EXPLORATORY DATA ANALYSIS COURSE PROJECT

Play Store Applications Performance Analysis

TEAM NO: 10

Guided By:
Dr. Karibasappa K G
Prof. Sunil G

TEAM DETAILS

NAME	USN	Roll No.
Pranavi Kulkarni	01FE20BCS118	253
Mayuri Kalmat	01FE20BCS095	230
Parag Hegde	01FE20BCS096	231
Pranav Jadhav	01FE20BCS099	234

Content:

Abstract:

Play Store Applications Performance Analysis is a project under the course Exploratory Data Analysis. A sample data set consisting of 10,840 records is taken from over millions of records present in the play store for the analysis of the various trends. We have filtered the data by applying suitable data preprocessing techniques. Later, we have demonstrated the visualizations by various graphs and plots. Various like users age, users' interests, highly downloaded categories and applications have been observed and recorded. At the end, prediction has been demonstrated for predicting the rating and downloads of the applications. This analysis helps the application developers to develop the better applications and the users to choose the best applications.

Introduction:

Google Play Store is home to millions of mobile applications used worldwide. Every day, millions of users download various applications that belong to a wide range of genres from Education to Finance, from beauty to communication, from shopping to sales. Every possible category is covered by the apps.

As there are several applications hosted on the play store for a single genre, it becomes essential to figure out which application among all performs well. Based on the data available such as number of downloads, ratings, reviews and so on, we can analyze the users' interests. This analysis helps the users to download the better applications and also the application developers to improvise and modify according to the user interests.

Problem Statement:

"Analyze the user interests and trends of Google Play Store applications."

Dataset description:

Categorical	Category,Type, Genres, ContentRating	4
Ordinal	Rating, LastUpdated, TimeStamp, date, month, year, CurrentVersion, AndroidVersion	8
Numerical	NoOfReviews,Size,NoOfInstalls,Price, UpdateFrequency	5
Binary	Advertisement, BetaVersion	2

Attributes	Description
Category	Denotes the category to which the app belongs.
Type	Denotes if the application is free to use or the paid one.
Genres	Denotes the genre to which the application belongs.

Content Rating	Denotes the group of users suitable to use the respective application.

Attributes	**Description**
Rating	Denotes the category to which the app belongs.
LastUpdated	Denotes the date on which the application was updated recently.
Timestamp	Denotes the time at which the data was recorded for analysis.
Attributes	**Description**
Date, Month, Year	The date on which the data was recorded for analysis.
CurrentVersion	Denotes the current version of application.
Android Version	Denotes the minimum version supported by the application.

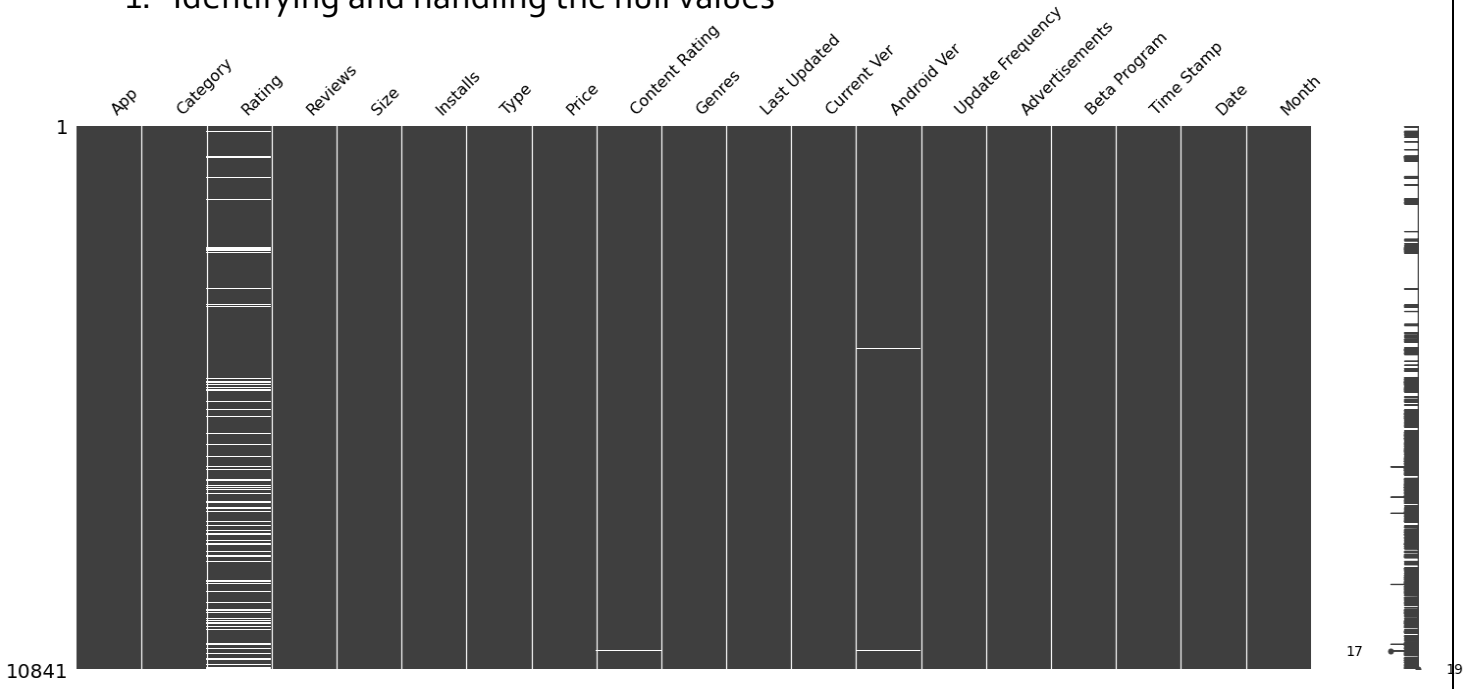
Attributes	**Description**

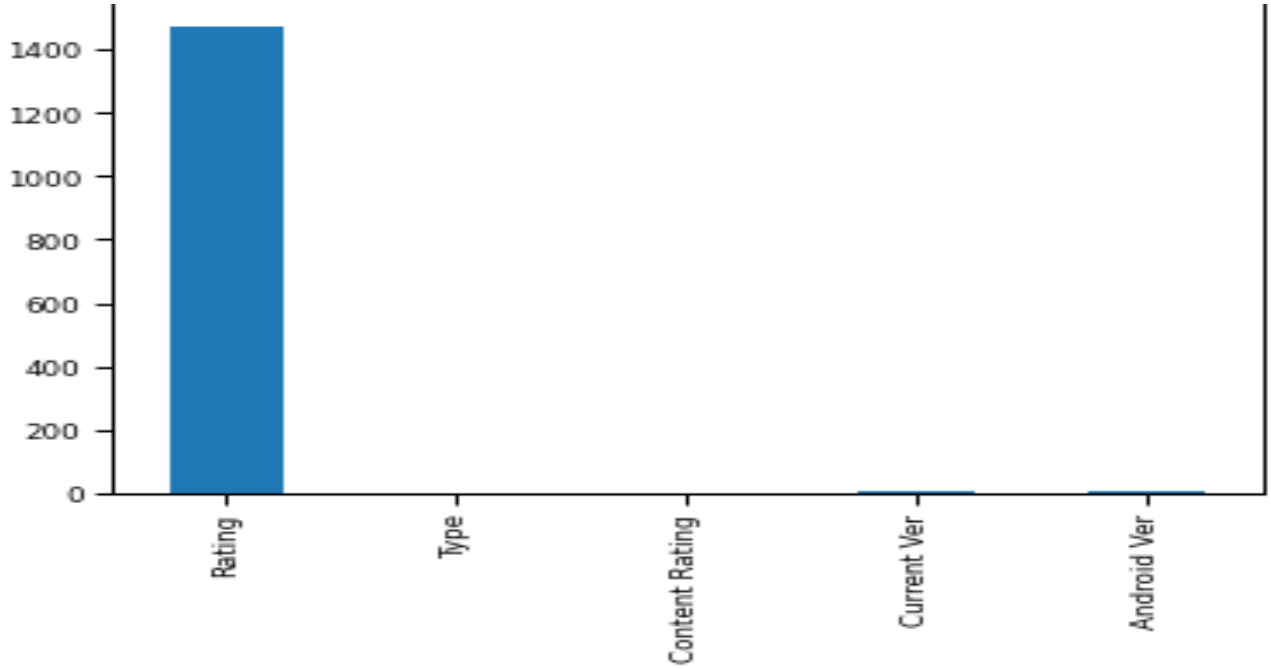
NoOfReviews	Denotes the count of reviews by the users to the application.
Size	Denotes the download size of the application.
NoOfInstalls	Denotes the number of downloads of the application throughout the globe.
UpdateFrequency	Denotes the average time gap between two successive updates to the application.

