

**Generating and Assessing Customer Profiles using Clustering Method**



Bala Durga Sridevi Kesavarapu

**1. Generating and Assessing Customer Profiles using Clustering Method:**

*Clustering* is also known as unsupervised classification and a method to classify data with an unknown target. As the class and number of classes of each case unknown, the goal of *clustering* is to describe the data. In the department store information provided, the goal is to group existing customers for profiling and future targeted marketing campaigns.

**DATA -** Macy's Department Store dataset contains customer demographics information as well as purchase counts by department for 1966 of its customers.

**AIM** -

* Identify 5 customer profiles using a cluster analysis model.
* Develop a predictive model for each cluster.

**CLUSTERING ANALYSIS**

*Building the Process Flow Diagram* ([Screenshot 1.0](#y5gb9akqebz0))

1. Add the MACYS data source to the project

* In the Project Panel, right-click **Data Sources** and click **Create Data Source**.
* In the Data Source Wizard — Metadata Source window, click **Next**.
* In the Data Source Wizard — Select a SAS Table and enter **MACYS** in the **Table** field. Click **Next**.
* In the Data Source Wizard — Table Information window, click **Next**.
* In the Data Source Wizard — Metadata Advisor Options window, click **Basic**. Click **Next**.
* In the Data Source Wizard — Column Metadata window, make the following changes ([Screenshot 1.1](#qm9ly4cb33t1))
  + Set ACCTNUM role to **Rejected**. This was done because account numbers are unique to each customer and cannot be clustered.
  + Set NTITLE to **Rejected**. This was done because we already have data fields indicated sex(SEX) and marital status (MARITAL).
  + Set STATECOD to **Rejected**. This was done because there are too many states which fragments the clusters. Trying to cluster 50 states into 5 profiles resulted in a large STATECOD segment of **OTHER** which contained customers outside the top 10 states.
  + Set MARITAL to Binary as Yes/No values only
  + Set SEX to Binary as Male/Female values only.
* Data Assumptions
  + AGE is account age in months, not customer age, otherwise the data would indicate 15-year-old mechanics with 2 cars and having kids.
  + STATECOD AE is Active Duty Military. Most of these entries were filtered out in the FILTER Node due to income being 0.
* Data Recommendations
  + STATECOD data would be more useful for clustering if the data were transformed to a nominal variable indicating geographic regions - Southeast, Midwest, Northeast, etc.
  + Future analysis could use the TRANSFORM Node to map STATECOD into geographic regions.

1. Create the FILTER Node
   * INCOME was filtered to be greater than 10,000 as the target market should have a viable income source ([Screenshot 1.2](#ng6fodm7ocso)).
   * HOMEVAL was filtered to be greater than 20,000 ([Screenshot 1.3](#g9lqva2lk72y)).

* The FILTER Node configuration resulted in 483 out of 1966 entries being filtered out. 1484 entries remained which is 75% of our starting data.

1. Create the CLUSTER Node

* Configure the CLUSTER Node to have a maximum of 5 clusters.
* The following variables are purchase history data and were not used for clustering purposes as we wanted to cluster based on customer demographics instead of purchase history ([Screenshot 1.4](#3184t5w1o3db)). However, we did create a separate path for the purchase cluster in case future analysis is required ([Screenshot 1.13](#qroznw2m8k4m)).
  + APPAREL, BLANKETS, COATS, DISHES, DOMESTIC, FLATWARE, HHAPPAR, HOMEACC, JEWELRY, KITCHEN, LAMPS, LEISURE, LINENS, LUXURY, MENSWARE, OUTDOOR, TOWELS, WAPPAR, WCOATS

1. Create the SEGMENT PROFILE Node

* Default values were used for the SEGMENT PROFILE Node

*Clustering Analysis Results*

The Cluster Analysis Model yielded 5 clusters as configured ([Screenshot 1.5](#vndluirydvh8)). Below are our cluster names and profile characteristics/descriptions. Also included are recommendations on how to target each clustering in a marketing campaign.

For each demographic based cluster, the department purchase history will be parsed using filter nodes to predict what types of items each cluster will purchase. With the predictive model developed determining which departments the customers purchase from by cluster, that information can then be utilized within marketing campaigns. As an example, if the “low income white females with few kids and long term accounts” tend to purchase kitchen wares, then during a sales marketing campaign or new product introduction, the accounts matching the demographics cluster would be targeted for an email notification or mail out flyers. This would be in conjunction with allowing customers to select their interests on the account setup process already existing.

An example using Cluster 3 (Males) has been generated ([Screenshot 1.14](#t5p9s6wsfqxz)). Without a target variable, the department purchase information was included and the demographics information removed for the new cluster generated. ([Screenshot 1.15](#x47jgse5pt3c)).

1. Low income white females with few kids and long term accounts ([Screenshot 1.6](#xs97ry15rmtu))
   * SEX: Female
   * NUMKIDS: 1 or fewer
   * HOMEVAL: less than 92012
   * INCOME: 17668 to 38118
   * RACE: White
   * AGE: 26.5 to 35.8 months
2. Low income minority homeowners with no college education ([Screenshot 1.7](#yixc6g6nbtmj))
   * RACE: Black or Hispanic
   * NUMCARS: 1 (lower than data set distribution)
   * INCOME: less than 27893
   * EDLEVEL: High School or lower
   * HOMEVAL: less than 71437
3. Men ([Screenshot 1.8](#ek67pr3naai7))
   * SEX: Male
4. Higher Income Earners ([Screenshot 1.9](#p8dd23brrkei))
   * INCOME: 33006 to 61703
   * HOMEVAL: 92012 to 230232
   * EDLEVEL: College or High School
   * JOB: Manufacturing, Clerical, Retail, or Sales
5. Multiple Kids ([Screenshot 1.10](#dktroz55jy1h))
   * NUMKIDS: 2 or more children

**PREDICTIVE MODELS**

After obtaining the clustering results, a predictive model was created to develop rules for assigning customers to one of the five clusters.

*Building the Process Flow Diagram (*[*Screenshot 1.11*](#18nmntp53270)*)*

Output from the Clustering Node was used as input for the predictive model. This was done by using the METADATA Node to set the SEGMENT (i.e. Cluster) id as the target variable for the predictive model.

