Capstone Project - 1
Play Store App Review analysis

Team Members

ToyajKumar Yadav
Mohammad Faiz
Data science Trainee , @ AlmaBetter.

WHY DO WE NEED TO ANALYZE THE PLAY STORE ?

Android Apps holds the most of the market of the apps around the world.

Apps market is increasing very rapidly with 21% growth every year. Growth increasing drastically since pandemic Covid-19

How an App become successful and popular among users. What are the key features ?

What are the common behavior and feedbacks of users related to the apps.

INTRODUCTION

- Google Play was launch on March 6, 2012, bringing together Android market marking a shift in Google's digital distribution strategy.
- Google Play generates the largest amount of worldwide app downloads. In the third quarter of 2022, users downloaded 27.1 billion apps from Google Play.
- Android is the most popular operating system in the world.
- Google Play store has more than 2.5 Million+ verified apps in it .
- Actionable insights can be drawn for developers to work and occupy the Android market.

- **The main Goal of our project is :**
1. The objective of our project is to Explore and analyze the data to discover key factors responsible for app's success.
2. The purpose of the project is to analyze detailed information of apps, their categories and their features which makes them successful at Play Store.

Problem Statement

- Two datasets are provided, one with **basic information** and the   other with **user reviews** for the respective app.
- We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app success.

- **So, what factors influence an app's success?**

   An app can be successful if it has :
i.   A max average-user rating of 5.
ii.   A good number of positive reviews .
iii.   Depends on the highest number of active users .
iv.   Highest number installs.

AGENDA :

➤ Introduction .

➤ Category wise play store apps installs.

➤ Which category (Genre) has the Rating, Review & downloads on the play store.

➤ Category wise most popular apps.

➤ Top 10 most installed apps in play store.

➤ Average sizes of apps in each category.

➤ Total number of Paid and free apps .

➤ Count of positive , negative and neutral reviews.

➤ Apps updating every year.

➤ Apps size (in MB) and their count.

➤ Sentiment Polarity and sentiment subjectivity of apps available on Play Store.

AI

DATASET PREPARATION

➢ **LOADING THE DATASETS:** There are two datasets, First Play store app and second dataset is User Reviews dataset.

➢ **IMPORT LIBRARIES :** NumPy, Pandas, datetime, Seaborn and Matplotlib (Pyplot sub-module)

➢ **DATA CLEANING:** Null values, Finding and removing Outliers, Removing duplicate data, removing wrong data..

➢ **DATA IMPUTATION :** Filling the missing categorical values with mode and numerical values with median.

➢ **EXPLORATORY DATA ANALYSIS:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

REMOVING DUPLICATES

CLEANING

DROPPING NaN CELLS

REFINED DATA

AI
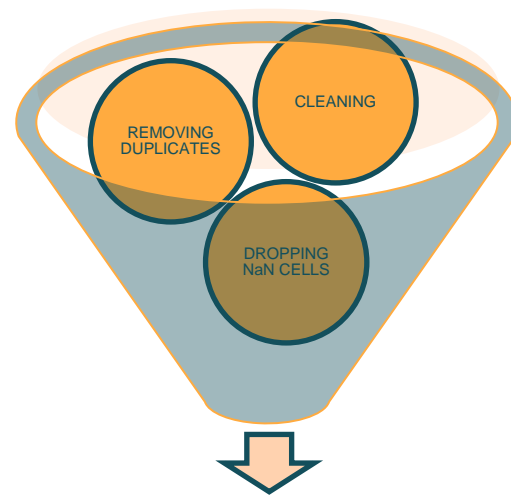
ATTRIBUTES IN GOOGLE PLAY STORE DATA

1.App : This column Contains the name of the app for each observation.

2.Category : This column Contains Category to which the app belongs.

3.Rating : This column contains the average rating for the app.

4.Reviews : This column contains the number of reviews that the app has received on the play store.

5.Size : This column contains the amount of memory the app occupies on the device.

6.Installs : This column contains the number of times that the app has been downloaded and installed from the play store.

7.Type : This column contains the information whether the app is free or paid.

8.Price: If the app is a paid app, this column contains the data about its price.

9.Content Rating: This column contains the maturity rating of the app i.e. the age group of the audience for which it is suitable.

