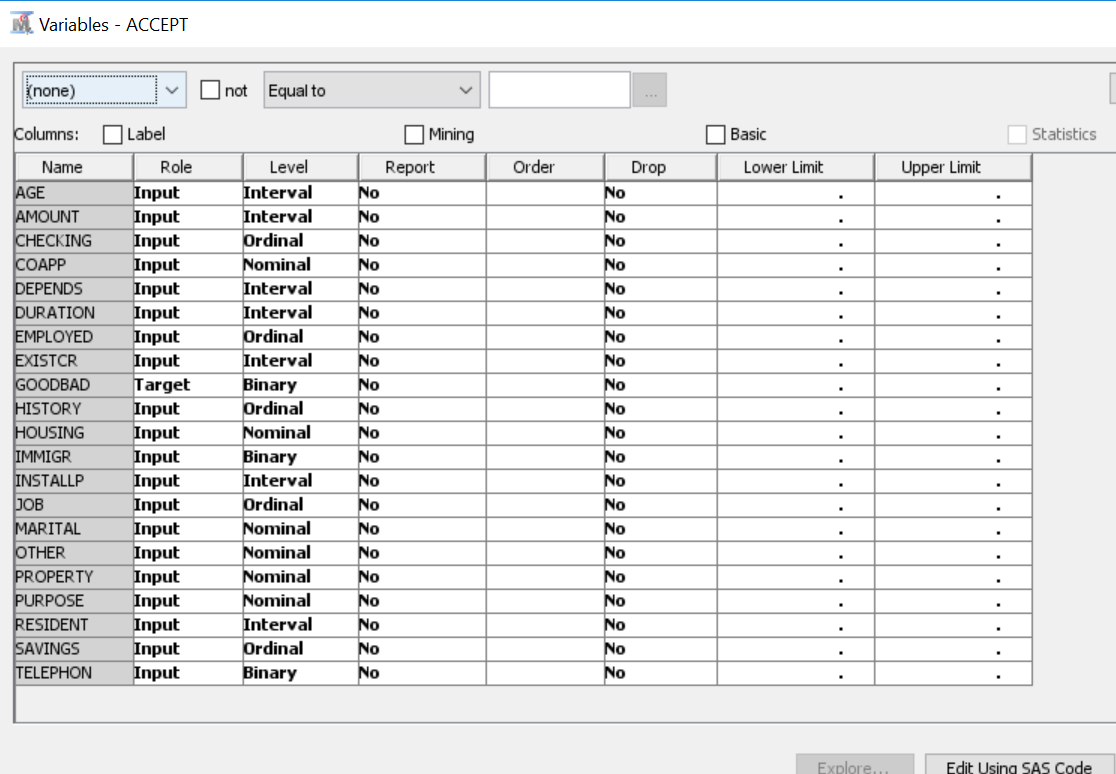
CREDIT SCORING INDIVIDUAL ASSIGNMENT

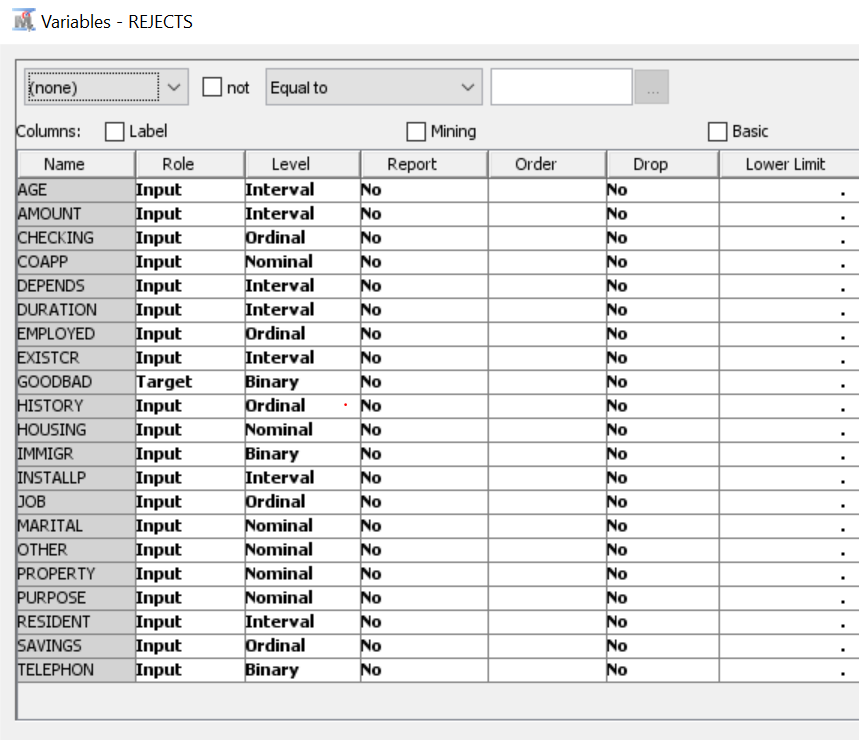
INTRODUCTION

Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit. These techniques describe who should get credit, how much credit they should receive, and which operational strategies will enhance the profitability of the borrowers to the lenders

Data import

Since the data was without headers, we manually define the headers and import it. We then change the