

Capstone Project

On

EDA on Play Store App Review

By

Ashokanand Chaudhary

Objective

- **To Perform Exploratory Data Analysis**
- **Sentiment Analysis of The Customer Reviews**
- **Discover the relationship between various variables**

Points for Discussion

- Summary of the dataset
- Data cleaning
- Word Cloud of App
- The Best App
- Top Category vs Broad Genres vs Sub Genres
- Top Category by Installs
- The number of App present in each category
- Size vs Price
- Facet Grid of Content's Subset Category

Continue.....

Points for Discussion

- Most popular Content Rating
- Decoding Rating
- The share of the app in Sub Genre
- Audience Sub-target relationship with sentiment polarity and size
- Review by Size and Type
- Visualization of the Category on the basis of type with the most Downloaded
- Sentiment analysis
- Feedback analysis according to the Sentiments
- Relationship between Sentiment Subjectivity and Sentiment Polarity
- Correlation Heatmap
-

Points for Discussion

- Pair Plot
- Challenges faced
- Solution to Business Objective
- Conclusion

Dataset Summary

- Both of them are publicly available datasets.
- **Store Dataset:**
- It has rows more than 10k rows and 13 column
- **Review Dataset:**
- It has more than 64k rows and 5 column

Columns of Store Dataset

- App
- Category
- Rating
- Reviews
- Size
- Install
- Type
- Price
- Content Rating
- Genres
- Last Updated
- Current Updated
- Android Ver

Columns of Review Dataset

- App
- Translated Review
- Sentiment
- Sentiment Polarity
- Sentiment Subjectivity

Data Cleaning

Handling Null

Handling Inappropriate field

- Column
- Translated_Review
- Sentiment
- Sentiment_Polarity
- Suintiment_Subjectivity
- Rating
- Content Rating
- Current Ver
- Android Ver

Data Cleaning

Handling Null

Handling Noise

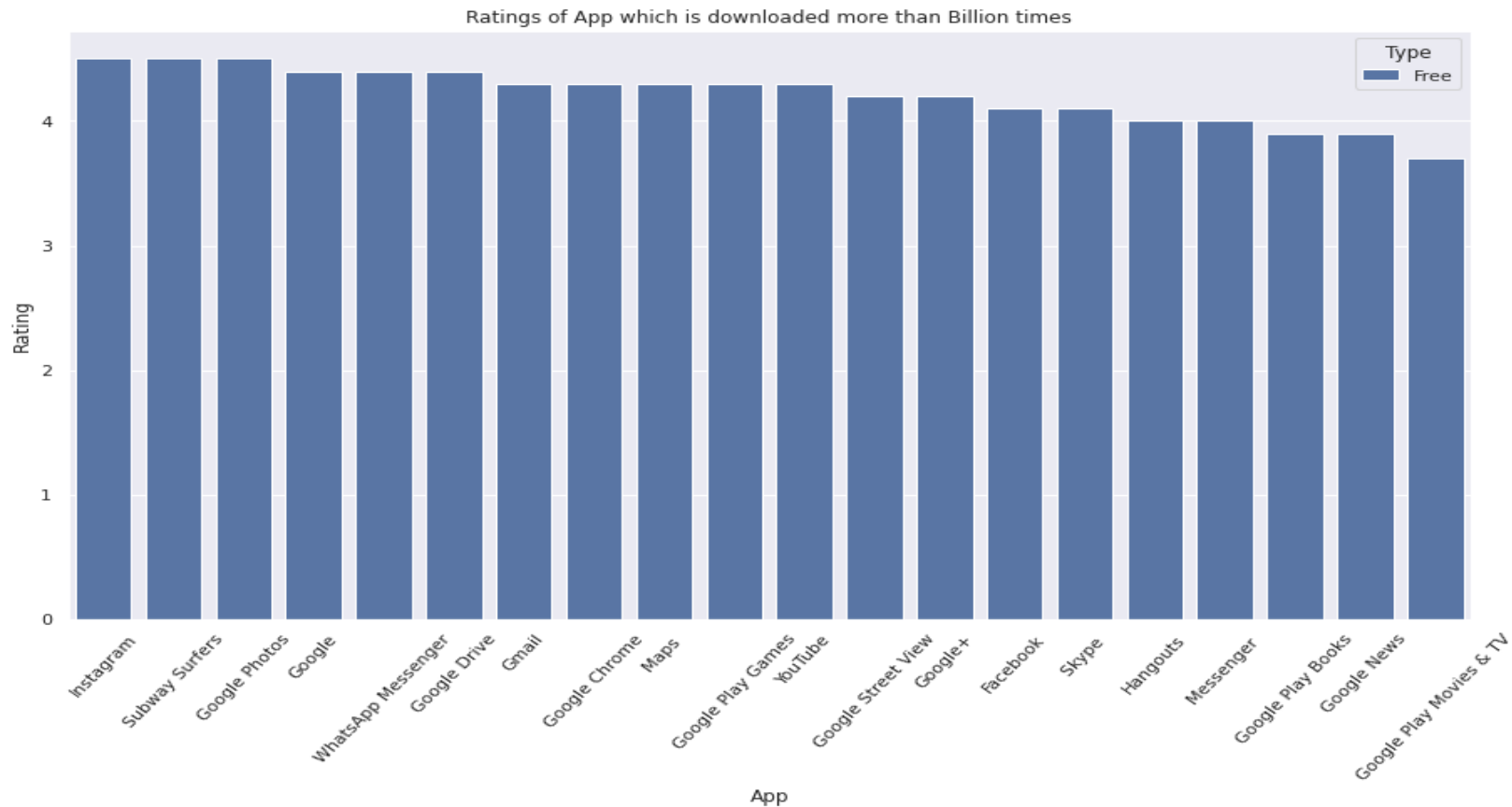
- App Short Name
- Rating Outlier, Int Type
- Size 'M' &'K', Int Type
- Install '+', ',', Int type
- Price '\$'
- Genres Two column

