

# Capstone Project - 1

## PLAY STORE APP REVIEW ANALYSIS

### Team Members

Saurabh Funde

Kalyani Motkari

Sameer Ambekar



Google Play is an online store where people go to find their favorite apps, games, movies, TV shows, books, and more. It provides more than 2 million apps and games to billions of people around the world, generating over \$120 billion in earnings for developers to date.

# EDA Approach

- Problem Statement
- Loading required packages for EDA
- Loading datasets
- Dataset's exploration
- Dataset's wrangling
- Dataset's merging
- Questions and answers for data analysis
- Data Visualization's

# Problem Statement

Play store is a marketplace (App) for downloading Android applications for smartphones. Smartphones sells increasing across world. It's a very lucrative market for App developers as users are looking for comfort and their needs increasing like entertainment, games etc.

It's important to find out what type of apps people downloading before developing an App for users and list that APP on Play Store for downloading.

How does Size, Price and Type of app important for apps success and engagement?

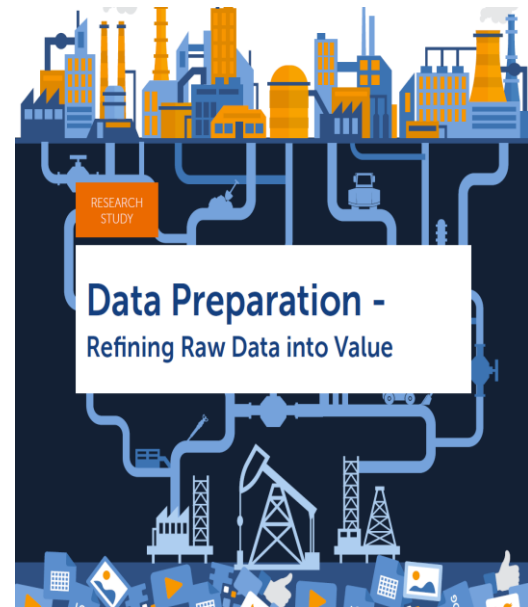
# Examining the Dataset

- 1. Play Store App Dataset:** This dataset contains 10841 rows and 13 columns.
- 2. User Reviews Dataset:** This dataset contains user reviews and sentiment of apps.



# Dataset Preparation

- **Loading the data sets:** Two datasets, First Play store app dataset and User Reviews dataset.
- **Import Libraries:** NumPy, Pandas, Seaborn and Matplotlib
- **Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data.
- **Data Imputation:** Filling the missing categorical values with mode and numerical values with median. Conversion of price, installs, reviews into numerical values.
- **Exploratory Data Analysis:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.



# Attributes in Google Play store Data

**App-** Application name

**Category-** Category the app belongs to

**Rating-** Overall user rating of the app

**Reviews-** Number of user reviews for the app

**Size-** Size of the app

**Installs-** Number of user downloads for the app

**Type-** Paid or Free

**Price-** Price of the app

**Content Rating-** Suitable age group for given app

**Genres-** An App can belong to multiple genres

**Last Updated-** Date of Last modification by app development team

**Current Ver-** It shows Current version of app

**Android Ver-** It tells the app can run on which Android version compatible



# Attributes in User reviews

1. **App-** Application name
2. **Translated Review-** User review
3. **Sentiment-** Positive/Negative/Neutral
4. **Sentiment Polarity-** Sentiment polarity score
5. **Sentiment Subjectivity-** Sentiment subjectivity score





# EDA

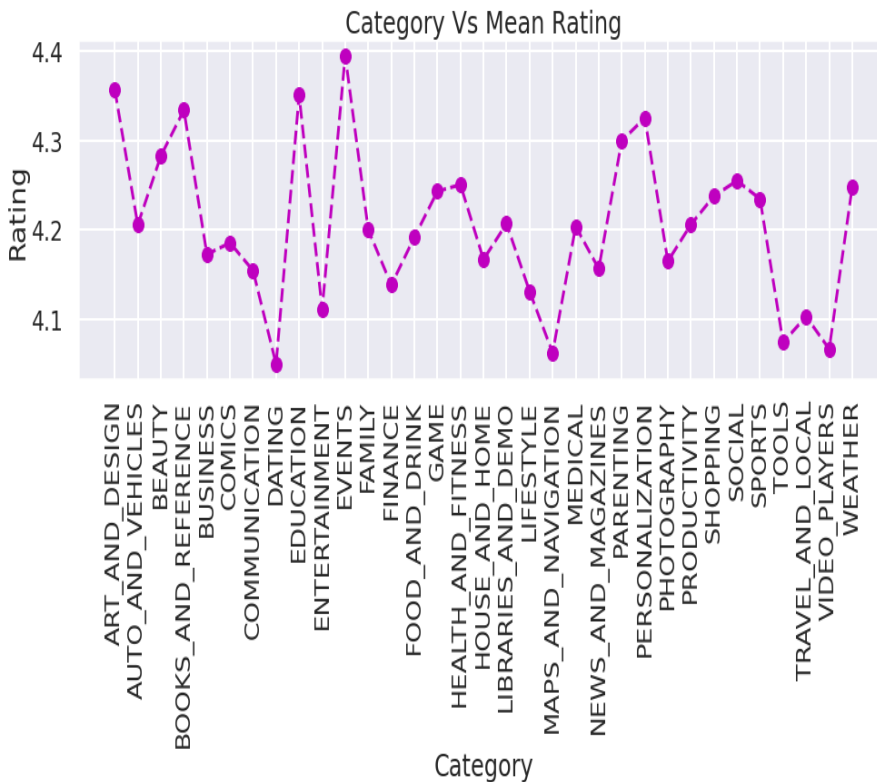
## Category wise VS Mean Rating

### 1. Highest Ratings apps:

Events, Art and Design, Education categories has got 4 and above ratings.

