

EDA

CAPSTONE PROJECT

Play Store App Review Analysis



By

Ram Manohar Thakur

Sayan Biswas

Shradha Lakhadive

Snehal Ragit

Data Science Trainee, Almabetter

CONTENTS :

- Introduction
- Steps Involved
- Objectives
- Agenda
- Data Understanding
- Data preparation
- Data Visualizations
- Conclusion



INTRODUCTION :



- Google Play, also branded as the Google Play Store and formerly Android Market, is the official distribution storefront for Android applications and other digital media, such a music, movies and books, from google.
 - Google has released a timeline of apps and games that transformed the lives of individuals over the last 10 years.
 - Google Play was launched in 2012, uniting Android Market, Google Music, Google eBookstore and Google Movies under a label. As per the Android Authority report
 - Google serves as a trusted platform for legit app developers and gets them the opportunity to earn. Google takes 30% of apps and in-app purchases, while 70% of the payment goes to the developers. Google recently halved its cut to 15% in 2021.
 - Lots of designers and developers work on it to make an app successful on the play store.
- The main moto of our project is to analyze the data of google play store to extract the features of play store reviews and analyses the opinion of play store app reviews.

STEPS INVOLVED :



- Importing the data sets : We are having two datasets, First one is Play store app dataset and second one is User Reviews dataset.
- Importing Libraries : NumPy, Pandas, Seaborn, Matplotlib, Datetir
- Data Understanding : This steps helps us to explore data in few aspects like rows, column, shape and data type etc.
- Data Preparation : Finding and removing Outliers, Null values, Removing duplicate data and transform the data to make efficient for analyzing and visualization.
- Data visualization : Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.



OBJECTIVES :

- Understand Consumer behavior and demand how they reacts to different Category Genres of Google Play store Apps.
- Find how small changes or update impacts on app performances.
- Find the most popular and trending apps in recent times.
- Analyze the Reviews, Ratings, Sentiments of people towards various apps in Play stores.
- Help developers or clients to recognize the gap to make the app better.



AGENDA for Analysis:

- Correlation Heatmap
- Categorical Analysis
- Top Free and paid apps
- App Rating Analysis
- App Reviews Analysis
- Most popular Apps
- Type and content Rating Analysis
- Average Price of Paid Apps In Each Category
- App Size Analysis

DATA UNDERSTANDING :

10841 Apps
13 Columns



- 1) App : This column Contains the name of Apps.
- 2) Category : The Category shows which type of app it is.
- 3) Rating : This shows the average rating for the app. The range of rating is in between 0 to 5.
- 4) Reviews : Number of Reviews shows by the play store that how many users voted on that app.
- 5) Size : User need required memory storage in android to install the app.
- 6) Installs : It shows that how many times that app downloaded and install from the play store by the user.
- 7) Type : It shows the information about that particular app whether the app is paid or free.
- 8) Price : It shows the data about the price if the app is paid.
- 9) Content Rating : This title shows content that may be suitable for all ages.
- 10) Genres : The columns of Genres indicates which sub-category the app belongs to.
- 11) Latest update : It shows the latest update of the app was released.
- 12) Current Version : It shows the current version of the app available on the play store.
- 13) Android Version : It gives the detailed about the android version, whether the app is supported or not.

Data type:

- String
- integer
- float
- object

Data Preparation:

- **Missing Value:**

We have 1476 null values

Rating has 99% of total missing values.

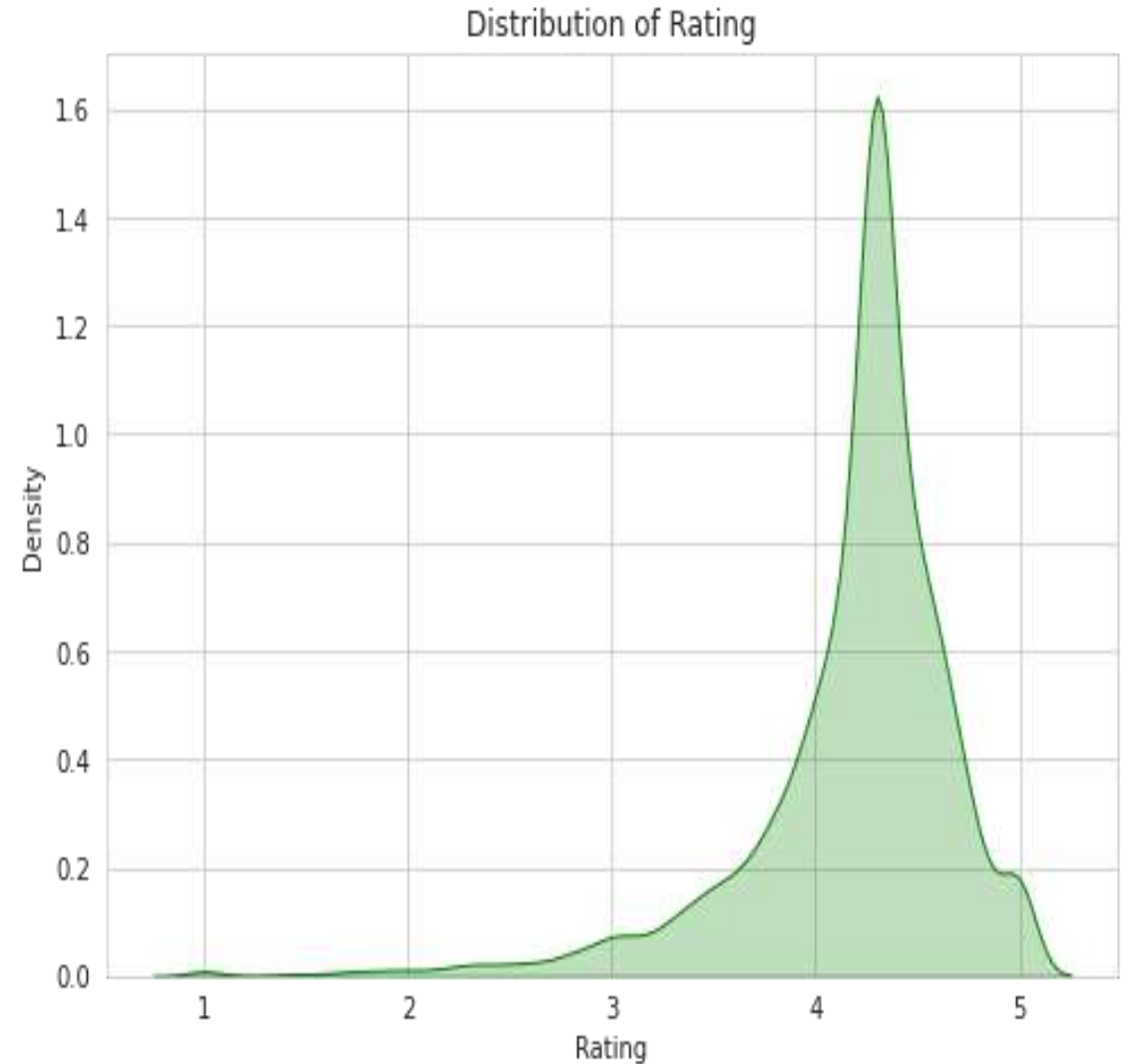
- **Treating with missing value:**

Analyzed features.

Replaced rating missing values with median value.

