

Data Description:

The dataset 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.

Domain:

Banking

Context:

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**.

Attribute Information:

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

Steps and tasks:

1. Import the datasets and libraries, check datatype, statistical summary, shape, null values or incorrect imputation.
2. EDA: Study the data distribution in each attribute and target variable, share your findings
 - Number of unique in each column?
 - Number of people with zero mortgage?
 - Number of people with zero credit card spending per month?
 - Value counts of all categorical columns.
 - Univariate and Bivariate
 - Get data model ready
3. Split the data into training and test set in the ratio of 70:30 respectively
4. Use the Logistic Regression model to predict whether the customer will take a personal loan or not. Print all the metrics related to evaluating the model performance (accuracy, recall, precision, f1score, and roc_auc_score). Draw a heatmap to display confusion matrix
5. Find out coefficients of all the attributes and show the output in a data frame with column names? For test data show all the rows where the predicted class is not equal to the observed class.
6. Give conclusion related to the Business understanding of your model?