CAPSTONE PROJECT

PLAY STORE APP REVIEW ANALYSIS

TEAM MEMBERS

Tushar Jain

Rishabh Kumar

Shivansh Yadav

Madhur Awasthi

(Data Science Trainees @ Almabetter)

CONTENT

- About Play store
- Agenda
- Data Summary
- Data Cleaning
- Visualization and Analysis
- Merging of data
- Conclusion
- Tools used

About Play store

How Play Store works

Google Play, also branded as the **Google Play Store** and formerly **Android Market**, is a digital distribution service operated and developed by Google. Where you can download or buy millions of apps, games, and other media onto your Android device. Applications are available on Google Play either free of charge or at a cost.

Agenda

The objective of this project is to **analyze** the Play store applications data ,finding **key factors** responsible for **success** of an application and deliver insights to provide customer demands and thus help developers and investors to create and popularize products.
Our analysis will be fruitful for investors and developers.

Data Summary

Our Play store data set consists of 13 columns ,detailed description of all features is given below

- **App**: The app column describes name of the application.
- **Category**: The category column describes the category of application i.e. from which category particular app belongs to.
- **Rating**: This column describes the rating of particular app. This gives us the idea of how popular the particular app is.
- **Reviews**: This column describes the reviews about app. From this customer get idea about app i.e. app reviews is good or not.
- **Size**: This column gives the size of app i.e. how much kb or mb of a app is .

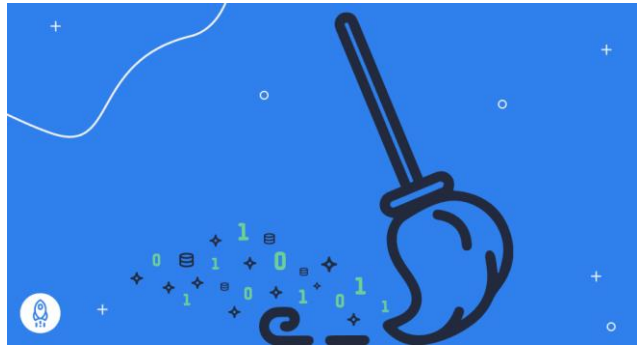
Data Summary

- **Installs**: This column gives the idea about installs i.e. how much times a particular app installs.
- **Type**: This column describes the type of data i.e. app is paid or free.
- **Price**: This column describes the price of paid app.
- **Content Rating**: This column describes content rating i.e. particular app is suitable for everyone, teen & kids etc.
- **Genres**: This column describes the category of app in much more details.
- **Last Updated**: This column will tell us when particular app was last updated.

Data Summary

- **Current Version**: This column will tell us the current version of particular app.
- **Android Version**: This column will tell us the android version of particular app.
- **Translated Review**: Reviews given by the users
- **Sentiment Polarity**: Sentiments of users on scale of [-1 to 1]

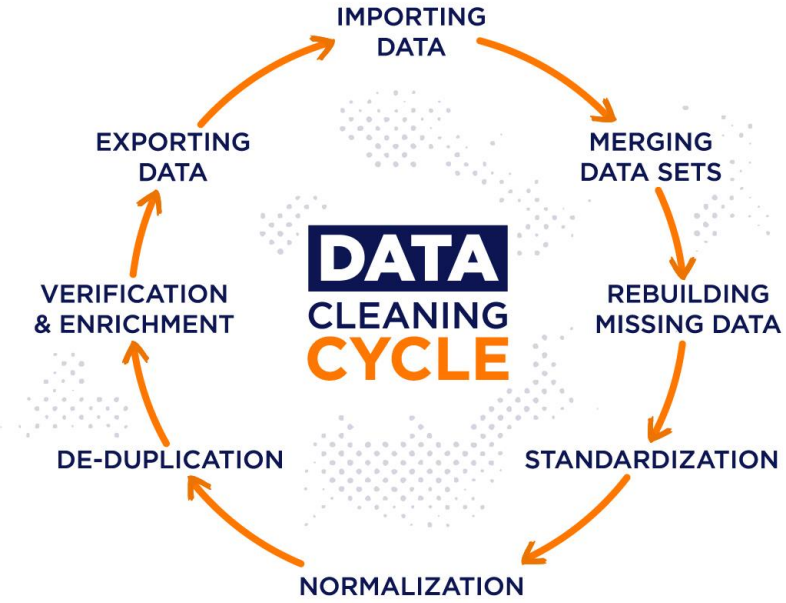
Data Cleaning



Cleaning of the dataset is first and foremost part of our project, getting rid of irrelevant
Data is very essentia'

