

Capstone Project 1

Play Store App Review Analysis

Presented By
SOURAV KARMAKAR



Contents

1. Introduction
2. Objective
3. problem Statement
4. Description of data
5. Cleaning the data
6. Data analysis and visualizations
7. Conclusion

Introduction

Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .

There are more than 5 million apps found on Google Play Store.

Android is the most popular operating system in the world, with over 3 billion active users spanning over 190 countries.

Lots of designers and developers work on it to make an app successful on the Play Store.

Objectives

- ➡ Understand Consumer behaviour and demand, how they reacts to different Category and Genres of Google playstore Apps
- ➡ Find the most popular and trending apps in recent times
- ➡ Find how small changes or update impacts on app performances
- ➡ Analyze the Reviews,Rating,Sentiments of people towards various apps in playstore
- ➡ Above all, help developers or Clients to recognize the gap, make the app better and meet customer expectations.

Problem statement

1. Top 5 Categories which have overall best ratings
2. Top 5 Genres which have best average ratings
3. Last Update Impacts on trend of Rating
4. Find Relationship between Rating and Average Reviews
5. Let's analyze the sentiment subjectivity and overall people's sentiment
6. Top 5 Categories Review Sentiments
7. Let's do sentiment analysis of each app genres
8. Let's Check if Category is Free or Paid
9. Most Reviewed Apps
10. Apps which have rating more than 4 and have reviews greater than 10000
11. Variation of review with month in the year 2018
12. Let's look the sentiment column according to Rating
13. Factors affecting installs
14. Top Content Rating Per Installations
15. co-relation between installs, ratings, and size

Description of Dataset

There are two dataset: Play Store Data & User Data

1. Play Store Data:-

App - Name of the Application

Category - Category of the Application

Rating - Rating given to the Application

Reviews - No of reviews given to the Application

Size - Size of the Application

Installs - No of downloads of the Application

Type - Free or Paid

Price - Price of the Application if it is paid

Description of Dataset

Content Rating-It is Age appropriate or Not

Genres - Type of Genre the Application belongs to

Last Updated - When the last time the Application is Updated

Current Ver - Current version of the Application

Android Version- Minimum Android version required to run the Application

Description of Dataset



2. User Review Data:

- **App** – An app name
- **Translated Review**:- Reviews being given by consumer
- **Sentiment** – Sentiment given to an app by users (i.e. Positive, Neutral, Negative)
- **Sentiment Polarity** – The polarity of sentiment measures how negative or positive the context is. In the data we have, the polarity ranges from +1(Positive) to -1(Negative).
- **Sentiment Subjectivity** - The subjectivity of a sentiment is how likely that sentiment is to be based on data or factual information, versus personal opinions or public notions

Data Cleaning

- Data cleaning not just means removing the incorrect data or erroneous data. Many times we get the data which has all kinds of values some of them will cause problems during the analysis of the data and make our predictions incorrect. So we have to make sure our data has no erroneous values.

- Data Cleaning Step:

Removing unwanted Values : Deleting of duplicate/incorrect or irrelevant values

Handling Missing Values: Handling missing values in our Dataset

Handling Structural Errors: Fixing mislabeled categories or classes, Types, Strange name conventions

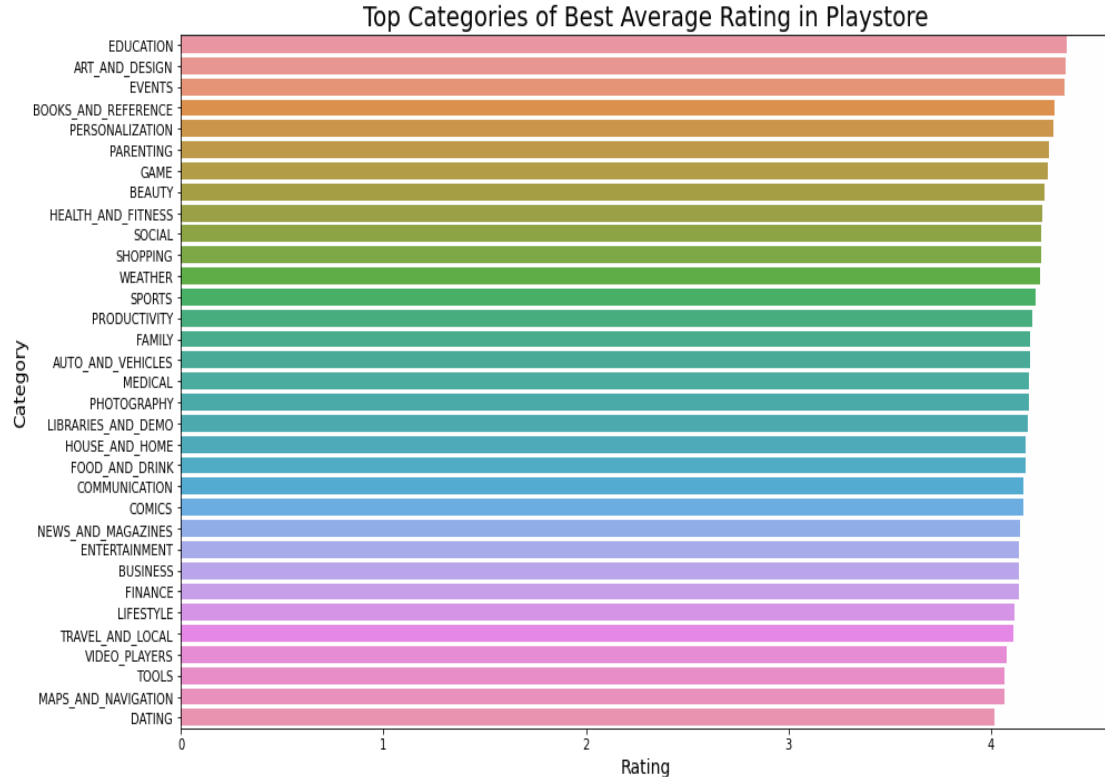
Filtering Unwanted Outliers: Removing incorrect or unwanted outliers

Replacing missing values with mean, median or mode: Replacing missing values with mean is the most popular method of replacing missing values

Data Analysis & Visualization

Top Categories with Best Average Rating in Play store

- From above graph we can conclude that EDUCATION is the top category which has average rating 4.37.
- Also top 5 categories of Application in the play store which has average best ratings are EDUCATION, ART_AND_DESIGN, EVENTS, BOOKS_AND_REFERENCE, PERSONALIZATION.
- FAMILY which category has maximum apps in play store(1931). It has overall average rating 4.19 and ranked 15th according to overall average rating. But interestingly, 2nd most maximum apps category 'GAME' ranked 5th with 4.27 average rating.

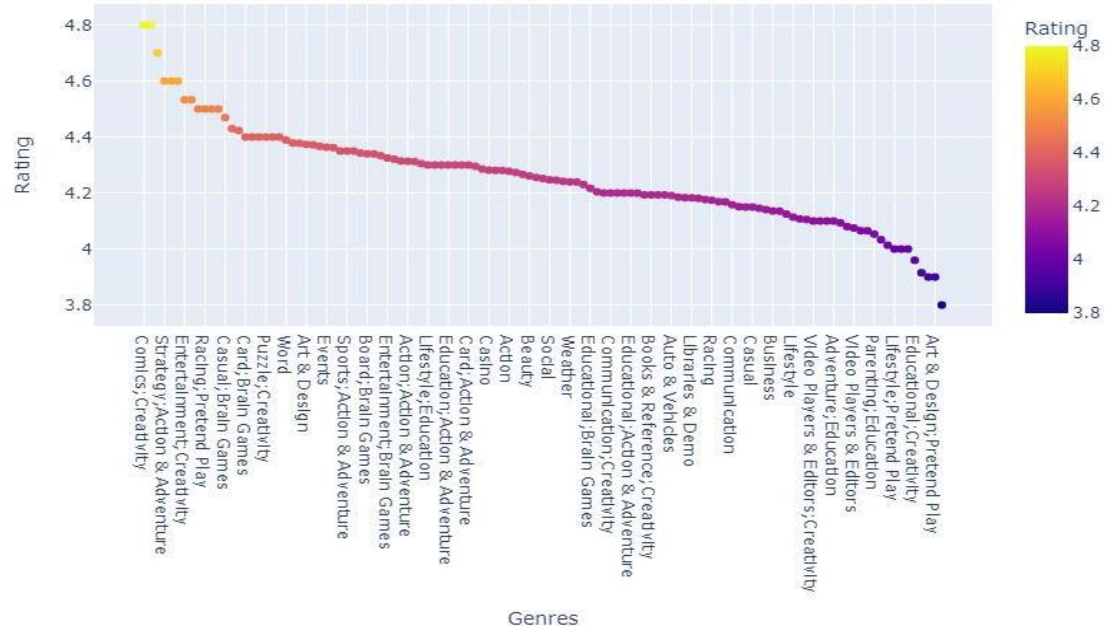


Data Analysis & Visualization(Cont..)

