Project 3 - Data Engineering Track

**Credit Score Classification**

Some topics, ideas and questions that we’ll explore based on our credit score classification dataset that will help with Project 3 - Data Engineering Track, factors influencing credit scores like financial features, occupation, age, and loan payment behaviour and other risk assessments affecting credit scores. Data extractions, SQL, ELT, HTML, CSS and JavaScript will be used for the data analysis.

**Klood Azam** - Question no. 1: Using MongoDB for storing and managing the data

How is the distribution of credit scores across different age groups?

How do the number of bank accounts and credit cards affect credit score?

Storing the raw data in the database, and then take it out and clean and store it w/ a different name

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

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

(and considering libraries for data streaming, cloud, data pipelines, or data validation)

HTML - thursday

Write title, menu for subcategories per group member

Graph

Column 1 - q1 & q2

ocupations

Clear column

Credit score

Poor

Standard

Good

Column 2 - q3 &4

Occupations

**Jennifer Mancheno** - Question no.4 Pandas DataFrame, 6

*Clean the data*

Might divide the csv files by quarterly

**Sukhwinder Singh** - Question no. 3

*What is a credit score by occupation?*

How the payment behaviour affect’s the credit score?

*Added new columns to the dataframe - credit\_history\_age to credit\_history\_age\_months*

* *Number of each type of loan*

Source

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

GitHub Link

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

* each one of us should have a written portion as to what we want to present