PLAY STORE APP REVIEW ANALYSIS

VINIT LADSE, PRATIKSHA KHARODE,

GAURAV BHAKTE

EDA CAPSTONE PROJECT

ALMABETTER,BANGLORE

ABSTRACT :

- The google play store is one of the largest and most popular 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 models that we used for prediction purposes and finding which one gives the highest accuracy.

- This journal also gives detailed information on feature extraction and the complete data visualization done on this data set.

PROBLEM STATEMENT :

- Google play store is mostly use app store worldwide also top global market share.

- My main objective is to find key factor responsible for app success and engagement of users.

- Thousands of new app regularly update play store of different Category.

- I find distribution of every app based on their size, installs, reviews and much more.

DATA SUMMARY :

- The dataset provided was cleaned but we found some of the missing and null values in the column. By using info( ) method with "play_store_data_copy.info()" which shows total 13 data columns. We found that –

- The dataset has a shape of (10841, 13) but for proper analysis we've taken 9360 columns and 13 rows.

- We have use the following column from the dataset:
    - App
    - Category
    - Rating
    - Review
    - Size
    - Installs
    - Type
    - Price
    - Content Rating
    - Genres
    - Last Updated
    - Current Ver
    - Android Ver

INTRODUCTION :

Mobile industry growing rapidly, competition for apps also grown significantly so developer need to do enough research to make app success.

The Google Play Store is found to be the largest app market in the world. It has been observed that although it generates more than double the downloads than the Apple App Store but makes only half the money compared to the App Store.

We perform Data Cleaning over the dataset. Further we divided our project in Four main parts i.e Analysis on Play Store Data and Reviews Data,Analysis based on Cancellations, General Analysis,

Data Visualization. After the data set is ready, we try to analyze the data set using different plots and remove the stuff not needed from the data set.

STEPS INVOLVED

- Understanding The Data

  Before get started with the project we firstly look up at dataset. The dataset was pretty cleaned. The column doesn't have any string, list and dictionary values. By looking to data we came up with many creative ideas that how we can deal with those column.

- Discusing Problem Statement

  After analyzing the datasets we discussed with every single problem to overcome it. We all decided to divide our task and initialized with our own problem statement. The problem statement were based on target variable we took for analysis.

- Data cleaning

  The next task was data cleaning which was easy with this dataset. As mentioned in above points the data were float64 dtype , int64 dtype, object dtype,datetime64. Some of the column like Android Ver, Current Ver ,etc was having null values. So, we decided to keep on hold for further analysis .

- Exploratory Data Analysis

  After data cleaning it was sure to target some important columns for Exploratory Data Analysis. Matching the data with correct suitable problem by python libraries to result some insightful visualization was great task. These also gives us a more information from different charts and graphs.

- Visualization of analysis :

  The EDA parts make more clear about data in a picture and graphical form. Mainly we perform matplotlib and seaborn libraries of python for the data analysis. The libraries helps a lot with bar charts, pie charts, dist plot, scatter plot, box plot, line charts and many more.

- Results and Conclusions

The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product.After analysing the dataset we have got answers to some of the serious and interesting questions which any of the android users would love to know.

