Capstone Project-1

EDA on Play Store App Review

Team Member-

Zubair Ahmad
Asif Ansari
MD Ejaz Ansari

Contents

1. **Introduction**
2. **Objective**
3. **Flow chart**
4. **Problem statement and Data exploration**
5. **Description of Data**
6. **Data Cleaning**
7. **Data Analysis and Visualization**
8. **Problem Challenges faced and how overcome**
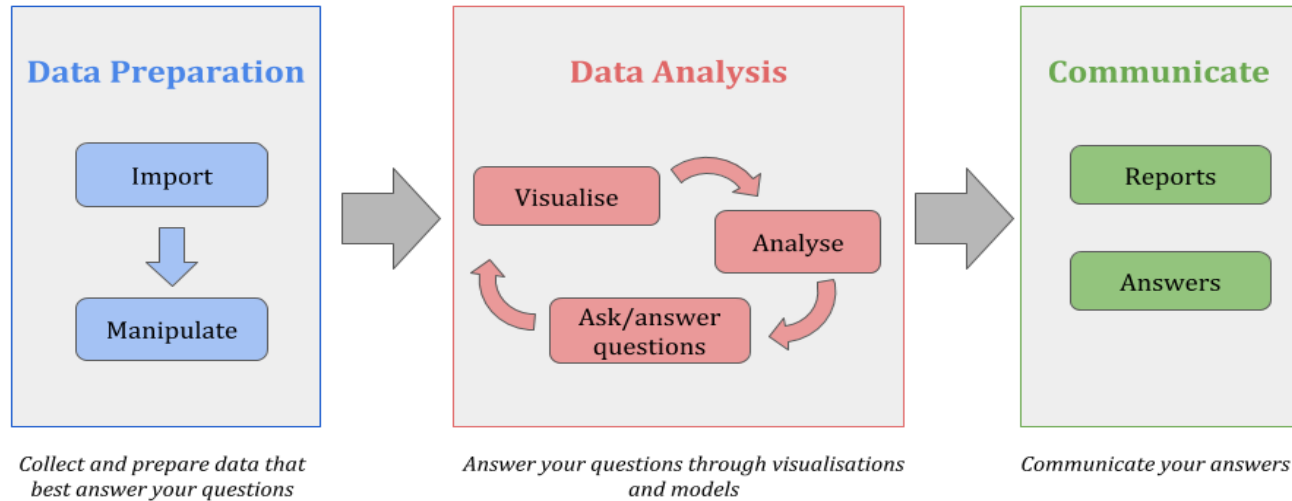9. **Conclusion**

INTRODUCTION

Google Play is an online store where people go to find and enjoy their favorite apps, games, movies, TV shows, novel books, and more on their Android devices. Google Play was launched on March 6, 2012 and now at that time Google Play is available in 190 countries and partners with developers from all over the world to distribute their apps and having developers from more than 190 different locations could distribute apps on Google Play. There are more then 5 million apps found on Google play store.

OBJECTIVE

- ➤ Convert the data from dataset
- ➤ Clean the data from the dataset
- ➤ Extract statistics form dataset
- ➤ Exploratory analysis of Dataset
- ➤ Question to be asked from the data
- ➤ Conclusion

FLOW CHART TO PROCESS

Data Preparation

Import

Manipulate

Collect and prepare data that
best answer your questions

Data Analysis

Visualise

Analyse

Ask/answer
questions

Answer your questions through visualisations
and models

Communicate

Reports

Answers

Communicate your answers

PROBLEM STATEMENT AND DATA EXPLORATION

We have two dataset provided from which one is **play store dataset** having basic information provided into it and other with **user reviews** for all the provided application.

We have to examine and evaluate the data in both datasets from which we have to find the important information from the data and their characteristics.

Play store dataset :- It contains basic details of all the app having 10841 rows and 13 columns and their feature are

- **App:** It contains name of the app with short description.
- **Category:** This column give the category to which an app belongs. This data set contains 33 categories.
- **Rating:** The average rating given by the users for the respective app.
- **Reviews:** The number of users that have dropped a review for this respective app.

DATA EXPLORATION

- **Size:** The disk space required to install the respective app.

- **Installs:** The approximate number of times the respective app was installed

- **Type:** It states whether an app is free to use or paid.

- **Price:** It gives the price payable to install the app. Price is 0 for free app.

- **Content Rating:** It states which age group is suitable to consume the content of the respective app.

- **Genres:** It gives the genres to which respective app belongs.

- **Last Updated:** It gives the date at which the latest update for the respective app was released.

