

Capstone Project

Play Store App Review Analysis

by
Girish R

Content

- Overview
- Problem Statement
- Data Summary
- Exploratory Data analysis
- Summary and Conclusion



Overview



With over one million apps available on the Google Play Store, app developers must work hard to create apps that stand out from the crowd. Businesses must identify key details that influence user decisions in order to stand out in this oversaturated market.

The goal of this project is to provide insights that will help developers better understand the app market and user needs, and thus develop the product based on the data insights to get the most out of the app.

Problem Statement

- *Saturation begins with availability, which we can determine by looking for categories with the most and least number of apps.*
- *Identifies the key factors that influence user decision making and proposes a relationship between app metrics.*
- *What are the best business models to choose from, and why are they the best choice for any given segment or category?*
- *Dive into the user review sentiments and follow up on their positivity, negativity, and neutrality in the best way possible to aid decision-making.*

Data Summary

> Dataset shape –

10841 Rows and 13 columns

> Exploring Columns-

- ✓ **App** - *Given the Official name of each application*
- ✓ **Category** - *These are group in which applications were divided (Presented 33 unique categories)*
- ✓ **Rating** - *Ratings are number varies from 0 – 5.0 calculated by play store based on ratings given by every unique user.*
- ✓ **Reviews**- *Number of Reviews given by every unique user who installed the app on their device.*
- ✓ **Size**- *Size of the particular app and is in the format of Mb and Kb .*

Data Summary (continue)

- ✓ **Installs** - *Number of App installs/downloads by users, also in millions.*
- ✓ **Type** – *Type refers to the business model, Free or Paid, given for each app.*
- ✓ **Price** – *Price of the app is provided if it's a Paid app else it's mentioned 0.0*
- ✓ **Genres** – *Genres are broader segments and an app can be included in multiple genres depending on the content it provides.*
- ✓ **Sentiment** – *Sentiments refers to the emotions, which are calculated based on the written review given by an user, they are of 3 types-'Positive', 'Negative' and 'Neutral'.*
- ✓ **Others** – *Some other columns such as 'Android ver', 'Current ver', 'Last updated', 'Content Rating' also presented, but they are not so useful in the process of decision making, so we have not used them.*

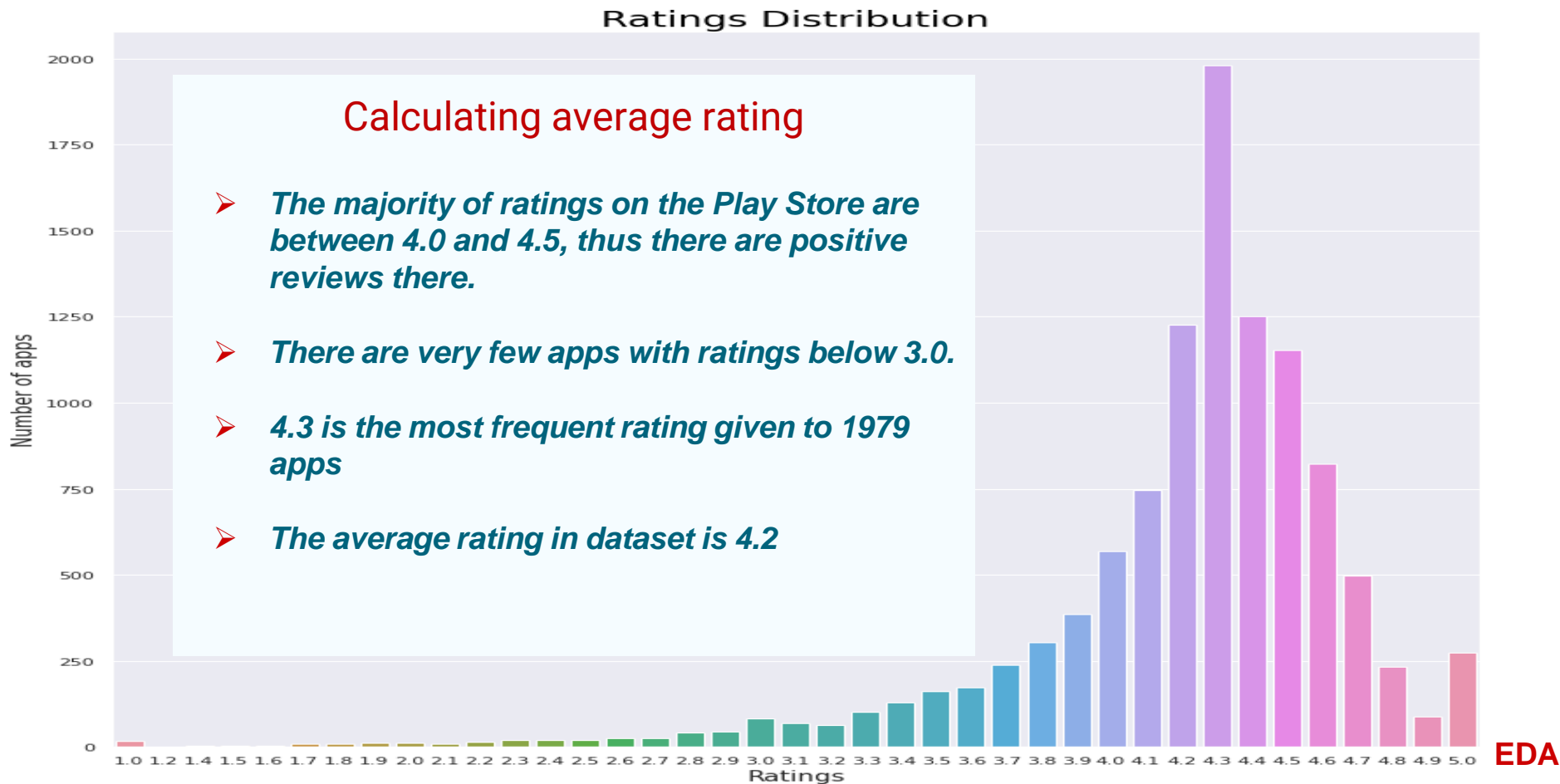
Exploratory Data Analysis

❖ The Data Cleaning Process in Four Steps

- Checking for outliers in important columns for analysis
- Dropping the outliers (if any)
- Taking care of Null Values
- Converting column data types for further use

