



Google Play
Store

AI

An Exploratory Data Analysis on

Playstore App Reviews

Capstone Project by:

Sanjeev Hegde

Data Science Trainee, **AlmaBetter**

Table of Contents:

- Need for the Analysis
- Problem Statement
- Introduction
- Attributes in the Data
- Data Cleaning and Preparation
- Statistical Walkthrough for numerical data
- Analysis of the Category of Apps
- Category and Installs
- Categories and Rating
- Correlation Heatmap, Reviews and Installs
- App size and Installs
- App Size and Rating
- Content Rating and Apps
- App(s) with Highest Rating, Price and Installs
- Sentiment Polarity Analysis from User Reviews
- Sentiment Subjectivity Analysis from User Reviews
- Sentiment Polarity, Sentiment Subjectivity and Apps
- Content Rating and User Reviews



AI



Why perform an analysis?

AI



For analysing various parameters affecting user preference- improvising app development

For studying shortfalls in apps providing services in specific categories

For analysing the user experience from reviews

For developing appropriate revenue model and/or pricing

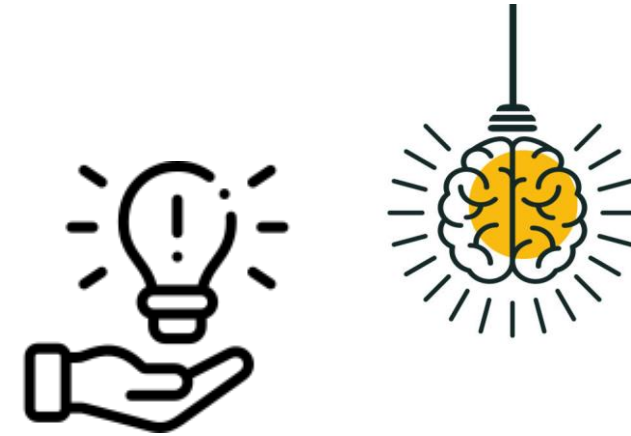


What makes a successful app?



There are numerous factors which need to be considered while developing an app in order to ensure it is successful.

- Is it the Category of the App?
- Does size of the App matter?
- Do app reviews help in fueling the installs?
- Priced or a Free App to develop?



These are some of the examples of questions that pop up in head while developing an app or while providing any service to customers through Apps on Playstore.

Let us analyze the given two Data sets and answer essential questions for better understanding the App market

Introduction

AI

- Google Play store is the official app store for devices which run android platform developed by Google
- Developers can get their certified apps hosted on the Playstore
- Applications can be either free or priced or even contain in-app purchases
- As on 2017 August, about 3 million android applications are on play store
- Apps include Services, Games, Books, Movies etc
- Key features include:
 1. Provide reviews and ratings for each app which are public
 2. Sign up for beta programs
 3. Make in-app purchases through google play balance



Google Play

Attributes in Data

1. Play store Data (data set-1)

1. **App:** The name of the application on which the data is provided
2. **Category:** The category to which the app belongs to
3. **Rating:** User Rating for the app out of 5
4. **Reviews:** Publicly available user reviews
5. **Size:** Size of the application
6. **Installs:** Total Number of installations
7. **Type:** Free or paid
8. **Price:** Price of the application
9. **Content Rating:** To whom the application has been developed for
10. **Genres:** Genre the application belongs to
11. **Last Updated:** Last update provided the developer for the app
12. **Current Ver:** Present version of the application
13. **Android Ver:** Android version that application is supported on

1.2 User Reviews (data set-2)

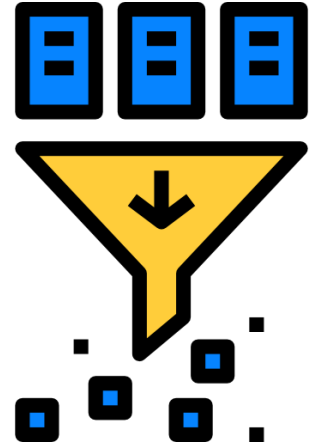
- 1. App:** The name of the application on which the data is provided
- 2. Translated Reviews:** The translated and processed content of the user review
- 3. Sentiment:** The sentiment of the user review
- 4. Sentiment Polarity:** A quantitate measure of the sentiment in user's review. Values range from -1 implying highest negative to +1 implying highest positive sentiment in the review.
- 5. Sentiment Subjectivity:** A quantitative measure of the subjectiveness of the review. Values Range from 0 to +1 wherein zero implies Objective review while +1 implies highly subjective review

Data Cleaning and Preparation

AI

Given data sets are cleaned and processed for exploratory data analysis. Following processes have been made:

1. Identification and modification of abnormal data with appropriate changes
2. Finding and filling null values in the data with suitable numeric values
3. Conversion of quantified parameters such as app size to Kilobytes
4. Eliminating non essential characters such as commas, +, \$ etc in numeric values
5. Converting string values of dates to date and time format
6. Converting all columns to their required data type
7. Dropped apps with same name and last update time as these are duplicates and are extremely likely be developed by the same Developer
8. Dropping Non essential columns such as Translated Review Parameter which has been already evaluated for Sentiment, polarity, subjectivity



Statistical Overview:

AI

- There are 10,841 entries in the Play Store Data set
- 9701 unique entries remain after data cleaning
- Average Rating for the apps is about 4.2 stars
- Average size of app is about 16MB
- Largest app is 100MB in size while smallest app is 1KB
- Costliest app costs \$400
- Standard deviation for ratings from mean is about 0.49 stars

