Google PlayStore review analysis

Nagendra Pratap Singh

Data Science Trainees,

Alma Better, Bangalore

**ABSTRACT**

One of the most well-known and significant Android software shops is the Google Play store. It contains a vast amount of data that, when correctly analysed, can be quite useful. The dataset for the Google Play Store has been made available to us.

Our major goal is to conduct exploratory data analysis on the provided data set in order to draw the proper conclusions about current trends and the kinds of problems that the provided data set can be used to address.

**DATASET: -**

Mobile applications are incredibly helpful in our daily lives and are made to make things easier. They are employed in practically every sphere of life, including business, social interaction, medicine, and many others. The key factor in their popularity is that over 83% of the apps can be downloaded for free and quickly. It gives the folks a sense of comfort. Almost 3.5 million apps are available in the Play store, and an additional 3000+ apps are being created daily. With 5 billion users worldwide, the market is much more widely used. Making an app while blindfolded without conducting any study or analysis can take too much time and be less effective. This could result in

the project's failure. As a result, EDA may be performed on this dataset to provide developers with a comprehensive understanding of the market, as well as dos and don'ts, in order to get the most out of it and effectively design a working application.

**INTEGRAL METHODOLOGY**

The following sections make up the entire analysis: Description, dataset breakdown, Locating and handling the missing and null values cleansing of data following that, we shall Analysing Exploratory Data.

**DATASET DISCRIPTION**

Now let’s take a look at our data which has total of 2 csv files

Playstore.csv: This csv file contains data about the apps in the Play Store, including information about the app's name, category, size, and version. There are 13 variables in this data set.

User review.csv: This dataset contains reviews of the app along with there positive, neutral and negative sentiments

Before going into data, we will analysis the following observation.

1. Checking Correlation
2. Getting the average Rating of the Apps
3. Total number of applications in each category
4. Check the number of installs in each

Category

1. Getting App Size Distribution
2. Check the number of installs on the

 basis of their size.

1. Let’s check the app on the basis of price in play store. (Free or paid)
2. let’s compare Category according

 there Install

1. Category and Rating
2. Category and Reviews
3. Review sentiments in all the app

dataset

1. Let's see a more depth understanding of sentiment Polarity and Subjectivity.
2. Does sentiments Polarity is proportional to sentiments subjectivity.
3. Sentiment analysis of user reviews
4. Word Cloud

**DATASET FEATURES**

There are total 13 feature in the Playstore.csv dataset. The feature include: -

* App: - Name of the App
* Category: - Category under which the App falls.
* Rating: - Application's rating on PlayStore
* Reviews: - Number of reviews of the App.
* Size: - Size of the App.
* Install: - Number of Installs of the App
* Type: - Whether the App is free/paid
* Price: - Price of the app (0 if it is Free)
* Content Rating: - Appropriate Target Audience of the App.
* Genres: - Genre under which the App falls.
* Last Updated: - Date when the App was last updated
* Current Ver:- Current Version of the Application
* Android Ver: - Minimum Android Version required to run the App

**EXAMINING NULL VALUES**

We looked at the dataset and obtained the data as well. Now we must remove any properties that are no longer relevant. One of the most crucial aspects of data analysis is data cleaning. All of the null values must be replaced with either the average of their columns or non-null values.

**DATA CLEANING: -**

In this, the dataset will be cleaned, which will facilitate analysis and visualisation for us. The values in this dataset were cleaned by removing the pre and post symbols, and then they were transformed into numerical form for usage in the future. The symbols we are mostly deleting in this case are $, M, +, K, etc.

**DATA VISUALISATION**

With more than 1 billion active users across 190 nations, Google Play is still a crucial distribution channel for expanding a global audience. This increases the fierce rivalry in the app business.

Also, we can argue that data cleaning is essential before beginning any kind of exploration because it improves accuracy and results.

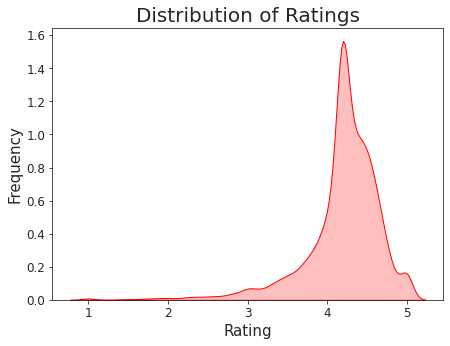
We can all see the various findings and inferences that come from data visualization, as well as how visualizations can make any graph, plot, chart, or map interesting.

