Exploratory Data Analysis on Playstore Dataset

Outline

Project Summary

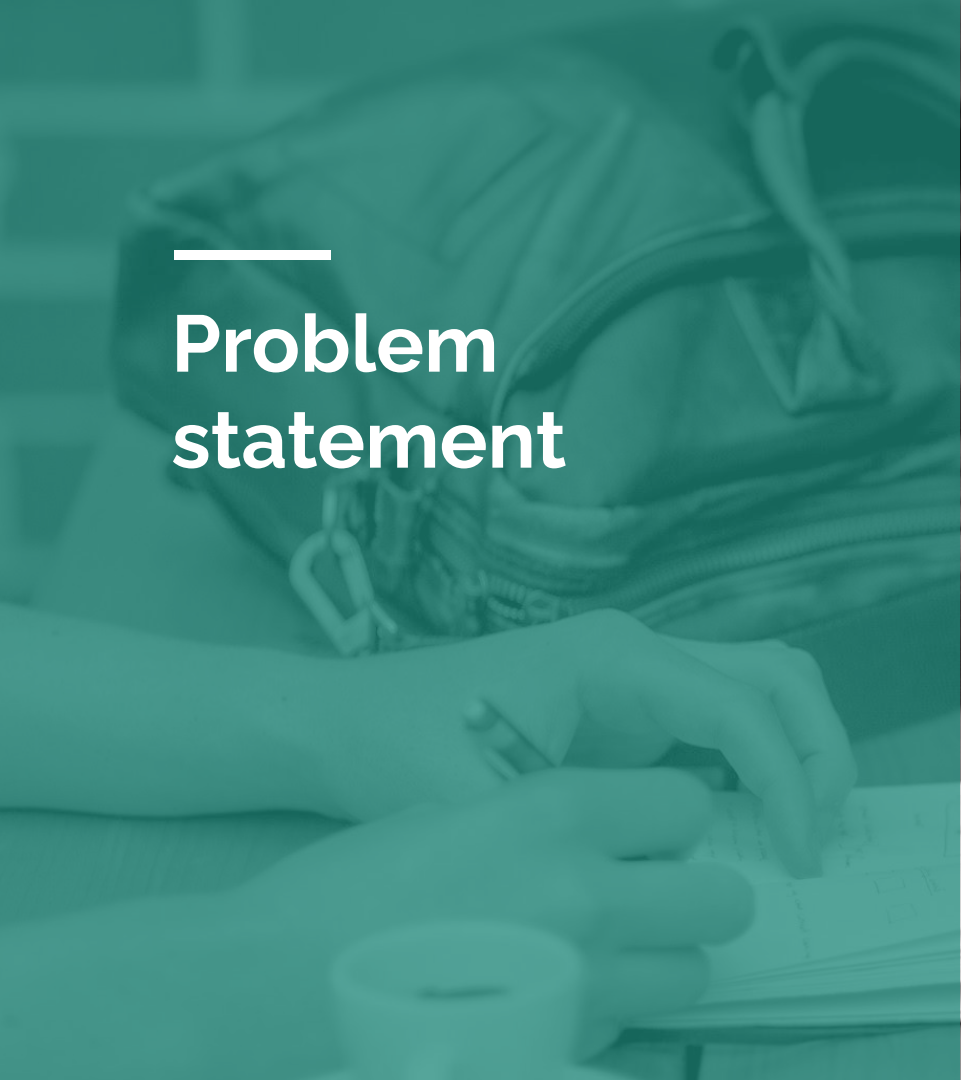
- We are working with a Play Store app review dataset. First, let's understand what the Play Store is. Basically, the Play Store is a type of application marketplace (according to Wikipedia). When we visit any market (such as a vegetable/fruit market or a mall), we typically purchase items, search for the things we need, compare prices with other sellers, read reviews about those items, rate them, and also assess the demand for those items. Similarly, when we use a smartphone, we require different types of applications to fulfill various needs. According to Wikipedia, the Play Store functions as a marketplace. In the Play Store, we download applications, check their type (free or paid), explore their uses, look at the genres of applications, write reviews for applications, rate them, and assess the demand for these applications (in terms of installs).

Objective

The objective is to gain actionable insights from the Play Store dataset to optimize app performance and user satisfaction. This involves understanding various aspects of app categories, user sentiments, pricing strategies, and user engagement.

More for Business objective we can analyze for below points from dataset,

- **Type of application Demand.**
- **Number of applications.**
- **Applications prefered category by users.**
- **Applications prefered Genres by users.**
- **Sentiment Analysis for ratings, category.**
- **Distribution of application over ratings and reviews.**



Problem statement

1. Find the number of unique applications and application categories available in the dataset.
2. Determine the number of installs for each application category and identify the category with the maximum number of installs. Visualize this using a pie chart.
3. Calculate and display the average rating of applications in each category.
4. Count the number of free and paid applications. Identify applications with the maximum and minimum prices.
5. Categorize applications based on content ratings and create a pie chart to show the distribution of applications in each rating category.
6. Categorize applications based on genres and create a pie chart to show the distribution of applications in each genre.
7. Identify the latest and oldest applications in the dataset based on their last update dates.
8. Determine which Android version most applications are compatible with.
9. Analyze how important the rating of an application is by exploring its relationship with other factors.

And More.....

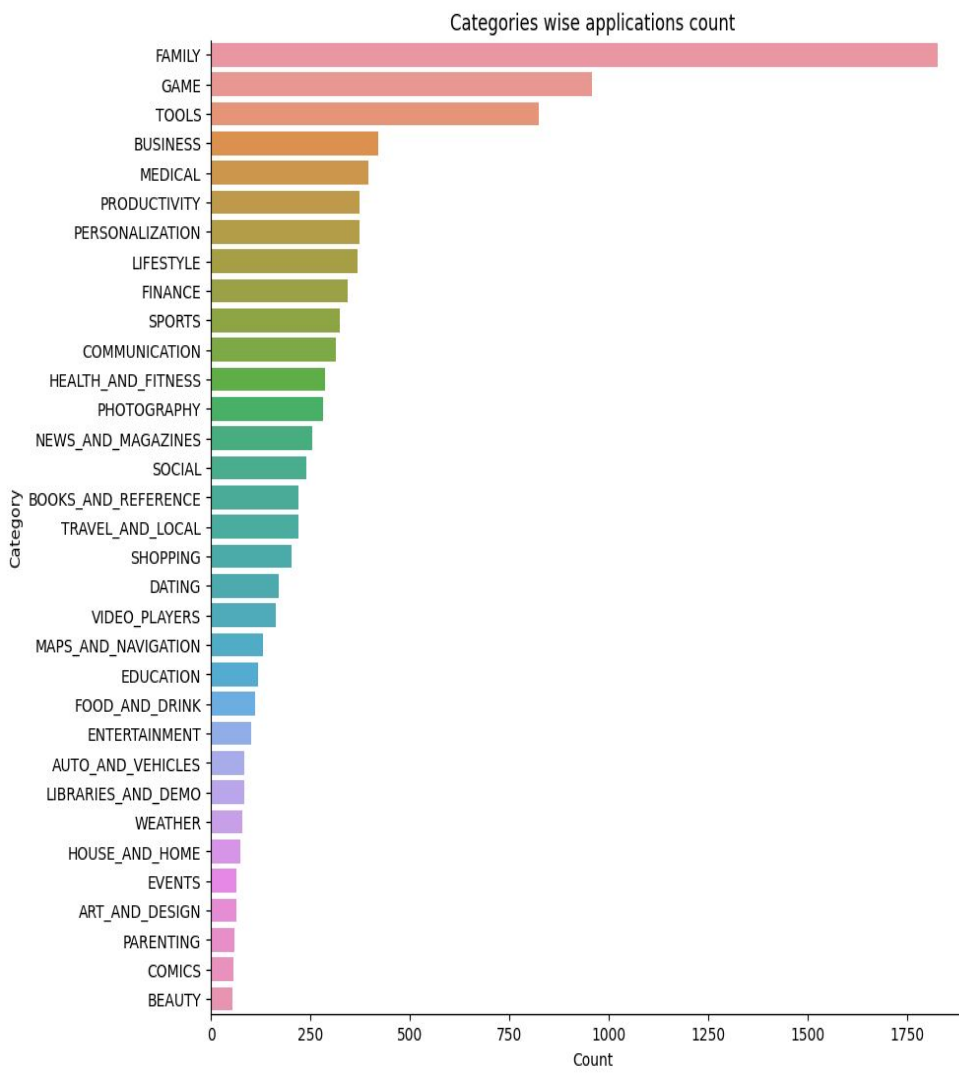
Categories and their application Count

Most Popular Categories: Family, Games, tools

App Diversity : Few Categories with large number of application and few with less numbers.

Opportunities : For Which categories having less number of application its having opportunities for development.

