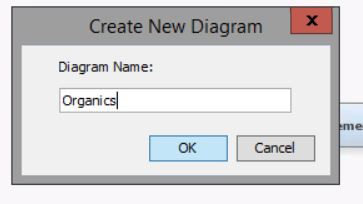
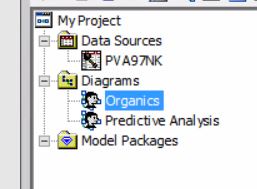
Exercise 15

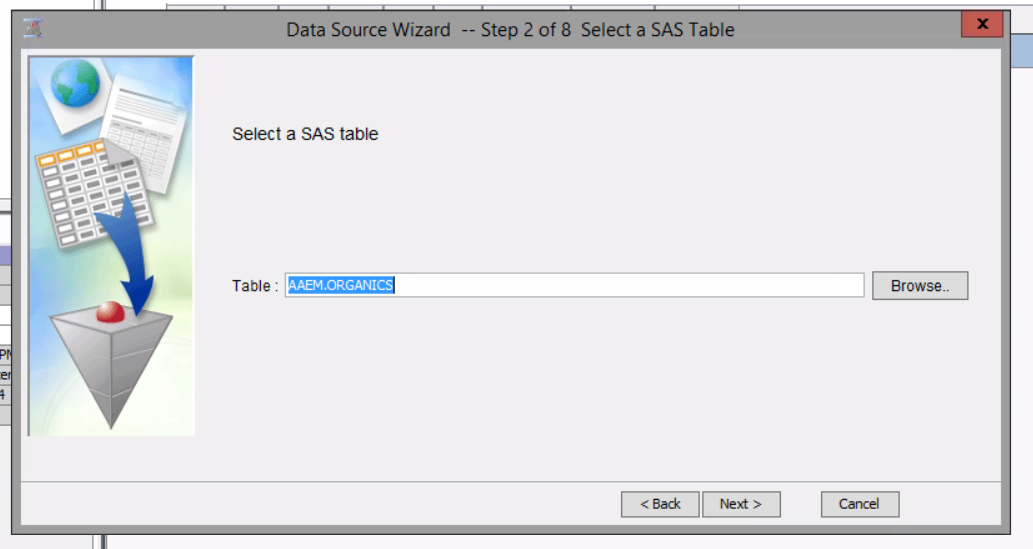
Problem 1

1. Creating a new diagram named **Organics**

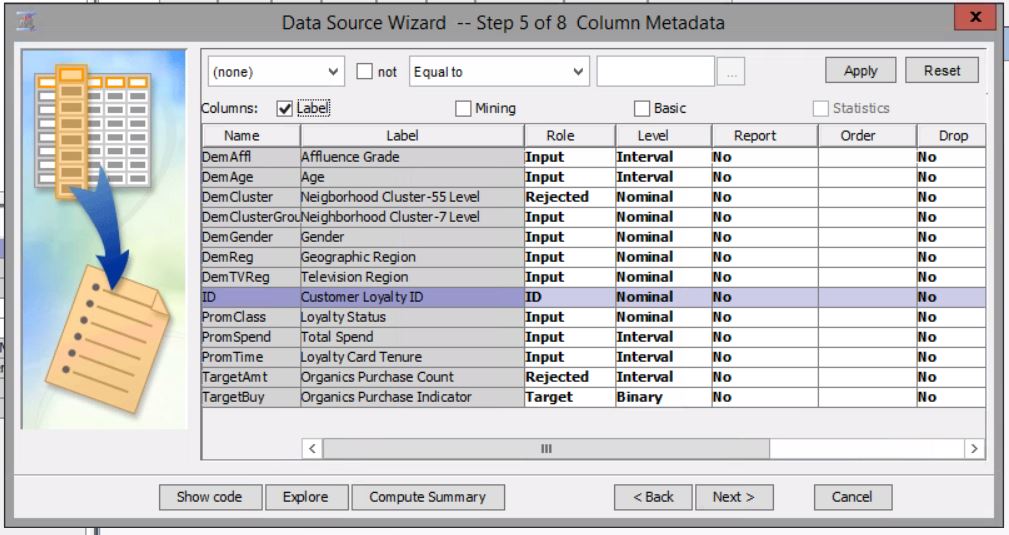




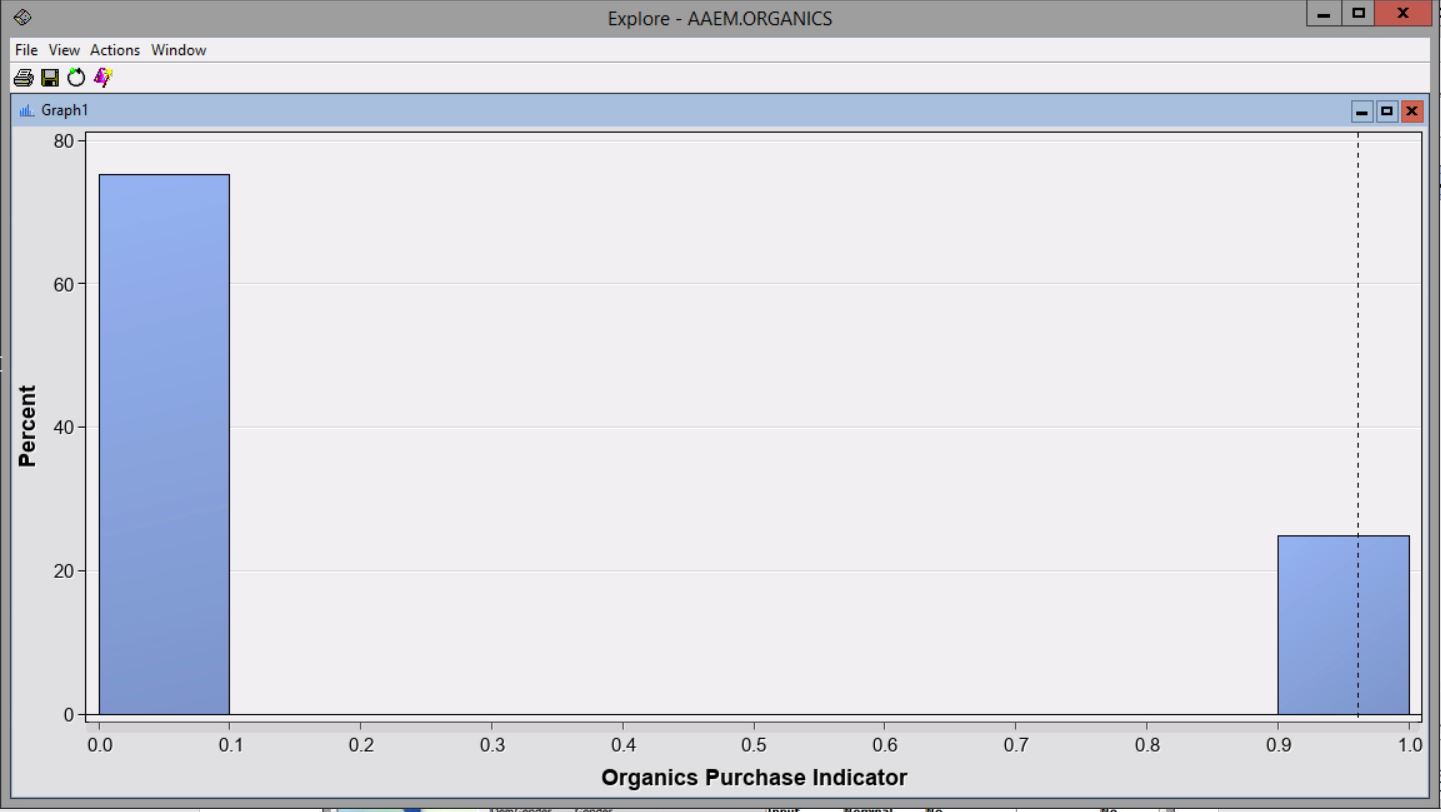
1. Define the data set **AAEM.ORGANICS** as a data source for the project



1. Set the model roles for the analysis variables

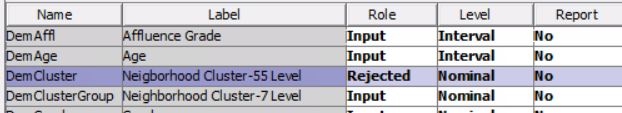


1. Examine the distribution of the target variable. What is the proportion of individuals who purchased organic products?