### 2. Lowest Ratings apps:

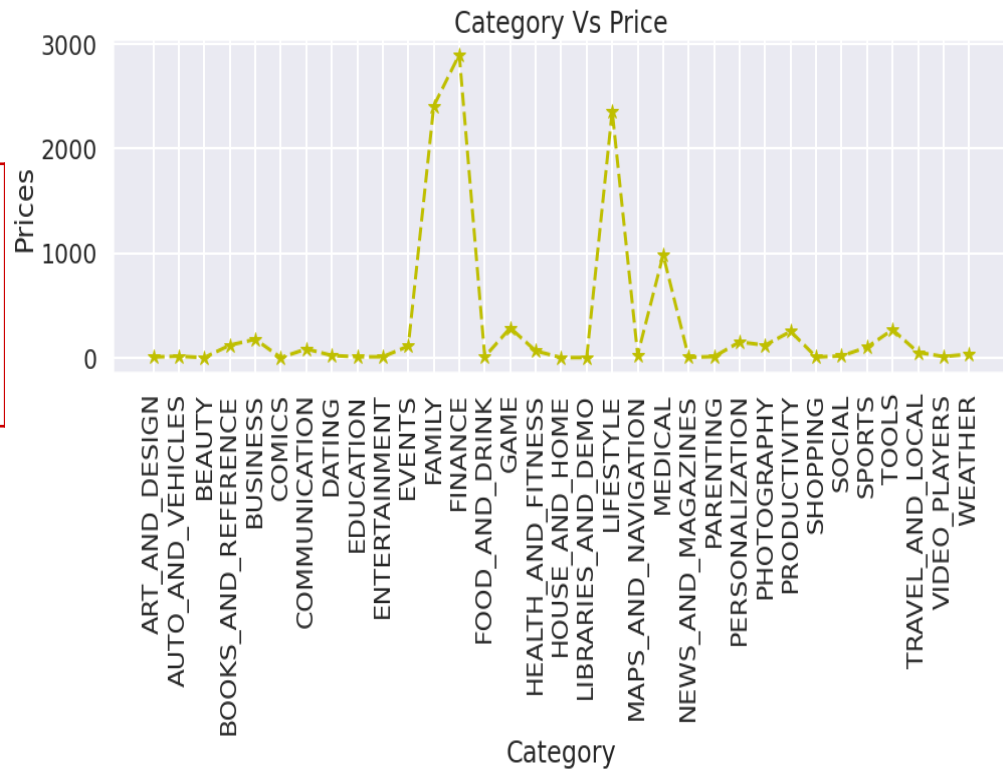
Dating, Maps and navigation, Video players category has got lowest ratings i.e. below 4.



# EDA(Cont.)

## Category wise Vs Pricing

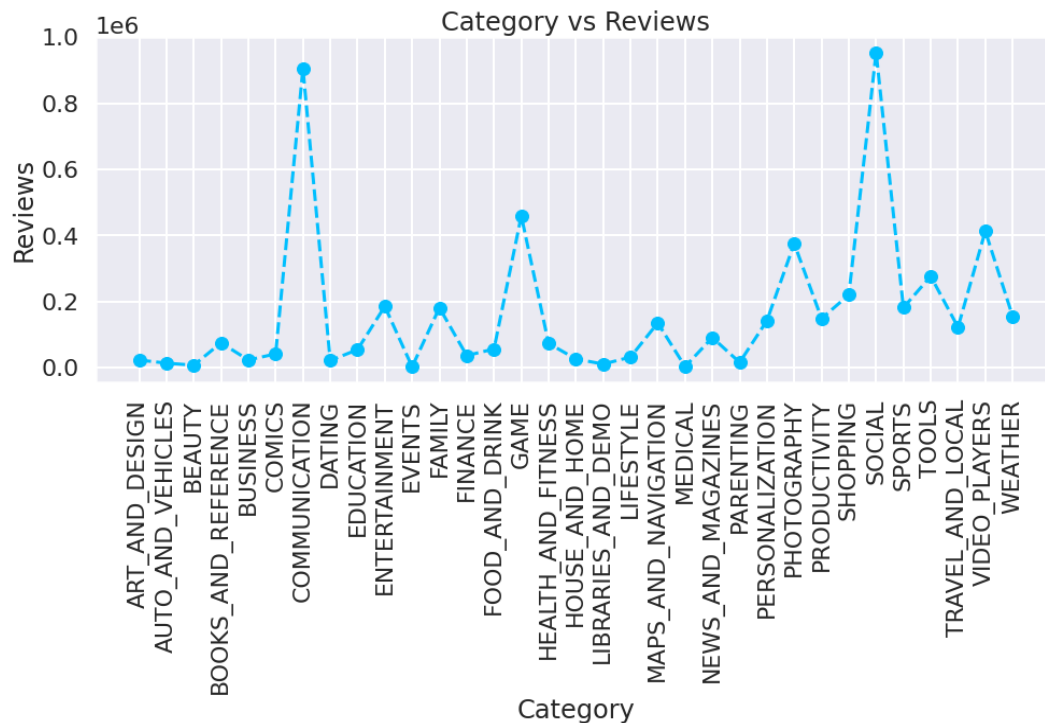
As we can see that **Finance, Lifestyle and Medical** Category has got the highest pricing.



# EDA(Cont.)

## Category wise Reviews

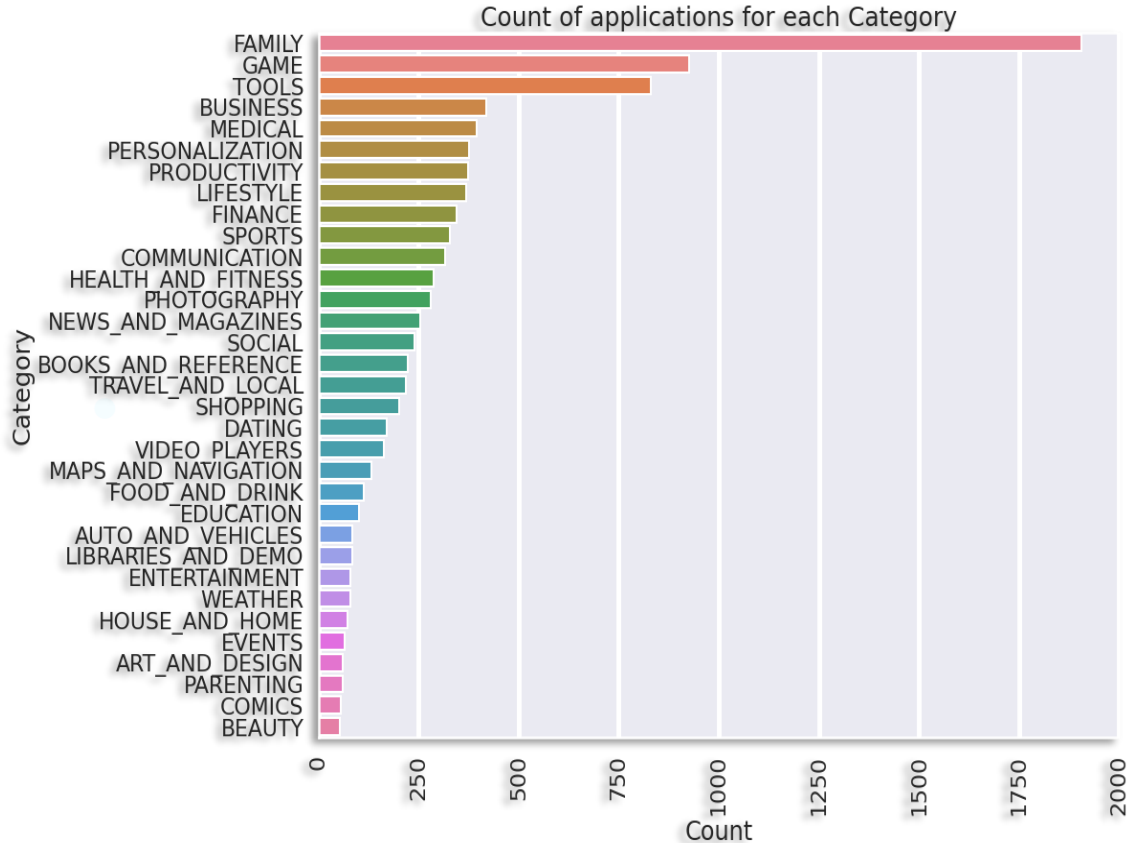
The **Social, Communication and Game** category got the highest reviews.



