

# Capstone Project - 1

## PLAY STORE APP REVIEW ANALYSIS

### Team Members

Ajit Varpe

Niharika Soni

Harisha Chennozswala

Satya Prakash

# Let's Review Play Store App:

1. Google Play Store App
2. Data pipeline
3. Data sets
4. Variables in Google Play store and User Reviews.
5. Exploratory Data Analysis
  - a) Category wise Analysis
  - b) Free Apps vs Paid Apps
  - c) Genre wise analysis
  - d) The highest earning app
  - e) User Review Analysis
6. Conclusion





The **Google Play Store** is the official pre-installed app on Android devices which provides access to the Google Play store. It allows users to browse and download music, books, movies and lots of applications from Google Play. It is the largest and most popular Android app store. With around three million different apps available for users to download. To differentiate themselves in this oversaturated market, they need to pinpoint essential factors that play a role in customers' decision-making process.

# DATA PIPELINE



## Data cleaning

- ❑ **Loading the data sets:** We have two different data sets, one is play store app data and the second data set with reviews of users of the apps.
- ❑ **Data cleaning:** Null values, Finding and removing Outliers, Data mismatch.
- ❑ **Data Imputation:** Impute Median, modes of categorical values, filling the missing categorical values with mode, conversion of price, installs, reviews into numerical values, duplicate values, identical data,
- ❑ **Exploratory Data Analysis:** Analysing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

# DATA SETS

1. **Play Store Data.csv:** Contains all the details of the applications on Google Play store. There are 13 features that describe a given app. It have 9659 unique apps
2. **User Review.csv:** Three features: Sentiment (Positive, Negative or Neutral), Sentiment Polarity and Sentiment Subjectivity.



# VARIABLES DESCRIPTION IN GOOGLE PLAY STORE:

Variable	Description
1. App	Application name
2. Category	Category the app belongs to
3. Rating	Overall user rating of the app
4.Reviews	Number of user reviews for the app
5. Size	Size of the app
6. Installs	Number of user downloads for the app
7. Type	Paid or Free
8. Price	Price of the app
9.Genres	An App can belong to multiple genres

# VARIABLE DESCRIPTION IN USER REVIEWS:

Variable	Description
App	Application name
Translated_Review	User review
Sentiment	Positive/Negative/Neutral
Sentiment_Polarity	Sentiment polarity score
Sentiment_Subjectivity	Sentiment subjectivity score



