

Capstone Project1

Play Store Apps Review Analysis

Team

Name: Web Crawlers

Members:

1. Rajat Chaudhary
2. Anukriti Shakyawar
3. Raman Kumar
4. Deepmala Srivastava

Content

- 1. Introduction**
- 2. Problem Statement**
- 3. Data Description**
 - a) Play store App Data**
 - b) User Reviews Data**
- 4. Data Processing**
- 5. Imported Libraries**
- 6. Data Cleaning & Filtering**
- 7. Exploratory & Visualization Analysis**
- 8. Conclusion**

Introduction

- Play store is one of the marketplaces for downloadable software programs with the highest growth in mobile applications.
- The play store apps data has a lot of potential to help app development companies succeed.
- Given the explosive rise of Android-based gadgets and applications, it will be useful to do data analysis on the collected information to gain insightful knowledge from this data.

Problem Statement

The objective of this project to performed exploratory data analysis to discover key factors responsible for app engagement and success.

Data Description



1 Play store App Data

It consists of 13 columns with 10841 Rows :-

1. **App:** Name of the App
2. **Category:** Category under which it falls
3. **Rating:** Application's rating on playstore
4. **Reviews:** Number of reviews of the app
5. **Size:** Size of the app
6. **Installs:** Number of Installation of the app
7. **Type:** Whether the app is free or paid
8. **Price:** Price of the app if it's a paid app (0 if it's a free app)
9. **Content Rating:** Appropriate target audience of the app
10. **Genres:** Genres under which the app falls
11. **Last Updated:** Date when the App was last updated
12. **Current Ver:** Current version of the App
13. **Android Ver.:** Minimum android version required to support the App.

Data Description



2. User Reviews Data

In this dataset the review given by the consumers were calculated on different parameters, as:

1. App

2. Translated review

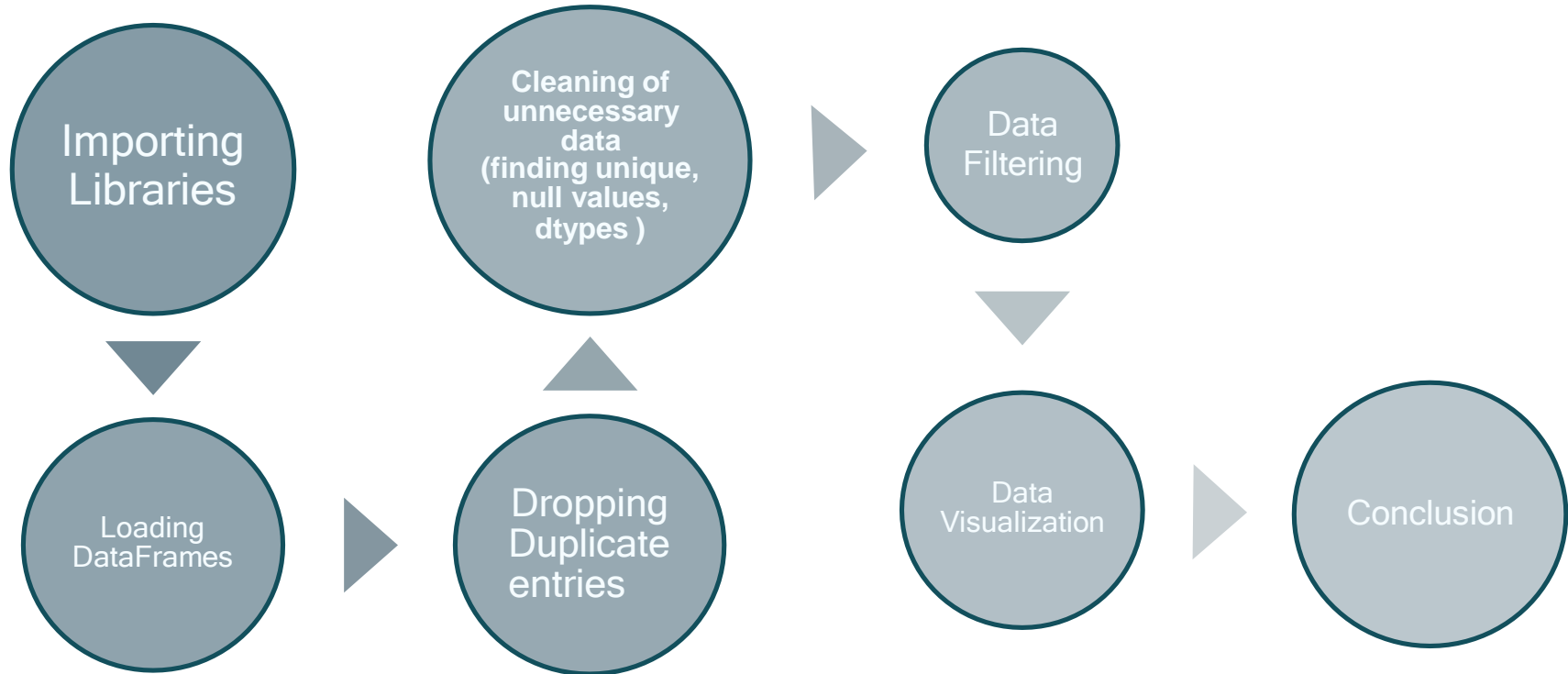
3. Sentiment

4. Sentiment Polarity

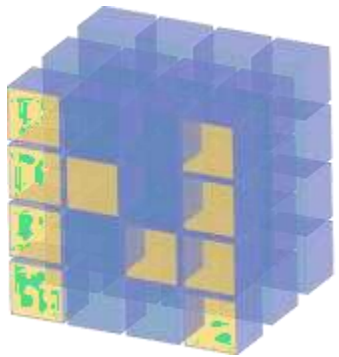
5. Sentiment Subjectivity

Data Processing

AI



Imported Libraries



NumPy

matplotlib



seaborn



pandas

Data Cleaning & Filtering

Play Store Dataset

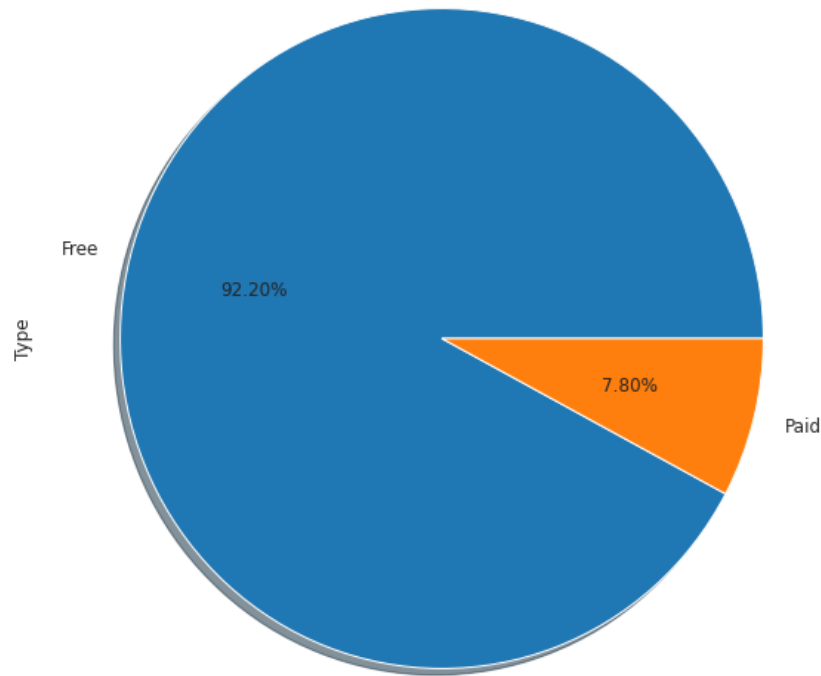
- Identifying and removing duplicate entries.
- Removing visual impurities like "+", "\$" sign.
- Comparing unique and non-null values to check how many entries were repeated.
- Filtering null values by operating on each column individually.