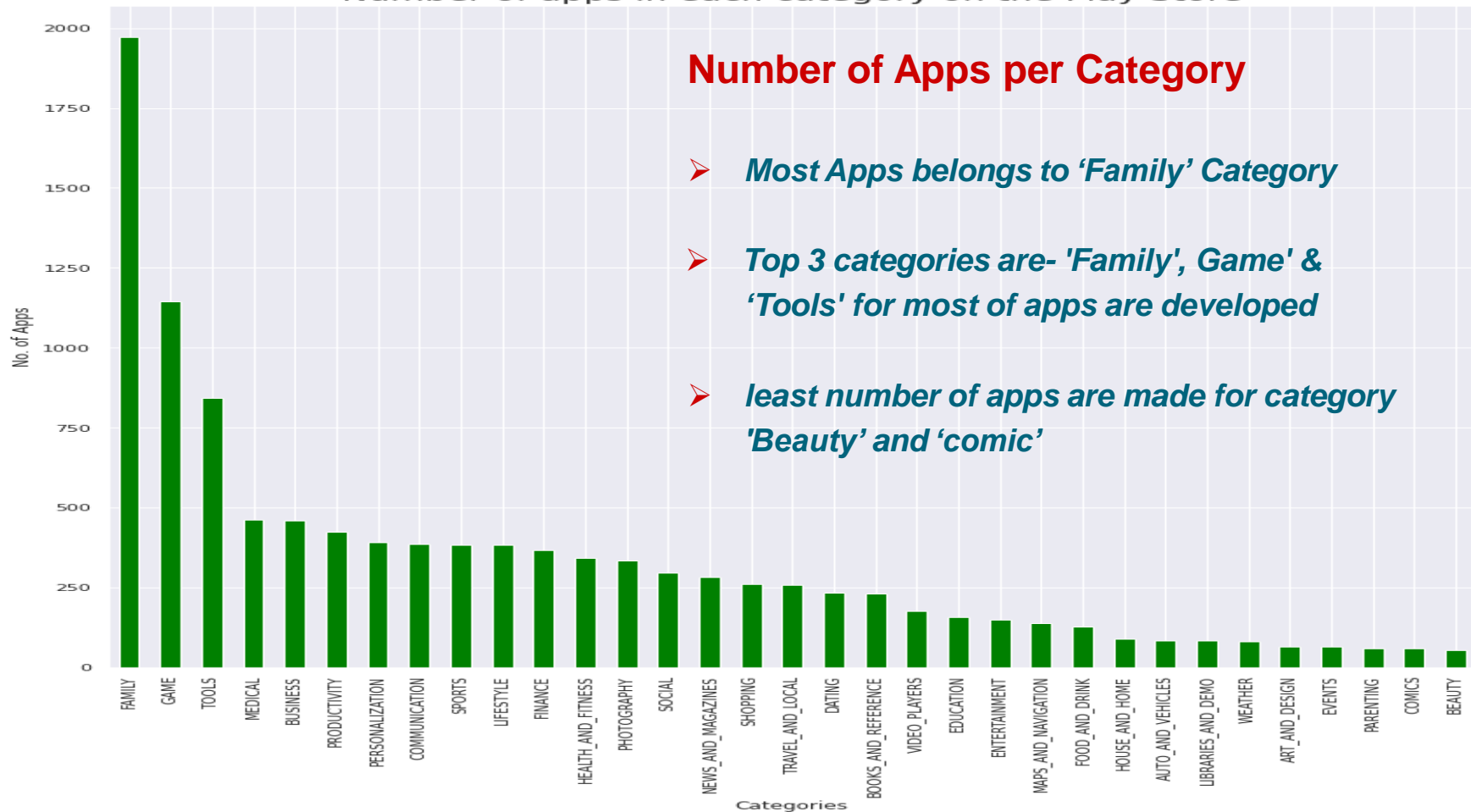


Measuring the amount of ratings that apps have received



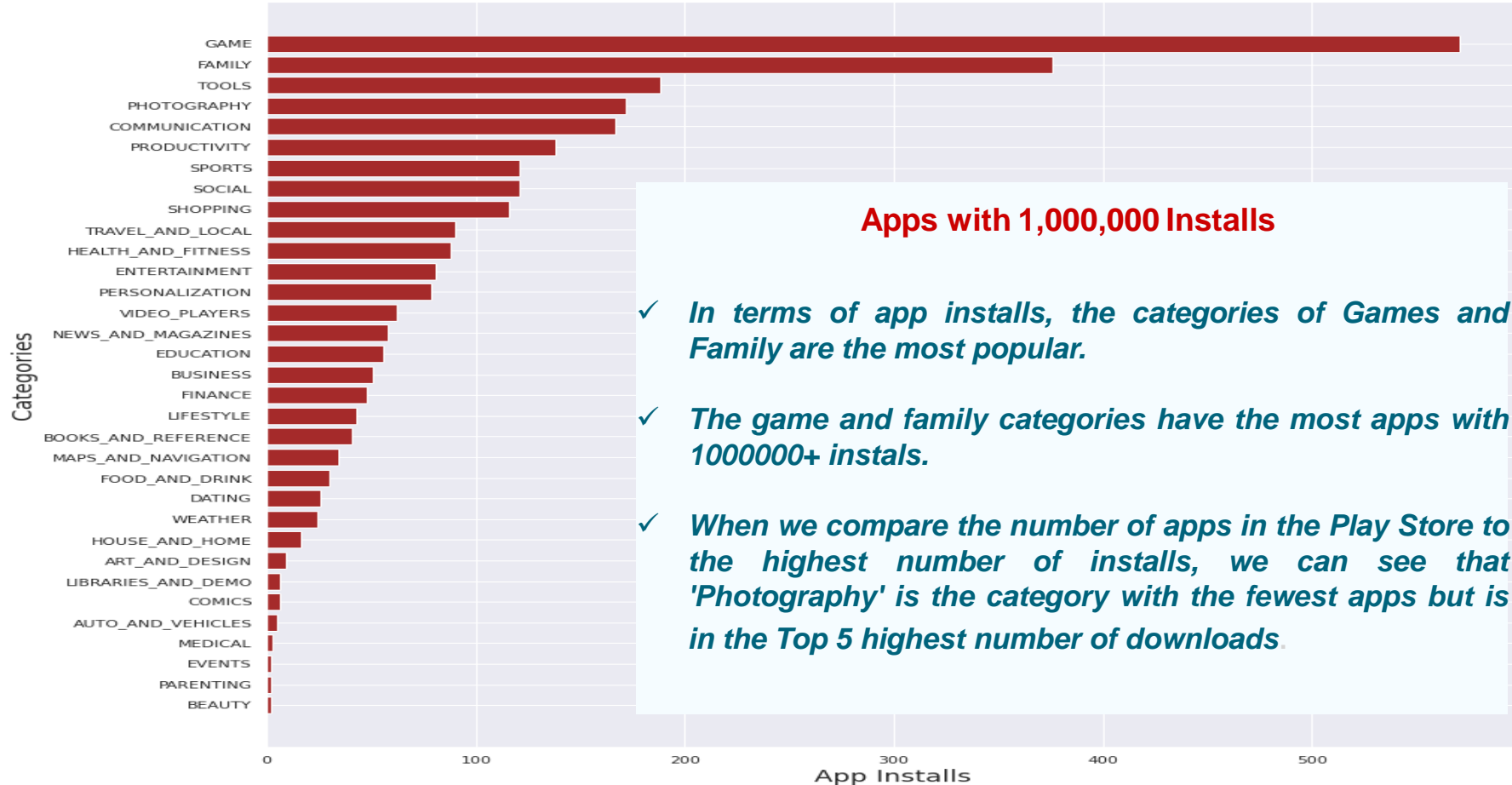
The category with the most apps in the Google Play Store

