Capstone Project
Play Store App review Analysis

By
Mr. Vinay V. Kulkarni

Outline of Presentation

- Summary of Dataset

- Packages utilized

- Data Cleaning

- Data Preparation

- Exploratory Data Analysis (EDA)

- Questions answered with EDA

- Conclusions

Data Summary

- 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.

- There are two Datasets provided for EDA of Play Store Apps.

- The Fist Dataset named "**Play Store Data**" consists of 13 columns:- App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current Ver., and Android Ver. with 10841 Rows.

- The second dataset named "**User Reviews**" consist of 5 columns:- 'App', 'Translated_Review', 'Sentiment', 'Sentiment_Polarity', 'Sentiment_Subjectivity'

Packages Required

- Following packages are utilized in analysis of the Play Store apps:

```python
# Imports
import pandas as pd
import numpy as np

import warnings
warnings.filterwarnings('ignore')
```

Data Cleaning

- After doing dataset study, it is observed that few columns have **NaN** values which need to removed.

```
App                  0
Category             0
Rating            1474
Reviews              0
Size                 0
Installs             0
Type                 1
Price                0
Content Rating       1
Genres               0
Last Updated         0
Current Ver          8
Android Ver          3
dtype: int64
```

After Cleaning the Dataset

```
printinfo()
```

	data_type	null_count	unique_count
App	object	0	9659
Category	object	0	33
Rating	float64	0	39
Reviews	object	0	6001
Size	object	0	461
Installs	object	0	21
Type	object	0	2
Price	object	0	92
Content Rating	object	0	6
Genres	object	0	119

Data Preparation


```
printinfo()
```

	data_type	null_count	unique_count
App	object	0	9659
Category	object	0	33
Rating	float64	0	39
Reviews	object	0	6001
Size	object	0	461
Installs	object	0	21
Type	object	0	2
Price	object	0	92
Content Rating	object	0	6
Genres	object	0	119

```
printinfo()
```

	data_type	null_count	unique_count
App	object	0	8434
Category	object	0	33
Rating	float64	0	39
Reviews	int64	0	4680
Size	float64	0	459
Installs	int64	0	20
Type	object	0	2
Price	float64	0	87
Content Rating	object	0	6
Genres	object	0	116