User Reviews Dataset

- Removing Duplicate values
- Filtering null values

Data Visualization

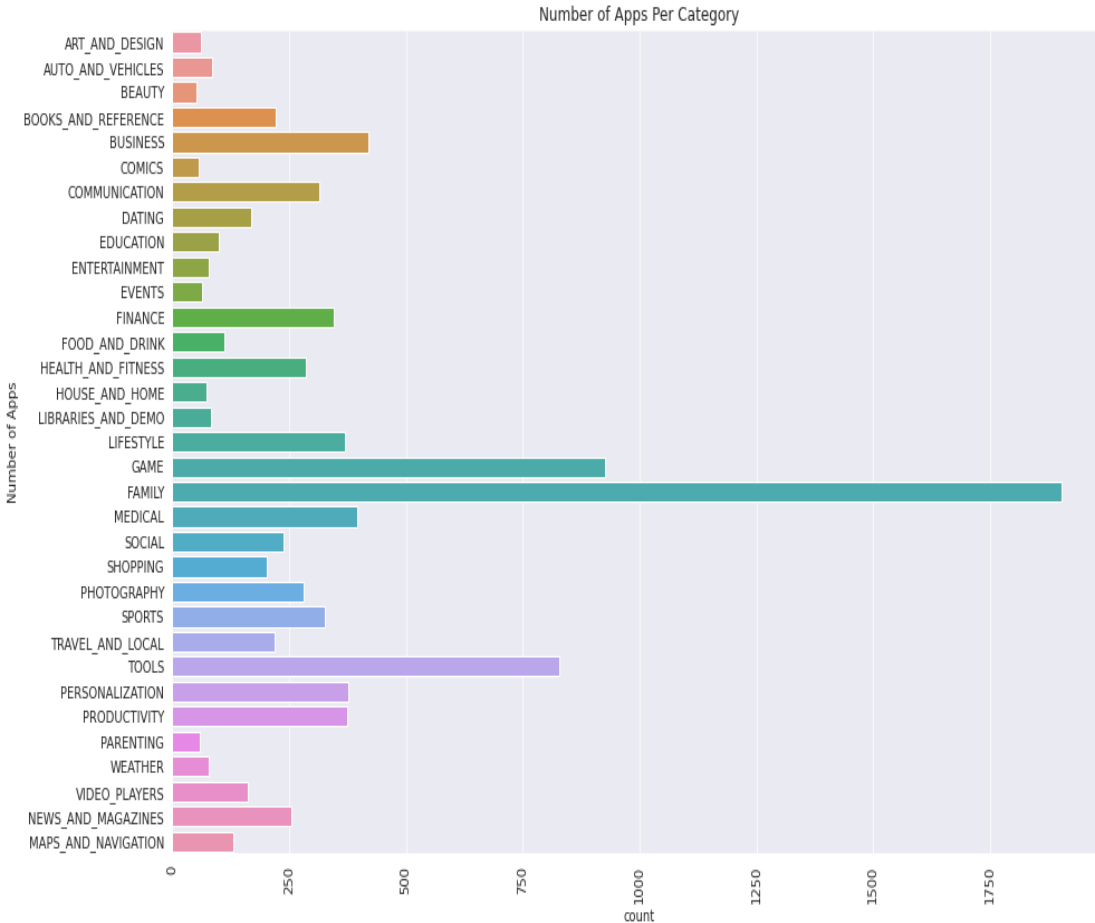
Distribution of Application Type



Play Store has 92.2% Free apps and only 7.80% are paid apps.

Number of Apps per Category

AI

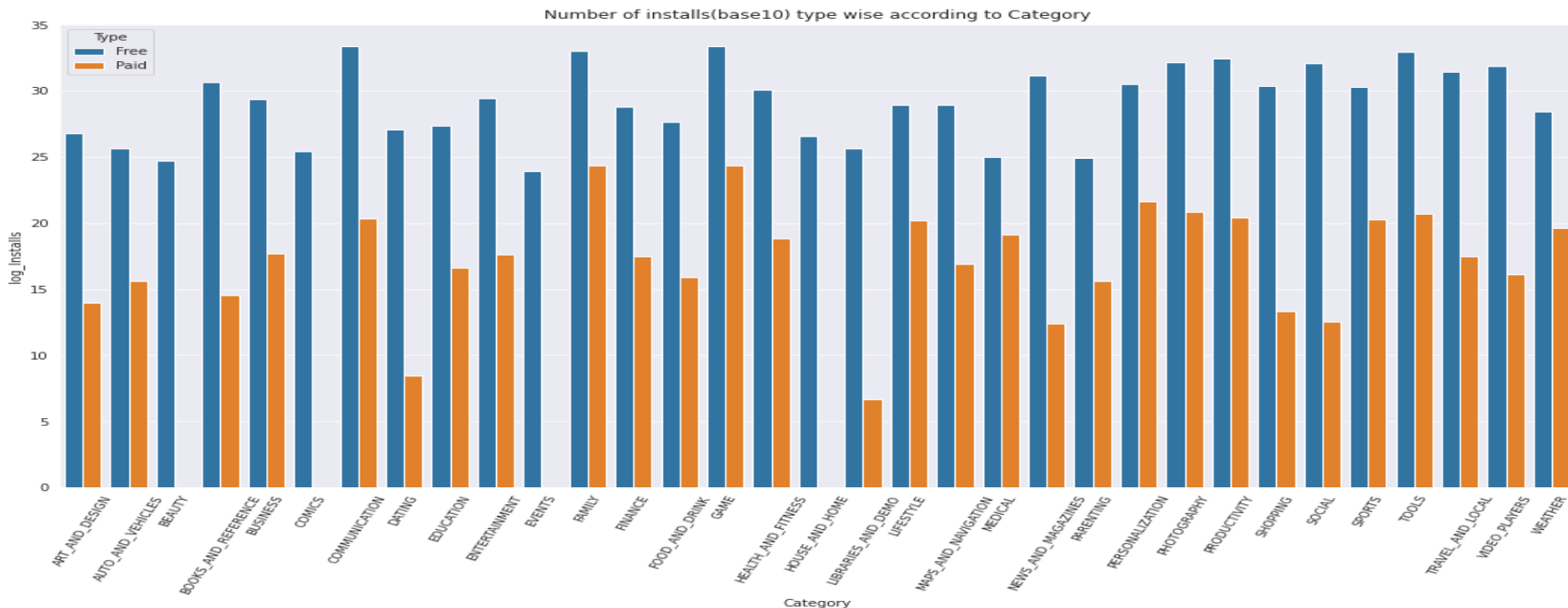


Most of the apps are from the categories of 'Family', 'Game' and 'Tools' category.

