**What is Personal Loan?**

A Personal Loan is an unsecured loan that may be short term or long term, given by a bank or non-banking finance company (NBFC) to meet their customers’ personal expenses. An individual can obtain a personal loan based on his income, credit score, repaying capacity, etc.

**Data Description:**

The .csv file contains data on 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (Personal Loan). Among these 5000 customers, only 480 (= 9.6%) accepted the personal loan that was offered to them in the earlier campaign.

**Content:**

This case is about a bank (Thera Bank) whose management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors). A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio with a minimal budget.

**This notebook includes the following:**

* Data cleaning
* Exploratory data analysis (EDA)
* Preparing the data to train a model
* Training and making predictions using various classification models
* Model evaluation
* Using various ensemble learning methods

**Objective:**

The classification goal is to predict the likelihood of a liability customer buying personal loans, which means we have to build a model which will be used to predict which customer will most likely accept the offer for a personal loan based on the specific relationship with the bank across various features given in the dataset. Here I will be using the Supervised Learning methods to predict which model is best for this problem amongst **Logistic Regression**, **Naive Bayes (NB)**, and **K-Nearest Neighbors (KNN)**.

**What Problem We Have and Which Metric to Use?**

* Based on the data and attribute Information, We have a classification problem.
* We will make a classification on the target variable **Personal Loan**
* We will build a model to get the best classification possible on the target variable.
* For that, we will look at the balance of the target variable.
* As we will see later, our target variable has imbalanced data
* For that reason, we are not going to use the Accuracy score,
* Based on the problem on hand, we will use the **Recall** score.

**Attribute Information:**

1. ID : Customer ID
2. Age : Customer's age in completed years
3. Experience : #years of professional experience
4. Income : Annual income of the customer ($000)
5. ZIP Code : Home Address ZIP code.
6. Family : Family size of the customer
7. CCAvg : Avg. spending on credit cards per month ($000)
8. Education : Education Level.  
   1: Undergrad;  
   2: Graduate;  
   3: Advanced/Professional
9. Mortgage : Value of house mortgage if any. ($000)
10. Personal Loan : Did this customer accept the personal loan offered in the last campaign?
11. Securities Account : Does the customer have a securities account with the bank?
12. CD Account : Does the customer have a certificate of deposit (CD) account with the bank?
13. Online : Does the customer use internet banking facilities?
14. Credit card : Does the customer use a credit card issued by