CAPSTONE PROJECT-1

PLAY STORE APP REVIEW ANALYSIS

By

- RAJESH MOHANTY
- SAIRINDHRI JENA
- MANISHWAR GUPTA
- ASHISH MAHARANA
- MANISHANKAR KUMAR SAW
- **DATA SCIENCE TRAINEE, ALMABETTER**

CONTENT

1. INTRODUCTION
2. PROBLEM STATEMENT
3. DATASET PREPARATION
4. DATA VISUALIZATION PLOTS
5. CHALLENGES FACED
6. CONCLUSIONS

INTRODUCTION

1. EXPLORATORY DATA ANALYSIS:

- Exploratory data analysis is the process of analyzing and interpreting datasets while summarizing their particular characteristics with the help of data visualization methods.

2. GOOGLE PLAY STORE:

- Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.
- Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy.
- Android is the dominant mobile operating system today more than 85% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store.
- There are more than 3.04 million apps found on Google Play Store.

3. TOOLS AND LIBRARIES USED FOR OUR EDA :

- Python, Numpy, Pandas, Matplotlib, and Seaborn

PROBLEM STATEMENT

- The objective is to explore and analyze the data to discover key factors responsible for app engagement and success.
- For that we are provided with two datasets, one is the play store dataset and another one is the user review dataset.
- We have to do actionable insights into the given database and should come up with the key factors that increased the number of users, long-term usage and etc.,

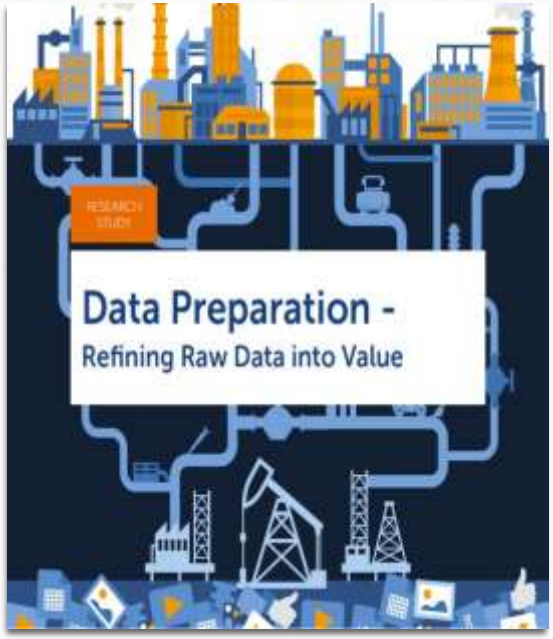
DATASET PREPARATION

1. **Importing the required libraries**:- Numpy, Panda, Seaborn, and Matplotlib.

2. **Loading the datasets**:- Mounting and reading the Play Store and User Review Datasets csv files given to us from the drive.

3. **Data cleaning:** Null values, Removing duplicate data.

4. **Exploratory Data Analysis:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.



ATTRIBUTES IN PLAY STORE DATA

1. **App:-** It tells us about the name of the application with a short description (optional).
2. **Category:-** It gives the category to the app.
3. **Rating:-** It contains the average rating the respective app received from its users.
4. **Reviews:-** It tells us about the total number of users who have given a review for the application.
5. **Size:-** It tells us about the size occupied by the application on the mobile phone.
6. **Installs:-** It tells us about the total number of installs/downloads for an application.
7. **Type:-** It states whether an app is free to use or paid.
8. **Price:-** It gives the price payable to install the app. For free type apps, the price is zero.
9. **Content Rating:-** It states whether or not an app is suitable for all age groups or not.
10. **Genres:-** It tells us about the various other categories to which an application can belong.
11. **Last Updated:-** It tells us about when the application was updated.
12. **Current Ver:-** It tells us about the current version of the application.
13. **Android Ver:-** It tells us about the android version which can support the application on its platform.

ATTRIBUTES IN USER REVIEW DATA

1. **App:-** Contains the name of the app with a short description (optional).

