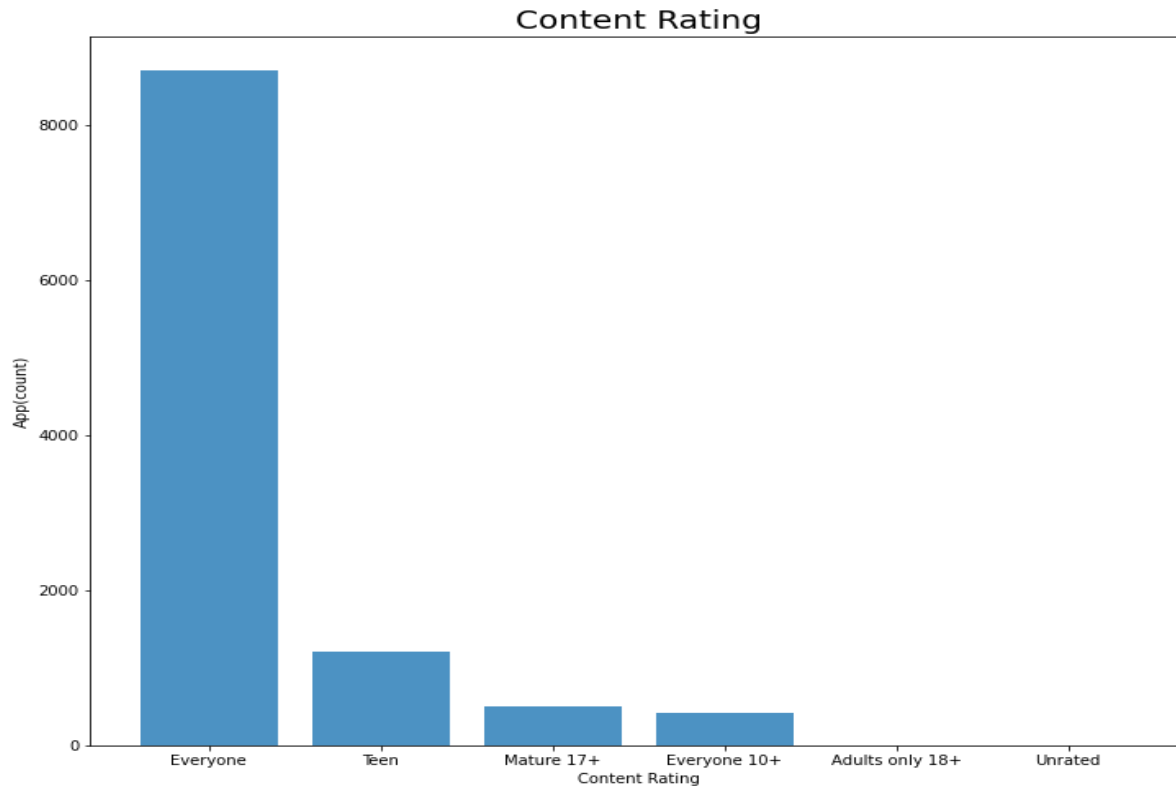
How many apps in each category?



How many ratings are getting for the apps give it's relation?