What we have done:

- Removed the duplicates
- Removed unnecessary columns
- Removed null values
- Removed unnecessary symbols
- Converted values to analyzable Data type



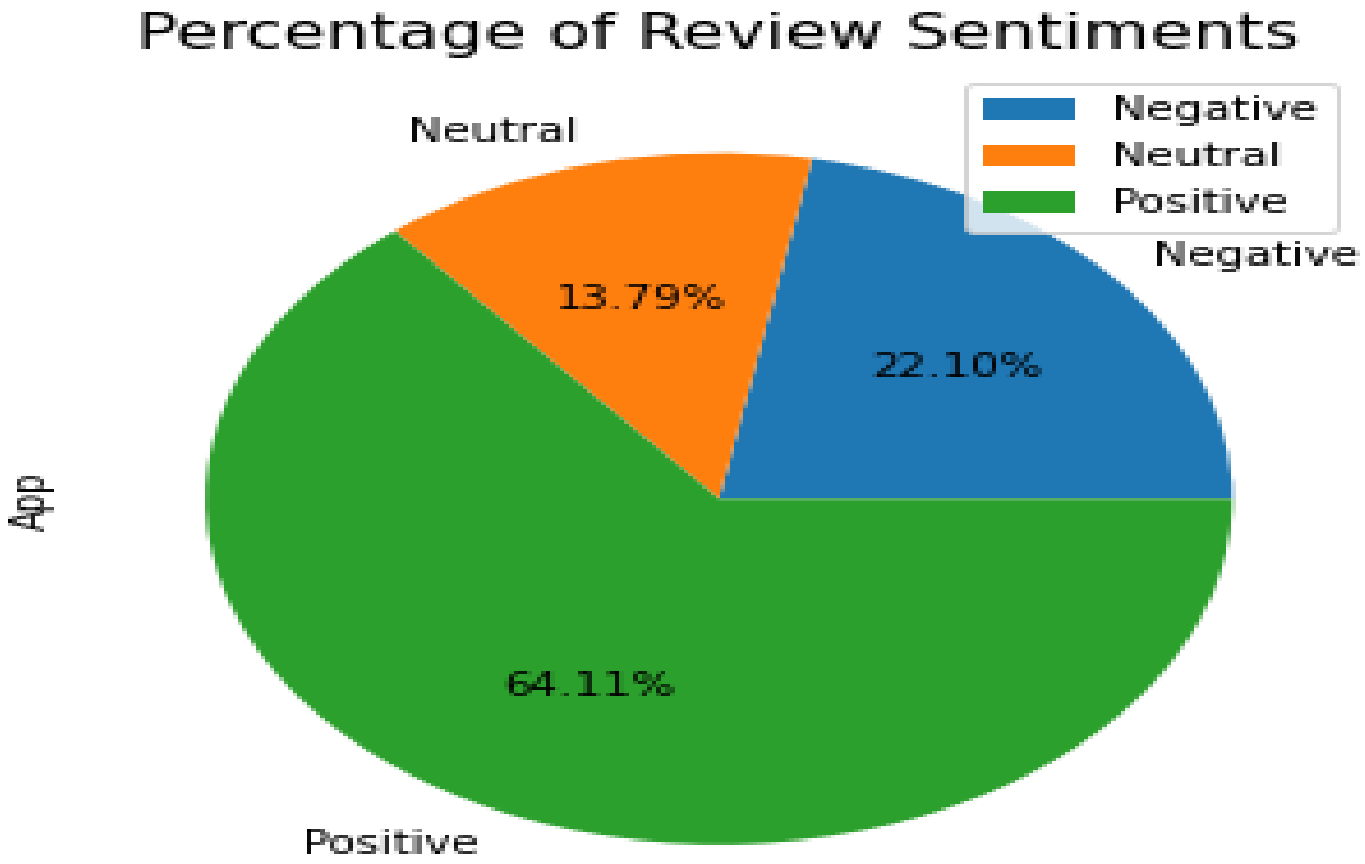
Visualization and Analysis

Visualization and analysis is the core part of our project, converting Tabular data into graphical form makes analysis easier and helps to find insights.