Attributes	**Description**
Advertisement	Denotes if the application contains advertisements.
Beta Version	Denotes if the beta version is available for the application.

Data pre-processing:

1. Identifying and handling the null values

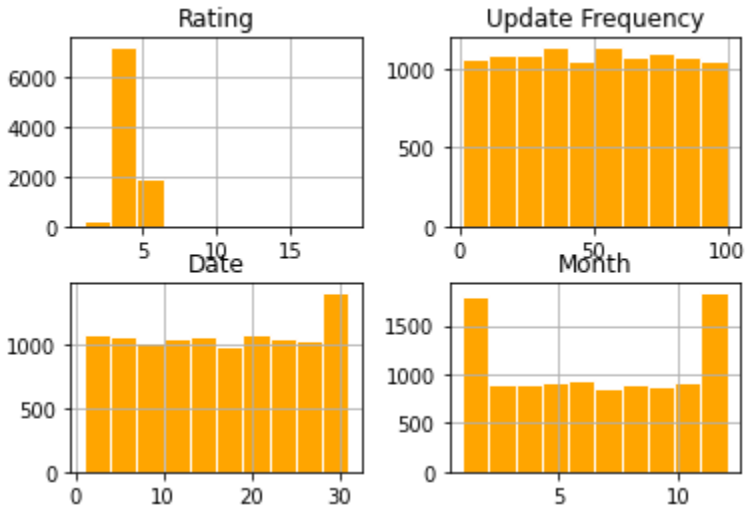




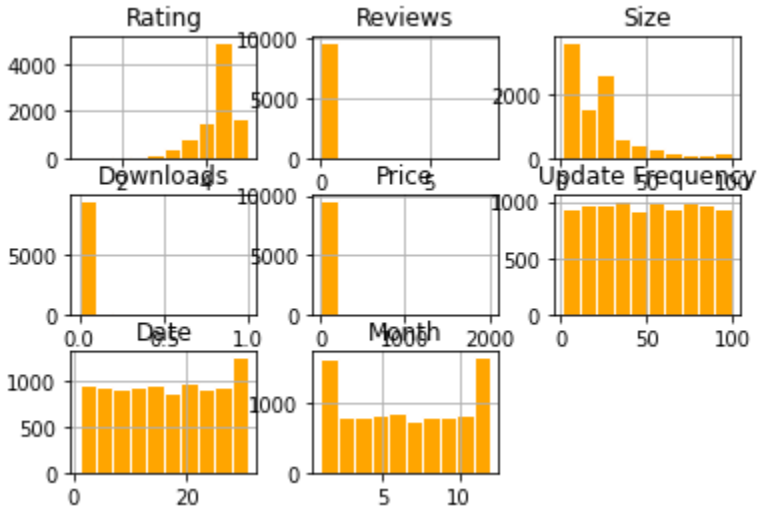
From the bar graph of Null Values, it is evident that attribute 'rating' has the highest number of null values. This also indicates that users are less active in giving the ratings to the applications they use.

The Impact of Preprocessing

Before preprocessing



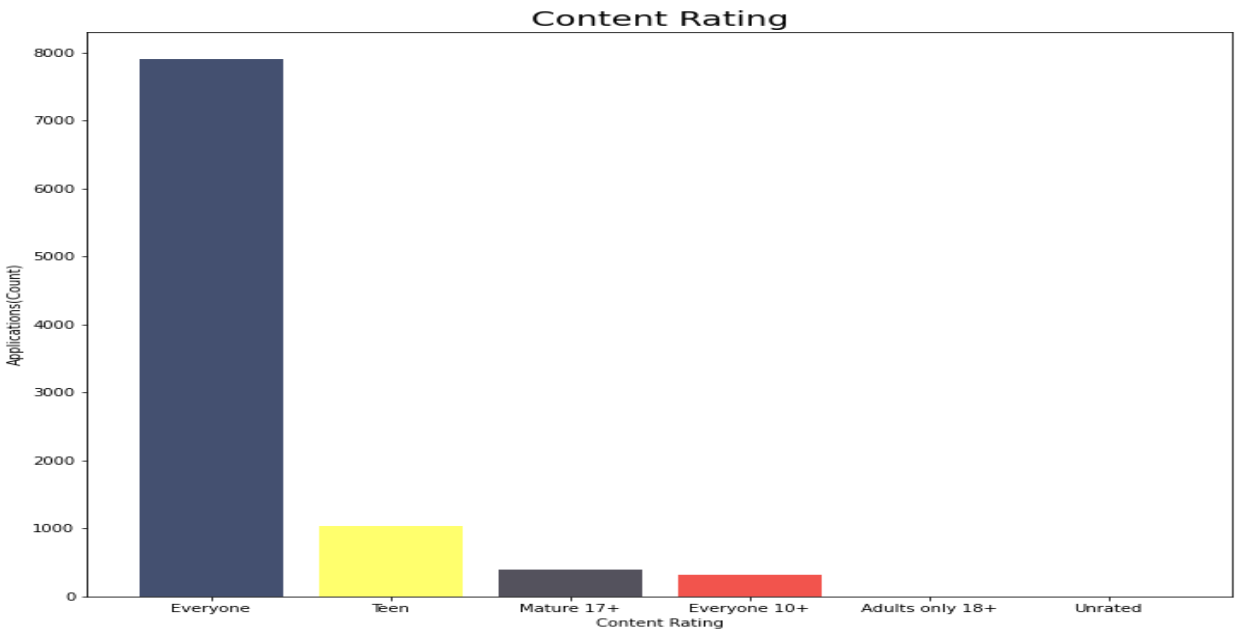
After Preprocessing



Inference: After the preprocessing, since the null values are removed, there is increase in the number of numeric attributes.

