**EDA on Play Store App Reviews**

**Tejaswi Padarthi, Bhasker Purimitla - Data Analysts**

**Glucksort Private Limited**

**Abstract:**

Google play store is the official app store for all devices operating on the Android OS. It allows the users to browse and download the apps that are developed with the android software development kit (SDK). Apart from offering android applications and games, it also serves as a digital media store offering music, books, movies, and television programs. User ratings and reviews can significantly increase the number of app downloads; hence it is important to analyse the parameters which lead to users giving positive feedback and higher rating. Though this exploratory data analysis, we can understand and discover the key factors responsible for app engagement and success.

**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.

**• Installs:** The approximate number of times the respective app was Installed.

**• Type:** It states whether an app is free to use or paid.

**• Content rating:** It states which age group is suitable to consume the

content of the respective app.

**• Genres**: It gives the genre(s) to which the respective app belongs.

**Data Summary:**

**We are provided with two datasets:**

**● Play\_store\_data:** It contains the basic details of the app like number of user reviews, ratings, etc.

**● User reviews:** It contains the user reviews and its sentiment score for the respective app. We need to explore and analyse the data to discover key factors responsible for app engagement and success.

We need to explore and analyse the data to discover key factors responsible for app engagement and success.

**The contents of play\_store\_data are:**

● **App:** It contains the name of the app with a short description (optional).

**● Category:** This section gives the category to which an app belongs. In this dataset, the apps are divided among 33 categories.

● **Size:** The disk space required to install the respective app.

● **Rating:** The average rating given by the users for the respective app.

It can be in between 1 and 5.

● **Reviews:** The number of users that have dropped a review for the

respective app.

● **Installs:** The approximate number of times the respective app was installed.

● **Type:** It states whether an app is free to use or paid.

● **Price:** It gives the price payable to install the app. For free type apps,

the price is zero.

● **Android Ver:** The 3 rows containing NaN values were dropped from the

dataset.

**The contents of User Reviews are:**

**• App:** It contains the name of the app with a short description (optional).

**• Translated Review:** It contains the English translation of the review dropped by the user of the app.

**• Sentiment:** It gives the attitude/emotion of the writer. It can be ‘Positive’, ‘Negative’, or ‘Neutral’.

**• Sentiment Polarity:** It gives the polarity of the review. Its range is [ 1,1], where 1 means ‘Positive statement’ and -1 means a ‘Negative statement’.

**• Sentiment Subjectivity:** This value gives how close a reviewer’s opinion is to the opinion of the general public. Its range is [0,1]. Higher the subjectivity, closer is the reviewer’s opinion to the opinion of the general public, and lower subjectivity indicates the review is more of a factual information.

**Data Cleaning:**

**• App:** The duplicate values in the dataset were dropped based on the this ‘App’ column.

**• Rating:** The 1463 NaN values were imputed with its category median value. Corresponding category median value.

**• Size:** Converted all the values into a single unit like (all values in (MB’s).

**• Type**: One row containing NaN value was replaced with a mode of the column.

**• Price:** The ‘$’ symbol was removed, and converted into numeric datatype.

**● Current Ver:** The 8 rows containing NaN values were dropped from the dataset.

**● Android Ver:** The 3 rows containing NaN values were dropped from the dataset.

**Data Visualization:**

It deals with the graphical representation of data, from which we can draw conclusions and take different business decisions.

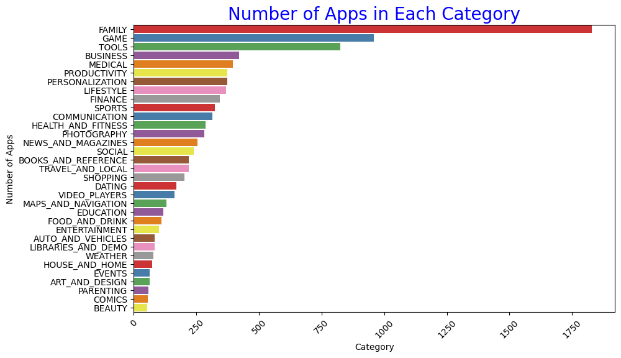
**Number of apps in each category:**

• The apps in the dataset are divided among various categories based on its applications and use-cases.