# EXPLORATORY DATA ANALYSIS(EDA)

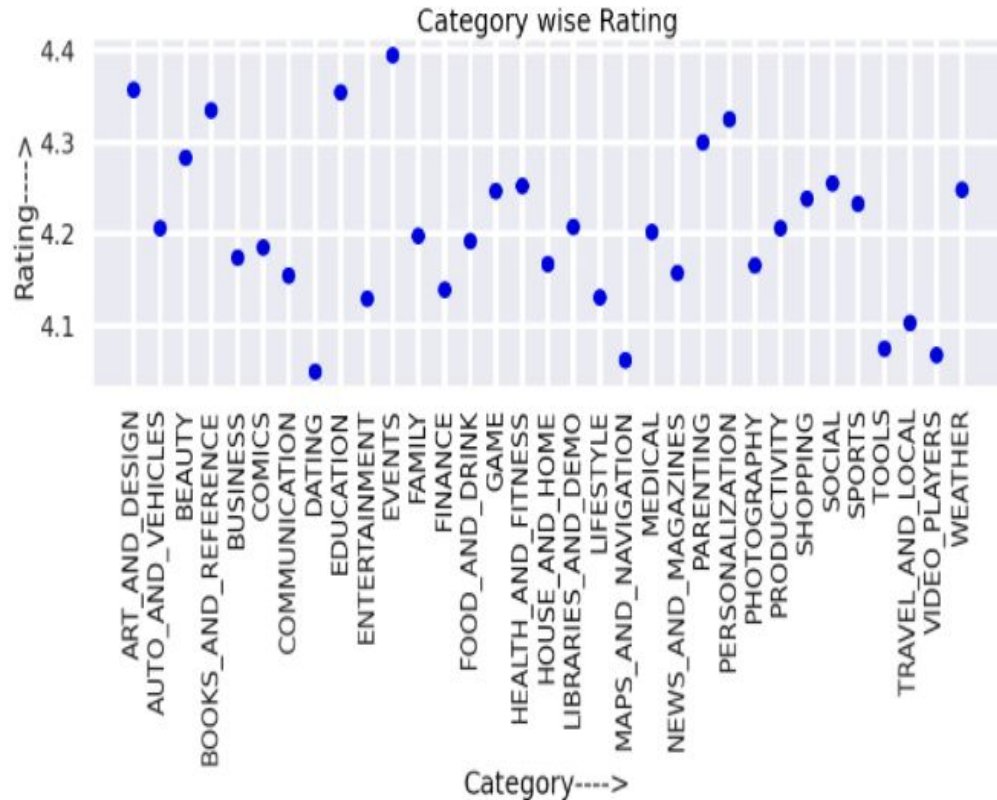
## CATEGORY WISE RATING

### Highest Ratings apps:

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

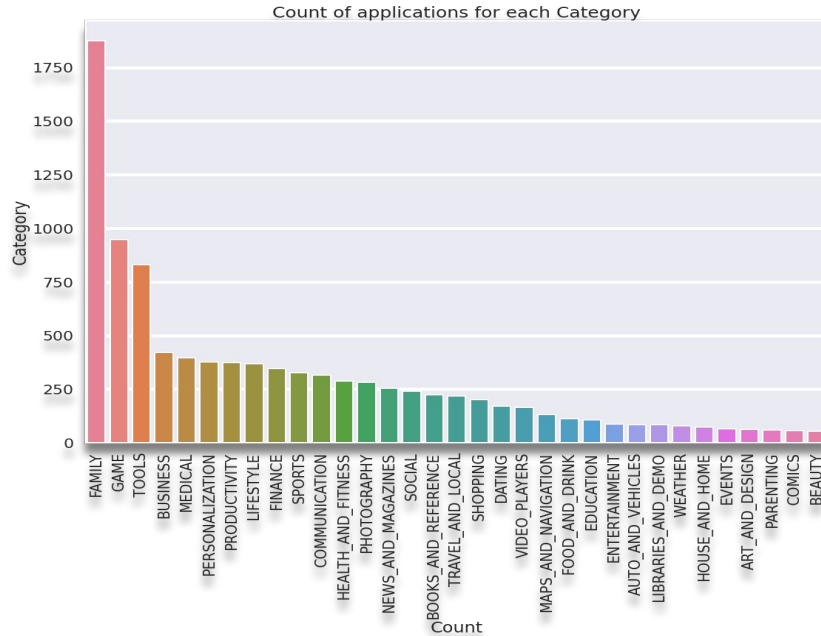
### Lowest Ratings apps:

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



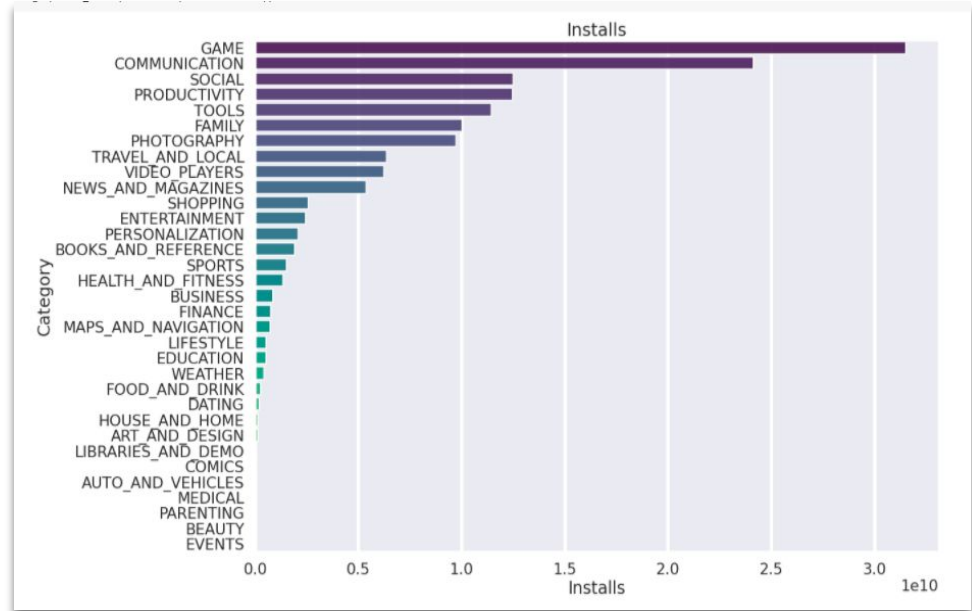


## Count of applications for each category



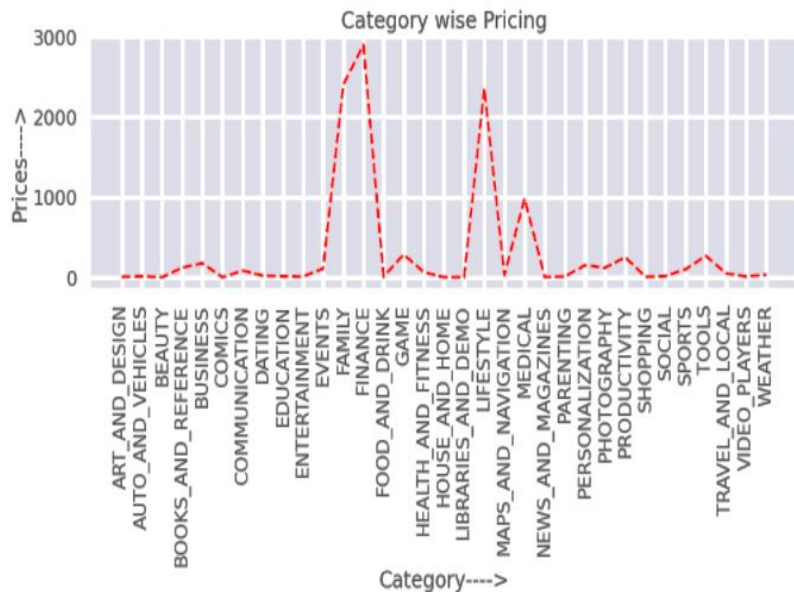
Family and Game apps have the highest market prevalence. Interestingly, Tools, Business and Medical apps are also at the top.

