

Capstone Project-1

EDA on Play Store App Reviews

By

USHA RANI BOLIMERA

Data Science Trainees, AlmaBetter

Points of Discussion

- Problem Statement
- Data summary
- Exploratory Data Analysis
- Correlation heatmap
- Type and Content Rating Analysis
- Categorical Analysis
- App Rating Analysis
- Top Free and Paid Apps
- Average Price of Paid Apps in Each Category
- Most Popular Apps
- App Size Analysis
- App Reviews Analysis
- Analysis Summary

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.

Data Summary

Play_Store_Data

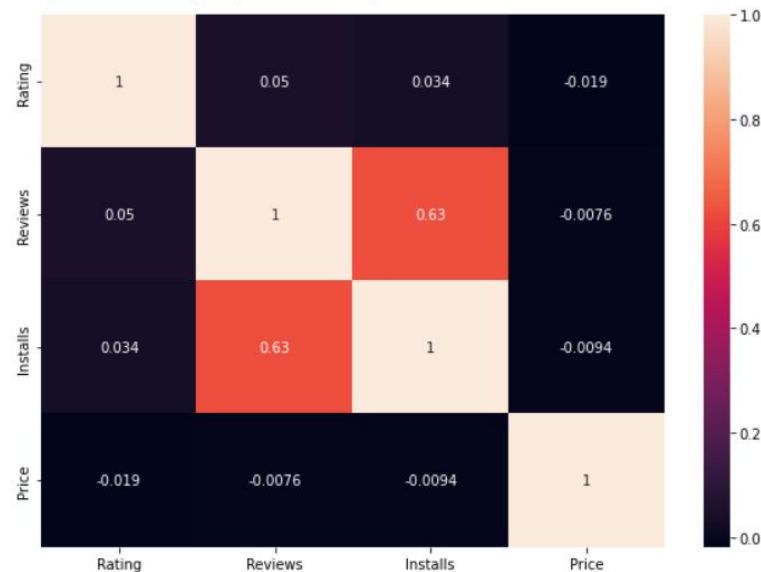
- App
- Category
- Size
- Rating
- Reviews
- Installs
- Type
- Price
- Content Rating
- Genres
- Last Updated
- Current Ver
- Android Ver

User_reviews

- App
- Translated
- Review
- Sentiment
- Sentiment_Polarity
- Sentiment_Subjectivity

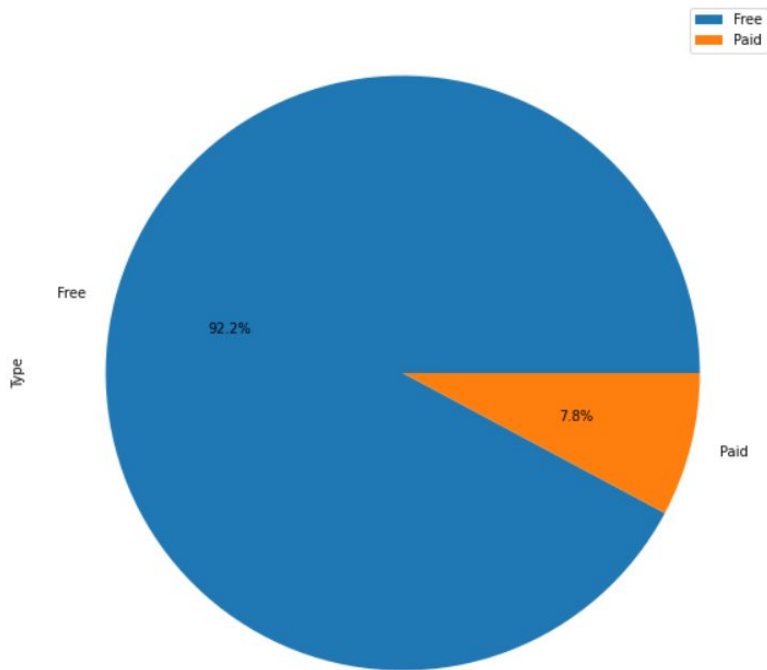
Correlation Heatmap

- There is a strong **positive** correlation between the **Reviews** and **Installs**.
- The Price is slightly **negatively** correlated with the **Rating**, **Reviews**, and **Installs**.
- The Rating is slightly positively correlated with the Installs and Reviews

