**Play Store App Review Analysis**

**Rohit Meshram,Rahul gaykwad,**

**Rutuja Hingankar,Prashik Ingle,**

**Nrayan borde**

**Data science trainees,**

**AlmaBetter, Pune**

**Abstract:**

The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product. The dataset is chosen from Kaggle. It is of 10k Play Store apps for analyzing the Android market. This dataset contains details of different applications and reviews from different users.

Our experiment can help understand what could be the reason for the classification of such labels by feature selection, data analysis and prediction with machine learning algorithms taking into account previous trends to determine the correct classification.

**Introduction**

The Google Play Store and formerly Android Market, is a digital distribution service operated and developed by google. It serves as the official app store for certified devices running on the android operating system and its derivatives as well as ChromeOS, allowing users to browse and download application. • The play store 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. • Taking into account billion of android users worldwide, mining this data has the potential to reveal user behaviours and trends in the whole global scope. This dataset is obtained from scraping Google Play Store.

**Problem Statement**

• Two datasets are provided, one with basic information and the other with user reviews for the respective app.

• We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success. So, what factors influence an app’s success? An app is said to be successful if it has :

• A high average user rating

• A good number of positive reviews

• A good numbers of monthly average users

• High revenue per customer and so on.

**Dataset Description**

Two different datasets provided for analysis:

1. **Play Store Data.csv**

App : Categorical, the app names.

Category : Categorical, category the app belongs to.

Rating : Numerical, range from 0.0 to 5.0, rating has received from the users.

Reviews : Numerical, the number of reviews that the app recived.

Size : Numerical, the size of the app. The suffix M- megabytes,kilobytes.

Installs : Numerical, describes the number of installs.

Type : Categorical, a label that indicates weather the app is free or paid.

Price : Numerical, the price value for the paid apps.

Content Rating : Categorical, a categorical rating that indicates the ages group for

user.

Genre : Categorical, a categorical rating that indicates the age group for user. Last update : Date format, the date at which the app was last updated.

**2.User Review.csv**

App :The app name.

• Translated\_Review :The review text in English

• Sentiment :The sentiment of the review, positive, neutral, or negative.

• Sentiment\_polarity :The sentiment in numerical form,ranging from-1.00 to1.00.

• Sentiment\_subjectivity : a measure of the expression of opinions, evalutions, feelings, and speculations.

**Data Prepration**

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data. This step is about getting to know the data and understanding what has to be done before the data becomes useful in a particular context. This can be done by reading the CSV file and doing initial statistical analysis. Though the dataset may seem to have the correct datatypes for each column, we need to check it. Inconsistent datatypes will create issues while dealing with problems.

**Data Processing**

Discussion of Google play store dataset will involve various steps such as:

loading the data into data frame cleaning the data extracting statistics from the dataset exploratory analysis and visualizations •questions that can be asked from the dataset, Library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes python programming simpler and convenient for the programmer.

**Cleaning data**

Cleanse and validate data This step is crucial for removing faulty data and filling gaps. Important tasks here includes: Removing extraneous data Filling in missing values. Conforming data to a standardized pattern. Dataset may contain duplicate values for particular application. This step is crucial for removing faulty data and filling in gaps. Important tasks here includes,Removing extraneous data ,Filling in missing values,Conforming data to a standardized pattern. Dataset may contain duplicate values for particular application:



## **Transform data**

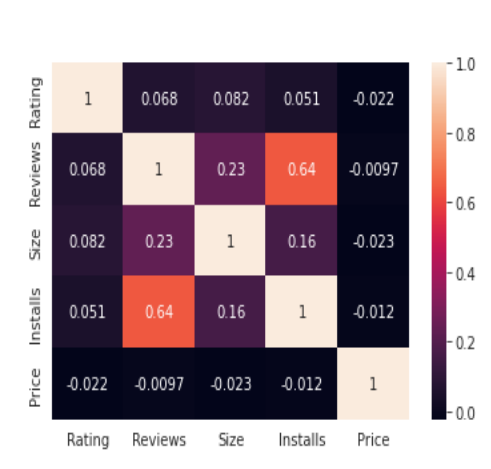
Transforming data is the process of updating the format or value entries in order to reach a well-defined outcome, or to make the data more easily understood by a wider audience. Since, size of the applications present in the datset are in MB and KB. Therefore, for ease in data processing, entire size column is converted to MB.