Number of Apps Vs Category

Apps available on play store for multiple purposes
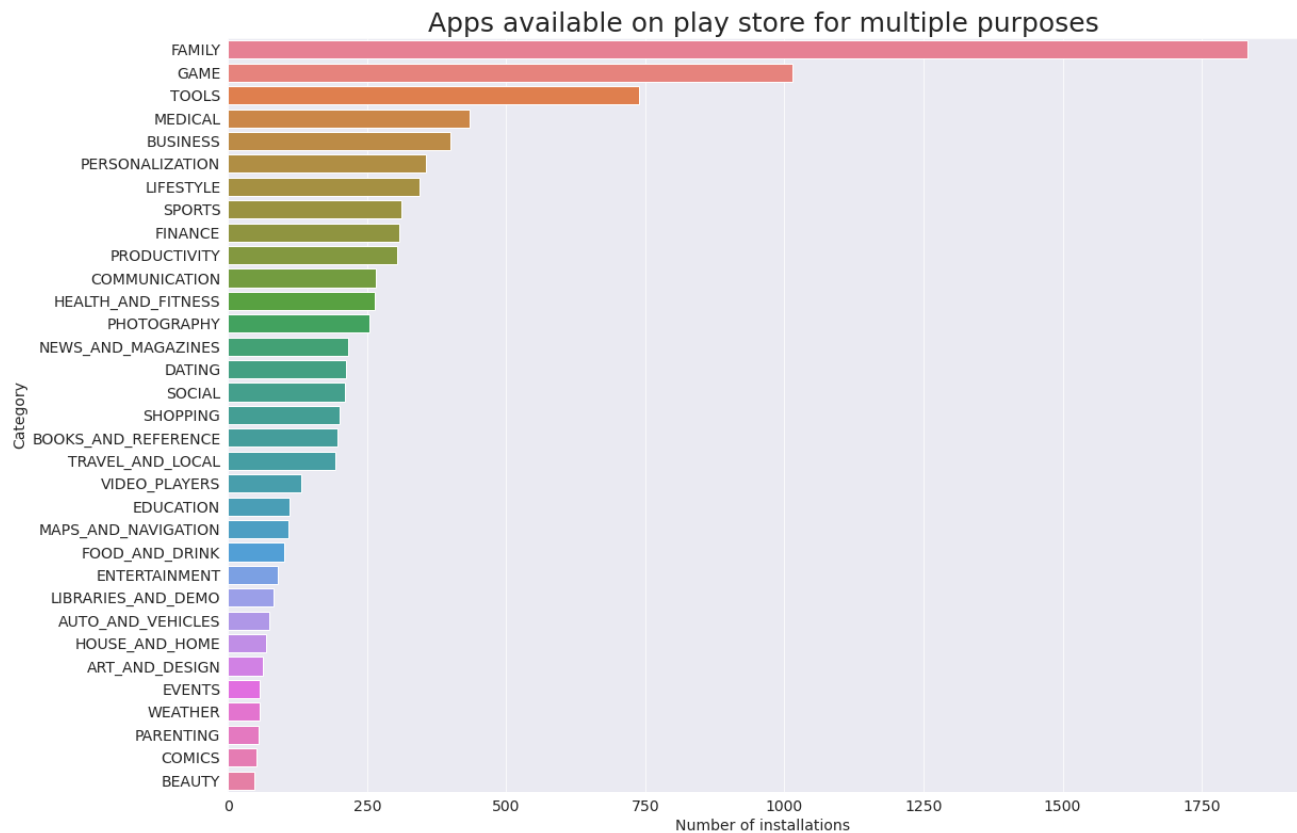
Content Rating Vs App Count
