

# Capstone Project 1

## Play Store App Review Analysis

By  
Prasad Kanagi

## Points for discussion

- ☐ Introduction
- ☐ Data Cleaning
- ☐ **Data exploration**
  - ☐ Category wise play store apps installs
  - ☐ Category wise most popular apps
  - ☐ Top 10 apps in play store considering all the parameters
  - ☐ Average installs, category wise
  - ☐ Most installed apps in communication category
  - ☐ Average sizes of apps in each category
  - ☐ Category wise percentage of paid apps
  - ☐ Category wise top installed paid apps
  - ☐ Average rating of paid apps
  - ☐ Correlation between Rating ,Installs and Price
  - ☐ Category wise installed apps with content rating
- ☐ Conclusions

## Introduction

- The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the Product.
- Play store and user reviews dataset contains details of different applications and reviews from different users.
- Our intension is to analyze the data and draw some insights through Python (Pandas) chart analysis , which can prove very beneficial to any Android app development company ,to understand what market demands and what kind or category/genre of apps they can go for as there product.

## **Data Prepration or Data Cleaning**

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

### **Gathering data**

This step is about getting to know the data and understanding what has to be done before the data becomes useful in a particular context. This can be done by reading the CSV file and doing initial statistical analysis. Though the dataset may seem to have the correct datatypes for each column, we need to check it. Inconsistent datatypes will create issues while dealing with problems.

This step is crucial for removing faulty data and filling in gaps. Important tasks here includes:

- Removing extraneous data
- Filling in missing values.
- Conforming data to a standardized pattern.

Dataset may contain duplicate values for particular application:

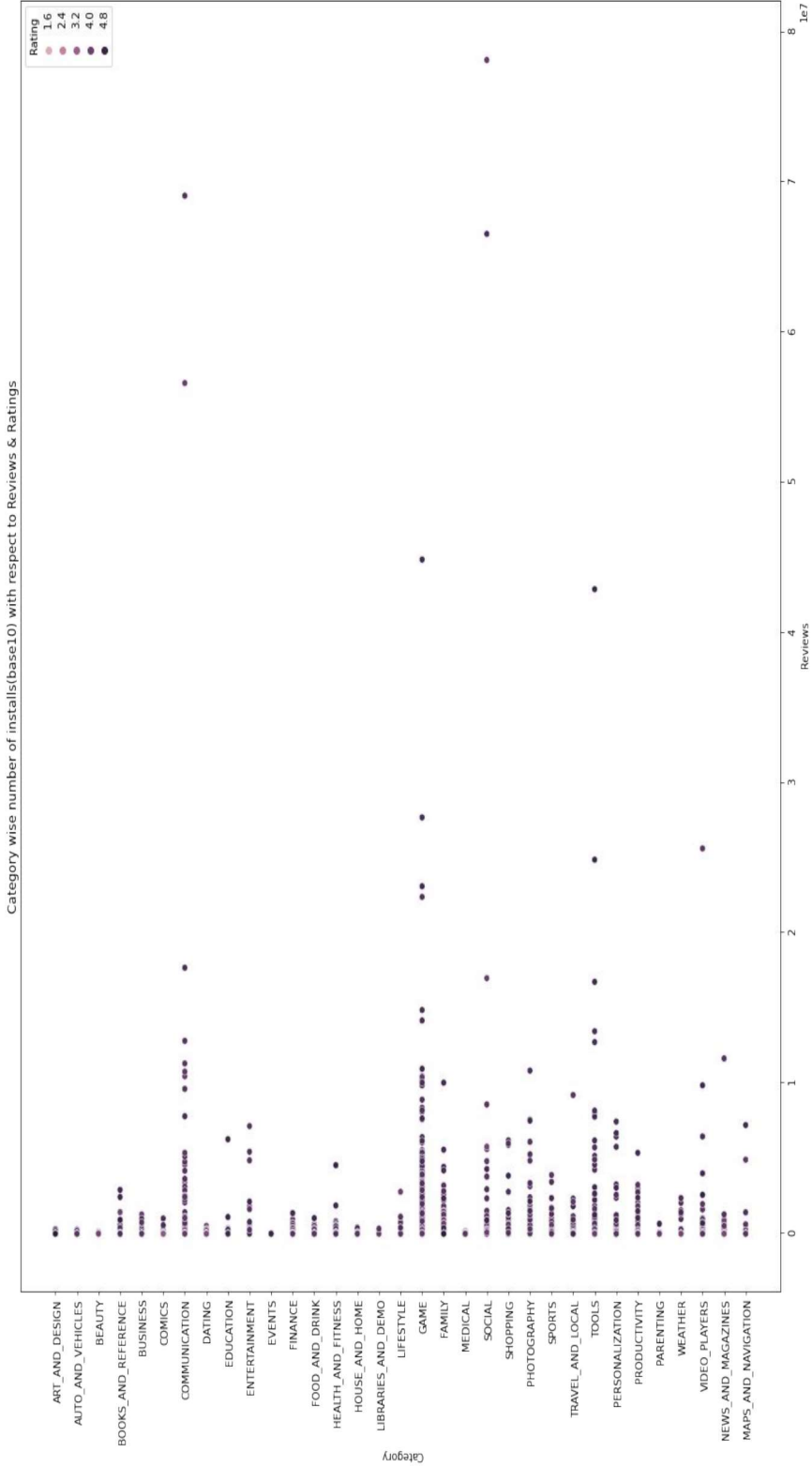
As we can see from the above example, ESPN app is having identical rows with difference in number of reviews.