DATA EXPLORATION

- **Current Ver:** It gives the current version of the respective app.

- **Android Ver:** It gives the android version of the respective app.

User reviews data set : It contains the user reviews and its sentiment score for the respective app having 64295 rows and 5 columns.

- **App:** It contains name of the app with short description.

- **Translated Review:** It contains the English translation of the review dropped by the user of the app.

- **Sentiment:** It gives the attitude/emotion of the writer. It can be 'Positive' , 'Negative' or 'Neutral'.

DATA EXPLORATION

- **<u>Sentiment Polarity:</u>** It gives the polarity of the review. Its range is [-1,1], where 1 means 'Positive statement' and -1 means a 'Negative statement'.

- **<u>Sentiment Subjectivity:</u>** This value gives how reviewer' s opinion is to the opinion of the general public. Its range is [0,1].

DISCRIPTION OF DATA

Data Description Document is to record all information about the data files and their contents so that someone can use the data in a future research project and understand the data's content and structure.

➢ The data which we have found is in the form of excel file.
➢ It's important that to convert data into data frame
➢ So, We have converted data into data frame

DATA CLEANING

- Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- When combining multiple data sources of data, there are many opportunities for data to be duplicated or mislabeled.
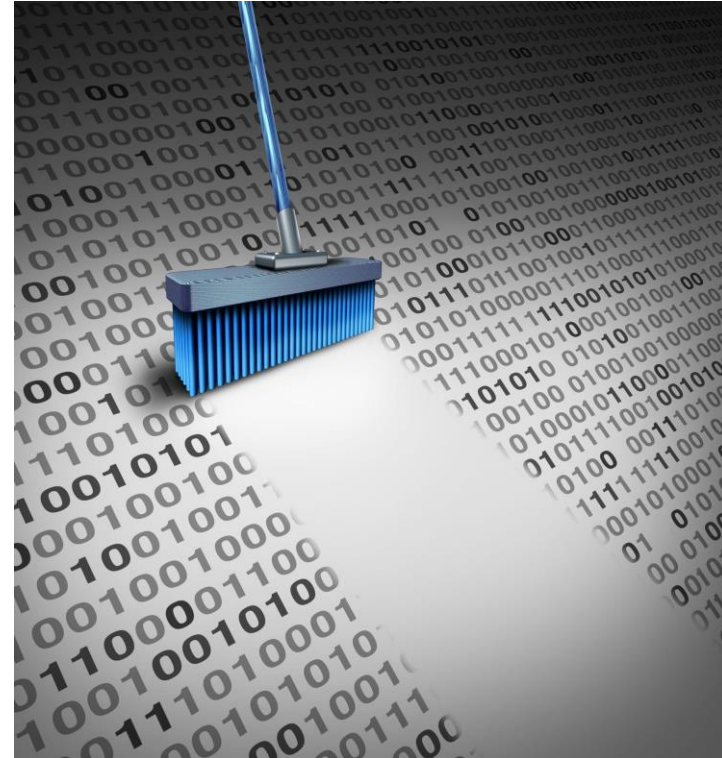
STEP TO CLEAN THE DATA

➤ Analyze the data which is impossible

➤ Drop the Data column which can't true

➤ Check the Data which have null values

➤ Drop the all Data which have null values

DATA ANALYSIS AND VISUALIZATION

☐ These are the top categories of Play store application which are used according to the given dataset. According to this graph we can say that-

❖ Family , Game, Tools are in the top category

❖ Photography, Dating, Shopping are in middle range.

❖ Weather, Comics and Beauty are in last.

