CAPSTONE PROJECT PROPOSAL: PREDICTING PERSONAL LOAN ACCEPTANCE FOR THERA BANK

**Background:**  
A substantial portion of Thera Bank's customer base, which is made up mostly of liability customers, is growing. By converting liability clients to personal loan customers while keeping them as deposits, the bank hopes to grow its asset customer base. The management is looking on ways to spot potential customers who are more likely to take out a personal loan. They seek to create a model that will aid in the identification of high-potential clients, boost campaign success rates, and do it at a lower cost.

**Objective:**  
This capstone project's main goal is to create a classification model that can determine if a Thera Bank liabilities customer is likely to accept a personal loan offer or not. By creating such a model, Thera Bank may raise their conversion rate and lower campaign expenditures while better targeting their marketing campaigns to high-potential clients.

**Data:**  
The 5000 observations in the data collection for this research each have fourteen variables broken down into four separate measurement categories. Five variables make up the binary category, including the personal loan goal variable as well as securities accounts, CD accounts, online banking, and credit cards. Age, experience, income, CC average, and mortgage are the five criteria that make up the interval category. The variables family and education are part of the ordinal category. The final group has a nominal ID and ZIP code. Since the variable ID doesn't contribute any noteworthy information, it won't be used in the analysis.

Methodology:  
Data cleaning, exploratory data analysis, feature engineering, model selection, and model evaluation are all steps in the project's conventional data science methodology. The data set will first be cleaned, with missing values, outliers, and unimportant variables being eliminated. To comprehend how the variables in the data set are distributed, exploratory data analysis will be carried out. Based on domain expertise and feature selection methods, feature engineering will be used to develop new variables. We'll construct and assess a variety of classification models, including logistic regression, decision trees, and random forests using the right performance metrics like recall, accuracy, and F1 score. Based on how well the model performed on the test set, it will be chosen as the best one.

**Expected Outcomes:**  
The expected outcome of this project is a classification model that can predict the likelihood of a liability customer of Thera Bank accepting a personal loan offer. The model will be used to identify high-potential customers and increase the success ratio of the marketing campaign while reducing the campaign costs.

Conclusion:  
Thera Bank can focus their marketing campaigns to high-potential clients and boost their conversion rate while lowering campaign expenses by creating a categorization model to forecast the chance of a particular customer accepting a personal loan offer. Thera Bank will benefit from this project's insightful data on customer behavior and be better able to make data-driven decisions to enhance business operations.