What we have done :

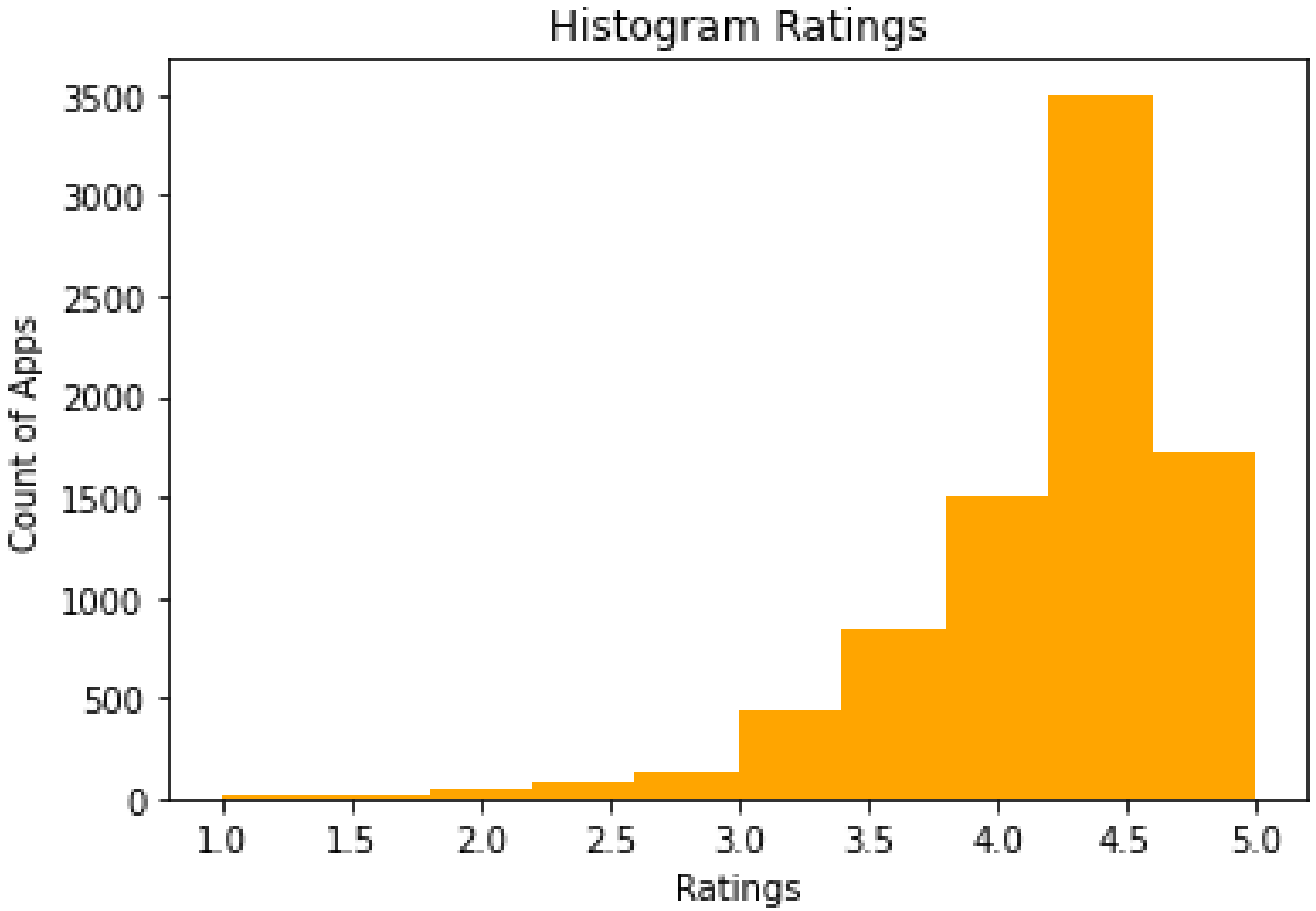
- Created different type of charts to visualize data :
  - Bar Chart
  - Pie Chart
  - Box Plot
  - Line Chart
  - Histogram