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
2959	ESPN	SPORTS	4.2	521138	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
3010	ESPN	SPORTS	4.2	521138	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
3018	ESPN	SPORTS	4.2	521138	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
3048	ESPN	SPORTS	4.2	521140	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
3060	ESPN	SPORTS	4.2	521140	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
3072	ESPN	SPORTS	4.2	521140	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up
4069	ESPN	SPORTS	4.2	521081	Varies with device	10,000,000+	Free	0	Everyone 10+	Sports	July 19, 2018	Varies with device	5.0 and up

- **Data exploration** is an approach similar to initial [data analysis](#), whereby a [data analyst](#) uses visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.<sup>[1]</sup> These characteristics can include size or amount of data, completeness of the data, correctness of the data, possible relationships amongst data elements or files/tables in the data.

Variable	Description
App	Application name
Category	Category the app belongs to
Rating	Overall user rating of the app
Reviews	Number of user reviews for the app
Installs	Number of user downloads for the app
Price	Price of the app
Last updated	Date when the app was last updated
Sales Revenue	Revenue generated from paid apps

# Category Wise Play Store Apps Installs

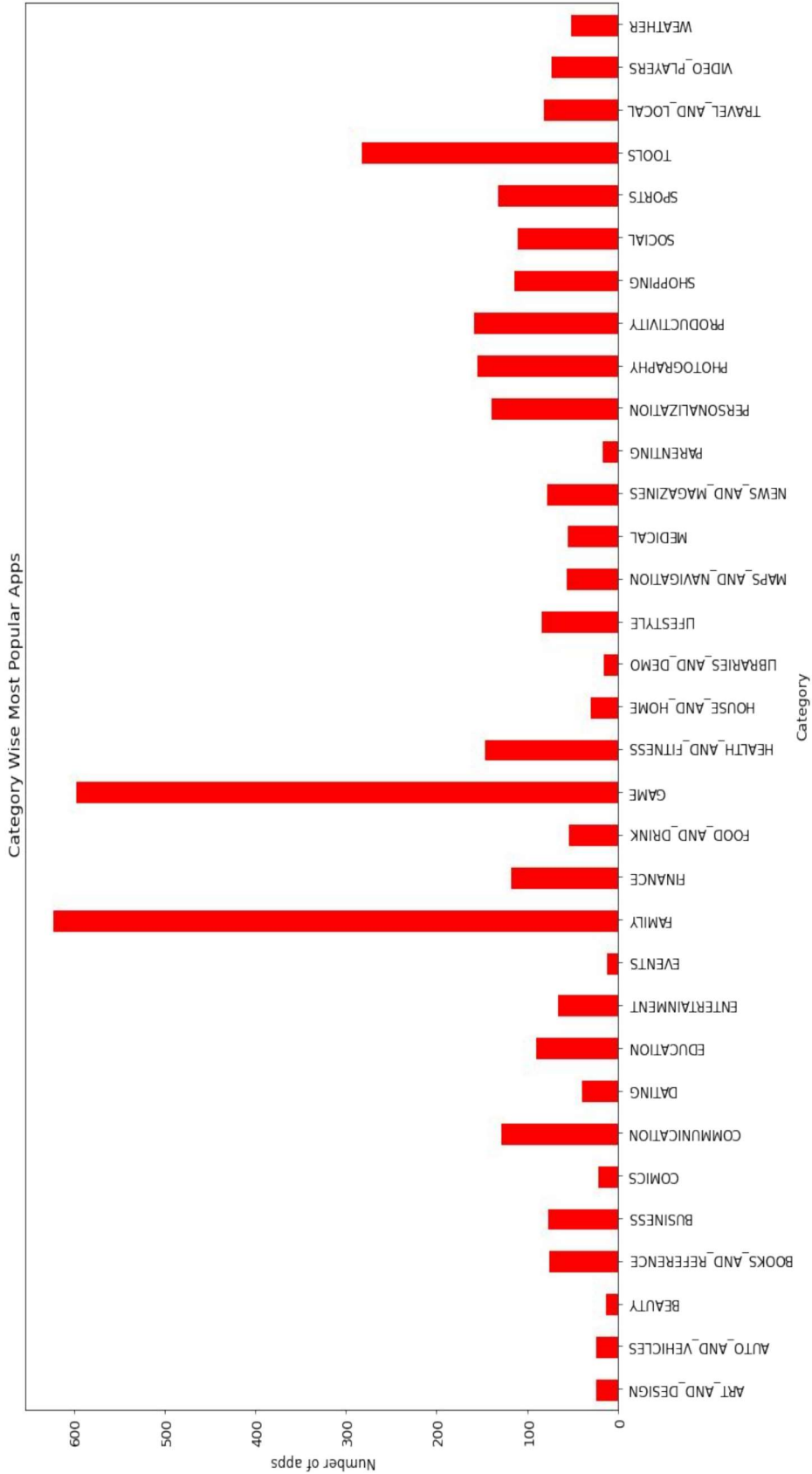


- Findings : There are more reviews and ratings for a growing number of apps in the Play Store.
  - 1) Communications, 2) Gaming, 3) Family, 4) Tools
- Recommendations:
  - i) Investors in app development might choose from the aforementioned categories for high marketing returns.
  - ii) Since they have become a need for every Android smartphone user, entrepreneurs and software solution providers may target these APP categories for marketing and contracting.