Market Fall : Can check market is falling towards the categories like Family type, Games, tools and Business while there is less market fall towards categories like beauty, comics,parenting and Art_and_design



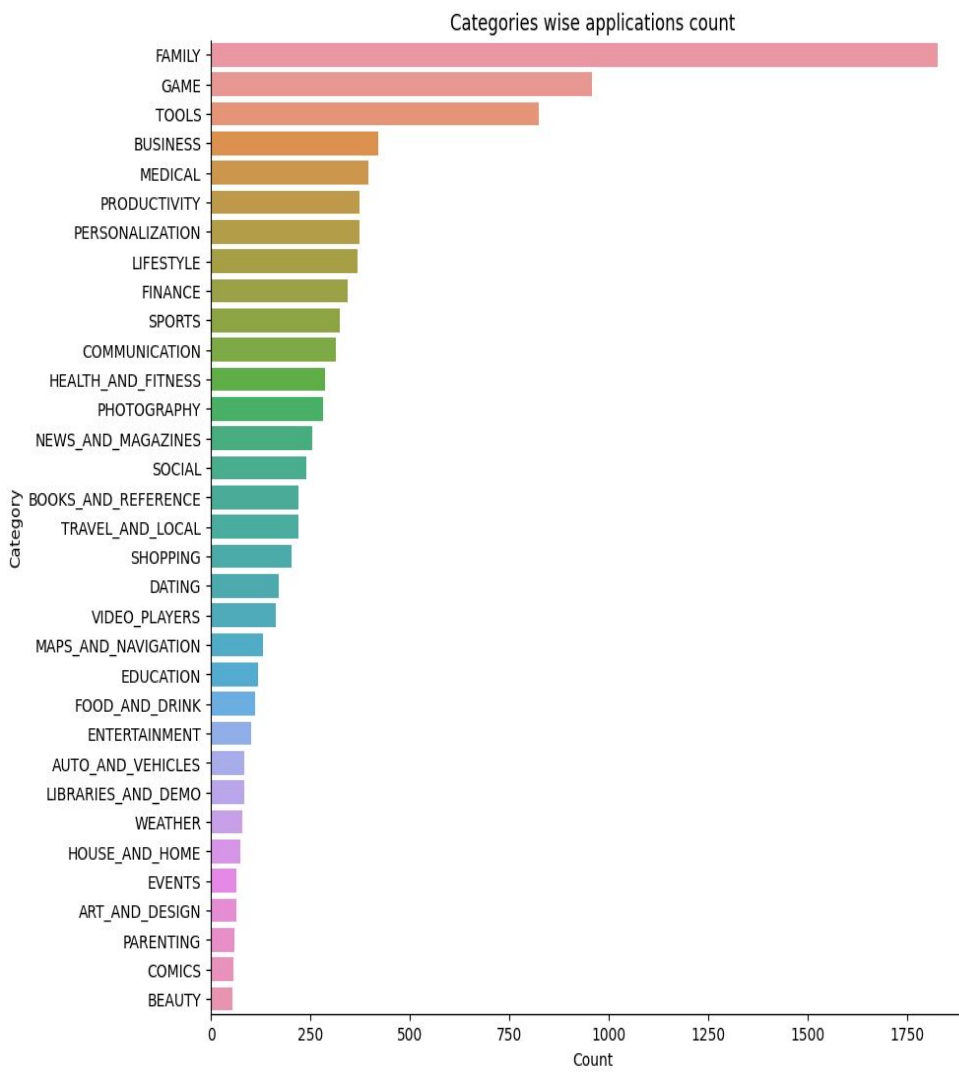
Categories and their application Count

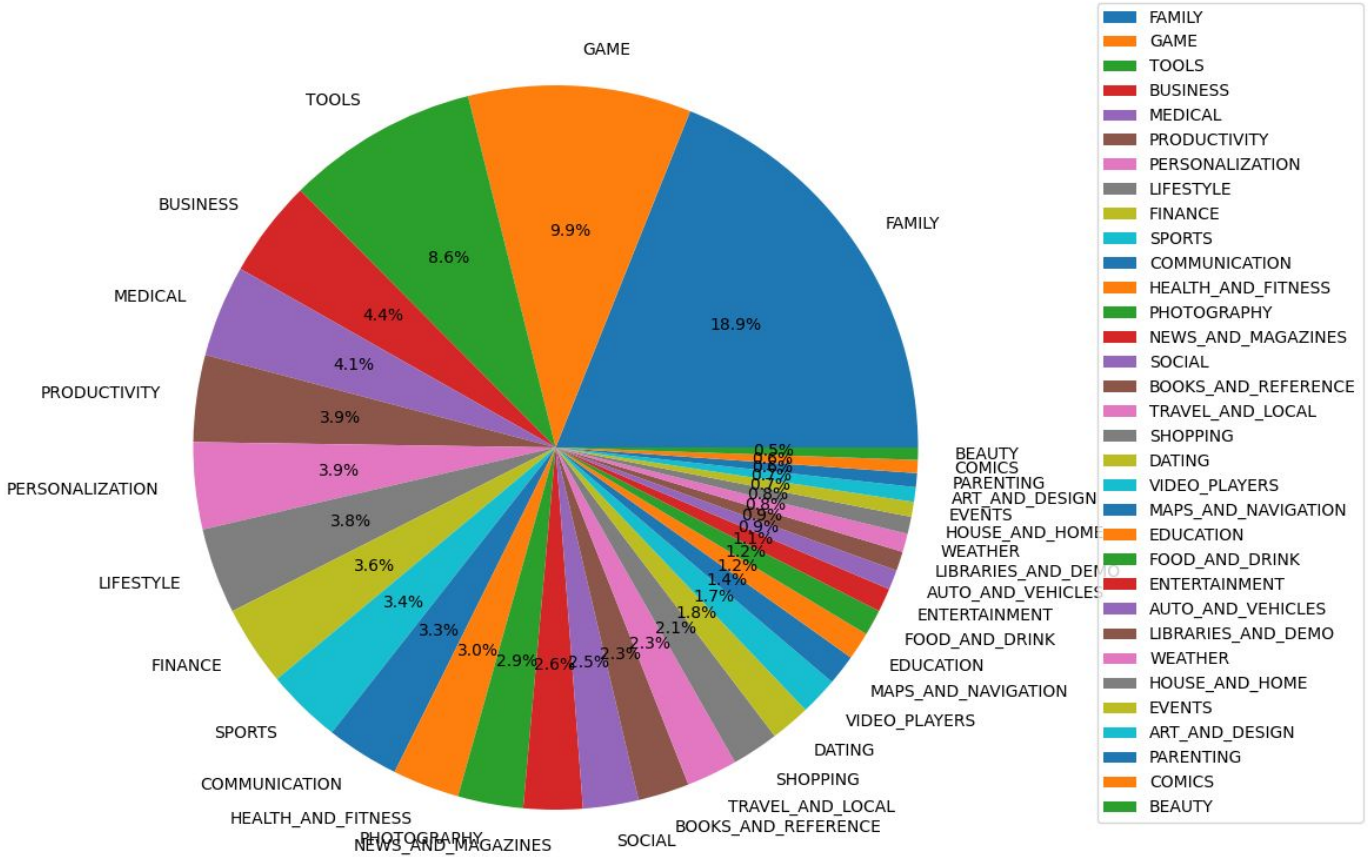
Most Popular Categories: Family, Games, tools

App Diversity : Few Categories with large number of application and few with less numbers.

Opportunities : For Which categories having less number of application its having opportunities for development.

Market Fall : Can check market is falling towards the categories like Family type, Games, tools and Business while there is less market fall towards categories like beauty, comics,parenting and Art_and_design





