Capstone Project
Play Store App Review Analysis

By: MANISH GUPTA

Points for discussion

1. **Defining the problem statement**
2. **Introduction to Play store Platform**
3. **Exploring the database**
4. **Features selection**
5. **Top 'Genres' and their analysis**
6. **Free apps VS Paid apps**
7. **Impact of 'Rating'**
8. **Users Subjectivity and Sentiment**
9. **Conclusion**

AI

Defining the problem statement

"**The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.**

Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.

Explore and analyze the data to discover key factors responsible for app engagement and success."

- From the problem statement given, we should analyze the given database and should come up with the key factors that increased the number of users, long term usage and etc.,

- The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

Introduction to Play store Platform

- Play store is an Android Market serves as the official app store for certified devices running on the Android Operating system.

- Developed and Operated by Google, launched on 6th March, 2012.

- Also called as Google Play store.

- Approximately 3.48 million apps are in Play store.

- Each apps are having separate product page with user review and rating system.

- All the apps are separated into different categories based on the purpose.

Exploring the database

We have been provided with two databases

- Play store app database

  - Shape of this database is (10841, 13).

  - Out of this thirteen columns we have numeric

    columns like Rating, Reviews, Size, Installs.

- User reviews database

  - Shape of this database is (64295, 5).

  - Here there are only two numeric values found,

   Sentiment Subjectivity, Sentiment Polarity.

Features selection

The columns are also known as features, one or more different features are grouped together for different analyses to form a data frame.

- **top_genres_df:** It contains the rows of top 'Genres' based on the app counts.
- **genres_free_apps_installs_df:** It holds the rows of top 'Genres' and corresponding total free app installs count.
- **genres_paid_apps_installs_df:** It holds the rows of top 'Genres' and corresponding total paid app installs count.
- **genres_ratings_df:** It takes the rows of top 'Genres' and their corresponding mean Rating.
- **top_50_genres_df:** This data frame contains the rows of features top Genres, Free app installs count, Paid app installs count and Rating.
- **category_type_installs_df:** It contains the rows of Category wise free and paid apps installs count.
- **price_df:** This holds the rows of paid apps' non null price values.
- **category_price_mean_df:** This holds the rows of Category wise mean price of paid apps.
- **rating_df:** This holds the rows of mean price for all the apps.
- **category_mean_rating_df:** It contains the rows of Category wise mean rating of all the apps.
- **content_rating_df:** It holds the rows of different aged people wise ratings count.
- **non_null_user_reviews_df:** It holds the different users' non null Sentiment type, Sentiment Subjectivity distribution and Sentiment Polarity.
- **sentiment_subjectivity_df:** It holds the different users' sentiment subjectivity distribution.

Top 'Genres' and their analysis

The analyze is done between the top 50 Genres which was having more number of downloads


Top 50 Genres VS App Counts
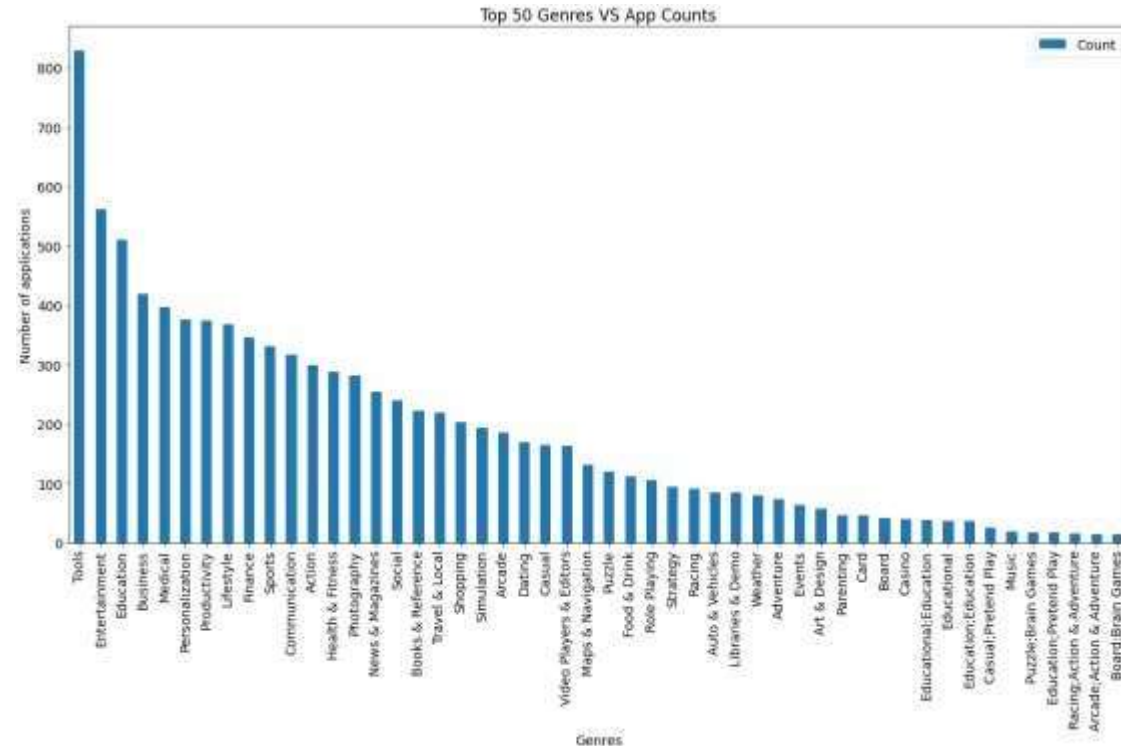
- As we saw here, the Play store having more number of applications in the genres like Tools, Entertainment, Education and etc.,

