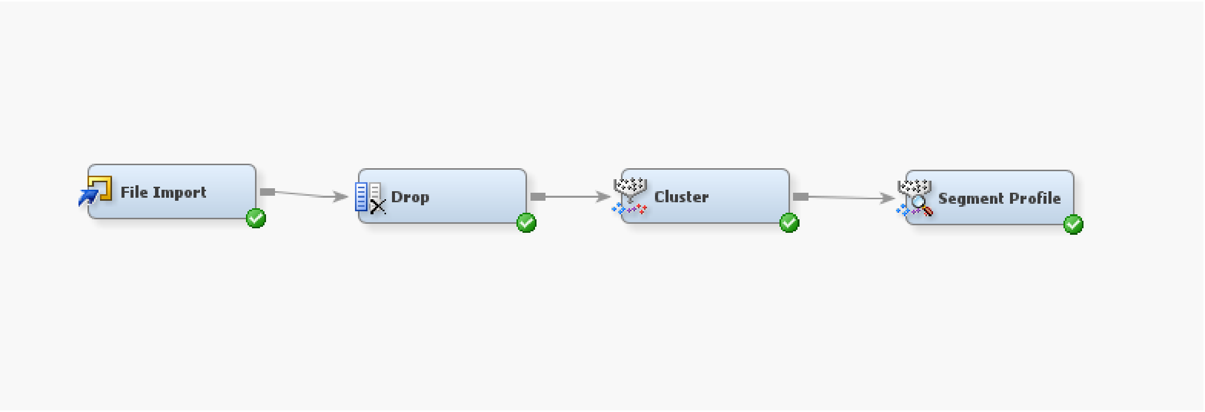
Objective 1:Identify the most profitable customer segment.

Purpose:

Identify high-value customer segment based on past purchase history. High-value customers represent the most profitable group to receive targeted marketing attention for their current and future value to the business. This will allow us to focus marketing efforts on the most profitable segments. The key advantage of focusing on high-value segments is marketing efficiency - effort and resources can be concentrated on the subset of customers that provide the largest profits.

Overview of the EM diagram



Methodology- Cluster analysis

The raw dataset contains a lot of irrelevant variables. Then, we need to drop some irrelevant variables like **Days_to_Ship Actually, Ship_Status, Category, Country, Order_Date, Order_ID, City, Postal_Code,** etc.. **Customer_name** and **Sales_per_Customer** are the input variables for clustering.

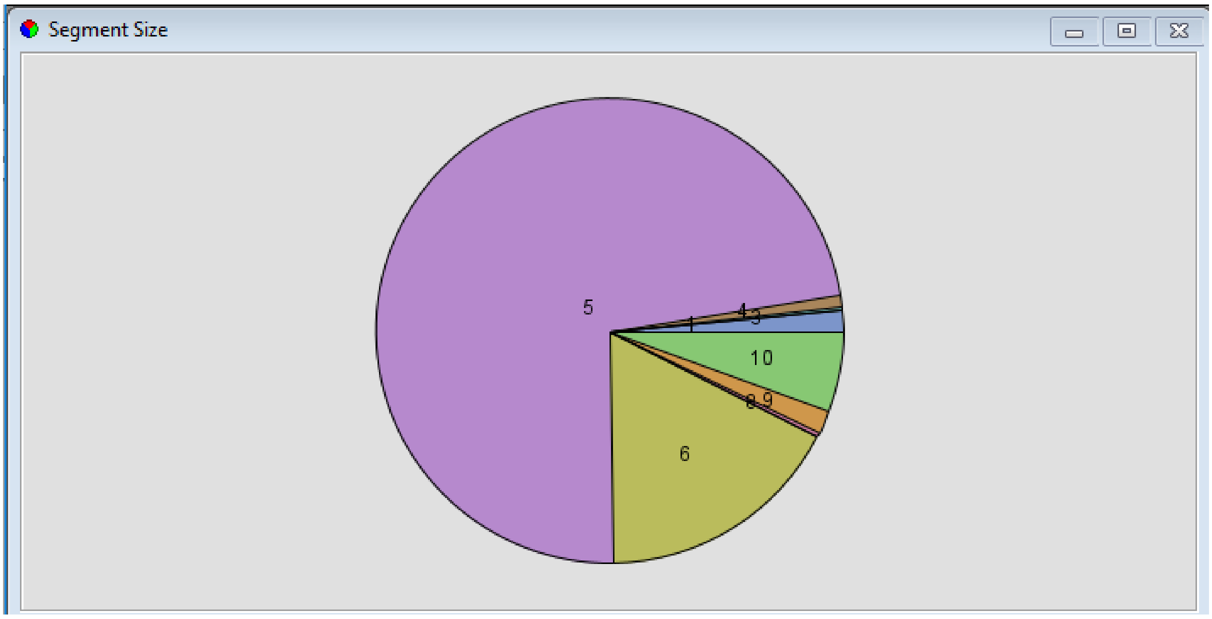
Input Variables:

Customer_name and **Sales_per_Customer.**

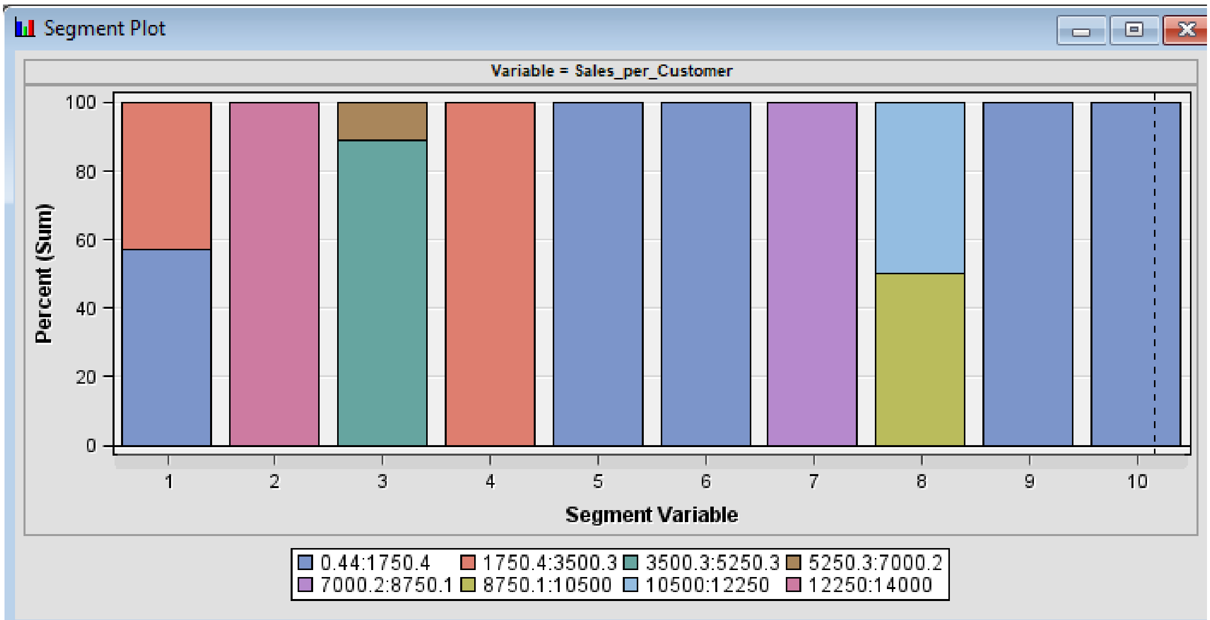
Name	Drop	Role	Level
Category	Yes	Input	Nominal
City	Yes	Input	Nominal
Country	Yes	Input	Nominal
Customer_Name	No	Input	Nominal
Days_to_Ship_A	Yes	Input	Interval
Days_to_Ship_S	Yes	Input	Interval
Discount	Yes	Input	Interval
Number_of_Rec	Yes	Input	Interval
Order_Date	Yes	Input	Nominal
Order_ID	Yes	Input	Nominal
Postal_Code	Yes	Input	Interval
Product_Name	Yes	Text	Nominal
Profit	Yes	Input	Interval
Profit_Ratio	Yes	Input	Interval
Profit_per_Orde	Yes	Input	Interval
Quantity	Yes	Input	Interval
Region	Yes	Input	Nominal
Sales	Yes	Input	Interval
Sales_Forecast	Yes	Input	Interval
Sales_per_Custo	No	Input	Interval
Satisfaction	Yes	Input	Interval
Segment	Yes	Segment	Nominal
Ship_Date	Yes	Input	Nominal
Ship_Mode	Yes	Input	Nominal
Ship_Status	Yes	Input	Nominal
State	Yes	Input	Nominal
Sub_Category	Yes	Input	Nominal

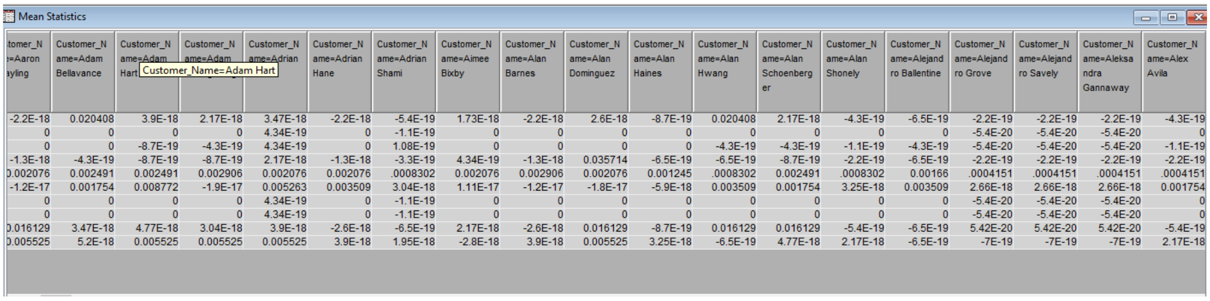
Findings

Results of cluster analysis

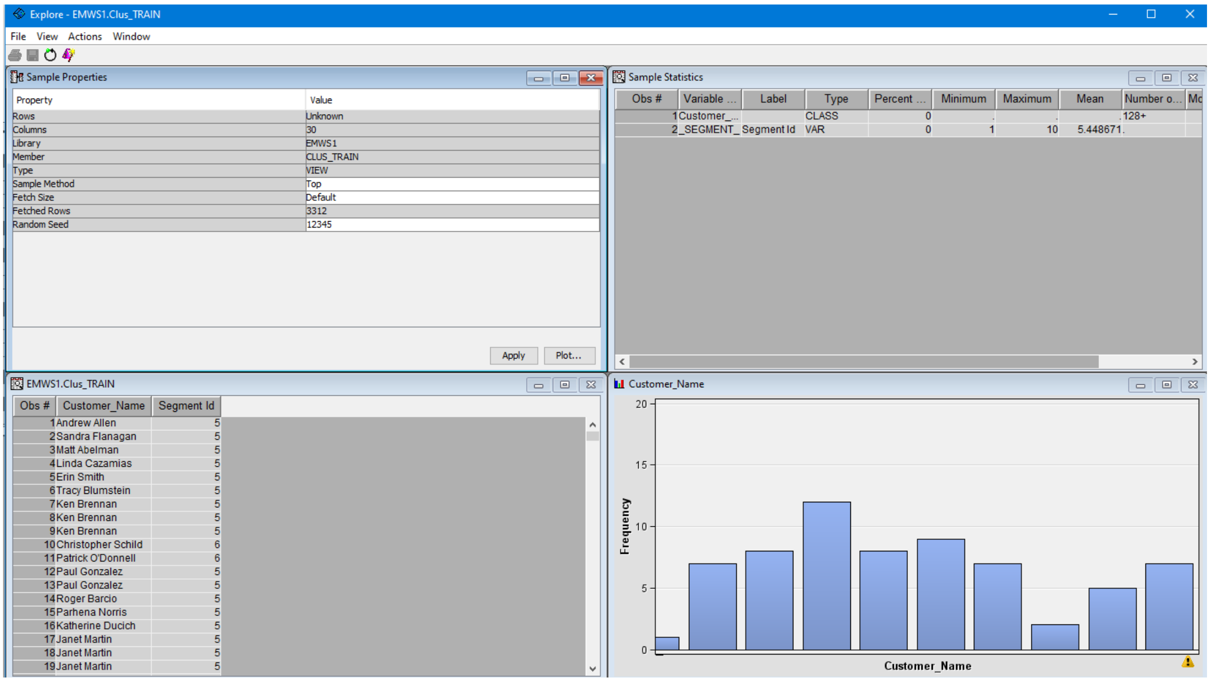


There are 10 clusters representing 10 different groups of spending.

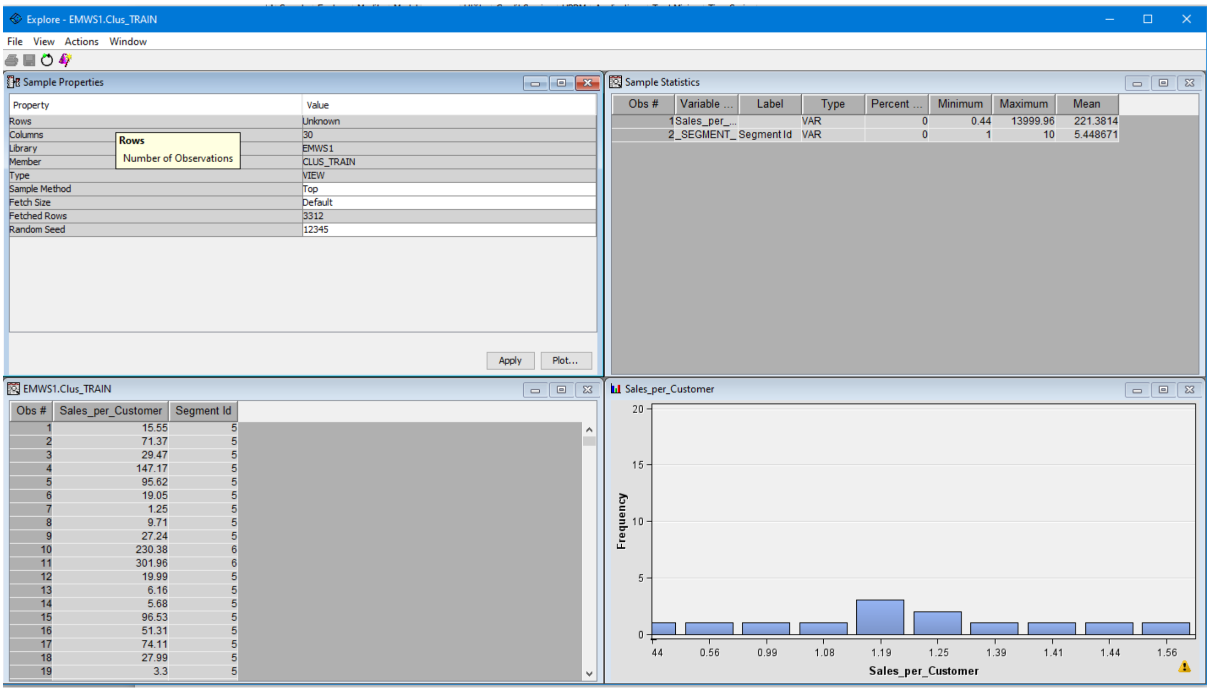




Segment profiling of Customers:



Segment profiling of sales per customers:

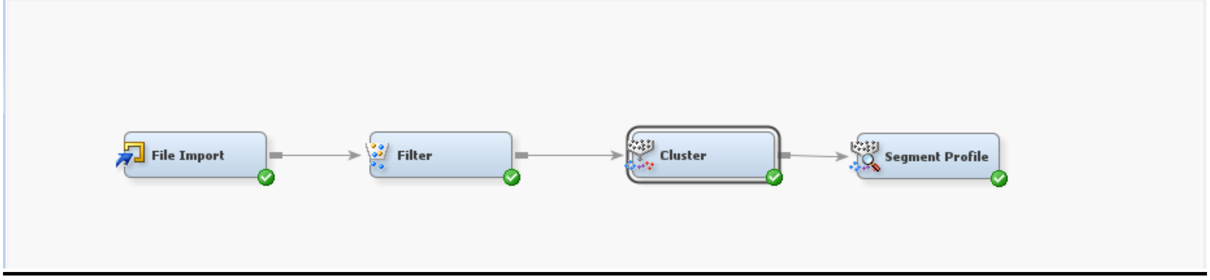


conclusion:

focus on marketing the highest percentage of the group(group 5). 72.7% , rest are non-profitable customers.

Objective 2: Identify the seasonal product.
For developing the tailored-marketing strategy, companies should familiarize with the top-sellers in various time-slots, i.e. seasons. Firstly, we should identify different types of products sold in the given time period. Therefore, we can spot the best seller in different seasons. Observations with similar characteristics would be grouped into the same segment by using cluster analysis.

