Capstone Project 1

Play Store App Review Analysis

Team Members

Shubham Chavhan

Shweta Shivhare

Abhishek Singh

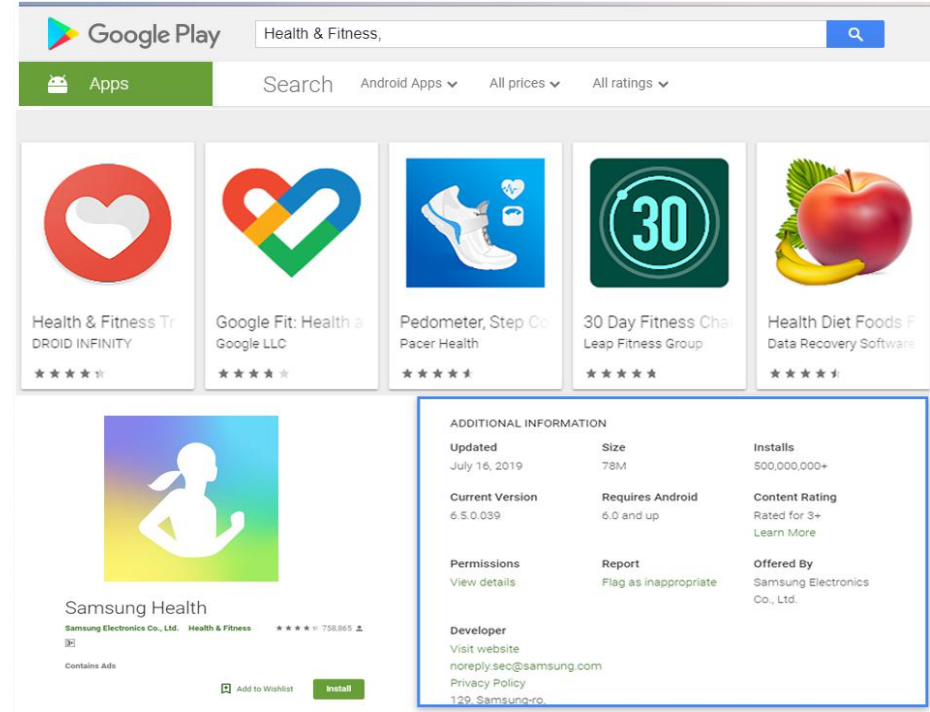Sushmita Chaudhary

Lets Explore

AI

Introdution

❖ The Google Play Store is the official pre-installed app on Android devices.

❖ It allows users to browse and download music, books, movies and lots of applications from Google Play.

❖ It is the largest and most popular Android app store with around three million different apps available for the users to download.

Data Pipeline
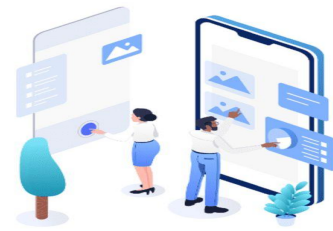
❖ **Data Wrangling-** The process of cleaning and unifying messy and complex data sets for easy access and analysis, which includes treatment of null values and removing duplicates.

❖ **Data Processing-** In this process, we manually examined each feature and processed the null values, by replacing them with the mean or median values, wherever required. We also plotted box plots to detect outliers and transformed data types into another, for easy interpretation.

❖ **EDA -** In this part, we did some exploratory data analysis(EDA) on the selected features to draw conclusions.

Data Summary

Variables Description In Google Play Store:

❖ **Apps -** Application name.

❖ **Categories -** Category of given application.

❖ **Rating -** Rating is a reflection of how users respond to the apps. Rating of app is between 0 and 5.

❖ **Reviews -** A review is judgement or discussion of the quality of applications. Review also means to go over a subject again as part of study or to look at other updated features from time to time.

❖ **<u>Size</u> -** The amount of space required to install the app. This feature shows the size of the application.

❖ **<u>Installs</u> -** This feature represents the number of installations on devices.

❖ **<u>Type</u> -** This feature tells us whether an app is free or paid.

❖ **<u>Price</u> -** This column shows the price of certain apps.

❖ **<u>Content Rating</u> -** Apps use a separate content rating system. Titles rated ALL, have content that may be suitable for all ages. Titles rated 12+ have content that is suitable for ages 12 and older. This feature separates users from accessing content that is irrelevant to them depending on their age.

❖ **<u>Genres</u> -** App genres help us determine the type of apps that are being built. It is a detailed description of the category of app.

❖ **<u>Last Updated</u> -** This feature represents the date when the application was last updated. Updating the apps gives us access to the latest features and it also improves security and stability of the applications.

❖ **<u>Current Version</u> -** A positive integer used as an internal version number. This number is used only to determine whether one version is more recent than the another, with higher numbers indicating more recent versions.

❖ **<u>Android Version</u> -** Android version indicates the version of Android platform. The bigger the version number, the newer the Android is.

Variable Descriptions In User Reviews:
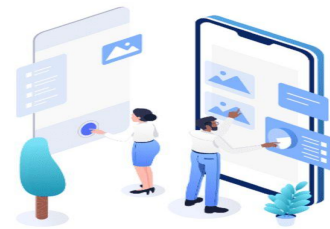
❖ **Apps -** Application name.

❖ **Translated Review -** Review for given app.

❖ **Sentiment -** Boolean values of user translated reviews; 'positive','negative and 'neutral'.

❖ **Sentiment Polarity -** Sentiment polarity for an app defines the orientation of the expressed sentiment. It is calibrated value of translated reviews that varies from -1 to 1.

❖ **Sentiment Subjectivity -** Sentiment Subjectivity is basically pitch of translated reviews that varies from 0 to 1.

