Capstone Project-1

Play Store App Review Analysis

AI

Team
Name: Web Crawlers

Members:

1. Rajat Chaudhary

2. Anukriti Shakyawar

3. Raman Kumar

4. Deepmala Srivastava

Content

1. Introduction
2. Problem Statement
3. Data Description
   a) Play store App Data
   b) User Reviews Data
4. Data Processing
5. Imported Libraries
6. Data Cleaning & Filtering
7. Exploratory & Visualization Analysis
8. Conclusion

Introduction

- Play store is one of the marketplaces for downloadable software programs with the highest growth in mobile applications.

- The play store apps data has a lot of potential to help app development companies succeed.

- Given the explosive rise of Android-based gadgets and applications, it will be useful to do data analysis on the collected information to gain insightful knowledge from this data.

Problem Statement

The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

Data Description
1. Play store App Data

The play store data is distinguished by 13 different variables, as:

1. **App**
2. **Category**
3. **Rating**
4. **Reviews**
5. **Size**
6. **Installs**
7. **Type**
8. **Prize**
9. **Content Rating**
10. **Genres**
11. **Updated**
12. **Current Version**
13. **Android Version**

Data Description
1. User Reviews Data

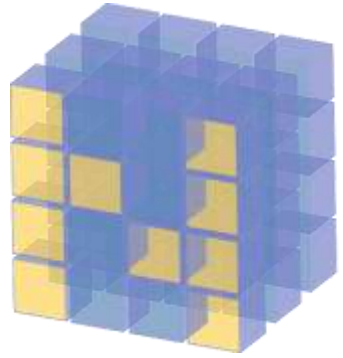In this dataset the review given by the consumers were calculated on different parameters, as:
1. App
2. Translated review
3. Sentiment
4. Sentiment Polarity
5. Sentiment Subjectivity

