

Capstone Project

Play Store App Review Analysis

By Ayush Shrivastava

Content

1. Problem statement
2. looking at the dataset
3. Dataset preparation
4. data overview 1. Attributes in Google playstore 2. User review
Attributes
5. Overview of analysis
6. Exploratory data analysis
7. Conclusion



- The Google Play Store is the centre of Google Play; from there, you can access the vast selection of apps and games created by Google
- Over 2 million apps and games are available on the google play store for the billions of users worldwide.
- Interestingly, the bulk of the apps in a variety of categories, including social media, entertainment, productivity, and photography, are available for free on the Play Store, which offers both free and paid apps.

➤ Problem Statement:

- The official distribution centre for Android apps and some other digital media, such as music, movies, and books, is the Google Play store. More than 2.5 million apps are available on the Google Play Store. To determine the crucial elements that affect app usage and success, We are going to do exploratory data analysis the data in both datasets.
- One dataset contains general information, whereas the other contains user reviews for the relevant Applications.

➤ looking at the dataset :

- 1 . **Playstore data_csv.**

includes all of the application's information from the Google Play Store. A certain app can be described by 13 different factors. it features 9660 unique apps.

- 2. **User Review_csv.**

5 attributes are included in this csv file: App, translated review, sentiment, sentiment polarity, and sentiment subjectivity.

➤ Dataset preparation :

- **Loading the data sets:** we have two different data set , first data set is playstore and second data set is users of reviews of the app.
- **Import Libraries:** Numpy , Pandas , matplotlib , seaborn
- **cleaning the data:** drop down missing value Identifying and removing outliers , drop down duplicate data.
- **Data Imputation:** Price, install, and reviews are converted into numerical values.
- **EDA :** combining graphs or charts and data visualisation techniques to analyse the data sets and highlight their key features.

➤ Attributes in Google Play store

- **App-** Name of Application.
- **Rating-** The overall app rating is shown in this column.
- **Size-** Size of Application .
- **Category-** The category to which the app belongs is contained in this column.
- **Reviews-** The number of reviews received by App.
- **Installs-** the number of times application has been installed.
- **Price-** Price of Application.
- **Type-** Whether the app is free or paid for is shown in this column.
- **Content Rating-** Suitable age for given application.
- **Genres-** The information about the app's genre is contained in this column.
- **Last Updated-** when the most recent app update was made available.
- **Current version-** it shows the current version of the application.
- **Android version-** Information about the Android versions that the app is compatible with is contained in this column.

➤ User Review Attributes

- **App-** Name of App
- **Translated Review-** User review
- **Sentiment-** Positive/Negative/Neutral
- **Sentiment Polarity-** Sentiment polarity score
- **Sentiment Subjectivity-** Sentiment subjectivity score

➤ OVERVIEW OF ANALYSIS

Data Cleaning

- Understand the structure of the dataset and clean data before analysis

Data Exploration

- Visual exploration is used to identify initial patterns, traits, and points of interest.

Predictive Modeling

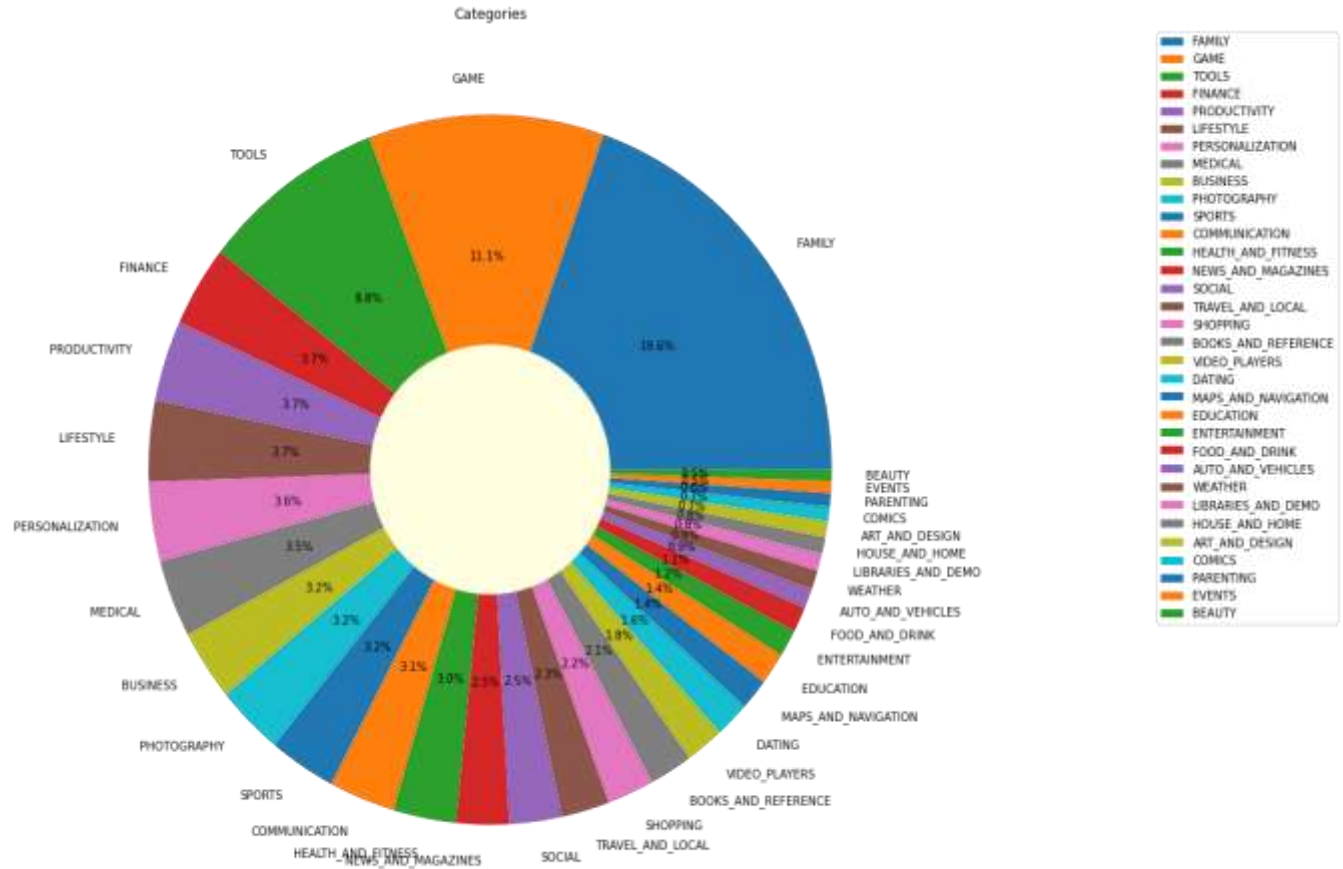
- Formulate a statistical model to forecast an outcome using relevant predictors