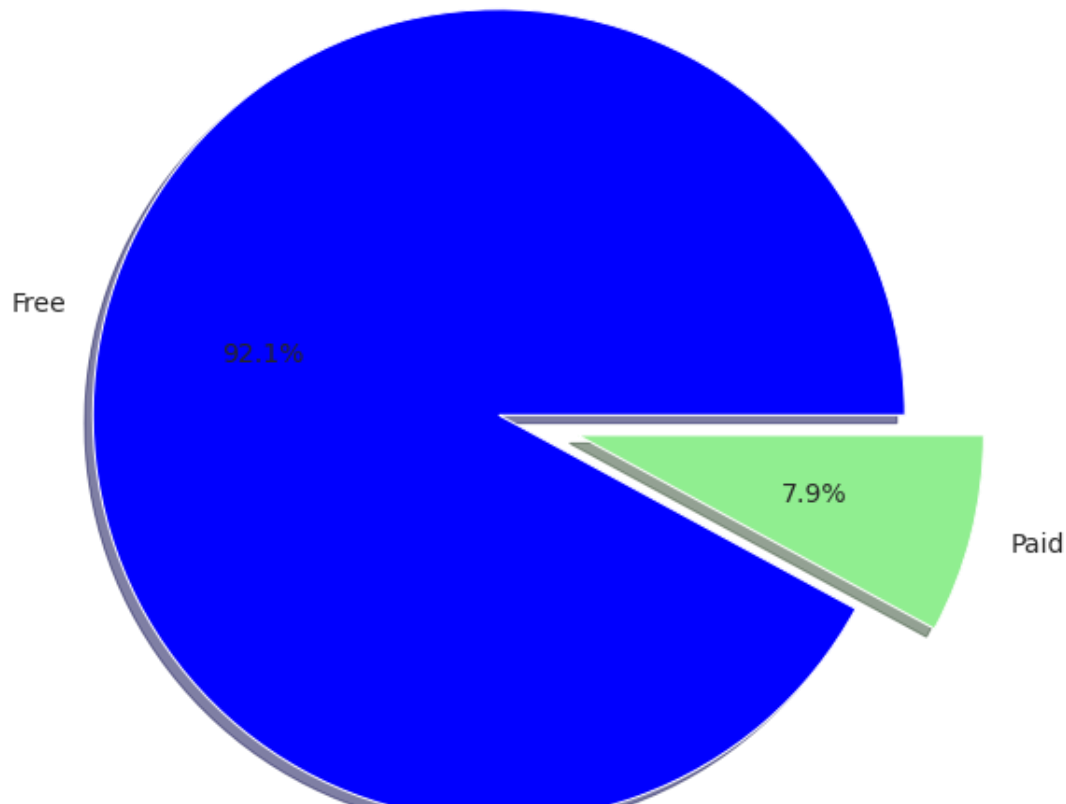
Distribution of Rating
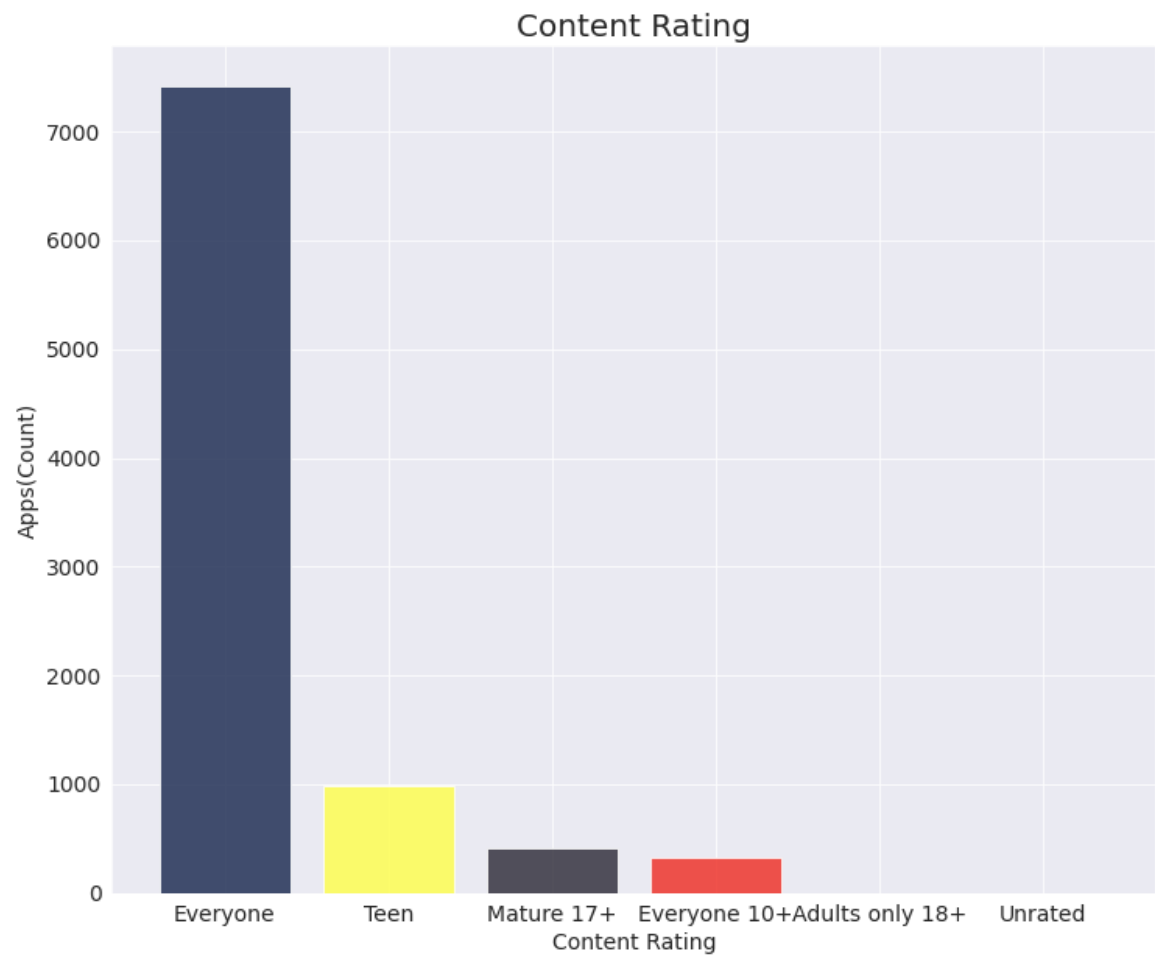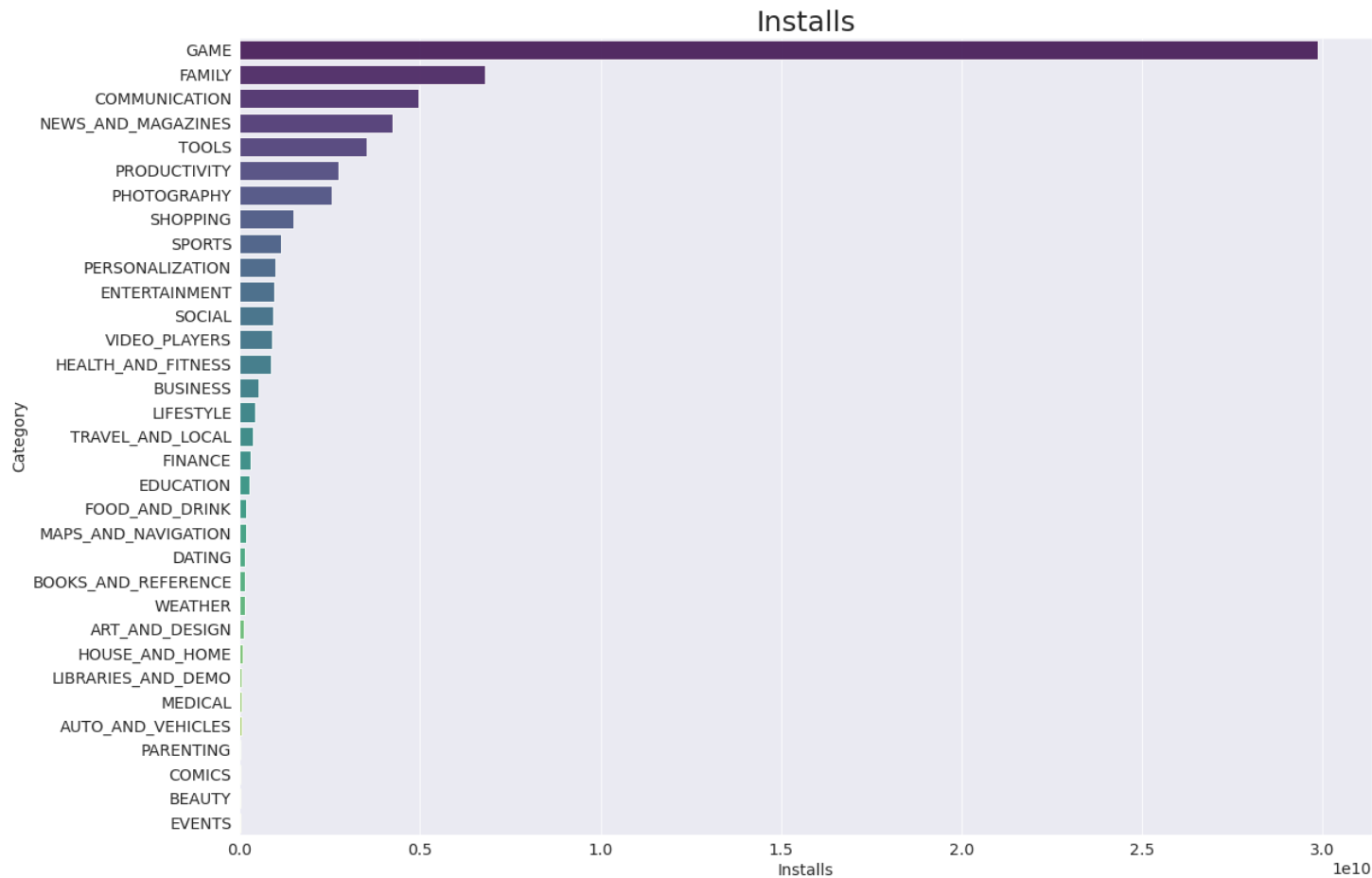

