

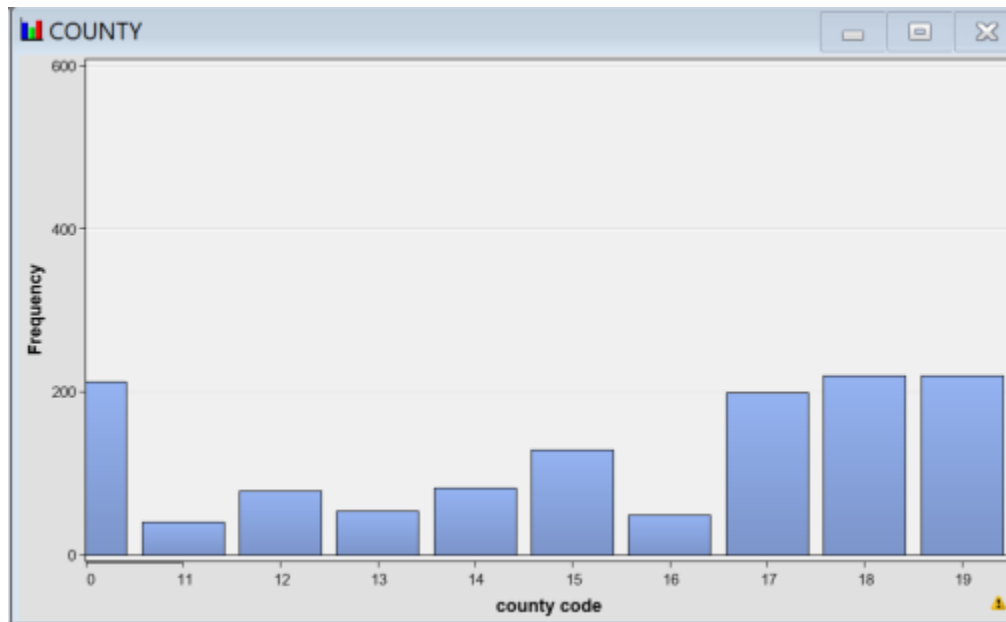
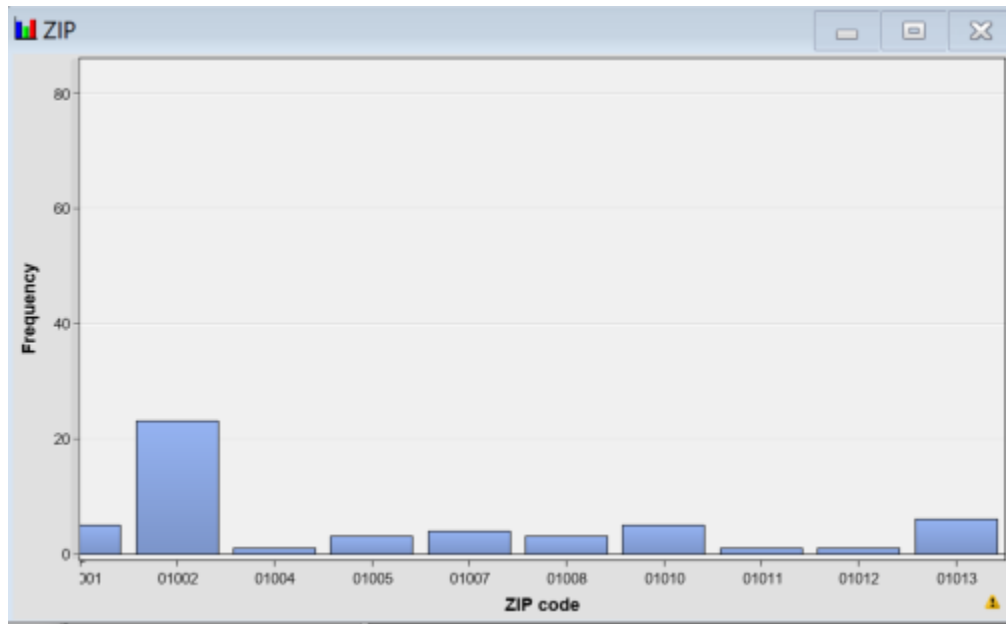
## Project 2

Ben Baker

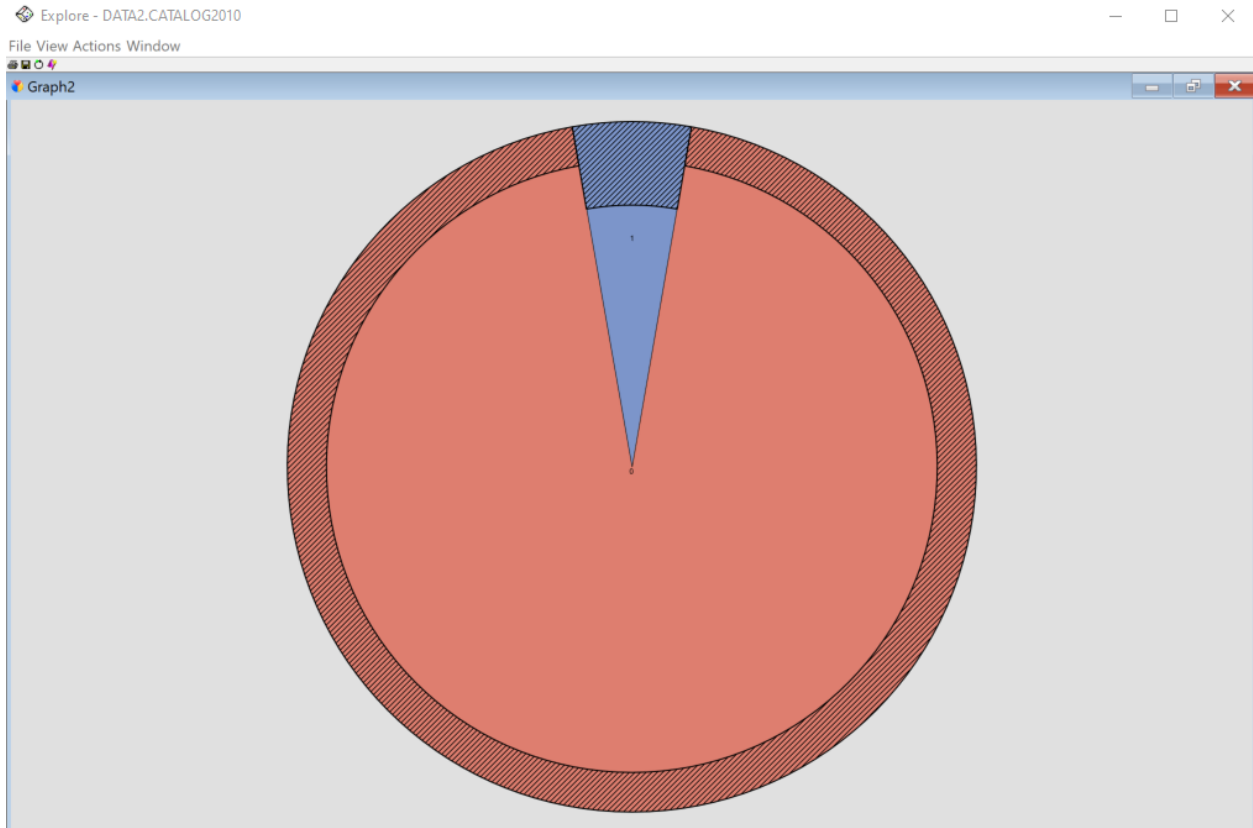
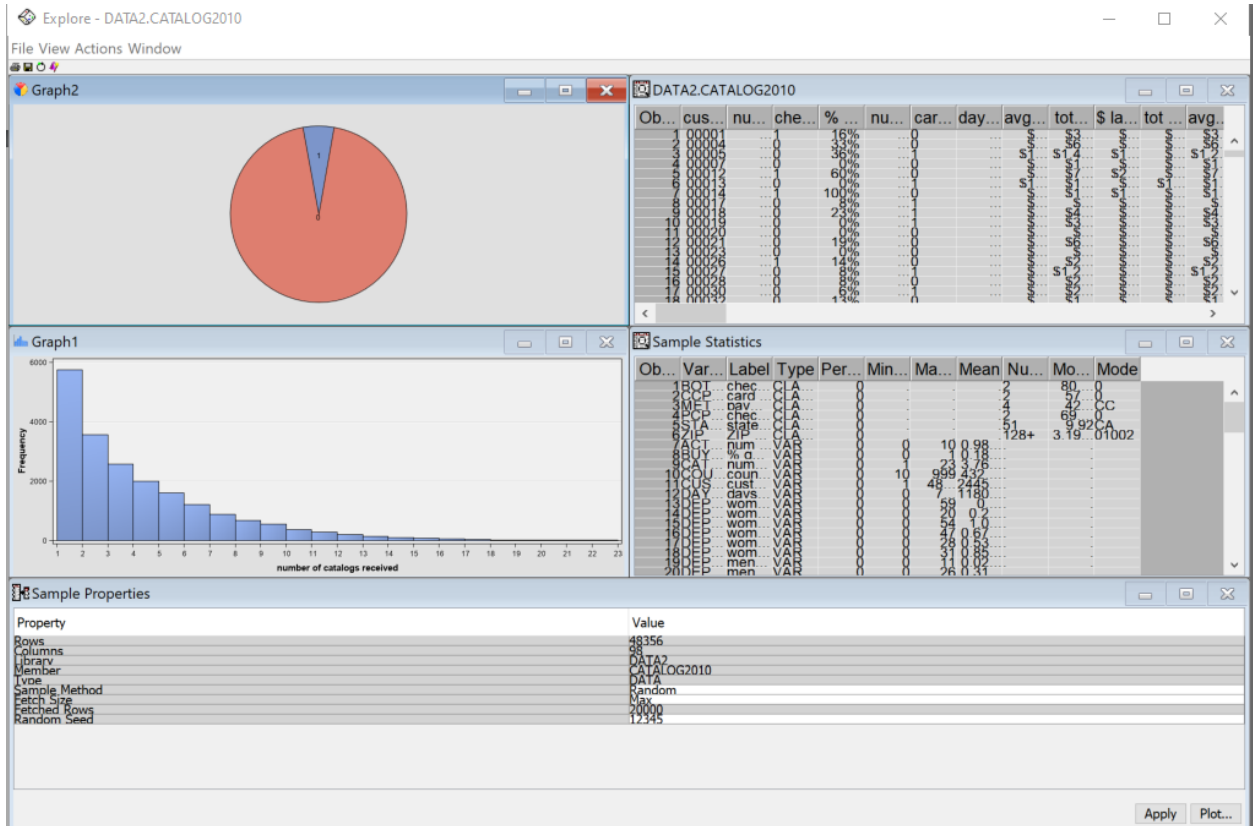
Robbie Brouillard

### Part 1-Predictive Analytics

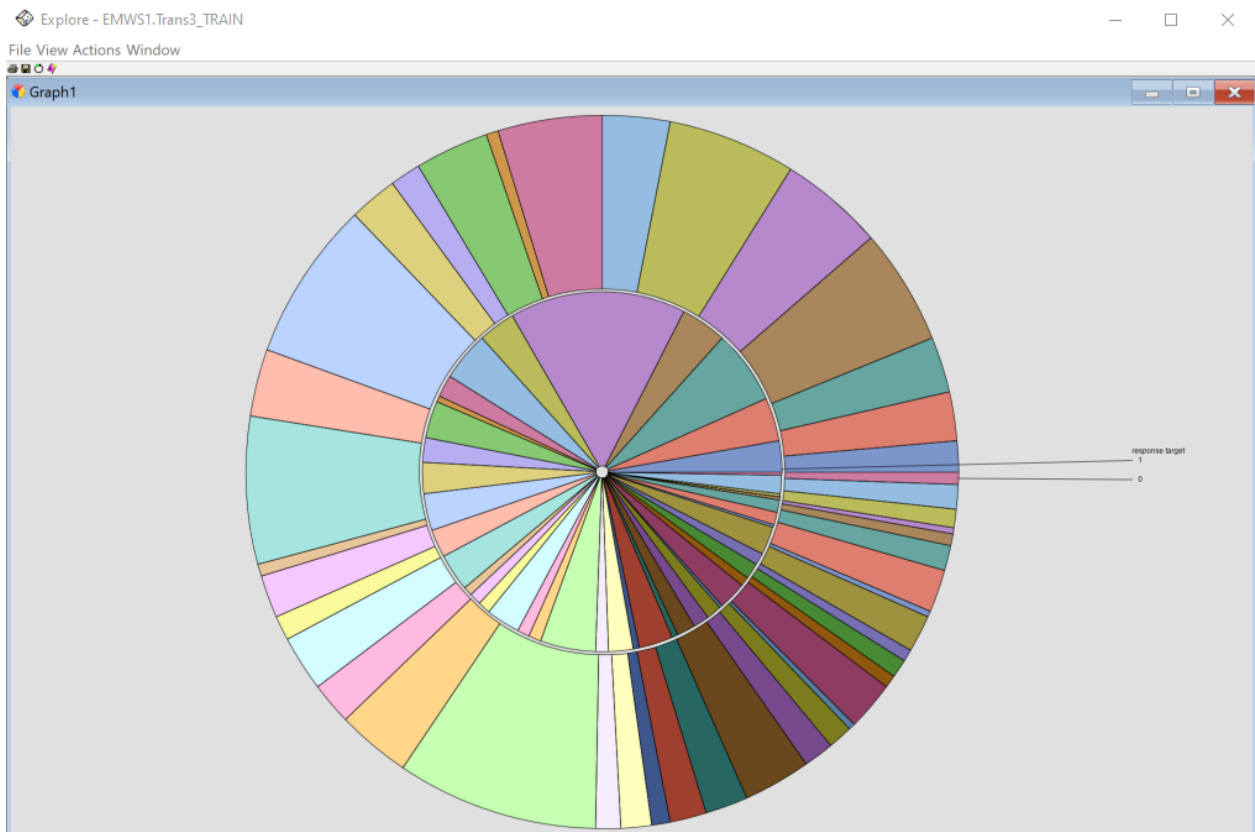
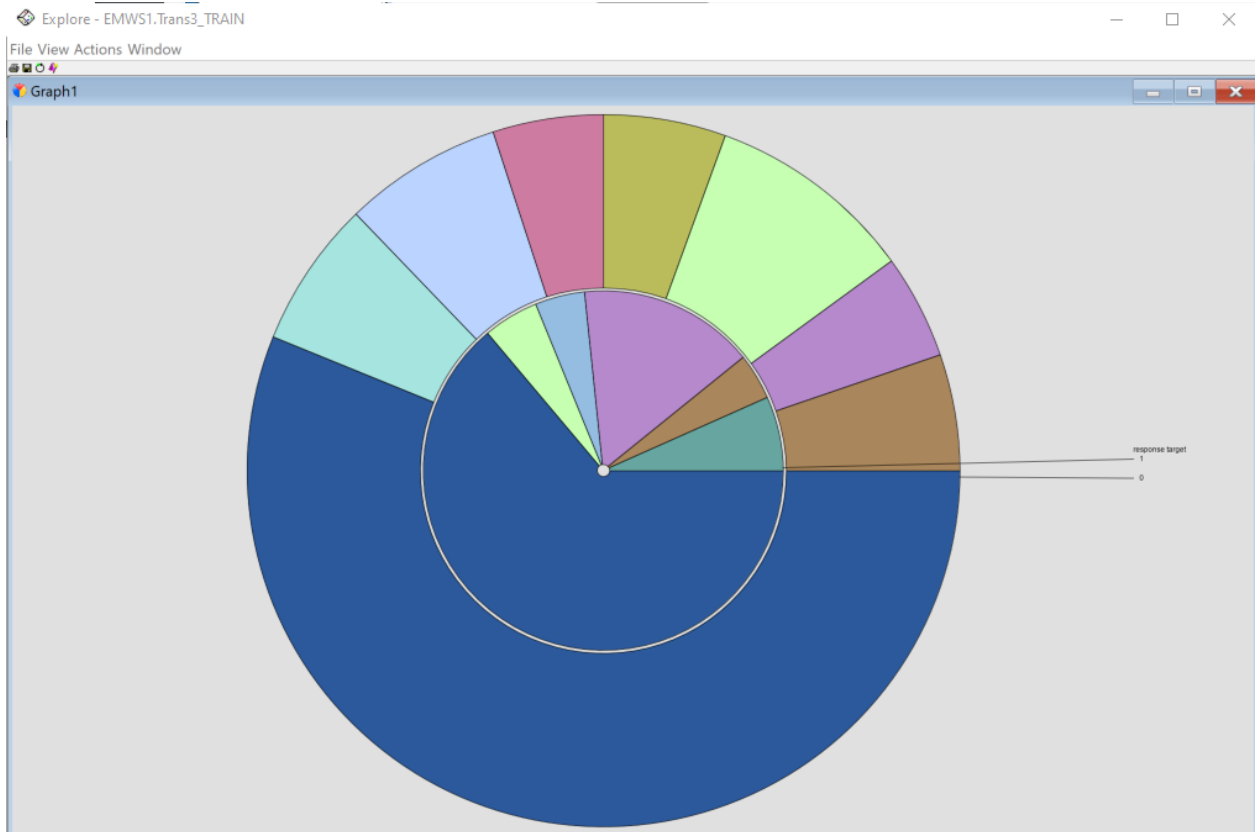
(Tutorial):

