Capstone Project -1

Play Store App Review Analysis

Individual Project

Saurabh Shinkar

Problem Statement

- *The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.*
- *Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.*
- *Explore and analyze the data to discover key factors responsible for app engagement and success.*

Understanding Of The Play Store Data

- **App:** Application Name
- **Category:** Category of the Application
- **Rating:** Rating of the Application
- **Review:** Review of the Application
- **Size:** Size of the Application
- **Installs:** No. of installs of the Application
- **Type:** Whether the Application is free or paid
- **Price:** Price of the Application
- **Content Rating:** For whom the content has been created
- **Genres:** Genres to which Application belong
- **Last updated:** Application recently updated Date
- **Current Ver:** Current version of application
- **Android Ver:** Android version of application

Understanding Of The User Reviews Data

App: Application Name
Translated Review: Review In Words
Sentiment: An attitude, thought, or judgment prompted by feeling. It is Positive, Negative or Neutral in our Data.
Sentiment Polarity: It is the expression that determines the sentimental aspect of an opinion
Sentiment Subjectivity: Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

Data Cleaning And Preparation

- **Handling missing values**
  - Looking for percentage of null values of each column
  - For columns containing large null values, replacing null with proper values.
  - For columns containing small null values, dropping those nulls.
  - For columns containing more than 50% null, dropping whole column.

- **Making each column values in proper format**
  - Removing unnecessary signs
  - Changing Type of values into float or integer from string and other format

- **Handling duplicate values**

Insights Into The Play Store Data

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19.0	10000	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14.0	500000	Free	0.0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
U Launcher Lite – FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7	5000000	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25.0	50000000	Free	0.0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8	100000	Free	0.0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

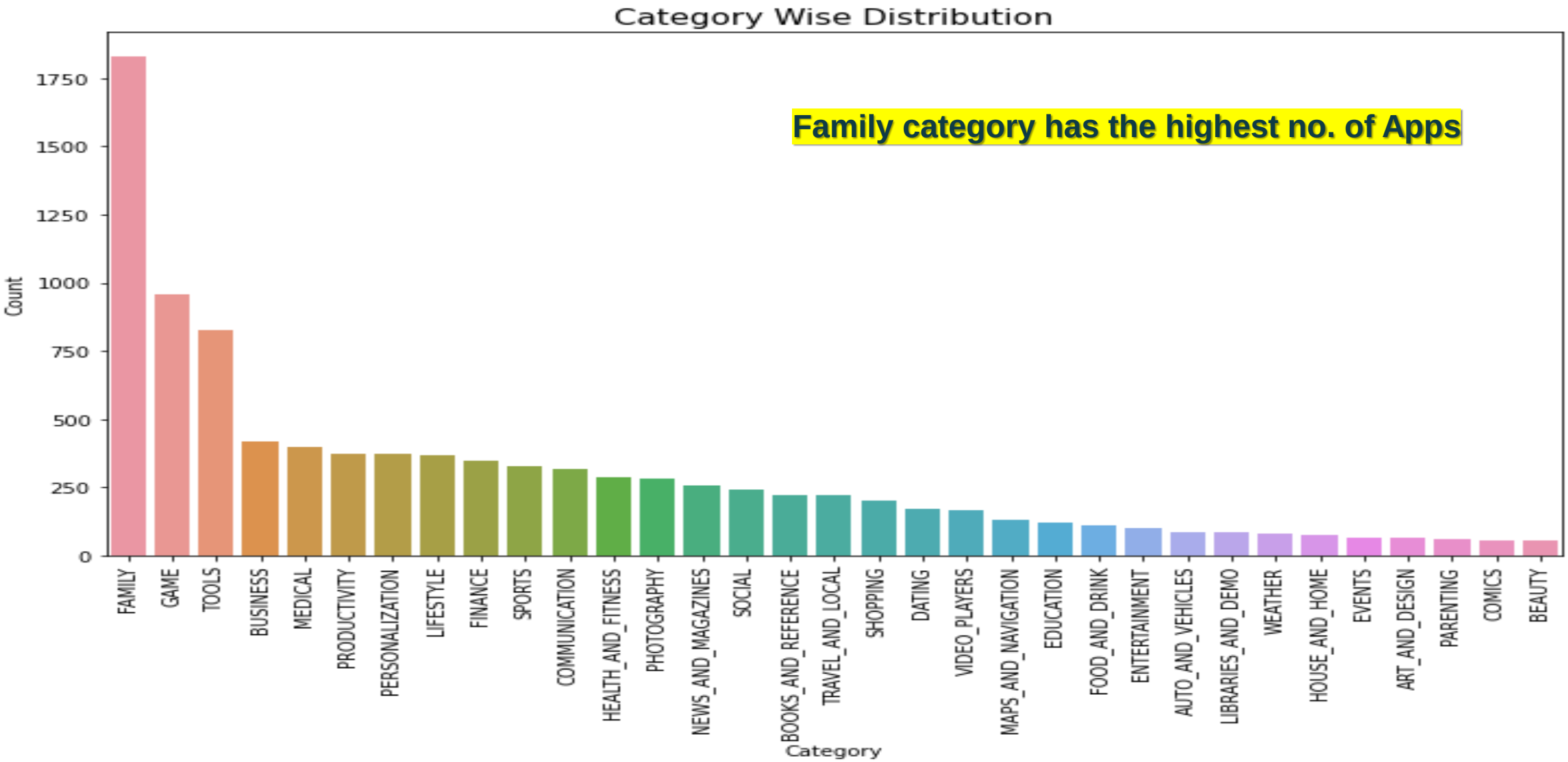
Range Index: 10841 entries, 0 to 10840
Data columns (total 13 columns)
7 Numerical columns and 6 Categorical columns

