Capstone Project

Play store app review analysis

Problem statement

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.

Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.

Explore and analyse the data to discover key factors responsible for app engagement and success.

Data summary

The dataset spans over nine years :
2010,2011,2012,2013,2014,2015,2016,2017,2018.

Important features:

1] App :- Name of the application.

2]Category :- The category to which an application belongs to

3]Content Rating:- Content rating tells us the app is available for everyone or it belongs to the people of particular age group.

4]Type:- An application is free or paid.

5]Reviews:- People reactions corresponding to an application.

6]Last update:- On which date an application receives its latest update.

7]Rating:- Users can rate your app on a scale of one to five stars.

Content

- a

Analysis based on :

- Category
- App
- Installs
- Ratings
- Reviews
- Size
- Type
- Content rating
- Sentiments

Refinement of dataset

Refining a dataset is the most crucial stage in data analysis. Refinement of a dataset refers to the process of cleaning data and handling the null values/ missing data. If your dataset is not properly refined then you may not achieve the optimum results from the dataset. Unrefined data were more prone to creating errors in the results. Sometimes you get an error or gets the wrong results.

Refine the dataset:

Removing Duplicate Apps

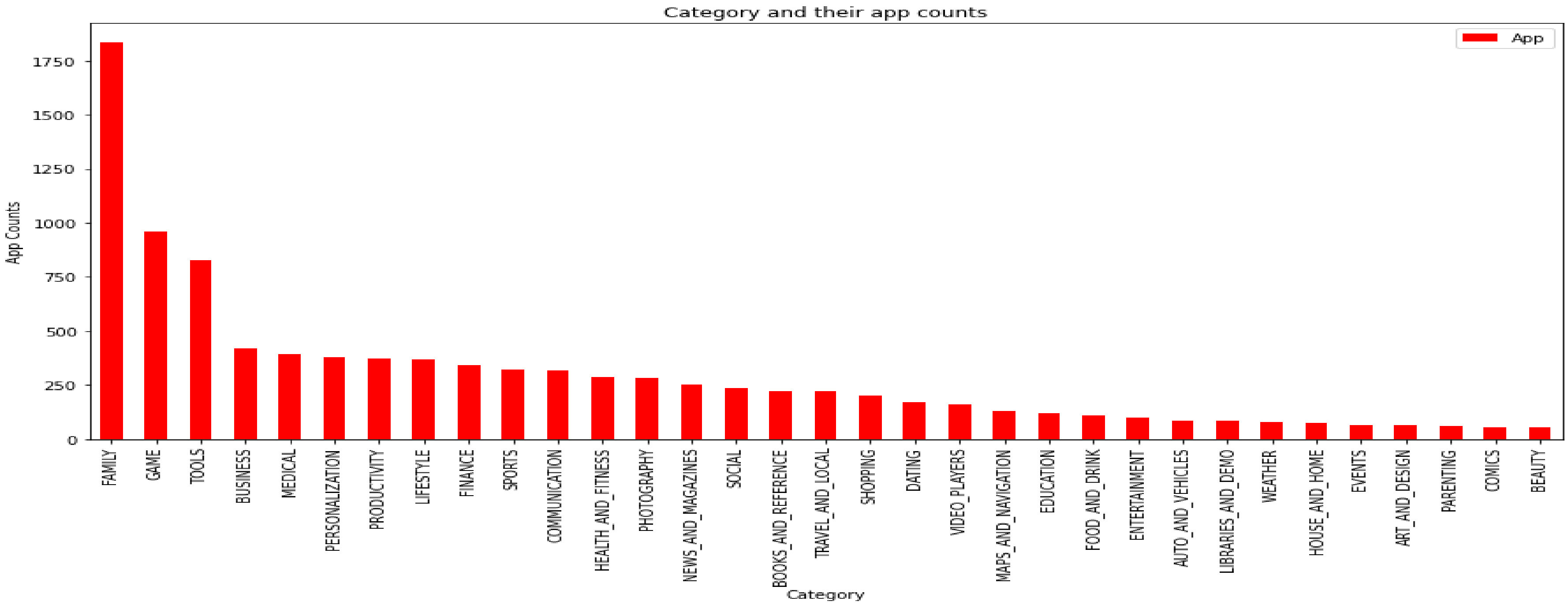
Removing 19.0 from the column rating

Removing 1.9 from the column category

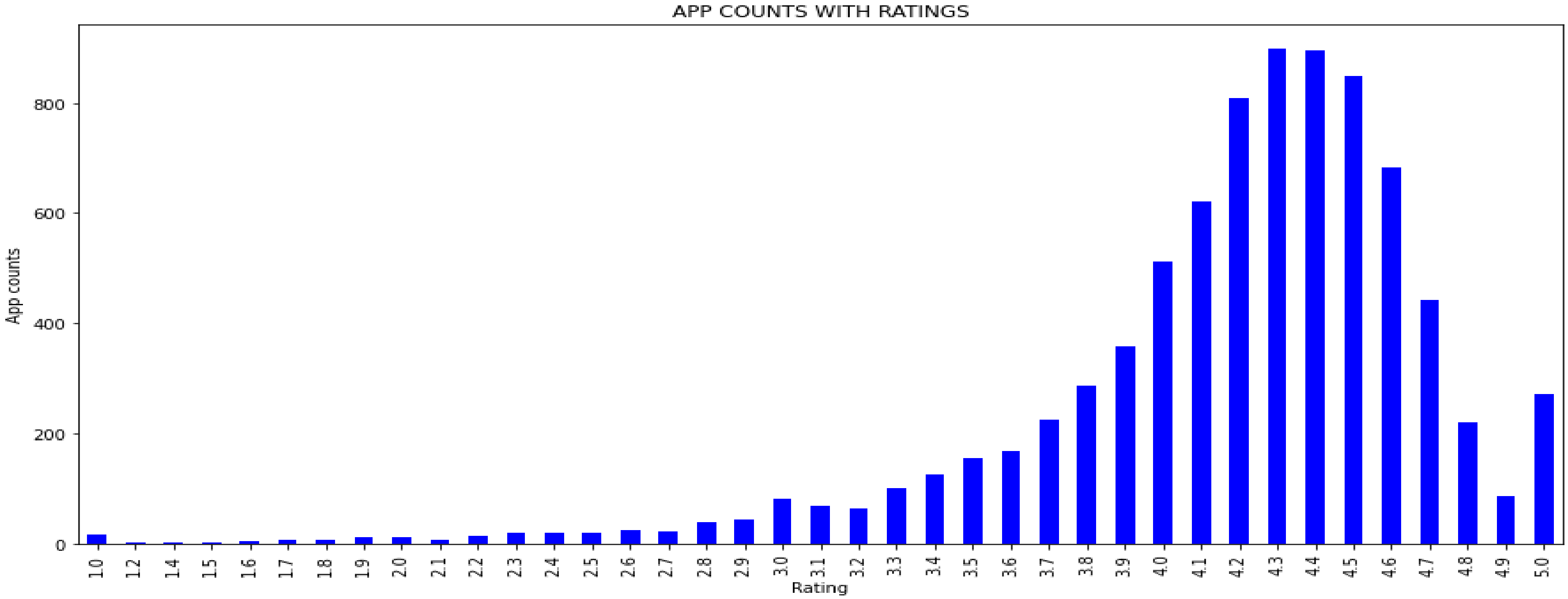
Convert the dates in datetime format.

Converting the data of install column into numeric data.

