

Capstone Project

Play Store App Reviews Analysis

by

Ifraz Qazi

Pranav Shirke

Akram Khan

Points For Discussion

- Data Summary
- Percentage reviews sentiment distribution
- Free vs Paid
- Updated Free Apps
- Updated Paid Apps
- Number of Free or Paid Application Add or Updated based on year
- Category vs Installs
- Does Sentiment Subjectivity Proportional to Sentiment Polarity
- Correlation Map
- Price vs Ratings
- Reviews vs Ratings
- Number of Installs for each Category in a year
- Types of Reviews for each Category
- Number of Applications according to Content Rating
- Installs vs Size
- Conclusion

Data Summary

Problem statement: The Google Play Store is the largest app market in the world. It generates more than double the downloads of the Apple App Store but makes only half the money as the Apple Store. Explore and analyze the data to discover the key factor responsible for app engagement and success.

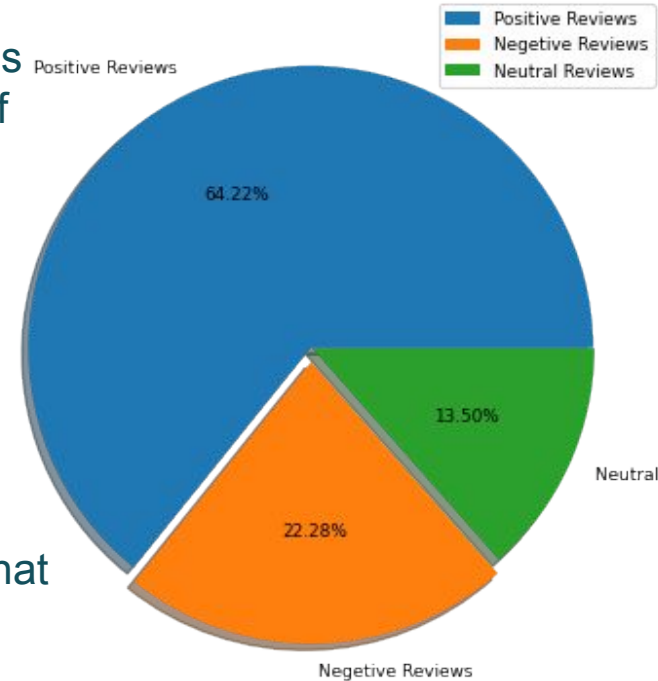
Play_store_df : This data frame is having a shape (10841,13). It holds the 13 features which include apps, category, rating, reviews, size, installs, type, price, content rating, genres, last update, current version, android version. This 13 features and 10841-row labels contains all sort of information which can be analyzed, interpreted, and implemented in order to take better business decision.

User_review_df : This data frame is having a shape (64295, 5). It consist of 5 features which involves app, translated review, sentiment, sentiment subjectivity, sentiment polarity.

Merged_df : As the name suggest, this data frame is the result of merging above two data frame based on feature app. As app is the only feature that is common between these two data frame. merged_df encapsulate all the feature from play_store_df and user_review_df.

Percentage reviews sentiment distribution

A Pie Chart Representing Percentage of Review Sentiments



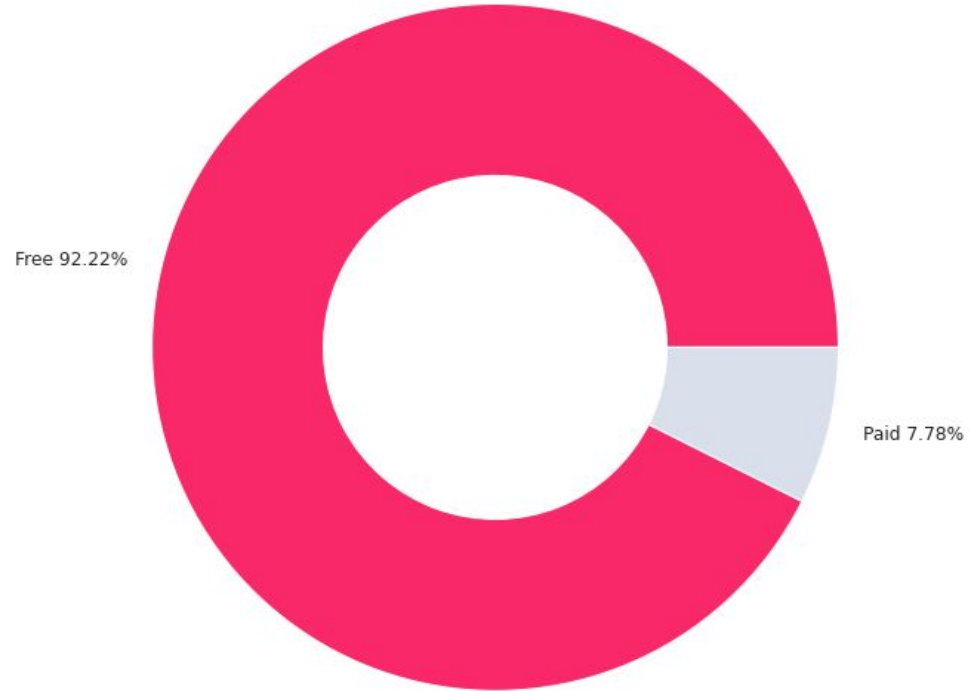
From the above pie chart, it can easily be understood that there is around 65% of user reviews sentiment is positive, around 22% of reviews sentiment is negative and the remaining around 14% of reviews sentiment is neutral.

If some apps have a higher percentage of positive reviews sentiments, then it is sure that the app is performing its intended work, and people are enjoying it, they may share the app with somebody thus increases the number of installations.

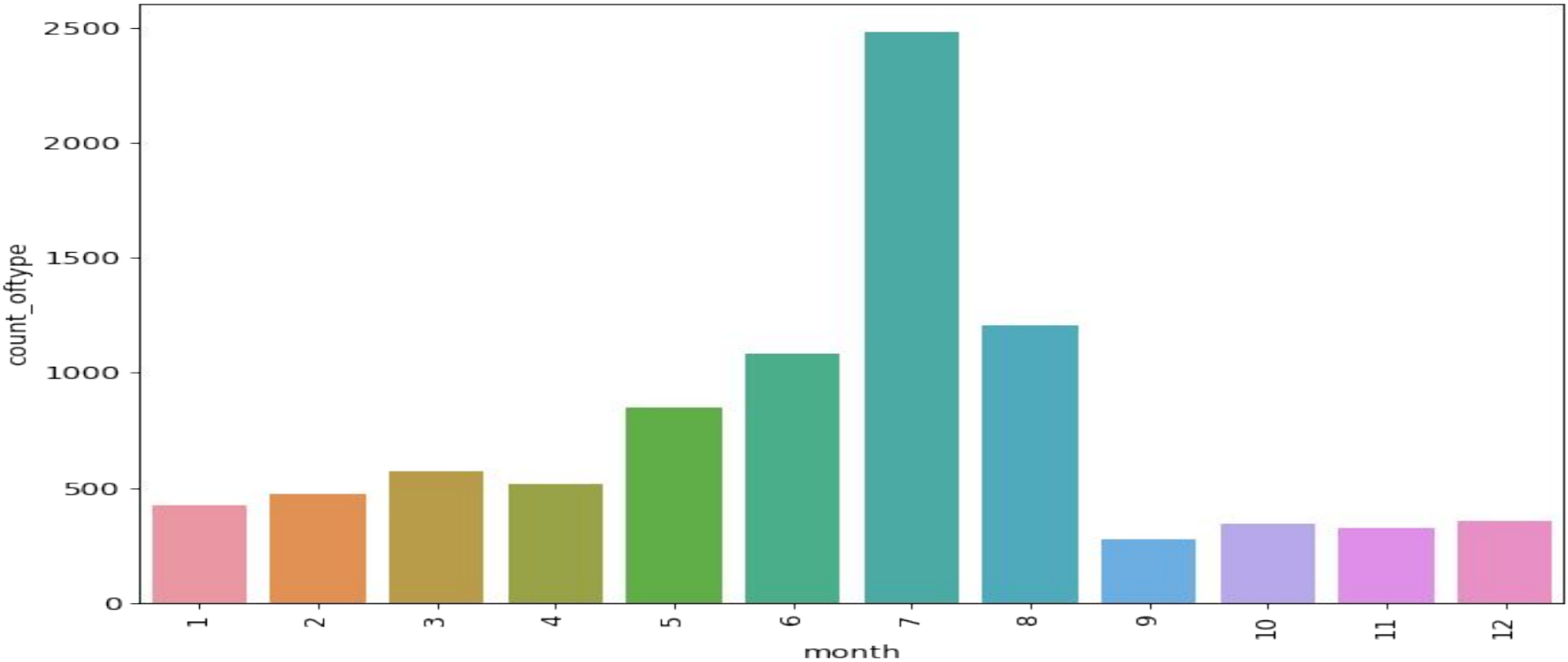
So, need to keep an eye-tracking on the review sentiment it is what decides whether the app is going to feature on google play store. By featuring I mean visibility of apps when someone searches for a category. If the app is not visible in the top 10 or 12 apps range then there are fewer chances of the app being installed.

