**App Review Analysis**

**Anjali Tidke, Shubham Dukare**

**Data science aspirants**

**Almabetter, Bangalore.**

**Abstract:**

The project “App review analysis” is for the analysis of the app data from Play Store. In this project, we have some data of applications and user reviews for analysis.

Google Play Store app is used to install most of the android application for the mobile.

In the play store, we can see lot of apps. For each app, there is name, its origin company name, version, it’s type, category, size, reviews etc. So, like this we have data as a input.

The analysis process includes data cleaning, removing or replacing null values, convert the data into proper format for analysis, data visualization etc. from data visualization, we get some understanding from which we can make business decisions.

***Keywords: App review analysis, play store, EDA.***

**1. Problem Statement**

 The data from play store app has enormous potential to success an app-driving business. Some insights can be drawn from this data to capture the android market.

We are provided with two datasets, first has the app data and 2nd has user reviews data of android apps.

We have to explore and analyse the data to find out some meaningful insights to make business decisions, to discover the key factors which are responsible for app engagement and success.

**2. Introduction**

Google play store is an application which is the best way to deliver the apps to android phone users. As we can see, most users use the play store to install android apps. Lot of data will be generated from these apps. From the data analysis of the apps, we can make some decisions to capture the market.

This EDA project is for the analysis of data from google play store dataset. The aim of the project is to finding out the key factors which are responsible for the app success and capture the android market.

Google Play-Store app Dataset columns description: -

* **App:** Application name
* **Category:** Category of the app
* **Rating:** Overall user rating of the app
* **Reviews:** Number of user reviews for the app
* **Size:** Size of the app
* **Installs:** Number of user downloads/installs for the app
* **Type:** Paid or Free
* **Price:** Price of the app
* **Content:** Rating Age group the app is targeted at - Children / Mature 21+ / Adult
* **Genres:** An app can belong to multiple genres (apart from its main category). For e.g., a musical family game will belong to Music, Game, or Family genres.
* **Last Updated:** Date when the app was last updated on Play Store
* **Current Ver:** Current version of the app is available on the Play Store
* **Android Ver:** Min required Android version

User Review Dataset column description: -

* **App:** Application name
* **Translate Reviews:** Reviews on various applications given by the user
* **Sentiment:** Positive, Negative, or Neutral
* **Sentiment Polarity:** Tell us about the polarity of the sentiment of the translated reviews that we received.
* **Sentiment Subjectivity:** Tell us about the subjectivity of the various reviews according to the translated review.

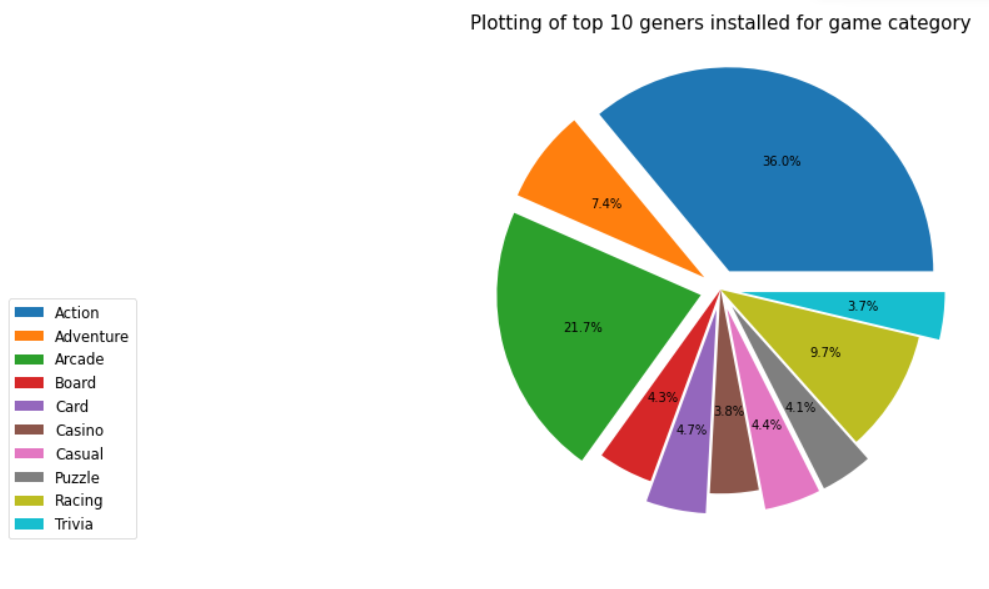
# 3. **Steps involved**

1. **Load the data**: - We have to access the data for the analysis. We are provided with two datasets. We imported those datasets in colab notebook using google drive.
2. **Basic data information:** - After loading the data, we checked the of both datasets.
   1. Play Store Data: - Play store data have total 10840 entries and 11 columns. By observing the data type of each column, we seen that data is not in proper data type format. One work for us was to make the data into proper data type format. Also Rating column have more null values. Type and Content rating column have some unique values.
   2. User review Data: - This dataset have total 64295 entries and 5 columns. This dataset has a greater number of null values and sentiment column have 3 unique values.

By observing the two datasets, we got some queries in our mind as below,

* + - Co-relation matrix
    - Is sentiment subjectivity proportional to sentiment polarity
    - User sentiment pie chart
    - Number of Installed applications for each category
    - Types of installs per category
    - Top 10 genres installed per category
    - Free apps vs paid apps
    - Number of installed applications according to content rating
    - Distribution of Rating
    - Installs per Category
    - Pricing vs Category
    - Number of Reviews per Category

| **Fig 1: Co-relation matrix** | **Fig 2: Is sentiment subjectivity proportional to sentiment polarity** |
| --- | --- |
| From the above analysis, we can see that Installs and Reviews are Positively Correlated. | we can see here from the above scatter plot analysis sentiment subjectivity is not proportional to sentiment polarity but in many cases,we can see a proportional behavior when variance is very low or high |



| **Fig 3: Sentiment analysis** | **Fig 4: Top 10 genres installed for game category** |
| --- | --- |
| We can see here, Postive reviews are more and neutral reviews are less | Above analysis shows that, Action have more 36% |

| **Fig 5: Types of reviews for each category** | | |
| --- | --- | --- |
| Here we can see game category have more number of positive reviews | | |
| **Fig 6: Types of installs per category** | | |
| The game category have more number of installed apps | | |
| **Fig 7: Free apps vs paid apps** | **Fig 8: Number of installed applications according to content rating** |
| Free apps have 92.22% and paid apps have 7.78% | from above plot we can see highest number of installed application Content Rating is everyone |

**4. Conclusion:**

That’s the analysis of dataset.

* Here we concluded as per above pie chart it can be easily understood positive user review sentiments is 65%, negative user review sentiment is 22% and neutral user review sentiments is 11.399%
* Positive reviews have more percentage than Negative and Neutral reviews
* health and fitness having high no of positive reviews as compare others
* total highest no of review has Game category and lowest no of reviews have comics category.
* highest number of app installed category is Game.
* Action And Adventure having more no of percentage as well as Trivia and Casio having lowest no percentage
* highest number of paid type category are family & Game
* highest number of free type category are Game & communication
* Paid Apps are having less bad reviews and Free apps having a equal distribution of Sentiment Polarity.
* highest number of installed application Content Rating is everyone.

**References-**

1. Almabetter course
2. Stack Overflow
3. medium