Capstone Project-1

Play Store App Review Analysis

Team Members

Naman Jain and Prachi Jain

Overview of Play Store app Analysis:

1. Introduction
2. Abstract
3. Problem Statement
4. Data Set
5. Pre Processing
6. Steps Involved
7. Exploratory Data Analysis
8. Conclusion

INTRODUCTION:

- Play Store is an online shop where users can discover and enjoy their favourite applications, games, movies, TV episodes, books, and other content for their Android devices.
- It contains more than 3.5 million Android Apps and has a user base of over 1 billion people.
- More than 48 billion apps have been installed from this store.

ABSTRACT:

- Google Play is a digital distribution service operated and developed by Google Inc. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android SDK and published through Google. Google Play also serves as a digital media store, offering music, books, movies, and television programs.

- This report employs multiple Python tools, including NumPy, pandas, matplotlib, and seaborn, to analyse numerous Play Store applications

PROBLEM STATEMENT:

- To explore and analyse the data in order to identify:
- Major highlights about the applications available in the Play Store.
- The variables responsible for app engagement and success.

Our goal is to help android developers to know what is the motivating factor for people to download an app. It will also help to find out the factors that affect someone’s decision to download an app. I would like to analyze categories, reviews, prices, ratings, and installs for this purpose and find out how they are interrelated.

DATA SET:

APPLICATION DATA

- **App:** Contains the names of the applications
- **Category:** Information about the category in which the application may be included to
- **Rating:** Average user rating of the application
- **Reviews:** Count of user reviews received by the application
- **Size:** Application size
- **Installs:** Rounded of values about the total installs of every application
- **Type:** Paid or Free
- **Price:** Cost of the application
- **Content Rating:** describes the maturity level of the content in an app
- **Genres:** Similar to category contains multi-category information about the application content
- **Last Updated:** Date on which the latest update of the application was rolled out
- **Current Ver:** Latest version of the app
- **Android Ver:** OS information on or above which the application can be used

REVIEW DATA:

- **App:** Application name
- **Translated_Review:** Processed reviews of the users which is done using steps involved with text pre-processing
- **Sentiment:** Positive or Negative depending on the tone of the review
- **Sentiment_Polarity:** Defines the inclination of the sentiment
- **Sentiment_Subjectivity**: Presenting the Personal or factual context of the review

PRE PROCESSING:

Variables removal or Data cleaning:

- Elimination of useless details such as Last Updated, Current Ver, and Android version.
- Convert the data type of variables such as Reviews, Size, Installs, and pricing, data imputations such as the replacement of "k" with "1000", "M" with "1,000,000", the removal of ", "$,“.
- Conversion of the object to int or float.

Null Value Imputation:

- Mean, mode and Median imputation can be used but in this dataset, the null values have been removed as imputation will not do justice to the variables.

Reduced dataset Size

- Application dataset: 9145*10
- User review dataset: 37427*5.