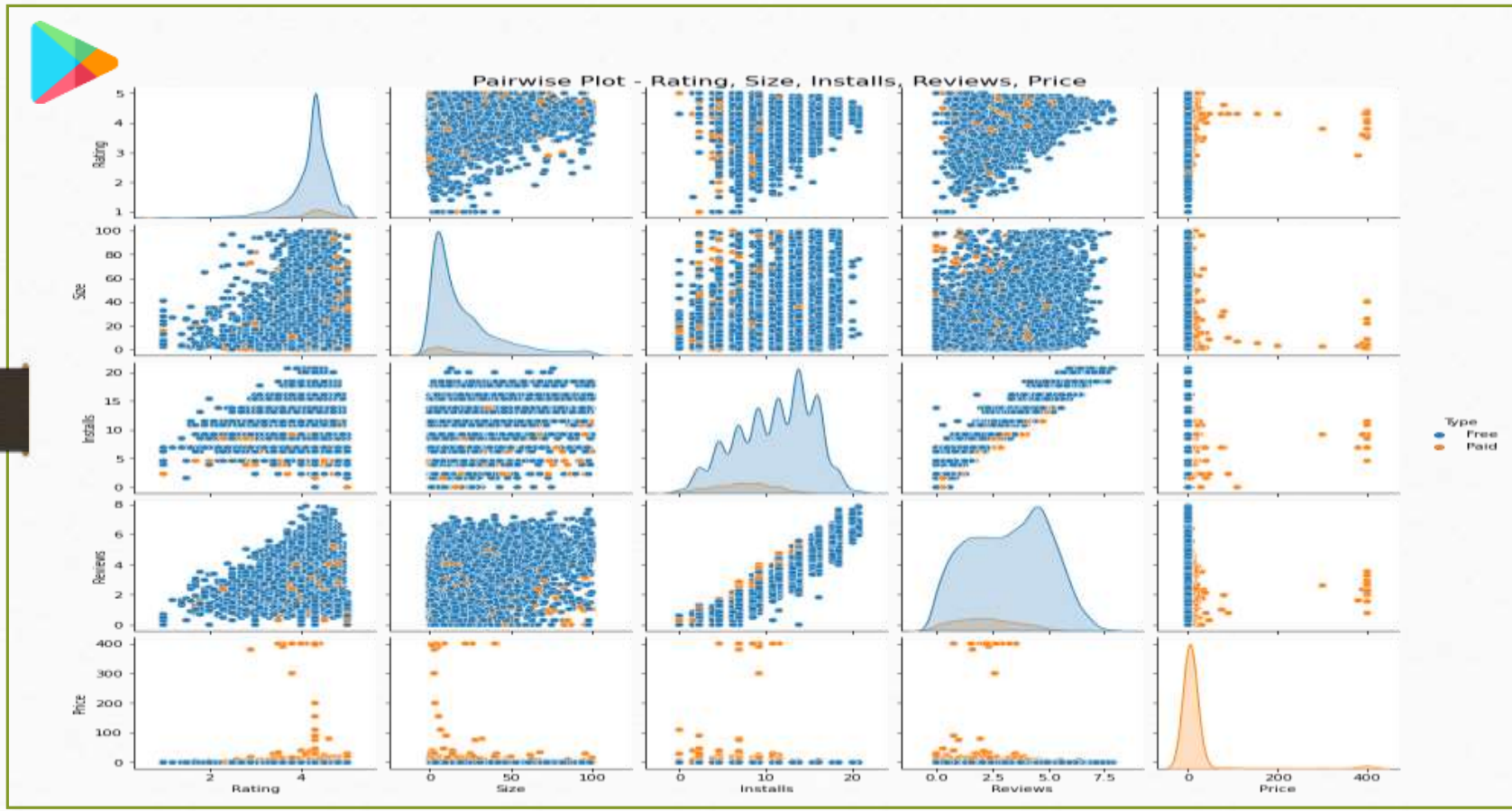
2. **Translated Review**:- It contains the English translation of the review dropped by the user of the app.

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

