KLE Technological University

Creating Value
Leveraging Knowledge

Exploratory Data Analysis

Play Store Applications Performance Analysis

School of Computer Science and Engineering

2021-22

Table of Contents

List of Figures

List of Tables

1. Course and Team Details

1.1 Course details

Course Name	Exploratory Data Analysis
Course Code	21ECSC210
Semester	IV
Division	B
Year	2021-22
Guide	Dr. Karibasappa K G Prof. Sunil V G

1.2 Team Details

NAME	USN
Mayuri Kalmat	01FE20BCS095
Parag Hegde	01FE20BCS096
Pranav Jadhav	01FE20BCS099
Pranavi Kulkarni	01FE20BCS118

2. 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.

3. 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.

4. Problem Statement

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

5. Dataset Description: talk about tables

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

Table 1 : Dataset Description

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.

Table 1.1 : Attributes

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

Table 1.2 : Attributes and their description

6. Data Pre-processing :

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format.

6.1 Objectives of Data Preprocessing :

I. Data Cleaning and conversion

- First, the null values were checked.
- 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.
- Handling null values present in different columns. By dropping the entries, entering the mode value of the column.
- To remove unwanted values like: '+', 'M', '$' from 'Installs', 'Size','Pricecolumns respectively in order to carry out the numerical analysis.
- Conversion of data in 'Installs' , 'Size', 'Paid' column to numeric, which they were not previously.
- Further handle the null values in the converted column by entering the mean value into it

II. Data Integration and Data Reduction

- Integration Of Data: We have manually integrated the data collected from Google Play Store and Kaggle.com
- Elimination: Because we have two columns i.e Category and Genre which almost have the same content.So, we can consider eliminating Genre to reduce redundancy.
- Dropping Of Columns: Through brief runthrough through the csv, we realised that no analysis can be made using the column Beta version, Current version, Last Updated,Android version columns. Hence, we drop these columns.

III. Data transformation/ Making predictions

- Here we are considering making predictions

6.2 Impact of Preprocessing :

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format.

Before Pre-processing:

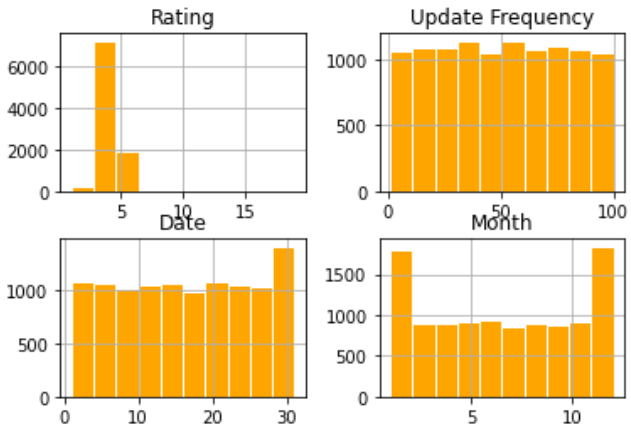

Figure 1.1: Trends before Preprocessing

After Pre-processing:

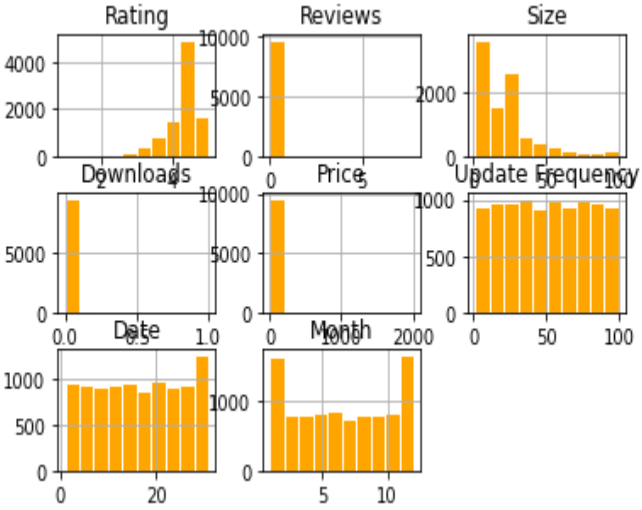

Figure 1.2 : Trends after Preprocessing

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

7. Exploratory Data Analysis :

7.1.To know the count of applications as per their user type.