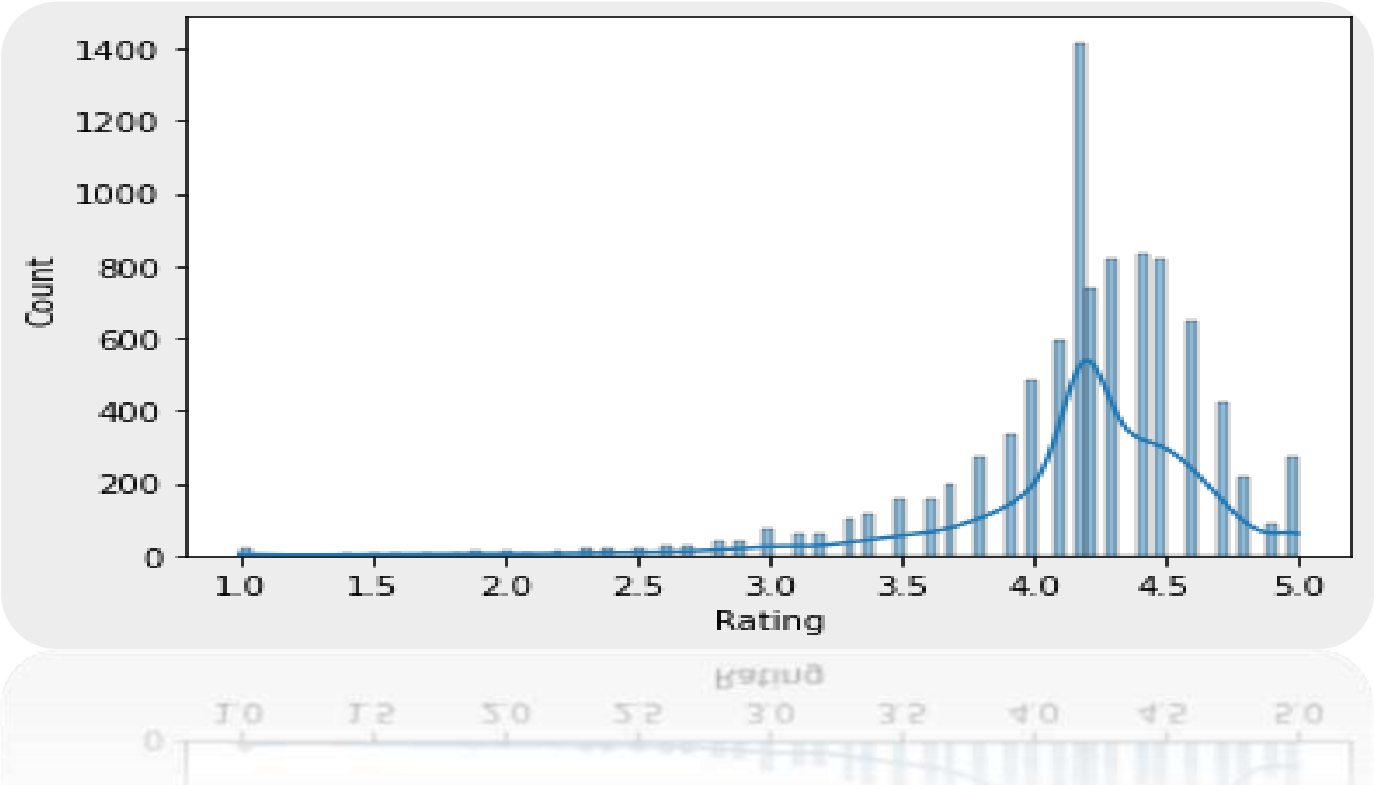
STEPS INVOLVED:

- 1. Data Wrangling
- 2. Performed Analysis
- 3. Checked correlation between each feature
- 4. Treatment of Null values
- 5. Treatment of Outliers
- 6. Explored Visualization
- 7. Transformation