Exploratory Data Analysis

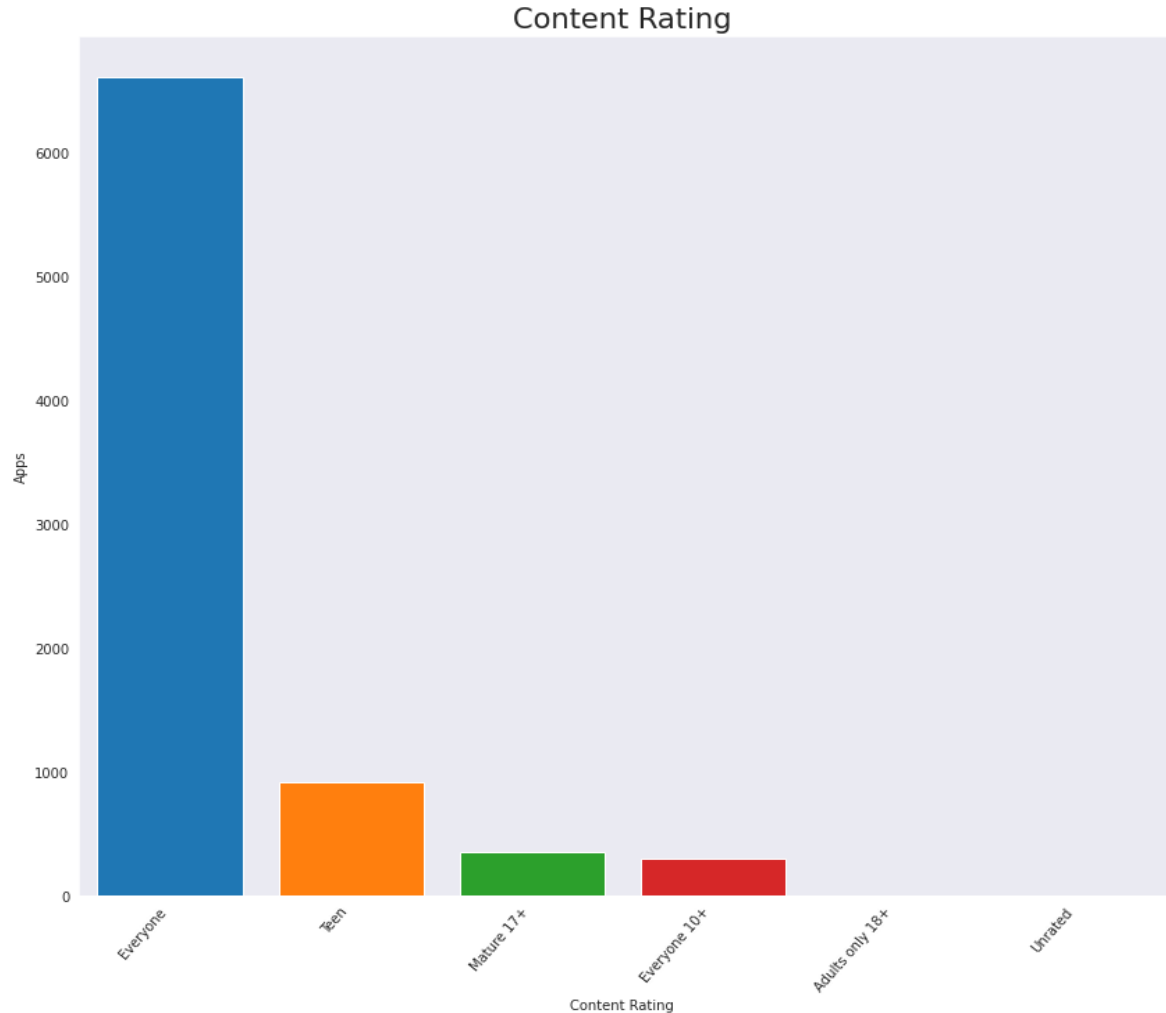
AI

- Uncertain , finance , Productivity, and lifestyle, apps are also among the most popular ones.

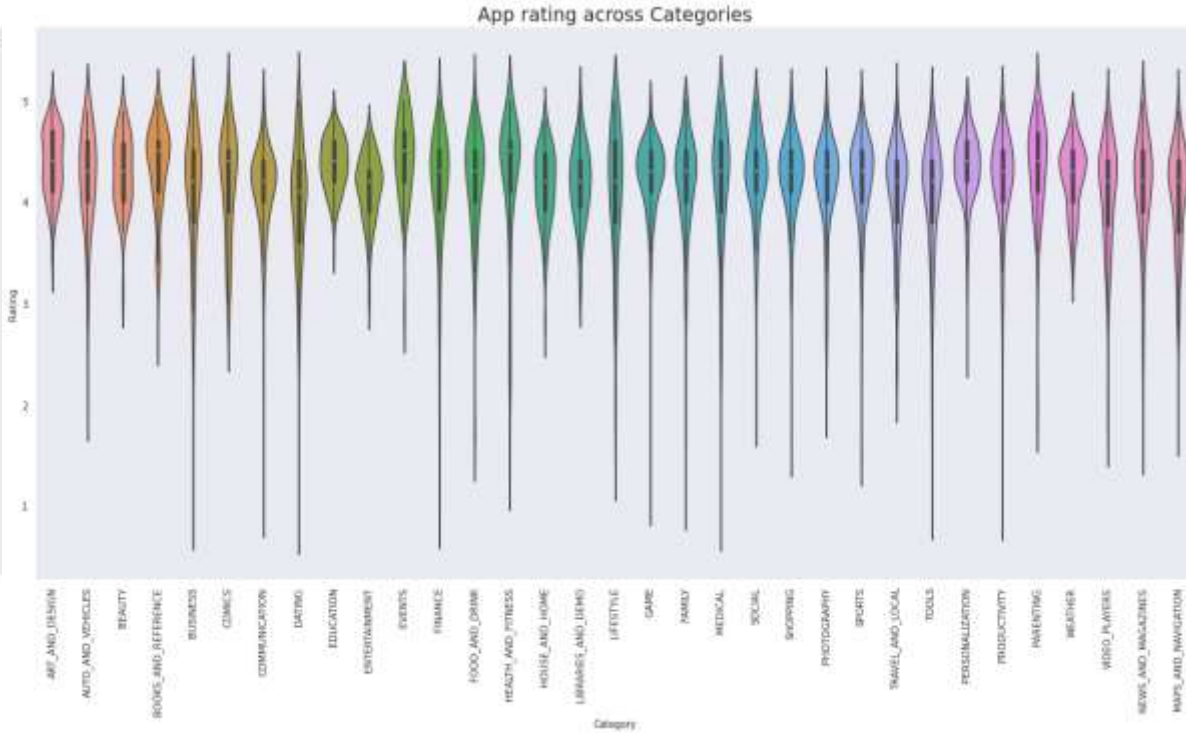
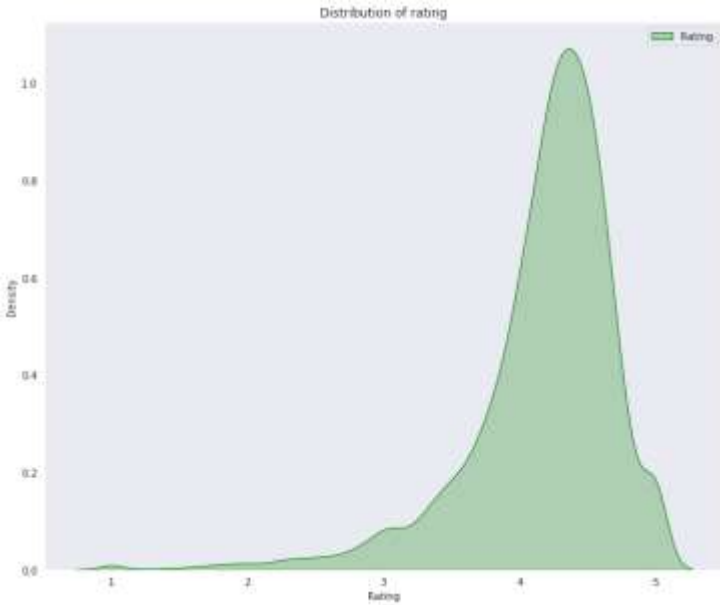


Content Rating

- *The (Everyone) category has the most apps, as can be seen from the graph*
- In the Google Play store, the bulk of the apps (81.8%) are universally usable. Different age restrictions apply to the remaining apps.