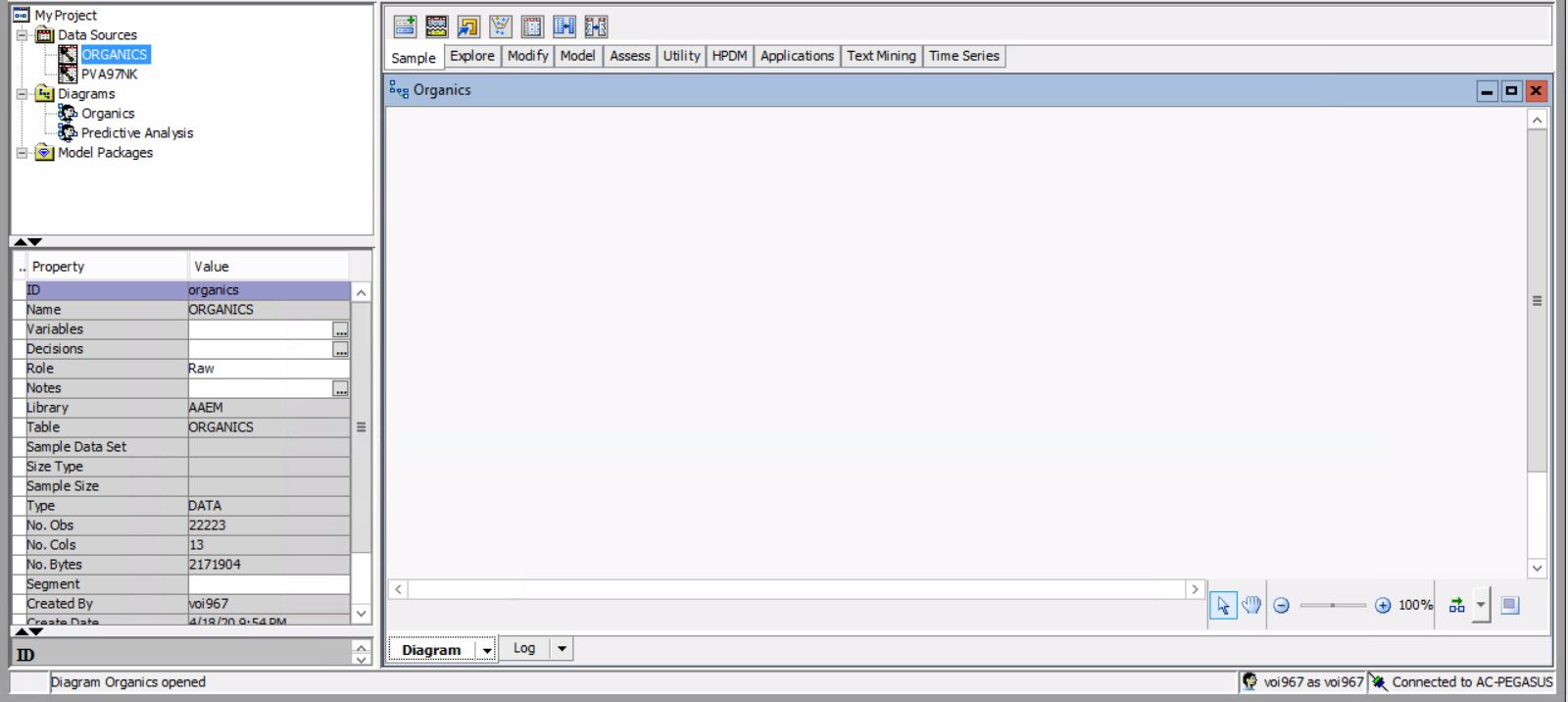
* One out of every four inputs or about 25% of individuals purchased organic products