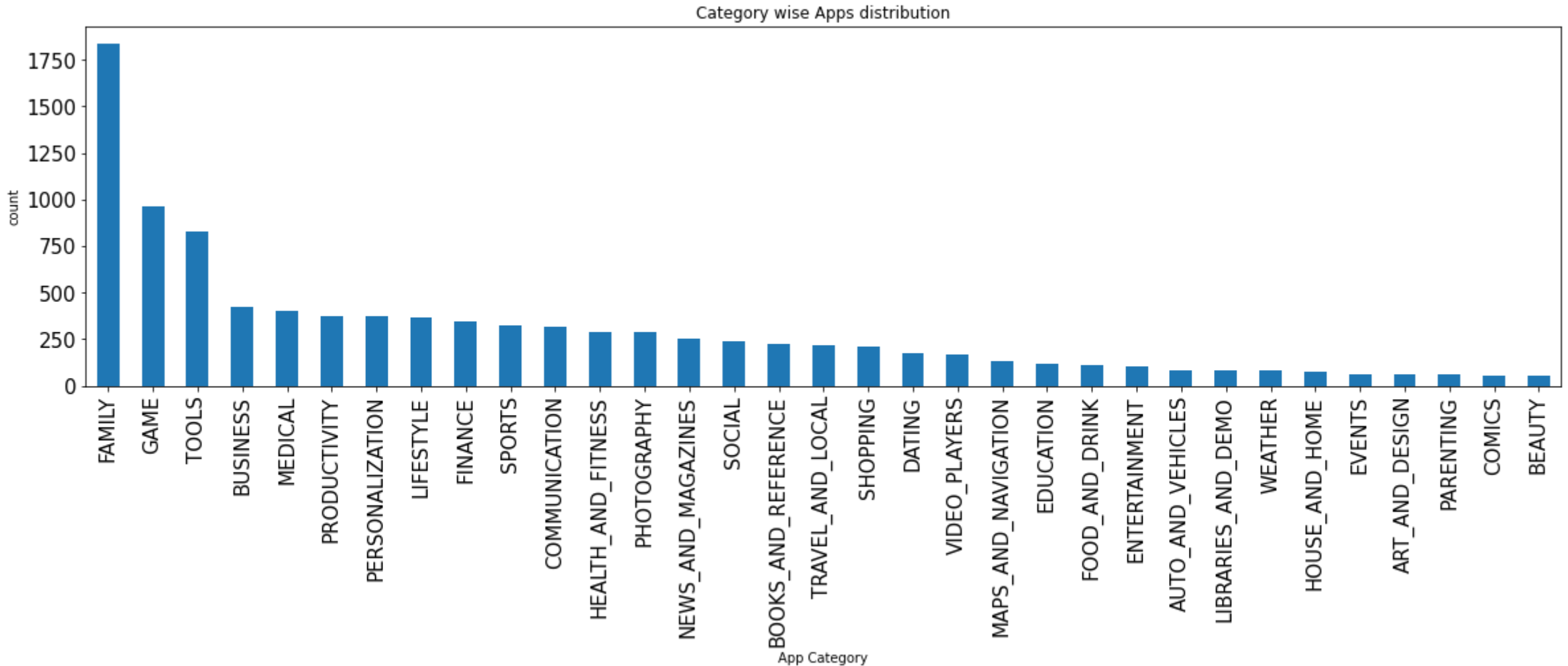


OVERVIEW

Category wise apps



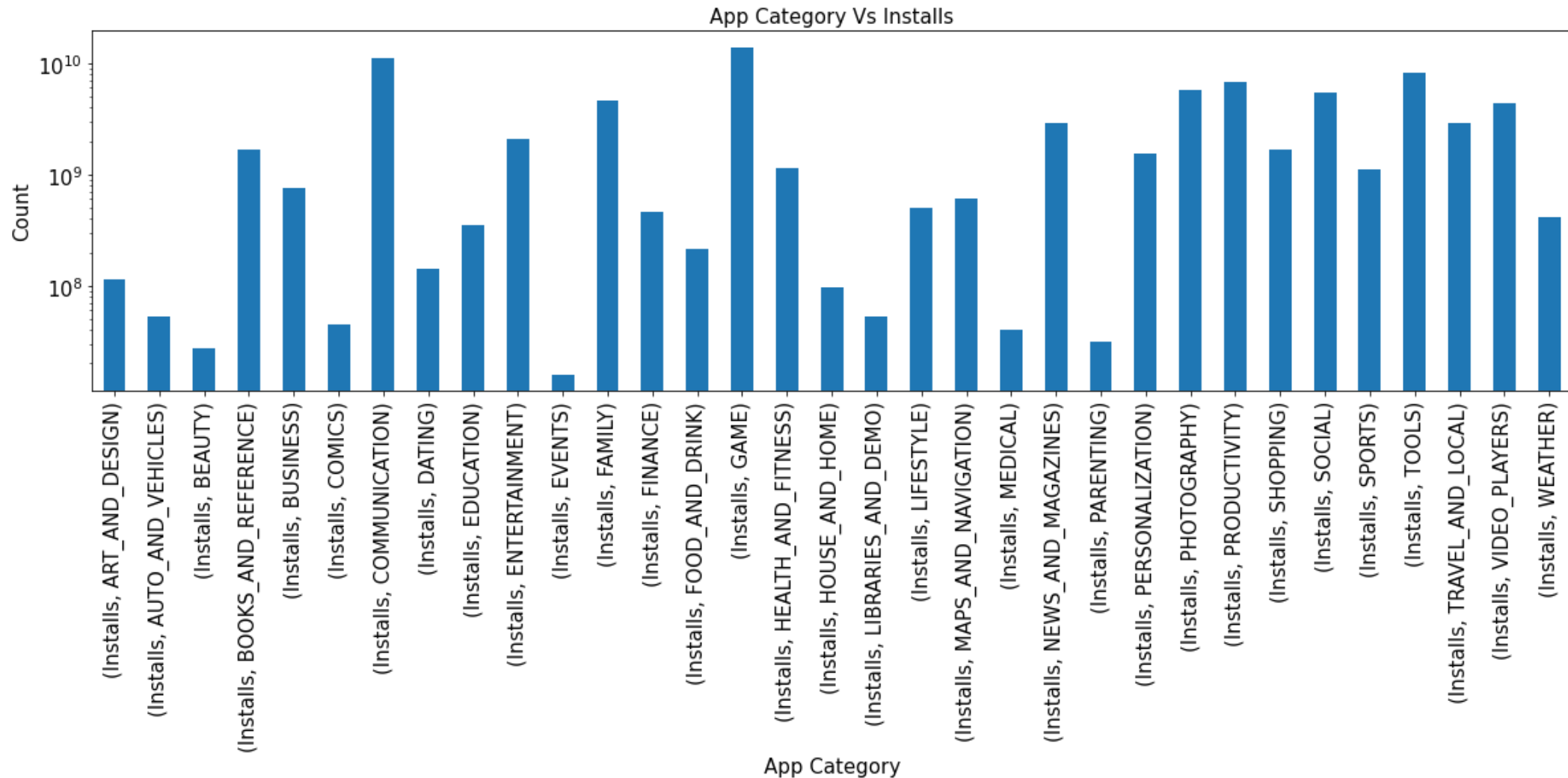
- It can be seen that Family category apps are present the most followed by Games and Tools categories