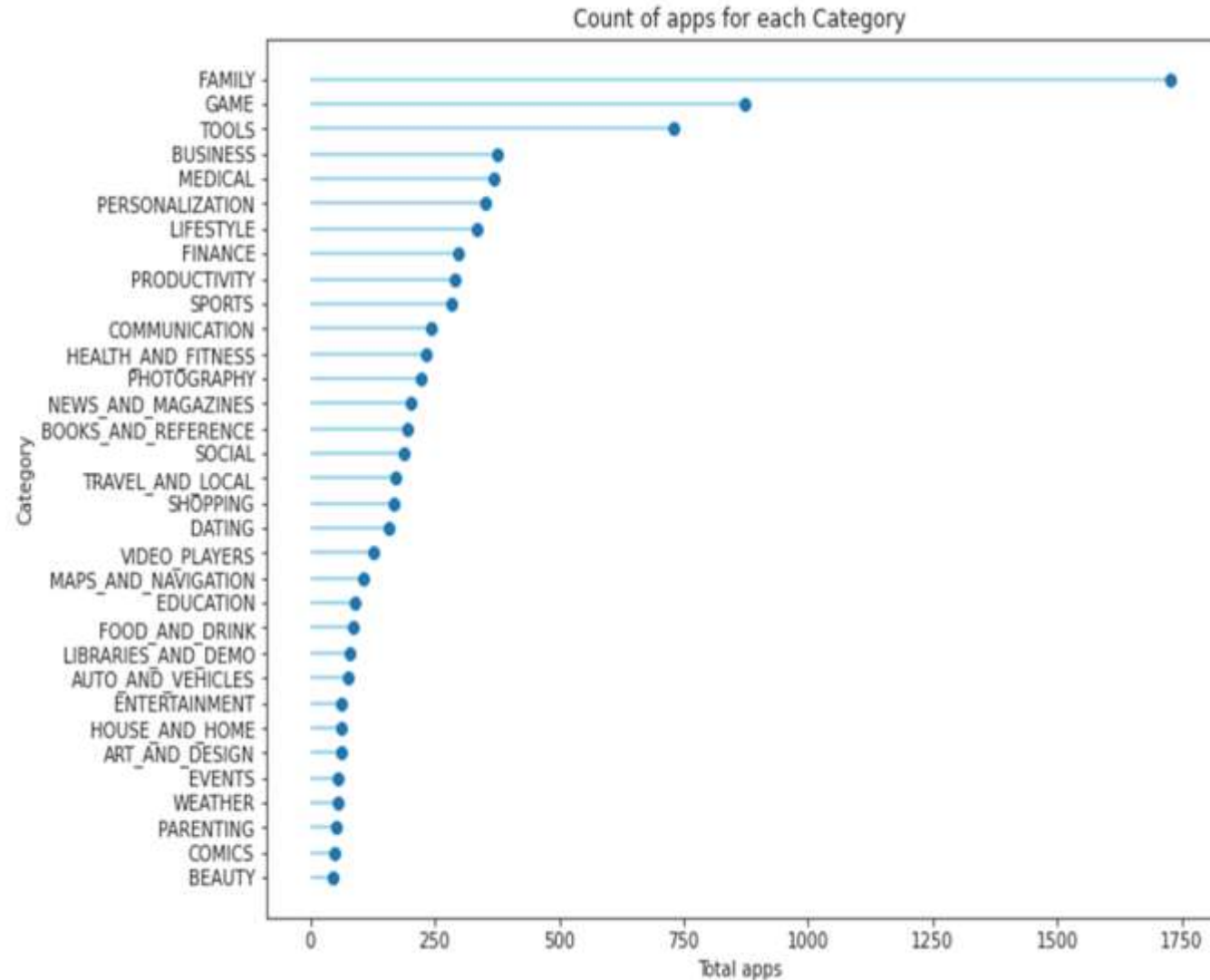
Distribution of Rating :

- From the above plot we can see that the rating from 3 to 5 is higher
- Median of the graph is 4.3



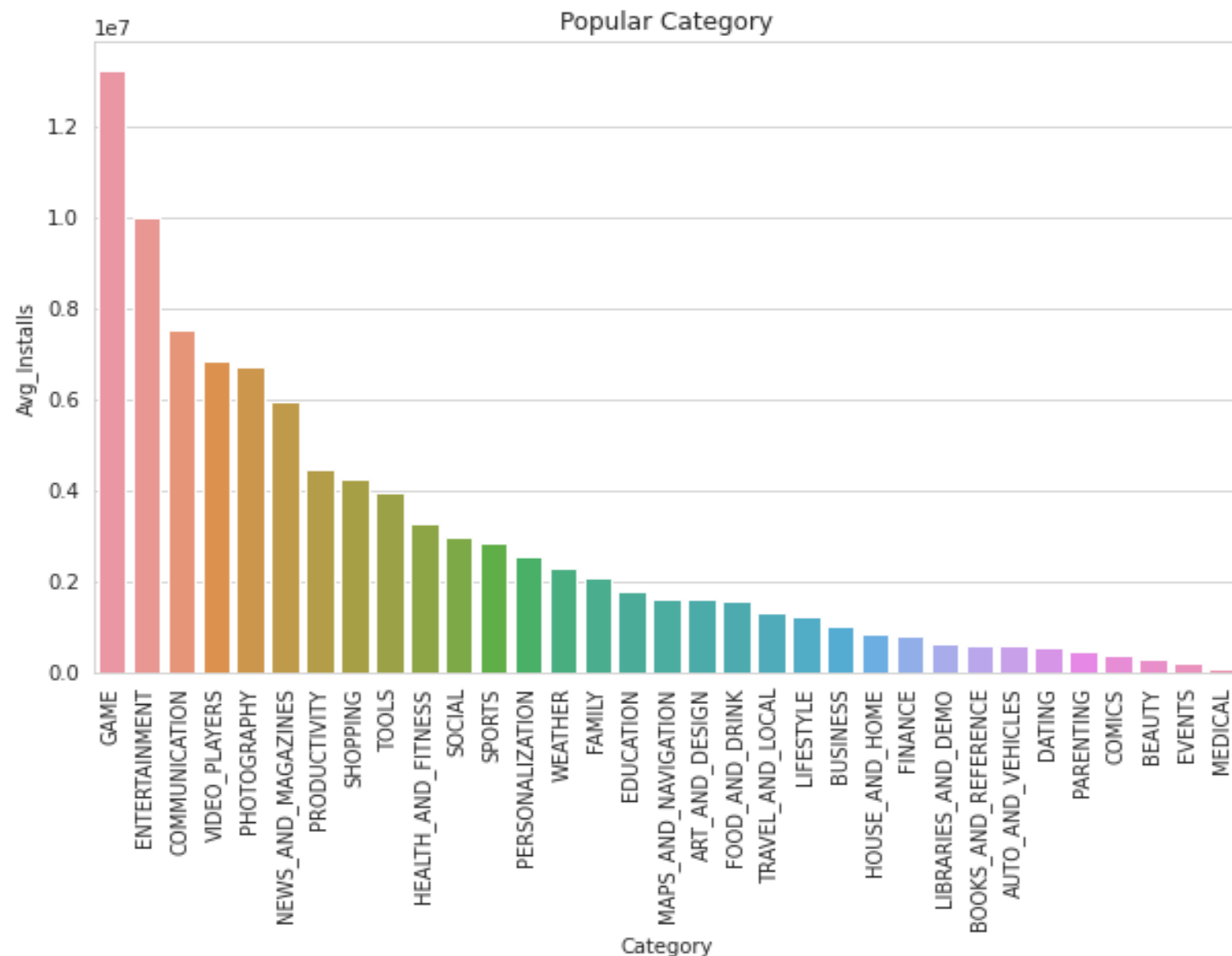
Count of Apps for Each Category :

- Total apps=33 categories.
- Family and Games apps have the highest count of applications than the others categories app.
- Family category has 1726 apps
- Games and tools have 873 and 731 apps respectively.



