Capstone Project 1

Google Play Store App Review Analysis

Presented By:

TEAM- Scraper Nerds
Navneet Sawhney
Tinu J Pooppally
Biswanath Das
Manoj Chetry
Shubham V Jadav

What is Google Play Store and why Analyse it?

- An online platform for downloading and purchasing mobile apps and games for Android devices.
- It is the official app store for the Android operating system and is maintained by Google.
- Google Play Store is one of the most popular app stores in the world, with over 2.9 million apps available for download as of 2021.
- According to a study by App Annie, Google Play Store accounted for 85% of all app downloads globally in 2020. This is primarily due to the large market share of Android devices, which make up over 75% of the global smartphone market. (The state of mobile 2020 (2020). App Annie)
- As per latest Google Play stats, there are 3.6 million apps currently at the Google Play Store. The number is constantly rising as around 3,739 apps are added to the Play Store every single day. ()

Problem Statement

To perform exploratory data analysis (EDA) on the Google Play Store in order to understand the trends and patterns in app downloads, user ratings, and revenue generated by apps with special focus on the medical category.

AImaBetter

What is Exploratory Data Analysis (EDA)?

Columns /Attributes in Google Play store Dataset 1

The data set contains the following columns:

- **App:-** This Column contains the name of the Apps
- **Category:-** This contains the category to which the App belongs. The category column contains 33 unique values.
- **Rating:-** This column contains the average value of the individual rating the App has received on the Play store. Individual rating values can vary between 0 to 5.
- **Reviews**:- This column contains the number of people that have given their feedback for the App.
- **Size:-** This column contains the size of the app i.e. The memory space that the App occupies on the device after installation.
- **Installs:-** This column indicates the number of times that the App has been downloaded from the play store, these are approximate values and not absolute values.
- **Type:-** This column contains only two values- free and paid. They indicate whether the user must pay money to install the app on their device or not.
- **Price:-** For paid apps, this column contains the price of the app, for free apps, it contains the value 0.
- **Content Rating**:- It indicates the targeted audience of the app and their age group.
- **Genre**:- This column contains which genre the app belongs to, genre can be considered as a sub-division of a Category.
- **Last updated:-** This column contains the info about the date on which the last update for the app was launched.
- **Current version:-** Contains information about the current version of the app available on the play store.
- **Android version:-** Contains information about the version of the Android OS on which the app can be installed.

Columns /Attributes in User Review Dataset 2

User Reviews Dataset has 64295 rows and 5 columns. The 5 columns are identified as follows:

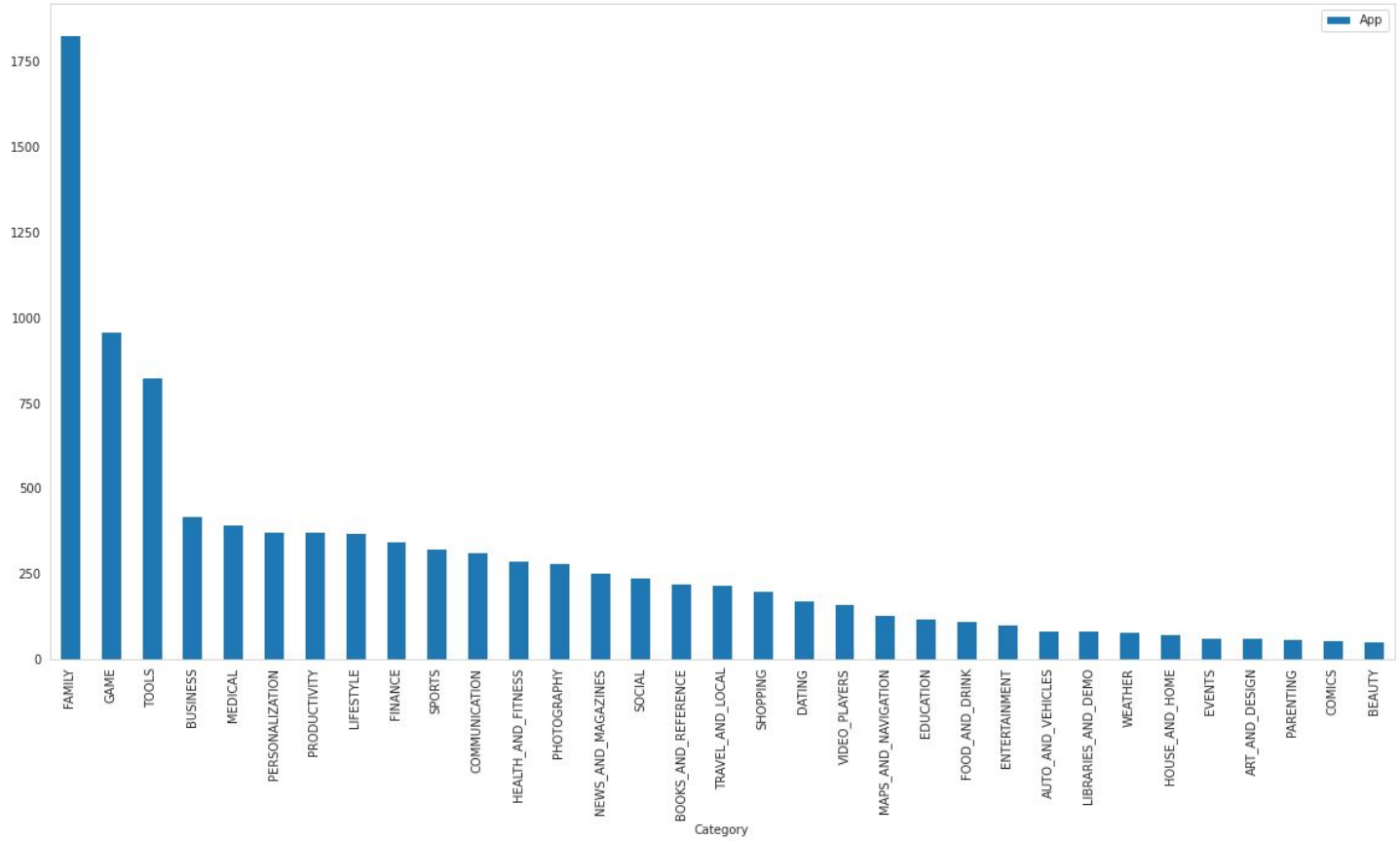
- **App:** Contains the name of the App.
- **Translated Review:** It contains the English translation of the review dropped by the user of the App.
- **Sentiment:** It gives the attitude/emotion of the writer. It can be ‘Positive’, ‘Negative’, or ‘Neutral’.
- **Sentiment Polarity:** It gives the polarity of the review. Its range is [-1,1], where 1 means ‘Positive statement’ and -1 means a ‘Negative statement’.
- **Sentiment Subjectivity:** A value from 0 to 1 indicating the subjectivity of the review. Lower values indicate the review is based on factual information, and higher values indicate the review is based on personal or public opinions or judgement.

Point of discussion

- Evaluating apps based on their **category and rating**
- Evaluating apps based on the **version of the app**
- Evaluating apps based on their **Genres**
- Evaluating apps based on the **sentiment of reviews**
- Evaluating Apps in **Medical Category**

Category and Rating

Number of Apps per category



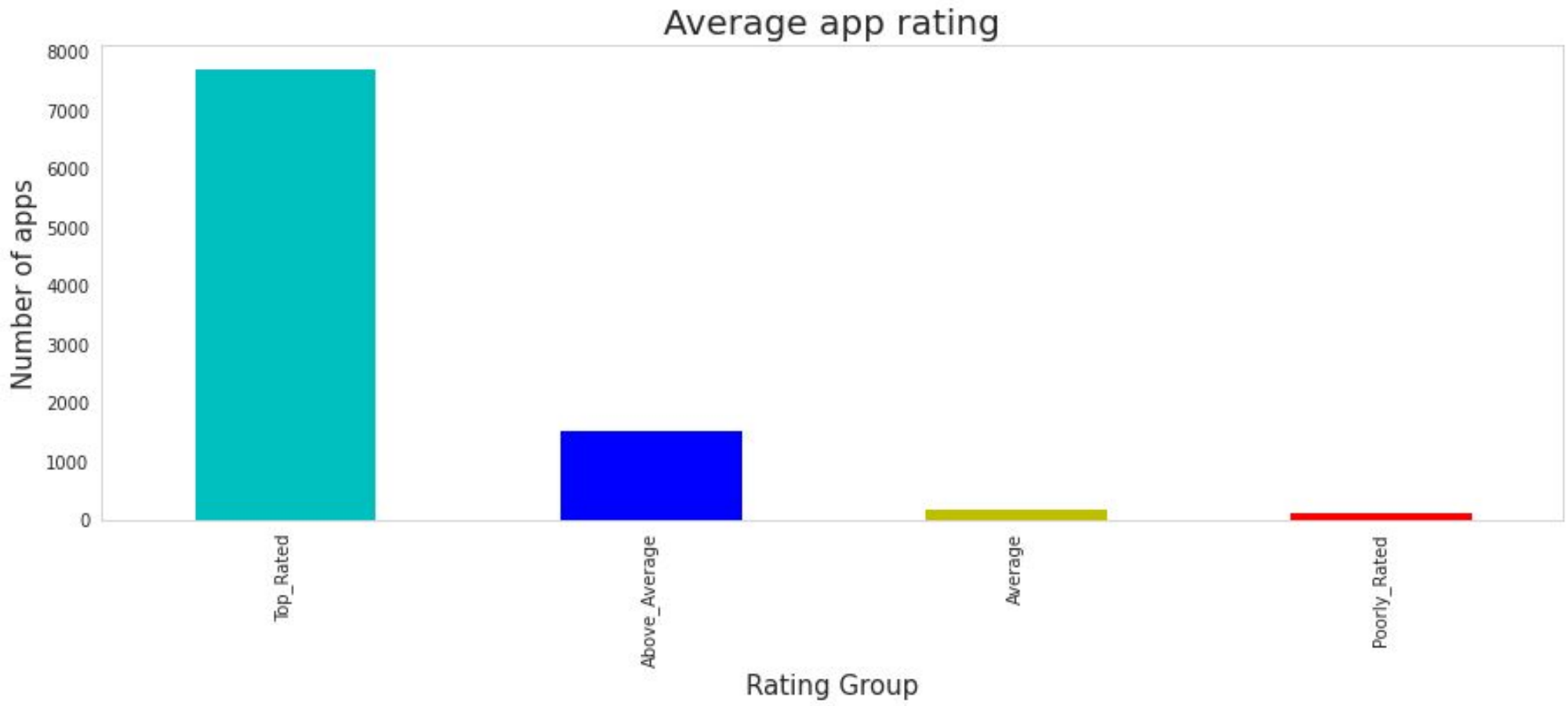
The above bar graph represents the distribution of several apps in different categories in the Play Store. It can be inferred that FAMILY Category has the maximum number of Apps.

