CAPSTONE PROJECT

Play Store App Review Analysis

Done By:

PARUL
SRIVASTAVA

CONTENT

INTRODUCTION

- Google play store is widely used by users to download the required applications onto their android devices.
- Here we are provided with two data sets one is of play store data and user review data to draw inferences from it.
- And find out how developers can work on enhancing it more to capture market.
- For that we will be using the provided data frame and will draw insights from it.

OBJECTIVE

- 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.

- We have to explore and analyze the data to discover key factors responsible for app engagement and success.

- And perform visualization to draw conclusions at the end.

OBJECTIVE

- To perform various steps for the better data understanding which includes data wrangling, data cleaning and data analyzing.
- Each (column) needs to be cleaned to remove unnecessary data which include null values or any other type.
- Further do visualization to draw conclusions for developers as what can be done in near future for more successful play store application.
- Inculcating necessary changes will lead to more user engagement and more usage of play store in near future.

DATA CLEANING

- Data cleaning is done by removing any null, nan or replicated values.
- Duplicated values are also removed to reduce redundancy in data.
- And after that we performed mean operations on column to fill any missing values.
- Like in case of content rating, current version, android version replicated values are removed etc.
- After this step data can be analyzed in a better way.

DATA ANALYZING

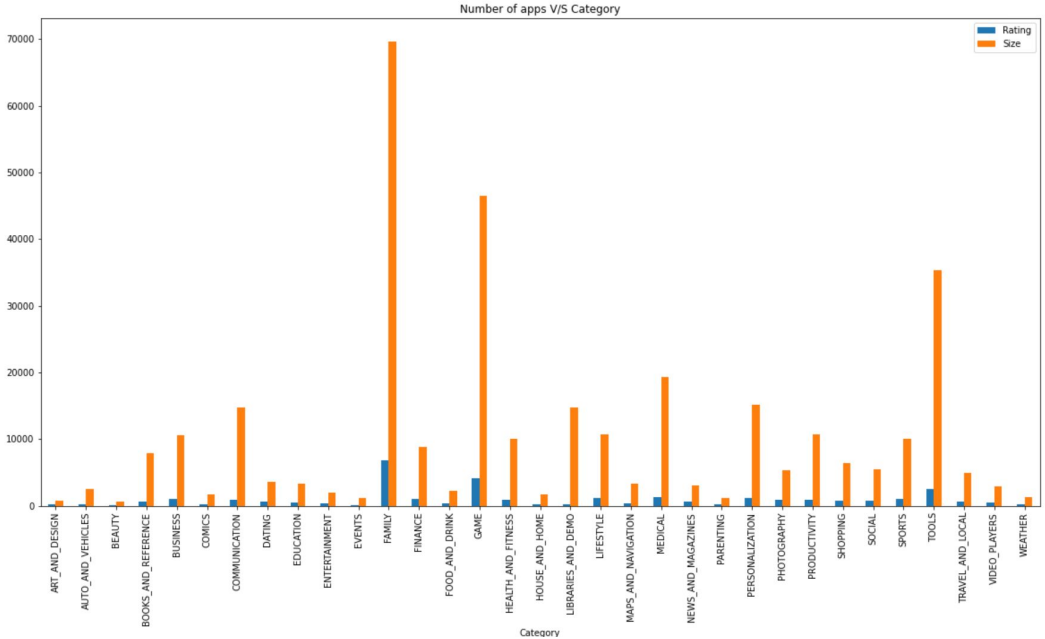
- Firstly, in our capstone project we started with data analyzing by using data wrangling operations to know more about our data.
- Further we performed various operations on raw data to gather process and transform it and draw insights from it.
- For which we began with importing the data and csv file from the google drive.
- After that we did basic operations like getting the head and tail of data, describing the data, knowing about what columns it have etc.

