Capstone Project-1

Play Store App review analysis

Member

Sourabh Pramanik

Contents

Problem Statement

Here in this project we have two datasets-

1. Play Store App Data
2. User Review Data

In today's world there are lots of number of applications present in Play Store.

In this project we will try find some conclusion like The popularity of applications, The type of applications developers should develop etc.

Data Summary

This project contains two Datasets-

1. Play Store Dataset
2. User Review Dataset

We will complete this project by using following steps-

- After reading the data we will do the Data Cleaning operation.
- We will perform the some statistical operation on our dataset.
- We will do the exploratory analysis and data visualization.
- After all the analysis we will come to the conclusion.

Data Summary

Outcomes of this Project -

- Installation of application by users according to the categories.
- The mostly demanded applications in playstore.
- Factors which affect the installation of application by the user.

Dependent Variables

Play store csv file -

- Category
- Ratings
- Reviews
- Size
- Type
- price

Dependent Variables

User review csv file -

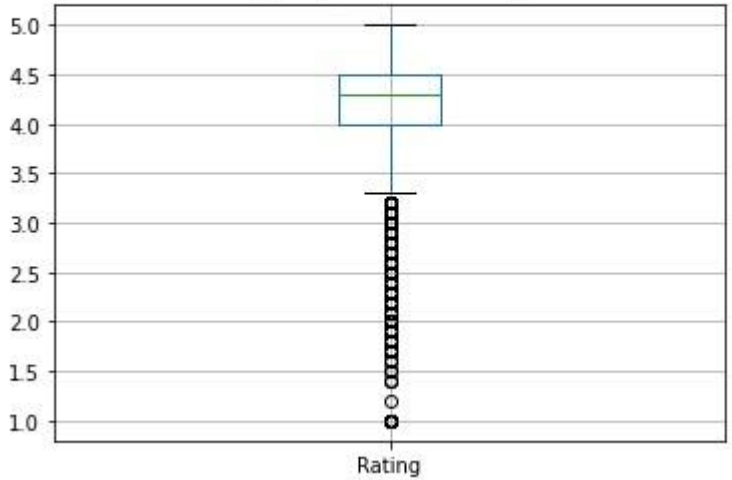
- Sentiment
- Sentiment Subjectivity

Data Cleaning

Ratings

```
[ ]  Play_store_data_df.boxplot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f02569752d0>
```



Data Cleaning

Null Values

```
Play_store_data_df.isnull().sum()
```

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

Data Cleaning

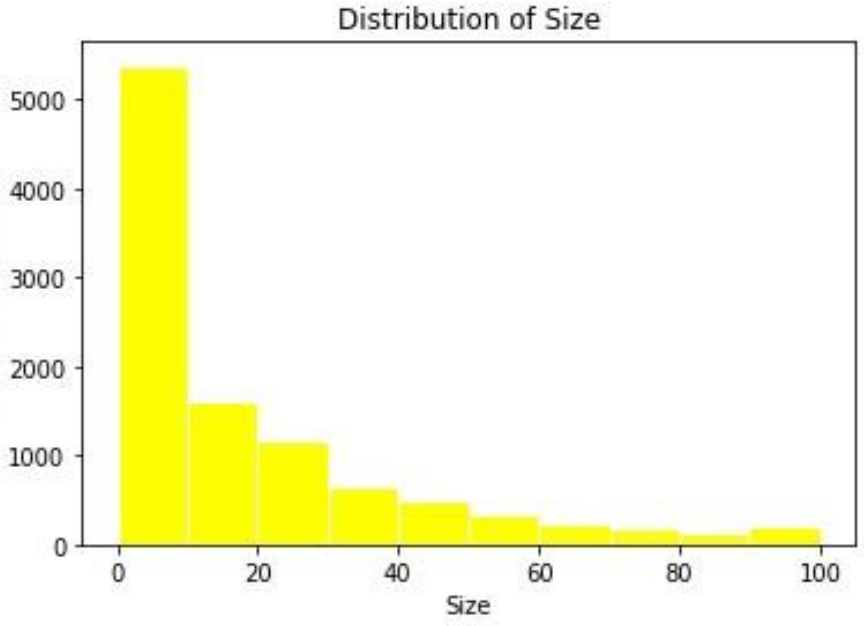
```
[ ] Play_store_data_df[Play_store_data_df['App']=='Candy Crush Saga']
```