# EDA(Cont.)

## Count of Application in each category

**Family and Game** apps have the highest market prevalence. Surprisingly **Tools, Business and Medical** apps are also at the Top Count of applications.



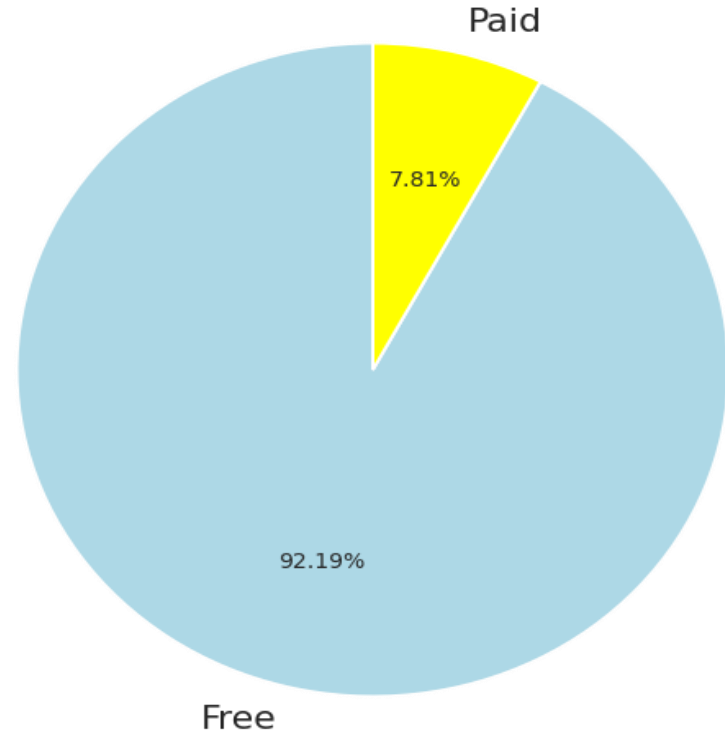
# EDA(Cont.)

## Free Vs Paid Application's Type

We Observed that **92.19% of Apps are free** and only **7.81% of Apps are paid** in Play store.