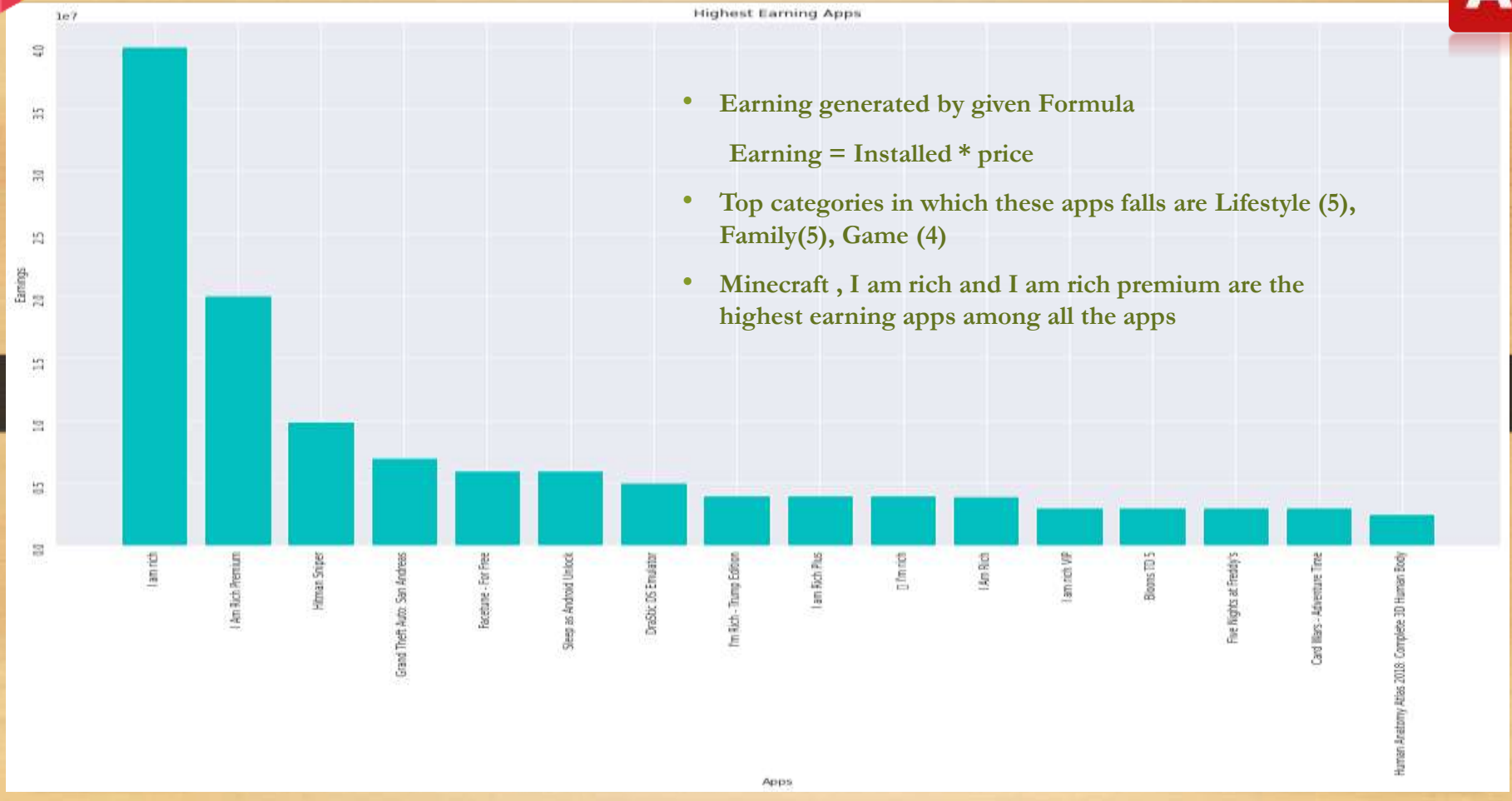
4. **Sentiment Polarity: -** It gives the polarity of the review i.e. ‘Positive statement’ and ‘Negative statement’.