Number of apps in each category on the Play Store



Top Categories with over 1 million app installs

Apps per category with greater than 1,000,000 Installs

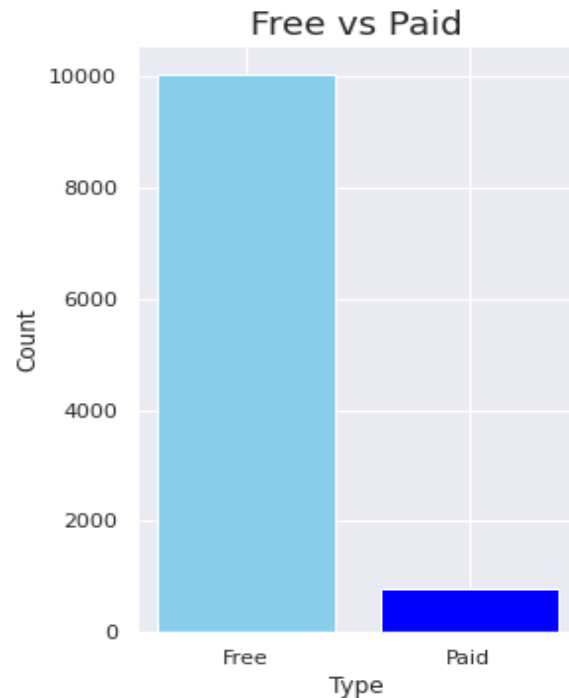


Apps with 1,000,000 Installs

- ✓ *In terms of app installs, the categories of Games and Family are the most popular.*
- ✓ *The game and family categories have the most apps with 1000000+ instals.*
- ✓ *When we compare the number of apps in the Play Store to the highest number of installs, we can see that 'Photography' is the category with the fewest apps but is in the Top 5 highest number of downloads.*

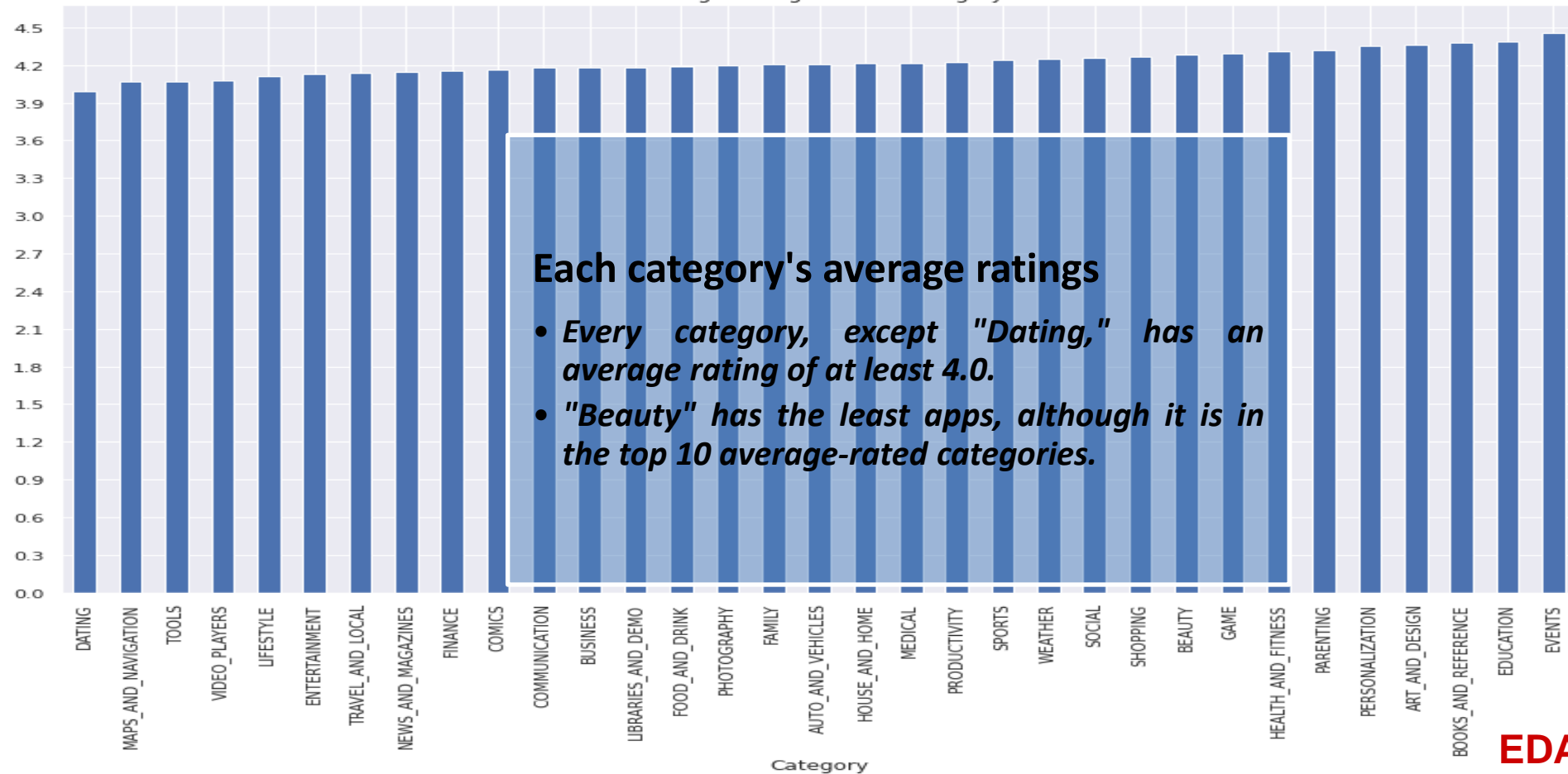
Paid vs Free apps

- ❖ This graph compares the number of apps in the "free" and "paid" categories that are available. This will make it easier for us to comprehend the dataset's distribution of free and premium apps.
- ❖ From the graph we can see that Free app is more compared to Premium app



