Data Science Capstone Project

**CREDIT-WORTHINESS OF SMALL BUSINESSES**

horizontal line



### Project Collaborators

Brian Njau  
Grace Nyongesa  
Josiah Okumu  
Leo Kariuki  
Lynne Mutwiri  
Stella Kitur

# Business Overview

The Small Business Administration (SBA) is a crucial organization in the United States that supports and promotes small enterprises by providing access to credit. Through loan guarantees, the SBA encourages lenders to credit small businesses, which play a vital role in job creation and reducing unemployment. However, there have been instances of loan defaults, posing a challenge for the SBA in accurately assessing the creditworthiness of applicants and mitigating risks.  
To address this, the SBA focuses on evaluating creditworthiness by considering factors such as *financial history, business plans, collateral, and projected cash flows*. They also strive to adapt to market dynamics, embracing innovative approaches and technology for informed lending decisions that minimize the likelihood of defaults. Overall, the SBA's commitment to supporting small businesses goes beyond loan guarantees, encompassing mentorship programs, training resources, government contracting opportunities, and disaster recovery assistance to foster an environment conducive to small business success.  
  
***Stakeholders***: Small Businesses, Government, and Lending institutions such as the SBA

## Business Challenges

* Small businesses, particularly those in their early stages often face challenges in accessing credit due to limited financial resources and a lack of collateral. When starting a new business, entrepreneurs typically have minimal or no track record of financial performance, making it difficult for traditional lenders to assess their creditworthiness.
* Additionally, most banks want to see a business's growth projections before financing them. However, small businesses often grow slowly, which makes it hard for banks to service their loans.
* Understanding the creditworthiness of small businesses is crucial for lenders, policymakers, and investors. Assessing the viability of small businesses ensures that lenders can mitigate risks and make informed decisions when extending credit as well.

## 

## 

## Project Overview

This project aims to develop a predictive model for SBA-guaranteed loan applications. It will analyze financial indicators and metrics to accurately assess the likelihood of loan approval, helping small business owners make informed decisions. Leveraging machine learning and statistical techniques, the model will streamline approvals and identify key financial factors. Ultimately, the project addresses credit access challenges for small businesses, providing a data-driven tool for assessing loan approval probabilities. By streamlining the process and fostering informed decision-making, it supports small business growth while mitigating risks for lenders and policymakers.

## Business Problem Statement

By analyzing the financial indicators and metrics that influence loan approval decisions, our goal is to develop a reliable predictive model that accurately assesses whether an SBA-guaranteed loan application will be approved or rejected. This project aims to address the challenge faced by small business owners in understanding their chances of securing funding and provide them with a tool to make informed decisions, while also streamlining the loan approval process for the SBA and identifying key financial factors that impact loan approvals.

## Project Objectives

Building a loan classifier that will help businesses (and their lenders) improve their creditworthiness, translating to a higher financial muscle for business sustainability.

1. Conduct a comprehensive analysis of the dataset from the U.S. Small Business Administration to identify patterns and trends for accurate loan approval predictions.
2. Develop a robust machine learning model that utilizes the identified patterns and trends to predict loan approval outcomes effectively.
3. Deploy and evaluate the developed machine learning model in the loan approval process of the U.S. Small Business Administration, continuously optimizing its predictive capabilities for informed lending decisions.

## 

## Success Metrics

1. Accuracy: 85% accuracy which measures the overall correctness of the model's predictions by classifying loan applications as Approved or Rejected.
2. Precision: Precision is the proportion of true positive predictions (correctly classified loans) out of all positive predictions made by the model. It is useful when the cost of misclassifying a good loan as bad (false positives) is high.

## Deployment

Using locally available data / or at least using data from other African countries…  
The following columns will be helpful in the deployment process

LoanNr\_ChkDgt: This column can be used as a unique identifier for each loan.

LowDoc (Low Documentation): Low Doc provides flexible financing solutions for self-employed people. Low Doc loans are designed for customers who have an income and assets, but are unable to provide the usual verification documentation like financial statements and in some cases, tax returns.

NAICS: This column contains information about the industry or sector that the small business operates in, which can be relevant for predicting loan defaults.

ApprovalDate: This column can be used to analyze trends in loan approvals over time.

Term: This column contains information about the length of the loan term, which can be relevant for predicting loan defaults.

NoEmp: This column contains information about the number of employees that the small business has, which can be relevant for predicting loan defaults.

FranchiseCode: This column contains information about whether the small business is a franchise, which can be relevant for predicting loan defaults.

UrbanRural: This column contains information about whether the small business is located in an urban or rural area, which can be relevant for predicting loan defaults.

RevLineCr: This column contains information about whether the small business has a revolving line of credit, which can be relevant for predicting loan defaults.