5. **Sentiment Subjectivity:-** This value gives how close a reviewer’s opinion is to the opinion of the general public.

PAIRWISE PLOT - RATING, SIZE, INSTALLS, REVIEWS, PRICE



HIGHEST EARNING APP

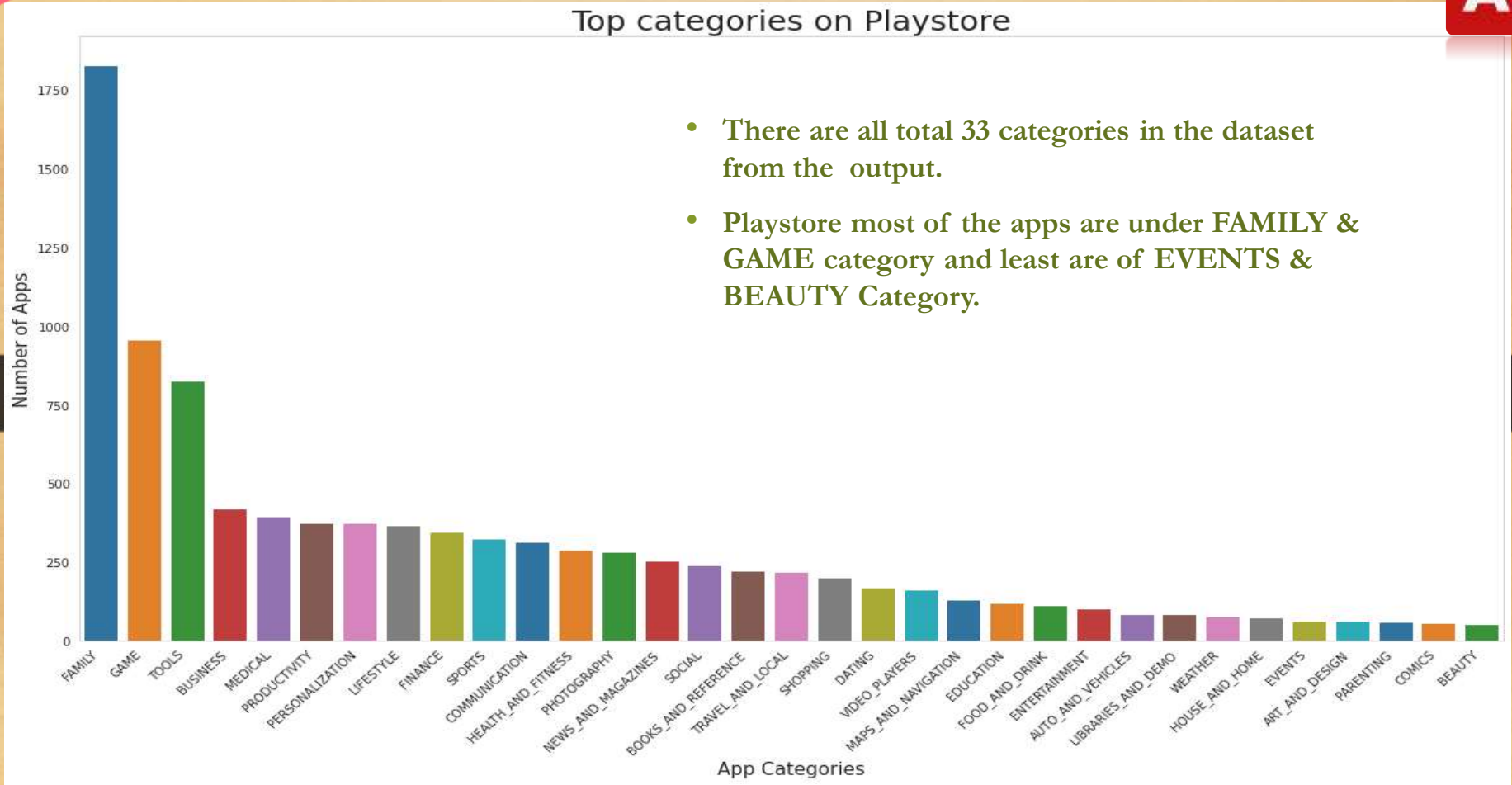


- Earning generated by given Formula

  Earning = Installed * price

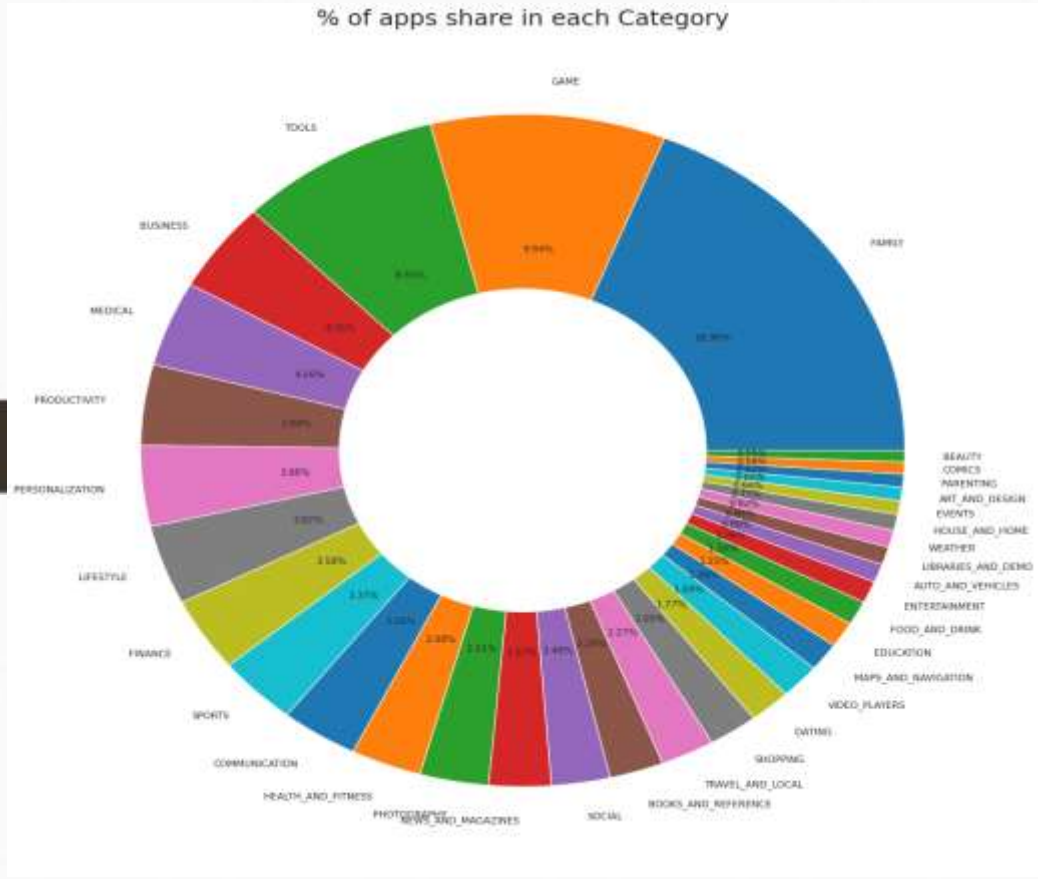
- Top categories in which these apps falls are Lifestyle (5), Family(5), Game (4)
- Minecraft , I am rich and I am rich premium are the highest earning apps among all the apps

TOP CATEGORIES ON PLAYSTORE



- There are all total 33 categories in the dataset from the output.
- Playstore most of the apps are under FAMILY & GAME category and least are of EVENTS & BEAUTY Category.