- The developers are mostly focusing on these, genres because of the people daily basis requirements.

- Genres likes Educational, Parenting, Music are having comparatively less amount of apps count.

# Free apps VS Paid apps Installs (Top genres based)
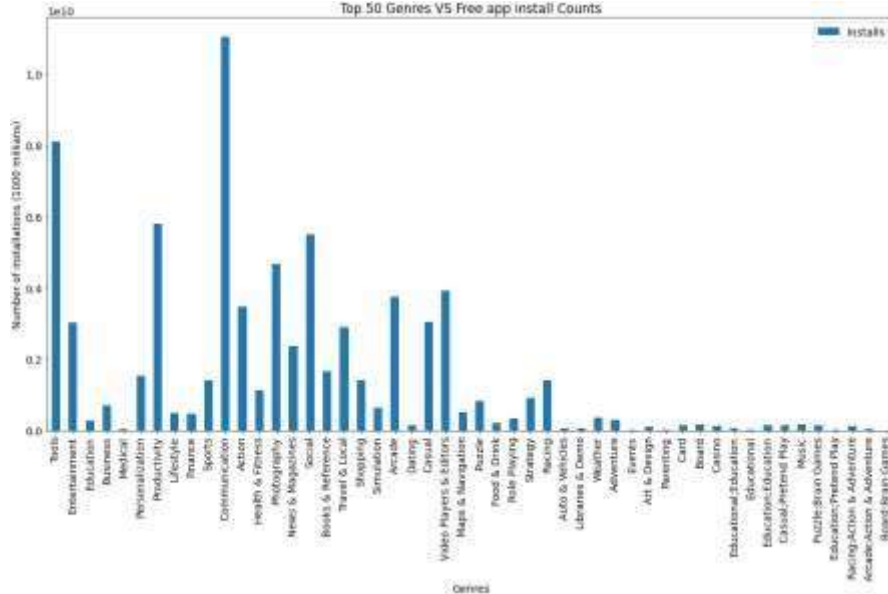
- When comparing the both plots, people are showing more interest on free apps like communication, Tools.

- Again the Educational, Parenting and Music are the genres in the least.
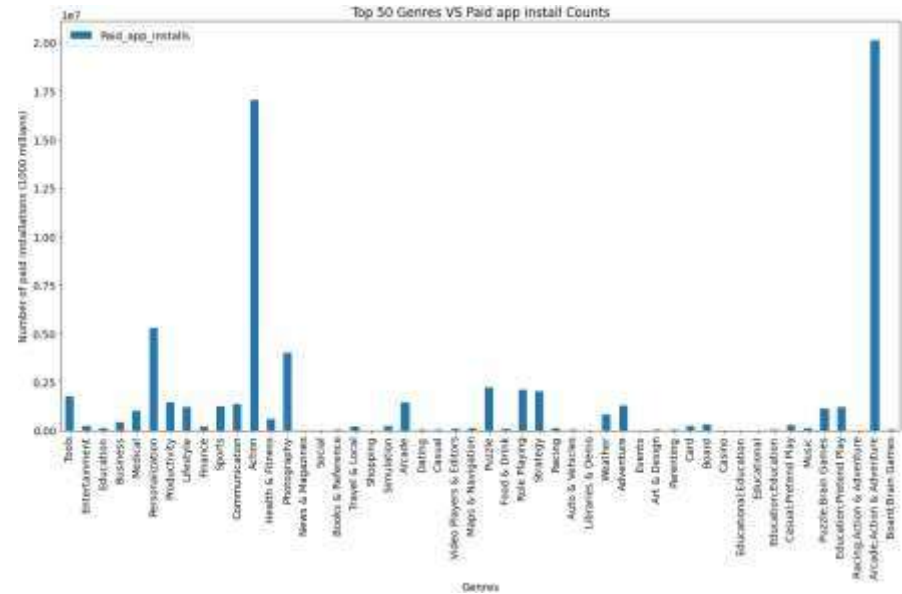
- When it's coming commercial people are preferring the apps like Games, Photography, Personalization.