AI

Data Processing
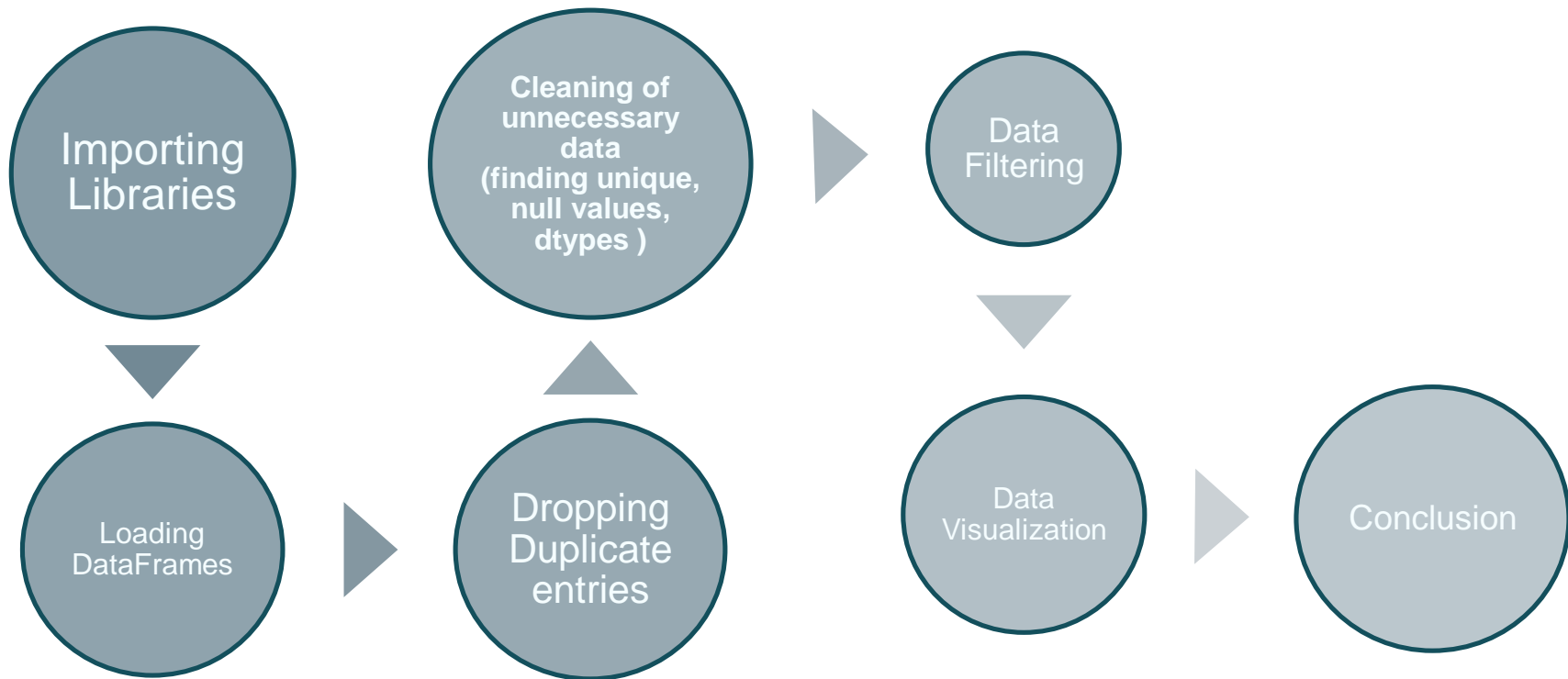
# Imported Libraries

Data Cleaning & Filtering

Play Store Dataset

- Identifying and removing duplicate entries.
- Removing visual impurities like "+", "$" sign.
- Comparing unique and non-null values to check how many entries were repeated.
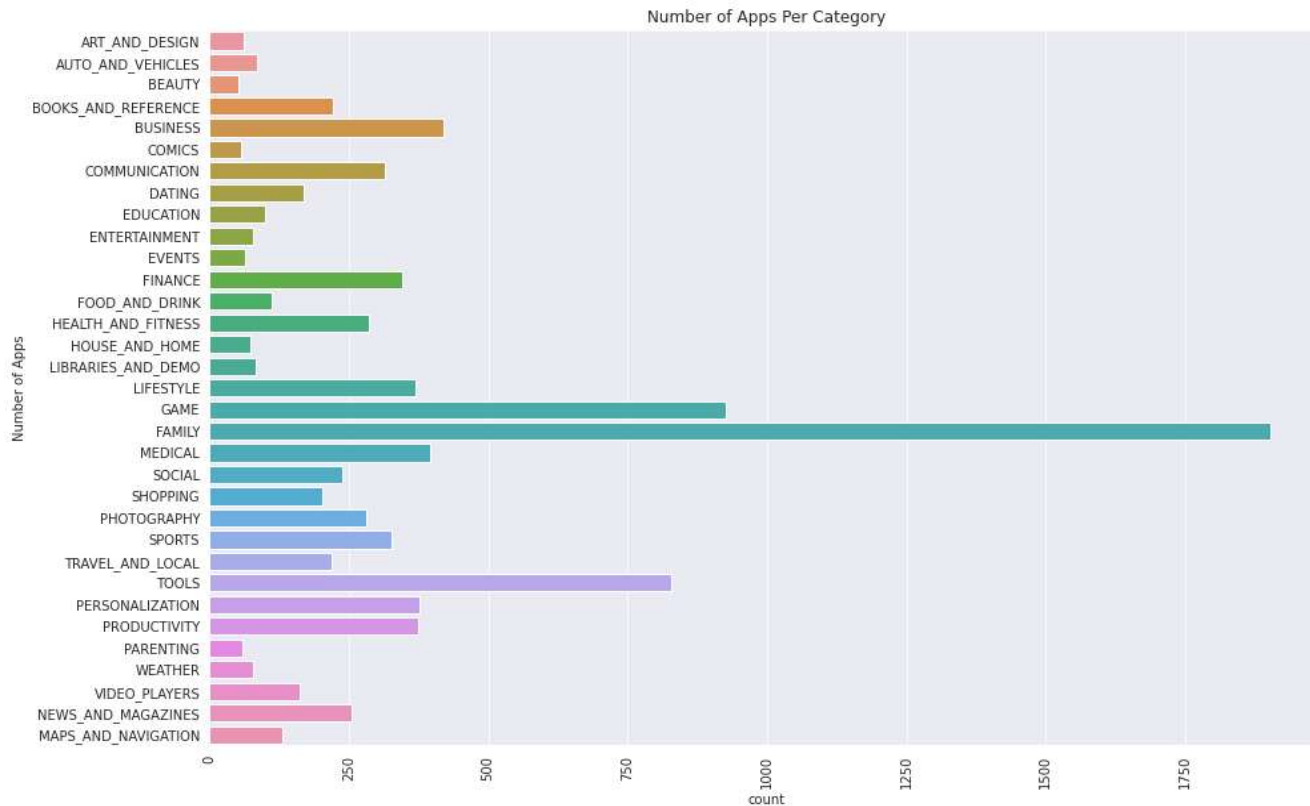- Filtering null values by operating on each column individually.

User Reviews Dataset

- Removing Duplicate values
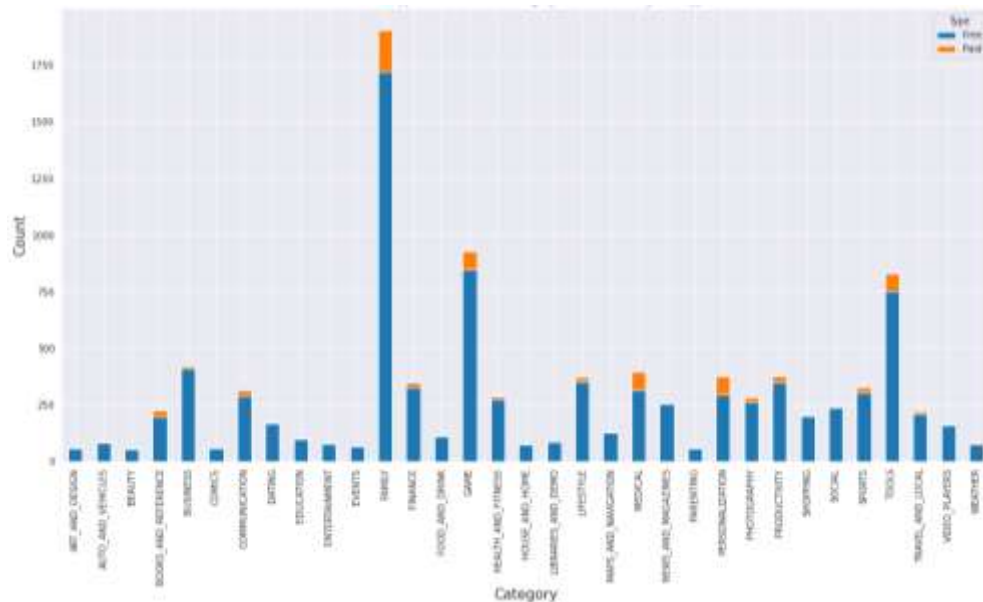- Filtering null values

Data Visualization

Distribution of Application Type

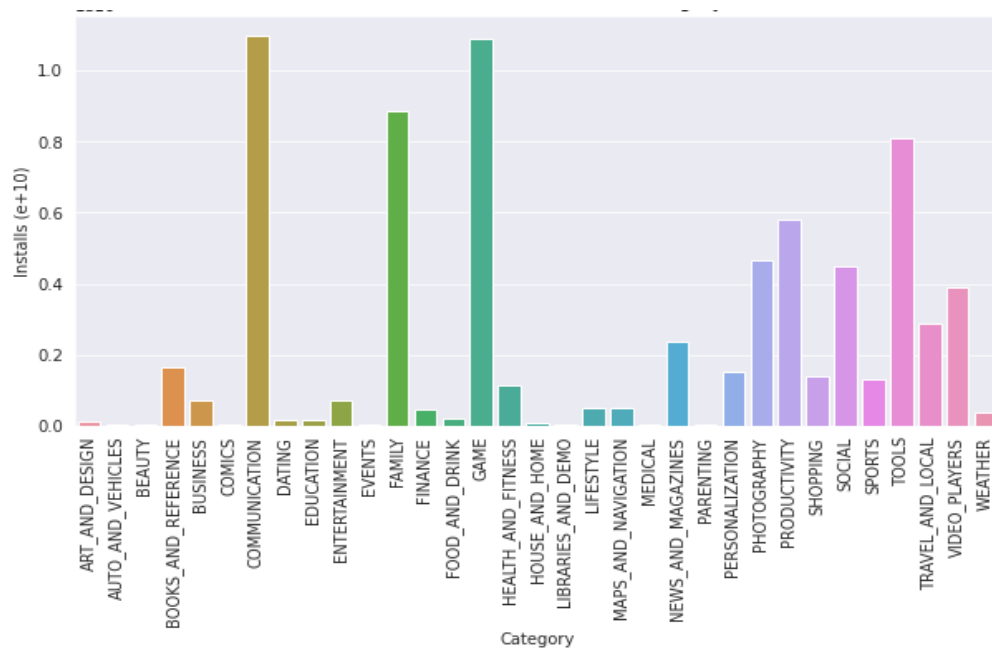# Number of Apps per Category
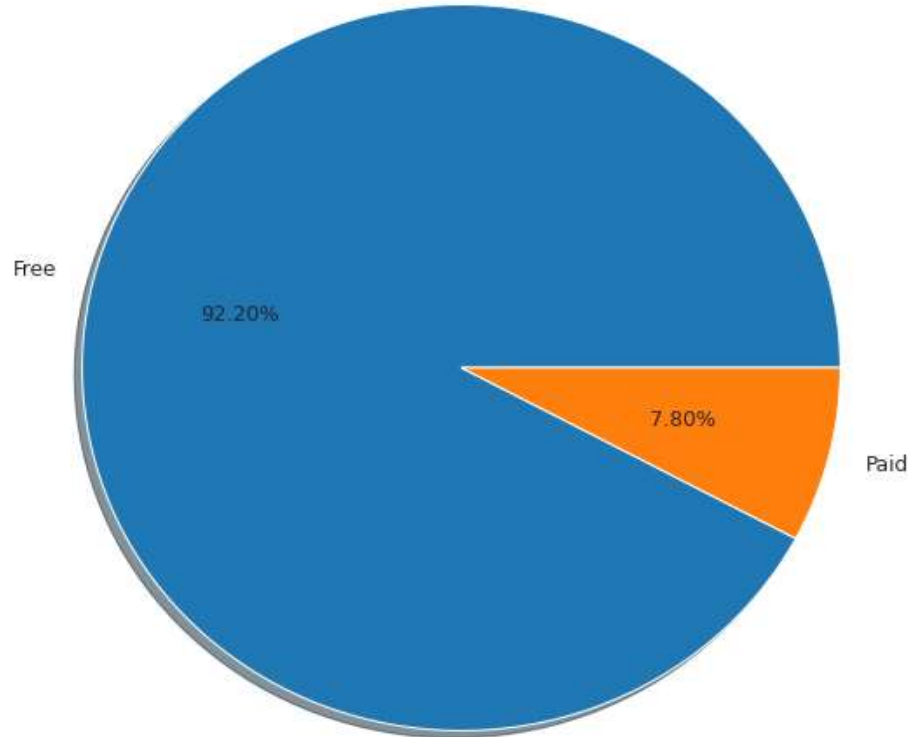
Number of Apps Per Category