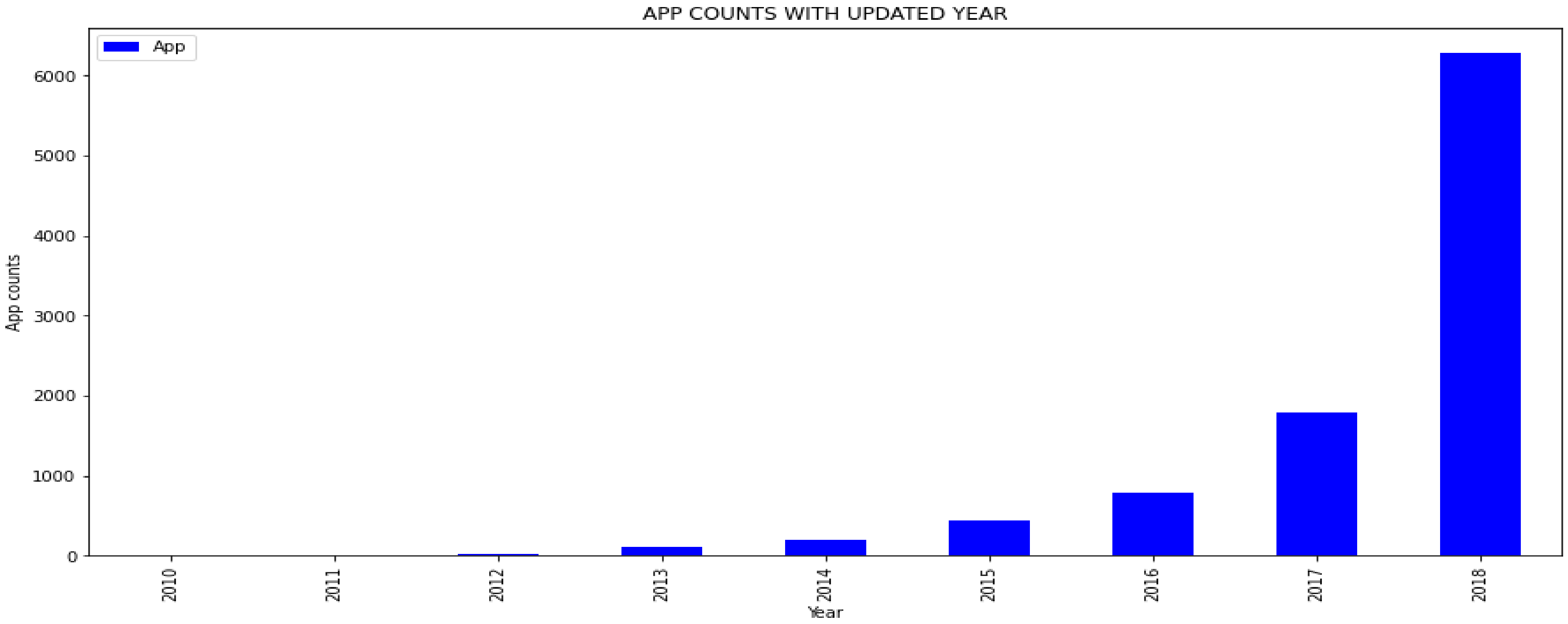
Category and their app counts



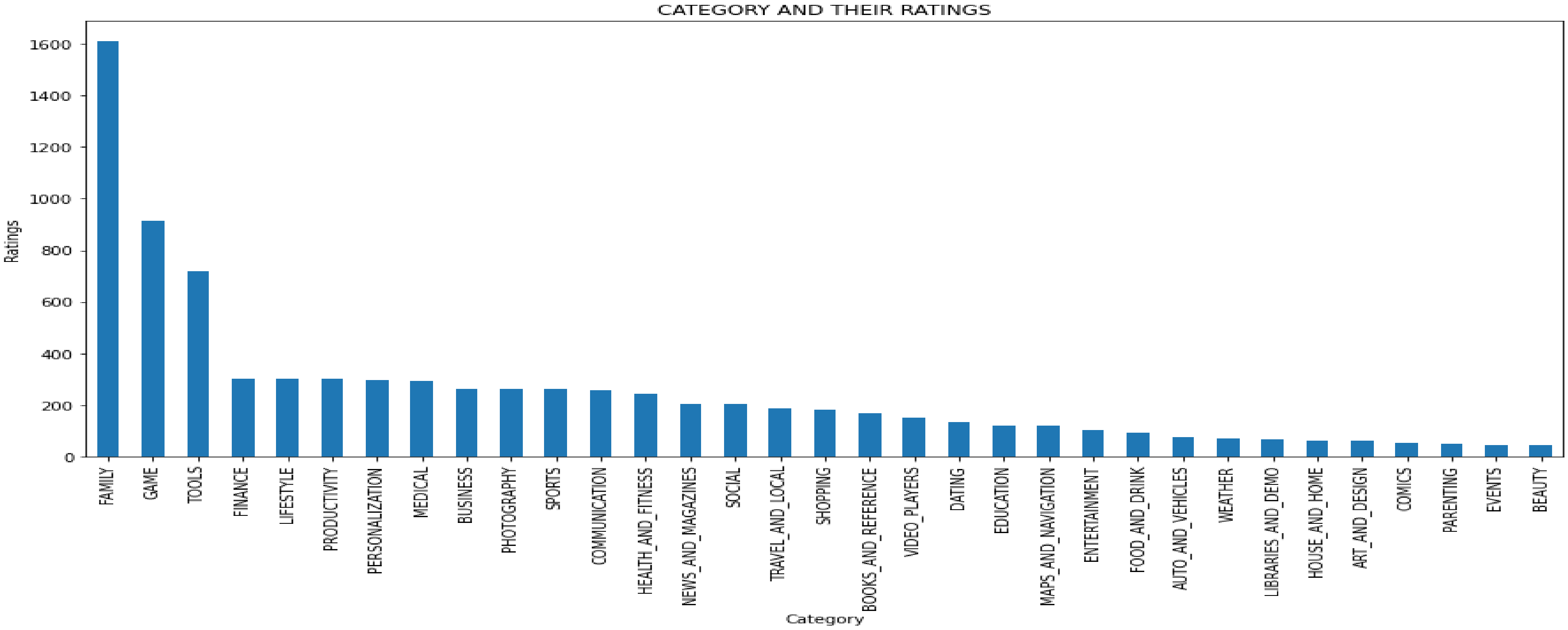
App counts with ratings



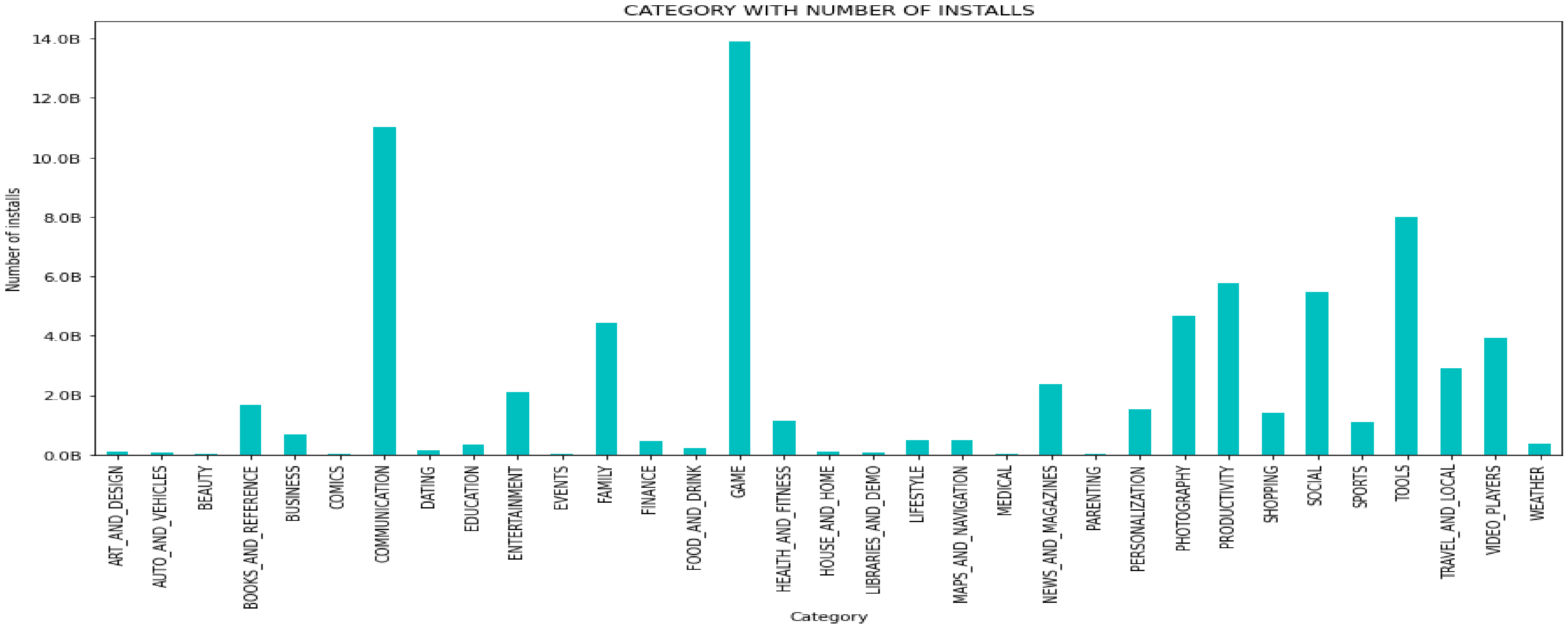
