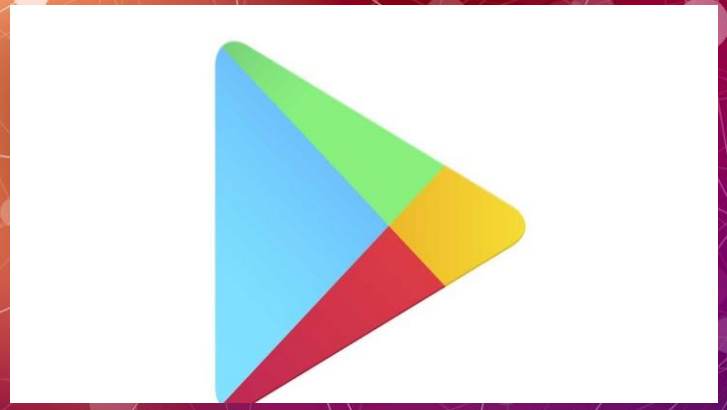
Capstone Project



Play Store App Review Analysis

Presented By-

Sriyanka Dhal

Department of Computer Science

College of Engineering Bhubaneswar

Outline

- Problem Statement
- Proposed system/solution
- System development approach
- Result
- Conclusion
- Future Scope
- References

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.

Proposed Solution

Description of App Dataset columns App : The name of the app, Category : The category of the app, Rating : The rating of the app in the Play Store, Reviews : The number of reviews of the app, Size : The size of the app, Install : The number of installs of the app, Type : The type of the app (Free/Paid), The price of the app (0 if it is Free), Content Rating :The appropiate target audience of the app, Genres: The genre of the app, Last Updated : The date when the app was last updated, Current Ver : The current version of the app, Android Ver : The minimum Android version required to run the app.

Cleaning Data-Dataset can contain missing data, numerical string value, various cues. If we can clean them, we can make easy our analysis.

Exploratory Data Analysis-corr() : It returns correlation. describe (): It returns number of entries, average of entries, outlier values, standard deviation, minimum and maximum entry.


Genres
Ver
Current
Price
Size
Content
Reviews
Type
Android
Updated
Rating
App
Category
Installs
Last

System Approach

The system approach section outlines the overall strategy and methodology for developing and implementing the Play Store app review analysis. Here is a suggested structure for this section.