Exploratory Data Analysis

1.Availability based on categories

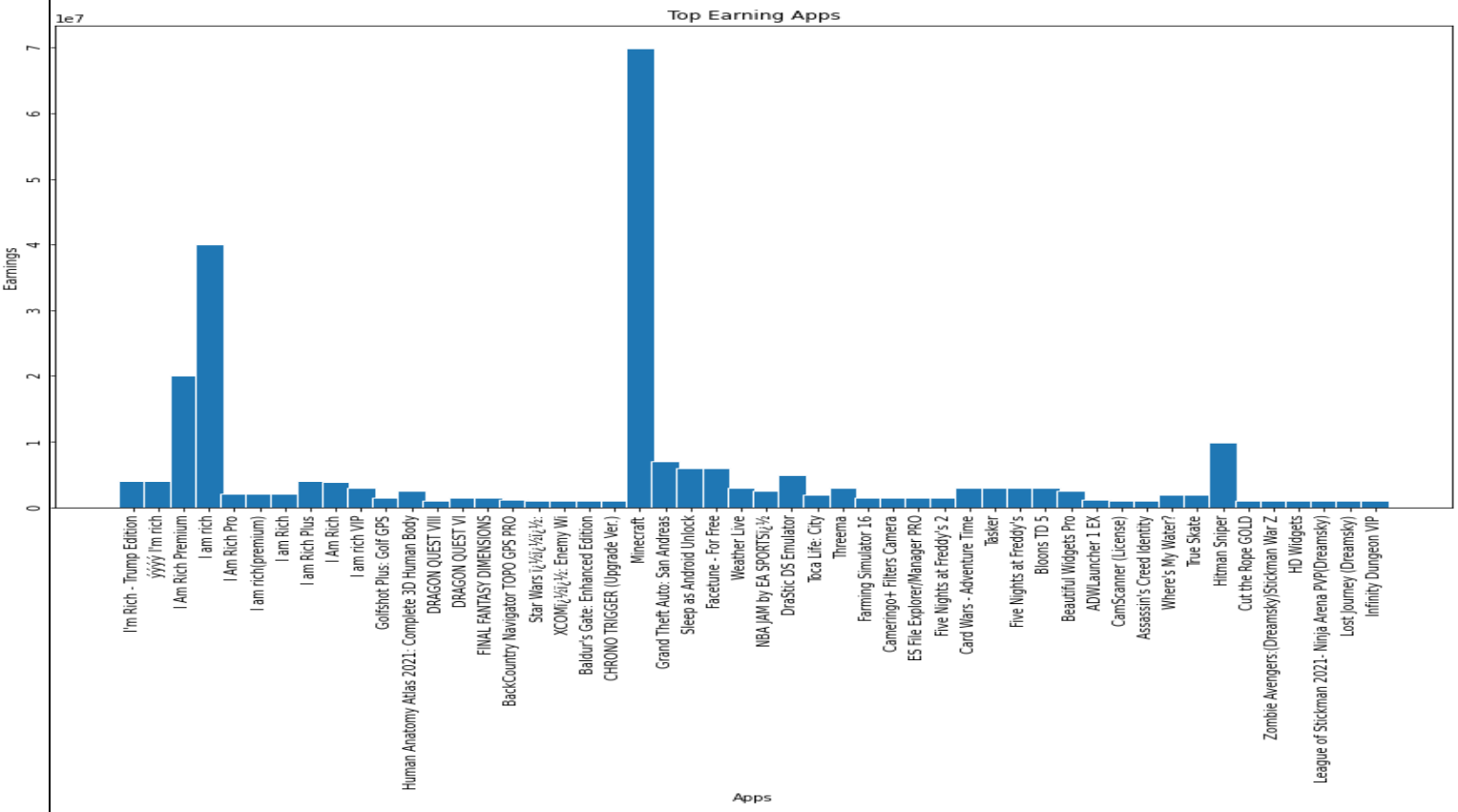


From the bar graph, it is evident that the applications which are available for everyone are the highest downloaded ones.

Followed by them, there are the apps suitable for teens, mature 17+ etc.,

This shows that most of the users download the applications that are made to be used and available for everyone.

2.The Highest Earning Applications

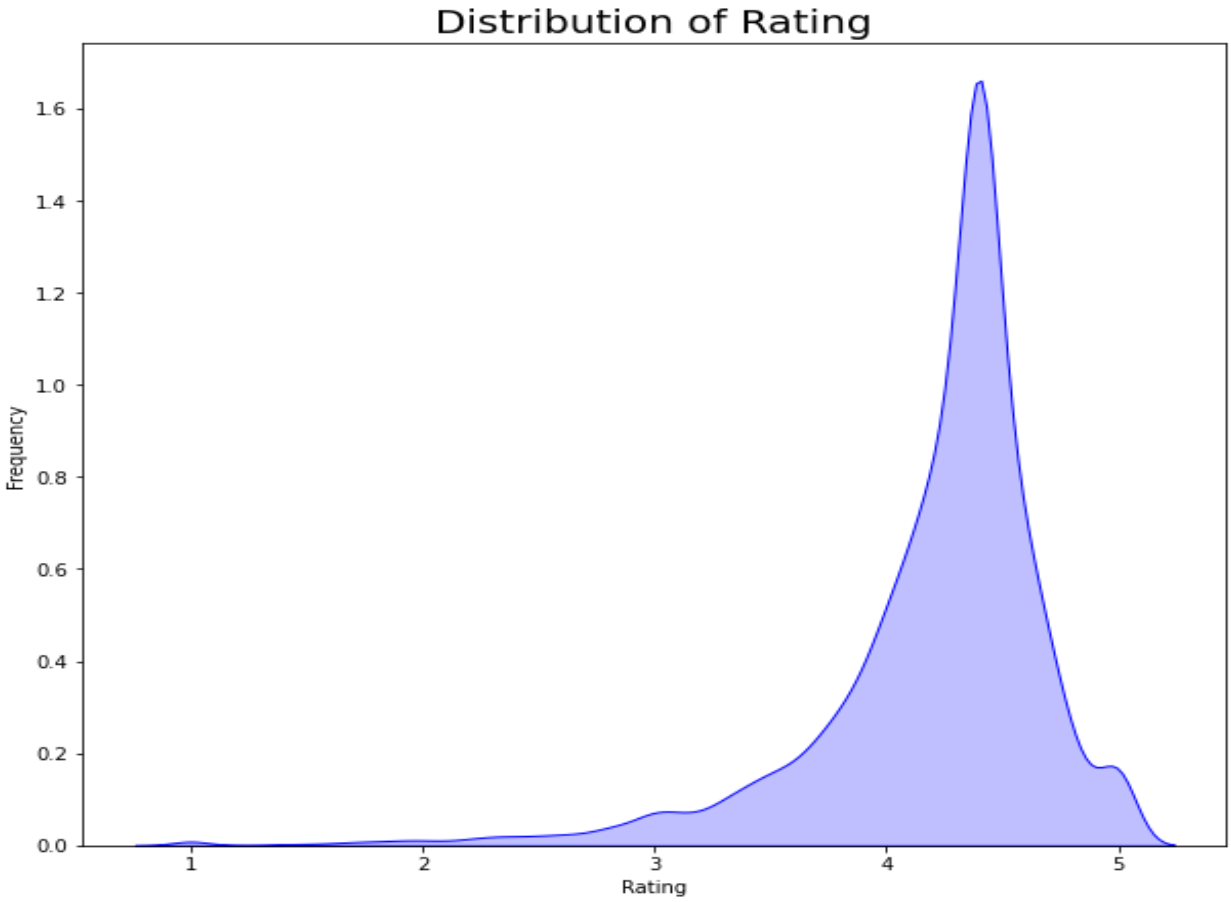


When a bar graph was plotted between the paid apps and their earnings, it was then inferred that:

the Minecraft is the highest earning app.

