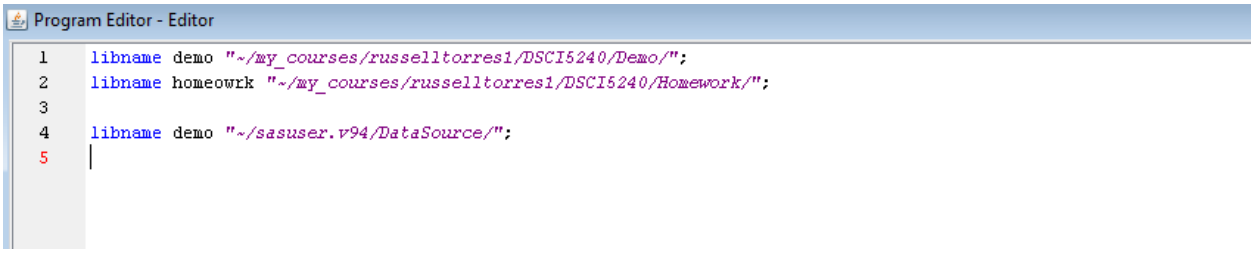
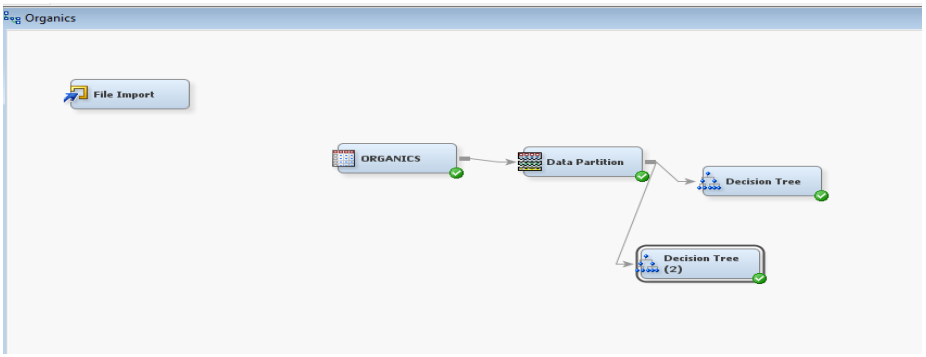
Homework 4- Masters Group 9

Group Members:

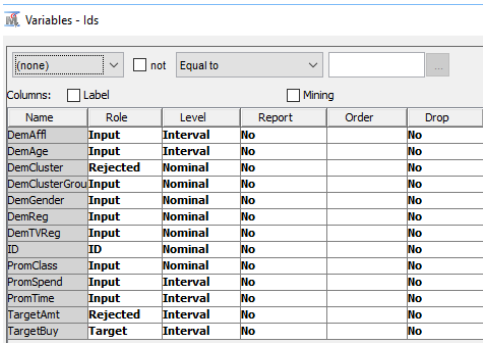
A) Oluwaseyitan Awojobi
B) Mohamed Elgendy
C) Ramya Koya
D) Karthik Sagar Tadi
E) Vijay Tulluri

SOLUTION



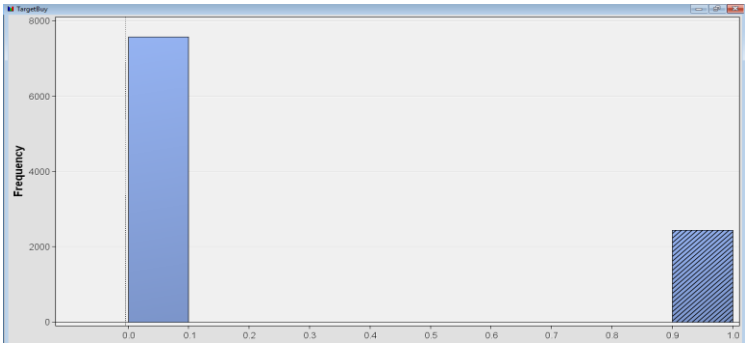


4b. With a total of 10,000 variables and using the explore button on SEM, it **appears 2,434 persons** bought organic products while 7,566 did not purchase. This means we had **24.34%** of people purchase the products.