Top Genres in the Playstore based on Average Rating

- From the scatter plot from plotly we can conclude that both THE GENRE- 'BOARD;PRETEND PLAY' AND 'COMICS;CREATIVITY' is having the highest Average Rating of 4.8.
- There is no direct relation between quantity of apps under a genre and rating.
- Here Top 5 genres according to maximum average reviews are :- COMICS;CREATIVITY BOARD;PRETEND PLAY,HEALTH & FITNESS,EDUCATION,STRATEGY,AC TION & ADVENTURE, PUZZLE;EDUCATION.
- GENRE-'PARENTING;BRAIN AND GAMES' has got lowest Average Rating.

Average Rating of each Genres

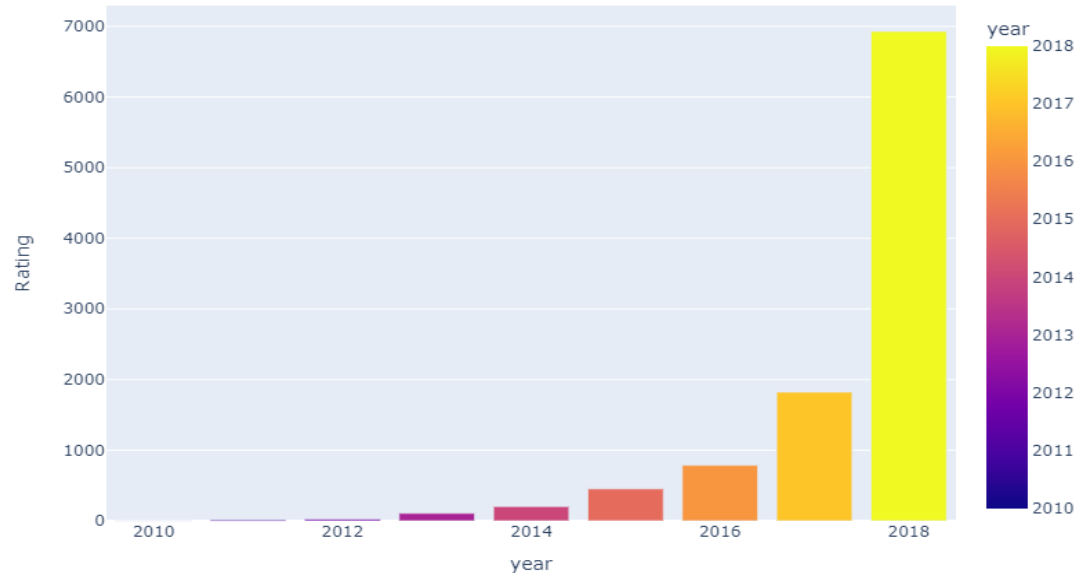


Data Analysis & Visualization(Cont..)

Last Update impacts on Trend of Rating

- From this graph from plotly we can conclude that
- there is more number ratings given the application which are updated recently in 2018 no. of rating is 6929 than those application which were updated in 2017 no. of rating 1823.
- This shows with the latest update user reviews response increases for both less or more ratings.

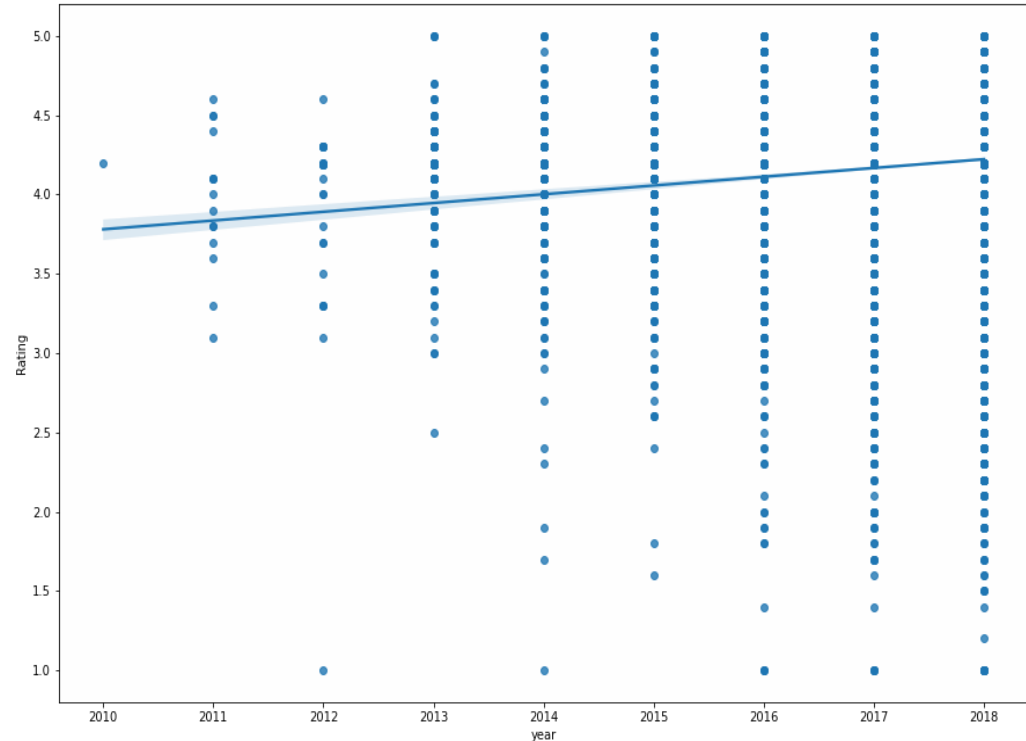
Ratings vs Year



Data Analysis & Visualization(Cont..)

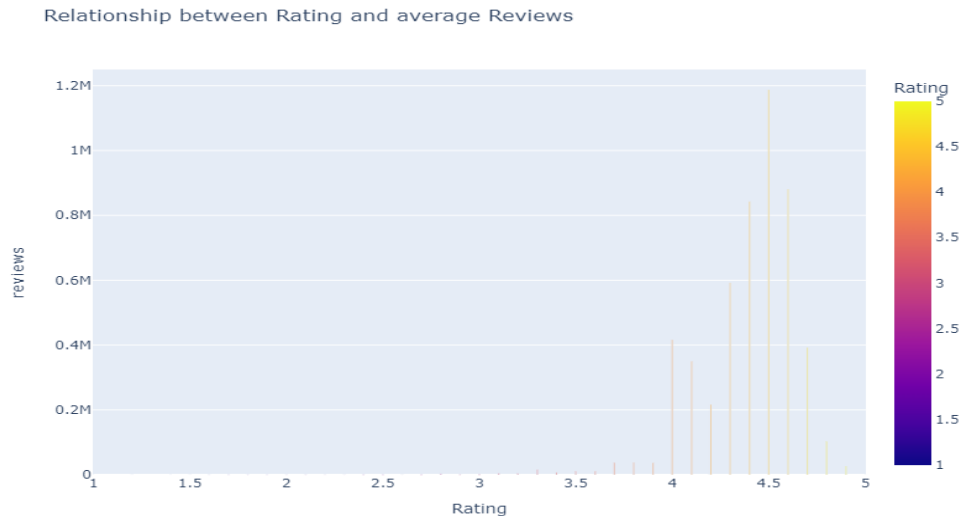
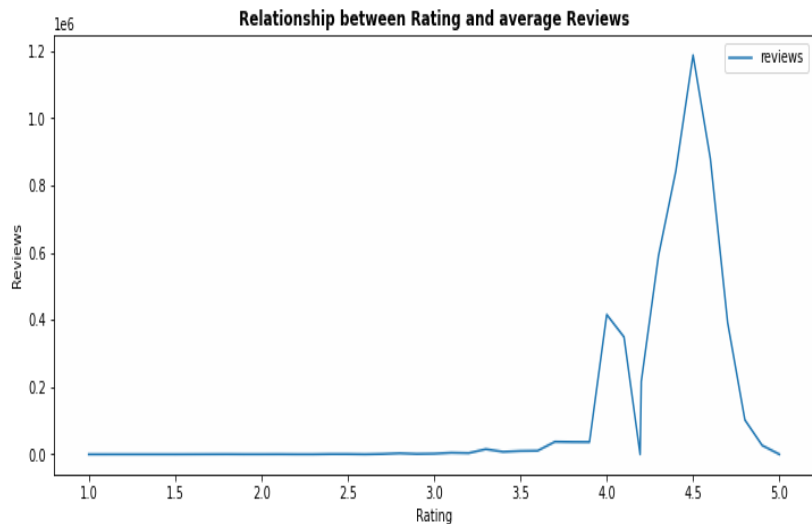
How last update impacts on trend ?

- From these above graphs we can conclude that, rating is increasing in a proportionate manner with the last updated time.
- So from this we can be sure that with the latest updation the reviewers are giving better rating.



Data Analysis & Visualization(Cont..)

