Play Store App Review Analysis

Presented by – Gaurav Kumar

Abstract

Mobile apps are everywhere. They are easy to create and can be lucrative. Because of these two factors, more and more apps are being developed. In this project, we will do a comprehensive analysis of the Android app market by comparing over ten thousand apps in Google Play across different categories. We'll look for insights in the data to devise strategies to drive growth and retention. This data set have two csv file:

- Play store apps.csv: contains all the details of the applications on Google Play. There are 13 features that describe a given app.

- user_reviews.csv: contains 100 reviews for each app, most helpful first. The text in each review has been pre-processed and attributed with three new features: Sentiment (Positive, Negative or Neutral), Sentiment Polarity and Sentiment Subjectivity.

Problem Statement:

Android is expanding as an operating system. It has captured around 74% of the total market which is a true indicator of the huge amount of population using android. Our goal is to help android developers to know what the motivating factor for people is to download an app. It will also help to find out the factors that affect someone's decision to download an app. I would like to Analyze category, reviews, price, ratings and installs for this purpose and find out how they are interrelated.

Introduction

Google Play, formerly known as Android Market, is the official distribution storefront for Android applications and other digital media, such as music, movies and books, from Google. It is availableon mobile devices and tablets that run the Android operating system (OS), supported Chrome OS devices and on the web. Users can access Google Play to browse, purchase and download software applications from Google and third-party developers.

INTEGRAL METHODOLOGY:

Discussion of Google play store dataset will involve various steps such as:

- loading the data into data frame

- cleaning the data

- extracting statistics from the dataset

- exploratory analysis and visualizations

- questions that can be asked from the dataset

- Conclusion

Data Cleaning

Data cleaning is the foremost step in any data science project. Cleaner the data, better are the results. We handled the missing values in some of the columns by either dropping the rows or imputing them with the median values, depending on the percentage of nulls in each feature. Then we remove duplicates values from the columns. We also remove special characters like ('+','$', ',') in our dataset. We transformed some of the columns like Installs, Size and Price to numeric type for ease of analysis.