Categories and their application Count

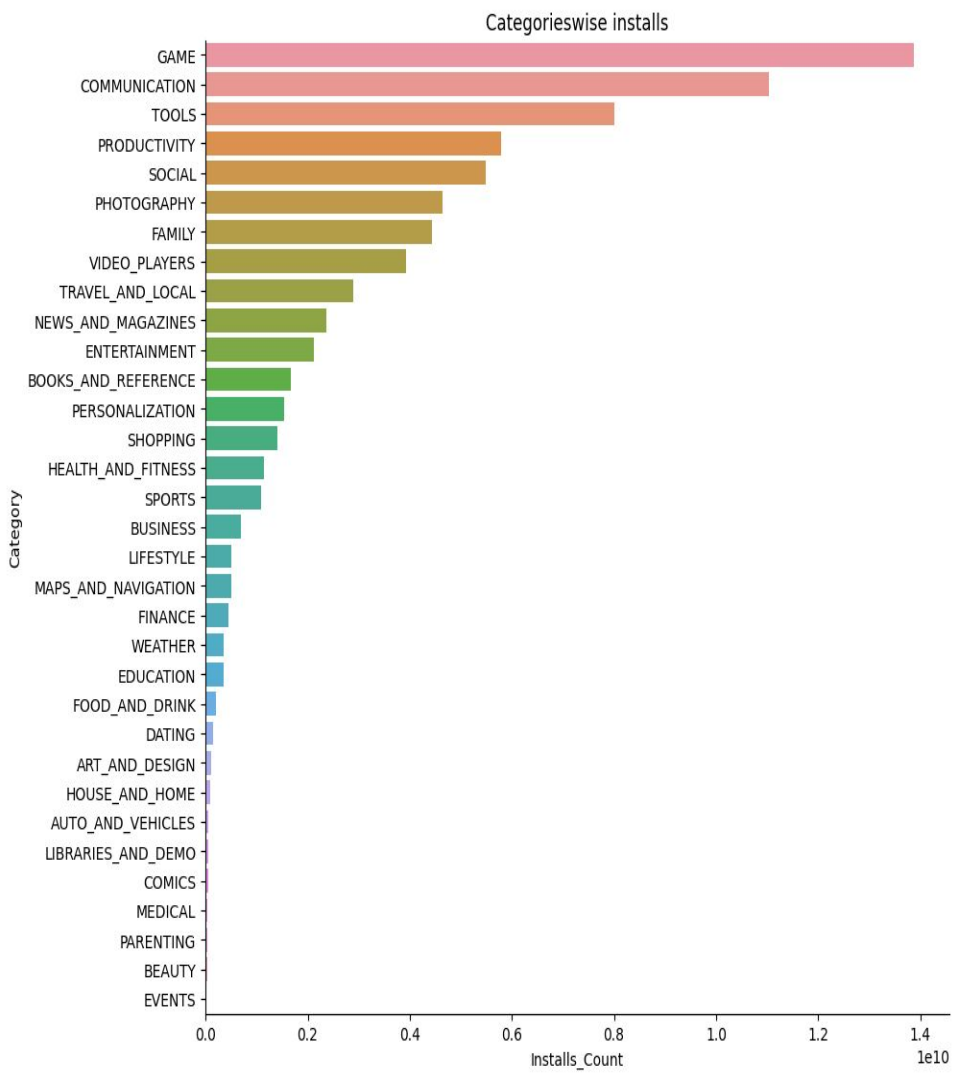
Categories and their application Count

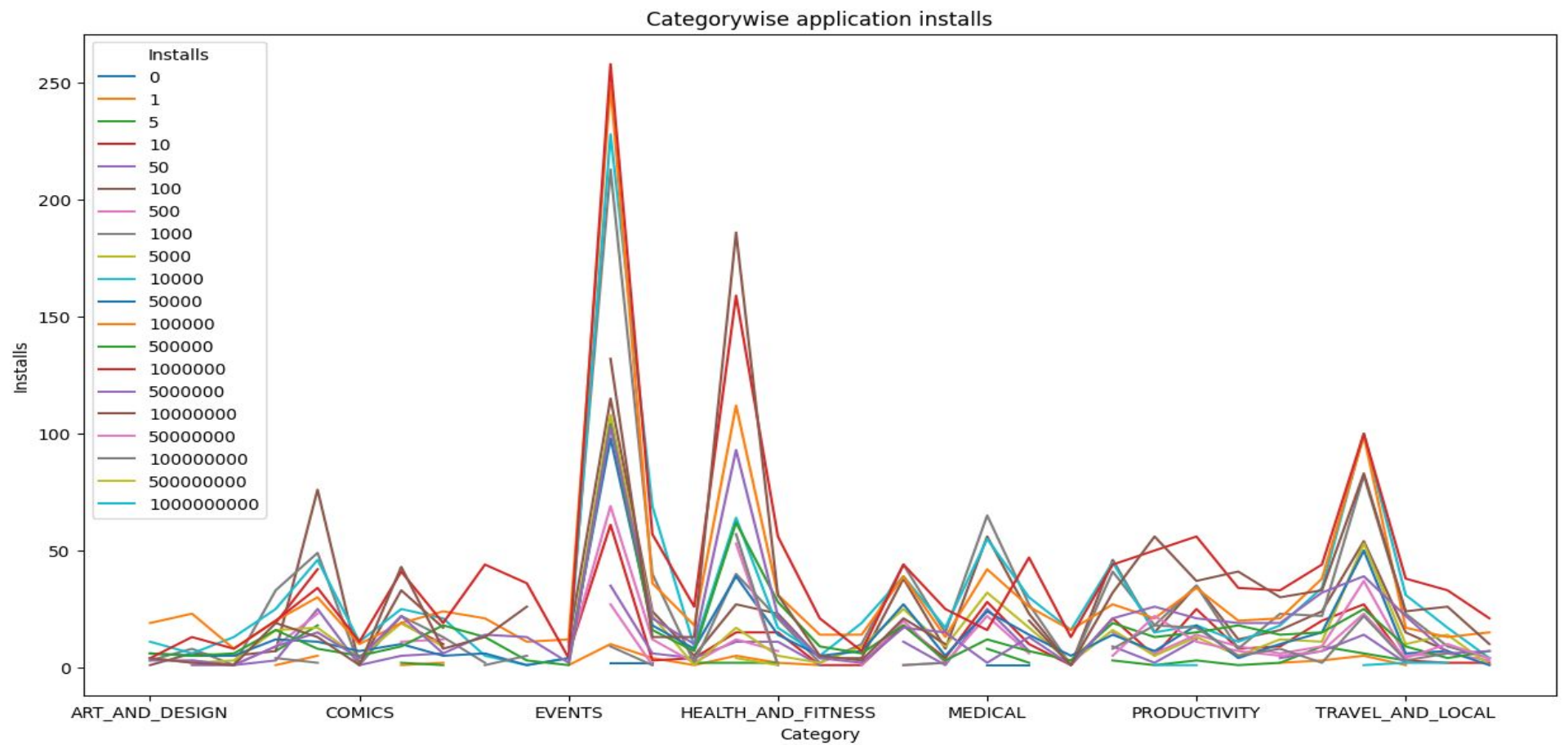
Most installed Categories: Game, Communication, Tools.

App Diversity : Few Categories with large number of installs and few with less numbers.

Opportunities : For Which categories having less number of installs its having opportunities for development.

Market Fall : Can check market is falling towards the categories like Family Game, Communication, Tools while there is less market fall towards categories like Event, Beauty ,parenting and Medical.



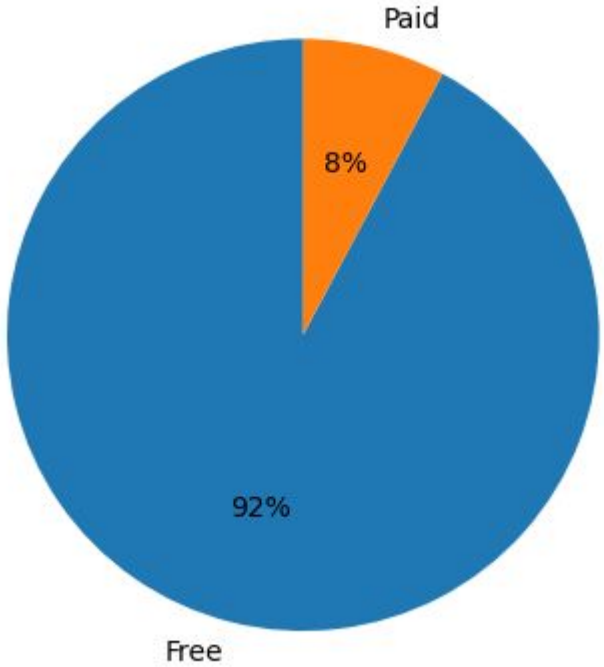


Categories and their application Count

Type of app and its count

Proportion of Free and Paid Apps: Free apps 92% and Paid Apps 8%.

Comparison of App Types:"Free" segment is much larger than the "Paid" segment, it indicates that there are more free apps available in the dataset.



