Figure 1: Process Flow Diagram
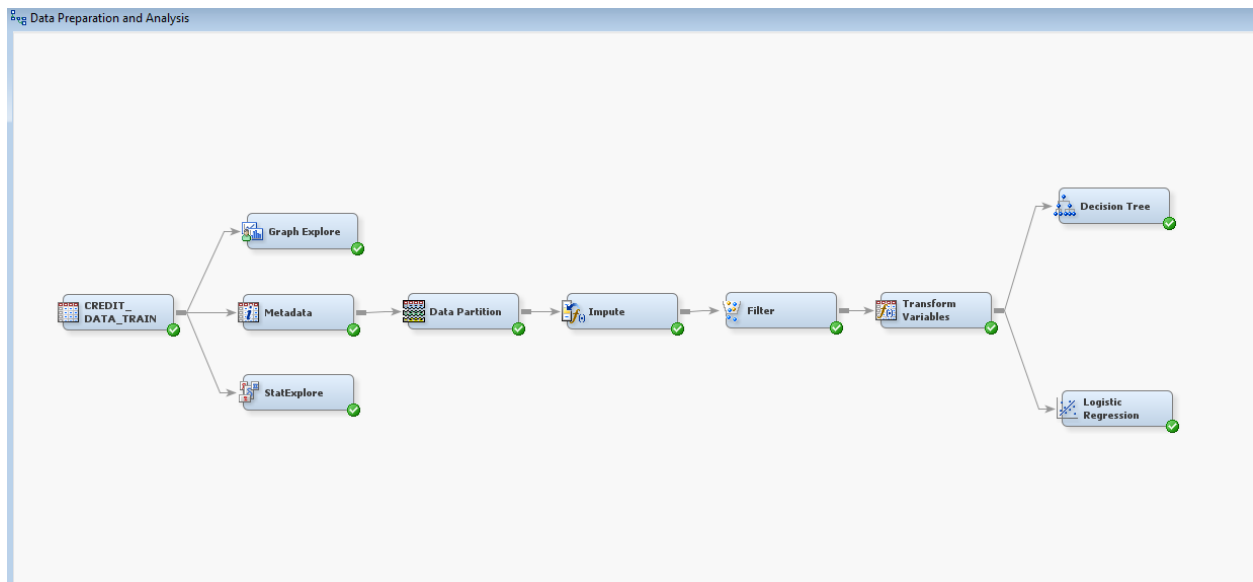
Figure 2: Data Partition Results

```
24
25    Partition Summary
26
27                                    Number of
28    Type             Data Set       Observations
29
30    DATA        EMWS2.Ids_DATA          150000
31    TRAIN       EMWS2.Part_TRAIN         89998
32    VALIDATE    EMWS2.Part_VALIDATE      60002
33
34
35    *-----------------------------------------------------------*
36    * Score Output
37    *-----------------------------------------------------------*
38
39
40    *-----------------------------------------------------------*
41    * Report Output
42    *-----------------------------------------------------------*
43
44
45
46
47    Summary Statistics for Class Targets
48
49    Data=DATA
50
51                     Numeric    Formatted    Frequency
52        Variable      Value       Value        Count      Percent        Label
53
54    SeriousDlqin2yrs     0           0         139974      93.316     SeriousDlqin2yrs
55    SeriousDlqin2yrs     1           1          10026       6.684     SeriousDlqin2yrs
56
57
58    Data=TRAIN
59
60                     Numeric    Formatted    Frequency
61        Variable      Value       Value        Count      Percent        Label
62
63    SeriousDlqin2yrs     0           0          83983     93.3165     SeriousDlqin2yrs
64    SeriousDlqin2yrs     1           1           6015      6.6835     SeriousDlqin2yrs
65
66
67    Data=VALIDATE
68
69                     Numeric    Formatted    Frequency
70        Variable      Value       Value        Count      Percent        Label
71
72    SeriousDlqin2yrs     0           0          55991     93.3152     SeriousDlqin2yrs
73    SeriousDlqin2yrs     1           1           4011      6.6848     SeriousDlqin2yrs
74
```
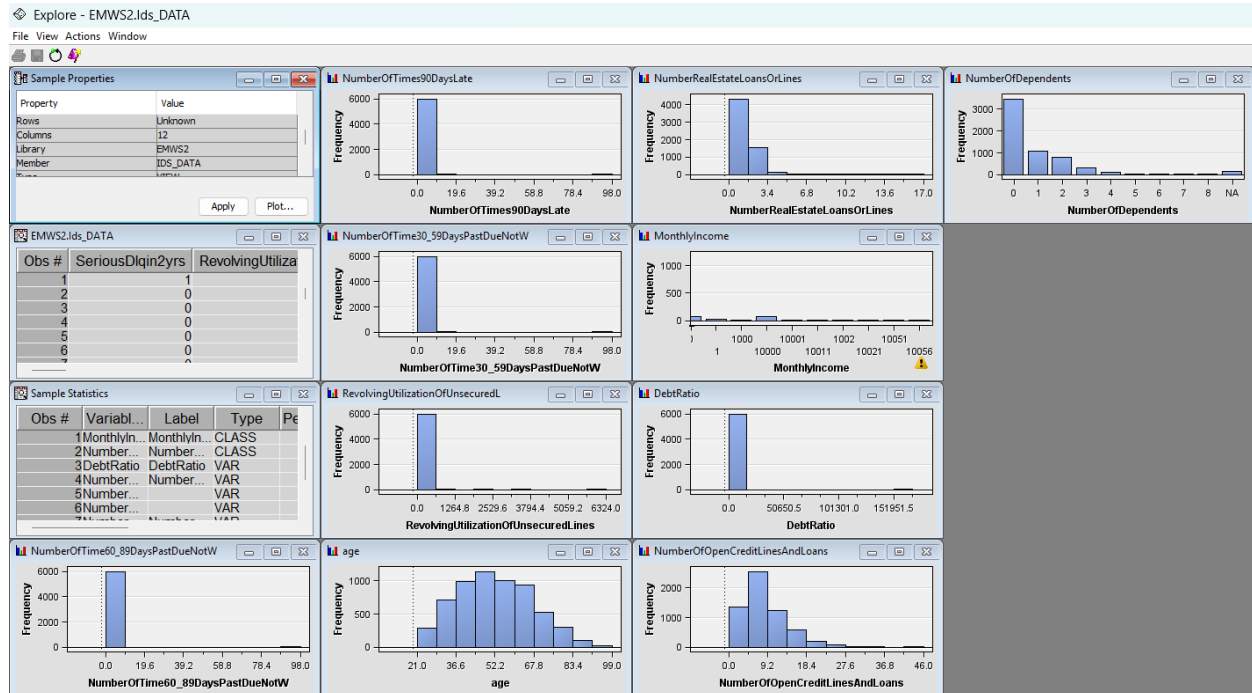
Figure 3: Histograms of Input Variables