App counts with updated year



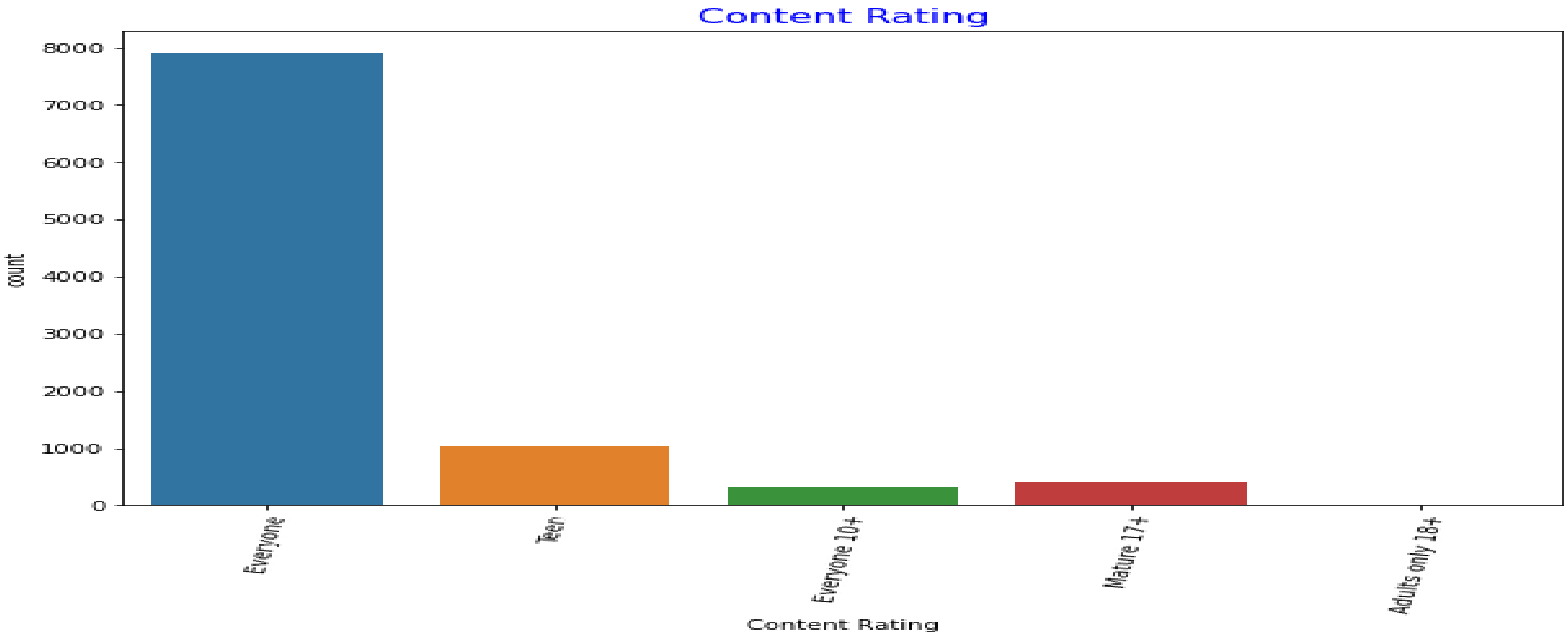
Category and their ratings



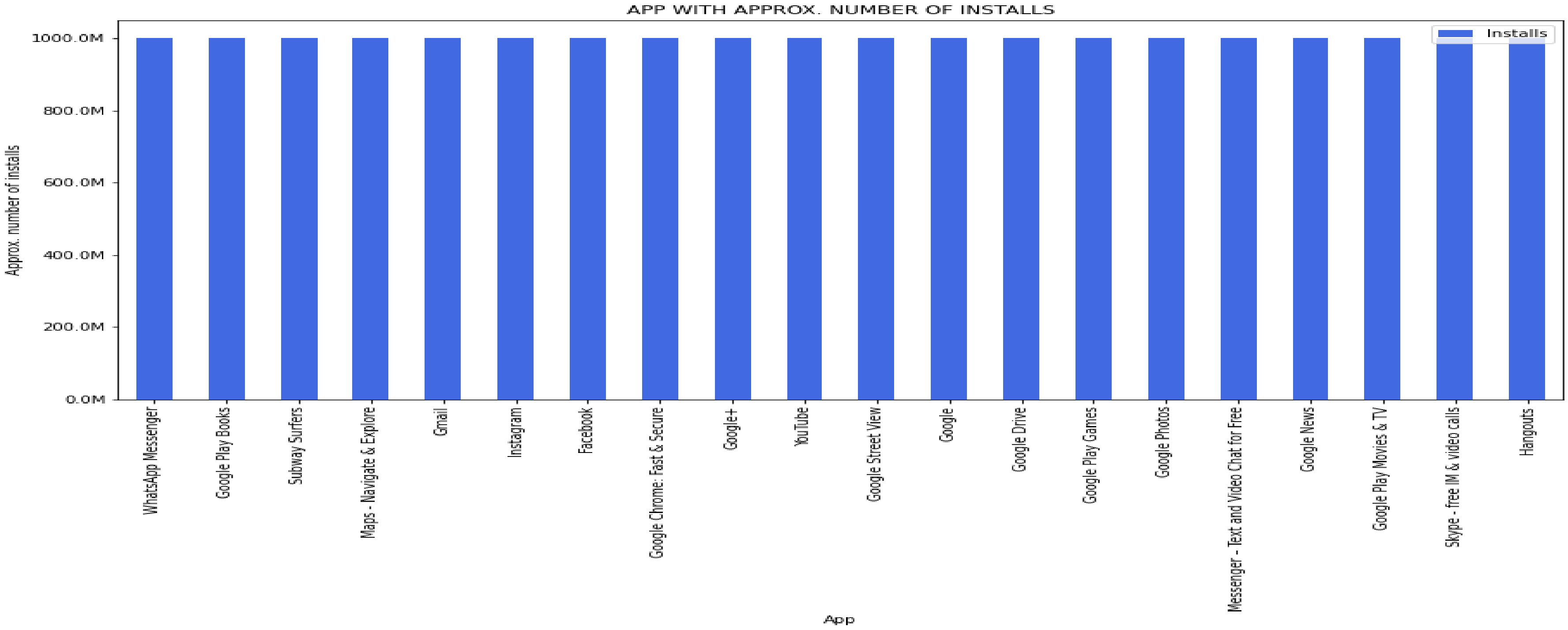
Category with number of installs



