

Capstone Project-1

Play Store App Review Analysis

By

Aniket Deshmukh
Mahesh Landage
Prajakta Dangale

Introduction

Google Play Store or formerly Android Market, is a digital distribution service developed and operated by Google. It is an official apps store that provides variety content such as apps, books, magazines, music, movies and television programs. It serves as a platform to allow users with 'Google certified' Android operating system devices to download applications developed and published on the platform either with a paid or free of cost. With the rapidly growth of Android devices and apps, it would be interesting to perform data analysis on the data to obtain valuable insights.



Problem Statement

- ❖ The Play Store apps data has enormous potential to drive app-making businesses to success. Android is expanding as an operating system and Mobile app industry is increasing significantly and thus giving rise to more competitions to the one's that are creating applications.
- ❖ Due to the competition in the market and also expansion in order to help our developer understand what kinds of apps are likely to attract more users and what is the motivating factor for the people to download an app we analyze and research relevant data.
- ❖ For the app development industry where they can analyse the downloads and demand off app download in the industry.
- ❖ We aim on providing doing sentimental analysis on the apps that generated most positive and negative sentiments and sustainability of app in market on basis of previous data and current market.

Dataset Preparation

- **Import Libraries:** NumPy, Pandas, Seaborn, Matplotlib and Datetime.
- **Loading files:** Add both files Play store app dataset and User Reviews dataset.
- **Explore and Analyze the data:** Find useful information and describe it briefly.
- **Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data and Correcting Data Types.
- **Find Insights:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

First Dataset : Play Store Data

1. App : The name of the app
2. Category : The category of the app
3. Rating : The rating of the app in the Play Store
4. Reviews : The number of reviews of the app
5. Size : The size of the app
6. Install : The number of installs of the app
7. Type : The type of the app
8. Price : The price of the app
9. Content Rating : The rating given by user
10. Genres: The genre of the app
11. Last Updated : The date when the app was last updated
12. Current Ver : The current version of the app
13. Android Ver : The minimum Android version required to run the app



Second Dataset: User Reviews

1. App : The name of the application
2. Translated_Review : User comment
3. Sentiment : User view or opinion
4. Sentiment_Polarity : Strongness of Sentiment polarity
5. Sentiment_Subjectivity : It is a public opinion and not a factual information

