

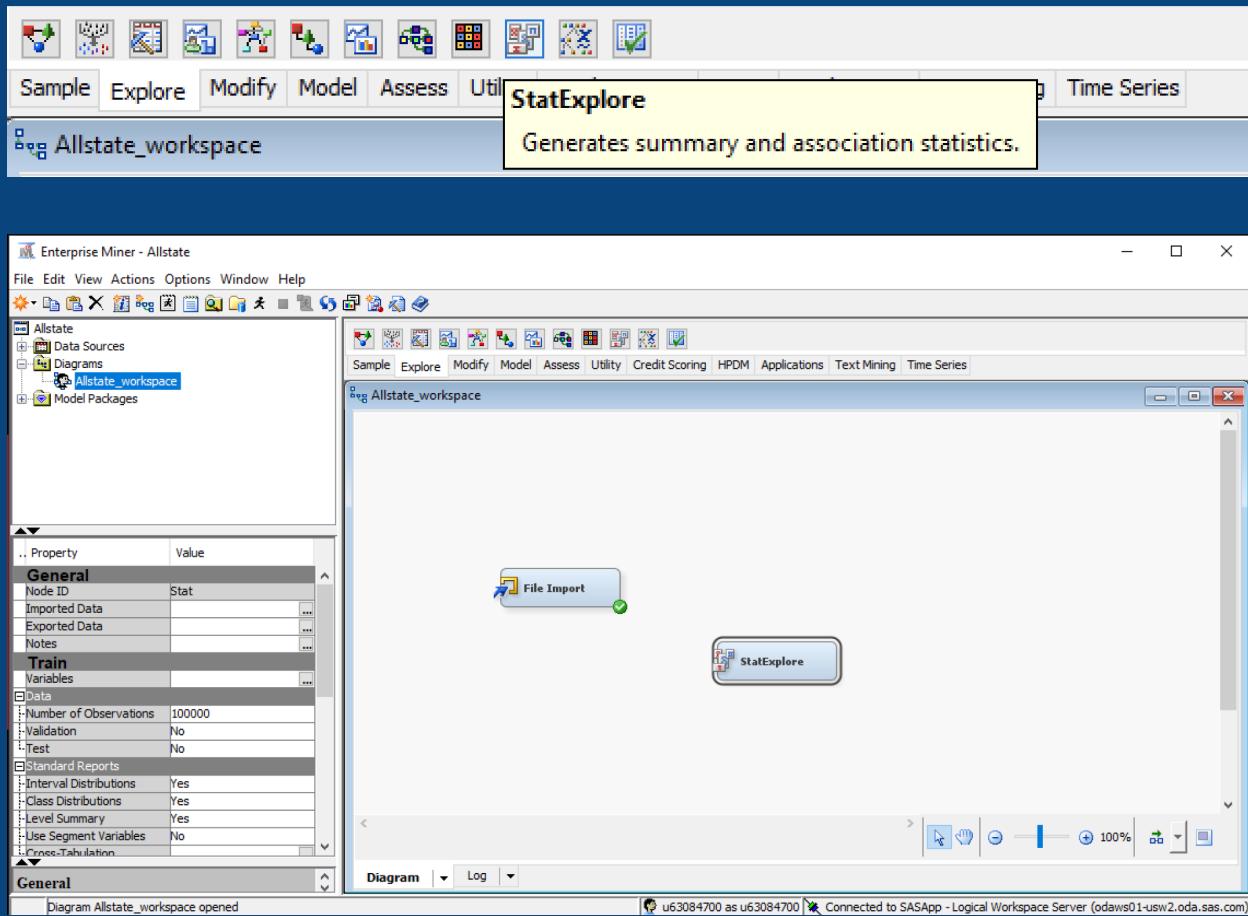
Model Building in SAS Enterprise Miner

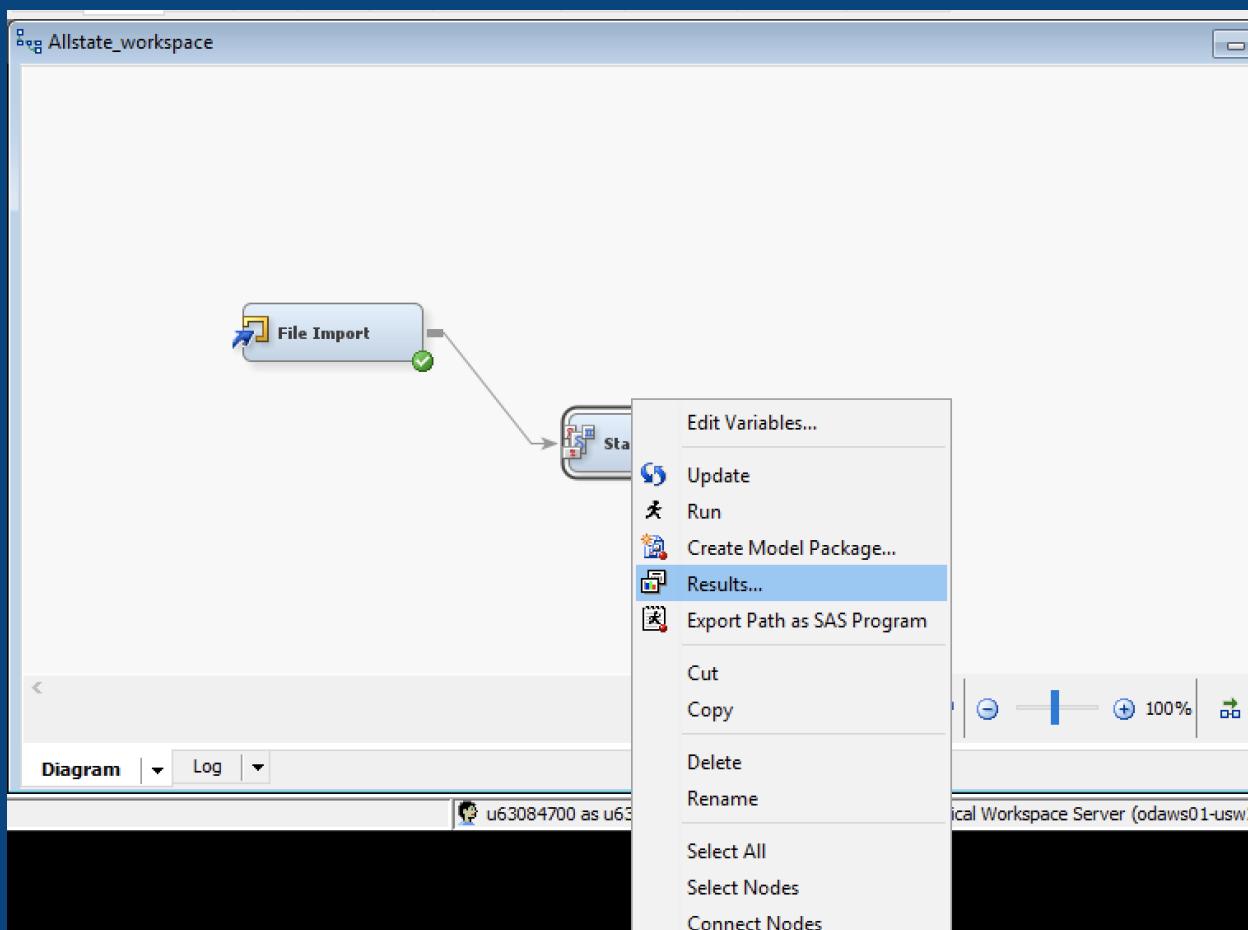
Content

Exploring the Data, Step 1 - p.1
Exploring the Data, Step 1 - p.9
Decision Tree Modelling - p.16
 Interactive Tree - p.18
 Maximal Tree - p.25
 Optimal Tree - p.33
Regression Modeling - p.38
 Forward Regression - p.41
 Fordawd Regression, best sequence model - p.49
 Stepwise regression - p.59
 Stepwise Regression, Adjusted Entry/Stay Levels - p.60
Stepwise Regression, Adjusted Entry/Stay Levels, Best sequence model - p.66

Exploring the Data, Step 1

Now we can explore our data and run some basic statistics





Here we see the table with nominal variables. We see there are two variables (time and location) that have 513 levels which is an issue for us when we run the regression modeling.

```

34
35 Class Variable Summary Statistics
36 (maximum 500 observations printed)
37
38 Data Role=TRAIN
39
40
41 Data
42   Role      Variable Name    Role      Number
43           Levels      of      Mode      Mode2
44   TRAIN     A          INPUT    3        0       0       48.88    1       39.88
45   TRAIN     B          INPUT    2        0       0       82.72    1       17.29
46   TRAIN     C          INPUT    4        0       1       47.20    3       27.49
47   TRAIN     C_previous  INPUT    5        0       3       40.89    1       25.63
48   TRAIN     D          INPUT    3        0       3       44.35    1       29.70
49   TRAIN     E          INPUT    2        0       0       99.59    1       0.41
50   TRAIN     F          INPUT    4        0       0       52.11    2       28.63
51   TRAIN     G          INPUT    4        0       2       46.94    1       24.96
52   TRAIN     VAR1       INPUT   513      0       48600   0.61    50760   0.61
53   TRAIN     car_value  INPUT   10      669     e       35.44    f       23.24
54   TRAIN     day        INPUT   7       0       0       21.71    1       20.15
55   TRAIN     homeowner  INPUT   2       0       0       96.27    1       3.73
56   TRAIN     location   INPUT   513      0       10690   1.10    10853   0.98
57   TRAIN     married_couple INPUT  2       0       0       92.93    1       7.07
58   TRAIN     risk_factor INPUT  5       0       NA      48.12    3       19.33
59   TRAIN     shopping_pt INPUT  12      0       1       18.15    2       17.20
60   TRAIN     state      INPUT  36      0       NY      11.14    PA      8.67
61   TRAIN     record_type INPUT  2       0       0       99.86    1       0.14
62
63
64
65 Distribution of Class Target and Segment Variables
66 (maximum 500 observations printed)
67

```

That's why we returned to Tableau Prep and got rid of minutes for time so that we can reduce the number of its levels down to 23. We will discuss what we can do with the location variable a few moments later.

```

# Output
34
35
36 Class Variable Summary Statistics
37 (maximum 500 observations printed)
38
39 Data Role=TRAIN
40
41             Number
42 Data          of
43 Role   Variable Name   Role  Levels  Missing  Mode  Mode2  Percentage
44
45 TRAIN    A      INPUT    3      0      0    45.72    1    43.04
46 TRAIN    B      INPUT    2      0      0    84.66    1    15.34
47 TRAIN    C      INPUT    4      0      1    44.99    3    29.30
48 TRAIN  C_previous  INPUT    5    5124    3    33.93    1    33.47
49 TRAIN    D      INPUT    3      0      3    45.57    1    28.17
50 TRAIN    E      INPUT    2      0      0    99.59    1      0.41
51 TRAIN    F      INPUT    4      0      0    49.82    2    29.64
52 TRAIN    G      INPUT    4      0      2    46.86    1    24.92
53 TRAIN  car_value  INPUT   10     600    e    35.48    f    23.58
54 TRAIN    day     INPUT    7      0      0    21.11    1    20.53
55 TRAIN  homeowner  INPUT    2      0      0    86.15    1    13.85
56 TRAIN  location   INPUT   513     0    10690   1.68    12898   1.33
57 TRAIN  married_couple  INPUT    2      0      0    96.23    1      3.77
58 TRAIN  risk_factor  INPUT    5      0    NA    46.51    3    19.17
59 TRAIN  shopping_pt  INPUT   12     0      1    18.49    2    17.38
60 TRAIN  state      INPUT   36     0    NY    11.66    FL    9.00
61 TRAIN  time       INPUT   23     0     14    12.53    15    12.13
62 TRAIN  record_type  TARGET   2      0      0    99.86    1      0.14
63
64
65
66 Distribution of Class Target and Segment Variables
67 (maximum 500 observations printed)
<

```

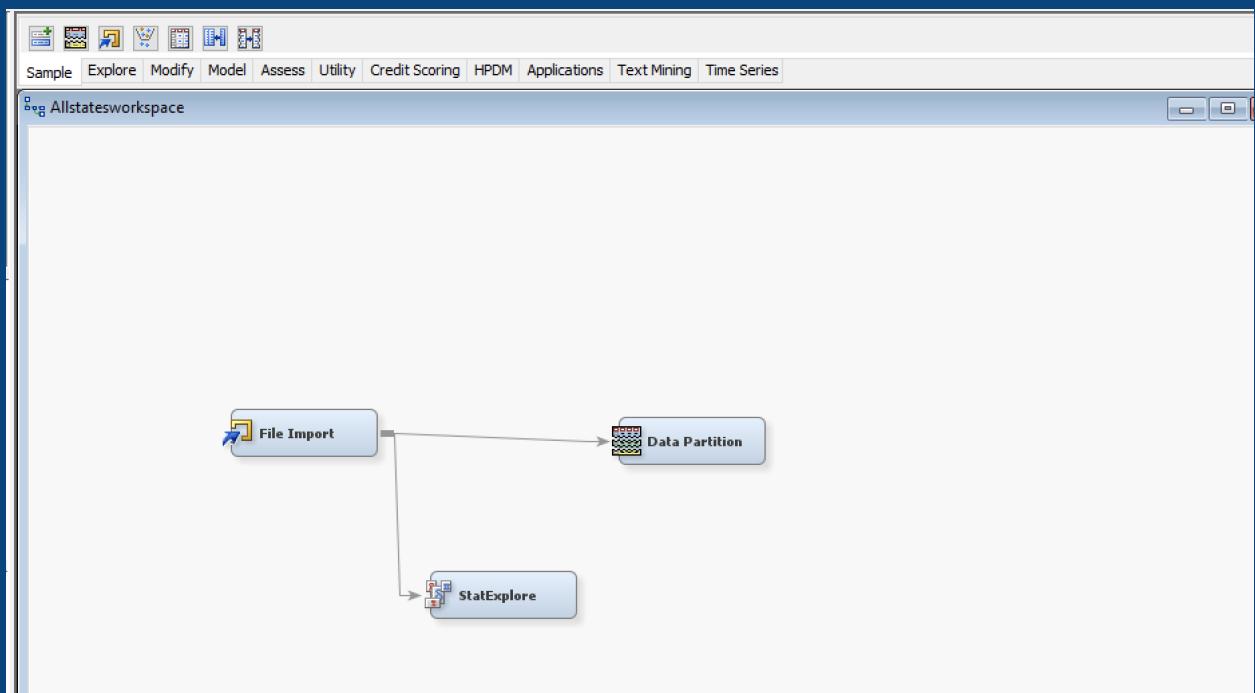
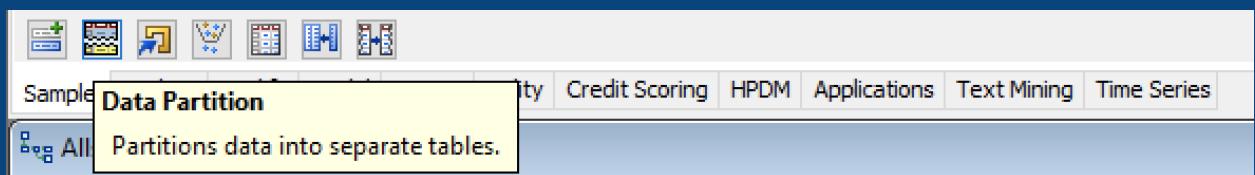
And here we see the table with interval variables

```

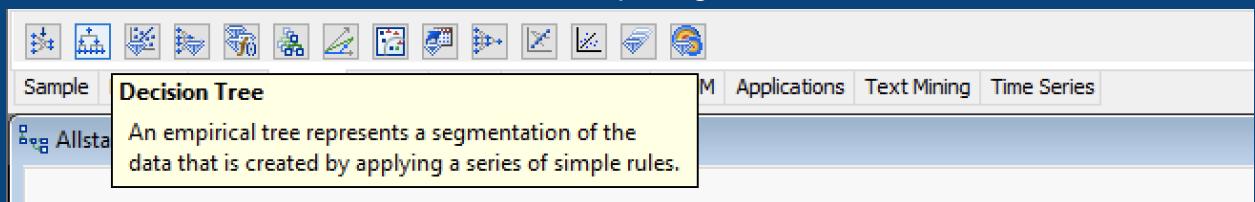
# Output
71 Data  Variable
72 Role  Name   Role  Level  Frequency
73           Count  Percent
74
75 TRAIN  record_type  TARGET    0      99859  99.859
76 TRAIN  record_type  TARGET    1       141  0.141
77
78
79 Interval Variable Summary Statistics
80 (maximum 500 observations printed)
81
82 Data Role=TRAIN
83
84             Standard  Non
85 Variable        Role  Mean  Deviation  Missing  Missing  Minimum  Median  Maximum  Skewness  Kurtosis
86
87 age_oldest     INPUT  37.50674 16.00878 100000    0      18      32      75  0.785494 -0.53625
88 ageyoungest    INPUT  36.38584 15.81086 100000    0      16      30      75  0.882901 -0.34521
89 car_age        INPUT 10.00357  6.247021 100000    0      0      10      54  0.825178 1.889062
90 cost            INPUT 631.8653  48.2027 100000    0     260      631  882  0.124235 0.875078
91 duration_previous INPUT 4.973513 4.165755 94876    5124    0      4      15  1.094918 0.309696
92 group_size      INPUT 1.06877  0.264086 100000    0      1      1      4  3.899436 15.6846
93
94
95
96 Class Variable Summary Statistics by Class Target
97 (maximum 500 observations printed)
98
99 Data Role=TRAIN Variable Name=A
100
101             Number
102 Target  Level  Levels  Missing  Mode  Mode2  Percentage
103
104 <

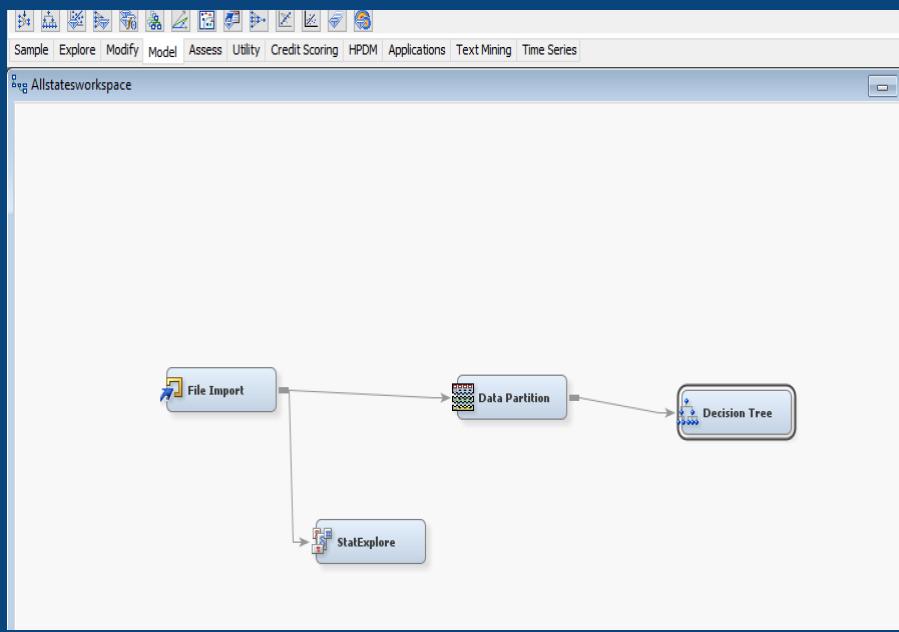
```

Data Partition creation:

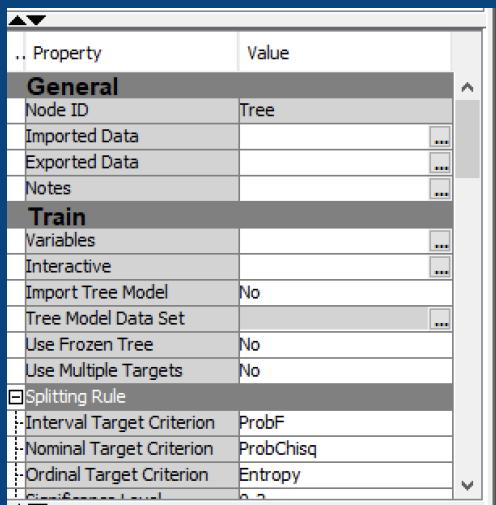


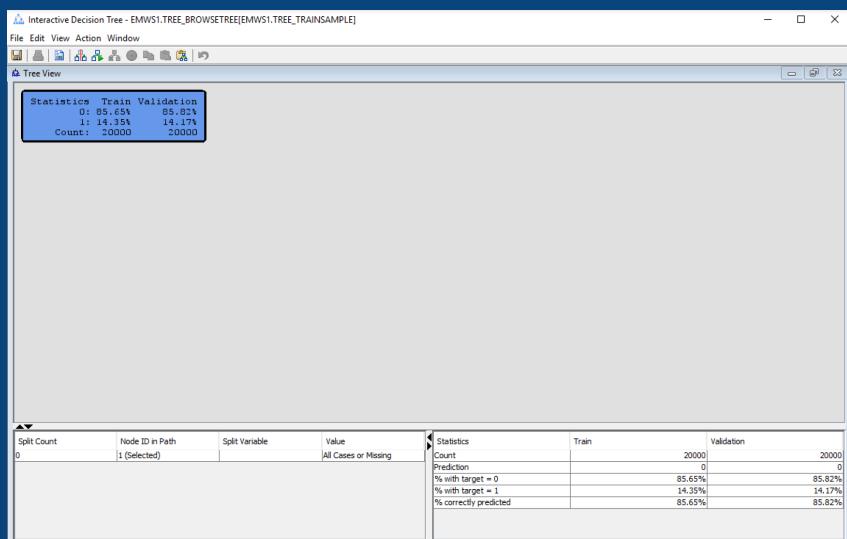
Then we do a decision tree with a manual splitting rule



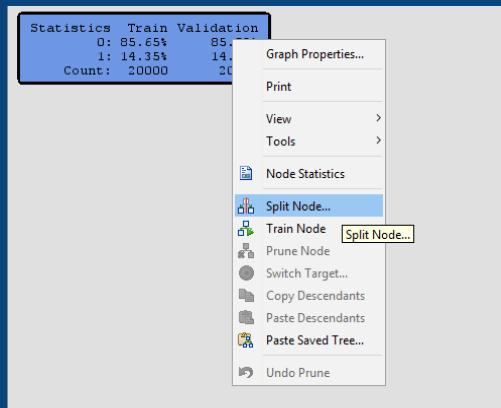


Property panel of DT -> Ellipsis of 'Interactive'





Now we create a splitting rule



12 variables have a logworth higher than 0.7 (the threshold for the first splitting step of DT).

Split Node 1

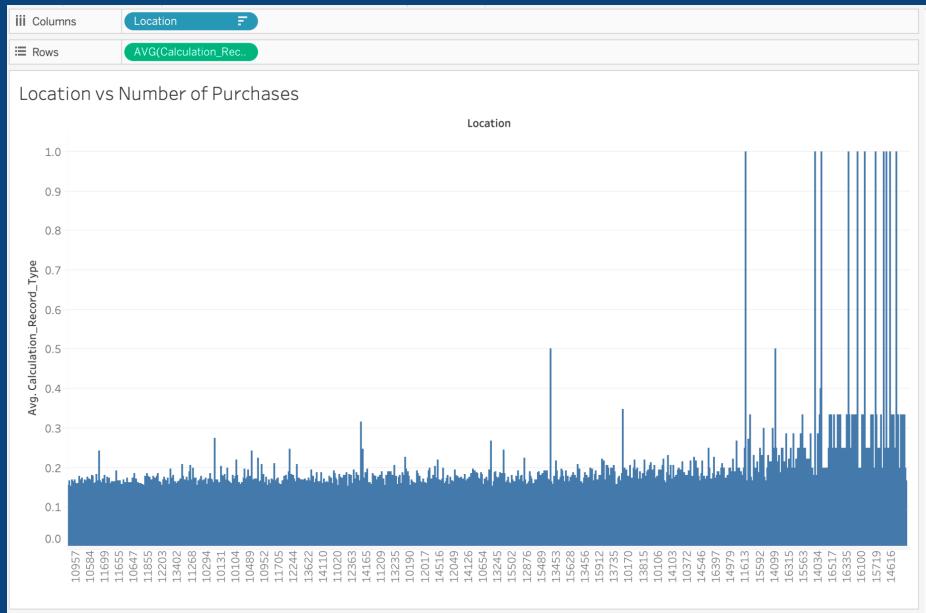
Target Variable: record_type

Variable	Variable Description	-Log(p)	Branches
shopping_pt	shopping_pt	702.9575	2
time	time	17.6498	2
C_previous	C_previous	11.6728	2
duration_previous	duration_previous	11.6448	2
A	A	4.9961	2
cost	cost	2.1726	2
C	C	2.0981	2
D	D	1.8485	2
F	F	1.0655	2
G	G	0.8873	2
day	day	0.8855	2
B	B	0.7732	2
state	state	0.525	2
car_value	car_value	0.3815	2
car_age	car_age	0.3192	2
risk_factor	risk_factor	0.2444	2
E	E	0.1506	2
married_couple	married_couple	0.0879	2
homeowner	homeowner	0.0819	2
group_size	group_size	0.0027	2
age_oldest	age_oldest	0.0	2
age_youngest	age_youngest	0.0	2

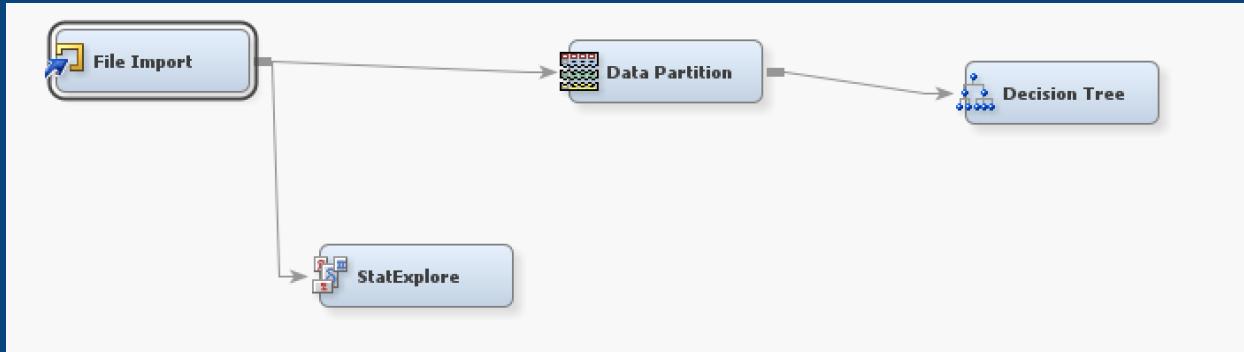
Edit Rule... OK Cancel Apply Refresh

Variable 'Location' is not included in this list, so we can reject it since it will entail 512 dummy variables for us as we get to regression modeling, which will cause many issues for us.

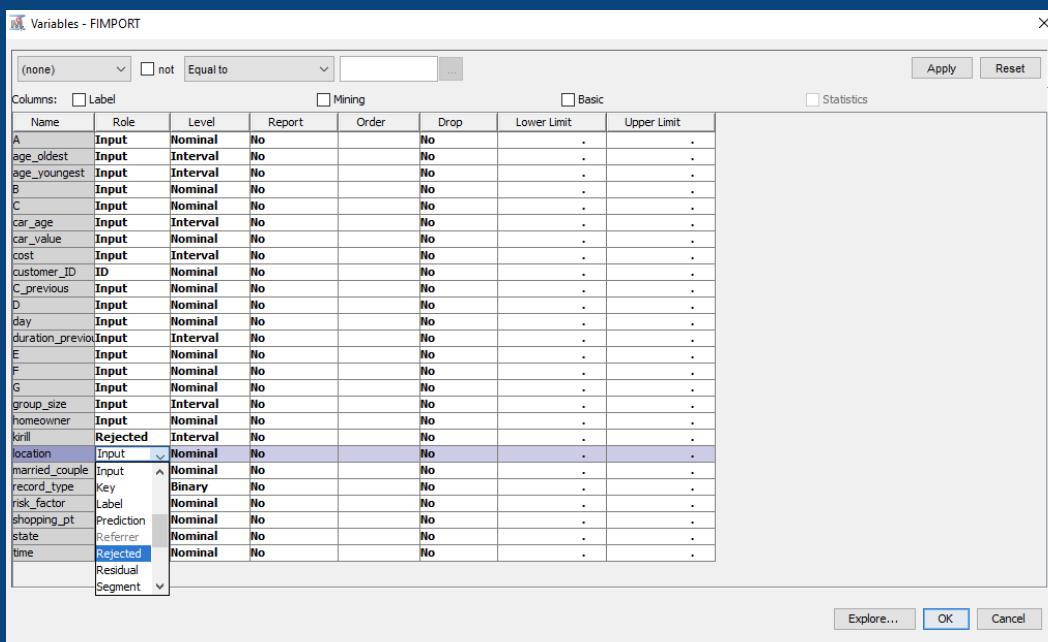
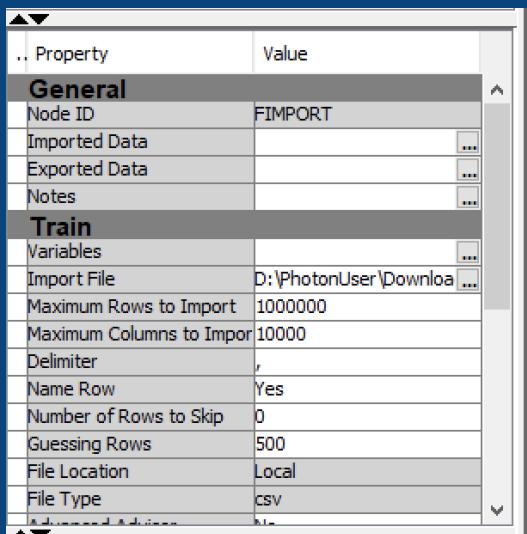
In addition, here is the chart of the average number of sales per location overall time we built in Tableau, and we also see no valuable correlation here (except for a little number of outliers).



So now we can go to the metadata table and reject our 'location' variable.



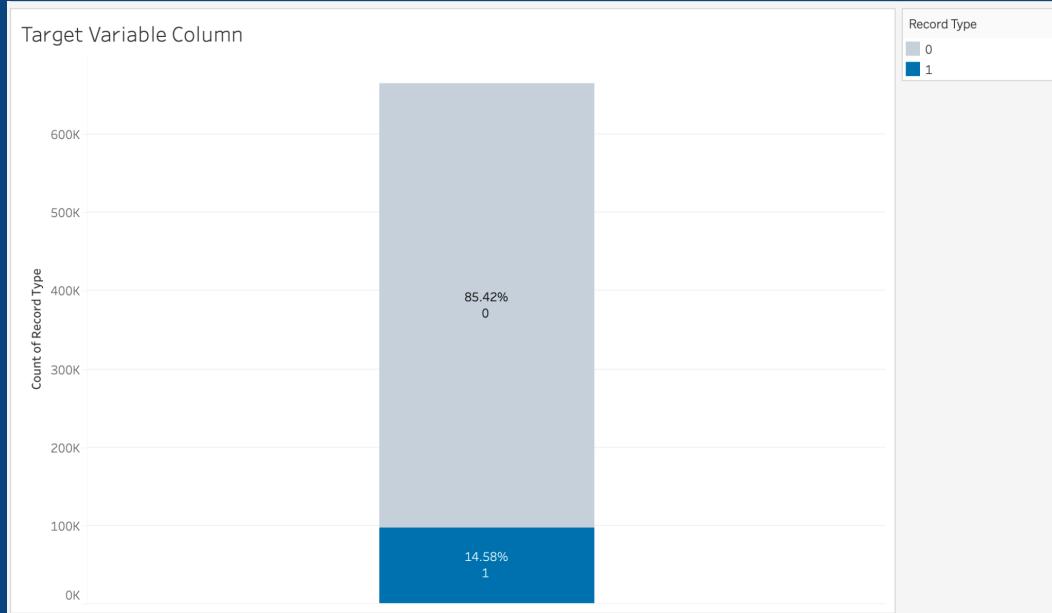
File Import -> property panel: variables



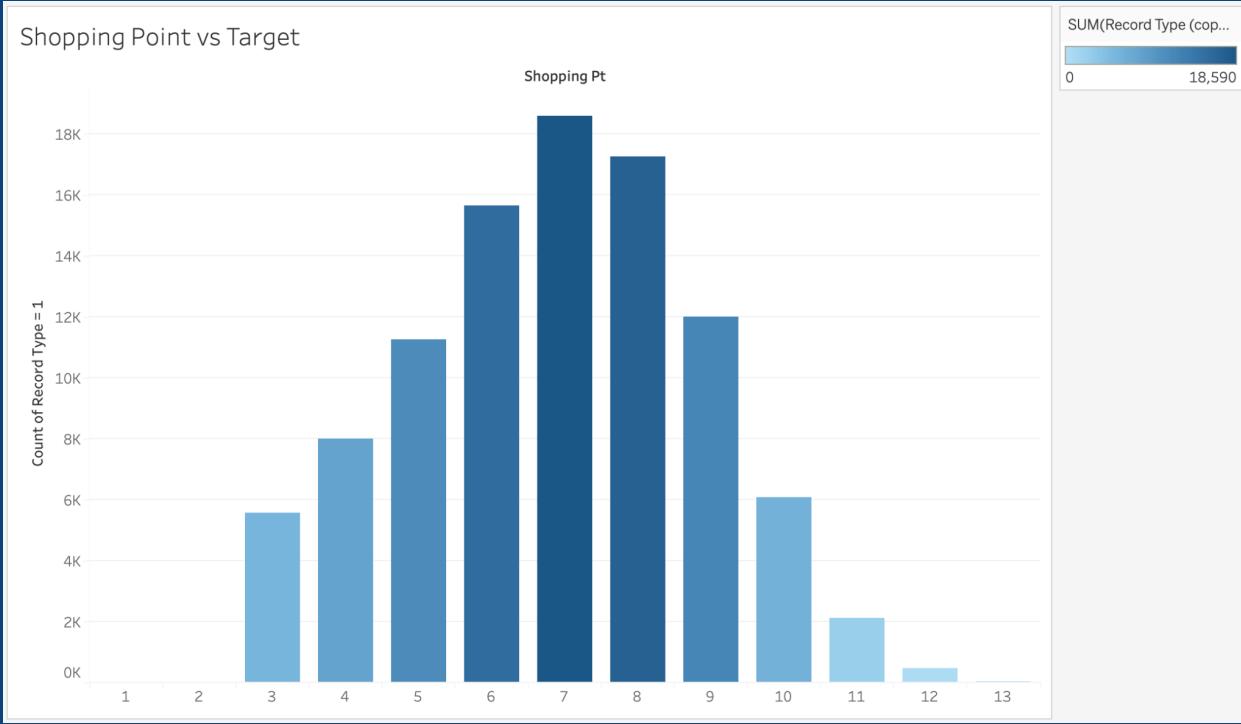
Exploring the Data, Step 2

We will see if there are any explicit correlations between statistically significant variables and our target variable (record_type). In addition, we are going to examine our target variable column visually.