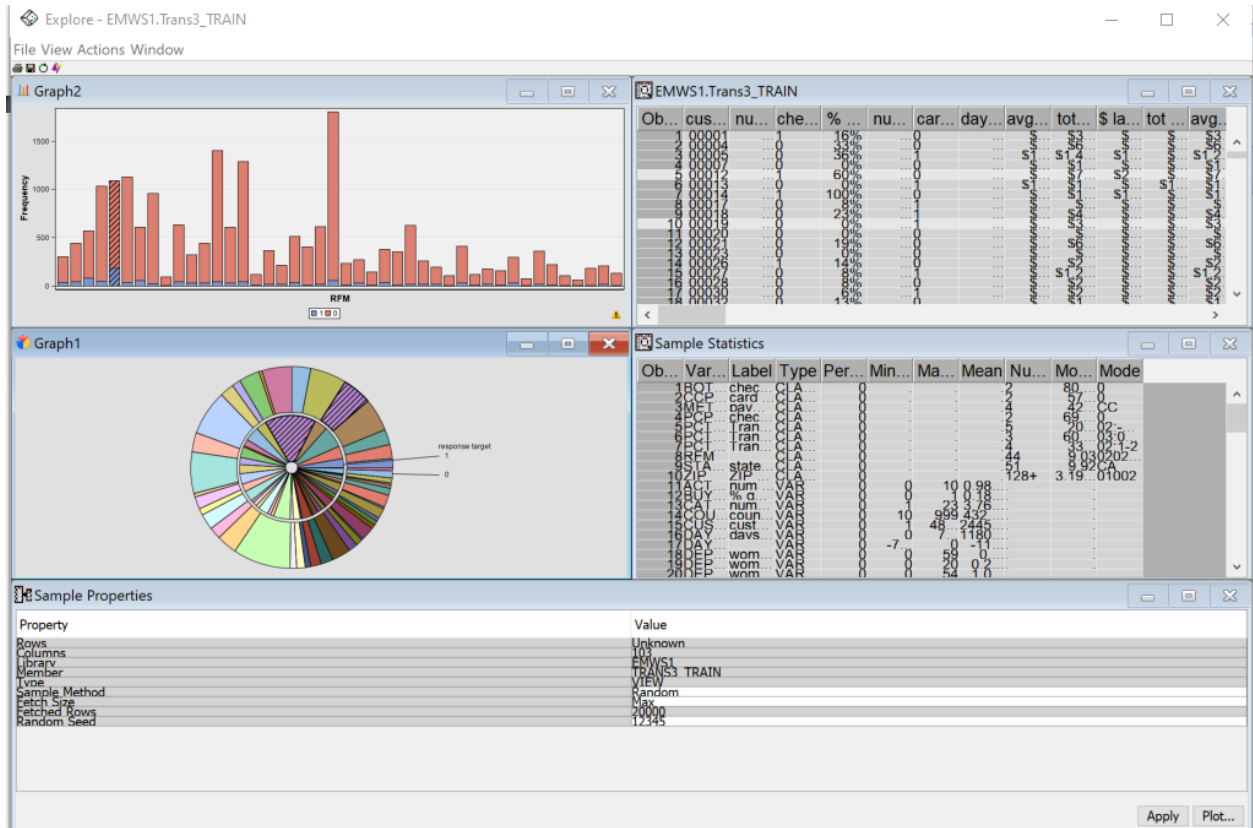
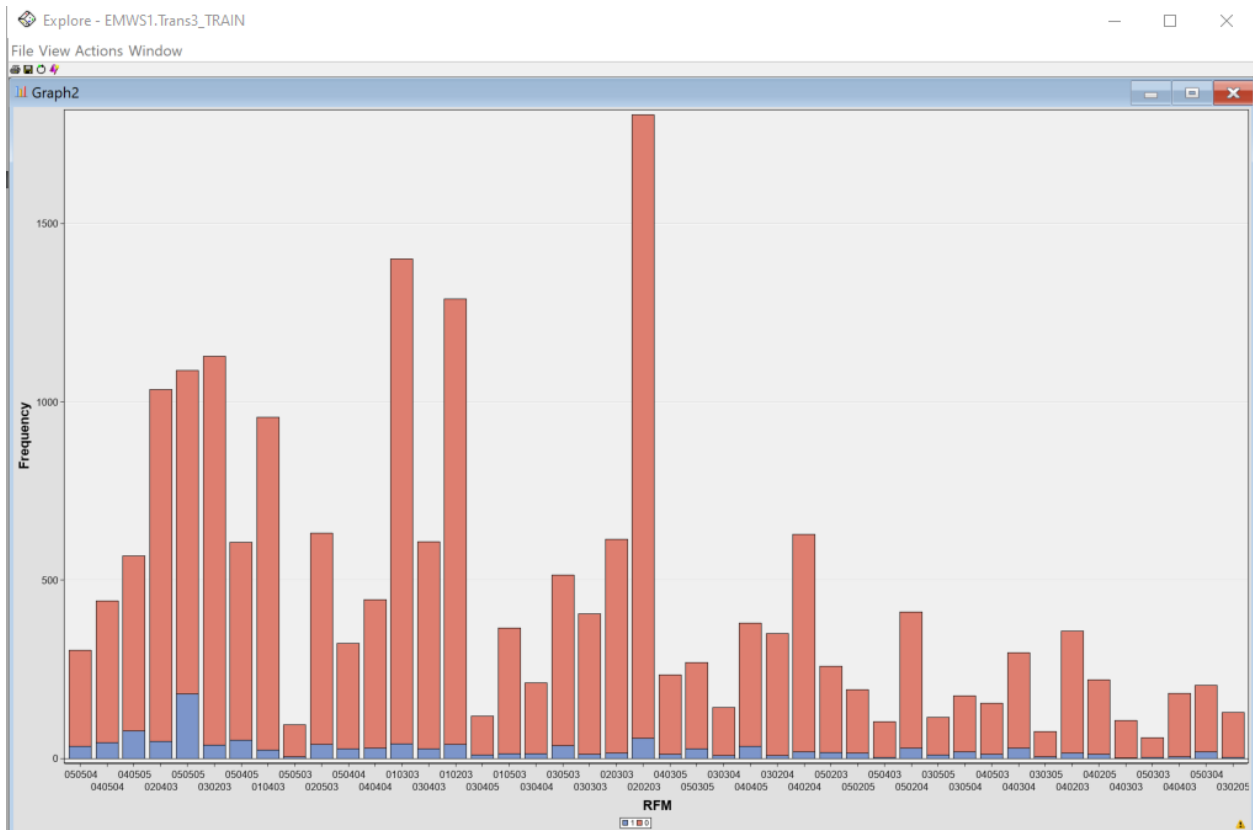






DAYLAS...	Transfor...	Transfor...	Transfor...	
-14605:-256-high	04:24.75-7...	05:6-high	05:6-high	^
-27604:-573--256	04:24.75-7...	05:6-high	05:6-high	
-56404:-573--256	05:71.6-high	05:6-high	05:6-high	
-131202:-1826--1...	03:0-24.75	04:3-6	04:3-6	
-10105:-256-high	05:71.6-high	05:6-high	05:6-high	
-78203:-1008--5...	03:0-24.75	02:1-2	02:1-2	
-13705:-256-high	05:71.6-high	04:3-6	04:3-6	
-131702:-1826--1...	03:0-24.75	04:3-6	04:3-6	
-5005:-256-high	04:24.75-7...	05:6-high	05:6-high	
-17305:-256-high	05:71.6-high	05:6-high	05:6-high	
-442401:low--1826	03:0-24.75	04:3-6	04:3-6	
-31304:-573--256	05:71.6-high	05:6-high	05:6-high	
-311201:low--1826	03:0-24.75	04:3-6	04:3-6	
-19305:-256-high	03:0-24.75	05:6-high	05:6-high	
-152502:-1826--1...	03:0-24.75	05:6-high	05:6-high	
-130302:-1826--1...	03:0-24.75	04:3-6	04:3-6	
-7105:-256-high	04:24.75-7...	04:3-6	04:3-6	
-290401:low--1826	03:0-24.75	04:3-6	04:3-6	v
				>





**(Task 1):**

Training Code - Code Node

File Edit Run View

Macro

**Train**

- Utility
  - EM\_REGISTER
  - EM\_REPORT
  - EM\_DATA2CODE
  - EM\_DECDATA
  - EM\_CHECKMACRO
  - EM\_CHECKSETINIT
  - EM\_ODSLISTON
  - EM\_ODSLISTOFF
- Variables
  - EM\_INTERVAL
  - EM\_CLASS
  - EM\_TARGET
  - EM\_TARGET\_LEVEL
  - EM\_PRIMARY\_TARGET

Macros Macro Variables Variables

Training Code

```
data catalog;
  set 'C:\Users\bbaker48\Desktop\project_mbad6201\SAS_DATA_BA\catalog2010.sas7bdat';
run;

proc univariate data=catalog normal plot;
  var monlast dolindet;
run;

/*
monlast: 52+(1.5*43)=116.5
dolindet: 223.2+(1.5*179.2)=492
*/

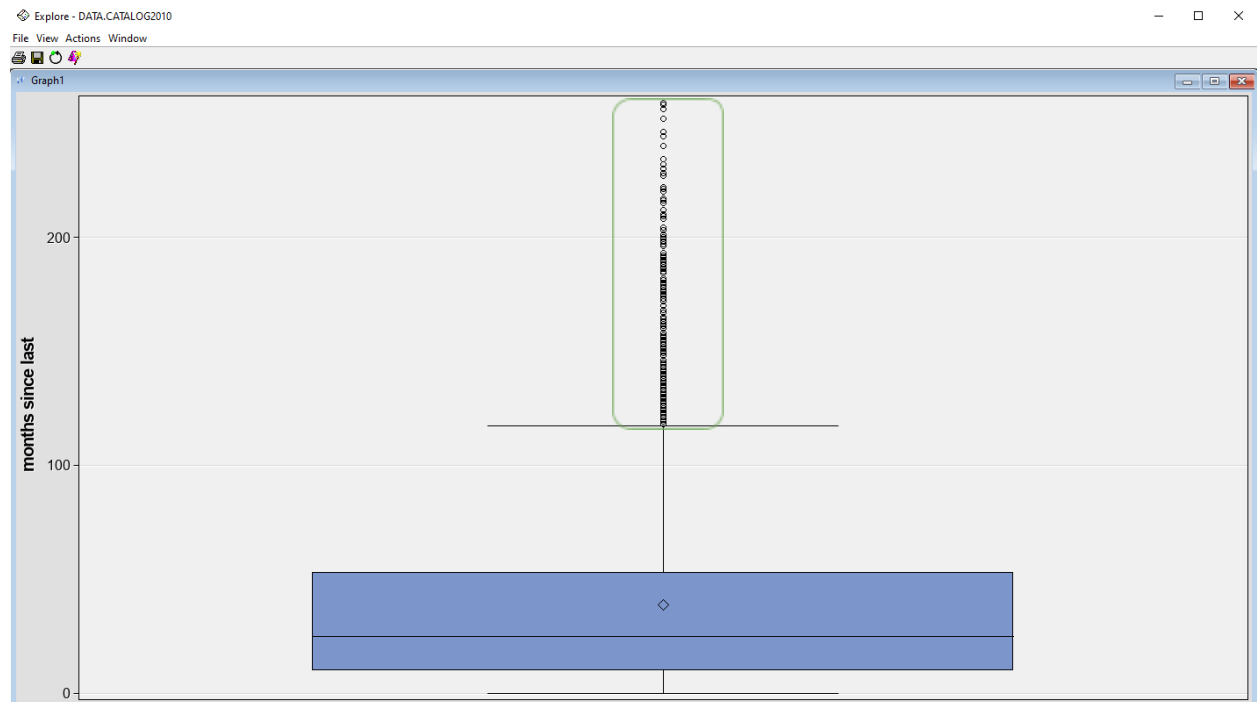
data mod_catalog;
  set catalog;
  if monlast>116.5 then delete;
  if dolindet>492 then delete;
run;
```

Output Log Result Log

1  
2

bbaker48 as bbaker48 - project2\_part1 - Exploratory Analysis - EMCODE - Complete

Type here to search

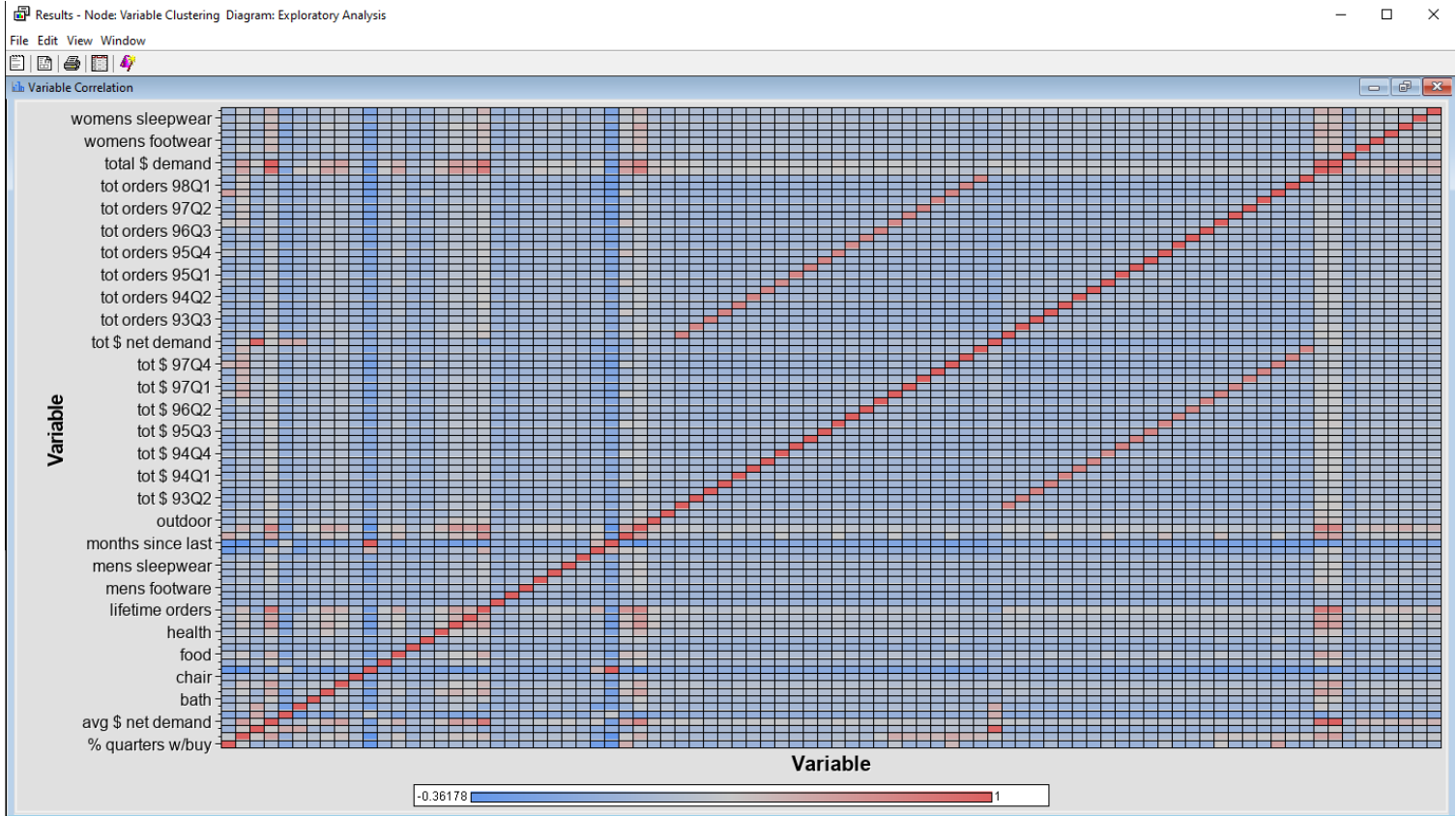


This is a box plot for all values of 'MONLAST' in the data, all values greater than  $(Q3 + 1.5 \cdot (IQR)) = 116.5$  are considered outliers and are circled in green. We then removed these outliers in the sas code above.



This is a box plot for all values of 'DAYLAST' in the data, all values greater than  $(Q3 + 1.5 \cdot (IQR)) = 492$  are considered outliers and are circled in green. We then removed these outliers in the sas code above.

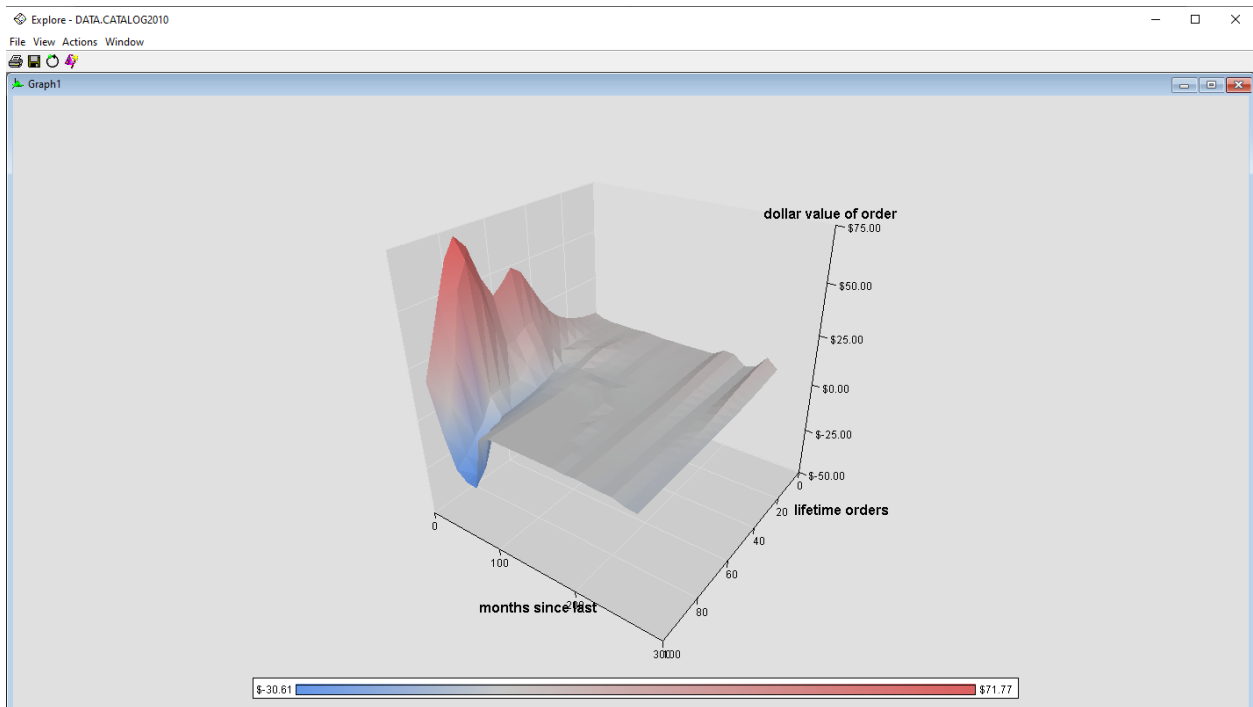
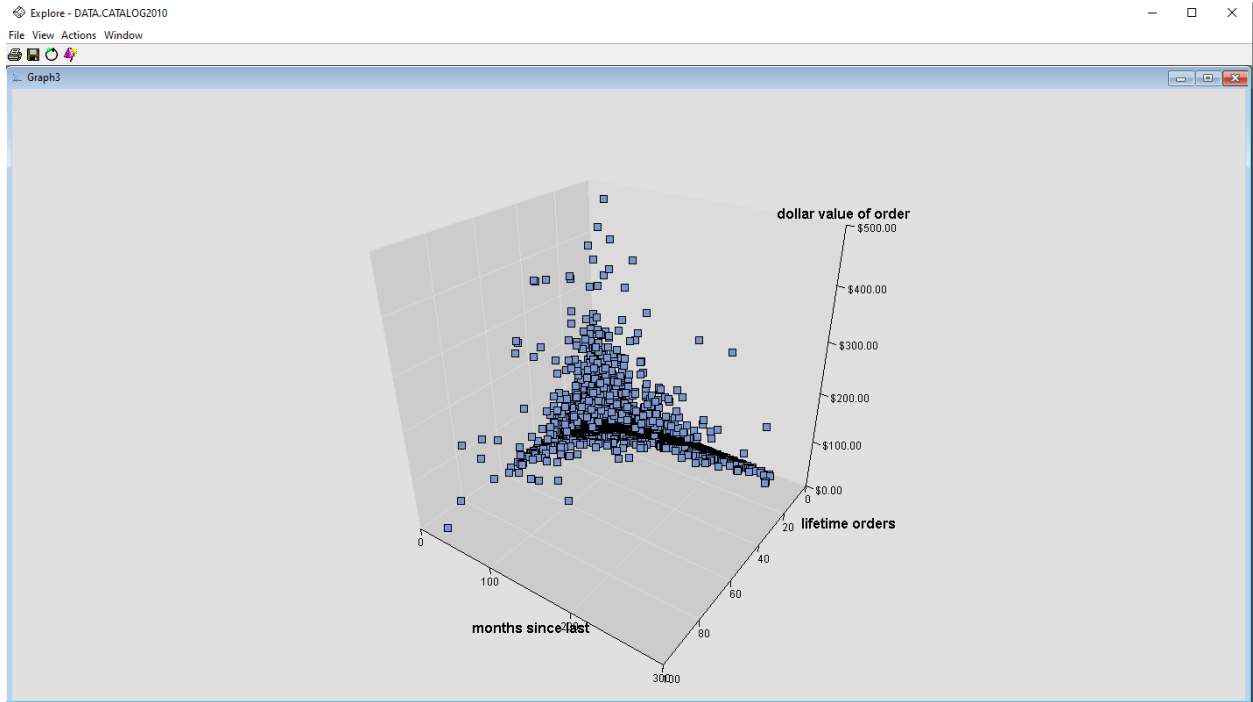




