Google Play

Sarah Al-Hussain

T5: Data Science Bootca

mp

**Google Play Store App Dataset
Project**

GOAL:

This project aims to help new developers
To decide what App types and characters
People prefer.

PROJECT PLAN:

EDA

This section will answer the develop questions

Model

This section will help the developer to predict the Rate and Installs
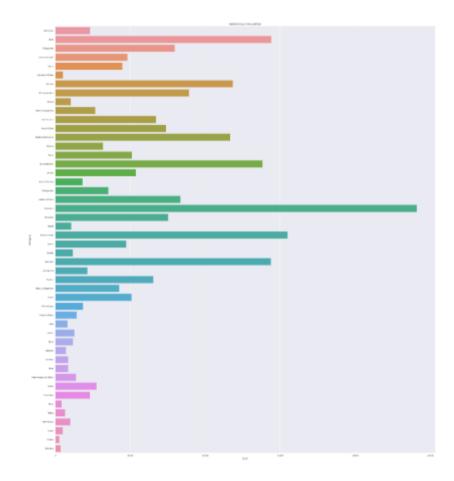
DATASET:

- The dataset contains over 1 million instances with 24 features for each.

- Features include numerical and categorical types, such as Rating, type, Developer, etc.

- Dataset founded through the Kaggle website.

- Data cleaning.

EDA

WHICH CATEGORY HAS MORE APPS?

- The most common Category is "Education" Followed by "Music & Audio" and "Comics" is The Least common category.

WHICH CATEGORY HAS MORE APPS?
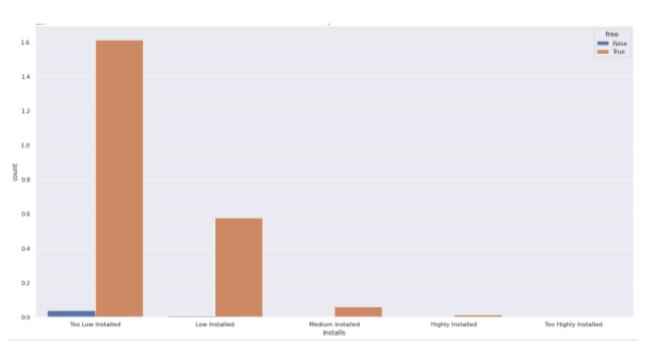
TO WHICH INSTALLS LABEL THE MAXIMUM INSTALLS APPS BELONGS?

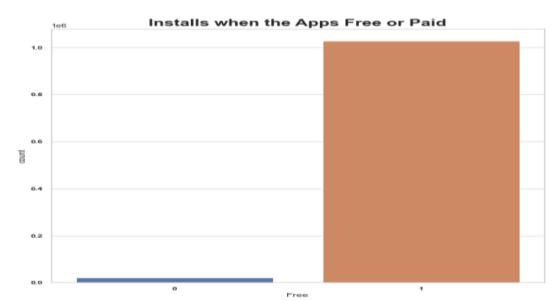The maximum Installed Apps belongs to "Too Low Installed " whether they Free or Paid

WILL THE PRICE AFFECT NUMBER OF INSTALLATIONS?
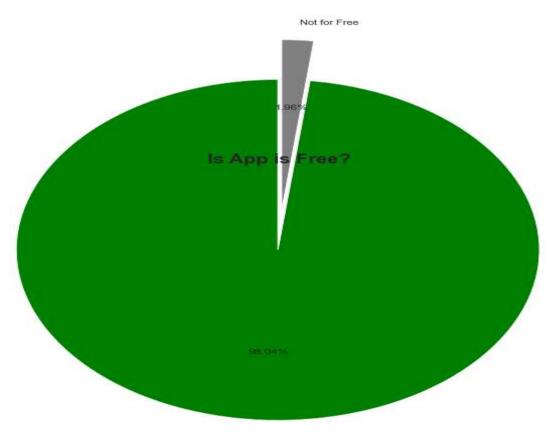
The number of installs when the app free is clearly higher than paid Apps.

```
total number of installs when it is free: 318872231712
total number of installs when it is not free: 373907463
```