Free vs Paid

Here we can see that 92.22% apps are free and 7.78% apps are paid on Google Play Store, so we can say that Most of the apps are free on Google Play Store.

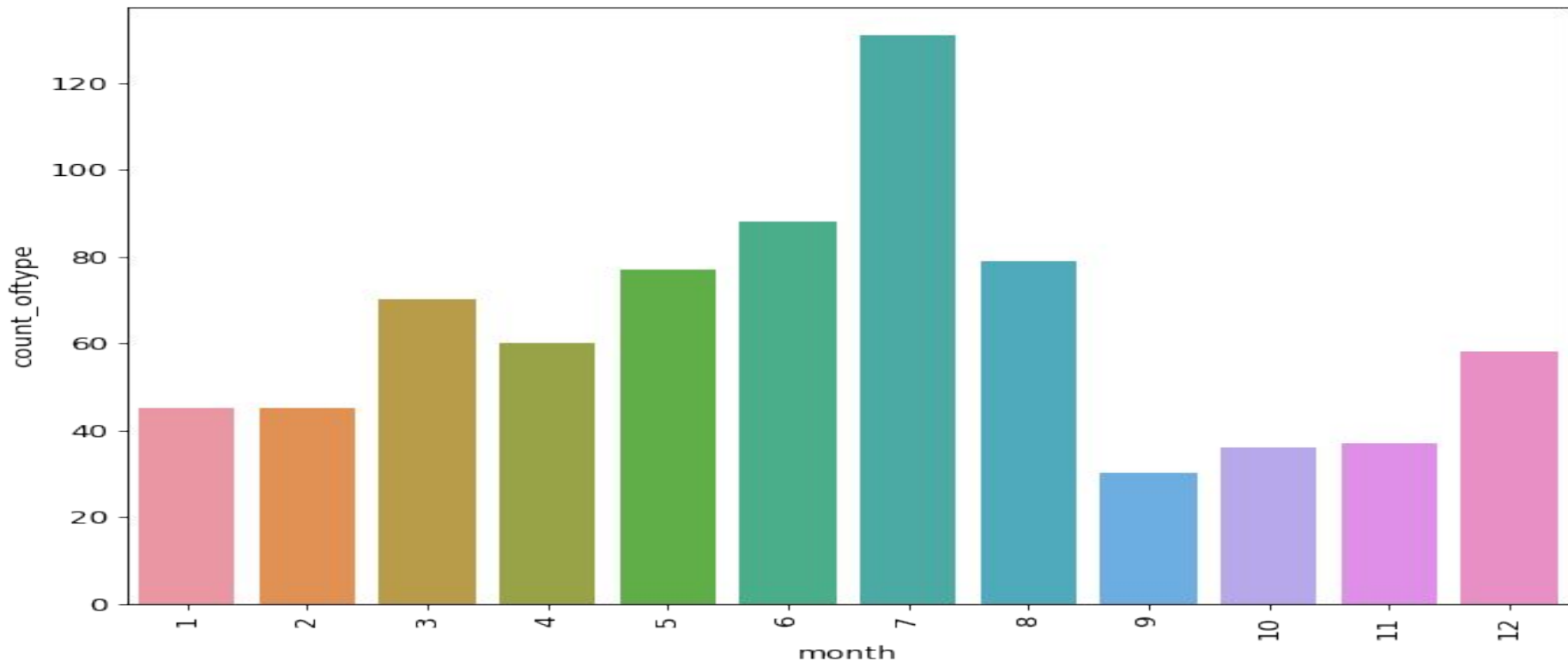


Updated Free Apps



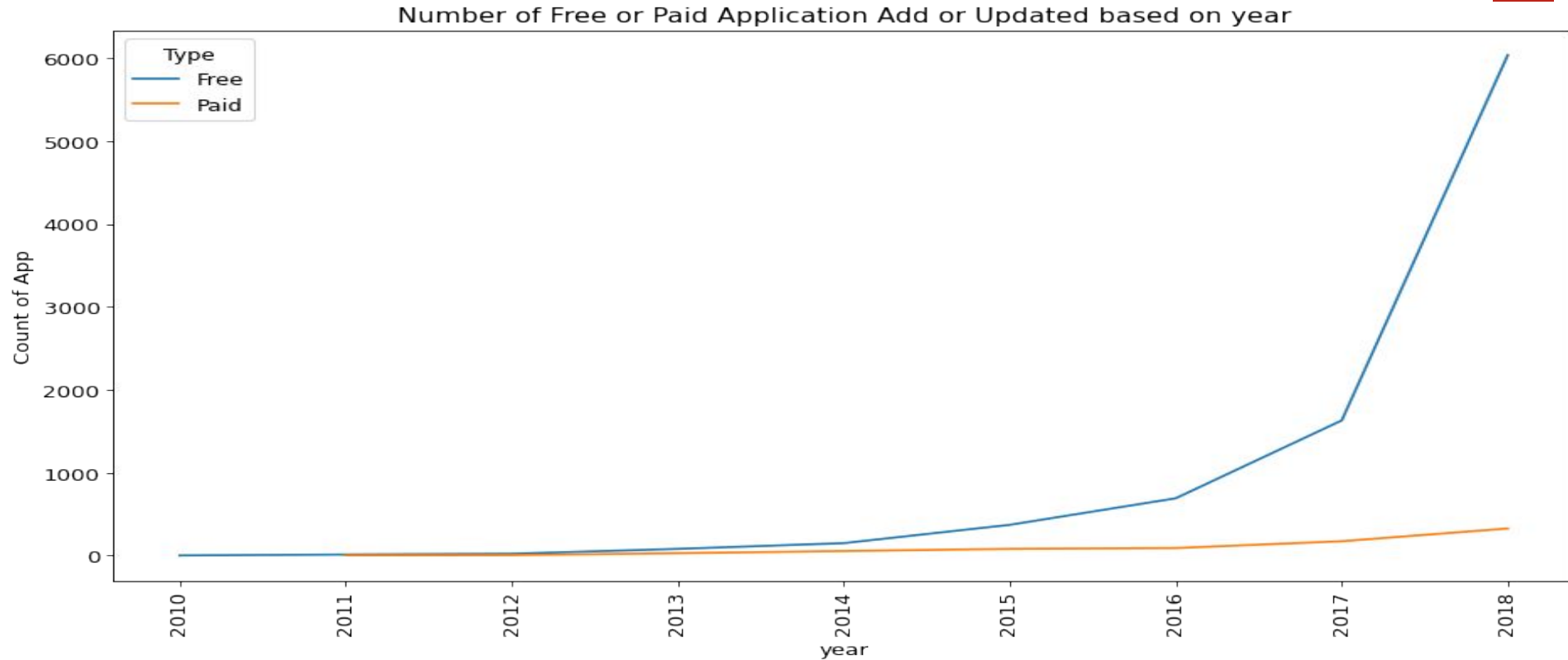
In this data almost 50% apps are added or updated on the month of July, 25% of apps are updated or added on the month of August.

