**PLAY STORE APP REVIEW ANALYSIS**

**PARUL SRIVASTAVA**

**Data science trainee,**

**AlmaBetter, Bangalore**

**Abstract:**

The play store application is the popularly used application by the android users to install wide variety of applications with ease. Here we were provided with data set of google play store app and user view and the analysis was done based on that to find out the play store application success.

With the increasing competition it becomes crucial for us to analyze this data and increase the success ratio of google play store application further more in the near future.

***Data Frame: Play store app review analysis, user reviews.***

**1.Problem Statement: -**

Data provided by google play store application gives us brief about the various category available on it, applications with different size, the number of installs, rating provided by the users and how it has affected the google play store. What can be done further to increase the demand of google play store in market place, what are all the factors responsible for its success.

Because at the end for developers all they need is that their app should perform well or else if there is nothing new or better then the rest then there is no point in building the application.

The main objective is to do exploratory data analysis, which could help them in predicting the user’s requirements proactively. The data frame had the following fields: -

* Applications: Comprises of various applications present in data frame.
* Category: The category of applications.
* Rating: The rating provided by the customers.
* Size: The size of the application in mb/kb.
* Installs: Installs of each application and its count.
* Type: The application type whether it is free or paid.
* Price: Price of the application.
* Content Rating: Given by the users.
* Genres: Genres is about application field type whether it is lifestyle, gaming, education etc.

**2. Introduction: -**

### With the advancement of technology, the need for the mobile applications for any use is also increasing rapidly. So, the developers are also interested in developing applications that add a value content to the play store.

### Out of all the applications google play store is preferred more because of its increasing popularity and rapid growth in all the years. Developers and users are the main factors ion deciding how play store will impact the future. We have focused on the dynamics of google app play store in finding out the factors that had make it successful and improvements that can be inculcated for much better performance.

### Our goal here is to firstly perform data wrangling and to know about data frame. After that the data cleaning is done to remove and detect any replicated values or null/nan values which are then removed to avoid redundancy. Once this is done then the observations are made that which is the highest rated apps, what are all the apps with five -star rating, whether the app is paid or not and several others. After the basic observations we performed data visualization to visualize the data with the help of the bar chart, pie chart, scatter plot, joint plot and bar graphs.

## **3. Analysis: -**

* Data Cleaning
* Data Visualization (Play store data)

### In today’s scenario with the increasing demand of the mobile phone the demand for the mobile applications is also increasing.

### The reasons for increase in demand are:

* More mobile users
* Increase in mobile availability
* Wide range of applications to choose
* Freely available.

# **4. Google play store data set: -**

The data set provides all the necessary information that is required by any data scientist to perform analysis.

Data set is usually taken from Kaggle website where various data sets are available for analysis to be performed by anyone. The data set can be used to draw meaningful insights from it for future purposes. Like how developers can develop an application that will lead more user attention and more downloads from play store in return, hence increasing its revenue in return. Not only this it will help them to know more about the customer requirements to customize any application if and when needed.

Whether an application is paid or not will it succeed or not all the relevant inferences can be drawn by analyzing the data set. Also, will it increase the overall performance of google play store or not relating with the current scenario is done. This will play an important role in deciding what can be done more and in a better way by the developers which will benefit not only users but also the play store application, in increasing the revenue of it in the near future.

**5. Steps involved:**

* **Exploratory Data Analysis:**

After loading the data set, we performed this method by comparing our target variable that is Content rating with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

* **Basic observations**

We got the inference that which app has the maximum reviews, which apps have the highest rating, whether application is paid or free, all the apps with five - star rating, apps with the maximum installs, apps with the maximum reviews and average rating of the apps.

* **Data mining**

Data mining is the process of extracting data from the scratch. Here we imported the csv data provided to us and performed analysis on it.

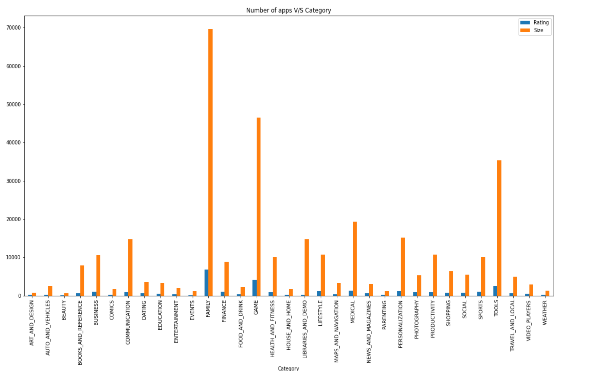
Next, we analyzed how this data will help us in determining factors about content rating, size, installs, reviews and various others.

* **Sentiment analysis:**

It is performed to know the sentiment behind the data. Generally, the sentiment can be positive, negative or neutral based on the data.