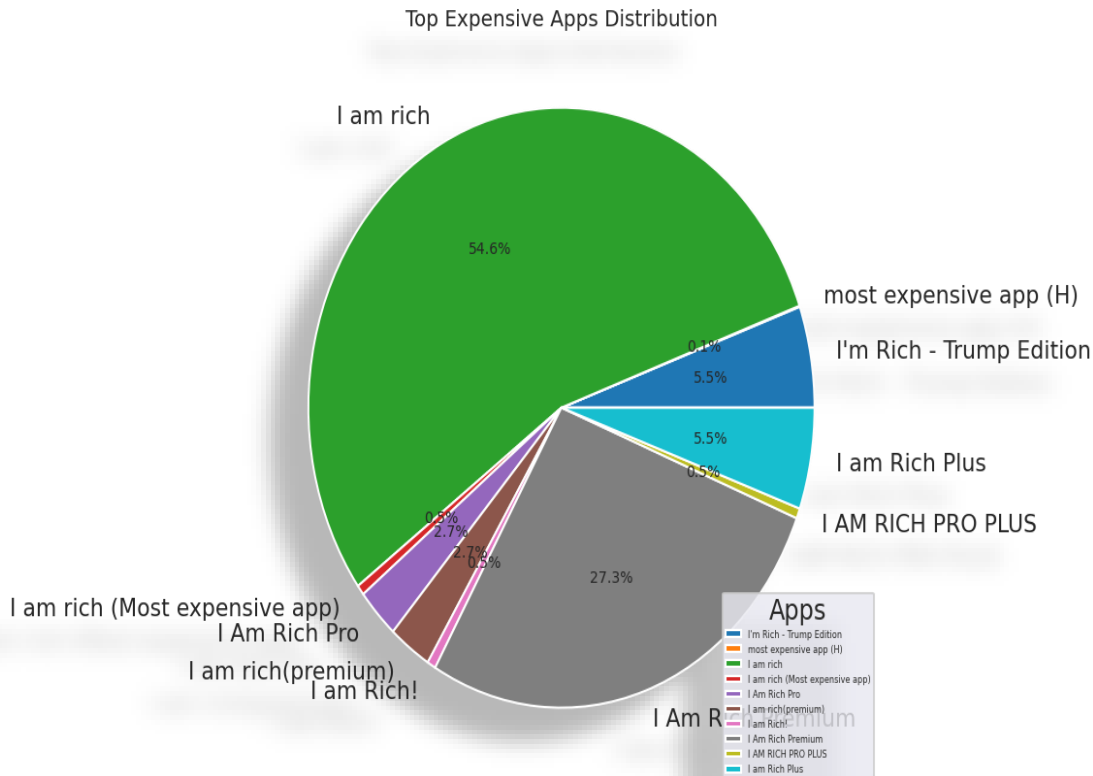
Percent of Free Vs Paid Apps in store



# EDA(Cont.)

## Top Expensive Apps

**I am rich** is the most expensive Application in the google play store.

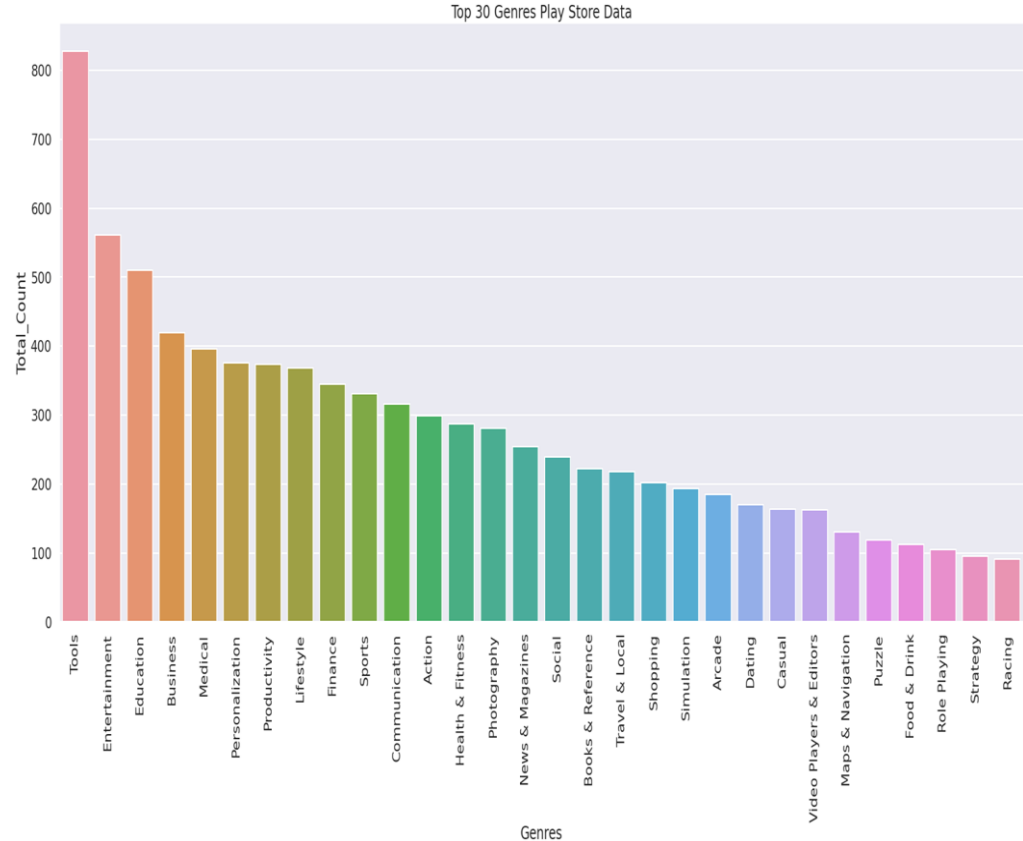


# EDA(Cont.)

## Top 30 Genres

The highest number of apps are from **Tools** genre.

**Tools, Entertainment, Education, Business** and **Medical** are top Genres for App.



# EDA(Cont.)

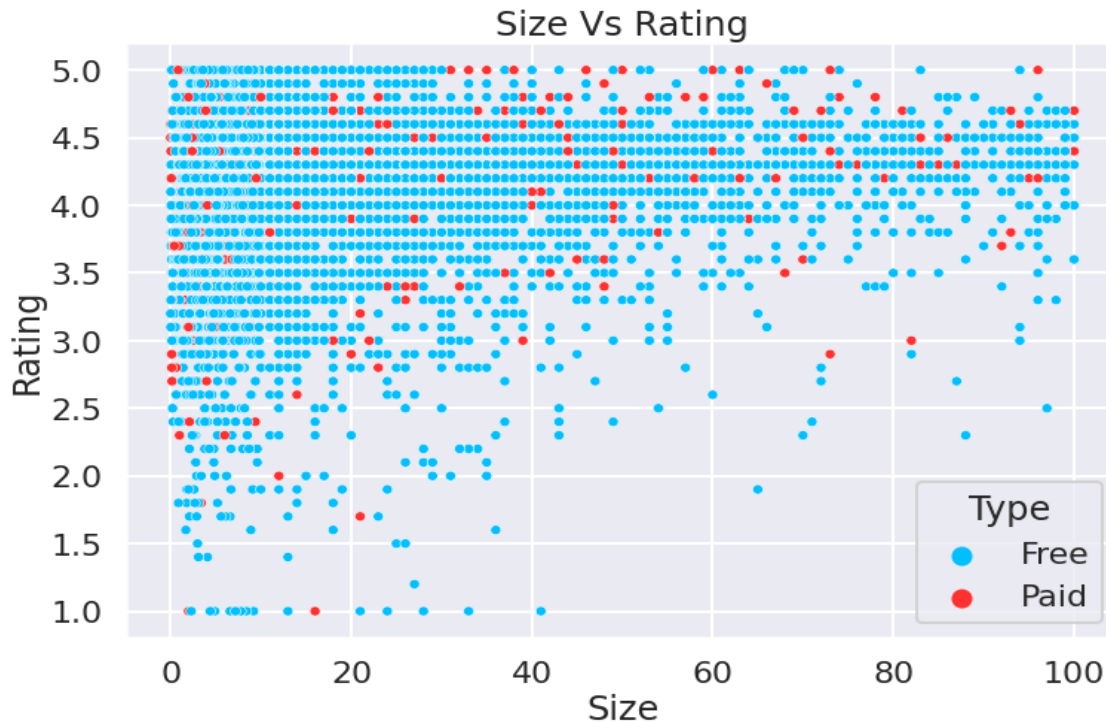
## Distribution of Size Vs Rating

Large number of apps are in small size.

**92.19%** apps are Free on Play Store.

**8783** Apps are below of 50MB in size including Free and Paid.

**7749** Apps have rating more than 4.0 including Free and Paid.





## EDA(Cont.)

### Relationship viz Price and Rating

**8905** Apps are **Free** in type.

The **mean** rating of **Free** apps is **4.18**.