Updated Paid Apps



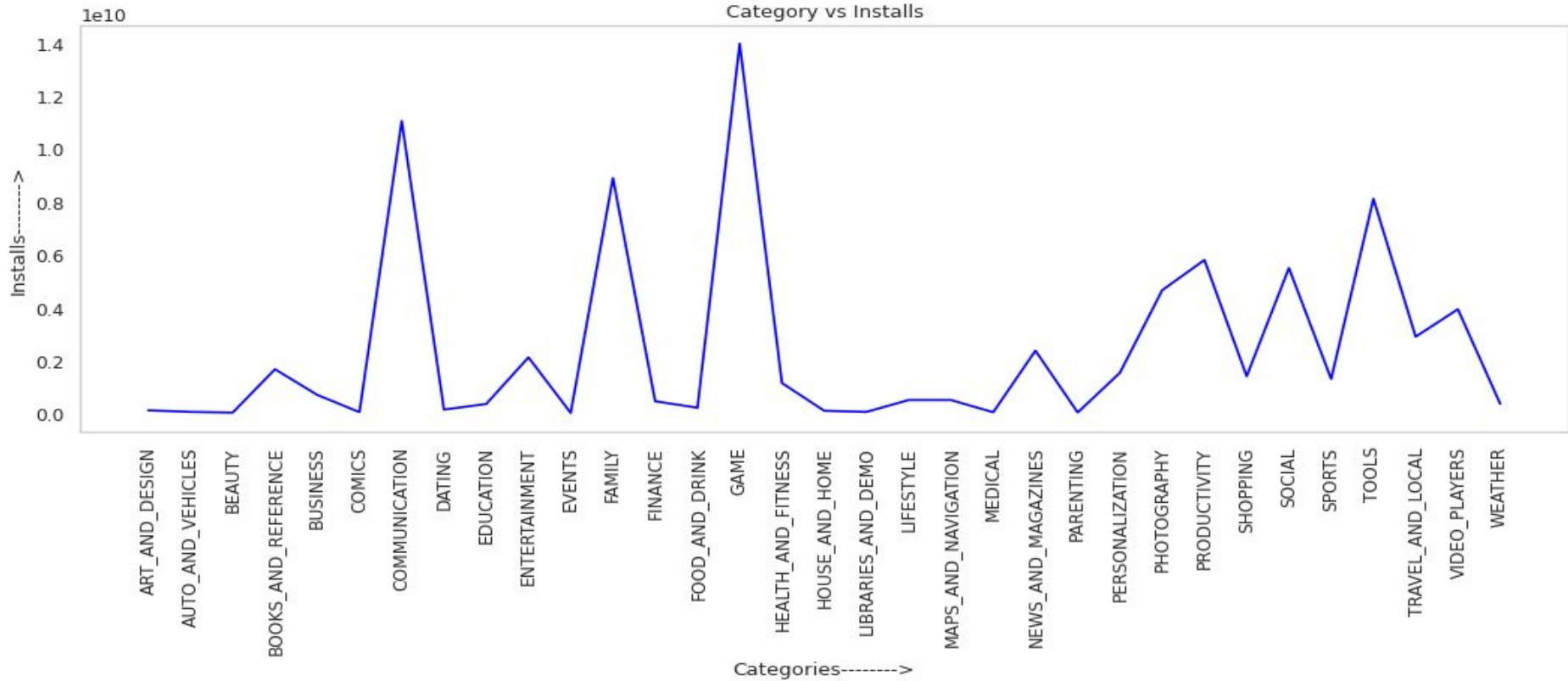
Same as free apps most of the paid apps too updates in the month of July.

Number of Free or Paid Application Add or Updated based on year



By observing above plot we can conclude that before 2011 there were no paid apps, but with the years passing free apps has been added more in comparison to paid apps.

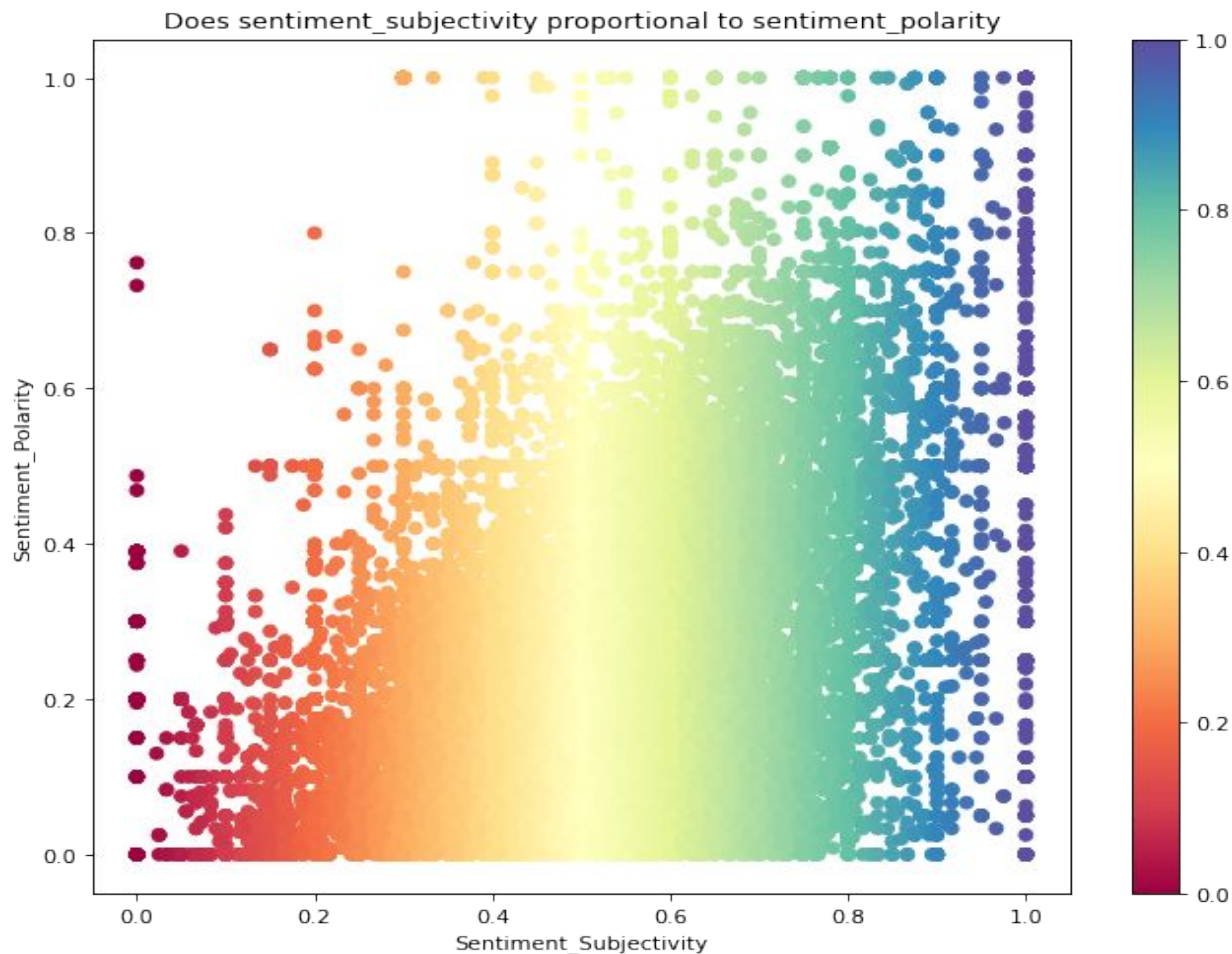
Category vs Installs



In This plot we can see that Game, Communication, Family, Tools Category is having highest no of installs.

Does sentiment_subjectivity proportional to sentiment_polarity

From the above scatter plot it can be concluded that sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behaviour, when variance is too high or low.



Correlation map

We can determine from above that **Installs** and **Reviews** are Highly Positively Correlated.