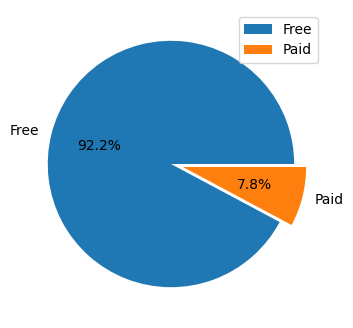
• In this dataset, the apps are divided into a total of 33 categories.

• The higher the number of apps in a category, the more competitive it is to launch an app in the said category.

• From the bar graph below, we can say that the “Family” category has the highest number of apps, followed by the “Game” and “Tools” category. From this we can say that these categories are the most competitive to get in to.

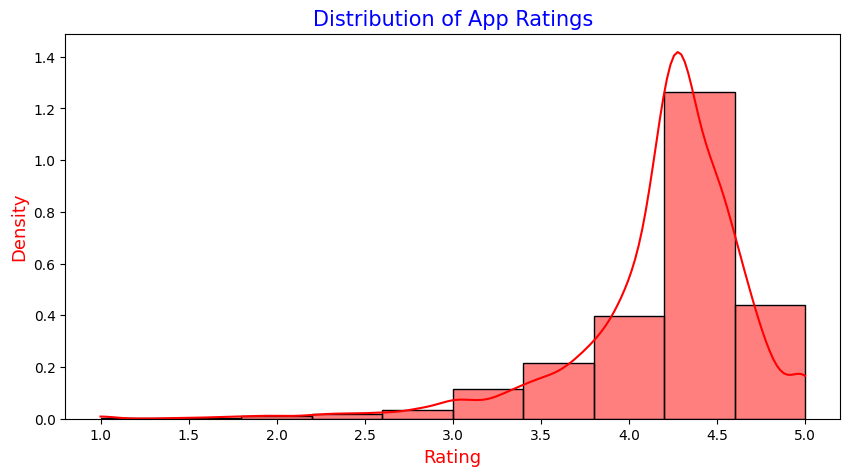


**Free apps and paid apps in the database:**



* Approximately 92% of the apps in the play store are free to install**.**

**Distribution of App Ratings:**



**•** The majority of ratings are concentrated between 4 and 4.5 stars, indicating

strong user satisfaction.

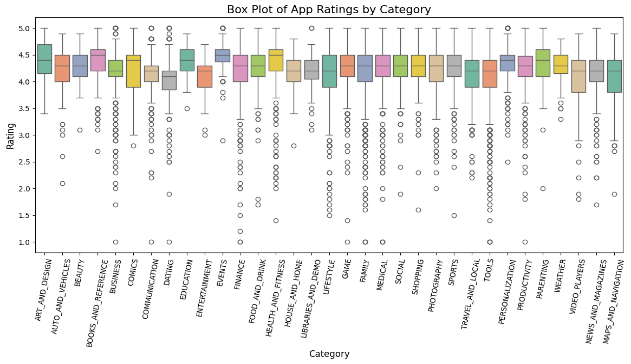
• Positive ratings highlight the app's success in meeting user expectations.

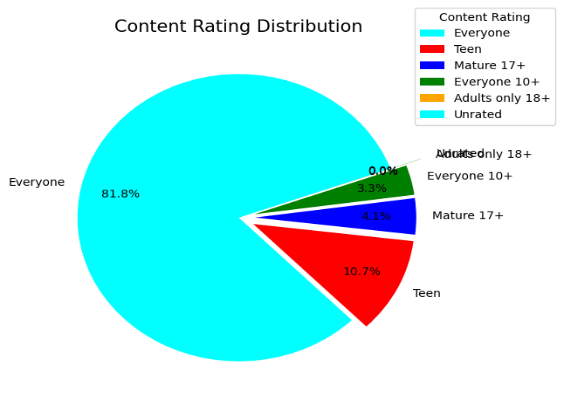
• This trend can be leveraged to attract new users and build credibility.

**Rating distribution across different Categories**:

● The Distribution of user rating for all the categories as shown below.

This metric can be used by a developer to find and study the categories which are not popular among the users and see what mistakes they are doing. Also, this metric can be used to find and study the categories which are popular among the users based on Rating and implement some strategies in their app.