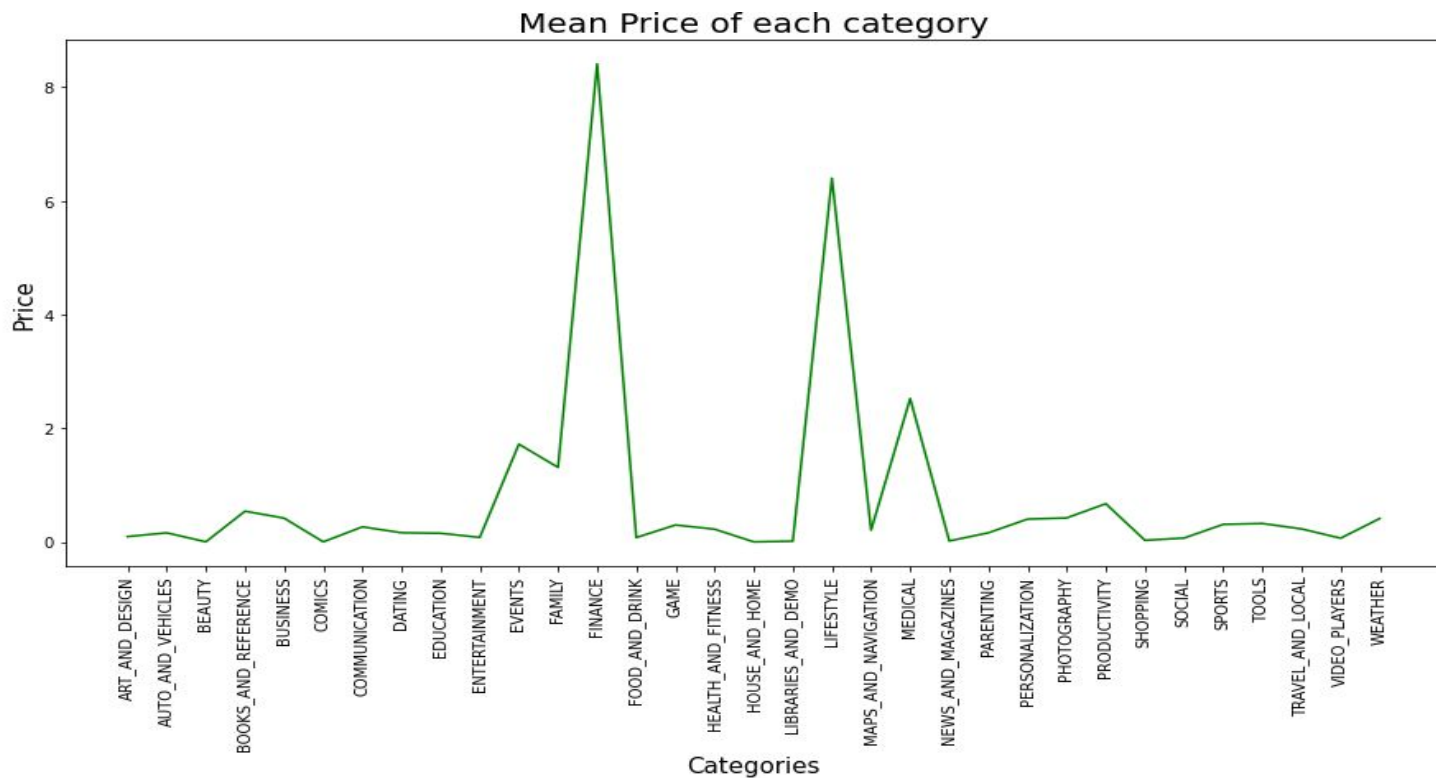


Data Cleaning Methods and Process



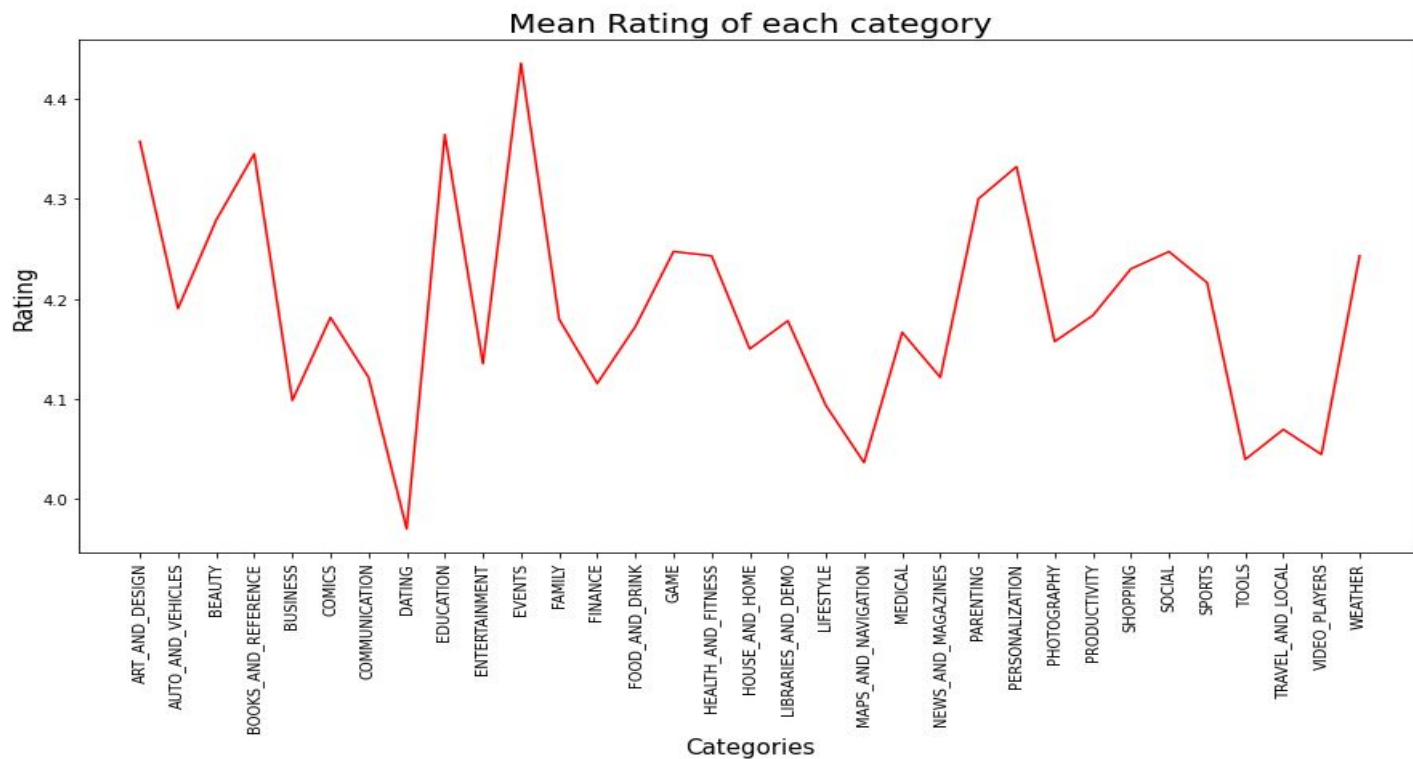
Dataset	Columns	Methods
Play store app reviews data	Category	Drop non 'str' values.
	Rating	1.Place 'NaN' with null. 2. Transform to numbers.
	Reviews	Transform to numbers.
	Size	1. Change 'MB' and 'KB' to bytes. 2. Transform to numbers.
	Installs	1. Remove '+' and ',' . 2. Transform to numbers.
	Type	1.Place 'NaN' with null. 2. Remove Duplicates.
	Price	1. Remove '\$' . 2. Transform to numbers.
	Genres	1. Split into to columns by ';'.
User reviews data	Sentiment Polarity	1.Place 'NaN' with null. 2. Transform to numbers.
	Sentiment Subjectivity	1.Place 'NaN' with null. 2. Transform to numbers.

