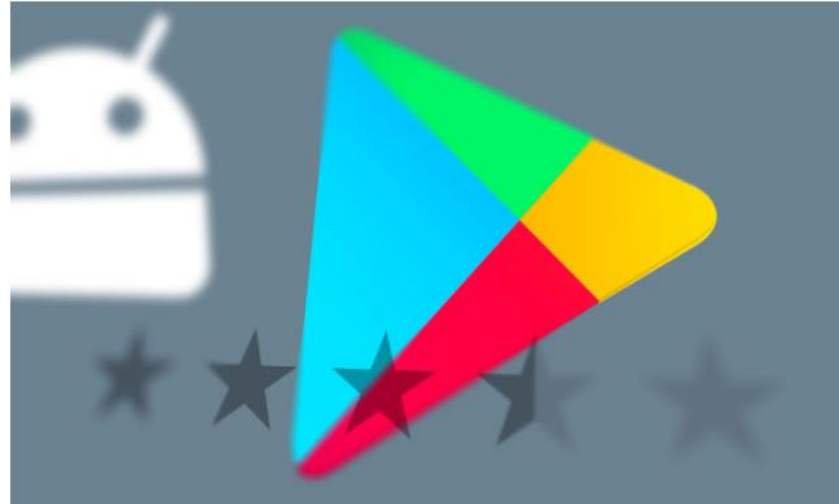# Capstone Project
Play Store Review Analysis
Ritik Gupta

Content

- Overview

- Problem Statement

- Data Summary

- Exploratory Data analysis

- Summary and Conclusion

Overview

With more than million apps accessible on the Google Play Store, app developers face a difficult task in creating apps that stand out from the crowd. To stand out in this oversaturated market, businesses must identify key details that influence users decisions.

The goal of this project is to provide insights that will help developers to better understand App market, user needs and hence develop the product based on the insights through the data and extract best out of the app.

Problem Statement

- *Saturation begins with availability and to find that we have to find categories with most and least number of apps.*

- *Finding the key factors that impact on user decision making and proposes relation between app metrices.*

- *What to choose from different type of business model and what makes them the best choice for any particular segment or category ?*

- *Dive and observe the user review sentiments and follow up with their positivity, negativity and neutrality in best way, that help in 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 .*
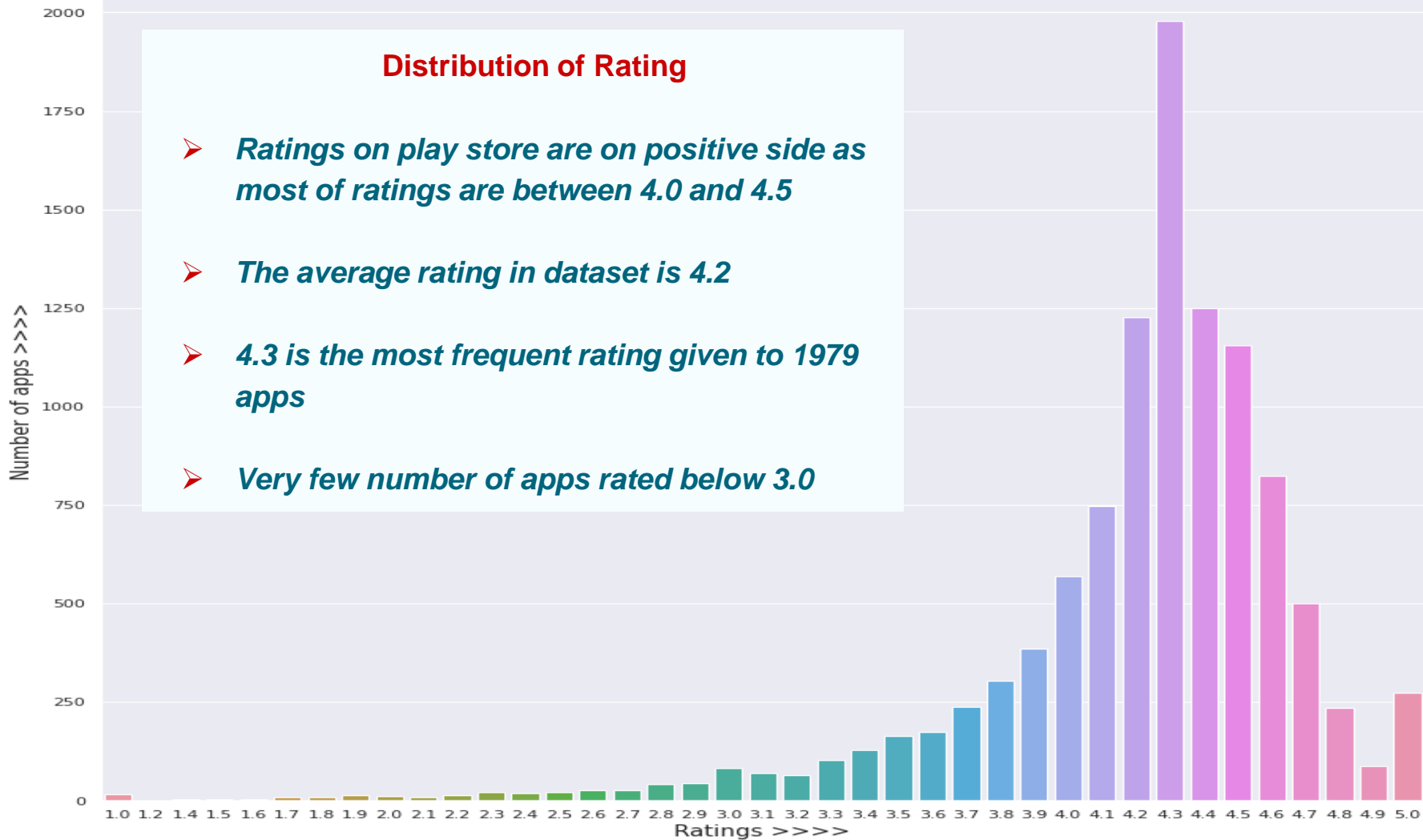
Data Summary (continue)

