



Google Play

Capstone Project

Play Store App Review Analysis

By

Manjeet Srivastava

Vivek Kumar

Rishabh Srivastav

Shubham Bharti

Durga Shankar

Data Science Trainee, AlmaBetter



WHY ANALYZE THE GOOGLE PLAY STORE?



Mobile App Market
is set to grow 20%
by 2023



Android Apps
comprise 90% of the
Mobile App Market



What makes an App
popular? Can we predict
how popular it's going to
be?



What are some
interesting patterns in
user behavior related to
app usage & feedback?



Introduction

- Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.
 - Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .
 - Android is the dominant mobile operating system today more than 85% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store.
 - There are more than 3.04 million apps found on Google Play Store.
 - The Play Store apps data has enormous potential to drive app-making businesses to success.
 - Actionable insights can be drawn for developers to work on and capture the Android market. The main goal of our project is–
- 1) The purpose of our project is to gather and analyze detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.
 - 2) The Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.



Problem Statement

- ❑ Two datasets are provided, one with **basic information** and the other with **user reviews** for the respective app.
- ❑ We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success.

So, what factors influence an app's success?

An app is said to be successful if it has:

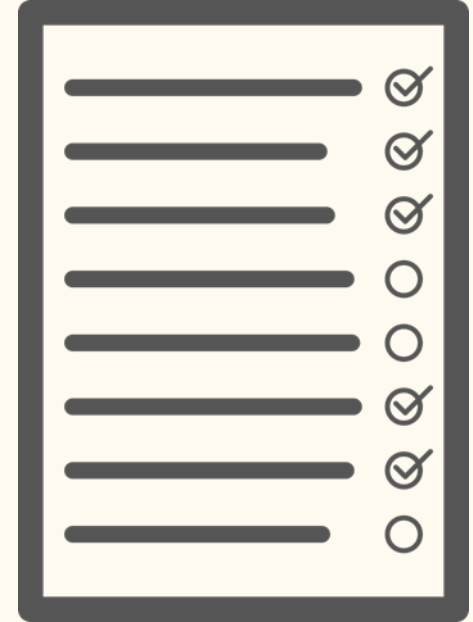
- ❑ A high average user rating
- ❑ A good number of positive reviews
- ❑ A good number of monthly average users
- ❑ High revenue per customer and so on.



- ☐ NUMBER OF APPS IN EACH CATEGORY
- ☐ MOST OF THE FALLS UNDER WHICH GENRE
- ☐ NUMBER OF FREE AND PAID APPS IN EACH CATEGORY
- ☐ PERCENTAGE OF FREE AND PAID APPS
- ☐ NUMBER OF APPS INSTALLS IN EACH CATEGORY
- ☐ DISTRIBUTION OF PAID APPS
- ☐ Distribution of FREE APP
- ☐ TOP 10 INSTALLED APP IN SPORTS CATEGORY
- ☐ TOP 10 PAID APPS
- ☐ AVERAGE PRICE OF PAID APPS IN EACH CATEGORY
- ☐ TOTAL PRICE IN EACH CATEGORY(SUM)
- ☐ WHAT IS THE FREE AND PAID APPS SIZE (IN MB)
- ☐ DATA SHAPE
- ☐ DISTRIBUTION OF RATING



- ☐ POINT PLOT OF AVERAGE RATING IN EACH CATEGORY
- ☐ COUNTPLOT OF RATING(TOP10)
- ☐ PLOT THE BARGRAPH FOR CONTENT RATING
- ☐ BARPLOT DISPLAYING THE RATING FOR EACH CONTENT RATING
- ☐ PLOTTING THE OVERALL SENTIMENT OF REVIEWS
- ☐ TOP APPS HAVING HIGHER NUMBER OF REVIEWS
- ☐ PLOTTING OF POSITIVE REVIEWS
- ☐ WORD CLOUD FOR POSITIVE REVIEWS
- ☐ PLOTTING OF NEGATIVE REVIEWS
- ☐ WORD CLOUD OF NEGATIVE REVIEWS
- ☐ CORRELATION BETWEEN RATING,REVIEWS,SIZE AND INSTALLS

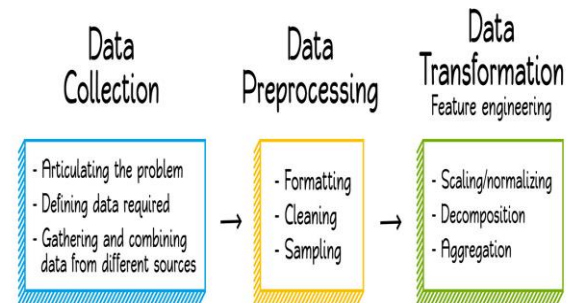




Dataset Preparation

- **Loading the data sets**: Two datasets, First Play store app dataset and User Reviews dataset.
- **Import Libraries**: NumPy, Pandas, Seaborn and Matplotlib
- **Data cleaning**: Null values, Finding and removing Outliers, Removing duplicate data.
- **Data Imputation**: Filling the missing categorical values with mode and numerical values with median. Conversion of price, installs, reviews into numerical values.
- **Exploratory Data Analysis**: Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

Data Preparation Process





Attributes in Google Play store Data

- 1.**App**:This column Contains the name of the app for each observation.
- 2.**Category** :This column Contains Category to which the app belongs.
- 3.**Rating** : This column contains the average rating for the app.
- 4.**Reviews**:This column contains the number of reviews that the app has received on the play store.
- 5.**Size**:This column contains the amount of memory the app occupies on the device.
- 6.**Installs**:This column contains the number of times that the app has been downloaded and installed from the play store.
- 7.**Type**:This column contains the information whether the app is free or paid.
- 8.**Price**: If the app is a paid app, this column contains the data about its price.
- 9.**Content Rating**: This column contains the maturity rating of the app ie. the age group of the audience for which it is suitable.
- 10.**Genres**: This column contains the data about to which genre the app belongs. Genres can be considered as a further division of the group of Category.
- 11.**Last Updated**: Contains the date on which the latest update of the app was released.
- 12.**Current Version**: Contains information on the current version of the app available on the play store.
- 13.**Android Version**: Contains information about the android versions on which the app is supported.



Attributes in User reviews

1. **App**– Application name
2. **Translated Review**– User review
3. **Sentiment**– Positive/Negative/Neutral Reviews
4. **Sentiment Polarity**– Sentiment polarity score
5. **Sentiment Subjectivity**– Sentiment subjectivity score





OVERVIEW OF ANALYSIS

Data Cleaning



Understand the structure of the dataset and clean data before analysis

Data Exploration



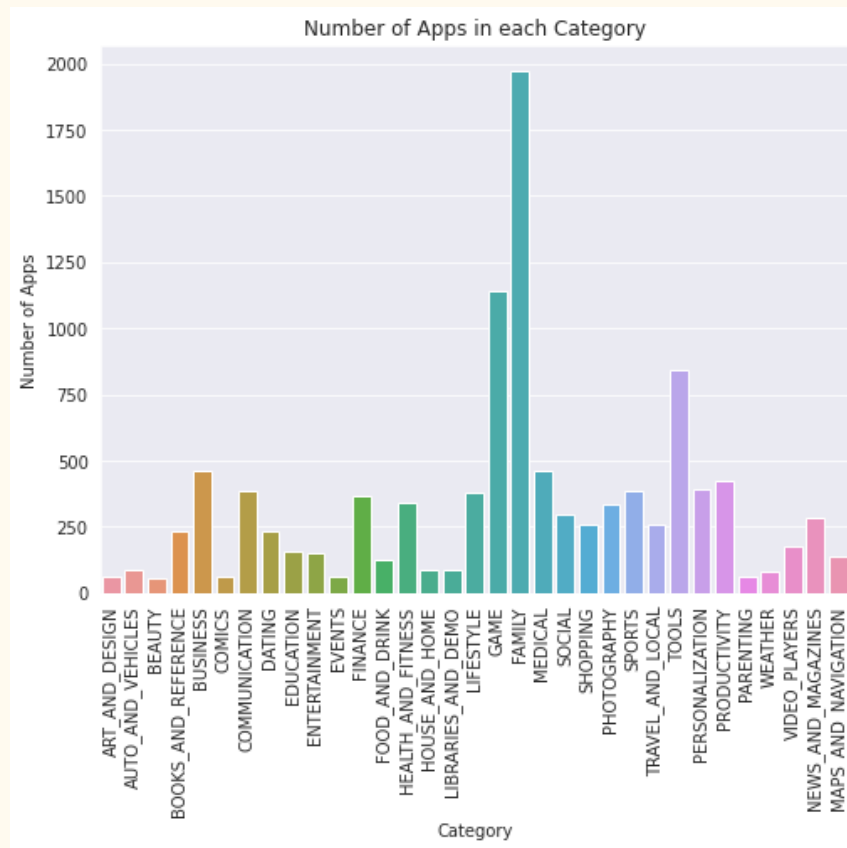
Uncover initial patterns, characteristics, and points of interest using visual exploration

