

Capstone Project - 1

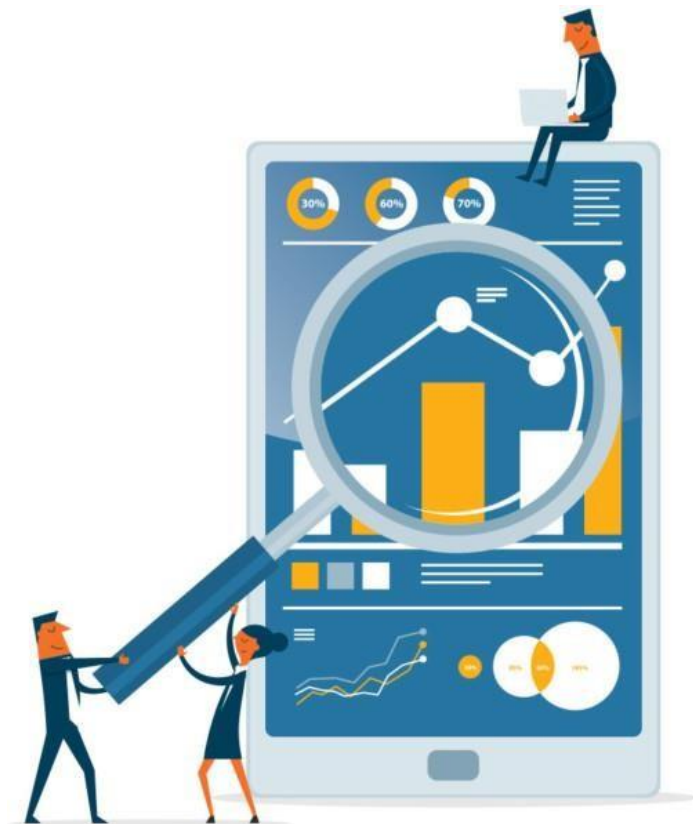
Play Store App Review Analysis

By:

- **Bhoopalan M.S**
- **Rajakumaran S**
- **Gowtham R**
- **Ranjith R**

Roadmap

- **Introduction**
- **Defining Problem Statement**
- **Exploratory data analysis**
- **Conclusion**



Introduction

➤ Importance of apps

- Explore to different application on single platform
- Multiple choices sophisticate different users

➤ Benefits to a business

- Attract the application developers
- User friendly with feedbacks

➤ Challenges and Opportunities

- Eliminate Offshoot apps
- Fake ratings escalate and deteriorate the app review

➤ Google Play Store dataset

- Apps and features
- User Reviews and feedback

Defining Problem Statement

To closely analyse the play store data in reducing error percentage, removing irrelevant data, better visualization with comparison, this is done by analysing in different phases like data extraction, data cleaning , data visualization, use of different models

Data Summary (for App and Features dataset)

This Dataset contains data on the Google Play Store applications. It has 10,841 rows of app and corresponding details in Following Columns:

App Category: Category of the app. like Family, Sports, Education, Entertainment, Shopping...etc.

Rating: This is How users rate the app out of 5, which has 1 means Lowest rating and 5 has the Highest rating.

Reviews: Total number of user reviews each app.

Size: The memory size the application.

Installs: Total number of times each application has been installed by users.

Data Summary (Cont.)

Type: Whether the app is free or a paid app.

Price: The price of the app to buy.

Content Rating: This column specifies who can use the app. it Can be for teens, mature audience, or everyone.

Genres: Sub category and purpose of the app.

Last Updated: last recent update for the app.

Current Ver: Current version of the app.

Android Ver: the app supported from which Android OS version

Data Cleaning

The first and most crucial step without which one cannot hope to proceed. Before diving straight away into the analysis part, it is crucial to make the dataset analysis-ready first. In this step, I have done the following:

1. Removing unnecessary rows and duplicates
2. Checking and converting the datatype of variables
3. Treatment of null values
4. Ensuring one row for each app