Exploratory Data Analysis

Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images.
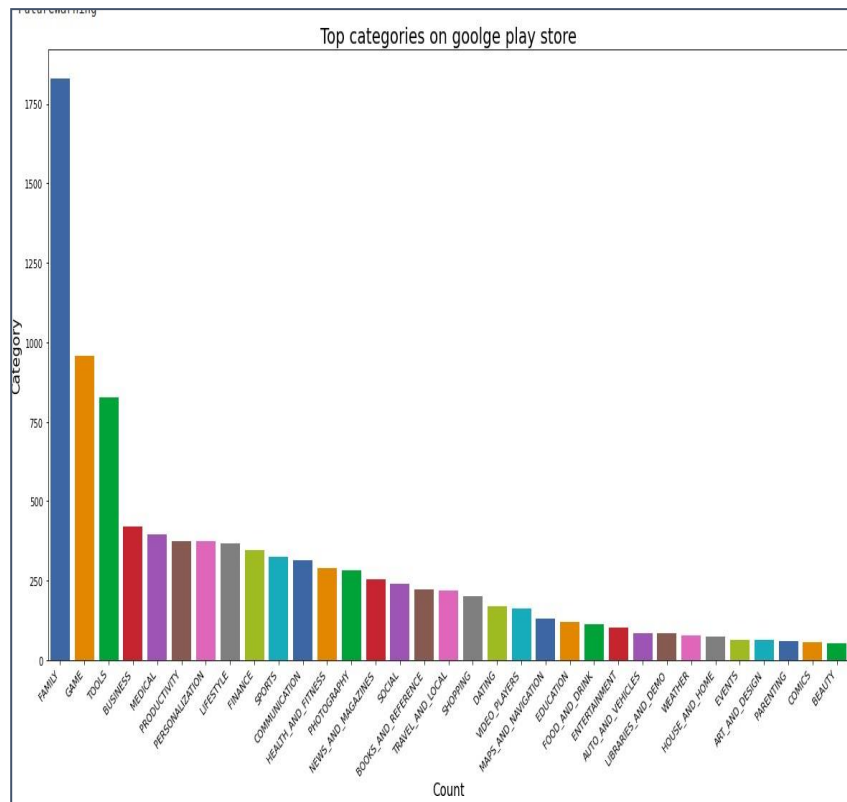
1. Top 10 Installed Apps
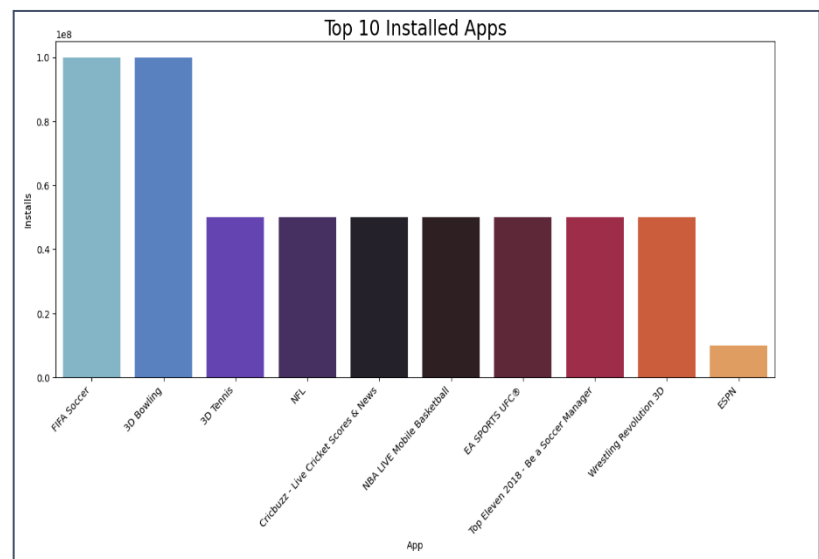
From the above graph we can see that in the Sports category FIFA Soccer and 3D Bowling havethe highest installs.
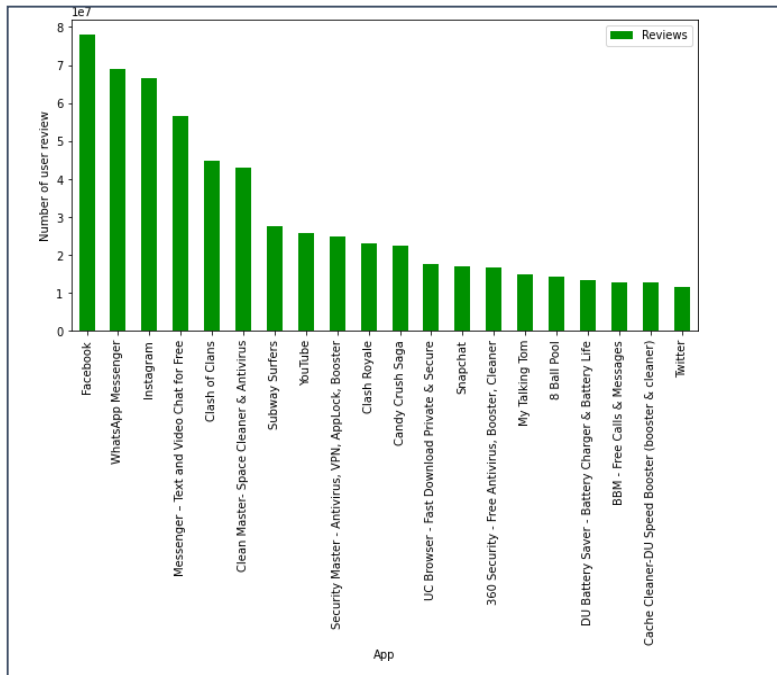
2. Paid vs Free