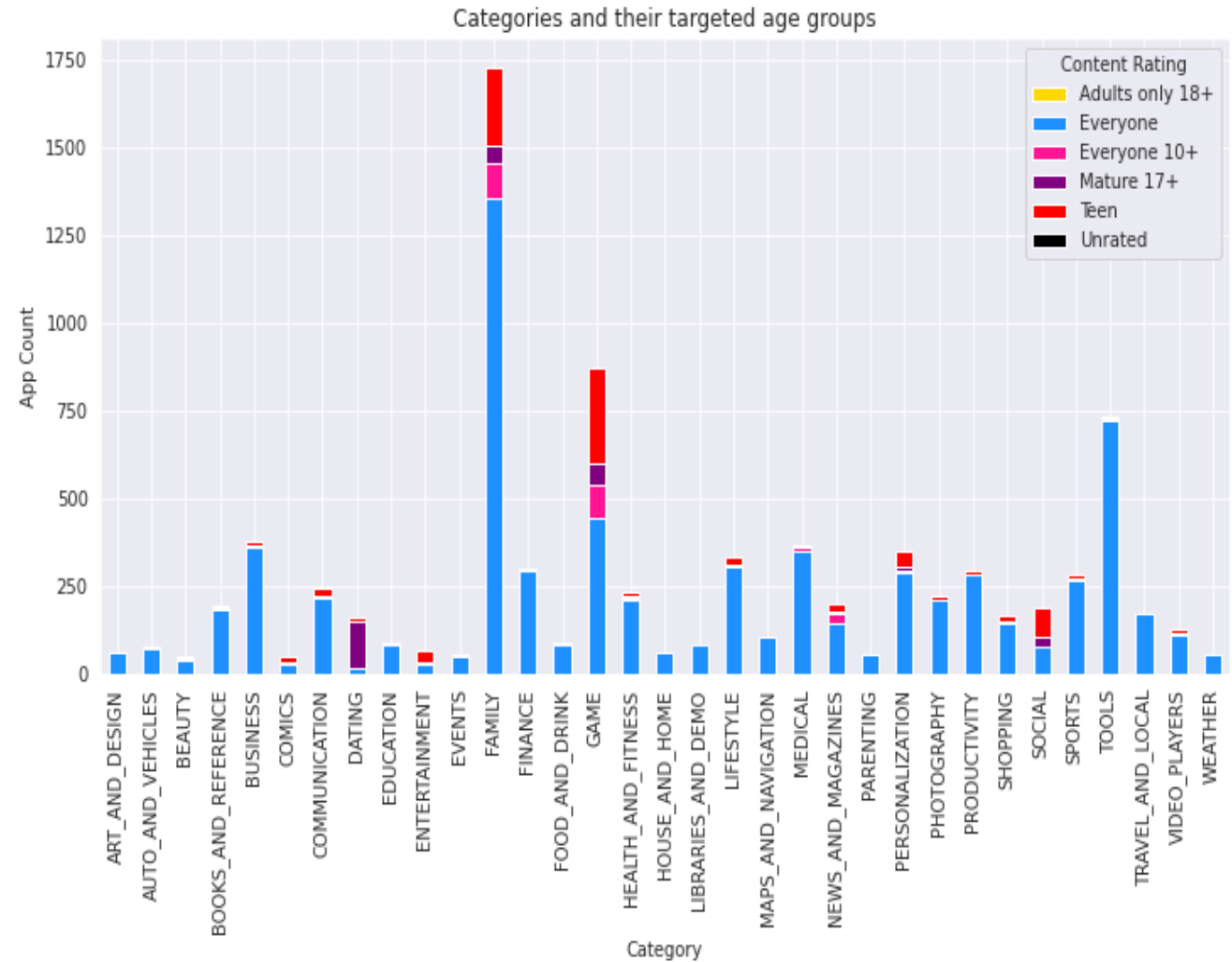
Popular Category :

- Most popular app-Game
- We can see that in this graph the Entertainment, and Communication have highest installs than other categories.
- Medical category has least number of installs.



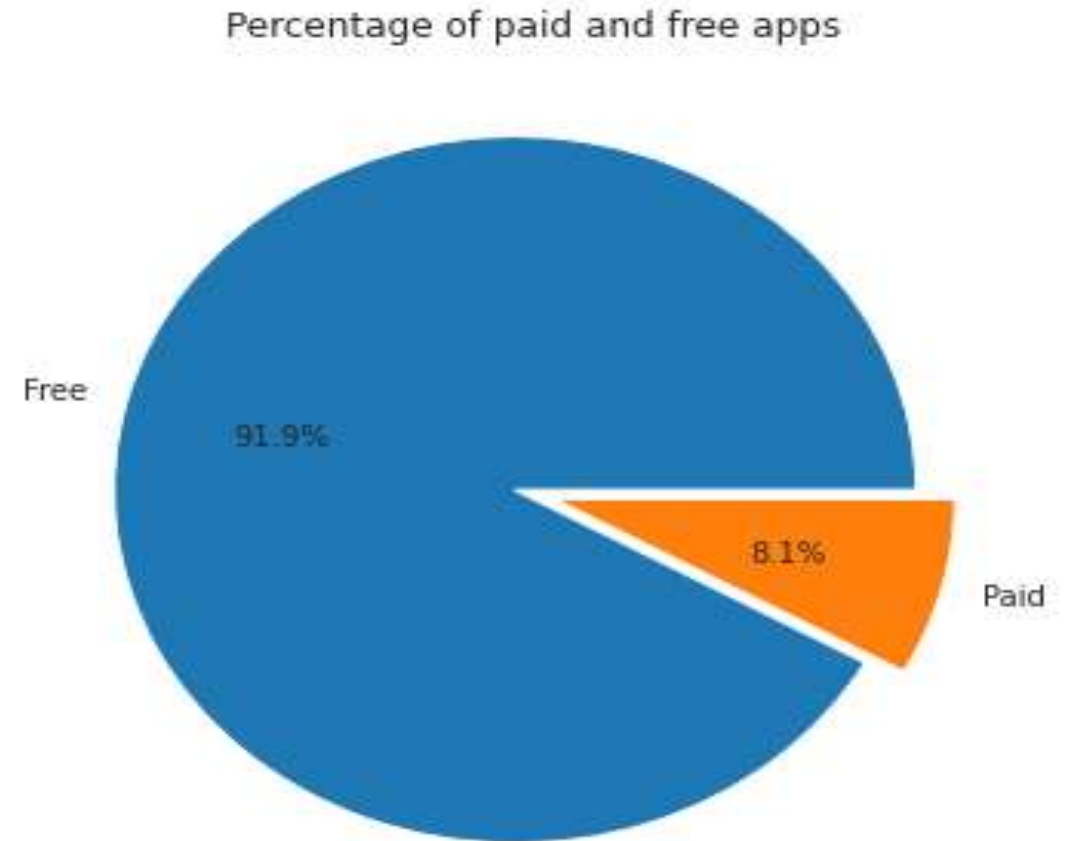
Categories And Their Targeted Age Groups :

- This graph shows targeted age groups of different categories.
- Every category has all age group people.
- Dating category is only for mature people.