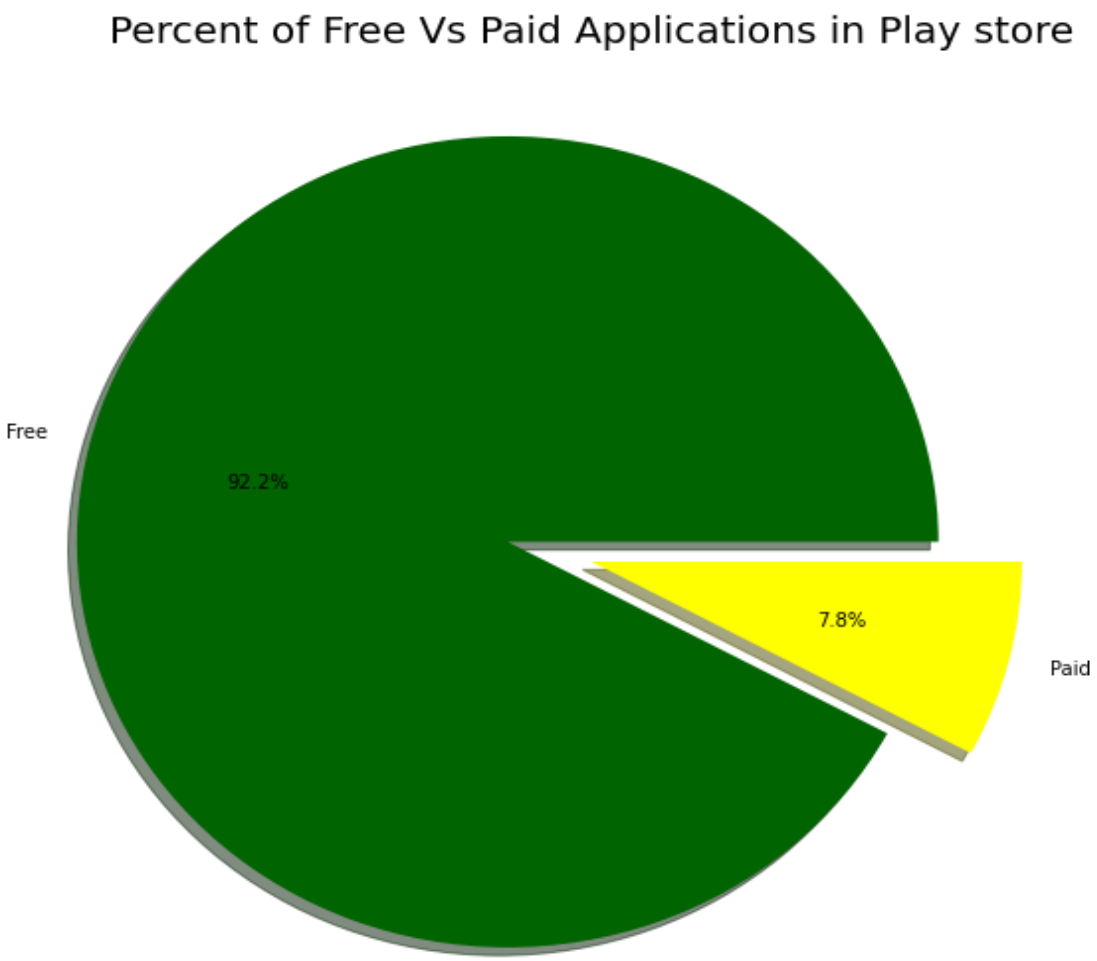
It is followed by 'I am Rich', 'I Am Rich Premium', 'Hitman Sniper', and 'Grand Theft Auto: San Andreas'.

3.The Distribution of Ratings



From the graph, it is inferred that majority of the applications are rated between 4.0 and 4.5. The applications with ratings below 3.5 are the too less in number.

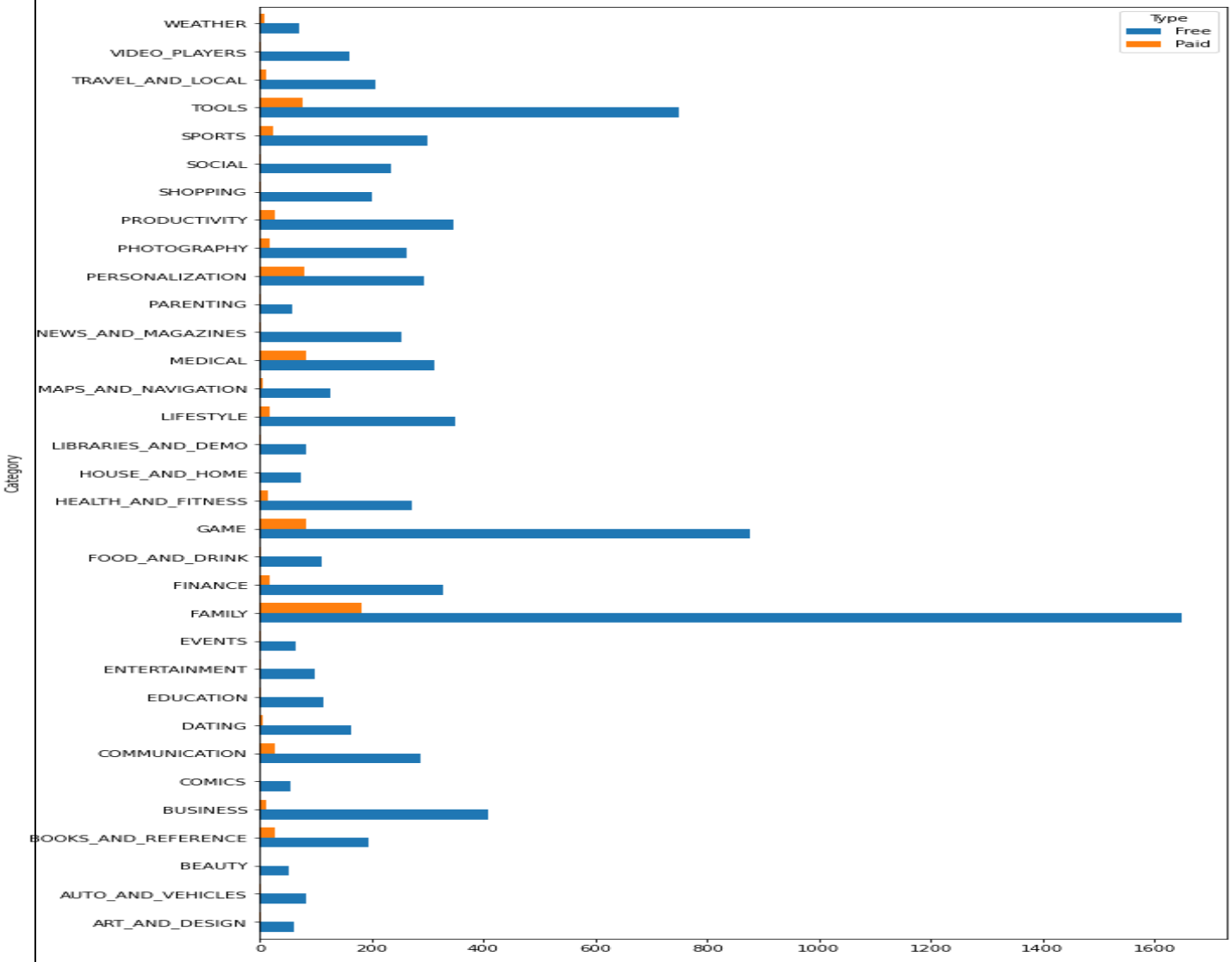
4.Distribution Of paid and free applications in Play Store



From the pie chart, it is clearly evident that the google play store is dominated by the free applications.