BASIC OBSERVATIONS

- All the applications with the five star rating.
- Average overall rating for all the applications of play store which turned around 4.16
- Top 5 categories getting the highest average rating.
- Count of apps whether free or paid was found out.
- Applications with the maximum reviews were shown.
- Top five applications having maximum number of installs followed by maximum number of reviews.

DATA VISUALIZATION

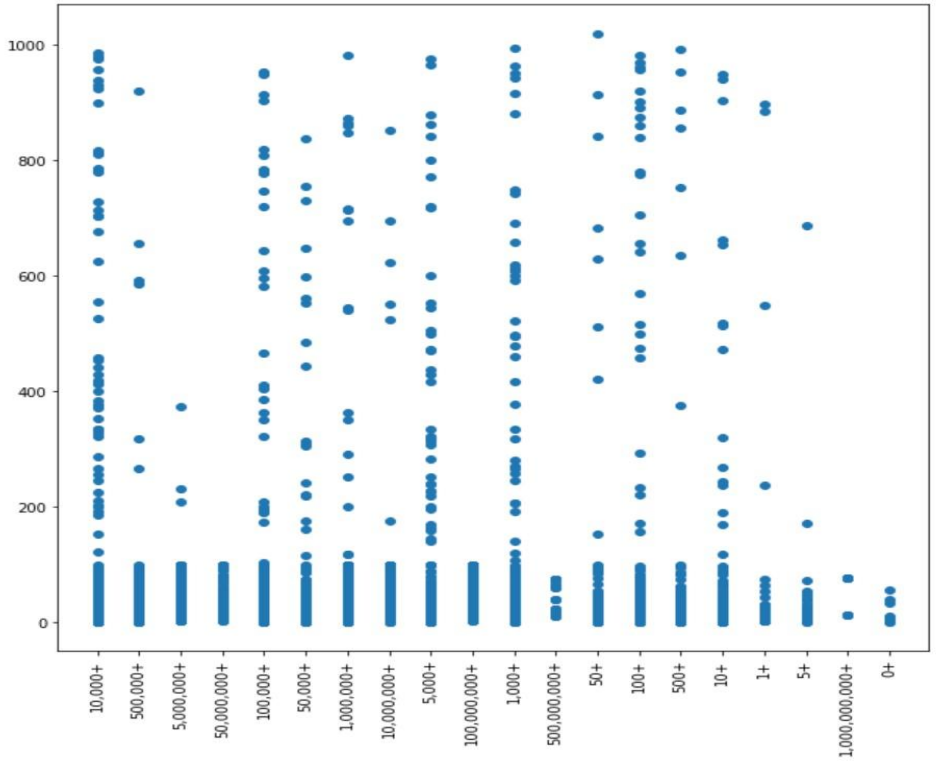
Which category was most famous among Play store users?



- The most famous category was found out to be Family.
- After that on the second highest position it was Games.
- Followed by the tools and medicine afterwards.
- From this plot observation can be drawn was that the developers can use this category wise count of apps for future references.