Relationship between Rating and Reviews



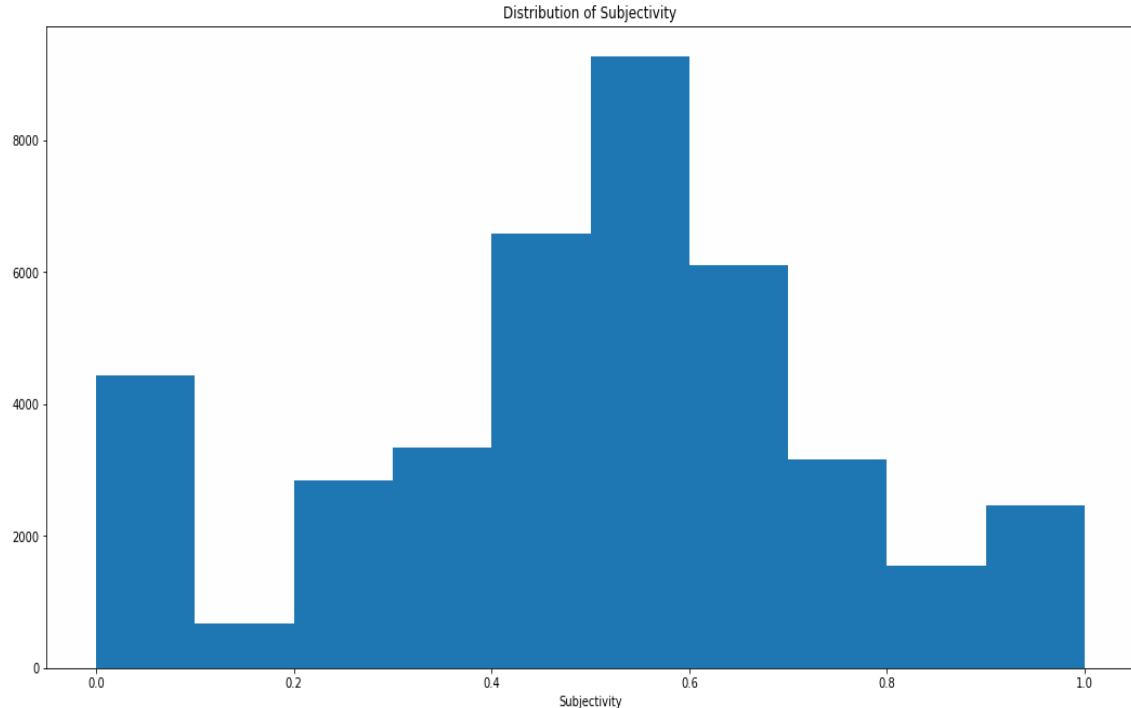
From the above graph we can conclude that

- As the rating increases the average reviews for each ratings also increases .
- Also we can observe some deviation after 4.5 rating as we can see 4.5 rating has maximum no.of reviews (1.1 Million) .
- But after 4.5 rating reviews eventually decreases and we can see only around 26586 reviews.

Data Analysis & Visualization(Cont..)

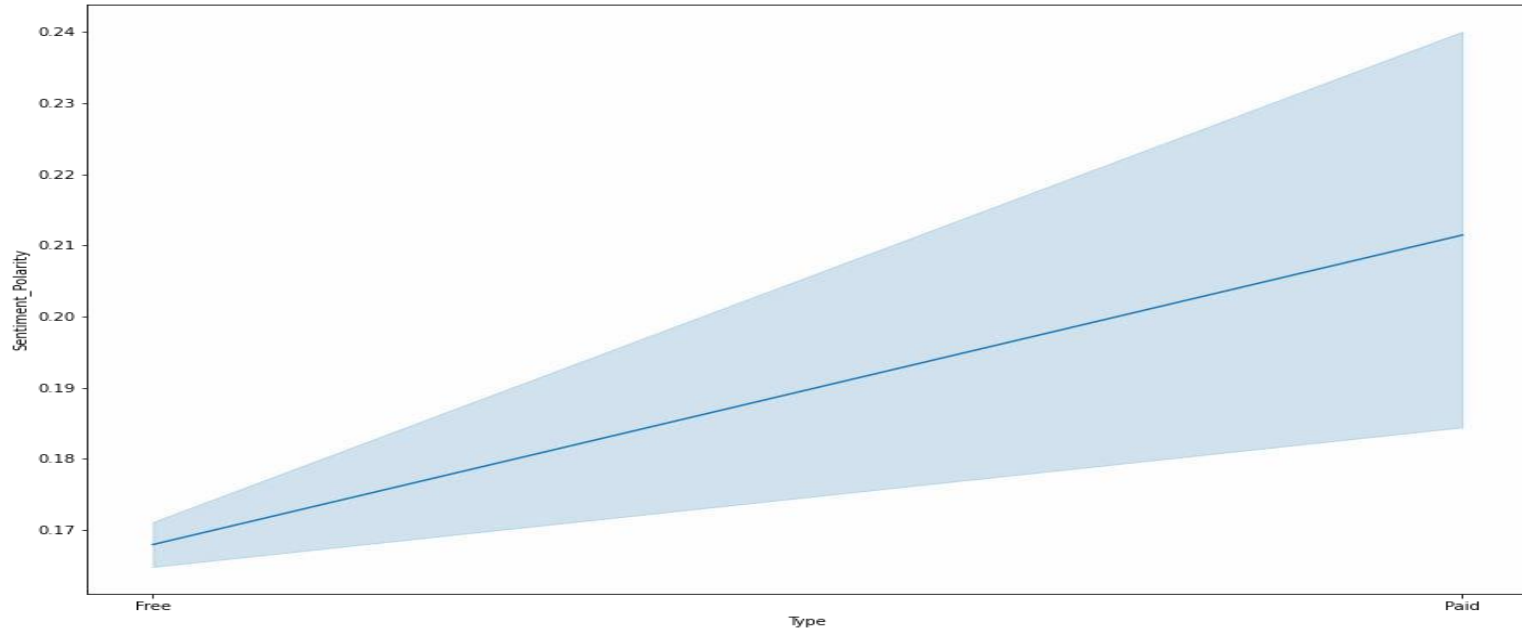
Distribution of Sentiment Subjectivity

- It can be seen that maximum number of sentiment subjectivity lies between 0.4 to 0.7.
- From this we can conclude that maximum number of users give reviews to the applications, according to their experience.



Data Analysis & Visualization(Cont..)

How Sentiment Polarity varies with Free and Paid Apps:



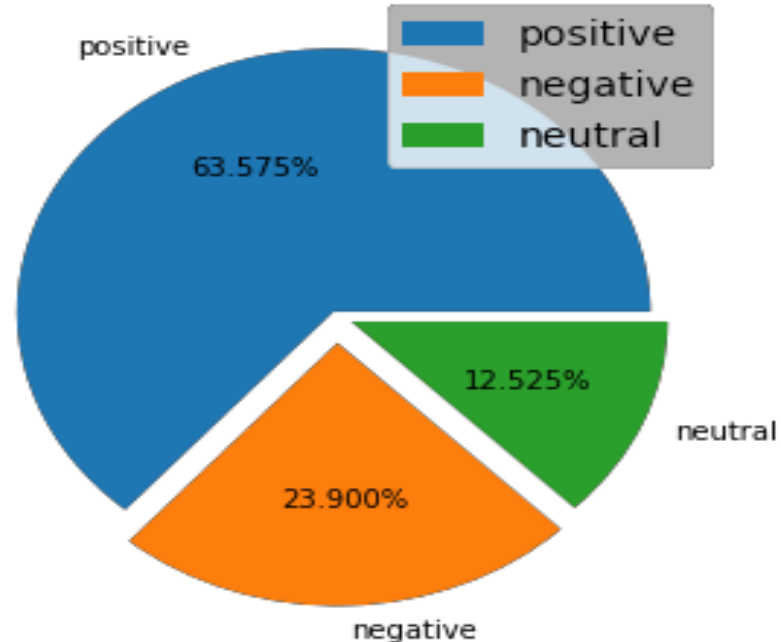
So people express more sentiment towards paid app rather than free app

Data Analysis & Visualization(Cont..)

Overall Percentage of Review Sentiment

Overall Percentage of Review sentiment

- Most of the sentiment reviews given by the user are positives with 63.575%.
- But also there is a negative sentiment percentage of 23.900% %.
- This means app developers needs to convert more negative sentiments to neutral or positive sentiments with their Hard work

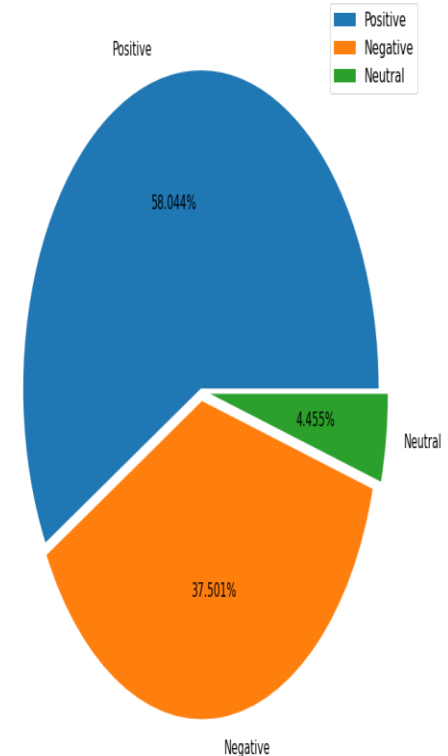


Data Analysis & Visualization(Cont..)