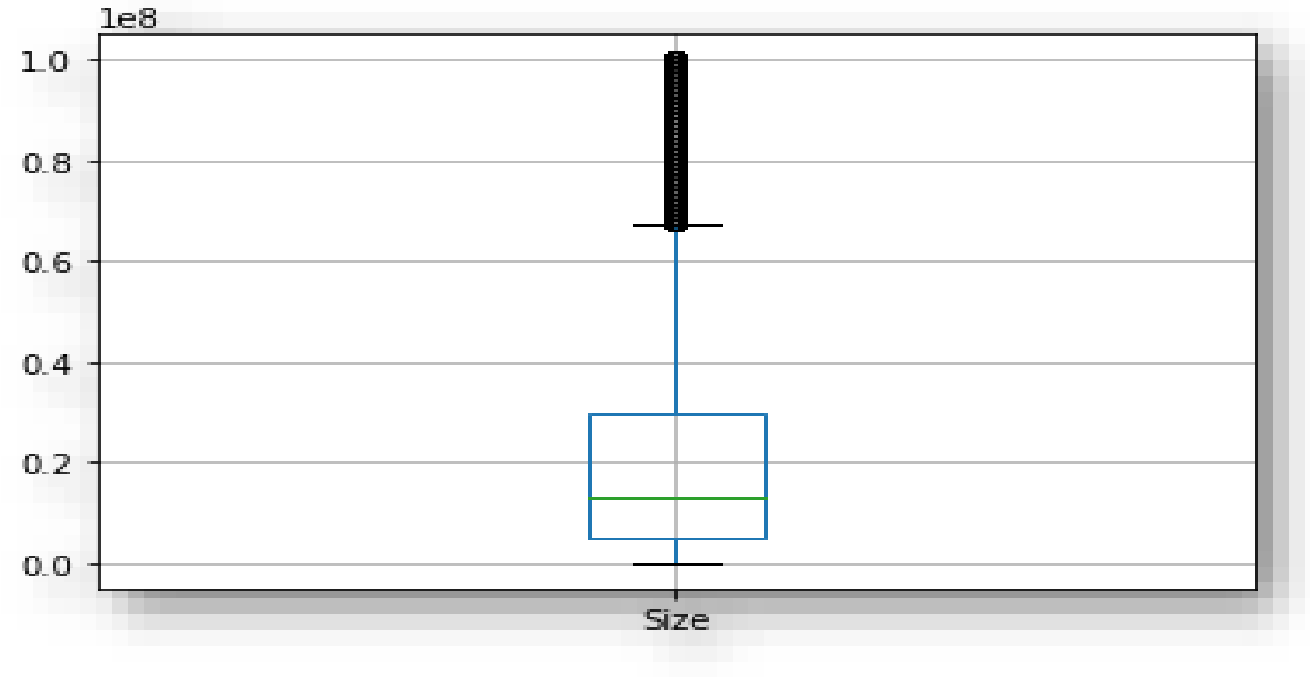
EXPLORATORY DATA ANALYSIS:

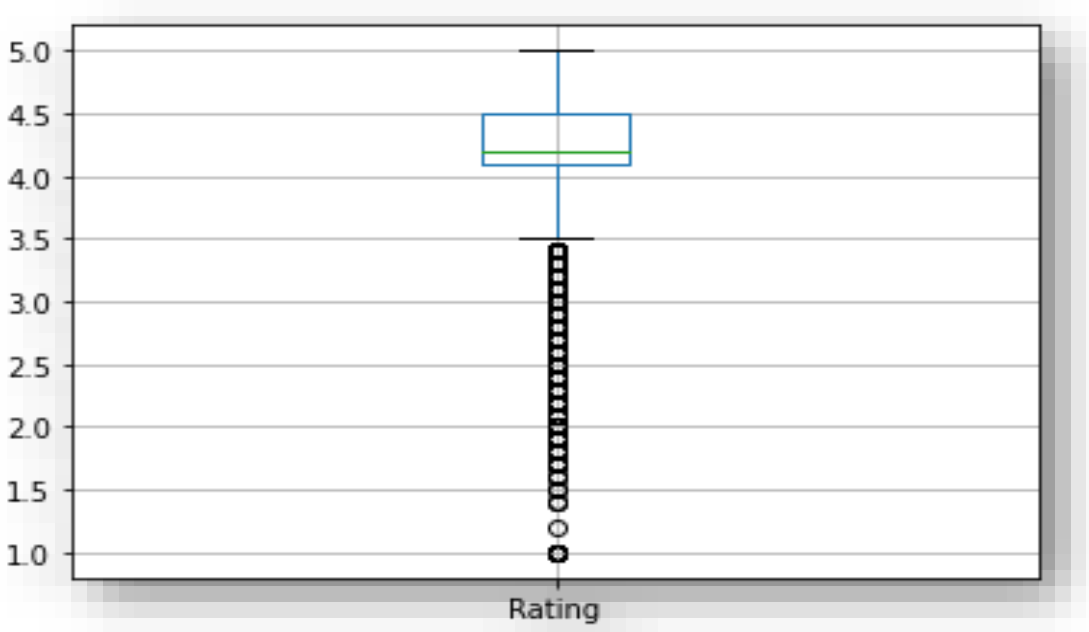
- Using the panda’s package, The data is loaded. Operations were performed such as casting the columns into their appropriate data types, to better comprehend the data.
- Statistical data, univariate analysis, and the accuracy of data in each column were all investigated.
- Understanding the relationship of all numerical and categorical columns. In addition, we performed Bivariate Analysis on both numerical and categorical data.
- Performed Univariate analysis for numerical columns to grasp distribution and information and see if it follows a Gaussian distribution.
- Used a variety of plots to gain a better understanding of the data and presented it in an attractive way.