Number of Installs Type-wise according to Categories

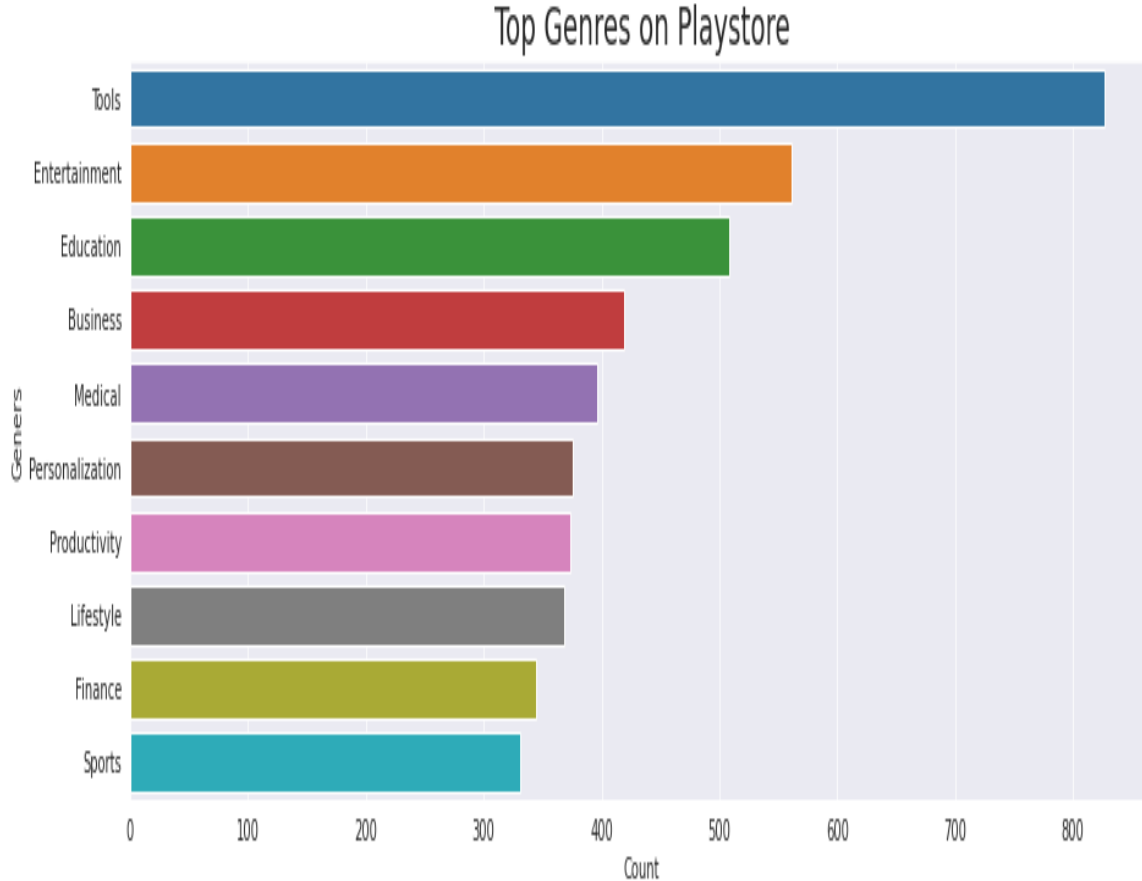


App installations have a significantly higher proportion of free software than paid ones.



Top Genres in Playstore

AI

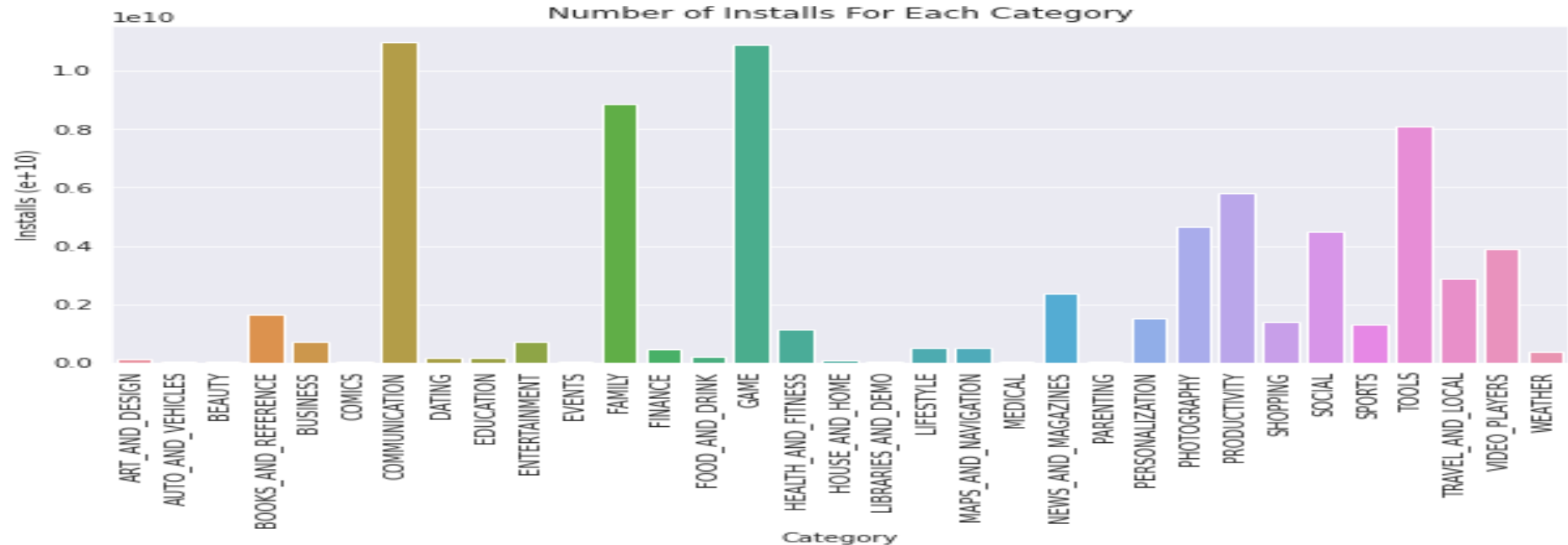


The top genres in our data set is Tools followed by entertainment, education, business, medical apps and so on.

Number of installs for each category

AI

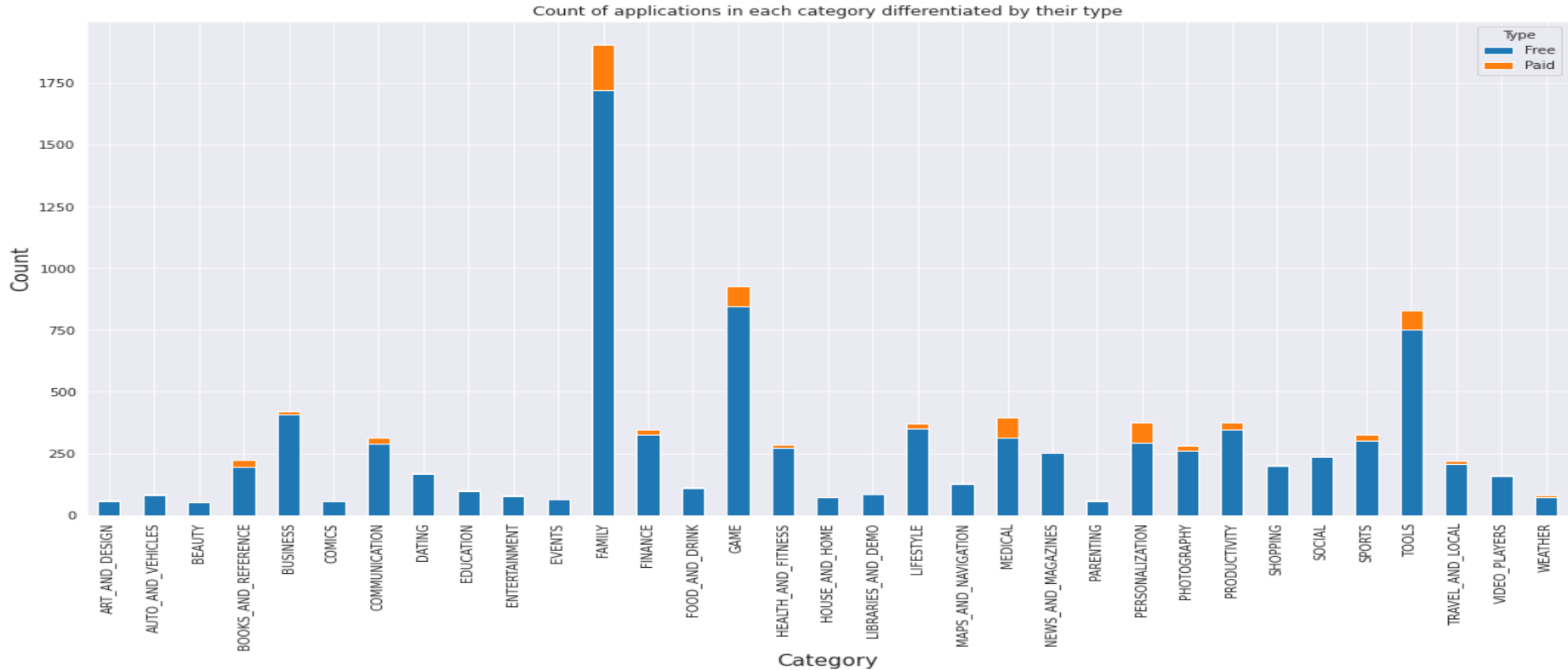
- Games, communication, Family and tools have the most installed apps.



