Project Report

Google Play Store Data Analysis

Abstract

The google play store is one of the largest and most pop-

ular Android app stores. It has an enormous amount of data

that can be used to make an optimal model. We have used a

raw data set of Google Play Store from the Kaggle website.

This data set contains 13 different features that can be used

for predicting whether an app will be successful or not using

different features. This data set is scraped from the Google

Play Store. This journal talks about different classiﬁer mod-

els that we used for prediction purposes and ﬁnding which

one gives the highest accuracy. This journal also gives de-

tailed information on feature extraction and the complete

Data visualization done on this data set. Our project code

can be found at https://github.com/Rimshamaredia/CSCE-

421-Project.

Abstract

The google play store is one of the largest and most pop-

ular Android app stores. It has an enormous amount of data

that can be used to make an optimal model. We have used a

raw data set of Google Play Store from the Kaggle website.

This data set contains 13 different features that can be used

for predicting whether an app will be successful or not using

different features. This data set is scraped from the Google

Play Store. This journal talks about different classiﬁer mod-

els that we used for prediction purposes and ﬁnding which

one gives the highest accuracy. This journal also gives de-

tailed information on feature extraction and the complete

Data visualization done on this data set. Our project code

can be found at https://github.com/Rimshamaredia/CSCE-

421-Project.

Abstract

The google play store is one of the largest and most pop-

ular Android app stores. It has an enormous amount of data

that can be used to make an optimal model. We have used a

raw data set of Google Play Store from the Kaggle website.

This data set contains 13 different features that can be used

for predicting whether an app will be successful or not using

different features. This data set is scraped from the Google

Play Store. This journal talks about different classiﬁer mod-

els that we used for prediction purposes and ﬁnding which

one gives the highest accuracy. This journal also gives de-

tailed information on feature extraction and the complete

Data visualization done on this data set. Our project code

can be found at https://github.com/Rimshamaredia/CSCE-

421-Project.

**Overview:**

The google play store is one of the largest and the most popular android app stores. It has an enormous amount of data that can be used to make the various prediction and Analytical models. This dataset consists of 13 different features which is used to for data analysis.

**Introduction:**

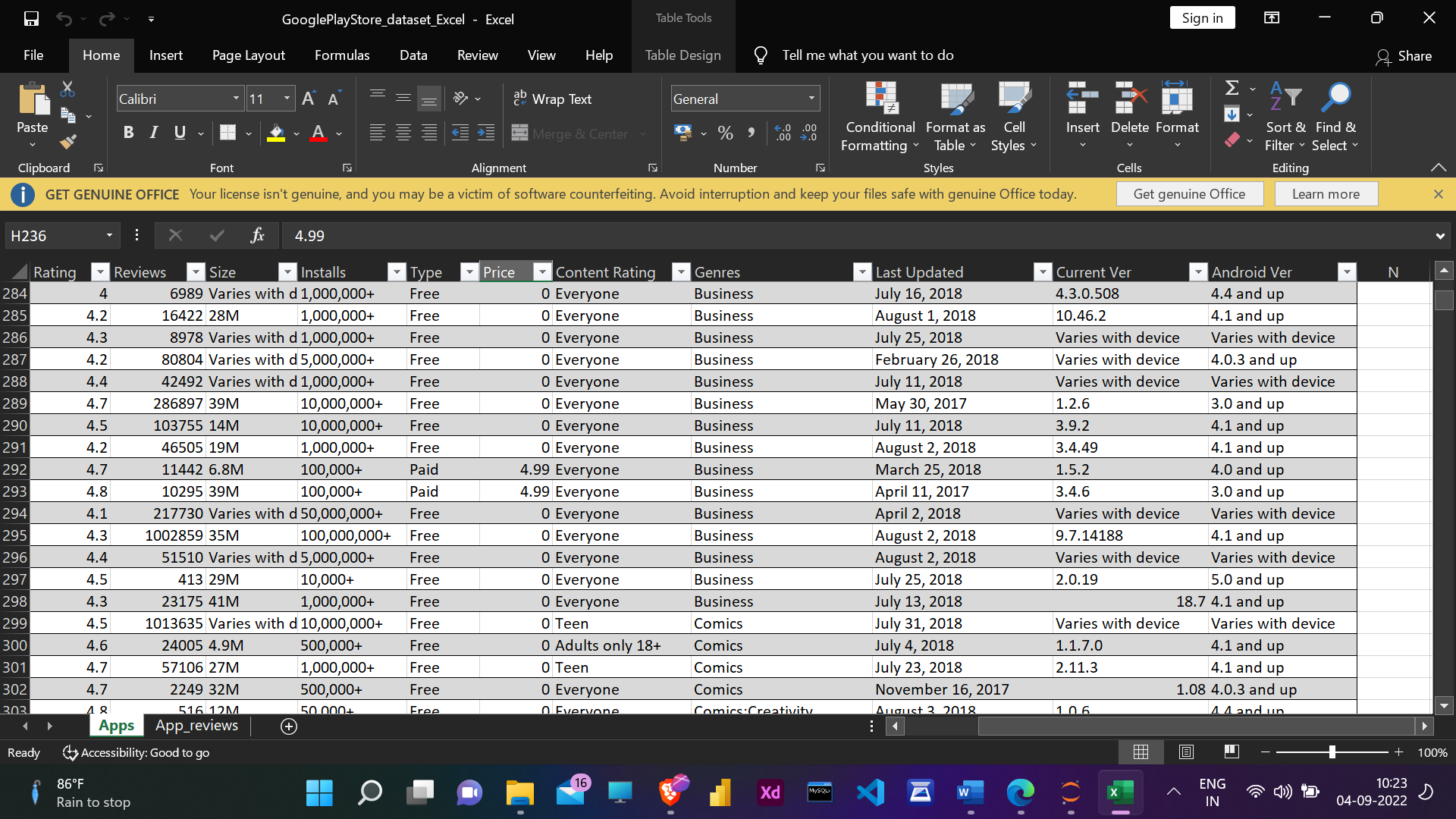
Mobile application are one of the fastest growing segments of downloadable software application market. Out of all of the markets I choose Google Play Store due to its increasing popularity and recent fast growth. One of the main reasons of its popularity is the fact the about 81% of the apps are free of cost. Developers and users play key roles in determining the impact that market interactions have on future technology. However, the lack of a clear understanding of the inner working and dynamic of popular app markets impacts both the developers and users.