Mean Price vs Category



Highest Mean price is for **FINANCE** category and after that **LIFESTYLE** category.

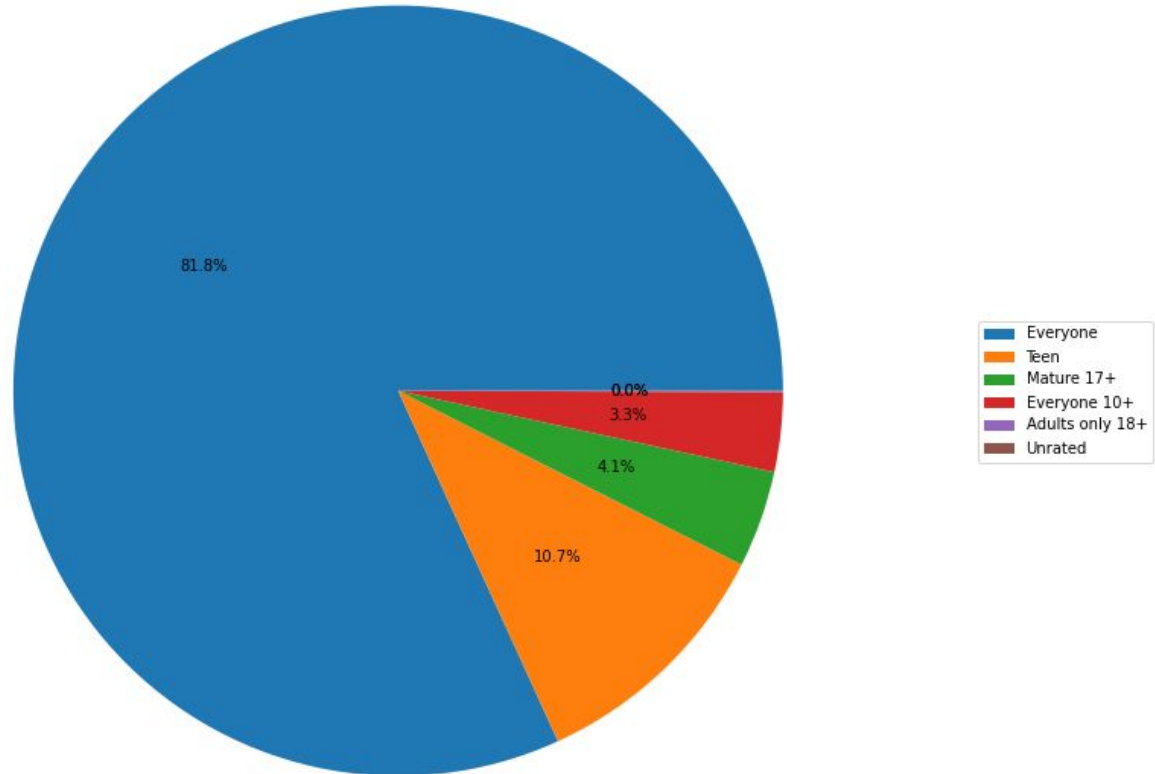
Mean Rating of each Category



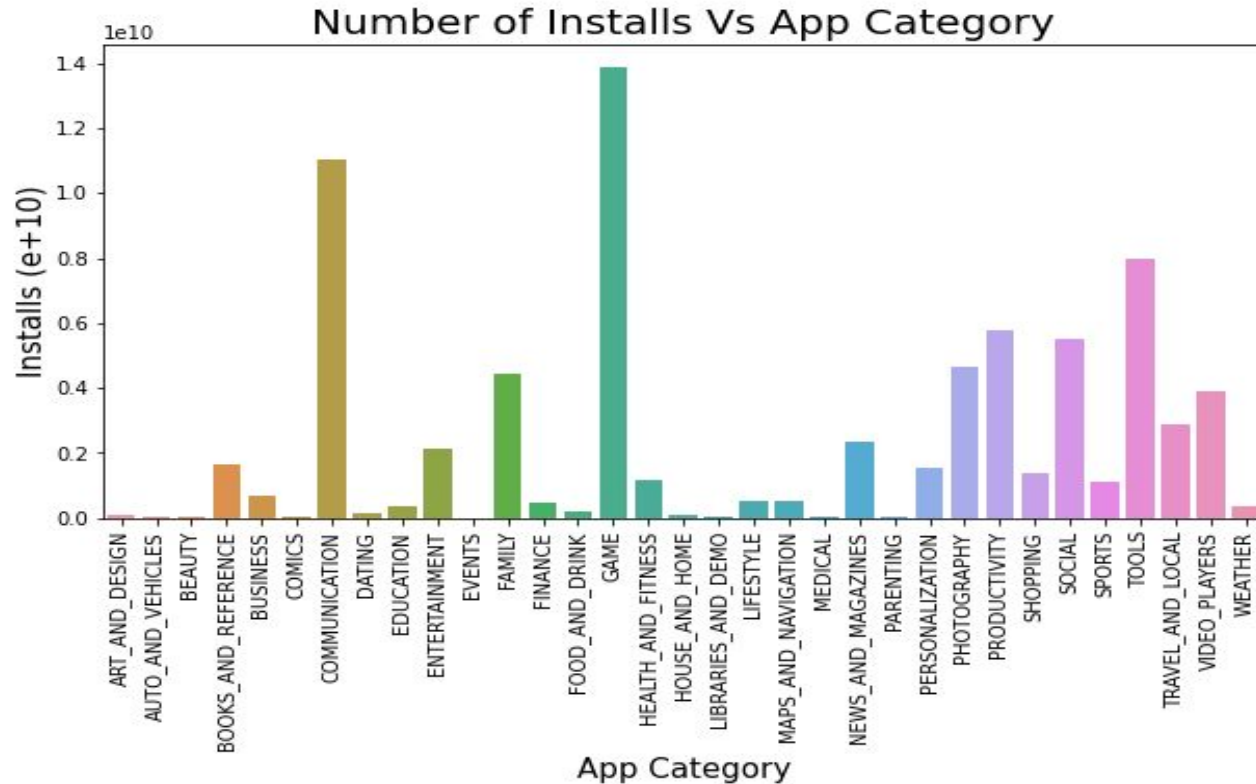
Highest Rating is for **EVENTS** category and Lowest rating for **DATING** category.