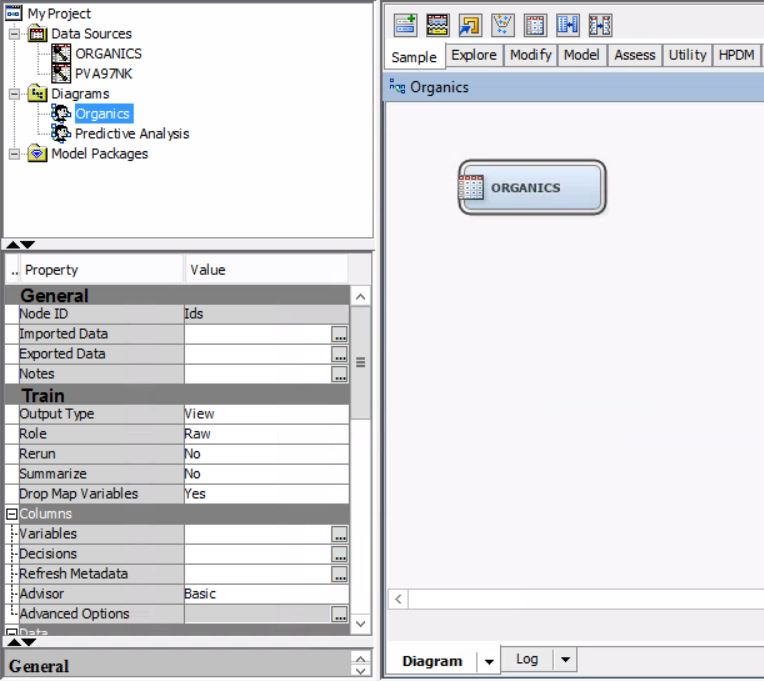
1. The variable **DemClusterGroup** contains collapsed levels of the variable **DemCluster**. Presume that, based on previous experience, you believe that **DemClusterGroup** is sufficient for this type of modeling effort. Set the model role for **DemCluster** to Rejected.



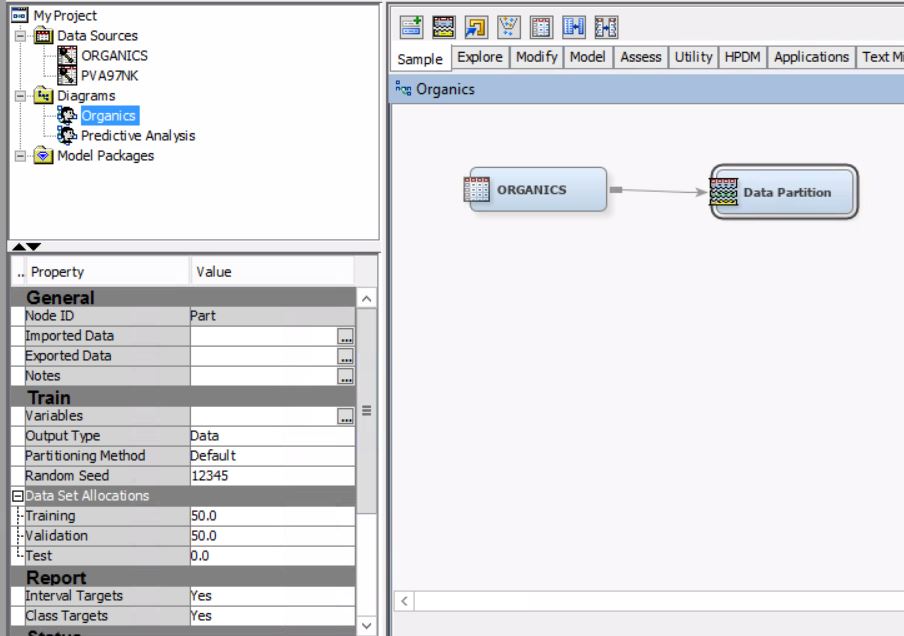
1. As noted above, only **TargetBuy** can be used for this analysis and should have a role of Target. Can **TargetAmt** be used as an input for a model used to predict **TargetBuy**? Why or why not?