Sentiment Division



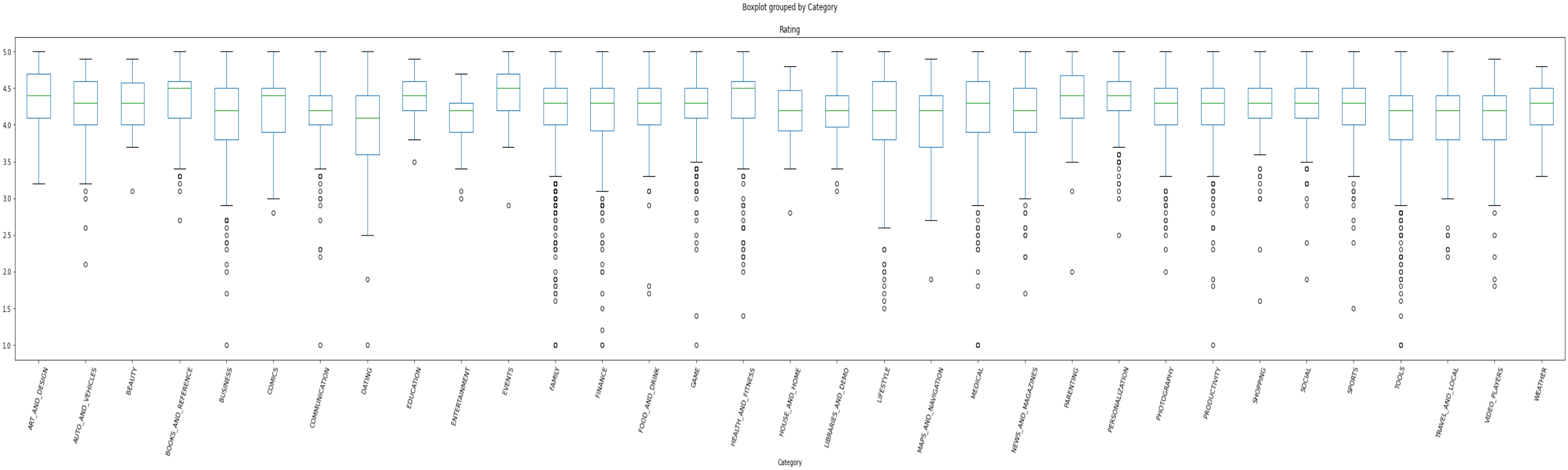
It represents the **Sentiment Division** in percentage among applications

Installation rating wise



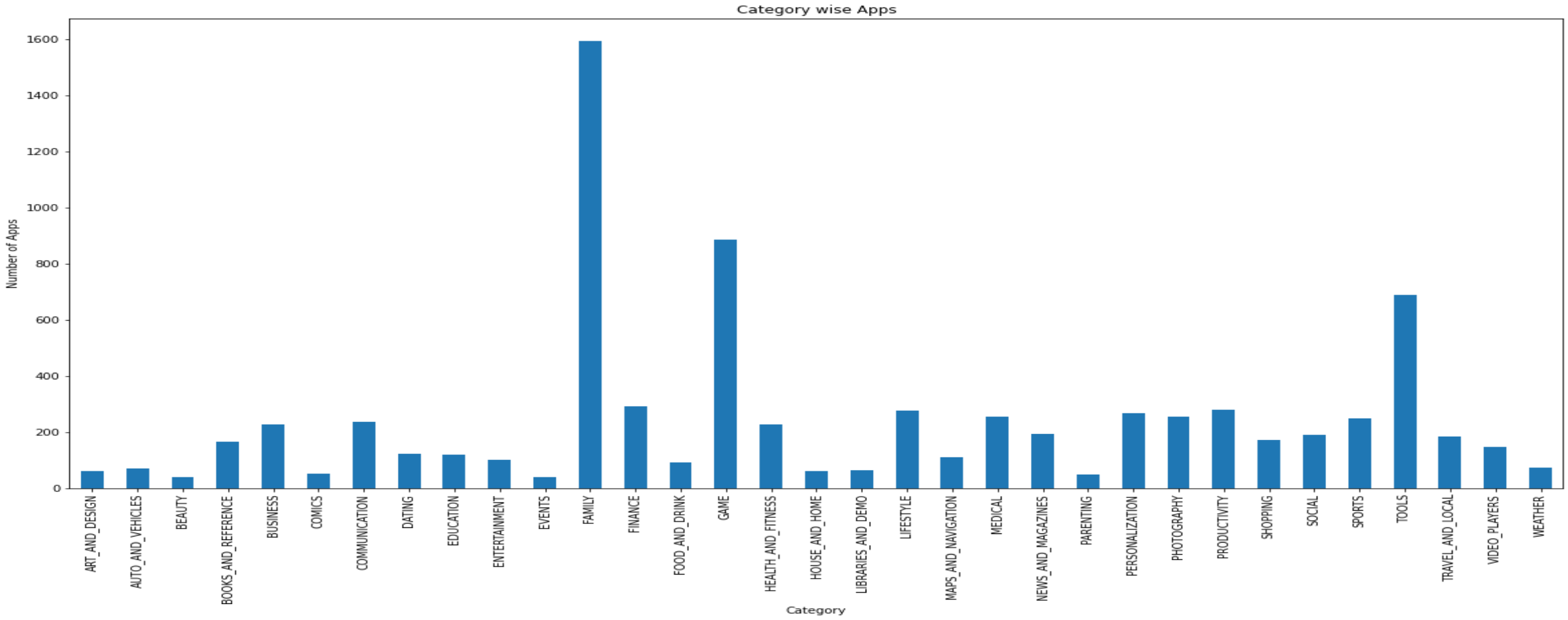
It represents **no of applications** having particular **Rating**, here we can see **maximum** number of applications have **4.5 rating**