- List of Highly Correlated Variables

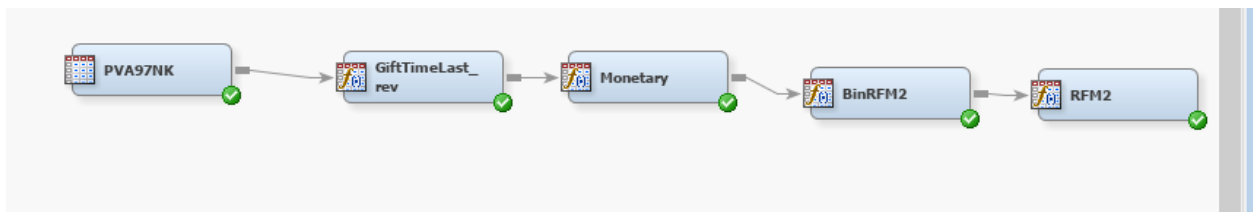
- MONLAST
- DAYLAST
- DOLINET
- DOLNETDT
- DOLINDEA
- DOLNETDA

- It may be possible to drop 'daylast' as it is perfectly collinear with 'monlast'
- it may also be possible to drop 'DOLNET' as it is perfectly collinear with 'DDOLNETDT'



(Task 2):

B.)



**Formulas**

Frequency

Age

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Method	Number of Bins
DemAge	Input	Interval	DEFAULT	4.0
DemCluster	Input	Nominal	DEFAULT	4.0
DemGender	Input	Nominal	DEFAULT	4.0
DemHomeOwn	Input	Binary	DEFAULT	4.0

**Inputs**

Frequency

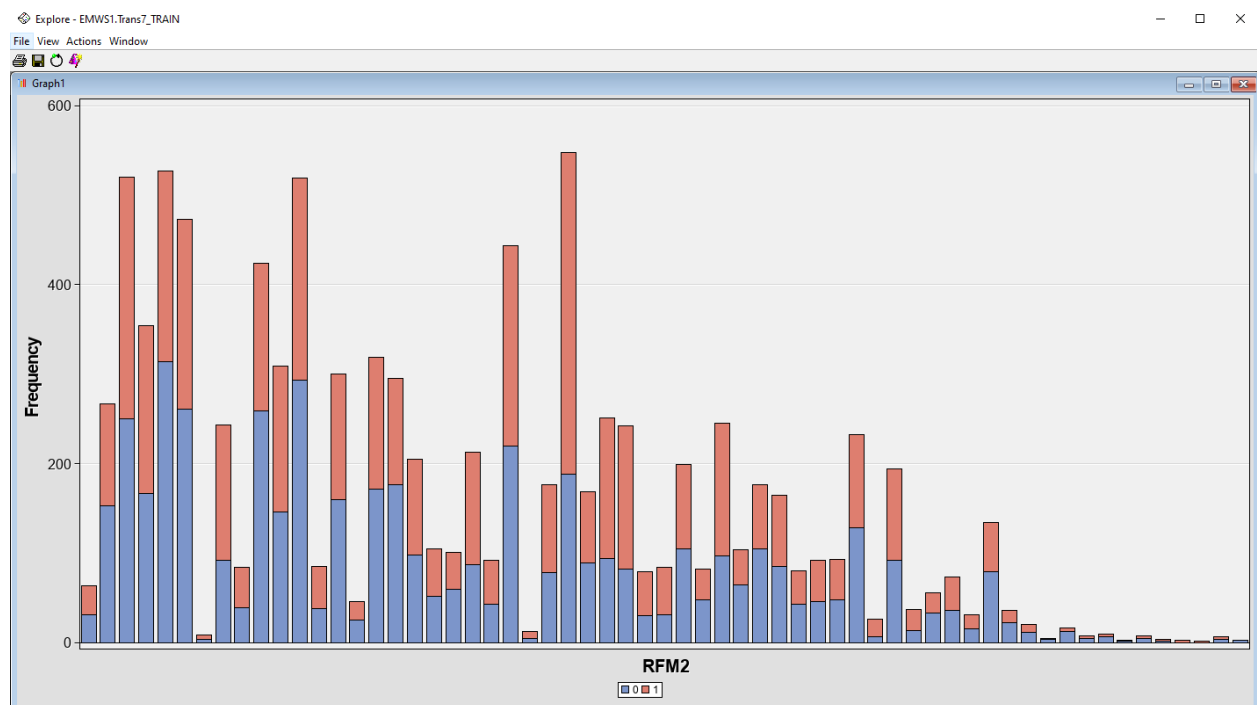
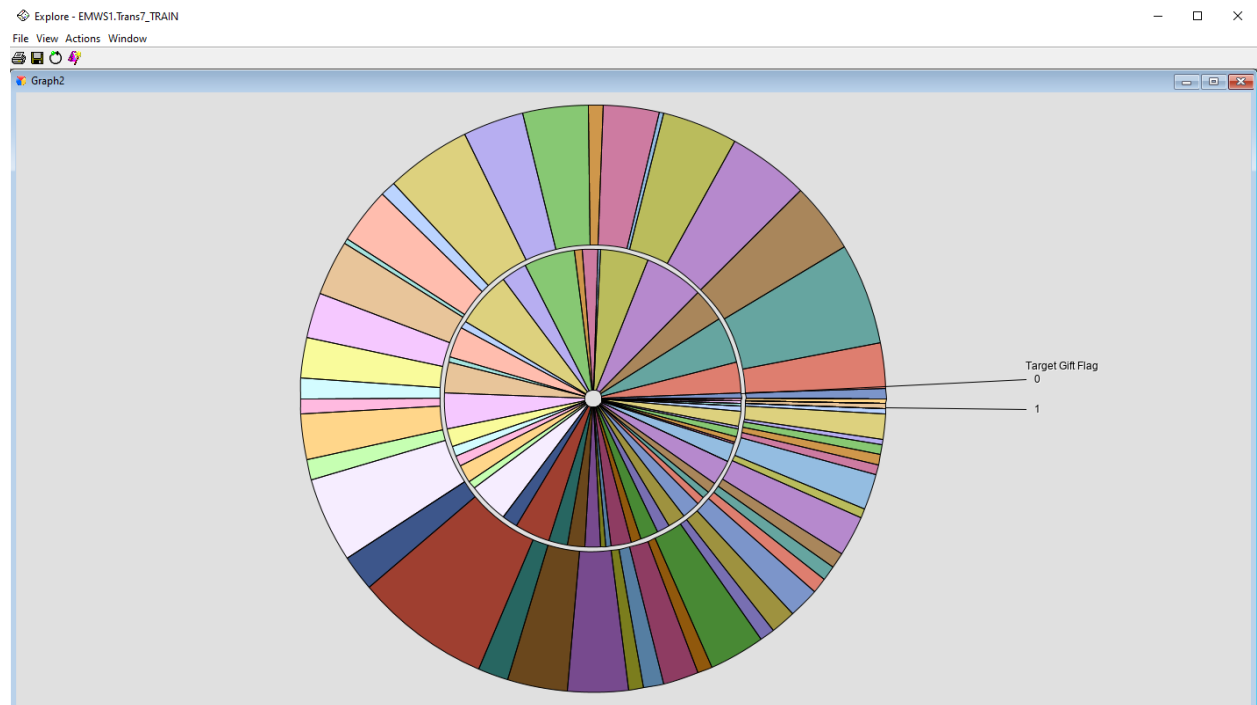
RFM2

Name	Type	Length	Format	Level	Formula	Label	Role	Report
RFM2	Numeric	8		Interval	substr(pctl_GiftTimeL...		Input	No

**Outputs** Sample Log

Preview OK Cancel

C.)



**D.)**

Response rate (040404 :

$$(188 / (188 + 359)) = 0.344$$

**Response rate for Group 040404 = 34.4%**

Response rate (030303):

$$(219 / (219+224)) = 0.4944$$

**Response rate for Group 030303 = 49.44%**

**E.)**

Each Promotional mail cost = \$2.30

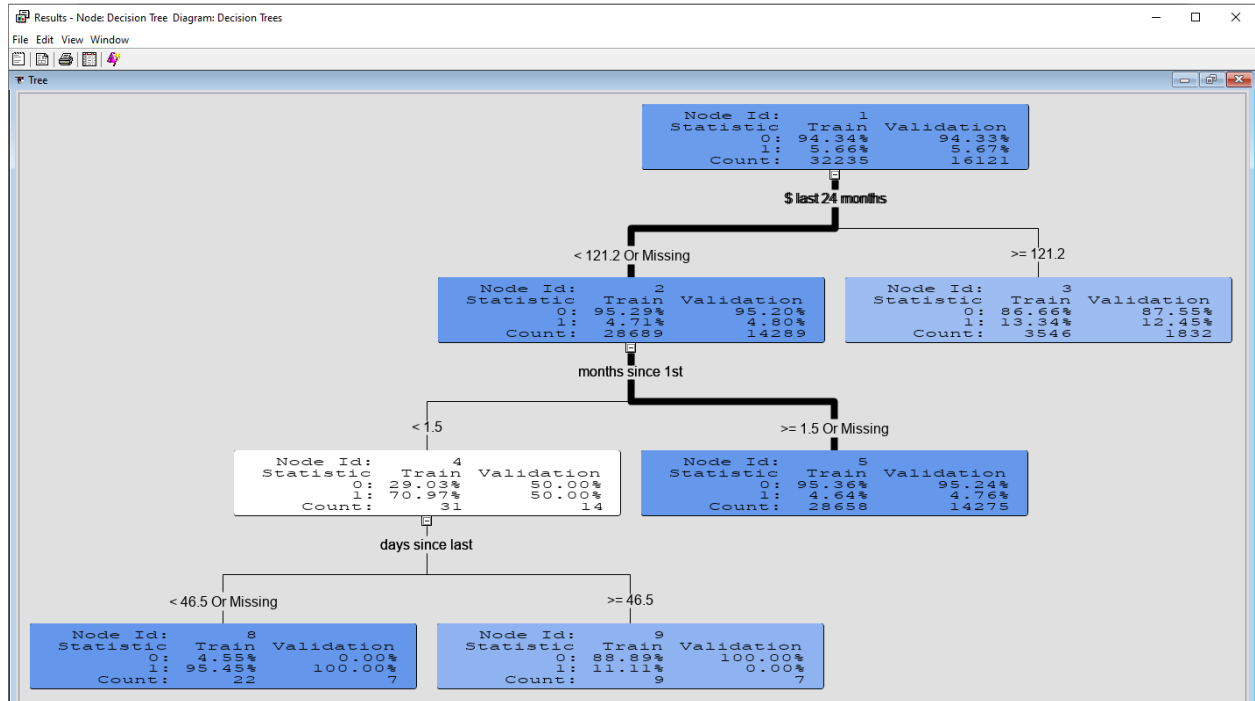
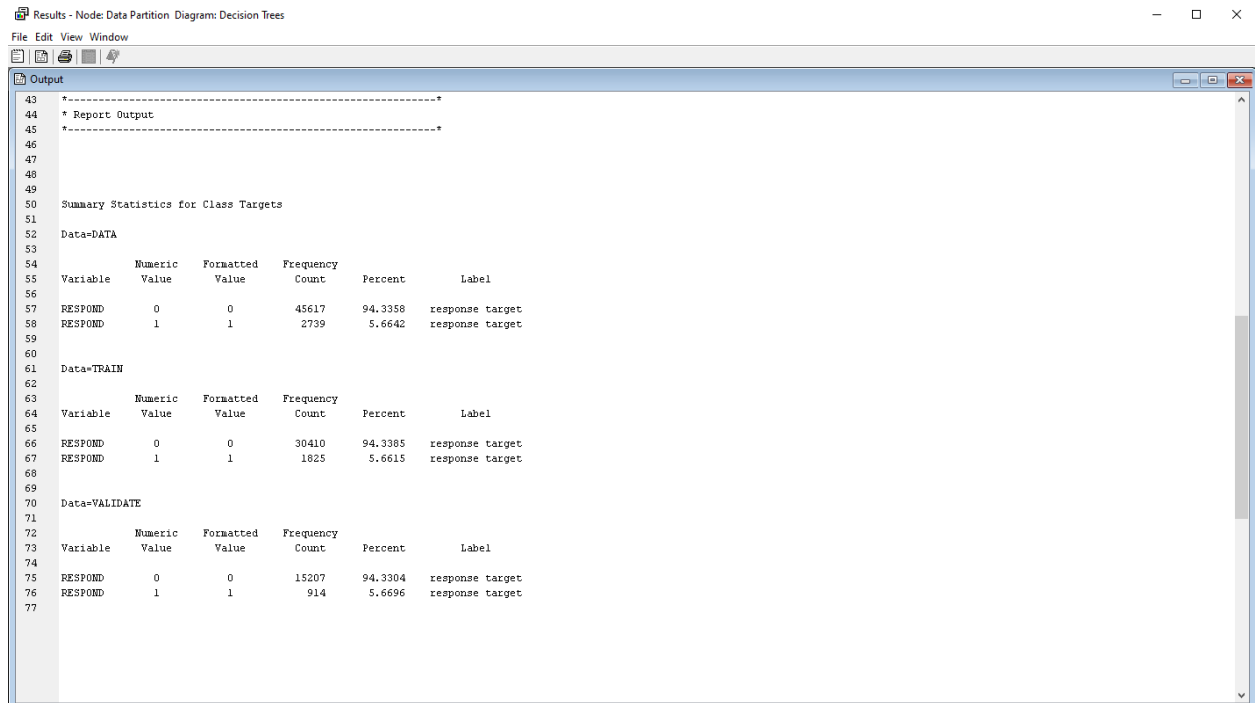
Average donation = \$21.00