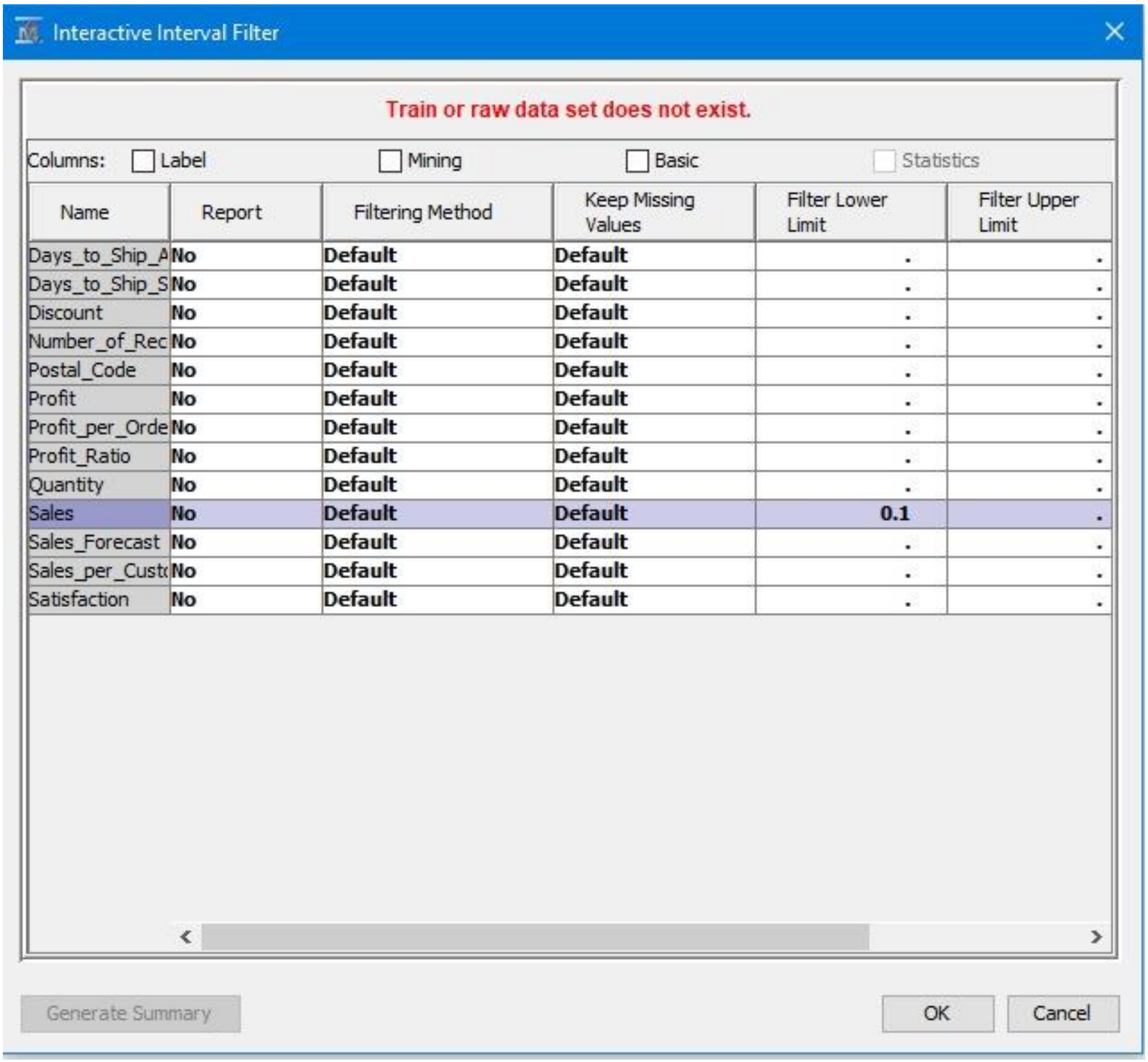
Overview of the EM diagram



2.1 Methodology

Firstly, we do data cleansing in order to eliminate the missing values. We set the user-specific for class variable in the filter: Category, City, Country, Custer name, Order date, order ID, region, segment, ship date, ship mode, ship status, state, sub category. And then, we also set the user-specific interval variables: day to ship actual, day to ship scheduled, discount, number of records, postal code, profit, profit ratio, profit per order, quantity, sales, sales forecast, sales per customer. Also, we set the lower limit for Sales is 0.1 as companies will not generate zero sales in each sales record.

Name	Use	Report	Role	Level
City	No	No	Input	Nominal
Country	No	No	Input	Nominal
Customer_Name	No	No	Input	Nominal
Days_to_Ship_A	No	No	Input	Interval
Days_to_Ship_S	No	No	Input	Interval
Discount	No	No	Input	Interval
Number_of_Rec	No	No	Input	Interval
Order_Date	Default	No	Input	Nominal
Order_ID	No	No	Input	Nominal
Postal_Code	No	No	Input	Interval
Profit	Default	No	Input	Interval
Profit_Ratio	No	No	Input	Interval
Profit_per_Orde	No	No	Input	Interval
Quantity	Default	No	Input	Interval
Region	No	No	Input	Nominal
Sales	Default	No	Input	Interval
Sales_Forecast	No	No	Input	Interval
Sales_per_Custo	No	No	Input	Interval
Segment	Default	No	Input	Nominal
Ship_Date	No	No	Input	Nominal
Ship_Mode	No	No	Input	Nominal
Ship_Status	No	No	Input	Nominal
State	No	No	Input	Nominal
Sub_Category	Default	No	Input	Nominal



Interactive Interval Filter

Train or raw data set does not exist.

Columns: Label Mining Basic Statistics

Name	Report	Filtering Method	Keep Missing Values	Filter Lower Limit	Filter Upper Limit
Days_to_Ship_A	No	Default	Default	.	.
Days_to_Ship_S	No	Default	Default	.	.
Discount	No	Default	Default	.	.
Number_of_Rec	No	Default	Default	.	.
Postal_Code	No	Default	Default	.	.
Profit	No	Default	Default	.	.
Profit_per_Orde	No	Default	Default	.	.
Profit_Ratio	No	Default	Default	.	.
Quantity	No	Default	Default	.	.
Sales	No	Default	Default	0.1	.
Sales_Forecast	No	Default	Default	.	.
Sales_per_Custo	No	Default	Default	.	.
Satisfaction	No	Default	Default	.	.

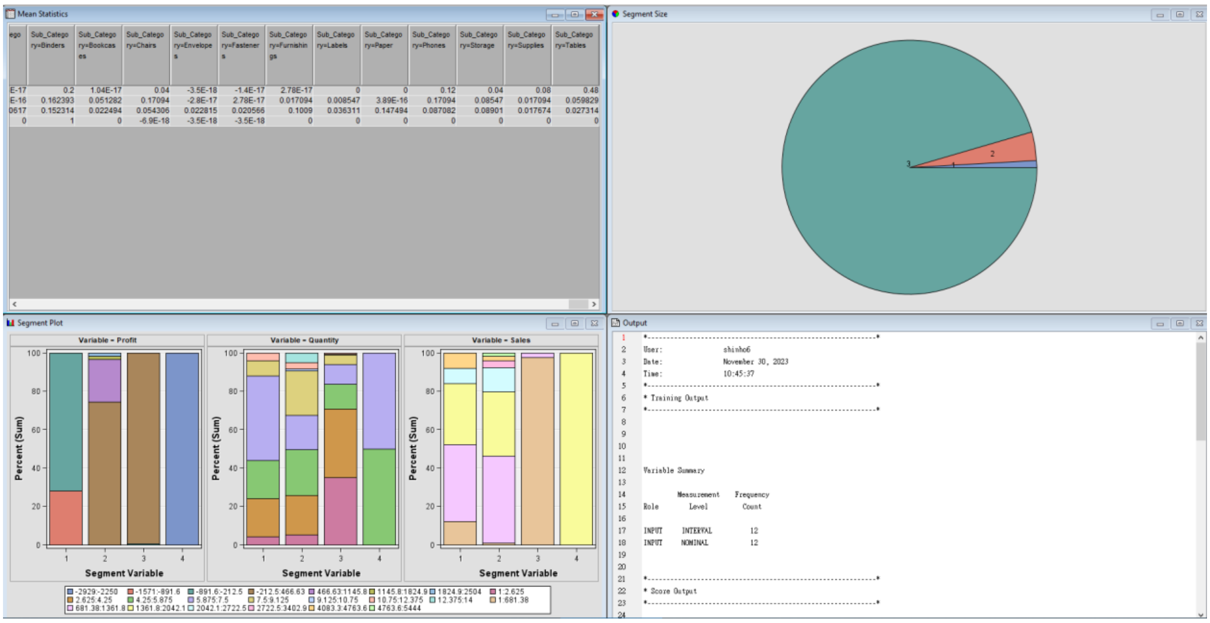
Generate Summary OK Cancel

Input Variables:

Order Date, Profit,Quantity, Sales, Sub Category, Segment

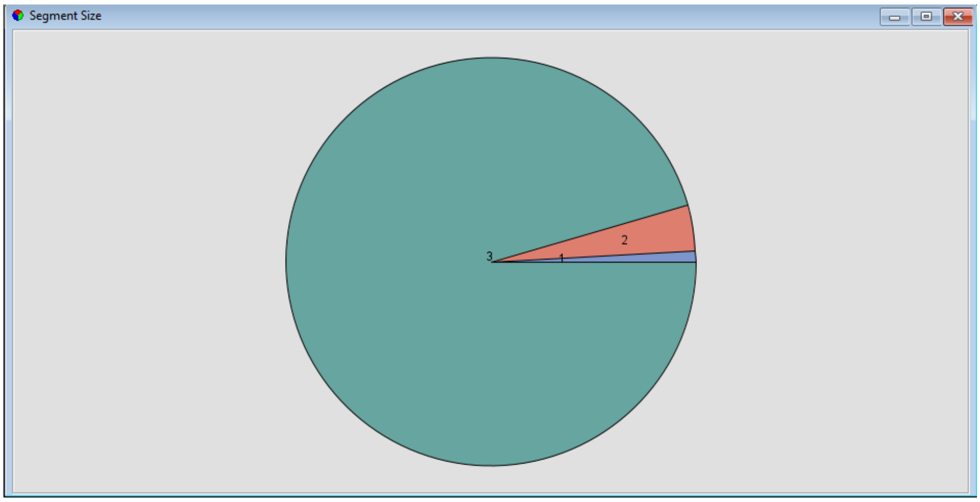
2.2 Findings

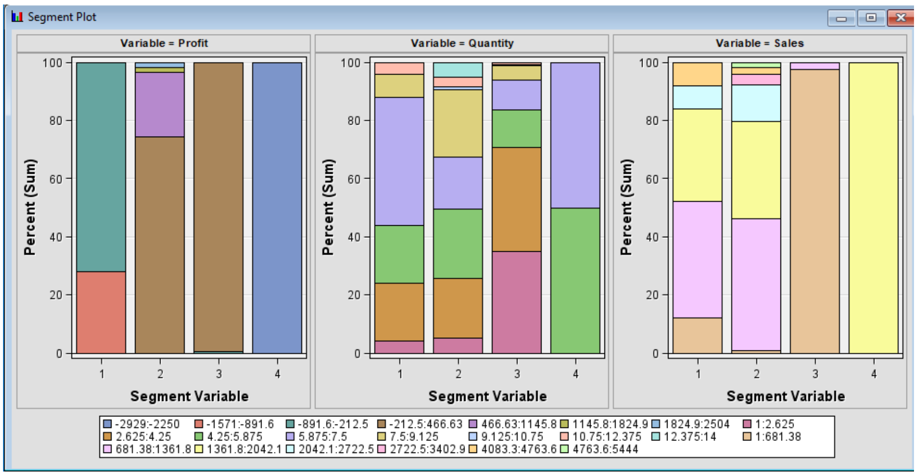
Results of cluster analysis



After data cleansing, we investigate the result by the order date. There are 4 clusters representing 4 different seasons. Hence, we will focus on analyzing these four groups.

<u>Cluster 1</u>





Output

```
*------------------------------------------------------------*
User:                shinho6
Date:                November 30, 2023
Time:                10:45:37
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary

         Measurement    Frequency
Role        Level         Count

INPUT     INTERVAL         12
INPUT     NOMINAL          12


*------------------------------------------------------------*
* Score Output
*------------------------------------------------------------*


*------------------------------------------------------------*
* Report Output
*------------------------------------------------------------*




Variable Importance

                             Number of     Number of
Variable                     Splitting     Surrogate
  Name         Label           Rules         Rules       Importance

Profit       Profit              4             2           1.00000
Sales        Sales               2             0           0.86643
Quantity     Quantity            0             1           0.13764

```

Segment Id	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Profit	Quantity	Sales	Segment=Consumer	Segment=Corporate	Segment=Home Office
1	117	0.811...	17.40858	2	4.832...	401.4786	6.358974	1654.256	0.418803	0.410256	0.17094
2	31...	0.392...	4.975241	1	4.832...	14.84277	3.65627	125.6309	0.509646	0.290354	0.2
3	27	0.858...	8.814878	2	5.622...	-616.556	5.740741	1450.481	0.518519	0.296296	0.185185
4	2	0.755...	2.447889	3	13.94...	-2608.5	5.5	1707.5	0.5	0.5	0

Segment Id	Sub_Category=Accessories	Sub_Category=Appliances	Sub_Category=Art	Sub_Category=Binders	Sub_Category=Bookcases	Sub_Category=Chairs	Sub_Category=Envelopes	Sub_Category=Fasteners	Sub_Category=Furnishings	Sub_Category=Labels	Sub_Category=Paper	Sub_Category=Phones	Sub_Category=Storage	Sub_Category=Supplies	Sub_Category=Tables
1	0.153846	0.102564	1.8E-16	0.162393	0.051282	0.17094	-2.8E-17	2.78E-17	0.017094	0.008547	3.89E-16	0.17094	0.08547	0.017094	0.059829
2	0.082637	0.048553	0.090675	0.15209	0.022508	0.054341	0.02283	0.020579	0.100643	0.036334	0.147588	0.087138	0.089068	0.017685	0.027331
3	-1.4E-17	0.037037	4.16E-17	0.222222	6.94E-18	0.037037	0	-1.4E-17	0.037037	-6.9E-18	0	0.111111	0.037037	0.074074	0.444444
4	0	6.94E-18	0	1	0	-6.9E-18	-3.5E-18	-3.5E-18	0	0	0	0	0	0	0

Segment 1 generated $1654.256 total sales. In this segment consumers and corporate buy most as they shared the biggest proportion in this segment. The category of accessories, binders, chairs and phones shared the highest frequency from the statistical table. Therefore, the products in accessories, binders, chairs and phones group is the best-sellers in this season. This season is heavily impacted by the festivals, like Christmas, New Years Holiday. As we need to buy some accessories to decorate and buy new chairs during the New Years Holiday. Also, the sales revenue is quite optimistic during these festivals.

Segment 2 generated $125.6309total sales. It is quite a disaster for this segment as it has the lowest sales volume among all segments. In this segment consumers buy most. The category of binders, furnishing, and paper shared the highest frequency from the statistical table. Therefore, the products in the tables group is the best-sellers in this season. We predict this is the season of Autumn as the category of Paper shares the highest proportion. It is not affected by any special festivals.

Segment 3 generated $1450.481 total sales. In this segment consumers also buy most as it shares the biggest proportion in this segment. The category of binders, tables and phones shared the highest frequency from the statistical table. Here we predicted that segment 3 is summer as the demand for tables is extremely huge. Before getting into September, students may purchase a new table and chair for a new semester.

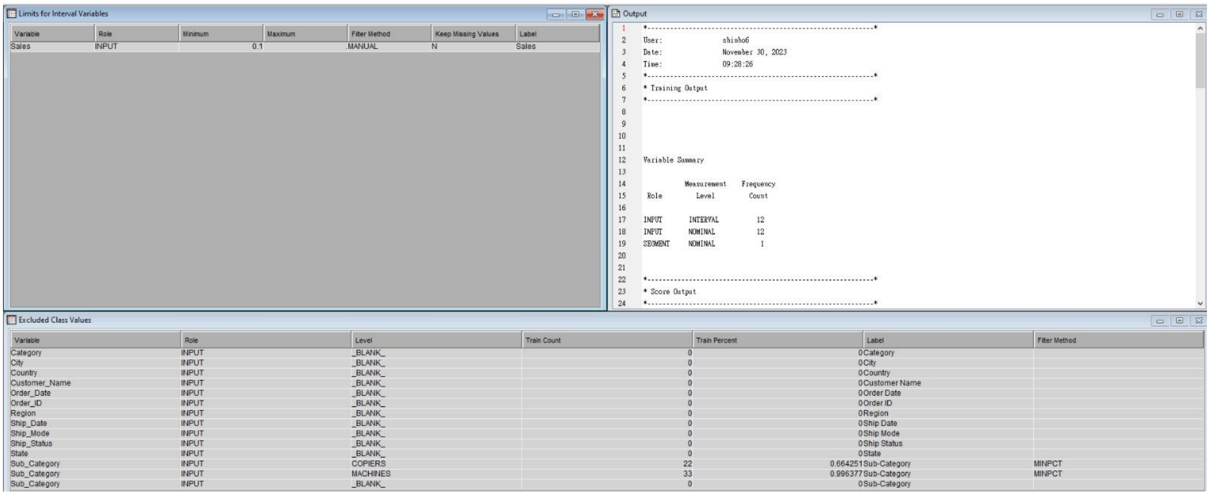
Segment 4 generated $1707.5 total sales, In this segment consumers and corporate also buy most as they shared the same proportion in this segment. The category of binders dominated in this segment, it shows 1 in this category group. So, we assume this segment is spring.