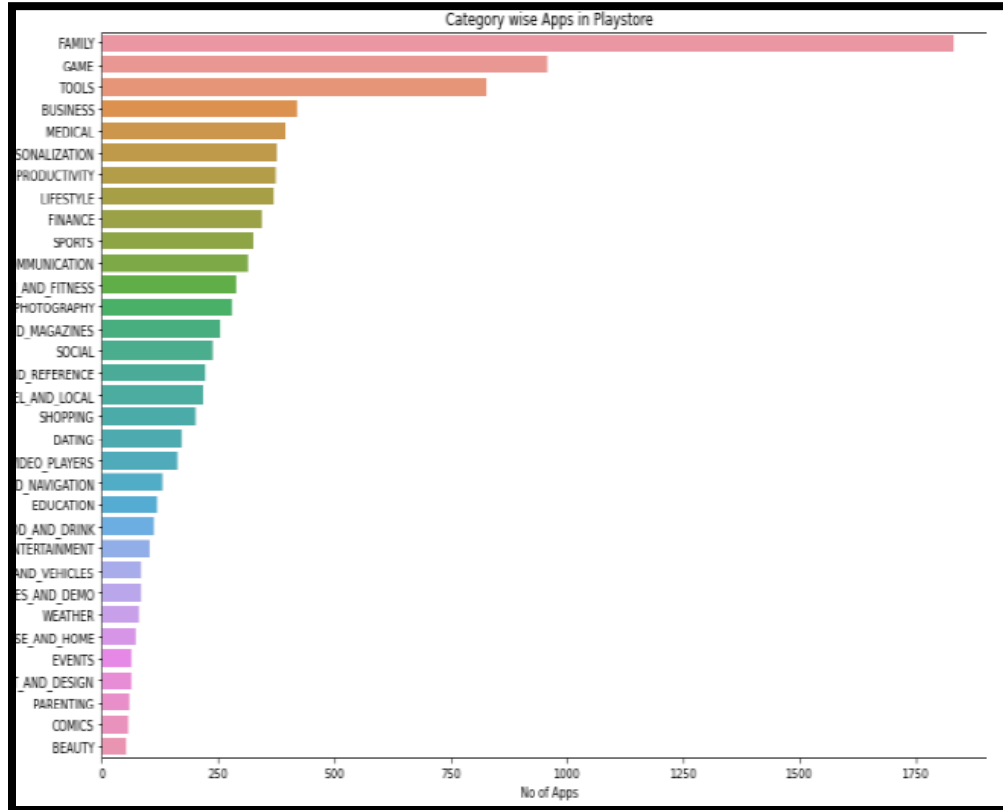


Exploration and Visualization

The next step involves basic exploration of the dataset, its variables; both dependent and independent, and the relation between them. In the next few slides, the visualization of the same has been covered.

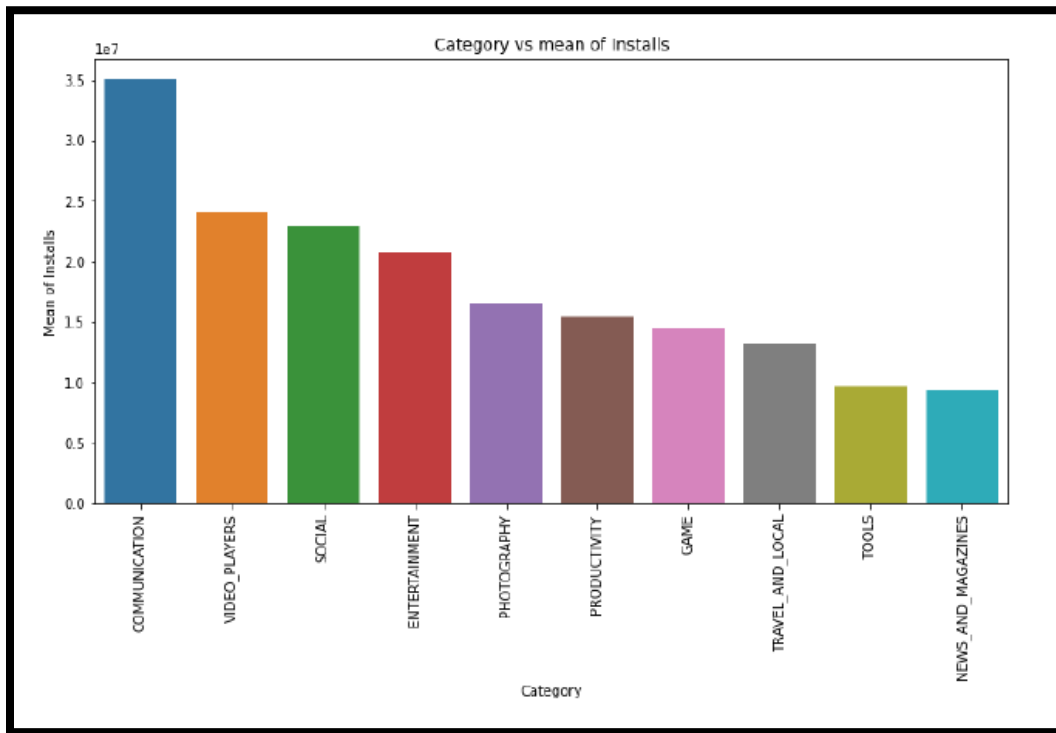


Exploration and Visualization-Top Categories in Play Store



- Maximum apps offered by: Family, Game, and Tools
- Least apps offered by: Parenting, Comics, and Beauty