System Requirements

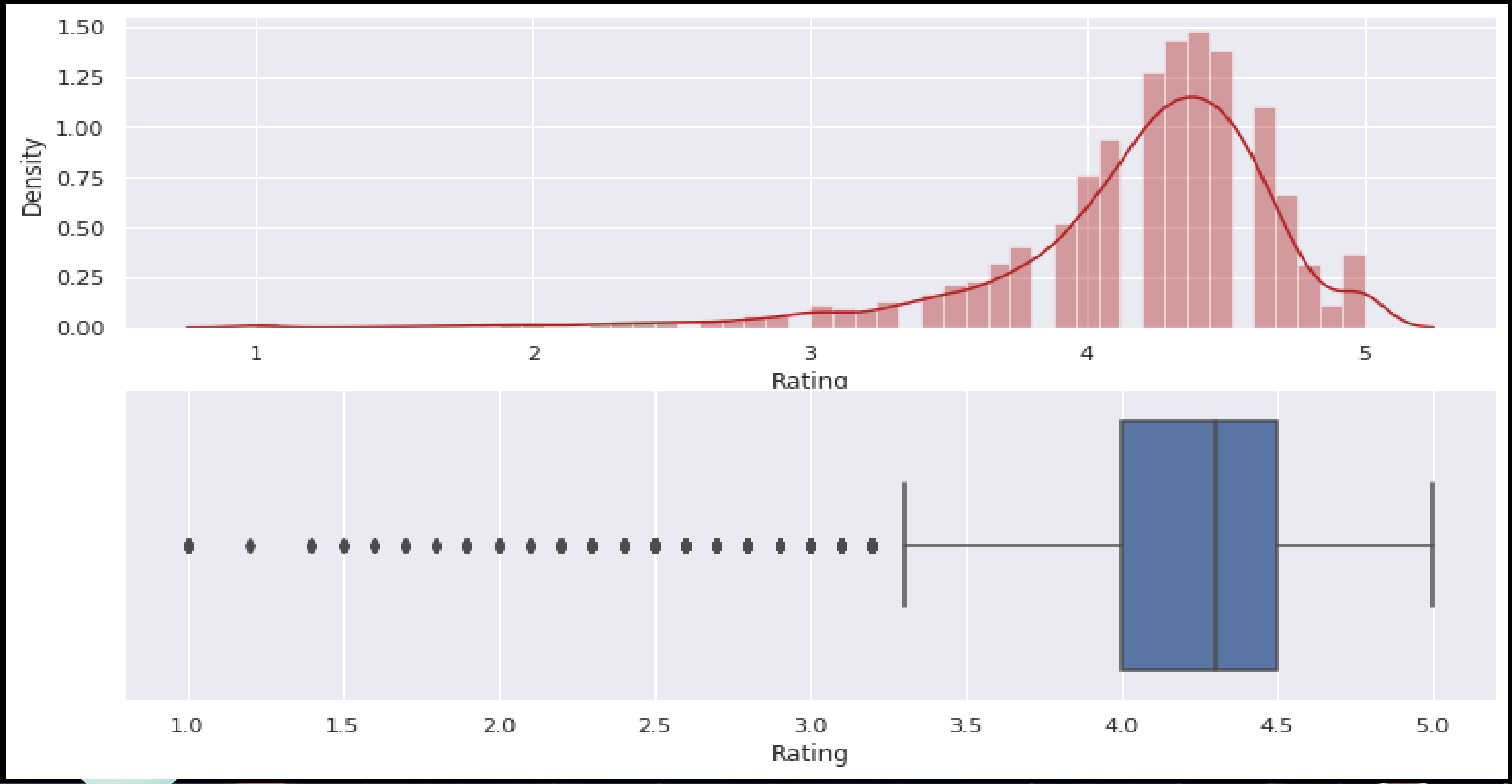
--> Data collection and integration(collect historical data, including logs, database and APIs provided, integrate real-time data pools such as weather APIs, ensure data integrity and consistency across different sources by implementing robust data collection and integration pipelines). -->Data preprocessing and feature engineering(perform data cleaning tp handle missing values, outliers and inconsistencies in the data set, conduct exploratory data analysis(EDA) to get the required insights, engineer features that capture relevant factors influencing app demand, normalize or scale numerical factors and encode categorical variables, to prepare data for modelling.

