Capstone Project

EDA on Play Store App Reviews

By

Anson Sibi

Data Science Trainee , Almabetter

Problem Statement

❏ Two datasets are provided, one with basic information and the other with user reviews for the respective app.

❏ We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success.

So, what factors influence an app's success?

An app is said to be successful if it has:

❏ A high average user rating
❏ A good number of positive reviews
❏ A good number of users
❏ High revenue per customer and so on.

Data Summary

Play store Data

❏ App ❏ Category ❏ Size ❏ Rating ❏ Reviews ❏ Installs ❏ Type
❏ Price ❏ Content Rating ❏ Genres ❏ Last Updated ❏ Current Ver
❏ Android Ver

User Review Data

❏ App ❏ Translated Review ❏ Sentiment ❏ Sentiment Polarity
❏ Sentiment Subjectivity

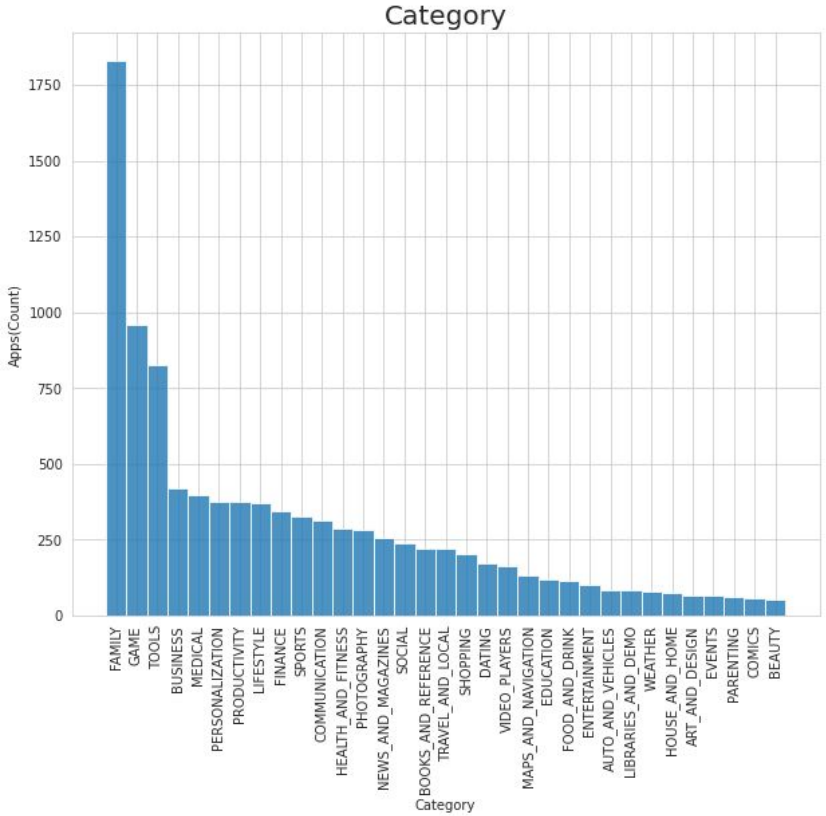
Pairplot to get an overall idea of the dataset

Correlation Heatmap

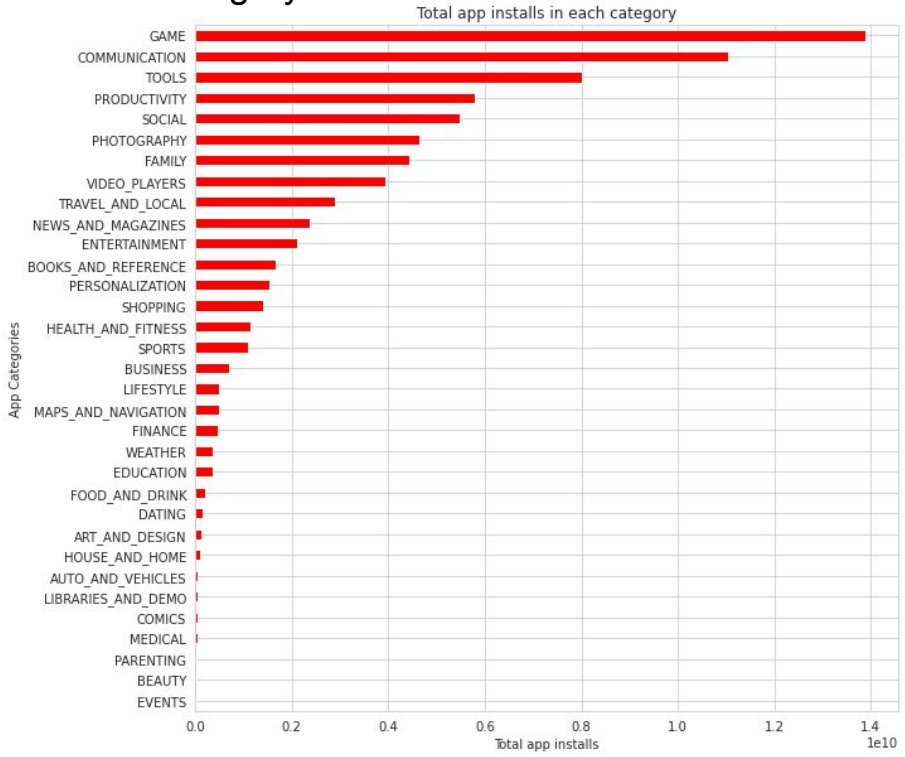
- There is a strong positive correlation between the **Reviews** and **Installs.**
- The Price is slightly negatively correlated with the **Rating, Reviews,** and **Installs.**
- The **Rating** is slightly positively correlated with the **Installs** and **Reviews.**

Categorical Analysis.