User Rating is Positively skewed with the mean at 4.1 with the highest possible rating being 5.



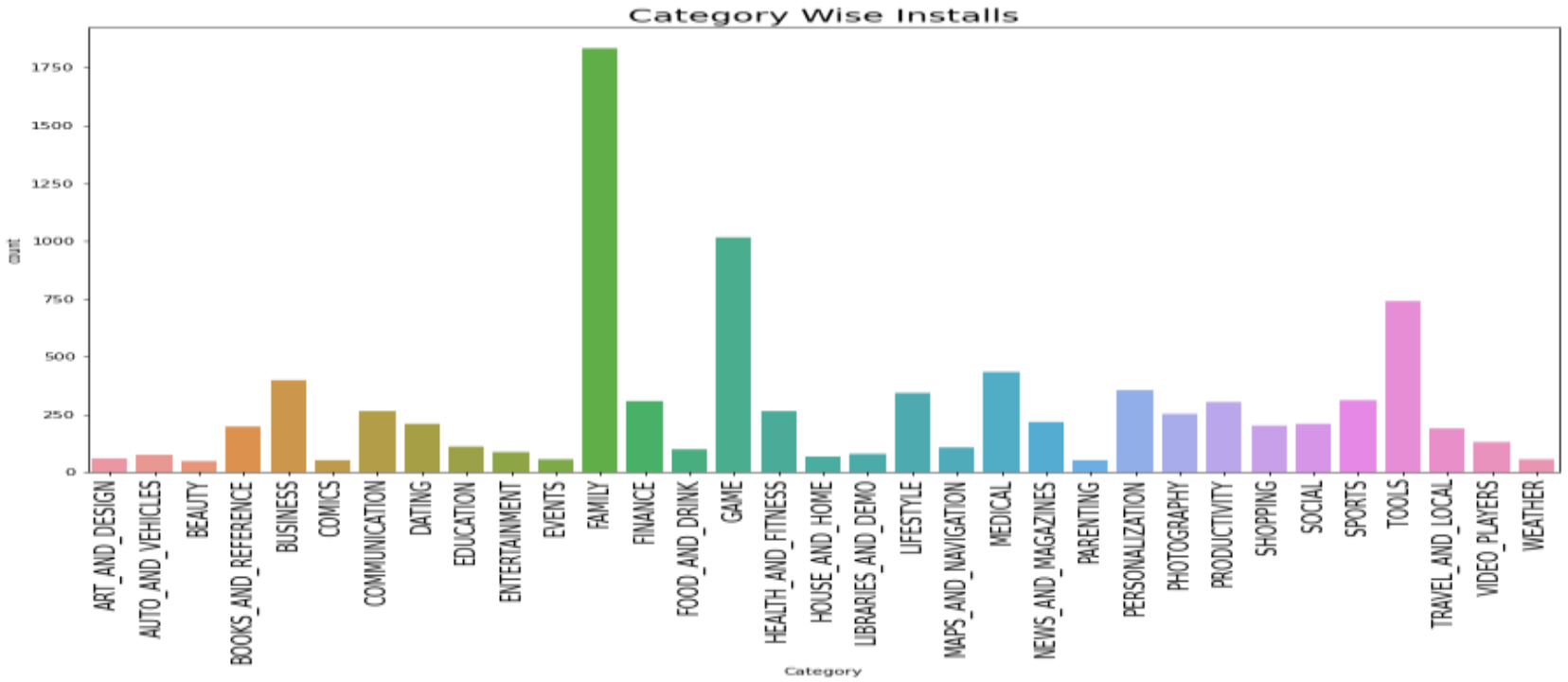
1.Size is an important factor that is taken into consideration
2.Users are reluctant about downloading an application of a larger size