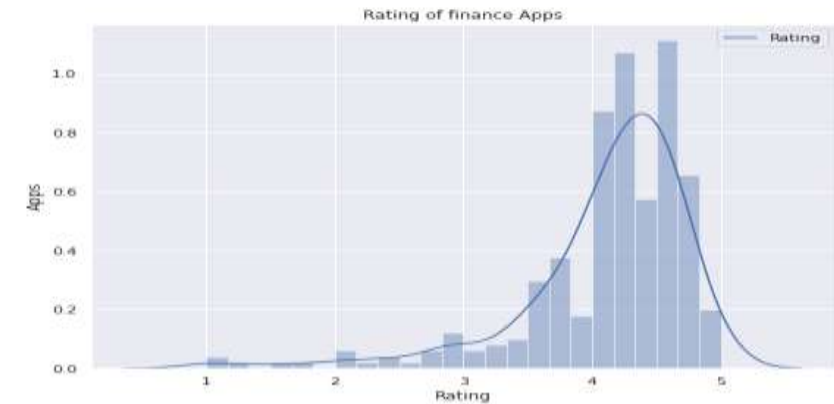
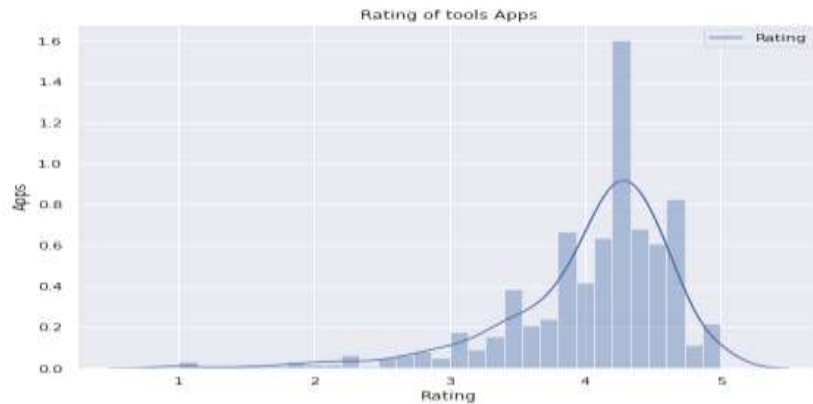
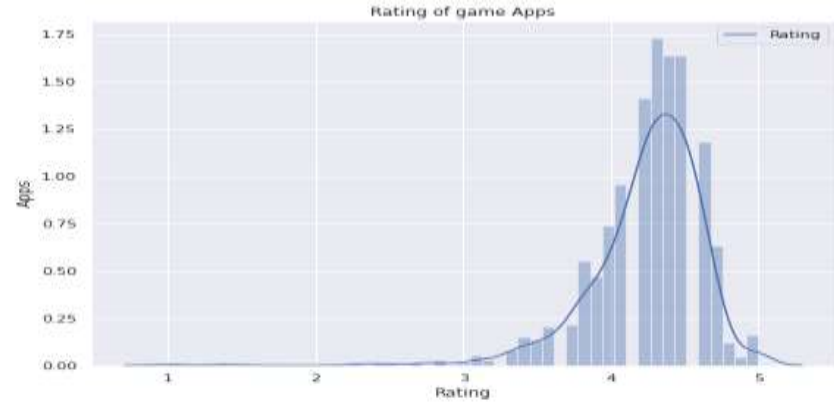
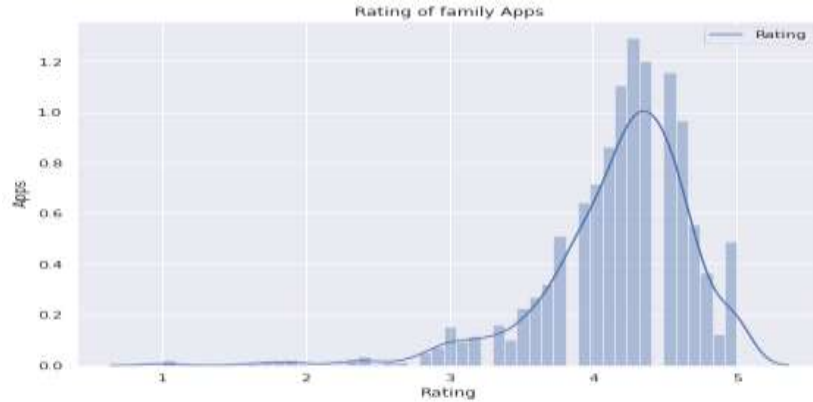
Rating Distribution



According to the graph above, the majority of applications in the Google Play store have ratings between 3.5 to 4.8.

According to the graph above, the two categories of applications with the greatest quality and 50% of apps with ratings greater than 4.5 are health and fitness and book and reference applications, while the dating category is lower than Average.

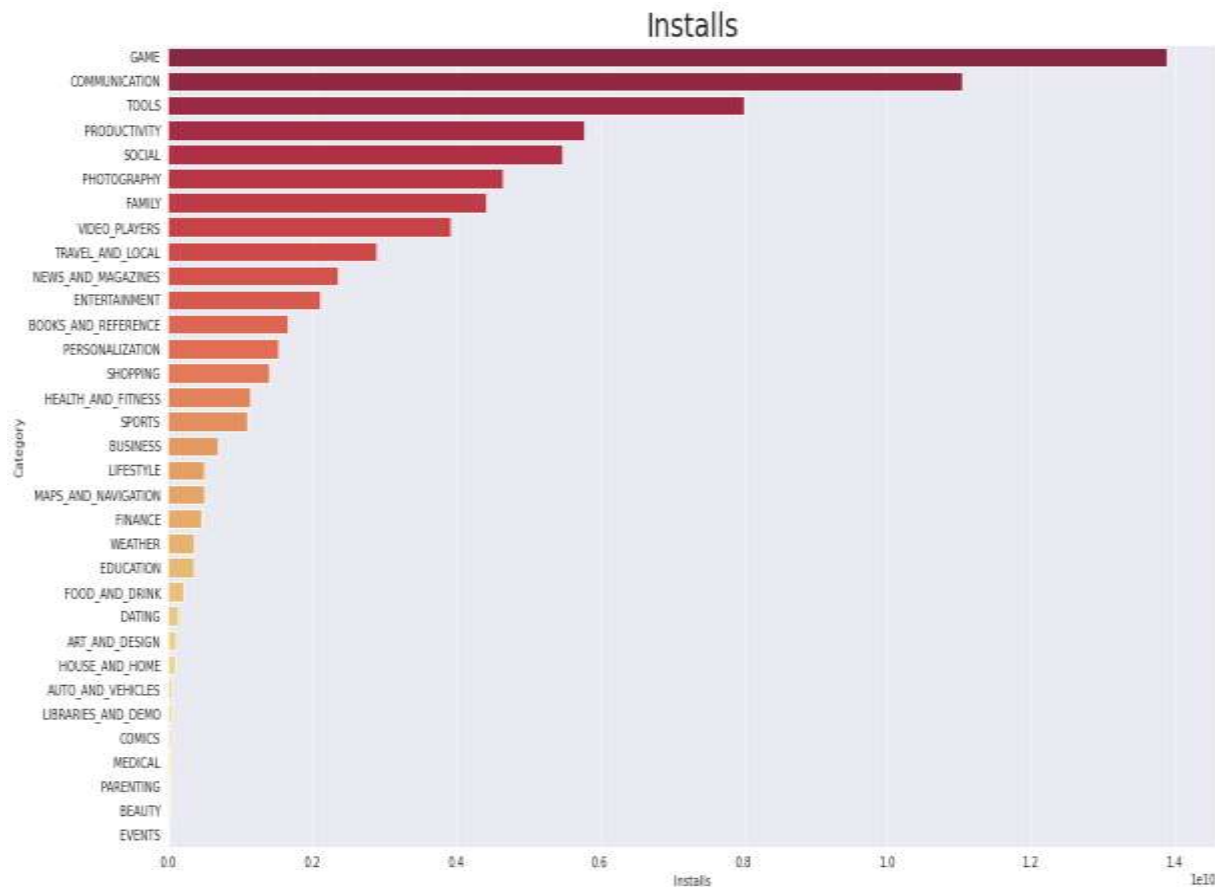
Top four categories with highest rating



According to the graph above, the family category has higher ratings between 3 to 5, games have between 3.5 to 4.8, tools have the greatest ratings between 3.8 to 4.9, and financial category applications have ratings between 3.6 and 5. It is concluded that all the top four category has average rating between 4 to 4.8.

