**Play Store App Review Analysis**

**Team Members**

**Saugata Deb, Harshad Savle, Prasoon Kumar**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Android is an operating system which has

Currently 2.8 billion users are active around the globe. In this project we will be doing analysis on Play Store App Review Data. With the development of android initially it was an operation system to be used in camera but when google acquired android in 2005. They started giving android as an operating system to smartphones in 2007 which was later going to dominate 1/3rd of the world's total population as what we are experiencing today. Android is basically an open source OS giving the developers an opportunity to develop applications in this data ridden world because of this it has become necessary for the data analytics professionals to analyze those data so to help companies realize how they can grow their business effectively in every aspect and learn the needs of the users for the betterment of their products and also help App developers how they can improve their products in different categories.

**1.Problem Statement**

Data has been provided by ABC Advertising Pvt. Ltd. to analyze the data for different App in different categories and help them satisfy their customers needs who possibly

resides around the world. From these analyses they can understand the needs of their customers in different aspects of product development and it will help solve the problems with the market superiors.

The main objective of this project is to deal with the data provided by the company and to analyze the data in every aspect possible so to help them match their idea and help them to convert new and retain old customers and make reasonable growth.

We have been provided with two datasets:

1. Play Store Data.csv
2. User Reviews.csv

**Play Store Data.csv** column elements:-

1. **App** - Name of the Application
2. **Category** - Category of the Application
3. **Rating** - Rating given to the Application
4. **Reviews** - No of reviews given to the Application
5. **Size** - Size of the Application
6. **Installs** - No of downloads of the Application
7. **Type** - Free or Paid
8. **Price** - Price of the Application if it is paid
9. **Content Rating**-It is Age appropriate or Not
10. **Genres** - Type of Genre the Application belongs to
11. **Last Updated** - When the last time the Application is Updated
12. **Current Ver** - Current version of the Application
13. **Android Version**- Minimum Android version required to run the Application

**User Review.csv** column elements:

1. **App**:- Type of Applications
2. **Translated\_Review**:- Reviews being given by consumer
3. **Sentiment**:- Sentiment of trust from customer
4. **Sentiment\_Polarity**:- It determines sentimental expression of the customer's opinion
5. **Sentiment\_Subjectivity**:- Sentimental Subjectivity in terms is a personal opinion and it falls in range [0,1].

**Problem Questions:**

1. Top Categories in Playstore?
2. Top Genres in the Playstore?
3. Top Content Rating per installation?
4. What is the percentage of free and paid Apps in the Play Store?
5. What is the effect of the last update on rating?
6. How does the last update have an effect on the trend of rating?
7. Effect on rating when the application was of type free and paid?
8. Relationship between reviews and rating?

## relationship between Rating and Average Reviews

1. Average Rating of each App category

## Average Rating for each genre

1. What is the distribution of sentiment subjectivity?
2. How sentiment polarity varies with Free and Paid Apps?
3. Different percentages of review sentiments based on two Datasets provided?
4. Different percentages of sentiment analysis on top 5 Reviewed App Categories?
5. Sentiment Analysis on each App Category
6. Sentiment Analysis on the basis of Genres

**2. Introduction**

### Android as we know is a huge marketplace where every app developer has the opportunity. When Google acquired Android in 2005. Google incorporated the market with android in 2007 which have let developers develop android applications and in this way with very less association with their own developer and more on open source developers it opened the markets of opportunities for both the makers and the user as well what we can easily speculate today. This has also helped in the development of many new businesses and also several new professions which we can experience now.

### There are basically three kinds of Applications in the Android Play Store. ‘**Background Services and Intent Receivers Applications**’,’**Foreground Background Application**s’ and ‘**Intermittent Applications’**.

The type of Applications which we will use to analyze in Play Store App Review Analysis mostly will be Intermittent type Applications.

In this project of Play Store App Review Analysis we will be analyzing each and every prospect of the data available with us and on the basis of this data we will try to solve every problem associated with the real world problem which lets them achieve customer prosperity.

## **3. Types of Android Applications**

### Background Services and Intent Receivers Applications

### Foreground Background Applications

### Intermittent Applications

### The type of Application where we will do the analysis can be any of the above.

## **4. Steps Involved in Data Cleaning:**

## **EDA(**Exploratory Data Analysis) **on two given Datasets.**

1. Play Store Data**(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', 'Android Ver'])**

1. User review**(['App', 'Translated\_Review', 'Sentiment', 'Sentiment\_Polarity', 'Sentiment\_Subjectivity'])**

After analyzing both of these datasets:

1. There are **10841** rows and **13** columns in **Play Store Data** and **64295** rows and **5** columns in **User Review.**
2. Except for **Rating** all the column values are object type which is type float.
3. Value of the **last Updated** column is in string.
4. Value of ‘**Size**’ is in string.
5. So **float64** type column value is **1** and **object** type column value is **12.**

* **Treatment of null value**

Removal of null values is necessary for each row to get effective analysis output in order to terminate negative data and work on useful data.

* **Treatment of Duplicate Rows**

After removal of null values another important task is the removal of duplicate row values if they exist. With the removal of duplicates we can now move forward with the problem statement without having unnecessary multiplied data.

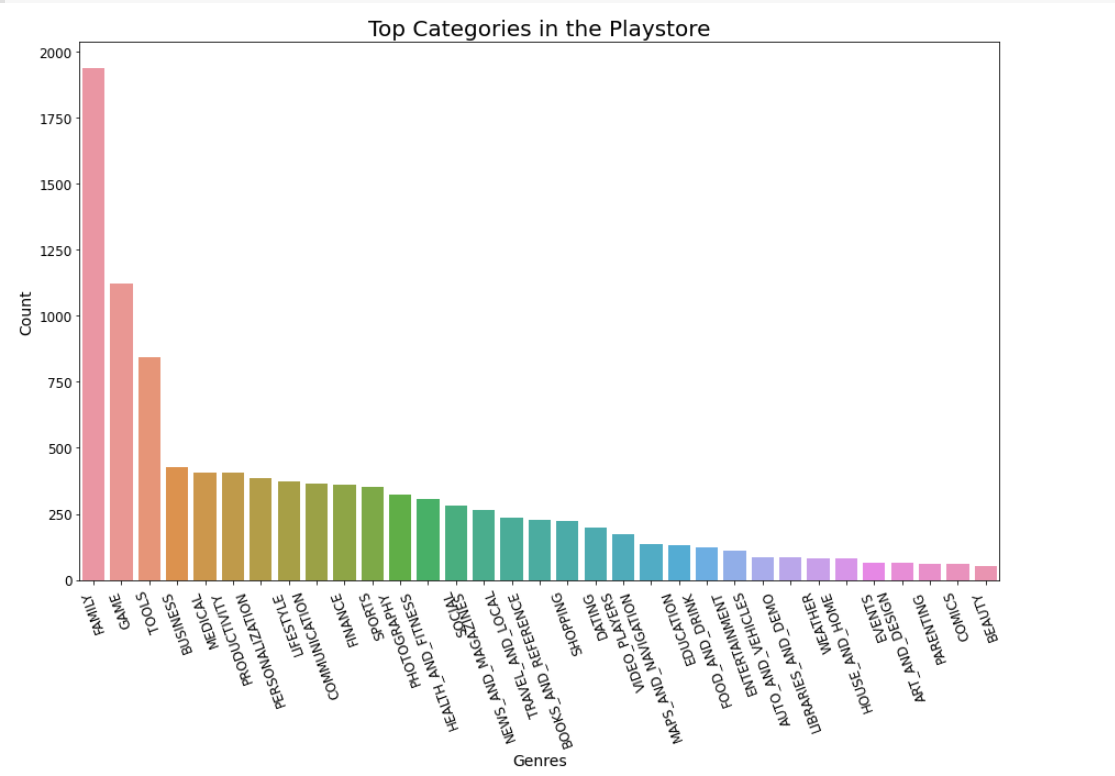
* **Conversion of necessary column values from string to required datatype:**

This is the very important task in order to get proper data visualization and analysis as well. As in our case except **Rating** every other column has a string type data type which is not required.For example: **Size** we need to have this column in float or int in order to analyze, similarly with **Last Updated** column we need this column value in **Datetime Format** and not string to get requisite data visualization.

* **Replacing the missing values with mean,median or mode**

This method is also important in terms of getting output visualization. We can replace the missing value with either of the three depending on the data in the dataset. In this project we are replacing the missing values with median value in **Rating** column.

