Google Play

Google play store App Prediction

Amal Alrashid

Table of Contents

01

Introduction

Mobile apps have become so prevalent, It has become important for developers to be able to predict the success of their apps. Google Play Store has countless apps, about 2 million in 2018.
The objective of this project is to understand the main features affecting Google Play Store Apps

02

Question

Can we predict the most used apps in Google Play Store?
Which app Category do users tend to download?
Can we predict app rating?

03

Dataset

Google Play

Dataset from Kaggle

Contains 10841 Apps

Contains 13 features

Contain numeric & categorical types

04

Data Cleaning

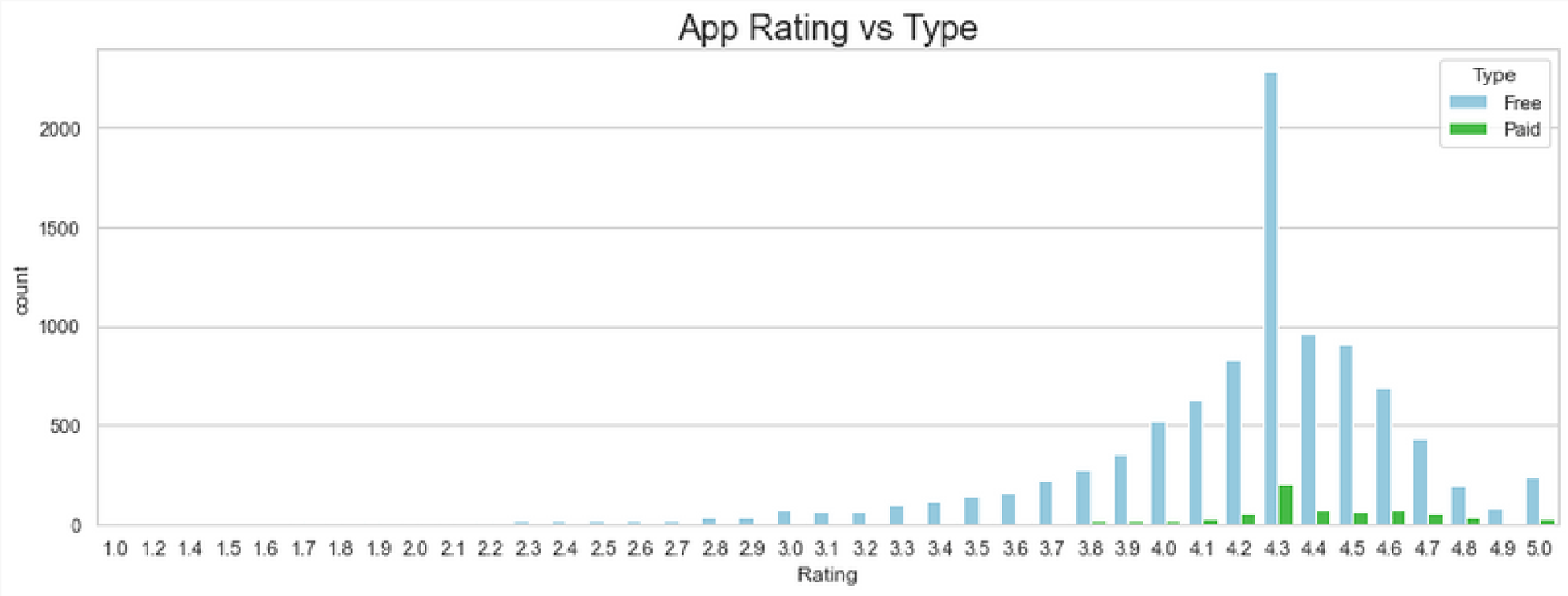
- Rename some features
- Convert categorical features into numerical features
- Convert categorical Values into numerical
- Fill missing value
- Rtype some features type
- Remove duplicate data