Predictive Modeling

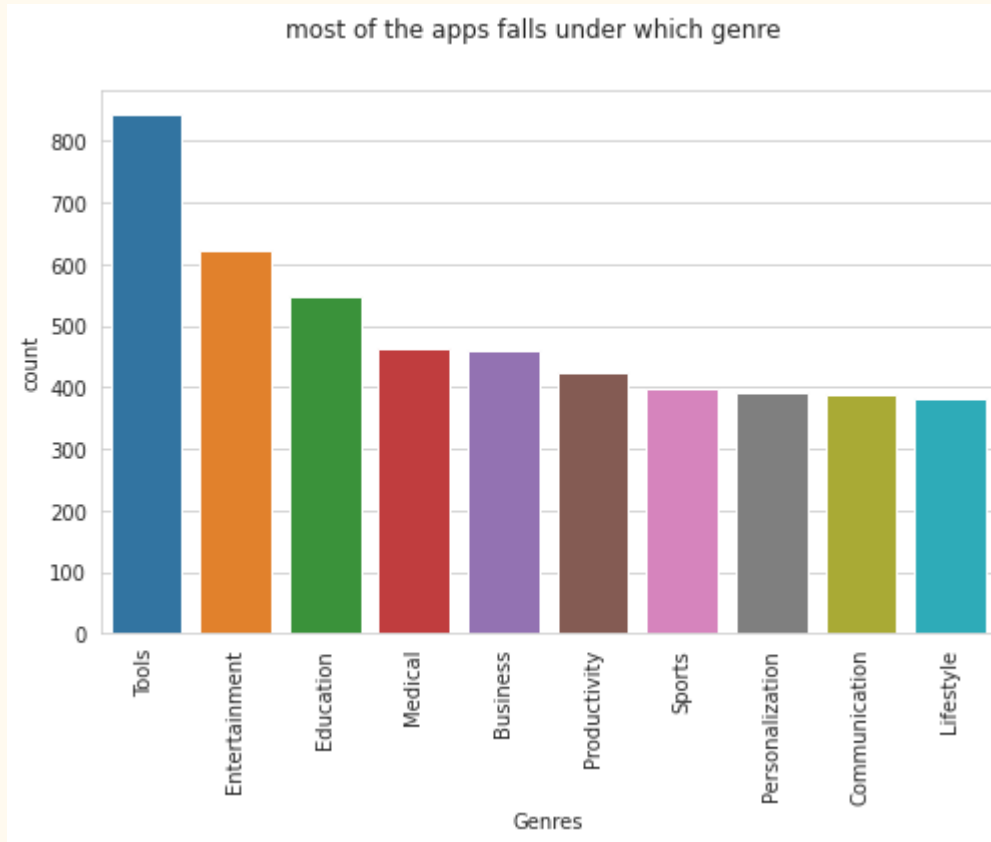


Formulate a statistical model to forecast an outcome using relevant predictors

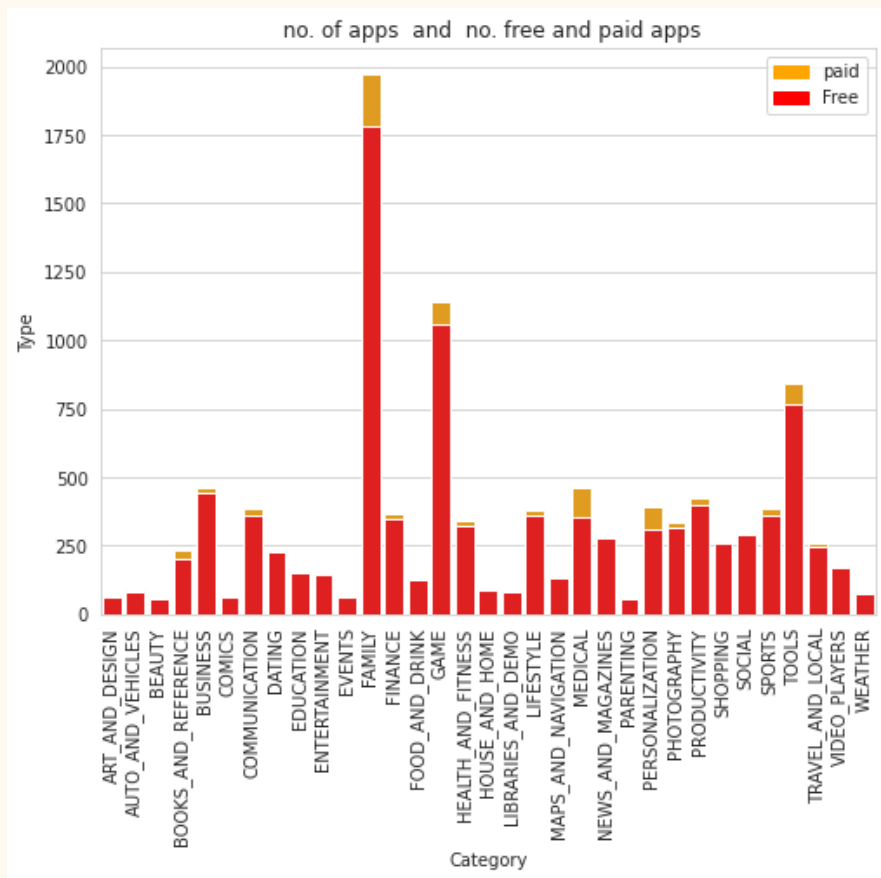
1. NUMBER OF APPS IN EACH CATEGORY



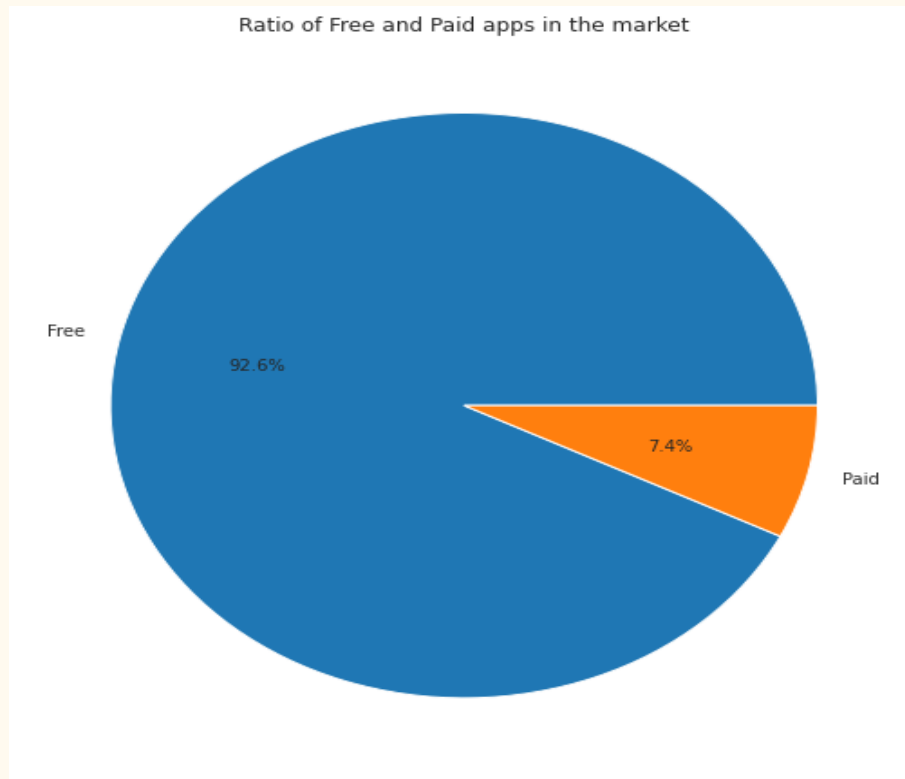
