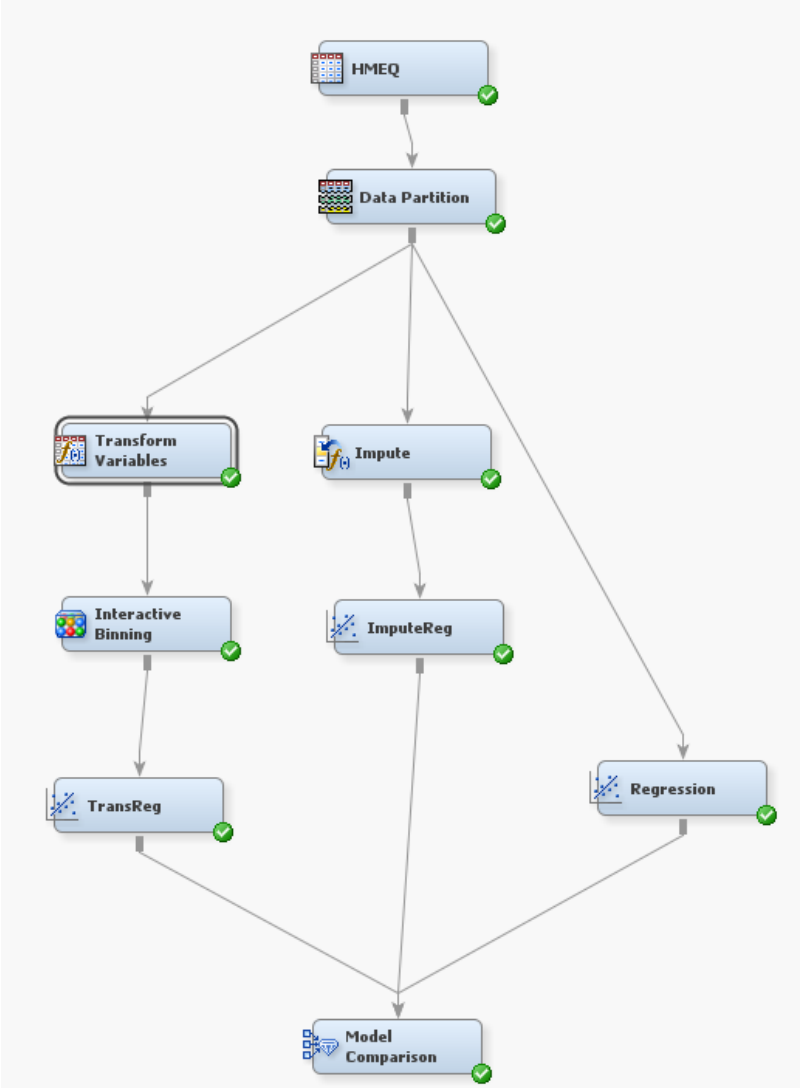
Diane Nguyen
Class Assignment 3
MSBA 645
4/2/19
Part 1

In this part of the assignment you will expand the model created in Class Exercise 1 by:

- **The final flow diagram as given on Page 58 of the Case Study book.**



- **Follow the instructions in the book and add Impute, Transform Variables, and Interactive Binning nodes. Provide a brief description of what you have done for each node and why.**

    - We partitioned the HMEQ dataset in the following was: Training: 40%, Validation: 30%, and Test: 30%. Then we realized some data was missing, in order to deal with the missing data we had to use the impute node in order to increase the performance of the nodes. We also added a Transform Variables Node to deal with the Skewness that existed, below this node wee added the interactive binning node to categorize variables into bins in order to deal with some of the non-linear relationships exist. Finally we added 3 regression nodes to compare the performances of the impute node and the interactive binning nodes and we added the model comparison node at the very bottom to see which of the 3 regressions performed the best. The Regression that Transformed Variables and used Interactive Binning performed the best.

