Capstone Project - 1
PLAY STORE APP REVIEW ANALYSIS

Team Members
Ankit Patil
Mihir Kulkarni
Ram Bakale

AI

Let's Review Play Store:

1. Overview of Google Play Store
2. Available Data sets
3. Variables used in Google Play store and User Reviews data sets
4. Exploratory Data Analysis
   a) Apps per Category
   b) Average Rating across categories
   c) Average Price according to Categories
   d) Average Price according to Categories
   e) Number of Installs per Category Type wise
   f) Free Vs Paid Apps Analysis
   g) Apps per Age Group
   h) 5 Star Rated Apps per Category
   i) Top 10 most Expensive Apps
   j) Top 10 highest rated Apps
   k) Top 10 Revenue making Apps
   l) Distribution of Sentiments as per Installs
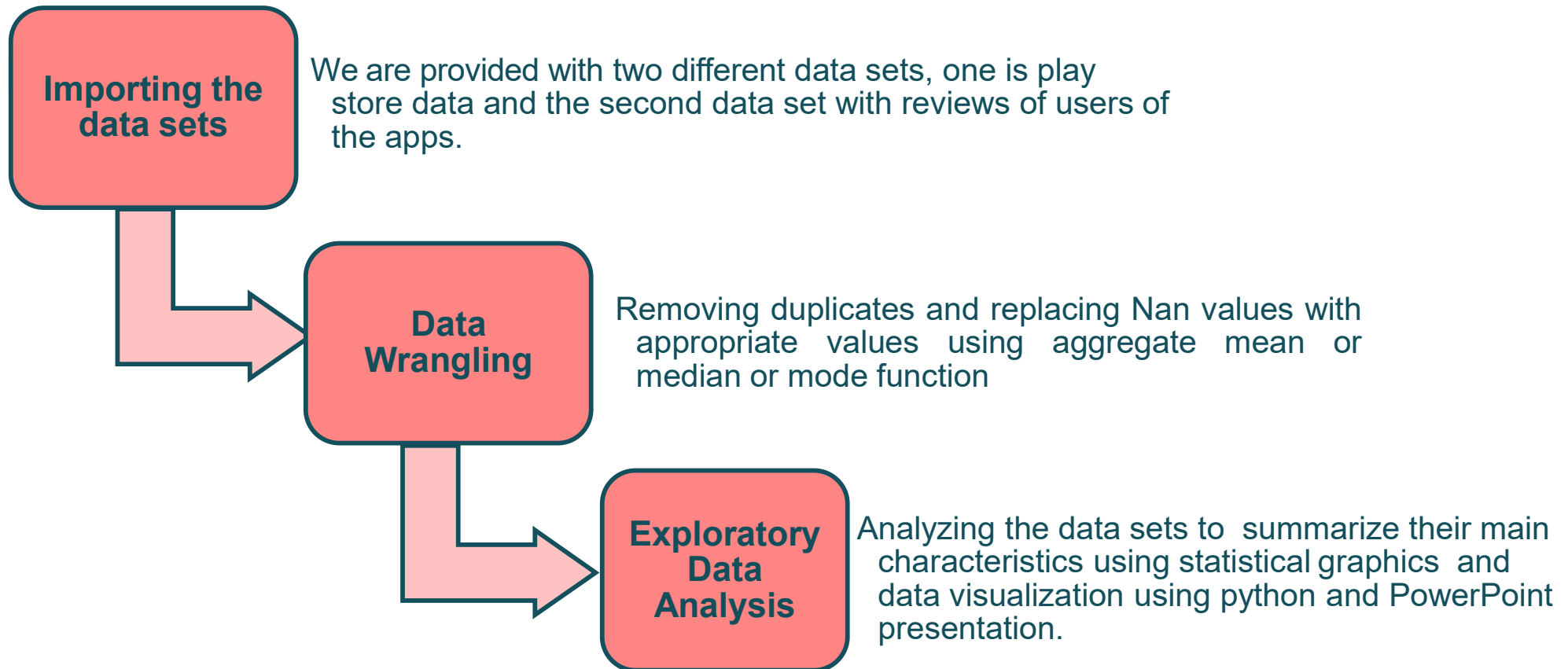5. Summary
6. Conclusion

Overview of Google Play Store

- The Google Play Store is an App that provides access to the collection of Apps that are available on Servers at google for Android devices.
- These apps further may be classified as per the category it falls under.
- Google keeps record for each User about apps installed, ratings given, reviews posted etc.
- This data is further used and analyzed by Google to enhance user experience by improving App engagement

AI

Process flow

Importing the data sets

We are provided with two different data sets, one is play store data and the second data set with reviews of users of the apps.

Data Wrangling

Removing duplicates and replacing Nan values with appropriate values using aggregate mean or median or mode function

Exploratory Data Analysis

Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualization using python and PowerPoint presentation.

Available Data sets

1. **Play Store Data.csv:** In this file the 9659 unique apps from free and paid category are described against 13 columns as characteristics comparable amongst them.

2. **User Review.csv:** It contains 3 columns as characteristics :
   ❖ Sentiment (Positive, Negative or Neutral),
   ❖ Sentiment Polarity
   ❖ Sentiment Subjectivity.

VARIABLES DESCRIPTION IN GOOGLE PLAY STORE:

Variable	Description
1. App	Name of Application
2. Category	Category to which the app belongs to
3. Rating	Overall user rating
4.Reviews	Number of user reviews for the app
5. Size	Size of the app
6. Installs	Number of user downloads for the app
7. Type	Paid or Free
8. Price	Price of the app
9.Genres	An App can belong to multiple genres

VARIABLE DESCRIPTION IN USER REVIEWS:

Variable	Description
App	Application name
Translated_Review	User review
Sentiment	Positive/Negative/Neutral
Sentiment_Polarity	Sentiment polarity score
Sentiment_Subjectivity	Sentiment subjectivity score

EXPLORATORY DATA ANALYSIS(EDA)