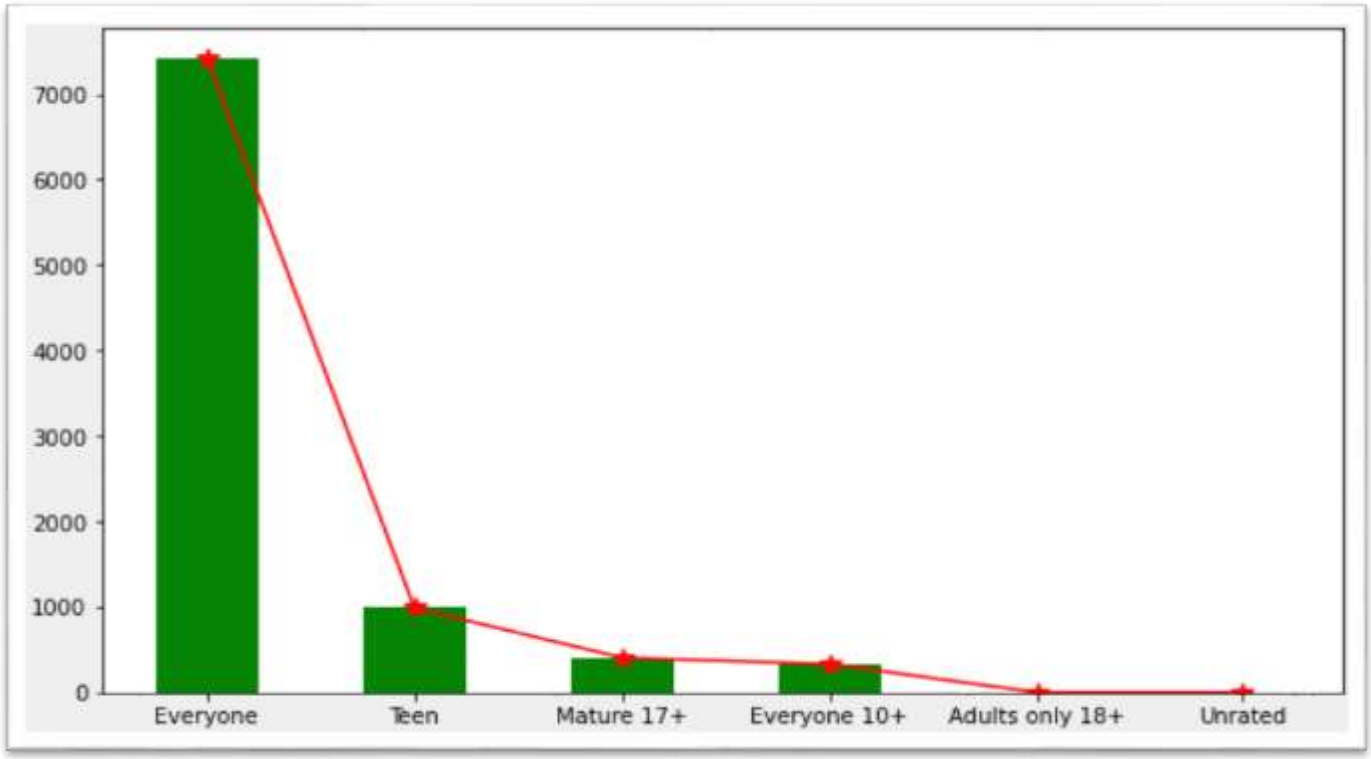




- First quartile has a rating of 4.1
- 25% of people rated the application to be between 4.1 and 4.2