**Analysis Methodology:**

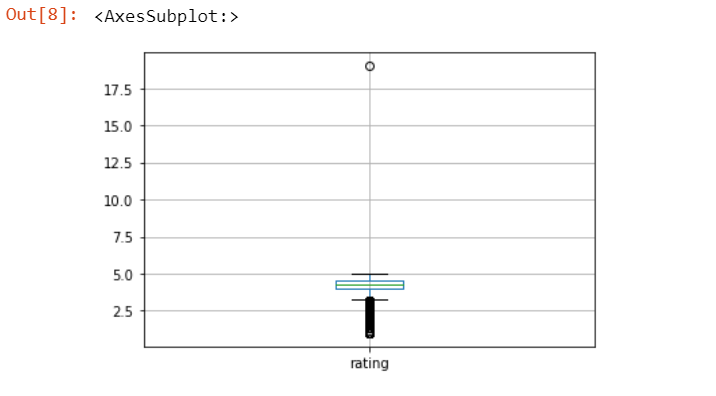
Our analysis approach is divided into three phases:

1. Data extraction (From the provided dataset)
2. Data cleansing
3. Visualization
4. Dashboard / Report Making

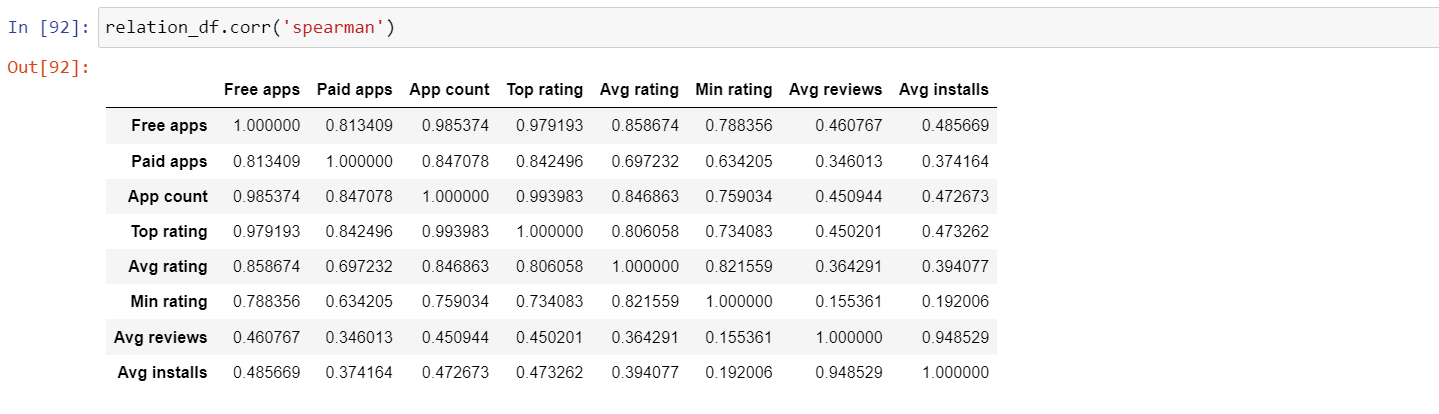
**Data Extraction** 🡪 This process deals with data collection in this we gathered the dataset provided by Ineuron.