Top categories on Google Playstore
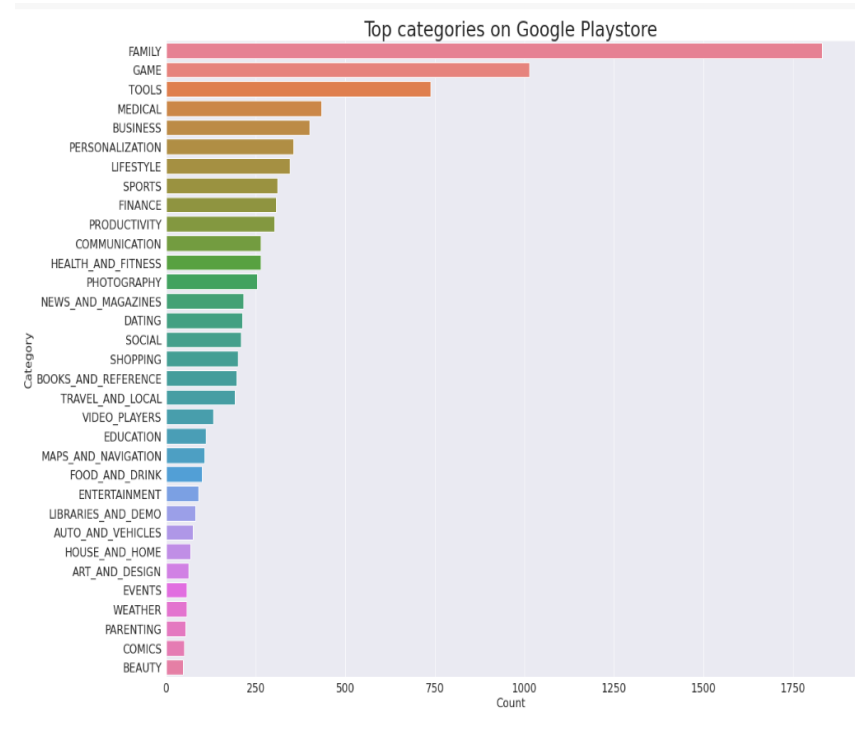
DATA ANALYSIS AND VISUALIZATION

The graph shows that which type of application are mostly install and used.

- These graph help to understand that which application should have to be in priority list

- **Family, Game , Tools , Medical** application are installed by highest number of user.

- **Printing, Comic, Beauty** application have very less user.
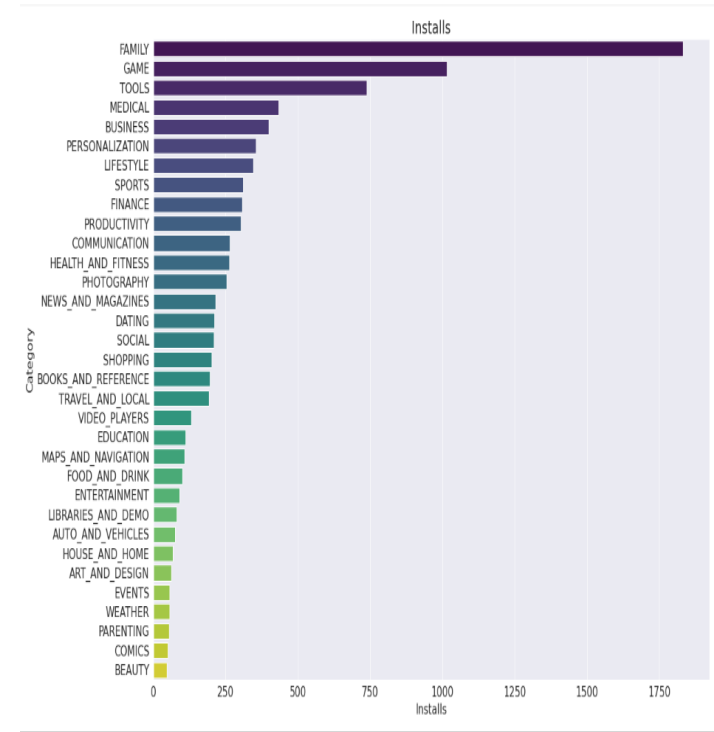
DATA ANALYSIS AND VISUALIZATION

**From the given figure we have found that these are the list of paid and free app install count.**

- Family, Games, Photography Personalization have higher paid user.

