

# Capstone Project

## Play Store App Review (EDA)

By

**Piyush M. Sonavane**

**Data Scientist Trainee , Almabetter**

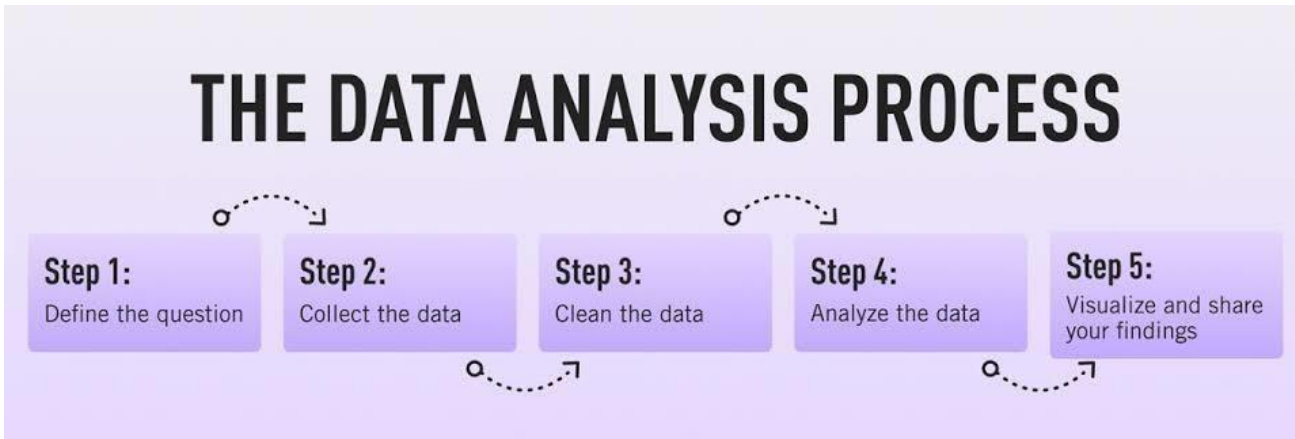
# Content

- Introduction
- Data Analysis Process
- Problem statement
- Data summary
- Data cleaning and exploring
- Data visualization :
  - I. Number of apps in each category
  - II. Top genres with highest number of apps
  - III. Free vs paid app via categories
  - IV. Top categories on Play Store
  - V. App vs content rating
  - VI. App vs Android version
  - VII. Sentiment analysis by user review
  - VIII. Sentiment analysis through categories
- Conclusion

# Introduction

- Google Play Store is the official online app store for Android devices. You can download various forms of media on to your Android devices through Google Play Store
- The Google Play Store is the largest and most popular Android app store. There are more than 3.04 million apps found on play store. The Play Store app data has enormous potential to drive app-making business to success
- Users can install the apps from the Google Play Store and they can give reviews and ratings to the apps based on their experience
- Exploratory Data Analysis (EDA) involves using statistics and visualizations to analyze and identify trends in data sets. It is important to get the maximum insights from a dataset

# Data Analysis Process



- 1) Define the question
- 2) Collect the data
- 3) Clean the data
- 4) Analyze the data
- 5) Data Visualization

# Objective

- The objective of this project is to deliver insights to understand customer demands better and thus assist developers to popularize the product. To clarify, the 'popular' in this project means a high number of installations.



# Problem statement

- i have two datasets ( 1. Playstore dataset , 2. User review dataset ) , one with basic information of different play store apps and the other with userreviews for the respective app
- i examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success
- Due to the presence of such wide variety of apps and the data associated with it.
- i need to extract meaningful insights which are responsible for the apps success and will help developer to capture the android market

# Data Summary

## 1) Play Store data

- Apps – App names.
- Category – The category to which app belongs.
- Rating – Rating of the app.
- Reviews – Number of reviews given to each app.
- Size – Size of the app.
- Installs – Number of installs of each app.
- Type – Free or Paid
- Price – Price of the app in \$.
- Content Rating – Age restriction for each app.
- Genres – Genre the app belongs to.
- Last Updated – When the app is last updated.
- Current Ver – Current version of the app.
- Android Ver – Android version on which the app is supported