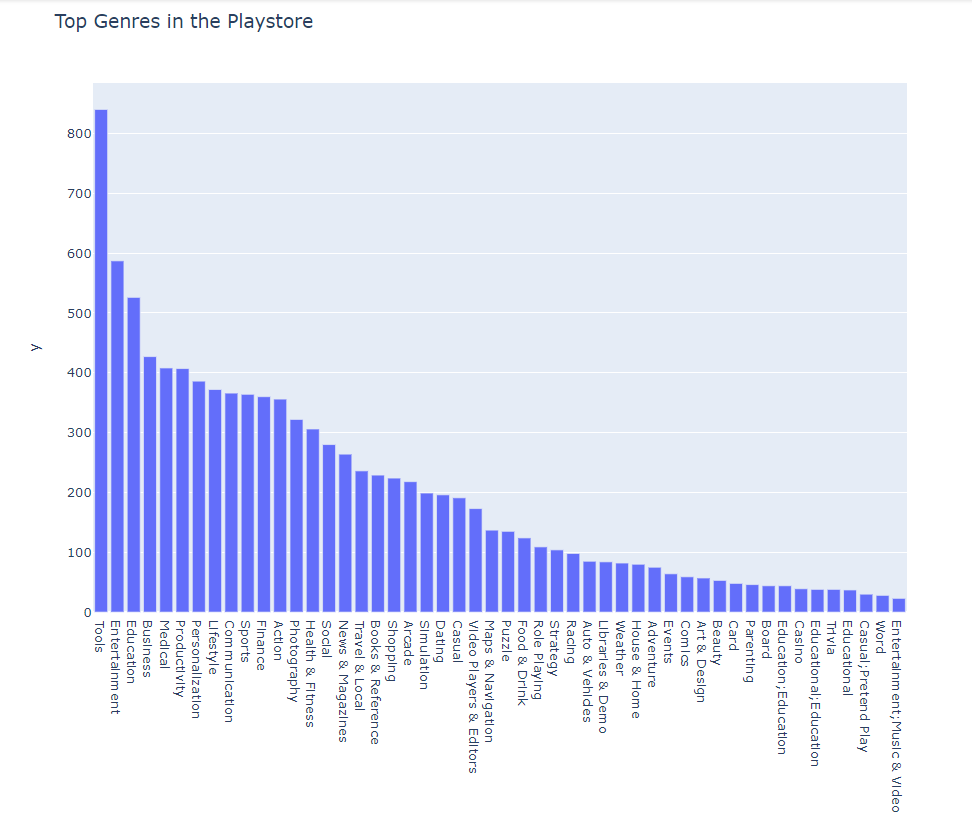
**5. Distribution of top App Categories in playstore**



* From the above graph we can conclude that there are 1939 applications which fall under the FAMILY category.
* Also top 5 categories of Application in the playstore are FAMILY,GAME,TOOLS,,BUSINESS,MEDICAL.
* Also there are only 53 applications which fall under the BEAUTY category.
* We can also conclude that there is significant difference between the top two categories FAMILY 1939 Apps and GAME 1121 Apps. This shows that there are more application

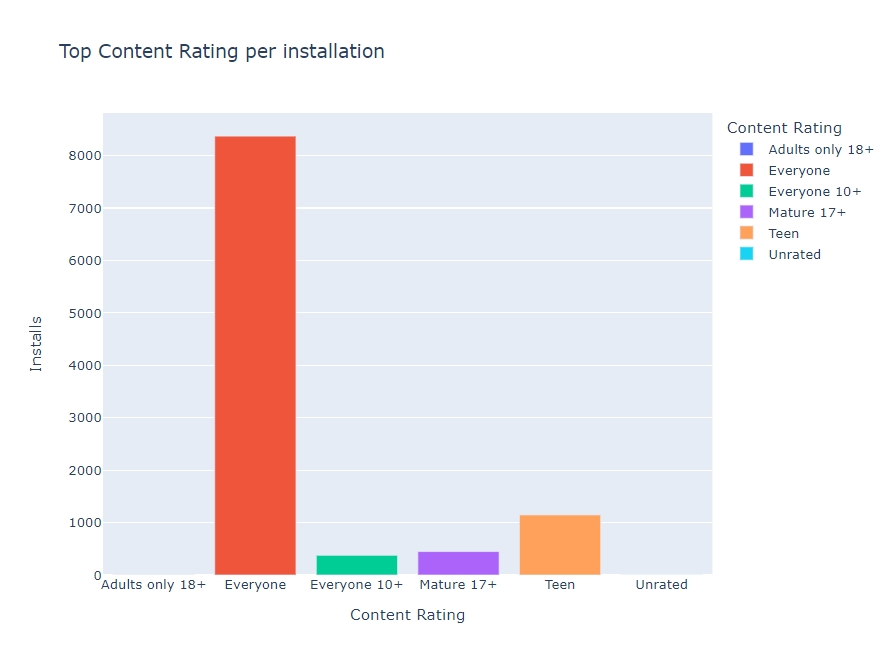
developers who develops Apps under FAMILY Category.

**6. Distribution of top App Genres in playstore**

****

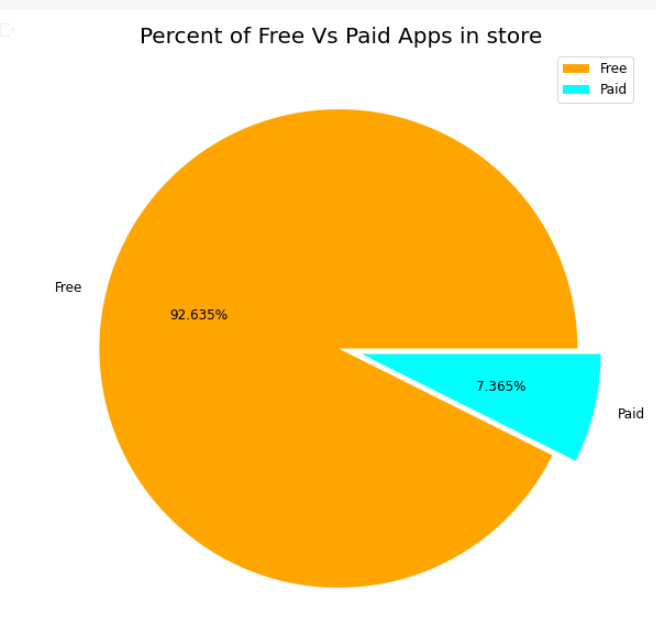
* From above bar plot in plotly we can conclude that maximum application which have been developed falls under App Genre Tools 840.
* Also from above we can observe from the above plot that least applications were developed under App Genre Entertainment,Music & Video 23.
* Also top 5 Genres are Tools,Entertainment,Education,Business and Medical which are 840,587,427,407,460 in Top 5 Genre name order.

**7. Top Content Rating per installation**

****

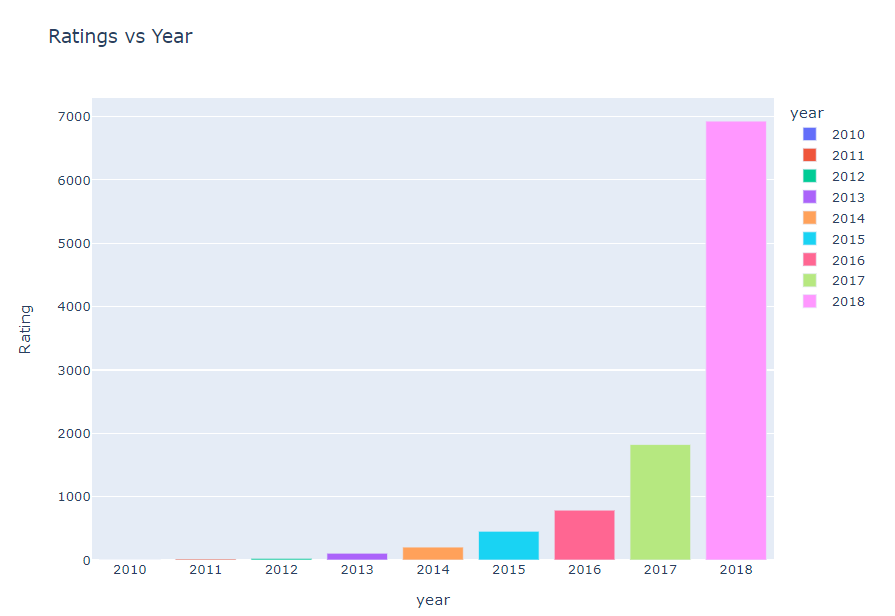
* From the above graph plotly we can conclude that App Having 'Content Rating' - **Everyone** is having maximum number installation of **100.228B**.
* Also we can conclude that there are **4.19B** number of installations for applications having Content Rating of **Everyone 10+**.
* We can also conclude that there are no such noticeable installations for Application having Content Rating of **Adults only 18+** and **Unrated**.