Word Cloud of App

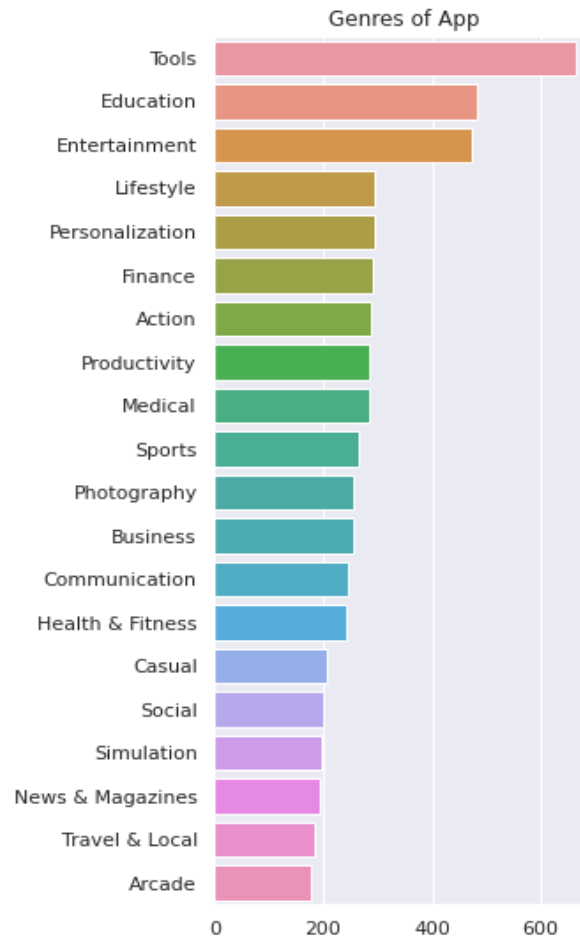
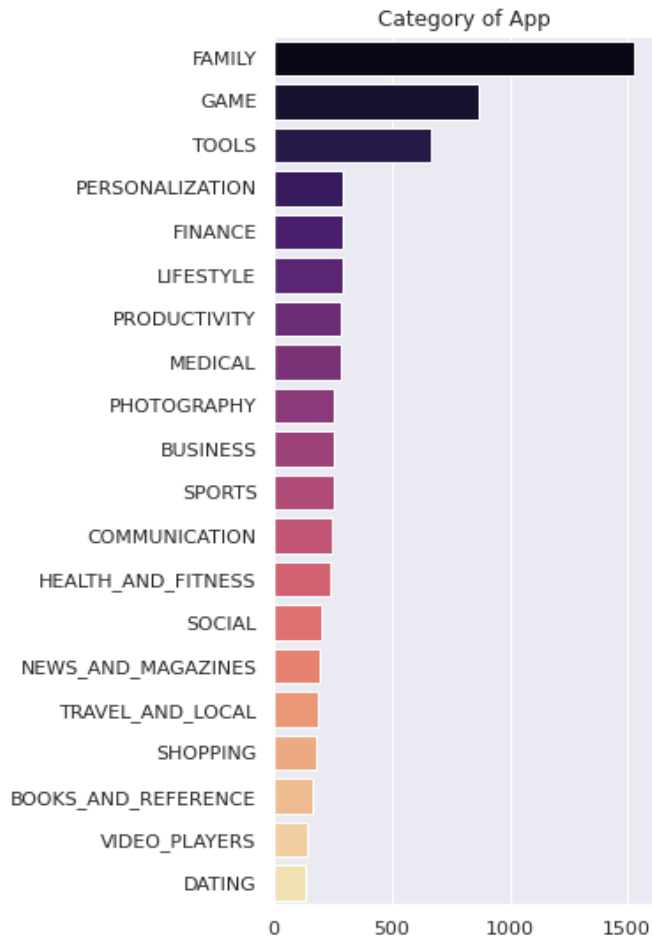
- Most used word in apps are Free, mobile, Game, Pro, Live, Phot, Video, Launcher



