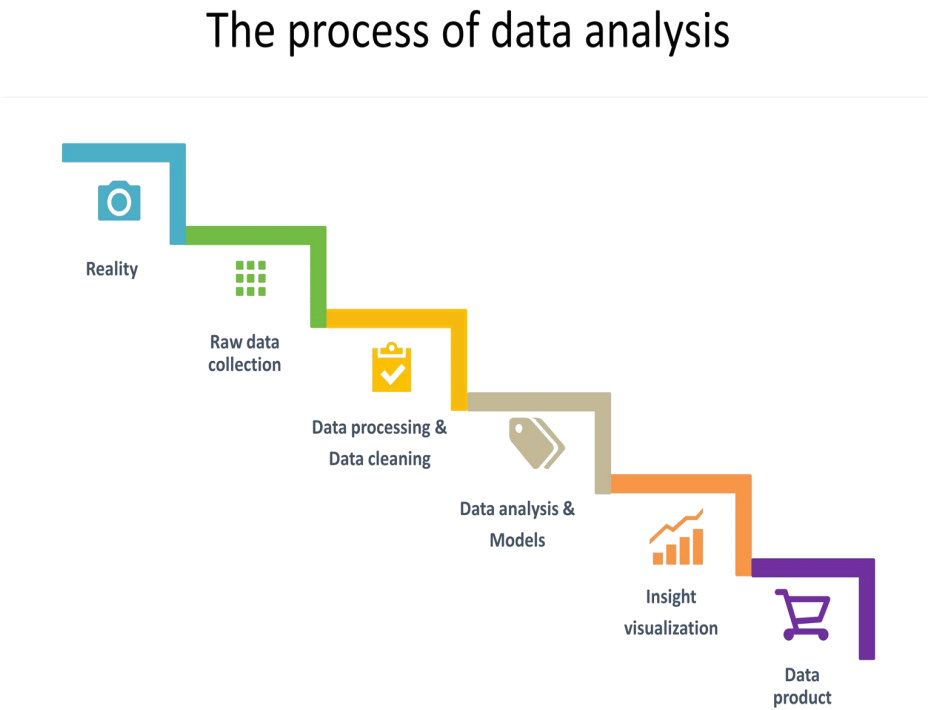
Capstone Project

Play Store App Review Exploratory Data Analysis(EDA)

By
Devyani Chaturvedi,
Data Trainee, AlmaBetter

Content

- Introduction
- Problem Statement
- Data Summary
- Data Exploring & Cleaning
- Data Visualization
  - Categorywise play store appVs Installs
  - Type and Content Rating Distribution
  - Size Distribution
  - Rating Distribution
  - Content Rating
  - Genres
  - Correlation Heatmap
  - App count Vs Categorywise Sentiment reviews
  - Sentiment
  - Sentiment Polarity and Sentiment Subjectivity
  - Popular Category
- Conclusion



Introduction

- *Google Play Store is the official online app store for Android devices.You can download various forms of media onto your Android devices through Google Play.*
- *The Google Play Store is the largest and most popular Android app store. There are more than 3.04 million apps found on play store.The Play Store app data has enormous potential to drive app-making business to success.*
- *Users can install the apps from the Google Play Store and they can give reviews and ratings to the apps based on their experience.*
- *The Objective of this project is to perform EDA on Play Store data to explore and analyze the data and discover key factors responsible for app engagement and success.*
- *Exploratory Data Analysis (EDA) involves using statistics and visualizations to analyze and identify trends in data sets. It is important to get the maximun insights from a data set.*

Problem Statement

We have two datasets, one with **basic information of different play store apps** 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.

- Due to the presence of such wide variety of apps and the data associated with it, we need to extract meaningful insights which are responsible for the apps success and will help developer to capture the android market.

Data Summary

1) Play Store Data :

- Apps – App names.
- Category – The category to which app belongs.
- Rating – Rating of the app.
- Reviews – Number of reviews given to each app.
- Size – Size of the app.
- Installs – Number of installs of each app.
- Type – Free or Paid
- Price – Price of the app in $.
- Content Rating – Age restriction for each app.
- Genres – Genre the app belongs to.
- Last Updated – When the app is last updated.
- Current Ver – Current version of the app.
- Android Ver – Android version on which the app is supported.

2) User Reviews :

- Apps – App names
- Translated_Review – Reviews given to each app.
- Sentiment – Sentiment of reviews Positive/Negative/Neutral.
- Sentiment_Polarity – Sentiment polarity score from -1 to 1.
- Sentiment_Subjectivity – Sentiment subjectivity score.