The top categories on Play Store based on the rating

From the analysis, we can conclude,

1.For all the categories, the average rating is above 4 stars.

2.Most of the Apps are "Top Rated" as we can see in the above visualization



The Average App Rating

An application that has the highest number of user engagement in the category which has the highest number of applicants.

- We already know that the 'FAMILY' category has the highest number of application

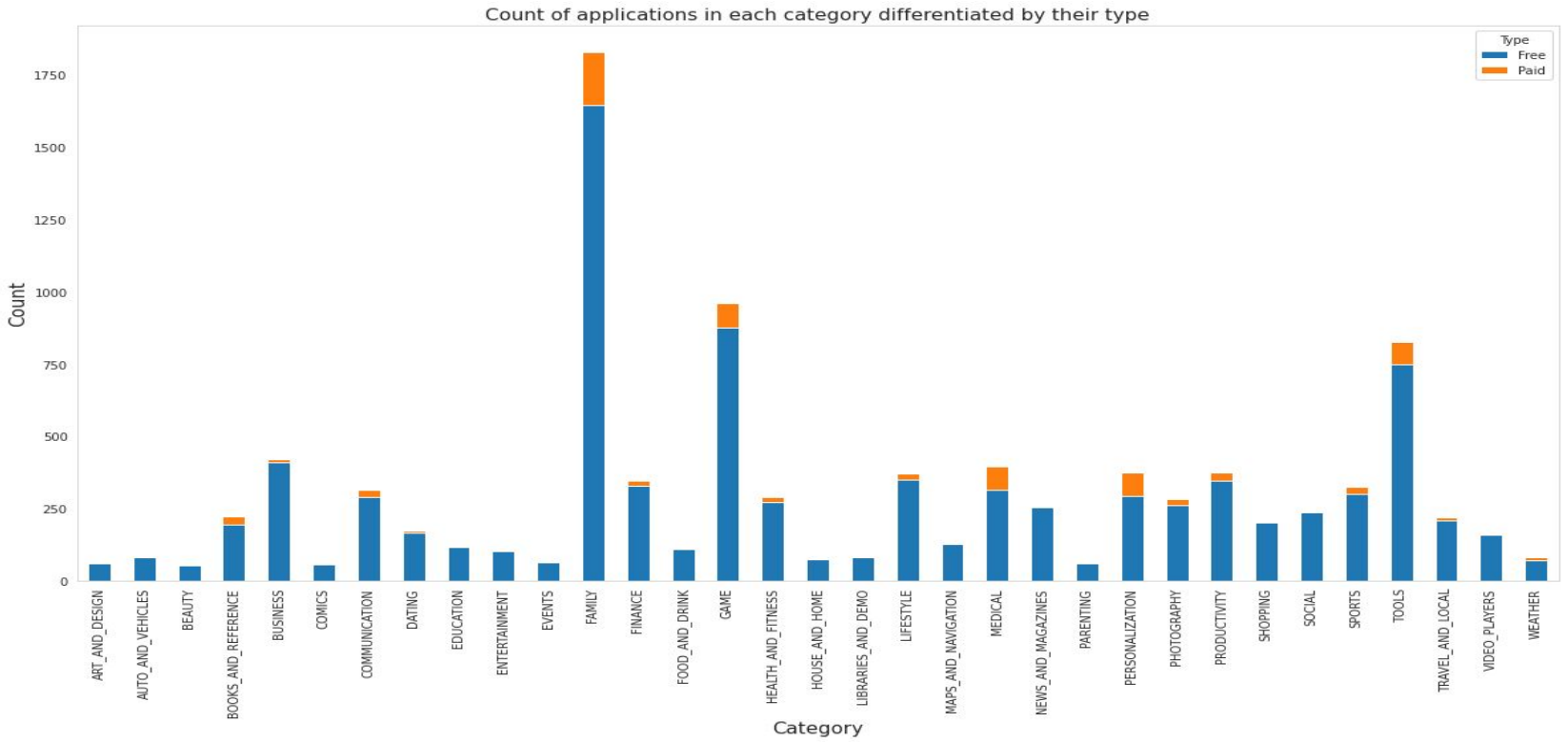
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Rating_Group
3975	Hay Day	FAMILY	4.5	10053186.0	94.0	100000000.0	Free	0.0	Everyone	Casual	2018-06-28	1_39_93	4.0.3 and up	Top_Rated
4111	Talking Tom Cat 2	FAMILY	4.3	3213548.0	55.0	100000000.0	Free	0.0	Everyone 10+	Entertainment	2018-07-25	5.3.5.16	4.1 and up	Top_Rated
4217	Talking Ginger	FAMILY	4.2	1752017.0	52.0	100000000.0	Free	0.0	Everyone	Entertainment	2018-07-05	2.5.6.11	4.1 and up	Top_Rated
4229	Where's My Water? Free	FAMILY	4.4	1372013.0	57.0	100000000.0	Free	0.0	Everyone	Puzzle;Brain Games	2017-11-21	1.10.0	4.1 and up	Top_Rated
4247	Talking Tom Cat	FAMILY	4.3	1838090.0	NaN	100000000.0	Free	0.0	Everyone 10+	Casual	2018-07-27	Varies with device	4.1 and up	Top_Rated
4329	Tom Loves Angela	FAMILY	4.1	1111915.0	50.0	100000000.0	Free	0.0	Everyone	Entertainment	2018-04-30	2.2.1.3	4.1 and up	Top_Rated
4797	Talking Tom & Ben News	FAMILY	4.4	1131937.0	41.0	100000000.0	Free	0.0	Everyone	Entertainment	2018-04-30	2.4.0.7	4.1 and up	Top_Rated
6269	Bitmoji – Your Personal Emoji	FAMILY	4.6	2312084.0	NaN	100000000.0	Free	0.0	Teen	Entertainment	2018-07-25	Varies with device	4.3 and up	Top_Rated
6849	Bubble Witch 2 Saga	FAMILY	4.3	2838064.0	NaN	100000000.0	Free	0.0	Everyone	Casual	2018-08-06	Varies with device	Varies with device	Top_Rated

- Highest number of installs belong to the free app as expected.
  Among the lowest number of installs which is zero, most of the apps belong to a paid version of the app.

AlmaBetter

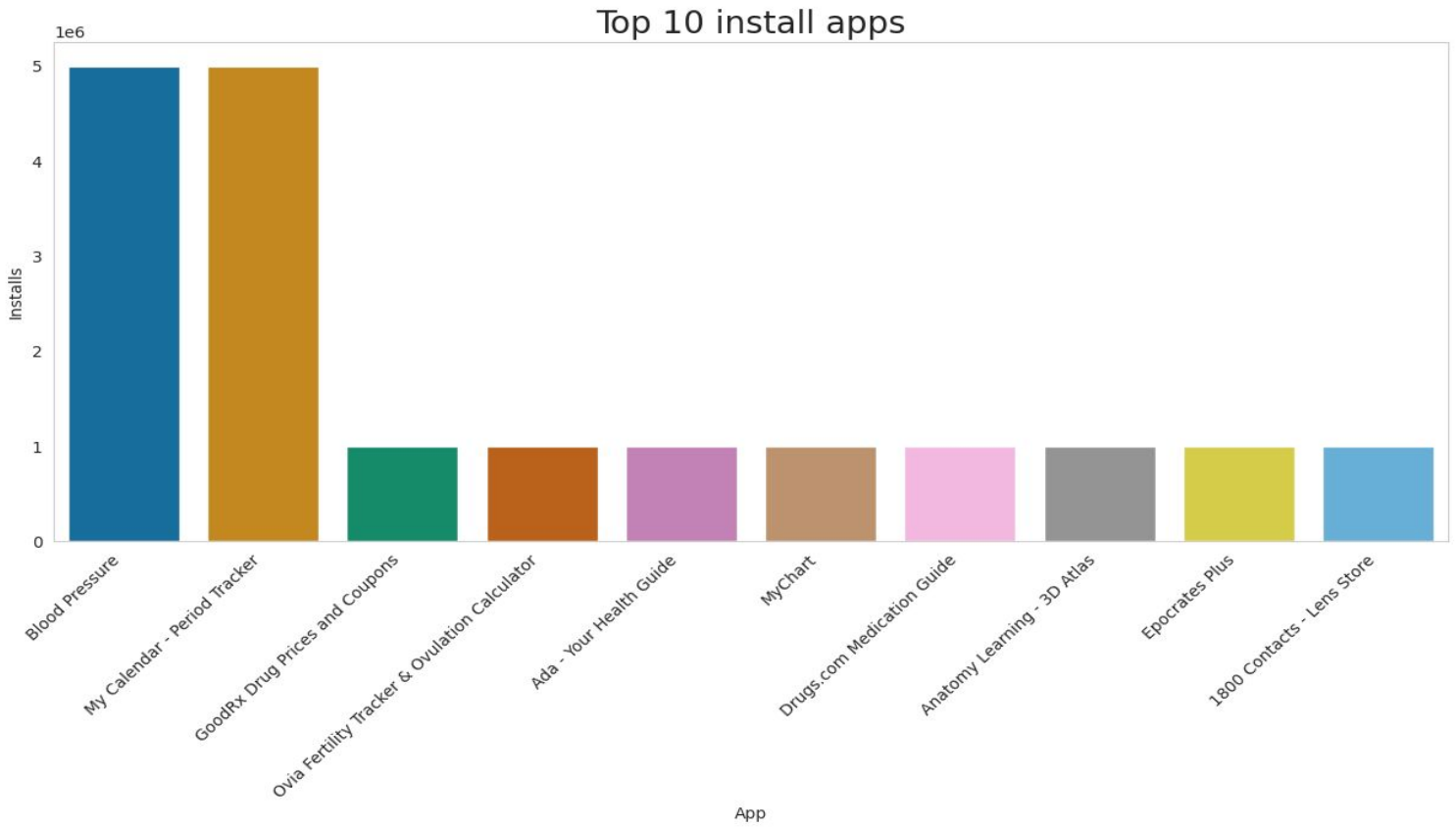
Count of applications in each category differentiated by their type.