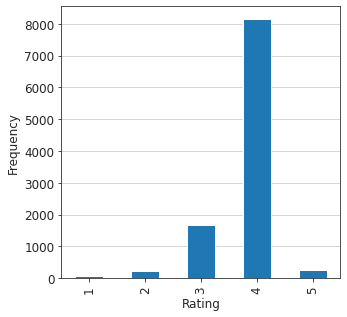
This make the data simpler and make it simpler to comprehend the function of the data and its component parts in the varied world of the play store.

While there are twice as many available applications from these categories as there are from the category FAMILY, according to the graph visualisations above, the majority of trending apps (in terms of users' installations) are from the categories of GAME , COMMUNICATION , and TOOL. These applications' popularity is largely owing to the fact that they may amuse or help the user by nature. Also, it demonstrates a positive trend in which we can observe that developers from these categories are placing more emphasis on the calibre than the number of the apps.

**OBSERVATION 1**

The first visualisation shows how the apps in the dataset were rated overall. The dataset's average rating will be determine

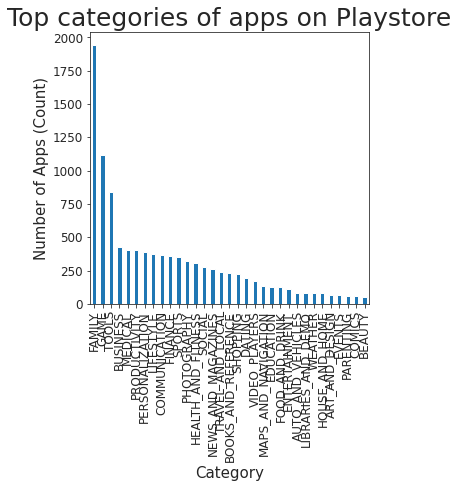




The graphical depiction shows a progressive rise in the ratings .We can observe that the majority of applications overall have ratings between 4.1 and 4.4.

It may also be claimed that when app frequency rises, so do ratings, or vice versa.

**OBSERVATION 2**

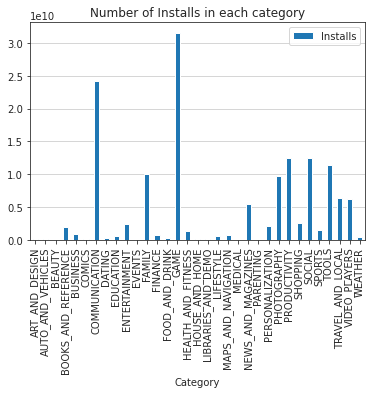
Finding the overall number of applications in each category becomes the second observation.

Following the visualisation, we can see that the two categories of "Family" and "Games" have the biggest number of apps available for download from the Play Store.

In contrast, there aren't many programmes available for download in the parenting, cosmetics, and comics categories on the Google Play Store.

**OBSERVATION 3**

following the categorization of the number of applications. We should also be aware of the installs in the Playstore in relation to the available categories. We will learn which categories have the most downloads through this observation.

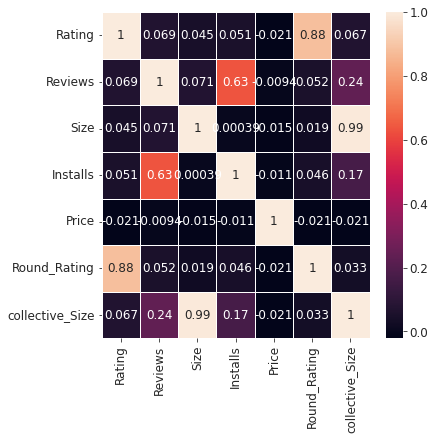


We can see that this observation is significantly different from the application count. We can see from the data set we've collected that even though families had the most applications in their category installed, there weren't many of them.

We can see that Gaming and Communication have the greatest install rates, which means that more users are installing these two types of applications.

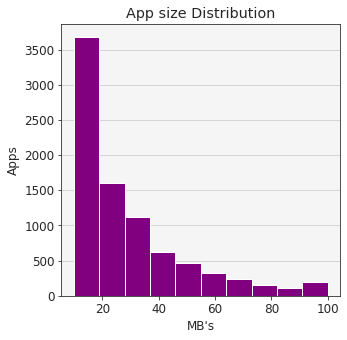
As predicted, comedy, events, medical, and many more categories had the lowest installations ever.