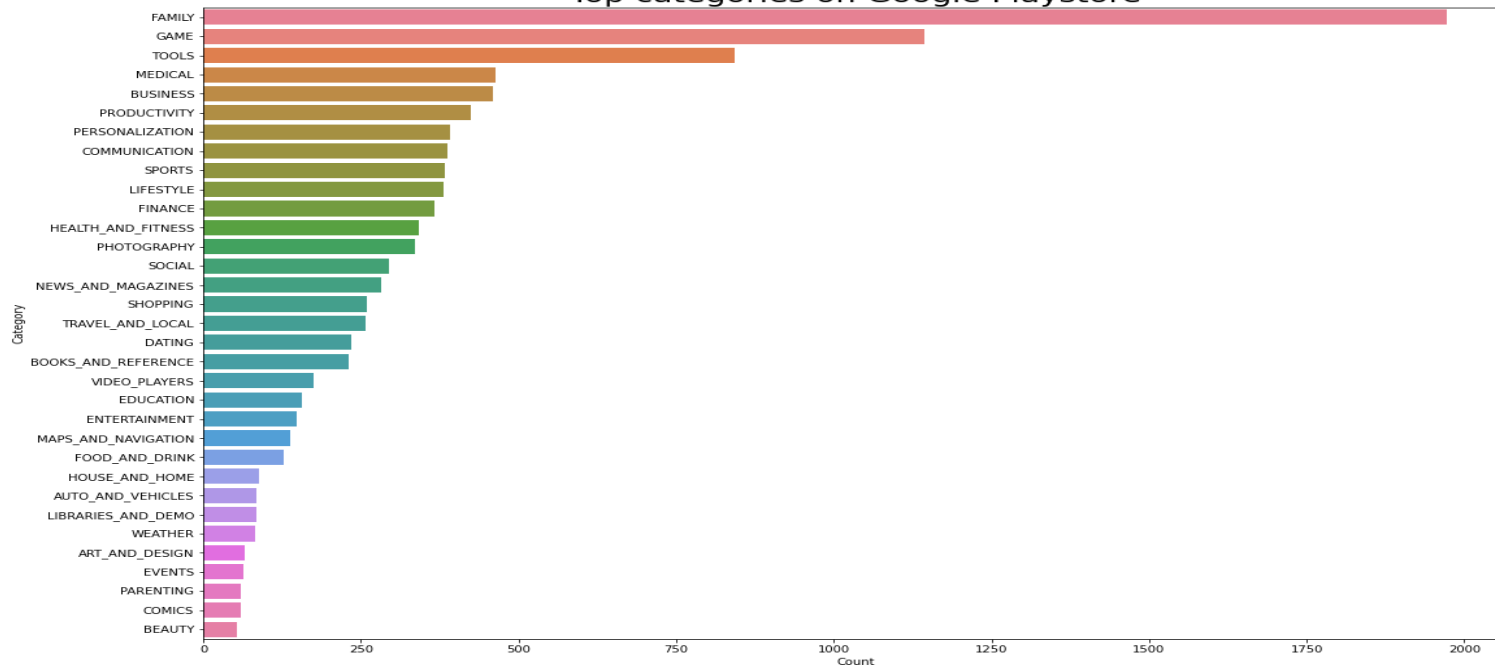


Which content rating is having higher number of apps?

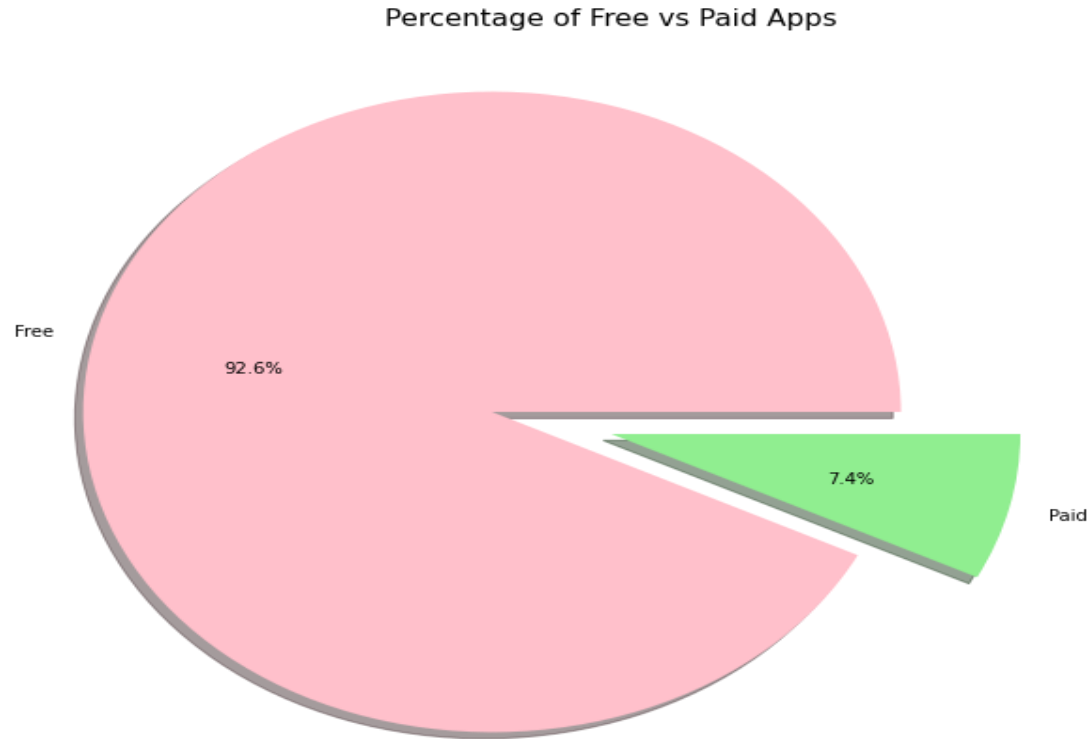


Which are the top categories in Play Store?