**Content Rating:**



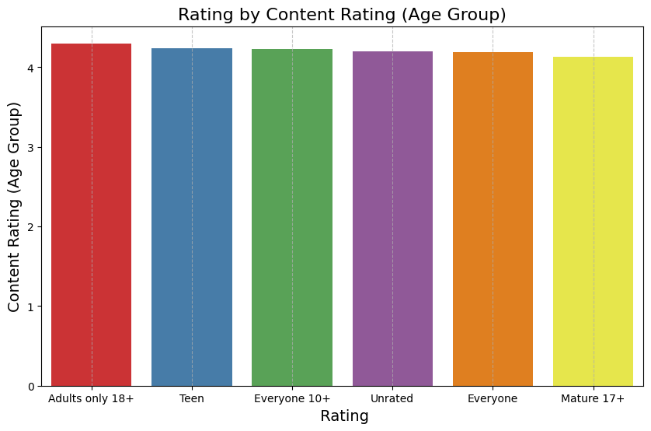
● Approximately 82% of the app in the play store have No age restrictions

to install and use the app.

● The rest of the apps have certain age restrictions on it.

• Around 4% of the apps are rated as “Mature 17+”, and around 3% of the apps as “Everyone 10+”.

**Rating Given by different Age group:**



• The graph illustrates the ratings provided by different age groups, offering

insights into user satisfaction across demographics.

• It is evident that the "Adults Only 18+" category have given higher ratings

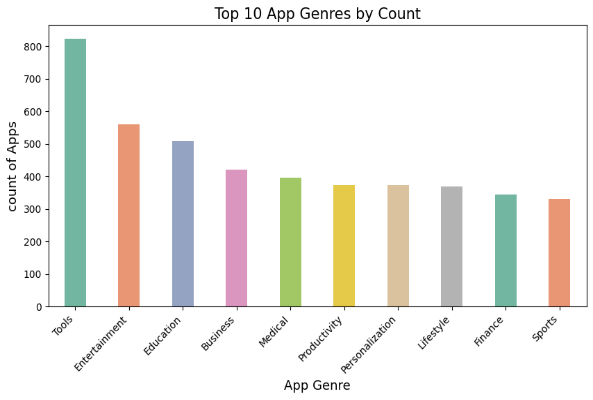
compared to other age groups, suggesting stronger satisfaction or alignment

with app content in these groups.

• Interestingly, the ratings across most age groups remain relatively similar,

indicating a consistent user experience regardless of age.

**Top 10 Genres Based on App Count**



• The graph highlights the top 10 genres with the highest number of apps on the

Play Store, showcasing their popularity and dominance.

• It is evident that most apps belong to these genres, indicating user interest and

developer focus in these categories.

• This trend suggests that developers have extensively explored these genres,

leading to a competitive landscape within them.

• To stand out and address untapped opportunities, developers should consider

focusing on less saturated genres to applications. Create innovative

applications.

**Price Distribution of Apps**

• The histogram reveals that the majority of app prices are concentrated between

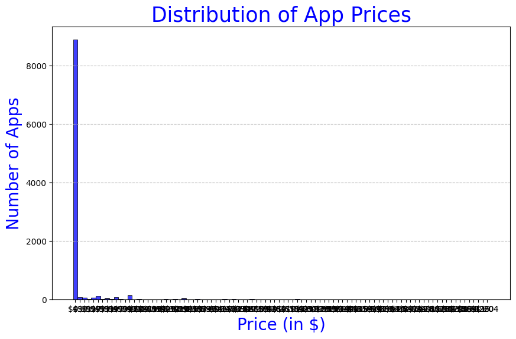
$0 and $50, indicating affordability for most users.

**•** This price distribution suggests that developers target a broad audience by

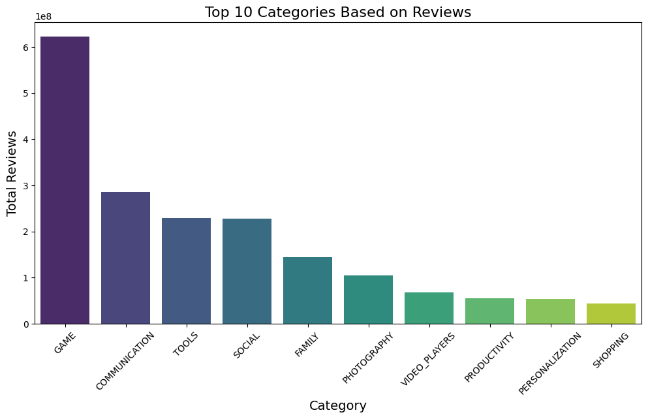
keeping app prices accessible.