Number of apps in each category.



Number of apps installed in each category.



Contind..

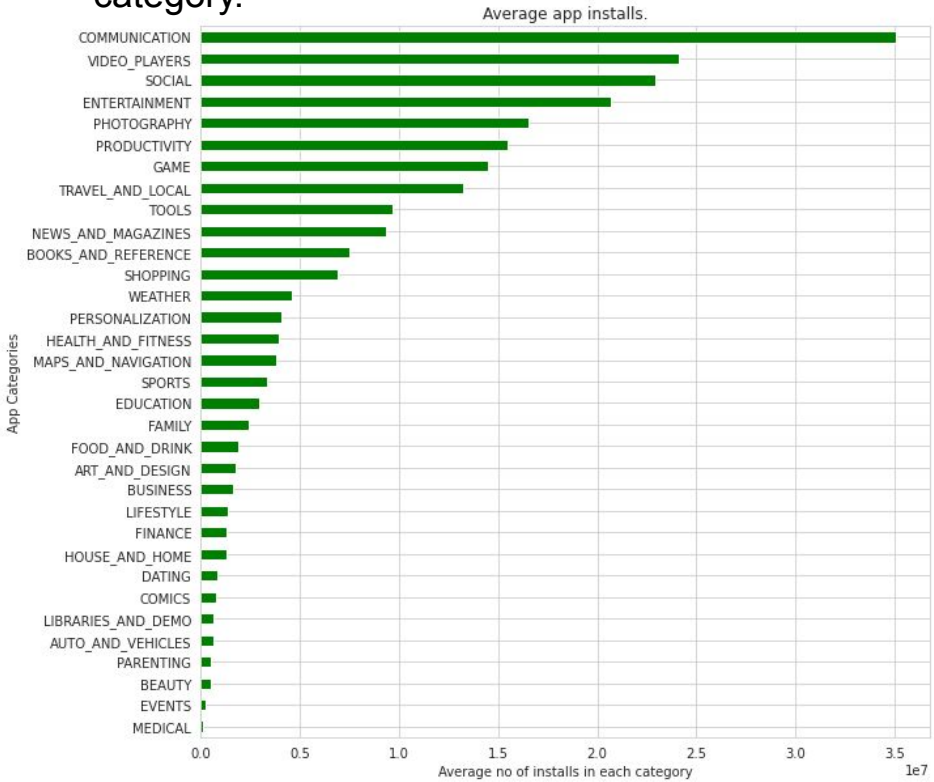
- The first five categories which contain the highest no of apps present in play store are **Family , Games , Tools , Medical** and **Business**.

- The first five categories which contain the highest no of installations in play store are **Games , Communication , Productivity , Social** and **Tools.**

- The first five categories which contain the highest average no of installs in play store are **Communication , Social , Video players , Productivity** and **Games.**

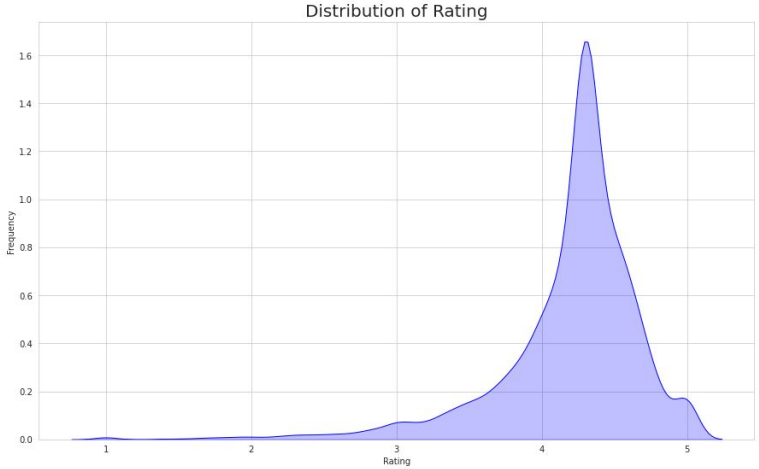
- **Communication , Video players ,Social ,Entertainment** and **Photography** categories contain the higher percentage of installation with respect to no of apps . So deploying apps in that categories may result in success.