Number of apps per category differentiated by type



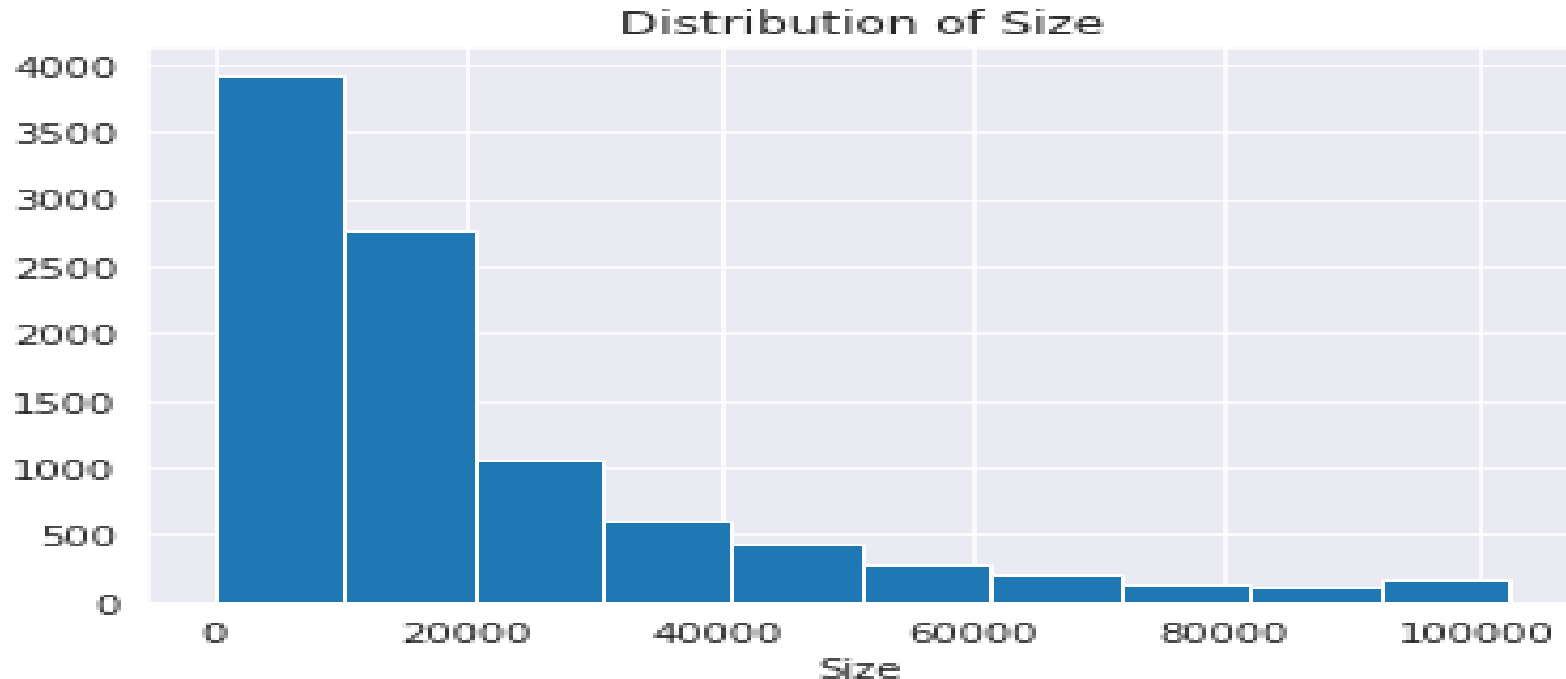
- Most free app download categories: Family, Games, Tools and Social
- Most paid apps download categories: Family, Personalization, and Medical



Distribution of Size of app



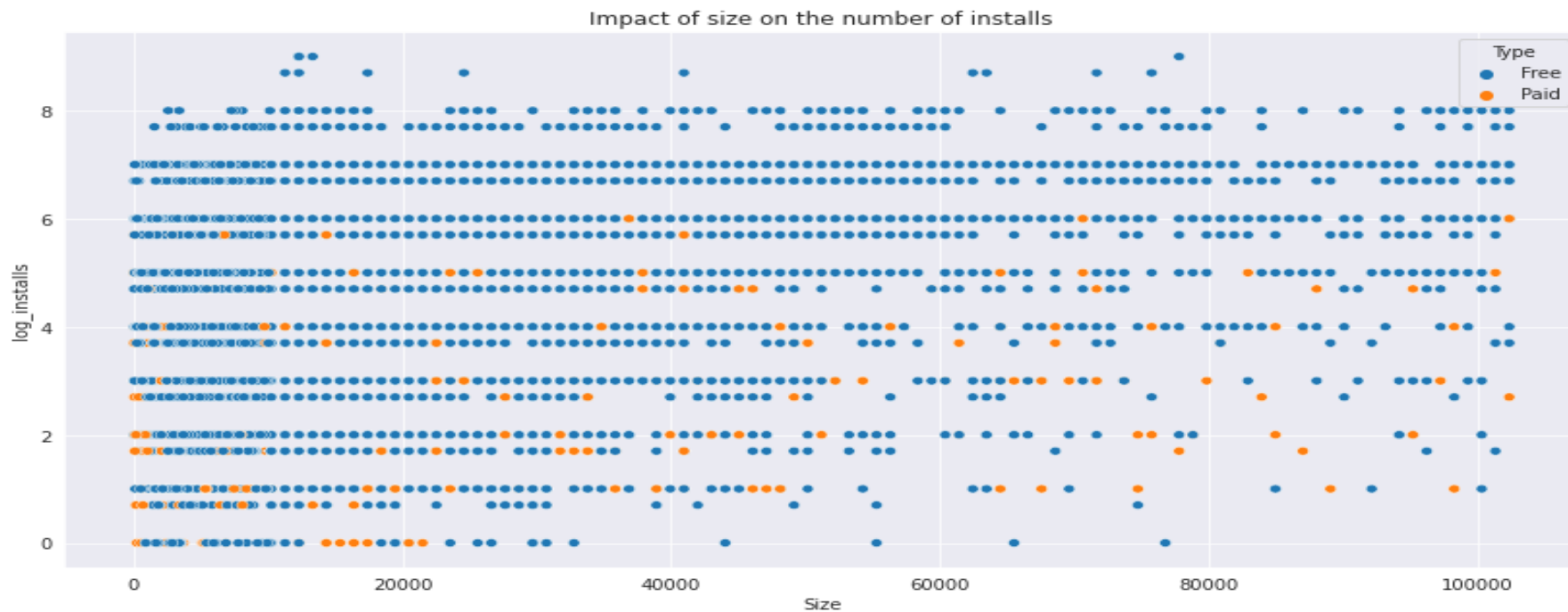
- Size of Maximum number of Apps are in the Range of 0 KB to 2000 KB.
- Most of the apps present in Play Store are smaller in size and consume less memory.