✓ **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.*
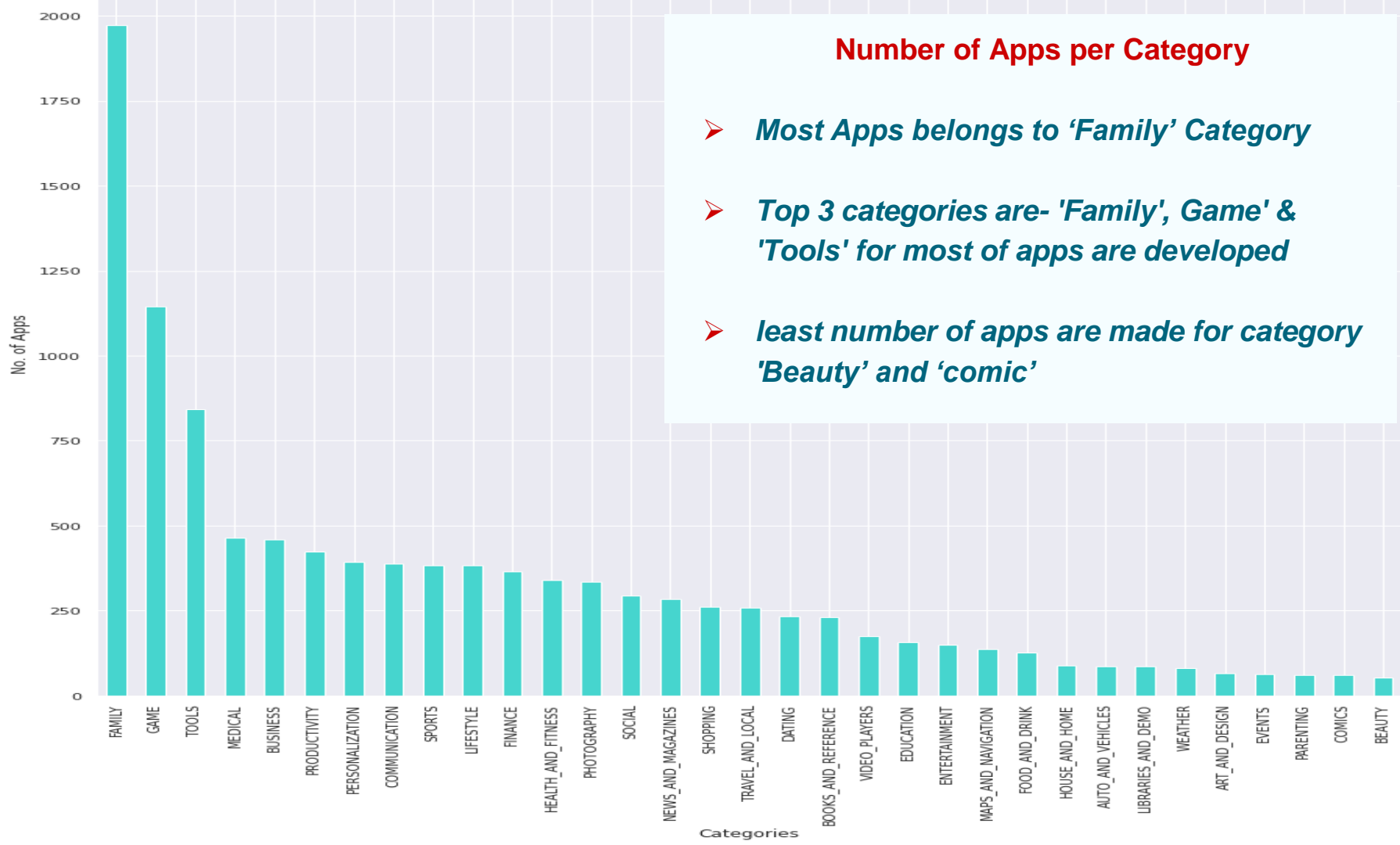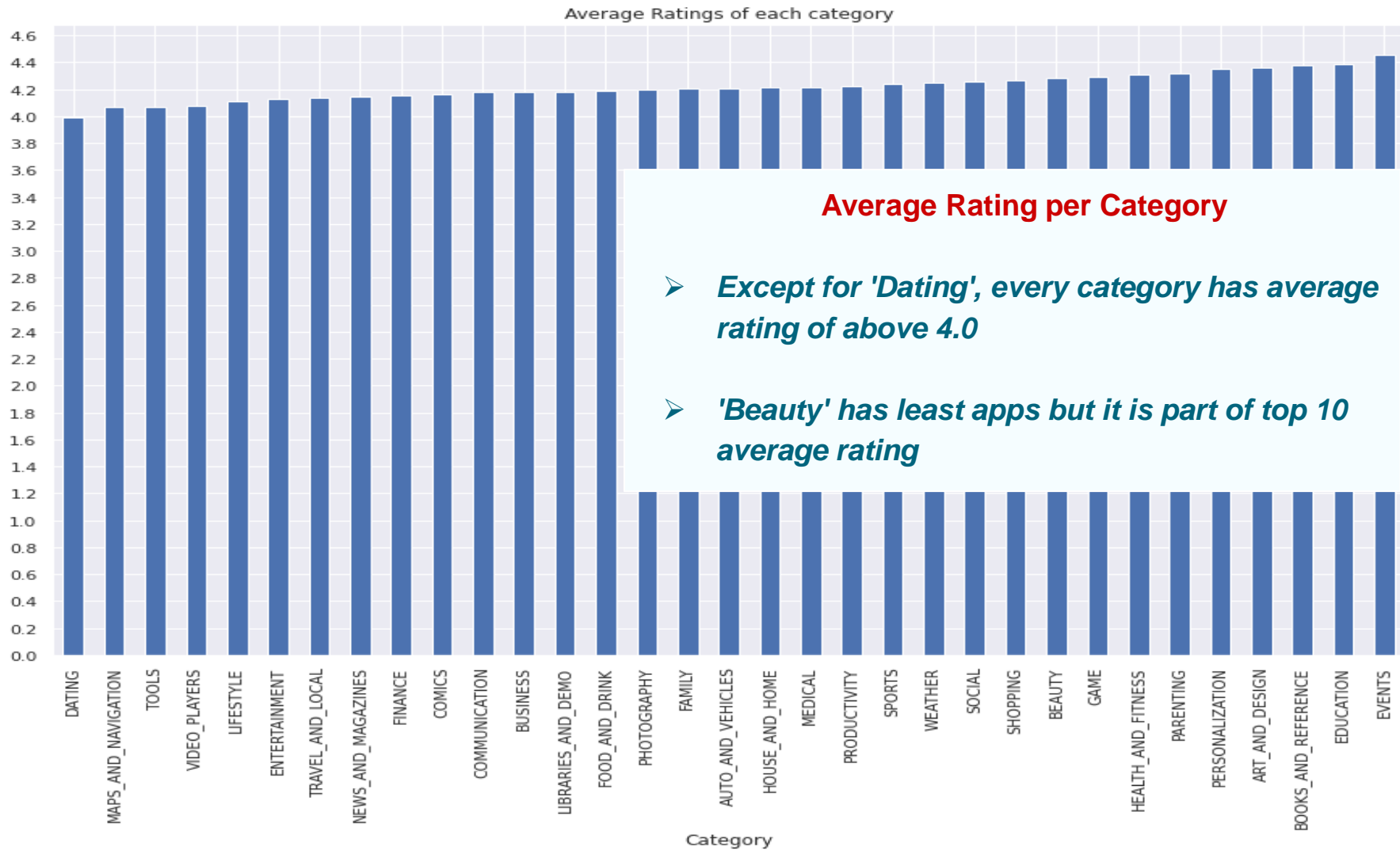
AI