Distribution of Rating
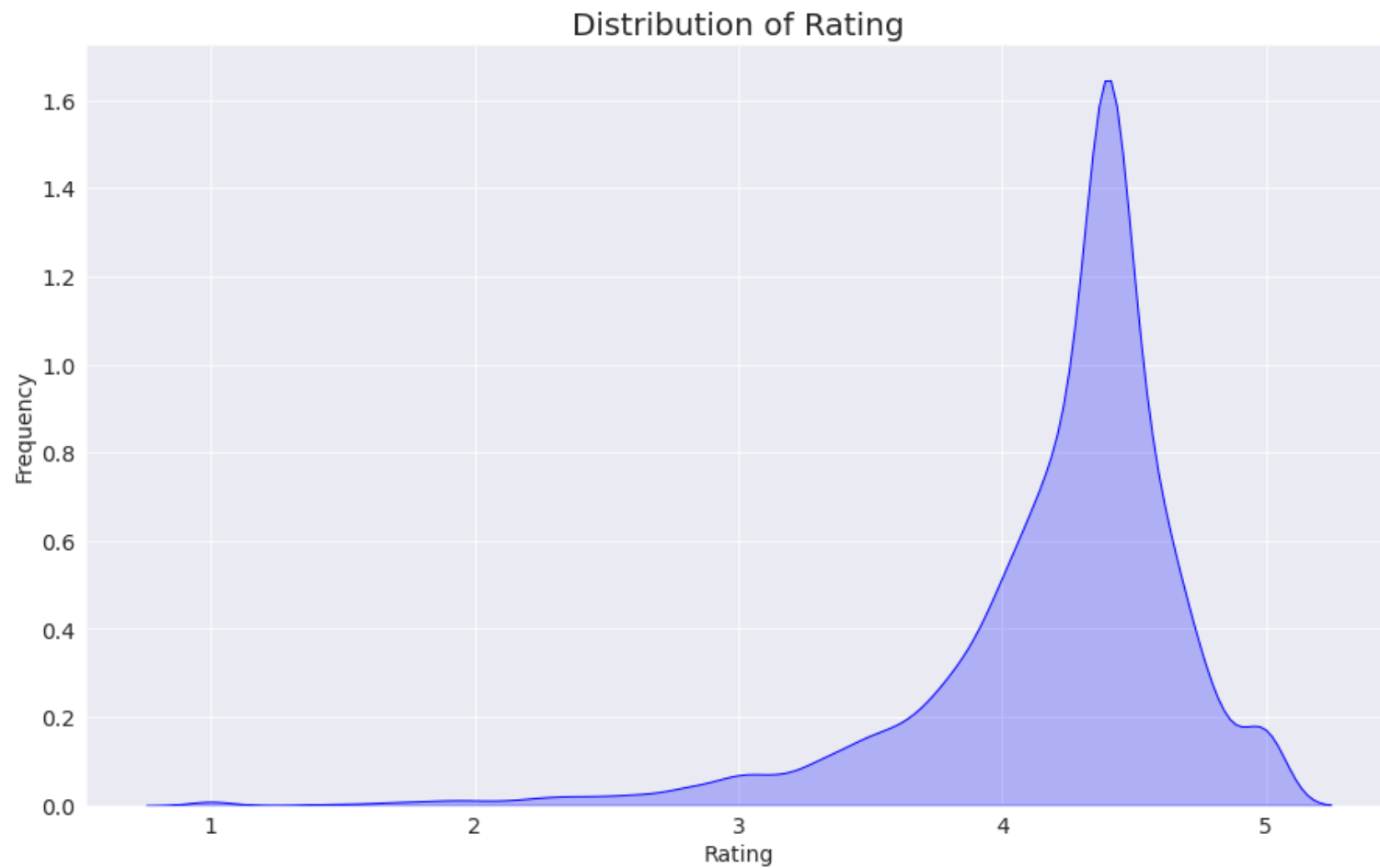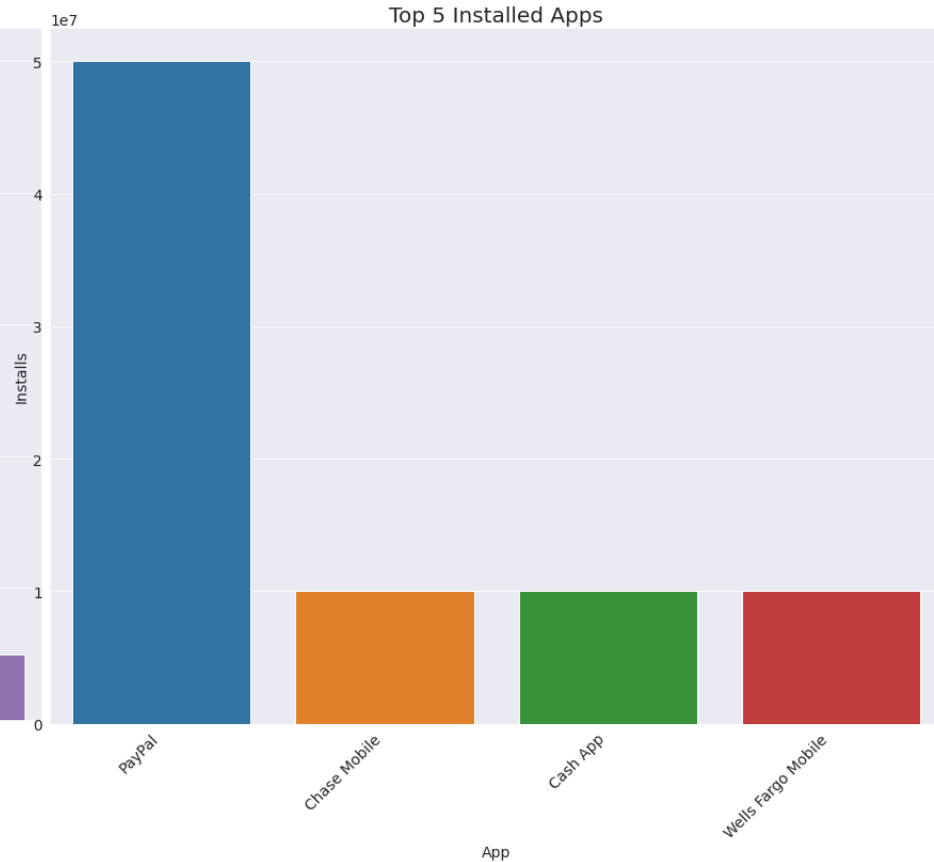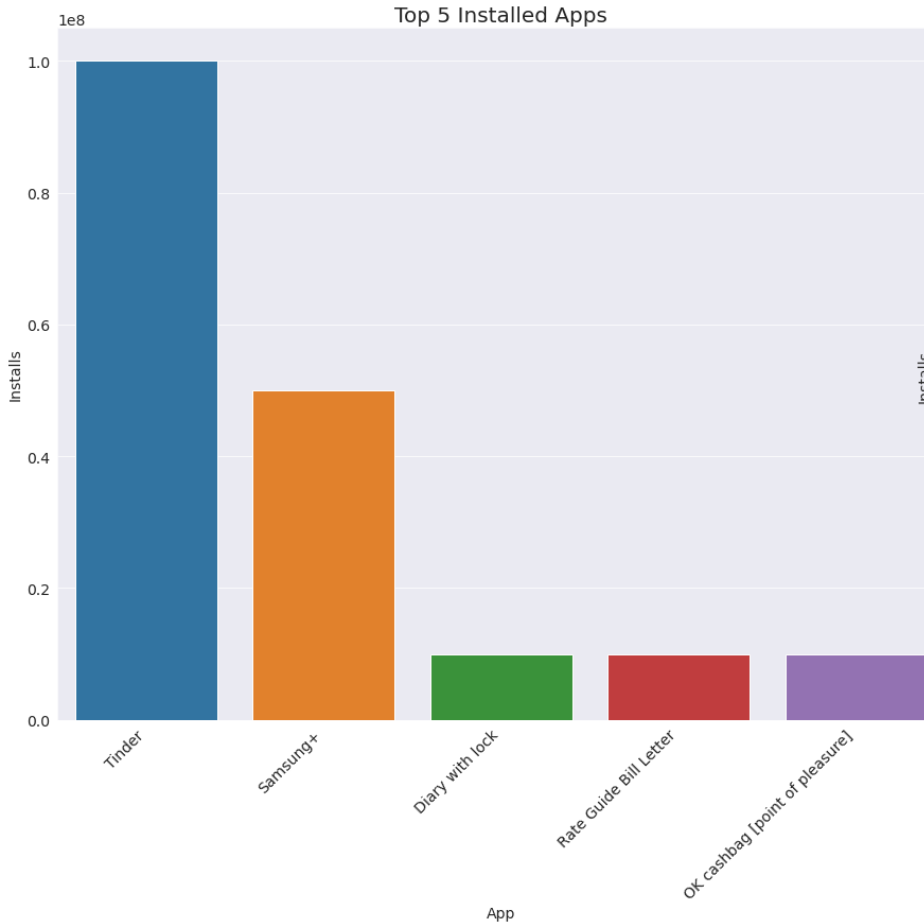
Paid Vs Free Apps in Play Store