Insights Into The User Reviews Data

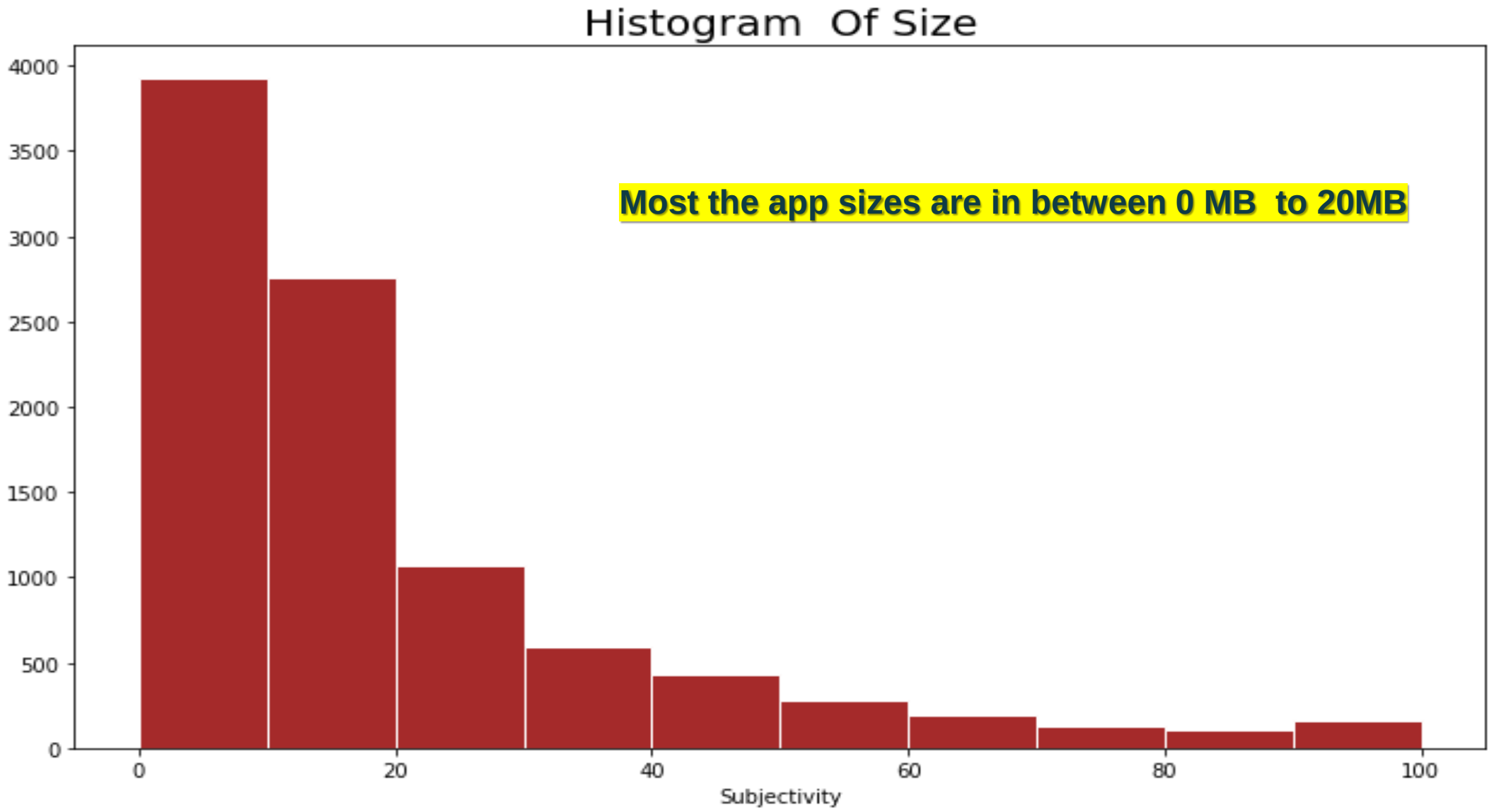
	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking food myself, case "10 Best Foods" helps lot, also "Best Before (Shelf Life)"	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
2	10 Best Foods for You	NaN	NaN	NaN	NaN
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000

Range Index: 64295 entries, 0 to 64294
Data columns (total 5 columns)
2 Numerical and 3 Categorical features