DATA ANALYSIS BY ALL TEAM MEMBERS

Analysis Based On Play Store Data By Vinit Ladse

General Analysis

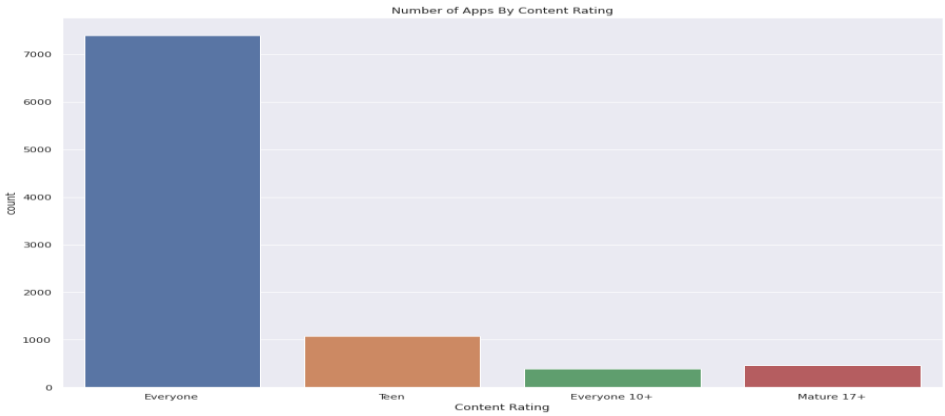
Data Description

- By using head() method we explore the top rows.
- By using info() method we find out the data type of all columns.
- By using shape function we find out number of rows and columns.
- By using describe() method compute and display summary statistic for a python dataframe.

Sanity Checks

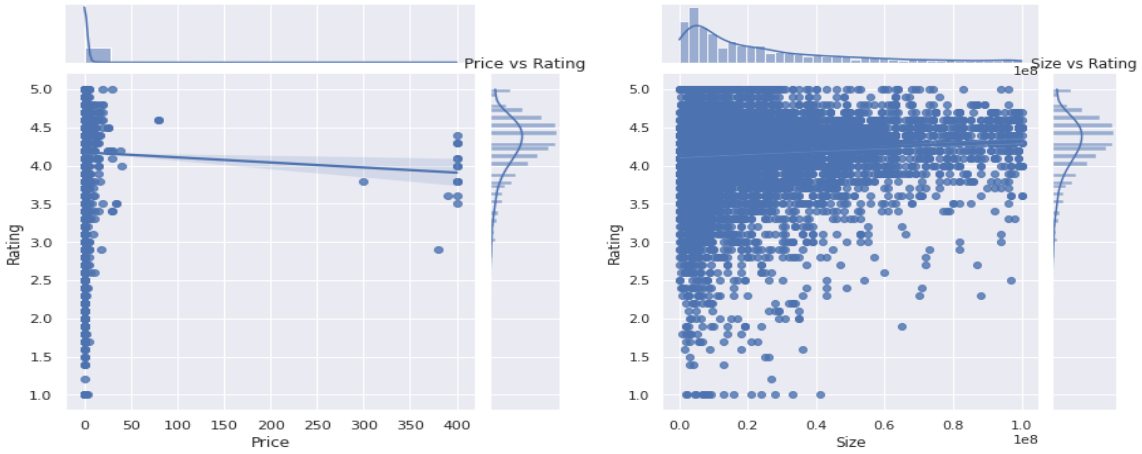
- Check for the following and handle accordingly.

Top Content Rating



In the following plot of top content rating here adult only 18+ and unrated apps have very few record so drop dose as they won't help in the analysis here everyone , teen , everyone 10+ , mature 17+ those are top content rating values .

Effect Of Price And Size Vs Rating



Joint plot to understand the effect of size on rating.The apps of lower size are rated distributely

They also contribute to the lowest rated apps of greater size are rated better as compare to the app of small sizes the frequency of the small app is also weight to apps higher then the app with large sizes .The maximum number of apps is around twenty thousand in size the regression line shows on upward trend as the size of the app increases which shows that the apps with greater size provide a better experience to the user and price goes up but there are very few apps with higher but they are very good.

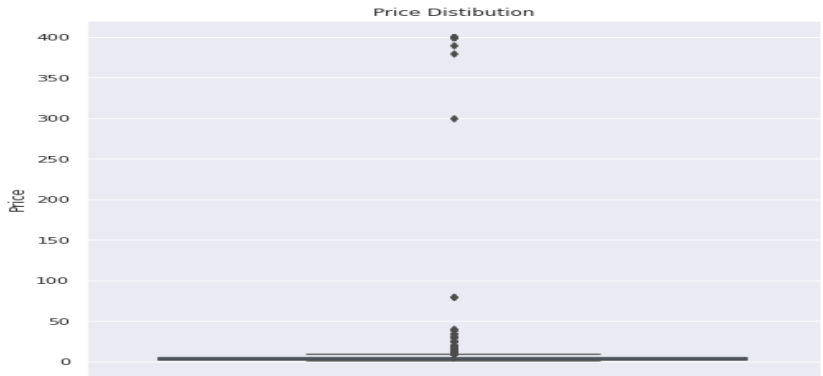
Analysis On Play Store Data By Pratiksha Kharode