Content Ratings

From the above plot we can see that Everyone category having majority of apps count. A majority of the apps (**81.80%**) in the play store are can be used by everyone. The remaining apps have various age restrictions to use it.

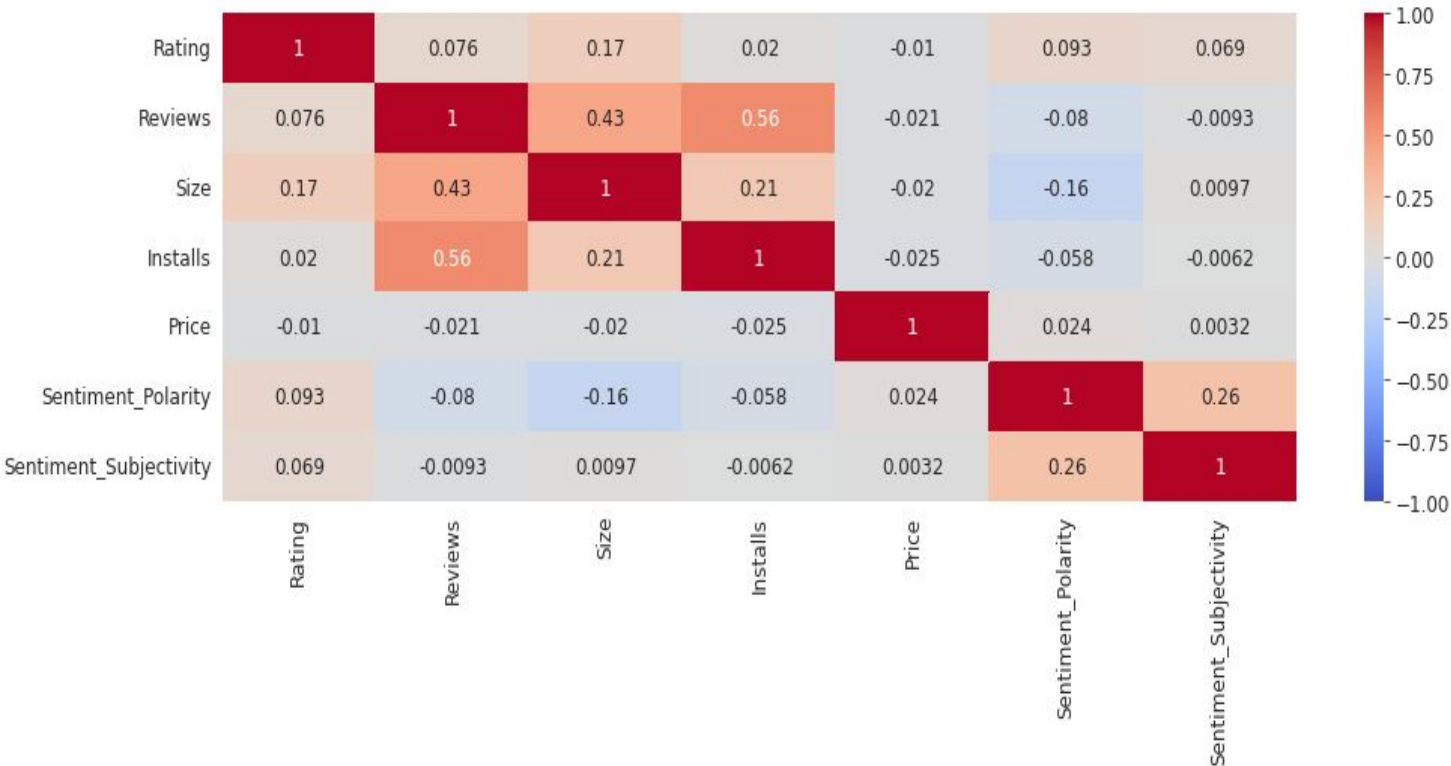


Category wise Installs of apps



Gaming have maximum number of installed app. On second Position **communication** have maximum number of apps installed.

