

Banking Segmentation Case Study

Case Study Description

A consumer bank sought to segment its customers based on historic usage patterns. Segmentation was to be used for improving contact strategies in the Marketing Department.

A sample of 100,000 active consumer customers was selected. An *active consumer customer* was defined as an individual or household with at least one checking account and at least one transaction on the account during a three-month study period. All transactions during the three-month study period were recorded and classified into one of four activity categories:

- traditional banking methods (TBM)
- automatic teller machine (ATM)
- point of sale (POS)
- customer service (CSC)

A three-month activity profile for each customer was developed by combining historic activity averages with observed activity during the study period. Historically, for one CSC transaction, an average customer would conduct two POS transactions, three ATM transactions, and 10 TBM transactions. Each customer was assigned this initial profile at the beginning of the study period. The initial profile was updated by adding the total number of transactions in each activity category over the entire three-month study period.

The **PROFILE** data set contains all 100,000 three-month activity profiles. This case study describes the creation of customer activity segments based on the **PROFILE** data set.

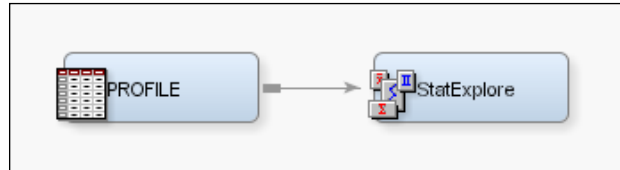
(You can find the **PROFILE** dataset in [Metadata Repository](#)→[Shared Data](#)→[Libraries](#)→[AAEM](#))

Case Study Data

Name	Model Role	Measurement Level	Description
ID	ID	Nominal	Customer ID
CNT_TBM	Input	Interval	Traditional bank method transaction count
CNT_ATM	Input	Interval	ATM transaction count
CNT_POS	Input	Interval	Point-of-sale transaction count
CNT_CSC	Input	Interval	Customer service transaction count
CNT_TOT	Input	Interval	Total transaction count

Accessing and Assaying the Data

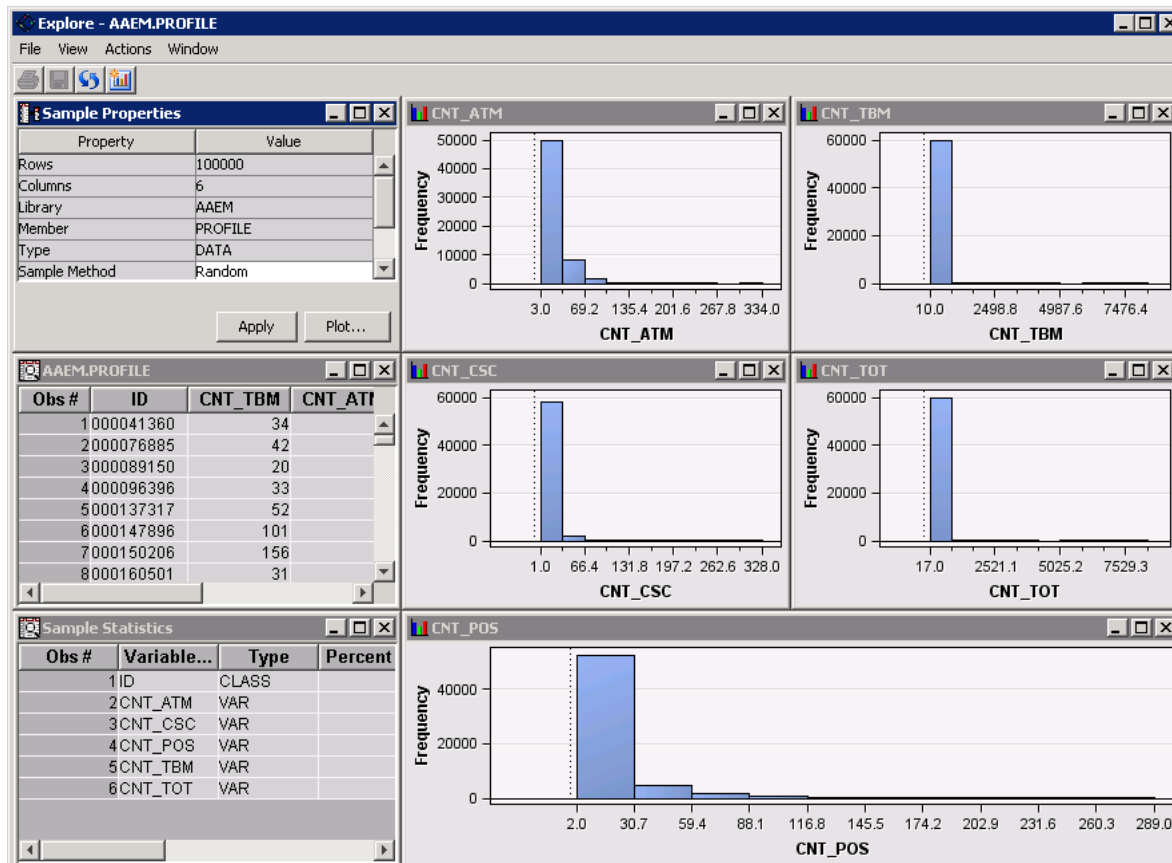
A SAS Enterprise Miner data source was defined using the metadata settings indicated above. The StatExplore node was used to provide preliminary statistics on the input variables. (Click on the Explore tab, you will see the StatExplore tool, which is the 3rd to the last/right one)