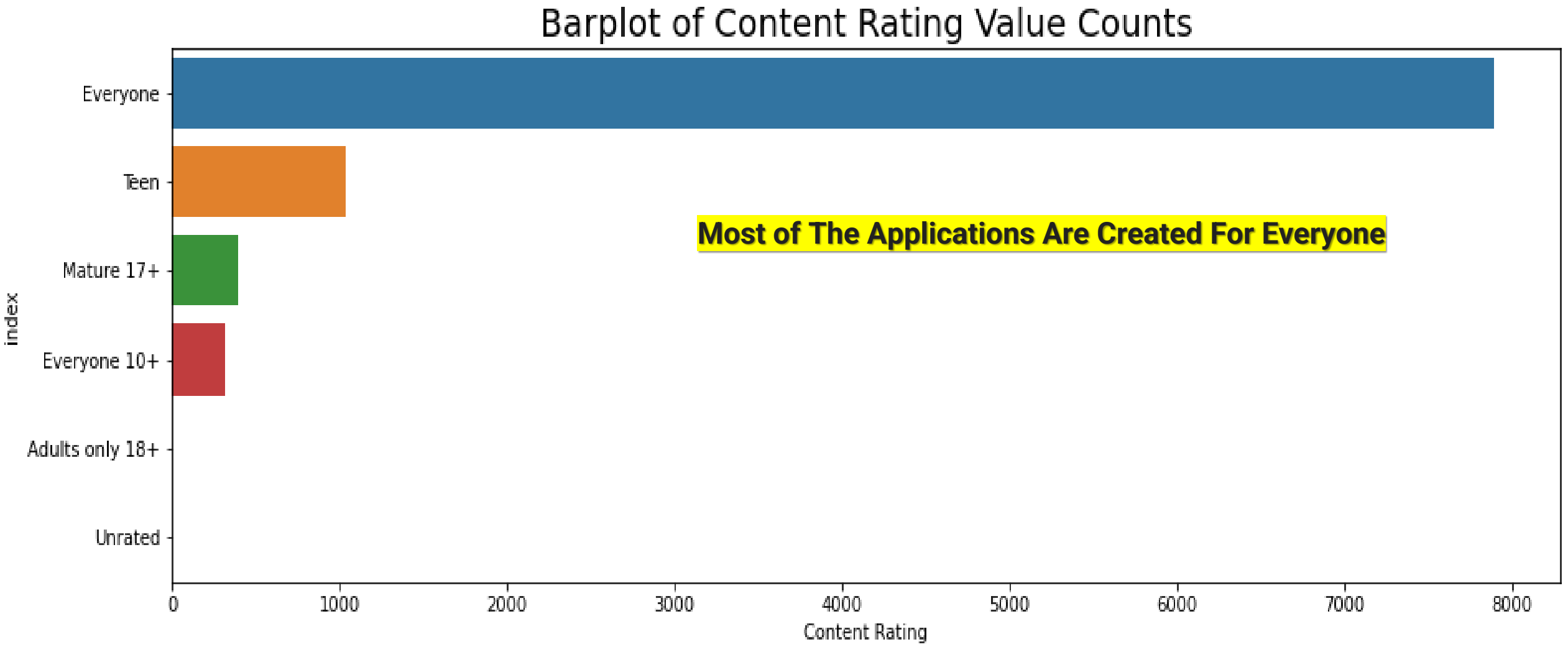
Category Wise Distribution Of The Apps



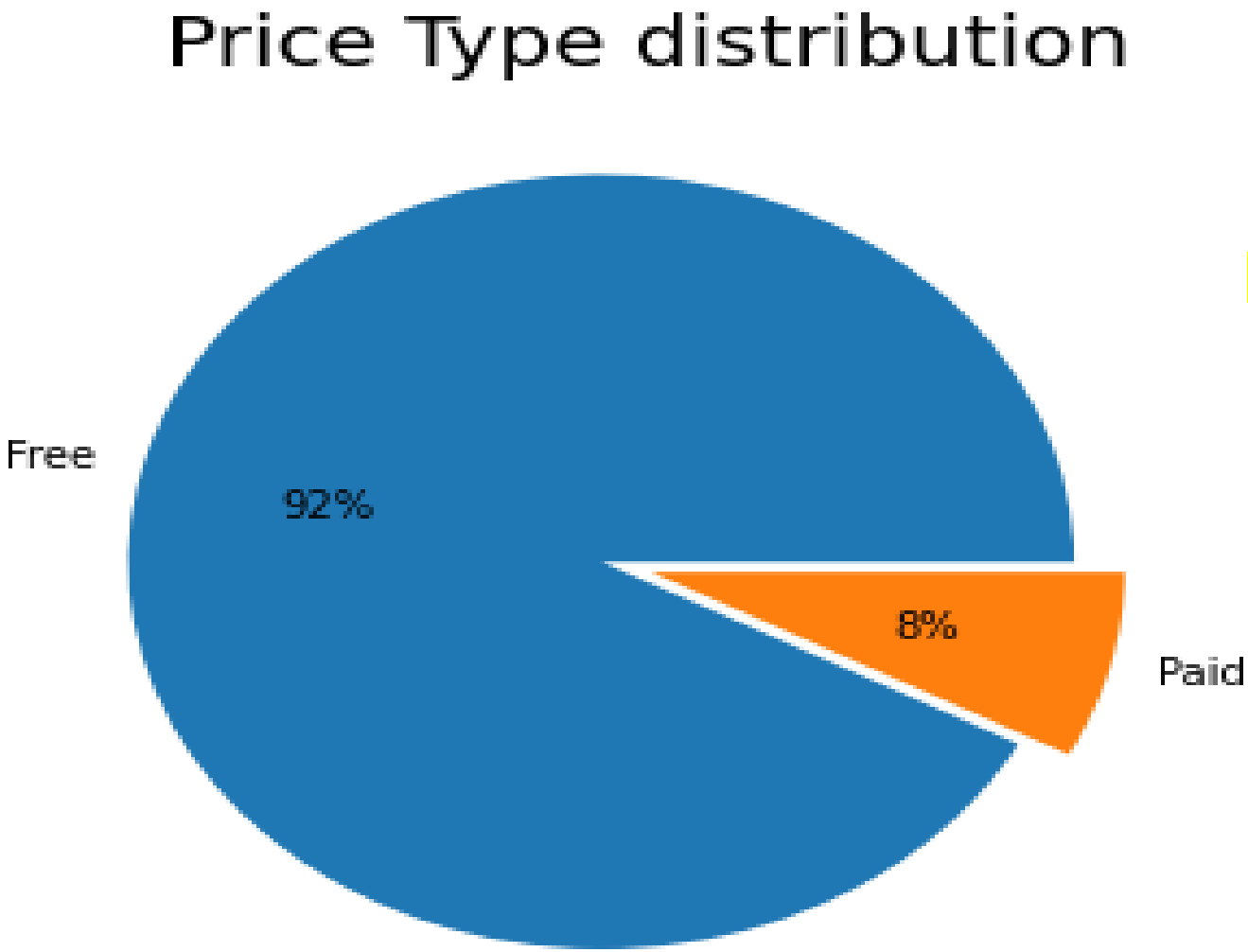
Distribution Of Size



Content Rating Value Counts

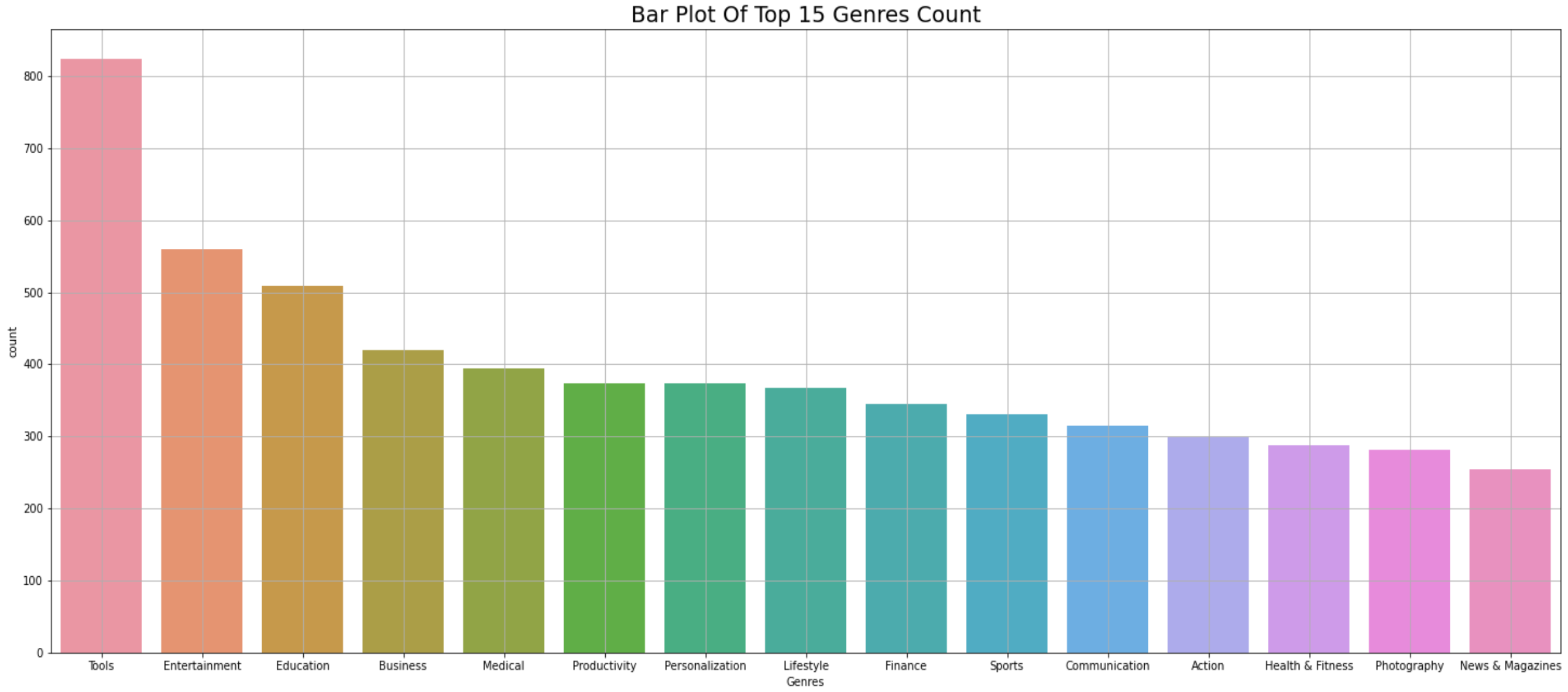