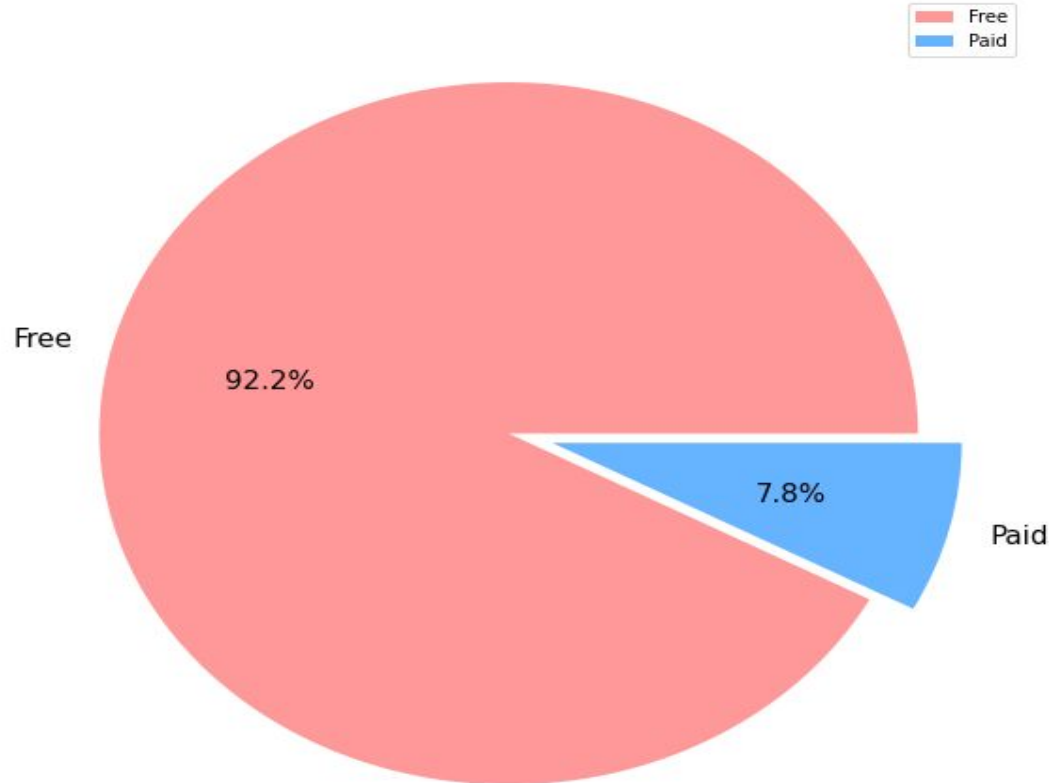
Co-Relation in Merged Dataset.



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.