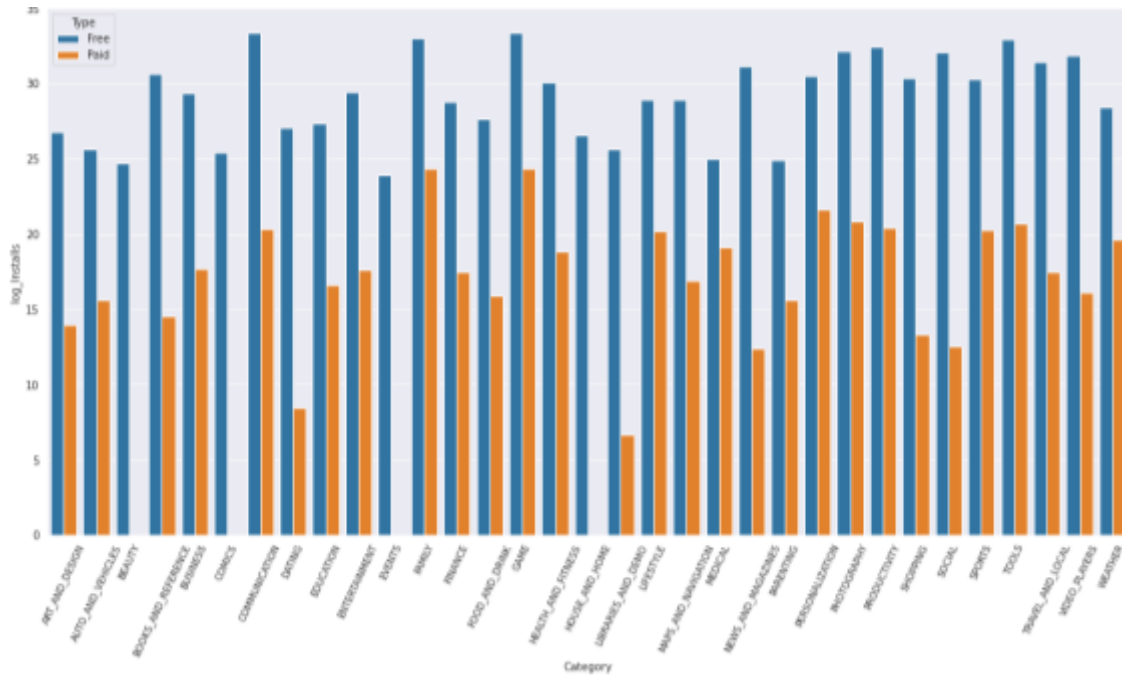# Number of apps per category differentiated by type

- Most free app download categories: Family, Games, Tools and Social
- Most paid apps download categories: Family, Personalization, and Medical

Number of installs for each category

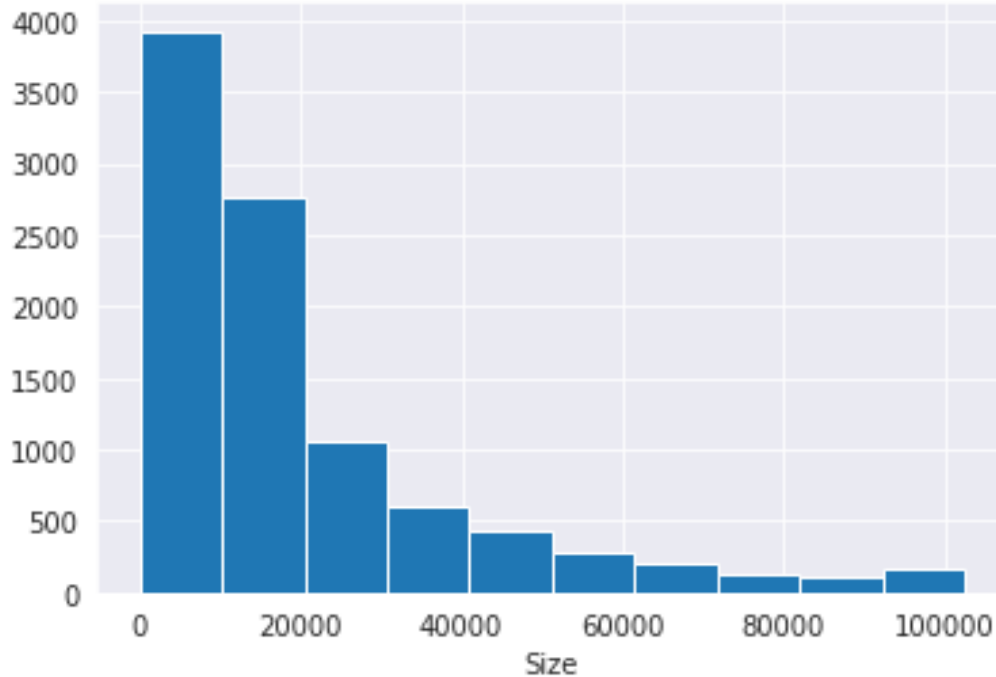Games, communication, Family and tools have the most installed apps.