The Interval Variable Summary from the StatExplore node showed no missing values but did show a surprisingly large range on the transaction counts.

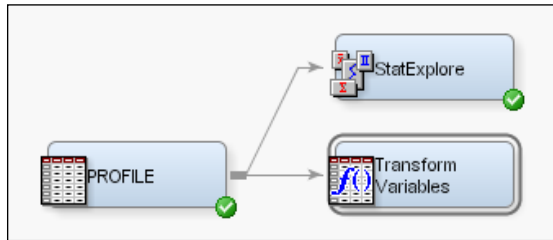
32	Interval Variable Summary Statistics									
33	(maximum 500 observations printed)									
34										
35	Data Role=TRAIN									
36										
37				Standard	Non					
38	Variable	Role	Mean	Deviation	Missing	Missing	Minimum	Median	Maximum	Skewness
39										
40	CNT_ATM	INPUT	19.49971	20.8561	100000	0	3	13	628	2.357293
41	CNT_CSC	INPUT	6.68411	12.12856	100000	0	1	2	607	6.236494
42	CNT_POS	INPUT	11.9233	20.73384	100000	0	2	2	345	3.343805
43	CNT_TBM	INPUT	68.13696	101.1542	100000	0	10	52	14934	53.05219
44	CNT_TOT	INPUT	106.2441	113.3704	100000	0	17	89	15225	39.2061
45										

A plot of the input distributions showed highly skewed distributions for all inputs.



It would be difficult to develop meaningful segments from such highly skewed inputs. Instead of focusing on the transaction counts, it was decided to develop segments based on the relative proportions of transactions across the four categories. This required a transformation of the raw data.

A Transform Variables node was connected to the **PROFILE** node. (Click on the Modify tab, you will see the Transform Variable tool, which is the right most one)