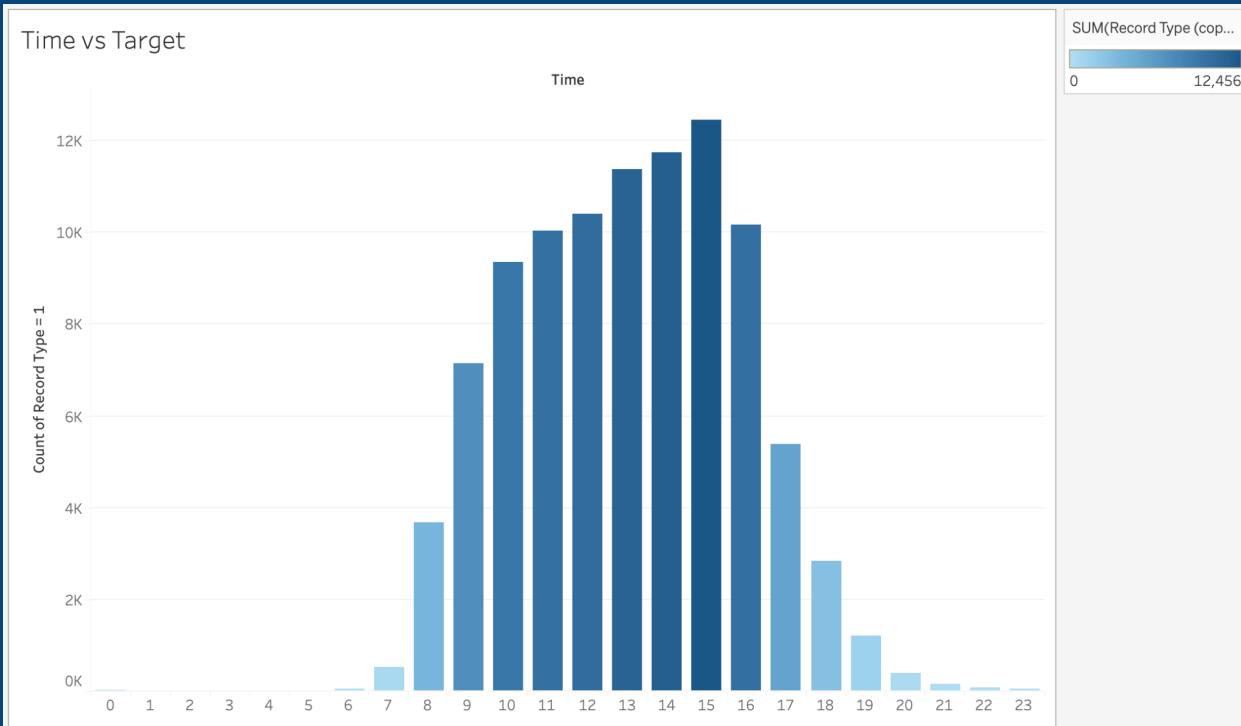
- Here we scaled out the target variable(record_type), and observed record type 0 to hold 85.42% and record type 1 to be 14.58%



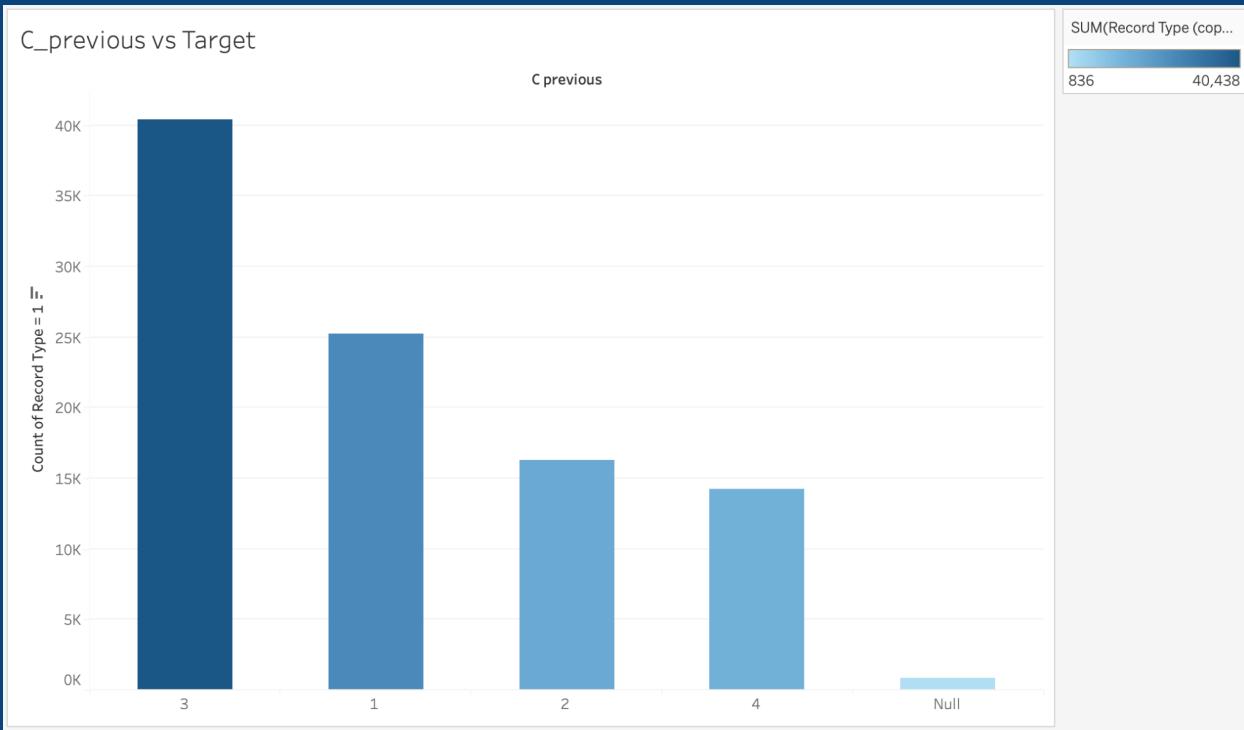
- Scaling the variable shopping_pt against the target variable (record type1), shopping_pt 7 is observed to be the highest



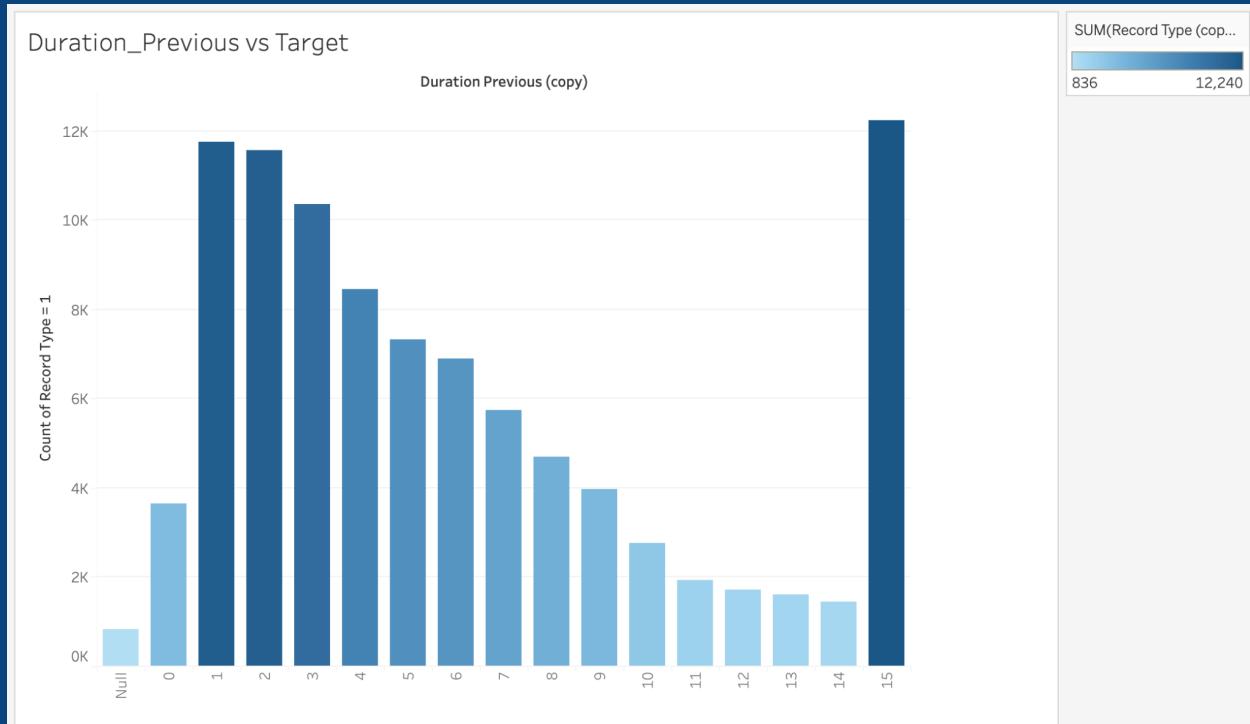
- Scaling the variable Time(in hrs) against the target variable (record type1), it is observed that most transactions occur between the 9th and 16th hour of the day



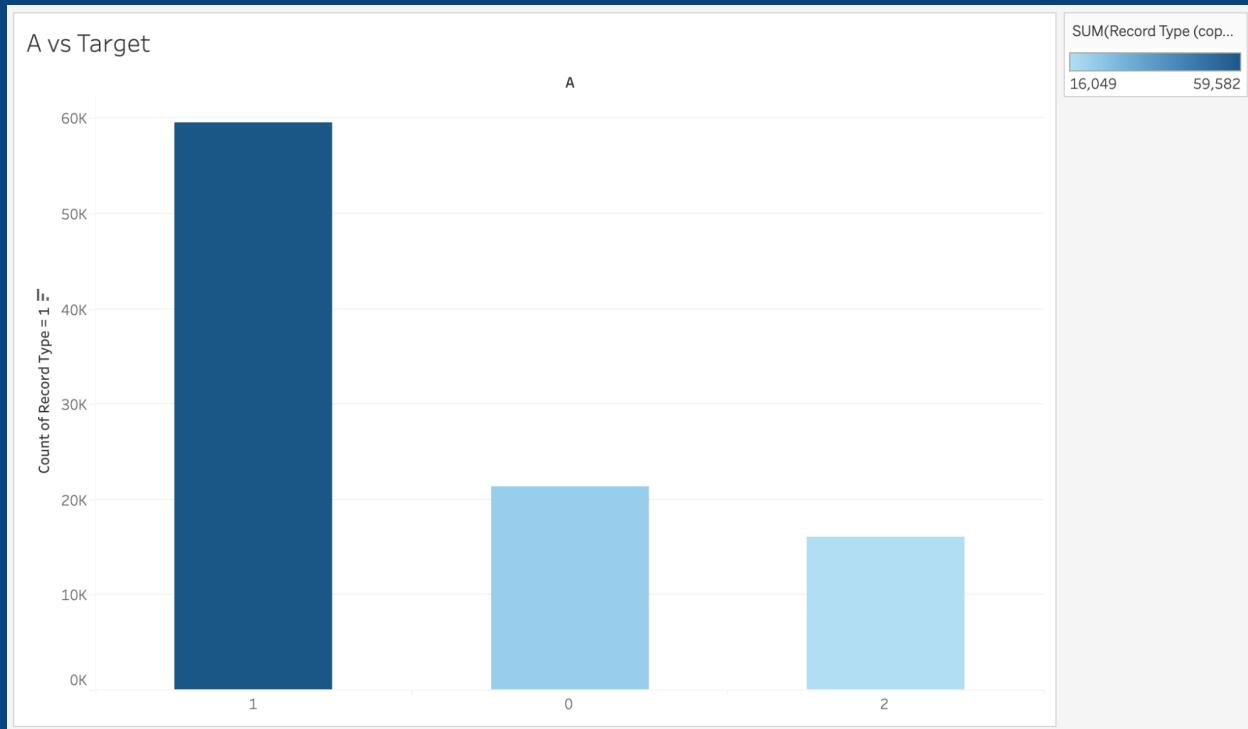
- Comparing the variable C_previous against the target variable (record type1), product option C =3 has the highest rate of sales.



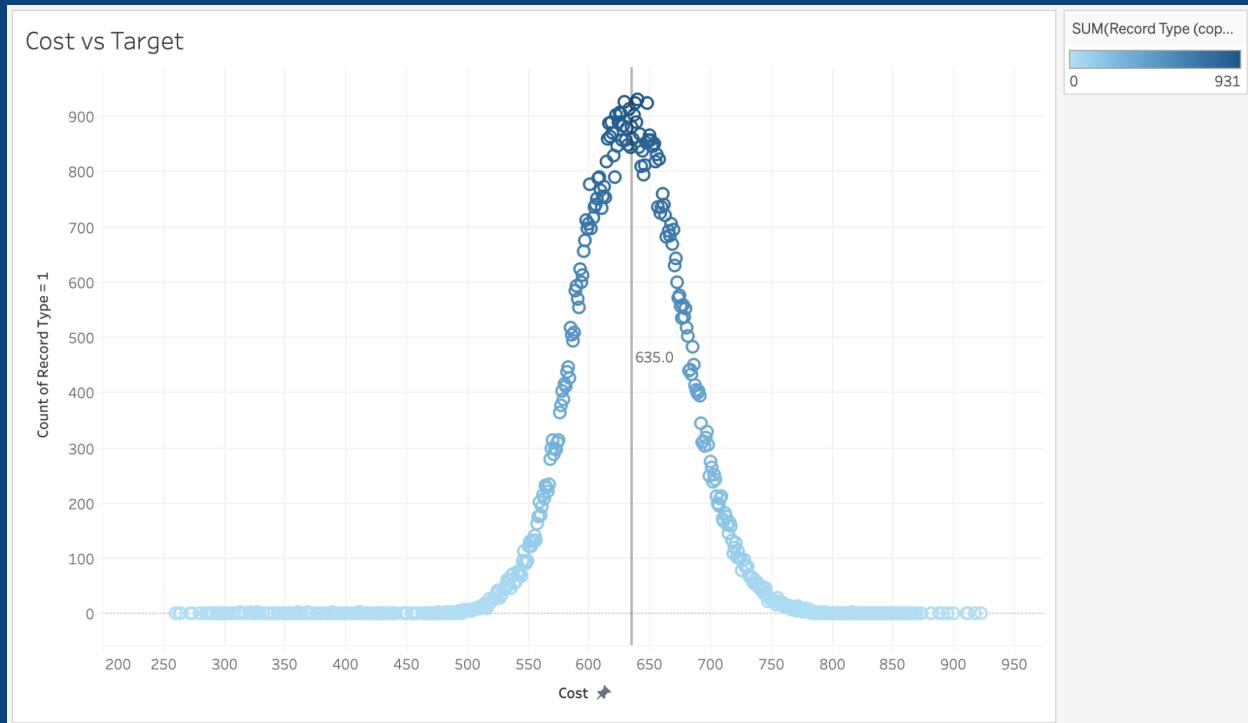
- Comparing the variable Duration_previous against the target variable (record type1), the highest number of policy purchases were done by people who had cover for 1,2 or 15 years.



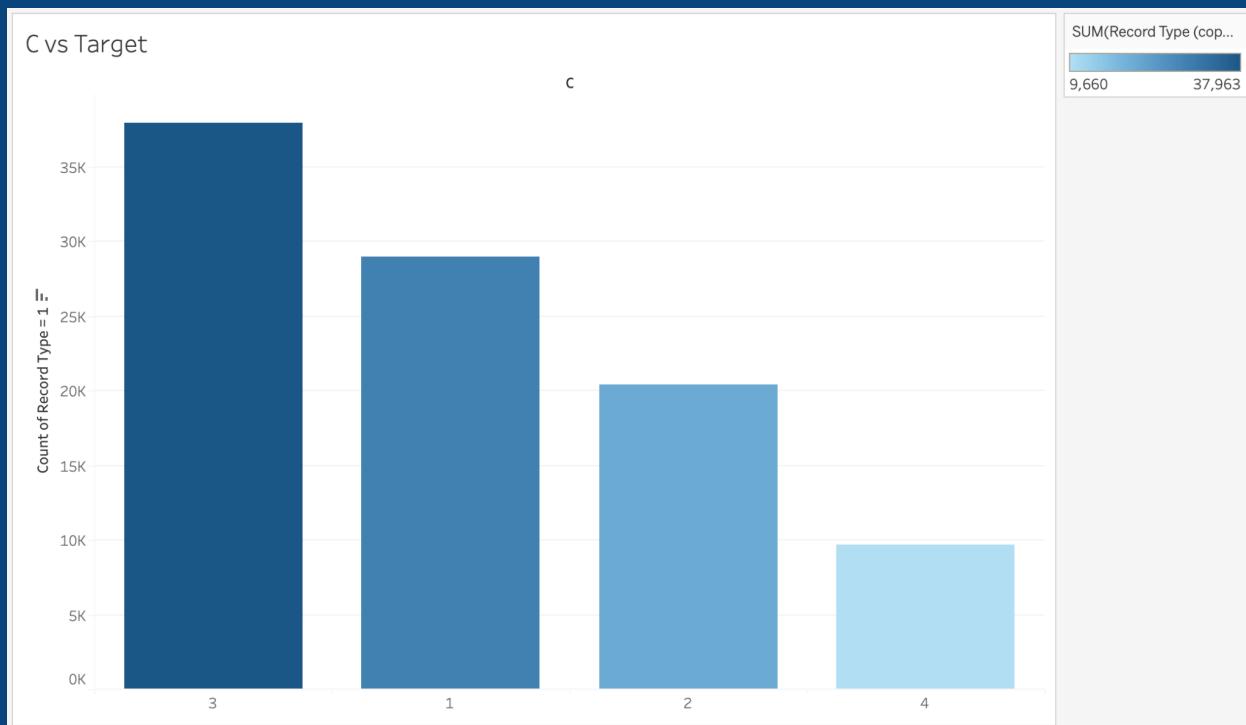
- Comparing variable A against the target variable (record type), product options A = 1 have the highest sales rate.



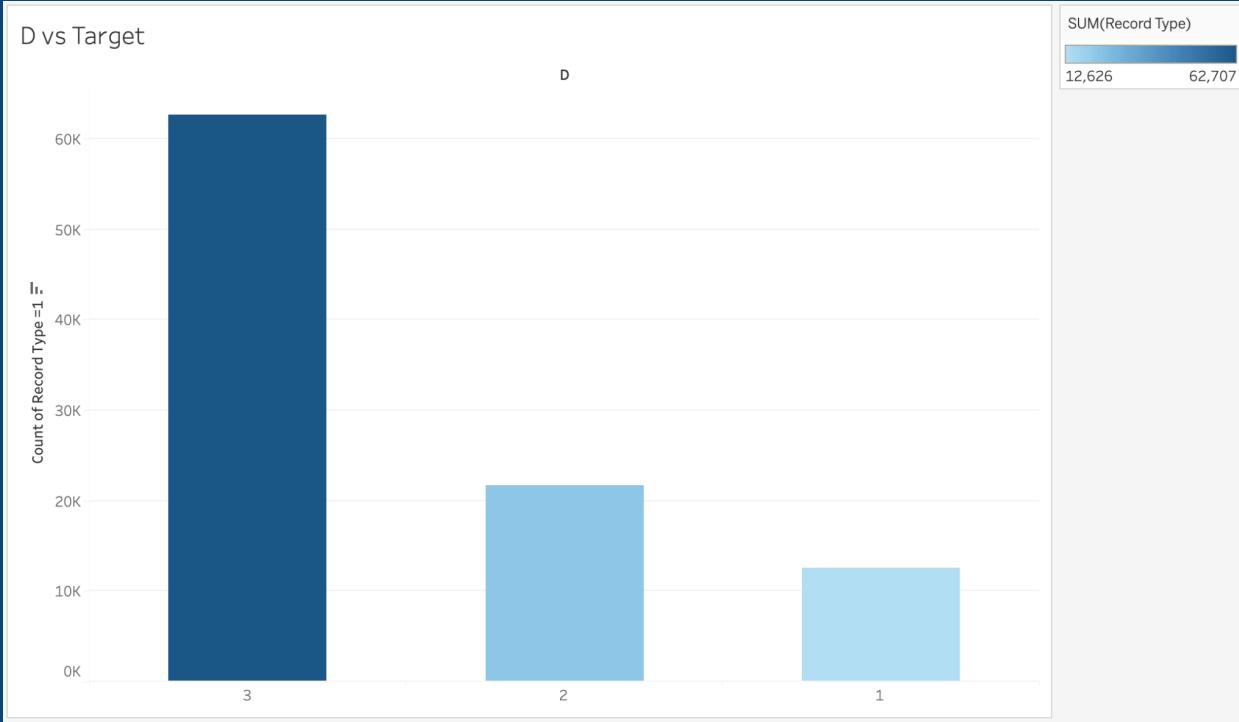
- Comparing variable cost against the target variable (record type), it is exposed that the highest sales rate is around \$635 for cost.



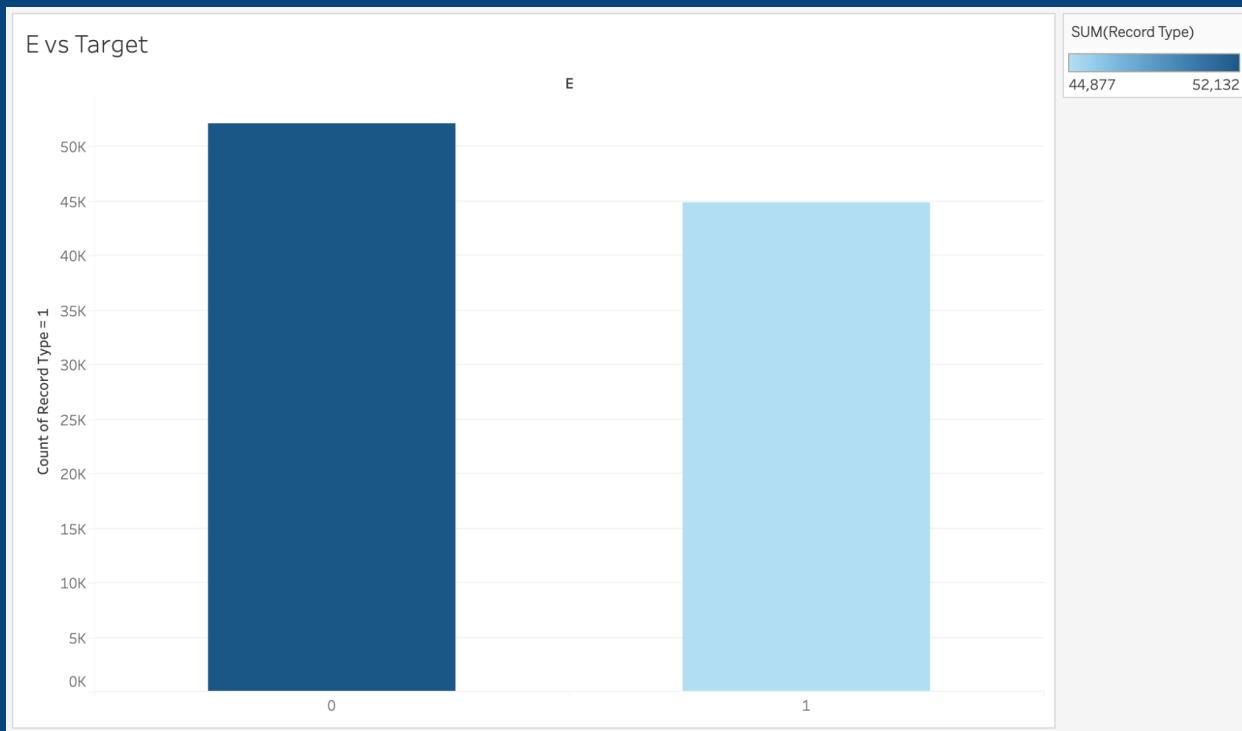
- Comparing variable C against the target variable (record type), it is exposed that option 3 has the highest sales rate.



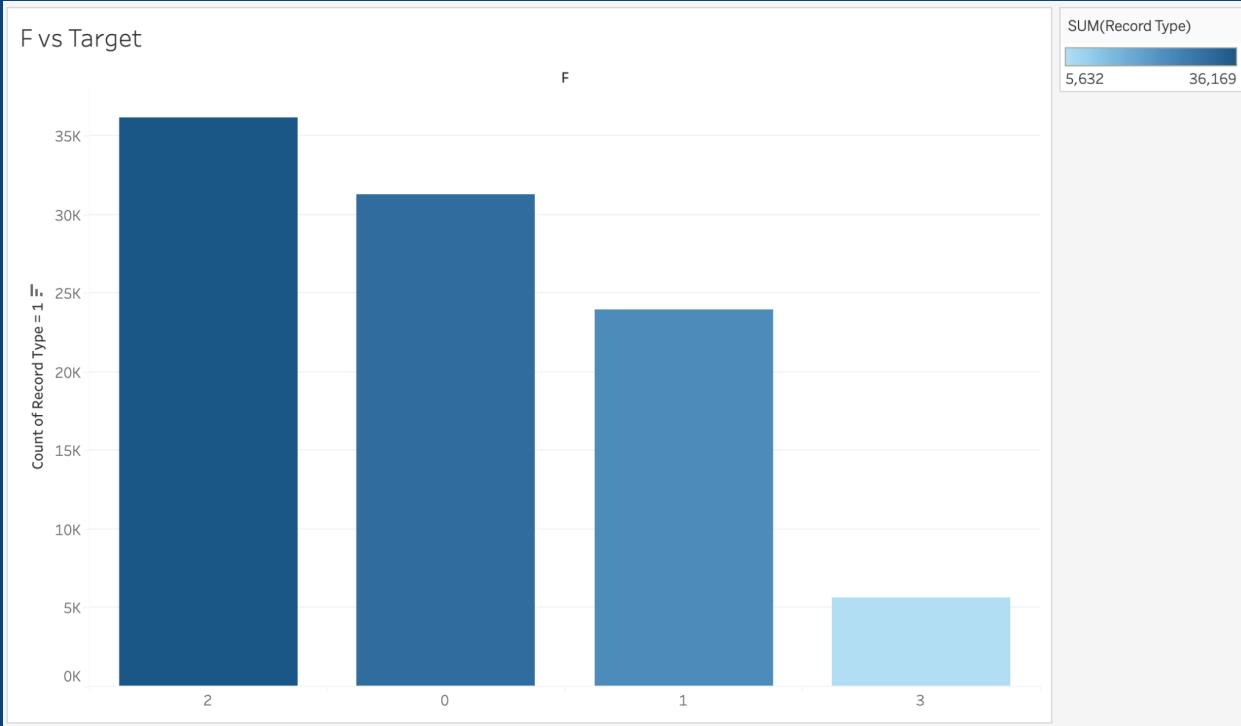
- Comparing variable D against the target variable (record type), it is exposed that option 3 has the highest sales rate.



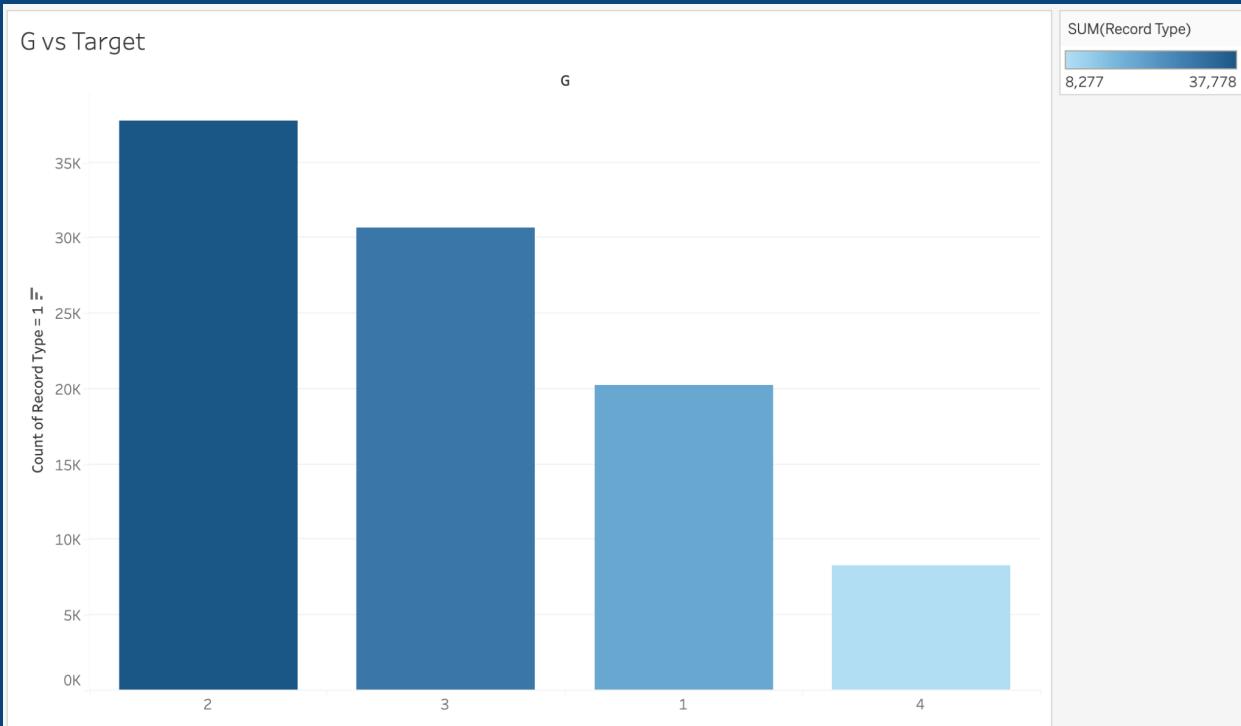
- Comparing variable E against the target variable (record type), it is exposed that option 0 has the highest sales rate.



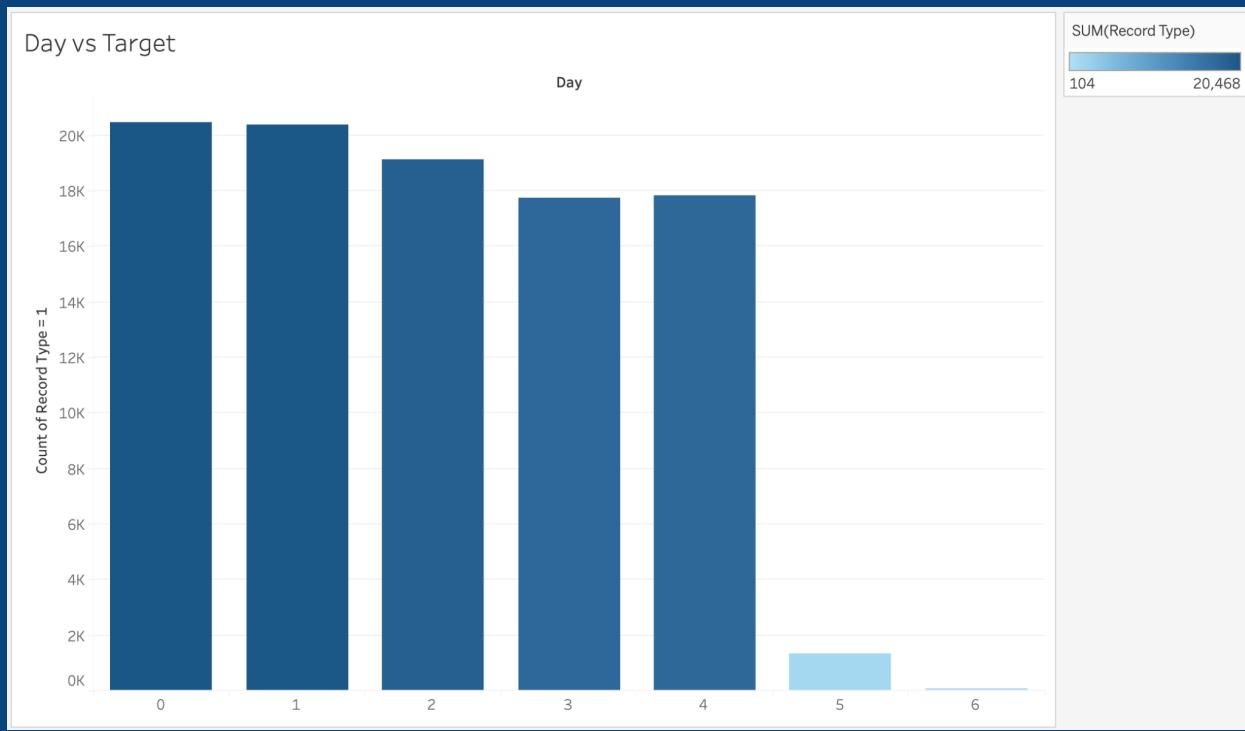
- Comparing variable F against the target variable (record type), it is exposed that option 2 has the highest sales rate.



- Comparing variable G against the target variable (record type), it is exposed that option 2 has the highest sales rate.

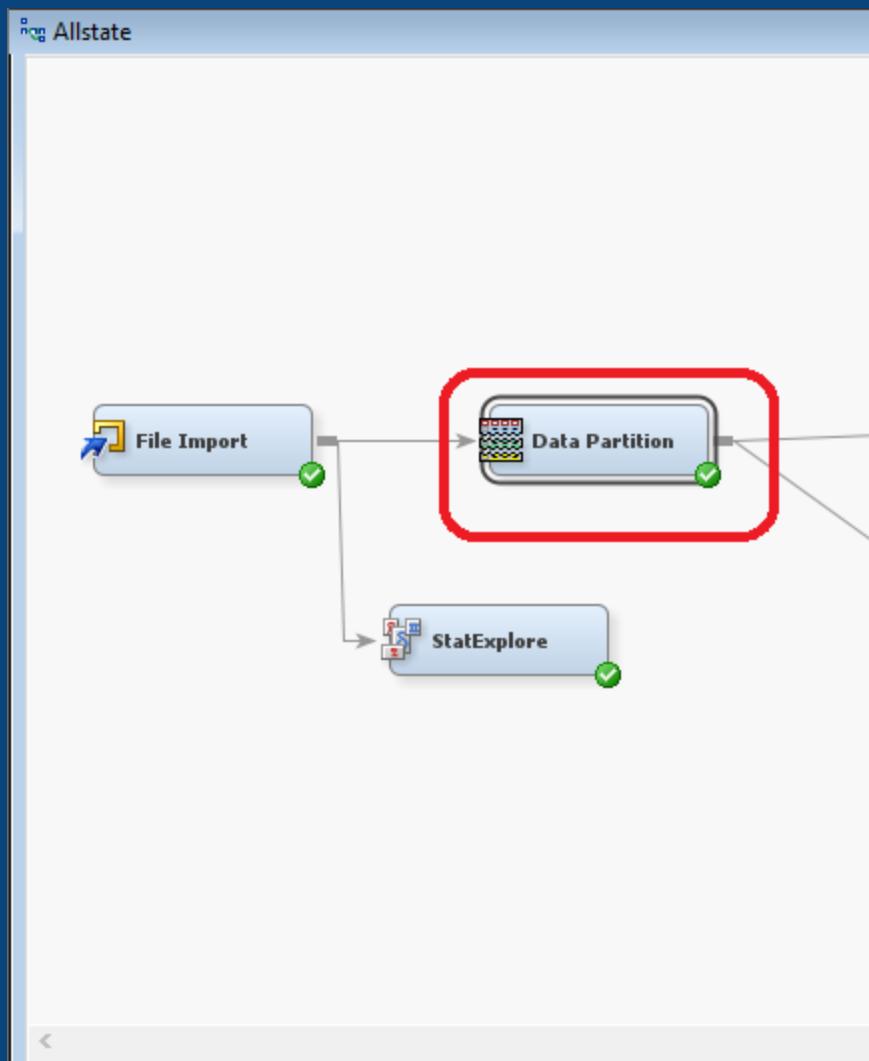


- Comparing variable Day against the target variable (record type), it is exposed that the time range from Monday to Friday is associated with the highest sales rate.



