**Title: Can we Predict Customers’ Behavior?**

**Students: Kurt Yokley & Brandon Robertson**

1. **Business Problem**

A supermarket is offering a new line of organic products. The supermarket's management wants   
to determine which customers are likely to purchase these products.

The supermarket has a customer loyalty program. As an initial buyer incentive plan, the supermarket provided coupons for the organic products to all of the loyalty program participants and collected data that includes whether these customers purchased any of the organic products.

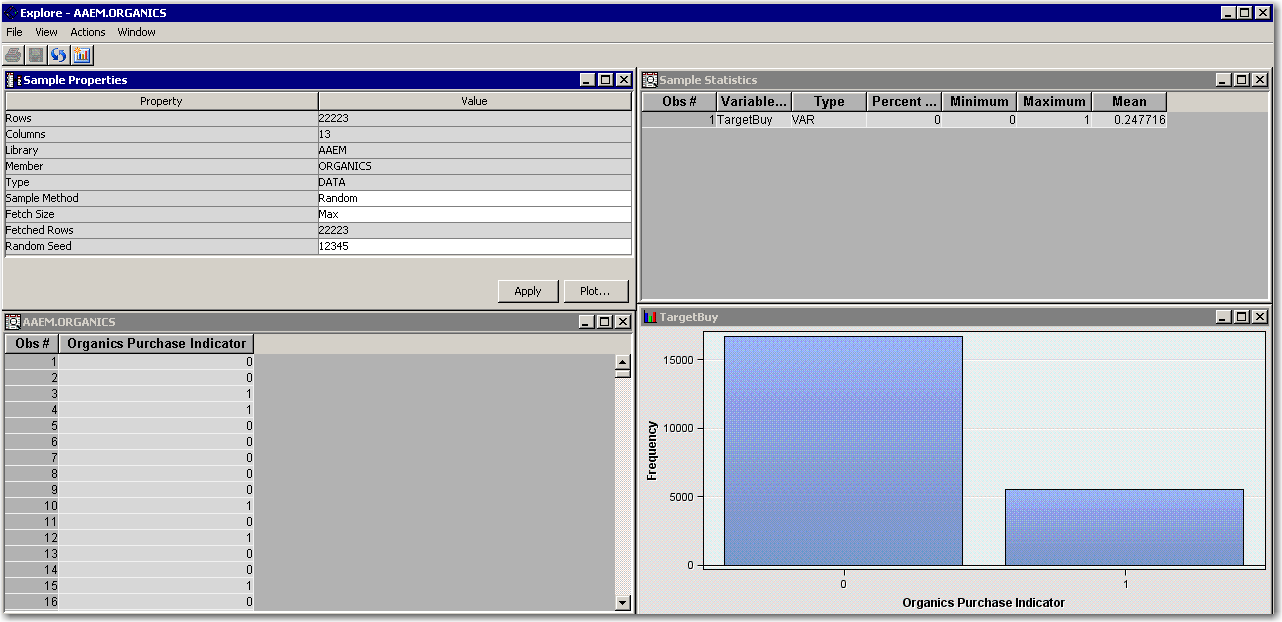
1. **Research Questions & Building Model**
   * How do the purchase orders vary by gender?
   * Does age play a role in purchasing orders?
2. **Data description**

The ORGANICS data set contains:

* 13 variables and
* over 22,000 observations.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Model Role | Measurement Level | Description |
| ID | ID | Nominal | Customer loyalty identification number |
| DemAffl | Input | Interval | Affluence grade on a scale from 1 to 30 |
| DemAge | Input | Interval | Age, in years |
| DemCluster | Rejected | Nominal | Type of residential neighborhood |
| DemClusterGroup | Input | Nominal | Neighborhood group |
| DemGender | Input | Nominal | M = male, F = female, U = unknown |
| DemRegion | Input | Nominal | Geographic region |
| DemTVReg | Input | Nominal | Television region |
| PromClass | Input | Nominal | Loyalty status: tin, silver, gold, or platinum |
| PromSpend | Input | Interval | Total amount spent |
| PromTime | Input | Interval | Time as loyalty card member |
| TargetBuy | Target | Binary | Organics purchased? 1 = Yes, 0 = No |
| TargetAmt | Rejected | Interval | Number of organic products purchased |

1. **Data Exploration**



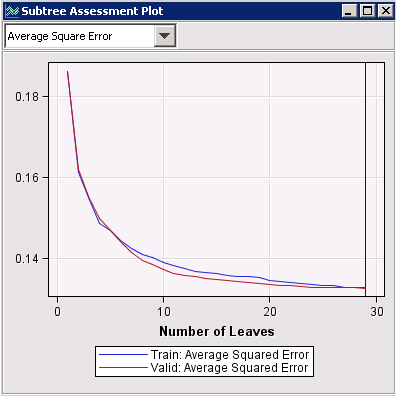
* The proportion of individuals who purchased organic products appears to be 25%.

1. **Data Partition:**

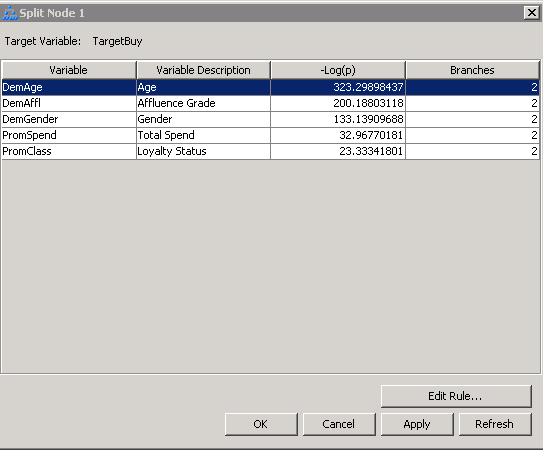
* Assign 50% of the data for training and 50% for validation.

1. **Implementations and Findings**

We created a decision tree model. We used average square error as the model assessment statistic. The easiest way to determine the number of leaves in your tree is via the iteration plot. We also used Subtree Assessment Plot to assess the model.



* Using average square error as the assessment measure results in a tree with 29 leaves.



* Age is used for the first split.
* Competing splits are Affluence Grade, Gender, Total Spend, and Loyalty Status.

1. **Conclusion**

Our data showed us that the solutions to acquiring more customer purchases is to focus on three areas of business. The first is to focus on women since the data shows they have a higher purchase rate. The second is to market towards middle aged customers instead of the older generation. The third is the silver and tin loyalty members. These customers are the top buyers and those are the ones that need to be targeted.