**Play Store App Review analysis**

**Trupti Patil**

**Data science trainees,**

**Alma Better, Bangalore**

**Abstract:**

Play store app is most popular and largely used Android app. The play store has millions of applications with different – different categories. Now a days number

of application are present in play store which are downloaded by population and they have enormous amount of data present in it.

The object of this project is to understand peoples demands and also area of interest.

For analysis and visualization, we use Python, Pandas, NumPy, Matplotlib, seaborn libraries

***Keywords: EDA – Exploratory Data Analysis***

## **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.

The population can download different app on smartphones. By doing analysis it is very helpful for developers to understand the performance of android market. There are lots of categories through which we can know what the needs of the population.

Here each app has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps which contain Application name, Review, Sentiment, Polarity, Subjectivity.

### **2. Introduction**

### In today’s world mobile apps playing an important role in any individuals life. There are millions of applications which provides wide collection of data on features like rating, price, number of downloads, app categories and so on. To know trend of market and which apps are loved by customer analysis is important. Hence proposing analyzed data to developer helps to know what customer is likely to download, which category got maximum downloads which plays very important role.

### Play store is digital distribution service developed and operated by Google.it is official app store which provides variety of content like book, music, magazine, exercise, meditation, recipes etc. with the growth in Android device and application, it would be exciting and interesting to perform data analysis.

### This EDA will help developer to know the success rate and help them to decide what features should be added or modify and what should be maintained according to the current state of their app.

### 

## **3. Datasets Provided**

● Play Store App Data

● User Review Data

Play store app contains the basic details of the app like number of user reviews, ratings, etc. user review data contains the user reviews and its sentiment score for the respective app. The dataset that is going to be used is play store data. The tools that are going to use for this EDA would be NumPy, pandas, matplotlib and seaborn. 

## **4. Uploading Data**

Well first thing first we will importing necessary packages and libraries. Then upload both the datasets we will be loading the Play Store Apps Data stored in csv using pandas which is a fast and powerful python library for data analysis and easy data manipulation in pandas Data Frame object. It usually used for working with the tabular data (e.g., data in spreadsheets) in various formats such as csv, Excel

spreadsheets, JSON etc. we will perform some data preparation and also cleaning it.

## **5.Description of Dataset**

## 1) **Play Store App Data**

1.**App:** The name of the app

2.**Category**: The category of the app

3.**Rating**: The rating of the app in the Play Store

4.**Reviews**: The number of reviews of the app

5.**Size**: The size of the app

6.**Install**: The number of installs of the app

7.**Type**: The type of the app (Free/Paid)

8.**The price of the app** (0 if it is Free)

9.**Content Rating**: The appropriate target audience of the app

10.**Genres**: The genre of the app

11.**Last Updated**: The date when the app was last updated

12.**Current Version**: The current version of the app

13.**Android Version**: The minimum Android version required to run the app

## 2) **User Review Data**

1. **App**: It gives name if application  
Translated Review - This gives information about what the users feedback is about the application.  
2. **Sentiment**: This column shows users reviews are positive or negative or neutral in words  
3.**Sentiment\_Polarity**: This shows sentiment column by value count  
4.**Sentiment\_Subjectivity**: This shows degree of emotions/sentiments expressed in a given text.

**6.** **Steps involved EDA**

* Basic information about data frames
* Duplicate values.
* Unique values in the data.
* Visualize the Unique counts.
* Find the Null values.
* Replace the Null values.
* Plotting the graph
* **Basic information of Data frame**

1.In play store data has 13 columns of properties with 10841 rows of data.

2. Here most of columns like ‘App’, ‘Category’, ‘Reviews’, ‘Installs’, ‘Type’, ‘Price’, ‘Content Rating’, ‘Genres’, ‘Last Updated’, ‘Current Ver’, ‘Android Ver’ are in the dtype of ‘object’

3. Values of column ‘Rating’ and ‘Size’ are in ‘float’ dtype. Representing size in ‘M’ as Megabytes, 'k' as Kilobytes and also ‘Varies with devices’.

4. Values of column ‘Installs’ representing install amount with symbols such as ‘,’and ‘+’.

5. Values of column ‘Price’ representing price with symbol '$'. Hence, we will need to do some data cleaning.

6. Another dataset contain customer review on android app which has 5 columns with 64295 rows of data.

7. Here most of columns like ‘App’, ‘Translated Review’, ‘Sentiment’ Ver’ are in the dtype of ‘object’. Remaining 'Sentiment Polarity', 'Sentiment Subjectivity' are float type

* **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.

* **Missing Value Treatment**

Missing values in a dataset can occur due to unavailable data or manual errors. It is very important to consider & treat the missing values in any dataset since it is determinantal to model performance

* **Duplicate Value Treatment**

Here the duplicate values are just deleted from the dataset. Duplicates can reduce the effectiveness of the model created.

* **Plotting the graph**

By plotting graph, we get the visualization of data in better way Gives a better understanding of the data set. Helps detect outliers or anomalous events. Helps understand data set variables and the relationship among them. For plotting graph commonly use matplotlib.pyplot library.

# **8. Questionary with answer**

## 1.Number of apps category wise

Answer: The graphs give the information about how many apps are there by the category wise. Here game category having maximum number of apps followed by family category and then dating category have maximum number of apps.

## 2.Distribution of the ratings of the apps Answer: Maximum apps having is 4.4 rating followed by 4.5 and 4.3.

## 3.Compare the app category wise with rating.

Answer: Highest quality apps above average rating are very few. Here one thing is must observe that maximum app having rating between 3.5 to 4.5 Apps with a rating above red line are game and having max pat above red line are family, auto and vehicles which indicates positive feedback for these categories. Performance of most app categories are below the average However, the apps in productivity and maps and navigation, comics have no body to describe them.

4. Proportion of apps in play store paid vs free

Answer: play store 17% are paid apps and 98.3% apps are free to install and use. 5.Top10 apps present in play store Genres wise Answer: In the top 10 Genres installed app at first three Genres are action, casual, dating. Maximum number of apps present in casual followed by News & Magazine followed by action.

6. View of sentiment polarity vs sentiment subjectivity Answer: sentiment subjectivity is not evenly distributed with sentiment polarity but in maximum number of cases, shows some kind proportional behavior, when variance is too high or low

## 7.Histogram of subjectivity Answer: Maximum number of sentiment subjectivity lies between 0.4 to 0.7. From this we can conclude that maximum number of users give reviews to the applications, according to their experience.

## 8.Percentile distribution of review sentiments of dataset Answer: Populations positive reviews are 62.52% which shows the good side. Negative reviews are 26.30% shows area of improvement and neutral 11.18%.

## 9.Correlation between the columns of data frame. Answer: This gives fair idea of the correlation strength between different variables.

## **9. Conclusion:**

That's it! We reached the end of our exercise.

Starting with loading the data for play store and user reviews then we done for each dataset data cleaning , null values treatment, changing the data type of categorical columns, merging both the data set, visualize and analyze through questions and answers.

**References-**

1. Geeks for Geeks
2. Analytics Vidhya