Category and Installs

AI

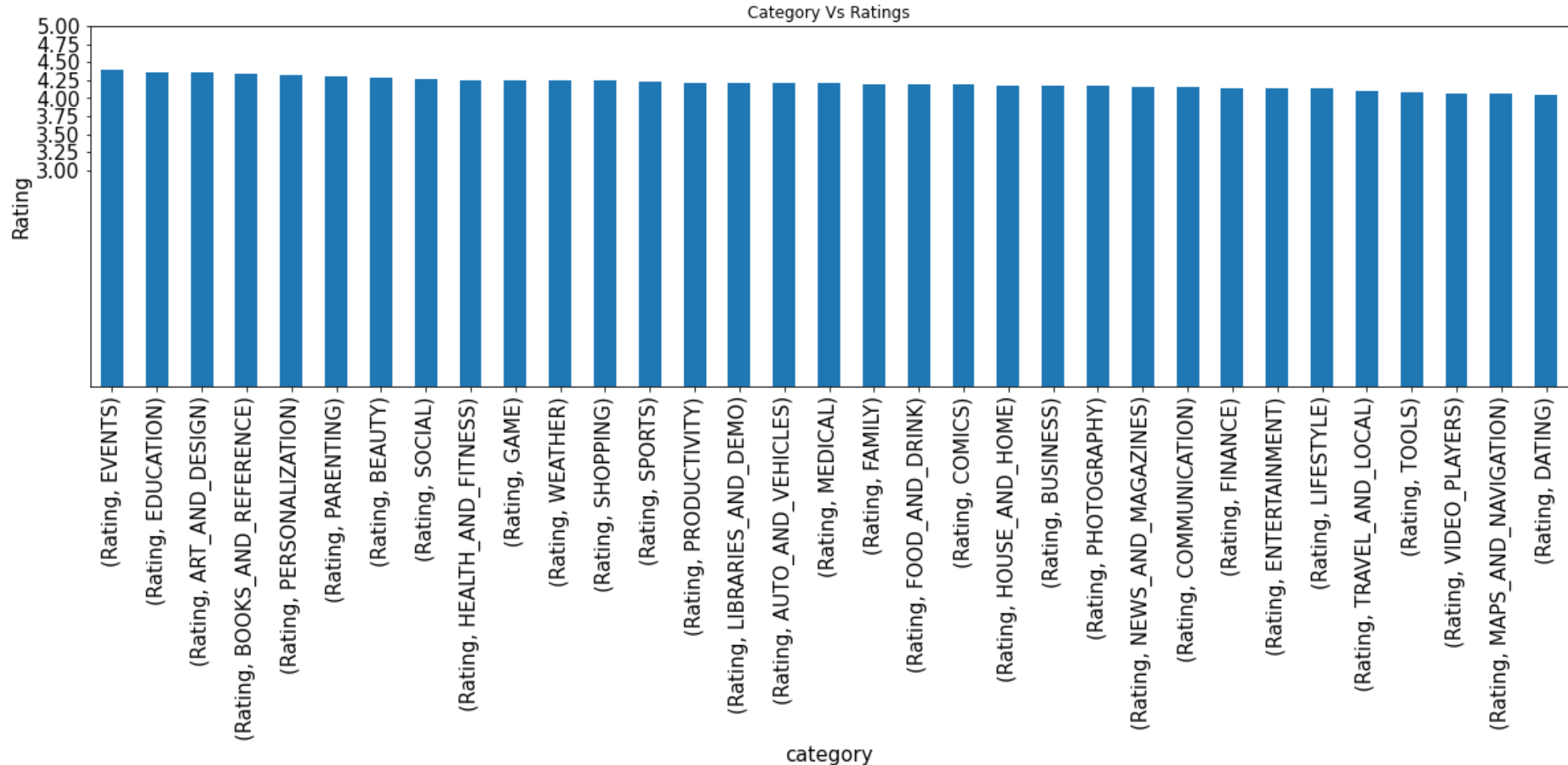
- It can be seen that Game category apps have highest installs followed by Communication apps



