Capstone Project

Play Store App Review Analysis

By
Adil Imam
Asadullakhan Pathan
Md Sazil Sharif
sushil kumar singh
Madhulika Kumari
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:

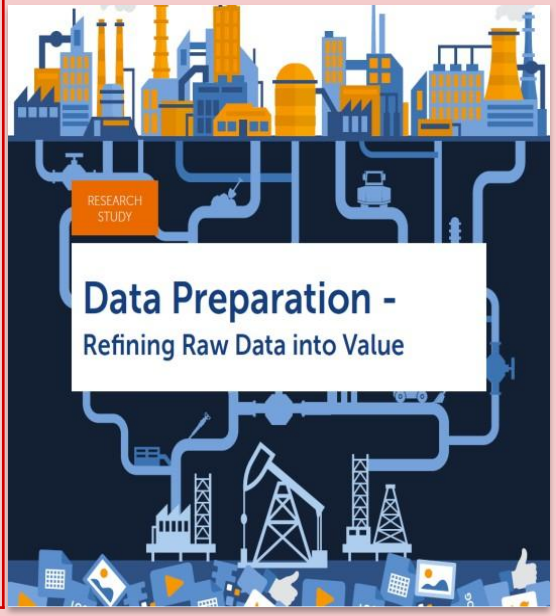
- A high average user rating
- A good number of positive reviews
- A good number of monthly average users
- High revenue per customer and so on.

Problem Statement

- Which are Top categories on Google Play store?
- Which category of Content is found more?
- Distribution of the ratings of the apps?
- What percentage of apps are Free and Paid?
- Which category of apps has the greatest number of installs?
- What are the Top 10 installed apps in different categories?
- Which are the top expensive Apps?
- Which are the Apps with the highest number of reviews?
- Count of Apps found in different genres?
- Which are the apps that have made the highest-earning?

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.



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 i.e. 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
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

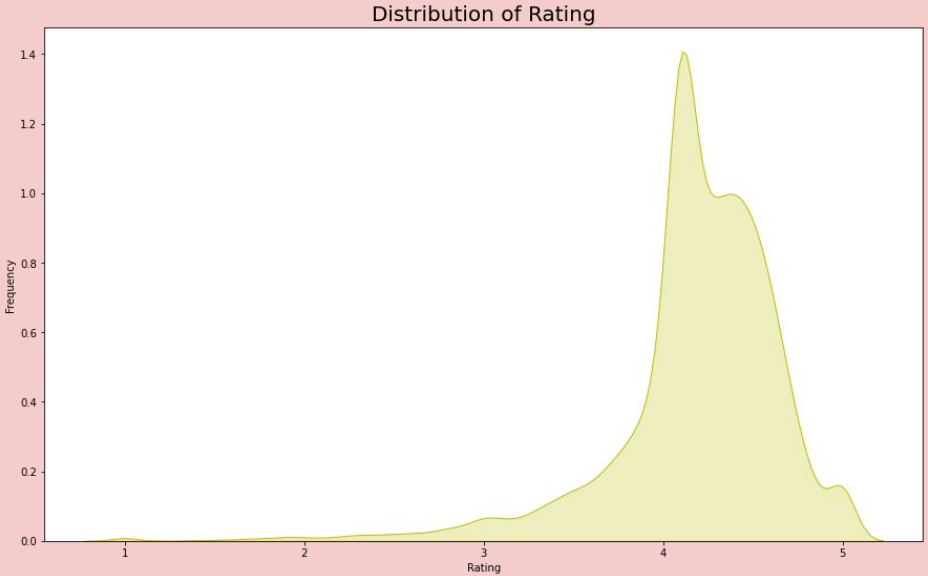
Uncover initial patterns, characteristics, and points of interest using visual exploration