Rating based on Category



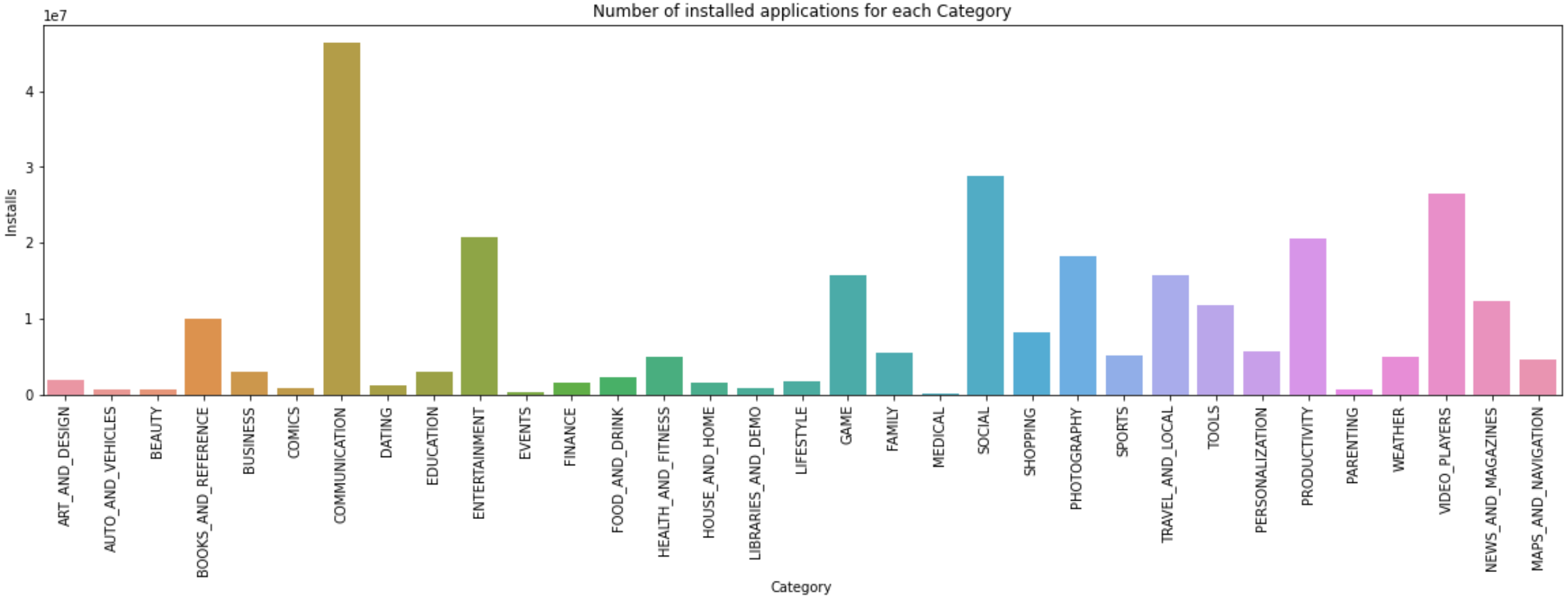
This is a **Boxplot** representing percentile division of **rating** with respect to different **categorie**s and shows **outliers**