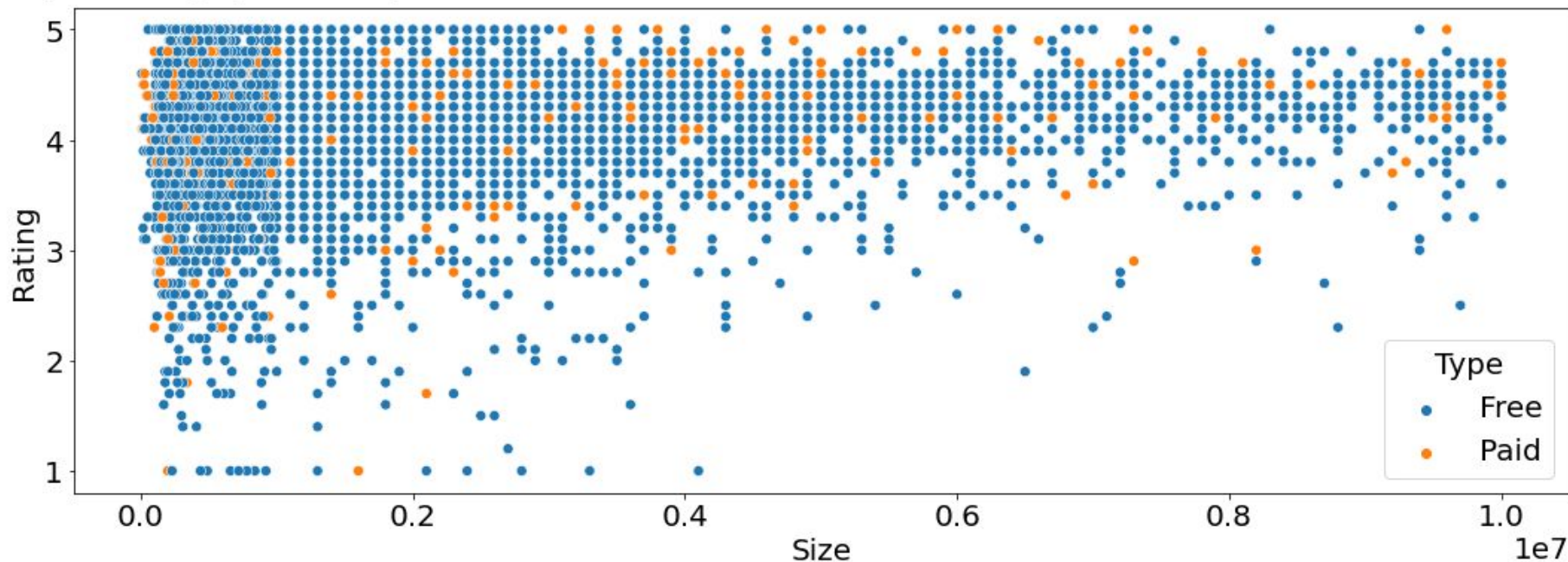
Distribution of Free and Paid apps

Available types of Paid and Free apps



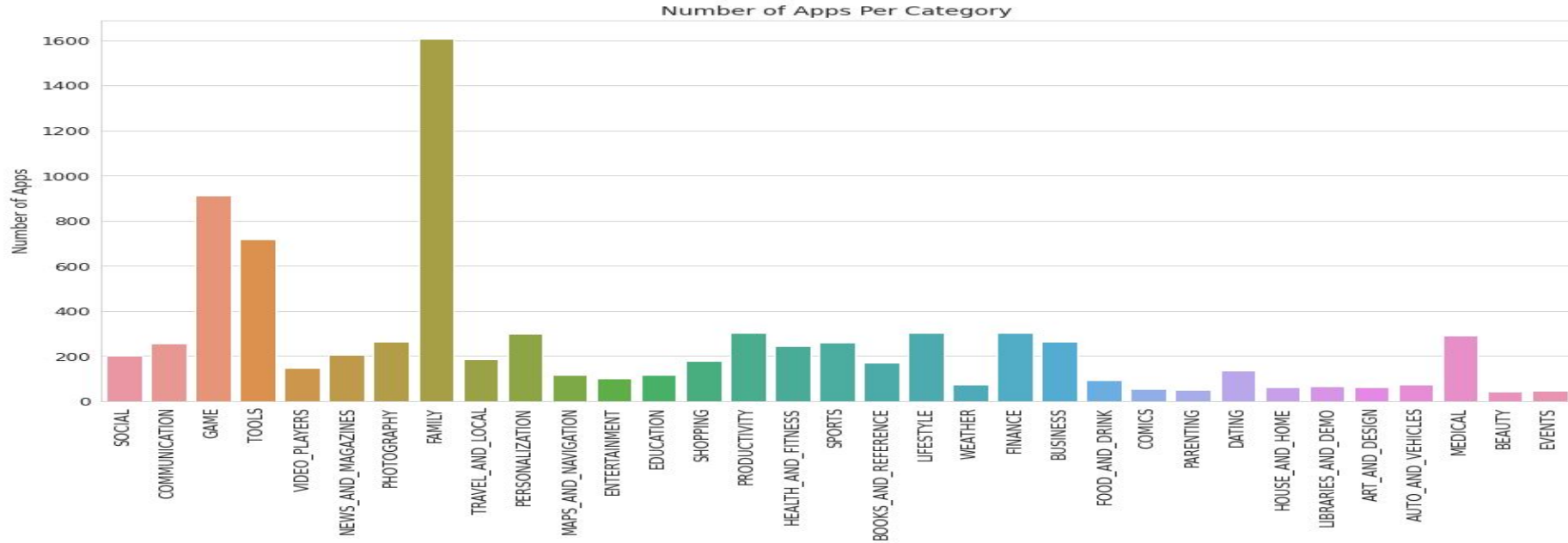
We Observed that **92.2%** of Apps are free and only **7.8%** of Apps are paid in Play store.

Relation between Rating, Size and Type of app




From this scatter plot, we can imply that majority of the **free apps** are **small in size** and having **high rating**. While for **paid apps**, we have quite **equal** distribution in term on **size and rating**.

Number of apps in Category




Maximum number of apps present in '**FAMILY**' category then '**GAME**' category and then '**TOOLS**'

Top 5 Expensive apps

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content_Rating	Genres	Last_Updated	Current_Ver	Android_Ver
4367	I'm Rich - Trump Edition	LIFESTYLE	3.6	275	730000.0	10000.0	Paid	400.00	Everyone	Lifestyle	May 3, 2018	1.0.1	4.1 and up
5364	I am rich (Most expensive app)	FINANCE	4.1	129	270000.0	1000.0	Paid	399.99	Teen	Finance	December 6, 2017	2	4.0.3 and up
4362	 I'm rich	LIFESTYLE	3.8	718	2600000.0	10000.0	Paid	399.99	Everyone	Lifestyle	March 11, 2018	1.0.0	4.4 and up
5362	I Am Rich Pro	FAMILY	4.4	201	270000.0	5000.0	Paid	399.99	Everyone	Entertainment	May 30, 2017	1.54	1.6 and up
4197	most expensive app (H)	FAMILY	4.3	6	150000.0	100.0	Paid	399.99	Everyone	Entertainment	July 16, 2018	1.0	7.0 and up

From above data we can conclude that-

The top 5 most expensive apps in the store are I'm Rich - Trump Edition, I am rich (Most expensive app),  I'm rich, I Am Rich Pro, most expensive app (H).