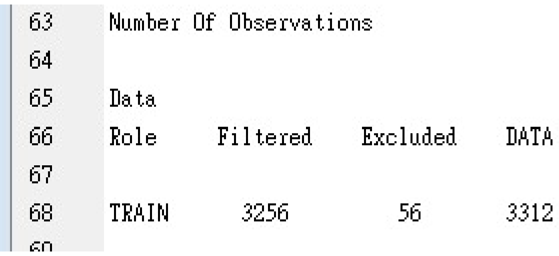
Objective 2 Setting:

<u>2.1 Data preparation</u>

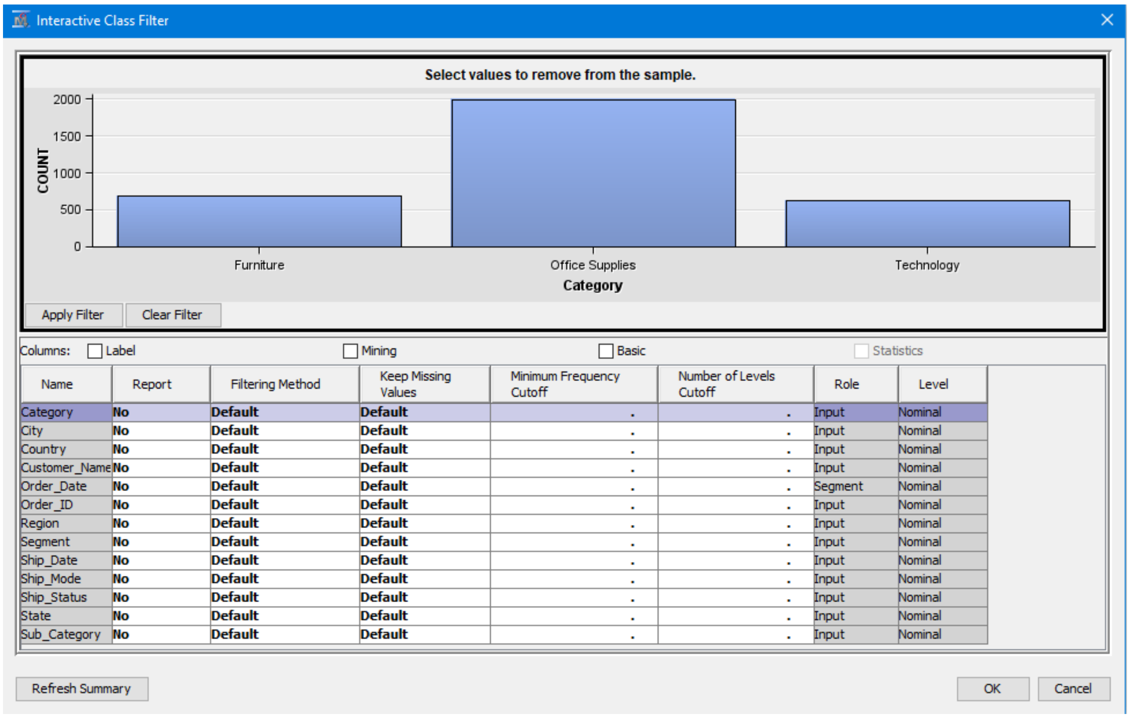
2.1.1 Filter

Property	Value
Export Table	Filtered
Tables to Filter	Training Data
Distribution Data Sets	Yes
Class Variables	
Class Variables	...
Default Filtering Method	Rare Values (Percentage)
Keep Missing Values	No
Normalized Values	Yes
Minimum Frequency Cutoff	1
Minimum Cutoff for Percent	0.01
Maximum Number of Levels	25
Interval Variables	
Interval Variables	...
Default Filtering Method	User-Specified Limits
Keep Missing Values	No
Tuning Parameters	...
Score	
Create Score Code	Yes
Update Measurement Level	No
Status	

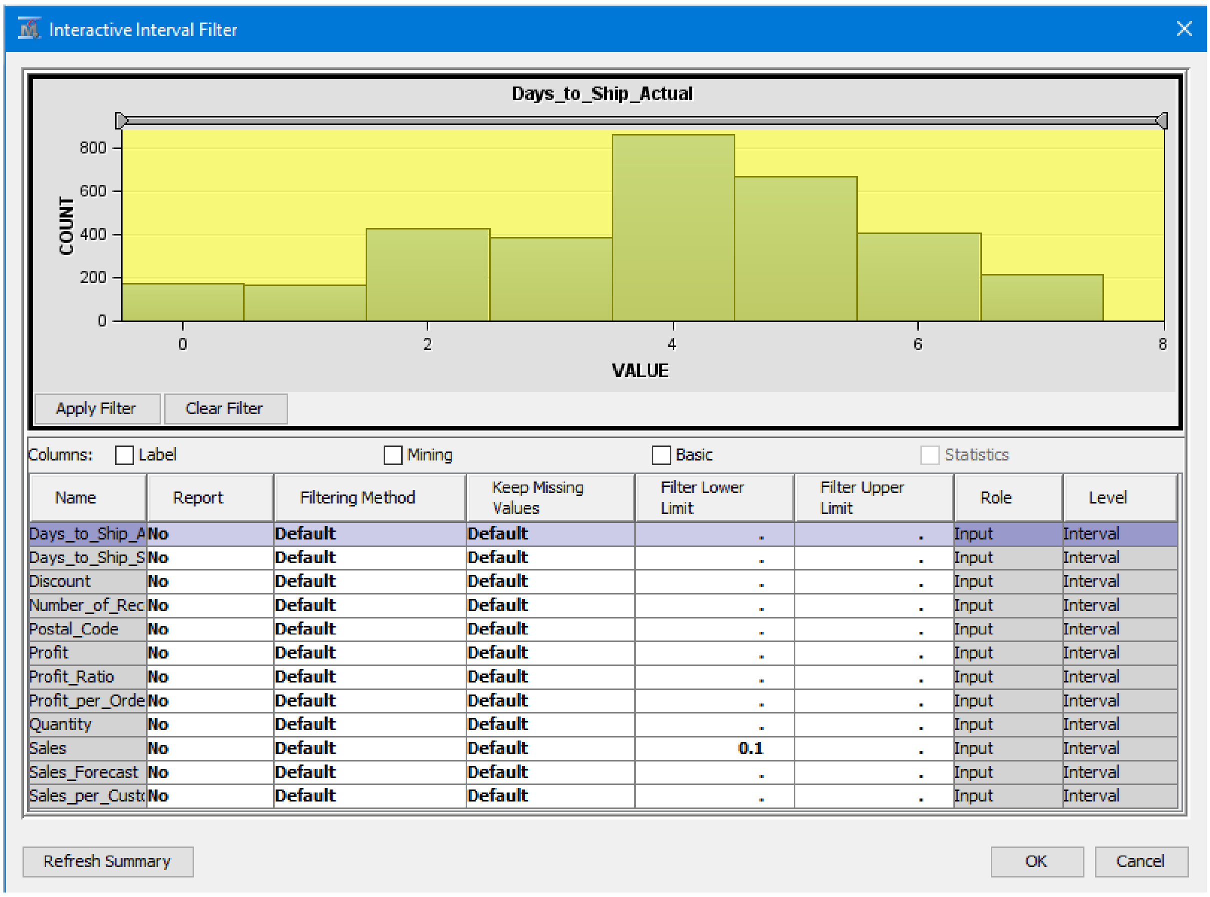




Class variable

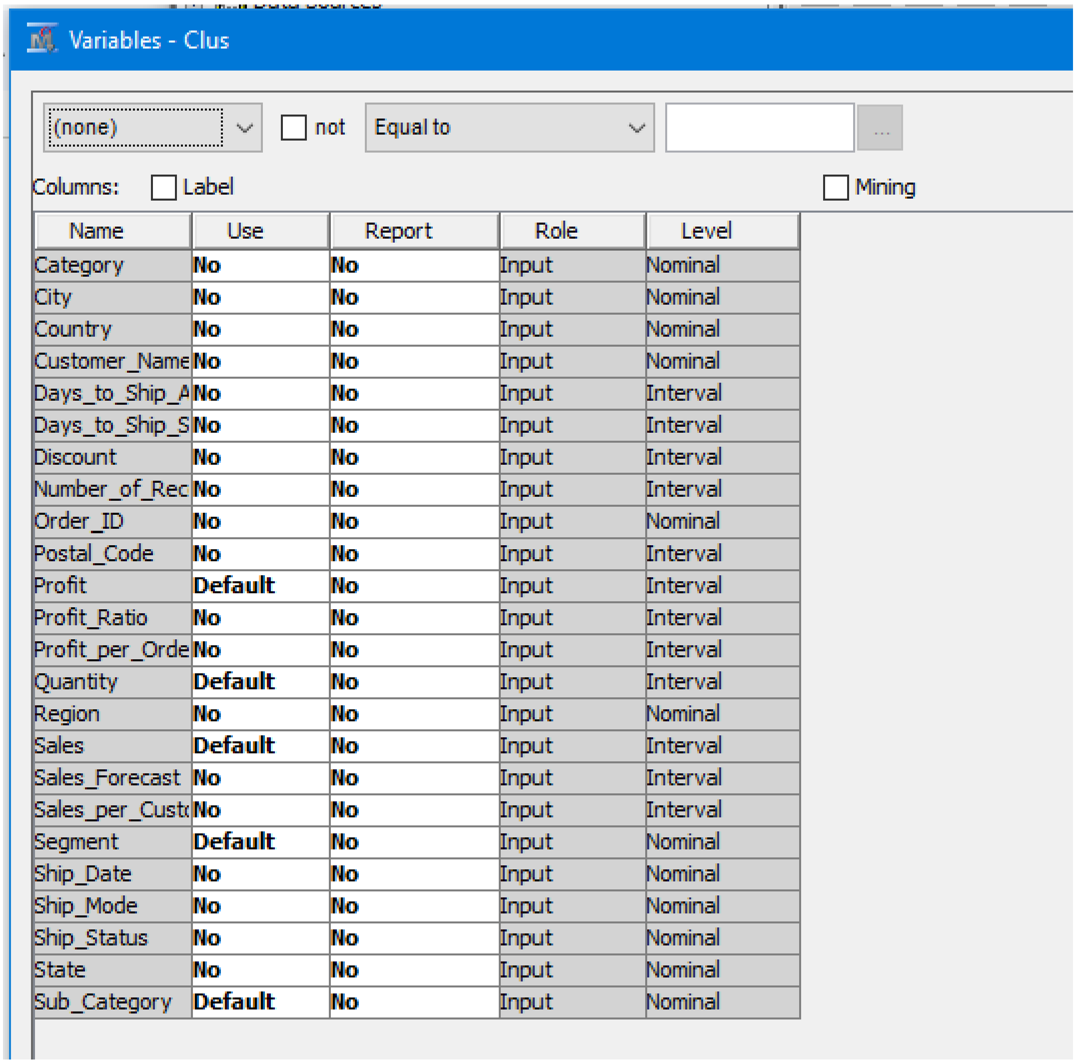


Interval variable



Name	Report	Filtering Method	Keep Missing Values	Filter Lower Limit	Filter Upper Limit	Role	Level
Days_to_Ship_A	No	Default	Default	.	.	Input	Interval
Days_to_Ship_S	No	Default	Default	.	.	Input	Interval
Discount	No	Default	Default	.	.	Input	Interval
Number_of_Rec	No	Default	Default	.	.	Input	Interval
Postal_Code	No	Default	Default	.	.	Input	Interval
Profit	No	Default	Default	.	.	Input	Interval
Profit_Ratio	No	Default	Default	.	.	Input	Interval
Profit_per_Orde	No	Default	Default	.	.	Input	Interval
Quantity	No	Default	Default	.	.	Input	Interval
Sales	No	Default	Default	0.1	.	Input	Interval
Sales_Forecast	No	Default	Default	.	.	Input	Interval
Sales_per_Custo	No	Default	Default	.	.	Input	Interval

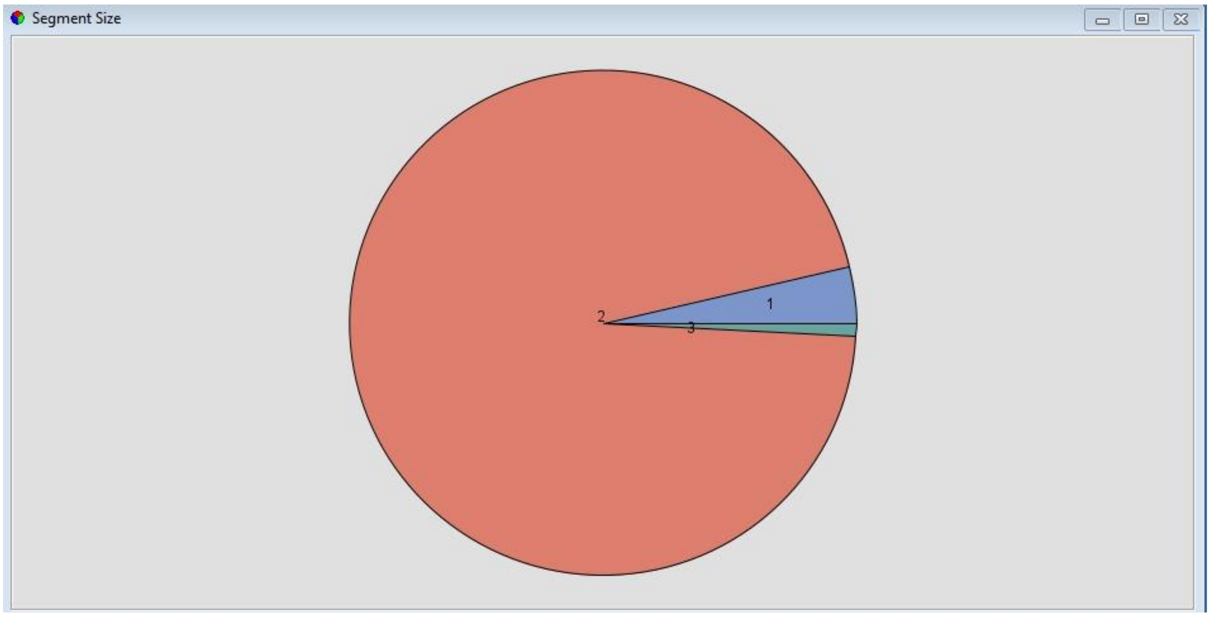
- Cluster 1



Name	Use	Report	Role	Level
Category	No	No	Input	Nominal
City	No	No	Input	Nominal
Country	No	No	Input	Nominal
Customer_Name	No	No	Input	Nominal
Days_to_Ship_A	No	No	Input	Interval
Days_to_Ship_S	No	No	Input	Interval
Discount	No	No	Input	Interval
Number_of_Rec	No	No	Input	Interval
Order_ID	No	No	Input	Nominal
Postal_Code	No	No	Input	Interval
Profit	Default	No	Input	Interval
Profit_Ratio	No	No	Input	Interval
Profit_per_Orde	No	No	Input	Interval
Quantity	Default	No	Input	Interval
Region	No	No	Input	Nominal
Sales	Default	No	Input	Interval
Sales_Forecast	No	No	Input	Interval
Sales_per_Cust	No	No	Input	Interval
Segment	Default	No	Input	Nominal
Ship_Date	No	No	Input	Nominal
Ship_Mode	No	No	Input	Nominal
Ship_Status	No	No	Input	Nominal
State	No	No	Input	Nominal
Sub_Category	Default	No	Input	Nominal

2.2 Cluster Result

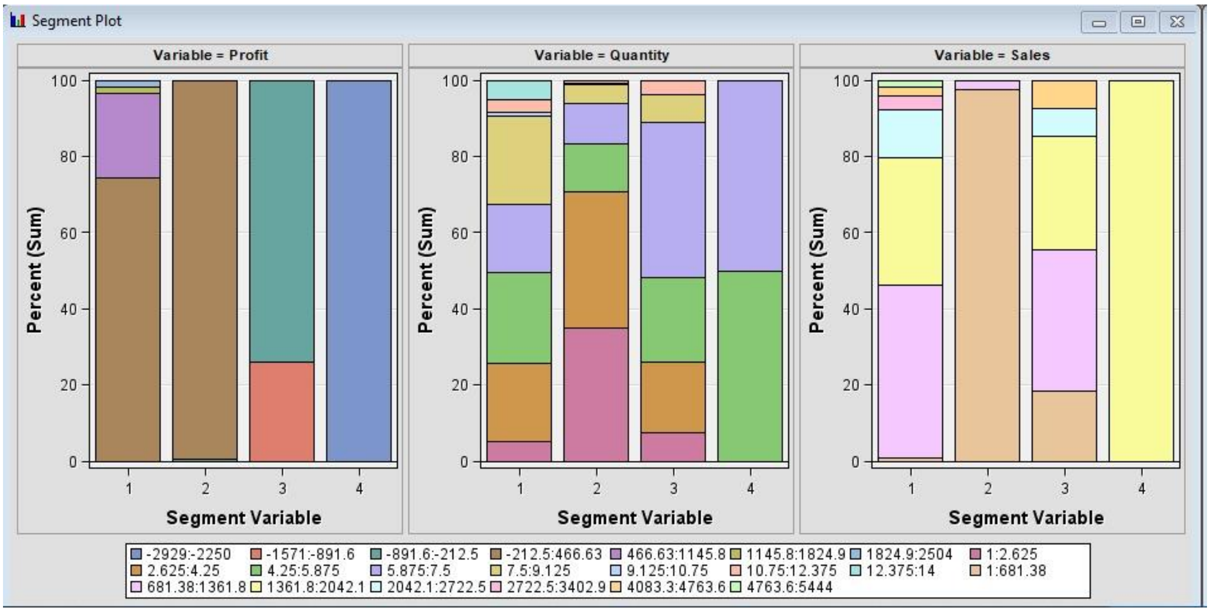
- Cluster1



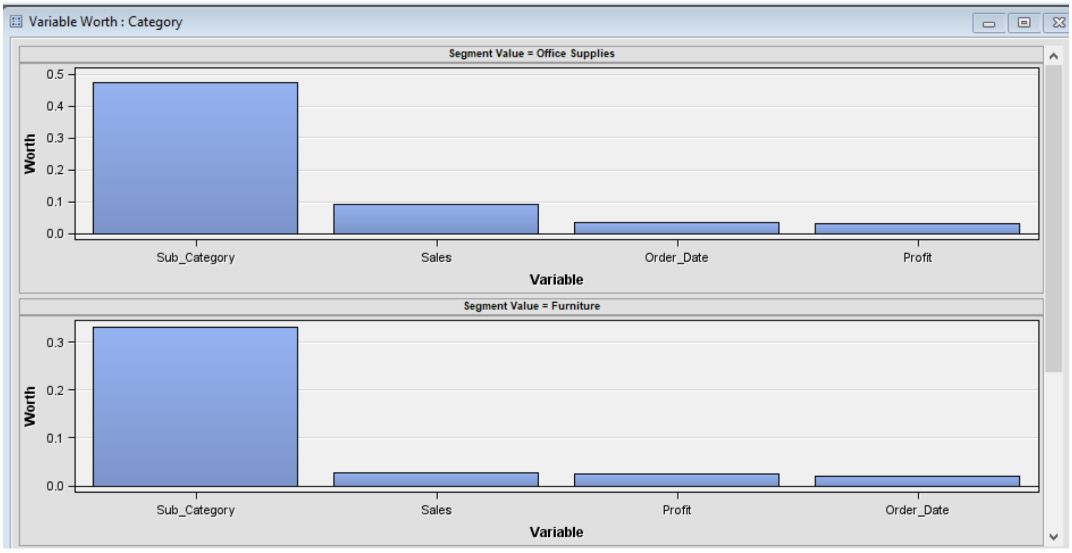
Segment Id	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Profit	Quantity	Sales	Segment=Consumer
1	117	0.811...	17.40858	2	4.832...	401.4786	6.358974	1654.256	0.418803
2	31...	0.392...	4.975241	1	4.832...	14.84277	3.65627	125.6309	0.509646
3	27	0.858...	8.814878	2	5.622...	-616.556	5.740741	1450.481	0.518519
4	2	0.755...	2.447889	3	13.94...	-2608.5	5.5	1707.5	0.5