Number of apps category wise



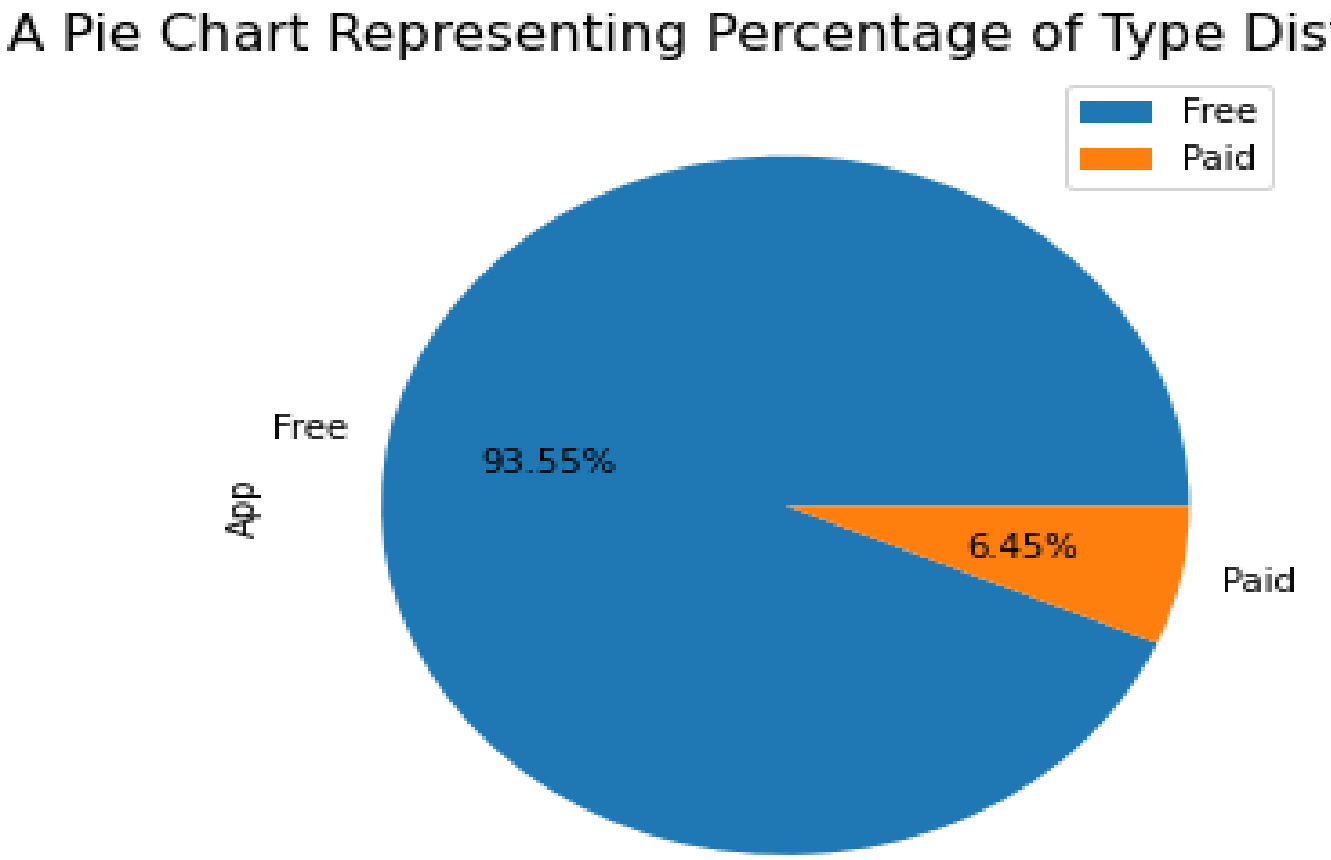
This **Bar-Graph** represents total no of apps category wise