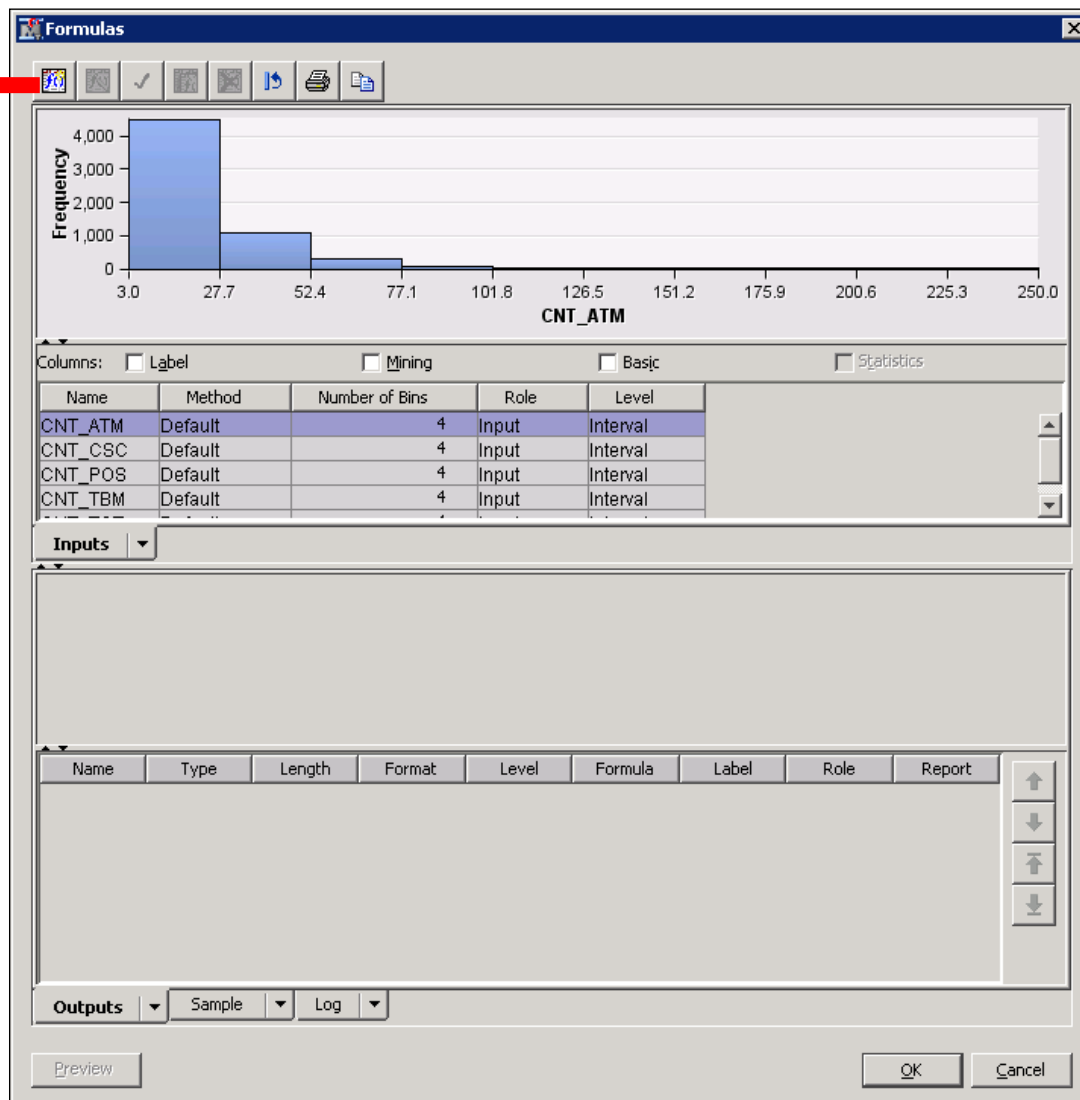


The Transform Variables node was used to create *category logit scores* for each transaction category.

$$\text{category logit score} = \log(\text{transaction count}_{\text{in category}} / \text{transaction count}_{\text{out of category}})$$

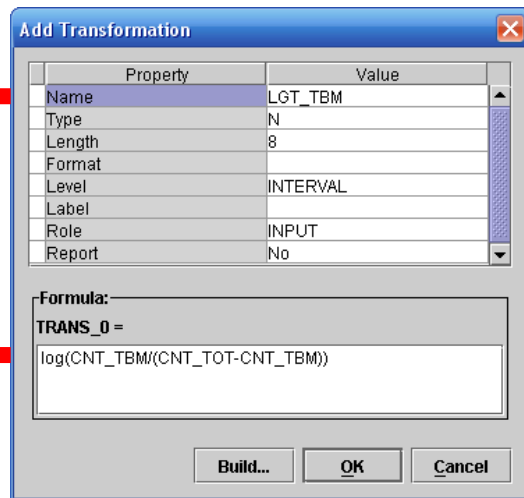
The transformations were created using these steps:

1. Select **Formulas** in the Transform Variable node's Properties panel. The Formulas window appears.



2. Select the Create icon as indicated above.

The Add Transformation dialog box appears.