AI

Count of applications for each category

Applications per Category
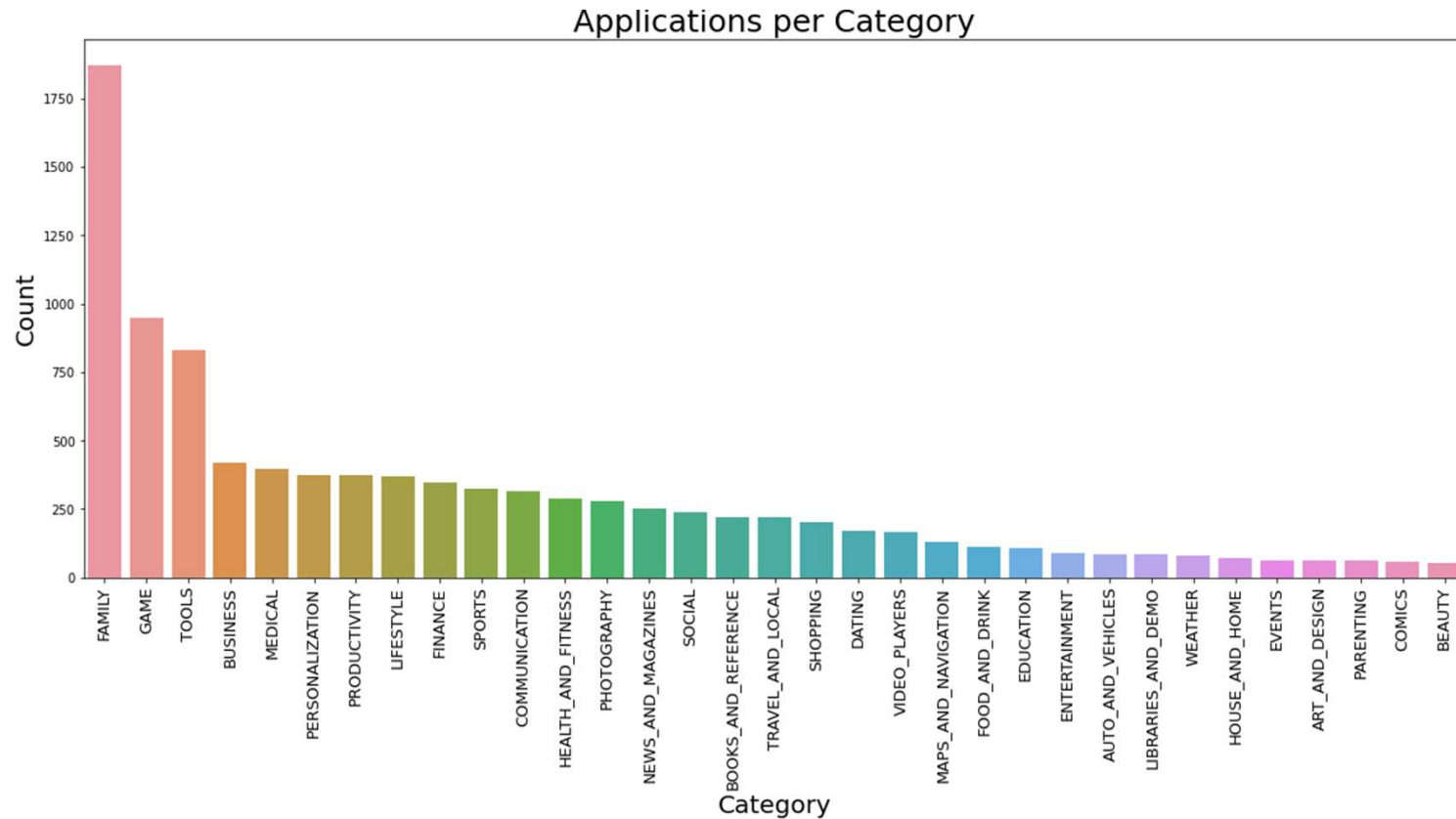
Family has most number of apps present on Play Store, followed by Game and then Tools
Beauty has least number of apps present on Play Store
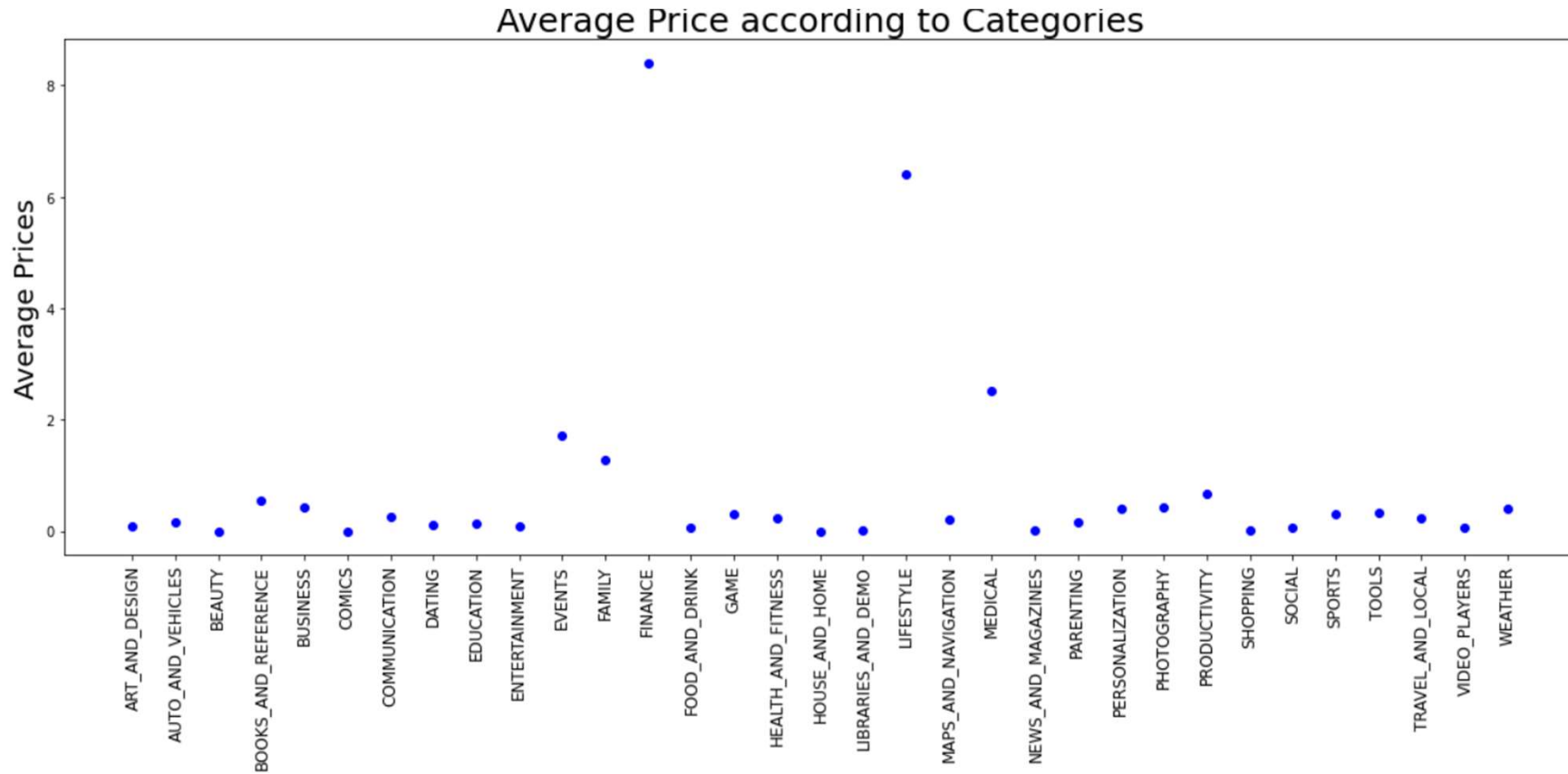
EDA(continued)

Average Rating Across Categories:

Highest Rated apps:

Events, Art and Design, Education Categories has got maximum i.e. 4.35 and above ratings.