Top 5 Categories Review Sentiments

- 'GAME','FAMILY','HEALTH_AND_FITNESS','TRAVEL_AND_LOCAL','TOOLS' as most popular category which has most translated reviews
- 'HEALTH_AND_FITNESS' Category has got highest positive percentage of 77.119% and negative sentiment percentage of 11.494% and neutral percentage of 11.388%.Also we found that top Category GAME has less positive sentiment percentage of 58.04% than it's competitor.
- Most negative sentiments from the top translated app category has been received by GAME CATEGORY THIS shows that even if GAME app has the highest translated reviews but in positive sentiment percentage it is lower than it's counterparts.

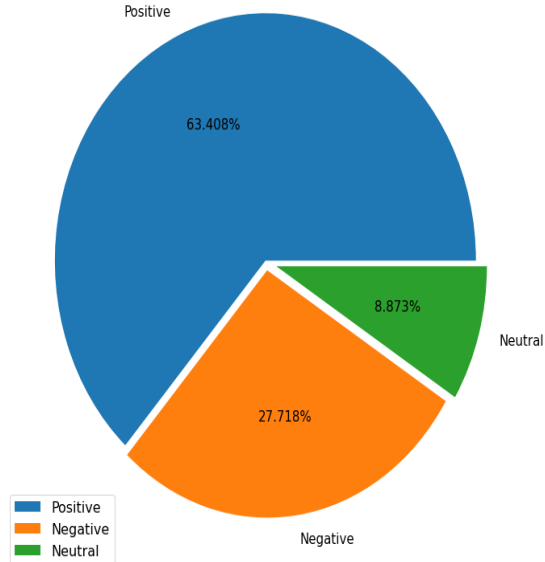
1.Pie chart representing sentiment percentage of GAME category



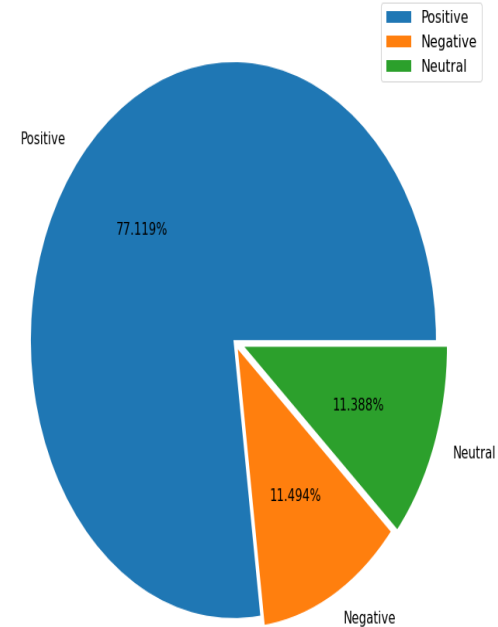
Data Analysis & Visualization(Cont..)

Top 5 Categories Review Sentiments

2.Pie chart representing sentiment percentage of FAMILY category



3.Pie chart representing sentiment percentage of HEALTH_AND_FITNESS category

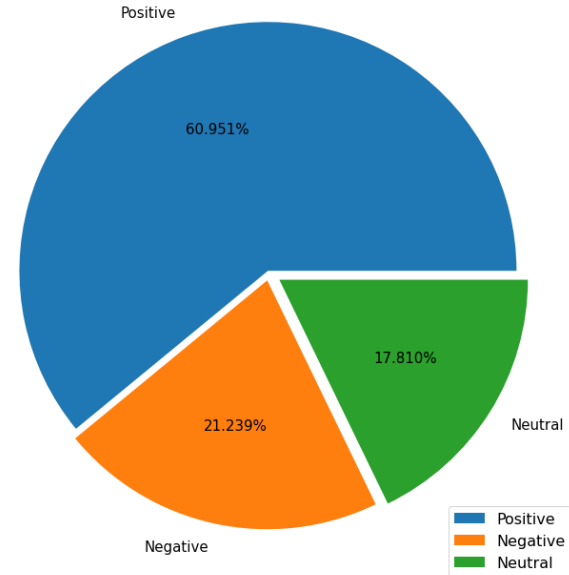
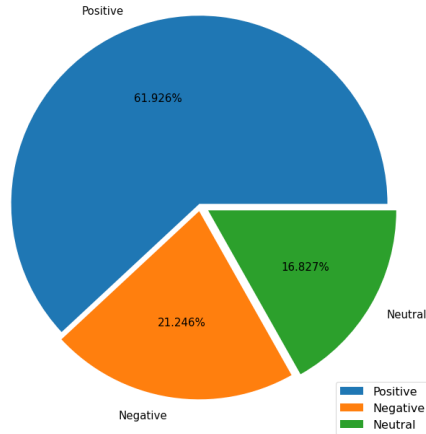


Data Analysis & Visualization(Cont..)

Top 5 Categories Review Sentiments

5.Pie chart representing sentiment percentage of TOOLS category

4.Pie chart representing sentiment percentage of TRAVEL_AND_LOCAL category

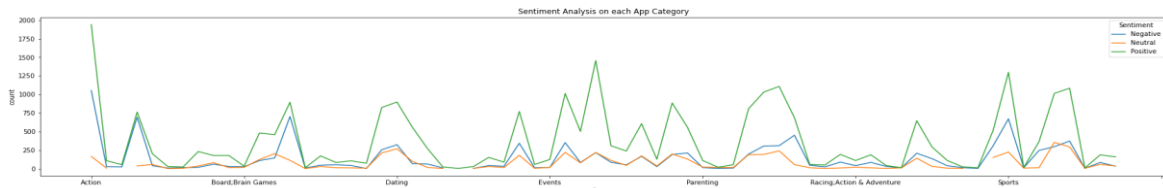


Data Analysis & Visualization(Cont..)

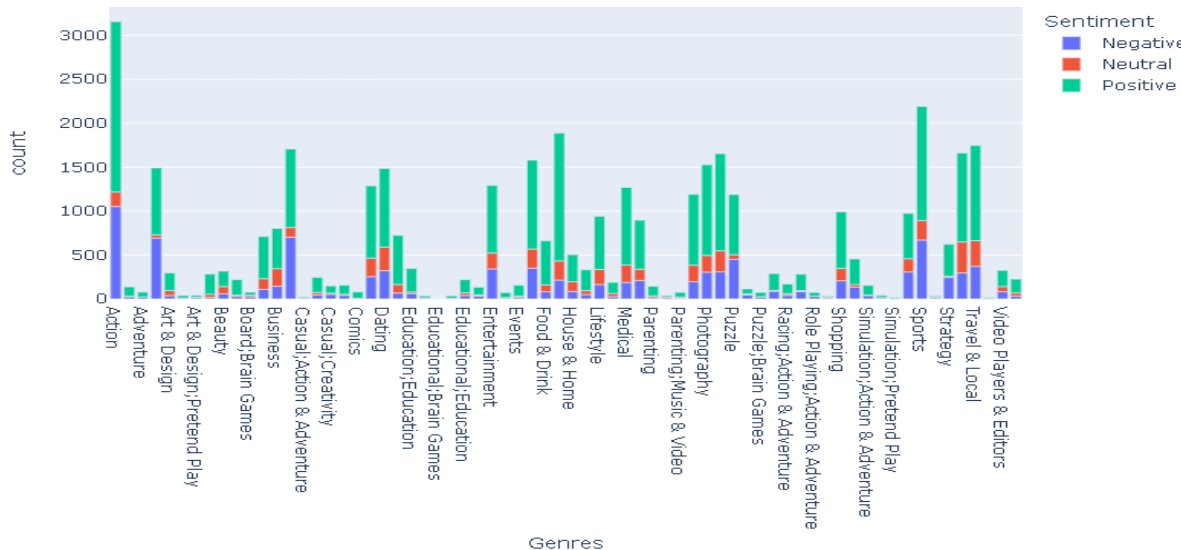
Let's Do Sentiment Analysis of Each Genres

From the graph we can conclude that

- Action Genres has got most positive sentiment in terms of count and it has also highest positive vs negative ratio. So undoubtedly Action Genre apps are most popular.
- Again if we count on the basis of the count of positive sentiment we can see that Casual, Simulation, Educational; Brain games these genres have got least number of positive sentiment and no negative sentiment as such..