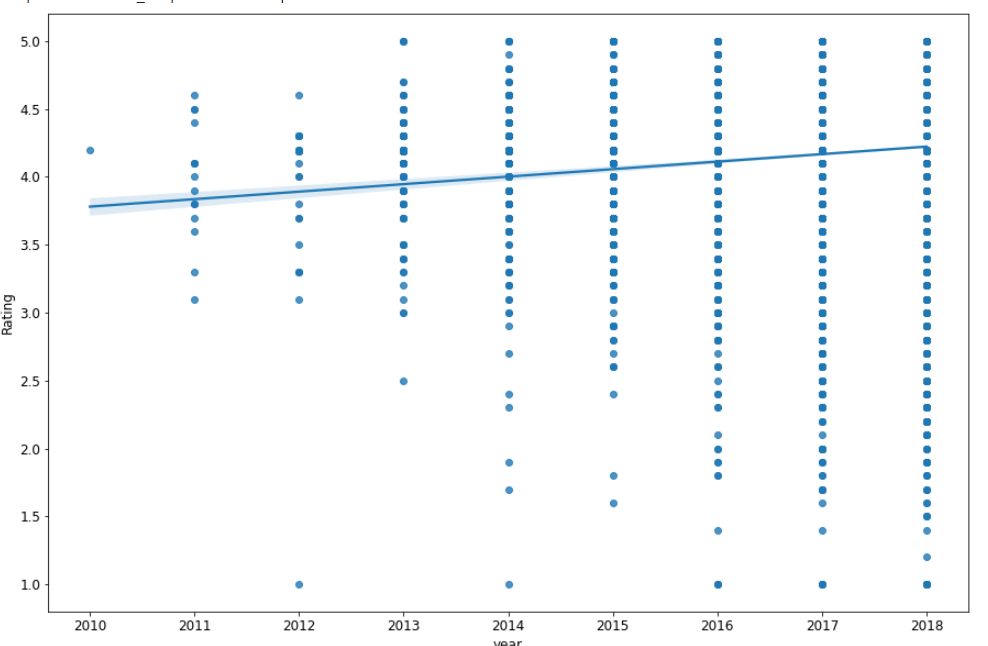
**8. Percentage of the type of App present in Play Store Free vs Paid.**

****

* From the above chart we can conclude that most of the apps available on the playstore are free which are enjoyed by most of the users.
  1. Free Apps= 92.635%
  2. Paid Apps= 7.365%

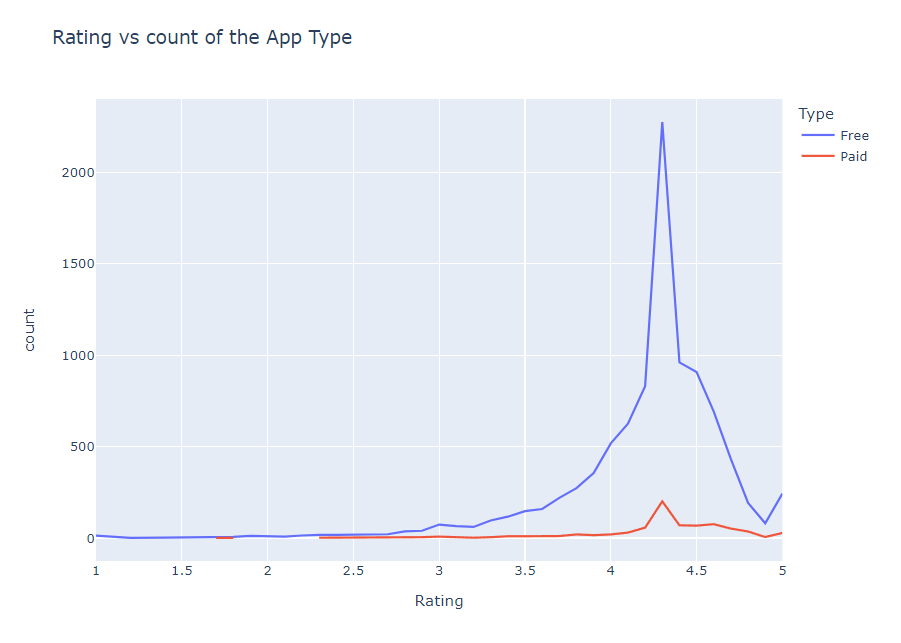
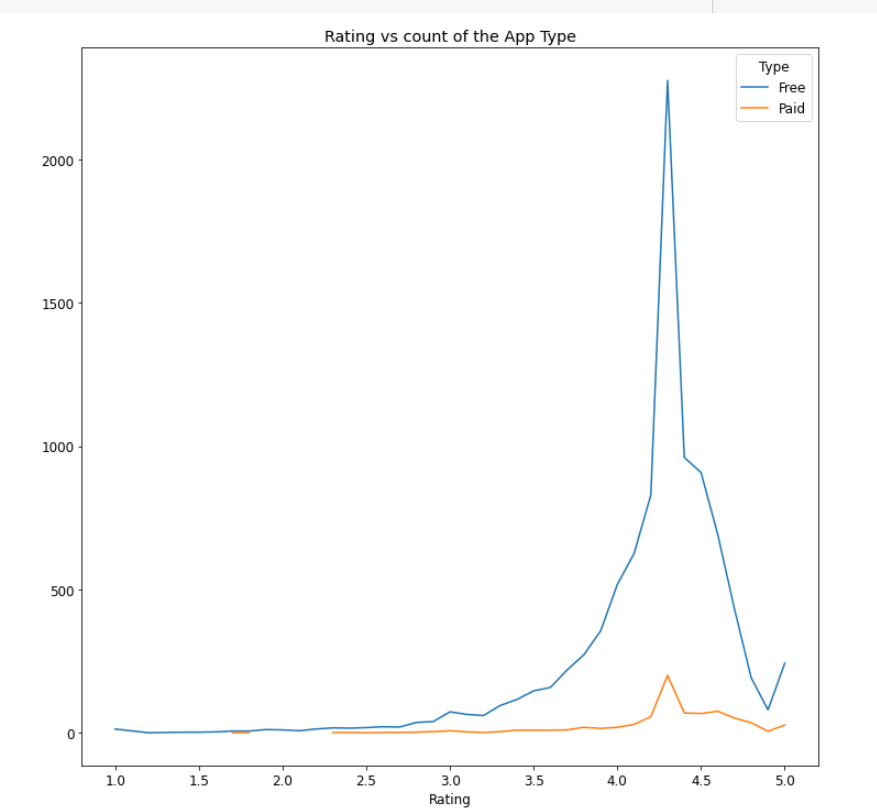
9.**Effect of last update on rating**