10.Genres: This column contains the data about to which genre the app belongs. Genres can be considered as a further division of the group of Category.

11.Last Updated: Contains the date on which the latest update of the app was released.

12.Current Version: Contains information on the current version of the app available on the play store.

13.Android Version: Contains information about the android versions on which the app is supported.

ATTRIBUTES IN USER REVIEWS

1. **App-** Application name

2. **Translated Review-** User review

3. **Sentiment-** Positive/Negative/Neutral

4. **Sentiment Polarity-** Sentiment polarity score

5. **Sentiment Subjectivity-** Sentiment subjectivity score

OVERVIEW OF ANALYSIS

| DATA CLEANING	DATA EXPLORATION
Understanding the structure of the dataset and cleaning data before analysis.	Uncover initial patterns, characteristics, and points of interest using visual exploration.

AI

CORRELATION HEATMAP

- ➢ There is a strong **positive** correlation between the **Reviews** and **Installs**.

- ➢ The Installs are slightly **negatively** correlated with the **Ratings, Reviews** and **size** .

- ➢ The **Rating** is slightly **positively** correlated with the **size** and **reviews** .
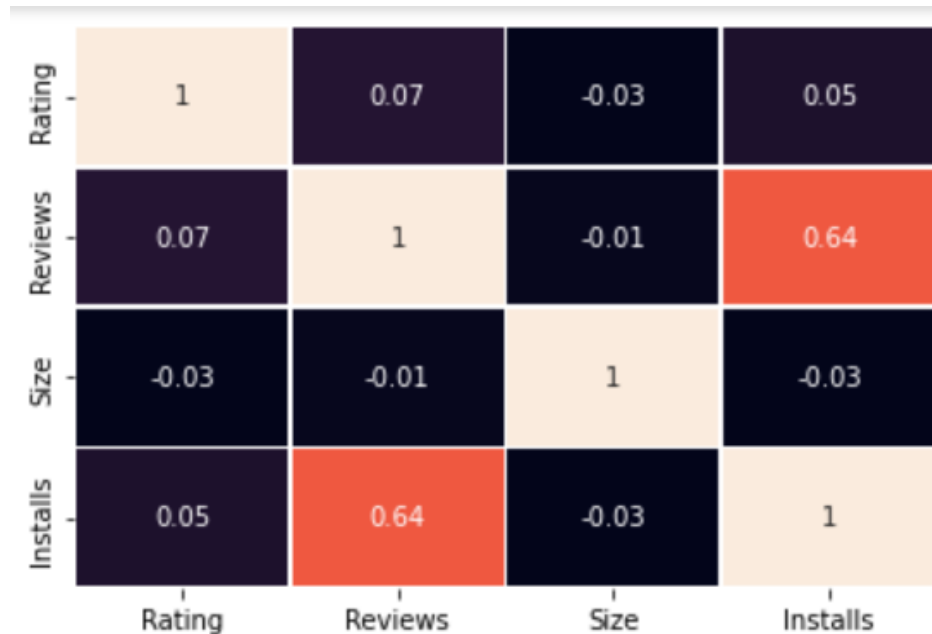
PERCENTAGE OF PAID APPS v/s FREE APPS

We Analyzed that **93.0% of Apps are free**

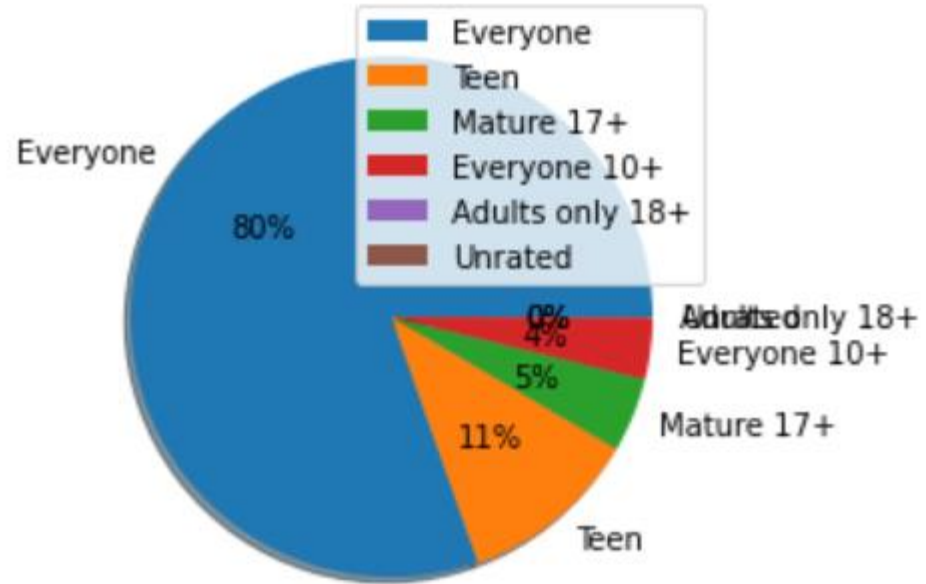 and only **7.0% of Apps are paid** in Play