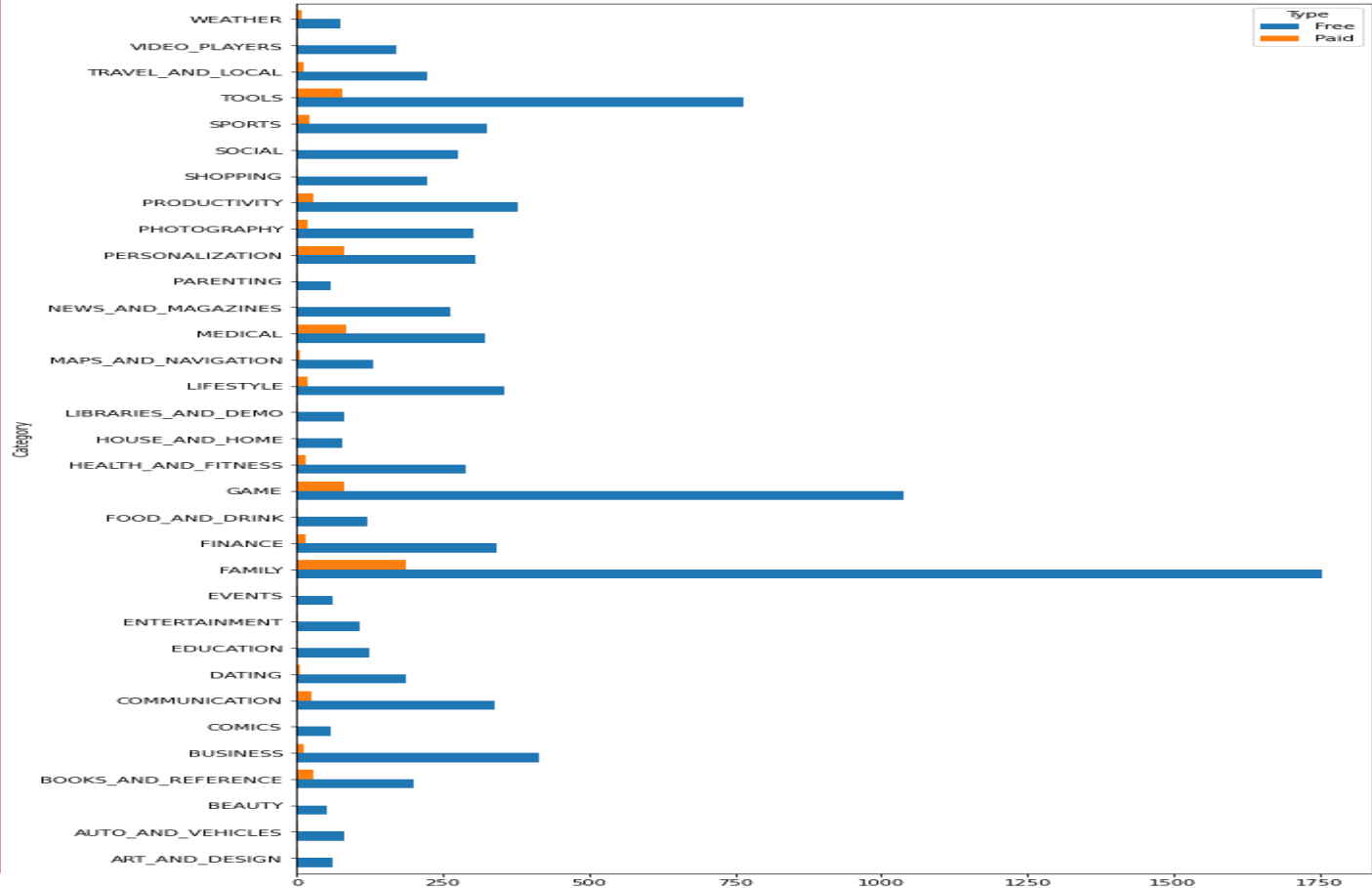
Sentiment Analysis on each App Genres



TO FIND OUT APP IS PAID OR FREE

From the above graph we can conclude that FAMILY category has the most free and paid app on the google play store.

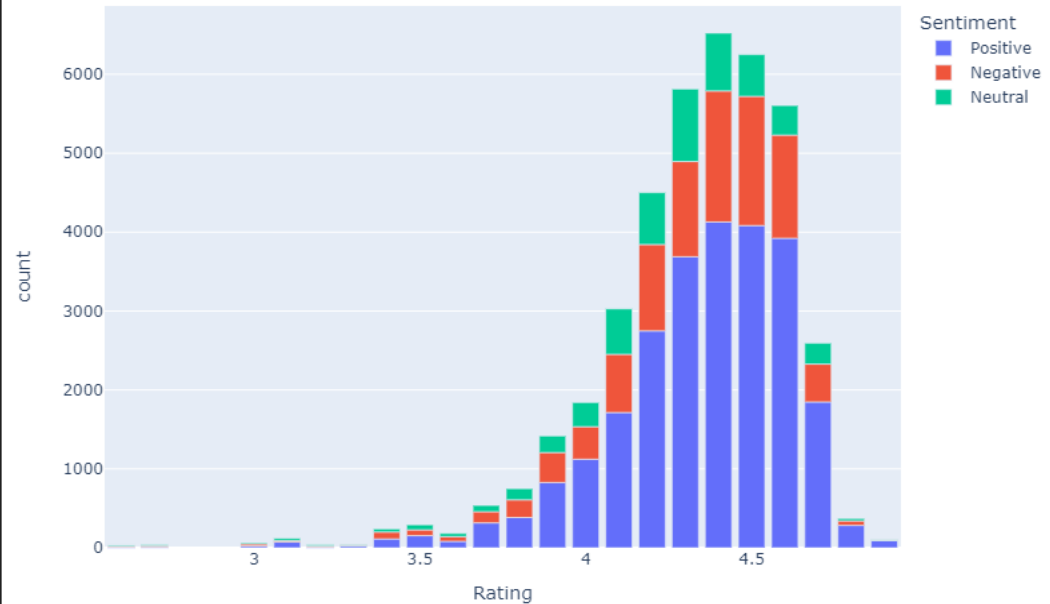
- We can see SOCIAL app is always free, like ENTERTAINMENT, EVENT, EDUCATION, COMIC, and more.
- The MEDICAL has a high amount of paid app considering quantity of MEDICAL app is not that much.
- The second place in paid app is GAME category app followed by the TOOLS category app.
- Google Play Store has more free apps than paid apps.



Sentiment Analysis of each Rating

- From the above graph we can conclude that most of the reviews are Positive.
- Neutral reviews are in the middle range.
- Less people has commented negatively about google play store apps.
- 4.4 rating apps have most positive reviews from 2.5 to 3.1 there are almost equal number of positive,neutral,negative reviews.

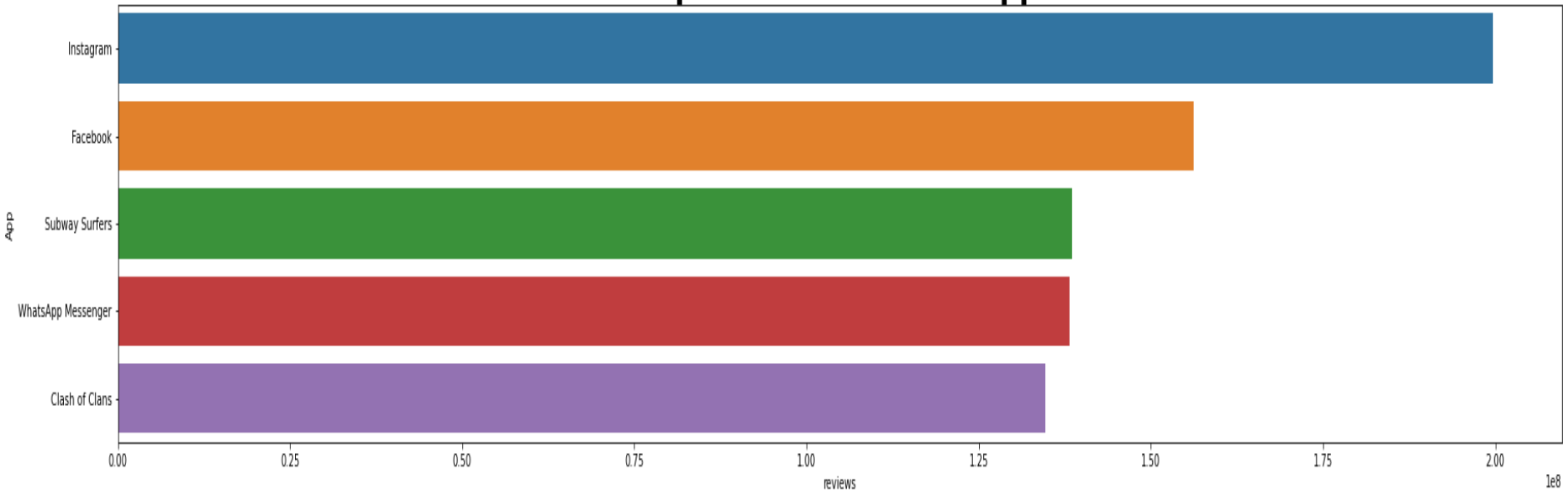
Sentiment analysis of each Rating



Data Analysis & Visualization(Cont..)