A summary of the final comparative results:

- **Misclassification rates for the four models in ascending order of misclassification rates.**

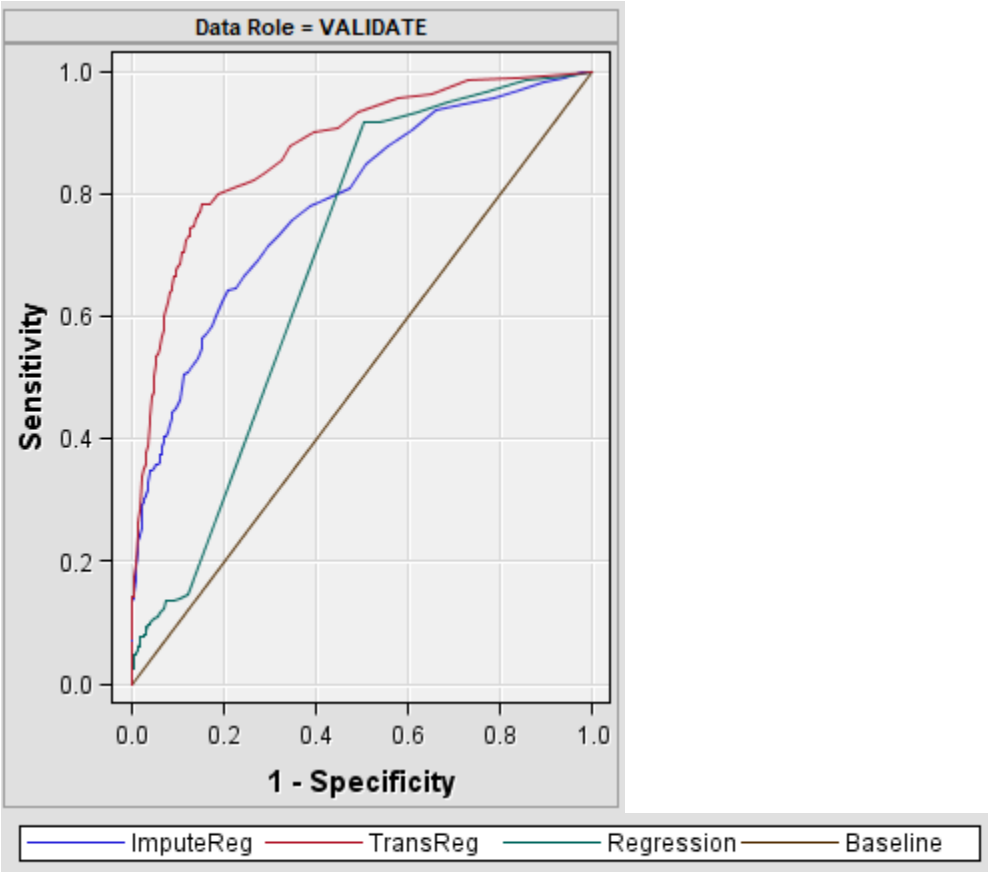
Selected Model	Predecessor	Model	Model Description	Target Variable	Valid: Misclassification Rate	Train: Misclassification Rate
Y	Reg2	Reg2	TransReg	BAD	0.14038	0.130563
	Reg3	Reg3	ImputeReg	BAD	0.16443	0.157011
	Reg	Reg	Regression	BAD	0.196868	0.192695

Selected Model

Model Selection based on Test: Misclassification Rate (_TMISC_)

Selected Model	Model Node	Model Description	Test: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
Y	Reg2	TransReg	0.13575	0.09838	0.13056	0.10101	0.14038
	Reg3	ImputeReg	0.16760	0.11745	0.15701	0.12552	0.16443
	Reg	Regression	0.19497	0.16068	0.19270	0.16643	0.19687

- **The ROC chart for the Validation models**



- **Your decision on which model to use given the misclassification rates and the ROC chart. Justify your decision.**

Based on the misclassification rates, the ROC charts I would choose the "TransReg" which is the Regression that involved transforming variables and interactive binning. The "TransReg" has the lowest misclassification rate as compare to the other regressions. In the ROC chart the "TransReg" is in the most upper left-hand corner of the chart as compared to the rest which means that it has the best performance in predicting the target "BAD".