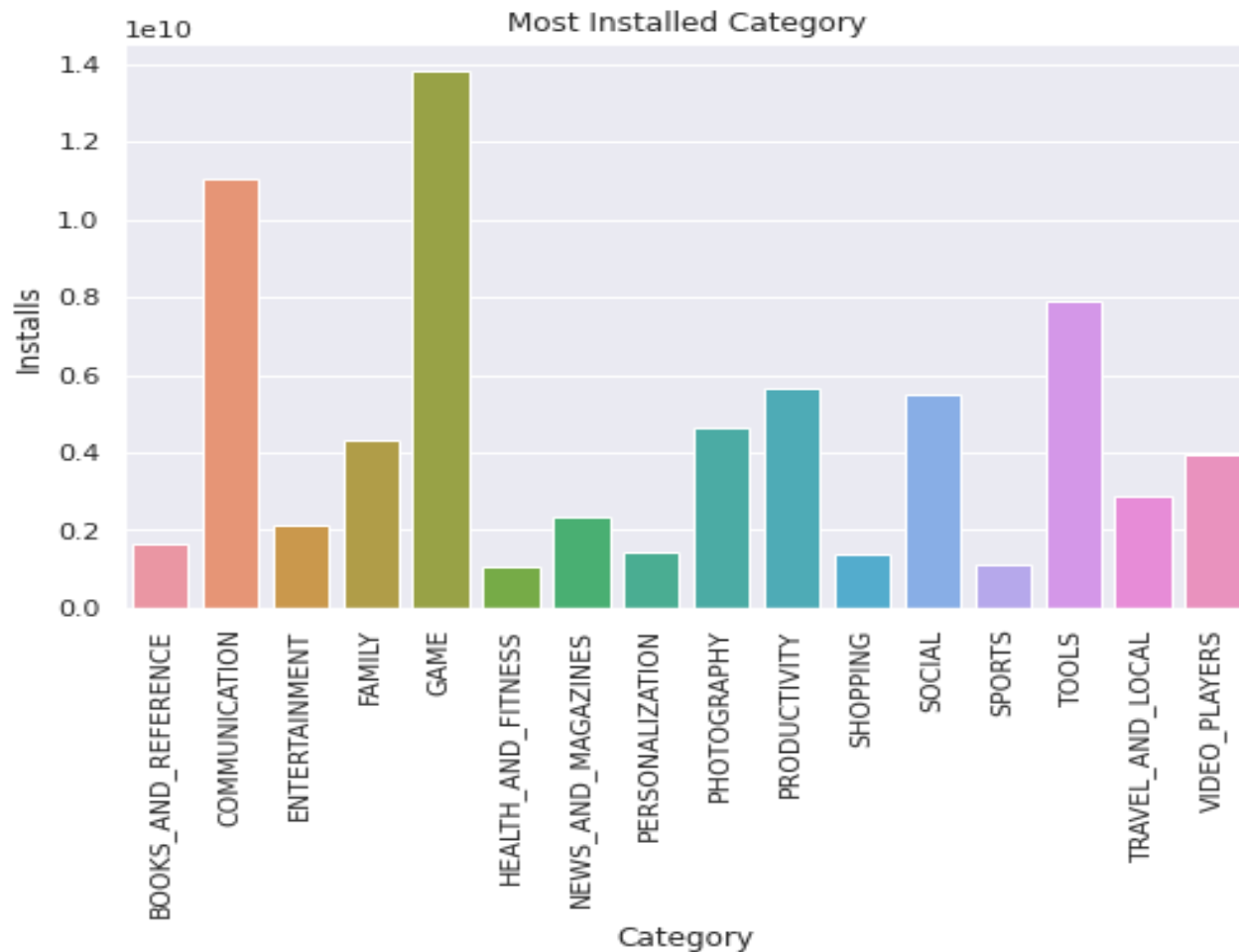
Best App



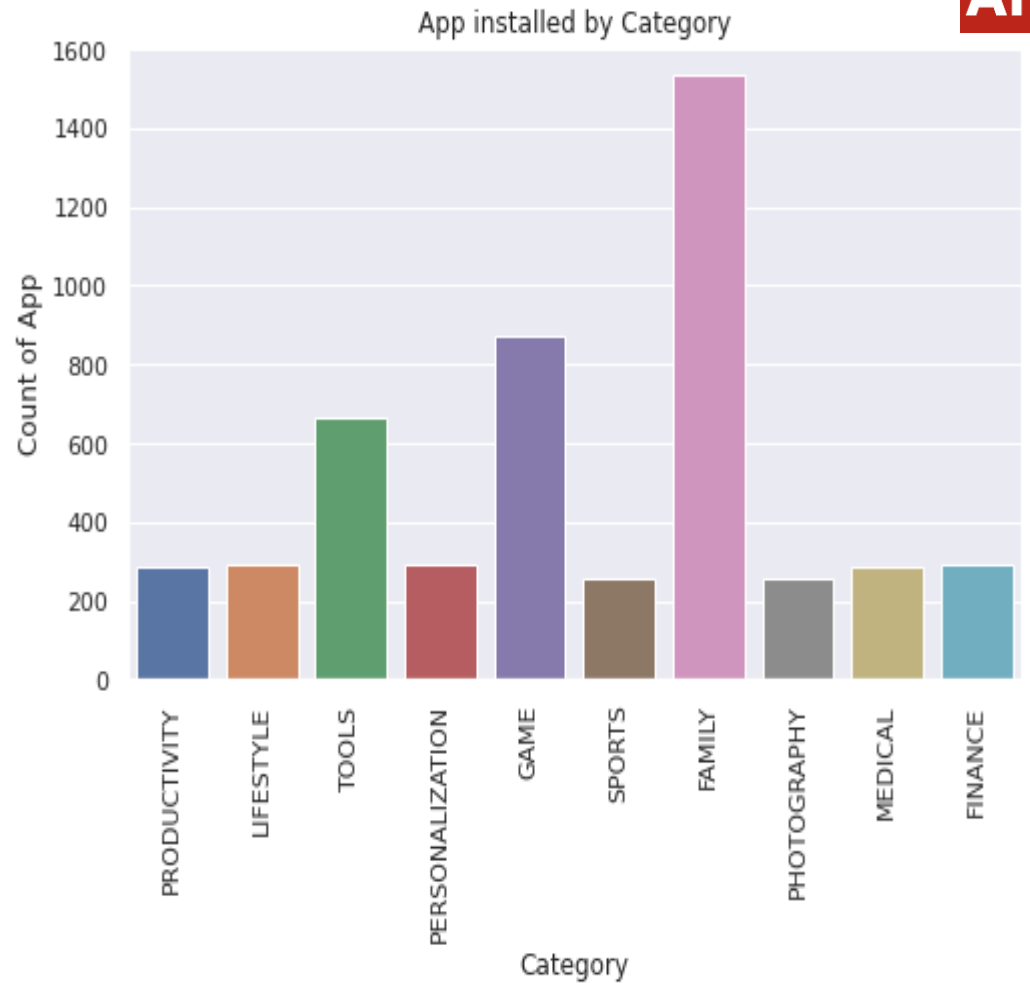
Category vs Genres vs Sub Genres



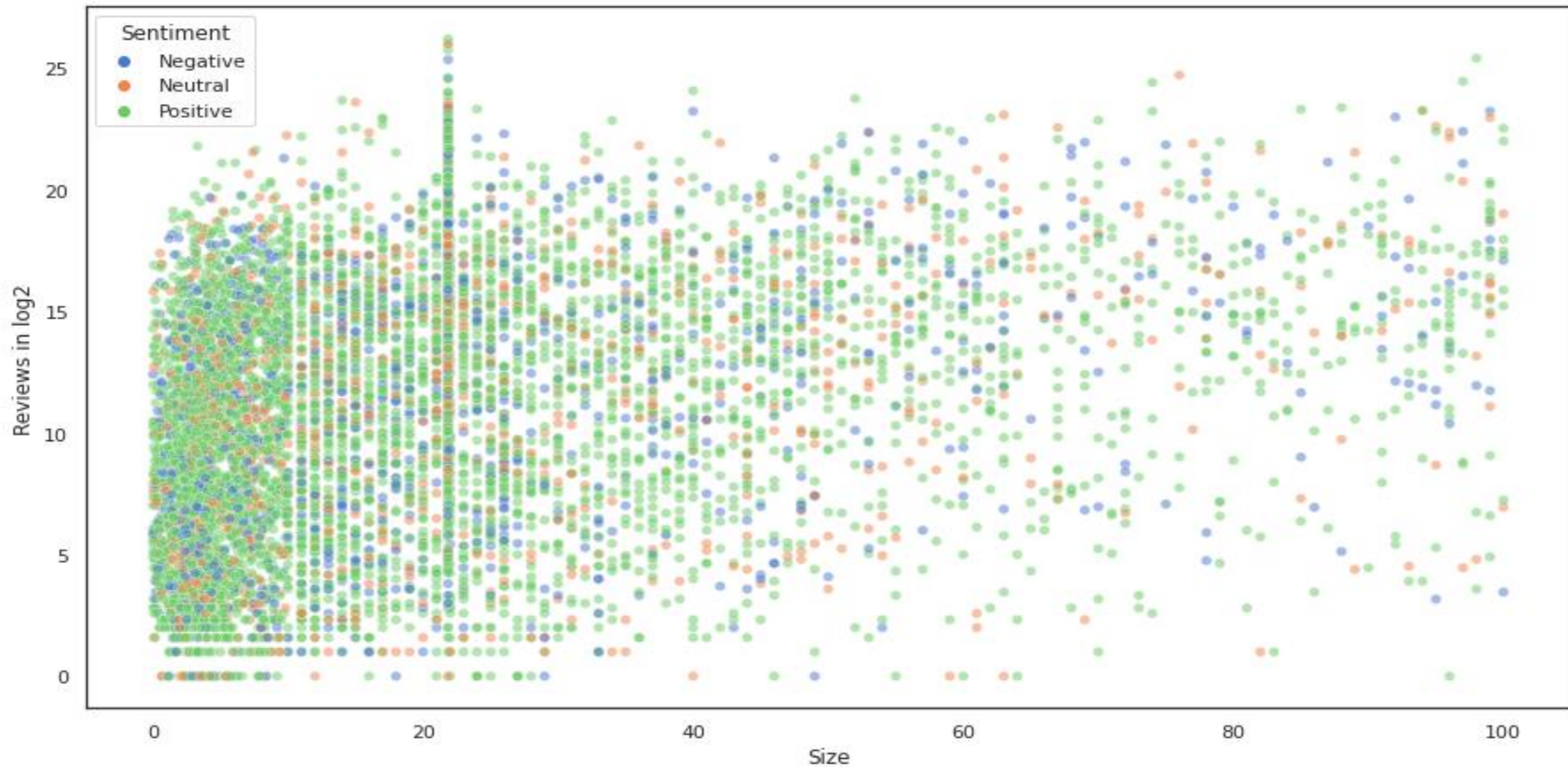
The most popular category is Games Followed by communication and Tools.