- Here we see that, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free & Clash of Clans have highest reviews .
- All these apps have overall reviews between $4.4e^7$ and $7.9e^7$.

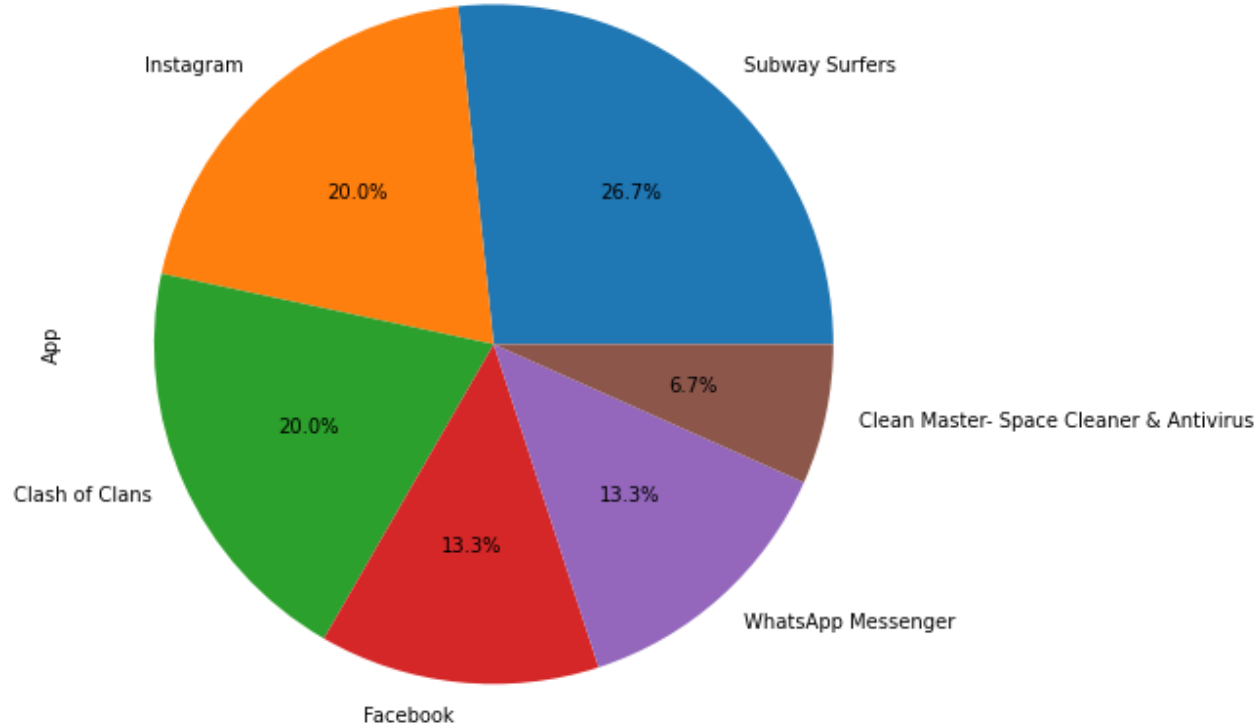
Top 5 Most Reviewed App



Data Analysis & Visualization(Cont..)

Apps which having rating greater than 4 and reviews greater than 1000

- From the above pie chart we found that Subway Surfers app have highest rating along with highest reviews.
- From most reviewed apps, we have seen that Facebook has more reviews but when both rating and reviews comes into picture Instagram app capture it.

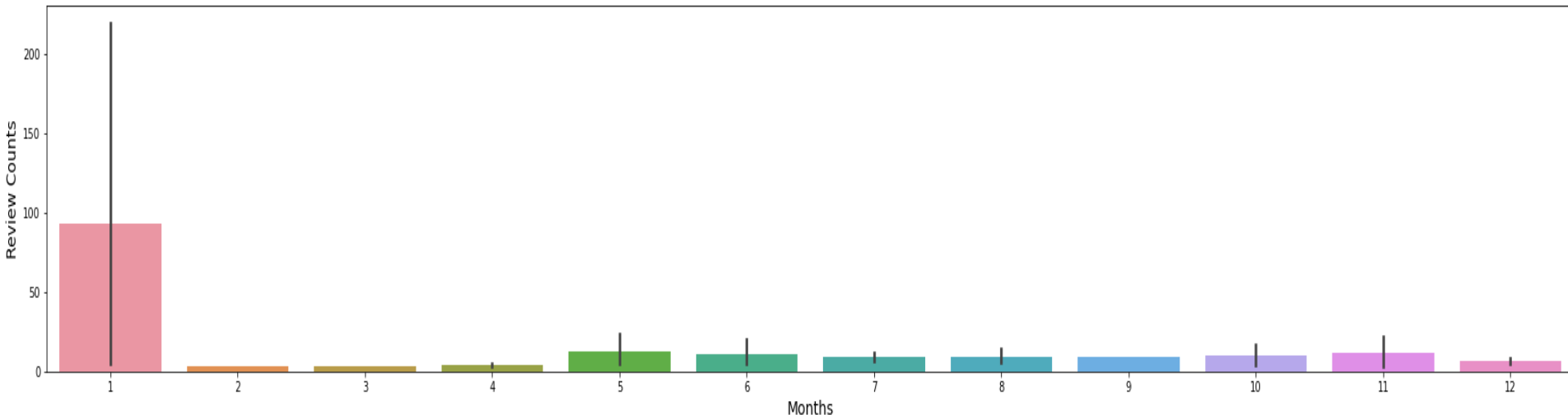


Data Analysis & Visualization(Cont..)

Variation of review with month

- From this bar plot we come to the point that apps have highest review in the month of January and lowest review in February and March.

Variation of review with month

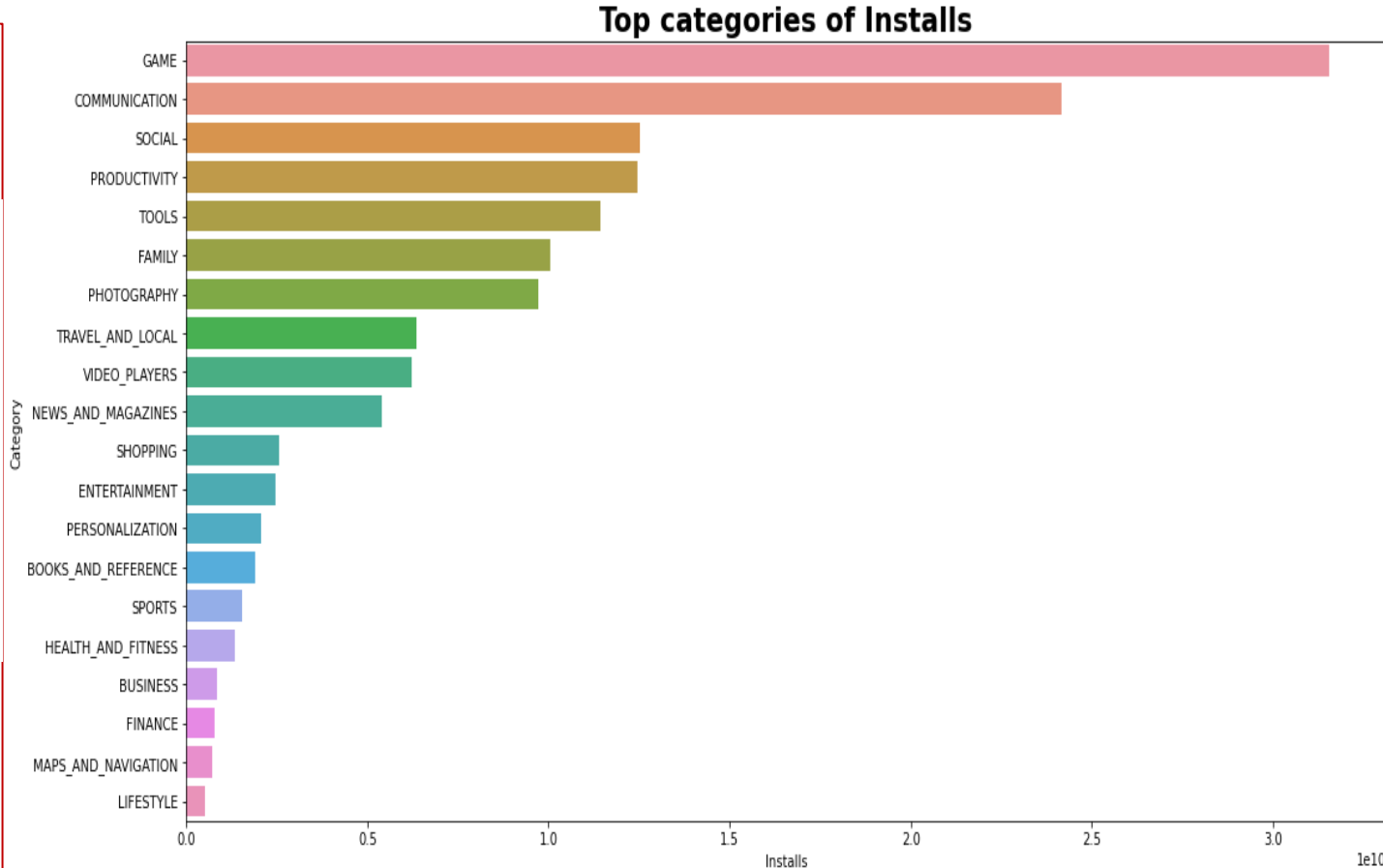


Data Analysis & Visualization(Cont..)