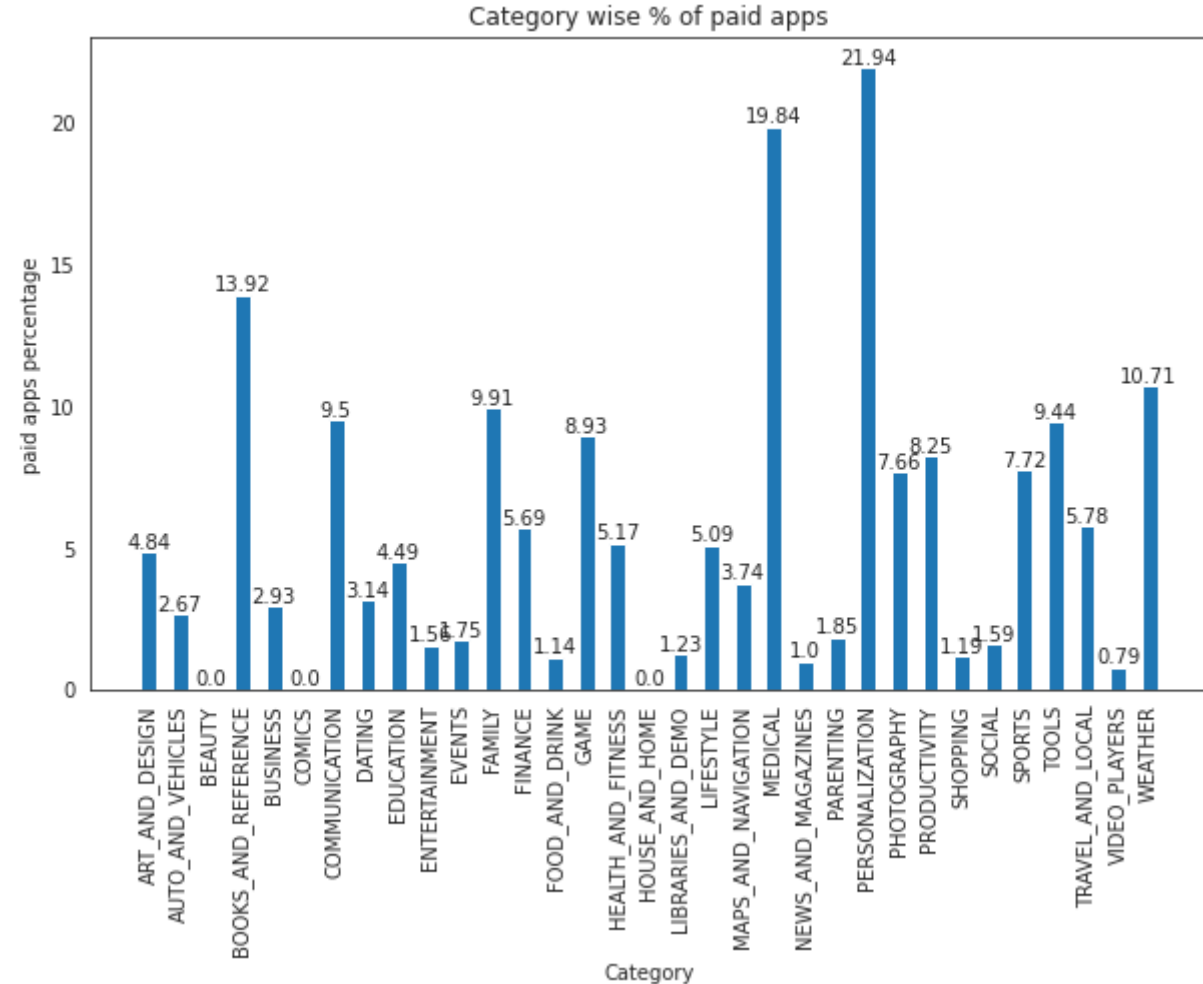
Percentage of Free and Paid Apps :

- The percentage of free apps is 91.9% and paid apps is 8.1%.
- We can say that the percentage of free app is more than the paid apps.



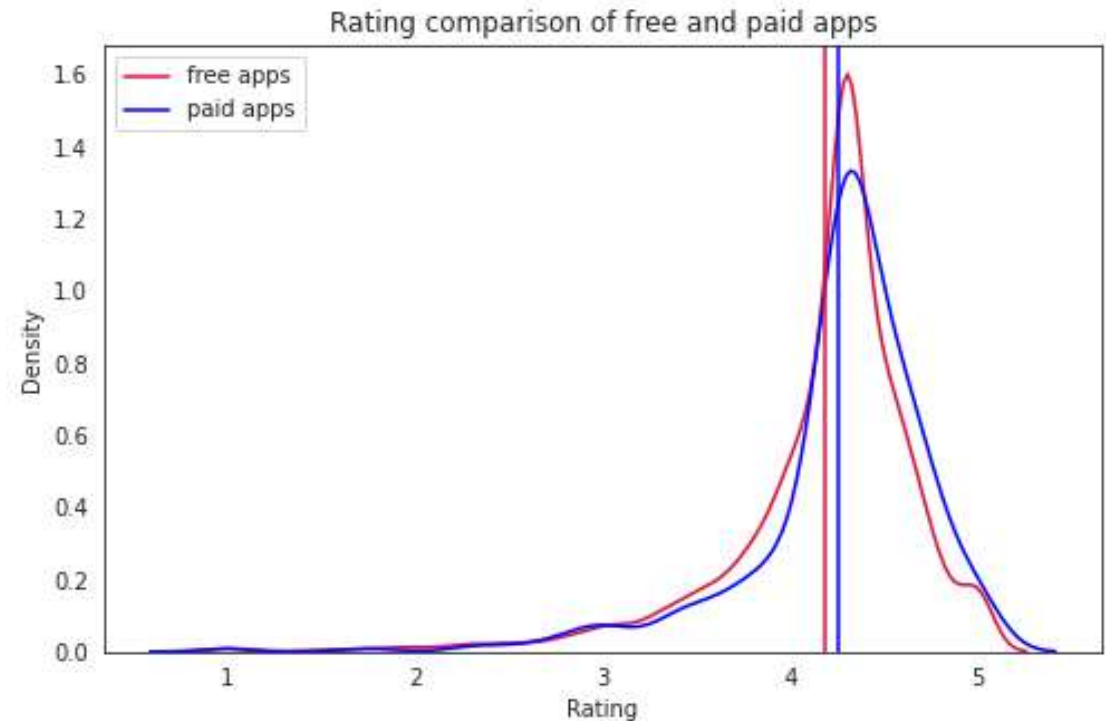
Percentage of Paid Apps :

- The apps which we have paid in that percent of personalization is 21.94 which is higher than the other applications.
- Beauty,comics,house and home having least percentage.



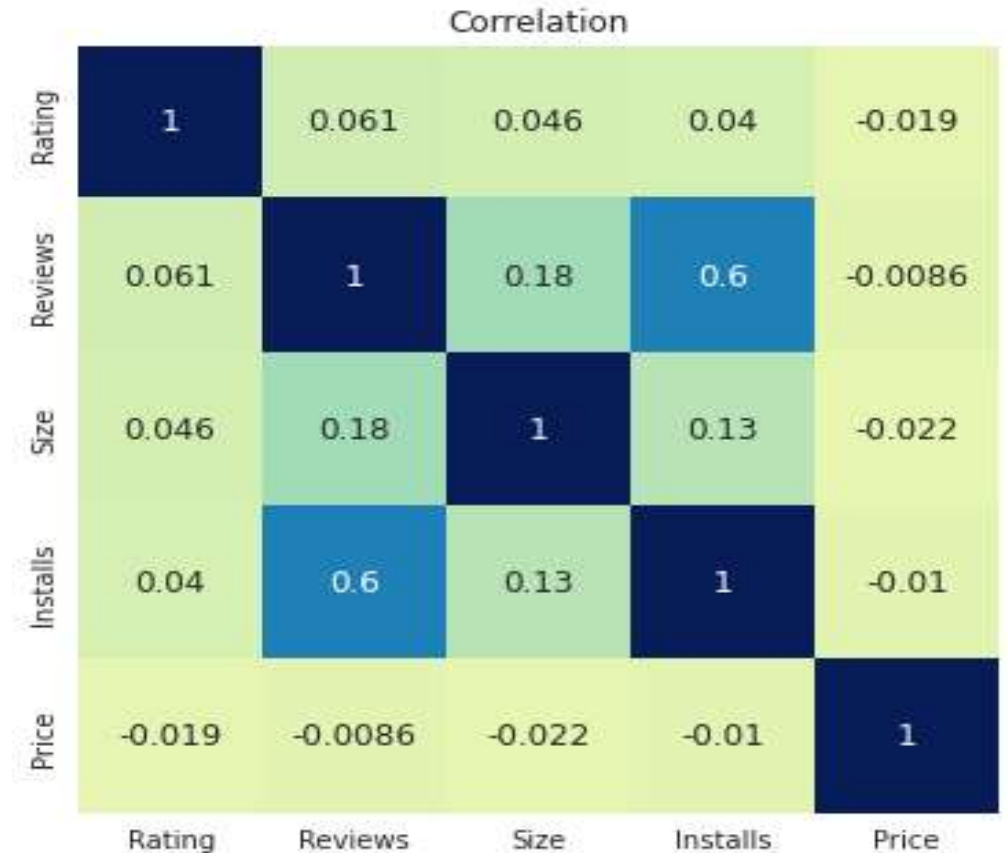
Rating Comparison Between Free App And Paid Apps :