DATA VISUALIZATION

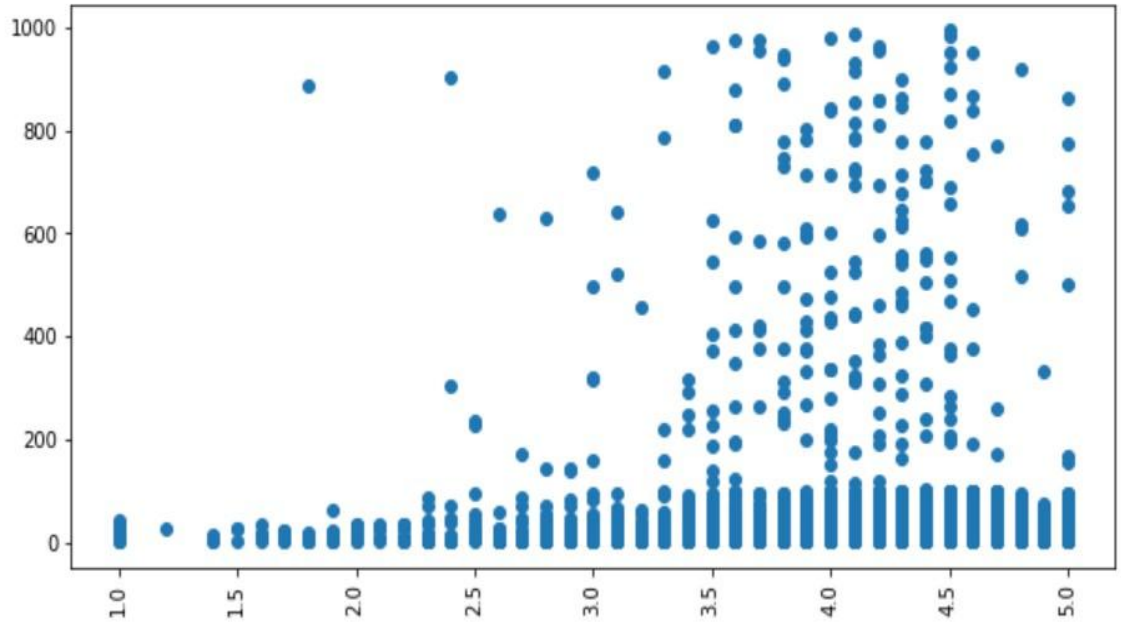
We have taken three columns under consideration to draw inferences:-



- From this scatter plot the inference made is that lesser the size of an application, more likely it is to be installed as compared to the applications with more size.
- As we can see with x-axis number of installs are given while with y-axis size of application is given.
- And as size increases the number of installs decreases.

DATA VISUALIZATION

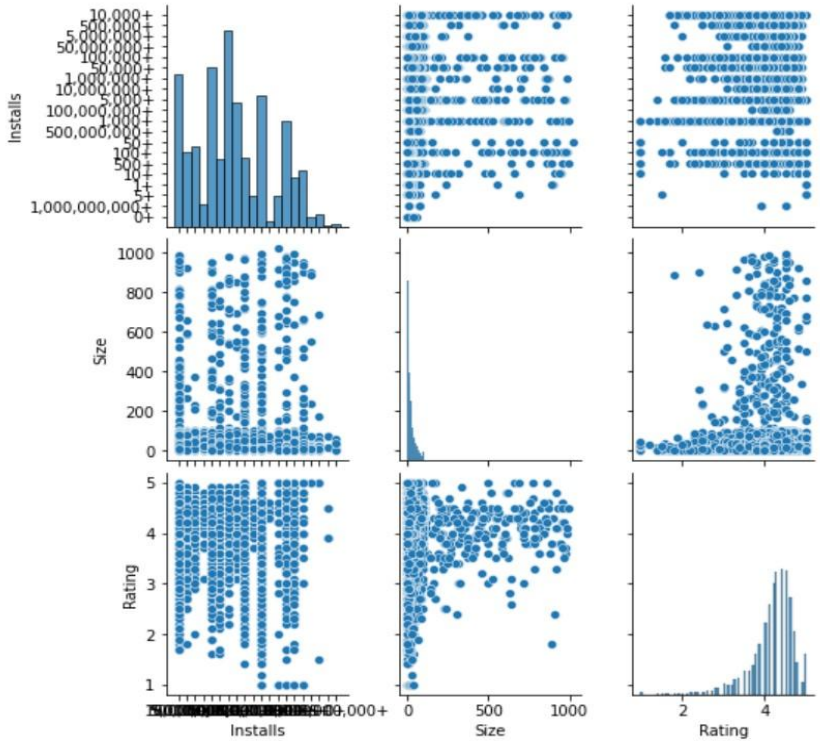
We have taken three columns namely installs, size and rating under consideration to draw inferences:-



- From this scatter plot the inference made is that lesser the size of an application, more better is its rating.
- As we can see with x-axis ratings are given while with y-axis size of application is given.
- And as size increases the ratings decreases.