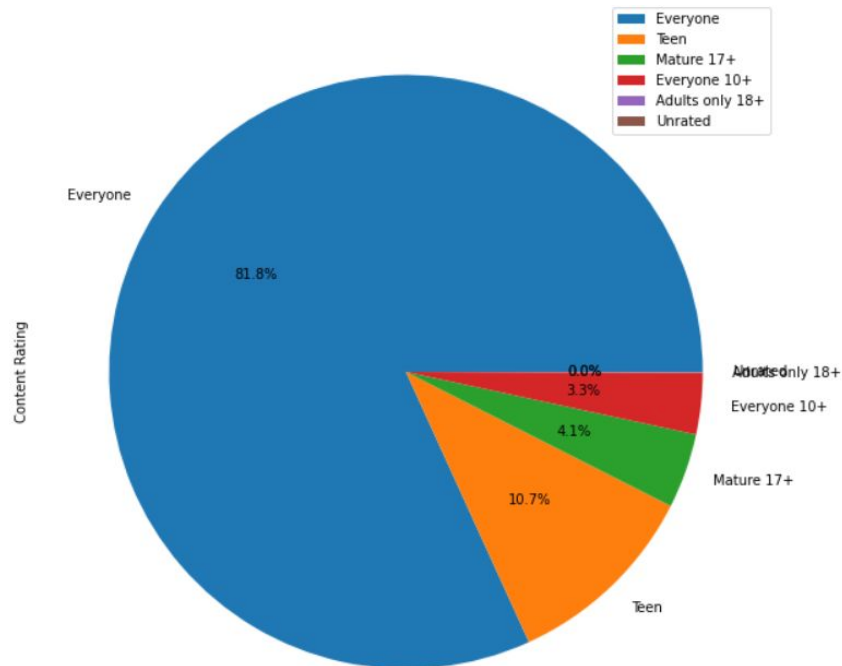


Type and Content Rating

Free and paid apps in the df

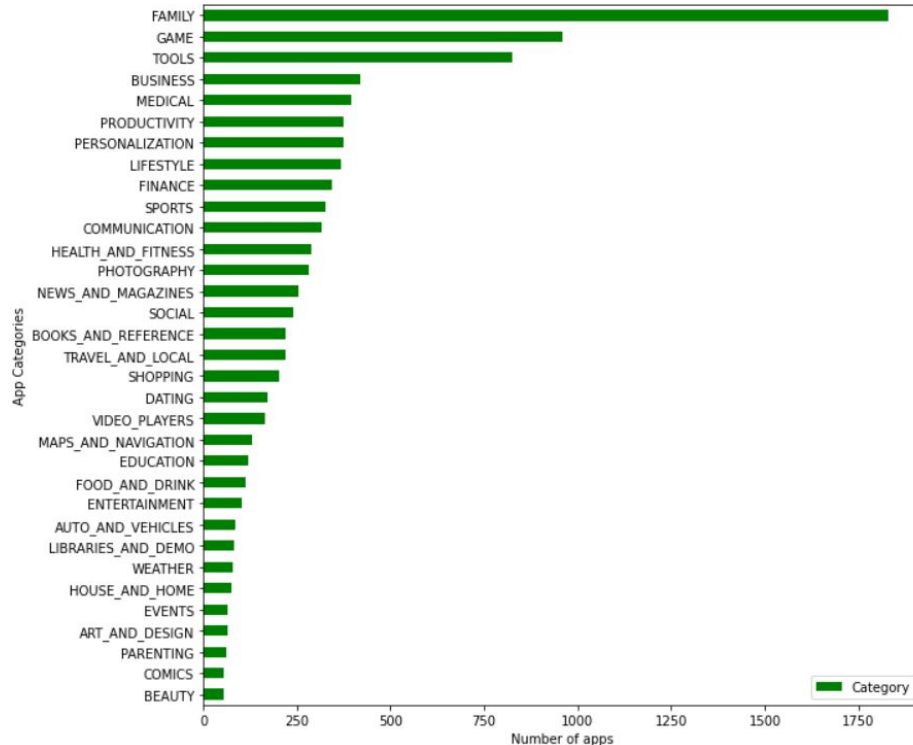


Content rating types in the df

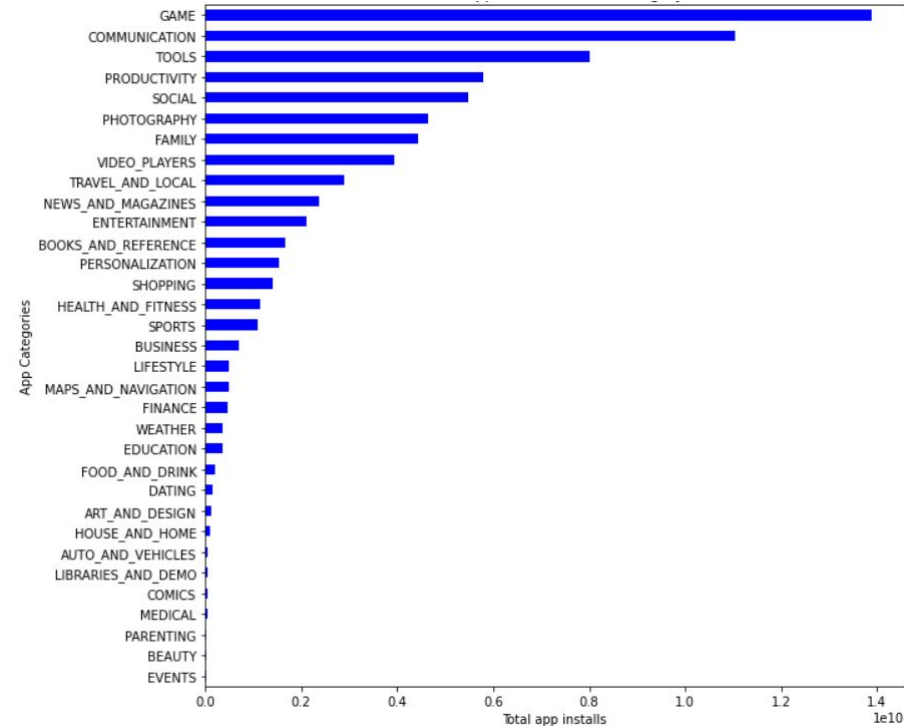


Categorical Analysis