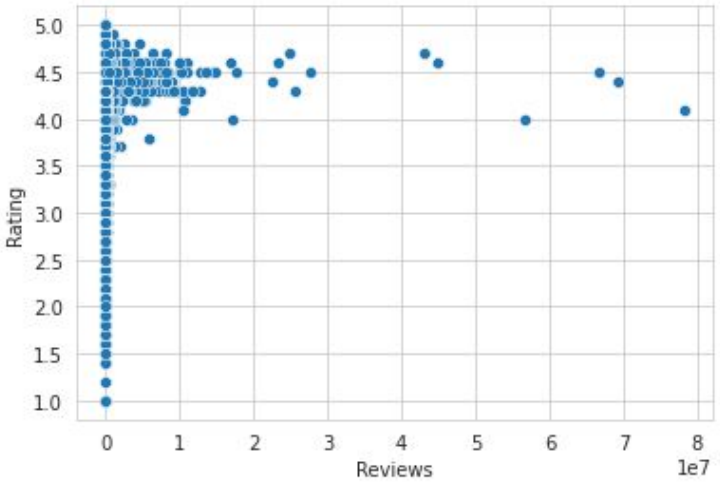
Average number of app installs in each category.



App rating analysis.

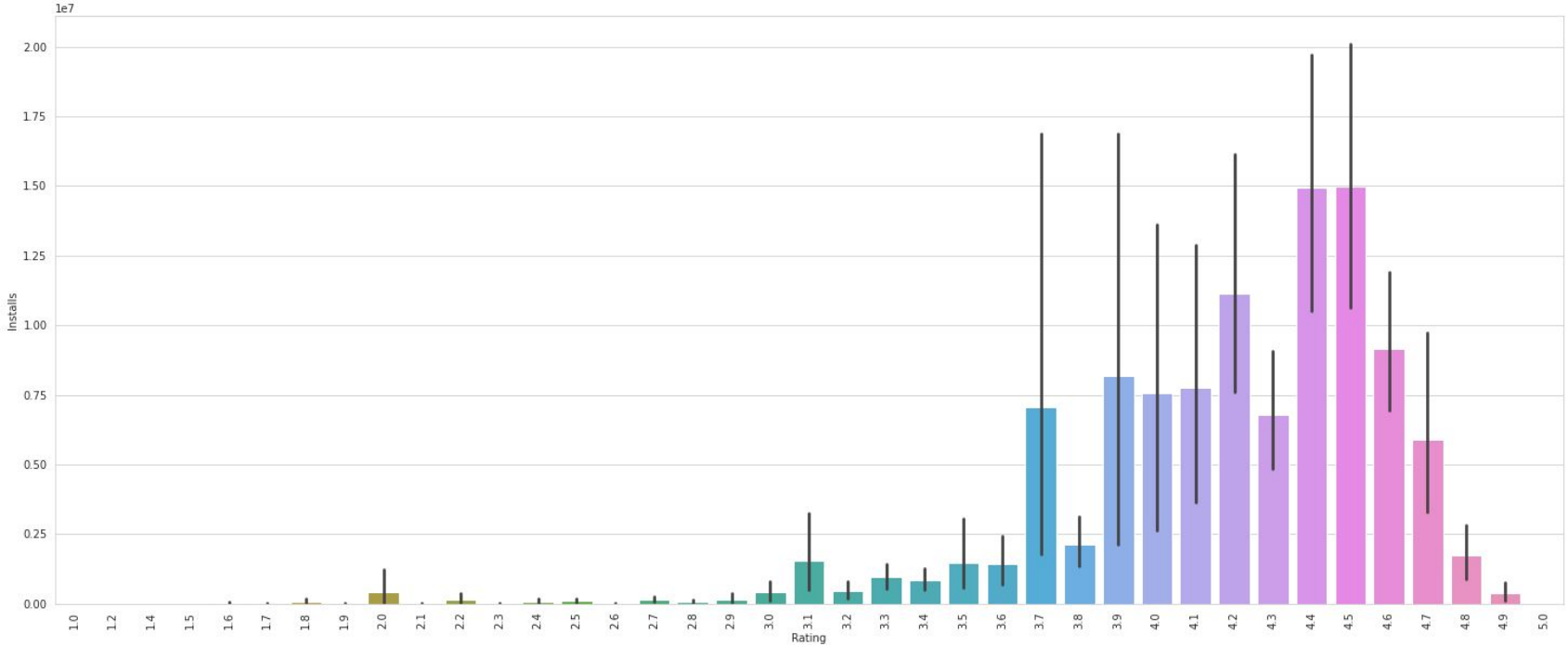
- From the distribution plot we get to know that majority apps are rated between 4 and 5.
- Average rating of apps(mean)=4.192. Median value of rating of apps =4.3.
- From the scatter plot , it's evident that as the number of reviews increase the rating also increases.
- This implies that majority of the users are happy with the service received via the respective app.