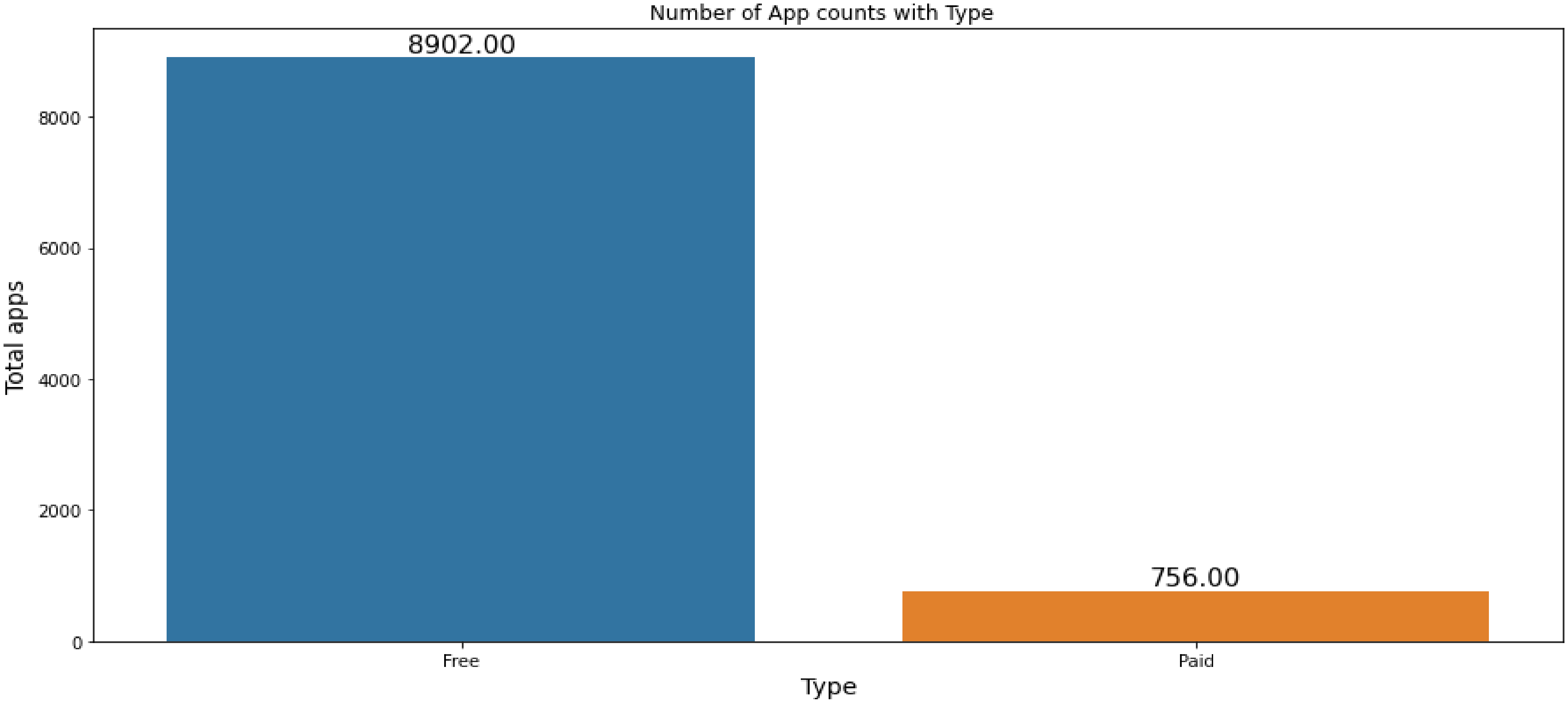
Content rating and their counts



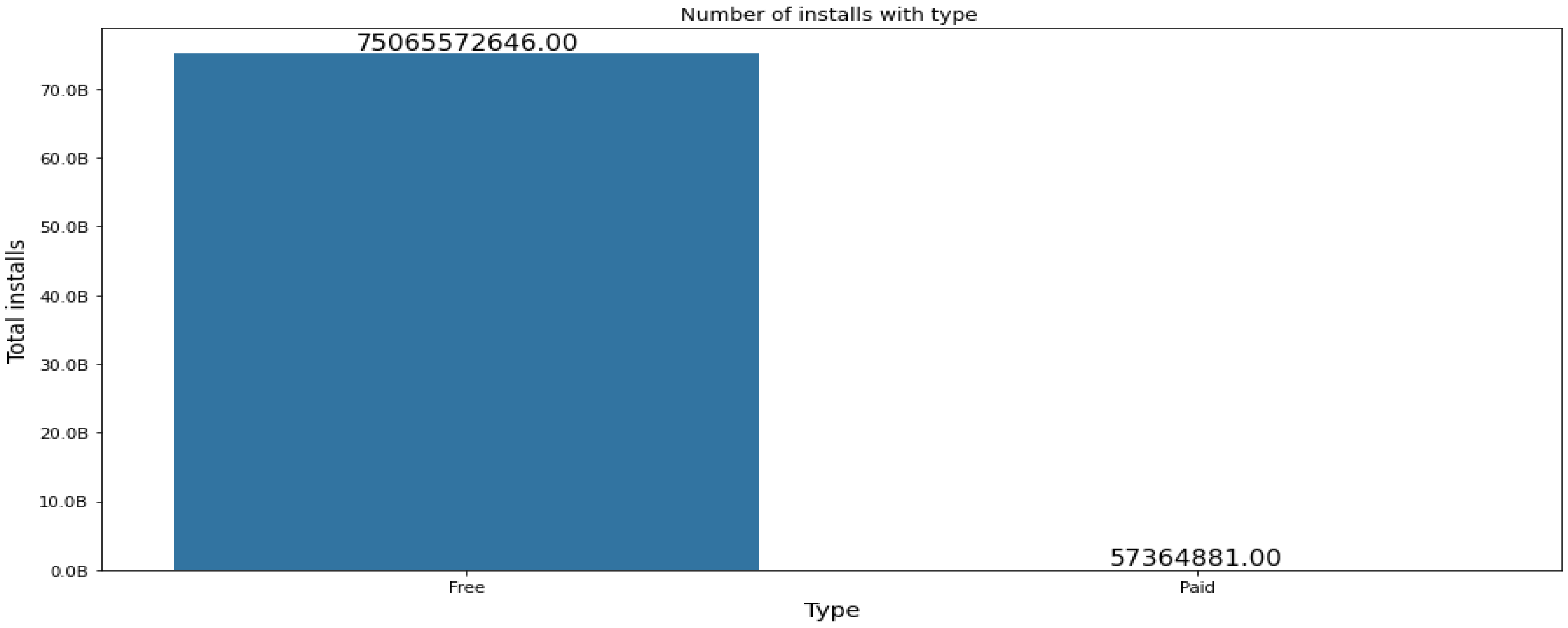
App with approx. installs



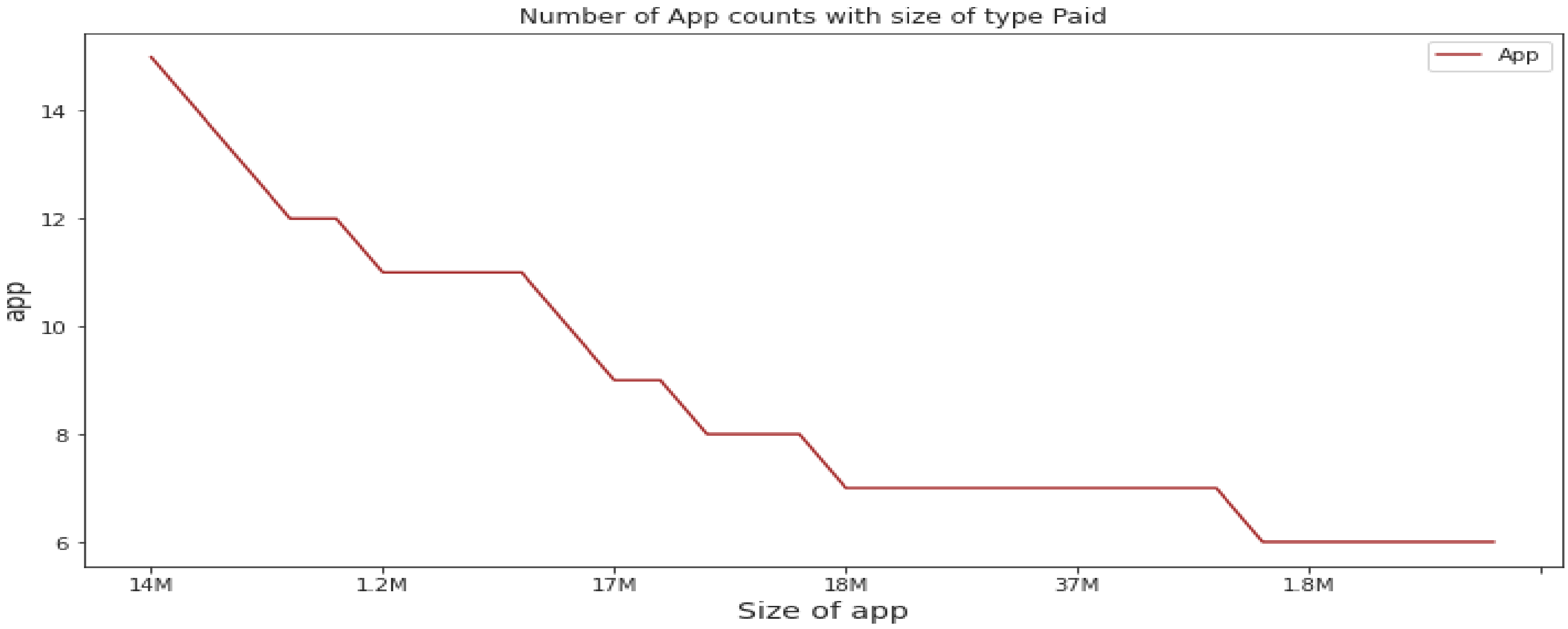