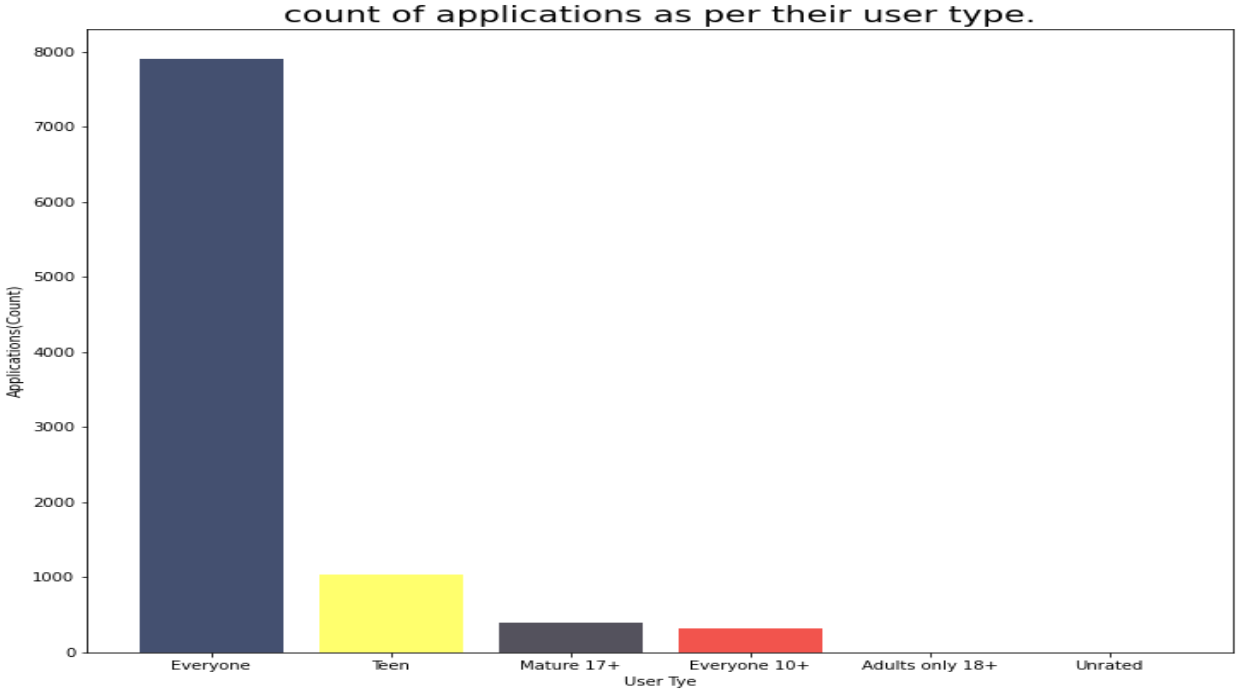

Figure 2 : Count of applications as per their types.

The applications that are under everyone's use include music, social media, navigation etc. They are widely used by all and sundry everyday. Hence, there are a greater number of applications under the user type 'everyone'.