Lowest Rated apps:

Dating, Maps and navigation, Video players category has got least ratings as per the available data.
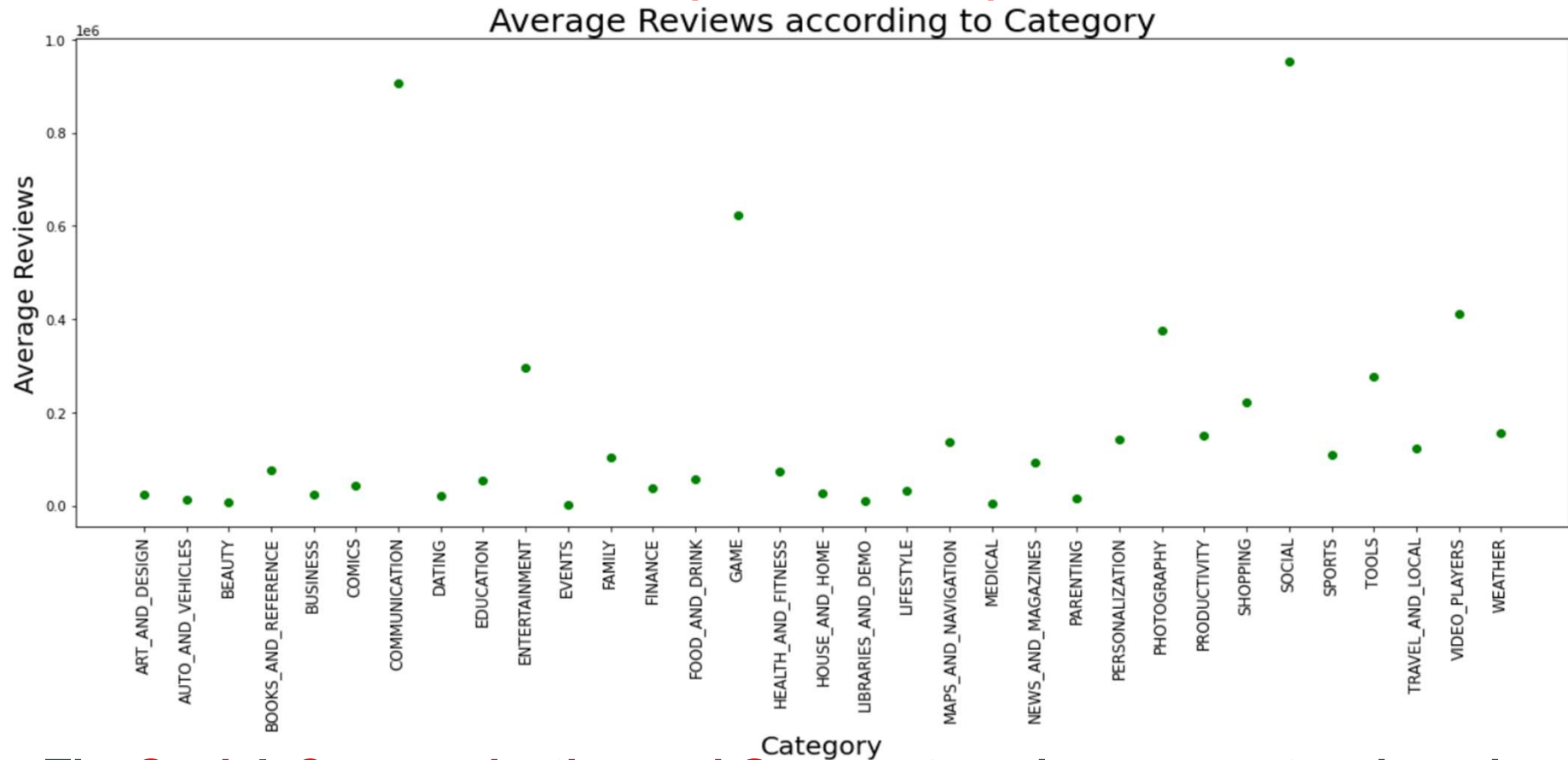

Average Rating according to Category
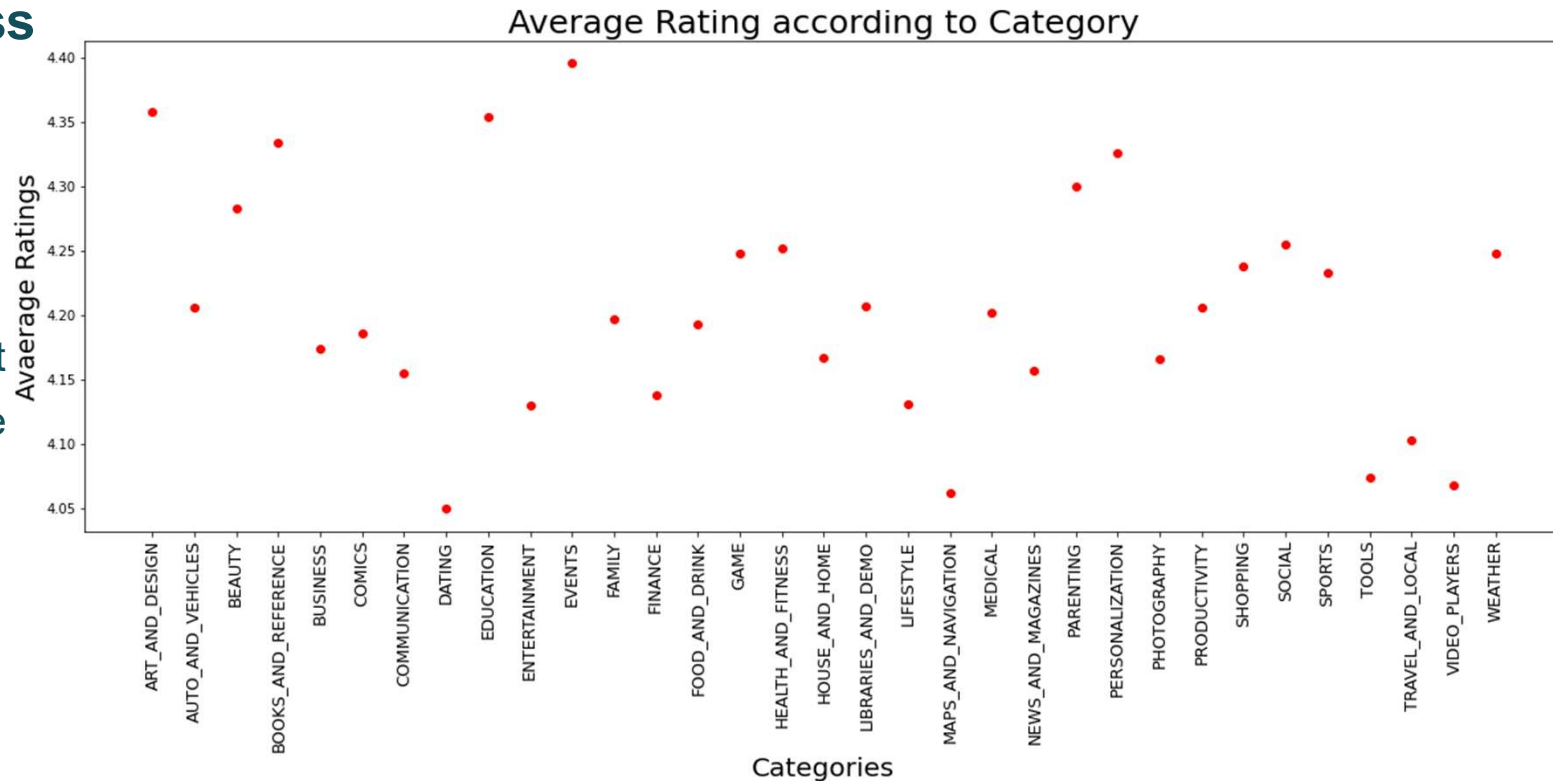
EDA(continued)