**754** Apps are **Paid** in type.

The **mean** rating of **Paid** apps is **4.26**

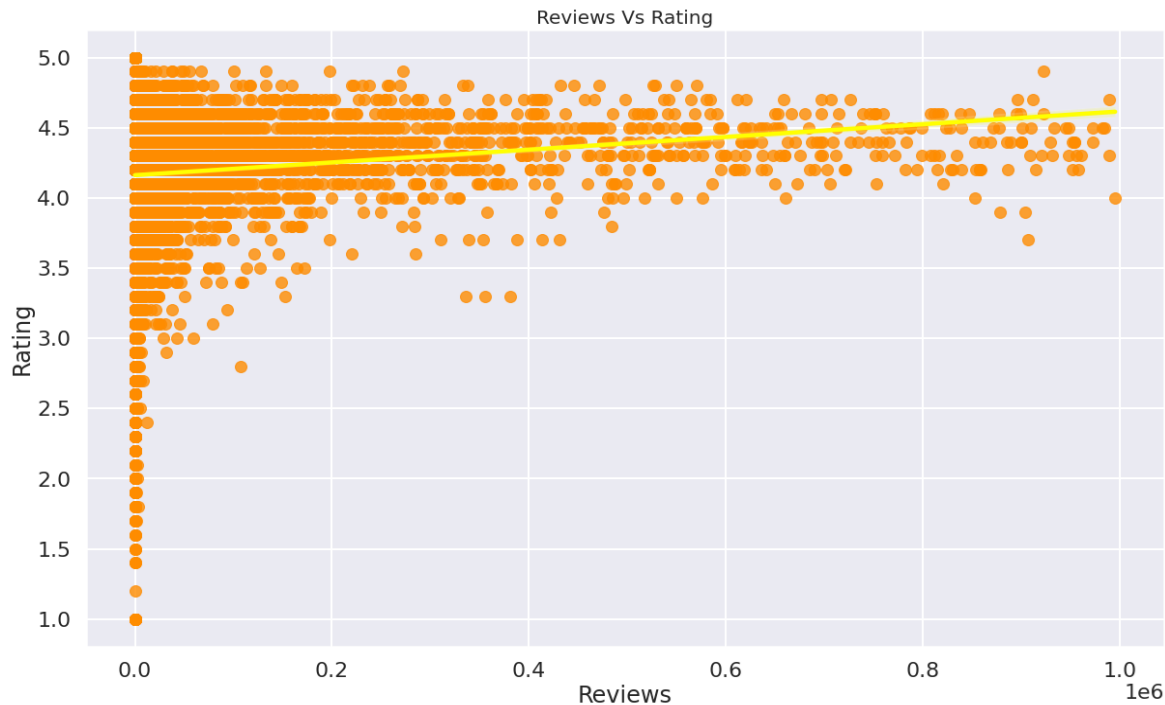


# EDA(Cont.)

## Relationship viz Reviews and Rating

9309 Apps have less than and equal to 1000000 reviews count.

Rating's of an increasing slightly as more people giving Reviews.