7.2. To know the Application with the highest earning :

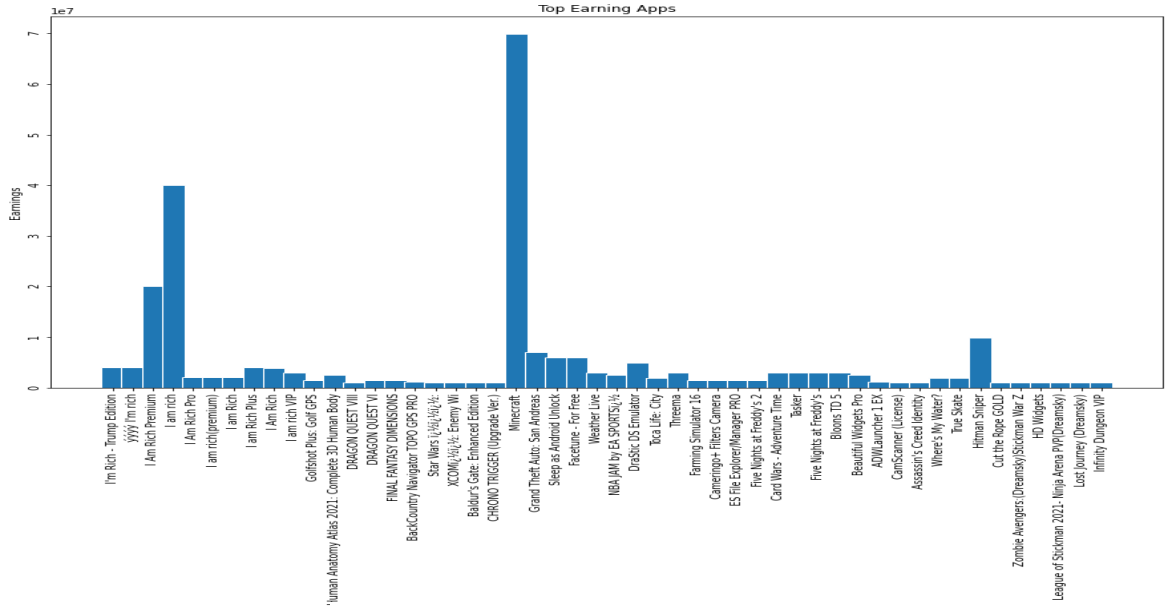

Figure 3 : Application with the highest rating.

Of all the paid applications, Minecraft is the highest earning application because of its versatility, creativity and quirky graphics with which it attracts more users.

7.3. To know the Range in which most application Reviews lie :

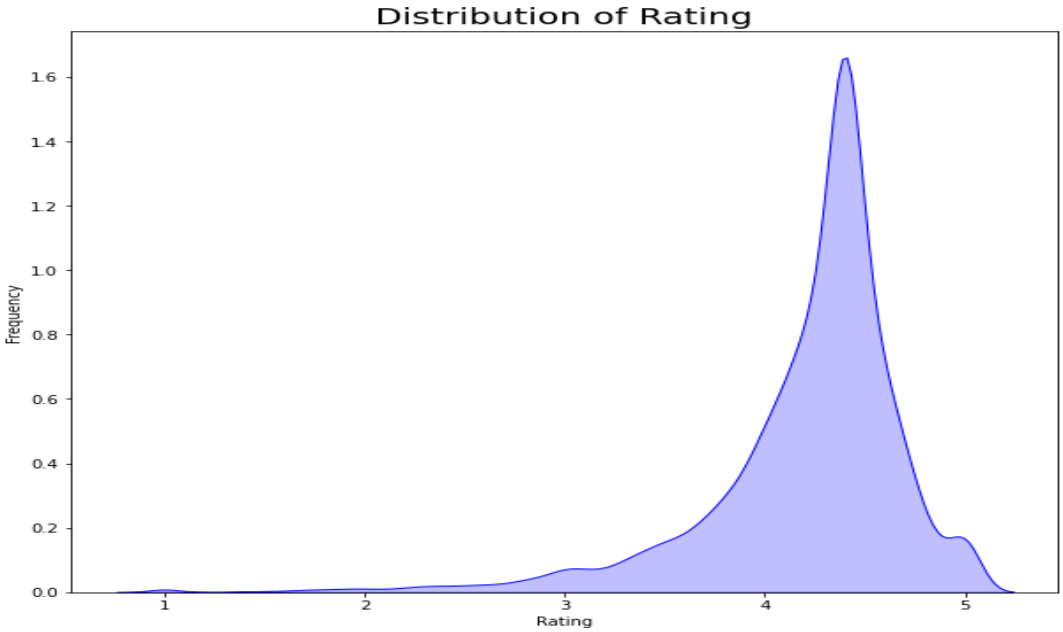

Figure 4 : Distribution of Rating

From the graph, it is inferred that majority of the applications are rated between 4.0 and 4.5. This indicates that majority of applications are user satisfactory. The applications with ratings below 3.5 are not upto to the users' expectations.

