**Play Store Apps Review**

**Sahib Hingorani**

**Data Science trainee, AlmaBetter**

**ABSTRACT:**

Google Play Store, also known as Android Market place, where the user can install different kinds of applications according to his/her needs It is the official distribution storefront for Android applications and other digital media, such a music, movies and books, from Google. User has ample number of choices for the apps for a particular usage.

This experiment can help users as well as app developers to understand the data according to their needs.

For users it will help them to understand which application is better among the rest.

For app developers it will help them to understand the market and current needs of the user.

**1. Problem Statement:**

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.

Each app (row) has values for category, rating, size, and more. Another dataset containscustomer reviews of the android apps.

* Play store app reviews dataset was divided into 13 subsets.
* User reviews dataset was divided into 5 subsets.

**2. Introduction:**

Google Play Store, also known as Android Market place, where the user can install different kinds of applications according to his/her needs It is the official distribution storefront for Android applications and other digital media, such a music, movies and books, from Google.

User has ample number of choices for the apps for a particular usage.

**3. Why Analysis is necessary?**

* **Users -** Through analysis they can check which apps are trending and which applications has the highest rating among the similar kinds of apps. As there are multiple numbers of applications present in the market with similar functions.
* **App Developers-**Through analysis they can check the reviews and decide what is the current need of the hour and develop or upgrade the applications according to the needs of user

**4. Steps Involved:**

* **Importing libraries and add path of given data set-**

This project is started with importing multiple libraries and then mount google drive and copy the path to the colab notebook.

* **Data Exploration-**

The very first step was to analyse both the datasets namely play store apps reviews data and users review data and check for the faulty data in rows and columns.

The dataset was analysed and found the numeric data i.e Reviews, size, installs etc were in type object, So to make the data presentable and cleaned the data by converting them into Float.

For that for loops and defining functions method were used.

Also there were values with some special characters such as $ and also some denotations like k and m. So, that also had to be removed to perform analysis seamlessly and make it more readable for users as well as for me.

* **Shifting a row-**

While performing data cleaning by converting objects into float, It was found that there was a row in play store app review dataset which is been shifted and uncleaned. So, to find that row number pivot table was used and got to know it is row no. 10475 according to my excel which I made presentable for better analysis and shifted that row by 1 so that all the values can fit into their own subsets.

After shifting that rest of the data cleaning was done smoothly.

* **Removing null values:**

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

1. **Data Analysis-**
2. The first analysis was done was to check how many applications are there in a single type of category. which concluded as the maximum number of apps are from family category which was followed by games and then tools.

For the same Seaborn countplot was used to show it visually and make it quite obvious for the users as well as app developers.

1. The second analysis was done to check the audience which are generally targeted by the app developers. Through the analysis it was found Everyone section is been targeted at most, means it does not contain any age group for which a pie chart was added.

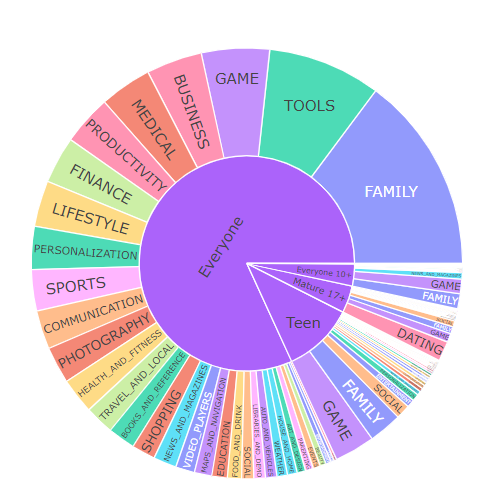
Further on the same analysis a sunburst chart was added which showed the app distribution between content rating.

EVERYONE- Family apps were leading.