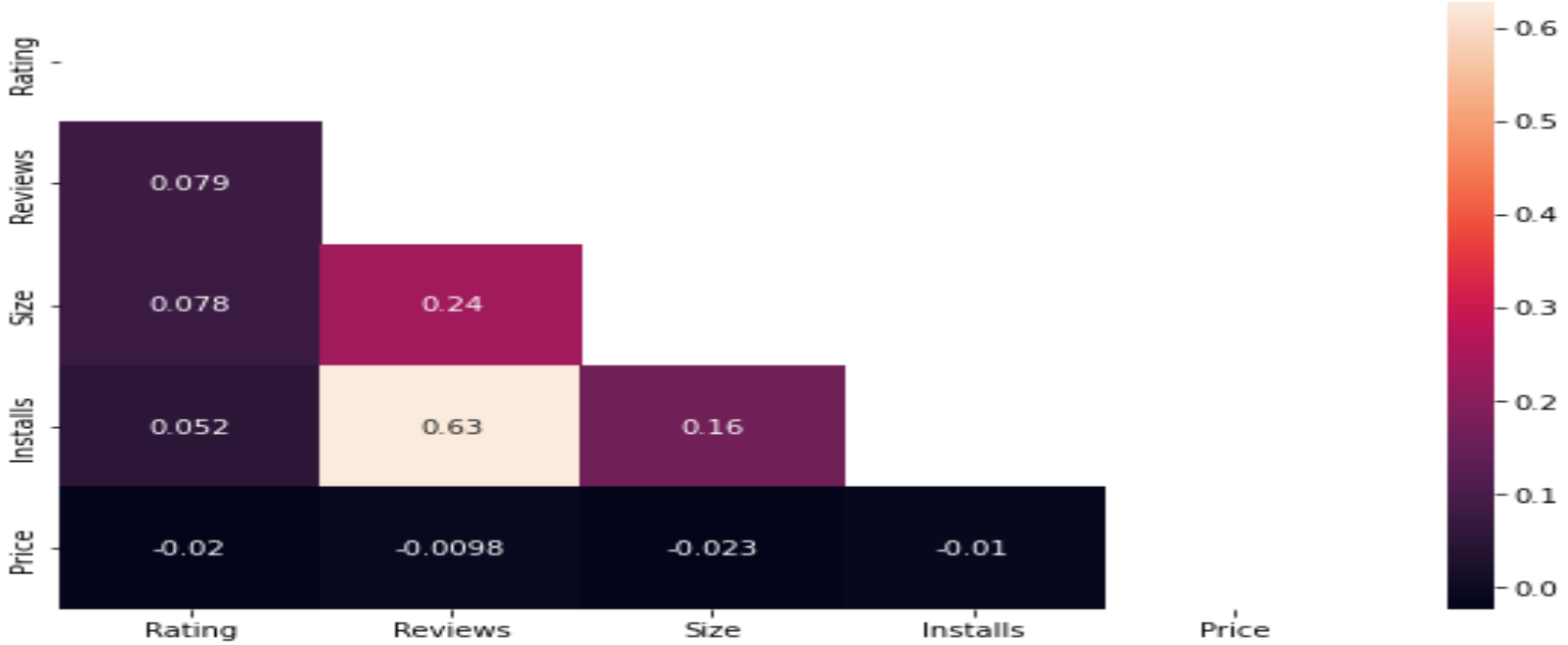
Most favourite application categories are Family, Games, and Tools.