Each category's average ratings

Average Ratings of each category



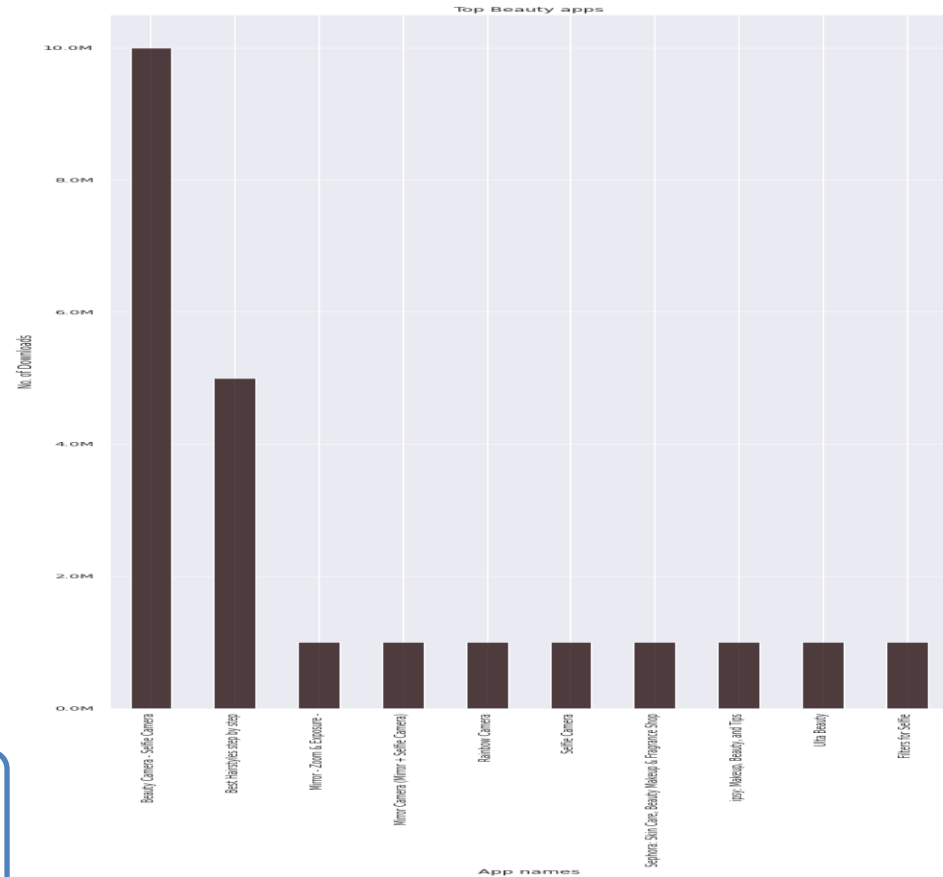
Analyzing some beauty-related apps since that is the least-competitive category

Top Beauty Apps

Beauty apps are the most downloaded apps.

Two beauty apps have reached 5 million and 10 million downloads, respectively.

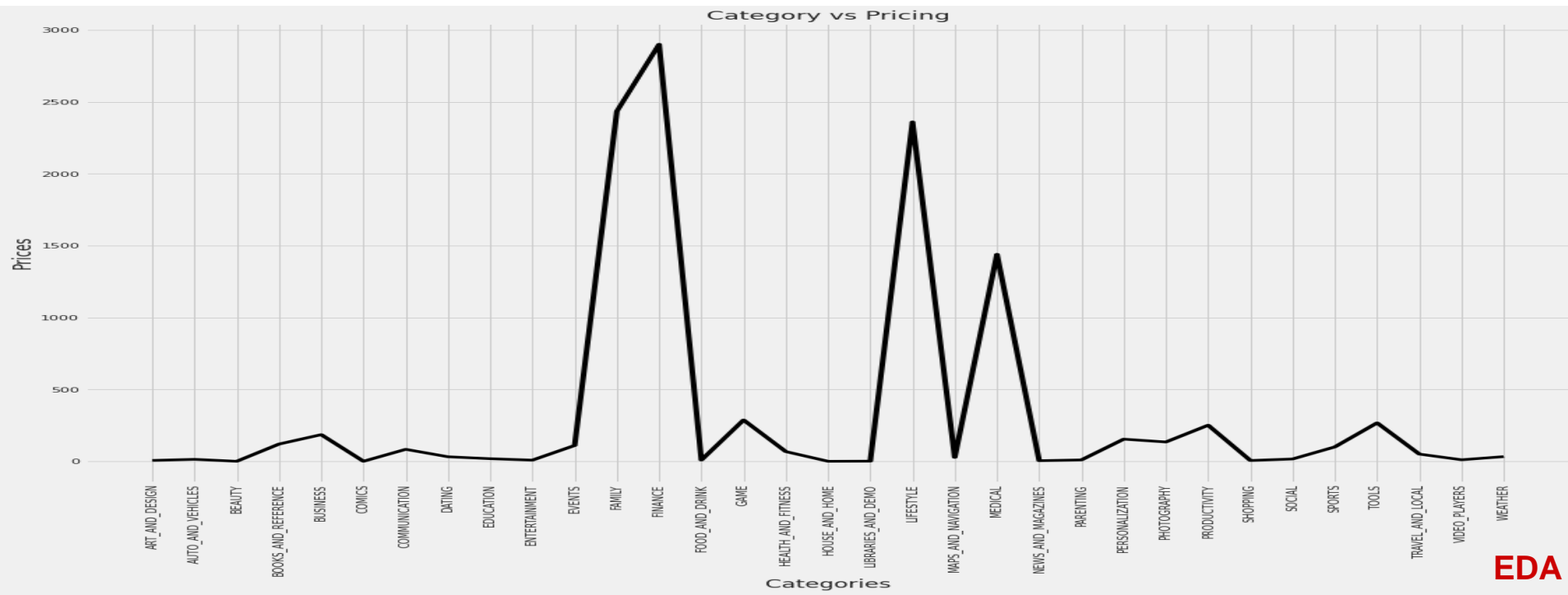
This shows a lot of potential in the beauty category due to low competition and high ratings with high app downloads.