Exploratory Data Analysis
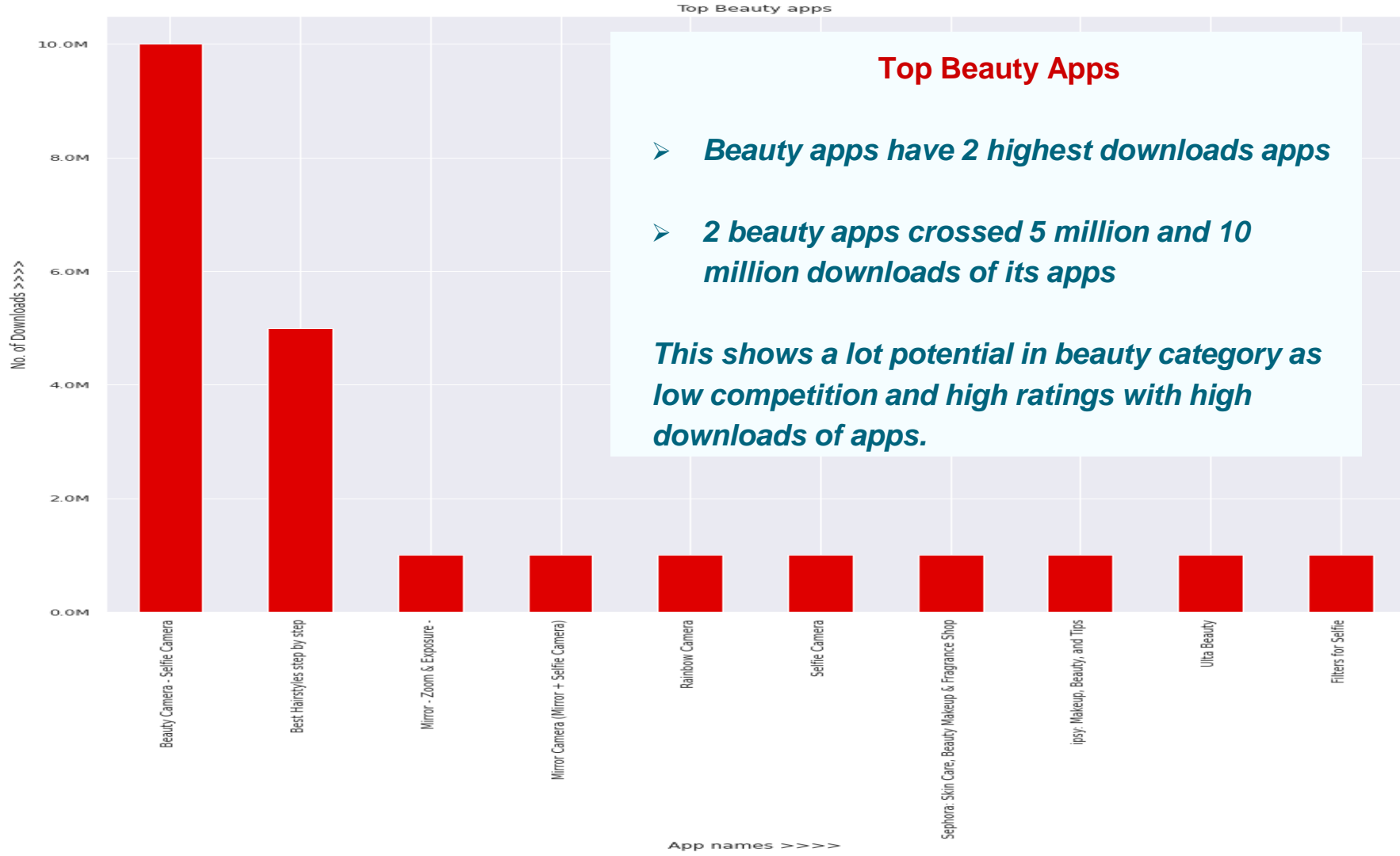
AI

❖ **4 Step Process of Data cleaning**

➢ Checking for outliers in important column with respect to analysis

➢ Dropping the outliers (if any)