TEEN- Games followed by Family

MATURE 17+ - Dating apps lead the way

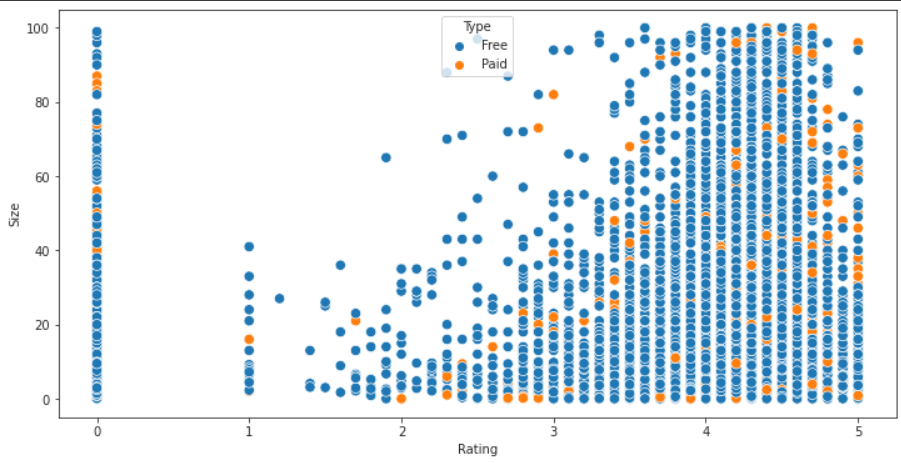
EVERYONE10+ - Family and then Games.



1. The third analysis was done to check amount of free and paid apps present in the play store market.

Through the analysis it was observed that, every 9 apps out of 10 were free to users. To represent this data two different kinds of visualization technique was used.

One was countplot and the other was scatterplot.

****

The scatter plot gives us a rough idea that in whole market there are very less number of paid apps, rest of them are free to users.

Blue= Free Apps

Yellow= Paid Apps.

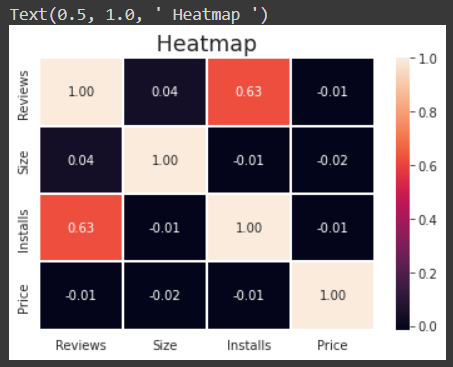
1. The fourth analysis was on Size of the apps per category. Through the analysis it was observed that Family and Tools has the highest amount of sizes as compared to the other categories.

Also, Art\_and\_Design & Comics has the lowest amount of app sizes as compared to the other categories.

1. Then, the analysis was taken forward to check the correlation between Rating, Reviews, Installs and Price.

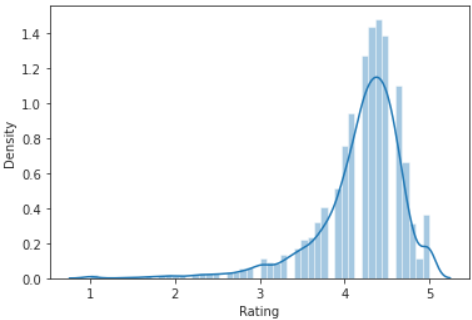
For the same Heatmap was used and from that It was found that there is strong positive correlation between reviews and installs.

And rest of relationships are either slightly positive or slightly negative. Or we can say Rest of them are as good as neutral.



1. Then some analysis was done on rating, and through what it was concluded that maximum number of ratings are in between 4 and 4.5 out of 5. To be precise with the data it was 4.19 out of 10.

To visualize the same a histogram was used.



1. And lastly there was an analysis done on sentiment of users while writing their reviews.

And for the same it was found that users are pretty much write positive reviews for the apps followed by negative reviews and then neutrals one.

1. **Conclusion-**

That’s it! That is the end of our exercise.

Starting with loading the data so far, we have done EDA, null values treatment, cleaning the dataset and then doing all the analyse.

This exercise has a very detailed information for users as well as app developers.

1. **References-**

* GeeksforGeeks
* RealPython
* Google
* Seaborn.pydata
* Askpython
* Moderator’s sample colab
* GitHub