- It can be inferred that certain app categories have more free apps available for download than others. In our dataset, the majority of apps in Family, Games, and Tools, as well as Social categories are free for users to install.



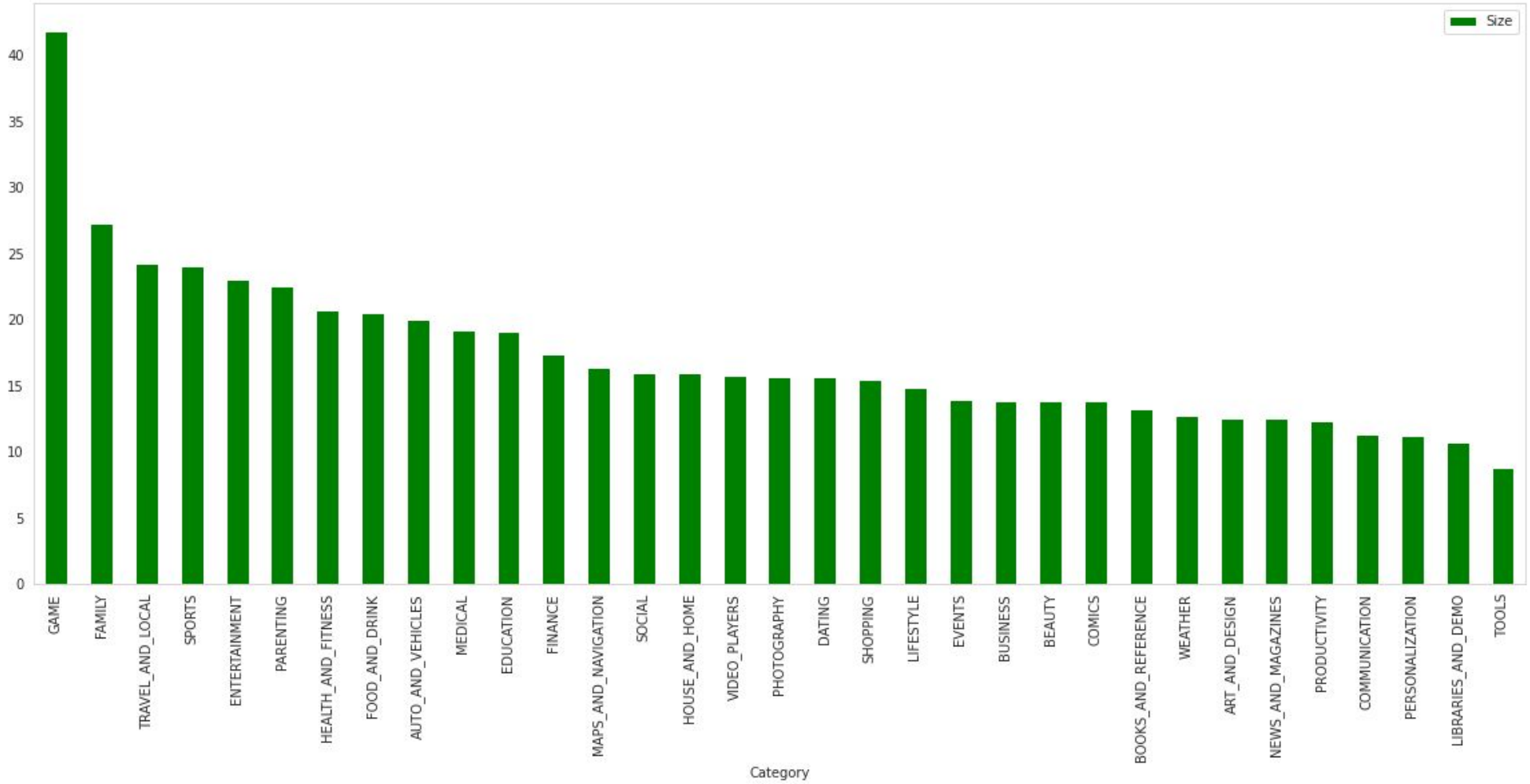
Top 10 Medical apps in any category based on their install count.

- We can see the top 10 Medical Category apps in the above visualization.

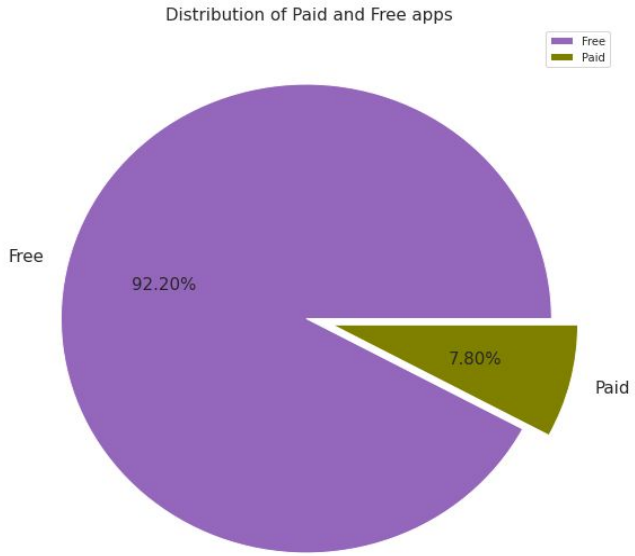


Optimal App Size

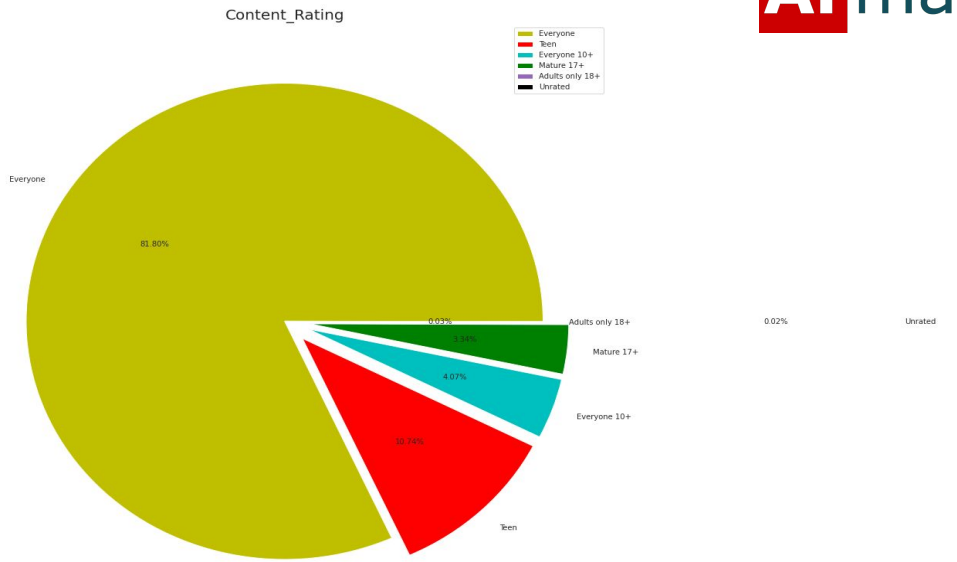
- The average App Size is in the range of 10-25 Mb.
- The Game category has the maximum average App Size.
- Tools has the least average App Size.
- The average App Size of the Medical category stands at 19 Mb.



Percentage of free apps over paid apps? Also Checks the distribution of Content Rating

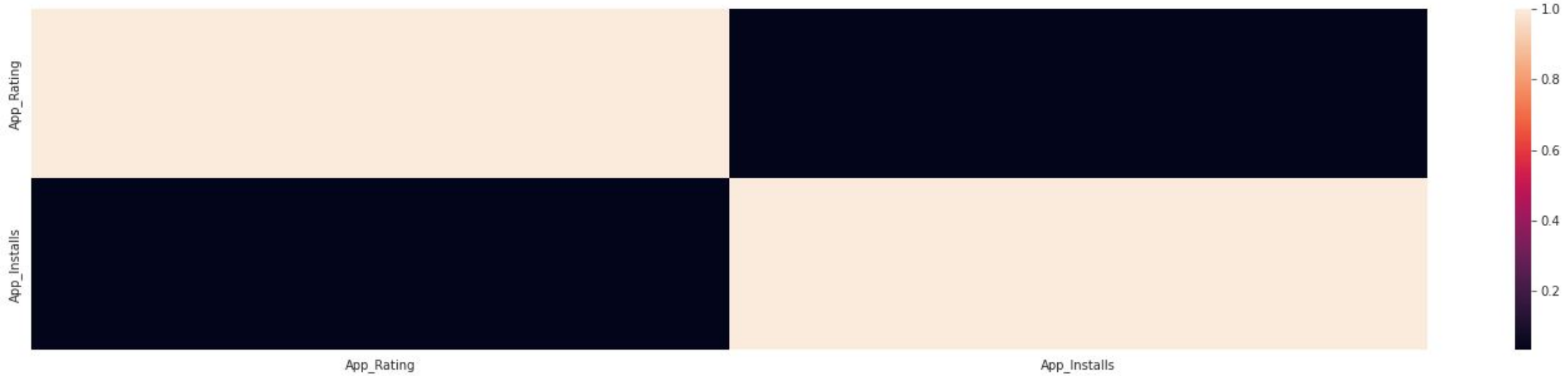


- It can be inferred that around 92% of Apps are free and 8% of apps are paid.