Decision Tree Modelling

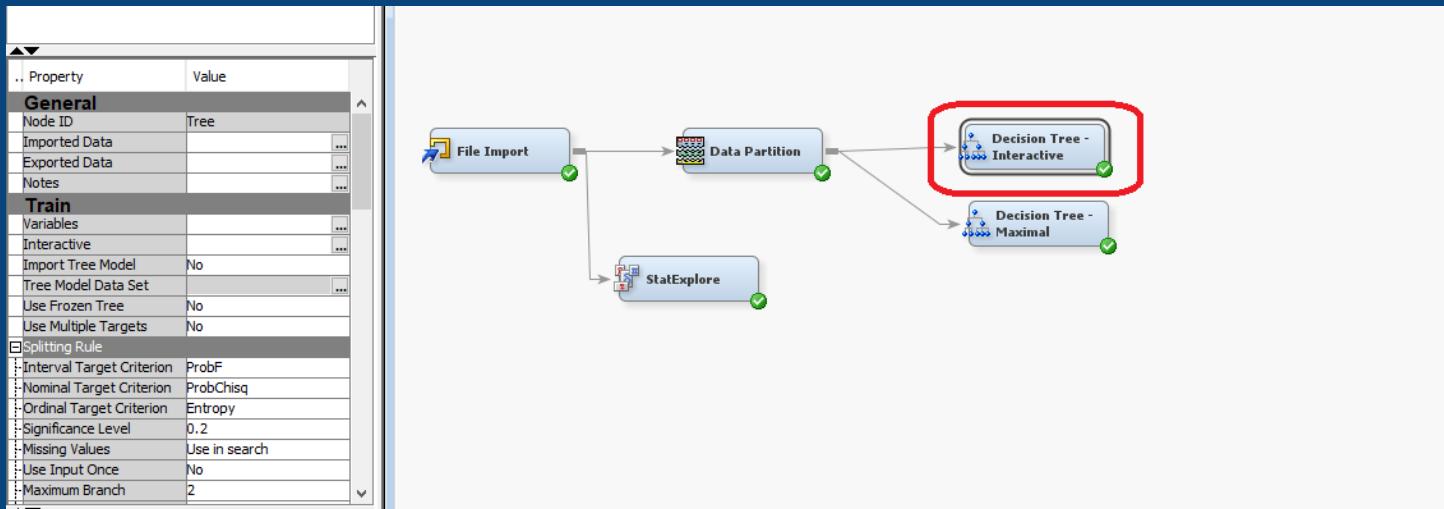
We started by creating the data partition node and splitting optimally Training/Validation/Test as 50/50/0. Why do we allocate our data in such a way? Our top priority is to train a reliable model, that can be applicable to other cases but not only cases we used in the training process, that's why we used 50% of data for training and 50% of data for validation. We left 0% of data for testing, because we are able to assess the degree of optimality and applicability of a model by the validation process.



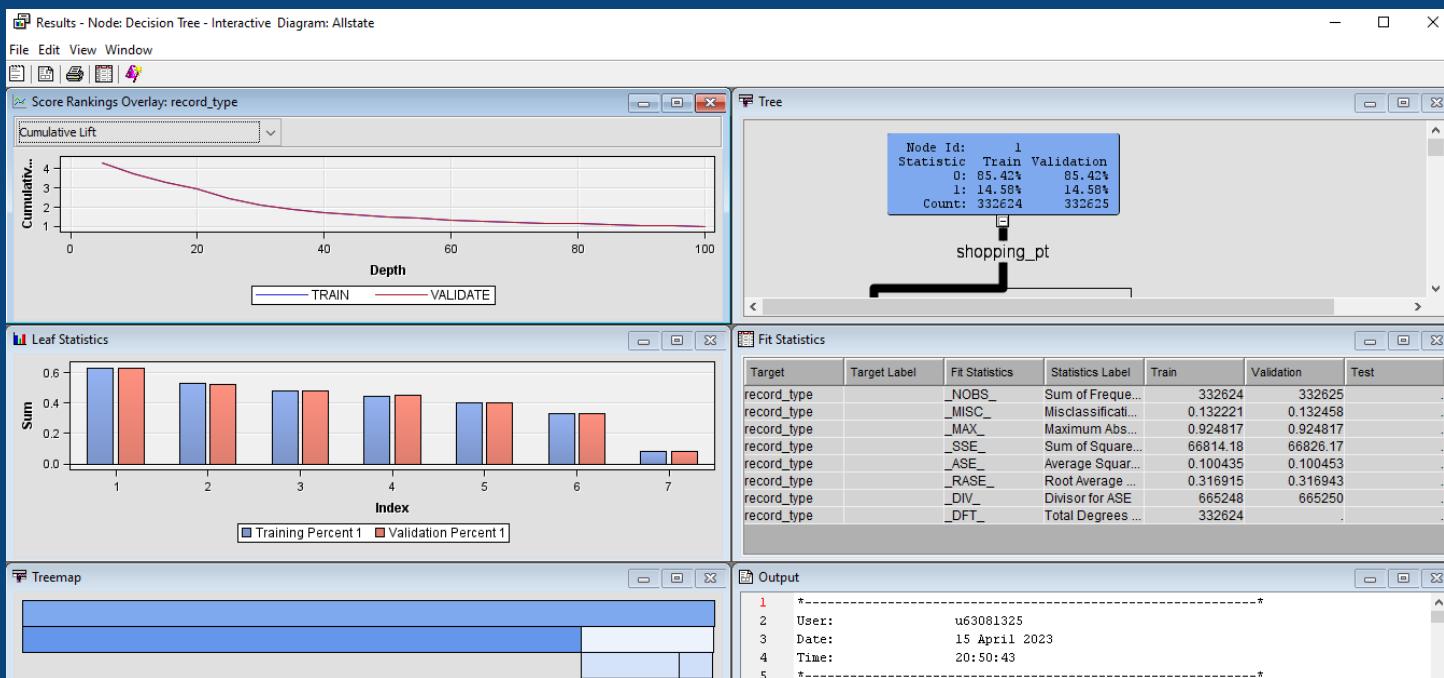
.. Property	Value
General	
Node ID	Part
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Partitioning Method	Default
Random Seed	12345
Data Set Allocations	
Training	50.0
Validation	50.0
Test	0.0
Report	
Interval Targets	Yes
Class Targets	Yes
Status	
Create Time	4/14/23 10:45 PM
Run ID	bb0c2814-65fe-eb4d-9ae5

Interactive Tree

Then creating the interactive decision tree

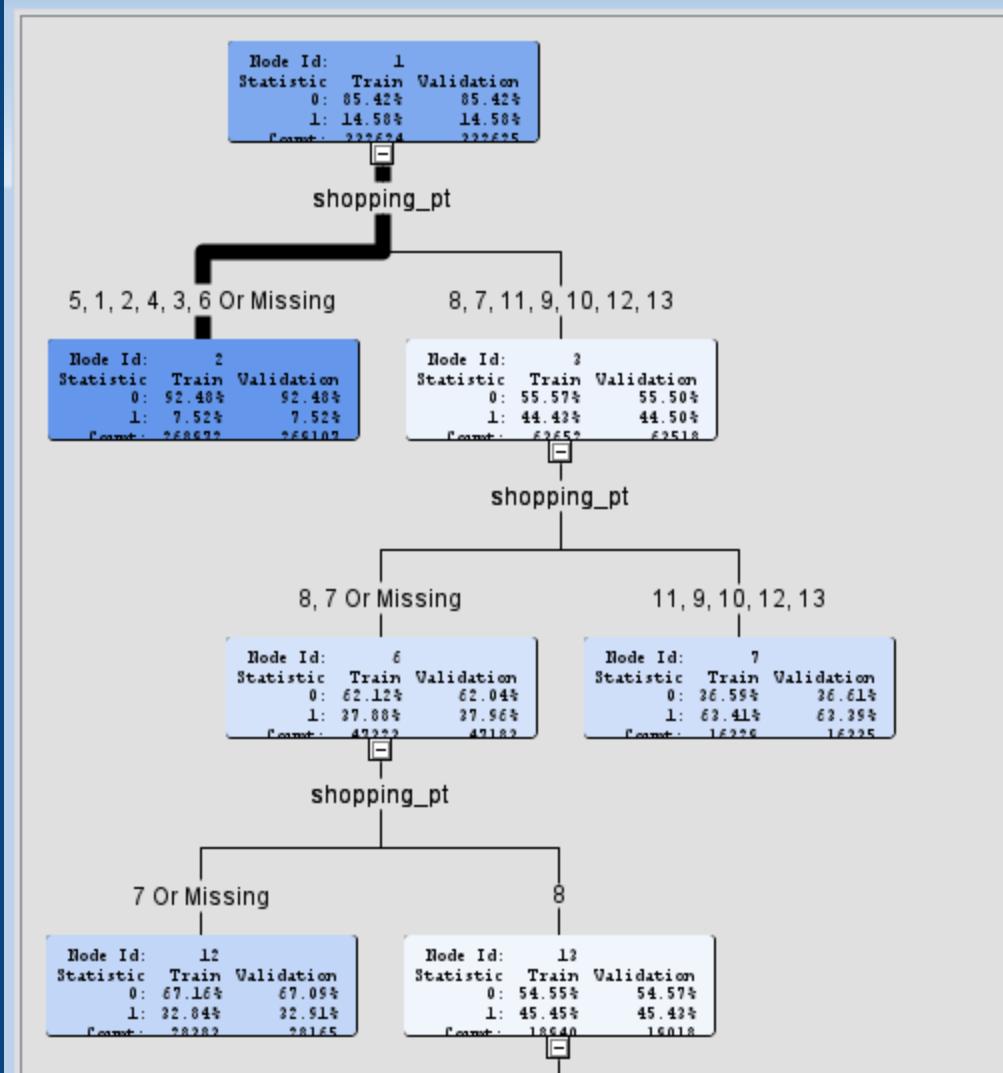


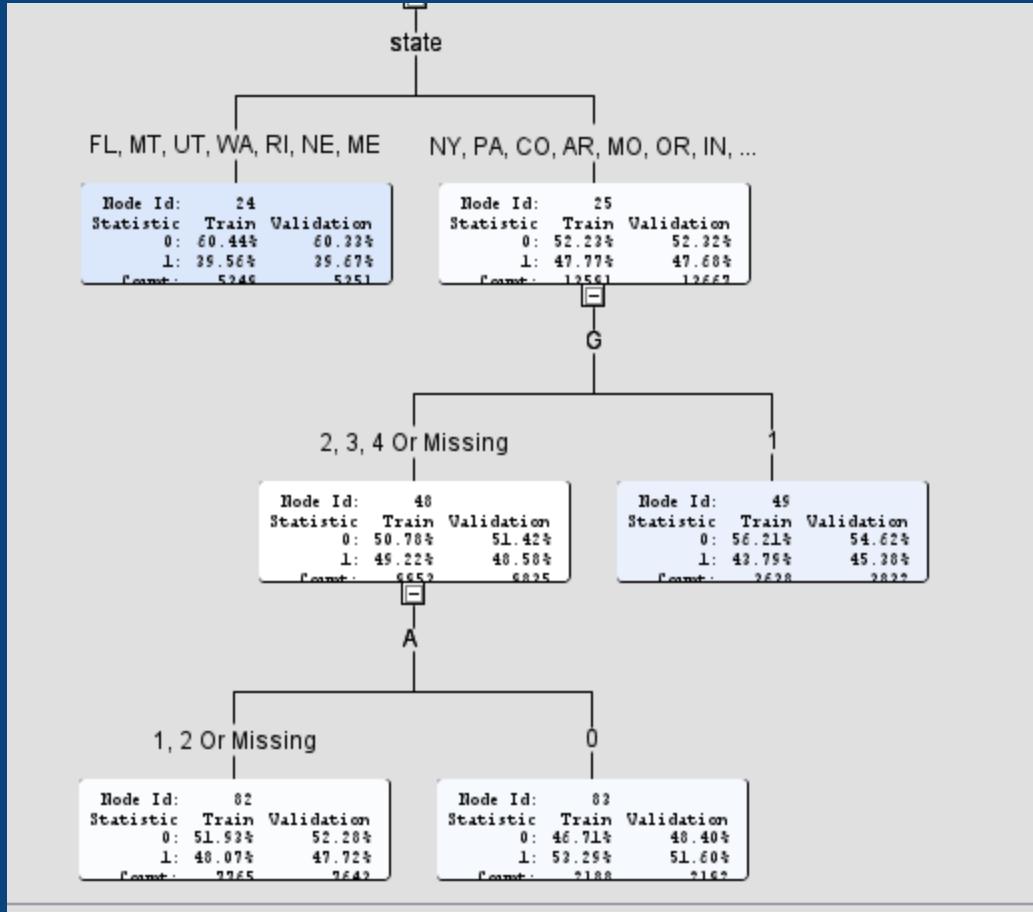
The result window of interactive decision tree



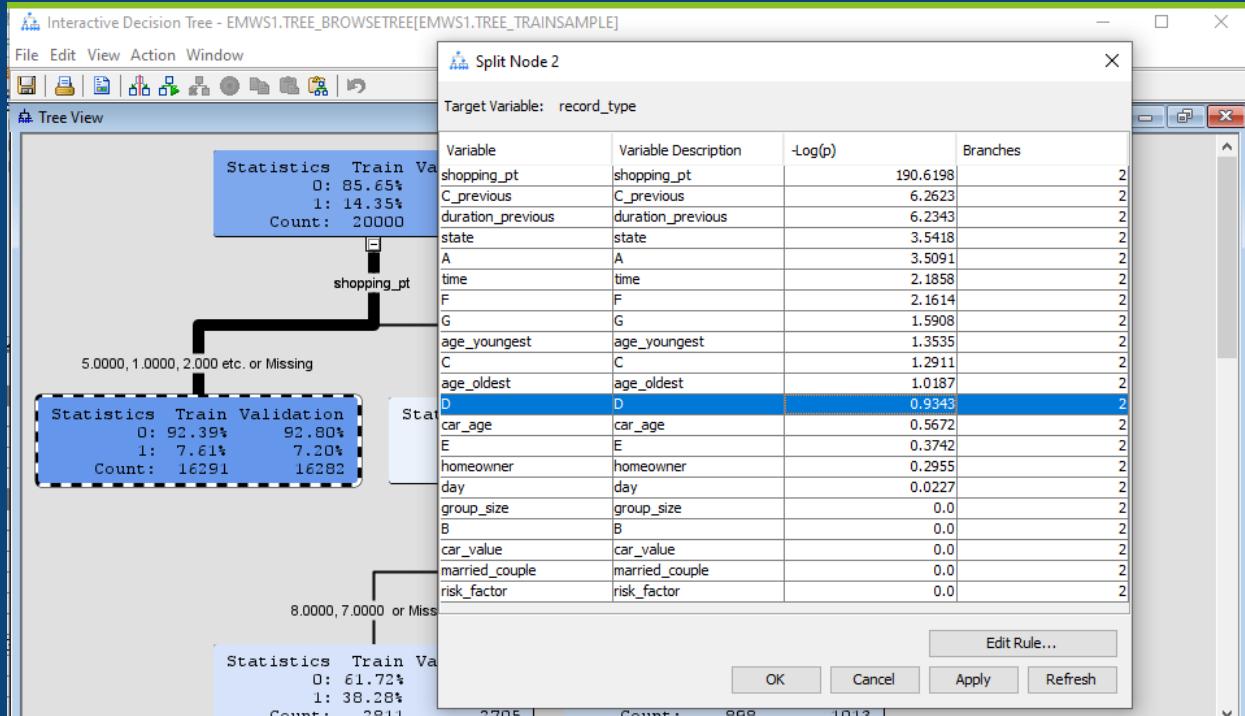
The tree view:

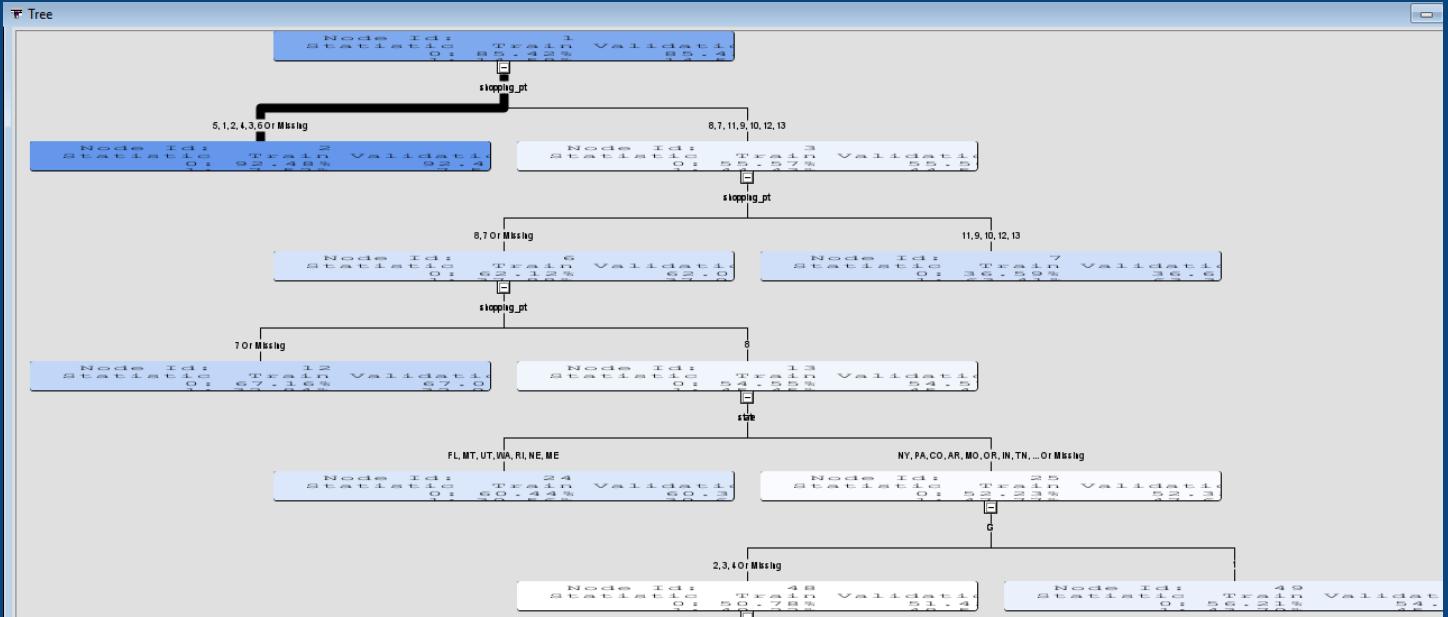
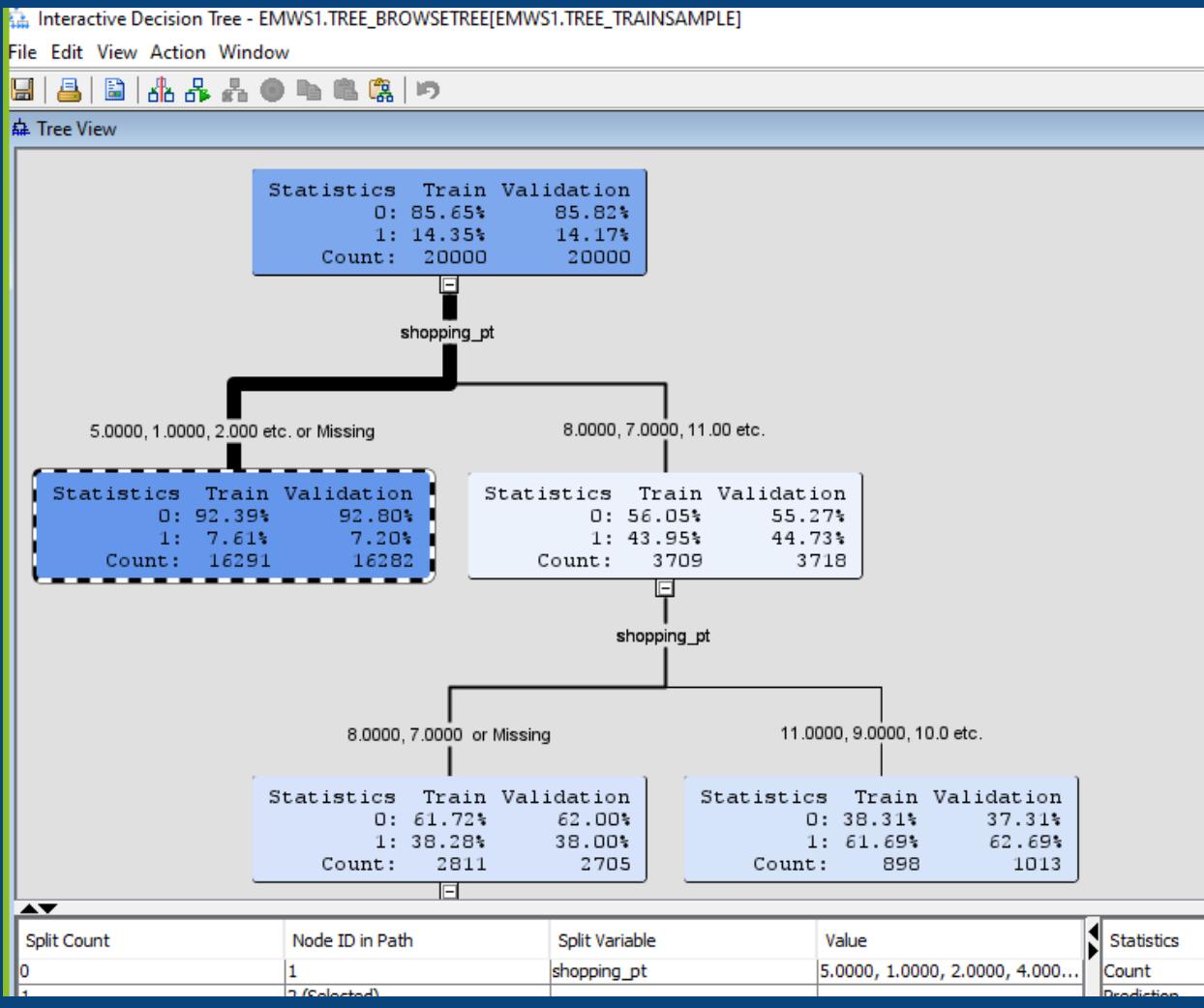
Tree Print





Then splitting the node:





Misclassification rate - validation (0.132458)

Results - Node: Decision Tree - Interactive Diagram: Allstate

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type	_NOBS_		Sum of Frequencies	332624	332625	
record_type	_MISC_		Misclassification Rate	0.132221	0.132458	
record_type	_MAX_		Maximum Absolute Error	0.924817	0.924817	
record_type	_SSE_		Sum of Squared Errors	66814.18	66826.17	
record_type	_ASE_		Average Squared Error	0.100435	0.100453	
record_type	_RASE_		Root Average Squared Error	0.316915	0.316943	
record_type	_DIV_		Divisor for ASE	665248	665250	
record_type	_DFT_		Total Degrees of Freedom	332624		

Average squared error

Results - Node: Decision Tree - Interactive Diagram: Allstate

File Edit View Window

Fit Statistics

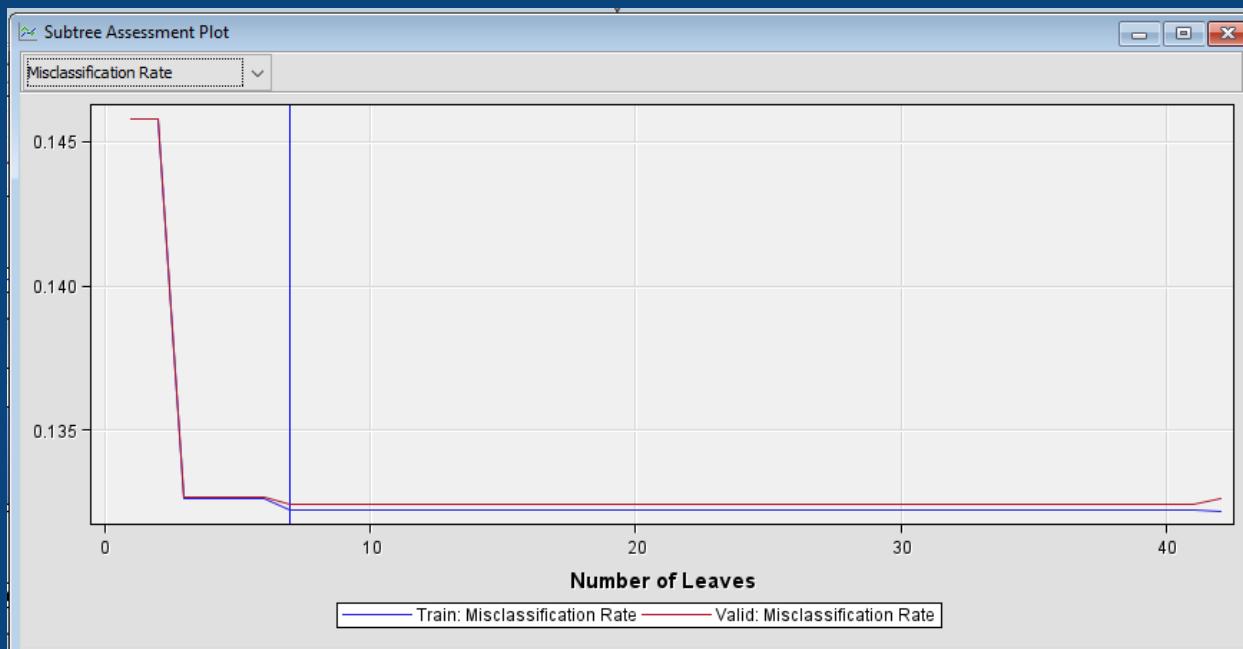
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type	_NOBS_		Sum of Frequencies	332624	332625	
record_type	_MISC_		Misclassification Rate	0.132221	0.132458	
record_type	_MAX_		Maximum Absolute Error	0.924817	0.924817	
record_type	_SSE_		Sum of Squared Errors	66814.18	66826.17	
record_type	<u>_ASE_</u>		Average Squared Error	0.100435	0.100453	
record_type	<u>_RASE_</u>		Root Average Squared Error	0.316915	0.316943	
record_type	_DIV_		Divisor for ASE	665248	665250	
record_type	_DFT_		Total Degrees of Freedom	332624		

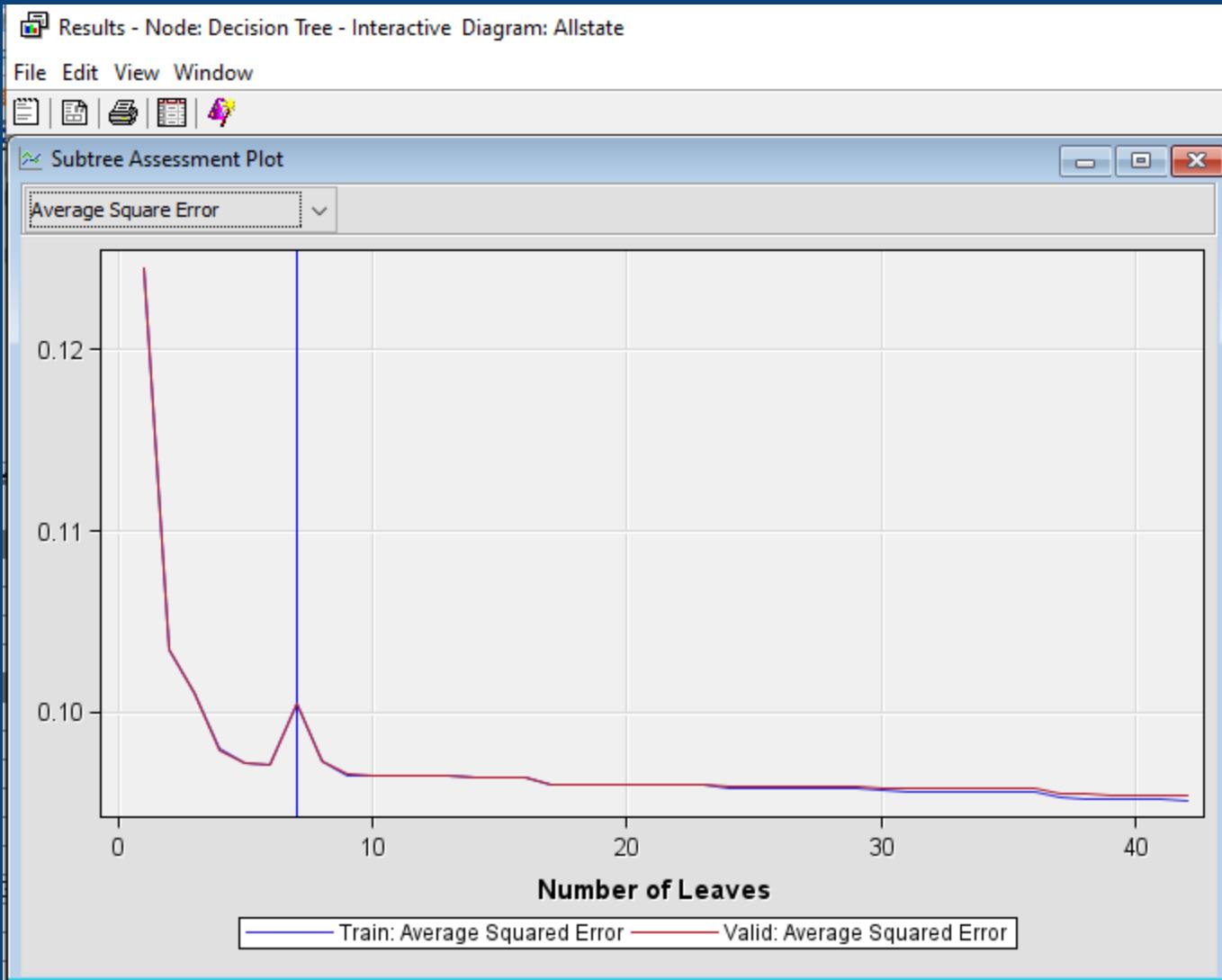
Assessment measure: Decision

Property	Value
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Import	
Observation Based Import No.	