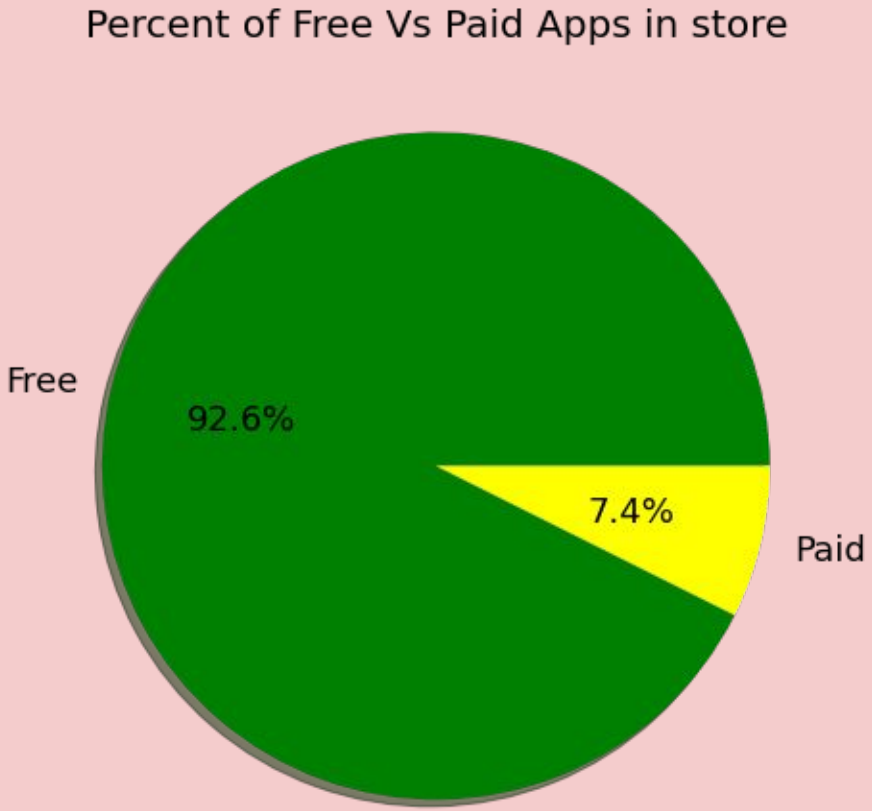
Predictive Modeling

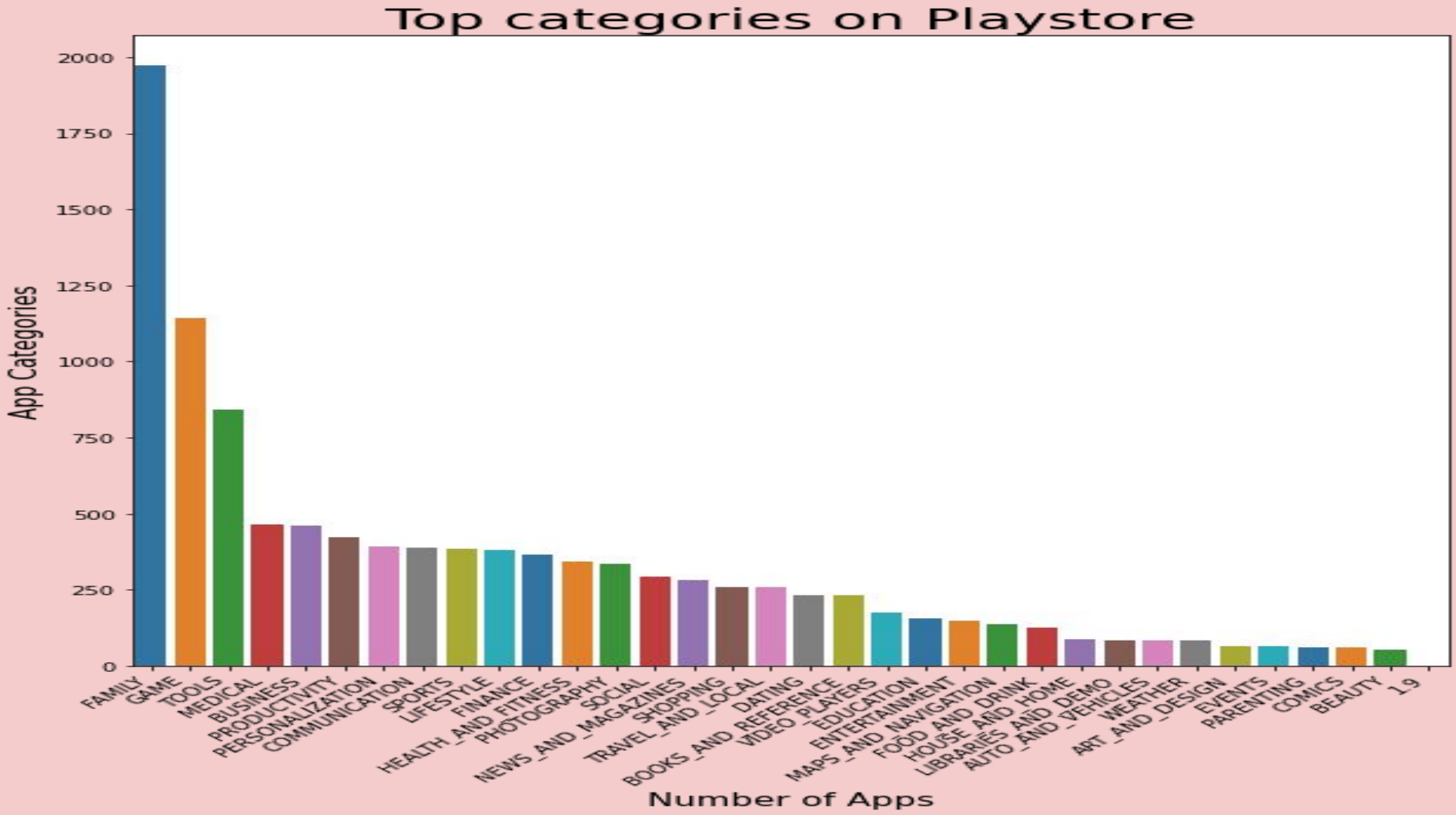
Formulate a statistical model to forecast an outcome using relevant predictors

Percentage of Paid apps v/s Free apps

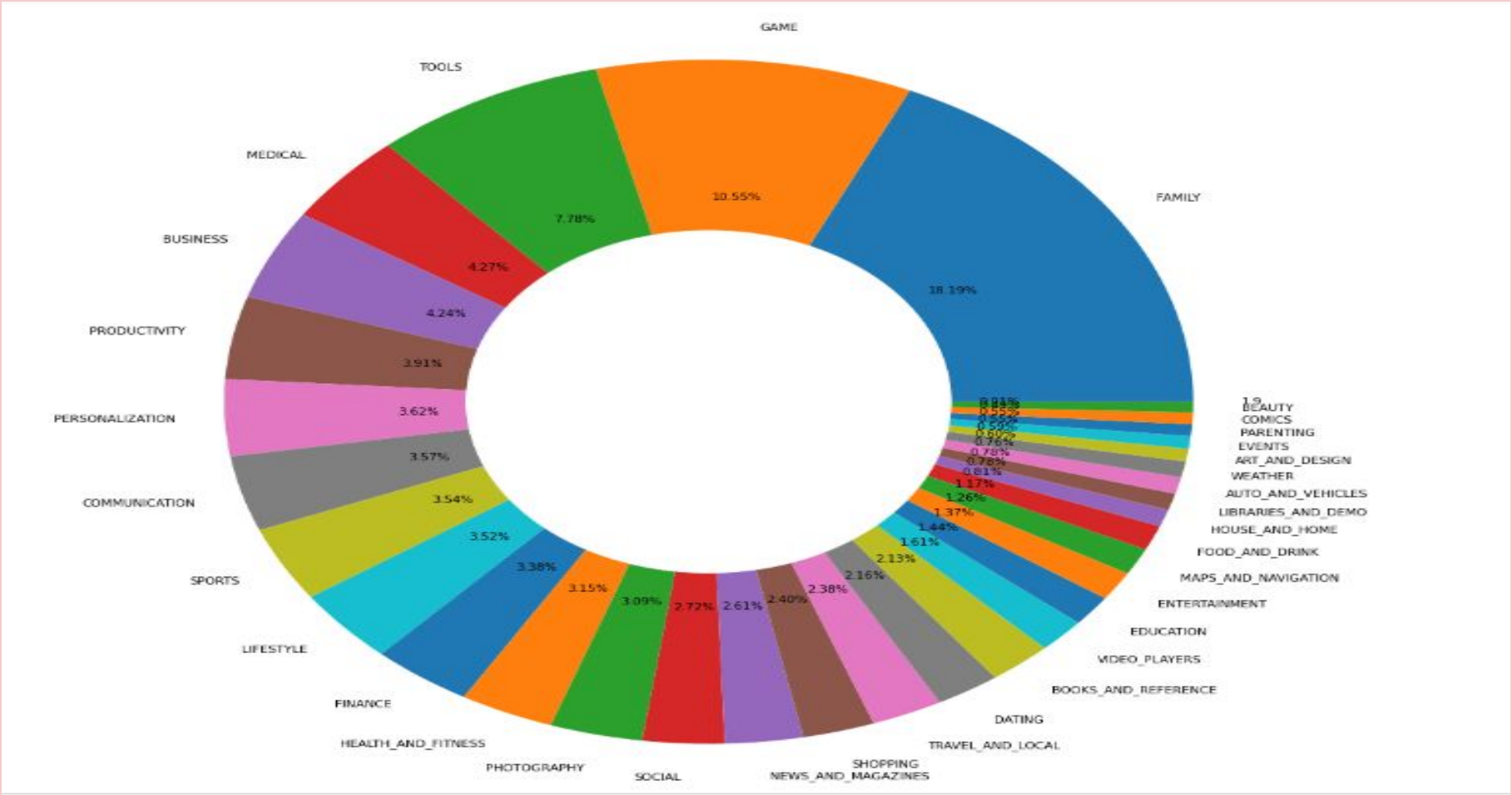
We Observed that **92.20% of Apps are free** and only **7.80% of Apps are paid** in Play store.