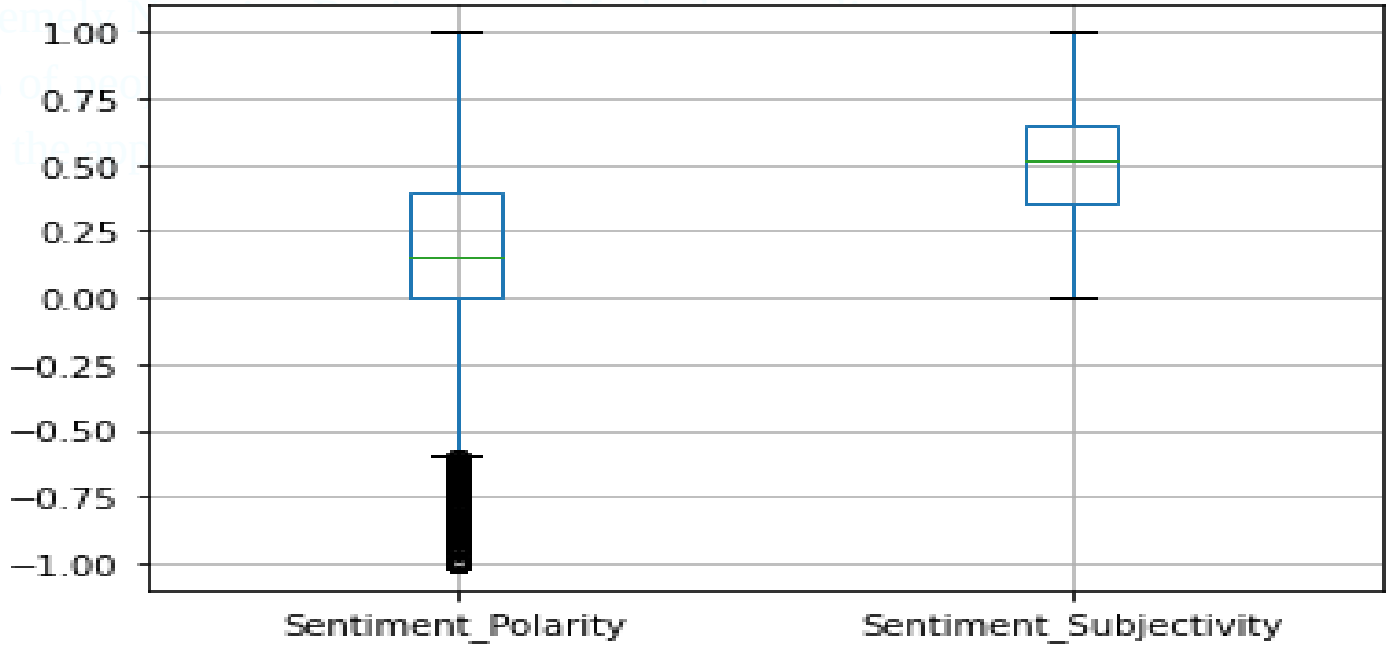




- More than 75% of the programs that can be downloaded from the Google Play store are categorized as suitable for use by everyone.
- Less than one percent of the apps specifically focus on users aged 18 or older

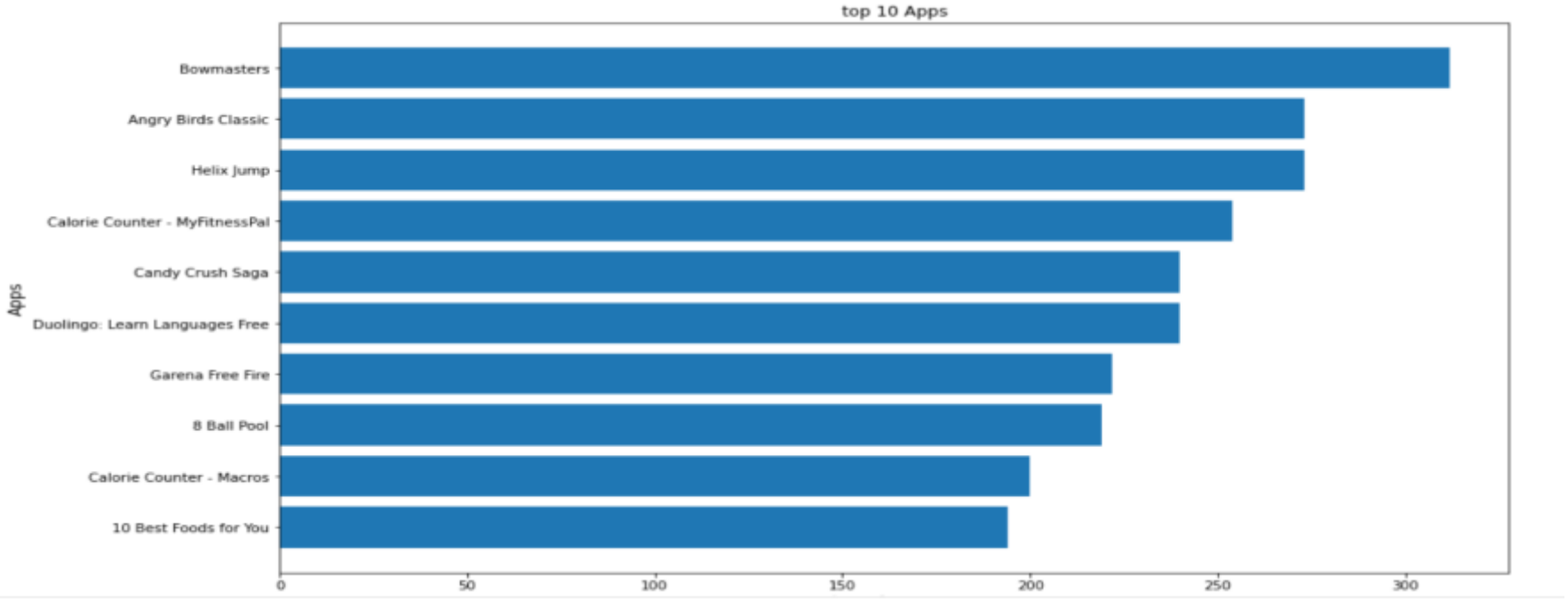
Out of numeric variables which are rating, reviews, size, installs, and price only Reviews and installs are correlated implying that users may be checking the user reviews while installing the app