05

Visualization

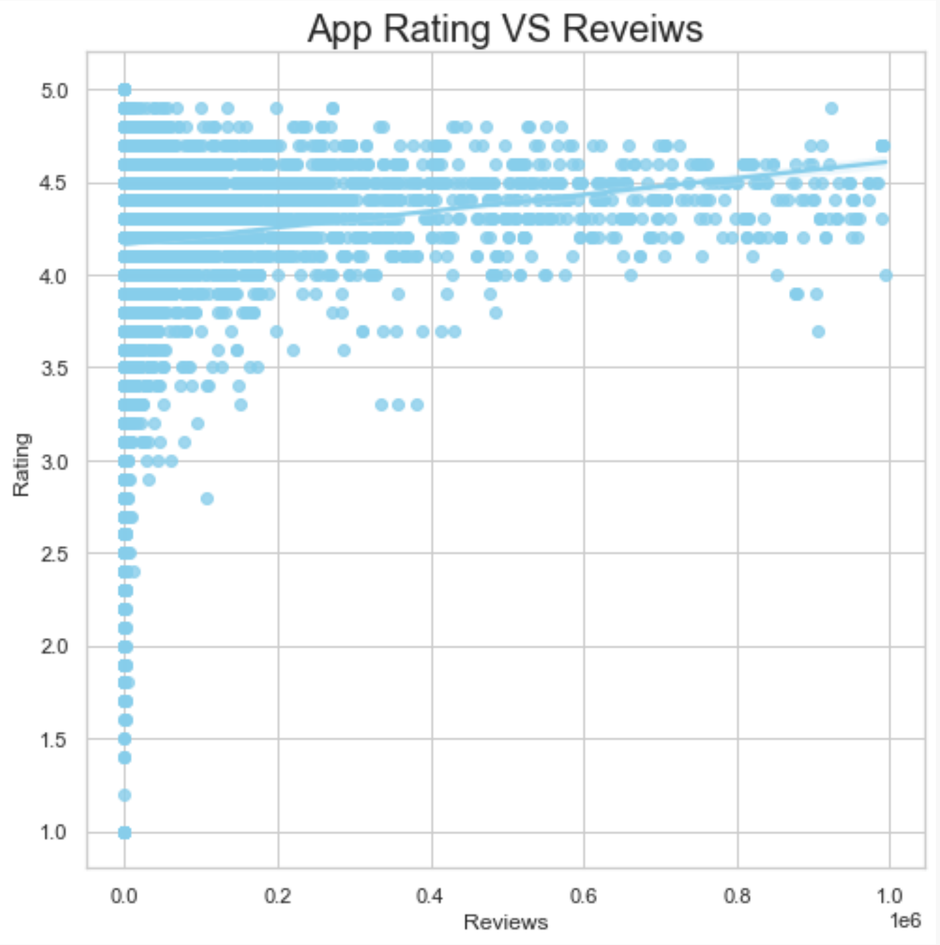
05

4.3 is the most common rating in Google Play Apps



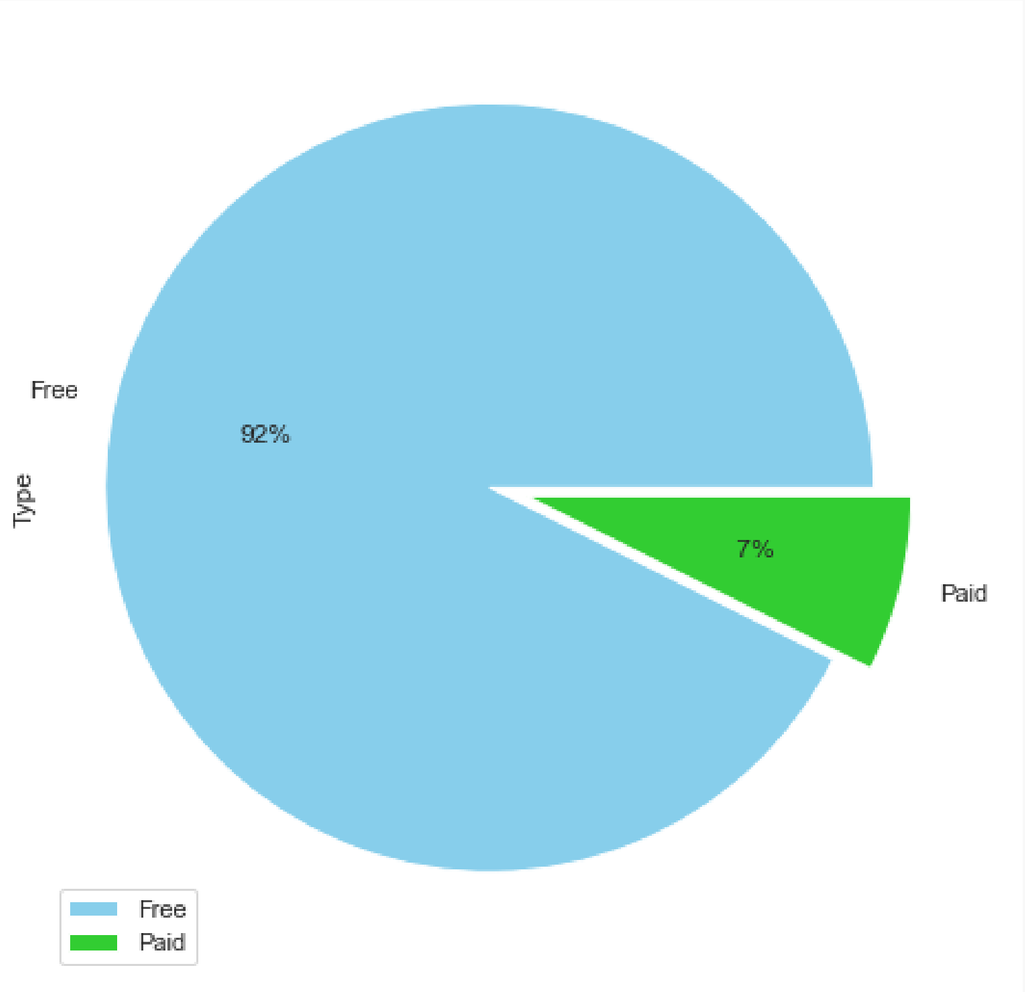