3. Top categories on play store
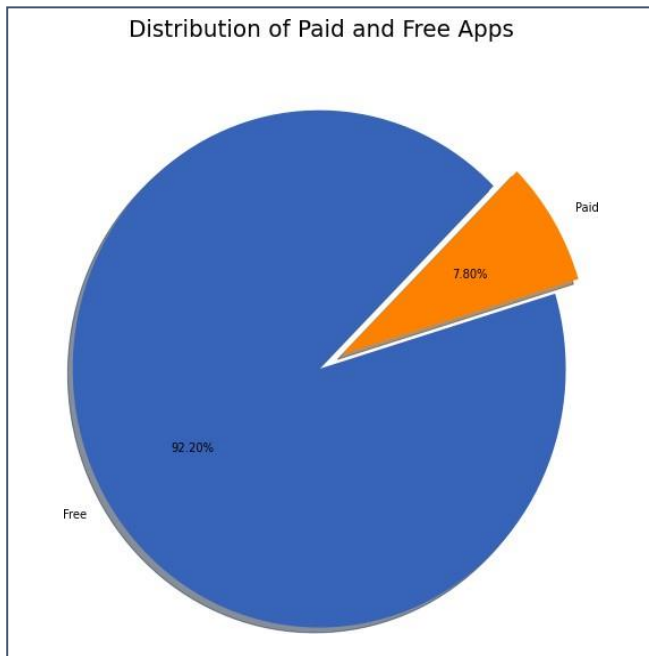
Free apps in play store are 92.20% and

Paid apps in play store are 7.80%

There are a total of 33 categories in the dataset from the above output. We can conclude that in play store most of the apps are under` FAMILY & GAME` category and least are of `EVENTS & BEAUTY` Category.