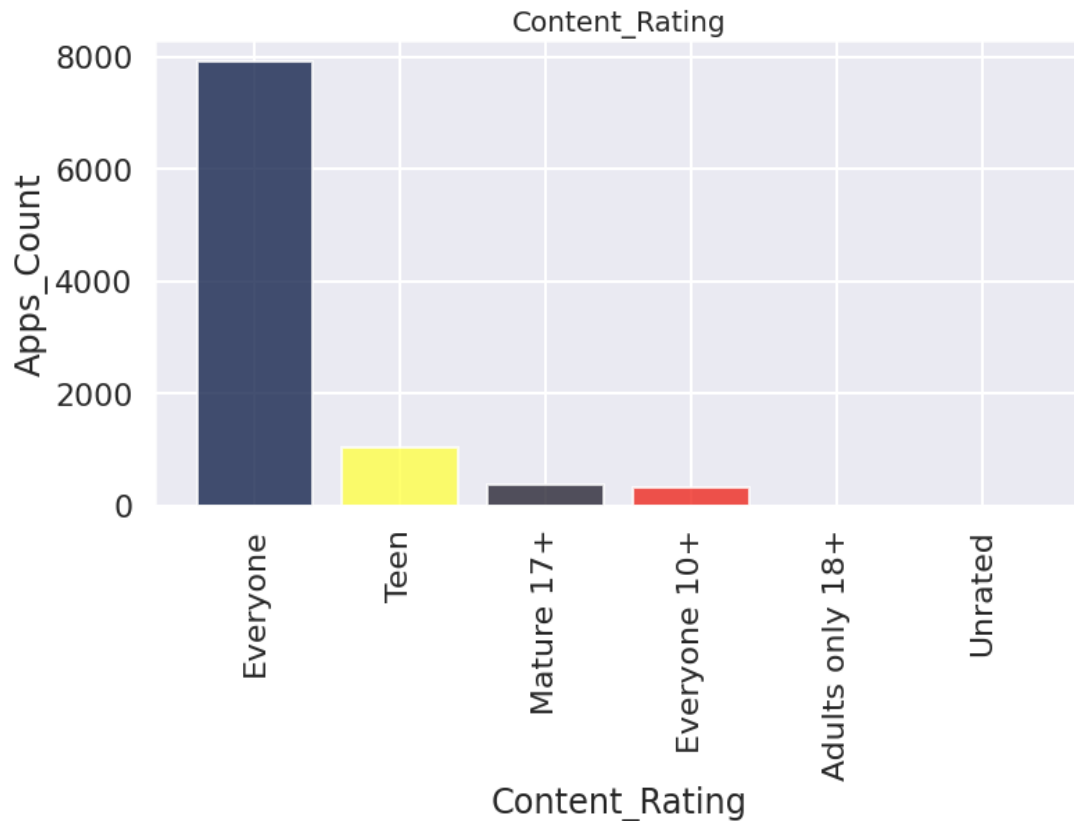


Viz for Apps with less than 1000000 Reviews

# EDA(Cont.)

## Content Rating Vs Count

**81%** Apps are for Everyone content and **10%** for Teens.

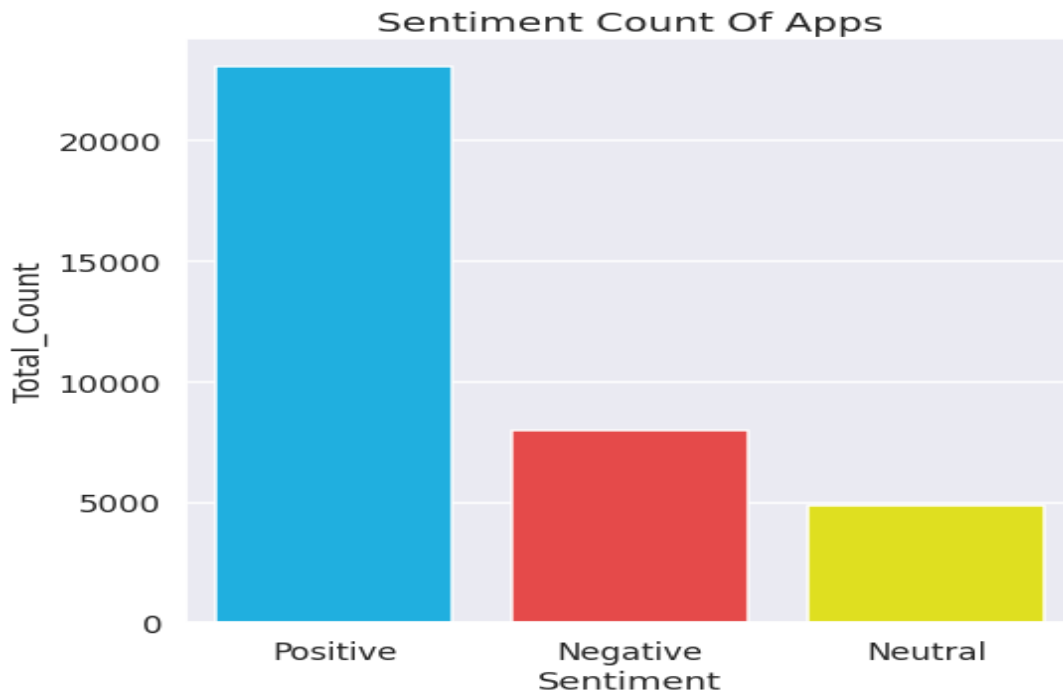


## EDA(Cont.)

### Overall Sentiment Count

The number of **Unique** Apps from Play store and User reviews merged dataset are **816**.

From Sentiment column, **64%** are **Positive**, **22%** are **Negative** and **13%** are **Neutral** values.

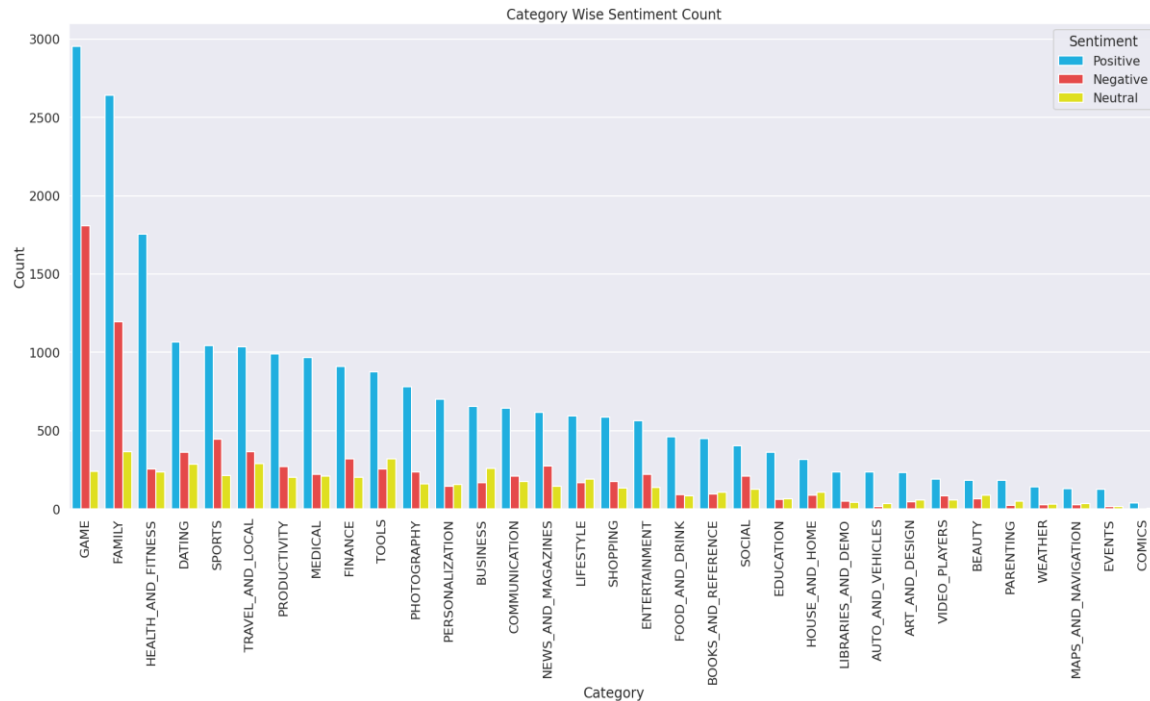


## Category wise Sentiment Count

The total number of Categories are 33.

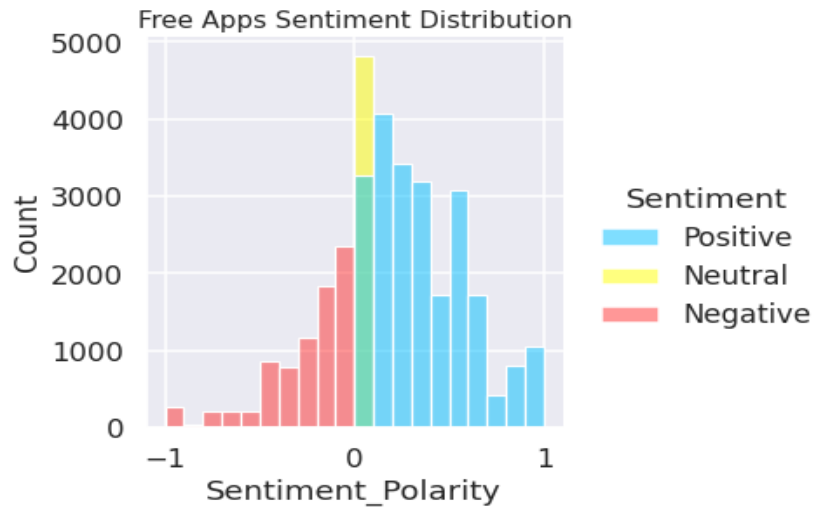
**Game** category has highest **Positive (2955)** and **Negative (1809)** sentiment count.