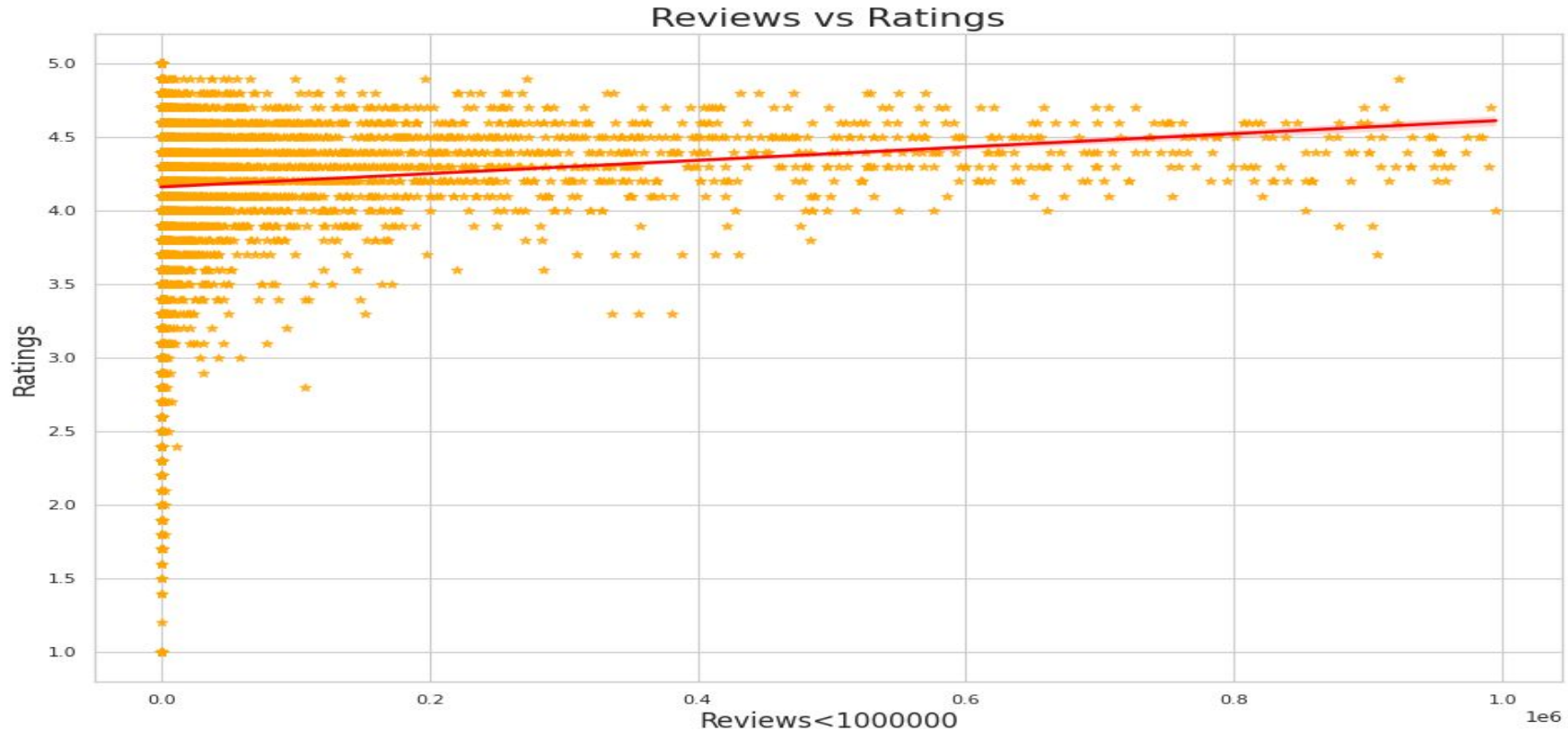


Price vs Ratings



Generally increasing the Prices, doesn't have significant effect on Higher Rating. For Higher Price, Rating is High and almost constant i.e. greater than 4.

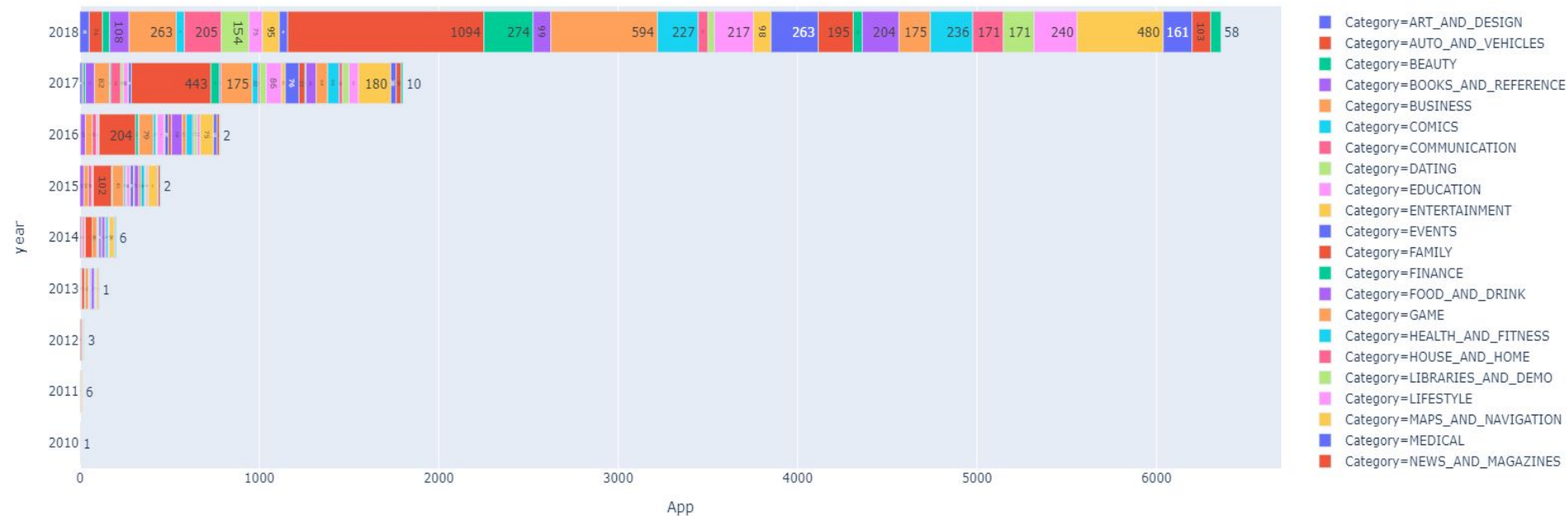
Reviews vs Ratings



Obviously by looking at above scatter plot with trendline we are able to conclude that lesser the reviews on applications lesser the rating as well.

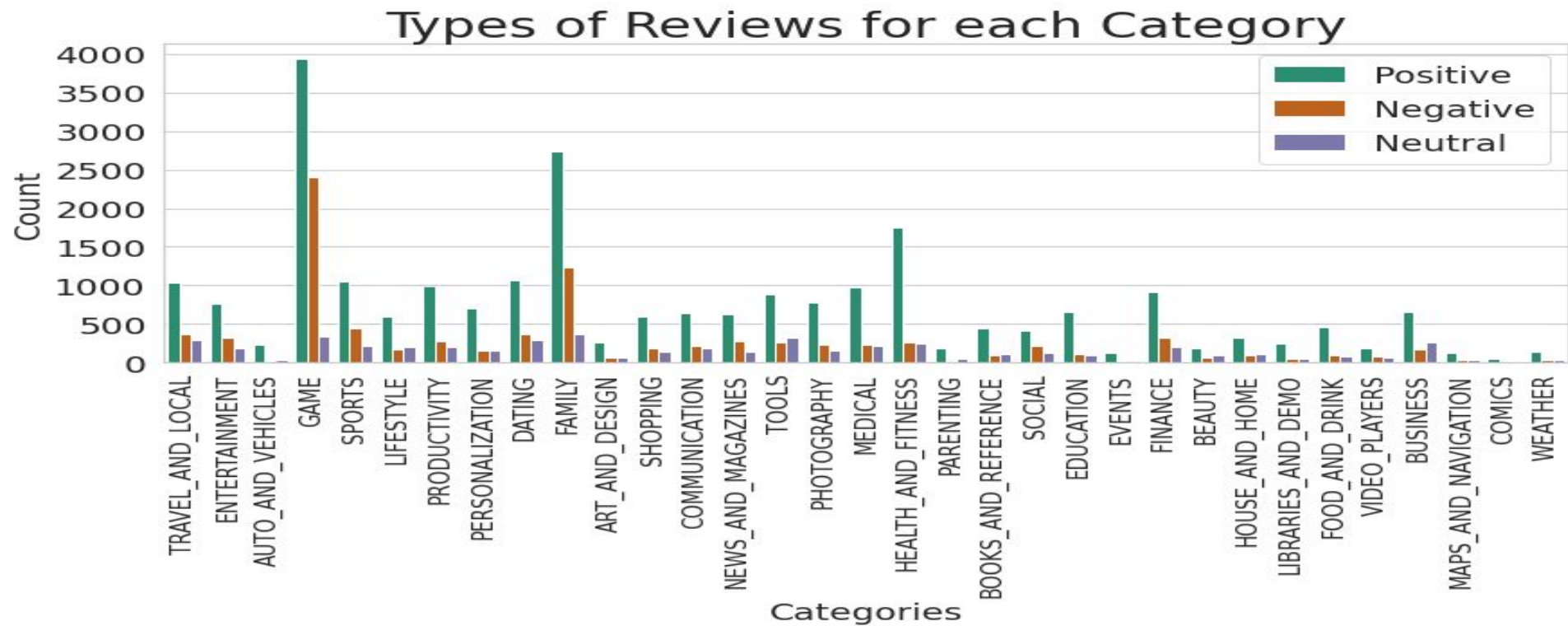
Category wise Count of apps per year

Category wise Count of apps per year



Family Category have high number of Application and if we move from year then we can also see the change in number of apps.