➢ Taking care of Null Values

➢ Converting data type of columns for further operation

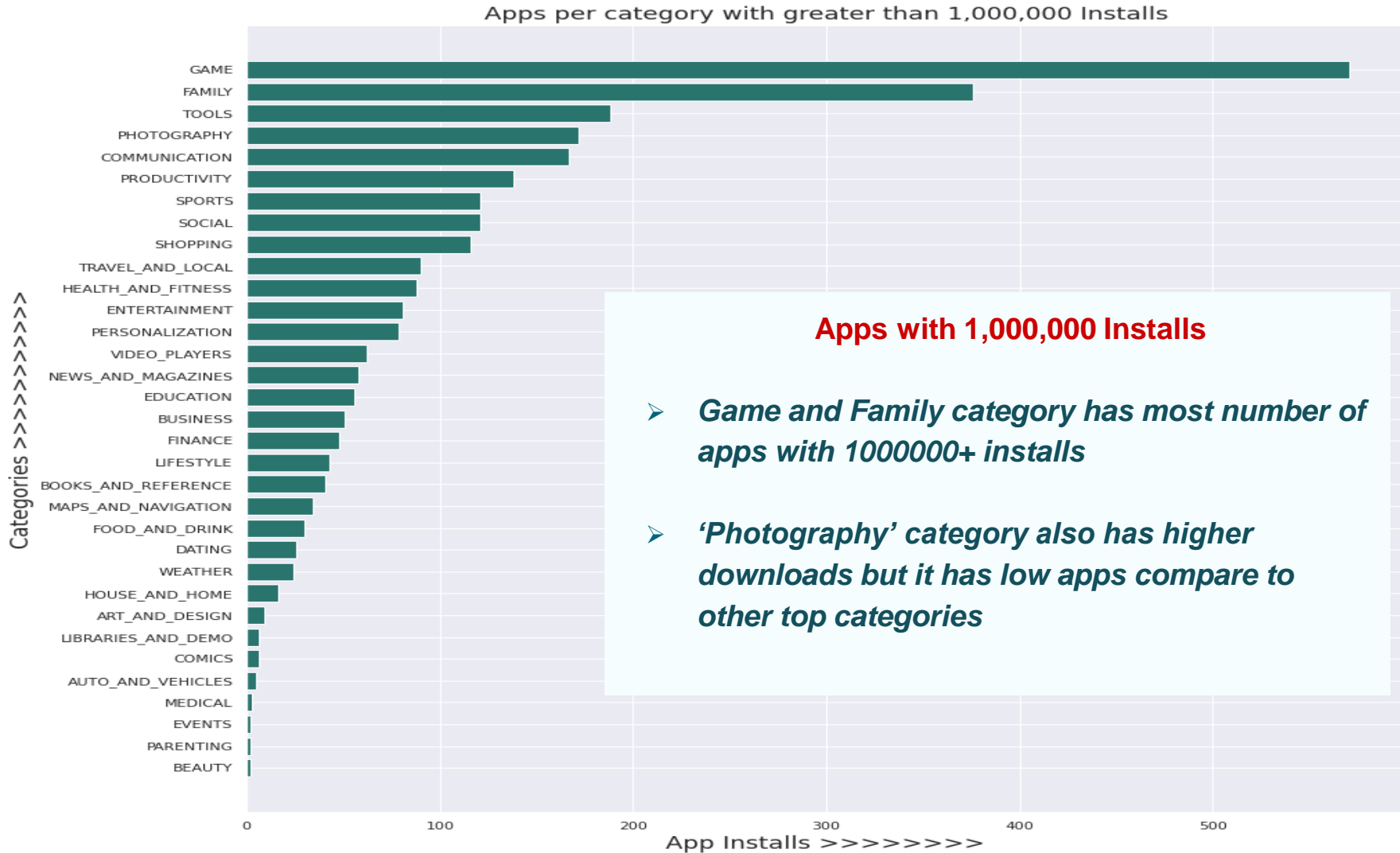**Distribution of Rating**

➤ *Ratings on play store are on positive side as most of ratings are between 4.0 and 4.5*

➤ *The average rating in dataset is 4.2*

➤ *4.3 is the most frequent rating given to 1979 apps*

➤ *Very few number of apps rated below 3.0*

Number of Apps per Category

➢ *Most Apps belongs to 'Family' Category*

➢ *Top 3 categories are- 'Family', Game' & 'Tools' for most of apps are developed*

➢ *least number of apps are made for category 'Beauty' and 'comic'*

EDA

Average Ratings of each category

Average Rating per Category

➢ *Except for 'Dating', every category has average rating of above 4.0*

➢ *'Beauty' has least apps but it is part of top 10 average rating*

AI

EDA

Top Beauty apps

Top Beauty Apps

➢ *Beauty apps have 2 highest downloads apps*

➢ *2 beauty apps crossed 5 million and 10 million downloads of its apps*

This shows a lot potential in beauty category as low competition and high ratings with high downloads of apps.