Percent of Free Vs Paid Apps in store
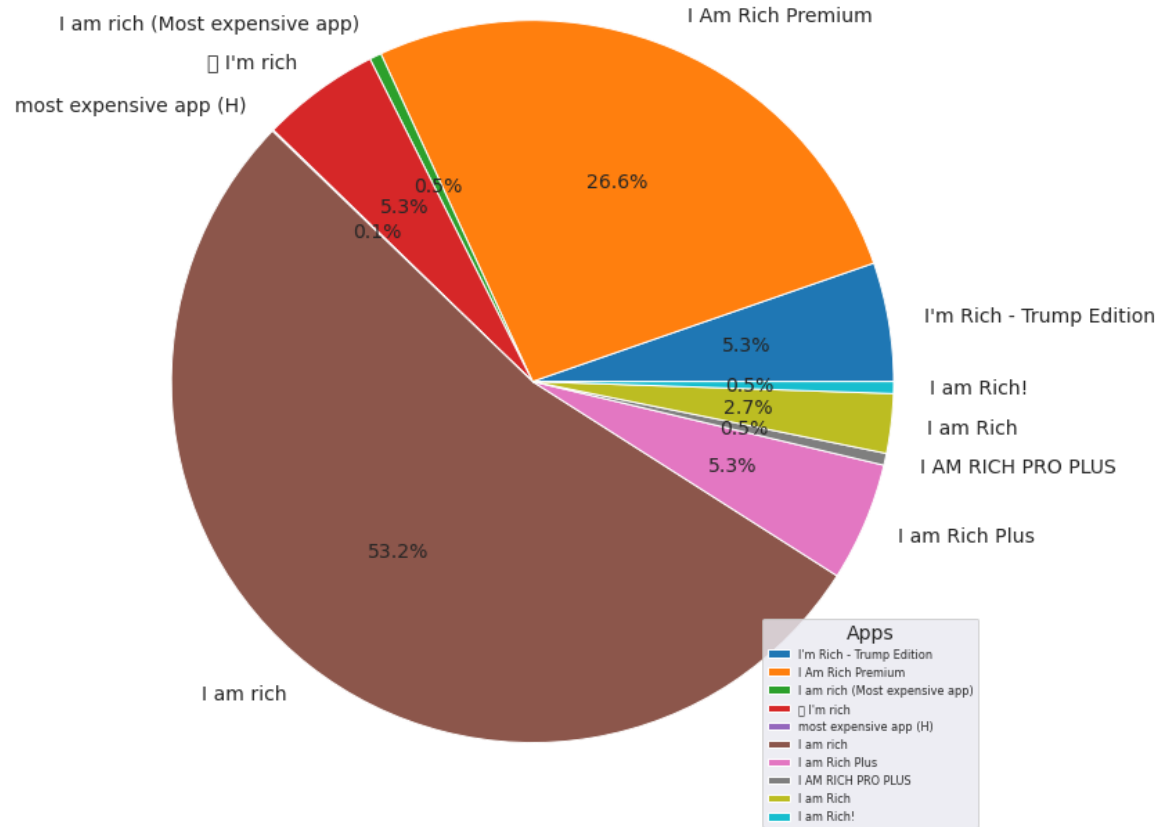
Category Vs Number of Installs