-

7.4. To know the distribution of paid and free applications in Play Store :

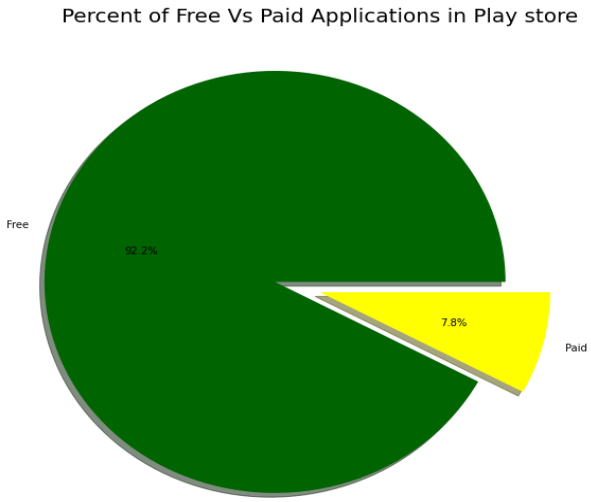

Figure 5 : Percentage of the Free vs Paid applications in Play Store

This chart tells us that majority of users prefer free applications over paid applications as they do not wish to spend money to access these applications. Moreover, most of the paid

applications are expensive due to which many of the users cannot afford them. Hence, they are lesser in count than that of free applications.