Data Exploring & Cleaning

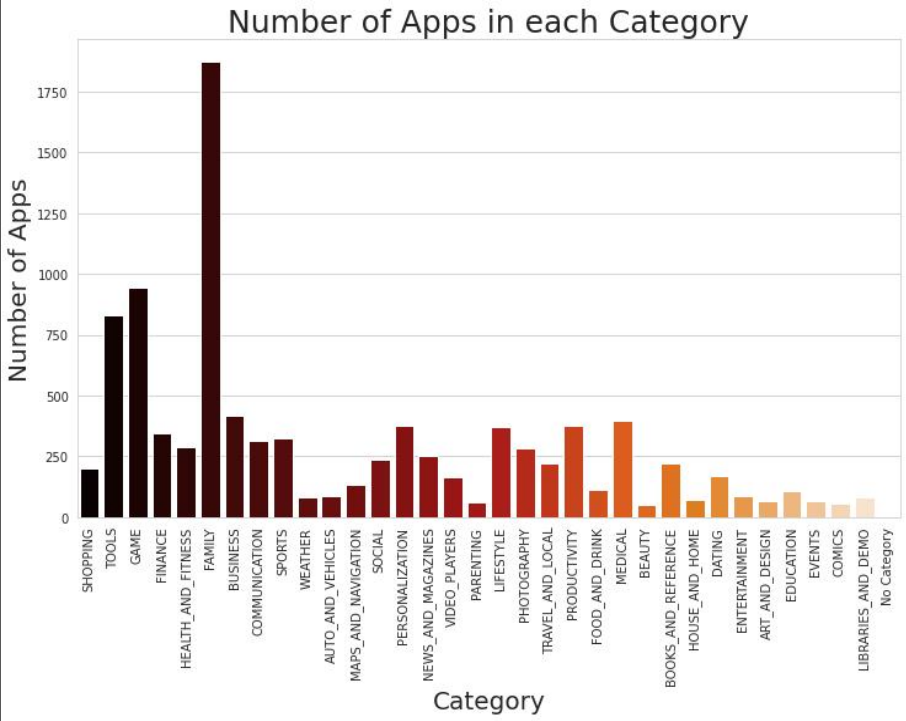
This step involves the following -

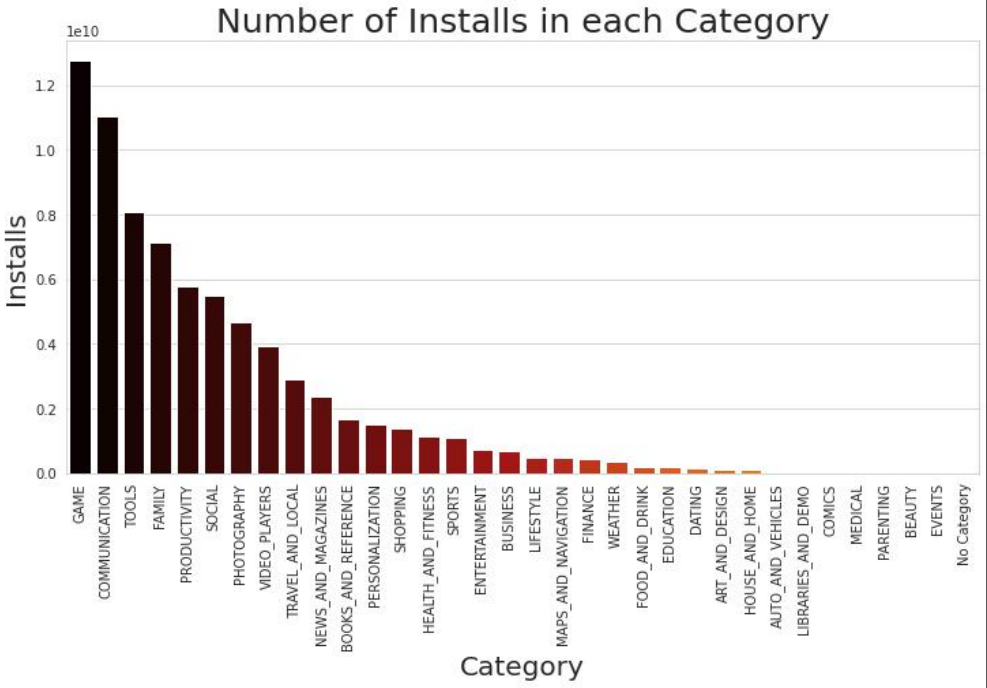
- ❑ Exploring and understanding the characteristics of datasets
- ❑ Converting columns with numeric data into integer/float datatype
- ❑ Handling the null values
- ❑ Checking and removing duplicates

After cleaning and organizing the datasets to make it accessible for visualization in order to analyze it, we'll perform data visualization.

Data Visualization

Apps Vs Category



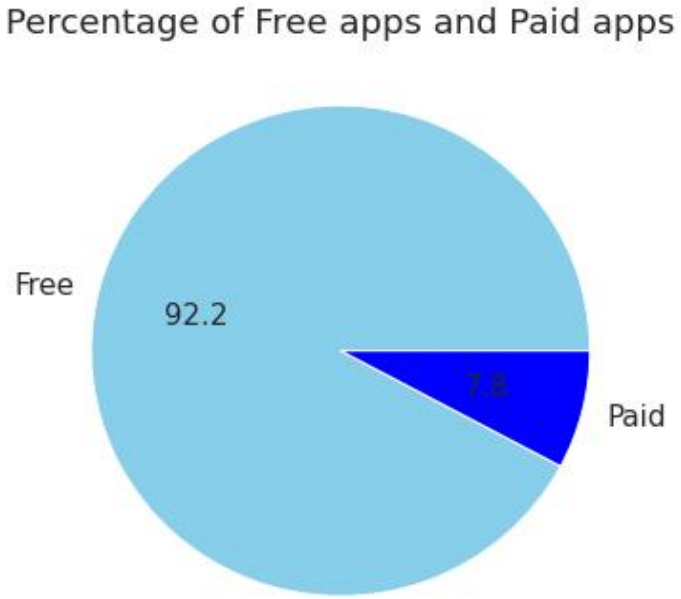


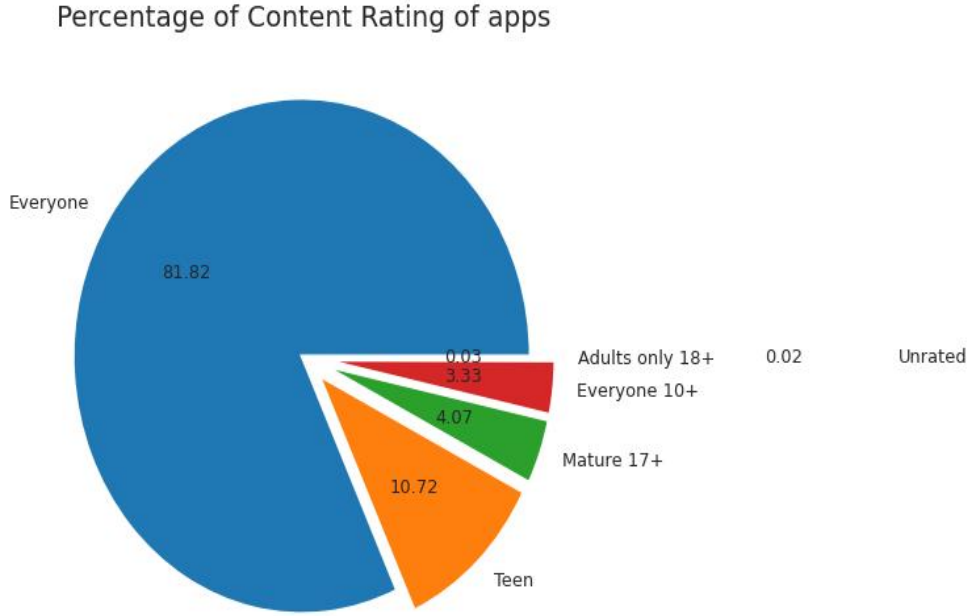
- Family, Game, Tools, Business & Medical are the top categories having maximum number of apps.

Game, Communication, Tools, Family and Productivity are the top categories having the highest number of installs.