Of the total applications, 92.2% applications are free to download and use. Rest 7.8% applications are the paid applications.

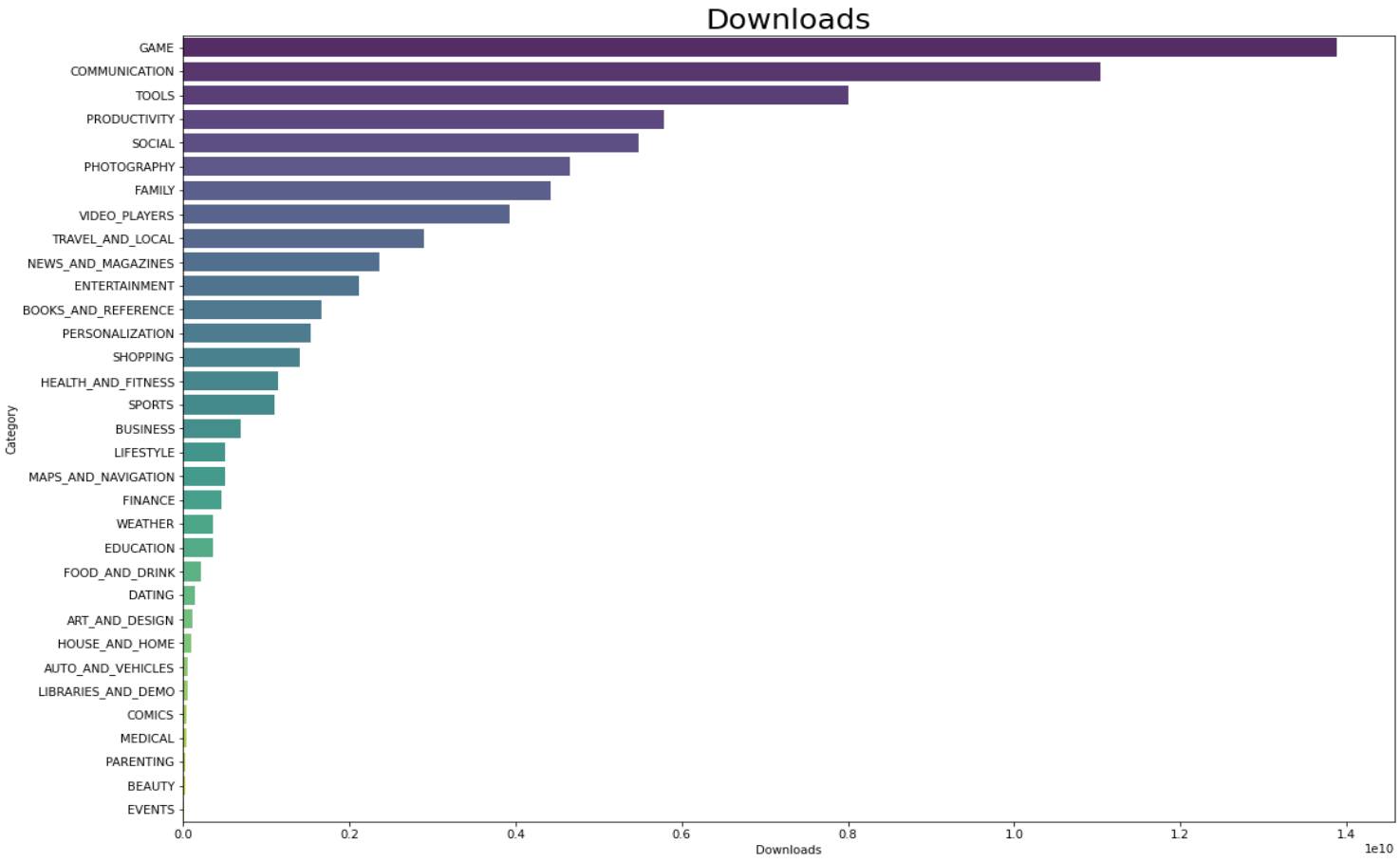
5.Number of applications that are free and paid in each category



'Family' is the category with the highest number of paid applications. It is also the category with the highest number of free applications.

The categories 'Game' and 'Tools' have almost same number of free applications.

6. To know which category has the highest number of downloads in Play Store



When it comes to the number of downloads, the category **'Game'** acquires the first position and can be called as **the highest downloaded category**.

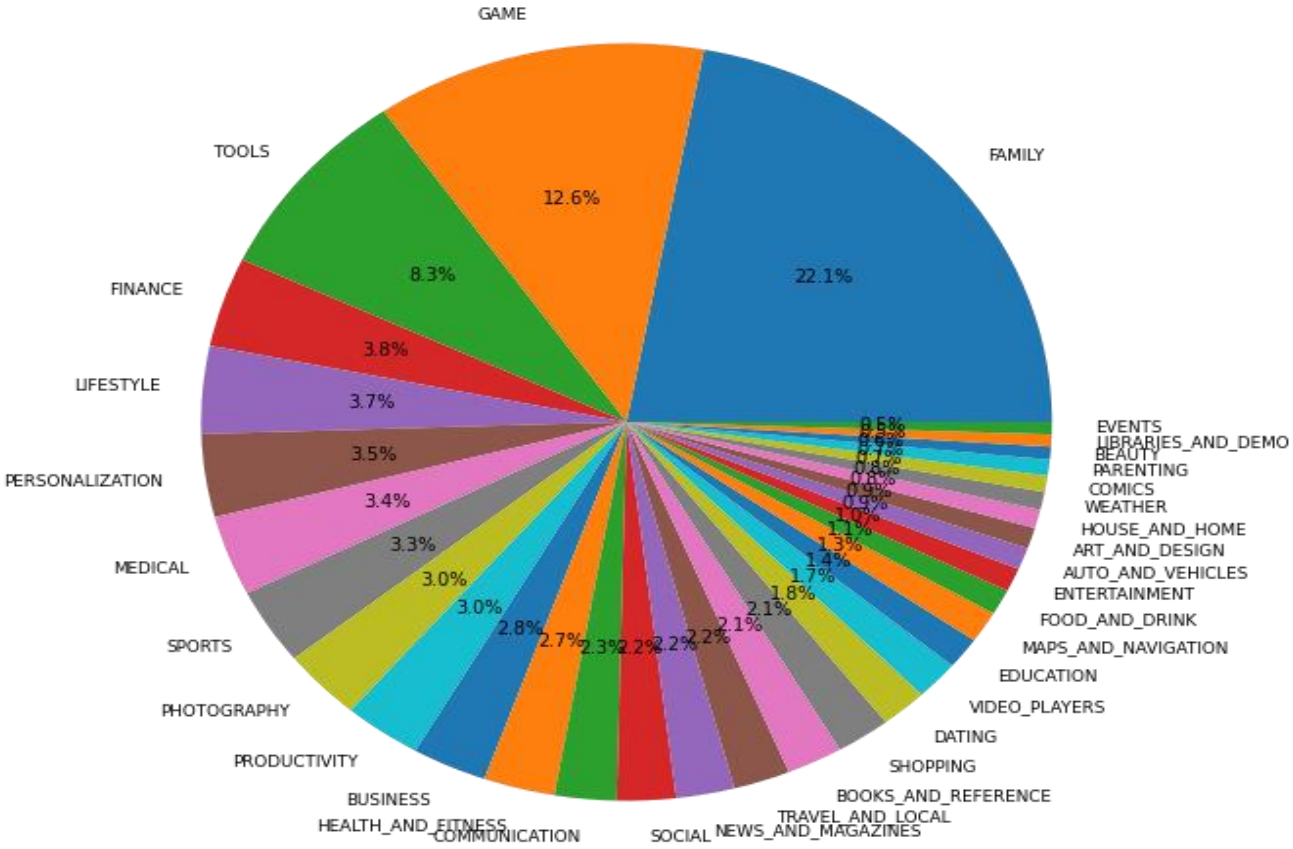
It is then followed by Communication, Tools, Productivity, and Social.