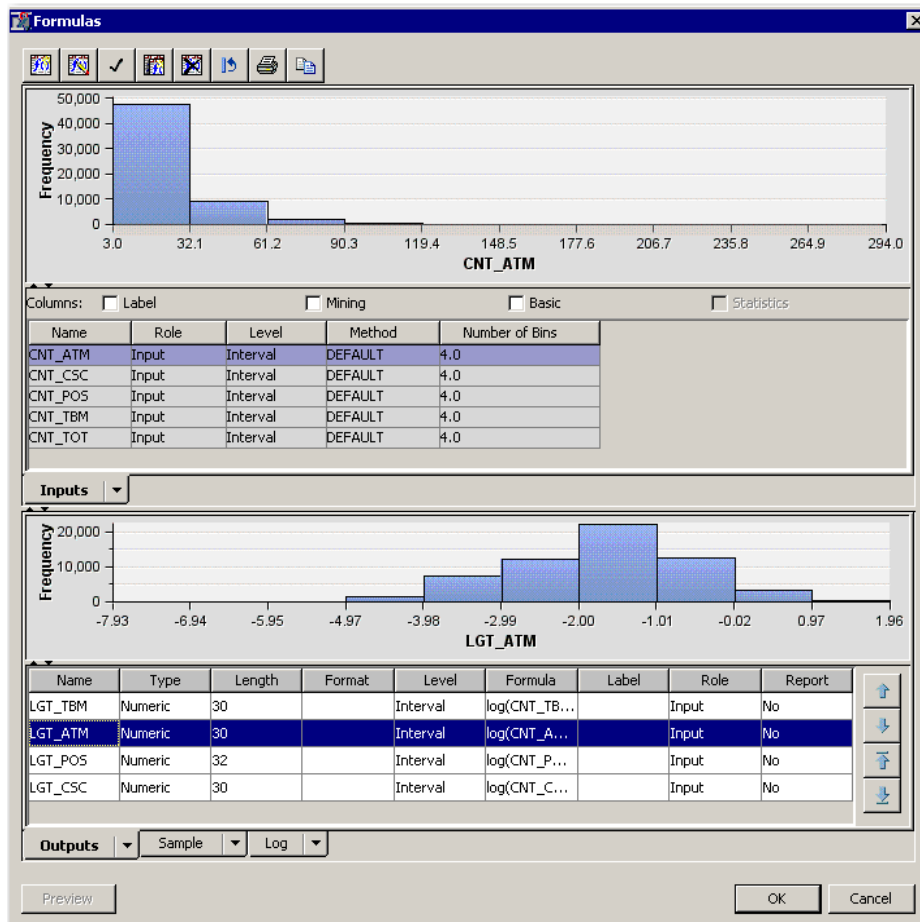
The Add Transformation dialog box is shown. It has a table with properties and values, and a formula field.

Property	Value
Name	LGT_TBM
Type	N
Length	8
Format	
Level	INTERVAL
Label	
Role	INPUT
Report	No

Formula:
TRANS_0 =
 $\log(\text{CNT_TBM}/(\text{CNT_TOT}-\text{CNT_TBM}))$

Buttons: Build..., OK, Cancel


3. For each transaction category, type the name and formula as indicated.
4. Select **OK** to add the transformation. The Add Transformation dialog box closes and you return to the Formula Builder window.
5. Select **Preview** to see the distribution of the newly created input.