goodbad as target as that’s what we are trying to predict that. Also some of the variables level’s were changed





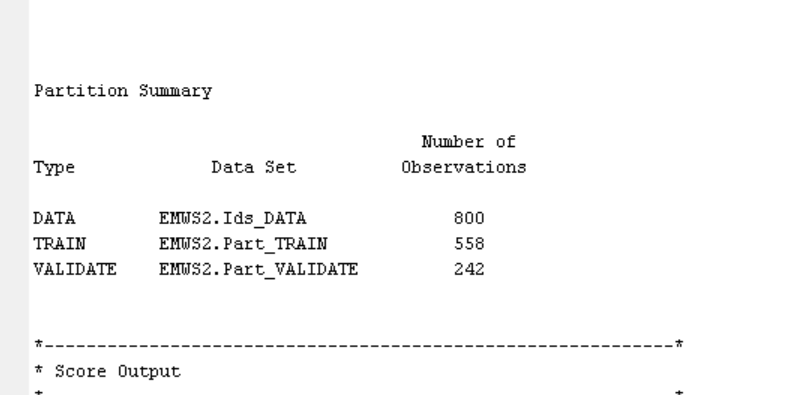
Process Flow

INPUT DATA NODE

We create a new diagram, then drag ACCEPTS file which was imported before. This action creates input data node.

Partition Data

We partition the data into train, validation, test in 70, 30, 0



Explore data and create Score Card

Here we will create a score card by performing the following tasks:

1) Group the characteristic variable into attribute

2) Logistic regression model to create initial scorecard

3) create a reject interface on logistic regression model

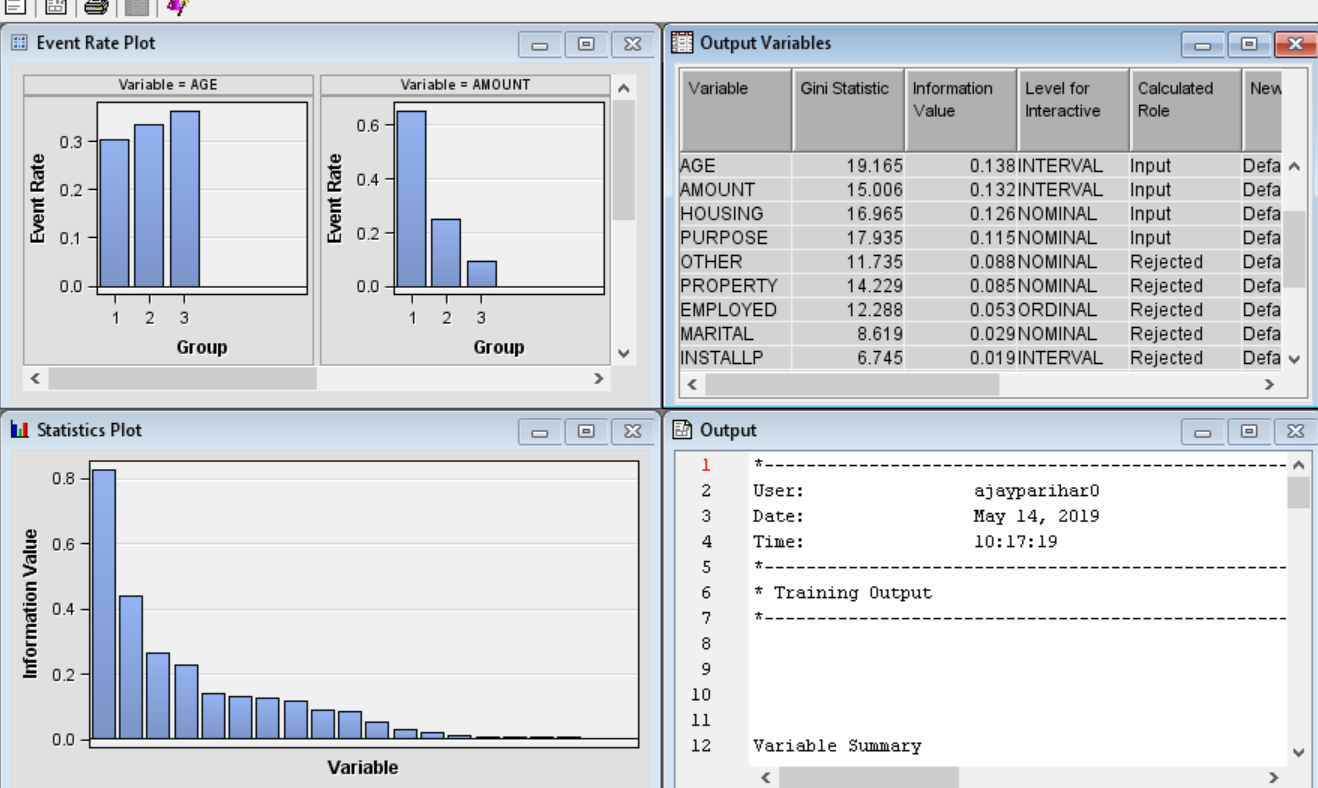
4) Create final scorecard

Grouping the characteristic variables into attributes

1. From the Credit Scoring tab, drag an Interactive Grouping node to the Diagram Workspace. Connect the Data Partition node to the Interactive Grouping node. The Interactive Grouping node performs the initial grouping automatically. You can use these initial groupings as a starting point to modify the classes interactively. By default, the unbinned interval variables are grouped into 20 quantiles (also called 9 bins), which are then grouped based on a decision tree. The Interactive Grouping node enables you to specify the properties of this decision tree.
2. Select the Interactive Grouping node in the Diagram Workspace. Set the value of the Interval Grouping Method property and the Ordinal Grouping Method property to Monotonic Event Rate. Set the value of the Maximum Number of Groups property to 10.
3. Right-click the Interactive Grouping node and click Run. In the Confirmation window that appears, click Yes. In the Run Status window that appears, click Results.

The Output Variables window displays each variable’s Gini Statistic and information value (IV). Note that a variable receives an Exported Role of Rejected if the variable’s IV is less than 0.10. Recall that IV is used to evaluate a characteristic’s overall predictive power (that is, the characteristic’s ability to differentiate between good and bad loans).