****

****

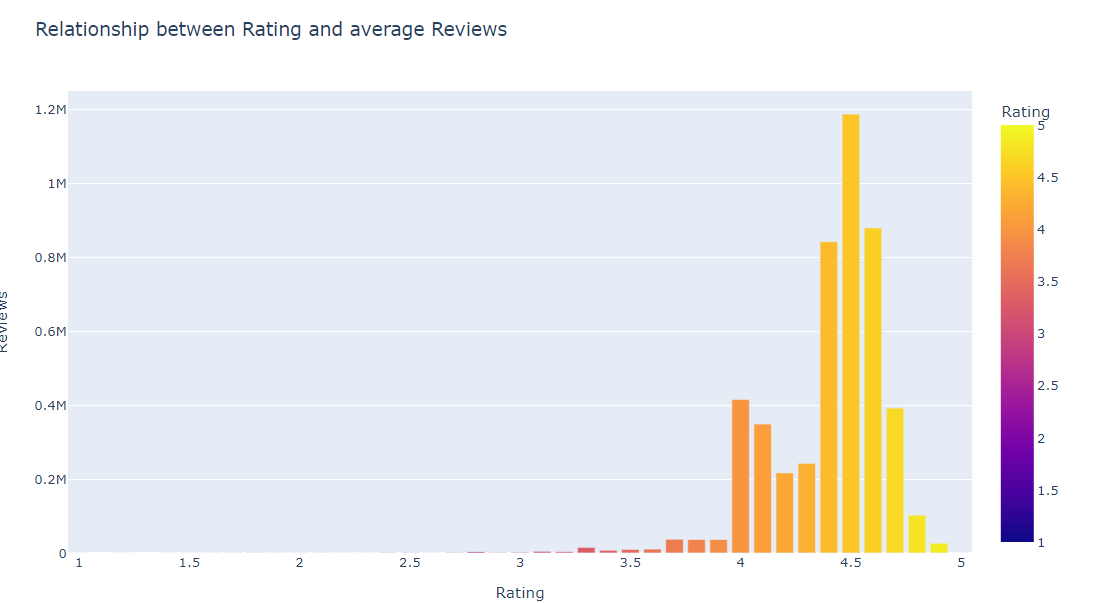
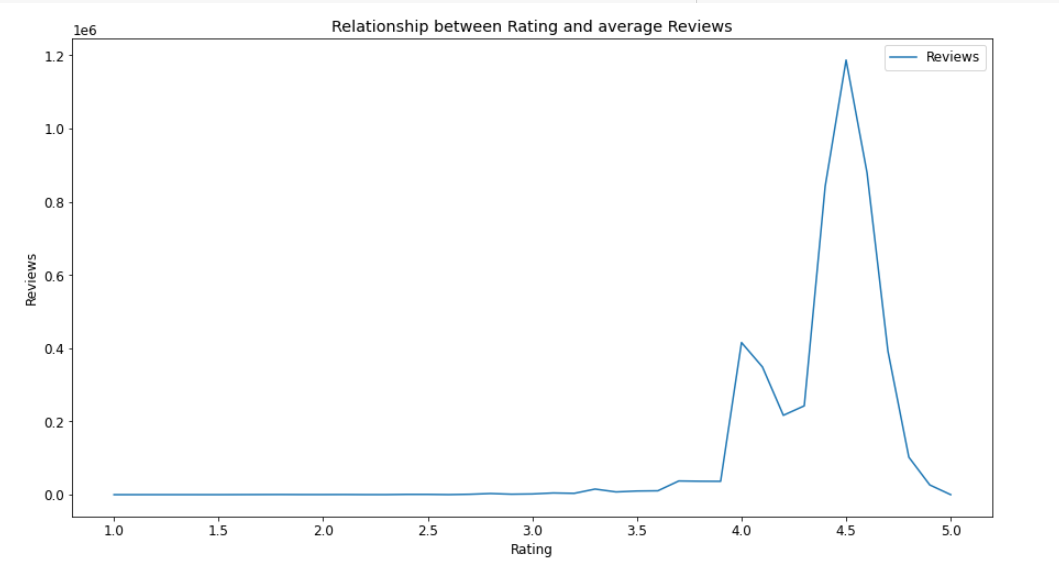
* From the above graph from plotly we can conclude that there are more number ratings given to the applications which were updated recently in 2018, no. of rating is 6929 than those applications which were updated in 2017 no.of rating 1823. This shows with the latest update user reviews response increases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

**10.Effect on rating when the application was of type free and paid**

****

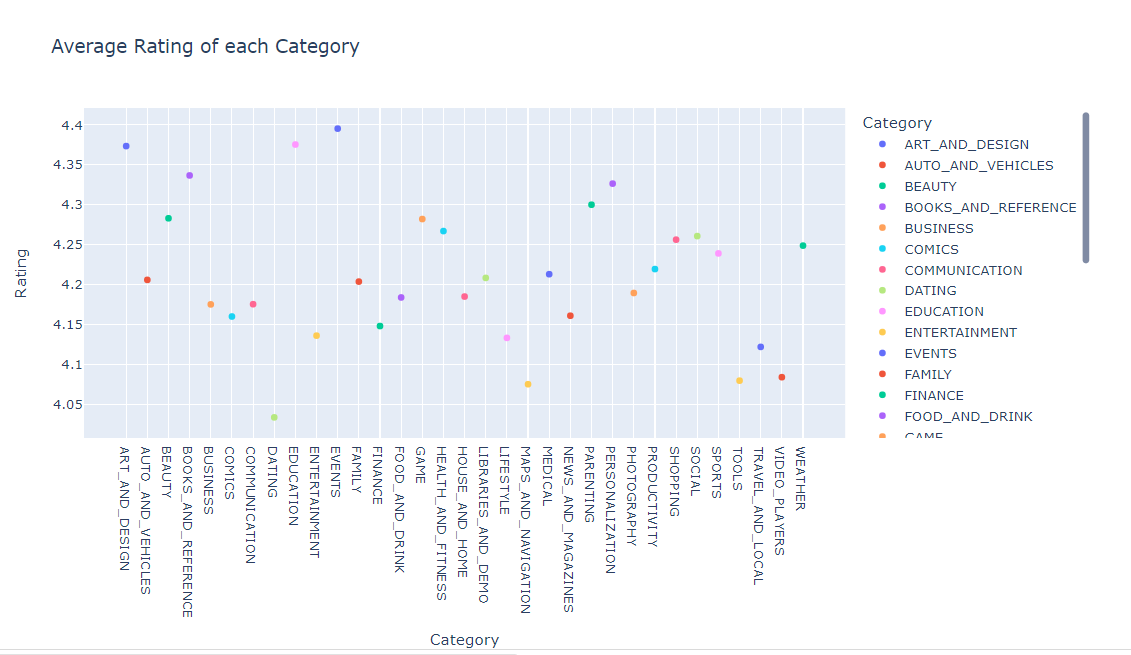
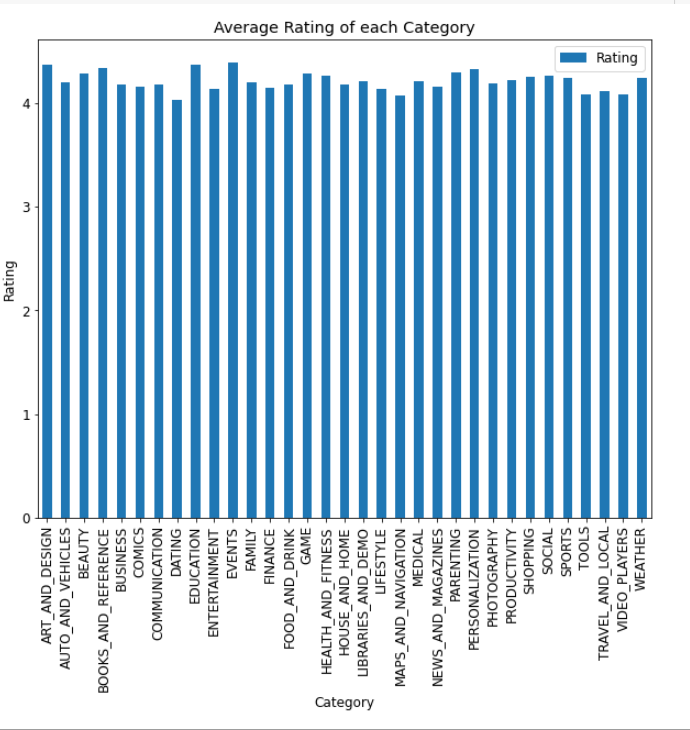
* From the above graph using plotly we can conclude that the Free Apps has got more rating in terms of number of number of ratings. From this we can also say the users of free apps are way higher than those who use paid apps.
* But when we try to watch very clearly the highest rating for both the Free and Paid Apps are 5. But if we go on finding the average of both the types of App we can see that the Average rating of the Free Apps will be less as compared to that of the Paid Apps . Again for the same that no. of users in Free Apps are way too high . For Example Free App for Rating 4.3 has a number count of 2275 as compared to only 201 count of rating for paid Apps with same Rating 4.3.

**11.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 of 1.1M and reviews increases in quite proportionate manner with the increase of rating but after 4.5 rating reviews eventually decreases and we can see only around 26586 reviews.