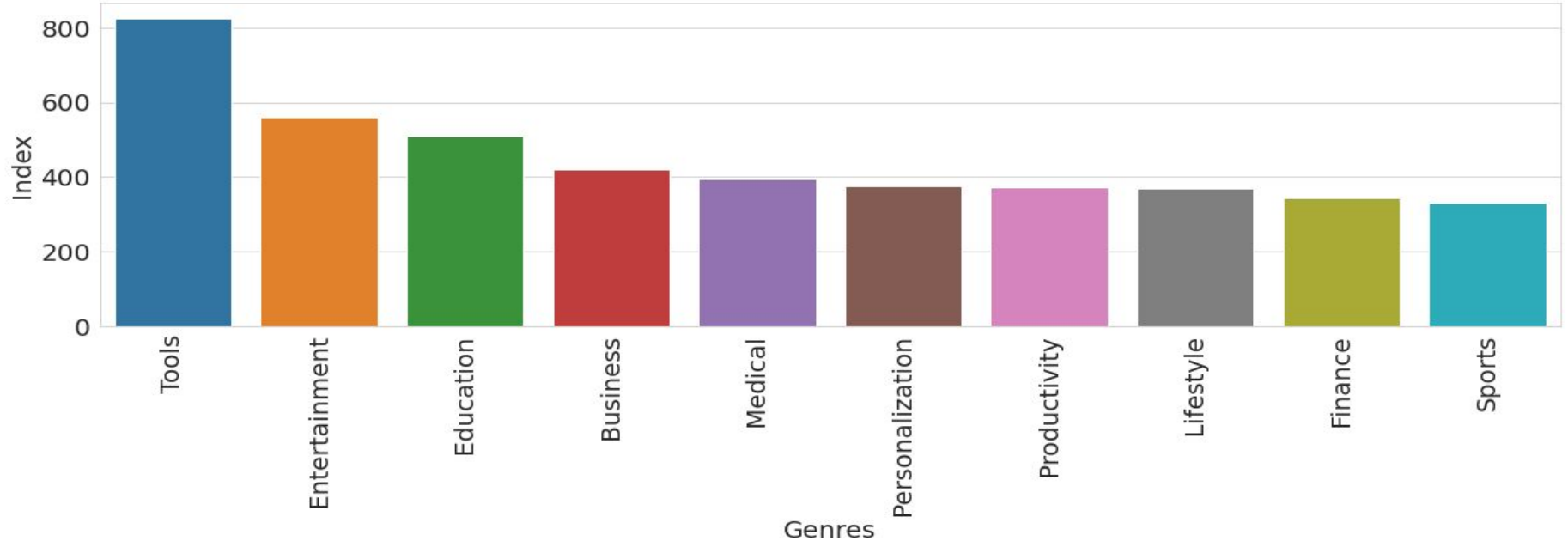
Top 5 Reviewed Apps

	App	Category	Rating	Reviews
3432	Facebook	SOCIAL	4.1	78158306
7665	WhatsApp Messenger	COMMUNICATION	4.4	69119316
4422	Instagram	SOCIAL	4.5	66577313
5096	Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56642847
2073	Clash of Clans	GAME	4.6	44891723

From the above data we can conclude that-

The 5 apps that have the most number of total reviews are Facebook, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free, Clash of Clans.