Data Visualization

Type and Content Rating Distribution Percentage



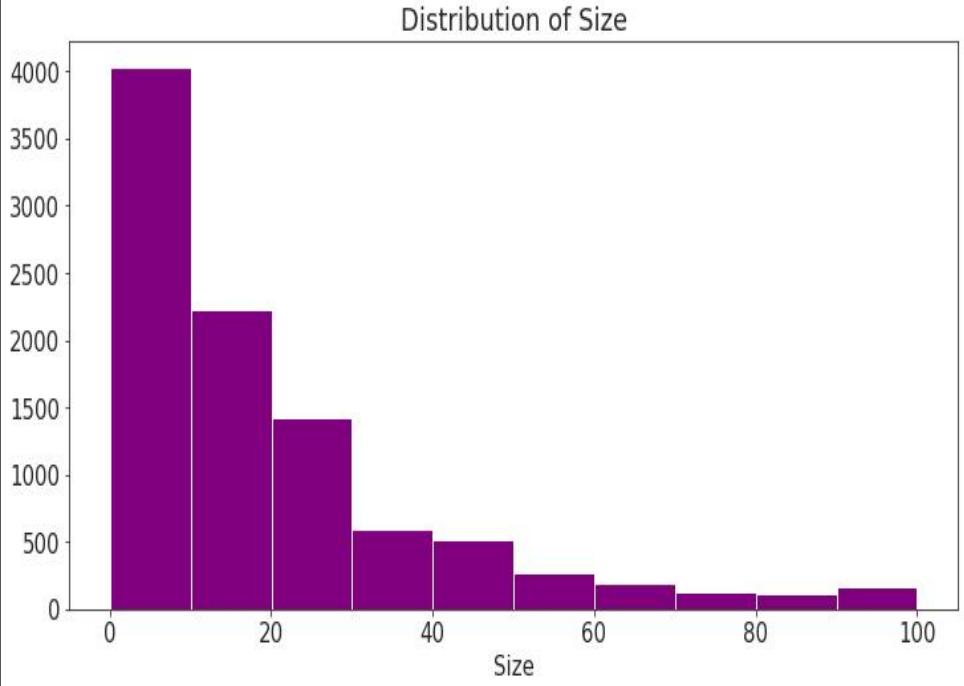


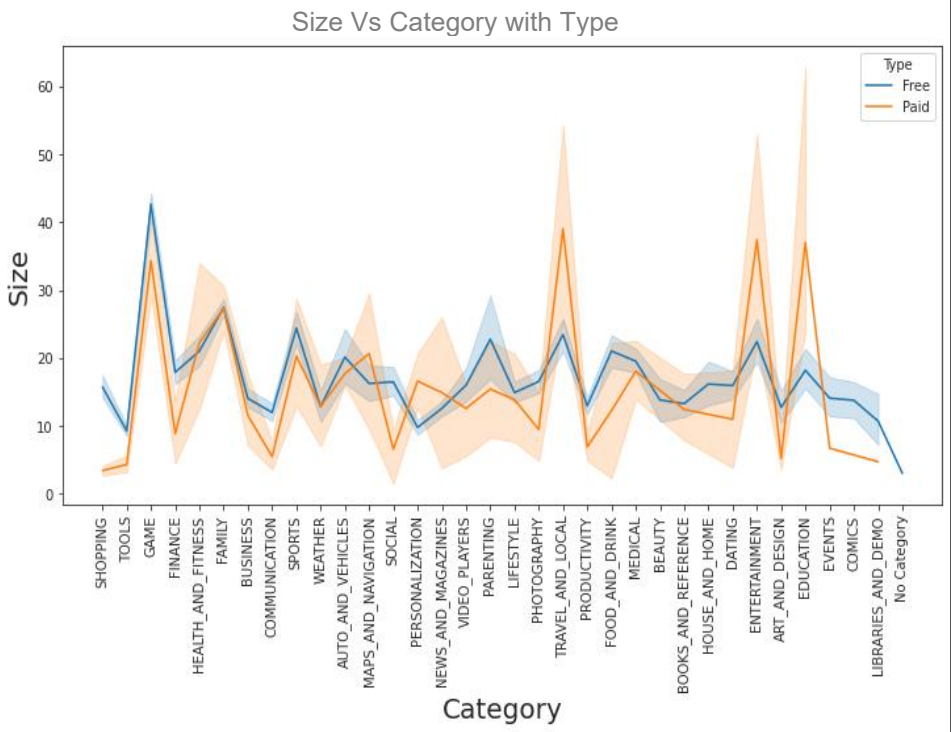
- **92.2% apps on play store are available for Free, only 7.8% apps are Paid.**

~82% apps are rated for Everyone, ~11% apps are rated for Teen and rest have other Content Ratings.