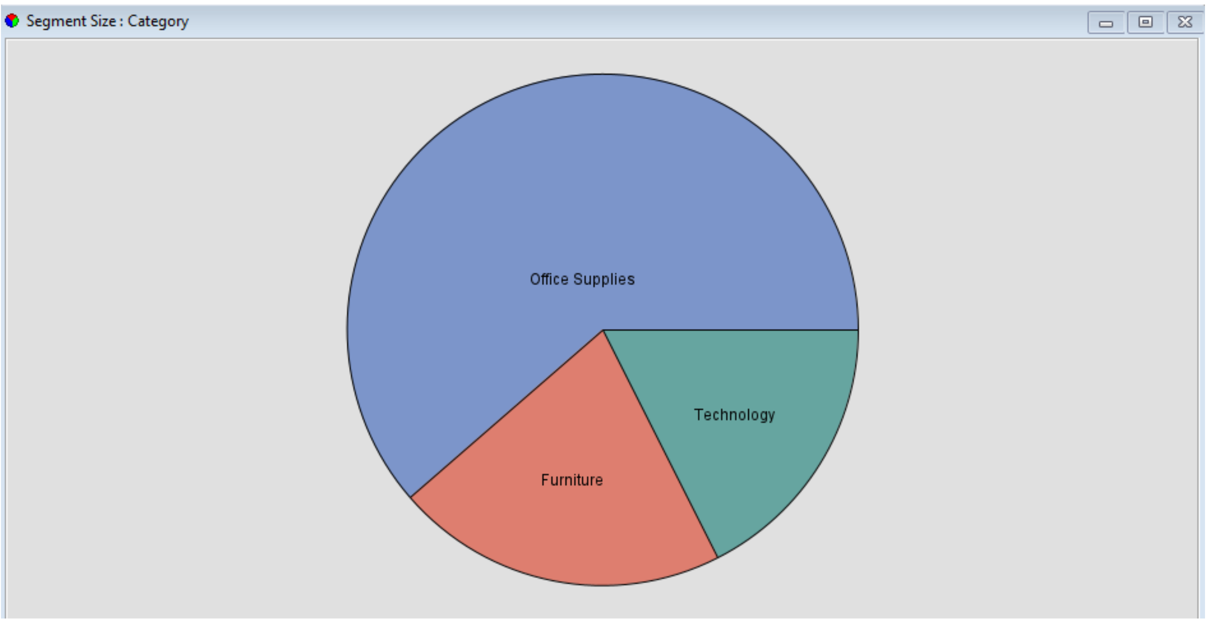
Segment Id	Segment=Corporate	Segment=Home Office	Sub_Category=Accessories	Sub_Category=Appliances	Sub_Category=Art	Sub_Category=Binders	Sub_Category=Bookcases	Sub_Category=Chairs	Sub_Category=Envelopes
1	0.410256	0.17094	0.153846	0.102564	1.8E-16	0.162393	0.051282	0.17094	-2.8E-17
2	0.290354	0.2	0.082637	0.048553	0.090675	0.15209	0.022508	0.054341	0.02283
3	0.296296	0.185185	-1.4E-17	0.037037	4.16E-17	0.222222	6.94E-18	0.037037	0
4	0.5	0	0	6.94E-18	0	1	0	-6.9E-18	-3.5E-18

Segment Id	Sub_Category=Fasteners	Sub_Category=Furnishings	Sub_Category=Labels	Sub_Category=Paper	Sub_Category=Phones	Sub_Category=Storage	Sub_Category=Supplies	Sub_Category=Tables
1	2.78E-17	0.017094	0.008547	3.89E-16	0.17094	0.08547	0.017094	0.059829
2	0.020579	0.100643	0.036334	0.147588	0.087138	0.089068	0.017685	0.027331
3	-1.4E-17	0.037037	-6.9E-18	0	0.111111	0.037037	0.074074	0.444444
4	-3.5E-18	0	0	0	0	0	0	0

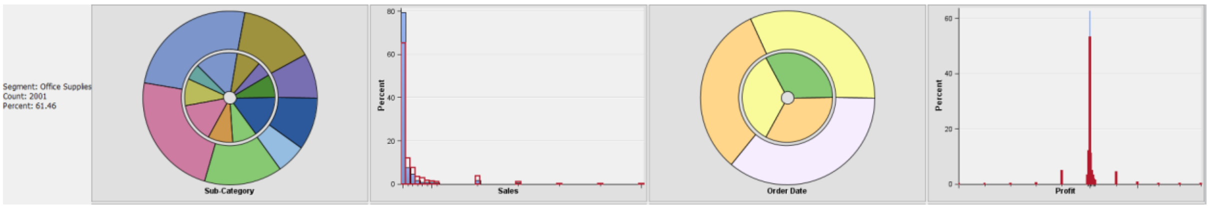


Product Type

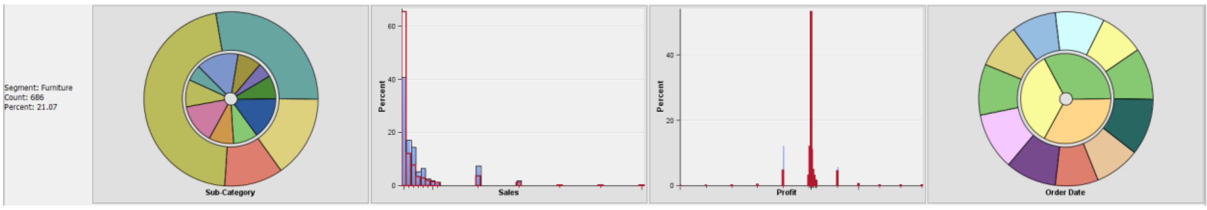




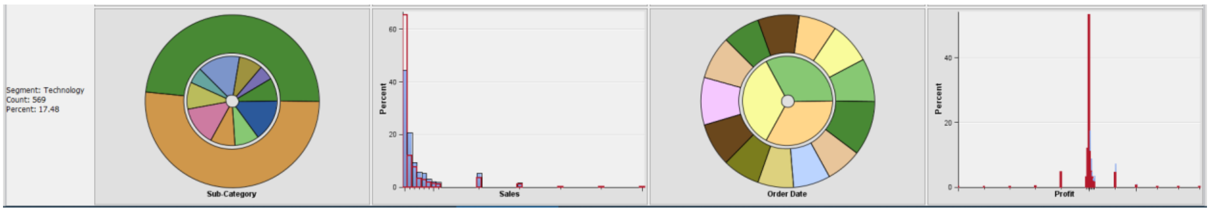
- Segment 1



- Segment 2



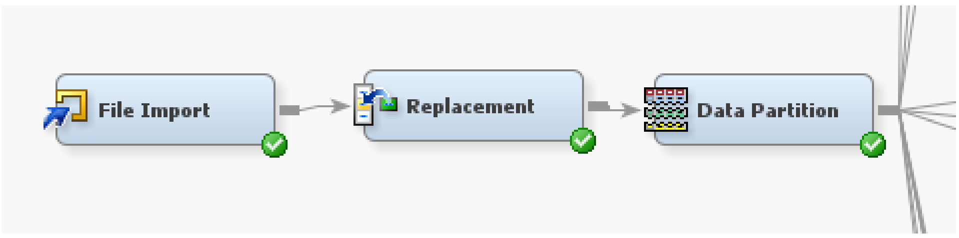
- Segment 3



Objective 3: Analysis customer preference for shipping

Purposes: Finding which factors affect the customer preference on selecting shipping mode(method)

3.1 Data Preparation



3.1.1 File Import

Name	Role	Level
Order_ID	ID	Nominal
Segment	Input	Ordinal
Discount	Input	Interval
Region	Input	Ordinal
Quantity	Input	Interval
Sales	Input	Interval
City	Input	Ordinal
Sub_Category	Input	Ordinal
Category	Input	Ordinal

Input variable {Segment, Discount, Region, Quantity, Sales, City, Sub_Category, Category}

Target variable {Ship_Mode}

3.1.2 Replacement

Since suppose the Sales should not be zero, it is an unreasonable data. Hence we would like to replace Sales with zero value missing. And set 1 as the lower limit field for Sales in the Interactive Replacement Interval Filter.

Property	Value
General	
Node ID	Repl
Imported Data	...
Exported Data	...
Notes	...
Train	
Interval Variables	
Replacement Editor	...
Default Limits Method	None
Cutoff Values	...
Class Variables	
Replacement Editor	...
Unknown Levels	Ignore
Score	
Replacement Values	Missing
Hide	No
Report	
Replacement Report	Yes
Status	
Create Time	11/29/23, 8:11 PM
Run ID	d9d0a384-9e4a-406a-a8c4-8
Last Error	
Last Status	Complete
Last Run Time	11/29/23, 9:35 PM
Run Duration	0 Hr. 0 Min. 1.84 Sec.
Grid Host	
User-Added Node	No

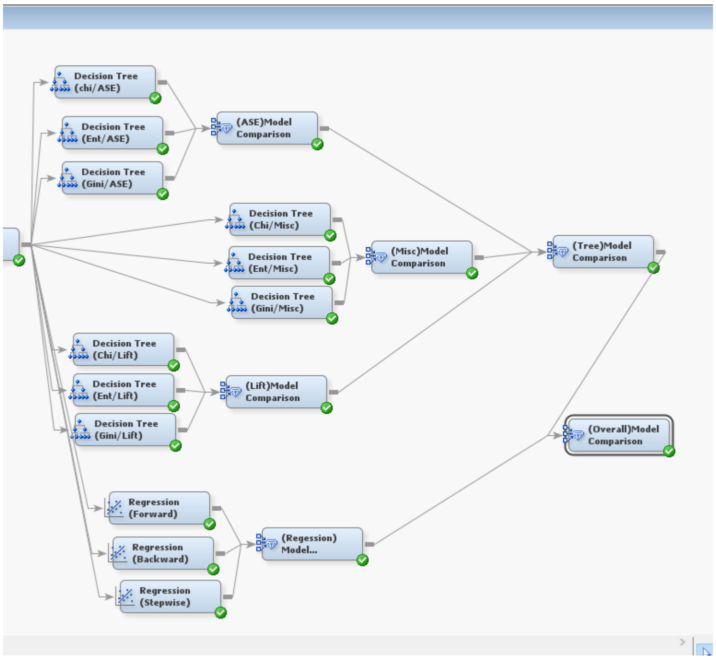
3.1.3 Data Partition

The data is split into 2 data sets, which is the Training and Validation data set. The proportion was set to be 50% each.

Random Seed	12345
Data Set Allocations	
Training	50.0
Validation	50.0
Test	0.0
Report	
Interval Targets	Yes

3.2 Methodolgy

We chose to use regression and decision trees to build different models and consider which has the best model by comparing with misclassification rate.



3.2.1 Decision tree

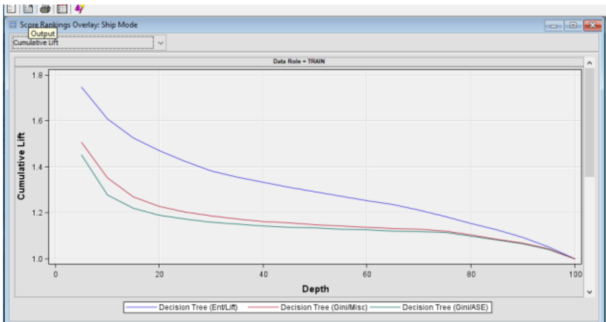
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	8
Maximum Depth	6
Minimum Categorical Size	5
Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Average Square Error
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345

- data are build with different criterion and assessment measure

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error	Valid: Sum of Squared Errors	Valid: Average Squared Error	Valid: Root Average Squared Error	Valid: Divisor for VASE	Train: Roc Index	Train: Gini Coefficient	Train: Kolmogorov-Smirnov Statistic
Y	Tree3	Tree3	Decision Tr...	Ship_Mode	Ship Mode	0.413398	1655	0.394562	0.977099	911.4304	0.137678	0.37105	6620	4965	1657	0.413398	1	971.1301	0.146519	0.382778	6628	0.628	0.255	0.201
	Tree	Tree	Decision Tr...	Ship_Mode	Ship Mode	0.425468	1655	0.409668	0.958904	933.063	0.140946	0.375428	6620	4965	1657	0.425468	1	990.2483	0.149404	0.386528	6628	0.627	0.254	0.193
	Tree2	Tree2	Decision Tr...	Ship_Mode	Ship Mode	0.429692	1655	0.38852	0.980392	870.2704	0.131461	0.362575	6620	4965	1657	0.429692	1	975.9758	0.14725	0.383732	6628	0.68	0.361	0.23

- (ASE) model comparision



Output

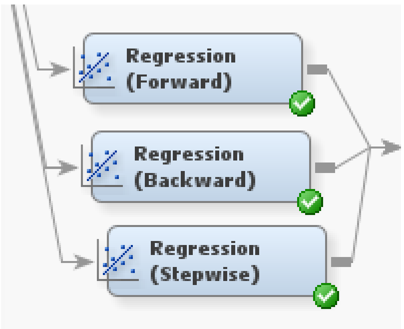
```
*------------------------------------------------------------*
User:               iajonuma2
Date:               November 29, 2023
Time:               21:43:40
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary

         Measurement    Frequency
 Role        Level         Count

TARGET     NOMINAL          1






                                                        BINNED_KS_
                                         _KS_PROB_         PROB_
```

Fit Statistics

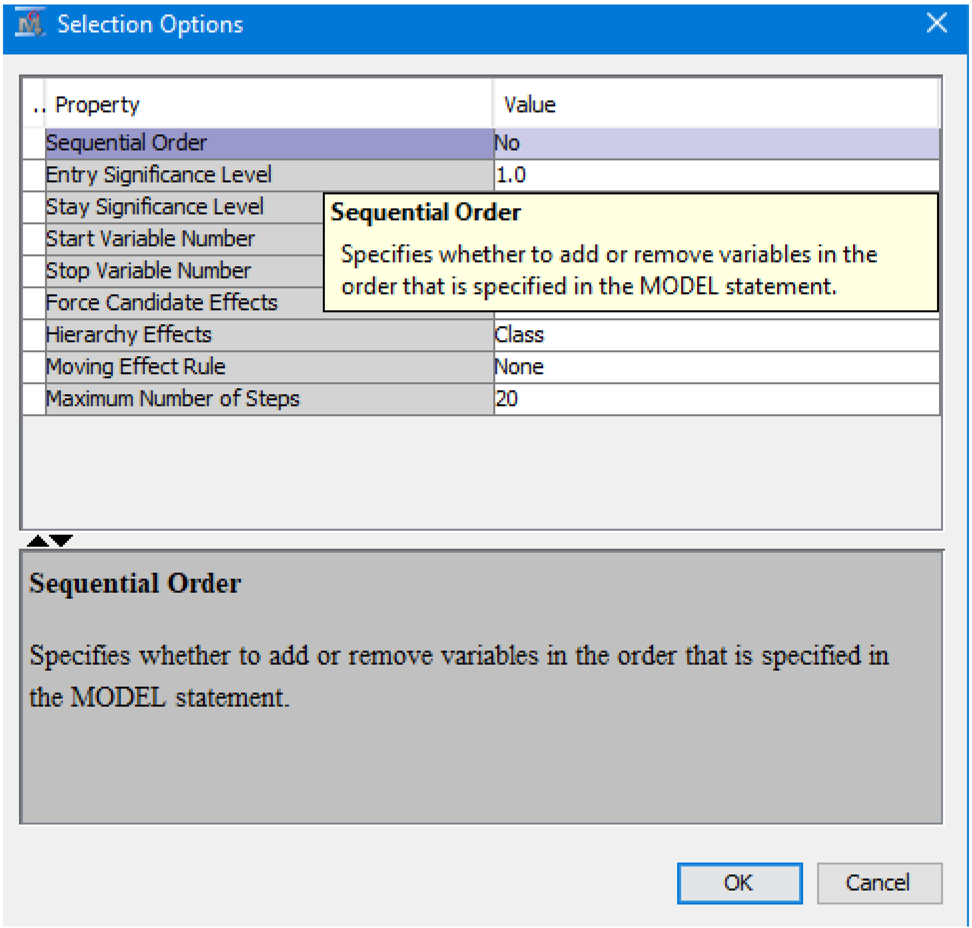
Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error	Valid: Sum of Squared Errors	Valid: Average Squared Error	Valid: Root Average Squared Error	Valid: Divisor for VASE	Train: Roc Index	Train: Gini Coefficient	Train: Kolmogorov-Smirnov Statistic
Y	MdlComp	Tree3	Decision Tr...	Ship_Mode	Ship Mode	0.146519	1655	0.394562	0.977099	911.4304	0.137678	0.37105	6620	4965	1657	0.413398	1	971.1301	0.146519	0.382778	6628	0.628	0.255	0.201
	MdlComp2	Tree6	Decision Tr...	Ship_Mode	Ship Mode	0.148139	1655	0.385498	0.977099	893.8401	0.135021	0.367452	6620	4965	1657	0.410984	1	981.866	0.148139	0.384888	6628	0.642	0.284	0.212
	MdlComp3	Tree8	Decision Tr...	Ship_Mode	Ship Mode	0.15238	1655	0.350453	0.97619	783.3555	0.118332	0.343994	6620	4965	1657	0.423657	1	1009.977	0.15238	0.390359	6628	0.75	0.501	0.358