## App's have the most number of installs



The highest installs are Game, Communication, Tools & Family.

## Category wise Pricing



The Finance, Lifestyle and Medical Category has got the highest pricing.

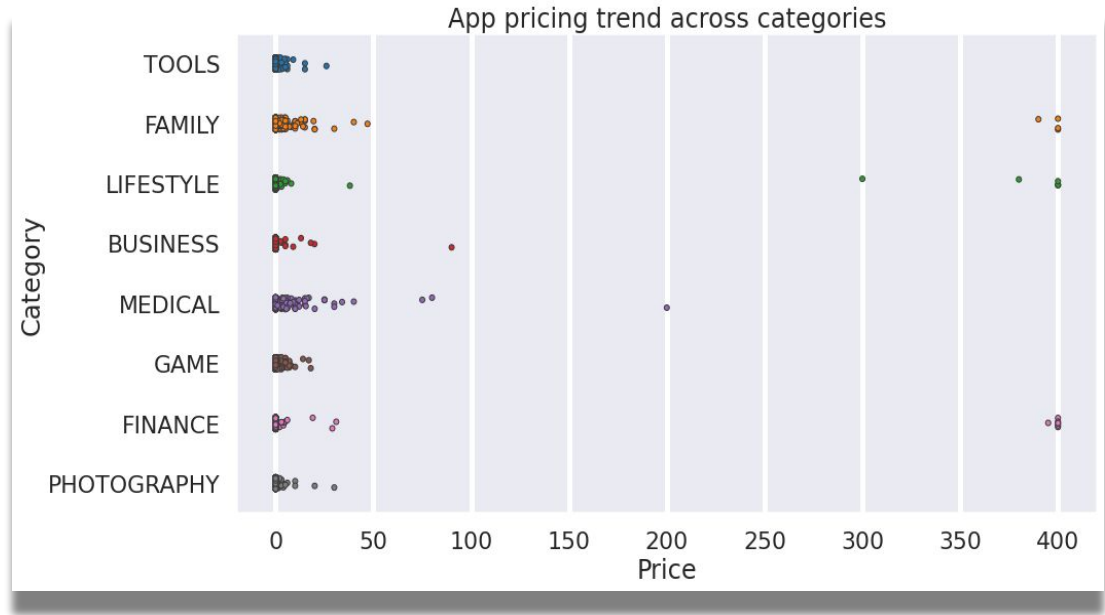
## Category wise Reviews



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

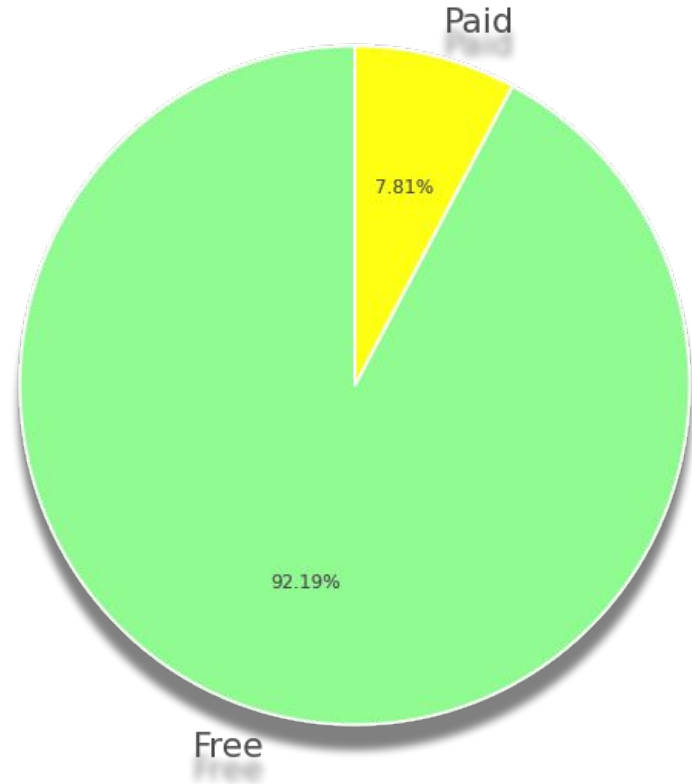
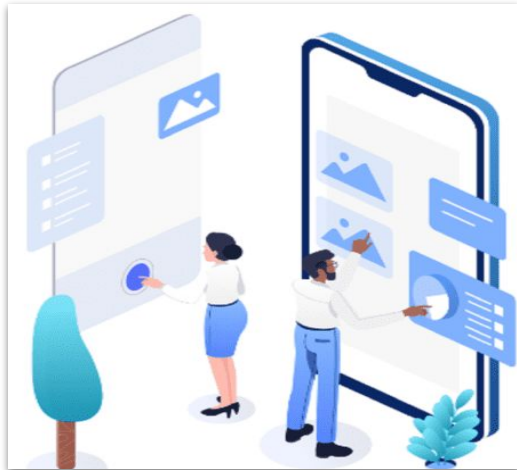
# RELATION BETWEEN APP CATEGORY AND APP PRICE