# Category Wise Most Popular Apps

AI



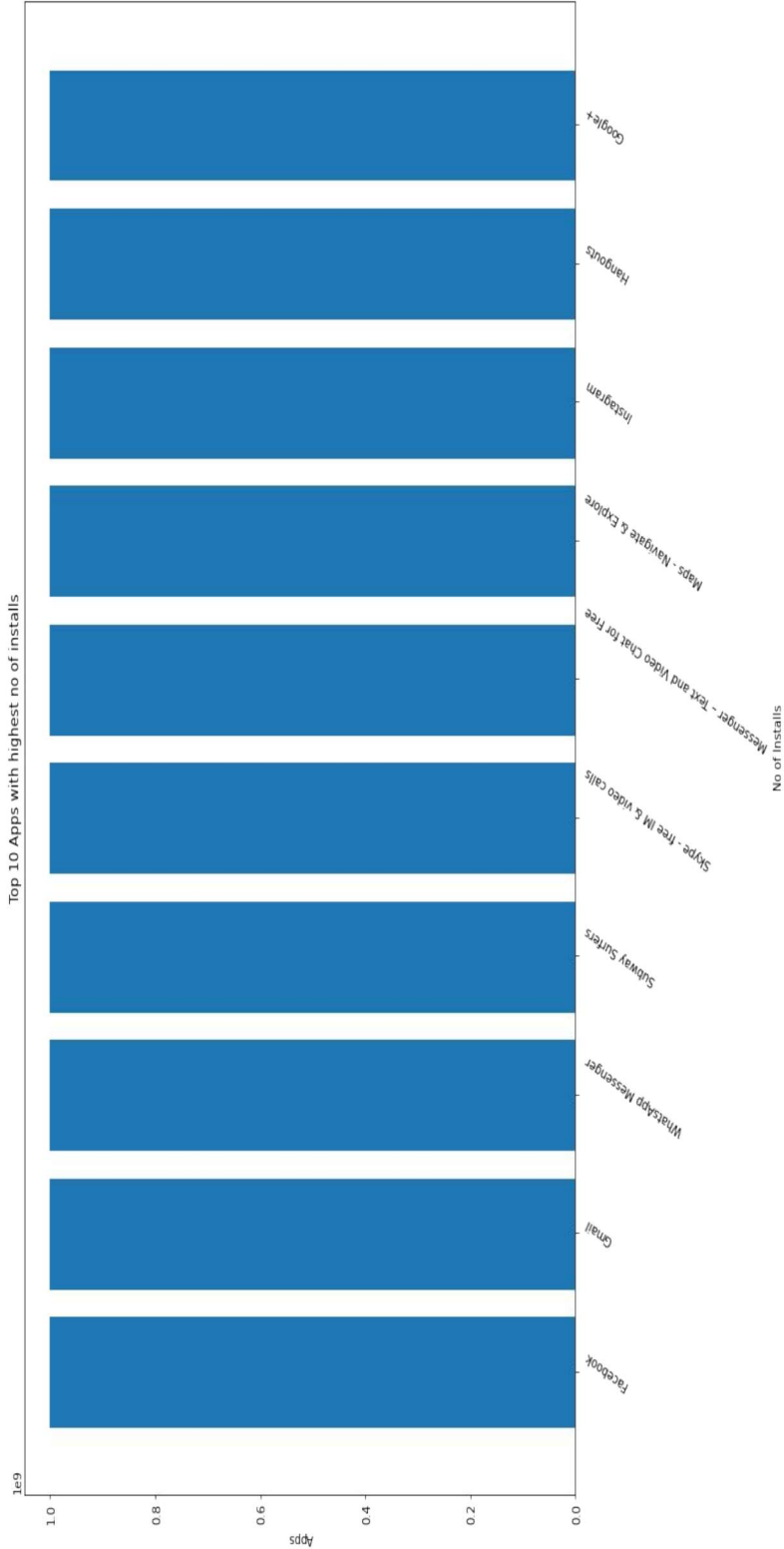
Findings : From the above bar Chart we filtered negligible count of Reviews (considered only which having more than 1000 reviews) and Ratings (Which rated more than 4). So that we can infer the most popular App Category.

1) Family, 2) Gaming

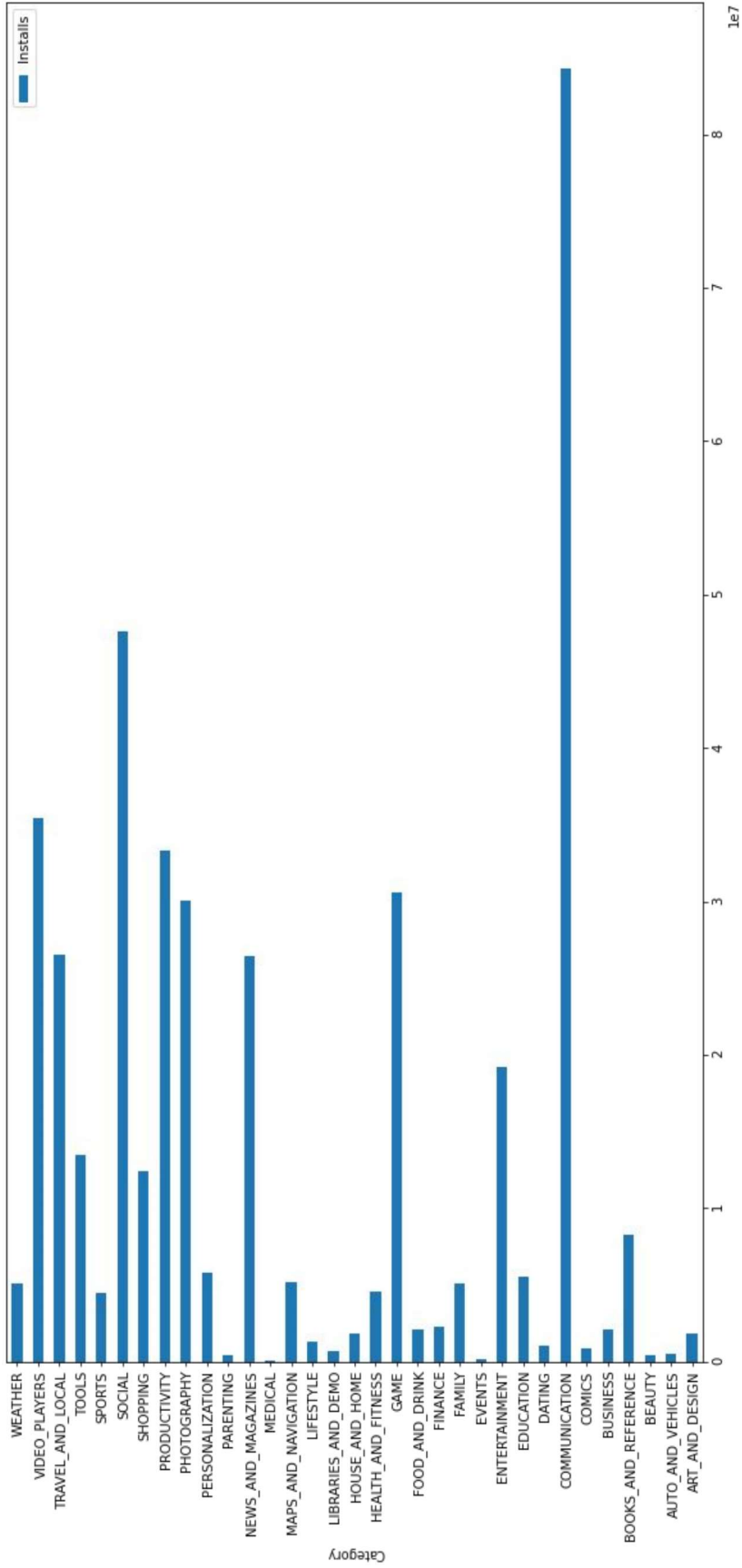
Recommendations:

- i) Investors in app development might choose from the aforementioned categories for high marketing returns.
- ii) Since they have become a need for every Android smartphone user, entrepreneurs and software solution providers may target these APP categories for marketing and contracting.

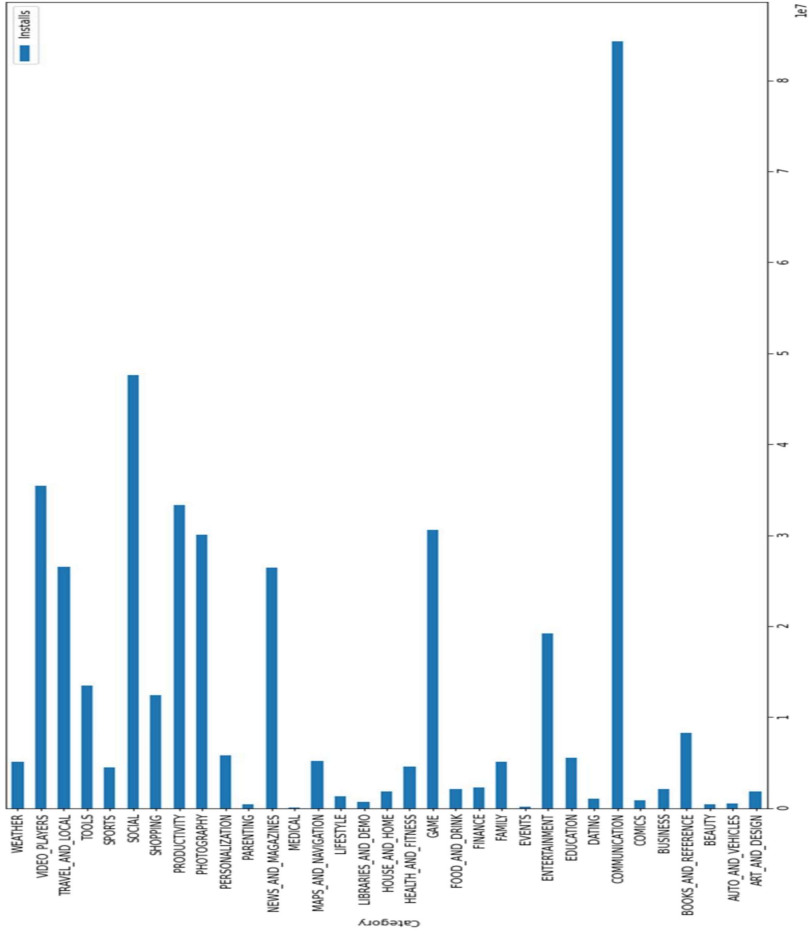
Top 10 best Apps in the Play Store considering all the parameters