- (Tree) model comparison

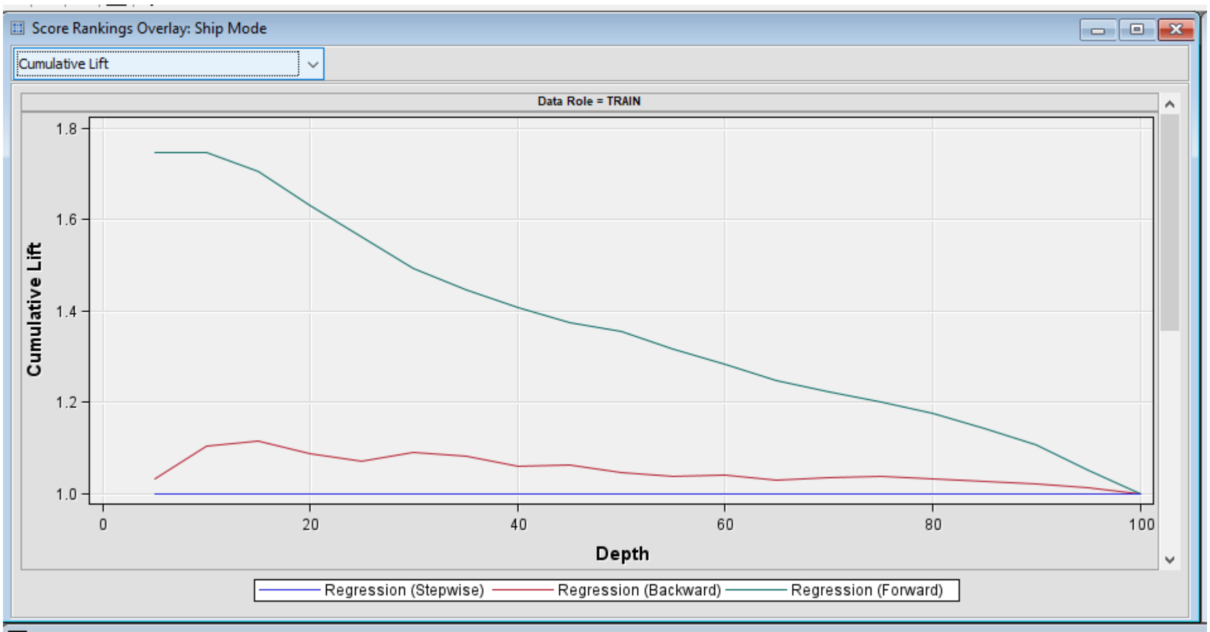
Regression model



⊟ Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
⊟ Model Options	
Suppress Intercept	No
Input Coding	Deviation
⊟ Model Selection	
Selection Model	Forward
Selection Criterion	Default
Use Selection Defaults	No
Selection Options	...
⊟ Optimization Options	


Selection Options
Property
Value
Sequential Order
No
Entry Significance Level
1.0
Stay Significance Level
Start Variable Number
Stop Variable Number
Force Candidate Effects
Hierarchy Effects
Class
Moving Effect Rule
None
Maximum Number of Steps
20
Sequential Order
Specifies whether to add or remove variables in the order that is specified in the MODEL statement.
Sequential Order
Specifies whether to add or remove variables in the order that is specified in the MODEL statement.
OK
Cancel

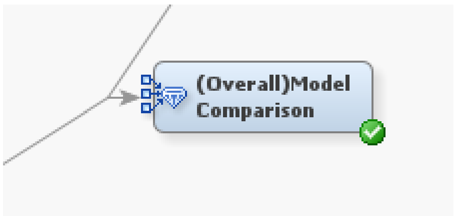
Property	Value
General	
Node ID	MdlComp5
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Assessment Reports	
Number of Bins	20
ROC Chart	Yes
Recompute	No
Model Selection	
Selection Data	Default
Selection Statistic	Average Squared Error
HP Selection Statistic	Default
SAS Viya Selection Statistic	...
Selection Table	Validation
Selection Depth	10
Score	
Selection Editor	...
Report	
Selected Model	
Target	Ship_Mode
Model Node	Reg
Model Description	Regression (Forward)
Selection Criteria	Valid: Average Squared Err
Status	
Create Time	11/29/23, 9:46 PM
Run ID	f4e05b61-c9db-43bf-82b1-
Last Error	
Last Status	Complete
Last Run Time	11/29/23, 10:31 PM
Run Duration	0 Hr. 0 Min. 3.39 Sec.
Grid Host	



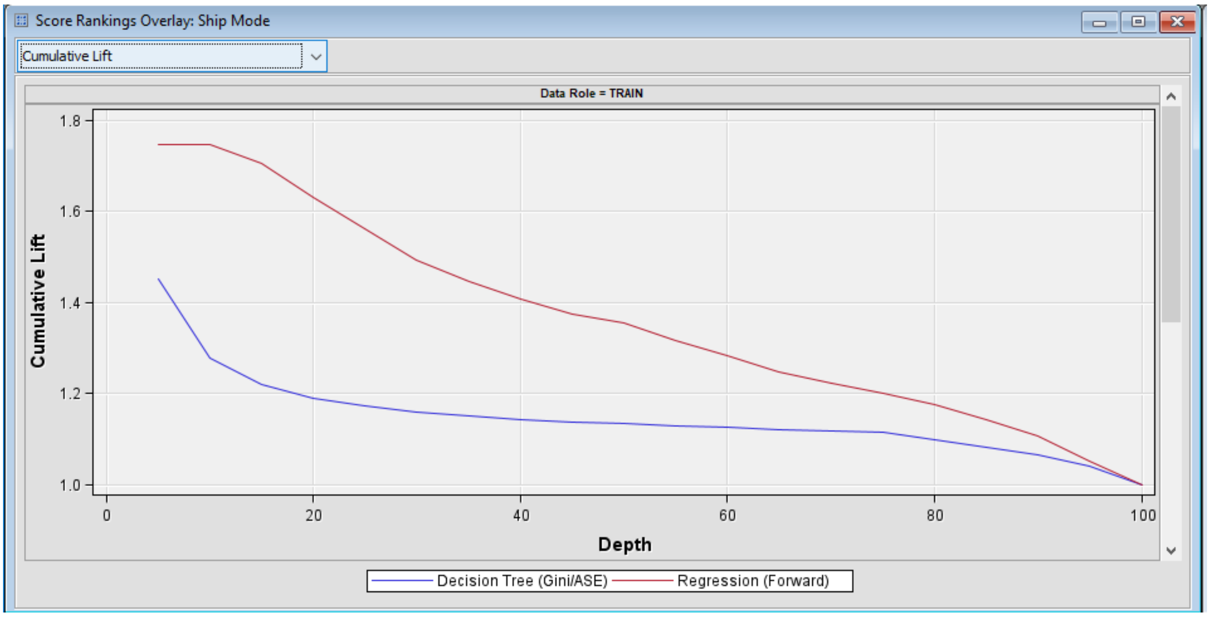
Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Total Degrees of Freedom	Train: Divisor for ASE	Train: Error Function	Train: Final Prediction Error	Train: Maximum Absolute Error	Train: Mean Square Error	Train: Sum of Frequencies	Train: Number of Estimate Weights	Train: Root Average Sum of Squares	Train: Root Final Prediction Error	Train: Root Mean Squared Error	Train: Schwarz's Bayesian Criterion	Train: Sum of Squared Errors
Y	Reg	Reg	Regression...	Ship_Mode	Ship Mode	0.142811	4297.077	0.106047	0.371764	4047	918	4965	6620	2461.077	0.154157	0.991412	0.130102	1655	918	0.325648	0.392628	0.360696	10273.41	702.028
	Reg3	Reg3	Regression...	Ship_Mode	Ship Mode	0.149906	3663.887	0.149898	0.552551	4962	3	4965	6620	3657.887	0.150079	0.943807	0.149988	1655	3	0.387166	0.3874	0.387283	3683.418	992.3215
	Reg2	Reg2	Regression...	Ship_Mode	Ship Mode	0.150808	3660.192	0.148578	0.544742	4938	27	4965	6620	3606.192	0.150203	0.993079	0.14939	1655	27	0.385458	0.38756	0.38651	3835.966	983.585

Overall comparison



.. Property	Value
General	
Node ID	MdlComp6
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Assessment Reports	
Number of Bins	20
ROC Chart	Yes
Recompute	No
Model Selection	
Selection Data	Default
Selection Statistic	Average Squared Error
HP Selection Statistic	Default
SAS Viya Selection Statistic	...
Selection Table	Train
Selection Depth	10
Score	
Selection Editor	...
Report	
Selected Model	
Target	Ship_Mode
Model Node	Reg
Model Description	Regression
Selection Criteria	Train: Average Squared Err
Status	
Create Time	11/29/23, 9:57 PM
Run ID	a5005d17-c1ac-43d4-a4ee
Last Error	
Last Status	Complete
Last Run Time	11/29/23, 10:00 PM
Run Duration	0 Hr. 0 Min. 3.42 Sec.
Grid Host	



Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Train: Average Squared Error	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error	Valid: Sum of Squared Errors	Valid: Average Squared Error	Valid: Root Average Squared Error	Valid: Divisor for VASE	Train: Akaike's Information Criterion	Train: Average Error Function
Y	MdlComp5	Reg	Regression (Forward)	Ship_Mode	Ship Mode	0.106047	1655	0.316012	0.991412	702.028	0.106047	0.325648	6620	4965	1657	0.395293	0.999899	946.549	0.142811	0.377903	6628	4297.077	0.371
	MdlComp4	Tree3	Decision Tree (Gini/ASE)	Ship_Mode	Ship Mode	0.137678	1655	0.394562	0.977099	911.4304	0.137678	0.37105	6620	4965	1657	0.413398	1	971.1301	0.146519	0.382778	6628	.	

Fit Statistics

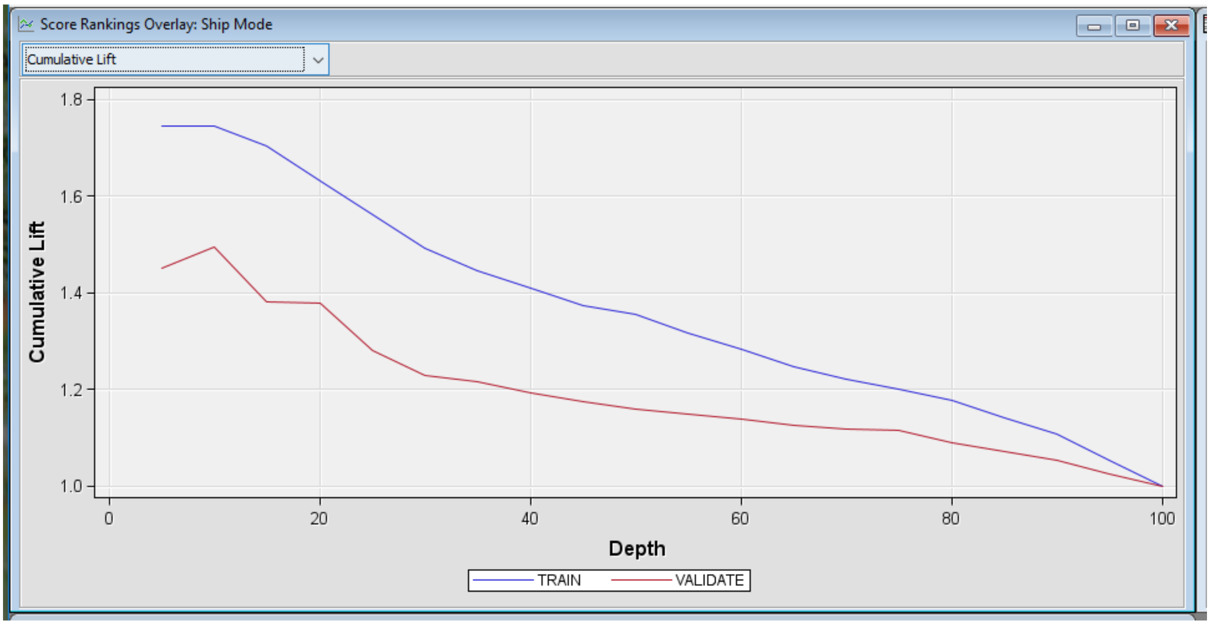
Model Selection based on Train: Average Squared Error (_ASE_)

Selected Model	Model Node	Model Description	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
Y	Reg	Regression	0.10605	0.31601	0.14281	0.39529
	Tree3	Decision Tree (3)	0.13768	0.39456	0.14652	0.41340

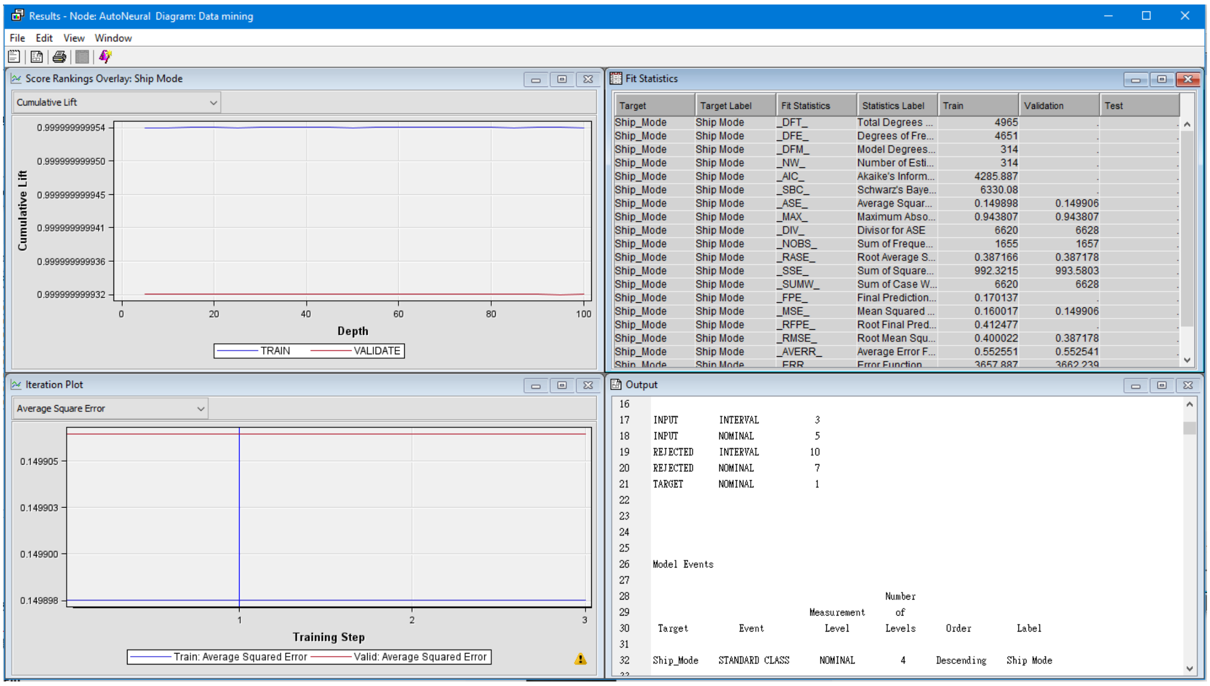
Best model
(regression forward)

