**Assignment 6 - SAS #1 – Introduction to Working with SAS**

For this assignment, you’ll be adding a new diagram ‘Organics’ under the project built in in-class exercise ‘My Project’. Below is the path of the data source, after chose Metadata Repository:

Shared Data 🡪 Libraries 🡪 AAEM 🡪 Organics

You can fill in the blanks right on the sheet. You can insert the required screen shots right in the document.

**Guidelines**

* You must submit the answers to the questions on the last page of this document.
* You must include your name at the top of the document.
* Your answers should be submitted at Blackboard.
* It must be sent by the start of class the day the assignment is due.

***If you do not follow these instructions, your assignment will be counted late.***

**Evaluation**

Your submission will be graded based on whether the answers provided for each question are correct.

**Name:**

1. **Initial Data Exploration**

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.

The **ORGANICS** data set contains 13 variables and over 22,000 observations. The variables in the data set are shown below with the appropriate roles and levels:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **ModelRole** | **MeasurementLevel** | **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 |

**🖉** Although two target variables are listed, this exercise concentrates on the binary variable **TargetBuy**.

* 1. Create a new diagram named **Organics**.
	2. Define the data set **AAEM.ORGANICS** as a data source for the project.
		1. Set the roles for the analysis variables as shown above.
		(You can go back and modify variable roles even after you complete the wizard by right-clicking on the **Organics** data source and selecting **Edit Variables…**)

		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.

**Include a screen shot showing DemCluster is rejected (as on page 17 of the in-class exercise).**

* + 1. Examine the distribution of the target variable **TargetBuy**. You can do this by clicking on that variable in the Column Metadata (step 6 of 9 in the wizard) and then clicking the **Explore** button.

		What is the proportion of individuals who purchased organic products (hint: take a look at the “Sample Statistics” window)?

		**ANSWER:**
		2. Finish the **Organics** data source definition.
	1. Add the **AAEM.ORGANICS** data source to the **Organics** diagram workspace.
	2. Explore the **Organics** data source and answer the following questions:

	Are there more males or more females in the sample? **ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
	(Hint - plot **DemGender** using a bar chart)

	**Include a screen show showing the bar chart (as on page 31 of the in-class exercise)**.

	What is maximum age of the people in the sample? **ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

	Plot **DemAffl** using a histogram with 30 “X” bins. Based on that, is it more common to be very affluent (wealthy) or very poor?
	 **ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

	Looking only at the **DemAffl** histogram, people in this sample most frequently receive an affluence grade between
	 **ANSWER:** \_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_