Exploratory Data Analysis

Category Wise App Count

❖ Family, Games & Tools were the most common categories, accounting for about 19.5%, 10% and 8.5% of the total number of apps in our dataset, respectively. Beauty was the least prevalent category.
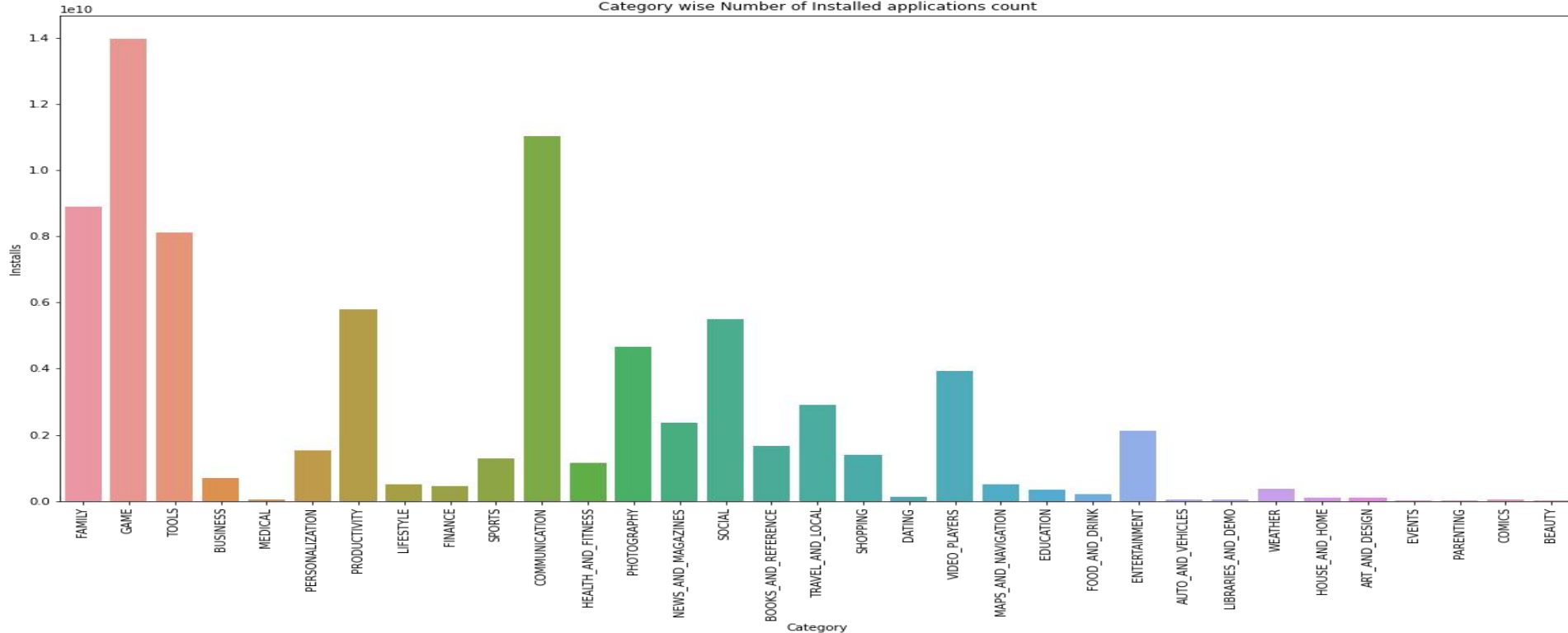

Category wise Play store App count
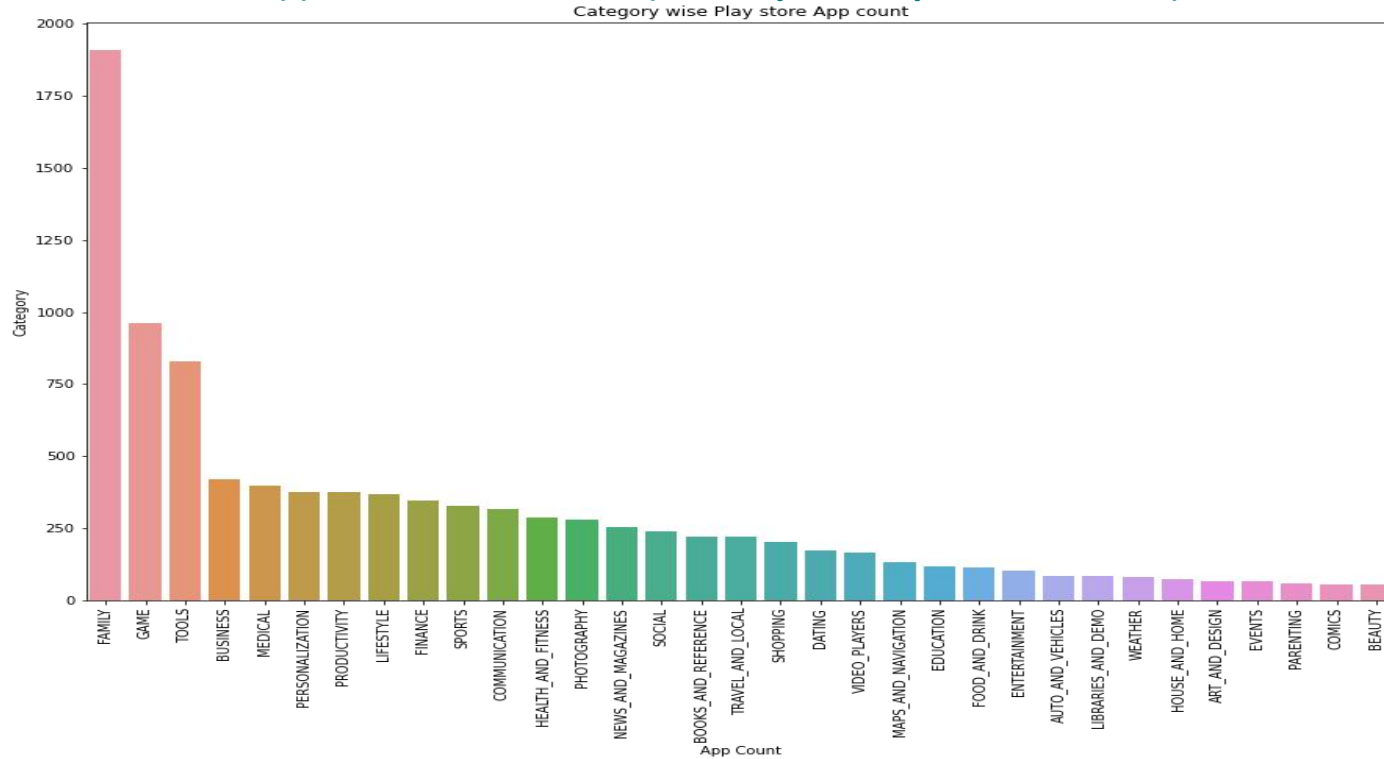
Comparing Category And App Type (Fee, Paid) by count of installs.

We can deduce that The Game and Family category has the highest number of Installation.

No of installs according to App Category and Type
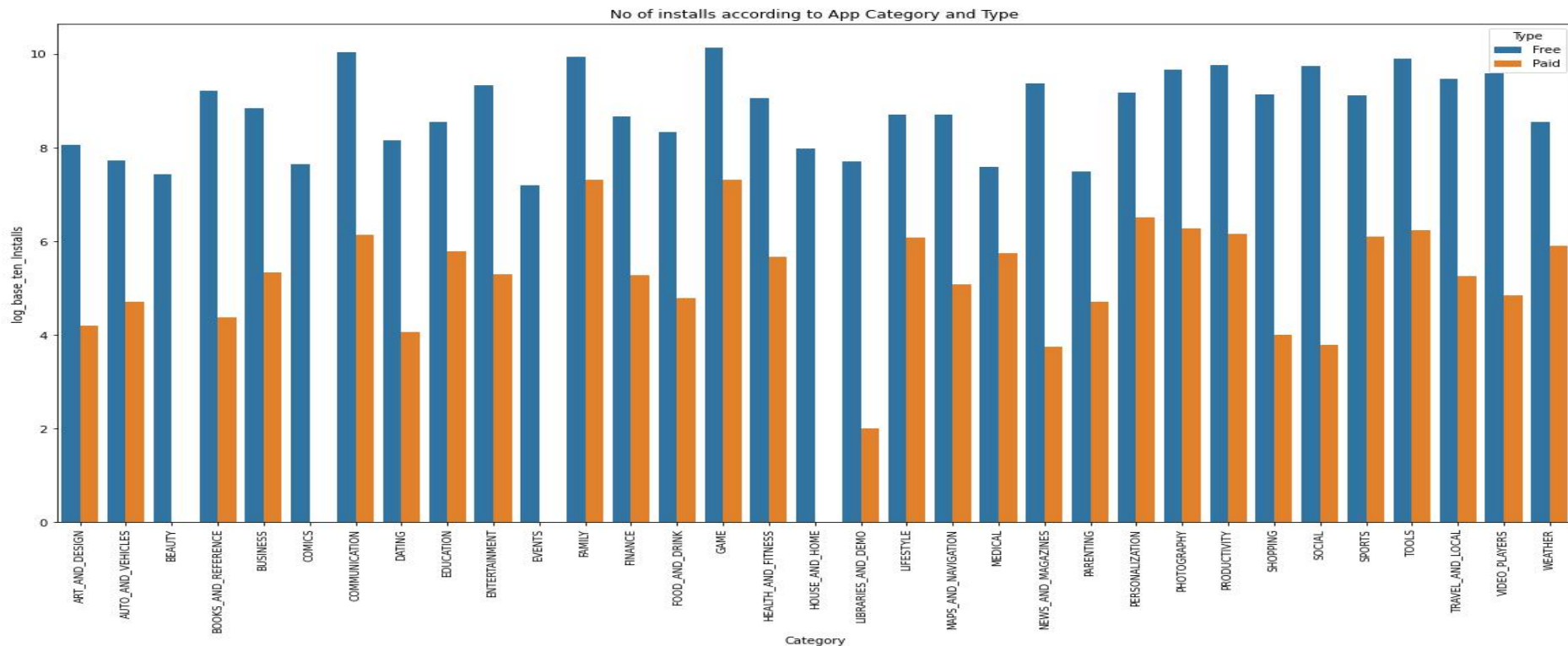
Number of Installed Applications