- Rating of free app is maximum between 3 to 5.
- Rating of paid app is also maximum between 3 to 5 and their density is slighter less than free app density.



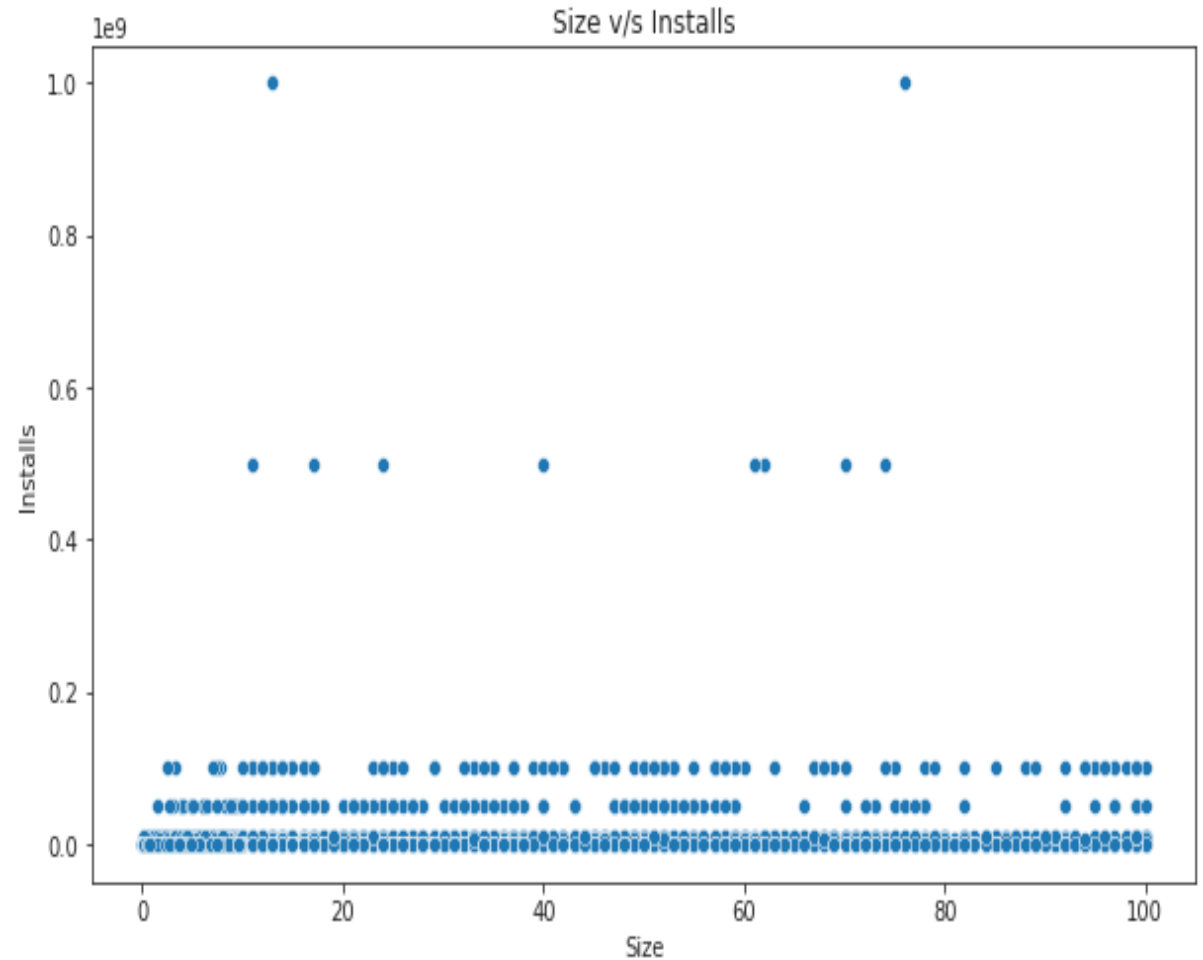
Correlation Heatmap :

- Installs and reviews have the strongest correlation
- Rating has negative relation with price.
- Installs has positive relation with size but it is very weak



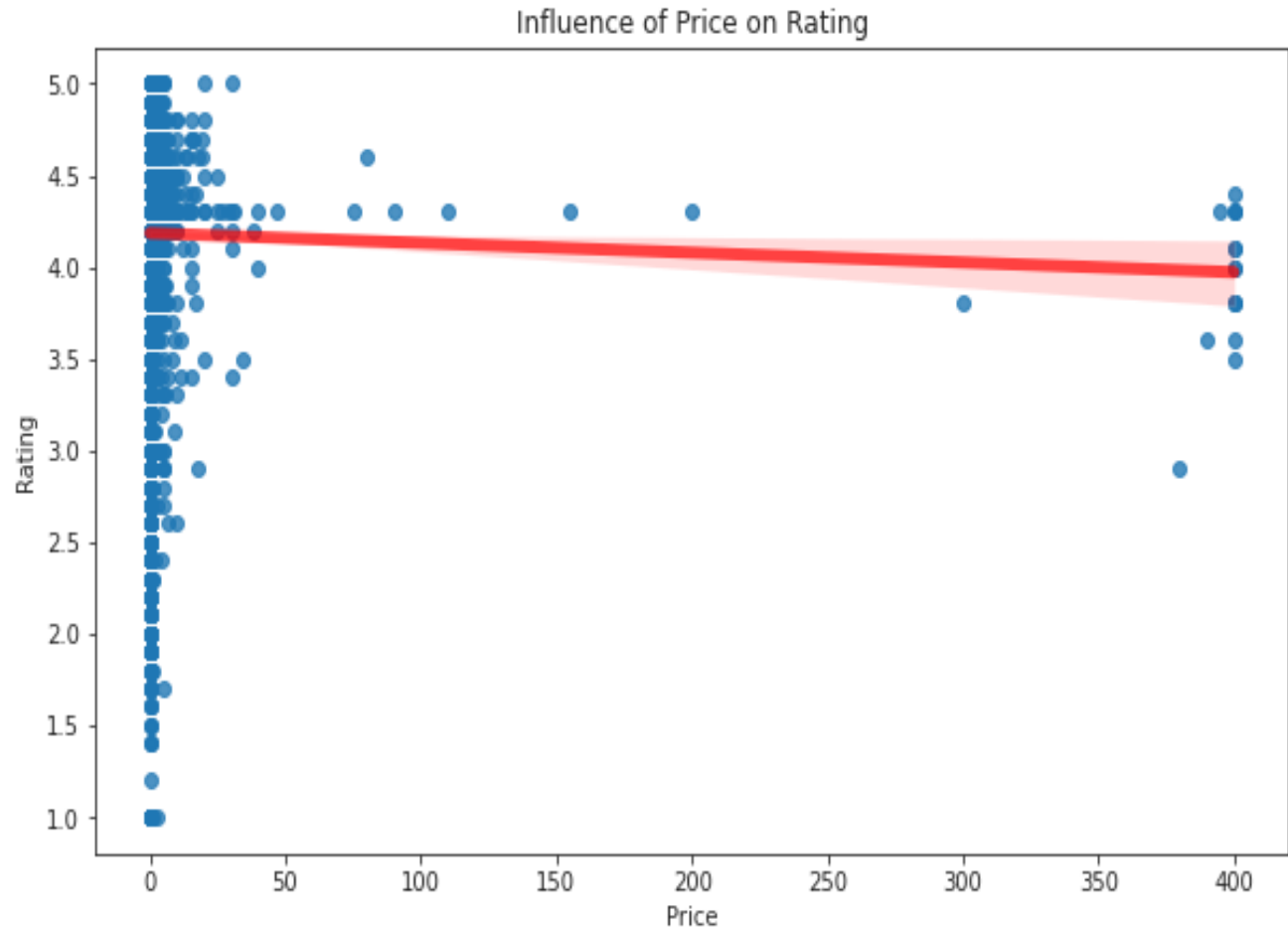
Size vs Installs :