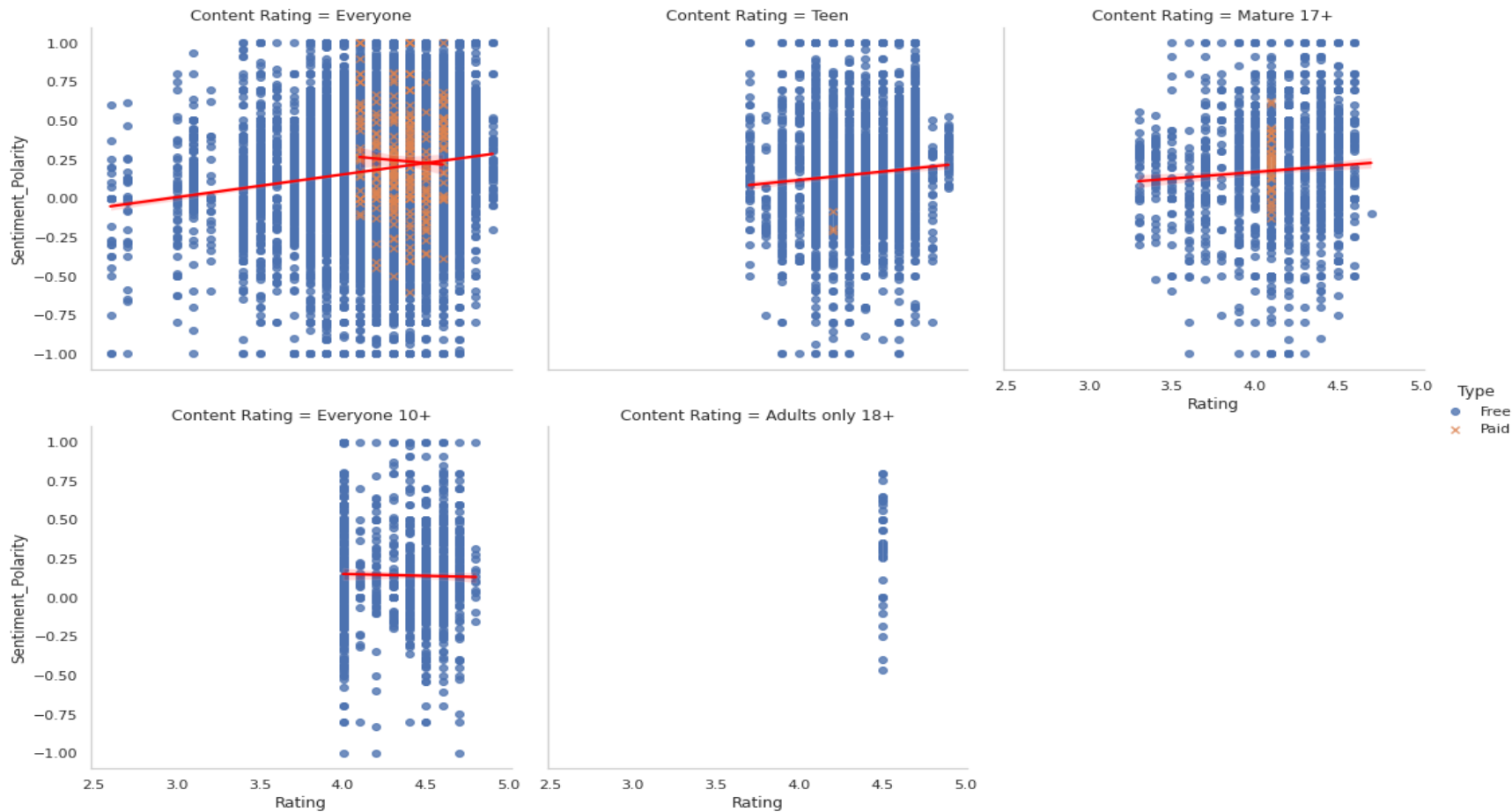
Family Category has most installed in the device