Impact of size on the number of installs

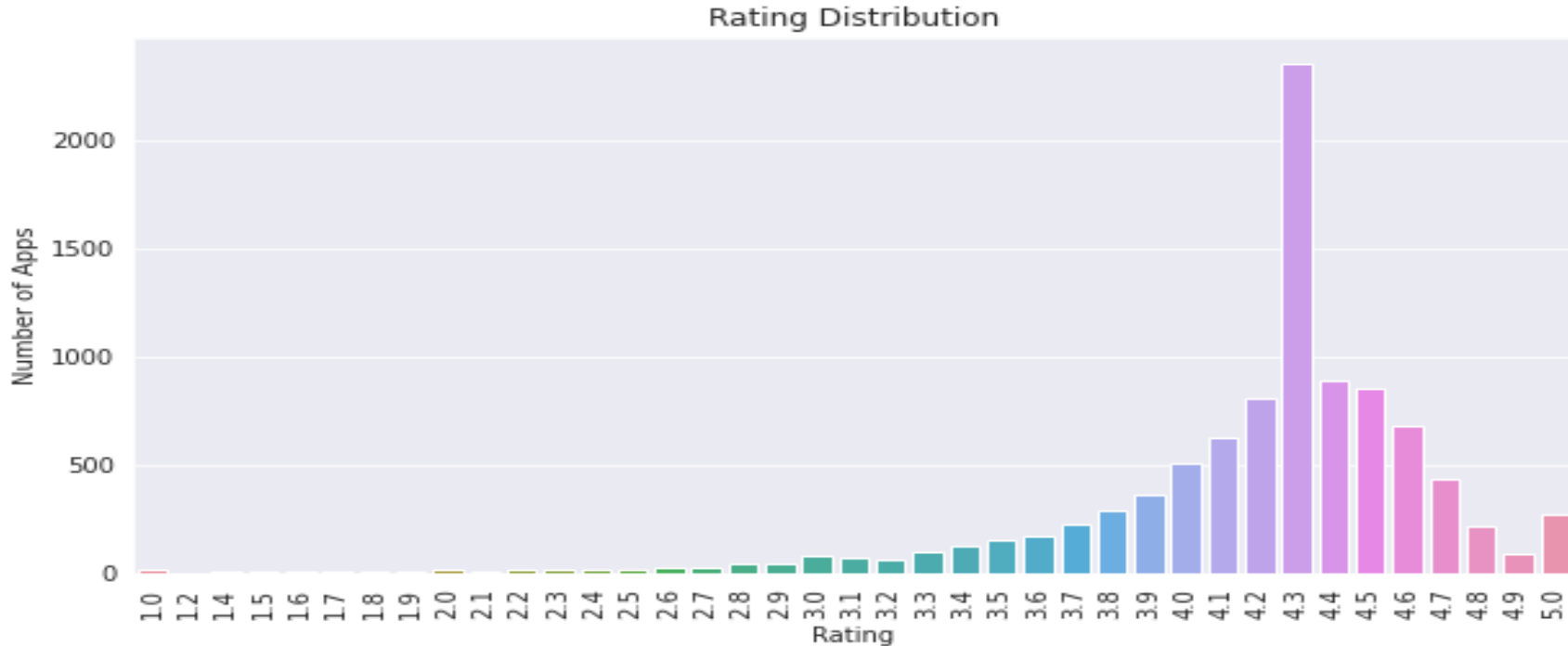


- Size impacts the no. of installs.
- Bulky apps tend to be installed less.



Distribution of App Rating

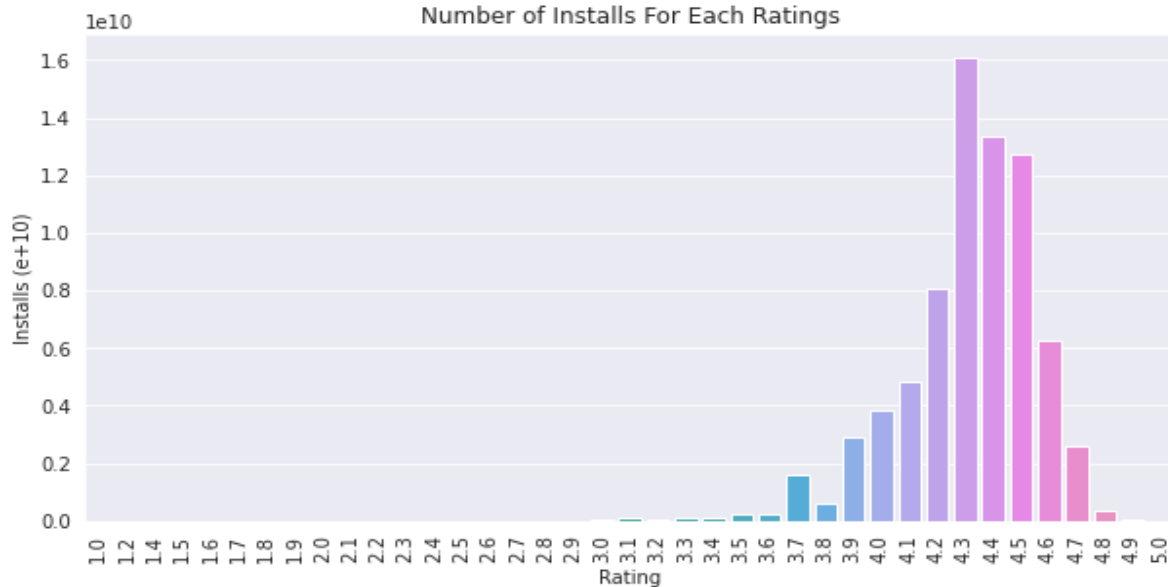
- Most of the apps in the Play Store are having rating higher than 4 or in the range of 4 to 4.7.



Number of Installs per Rating



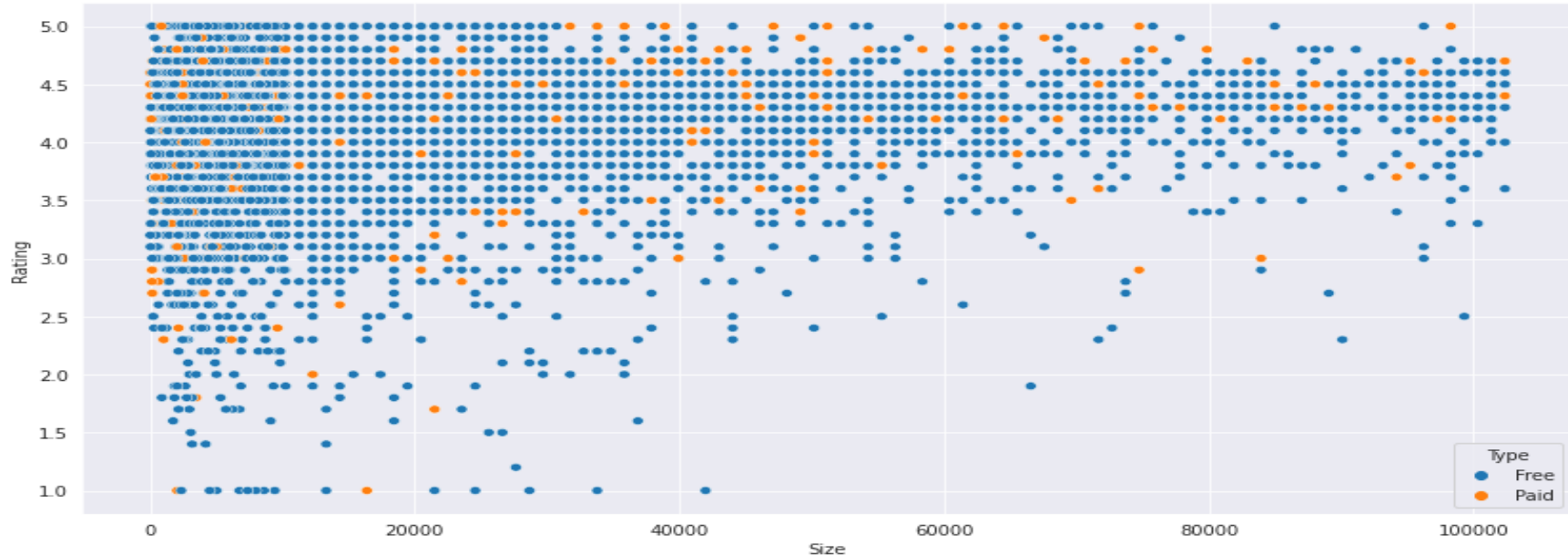
- Most of the apps downloaded by the customers are of higher rating (4.2 to 4.6 ratings).
- Ratings of 4.3, 4.4 & 4.5 have maximum number of installation.