- People are preferring less on Educational, Event, Art & Design.

Free apps VS Paid apps Installs (Category based)

Paid app install Counts(in millions) | Free app install Counts(in 100 millions)
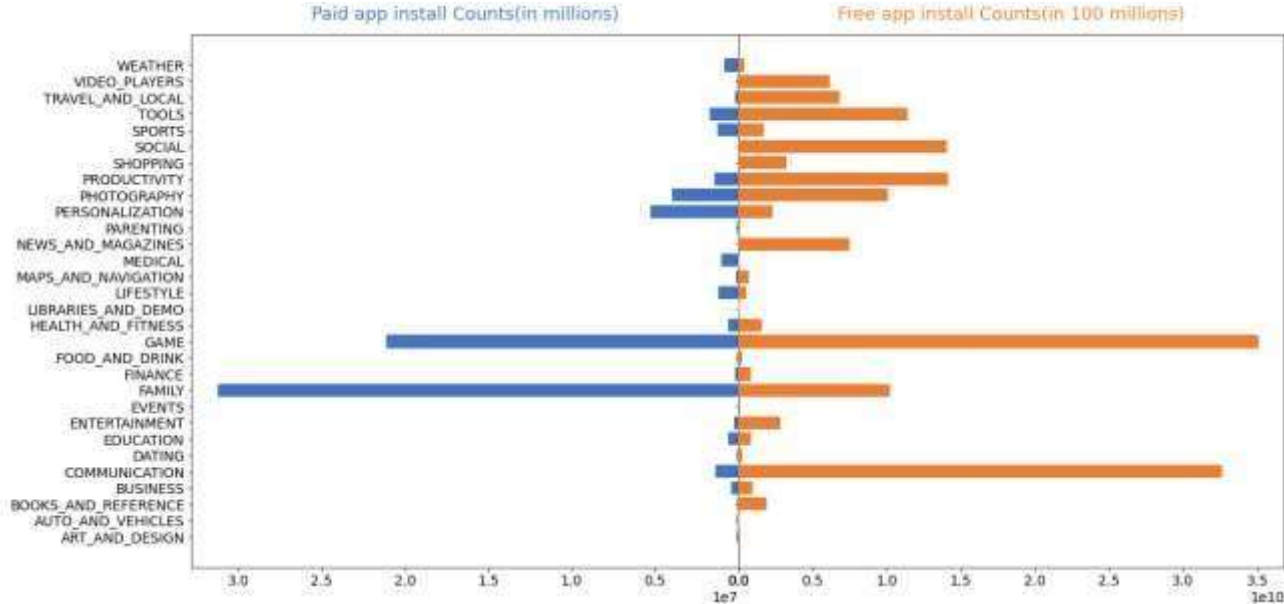
- Each categories are showing huge difference between their downloads comparison.
- People are spending more money in Family and Game.

- The overall category wise difference can be seen in Game.

- Communication also having some considerable amount of installed number in Free apps.