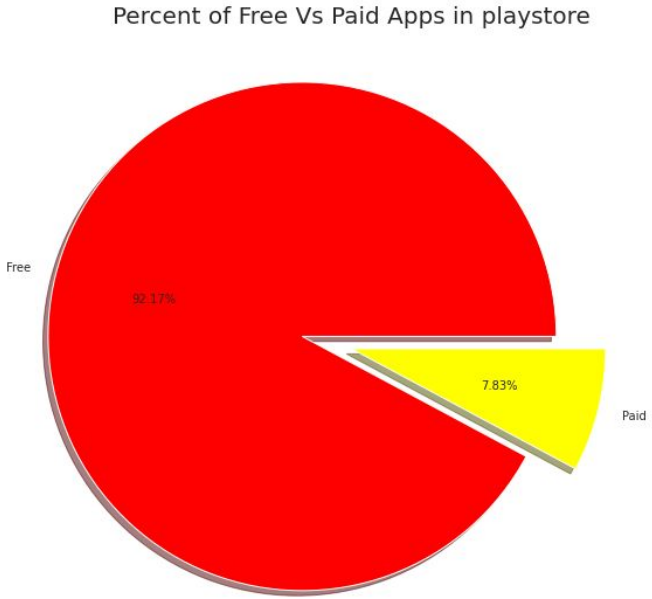
App rating analysis (Contind..)

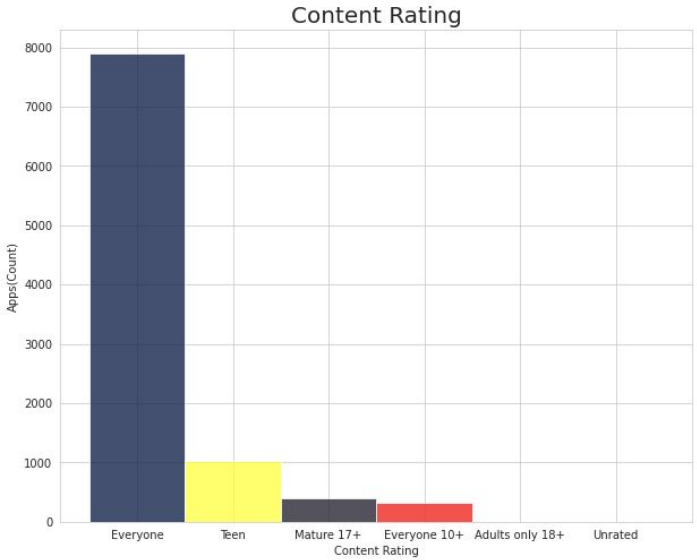
- Its obvious from the plot the rating is rightly skewed and as the no of installs increases rating also increase . We can conclude that more installed apps maintain there quality as a result they receive good rating.



Type and Content Rating.

- Almost 92.17% of all the apps in play store are free. And there is 7.83% of apps that requires to be paid.

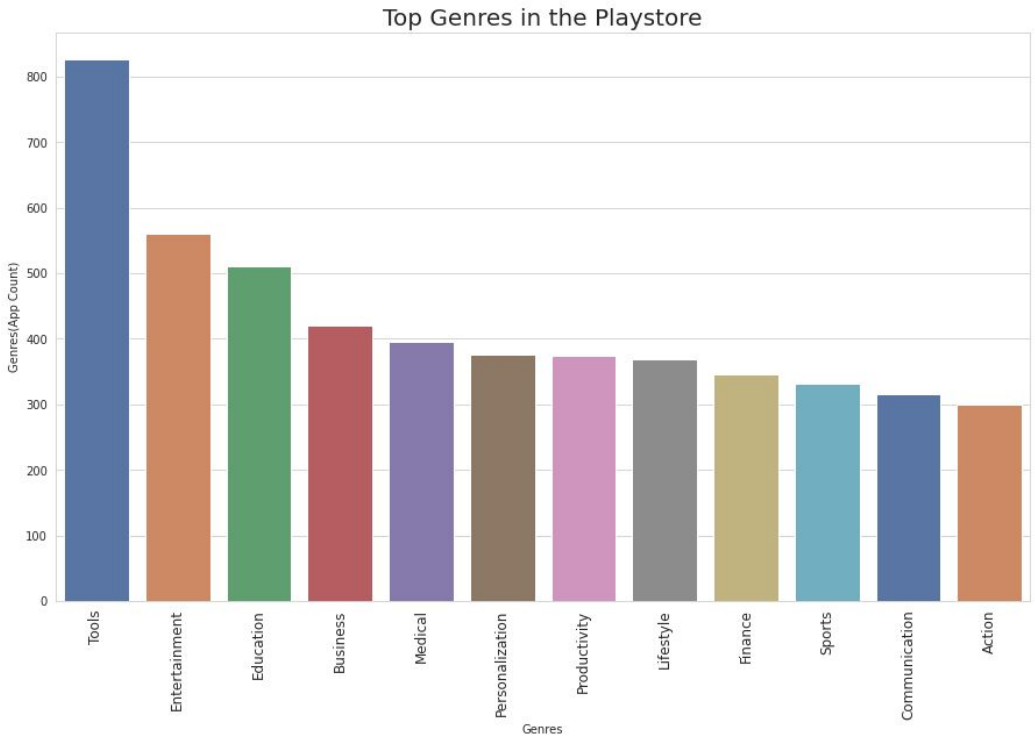




- From the content rating plot its clearly observable there are large number of apps under the tag Everyone(means there is no age restriction to use it).