Number of Installs Type-wise according to Categories

App installations have a significantly higher proportion of free software than paid ones.
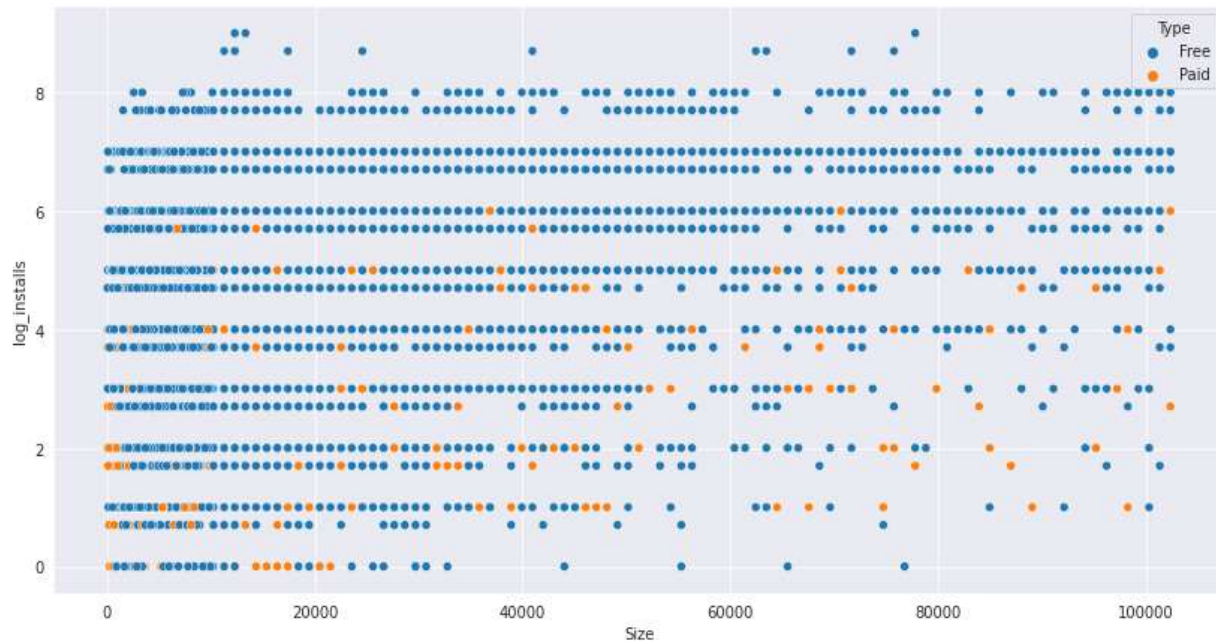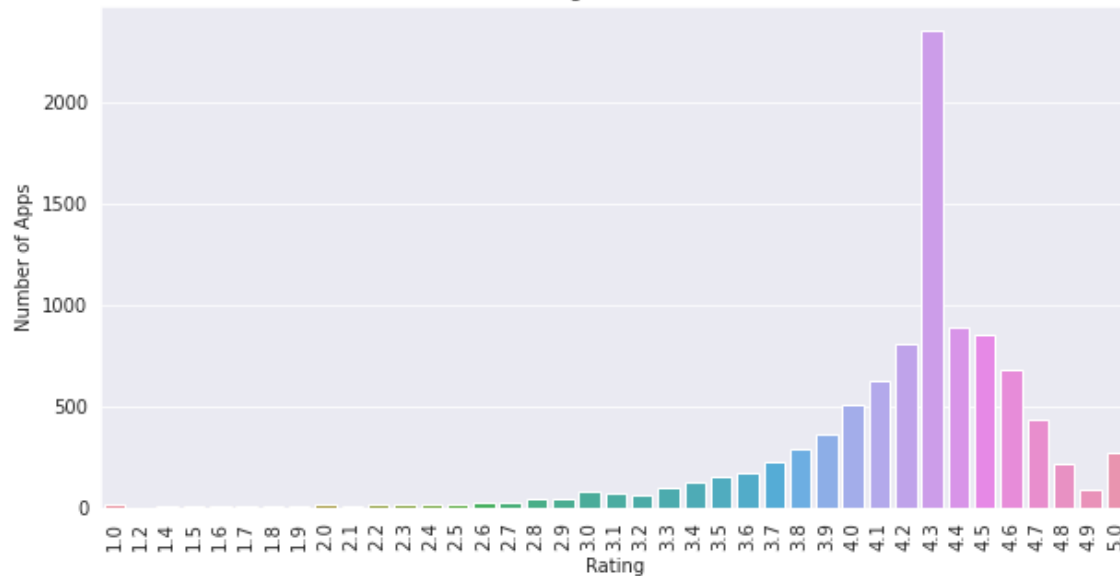
Distribution of Size of Apps

Most of the apps present in Play Store are smaller in size and consume less memory
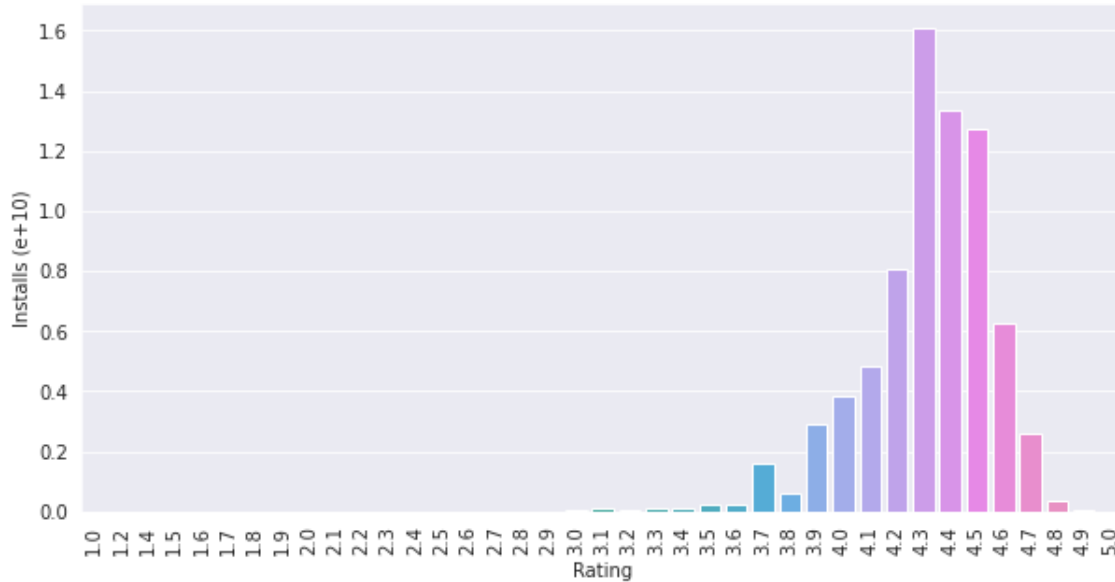
Impact of size on the number of installs

