Capstone Project

EDA capstone on Play Store App Review Analysis

By - Dipanshu Kumar

Content:

- Understanding Data Analysis
- Problem Statement and Challenges
- Steps involved
- Data cleaning
- Purpose of this Project
- Data visualization with Plots.
- Conclusion

- **What Is Data Analysis?**

- Data analytics is the science of analyzing raw data to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithm that work over raw data for human consumption.

Data Analysis Steps :

1. The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demographic, income, or gender. Data values may be numerical or be divided by category.

2. The second step in data analytics is the process of collecting it. This can be done through a variety of sources such as computers, online sources, cameras, environmental sources, or through personnel.

3. Once the data is collected, it must be organized so it can be analyzed. This may take place on a spreadsheet or other form of software that can take statistical data.

4. The data is then cleaned up before analysis. This means it is scrubbed and checked to ensure there is no duplication or error, and that it is not incomplete. This step helps correct any errors before it goes on to a data analyst to be analyzed.

World of Applications

- No, doubt Mobile Applications are trending, ever wonder why? The reasons are well known:
- Makes life easy
- Apps are fun
- Communication and Engagement
- Smart marketing, Lead generations and many more.

Problem Statements:

- Mobile applications have been one of the main drivers of the technological development of companies. The reception of mobile technology is the result of the recognition of the benefits that mobile applications provide to companies, which can certainly be used by companies of any size.
- However, not everything is rosy. Among so many apps only a third of mobile applications are established in the market. The hard truth is one out of every three developments is successful, this implies that companies must be aware of the risks, but more importantly of the way in which these risks can be identified and diminished.
- For which it includes Exploratory Data Analysis to highlight the trend and provide a brief understanding.

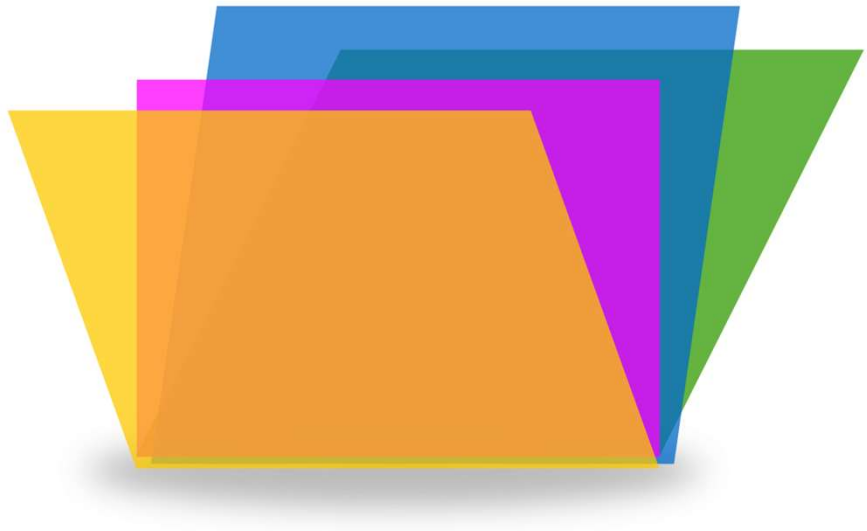
CSV Files

1.Play Store Data

- App: Name of the App.
- Category: Category under which it falls.
- Rating: Application's rating on play store.
- Reviews: No. of reviews of the app.
- Size: Size of the app.
- Installs: No. of Installs of the app.
- Type: If the App is free/ paid.
- Price: Price of the App (o if it is free).
- Content Rating: Appropriate target audience of the App.
- Genres: Genres under which the App falls
- Last Updated: Date when the App was last updated.
- Current Version: Current version of the App.
- Android Version: Minimum Android version required to run the App.