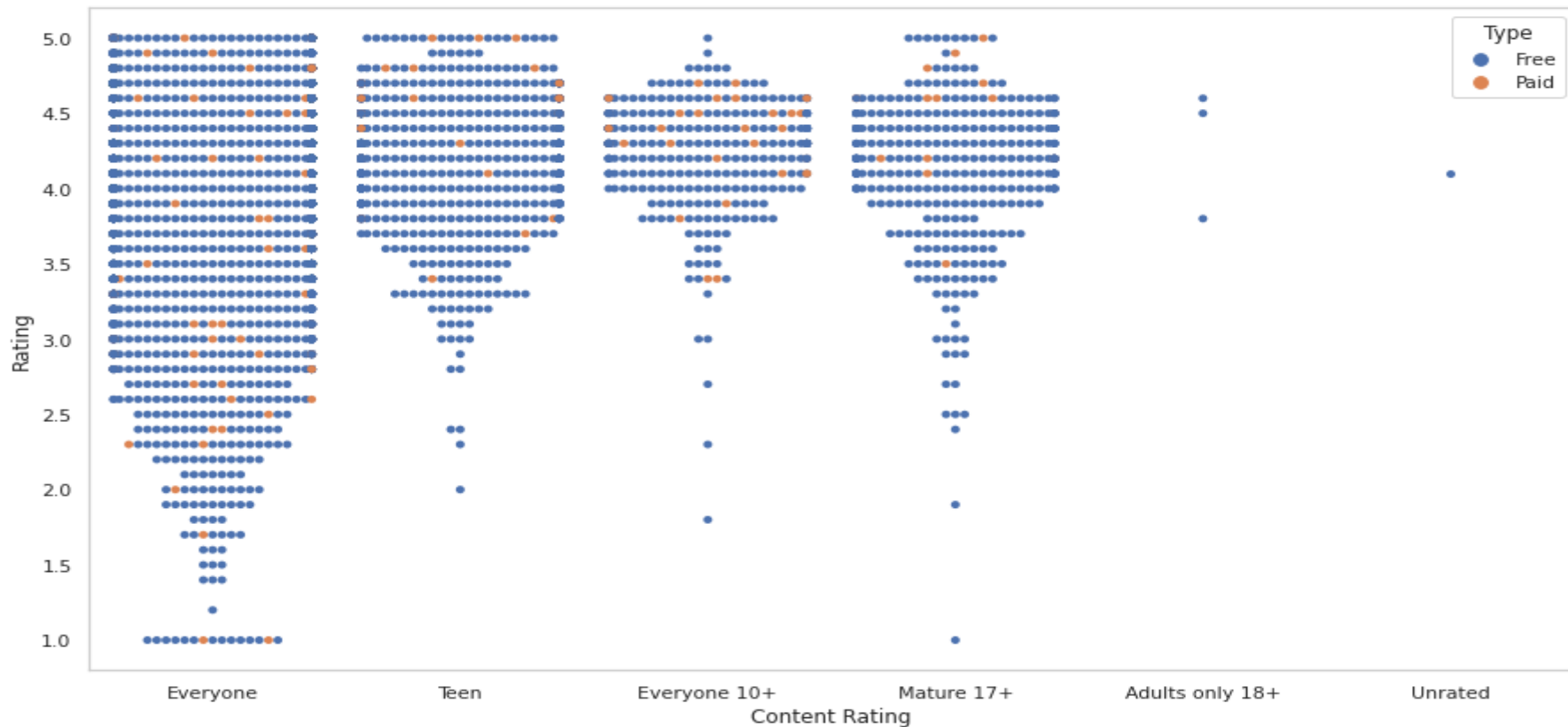
Size vs Reviews



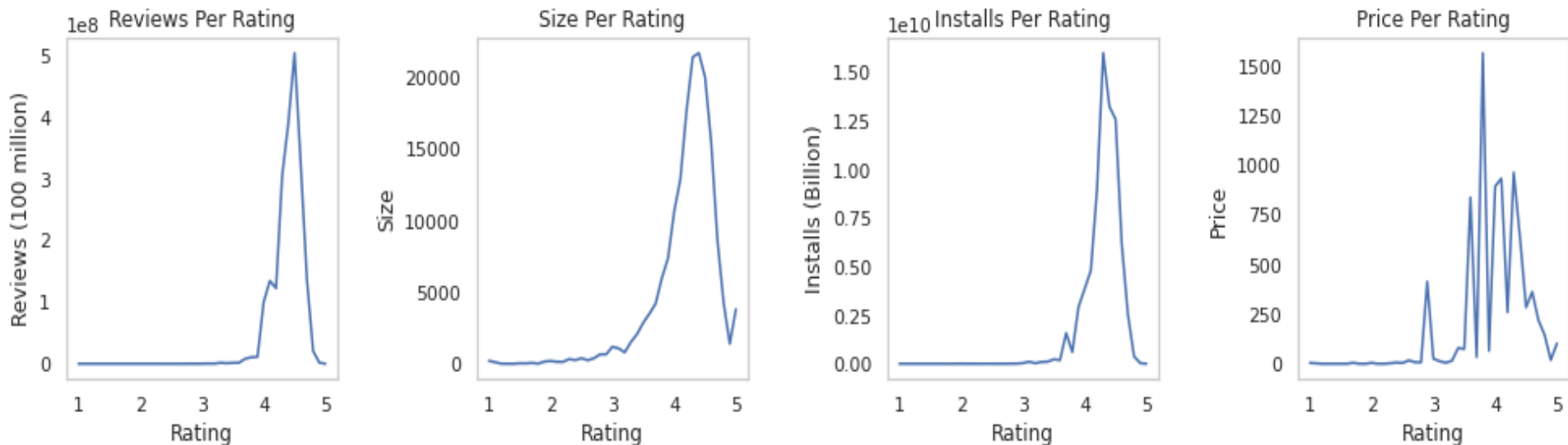
Facet Grid of Content Audience Type



Swarm plot between Rating and Content Rating

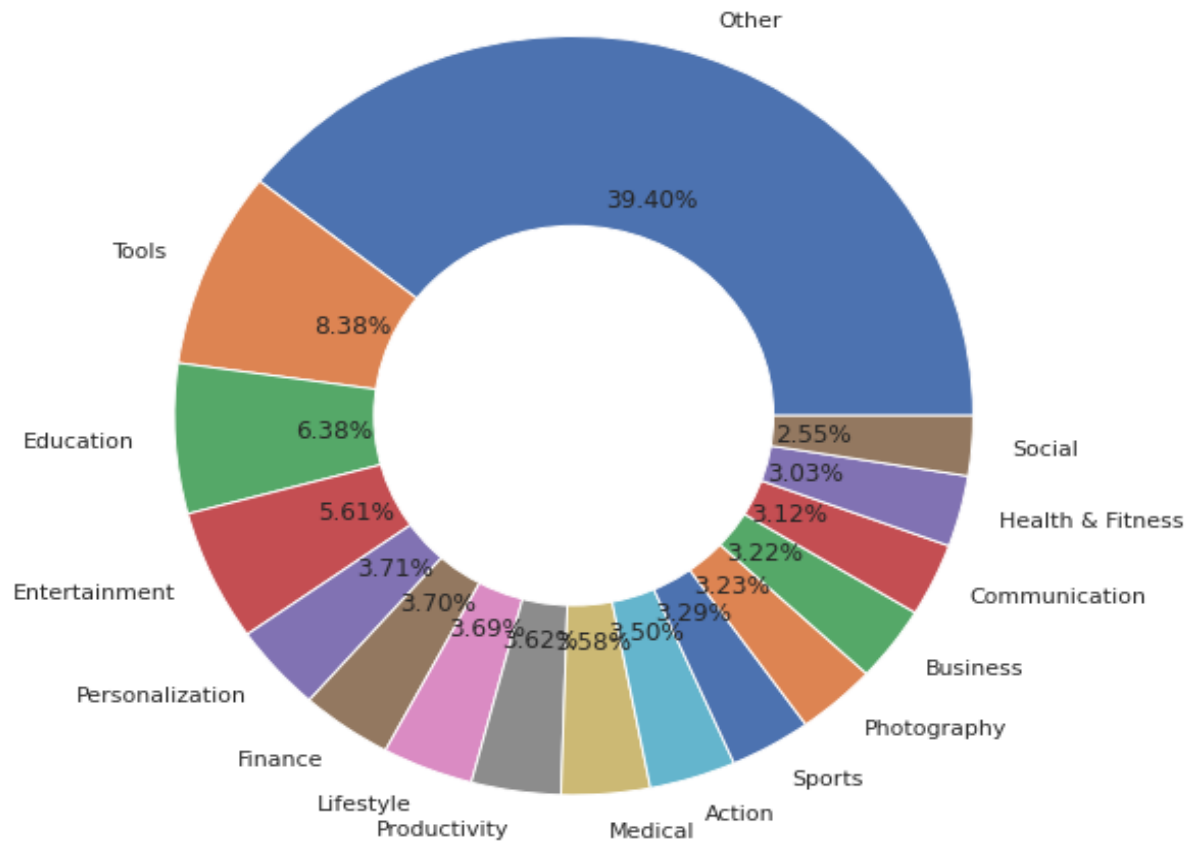


Decoding the Rating