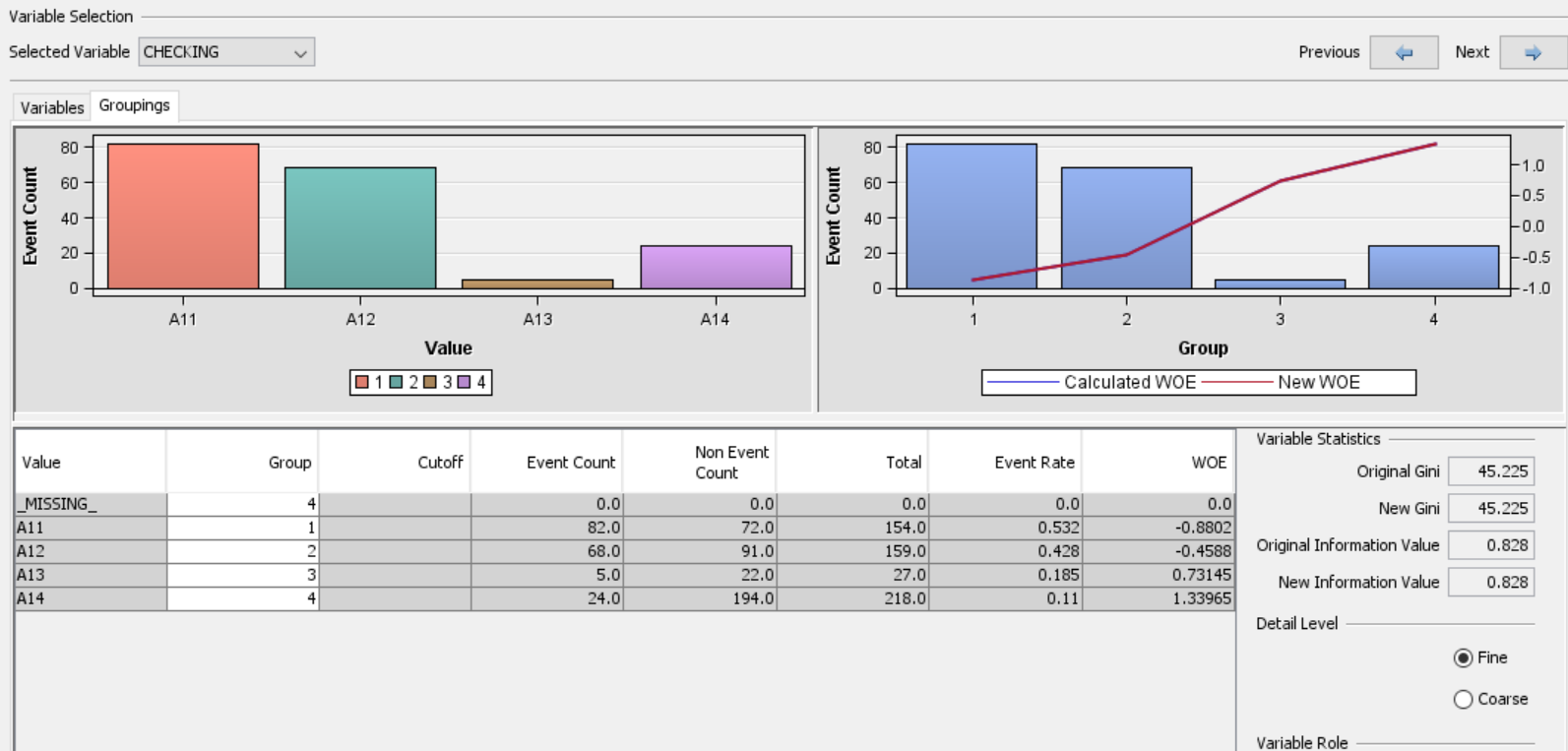
Here L is the number of attributes for the characteristic variable. In general an IV less than 0.02 is unpredictive, a value between 0.02 and 0.10 is weakly predictive, a value between 0.10 and 0.30 is moderately predictive, and a value greater than 0.30 is strongly predictive. The Gini statistic is used as an alternative to the IV.



Based on IV, we see following candidates for the final input variables

*CHECKING, HISTORY, SAVING, DURATION, AGE, AMOUNT, HOUSING, PURPOSE*

By default, the variables are sorted by their information value, given in the Original Information Value column. Also, the variable that is selected by default is the variable with the greatest IV. In this example, that variable is CHECKING.

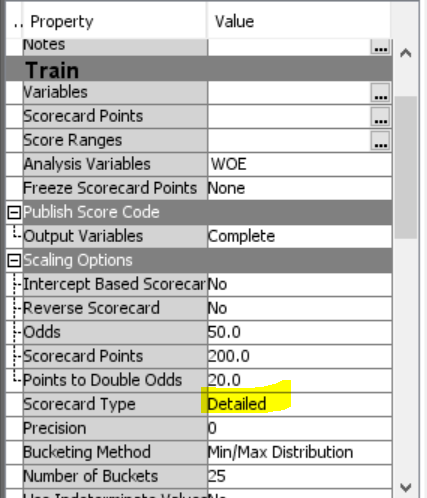


We tried splitting bins, it didn’t help much.