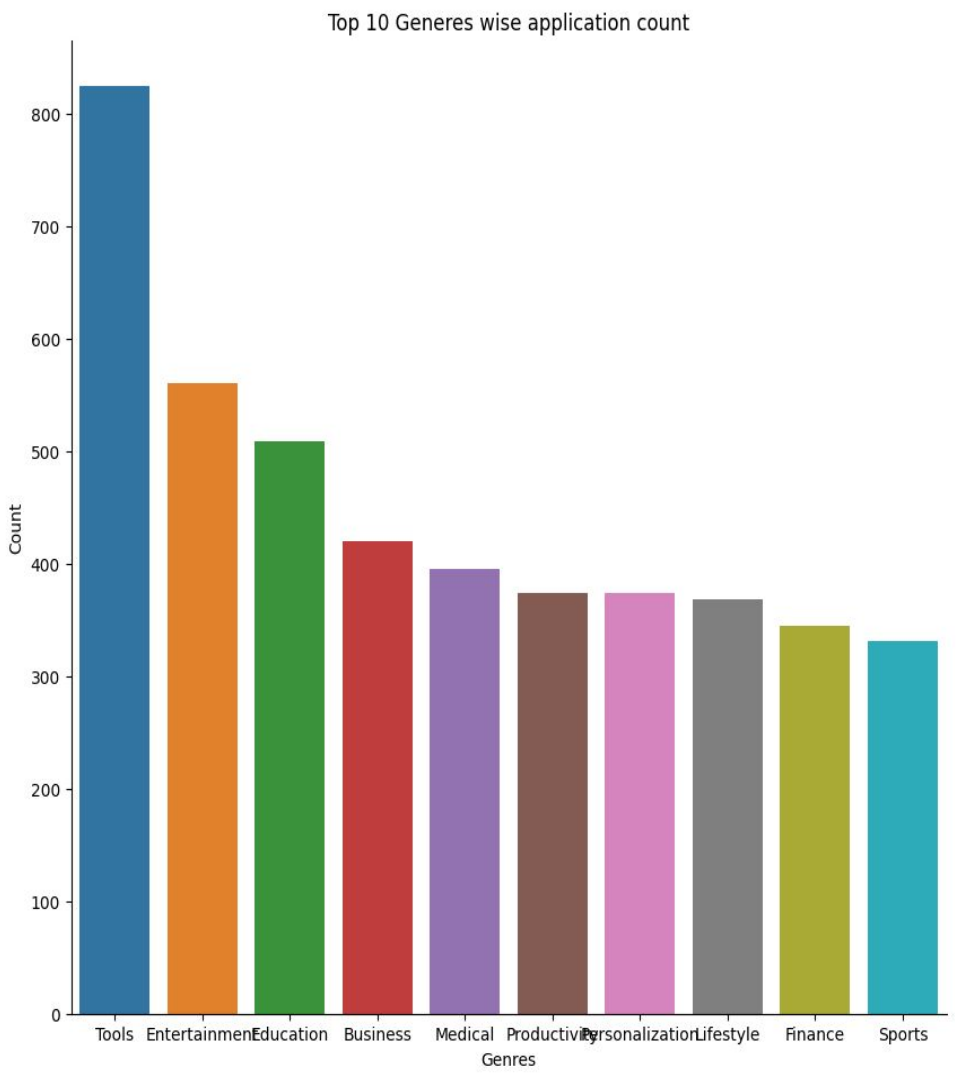
Top 10 Genres and application count

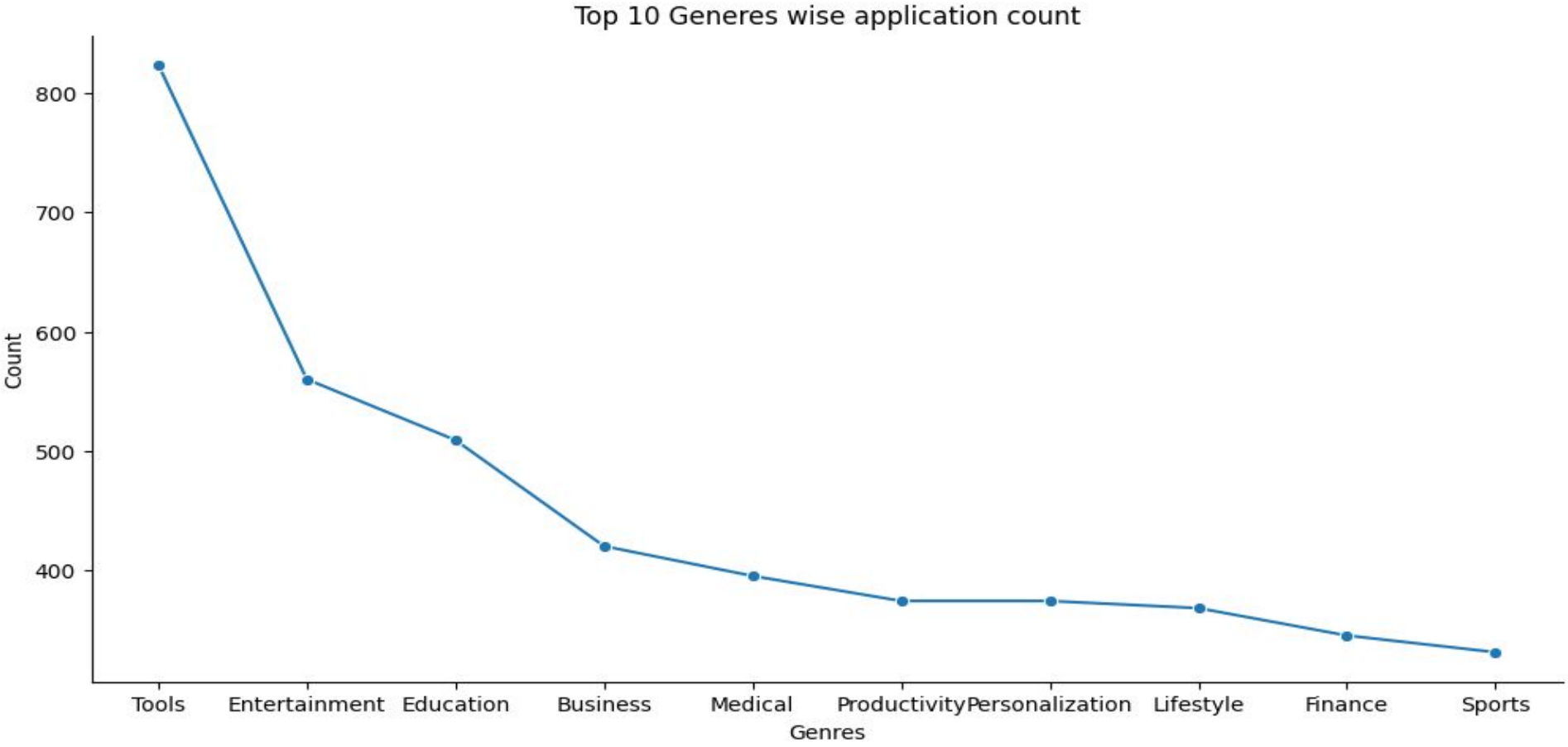
Most Popular Genre: Tools,Entertainment,Education.

App Diversity : Few Genre with large number of installs and few with less numbers.

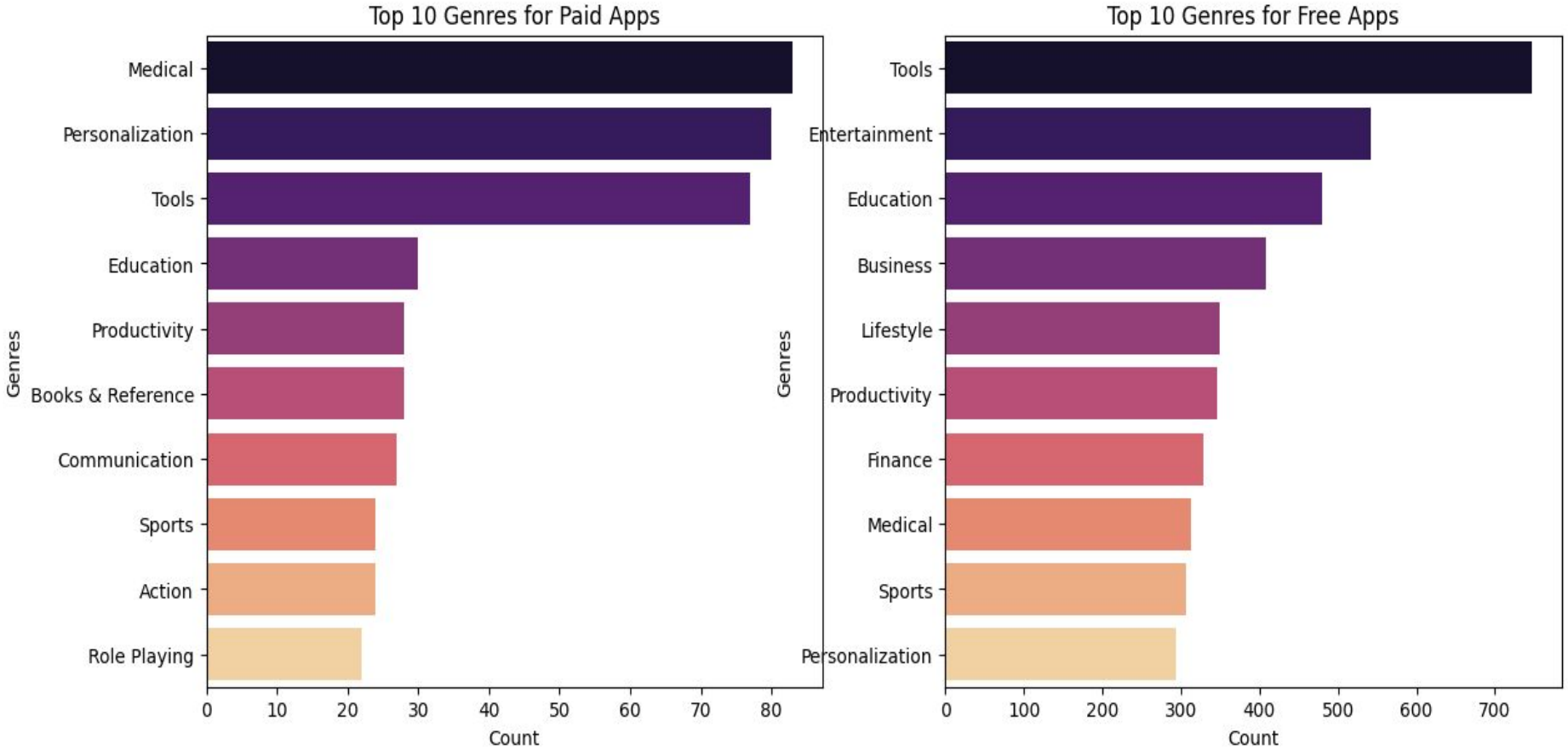
Opportunities : For Which Genre having less number of installs its having opportunities for development.

Market Fall : Can check market is falling towards the Genre like Tools,Entertainment,Education while there is less market fall towards categories like Sport, Finance and Lifestyle.