1. Add the METADATA Node to the diagram. Link the output from the CLUSTER Node as input to the METADATA Node.
2. Configure the METADATA Node
   1. In the node properties panel, under the Train -> Variables -> Train, set \_SEGMENT\_ variable NEW ROLE as Target ([Screenshot 1.12](#ihsf9n8jutf9)).
3. Add the DATA PARTITION Node
   1. Partition the data as below
      1. Training: 65%
      2. Validation: 35%
      3. Test: 0%
4. Add DECISION TREE Node
   1. Configure the Decision Tree Assessment Measure property to be “Misclassification”

*Predictive Model Results*

The Decision Tree model resulted in 13 rules to predict which cluster a customer would belong to. This model had a 7% Validation misclassification rate. The ruleset can be found in the appendix ([Appendix 1](#tn0t1q3j97yw)).

**2. Generating and Assessing Market Basket Analysis using Association Rule Discovery Method:**

*Market Basket Analysis* (also known as *association rule discovery* or *affinity analysis*) is a popular data mining method. In the simplest situation, the data consists of two variables: a *transaction* and an *item*. The aim of the analysis is to determine the strength of all the association rules among a set of items.

**DATA -** In our case, the ASSOCIATIONS dataset contains buying information for about 1000 of BuyLow Grocery store customers.

**AIM** –

* To develop at least 5 “meaningful” rules using the association node in EM and describe each of the rules that are generated and its implication.
* To develop an item taxonomy based upon the generated rules.

**ASSOCIATION ANALYSIS**

*Building the Process Flow Diagram*

* First, you need to add the ASSOCIATION data source to project.
* In the Project Panel, right-click **Data Sources** and click **Create Data Source**.
* In the Data Source Wizard — Metadata Source window, click **Next**.
* In the Data Source Wizard — Select a SAS Table and enter **ASSOCIATIONS** in the **Table** field. Click **Next**.
* In the Data Source Wizard — Table Information window, click **Next**.
* In the Data Source Wizard — Metadata Advisor Options window, click **Basic**. Click **Next**.
* In the Data Source Wizard — Column Metadata window, make the following changes:
* For the variable CUSTOMER, set the **Role** to **ID**.
* For the variable PRODUCT, set the **Role** to **Target**.
* For the variable TRANSACTION, set **level** to **Interval** and set the **Role** to **Rejected**.
* Under the properties panel for ASSOCIATION node, specify **Export Rule by ID** to **Yes** and change the number of rules to keep from **200 to 50.**

The variable TRANSACTION identifies the sequence in which the products were purchased. In this example, all of the products were purchased at the same time, so the order relates only to the order in which they are scanned at the register. When order is considered, association analysis is known as *sequence analysis*. Sequence analysis is demonstrated in the further section. ([Screenshot 2.0](#kix.nvtal157lgiu))

* In the Data Source Wizard — Decision Configuration window, click **Next**.
* In the Data Source Wizard — Create Sample window, click **Next**.
* In the Data Source Wizard — Data Source Attributes window, set the **Role** of the data source to **Transaction**. Click **Next**. ([Screenshot 2.1](#kix.g8pzzog72l0w))
* In the Data Source Wizard — Summary window, click **Finish**.
* In the Project Panel, drag the **ASSOCIATIONS** data source to your diagram workspace. On the **Explore** tab, drag an **Association** node to your diagram workspace. Connect the **ASSOCIATIONS** data source to the **Association** node. ([Screenshot 2.2](#kix.4stn9unqndj2))
* Set to Use as **NO** for transaction variable. ([Screenshot 2.3](#kix.831f7qw0x0vv))

We have assigned ID Role to CUSTOMER variable when we created the data source to perform association discovery because input data set must have a separate observation for each product purchased by each customer.

*Running the Association Node*

In your diagram workspace, right-click the **Association** node and click **Run**. In the Confirmation window, click **Yes**. Click **Results** in the Run Status window. ([Screenshot 2.4](#kix.4stn9unqndj2))

In the Results window, select **View** => **Rules** => **Rules Table** from the main menu.

The Rules Table displays information about each rule that was created. This includes the confidence, support, lift, number of occurrences, and the items in the rule. To explain confidence, support, and lift, consider the rule A => B where A and B each represent one product.

The support percentage for A => B is the percentage of all customers who purchased both A and B. Support is a measure of how frequently the rule occurs in the database.

The confidence percentage for A => B is the percentage of all customers who purchased both A and B, divided by the number of customers who purchased A.

The lift of A => B is a measure of the strength of the association. For example, if the lift is 2 for A => B, then a customer who purchased A is twice as likely as a customer chosen at random to purchase B.

Sort the Rules Table by clicking the **Support (%)** column heading. Notice that the top four rules are

* **Ice cream & chicken ==> soda & chocolate**
* **soda & chocolate ==> ice cream & chicken**
* **ice cream & chocolate => soda & chicken**
* **soda & chicken => ice cream & chocolate**

All with the support of 11.59%. ([Screenshot 2.5](#kix.831f7qw0x0vv))

The confidence for **ice cream & chicken ==> soda & chocolate** is 82.86%, which indicates that 82.86% of customers who purchased **ice cream** **& chicken** then purchased **soda & chocolate.**

For the rule **soda & chocolate ==> ice cream & chicken** the confidence is 78.91%. This means that 78.91% of the customers who purchased **soda & chocolate** then purchased **ice cream & chicken.**

For the rule **ice cream & chocolate ==> soda & chicken** the confidence is 76.82%. This means that 78.91% of the customers who purchased **ice cream & chocolate** then purchased **soda & chicken.**

For the rule **soda & chicken ==> ice cream & chocolate** the confidence is 83.45%. This means that 78.91% of the customers who purchased **soda & chicken** then purchased **ice cream & chocolate.**

Having a look at the second highest support (%). We noticed that there are two rules with support % of 10.39%

* **Turkey & Deli ==> Olives & Ham**
* **Olives & Ham ==> Turkey & Deli**

The confidence for **Turkey & Deli ==> Olives & Ham** is 77.61%, which indicates that 77.61% of customers who purchased **Turkey & Deli** then purchased **Olives & Ham**

For the rule **Olives & Ham ==> Turkey & Deli** the confidence is 65.82%. This means that 65.82% of the customers who purchased **Olives & Ham** then purchased **Turkey & Deli**

Lift, in the context of association rules, is the ratio of the confidence of a rule to the expected confidence of the rule. The expected confidence is calculated under the assumption that the left-hand side of a rule is independent from the right-hand side of the rule. Consequently, lift is a measure of association between the left-hand side and right-hand side of the rule. Values that are greater than one represents positive association between the left and right-hand sides. Values that are equal to one represent independence. Values that are less than one represents negative association between the left and right-hand sides.