Number of app counts with type



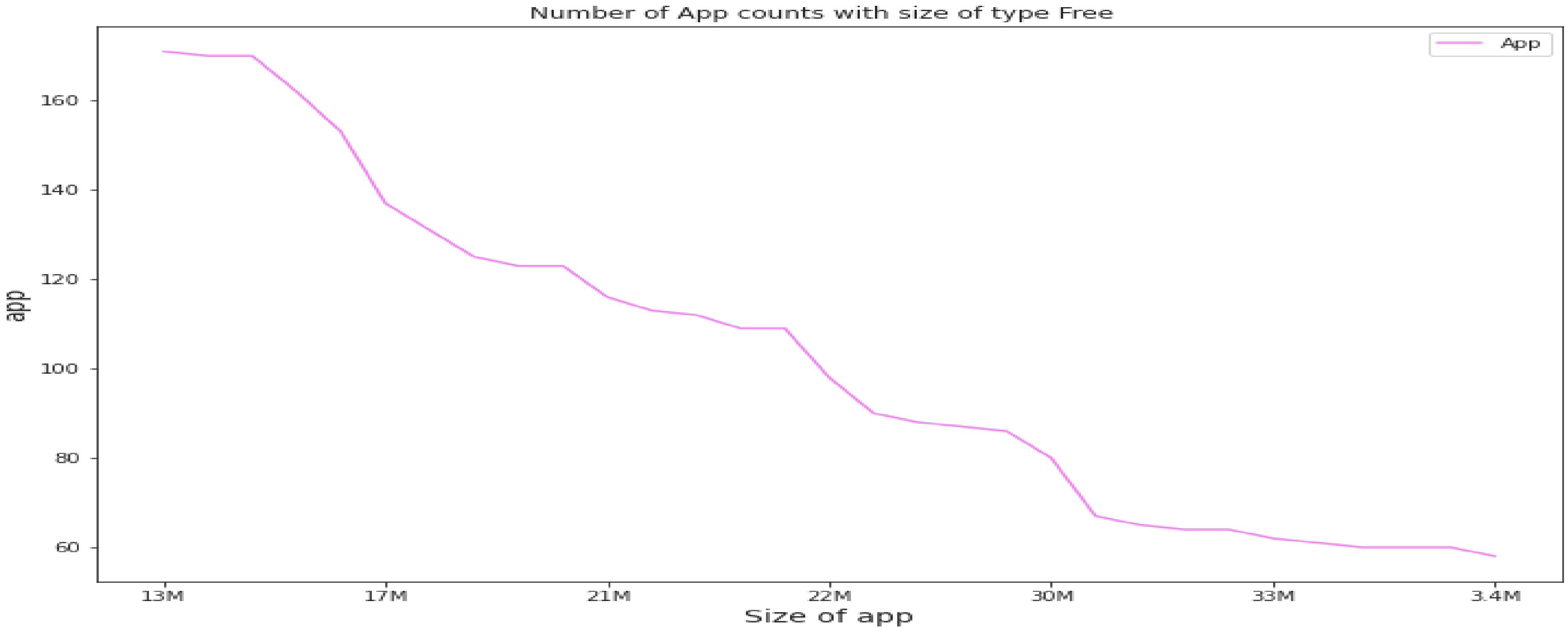
Number of installs with types



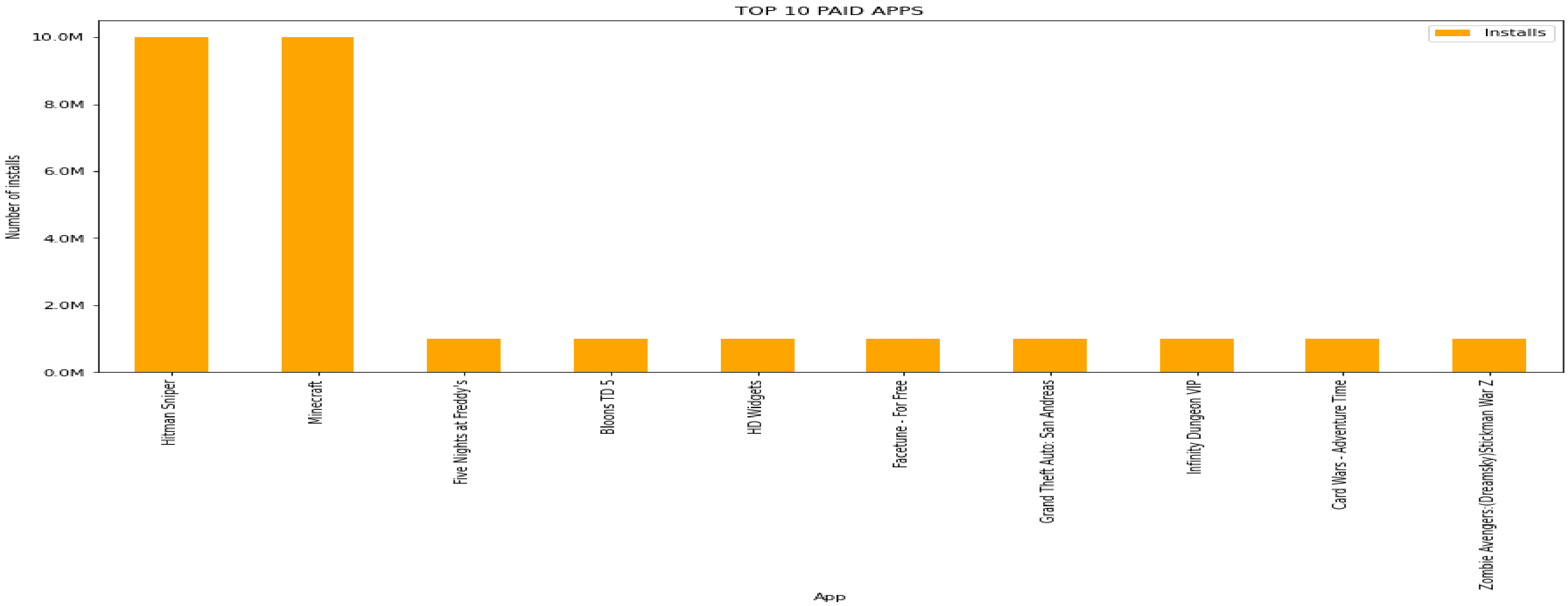
Number of app counts with size of type paid