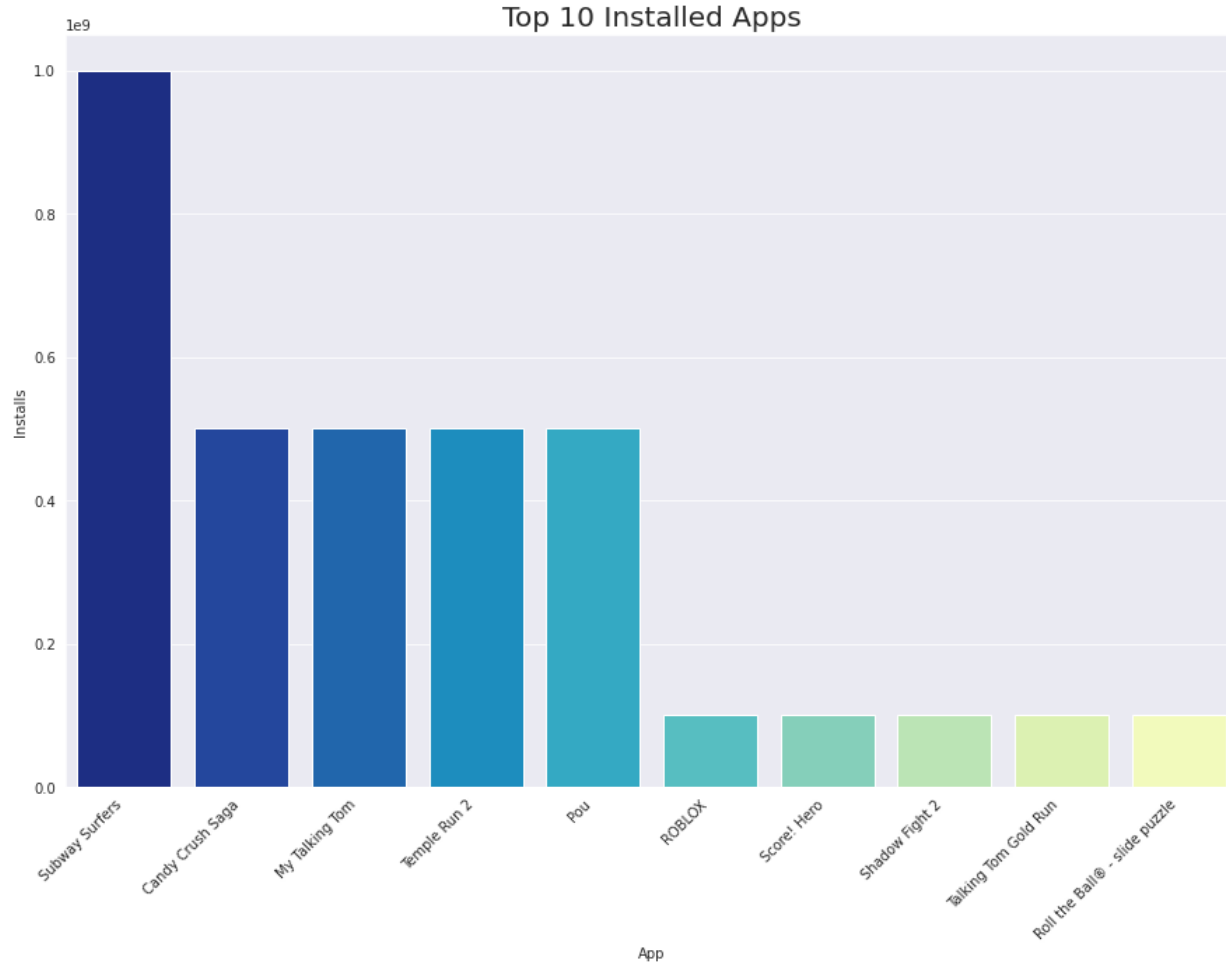
Categories with the most installations

As we have seen from the above plot that (Game) , (communiucation) and (tools) category has the most number of install



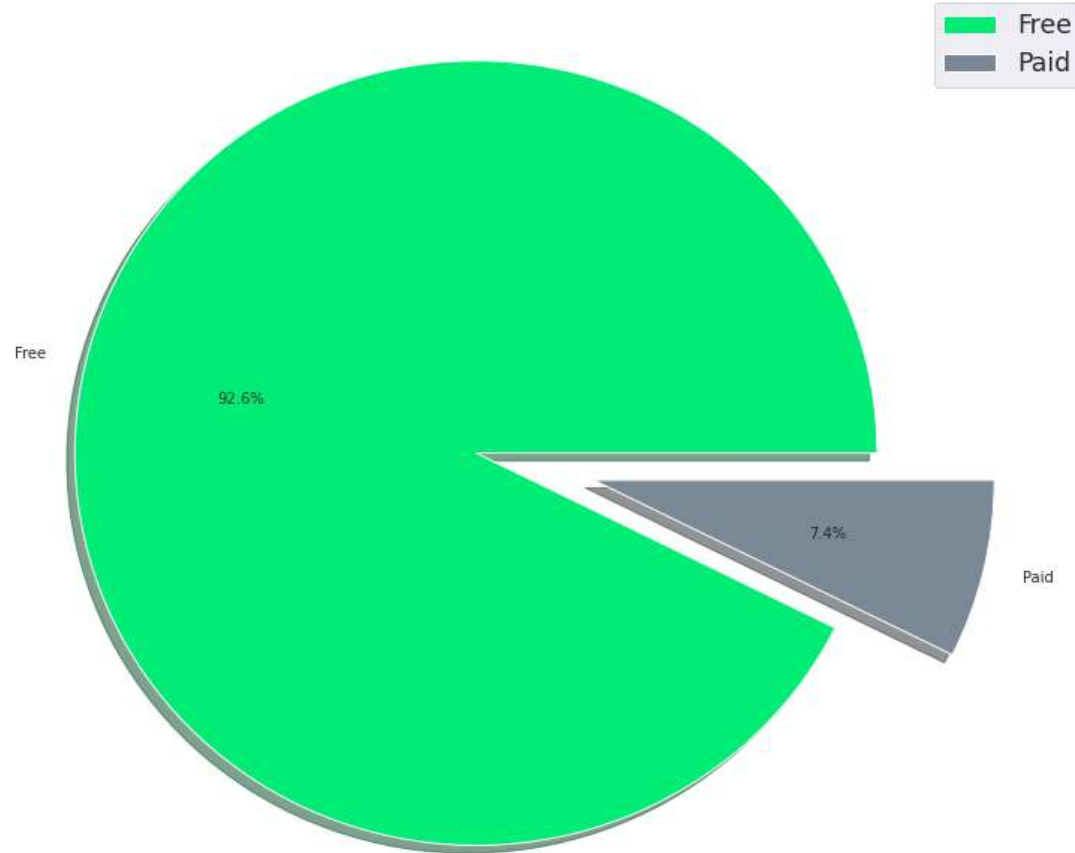
Top 10 Installed Applications

As we have seen in this bar chart, (Temple Run 2), (Candy Crush Saga), and (Subway Surfers) have the most number of installations in the game category.



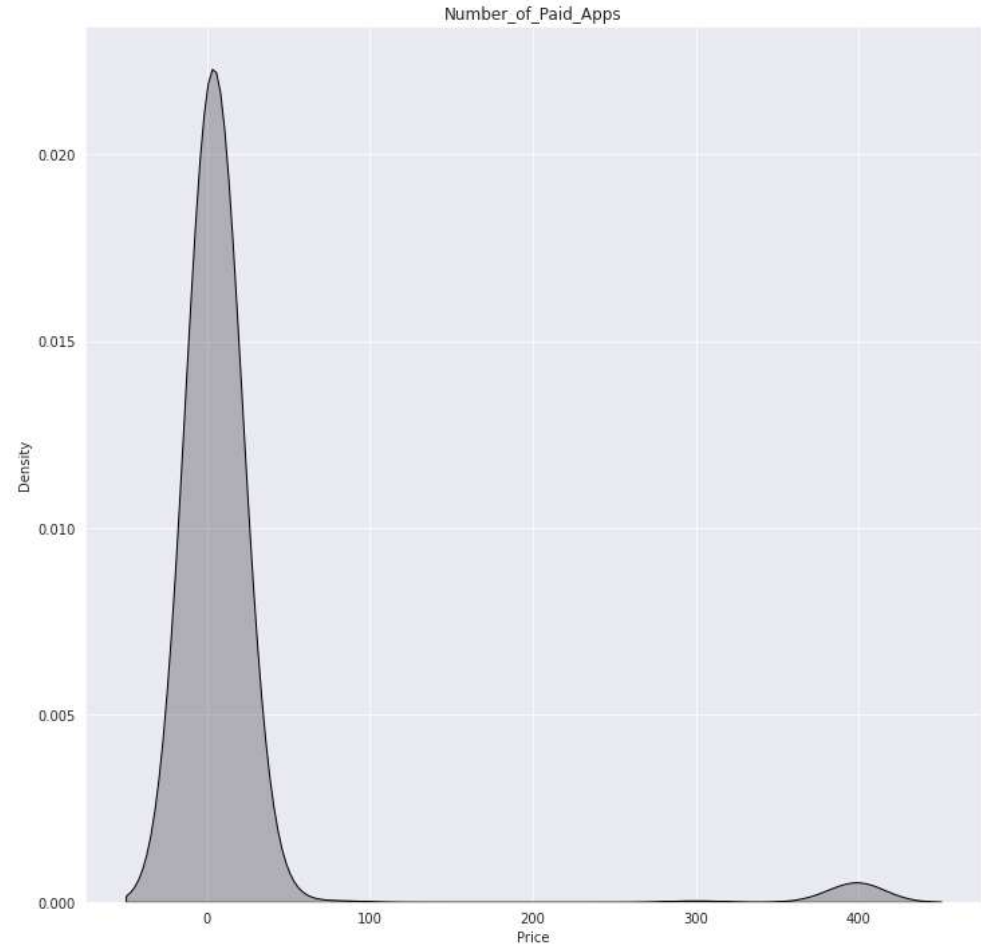
❑ Distribution of free and paid apps in the Google Play store

We have observed that 92.6% of the apps in the Google Play store are free, while 7.4% are paid. it means that on playstore the certain app categories have more free apps available for download than others