No. of Downloads >>>>

App names >>>>

AI

EDA

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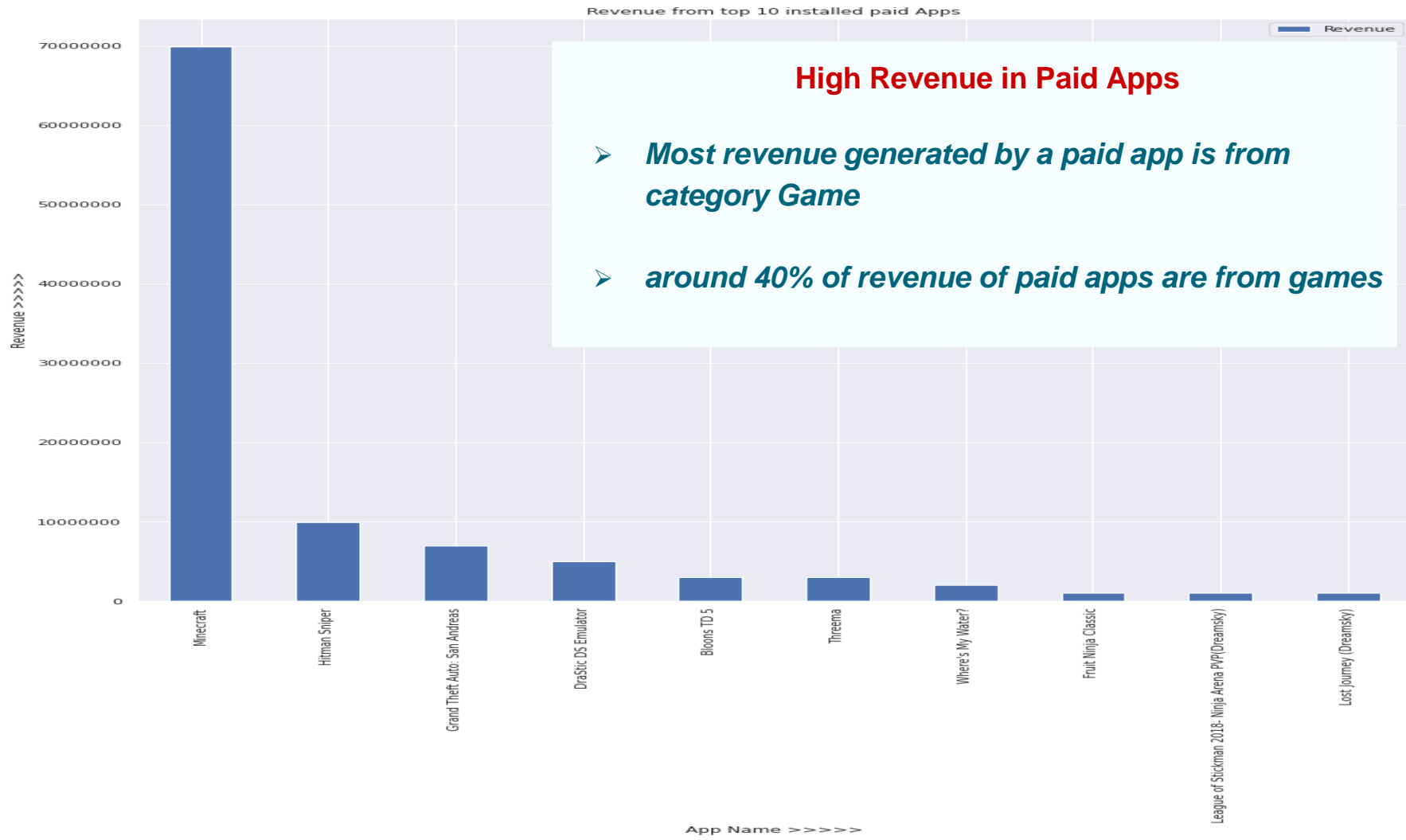
Apps with 1,000,000 Installs

➢ *Game and Family category has most number of apps with 1000000+ installs*

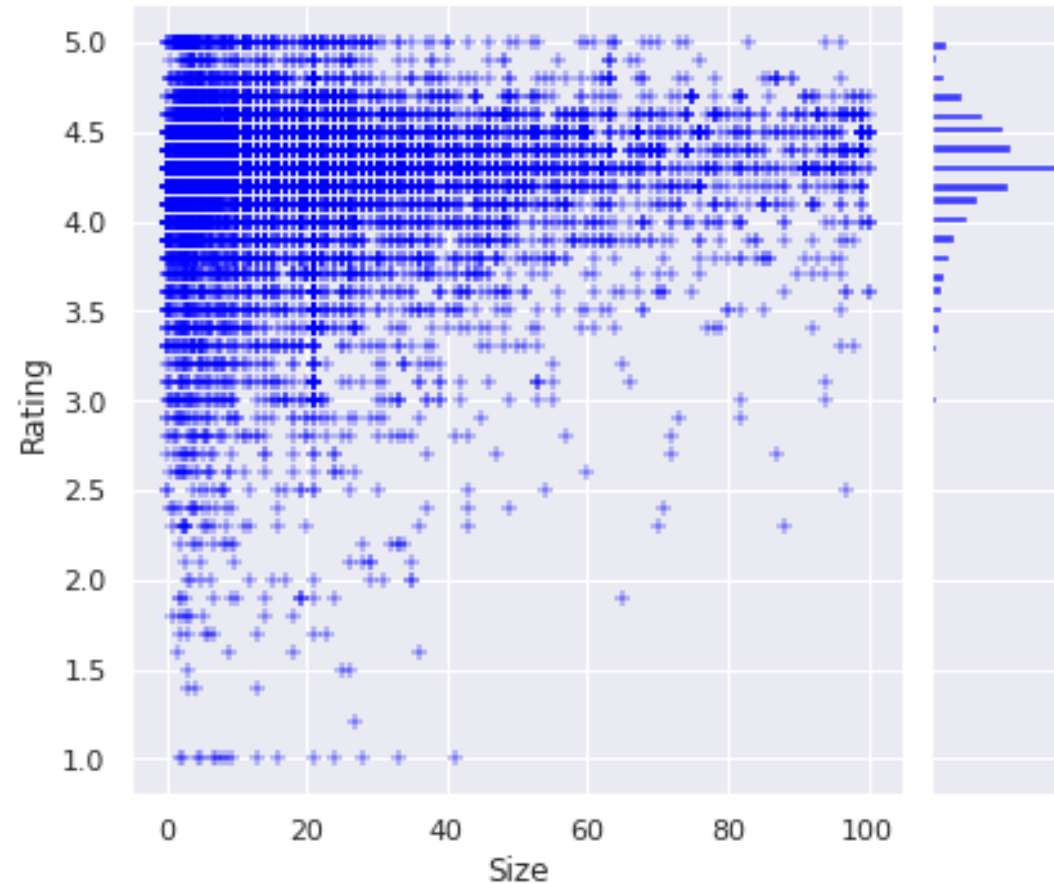➢ *'Photography' category also has higher downloads but it has low apps compare to other top categories*