2. MOST OF THE FALLS UNDER WHICH GENRE



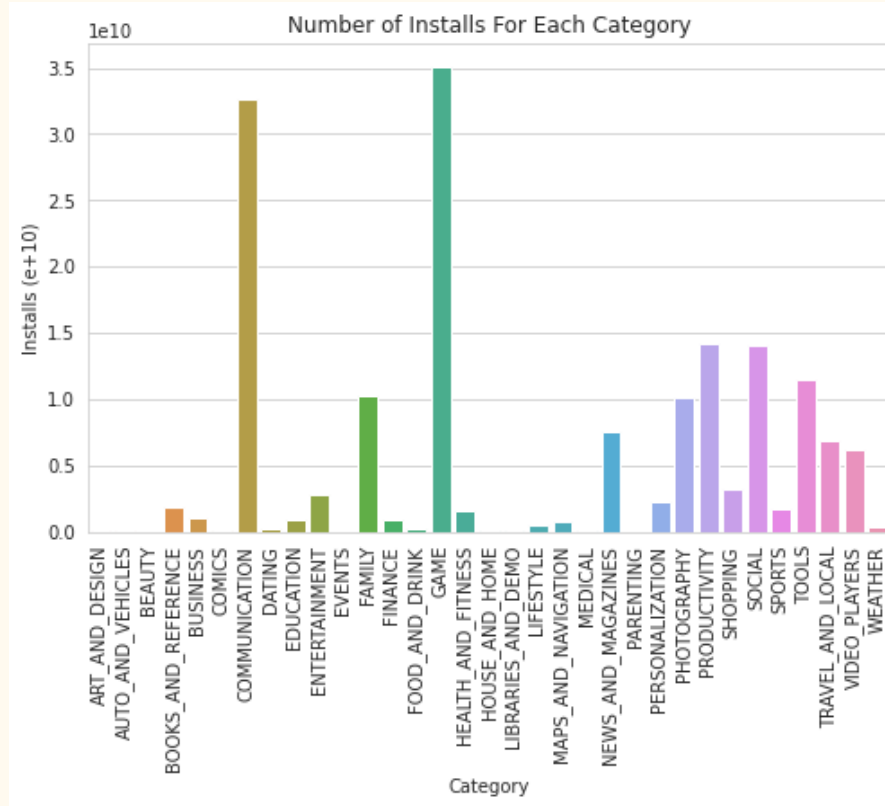
3. NUMBER OF FREE AND PAID APPS IN EACH CATEGORY



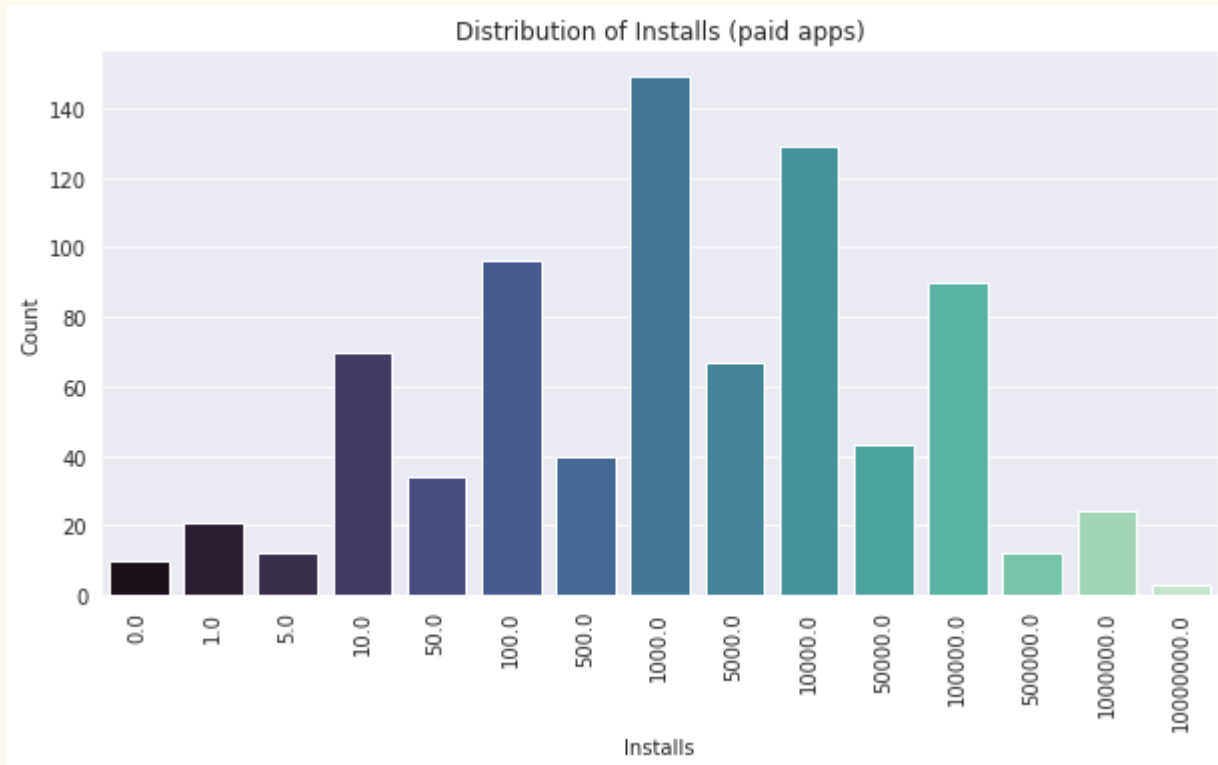
4. RATIO OF FREE AND PAID APPS



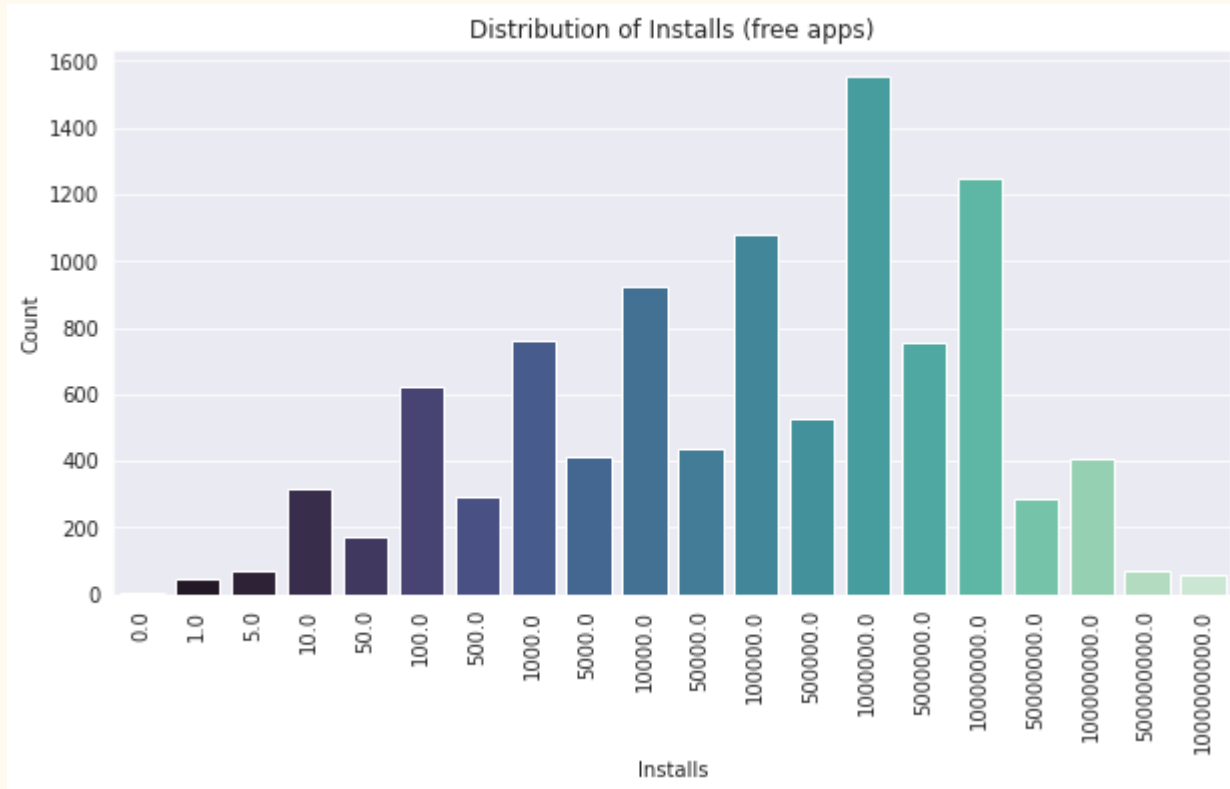
5. NUMBER OF APPS INSTALLS IN EACH CATEGORY