Finance has highest average price followed by Lifestyle and then Medical Category for apps on Play Store.

Average Price according to Categories

EDA(continued)

Average Reviews according to Category

The **Social, Communication and Game** categories are most reviewed on the play store.

We have converted numeric values to logarithmic values to compensate for skewness in number of installs for different categories

Number of installs type wise according to Category
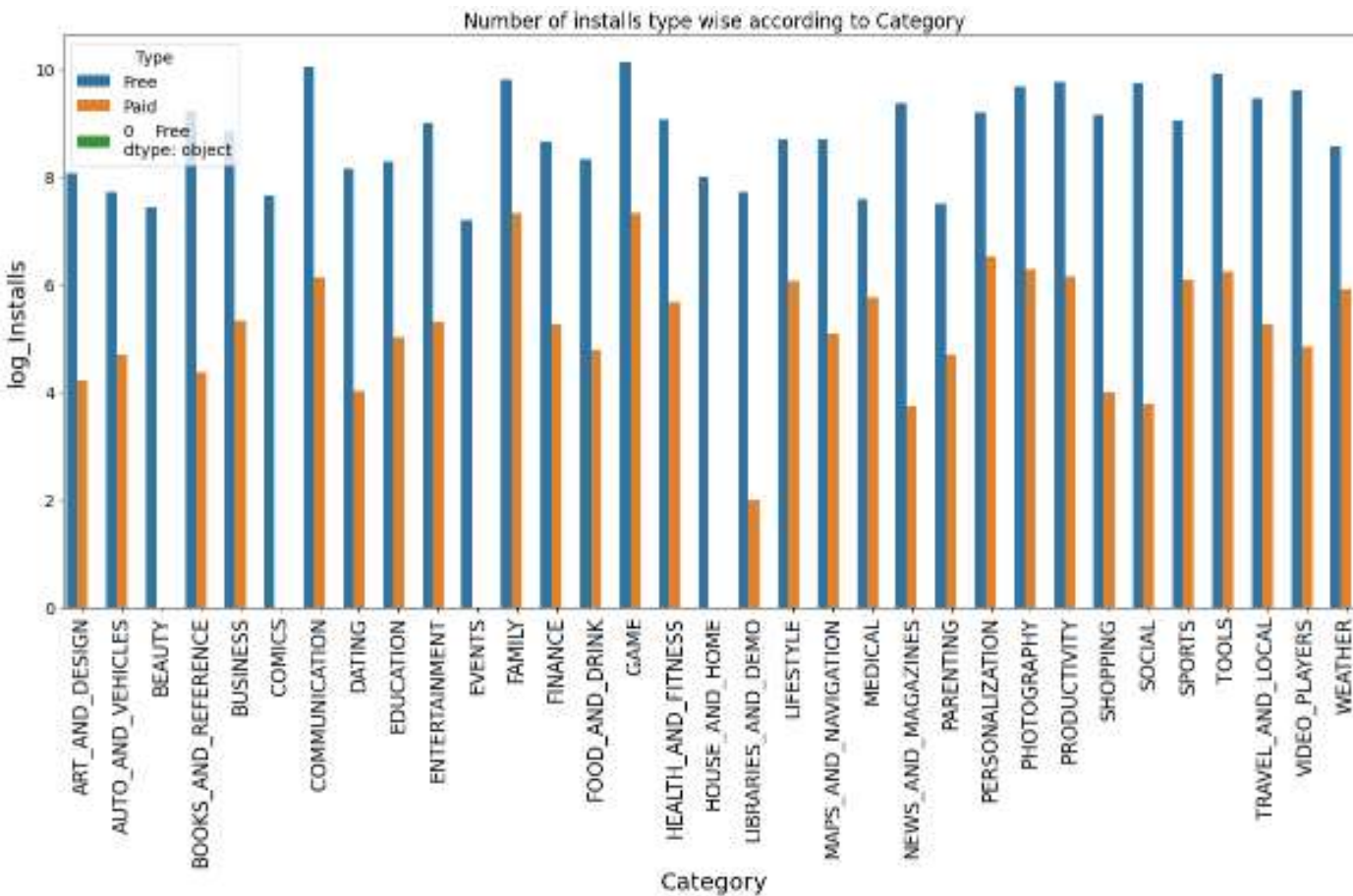
In our dataset, the majority of apps in **Family, Game** and **Communication** categories are free to install.