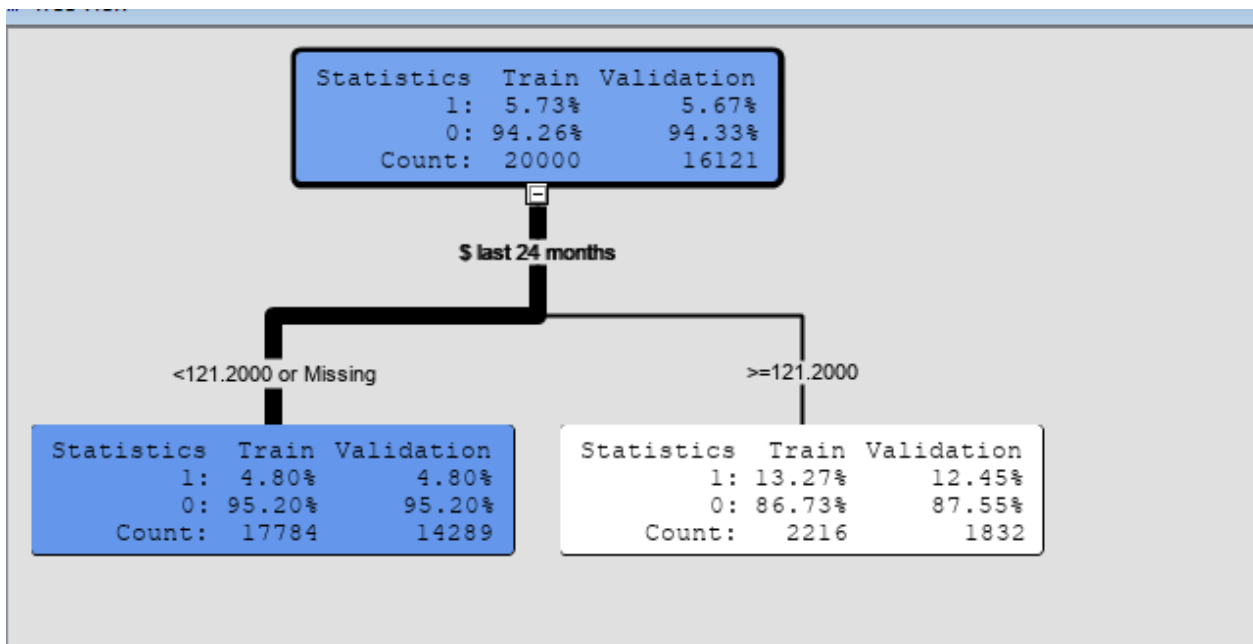
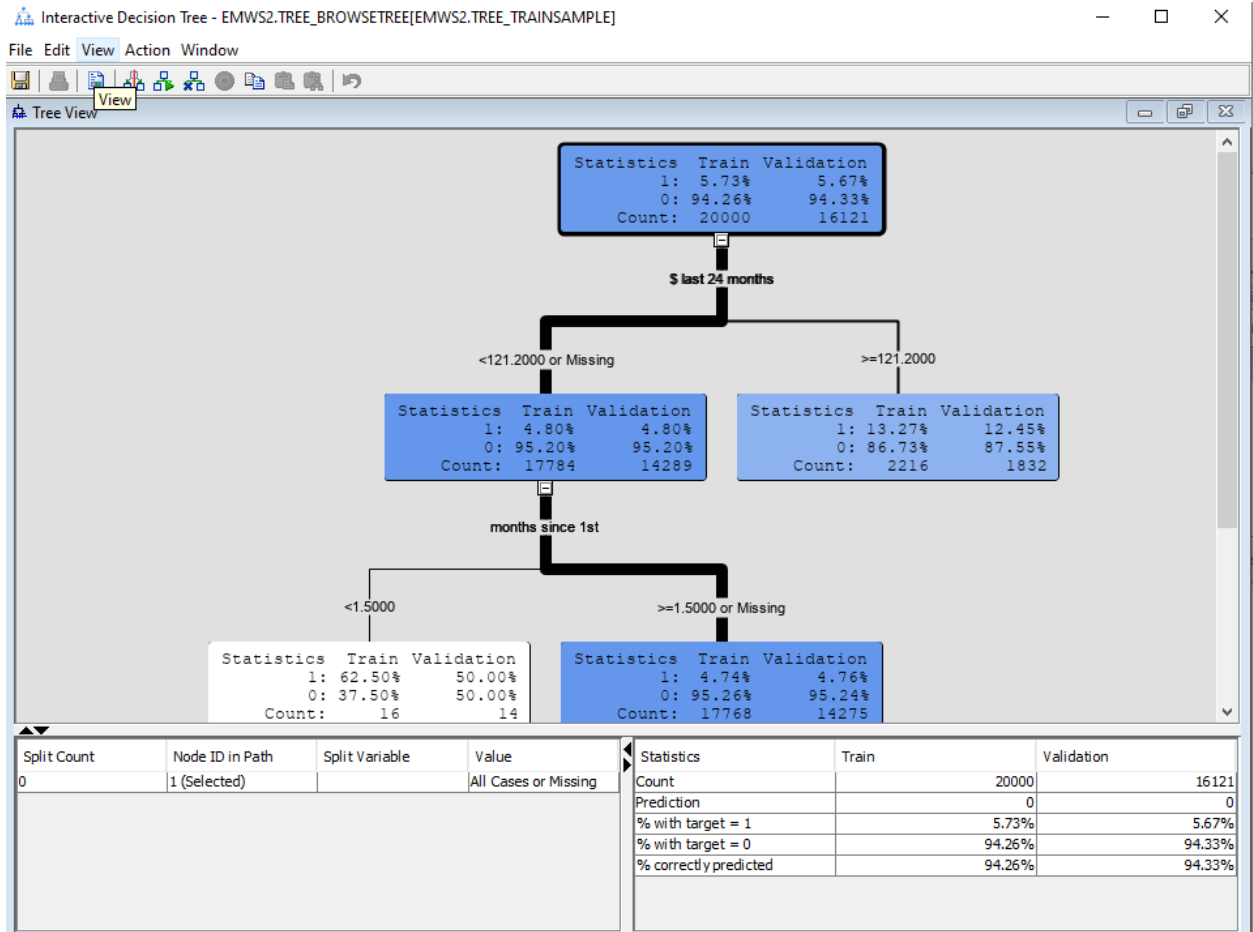
Break Even Response Rate =  $( 2.3 / 21 ) = 0.1905 = 10.95\%$

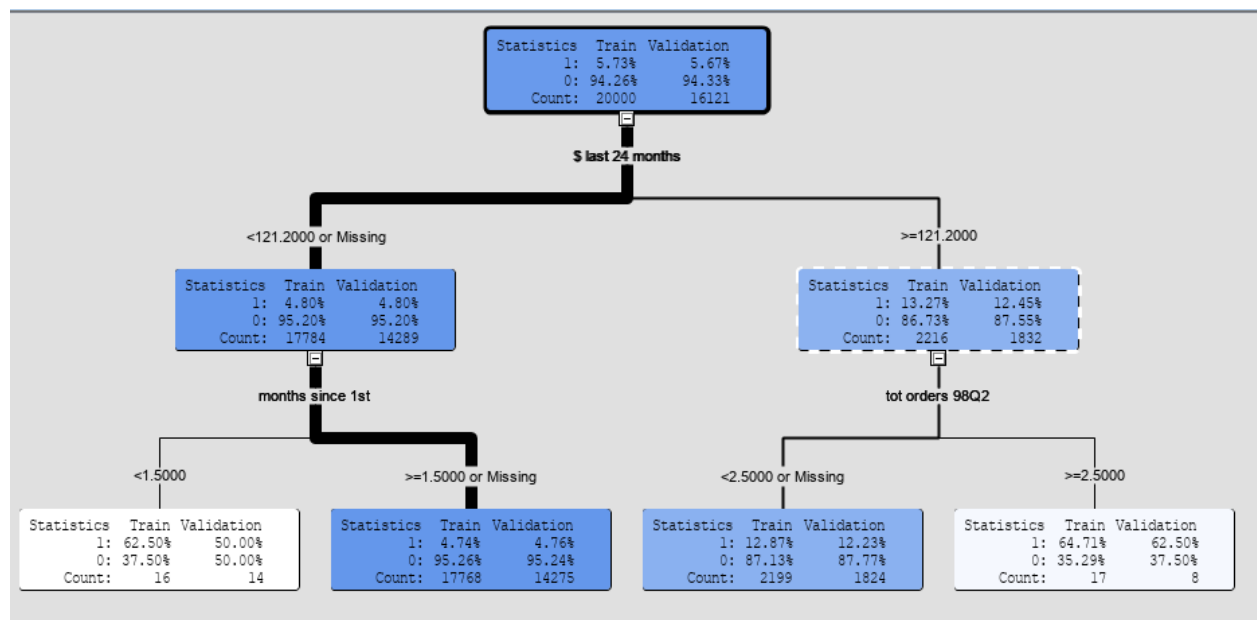
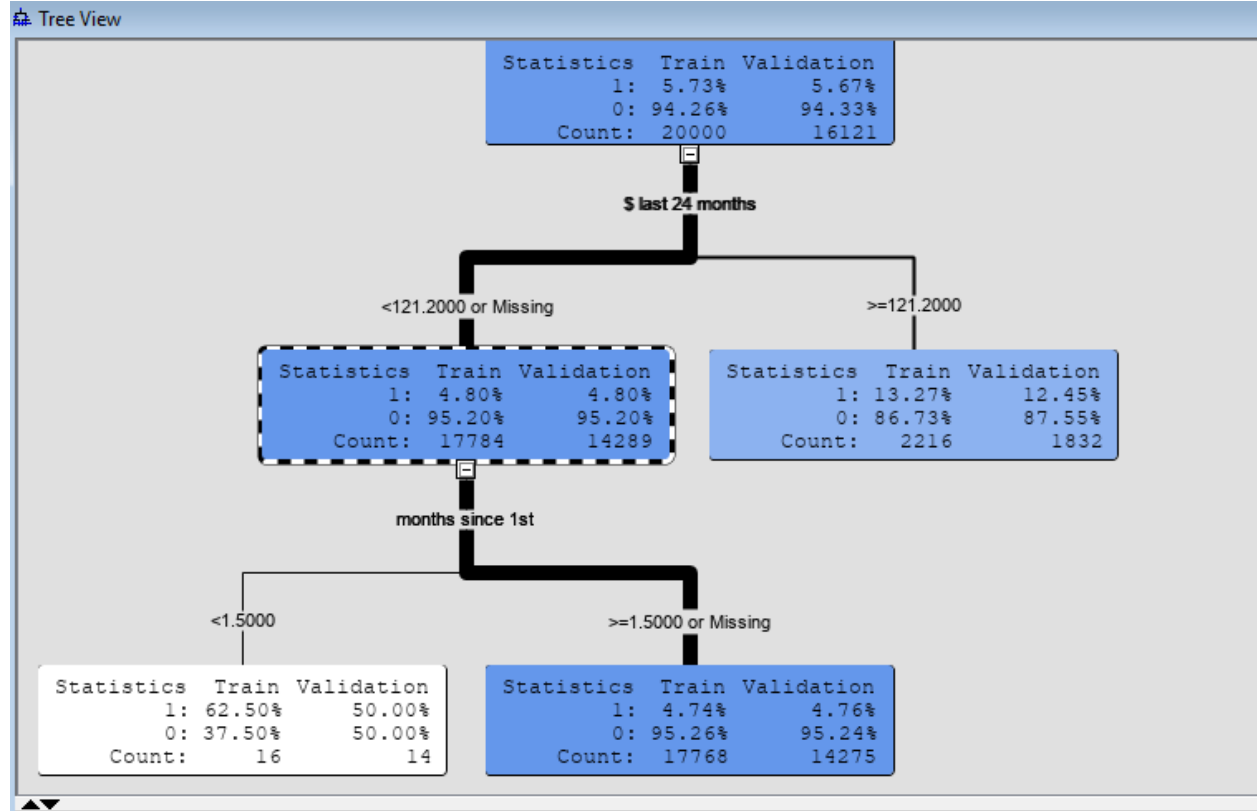
**Break Even Response rate for this promotion = 10.95%**

From the RFM and Frequency stacked bar chart, we see that many of the blue and red sections of bars are equal. Suggesting that the response:non-response rate is close to 50%. Therefore most of the RFM cells exceed this response rate of 10.95%. However, there are some cells for larger RFM values that do not exceed the minimum rate.

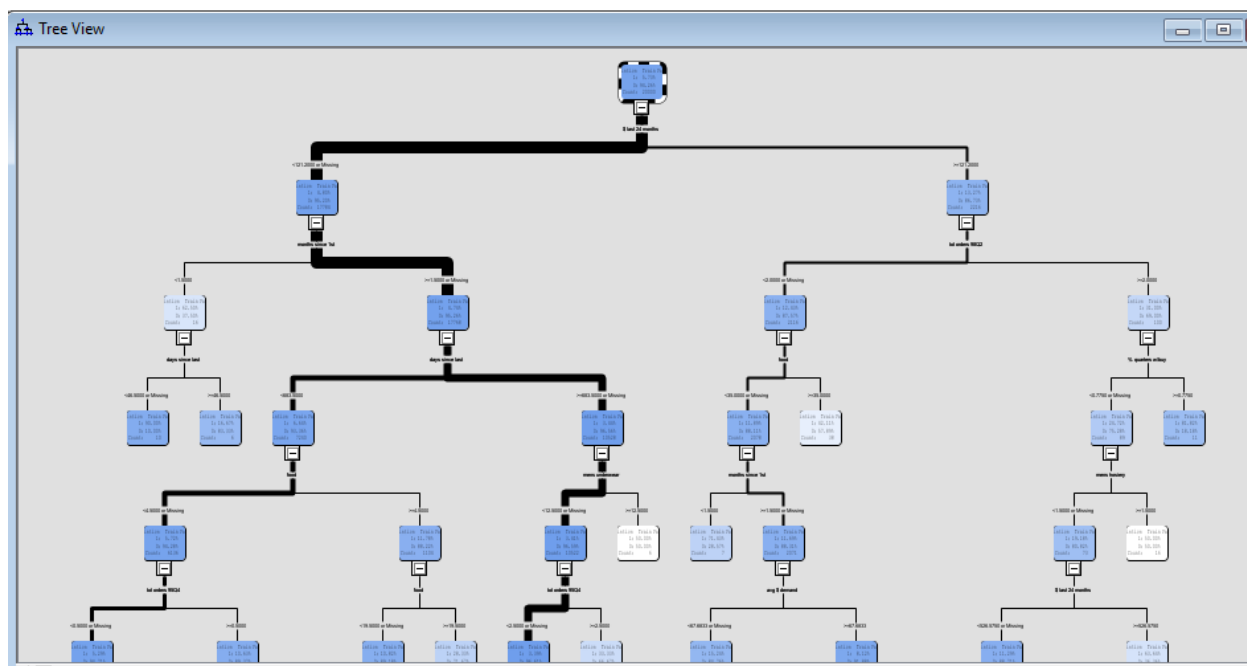
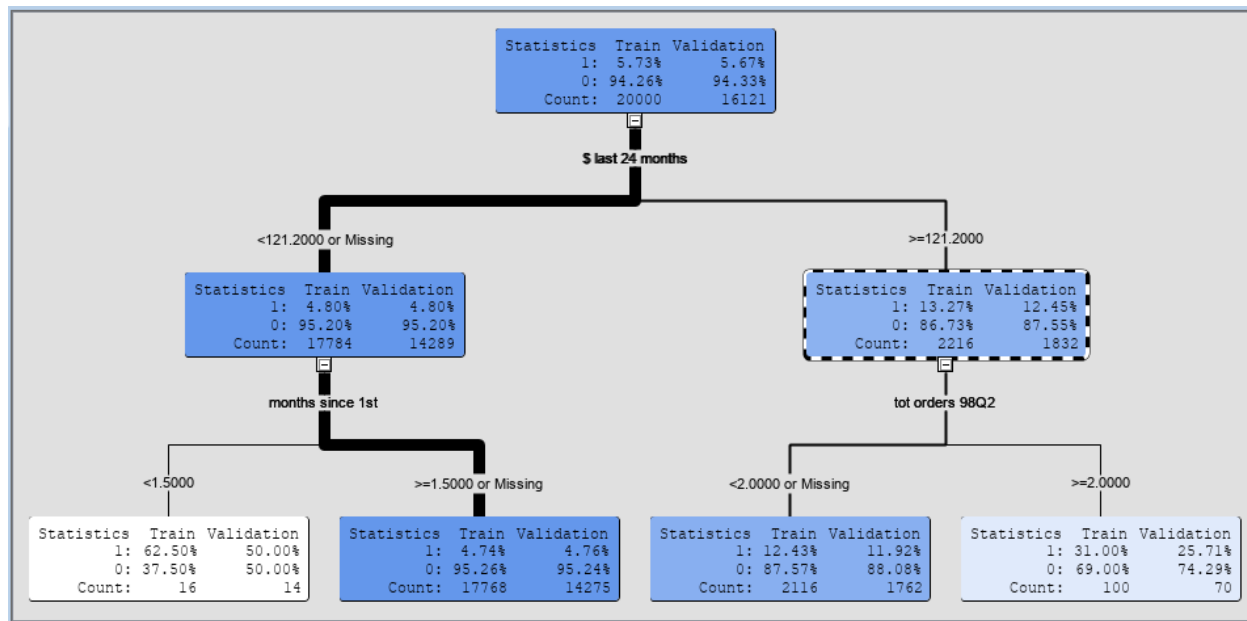
**(Tutorial):**