**OBSERVATION 4**

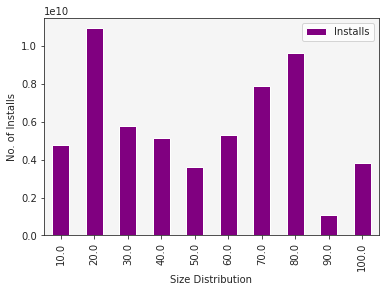
The correlation graph is crucial for the developer and analyst to have. With the use of this graph, we can determine whether there is a correlation between the traits, or whether they are mutually reliant. This may make it tough to visualise.

As we can see, practically all of the variables are unrelated, with the exception of reviews. The amount of installs seems to be impacted in some way by changes in reviews.

**OBSERVATION 5**

The distribution of app sizes is the next observation. This visualisation will teach us about the various size divisions and the amount of applications that are present in each distribution.We can see from the visualisation that the most apps on the Play store are between 10 and 20 MB in size, and that as the app size grows, the number of apps steadily declines, with the bulk of apps falling between 10 and 40 MB in size.

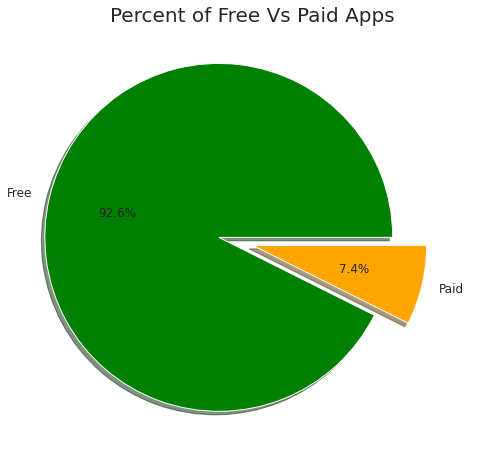
**OBSERVATION 6**

The following step is to obtain the install count in relation to application size once we have obtained the observation of the app size distribution. The fact that fewer apps are installed in the 10 mb range while having a better app distribution is an intriguing finding.

The 20 megabytes bracket has the highest installs overall, whereas the 70 to 80 mb bracket has greater installs overall.

**OBSERVATION 7**

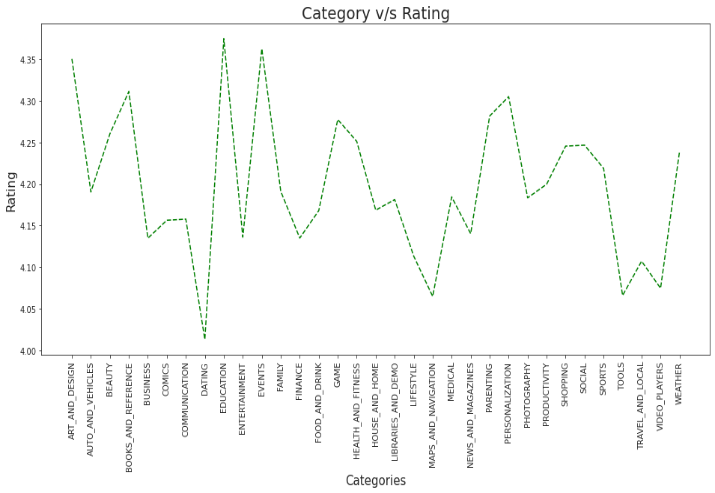
The PlayStore's applications are primarily split into two categories: free and paid. We will have a better knowledge of the dataset after visiting both free and premium apps.



We may infer from the plot that more PlayStore apps are free, which has a significant impact on downloads as well as reviews and ratings.