- As per the above pie chart, around 82% of Apps are created for Everyone the least are Unrated i.e. 0.02 followed by adults only 18+ i.e. 0.03.

Does the application rating affect its user engagement

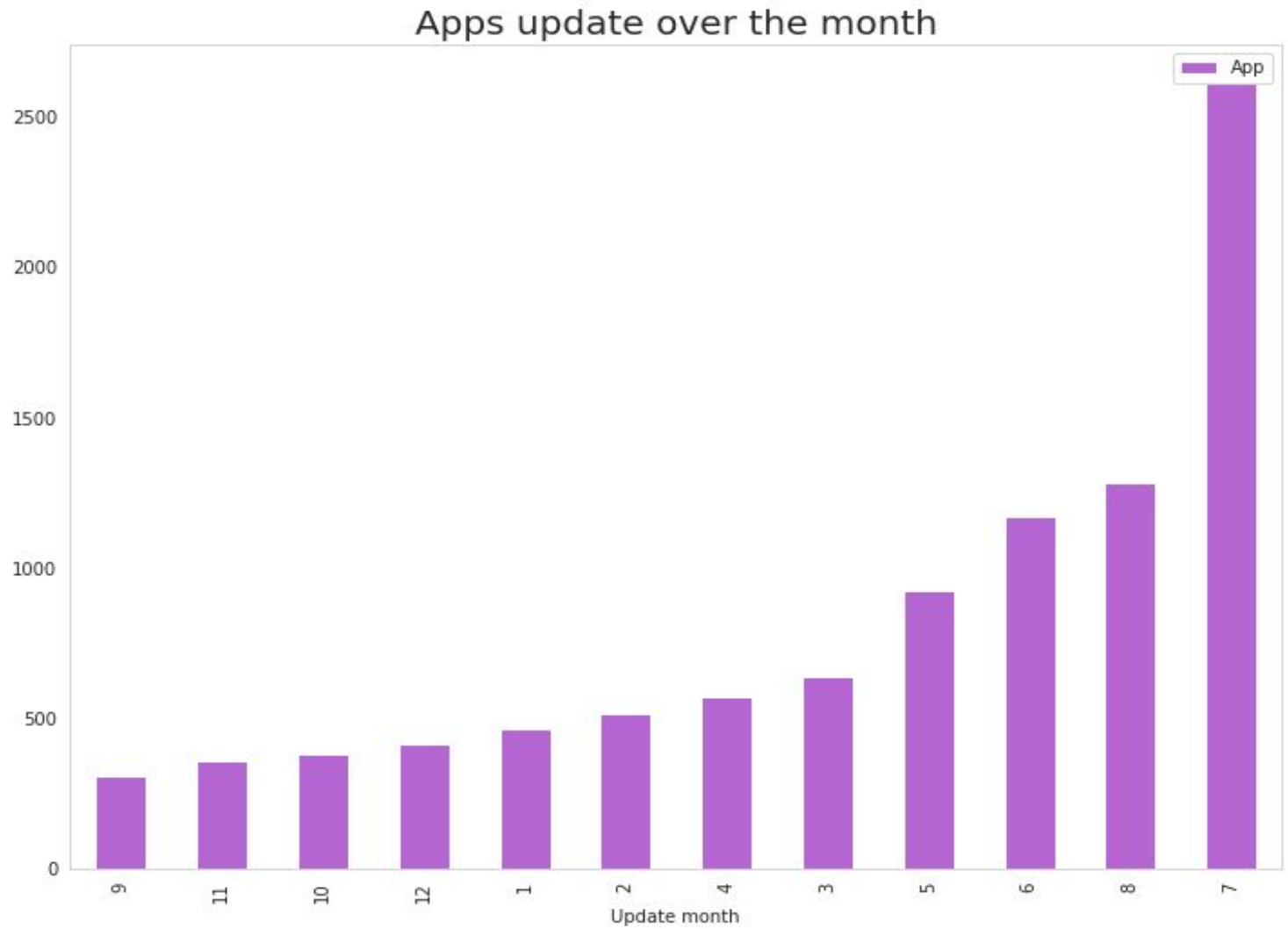


- The linear correlation coefficient of approximately 0.039 suggests that there is no appreciable linear correlation.

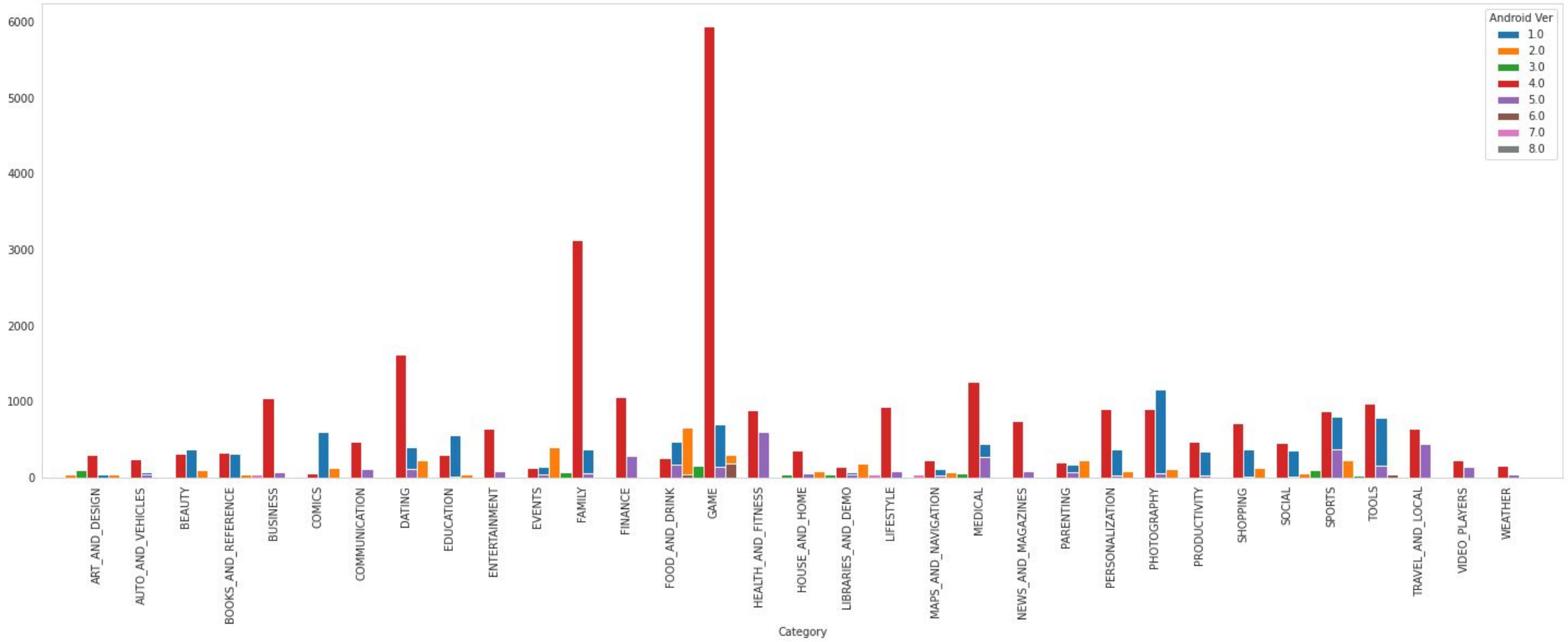
App Versions

Distribution of apps updated over the Month

- It can be inferred that the month of July records the highest App updates which are followed by August and June showing that the majority of the updates happen in the middle of the year.