**12.Average Rating of each App category**

****

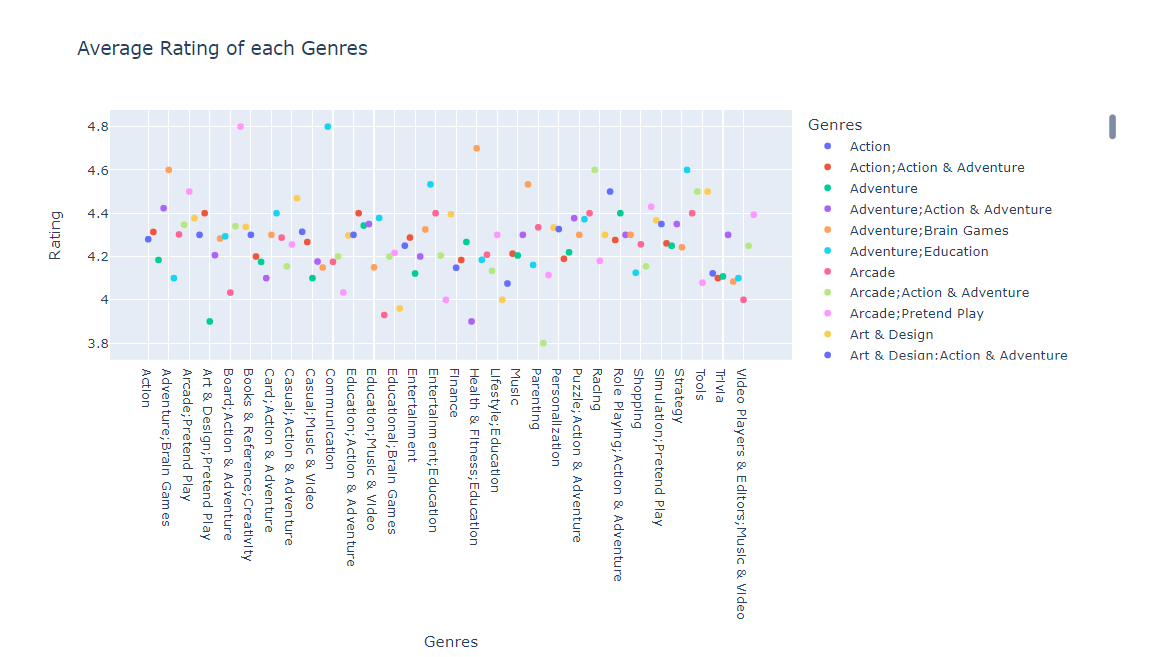
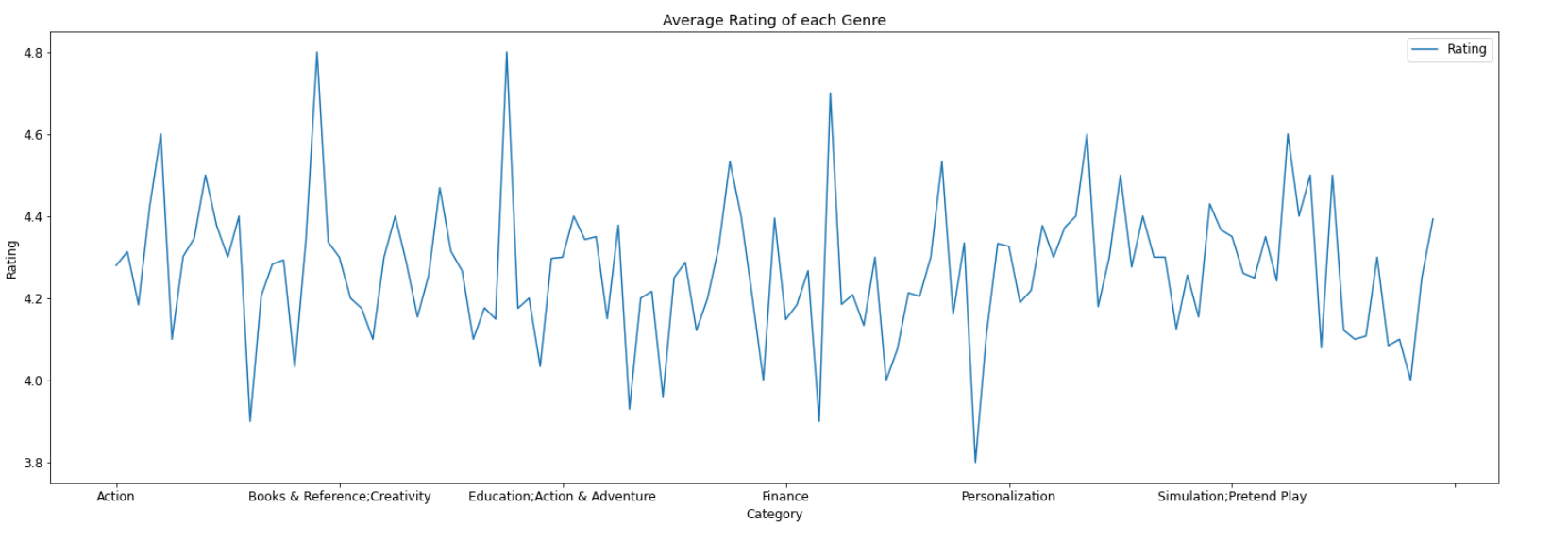
* From the above plot from plotly we can conclude that the 'Event' category has got the highest Average Rating of 4.39.
* Also we can see that 'Dating' Category has got Lowest Average Rating of '4.033673'