Bar Graph showing average installs ,Category Wise



# Bar Graph showing average installs ,Category Wise



## Findings & Recommendations

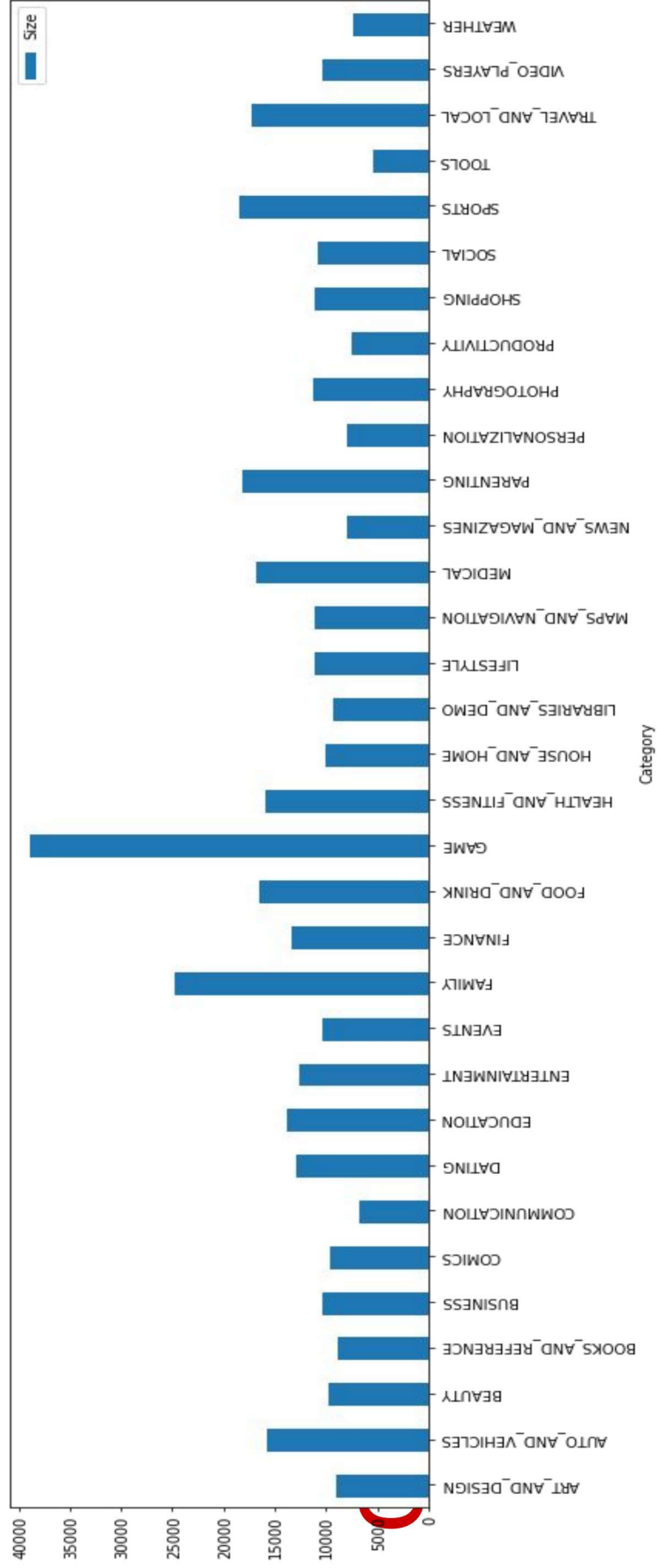
We recommend to develop an app in top 3 max installed categories of apps.

- 1) Communication or
- 2) Social or
- 3)Online Video Players

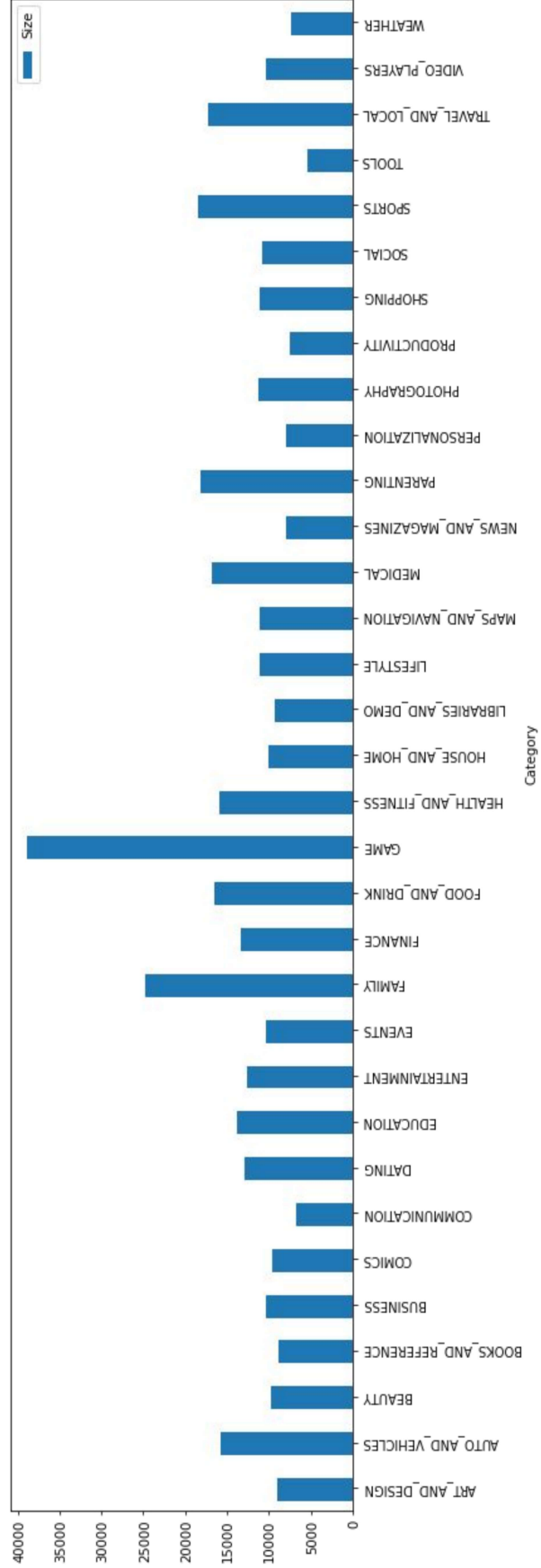
We find Parenting , Medicals ,Events, Auto & Vehicles category are least installed so we don't recommend to develop such apps.