Android version based on each category

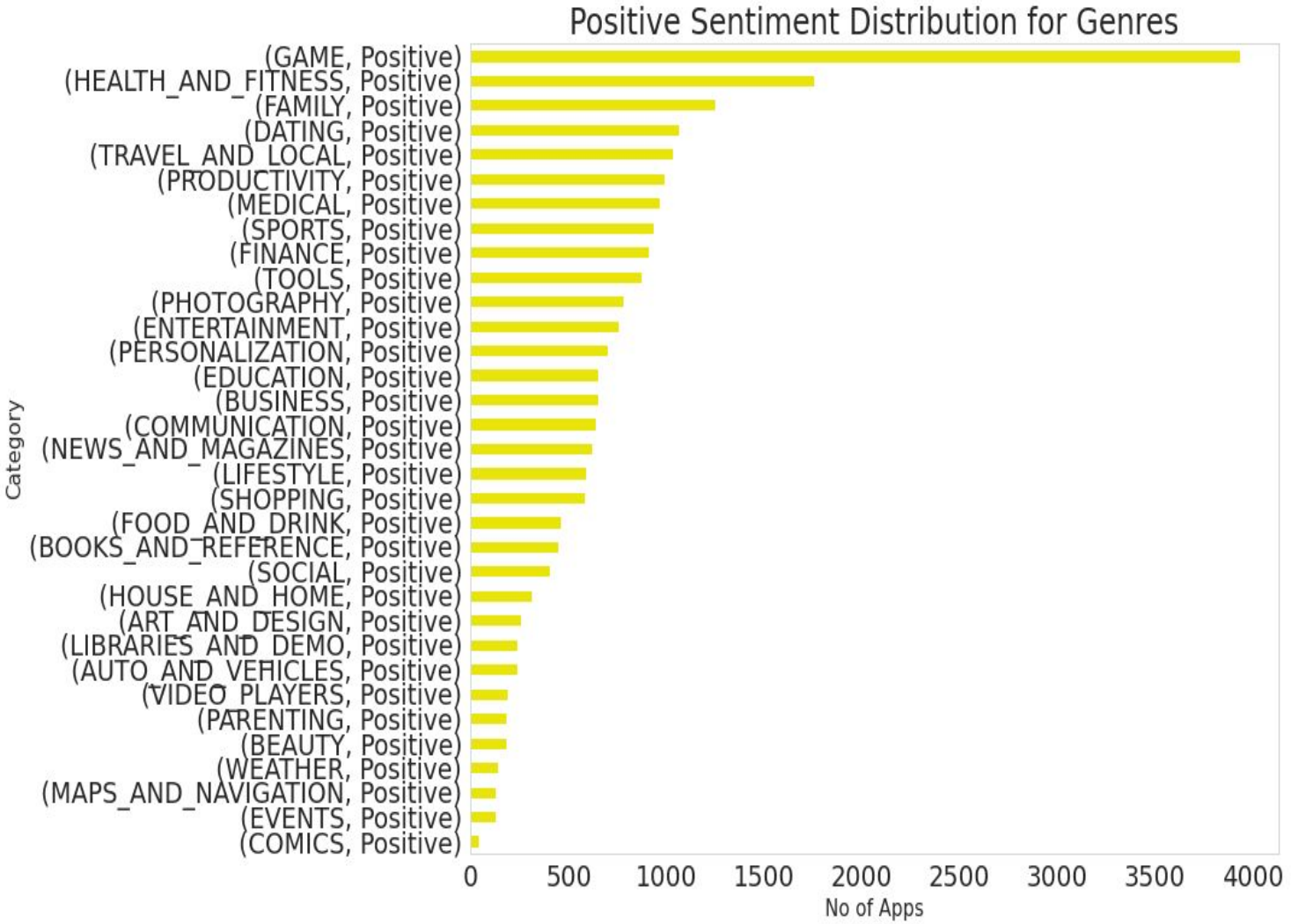


- It is evident from the above plot that the majority of the apps are working on Android_Ver 4.0 and up.

Genres

The genre that has the most positive sentiment

- It is likely that the game genre has the most positive reviews in the Play Store, as games tend to be highly engaging and entertaining for users. With more than 3900+ positive reviews
- The Health and fitness genre of apps on the Play Store has the second-highest number of positive reviews. With 2400+ positive reviews
- The Family genre of apps on the Play Store has the third-highest number of positive reviews. With 1500+ positive reviews
- Comics genre of apps on the Play Store has the lowest number of positive reviews with less than 50



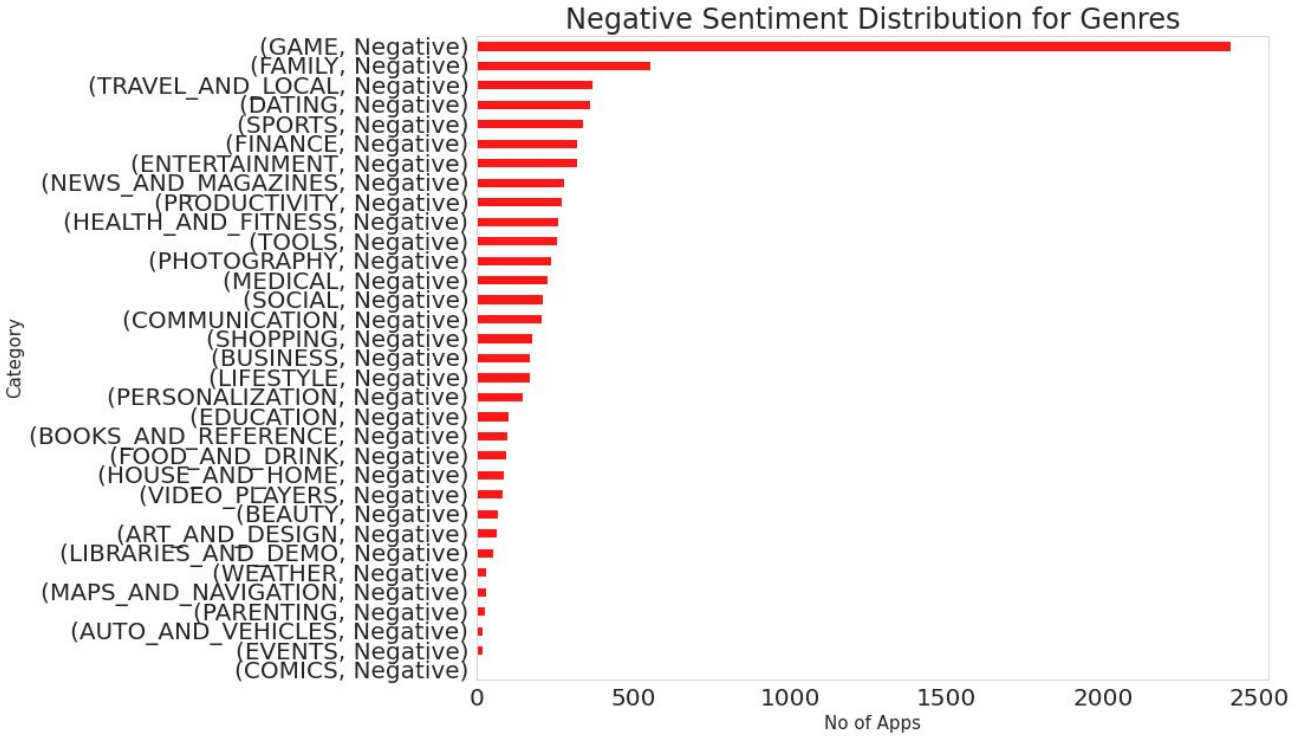
Genre that has most negative reviews

- The game genre may have the most negative reviews in the Play Store due to the high level of competition and the high expectations of users for gaming content. Additionally, games are often purchased or downloaded, which may lead to a higher rate of dissatisfaction among users who have spent money on a product that does not meet their expectations. With 2400+ negative reviews

- The Family genre of apps on the Play Store has the second-highest number of negative reviews. With 600+ negative reviews

- Travel and local genres of apps on the Play Store have the third-highest number of negative reviews. With 500 negative reviews

- Comics genre of apps on the Play Store has the lowest number of negative reviews with less than 50 reviews



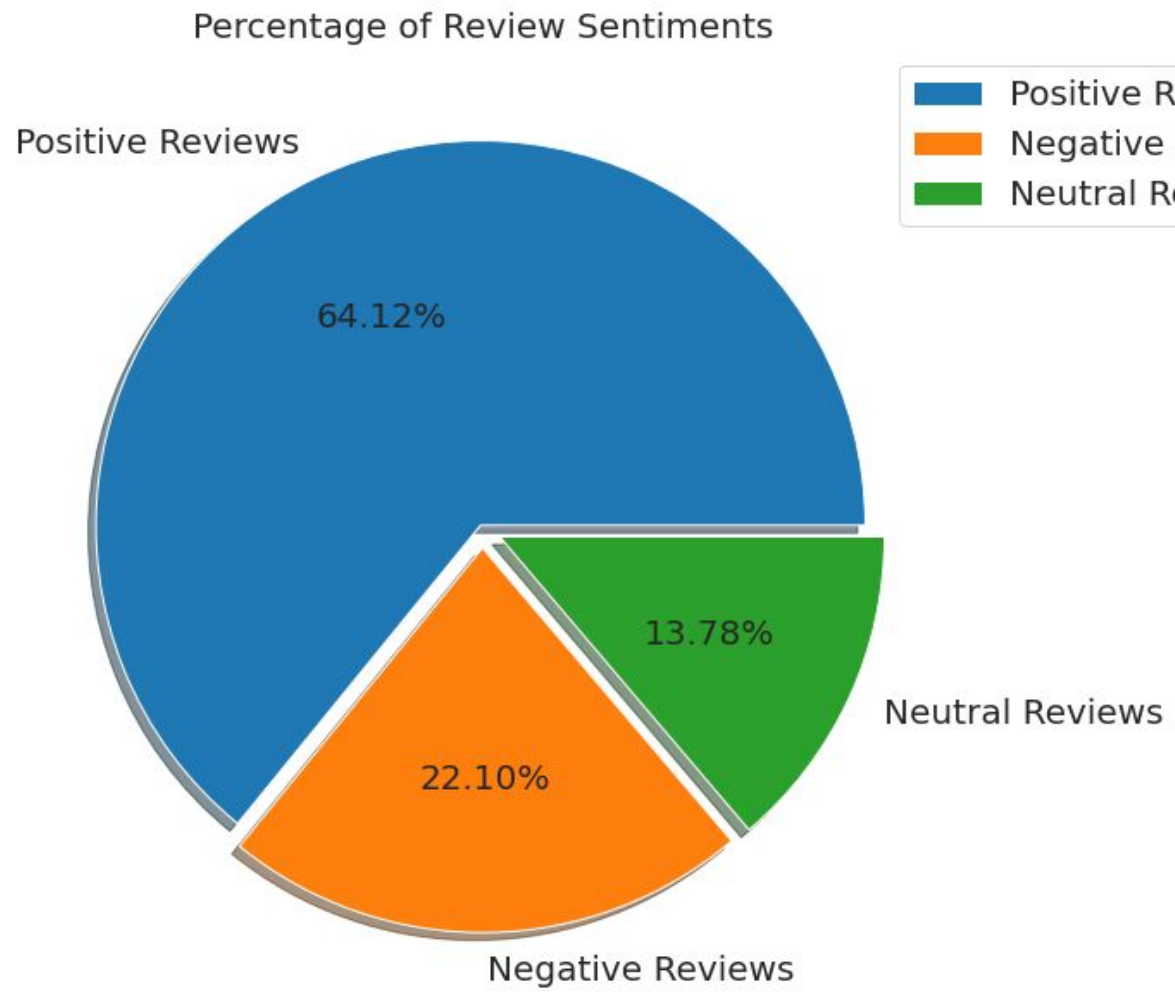
Reviews Sentiments

Type of sentiment that dominates the most

1.Among all the reviews the positive reviews dominate the most with a share of **64.12%**.