Removal of duplicate Data

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
1655	Candy Crush Saga	GAME	4.4	22426677	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
1705	Candy Crush Saga	GAME	4.4	22428456	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
1751	Candy Crush Saga	GAME	4.4	22428456	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
1842	Candy Crush Saga	GAME	4.4	22429716	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
1869	Candy Crush Saga	GAME	4.4	22430188	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
1966	Candy Crush Saga	GAME	4.4	22430188	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up
3994	Candy Crush Saga	FAMILY	4.4	22419455	74.0	500000000	Free	0.0	Everyone	Casual	July 5, 2018	1.129.0.2	4.1 and up

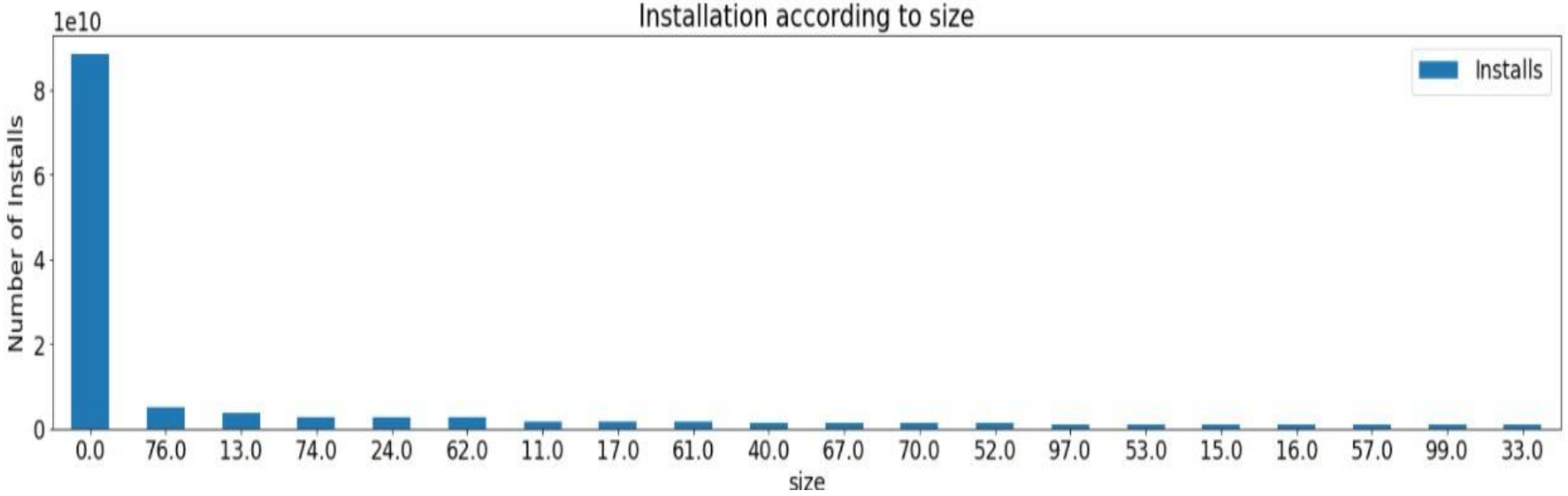
Data Visualization

Most of the application's size between 0 to 40.