# Exploratory Analysis and Visualization

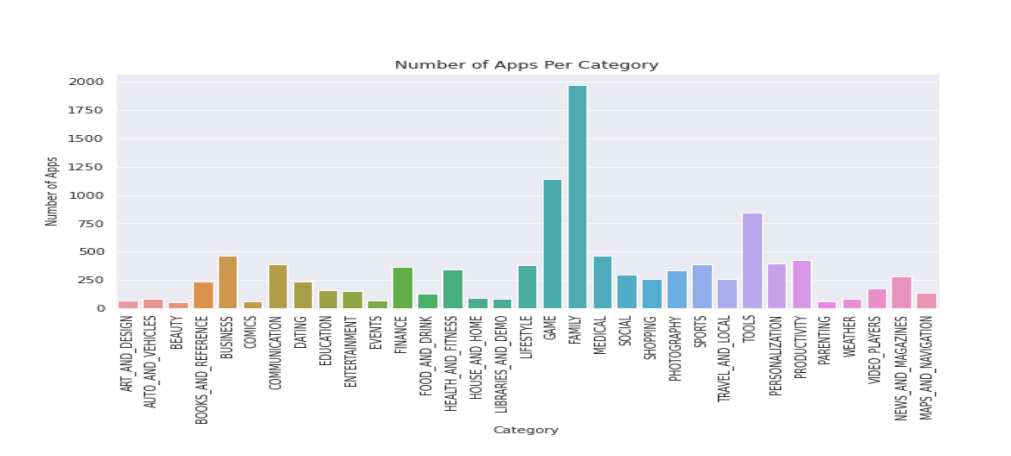
Exploratory data visualizations (EDVs) are the type of visualizations we assemble when we do not have a clue about what information lies within our dataset.

**Correlation Heatmap**

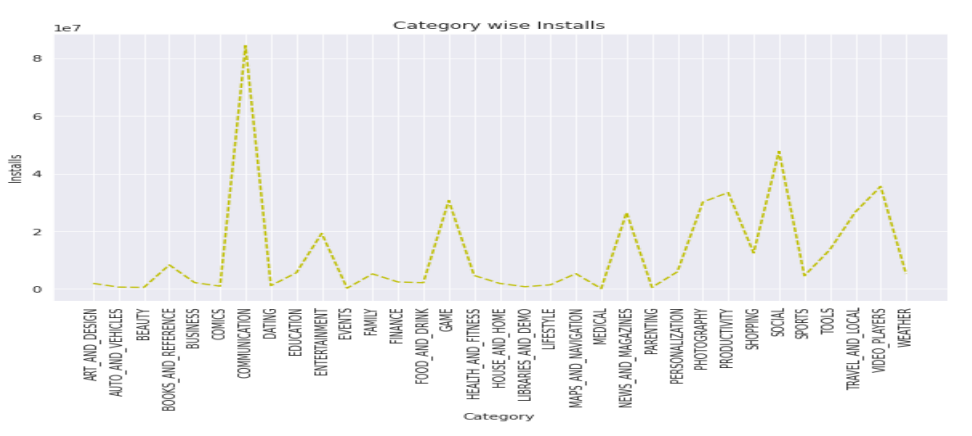
There is a strong positive correlation between the Review and Installs ,The price is slightly negatively corrected with the Rating , Review, and Installs., The rating is slightly positively correlation with the Installs and Reviews.



**Categorical Analysis**

•The Family, Game, and Tools category has the highest number of apps.