- Games, Business have higher free user
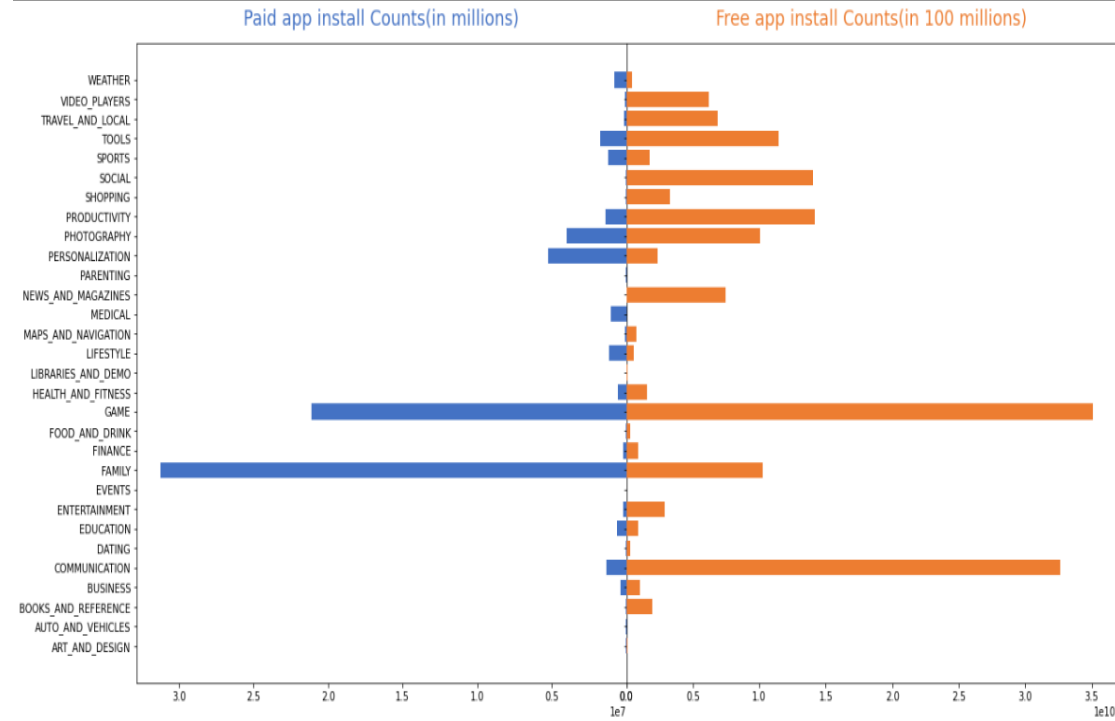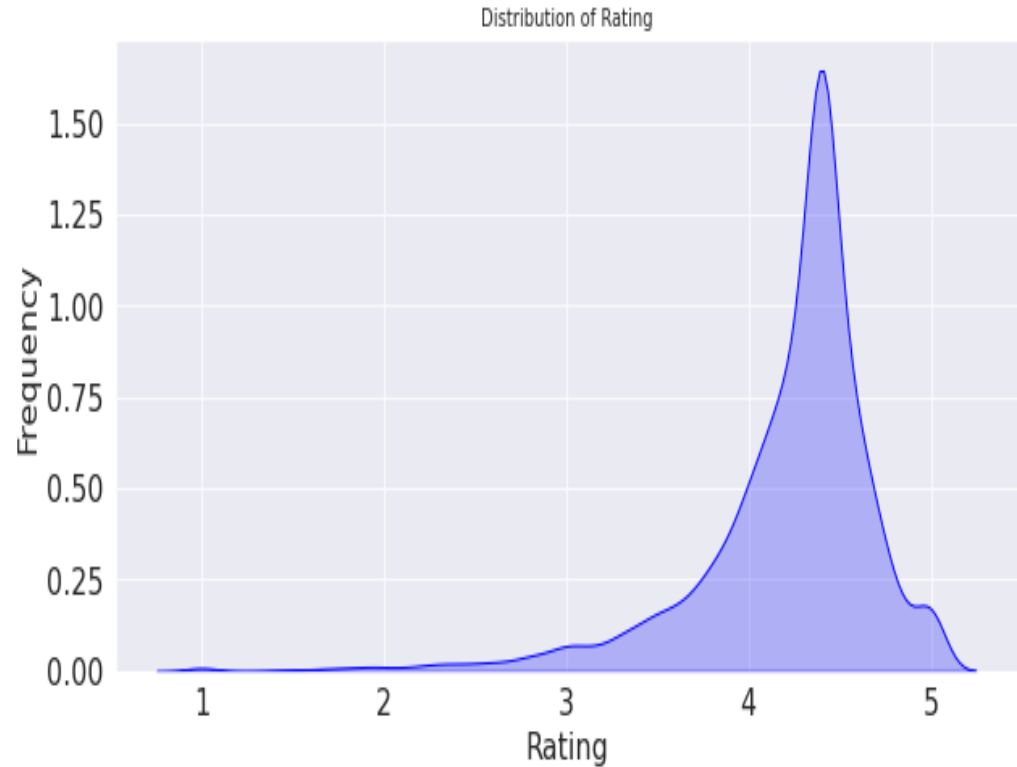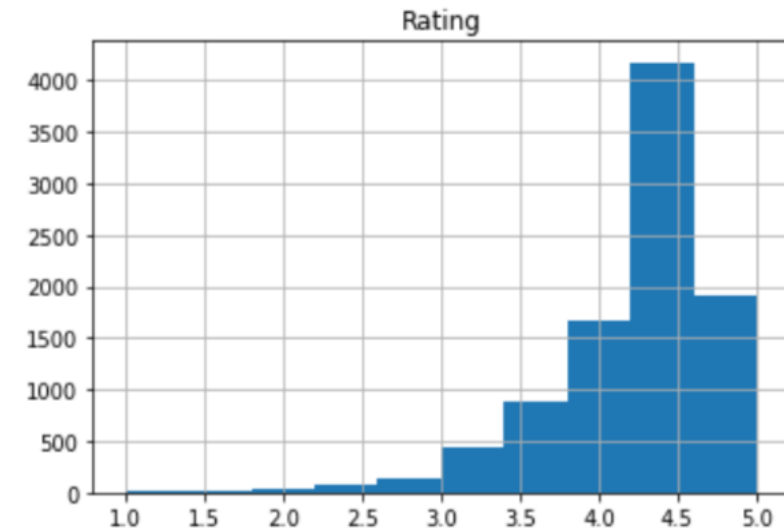
DATA ANALYSIS AND VISUALIZATION