7.5. To know number of applications that are free and paid in each category :

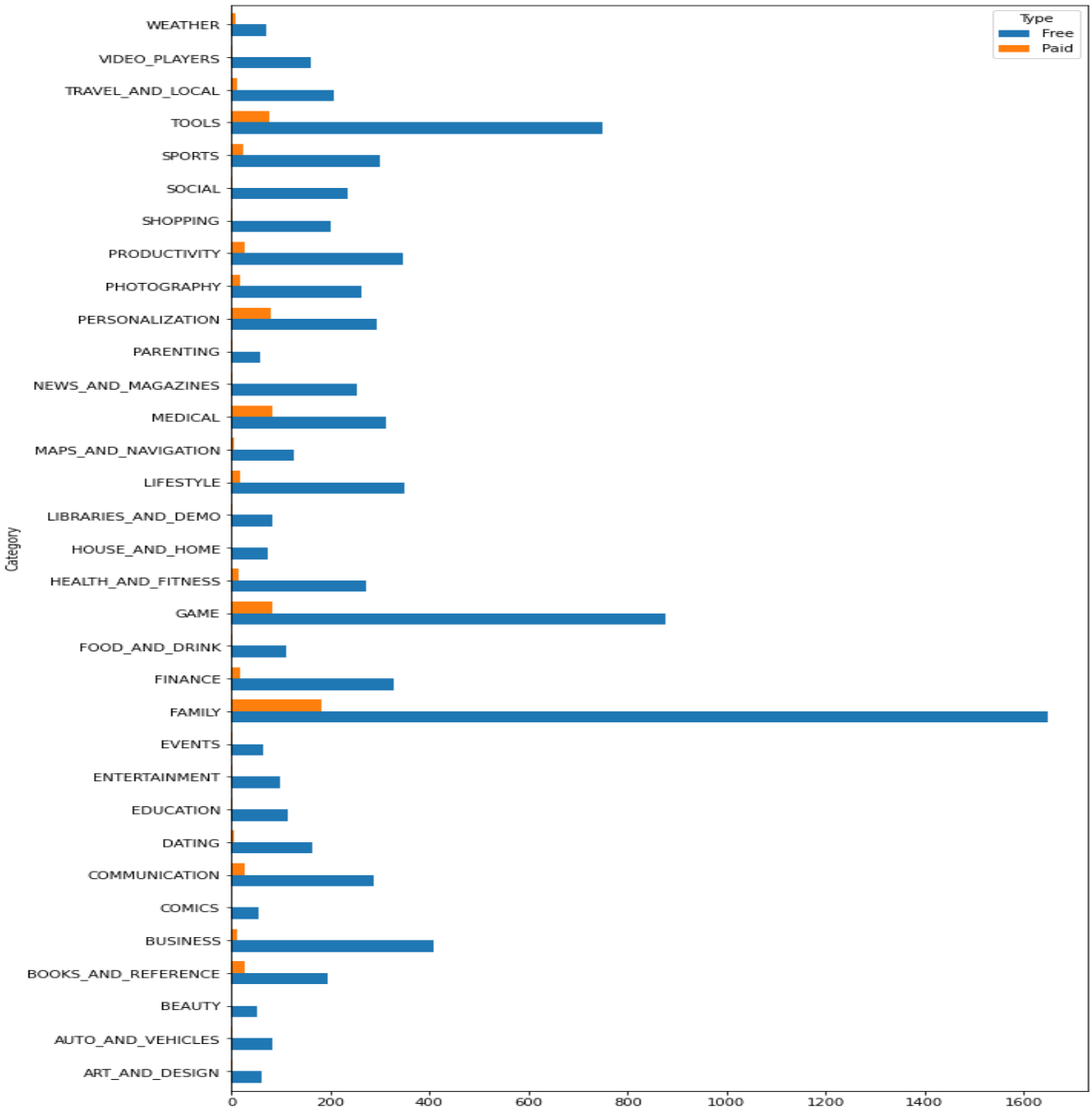

Figure 6 : Free and paid applications in each category

With 183 paid applications, Family is the category with the highest number of paid applications. With 1647 free applications It is also the category with the highest number of free applications. With 876 and 749 free applications respectively, the Game and Tools stand next to the Family with respect to number of free applications.