Top categories on Google Playstore

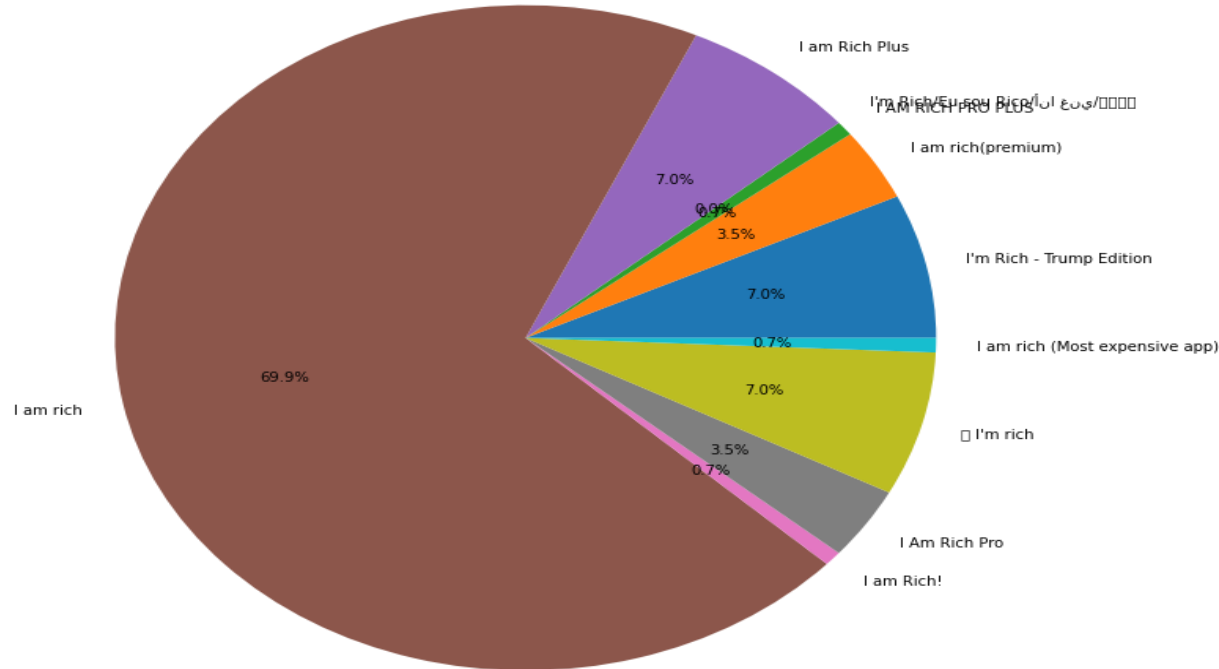


What is percentage of the paid and free apps?

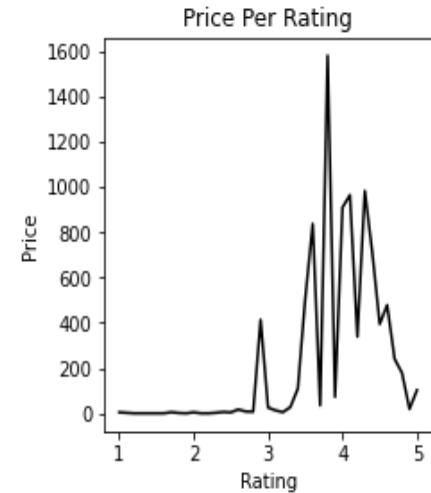
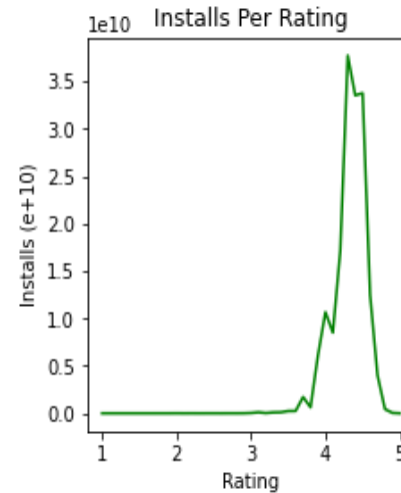
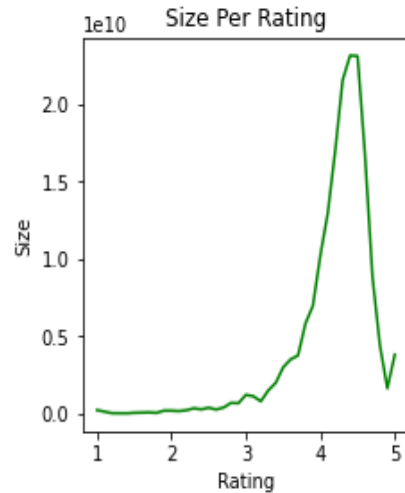
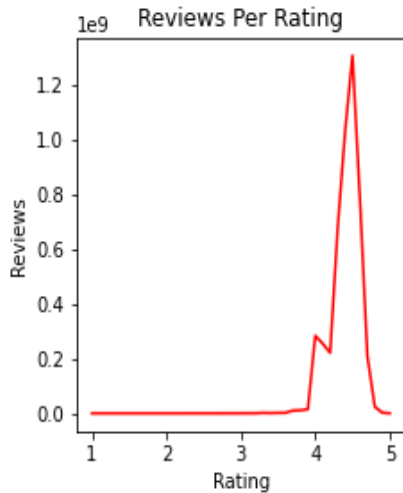


Which are the top 10 expensive apps?

