

Capstone Project Submission

Instructions:-

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name	Email
Varpe Ajit	ajitvarpe8652@gmail.com
Niharika soni	niharika.soni1851@gmail.com
Harisha Chennozswala	tasharisha@gmail.com
Satya Prakash	satya.prakash009@gmail.com

Contributor Roles:-

1. Niharika soni - Email: niharika.soni1851@gmail.com

A. Data loading

B. Data Wrangling

1. Deleting the duplicate rows.
2. Removing Outliers.

C. Data Analysis and Visualization.

1. Count of apps in each category.
2. Highest and Lowest rated Genres.
3. Total number of apps having 5 star rating.
4. Relation between app category and app price.
5. The app got highest and lowest price.
6. The apps got maximum and minimum reviews.

D. Objectives of EDA :-

1. What is the Count of apps in each category.
2. What are top 10 expensive app.
3. How sentiment is divided for different type of reviews?
4. Does sentiment_subjectivity proportional to sentiment_polarity?

2. Harisha Chennozswala Email: tasharisha@gmail.com

A . Data Wrangling.

1. Filling the null values with appropriate values using aggregate functions such as mean, median or mode.
2. Filling the missing categorical values with mode.
3. Conversion of Price, Installs into numerical values.
4. Missing values.

B. Data Analysis and Visualization.

1. Grouping all Apps storing in group.
2. Category wise rating.
3. Category wise Pricing.
4. Category wise Reviews.
5. Total number of free and paid apps.

6. The average rating of free and paid apps
7. What category of apps available for different age groups.

3. Ajit Varpe Email :- ajitvarpe8652@gmail.com

A. Data Wrangling :-

1. Checking duplicate values for same App name .
2. Removing +,\$,M and converting it to numeric value.
3. Converting review column to number.

B. Data Analysis and Visualization :-

1. Distribution of app rating.
2. Distribution of App Size.
3. Distribution of Subjectivity.
4. Number of Installed applications for each category.
5. Percentage of Review Sentiments.
6. How many apps were installed according to its type.
7. Finding avg rating of apps age group category wise.

C. Objectives of EDA :-

1. Count of applications in each category differentiated by their type
2. How many apps were installed according to its type
3. How does size impact the number of installs of any application?

4. Satya Prakash Email :- satya.prakash009@gmail.com

A. Data Wrangling :-

1. Converting all size of apps into a single unit of size i.e. MB.
2. converting into numeric value.

B. Data Analysis and Visualization :-

1. Histogram of log installs.
2. Histogram of size .
3. Average rating of free apps.
4. Average rating of paid apps.
5. Checking the average rating.
6. Which category has the highest average rating.
7. Finding apps available for age group category wise.

C. Objectives of EDA :-

1. top 20 apps present in the google play store as per their Genres.
2. Genres that are getting installed the most in top 20.
3. Does the category of app matter in the success probability of app.

Please paste the GitHub Repo link.

Personal Github Link:-

<https://github.com/AjitVarpe/Data-Analysis-of-playstore>

Teams Github Link:-

<https://github.com/Team-AlmaBetter/playstore-analysis>

Personal Drive Link:-

<https://drive.google.com/drive/folders/12YO5uaHTtcI4i5C9AGIDISK2eYZ4VtUa?usp=sharing>

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 Statement:-

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

So as the first step we had checked the sanity of data:

In this we had remove all the null values as well as garbage value so that we can get appropriate data.

So as the second step we divided our task in four sections:

- Analysis of Playstore data.
- Analysis of Reviews data.
- Analysis of distribution of type of reviews, category wise in the dataset.
- Analysis of Most download apps category and User interest category.

1. Analysis of Playstore data :-

In this part we have seen that which app category has maxximum download and which category have maxximum positive reviews.
Generally Checking and finding category that have maxximum download and reviews .

Conclusion: How rating and Download affect the apps.

2. Analysis of Reviews data :-

In this part we have seen that sentiment_subjectivity proportional to sentiment_polarity and does it affect the in apps success probability .

conclusion :-

sentiment_subjectivity is not proportional to sentiment_polarity.

3. Analysis of distribution of type of reviews, category wise in the dataset :-

In this part we have performed different EDA operation to retrieve more information form the datasets about the rating distribution and reviews .

conclusion :-

collected different information retrieved form the datasets and visualisation .

4. Analysis of Most download apps category and User interest category :-

In this part we have Analyzed most downloaded apps category and reviews sentiments and along with them we also analyzed the interconnection of sentiments subjectivity and sentiment polarity.

conclusion :- The most downloaded app category is Gaming . which has highest downloading .

So Third and Final step is about visualization:

Here we used in built libraries like **Matplotlib, Seaborn** for our data visualization. We have use different type of graph to visualize our data like bar plot bar, histogram, pie chart etc.

Approach:-

- Importing dataset of Playstore and reviews dataset in google colab notebook.
- Understanding the NaN, missing data's, duplicates data using .info, .describe, columns, .value_counts(), shape, Grouping dataset
- Merging dataset
- performing different EDA operation to retrieve more information form the datasets.

Conclusion :-

The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product. Dataset can also be used to look whether the original ratings of the app matches the predicted rating to know whether the app is performing better or worse compared to other apps on the Play Store.

It is not limited to the problem taken into consideration for this project. Many other interesting possibilities can be explored using this dataset.