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

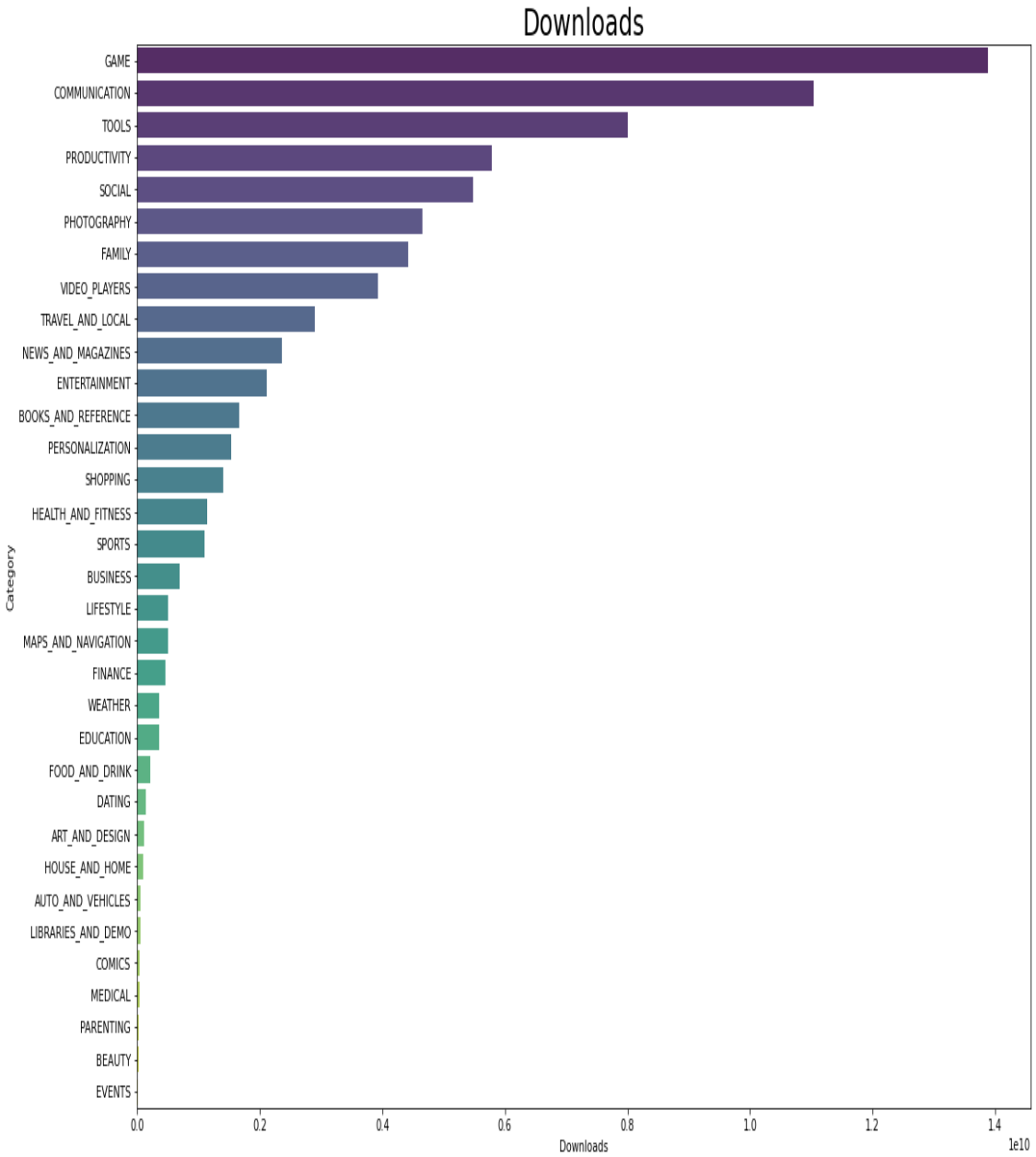

Figure 7 : Category with their downloads

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.

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

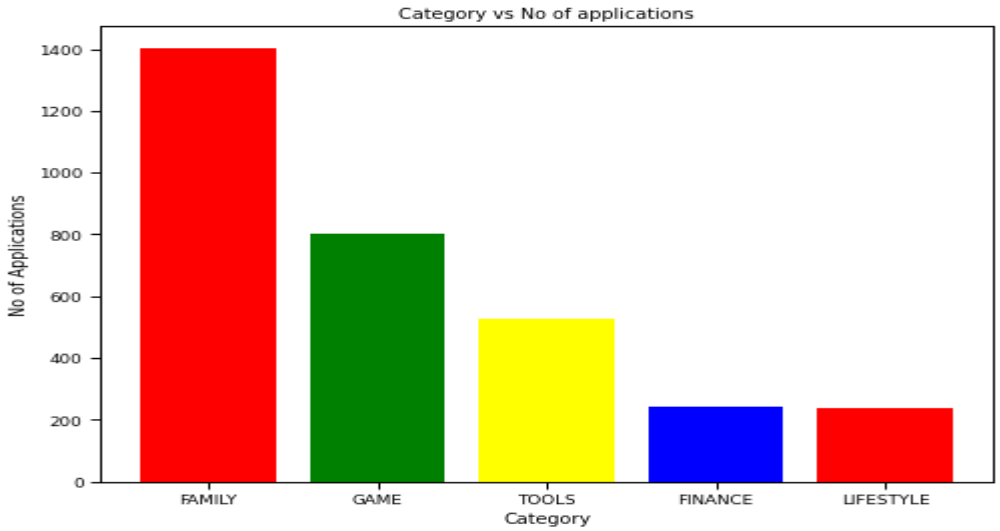

Figure 8.1 : Categories Vs No. of Applications (Bar Graph)