AI

EDA

Paid apps categories

No. of apps >>>>

Categories >>>>

FAMILY, MEDICAL, GAME, PERSONALIZATION, TOOLS, PRODUCTIVITY, BOOKS_AND_REFERENCE, COMMUNICATION, SPORTS, PHOTOGRAPHY
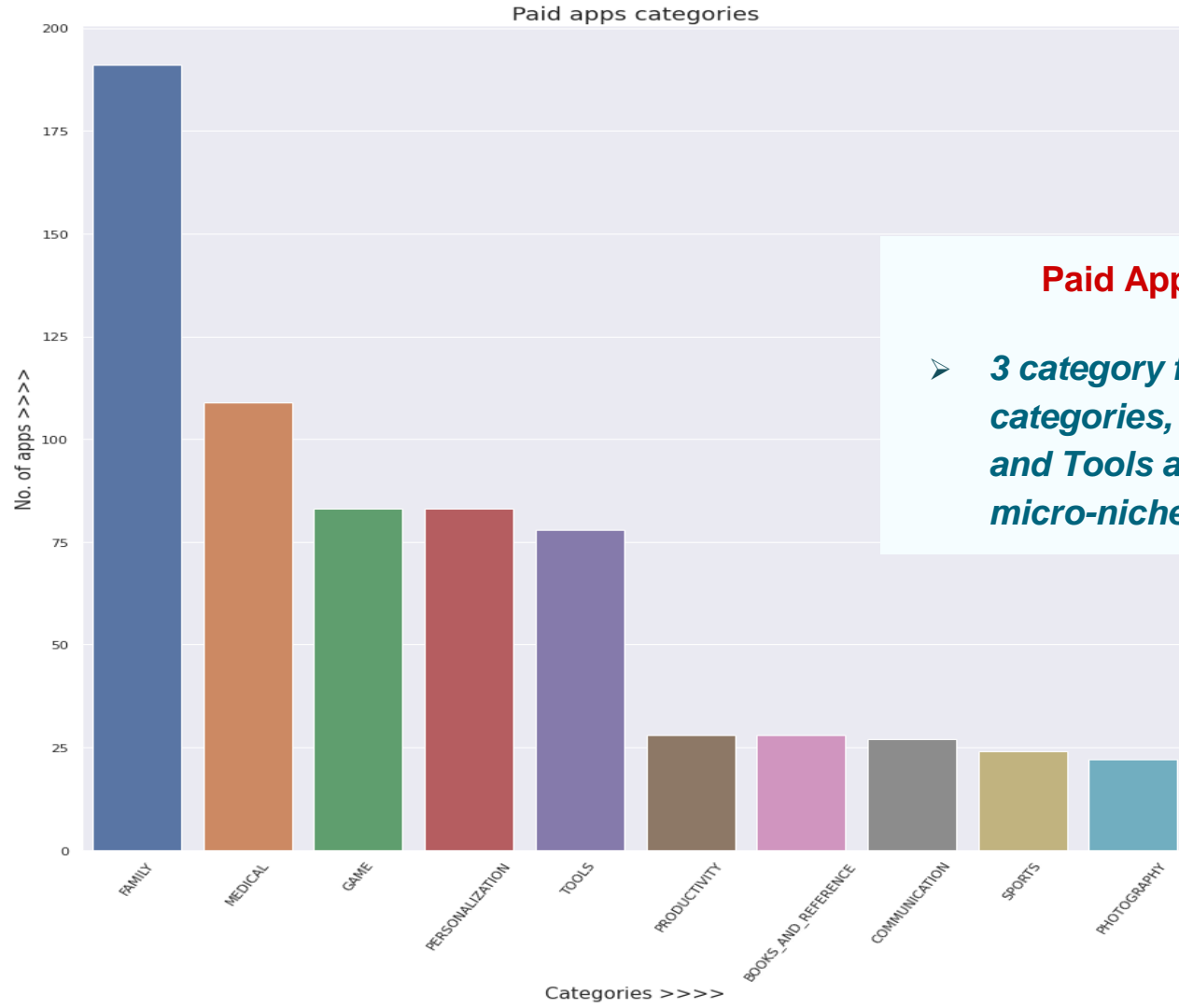
Paid App Categories

➢ *3 category from top 5 categories, Medical, Games and Tools are of specific micro-niche*

AI

EDA

Revenue from top 10 installed paid Apps

High Revenue in Paid Apps

➢ *Most revenue generated by a paid app is from category Game*

➢ *around 40% of revenue of paid apps are from games*

AI

EDA

Relation 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*

EDA
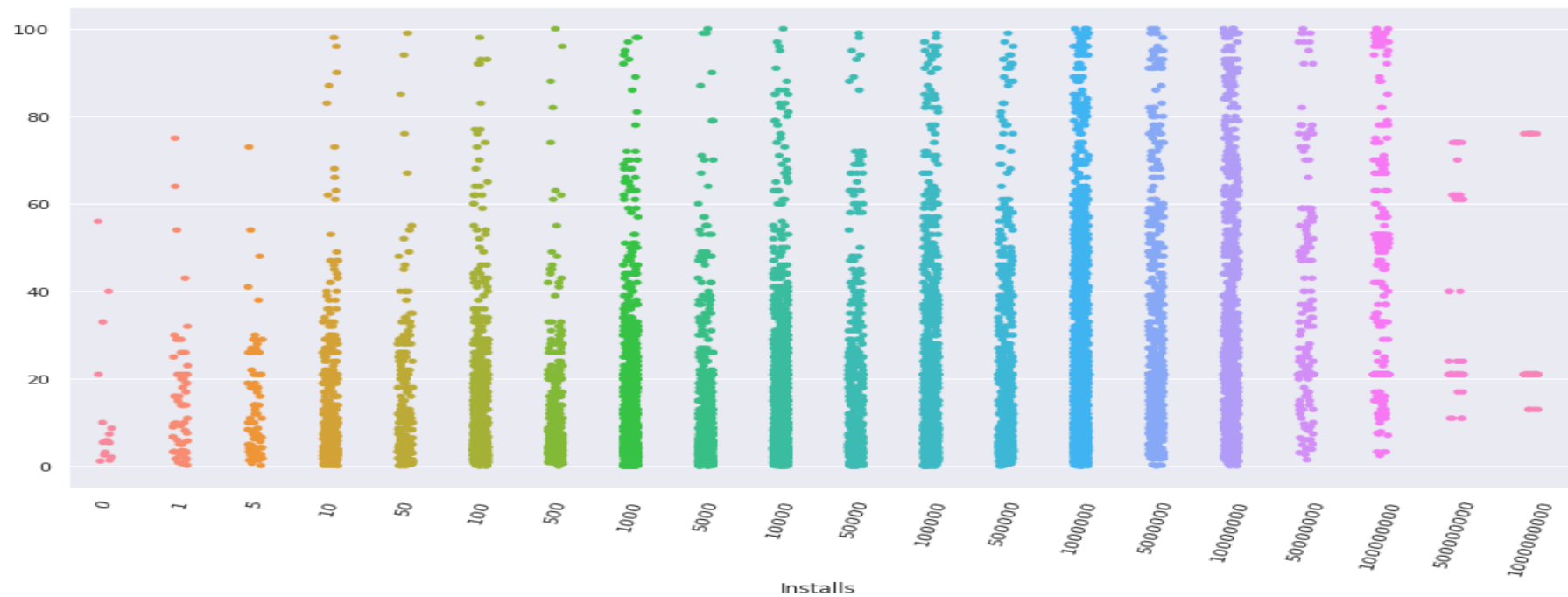
Relation Between App Size and Installs

➢ *The installation are decreasing as going to above 60 MB of size*
➢ *we can conclude that size is a key factor for app installs, users installs became lower in case of high app size.*

EDA

Size of Free and Paid Apps

The size of paid apps are comparatively lower then Size of free apps, it may because of paid apps developers charging for efficiency of apps and free app developers wants to cut the cost of development by neglecting size efficiency.