Showing Average sizes of apps in each category



# Showing Average sizes of apps in each category



## Findings & Recommendations

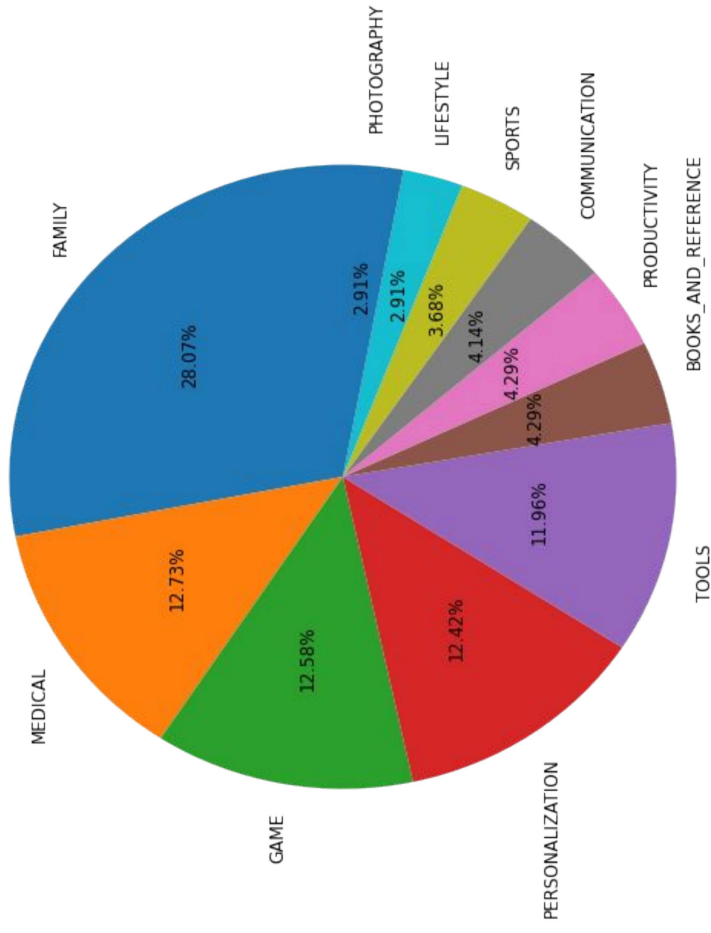
- Our Study shows avg sizes of all Highly installed categories of apps are below or around 12 MB, not very big in size .
- We recommend to target your apps max size itthin 14 MB.



## Category wise percentage of paid apps

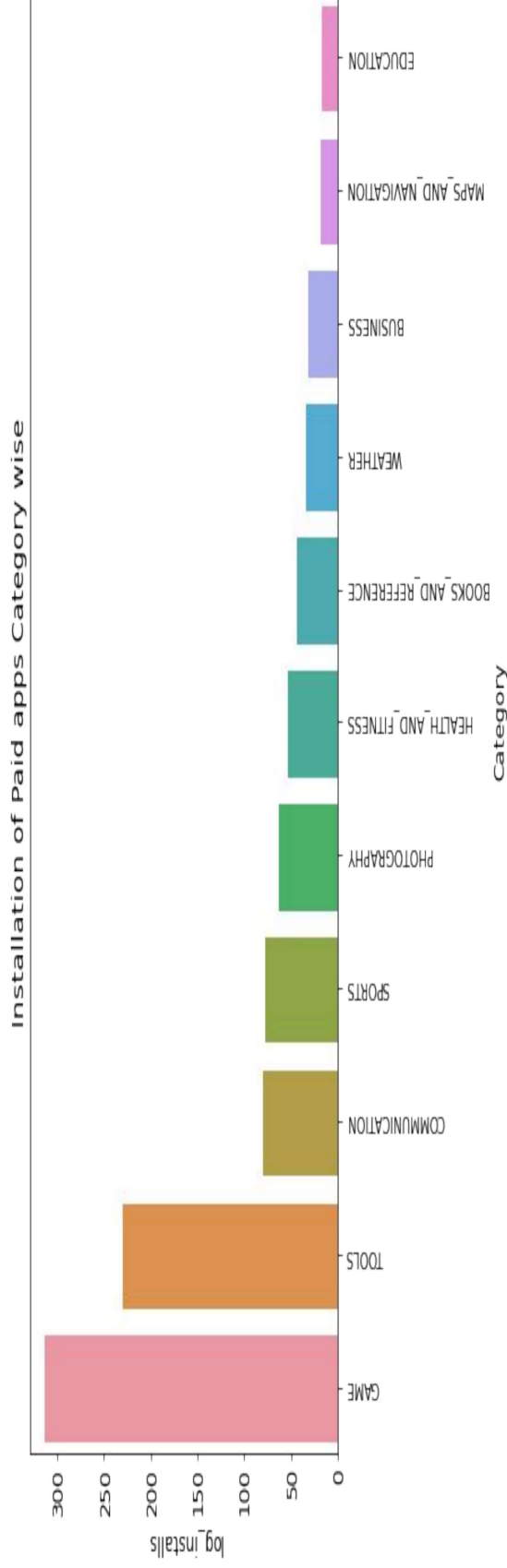
After filtering only paid apps from the data set, we can observe that more than 50% of the apps that are not free were concentrated in the Family, Medical and Game Category

Top 10 Paid apps Category Wise



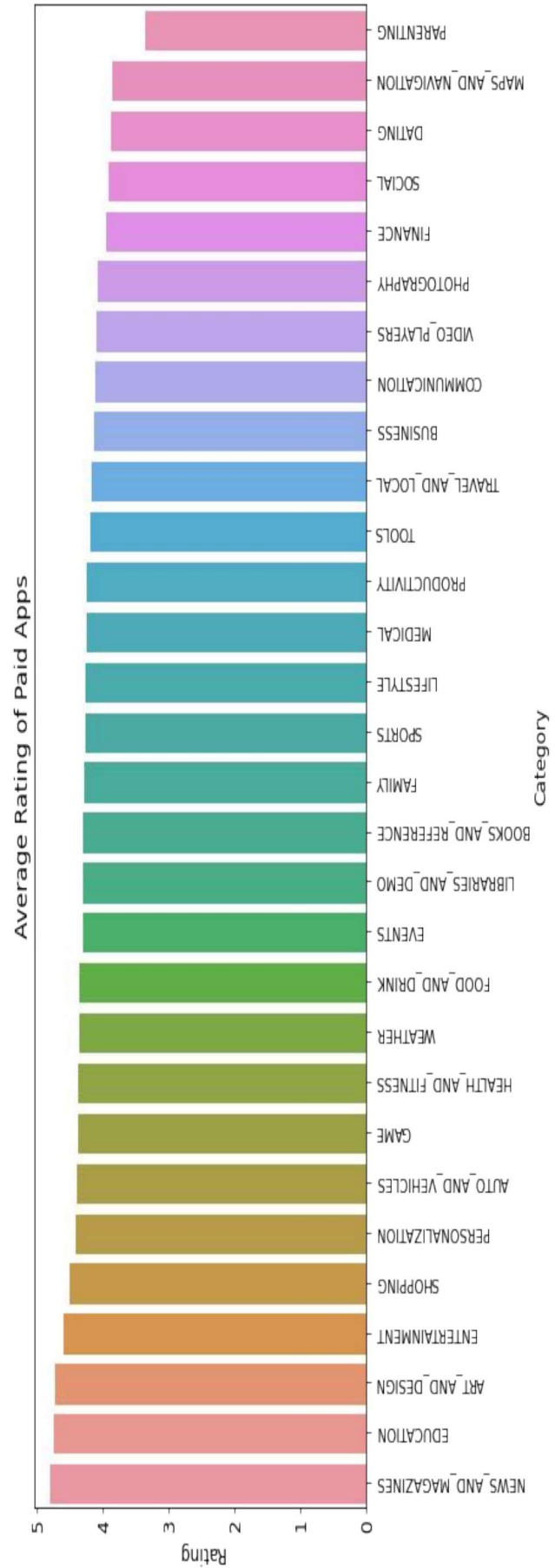
## Category wise top installed Paid apps