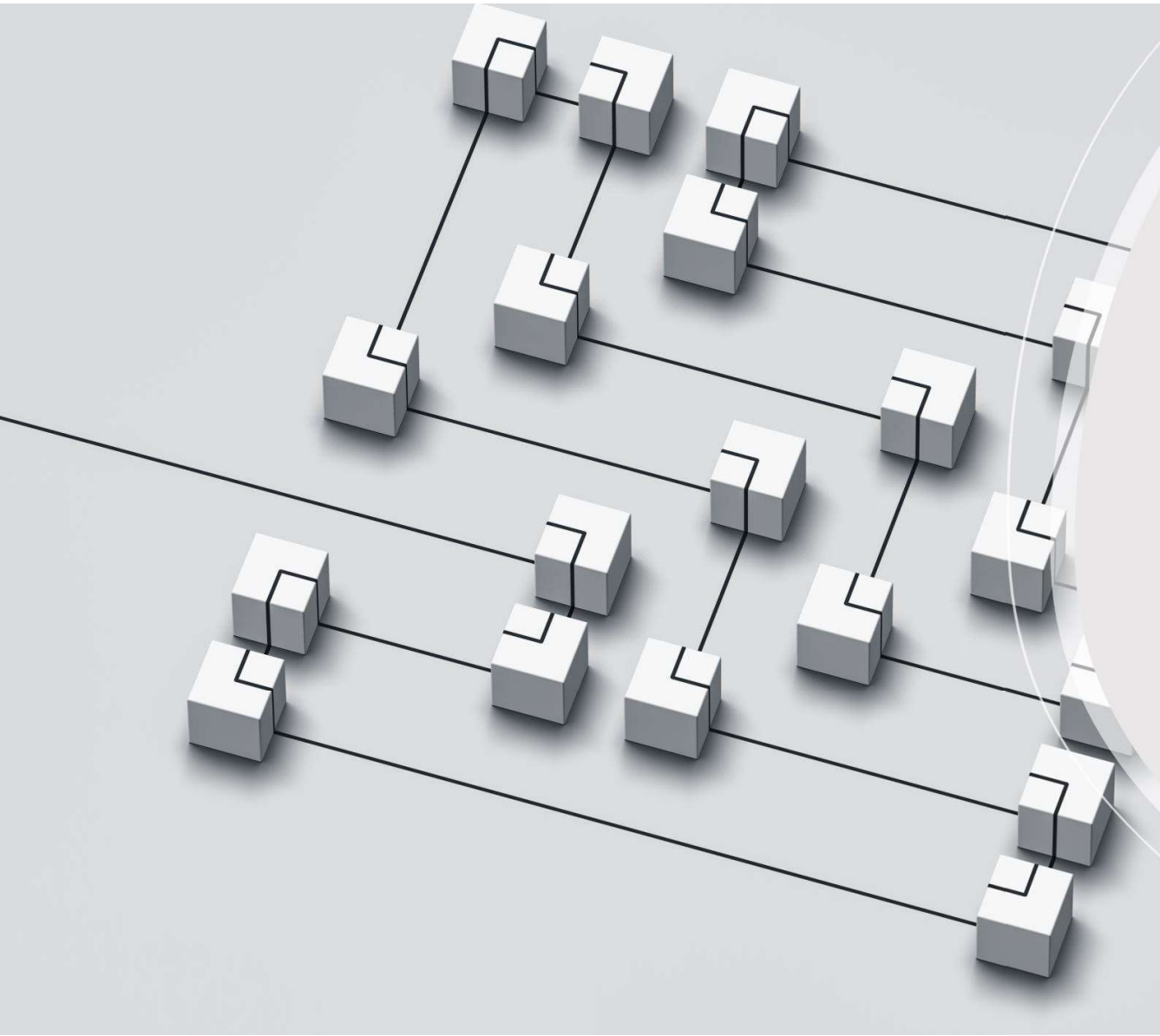
2.User Reviews

- Translated Reviews : To think or talk about something again and that's it(here we are taking about reviews corresponding to the apps)
- Sentiments : An attitude or opinion that is often caused or influenced by emotion here it has three categories we'll discuss this in an interesting manner.
- Sentiment Polarity : For an element it defines the orientation of the expressed sentiment.
- Sentiment subjective : An objective sentence expresses some factual information about the Apps.



What is data cleaning?

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.



- **Step 1: Remove irrelevant observations:**

- Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations. Duplicate observations will happen most often during data collection. When you combine data sets from multiple places, scrape data, or receive data from clients or multiple departments, there are opportunities to create duplicate data.

- **Step 2: Fix structural errors :**

- Structural errors are when you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization. These inconsistencies can cause mislabeled categories or classes. For example, you may find “N/A” and “Not Applicable” both appear, but they should be analyzed as the same category.

Step 3 : Handle missing data

- You can’t ignore missing data because many algorithms will not accept missing values. There are a couple of ways to deal with missing data. Neither is optimal, but both can be considered.