## 

## 

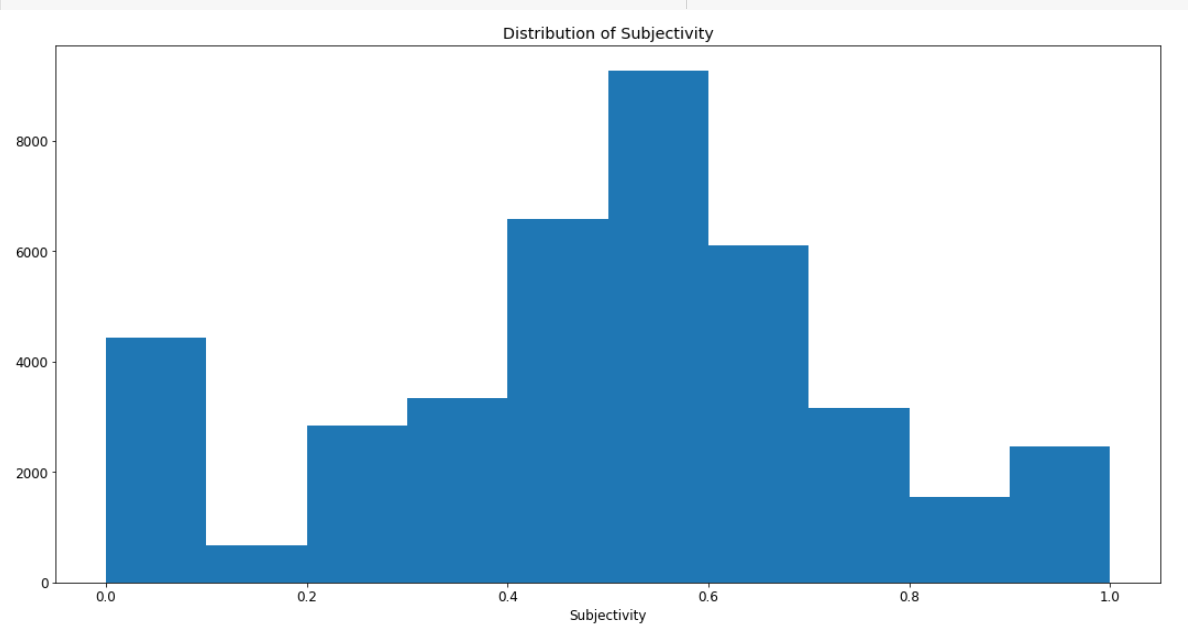
## 

## **13.Average Rating of each App Category Genre**



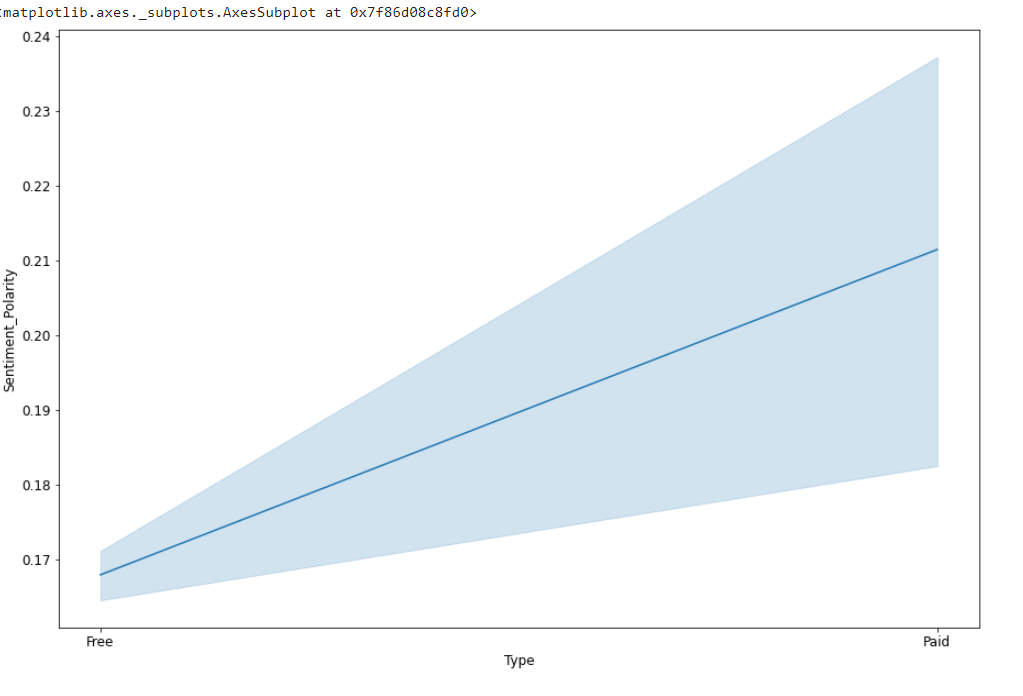
* From the scatter plot from plotly we can conclude that both the Genre-'Board;Pretend Play' and 'Comics;Creativity' have the highest Average Rating of 4.8 and Genre-'Parenting;Brain and Games' has the lowest Average Rating.

**14.Distribution of sentiments subjectivity:**

****

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

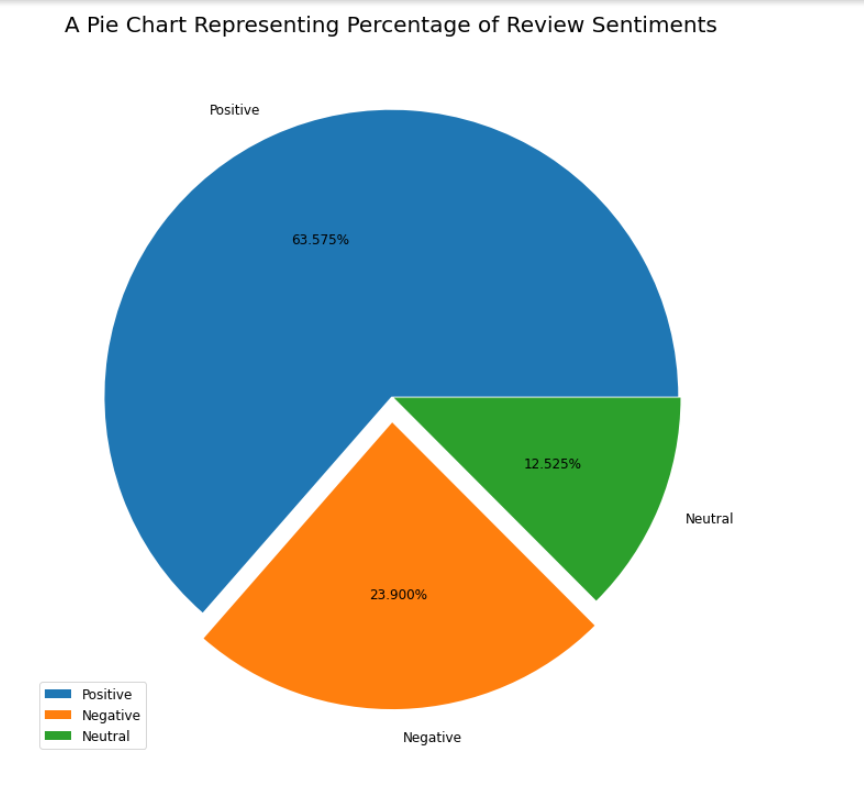
**15.Sentiment polarity variation with Free and paid Apps**

****

From the above line plot we can conclude that with an increase in sentiment polarity ,the sentiment polarity for the paid app is higher than the sentiment polarity for the free App. This means people have more sentiment towards paid apps than free apps.