**•** The concentration in this range highlights a potential opportunity to explore

premium pricing for niche or high-value apps.



**Top 10 Categories with Highest no of Reviews**



• The graph highlights the top 10 categories based on the number of reviews, providing a clear view of user engagement across categories.

• Categories such as Games and Social receive the highest number of reviews, reflecting their immense popularity and widespread user interaction.

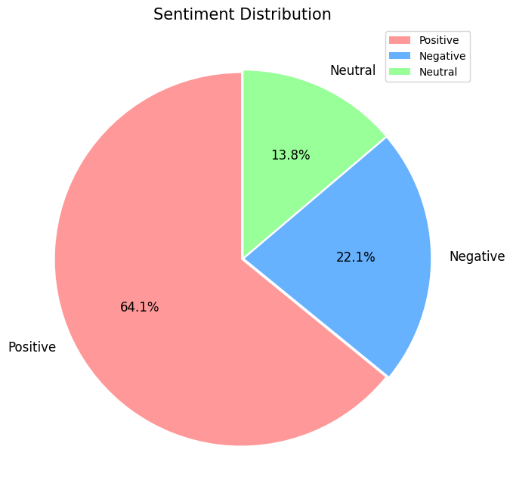
• This indicates that apps in these categories are highly effective in capturing user attention and encouraging feedback. • Developers aiming for higher user engagement and visibility can benefit from focusing on these popular categories or creating innovative apps within them.

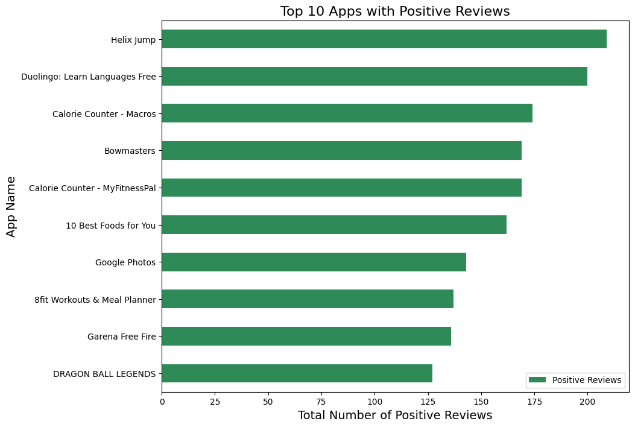
**Sentiment Distribution in reviews Dataset:**

• The graph shows that 64.1% of reviews express positive sentiment, highlighting strong user satisfaction with most apps.

• Negative sentiment, at 22.1%, points to areas where apps can be improved to better meet user expectations.

• Neutral sentiment makes up 13.8%, indicating an opportunity to engage these users and convert them into positive reviewers.



**Top 10 apps with positive reviews.**

• The graph highlights the top 10 apps with the highest number of positive reviews, providing insights into user satisfaction for each app.

• Helix Jump app leads the list with the highest number of positive reviews, followed by Duolingo: Learn Languages Free, indicating strong user approval for these apps.

• On the other hand, Dragon Ball Legends ranks last among the top 10 apps in terms of positive reviews, suggesting relatively lower user satisfaction compared to the others.

• These insights emphasize the importance of sustaining positive feedback while addressing concerns to improve the overall user experience.

• These findings can help developers understand user preferences and identify areas of improvement for apps with lower positive feedback.

**Conclusion:**

These are some of the aspects that the developer should research before proceeding with the app development. By conducting a simple exploratory data analysis. (EDA) on the play store dataset, we not only eliminate avoidable risks of failure, but we may also be able to provide better ideas for building the app.

* Most of the apps are **Free** with 92%.
* The category with the highest average rating is Events.
* The category with lowest average rating is Dating.
* Almost all apps are targeting everyone with 81.8%.
* All age groups have almost some average ratings.
* The Genre Tools has highest no of apps.
* Most competitive category: Games
* Most of the apps contains positive reviews rather than negative reviews.
* Helix Jump has highest positive reviews.
* Angry Birds Classic has highest negative reviews.
* Majority of the prices are between 0 to 5 dollars.
* The app with ratings falls between 3.0 to 4.5
* The app with highest no of installations.

Without any price (Free) is Hitman Sniper.

**Resources:**

* Greeksforgreeks
* ChatGPT
* W3 School Python