# Credit Scoring and Segmentation using Machine Learning

## Project Report

Soumyadeep Poddar

## Introduction:

This project focuses on credit scoring and customer segmentation using Python. We have utilized a dataset containing various credit-related features, including credit scores, payment history, and employment status. The analysis includes data preprocessing, credit score calculation using the FICO method, and customer segmentation using K-Means clustering. The resulting visualization showcases distinct customer segments based on their credit scores, providing valuable insights for informed decision-making in the financial domain.

## Objective:

The objective of this project is to perform comprehensive credit scoring and customer segmentation. We aim to calculate credit scores using the FICO method, incorporating various credit-related factors. Subsequently, we utilize K-Means clustering to segment customers based on their credit scores. By achieving these objectives, we seek to gain valuable insights into customer creditworthiness and behaviour, enabling data-driven decisions for financial services and risk assessment.

## Methodology:

1) Data Collection & Loading:
    - ✓ The credit scoring dataset is taken from dataset.
    - ✓ The given dataset includes features such as age, gender, marital status, education level, employment status, credit utilization ratio, payment history, number of credit accounts, loan amount, interest rate, loan term, type of loan, and income level.
2) Data Exploration:
    - ✓ Examination of dataset columns.
    - ✓ Descriptive statistics summary of dataset columns.
3) Data Visualisation:
    - ✓ Boxplot to visualise the distribution of Credit Utilisation Ratio.
    - ✓ Histogram to visualize the distribution of Loan Amount.
4) Data Analysis:
    - ✓ Correlation analysis between the variables using a heatmap.
5) Credit Scoring:
    - ✓ Calculation of credit scores using the FICO method.
6) Customer Segmentation:
    - ✓ Segmentation of customers based on credit scores using K-Means clustering.

## Conclusion:

This project successfully calculated credit scores using the FICO method and applied K-Means clustering for customer segmentation. By combining these techniques, we gained valuable insights into creditworthiness and customer behaviour. These insights empower data-driven financial decisions, enhancing risk management and customer service strategies.

## Limitations:

One limitation of this project is that it relies on a simplified credit scoring model using the FICO method. While it provides valuable insights into creditworthiness, it may not capture all the complexities of real-world credit assessments. Additionally, the effectiveness of the segmentation depends on the choice of features and clustering algorithms. Further refinement and consideration of additional data sources may be necessary for a more accurate representation of credit risk and customer behaviour.