% OF APPS SHARE IN EACH CATEGORY



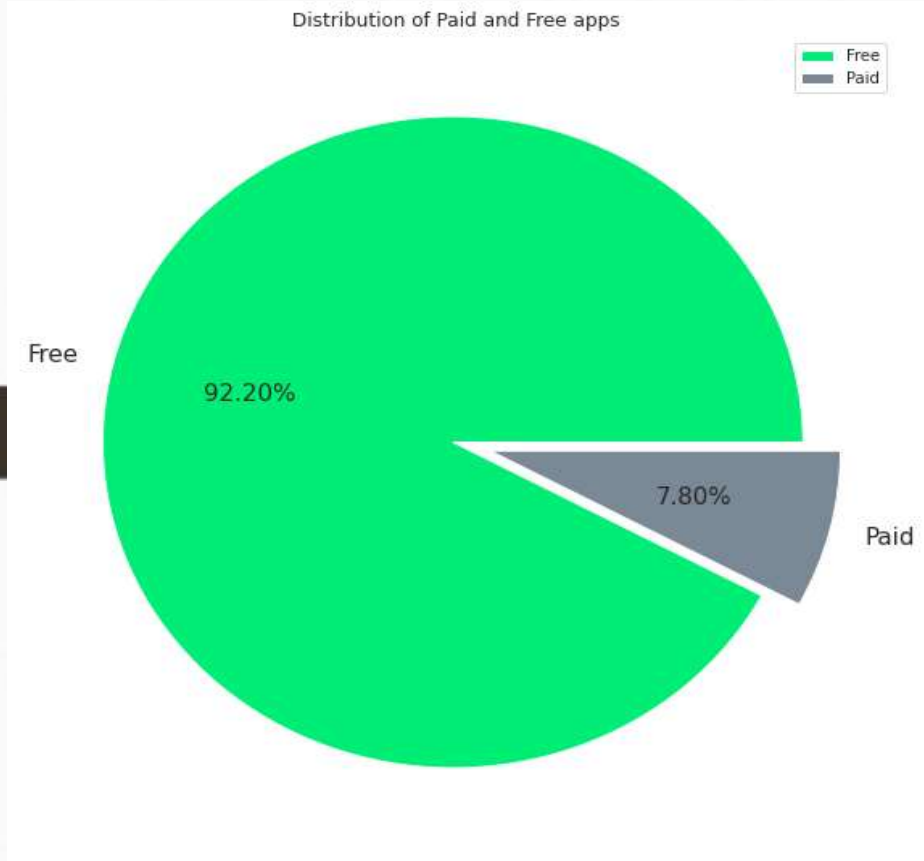
- The top 3 apps share based on % of categories are-