Free And Paid Apps

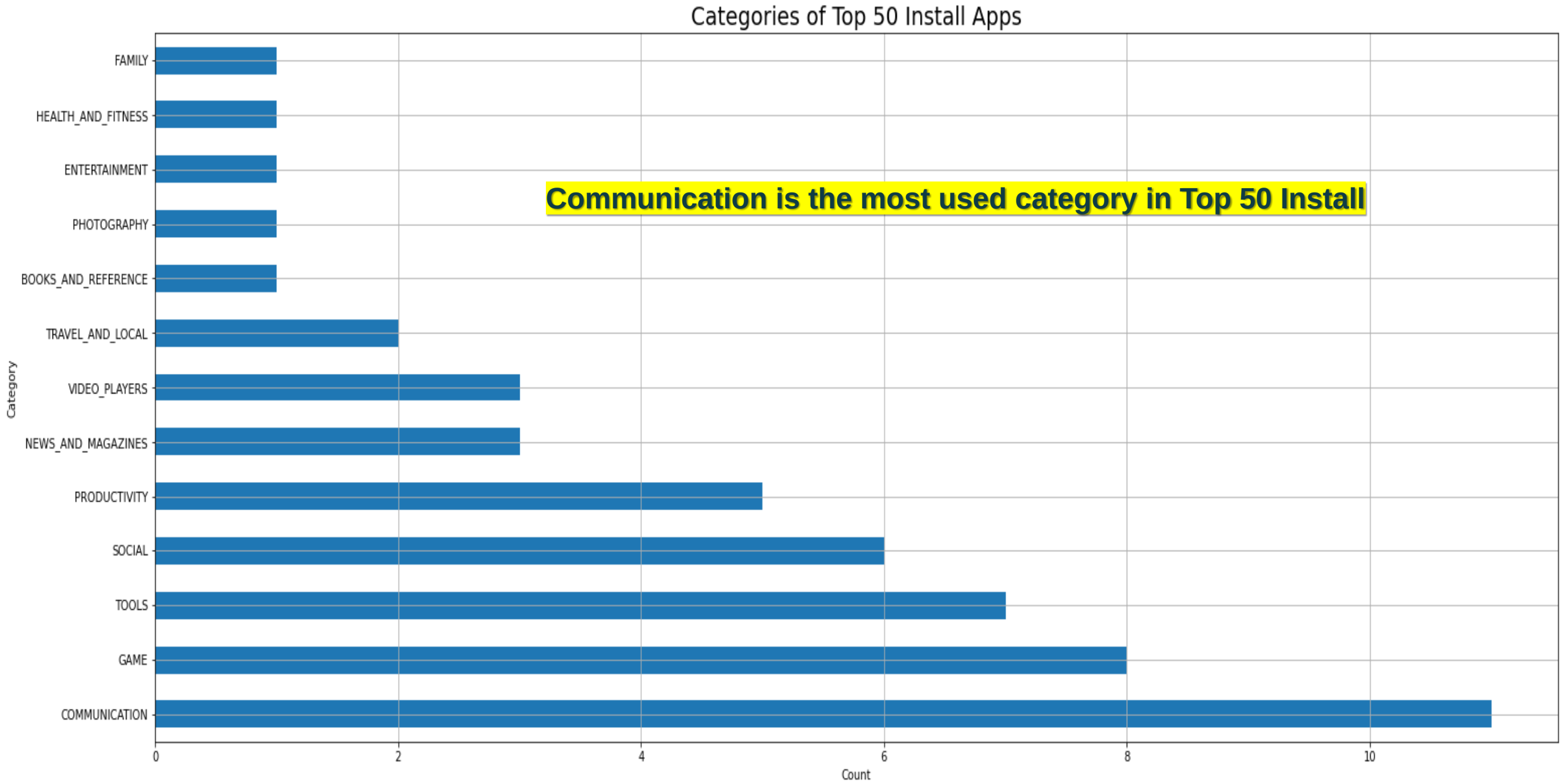


Approximately 92% apps are free

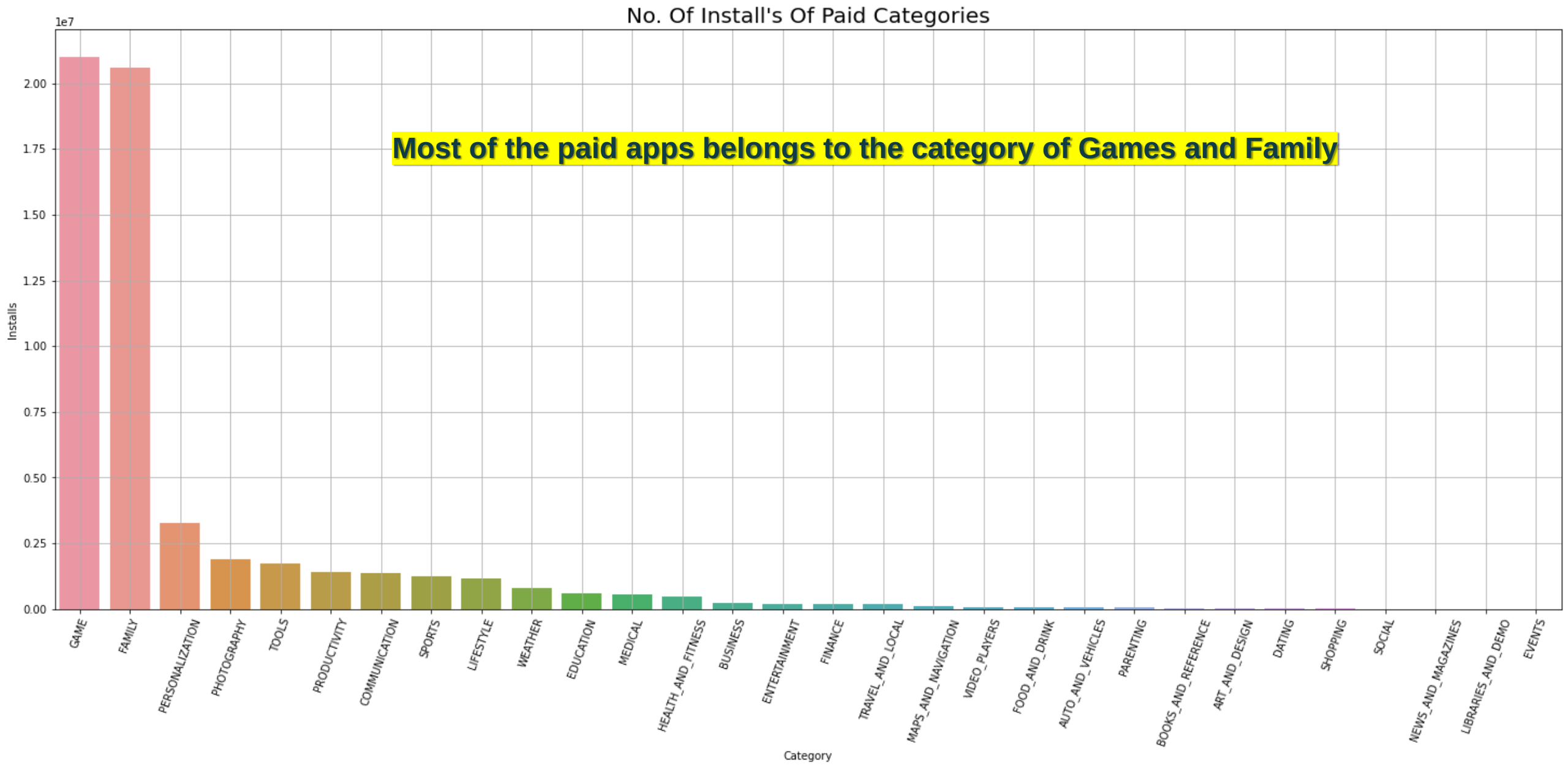
Top 15 Genres Count



Top 50 Install App's Category