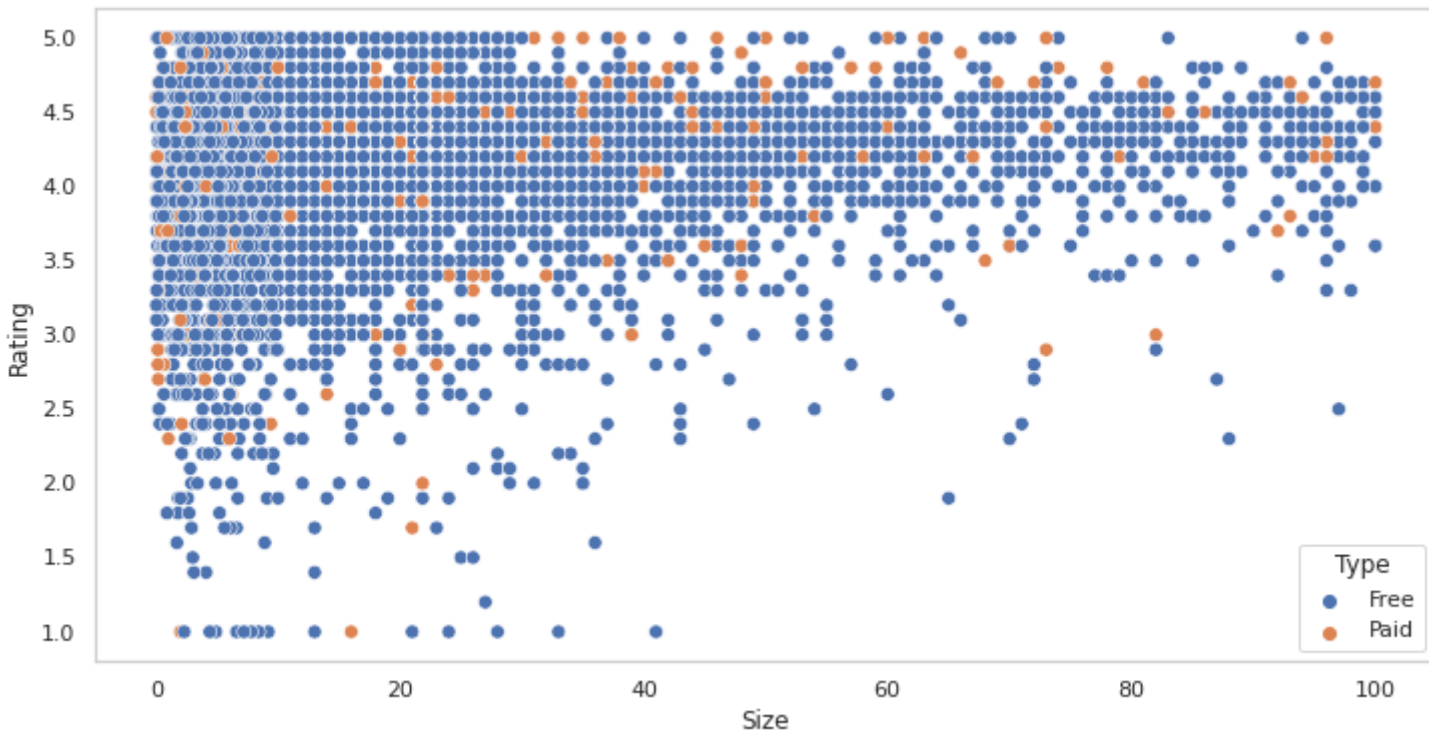


Apps having a rating in the range of 4 to 5 has the most reviews, are more bulk in size, are installed more and priced is high

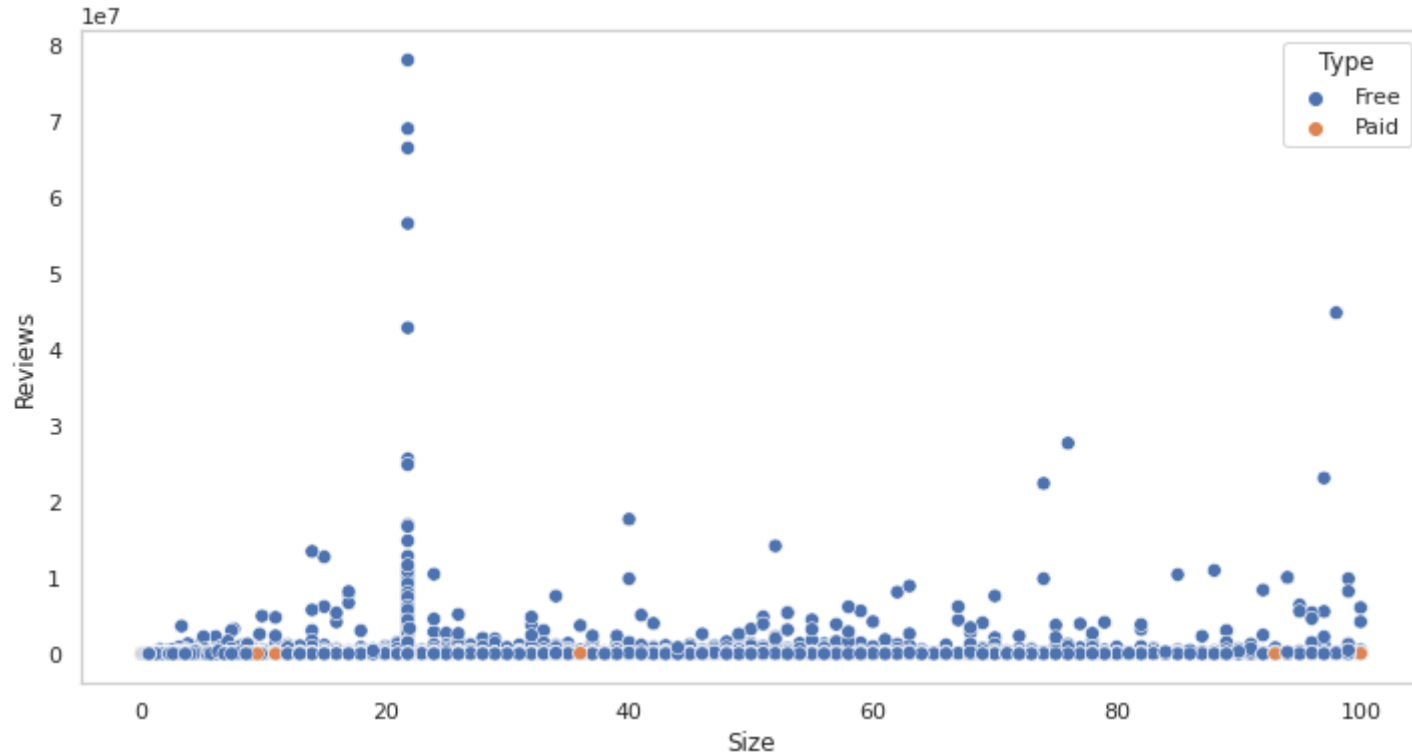
Percentage of apps share in Every Sub Genres