At the same time **Family, Game, and Personalization** categories had the biggest number of paid apps available on play store.
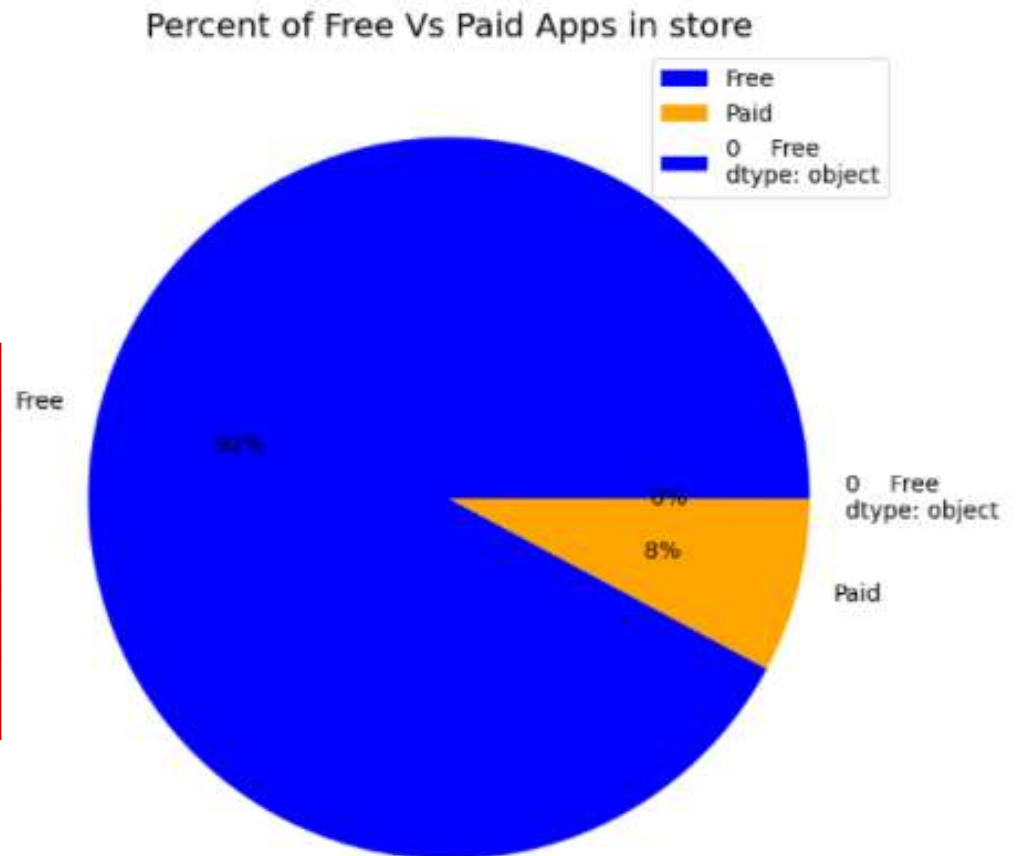
Comics, Events and **House And Home** Category has no paid apps

EDA(continued)

Free Vs Paid Apps

Here we can see that there are 92% of apps are free and only 8% of Apps are paid on Play store.

|	Free Apps	Paid Apps
Number	8903	755
Average Rating	4.186027	4.269934
Installs	75216062646	57363881
Reviews	2087014255	6595973

Percent of Free Vs Paid Apps in store

Number of Free Apps present on Play Store is in multiple folds of Paid Apps

EDA(continued)

Apps classified by different age group

Apps rated for Everyone are the highest number of Apps followed by Apps rated for Teens present on Play Store

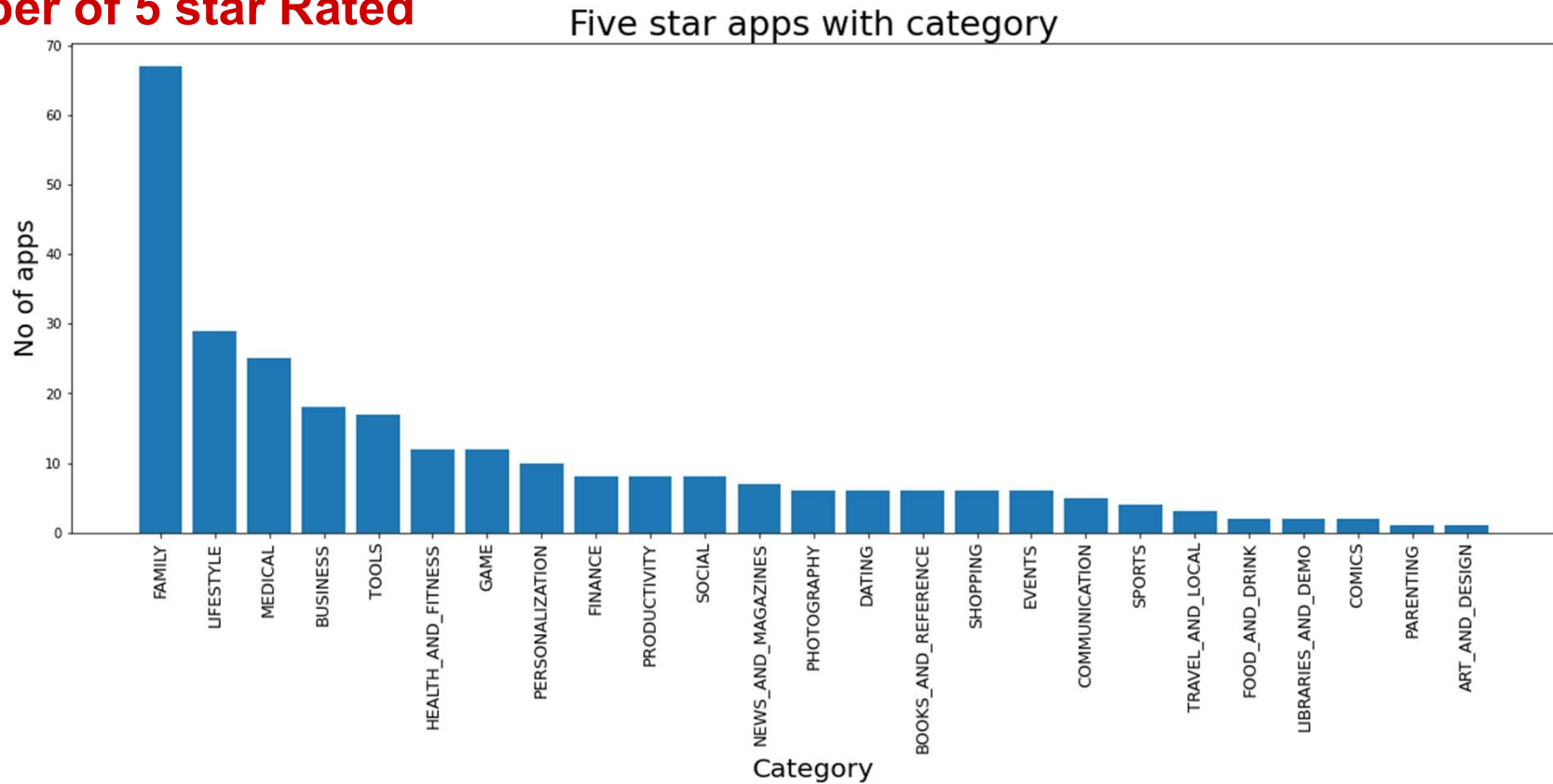Apps availabilty by ages
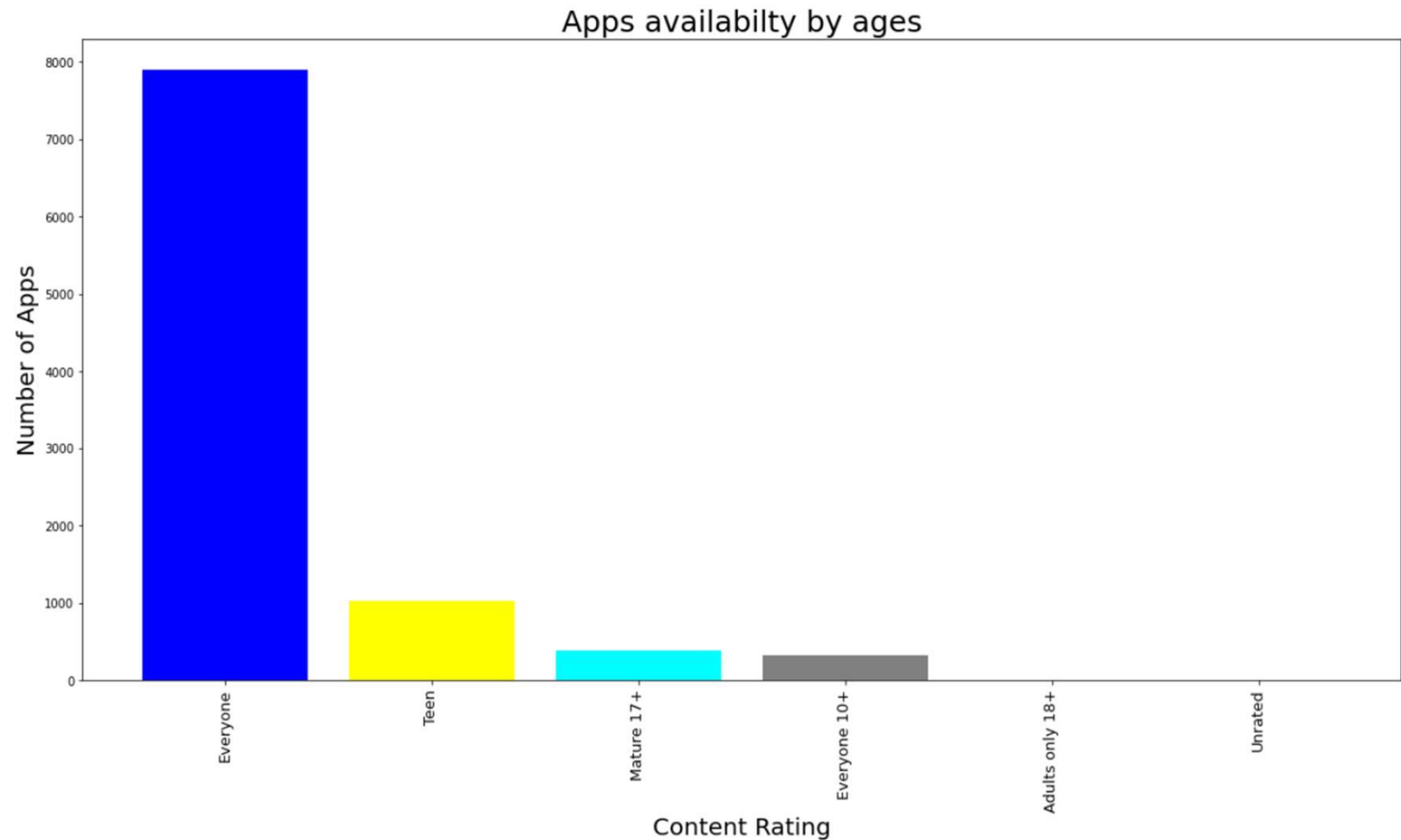
EDA(continued)