05

Apps with higher ratings tend to get more reviews



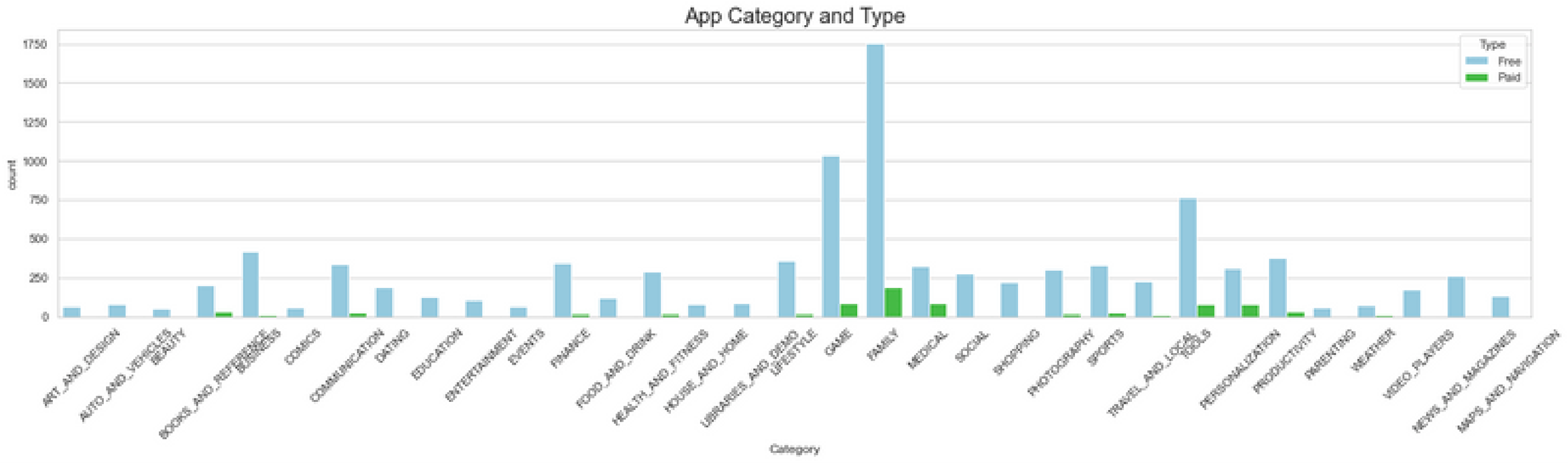
05

Most of the apps in the google store are free,