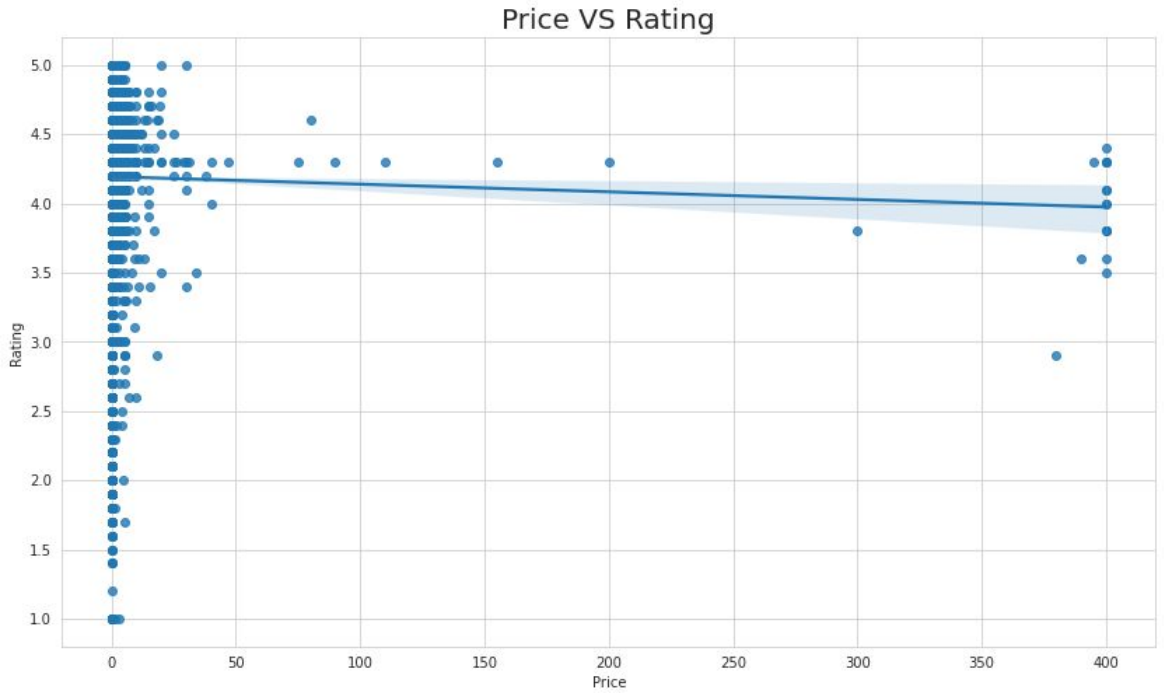
Top Genres In Play Store

- Tools , Entertainment , and Education are the top three genres by the number of apps.
- It can be concluded that , higher number of apps in these genres are the result of higher number of demand from the customers. Therefore choosing the right genre is a key factor for app success.



Let's See : How **Price** of apps affects its **Rating**..

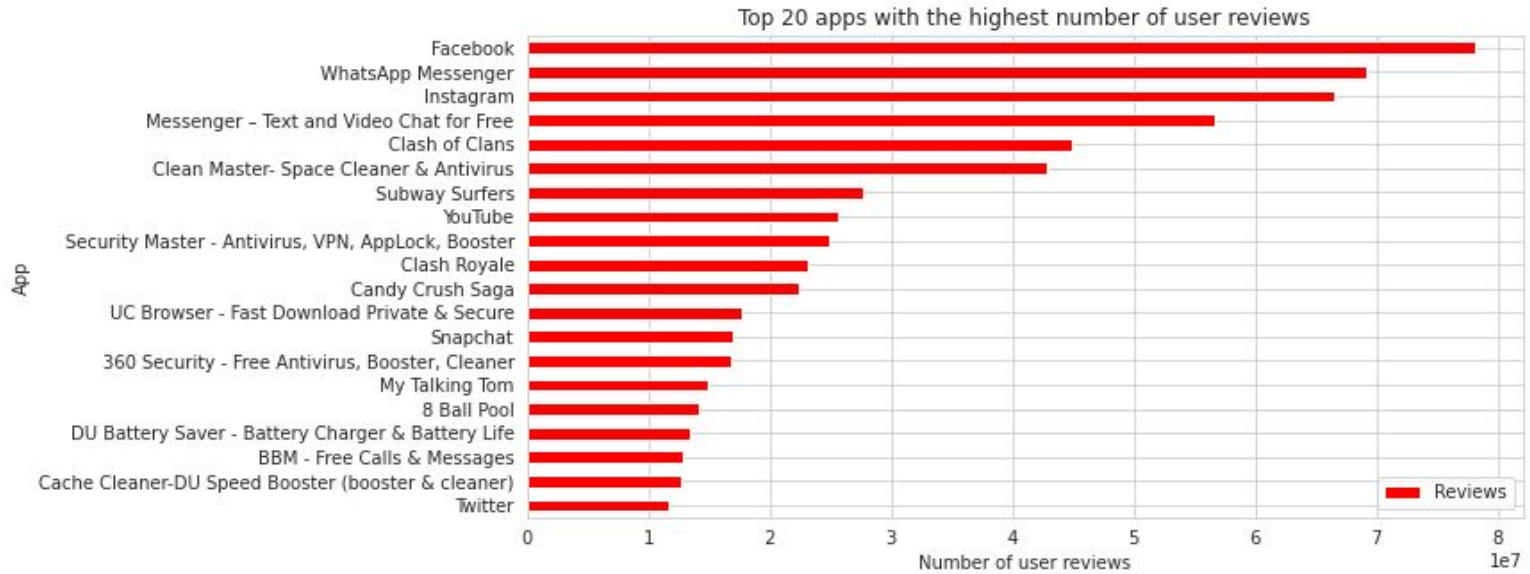
- As the price increases ratings received seems to decrease even below the average rating in the Appstore.
- This suggest that customers are more conscious about the quality of apps , when they are paid as compared to free apps.