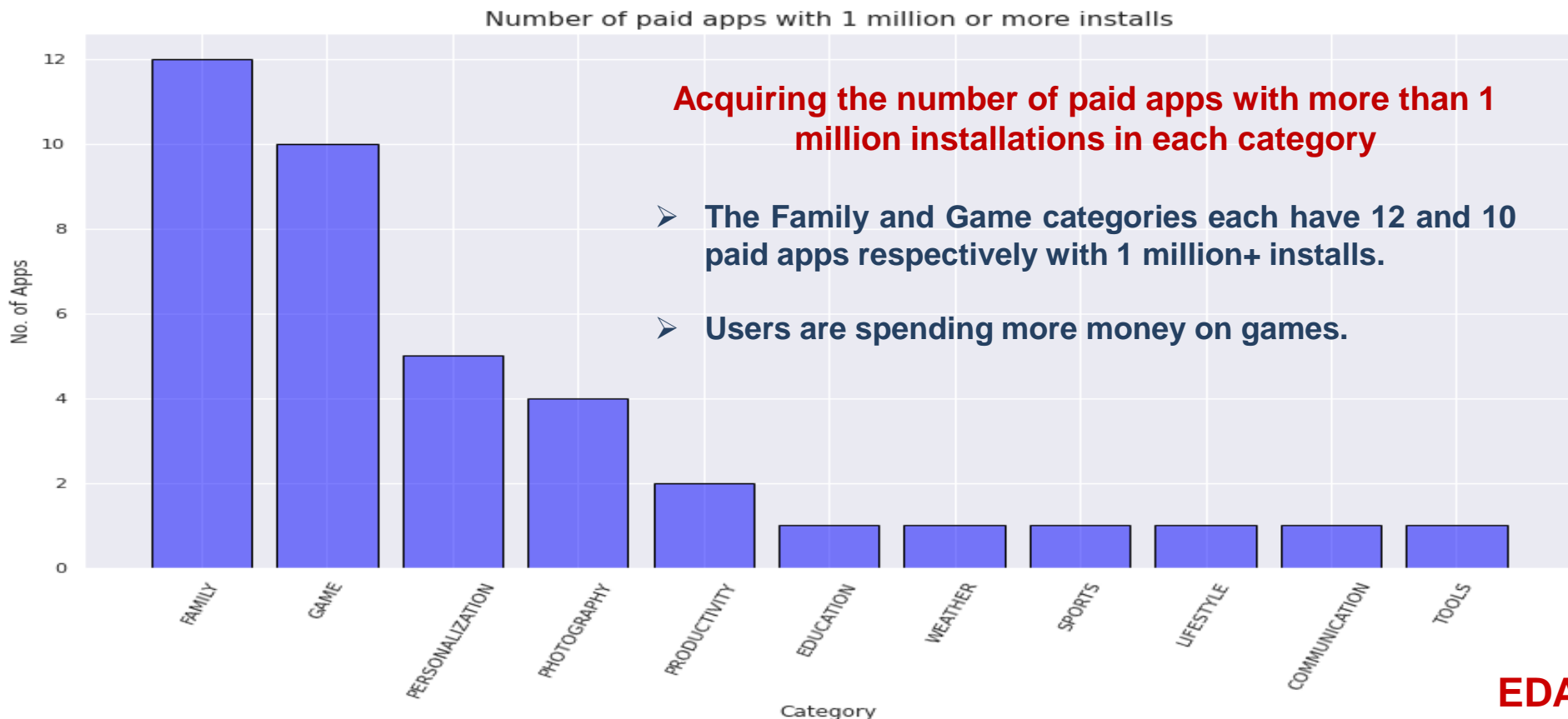
Exploring Paid Apps

Category vs Pricing

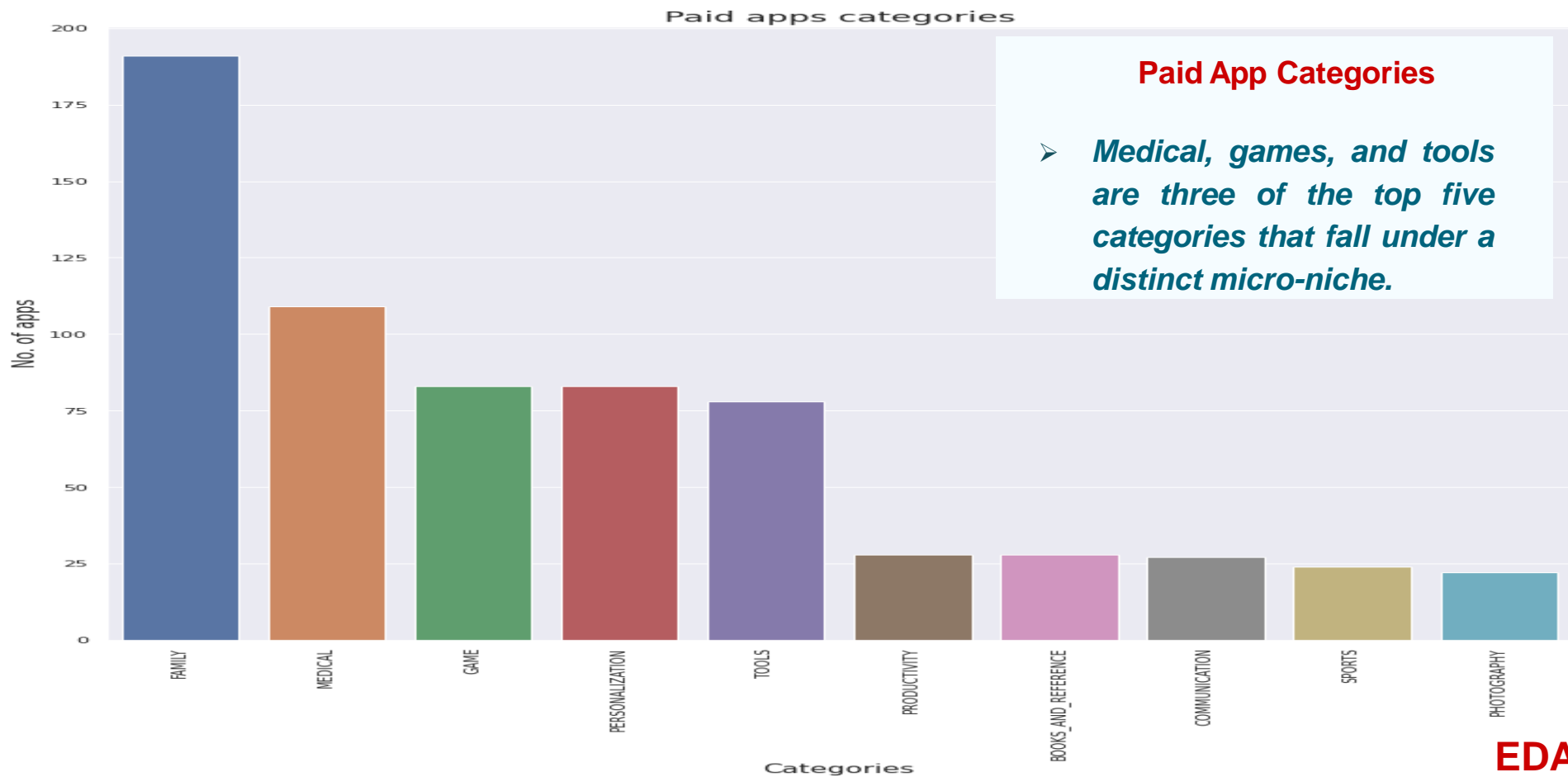
❑ In this graph we will know about which category has more paid apps