Subtree assessment plot showing number of trees:

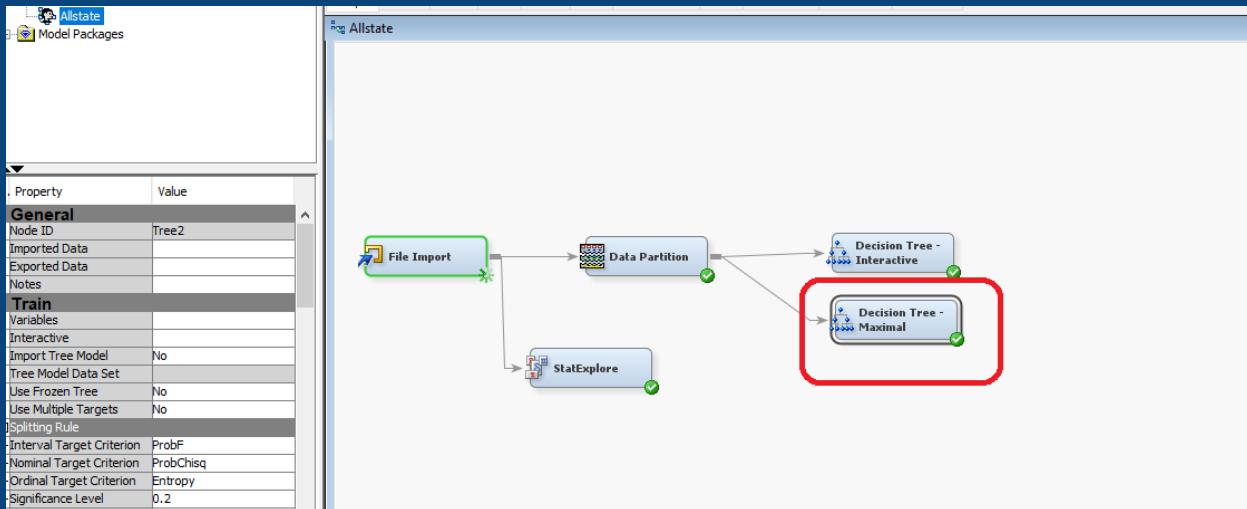




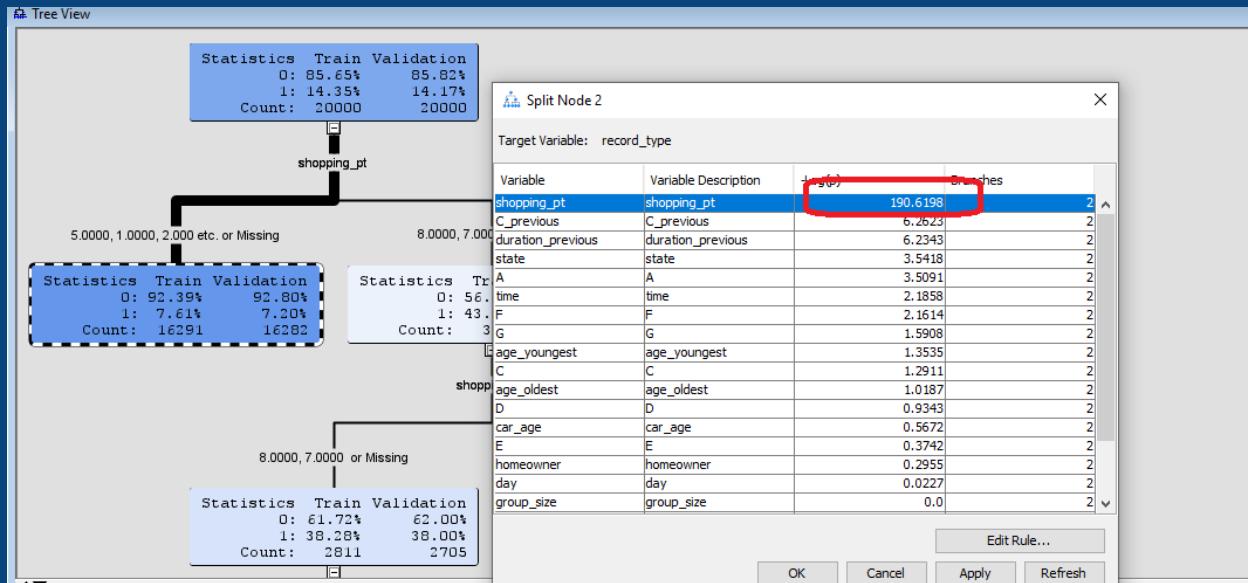


Assessment measure: Average square error

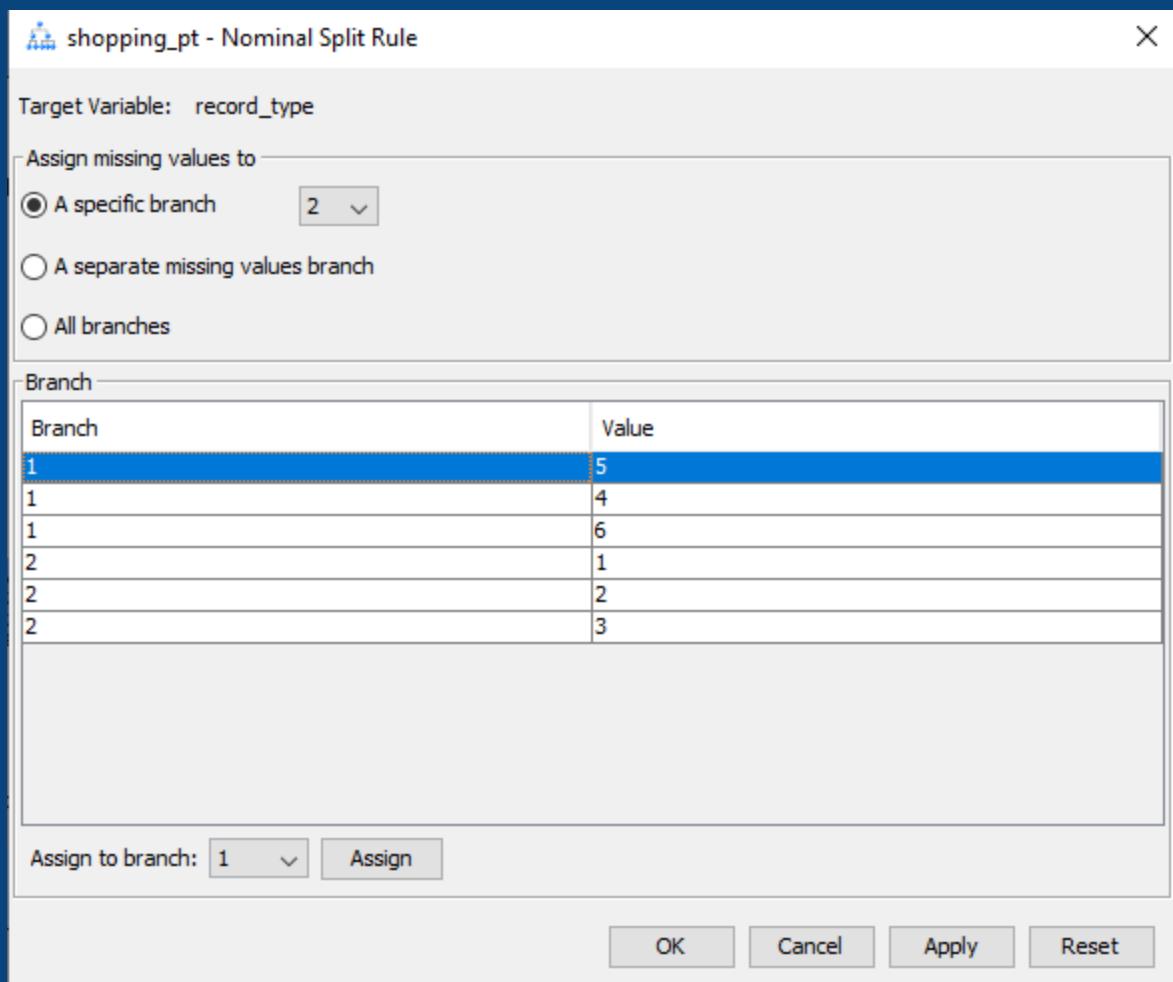
Creating the Maximal Tree



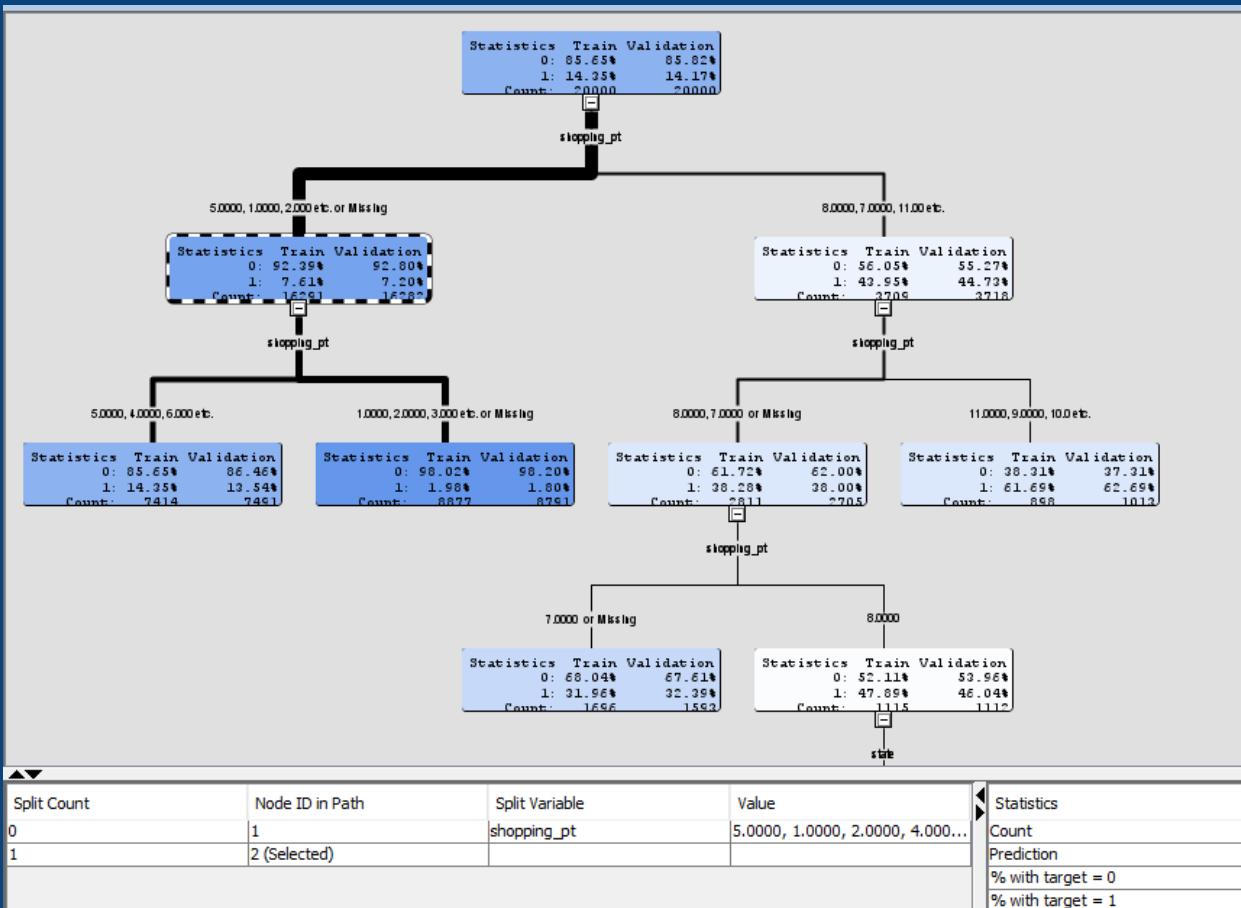
Highest logworth of maximal tree split node is shopping_pt, followed by C_previous and duration_previous



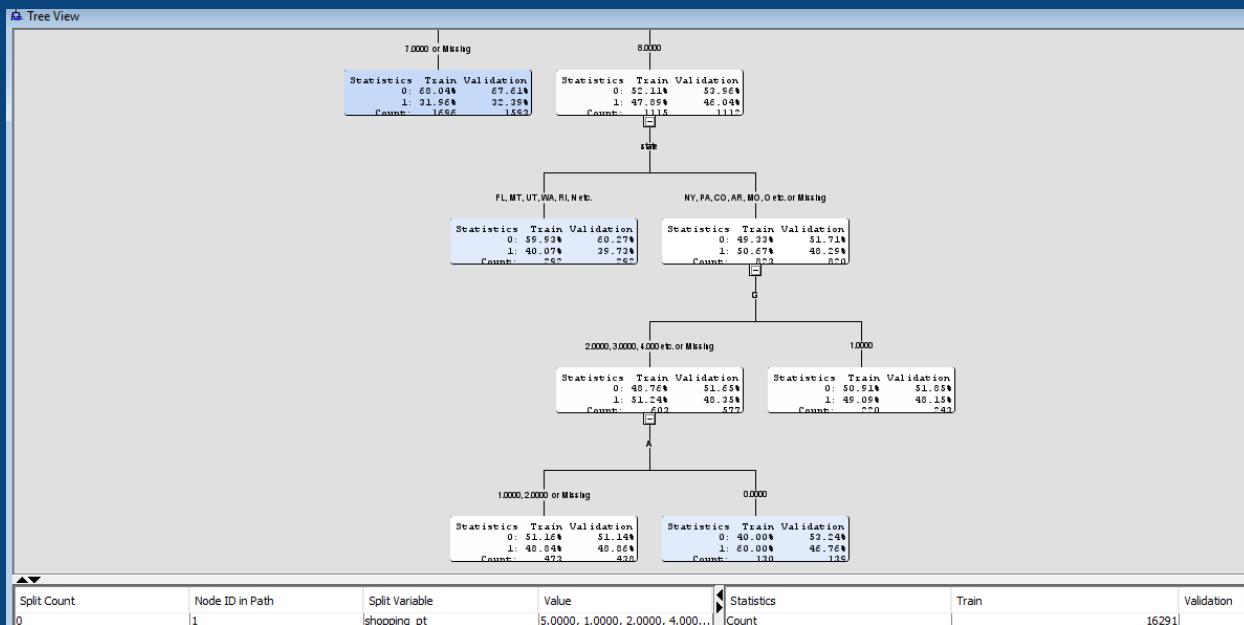
Shopping-pt Interval Splitting Rule dialog box which shows how the training data is partitioned using the input shopping_pt. Six branches are created.



New tree view, showing additional split:

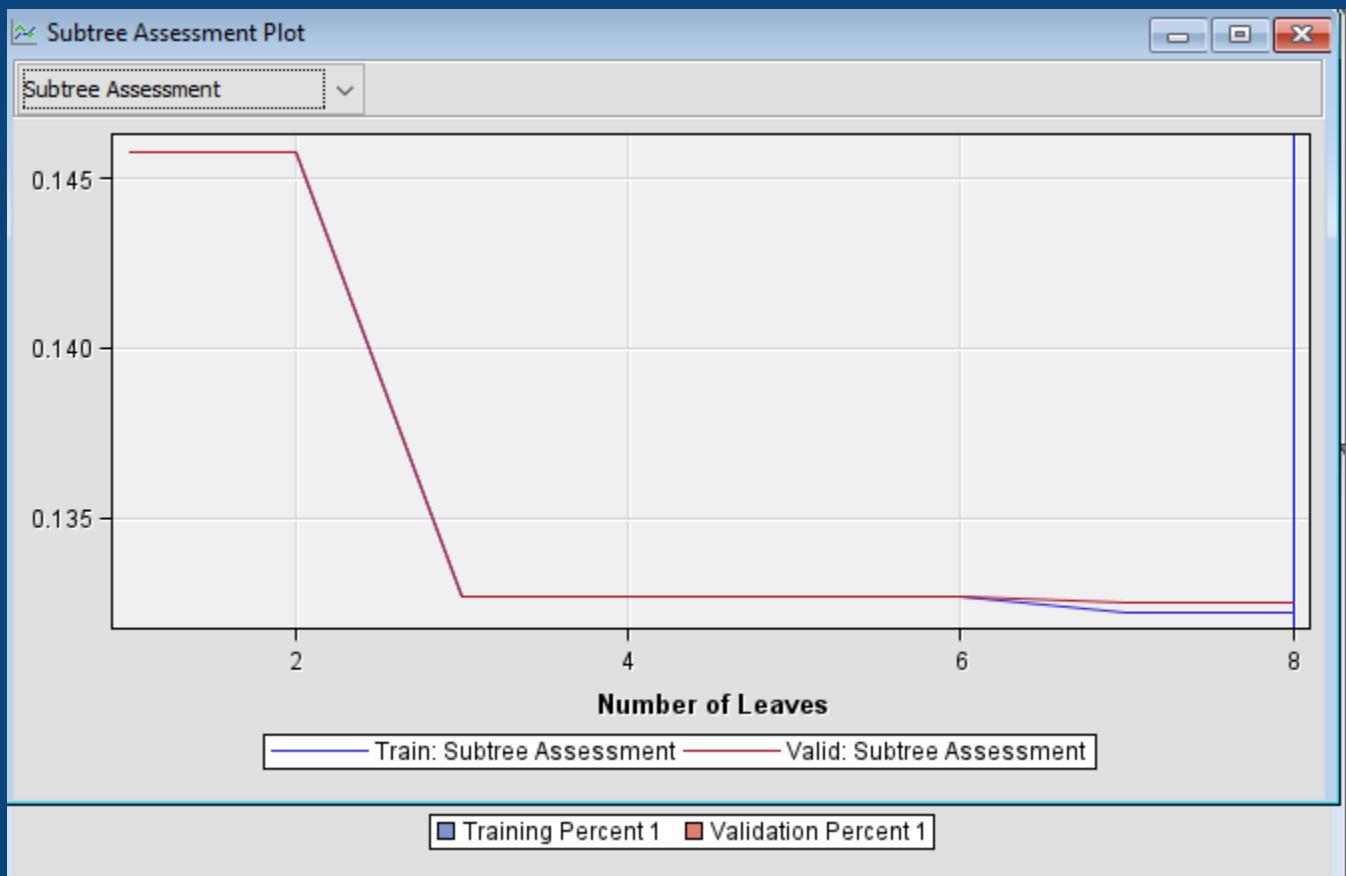


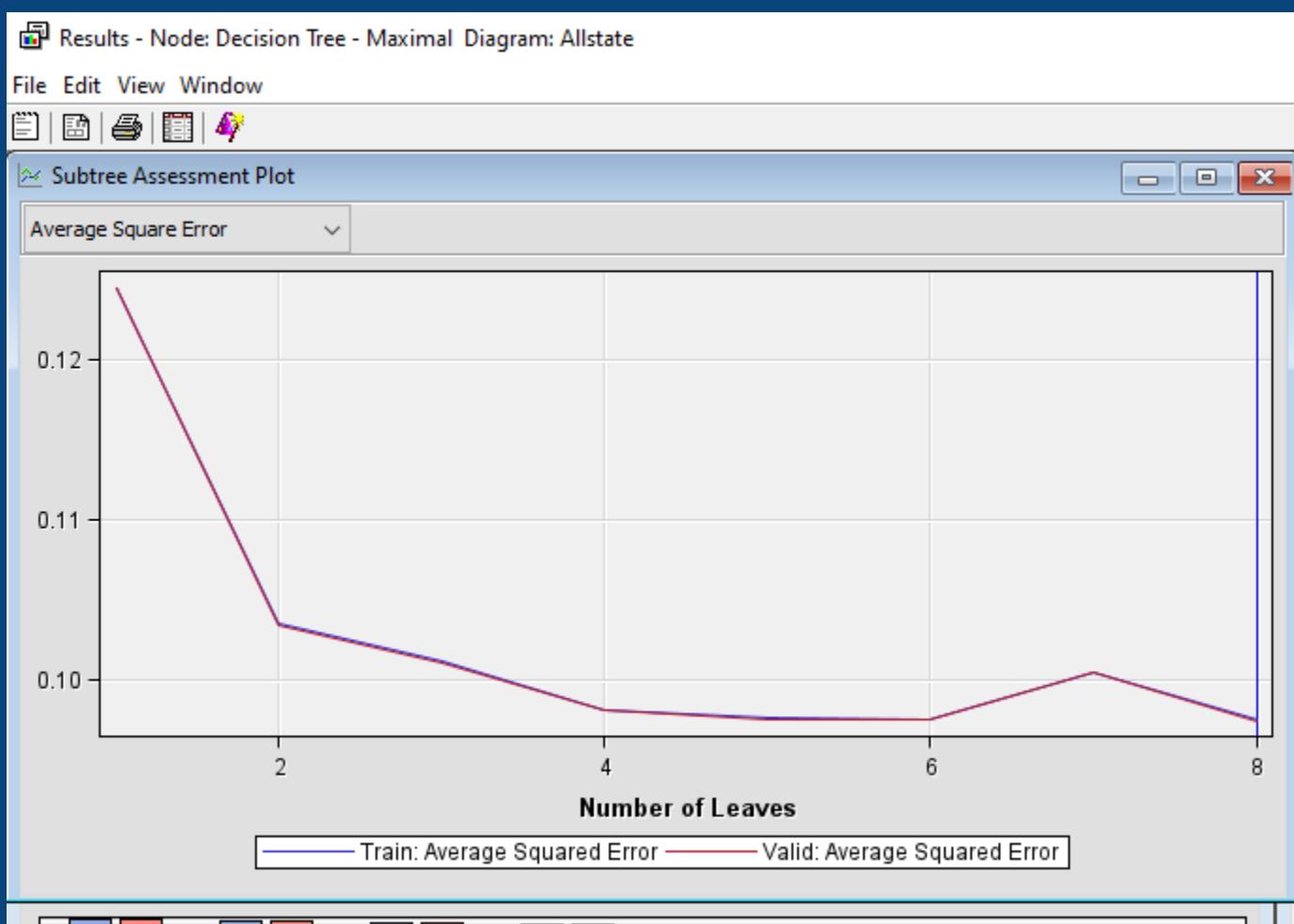
Split Count	Node ID in Path	Split Variable	Value	Statistics
0	1	shopping_pt	5.0000, 1.0000, 2.0000, 4.000...	Count
1	2 (Selected)			Prediction % with target = 0 % with target = 1



Subtree assessment plot:
Assessment measure = Decision

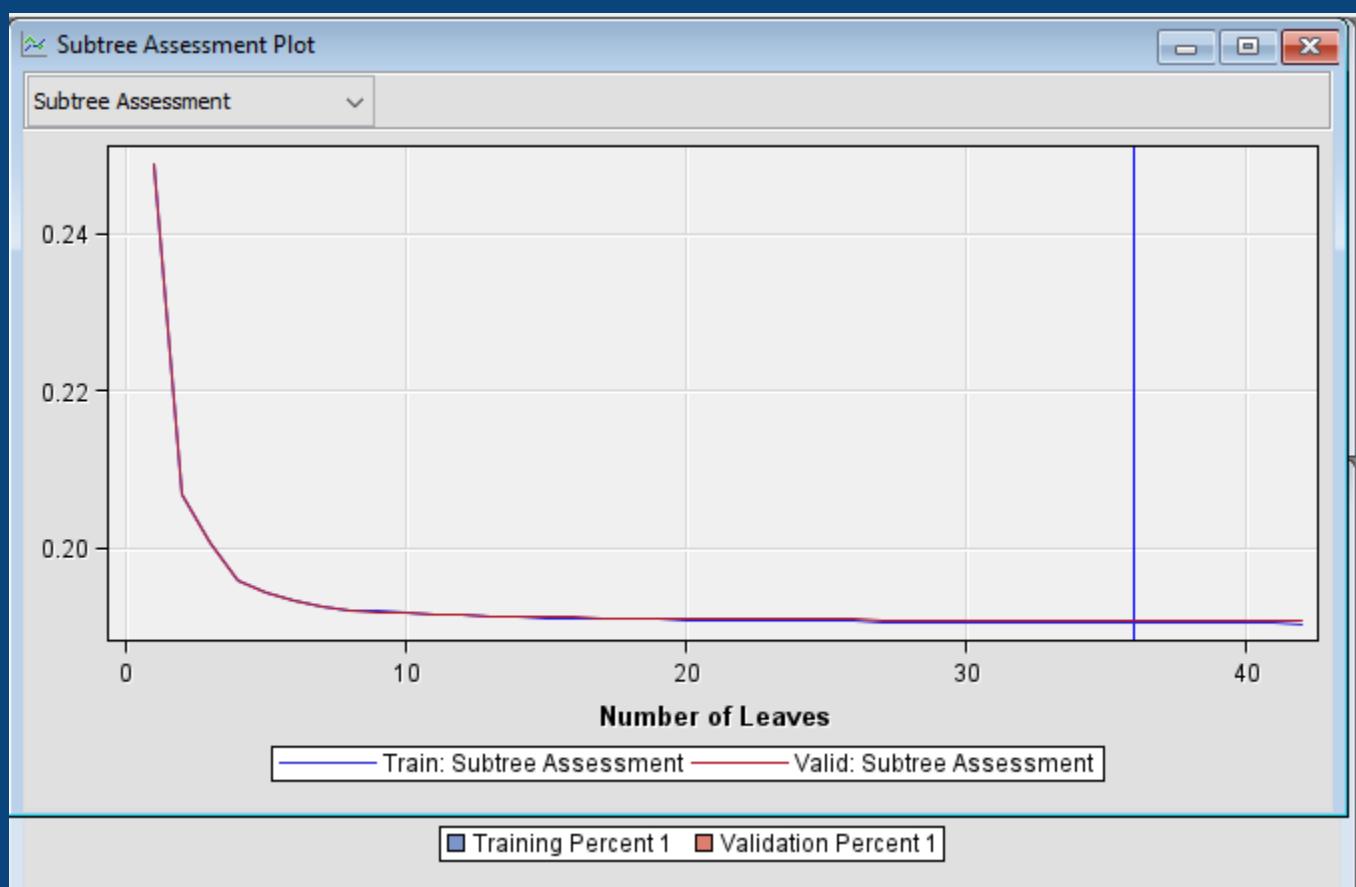
Property	Value
--Leaf Size	5
--Number of Rules	5
--Number of Surrogate Rules	0
--Split Size	.
Split Search	
--Use Decisions	No
--Use Priors	No
--Exhaustive	5000
--Node Sample	20000
Subtree	
--Method	Assessment
--Number of Leaves	1
--Assessment Measure	Decision
--Assessment Fraction	0.25
Cross Validation	
--Perform Cross Validation	No
--Number of Subsets	10
--Number of Repeats	1
--Seed	12345
Observation Based Import	

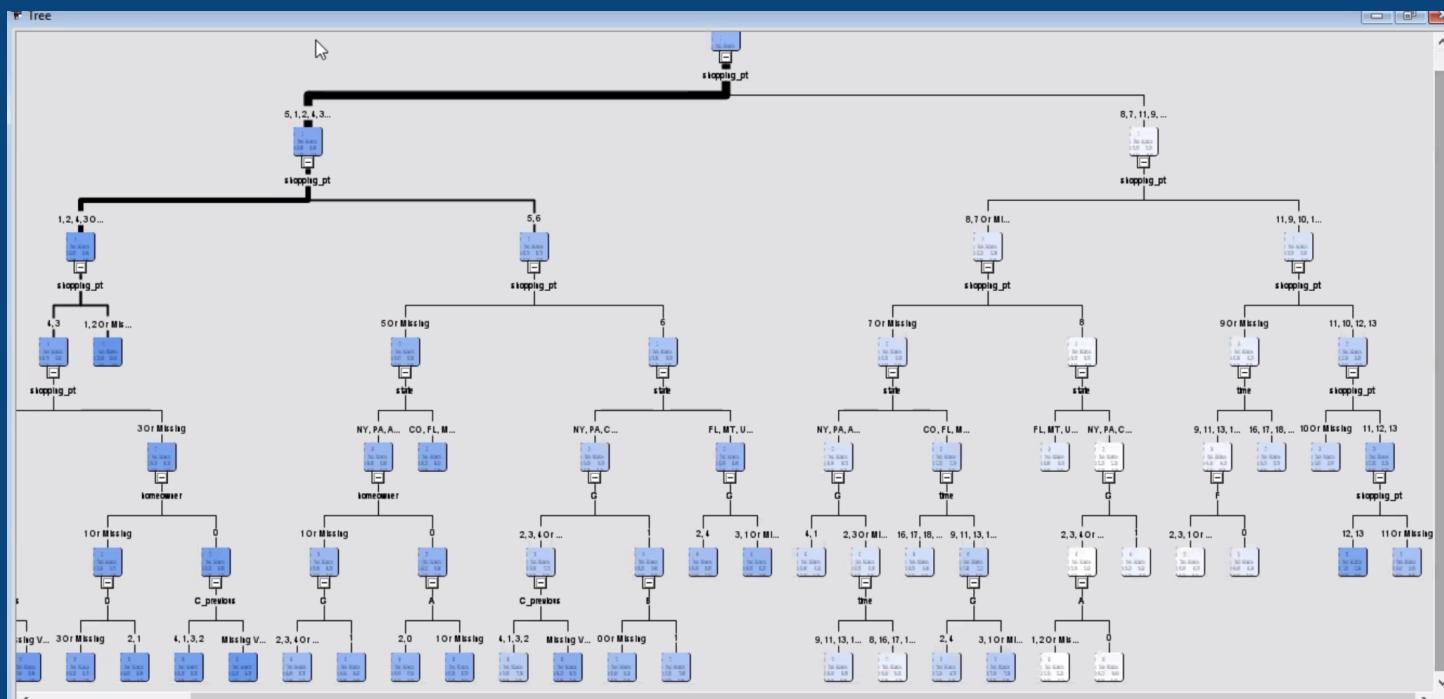
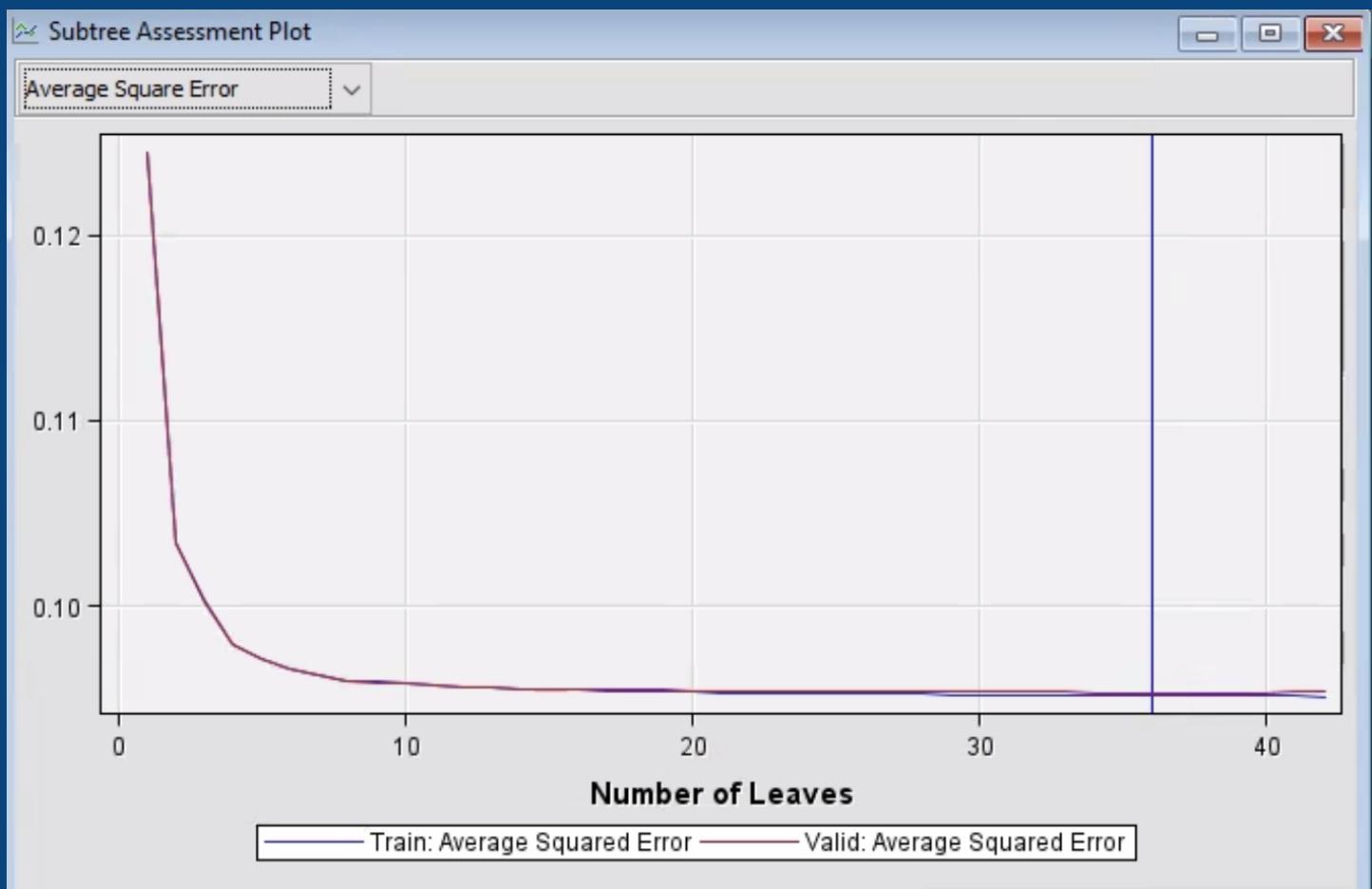




Assessment measure = Average squared error

Property	Value
-Ecorr Size	5
-Number of Rules	5
-Number of Surrogate Rules	0
-Split Size	.
] Split Search	
-Use Decisions	No
-Use Priors	No
-Exhaustive	5000
-Node Sample	20000
] Subtree	
-Method	Assessment
-Number of Leaves	1
-Assessment Measure	Average Square Error
-Assessment Fraction	0.25
] Cross Validation	
-Perform Cross Validation	No
-Number of Subsets	10
-Number of Repeats	1
] Seed	12345



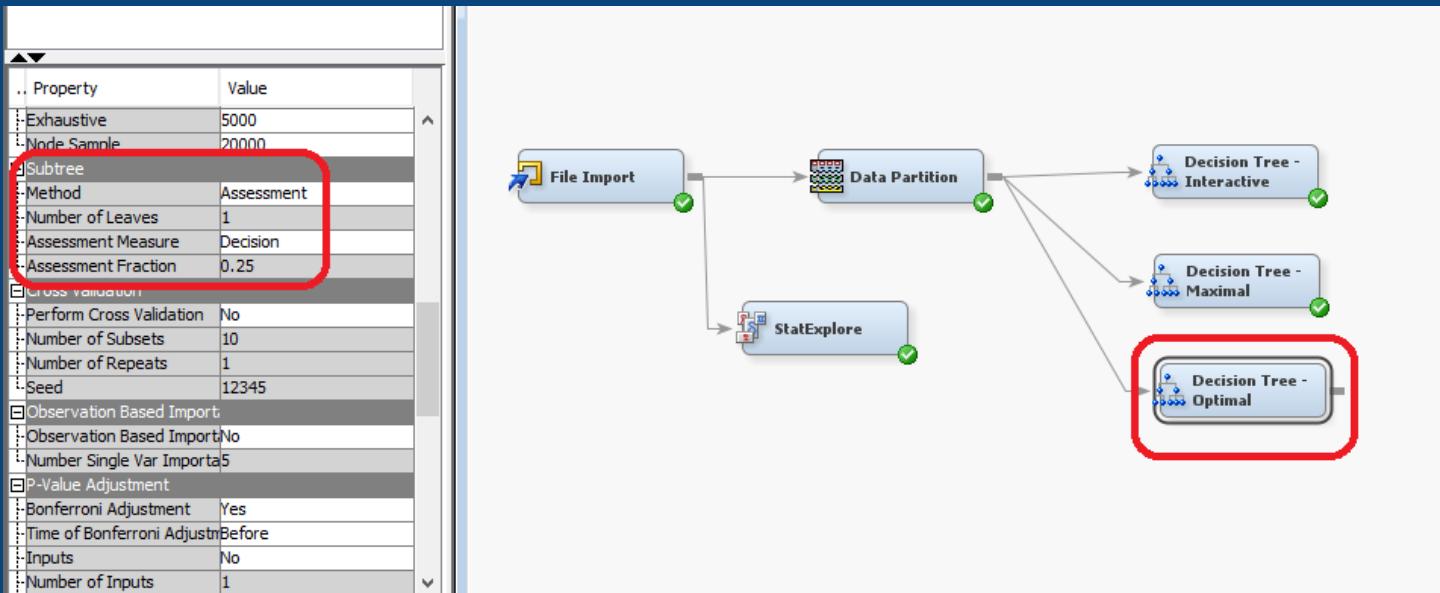


Misclassification rule and average squared error of the maximal tree:

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type	_NOBS_	Sum of Frequencies		332624	332625	
record_type	_MISC_	Misclassification Rate		0.132221	0.132458	
record_type	_MAX_	Maximum Absolute Error		0.980019	0.980619	
record_type	_SSE_	Sum of Squared Errors		64835.61	64820.28	
record_type	_ASE_	Average Squared Error		0.097461	0.097437	
record_type	_RASE_	Root Average Squared Error		0.312187	0.31215	
record_type	_DIV_	Divisor for ASE		665248	665250	
record_type	_DFT_	Total Degrees of Freedom		332624		

Optimal Tree

Creating optimal tree:



Misclassification rule:

Results - Node: Decision Tree - Optimal Diagram: Allstate						
File Edit View Window						
Fit Statistics						
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type	_NOBS_	Sum of Frequencies		332624	332625	
record_type	_MISC_	Misclassification Rate		0.132221	0.132458	
record_type	_MAX_	Maximum Absolute Error		0.924817	0.924817	
record_type	_SSE_	Sum of Squared Errors		66814.18	66826.17	
record_type	_ASE_	Average Squared Error		0.100435	0.100453	
record_type	_RASE_	Root Average Squared Error		0.316915	0.316943	
record_type	_DIV_	Divisor for ASE		665248	665250	
record_type	_DFT_	Total Degrees of Freedom		332624		

Assessment measure: Decision

Average squared error

Results - Node: Decision Tree - Optimal Diagram: Allstate

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type	_NOBS_		Sum of Frequencies	332624	332625	
record_type	_MISC_		Misclassification Rate	0.132221	0.132458	
record_type	_MAX_		Maximum Absolute Error	0.924817	0.924817	
record_type	_SSE_		Sum of Squared Errors	66044.40	66026.17	
record_type	_ASE_		Average Squared Error	0.100435	0.100453	
record_type	_RASE_		Root Average Squared Error	0.310919	0.310943	
record_type	_DIV_		Divisor for ASE	665248	665250	
record_type	_DFT_		Total Degrees of Freedom	332624		