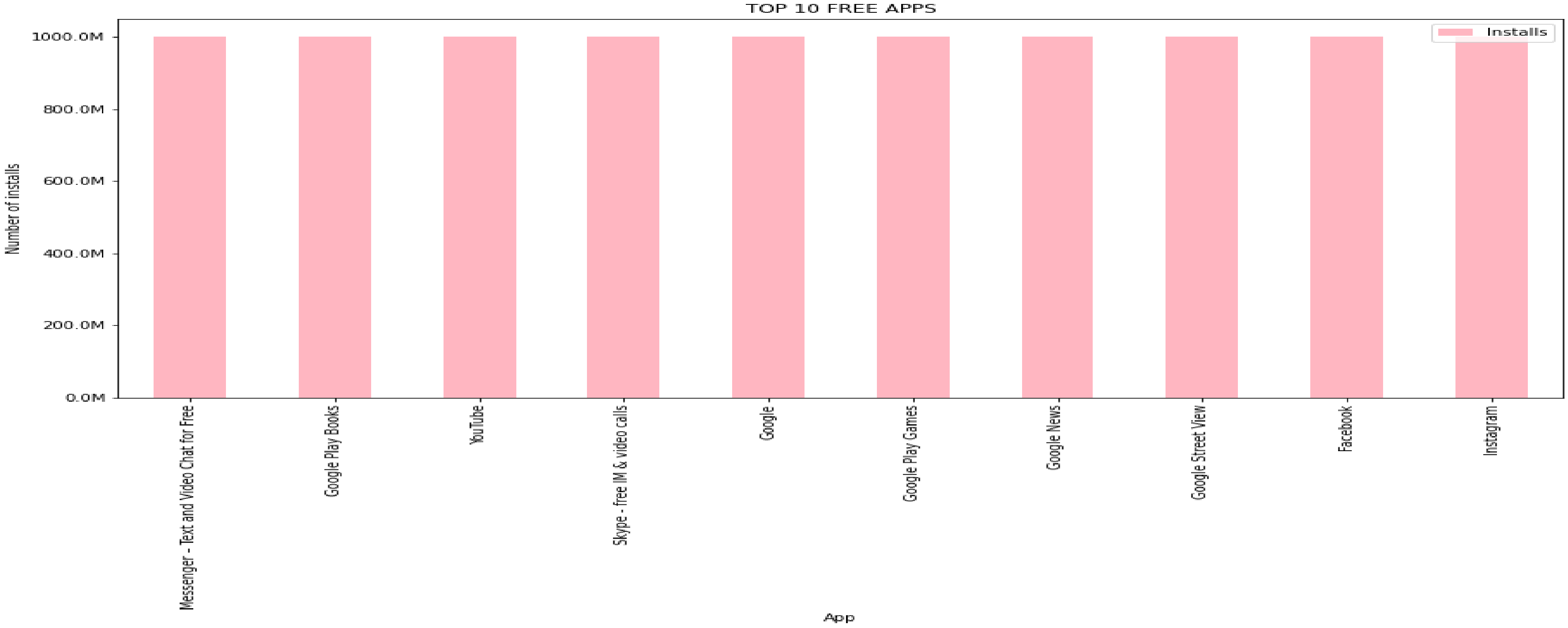
Number of app counts with size of type free