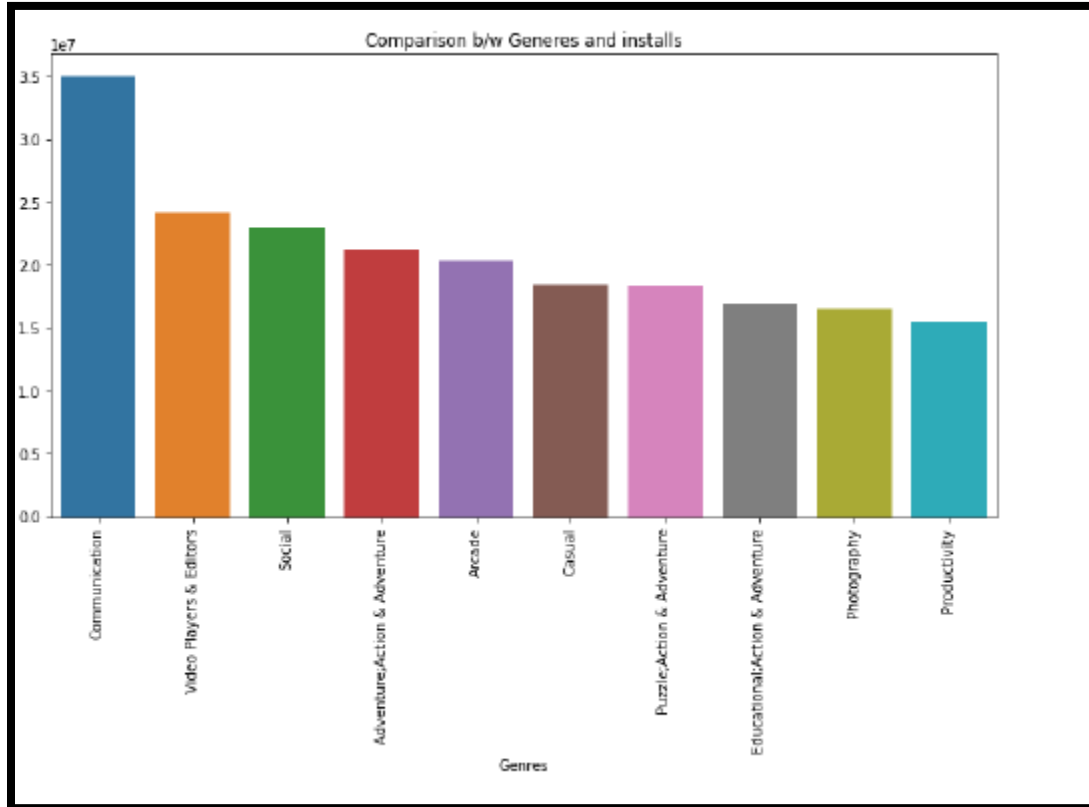
Exploration and Visualization- Categories vs Installation



- Most installed categories: Game, Communication, and Tools
- Least installed categories: Medical, Parenting, Events, and Beauty