**16.Different percentages of review sentiments based on two**

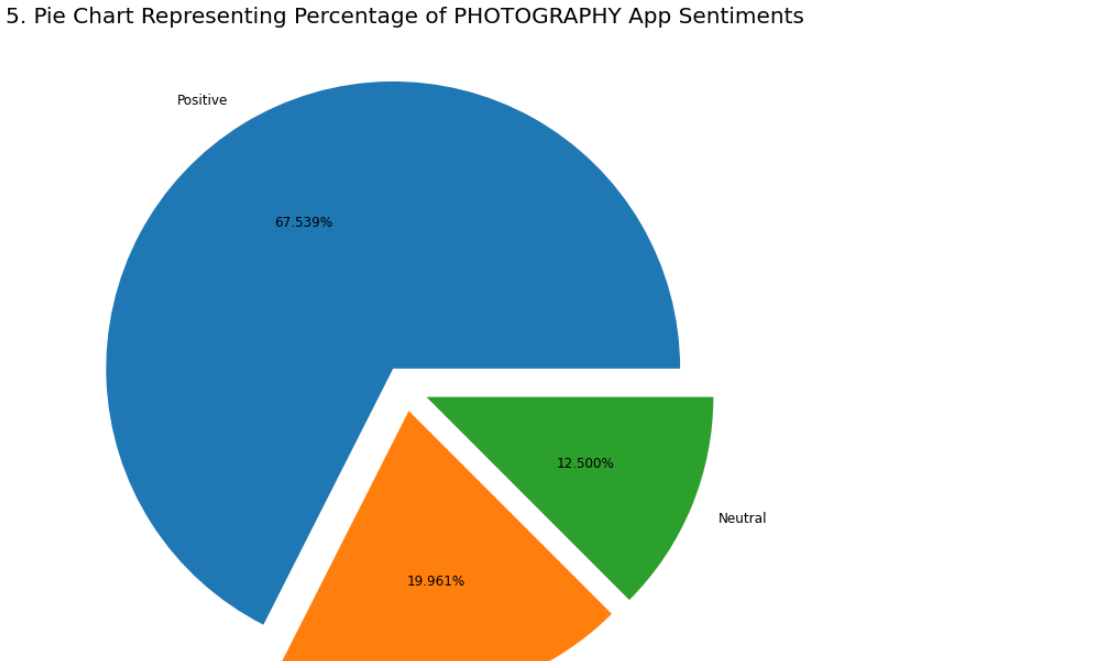
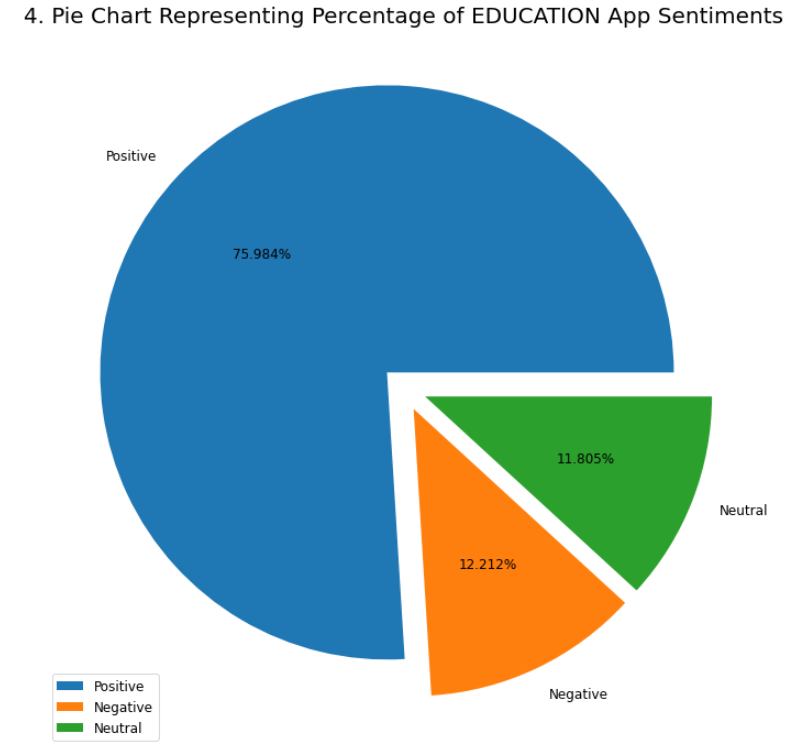
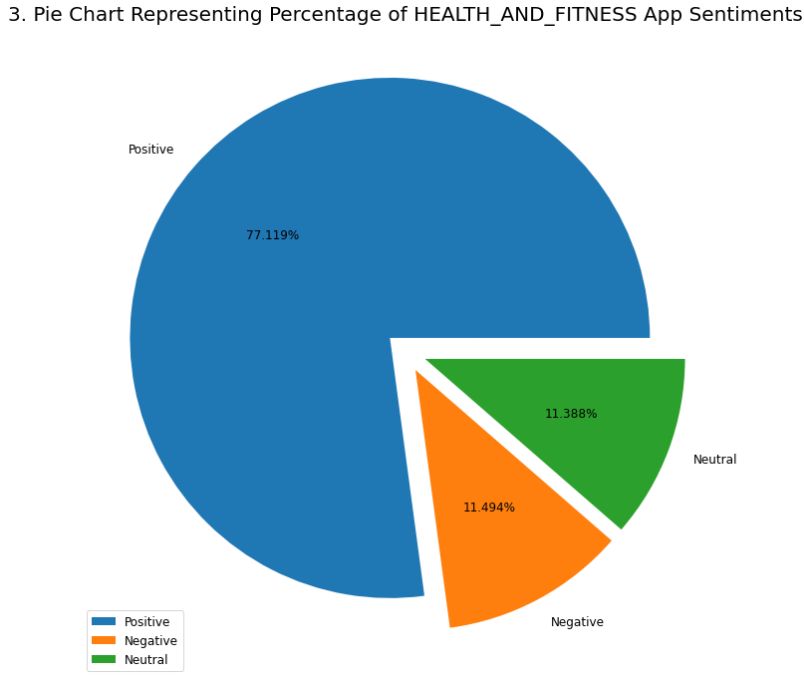
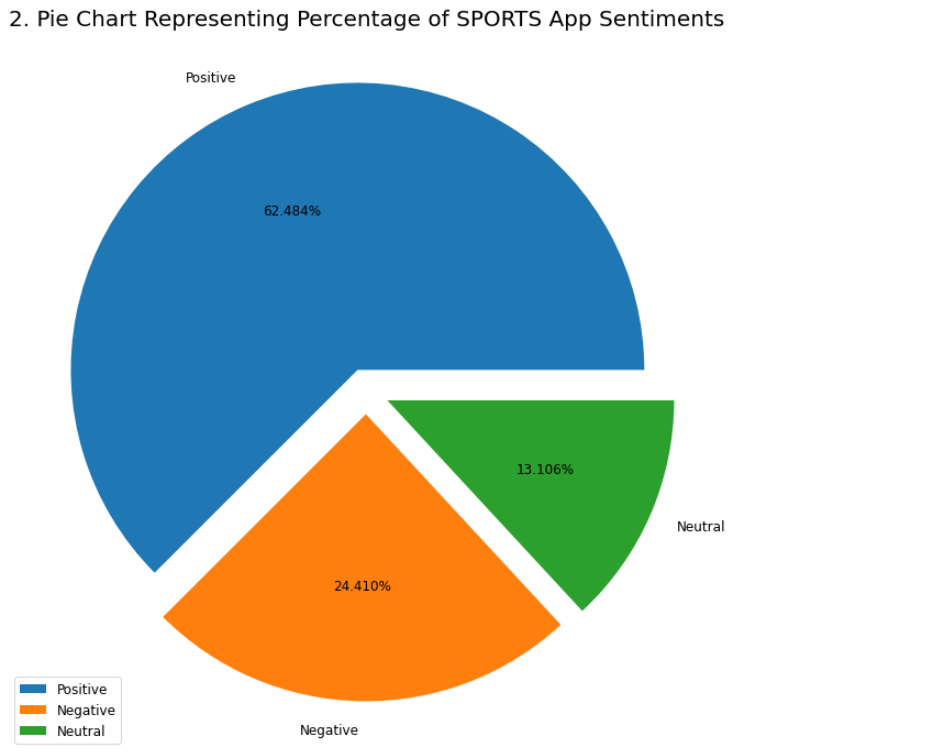
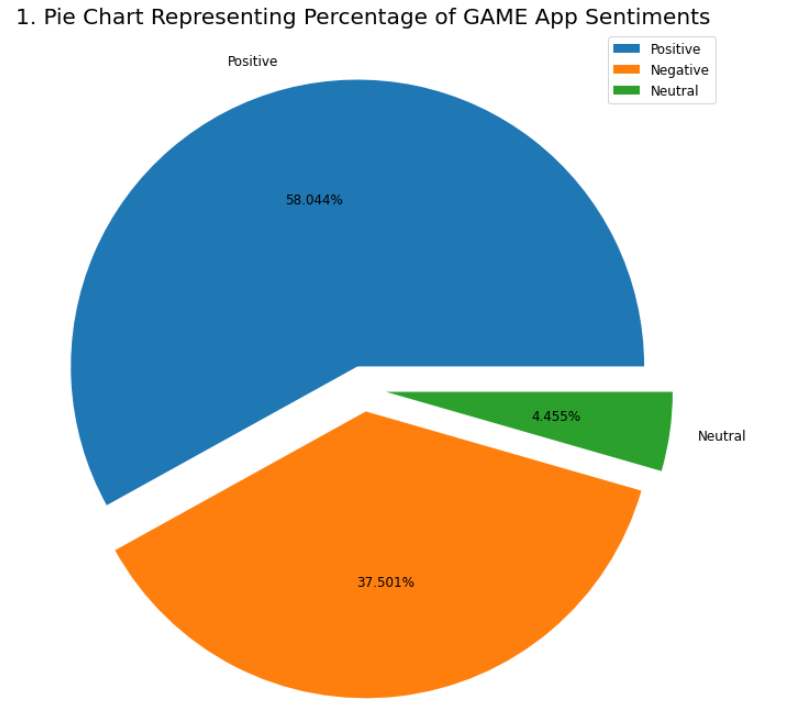
**Datasets:**

****

1. Positive are 63.625%
2. Negative are 24.976%
3. Neutral are 11.399%

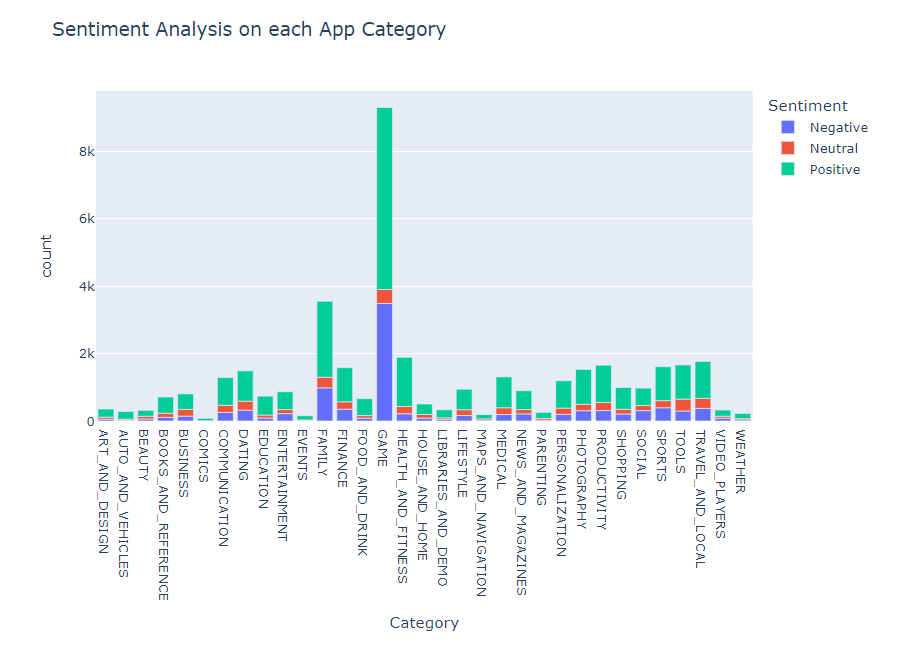
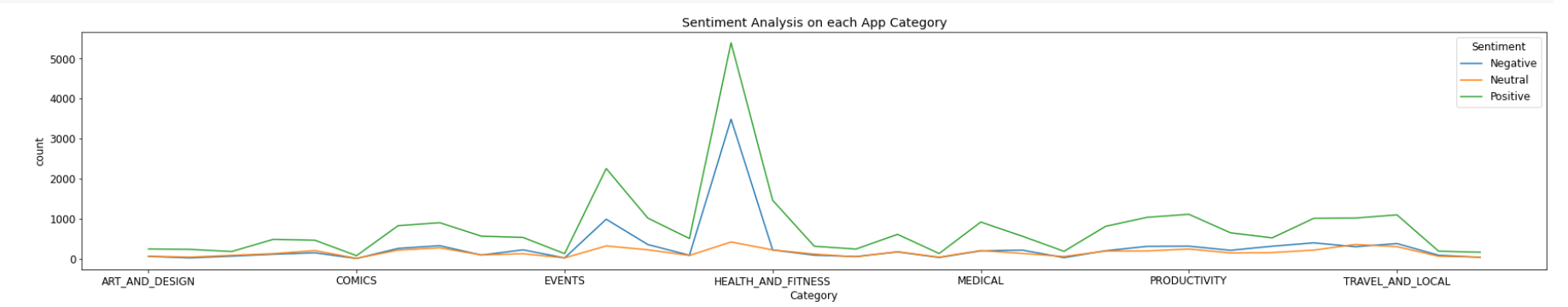
* From this pie Chart we can conclude that most of the sentiment reviews given by the user are positive with 63.625%. But also there is a negative sentiment percentage of 24.976% which is higher than the one with the neutral sentiments with 11.399%. This means app developers need to convert more negative sentiments to positive sentiments with their Hard work.