- Size does impact the no. of installs.
- Bulky apps tend to be installed less.
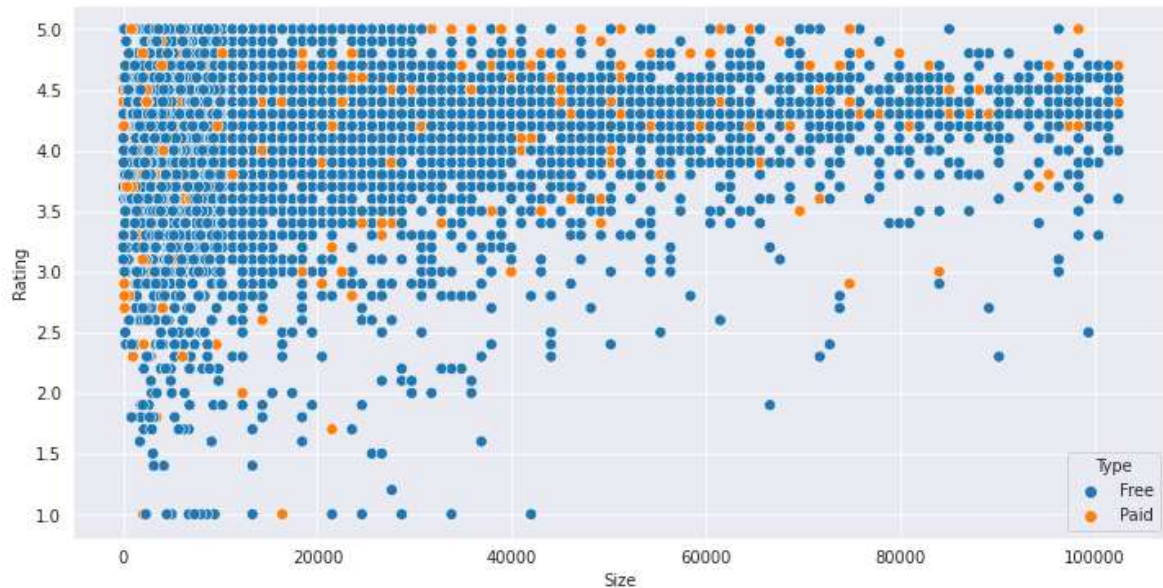
Distribution of App Ratings

Most of the apps in the Play Store are having rating higher than 4 or in the range of 4 to 4.7.

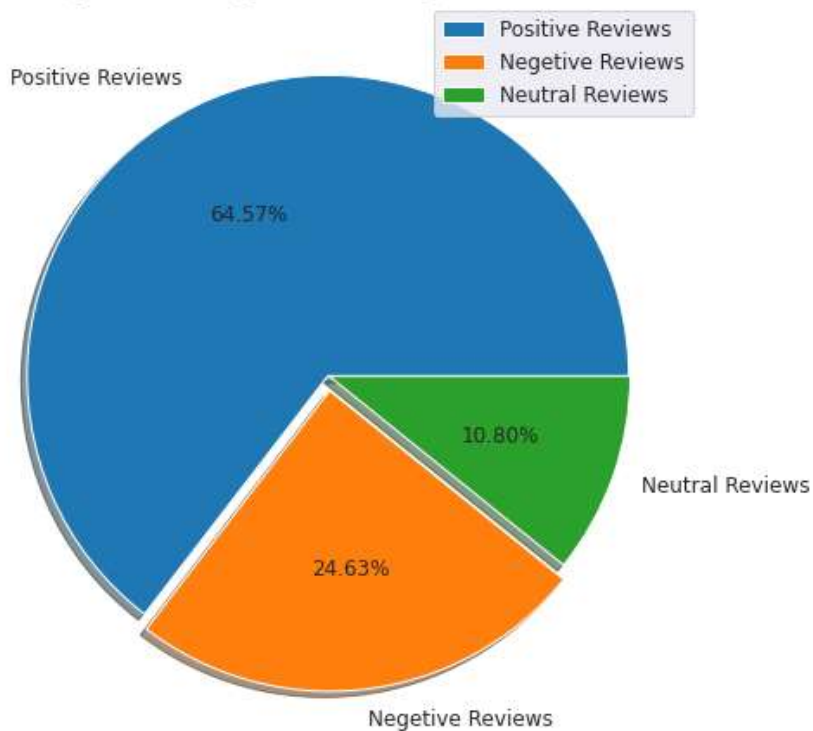# Number of Installs per Rating

Most of the apps downloaded by the customers are of higher rating (between 4.2 to 4.6 ratings).

# Distribution of Apps in terms of their Ratings, Size and Type

- The 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.
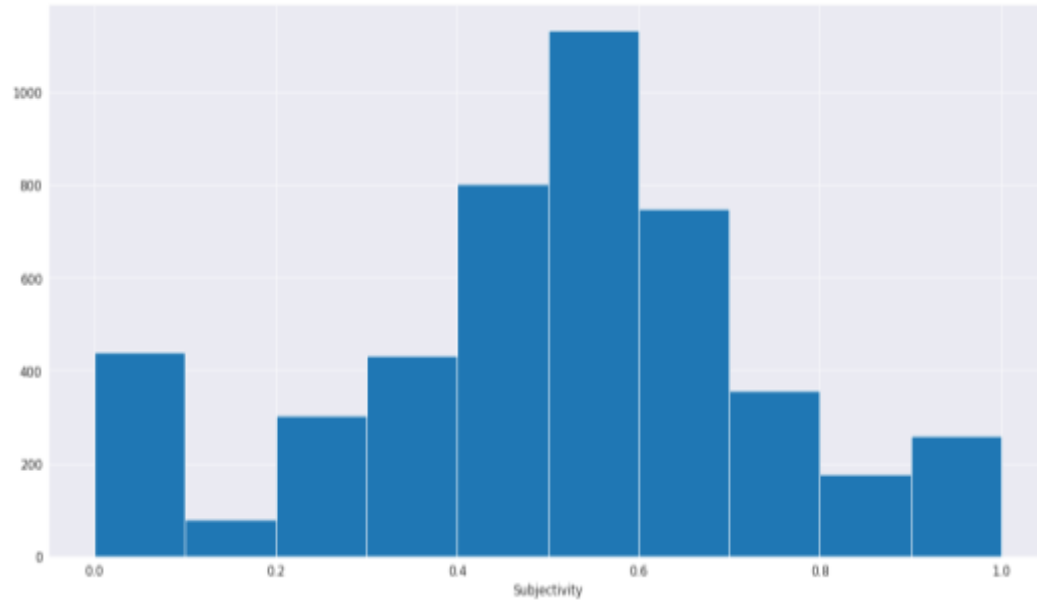
Sentiment of Reviews

- Most of the reviews are positive in nature with 64.57%
- Negative reviews are only 24.63%.
- Neutral reviews are only 10.8%.