Category and Rating

AI

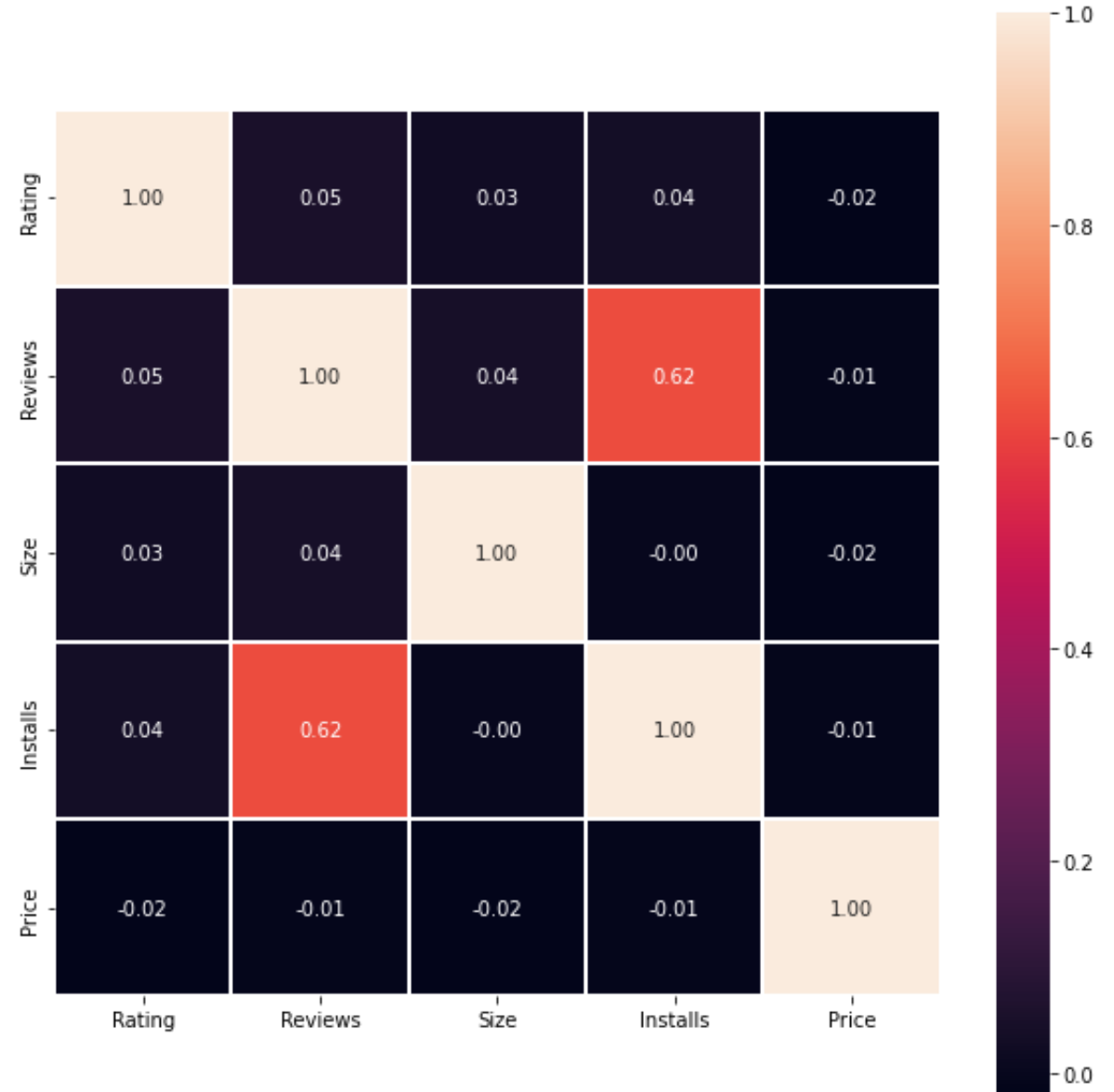
- It can be seen that Event category apps have highest Ratings about 4.4. All the category of apps have ratings above 4 Stars. There are 271 apps with 5 stars rating