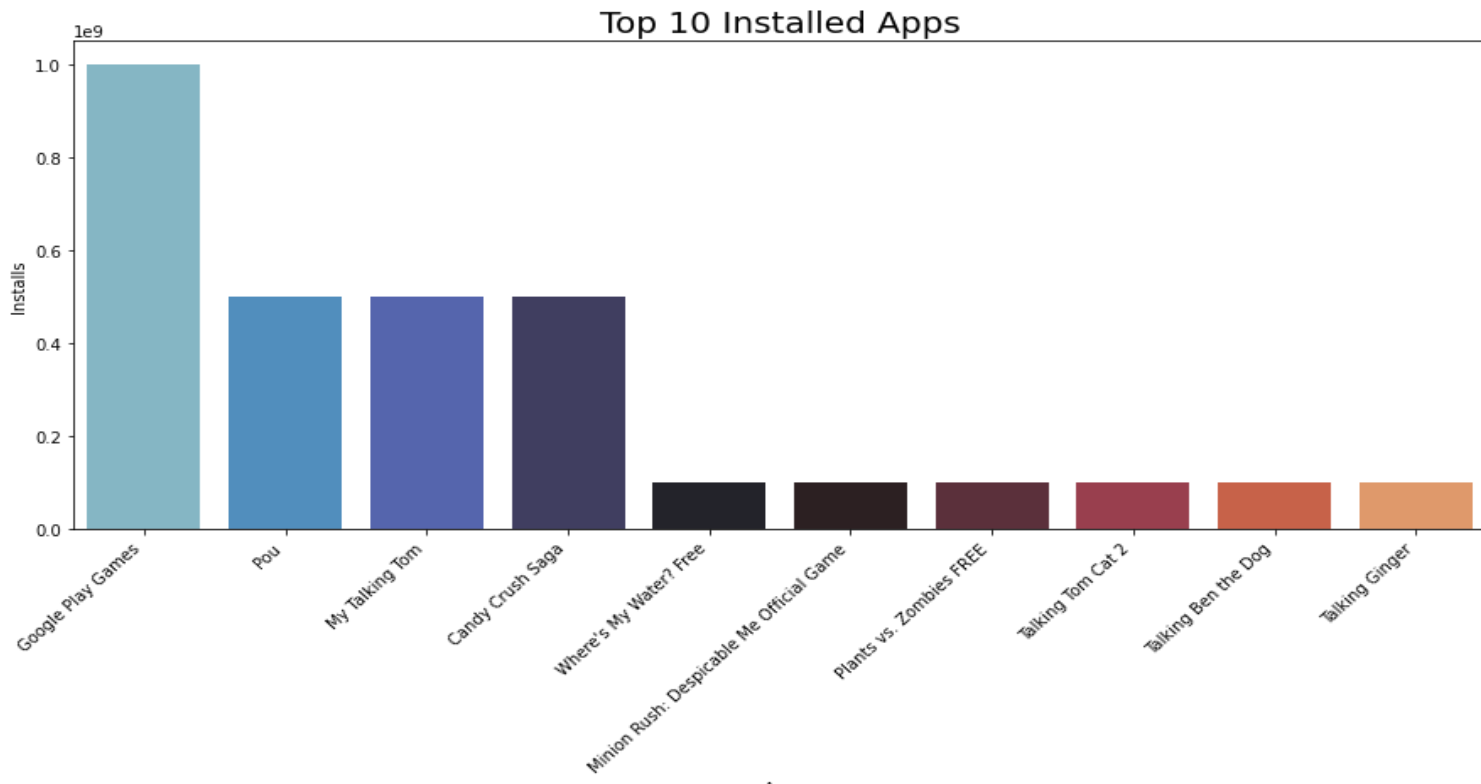
Top Expensive Apps Distribution



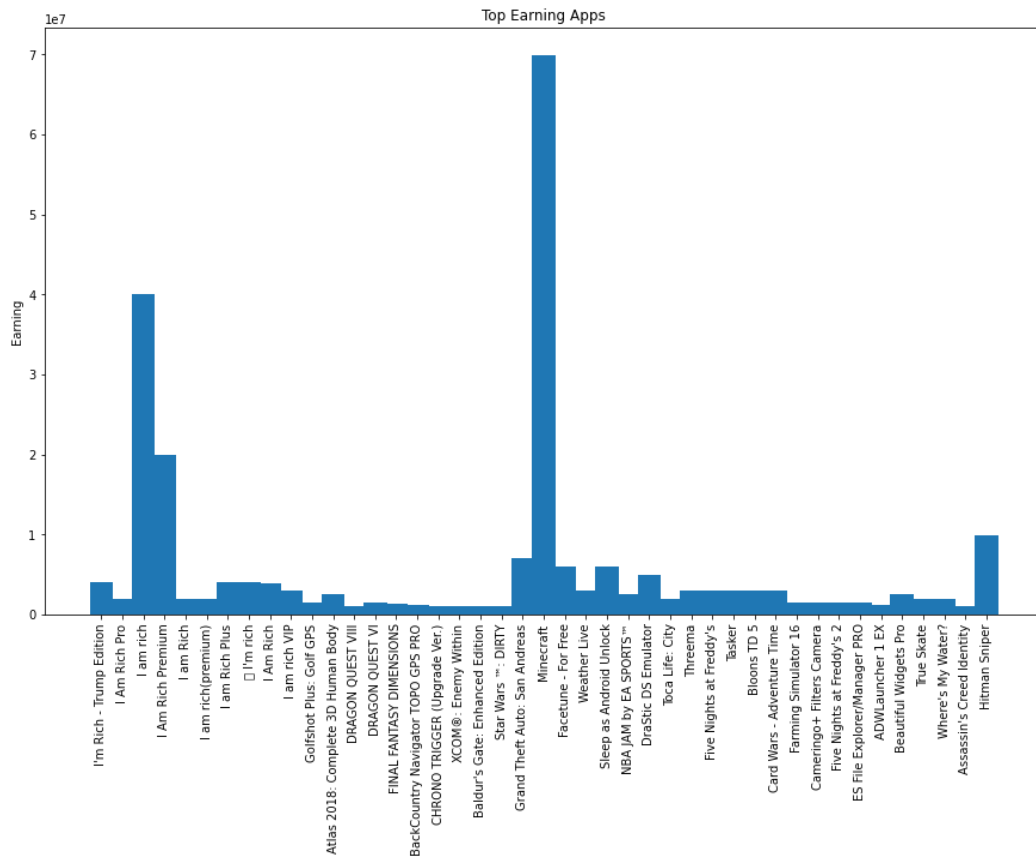
What is relation between rating and it's affecting factors?