DATA VISUALIZATION

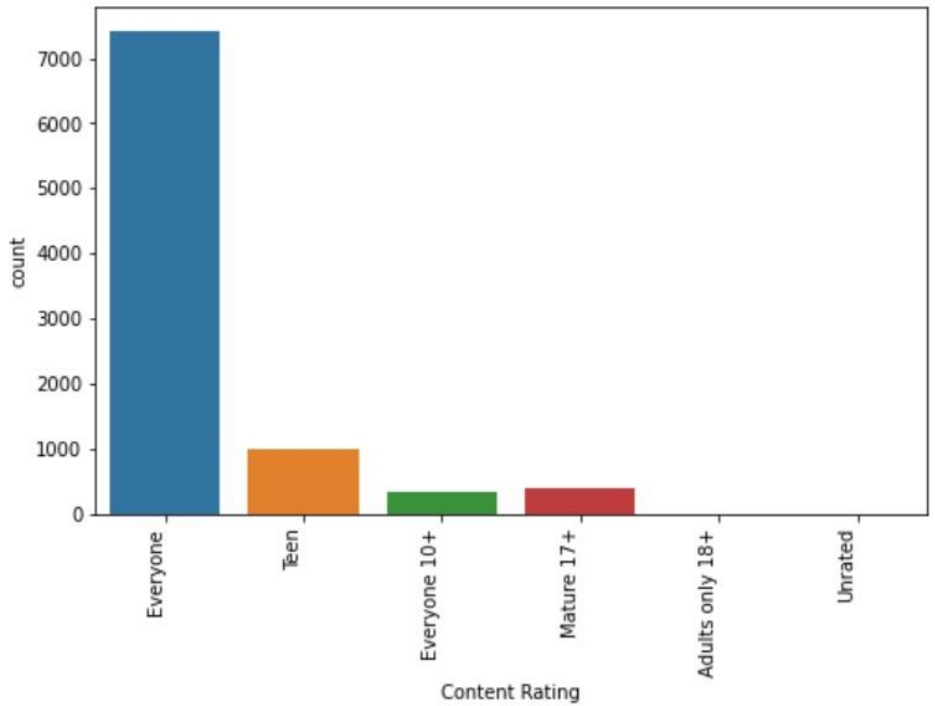
Pair plot of Installs, Size and Rating in data frame.



- From this scatter plot the inference drawn is the relationships among the various variables namely Installs, Size and Ratings.
- Dots is used to represent them, they are basically used to monitor how changing one variable affects the others.
- As here we can see greater the size is more are the number of installs as well as reviews.