Find top 5 installed apps (in any category)

Find Top 10 Expensive App


Top Expensive Apps Distribution

Find Apps with highest number of Reviews

	App	Category	Rating	Reviews
1879	Clash of Clans	GAME	4.6	44893888
1704	Clash of Clans	GAME	4.6	44891723
1670	Clash of Clans	GAME	4.6	44891723
3986	Clash of Clans	FAMILY	4.6	44881447
1872	Subway Surfers	GAME	4.5	27725352
1917	Subway Surfers	GAME	4.5	27725352
1750	Subway Surfers	GAME	4.5	27724094
1700	Subway Surfers	GAME	4.5	27723193
1654	Subway Surfers	GAME	4.5	27722264
3896	Subway Surfers	GAME	4.5	27711703

Find Count of Apps in different genres

Top Genres in the Playstore
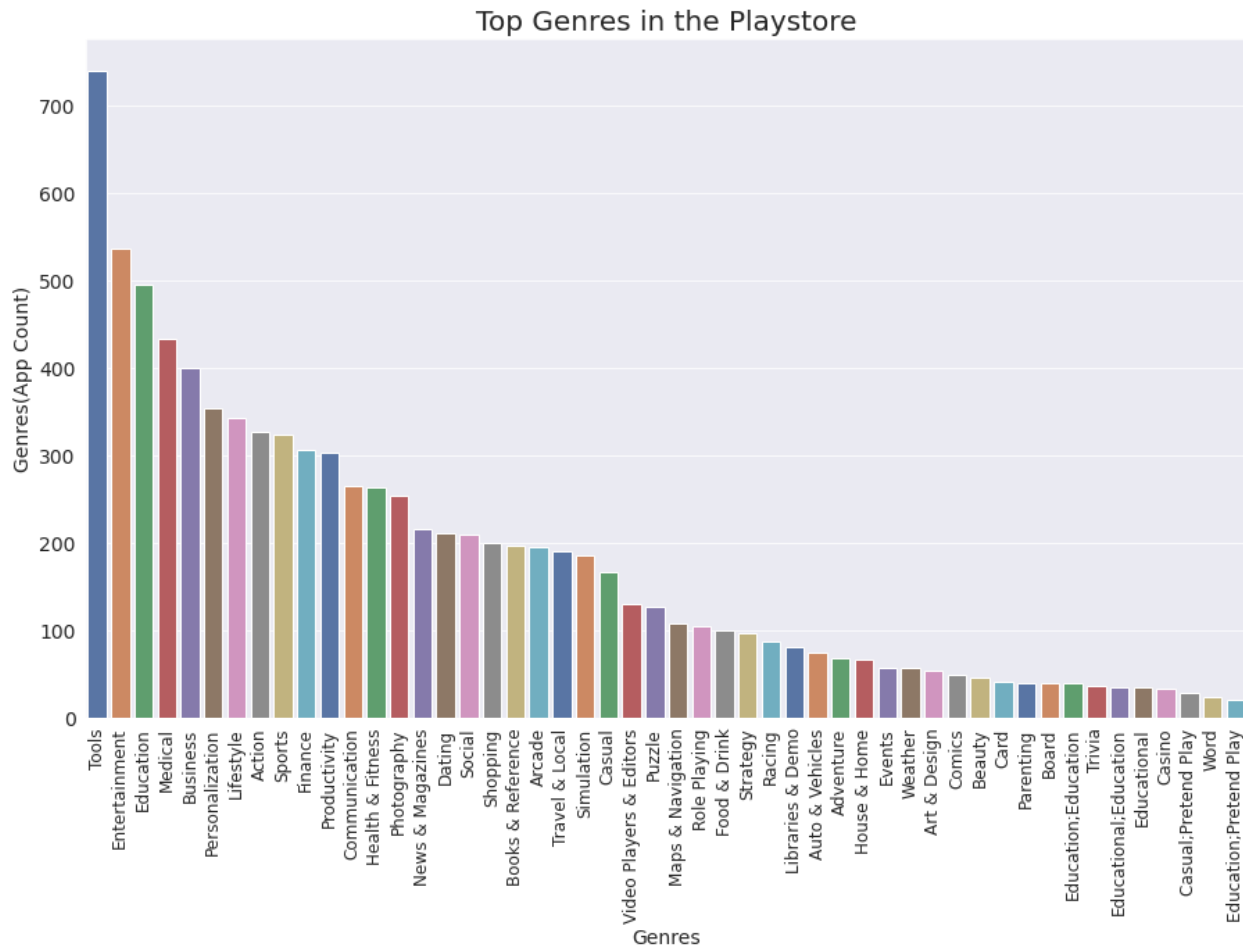
Highest Money making Apps

Top Earning Apps
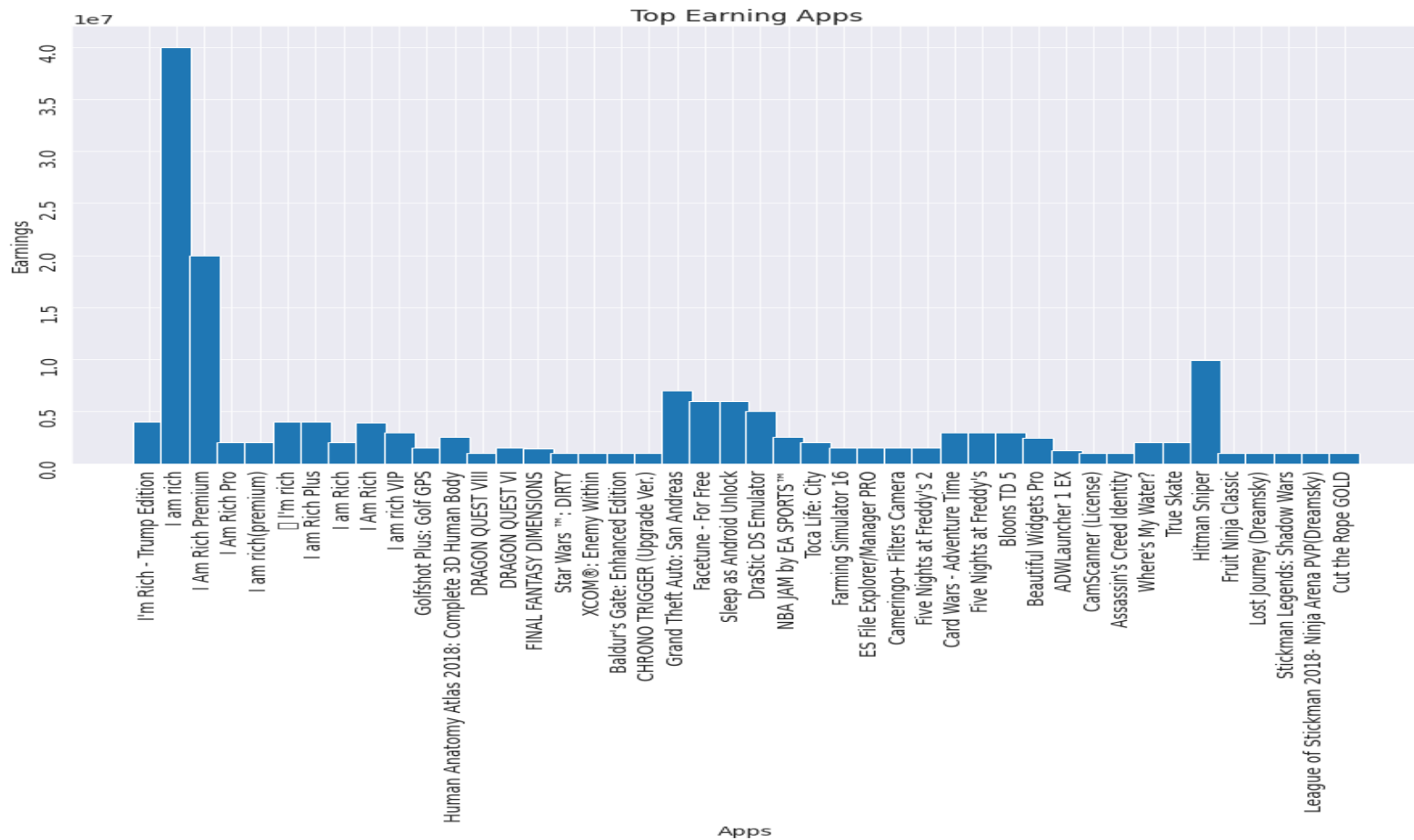
Data Analysis of "Review Dataset"