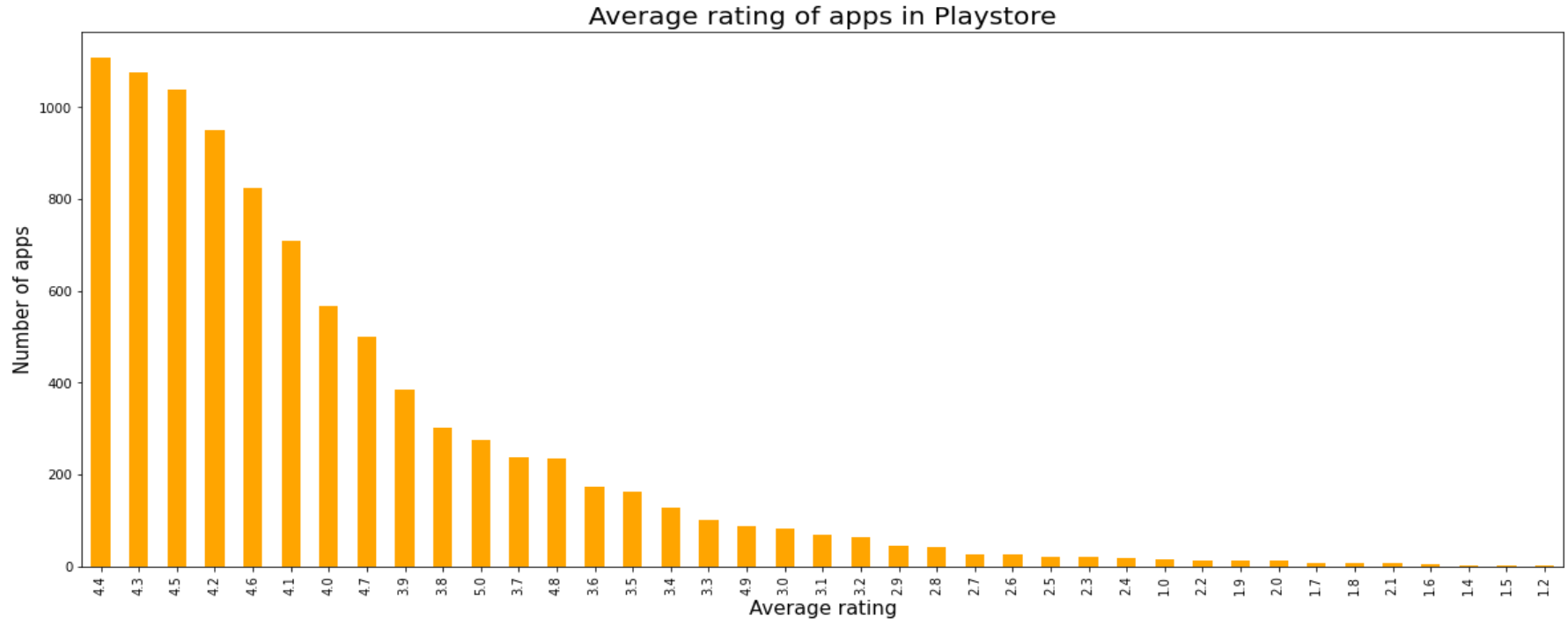
What are the Top 10 installed apps in any category?