# Data Summary

## 2) User Reviews Data :

- Apps – App names
- Translated Review – Reviews given to each app.
- Sentiment – Sentiment of reviews  
Positive/Negative/Neutral.
- Sentiment Polarity – Sentiment polarity score from -1 to 1.
- Sentiment Subjectivity – Sentiment subjectivity score





# Data Cleaning

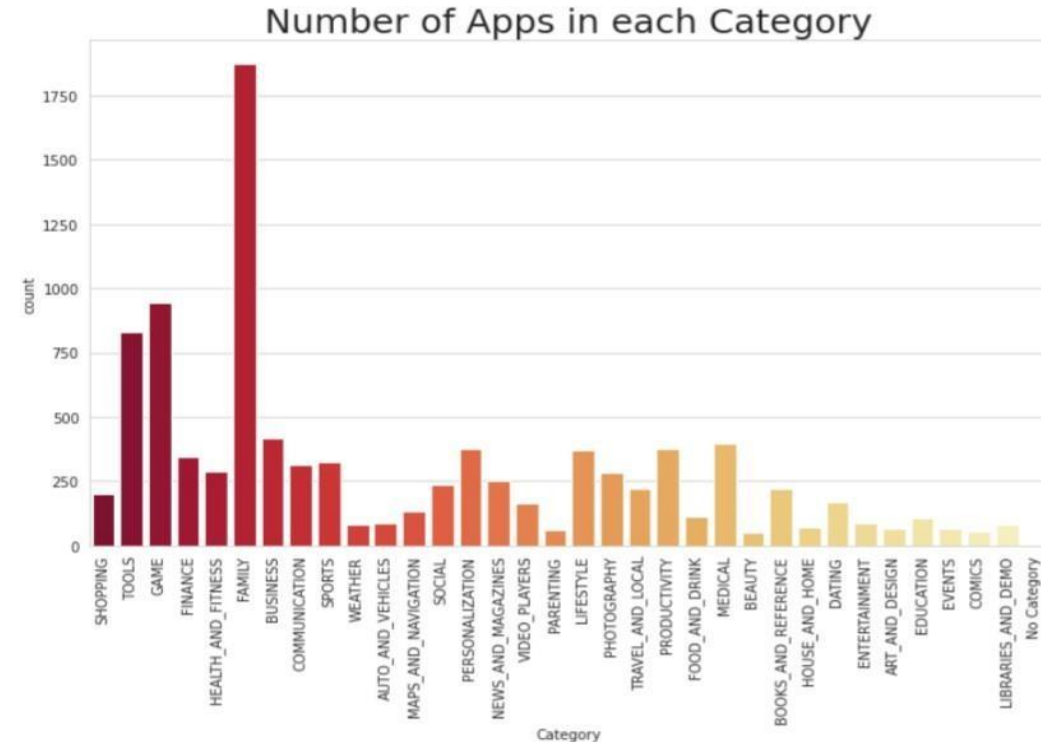
## **This step involves**

- Understanding and exploring the characteristics of datasets
- Converting columns with numeric data into integer/float datatype.
- Checking and removing duplicates.
- Handling the null values.

**After cleaning and organizing the datasets to make it accessible for Visualization in order to analyze it, i'll ready to perform data visualization.**