This shows that majority of users are interested in Gaming category.

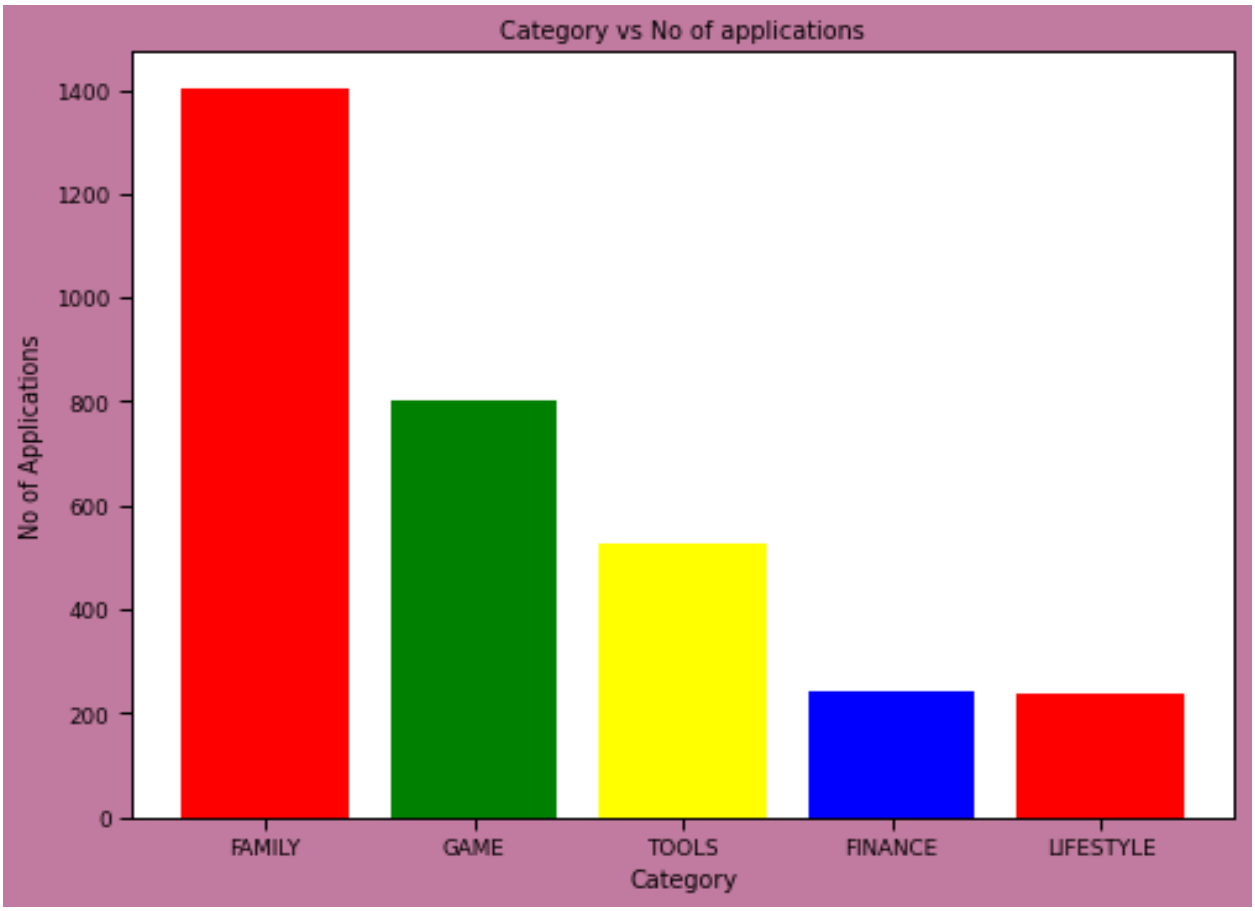
It's more likely that users might range between **teenagers to mid 20s**.

Very **minimum** downloads has been observed from the category **Events**.

7. To know relation between Category, count of application and their highest rating



This chart displays all the categories, partitioned according to the proportions of their quantity.



The bar graph of Category vs No of applications reveals that

the applications under the category 'Family' are downloaded the most. This category is followed by 'Game', 'Tools', 'Medical', and 'Finance'. Together, these five categories represent the top five most downloaded categories.

FAMILY

App Names	Ratings
Fr. Mike Schmitz Audio Teachings	5.000000
Safe Santa Fe	5.000000
BM SPM Practice	5.000000
WPBS-DT	5.000000
COMSATS BOOK STORE FOR BS(CS)	5.000000

GAME

App Names	Ratings
Axe Champs! Wars	5.000000
Mad Dash Fo' Cash	5.000000
Santa's Monster Shootout DX	5.000000
211:CK	5.000000
Ra Ga Ba	5.000000

TOOLS

App Names	Ratings
CZ Kompas	5.000000
CE Smart	5.000000
CR Tracker for Chests	5.000000
CL REPL	5.000000
BM speed test	5.000000