Distribution of App sizes

It can be seen from the distribution size plot that maximum applications' size is between 0–10 MB.

Distribution of Size

EDA(continued)

AI

Number of 5 star Rated apps

Five star apps with category
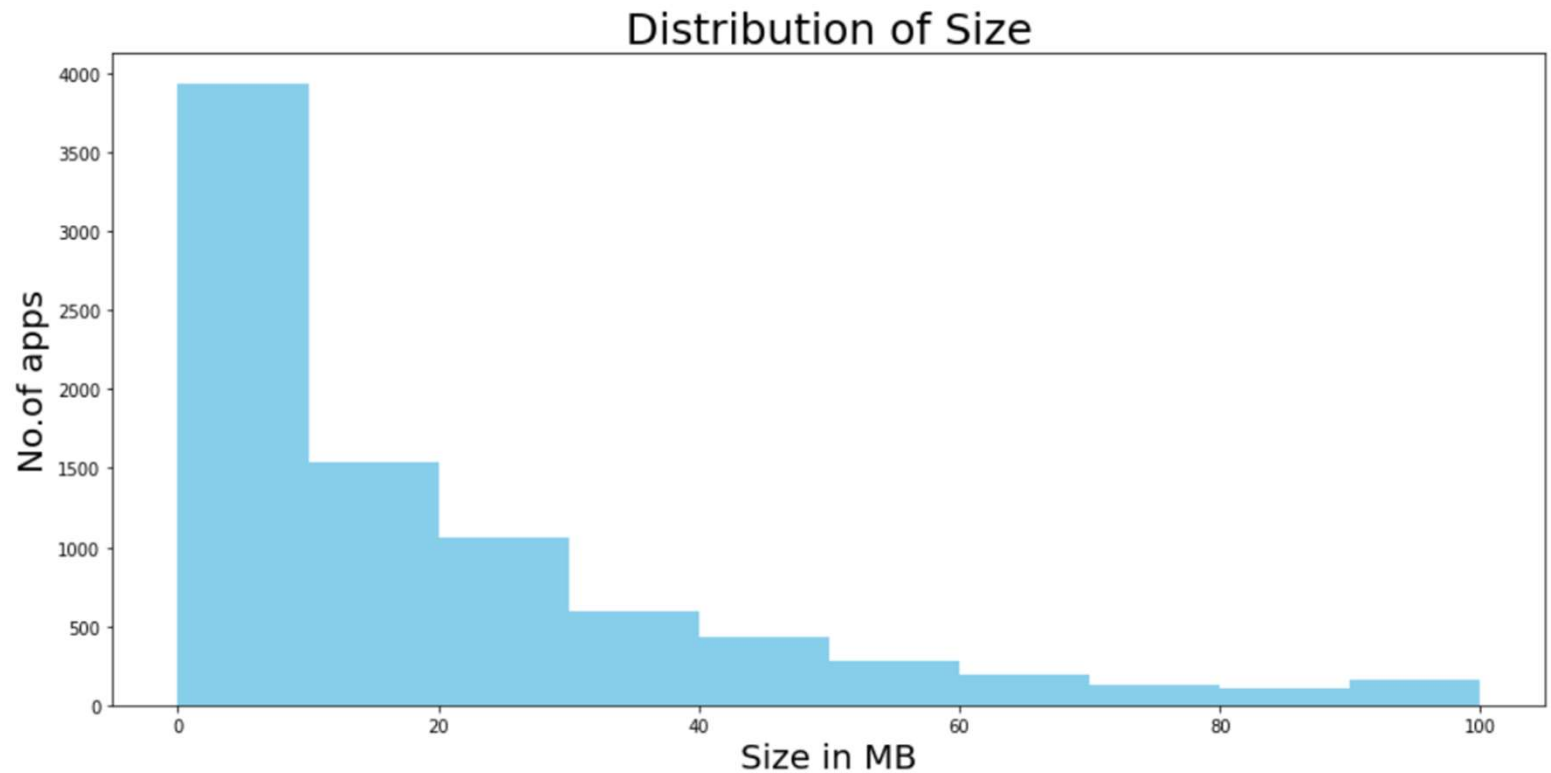
Family Category has most 5 star rated Apps followed by Lifestyle, whereas Art and Design has least 5 star rated Apps