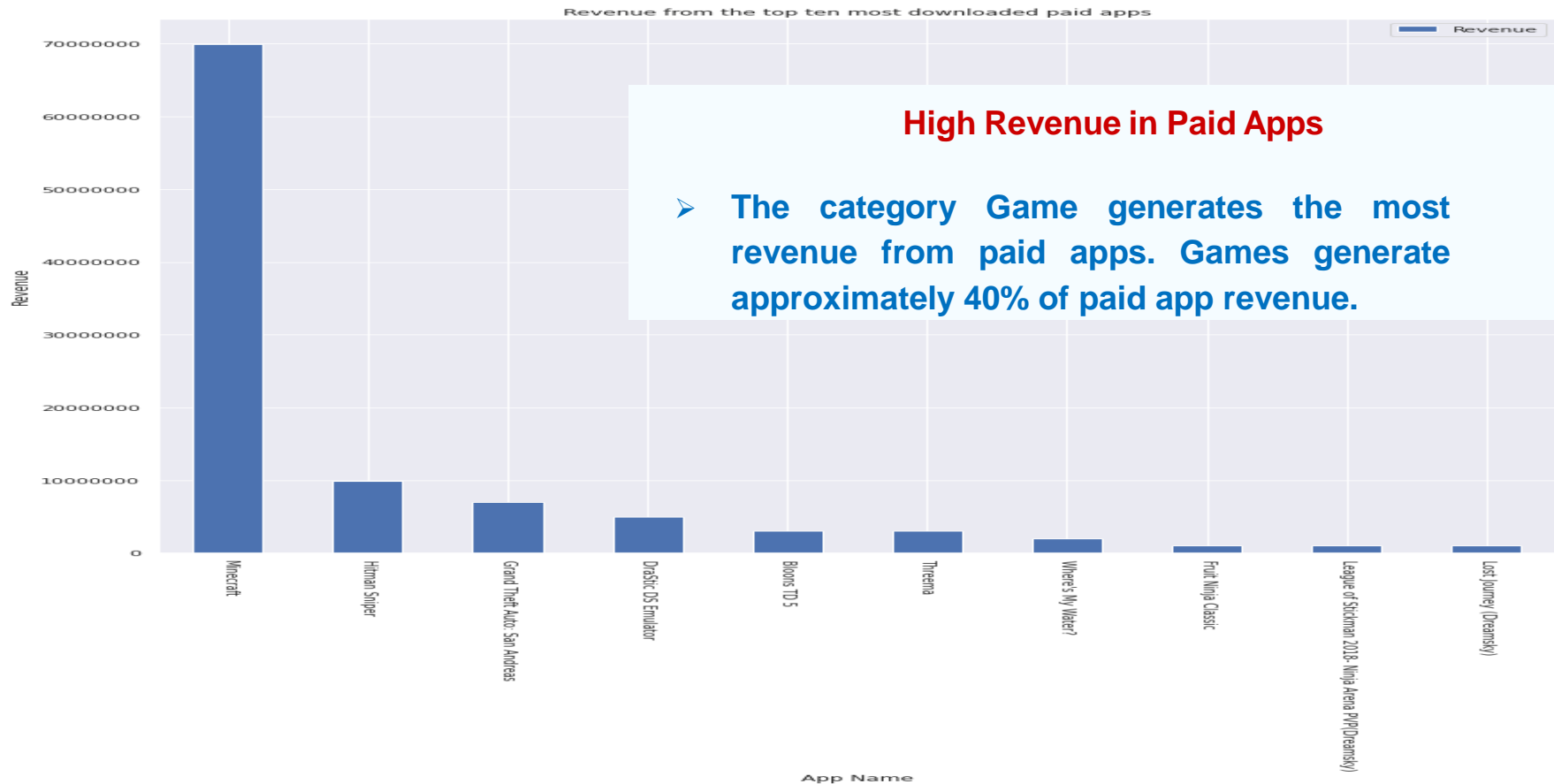
Acquiring the number of paid apps with more than 1 million installations in each category



Top ten categories with the highest number of paid apps AI



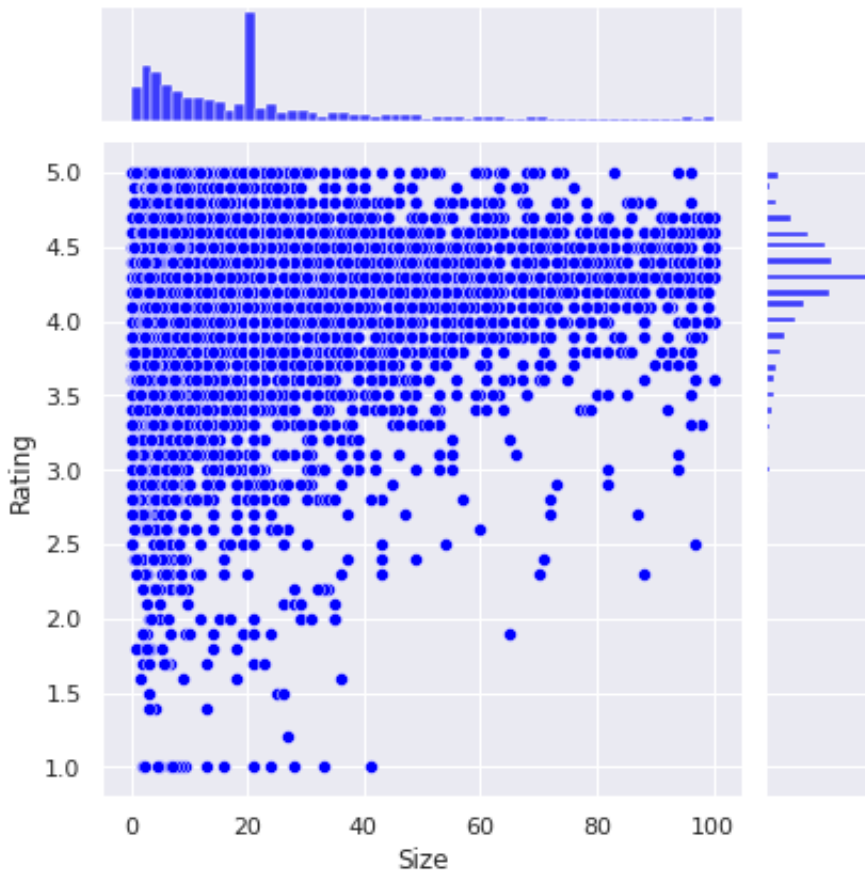
Top ten paid apps with the highest income