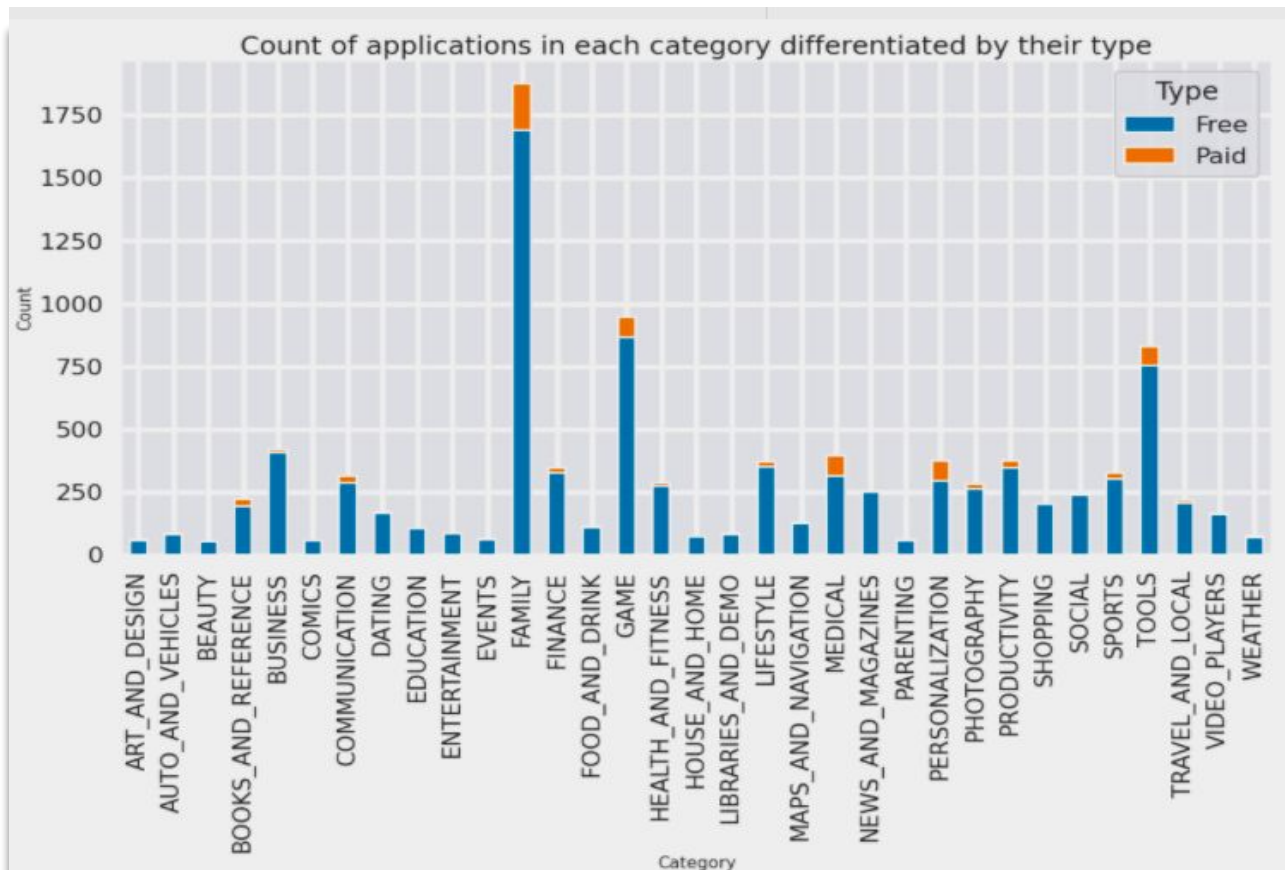
1. **Medical and Family apps are the most expensive.**
  2. **Some medical apps extend even up to \$80.**
- All game apps are reasonably priced below \$20.**



# FREE VS PAID APPS

Here we can see that there are 92.19% of apps are free and only 7.81% of Apps are paid on Playstore.