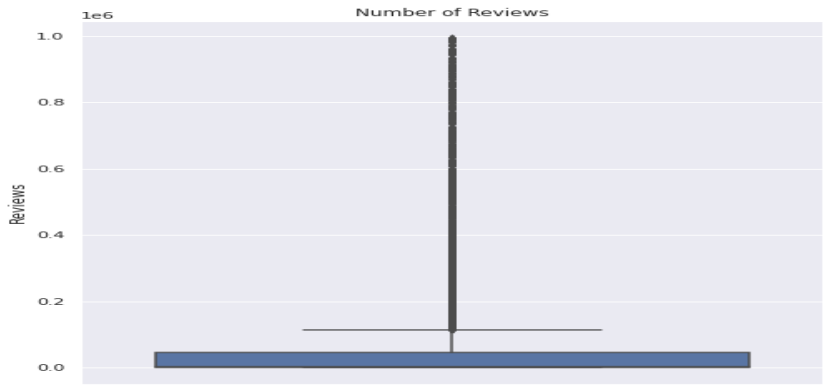
Analysis Based on Cancelation

Drop the missing values

- By using dropna() method drop the missing values in dataframe

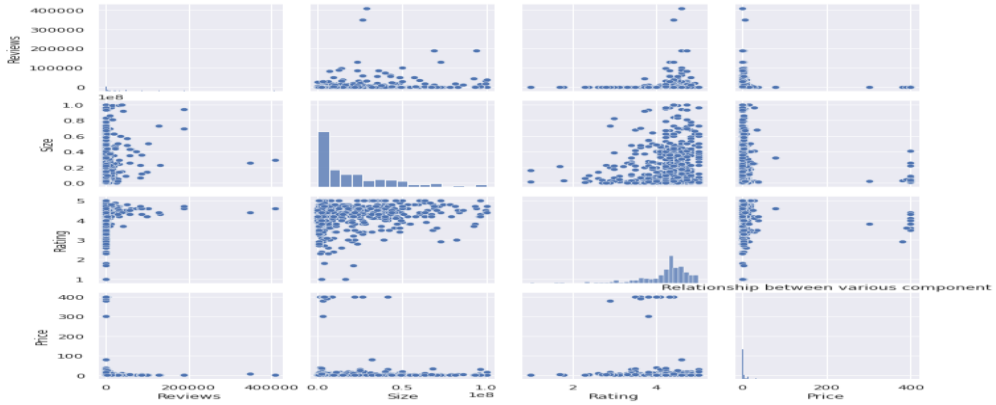
Cancellation of outliers





- Price distribution suitable graph to indentify outlier in price
- Number of Reviews here we can find out the maximum outliers in reviews and minimum outliers in reviews .

Make a pair plot with the column



- Further pair plot read with the column reviews size , rating , price .

Create five bucket based on size

Analysis On Play Store Data Reviews By Gaurav Bhakte

Google play store review sentiment analysis it shows sentiment polarity on x axis an sentiment subjectivity on y axis
In this scatered plot shows :

- Positive sentiment as red colour
- Negative sentiment as blue colour
- Natural sentiment as white colour

From the above scatered plot it can be concluded that sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of cases, it show a proportional behaviour when variance is to high or low .

Inside From Data

- ### World cloud
  World cloud is a data visuilization technique use for representing text data in which the size of each word indicate its frequency importance.

- ### Senitment Polarity
  The polarity of a sentiment measure how negative or positive the content is in the data that we have , the polarity ranges from -1 (most negative) to +1(most positive).

CONCLUDING STATEMENT :

The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product. Dataset can also be used to look whether the original ratings of the app matches the predicted rating to know whether the app is performing better or worse compared to other apps on the Play Store.