Distribution of Apps in terms of their Ratings, Size and Type

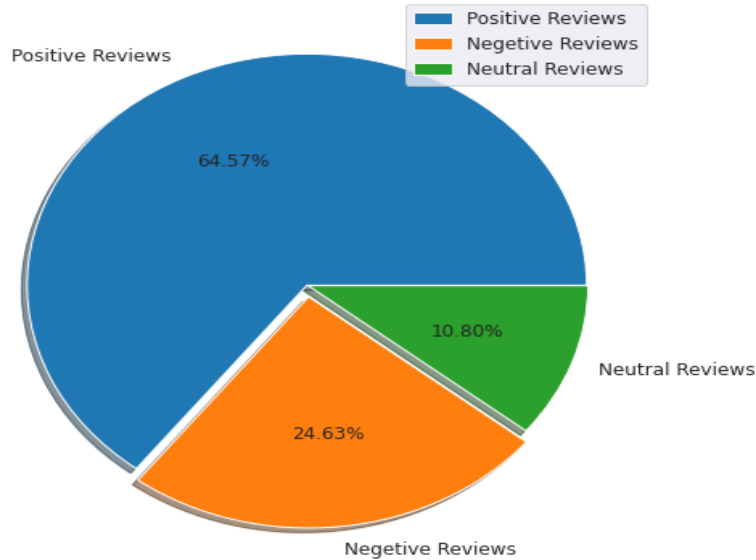


- The majority of the free apps are small in size and having high rating.
- While for paid apps, we have quite equal distribution in term on size and rating.



Reviews Sentiment

A Pie Chart Representing Percentage of Review Sentiments

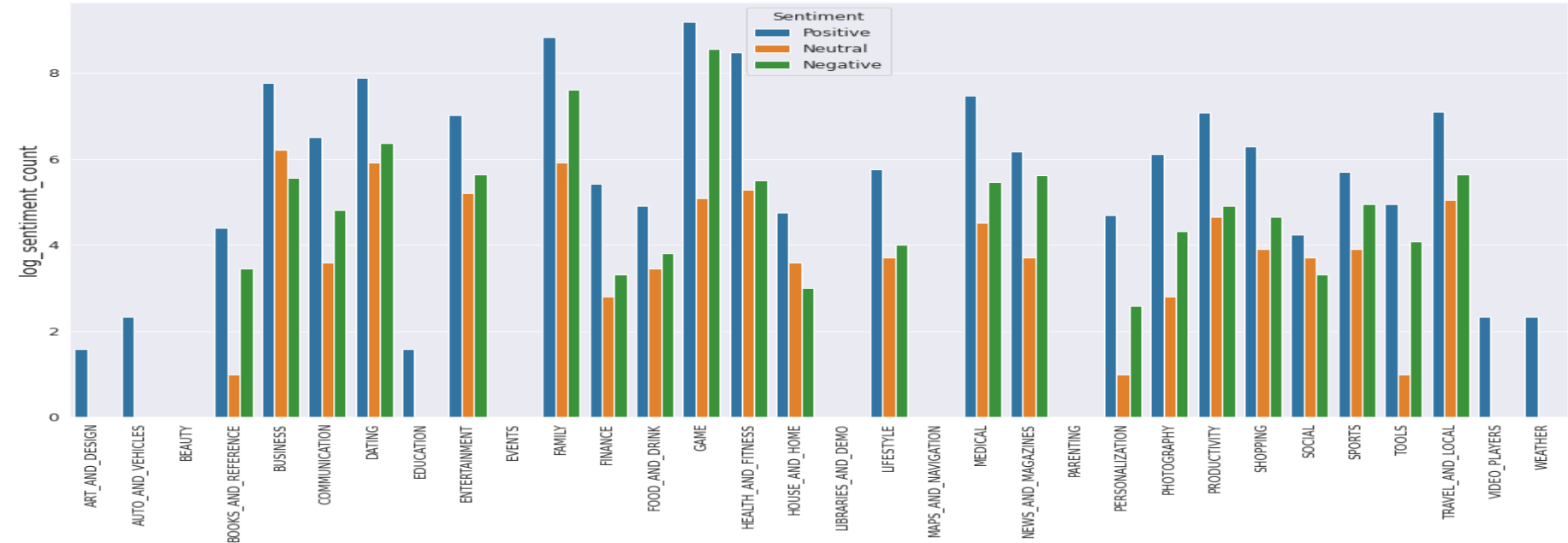


- Most of the reviews are positive in nature with **64.57%**
- Negative reviews are only **24.63%.**
- Neutral reviews are only **10.8%.**

Distribution of type of reviews as per categories AI

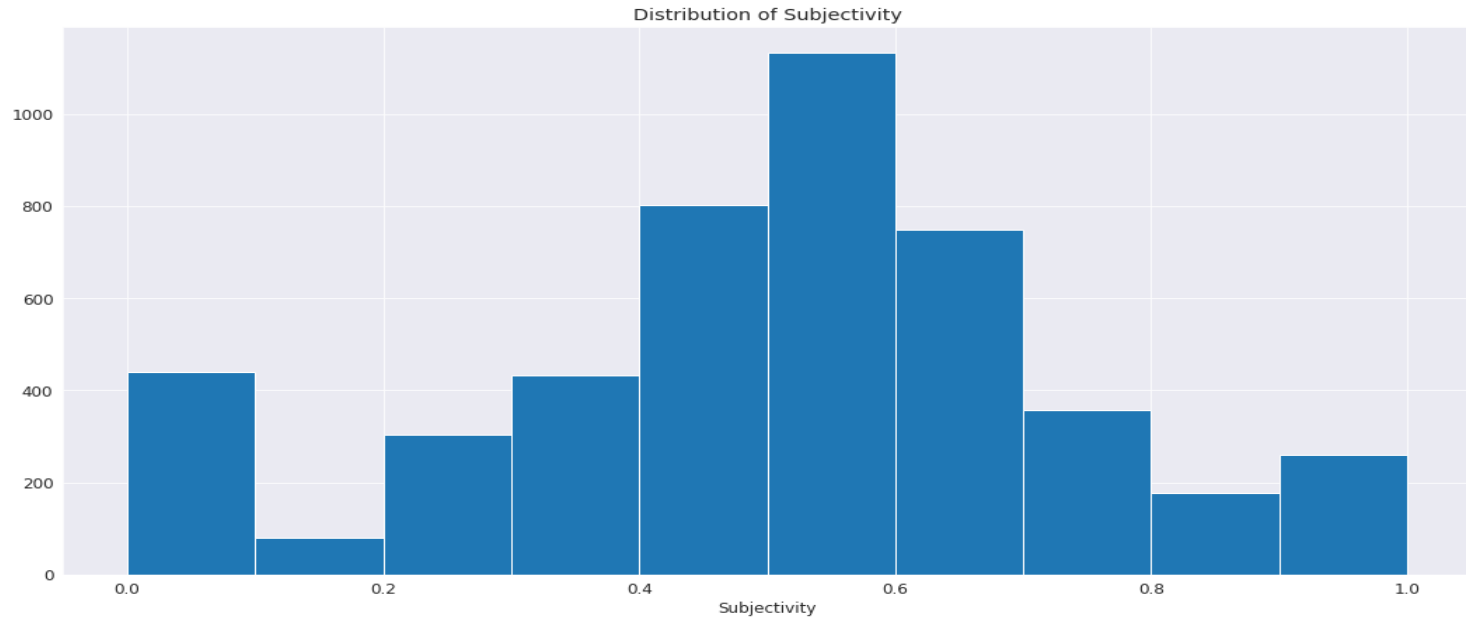
- Positive reviews are higher in each category of Play Store data,
- Some categories such as Game and Business have very less difference in the positive and negative reviews.
- Some categories such as House and Home and Business where neutral reviews are also very large as compared to negative reviews

Number of installs type wise according to Genres



Distribution of Subjectivity

- Maximum number of sentiment subjectivity lies between 0.4 to 0.7.
- These reviews comes from the experience from the users while using these apps.