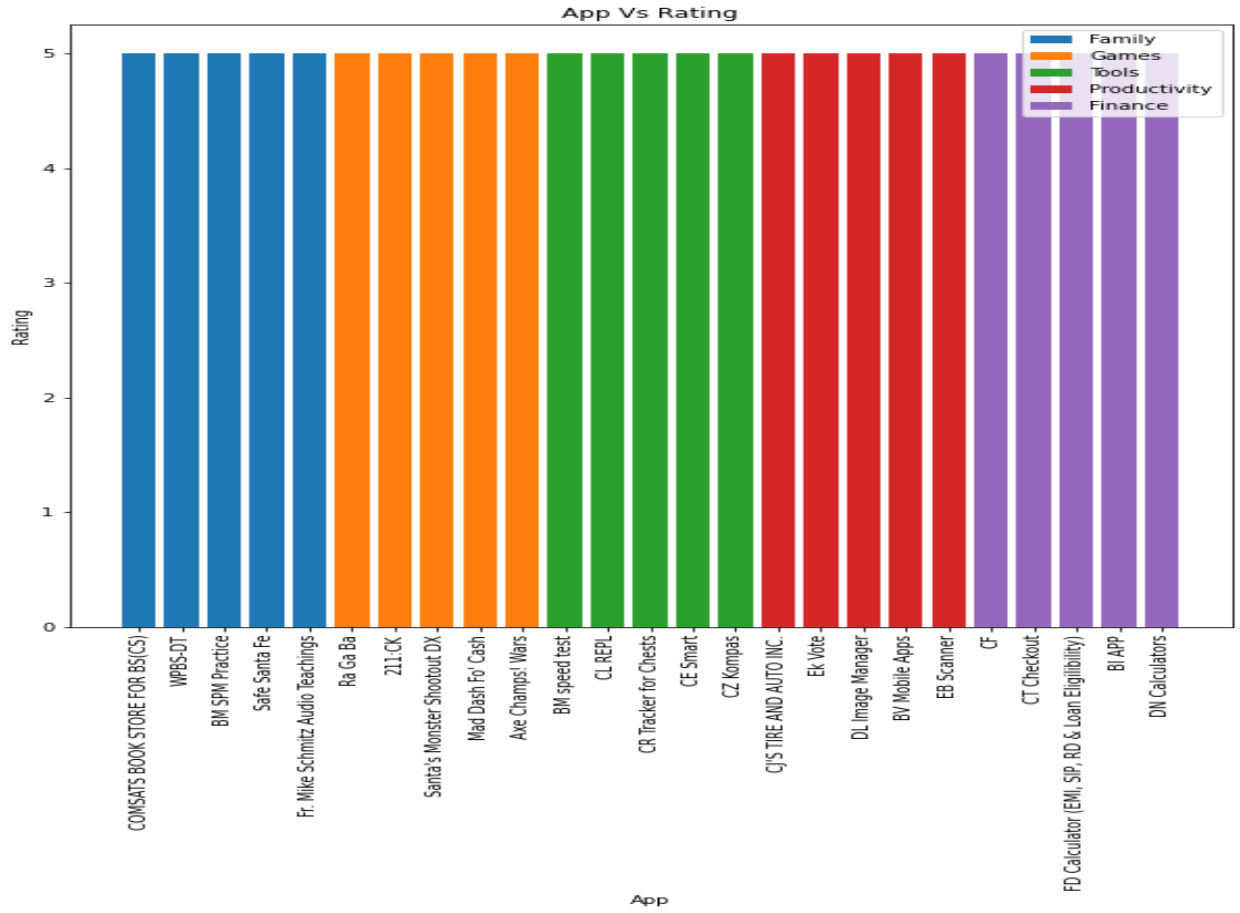
FINANCE

App Names	Ratings
EB Scanner	5.000000
BV Mobile Apps	5.000000
DL Image Manager	5.000000
Ek Vote	5.000000
CJ'S TIRE AND AUTO INC.	5.000000

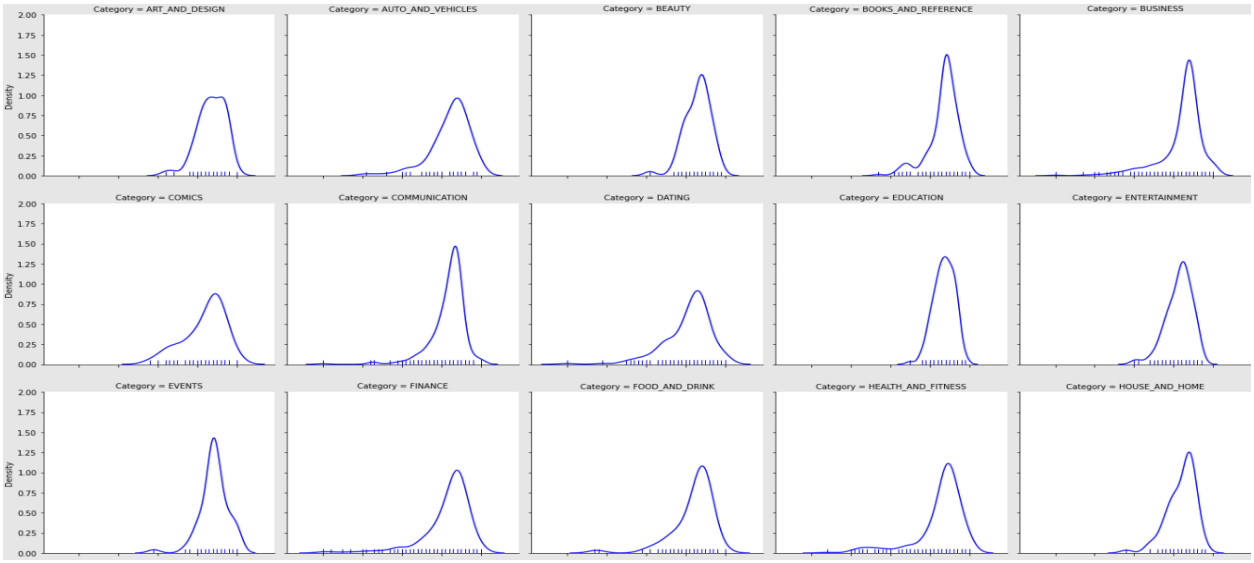
LIFESTYLE

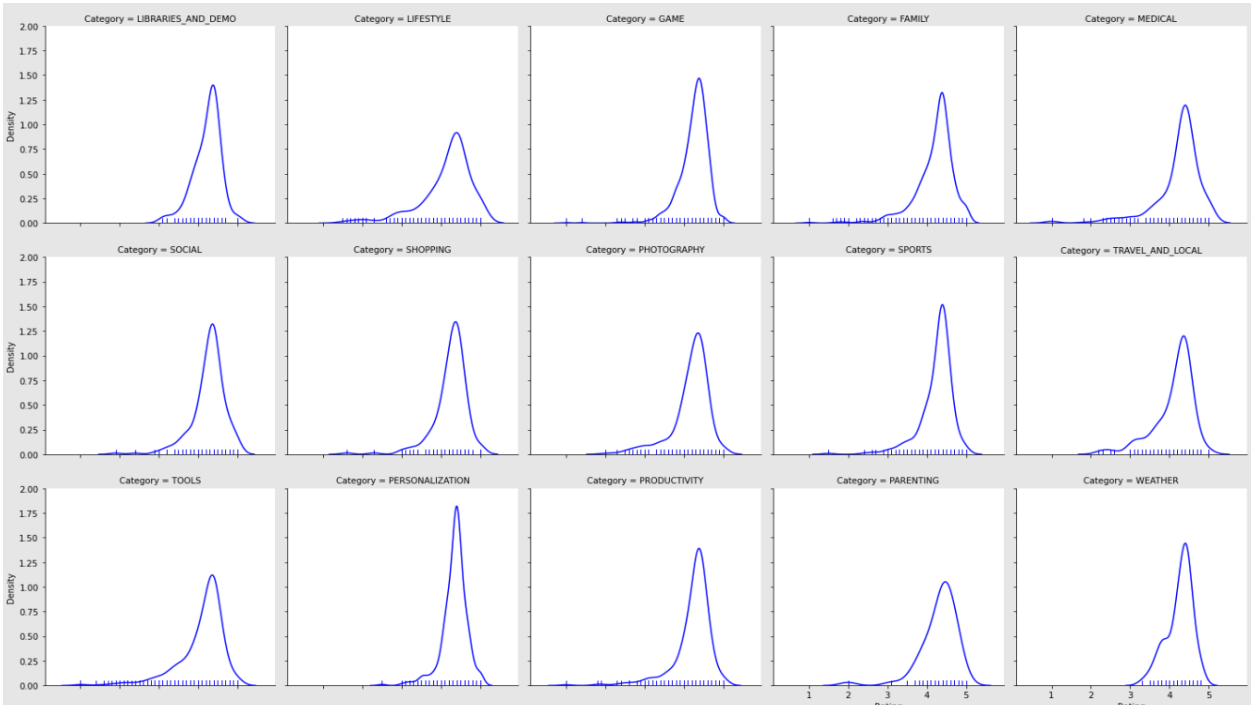
App Names	Ratings
DN Calculators	5.000000
BI APP	5.000000
FD Calculator (EMI, SIP, RD & Loan Eligibility)	5.000000
CT Checkout	5.000000
CF	5.000000

