**Play Store App Data Analysis**

**Vashu Garg (L), Palak Bindal,**

**Deepika Gupta, Somya Jain, Jyoti Singh**

**Data Science Trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

The Google Play Store apps data analysis provides enough potential to drive apps making businesses to succeed. Actionable stats can be drawn for developers to work on and capture the Android market. The data set that we have taken in this article is a web scraped data of 10 thousand Playstore applications to analyze the data to discover key factors responsible for app engagement and success.

Another dataset contains customer reviews of the android apps.

This project is about Google play store apps rating,review,content rating,installous,type ,price analysis etc. In this project we have done data exploration, data wrangling, data visualization, studied the relation between installs and ratings variables using scatter plot.plot the pie chart to see sentiment polarity rate category wise.Importing the required libraries. Fetch the data set from a given format Doing some operation on the data set.Data filtering and store in its proper format(sort column,drop null value,drop some column,change in required data type).Some important observation: find out some relation in the dataset through the graph.

The project was done in colab notebook.

***Keywords: (Data Visualization, Playstore app, Data Analysis, data filtering, data wrangling)***

**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 contains customer reviews of the android apps.Explore and analyze the data to discover key factors responsible for app engagement and success.we have to analyze data for future prediction.We aim on providing doing sentiment analysis on the apps that generated most positive and negative sentiments and sustainability of app in market on basis of previous data and current market.

Objective: Make a model to predict the app rating, with other information about the app provided.

**2. Introduction**

In today’s era , the Google Play Store is the largest and most popular android app store. It is flooded with millions of applications and it provides a wide collection of data.

The project aims at doing this with the help of sentimental analysis that will analyze customer needs and suggest the developers best app for developing . The analysis is achieved using the survey of the user downloads behavior on the apps across all the on the Google Play Store. Mobile app industry is increasing significantly and thus giving rise to more competitions to the one that is creating applications.

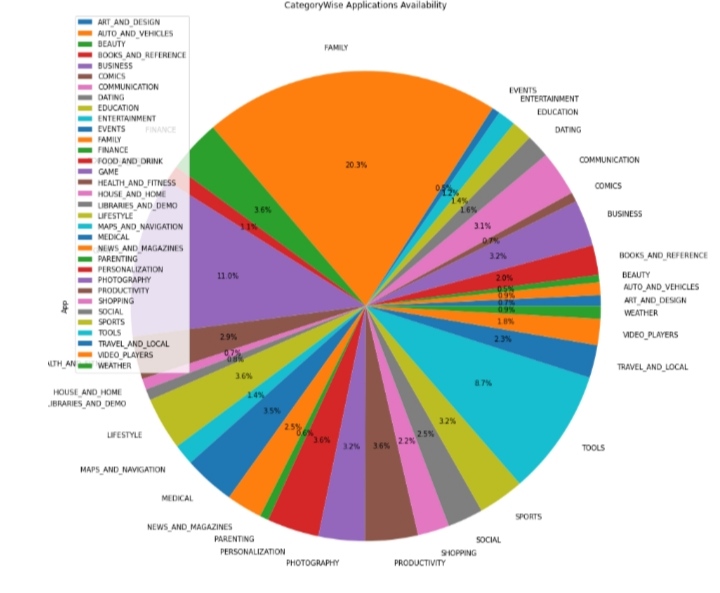
**About Dataset:**

| Parameters | Descriptions |
| --- | --- |
| Apps | Application Name |
| Category | Category the apps belongs to |
| Rating | Overall rating of the app |
| Reviews | Number of user reviews for the app |
| Size | Size of the app |
| Installs | No. of user downloads / installs the app |
| Type | Paid or Free |
| Price | Price of the app |
| Content rating | Age group the app is targeted at children at children / Mature / Adult |
| Genres | An app can belongs to multiple genres |
| Last Updated | Date when the app was last updated on google play store. |
| TranslatedReview | Comments given by user for different apps |
| Sentiment | Response of the user |
| Sentiment Polarity | Aims to differentiate the opinion into 'positive' and 'negative'. |
| Sentiment Subjectivity | Aims to remove 'factual' or 'neutral' content |

**3. Steps involved:**

* **Exploratory Data Analysis**

After loading the data file using pandas.It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.



* **Null values Treatment**

Check for null values in the data. Get the number of null values for each column. Drop records with nulls in any of the columns.

* **To omit the signs**

Variables seem to have incorrect type and inconsistent formatting. We need to fix them:Convert it to numeric (int/float). Installs field is currently stored as string and has values like 1,000,000+. Treat 1,000,000+ as 1,000,000 remove ‘+’, ‘,’ from the field, convert it to integer Price field is a string and has a $ symbol. Remove the ‘$’ sign, and convert it to numeric.

* **Outlier treatment:**

Average rating should be between 1 and 5 as only these values are allowed on the play store. Drop the rows that have a value outside this range. Reviews should not be more than installed as only those who installed can review the app. If there are any such records, drop them. For free apps (type = “Free”)

**4. Conclusions**

1. DS Creator 2.0 is of Category ‘Tools’
2. Apps of Category ‘Weather’ has maximum reviews
3. Best Food and drink app baMost of the apps available on play store belongs to Category ‘Family’ i.e, 20.3%Best gaming app based on rating is Monster Ride Pro
4. Best Art and Design app based on rating is Spring flowers theme couleurs d t space
5. Apps which have maximum installs are mostly free.
6. Total No. of Free apps are 8719 & Paid apps are 647
7. App with least rating is sed on rating is Bar B-Q Rib House
8. App which generated the highest revenue is Minecraft
9. Average no. of words used in Translated reviews is 18
10. Most of user sentiments are Positive
11. Game Category has the highest sentiment polarity rate.
12. Maximum installed apps are Subway surfers,Facebook,Messenger & Google Drive
13. Category with has max no. of apps with 4+ rating is ‘Family’
14. Most expensive app on Play store is “I’m Rich” priced at 400$
15. 1208 apps are famous among teens out of 10841 apps
16. Highest revenue generated app belongs to Category ‘Family’
17. Most of the apps has got the rating between 3.5 to 5
18. Maximum no. of installations belongs to Category ‘Communication’ and ‘Social Media’
19. Application reviews and installs are highly correlated i.e, 0.64
20. 48 apps are Google listed and their average rating is 4.3
21. Genre ‘Action’ has highest - 358 Gaming applications
22. Maximum no. of applications got updated in year 2018
23. Out of all Categories, ‘Game’ has got the most positive sentiments