Category vs No. of Apps



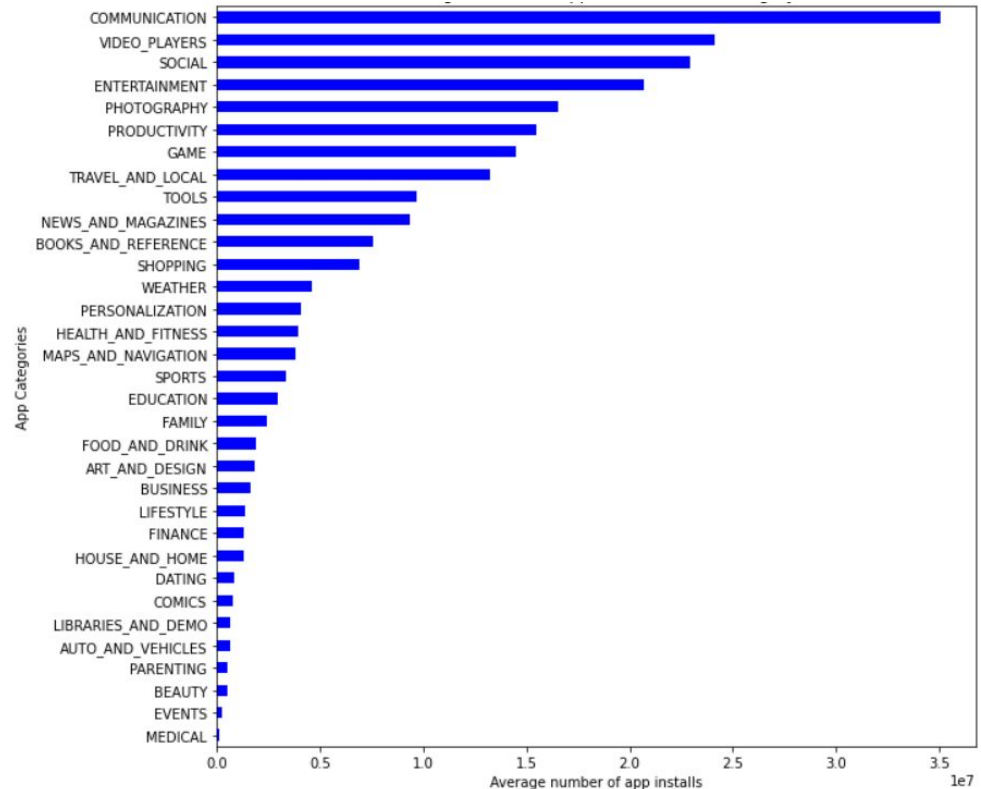
Category vs Total App Installs



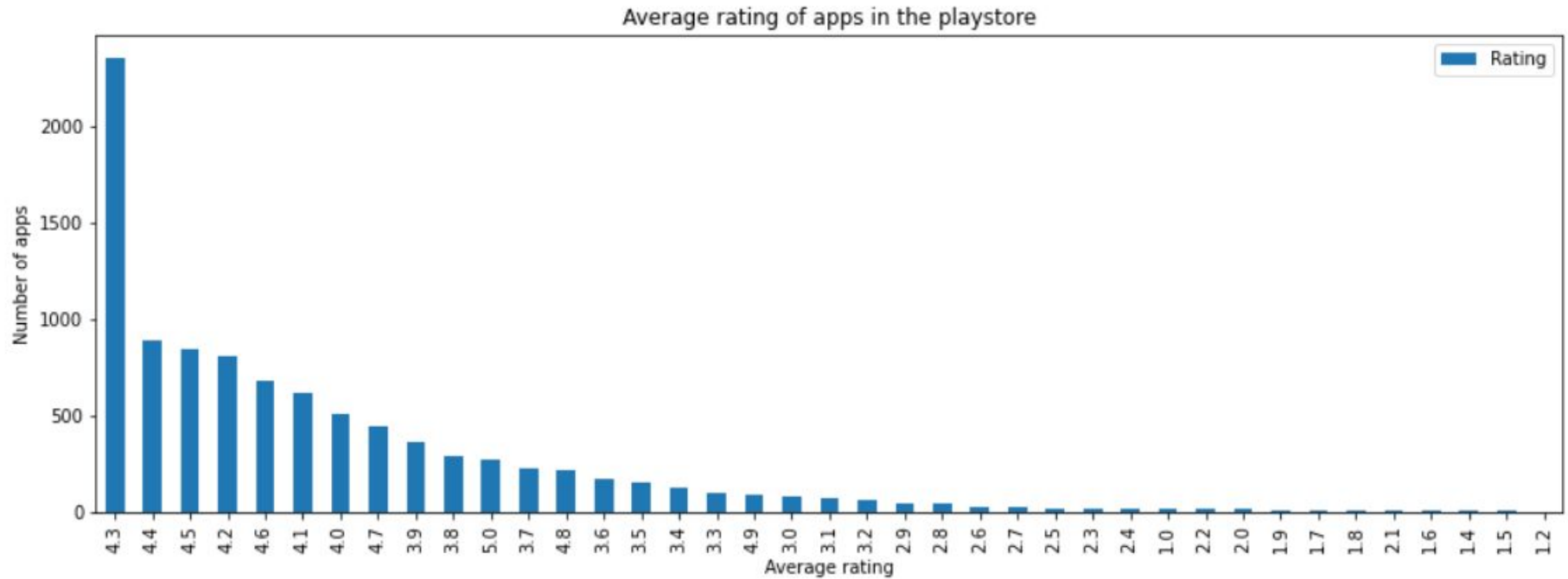
Categorical Analysis (Contd.)

- The **Family, Game, and Tools** category has the highest number of apps.
- The **Game, Communication, and Tools** category has the highest number of app installs.
- The **Communication, Video Players, and Social** category has the highest number of average app installs