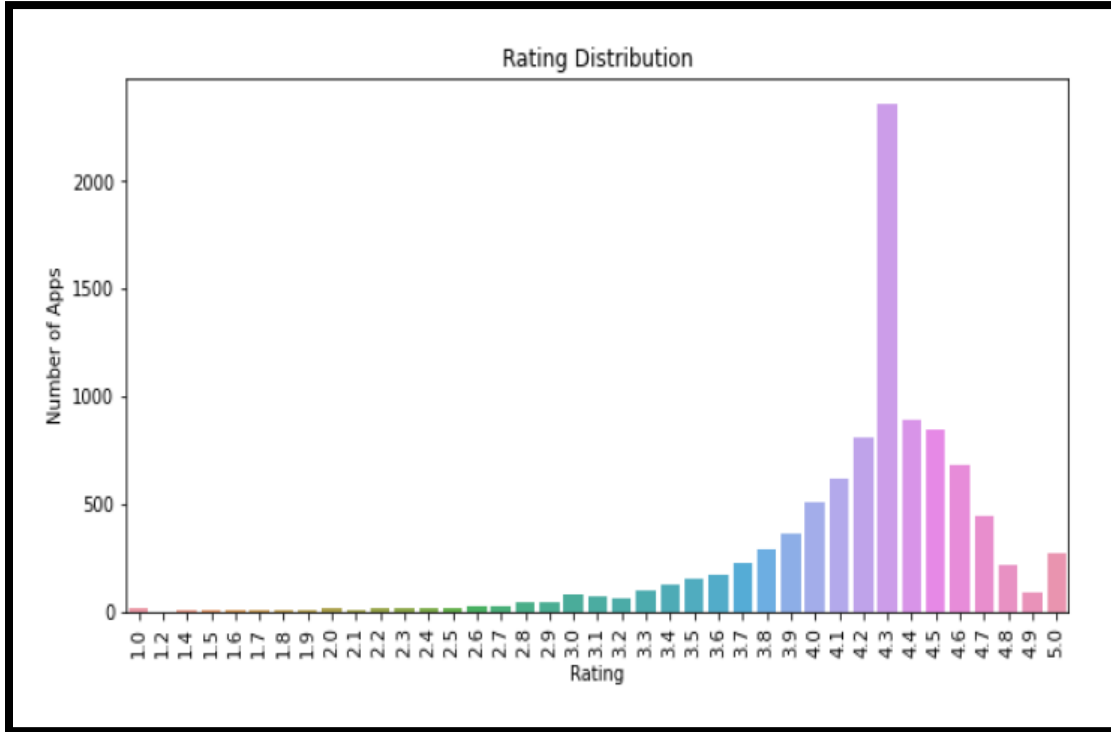
Exploration and Visualization

Genres vs Installation



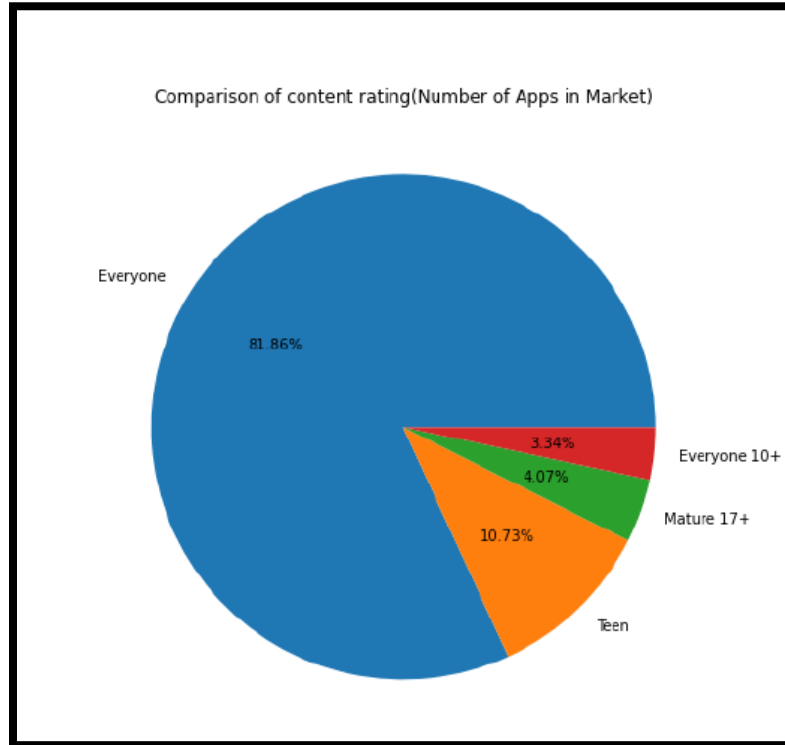
➤ Genre wise Most number of times users installed Communication Category apps and then Video Players, Social category and then Adventure; Action & Adventure etc.

Exploration and Visualization- Rating among apps



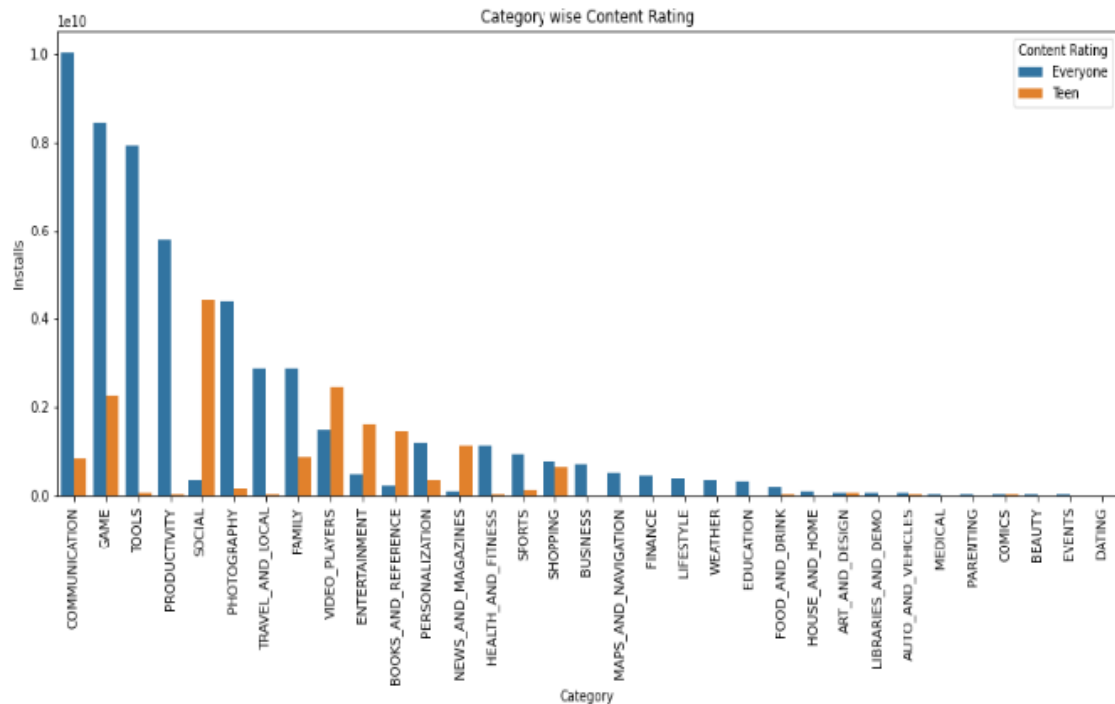
- Most of the apps lie between 4 and 5, and majority on 4.3

Exploration and Visualization- Content Ratings



- Everyone content Category having Highest number of apps and then Teen, Mature 17+, Everyone 10+