APPLICATION RATING ANALYSIS BASED ON AGE

- Everyone having maximum number installation of 700 Billion.
- Application having content rating of adult only 18+ and unrated.
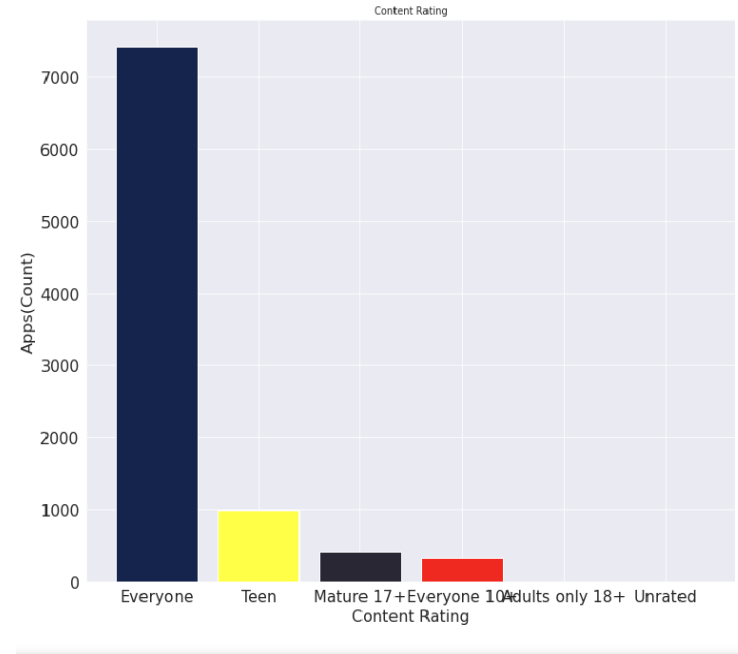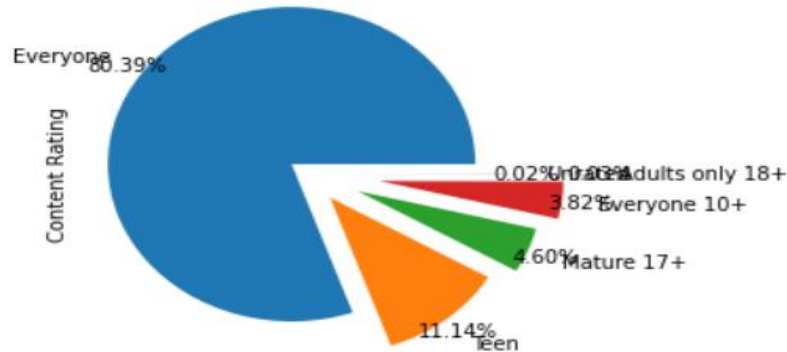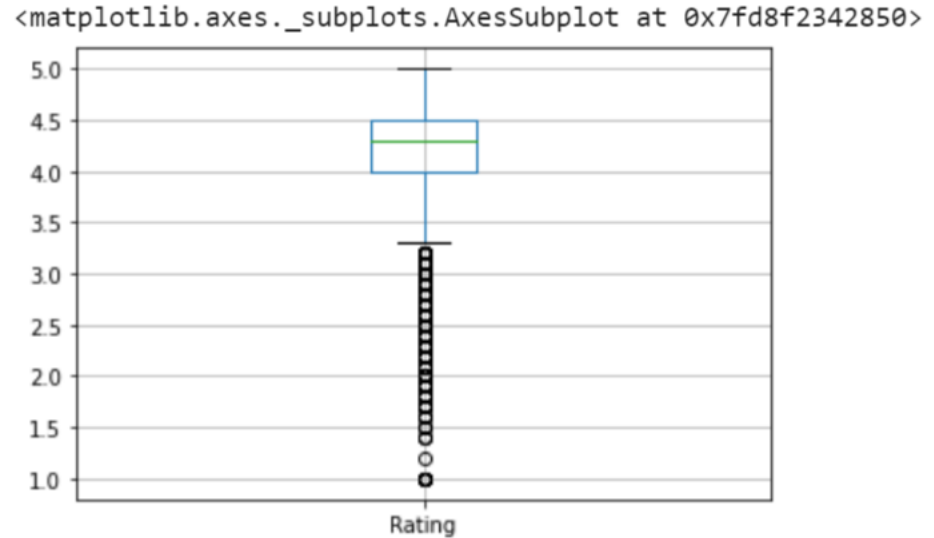
DATA ANALYSIS AND VISUALIZATION