Distribution of type of reviews as per categories
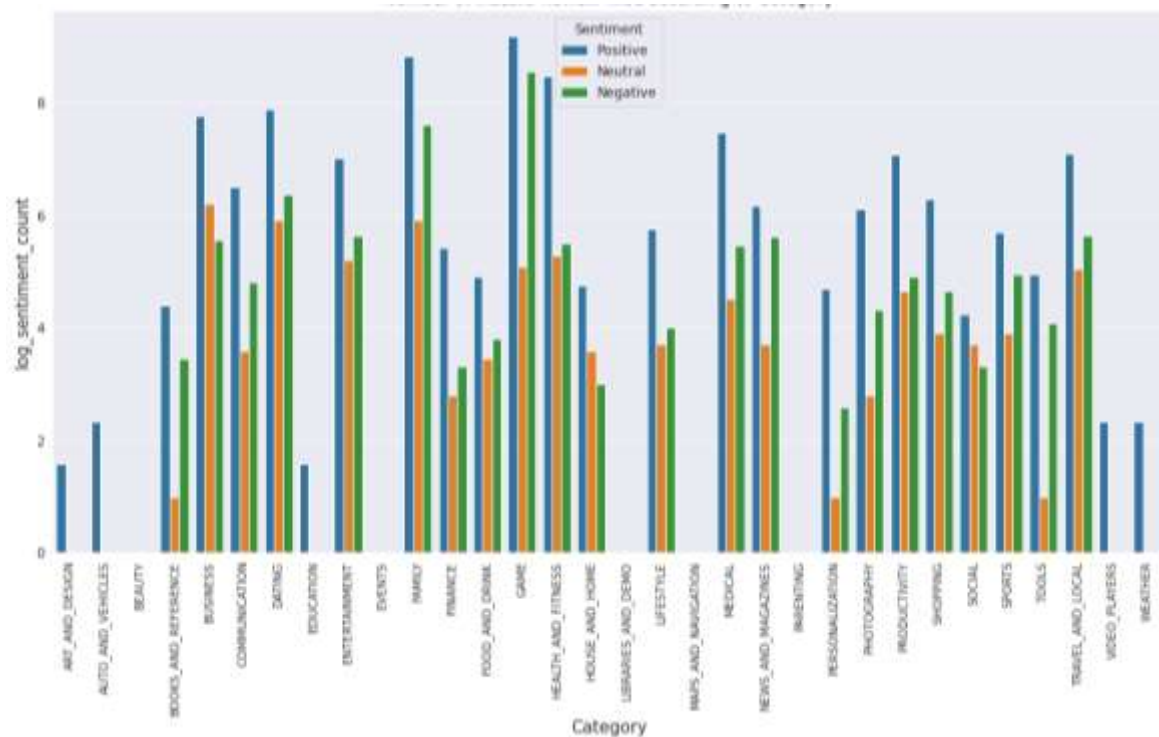
- The number of positive reviews are way higher than negative and neutral ones.
- Some categories did show equally higher no. of negative and neutral review.
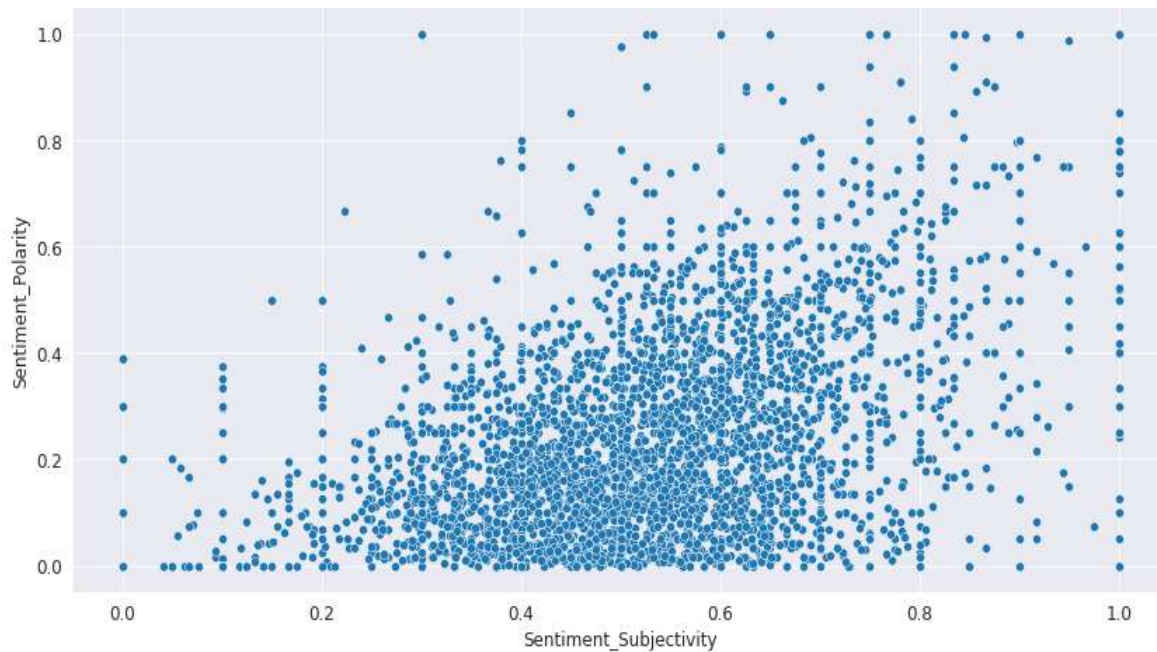
Distribution of Subjectivity