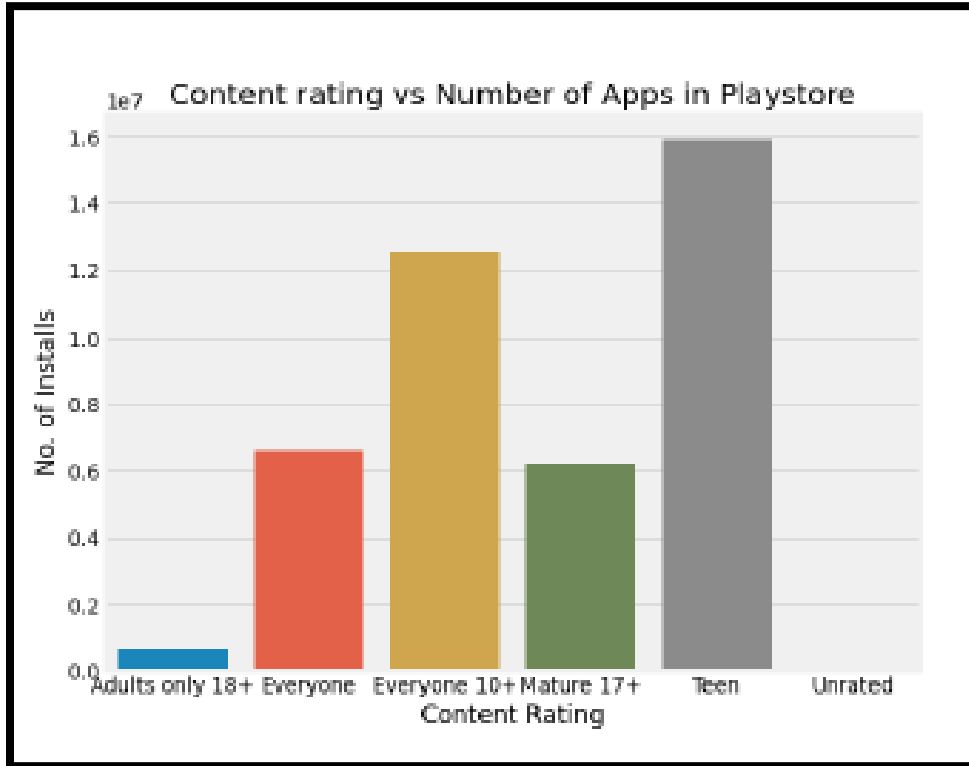
Exploration and Visualization- Category wise content rating



Everyone content
Category having Highest
number of installs of
Communication, game,
tools

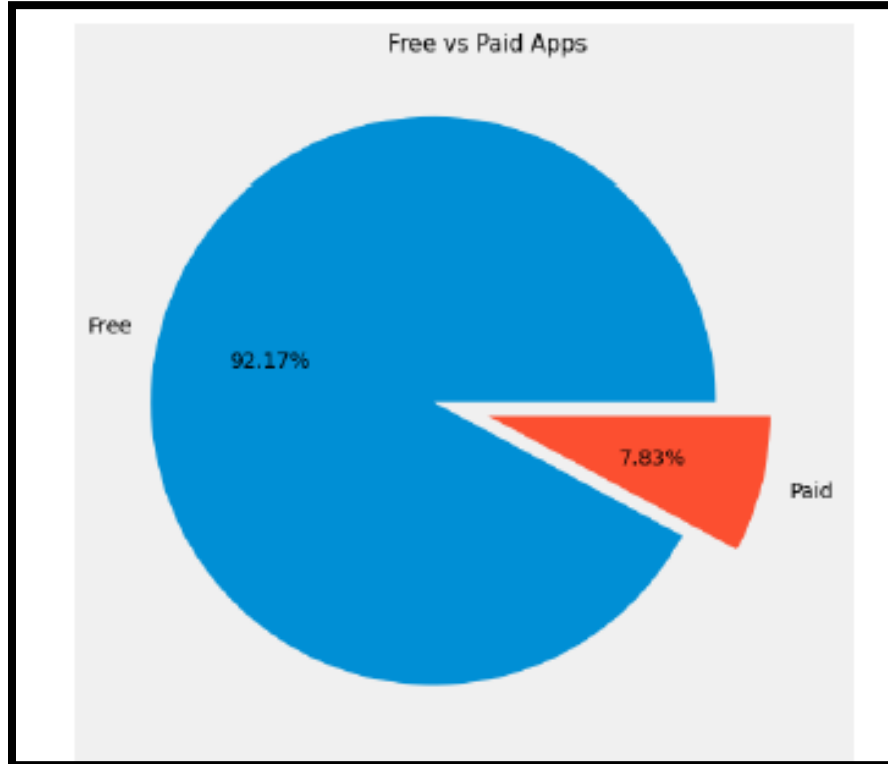
Teen content category
having highest no of
installs of Social, Video
player and Game.

