Play Store App Review Analysis

Presented by-Sarthak Arora, Jay Nandasana, Arshi Wani, Pranjali Tete, Madhavi Mali

ABSTRACT

Google play store is simply entertainment at our fingertips. It's an official app store and a digital media store having enormous things to offer like books, movies, programs and music. Applications are either free or paid. Our team has worked on playstore data. This dataset contains 13 features and 10840 observations. This information can be used for predicting key factors responsible for app engagement & success story.

INTRODUCTION

Play store is not just an app store, it's a platform offering various digital content to its consumers. The Google Play Store is home to android applications, music, movies, books, games and television programs. 81% of the apps are free of cost which has led to immense popularity of this platform. As per google survey report 3000+ apps are being added every other day. The Google Play Store contains applications for the Android system only. This document reveals the dynamics of the Play Store app and gives actionable insights for the developers to work on and rule the Android market.

INTEGRAL METHODOLOGY:

First, we investigated some basic information of our dataset. On doing so we found out that our data needed some cleaning, some values were missing, and some datatypes were incorrect. We started with data cleaning and correcting the data types, followed by data visualisation. We removed some unnecessary features and made it ready for analysis using different plots.

DATASET DESCRIPTION:

This dataset has 13 features and 10840 observations

App	Name of the App
Category	Category under which it falls
Rating	Application's rating on playstore
Reviews	Number of reviews of the app
Size	Size of the app
Installs	Number of Installation of the app
Type	Whether the app is free or paid
Price	Price of the app if it's a paid app (0 if it's a free app)
Content Rating	Appropriate target audience of the app
Genres	Genres under which the app falls
Last Updated	Date when the App was last updated
Current Version	Current version of the App
Android Version	Minimum android version required to support the App.

BREAKDOWN OF DATASETS

Before proceeding to data visualisation, we need to perform the following steps:

1. Importing required packages for future analysis.
2. Mounting drive and reading data files from Google drive.
3. Removing future warnings in seaborn plots.
4. Viewing all data information.
5. Dropping duplicate.
6. Removing special characters
7. Checking unique values, null count and datatypes of each column.

8. Segregation of numerical and categorical data.

EXAMINING NULL / MISSING VALUES

Some values in our dataset are null or missing. These values affect the accuracy and performance of the models that predict the outcome, so these need to be handled. While analysing our dataset the first thing we will do is to examine the null or missing values in our dataset. This makes our result accurate.Missing values are more in Size & Rating columns as can be seen by plotting graphs. Hence several methods are used to remove these values.

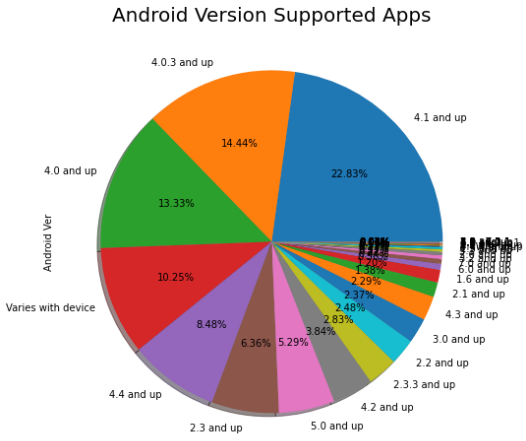
DATA CLEANING

Data cleaning is the foremost step in any data science project. Cleaner the data, better are the results. As the proverb goes by saying “More Data beats clever algorithm, but better data beats more Data” – Peter Norvig. To begin with our data cleaning, first we remove the duplicate values. Then we remove unnecessary characters in our dataset.
After doing so we find the unique values of each column and make the necessary changes in each column like converting datatypes, removing the null and ‘nan’ values. Lastly, we have done exploratory data analysis of our dataset.

DATA VISUALIZATION

Observation 1:

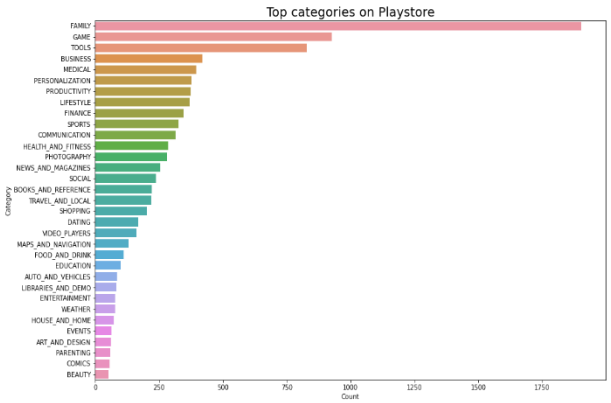
We plotted a graph of android version supported apps.



Android 4.0 and above version supported apps ratio is very high, more than 60% app's require android 4.0 version

Observation 2:

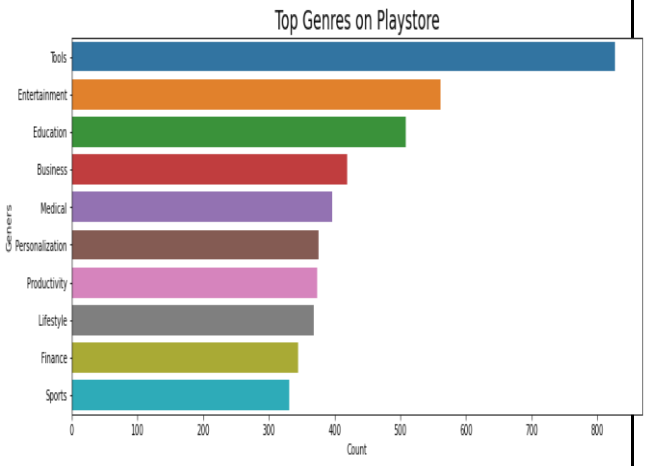
We plotted a graph of top categories on playstore.



