

CAPSTONE PROJECT SUBMISSION

Instructions

- I. Please fill in all the required information.
- II. Avoid grammatical errors.

TEAM MEMBER'S NAME, EMAIL AND CONTRIBUTION:

TEAM - DATA DEFENDERS

S No	Member Name	Email	Contribution
1.	Lokesh Tokas	lokesh.you@gmail.com	Data Exploration & Plotting
2.	Saraswat Mukherjee	mae21saraswat@gmail.com	Data Assessment
3.	Shubham Verma	shubhamvermavki@gmail.com	Data Relations
4.	Shubham Sartape	shubhamns19.pumba@gmail.com	Data Exploration and Plotting, Presentation

PLEASE PASTE THE GITHUB REPO LINK.

GitHub Link - <https://github.com/shubhamsartape/Play-Store-App-Review-Analysis---Capstone-Project>

PLEASE WRITE A SHORT SUMMARY OF YOUR CAPSTONE PROJECT AND ITS COMPONENTS.

DESCRIBE THE PROBLEM STATEMENT, YOUR APPROACHES AND YOUR CONCLUSIONS. (200-400 WORDS)

PROBLEM

The Play Store apps 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.

Each app (row) has values for category, rating, size, and more. Another dataset contains customer reviews of the android apps.

Explore and analyse the data to discover key factors responsible for app engagement and success.

APPROACH

DATA ASSESSMENT

- info()
- describe()
- head()
- tails()
- isnull().sum()
- outliers

DATA CLEANING/MANIPULATION BASED ON INITIAL ASSESSMENT

- Fixing 'Rating'
- Cleaning 'Install'
- Filling NA Value in 'Ratings' with mode value.
- Checking and fixing 'null' value in 'Type'
- Checking overall null values in data
- Fixing NA value in 'Type', 'Current_Ver', 'Android_Ver' with mode()
- Cleaning 'Price'
- Cleaning 'Size' :
- Dropping rows with Duplicate APP data
- Fixing 'Reviews' data type

EXPLORATION

- Top Category in Play Store.
- Top 3 Categories in Play Store by Content Rating
- Top 10 most installed paid app by installation no and their ratings.
- Paid vs Free app in Play Store
- Paid and free app by category in play store.
- App installs in each category.
- Top 10 category with Rating 4.0 and above.
- Top 10 category with highest average number of reviews.
- Paid app in each category by Price and no of installation.
- App installation by Size.
- Top App category with most reviews.
- Percentage of Review Sentiments
- Top 10 Apps with Maximum Positive Reviews
- Is sentiment subjectivity proportional to sentiment polarity?

CONCLUSION

After careful examination of the play store data and app reviews, we can see there are some key factors that are responsible for app engagement and success :

1. Category – Apps in **Family, Tools & Game** category are more likely to have higher success rate.
2. Content Rating – Apps with content rating of **Everyone** has as dominating market share.
3. Paid apps – Top installed apps in **Paid** type are mostly games with average rating of 4+ with installation numbers above 1M.
4. Market Share – **Free** type apps dominates the play store market share at **92.2%** while **Paid** types are only at **7.83%**
5. Instillation – Apps in **Game** category has overall highest no of installations.
6. Rating - Apps in **Family, Game & Tools** category are more likely to have higher user ratings.
7. User Engagement – App in **Social & Communication** category have very higher user engagement ratio as seen by comparing Review data.
8. Size – Apps with **size** between **7-24MB** have higher installation numbers.
9. User Comments – User are more compelled to give a feedback for apps in **Game** category, leading to an improved app experience.
10. 'Helix Jump', 'Duolingo: Learn Languages Free', 'Calorie Counter – Macros', These Apps have highest number of Positive reviews.
11. Sentiment Subjectivity is not always proportional to Sentiment Polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.