Problem Statement:

## **Credit Card Lead Prediction**

Happy Customer Bank is a mid-sized private bank that deals in all kinds of banking products, like Savings accounts, Current accounts, investment products, credit products, among other offerings.

The bank also cross-sells products to its existing customers and to do so they use different kinds of communication like tele-calling, e-mails, recommendations on net banking, mobile banking, etc.

In this case, the Happy Customer Bank wants to cross sell its credit cards to its existing customers. The bank has identified a set of customers that are eligible for taking these credit cards.

Now, the bank is looking for your help in identifying customers that could show higher intent towards a recommended credit card, given:

* Customer details (gender, age, region etc.)
* Details of his/her relationship with the bank (Channel\_Code,Vintage, 'Avg\_Asset\_Value etc.)

**Step by step problem solving explanation as follows**

**Step 1: Importing and Merging Data**

* Importing Pandas and NumPy
* Importing Train data set

### Step 2: Inspecting the Dataframe

* **let's look at the statistical aspects of the dataframe**
* **Let's check the dimensions of the dataframe**
* **Let's see the type of each column**

### Step 3: Data Preparation

#### Converting some binary variables (Yes/No) to 0/1

#### For categorical variables with multiple levels, create dummy features (one-hot encoded)

#### Dropping the repeated variables

#### Checking for Outliers

#### Checking for Missing Values and Inputing Them

### Step 4: Test-Train Split

* Putting feature variable to X
* Putting response variable to y
* Splitting the data into train and test

### Step 5: Feature Scaling

### Step 6: Looking at Correlations

### check the correlation matrix

#### Dropping highly correlated dummy variables

### Step 7: Model Building

* Let's start by splitting our data into a training set and a test set.

#### Running our First Model: Logistic regression model

* Summary:
  + Z - Gives you significance of the variable
  + p is hypothesis testing of significance of Z
  + H0: varibale is not significant
  + H0: var =0
  + H1: var !=0
  + P high implies that the variable is not significant
  + so
  + Lower the p value more the significant of the variable
  + Higher the P value less significant the variable

### Step 8: Feature Selection Using RFE

##### *Assessing the model with StatsModels*

* Getting the predicted values

##### *Creating a dataframe with the actual and the predicted probabilities*

* Confusion matrix
* Overall accuracy

## Metrics beyond simply accuracy

* see the sensitivity of our logistic regression model
* calculate specificity
* precision and recall

### Step 9: Plotting the ROC Curve

### Step 10: Finding Optimal Cutoff Point