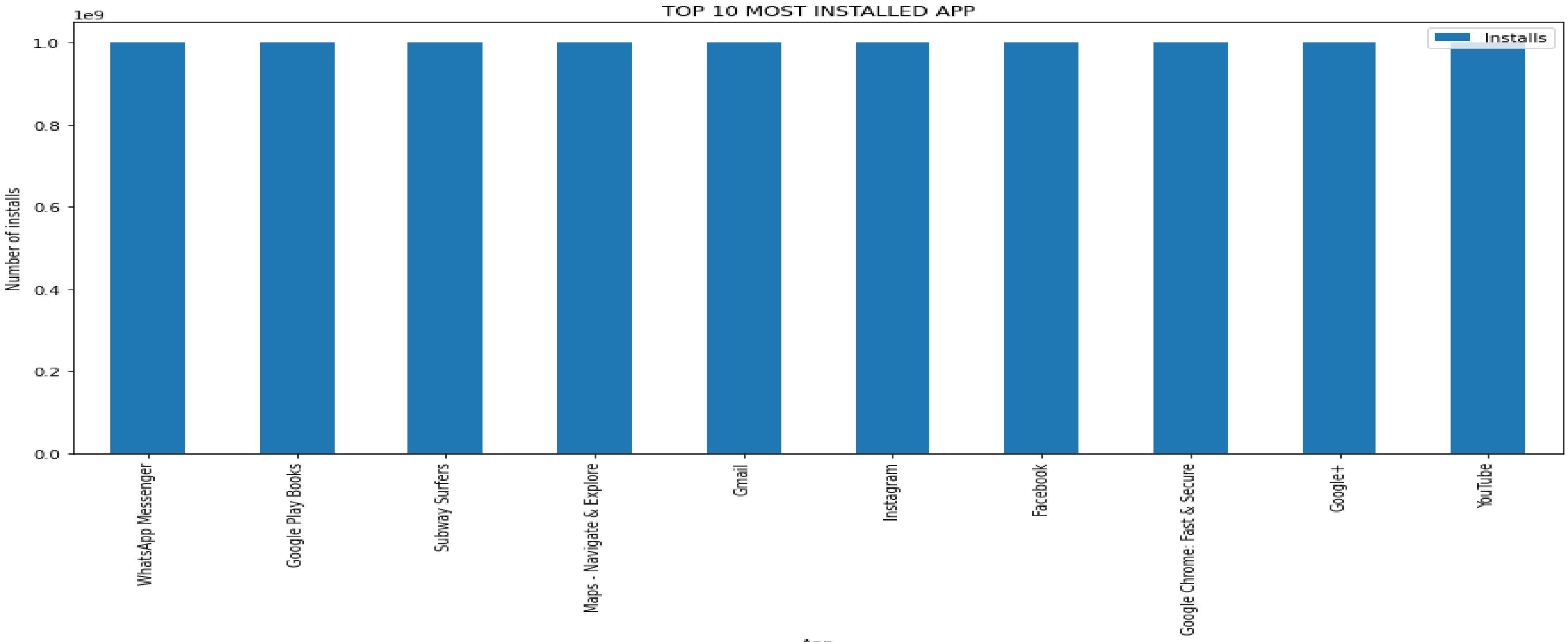
Top 10 paid apps



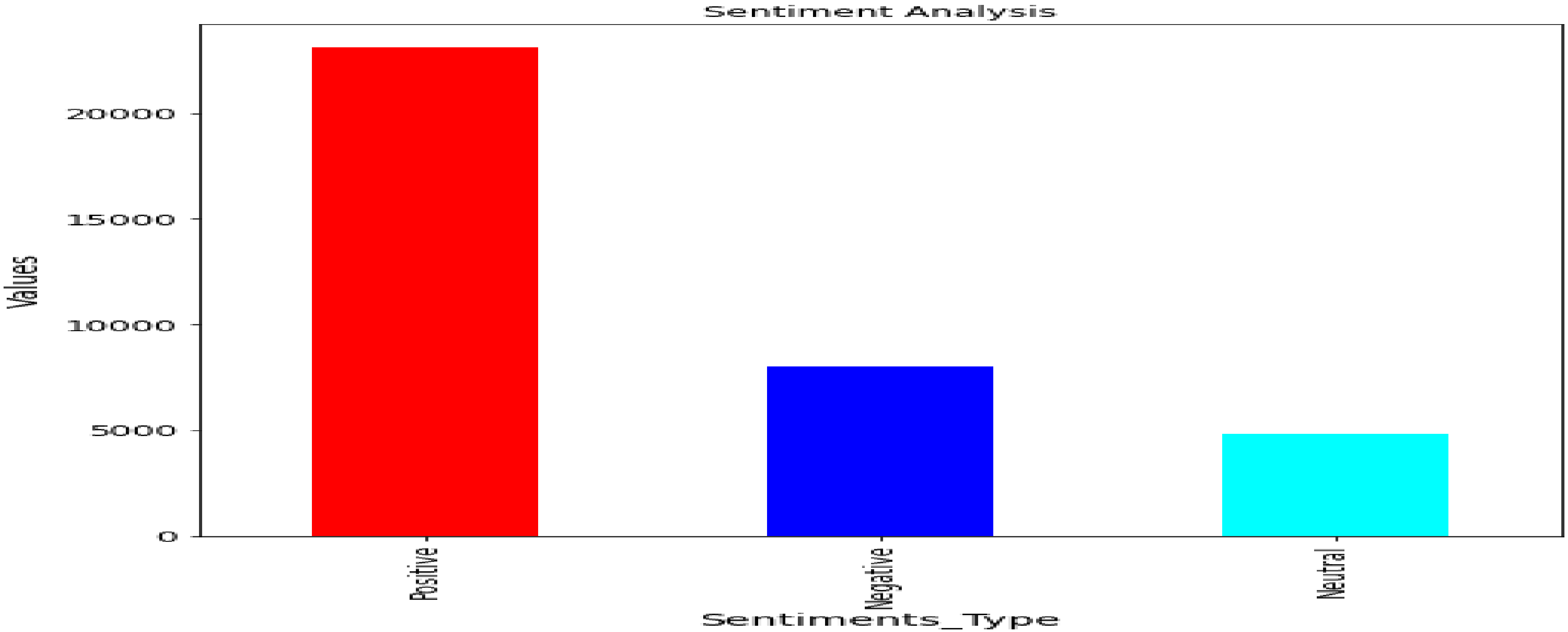
Top 10 free apps



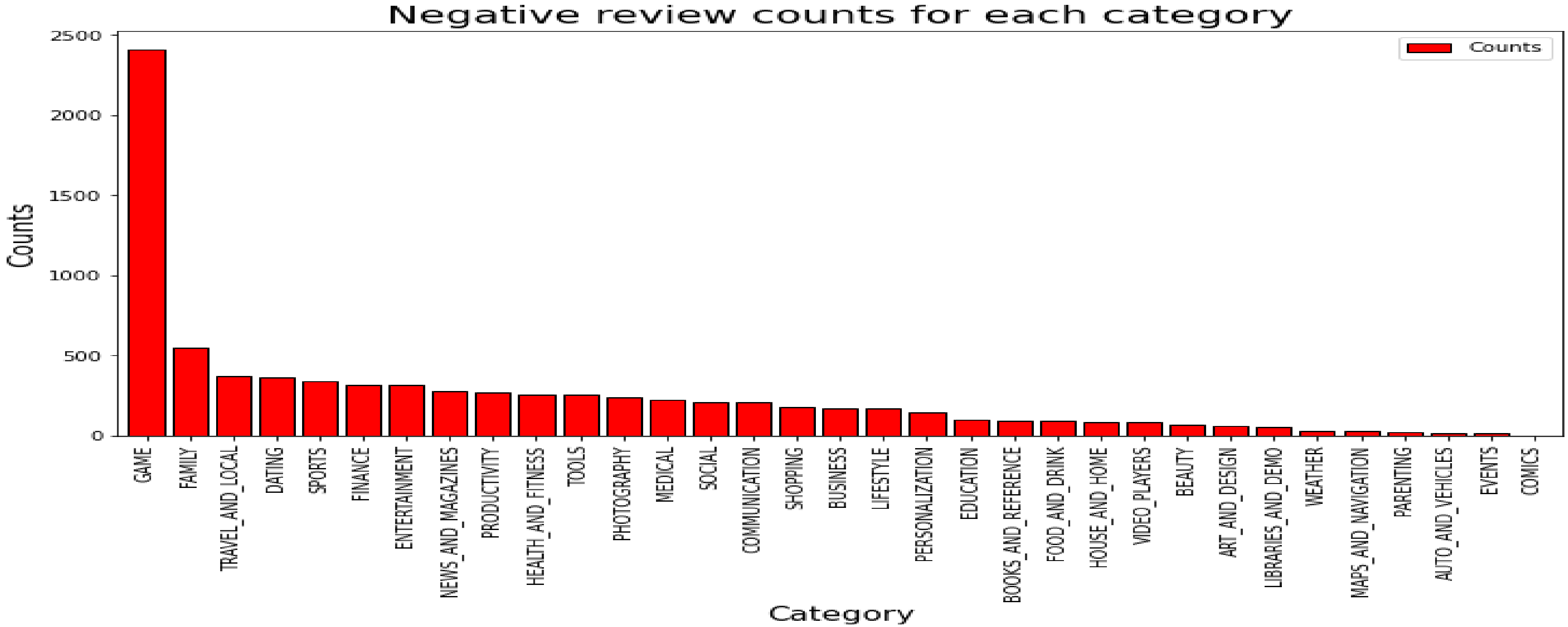
Top 10 most installed apps



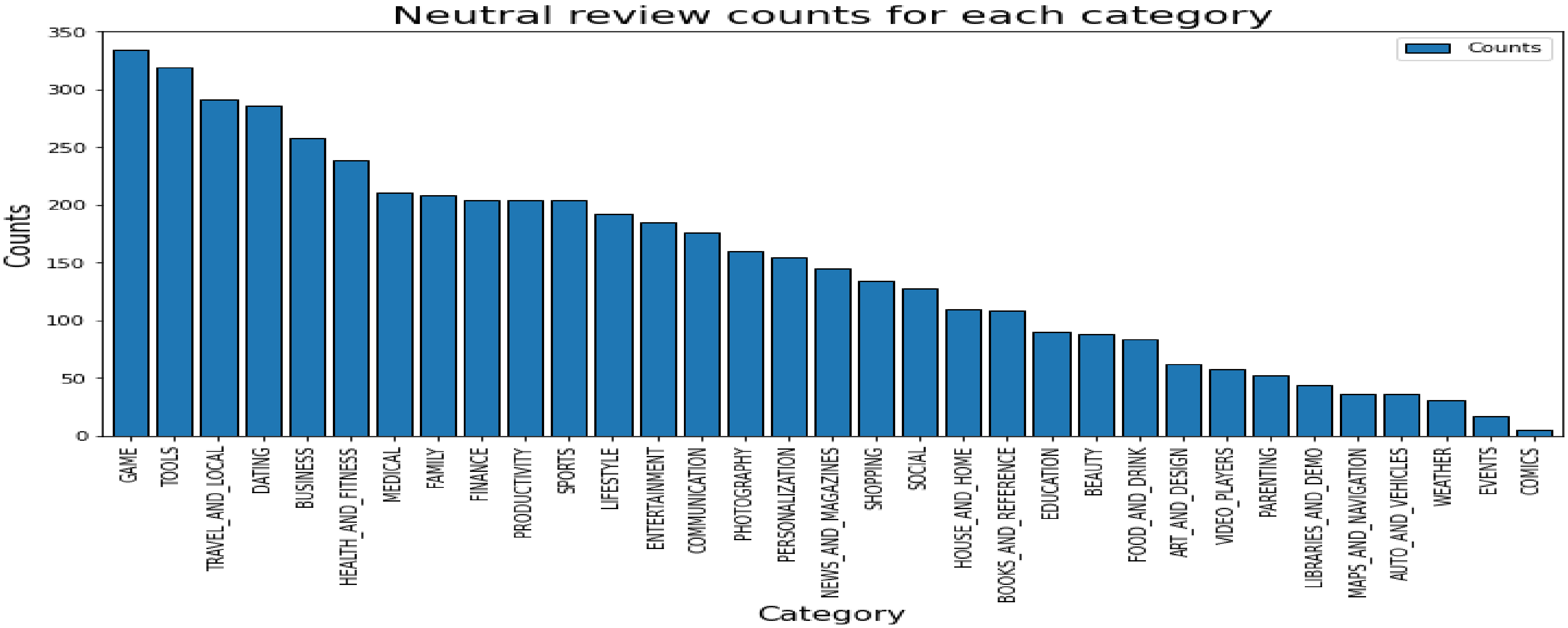
Sentiment Analysis



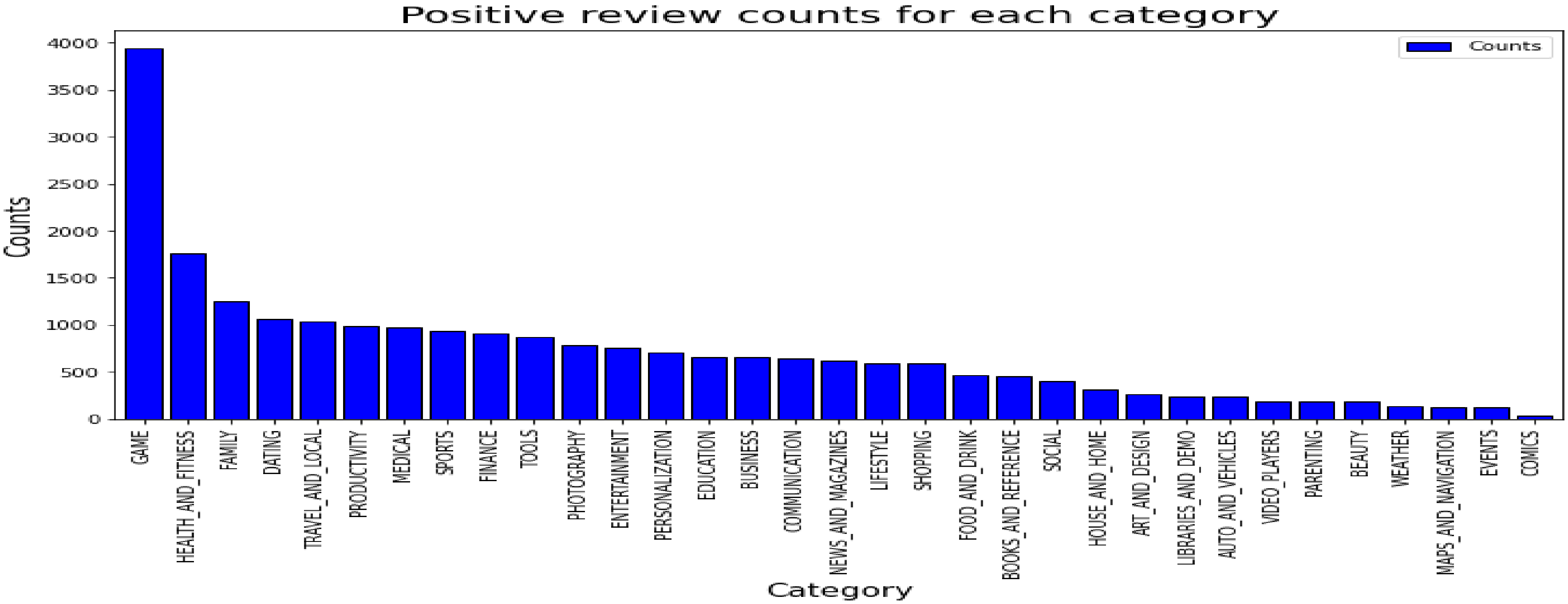
Negative review counts for each category