EDA(continued)

Top 10 Expensive Apps

Application	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
I'm Rich - Trump Edition	LIFESTYLE	3.6	275.0	7.3	10000.0	Paid	400.00	Everyone	Lifestyle
I Am Rich Pro	FAMILY	4.4	201.0	2.7	5000.0	Paid	399.99	Everyone	Entertainment
I am Rich	FINANCE	4.3	180.0	3.8	5000.0	Paid	399.99	Everyone	Finance
I'm Rich/Eu sou Rico//أنا غني/我 很 有錢	LIFESTYLE	4.3	0.0	40.0	0.0	Paid	399.99	Everyone	Lifestyle
most expensive app (H)	FAMILY	4.3	6.0	1.5	100.0	Paid	399.99	Everyone	Entertainment
I Am Rich Premium	FINANCE	4.1	1867.0	4.7	50000.0	Paid	399.99	Everyone	Finance
I am rich (Most expensive app)	FINANCE	4.1	129.0	2.7	1000.0	Paid	399.99	Teen	Finance
I AM RICH PRO PLUS	FINANCE	4.0	36.0	41.0	1000.0	Paid	399.99	Everyone	Finance
I am Rich Plus	FAMILY	4.0	856.0	8.7	10000.0	Paid	399.99	Everyone	Entertainment
I am Rich!	FINANCE	3.8	93.0	22.0	1000.0	Paid	399.99	Everyone	Finance

I am rich – Trump Edition is the most expensive app followed by I am Rich Pro on the google play store.

EDA(continued)

AI

Top 10 Highest Rated Apps

Application	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
Ríos de Fe	LIFESTYLE	5.0	141.0	15.0	1000.0	Free	0.0	Everyone	Lifestyle
FD Calculator (EMI, SIP, RD & Loan Eligilibility)	FINANCE	5.0	104.0	2.3	1000.0	Free	0.0	Everyone	Finance
Oración CX	LIFESTYLE	5.0	103.0	3.8	5000.0	Free	0.0	Everyone	Lifestyle
Barisal University App-BU Face	FAMILY	5.0	100.0	10.0	1000.0	Free	0.0	Everyone	Education
Master E.K	FAMILY	5.0	90.0	NaN	1000.0	Free	0.0	Everyone	Education
CL REPL	TOOLS	5.0	47.0	17.0	1000.0	Free	0.0	Everyone	Tools
AJ Cam	PHOTOGRAPHY	5.0	44.0	2.8	100.0	Free	0.0	Everyone	Photography
AI Today : Artificial Intelligence News & AI 101	NEWS_AND_MAGAZINES	5.0	43.0	2.3	100.0	Free	0.0	Everyone	News & Magazines
CS & IT Interview Questions	FAMILY	5.0	43.0	3.3	1000.0	Free	0.0	Everyone	Education
Ek Vote	PRODUCTIVITY	5.0	43.0	6.2	500.0	Free	0.0	Everyone	Productivity

Rios de Fe is the Highest Rated app followed by FD Calculator (EMI, SIP, RD & Loan Eligilibility) the google play store by Reviews.
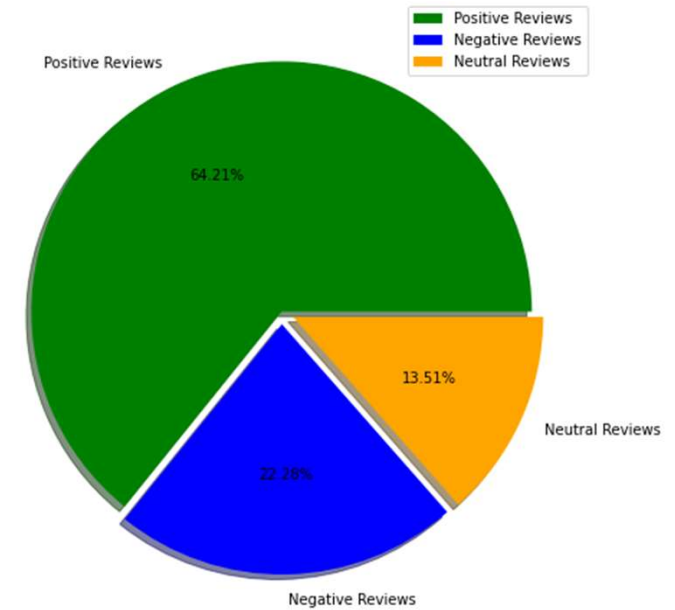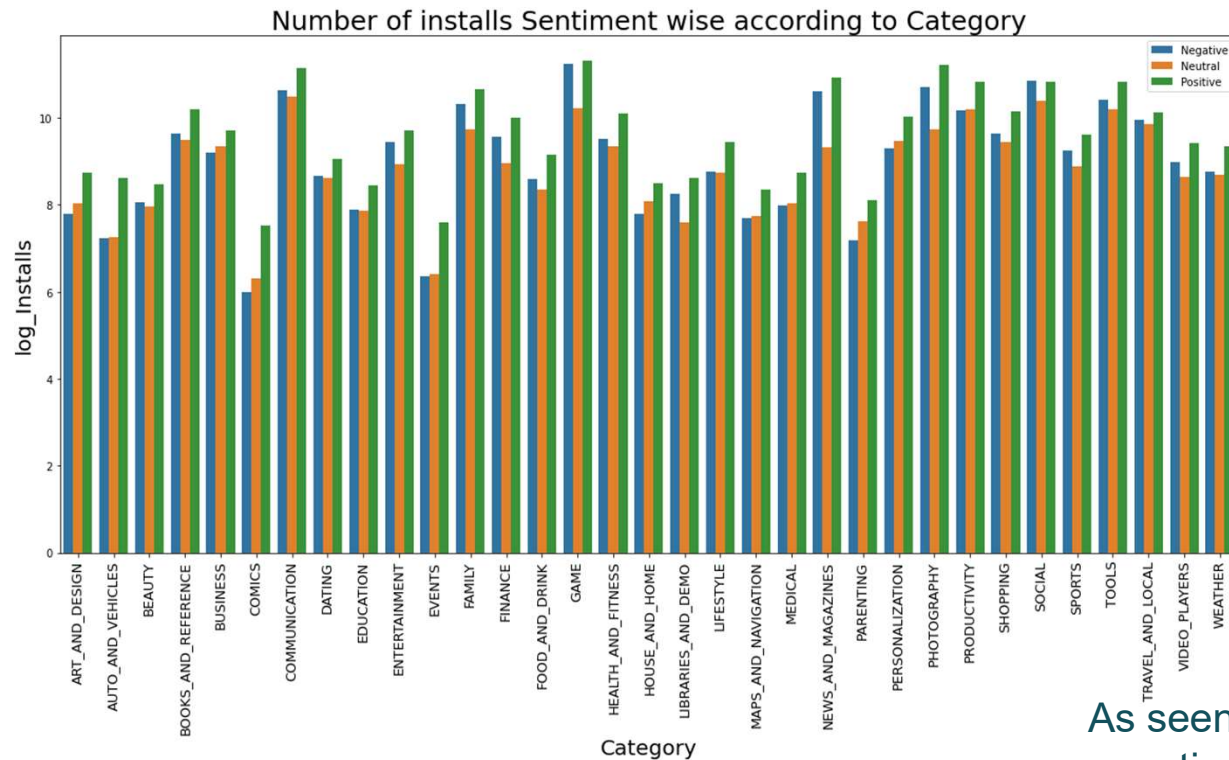
EDA(continued)