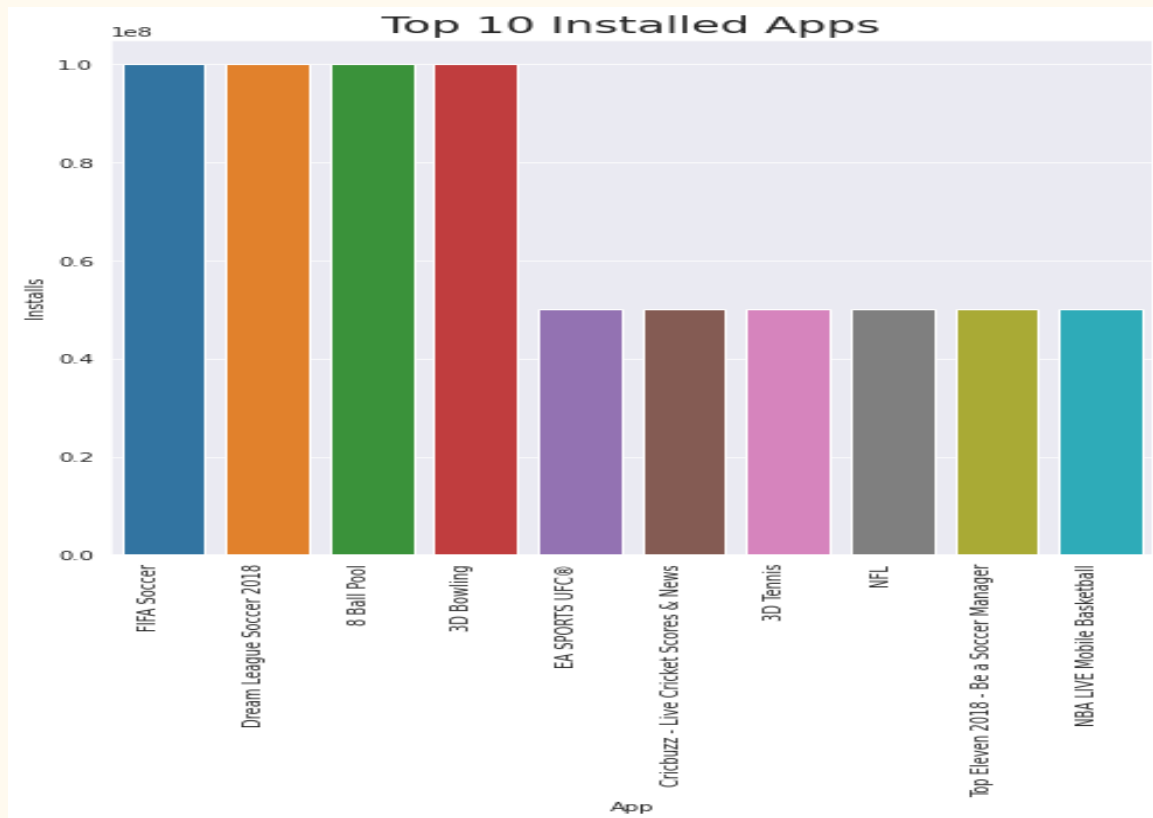
6. DISTRIBUTION OF PAID APPS



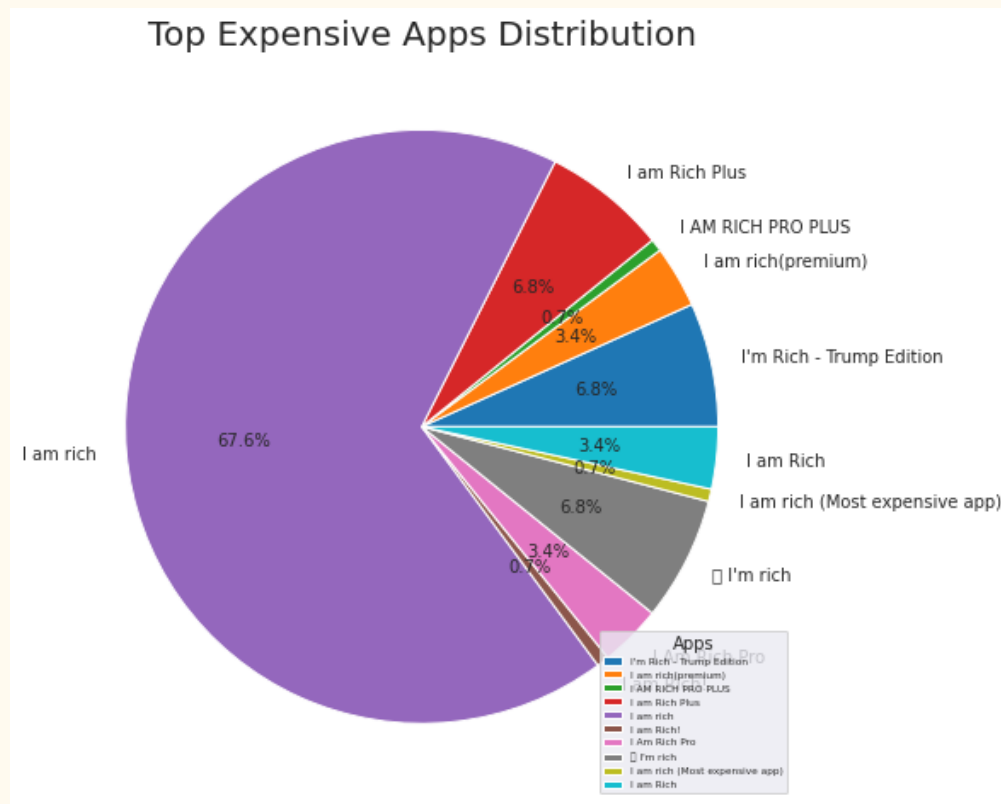
7.DISTRIBUTION OF FREE APPS



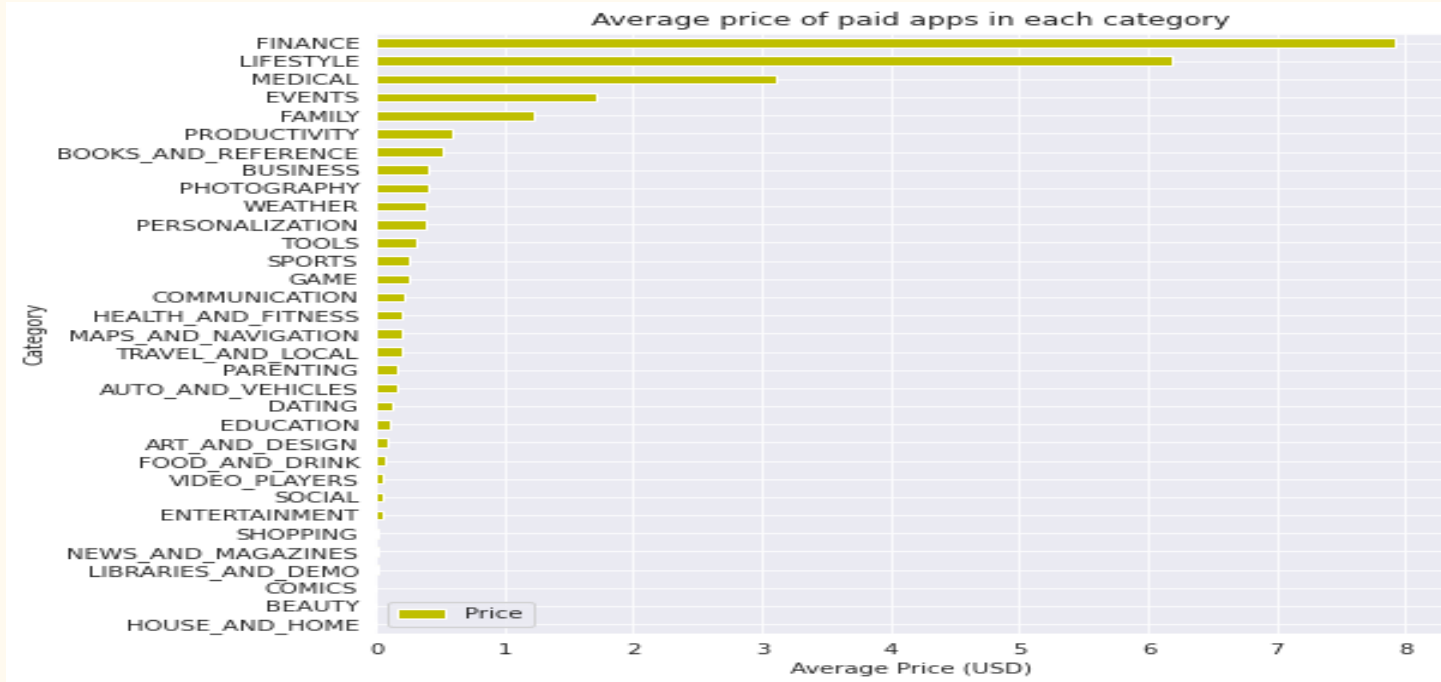
8. TOP 10 INSTALLED APP IN SPORTS CATEGORY



