CAPSTONE EDA PROJECT
ON
PLAY STORE APPS REVIEW ANALYSIS

MUJTABA ALI, KIRAN JOSHI, PRATEEK SACHDEVA,
SAMRAT CHOWDHURY & SUBHAM SINHA
Data Science Trainees,
AlmaBetter, Bangalore

ABSTRACT

The Google Play Store is one of the largest and most popular Android app stores. It has an enormous amount of data that can be used to make an optimal model. A recent study analyzes that user preferences, user opinion for improvement, user sentiment about particular feature and detail with descriptions of experiences are very useful for an application developer. However, many application reviews are very large and difficult to process manually. This data set contains 13 different features that can be used for predicting whether an app will be successful or not using different features. This data set is scraped from the Google Play Store. This journal gives detailed information on feature extraction and the complete Data visualization done on this data set.

1. PROJECT GOALS

In this project, we investigate the different variables of Apps on Google Play Store that why and how certain applications succeed and others. Also, what is required for an application to be considered as successfully topping the charts. We attempt to use our analysis to answer the following problem statements:

- Find out the app category which application has been installed maximum number of times.
- What are the factors affecting the maximum downloads?
- Exploring the top 5 highest rated free and paid apps.
- Exploring the average rating for each category.
- Understanding the sentiment polarity of the users.

Data Collected

- App : Name of application.
- Category : Category of application.
- Rating : Rating of an Application (between 0.0 to 5.00).
- Reviews : Reviews given by users.
- Size : Size of app (Kb or Mb).
- Installs : Total No. Of installations.
- Type : Type of app (Free/ paid).
- Price : Price of app.
- Content Rating : the rating given by targeted audience.

- Genres : Genre of App.
- Last Updated : Date when app is updated last time.
- Current Ver : Current version of App.
- Android Ver : Android version of device.

2. DATA CLEANING

The structure of the dataset and clean data before analysis. Cleaning on the data provided to bring it in a cleaner, representable form. Missing values were also removed in this process.. We started the analysis by creating a data dictionary to understand the structure of the dataset and what each feature represents. We found out the unnecessary raw data that were making the dataset clumsy and then we removed and drop them from the data set to sort the data and applied different methods to find out the results of the problem statements mentioned above.

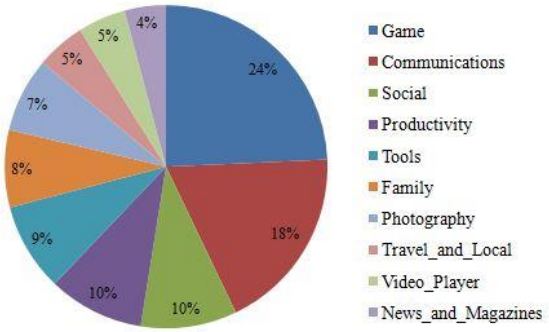
3. DATA EXPLORATION

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

Unique values from the given datasets. We applied methods like unique(), shape() and groupby() methods to understand to figure out the following inferences that game cat app has been installed max no of times.

i. Tops Apps Installed By Category

- Game, Communication, Productivity, Tools, Family, Photography, News & Magazines, Travel & Local, Video Players are untapped free app categories.
- High installs because of high underlying demand.
- Low user satisfaction.



Chart(a): Top Installed Apps By Category

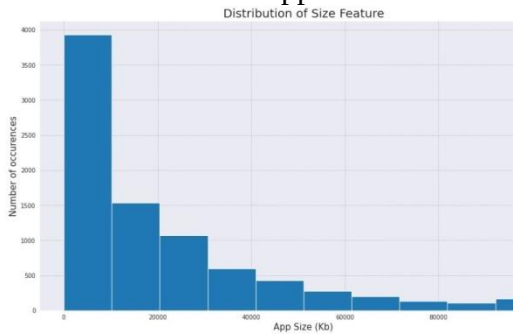
ii. Maximum Downloads

- The maximum no. of downloads is between 4.0 to 4.6.
- If the app is free to use then there is high probability to install.
- Paid apps are less likely to be installed.
- For the paid apps category maximum Installations are upto 20 dollars.
- Users prefer to pay for apps that are light-weighted. As app size get larger conversion rate decreases.
- If Developers are updating theirs app regularly then there is high chance to install.
- As you can see in Histplot, developer should give updates within 1 to 2 years.