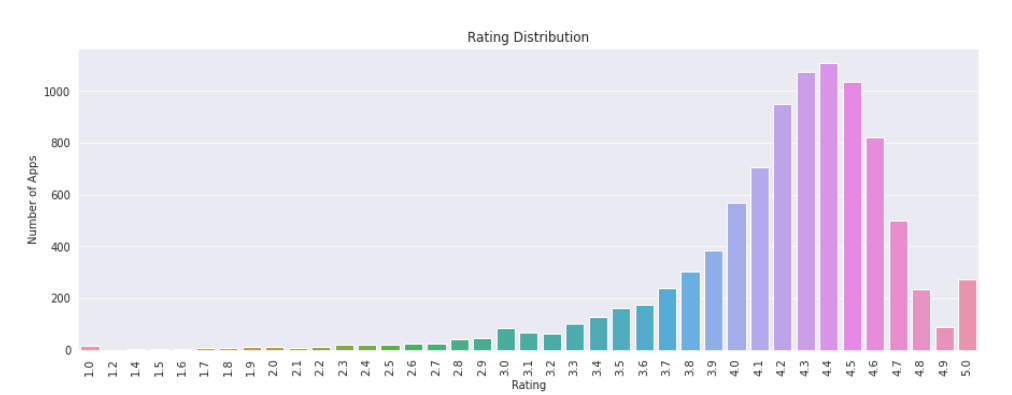
**Categorical Analysis**

• The Communication, Video players, and Social category has highest number of app installs. 

**Distribution of app ratings**

• Scale is from 1 to 5 Most of the apps, clearly hold a rating of 4.4

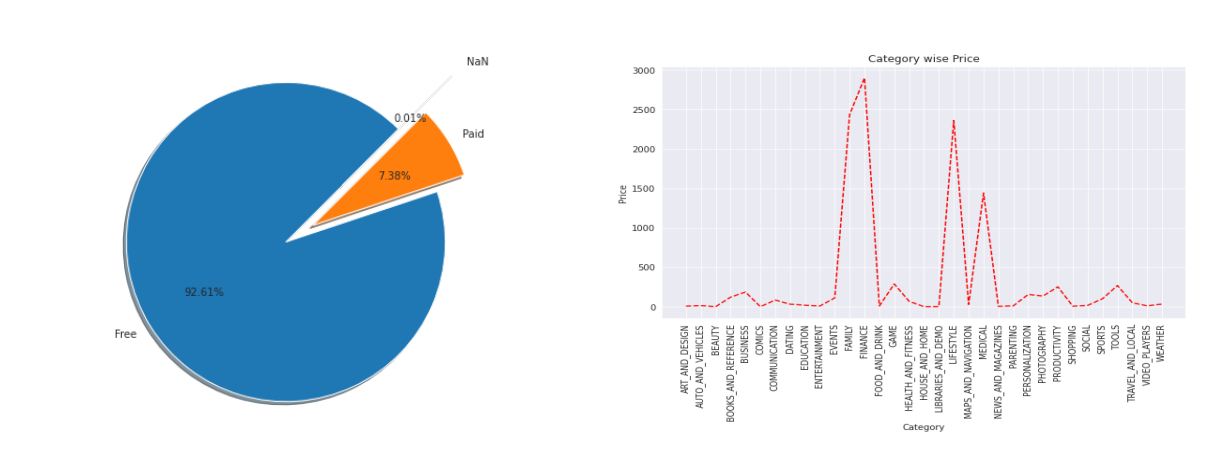
• This implies that the majority of the users are happy with the service received via the respective app.



**Paid app in each category**

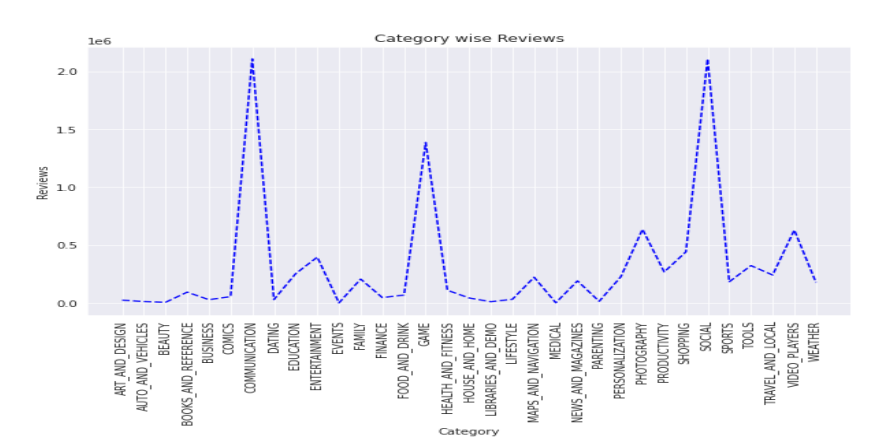
• It looks like certain app categories have more free apps available for download than others

• The paid apps in the finance, Lifestyle, and Event category are on average significantly more expensive than paid apps in other categories.