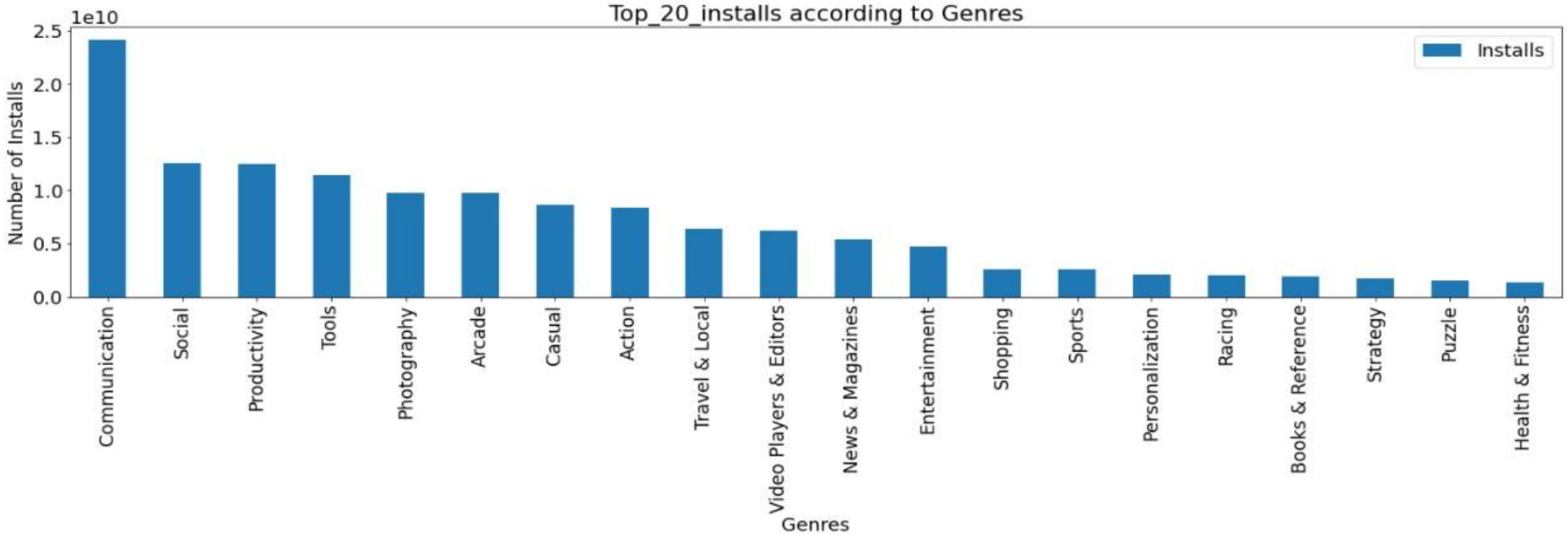
Data Visualization

According to the previous data we can see that most of the applications installed are of small sized application.



