Problem Statement

One of the most lucrative - and riskiest - banking activities is lending. This truth is so evident that most private citizens trust their future financial well-being to this activity in the form of investments. Banks, however, rely on lending not only as a long-term background activity for success, but as a daily reality of business. Therefore, one of the greatest dangers to any given bank is a default on a loan, and one of the most important missions to those same banks should be better identification of risk in lending.

The aim of this project is exactly that - identifying any and all factors that have a significant correlation to defaulting on loans, ranking them, and putting together a realistic roadmap on which loans to avoid, and how to mitigate risk on ambiguous loans that are worth the chance. These insights should be put into practice immediately, as each loan underwritten on shaky risk analysis is a potential money sink for any given bank, and we should be able to accurately predict loans which have defaulted through risk factors they share.

This project will lean on the following datasets, with the potential to add more if it is useful without increasing the scope of the project:
Loan Default Dataset - Kaggle
https://www.kaggle.com/datasets/yasserh/loan-default-dataset
Dataset including various details including loan amounts, pre-approval, and creditworthiness of the applicant

Loan Default Prediction - Kaggle
https://www.kaggle.com/datasets/kmldas/loan-default-prediction
A simplified dataset to help with our initial correlative hypotheses. Information includes whether the applicant defaulted, balance of loan, applicant's employment status and their salary

Loan Default Prediction Dataset - Kaggle
https://www.kaggle.com/datasets/nikhil1e9/loan-default
A more robust dataset including salaries, creditworthiness, age, and default among other details

These three datasets together will enable us to explore many different relationships among characteristics of loan applicants, and should give us a stronger understanding of how to look at the full picture of an applicant to give us better predictive power on the risk of a loan. The goal is not perfect prediction, but a combination of features which gives us enough confidence to approve or deny a loan application.