Most Popular Apps

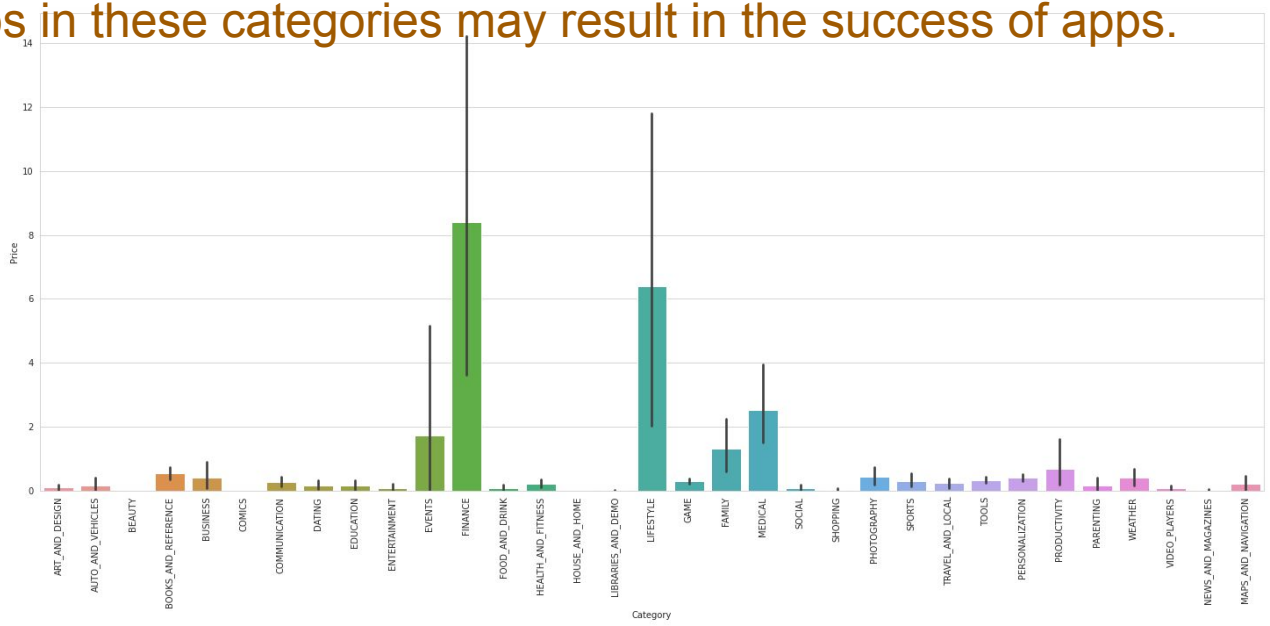
❏ We can state that the apps with more reviews, whether positive, negative, or neutral, are more popular than the others.

❏ This is because the number of user reviews indicates that these individuals have engaged with the app's content and have written their opinions on it.

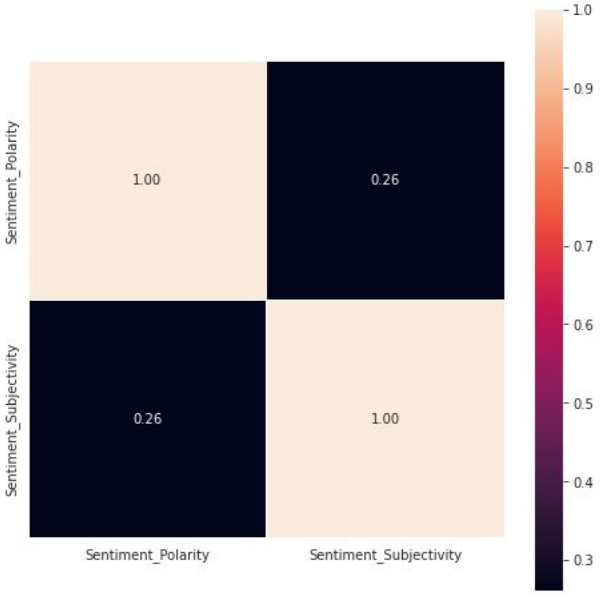


Which categories of apps are most expensive..?