Subtree assessment measure for optimal tree:

Assessment measure = Decision

.. Property Value

- ECT Size
- Number of Rules 5
- Number of Surrogate Rules 0
- Split Size .
- Split Search

 - Use Decisions No
 - Use Priors No
 - Exhaustive 5000
 - Node Sample 20000

- Subtree

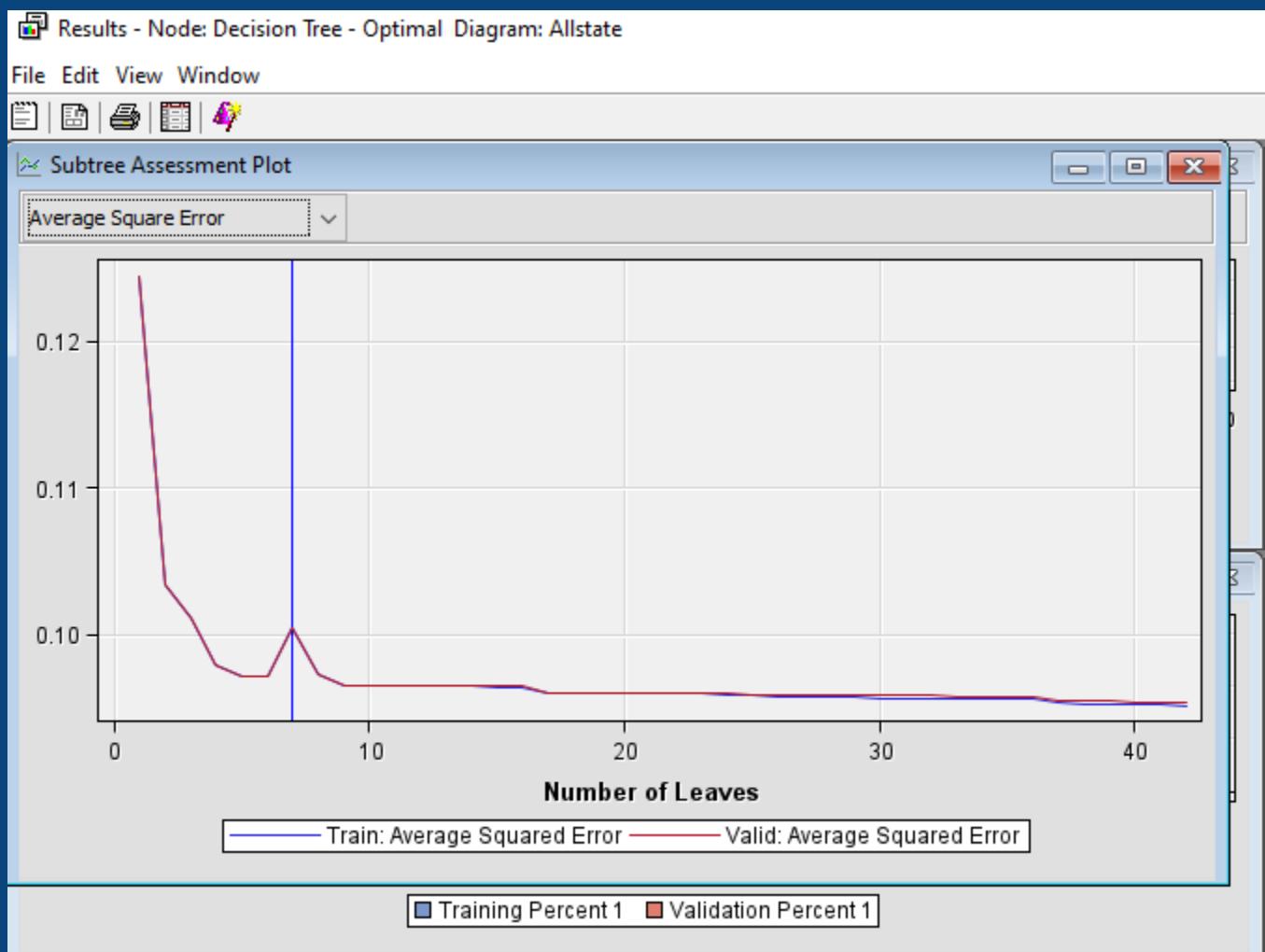
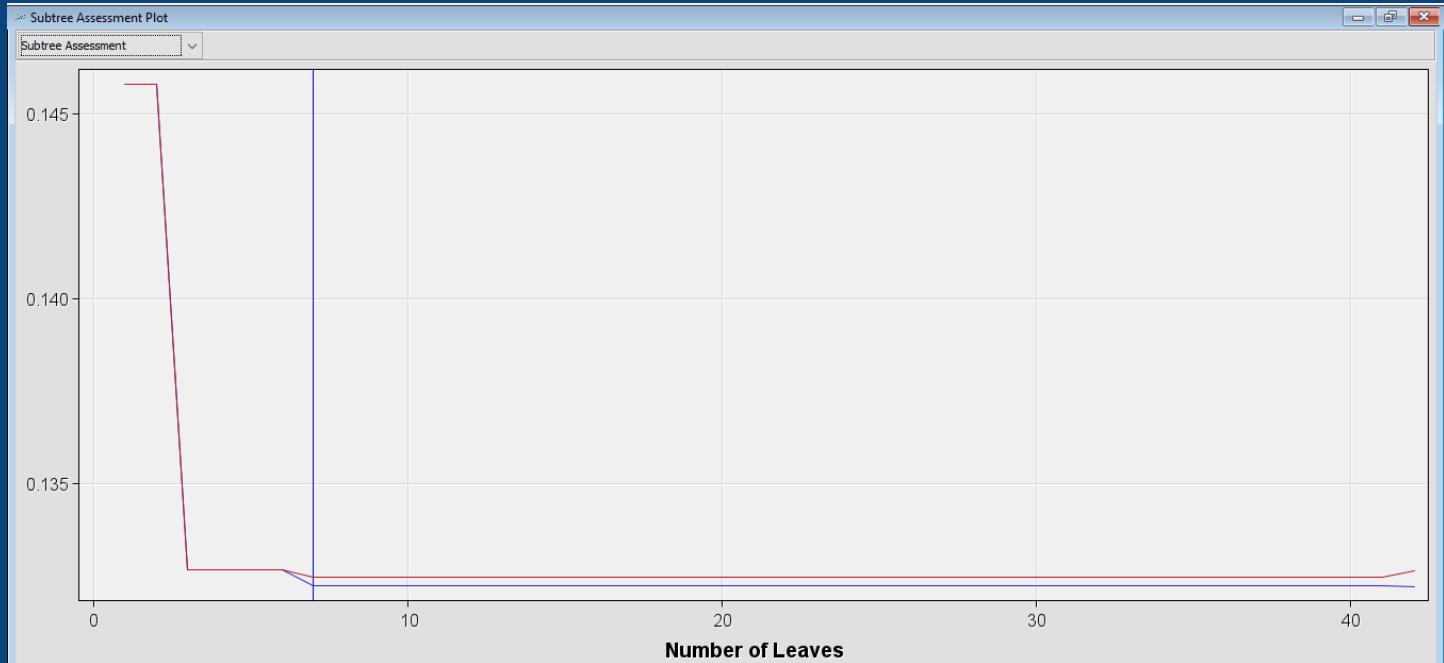
 - Method Assessment
 - Number of Leaves 1
 - Assessment Measure Decision **Red Box**
 - Assessment Fraction 0.25

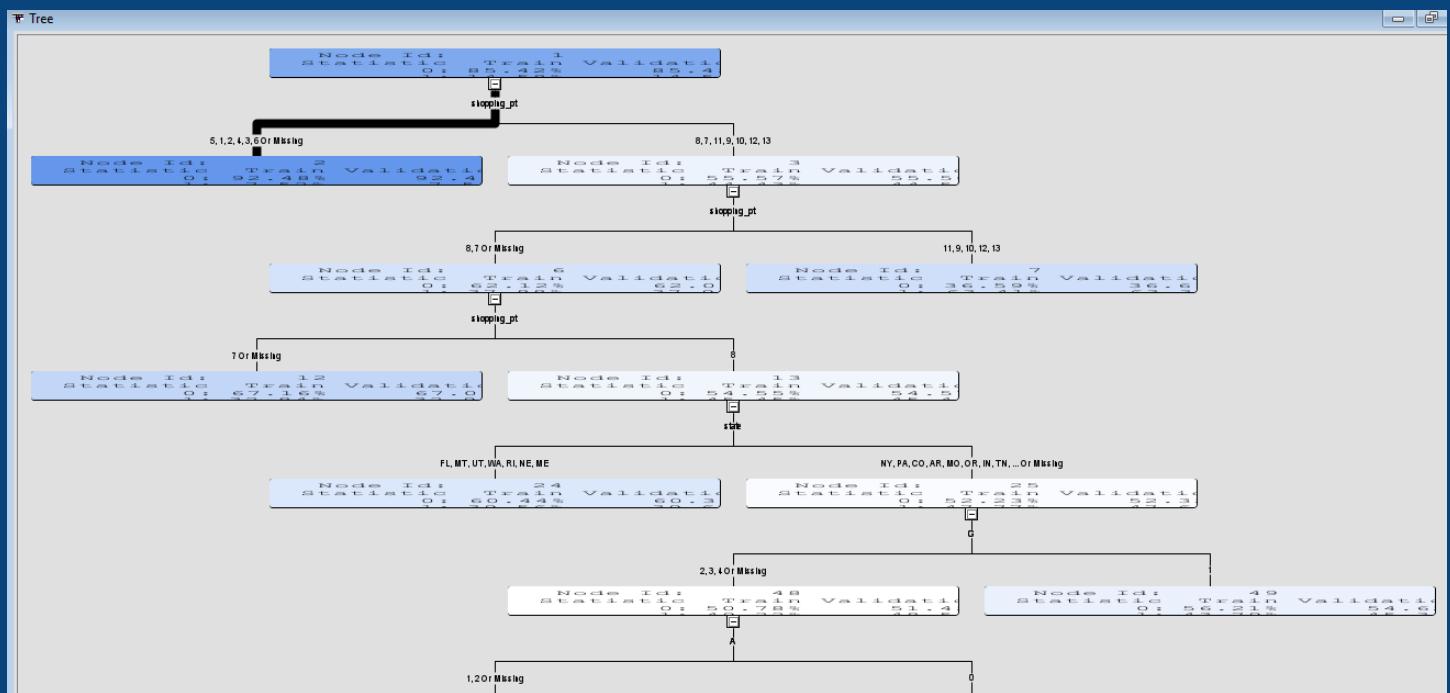
- Cross Validation

 - Perform Cross Validation No
 - Number of Subsets 10
 - Number of Repeats 1
 - Seed 12345

- Observation Based Import

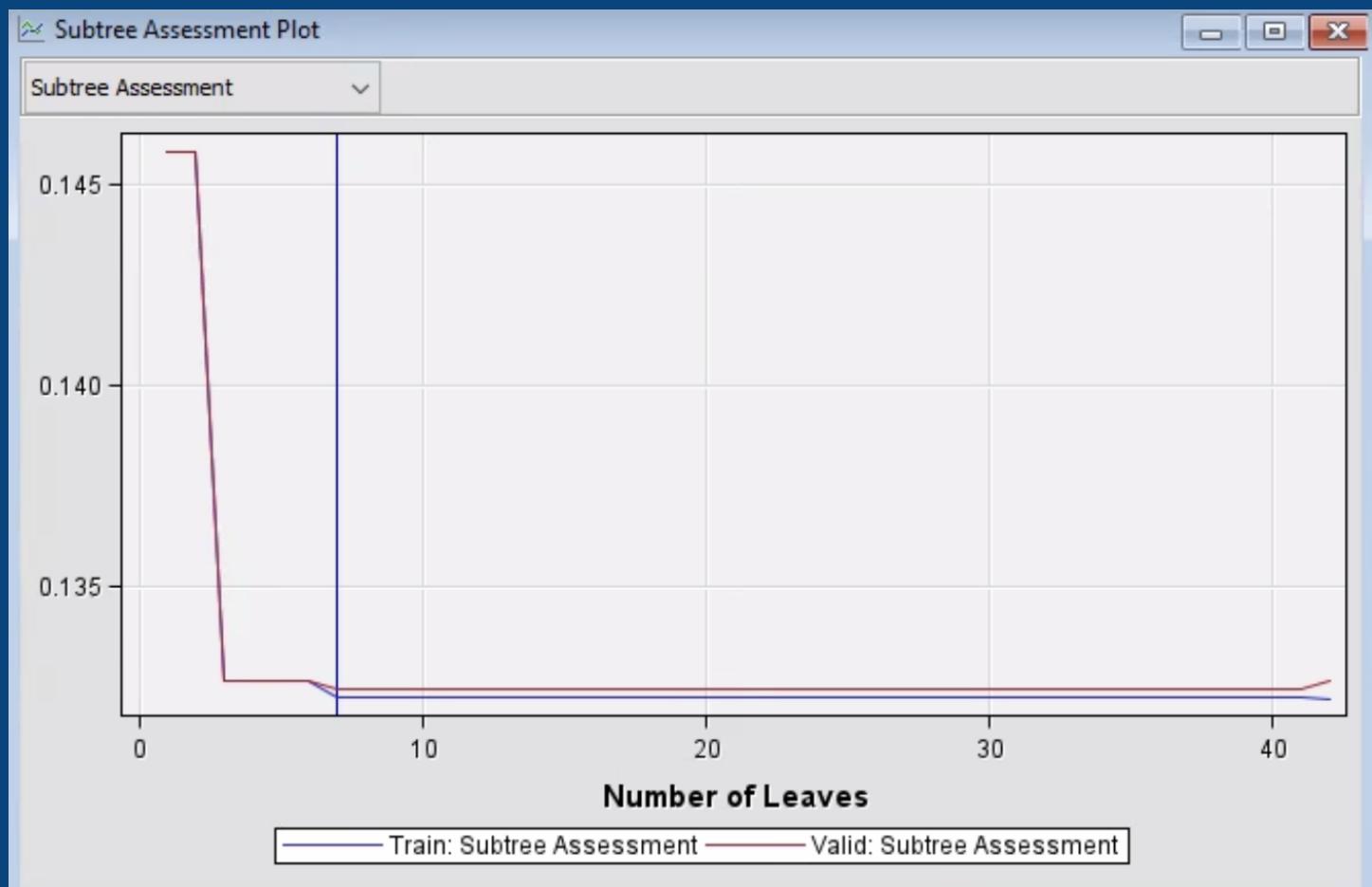
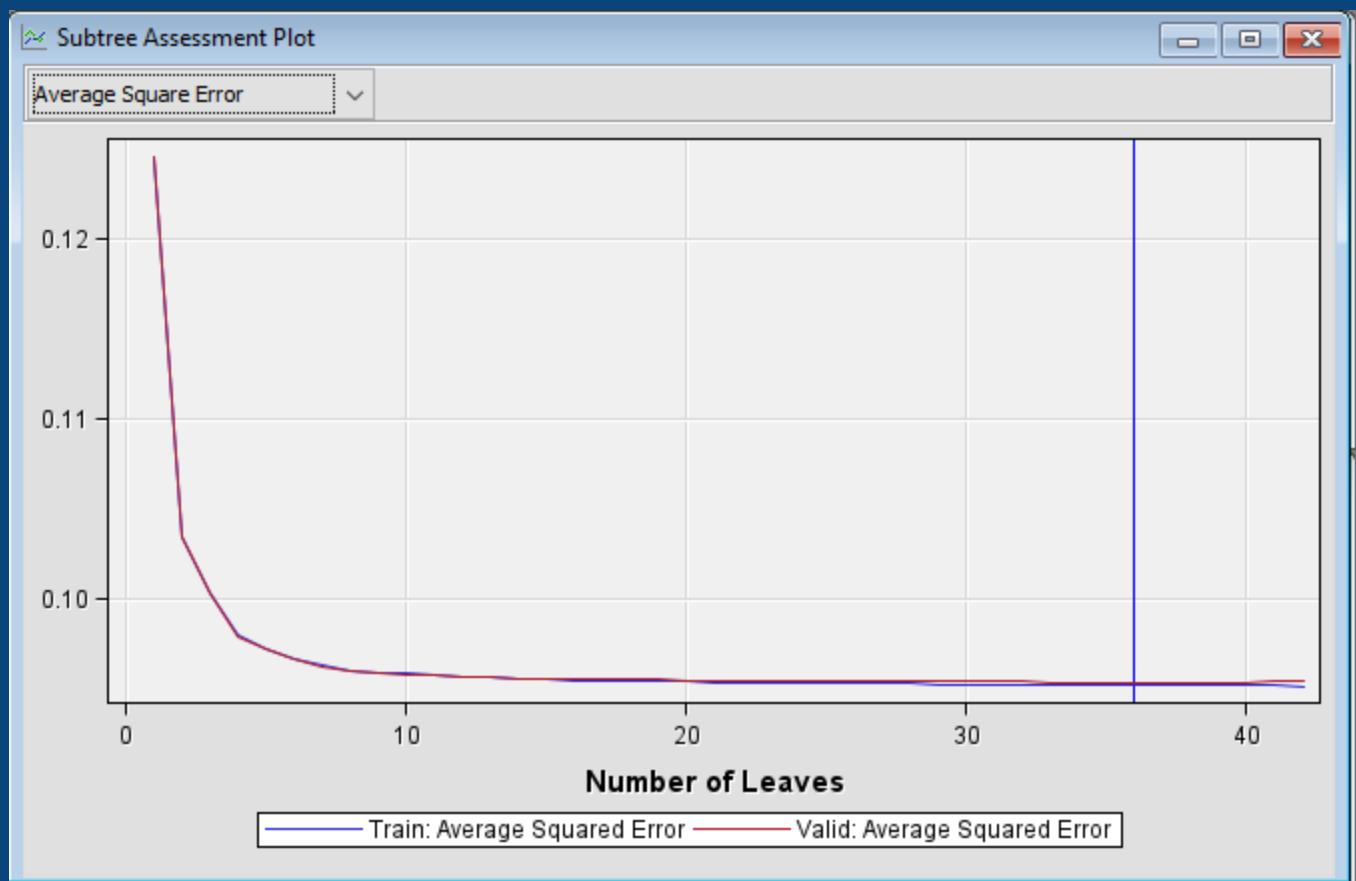
 - Observation Based Import No

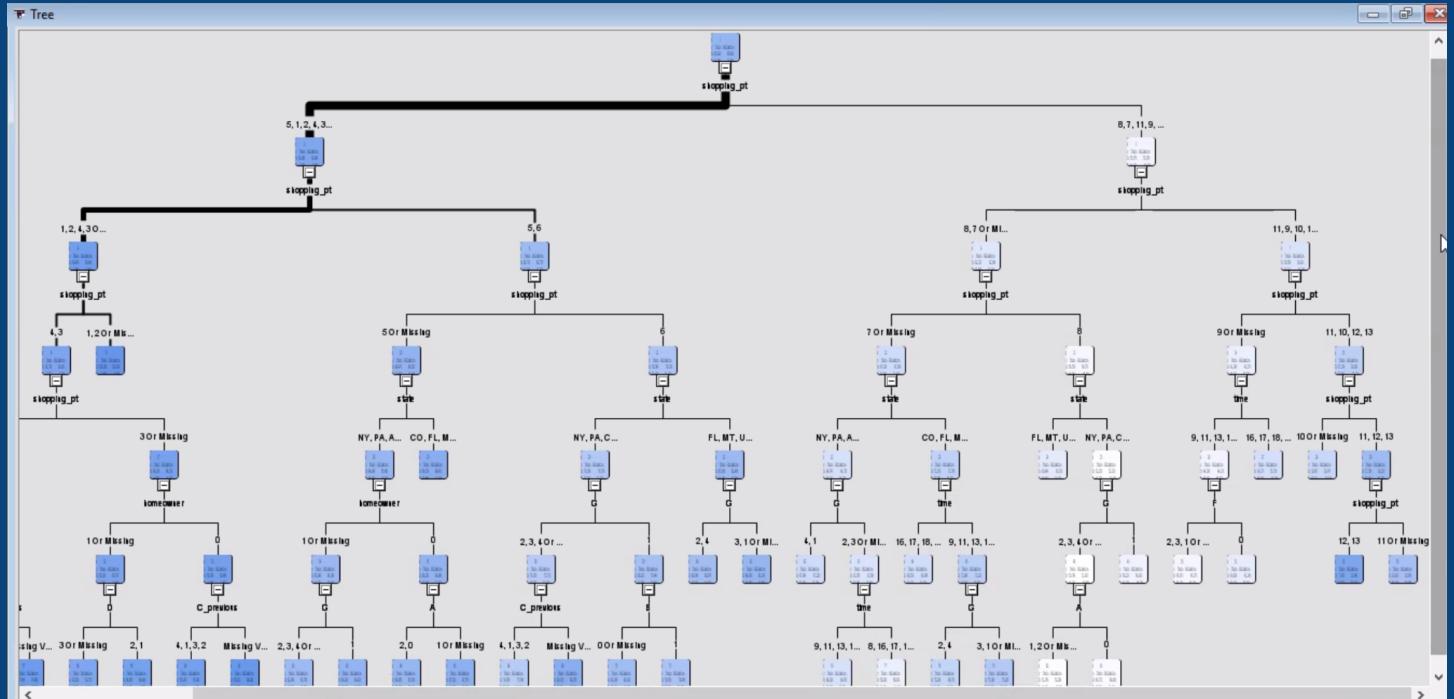




Assessment measure = Average squared error

.. Property	Value
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
<input checked="" type="checkbox"/> Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
<input checked="" type="checkbox"/> Subtree	
Method	Assessment
Number of Leaves	1
<input checked="" type="checkbox"/> Assessment Measure	Average Square Error
<input checked="" type="checkbox"/> Assessment Fraction	0.25
<input checked="" type="checkbox"/> Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
<input checked="" type="checkbox"/> Observation Based Import	
Observation Based Tree	





Comparison of the decision trees:

Decision Tree	Misclassification rule (validation)	ASE (validation)
Interactive	0.132458	0.100453
Maximal	0.132458	0.097437
Optimal	0.132458	0.100453

Regression Modeling

Firstly, let's check if there are missing values - we know we have some missings:

- C_Previou - 5124,
- car_value - 600,

- duration_previous - 5124.

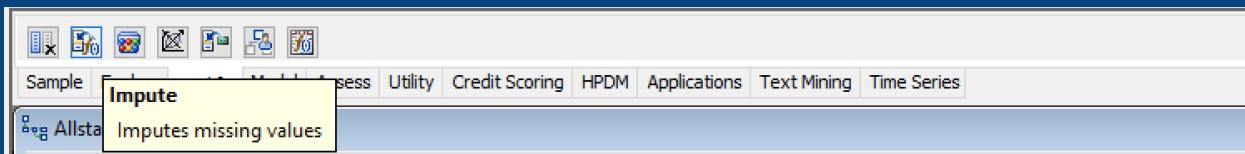
We also can estimate the percentage of missing by clicking on Data Partition Node -> Exported Data -> Train Data -> Explore:

Obs #	Variable ...	Label	Type	Percent Missing	Minimum	Maximum	Mean	Number o...	Mode Per...	Mode
1	car_value		CLASS	0.275	.	.	.10	32.905E	.	
2	risk_factor		CLASS	0	.	.	.5	35.65NA	.	
3	state		CLASS	0	.	.	.36	15.88FL	.	
4	A		VAR	0	0	2	0.92175	.	.	
5	B		VAR	0	0	1	0.45575	.	.	
6	C		VAR	0	1	4	2.27745	.	.	
7	C_previous		VAR	2.97	1	4	2.436721	.	.	
8	D		VAR	0	1	3	2.4558	.	.	
9	E		VAR	0	0	1	0.44225	.	.	
10	F		VAR	0	0	3	1.15775	.	.	
11	G		VAR	0	1	4	2.26935	.	.	
12	_dataobs_	Observatio...	VAR	0	49	665245	333677.6	.	.	
13	age_oldest		VAR	0	18	75	45.04235	.	.	
14	age_young...		VAR	0	16	75	42.68755	.	.	
15	car_age		VAR	0	0	64	8.12085	.	.	
16	cost		VAR	0	264	857	634.9224	.	.	
17	customer_ID		VAR	0	10000016	10152702	10076335	.	.	
18	day		VAR	0	0	6	1.96245	.	.	
19	duration_pr...		VAR	2.97	0	15	5.966557	.	.	
20	group_size		VAR	0	1	4	1.23455	.	.	
21	homeowner		VAR	0	0	1	0.5395	.	.	

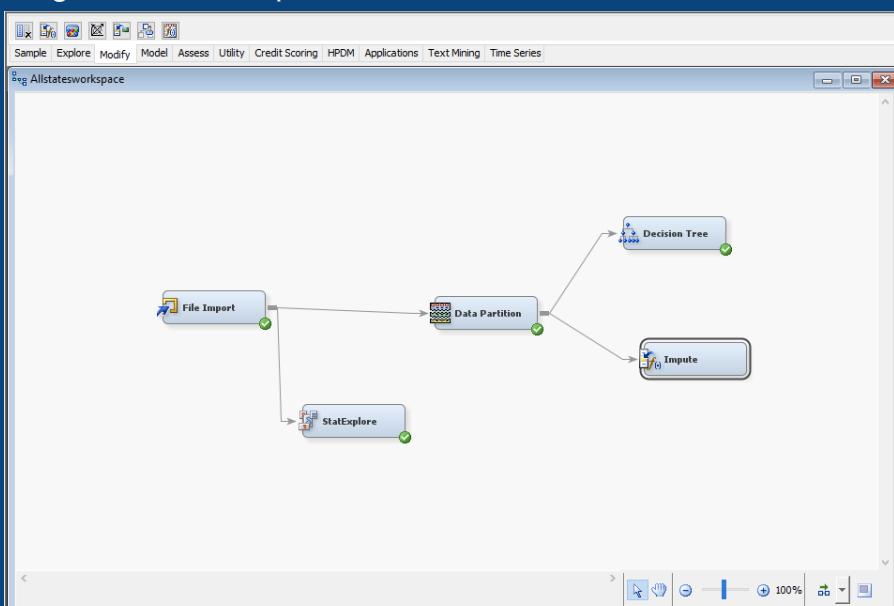
Since missing values bring severe issues for regression models, we will utilize synthetic estimation method to fill our missings with mean value for interval variables and mod value for categorical values. Then we will create missing indicator to flag all imputed values.

Firstly, let's impute synthetic values into the missings:

Modify Tab -> Impute Tool



Drag it to the workspace and connect to the Data Partition Node



Property Panel of Impute Node ->

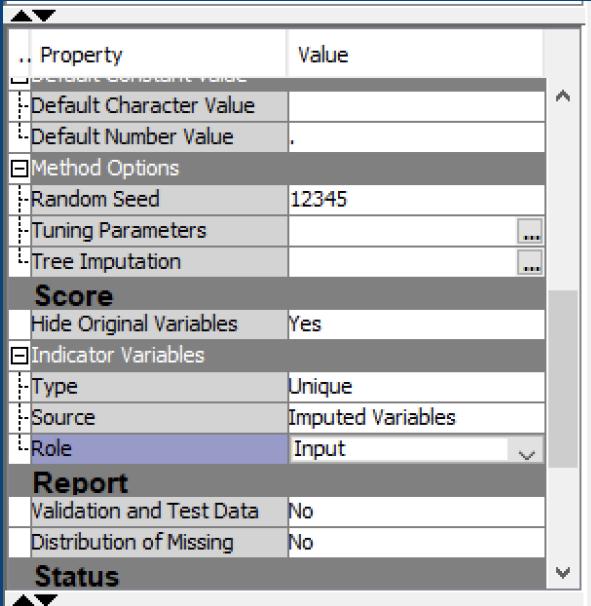
.. Property	Value
Exported Data	
Notes	
Train	
Variables	
Nonmissing Variables	No
Missing Cutoff	50.0
Class Variables	
Default Input Method	Count
Default Target Method	None
Normalize Values	Yes
Interval Variables	
Default Input Method	Mean
Default Target Method	None
Default Constant Value	
Default Character Value	
Default Number Value	.

Make sure we have such properties as on the screenshot (so that we replace numeric missings with the mean value and categorical missings with the mod value).

Now we need to set the missing indicator. We go for that to the Score Section of the Property Panel.

.. Property	Value
Percent Constant Value	
Default Character Value	
Default Number Value	.
Method Options	
Random Seed	12345
Tuning Parameters	
Tree Imputation	
Score	
Hide Original Variables	Yes
Indicator Variables	
Type	None
Source	Imputed Variables
Role	Rejected
Report	
Validation and Test Data	No
Distribution of Missing	No
Status	

We need to select 'Unique' for Type and 'Input' for Role in Indicator Variable so that we have these properties:



At this point we run the Impute Node and can take a look at the results:

Variable Name	Impute Method	Imputed Variable	Indicator Variable	Impute Value	Role	Measurement Level	Label	Number of Missing for TRAIN
C_previous	COUNT	IMP_C_previous	M_C_previous	3	INPUT	NOMINAL		9312
car_value	COUNT	IMP_car_value	M_car_value	8	INPUT	NOMINAL		774
duration_previous	MEAN	IMP_duration_previous	M_duration_previous	6.0073953333	INPUT	INTERVAL		9312

```

34
35  Imputation Summary
36  Number Of Observations
37
38
39      Impute
40  Variable Name    Method   Imputed Variable     Indicator Variable   Impute Value   Role   Measurement Level   Label   Number of Missing for TRAIN
41
42  C_previous       COUNT    IMP_C_previous      M_C_previous        3           INPUT  NOMINAL          9312
43  car_value         COUNT    IMP_car_value       M_car_value        8           INPUT  NOMINAL          774
44  duration_previous MEAN     IMP_duration_previous M_duration_previous 6.0073953333  INPUT  INTERVAL          9312
45
46
47
48
49  Variable Distribution Training Data
50

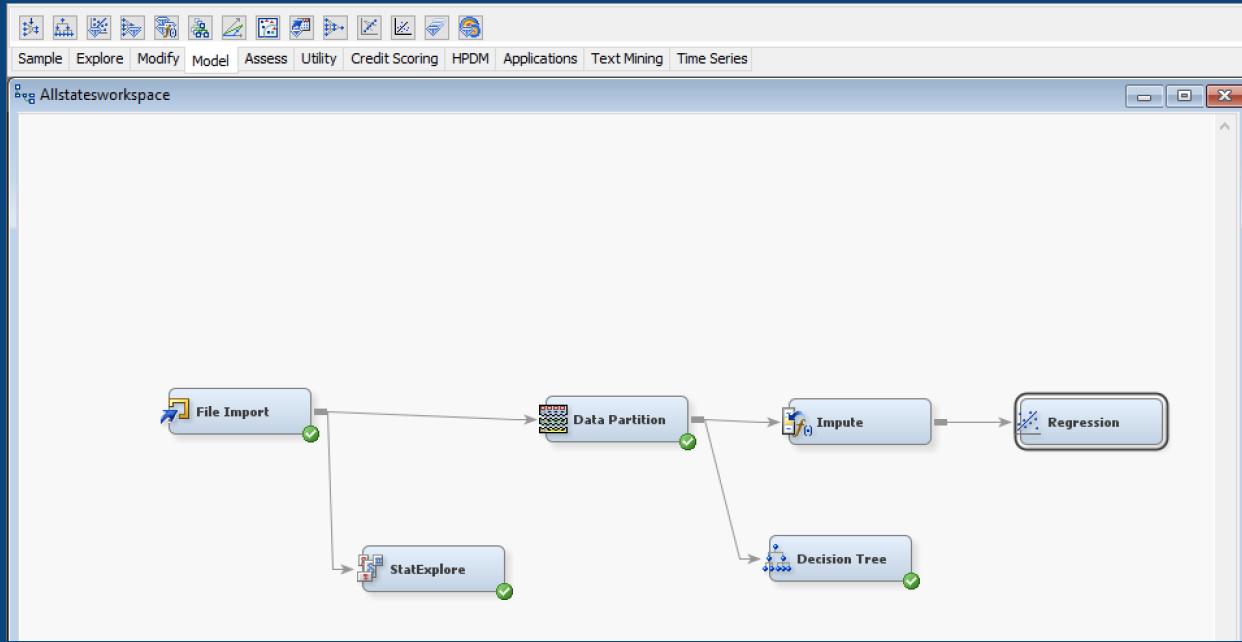
```

At this point, we can get to building the logistic regression model.

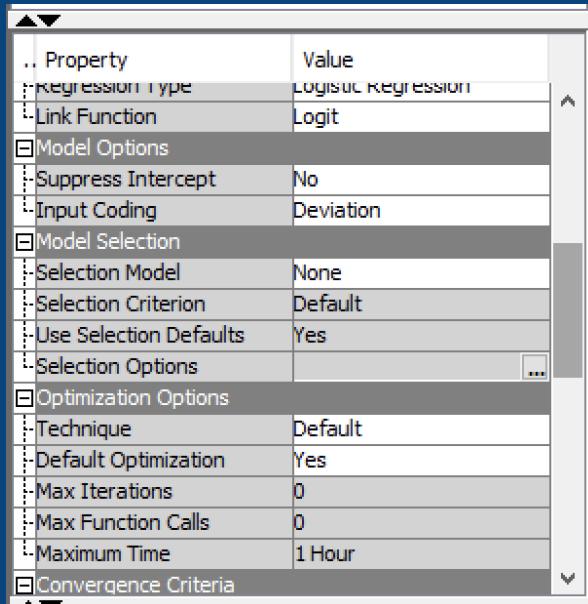
Forward Regression

Firstly, we run Forward regression.