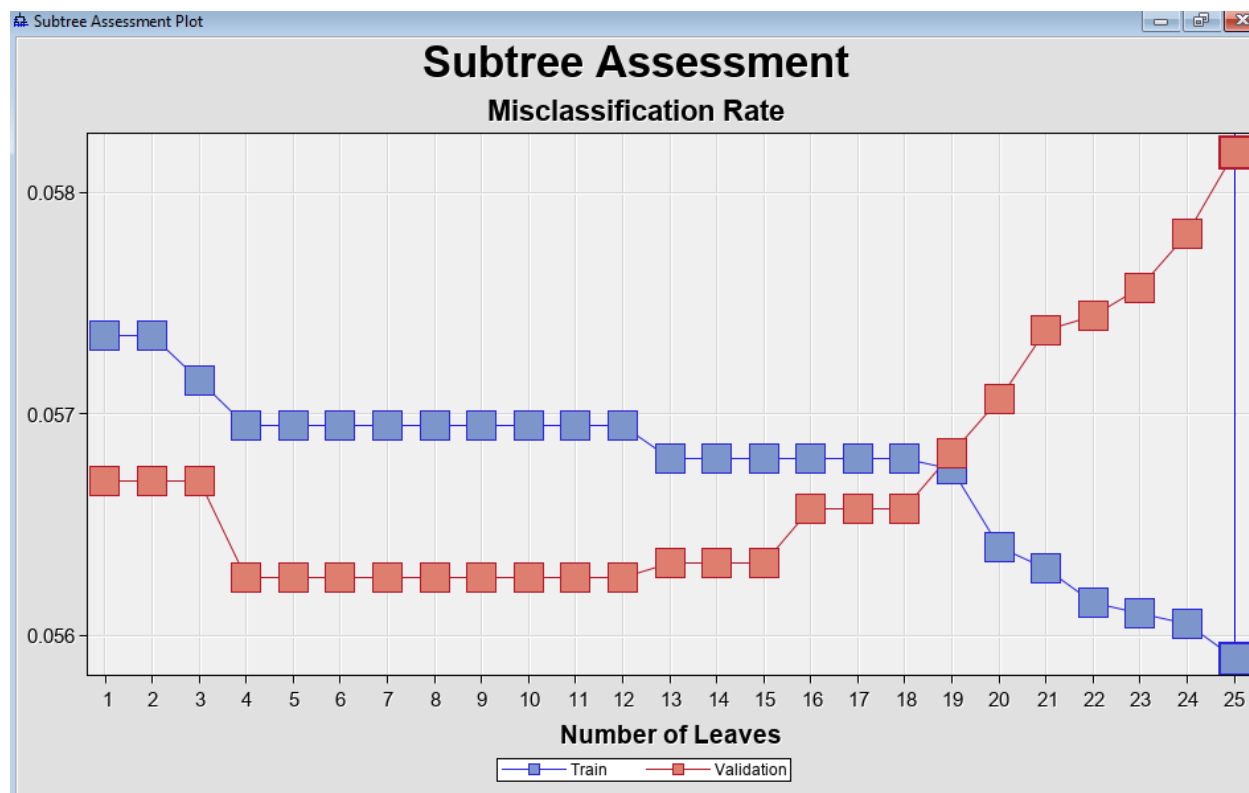




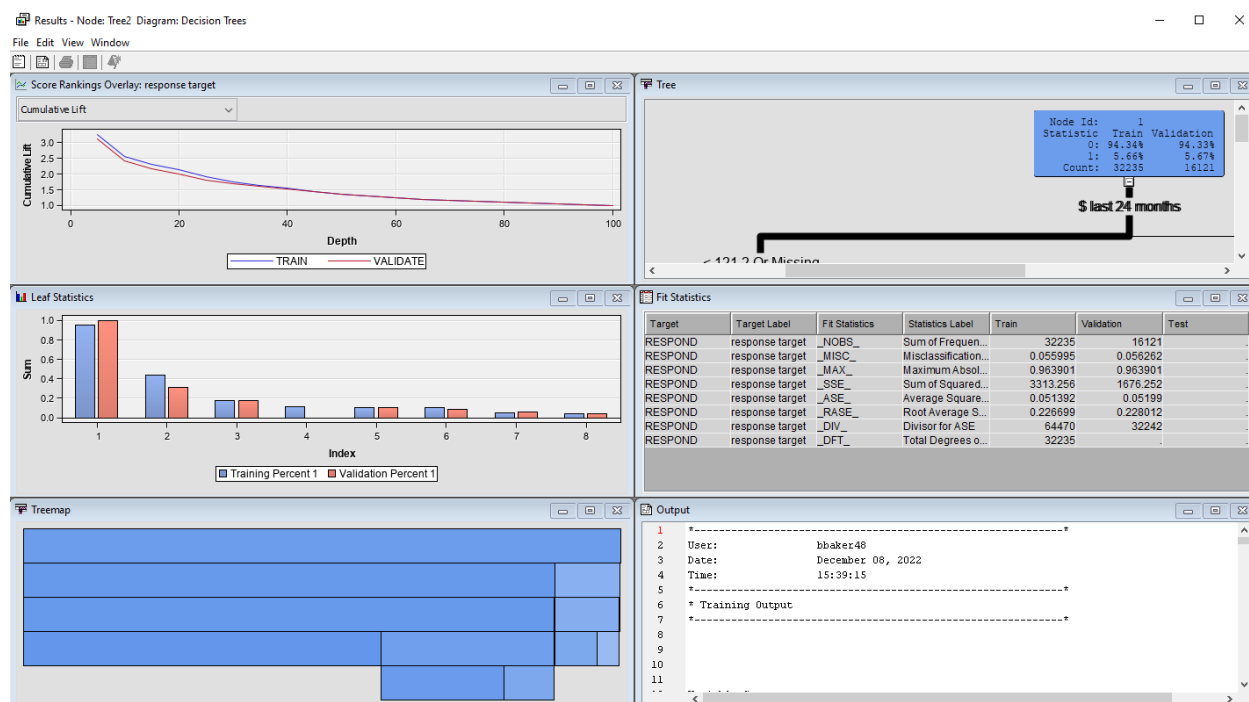


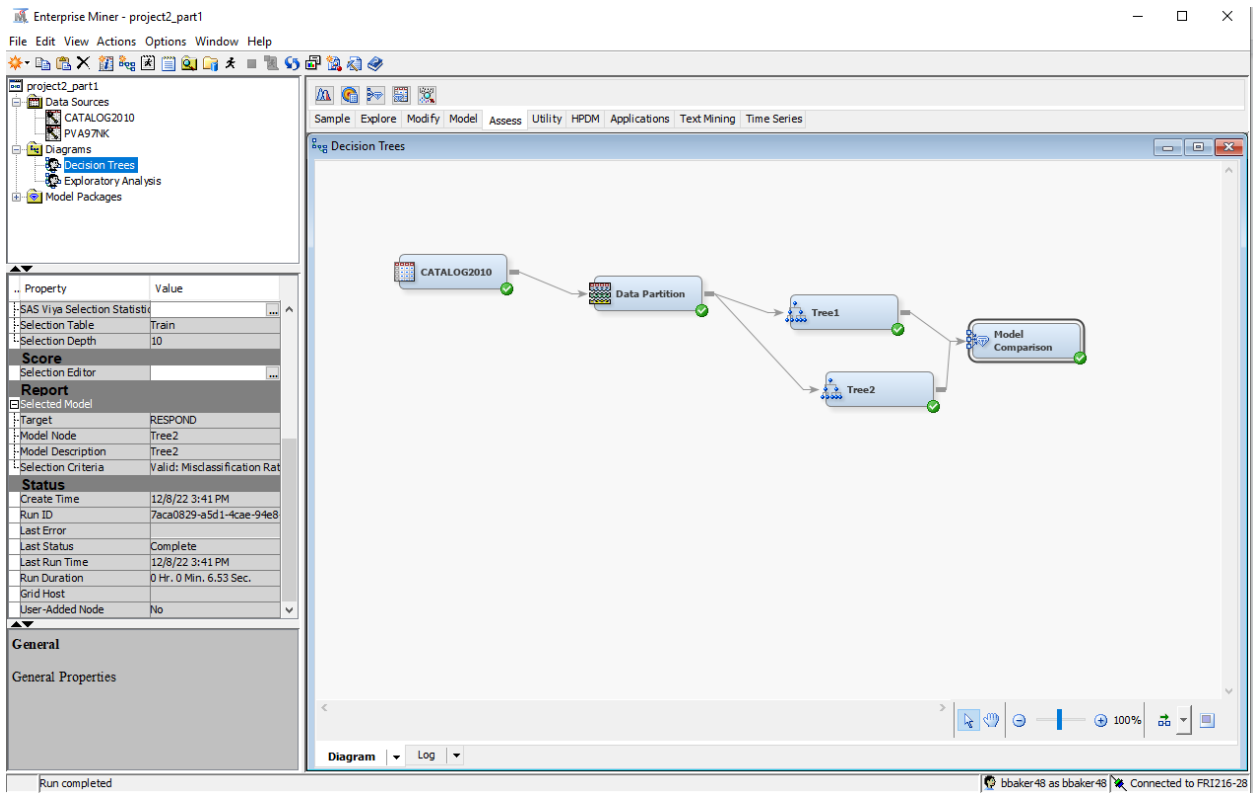
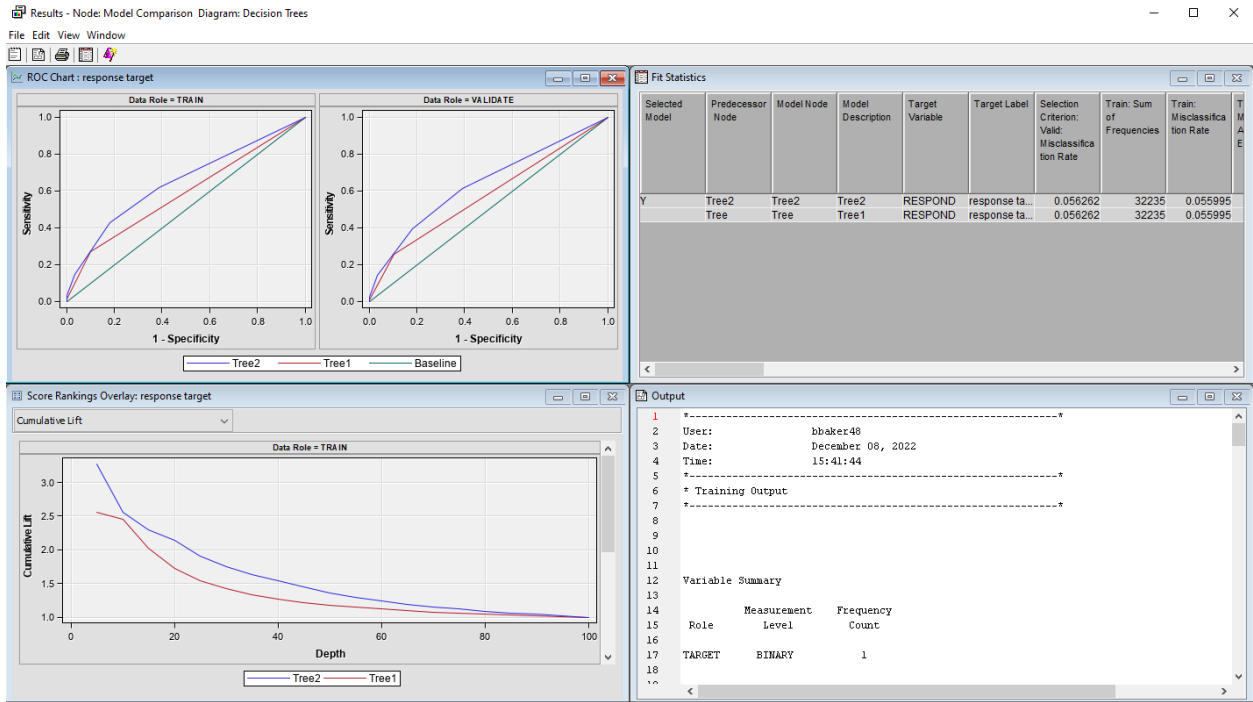






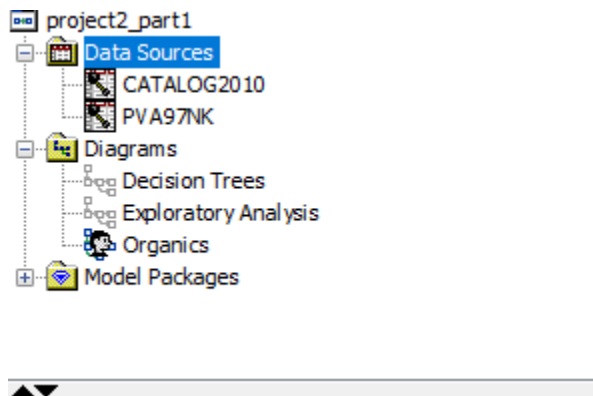
The plot above shows that the best tree, based on the validation data, has 4 leaves. It also shows that the maximum 25-leaf tree generates a lower misclassification rate than any of its simpler predecessors.





**(Task 3):**

a.



b.

1)

Data Source Wizard -- Step 5 of 8 Column Metadata

(none) ☐ not Equal to ☐ Apply ☐ Reset

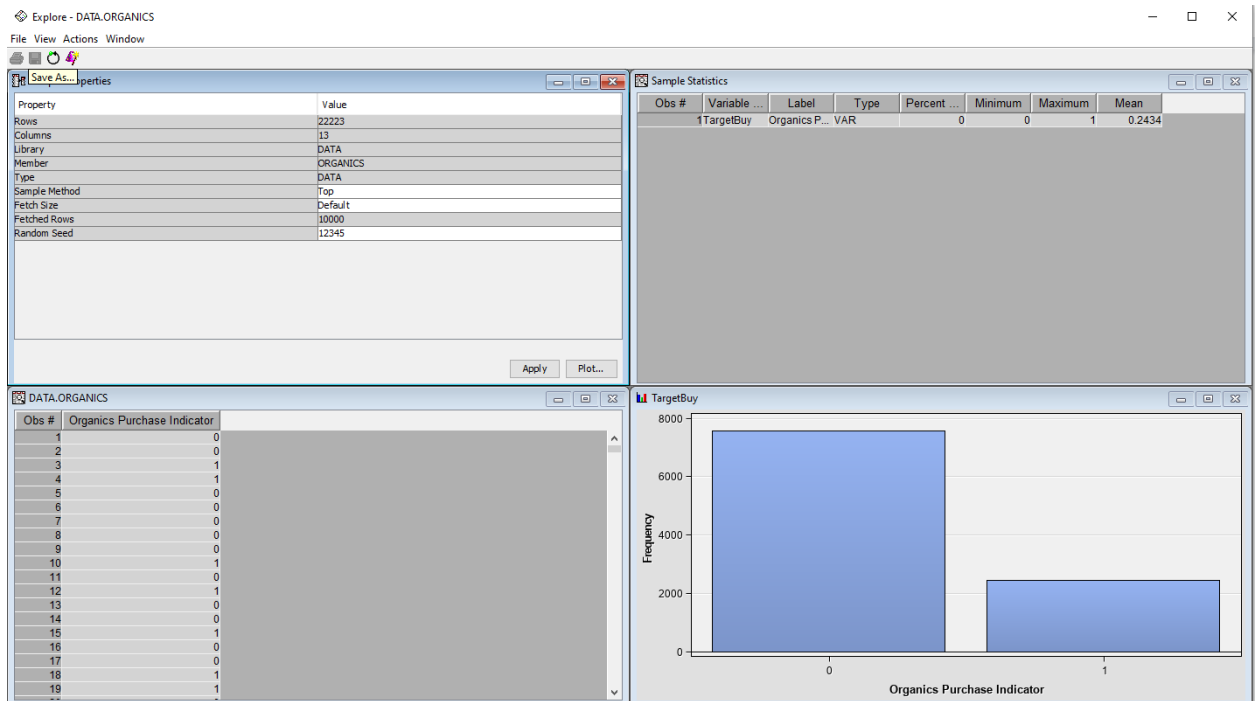
Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
DemAffl	Input	Interval	No		No	.	
DemAge	Input	Interval	No		No	.	
DemCluster	Rejected	Nominal	No		No	.	
DemClusterGroup	Input	Nominal	No		No	.	
DemGender	Input	Nominal	No		No	.	
DemReg	Input	Nominal	No		No	.	
DemTVReg	Input	Nominal	No		No	.	
ID	ID	Nominal	No		No	.	
PromClass	Input	Nominal	No		No	.	
PromSpend	Input	Interval	No		No	.	
PromTime	Input	Interval	No		No	.	
TargetAmt	Rejected	Interval	No		No	.	
TargetBuy	Target	Binary	No		No	.	

< >

Show code Explore Compute Summary < Back Next > Cancel

2)



- The distribution of TargetBuy is displayed by the bar graph in the bottom right. The frequencies are 7566 for no organics bought and 2434 for organics bought. The proportion of those who bought organics is:

$$2434 / (2434 + 7566) = 0.2434 \text{ or } 24.34\%$$

3)

(none)

▼

☐ not

Equal to

▼

...

Apply

Reset

Columns:

☐ Label

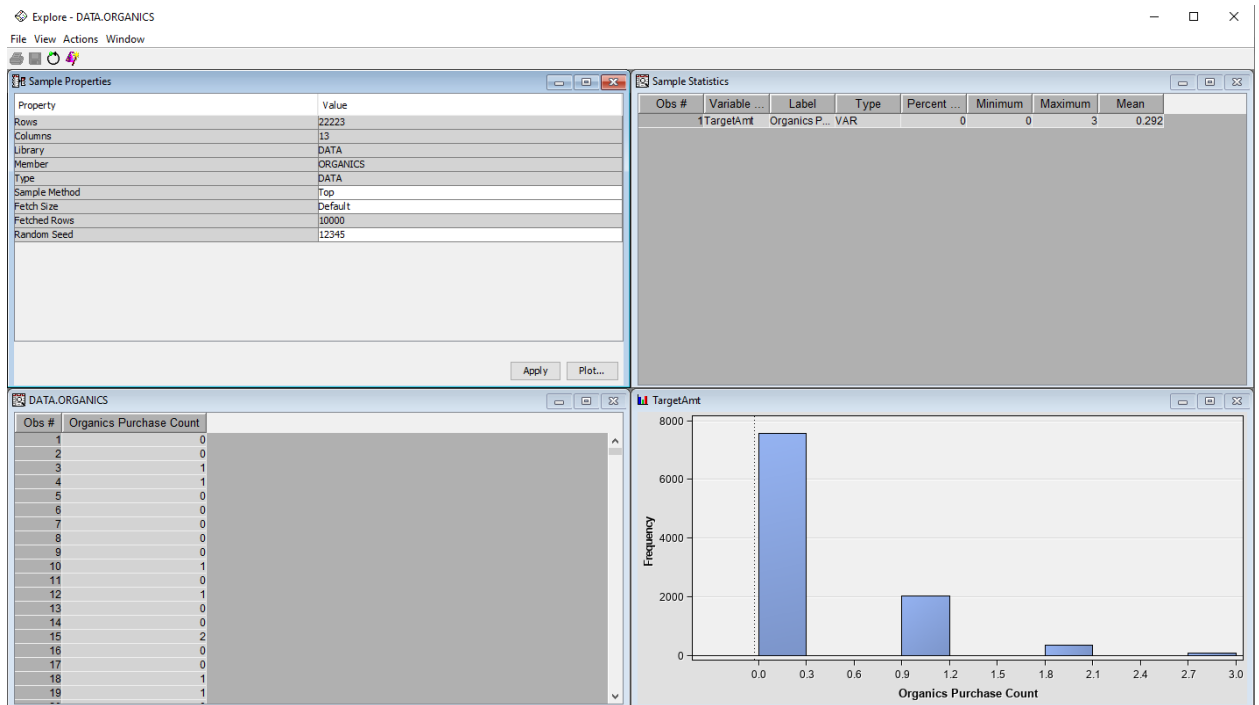
☐ Mining

☐ Basic

☐ Statistics

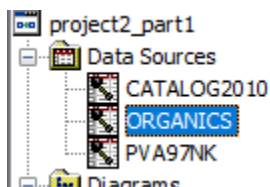
Name	Role	Level	Report	Order	Drop	Lower Limit	Upper
DemAffl	Input	Interval	No		No	.	
DemAge	Input	Interval	No		No	.	
DemCluster	Rejected	Nominal	No		No	.	
DemClusterGroup	Input	Nominal	No		No	.	
DemGender	Input	Nominal	No		No	.	
DemReg	Input	Nominal	No		No	.	
DemTVReg	Input	Nominal	No		No	.	
ID	ID	Nominal	No		No	.	
PromClass	Input	Nominal	No		No	.	
PromSpend	Input	Interval	No		No	.	
PromTime	Input	Interval	No		No	.	
TargetAmt	Rejected	Interval	No		No	.	
TargetBuy	Target	Binary	No		No	.	

4)

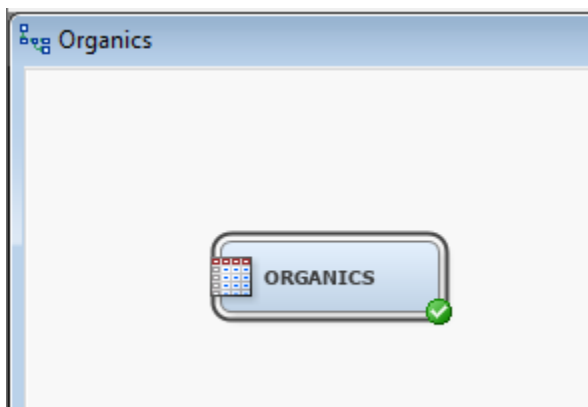


- We should probably not use TargetAmt to help predict TargetBuy because they are more than likely highly correlated. In the bar graph for TargetAmt, we see that the zero category has the same amount of observations in it as TargetBuy, while it splits it up the TargetBuy=1 category.

