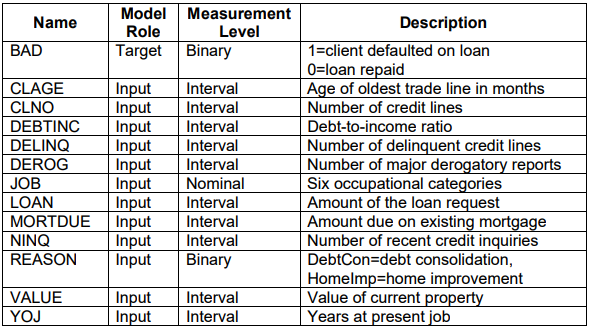
**Avail Finance - Internship Assignment**

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 Equal 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 to provide a reason for any adverse actions (rejections).

The Home Equity dataset (HMEQ) 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.



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 more than the threshold will be recommended for rejection.

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

1. **Data Partition:** Partition the data in such a way that 67% are for the training and 33% for the validation.
2. **Impute:** Impute the missing values based on the entire dataset. Explore the input variables and mention some insights about the input variables.
3. **Transform Variables:** Perform some variable transformations that you think will help.
4. **Variable Selection:** Filter out variables to select the optimum number of variables for modeling. Focus on avoiding multicollinearity and choosing variables that are highly informative regarding the dependant variable
5. **Modeling:** Apply any relevant modeling method/s and report the model parameters with the results of the modeling process. Give the reasoning on why a particular model was chosen by you
6. **Comparison:** Assess the model/s using the validation set and report whatever insights you can.
7. **Statistical tests:** Give a brief on possible tests we should perform to check the stability and significance of the variables used in the model.

You are free to choose your own methods and processes. You should clearly document all the steps and assumptions made during the whole workflow. Submit all the codes and the results along with the document in a zipped folder