Installs when the Apps Free or Paid

THE NUMBER OF INSTALLS WHEN THE APP FREE IS CLEARLY **HIGHER** THAN PAID APPS**.**
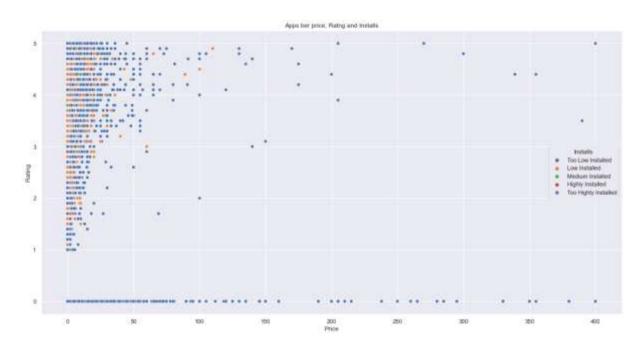
WHAT IS THE MOST DOWNLOADED APP?
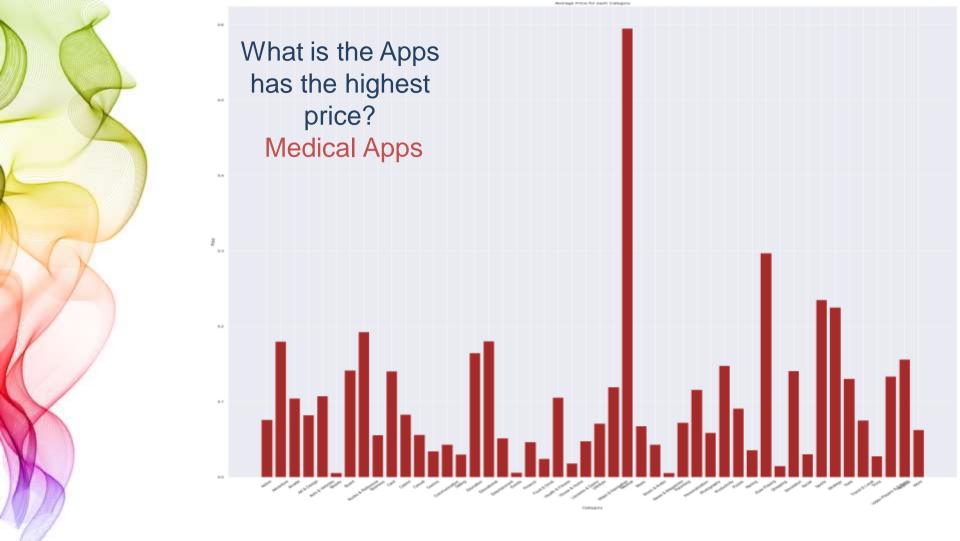
"YouTube" is the most downloaded app.

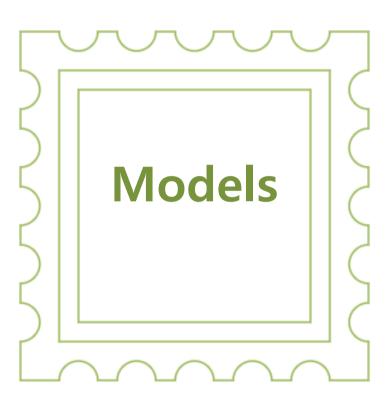| App Name	App Id	Category	Rating	Rating Count	Installs	Minimum Installs	Maximum Installs	Free	Price
YouTube	com.google.android.youtube	Video Players & Editors	4.4	112440547.0	Too Highly Installed	5.000000e+09	9766230924	True	0.0

From this scatterplot we can notice most of Apps that have higher price and have high rate (3 and above), and most of them belongs to "Too Low Installed" Apps.

Apps ber price, Rating and Installs

What is the Apps has the highest price?
Medical Apps

Models

Models

1.85   Linear Regression

0.49   Decision Tree

0.48   Random Forest

CHALLENGES:

Individual Work   Big Data   Model

THANK YOU FOR LISTINING