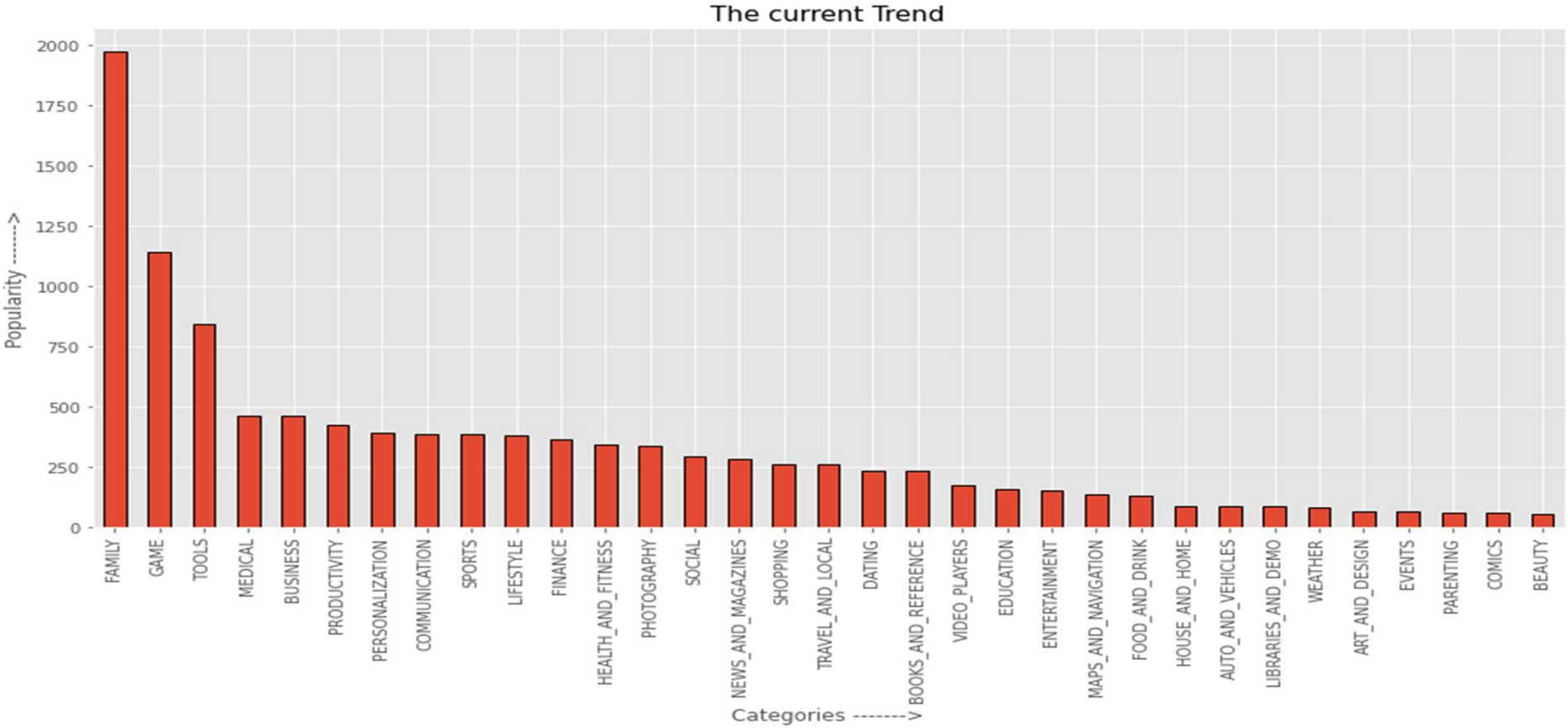
The Impact of Reviews

Reviews not only have the power to influence consumer decisions but can strengthen a company's credibility. Reviews have the power to gain customer trust, and they encourage people to interact with the company. Customer interaction ultimately leads to improved profits for businesses.