Assessing the link between App Size and Ratings

Assessing the link between App Size and Ratings

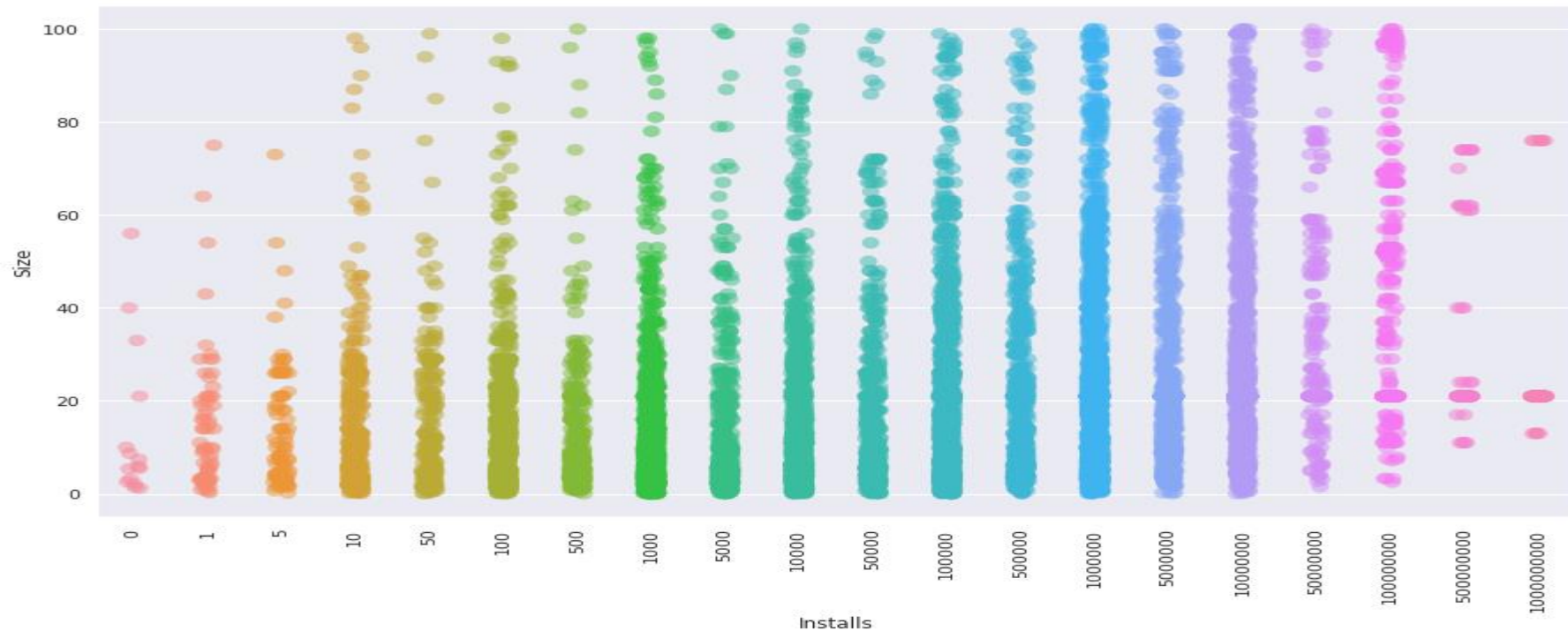
- *When sized between 0-60, the ratings are mostly between 3.7 and 4.8*
- *When the size increase from 60, the number of rating drops*



Examining the effect of size on app installation

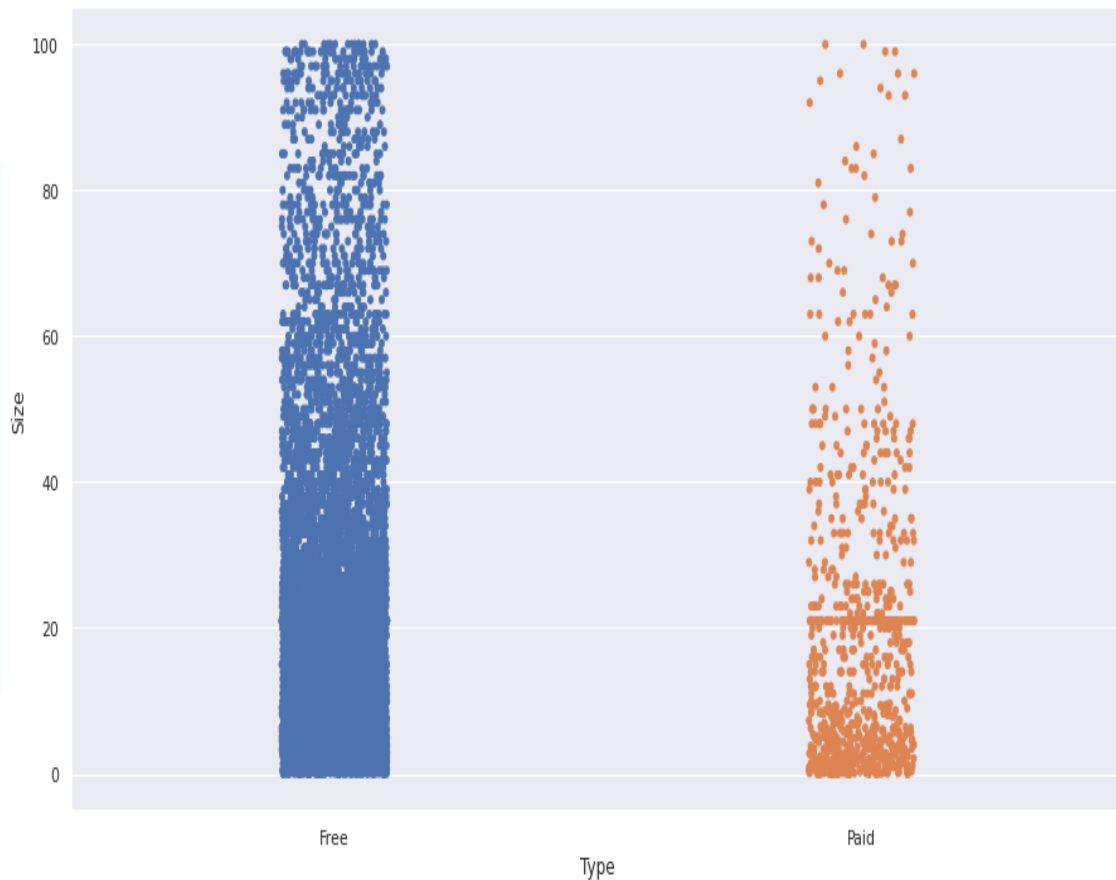
Examining the effect of size on app installation

- ✓ As we can see, the installation rate decreases as the app size increases above 60 MB, implying that size is an important factor for app installs, with users installing less frequently as the app size increases