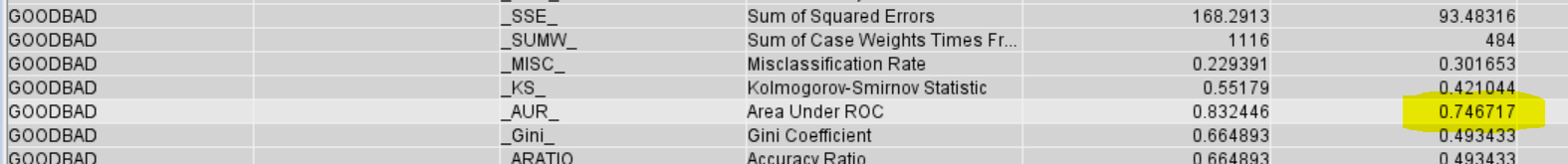
CREATING SCORECARD WITH LOGISTIC REGRESSION

From the Credit Scoring tab, drag a Scorecard node to the Diagram Workspace. Connect the Interactive Grouping node to the Scorecard node.

Also we change the scorecard type: detailed

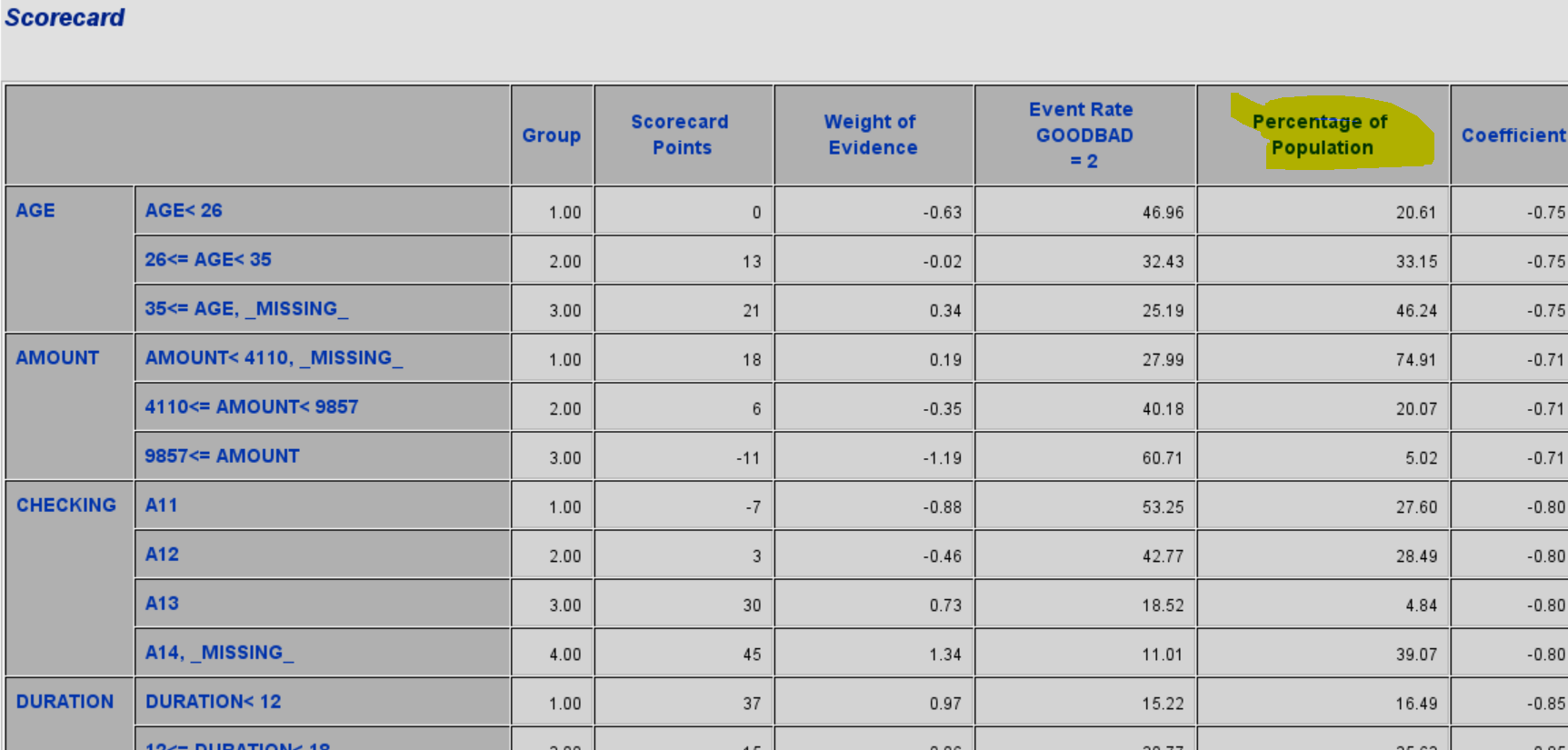


Run the scorecard and maximize the fit statistic window, we see area under ROC curve as .74