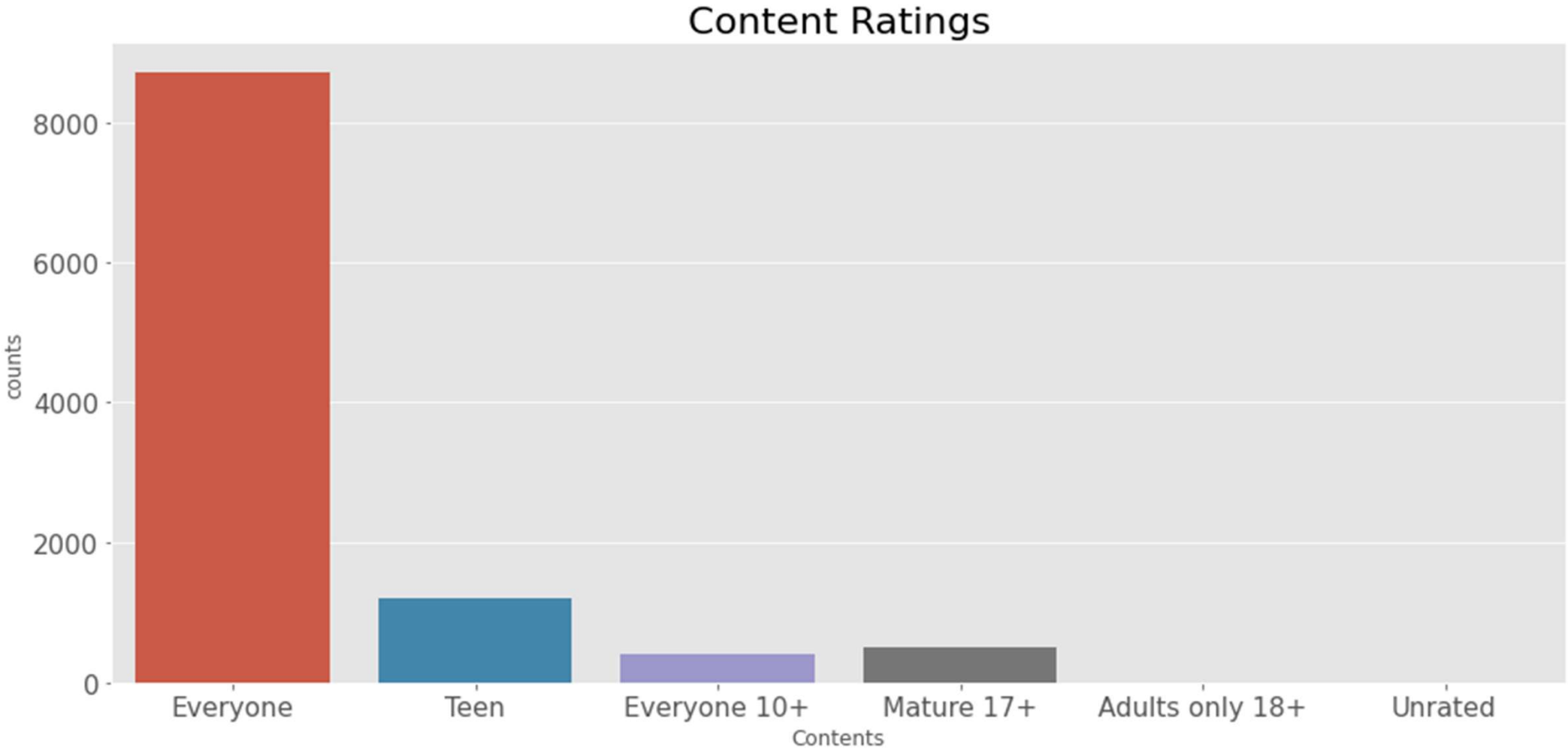
Purpose of this project

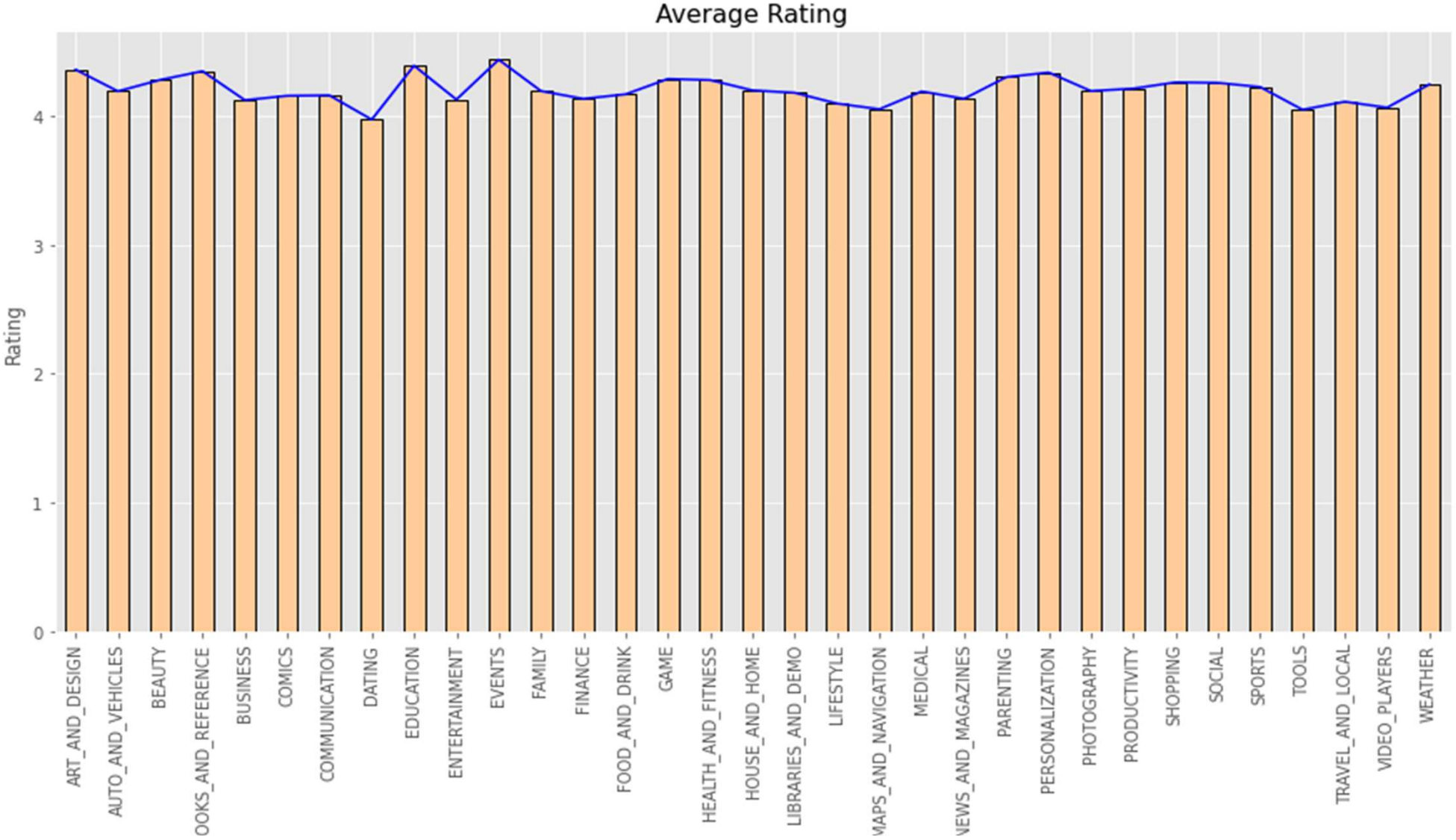
This EDA project aims to discover patterns that lead to a successful application on the Play Store. This will be done by analyzing the historical data collected from the Play Store according to dataset source. We hypothesize that there are patterns within the data that lead to a successful app. We are trying find the successful pattern to develop an app that may be in the high ranks one day, help ad providers know which apps to post their adds on.