DATA VISUALIZATION

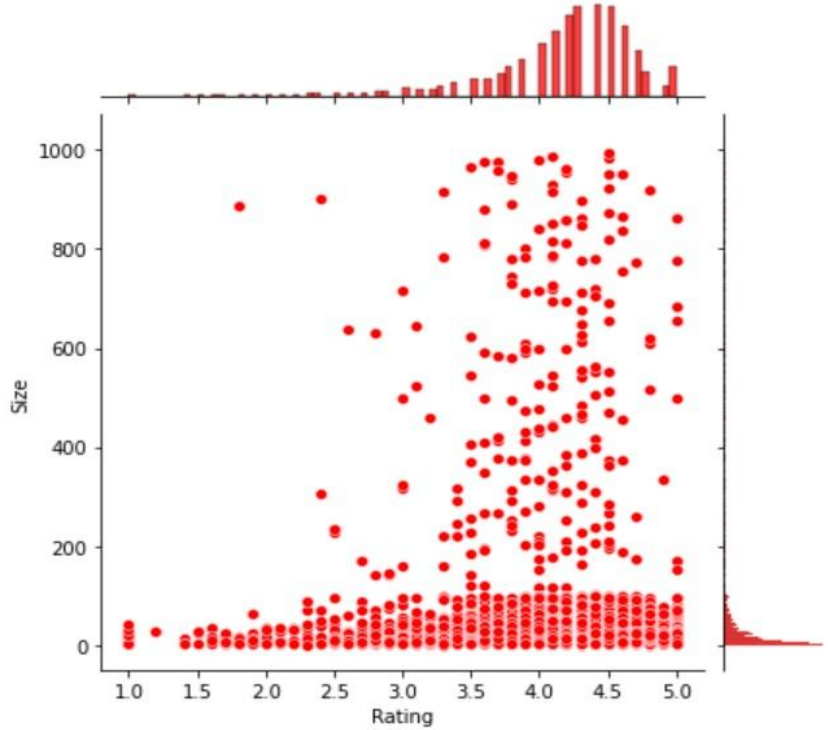
CONTENT RATING CATEGORY:-



- From this count plot of Category "everyone" is the widely used set of content rating used by the developers.
- Which makes it more age-friendly i.e. to be used by any age of people using play store application.
- Most of the applications can be accessed by every age group.

DATA VISUALIZATION

JOINTPLOT: Showing relationship between Rating and Size:-



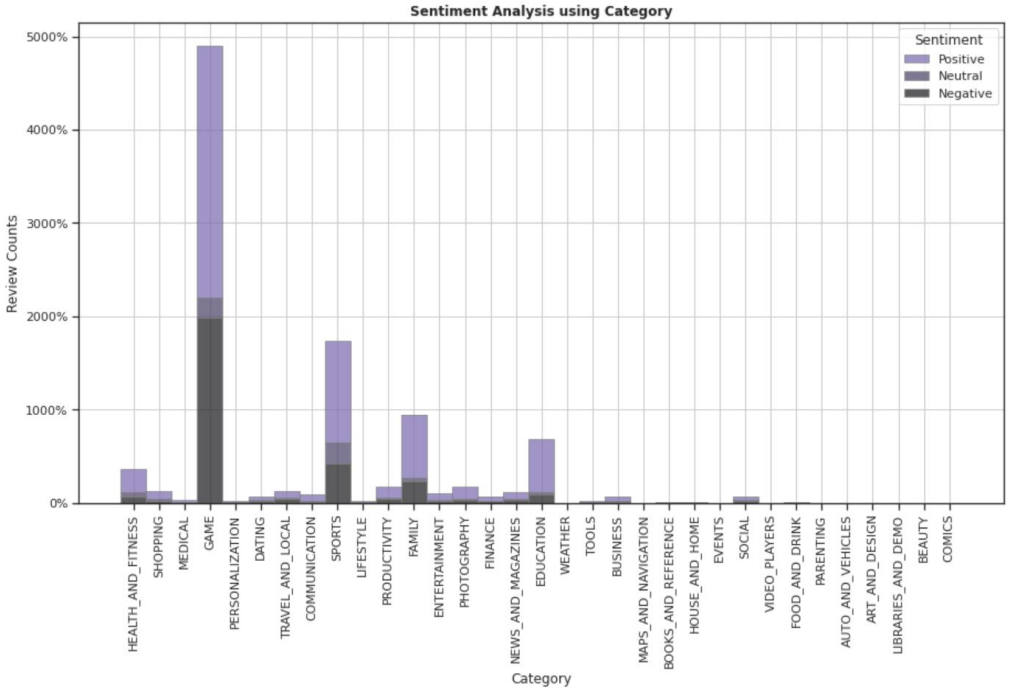
- From this joint plot the inference is that the applications have size around 30MB had the rating between 4.0 to 4.5MB.
- And so, we can conclude that applications which are smaller in size, like less than 20MB, can have 5.0 rating.
- Since, it is more preferred by users because of easy to download and occupying lesser space.