Audience Sub-target relationship with sentiment polarity and size

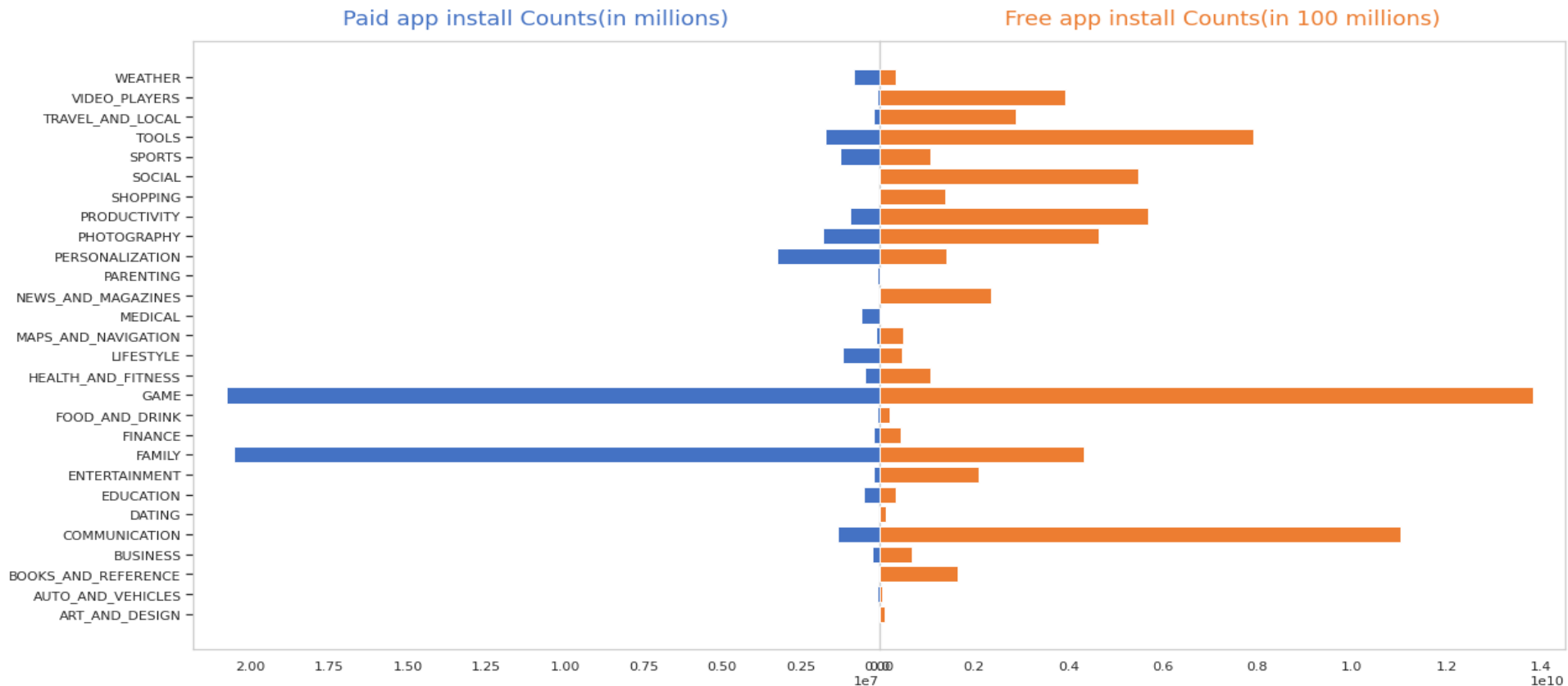


Review by Size and Type



- Reviews and sizes have very less correlation
- Paid apps get very less review
- Apps that are close to 20 Mb get more reviews.
- App in bulk size have more reviews