❖ Here Game, Communication & Family categories have the highest number of apps installed.

❖ Among the apps that had been downloaded between range 0.8 to 1.4+ times were Instagram, Facebook, Google + and Google News as per the present scenario.
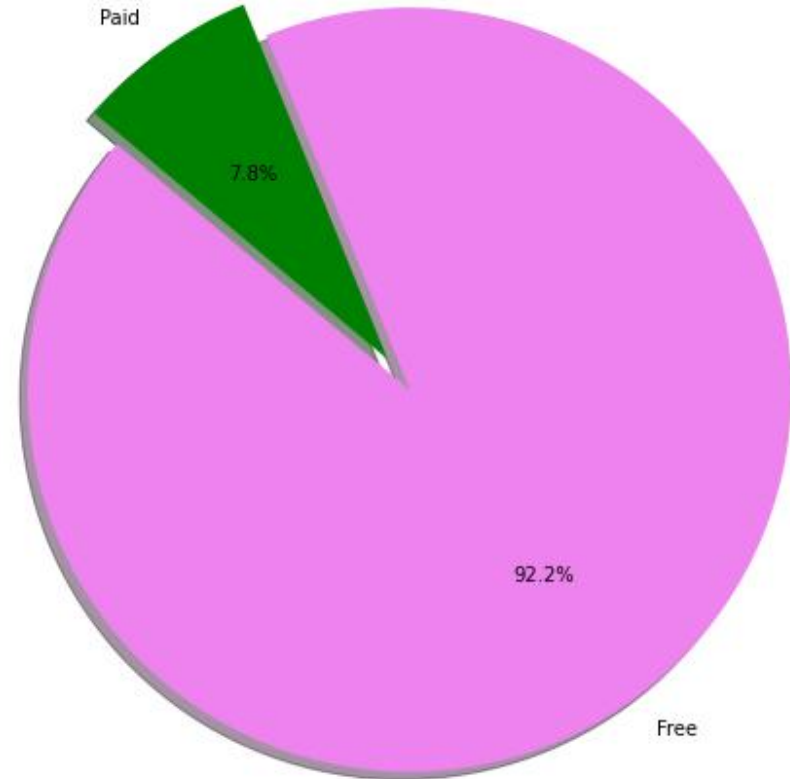

Category wise Number of Installed applications count

Free Vs. Paid Apps

❖ Here we can see that there are 92.19% of apps are free and only 7.81% of Apps are paid on Playstore.
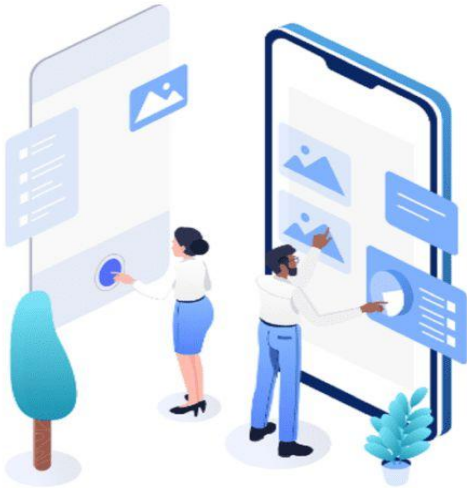
Paid vs Free Apps Percentage in store

Paid

7.8%

92.2%

Free

Category Wise Pricing Visualization

❖ The majority of apps except Finance, Lifestyle, Events & Medical categories were free to install.

❖ Lifestyle, Finance, and Medical categories had the biggest number of paid apps available for download.

Category wise Pricing
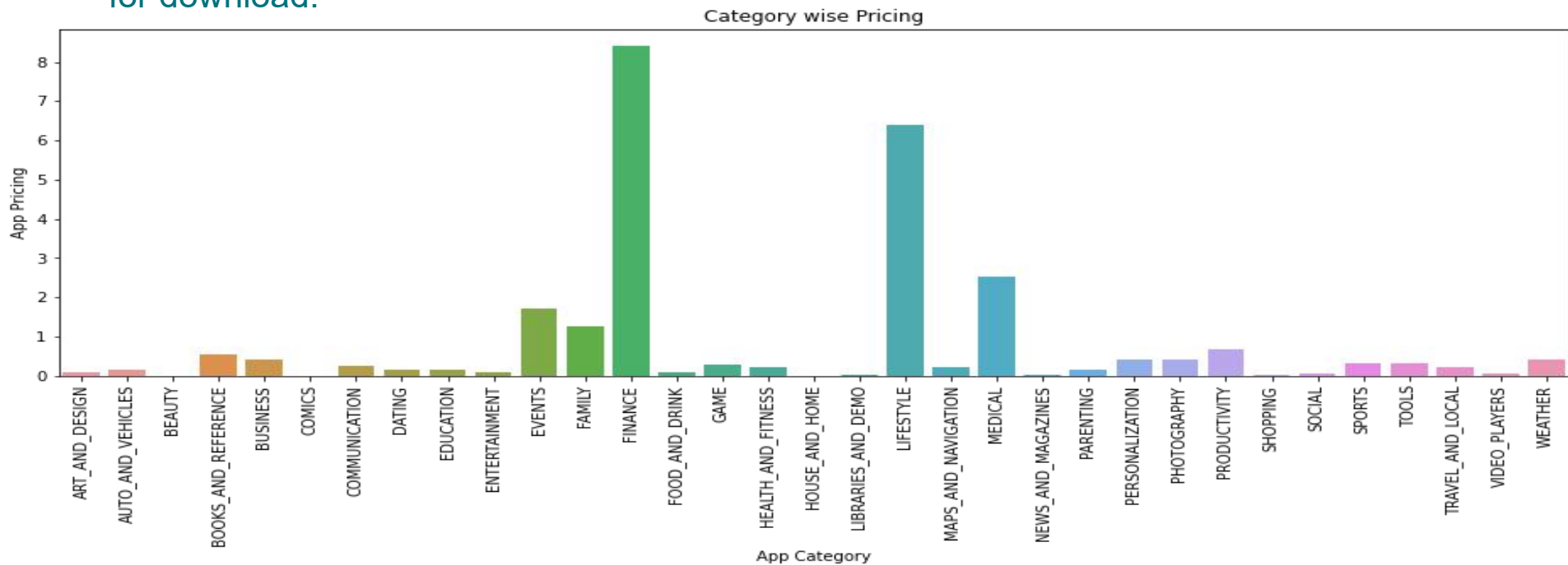
Distribution of User Ratings