Data Visualization

Size Distribution



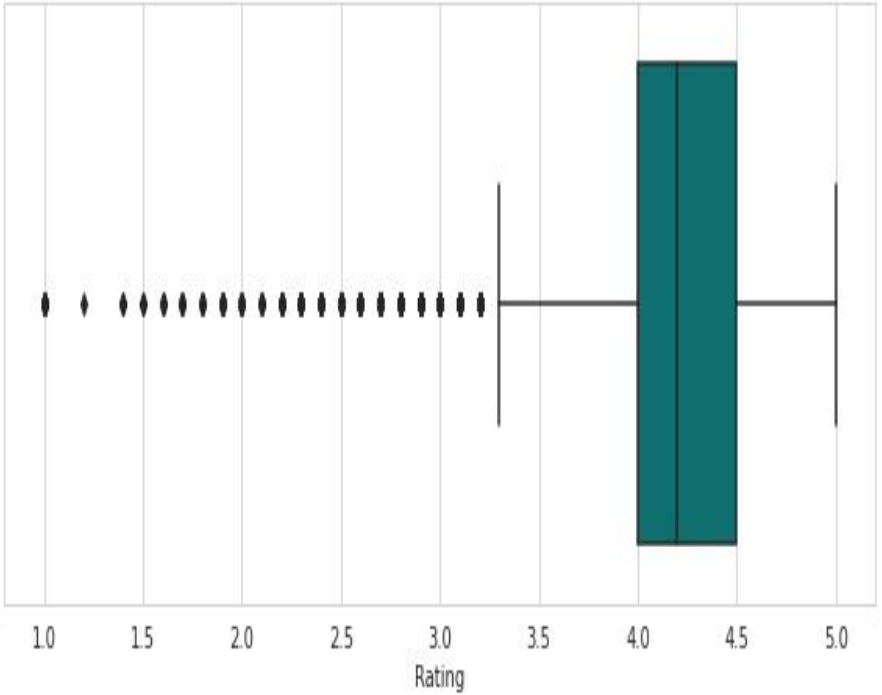


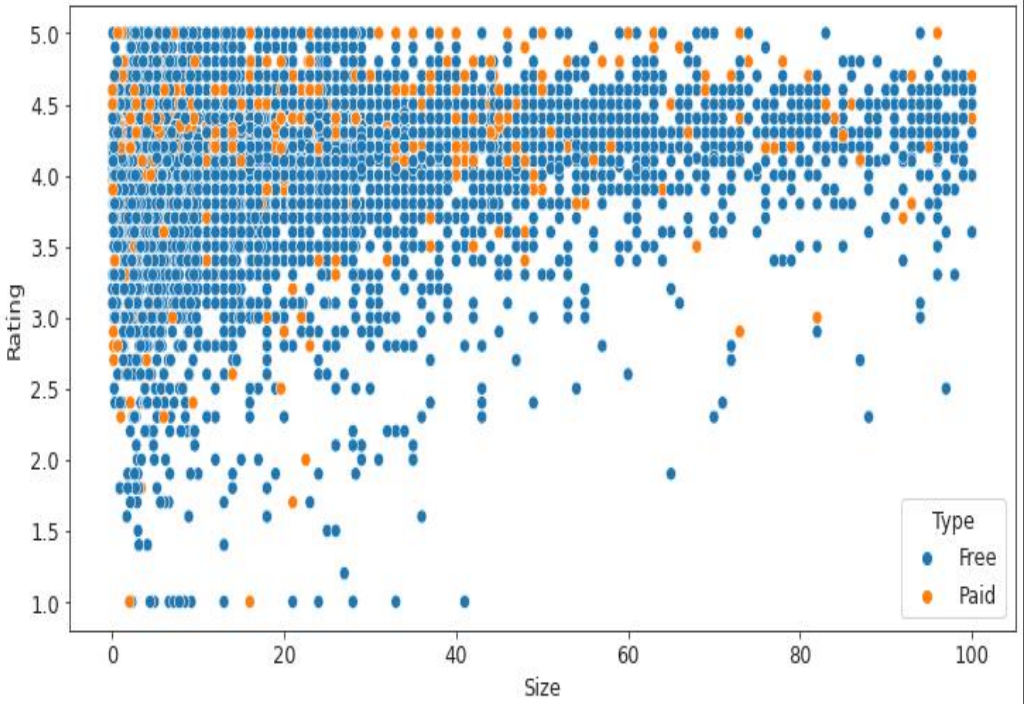
- A large number of apps are small in size.
  More than 6000 apps have size less than 20 MB.

Average size of most of the categories is less than 30 MB whether it is Free or Paid.

Data Visualization

Rating Distribution



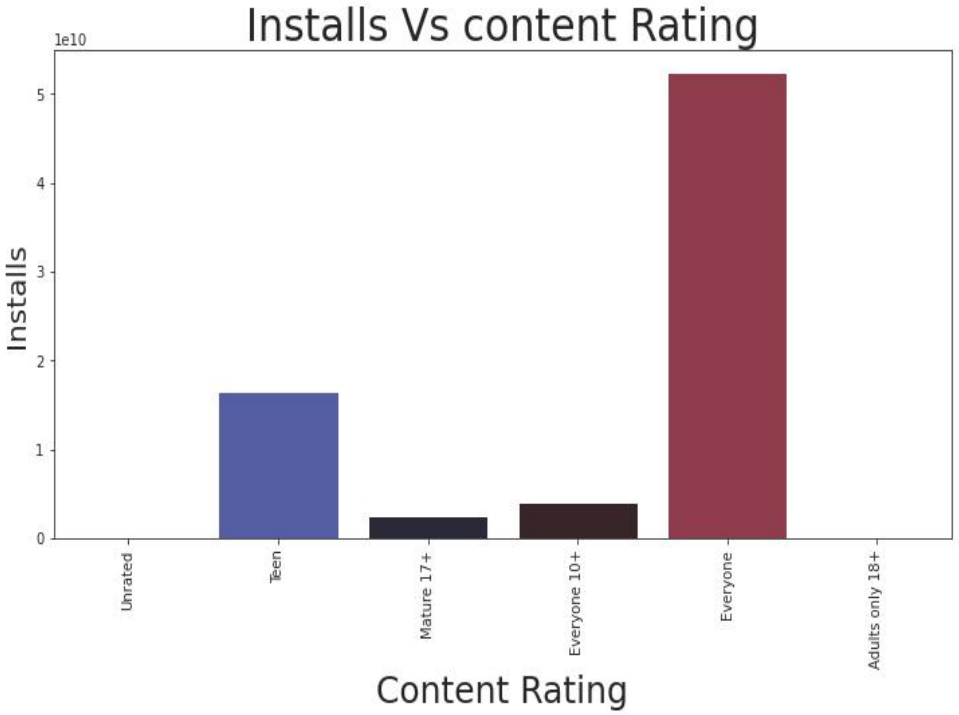


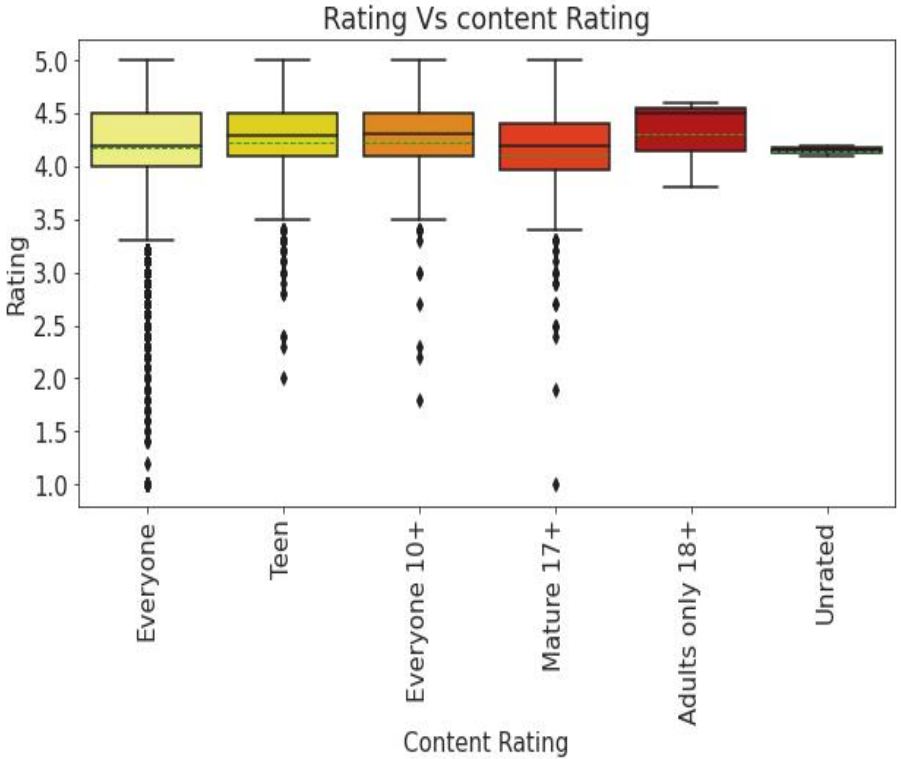
- 50% Ratings are between 4 and 4.5 and Median of Rating is 4.2.