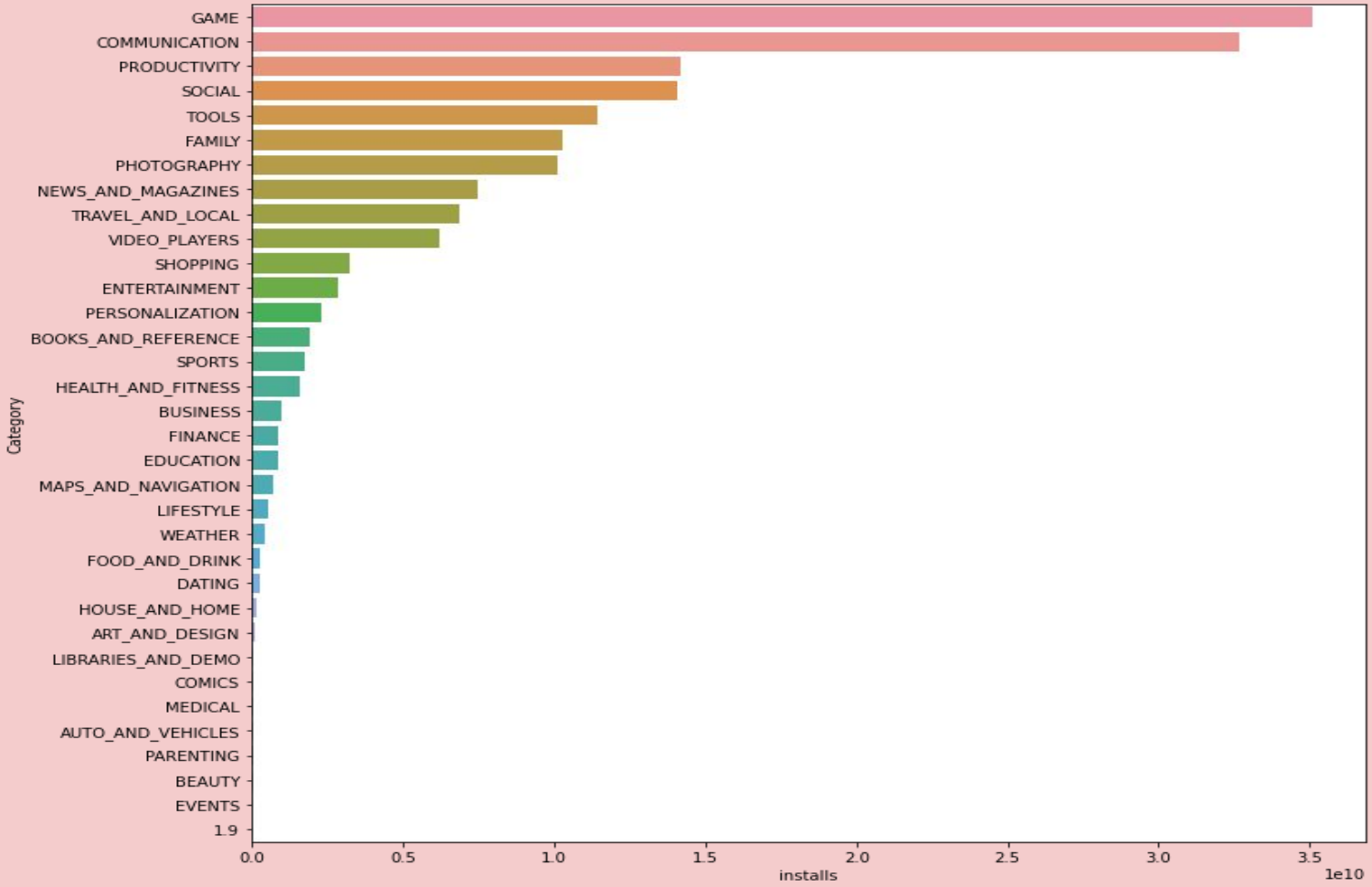


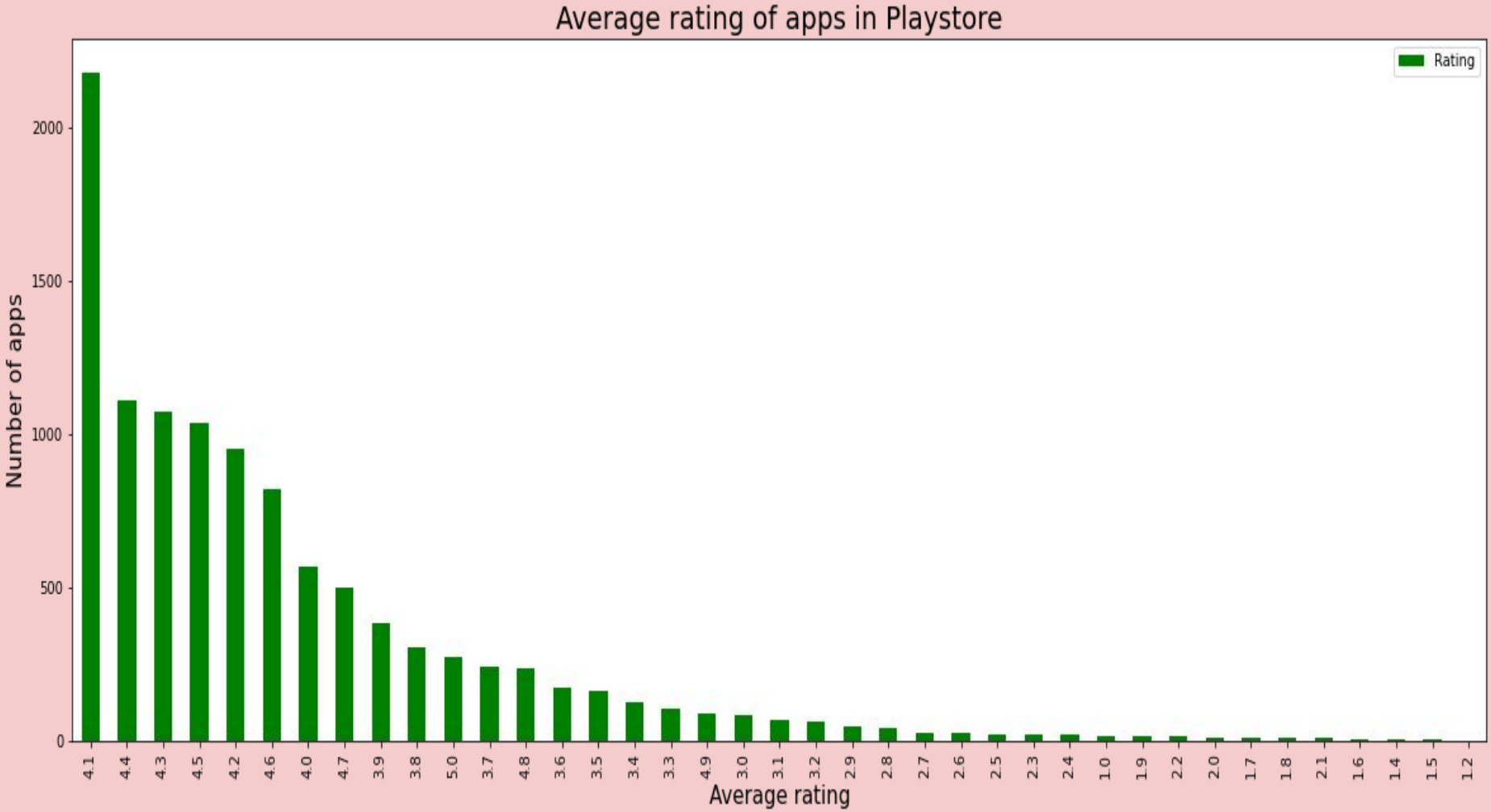

Top categories on Playstore
App Categories
Number of Apps
2000
1750
1500
1250
1000
750
500
250
0
FAMILY
GAME
TOOLS
MEDICAL
BUSINESS
PRODUCTIVITY
PERSONALIZATION
COMMUNICATION
SPORTS
LIFESTYLE
FINANCE
HEALTH_AND_FITNESS
PHOTOGRAPHY
SOCIAL
NEWS_AND_MAGAZINES
