

# Data gathering and defining stakeholders + KPIs

## 1 Dataset Overview

- Dataset Name: Data from WellsFargo Bank.
- Data Composition: The training set consists of approximately 21,000 accounts and the evaluation set consists of approximately 5400 accounts.

The dataset includes:

- Binary variable, `bad_flag` (Loan performance in recent 12 months. =1 Never delinquent or 2 times or less 30 DPD. =0 One-time 60 DPD or worse, charge off, bankruptcy and repossession.)
- Predictor variables like the FICO scores, payment-to-income ratios, loan-to-values ratios, and other credit and loan features.
- Categorical features like gender and race to evaluate fairness.

## 2 Problem Description

This project focuses on developing a machine learning model to improve the decision-making process for auto loan approvals. The goal is to create a predictive model that assesses the creditworthiness of loan applicants based on financial features, helping lenders make more accurate and data-driven lending decisions. We will ensure the models do not unfairly discriminate against applicants based on gender or race.

## 3 Stakeholders of the Project

- Internal Stakeholders:
  - Product Managers: Responsible for designing the system's features and ensuring it meets business goals.
  - Data Analysts: Analyze application data, trends, and loan outcomes.
  - Customer Service Team: Handles customer inquiries and issues related to loan applications.
  - Executives: Oversee the overall performance of the loan application system, focusing on KPIs like profitability and growth.
- External Stakeholders:
  - Customers: Individuals applying for loans, whose satisfaction and feedback are crucial.
  - Regulators: Government bodies ensuring that the loan system adheres to financial regulations.
  - Credit Agencies: Provide credit scores that help in making decisions on loan approvals.
  - Investors/Shareholders: Interested in the profitability and growth of the loan system.

## 4 Company Key Performance Indicators (KPIs)

- Approval Rate: Percentage of loan applications approved out of the total received.
- Processing Time: Average time taken to process a loan application.
- Error Rate: The percentage of loan applications that require manual intervention due to data entry issues or system errors.
- Loan Default Rate: Percentage of loans that are not paid back according to the terms.