**Capstone 3:**

Project Proposal

Data: [Credit Risk Analysis | Kaggle](https://www.kaggle.com/datasets/ranadeep/credit-risk-dataset?select=loan)

Credit default risk analysis plays a vital role in enabling creditors to assess the likelihood of loan repayment. Its significance lies in providing financial institutions with the necessary tools to distinguish between customers who are likely to pay back their loans and those who may default. This project aims to construct a predictive model capable of effectively discerning such customers, utilizing data obtained from an undisclosed financial institution. The dataset comprises relevant information about the customers that I will utilize to gauge their probability of defaulting on loans.

The primary stakeholders in this project are the credit firm that provided the dataset and other credit firms operating in similar positions, equipped with comparable datasets. By developing the most simplistic yet successful model based on this dataset, the intention is to extend the model's applicability to other firms confronting similar scenarios. This would allow them to utilize the model as a valuable tool in identifying customers who are more likely to default on their loans.

However, the dataset's most significant limitation lies in the presence of missing values. Numerous columns within the dataset have a substantial proportion of missing values. In cases where these missing values are integral to the model, I will have to impute values that may influence the model in unknown ways.

The goal for this project is to produce classifiers that are effective enough for the credit company to turn a profit. The first classifier is a binary classifier that can tell if the loan was good or bad for the credit firm. The second classifier will take a more in-depth approach to predict the exact outcome of the loan.