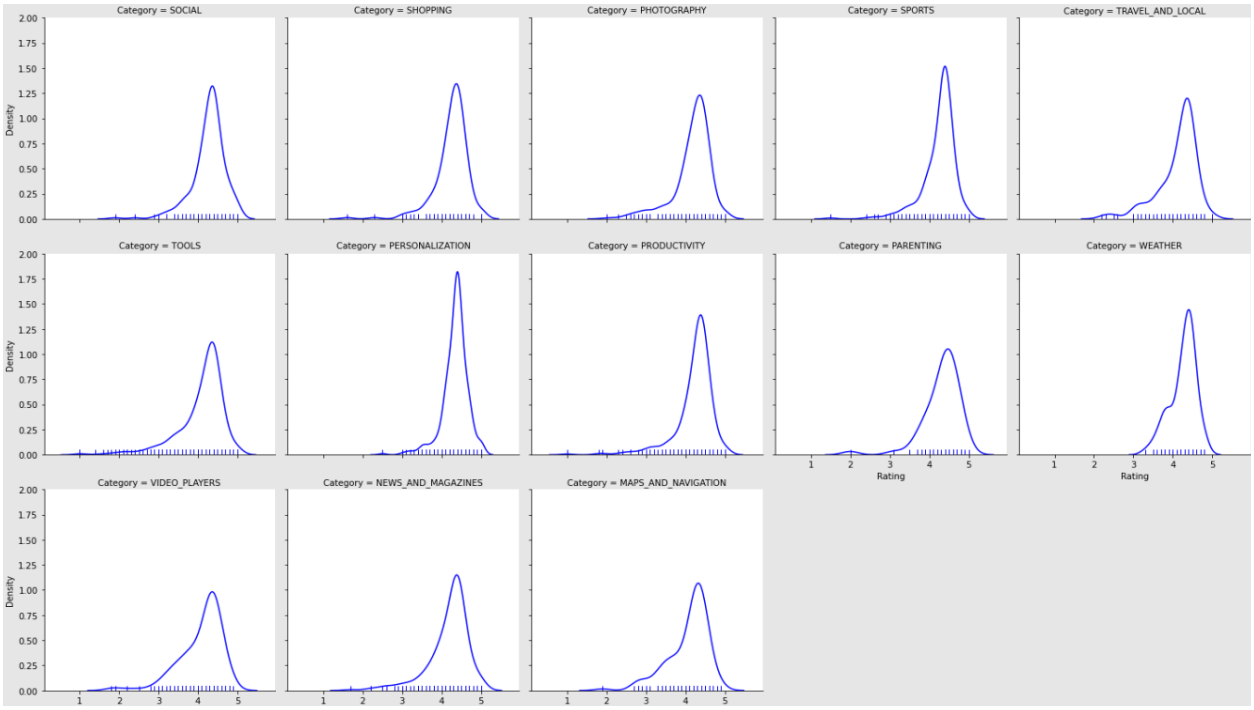
RATINGS OF TOP FIVE APPLICATIONS OF FIVE CATEGORIES WITH HIGHEST RATINGS.



8. Category wise distribution of Rating

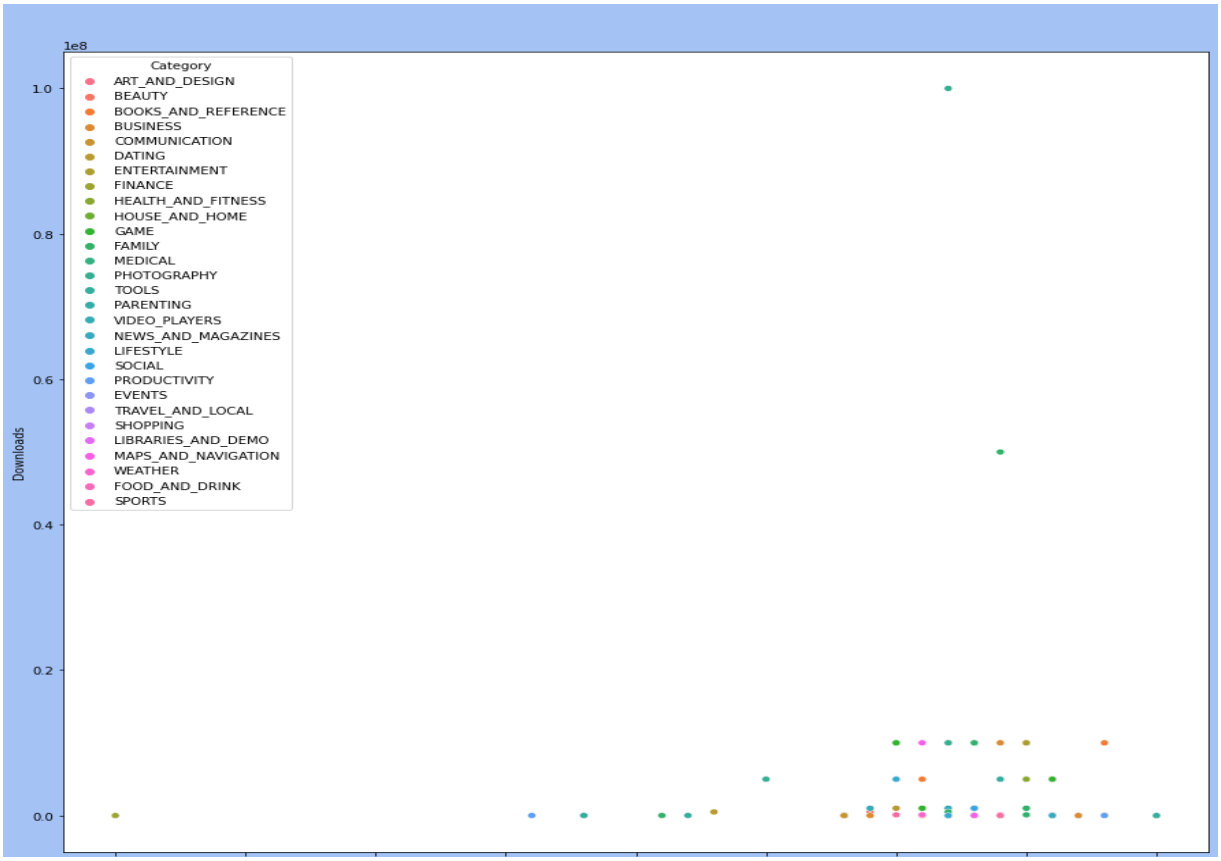






Everyday updated apps give optimization and better applications hence its ratings are high.

9. To know the rating of the application when update frequency is 1(everyday update).



Everyday updated apps gives optimisation and better applications hence its ratings are high

Predictions:

1. For the given number of reviews, size of the applications, and number of downloads predict the Rating of the application

- For given number of reviews=**100000,** size of the applications=20, number of downloads=1000000

  The predicted rating is 4.13

2. To predict the number of downloads of an application for the given rating, reviews and size

- For the given rating=5, reviews=1000 and size=100000

  The predicted number of downloads is 1.5031*10^9