Rating analysis by the given figure

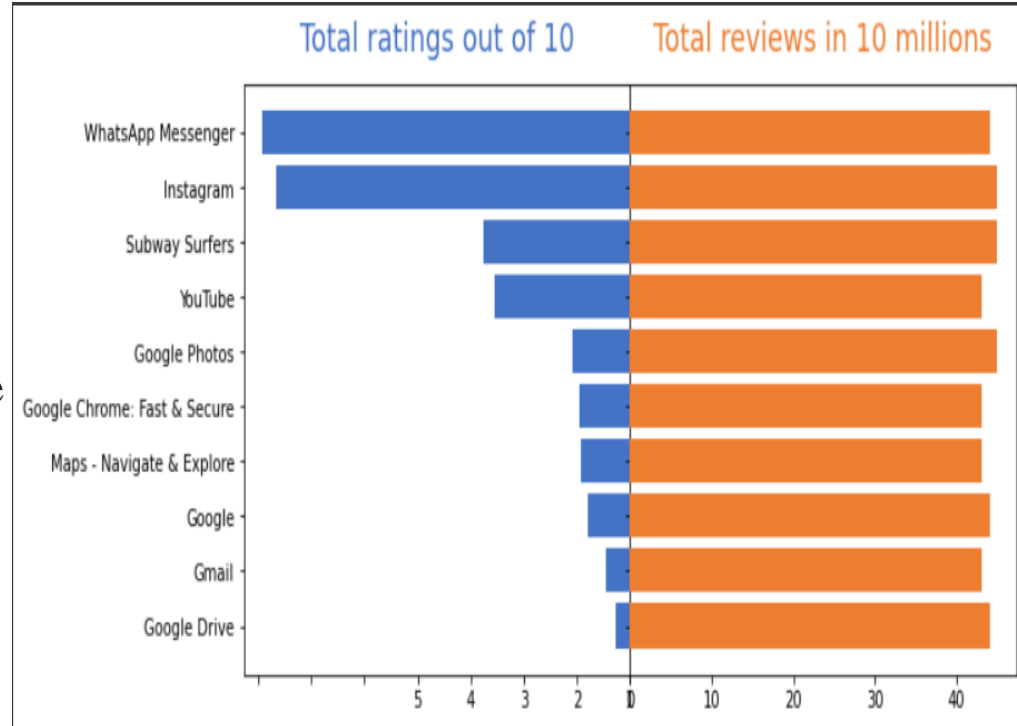# DATA ANALYSIS AND VISUALIZATION

Rating Analysis by Outlier

From the given graph we can say that
outlier have highest rating is 5
and maximum rating by the user is
between 4 to 4.5 and average rating is
4.3

`<matplotlib.axes._subplots.AxesSubplot at 0x7fd8f2342850>`

DATA ANALYSIS AND VISUALIZATION

From the given figure we have found that these are the top 10 application and their view in 10 millions