EDA
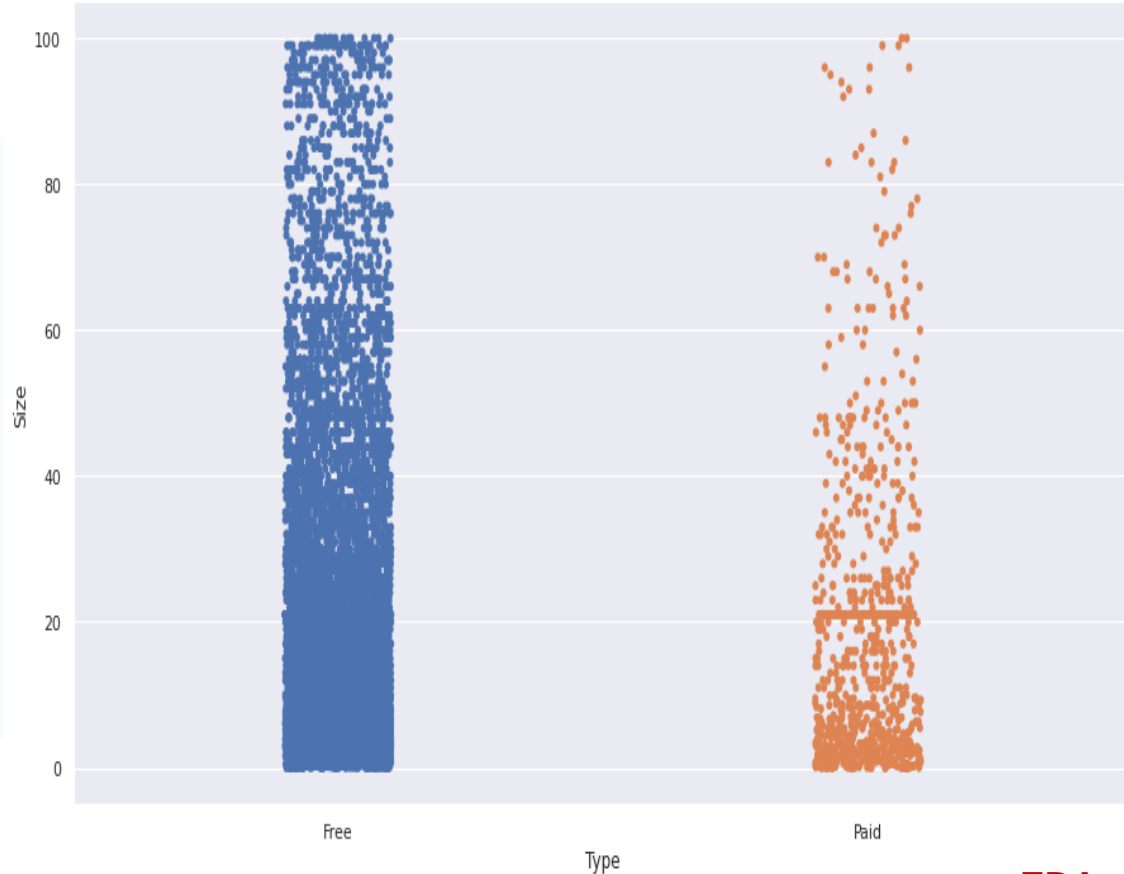
Sentiments per Category

➢ *Positive and Negative sentiments are more and Neutral are least*

➢ *'Game' has most number of positive as well as negative reviews*

➢ *No other category has high number of Negative reviews so we can say no category is performing negative particularly*