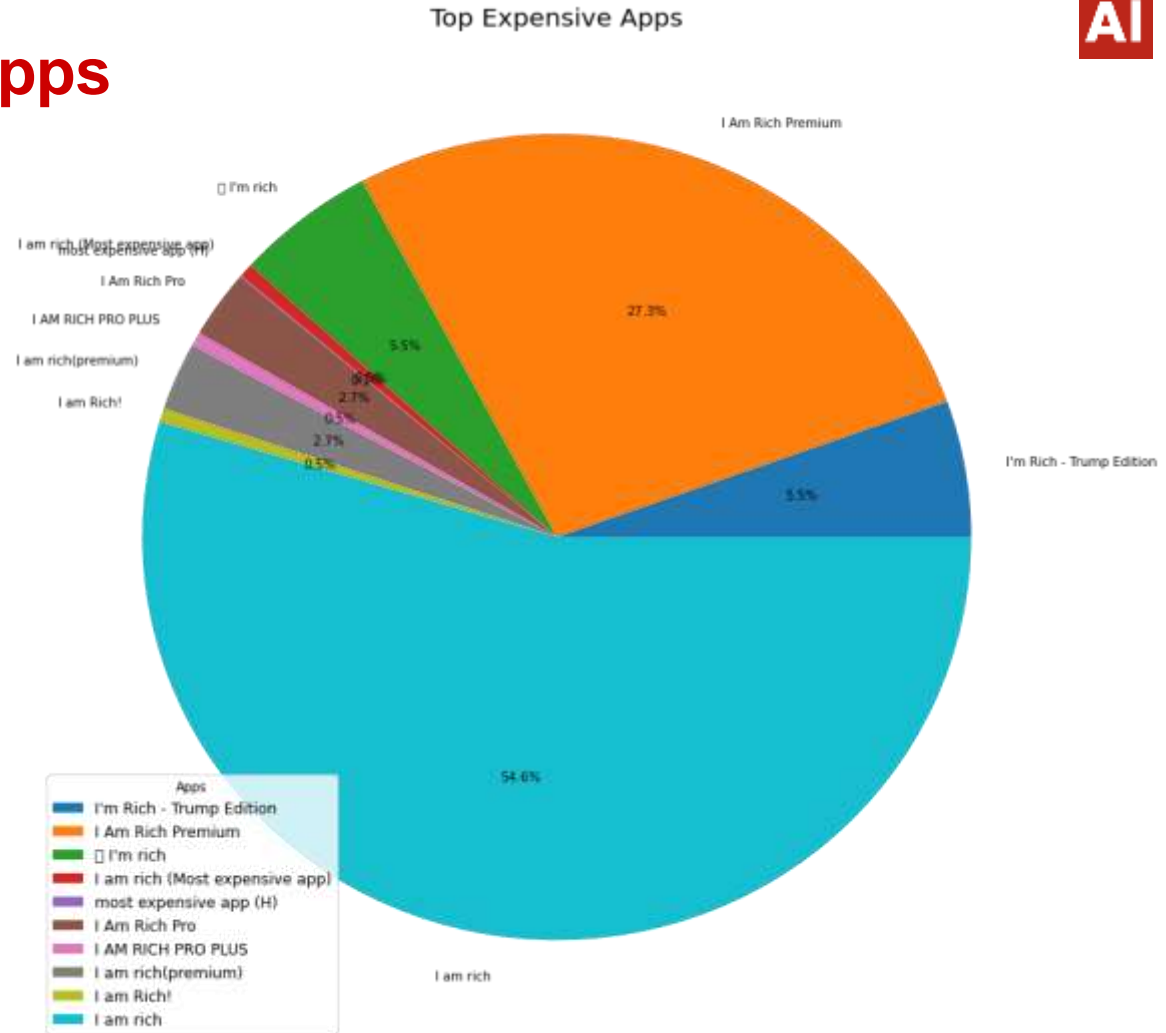
Number of Paid Apps

According to the graph , there are 645 paid apps overall, with the majority of them costing between 0 to 100 rupees, and less of them are priced between 300 to 400 rupees.



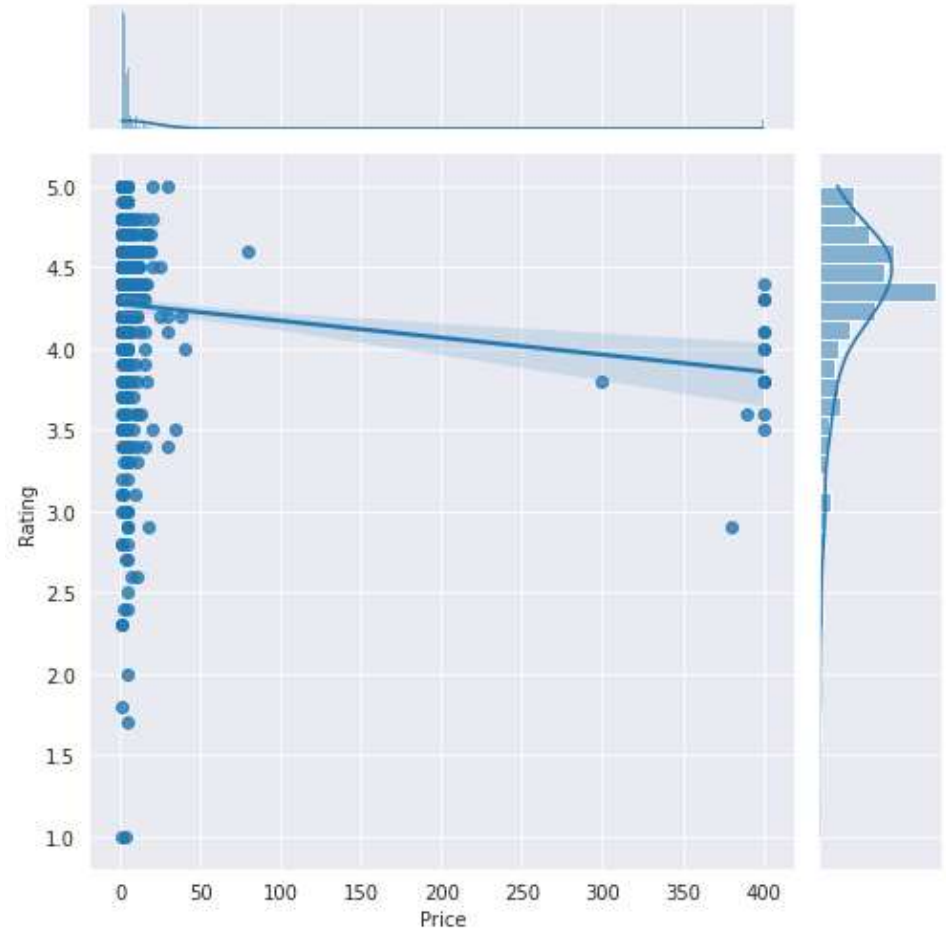
Top 10 expensive Apps

We can see from the following graph that the Google Play Store's most expensive app is "I am Rich," followed by "I am Rich Premium."