Correlation Heatmap

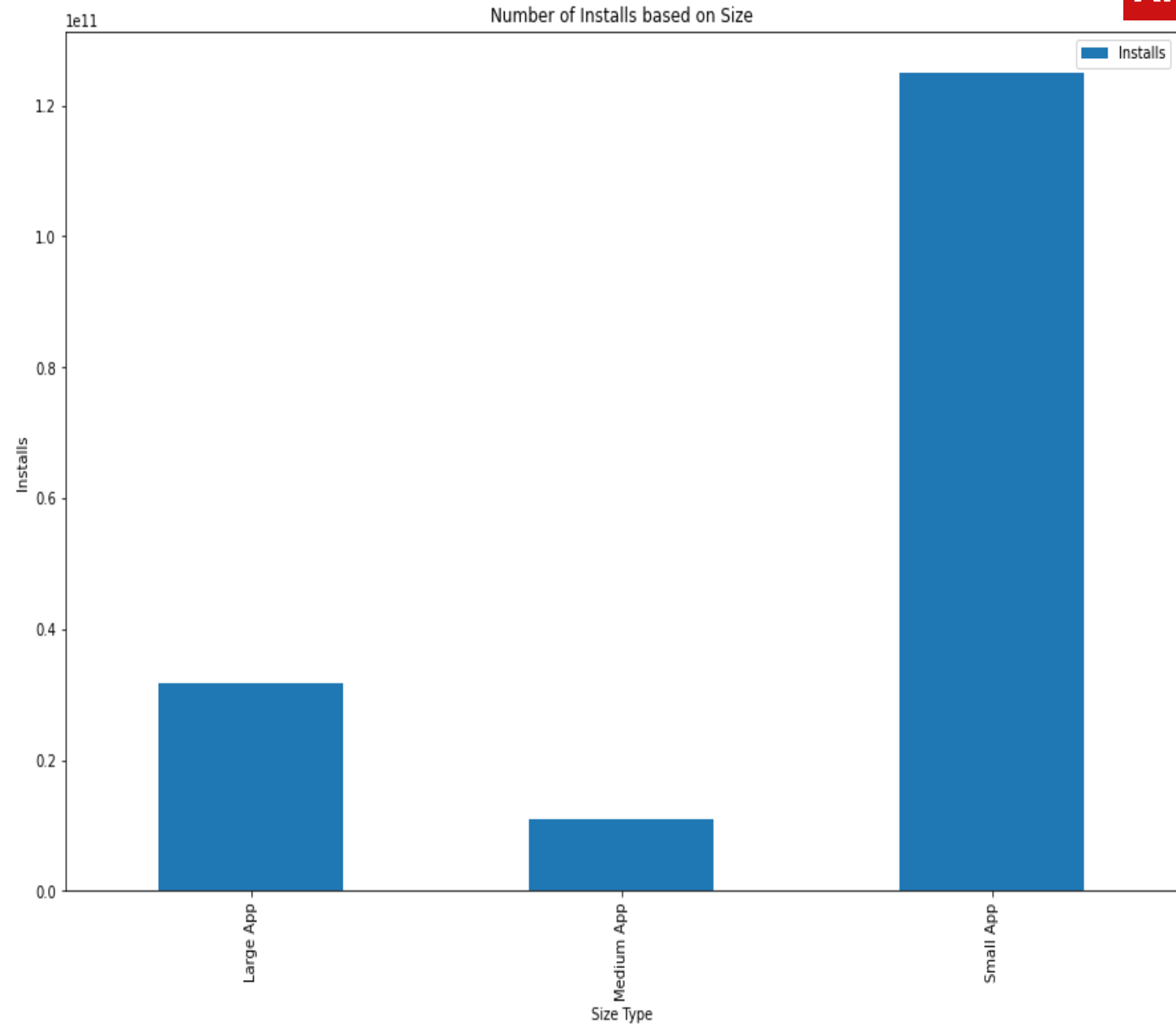
AI

- There is a direct correlation between Reviews and Installs
- The Correlation is 0.62
- Users tend to install apps with higher number of reviews in comparison to apps with fewer reviews



