Play Store App Review - EDA
Ritik Gupta
Data Science Trainee,
AlmaBetter

## Abstract :

Android app market is growing exponentially and becoming more competitive at the same time, developer are pushing apps on regular basis and varieties of apps are increasing.

In this competitive market, Developer needs to choose or target the segment according to its user's behaviour, because at the end they are the users who installs and use the app and give the rating based on the performance and we can also conclude rating as the satisfaction level of customers and a KPI to make decision.

This type of decision can be taken by developers with help of analysing and reviewing the dataset of apps which are presented and operational in market.

**Keywords**: Android Apps, Play Store, EDA, Reviews, App ratings, Free Apps, Paid Apps, User Reviews

## Problem Statement :

Android app market growing tremendously, so does the confusion, and app developers needs to find out which app they need to build, while considering popular choice of users, profitable category to develop apps, competition and appropriate type of business model that gives them higher returns?

## About Dataset

- **App**: Name of apps are given
- **Category**: Category name for particular app, 33 unique category presented.
- **Rating**: Rating of each app given between 1.0 - 5.0

- **Reviews**: Number/count. of reviews given to each app
- **Size**: Size of app in MB and KB
- **Installs** – number of app installs
- **Type** – Free or Paid type
- **Price** – the price of app if its Paid app, else 0
- **Genres** - genres of app, can be multiple
- **Sentiment** – sentiment of review posted by user, weather Positive, Negative or Neutral
- Includes Some other parameters such as 'Android version', 'Current version', 'Last updated', 'Content Rating'.

## Steps involved in EDA:

1. **Treatment of Outliers**: In the column of Rating, values should be between 0 - 5.0, and outside that are all outliers or misinformation, so removed the value which has above 5.0 rating

2. **Null value treatment:** Using info, we found null values in 'Rating', 'Type', 'Android ver' and 'Current ver', we filled the null values of 'Rating' with median of all ratings, and other columns with the mode of themselves, as they are very few.

3. **Data analysis**: Performed analysis with important columns and gather key details using visualisation tools and further used that information to dive deep and find out more insights and create a conclusion.

# Observations :

- **Ratings**: Ratings are considered as performance matrices and we have seen that-
  - ➢ 4.2 is the Average Rating in the whole data.
  - ➢ The most frequently given Rating is 4.3, so we can see that ratings are on positive side.
  - ➢ Every category has more than 4.0 average rating except for 'Dating'.

- **Content Rating**: In the whole Data of apps, almost 80% apps belong to content rating of 'Everyone'.

- **Popular Category:**
  - ➢ Category with highest number of apps are 'Family', 'Game' and 'Tools', others are too low compare to these three categories.
  - ➢ These categories has most number of apps with above 1,000,000 installs
  - ➢ The Category 'Photography' has low number of apps, but its in top five most installed app category.

- **Free and Paid Apps:**
  - ➢ From the data collected, 92% apps are free.
  - ➢ Despite of low no. of apps 'Photography' performed good in terms of paid apps with more than 100000 installs.
  - ➢ 40% of Paid Apps revenue is occur from 'Game'.
  - ➢ 3 category from "top 5 categories with most number of paid apps", Medical, Games and Tools are of micro-niche, targeting a specific micro segment of users.
  - ➢ the size of paid apps are comparatively lower then Size of free apps, it may because of paid apps

developers charging for efficiency of apps and free app developers wants to cut the cost of development by neglecting size efficiency.

- **Size, Rating and Installation**: Size is also to be considered a key factor in decision making because-
  - ➢ When Sizes are between 0-60, the ratings are mostly between 3.7 and 4.8 but when the size increase from 60 number of rating drops.
  - ➢ The installations are decreasing as going to above 60 MB of size, so we can conclude that size is a key factor for app installs, user installs became lower in case of high app size.

- **Sentiments:**
  - ➢ Positive sentiments are more compare to negative, and neutral are less, and specifically Gaming has largest reviews with positive and negative sentiment of approx. 60% and 40% respectively, all other category has following the same pattern of positive reviews more than negative reviews and neutral are the lowest, by which we conclude that none of category is performing negative particularly.

  - ➢ There are 5 categories which has more than 70% of positive reviews and we can conclude that these categories are already satisfying its users, and the categories with less positivity are yet to receive the user satisfaction, because a negative review are referred to dissatisfied user.

- **Least Competitive Category**:
  - ➢ 'Beauty' category has lowest apps in market, but the category received above average Rating, also apps in this category didn't satisfy users that's why it has 54% positive reviews.

  - ➢ 'Comics' category has highest positive reviews of 90% but also no. of apps is lowest in market, that makes it a least competitive category

  - ➢ 'Photography' has also low number of apps, but its most of apps have greater than 1000000 installs, that makes it also a least competitive category with huge potential.

## Summary and Suggestions :

- 'Game' and 'Family' category are already dominating the market, but there are some low competitive categories which have lots of potential are- 'Beauty', 'Comic' and 'Photography' developers should focus on that.

- Size of app will create impact of user behaviour so its best choice for developers to create apps with low size as possible.

- Paid apps are not suitable for broad users, they are meant to be made for a specific niche or segment and focus only on particular segment for better profits.

- For paid games, we seen a pattern that if it was already available for different platforms and is popular, then it available on play store, it will receive tremendous response from user. For example;- Minecraft, GTA vice city. Generated highest revenue.

## References:

- Matplotlib Documentation
- Seaborn Documentation
- Towards data science blogs
- EDA: Exploratory Data Analysis | Introduction to Exploratory Data Analysis (analyticsvidhya.com)