Price distribution
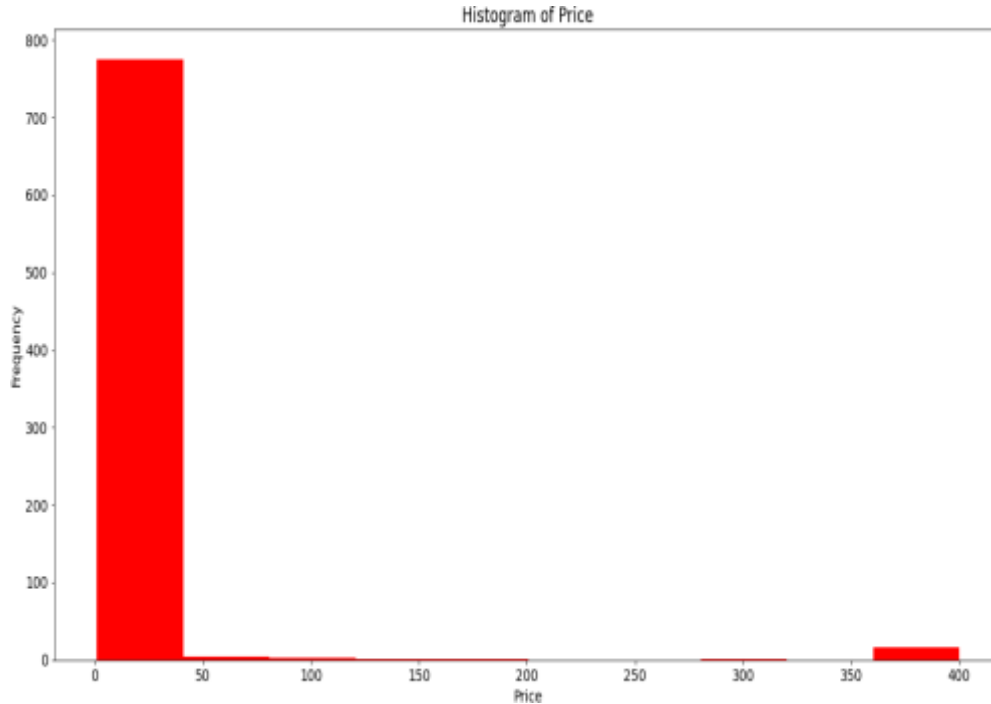
- The distribution of Price plotted as we see here.

- Most of the paid apps are in between the price range of 0 to 50 USD.

- Few apps are in the price range of 350 to 400 USD.

- Very few of the apps pricing between 75 to 200 USD.

Impact of 'Rating'

Histogram of Rating
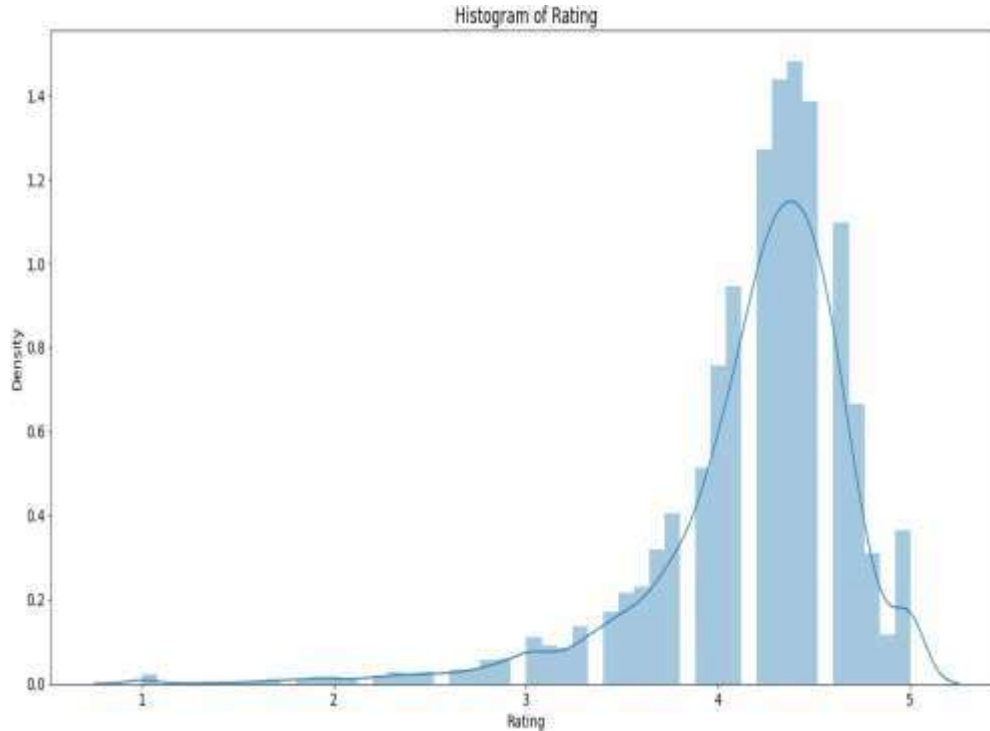
- In general Rating is the main scale factor, measures that how much people are satisfied with the product.

- Here, most of the people rated the apps between the range of 4 to 5, which are can be considered as good.