Only 7% of are paid apps



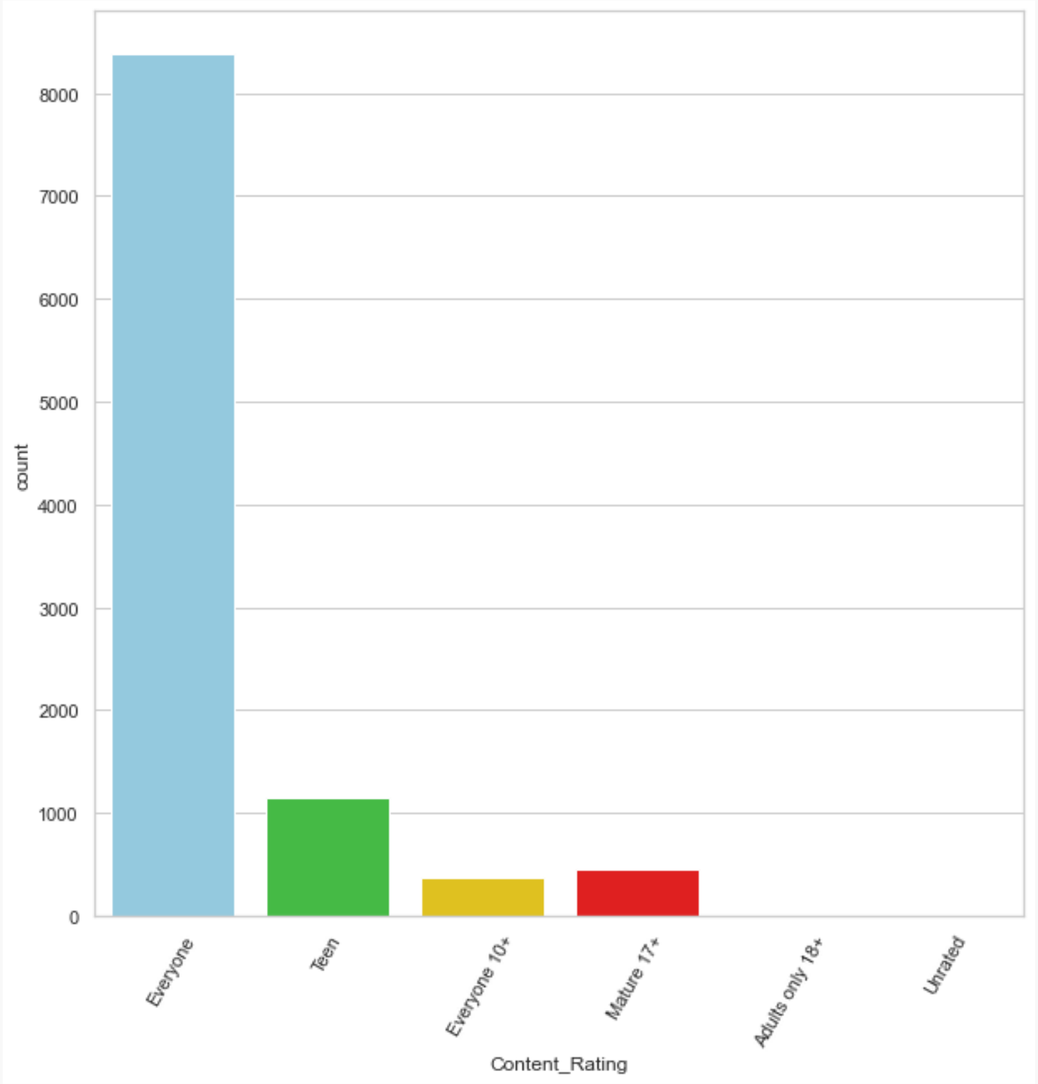
05

Family category is the most common category in the