* No, because TargetAmt is redundant and irrelevant to the target variable being tested. The target variable is only checking whether individuals are buying organic products or not.

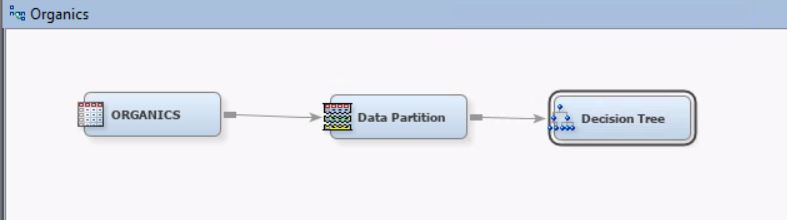
1. Finish the **Organics** data source definition.
2. Add the AAEM.ORGANICS data source to the Organics diagram workspace.



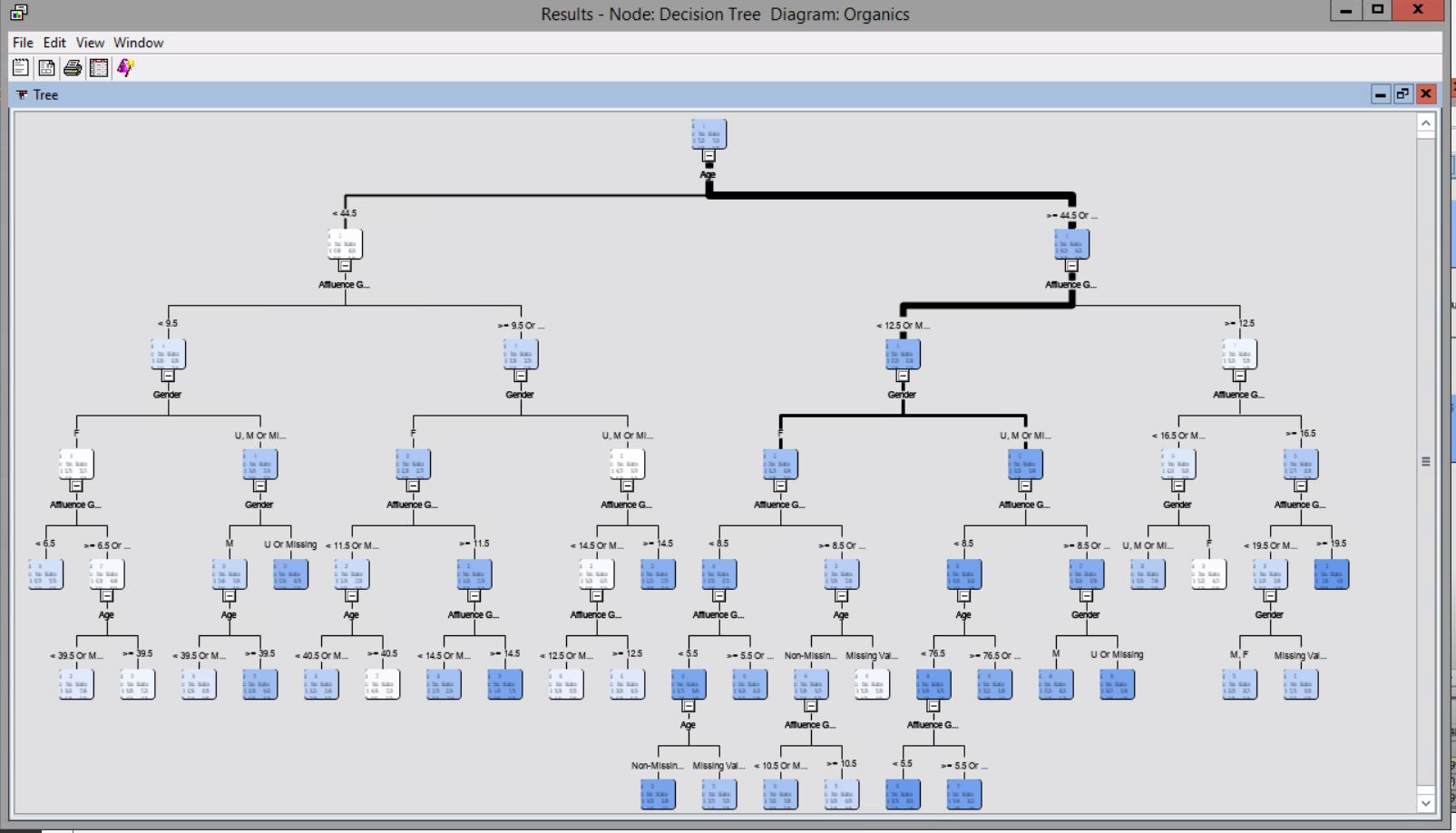
1. Add a **Data Partition** node to the diagram and connect it to the **Data Source** node. Assign 50% of the data for training and 50% for validation.