BARPLOT ANALYSIS

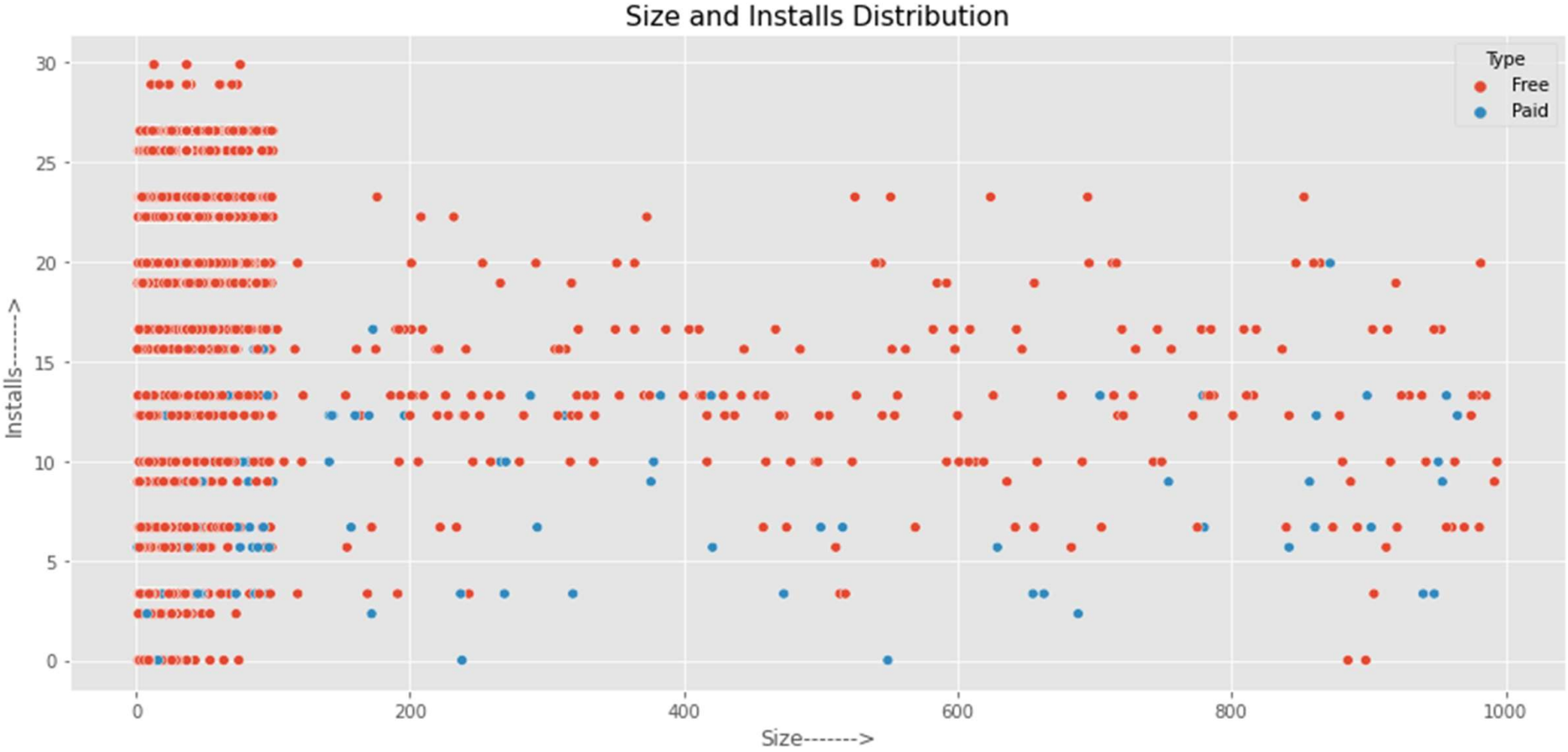


COUNTPLOT ANALYSIS

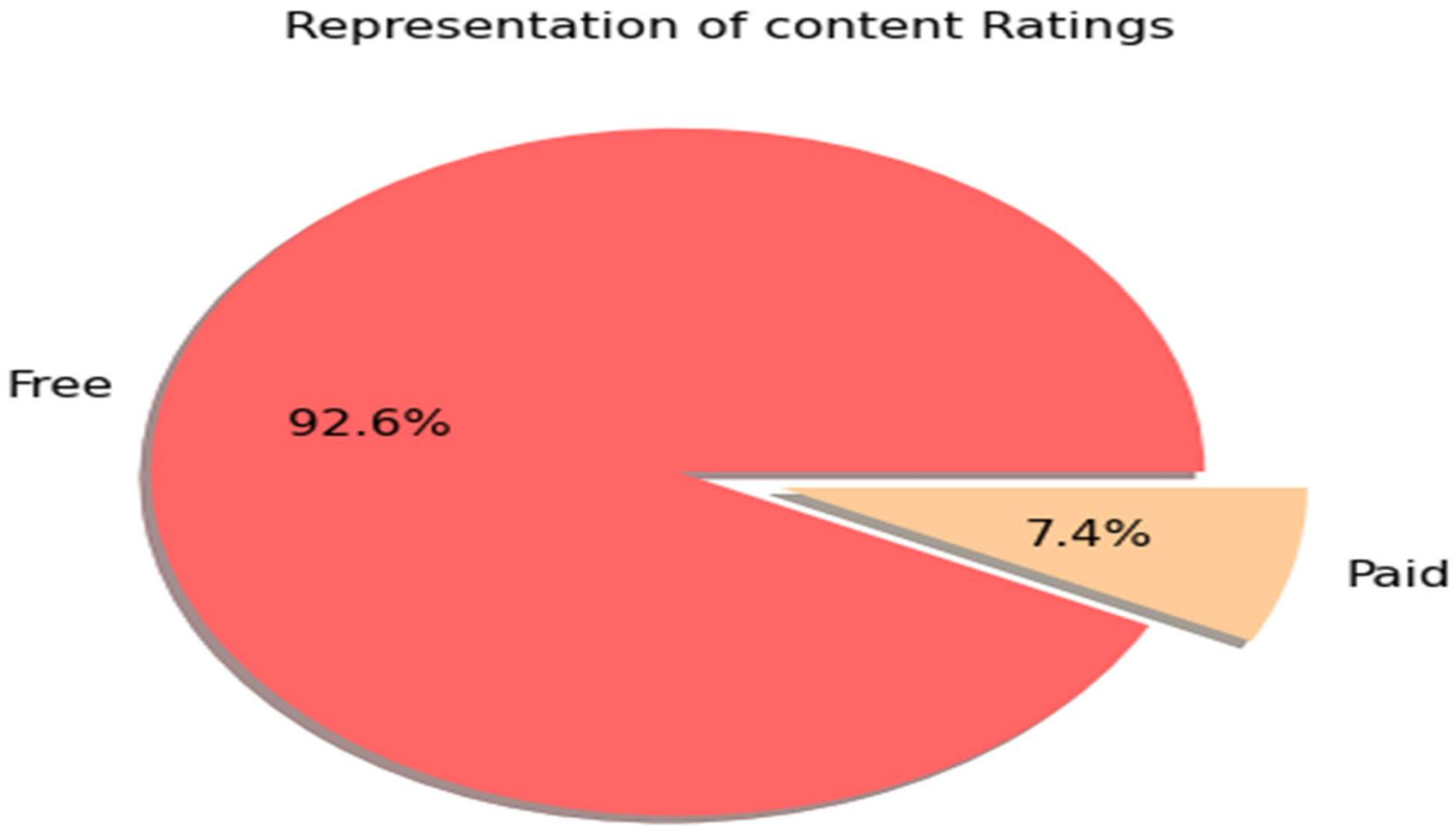



Average Rating
Rating
0
1
2
3
4
ART_AND_DESIGN
AUTO_AND_VEHICLES
BEAUTY
OOKS_AND_REFERENCE
BUSINESS
COMICS
COMMUNICATION
DATING
EDUCATION
ENTERTAINMENT
EVENTS
FAMILY
FINANCE
FOOD_AND_DRINK
GAME
HEALTH_AND_FITNESS
HOUSE_AND_HOME
LIBRARIES_AND_DEMO
LIFESTYLE
MAPS_AND_NAVIGATION
MEDICAL
NEWS_AND_MAGAZINES
PARENTING
PERSONALIZATION
PHOTOGRAPHY
PRODUCTIVITY
SHOPPING
SOCIAL
SPORTS
TOOLS
TRAVEL_AND_LOCAL
VIDEO_PLAYERS
WEATHER