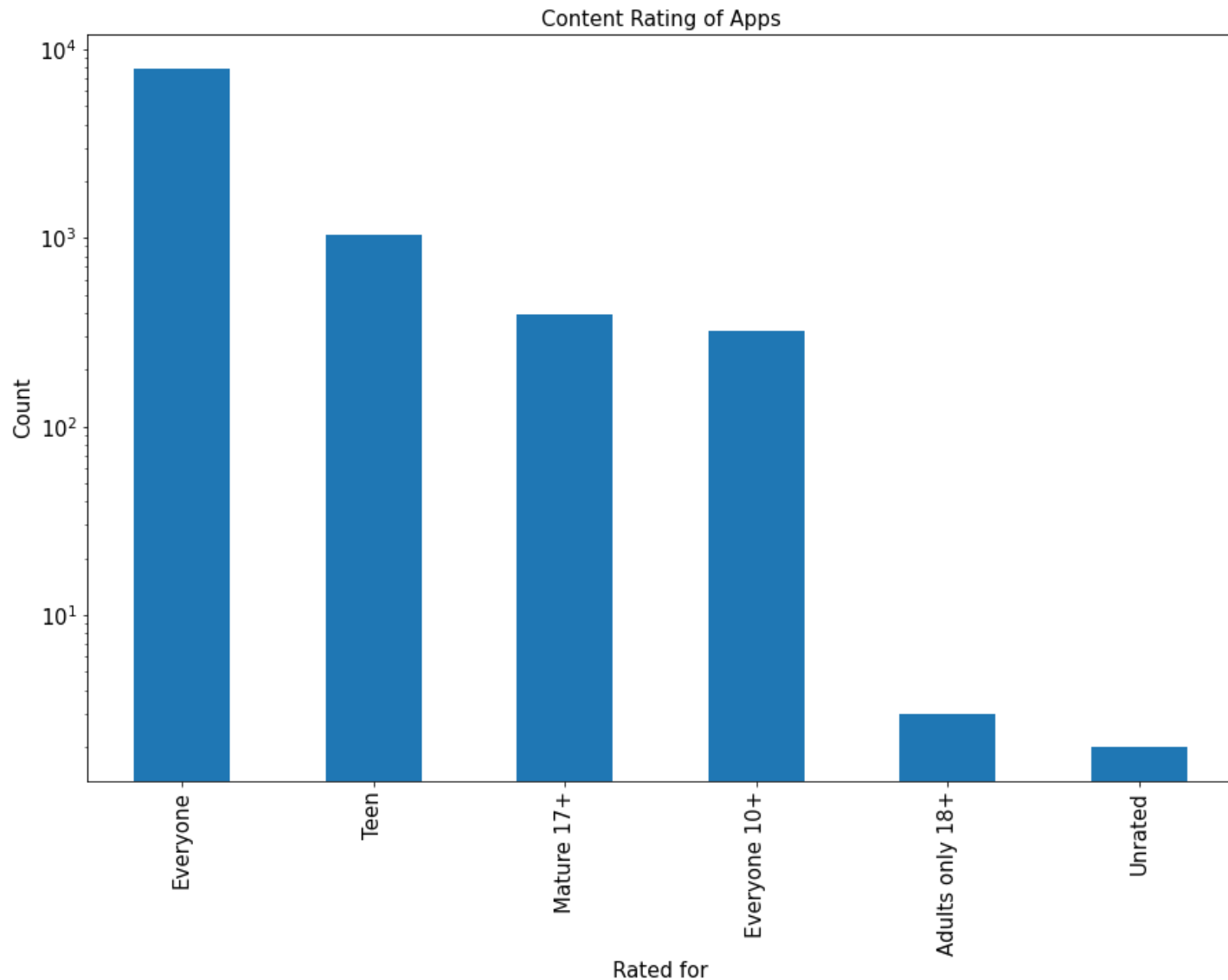
Installs and Size

- Initially apps are categorized according to their size wherein:
 - Small: 0MB to 25MB(excluding)
 - Medium: 25MB to 50MB(excluding)
 - Large: more than 50MB(including)
- It can be seen that Small apps have been installed the most
- Larger apps are the second most installed type of apps with size more than 50MB



Content Rating and Apps

- The given data set has highest number of apps rated for Everyone followed by Teen
- It can be noted that there are few unrated apps



Some more insights:

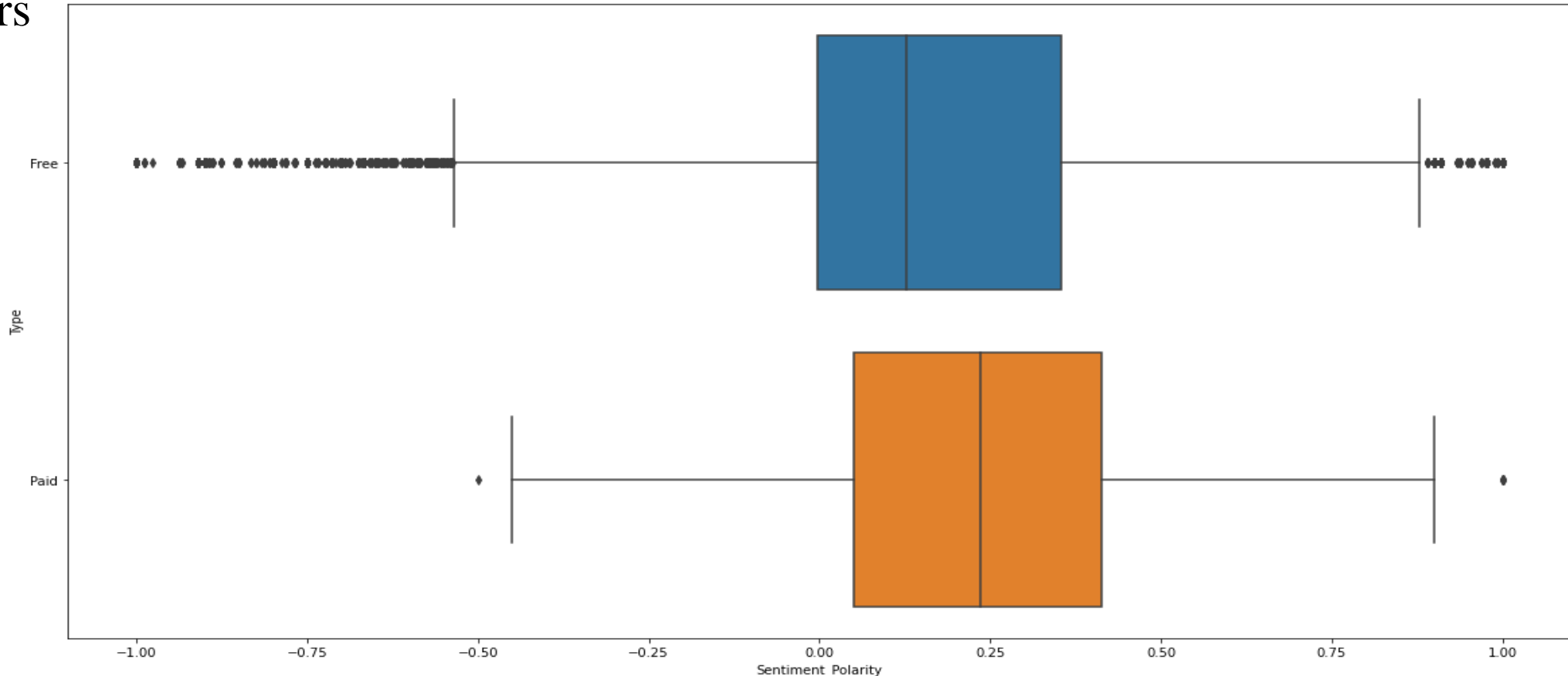
AI

- The Costliest app is I'm Rich - Trump Edition with price tag of 400\$ and 3.6 star rating with more than 10,000 installs.
- Free apps have been downloaded significantly higher compared to paid apps. There are 79,16,18,73,646 installs for Free apps in Total while paid apps have 5,73,63,881 downloads.
- Top installed apps have been downloaded more than 10^9 times which includes apps such as Facebook, WhatsApp, Instagram, Subway Surfers etc.
- There are 271 apps with 5 star rating and top app with highest average user rating is "Ek Bander Ne Kholi Dukan" belonging to Family category and entertainment Genre with 10,000 plus downloads