  Family=18.96%

  Game =9.94%

  Tools =8.55%

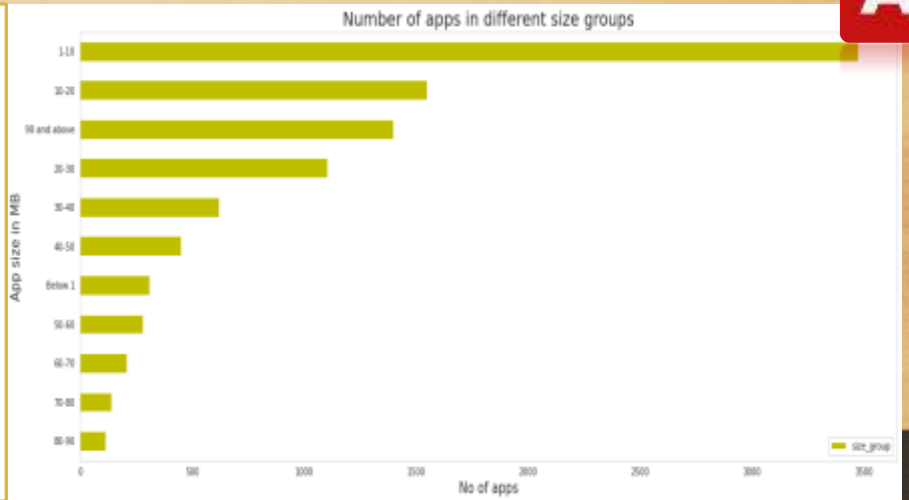
DISTRIBUTION OF PAID AND FREE APP

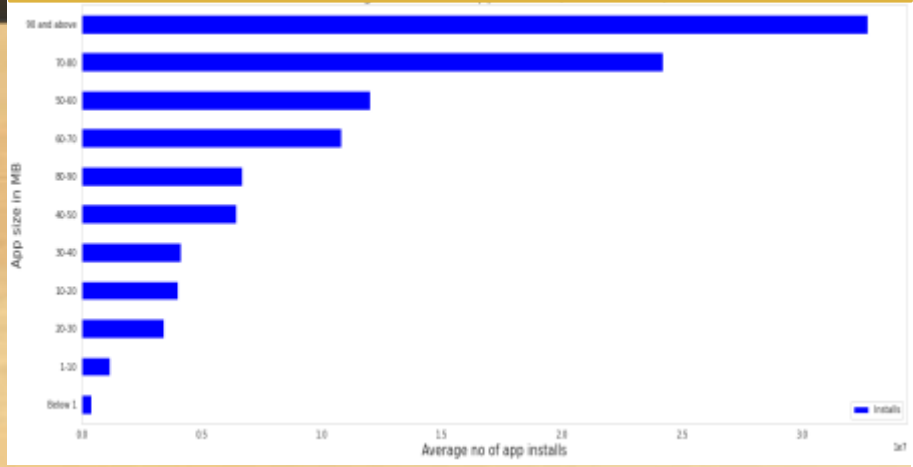


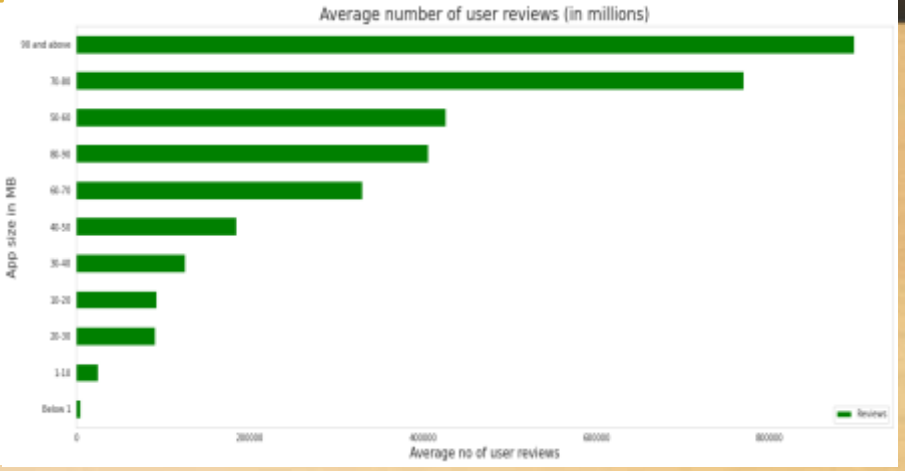
We observed that 92.20% of Apps are free and 7.80% of Apps are paid in the play store

ANALYSIS OF APP SIZE

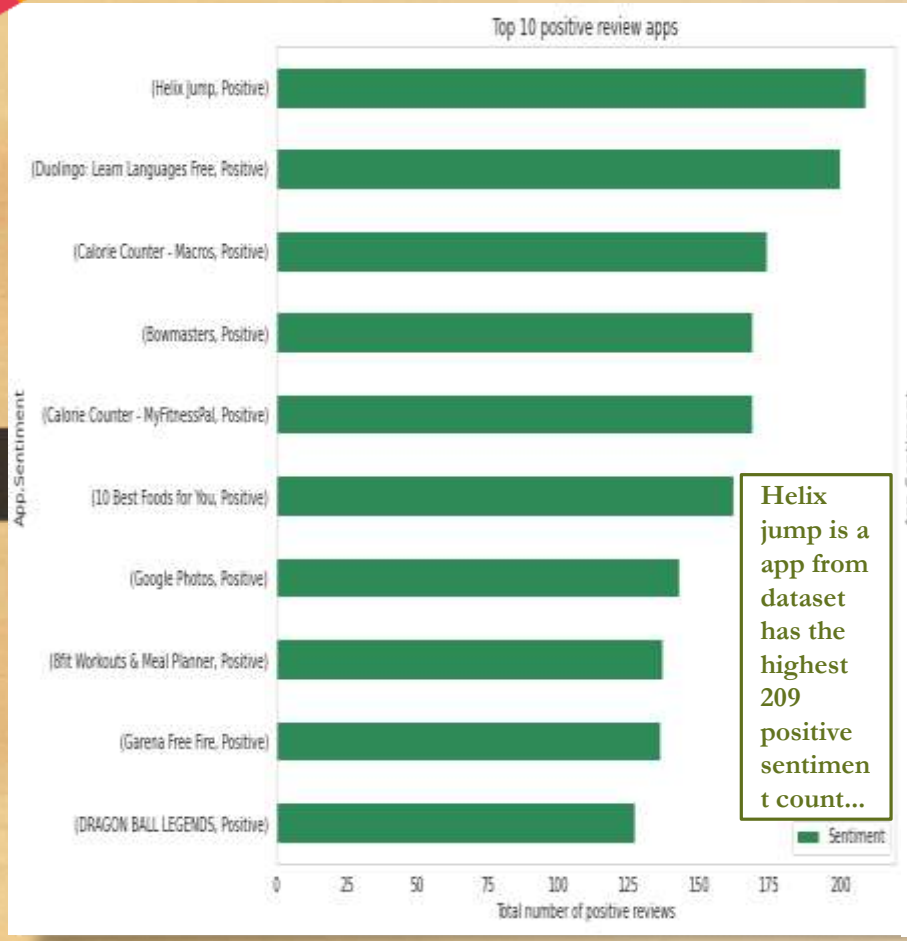
- The sizes of the majority of the apps range between 1 and 20 MB.
- The apps are categorized based on their size between 1 to 100 MB with intervals of 10 MB each.
- The apps which are smaller in size on average have lower no of app installs and user reviews