❖ This distribution plot shows us the distribution of ratings. Average ratings of the apps are between 4.2 to 4.3

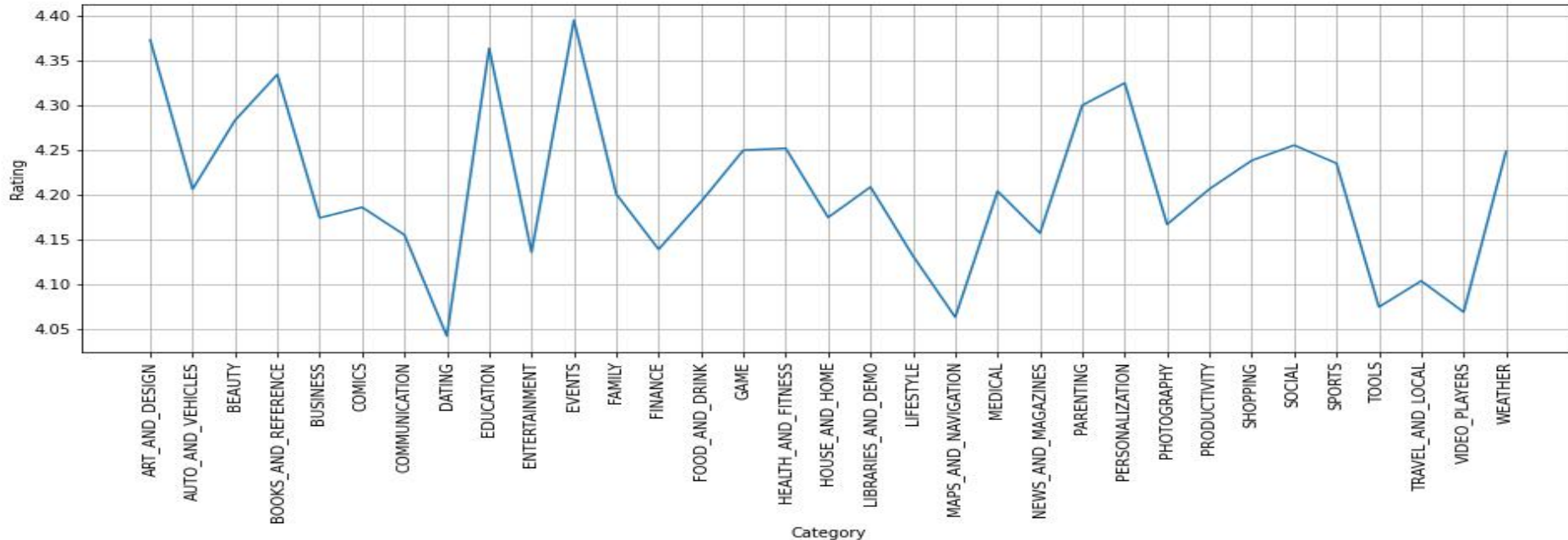❖ Most of the apps are rated within the range of 4.0 to 4.6.
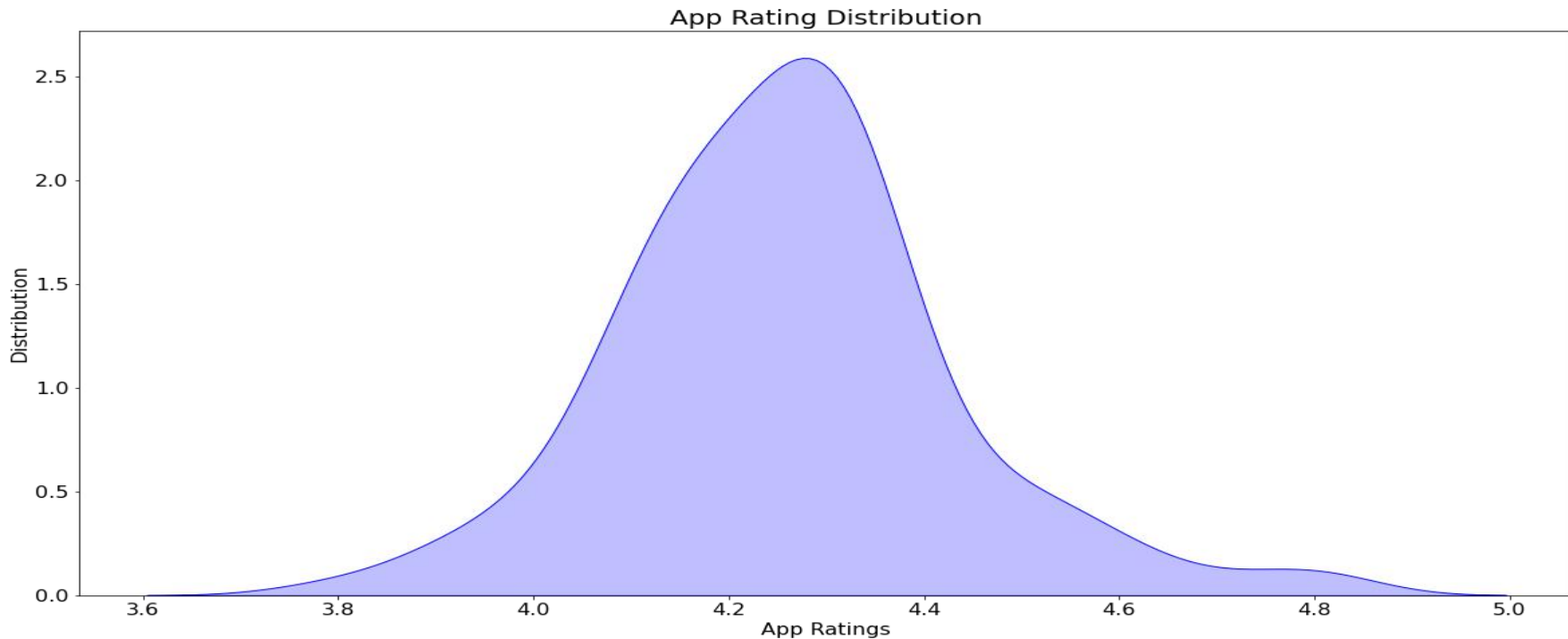
Category Wise User Ratings

AI

A. Highest Ratings apps:
Events, Art and Design, Education categories has got 4 and above ratings.

B. Lowest Ratings apps:
Dating, Maps and navigation, Video players category has got lowest ratings i.e. below 4.

Here Most of the apps are rated between 4.15 to 4.30.