Installs Of Paid Category



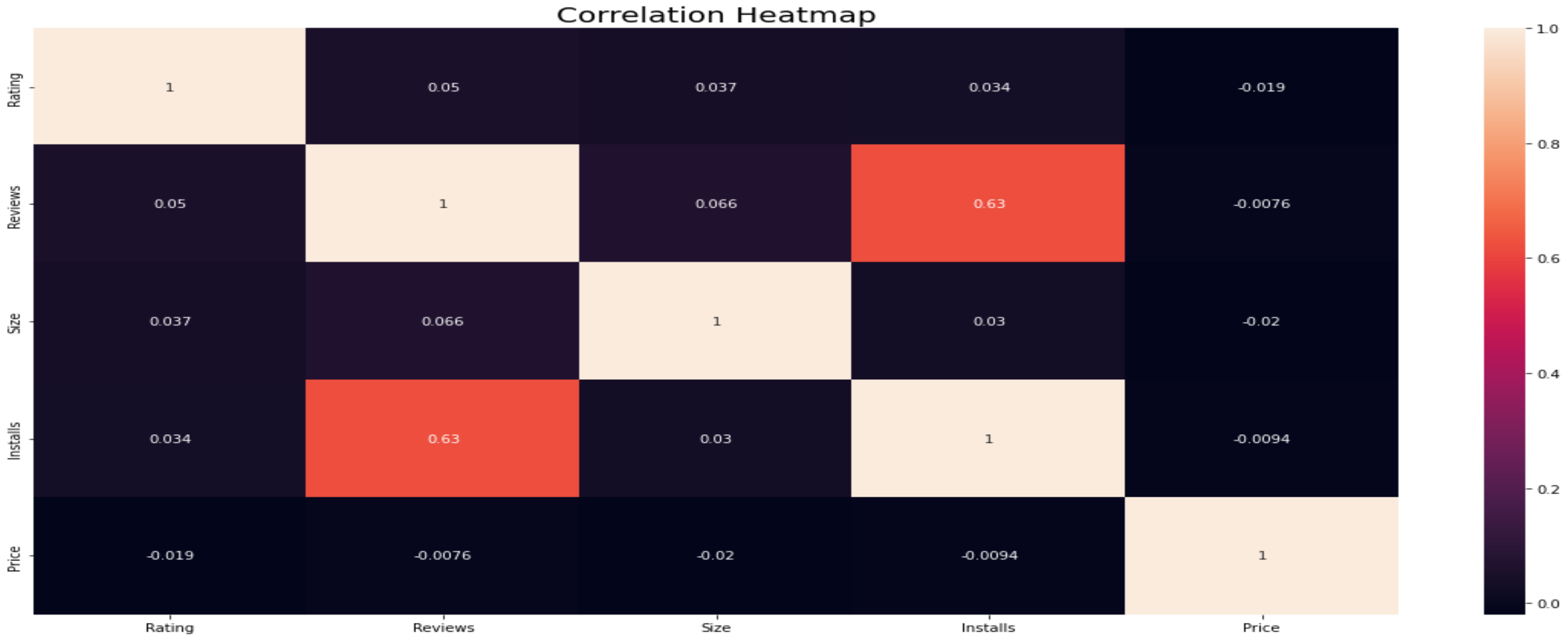
Category Sorted By There Mean Price

Category	Mean Price
FINANCE	8.408203
LIFESTYLE	6.398022
MEDICAL	2.520759
EVENTS	1.718594
FAMILY	1.312292