**OBSERVATION-8**

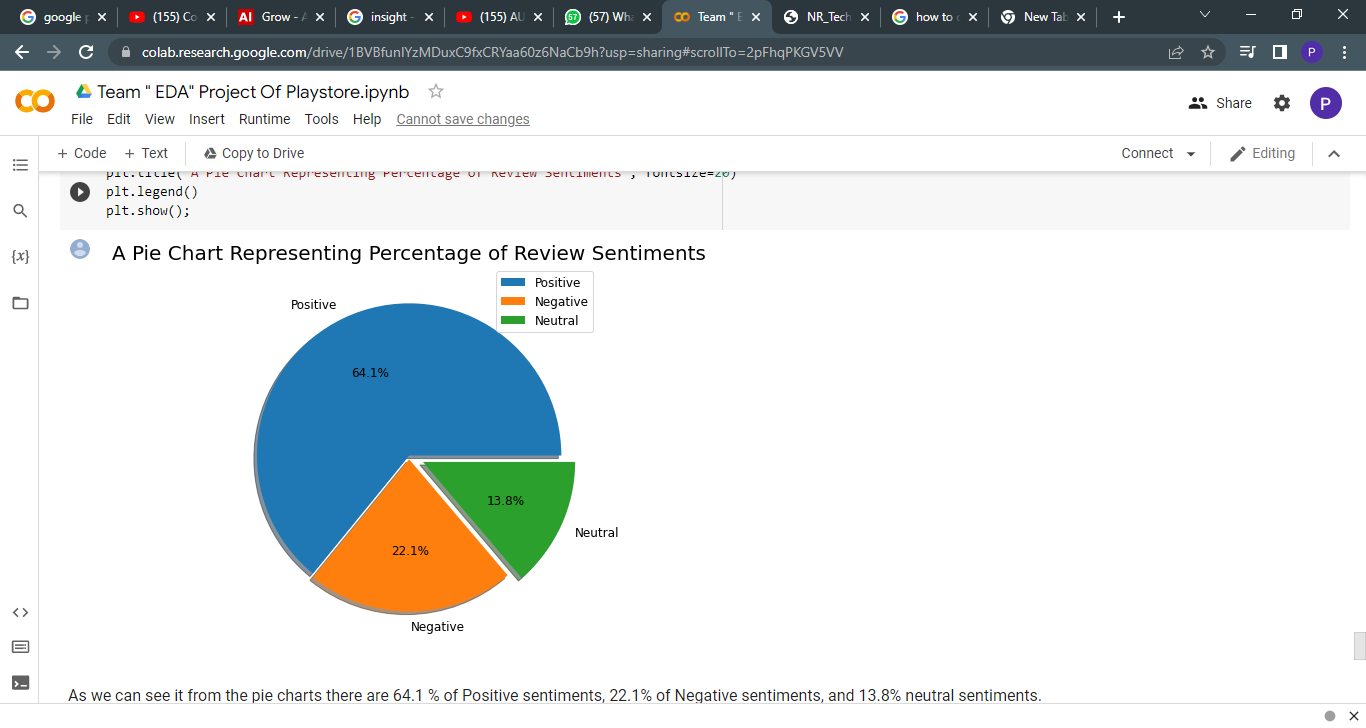
Let's now examine the relationship between category and rating.



Now we visualize the app category and there ratings we can see that dating apps are at low ratings but education apps are maximum ratings so we can say that people are involved in education app as compare to dating apps

**OBSERVATION-9**

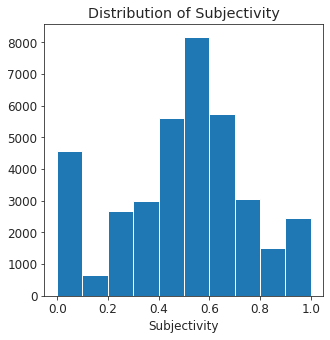
Now let us have a look at Review sentiments in all the app dataset.



From the pie charts, it is clear that there are 64.1% positive, 22.1% negative, and 13.8% neutral attitudes.