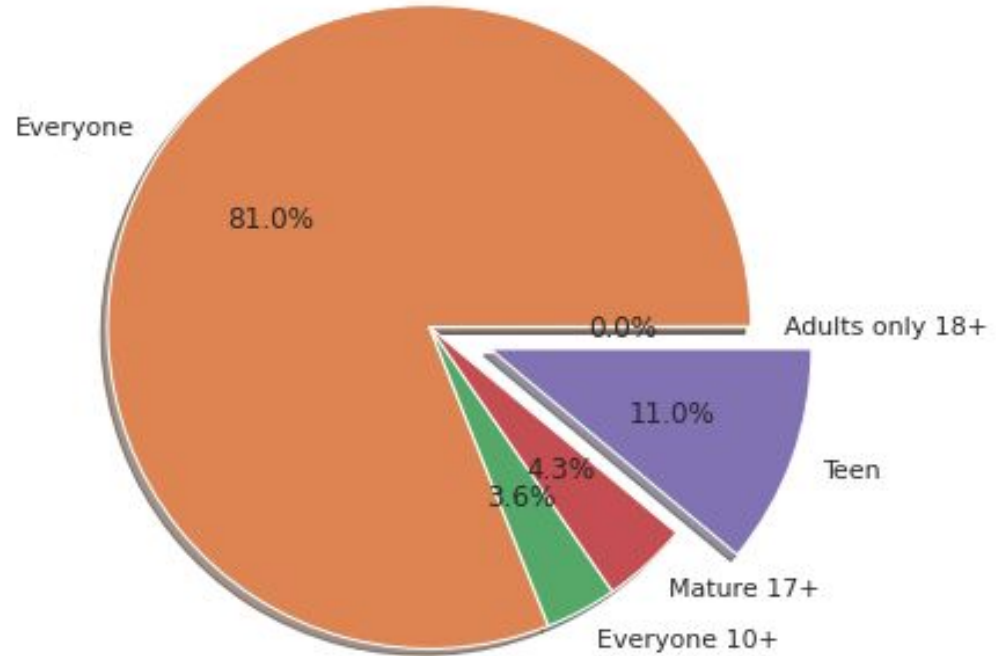
Types of Reviews for each Category



Game Category appears to have the most Positive and Negative Reviews. Shows volatility of Game Category. **Health and Fitness** appears to have least Negative reviews in relation to its positive reviews.

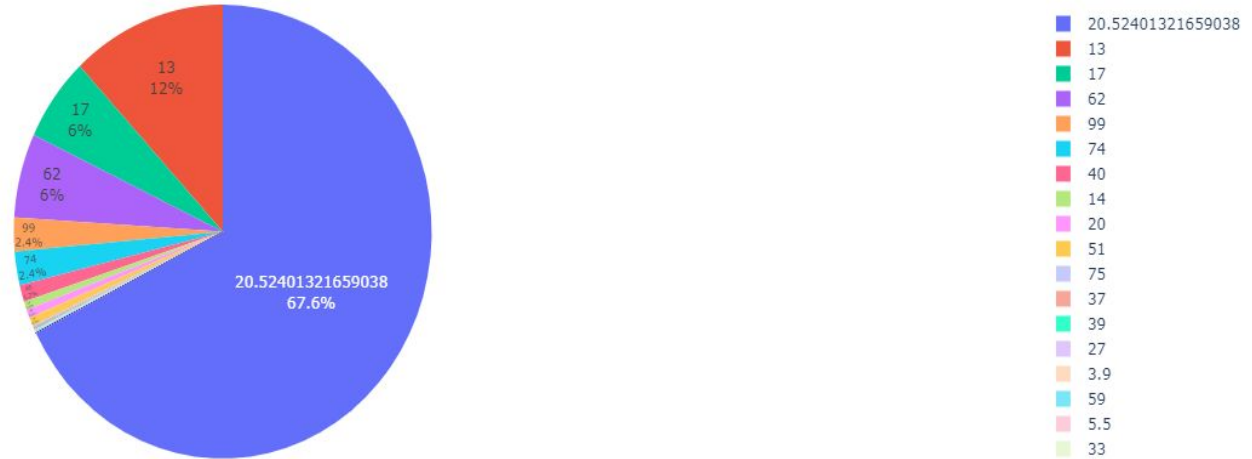
Number of Applications according to Content Rating

It can be seen from above plot that 81% Of Apps are Targeted For Content Rating 'Everyone'.



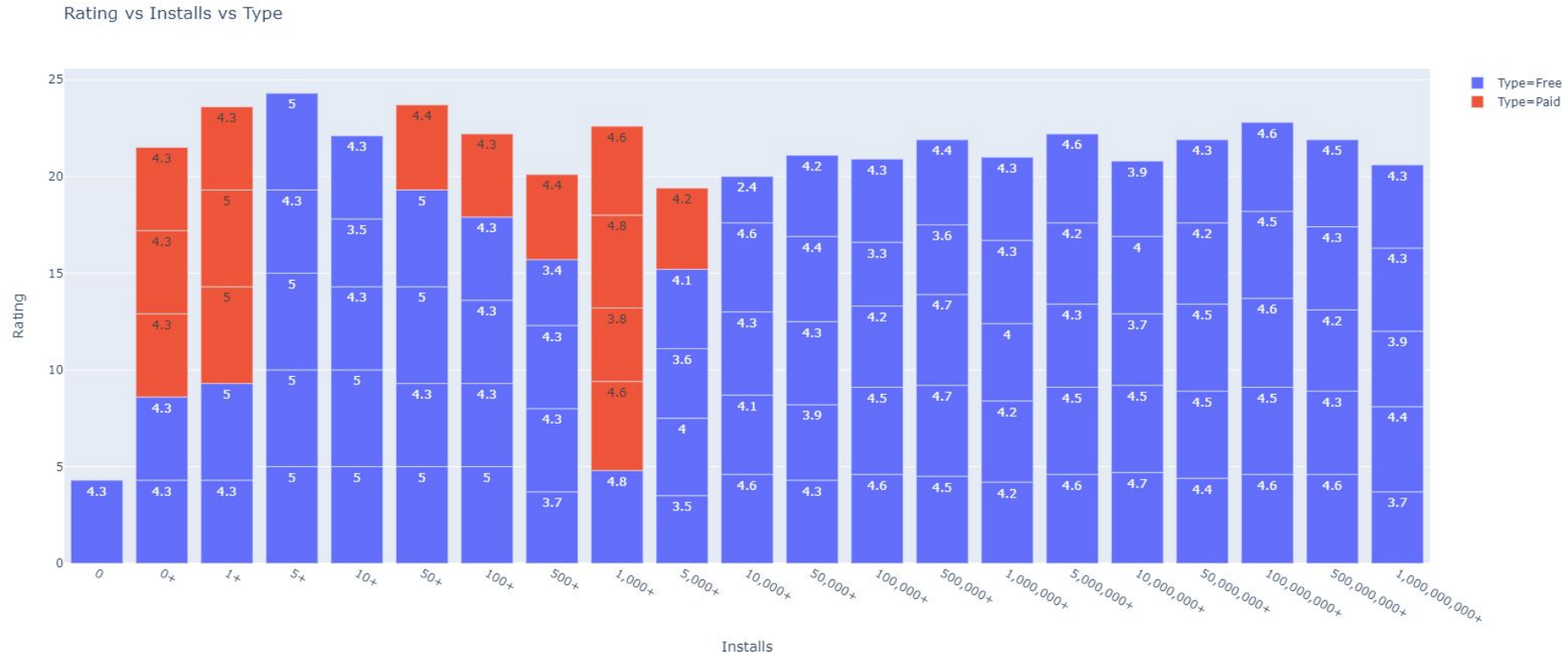
Installs vs Size

Installs vs Size



App Size will be one of the most focused area when developing the app. From Pie chart there is something developer needs to keep in mind that our app size should be in the range of 20 to 30 MB. Users preferred to download the apps which has minimum size.