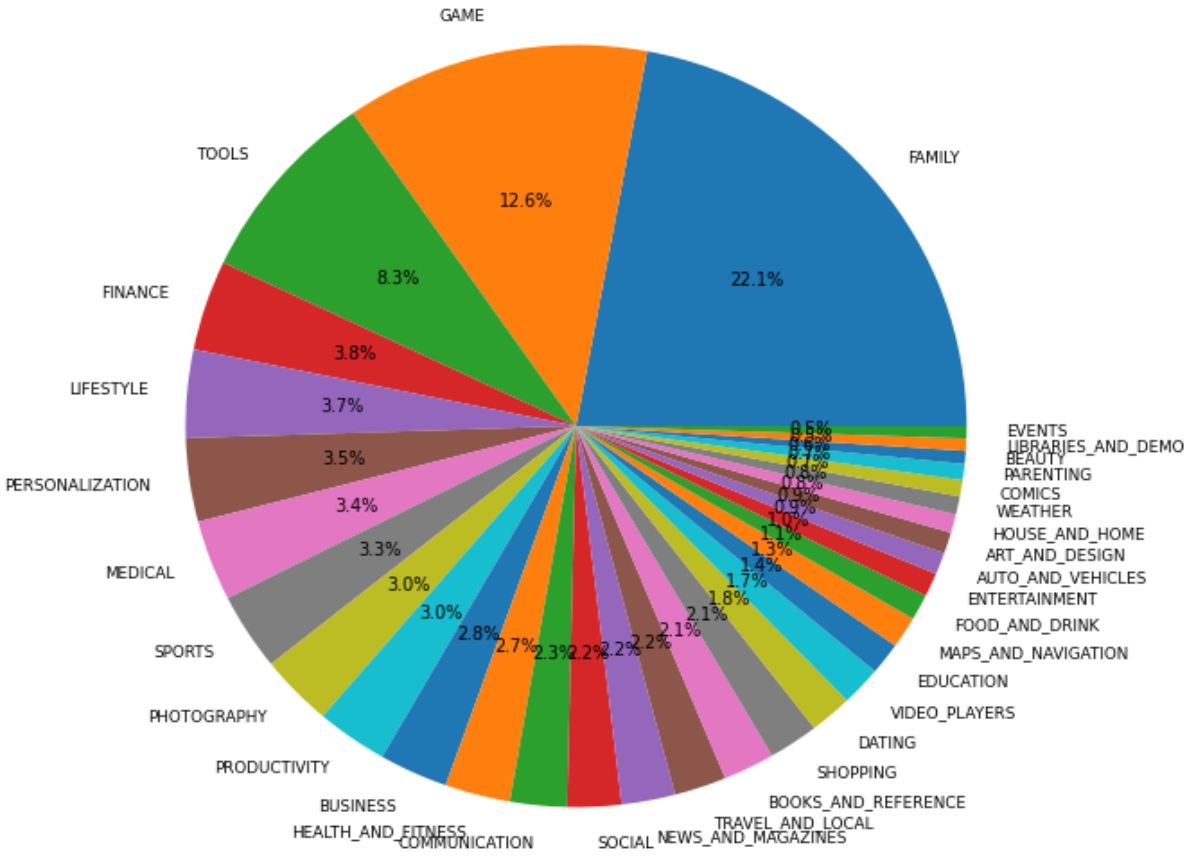

Figure 8.2 : Categories Vs No. of Applications (Pie Chart)

The category 'Family' has the greatest number of applications as in majority of households, there will be at least one kid and many users use it. More active mobile users are the youngsters and they are interested in playing games, so it contains the second largest number of applications. Next are the categories of Tools, finance and lifestyle.

Correlation between various numerical attributes :