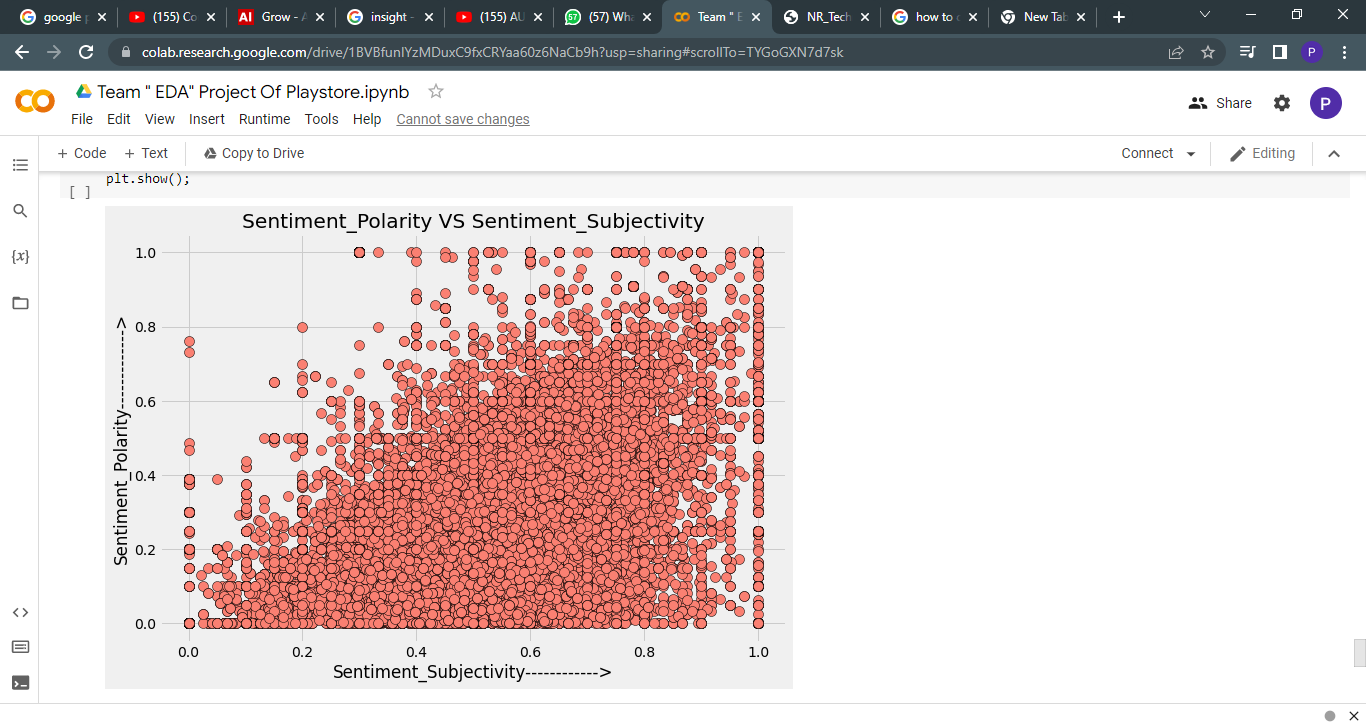
**OBSERVATION 10**

Let's see a more depth understanding of sentiment Polarity and Subjectivity.

It is clear that between 0.4 and 0.7 is when sentiment subjectivity is at its highest. This leads us to the conclusion that the majority of the audience rates the applications in accordance with their personal experiences.

**OBSERVATION-11**

Let’s have a look at our last observation which is sentiment polarity vs sentiment subjectivity



From the aforementioned scatter plot, it can be inferred that, while not always proportional to sentiment polarity, sentiment subjectivity generally exhibits a proportionate behaviour.

**CONCLUSION AND FUTURE WORK**

With more than 1 billion active users across 190 nations, Google Play is still a crucial distribution channel for expanding a worldwide audience. This increases the fierce rivalry in the app business.

Also, we can argue that data cleansing is essential before beginning any sort of study since it improves accuracy and results.

We can all see the many findings and inferences that come from data visualisation, as well as how visualisations can make any graph, plot, chart, or map interesting.

This make the data simpler and make it simpler to comprehend the function of the data and its component parts in the varied world of the play store.

According to the graph visualisations above, even though there are twice as few applications in these categories as there are in the category FAMILY, the majority of the trending apps (in terms of user installations) come from the categories of GAME, COMMUNICATION, and TOOL. These applications' popularity is largely owing to the fact that they may amuse or help the user by nature. Also, it demonstrates a positive trend in which we can observe that developers from these categories are placing more emphasis on the calibre than the number of the apps.

In addition to this, it can be noted that most applications have a larger number of user installations or positive reviews, with the majority of apps having ratings around 4. (8000 apps). Also, the app's size and cost make a small difference but do not significantly affect its positive ratings and reviews, even if they are high. Yet, we receive a bigger number of user installations and ratings as a result of the free in charges.

The figure also demonstrates the importance of feelings, with the highest percentage of favourable sentiments (64.1%), which are found between subjectivity ranges of 0.4 and 0.7. This leads us to the conclusion that the majority of the audience rates the applications in accordance with their personal experiences.

We can see that data exploration is crucial before beginning to create ML models.

As a result, we discovered that the majority of the popular Android applications fall into the categories of helping, communicating, or amusing.

**Thank YOU**