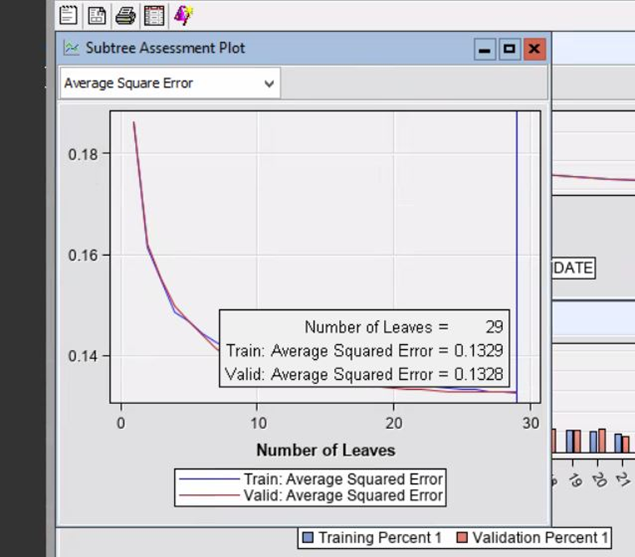
1. Add a **Decision Tree** node to the workspace and connect it to the **Data Partition** node.



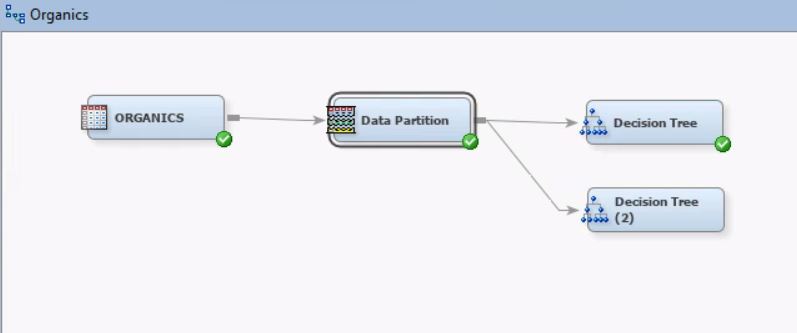
1. Create a decision tree model autonomously. Use average square error as the model assessment statistic.



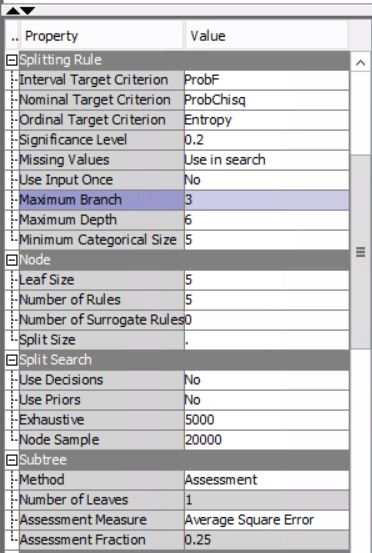
1. How many leaves are in the optimal tree?
   * 29 leaves



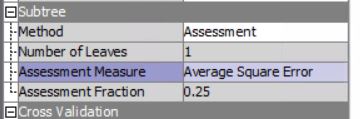
1. Which variable was used for the first split?
   * Age is the first split
2. What were the competing splits for the first split?
   * Competing splits are Affluence Grade and Gender
3. Add a second **Decision Tree** node to the diagram and connect it to the **Data Partition** node.