- From the above scatterplot we can observe that user install less app that which app size having high mb



Influence of Price on Rating :

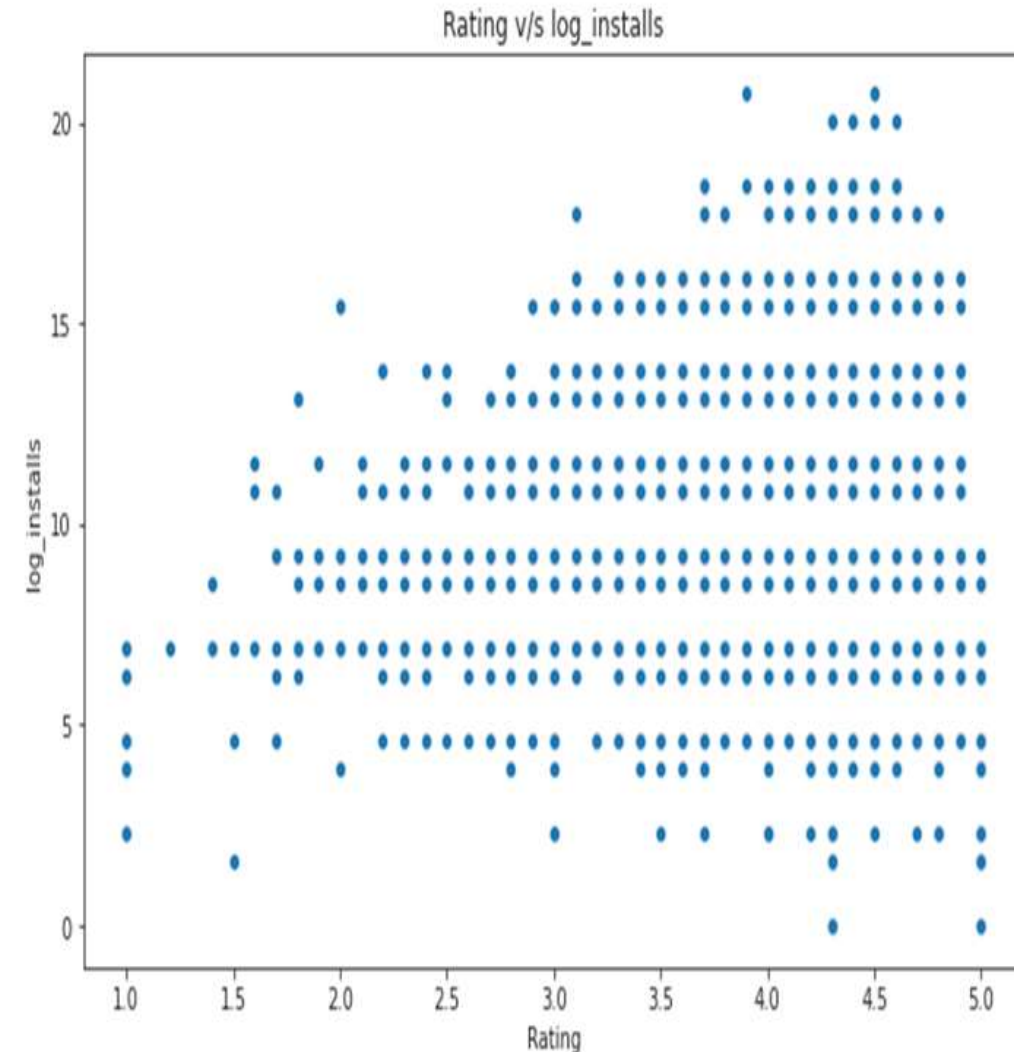
- Maximum app cost Below \$100.
- Price of apps in the range between 0 to 50.
- Rating decreases with increasing price.

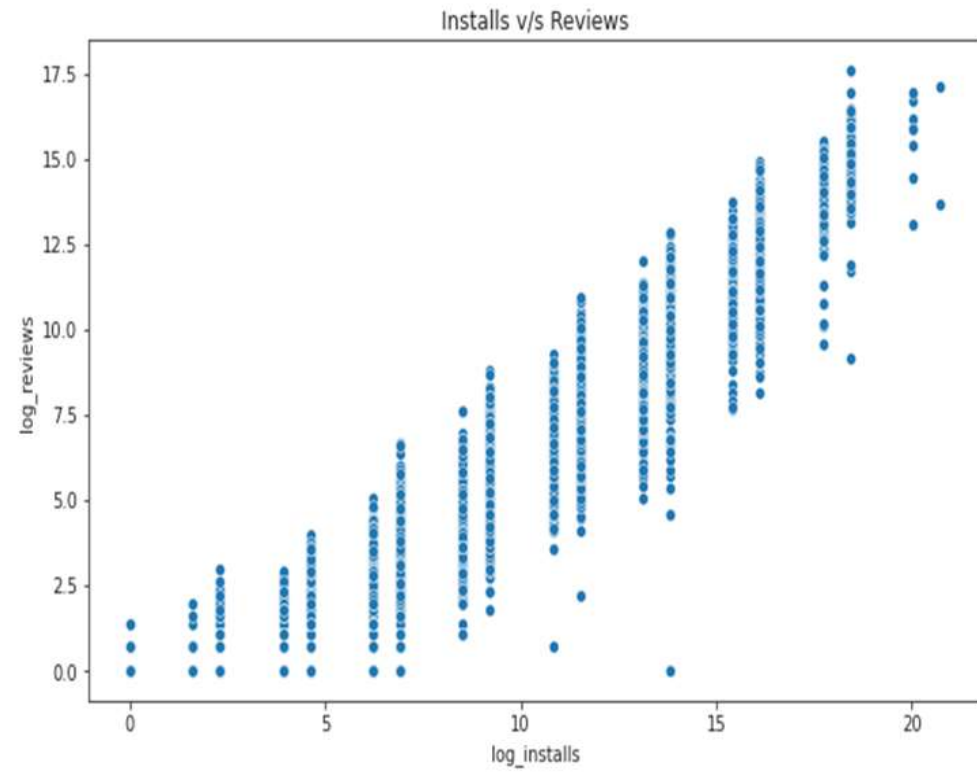
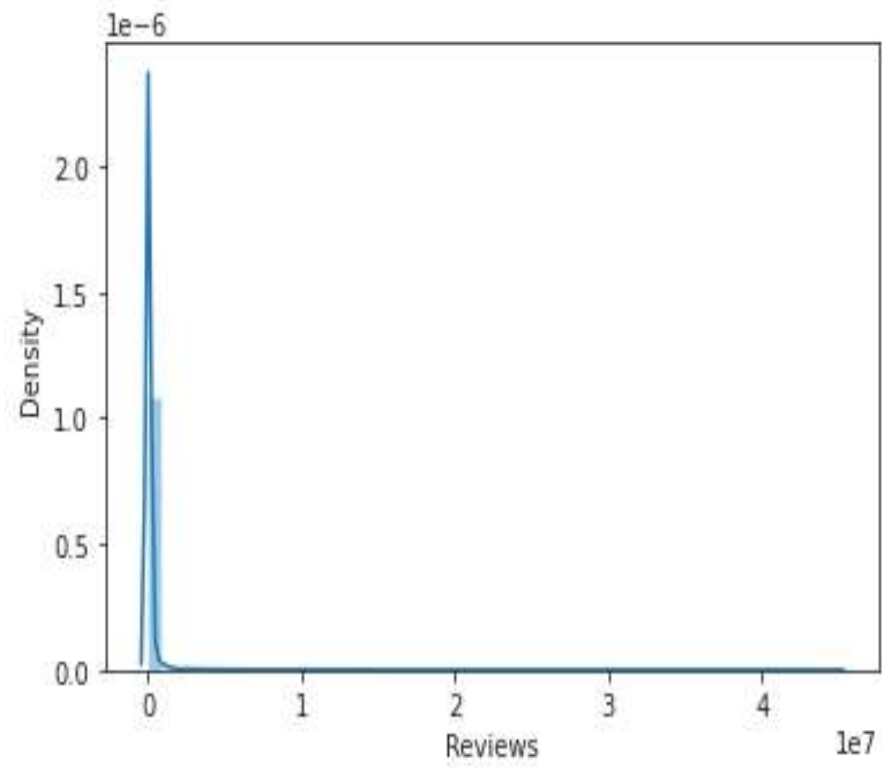