SHOPPING
TRAVEL_AND_LOCAL
DATING
BOOKS_AND_REFERENCE
VIDEO_PLAYERS
EDUCATION
ENTERTAINMENT
MAPS_AND_NAVIGATION
FOOD_AND_DRINK
HOUSE_AND_HOME
LIBRARIES_AND_DEMO
AUTO_AND_VEHICLES
WEATHER
ART_AND_DESIGN
EVENTS
PARENTING
COMICS
BEAUTY
1.9

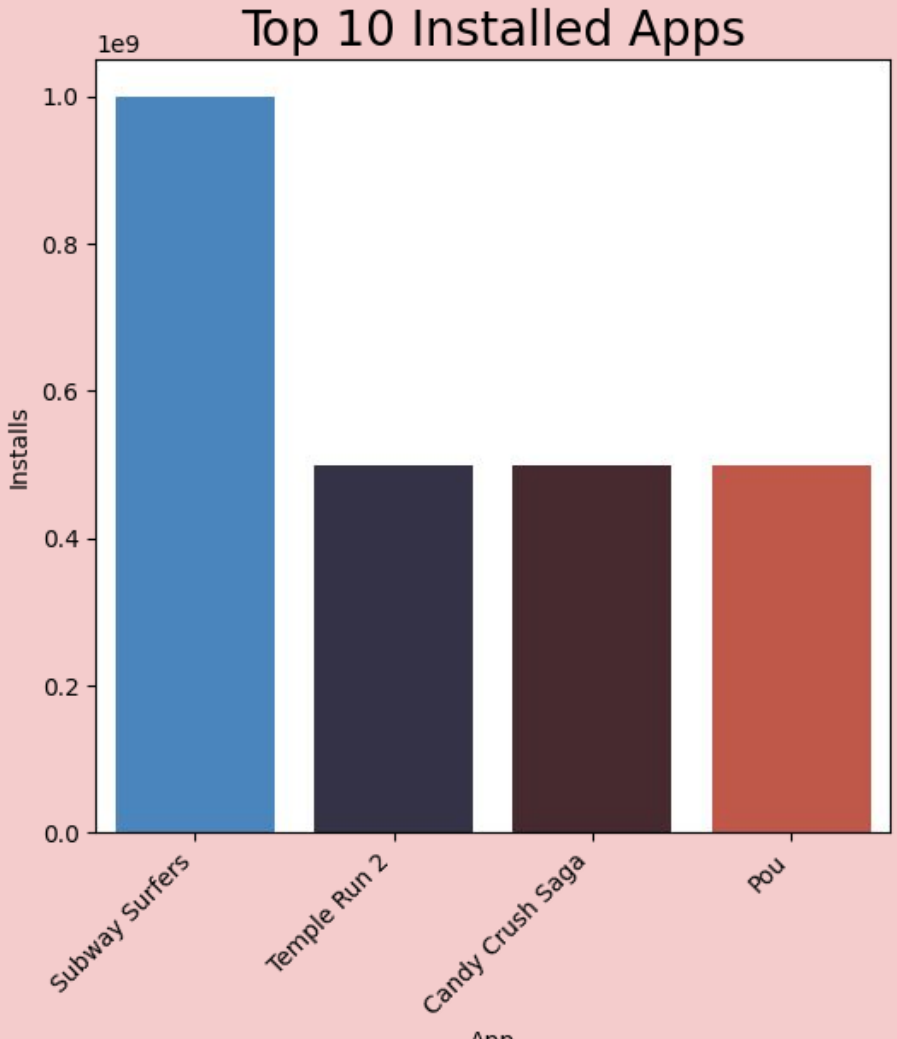
Percentage of apps belonging to each category in the playstore