Exploration and Visualization - Content rating vs Installs



- Teen and Everyone content Category having Most number of apps installed by users and then Everyone, Mature 17+, etc

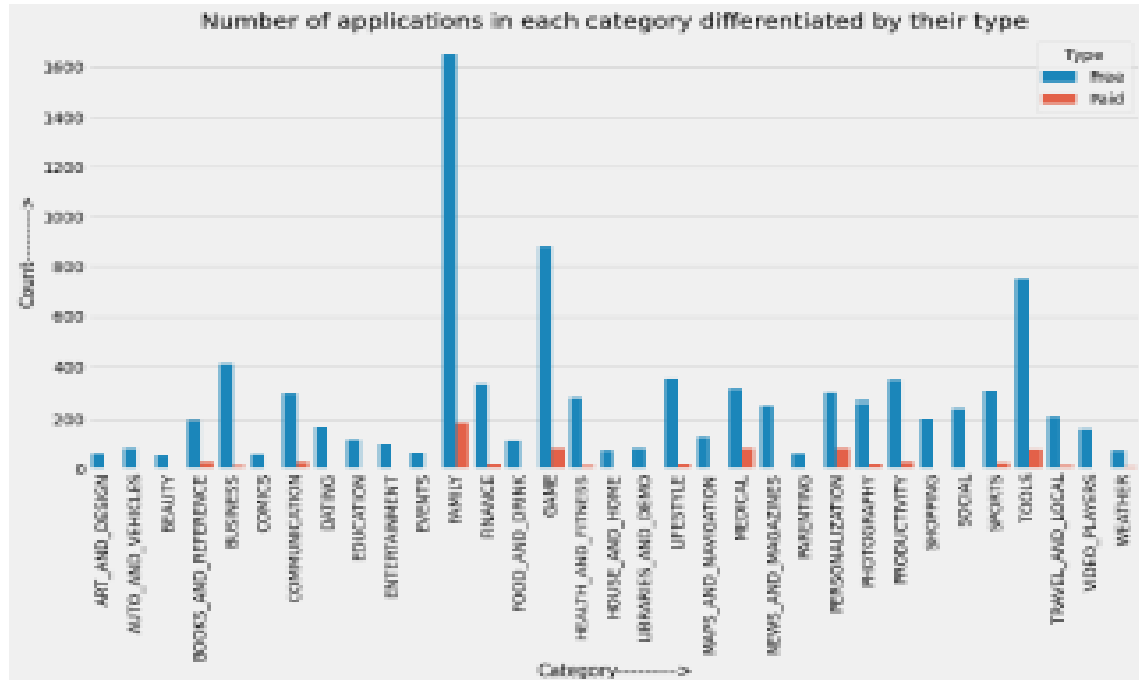
Exploration and Visualization-Paid vs Free Apps:



Free Type apps have highest number of apps 92.17% present in the Play store compared to Paid apps

