**Woe Node - these are some settings in the WOE node that scoring team uses**

tree growth settings: 3% percentage of training data, minimum number of good/bad – 30. This will help the WOE know when to stop binning (probably makes it run faster)

split parameters: “include info value” – this will allow us to see the information value of the variable

generate woe transformation report (record type excel) “ – this will output an excel sheet with the variable binning (its good to save this so we can easily reference the binning of the variables in case we lose them)

**logistic regression node**

To build the model, select all the WOE variables with the following settings:

Stepwise, entry 0.01, exit 0.01, round trip 3 (the last 3 can be changed as we see fit. But that’s the general settings that the scoring team will use)

After the stepwise runs, go to “currently selected sequence” and review the model. This will be the model that satisfies the conditions above using the stepwise function. You want to review the model and look at the following:

c statistic is AUC (youll want to note this and track is as you made changes to the model)

look for negative coefficient in model parameter . If there are negative coefficient this variable should not be used (the model trend flipped). Remove those variables and re-run a model. Again, go to “currently selected sequence” and review the model. There should be no negative coefficients (if there are, remove and repeat – remember to keep track of AUC in excel). Then you want to review:

variance inflation factors - value should be 2/3 or less. This is a value that determines if you have multicollinearity in your model. If you noticed variables with VIF of 4/5, look at which has a higher IV (scientific approach) or better trend (art approach) to determine which to keep. Remove other variables and re-run the model (remember to keep track of AUC in excel).

You can finally be satisfied with the model if the AUC is still decent, there are no negative model parameters and the correlation between variables is minimal (may want to also look at correlation between variables but tracking VIF should take care of this).