In this graph we can observe that,most of the users had installed apps in the the Game and Tools category even though they were paid.So, app development business holders have scope to develop apps in these categories with the paid criteria.



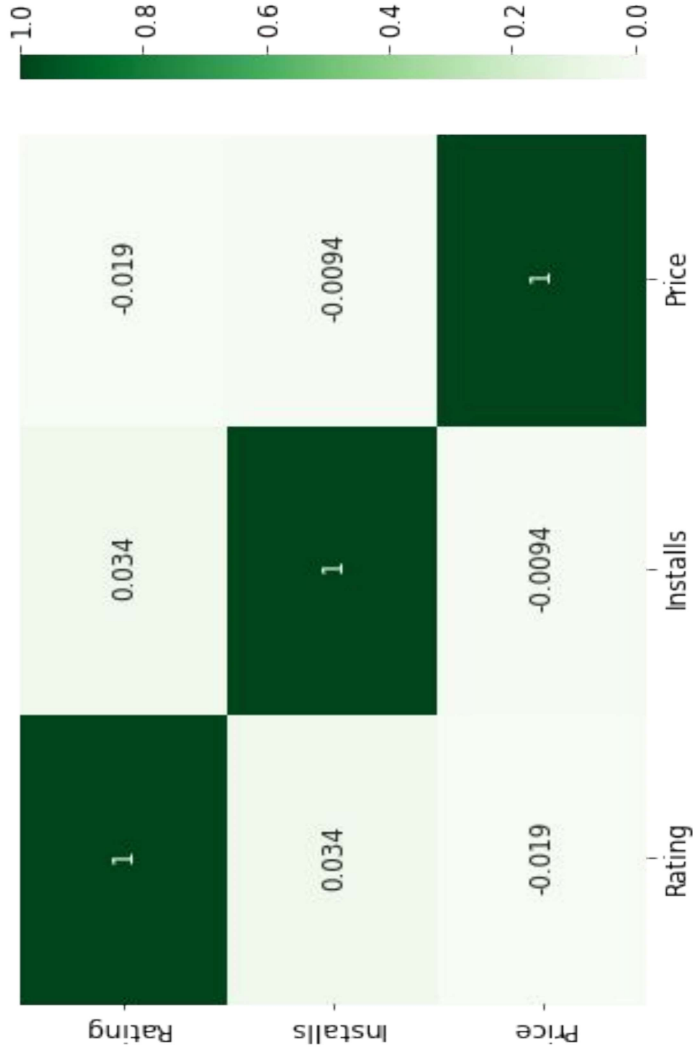
# Average Rating of Paid Apps

Again in the paid type, on an average News &Magazines, Education have high ratings. More concentration can be given to the lowest rating apps for futher modifications and updations to get inserted into the market.



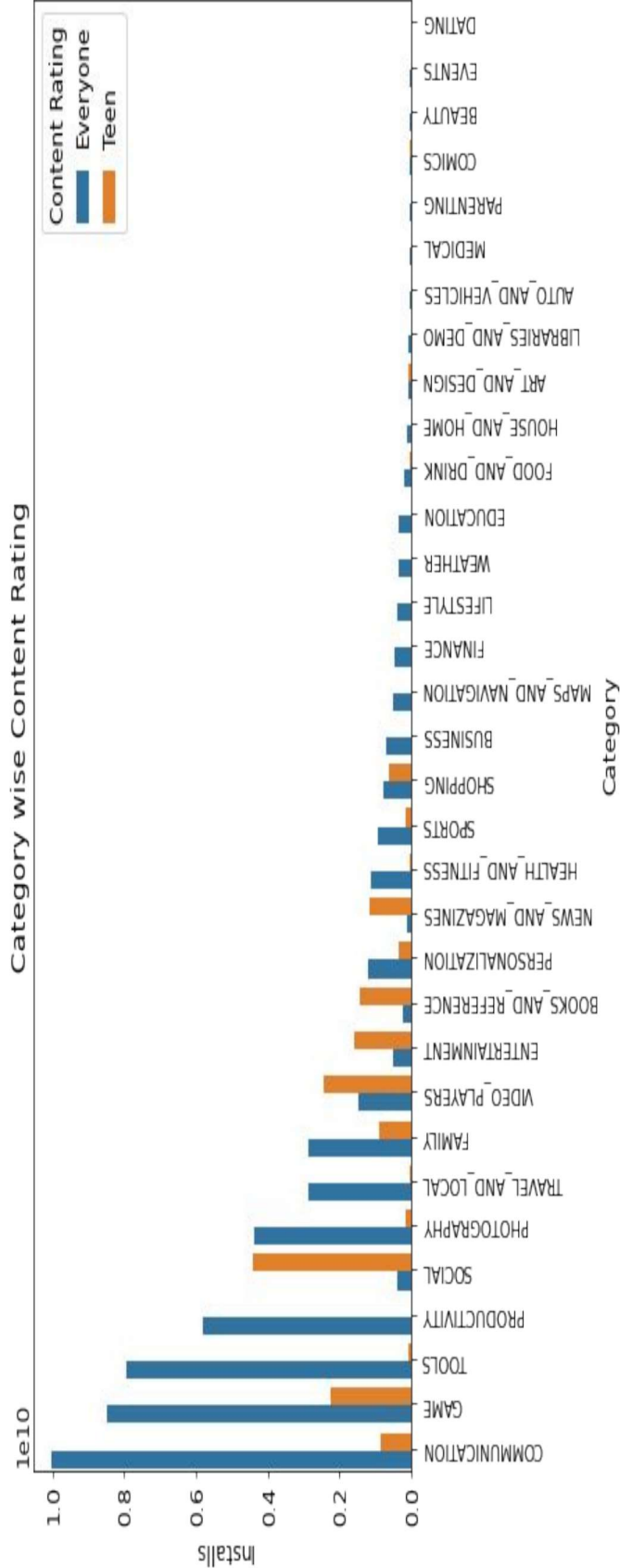
## Correlation between Ratings ,installs and Price