Name	Role	Level	Report	Order	Drop
DemAffl	Input	Interval	No		No
DemAge	Input	Interval	No		No
DemCluster	Rejected	Nominal	No		No
DemClusterGrou	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	Interval	No		No

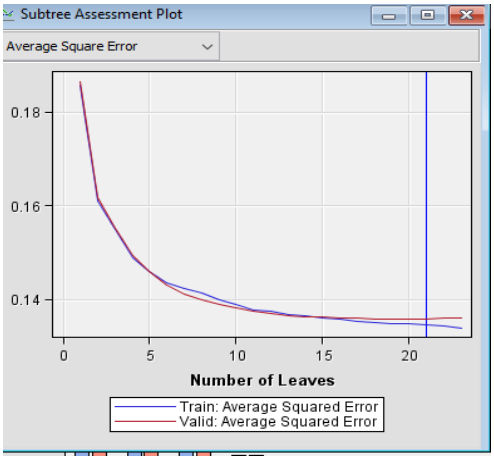
4d. TargetAmt variable should not be included because it contains **repeated values** which are available in the TargetBuy hence makes it redundant.

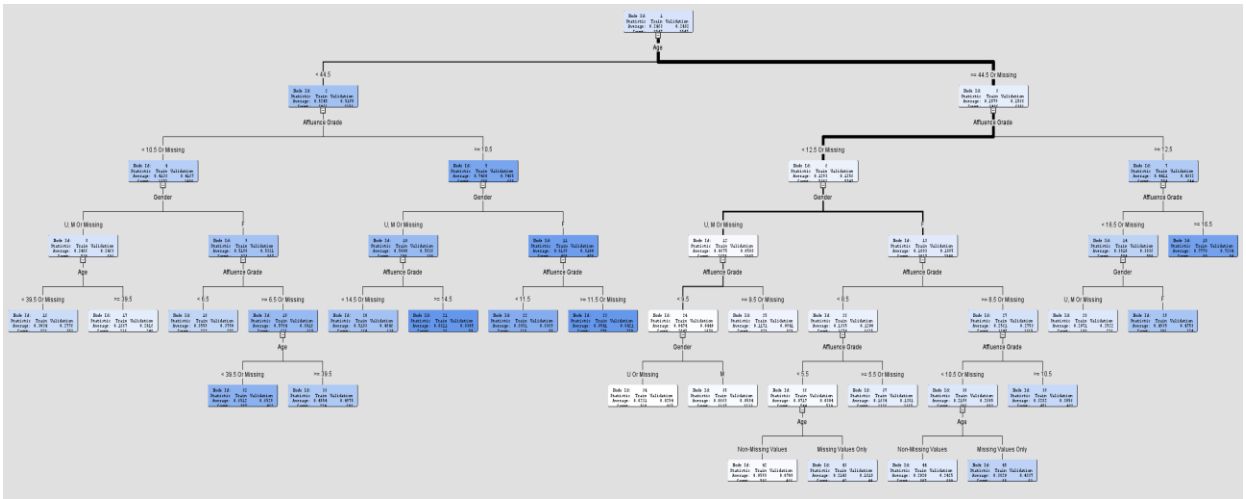


8. There are 45 leaf's here and using the average square error as the model assessment statistics, there are **21 leaves** in **the optimal tree.**

8b. The age variable was used for the first split.