Chart(b): Bar graph of maximum downloads

iii. Factors Affecting Maximum Number Of Downloads

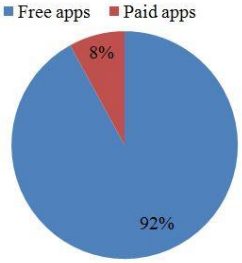
- The size of app indirectly proportional to No. of Installation.
- If the size of app is less than or equal to 10MB then there is high chance to install that application.



Chart(c):Distribution of size features

iv. Free vs Paid Apps

- 92% of apps on play store are free to download and rest are paid



Chart(d): Free vs Paid Apps

- Top 5 Highest Rated Free Apps
  - DF Glue Board
  - CN Resident
  - CL Strength
  - Color CL
  - CP Trivia
- Top 5 Highest Rated Paid Apps
  - USMLE Step 2 CK Flashcards
  - 211:CK
  - 30WPM Amateur ham radio Koch CW Morse code tra...
  - Morse Player
  - AC DC Power Monitor

v. Rating

- The average ratings across categories is statistically different.
- The Events, Education, Art and Design, Books and Reference produce the best apps with 50% apps having a rating greater than 4.5. Dating apps have a rating lower than the average.

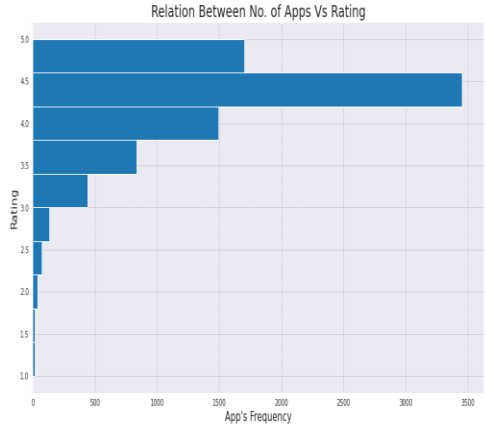
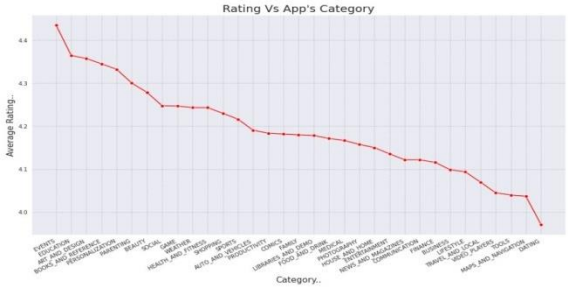


Chart (e): Relation between number of Apps and Rating

Average Rating for Each Category

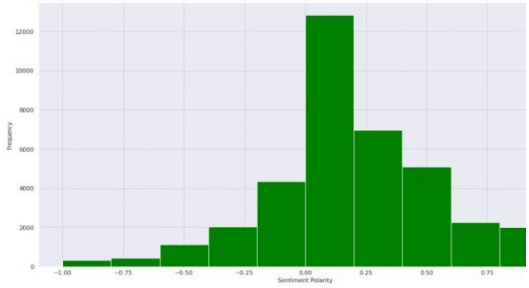
Event category has the maximum number of average ratings.



Chart(f): Rating vs Apps category

vi. Sentiment Polarity

- Maximum Apps around 40000 apps reviews Sentiment polarity lies between 0.0 to .18 (Neutral to towards Positive)
- Around 3000 apps reviews Sentiment Polarity lies between -1 to 0.0 means About 3000 apps have negative sentiment polarity.
- Mostly Apps Sentiment polarity lie between -0.47 to +0.8



Chart(g): Sentiment Polarity Graph of Users

4. CONCLUSION

Results we got as per our analysis :

- **Game** category has the most number of downloads and ratings, and **Event** category has the maximum number of average rating.
- The size and free/paid apps category affected the maximum number of downloads.
- We did not find any correlation between paid apps and number of downloads or ratings.
- Users prefer to pay for apps that are light-weighted. As app size get larger conversion rate decreases.