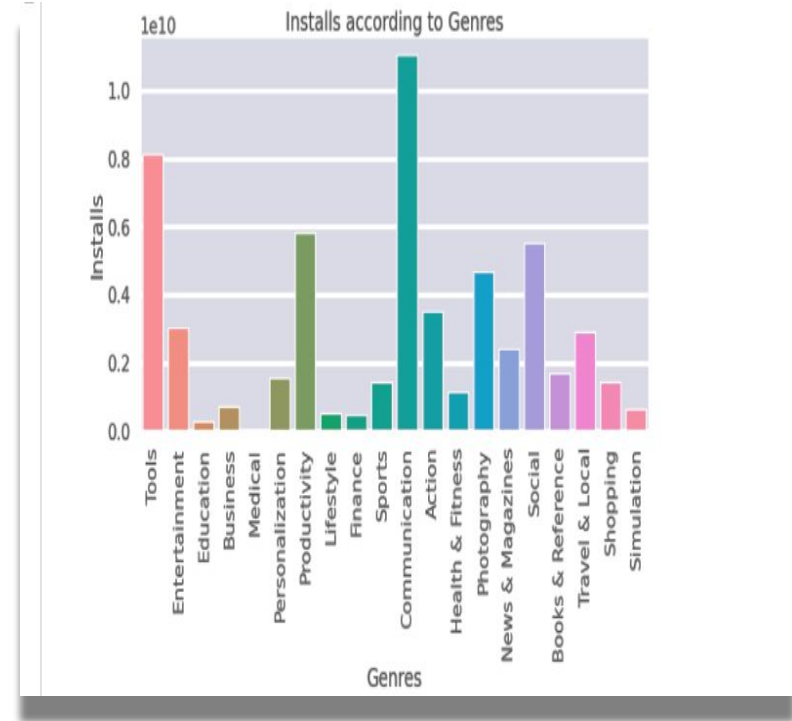
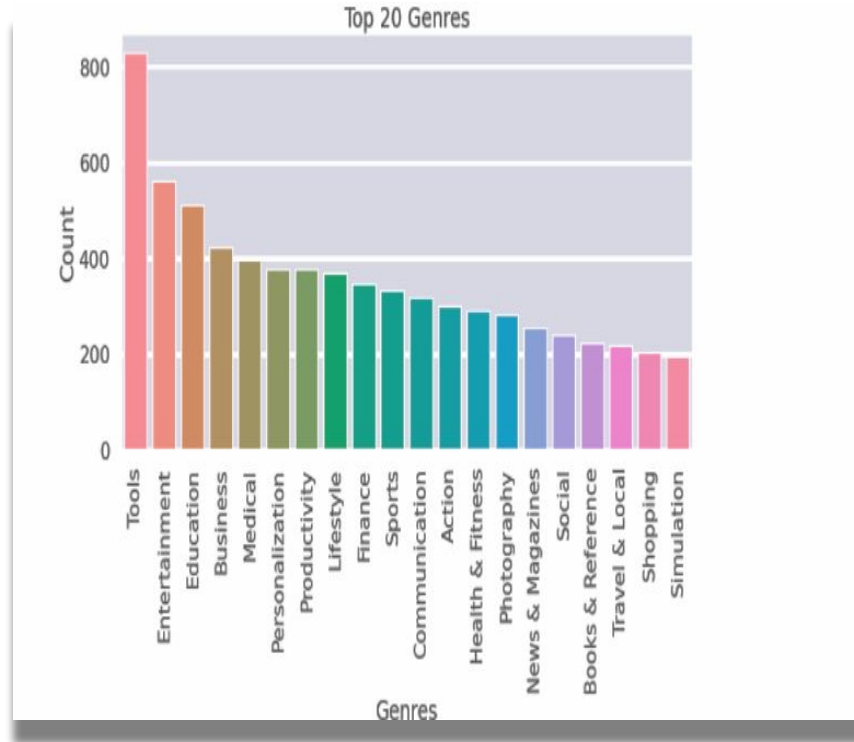




In our dataset, the majority of apps in **Family, Game and Tools**, as well as **Business** categories were free to install.

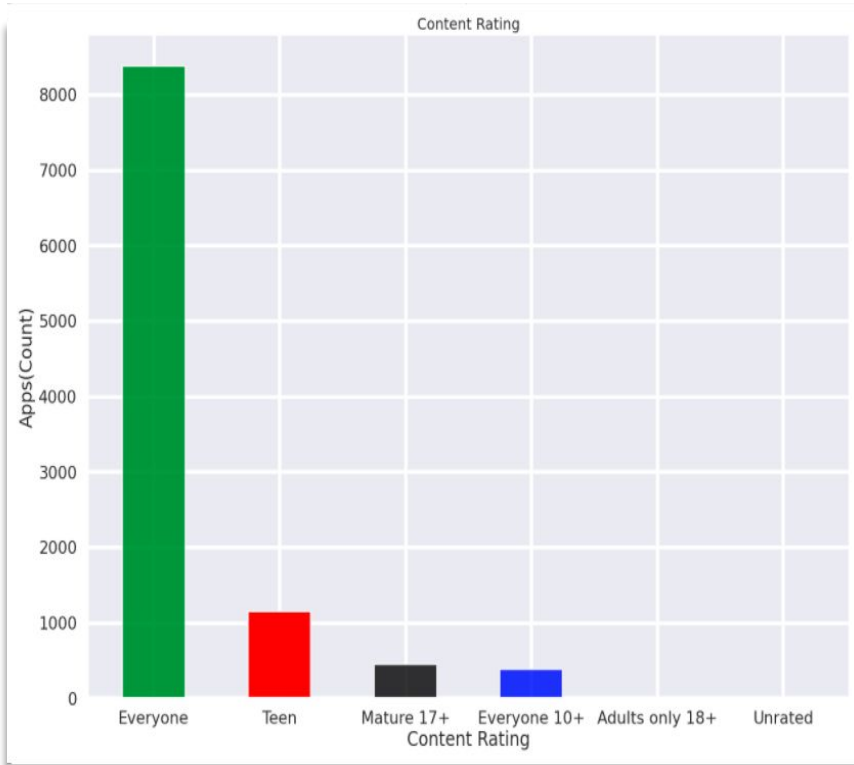
At the same time **Family, Game, Tools and Medical** categories had the biggest number of paid apps available for download.