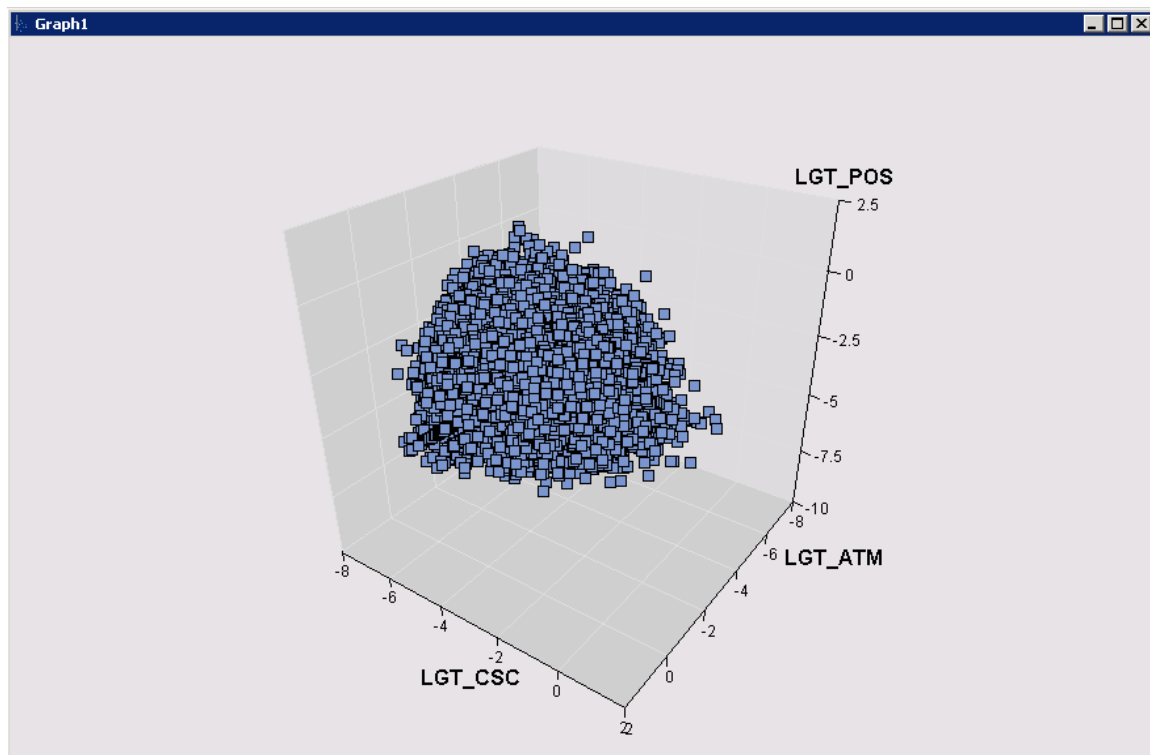


6. Repeat Steps 1-5 for the other three transaction categories.
7. Select **OK** to close the Formula Builder window.
8. Run the Transform Variables node.

Segmentation was to be based on the newly created category logit scores. Before proceeding, it was deemed reasonable to examine the joint distribution of the cases using these derived inputs. A scatter plot using any three of the four derived inputs would represent the joint distribution without significant loss of information.

A three-dimensional scatter plot was produced using the following steps:

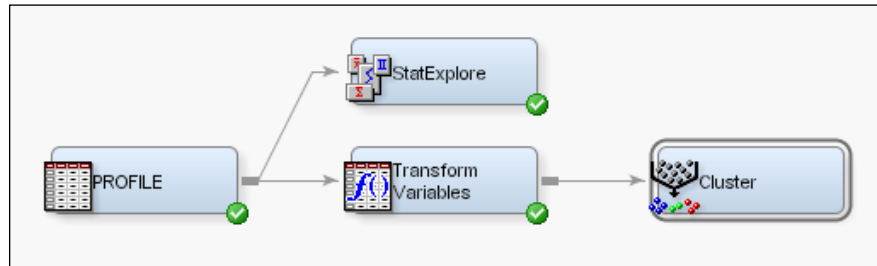
1. Select **Exported Data** from the Properties panel of the Transform Variables node. The Exported Data window appears.
2. Select the **TRAIN** data and select **Explore**. The Explore window appears.
3. Select **Actions** ⇒ **Plot...** or click  (the Plot Wizard icon). The Plot Wizard appears.
4. Select a three-dimensional scatter plot.
5. Select **Role** ⇒ **X, Y, and Z** for **LGT_ATM**, **LGT_CSC**, and **LGT_POS**, respectively.
6. Select **Finish** to generate the scatter plot.