# Data Visualization

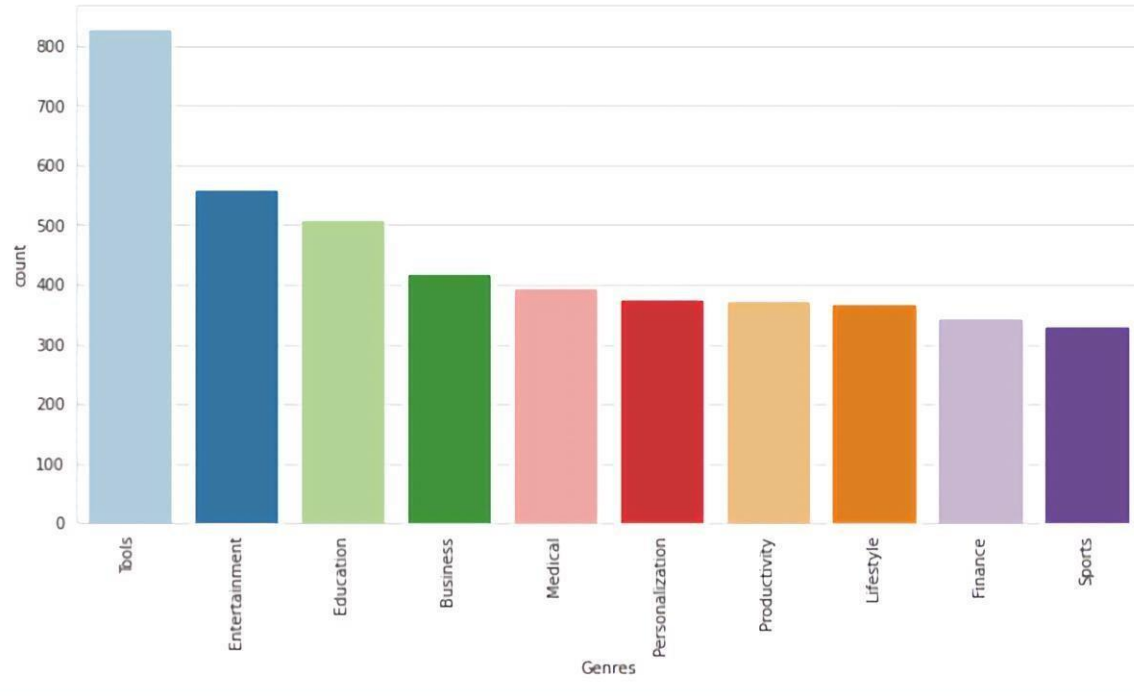
## 1) Number of apps in each category



**Family , Games, Tools are top 3 categories which have maximum number of Apps**

# Data Visualization

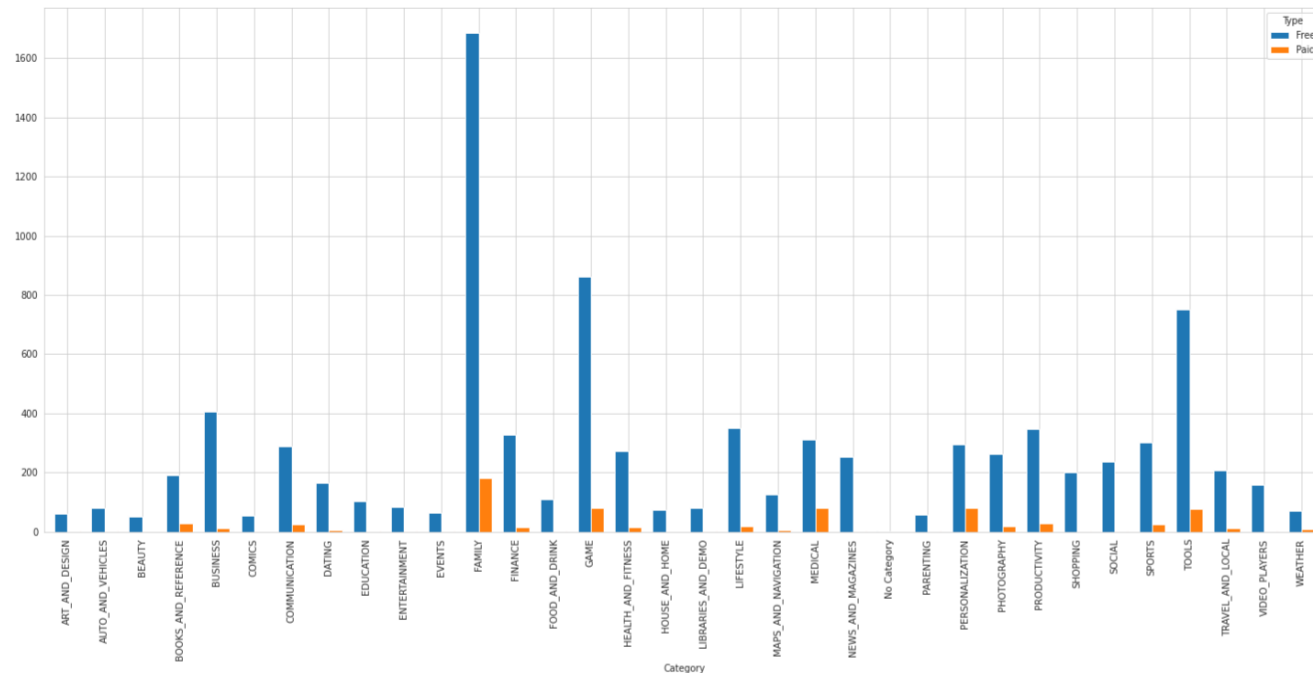
## 2) Top Geners with highest number of apps