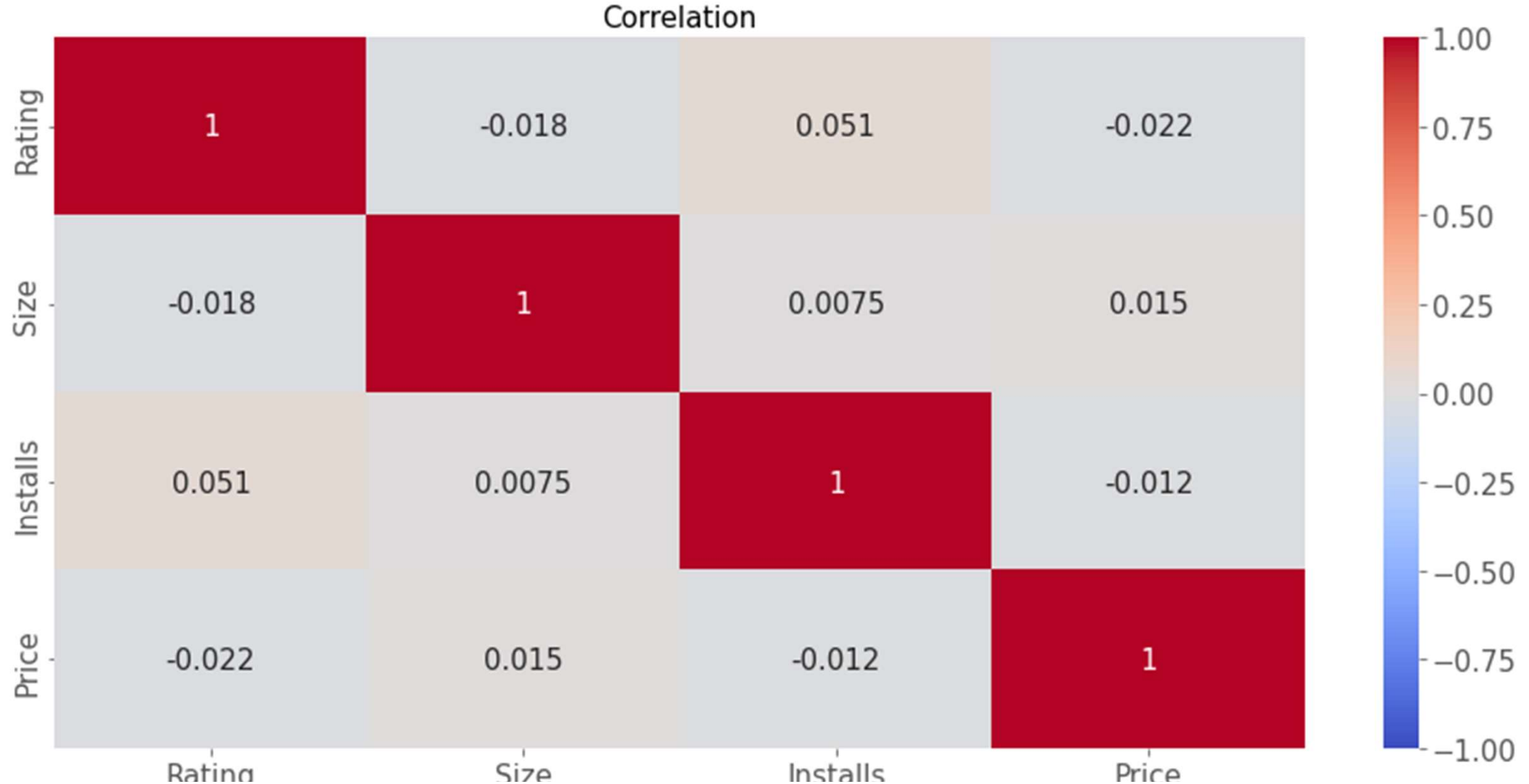
SCATTERPLOT ANALYSIS



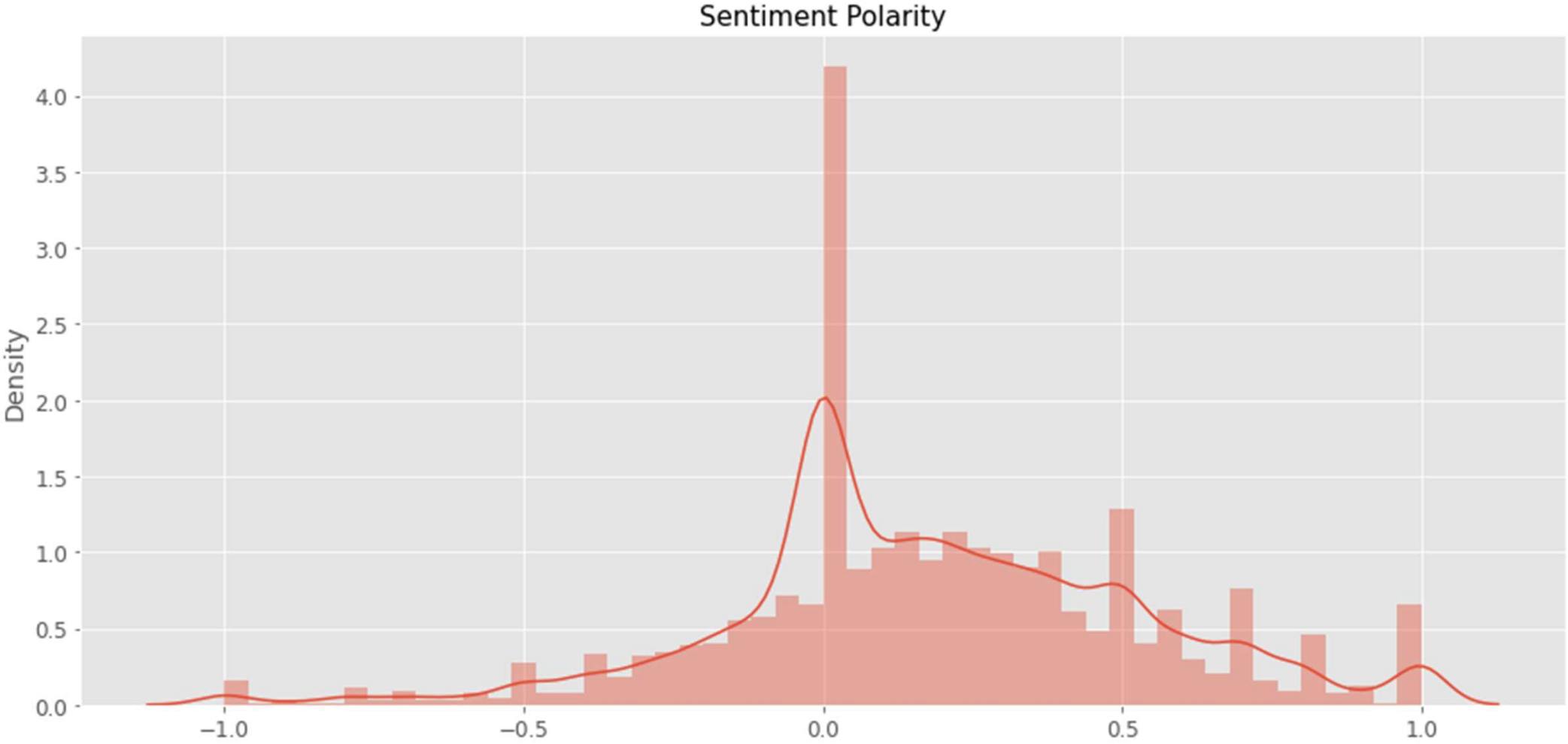
PIECHART ANALYSIS

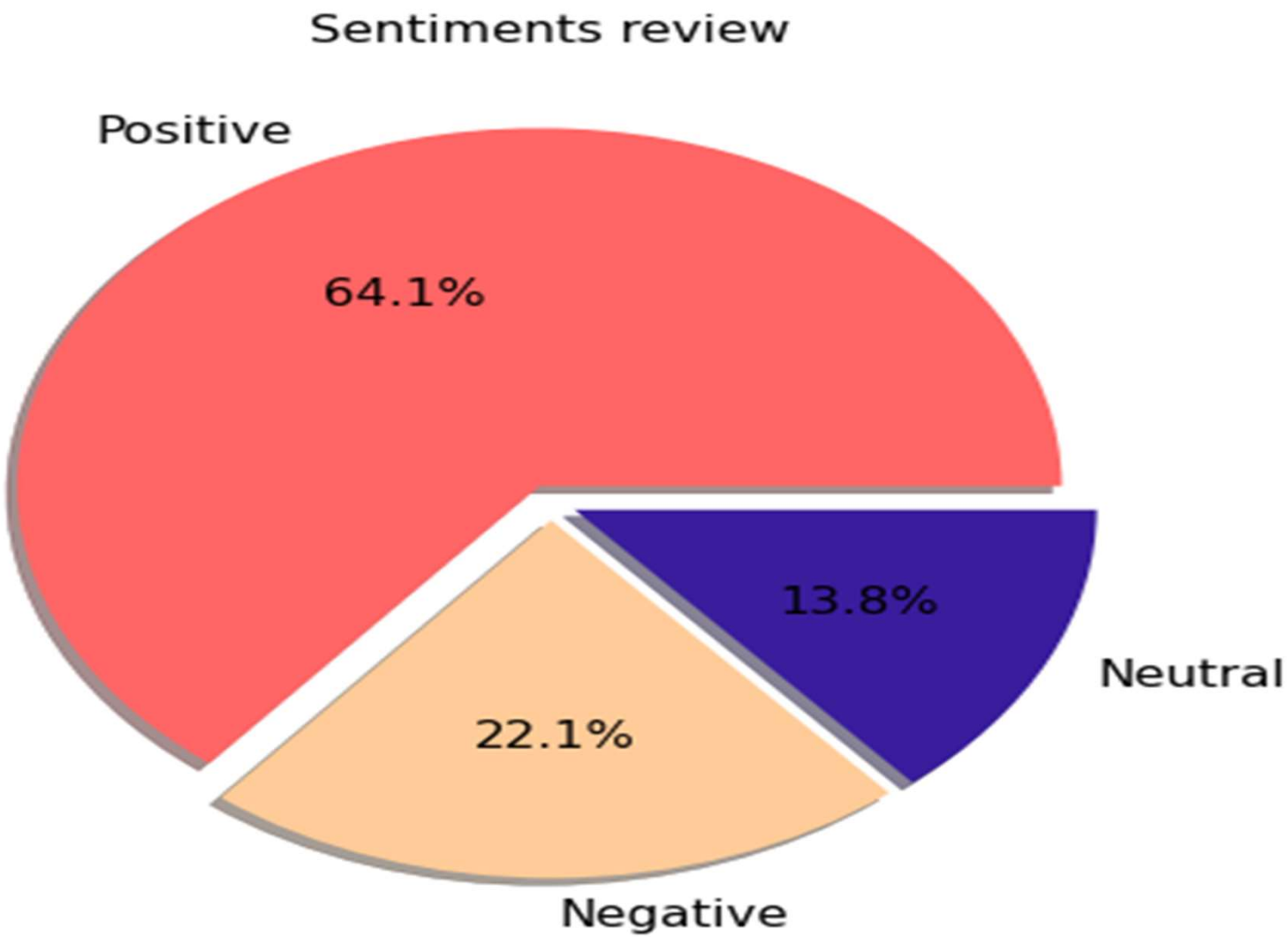


HEATMAP ANALYSIS

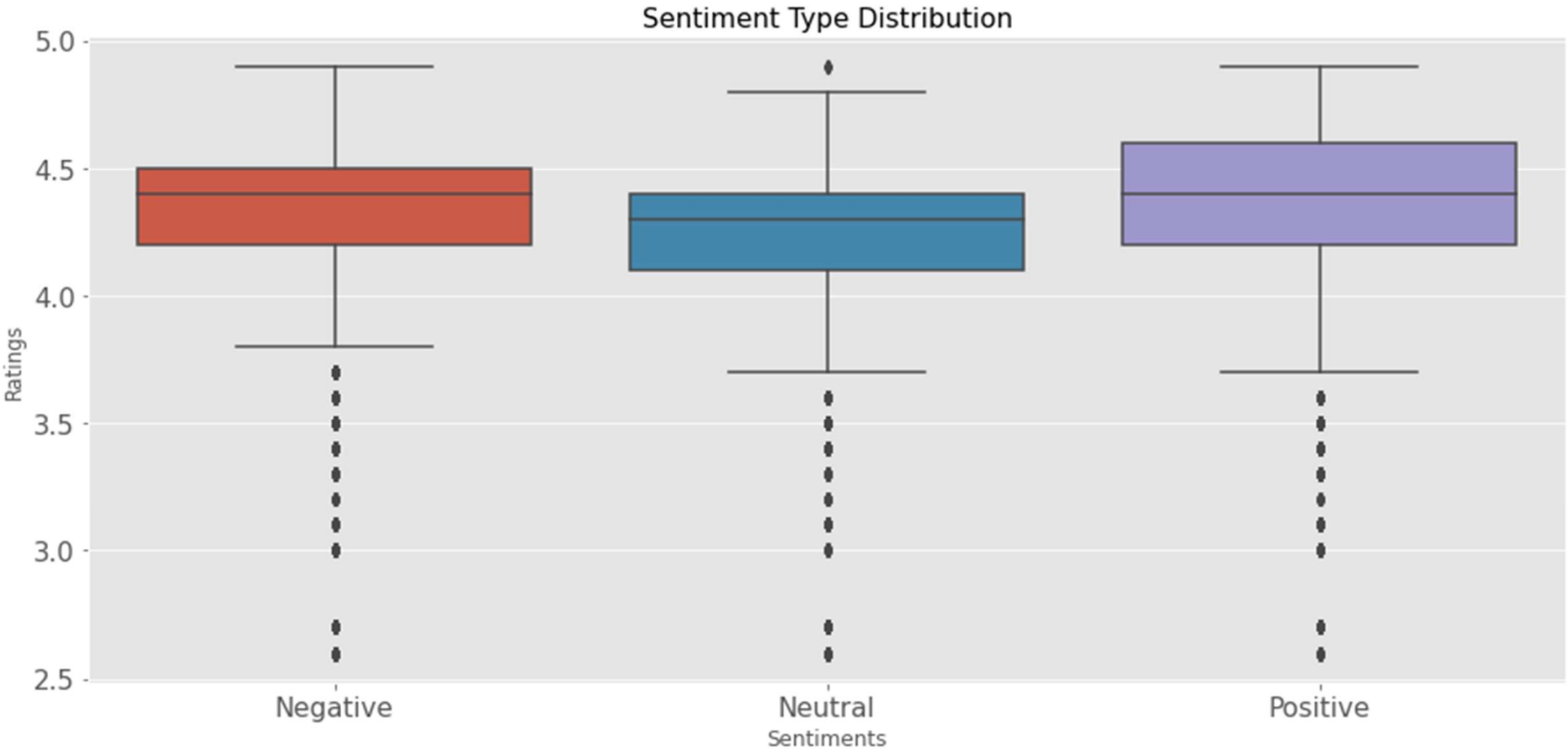


DISTPLOT ANALYSIS




Sentiments review
Positive
64.1%
13.8%
Neutral
22.1%
Negative

BOXPLOT ANALYSIS



Conclusion:

- The major purpose of this study was to evaluate the impact of the apps reviews, where a number of models were proposed, intended to explain the relationship between the apps and the users which provides knowledge management practice.

- Significant and positive results were found. We have seen that what ever are the category i.e. genre the apps are built, people will review and rate it positively. It totally depends on the organization to make it engaging, people will move towards those apps which came with something interesting.

- Also positive reviews have the power to motivate the developers to be more innovative to enhance retention and loyalty.

Thank You