Average App Installs in Each Category



Average rating of the apps

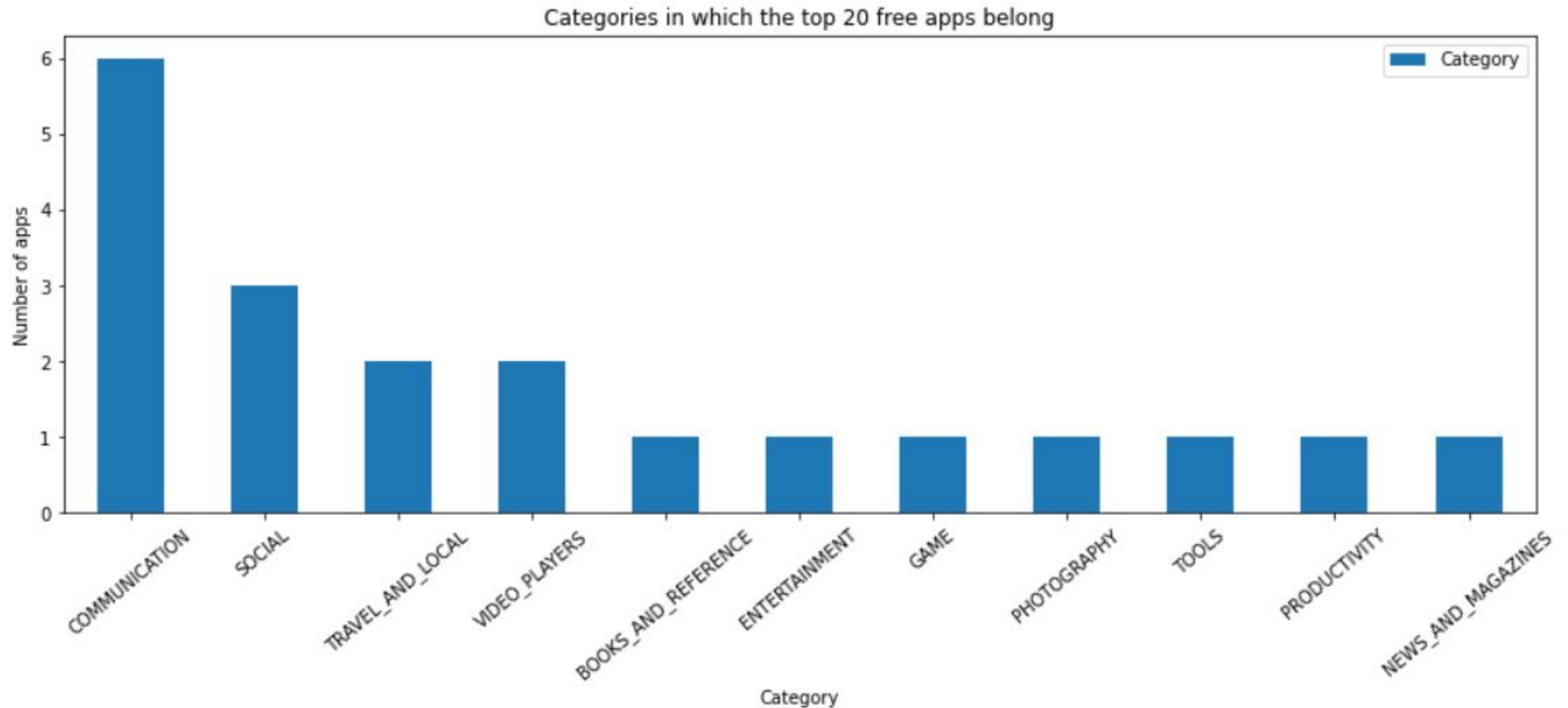


Top 10 Free Apps

- There are a total of 10 free apps with over one billion installs.
- The top categories in which these apps fall are Communication(6), Social(1), Video Players(2), Travel and Local(2).

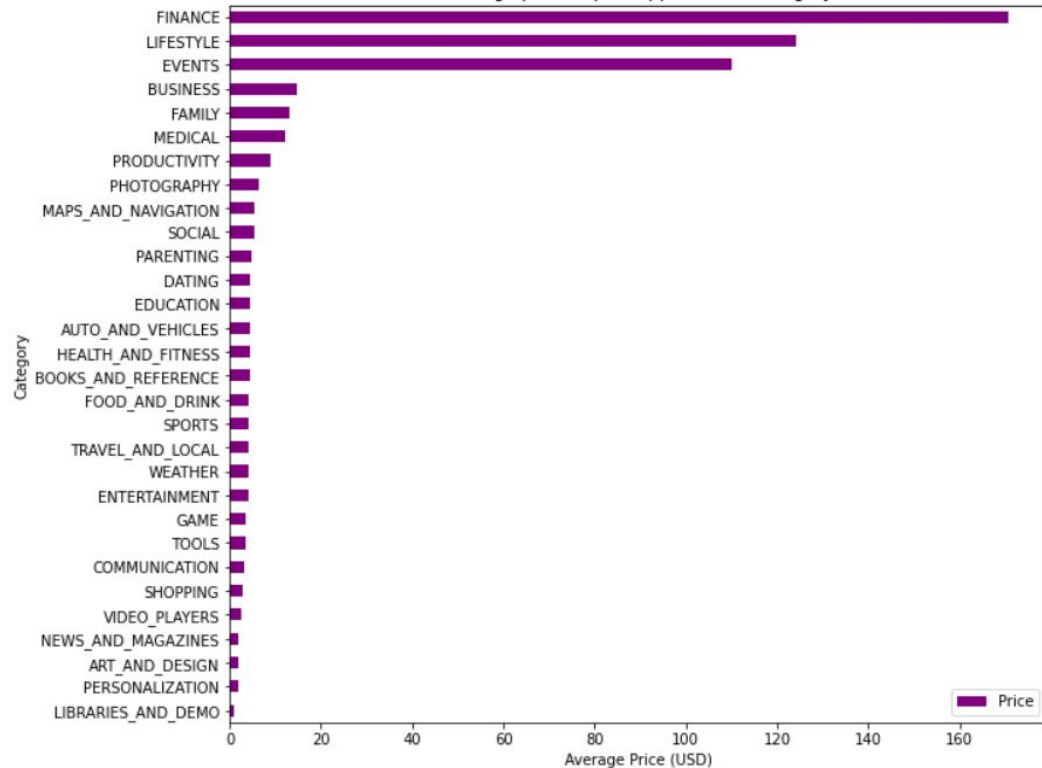
App	Category
Google Play Books	BOOKS_AND_REFERENCE
Messenger – Text and Video Chat for Free	COMMUNICATION
WhatsApp Messenger	COMMUNICATION
Google Chrome: Fast & Secure	COMMUNICATION
Gmail	COMMUNICATION
Hangouts	COMMUNICATION
Skype - free IM & video calls	COMMUNICATION
Google Play Games	ENTERTAINMENT
Subway Surfers	GAME
Facebook	SOCIAL