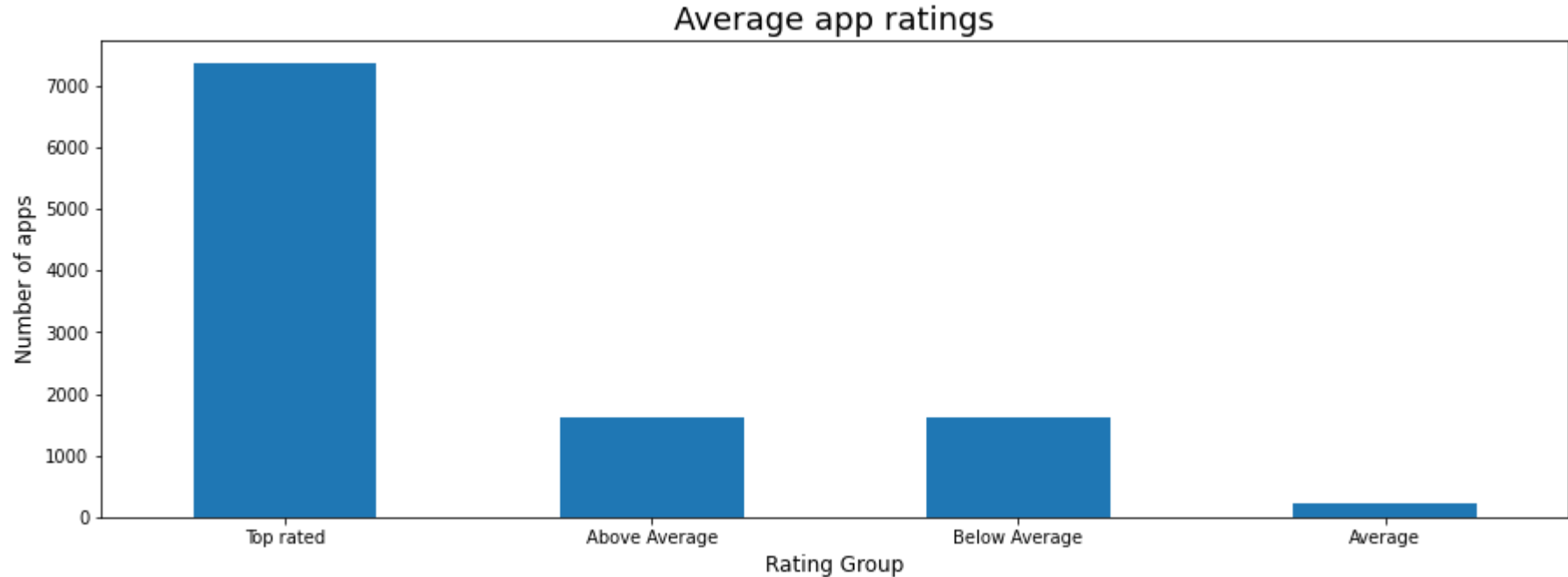
Which apps giving the highest earning?



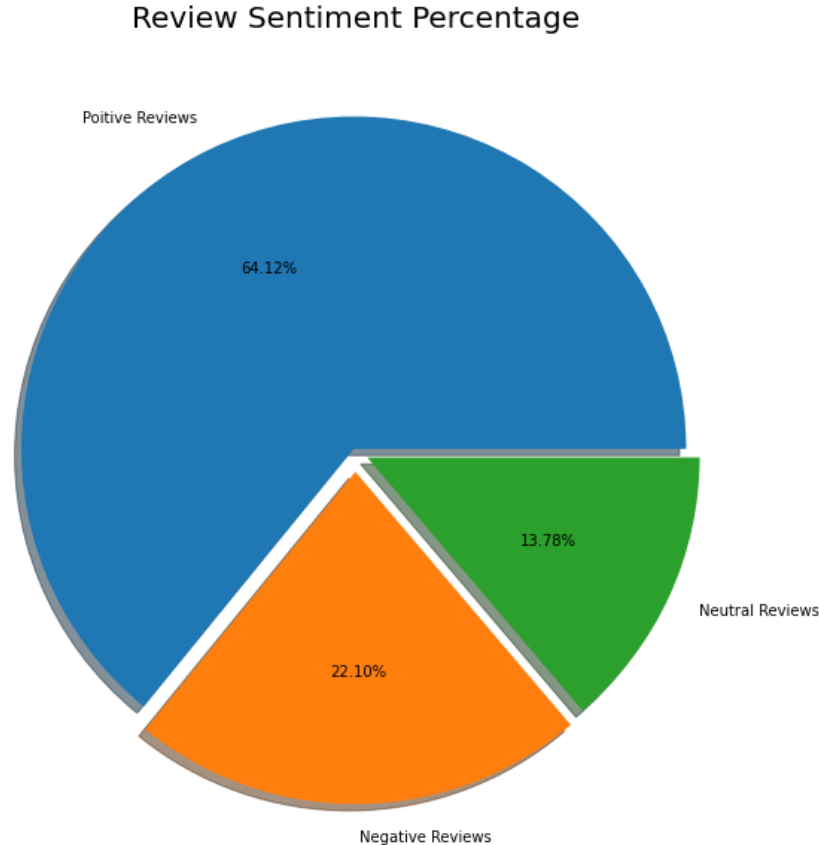
What is average rating of the App in Playstore?