Top 10 Genres and application count



Top 10 Genres For Paid and Free app

Ratings Distribution Plot

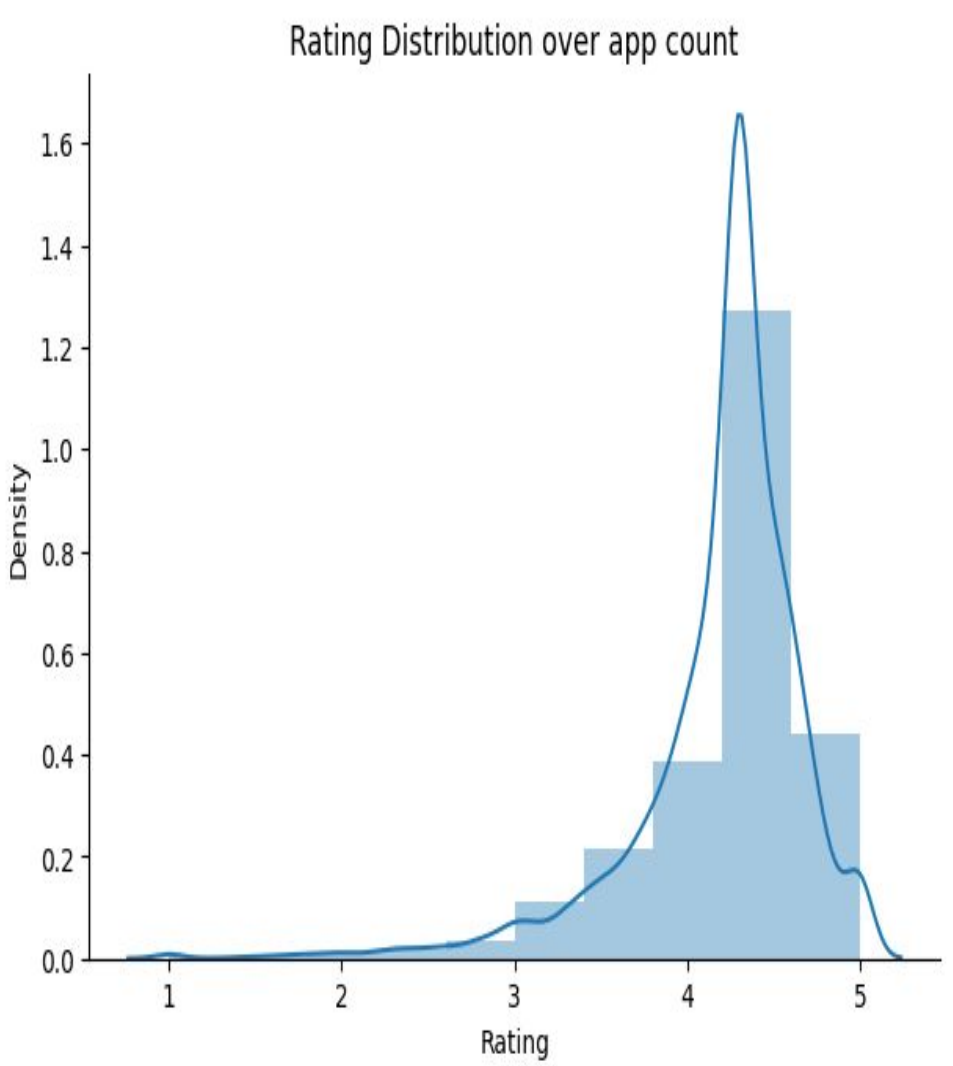
Top Rated Categories: Family, Game, Tools.

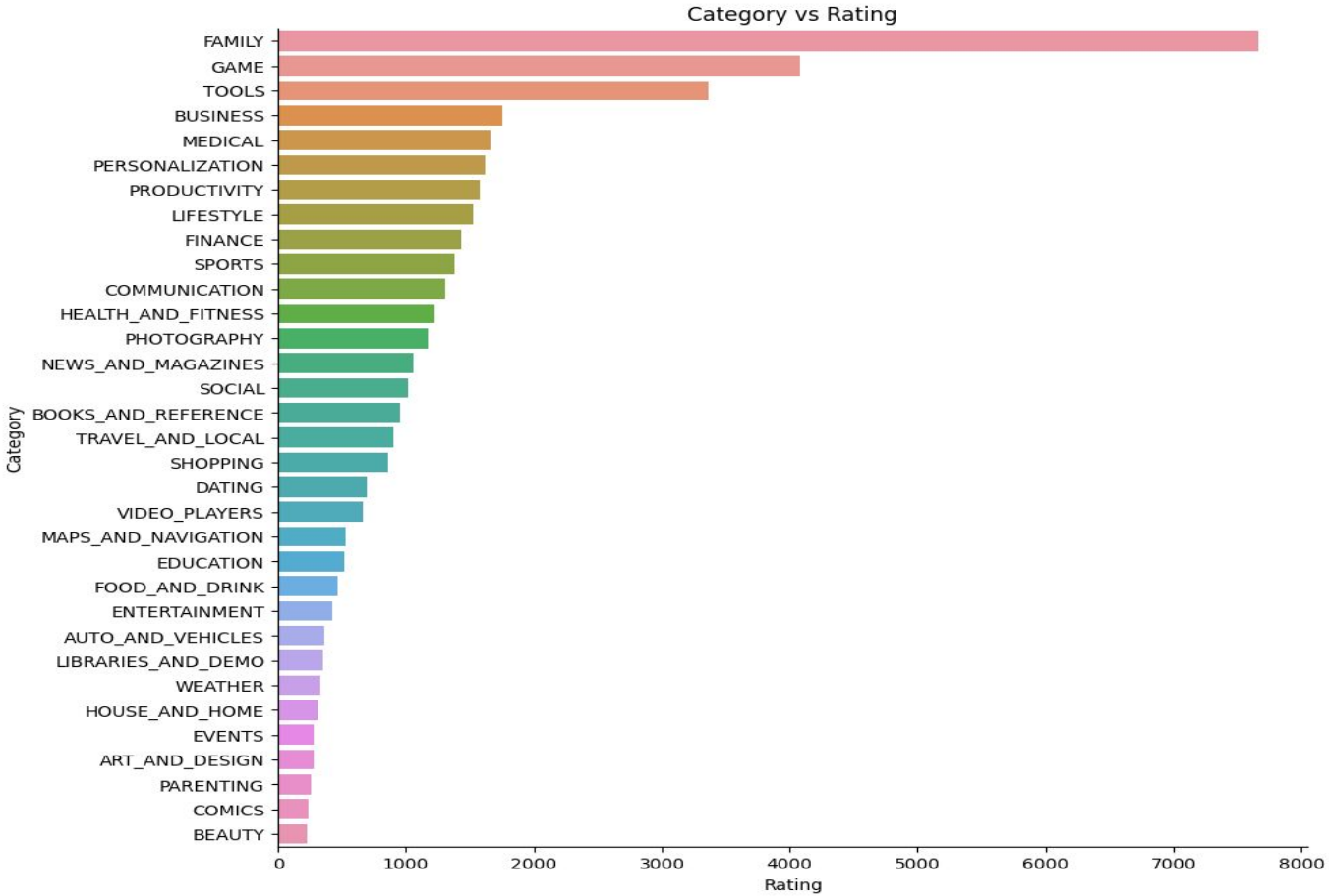
Bottom Rated Categories: Beauty, Comics,Parenting

App Diversity : Few categories with large number of rating and few with less numbers.

Opportunities : For Which categories having less number of rating it's having opportunities for development.

Market Fall : Can check market is falling towards the Genre Family, Game, Tools. while there is less market fall towards categories like Beauty, Comics,Parenting.



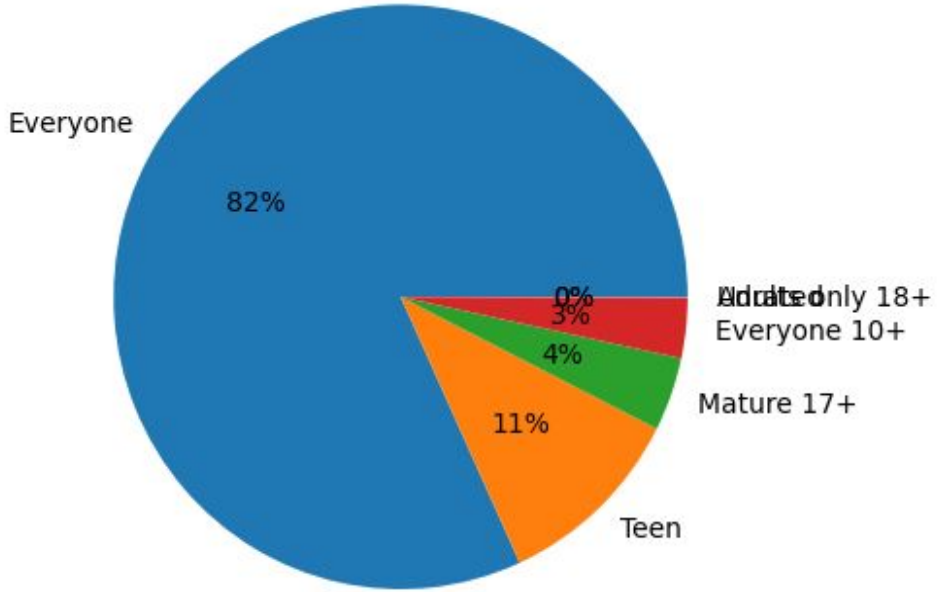


Category Vs Rating

Content Rating vs Application Count

There are 82% applications which are belong to 'Everyone' while 11%,4% belongs to 'teen' and 'Mature 17' Content Rating. There are very less applications for 'Adults only 18+'.

Opportunities : More Opportunities for Content rating for 'Adults only 18+','Everyone 10+'.