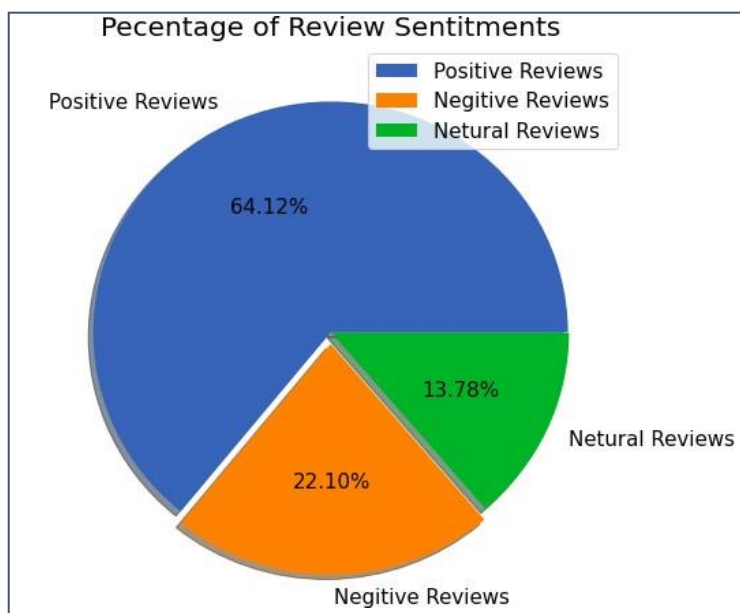## 4. Highest number of reviews

Top 20 apps with the highest number of reviews. Facebook has the highest number of userreview
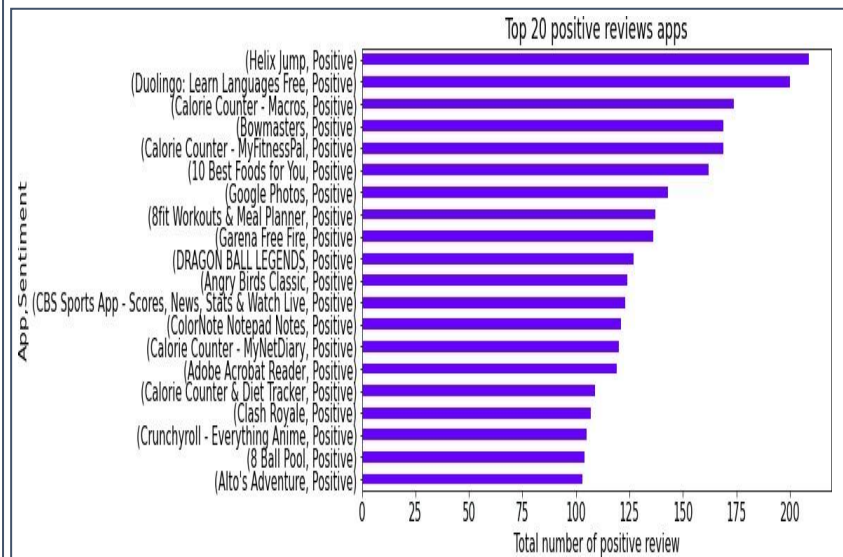
5. Top 10 Expensive apps
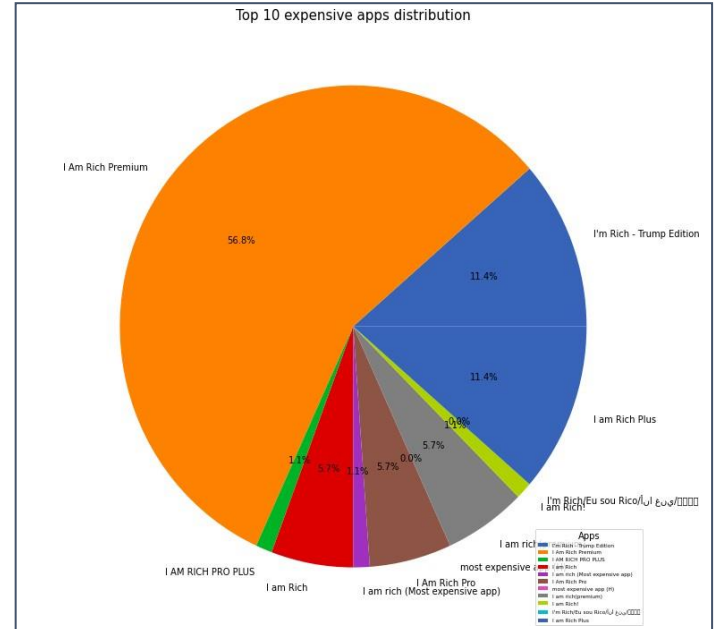
From the above graph we can interpret that the I am Rich Premium app is the most expensive appin the play store.
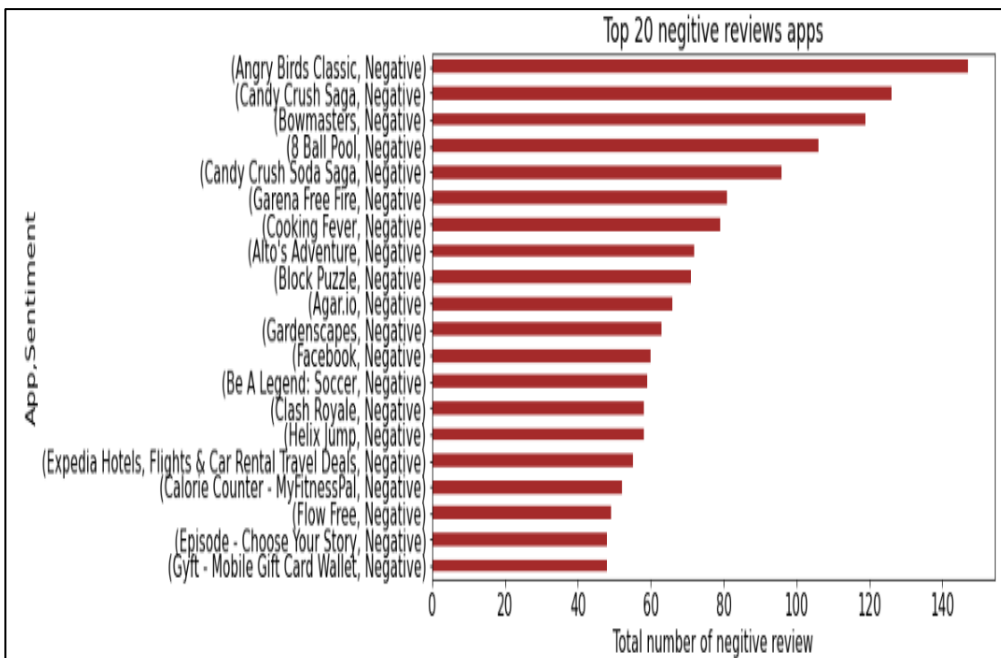
6. Percentage of review sentiments

- Positive reviews are 64.12%
- Negative reviews are 22.10%
- Neutral reviews are 13.78%

7. Highest positive reviews

Helix jump as max no. of positive review

8. Highest Negative review

Top 20 negitive reviews apps

Angry Birds Classic has got highest negative reviews

Conclusions

- In the Sports category FIFA Soccer and 3D Bowling has the highest number of installations.
- Percentage of free apps = 92.20%
- Maximum apps in the play store are from Family category
- Category with the highest number of installs: Game
- Most popular app in the Play Store based on the number of reviews: Facebook
- I am Rich Premium app is the most expensive app in the play store.
- Overall percentage of review sentiment in which Positive sentiment count is 64%, Negative 22% and Neutral 14%.

- Helix Jump has the highest number of positive reviews
- Angry Birds Classic has the highest number of negative reviews

References:
- GeeksforGeeks
- AlmaBetter Class material
- YouTube
- W3schools.com