Top 20 Free Apps Categories



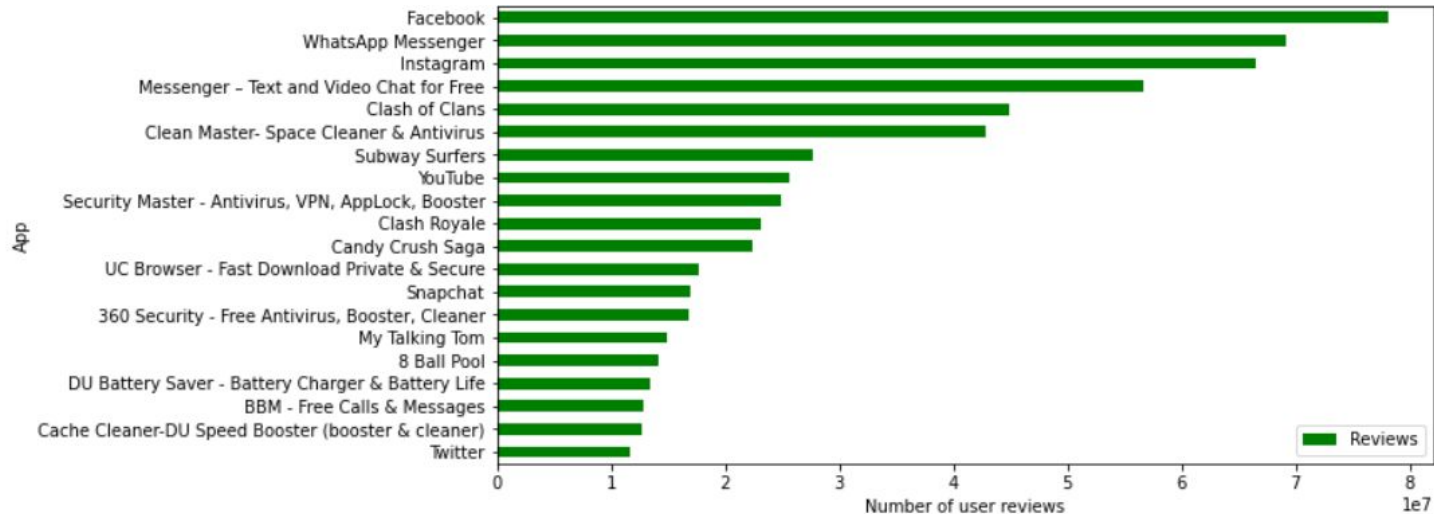
Average Price of Paid Apps in Each Category

- The paid apps in the Finance, Lifestyle, and Events category are on average significantly more expensive than the paid apps in other categories.