5)

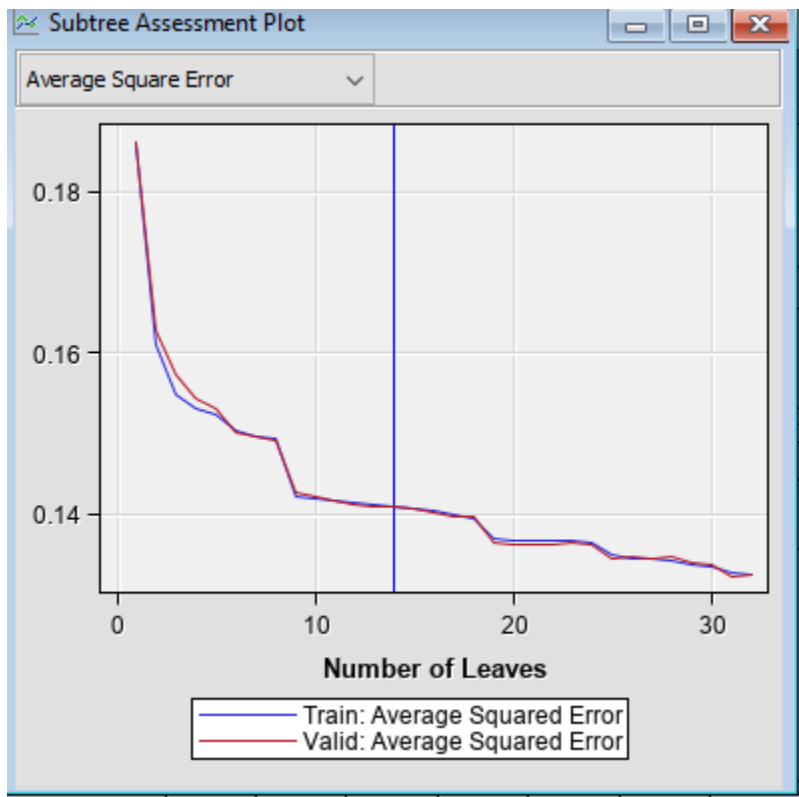


C.





1) a



- Number of leaves in the optimal tree is 14.

2)

- The variables used for the first splits were Age and Affluence Grade. The competing splits for the first split was Affluence grade less than 44.5 or Affluence grade greater than or equal to 44.5 or missing.

3)

- The second split used Affluence grade. Affluence grade split by  $<9.5$ ,  $\geq 9.5$ ,  $<12.5$  or missing and  $\geq 12.5$ . When Affluence grade is  $<9.5$  and  $\geq 9.5$  the next node is gender, while when affluence grade is  $<12.5$  or missing and  $\geq 12.5$  split again into affluence grade.

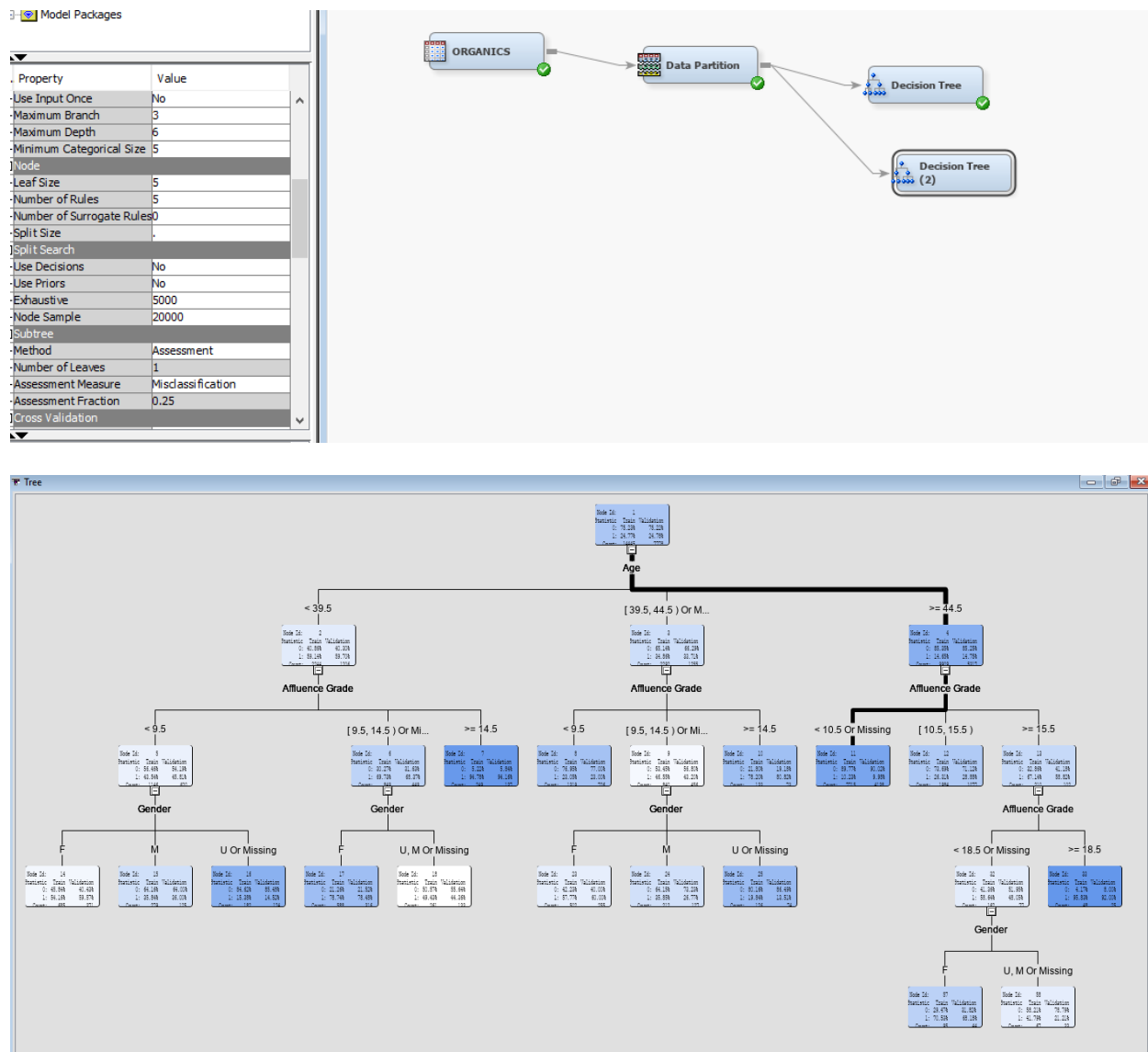
4)

- The root node of the Decision tree is 'Age'. This is dependent on the respondents age, either ' $<44.5$ ' or ' $\geq 44.5$  or missing'. Affluence grade is the next category the respondent is split into. For Age  $< 44.5$  the affluence grade is split between: ' $< 9.5$ ', ' $\geq 9.5$  or missing'. For the rest of age values, affluence grade is split between: ' $<12.5$  or missing' and ' $\geq 12.5$ ', and they continue on their own series of splits. Using the Affluence Grade split: ' $<9.5$ ' and ' $\geq 9.5$ ', when we connect the second node, 'gender', we get that 'U, M, or missing' and 'Females' must meet different criteria for the followign decision. For Females with Affluence Grade ' $<9.5$ ' they split into another subtree of Affluence Grade, that is: ' $<6.5$ ', ' $\geq 6.5$  or missing'. The ones with grade ' $<6.5$ ' reach their final node. For those with affluence grade ' $\geq 6.5$  or missing' they are then split between 'Age  $< 39.5$  or

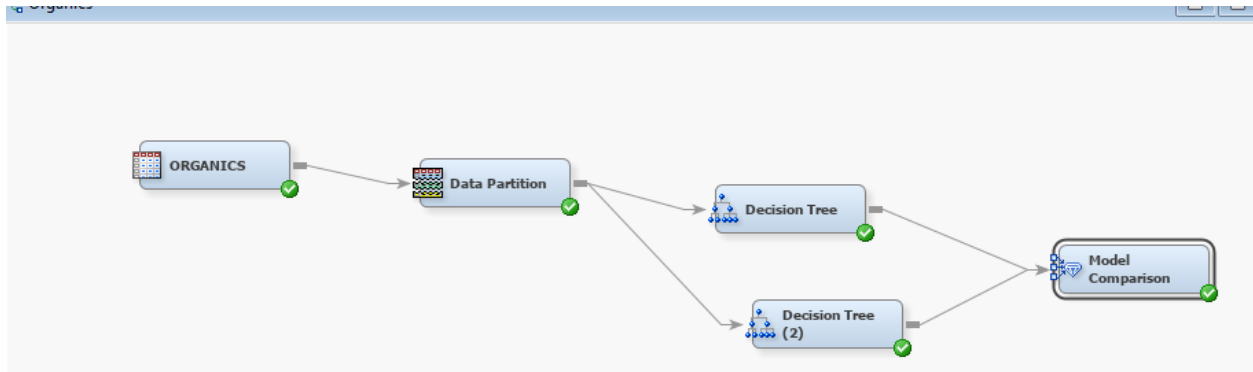


missing' and 'Age  $\geq 39.5$ '. Then, reaching the next gender subtree, both 'U,M or missing' and 'Females' who took this series of splits reach their final node.

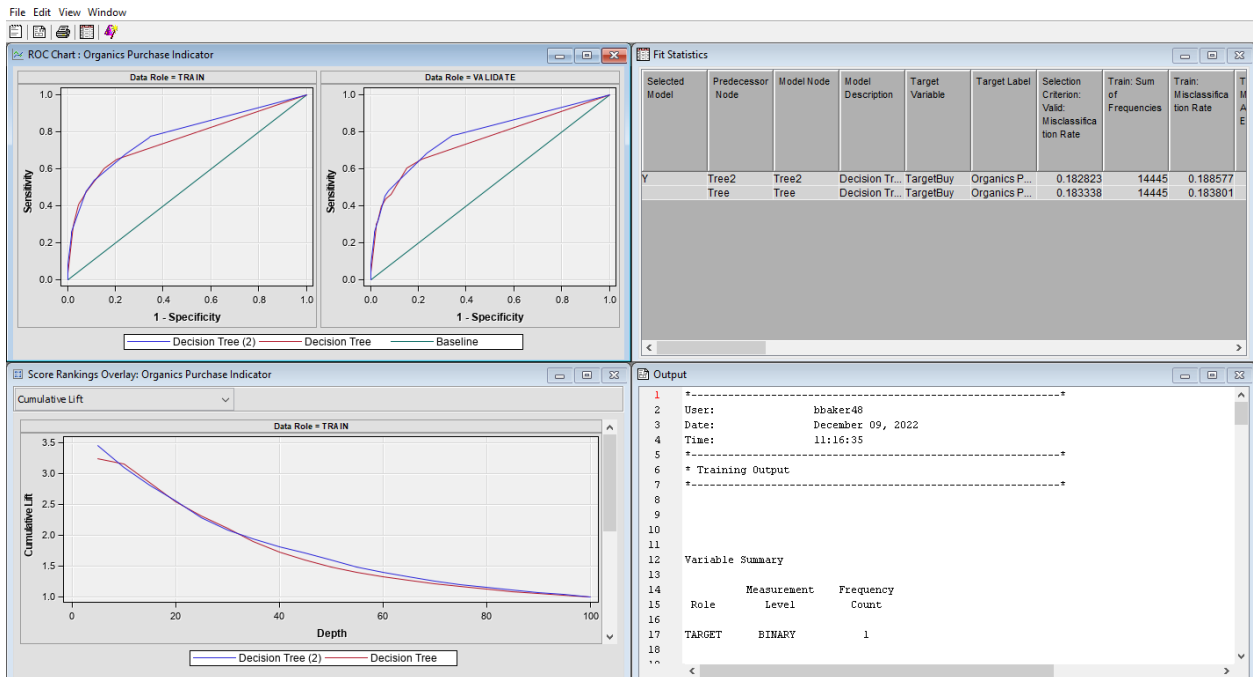
g.



- The optimal number of leaves in this tree is 16.



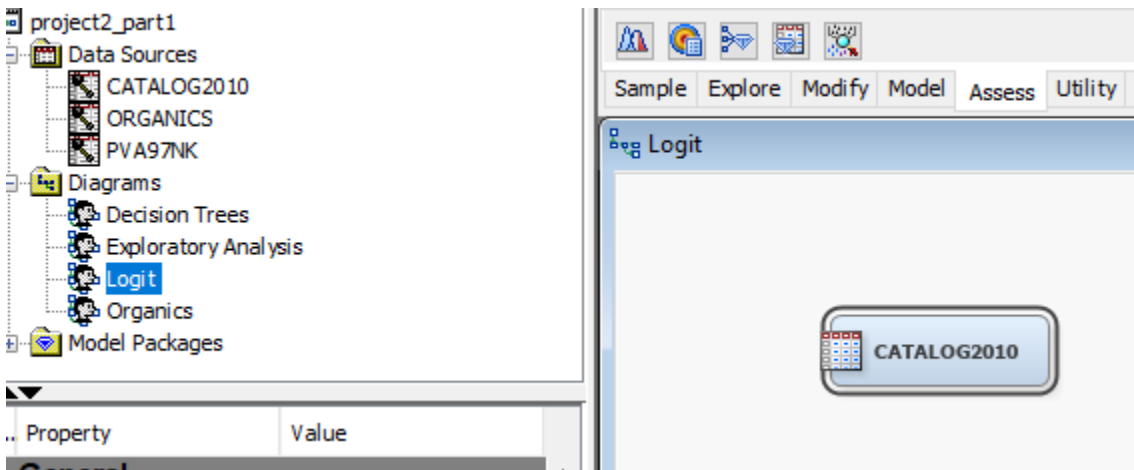
Results - Node: Model Comparison Diagram: Organics



- The second tree is the better decision tree as it has a higher ROC maximum point than decision tree 1 and it also has a lower misclassification rate, along with a lower averaged squared error.

Part 3 - Logit

1.



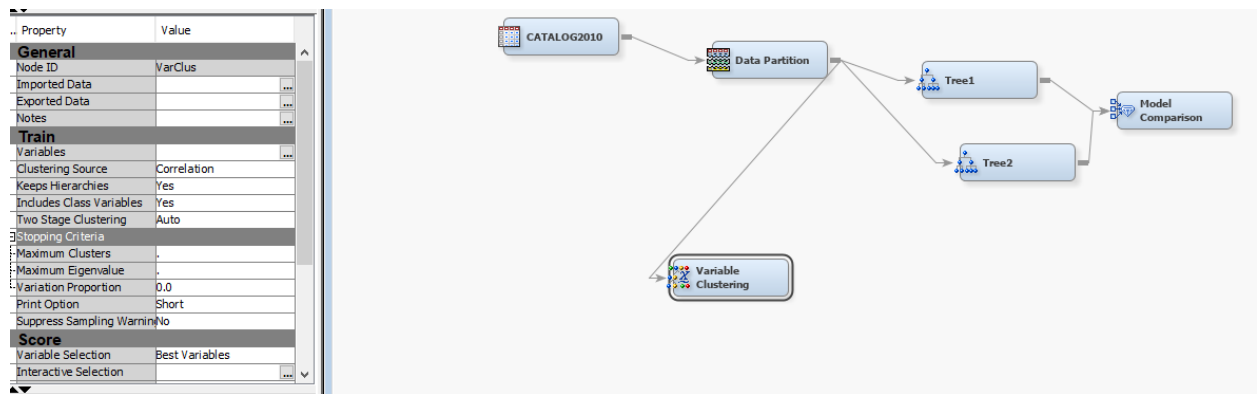
2.

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit	Number of Levels	Percent Missing	Minimum	Maximum	Mean	Standard Deviation
ACTBUY	Input	Interval	No		No	.	.	11	0	.	.	.	.
BOTHPAYM	Input	Binary	No		No	.	.	2	0	.	.	.	.
BUYPROP	Input	Interval	No		No	.	.	.	0	0	.	0.18858	0.256956
CATALOGCNT	Input	Interval	No		No	.	.	.	0	1	27	3.76582	3.133252
CCPAYM	Input	Binary	No		No	.	.	2	0	.	.	.	.
COUNTY	Rejected	Nominal	No		No	.	.	.	0	10	999	426.4056	315.3986

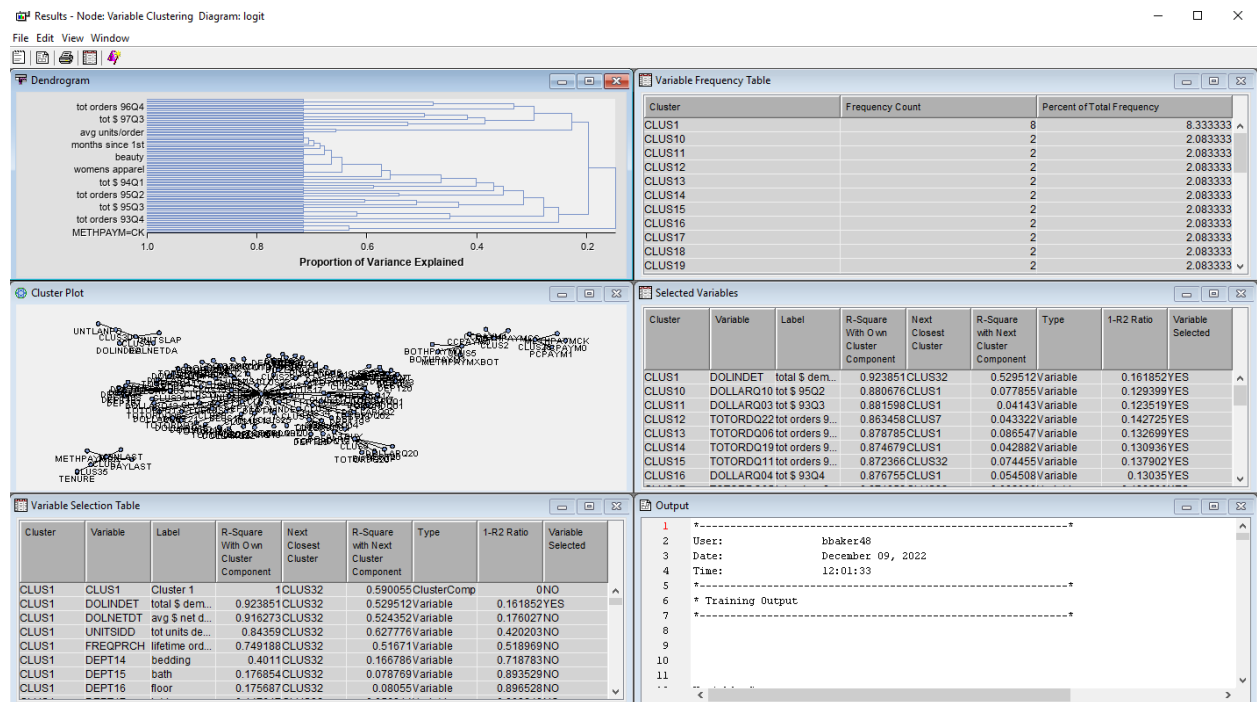
3.