app store, then Game category



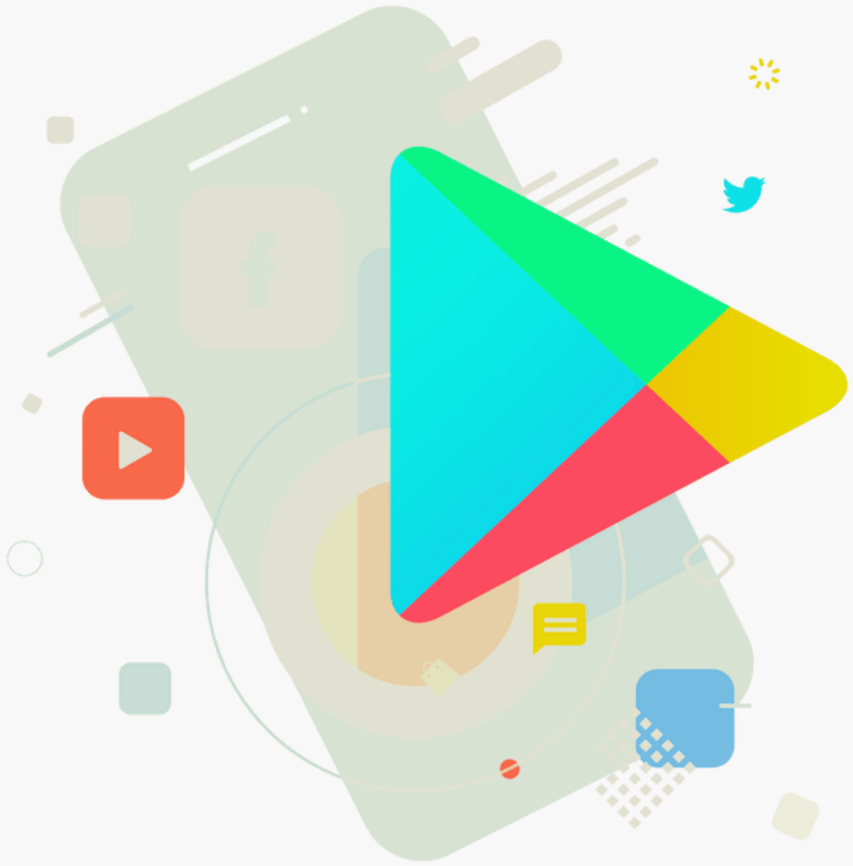
05

In terms of content rating, most Apps are rated as "Everyone"