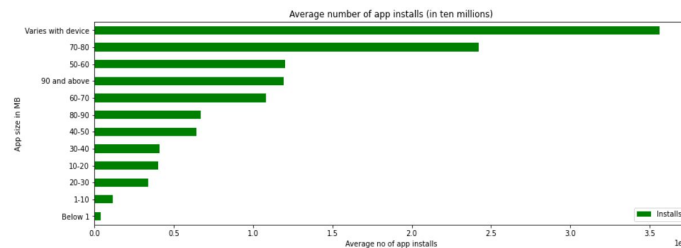
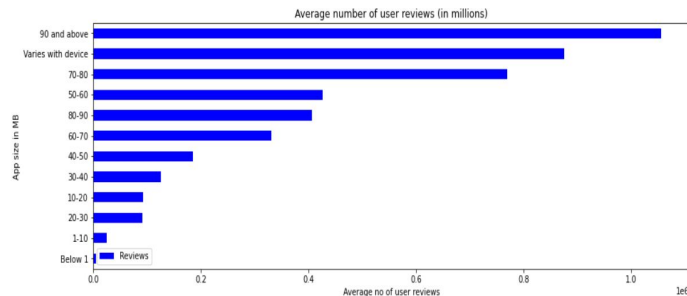
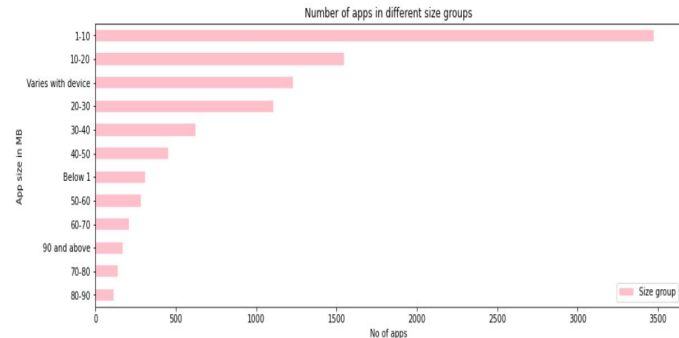
Most Popular Apps

- We can state that the apps with more reviews, whether positive, negative, or neutral, are more popular than the others.
- This is because the number of user reviews indicates that these individuals have engaged with the app's content and have written their opinions on it.