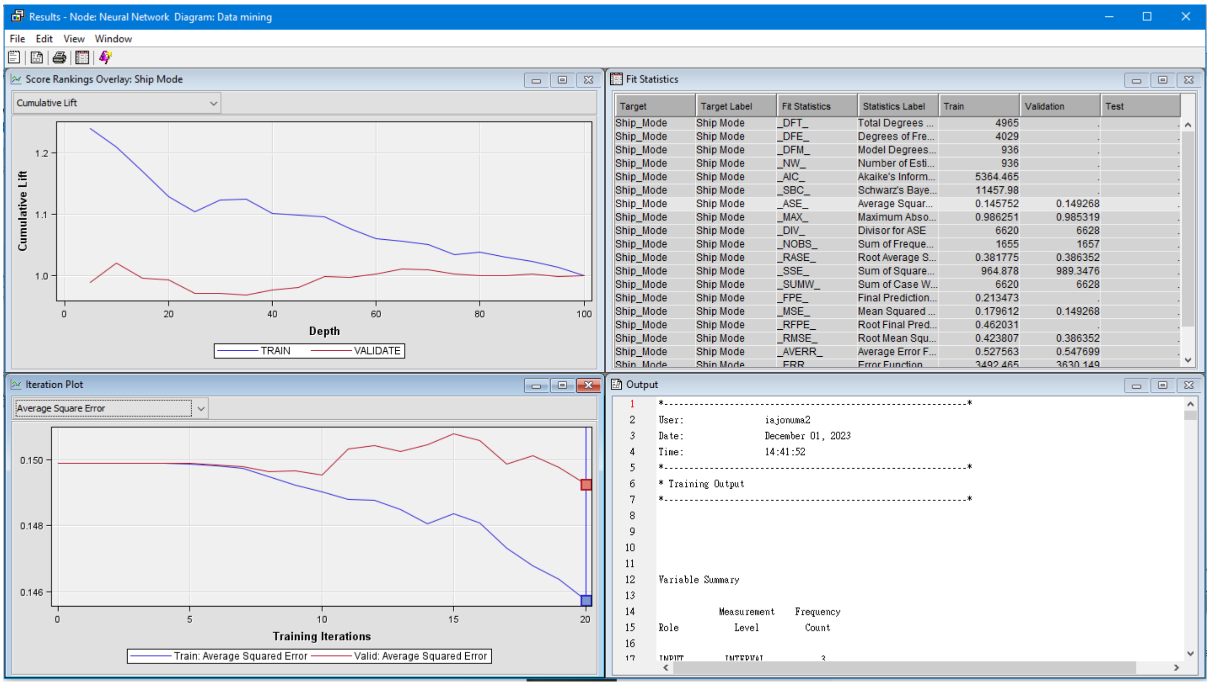
Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Category	6	7.3814	0.2870
City	842	417.4008	1.0000
Discount	3	1.9610	0.5806
Quantity	3	8.6327	0.0346
Region	9	15.1936	0.0858
Sales	3	1.4217	0.7004



NN(auto)-follow the tutorial




Results - Node: Neural Network Diagram: Data mining
File Edit View Window
Score Rankings Overlay: Ship Mode
Cumulative Lift
Cumulative Lift
1.2
1.1
1.0
0 20 40 60 80 100
Depth
TRAIN
VALIDATE
Fit Statistics
Target | Target Label | Fit Statistics | Statistics Label | Train | Validation | Test
Ship_Mode Ship Mode _DFT_ Total Degrees ... 4965 . .
Ship_Mode Ship Mode _DFE_ Degrees of Fre... 4029 . .
Ship_Mode Ship Mode _DFM_ Model Degrees... 936 . .
Ship_Mode Ship Mode _NW_ Number of Esti... 936 . .
Ship_Mode Ship Mode _AIC_ Akaike's Inform... 5364.465 . .
Ship_Mode Ship Mode _SBC_ Schwarz's Baye... 11457.98 . .
Ship_Mode Ship Mode _ASE_ Average Squar... 0.145752 0.149268 .
Ship_Mode Ship Mode _MAX_ Maximum Abso... 0.986251 0.985319 .
Ship_Mode Ship Mode _DIV_ Divisor for ASE 6620 6628 .
Ship_Mode Ship Mode _NOBS_ Sum of Freque... 1655 1657 .
Ship_Mode Ship Mode _RASE_ Root Average S... 0.381775 0.386352 .
Ship_Mode Ship Mode _SSE_ Sum of Square... 964.878 989.3476 .
Ship_Mode Ship Mode _SUMW_ Sum of Case W... 6620 6628 .
Ship_Mode Ship Mode _FPE_ Final Prediction... 0.213473 . .
Ship_Mode Ship Mode _MSE_ Mean Squared ... 0.179612 0.149268 .
Ship_Mode Ship Mode _RFPE_ Root Final Pred... 0.462031 . .
Ship_Mode Ship Mode _RMSE_ Root Mean Squ... 0.423807 0.386352 .
Ship_Mode Ship Mode _AVERR_ Average Error F... 0.527563 0.547699 .
Iteration Plot
Average Square Error
0.150
0.148
0.146
0 5 10 15 20
Training Iterations
Train: Average Squared Error
Valid: Average Squared Error
Output
User: iajonuma2
Date: December 01, 2023
Time: 14:41:52
* Training Output
Variable Summary
Measurement Frequency
Role Level Count

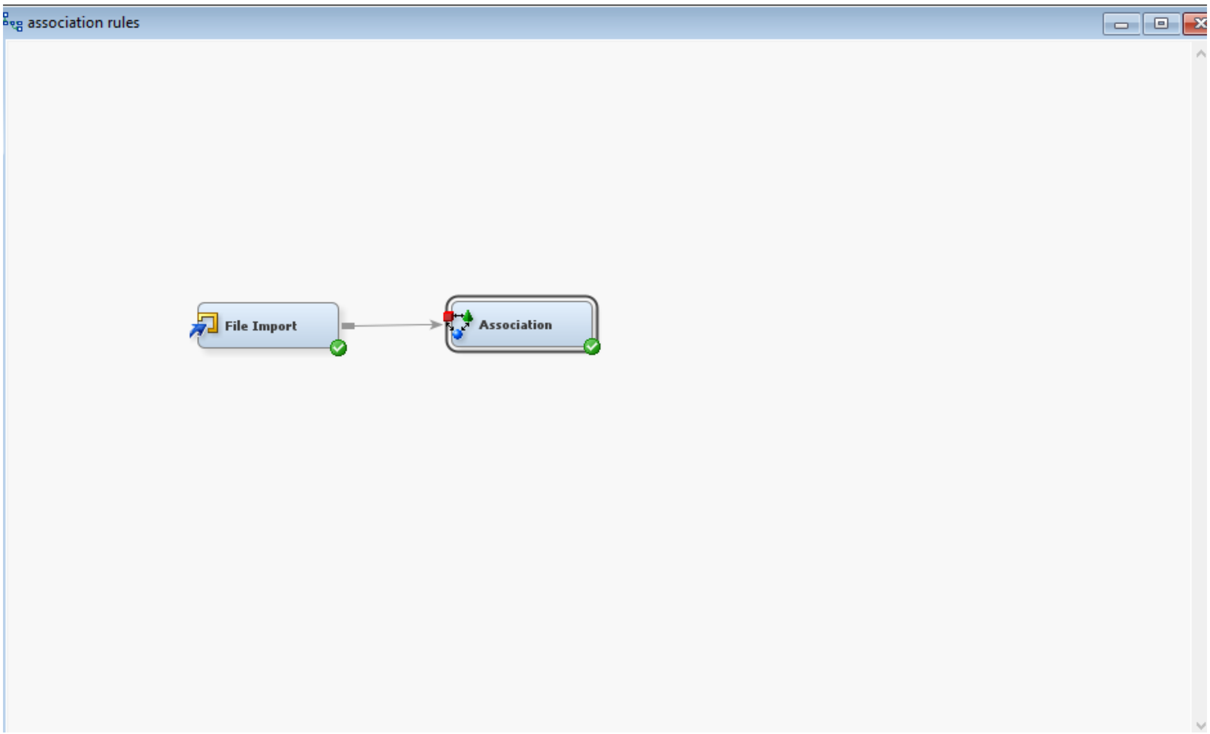
Objective 4: Identify shipping lanes with repeated delays.

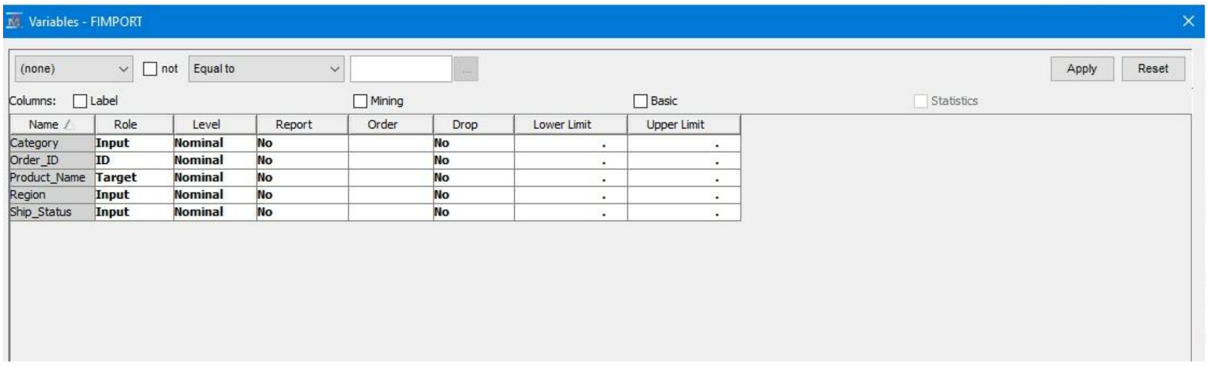
Purpose:

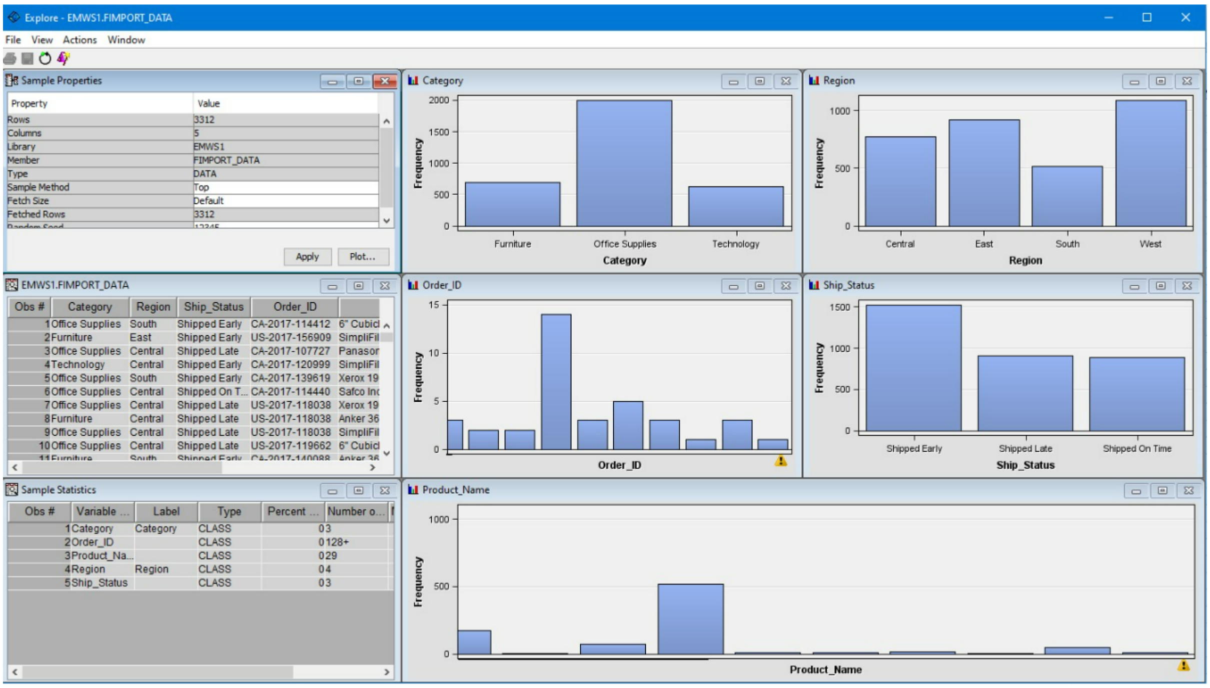
Identify shipping lanes with repeated delays based on the shipping preference and shipping status. To collect the customer information through interactions to reach our purposes, such as demographic information, purchasing behavior, and geographic information. Therefore, the company could give a better prediction to consumers to predict when they would receive their package. In the long term, this will reduce customer dissatisfaction and attract them to consume more in our company.

Methodology:

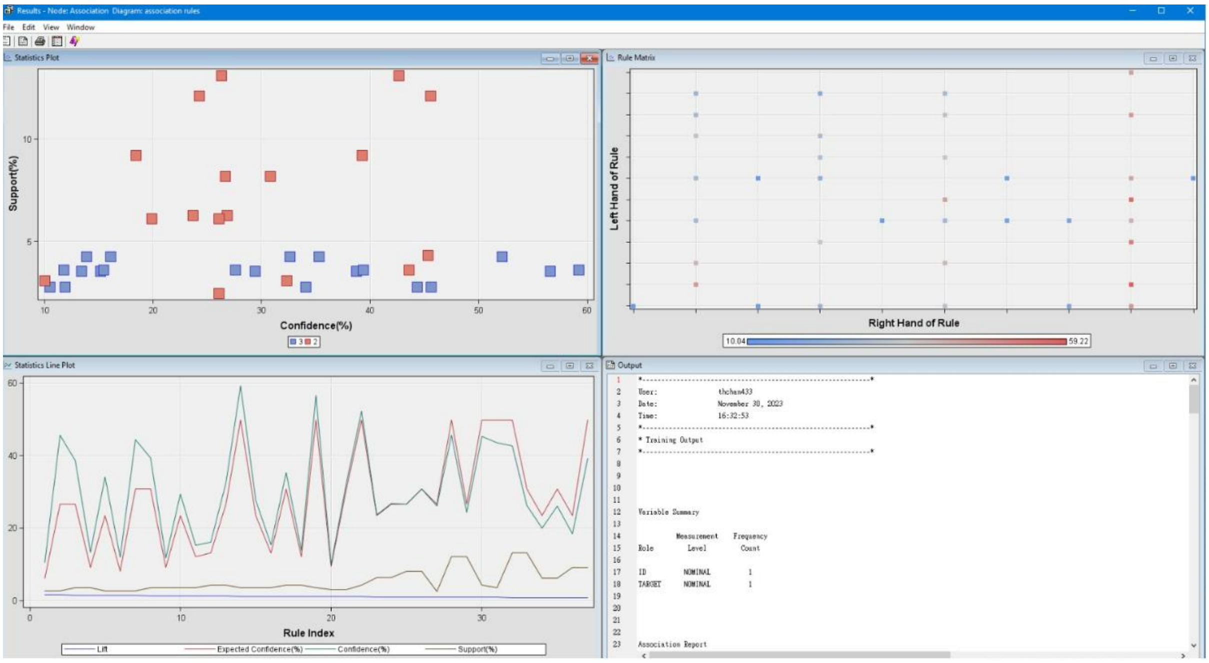
Input Variables: Product categories, region, ship status, order ID, product name




Variables - FIMPORT
(none)
not
Equal to
Apply
Reset
Columns:
Label
Mining
Basic
Statistics
Name
Role
Level
Report
Order
Drop
Lower Limit
Upper Limit
Category Input Nominal No No
Order_ID ID Nominal No No
Product_Name Target Nominal No No
Region Input Nominal No No
Ship_Status Input Nominal No No


Explore - EMWS1.FIMPORT_DATA
File View Actions Window
Sample Properties
Property
Value
Rows 3312
Columns 5
Library EMWS1
Member FIMPORT_DATA
Type DATA
Sample Method Top
Fetch Size Default
Fetched Rows 3312
Apply
Plot...
Category
Frequency
Furniture
Office Supplies
Technology
Region
Central
East
South
West
EMWS1.FIMPORT_DATA
Obs #
Category
Region
Ship_Status
Order_ID
Order_ID
Ship_Status
Shipped Early
Shipped Late
Shipped On Time
Sample Statistics
Obs #
Variable
Label
Type
Percent
Number o
1 Category Category CLASS 0 3
2 Order_ID CLASS 0 128+
3 Product_Na... CLASS 0 29
4 Region Region CLASS 0 4
5 Ship_Status CLASS 0 3
Product_Name

.. Property	Value
General	
Node ID	FIMPORT
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Import File	C:\Users\thchan433\Docu ...
Maximum Rows to Import	1000000
Maximum Columns to Import	10000
Delimiter	,
Name Row	Yes
Number of Rows to Skip	0
Guessing Rows	500
File Location	Local
File Type	xlsx
Advanced Advisor	No
Rerun	No
Score	
Role	Transaction
Report	
Summarize	No
Status	
Create Time	11/30/23, 4:29 PM
Run ID	
Last Error	
Last Status	
Last Run Time	
Run Duration	
Grid Host	
User-Added Node	No