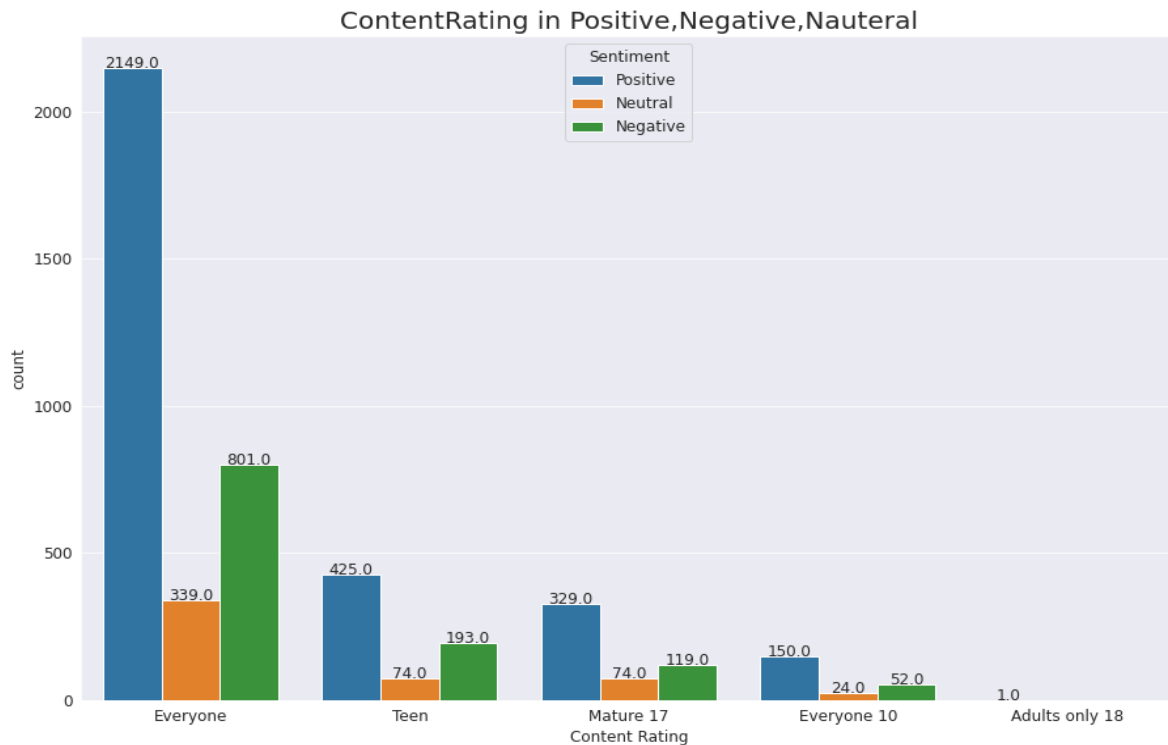


Sentiment Subjectivity and Sentiment Polarity AI



- Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of cases, it shows a proportional behaviour when variance is too high or low.

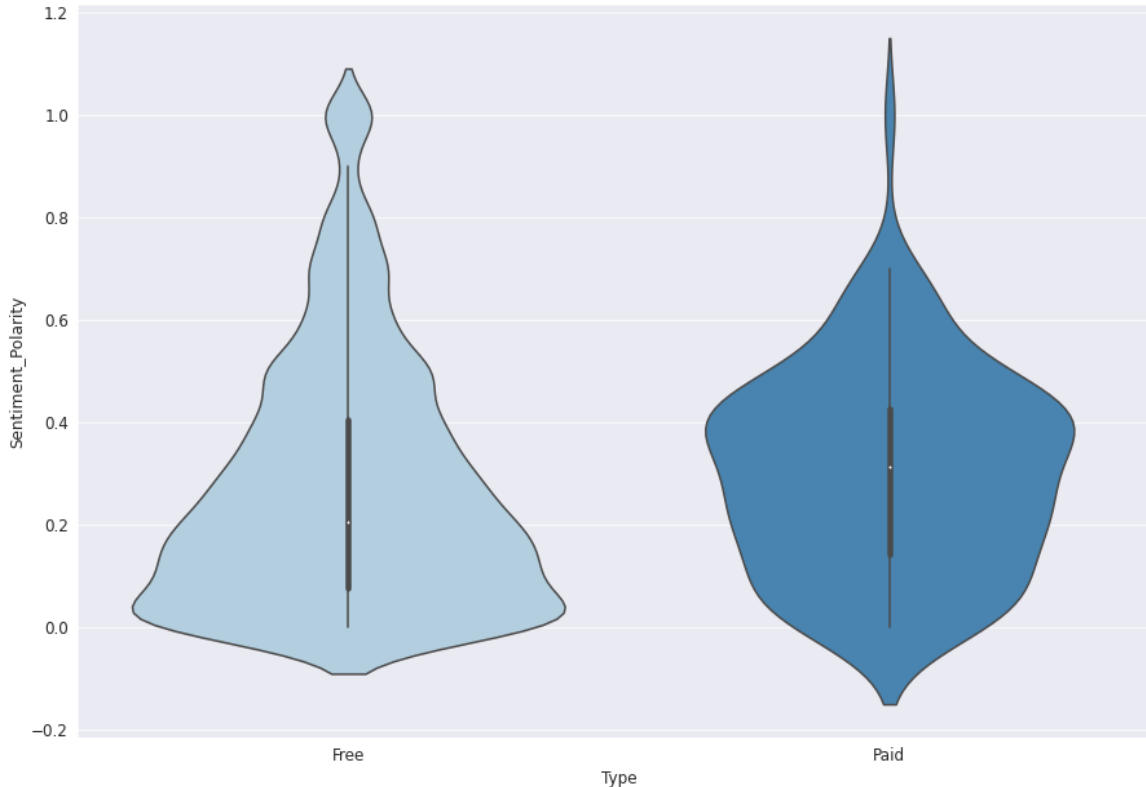
Content Rating on the basis of Age



- Most of the positive ratings came from the everyone category apps with 2149 ratings and 801 negative ratings.
- Neutral ratings are also low in comparison to positive and negative reviews.