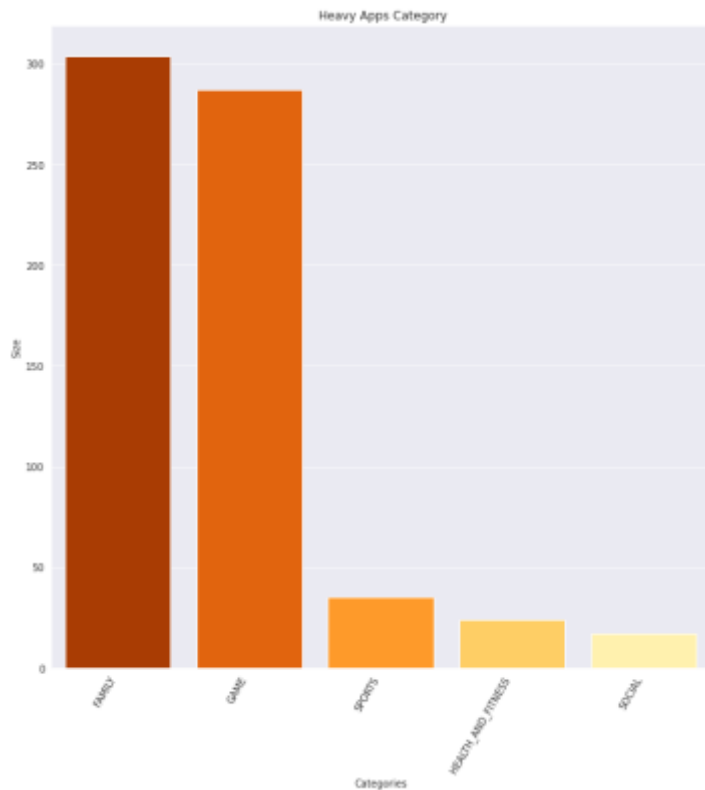


□ Price Vs Rating

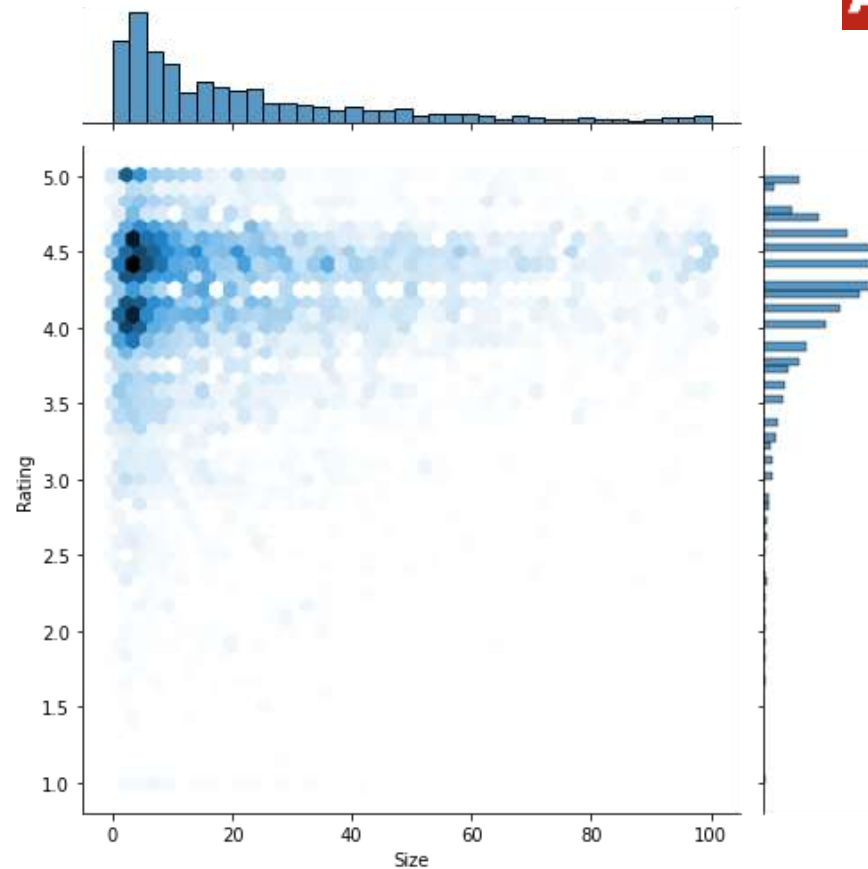
As we have seen the correlation between price and rating in the above graph found that The majority of highly rated apps have best pricing between ~1 to 30 . there are very few Apps above 350.



Size Analysis



The majority of large size Applications (>50MB) fall within the game and family Category.



We discover that the bulk of highly rated apps (rated over 4) fall between 2 MB to 40 MB in size.

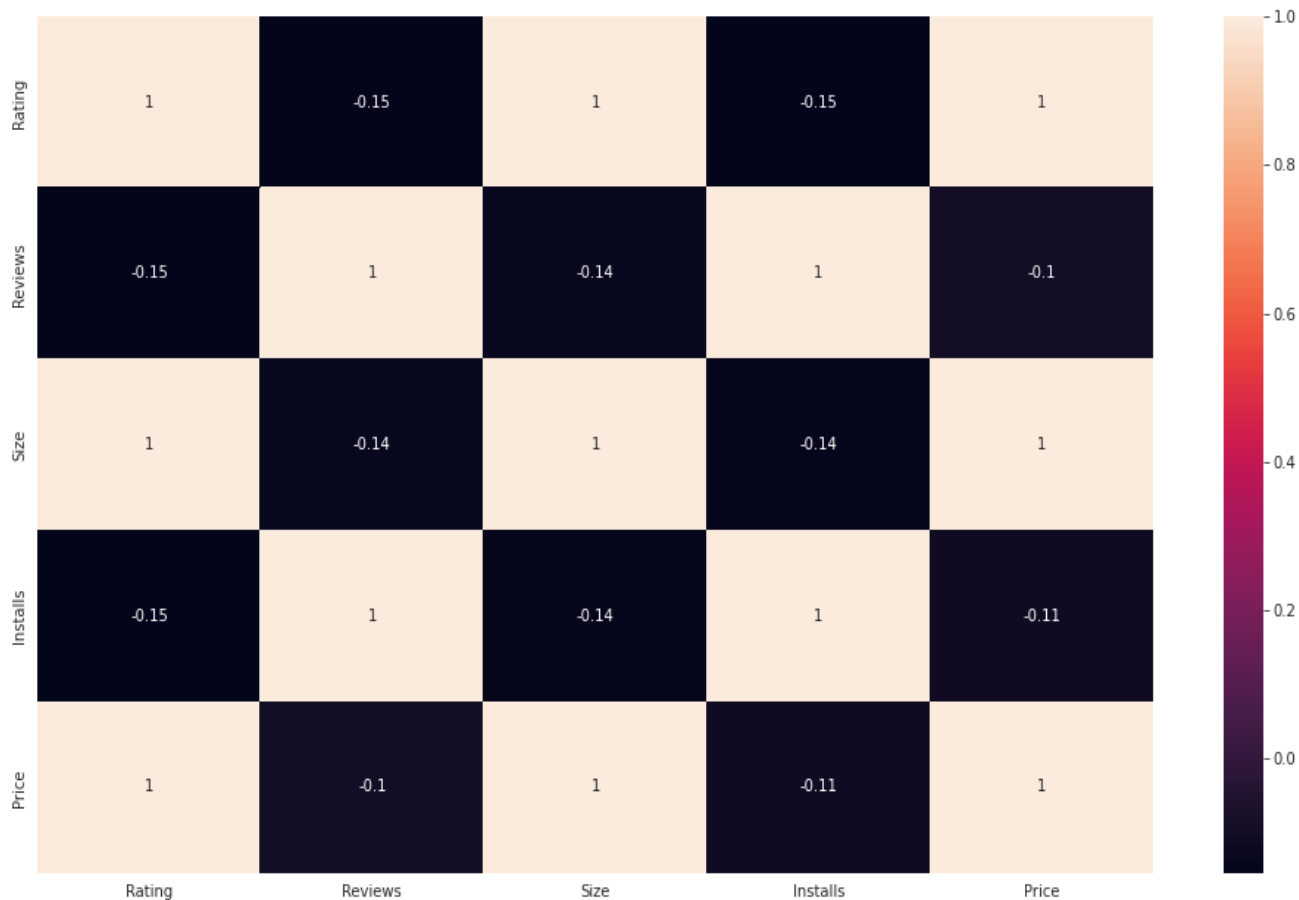
Pairwise plot of all the numerical interactions.