- The apps belonging to Finance , Lifestyle ,Medical categories are the most expensive ones . Launching apps in these categories can get customers willing to spend money on apps.
- Launching premium apps in these categories may result in the success of apps.
- We can also conclude that Finance category contains large number of apps which are of high prized



User Reviews

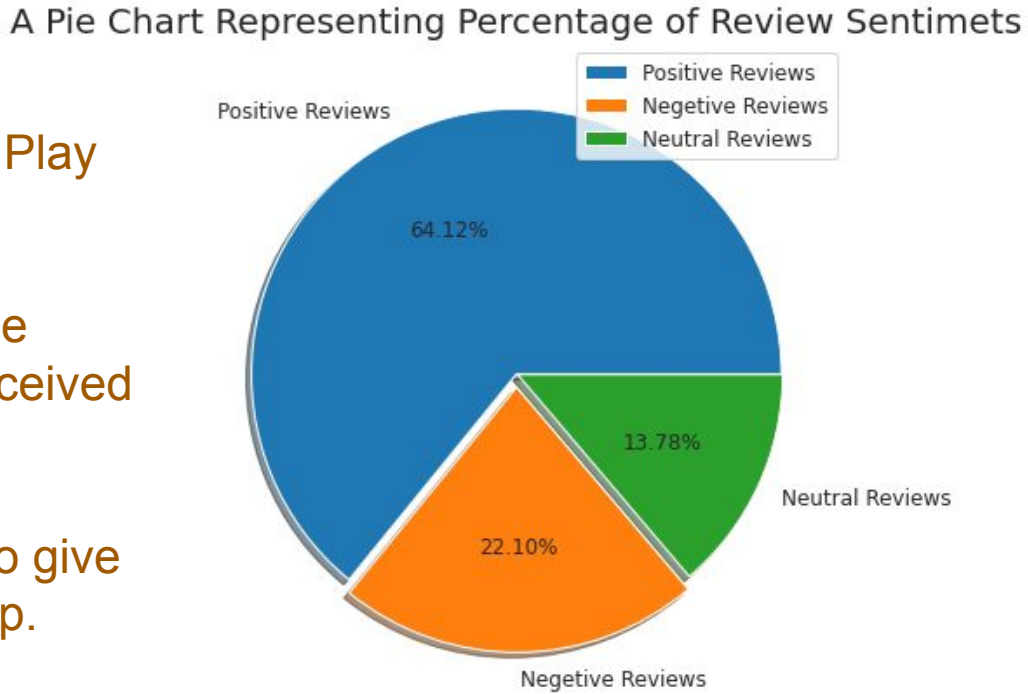


- The first correlation heat map is of user review dataset. There is a strong linear relationship between sentiment Subjectivity and Sentiment Polarity.

- The second correlation heat map is of the merged two datasets.It gives us an overall idea regarding the nature of whole dataset.

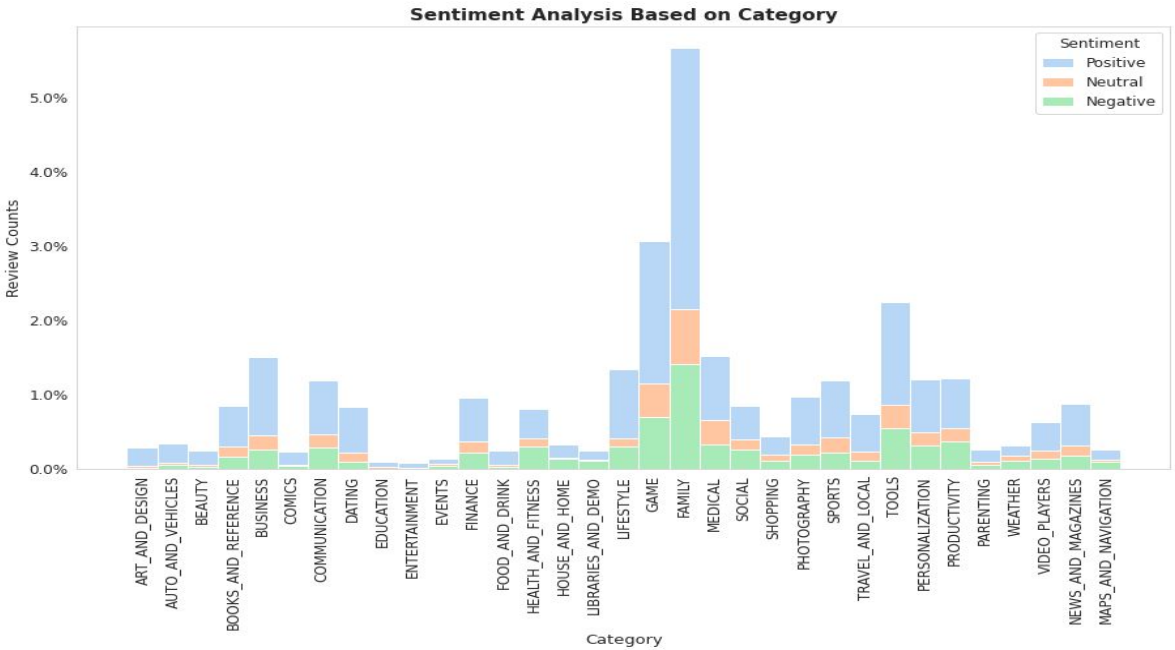
Sentiment Analysis.