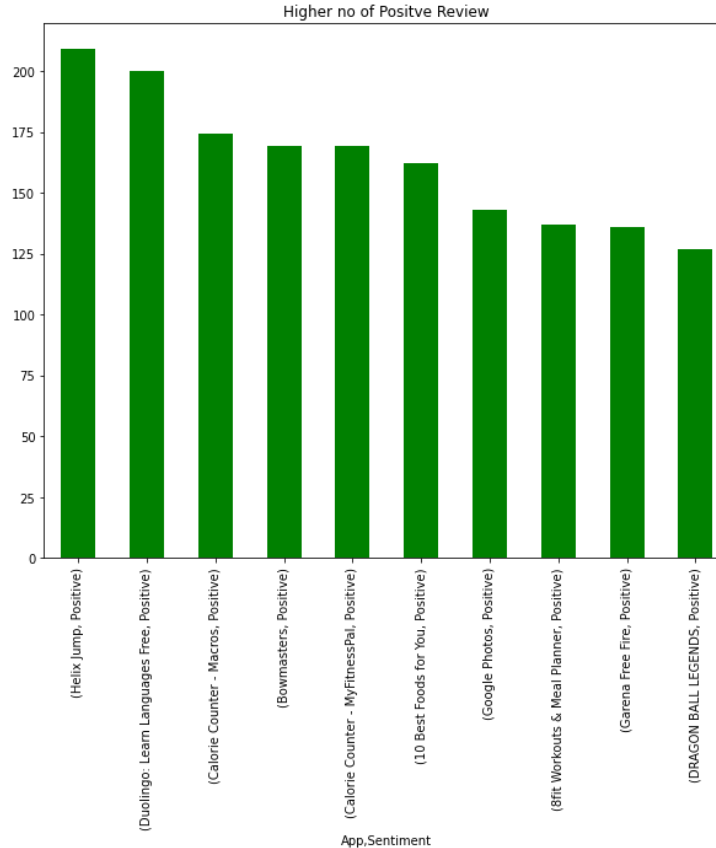
From avg rating of app can we group as Top rated, Above average, Average, Below average?



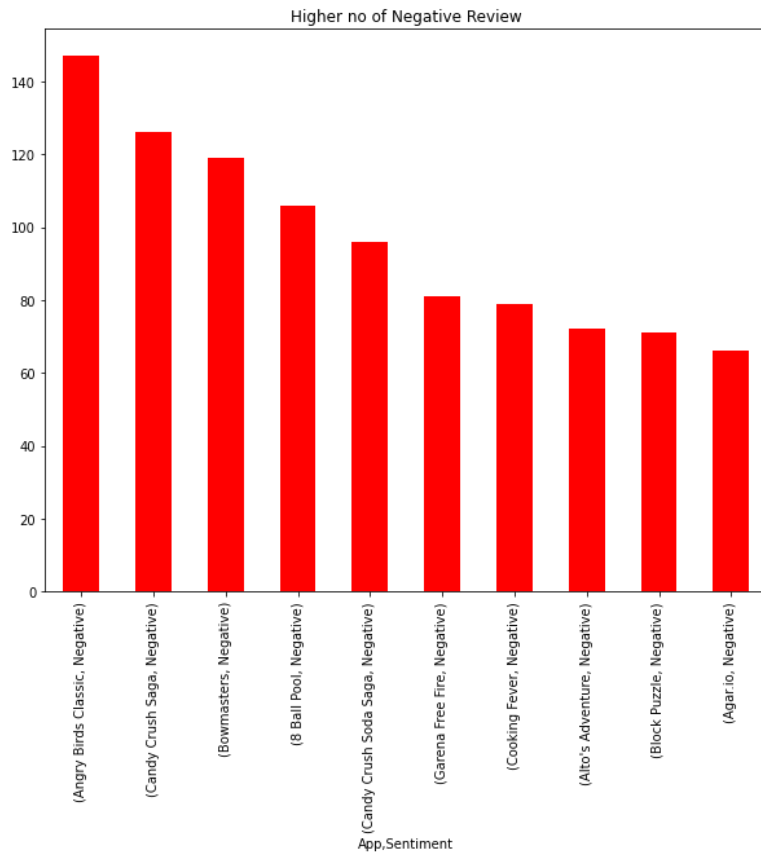
What is percentage of the user review sentiments?



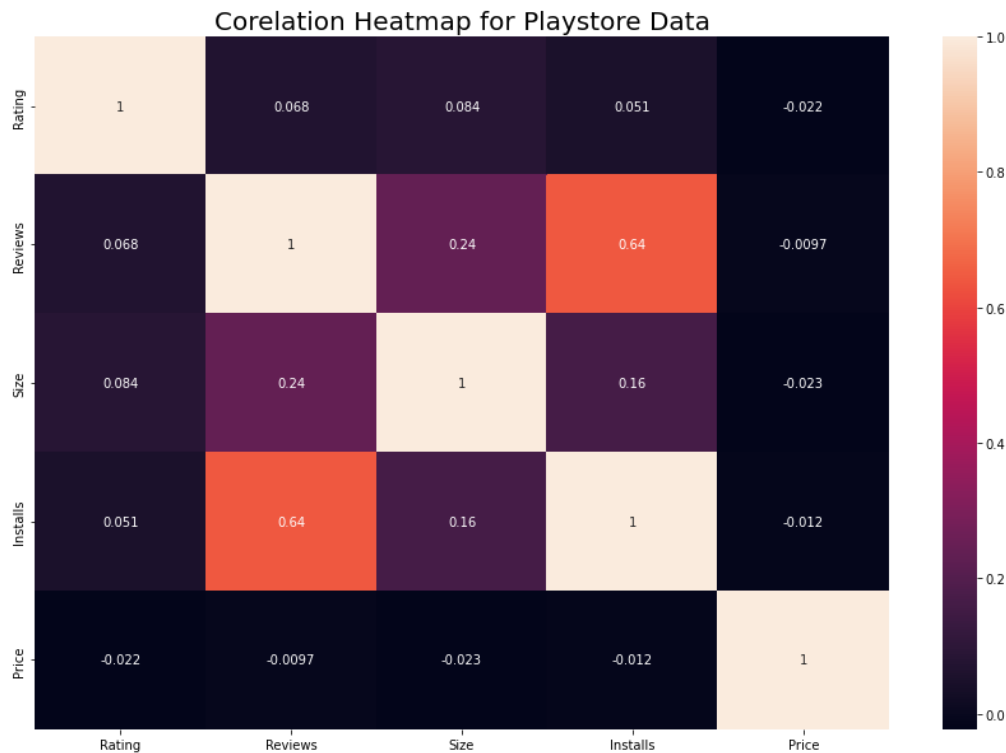
Which app has higher number of positive review?