Rating vs Installs

Apps with rating between 3 to 5 have more installs

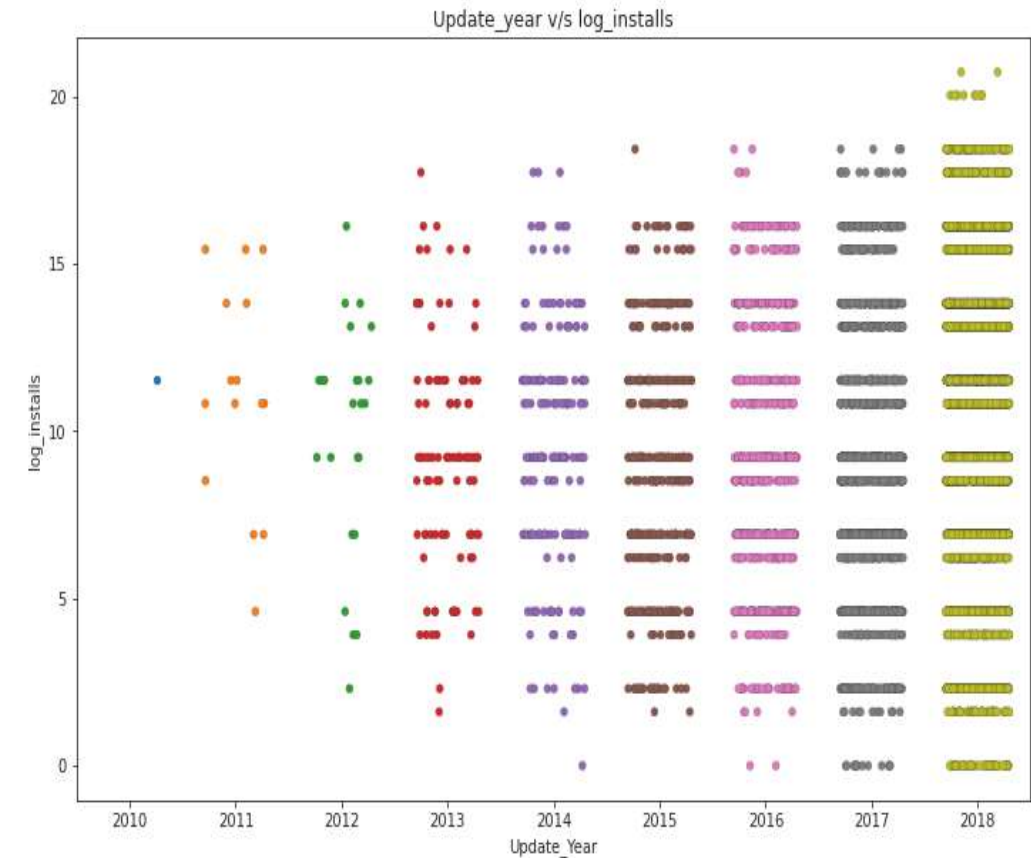
- users prefer highly rated app to download
Above graph shows positive relationship between Rating and Installation.
- The higher rating does contribute to more installations.
- People have a natural tendency to trust the opinion of those around them.





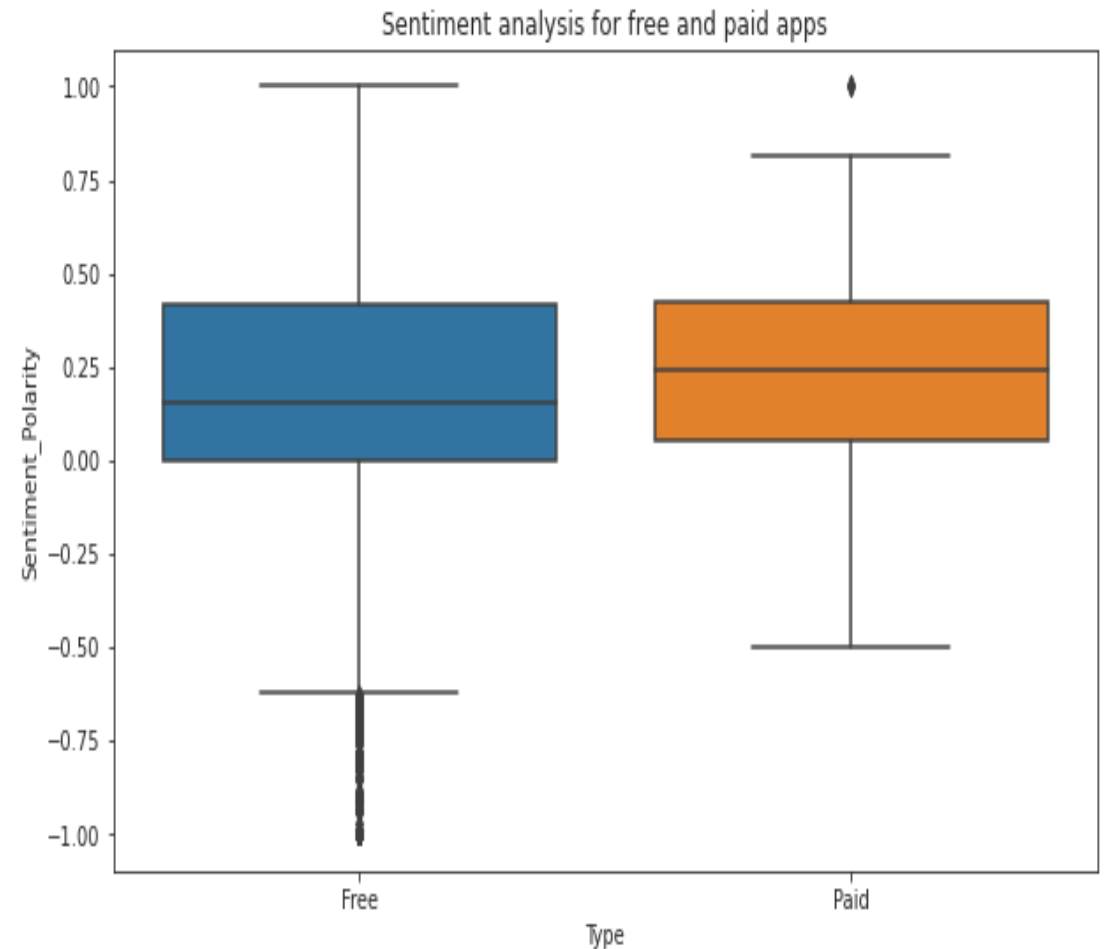
Update Year vs Log Installs :

- Most apps get frequent updates.
- Updated app has more installations
- so, 2018 apps has higher installation among all.