Category Wise User Reviews

❖ This plot shows the mean of each categories with respect to user reviews.
❖ Users tend to leave more reviews for apps in communication, game & social categories.

Revenue And Top Expensive Apps

AI

- ❖ The majority of apps, which generate revenue is between 0 to 10 million .
- ❖ Minecraft is generating highest revenue among all, which is about 70 million.
- ❖ "I am rich" and "I am Rich Premium" are the two most used expensive Apps.

Top Expensive Apps
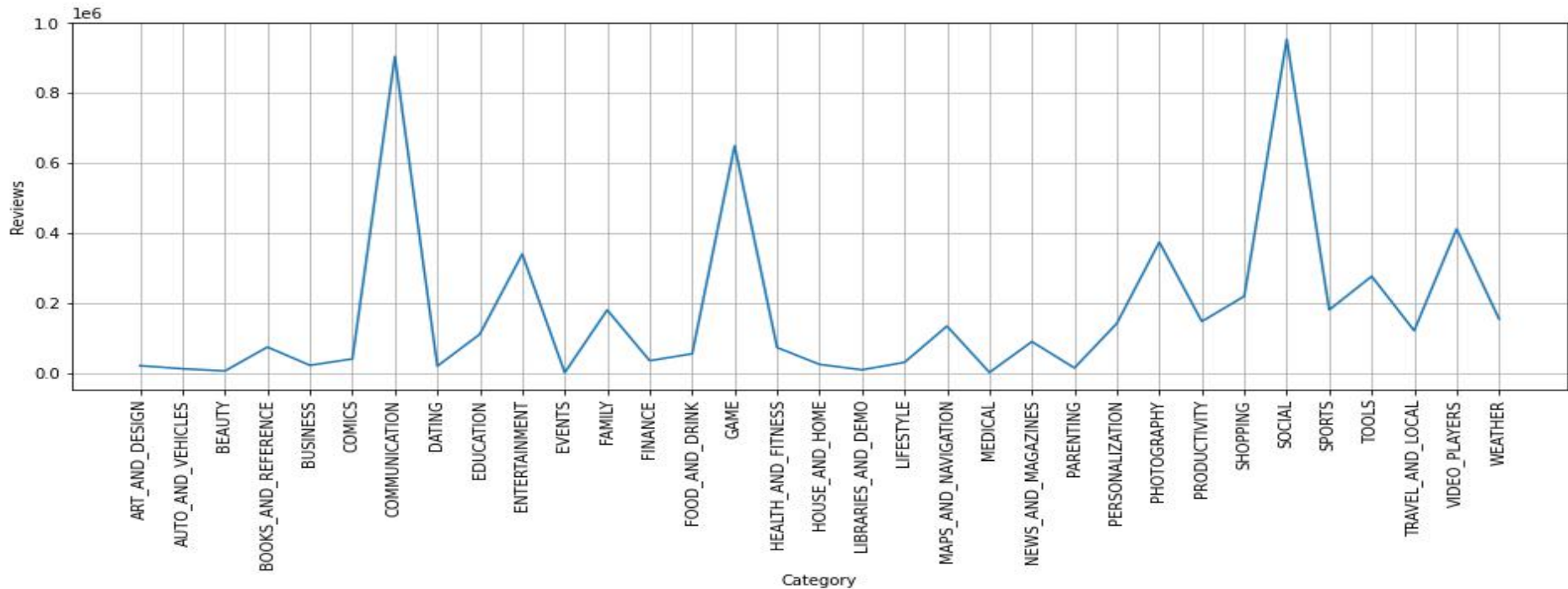
Revenue Generating Apps Graph

Genre wise App rating

❖ Average rating for apps, genre wise is between 4 and 5

Genres wise App Ratings
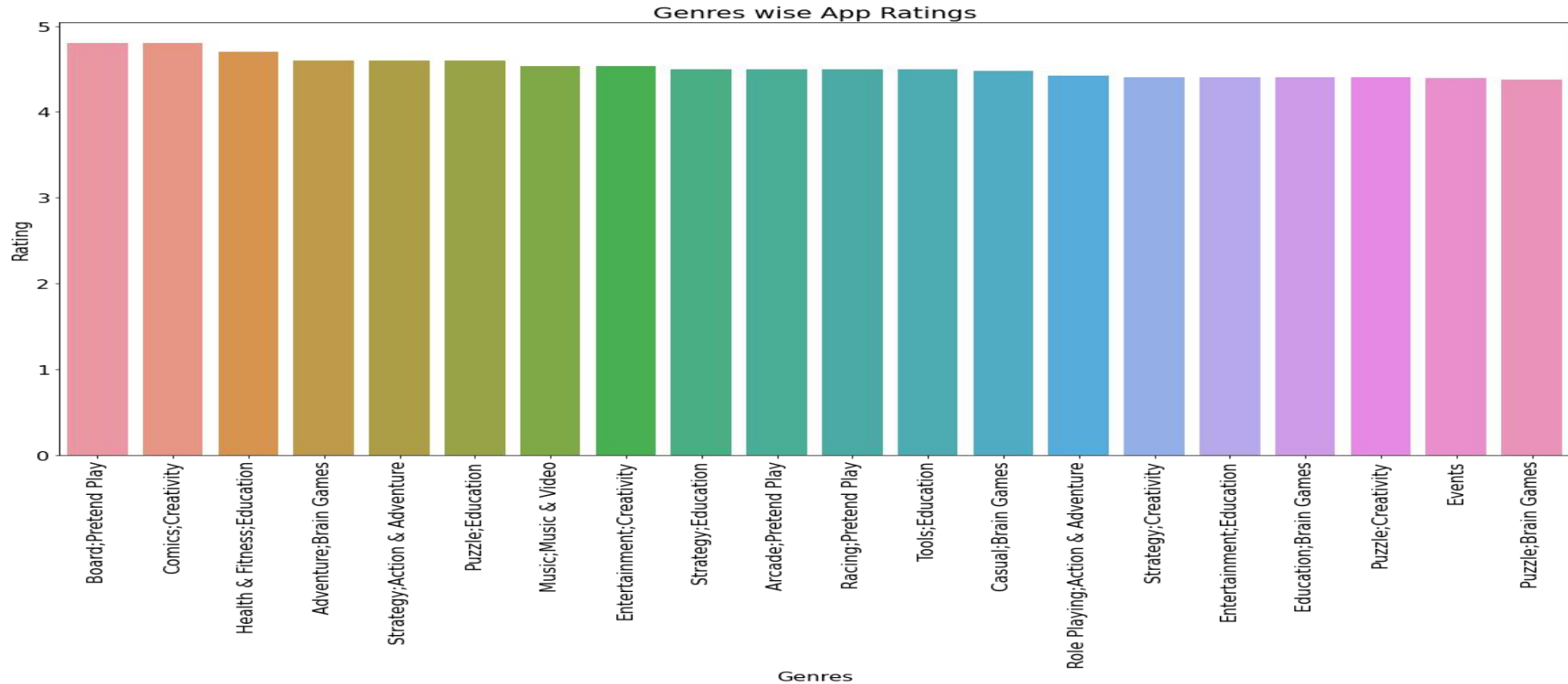
Size Distribution