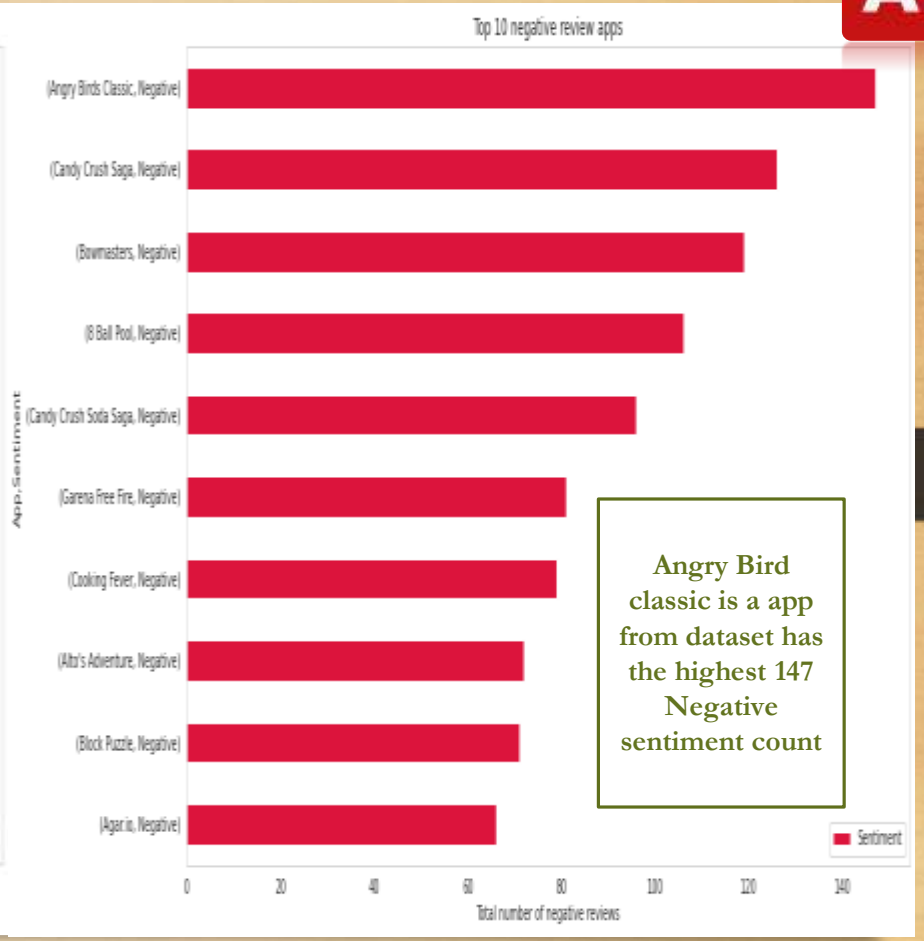




TOP 10 POSITIVE/NEGATIVE REVIEW APP

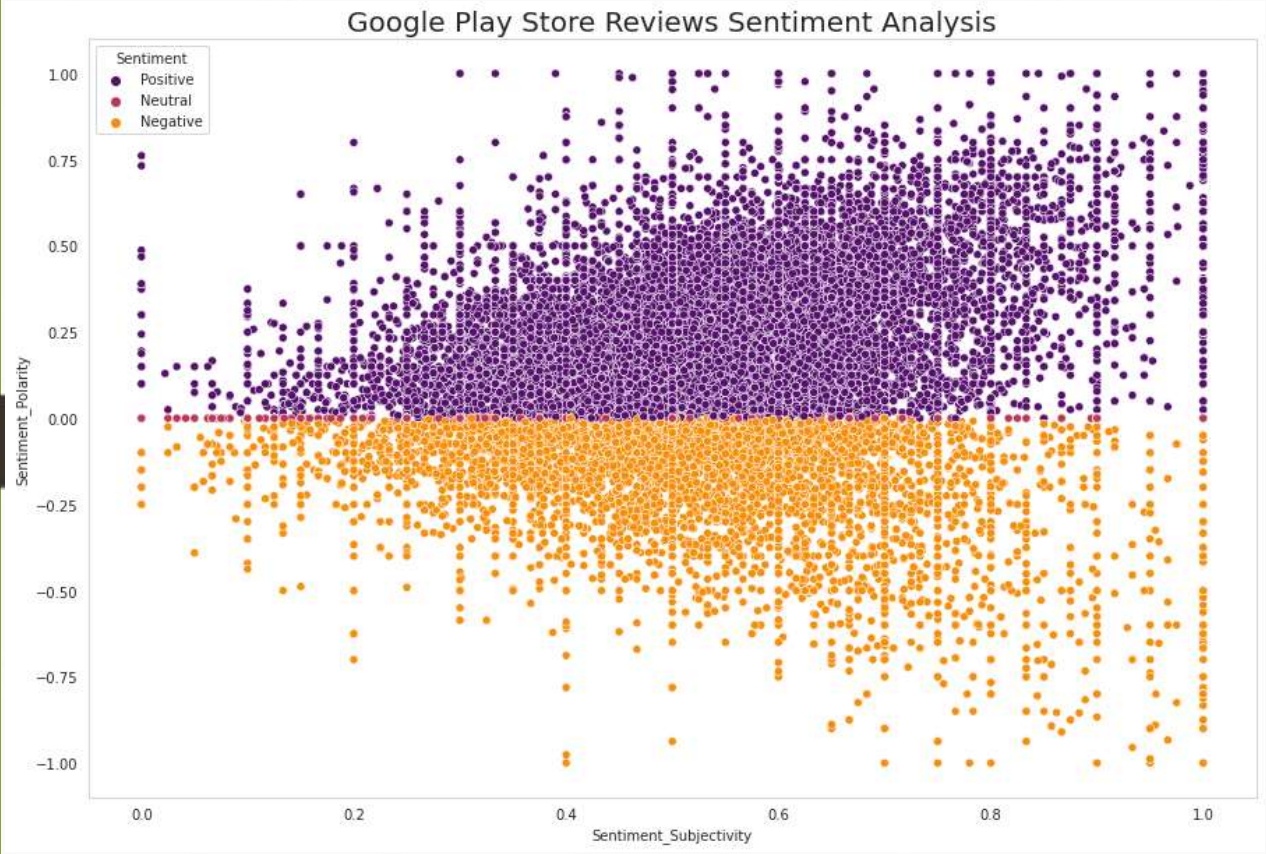


Helix jump is a app from dataset has the highest 209 positive sentiment count...