Rating vs Installs vs Type



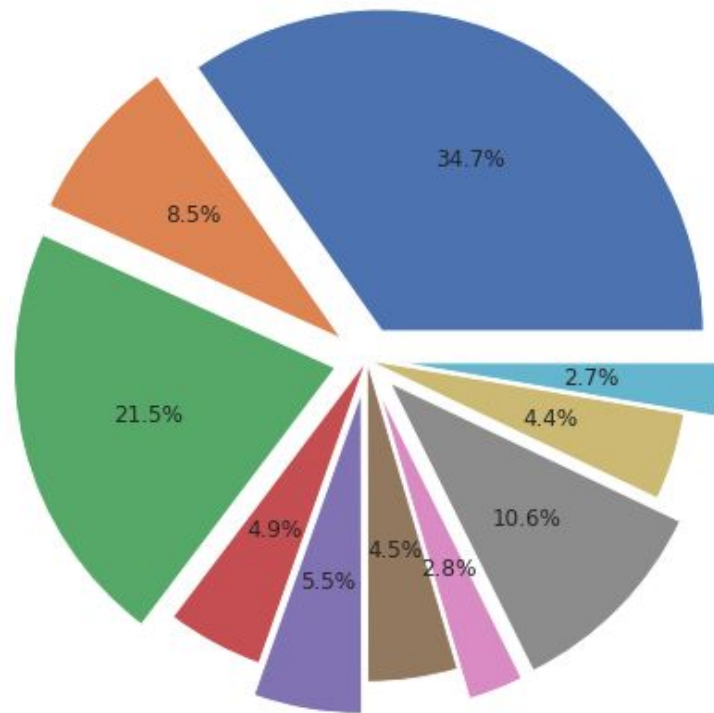
From above graph, we clearly make up a point that Free apps having large number of installs with average rating greater than 4. Also here it shown us that paid apps having less installs upto 5000+ therefore it implements that to gain higher users- you will have to make sure your app works flawlessly and offers a full set of features. Another reason for people ignoring paid apps even with lots of great reviews, people are less likely to purchase an app if they can't try it out first.

Top 10 Installed Genres for Game Category

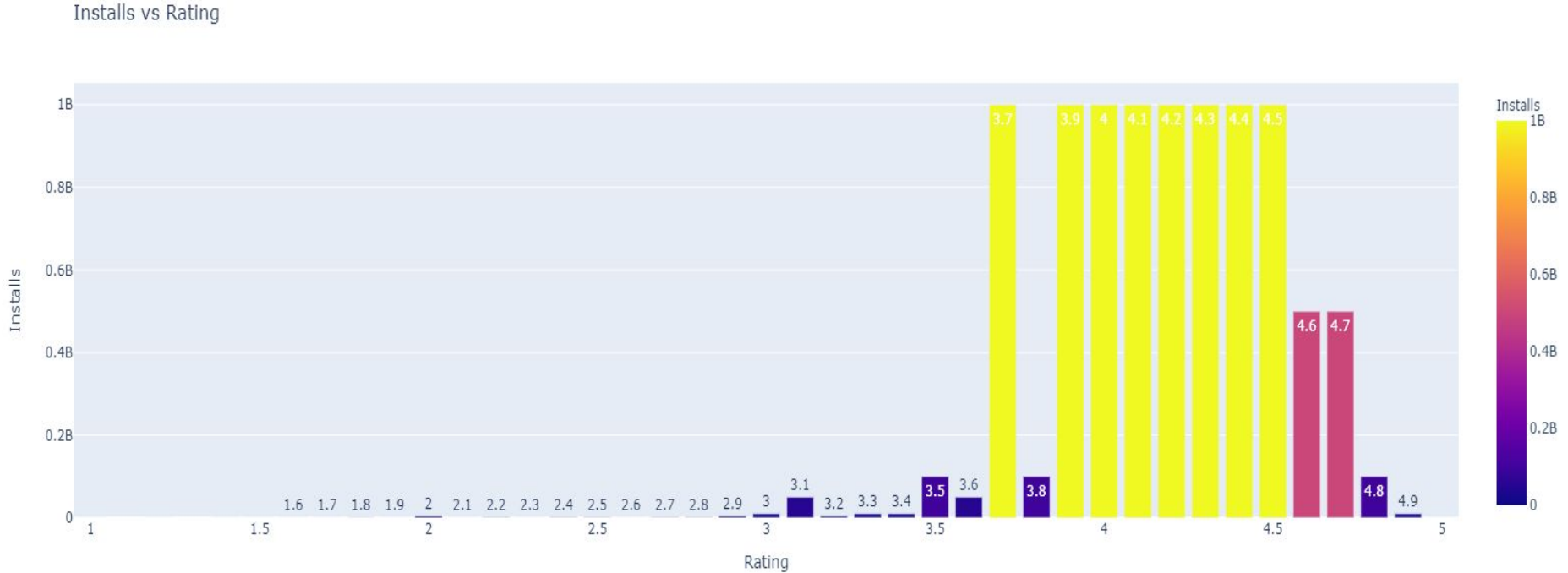
Here we can conclude from graph that **Action** Genres for Game have more number of downloads followed by **Arcade**.



Top 10 Installed Genres for Game Category

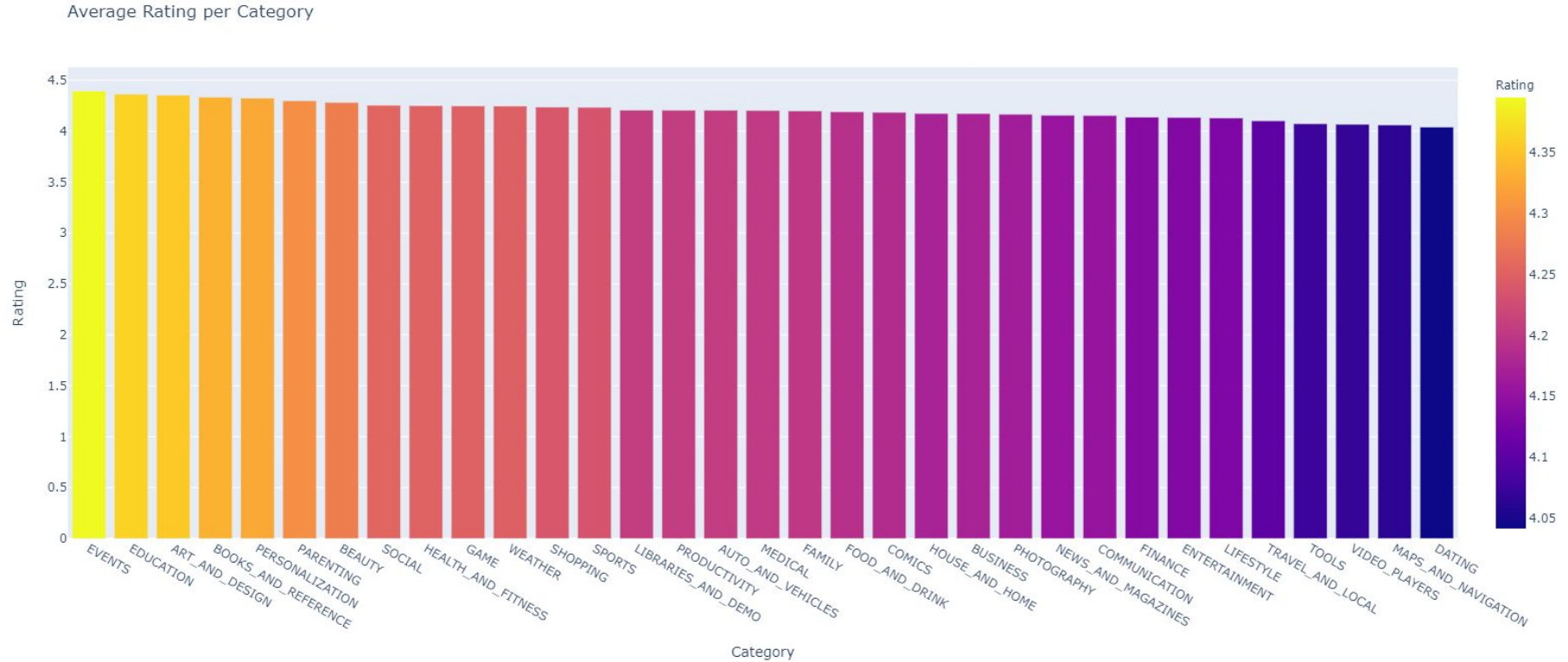


Installs vs Rating



Apps with approximate Rating range between 4 to 4.5 have more likely 1B+ installs.

