

## Project 1 The Credit Scoring Model

The consumer credit department of a bank wants to automate the decision making process for approval of home equity lines of credit. To do this, they will follow the recommendations of the Equity Credit Opportunity Act to create an empirically derived and statistically sound credit scoring model. The model will be based on data collected from recent applicants granted credit through the current process of loan underwriting. The model will be built from predictive modeling tools, but the created model must be sufficiently interpretable so as to provide a reason for any adverse actions (rejections).

The **HMEQ** data set contains baseline and loan performance information for 5,960 recent home equity loans. The target (**BAD**) is a binary variable indicating whether an applicant eventually defaulted or was seriously delinquent. This adverse outcome occurred in 1,189 cases (20%). For each applicant, 12 input variables were recorded.

<i>Name</i>	<i>Model Role</i>	<i>Measurement Level</i>	<i>Description</i>
<b>BAD REASON</b>	Target Input	Binary Binary	1=default on loan, 0=paid back loan HomeImp=home improvement DebtCon=debt consolidation
<b>JOB</b>	Input	Nominal	Six occupational categories
<b>LOAN</b>	Input	Interval	Amount of loan request
<b>MORTDUE</b>	Input	Interval	Amount due on existing mortgage
<b>VALUE</b>	Input	Interval	Value of current property
<b>DEBTINC</b>	Input	Interval	Debt to income ratio
<b>YOJ</b>	Input	Interval	Years at present job
<b>DEROG</b>	Input	Interval	Number of major derogatory reports
<b>CLNO</b>	Input	Interval	Number of trade lines
<b>DELINQ</b>	Input	Interval	Number of delinquent trade lines
<b>CLAGE</b>	Input	Interval	Age of oldest trade line in months
<b>NINQ</b>	Input	Interval	Number of recent credit inquiries

The credit scoring model will give a probability of a given loan applicant defaulting on loan repayment. A threshold will be selected such that all applicants whose probability of default is in excess of the threshold will be recommended for rejection.

For this project, you are required to perform the following tasks:

1. Assemble a diagram that includes Input Data Source, Data Partition, Insight, Transform Variables, Replacement, Regression, and Assessment nodes.
2. Partition the data in such a way that 67% are for the training and 33% for the validation.

3. Explore the input variables using Insight nodes and make some comments about the input variables.
4. Perform some variable transformations that you think it will help. You are recommended to do the followings:
  - Perform the log transformation on YOJ.
  - Convert DEROG from interval variable to nominal variable with three levels: 0, 0-2.5, and >2.5.
  - Convert DELINQ from interval variable to nominal variable with three levels: 0, 0-3.5, and >3.5.
  - Convert NINQ from interval variable to nominal variable with three levels: 0-1, 2-3, and >3.
5. Impute the missing values based on the entire data set.
  - For interval variables, select **tree imputation** as the imputation method.
  - For nominal variable, replace the missing value by a default constant. (you need to define this constant)
6. Fit the regression model using Stepwise Regression procedure. Report the fitted model, i.e. write out the fitted model.
7. Based on the information provided by Assessment nodes, answer the following questions:
  - How good is the predicted probability model compared to the baseline rate model (20%) if the top 20% (in highest probability of default) is chosen as the threshold to reject the loan application?
  - By rejecting the worst 20% of all applications using the fitted model, how many bad loans have been eliminated from the loan portfolio?

Note: For this analysis, you do not need to set up a target profile. The data were not oversampled so there was no need to specify prior information. Your study will focus on accuracy so the default target profile is appropriate.

In your project, you should clearly present the steps of performing the analysis by including the SAS-enterprise miner procedures