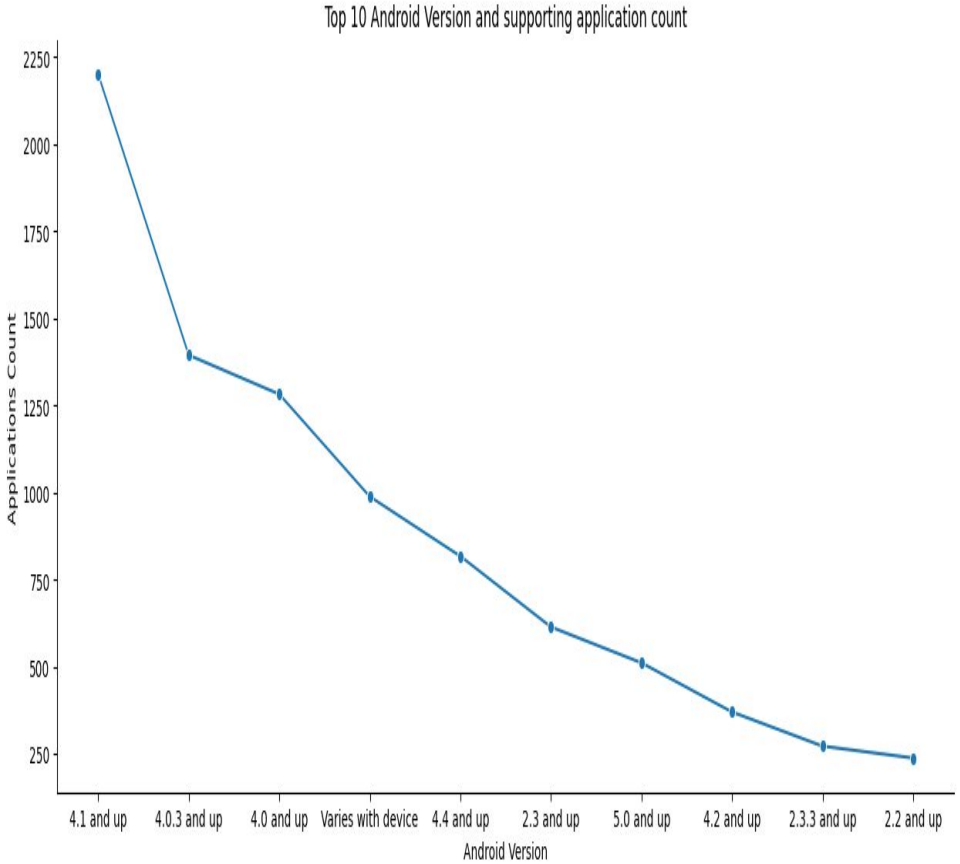


Count of application and supporting android Version

Most used Android Versions: '4.1 and up' and '4.0.3 and up'.

Most Unused Android Version : '2.2 and up' and '2.3.3 and up'

Most users have moved on to supporting more recent versions.



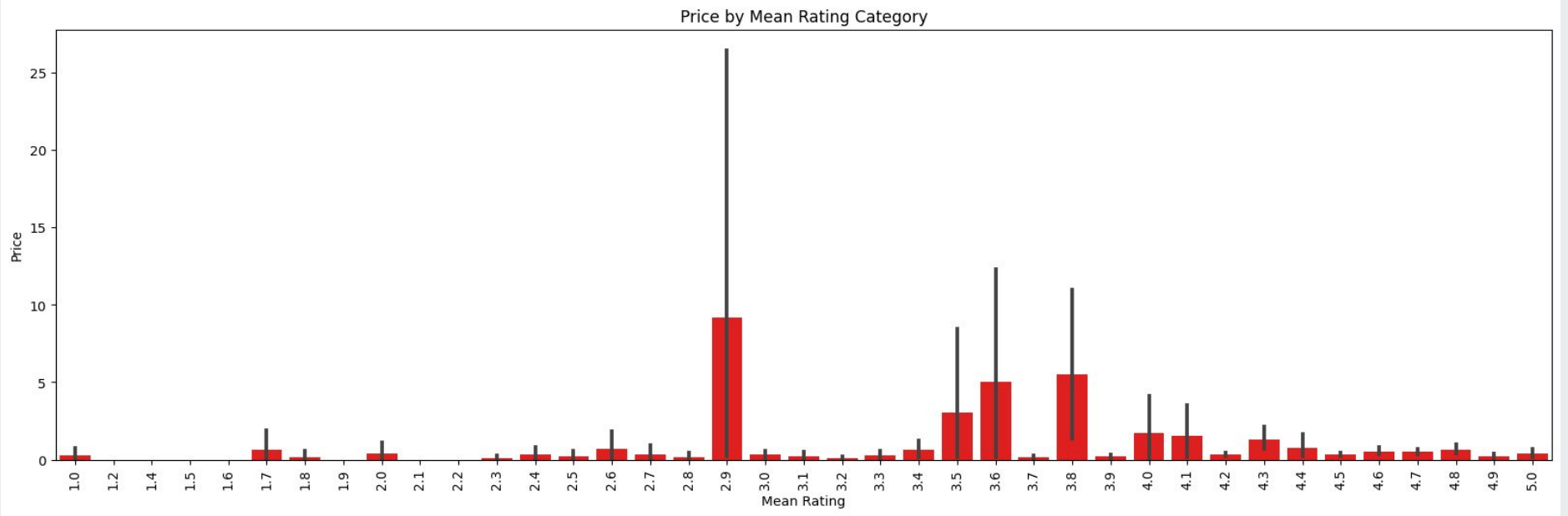
Price Vs Rating Comparison

Free apps tend to have lower ratings compared to paid apps, it may suggest that users have higher expectations for paid apps.

Bars show a clear trend of higher ratings for higher-priced apps, it suggests that users may associate higher prices with better quality.

There is no clear trend or if free apps have similar ratings to paid apps

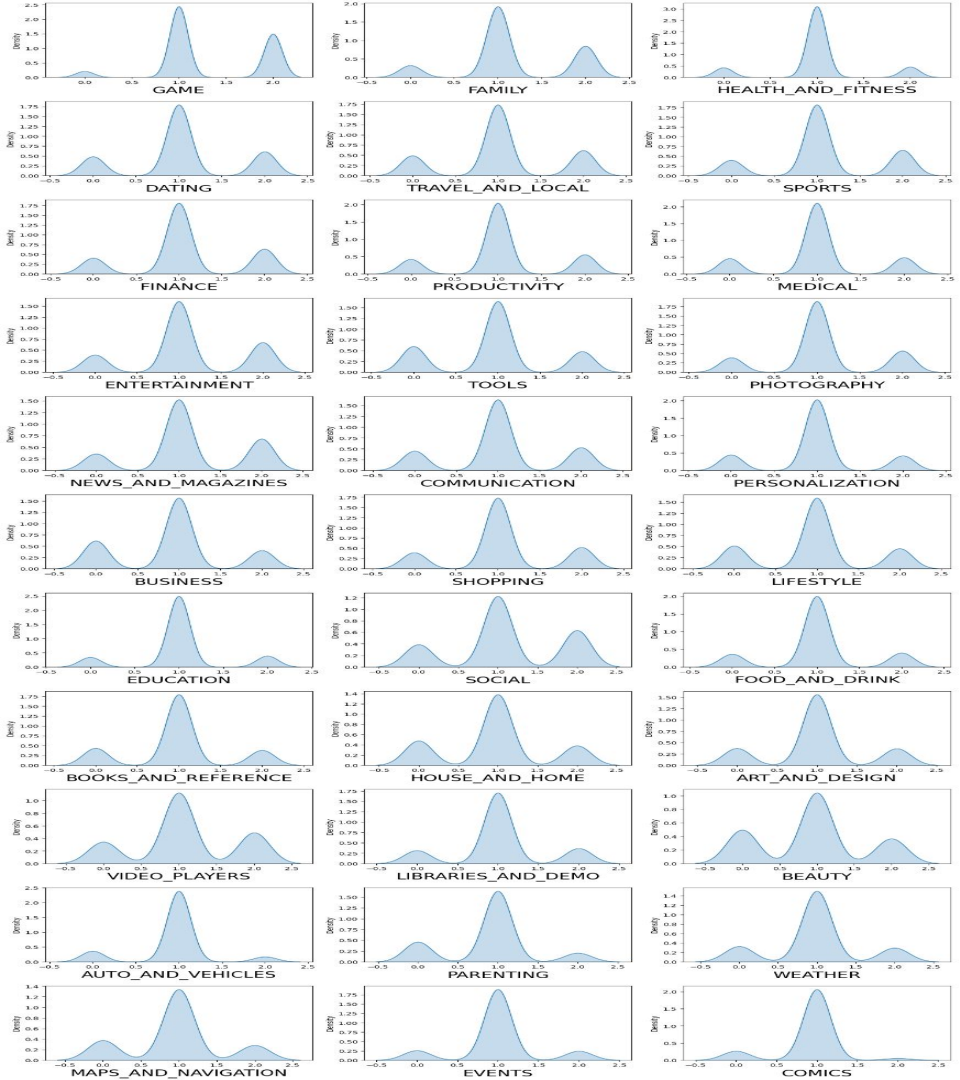
We can see that highly expensive apps are not necessarily well rated.



Sentiment Analysis for Categories

Positive Sentiment Categories: Game, Health&Fitness, Education this Categories have more positive sentiment and have less negative sentiment.