Installs based on Category



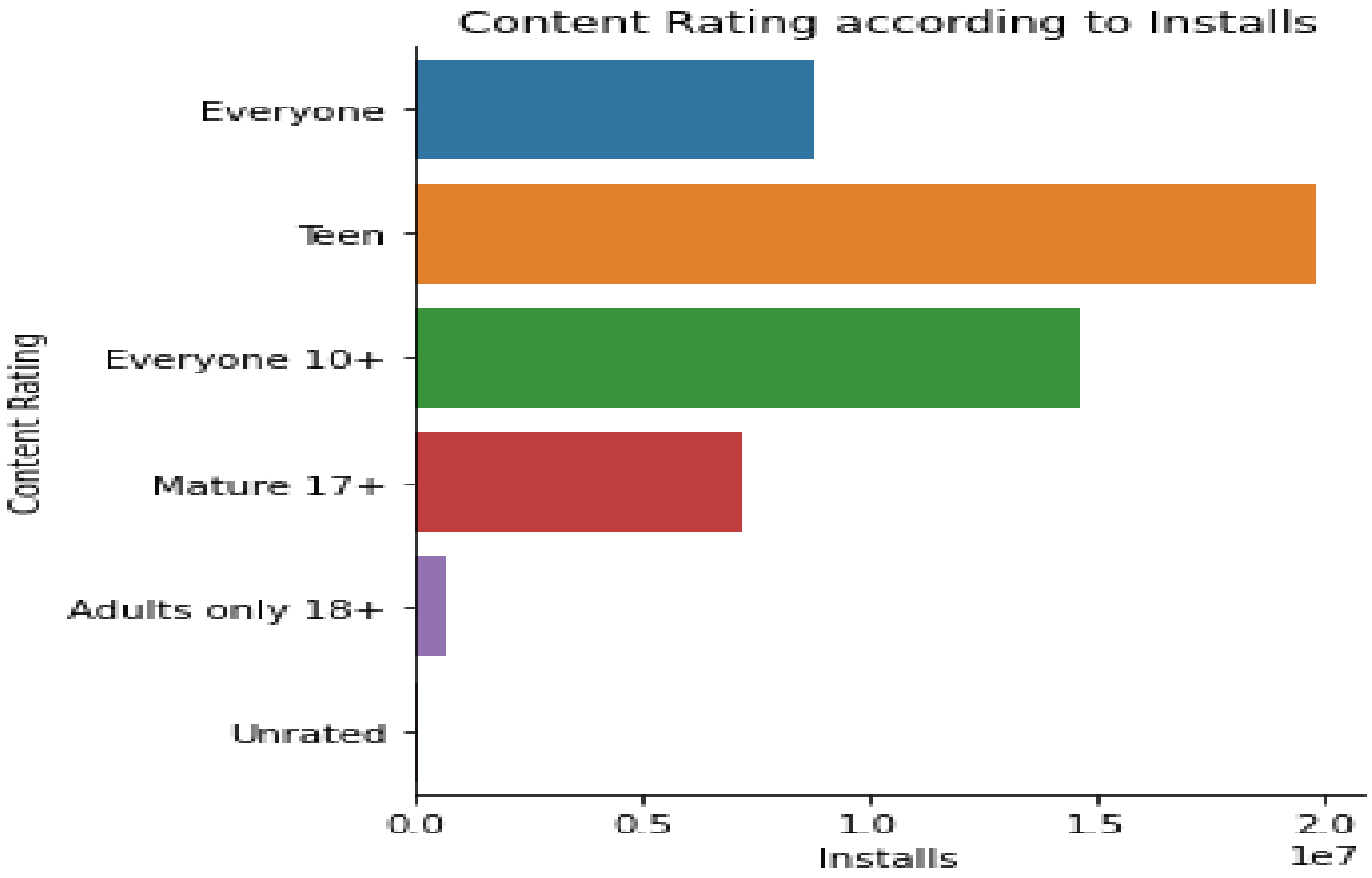
This **Bar-Graph** represents no of **installs** in millions category wise in **millions**