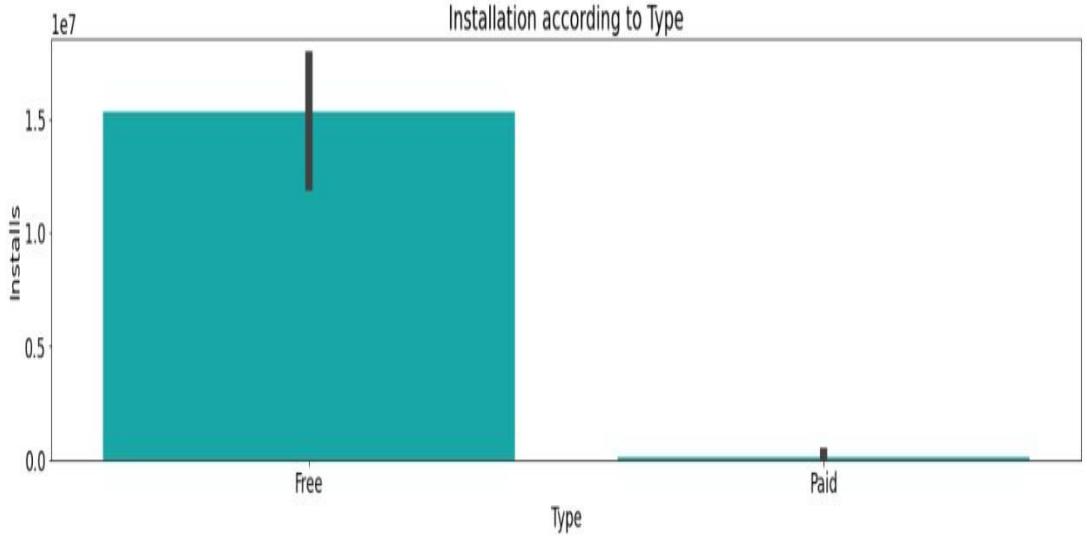
Data Visualization

The most popular applications are under communication genres



Data Visualization

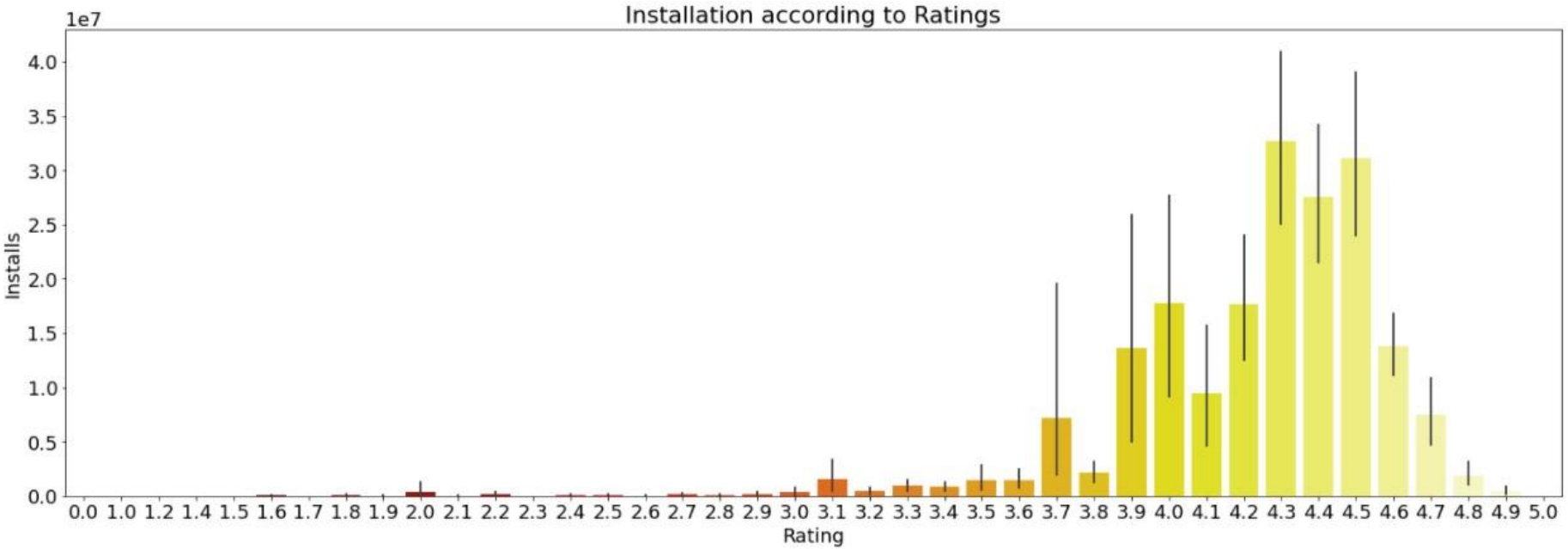
Apps installation according to Type -



Free applications installation graph is high.