**7.1. EDA:**

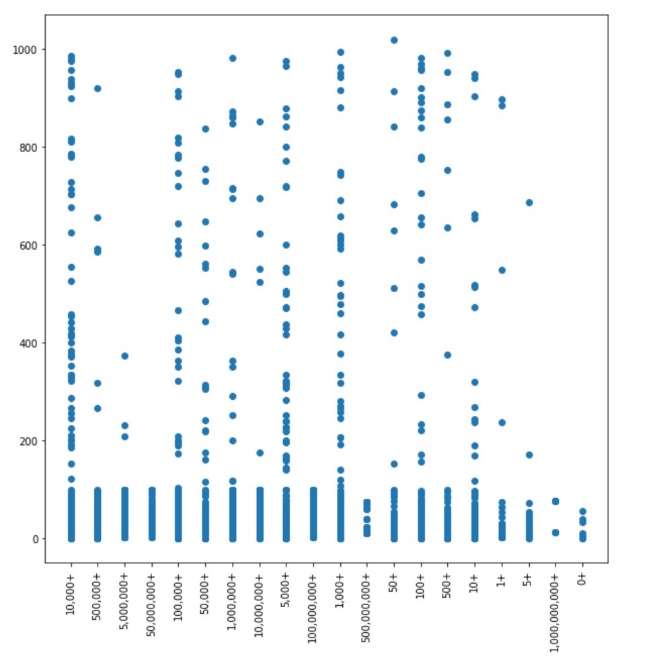
1. **Topmost categories on play store: -**

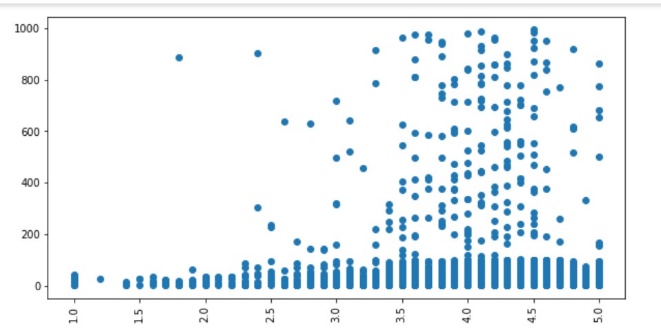


The observations made from this is that the category Family is the most widely used category by play store users. After that, Game and Tools is the highest used category by play store users, followed by business medicals etc.

1. **Relationship between Installs, Size and Rating column:**

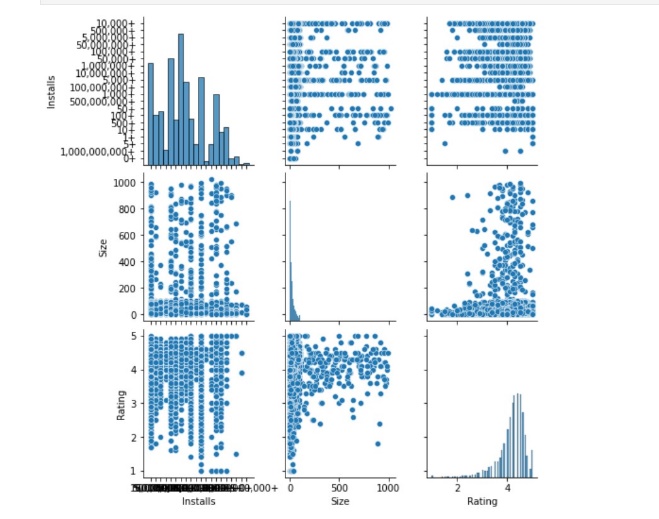
From the below first scatter plot, the inference made is that lesser the size of an application is the more likely it is to be installed as compared to the applications with more size. And from the second scatter plot inference is that higher rating is obtained by low smaller size applications compared to larger size applications.





1. **Using pair plot to further depict relationship between size rating and install: -**

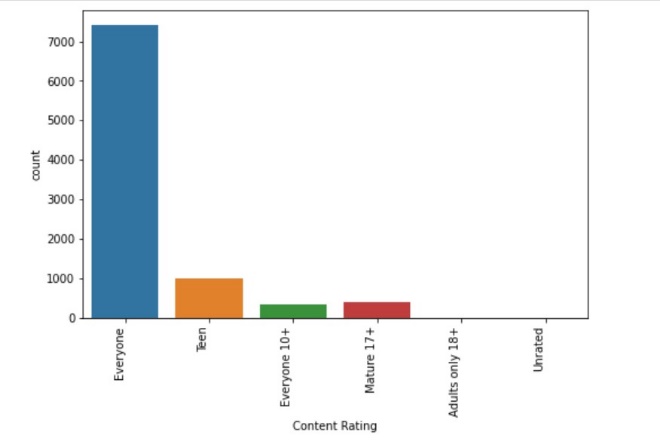
To understand relationship, we have following pair plot-

****

From the above scatter plot the inference drawn is the relationships among the various variables namely Installs, Size and Ratings. Dots is used to represent them, they are basically used to monitor how changing one variable affects the others. As here we can see greater the size is more are the number of installs as well as reviews.