Sentiment Polarity Analysis

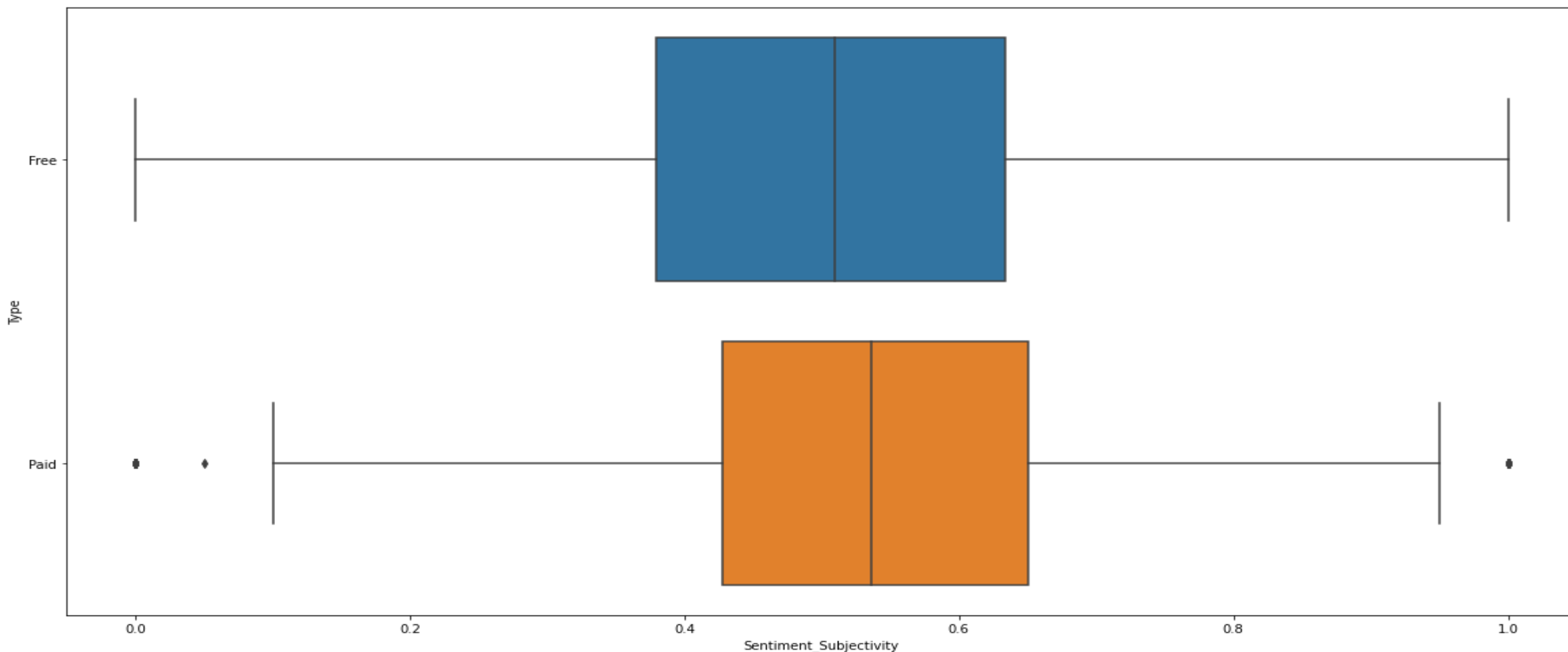


- median sentiment polarity for free apps is about 0.15 implying slightly positive sentiment among users in general
- median sentiment polarity for paid apps is about 0.23
- Paid app users have reviewed the installed apps in more positive way than Free app users



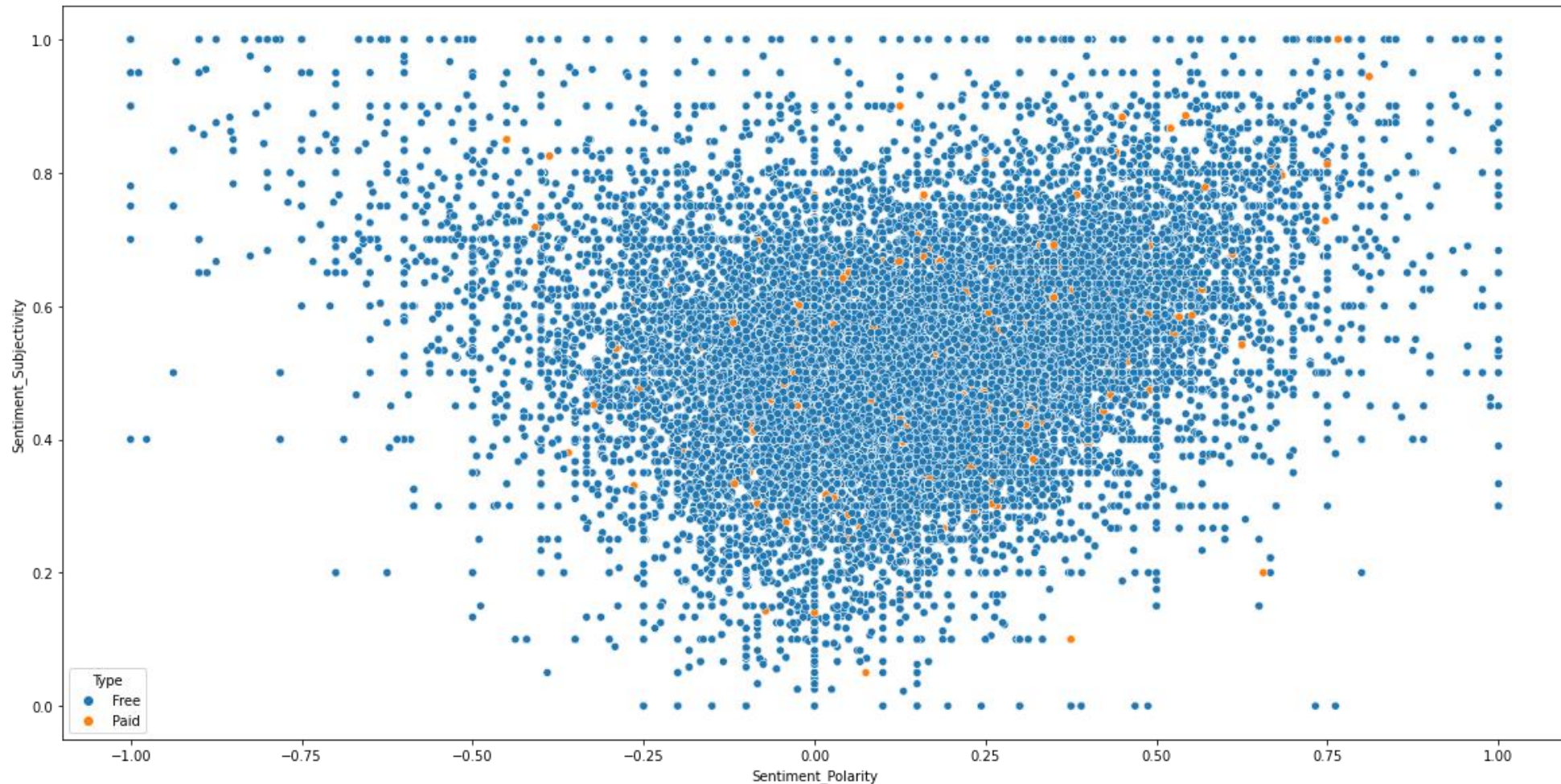
Distribution of apps based on Sentiment Subjectivity