Most of the apps are small in Size and have high Rating irrespective of their Type.

Data Visualization

Content Rating



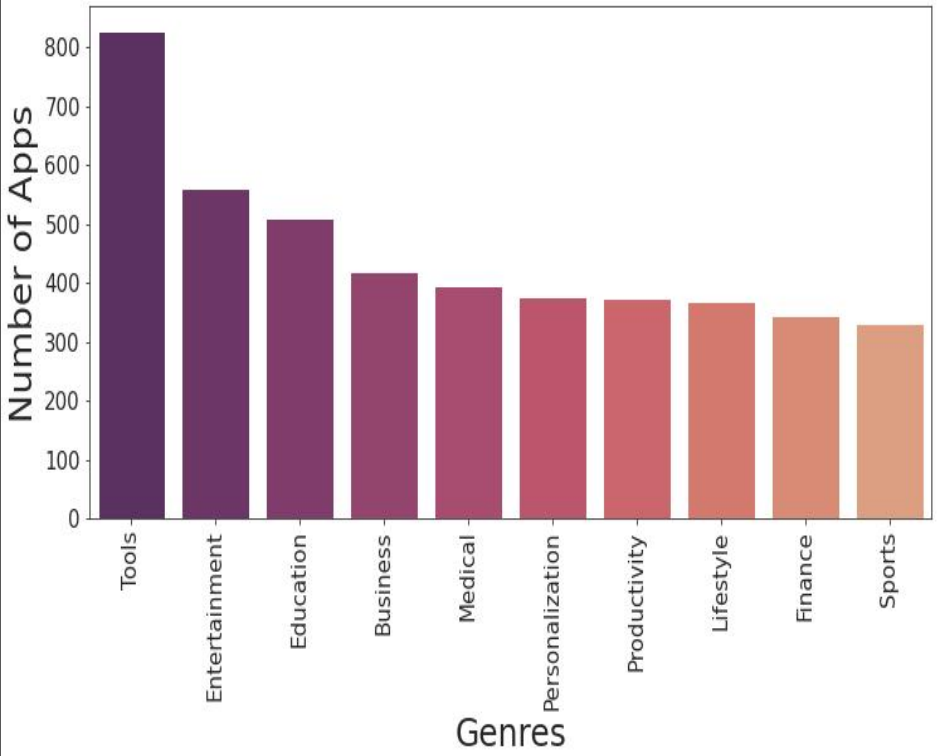


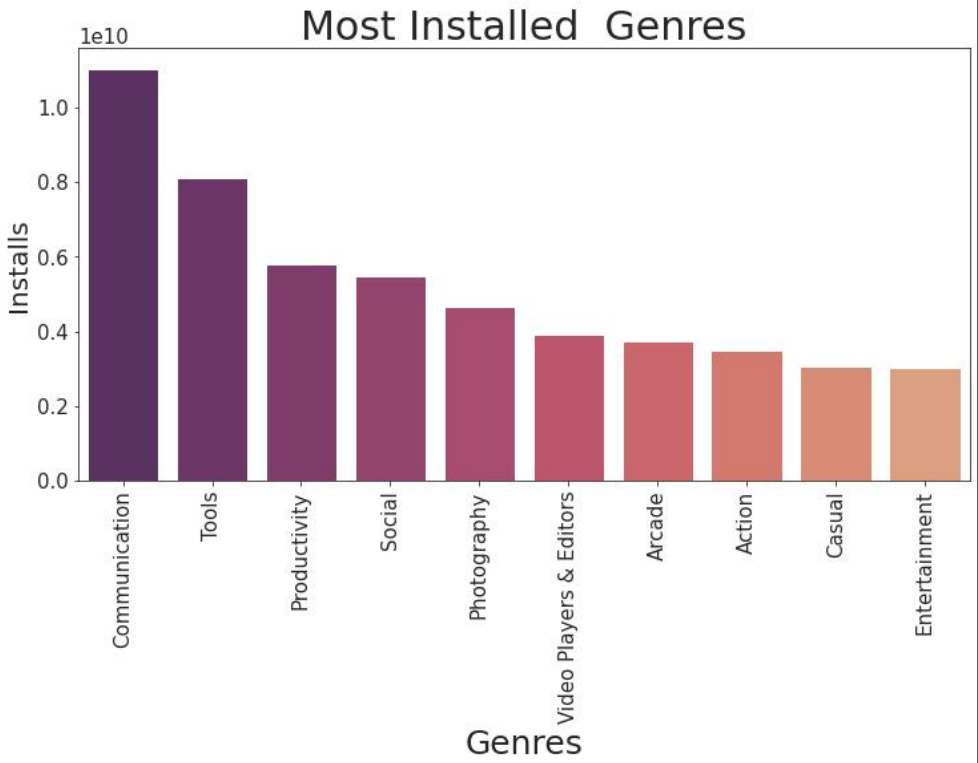
- **Apps Rated for Everyone has the highest number of Installs followed by apps rated forTeen**

Apps rated for Adults only 18+ and Teen have the highest average Rating

Data Visualization

Genres(Top 10)





- These are the top ten Genres out of 119 Genres having the highest no. of apps.