The scatter plot showed a single clump of cases, making this analysis a segmentation (rather than a clustering) of the customers. There were a few outlying cases with apparently low proportions on the three plotted inputs. Given that the proportions in the four original categories must sum to 1, it followed that these outlying cases must have a high proportion of transactions in the non-plotted category, TBM.

Creating Segments

Transactions segments were created using the Cluster node.



Two changes to the Cluster node default properties were made, as indicated below. Both were related to limiting the number of clusters created to 5.

Train	
Variables	...
Cluster Variable Role	Segment
Internal Standardization	None
<input checked="" type="checkbox"/> Number of Clusters	
Specification Method	User Specify
Maximum Number of C5	

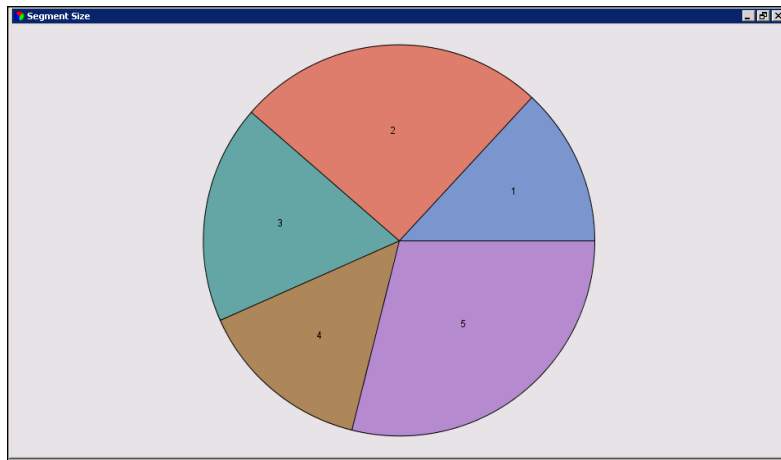


Because the inputs were all on the same measurement scale (category logit score), it was decided to **not** standardize the inputs.

Only the four LGT inputs defined in the Transform Variables node were set to **Default** in the Cluster node.

Variables - Clus				
(none)	<input type="checkbox"/> not	Equal to		...
Columns:	<input type="checkbox"/> Label	<input type="checkbox"/> Mining	<input type="checkbox"/> Basic	<input type="checkbox"/> Statistics
Name	Use	Report	Role	Level
CNT_ATM	No	No	Input	Interval
CNT_CSC	No	No	Input	Interval
CNT_POS	No	No	Input	Interval
CNT_TBM	No	No	Input	Interval
CNT_TOT	No	No	Input	Interval
ID	Yes	No	ID	Nominal
LGT_ATM	Default	No	Input	Interval
LGT_CSC	Default	No	Input	Interval
LGT_POS	Default	No	Input	Interval
LGT_TBM	Default	No	Input	Interval

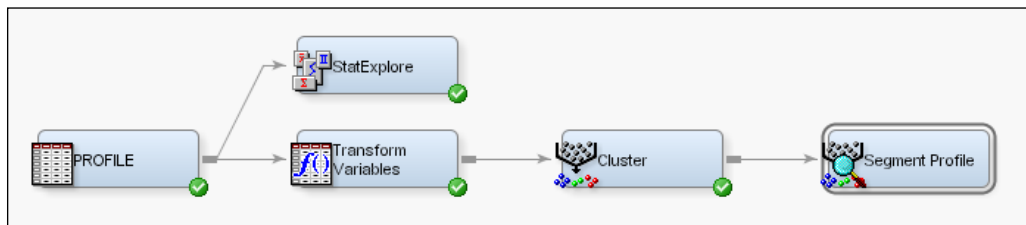
Running the Cluster node and viewing the Results window confirmed the creation of five nearly equally sized clusters.