Sort the Rules Table by **Lift**. ([Screenshot 2.6](#kix.otze78yoqope))

Notice that two rules

* **peppers & fruit ==> chocolate & apples**
* **chocolate & apples ==> peppers & fruit**

has the greatest lift of 5.67.

This indicates that customers who buy peppers & fruit are 5.67 times more likely to buy chocolate & apples than a customer chosen at random. Similarly, customers who buy chocolate & apples are 5.67 times more likely to buy peppers & fruit than a customer chosen at random.

Having a look at the second highest Lift, we noticed that two rules

* **Ice cream & chicken** **==> soda and chocolate**
* **soda & chocolate ==> ice cream & chicken**

has the lift of 5.64.

This indicates that customers who buy ice cream & chicken are 5.64 times more likely to buy soda & chocolate than a customer chosen at random. Similarly, customers who buy soda & chocolate are 5.67 times more likely to buy ice cream & chicken than a customer chosen at random.

Based upon our analysis, we have the following rules

* **Ice cream & chicken ==> soda & chocolate**
* **soda & chocolate ==> ice cream & chicken**
* **ice cream & chocolate => soda & chicken**
* **soda & chicken => ice cream & chocolate**
* **Turkey & Deli ==> Olives & Ham**
* **Olives & Ham ==> Turkey & Deli**
* **peppers & fruit ==> chocolate & apples**
* **chocolate & apples ==> peppers & fruit**
* **Ice cream & chicken ==> soda and chocolate**
* **soda & chocolate ==> ice cream & chicken**

As per the rules we have, it is advised to the BuyLow grocery store that

* They should put ice cream, chicken, soda, and chocolate together in one area in the store as people are more likely to buy soda and chocolate if they buy ice cream and chicken.
* The store can even raise the price of one (say olives and ham) and lower it on the other (say turkey and deli).
* To save some money and do better budgeting, the store can decide not to advertise the products falling under an association rule together. This will help them in saving money and do effective advertising as they know people tend to buy some items based upon the items they have already added in their basket.

All these measures can help the store to increase its sales and profit and also make better inventory holding decisions.

***Sequence Analysis***

Sequence variables in the data set enables you to conduct a sequence analysis.

* Drag the **ASSOCIATION** node, rename it as **SEQUENCE** and connect it with the **ASSOCIATION** node. ([Screenshot 2.7](#eduf9318mvcc))
* Under the properties panel, specify support type as count and **export rule ID** to **Yes** and rule numbers to keep to **50 from 200.**
* Change the **sort criterion** to **Count.**
* Right click sequence node and edit the use of the **transaction variable** to **YES.** ([Screenshot 2.8](#d5q38hu1jpnw))
* Now, Update and Run it.
* Results ([Screenshot 2.9](#3584m9ucn1qf))

In the Results window, select **View** => **Rules** => **Rules Table** from the main menu.

Sort the Rules Table by clicking the **Support (%)** column heading. The rule with highest support

* **Chips ⇒ Beer,** with the support of 45.45%, with transaction count 455 and Lift as 1.18. ([Screenshot 2.10](#kix.831f7qw0x0vv))
* **Of those customers who purchase chips, of them 93.24% will purchase a beer.**

Sort the Rules Table by clicking the **Lift** column heading. The rule with highest lift

* **Ice cream ⇒ Soda,** with the lift of 2.38, with transaction count 220 and support as 21.98%. ([Screenshot 2.11](#kix.831f7qw0x0vv))
* **Of those customers who purchase ice cream, of them 70.29% will purchase a soda.**

***Taxonomy***

**Beverages**

* Alcoholic
* Liquor
* Beer
* Non-alcoholic
* Soda
* Diet Soda

**Food**

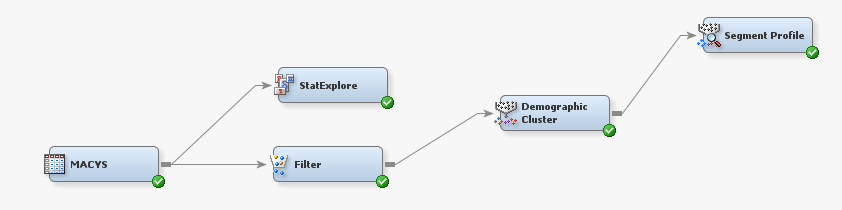
* Meats
* Chicken
* Deli
* Fish
* Ham
* Steak
* Turkey
* Vegetable
* Peppers
* Fruit
* Tomatoes
* Olives
* Apples
* Others