Figure 4: User-specified Filter Results

Results - Node: Filter Diagram: Data Preparation and Analysis

File Edit View Window

Output

```
45      Number Of Observations
46
47      Data
48      Role          Filtered      Excluded      DATA
49
50      TRAIN           85259          4739        89998
51      VALIDATE        56897          3105        60002
52
53
54
55      Statistics for Original and FILTERED Data
56      (maximum 500 observations printed)
57
58      Data Role=TRAIN Variable=NumberOfOpenCreditLinesAndLoans
59
60      Statistics              Original      Filtered
61
62      Non Missing             89998.00      85259.00
63      Missing                     0.00          0.00
64      Minimum                     0.00          0.00
65      Maximum                    56.00         18.00
66      Mean                        8.46          7.82
67      Standard Deviation          5.14          4.13
68      Skewness                    1.18          0.40
69      Kurtosis                    2.75         -0.47
70
71
72      Data Role=TRAIN Variable=NumberOfTime30_59DaysPastDueNotW
73
74      Statistics              Original      Filtered
75
76      Non Missing             89998.00      85259.00
77      Missing                     0.00          0.00
78      Minimum                     0.00          0.00
79      Maximum                    98.00          5.00
80      Mean                        0.43          0.23
81      Standard Deviation          4.26          0.63
82      Skewness                   22.24          3.65
83      Kurtosis                  505.62         15.95
84
```