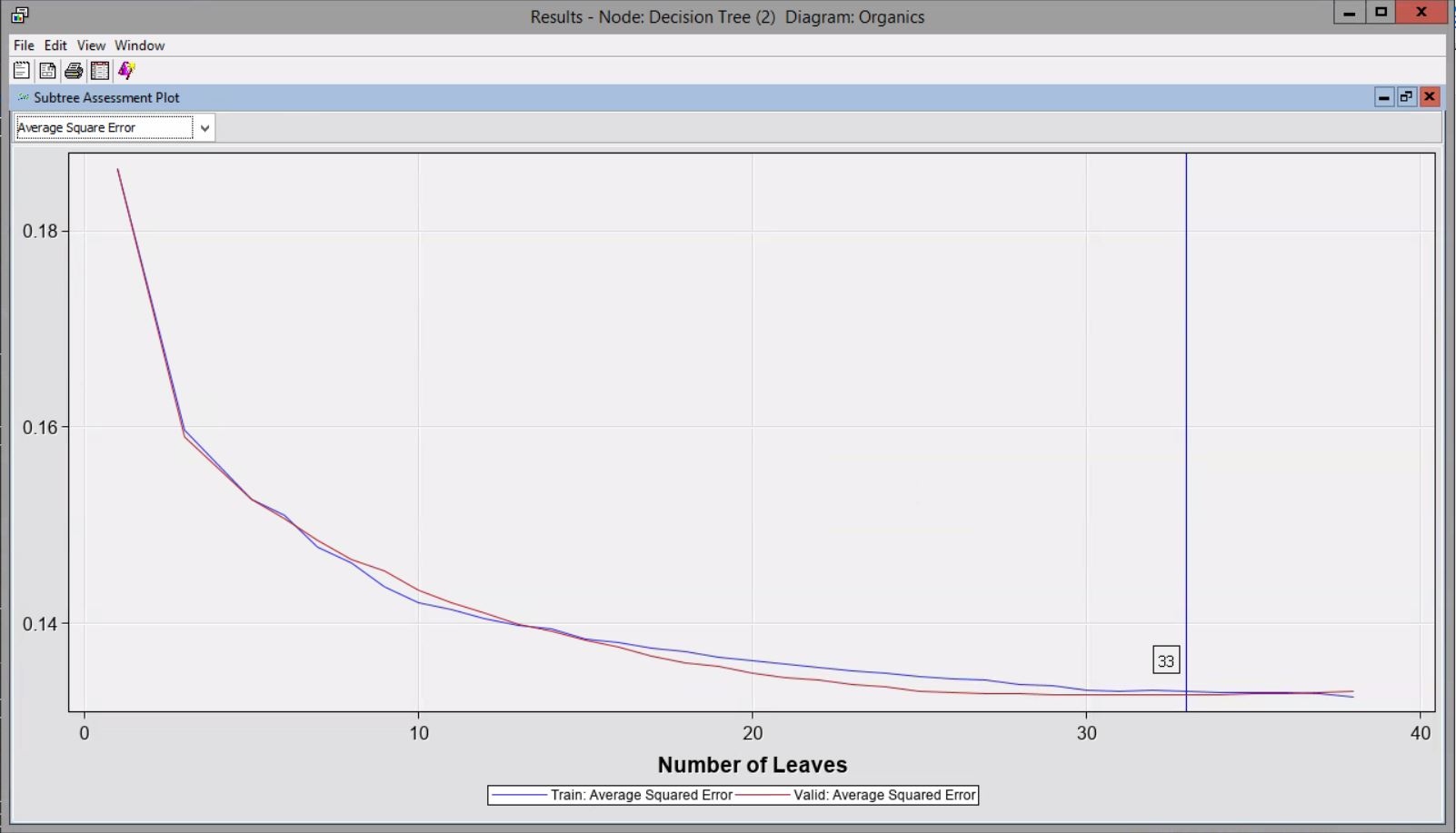
1. In the Properties panel of the new Decision Tree node, change the maximum number of branches from a node to 3 to enable three-way splits.