Ice cream

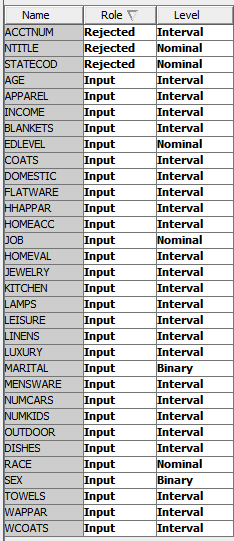
* Chocolate
* Baguette
* Chips

**Appendix**

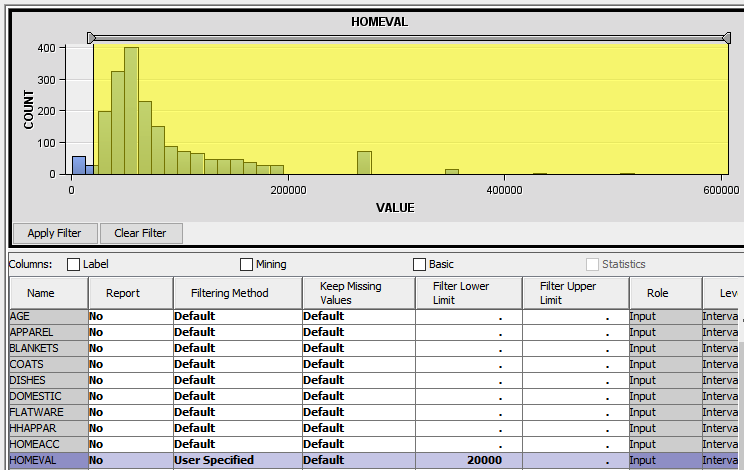
Screenshot 1.0



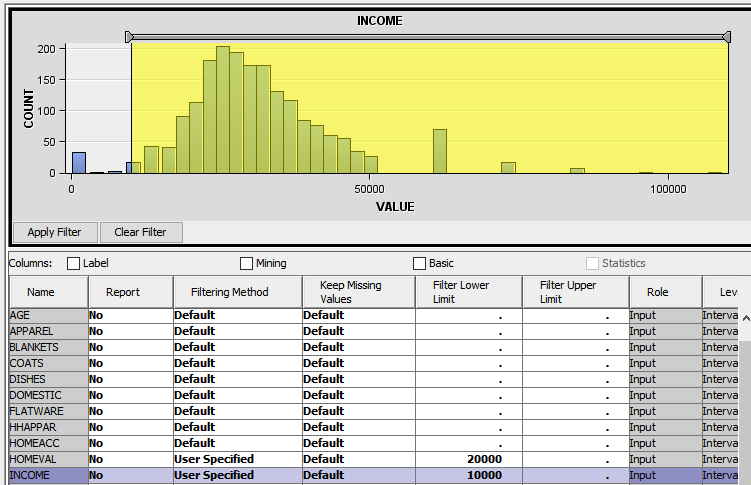
Screenshot 1.1



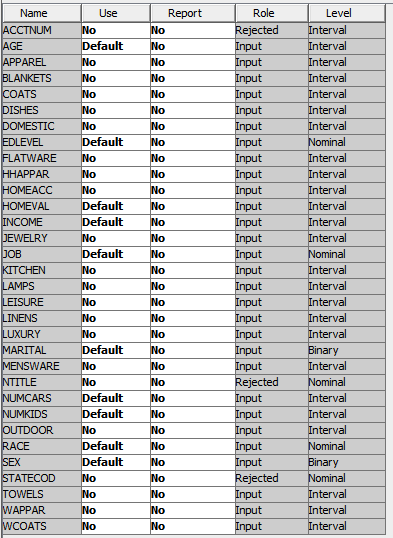
Screenshot 1.2



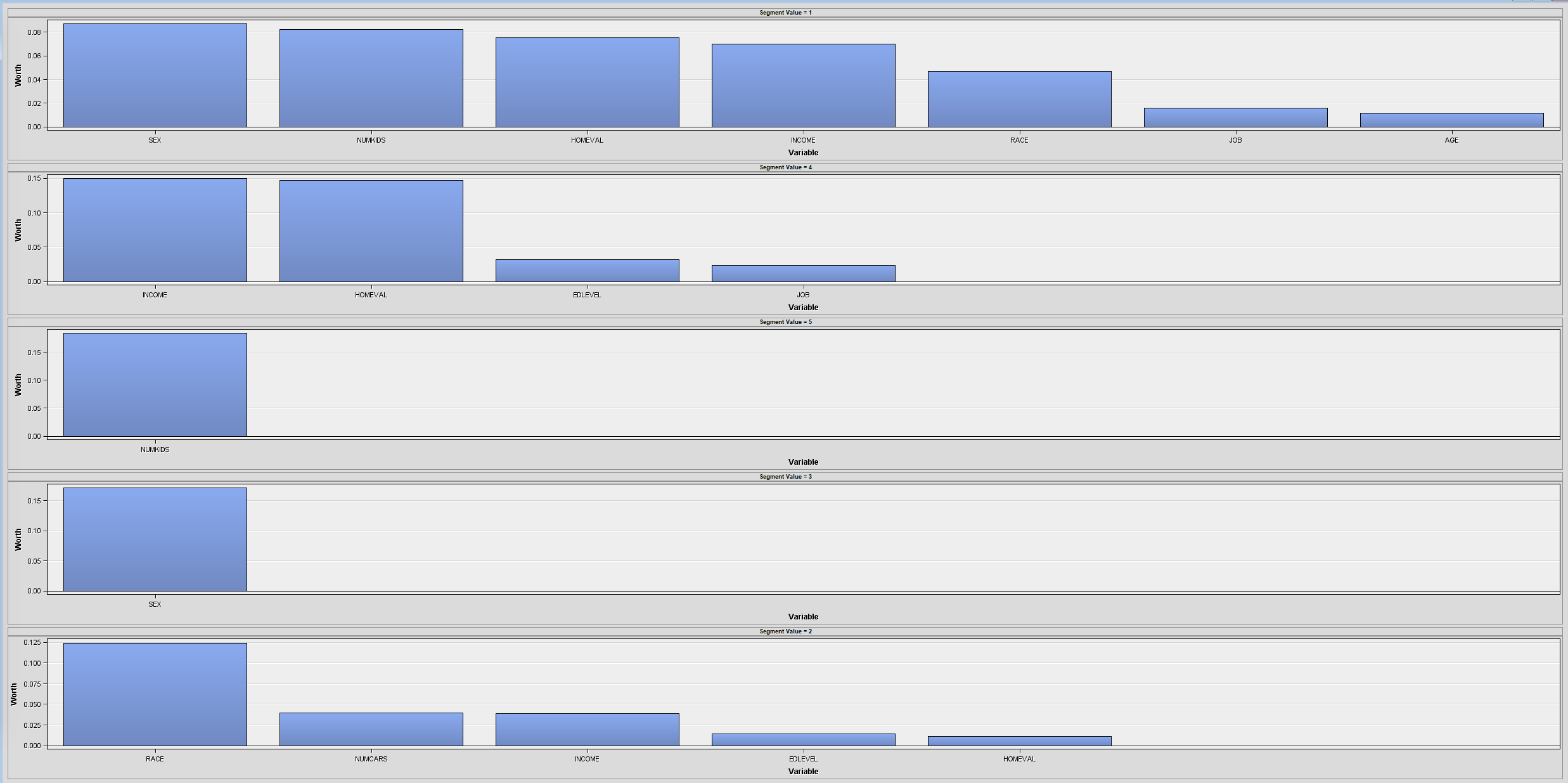
Screenshot 1.3



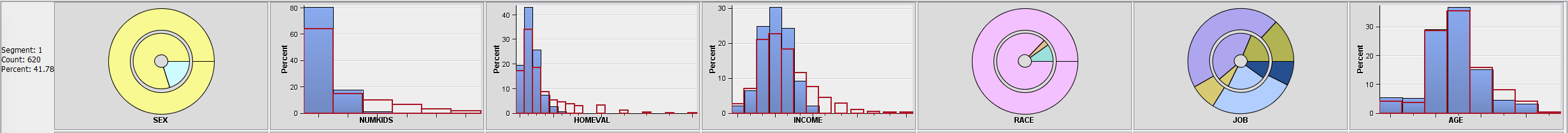
Screenshot 1.4



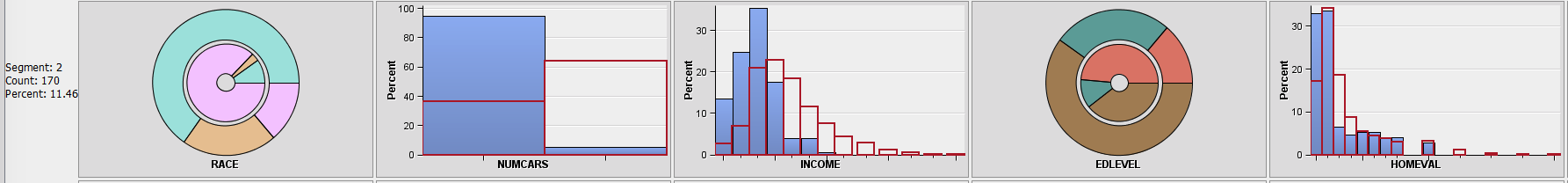
Screenshot 1.5



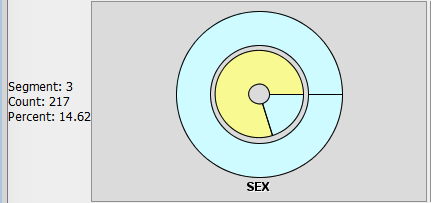
Screenshot 1.6



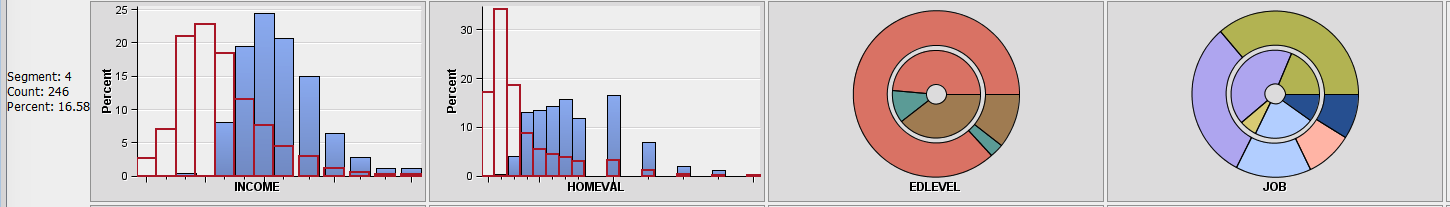
Screenshot 1.7



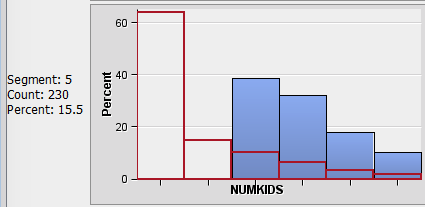
Screenshot 1.8



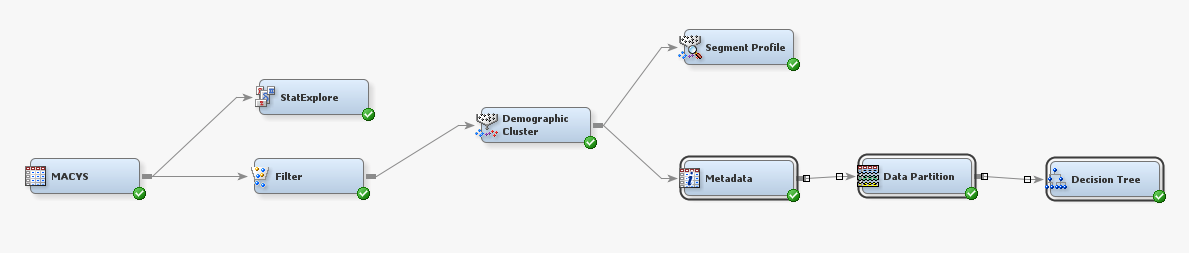
Screenshot 1.9



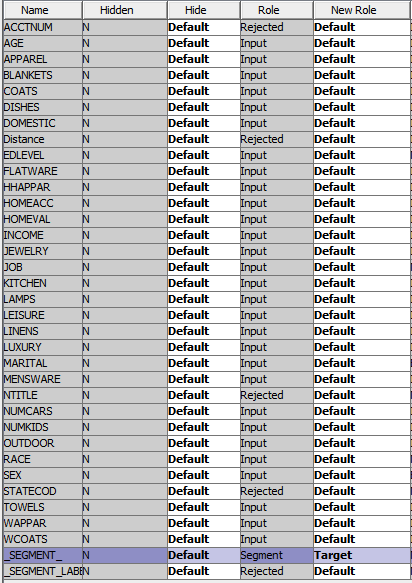
Screenshot 1.10



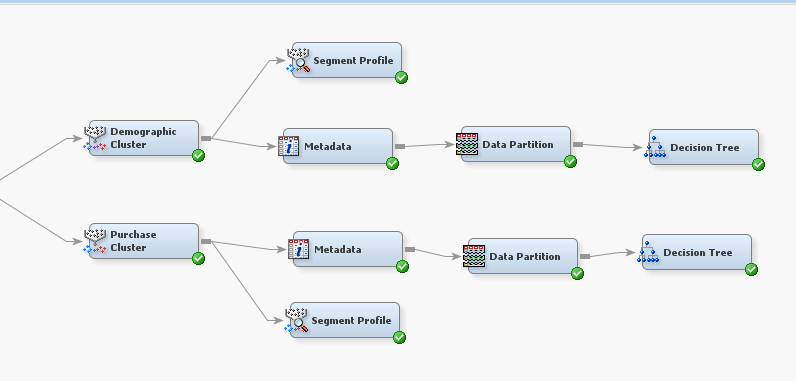
Screenshot 1.11



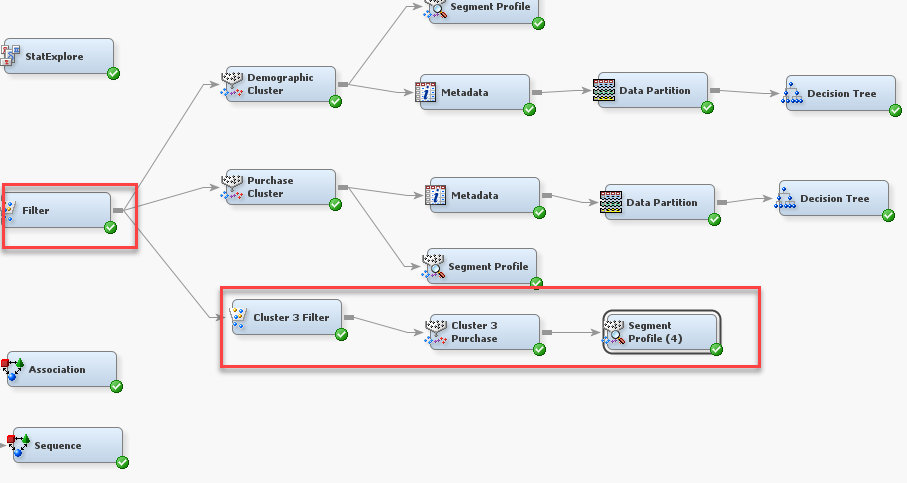
Screenshot 1.12



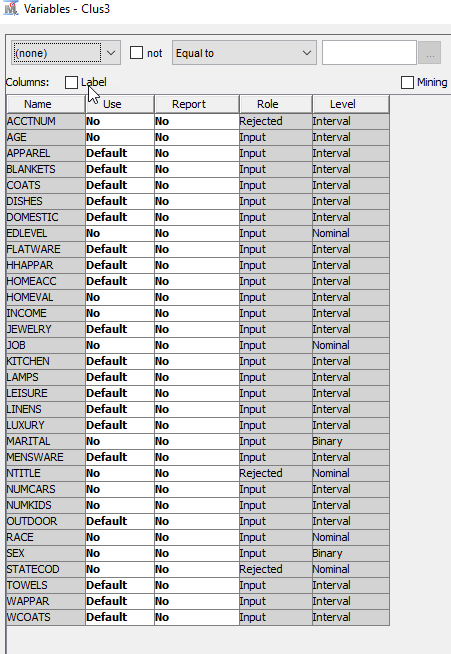
Screenshot 1.13



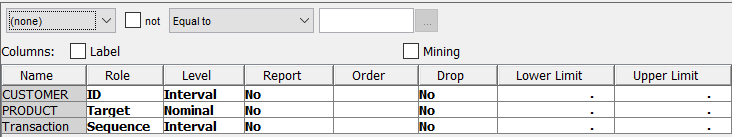
Screenshot 1.14



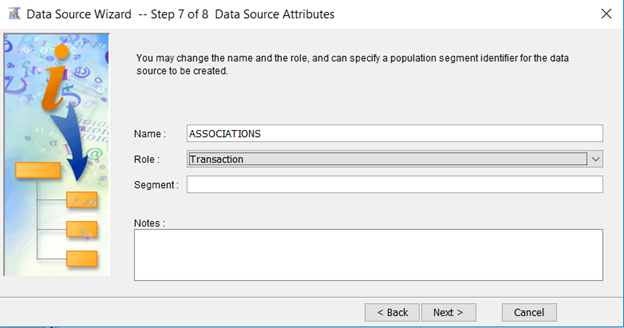
Screenshot 1.15



Screenshot 2.0

1

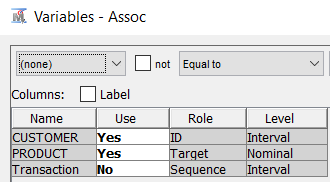
Screenshot 2.1



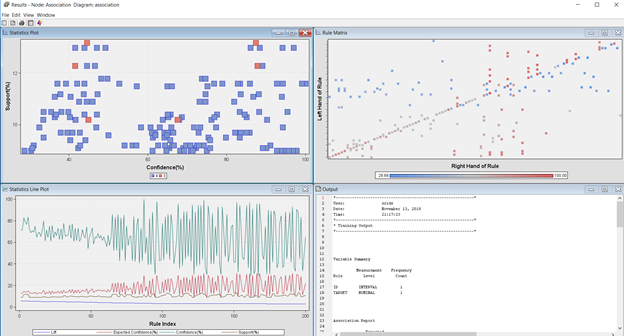
Screenshot 2.2



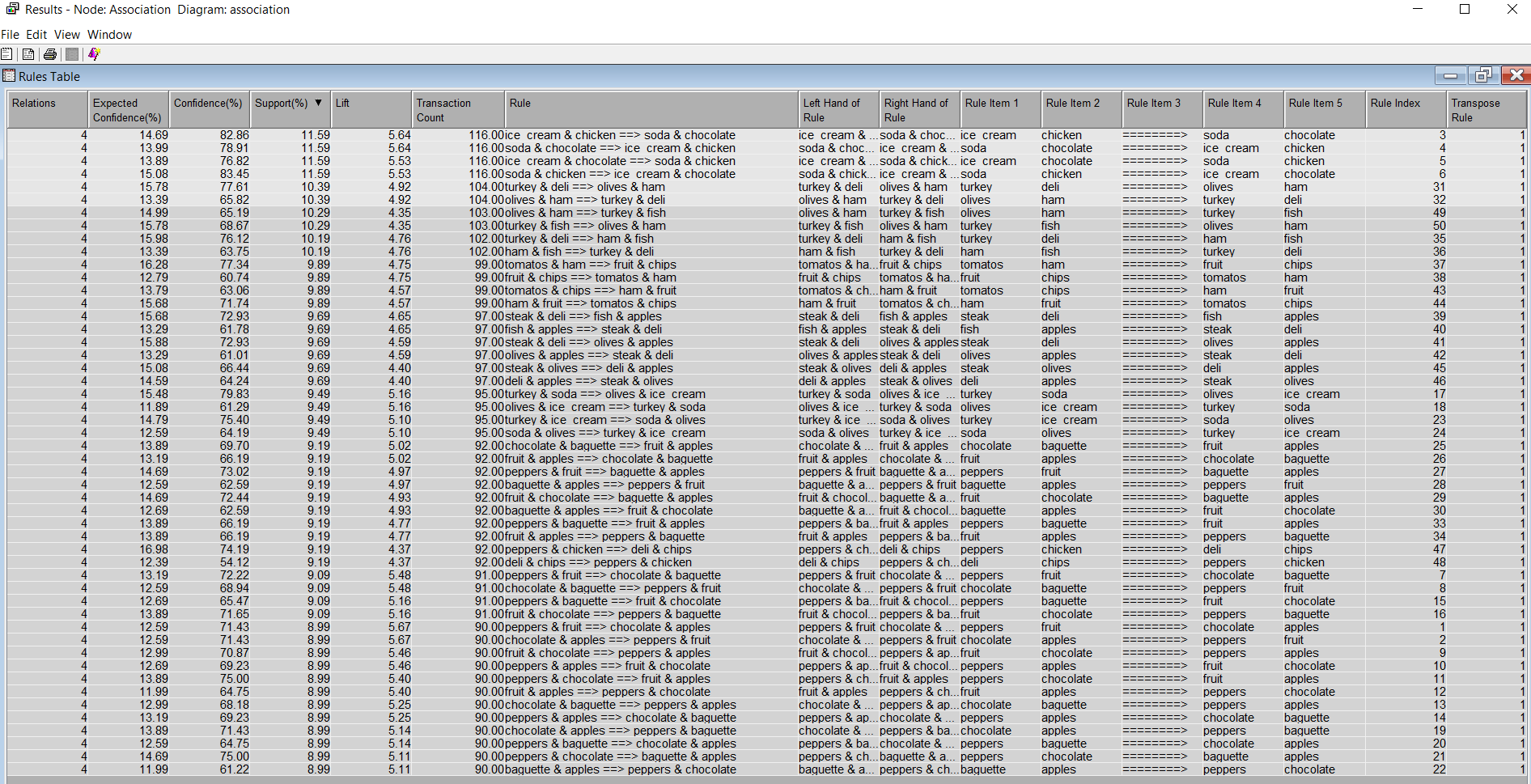
Screenshot 2.3



Screenshot2.4



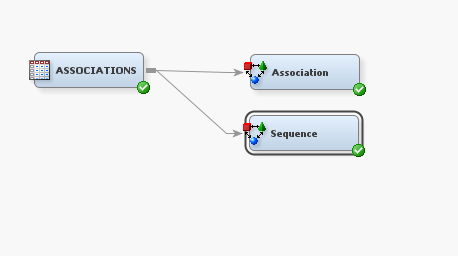
Screenshot 2.5



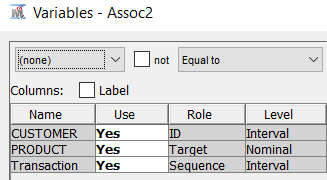
Screenshot 2.6



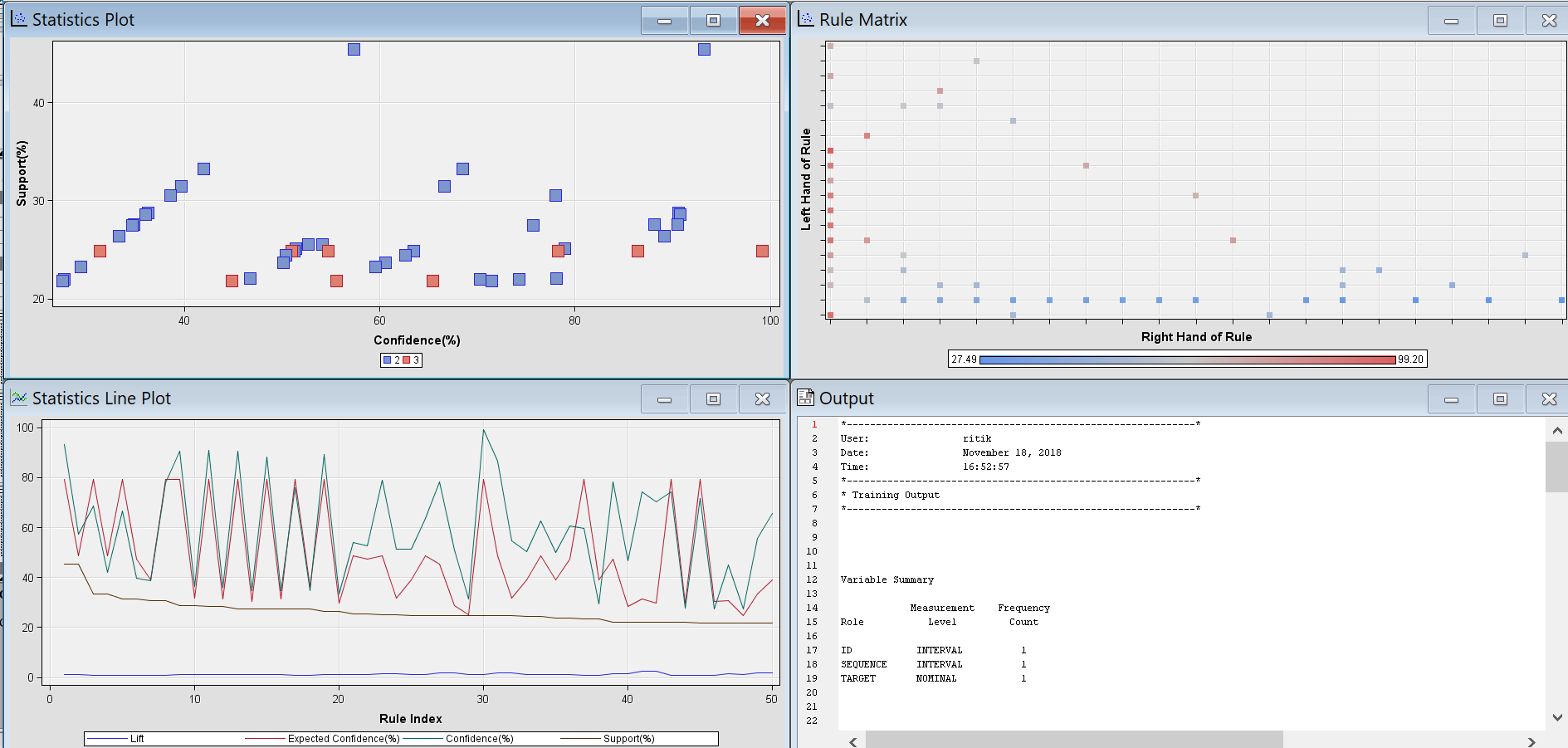