Figure 5: Transform Variables Formulas Window

Figure 6: Logistic Regression Fit Statistics

Figure 7: Logistic Regression Iteration Plot
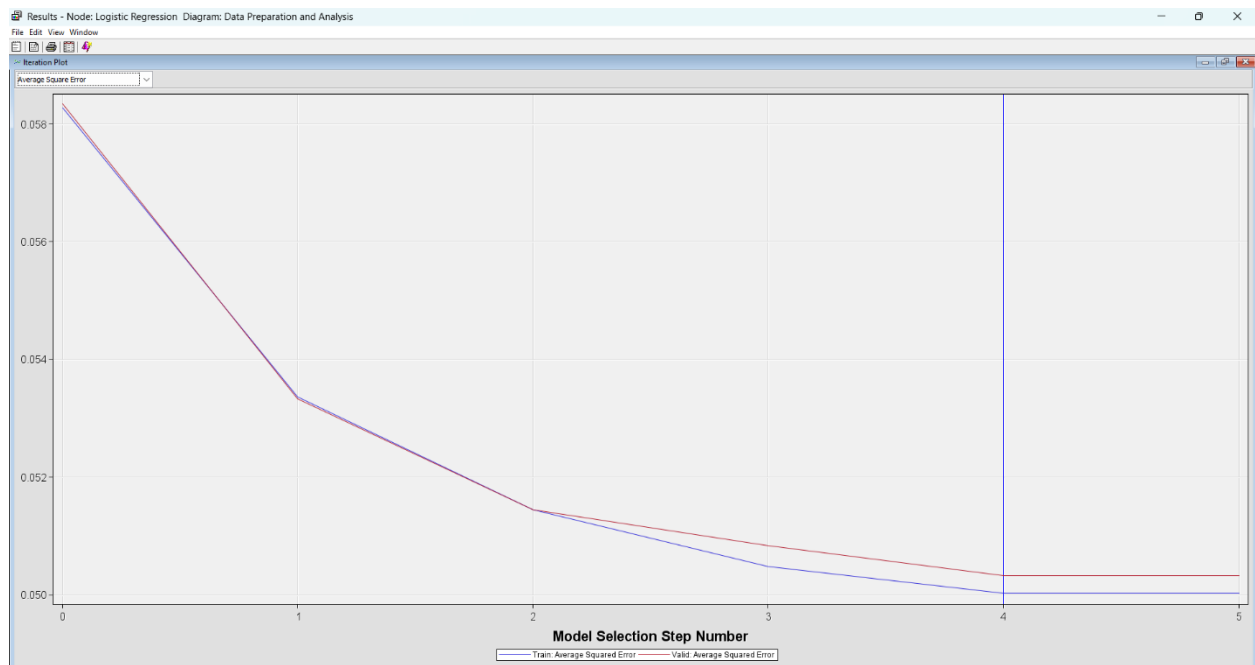
Figure 8: Logistic Regression Cumulative Lift
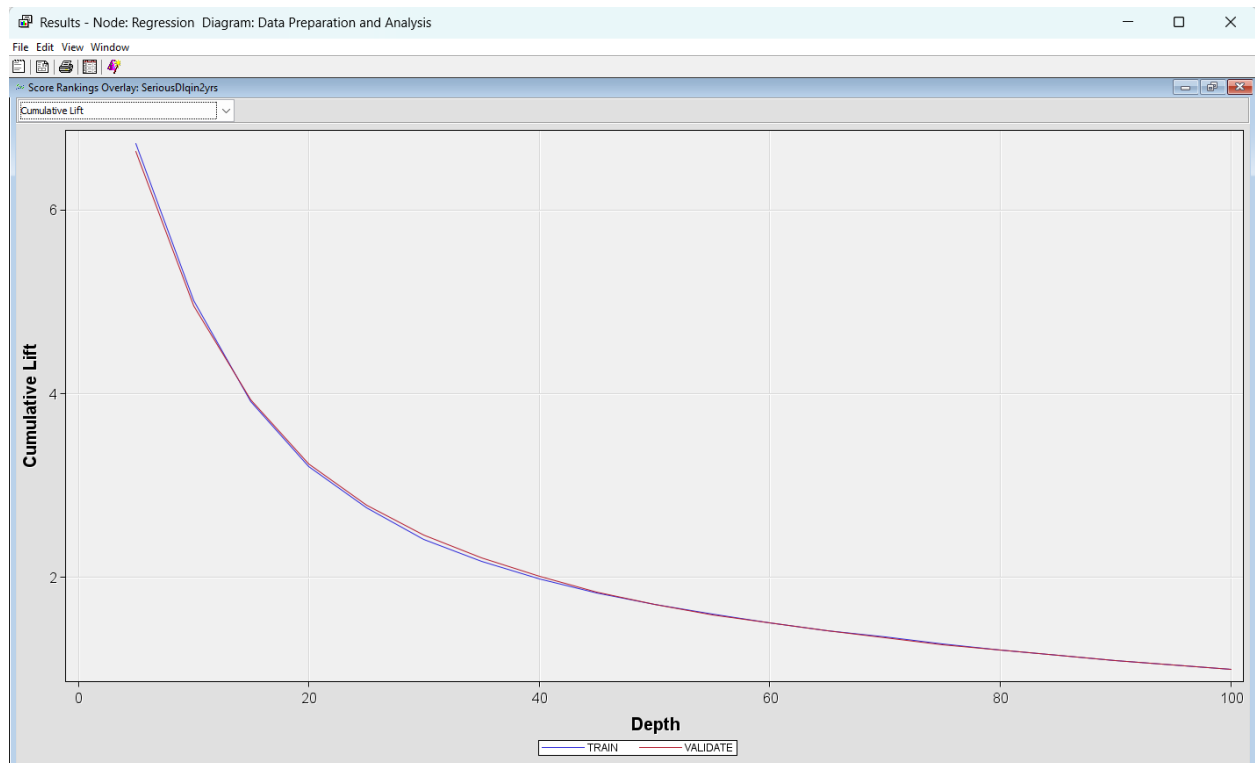
Figure 9: Decision Tree Variable Importance Table

| Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
NumberOfTimes90DaysLate	NumberOfTimes90DaysLate	3	1.0000	1.0000	1.0000
RevolvingUtilizationOfUnsecuredL	RevolvingUtilizationOfUnsecuredLines	11	0.6308	0.6270	0.9940
NumberOfTime60_89DaysPastDueN...		5	0.4828	0.4625	0.9580
NumberOfTime30_59DaysPastDueN...		9	0.4272	0.4492	1.0516
DebtRatio	DebtRatio	4	0.1217	0.0869	0.7138
age	age	1	0.0901	0.0861	0.9556
NumberOfOpenCreditLinesAndLoans	NumberOfOpenCreditLinesAndLoans	2	0.0583	0.0154	0.2641
NumberRealEstateLoansOrLines	NumberRealEstateLoansOrLines	1	0.0470	0.0591	1.2575

Figure 10: Decision Tree Fit Statistics

| Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
SeriousDlqin2yrs	SeriousDlqin2yrs	NOBS	Sum of Frequencies	85259	56897
SeriousDlqin2yrs	SeriousDlqin2yrs	MISC	Misclassification Rate	0.059736	0.061444
SeriousDlqin2yrs	SeriousDlqin2yrs	MAX	Maximum Absolute Error	0.988977	0.988977
SeriousDlqin2yrs	SeriousDlqin2yrs	SSE	Sum of Squared Errors	8071.066	5494.126
SeriousDlqin2yrs	SeriousDlqin2yrs	ASE	Average Squared Error	0.047333	0.048281
SeriousDlqin2yrs	SeriousDlqin2yrs	RASE	Root Average Squared Error	0.217561	0.21973
SeriousDlqin2yrs	SeriousDlqin2yrs	DIV	Divisor for ASE	170518	113794
SeriousDlqin2yrs	SeriousDlqin2yrs	DFT	Total Degrees of Freedom	85259	