- The majority of the reviews in the Play Store (~64%) are positive reviews.
- This implies that the majority of the users are happy with the services received via the respective app.
- From the pie plot customers like to give positive reviews if they like the app.
- About 22.10% of all the reviews are negative.



Sentiment Analysis based on category.

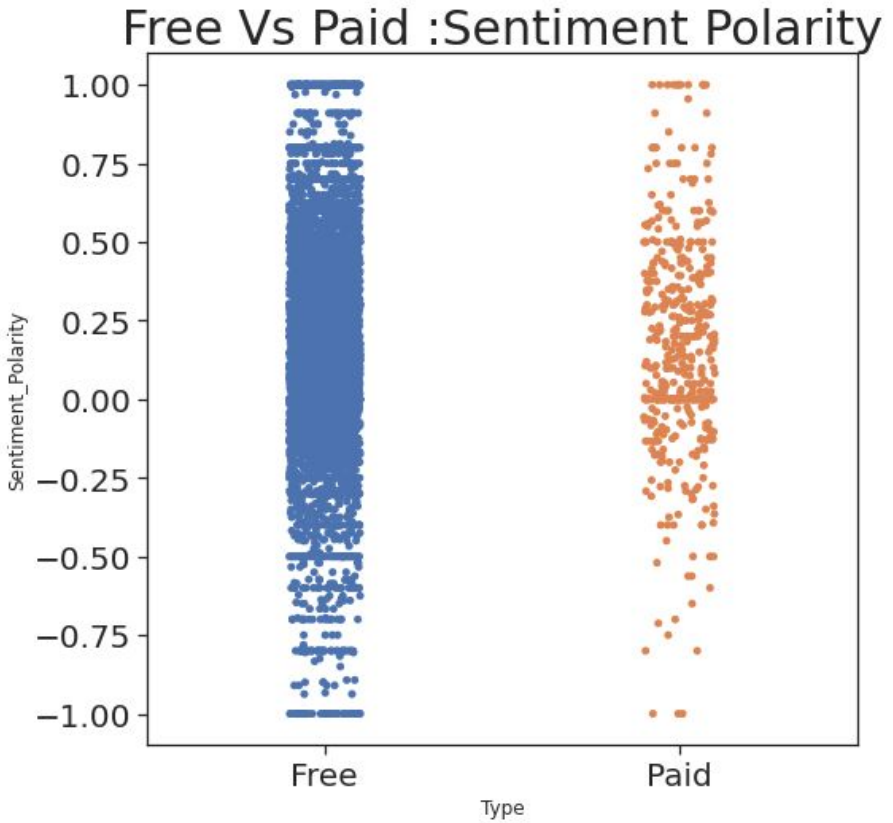
- The graph shows fraction of reviews with positive , negative and neutral characteristics with respect to each categories.
- From the graph it is evident that majority of apps in each category has higher no of positive reviews than the negative and neutral reviews.



Sentiment polarity

- Type vs Sentiment Polarity graph clearly shows that Paid Apps are less likely to receive bad reviews .

- Free apps have a more equal distribution of Sentiment Polarity.



Challenges Faced

❏ Reading the dataset and comprehending the problem statement.

❏ Examining the business KPIs for app development and devising a solution to the problem.

❏ Handling the error, duplicate and NaN values in the dataset.

❏ Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.

Analysis Summary

- Percentage of free apps = ~92.62%
- Percentage of apps with no age restrictions = ~82
- Most competitive category: Family
- Category with the highest number of installs: Game
- Category with the highest average app installs: Communicaction
- Average rating of apps(mean)=4.192
- Median value of rating of apps =4.3
- The top 10 apps by the number of installations are free.
- Category in which the paid apps have the highest average installation fee: Finance
- Most popular app in the Play Store based on the number of reviews: Facebook
- Top three Genres by the number of apps are Tools , Entertaiment and Education.
- The app which having highest number of reviews is Facebook.
- 64.12% of all reviews are positive.

Conclusion and inferences.

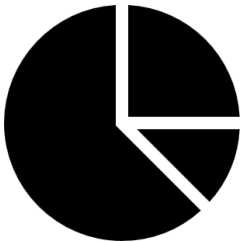
- Apps which have more no of reviews are tend to be downloaded more by people . Therefore getting higher number of reviews is a key factor for app success.

- Communication , Social , Video players , Productivity and Games categories contain the higher percentage of installation with respect to no of apps . So deploying apps in that categories may result in success.

- Number of installation of free app is more compare to Paid app, so developer can also consider this point for high reach.

- As the top 10 of most installed apps are free . So free apps have a wider set of audience as compared to paid apps.

- As the price increases ratings received seems to decrease even below the average rating in the appstore. Social and Communication apps like Facebook , Instagram and whatsapp have the highest number of downloads ,apps in this category has higher chance of success.

- Type vs Sentiment_Polarity graph clearly shows that Paid Apps are less likely to receive bad reviews and Free apps have a more equal distribution of Sentiment Polarity.

THANK YOU