# TOP 20 APPS PRESENT AND GETTING INSTALLED THE MOST IN THE PLAY STORE AS PER THEIR GENRES



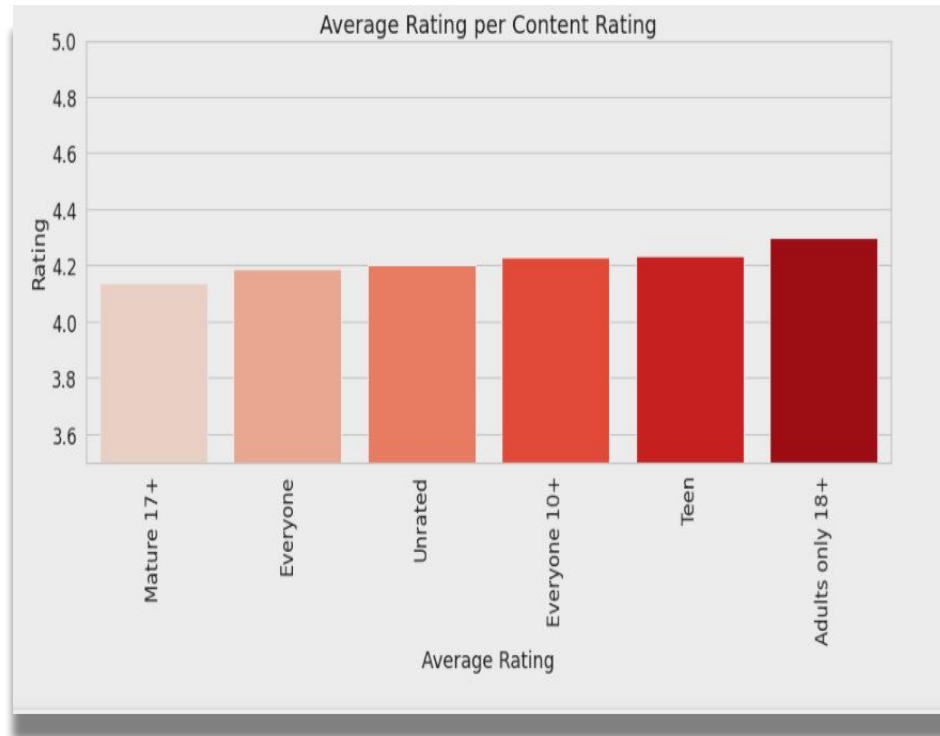
**From the above two plots:** Maximum number of apps present in play store comes under Tools, Entertainment and Education Genres but as per the installation and requirement in the market plot, Maximum installed apps comes under Communication, Tools and Productivity Genres.