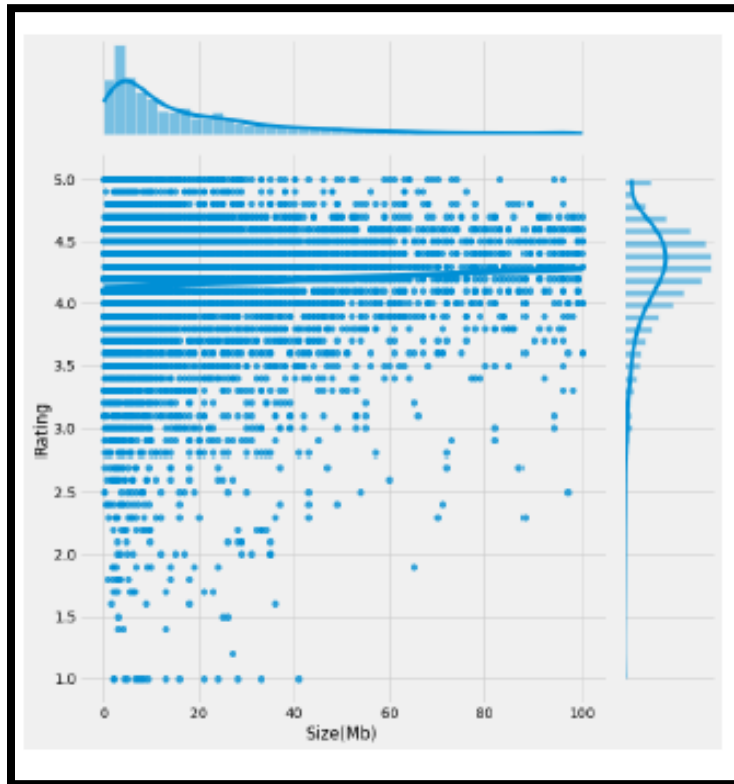
Exploration and Visualization

Category wise Free and paid apps count

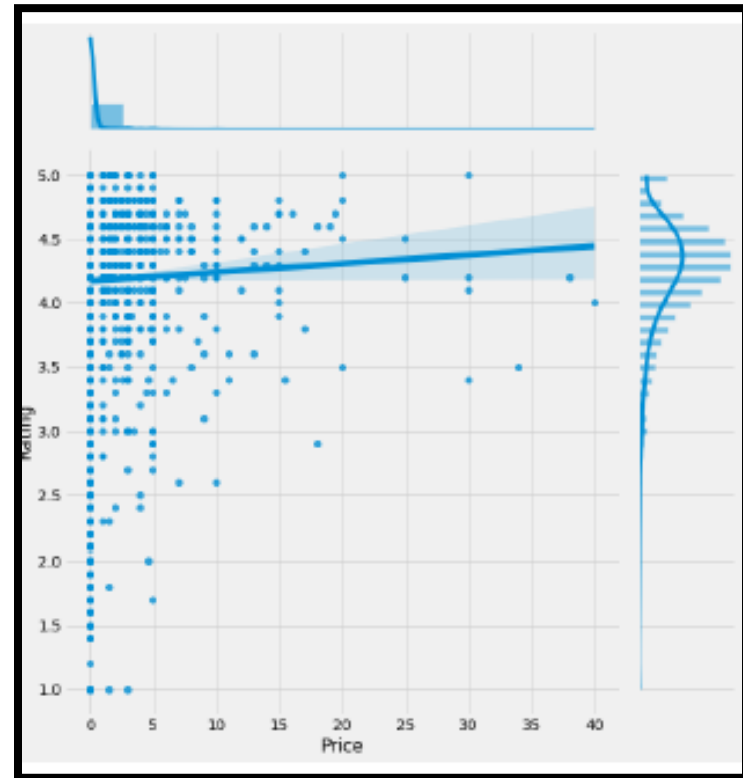


Exploration and Visualization

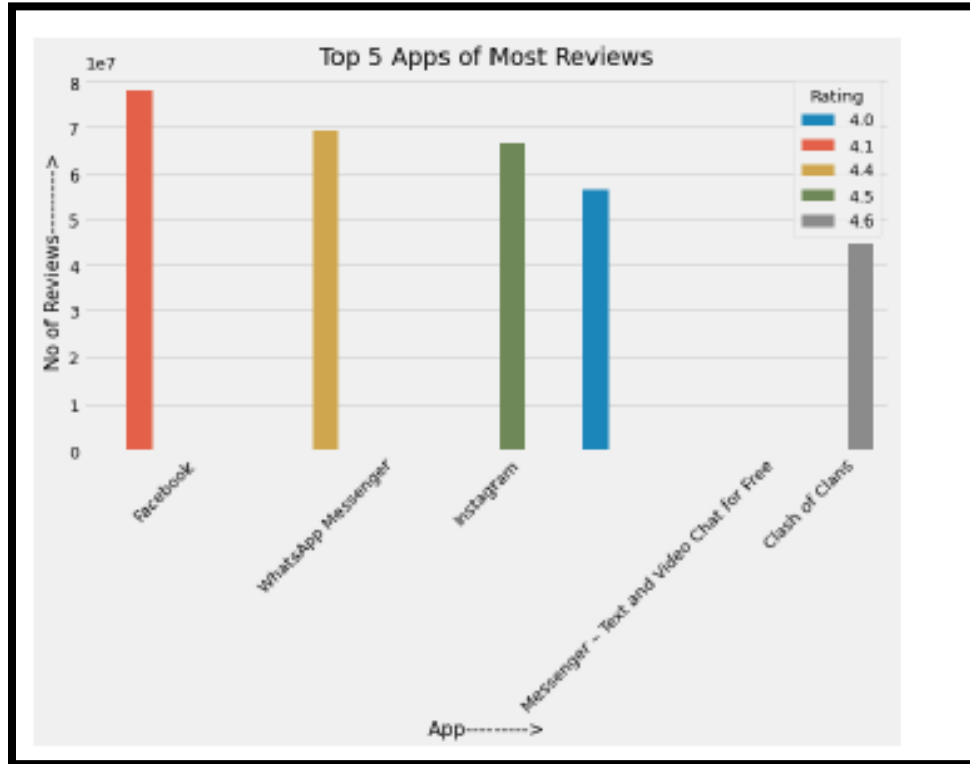
Effect of size on Rating



Effect of price on Rating



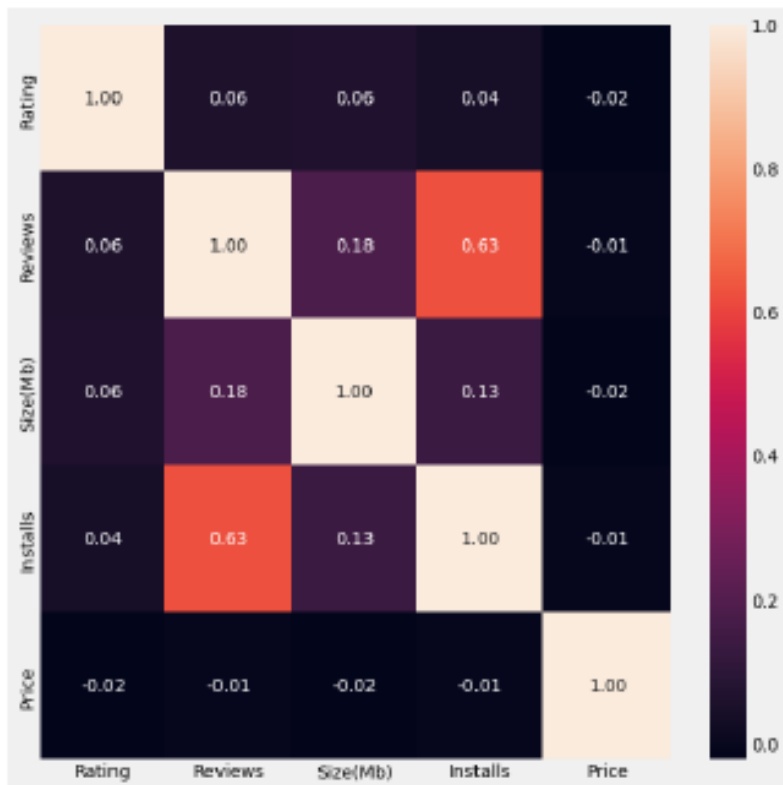
Exploration and Visualization-Top 5 most reviewed apps