**Analysis of user reviews**

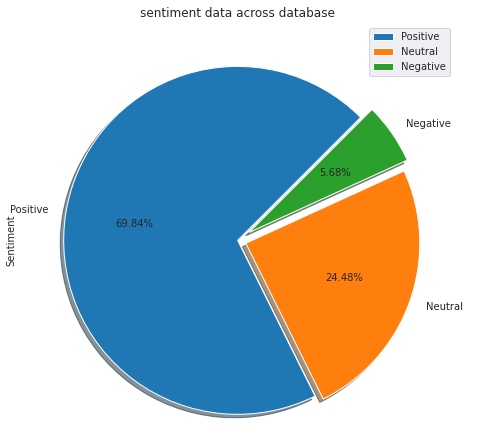
Apps related to communication ,Games and Social seem to have highest number of reviews.

****

**Positive and Negative Review**

The percentage of positive reviews of the apps in the play store is much greater than the negative and neutral reviews which are 69.84%.

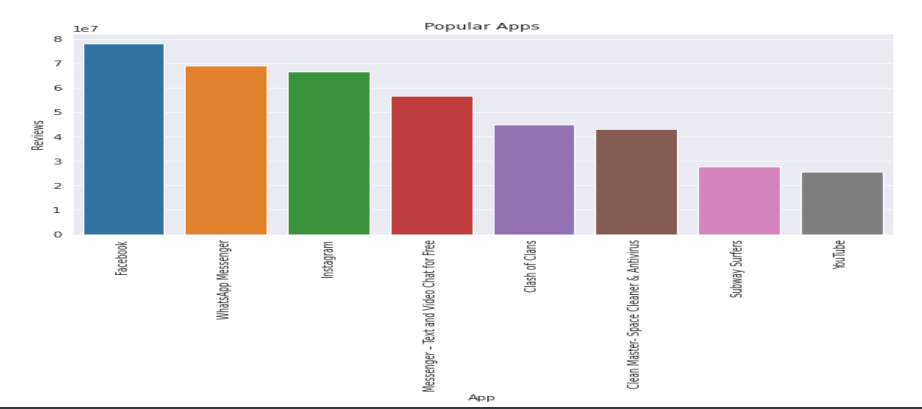
• Negative percentage of apps are 5.68%

**** • Neutral percentage of apps are 24.48%

**Most Review Apps**

• This are some most review apps in the play store. We can state that the apps with more review, whether positive, negative, or neutral, are more popular than the other.

• This is because the number of user reviews indicates that these individuals have engaged with the app’s content and have written their opinion on it.



**Challenges Faced**

• Reading the dataset and comprehending the problem statement.

• Examining the business KpIs for app development and devising a solution to the problem.

• Handling the error, duplicate and NaN values in the dataset.

• Designing multiple visualization to summarize the information in the dataset and successfully communicate the result and trends to the reader.

**Conclusion**

In this project of play store app review analysis we've drawn many intresting inferences and some great results, here's a conclusion of few of them:

• Maximum number of apps belongs to the family category and gaming category.

• Most of the apps are free very few apps are paid apps in which family category has the most free and paid apps and social apps like entertainment are almost free.

• The rating is very high with atleast 75% of the apps in the range 4.0 to 4.7.

• Among all rating of events apps is highest whereas dating apps has lowest rating.

• Communication and social category apps has high reviews

. • Mostly downloaded apps are communication apps.

• Most of the paid apps belongs to finance category.

• There are more number of positive reviews of apps in which facebook, whatsapp, Instagram are the most review apps

**Dataset reference**: <https://www.kaggle.com/lava18/google-play-store-apps>

**Jupyter notebook reference**: <https://jovian.ml/ritz1602-rs/course-project-google-play-store-dataset>