Average Rating per Category

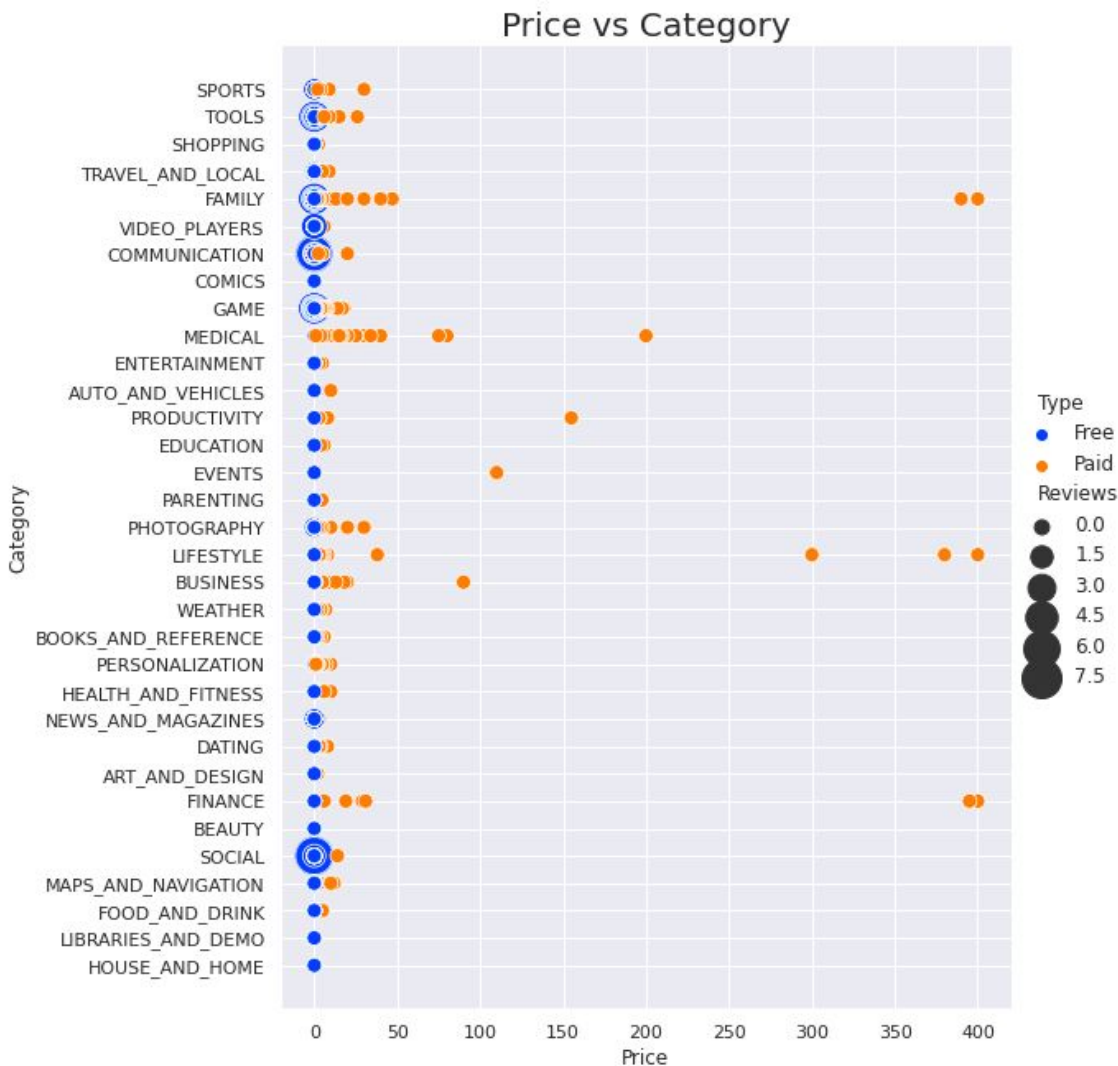


We can see from the Category vs Average Rating graph that **Events** Category have average number of ratings with 4.39 as more and **Dating** Category with 4.04 as less.

Price vs Category

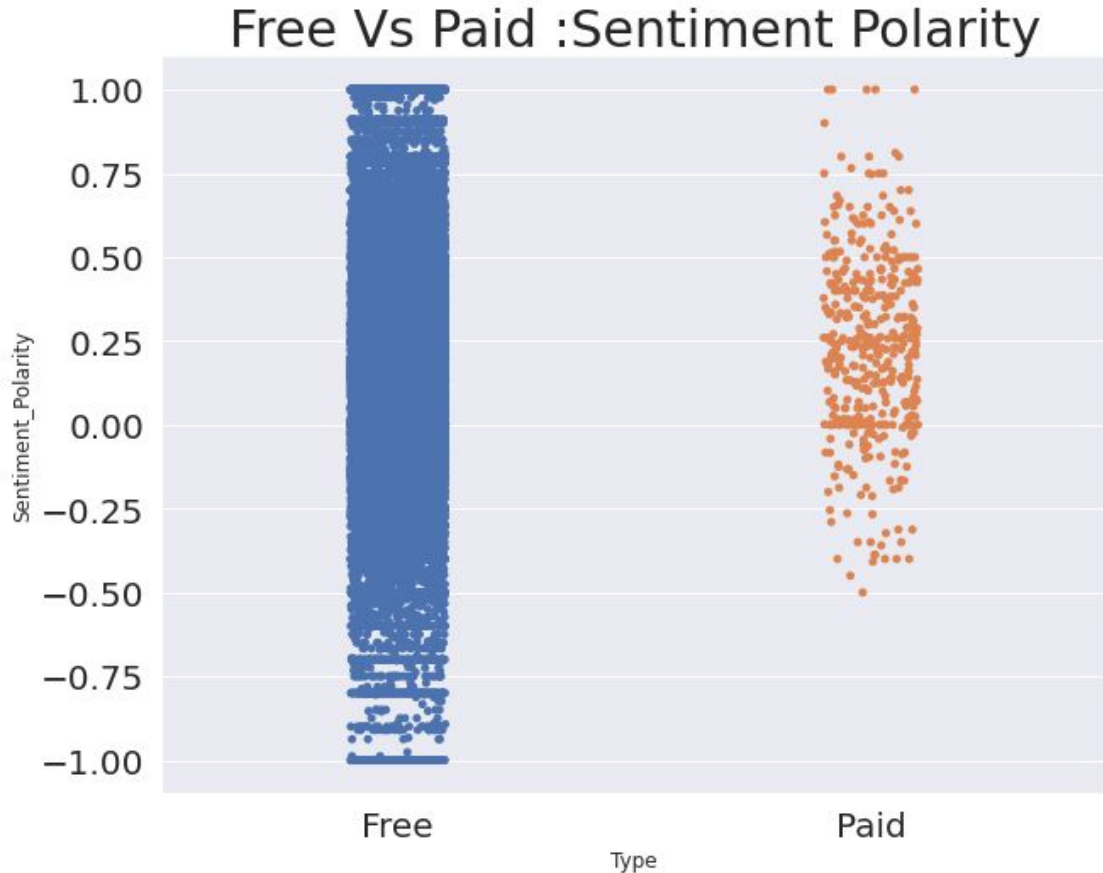


Price Vs Ratings plot delivers that Type of Apps is either PAID or FREE, Circle is larger if Reviews are more. **Social, Family, Communication, Family, Tools** Categories have highest number of reviews. **Finance, Lifestyle, Family** have high price in paid apps. There we can observe from plot points that **Medical** and **Family** Category have more number of paid apps.