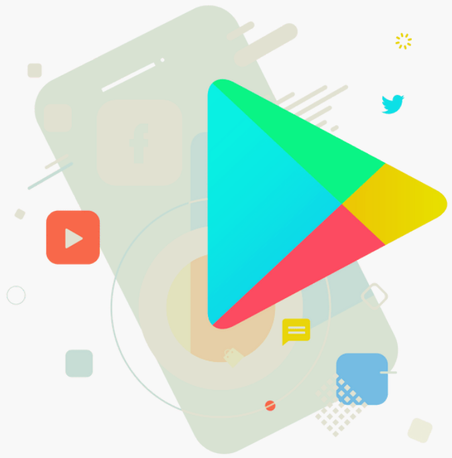
05

The most downloaded apps in google store



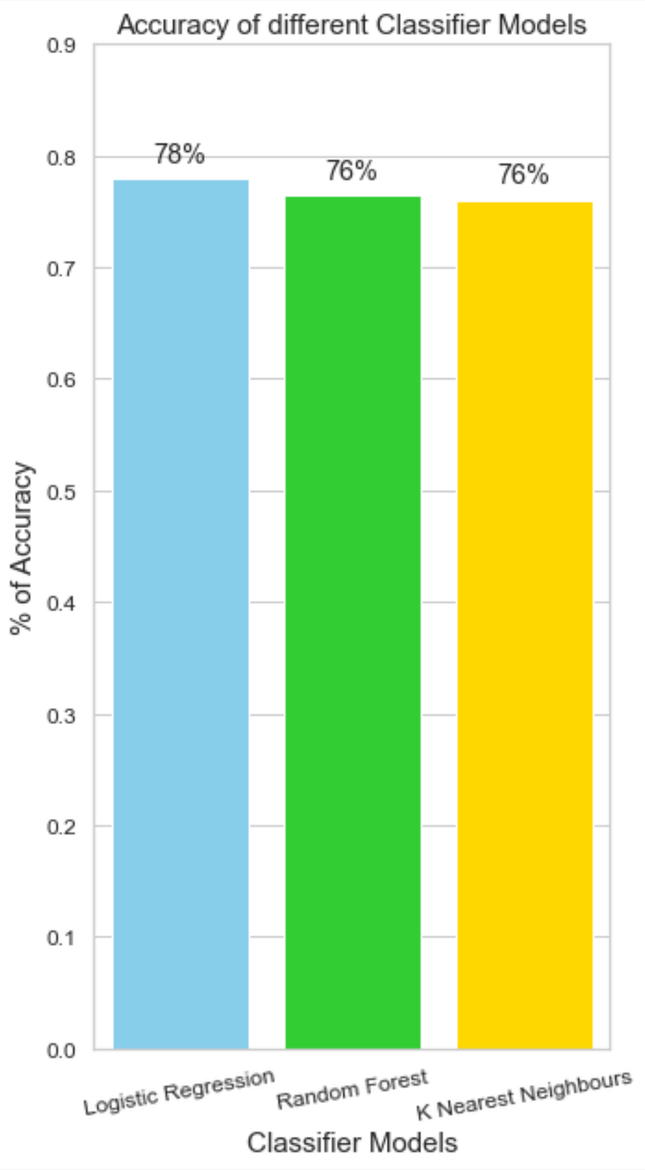
06

Data Model



06 The Model Used

- Logistic Regression
- Random forest
- KNeighbors Classifier

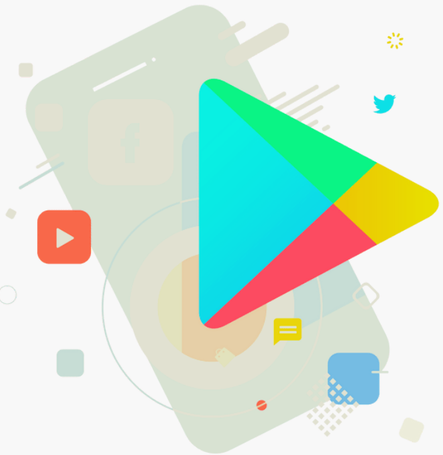


06 Model Accuracy

Logistic Regression :78%

Random forest : 76%

KNeighbors Classifier :76%



07

Conclusion & Tools

After applying three different models to the dataset, we can conclude that the Logistic Regression model ranked the highest accuracy among the other models

07

Conclusion & Tools

For Data Processing

Pandas & Numpy

For modeling library

Sklearn library

For Visualization

seaborn & matplotlib

Google Play

Thank you