Screenshot 2.7



Screenshot 2.8



Screenshot 2.9



Screenshot 2.10



Screenshot 2.11



APPENDIX 1

\*------------------------------------------------------------\*

Node = 5

\*------------------------------------------------------------\*

if SEX IS ONE OF: MALE

AND NUMKIDS >= 3.5

then

Tree Node Identifier = 5

Number of Observations = 6

Predicted: \_SEGMENT\_=5 = 0.83

Predicted: \_SEGMENT\_=4 = 0.00

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.17

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 9

\*------------------------------------------------------------\*

if SEX IS ONE OF: MALE

AND NUMKIDS < 3.5 or MISSING

AND INCOME >= 49700

then

Tree Node Identifier = 9

Number of Observations = 6

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 1.00

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 13

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND NUMKIDS >= 1.5

AND HOMEVAL >= 205350

then

Tree Node Identifier = 13

Number of Observations = 9

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 1.00

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 15

\*------------------------------------------------------------\*

if SEX IS ONE OF: MALE

AND NUMKIDS < 3.5 or MISSING

AND INCOME < 49700 or MISSING

AND HOMEVAL >= 163250

then

Tree Node Identifier = 15

Number of Observations = 9

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.44

Predicted: \_SEGMENT\_=3 = 0.11

Predicted: \_SEGMENT\_=2 = 0.44

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 18

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: BLACK, HISPANIC

AND NUMKIDS < 1.5 or MISSING

AND INCOME < 41100 or MISSING

then

Tree Node Identifier = 18

Number of Observations = 68

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.04

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.90

Predicted: \_SEGMENT\_=1 = 0.06

\*------------------------------------------------------------\*

Node = 19

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: BLACK, HISPANIC

AND NUMKIDS < 1.5 or MISSING

AND INCOME >= 41100

then

Tree Node Identifier = 19

Number of Observations = 9

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 1.00

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 20

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND NUMKIDS >= 1.5

AND INCOME < 51600 or MISSING

AND HOMEVAL < 205350 or MISSING

then

Tree Node Identifier = 20

Number of Observations = 162

Predicted: \_SEGMENT\_=5 = 0.89

Predicted: \_SEGMENT\_=4 = 0.01

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.06

Predicted: \_SEGMENT\_=1 = 0.05

\*------------------------------------------------------------\*

Node = 21

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND NUMKIDS >= 1.5

AND INCOME >= 51600

AND HOMEVAL < 205350 or MISSING

then

