Project 3 - Data Engineering Track

**Credit Score Classification**

The purpose of this project is to analyse and classify credit scores using a dataset that reflects various factors influencing an individual’s credit score. By investigating key financial features such as the number of bank accounts, credit cards, occupation, age, and loan payment behaviour, we aim to gain insights into how these elements impact credit risk assessment. Through data cleaning, visualization, and analysis, we will develop a comprehensive understanding of the data, utilizing tools such as SQL, ELT processes, and Python libraries like PyTorch to automate the pipeline. The goal is to create an integrated solution that offers both a technical and business perspective on credit score classification, with results presented in an accessible HTML format

**Roles:**

**Kloood Azam** - Question no. 1: Using SQLite for storing and managing the data

1. How is the distribution of credit scores across different age groups?
2. How do the number of bank accounts and credit cards affect credit score?

**Jennifer Mancheno**- Question no.4 & Data Frame

* Clean the data and Analyse quarterly reports to create a report for critical change management

**Deniza Robinson** - Question no. 5 & HTML visualisation

* Pytorch Library (data pipeline) - the visualisations from the data pipeline will be used in the HTML

**Sukhwinder Singh** - Question no. 3 ETL

1. What is a credit score by occupation?
2. How the payment behaviour affect’s the credit score?

**Visualizations:**

A screenshot of a computer

Description automatically generated

A table with numbers and a few objects

Description automatically generated with medium confidence

Source

<https://www.kaggle.com/datasets/parisrohan/credit-score-classification/data>

GitHub Link

<https://github.com/Jennifer-JM/Project-3>