Play Store App Review Analysis- EDA

Girish R
Data Science Trainee,
Alma Better

Abstract:

The Android app market is expanding exponentially while also becoming more competitive; developers are releasing apps on a regular basis, and the variety of apps is expanding.

In this competitive market, developers must choose, or target segments based on their users' behaviour, because it is the users who instal and use the app and give ratings based on performance, and we can also conclude ratings as customer satisfaction and a KPI to make decisions.

Developers can make this type of decision by analysing and reviewing the dataset of apps that are currently available and operational in the market.

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

Problem Statement:

The Android app market is rapidly expanding, and with it comes confusion. As a result, app developers must decide which apps to create while taking the preferences of users, lucrative app categories, market competition, and the best business models into account in a store.

About Dataset

- ***App***: Name of apps are given
- ***Category***: Category name for particular app unique categories presented.
- ***Rating***: Rating of each app givenbetween 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 zero
- ***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’, ‘Lastupdated’, ‘Content Rating’.

Steps involved in EDA:

1. **Treatment of Outliers**: Values in the Rating column should be between 0 and 5.0, and anything outside of that is an outlier or misinformation, so remove any value with a rating higher than 5.0.
2. **Null value treatment:** Using information, we discovered null values in 'Rating,' 'Type,' 'Android ver,' and 'Current ver,' and we filled the null values in 'Rating' with the median of all ratings, and other columns with the mode of themselves, as they are few.
3. **Data analysis**: Analyzed key columns and gathered key details using visualization tools, then used that information to delve deeper and discover more insights and draw a conclusion.

Observations:

- **Ratings**: Ratings are thought of as performance matrices, and we can see 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**: Nearly 80% of the apps in the data had a "Everyone" content rating.

- **Popular Category:**
  - Family, Games, and Tools are the categories with the most apps; other categories pale in comparison to these three.
  - These categories have the most apps with more than 1,000,000 installs.
  - Despite having a small number of apps, the category "Photography" is among the top five most downloaded app categories.

- **Free and Paid Apps:**
  - From the data collected, 92% apps are free.
  - 'Photography' did well among premium applications with more than 100,000 installs despite the low number of apps.
  - 40% of the money made from paid apps comes from games.
  - Three of the "top 5 categories with the most purchased apps" Medical, Games, and Tools are micro-niche categories that cater to a narrow user base.
  - The size of paid apps is 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 should be taken into consideration when making decisions also because -
  - Most ratings fall between 3.7 and 4.8 when sizes range from 0 to 60, although the number of ratings rises when sizes exceed 60.
  - We may infer that size is a crucial element for app installs because the number of installations drops off as the program size increases over 60 MB. Users are also less likely to install large apps.

- **Sentiments:**
  - Positive reviews outnumber negative reviews, and neutral reviews come in last. This means that no category is performing particularly poorly. Gaming has the most reviews, with positive and negative sentiments at about 60% and 40%, respectively. All other categories have also followed this pattern, with positive reviews coming in at a higher rate than negative reviews and neutral reviews coming in at the lowest.

  - The categories with less positivity have yet to receive customer satisfaction because a negative review is referred to as a dissatisfied user. There are 5 categories with more than 70% of positive evaluations, and we can assume that these categories are already fulfilling their users.

- **Least Competitive Category**:
  - ✓ The 'Beauty' category has the fewest apps in the market, but it received an above-average rating; additionally, apps in this category did not satisfy users, which is why it has 54% positive reviews.

  - ✓ The 'Comics' category has the highest positive reviews of 90%, but it also has the fewest apps in the market, making it the least competitive category.

- ✓ 'Photography' has fewer apps, but most of them have more than 1000000 installs, making it the least competitive category with the most potential.

Summary and Suggestions:

- ✓ The "game" and "family" categories already rule the market, but "beauty," "comic," and "photography" are three low-competitive categories with a lot of promise that developers could concentrate on.
- ✓ Paid apps are not appropriate for general users; instead, they are designed for a certain niche or market sector and concentrate solely on that market segment to maximize earnings.
- ✓ When it comes to premium games, we've noticed a pattern: if a game is already well-liked and widely accessible across platforms, users will respond enthusiastically when it debuts in the Google Play store. For instance, consider Minecraft; GTA Vice City produced the most money.
- ✓ Because user behavior is influenced by an app's size, it is advisable for developers to make their apps as small as possible.

References

1. Geeks For Geeks
2. Matplotlib Documentation
3. Seaborn Documentation
4. Towards data science blogs
5. EDA: Exploratory Data Analysis | Introduction to Exploratory Data Analysis (analyticsvidhya.com)