- ❖ Maximum apps are in the size category of 10000 KB to 30000 KB i.e. 10MB to 30 MB
- ❖ size category of 0 KB (varies with devices) showing highest count
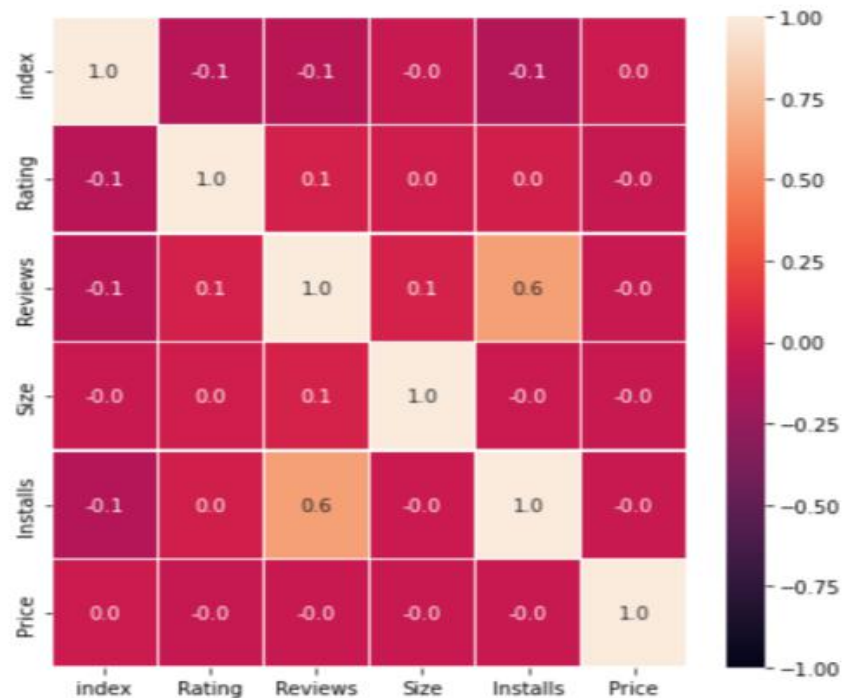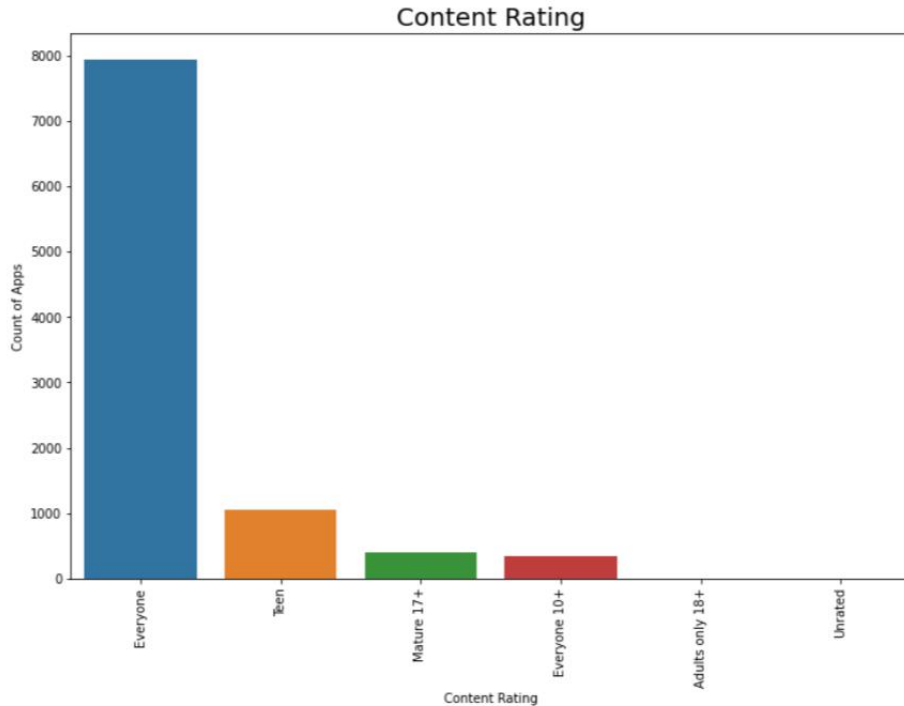
Correlation Between Different Parameters

- ❖ We can see that installs and reviews have the strongest correlation.
- ❖ There is no correlation found between installs and other features.

Size distribution of all the Apps

Content Rating


Content Rating

Content Rating wise Average rating of apps

❖ we can see from the graph that most of the apps are for "everyone" (Approx 7900) so that it could capture the most of the customer-base.
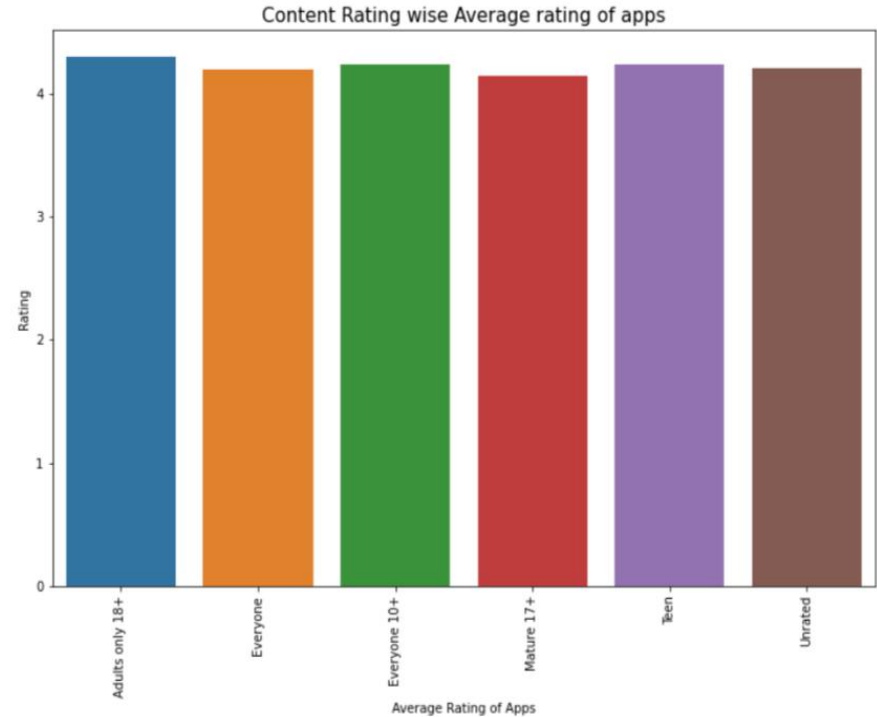
❖ Apps made only for Adults are liked most with an average rating given by them is 4.3