Factors affecting installs

1. Installs vs category

- Top Categories are Games, Communication, Social, Productivity, Tools etc.
- Top 5 categories have over 1cr installations in play store

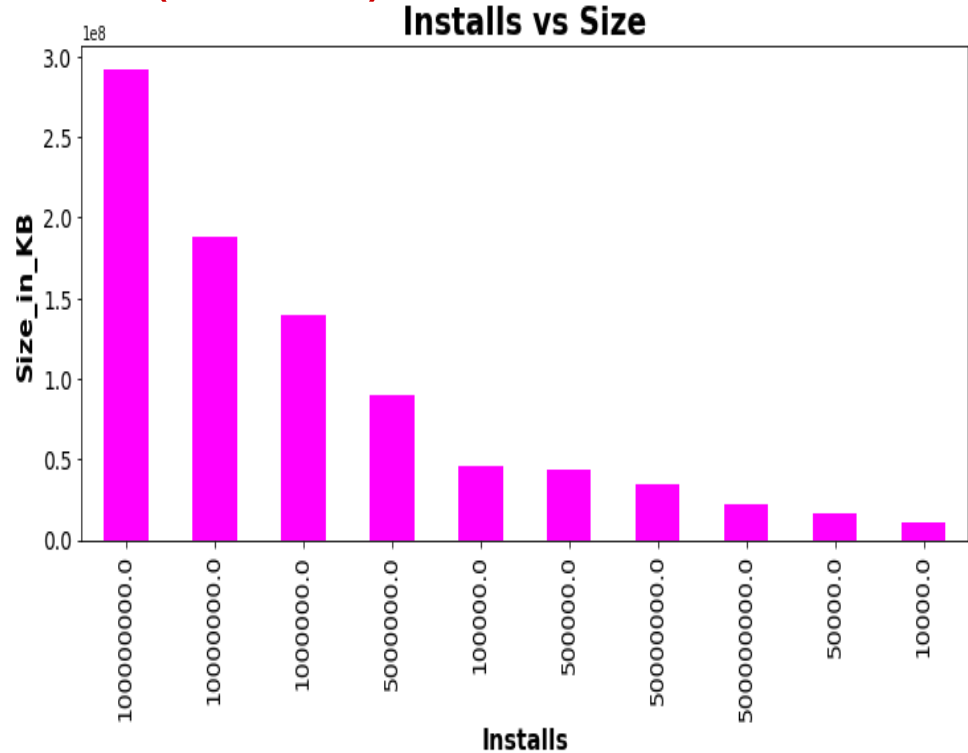


Data Analysis & Visualization(Cont..)

Factors affecting installs

2. Installs vs size

- This graph concludes that Heavy app does not less attractive for customer for installation.
- AS amongst top 10 heavy apps 4 apps are over 1 crore installations and 4 are above 1 lakh installations

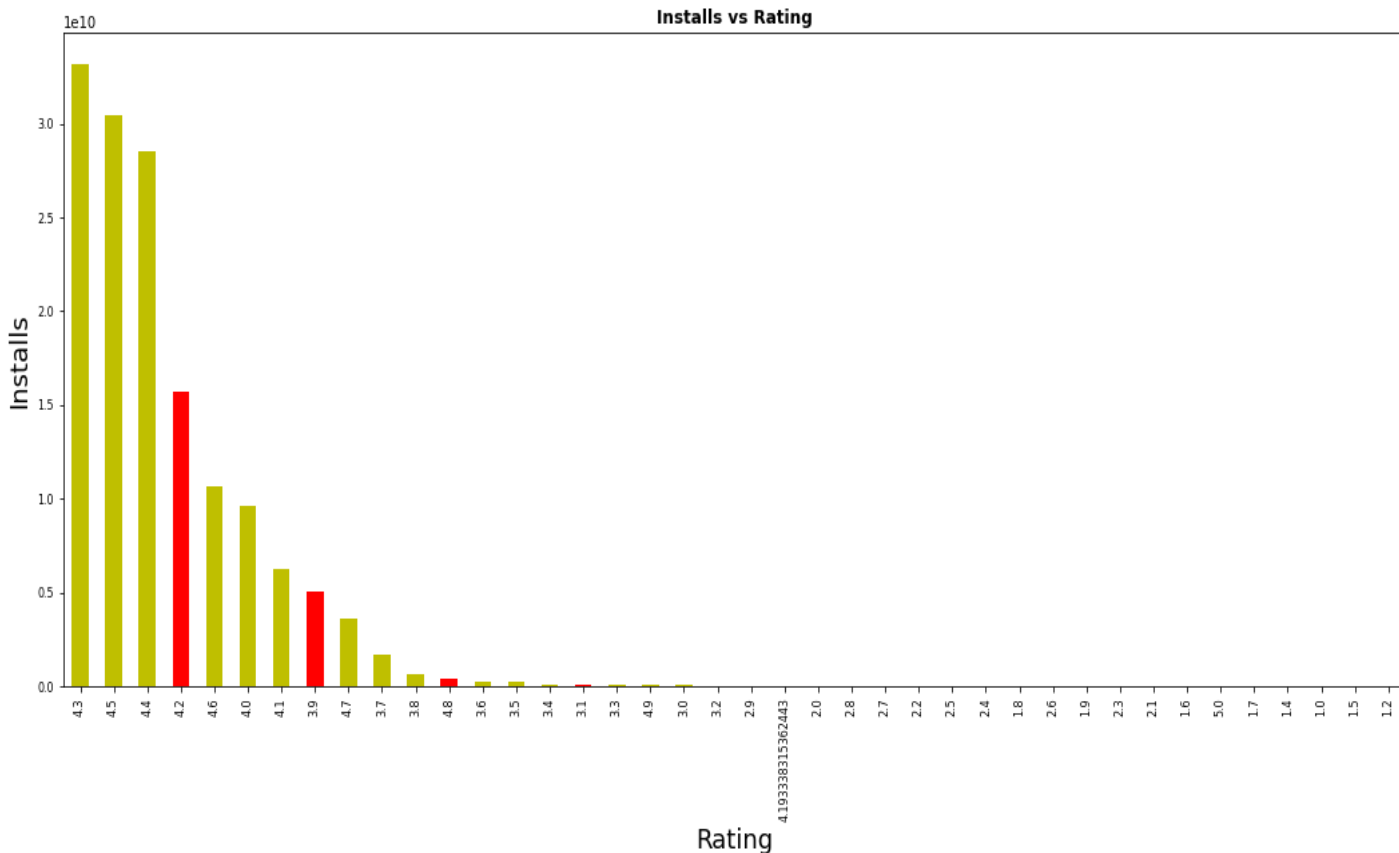


Data Analysis & Visualization(Cont..)

Factors affecting installs

3. Install vs Rating

- Top 3 Ratings are 4.3,4.5,4.4 (yellow bars)
- It suggests good rating does not implies good installations. Certain Rating apps are only enjoys most installation
- Top Ratings apps which have most installs 4.3, 4.5, 4.4, 4.2, 4.6, 4.0, 4.1, 3.9, 4.7, 3.7
- Interestingly below 3.7 and above 4.7 rated every apps enjoys very less installs compared to top 10 ratings