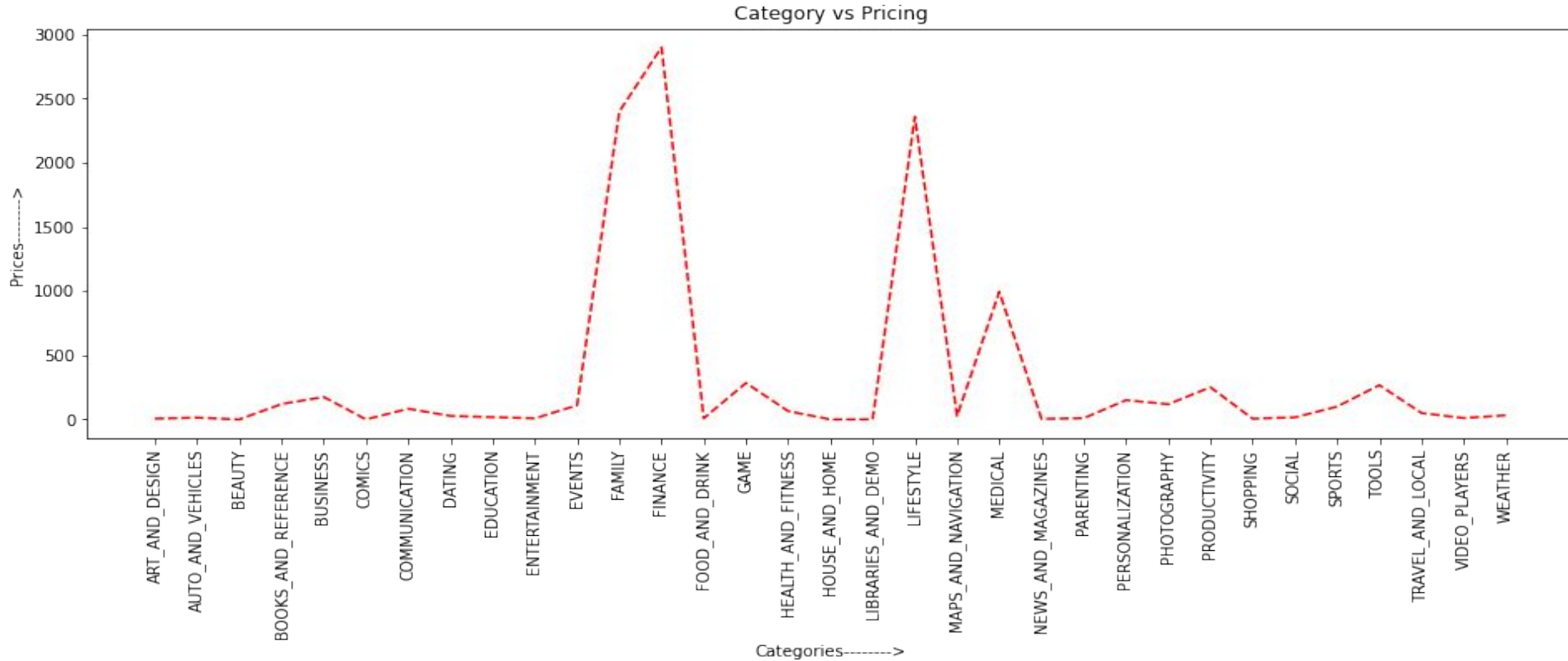


Free vs Paid: Sentiment Polarity

Type vs
Sentiment_Polarity
graph clearly shows
that **Paid Apps** are less
likely to receive bad
reviews and **Free apps**
have a more equal
distribution of
Sentiment Polarity.

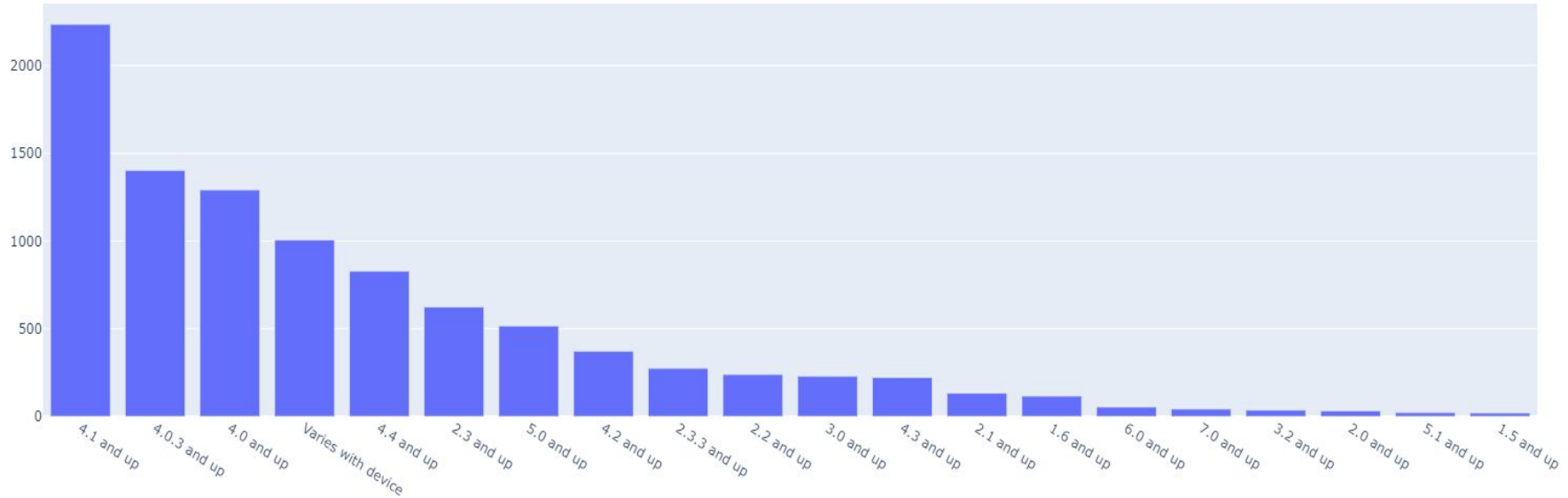


Category vs Pricing



In This plot we can see that prices are different for different Categories, Finance, Lifestyle, Medical is having highest prices.

Android Version Requirements of Apps



By observing above plot we can conclude that most of the Applications requirement of Android version is 4.1 and up.

Conclusion

As per our EDA, an ideal application on the google play store should obey the following properties/characteristics

1. **Category Type:** Before 2011 there were no paid apps, but with the years passing free apps has been added more in comparison to paid apps, By comparing the apps updated or added in the year 2011 and 2018 free apps are increased from 80% to 96% and paid apps are gone from 20% to 4%. So, we can conclude that most people are after free apps.
2. **Installs vs Rating:** Some Semi-popular apps are crossing a Million and then 10 Million, and only a few apps are able to go beyond the 500M and 1B mark. Popular apps under the Social media category like Facebook, WhatsApp, Instagram have more than 1B installations.
3. **Rating vs Installs vs Type:** Free apps are preferred over paid apps and the rating are in the range of 4 to 5. Therefore the free Apps are most downloaded as compared to paid apps, Hence Developer should focus more on free apps. User prefer more of free apps. Most of the apps present in play store are more or less of same size so size doesn't affect their decision much.
4. **The most installed category:** As we have explored applications belong to the category gaming and followed by communication are being installed the most, this is the clue to choose the category.
5. **User Reviews:** User Reviews often contain valuable feedback and suggestions for improving the app. Analyzing the top trends and issues for the app enables us to assess which parts of the app or game to focus on. It also shows where to invest in the business to improve the offering and create a more successful app or game.
6. **Ratings:** App rating is a reflection of how users respond to the app. Learn what affects rating and what we can do to influence it.

Thank You!