- Maximum number of sentiment subjectivity lies between 0.4 to 0.7.
- These reviews comes from the experience from the users while using these apps.
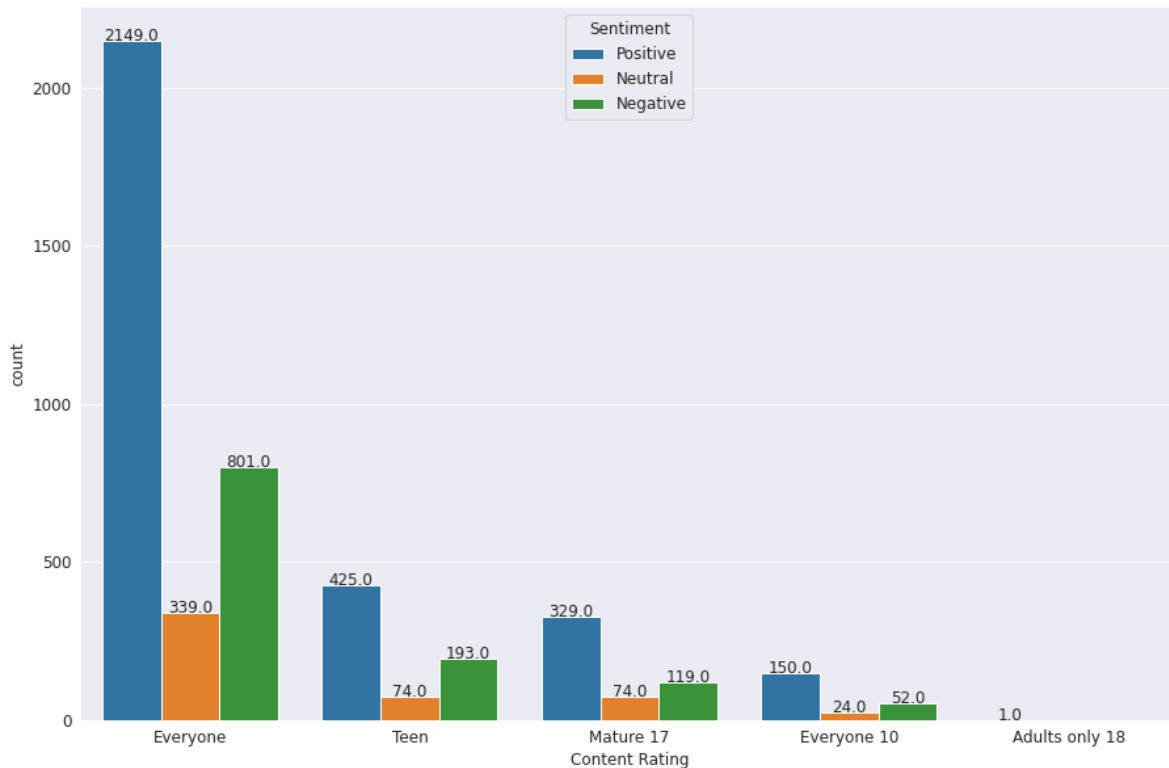
# Sentiment subjectivity v/s Sentiment Polarity

Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of cases, it shows a proportional behavior when variance is too high or low.

Content Rating on the basis of Age

- Most of the positive ratings came from the everyone category apps with 2149 ratings and 801 negative ratings.
- Neutral ratings are also low in comparison to positive and negative reviews.

# Sentiment polarity relationship with Type
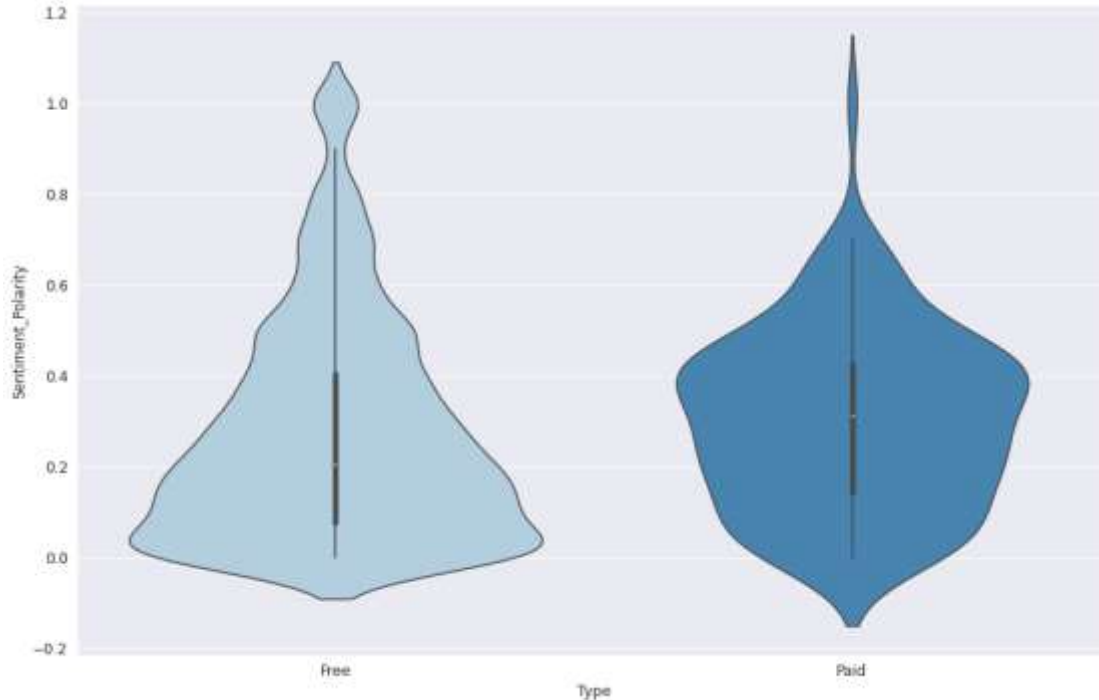
- In Free apps the sentiment polarity lies majorly in 0.1 and decreases after that.
- The sentiment polarity of paid apps which falls majorly between 0.1 to 0.4 with highest in 0.4.

# Content Rating relation with Sentiment Polarity
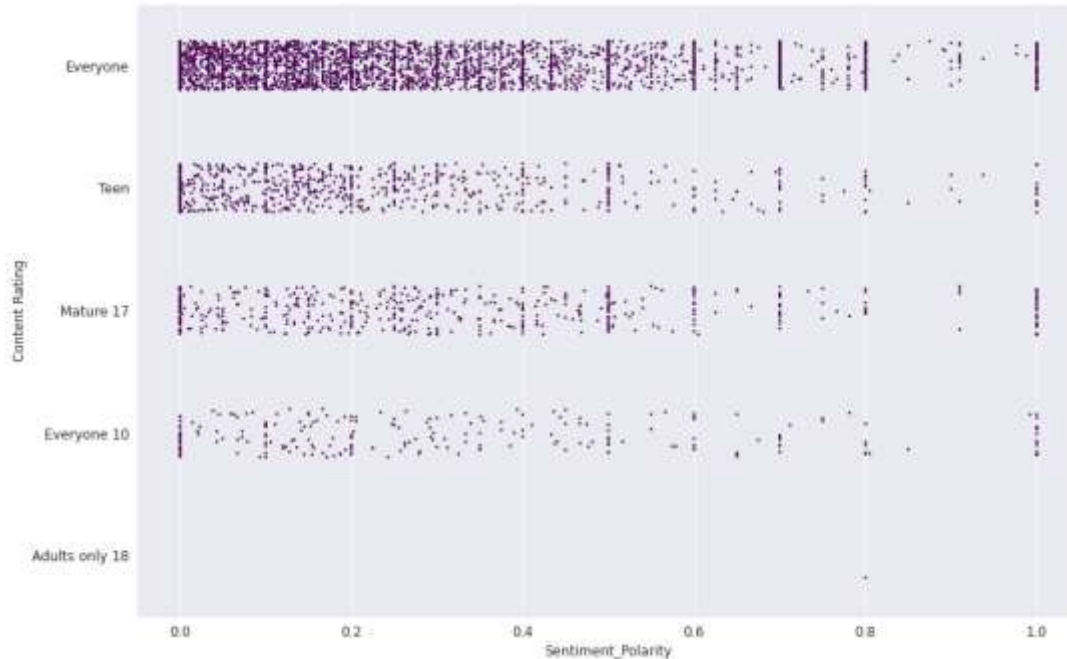
- Sentiment polarity is low in most of the categories except for Everyone category.
- Polarity is evenly distributed.