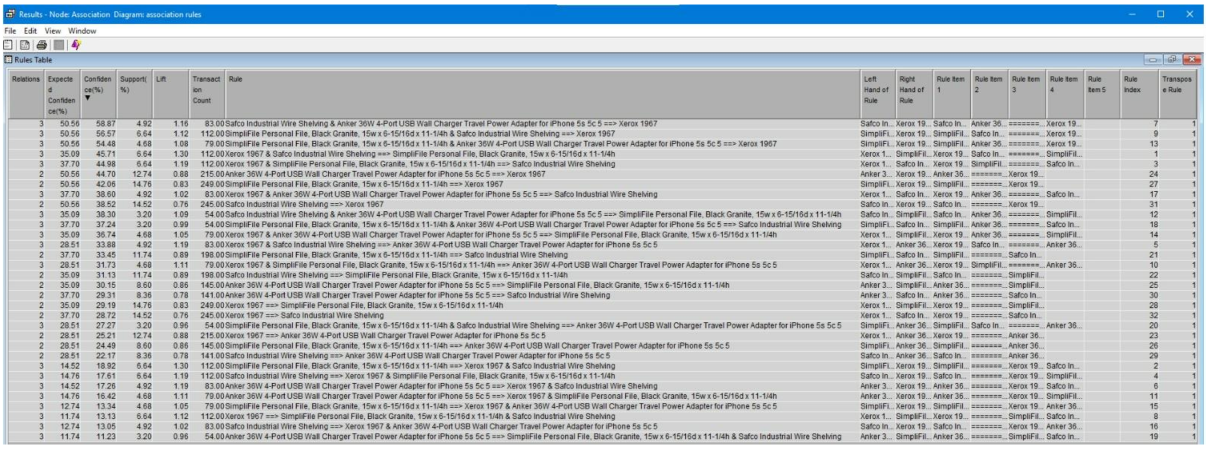


Explore - EMWS1.FIMPORT_DATA

File View Actions Window

EMWS1.FIMPORT_DATA

Obs #	Category	Region	Ship_Status	Order_ID	Product_Name
1	Office Supplies	South	Shipped Early	CA-2017-114412	Xerox 1967
2	Furniture	East	Shipped Early	US-2017-156909	Safco Industrial Wire Shelving
3	Office Supplies	Central	Shipped Late	CA-2017-107727	Xerox 1967
4	Technology	Central	Shipped Early	CA-2017-120999	Safco Industrial Wire Shelving
5	Office Supplies	South	Shipped Early	CA-2017-139619	Xerox 1967
6	Office Supplies	Central	Shipped On T...	CA-2017-114440	Xerox 1967
7	Office Supplies	Central	Shipped Late	US-2017-118038	Xerox 1967
8	Furniture	Central	Shipped Late	US-2017-118038	Economy Binders
9	Office Supplies	Central	Shipped Late	US-2017-118038	Xerox 1967
10	Office Supplies	Central	Shipped Late	US-2017-119662	SimpliFile Personal File, Blac...
11	Furniture	South	Shipped Early	CA-2017-140088	Xerox 1967
12	Technology	Central	Shipped Late	CA-2017-155558	Anker 36W 4-Port USB Wall C...
13	Office Supplies	Central	Shipped Late	CA-2017-155558	Safco Industrial Wire Shelving
14	Office Supplies	West	Shipped On T...	US-2017-109484	SimpliFile Personal File, Blac...
15	Furniture	East	Shipped Early	CA-2017-161018	Anker 36W 4-Port USB Wall C...
16	Office Supplies	West	Shipped Late	CA-2017-157833	SimpliFile Personal File, Blac...
17	Technology	South	Shipped Early	CA-2017-119004	Safco Industrial Wire Shelving
18	Technology	South	Shipped Early	CA-2017-119004	Xerox 1967
19	Office Supplies	South	Shipped Early	CA-2017-119004	Xerox 1967
20	Furniture	East	Shipped Early	CA-2017-146780	Xerox 1967
21	Office Supplies	West	Shipped Late	US-2017-107272	Xerox 1967
22	Office Supplies	West	Shipped Late	US-2017-107272	SimpliFile Personal File, Blac...
23	Technology	East	Shipped Late	US-2017-164147	Safco Industrial Wire Shelving
24	Office Supplies	East	Shipped Late	US-2017-164147	Sanyo 2.5 Cubic Foot Mid-Siz...
25	Office Supplies	East	Shipped Late	US-2017-164147	Xerox 1967
26	Office Supplies	West	Shipped Early	CA-2017-106180	SimpliFile Personal File, Blac...
27	Office Supplies	West	Shipped Early	CA-2017-106180	Xerox 1967
28	Office Supplies	West	Shipped Early	CA-2017-106180	Xerox 1967
29	Office Supplies	Central	Shipped Early	CA-2017-155376	Safco Industrial Wire Shelving
30	Office Supplies	Central	Shipped Late	US-2017-152366	Safco Industrial Wire Shelving
31	Office Supplies	East	Shipped Late	CA-2017-107720	Safco Industrial Wire Shelving
32	Office Supplies	East	Shipped Late	US-2017-124303	Safco Industrial Wire Shelving
33	Office Supplies	East	Shipped Late	US-2017-124303	Xerox 1967
34	Office Supplies	East	Shipped Early	CA-2017-105074	SimpliFile Personal File, Blac...
35	Office Supplies	Central	Shipped Late	US-2017-116701	Safco Industrial Wire Shelving
36	Furniture	South	Shipped Early	CA-2017-126382	Safco Industrial Wire Shelving
37	Technology	West	Shipped Early	CA-2017-108329	Anker 36W 4-Port USB Wall C...
38	Office Supplies	Central	Shipped On T...	CA-2017-135860	Xerox 1967
39	Technology	Central	Shipped On T...	CA-2017-135860	SimpliFile Personal File, Blac...
40	Office Supplies	Central	Shipped On T...	CA-2017-135860	Safco Industrial Wire Shelving
41	Office Supplies	Central	Shipped On T...	CA-2017-135860	SimpliFile Personal File, Blac...
42	Office Supplies	Central	Shipped On T...	CA-2017-135860	Anker 36W 4-Port USB Wall C...

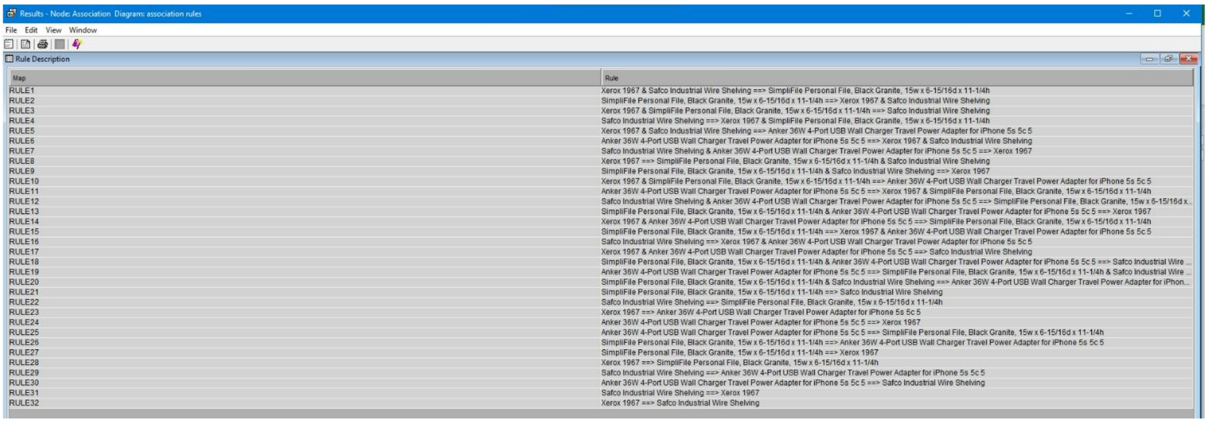


Results - Node: Association Diagram: association rules

File Edit View Window

Rules Table

Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule Index	Transpose Rule
3	50.56	58.87	4.92	1.16	83.00	Safco Industrial Wire Shelving & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967	Safco In...	Xerox 19...	Safco In...	Anker 36...	=======...	Xerox 19...		7	1
3	50.56	56.57	6.64	1.12	112.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving ==> Xerox 1967	SimpliFi...	Xerox 19...	SimpliFil...	Safco In...	=======...	Xerox 19...		9	1
3	50.56	54.48	4.68	1.08	79.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967	SimpliFi...	Xerox 19...	SimpliFil...	Anker 36...	=======...	Xerox 19...		13	1
3	35.09	45.71	6.64	1.30	112.00	Xerox 1967 & Safco Industrial Wire Shelving ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Xerox 1...	SimpliFil...	Xerox 19...	Safco In...	=======...	SimpliFil...		1	1
3	37.70	44.98	6.64	1.19	112.00	Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Safco Industrial Wire Shelving	Xerox 1...	Safco In...	Xerox 19...	SimpliFil...	=======...	Safco In...		3	1
2	50.56	44.70	12.74	0.88	215.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967	Anker 3...	Xerox 19...	Anker 36...	=======...	Xerox 19...			24	1
2	50.56	42.06	14.76	0.83	249.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967	SimpliFi...	Xerox 19...	SimpliFil...	=======...	Xerox 19...			27	1
3	37.70	38.60	4.92	1.02	83.00	Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire Shelving	Xerox 1...	Safco In...	Xerox 19...	Anker 36...	=======...	Safco In...		17	1
2	50.56	38.52	14.52	0.76	245.00	Safco Industrial Wire Shelving ==> Xerox 1967	Safco In...	Xerox 19...	Safco In...	=======...	Xerox 19...			31	1
3	35.09	38.30	3.20	1.09	54.00	Safco Industrial Wire Shelving & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Safco In...	SimpliFil...	Safco In...	Anker 36...	=======...	SimpliFil...		12	1
3	37.70	37.24	3.20	0.99	54.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire Shelving	SimpliFi...	Safco In...	SimpliFil...	Anker 36...	=======...	Safco In...		18	1
3	35.09	36.74	4.68	1.05	79.00	Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Xerox 1...	SimpliFil...	Xerox 19...	Anker 36...	=======...	SimpliFil...		14	1
3	28.51	33.88	4.92	1.19	83.00	Xerox 1967 & Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	Xerox 1...	Anker 36...	Xerox 19...	Safco In...	=======...	Anker 36...		5	1
2	37.70	33.45	11.74	0.89	198.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Safco Industrial Wire Shelving	SimpliFi...	Safco In...	SimpliFil...	=======...	Safco In...			21	1
3	28.51	31.73	4.68	1.11	79.00	Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	Xerox 1...	Anker 36...	Xerox 19...	SimpliFil...	=======...	Anker 36...		10	1
2	35.09	31.13	11.74	0.89	198.00	Safco Industrial Wire Shelving ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Safco In...	SimpliFil...	Safco In...	=======...	SimpliFil...			22	1
2	35.09	30.15	8.60	0.86	145.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Anker 3...	SimpliFil...	Anker 36...	=======...	SimpliFil...			25	1
2	37.70	29.31	8.36	0.78	141.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire Shelving	Anker 3...	Safco In...	Anker 36...	=======...	Safco In...			30	1
2	35.09	29.19	14.76	0.83	249.00	Xerox 1967 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Xerox 1...	SimpliFil...	Xerox 19...	=======...	SimpliFil...			28	1
2	37.70	28.72	14.52	0.76	245.00	Xerox 1967 ==> Safco Industrial Wire Shelving	Xerox 1...	Safco In...	Xerox 19...	=======...	Safco In...			32	1
3	28.51	27.27	3.20	0.96	54.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	SimpliFi...	Anker 36...	SimpliFil...	Safco In...	=======...	Anker 36...		20	1
2	28.51	25.21	12.74	0.88	215.00	Xerox 1967 ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	Xerox 1...	Anker 36...	Xerox 19...	=======...	Anker 36...			23	1
2	28.51	24.49	8.60	0.86	145.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	SimpliFi...	Anker 36...	SimpliFil...	=======...	Anker 36...			26	1
2	28.51	22.17	8.36	0.78	141.00	Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	Safco In...	Anker 36...	Safco In...	=======...	Anker 36...			29	1
3	14.52	18.92	6.64	1.30	112.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967 & Safco Industrial Wire Shelving	SimpliFi...	Xerox 19...	SimpliFil...	=======...	Xerox 19...	Safco In...		2	1
3	14.76	17.61	6.64	1.19	112.00	Safco Industrial Wire Shelving ==> Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Safco In...	Xerox 19...	Safco In...	=======...	Xerox 19...	SimpliFil...		4	1
3	14.52	17.26	4.92	1.19	83.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967 & Safco Industrial Wire Shelving	Anker 3...	Xerox 19...	Anker 36...	=======...	Xerox 19...	Safco In...		6	1
3	14.76	16.42	4.68	1.11	79.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h	Anker 3...	Xerox 19...	Anker 36...	=======...	Xerox 19...	SimpliFil...		11	1
3	12.74	13.34	4.68	1.05	79.00	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	SimpliFi...	Xerox 19...	SimpliFil...	=======...	Xerox 19...	Anker 36...		15	1
3	11.74	13.13	6.64	1.12	112.00	Xerox 1967 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving	Xerox 1...	SimpliFil...	Xerox 19...	=======...	SimpliFil...	Safco In...		8	1
3	12.74	13.05	4.92	1.02	83.00	Safco Industrial Wire Shelving ==> Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5	Safco In...	Xerox 19...	Safco In...	=======...	Xerox 19...	Anker 36...		16	1
3	11.74	11.23	3.20	0.96	54.00	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving	Anker 3...	SimpliFil...	Anker 36...	=======...	SimpliFil...	Safco In...		19	1

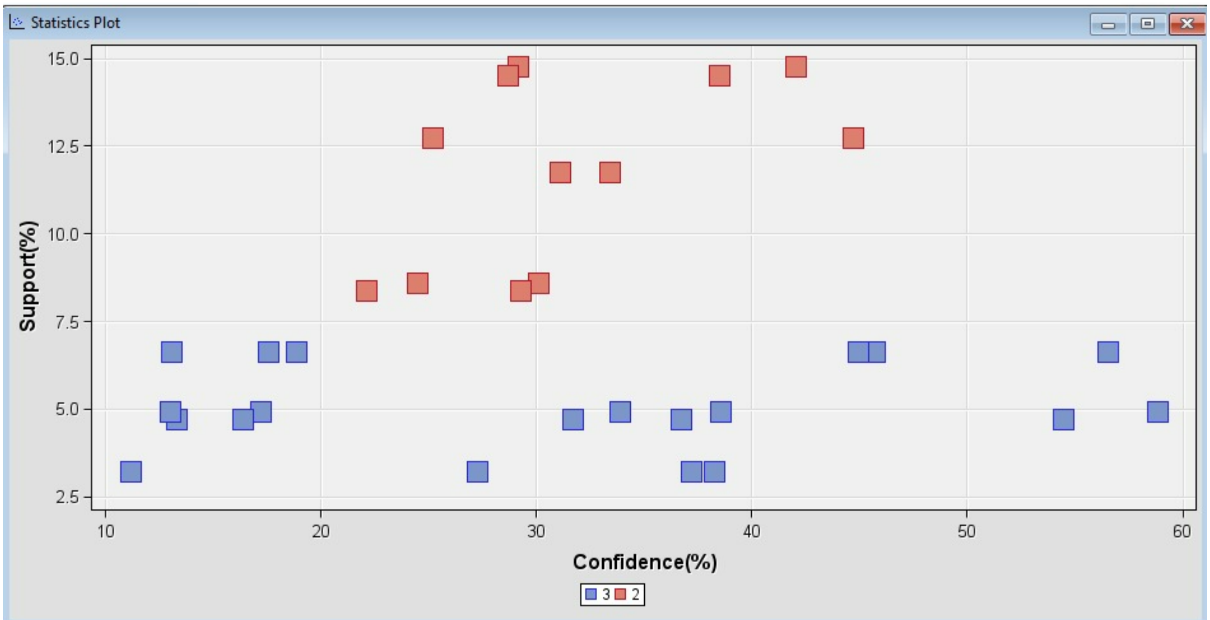


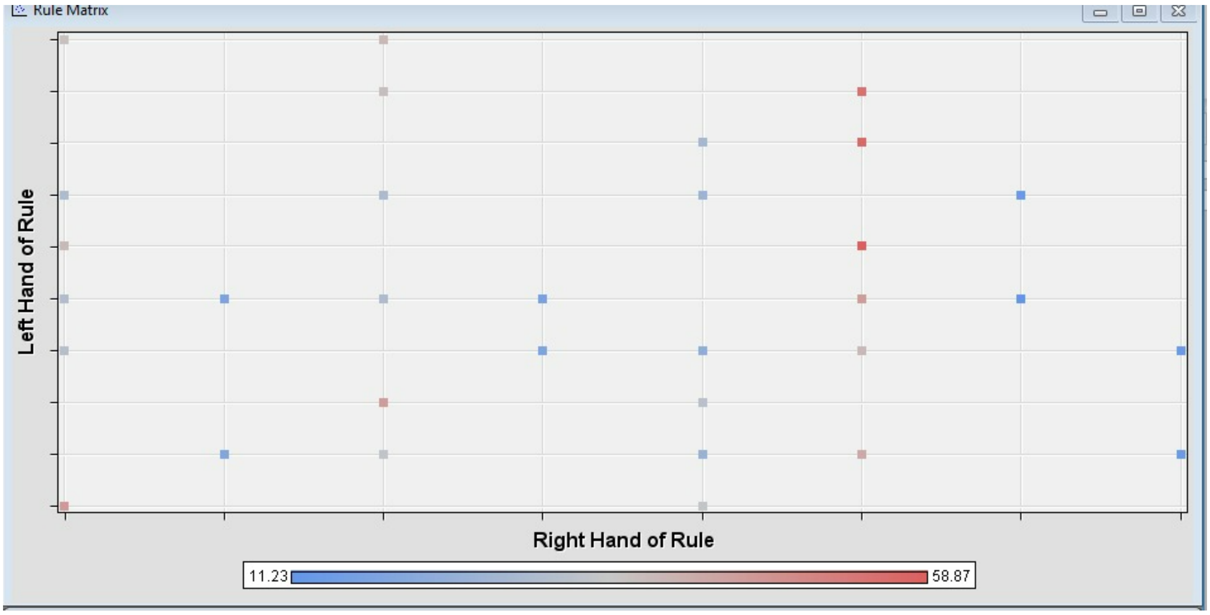
Results - Node: Association Diagram: association rules

