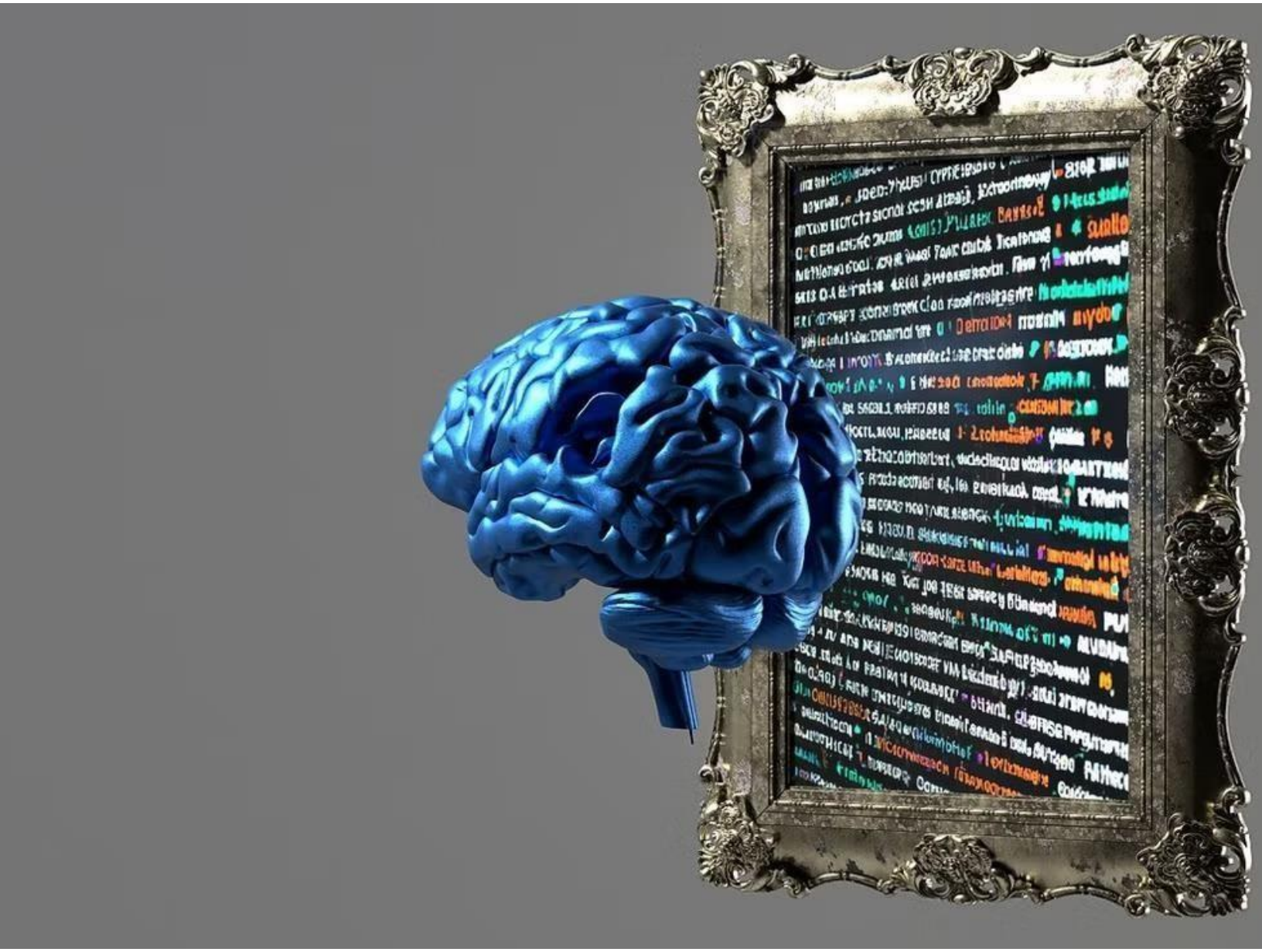
Data Science Projects



https://www.github.com/EntimemaInsight

I. Power BI Projects

1. Monthly Review: Credit Card Company Analysis using PBI: GitHub Link

- This project provides insights and recommendations for a credit card company's future strategy. It includes interactive dashboards, Excel data files, SQL query files, and visualizations to facilitate data exploration. The key insights cover revenue analysis, and collected vs. withdrawal analysis, enabling informed decision-making for the company's operations and growth.

2. Financial Performance and KPI Overview using PBI: GitHub Link

- This project focused on financial performance and key performance indicators (KPIs) for a credit card company. The dashboard provides a comprehensive overview of the company's financial performance, including revenue and expense analysis, tracking budget vs. actual figures, and key metrics such as net revenue, EBITDA, assets, and liabilities. The KPIs overview offers insights into various performance areas, enabling informed decision-making for strategic investments and financial management.

3. Credit Limit Increasing Simulation using PBI: GitHub Link

- This project conducts a credit limit increasing simulation using Power BI, leveraging customer data such as credit scores, payment history, and credit utilization to analyze creditworthiness. The simulation provided insights into factors impacting creditworthiness and allowed from projections of revenue growth and transactional amount based on multiple linear regression analysis.

II. R-automation Projects

1. KPI Reporting R-automation: GitHub Link

- This project automates KPI reporting using R scripts. It transposes SQL files into an Excel file with three sheets named KPI_Report_ES_03. Additionally, the project generates an additional file named ES_KPI_Daily_Summary, which summarizes the information from KPI_Report_ES_03 on a daily basis. The Daily KPI Summary file provides an overview of the performance metrics for the day and serves as a quick reference for stakeholders. The R scripts execute SQL queries, transpose the result data frames to Excel worksheets, and automate the process.

III. Python ML Projects

1. Automobile Sales Forecasting using Linear Regression: GitHub Link

- This project utilizes linear regression to forecast the sales prices of vehicles based on various features such as mileage, engine volume, brand, and body type. The dataset used in this analysis consists of information on used cars sold in the European market. The project involves data preprocessing steps to handle missing values and outliers, encoding categorical variables, and transforming variables to meet linear regression assumptions. The performance of the linear regression model is evaluated using metrics such as R-squared, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). The project also includes exploratory data analysis techniques to gain insights into the relationships between variables and to identify potential issues with the data. The project aims to provide accurate sales price predictions and valuable insights into the factors influencing vehicle sales prices.

2. **Credit Cards Forecasting using Multiple Linear Regression: GitHub Link**

- The project aims to analyze a dataset of credit card holders and develop a predictive model to estimate a cardholder's credit card balance based on their characteristics. The project provides insights into the factors influencing credit card balances and their significance for credit card companies.

3. **Naïve-Bayes-Classifier-for-YouTube-Comments: GitHub Link**

- This project implemented a Naïve Bayes Classifier for YouTube comments with the aim of classifying comments as „Ham “(non-spam) or „Spam “. The project involved data preparation, text classification, model evaluation, and the visualization of prediction probabilities. By utilizing the Multinomial Naïve Bayes Classifier, the project successfully achieved high accuracy in distinguishing between spam and non-spam comments. The project also included making predictions on new data and saving the results in an output file.