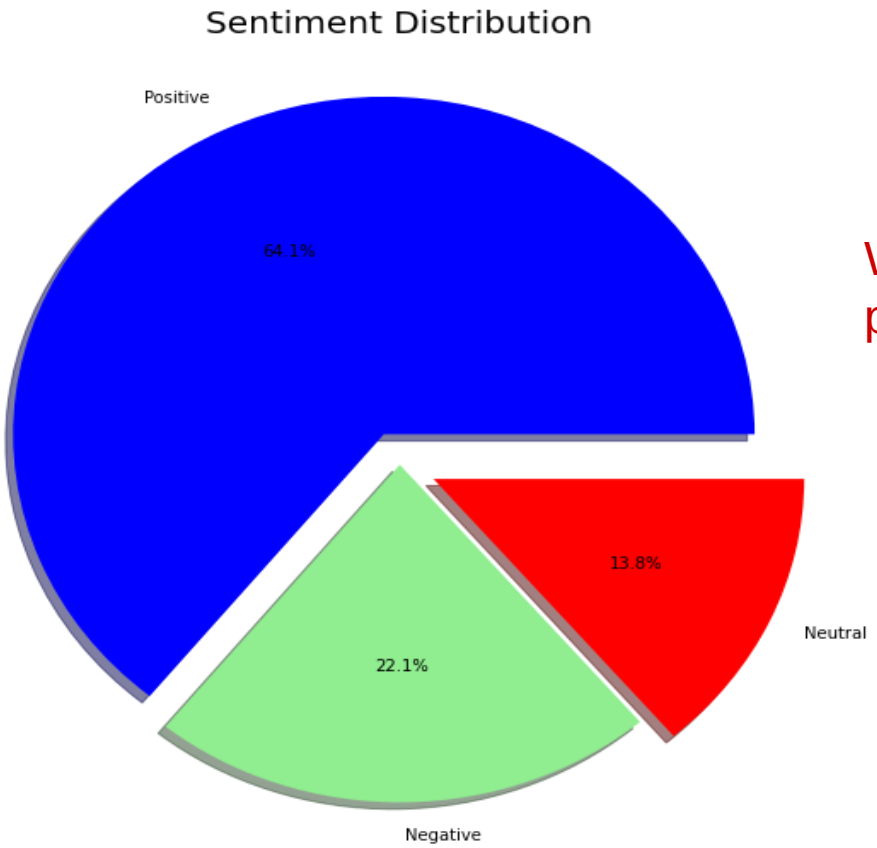




- Extremely Negative Reviews are Marked as outliers
- 25% of people are extremely positive with the application usage, and they are pleased with the application's performance

Top 10 Apps to their occurrence in the reviews





We can see that 64.1 percent got positive sentiments.

CONCLUSION:

From analyzing the data, we were able to deduce the answers to some of the most pressing and intriguing questions that could arise in the mind of an Android user.

Besides that, the data presented above demonstrates that high-quality apps with ratings above 4.0 are more likely to get more reviews and downloads. The figures illustrate that there are rises in both app size and cost, but this does not mean that highly rated apps are typically very large and expensive. Useful data about the Play Store's most downloaded apps can be found in the Google Play Store Apps report. They're probably useful or entertaining, which is why these apps have gained such a large user base. It also demonstrates that creators in these fields are prioritizing quality over quantity when it comes to app production.

THANK YOU