**Family** category has highest **Neutral (367)** sentiment count.

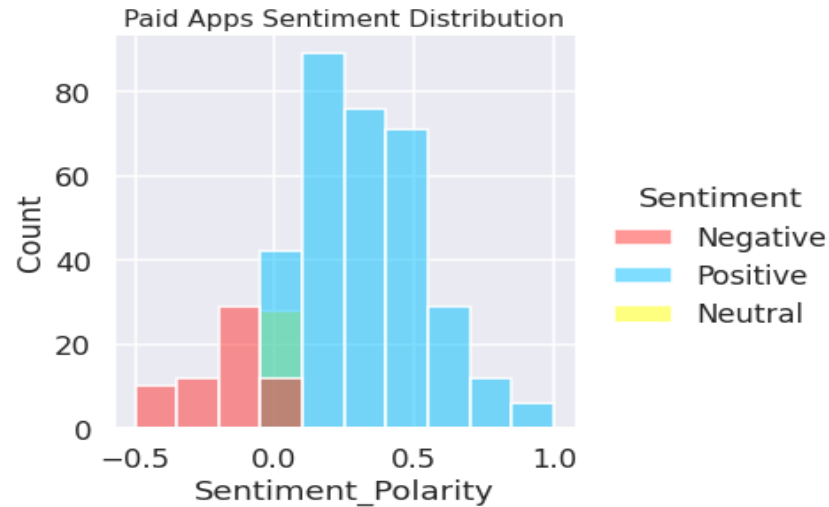


# EDA(Cont.)

## Sentiment Polarity Distribution



**Positive: 22748**  
**Negative: 07942**  
**Neutral: 04823**

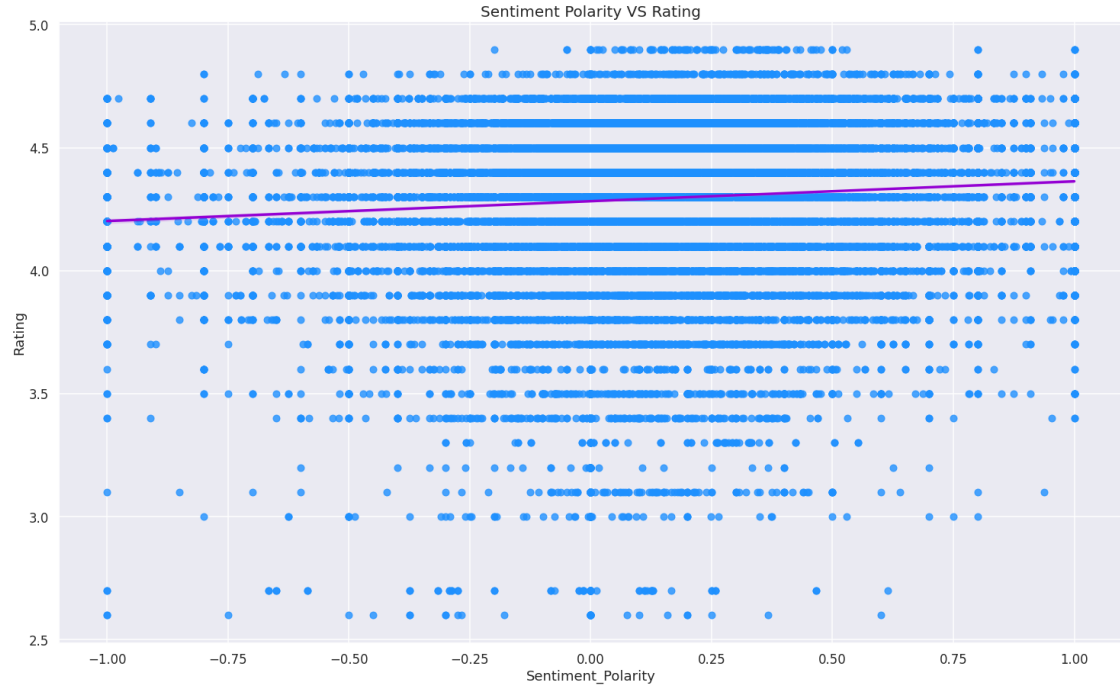


**Positive: 325**  
**Negative: 63**  
**Neutral: 28**

# EDA(Cont.)

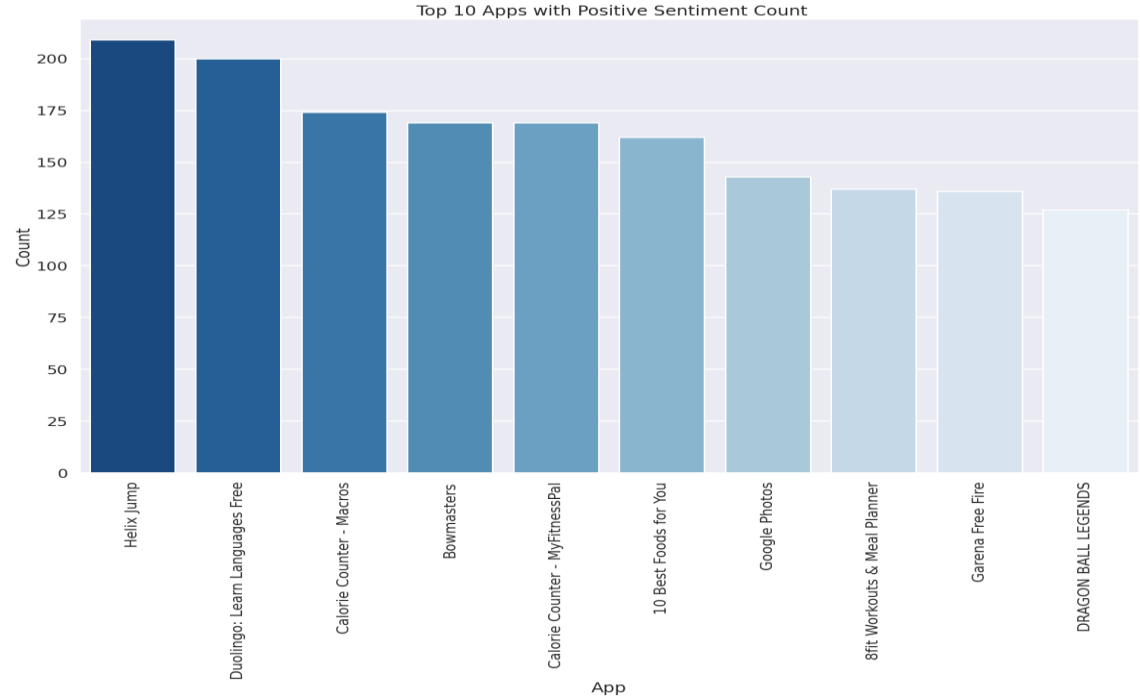
## Sentiment Polarity Vs Rating's

Sentiment Polarity vs Rating plot shows us, There is no relation between two feature.