Name	Role	Level	Report
STATE	Rejected	Nominal	No
TENURE	Input	Interval	No

4.



5.



Results - Node: Variable Clustering Diagram: logit

File Edit View Window

Selected Variables

Cluster	Variable	Label	R-Square With Own Cluster Component	Next Closest Cluster	R-Square with Next Cluster Component	Type	1-R2 Ratio	Variable Selected
CLUS1	DOLINDET	total \$ demand	0.923851	CLUS32	0.529512	Variable	0.161852	YES
CLUS10	DOLLARQ10	tot \$ 95Q2	0.880676	CLUS1	0.077855	Variable	0.129399	YES
CLUS11	DOLLARQ03	tot \$ 93Q3	0.881598	CLUS1	0.04143	Variable	0.123519	YES
CLUS12	TOTORDQ22	tot orders 98Q2	0.863458	CLUS7	0.043322	Variable	0.142725	YES
CLUS13	TOTORDQ06	tot orders 94Q2	0.878785	CLUS1	0.085547	Variable	0.132699	YES
CLUS14	TOTORDQ19	tot orders 97Q3	0.874679	CLUS1	0.042882	Variable	0.130936	YES
CLUS15	TOTORDQ11	tot orders 95Q3	0.872366	CLUS32	0.074455	Variable	0.137902	YES
CLUS16	DOLLARQ04	tot \$ 93Q4	0.876755	CLUS1	0.054508	Variable	0.13035	YES
CLUS17	TOTORDQ05	tot orders 94Q1	0.871852	CLUS32	0.082009	Variable	0.139596	YES
CLUS18	DOLLARQ16	tot \$ 96Q4	0.866462	CLUS1	0.084269	Variable	0.145827	YES
CLUS19	TOTORDQ18	tot orders 97Q2	0.879799	CLUS1	0.059933	Variable	0.127729	YES
CLUS2	CCPAYM0	CCPAYM=0	1	CLUS28	0.3108	Variable	0	YES
CLUS20	TOTORDQ14	tot orders 96Q2	0.84481	CLUS32	0.057243	Variable	0.164613	YES
CLUS21	TOTORDQ21	tot orders 98Q1	0.860266	CLUS1	0.043468	Variable	0.146084	YES
CLUS22	DOLLARQ09	tot \$ 95Q1	0.873589	CLUS1	0.074568	Variable	0.136597	YES
CLUS23	DOLLARQ02	tot \$ 93Q2	0.869018	CLUS1	0.092496	Variable	0.144333	YES
CLUS24	TOTORDQ01	tot orders 93Q1	0.875695	CLUS1	0.105666	Variable	0.138991	YES
CLUS25	TOTORDQ07	tot orders 94Q3	0.869525	CLUS1	0.082185	Variable	0.142158	YES
CLUS26	TOTORDQ13	tot orders 96Q1	0.845288	CLUS32	0.070058	Variable	0.166366	YES
CLUS27	DOLLARQ08	tot \$ 94Q4	0.853544	CLUS1	0.094627	Variable	0.161764	YES
CLUS28	METHPAYMCK	METHPAYM=CK	1	CLUS2	0.3108	Variable	0	YES
CLUS29	DEPT03	womens underwear	0.473772	CLUS1	0.200941	Variable	0.658559	YES
CLUS3	DOLLARQ17	tot \$ 97Q1	0.81237	CLUS1	0.078702	Variable	0.203659	YES
CLUS30	UNTLANPO	avg units/order	1	CLUS4	0.123804	Variable	0	YES
CLUS31	DEPT12	mens misc	0.367295	CLUS1	0.090206	Variable	0.695438	YES
CLUS32	CATALOGCNT	number of catalogs received	0.789991	CLUS1	0.51167	Variable	0.541574	YES
CLUS33	DEPT21	light	1	CLUS1	0.010229	Variable	0	YES
CLUS34	DEPT19	window	0.532576	CLUS1	0.026522	Variable	0.480159	YES
CLUS35	TENURE	months since 1st	1	CLUS7	0.192899	Variable	0	YES
CLUS4	DOLINDEA	avg \$ demand	0.912853	CLUS30	0.262021	Variable	0.118089	YES
CLUS5	BOTHPAYM0	BOTHPAYM=0	1	CLUS2	0.171641	Variable	5.36E-16	YES
CLUS6	TOTORDQ12	tot orders 95Q4	0.732581	CLUS32	0.077932	Variable	0.290021	YES
CLUS7	MONLAST	months since last	0.95065	CLUS35	0.197535	Variable	0.061498	YES
CLUS8	TOTORDQ15	tot orders 96Q3	0.872407	CLUS1	0.064934	Variable	0.136453	YES
CLUS9	TOTORDQ20	tot orders 97Q4	0.821266	CLUS7	0.054505	Variable	0.189037	YES

Above are variables chosen as candidates for logistic regression, from the clusters, using the '1-R<sup>2</sup>' ratio

The following is the complete list of which variables are in which cluster; variables in the same cluster were similar in the analysis. The variable with the lowest '1-R<sup>2</sup>' ratio value are selected as the representative for the cluster.

Output						
5168						
5169						
5170			R-squared with			
5171	35 Clusters		-----			
5172			Own	Next	1-R**2	Variable
5173	Cluster	Variable	Cluster	Closest	Ratio	Label
5174			-----			
5175	Cluster 1	DEPT14	0.4011	0.1668	0.7188	bedding
5176		DEPT15	0.1769	0.0788	0.8935	bath
5177		DEPT16	0.1757	0.0805	0.8965	floor
5178		DEPT17	0.1476	0.0563	0.9032	table
5179		DOLINET	0.9239	0.5295	0.1619	total \$ demand
5180		DOLNETDT	0.9163	0.5244	0.1760	avg \$ net demand
5181		FREQPRCH	0.7492	0.5167	0.5190	lifetime orders
5182		UNITSIDD	0.8436	0.6278	0.4202	tot units demand
5183			-----			
5184	Cluster 2	CCPAYMO	1.0000	0.3108	0.0000	CCPAYM=0
5185		CCPAYM1	1.0000	0.3108	0.0000	CCPAYM=1
5186		METHPAYMCC	1.0000	0.3108	0.0000	METHPAYM=CC
5187			-----			
5188	Cluster 3	DOLL24	0.4767	0.3047	0.7526	\$ last 24 months
5189		DOLLARQ17	0.8124	0.0787	0.2037	tot \$ 97Q1
5190		TOTORDQ17	0.7392	0.0625	0.2782	tot orders 97Q1
5191			-----			
5192	Cluster 4	DOLINDEA	0.9129	0.2620	0.1181	avg \$ demand
5193		DOLNETDA	0.9039	0.2546	0.1290	tot \$ net demand
5194		UNITSLAP	0.4720	0.1137	0.5957	avg price/unit
5195			-----			
5196	Cluster 5	BOTHPAYMO	1.0000	0.1716	0.0000	BOTHPAYM=0
5197		BOTHPAYM1	1.0000	0.1716	0.0000	BOTHPAYM=1
5198		METHPAYMXBOT	1.0000	0.1716	0.0000	METHPAYM=XBOT
5199			-----			
5200	Cluster 6	ACTBUY	0.4676	0.3380	0.8042	num qtrrs w/buy
5201		DEPT25	0.3441	0.2087	0.8289	food
5202		DOLLARQ12	0.6580	0.0974	0.3789	tot \$ 95Q4
5203		TOTORDQ12	0.7326	0.0779	0.2900	tot orders 95Q4
5204			-----			
5205	Cluster 7	DAYLAST	0.9506	0.1976	0.0615	days since last
5206		MONLAST	0.9506	0.1975	0.0615	months since last
5207		METHPAYMDK	0.7497	0.1181	0.2838	METHPAYM=DK
5208			-----			

Output

5207		METHPAYMDK	0.7497	0.1181	0.2838	METHPAYM=DK
5208						
5209	Cluster 8	DOLLARQ15	0.8724	0.0862	0.1396	tot \$ 96Q3
5210		TOTORDQ15	0.8724	0.0649	0.1365	tot orders 96Q3
5211						
5212	Cluster 9	BUYPROP	0.5388	0.1214	0.5250	% quarters w/buy
5213		DEPT26	0.1761	0.0493	0.8666	gift
5214		DOLLARQ20	0.7222	0.0645	0.2969	tot \$ 97Q4
5215		TOTORDQ20	0.8213	0.0545	0.1890	tot orders 97Q4
5216						
5217	Cluster 10	DOLLARQ10	0.8807	0.0779	0.1294	tot \$ 95Q2
5218		TOTORDQ10	0.8807	0.0810	0.1298	tot orders 95Q2
5219						
5220	Cluster 11	DOLLARQ03	0.8816	0.0414	0.1235	tot \$ 93Q3
5221		TOTORDQ03	0.8816	0.0436	0.1238	tot orders 93Q3
5222						
5223	Cluster 12	DOLLARQ22	0.8635	0.0485	0.1435	tot \$ 98Q2
5224		TOTORDQ22	0.8635	0.0433	0.1427	tot orders 98Q2
5225						
5226	Cluster 13	DOLLARQ06	0.8788	0.0925	0.1336	tot \$ 94Q2
5227		TOTORDQ06	0.8788	0.0865	0.1327	tot orders 94Q2
5228						
5229	Cluster 14	DOLLARQ19	0.8747	0.0500	0.1319	tot \$ 97Q3
5230		TOTORDQ19	0.8747	0.0429	0.1309	tot orders 97Q3
5231						
5232	Cluster 15	DOLLARQ11	0.8724	0.0754	0.1380	tot \$ 95Q3
5233		TOTORDQ11	0.8724	0.0745	0.1379	tot orders 95Q3
5234						
5235	Cluster 16	DOLLARQ04	0.8768	0.0545	0.1303	tot \$ 93Q4
5236		TOTORDQ04	0.8768	0.0673	0.1321	tot orders 93Q4
5237						
5238	Cluster 17	DOLLARQ05	0.8719	0.0831	0.1398	tot \$ 94Q1
5239		TOTORDQ05	0.8719	0.0820	0.1396	tot orders 94Q1
5240						
5241	Cluster 18	DOLLARQ16	0.8665	0.0843	0.1458	tot \$ 96Q4
5242		TOTORDQ16	0.8665	0.0874	0.1463	tot orders 96Q4
5243						
5244	Cluster 19	DOLLARQ18	0.8798	0.0601	0.1279	tot \$ 97Q2
5245		TOTORDQ18	0.8798	0.0589	0.1277	tot orders 97Q2
5246						
5247	Cluster 20	DOLLARQ14	0.8448	0.0663	0.1662	tot \$ 96Q2

Output

5246						
5247	Cluster 20	DOLLARQ14	0.8448	0.0663	0.1662	tot \$ 96Q2
5248		TOTORDQ14	0.8448	0.0572	0.1646	tot orders 96Q2
5249						
5250	Cluster 21	DOLLARQ21	0.8603	0.0475	0.1467	tot \$ 98Q1
5251		TOTORDQ21	0.8603	0.0435	0.1461	tot orders 98Q1
5252						
5253	Cluster 22	DOLLARQ09	0.8736	0.0746	0.1366	tot \$ 95Q1
5254		TOTORDQ09	0.8736	0.0761	0.1368	tot orders 95Q1
5255						
5256	Cluster 23	DOLLARQ02	0.8690	0.0925	0.1443	tot \$ 93Q2
5257		TOTORDQ02	0.8690	0.0978	0.1452	tot orders 93Q2
5258						
5259	Cluster 24	DOLLARQ01	0.8757	0.1086	0.1395	tot \$ 93Q1
5260		TOTORDQ01	0.8757	0.1057	0.1390	tot orders 93Q1
5261						
5262	Cluster 25	DOLLARQ07	0.8695	0.0824	0.1422	tot \$ 94Q3
5263		TOTORDQ07	0.8695	0.0822	0.1422	tot orders 94Q3
5264						
5265	Cluster 26	DOLLARQ13	0.8453	0.0930	0.1706	tot \$ 96Q1
5266		TOTORDQ13	0.8453	0.0701	0.1664	tot orders 96Q1
5267						
5268	Cluster 27	DOLLARQ08	0.8535	0.0946	0.1618	tot \$ 94Q4
5269		TOTORDQ08	0.8535	0.1106	0.1647	tot orders 94Q4
5270						
5271	Cluster 28	PCPAYM0	1.0000	0.3108	0.0000	PCPAYM=0
5272		PCPAYM1	1.0000	0.3108	0.0000	PCPAYM=1
5273		METHPAYMCK	1.0000	0.3108	0.0000	METHPAYM=CK
5274						
5275	Cluster 29	DEPT01	0.3956	0.1594	0.7189	womens apparel
5276		DEPT02	0.3687	0.1327	0.7279	womens sleepwear
5277		DEPT03	0.4738	0.2009	0.6586	womens underwear
5278		DEPT04	0.3702	0.1596	0.7494	womens hosiery
5279		DEPT05	0.3208	0.1479	0.7971	womens footwear
5280						
5281	Cluster 30	UNTLANP0	1.0000	0.1238	0.0000	avg units/order
5282						
5283	Cluster 31	DEPT07	0.1185	0.0231	0.9024	mens apparel
5284		DEPT08	0.2880	0.0951	0.7868	mens sleepwear
5285		DEPT09	0.2913	0.0482	0.7446	mens underwear
5286		DEPT10	0.3128	0.0804	0.7473	mens hosiery



Output						
5276		DEPT02	0.3687	0.1327	0.7279	womens sleepwear
5277		DEPT03	0.4738	0.2009	0.6586	womens underwear
5278		DEPT04	0.3702	0.1596	0.7494	womens hosiery
5279		DEPT05	0.3208	0.1479	0.7971	womens footwear
5280						
5281	Cluster 30	UNTLANP0	1.0000	0.1238	0.0000	avg units/order
5282						
5283	Cluster 31	DEPT07	0.1185	0.0231	0.9024	mens apparel
5284		DEPT08	0.2880	0.0951	0.7868	mens sleepwear
5285		DEPT09	0.2913	0.0482	0.7446	mens underwear
5286		DEPT10	0.3128	0.0804	0.7473	mens hosiery
5287		DEPT11	0.2010	0.0474	0.8388	mens footwear
5288		DEPT12	0.3673	0.0902	0.6954	mens misc
5289						
5290	Cluster 32	CATALOGCNT	0.7897	0.6117	0.5416	number of catalogs received
5291		DEPT06	0.3359	0.1902	0.8201	womens misc
5292		DEPT13	0.4801	0.2840	0.7261	kitchen
5293		DEPT20	0.0326	0.0169	0.9841	furniture
5294		DEPT22	0.5862	0.3539	0.6406	household
5295		DEPT23	0.4425	0.2303	0.7243	beauty
5296		DEPT24	0.3272	0.1334	0.7764	health
5297		DEPT27	0.2906	0.1364	0.8214	outdoor
5298						
5299	Cluster 33	DEPT21	1.0000	0.0102	0.0000	light
5300						
5301	Cluster 34	DEPT18	0.5326	0.0455	0.4897	chair
5302		DEPT19	0.5326	0.0265	0.4802	window
5303						
5304	Cluster 35	TENURE	1.0000	0.1929	0.0000	months since 1st
5305						

There are 35 clusters; therefore, 35 selected variables.