Mean Price And Mean Rating of Top 5 Categories By No. Of Installs

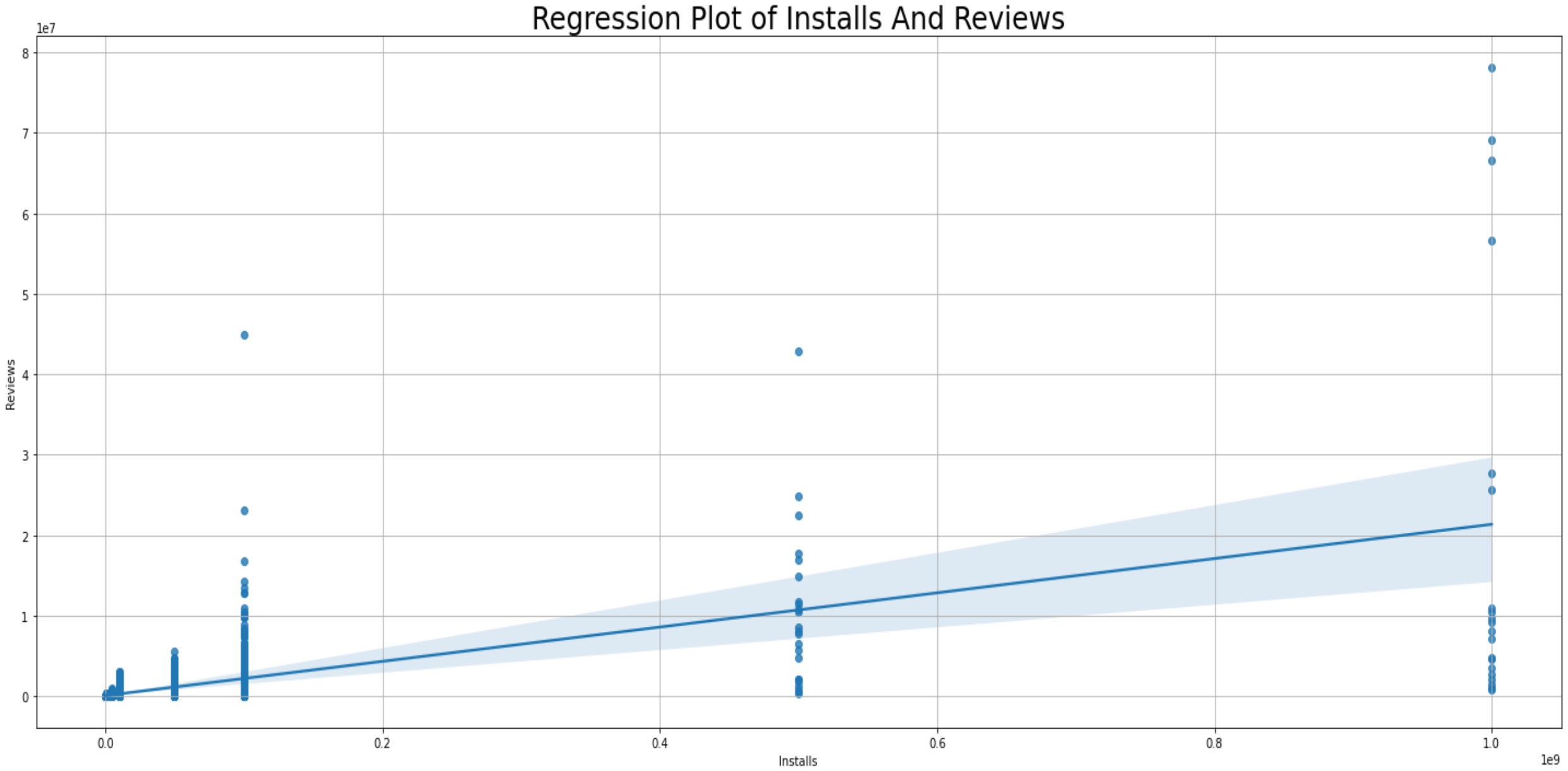
Category	Total Installs	Mean Rating	Mean Price
GAME	13878924415	4.249948	0.296465
COMMUNICATION	11038276251	4.154921	0.263937
TOOLS	8001271905	4.073455	0.322739
PRODUCTIVITY	5793091369	4.206150	0.670936
SOCIAL	5487867902	4.255230	0.066820

Correlation Heatmap Of The Play Store Data



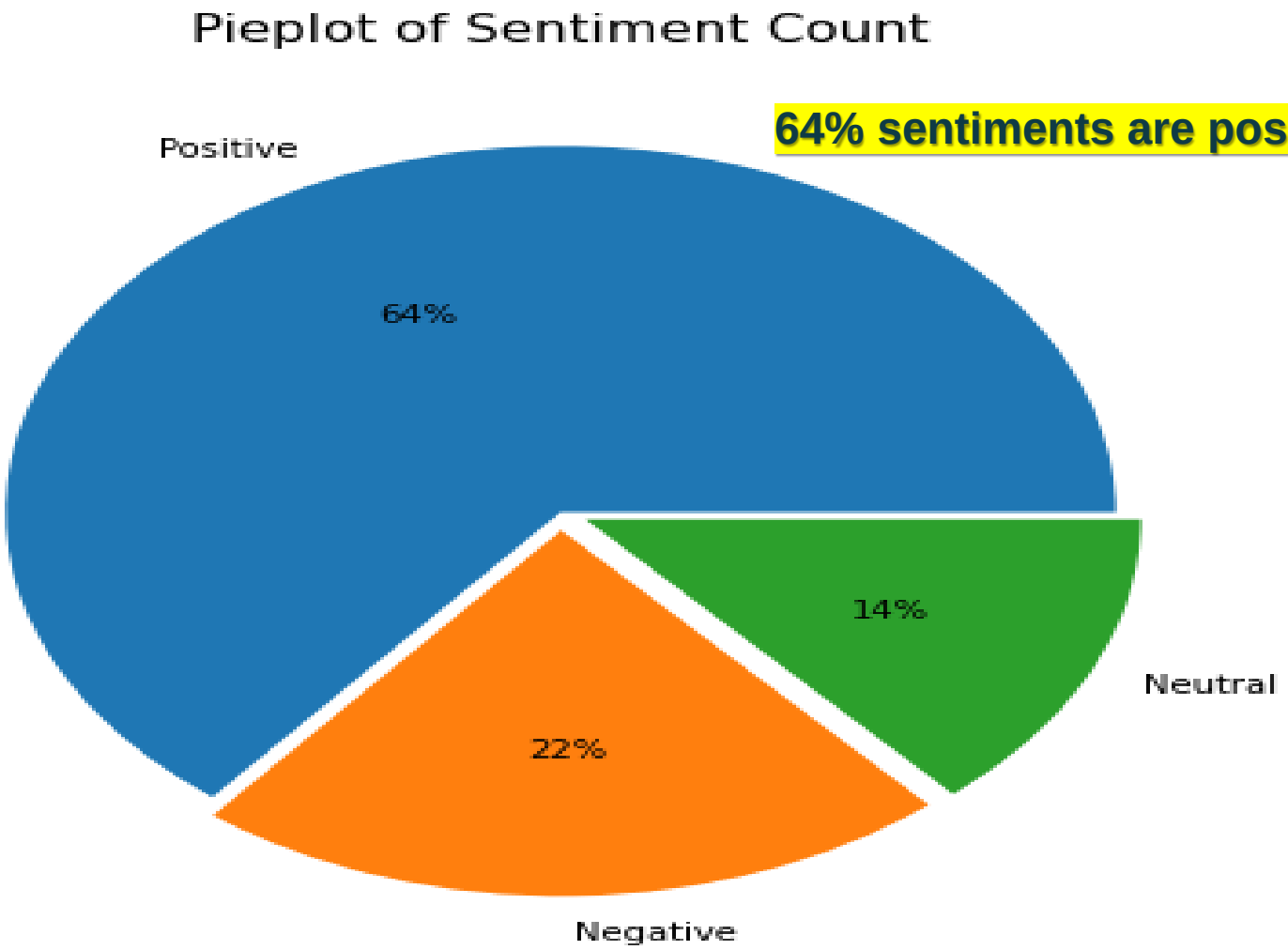
Installs are positively correlated with reviews with correlation value of 0.63