Categories Relation with Sentiment Subjectivity
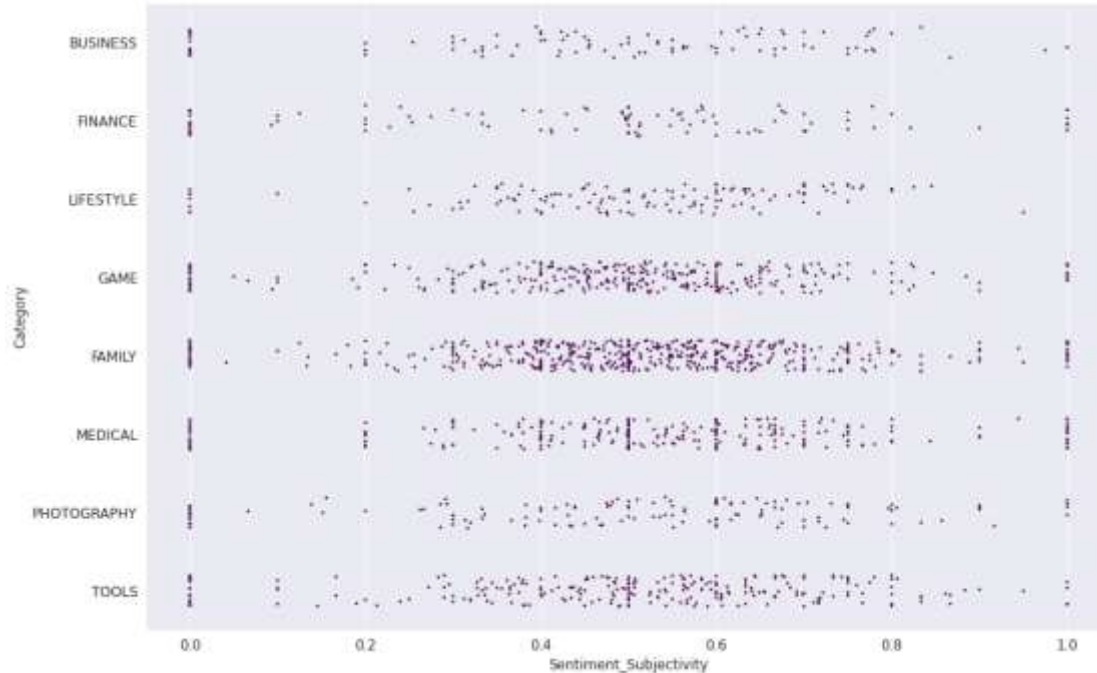
- Subjectivity lies mostly between 0.2 to 0.8 in all Categories.
- Family Category has the higher number of reviews in which Subjectivity lies between 0.4 to 0.6

Word Cloud of Reviews
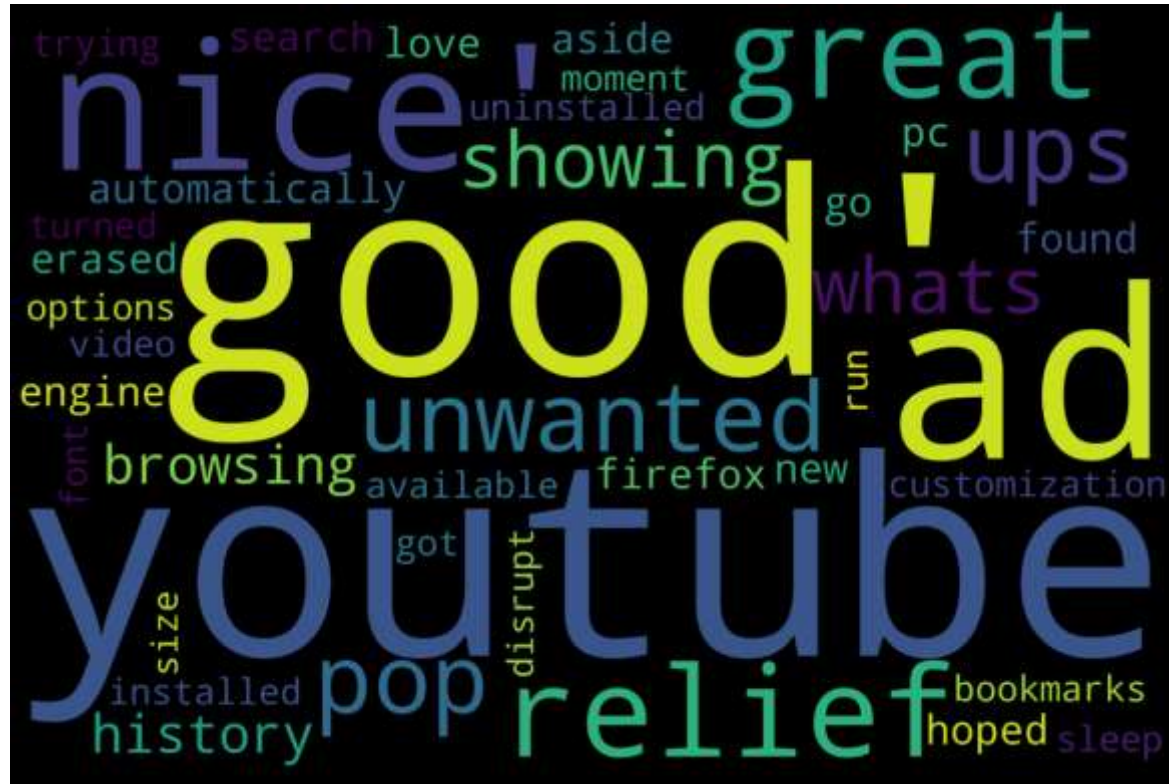
Asking Questions from Dataset

- 1. What is the top 5 apps on the basis of installs?
- 2. What is the top 5 reviewed apps?
- 3. What is the top 5 expensive apps?
- 4. What is the top 3 most installed apps in Game category?
- 5. Which 5 apps from the 'FAMILY' category are having the lowest rating?

Conclusion

❖ Most of the trending apps are from the categories like family, tools, and games.

❖ Most preferred Apps by users in point of size or weight are light size apps.

❖ Users also installed apps on the basis of their ratings.

❖ These ratings are defined on 2 points- Polarity and Subjectivity.

❖ This analysis will help developers while preparing for their next apps.