- Facebook tops the most reviewed app
- Following WhatsApp and Instagram

Exploration and Visualization

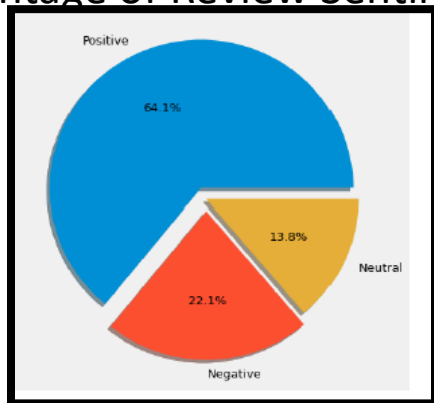
Correlation between columns



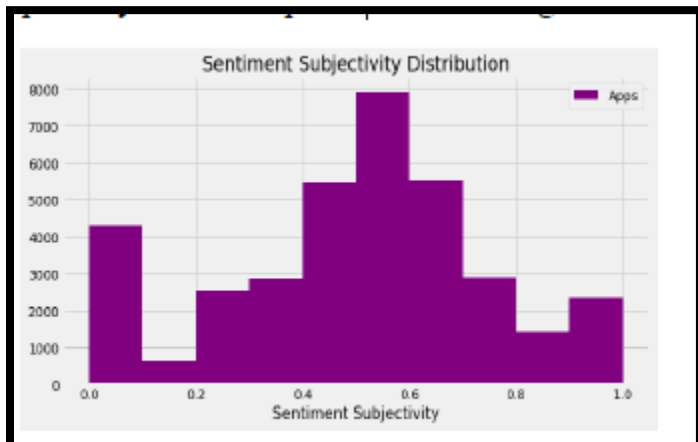
- from the above Heat Map shows high correlation in column between Installs and Reviews.

Exploration and Visualization

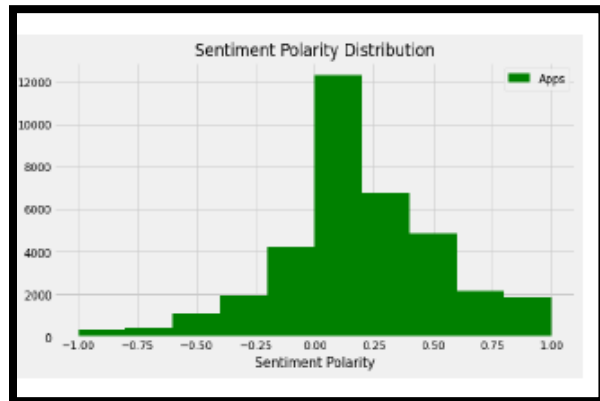
Percentage of Review Sentiment



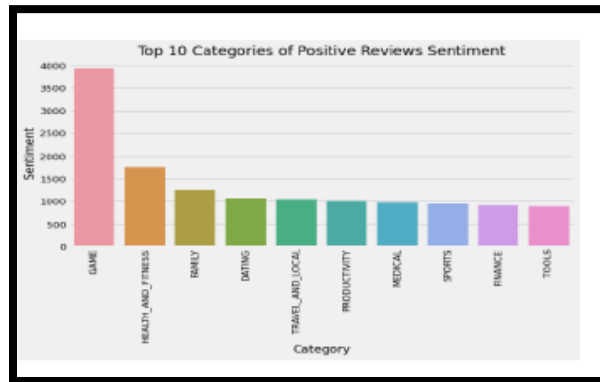
Sentiment Subjectivity



Sentiment Polarity



Top 10 positive review sentiment



Important Inferences and Conclusion



- The highest amount of application base on categories is family.
- In Communication Category highest is Video Players
- The highest amount of reviews in total is achieved by game category, but the highest amount of reviews/app is achieved by social.
- In Genre with installs of application Communication Category apps have highest
- In this distribution ratings majority of the apps had a rating of 4.3
- Free Type apps have highest number of apps present in the Play store compared to Paid.
- Reviews of the apps high installation also high, because users install only popular and most reviewed apps.
- In review sentiment analysis, most of the apps got positive reviews and 'Game' categories had highest positive reviews
- The maximum number of sentiment subjectivity of Apps between 0.4 to 0.7. From this, we can conclude that the maximum number of the Users give reviews to the applications, according to their own experience.

• **THANK YOU**