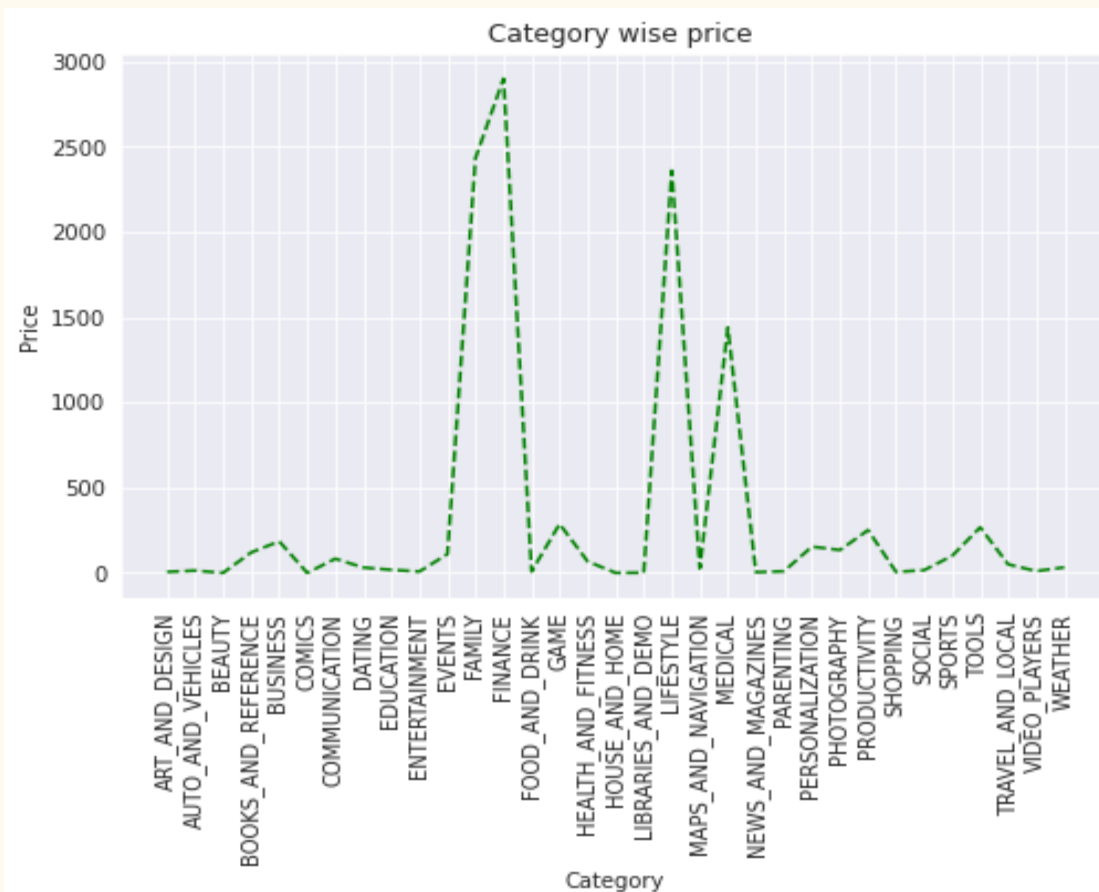
9. TOP 10 PAID APPS



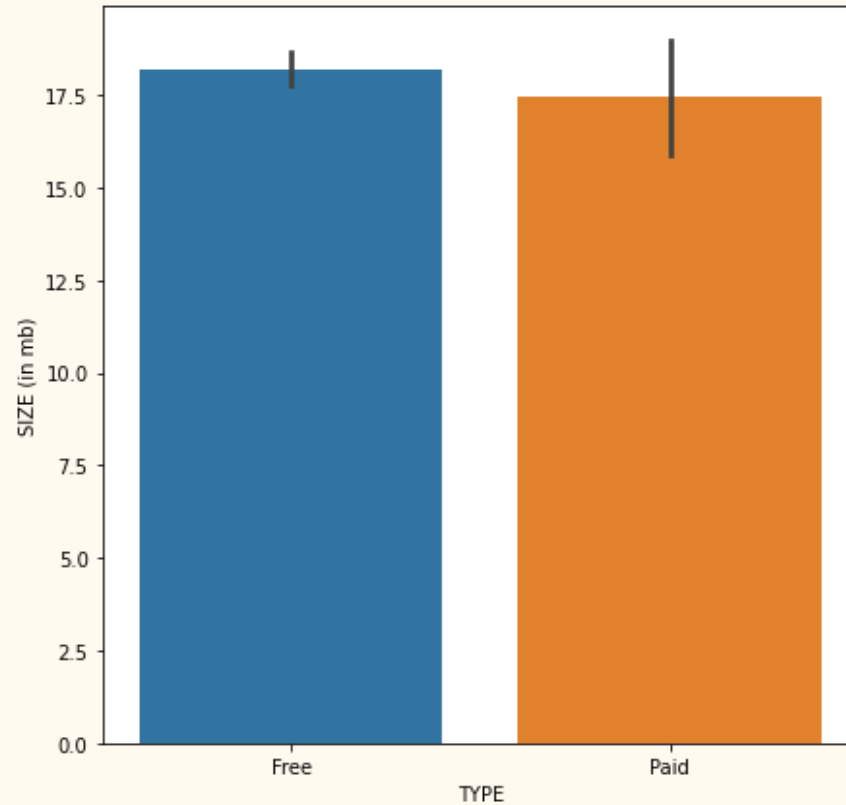
10. AVERAGE PRICE OF APPS IN EACH CATEGORY(IN USD)