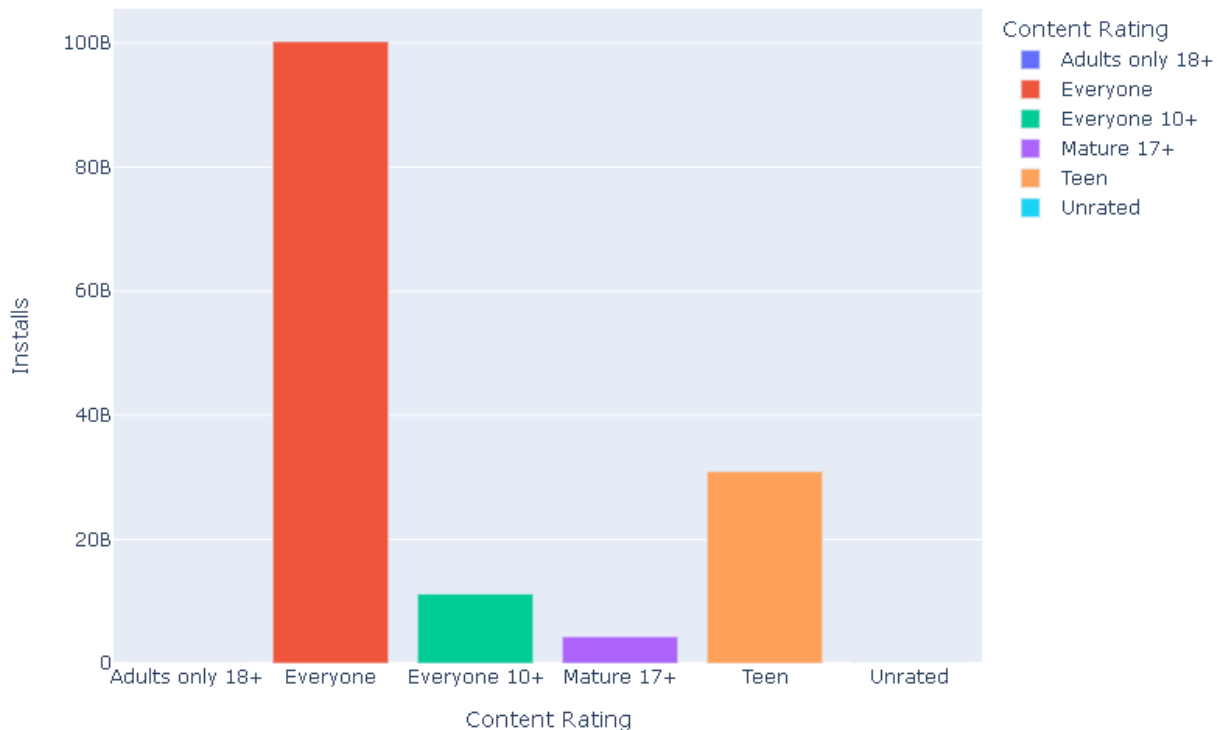


Data Analysis & Visualization(Cont..)

Top Content Rating per Installation

- App Having 'Content Rating' - Everyone is having maximum number installation of 100.228 Billion
- Also we can conclude that there are 4.29 Billion installation for application having Content Rating of Everyone 10+.
- We can also conclude that there are no such noticeable installation for Application having Content Rating of Adults only 18+ and Unrated.

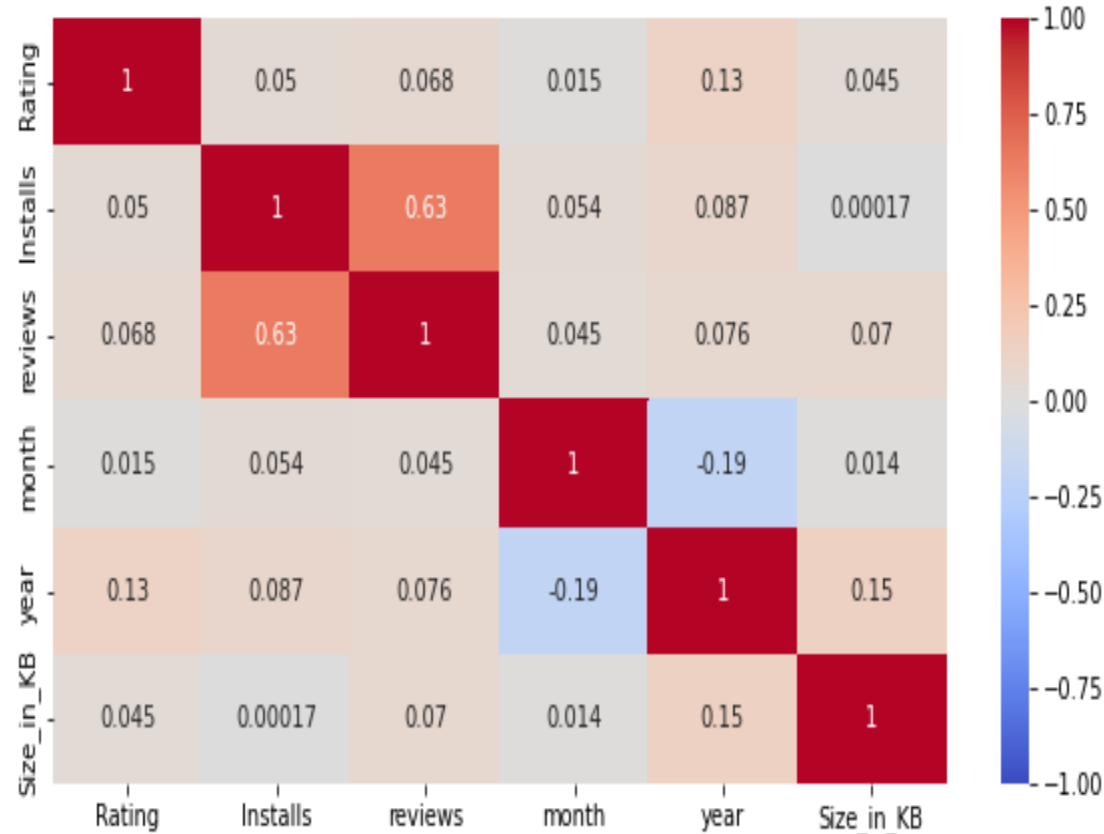
Top Content Rating per installation



Data Analysis & Visualization(Cont..)

Correlation

- Above graph has only two good correlations of 0.63 between reviews and installs.
- It is positively co-related means Most installed apps got most reviews and less installed app got less reviews.



Conclusion

- category Family and the genre tools are in large numbers. But interestingly Category EDUCATION and Genres 'Board;Pretend Play' and 'Comics;Creativity' both get the best average rating in playstore.
- App Rating is directly proportional with the recent update. From this we can be sure that with the latest updation the reviewers are giving better rating.
- Rating are directly proportional with Average Reviews till 4.5. But after that we observe some deviation and eventually reviews decreases after 4.5 ratings. 4.5 Rating apps have maximum 1.1 million reviews. Positive review sentiments are 63% and negatives are 24.9%, so it concludes that developers has to work hard to convert more negative sentiments to neutral or positive.
- most of the opinion on sentiment subjectivity lies high in the range 0.4 to 0.7.It means people give review after some experience.
- After analyze the data we found, though GAME category has maximum no of translated reviews but, Most positive percentage acquired by HEALTH_AND_FITNESS category and GAME has most negative reviews among top 5 categories. Sports App Category has received the highest neutral sentiments. so it concludes that most reviews does not suggests most positive sentiment percentage.
- We found that Action Genres got most strong positive sentiment among all genres. It suggest that people prefers Action genres app far more than any other genres.

Conclusion

- We also found people's sentiment are more towards paid app and most of their sentiments are positive towards playstore app. It concludes that they are happy with playstore apps and they expect much from paid apps.
- google playstore has more free apps than paid. and Family category has most free and paid apps in playstore. MEDICAL category has highest amount of paid apps considering quantity followed by GAME,TOOLS Category.
- Social,entertainment,event,education,comic* etc these popular category apps are always free. It concludes that people are not willing to pay for entertainment,education apps but for health and medical purposes people are ready to pay.
- We found that in terms of both Top Total and Average Apps Reviews outcomes Facebook,Instagram,whatsapp messenger,Clash of Clans these apps are common. so we can conclude that these 4 apps are most populars and trending amongs masses.
- Subway Surfers app has highest rating along with highest reviews followed by Instagram,clash of clans,Facebook,whatsapp messenger,clean master.
- After analyze the data we got highest reviewed apps are those which are last updated in january,2018 and got lowest review in february and march updation. It concludes that Most of the company did major updation of their apps in january month, so in that months they get most reviews.
- 4.4 rated apps are most reviwed apps in google playstore and from 2.5 to 3.1 there are almost equal numbers of positive,negative,neutral reviews. It states that neither top rating nor less rating makes an app good or bad. Most positive sentiment apps are in middle range of 4 and 4.5 rating always.

Conclusion

- We found that GAME, COMMUNICATION, SOCIAL these category apps got most installation (above 1 crore).
- Most heavy app got 10cr installations, amongs top 10 heavy apps 4 apps got above 1 cr installations and 4 apps got above 1 lakhs installations. so, it concludes that heavy apps are still not obstructs customer from installing.
- Most Installed apps are also rated between mostly 4 and 4.5. More specifically they are 4.3, 4.5, 4.4 rated mostly. Less installed apps are either below 3.7 rated or above 4.7 rated. It concludes high rated near to 5.0 rated apps have not much organic reviews and below 3.9 rated apps have failed to satisfy customers so it gets less installation and rating.
- We can also conclude that most of the apps which are used by the users have a content rating of 'Everyone'.
- From install, rating, reviews co-relation graph we also found install and reviews are positively co-related means when Installation is more reviews are more, when it lacks, reviews are also less.
- **Important Points to be noted:**
 - All active apps on playstore has an an Average Rating of 4.32.
 - Also we can see that after merging both of the dataset the maximum Average Rating is 4.9.
 - Also the average sentiment Polarity is 0.16 and average sentiment_subjectivity is 0.49
 - Also we have noticed that the average size of the Application available on playstore is 21933.38 KB
 - lastly, all of the calculations and graphs in this project have accuracy in the range of 75% to 80%.

THANK YOU