Top 10 Revenue generating Apps Present on Play store based on their prices and installs.
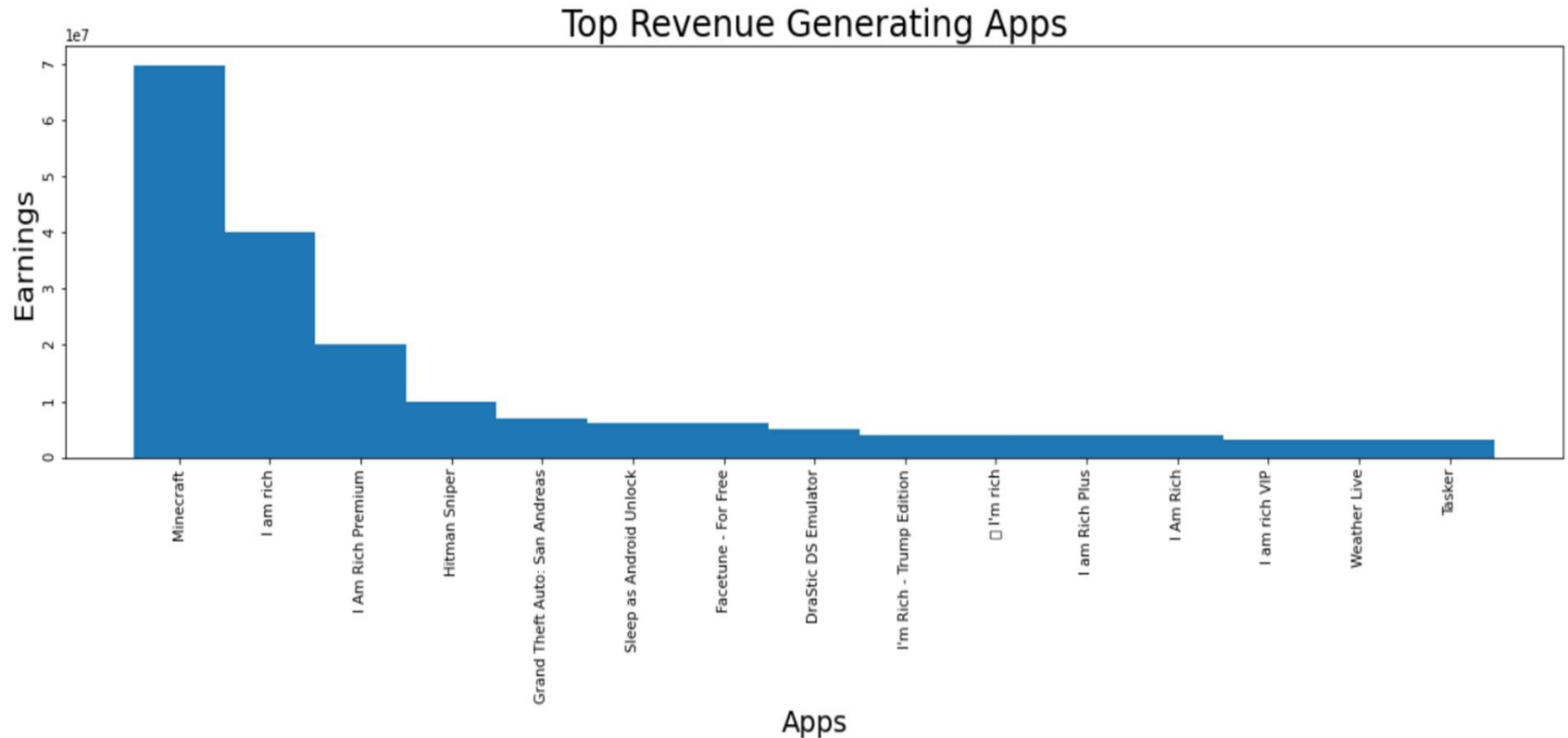
Minecraft has made the most revenue compared to other apps present on Play Store

Distribution of Sentiments according to Installs

Number of installs Sentiment wise according to Category

As seen from the pie chart, positive reviews are 64.08%,
negative reviews are 22.85%
and neutral reviews are 13.06%
Majority of users are satisfied with the apps

* We have converted Installs values to Log Installs values to compensate for wide range in installs across all categories

SUMMARY

Average Rating	**Free: 4.19** **Paid: 4.27**
Number of Installs	**Free: 75216062646** **Paid: 57363881**
Total 5 Star Rated Apps	**271**
Category with most number of apps	**Family**
Category with avg Highest rating	**Events**
Category with highest avg Price	**Finance**
Category with highest avg Reviews	**Social**
Category with highest number of 5 star rated Apps	**Family**
Most Expensive App	**I'm Rich - Trump Edition**
Highest Rated app sorted with Highest number of Reviews	**Ríos de Fe**
Highest Revenue making app	**Minecraft**
Sentiments in Percentage	**Positive: 64.21** **Neutral: 13.51** **Negative: 22.58**

CONCLUSION:

Goal is to provide a better understanding of how App's ratings and reviews measure up to the industry as a whole.

1. User mostly prefers more of free apps. Most of the apps present in playstore are more or less of same size so size doesn't affect their decision much

2. The apps which are available for everyone are having the ratings 4 and above out of 5.

3. Most popular category of apps are on two basis - Number of Installs and number of reviews.

4. Family, Game, Communication, Social, Tools, Medical, Events Apps are the highest earning apps and at the same time Medical, Family, game, Tools apps are the expensive apps. This helps us to understand the application market and invest money in the right categories to make good profits.

5. Also, customer ratings plays a huge role while releasing the application.

6. The sentiment does not influence the final rating of the app because of the proportional behaviour.

7. Popular categories have a more positive sentiment.

8. Positive reviews are higher than negative and neutral but the sentiment subjectivity is not always proportional to sentiment polarity.

THANK YOU