Drag it to the workspace and connect it to the Impute Node



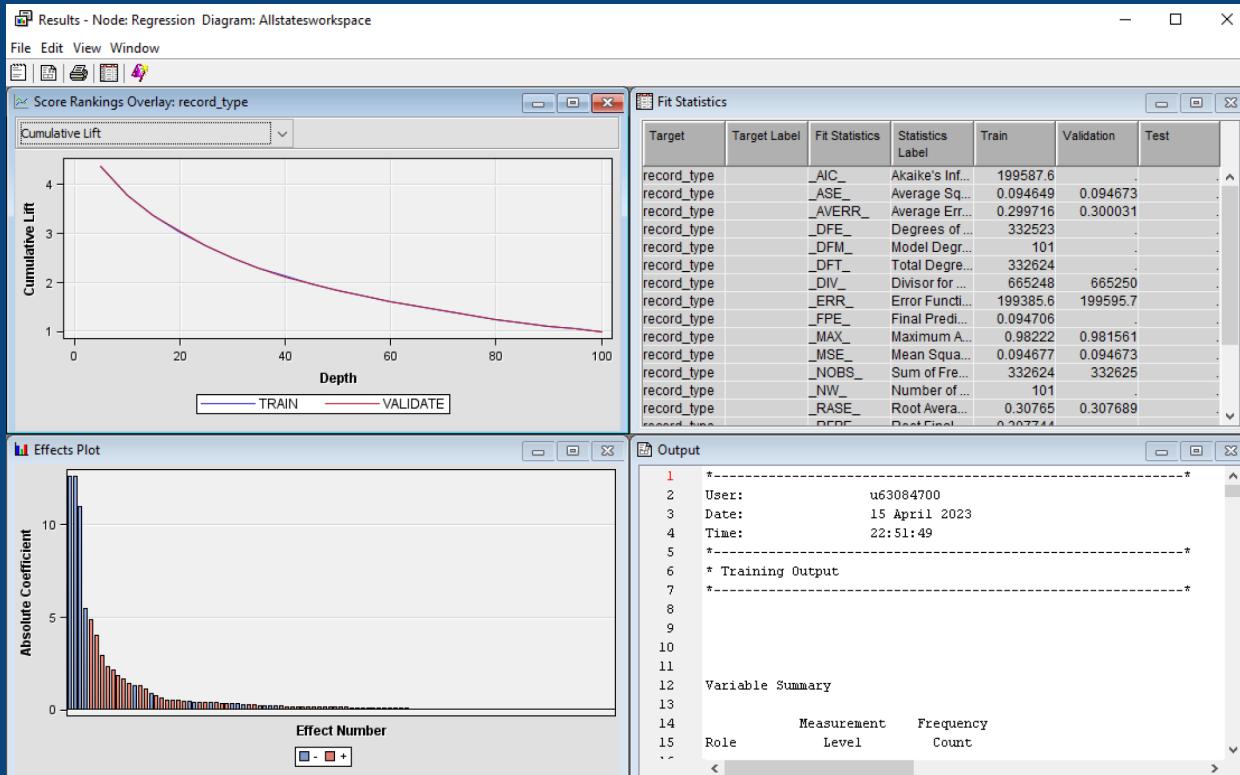
Open its property panel



And select 'Forward' for Selection Model in Model Selection Section

... Property	Value
Regression type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Forward
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour
Convergence Criteria	

Now we can run our Regression Node. And see the results.



Let's open the Output window and find the Type 3 Analysis of Effects, Analysis of Maximum Likelihood Estimates, and Odds Ratio Estimates:

Type 3 Analysis of Effects:

Type 3 Analysis of Effects				
		Wald		
4383	Effect	DF	Chi-Square	Pr > ChiSq
4385	A	2	34.0513	<.0001
4386	C	3	66.3919	<.0001
4387	D	2	52.5871	<.0001
4388	E	1	9.0459	0.0026
4389	G	3	299.8537	<.0001
4390	IMP_C_previous	3	20.5938	0.0001
4391	IMP_duration_previous	1	35.3420	<.0001
4392	M_C_previous	1	201.0496	<.0001
4393	M_car_value	1	0.0000	1.0000
4394	cost	1	149.8297	<.0001
4395	day	6	24.2502	0.0005
4396	homeowner	1	82.7288	<.0001
4397	married_couple	1	22.6081	<.0001
4398	risk_factor	4	9.7228	0.0454
4399	shopping_pt	12	30864.7317	<.0001
4400	state	35	1025.6215	<.0001
4401	time	23	674.0740	<.0001

Here we see the statistically significant inputs for our model.

Analysis of Maximum Likelihood Estimates:

Analysis of Maximum Likelihood Estimates								
4403	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
4404	Intercept	1	-5.4835	12.7465	0.19	0.6671	0.004	
4405	A	0	0.0500	0.0119	17.58	<.0001	1.051	
4406	A	1	-0.0488	0.00687	30.24	<.0001	0.952	
4407	C	1	0.0605	0.0153	15.55	<.0001	1.062	
4408	C	2	-0.00835	0.0140	0.35	0.5524	0.992	
4409	C	3	-0.0933	0.0122	58.66	<.0001	0.911	
4410	D	1	-0.0859	0.0135	40.43	<.0001	0.918	
4411	D	2	0.0129	0.0103	1.56	0.2110	1.013	
4412	E	0	-0.0200	0.00664	9.05	0.0026	0.980	
4413	G	1	-0.2154	0.0127	289.50	<.0001	0.806	
4414	G	2	0.0174	0.0105	2.76	0.0968	1.018	
4415	G	3	0.0941	0.0117	65.10	<.0001	1.099	
4416	IMP_C_previous	1	-0.00439	0.0130	0.11	0.7350	0.996	
4417	IMP_C_previous	2	0.0145	0.0140	1.07	0.3017	1.015	
4418	IMP_C_previous	3	0.0423	0.0111	14.53	0.0001	1.043	
4419	IMP_duration_previous	1	0.00772	0.00130	35.34	<.0001	0.0196	1.008
4420	M_C_previous	0	0.4075	0.0287	201.05	<.0001		1.503
4421	M_car_value	0	1	4.8435	.	.	.	126.918
4422	cost	1	-0.00211	0.000172	149.83	<.0001	-0.0534	0.998
4423	day	0	1	0.00551	0.0257	0.05	0.8301	1.006
4424	day	1	1	0.0545	0.0257	4.49	0.0341	1.056
4425	day	2	1	0.0153	0.0258	0.35	0.5521	1.015
4426	day	3	1	0.0276	0.0260	1.12	0.2891	1.028
4427	day	4	1	0.000870	0.0260	0.00	0.9733	1.001
4428	day	5	1	0.1672	0.0473	12.49	0.0004	1.182
4429	homeowner	0	1	-0.0572	0.00628	82.73	<.0001	0.944
4430	married_couple	0	1	0.0338	0.00711	22.61	<.0001	1.034
4431	risk_factor	1	1	-0.00652	0.0153	0.18	0.6697	0.994
4432	risk_factor	2	1	-0.0314	0.0136	5.35	0.0207	0.969
4433	risk_factor	3	1	-0.00205	0.0121	0.03	0.8655	0.998
4434	risk_factor	4	1	0.00363	0.0127	0.08	0.7747	1.004
4435	shopping_pt	1	1	-12.6075	10.8329	1.35	0.2445	0.000
4436	shopping_pt	2	1	-12.6230	10.8313	1.36	0.2438	0.000

# Output								
4442	shopping_pt	3	1	-1.3263	9.7202	0.02	0.8915	0.265
4443	shopping_pt	4	1	-0.8920	9.7202	0.01	0.9269	0.410
4444	shopping_pt	5	1	-0.3946	9.7202	0.00	0.9676	0.674
4445	shopping_pt	6	1	0.1781	9.7202	0.00	0.9854	1.195
4446	shopping_pt	7	1	0.7628	9.7202	0.01	0.9374	2.144
4447	shopping_pt	8	1	1.3109	9.7202	0.02	0.8927	3.710
4448	shopping_pt	9	1	1.8344	9.7202	0.04	0.8503	6.262
4449	shopping_pt	10	1	2.3339	9.7202	0.06	0.8102	10.318
4450	shopping_pt	11	1	2.9484	9.7204	0.09	0.7616	19.075
4451	shopping_pt	12	1	3.9962	9.7223	0.17	0.6810	54.390
4452	state	AL	1	0.1015	0.0309	10.82	0.0010	1.107
4453	state	AR	1	0.1649	0.0447	13.60	0.0002	1.179
4454	state	CO	1	-0.1477	0.0310	22.70	<.0001	0.863
4455	state	CT	1	0.1831	0.0346	28.04	<.0001	1.201
4456	state	DC	1	0.1288	0.0732	3.09	0.0785	1.137
4457	state	DE	1	0.1519	0.0712	4.55	0.0329	1.164
4458	state	FL	1	-0.4248	0.0239	315.90	<.0001	0.654
4459	state	GA	1	-0.1114	0.0507	4.82	0.0281	0.895
4460	state	IA	1	0.0110	0.0667	0.03	0.8687	1.011
4461	state	ID	1	-0.2187	0.0602	13.21	0.0003	0.804
4462	state	IN	1	0.1616	0.0312	26.86	<.0001	1.175
4463	state	KS	1	0.1328	0.0595	4.97	0.0257	1.142
4464	state	KY	1	0.1575	0.0369	18.22	<.0001	1.171
4465	state	MD	1	0.1687	0.0304	30.82	<.0001	1.184
4466	state	ME	1	0.0836	0.0571	2.15	0.1428	1.087
4467	state	MO	1	-0.1072	0.0382	7.86	0.0051	0.898
4468	state	MS	1	0.0280	0.0452	0.39	0.5344	1.028
4469	state	MT	1	-0.3235	0.0817	15.67	<.0001	0.724
4470	state	ND	1	0.3618	0.0941	14.80	0.0001	1.436
4471	state	NE	1	-0.0132	0.0834	0.02	0.8747	0.987
4472	state	NH	1	-0.0536	0.0521	1.06	0.3033	0.948
4473	state	NM	1	0.0856	0.0535	2.56	0.1094	1.089
4474	state	NV	1	-0.4185	0.0397	111.33	<.0001	0.658
4475	state	NY	1	0.1379	0.0223	38.38	<.0001	1.148
4476	state	OH	1	-0.0357	0.0256	1.93	0.1642	0.965
4477	state	OK	1	-0.0938	0.0405	5.36	0.0206	0.910
4478	state	OR	1	-0.1999	0.0392	26.07	<.0001	0.819
4479	state	PA	1	0.0261	0.0214	1.49	0.2225	1.026
4480	state	RI	1	0.1667	0.0595	7.84	0.0051	1.181
4481	state	SD	1	0.3535	0.1923	3.38	0.0660	1.424
4482	state	TN	1	0.0346	0.0355	0.95	0.3304	1.035
4483	state	UT	1	-0.2938	0.0392	56.17	<.0001	0.745
4484	state	WA	1	-0.3503	0.0316	122.60	<.0001	0.704
4485	state	WI	1	0.0408	0.0461	0.78	0.3759	1.042
4486	state	WV	1	0.2346	0.0630	13.85	0.0002	1.264
4487	time	0	1	2.1519	8.2632	0.07	0.7945	8.601
4488	time	1	1	0.2909	8.2698	0.00	0.9719	1.338
4489	time	2	1	1.6894	8.3055	0.04	0.8388	5.416
4490	time	3	1	-10.9379	189.6	0.00	0.9540	0.000
4491	time	4	1	1.4075	8.2696	0.03	0.8648	4.086
4492	time	5	1	0.1054	8.2657	0.00	0.9898	1.111
4493	time	6	1	1.1400	8.2499	0.02	0.8901	3.127
4494	time	7	1	0.5226	8.2453	0.00	0.9495	1.686
4495	time	8	1	0.0495	8.2450	0.00	0.9952	1.051
4496	time	9	1	-0.0537	8.2450	0.00	0.9948	0.948
4497	time	10	1	-0.0565	8.2450	0.00	0.9945	0.945
4498	time	11	1	-0.0292	8.2450	0.00	0.9972	0.971
4499	time	12	1	0.00433	8.2450	0.00	0.9996	1.004
4500	time	13	1	-0.0255	8.2450	0.00	0.9975	0.975
4501	time	14	1	-0.00867	8.2450	0.00	0.9992	0.991
4502	time	15	1	0.0475	8.2450	0.00	0.9954	1.049
4503	time	16	1	0.2573	8.2450	0.00	0.9751	1.293
4504	time	17	1	0.4090	8.2450	0.00	0.9604	1.505
4505	time	18	1	0.4764	8.2450	0.00	0.9539	1.610
4506	time	19	1	0.5364	8.2451	0.00	0.9481	1.710
4507	time	20	1	0.5255	8.2455	0.00	0.9492	1.691
4508	time	21	1	0.3726	8.2462	0.00	0.9640	1.452
4509	time	22	1	0.6151	8.2473	0.01	0.9406	1.850

# Output								
4477	state	OK	1	-0.0938	0.0405	5.36	0.0206	0.910
4478	state	OR	1	-0.1999	0.0392	26.07	<.0001	0.819
4479	state	PA	1	0.0261	0.0214	1.49	0.2225	1.026
4480	state	RI	1	0.1667	0.0595	7.84	0.0051	1.181
4481	state	SD	1	0.3535	0.1923	3.38	0.0660	1.424
4482	state	TN	1	0.0346	0.0355	0.95	0.3304	1.035
4483	state	UT	1	-0.2938	0.0392	56.17	<.0001	0.745
4484	state	WA	1	-0.3503	0.0316	122.60	<.0001	0.704
4485	state	WI	1	0.0408	0.0461	0.78	0.3759	1.042
4486	state	WV	1	0.2346	0.0630	13.85	0.0002	1.264
4487	time	0	1	2.1519	8.2632	0.07	0.7945	8.601
4488	time	1	1	0.2909	8.2698	0.00	0.9719	1.338
4489	time	2	1	1.6894	8.3055	0.04	0.8388	5.416
4490	time	3	1	-10.9379	189.6	0.00	0.9540	0.000
4491	time	4	1	1.4075	8.2696	0.03	0.8648	4.086
4492	time	5	1	0.1054	8.2657	0.00	0.9898	1.111
4493	time	6	1	1.1400	8.2499	0.02	0.8901	3.127
4494	time	7	1	0.5226	8.2453	0.00	0.9495	1.686
4495	time	8	1	0.0495	8.2450	0.00	0.9952	1.051
4496	time	9	1	-0.0537	8.2450	0.00	0.9948	0.948
4497	time	10	1	-0.0565	8.2450	0.00	0.9945	0.945
4498	time	11	1	-0.0292	8.2450	0.00	0.9972	0.971
4499	time	12	1	0.00433	8.2450	0.00	0.9996	1.004
4500	time	13	1	-0.0255	8.2450	0.00	0.9975	0.975
4501	time	14	1	-0.00867	8.2450	0.00	0.9992	0.991
4502	time	15	1	0.0475	8.2450	0.00	0.9954	1.049
4503	time	16	1	0.2573	8.2450	0.00	0.9751	1.293
4504	time	17	1	0.4090	8.2450	0.00	0.9604	1.505
4505	time	18	1	0.4764	8.2450	0.00	0.9539	1.610
4506	time	19	1	0.5364	8.2451	0.00	0.9481	1.710
4507	time	20	1	0.5255	8.2455	0.00	0.9492	1.691
4508	time	21	1	0.3726	8.2462	0.00	0.9640	1.452
4509	time	22	1	0.6151	8.2473	0.01	0.9406	1.850

Odds Ratio Estimates:

⇒ Output

4512	Odds Ratio Estimates	
4513		
4514		
4515	Effect	Point Estimate
4516		
4517	A	0 vs 2 1.053
4518	A	1 vs 2 0.954
4519	C	1 vs 4 1.020
4520	C	2 vs 4 0.952
4521	C	3 vs 4 0.874
4522	D	1 vs 3 0.853
4523	D	2 vs 3 0.942
4524	E	0 vs 1 0.961
4525	G	1 vs 4 0.727
4526	G	2 vs 4 0.917
4527	G	3 vs 4 0.990
4528	IMP_C_previous	1 vs 4 1.049
4529	IMP_C_previous	2 vs 4 1.069
4530	IMP_C_previous	3 vs 4 1.099
4531	IMP_duration_previous	1.008
4532	M_C_previous	0 vs 1 2.259
4533	M_car_value	0 vs 1 999.000
4534	cost	0.998
4535	day	0 vs 6 1.318
4536	day	1 vs 6 1.385
4537	day	2 vs 6 1.332
4538	day	3 vs 6 1.348
4539	day	4 vs 6 1.312
4540	day	5 vs 6 1.550
4541	homeowner	0 vs 1 0.892
4542	married_couple	0 vs 1 1.070
4543	risk_factor	1 vs NA 0.958
4544	risk_factor	2 vs NA 0.934
4545	risk_factor	3 vs NA 0.962
4546	risk_factor	4 vs NA 0.968
4547	shopping_pt	1 vs 13 <0.001
4548	shopping_pt	2 vs 13 <0.001
4549	shopping_pt	3 vs 13 <0.001
4550	shopping_pt	4 vs 13 <0.001

⇒ Output

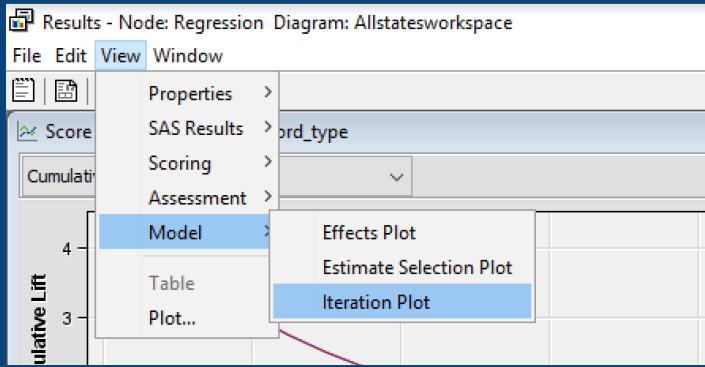
4550	shopping_pt	4 vs 13 <0.001
4551	shopping_pt	5 vs 13 <0.001
4552	shopping_pt	6 vs 13 <0.001
4553	shopping_pt	7 vs 13 <0.001
4554	shopping_pt	8 vs 13 <0.001
4555	shopping_pt	9 vs 13 <0.001
4556	shopping_pt	10 vs 13 <0.001
4557	shopping_pt	11 vs 13 <0.001
4558	shopping_pt	12 vs 13 <0.001
4559	state	AL vs WY 1.252
4560	state	AR vs WY 1.334
4561	state	CO vs WY 0.976
4562	state	CT vs WY 1.358
4563	state	DC vs WY 1.287
4564	state	DE vs WY 1.317
4565	state	FL vs WY 0.740
4566	state	GA vs WY 1.012
4567	state	IA vs WY 1.144
4568	state	ID vs WY 0.909
4569	state	IN vs WY 1.330
4570	state	KS vs WY 1.292
4571	state	KY vs WY 1.324
4572	state	MD vs WY 1.339
4573	state	ME vs WY 1.230
4574	state	MO vs WY 1.016
4575	state	MS vs WY 1.163
4576	state	MT vs WY 0.818
4577	state	ND vs WY 1.624
4578	state	NE vs WY 1.116
4579	state	NH vs WY 1.072
4580	state	NM vs WY 1.232
4581	state	NV vs WY 0.744
4582	state	NY vs WY 1.298
4583	state	OH vs WY 1.091
4584	state	OK vs WY 1.030
4585	state	OR vs WY 0.926
4586	state	PA vs WY 1.161
4587	state	RI vs WY 1.336
4588	state	SD vs WY 1.611

Output			
4588	state	SD vs WY	1.611
4589	state	TN vs WY	1.171
4590	state	UT vs WY	0.843
4591	state	WA vs WY	0.797
4592	state	WI vs WY	1.178
4593	state	WV vs WY	1.430
4594	time	0 vs 23	5.164
4595	time	1 vs 23	0.803
4596	time	2 vs 23	3.251
4597	time	3 vs 23	<0.001
4598	time	4 vs 23	2.453
4599	time	5 vs 23	0.667
4600	time	6 vs 23	1.877
4601	time	7 vs 23	1.012
4602	time	8 vs 23	0.631
4603	time	9 vs 23	0.569
4604	time	10 vs 23	0.567
4605	time	11 vs 23	0.583
4606	time	12 vs 23	0.603
4607	time	13 vs 23	0.585
4608	time	14 vs 23	0.595
4609	time	15 vs 23	0.630
4610	time	16 vs 23	0.776
4611	time	17 vs 23	0.904
4612	time	18 vs 23	0.967
4613	time	19 vs 23	1.026
4614	time	20 vs 23	1.015
4615	time	21 vs 23	0.871
4616	time	22 vs 23	1.111
4617			

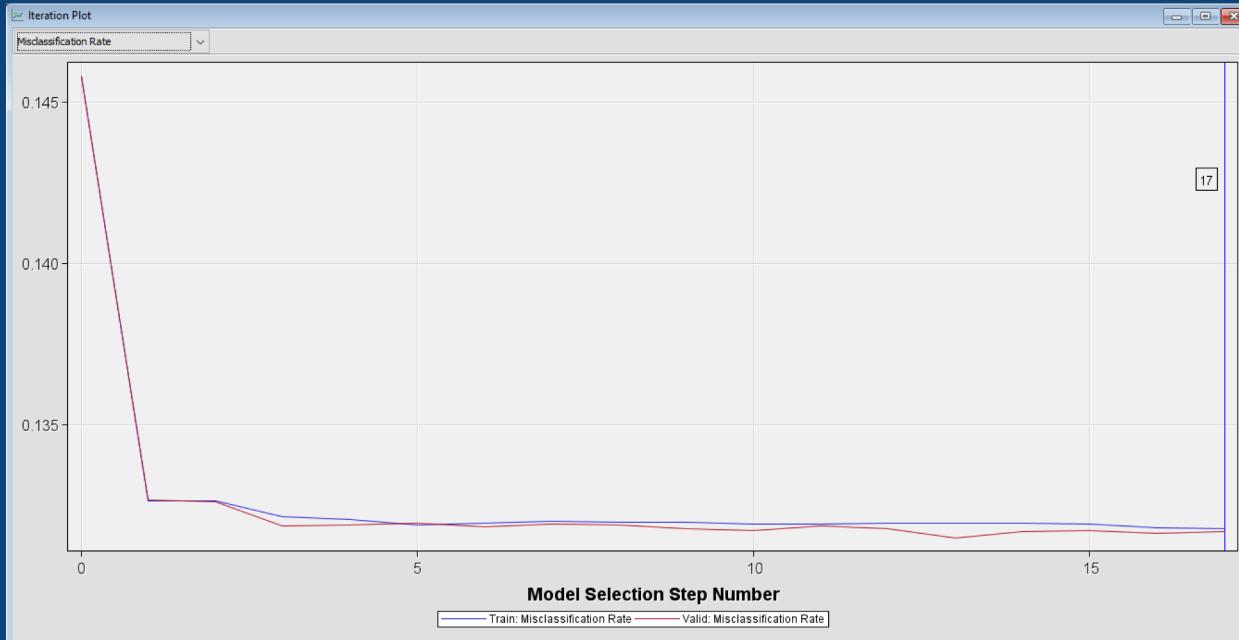
Here are the fit statistics of our Forward Regression:

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type		_AIC_	Akaike's Information Criterion	199597.6		
record_type		_ASE_	Average Squared Error	0.094649	0.094673	
record_type		_AVERR_	Average Error Function	0.299716	0.300031	
record_type		_DFE_	Degrees of Freedom for Error	332523		
record_type		_DFM_	Model Degrees of Freedom	101		
record_type		_DFT_	Total Degrees of Freedom	332624		
record_type		_DIV_	Divisor for ASE	665248	665250	
record_type		_ERR_	Error Function	199395.6	199595.7	
record_type		_FPE_	Final Prediction Error	0.094706		
record_type		_MAX_	Maximum Absolute Error	0.9822	0.981561	
record_type		_MSE_	Mean Square Error	0.094677	0.094673	
record_type		_NOBS_	Sum of Frequencies	332624	332625	
record_type		_NW_	Number of Estimate Weights	101		
record_type		_RASE_	Root Average Sum of Squares	0.30765	0.307689	
record_type		_RFPE_	Root Final Prediction Error	0.307744		
record_type		_RMSE_	Root Mean Squared Error	0.307697	0.307689	
record_type		_SBC_	Schwarz's Bayesian Criterion	200689.8		
record_type		_SSE_	Sum of Squared Errors	52964.78	62981.05	
record_type		_SUMW_	Sum of Case Weights Times Freq	665248	665250	
record_type		_MISC_	Misclassification Rate	0.1318	0.131713	

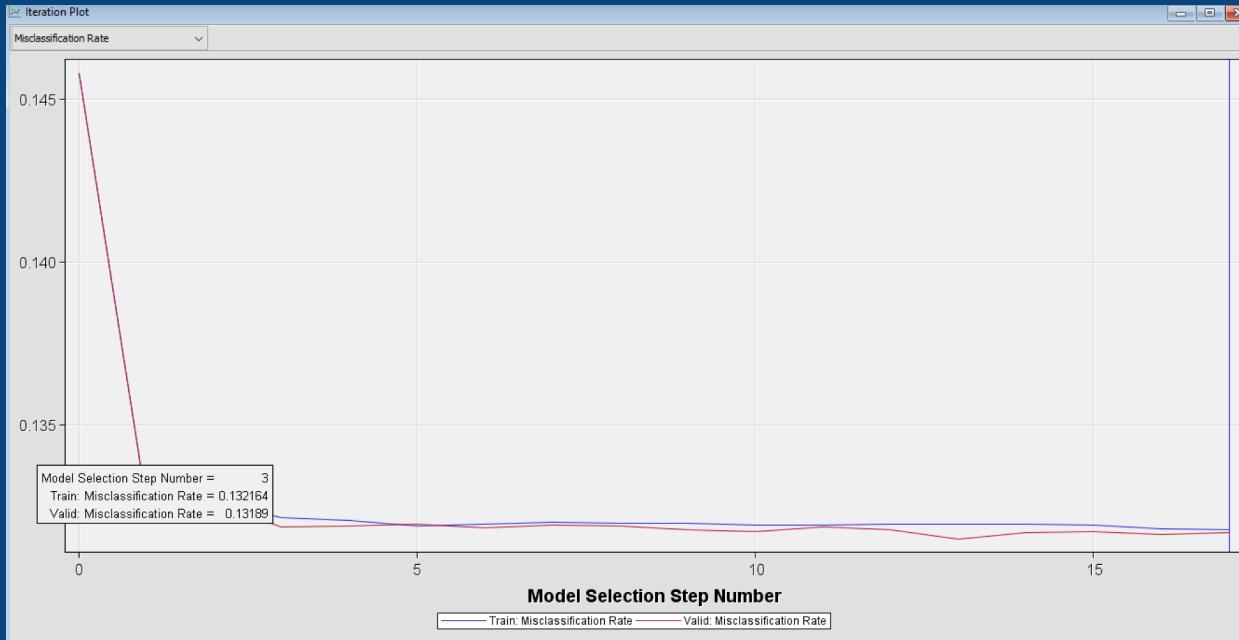
Now we need to expose the interaction plot for our Forward Regression. To do that, we need to select View -> Model -> Iteration Plot



Since our prediction is a decision (yes/no), we select Misclassification Rate for Iteration Plot:



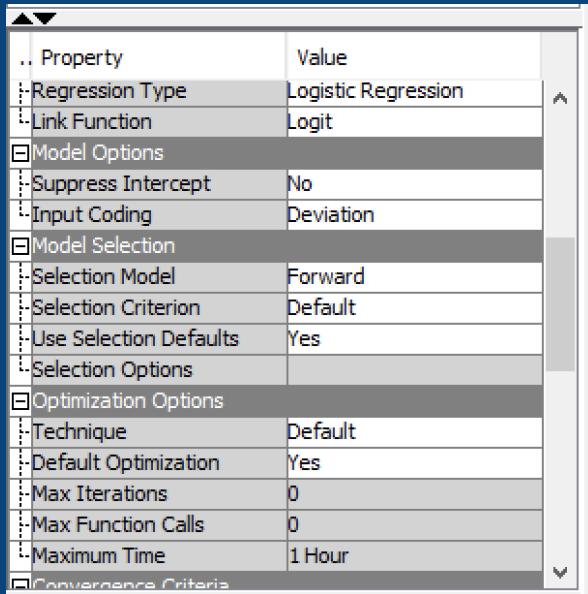
We see, that the final step in our model is 17, but apparently, an acceptable result appears at step 3:



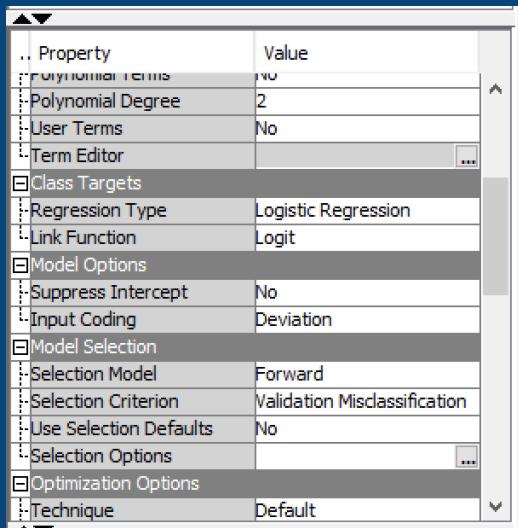
It means we can try to configure the regression node to select the model with the smallest fit statistic rather than the final forward selection interaction. That is, we will optimize the complexity of forward regression.

Fordawd Regression, best sequence model

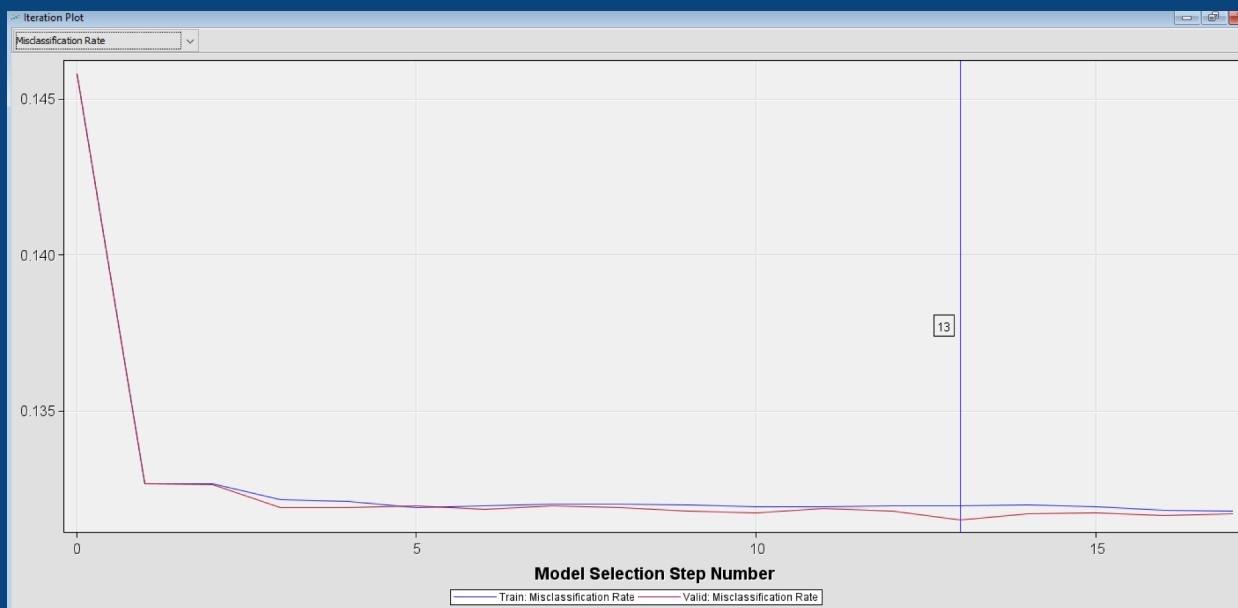
To do that, we close the results window and go back to the property panel of our regression node and find there a Model Selection Section



Select Selection Criterion -> Validation Misclassification, Use Selection Defaults -> No



Now we can run the forward regression node and take a look at the iteration plot.



Step 13 was chosen as the step with the lowest validation fit statistics.
We still assume that the model at step 3 could be the most optimal one.