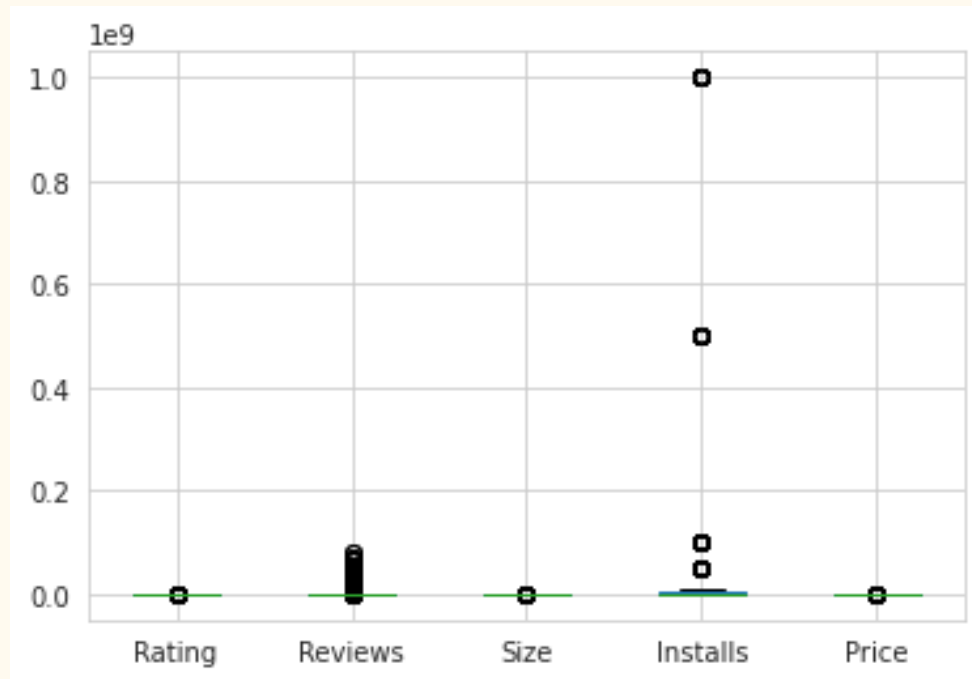


11. TOTAL PRICE IN EACH CATEGORY(SUM)

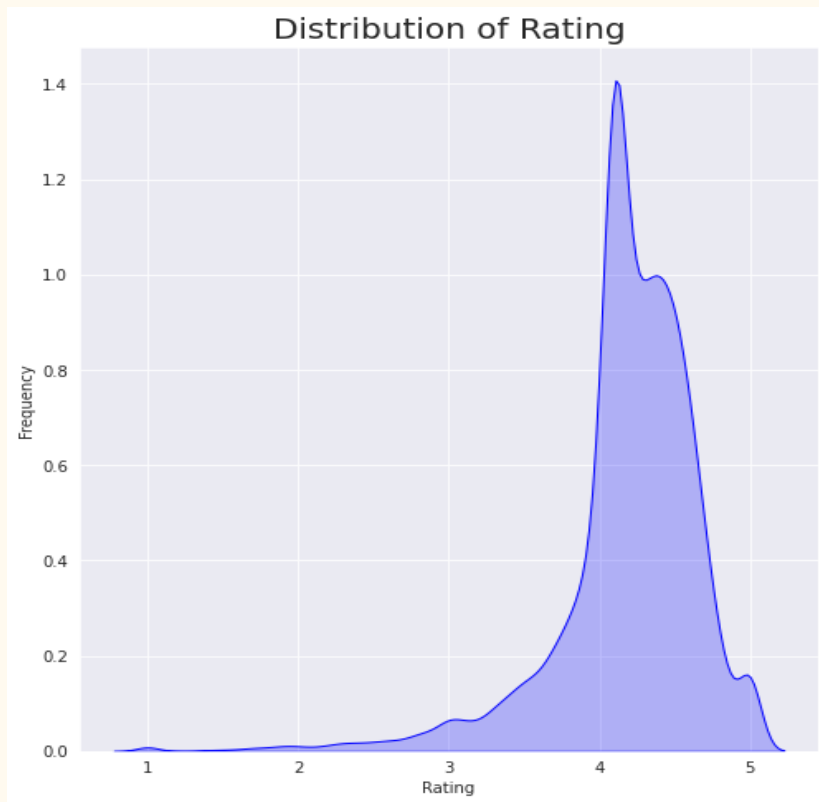


12. WHAT IS THE FREE AND PAID APPS SIZE (IN MB)

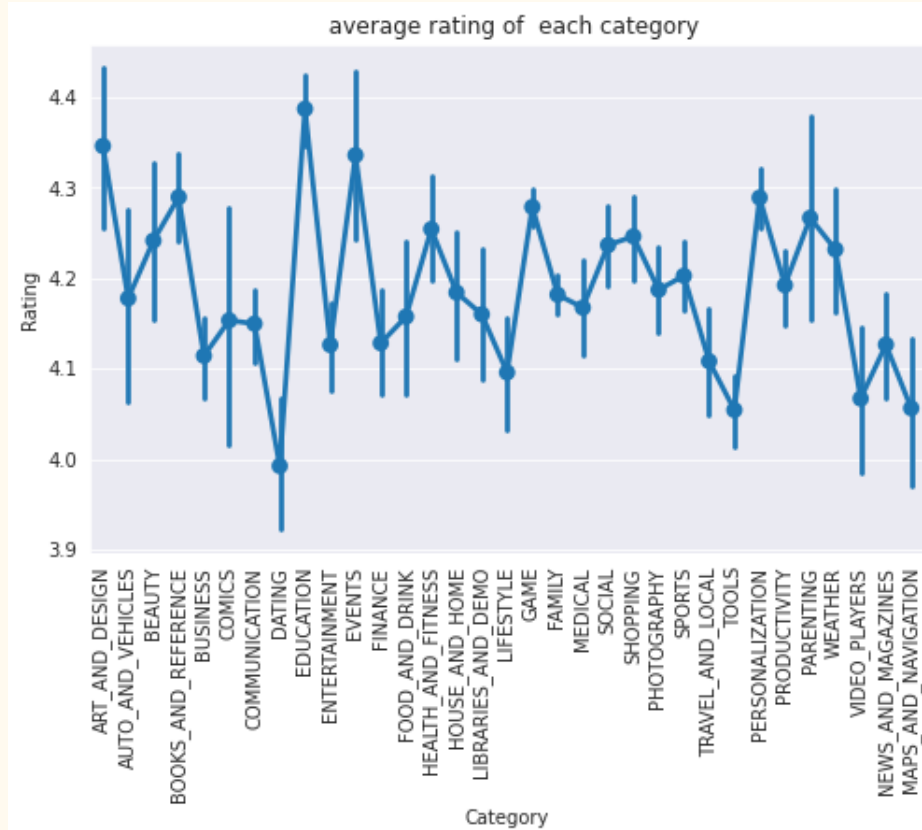




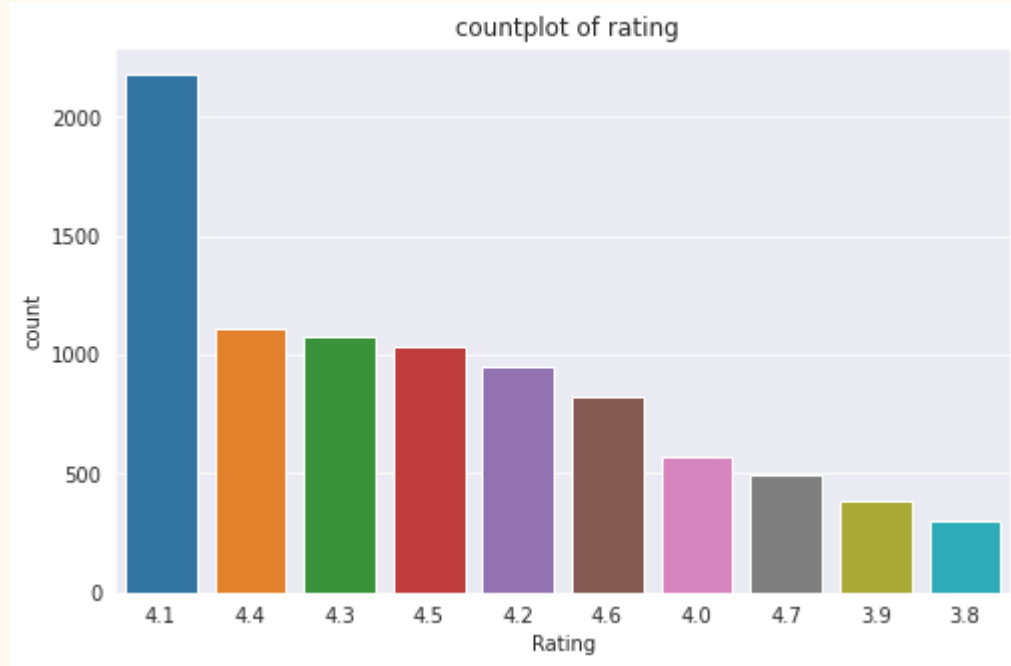
14. DISTRIBUTION OF RATING



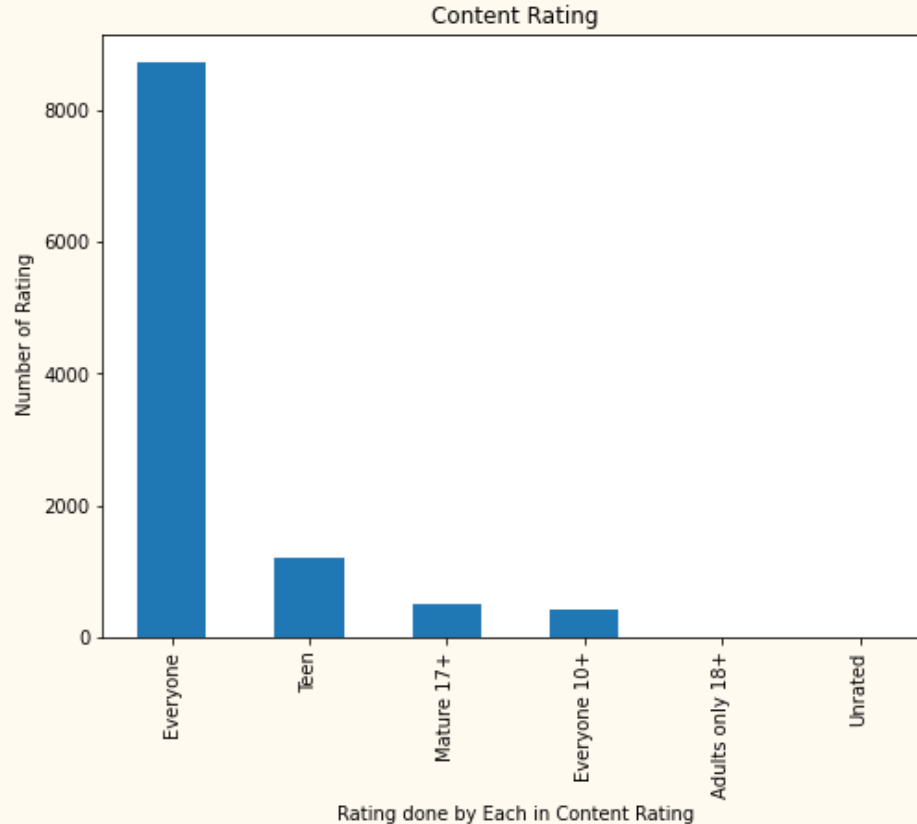
15. POINT PLOT OF AVERAGE RATING IN EACH CATEGORY