Negative Sentiment Categories: Beauty, Video_playes,Lifestyle, News and Magazine have more negative sentiment compared to others

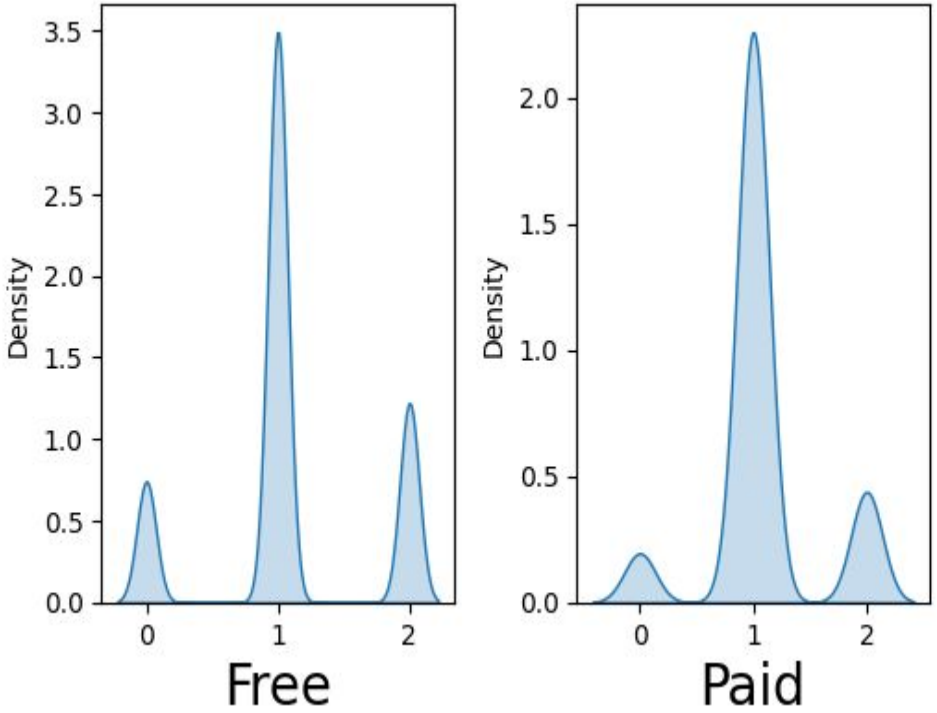


Sentiment Analysis for Type(Free and Paid)

For Free type negative sentiment is more as compared to paid type.

Also, for Free neutral sentiment is more as compared to paid type.

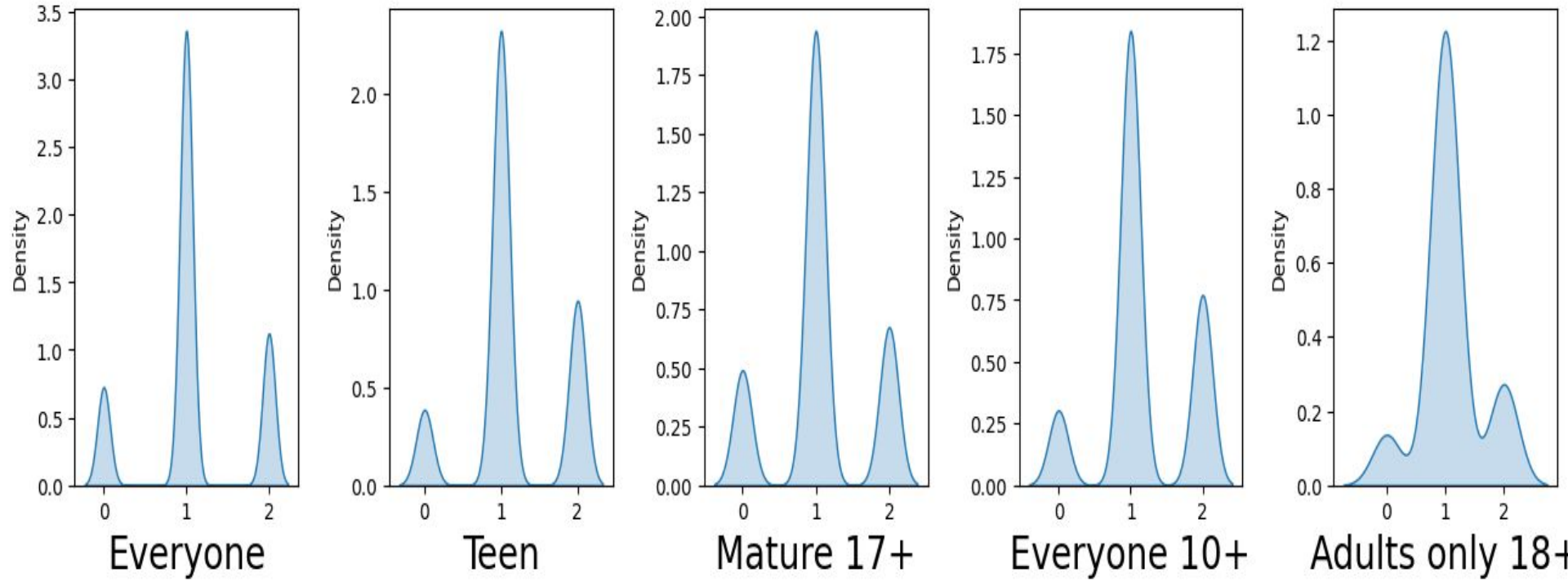
Positive sentiment is having more density for free type rather that paid type



Sentiment Analysis for Content Rating

Almost for every Content Rating sufficient negative sentiment

For 'Everyone' have more positive sentiment as compared to others.