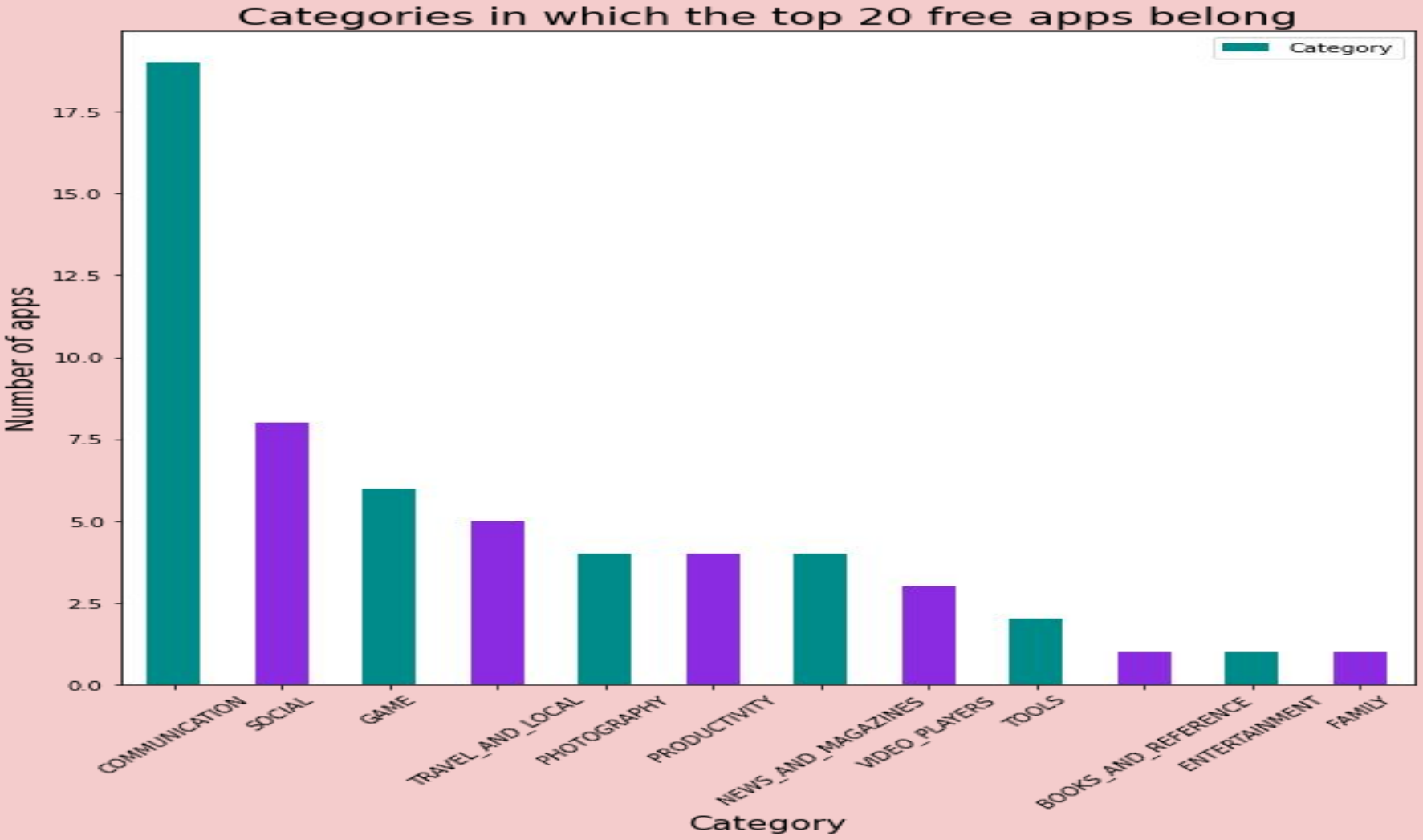


category Apps which are having most number of installs

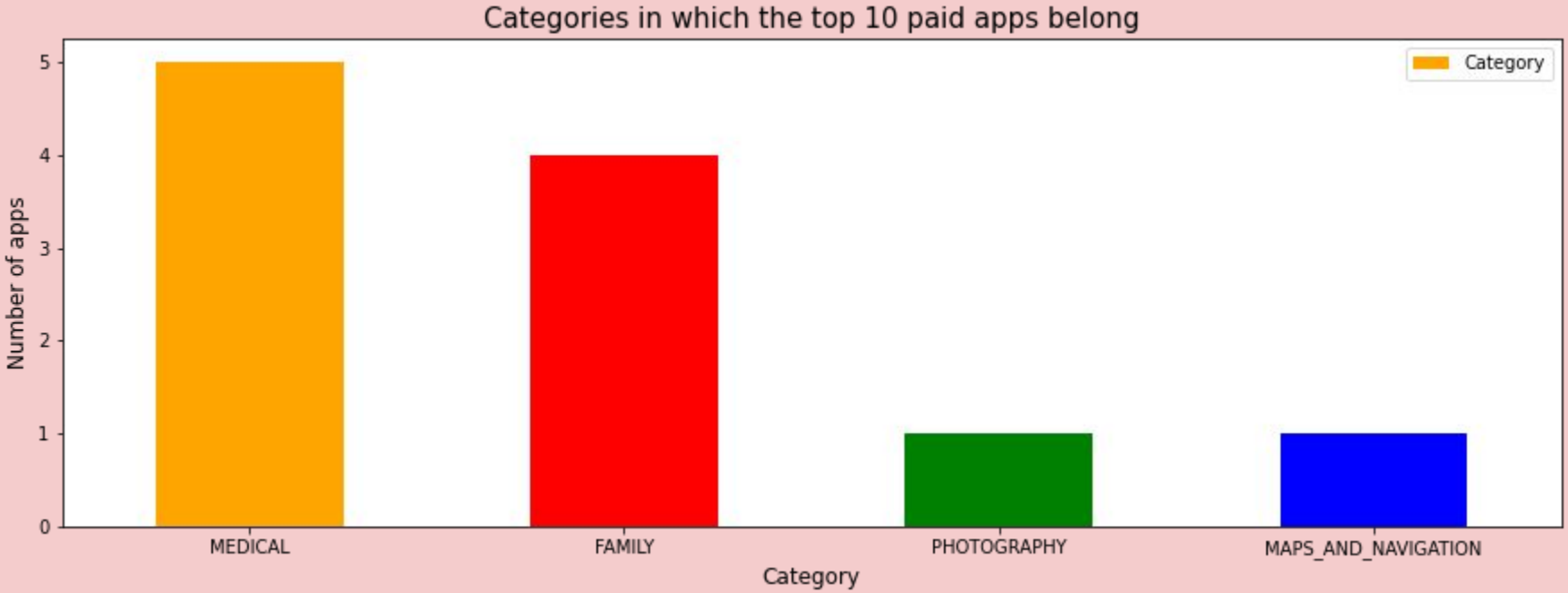


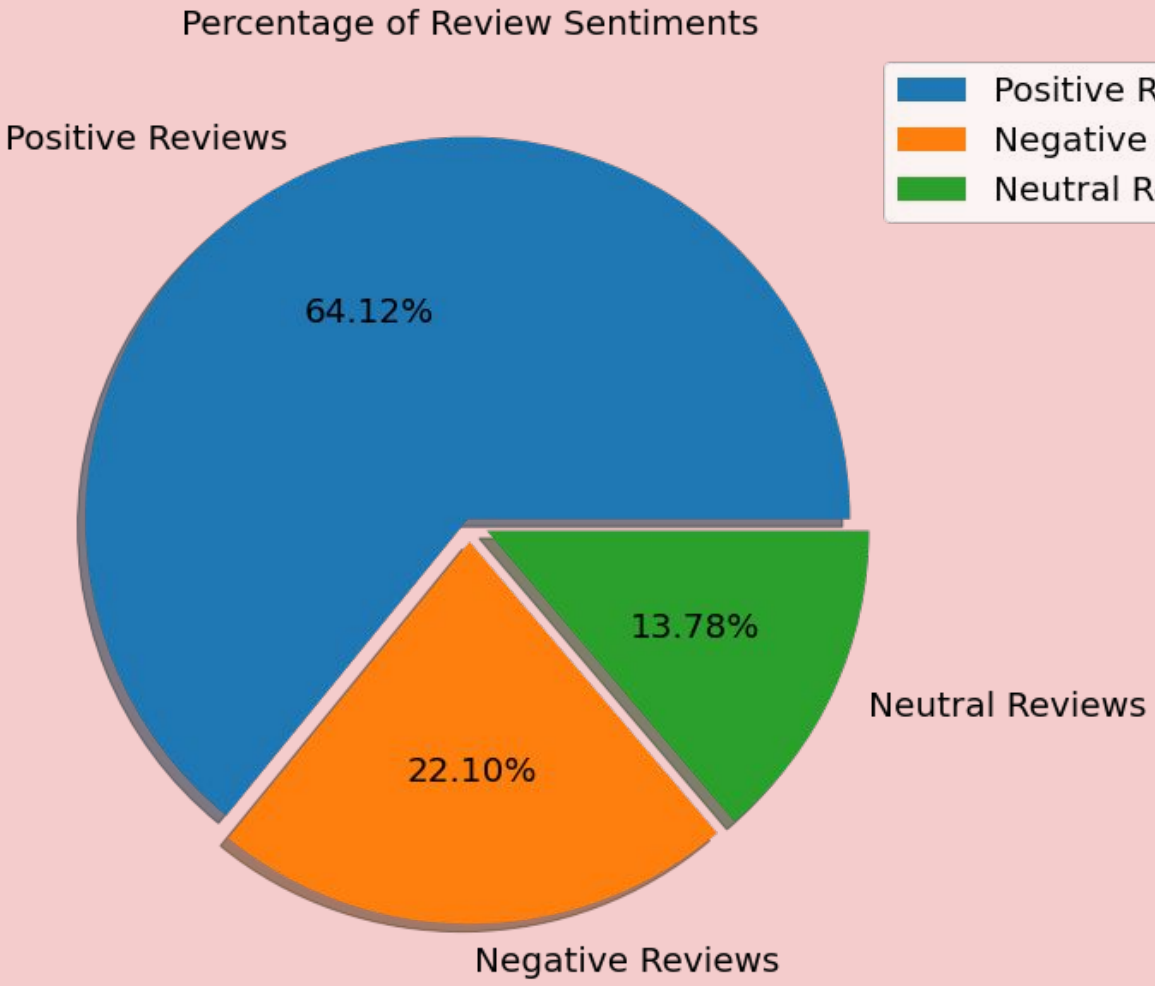

Average rating of apps in Playstore
Rating
Number of apps
2000
1500
1000
500
0
4.1
4.4
4.3
4.5
4.2
4.6
4.0
4.7
3.9
3.8
5.0
3.7
4.8
3.6
3.5
3.4
3.3
4.9
3.0
3.1
3.2
2.9
2.8
2.7
2.6
2.5
2.3
2.4
1.0
1.9
2.2
2.0
1.7
1.8
2.1
1.6
1.4
1.5
1.2
Average rating


Top 10 Installed Apps
1e9
1.0
0.8
0.6
0.4
0.2
0.0
Installs
Subway Surfers
Temple Run 2
Candy Crush Saga
Pou


Categories in which the top 20 free apps belong
Category
Number of apps
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
COMMUNICATION
SOCIAL
GAME
TRAVEL_AND_LOCAL
PHOTOGRAPHY
PRODUCTIVITY
NEWS_AND_MAGAZINES
VIDEO_PLAYERS
TOOLS
BOOKS_AND_REFERENCE
ENTERTAINMENT
FAMILY
Category

