***Part 1***

**Completion of the initial tutorial with Catalog Data (10points)**

**Task 1:**

**SAS Code (5points)**

* **Univariate Analysis**
* **Outlier**
* **Plotting**

You need to perform additional analysis on various variables and make a report

1. You might want to study which variables are highly correlated. If you find such variables you can suggest dimension reduction by dropping one of the variables. **6points**
2. You can study in if there are outliers in your variables **6points**
3. You can make 3-D plots to get a better sense of how independent variables affect the dependent variables **6points**

**Task 2: 10 points**

**SAS Code (5points)**

* **Univariate Analysis**
* **Outlier**
* **Plotting**

A national veterans’ organization seeks to better target its solicitations for donation. By soliciting only, the most likely donors, less money is spent on solicitation efforts and more money is available for charitable concerns. Solicitations involve sending a small gift to an individual and including a request for a donation. Gifts to donors include mailing labels and greeting cards.

The organization has more than 3.5 million individuals in its mailing database. These individuals are classified by their response behaviors to previous solicitation efforts. Of particular interest is the class of individuals identified as *lapsing donors*. These individuals made their most recent donation between 12 and 24 months ago. The organization seeks to rank its lapsing donors based on their responses to a greeting card mailing sent in June of 1997. (The charity calls this the 97NK Campaign.) With this ranking, a decision can be made to either solicit or ignore a lapsing individual in the June 1998 campaign.

***Part 2***

**Task 3 Organic Dataset Decision Tree**

**Completion and correct Diagrams (8 points)**

**SAS Code (2 points)**

1. Set the roles for the analysis variables as shown above.
2. Examine the distribution of the target variable. What is the proportion of individuals who purchased organic products?
3. 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.
4. 4)  As noted above, only **TargetBuy** is used for this analysis, and it should have a role of **Target**. Can **TargetAmt** be used as an input for a model that is used to predict **TargetBuy**? Why or why not?
5. Finish the **ORGANICS** data source definition.

Add the **ORGANICS** data source to the Organics diagram workspace.

Add a **Data Partition** node to the diagram and connect it to the **Data Source** node. Assign 65% of the data for training and 35% for validation.

Add a **Decision Tree** node to the workspace and connect it to the **Data Partition** node. Create a **decision tree** model autonomously. Use **Misclassification** as the model assessment statistic.

* 1. How many leaves are in the optimal tree? **2points**
  2. Which variables were used for the first split? What were the competing splits for this first split? **2points**
  3. Which variables were used for the second split for all branches from first split? **2points**
  4. Discuss the results and provide your insights? **2points**

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 allow for three-way splits.
2. Create a decision tree model using **Misclassification** as the model assessment statistic. **2points**
3. How many leaves are in the optimal tree? **2points**

Based on **Misclassification rate**, which of the decision tree models appears to be better?

***Part 3***

**Logit 10 Points**

Completion and correct procedure followed