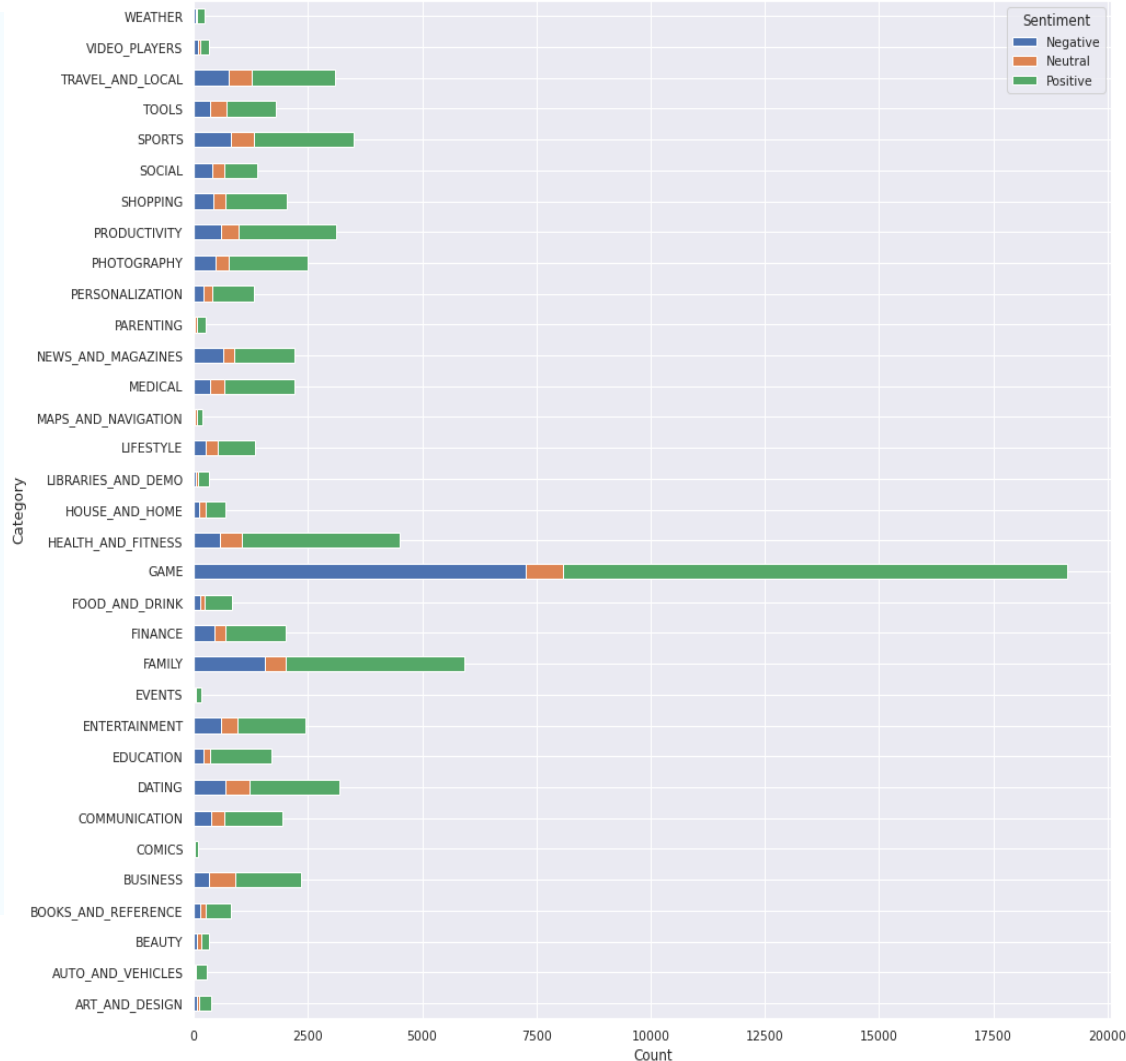
Size of App Analysis for Free and Paid Apps

❖ *Inferring from the graph that paid app developers charge for app efficiency while free app developers strive to reduce the cost of development by ignoring size efficiency, we may infer that the size of paid applications is relatively smaller than the size of free applications.*



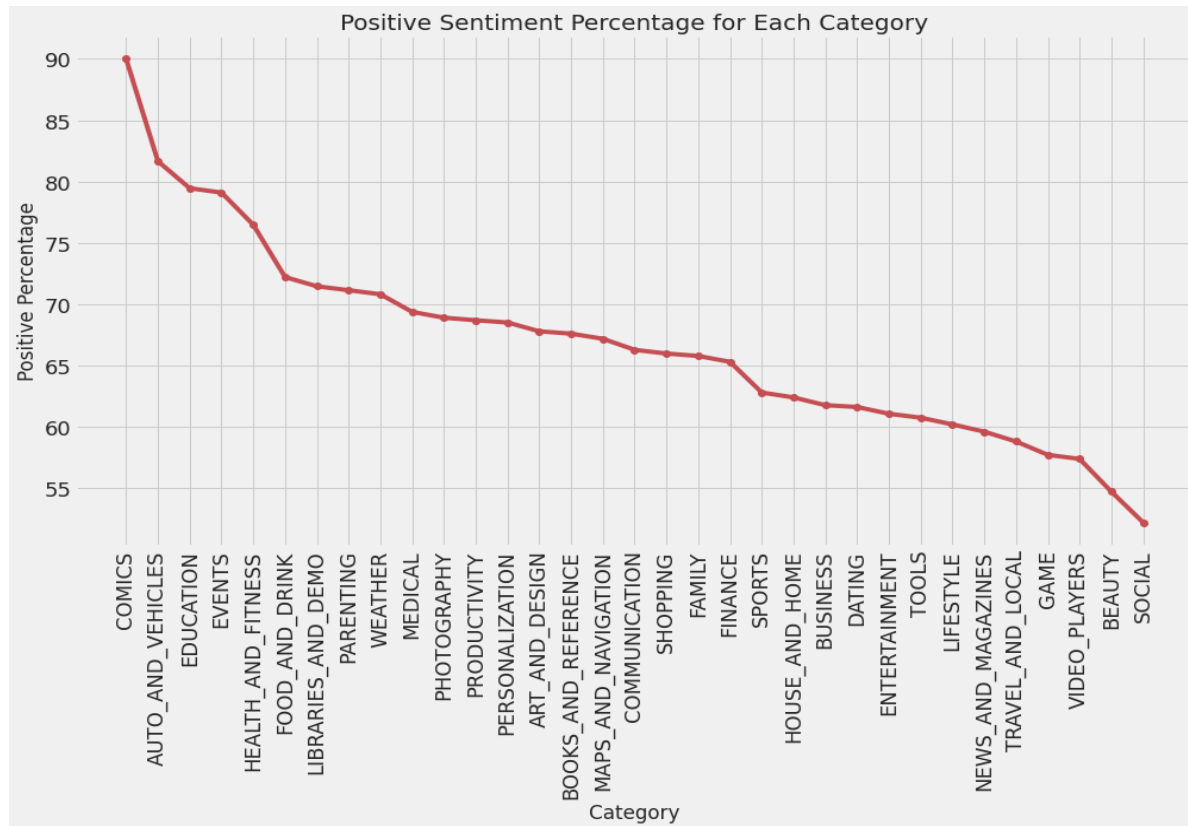
The number of positive, negative, and neutral emotions in each category is analyzed.

- Overall, positive sentiments are higher than negative and neutral sentiments in reviews, with gaming having the highest positive and negative sentiments at about 60% and 40%, respectively. All other categories also follow this pattern, with positive reviews being higher than negative reviews and neutral reviews being the lowest, leading us to the conclusion that no category is performing particularly negatively.



Calculating the positive sentiment percentage for each category from high to low

- The graphical representation shows that 5 categories have more than 75% of the total reviews that are favourable, indicating that these categories have already satisfied their customers. In contrast, the categories with lower levels of positivity have not yet achieved this level of customer satisfaction.



Summary and Conclusion

- ✓ *The "game" and "family" categories already rule the market, but "beauty," "comic," and "photography" are three low-competitive categories with a lot of promise that developers could concentrate on.*
- ✓ *Paid apps are not appropriate for general users; instead, they are designed for a certain niche or market sector and concentrate solely on that market segment to maximize earnings.*
- ✓ *When it comes to premium games, we've noticed a pattern: if a game is already well-liked and widely accessible across platforms, users will respond enthusiastically when it debuts in the Google Play store. For instance, consider Minecraft; GTA Vice City produced the most money.*
- ✓ *Because user behaviour is influenced by an app's size, it is advisable for developers to make their apps as small as possible.*

THANK YOU