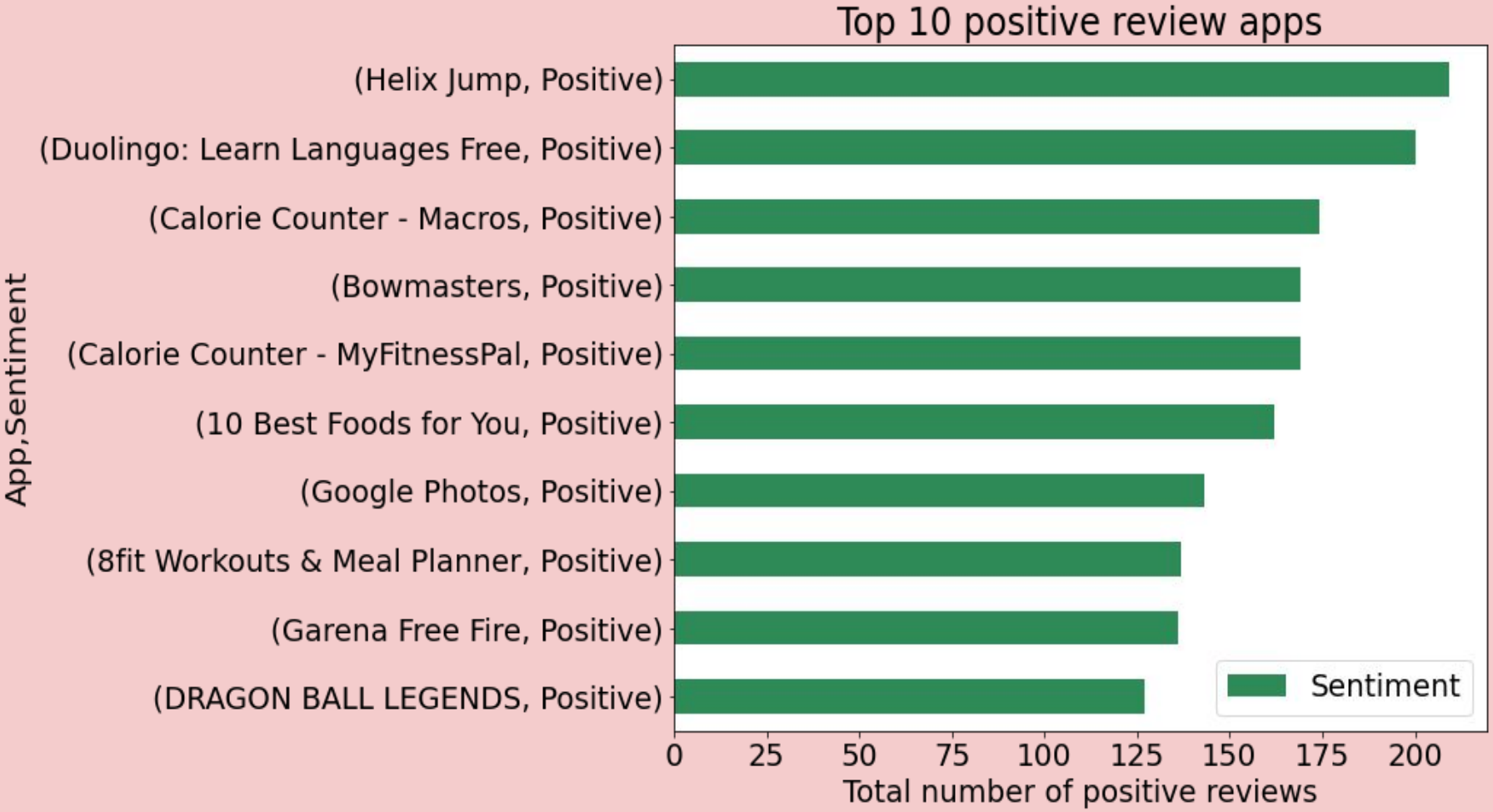
Category in which the top 10 paid app belong