- WhatsApp Messenger, Instragram, Subway Surfers, Youtube, Google Photos have highest rating and these are top 5 app.
- Google Chrome, Map- Navigate & Explore, Google, Gmail, Google Drive are in the list of top 5 to 10
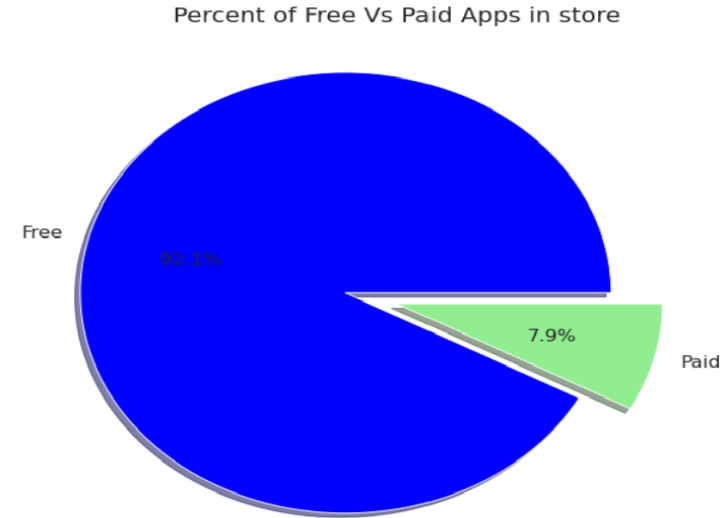
DATA ANALYSIS AND VISUALIZATION

User Analysis of Dataset

From the given figure we can say that free user is very high as compare to the free user.
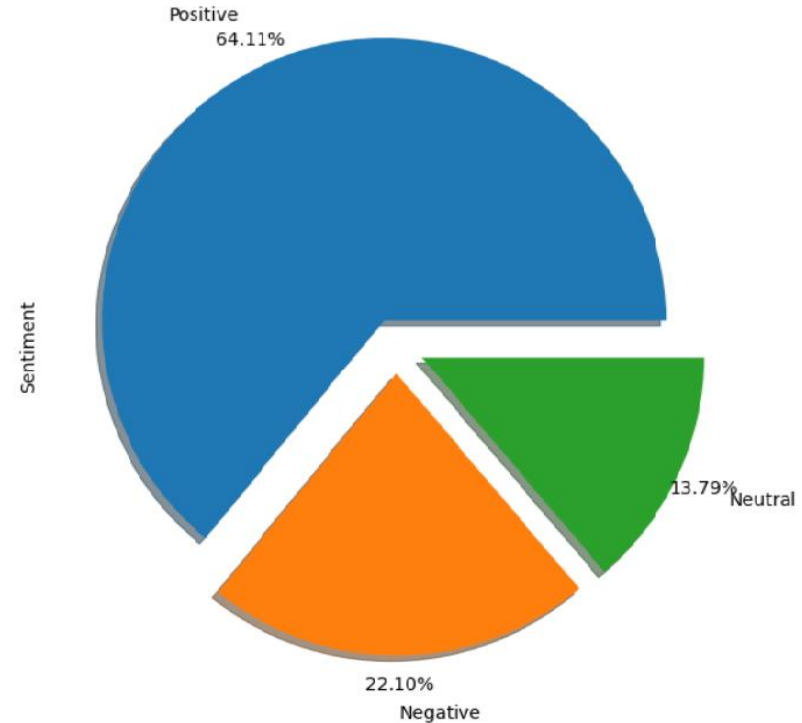
As we can see that only 7.9% of them are the paid user so demond of free app is very high.

Percent of Free Vs Paid Apps in store

Free 92.1%

7.9%

Paid

DATA ANALYSIS AND VISUALIZATION

Sentiment Percentage Of Reviews

From the given figure we can see that 64.11% of them have given positive review and 22.1% of them have given negative review and 13.79% of them are neutral