- Here we can interpret that price and installation of apps are negatively correlated to each other which means for increase in the prices there are slight decrease of installations
- And there is a slight positive correlation between Ratings and Installations

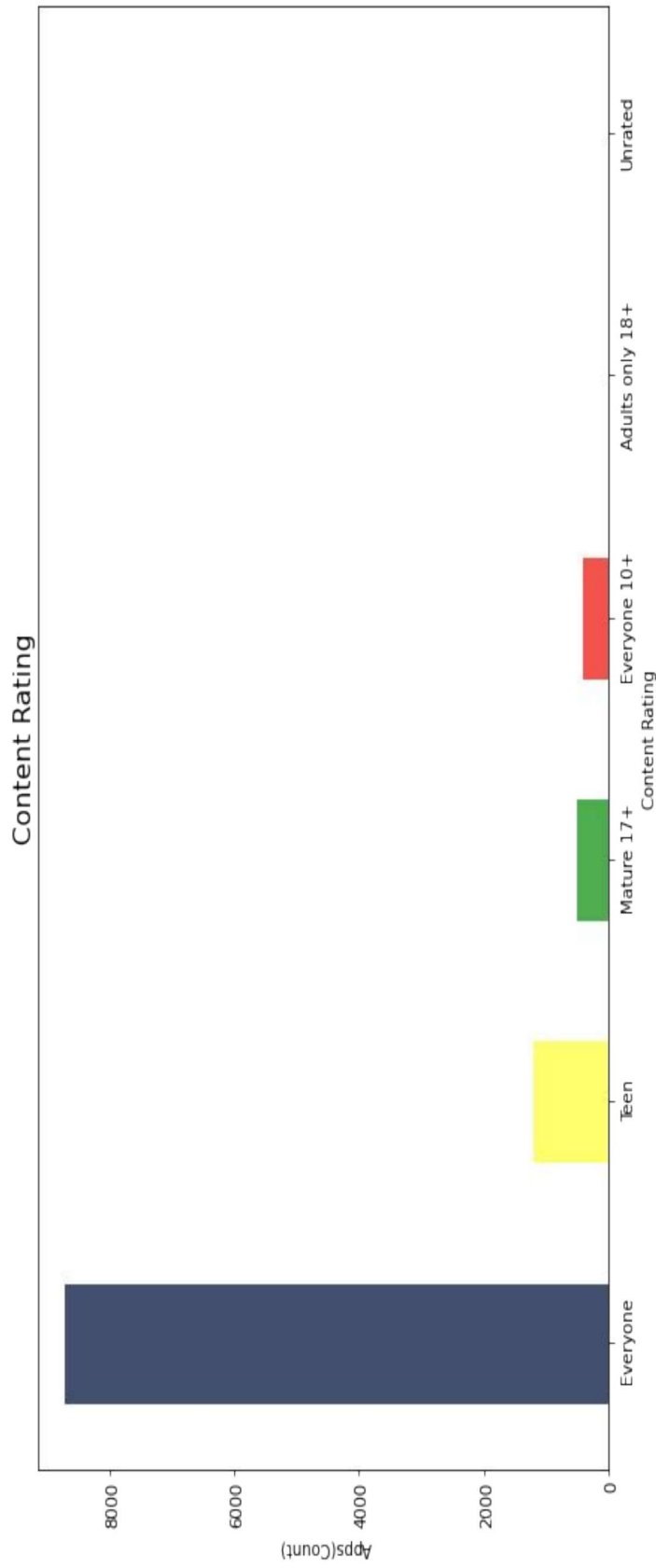


# Category wise Installed Apps with Content Rating Everyone and Teens

Here we can observe that installed apps in the Social, Games category are high with the content rating Teens



From here we can see that maximum no of apps are targeted for everyone



## Conclusions

Based on the above EDA research, we can conclude that

- The Play Store apps data has enormous potential to drive app-making businesses to success.
- App development business companies can be suggested to develop the applications with **paid criteria** in the categories like **game, tools, communication**. Developing in these can give them a better place in the market.
- Customers' essential needs, such as a few categories of apps, drive a large number of installations. These Apps stand in top 10 of the Google Play store
  - 1) Facebook
  - 2) Gmail
  - 3) Whatsapp Messenger
- Number of installation according to content rating Everyone are in communication and gaming categories. Whereas Social category apps like Facebook, linkedin, instagram are more confined to teen category.
- App seems to be which is updating more versions have more number of users.

# Q&A