Regression Line Of Installs And Reviews



Word Cloud Image Of Translated Reviews

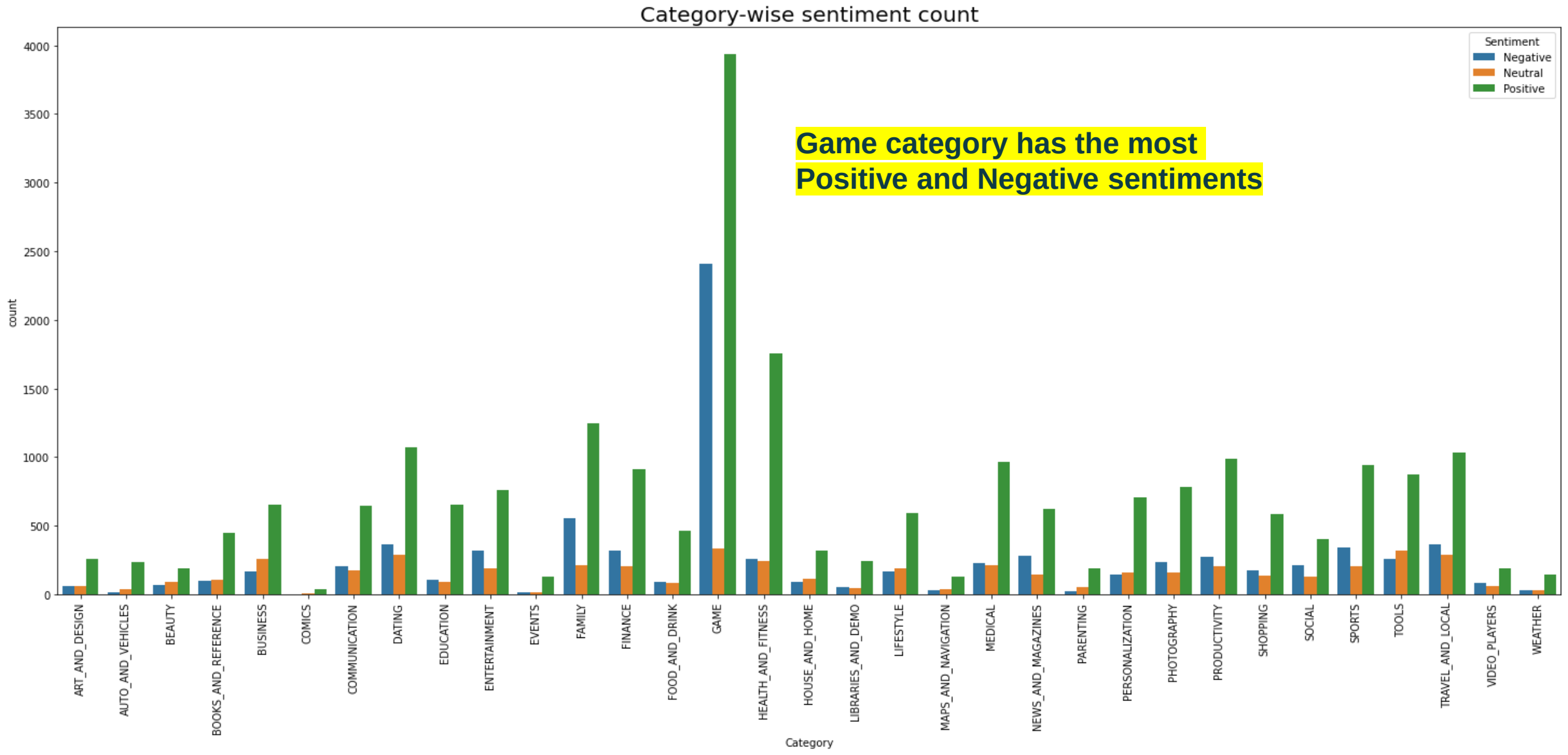
Sentiment Count



64% sentiments are positive, 22% are negative and 14% are neutral.

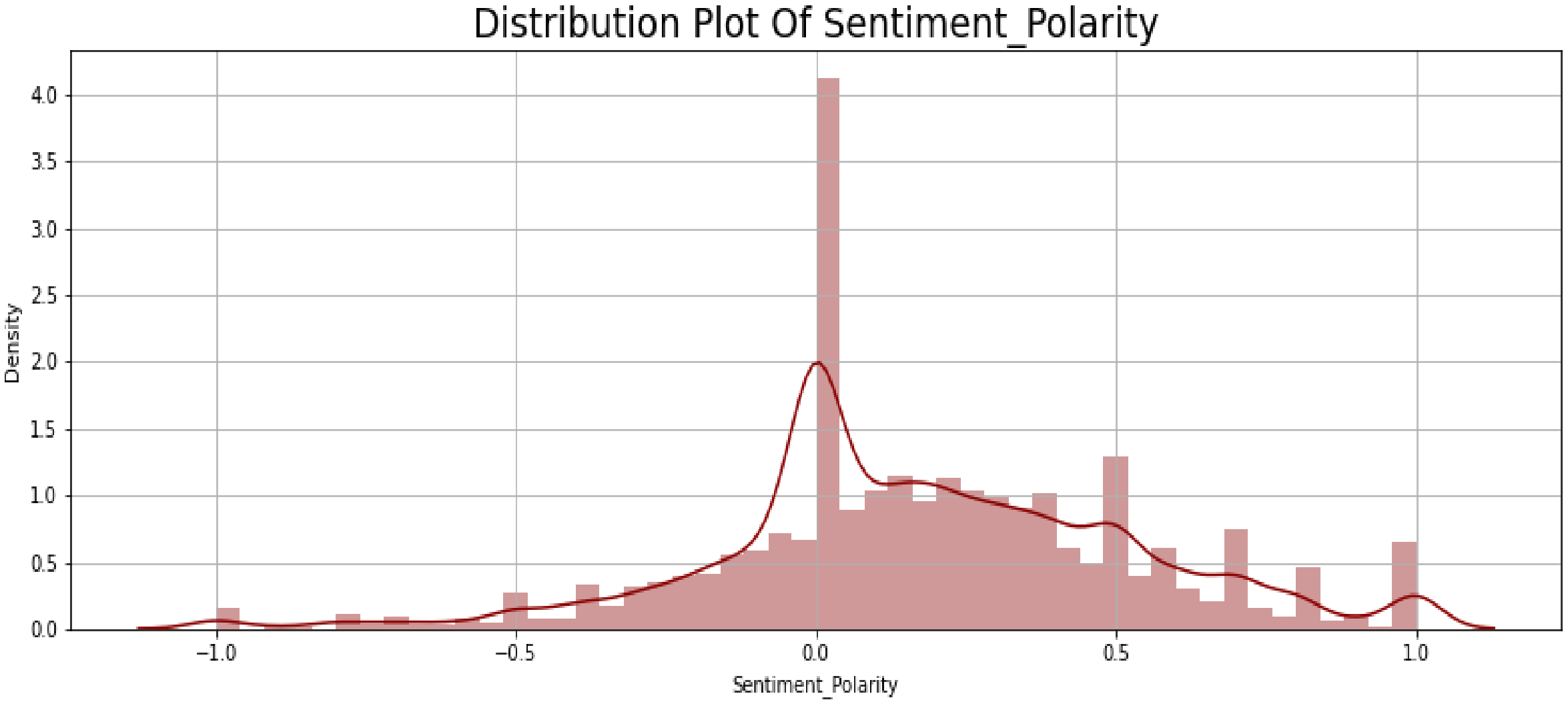
Sentiment	count
Positive	23073
Negative	8005
Neutral	4851