Playstore market is ruled over by family and game category followed by tools, business, medical, personalisation and so on.

Observation 3:

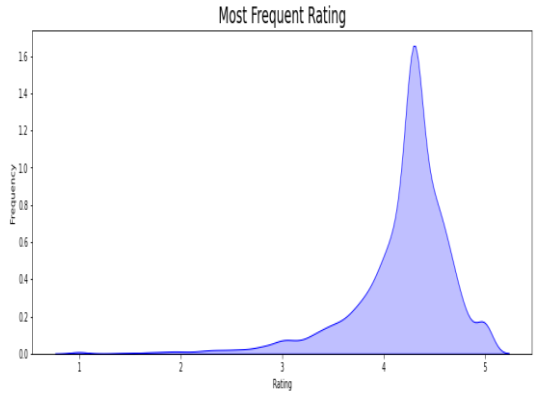
We plotted a graph of top genres on playstore.



As seen in the graph top 5 genres are- tools, entertainment, education, business, medical.

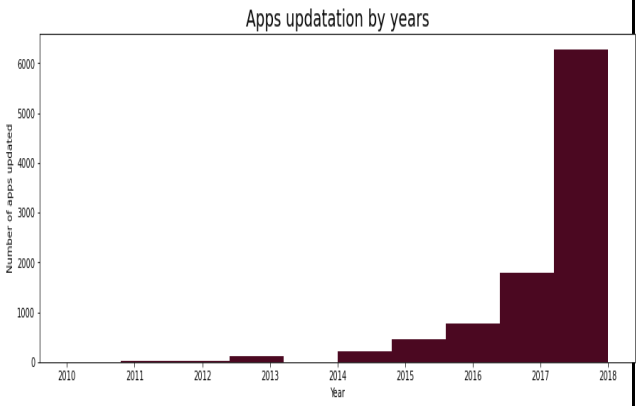
Observation 4:

We plotted a graph to check the most frequent rating of the apps.



Most apps have a rating between 4 to 5(mostly 4.3 approx.)

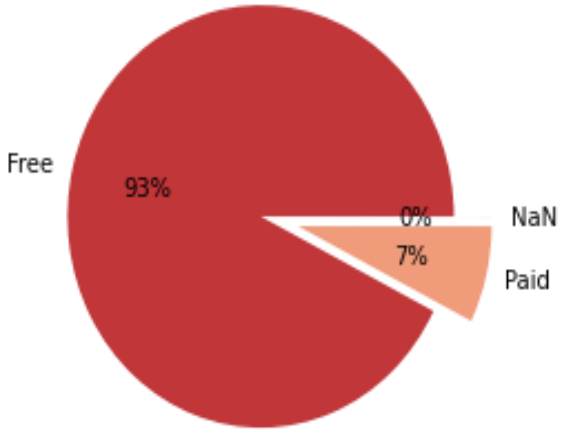
We plotted a graph between number of apps updated and year.



Most of the apps were updated during 2017-2018

Observation 5:

We plotted a pie chart to check whether the app is paid or free.

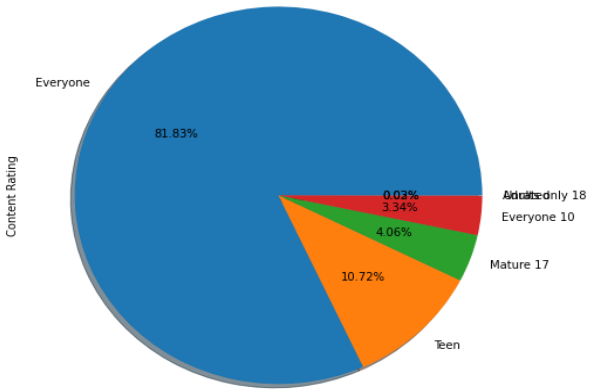


- 0.0 apps have Nan values
- 93.0% apps are free
- 7.0% apps are paid

Observation 6:

Observation 7:

We plotted a pie chart to check the content rating.



- Most of the apps are suitable for everyone
- 10.72% apps are suitable for teens
  4.06% apps suitable for mature 17

ADVANTAGES OF VISUALISATION:

- Visualised data is processed faster and easier.
- Better insights of the data are drawn which may be missed in traditional reports
- Helps us visualise trends which improve performance

CONCLUSION AND FUTURE WORK

The app developers can predict the outcome of the developed apps. Better insights are drawn from this visualisation. Apps which need to be improved can be worked upon by the developers. The dataset contains immense possibilities to improve business values and have a positive impact.
We could add a system that would create application on its own by using the data set and creating the best user interface by highly rated apps.

ACKNOWLEDGEMENT

This project is presented by-Sarthak Arora, Jay Nandasana, Arshi Wani, Pranjali Tete, Madhavi Mali.
Our sincere efforts have made us to accomplish the task of completing this project. We are extremely grateful to our instructors and mentors who have helped us to grow in this field. We would like to express our sincere gratitude to the celebrated authors whose phenomenal work has been consulted and referred in our project work. We also wish to convey our appreciation to our peers who provided encouragement and timely support in the hour of need. This project helped us improve our skills and enhanced our knowledge.

REFERENCES

https://seaborn.pydata.org/tutorial/color_palettes.html https://www.python-graph-gallery.com/196select-one-color-with-matplotlib
https://www.analyticsvidhya.com/blog/2021/05/10-colab-tips-and-hacks-for-efficient-use-ofit/