Here are the fit statistics for this best sequence model

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type		_AIC_	Akaike's Information Criterion	199623.2		
record_type		_ASE_	Average Squared Error	0.094675	0.094677	
record_type		_AVERR_	Average Error Function	0.299812	0.300061	
record_type		_DFE_	Degrees of Freedom for Error	332537		
record_type		_DFM_	Model Degrees of Freedom	87		
record_type		_DFT_	Total Degrees of Freedom	332624		
record_type		_DIV_	Divisor for ASE	665248	665250	
record_type		_ERR_	Error Function	199449.2	199815.8	
record_type		_FPE_	Final Prediction Error	0.094725		
record_type		_MAX_	Maximum Absolute Error	0.982204	0.982191	
record_type		_MSE_	Mean Square Error	0.0947	0.094677	
record_type		_NOBS_	Sum of Frequencies	332624	332625	
record_type		_NW_	Number of Estimate Weights	87		
record_type		_RASE_	Root Average Sum of Squares	0.307694	0.307696	
record_type		_RFPE_	Root Final Prediction Error	0.307774		
record_type		_RMSE_	Root Mean Squared Error	0.307734	0.307696	
record_type		_SBC_	Schwarz's Bayesian Criterion	200555.3		
record_type		_SSE_	Sum of Squared Errors	62982.65	62983.69	
record_type		_SUMW_	Sum of Case Weights Times Freq	665248	665250	
record_type		_MISC_	Misclassification Rate	0.131951	0.131517	

Lets open the Output window and find the Type 3 Analysis of Effects, Analysis of Maximum Likelihood Estimates, and Odds Ratio Estimates for this best sequence model:

Type 3 Analysis of Effects

4381	Type 3 Analysis of Effects				
4382					
4383					
4384	Effect	DF	Chi-Square	Pr > ChiSq	Wald
4385					
4386	A	2	30.4795	<.0001	
4387	C	3	55.4538	<.0001	
4388	D	2	53.9629	<.0001	
4389	G	3	298.5474	<.0001	
4390	IMP_duration_previous	1	36.8042	<.0001	
4391	M_C_previous	1	191.2193	<.0001	
4392	M_car_value	1	0.0000	1.0000	
4393	cost	1	144.9419	<.0001	
4394	homeowner	1	86.9384	<.0001	
4395	married_couple	1	21.8039	<.0001	
4396	shopping_pt	12	30879.6492	<.0001	
4397	state	35	1067.1794	<.0001	
4398	time	23	679.1303	<.0001	
4399					

Analysis of Maximum Likelihood Estimates

4401	Analysis of Maximum Likelihood Estimates					
4402						
4403						
4404	Parameter	DF	Estimate	Standard Error	Chi-Square	Wald
4405						Standardized Estimate
4406	Intercept	1	-5.5623	12.8057	0.19	0.6640
4407	A	0	0.0433	0.0112	14.99	0.0001
4408	A	1	-0.0431	0.00843	26.12	<.0001
4409	C	1	0.0653	0.0129	25.75	<.0001
4410	C	2	0.00367	0.0117	0.10	0.7545
4411	C	3	-0.0666	0.00999	44.49	<.0001
4412	D	1	-0.0871	0.0135	41.72	<.0001
4413	D	2	0.0135	0.0103	1.71	0.1908
4414	G	1	-0.2144	0.0126	289.24	<.0001
4415	G	2	0.0156	0.0104	2.24	0.1341
4416	G	3	0.0936	0.0116	64.62	<.0001
4417	IMP_duration_previous	1	0.00783	0.00129	36.80	<.0001
4418	M_C_previous	0	0.3908	0.0283	191.22	<.0001
4419	M_car_value	0	4.8594	.	.	128.944
4420	cost	1	-0.00193	0.000160	144.94	<.0001
4421	homeowner	0	0.1	-0.0583	86.94	<.0001
4422	married_couple	0	0.1	0.0332	0.0710	0.943
4423	shopping_pt	1	1	-12.6178	10.8895	1.34
4424	shopping_pt	2	1	-12.6316	10.8887	1.35
4425	shopping_pt	3	1	-1.3253	9.7731	0.02
4426	shopping_pt	4	1	-0.8910	9.7731	0.01
4427	shopping_pt	5	1	-0.3936	9.7731	0.00
4428	shopping_pt	6	1	0.1789	9.7731	0.00
4429	shopping_pt	7	1	0.7634	9.7731	0.01
4430	shopping_pt	8	1	1.3119	9.7731	0.02
4431	shopping_pt	9	1	1.8347	9.7731	0.04
4432	shopping_pt	10	1	2.3354	9.7731	0.06
4433	shopping_pt	11	1	2.9482	9.7733	0.09
4434	shopping_pt	12	1	3.9977	9.7753	0.17
4435	state	AL	1	0.1074	0.0308	12.17
4436	state	AR	1	0.1725	0.0447	14.92
4437	state	CO	1	-0.1504	0.0309	23.64
4438	state	CT	1	0.1917	0.0345	30.85
4439	state	DC	1	0.1318	0.0732	3.24

# Output							
4439	state	DC	1	0.1318	0.0732	3.24	0.0717
4440	state	DE	1	0.1314	0.0704	3.48	0.0622
4441	state	FL	1	-0.4463	0.0216	427.29	<.0001
4442	state	GA	1	-0.1041	0.0506	4.23	0.0397
4443	state	IA	1	0.0188	0.0664	0.08	0.7768
4444	state	ID	1	-0.2403	0.0597	16.21	<.0001
4445	state	IN	1	0.1429	0.0303	22.30	<.0001
4446	state	KS	1	0.1379	0.0595	5.38	0.0204
4447	state	KY	1	0.1622	0.0368	19.40	<.0001
4448	state	MD	1	0.1499	0.0289	26.86	<.0001
4449	state	ME	1	0.0947	0.0569	2.77	0.0960
4450	state	MO	1	-0.0998	0.0380	6.88	0.0087
4451	state	MS	1	0.0388	0.0450	0.74	0.3891
4452	state	MT	1	-0.3260	0.0817	15.93	<.0001
4453	state	ND	1	0.3638	0.0939	15.02	0.0001
4454	state	NE	1	-0.0156	0.0833	0.03	0.8518
4455	state	NH	1	-0.0470	0.0520	0.82	0.3664
4456	state	NM	1	0.0826	0.0534	2.39	0.1223
4457	state	NV	1	-0.4139	0.0396	109.10	<.0001
4458	state	NY	1	0.1151	0.0194	35.34	<.0001
4459	state	OH	1	-0.0249	0.0253	0.96	0.3266
4460	state	OK	1	-0.0892	0.0404	4.87	0.0274
4461	state	OR	1	-0.2038	0.0391	27.14	<.0001
4462	state	PA	1	0.0354	0.0213	2.77	0.0958
4463	state	RI	1	0.1696	0.0594	8.15	0.0043
4464	state	SD	1	0.3467	0.1923	3.25	0.0714
4465	state	TN	1	0.0496	0.0353	1.97	0.1600
4466	state	UT	1	-0.2863	0.0391	53.64	<.0001
4467	state	WA	1	-0.3539	0.0316	125.60	<.0001
4468	state	WI	1	0.0449	0.0458	0.96	0.3272
4469	state	WV	1	0.2400	0.0630	14.51	0.0001
4470	time	0	1	2.1476	8.2922	0.07	0.7956
4471	time	1	1	0.3377	8.2988	0.00	0.9675
4472	time	2	1	1.7257	8.3349	0.04	0.8360
4473	time	3	1	-10.9478	190.3	0.00	0.9541
4474	time	4	1	1.3614	8.2986	0.03	0.8697
4475	time	5	1	0.1231	8.2948	0.00	0.9882
4476	time	6	1	1.1378	8.2790	0.02	0.8907
4477	time	7	1	0.5209	8.2744	0.00	0.9498

# Output							
4477	time	7	1	0.5209	8.2744	0.00	0.9498
4478	time	8	1	0.0473	8.2741	0.00	0.9954
4479	time	9	1	-0.0582	8.2741	0.00	0.9944
4480	time	10	1	-0.0601	8.2741	0.00	0.9942
4481	time	11	1	-0.0320	8.2741	0.00	0.9969
4482	time	12	1	0.000309	8.2741	0.00	1.0000
4483	time	13	1	-0.0291	8.2741	0.00	0.9972
4484	time	14	1	-0.0119	8.2741	0.00	0.9989
4485	time	15	1	0.0451	8.2741	0.00	0.9957
4486	time	16	1	0.2547	8.2741	0.00	0.9754
4487	time	17	1	0.4074	8.2741	0.00	0.9607
4488	time	18	1	0.4741	8.2741	0.00	0.9543
4489	time	19	1	0.5338	8.2742	0.00	0.9486
4490	time	20	1	0.5292	8.2746	0.00	0.9490
4491	time	21	1	0.3741	8.2753	0.00	0.9639
4492	time	22	1	0.6110	8.2763	0.01	0.9411
4493	time						1.842

Odds Ratio Estimates:

4495	Odds Ratio Estimates	
4496		
4497	Point	
4498	Effect	Estimate
4499		
4500	A	0 vs 2 1.045
4501	A	1 vs 2 0.958
4502	C	1 vs 4 1.070
4503	C	2 vs 4 1.006
4504	C	3 vs 4 0.938
4505	D	1 vs 3 0.852
4506	D	2 vs 3 0.942
4507	G	1 vs 4 0.726
4508	G	2 vs 4 0.914
4509	G	3 vs 4 0.989
4510	IMP_duration_previous	1.008
4511	M_C_previous	0 vs 1 2.185
4512	M_car_value	0 vs 1 999.000
4513	cost	0.998
4514	homeowner	0 vs 1 0.890
4515	married_couple	0 vs 1 1.069
4516	shopping_pt	1 vs 13 <0.001
4517	shopping_pt	2 vs 13 <0.001
4518	shopping_pt	3 vs 13 <0.001
4519	shopping_pt	4 vs 13 <0.001
4520	shopping_pt	5 vs 13 <0.001
4521	shopping_pt	6 vs 13 <0.001
4522	shopping_pt	7 vs 13 <0.001
4523	shopping_pt	8 vs 13 <0.001
4524	shopping_pt	9 vs 13 <0.001
4525	shopping_pt	10 vs 13 <0.001
4526	shopping_pt	11 vs 13 <0.001
4527	shopping_pt	12 vs 13 <0.001
4528	state	AL vs WY 1.263
4529	state	AR vs WY 1.348
4530	state	CO vs WY 0.976
4531	state	CT vs WY 1.374
4532	state	DC vs WY 1.294
4533	state	DE vs WY 1.294

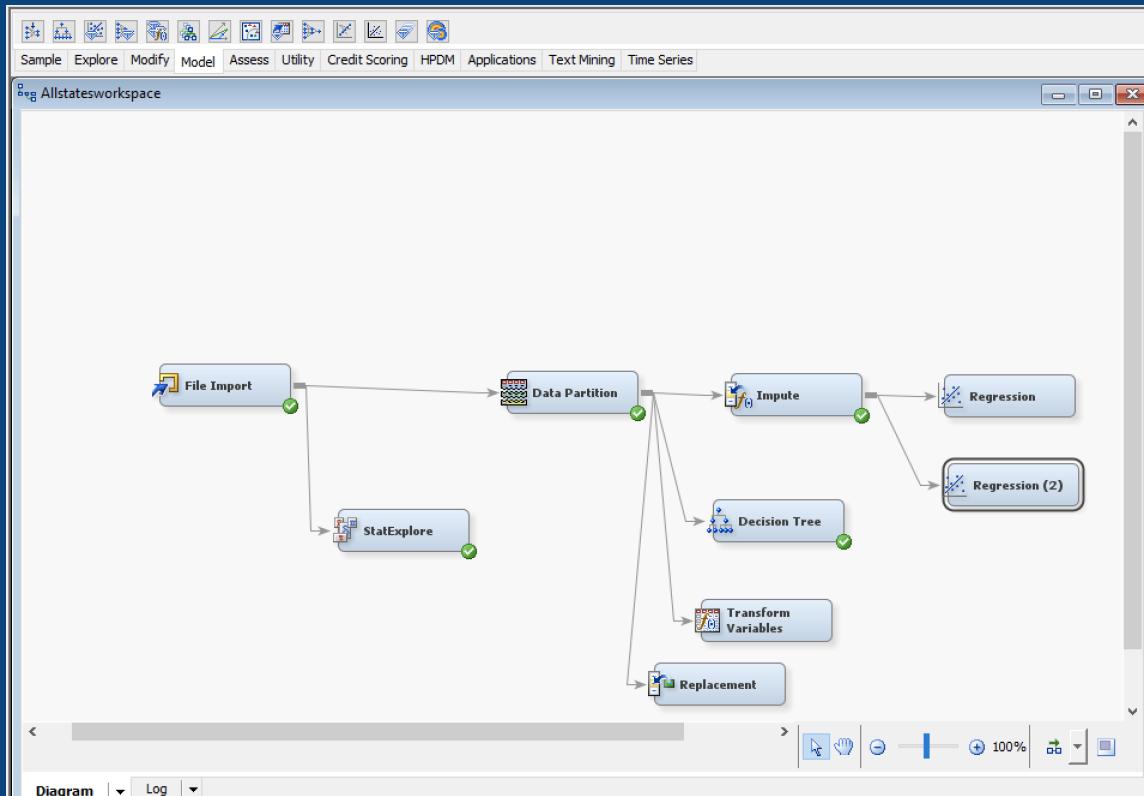
	Output	
4533	state	DE vs WY 1.294
4534	state	FL vs WY 0.726
4535	state	GA vs WY 1.022
4536	state	IA vs WY 1.156
4537	state	ID vs WY 0.892
4538	state	IN vs WY 1.309
4539	state	KS vs WY 1.302
4540	state	KY vs WY 1.334
4541	state	MD vs WY 1.318
4542	state	ME vs WY 1.247
4543	state	MO vs WY 1.027
4544	state	MS vs WY 1.179
4545	state	MT vs WY 0.819
4546	state	ND vs WY 1.632
4547	state	NE vs WY 1.117
4548	state	NH vs WY 1.082
4549	state	NM vs WY 1.232
4550	state	NV vs WY 0.750
4551	state	NY vs WY 1.273
4552	state	OH vs WY 1.107
4553	state	OK vs WY 1.038
4554	state	OR vs WY 0.925
4555	state	PA vs WY 1.175
4556	state	RI vs WY 1.344
4557	state	SD vs WY 1.605
4558	state	TN vs WY 1.192
4559	state	UT vs WY 0.852
4560	state	WA vs WY 0.796
4561	state	WI vs WY 1.187
4562	state	WV vs WY 1.442
4563	time	0 vs 23 5.154
4564	time	1 vs 23 0.844
4565	time	2 vs 23 3.380
4566	time	3 vs 23 <0.001
4567	time	4 vs 23 2.348
4568	time	5 vs 23 0.681
4569	time	6 vs 23 1.877
4570	time	7 vs 23 1.013
4571	time	8 vs 23 0.631

4571 Output			
4571	time	8 vs 23	0.631
4572	time	9 vs 23	0.568
4573	time	10 vs 23	0.567
4574	time	11 vs 23	0.583
4575	time	12 vs 23	0.602
4576	time	13 vs 23	0.585
4577	time	14 vs 23	0.595
4578	time	15 vs 23	0.630
4579	time	16 vs 23	0.776
4580	time	17 vs 23	0.904
4581	time	18 vs 23	0.967
4582	time	19 vs 23	1.026
4583	time	20 vs 23	1.022
4584	time	21 vs 23	0.875
4585	time	22 vs 23	1.109

Stepwise regression

At this point we can get to building the stepwise regression model.

We add a new regression node to our workspace:



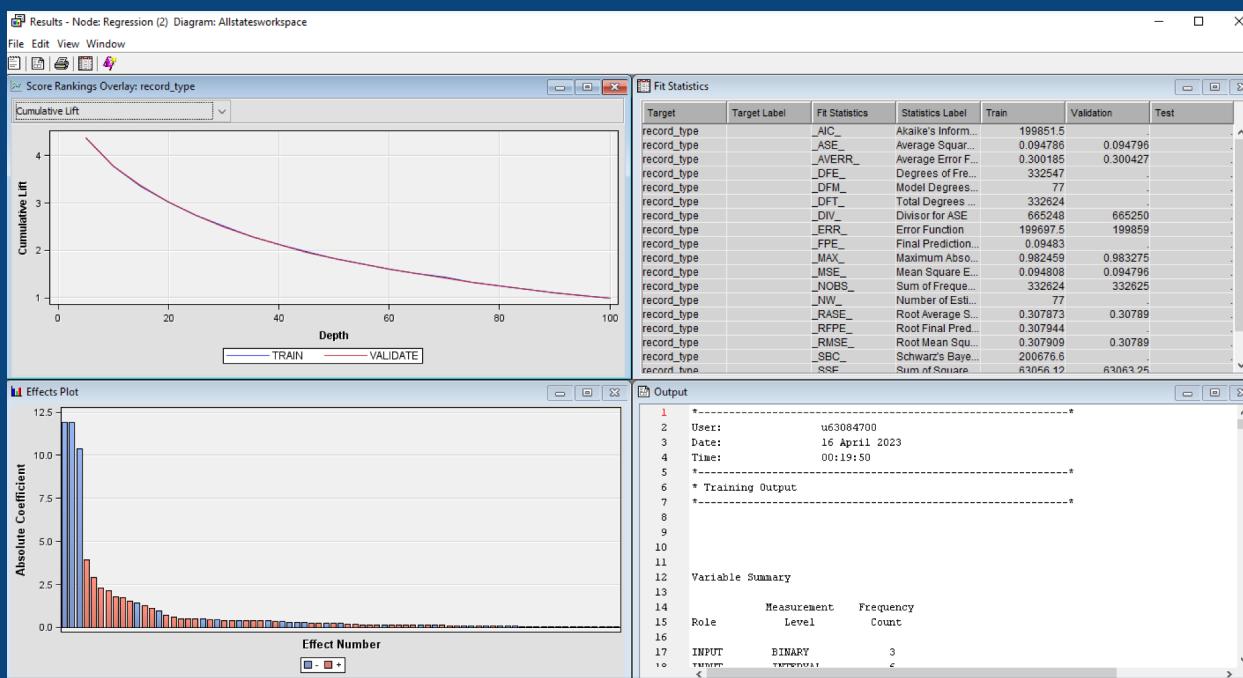
Open its property panel

Property	Value
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour
Convergence Criteria	

And select Selection Model -> Stepwise

Property	Value
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

Now we run the node and open results



Lets open the Output window and find the Type 3 Analysis of Effects, Analysis of Maximum Likelihood Estimates, and Odds Ratio Estimates:

Type 3 Analysis of Effects

Type 3 Analysis of Effects				
			Wald	
	Effect	DF	Chi-Square	Pr > ChiSq
2139	G	3	316.2179	<.0001
2140	M_C_previous	1	195.8588	<.0001
2141	cost	1	272.0308	<.0001
2142	homeowner	1	68.9907	<.0001
2143	shopping_pt	12	30999.2052	<.0001
2144	state	35	1060.7980	<.0001
2145	time	23	678.4634	<.0001
2151				

Analysis of Maximum Likelihood Estimates

Output

Analysis of Maximum Likelihood Estimates								
	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
2158	Intercept		1 -0.2712	9.3668	0.00	0.9769	0.762	
2159	G	1	1 -0.2173	0.0124	305.80	<.0001	0.805	
2160	G	2	1 0.0165	0.0103	2.55	0.1102	1.017	
2161	G	3	1 0.0971	0.0116	70.61	<.0001	1.102	
2162	M_C_previous	0	1 0.3918	0.0280	195.86	<.0001	1.480	
2163	cost		1 -0.00232	0.000141	272.03	<.0001	-0.0589	0.998
2164	homeowner	0	1 -0.0495	0.00596	68.99	<.0001	0.952	
2165	shopping_pt	1	1 -11.9351	7.9386	2.26	0.1327	0.000	
2166	shopping_pt	2	1 -11.9192	7.9260	2.26	0.1326	0.000	
2167	shopping_pt	3	1 -1.4003	7.2363	0.04	0.8466	0.247	
2168	shopping_pt	4	1 -0.9602	7.2363	0.02	0.8944	0.383	
2169	shopping_pt	5	1 -0.4618	7.2363	0.00	0.9491	0.630	
2170	shopping_pt	6	1 0.1100	7.2363	0.00	0.9879	1.116	
2171	shopping_pt	7	1 0.6955	7.2363	0.01	0.9234	2.005	
2172	shopping_pt	8	1 1.2434	7.2363	0.03	0.8636	3.468	
2173	shopping_pt	9	1 1.7638	7.2363	0.06	0.8074	5.834	
2174	shopping_pt	10	1 2.2623	7.2364	0.10	0.7546	9.605	
2175	shopping_pt	11	1 2.8754	7.2366	0.16	0.6911	17.732	
2176	shopping_pt	12	1 3.9076	7.2392	0.29	0.5893	49.778	
2177	state	AL	1 0.1001	0.0307	10.64	0.0011	1.105	
2178	state	AR	1 0.1517	0.0442	11.78	0.0006	1.164	
2179	state	CO	1 -0.1373	0.0308	19.81	<.0001	0.872	
2180	state	CT	1 0.2139	0.0343	38.81	<.0001	1.238	
2181	state	DC	1 0.1448	0.0730	3.93	0.0474	1.156	
2182	state	DE	1 0.1359	0.0703	3.74	0.0532	1.146	
2183	state	FL	1 -0.4209	0.0211	397.31	<.0001	0.656	
2184	state	GA	1 -0.1073	0.0501	4.58	0.0323	0.898	
2185	state	IA	1 -0.00740	0.0662	0.01	0.9111	0.993	
2186	state	ID	1 -0.2594	0.0596	18.94	<.0001	0.771	
2187	state	IN	1 0.1365	0.0301	20.62	<.0001	1.146	
2188	state	KS	1 0.1269	0.0593	4.58	0.0324	1.135	
2189	state	KY	1 0.1870	0.0366	26.05	<.0001	1.206	
2190	state	MD	1 0.1446	0.0287	25.34	<.0001	1.156	
2191	state	ME	1 0.0679	0.0566	1.44	0.2301	1.070	

Output

2191	state	ME	1 0.0679	0.0566	1.44	0.2301	1.070	
2192	state	MO	1 -0.1142	0.0378	9.10	0.0026	0.892	
2193	state	MS	1 0.0265	0.0444	0.36	0.5503	1.027	
2194	state	MT	1 -0.3077	0.0816	14.21	0.0002	0.735	
2195	state	ND	1 0.3689	0.0937	15.51	<.0001	1.446	
2196	state	NE	1 -0.0205	0.0832	0.06	0.8052	0.980	
2197	state	NH	1 -0.0621	0.0518	1.44	0.2304	0.940	
2198	state	NN	1 0.0794	0.0532	2.23	0.1351	1.083	
2199	state	NV	1 -0.3811	0.0392	94.43	<.0001	0.683	
2200	state	NY	1 0.1310	0.0188	48.57	<.0001	1.140	
2201	state	OH	1 -0.0276	0.0252	1.20	0.2727	0.973	
2202	state	OK	1 -0.0751	0.0401	3.51	0.0608	0.928	
2203	state	OR	1 -0.2136	0.0390	29.92	<.0001	0.808	
2204	state	PA	1 0.0371	0.0212	3.07	0.0799	1.038	
2205	state	RI	1 0.1771	0.0593	8.93	0.0028	1.194	
2206	state	SD	1 0.3835	0.1924	3.97	0.0462	1.467	
2207	state	TN	1 0.0354	0.0352	1.01	0.3148	1.036	
2208	state	UT	1 -0.2919	0.0389	56.29	<.0001	0.747	
2209	state	WA	1 -0.3607	0.0315	131.38	<.0001	0.697	
2210	state	WI	1 0.0408	0.0454	0.81	0.3695	1.042	
2211	state	WV	1 0.2255	0.0628	12.90	0.0003	1.253	
2212	time	0	1 2.1081	5.9721	0.12	0.7241	8.233	
2213	time	1	1 0.3741	5.9806	0.00	0.9501	1.454	
2214	time	2	1 1.7095	6.0309	0.08	0.7768	5.526	
2215	time	3	1 -10.3878	136.8	0.01	0.9395	0.000	
2216	time	4	1 1.4941	5.9798	0.06	0.8027	4.455	
2217	time	5	1 0.1065	5.9753	0.00	0.9658	1.112	
2218	time	6	1 1.0966	5.9537	0.03	0.8559	2.994	
2219	time	7	1 0.4853	5.9473	0.01	0.9550	1.625	
2220	time	8	1 0.0100	5.9468	0.00	0.9987	1.010	
2221	time	9	1 -0.0948	5.9468	0.00	0.9873	0.910	
2222	time	10	1 -0.0964	5.9468	0.00	0.9871	0.908	
2223	time	11	1 -0.0676	5.9468	0.00	0.9909	0.935	
2224	time	12	1 -0.0345	5.9468	0.00	0.9954	0.966	
2225	time	13	1 -0.0658	5.9468	0.00	0.9912	0.936	
2226	time	14	1 -0.0469	5.9468	0.00	0.9937	0.954	
2227	time	15	1 0.00840	5.9468	0.00	0.9989	1.008	
2228	time	16	1 0.2184	5.9468	0.00	0.9707	1.244	
2229	time	17	1 0.3696	5.9468	0.00	0.9504	1.447	
2230	time	18	1 0.4367	5.9469	0.01	0.9415	1.548	
2231	time	19	1 0.4979	5.9470	0.01	0.9333	1.645	
2232	time	20	1 0.4846	5.9475	0.01	0.9351	1.624	
2233	time	21	1 0.3384	5.9485	0.00	0.9546	1.403	
2234	time	22	1 0.5894	5.9499	0.01	0.9211	1.803	

Odds Ratio Estimates

Output

2237	Odds Ratio Estimates	
2238		
2239		
2240	Effect	Point Estimate
2241		
2242	G	1 vs 4 0.725
2243	G	2 vs 4 0.916
2244	G	3 vs 4 0.993
2245	M_C_previous	0 vs 1 2.189
2246	cost	0.998
2247	homeowner	0 vs 1 0.906
2248	shopping_pt	1 vs 13 <0.001
2249	shopping_pt	2 vs 13 <0.001
2250	shopping_pt	3 vs 13 <0.001
2251	shopping_pt	4 vs 13 <0.001
2252	shopping_pt	5 vs 13 <0.001
2253	shopping_pt	6 vs 13 <0.001
2254	shopping_pt	7 vs 13 <0.001
2255	shopping_pt	8 vs 13 <0.001
2256	shopping_pt	9 vs 13 <0.001
2257	shopping_pt	10 vs 13 <0.001
2258	shopping_pt	11 vs 13 <0.001
2259	shopping_pt	12 vs 13 <0.001
2260	state	AL vs WY 1.256
2261	state	AR vs WY 1.323
2262	state	CO vs WY 0.991
2263	state	CT vs WY 1.407
2264	state	DC vs WY 1.313
2265	state	DE vs WY 1.302
2266	state	FL vs WY 0.746
2267	state	GA vs WY 1.021
2268	state	IA vs WY 1.128
2269	state	ID vs WY 0.877
2270	state	IN vs WY 1.303
2271	state	KS vs WY 1.290
2272	state	KY vs WY 1.370
2273	state	MD vs WY 1.313
2274	state	ME vs WY 1.216
2275	state	MO vs WY 1.014

Output

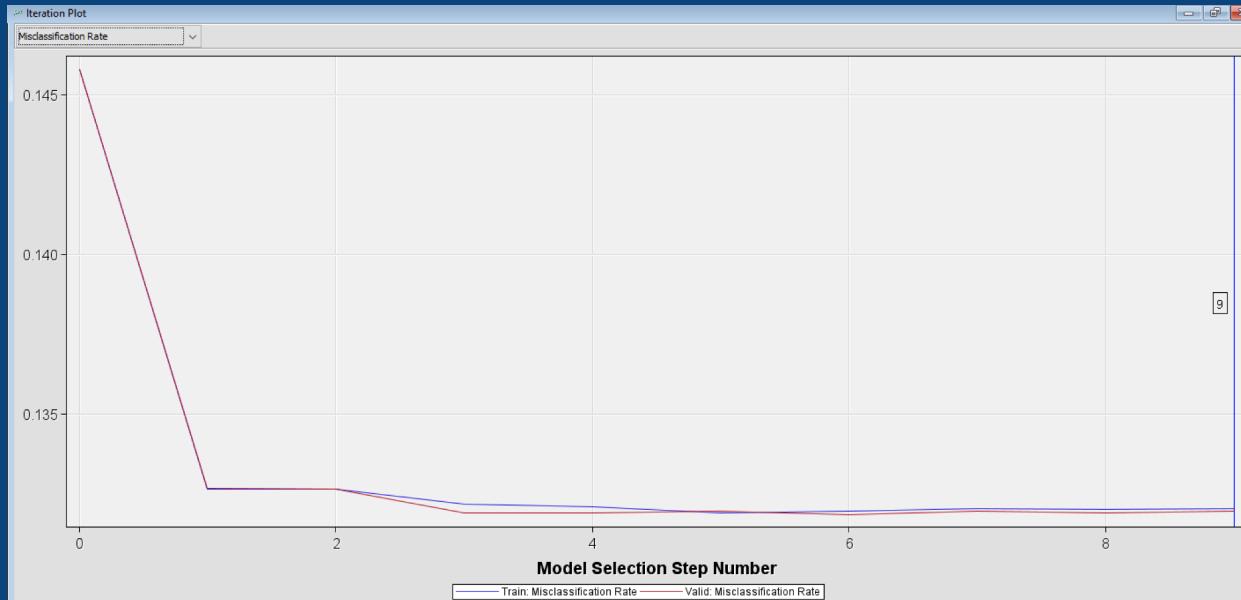
2275	state	MO vs WY 1.014
2276	state	MS vs WY 1.167
2277	state	MT vs WY 0.835
2278	state	ND vs WY 1.643
2279	state	NE vs WY 1.113
2280	state	NH vs WY 1.068
2281	state	NM vs WY 1.230
2282	state	NV vs WY 0.776
2283	state	NY vs WY 1.295
2284	state	OH vs WY 1.105
2285	state	OK vs WY 1.054
2286	state	OR vs WY 0.918
2287	state	PA vs WY 1.179
2288	state	RI vs WY 1.357
2289	state	SD vs WY 1.668
2290	state	TN vs WY 1.177
2291	state	UT vs WY 0.849
2292	state	WA vs WY 0.792
2293	state	WI vs WY 1.184
2294	state	WV vs WY 1.424
2295	time	0 vs 23 5.165
2296	time	1 vs 23 0.912
2297	time	2 vs 23 3.467
2298	time	3 vs 23 <0.001
2299	time	4 vs 23 2.795
2300	time	5 vs 23 0.698
2301	time	6 vs 23 1.878
2302	time	7 vs 23 1.019
2303	time	8 vs 23 0.634
2304	time	9 vs 23 0.571
2305	time	10 vs 23 0.570
2306	time	11 vs 23 0.586
2307	time	12 vs 23 0.606
2308	time	13 vs 23 0.587
2309	time	14 vs 23 0.599
2310	time	15 vs 23 0.633
2311	time	16 vs 23 0.781
2312	time	17 vs 23 0.908
2313	time	18 vs 23 0.971

2313	time	18 vs 23	0.971
2314	time	19 vs 23	1.032
2315	time	20 vs 23	1.019
2316	time	21 vs 23	0.880
2317	time	22 vs 23	1.131

Here are the fit statistics of our Stepwise Regression:

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type		_AIC_	Akaike's Information Criterion	199851.5		
record_type		_ASE_	Average Squared Error	0.094786	0.094796	
record_type		_AVERR_	Average Error Function	0.300185	0.300427	
record_type		_DFE_	Degrees of Freedom for Error	332547		
record_type		_DFM_	Model Degrees of Freedom	77		
record_type		_DFT_	Total Degrees of Freedom	332624		
record_type		_DIV_	Divisor for ASE	665248	665250	
record_type		_ERR_	Error Function	199697.5	199859	
record_type		_FPE_	Final Prediction Error	0.09483		
record_type		_MAX_	Maximum Absolute Error	0.982459	0.983275	
record_type		_MSE_	Mean Square Error	0.094808	0.094796	
record_type		_NOBS_	Sum of Frequencies	332624	332625	
record_type		_NW_	Number of Estimate Weights	77		
record_type		_RASE_	Root Average Sum of Squares	0.307873	0.30789	
record_type		_RFPE_	Root Final Prediction Error	0.307944		
record_type		_RMSE_	Root Mean Squared Error	0.307909	0.30789	
record_type		_SBC_	Schwarz's Bayesian Criterion	200676.6		
record_type		_SSE_	Sum of Squared Errors	63056.12	63063.25	
record_type		_SUMW_	Sum of Case Weights Times Freq	665248	665250	
record_type		_MISC_	Misclassification Rate	0.132026	0.131944	

Iteration Plot:



Stepwise Regression, Adjusted Entry and Stay Levels

Now we will run full model selection so that we can evaluate the largest sequence of regression models.

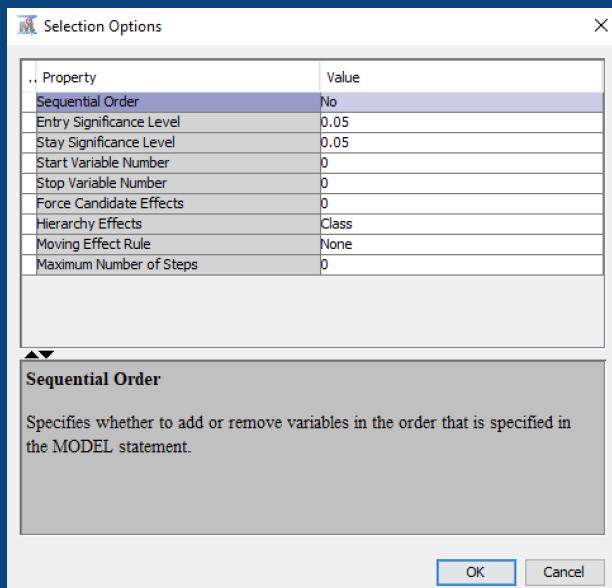
To do that, we need to close the results window and get back to the property panel of the stepwise regression node.

.. Property	Value
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Default
Use Selection Defaults	Yes
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

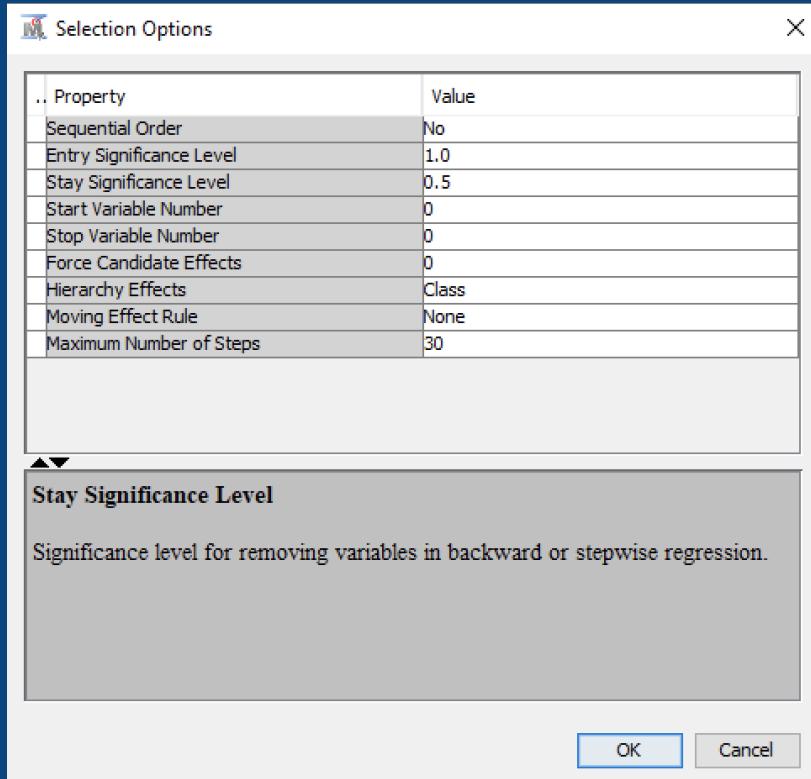
Select Use Selection Defaults -> No

.. Property	Value
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Default
Use Selection Defaults	No
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

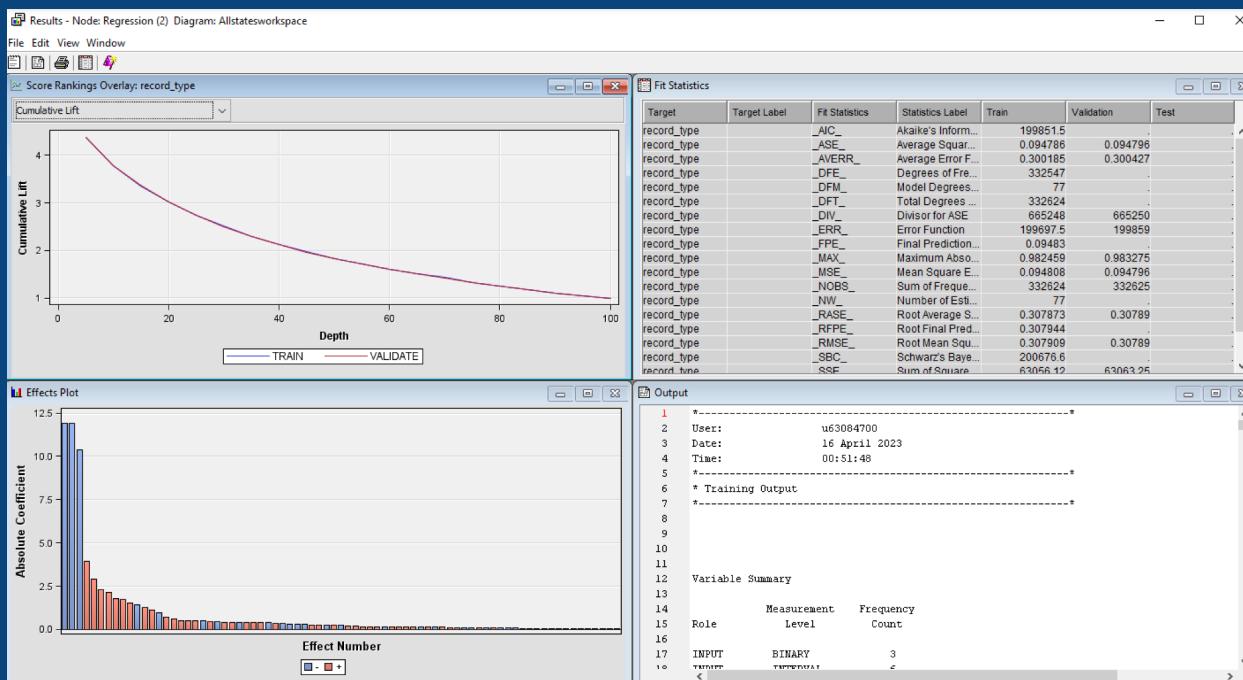
Then select Selection Options -> ...