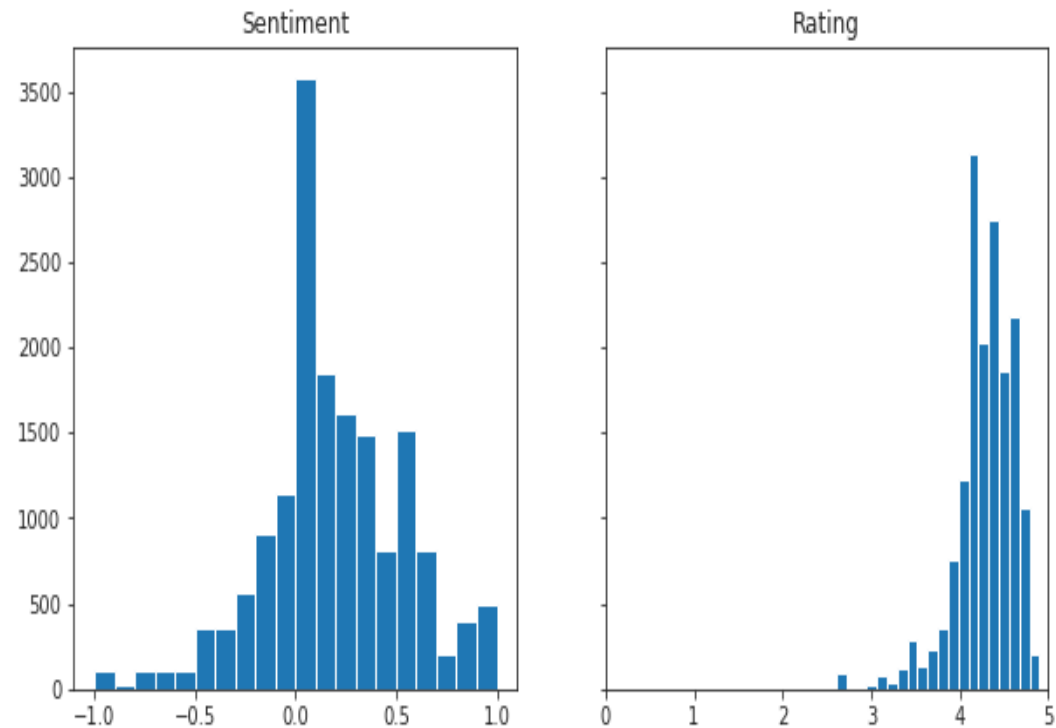
Sentiments Analysis For Free And Paid Apps :

- From the above scatter plot we can observe that the sentiment polarity is same in both the cases i.e in free app and paid app in the range between 0.00 to 0.50
- Free app has get more negative reviews as indicated by outliers on negative side.
- We can say that paid app has better quality reviews.
- Median polarity is higher for paid apps.



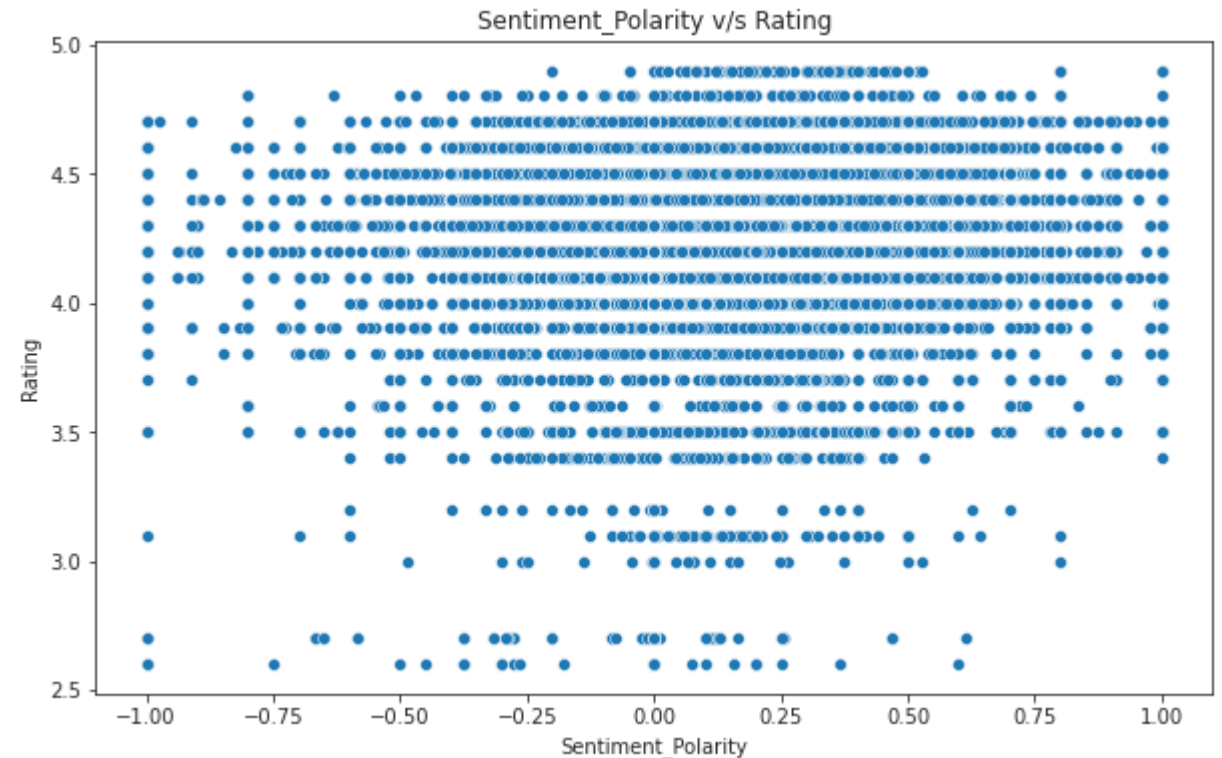
Sentiment Analysis for each Rating :

- More the rating more will be the sentiments



Sentiment Polarity vs Rating :

- Looking at the histograms, we can see a positive trend for both columns. In the Sentiment aspect, the majority of the values are above neutral (0), at around 0.2
- For the Rating column, most of the apps have got ratings above 3.5
- We can say that positive reviews about the app can strengthen new user's decision to download.



CHALLENGES FACED :

- ☐ Finding and removing the error, Duplicate and Null values in the data set.
- ☐ Understanding the multiple visualization to summarize the data.
- ☐ Having to read all reviews to analyze them.

Conclusion:

- Almost 91% apps are free and medical and Medical and personalization apps is good as well paid apps
- Users prefer the app having less size.
- Rating of app plays important role in downloading decision.
- Updating the apps improve the users decision towards downloading.
- Family category has more apps on play store.
- Game category is most popular among all apps.
- Sentiments in reviews is crucial in attracting users as other user's positive review strengthen the user decision towards downloading.

Q & A



THANK YOU