- These are the top ten Genres out of 119 Communication and Tools are the most
- Tools and entertainment having most Apps genres

# Data Visualization

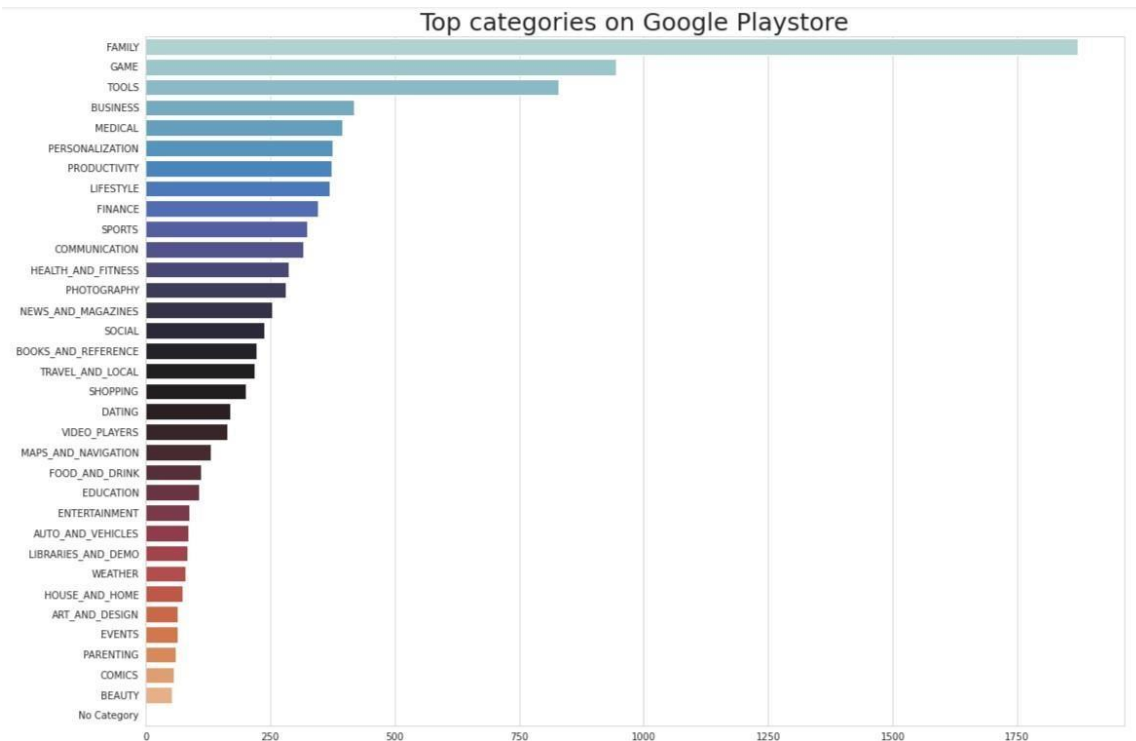
## 3) Free vs paid app via categories



**Family , Game  
Tools are top 3  
categories  
having maximum  
number of free  
and paid apps**

# Data Visualization

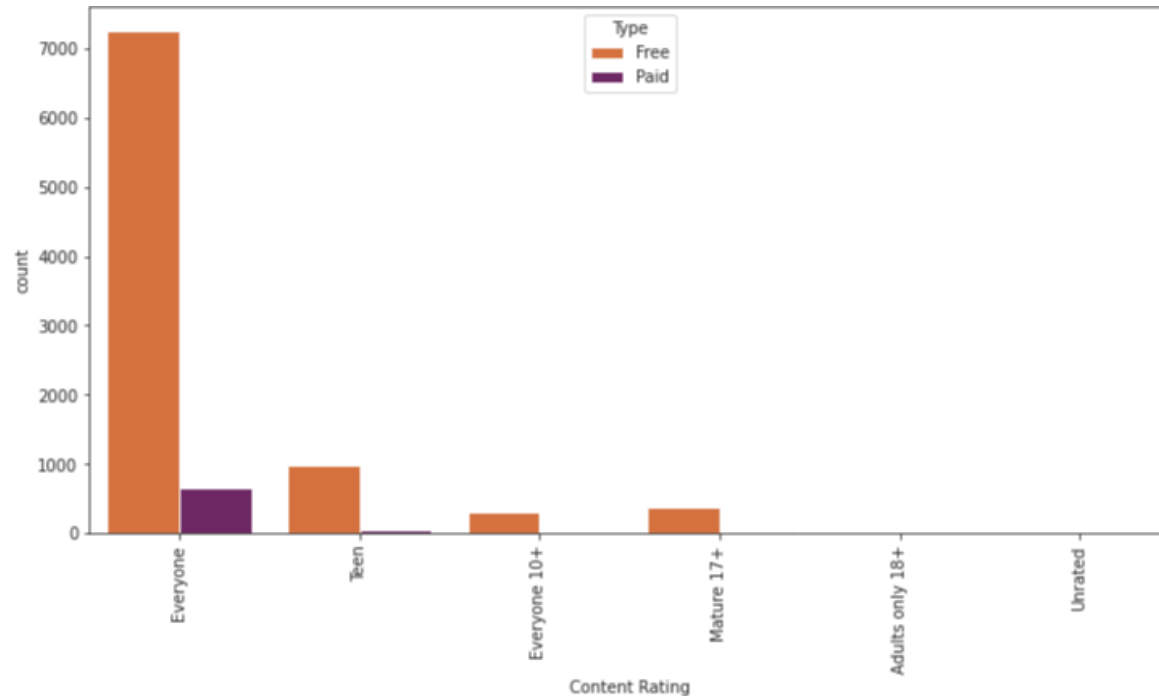
## 4) Top categories on Play Store



**FAMILY , GAME ,  
TOOLS ,  
BUSINESS ,  
MEDICAL Are top 5  
categories on Play  
Store**