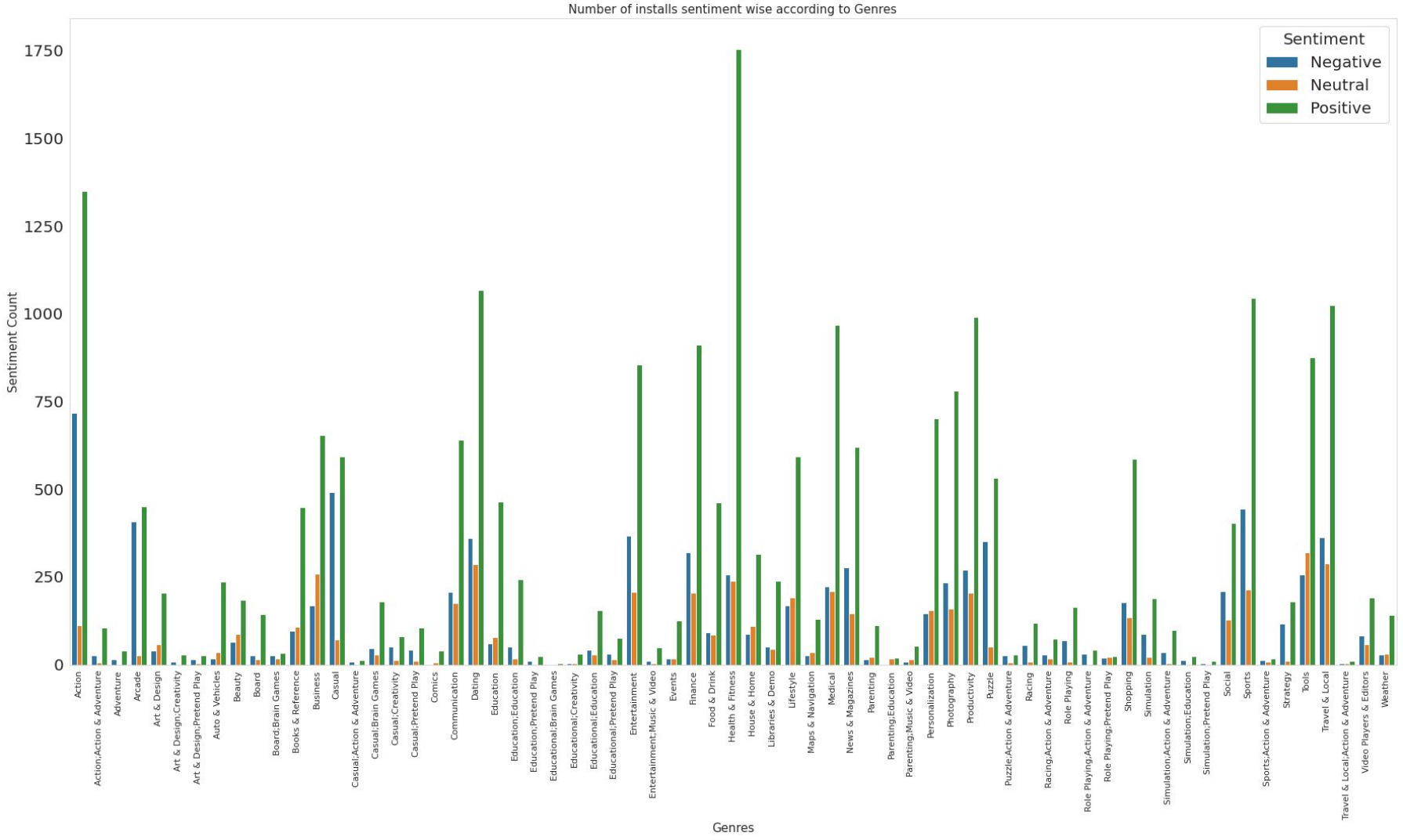
2.This is followed by negative reviews with a share of **22.10%**.

3. least number of reviews are neutral reviews with a share of **13.78%**.



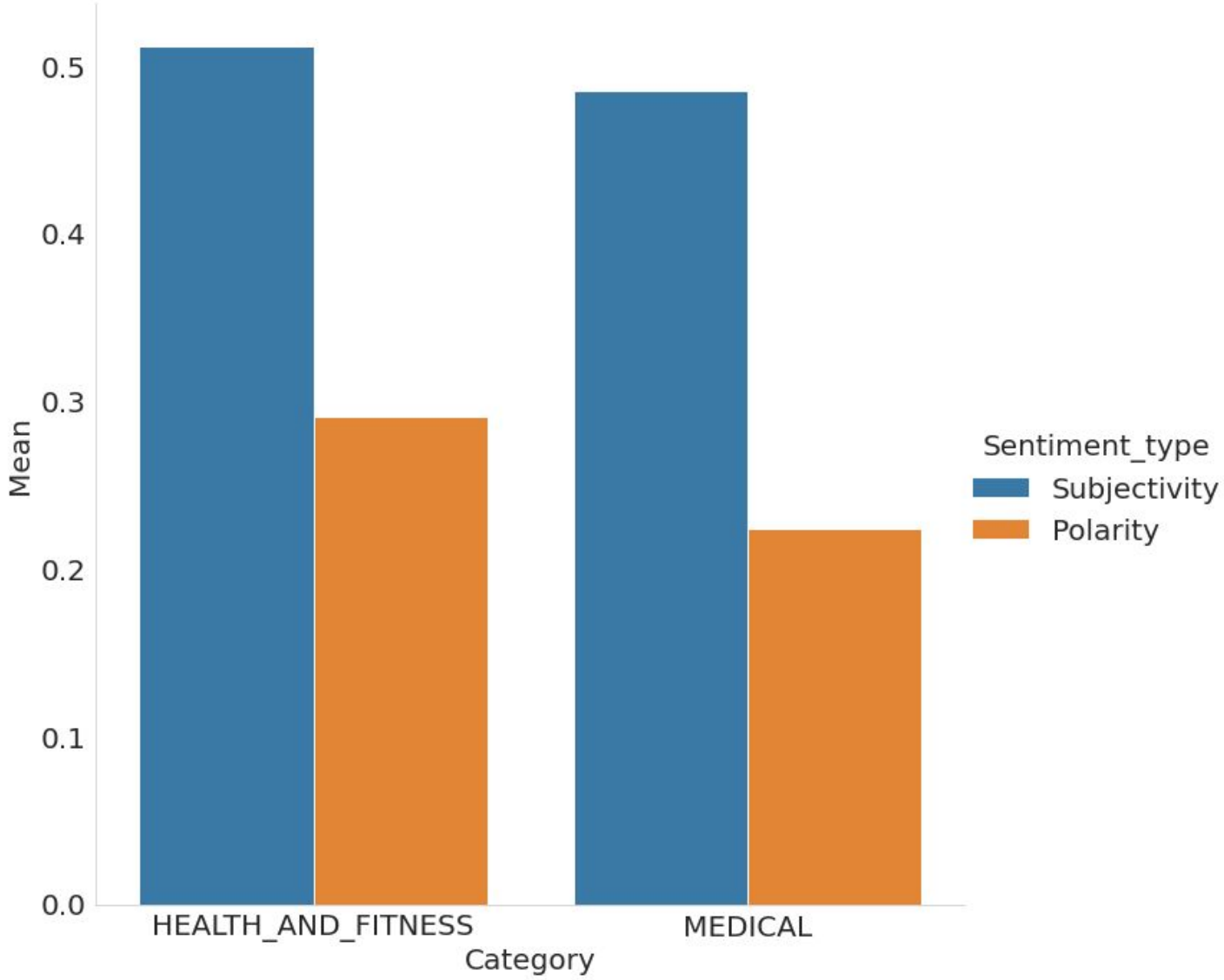
Distribution of sentiment within the different genres

- It can be seen that Health and Fitness have the highest number of positive reviews which is followed by Action. But it is also worthwhile to note that Action has the highest number of negative reviews as well while the ratio is much less for Health and Fitness.