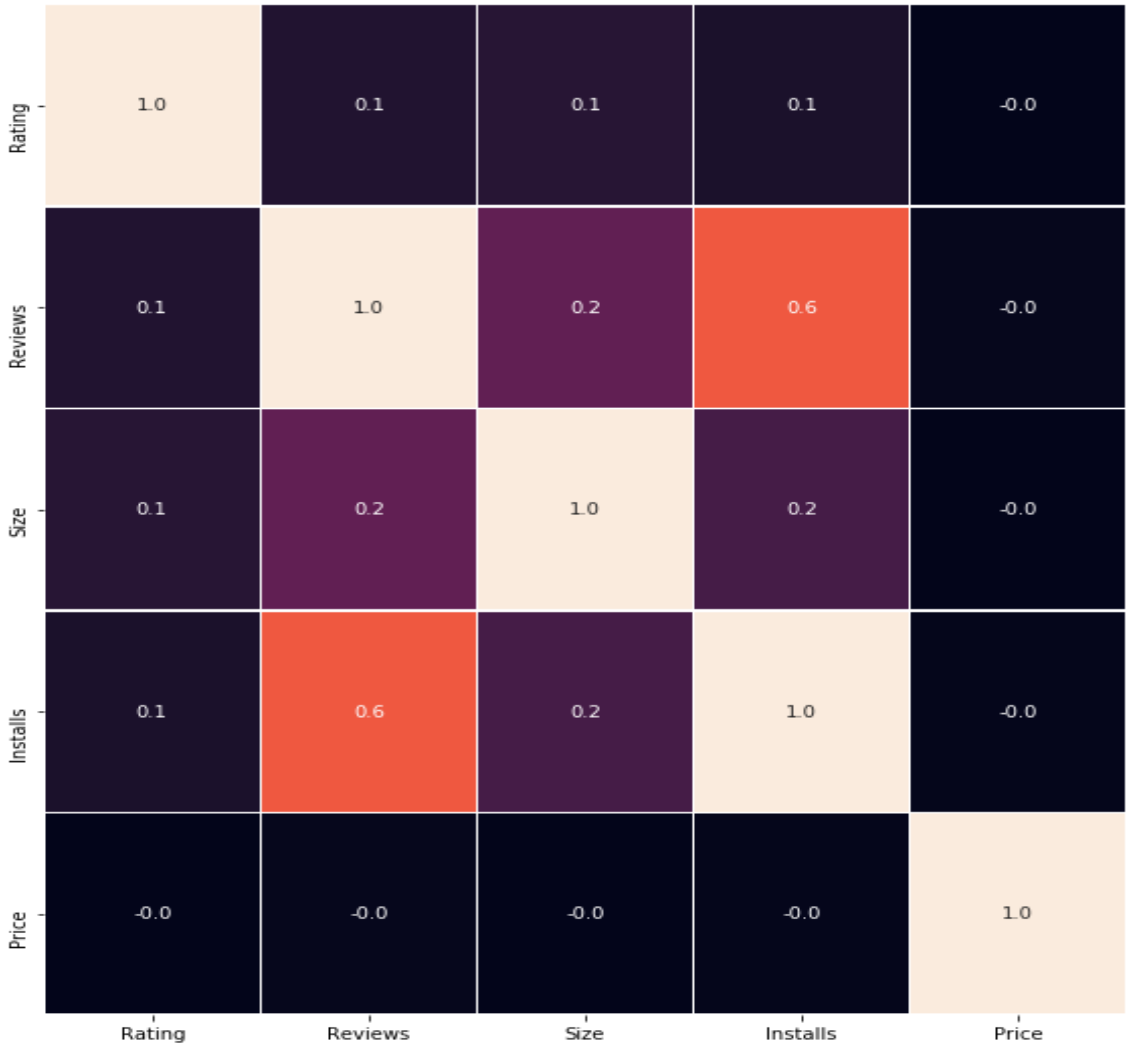
Library required to build the model

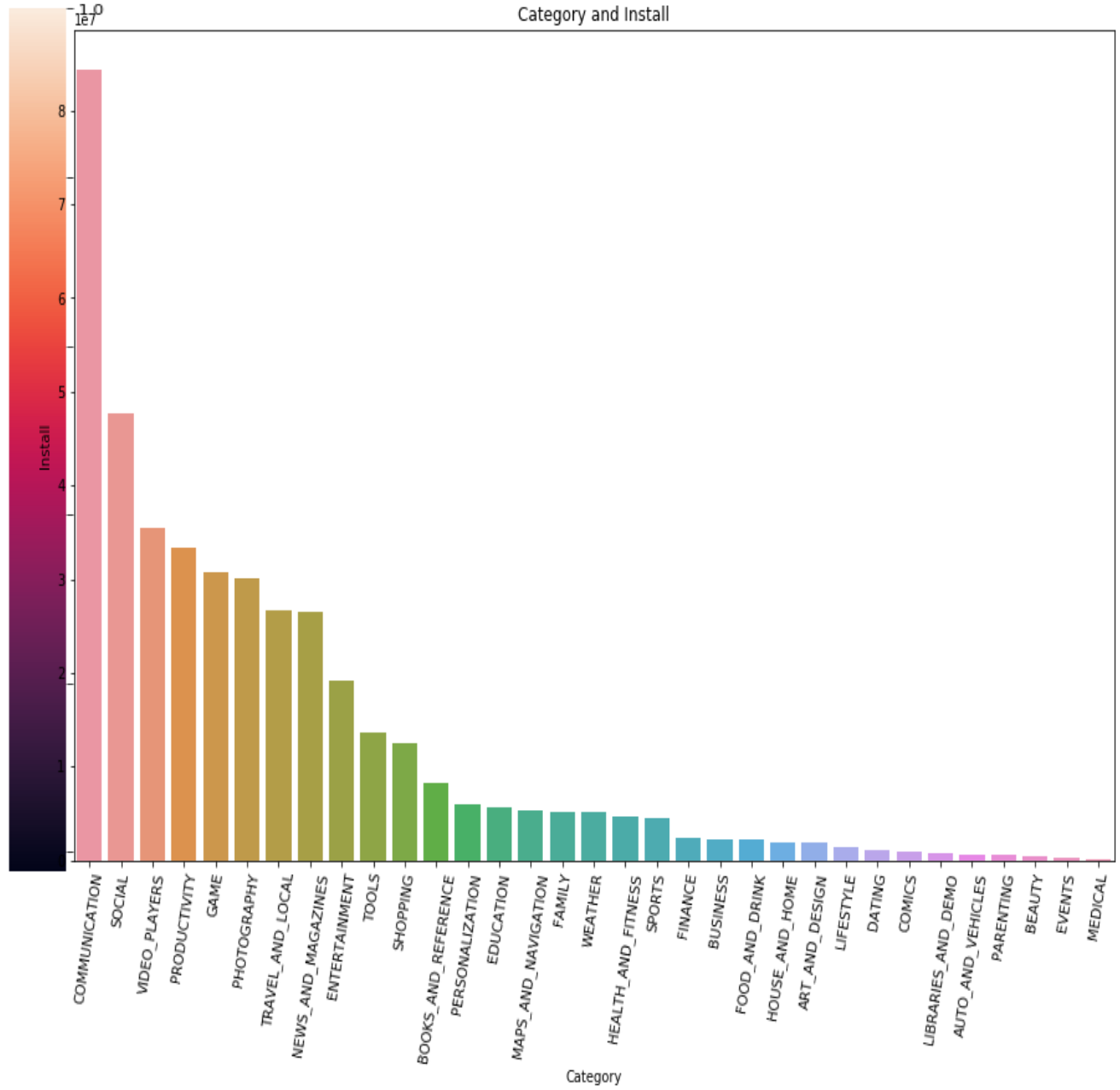
-> Data handling and manipulation(Pandas and Numpy). --> Data Visualization(Matpolib and Seaborn).