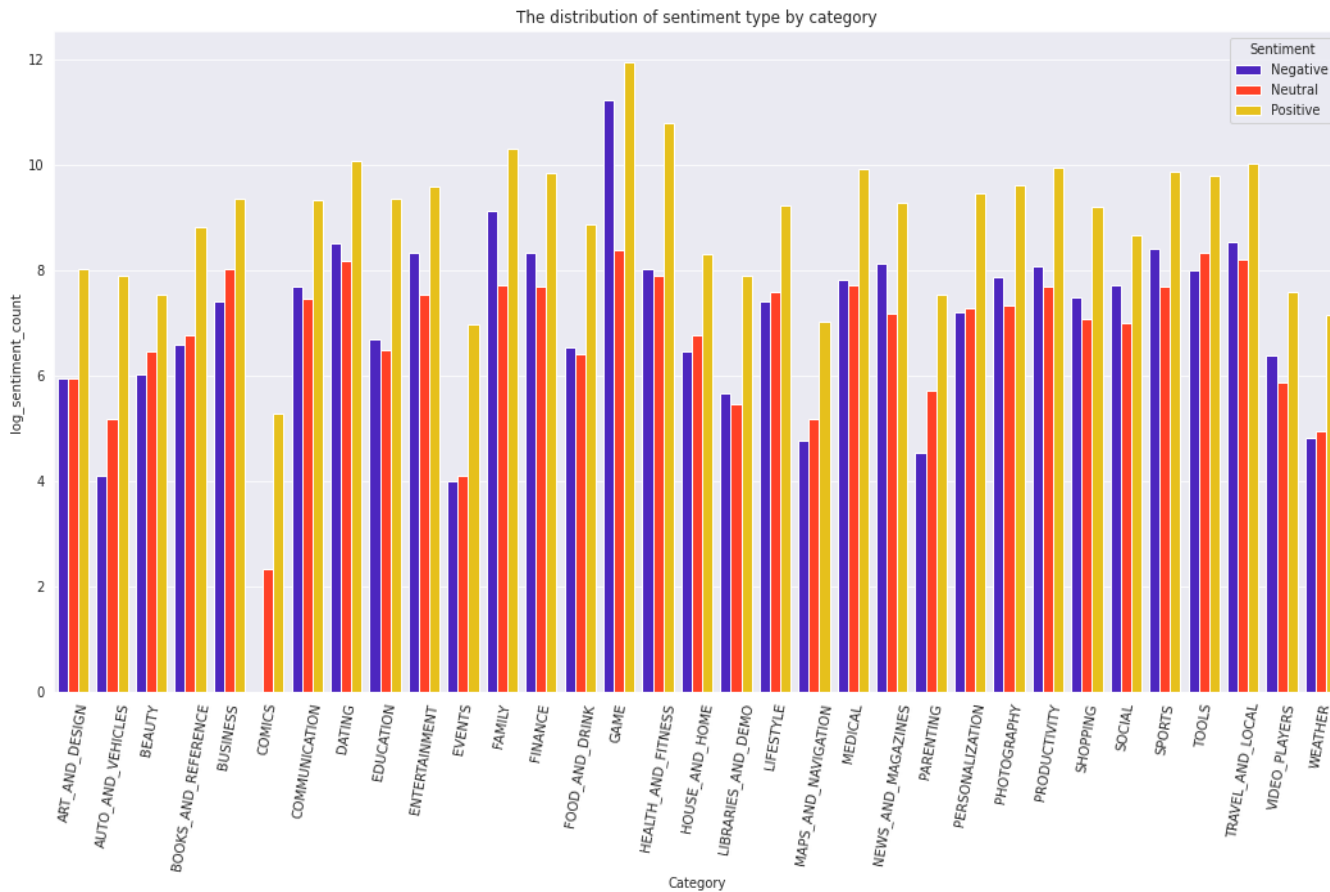
Heatmap Correlation

■ *There is a 0.15 moderately favourable relation between both the number of reviews and installations. That implies that users are more inclined to download an app if so many people have evaluated it.*

■ *Price is slightly negative correlated with rating , reviews and installs*



□ Type of review of sentiment distribution



▪ we found that health and fitness category has achieved highest number positive reviews followed by game category .

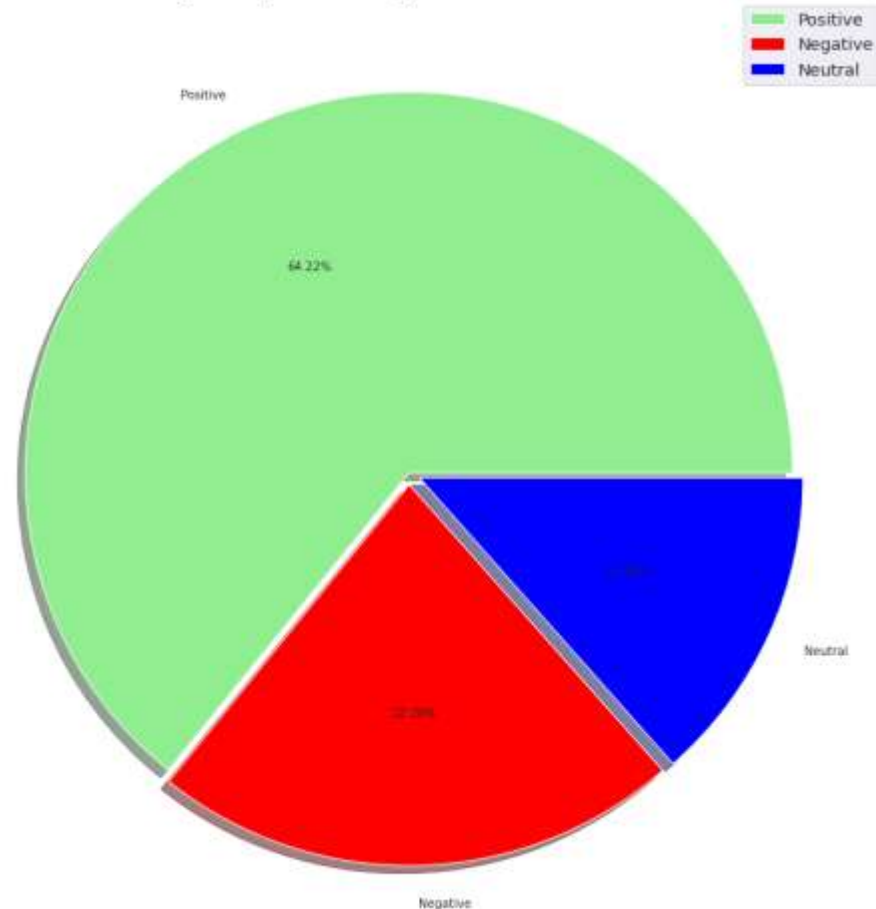
▪ Game is the category which contain larger amount of reviews but they have huge portion of both positive and negative sentiments

percentage of Review Sentiments

- There are 816 unique apps in the combined dataset of user reviews and Play store apps.
- Positive reviews outnumber negative and neutral ones, as can be seen from the plot, by a significant margin.

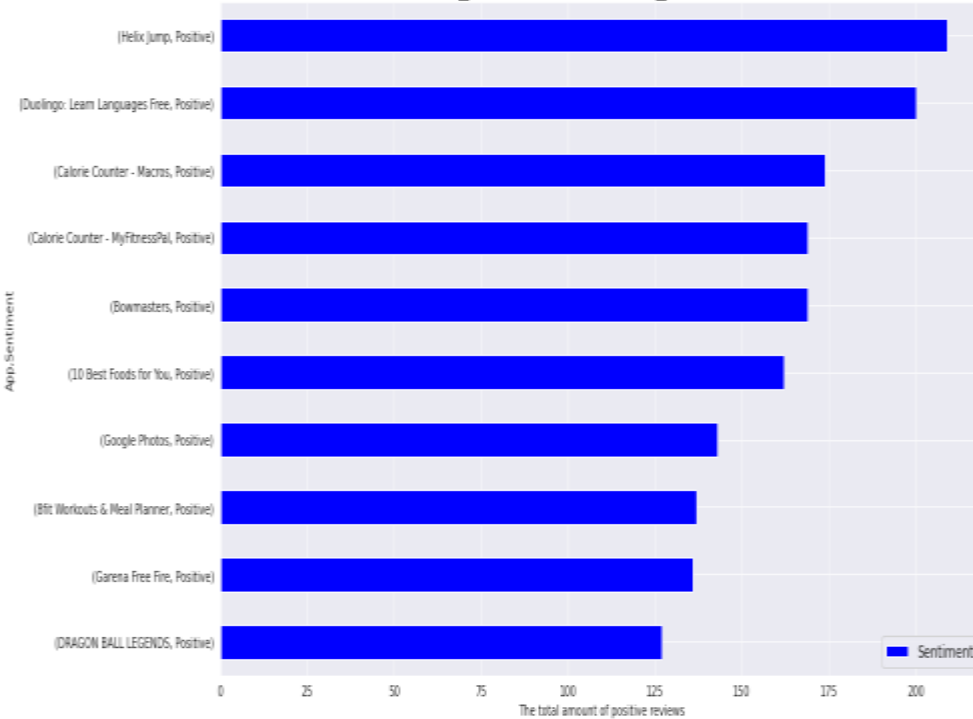
*Positive Reviews is **64.22%**.
negative Reviews is **22.28%**.
Neutral Reviews is **13.50%**.*