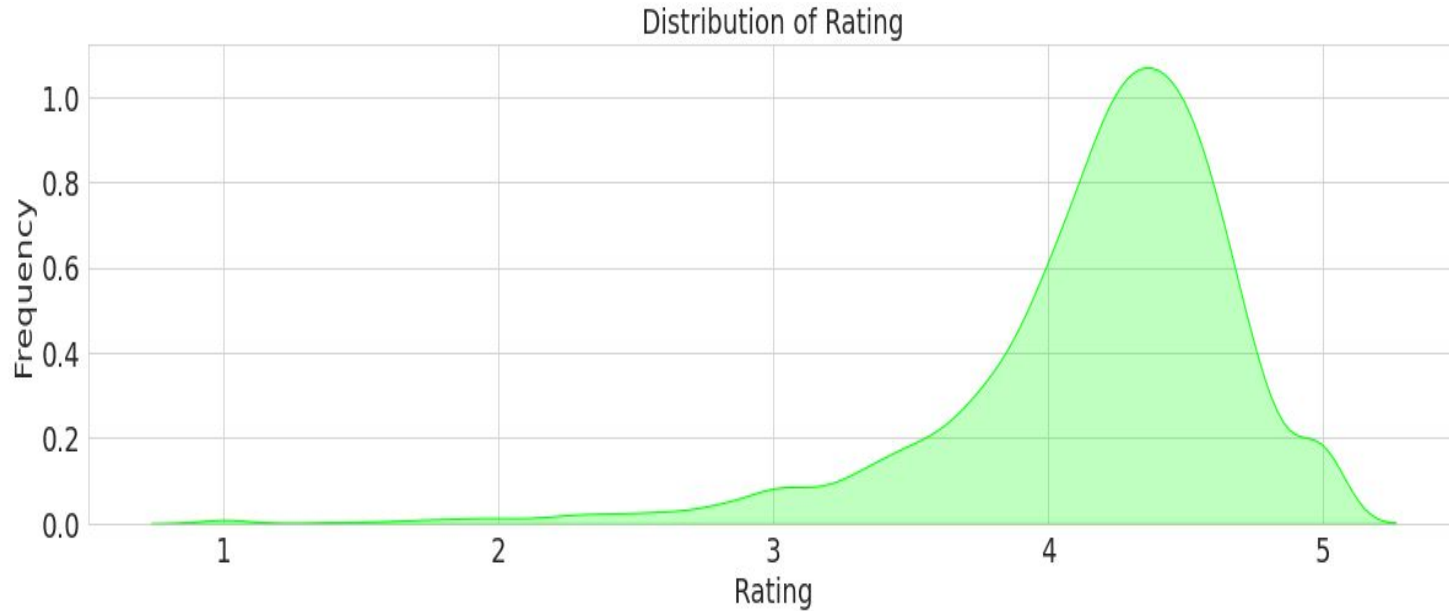
Top 10 Genres of App



Top three Genres are **Tools, Entertainment & Education** .

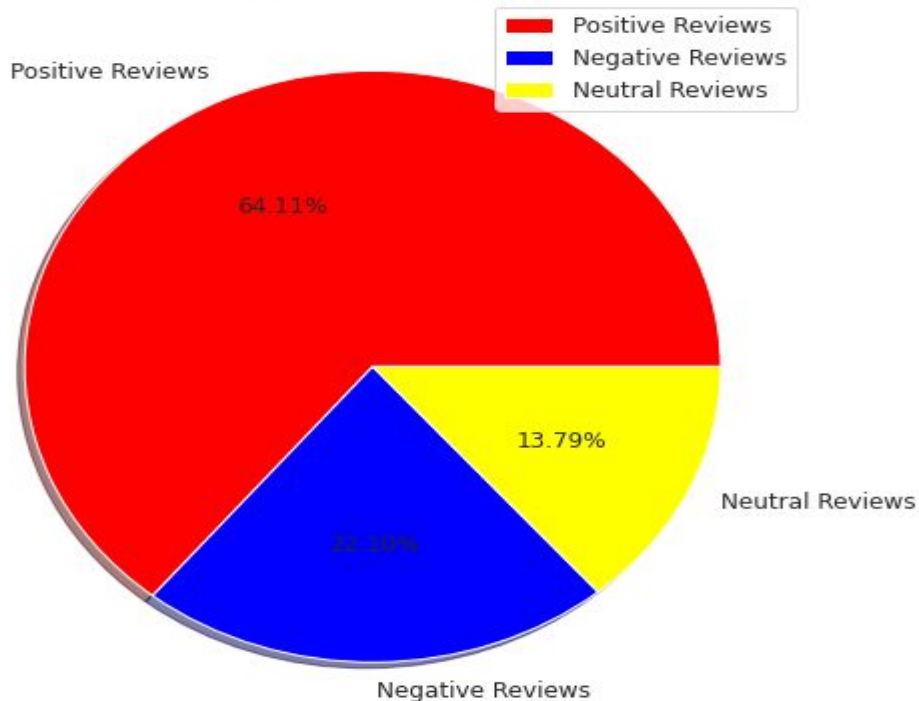
Distribution of Rating

From this distribution plotting, it implies that most of the apps in the Play Store are having rating higher than 4 or in the range of 4 to 4.7



User Sentiment Analysis

A Pie Chart Representing Percentage of Review Sentiments



From Sentiment column, **64% are Positive**, **22% are Negative** and **14% are Neutral** Reviews. Positive Sentiment have larger percentage, it has **3 times more** than negative sentiment.

Challenges

- Computation time.
- Reading the dataset and comprehending the problem statement. Our major challenge was data cleaning.
- Also dataset contains duplicate data and irrelevant data, need to clear that first was also problematic.



Conclusion

- Rating is very important factor for installation of apps as user mostly like to watch rating before using app, so developer should also work on updating their content as per the ratings.
- From all above we analyze rating and installation are related, so owners should encourage to write review of their app.
- Gaming have maximum number of installed app so anyone want to become developer can join this category.
- App category like events and beauty have not much reach, so one can also keep this consideration.
- Most of the apps are downloaded by teens, so users of other age category, must also be encouraged to install the apps.

Conclusion

- Maximum number of apps present in FAMILY category then GAME category and then TOOLS
- Highest Mean price is for FINANCE category and after that LIFESTYLE category.
- Highest Rating is for EVENTS category and Lowest rating for DATING category.
- 81.8% app can use every age group person.
- Adults only 18+ and unrated have very less number of apps.
- Gaming have maximum number of installed app.
- Around 92.2% app are freely available on play store.
- 4.3 to 4.5 have maximum average rating on play store.
- Positive Sentiment have larger percentage, it has 3 times more than negative sentiment.