Category And Sentiment Counts

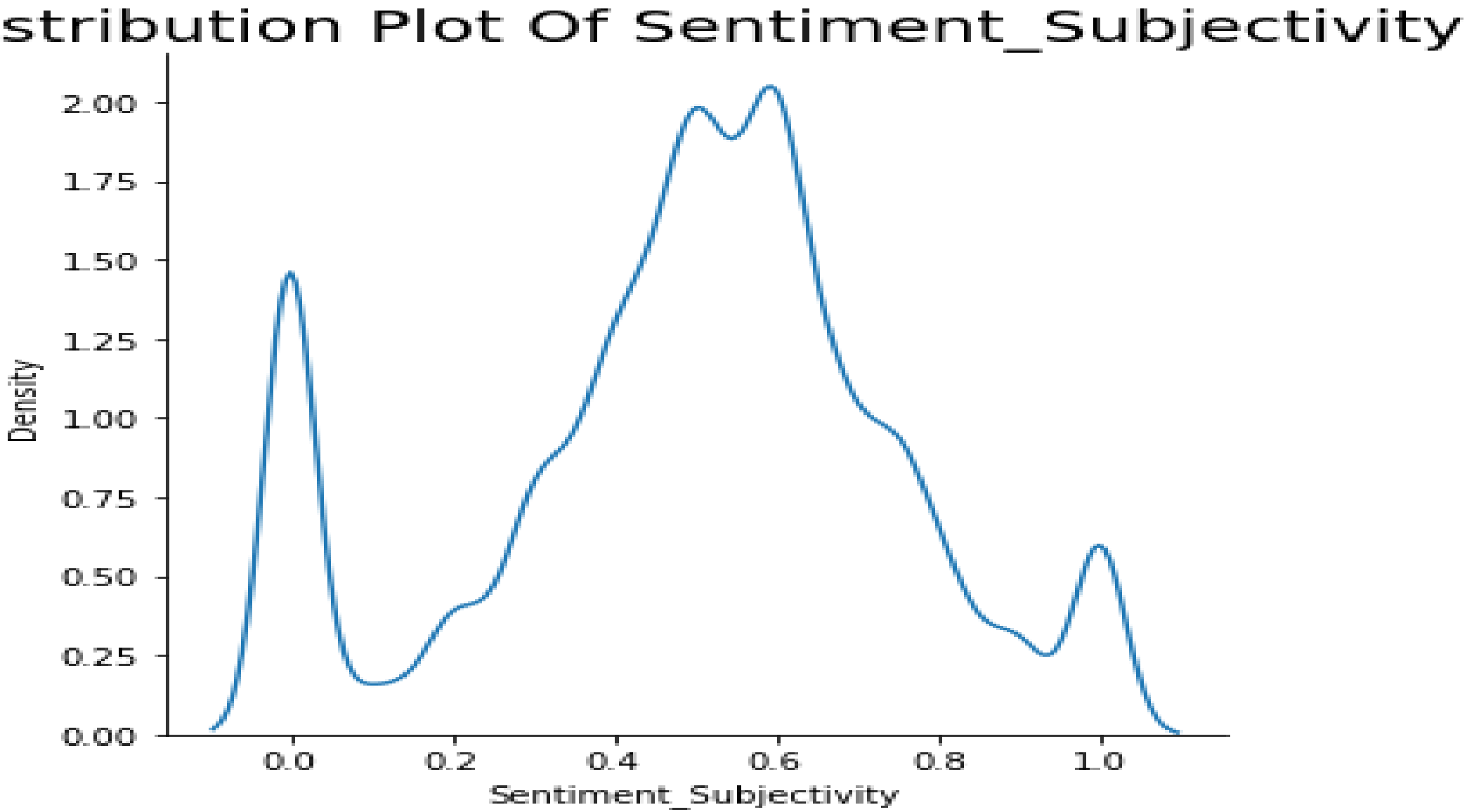


Game category has the most Positive and Negative sentiments

Sentiment Polarity



Sentiment Subjectivity



Conclusion’s

- Most of the app's belongs to Family category. They are approximately 19% among all categories.
- Ratings are negatively skewed with mean rating of 4.19
- Sizes are positively skewed with median size of 12 MB.
- Prices are positively skewed with mean price of 1,median price of 0 and max price of 400.
- Most of The Applications Are Created For Everyone
- Approximately 92% apps are free.
- Tools, Entertainment and Education are the top 3 genres.
- Approximately 64% sentiments are positive, 22% are negative and 14% are neutral.
- Mean Sentiment Polarity is 0.180904
- Among Top 50 install apps 22% belongs to the communication category and 16% belongs to the Game category.

Conclusion’s (continued)

- Events category have the maximum mean rating of 4.395313
- Genres of Communication, Tools, Productivity and Social have the maximum no. of installs.
- Apps belonging to Finance category have the highest mean price (8.408203) among all categories.
- Most of the app sizes are in between 0 MB to 20 MB
- Family category has the highest no. of paid apps.
- Game And Family Category Have The Highest No. Of Installs Of Paid Apps.
- Installs are positively correlated with reviews with correlation value of 0.63
- Mean Sentiment Subjectivity is 0.493767
- There are 271 apps which have 5 star rating. In which 67 apps belongs to the Family category.

Thank
you