# Data Visualization

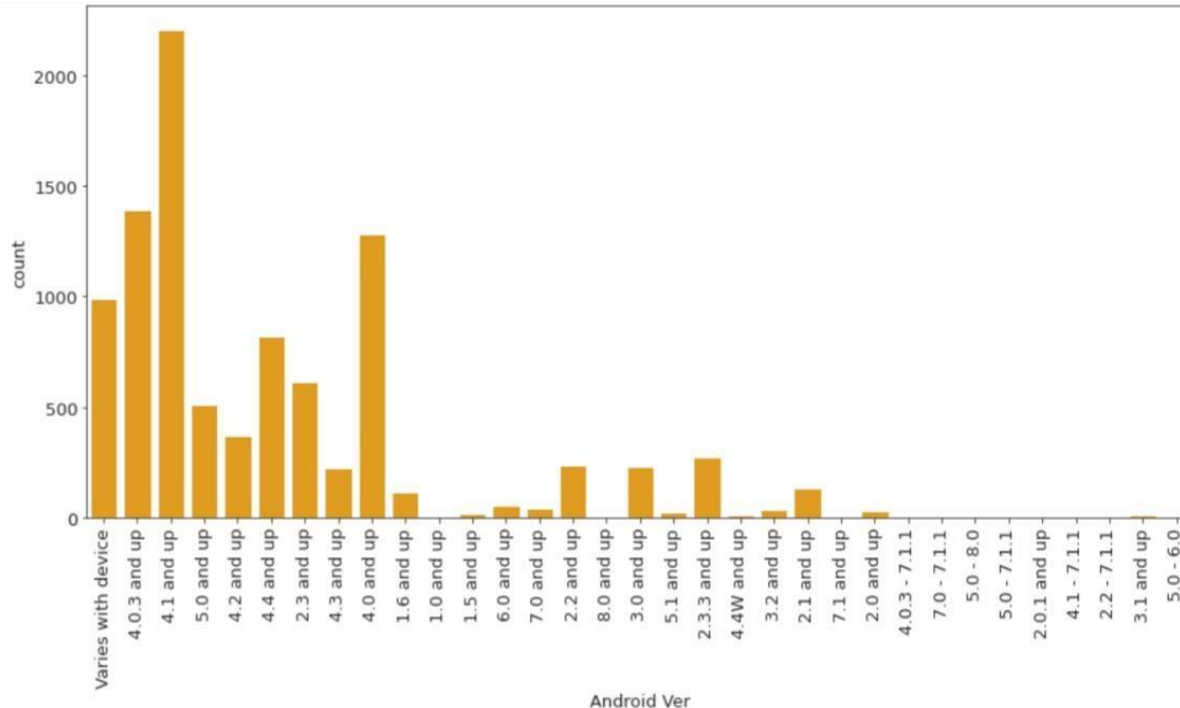
## 5) App vs content rating



**Apps rated for everyone has highest number of Installs followed by Apps rated for teen**

# Data Visualization

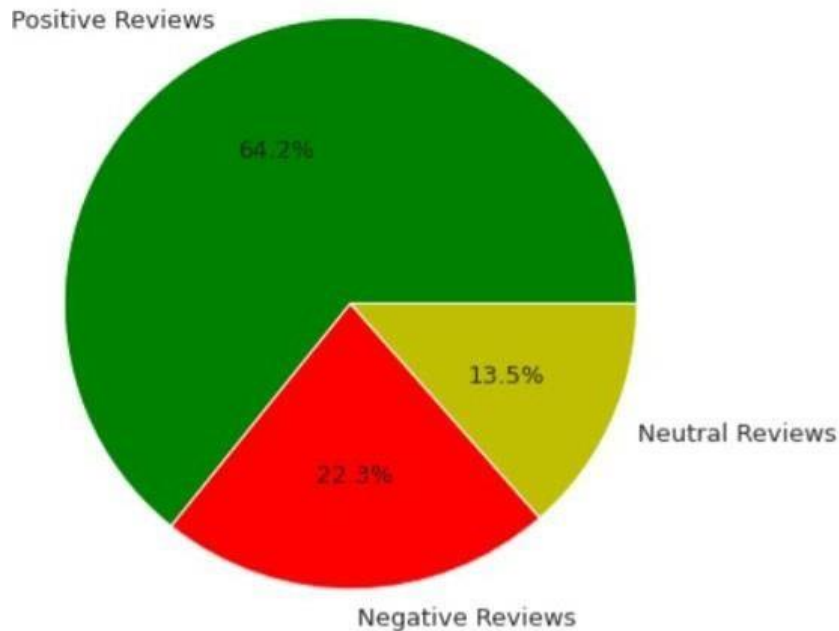
## 6) App vs Android version



**Most of the apps are on Android version 4.1 and up**

# Data Visualization

## 7) Sentiment analysis by user review

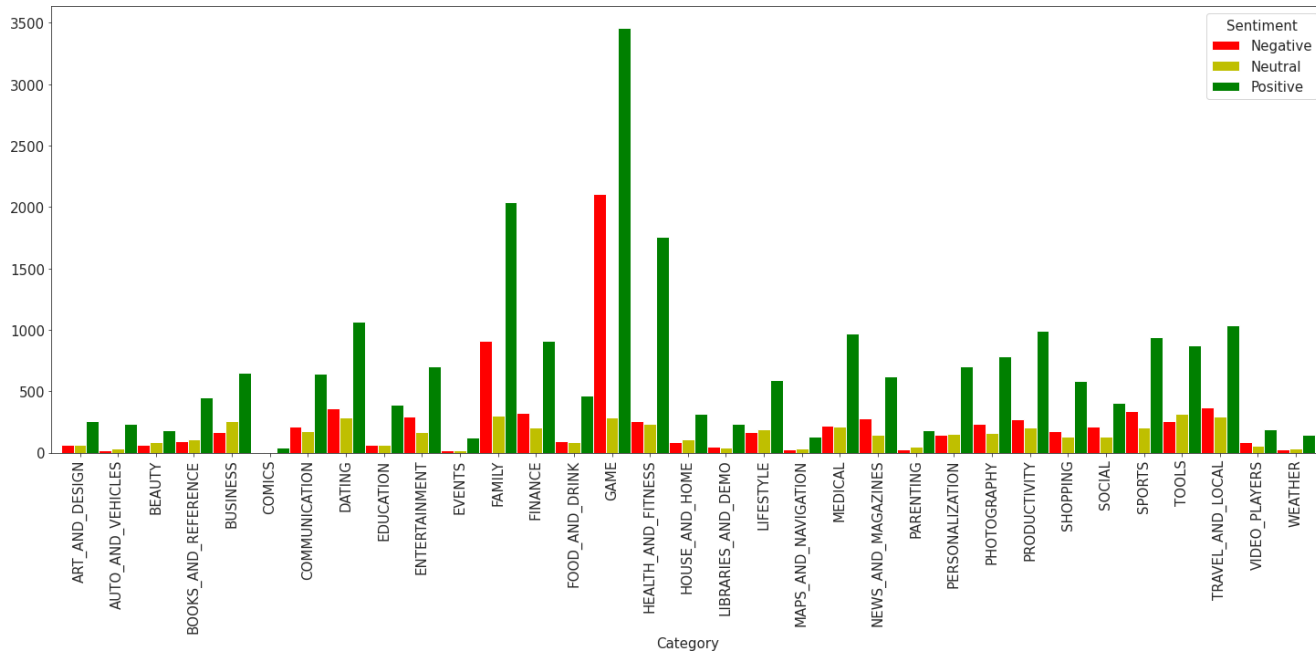


**64.2% More than half of the reviews from user review dataset Are Positive**



# Data Visualization

## 8) Sentiment analysis through categories



**Game category has most number of negative , neutral , and positive sentiments**

# Conclusion

- **Family , Games, Tools are top 3 categories which have maximum number of Apps There are more than 1750 App available in Family category.**
- **Tools and Entertainment Genres have the highest number of apps.**
- **Family , Game Tools are top 3 categories having maximum number of free and paid apps , And Family category has 1600 free Apps 200 paid Apps.**

# Conclusion

- **Family , Game , Tools , Business , Medical Are top 5 categories on Play Store.**
- **Apps rated for everyone has highest number of Installs Followed by Apps rated for teen.**
- **Most of the apps are on Android version '4.1 and up.**
- **64.2 % More than half of the reviews from user review dataset Are Positive.**

# Conclusion

- **The Sentiment Polarity of most of the users is towards the positive side.**
- **Game category has most number of negative , neutral , and positive sentiments.**

**THANK YOU**