# Summary of Final Cluster Solution:

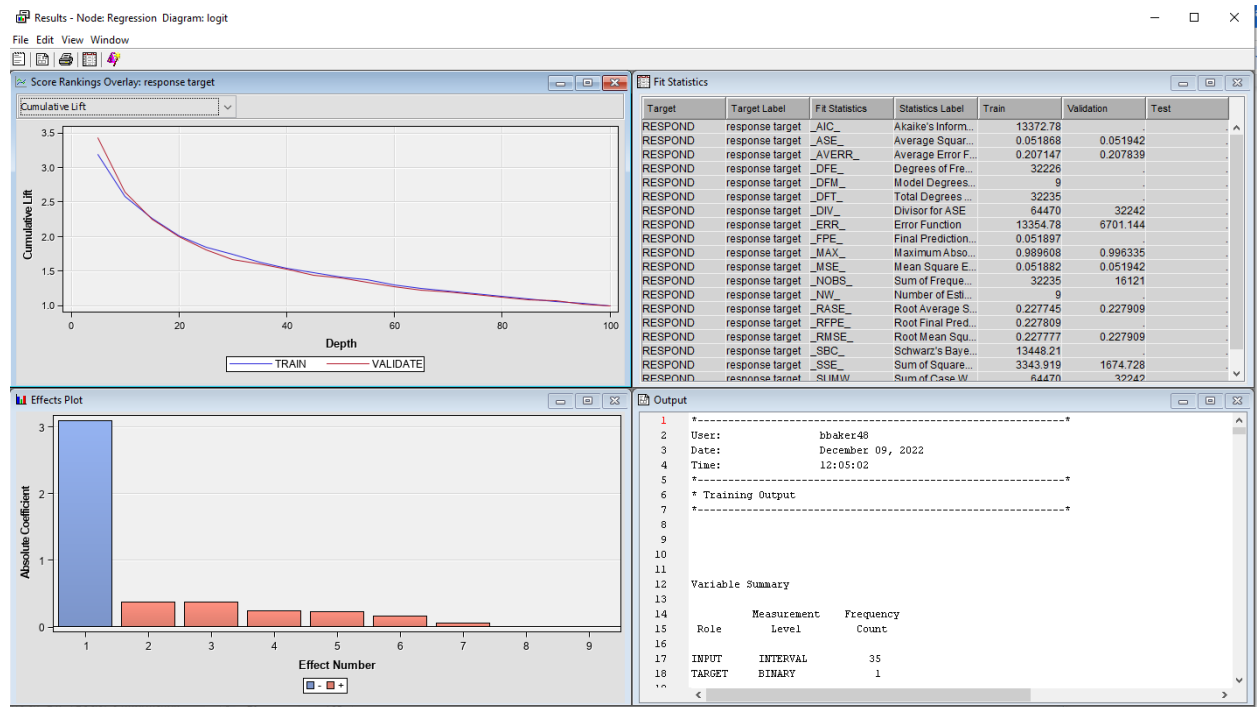
Output							
		Total	Proportion	Minimum	Maximum		Maximum
		Variation	of	Proportion	Second	Minimum	1-R**2
	Number	Explained	Variation	Explained	Eigenvalue	R-squared	Ratio
	of	by	Explained	by a	in a	for a	for a
	Clusters	Clusters	by Clusters	Cluster	Cluster	Variable	Variable
309							
310							
311	1	14.269655	0.1486	0.1486	5.075727	0.0043	
312	2	18.878088	0.1966	0.1578	3.699895	0.0106	0.9916
313	3	21.903078	0.2282	0.1877	2.748968	0.0111	0.9918
314	4	24.167157	0.2517	0.1877	2.623513	0.0111	0.9916
315	5	26.705701	0.2782	0.1893	2.073863	0.0113	0.9914
316	6	28.466963	0.2965	0.1958	2.023150	0.0113	0.9927
317	7	30.338115	0.3160	0.1958	1.906582	0.0113	1.0351
318	8	31.993414	0.3333	0.2145	1.829092	0.0117	1.0203
319	9	33.791416	0.3520	0.2145	1.734839	0.0117	1.1004
320	10	35.400976	0.3688	0.2252	1.661294	0.0119	1.0955
321	11	37.045861	0.3859	0.2333	1.645777	0.0120	1.0950
322	12	38.514796	0.4012	0.2333	1.641035	0.0120	1.0950
323	13	40.075066	0.4174	0.2462	1.578818	0.0122	1.0903
324	14	41.568625	0.4330	0.2462	1.555217	0.0122	1.0903
325	15	43.111311	0.4491	0.2462	1.548926	0.0122	1.0903
326	16	44.584612	0.4644	0.2462	1.545298	0.0122	1.0903
327	17	46.100224	0.4802	0.2538	1.537361	0.0124	1.0892
328	18	47.580873	0.4956	0.2538	1.522834	0.0124	1.0892
329	19	48.978020	0.5102	0.2538	1.518141	0.0124	1.0892
330	20	50.489323	0.5259	0.2538	1.507744	0.0124	1.0892
331	21	51.997062	0.5416	0.2538	1.501807	0.0124	1.0892
332	22	53.498395	0.5573	0.2538	1.499339	0.0124	1.0892
333	23	54.982973	0.5727	0.2615	1.482932	0.0126	1.0875
334	24	56.453832	0.5881	0.2693	1.480621	0.0125	1.0875
335	25	57.934102	0.6035	0.2693	1.474760	0.0125	1.0875
336	26	59.408841	0.6188	0.2693	1.400954	0.0125	1.0875
337	27	60.681938	0.6321	0.2693	1.327516	0.0125	1.0875
338	28	62.009454	0.6459	0.2693	1.255397	0.0125	1.0875
339	29	63.032289	0.6566	0.2862	1.237242	0.0133	1.0759
340	30	63.821130	0.6648	0.2862	1.203431	0.0133	1.0759
341	31	64.942290	0.6765	0.2631	1.137188	0.0138	1.0682
342	32	65.829529	0.6857	0.2631	1.043253	0.0167	1.0147
343	33	66.816439	0.6960	0.2631	1.043149	0.0326	1.0143
344	34	67.792590	0.7062	0.2631	1.014065	0.0326	1.0077
345	35	68.641746	0.7150	0.2631	0.984193	0.0326	0.9841

The clusters explained 71.5% of the variation in the data, circled above.

6.

Model Selection	
Selection Model	Forward
Selection Criterion	Validation Error
Use Selection Defaults	Yes
Selection Options	...
Optimization Options	

7.



# Variable Summary

Role	Measurement Level	Frequency Count
INPUT	INTERVAL	35
TARGET	BINARY	1

## Model Events

Target	Event	Measurement Level	Number of Levels	Order	Label
RESPOND	1	BINARY	2	Descending	response target

## The DMREG Procedure

### Model Information

Training Data Set	WORK.EM_DMREG.VIEW
DMDB Catalog	WORK.REG_DMDB
Target Variable	RESPOND (response target)
Target Measurement Level	Ordinal
Number of Target Categories	2
Error	MBernoulli
Link Function	Logit
Number of Model Parameters	36
Number of Observations	32235

### Target Profile

Ordered Value	RESPOND	Total Frequency
1	1	1825
2	0	30410

Shows the variables selected in the Forward Selection methods, 13 in total:

Summary of Forward Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Validation Error Rate
1	DOLINDET	1	1	418.2287	<.0001	6878.3
2	TOTORDQ20	1	2	178.2565	<.0001	6847.2
3	MONLAST	1	3	113.3610	<.0001	6781.4
4	TOTORDQ22	1	4	47.4870	<.0001	6751.4
5	CATALOGCNT	1	5	36.9828	<.0001	6727.5
6	TOTORDQ18	1	6	19.9779	<.0001	6719.0
7	TOTORDQ21	1	7	14.9769	0.0001	6712.7
8	TOTORDQ12	1	8	13.5709	0.0002	6701.1
9	TOTORDQ19	1	9	11.8344	0.0006	6702.1
10	DEPT03	1	10	10.4403	0.0012	6701.4
11	CCPAYM0	1	11	9.3003	0.0023	6709.1
12	TOTORDQ05	1	12	6.4600	0.0110	6716.5
13	DOLLARQ09	1	13	5.3211	0.0211	6717.9

The selected model, based on the error rate for the validation data, is the model trained in Step 8. It consists of the following effects:

Intercept CATALOGCNT DOLINDET MONLAST TOTORDQ12 TOTORDQ18 TOTORDQ20 TOTORDQ21 TOTORDQ22

Output

1245

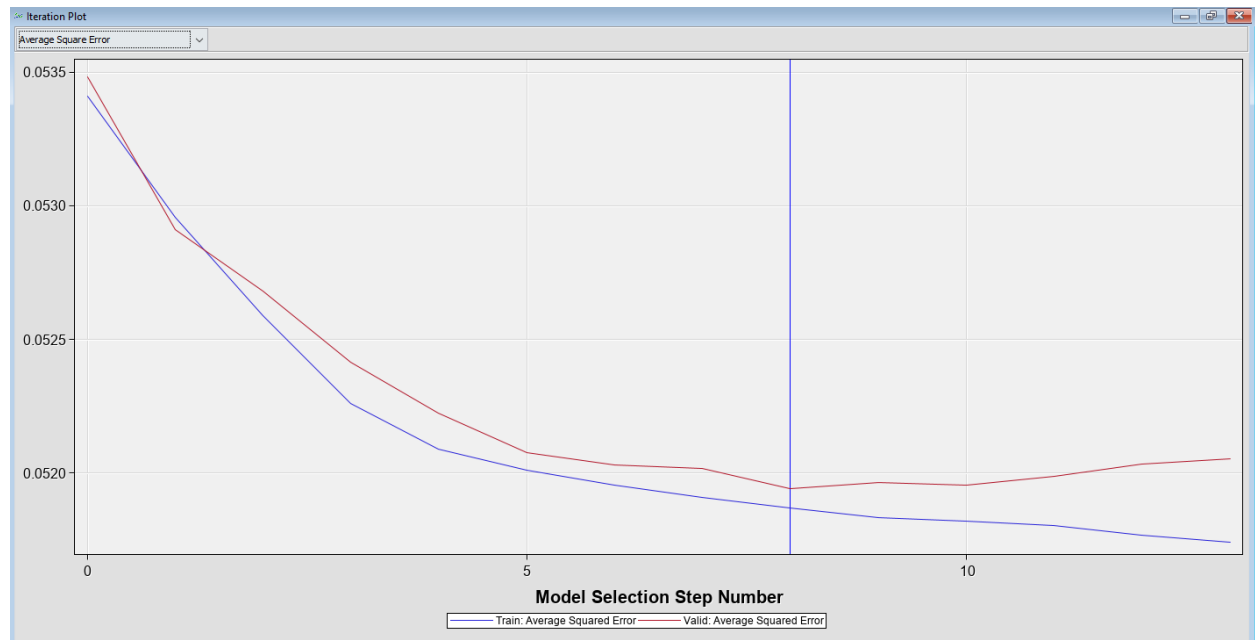
Data Role=VALIDATE Target Variable=RESPOND Target Label=response target

Posterior Probability Range	Number of Events	Number of Nonevents	Mean Posterior Probability	Percentage
0.95-1.00	1	1	0.98951	0.0124
0.90-0.95	1	0	0.91630	0.0062
0.85-0.90	0	1	0.86650	0.0062
0.70-0.75	0	1	0.71097	0.0062
0.65-0.70	1	2	0.67365	0.0186
0.60-0.65	1	0	0.62780	0.0062
0.55-0.60	2	1	0.59372	0.0186
0.50-0.55	2	3	0.51951	0.0310
0.45-0.50	2	2	0.48270	0.0248
0.40-0.45	2	1	0.42371	0.0186
0.35-0.40	4	9	0.36608	0.0806
0.30-0.35	6	15	0.31939	0.1303
0.25-0.30	15	24	0.27050	0.2419
0.20-0.25	14	69	0.22243	0.5149
0.15-0.20	41	163	0.17055	1.2654
0.10-0.15	115	748	0.11918	5.3533
0.05-0.10	383	5483	0.06734	36.3873
0.00-0.05	324	8684	0.03636	55.8774

Data Role=VALIDATE Target Variable=RESPOND Target Label=response target

Depth	Gain	Lift	Cumulative Lift	% Response	Cumulative % Response	Number of Observations	Mean Posterior Probability
5	243.140	3.43140	3.43140	19.4548	19.4548	807	0.18015
10	164.623	1.86007	2.64623	10.5459	15.0031	806	0.10221
15	125.304	1.46617	2.25304	8.3127	12.7739	806	0.08419
20	99.622	1.22546	1.99622	6.9479	11.3178	806	0.07390
25	80.710	1.05039	1.80710	5.9553	10.2456	806	0.06746
30	66.643	0.96286	1.66643	5.4591	9.4480	806	0.06347
35	59.719	1.18169	1.59719	6.6998	9.0555	806	0.05996
40	52.065	0.98474	1.52065	5.5831	8.6215	806	0.05518
45	44.409	0.83156	1.44409	4.7146	8.1875	806	0.05136
50	39.379	0.94098	1.39379	5.3350	7.9022	806	0.04757
55	33.472	0.74403	1.33472	4.2184	7.5674	806	0.04496
60	27.639	0.63461	1.27639	3.5980	7.2366	806	0.04298
65	22.366	0.59085	1.22366	3.3499	6.9377	806	0.04116
70	19.097	0.76591	1.19097	4.3424	6.7523	806	0.03973
75	15.680	0.67838	1.15680	3.8462	6.5586	806	0.03797
80	11.869	0.54708	1.11869	3.1017	6.3426	806	0.03586
85	8.378	0.52520	1.08378	2.9777	6.1446	806	0.03348
90	6.552	0.75497	1.06552	4.2804	6.0411	806	0.03016
95	2.268	0.25166	1.02268	1.4268	5.7982	806	0.02514
100	0.000	0.56896	1.00000	3.2258	5.6696	806	0.01863

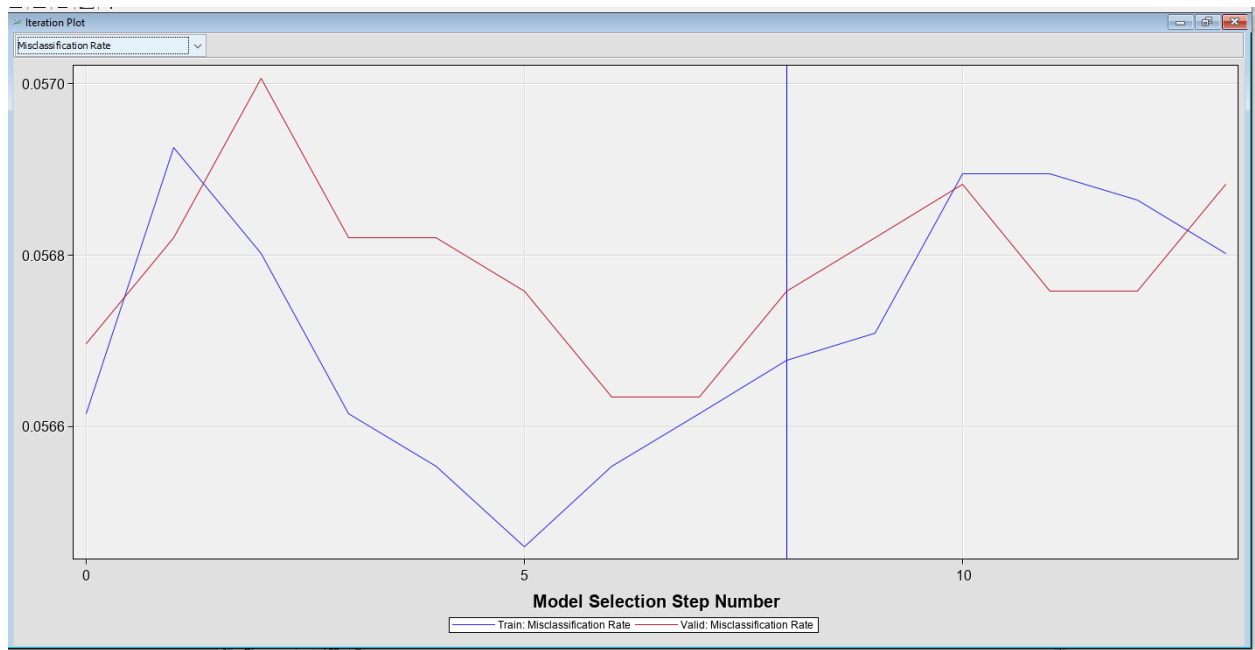
8.



Above shows that the smallest average squared error occurs in model 8, represented by the vertical line

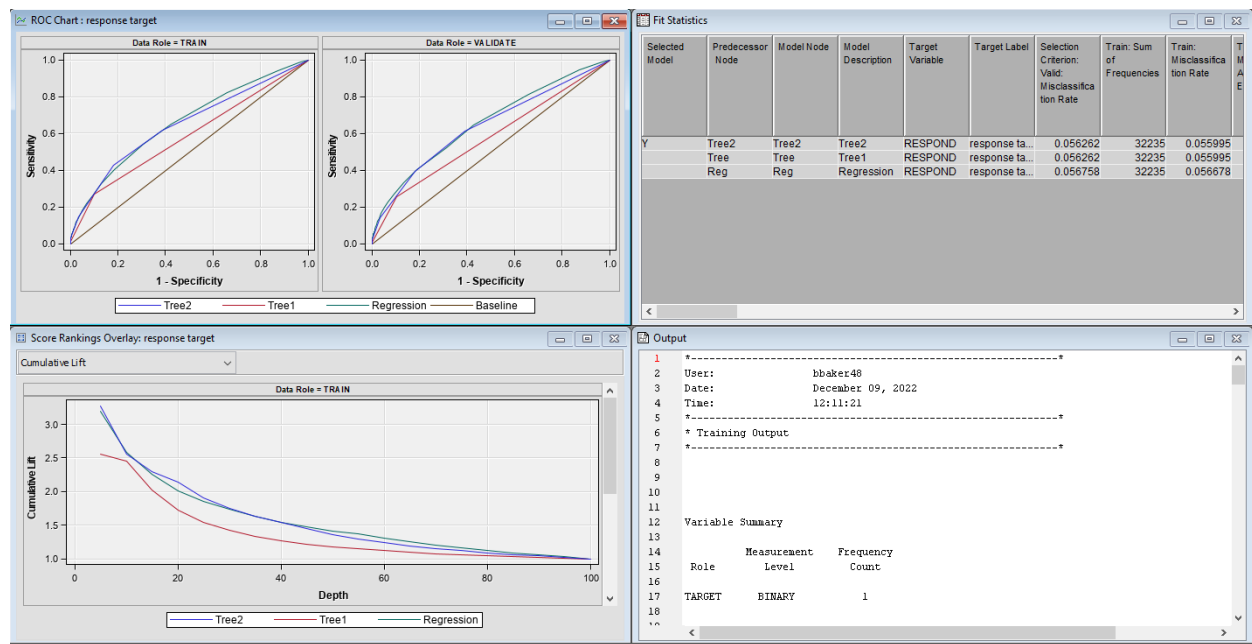


9.



Above shows that the model with the smallest misclassification occurs in steps 6 & 7, represented by the vertical line.

10.



The ROC chart above shows that two of the three models have good predictive accuracy as they deviate above the 45% angle. The logistic regression and Tree 2 models perform similarly on the validation data set, but, the logistic regression performs slightly better.

Data Role=Valid

Statistics	Tree2	Tree	Reg
Valid: Kolmogorov-Smirnov Statistic	0.22	0.15	0.22
Valid: Average Squared Error	0.05	0.05	0.05
Valid: Roc Index	0.64	0.58	0.65
Valid: Average Error Function	.	.	0.21
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.05	0.09	0.06
Valid: Cumulative Percent Captured Response	24.16	22.63	26.48
Valid: Percent Captured Response	8.53	10.97	9.30
Valid: Divisor for VASE	32242.00	32242.00	32242.00
Valid: Error Function	.	.	6701.14
Valid: Gain	141.49	126.21	164.62
Valid: Gini Coefficient	0.28	0.15	0.31
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.22	0.15	0.22
Valid: Kolmogorov-Smirnov Probability Cutoff	0.04	0.12	0.05
Valid: Cumulative Lift	2.41	2.26	2.65
Valid: Lift	1.71	2.20	1.86
Valid: Maximum Absolute Error	0.96	0.95	1.00
Valid: Misclassification Rate	0.06	0.06	0.06
Valid: Mean Square Error	.	.	0.05
Valid: Sum of Frequencies	16121.00	16121.00	16121.00
Valid: Root Average Squared Error	0.23	0.23	0.23
Valid: Cumulative Percent Response	13.69	12.83	15.00
Valid: Percent Response	9.67	12.45	10.55
Valid: Root Mean Square Error	.	.	0.23
Valid: Sum of Squared Errors	1676.25	1693.19	1674.73
Valid: Sum of Case Weights Times Freq	.	.	32242.00