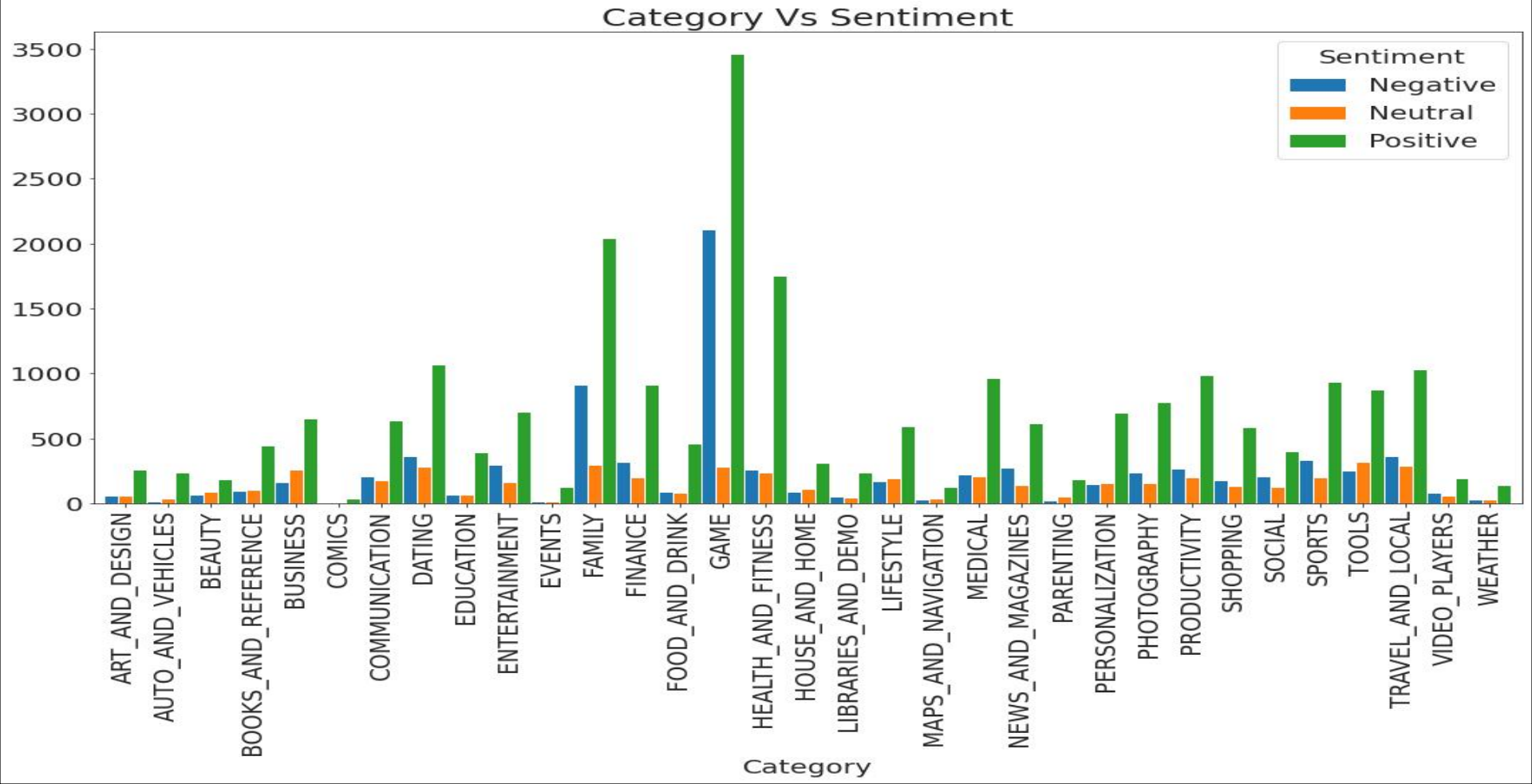
Communication and Tools are the most Installed Genres.

Data Visualization

Correlation Heatmap

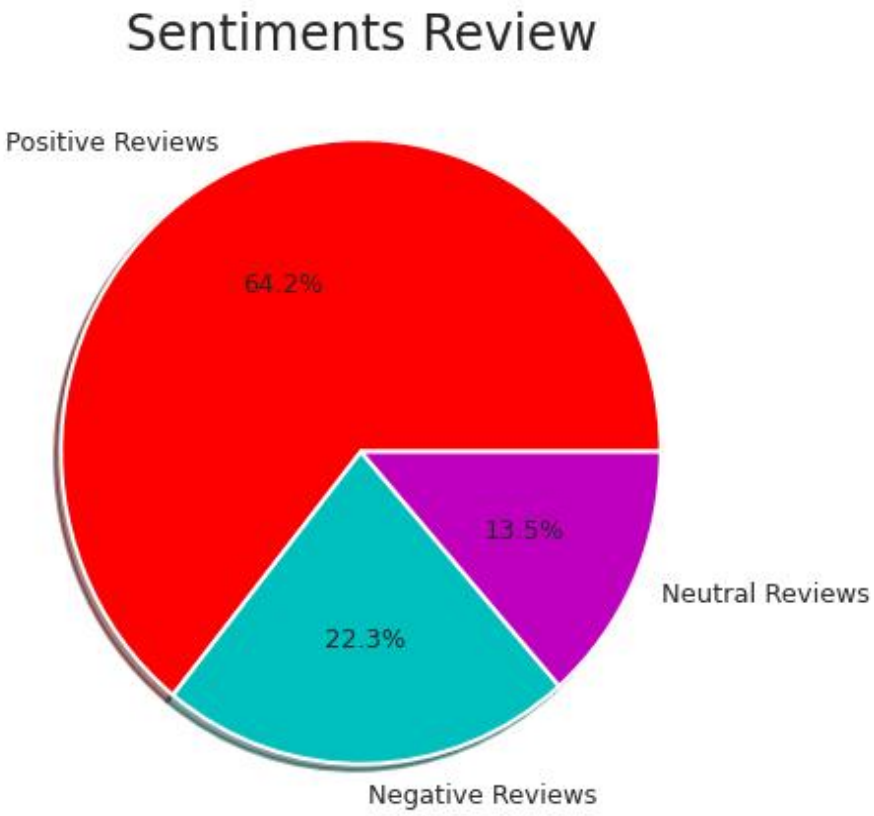
- Installs and Reviews are highly positively correlated with each other
- Size is slightly positively correlated with Reviews and Rating
- Price is slightly negatively correlated with Installs, Reviews and Rating
- Sentiment Polarity and Sentiment Subjectivity are not highly correlated