Application division based on pricing



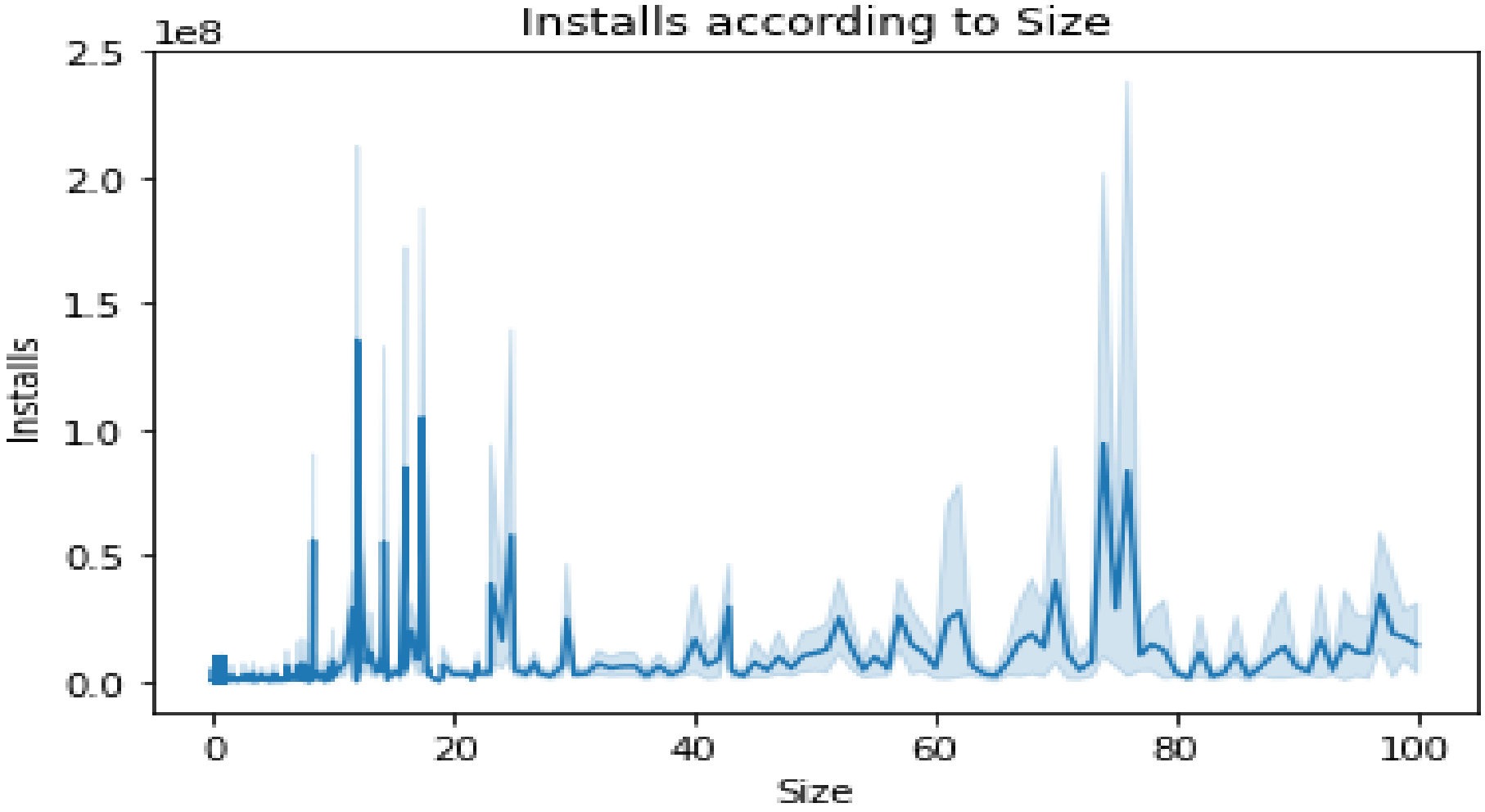
This **Pie chart** represent percentage of apps **free/paid**

Installs based on Content Rating



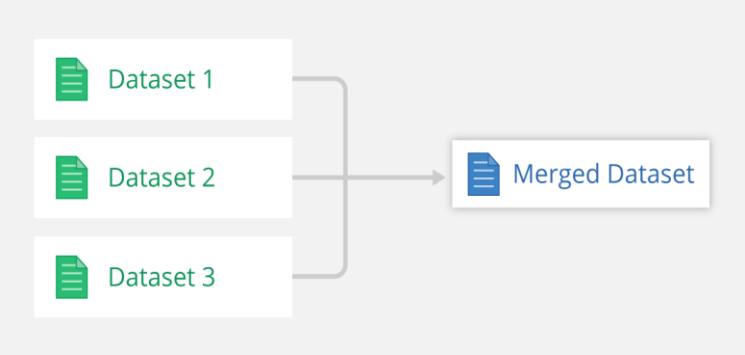
This Bar Graph represents **Installs** in millions **content rating** wise

Installs based on size



This Line Graph represents **Installs** with respect to **size**

Merge Data



- Merging of data helps us relating more than one data frame.

What we have done:

- Merged two data frames and analyzed sentiment polarity

Conclusion

- The applications **success** majorly perceived by its number of **'Installs'** and its **'Rating'** and there are few key factors responsible for both. So we found that apps with maximum no of 'Installs' belongs to Category**(Communication, Social, Video Players, Productivity, Entertainment)** Content Rating **(Teen, Everyone 10+, Everyone)**

- So if we want to build an app we should keep in mind the category as well as content rating. In our analysis of top 10 app we found the most successful app would be of Category**(Communication, Social)** and content rating **(Teen, Everyone)**

If we want to create an App, most **Favourable categories** are:

- Communication
- Social
- Video Players
- Productivity
- Entertainment

Most favourable content:

Teen
Everyone 10+
Everyone

- All most all of **top 10** apps are in **size** range **10-25 MB**
- Almost all top apps are **Free**
- As our **Top** application is **Subway Surfers** which comes into Category **Games** . So we can also consider a **Game** in Genre **Arcade**

we can deduce an application of

- Category **(Communication)**
- Content**(Teen)**
- Size **(10-25) MB**
- **Free**

would most probably **Succeed**

Tools used

Programming Language: Python

Libraries: Pandas (Data manipulation)

Numpy (Data manipulation)

Matplotlib (Data Visualization and analysis)

Seaborn (Data Visualization and analysis)

Thank you