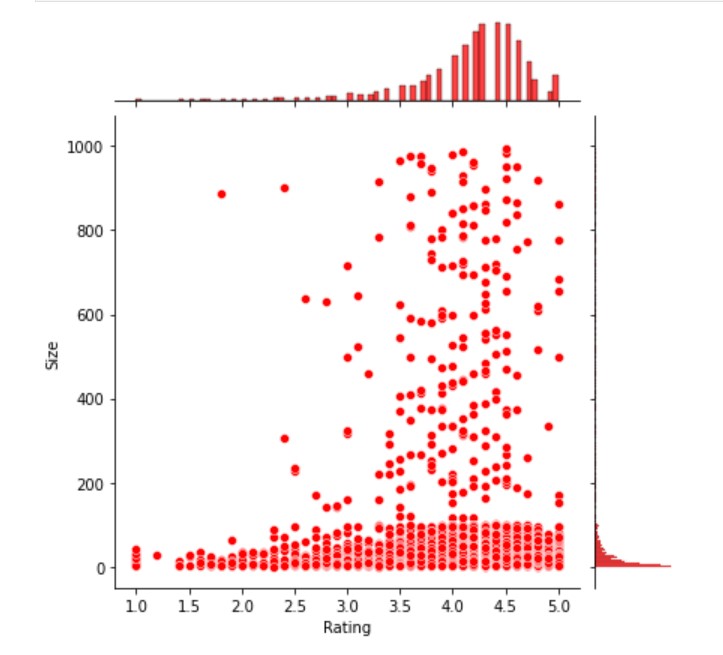
**4. Content Rating:-**

From the above count plot of Category "everyone" is the widely used set of content rating used by the developers making it more age-friendly. Most of the applications can be accessed by every age group.



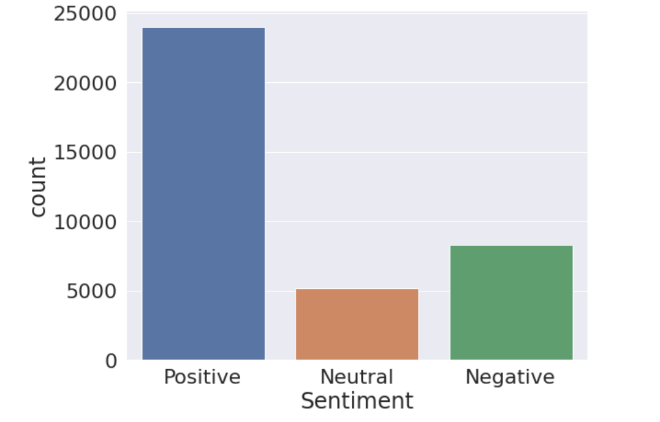
**5. Joint plot: -**

From this joint plot the inference is that the applications having size around 30MB had the rating between 4.0 to 4.5MB. And so we can conclude that applications which are smaller in size, like less than 20MB, can have a 5.0 rating.



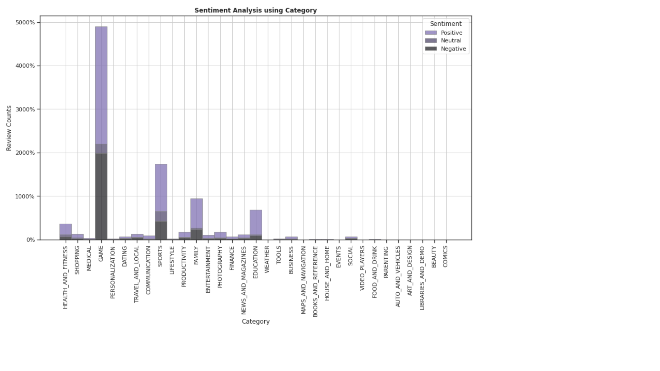
**6. Sentiment analysis :-**

The graph shows maximum sentiment are positive, positive sentiment shows user satisfaction with the application. Then comes the second highest as negative sentiment and neutral sentiment is the last one.



**6. Sentiment analysis using category data:-**

The maximum review count is of games be it positive or negative, followed by the sports and after that education. The following is the graph for it:



**Conclusions:**

* The conclusions drawn from the play store application is that it has various applications present for all the purposes.

* What makes an application more successful is that it needs to be age friendly. So, it can be used by every age group.
* Also, other factors which influenced users were positive sentiments, higher number of installs and better review.
* And free applications were more preferred and used by the users.
* The number of installs increases for the applications with higher number of ratings.
* Average rating of application turned out to be 4.16

**References:-**

1. AlmaBetter Video lectures and live classes.
2. GeeksforGeeks
3. Towards Data Science Website