Tree Node Identifier = 21

Number of Observations = 6

Predicted: \_SEGMENT\_=5 = 0.17

Predicted: \_SEGMENT\_=4 = 0.83

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 22

\*------------------------------------------------------------\*

if SEX IS ONE OF: MALE

AND RACE IS ONE OF: BLACK, HISPANIC

AND NUMKIDS < 3.5 or MISSING

AND INCOME < 49700 or MISSING

AND HOMEVAL < 163250 or MISSING

then

Tree Node Identifier = 22

Number of Observations = 21

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.00

Predicted: \_SEGMENT\_=3 = 0.14

Predicted: \_SEGMENT\_=2 = 0.86

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 23

\*------------------------------------------------------------\*

if SEX IS ONE OF: MALE

AND RACE IS ONE OF: WHITE or MISSING

AND NUMKIDS < 3.5 or MISSING

AND INCOME < 49700 or MISSING

AND HOMEVAL < 163250 or MISSING

then

Tree Node Identifier = 23

Number of Observations = 152

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.00

Predicted: \_SEGMENT\_=3 = 0.89

Predicted: \_SEGMENT\_=2 = 0.11

Predicted: \_SEGMENT\_=1 = 0.00

\*------------------------------------------------------------\*

Node = 24

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: WHITE or MISSING

AND NUMKIDS < 1.5 or MISSING

AND INCOME < 38300 or MISSING

AND HOMEVAL < 127550 or MISSING

then

Tree Node Identifier = 24

Number of Observations = 380

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.01