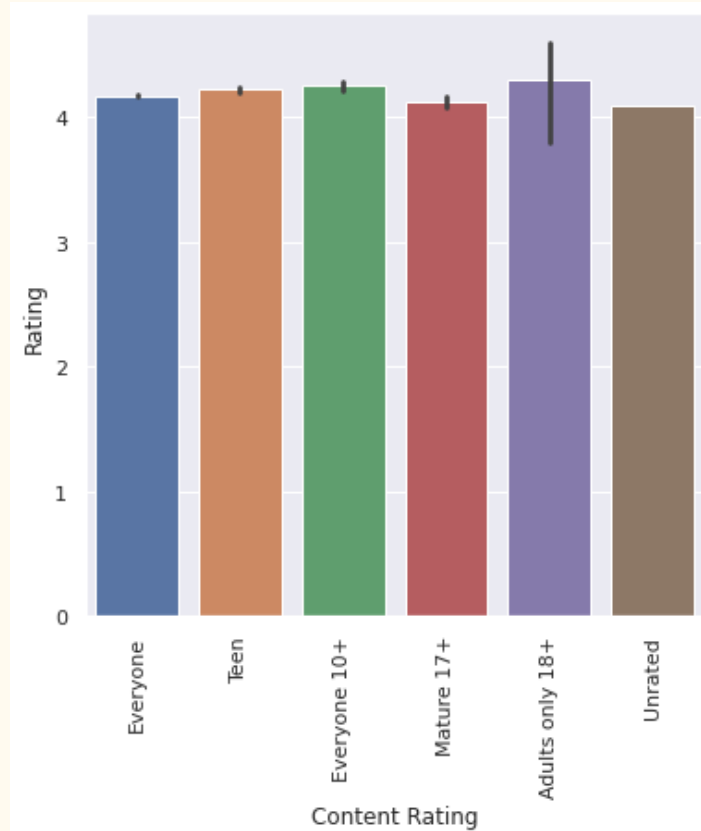
16. COUNTPLOT OF RATING(TOPIO)



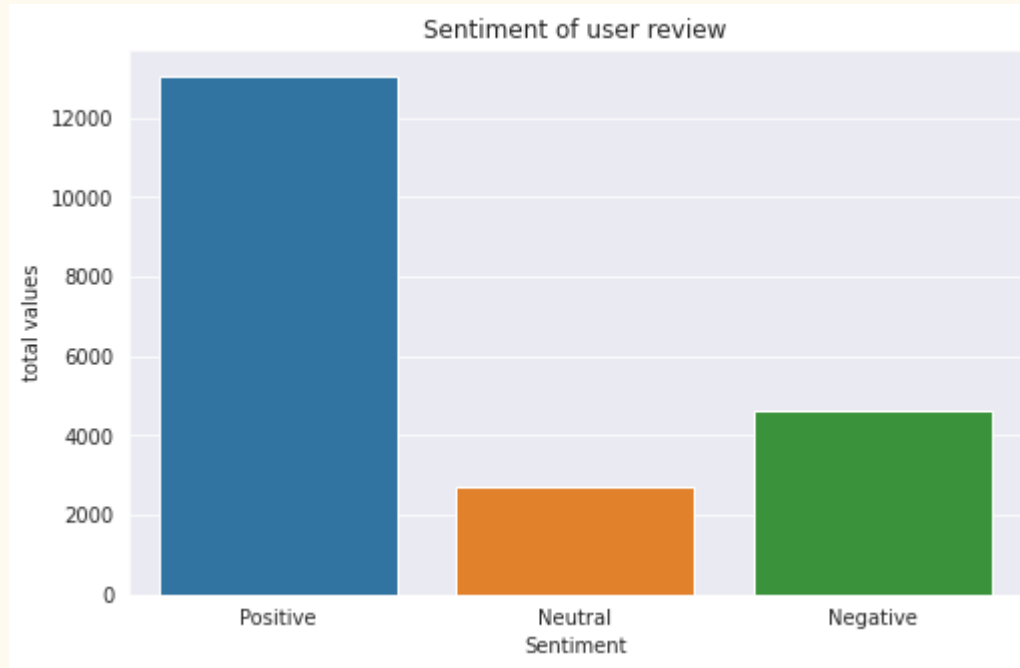
17. PLOT THE BARGRAPH FOR CONTENT RATING