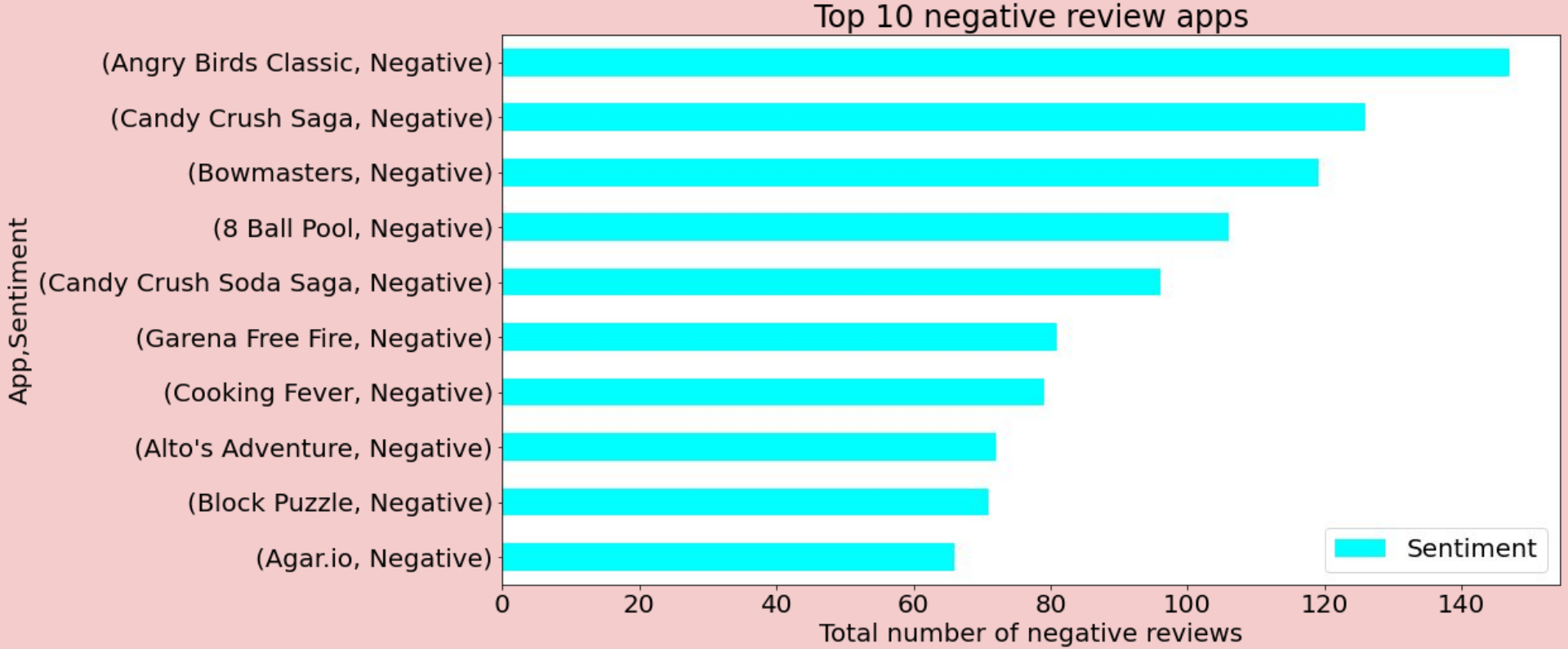


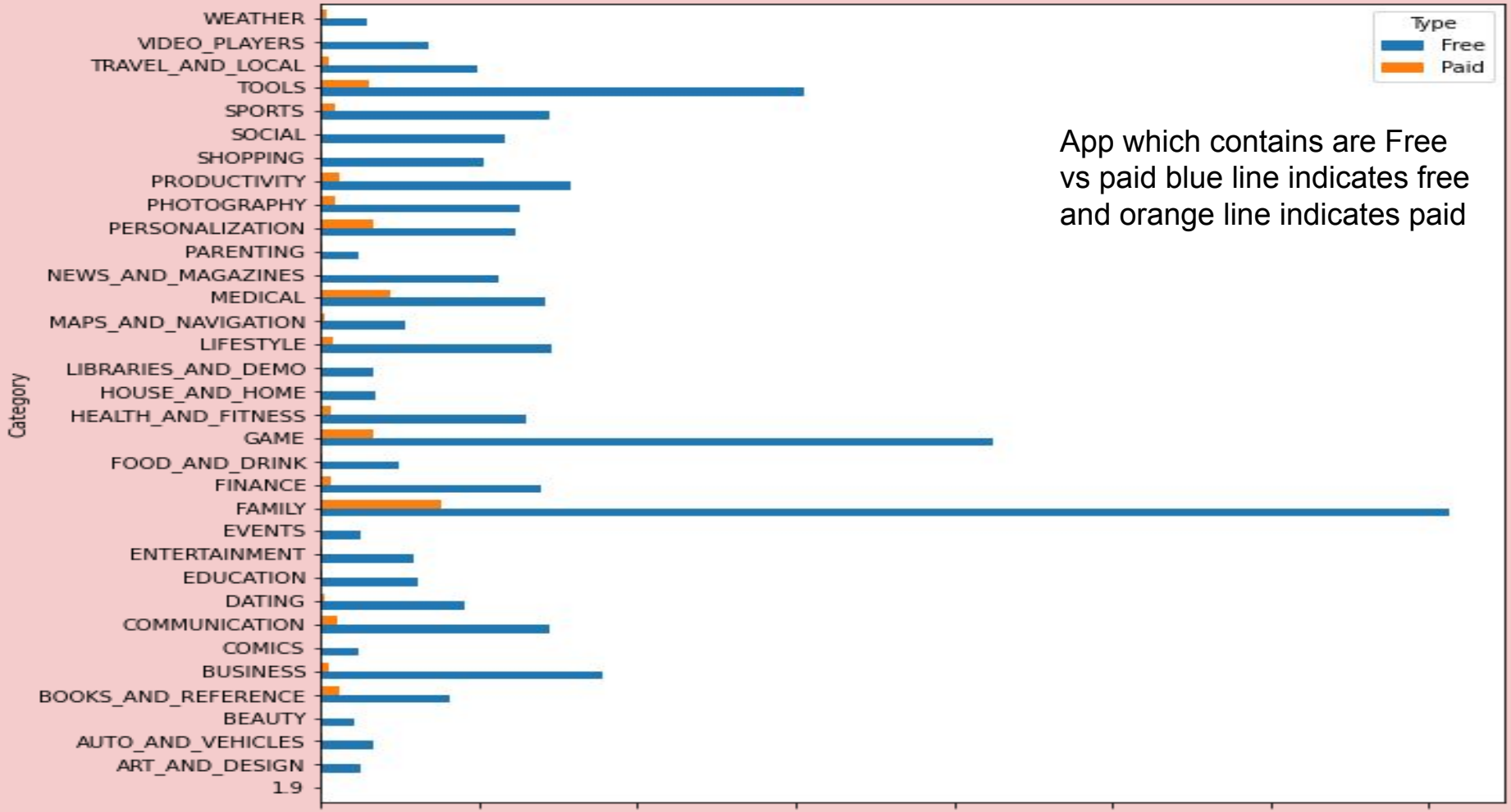


The number of **Unique** Apps from Play store and User reviews merged dataset are **816.**

From Sentiment column, **64%** are **Positive, 22%** are **Negative** and **14%** are **Neutral** values.


Top 10 positive review apps
(Helix Jump, Positive)
(Duolingo: Learn Languages Free, Positive)
(Calorie Counter - Macros, Positive)
(Bowmasters, Positive)
(Calorie Counter - MyFitnessPal, Positive)
(10 Best Foods for You, Positive)
(Google Photos, Positive)
(8fit Workouts & Meal Planner, Positive)
(Garena Free Fire, Positive)
(DRAGON BALL LEGENDS, Positive)
App,Sentiment
Sentiment
0
25
50
75
100
125
150
175
200
Total number of positive reviews


Top 10 negative review apps
(Angry Birds Classic, Negative)
(Candy Crush Saga, Negative)
(Bowmasters, Negative)
(8 Ball Pool, Negative)
(Candy Crush Soda Saga, Negative)
(Garena Free Fire, Negative)
(Cooking Fever, Negative)
(Alto's Adventure, Negative)
(Block Puzzle, Negative)
(Agar.io, Negative)
App,Sentiment
0
20
40
60
80
100
120
140
Total number of negative reviews
Sentiment


Type
Free
Paid
Category
WEATHER
VIDEO_PLAYERS
TRAVEL_AND_LOCAL
TOOLS
SPORTS
SOCIAL
SHOPPING
PRODUCTIVITY
PHOTOGRAPHY
PERSONALIZATION
PARENTING
NEWS_AND_MAGAZINES
MEDICAL
MAPS_AND_NAVIGATION
LIFESTYLE
LIBRARIES_AND_DEMO
HOUSE_AND_HOME
HEALTH_AND_FITNESS
GAME
FOOD_AND_DRINK
FINANCE
FAMILY
EVENTS
ENTERTAINMENT
EDUCATION
DATING
COMMUNICATION
COMICS
BUSINESS
BOOKS_AND_REFERENCE
BEAUTY
AUTO_AND_VEHICLES
ART_AND_DESIGN
1.9
0
250
500
750
1000
1250
1500
1750
App which contains are Free vs paid blue line indicates free and orange line indicates paid

Challenges Faced

- Reading the dataset and comprehending the problem statement.
- 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’s

92.19% apps are **Free** and 7.81% 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%.**

THANK
YOU