Enter 1.0 as the Entry Significance Level Value and enter 0.5 as the Stay Significance Level Value. Then let's change the Maximum Number of Steps to a large value (let it be 30). So that we have such Selection Options Window



Now we run the regression and view the results



Type 3 Analysis

Type 3 Analysis of Effects				
Effect	Wald			
	DF	Chi-Square	Pr > ChiSq	
G	3	316.2179	<.0001	
M_C_previous	1	195.8588	<.0001	
cost	1	272.0308	<.0001	
homeowner	1	68.9907	<.0001	
shopping_pt	12	30999.2052	<.0001	
state	35	1060.7980	<.0001	
time	23	678.4634	<.0001	

Analysis of Maximum Likelihood Estimates

Output

Analysis of Maximum Likelihood Estimates							
	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
2153	Intercept	1	-0.2712	9.3668	0.00	0.9769	0.762
2154	G	1	-0.2173	0.0124	305.80	<.0001	0.805
2155	G	2	0.0165	0.0103	2.55	0.1102	1.017
2156	G	3	0.0971	0.0116	70.61	<.0001	1.102
2157	M_C_previous	0	0.3918	0.0280	195.86	<.0001	1.480
2158	cost	1	-0.00232	0.000141	272.03	<.0001	-0.0589
2159	homeowner	0	-0.0495	0.00596	68.99	<.0001	0.952
2160	shopping_pt	1	-11.9351	7.9386	2.26	0.1327	0.000
2161	shopping_pt	2	-11.9192	7.9260	2.26	0.1326	0.000
2162	shopping_pt	3	-1.4003	7.2363	0.04	0.8466	0.247
2163	shopping_pt	4	-0.9602	7.2363	0.02	0.8944	0.383
2164	shopping_pt	5	-0.4618	7.2363	0.00	0.9491	0.630
2165	shopping_pt	6	0.1100	7.2363	0.00	0.9879	1.116
2166	shopping_pt	7	0.6955	7.2363	0.01	0.9234	2.005
2167	shopping_pt	8	1.2434	7.2363	0.03	0.8636	3.468
2168	shopping_pt	9	1.7638	7.2363	0.06	0.8074	5.834
2169	shopping_pt	10	2.2623	7.2364	0.10	0.7546	9.605
2170	shopping_pt	11	2.8754	7.2366	0.16	0.6911	17.732
2171	shopping_pt	12	3.9076	7.2392	0.29	0.5893	49.778
2172	state	AL	0.1001	0.0307	10.64	0.0011	1.105
2173	state	AR	0.1517	0.0442	11.78	0.0006	1.164
2174	state	CO	-0.1373	0.0308	19.81	<.0001	0.872
2175	state	CT	0.2139	0.0343	38.81	<.0001	1.238
2176	state	DC	0.1448	0.0730	3.93	0.0474	1.156
2177	state	DE	0.1359	0.0703	3.74	0.0532	1.146
2178	state	FL	-0.4209	0.0211	397.31	<.0001	0.656
2179	state	GA	-0.1073	0.0501	4.58	0.0323	0.898
2180	state	IA	-0.00740	0.0662	0.01	0.9111	0.993
2181	state	ID	-0.2594	0.0596	18.94	<.0001	0.771
2182	state	IN	0.1365	0.0301	20.62	<.0001	1.146
2183	state	KS	0.1269	0.0593	4.58	0.0324	1.135
2184	state	KY	0.1870	0.0366	26.05	<.0001	1.206
2185	state	MD	0.1446	0.0287	25.34	<.0001	1.156
2186	state	ME	0.0679	0.0566	1.44	0.2301	1.070
2187	state	MO	-0.1142	0.0378	9.10	0.0026	0.892
2188	state	MS	0.0265	0.0444	0.36	0.5503	1.027
2189	state	MT	-0.3077	0.0816	14.21	0.0002	0.735
2190	state	ND	0.3689	0.0937	15.51	<.0001	1.446
2191	state	NE	-0.0205	0.0832	0.06	0.8052	0.980
2192	state	NH	-0.0621	0.0518	1.44	0.2304	0.940
2193	state	NM	0.0794	0.0532	2.23	0.1351	1.083
2194	state	NV	-0.3811	0.0392	94.43	<.0001	0.683
2195	state	NY	0.1310	0.0188	48.57	<.0001	1.140
2196	state	OH	-0.0276	0.0252	1.20	0.2727	0.973
2197	state	OK	-0.0751	0.0401	3.51	0.0608	0.928
2198	state	OR	-0.2136	0.0390	29.92	<.0001	0.808
2199	state	PA	0.0371	0.0212	3.07	0.0799	1.038
2200	state	RI	0.1771	0.0593	8.93	0.0028	1.194
2201	state	SD	0.3835	0.1924	3.97	0.0462	1.467
2202	state	TN	0.0354	0.0352	1.01	0.3148	1.036
2203	state	UT	-0.2919	0.0389	56.29	<.0001	0.747
2204	state	WA	-0.3607	0.0315	131.38	<.0001	0.697
2205	state	WI	0.0408	0.0454	0.81	0.3695	1.042
2206	state	WV	0.2255	0.0628	12.90	0.0003	1.253
2207	time	0	2.1081	5.9721	0.12	0.7241	8.233
2208	time	1	0.3741	5.9806	0.00	0.9501	1.454
2209	time	2	1.7095	6.0309	0.08	0.7768	5.526
2210	time	3	-10.3878	136.8	0.01	0.9395	0.000
2211	time	4	1.4941	5.9798	0.06	0.8027	4.455
2212	time	5	0.1065	5.9753	0.00	0.9658	1.112
2213	time	6	1.0966	5.9537	0.03	0.8539	2.994
2214	time	7	0.4853	5.9473	0.01	0.9350	1.625
2215	time	8	0.0100	5.9468	0.00	0.9987	1.010
2216	time	9	-0.0948	5.9468	0.00	0.9873	0.910
2217	time	10	-0.0964	5.9468	0.00	0.9871	0.908
2218	time	11	-0.0676	5.9468	0.00	0.9909	0.935
2219	time	12	-0.0345	5.9468	0.00	0.9954	0.966
2220	time	13	-0.0658	5.9468	0.00	0.9912	0.936
2221	time	14	-0.0469	5.9468	0.00	0.9937	0.954
2222	time	15	0.00840	5.9468	0.00	0.9989	1.008
2223	time	16	0.2184	5.9468	0.00	0.9707	1.244
2224	time	17	0.3696	5.9468	0.00	0.9504	1.447
2225	time	18	0.4367	5.9469	0.01	0.9415	1.548
2231	time	19	0.4979	5.9470	0.01	0.9333	1.645
2232	time	20	0.4846	5.9475	0.01	0.9351	1.624
2233	time	21	0.3384	5.9485	0.00	0.9546	1.403
2234	time	22	0.5894	5.9499	0.01	0.9211	1.803

Odds Ratio Estimates

Output

2237		Odds Ratio Estimates	
2238			
2239		Point	
2240	Effect	Estimate	
2241			
2242	G	1 vs 4	0.725
2243	G	2 vs 4	0.916
2244	G	3 vs 4	0.993
2245	M_C_previous	0 vs 1	2.189
2246	cost		0.998
2247	homeowner	0 vs 1	0.906
2248	shopping_pt	1 vs 13	<0.001
2249	shopping_pt	2 vs 13	<0.001
2250	shopping_pt	3 vs 13	<0.001
2251	shopping_pt	4 vs 13	<0.001
2252	shopping_pt	5 vs 13	<0.001
2253	shopping_pt	6 vs 13	<0.001
2254	shopping_pt	7 vs 13	<0.001
2255	shopping_pt	8 vs 13	<0.001
2256	shopping_pt	9 vs 13	<0.001
2257	shopping_pt	10 vs 13	<0.001
2258	shopping_pt	11 vs 13	<0.001
2259	shopping_pt	12 vs 13	<0.001
2260	state	AL vs WY	1.256
2261	state	AR vs WY	1.323
2262	state	CO vs WY	0.991
2263	state	CT vs WY	1.407
2264	state	DC vs WY	1.313
2265	state	DE vs WY	1.302
2266	state	FL vs WY	0.746
2267	state	GA vs WY	1.021
2268	state	IA vs WY	1.128
2269	state	ID vs WY	0.877
2270	state	IN vs WY	1.303
2271	state	KS vs WY	1.290
2272	state	KY vs WY	1.370
2273	state	MD vs WY	1.313
2274	state	ME vs WY	1.216
2275	state	MO vs WY	1.014

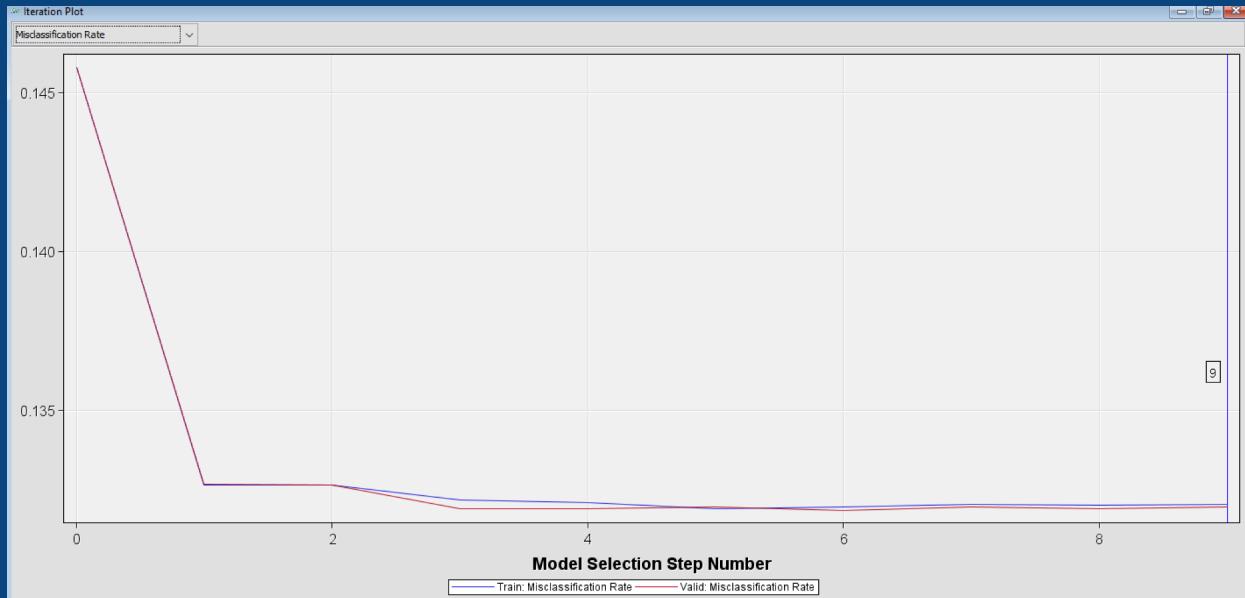
2276	state	MS vs WY	1.167
2277	state	MT vs WY	0.835
2278	state	ND vs WY	1.643
2279	state	NE vs WY	1.113
2280	state	NH vs WY	1.068
2281	state	NM vs WY	1.230
2282	state	NV vs WY	0.776
2283	state	NY vs WY	1.295
2284	state	OH vs WY	1.105
2285	state	OK vs WY	1.054
2286	state	OR vs WY	0.918
2287	state	PA vs WY	1.179
2288	state	RI vs WY	1.357
2289	state	SD vs WY	1.668
2290	state	TN vs WY	1.177
2291	state	UT vs WY	0.849
2292	state	WA vs WY	0.792
2293	state	WI vs WY	1.184
2294	state	WV vs WY	1.424
2295	time	0 vs 23	5.165
2296	time	1 vs 23	0.912
2297	time	2 vs 23	3.467
2298	time	3 vs 23	<0.001
2299	time	4 vs 23	2.795
2300	time	5 vs 23	0.698
2301	time	6 vs 23	1.878
2302	time	7 vs 23	1.019
2303	time	8 vs 23	0.634
2304	time	9 vs 23	0.571
2305	time	10 vs 23	0.570
2306	time	11 vs 23	0.586
2307	time	12 vs 23	0.606
2308	time	13 vs 23	0.587
2309	time	14 vs 23	0.599
2310	time	15 vs 23	0.633
2311	time	16 vs 23	0.781
2312	time	17 vs 23	0.908
2313	time	18 vs 23	0.971
2314	time	19 vs 23	1.032

2315	time	20 vs 23	1.019
2316	time	21 vs 23	0.880
2317	time	22 vs 23	1.131

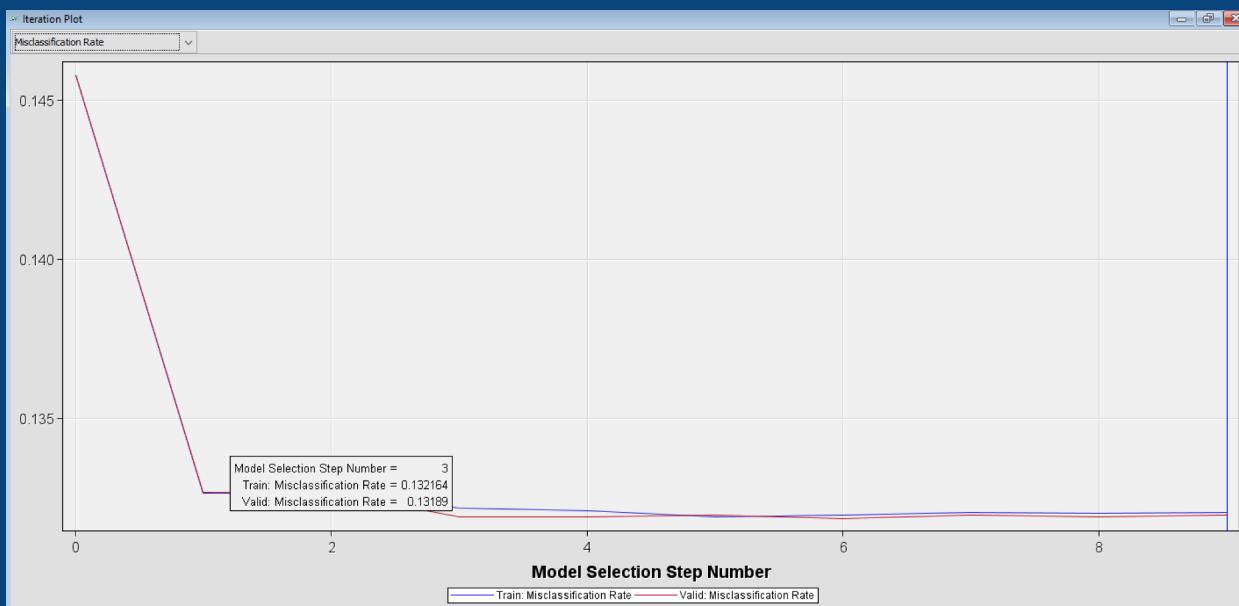
Fit statistics:

Fit Statistics						
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type		_AIC_	Akaike's Information Criterion	199851.5	-	-
record_type		_ASE_	Average Squared Error	0.094786	0.094796	-
record_type		_AVERR_	Average Error Function	0.300185	0.300427	-
record_type		_DFE_	Degrees of Freedom for Error	332547	-	-
record_type		_DFM_	Model Degrees of Freedom	77	-	-
record_type		_DFT_	Total Degrees of Freedom	332624	-	-
record_type		_DIV_	Divisor for ASE	665248	665250	-
record_type		_ERR_	Error Function	199897.5	199899	-
record_type		_FPE_	Final Prediction Error	0.09483	-	-
record_type		_MAX_	Maximum Absolute Error	0.982459	0.983275	-
record_type		_MSE_	Mean Square Error	0.094808	0.094796	-
record_type		_NOBS_	Sum of Frequencies	332624	332625	-
record_type		_NW_	Number of Estimate Weights	77	-	-
record_type		_RASE_	Root Average Sum of Squares	0.307873	0.30789	-
record_type		_RFPE_	Root Final Prediction Error	0.307944	-	-
record_type		_RMSE_	Root Mean Squared Error	0.307909	0.30789	-
record_type		_SBC_	Schwarz's Bayesian Criterion	200676.6	-	-
record_type		_SSE_	Sum of Squared Errors	63056.12	63063.25	-
record_type		_SUMW_	Sum of Case Weights Times Freq	665248	665250	-
record_type		_MISC_	Misclassification Rate	0.132026	0.131944	-

Iteration Plot:



We see, that the final step in our model is 9, but apparently an acceptable result appears at step 3:



It means we can try to configure the regression node to select the model with the smallest fit statistic rather than the final forward selection interaction. That is, we will optimize the complexity of stepwise regression.

Stepwise Regression, Adjusted Entry and Stay Levels, Best sequence model

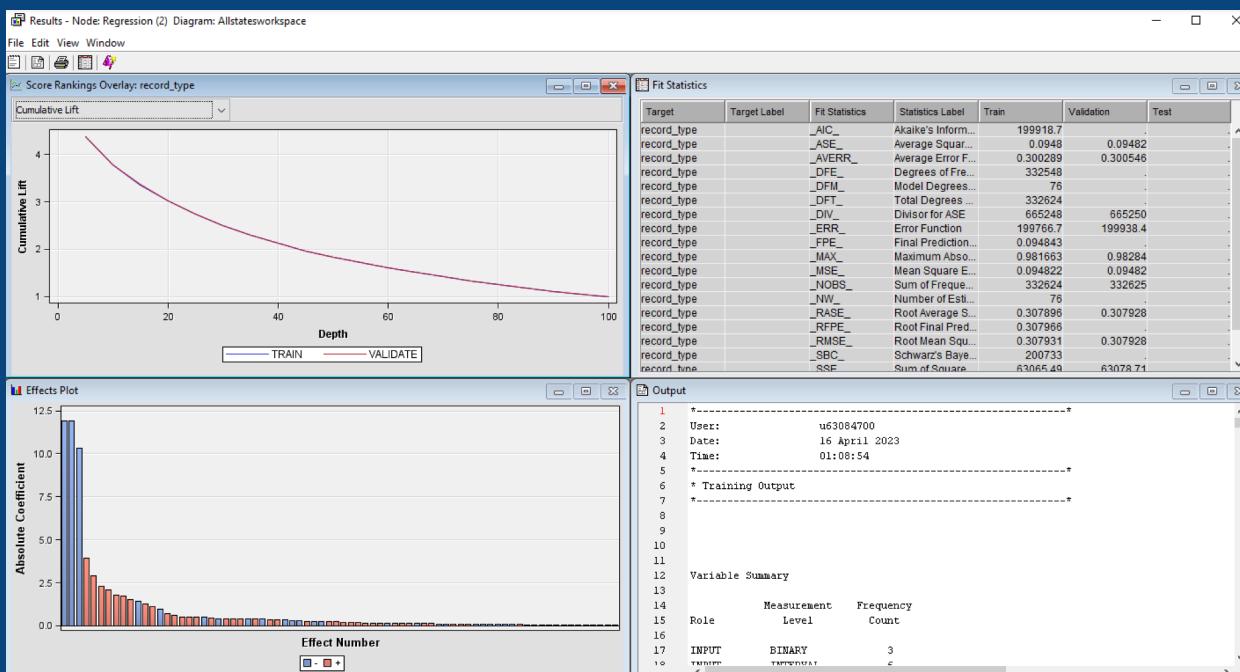
Getting back to the property panel of the stepwise regression node

Property	Value
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Default
Use Selection Defaults	No
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

Select Selection Criterion -> Validation Misclassification

Property	Value
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	Stepwise
Selection Criterion	Validation Misclassification
Use Selection Defaults	No
Selection Options	
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

Run the node and examine the results:



Type 3 Analysis

Type 3 Analysis of Effects				
		Wald		
2140	Effect	DF	Chi-Square	Pr > ChiSq
2141	G	3	326.7916	<.0001
2142	M_C_previous	1	197.1268	<.0001
2143	cost	1	365.1658	<.0001
2144	shopping_pt	12	30966.6859	<.0001
2145	state	35	1048.6589	<.0001
2146	time	23	680.2957	<.0001
2147				
2148				
2149				
2150				
2151				

Analysis of Maximum Likelihood Estimates

Output

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized		
						Estimate	Exp(Est)	
2158 Intercept	1	-0.0852	9.2625	0.00	0.9927	0.918		
2159 G	1	-0.2214	0.0124	317.89	<.0001	0.801		
2160 G	2	0.0133	0.0103	1.66	0.1972	1.013		
2161 G	3	0.0977	0.0116	71.53	<.0001	1.103		
2162 M_C_previous	0	0.3931	0.0280	197.13	<.0001	1.482		
2163 cost	1	-0.00261	0.000137	365.17	<.0001	-0.0662	0.997	
2164 shopping_pt	1	-11.9290	7.9316	2.26	0.1326	0.000		
2165 shopping_pt	2	-11.9108	7.9165	2.26	0.1325	0.000		
2166 shopping_pt	3	-1.3993	7.2338	0.04	0.8466	0.247		
2167 shopping_pt	4	-0.9598	7.2338	0.02	0.8944	0.383		
2168 shopping_pt	5	-0.4619	7.2338	0.00	0.9491	0.630		
2169 shopping_pt	6	0.1093	7.2338	0.00	0.9879	1.115		
2170 shopping_pt	7	0.6941	7.2338	0.01	0.9236	2.002		
2171 shopping_pt	8	1.2414	7.2338	0.03	0.8637	3.460		
2172 shopping_pt	9	1.7603	7.2339	0.06	0.8077	5.814		
2173 shopping_pt	10	2.2581	7.2339	0.10	0.7549	9.565		
2174 shopping_pt	11	2.8696	7.2341	0.16	0.6916	17.630		
2175 shopping_pt	12	3.9026	7.2367	0.29	0.5897	49.530		
2176 state	AL	0.1182	0.0306	14.93	0.0001	1.126		
2177 state	AR	0.1554	0.0442	12.36	0.0004	1.168		
2178 state	CO	-0.1390	0.0308	20.32	<.0001	0.870		
2179 state	CT	0.2076	0.0343	36.60	<.0001	1.231		
2180 state	DC	0.1318	0.0730	3.26	0.0709	1.141		
2181 state	DE	0.1470	0.0703	4.37	0.0365	1.158		
2182 state	FL	-0.4031	0.0210	368.39	<.0001	0.668		
2183 state	GA	-0.0918	0.0501	3.36	0.0669	0.912		
2184 state	IA	-0.0268	0.0662	0.16	0.6853	0.974		
2185 state	ID	-0.2768	0.0596	21.59	<.0001	0.758		
2186 state	IN	0.1298	0.0300	18.65	<.0001	1.139		
2187 state	KS	0.1244	0.0593	4.40	0.0359	1.133		
2188 state	KY	0.1854	0.0366	25.61	<.0001	1.204		
2189 state	MD	0.1436	0.0287	25.01	<.0001	1.154		
2190 state	ME	0.0571	0.0565	1.02	0.3122	1.059		
2191 state	MO	-0.1207	0.0378	10.19	0.0014	0.886		
2192 state	MS	0.0475	0.0443	1.15	0.2834	1.049		
2193 state	MT	-0.3147	0.0816	14.88	0.0001	0.730		
2194 state	ND	0.3473	0.0936	13.77	0.0002	1.415		
2195 state	NE	-0.0173	0.0832	0.04	0.8350	0.983		
2196 state	NH	-0.0649	0.0517	1.58	0.2095	0.937		
2197 state	NM	0.0955	0.0531	3.23	0.0723	1.100		
2198 state	NV	-0.3699	0.0392	89.10	<.0001	0.691		
2199 state	NY	0.1459	0.0187	60.80	<.0001	1.157		
2200 state	OH	-0.0309	0.0252	1.51	0.2169	0.970		
2201 state	OK	-0.0707	0.0400	3.12	0.0773	0.932		
2202 state	OR	-0.2251	0.0390	33.29	<.0001	0.798		
2203 state	PA	0.0451	0.0212	4.54	0.0331	1.046		
2204 state	RI	0.1808	0.0593	9.29	0.0023	1.198		
2205 state	SD	0.3848	0.1924	4.00	0.0455	1.469		
2206 state	TN	0.0371	0.0352	1.11	0.2911	1.038		
2207 state	UT	-0.3046	0.0389	61.43	<.0001	0.737		
2208 state	WA	-0.3634	0.0315	133.43	<.0001	0.695		
2209 state	WI	0.0255	0.0454	0.32	0.5741	1.026		
2210 state	WV	0.2429	0.0627	14.99	0.0001	1.275		
2211 time	0	2.0916	5.8103	0.13	0.7189	8.098		
2212 time	1	0.3875	5.8188	0.00	0.9469	1.473		
2213 time	2	1.6984	5.8708	0.08	0.7724	5.465		
2214 time	3	-10.2985	133.0	0.01	0.9383	0.000		
2215 time	4	1.5045	5.8182	0.07	0.7960	4.502		
2216 time	5	0.0980	5.8134	0.00	0.9865	1.103		
2217 time	6	1.0972	5.7913	0.04	0.8497	2.996		
2218 time	7	0.4800	5.7847	0.01	0.9339	1.616		
2219 time	8	0.00316	5.7843	0.00	0.9996	1.003		
2220 time	9	-0.1007	5.7842	0.00	0.9861	0.904		
2221 time	10	-0.0999	5.7842	0.00	0.9862	0.905		
2222 time	11	-0.0721	5.7842	0.00	0.9901	0.930		
2223 time	12	-0.0389	5.7842	0.00	0.9946	0.962		
2224 time	13	-0.0700	5.7842	0.00	0.9903	0.932		
2225 time	14	-0.0510	5.7842	0.00	0.9930	0.950		
2226 time	15	0.00467	5.7842	0.00	0.9994	1.005		
2227 time	16	0.2137	5.7842	0.00	0.9705	1.238		
2228 time	17	0.3667	5.7842	0.00	0.9494	1.443		
2229 time	18	0.4334	5.7843	0.01	0.9403	1.543		
2230 time	19	0.4957	5.7844	0.01	0.9317	1.642		
2231 time	20	0.4784	5.7850	0.01	0.9341	1.613		
2232 time	21	0.3402	5.7860	0.00	0.9531	1.405		
2233 time	22	0.5841	5.7875	0.01	0.9196	1.793		

Odds Ratio Estimates

Output

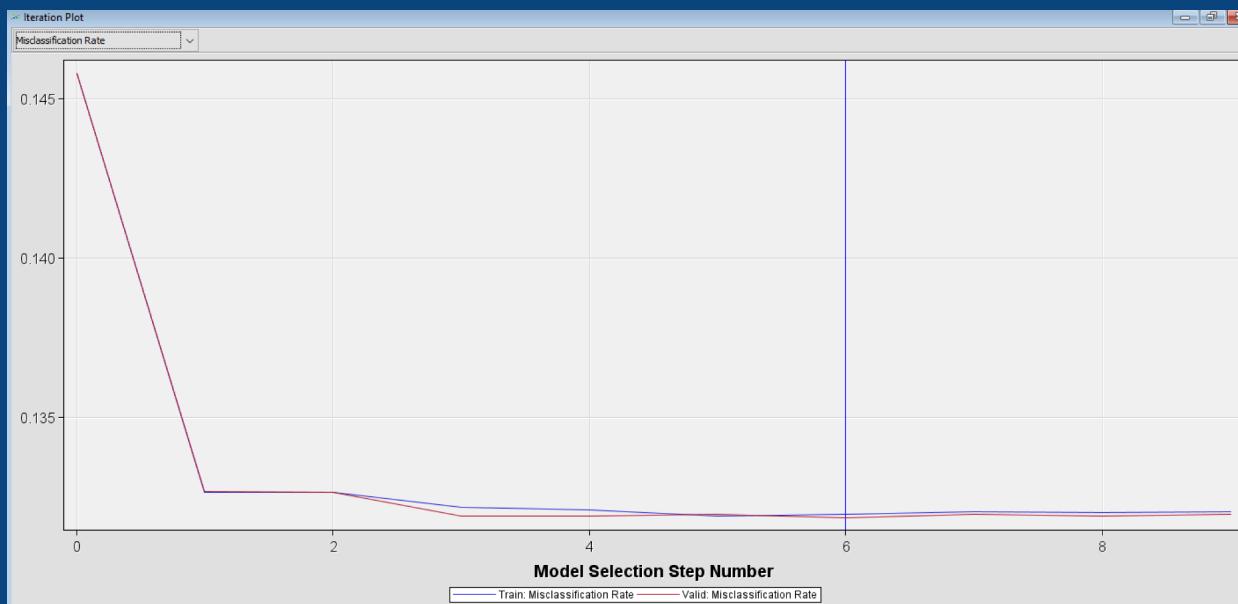
	Odds Ratio Estimates	
		Point Estimate
2236		
2237		
2238		Point
2239	Effect	Estimate
2240		
2241	G	1 vs 4 0.718
2242	G	2 vs 4 0.907
2243	G	3 vs 4 0.987
2244	M_C_previous	0 vs 1 2.195
2245	cost	0.997
2246	shopping_pt	1 vs 13 <0.001
2247	shopping_pt	2 vs 13 <0.001
2248	shopping_pt	3 vs 13 <0.001
2249	shopping_pt	4 vs 13 <0.001
2250	shopping_pt	5 vs 13 <0.001
2251	shopping_pt	6 vs 13 <0.001
2252	shopping_pt	7 vs 13 <0.001
2253	shopping_pt	8 vs 13 <0.001
2254	shopping_pt	9 vs 13 <0.001
2255	shopping_pt	10 vs 13 <0.001
2256	shopping_pt	11 vs 13 <0.001
2257	shopping_pt	12 vs 13 <0.001
2258	state	AL vs WY 1.285
2259	state	AR vs WY 1.334
2260	state	CO vs WY 0.994
2261	state	CT vs WY 1.405
2262	state	DC vs WY 1.303
2263	state	DE vs WY 1.323
2264	state	FL vs WY 0.763
2265	state	GA vs WY 1.042
2266	state	IA vs WY 1.112
2267	state	ID vs WY 0.866
2268	state	IN vs WY 1.300
2269	state	KS vs WY 1.293
2270	state	KY vs WY 1.374
2271	state	MD vs WY 1.318
2272	state	ME vs WY 1.209
2273	state	MO vs WY 1.012
2274	state	MS vs WY 1.197

2275	state	MT vs WY	0.834
2276	state	ND vs WY	1.616
2277	state	NE vs WY	1.122
2278	state	NH vs WY	1.070
2279	state	NM vs WY	1.256
2280	state	NV vs WY	0.789
2281	state	NY vs WY	1.321
2282	state	OH vs WY	1.107
2283	state	OK vs WY	1.064
2284	state	OR vs WY	0.912
2285	state	PA vs WY	1.195
2286	state	RI vs WY	1.368
2287	state	SD vs WY	1.678
2288	state	TN vs WY	1.185
2289	state	UT vs WY	0.842
2290	state	WA vs WY	0.794
2291	state	WI vs WY	1.171
2292	state	WV vs WY	1.456
2293	time	0 vs 23	5.143
2294	time	1 vs 23	0.936
2295	time	2 vs 23	3.471
2296	time	3 vs 23	<0.001
2297	time	4 vs 23	2.859
2298	time	5 vs 23	0.701
2299	time	6 vs 23	1.903
2300	time	7 vs 23	1.026
2301	time	8 vs 23	0.637
2302	time	9 vs 23	0.574
2303	time	10 vs 23	0.575
2304	time	11 vs 23	0.591
2305	time	12 vs 23	0.611
2306	time	13 vs 23	0.592
2307	time	14 vs 23	0.604
2308	time	15 vs 23	0.638
2309	time	16 vs 23	0.786
2310	time	17 vs 23	0.916
2311	time	18 vs 23	0.980
2312	time	19 vs 23	1.043
2313	time	20 vs 23	1.025
2314	time	21 vs 23	0.892
2315	time	22 vs 23	1.139

Fit statistics:

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
record_type		_AIC_	Akaike's Information Criterion	198918.7		
record_type		_ASE_	Average Squared Error	0.0948	0.09482	
record_type		_AVERR_	Average Error Function	0.300289	0.300546	
record_type		_DFE_	Degrees of Freedom for Error	332548		
record_type		_DFM_	Model Degrees of Freedom	76		
record_type		_DFT_	Total Degrees of Freedom	332624		
record_type		_DIV_	Divisor for ASE	665248	665250	
record_type		_ERR_	Error Function	199766.7	199938.4	
record_type		_FPE_	Final Prediction Error	0.094843		
record_type		_MAX_	Maximum Absolute Error	0.981663	0.98284	
record_type		_MSE_	Mean Square Error	0.094822	0.09482	
record_type		_NOBS_	Sum of Frequencies	332624	332625	
record_type		_NW_	Number of Estimate Weights	76		
record_type		_RASE_	Root Average Sum of Squares	0.307896	0.307928	
record_type		_RFPE_	Root Final Prediction Error	0.307966		
record_type		_RMSE_	Root Mean Squared Error	0.307931	0.307928	
record_type		_SBC_	Schwarz's Bayesian Criterion	200733		
record_type		_SSE_	Sum of Squared Errors	63065.49	63078.71	
record_type		_SUMW_	Sum of Case Weights Times Freq	665248	665250	
record_type		_MISC_	Misclassification Rate	0.131954	0.131857	

Iteration Plot:

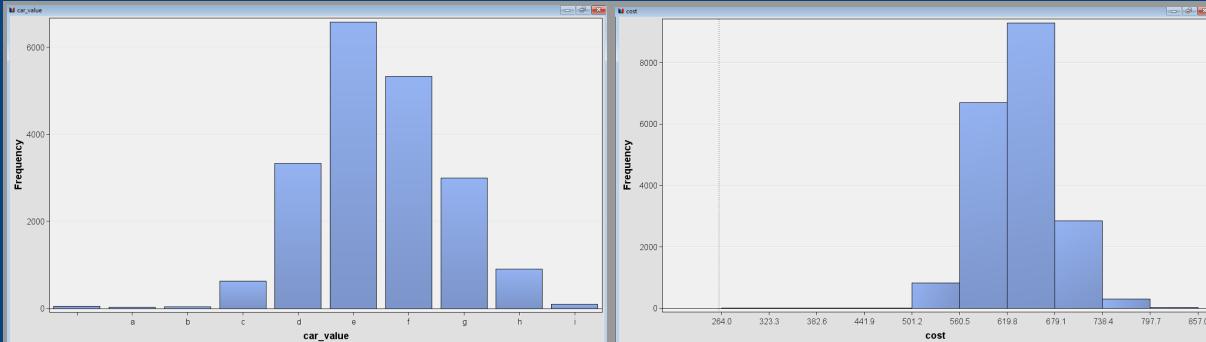


Question 1: Can we decode any categorical inputs so that we can reduce dummy variables?

Answer -> It seems like we don't know enough about the business context to decide if we can recode categorical variables.

Question 2: Do we need to run any log transformations of data?

Answer -> Only car_value and cost contain something that slightly reminds us of skewed distribution:



Nevertheless, we will assume there is no skew to avoid overcomplicated phases of interpreting the results of prediction.