DATA ANALYSIS AND VISUALIZATION

**These are the top 10 application which is installed in the category of game like**

- FIFA Soccer
- Dream League Soccer
- 3D Bowing
- 8 Ball pool
- EA SPORTS UFA
- Wrestling Revolution 3D
- NBA LIVE Mobile Basketball
- 3D Tennis
- Soccer 2018 and NBA
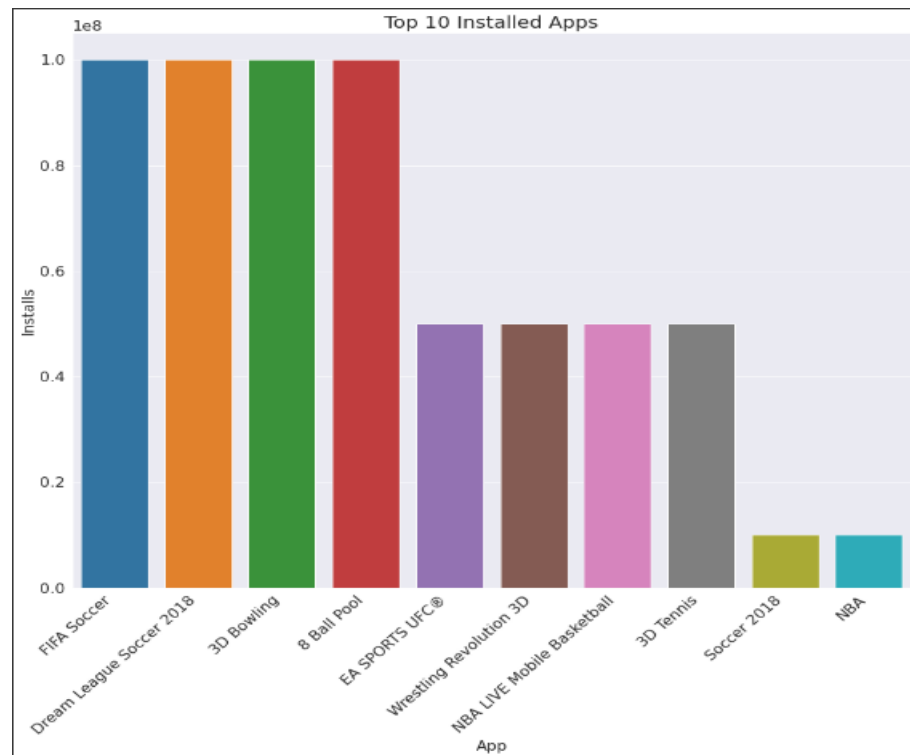
DATA ANALYSIS AND VISUALIZATION

This figure shows that all the categories of apps and their android Version used in the play store

- From this graph we see that android version 4 is found to be maximum
- Android Version 2 is found is at the second level
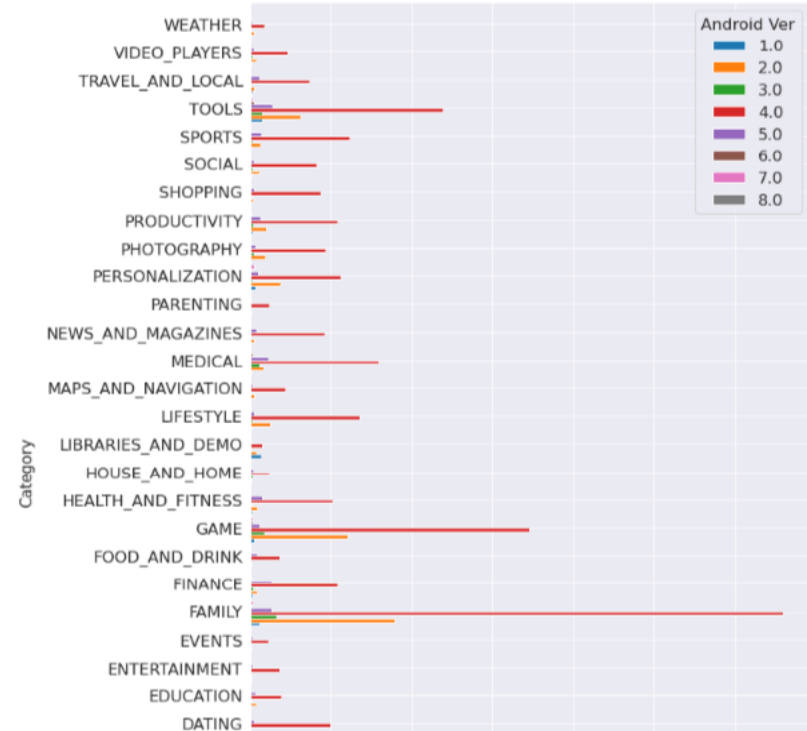
CHALLENGES FACED DURING PROJECT

- ➤ One of the major challenge we have faced is to clean the data from the datasets
- ➤ Find the error and duplicate and NaN values and handle them
- ➤ Drag the meaningful insight, we had to design multiple visualization.

HOW TO OVERCOME FROM THESE CHALLENGES

- ➤ Taken help from AlmaBetter Class Material
- ➤ Used Pandas, Numpy, Matplotlib, Seaboarn libraries
- ➤ GeeksforGeeks, W3School
- ➤ Youtube

CONCLUSION

➢ We have also found that most of the application which is used have a content rating of 'Everyone'

➢ Percentage of apps with no age restrictions 82%

➢ Most competitive category is family

➢ All the active application on play store has Average Rating is 4.3.

➢ Free user of application is 92% and paid user is 8%.

➢ We have found the list of 10 apps in a given category.

➢ We have top 10 gaming apps

REFRENCES

- ❑ Almabetter Deskboard
- ❑ Google colab notebook
- ❑ Researchgate.net
- ❑ Gethub.com
- ❑ Google.com