# CONTENT RATING



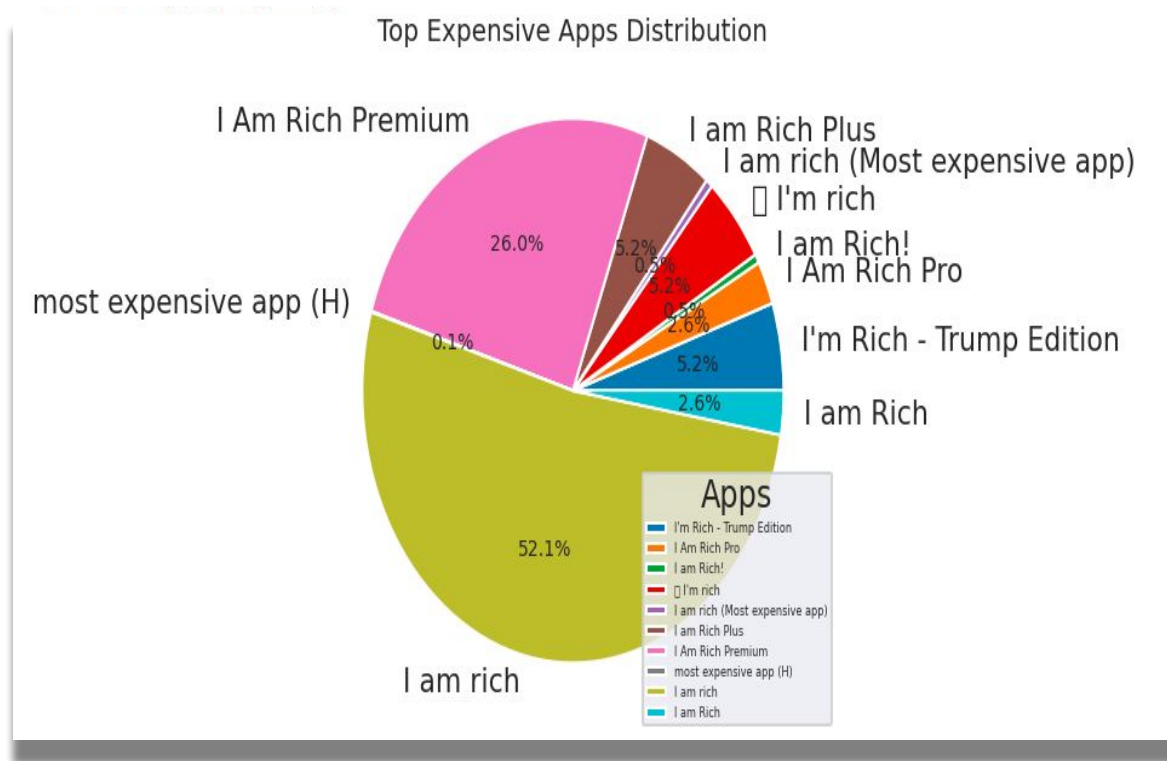
**Everyone category has the highest number of Apps**