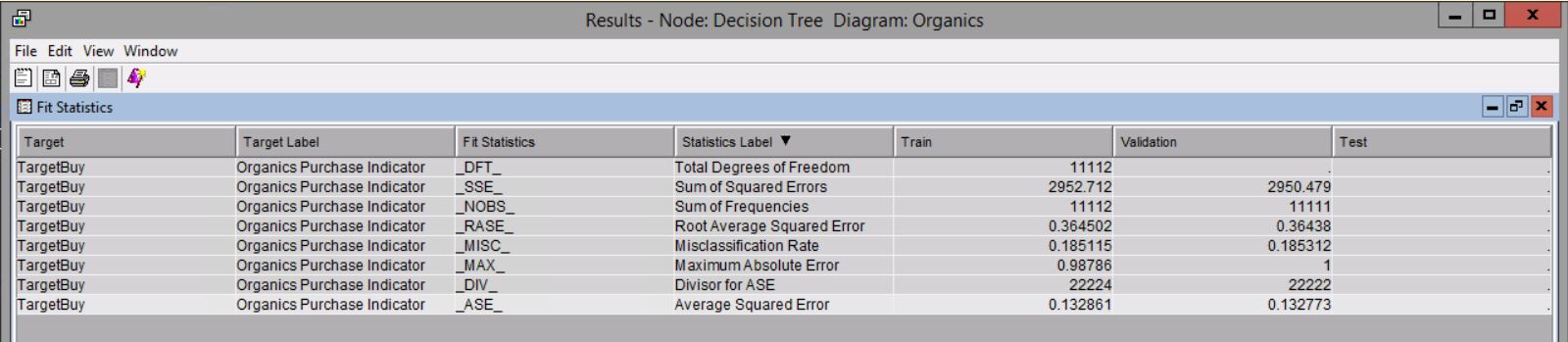
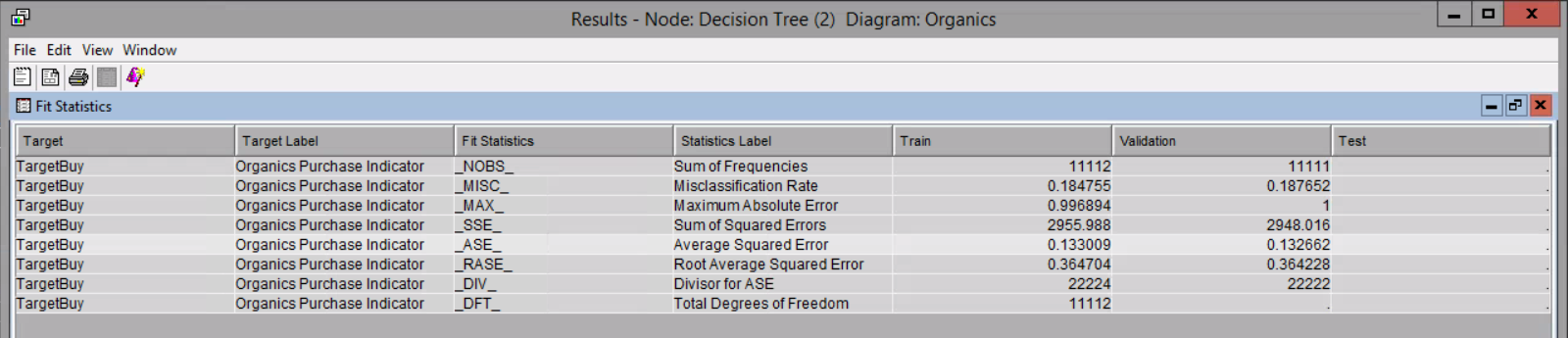


1. Create a decision tree model. Use average square error as the model assessment statistic.



1. How many leaves are in the optimal tree?
   * 33 leaves



1. Based on average square error, which of the decision tree models appears to be better?
   * Decision Tree
   * Decision Tree (2)
   * The second tree with the three way splits has a lower validation average square error and is the better model based on the fitted statistics results of both models.
2. Save the project as Exercise 15.