store.

PAID APPS
VS
FREE APPS
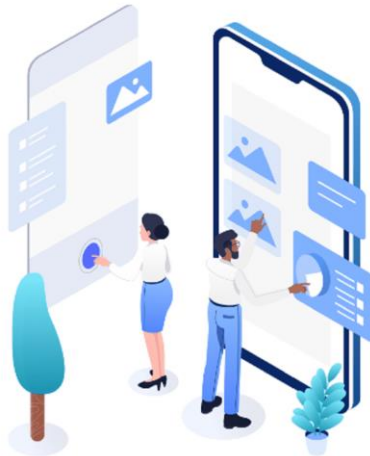
CONTENT RATING

From the plot we can see that Everyone category having majority of apps count.

Everyone apps are 80%,
11% apps are for teen,
5% apps Mature-17+,
4% apps are Everyone10+
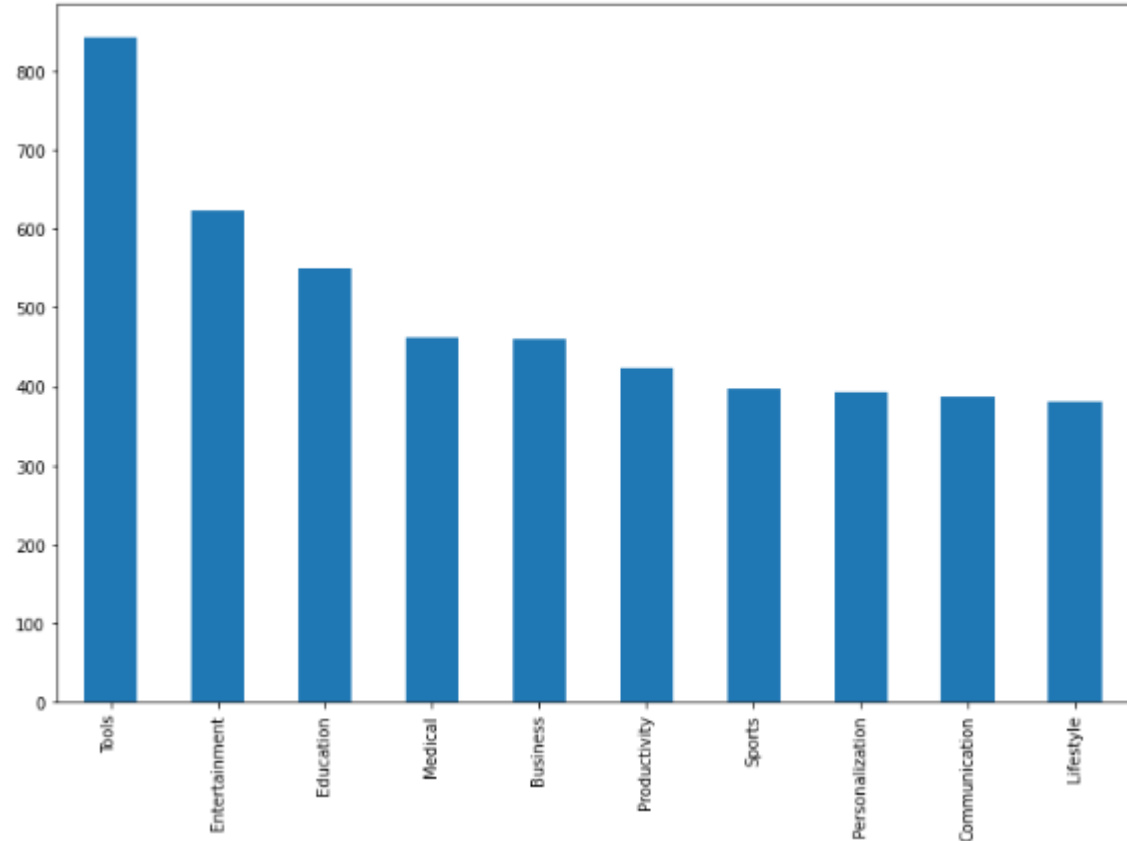and at last 0% app are Adult only18+ & Unrated.