Analysis of Review Dataset

- Here also the raw data need to be processed and remove Null values from dataset.

App		0
Translated_Review		26868
Sentiment		26863
Sentiment_Polarity		26863
Sentiment_Subjectivity		26863
dtype: int64		

App		0
Translated_Review		0
Sentiment		0
Sentiment_Polarity		0
Sentiment_Subjectivity		0
dtype: int64		

Before preprocessing   After Preprocessing

Review Sentiment Analysis

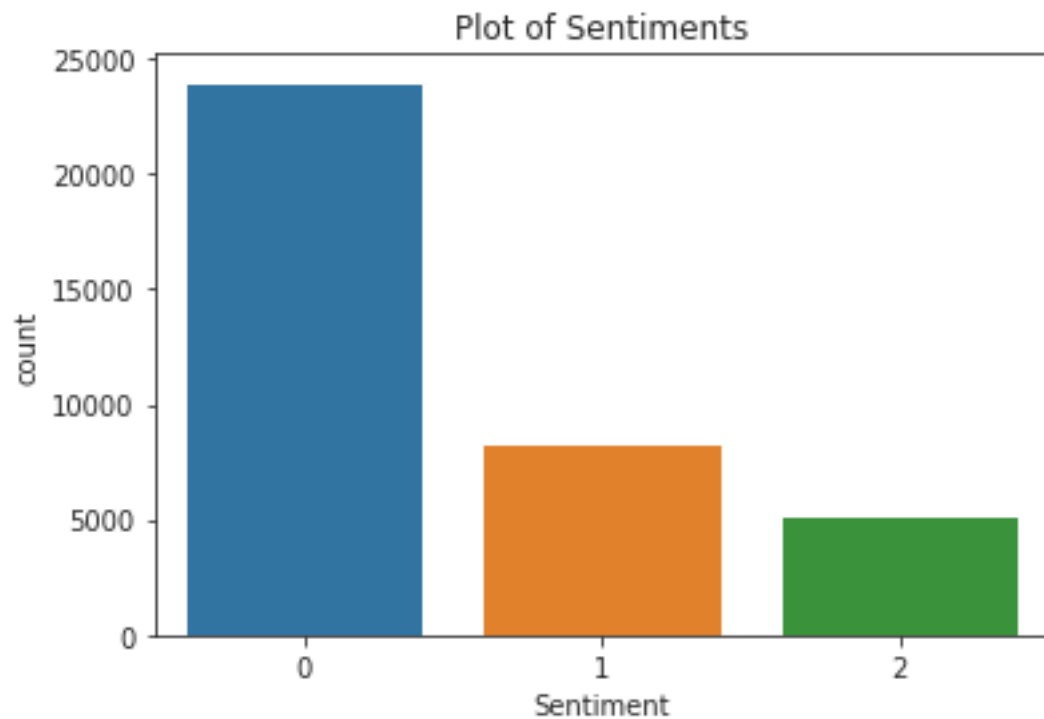Google Play Store Reviews Sentiment Analysis