Distribution Sentiment Types

AI

❖ The Sentiment Subjectivity is not always proportional to Sentiment Polarity, but in maximum number of cases, it shows a proportional behavior when variance is too high or too low.
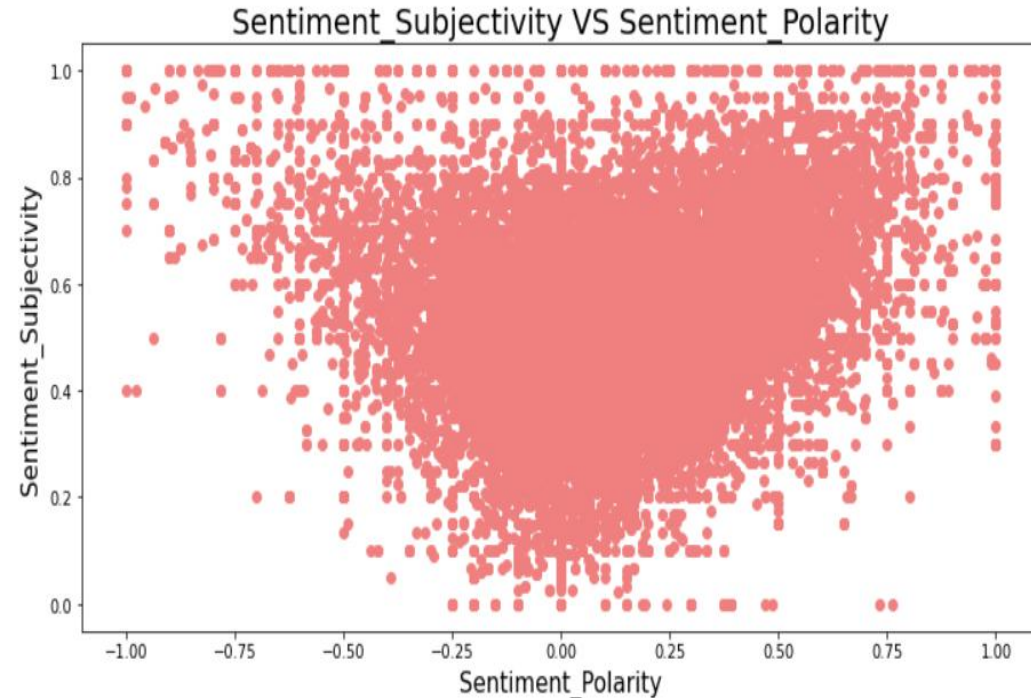
❖ Most of the review sentiments are positive and having review percent of 64.1

Sentiment_Subjectivity VS Sentiment_Polarity

SENTIMENTS OF USERS
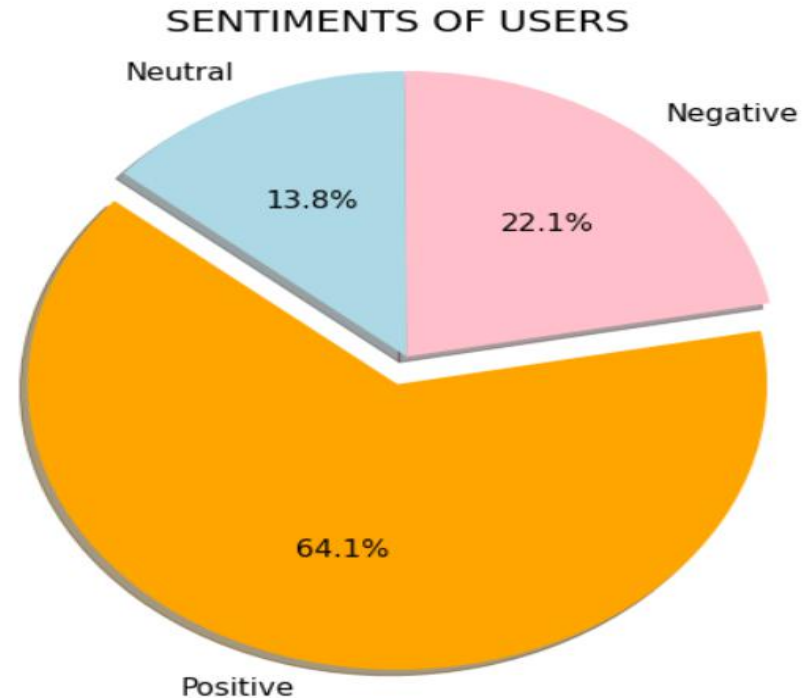
Relation Between Sentiment Polarity And Sentiment Subjectivity

❖ The Sentiment Subjectivity is not always proportional to Sentiment Polarity but in maximum number of case, it shows a proportional behavior when variance is too high or too low.
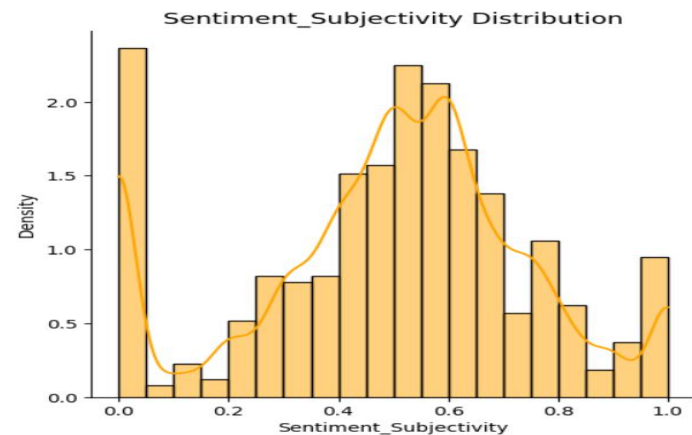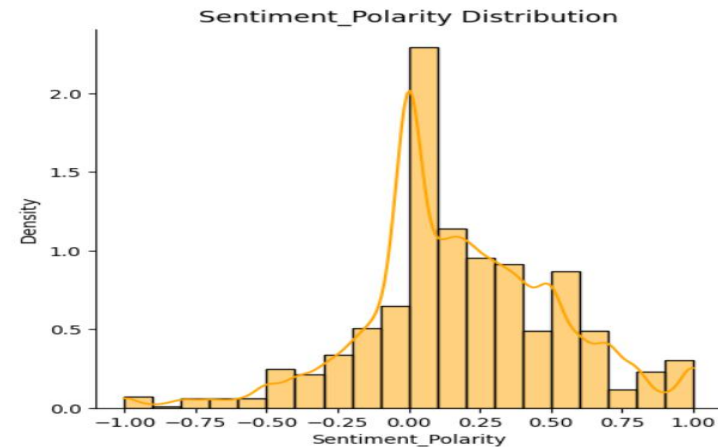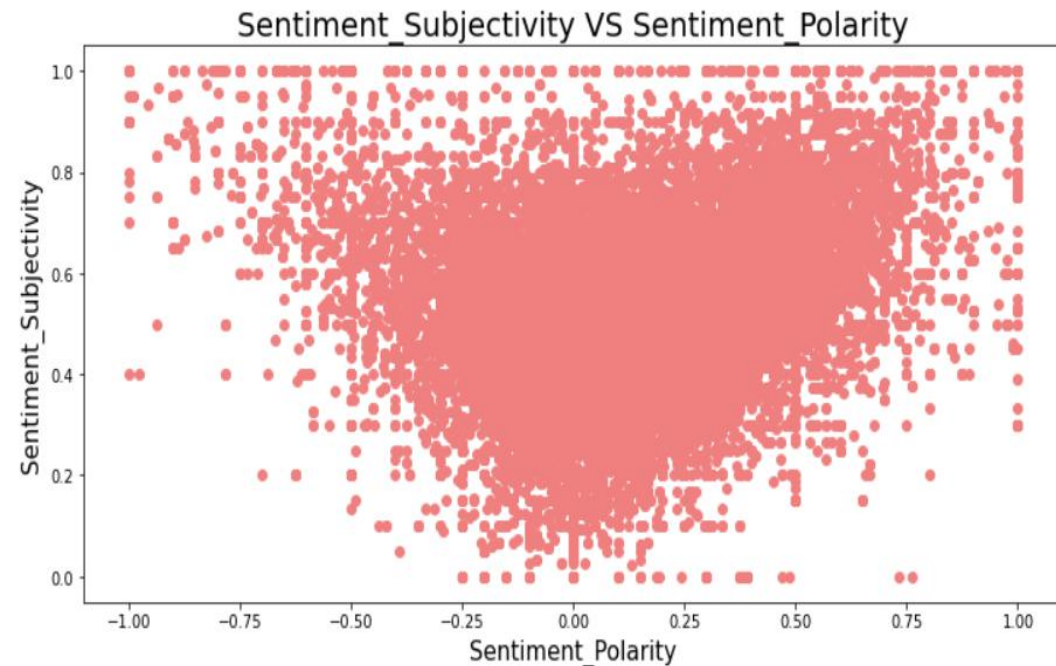
Relation Between Reviews And Sentiment Polarity/ Subjectivity