**17.Sentiment analysis on different top Reviewed Category on PlayStore**

****

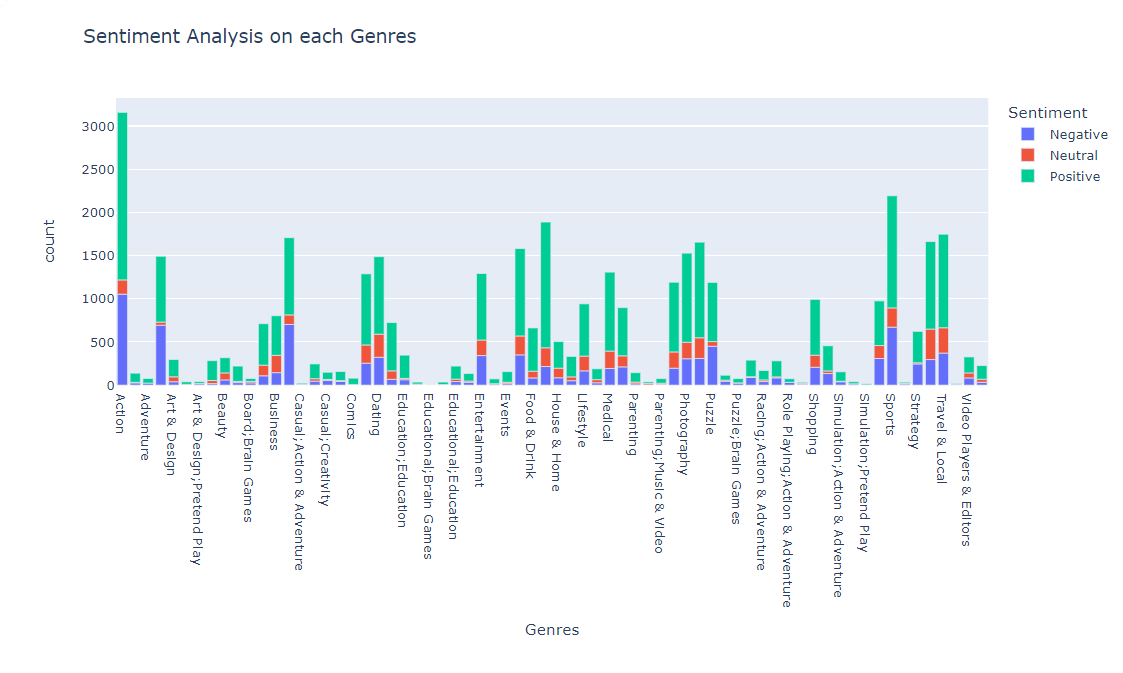
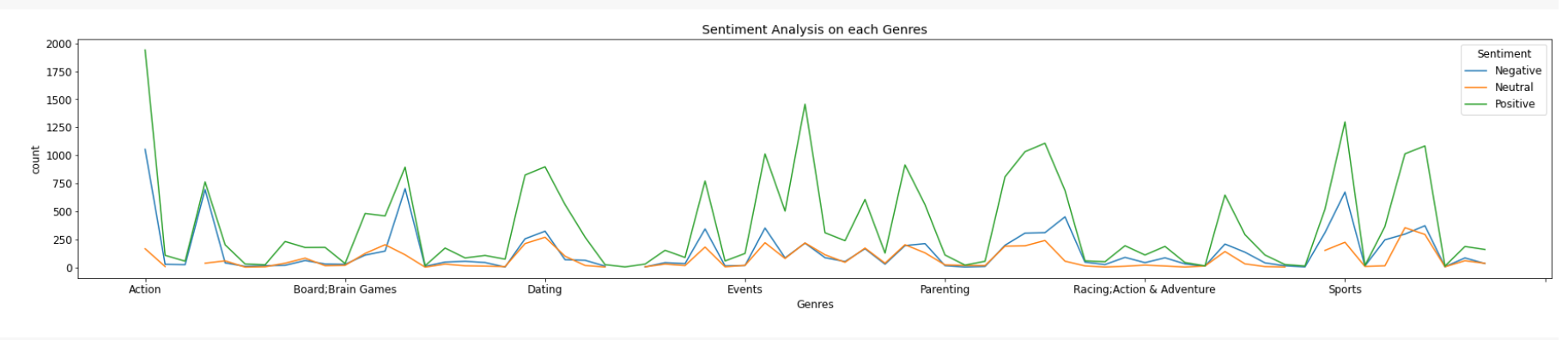
* From the above figure we conclude that 'HEALTH\_AND\_FITNESS' Category has got highest positive percentage of 77.119% and negative sentiment percentage of 11.494% and neutral percentage of 11.388%
* Also we found that top Category GAME has less positive sentiment percentage of 50.04% than it's competitor.
* 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 its counterparts.
* Highest percentages of neutral sentiments has been claimed by the SPORTS category from the list top 5 App category.

**18.Sentiment Analysis on each App Category:**

****

* From Above bar plot we can conclude that GAME Category has got the most positive sentiment in terms of count but if we take the percentage of positive vs negative sentiment we will find that HEALTH\_AND\_FITNESS has got the highest positive sentiment percentage.
* Again if we count on the basis of the count of positive sentiment we can see that Category comic has got least number of positive sentiment and no negative sentiment as such.

**19.Sentiment Analysis on the basis of Genres**

****

* From Above bar plot we can conclude that the Action Genre has got the most number of positive and negative sentiments.
* We can see that **Simulation;Pretend Play** has got least number of positive sentiments
* Also we can notice that **'Education;Brain and Games**' has no Sentiment at all.

**20.Conclusion:**

So here we come at the end of our project which is play store App Review Analysis.What we have done just take a short recap. First we have done the removal of null value from rows and columns and the same goes with the removal of duplicates from the datasets. Then we did the formatting for each of the required columns in each dataset.

After analyzing the data we conclude that App with the category Family and the genre tools are in large numbers. Also we can conclude that the number App Rating is directly proportional with the recent update. From this we can see that with all the major updates apps will get more ratings.

We can also conclude that most of the apps which are used by the users have a content rating of ‘Everyone’.

In percentage of Free and Paid App Available in the PlayStore we can assume that most apps being used by the users are Free. This shows very few users purchase Apps on playstore.

In rating vs count of App Type we conclude that rating is not get affected even if the app is paid or not but if we go on for finding the average rating we will find that free app will have less average rating compared to paid because of significantly high counts of free Apps as compared to Paid App available in App Store.

After moving forward when we performed analysis on sentiment subjectivity we found that most of the opinion on sentiment subjectivity lies high in the range 0.4 to 0.7.

When we analyzed sentiment polarity for paid and free Apps we noticed that sentiment polarity for free apps is way less than paid Apps.

In pie presenting the percentages of review sentiment we found that most of the sentiment are positive and neutral review is the lowest. Also in case finding the percentage of sentiments for top 5 Apps we found among top 5 App Category Health and Fitness has received the highest positive sentiments while Game app category has received the highest negative sentiments and Sports App Category has received the highest neutral sentiments.

**Important Points to be noted:**

* After merging both of the dataset we can find 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%.

**20.References**

* GeekforGeeks
* Kaggle
* Analytics Vidya