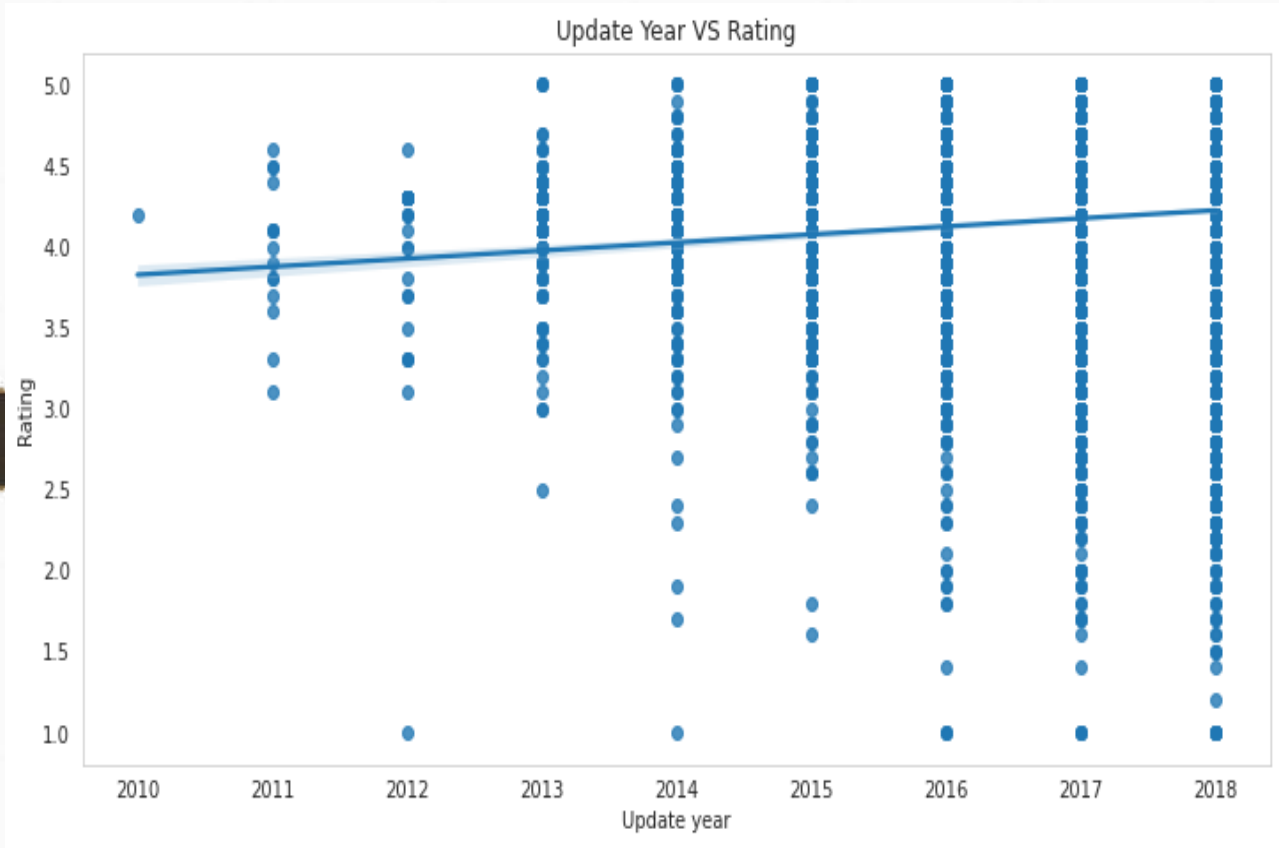
Angry Bird classic is a app from dataset has the highest 147 Negative sentiment count

GOOGLE PLAY STORE REVIEWS SENTIMENT ANALYSIS



- From the scatter plot we got to know that sentiment subjectivity is not always proportional to sentiment polarity but in a maximum no of cases a proportional behavior when variance is too high or low

UPDATE VS RATING



- From this plot, we get to know that the rating is good when the app is updating.

CHALLENGES FACED

- Reading and Comprehending the datasets given to us.
- Cleaning of the data and handling the NaN and duplicate values.
- Designing different data visualization plots to better understand the results and trends.

CONCLUSION (CONTD..)

- 92.20% of Apps are free and 7.80% of Apps are paid in the play store
- Minecraft, I am rich and I am rich premium is the highest-earning app among all the apps
- Playstore‘s most of the apps are under the FAMILY & GAME category and the least are in the EVENTS & BEAUTY Category.
- The apps which are smaller in size on average have lower no of app installs and user reviews
- Helix jump is an app from the dataset that has the highest 209 positive sentiment count…
- Angry Birds classic is an app from the dataset that has the highest 147 Negative sentiment count

CONCLUSION

- Most of the apps are free so developers should focus on creating free apps to have a huge customer base
- 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


Thank
You