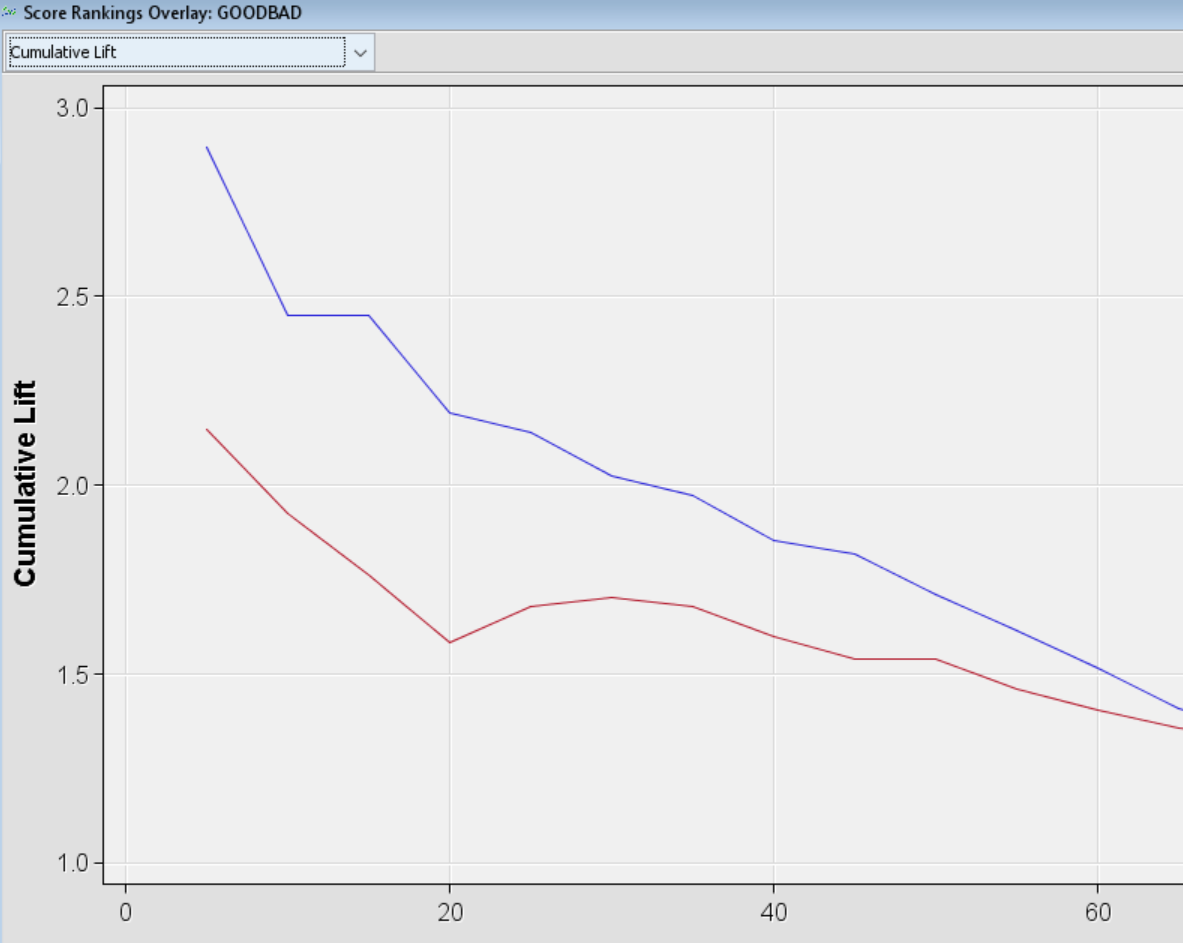


Scorecard

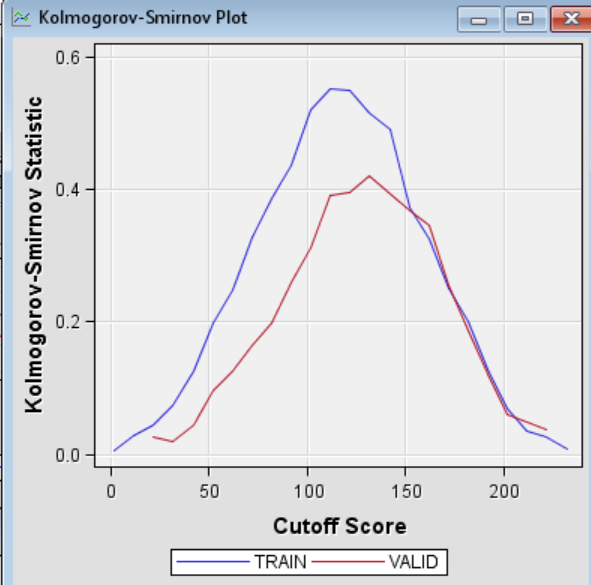
The percentage of population suggests total of bad applicants who have a score higher than the lower limit of the score range.



For lift and cumulative lift, the higher value in the lower deciles indicates a predictive scorecard model. Notice that both Lift and Cumilative have higher lift curve.



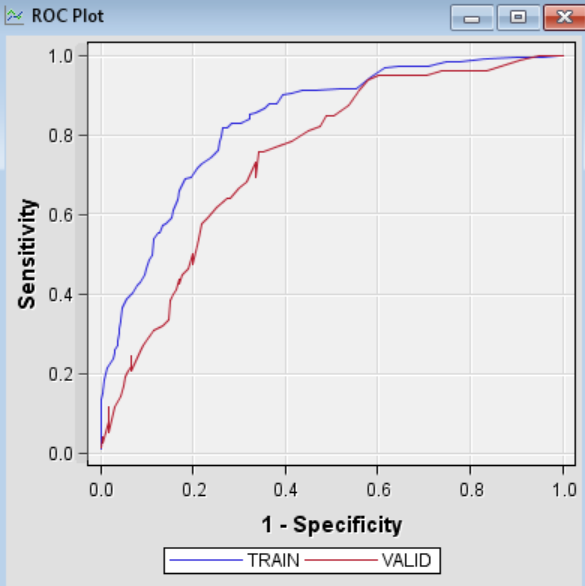
According to Kolmogorov plot the cutoff score is around 120



ROC curve

We now check the AUR curve, the value we got was .74

The AUR is generally regarded as providing a much better measure of the scorecard strength than the Kolmogorov-Smirnov statistic because the area being calculated encompasses all cutoff values. A scorecard that is no better than random selection has an AUR value equal to 0.50. The maximum value of the AUR is 1.0.



PERFORM REJECT INTERFACE ON MODEL

The preliminary scorecard that was built in the previous section used known good and bad loans from only the accepted applicants. The scorecard modeler needs to apply the scorecard to all applicants, both accepted and rejected. The scorecard needs to generalize the “through the door” population.

The following inference methods are supported in SAS Enterprise Miner:

1. Fuzzy
2. Hard cutoff
3. Parceling