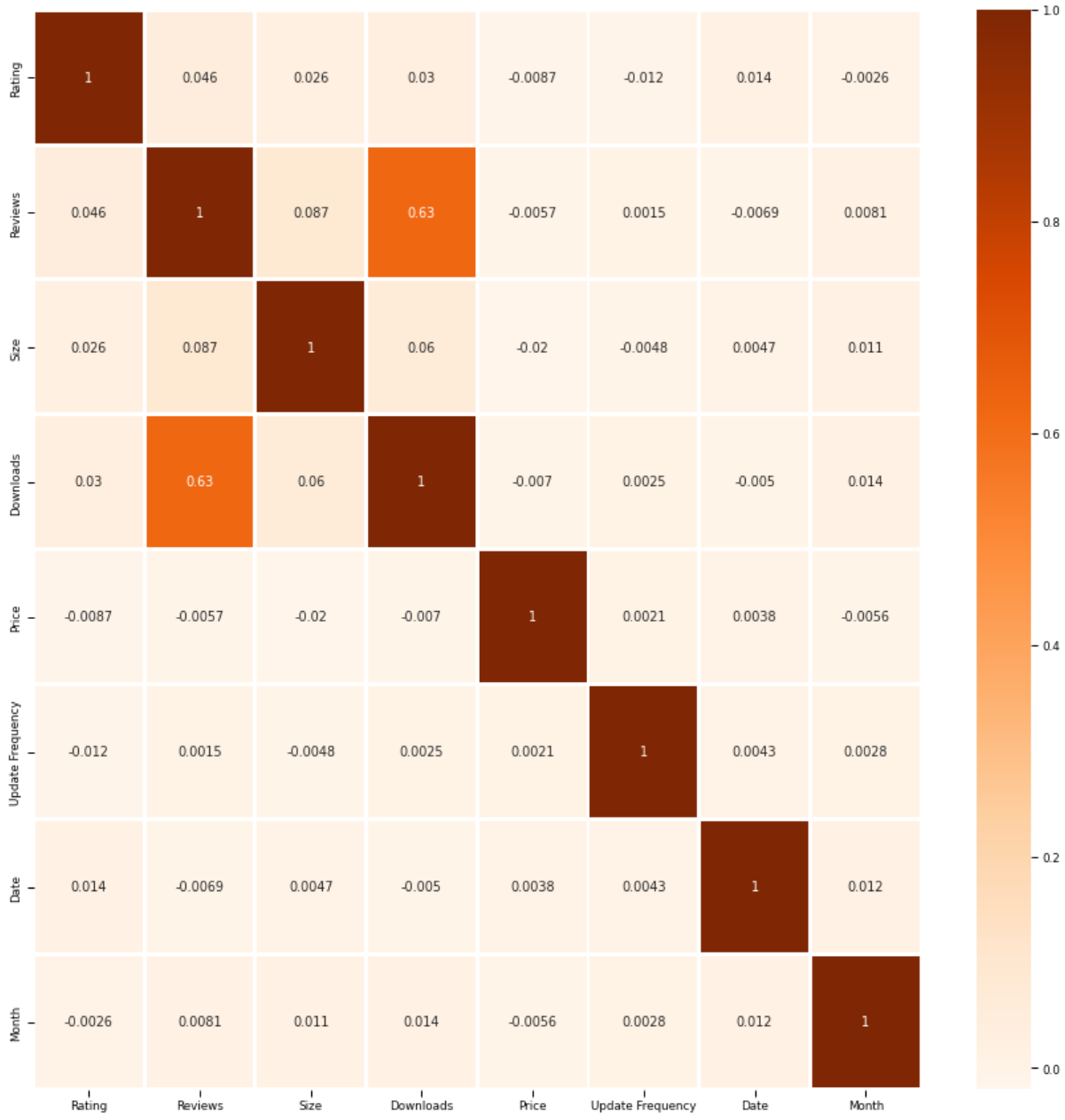

Figure 9 : Correlation between various numerical attributes.

From Figure 9, we get to know that Downloads and Reviews are highly correlated i.e the number of downloads affect the number of reviews. It is because, as number of downloads increase, many people review the applications since they have used it.

8. Predictions :

To predict the rating of the application, given the number of downloads, number of reviews and size of the application.
Number of downloads : 500000
Number of Reviews : 1000
SIze of the application: 20
Result : Predicted Rating value: 4.2
Error= 26%

9. References :

[1] Ceci L., (2022) 'Most popular Google Play app categories as of 1st quarter 2022, by share of available apps', Statista, May 18, Available at: https://www.statista.com/statistics/279286/google-play-android-app-categories/ (Accessed: 02 June, 2022).

[2] Bachchan V., (2020) '5 reasons why Minecraft is the best-selling video game of all time', sportskeeda, September 18, Available at: https://www.sportskeeda.com/esports/5-reasons-minecraft-best-selling-video-game-time (Accessed: 02 June, 2022).

[3] Norenko Y., (2018) 'Analysis of Apps in the Google Play Store', NYC Data Science Academy, August 08, Available at: https://towardsdatascience.com/data-science-a-deep-analysis-on-google-play-store-apps-from-kaggle-8283bbc508b0 (Accessed: 02 June, 2022).