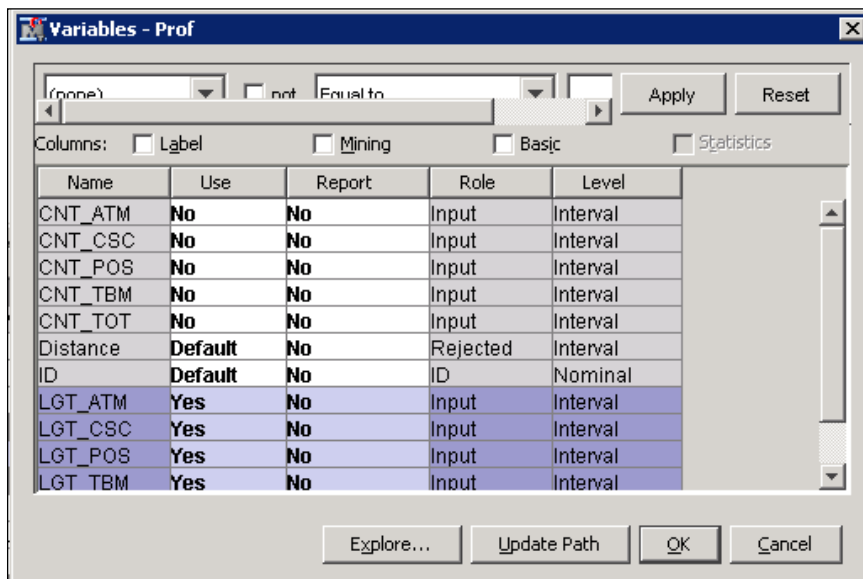
Additional cluster interpretations were made with the Segment Profile tool.

Interpreting Segments

A Segment Profile node attached to the Cluster node helped to interpret the contents of the generated segments.

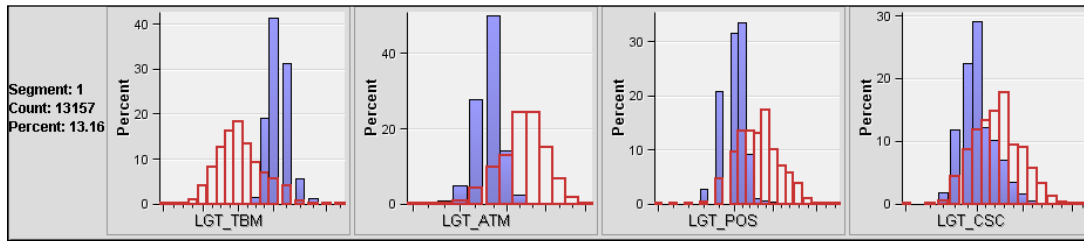


Only the **LGT** inputs were set to **Yes** in the Segment Profile node.

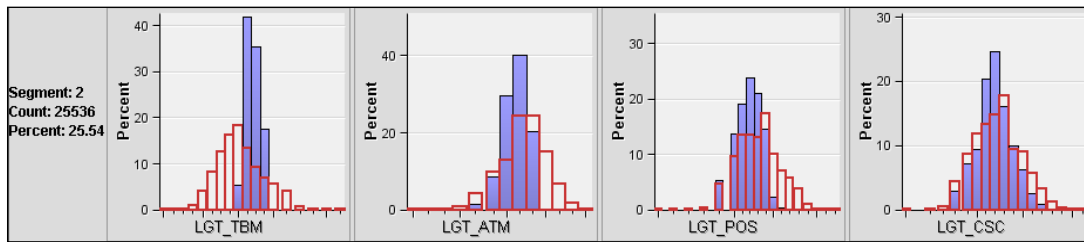


Name	Use	Report	Role	Level
CNT_ATM	No	No	Input	Interval
CNT_CSC	No	No	Input	Interval
CNT_POS	No	No	Input	Interval
CNT_TBM	No	No	Input	Interval
CNT_TOT	No	No	Input	Interval
Distance	Default	No	Rejected	Interval
ID	Default	No	ID	Nominal
LGT_ATM	Yes	No	Input	Interval
LGT_CSC	Yes	No	Input	Interval
LGT_POS	Yes	No	Input	Interval
LGT_TBM	Yes	No	Input	Interval