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.99

\*------------------------------------------------------------\*

Node = 25

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: WHITE or MISSING

AND NUMKIDS < 1.5 or MISSING

AND INCOME < 38300 or MISSING

AND HOMEVAL >= 127550

then

Tree Node Identifier = 25

Number of Observations = 17

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.82

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.18

\*------------------------------------------------------------\*

Node = 26

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: WHITE or MISSING

AND NUMKIDS < 1.5 or MISSING

AND INCOME >= 38300

AND HOMEVAL < 77850

then

Tree Node Identifier = 26

Number of Observations = 16

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.25

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.75

\*------------------------------------------------------------\*

Node = 27

\*------------------------------------------------------------\*

if SEX IS ONE OF: FEMALE or MISSING

AND RACE IS ONE OF: WHITE or MISSING

AND NUMKIDS < 1.5 or MISSING

AND INCOME >= 38300

AND HOMEVAL >= 77850 or MISSING

then

Tree Node Identifier = 27

Number of Observations = 101

Predicted: \_SEGMENT\_=5 = 0.00

Predicted: \_SEGMENT\_=4 = 0.99

Predicted: \_SEGMENT\_=3 = 0.00

Predicted: \_SEGMENT\_=2 = 0.00

Predicted: \_SEGMENT\_=1 = 0.01