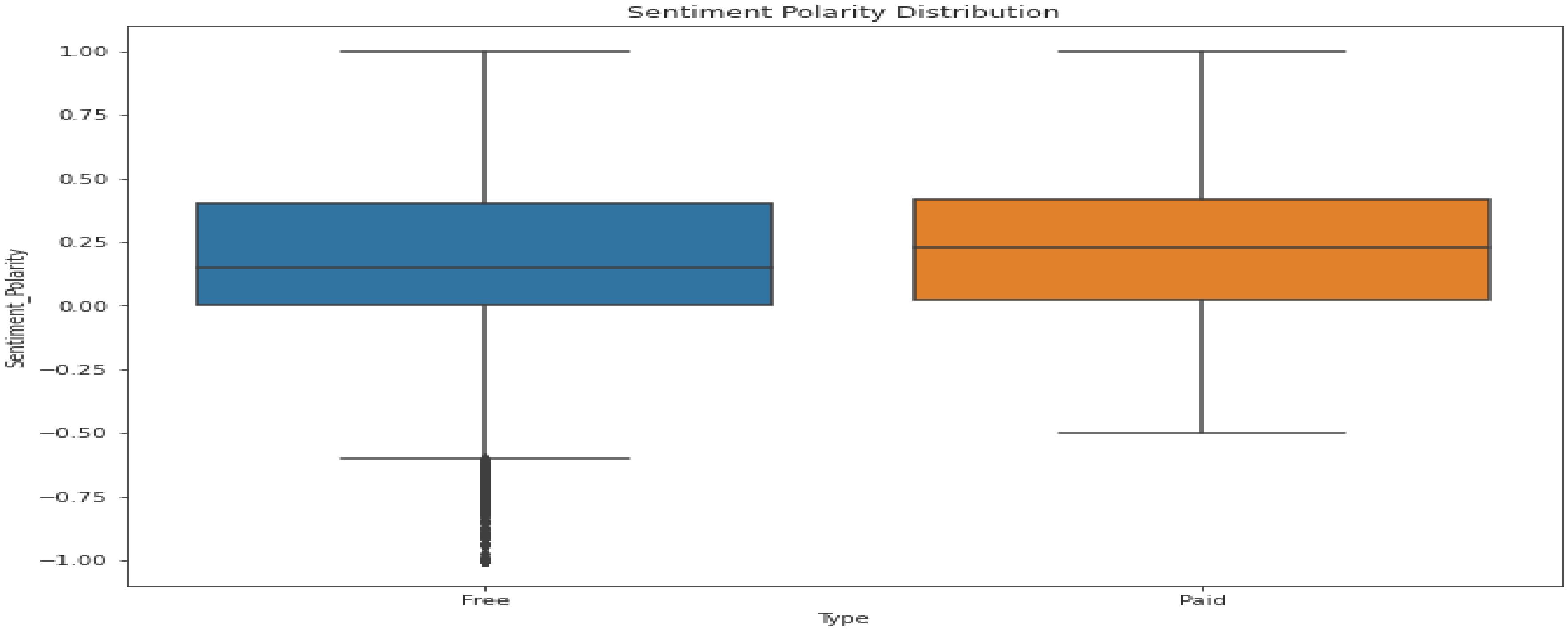
Neutral review counts for each category



Positive review counts for each category



Sentiment polarity distribution



Challenges

- ~ The biggest challenge is to find the dataset is redundant or not. The size of data set is large so it takes lots of time to understand the data is redundant or not.
- ~ The sentiment polarity and sentiment subjective cause troubles for us. It is not easy how to use that data and which plot is suitable for that type of data.
- ~ The number of installs in the data is 10,000,000+, most of the apps has the same number. It is quite confusing for us to take which app or not.
- ~ In case of size, more than 1000 apps size varies with devices because of that we cannot consider these apps in some cases to refine our results.

Observation

~ There is a total of 1832 apps that are from the family category which is the most and after that game category is at second with 959 apps.

~ Here we can see 897 apps are having a 4.4 rating which is the most and after that, in the second place 895 apps are having a 4.3 rating, most apps are having 4.3 ratings. Most of the apps having ratings between 3.5 to 4.8.

~ Here we can see the family category is having the most ratings which are 1608 and second there is the game category with 912 ratings.

~ Here we can see the game category is having the most number of installs and after that communication category is having the second most number of installs.

~Out of 9659 apps 6284 receive updates in the year 2018.

~ Here we can see that the 7903 app belongs to the content rating everyone and 1036 is for teens. For Mature 17+ there are 393 apps. Only 3 apps fall under the 18+ category.

~ Here, we can see the apps with 1000000000+ installs.

~ Mostly the size of free app lying between 2.3M to 33M. There are 171 app of size 13M.

~ The size of paid apps lying between 1.2M to 33M.

~ Here we can see there is a total of 8902 apps that are free and 756 apps are paid.

~ Here we can see a total of 75065572646 free apps are installed and a total of 57364881 paid apps are installed.

~ Here we can see the paid apps with the number of installs.

~ We observe that 23073 peoples share the positive reviews and 8005 people share the negative reviews of the apps. And, 4851 peoples share neutral reviews for the apps.

~ Games category dominates in terms of positive review.

~ Comics category revises very less amount of reviews. It falls at last in terms of review. But, there is no negative reviews for comics.

~ The application Helix Jump has 209 positive reviews.

~ The application Calorie Counter - MyFitnessPal has 169 negative reviews.

~ From the Boxplot, we can analyse that the free app receives many harsh comments of very high negativity, which can be analysed by the outliers on the negative y-axis. The paid apps also receive harsh comments but they are not extreme negative as free apps.

~ This may demonstrate something about application quality, i.e., paid applications being more excellent than free applications all things considered.

Conclusion

- The major key factor responsible for engagement and success is the size of the app, category, content rating, type, reviews, and rating. As we observe in the above graphs, the family and games category dominates the app counts. Most of the apps belong to these two categories. Similarly, the maximum app rating lying between 3.4 to 4.8. Also, the apps that are of type free are more popular than paid apps. Updates play an important role in the success of an app because the update provides the latest features and removes the bugs. Those apps whose sizes are less can be used on all devices. How many people installed the app is also a major factor because if more people install the app then it means more people give ratings and reviews for the apps. Games category app receives the highest positive reviews. But, the quality of paid apps is better than free. For example, the app Helix Jump receives the highest positive review having a rating of 4.2 and of type free and it is installed on more than 100000000 devices.
- Some factor varies from person to person such as those people who want quality in apps they opt for paid apps. In my opinion, the size of the app should be less and if the app is paid then the price of the app is economical. It would be better if the app is free. The app is available for all age groups. Nowadays people mostly used games and social apps.

THANK YOU !