Visualization of the Category based on Type with the Most Download



Sentiment Analysis

Positive Sentiments



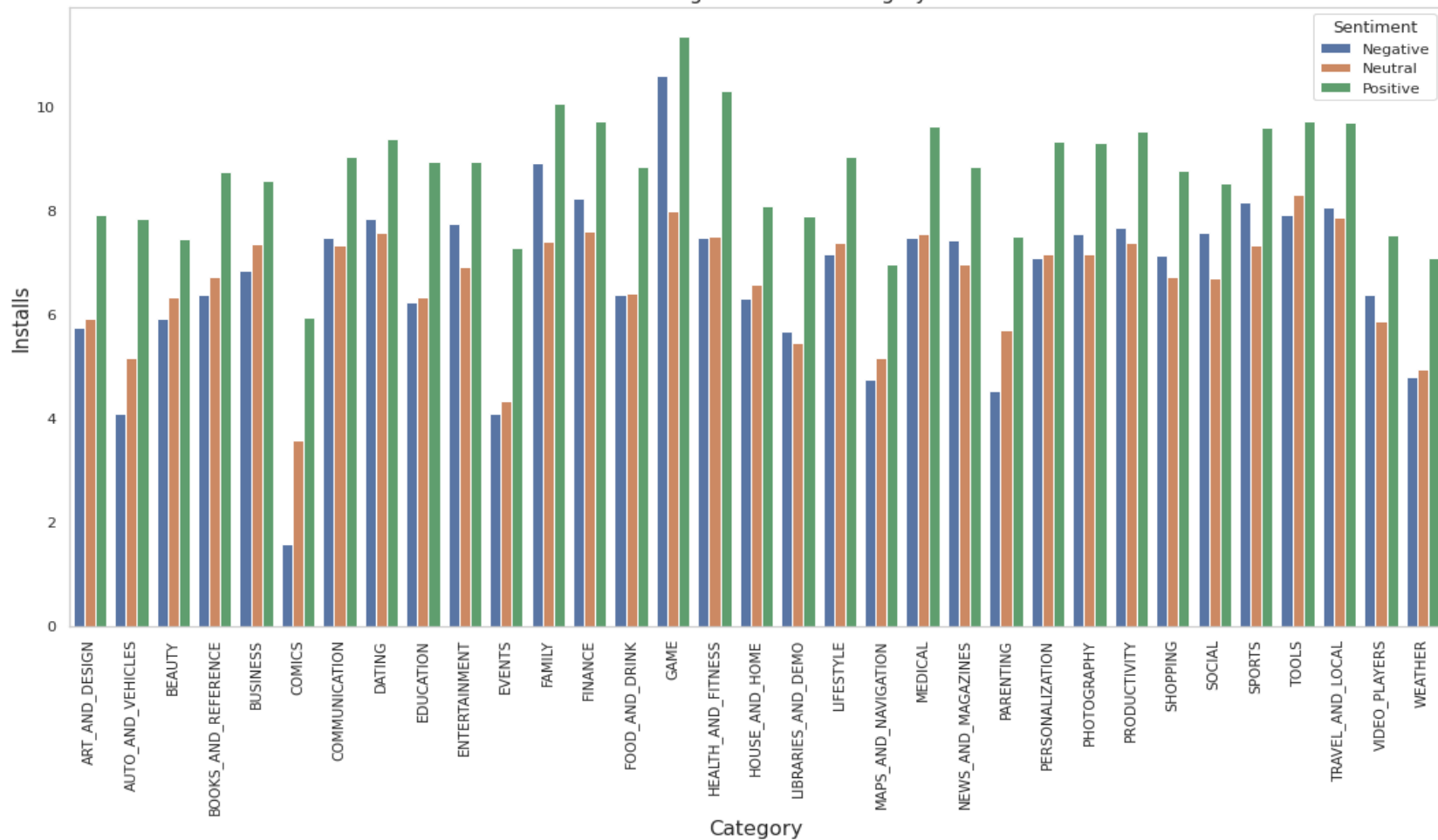
Negative Sentiments

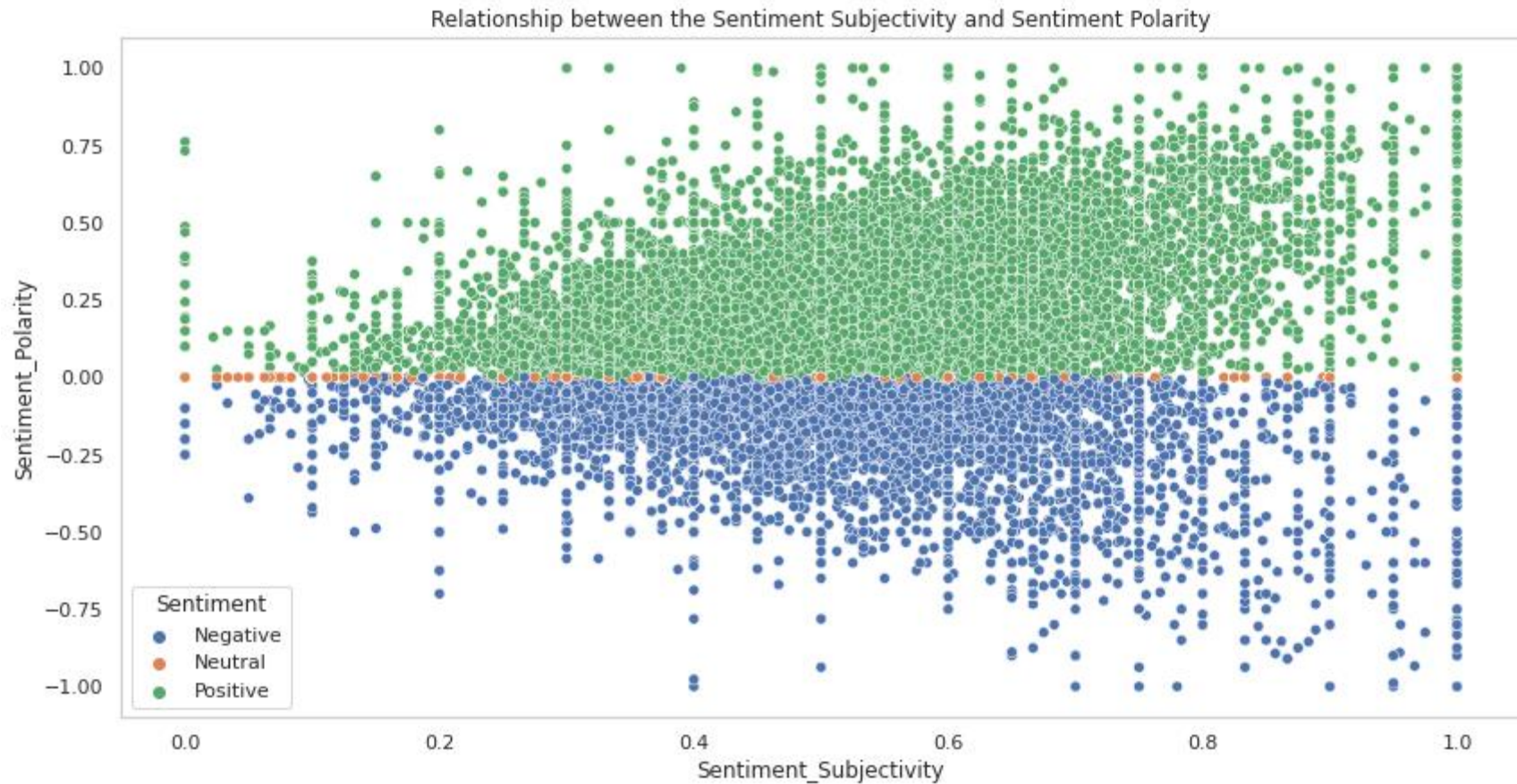


Neutral Sentiments



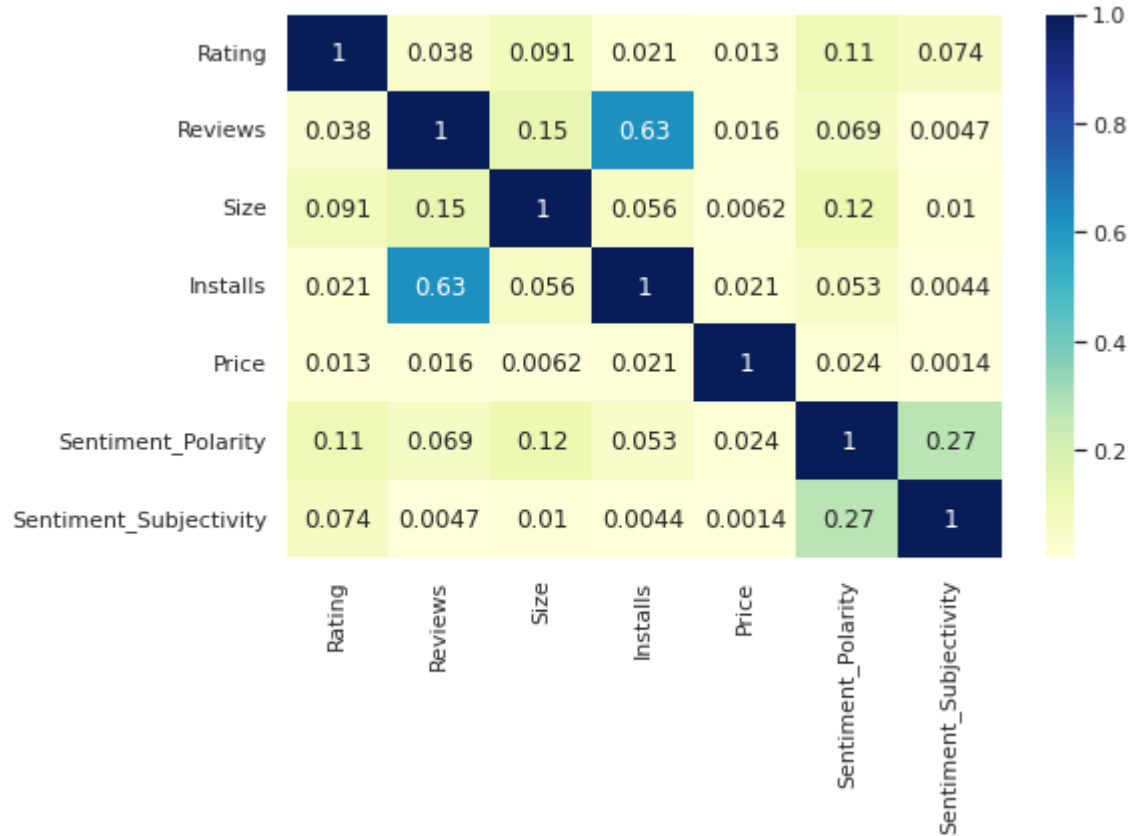
Number of install According to Different Category and Sentiments





Correlation Heatmap

- All the variables seem to be loosely correlated with each other
- Sentiment Polarity & Sentiment Subjectivity are slightly correlated with each other i.e. 0.27



Challenges Faced

- Presence of huge noises and anomaly
- Subset present in the attributes makes our analysis very hard
- Huge presence of categorical data
-

Solution to Business Solution...

- Should develop the Gaming app.
- Should stay away from the Comic category
- Should be available to everyone instead of deploying age restriction

Continue.....

Conclusion

- There are 20 app which has been downloaded more than a billion times has an average rating of 4.17
- Three best apps are Instagram, Subway surfers and Google Photos
- Family category has the highest number of apps.
- When more content restrictions are applied installation keeps on decreasing but the rating improves.

Conclusion

- Family app Category is the most downloaded paid app
- Game that is most installed has the most positive sentiments and also the most negative sentiments.
- Tool is a genres types app used mostly
- Tools has the most neutral sentiments.
- Comics have the least number of positive, negative and neutral sentiments

Continue.....

Conclusion

- Paid app gets very less reviews but has the high rating
- App that is close to 20 Mb gets more reviews.
- App in bulk size has more reviews
- The better rating apps are downloaded more and does not matter if it's costly or heavy in size.

**Thank you For Staying with me
with Patience**