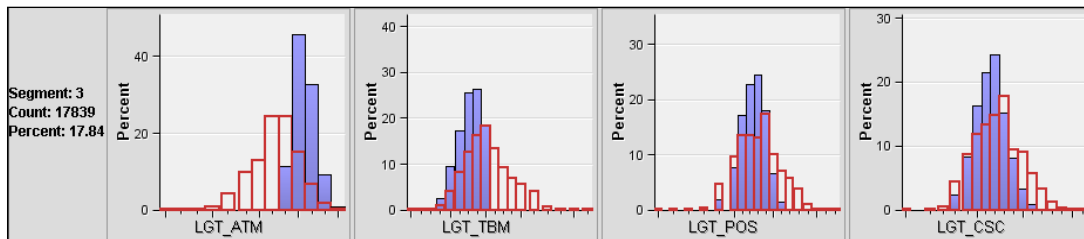
The following profiles were created for the generated segments:



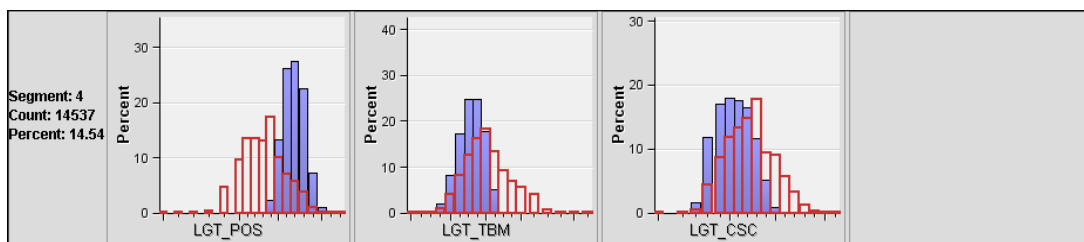
Segment 1 customers had a significantly higher than average use of traditional banking methods and lower than average use of all other transaction categories. This segment was labeled **Brick-and-Mortar**.



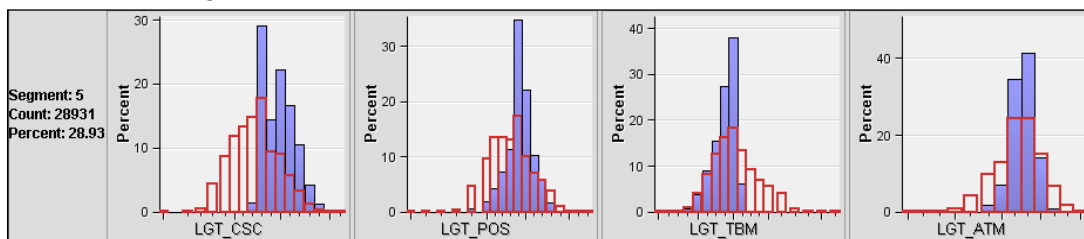
Segment 2 customers had a higher than average use of traditional banking methods but were close to the distribution centers on the other transaction categories. This segment was labeled **Transitional** because they seem to be transitioning from brick-and-mortar to other usage patterns.



Segment 3 customers eschewed traditional banking methods in favor of ATMs. This segment was labeled **ATMs**.



Segment 4 was characterized by a high prevalence of point-of-sale transactions and few traditional bank methods. This segment was labeled **Cashless**.



Segment 5 had a higher than average rate of customer service contacts and point-of-sale transactions. This segment was labeled **Service**.