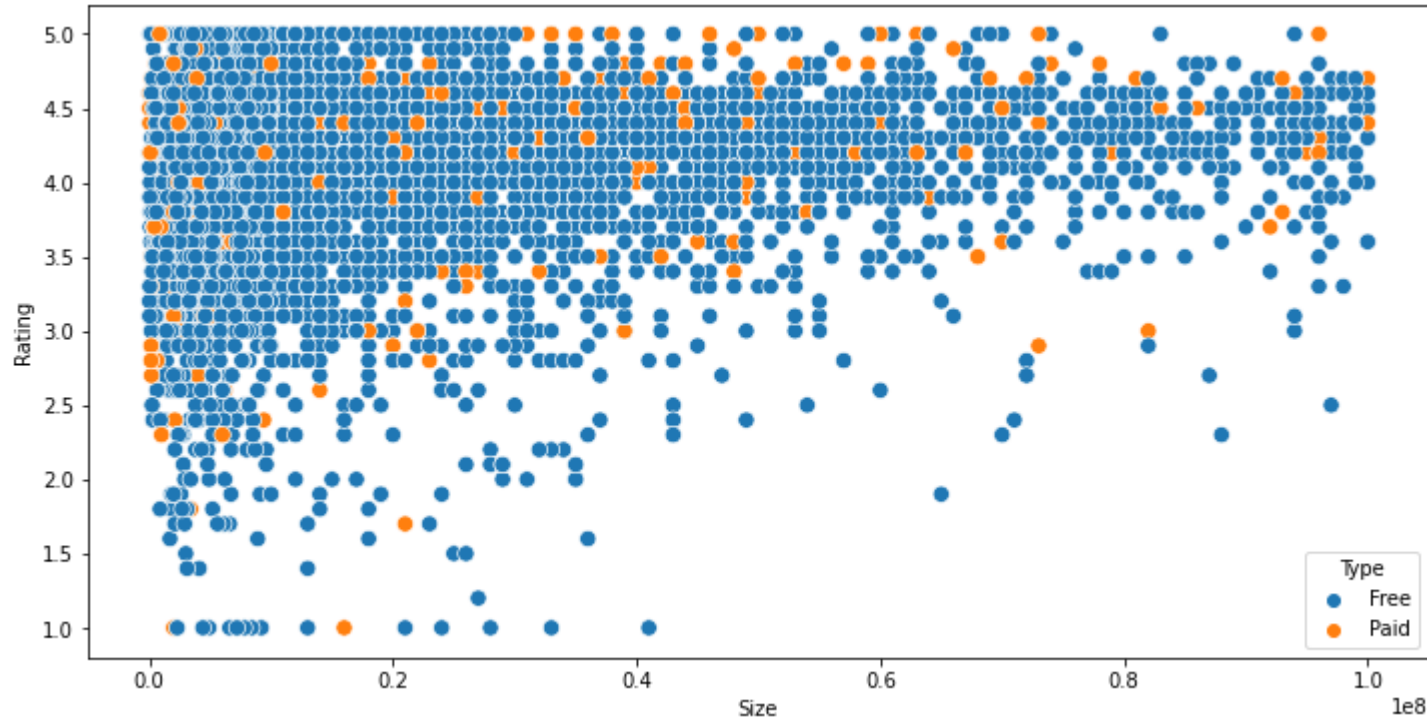
Which app has higher number of the negative review?



Correlation Heatmap for Play Store Data



Distribution of App on there Type, Size and Rating



Conclusion

- We have seen from the Exploratory data analysis that the 'family', 'game' 'tools', 'medical' this are the top category apps hence it is mostly used app.
- And the most number of that install is free approximately 92.6%, and paid is just 7.4%
- The review sentiments are showing the what reviews are getting for the app whether the app is good for the user perspective it show the sentiment of the users. Like in this data we have Positive Reviews = 64.12%, Negative Reviews = 22.10%, Neutral Reviews = 13.78%.

Thank You