## Top 10 Apps Positive Sentiment Count

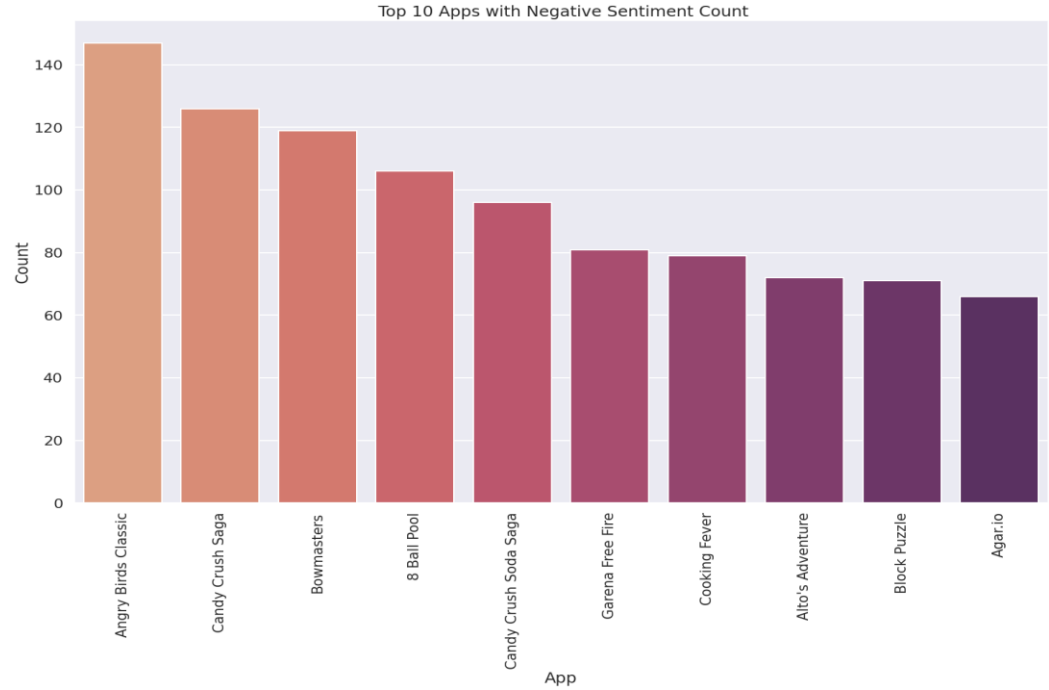
**Helix Jump** is a App from merged dataset has highest **209 Positive** sentiment count.





## Top 10 Apps Negative Sentiment Count

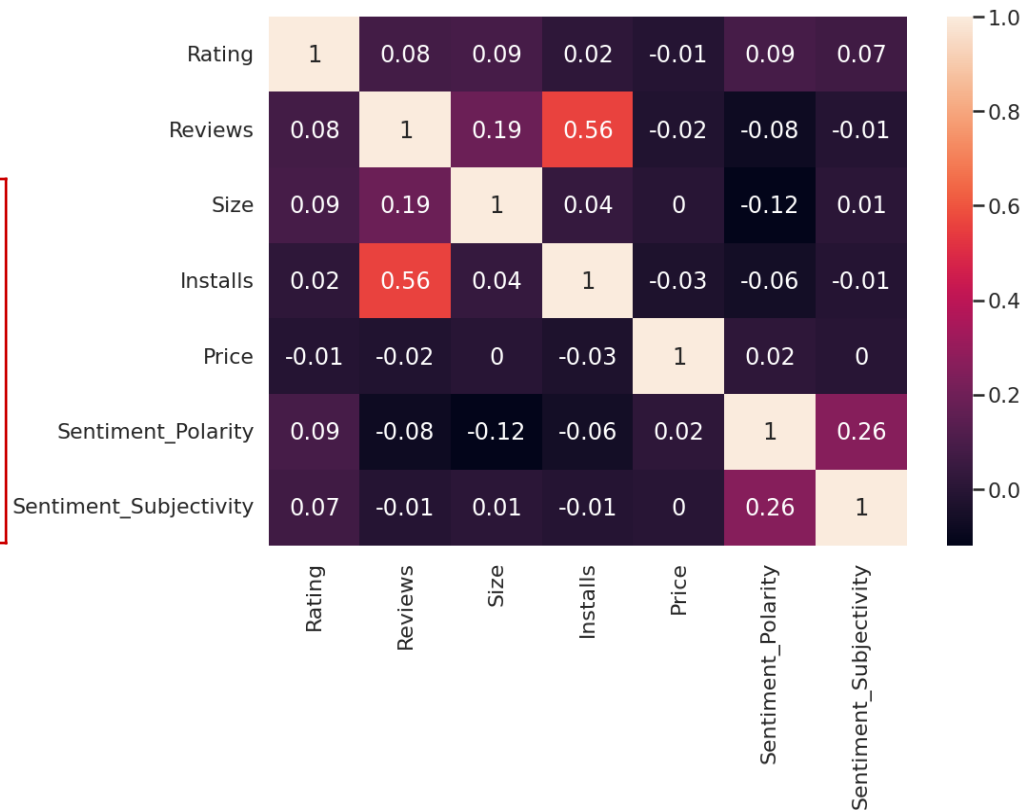
**Angry Bird Classic** is a app from merged dataset has highest **147 Negative** sentiment count.



# EDA(Cont.)

## Co-Relation in Merged Data frame

In this correlation matrix, There is not a significant relationship between Rating, Reviews, Size and Installs with respect to the Sentiment polarity and Sentiment subjectivity.



# Conclusion's:

- **92.19%** apps are **Free** and **7.81%** apps are paid in type.
- **81.10%** apps have **Everyone** content rating.
- **Events** category has a **highest mean rating of 4.39** and Dating category has lowest **4.05** rating.
- **Family, Game and Tools** are **top three** categories having **1906, 926 and 829** app count.
- **Tools, Entertainment, Education, Business and Medical** are top Genres.
- **8783 Apps** are having size less than or equal to **50 MB**.
- **7749 Apps** has rating **more than 4.0** including both type of app.
- **Overall sentiment count** of merged dataset in which **Positive sentiment count is 64%, Negative 22% and Neutral 13%.**
- **Price, Rating, Size has no or very less correlation** with **Sentiment Polarity.**
- It's good to develop a **Free type** app and having a content rating for **Everyone.**
- Most of the apps are fall in **0 to 50 MB Size group.** It's good to develop of app this size bracket.