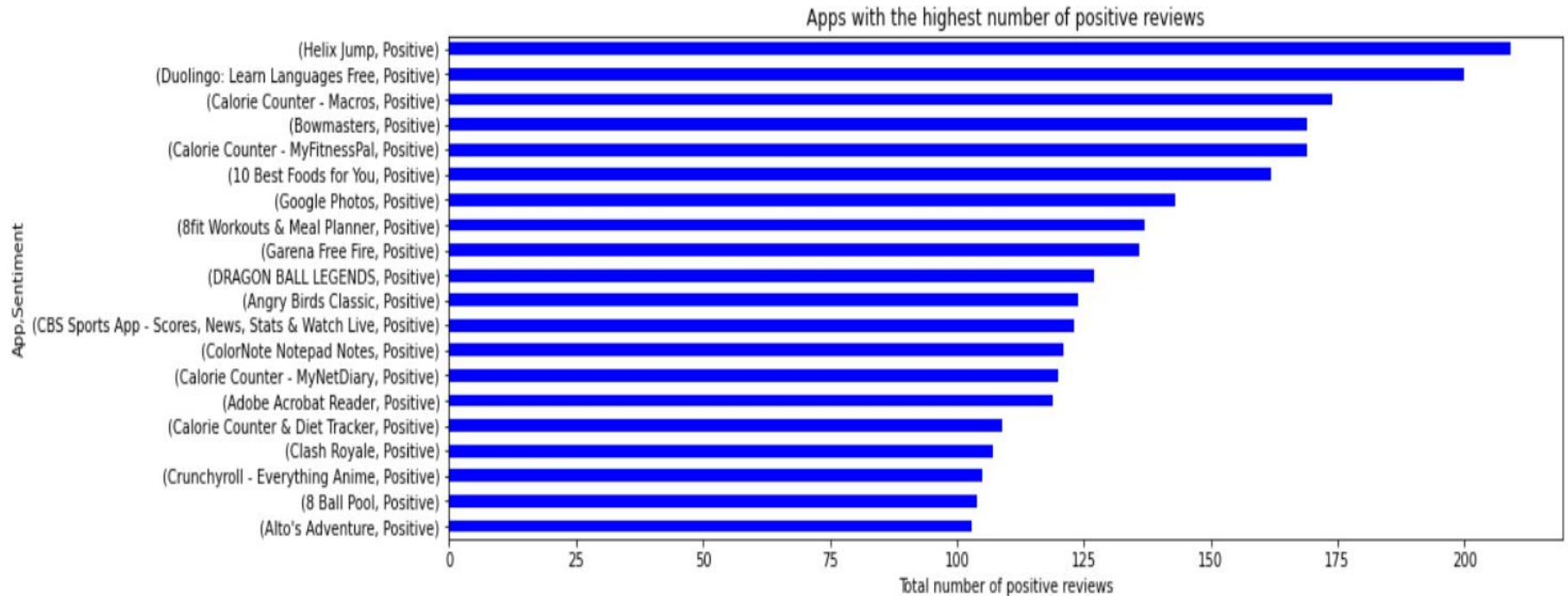


App Size Analysis

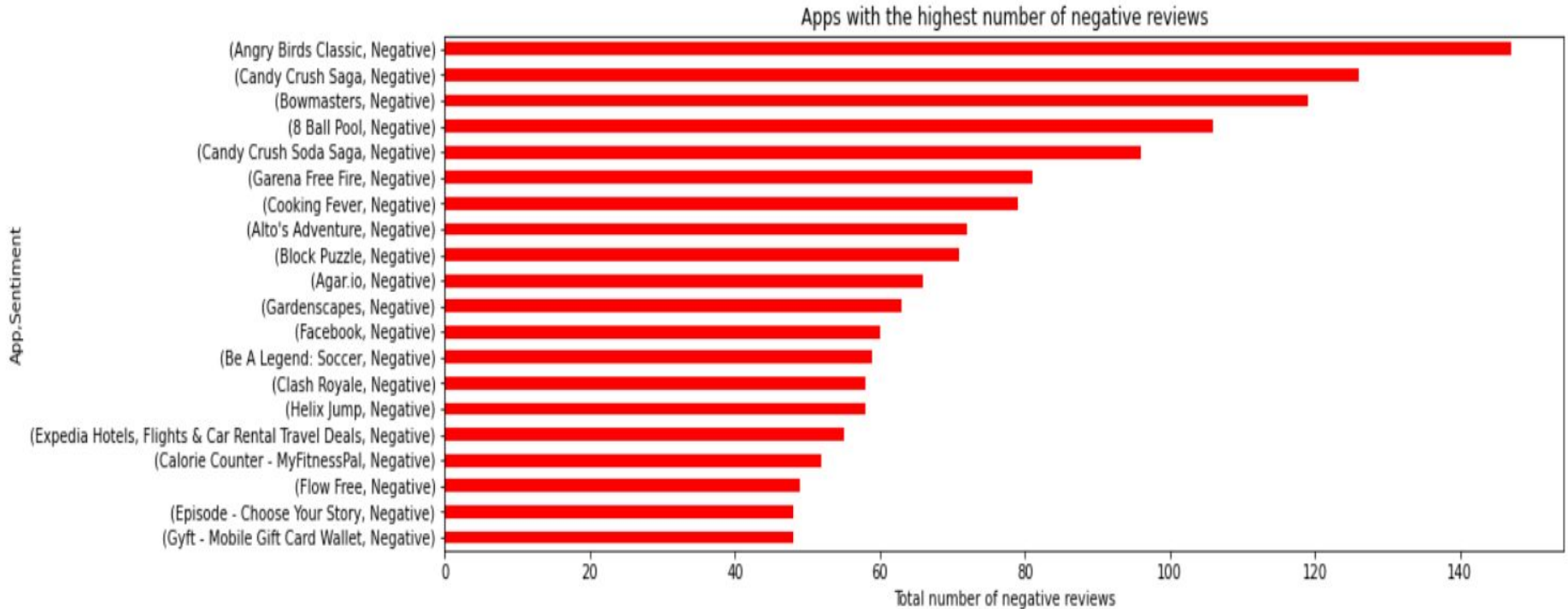
- The apps are categorized based on its size between ~0 to 100 MB in the intervals of 10 MB each.
- The total number of apps in each size category indicates the competition.
- Average number of user reviews and average app installs in each size category indicates the popularity of the respective app.



Positive Reviews



Negative Reviews



Analysis Summary

- Percentage of free apps = ~**92%**
- Percentage of apps with no age restrictions = ~**82%**
- Most competitive category: **Family**
- Category with the highest number of installs: **Game**
- Category with the highest average app installs: **Communication**
- There are **20** free apps that have been installed over a **billion** times
- There is a **positive** correlation between the **reviews** and **installs**. And also between **rating** with **installs** and **reviews**.
- **Price** is **negatively** correlated with the **rating**, **reviews**, and **installs**.

Analysis Summary (Contd.)

- Category in which the paid apps have the highest average installation fee: **Finance**
- Most popular app in the Play Store based on the number of reviews: **Facebook**
- 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