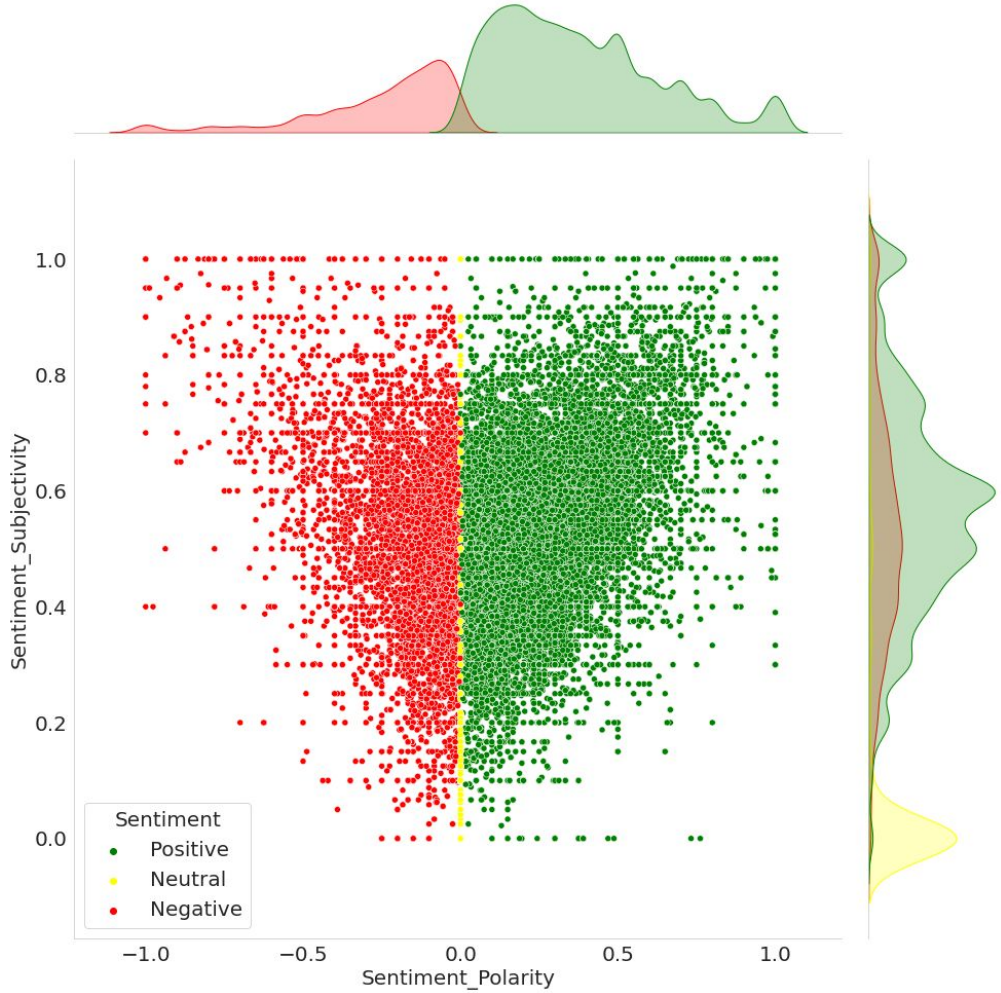
Sentiment Polarity within HEALTH_AND_FITNESS AND MEDICAL

This graph demonstrates how subjective the reviews are for the categories of health and fitness and medical. This further demonstrates that these apps are used by individuals more personally, hence any app falling under this category must be created in a way that may offer the best experience for the user, i.e. be user friendly



Sentiment Subjectivity proportional to Sentiment Polarity

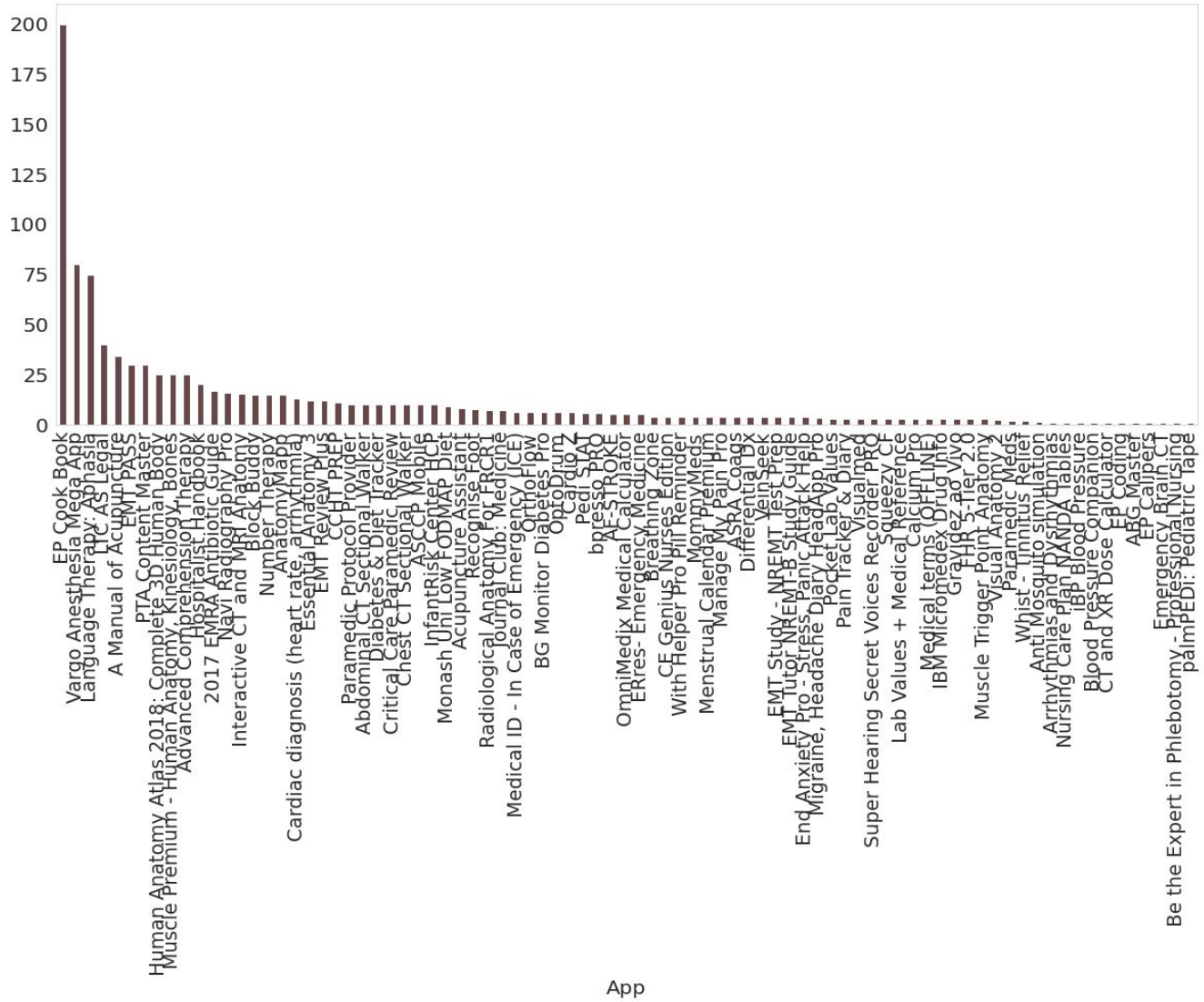
- From the above scatter plot, it can be concluded that sentiment subjectivity is not always proportional to sentiment polarity but in a maximum number of the case, shows a proportional behavior, when variance is too high or low.



Evaluating Apps in Medical Category

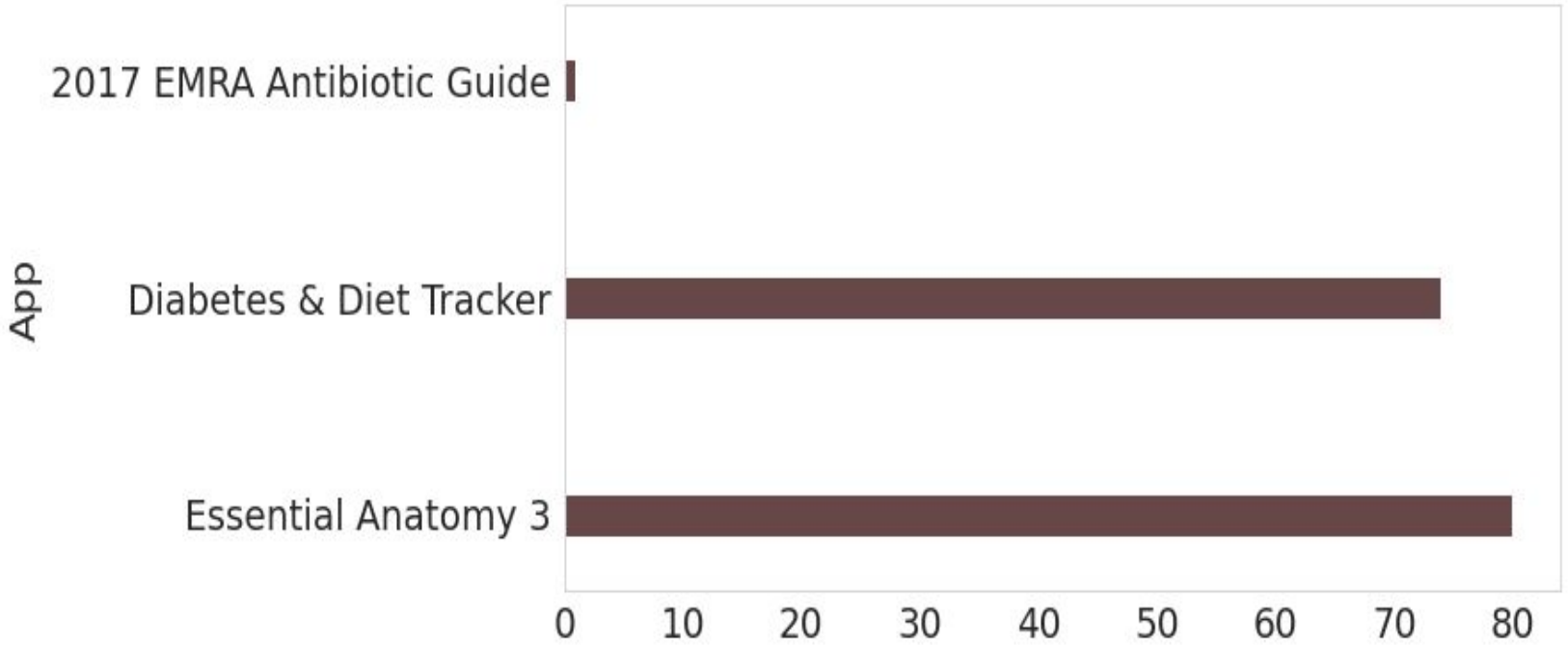
Price of paid apps in Medical Category

- Paid medical apps on the Google Play Store typically offer additional features and functionality not found in the free version. Some examples include advanced diagnostic tools, detailed explanations of medical conditions, and personalized treatment plans.
- We can see on a bar chart EP Cook Book is the most expensive app and its price is about 200$
- And the second most expensive app is Vargo Anesthesia Mega App and its price is about 80$
- After seeing a bar char we can say that most of the paid medical apps' prices are below 50$