- median for Free apps is about 0.5 while median for paid apps is about 0.55
- Relative to the paid app reviews, free app reviews are more objective in nature
- There are extremely objective and subjective reviews in paid reviews



Apps, Sentiment Polarity and Subjectivity

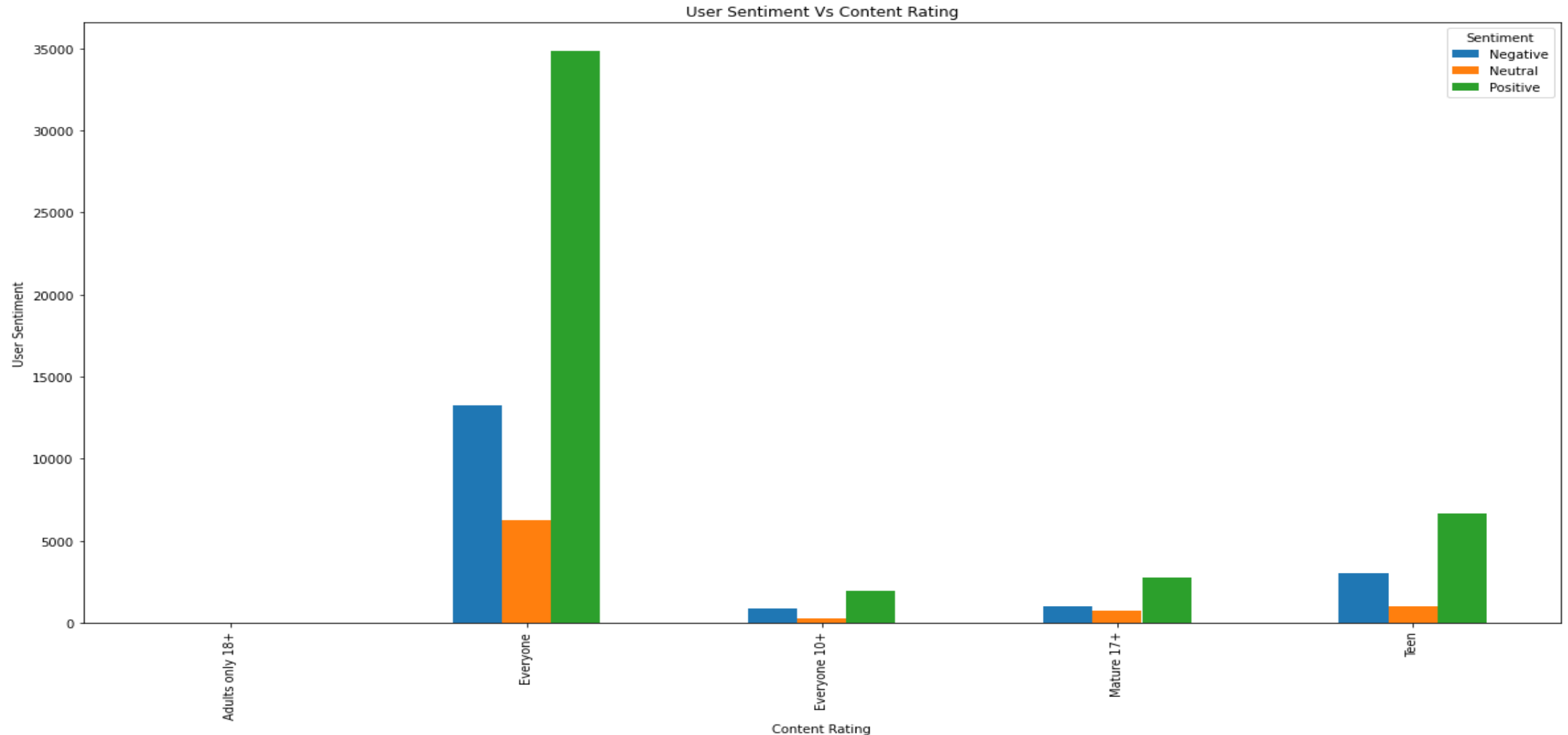
- It can be seen that more number of app reviews are subjective and positive in nature in comparison with other subjectivity and polarity parameters considered together.
- Least number of app reviews are objective and negative type



Content Rating and Type of Sentiment



- Apps made for everyone have the highest number of reviews with either neutral or positive sentiments



Thank You