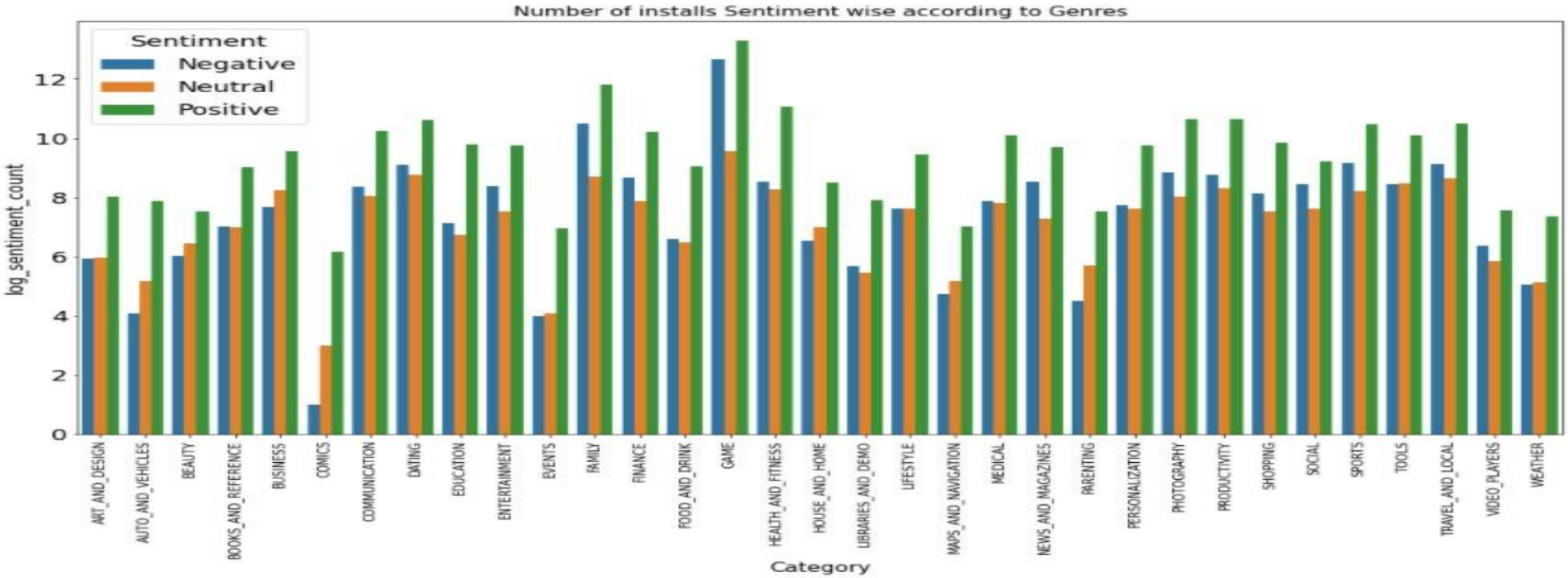
Data Visualization

User prefer the highest rating applications



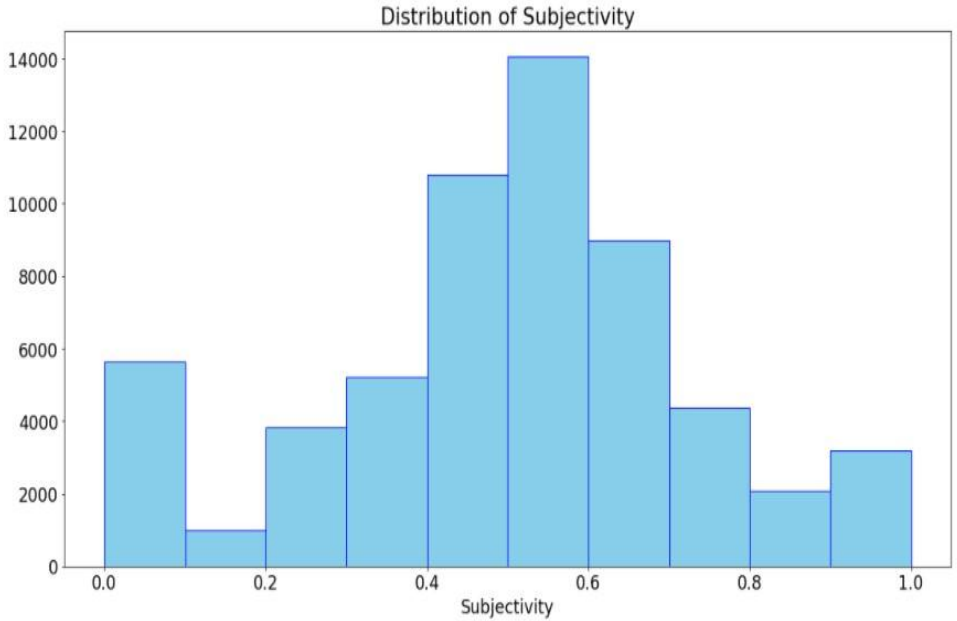
Data Visualization

Types of reviews category wise -



Data Visualization

Subjectivity - Personal opinion of the user about the application



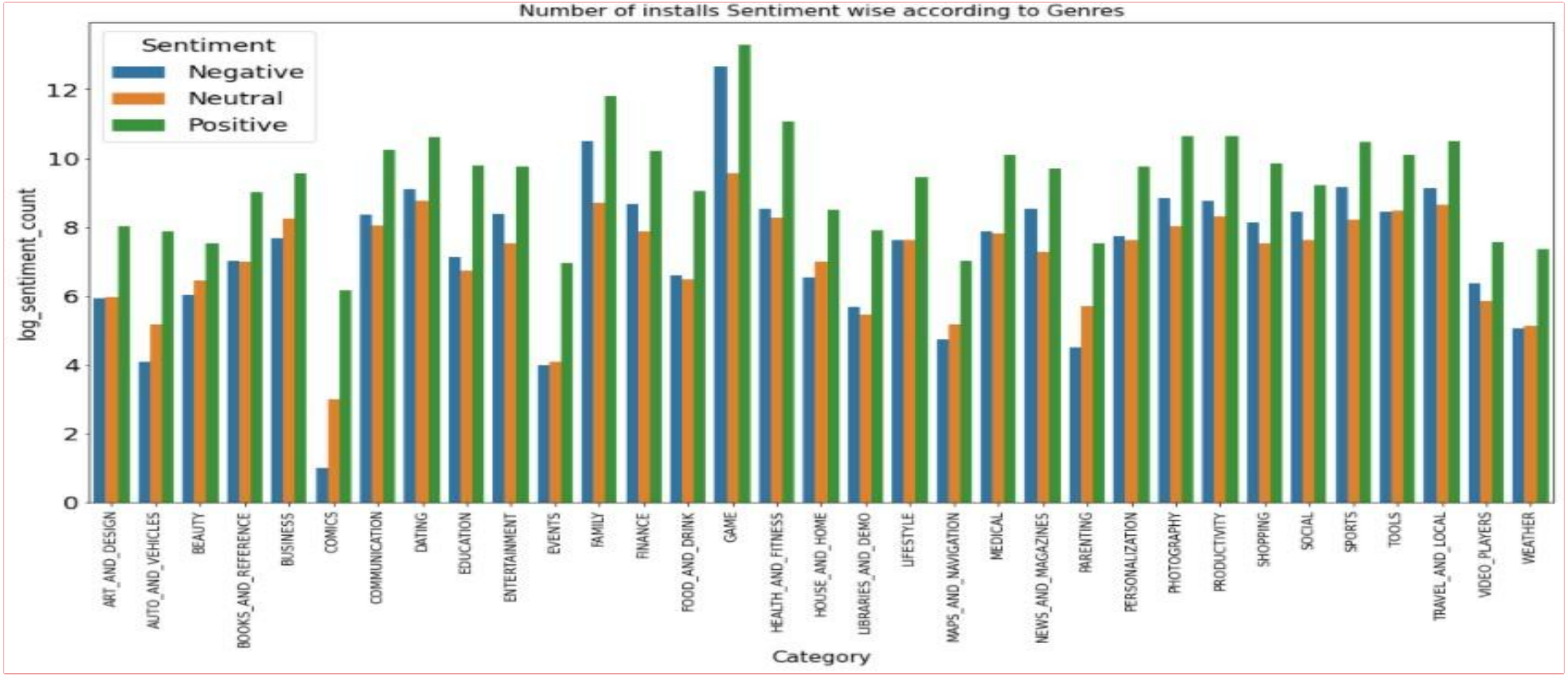
Challenges

Cleaning the null values from the dataset.

Understanding the sentiments of the user age wise.

Difference between installing the application and usage of the installed application.

Conclusion


Number of installs Sentiment wise according to Genres
Sentiment
Negative
Neutral
Positive
log_sentiment_count
0
2
4
6
8
10
12
ART_AND_DESIGN
AUTO_AND_VEHICLES
BEAUTY
BOOKS_AND_REFERENCE
BUSINESS
COMICS
COMMUNICATION
DATING
EDUCATION
ENTERTAINMENT
EVENTS
FAMILY
FINANCE
FOOD_AND_DRINK
GAME
HEALTH_AND_FITNESS
HOUSE_AND_HOME
LIBRARIES_AND_DEMO
LIFESTYLE
MAPS_AND_NAVIGATION
MEDICAL
NEWS_AND_MAGAZINES
PARENTING
PERSONALIZATION
PHOTOGRAPHY
PRODUCTIVITY
SHOPPING
SOCIAL
SPORTS
TOOLS
TRAVEL_AND_LOCAL
VIDEO_PLAYERS
WEATHER
Category

Conclusion

Users mostly prefer free applications.

Communication category apps are in high demand for all type of users. **(Source - External sources)**

A part of our population (18-30 yr age people) use Gaming category applications a lot and give reviews according to their Sentiments. So developers have to keep proper attention while developing those apps or making change.

Users installs the application depends according to previous rating and reviews.

References

https://seaborn.pydata.org/examples/index.html

https://matplotlib.org/3.1.1/index.html

Q & A