8c. Competing splits were **<44.5 and >= 44.5 and Affluence Grade**





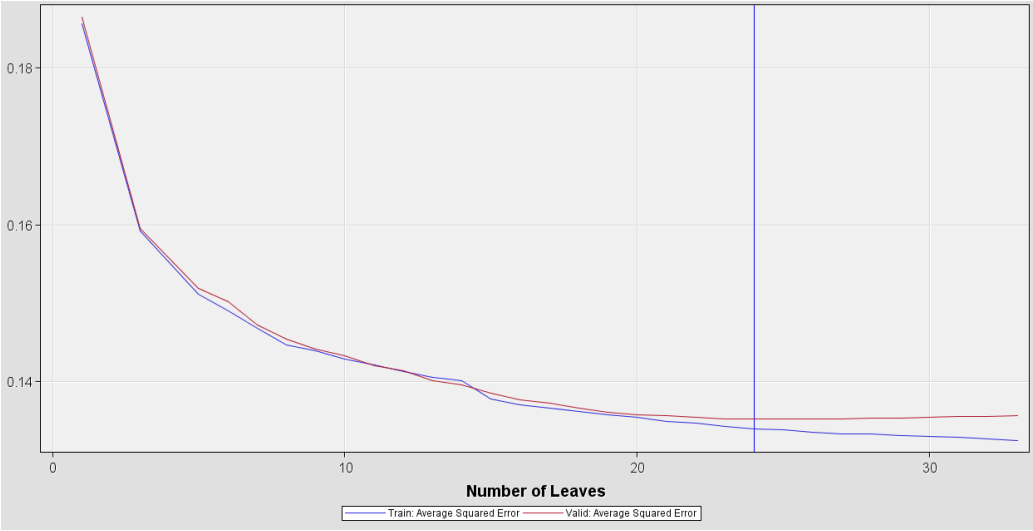
8. There are 55 leaves in the 3-way splits branch Decision Tree. Using the Average square error as the model assessment statistic, the **optimal tree** has **24 leaves.**

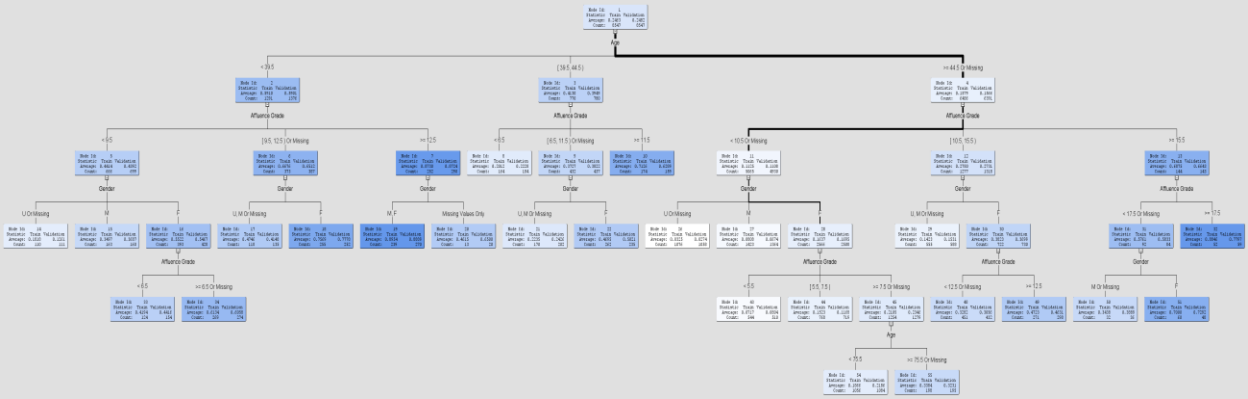
9b. Using the Average square error, while both appear to be similar with value around **0.134** for the training data and the validation data set.

For the Decision tree with 2 branches, Training had an error of **0.1346** while validation had an error of **0.1358.**

For the Decision tree with 3 branches, Training data had an error of **0.1340** and validation data of **0.1352.**

This tells us that the Decision tree with 3 branches is slightly better than that of 2 branches.





Using the model comparison node, it appears that the **second Decision tree is better**. The selection criterion however has very little difference. Decision tree 2 has an Average Squared error of 0.352 while Decision tree 1 has an Average square error of 0.1358.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Sum of Frequencies	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Maximum Absolute Error
Y	Tree2	Tree2	Decision Tr...	TargetBuy	Organics P...	0.135243	8547	0.967472	1145.516	0.134026	0.366095	8547	8547	8547	0.967472
	Tree	Tree	Decision Tr...	TargetBuy	Organics P...	0.13582	8547	0.976923	1150.836	0.134648	0.366944	8547	8547	8547	0.976923

Target Variable