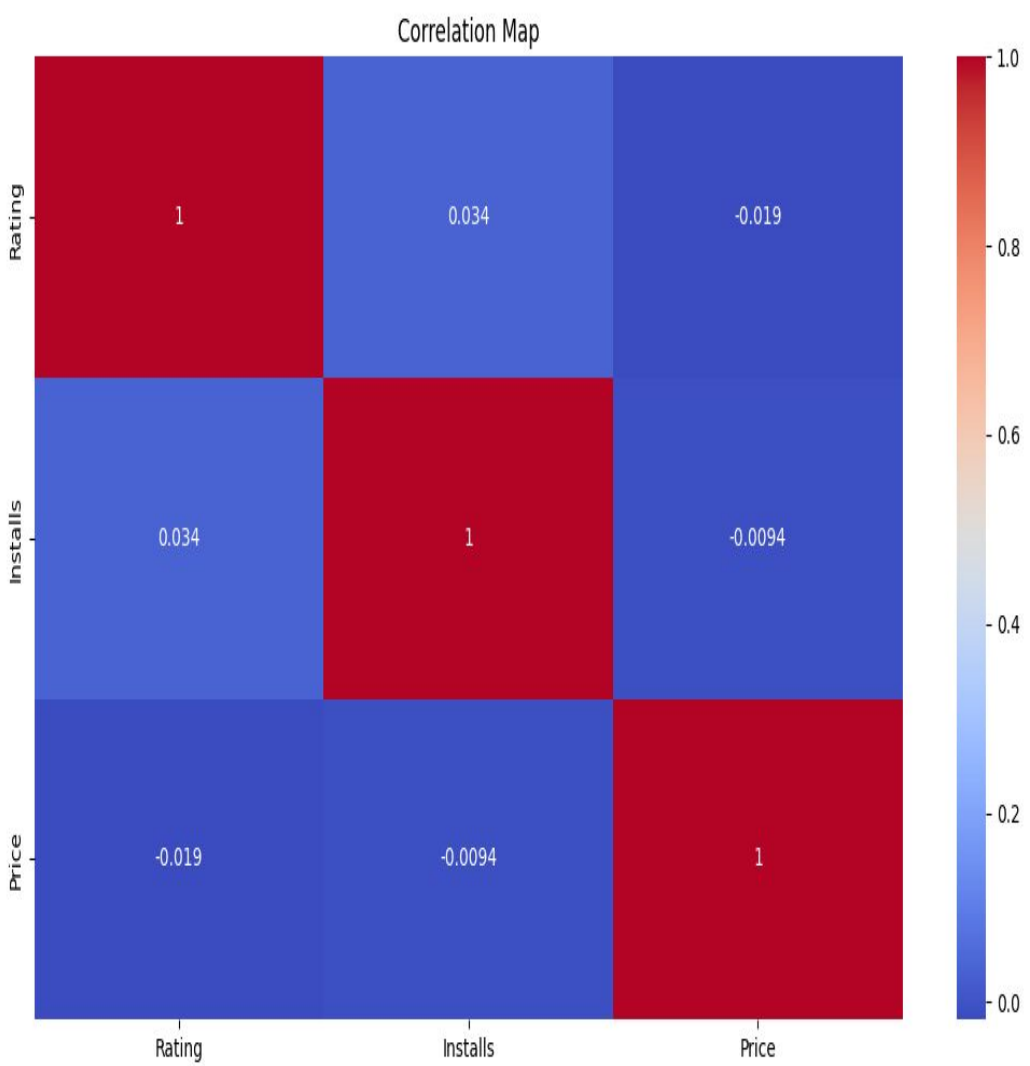


Correlation Heatmap for play store Dataset

Positive Correlation between Reviews and Installs

No Strong Correlation with Rating

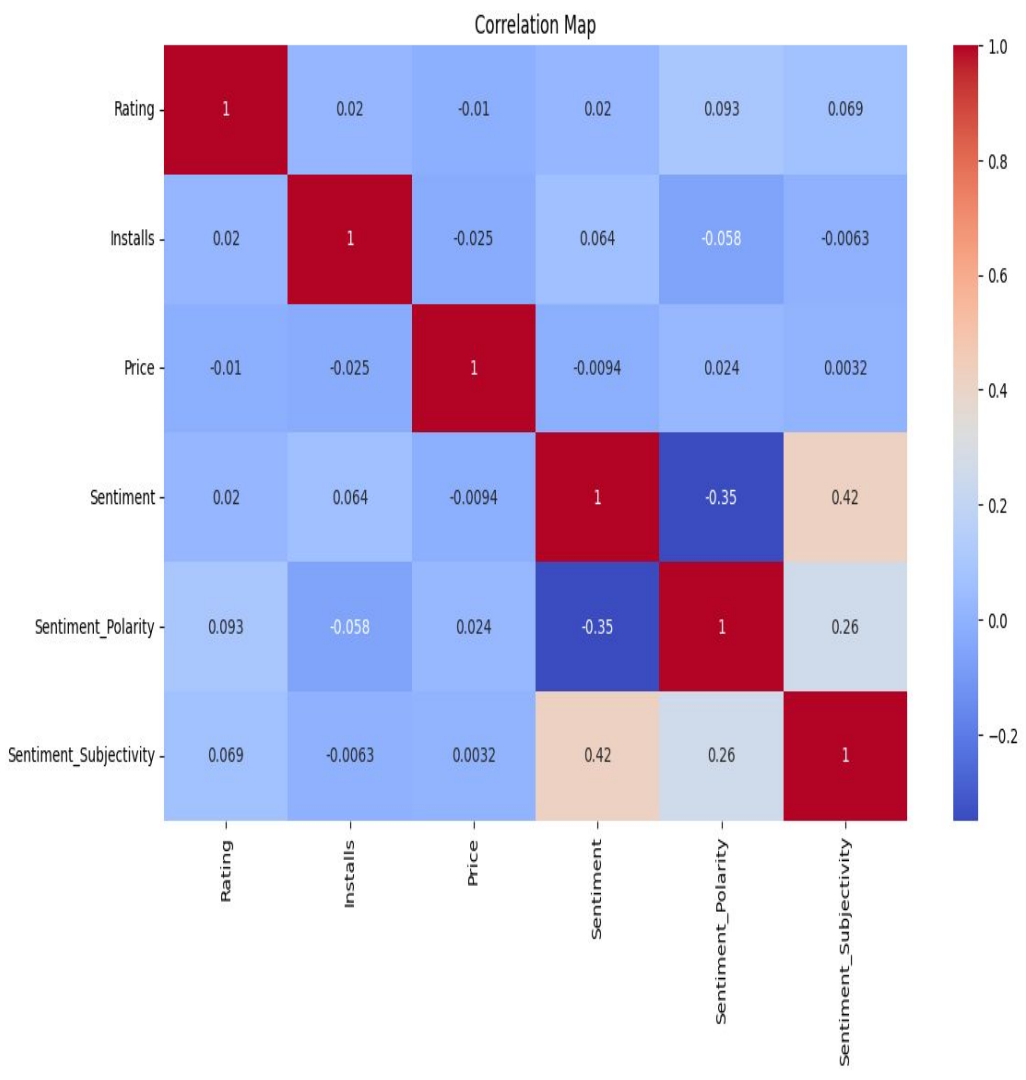
Price and Reviews are Weakly Correlated.



Correlation Heatmap for play store Dataset

Strong Positive Correlations for Reviews and Installs, Sentiment_Polarity and Sentiment_Subjectivity.

Strong Negative Correlations for Price and Rating



Pair Plot for play store dataset

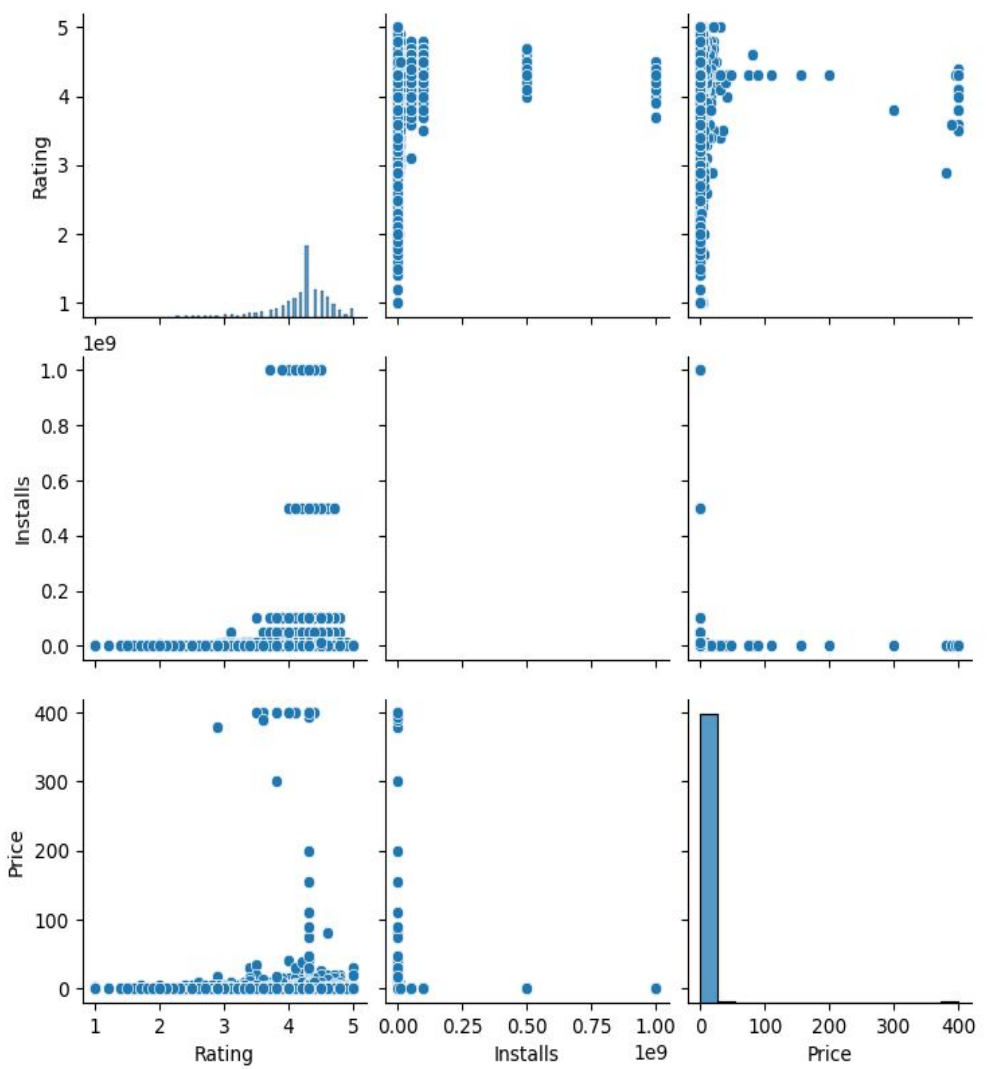
Rating vs Reviews: High rating app have more reviews.

Rating vs Installs:High rating app have more Installs.

Rating vs Price: High rating app have high price

Reviews vs Price: More reviews for highly paid app

Reviews vs Installs: High reviews app have More number of installs.



Solution to Business Objective

With Respect to mention business objective at beginning of this report, Found for meaningful insights, which will be helpful to client for identifying further business Strategies

Insights as below :

- **For Categories which have less number of application should be more highlighted for making change for current trend.**
- **For Categories which have less number of installs should be more focused for making change for current trend.**
- **Identity genre for less count of application with paid and free type and change for change for current trend.**
- **Try for focusing rating of applications.**
- **For Categories which have low rating should be more highlighted for making change for current trend.**
- **Focus for negative and neutral sentiment of Categories, rating and type of application.**

Conclusion

1. Most Popular Categories: Family, Games, tools.
2. Most Installed Categories: Family, Games, tools
3. Proportion of Free and Paid Apps: Free apps 92% and Paid Apps 8%. Comparison of App Types:"Free" segment is much larger than the "Paid" segment, it indicates that there are more free apps available in the dataset.
4. Most Popular Genre:Tools,Entertainment,Education.
5. Top 3 Genre for Free apps : Medical,Personalization and Tools
6. Top 3 Genre for Paid apps : Tools,Entertainment and Education
7. Chart 5 can get insight that, more number of applications rating lies near 4 to 4.5. Very Few application with rating from 1 to 3.
8. Top Rated Categories: Family, Game and Tools.
9. There are 82% applications which are belong to 'Everyone' while 11%,4% belongs to 'teen' and 'Mature 17' Content Rating. There are very less applications for 'Adults only 18+'.
10. Most used Android Versions: '4.1 and up' and '4.0.3 and up'.
11. Free apps tend to have lower ratings compared to paid apps, it may suggest that users have higher expectations for paid apps.There is no clear trend or if free apps have similar ratings to paid apps.
12. Positive Sentiment Categories: Game, Health&Fitness, Education this Categories have more positive sentiment and have less negative sentiment.

13. Negative Sentiment Categories: Beauty, Video_playes,Lifestyle, News and Magazine have more negative sentiment compared to others

14. For Free type negative sentiment is more as compared to paid type.for Free neutral sentiment is more as compared to paid type.

15. Almost for every Content Rating sufficient negative sentiment

16. Strong Positive Correlations for Reviews and Installs, Sentiment_Polarity and Sentiment_Subjectivity.
17. Positive Correlation between Reviews and Installs
18. Rating vs Reviews: High rating app have more reviews
19. Reviews vs Installs: High reviews app have More number of installs

Questions?

Thank-You