# AVERAGE RATING PER CONTENT RATING



**Apps made only for Adults are liked most with an average rating given by them is 4.3**

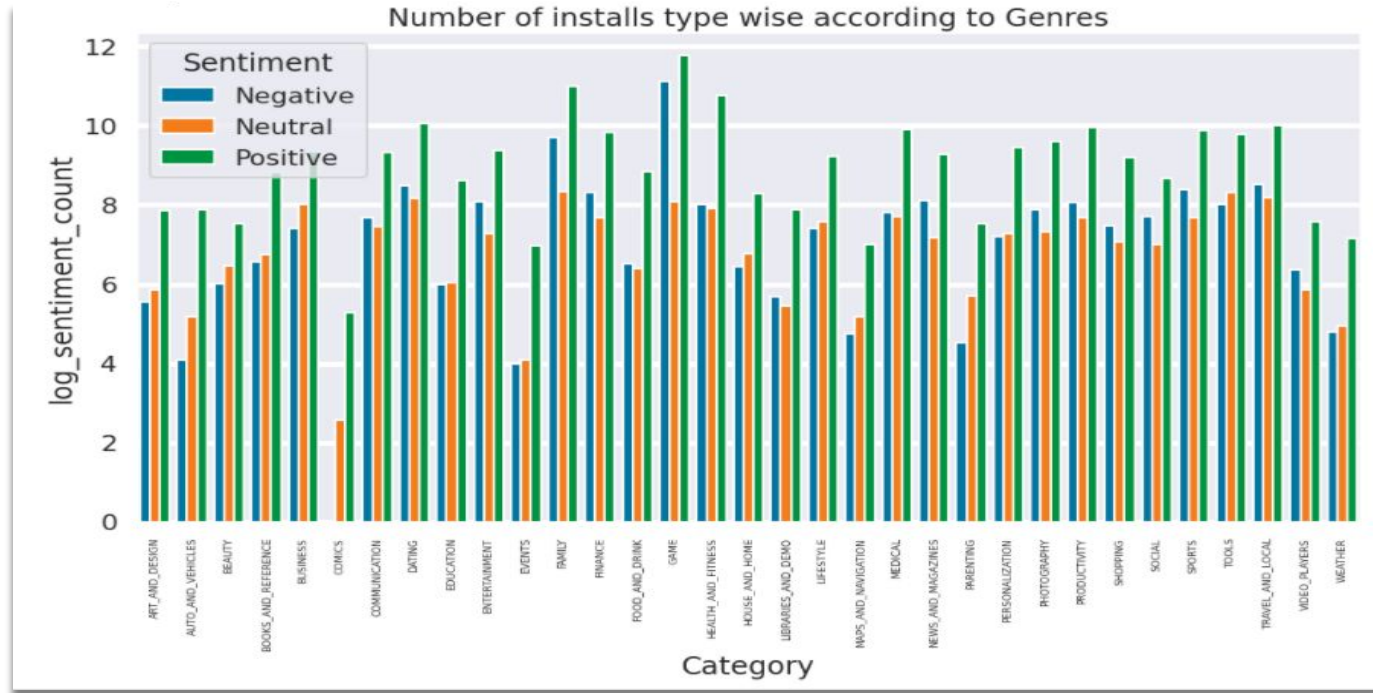
# TOP EXPENSIVE APPS



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

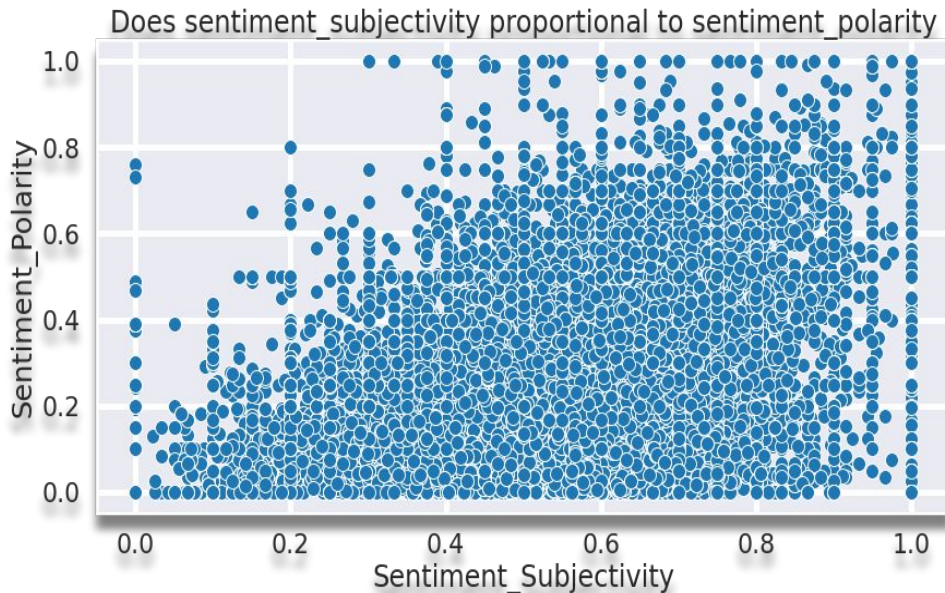


# DISTRIBUTION OF TYPE OF REVIEWS, CATEGORY WISE IN THE DATASET



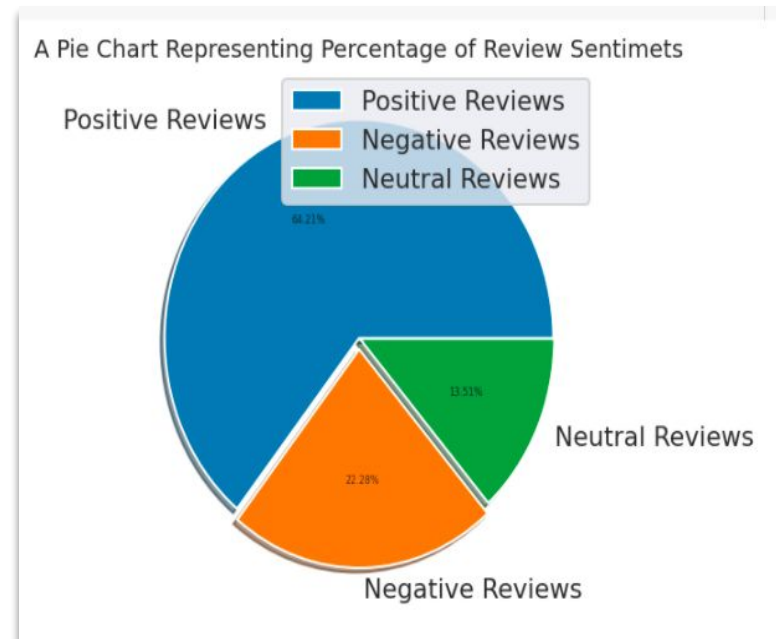
The number of positive reviews are way higher than the Negative and Neutral.

## How does Sentiment\_subjectivity proportional to sentiment\_polarity



The Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, it shows a proportional behavior when variance is too high or low.

## Division of sentiments for different type of reviews



The number of positive reviews are way higher than negative and neutral ones.

# INSIGHTS

Average number of stars for all apps:

4.1 ★★★★★

Total number of apps with 5 star ratings:

271

Category which has highest average rating:

Events, Art and Design, Education

Total number of Free and paid Apps

Free 8905

Paid 754

Average rating of Free and paid Apps:

Free 4.185918

Paid 4.269761

Maximum reviews - Facebook

Top Expensive App - I'm Rich





# CONCLUSION:

Goal is to provide a better understanding of how App's ratings and reviews measure up to the industry as a whole.

1. User mostly prefers more of free apps. Most of the apps present in playstore are more or less of same size so size doesn't affect their decision much.
2. The apps which are available for everyone are having the ratings 4 and above out of 5.
3. Most popular category of apps are on two basis - Number of Installs and number of reviews.
4. Family, Game, Communication, Social, Tools, Medical, Events Apps are the highest earning apps and at the same time Medical, Family, game, Tools apps are the expensive apps which helps us to understand the application market and invest money in the right categories to make good profits.
5. Also, customer ratings plays a huge role while releasing the application.
6. The sentiment does not influence the final rating of the app because of the proportional behaviour.
7. Popular categories have a more positive sentiment.
8. Positive reviews are higher than negative and neutral but the sentiment subjectivity is not always proportional to sentiment polarity.