EDA

Percentage of Positive sentiments for each Category
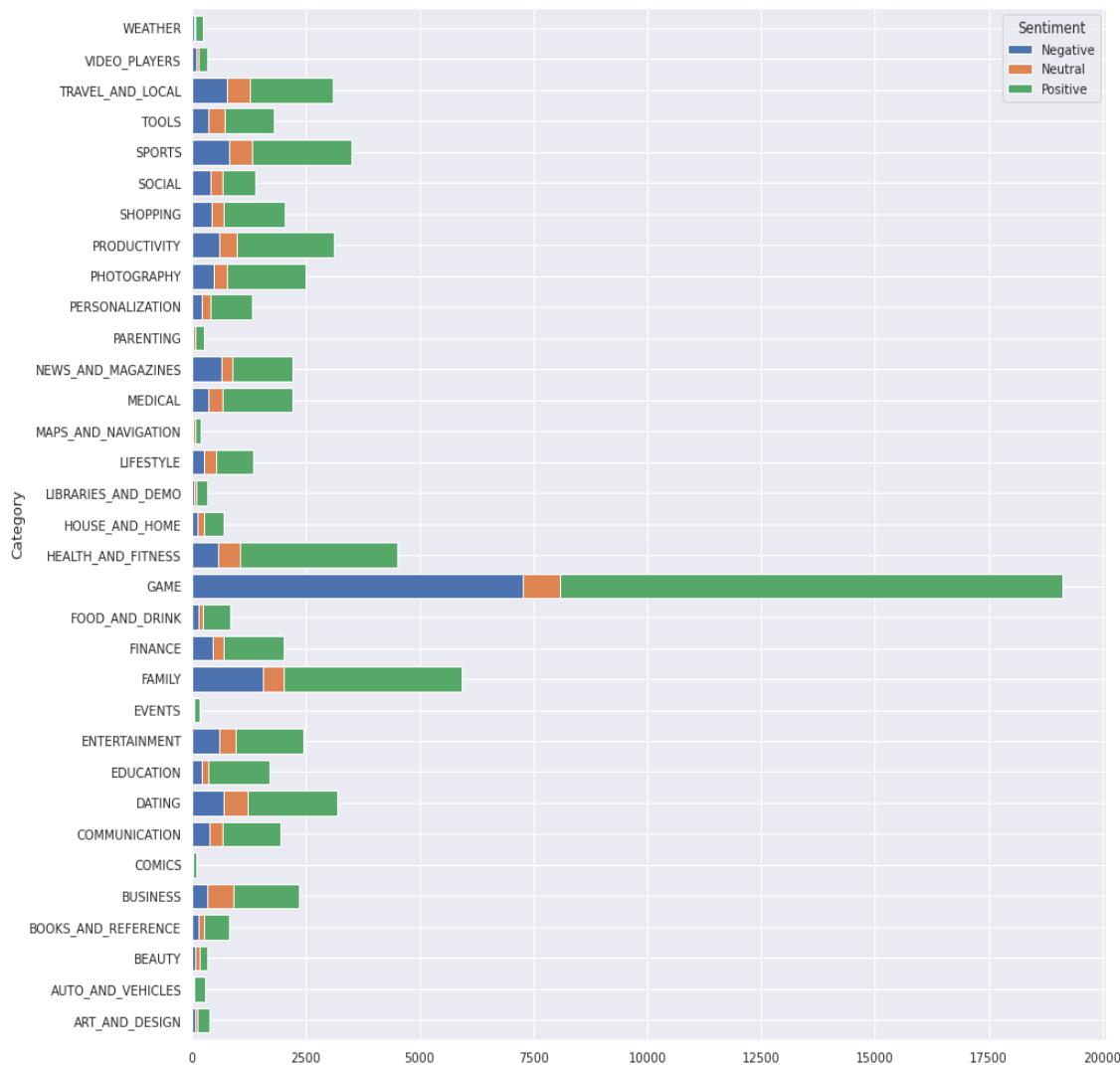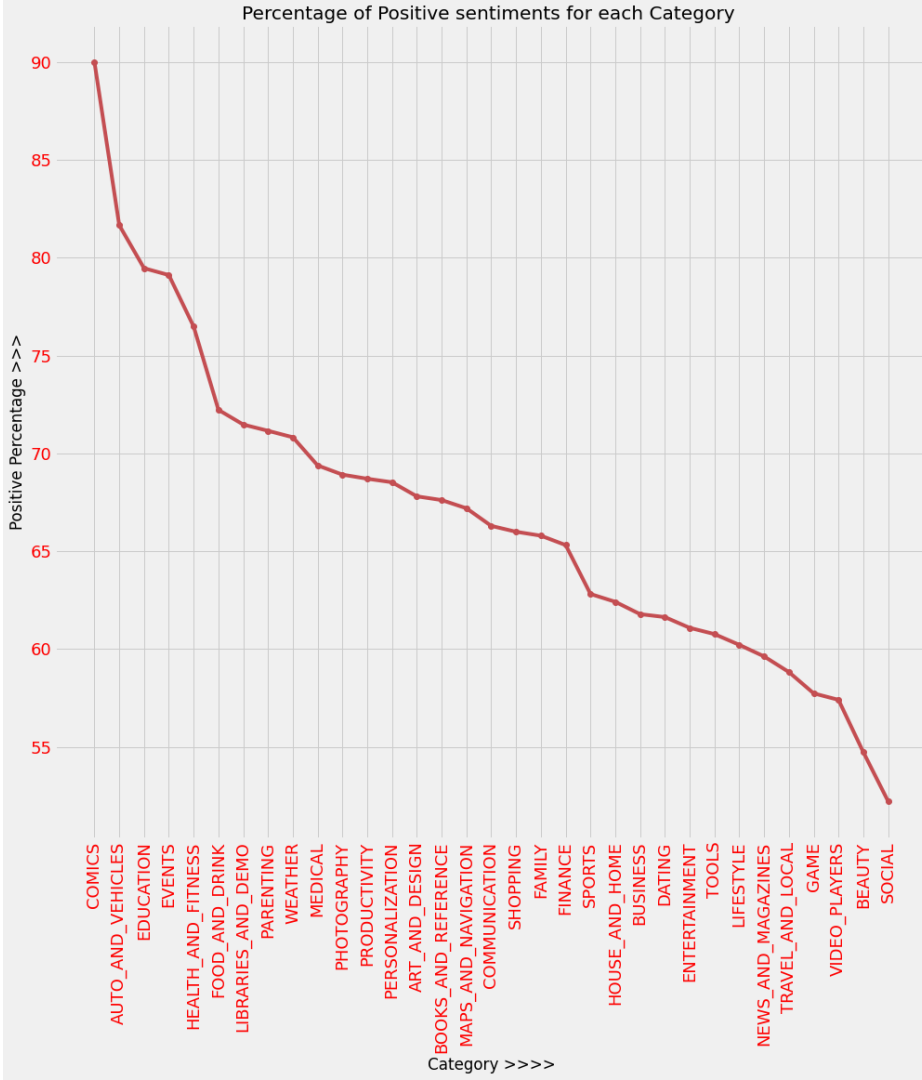
Top 10 Positive Percentage per Category

➢ *There are 5 categories which has more than 75% positive sentiment of total reviews*

➢ *These categories are already satisfying its users, and the categories with less positivity are yet to receive the user satisfaction, because a negative review are reference to dissatisfied user.*

AI

EDA

Summary and Conclusion

AI

- ❑ *'Game' and 'Family' category are already dominating the market, but there are some low competitive categories which have lots of potential are- 'Beauty', 'Comic' and 'Photography' developers should focus on that.*

- ❑ *Paid apps are not suitable for broad users, they are meant to be made for a specific niche or segment and focus only on particular segment for better profits.*

- ❑ *For paid games, we seen a pattern that if it was already available for different platforms and is popular, then if it launch on play store, it will receive tremendous response from user. For example;- Minecraft, GTA vice city, Generated highest revenue.*

- ❑ *Size of app will impact user 's behaviour, so its best choice for developers to create apps with low size as possible.*