Data Visualization

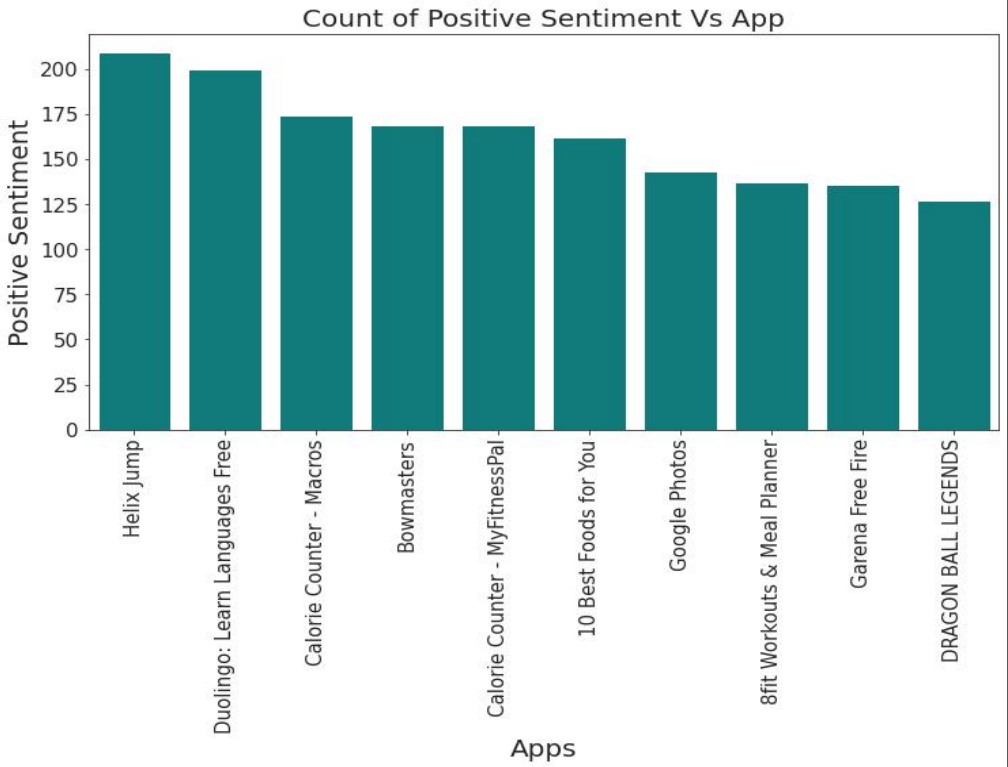


Data Visualization

Sentiment



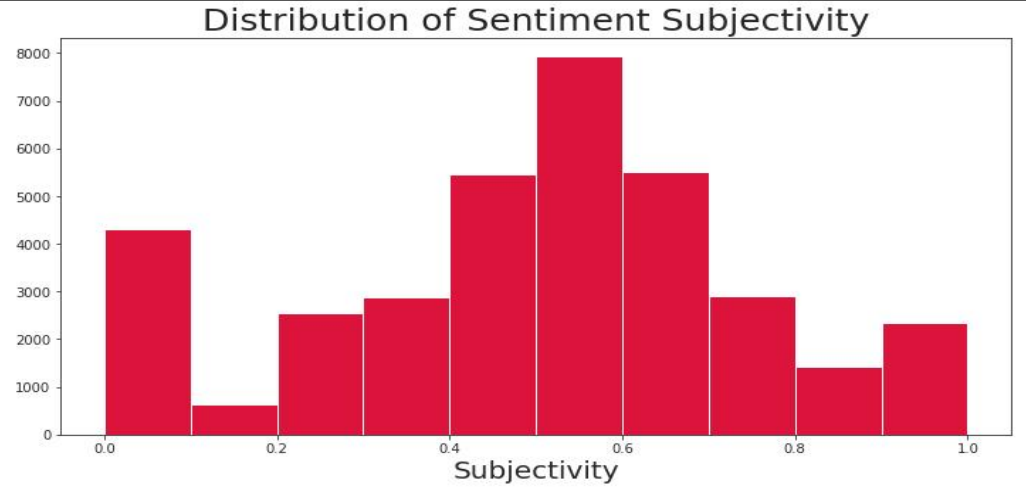
Top 10 apps with max. Positive Sentiment reviews



Data Visualization

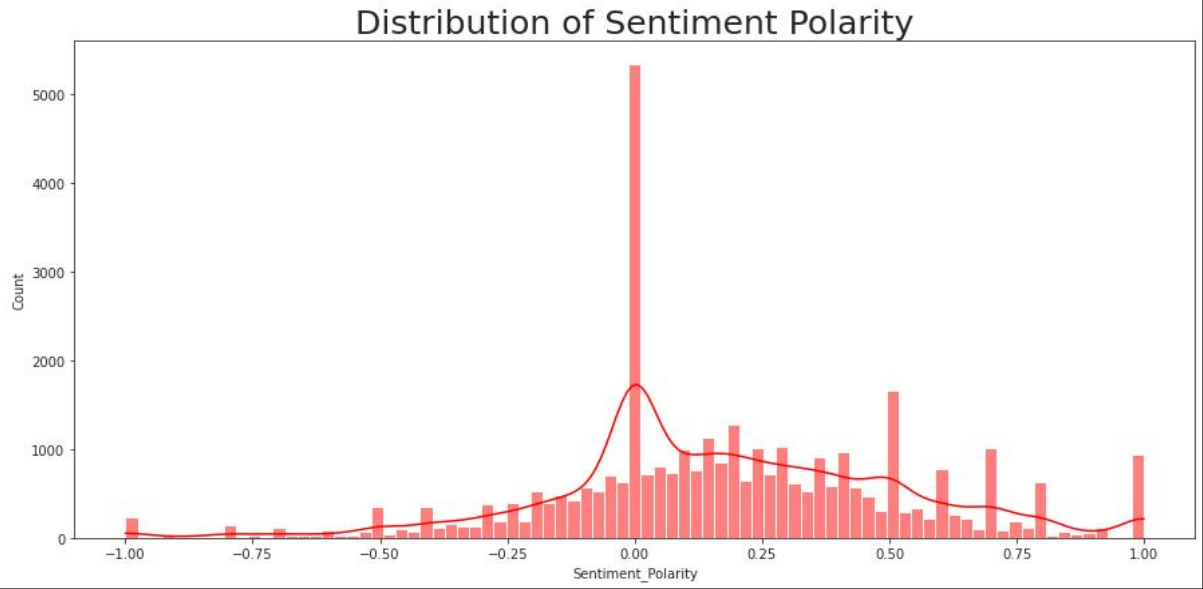
Sentiment Subjectivity

- Most of the sentiment subjectivity lies between 0.4 to 0.7 which shows that most of the reviews are towards subjective point of view of the users.