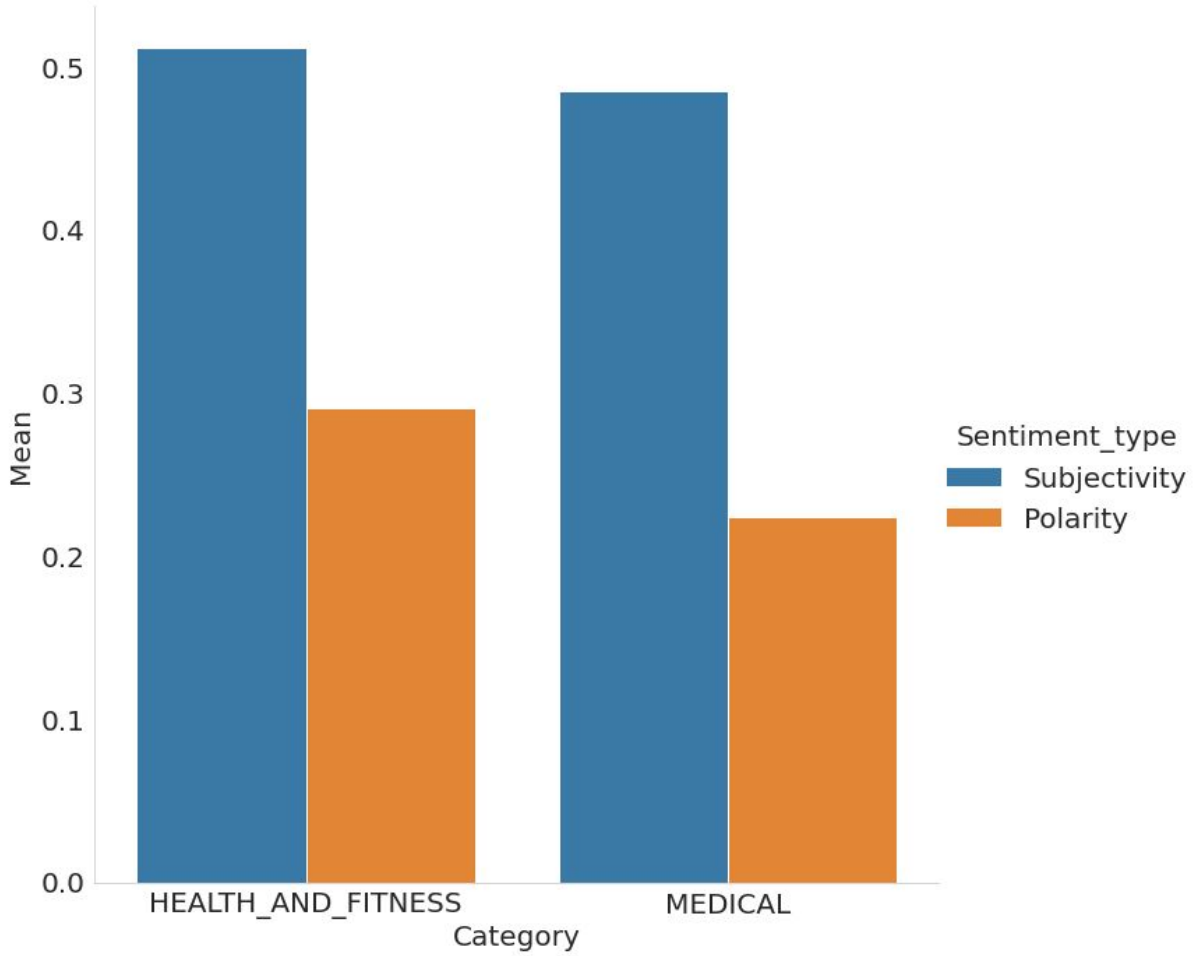
Top 3 paid medical apps based on number of Installs

Based on number of installs Essential Anatomy 3 clearly wins followed by Diabetes and Diet tracker.



Subjectivity of reviews related to Health & Fitness and Medical Category

This graph demonstrates how subjective the reviews are for the categories of health and fitness and medical. This further demonstrates that these apps are used by individuals more personally, hence any app falling under this category must be created in a way that may offer the best experience for the user, i.e. be user friendly.



Conclusion

As of now we know our findings and we have information in our hands. So it's time for us to put this valuable information in context and gain some knowledge.

For this project we have the context of a developer rather than an application which to be launched. So from our analysis, we should confirm the following factors to launch the successful application.

- The correlation between reviews and installs shows that reviews play an important role in determining the downloads of the apps.
- The average app size ranges from 10 to 25 and would be the optimum for best results.

- Most of the apps are rated in the range of 4 to 5 and on that note, the average rating of paid apps also stands at 4 showing that most of the paid apps are liked by the audience.
- For creating an app for the medical category, the app needs to be focused on providing personalized assistance as it is visible from the analysis that the majority of medical apps are reviewed with a high sentiment subjectivity.

- When looking at the best price point for a medical app we can see that the top installed apps are at an avg of 70 dollars while the avg of all paid apps is 11. So a price in between these would be suitable for good market behavior.

- So, from our analysis we suggest that the client must be aware of the above points while looking forward to creating a medical app.

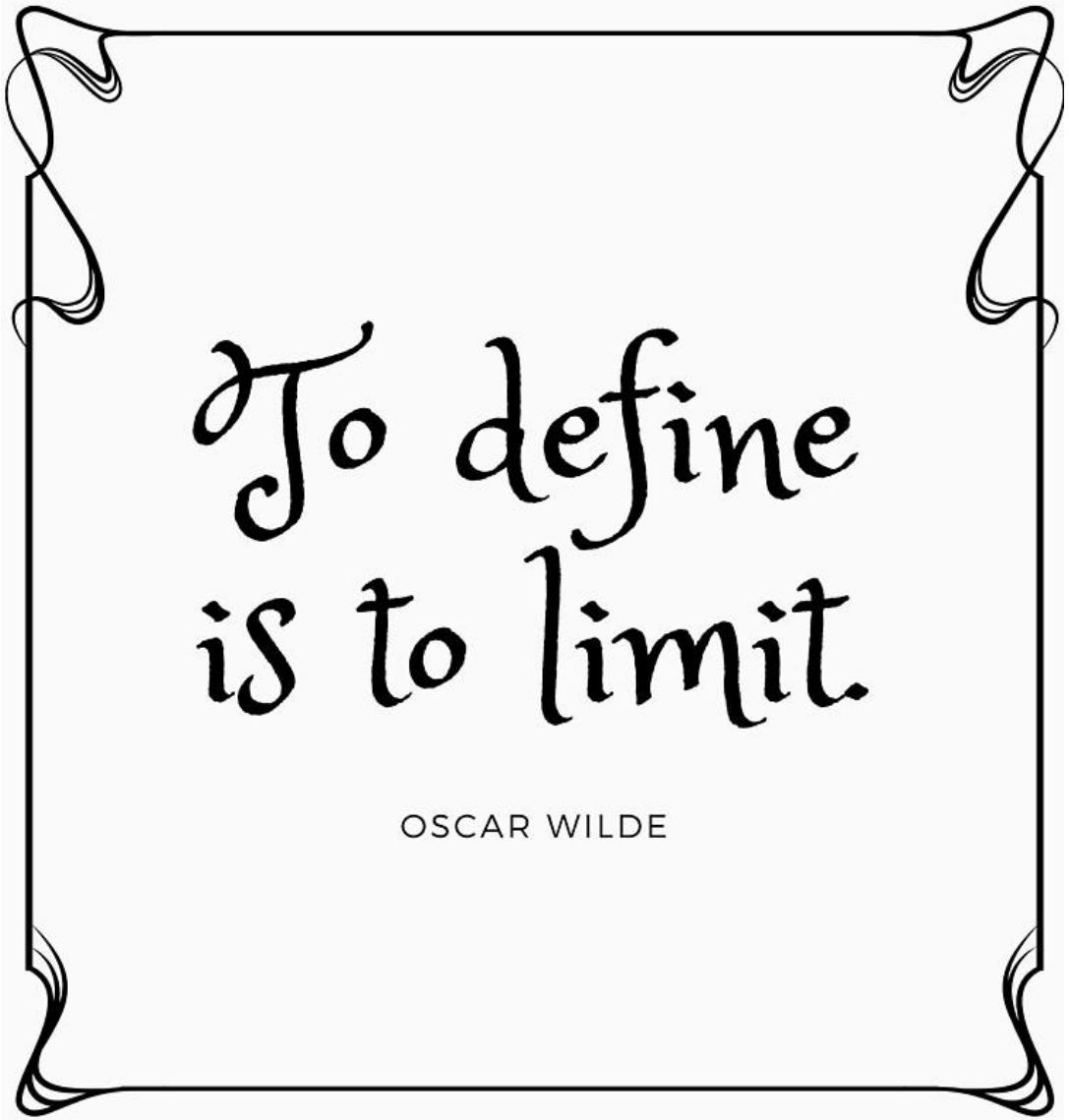
References

Ahmad, N., & Redzuan, N. B. (2020). Exploratory data analysis of Google Play Store mobile applications ratings and reviews. Journal of Physics: Conference Series, 1529(1), 012038. doi: 10.1088/1742-6596/1529/1/012038

Google Play Console Help. (n.d.). Analyze your Google Play data. Retrieved January 10, 2023, from https://support.google.com/googleplay/android-developer/answer/6083203?hl=en

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

The state of mobile 2020 (2020). App Annie.*URL https://www. appannie. com/en/insights/market-data/state-ofmobile-2020/. Last accessed November*.



THANK YOU!