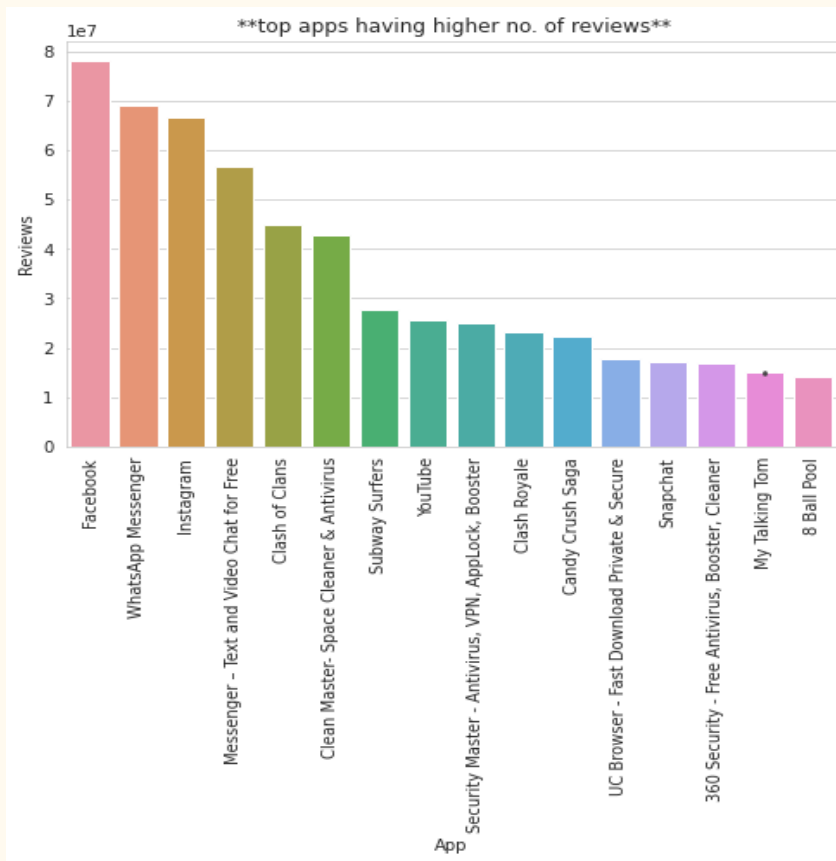
18. BARPLOT DISPLAYING THE RATING FOR EACH CONTENT RATING



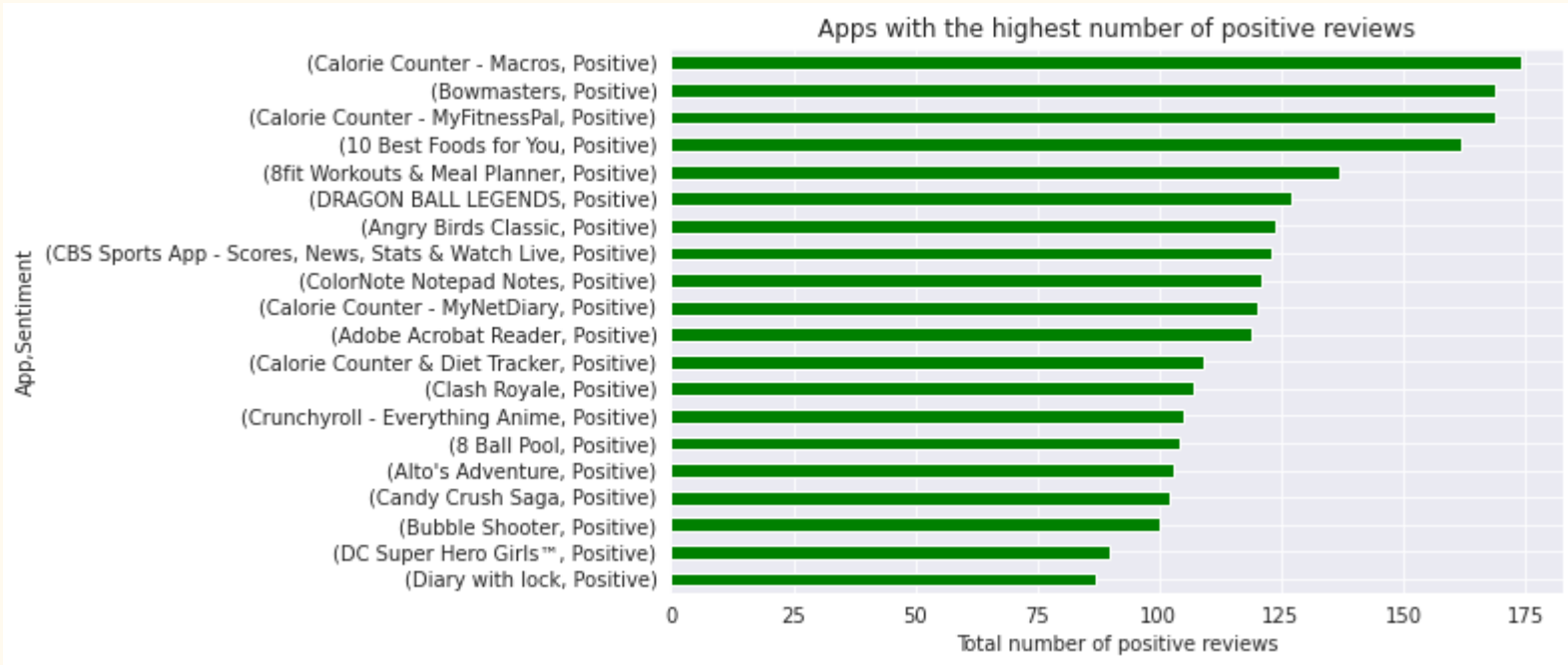
19. PLOTTING THE OVERALL SENTIMENT OF REVIEWS



20. TOP APPS HAVING HIGHER NUMBER OF REVIEWS

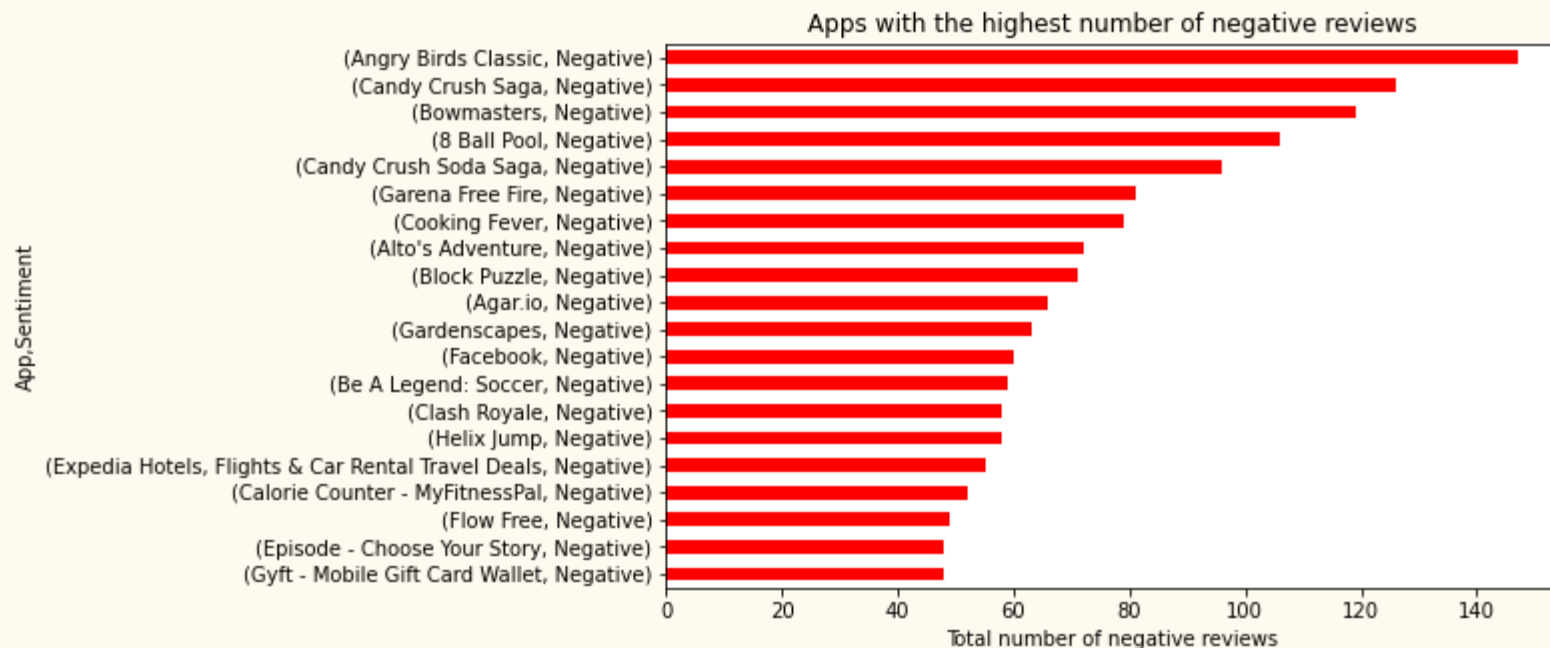


21. PLOTTING OF POSITIVE REVIEWS



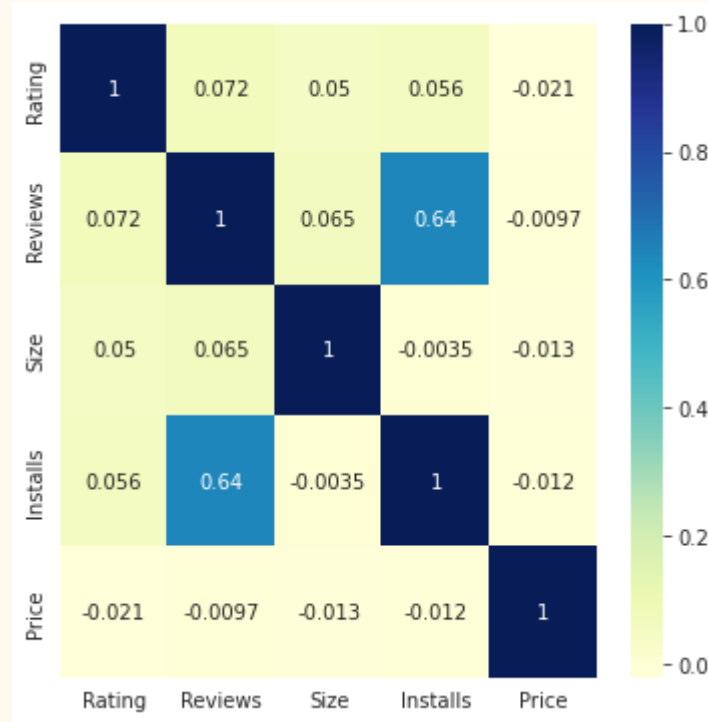


23. PLOTTING OF NEGATIVE REVIEWS





25. CORRELATION BETWEEN RATING,REVIEWS,SIZE AND INSTALLS



Challenges Faced

- ❑ Exploring the dataset and comprehending the problem statements.
- ❑ Examining the business KPIs for app development and devising a solution to the problem.
- ❑ Handling the error, duplicate and NaN values in the dataset.
- ❑ Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.



Conclusion

92.6% apps are **Free** and 7.4% apps are paid in type.

81.80% apps have **Everyone** content rating.

Events category has a **highest mean rating of 4.39** and Dating category has lowest 4.05 rating.

Family, Game and Tools are top three categories having 1906, 926 and 829 app count.

Most competitive category: **Family**

Category with the highest number of installs: **Game**

Tools, Entertainment, Education, Business and Medical are top Genres.

8783 Apps are having size less than or equal to **50 MB**.

7749 Apps has rating **more than 4.0** including both type of app.

Overall sentiment count of merged dataset in which **Positive sentiment count is 64%, Negative 22% and Neutral 14%.**

Conclusion

It's good to develop a **Free type** app and having a content rating for **Everyone**.

Percentage of apps that are top rated = **81.80%**

There are **20** free apps that have been installed over a **billion** times

Minecraft is the only app in the paid category with over **10M** installs, and also has produced the most revenue only from installation fee.

Price, Rating, Size **has no or very less correlation** with **Sentiment Polarity**.

The median size of the apps in the play store is 12 MB

The apps whose size **varies with device** has the highest number average app installs.

The apps whose size is **greater than 90 MB** has the highest number of average user reviews, ie, they are more popular than the rest.

Helix Jump has the highest number of positive reviews and **Angry Birds Classic** has the highest number of negative reviews.

THANK

YOU !