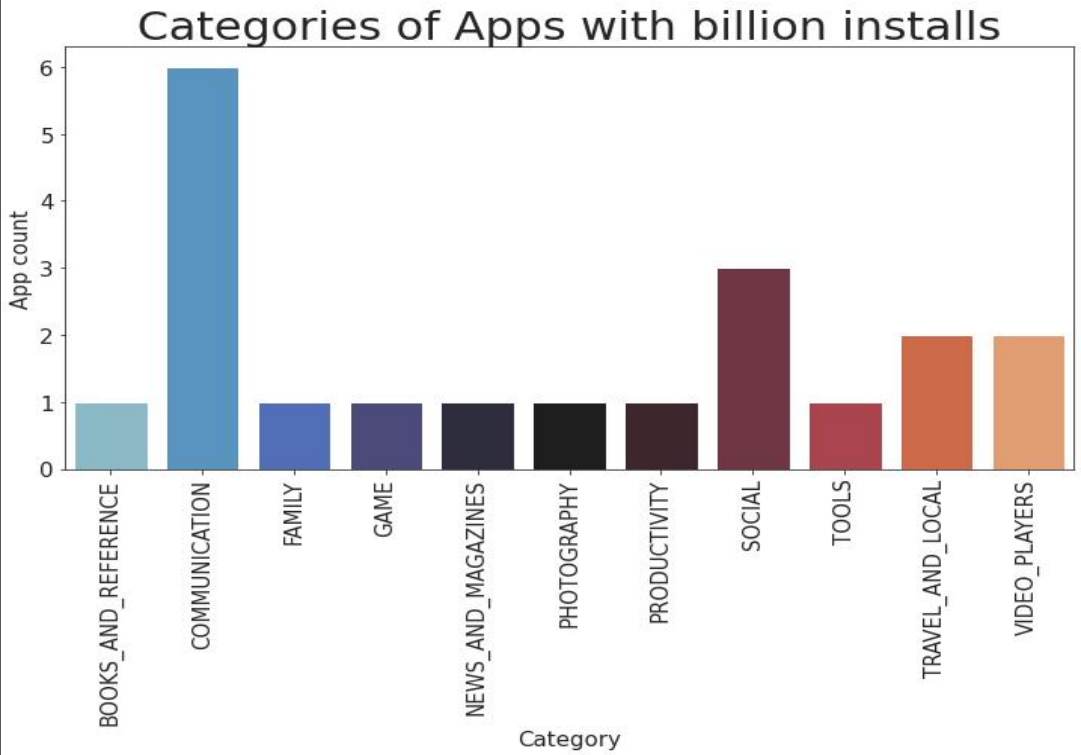
Sentiment Polarity

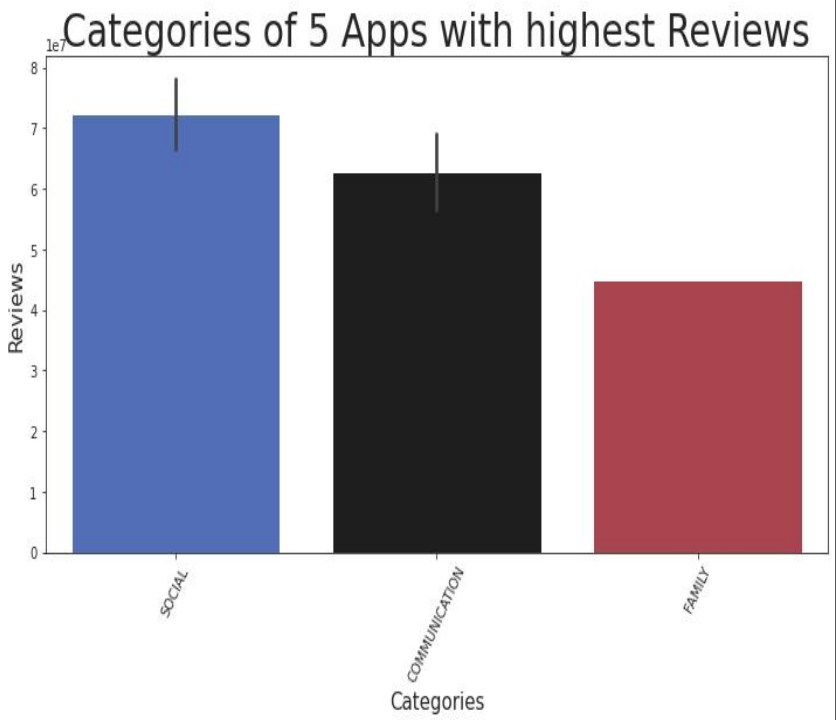
- The Polarity of most of the users is towards the positive side as we already saw in the pie chart.



Data Visualization

Popular Categories





- **Communication and Social categories are the top most installed and reviewed Categories**.

Conclusion

- Family Category has the highest number of apps on Play Store while Game Category has the highest number of installs, this shows that the categories which are more entertaining and interactive have more user's engagement and interest.
- 92.2% apps are Free and 7.8% apps are paid in type, also users tend to install Free apps way more than Paid apps.
- 81.8% apps have Everyone Content Rating and 10.7% apps have Teen Content Rating.
- Most of the apps have Size less than 50 MB including both Type. Thus, a paid app that is larger in size may not perform well in market as users prefer to pay for apps that are light-weighted.
- 75% apps have Rating between 4 and 5 of both Type.
- Most of the apps are running at Android Version 4.1 and above.Apps Rated for Everyone has the highest number of Installs followed by apps rated for Teen. Apps rated for Adults only 18+ and Teen have the highest average Rating.
- Content Rating Teen have quite good number of installs and Adults have the highest average rating which shows that the present youths are quite good at operating apps and thus developers can develop more apps which suits to the interest of the youth.
- Tools and Entertainment **Genres** have the highest number of apps whereas Communication and Tools Genres have the highest number of Installs.

Conclusion

- Apps having highest number of reviews are from the categories of **Social, Communication** and **Game** indicating the active participation of users on apps from these categories.
- **Facebook** is the most reviewed app with the review count of 78158306. **Helix Jump** has the highest number of Positive Reviews and **Angry Birds Classic** has the highest number of Negative Reviews.
- Most of the **sentiment subjectivity** lies between 0.4 to 0.7 which shows that most of the users give reviews to the application on the basis of their experience.
- The **Sentiment Polarity** of most of the users is towards the positive side.
- Installs and Reviews are highly positively correlated with each other. Size is slightly positively correlated with Reviews and Rating. Price is slightly negatively correlated with Installs, Reviews and Rating. Sentiment Polarity and Sentiment Subjectivity are not highly correlated. Price, Rating, Size **has no or very less correlation** with **Sentiment Polarity**.
- There are **20** apps that have been installed over a **billion** times and all of them are of Free Type. **Minecraft** is the only app in the Paid Type with over **10M** installs, and also has produced the most revenue only from installation fee.

These are some trends that we have observed from exploratory data analysis and made some assumptions that might lead to app success among the users in the play store and suggest categories of the apps.


THANK
YOU !