Sentiment polarity relation with type AI

of app

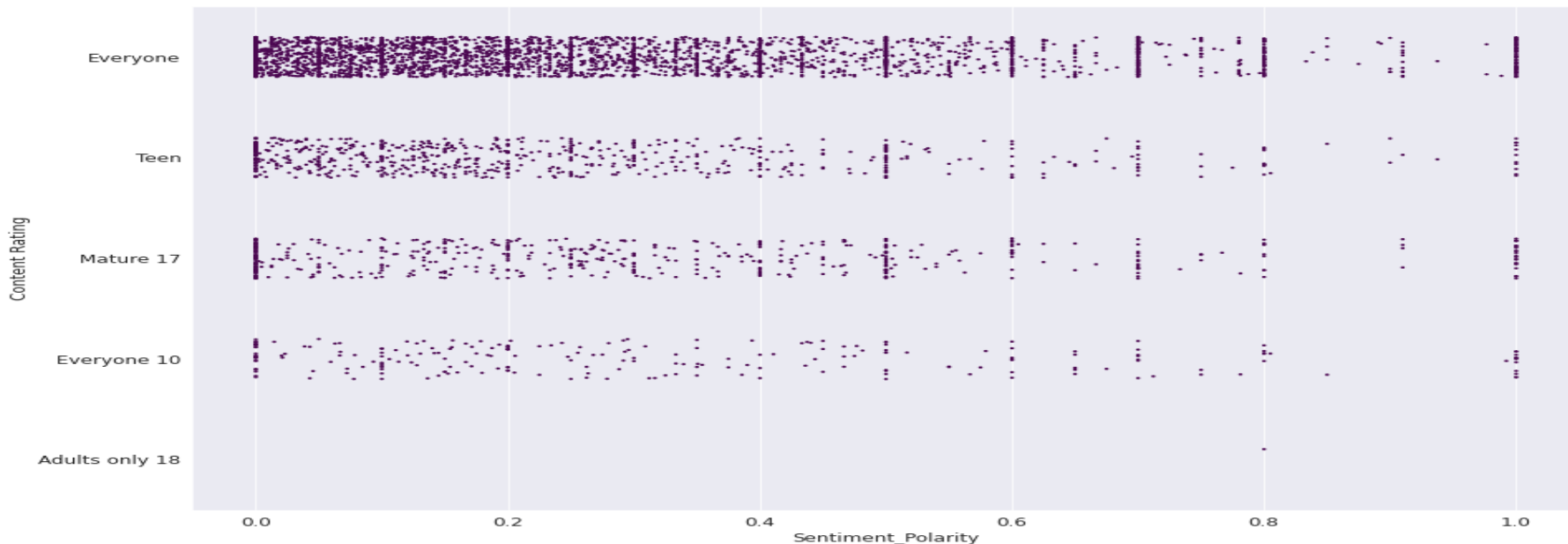


- In Free apps the sentiment polarity lies majorly in 0.1 and decreases after that.
- The sentiment polarity of paid apps which falls majorly between 0.1 to 0.4 with highest in 0.4.

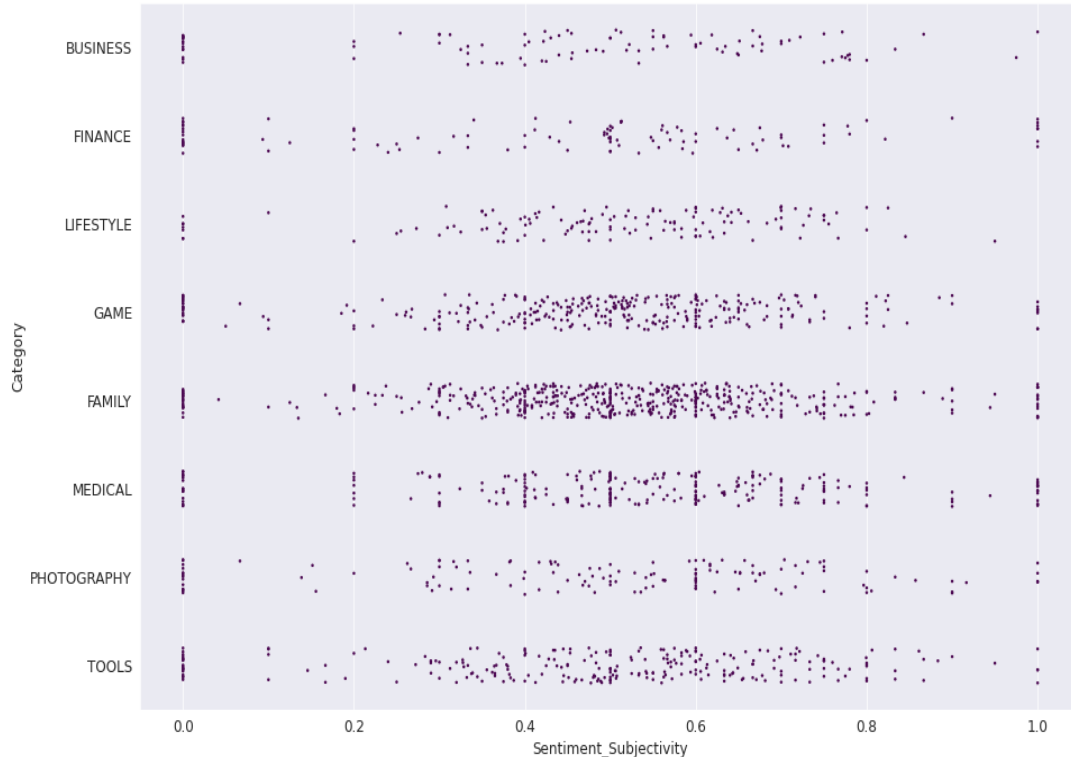
Content Rating relation with Sentiment Polarity



- Sentiment polarity is low in most of the categories except for Everyone category.
- Polarity is evenly distributed.



Categories Relation with Sentiment Subjectivity



- Subjectivity lies mostly between 0.2 to 0.8 in all Categories.
- Family Category has the higher number of reviews in which Subjectivity lies between 0.4 to 0.6

Asking questions from Dataset

1) What is the top 5 apps on the basis of installs?

3232	Google Street View	TRAVEL_AND_LOCAL	4.2	2129707	12288.0	1000000000	Free	0.0	Everyone
3909	Instagram	SOCIAL	4.5	66509917	12288.0	1000000000	Free	0.0	Teen
3687	Google Play Movies & TV	VIDEO_PLAYERS	3.7	906384	12288.0	1000000000	Free	0.0	Teen
5395	Google Photos	PHOTOGRAPHY	4.5	10847682	12288.0	1000000000	Free	0.0	Everyone
3904	WhatsApp Messenger	COMMUNICATION	4.4	69109672	12288.0	1000000000	Free	0.0	Everyone

The 5 apps with the most number of installs are:
Google Street View, Instagram, Google Play Movies & TV, Google Photos, WhatsApp Messenger

Asking questions from Dataset

2) What is the top 5 reviewed apps?



2. What is the top 5 reviewed apps?

```
df = play_store_3.groupby(by=['App', 'Category', 'Rating'])[['Reviews']].sum().reset_index()
df = df.sort_values(by=['Reviews'], ascending=False)
df.head(5)
```



	App	Category	Rating	Reviews
4325	Facebook	SOCIAL	4.1	78128208
9034	WhatsApp Messenger	COMMUNICATION	4.4	69109672
5395	Instagram	SOCIAL	4.5	66509917
6161	Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56646578
2561	Clash of Clans	FAMILY	4.6	44881447

The 5 apps that have the most number of total reviews are: Facebook, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free, Clash of Clans

Asking questions from Dataset

3) What is the top 5 expensive apps?

The top 5 most expensive apps in the store are: I'm Rich - Trump Edition, I am rich(premium), I'm Rich/Eu sou Rico/ أنا /غني 我很有錢, I am Rich Plus, I am Rich!

4) What is the top 3 most installed apps in Game category?

The top 3 most installed apps in the GAME category are:
Subway Surfers, Temple Run 2, Score! Hero

5) Which 5 apps from the 'FAMILY' category are having the lowest rating?

The 5 apps from the FAMILY category having the lowest rating are: FE Mechanical Engineering Prep, Truck Driving Test Class 3 BC, Speech Therapy: F, AC REMOTE UNIVERSAL-PRO, BG TV Ap

Conclusion

- ❖ Most of the trending apps are from the categories like family, tools, and games.
- ❖ Most preferred Apps by users in point of size or weight are light size apps.
- ❖ Users also installed apps on the basis of their ratings.
- ❖ These ratings are defined on 2 points- Polarity and Subjectivity.
- ❖ This analysis will help developers while preparing for their next apps.