Sentiment Polarity Distribution
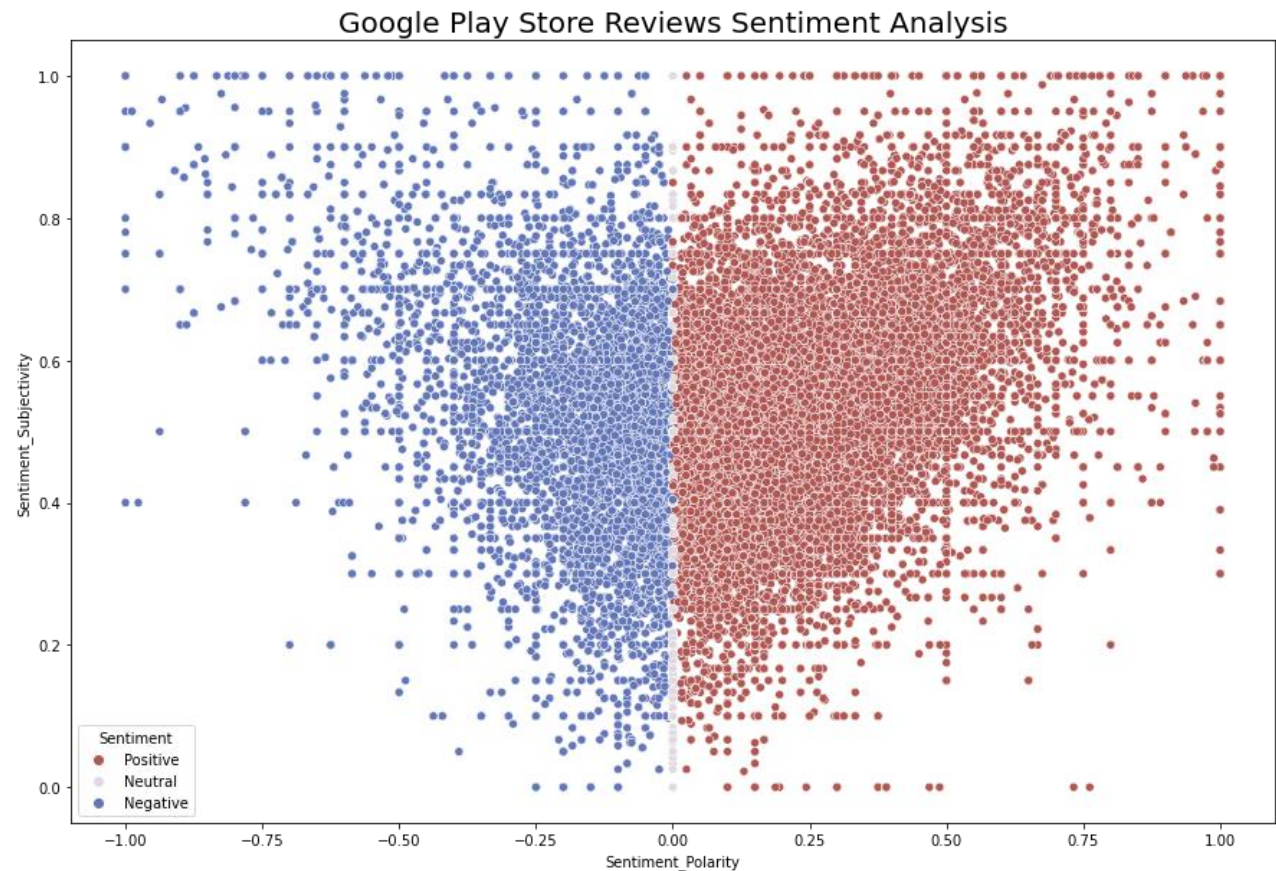
Sentiment Polarity Distribution
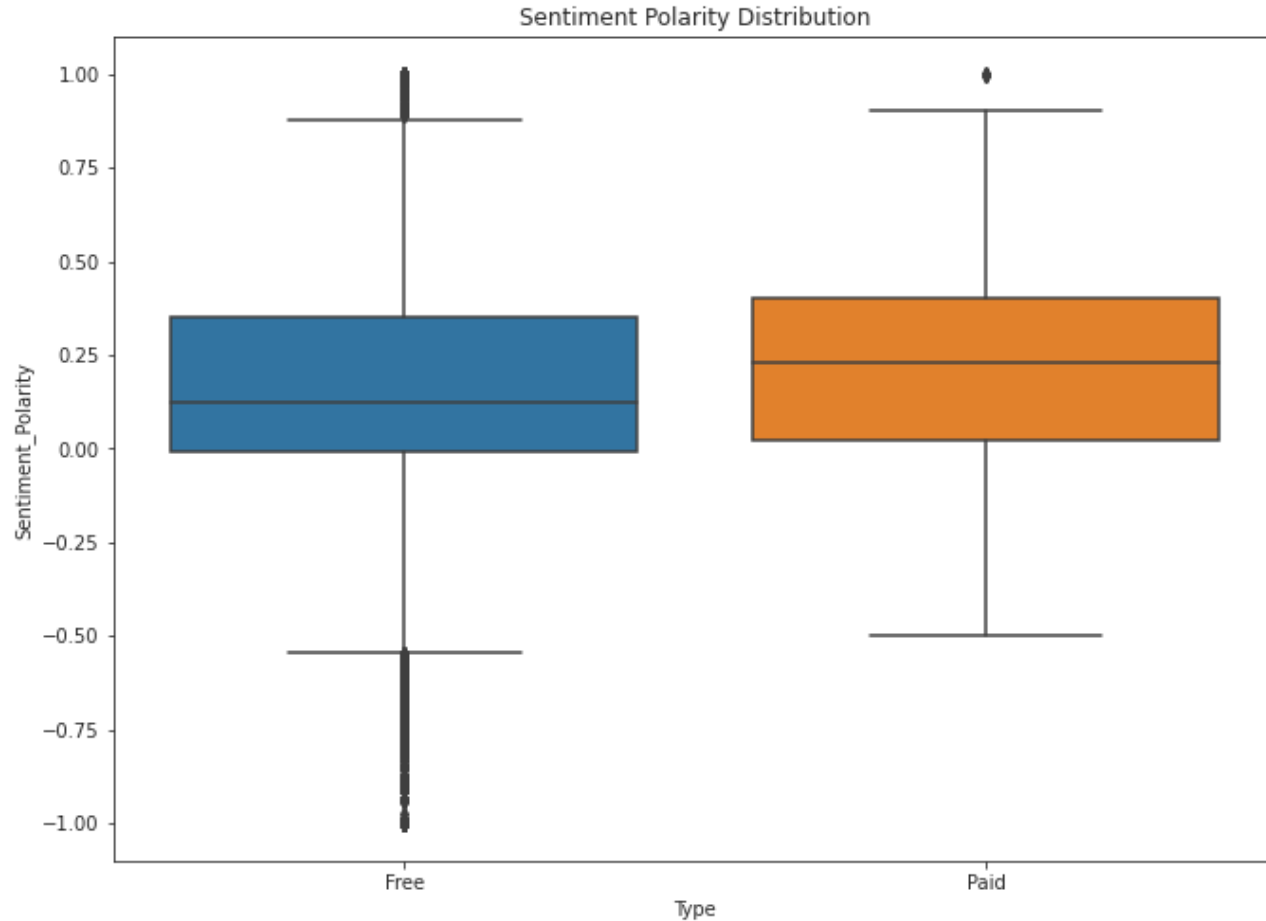
Sentiment Count

Conclusions

- The Play Store Dataset contains many NaN values, mixed multiple datatypes. Hence, it requires preprocessing and preparation part for clean and detailed data analysis.

- The third party packages like Pandas, Numpy helped to reveal unknown information from the both the dataset.

- Most of the apps are from category Family with average rating of 4.3/5 but still around 92% of the apps available on play store are free.

- The analysis of top installed apps with a particular category brings few interesting facts like Paypal app is mostly used in finance category, Tinder app is mostly used to connect socially.

Conclusions

- The app named "I am rich" is making most profit for play store compared to other apps. But "Clash of Clans" Game is having highest reviews and rating.

- The review dataset gives the genuine comments from users that gives the mostly positive reviews about apps in play store.