Figure 11: Decision Tree Diagram
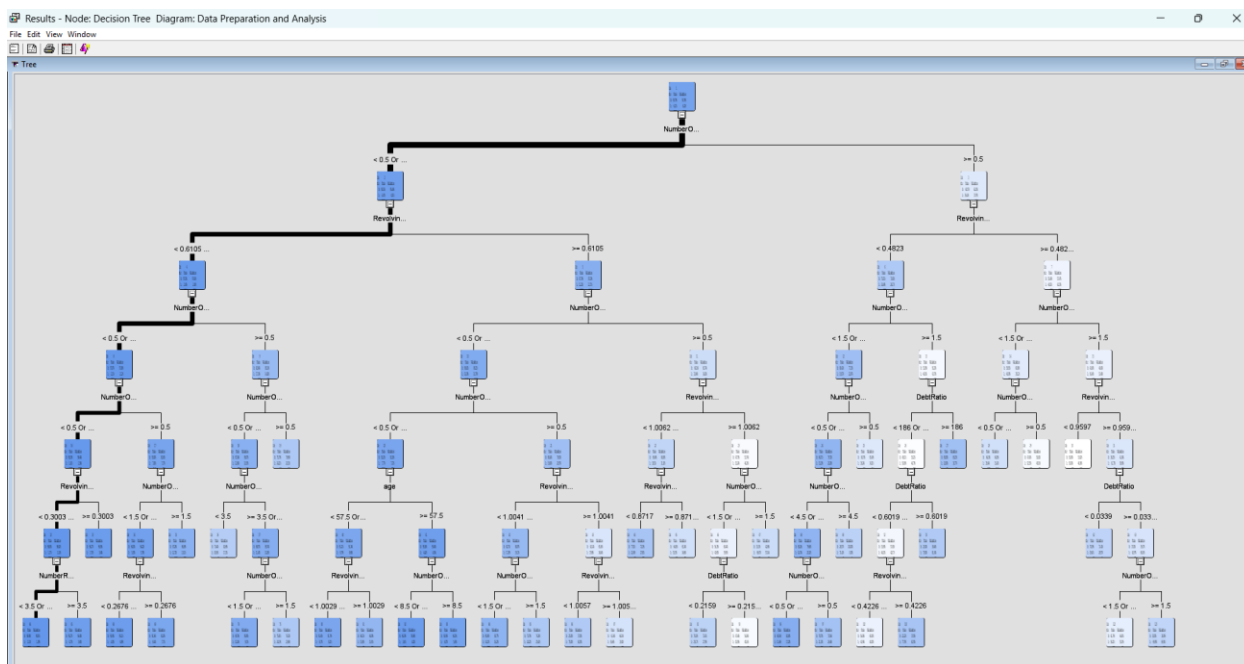
Figure 12: Decision Tree Cumulative Lift
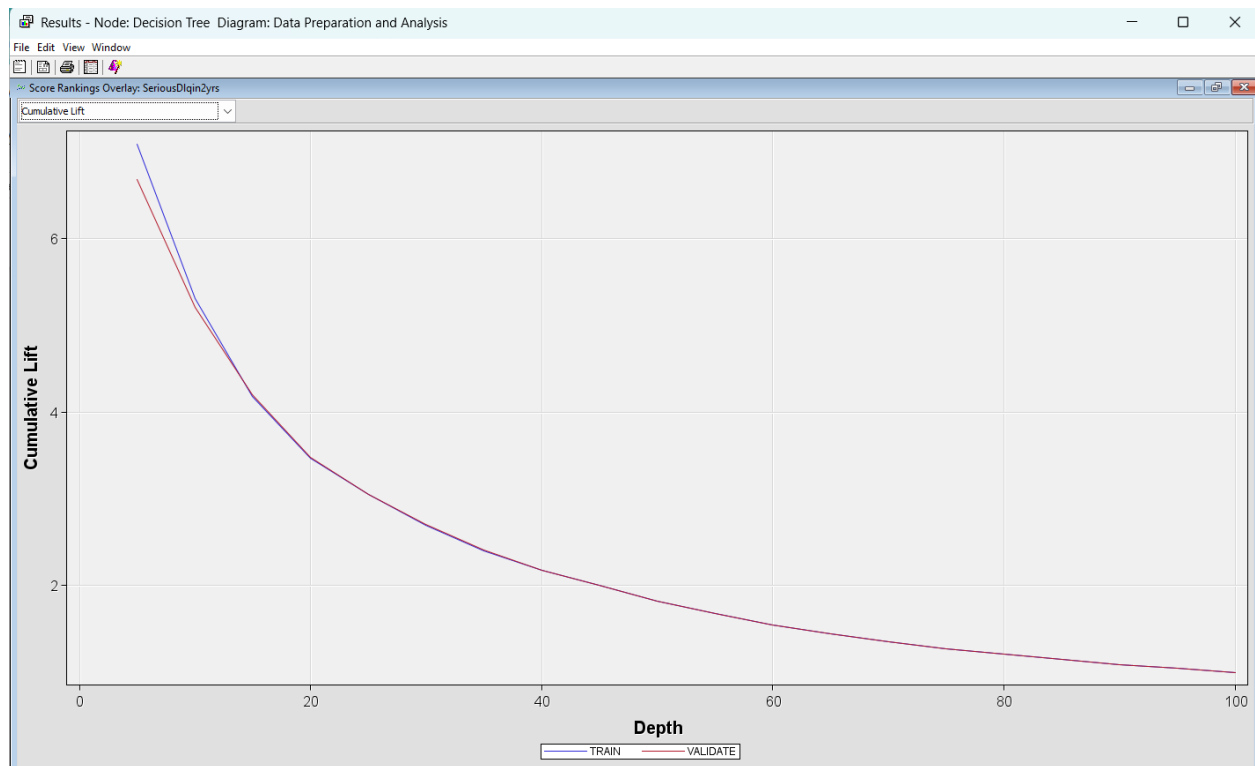