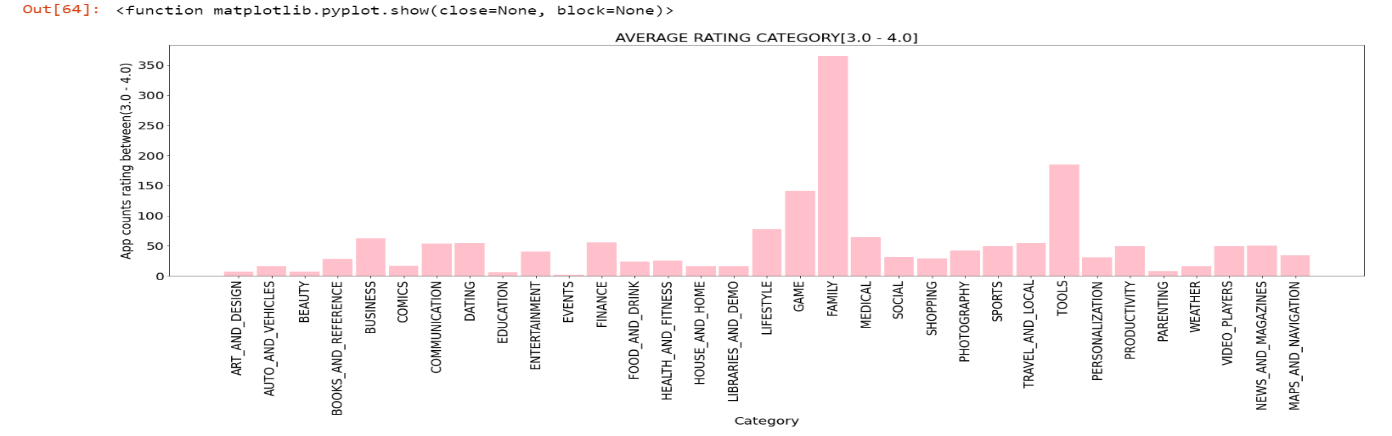


**Data Cleansing** 🡪 This process deals with data manipulation, data which is collected in the previous step consist of NaN values, unwanted Column, detecting outliers. Handling this error is also known as Exploratory Data Analysis (EDA).



**Visualization** 🡪 This step is also one of the part of EDA, Generally after data cleansing. Various observations are performed over the data in order to find some relationships amongst the (feature – feature) or (feature – labels). And during this process we use different types of chart, graph in order to make better visualizations.





**Story Lines covered through out the project:**

1. Mostly which category of apps are installed/ downloaded by year?
2. What is the count of apps in each category by year?
3. What is the count of Paid & Free apps in each category?
4. Revenue Generated from Paid apps by year?
5. What is the high, average, minimum rating category?
6. What is the Avg installs and Avg reviews of each category by year?
7. Find Strength and Direction of relationship between the features?

The market has in-

creased to over 845900 Apps and 226,500 unique sellers in

April 2013 [2].

C sd clk

The google play store is one of the largest and most pop-

ular Android app stores. It has an enormous amount of data

that can be used to make an optimal model. We have used a

raw data set of Google Play Store from the Kaggle website.

This data set contains 13 different features that can be used

for predicting whether an app will be successful or not using

different features. This data set is scraped from the Google

Play Store. This journal talks about different classiﬁer mod-

els that we used for prediction purposes and ﬁnding which

one gives the highest accuracy. This journal also gives de-

tailed information on feature extraction and the complete

Data visualization done on this data set. Our project code

can be found at https://github.com/Rimshamaredia/CSCE-

421-Project

The google play store is one of the largest and most pop-

ular Android app stores. It has an enormous amount of data

that can be used to make an optimal model. We have used a

raw data set of Google Play Store from the Kaggle website.

This data set contains 13 different features that can be used

for predicting whether an app will be successful or not using

different features. This data set is scraped from the Google

Play Store. This journal talks about different classiﬁer mod-

els that we used for prediction purposes and ﬁnding which

one gives the highest accuracy. This journal also gives de-

tailed information on feature extraction and the complete

Data visualization done on this data set. Our project code

can be found at https://github.com/Rimshamaredia/CSCE-

421-Pro