- Moderate rating is lies between 2.5 to 4.

- And below average or poor apps are in the range of 0 to 2.5 which are less in count.

Impact of 'Rating'

- The content rating shows the results for general contents as high.

- The content rating type 'Everyone' has the most percentage value of 80.39%.

- 'Teen' contents are second in the order with the percentage of 11.14%.

- Adults only and Unrated contents are least in this plot, 0.03% and 0.02% respectively.

- So it can be concluded that most of the contents are generic.

Users Subjectivity and Sentiment

Distribution of Subjectivity
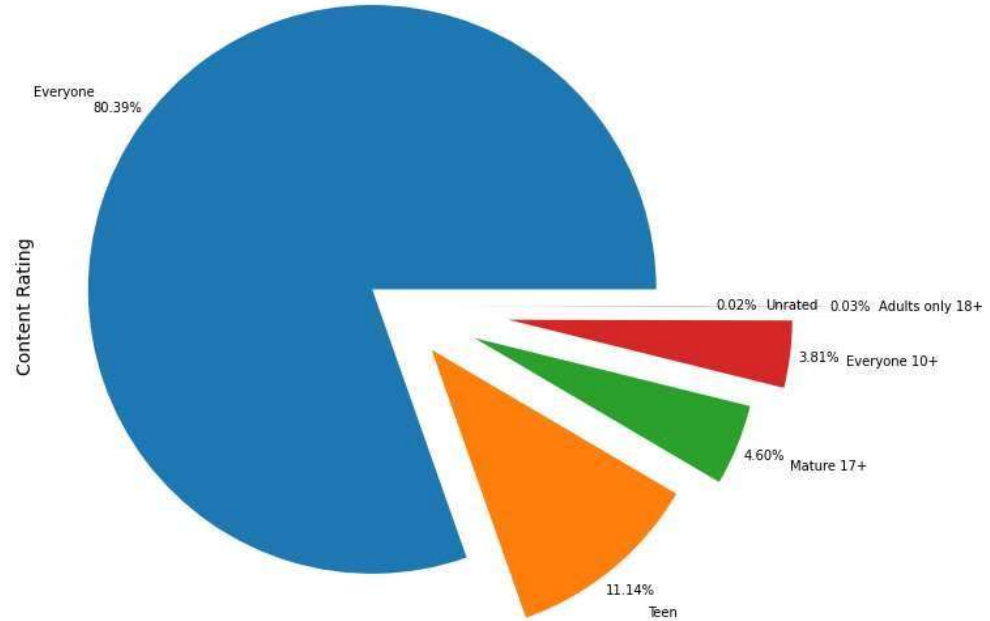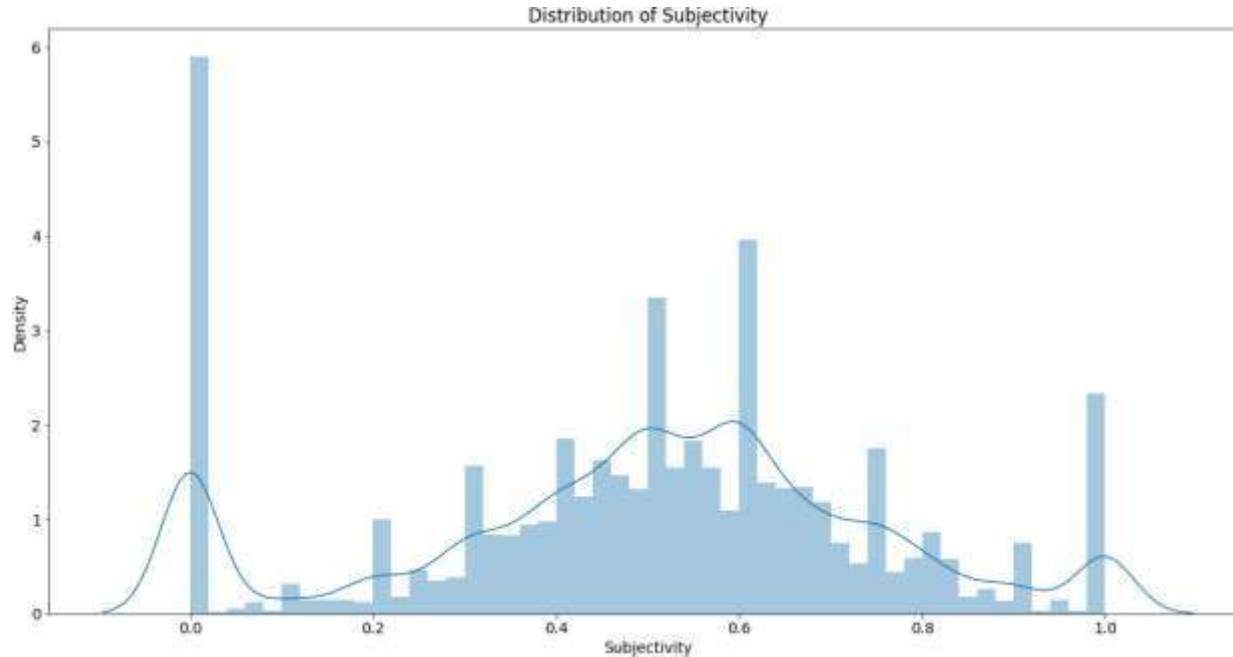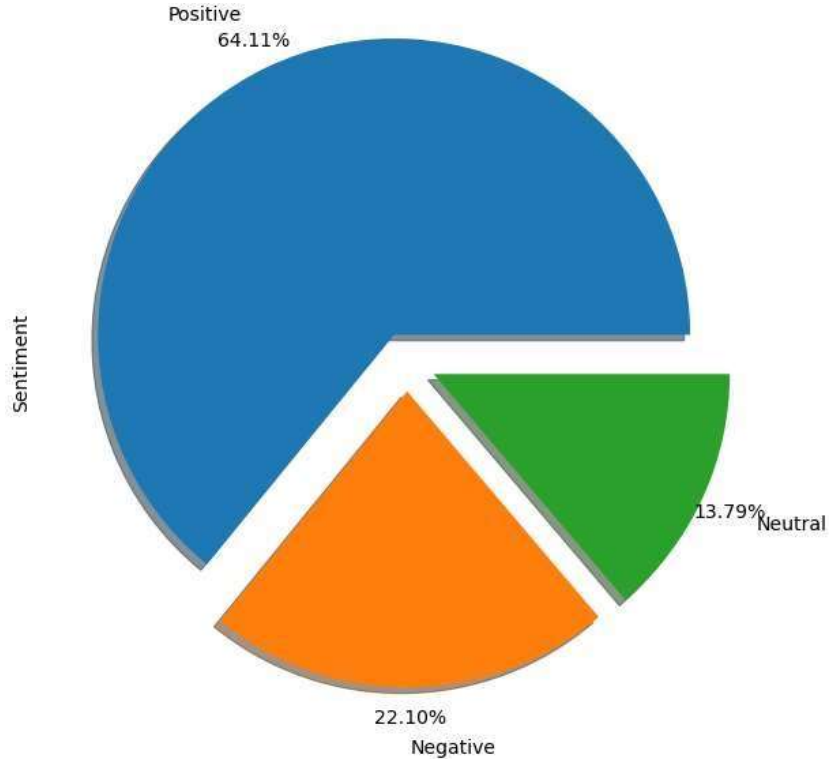
- Subjectivity are lies mostly between 0.5 and 0.65.
- It shows that the average content and apps reviews subjectivity are mostly relevant.
- Subjectivity of 100% has slightly occurs frequently.
- The nearly 0 subjectivity has a considerable amount of frequency.

Users Subjectivity and Sentiment (Continued)

- The sentiment plot shows the results for positive reviews as high.
- Sentiment type 'Positive' has the most percentage value of 64.11%.
- 'Negative' reviews are with the percentage of 22.10%.
- 'Neutral' percentage has 13.79%.

Conclusion

After analyzing the dataset we have got answers to some of the serious & interesting facts which any of the android users would love to know.

- Top Genres on Google Play store.
- Top Categories on Google Play store.
- Which Category of Content is downloaded more?
- Which category of apps has the most number of installs?
- What are the Top 10 installed apps in different categories?
- Distribution of the ratings of the apps
- Variation between Free and Paid apps
- Which are the top expensive Apps?
- Distribution of the Price of the apps
- Which are the apps that have made the highest-earning?
- Which are the Apps with the highest number of reviews?
- Developing apps related to the least categories as they are not explored much. Like events and beauty.
- Most of the apps are Free, so focusing on free app is more important.

Conclusion

- Focusing more on content available for Everyone will increase the chances of getting the highest installs.
- We have to focus on updating their apps regularly, so that it will attract more users.
- We have to keep in mind that the sentiments of the user keep varying as they keep using the app, so they should focus more on users needs and features.

Thank You