Figure 13: Decision Tree Leaf Statistics
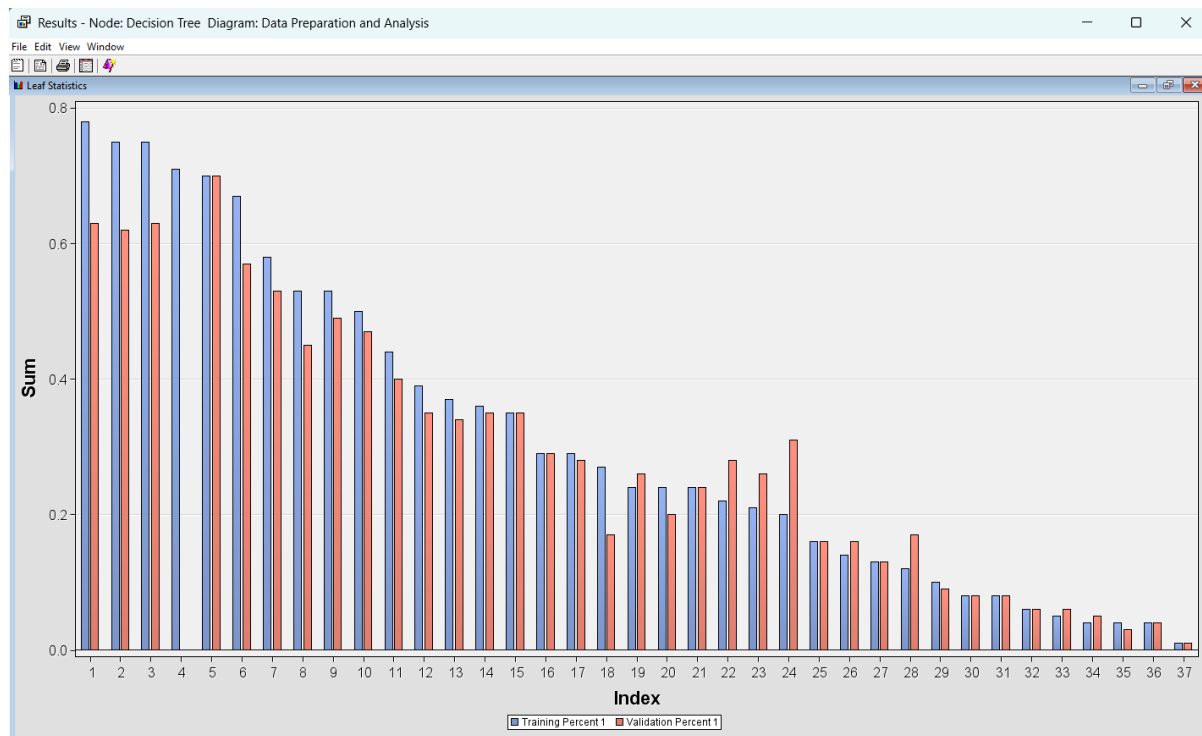
Figure 14: Decision Tree Subtree Assessment Pot