❖ No definitive relationship between sentiment polarity and number of reviews.
❖ Lot of variance in the values of polarity with respect to the number of reviews.

❖ Variance in the values of subjectivity with respect to the number of reviews is very high.
❖ There is no definitive relationship between sentiment subjectivity and number of reviews.

Sentiment Polarity vs Number of reviews

Sentiment Subjectivity vs Number of reviews
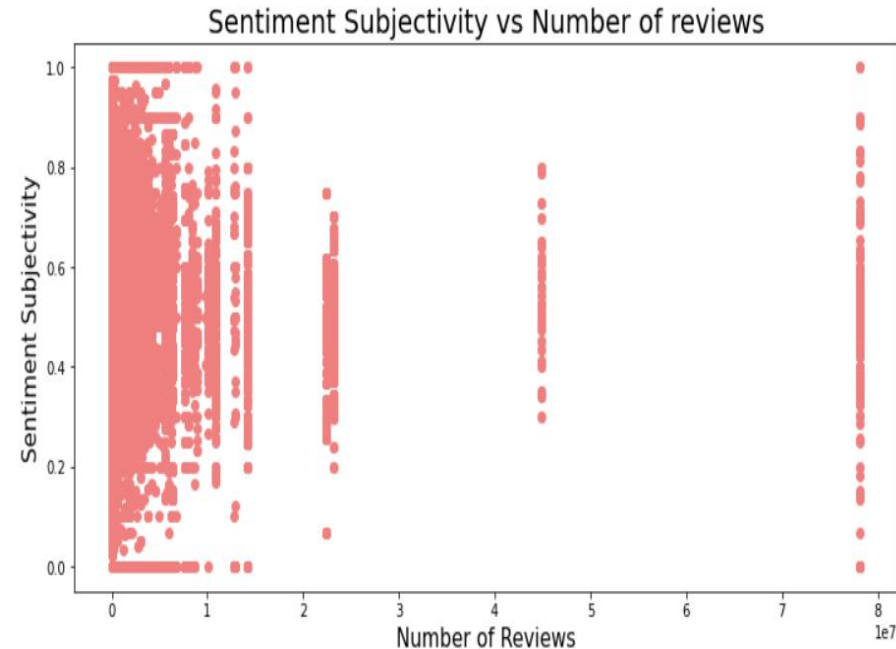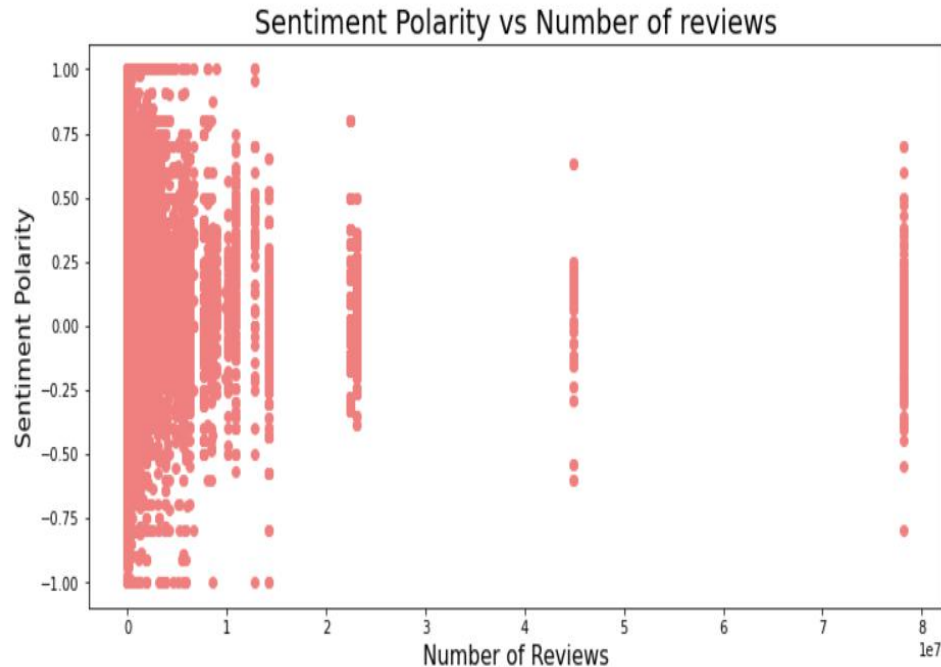
Insights
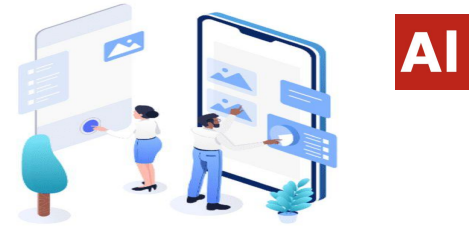
❖ Average number of stars for all apps : **4.1**

❖ Total number of apps with 5 star ratings: **271**

❖ Category which has highest average rating:
Events, Art and Design, Education

❖ Total number of Free and paid Apps
Free 8905   Paid 754

❖ Average rating of Free and paid Apps:
Free 4.1   Paid 4.2

❖ Maximum reviews - **Facebook**

❖ Top Expensive App - **I'm Rich**

Conclusion

❖ **The majority of users prefer free apps. The size of the app does not directly affect their decision to use paid or free apps.**

❖ **The highest number of apps have been published in the Family, Games, and Tools categories on the Play store, but the highest number of installations have come from the Games category.**

❖ **The apps which have the ratings above 4, are actually targeting all the people and not a certain age group.**

❖ **Dating, Maps, Travel and Video Player are the categories which have received the lowest ratings from the user and we know that these apps have high human interaction.**

❖ **On average the application size is in between 10 MB to 30 MB.**

❖ **Family, Game, Communication, Social, Tools, Medical, Events category apps have the highest earning.**

❖ **Medical, Family, game, Tools apps are the expensive apps.**

❖ **This helps us to understand the application market and tells us as to which would be the right category to invest money to make good profits.**

❖ **The sentiment does not influence the final rating of the apps because of the proportional behavior.**

❖ **Popular categories have a more positive sentiment.**

❖ **Positive reviews are higher than negative and neutral but the sentiment subjectivity is not always proportional to sentiment polarity.**