Results

- Sentiment analysis- By classifying user reviews into positive, negative or neutral sentiments, the project can provide an overall sentiment score for each app. This information can help developers gauge user satisfaction and identify areas of improvement.
- Feature importance- Through feature analysis and model interpretation, the project can identify the most influential factors affecting app success. This could include app category, rating, size, update frequency and specific keywords or phrases mentioned in reviews.
- Temporal trends- Analysis of review data over time can reveal temporal trends and patterns in user-sentiments. Developers can use this information to track changes in user-perception, identify seasonal variations and access the impact of updates of new features on app engagement.


1.50
1.25
1.00
0.75
0.50
0.25
0.00
Density
1
2
3
4
5
Rating
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating


Rating
Reviews
Size
Installs
Price
1.0 0.1 0.1 0.1 -0.0
0.1 1.0 0.2 0.6 -0.0
0.1 0.2 1.0 0.2 -0.0
0.1 0.6 0.2 1.0 -0.0
-0.0 -0.0 -0.0 -0.0 1.0


Category and Install
1e7
Install
Category
COMMUNICATION
SOCIAL
VIDEO_PLAYERS
PRODUCTIVITY
GAME
PHOTOGRAPHY
TRAVEL_AND_LOCAL
NEWS_AND_MAGAZINES
ENTERTAINMENT
TOOLS
SHOPPING
BOOKS_AND_REFERENCE
PERSONALIZATION
EDUCATION
MAPS_AND_NAVIGATION
FAMILY
WEATHER
HEALTH_AND_FITNESS
SPORTS
FINANCE
BUSINESS
FOOD_AND_DRINK
HOUSE_AND_HOME
ART_AND_DESIGN
LIFESTYLE
DATING
COMICS
LIBRARIES_AND_DEMO
AUTO_AND_VEHICLES
PARENTING
BEAUTY
EVENTS
MEDICAL

Conclusion

In this project of analysing play store applications, we have worked on several parameters which would help us to do well in launching their apps on the play store.
In the initial phase, we focused more on the problem statements and data cleaning, in order to ensure that we give them the best results out of our analysis.

Future Scope

There is so much more which can be explored. Like we have current version, android version available which can be explored in detail and we can come out with more analysis where we can tell how does these things effect and needs to be kept in mind while developing app for the users. We can explore the correlation between the size of the app and the version of Android on the number of installs. Machine learning can help us to deploy more insights by developing models which can help us interpret even more better. We have left this as future work as this is something where we can work on.

References

- Google Colab
- Matpolib
- Github
- Kaggle
- Play store real-time app revies and ratings

Certificate 1

In recognition of the commitment to achieve professional excellence

Getting Started with Enterprise Data Science

IBM

Sriyanka Dhal

Has successfully satisfied the requirements for:

Getting Started with Enterprise Data Science

Issued on: 29 FEB 2024

Issued by IBM

Verify: https://www.credly.com/go/PoDJl4ev

IBM

Certificate 2

Thank you!