SENTIMENT ANALYSIS

- The user review data was firstly cleaned by removing null and nan values.
- After that we merged the columns app and category with the user review data set to analyse it in a better way.
- Sentiment analysis using category graph was plotted using matplotlib library to find out the sentiment and visualize them properly.
- Followed by a graph for user review sentiment was plotted.
- The conclusions were drawn from this data that maximum sentiments were positive followed by negative and neutral at the end.

SENTIMENT ANALYSIS

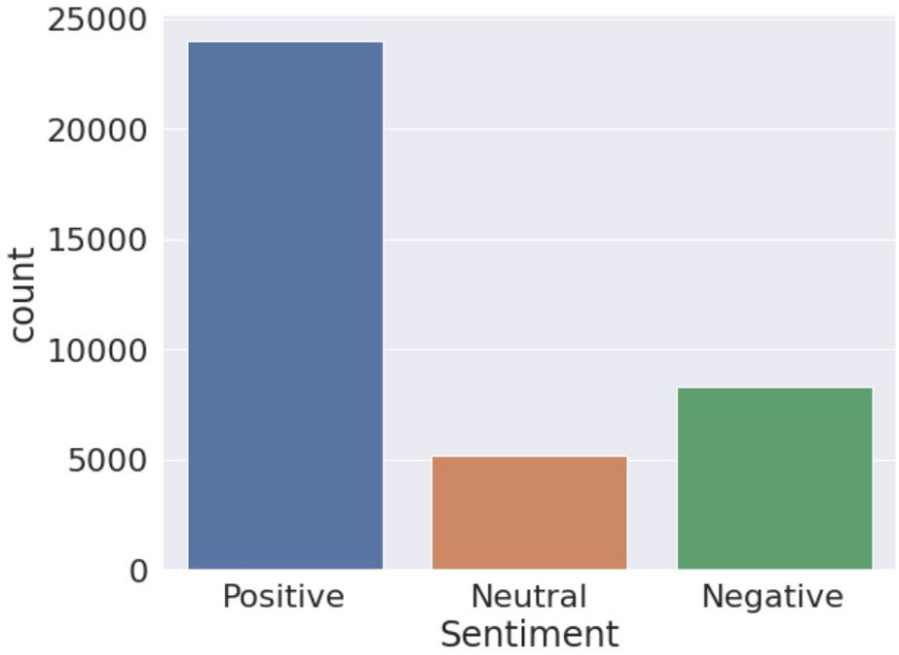
Sentiment analysis based on category



- The maximum review count is of games be it positive or negative.
- Followed by the second highest sports and after that education is on the third.
- It is performed to find out positive, negative or neutral sentiment.

DATA VISUALIZATION

SENTIMENT ANALYSIS OF USER REVIEWS:-



- The graph shows maximum sentiment of user reviews are positive.
- Positive sentiment shows user satisfaction with the application.
- Then comes the second highest as negative sentiment and neutral sentiment is the least one.

CONCLUSIONS

- From the Play Store App data we got to know that it has various applications present for all the purposes.
- What makes an application more successful is that it needs to be age friendly.
- Have a size that is light for users to use it with ease, with an average rating of 4.15
- Also, other factors which influenced users were positive sentiments, more number of installs, better reviews.
- Furthermore free applications were more preferred and used by the users of play store application.
- The number of installs increases for applications with more ratings.

FUTURE SCOPE

- The analysis can be used by developers to know how play store can engage more users compared to now.
- Also, developers can get an inference of what should be the size of an applications for users.
- The other scope is how positive sentiment helps in engaging more market place.
- Developers can make future applications age friendly as they are more successful.
- And key take away for developers can be that the applications should be free rather than paid ones as they are more preferred.

REFERENCES

For this exploratory data analysis we have taken help from sources as mentioned:-

- Almabetter Video Lectures and live classes.
- GeeksforGeeks Website.
- Towards Data Science Website.

THANK YOU