showing the percentage of Review Sentiments

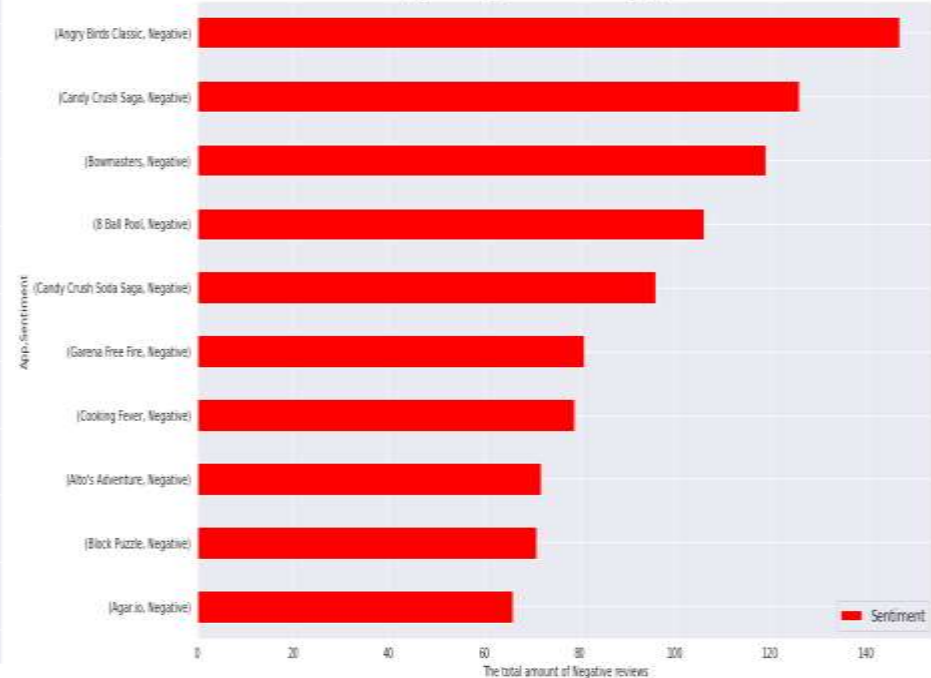


Top Positive and Negative reviews Apps

Top_10 Positive Review_Application



Top_10 Negative Review_Application

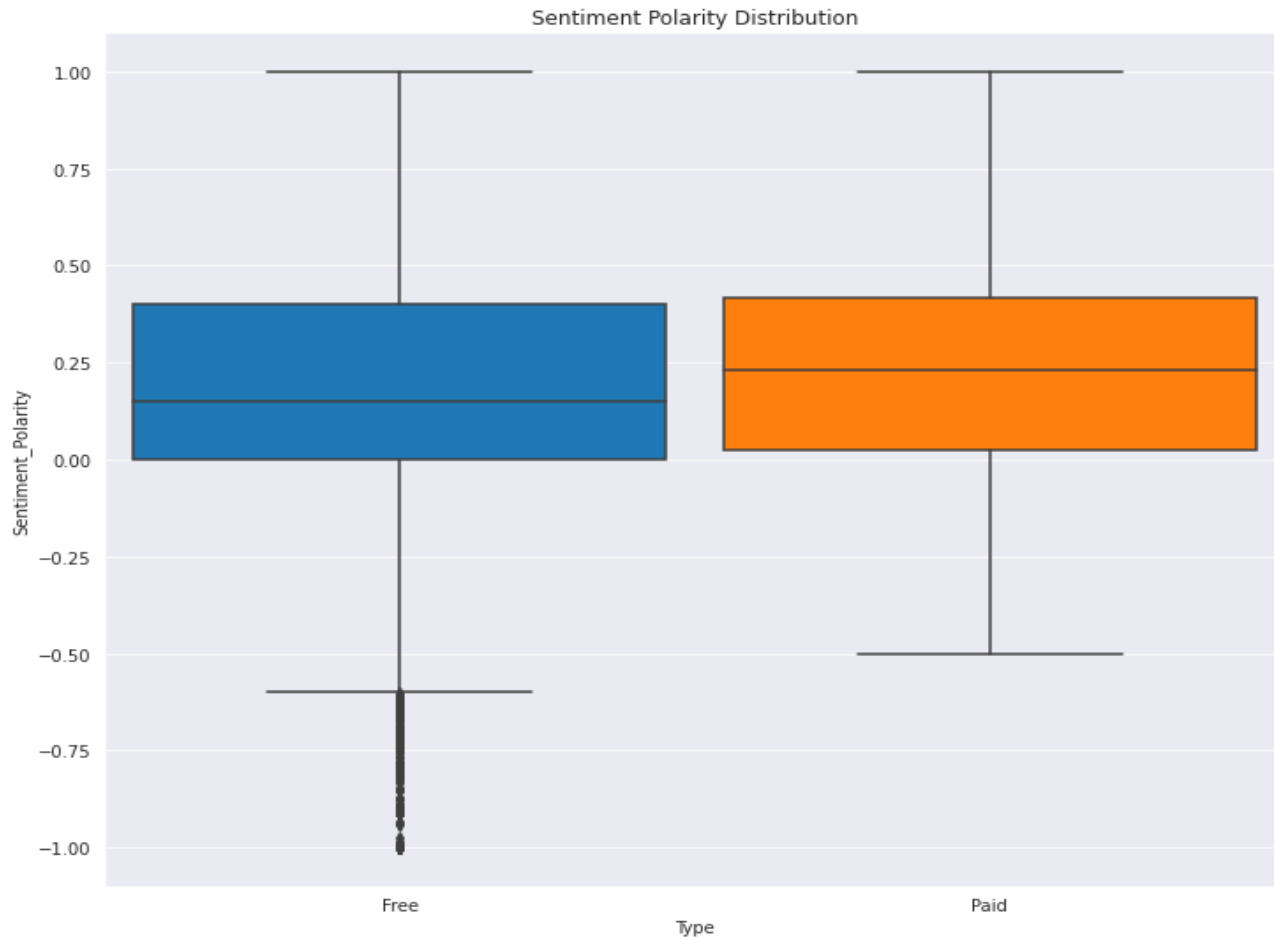


We can see from the graph that the apps with the most positive user reviews are the Helix Jump, Duolingo: Learn Language Free, and Calorie Counter.

As we are seeing that the apps with the most Negative reviews are (Angry Bird), (candy crush saga), (Bowmasters)

Sentiment Polarity Analysis

user reviews for both paid and free apps are analysed by plotting the sentiment polarity scores. The outliers on the negative y-axis in the following graph show how unpleasant remarks are more common for free apps, but paying customers are more patient.



Conclusions

- As observed in the above plots, the categories for Family, Games, and Tools contain the greatest number of application-based categories in the Google Play Store.
- According to the graph, the category "Everyone" has the most apps. The vast majority of apps in the Google Play store (81.8%) are accessible to all users.
- According to the above graph, the majority of the apps in the Google Play Store have ratings between 3.5 to 4.8, however the average rating for all apps is relatively high at 4.3.
- We have found from the analysis that the two categories of applications with the greatest quality and 50% of apps with ratings greater than 4.5 are health and fitness and book and reference applications, while the dating category is lower than Average
- it is concluded that all the top four category has average rating between 4 to 4.8
- We found that The majority of installed apps fall under the categories of games, communication, and tools
- The bulk of installed apps, according to research, fell into the game, communication, and tool categories, with Temple Run 2, Candy Crush Saga, and Subway Surfers receiving the most installations in the game category.
- The game and family Category includes the majority of big Applications (>50MB). Despite this, the excellent ratings for these hefty apps suggest that they are hefty for a reason.
- we have observed that 92.6% of the apps in the Google Play store are free, while 7.4% are paid. it means that on playstore the certain app categories have more free apps available for download than others

Conclusions

- According to the graph above, there are 645 paid apps overall, with the majority of them costing between 0 and 100 rupees, and less of them are priced between 300 and 400 rupees.
- I am Rich is the most costly app in the Google Play Store, followed by I am Rich Premium, as per the visualisation.
- found that The majority of highly rated apps have best pricing between ~1 to 30 . there are very few Apps above 350.
- we concluded from visualisation The paid apps have tiny size applications between 2 MB to 40 MB, and they are highly rated. (There is a relationship between size and rating, and trends can be seen between them. We discover that the bulk of highly rated apps (rated over 4) fall between 2 MB to 40 MB in size.)
- There is a 0.15 moderately favourable relation between both the number of reviews and installations. That implies that users are more inclined to download an app if so many people have evaluated it.
- we found that health and fitness category has achieved highest number positive reviews followed by game category , game is the category which contain larger amount of reviews but they have huge portion of both positive and negative sentiments
- Positive reviews outnumber negative and neutral ones, as can be seen from the plot above, by a significant margin.
- Calorie Counter, Duolingo: Learn Language Free, and Helix Jump have received the most favourable user reviews. the most unfavourable reviews received by games Candy Crush Saga and Angry Birds (Bowmasters)
- The outliers on the negative y-axis in the following graph show how unpleasant remarks are more common for free apps, but paying customers are more patient.