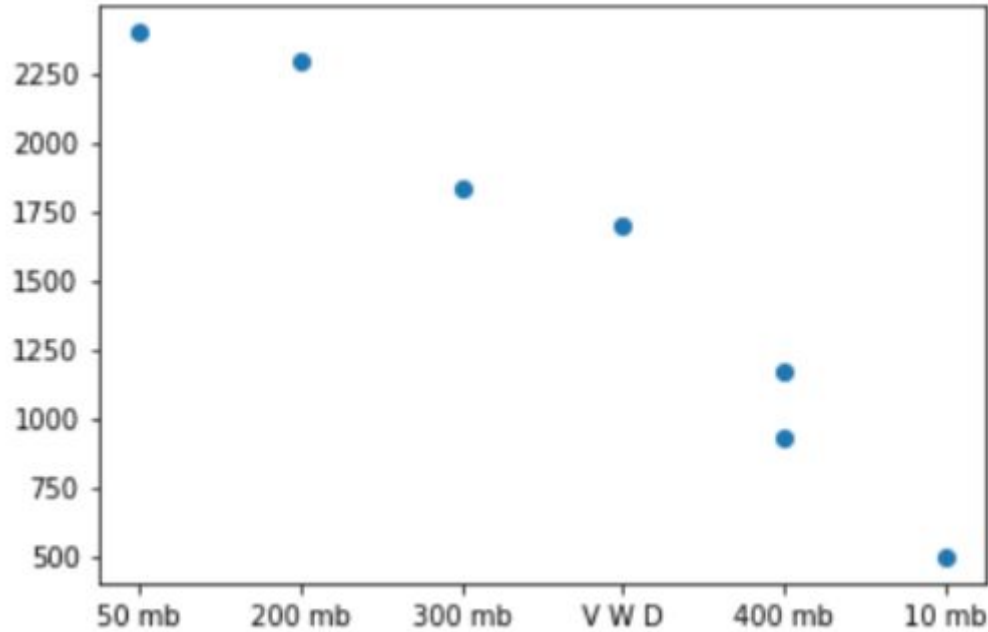# Top-10 (Genres) Categories wise play store apps Installs

AI

Tools, Entertainment, Education are the Top 3 Count of applications .

Later medical & business have almost same count .

and then count of remaining applications are in decreasing order.

App Size analysis

AI

This graph shows number of apps on the size basis.

50MB size apps are the most installed.

VWD = varies with device

Challenges Faced

❖ Reading the dataset and comprehending the problem statement.

❖ Examining the business KPIs for app development and devising a solution to the problem.

❖ Handling the error, duplicate and NaN values in the dataset.

❖ Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.

Conclusion's

93.0% apps are **Free** and **7.0%** apps are **paid** in type.

80.0% apps have **Everyone** content rating.

There is a **total of 23998 +ve Reviews, 8271 –ve Reviews and 5158 neutral Reviews** in sentiment .

Tools and Entertainment are top two categories having 842, and 623 app count.

App which the highest number of installs: **Google Play Book**

Tools, Entertainment, Education, Medical and business are top Genres.

2500 Apps are having size less than or equal to **50 MB.**

Conclusion's

It's good to develop a **Free type** app and having a content rating for **Everyone.**

Percentage of apps that are top rated = **81.80%**

There are **20** free apps that have been installed over a **billion** times

Minecraft is the only app in the paid category with over **10M** installs, and also has produced the most revenue only from installation fee.

Price, Rating, Size **has no or very less correlation** with **Sentiment Polarity**.

The median size of the apps in the play store is 12 MB

The apps whose size **varies with device** has the highest number average app installs.