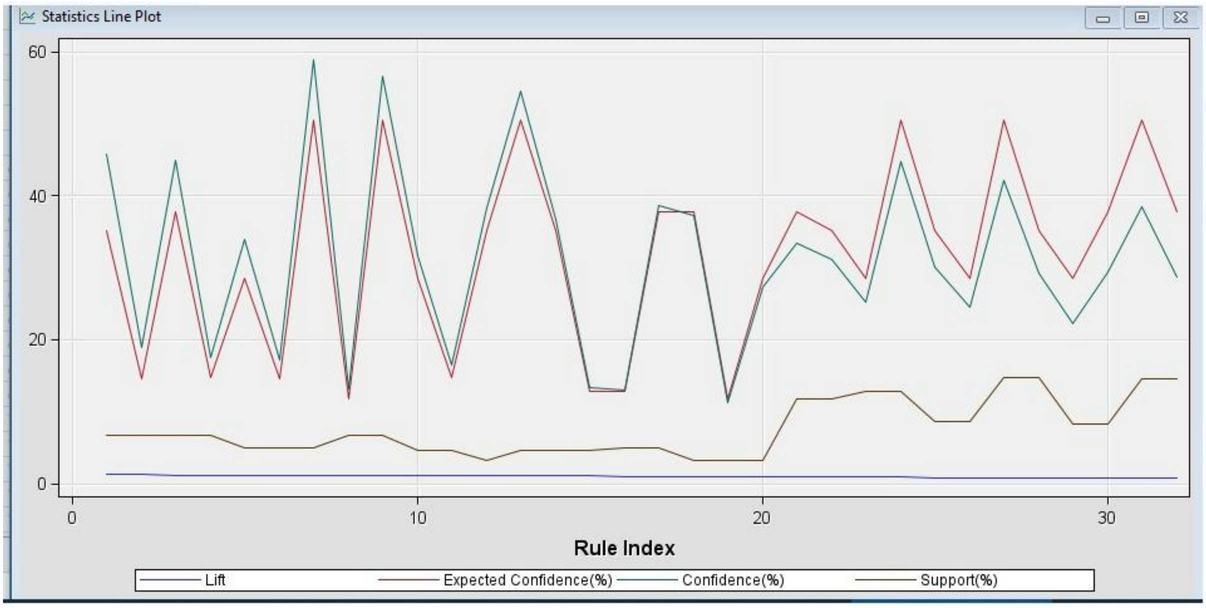
File Edit View Window

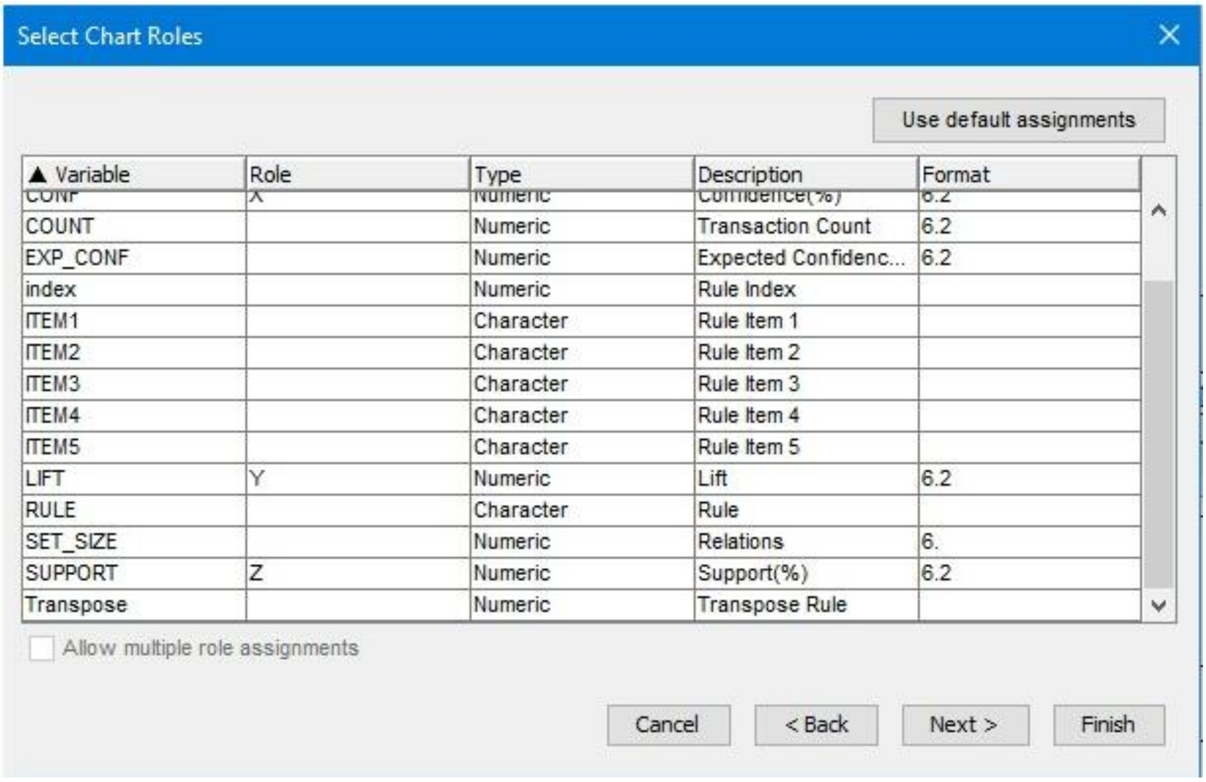
Rule Description

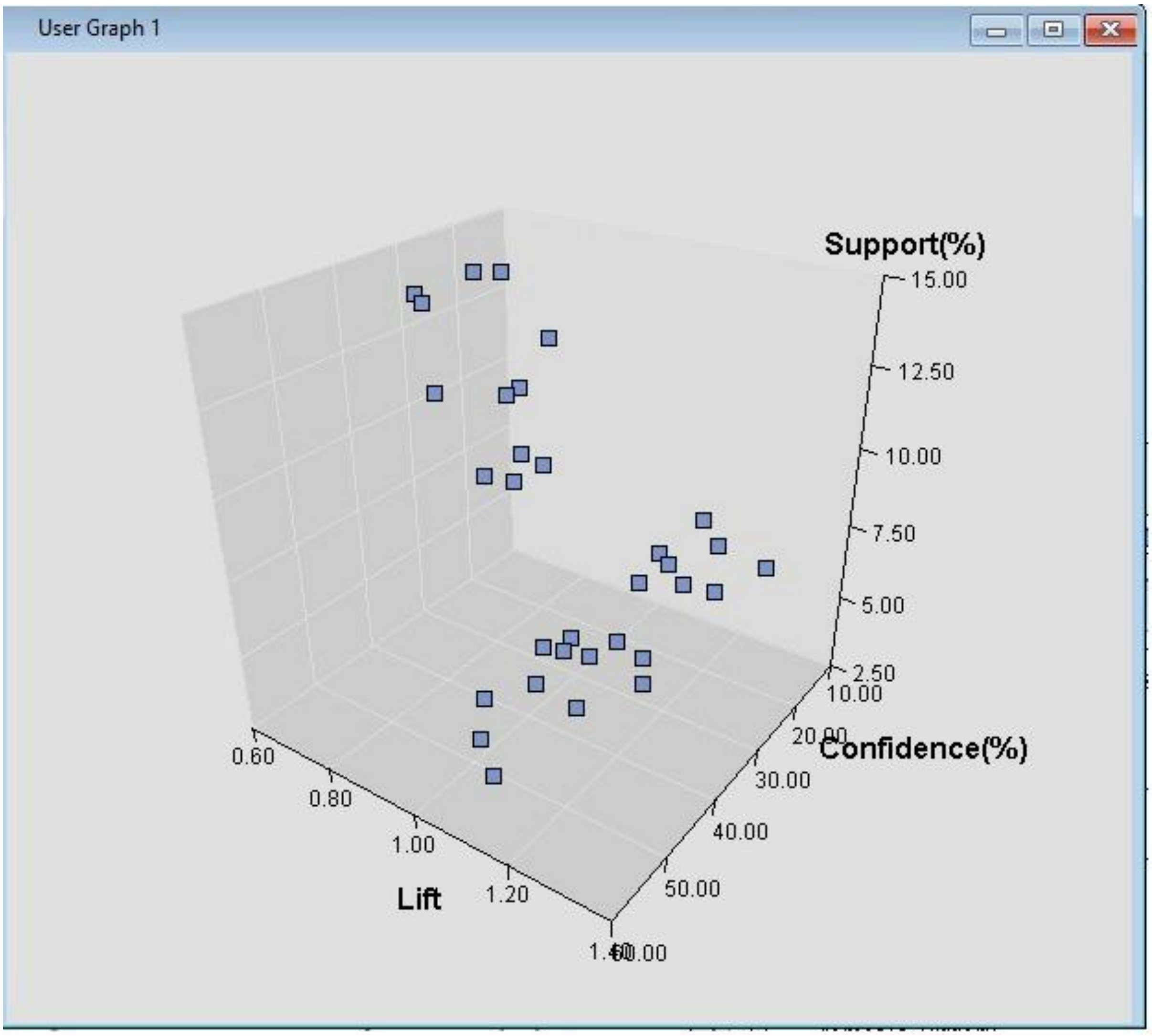
Map	Rule
RULE1	Xerox 1967 & Safco Industrial Wire Shelving ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE2	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967 & Safco Industrial Wire Shelving
RULE3	Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Safco Industrial Wire Shelving
RULE4	Safco Industrial Wire Shelving ==> Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE5	Xerox 1967 & Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE6	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967 & Safco Industrial Wire Shelving
RULE7	Safco Industrial Wire Shelving & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967
RULE8	Xerox 1967 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving
RULE9	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving ==> Xerox 1967
RULE10	Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE11	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967 & SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE12	Safco Industrial Wire Shelving & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x...
RULE13	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967
RULE14	Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE15	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE16	Safco Industrial Wire Shelving ==> Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE17	Xerox 1967 & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire Shelving
RULE18	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire ...
RULE19	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire ...
RULE20	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h & Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhon...
RULE21	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Safco Industrial Wire Shelving
RULE22	Safco Industrial Wire Shelving ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE23	Xerox 1967 ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE24	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Xerox 1967
RULE25	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE26	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE27	SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ==> Xerox 1967
RULE28	Xerox 1967 ==> SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h
RULE29	Safco Industrial Wire Shelving ==> Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5
RULE30	Anker 36W 4-Port USB Wall Charger Travel Power Adapter for iPhone 5s 5c 5 ==> Safco Industrial Wire Shelving
RULE31	Safco Industrial Wire Shelving ==> Xerox 1967
RULE32	Xerox 1967 ==> Safco Industrial Wire Shelving




Rule Matrix
Left Hand of Rule
Right Hand of Rule
11.23
58.87


Statistics Line Plot
60
40
20
0
0
10
20
30
Rule Index
Lift
Expected Confidence(%)
Confidence(%)
Support(%)


Select Chart Roles
Use default assignments
▲ Variable | Role | Type | Description | Format
CONF | X | Numeric | Confidence(%) | 6.2
COUNT | | Numeric | Transaction Count | 6.2
EXP_CONF | | Numeric | Expected Confidenc... | 6.2
index | | Numeric | Rule Index |
ITEM1 | | Character | Rule Item 1 |
ITEM2 | | Character | Rule Item 2 |
ITEM3 | | Character | Rule Item 3 |
ITEM4 | | Character | Rule Item 4 |
ITEM5 | | Character | Rule Item 5 |
LIFT | Y | Numeric | Lift | 6.2
RULE | | Character | Rule |
SET_SIZE | | Numeric | Relations | 6.
SUPPORT | Z | Numeric | Support(%) | 6.2
Transpose | | Numeric | Transpose Rule |
Allow multiple role assignments
Cancel
< Back
Next >
Finish

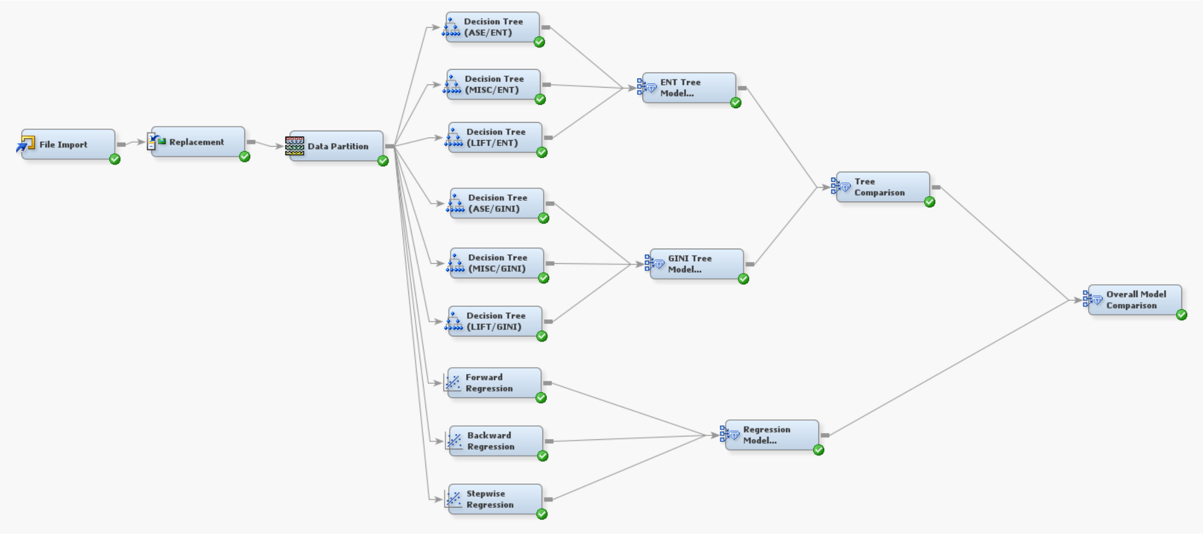


User Graph 1
Support(%)
15.00
12.50
10.00
7.50
5.00
2.50
10.00
20.00
Confidence(%)
30.00
40.00
50.00
60.00
0.60
0.80
1.00
Lift
1.20
1.40

<u>Objective 5:</u> To determine what factor influences customer's satisfaction.

Customer's satisfaction directly impacts customer loyalty and repeat purchase. It helps companies understand how well they are meeting customer's expectations and identify areas of improvement.By analyzing customer satisfaction level, companies can make strategic decisions to enhance their product and service. Finally, increase the profitability.

Overview of the EM diagram



Data Preparation

1. Replacment

Columns: Label Mining Basic Statistics

Name	Use	Limit Method	Replacement Lower Limit	Replacement Upper Limit	Replace Method
Days_to_Ship_A	Default	Default	.	.	Default
Days_to_Ship_S	Default	Default	.	.	Default
Discount	Default	Default	.	.	Default
Number_of_Rec	Default	Default	.	.	Default
Postal_Code	Default	Default	.	.	Default
Profit	Default	Default	.	.	Default
Profit_Ratio	Default	Default	.	.	Default
Profit_per_Orde	Default	Default	.	.	Default
Quantity	Default	Default	.	.	Default
Sales	Default	User Specified	1	.	Default
Sales_Forecast	Default	Default	.	.	Default
Sales_per_Cust	Default	Default	.	.	Default

-prevent missing value

2.Data Partition

The validation dataset will be occupied 50%, and the training dataset will be occupied 50%.

General	
Node ID	Part
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Partitioning Method	Default
Random Seed	12345
Data Set Allocations	
Training	50.0
Validation	50.0
Test	0.0
Report	
Interval Targets	Yes
Class Targets	Yes

Methodology

We will split the data into a (50%)training set and a (50%)testing set.

The target variable: **Satisfaction**

The Input variables {**Ship Status**, **Discount**, **Category**, **Sales per Customer**}

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Order_ID	ID	Nominal	No		No	.	.
Discount	Input	Interval	No		No	.	.
Ship_Status	Input	Nominal	No		No	.	.
Sales_per_Custo	Input	Interval	No		No	.	.
Category	Input	Nominal	No		No	.	.
Region	Rejected	Nominal	No		No	.	.
Sales	Rejected	Interval	No		No	.	.
Profit_per_Orde	Rejected	Interval	No		No	.	.
Quantity	Rejected	Interval	No		No	.	.
Ship_Date	Rejected	Nominal	No		No	.	.
State	Rejected	Nominal	No		No	.	.
Ship_Mode	Rejected	Nominal	No		No	.	.
Sales_Forecast	Rejected	Interval	No		No	.	.
Sub_Category	Rejected	Nominal	No		No	.	.
City	Rejected	Nominal	No		No	.	.
Days_to_Ship_S	Rejected	Interval	No		No	.	.
Number_of_Rec	Rejected	Interval	No		No	.	.
Customer_Name	Rejected	Nominal	No		No	.	.
Country	Rejected	Nominal	No		No	.	.
Days_to_Ship_A	Rejected	Interval	No		No	.	.
Profit	Rejected	Interval	No		No	.	.
Product_Name	Rejected	Nominal	No		No	.	.
Profit_Ratio	Rejected	Interval	No		No	.	.
Order_Date	Rejected	Nominal	No		No	.	.
Postal_Code	Rejected	Interval	No		No	.	.
Segment	Segment	Nominal	No		No	.	.
Satisfaction	Target	Ordinal	No		No	.	.

1. Decision Tree

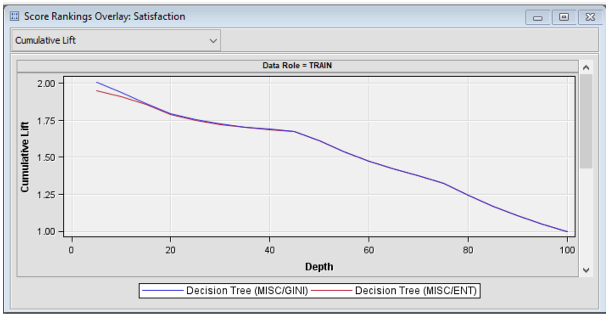
We will use those input variables to build 2 tree to determine those variable importance towards satisfaction. Then, We will use those splitting criterion (Entropy, and Gini) for a decision tree as impurity measures to evaluate a split at each node. Besides, (misclassification, Average Square Error and Lift) is the assessment measure.

General	
Node ID	Tree
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Interactive	...
Import Tree Model	No
Tree Model Data Set	...
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2

Property	Value
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Average Square Error
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Importa	
Observation Based Importa	No

Tree Model Comparison

Initially, we tested two different nominal splitting rules (Entropy, and Gini) for a decision tree with a maximum of two splits, and the Both Entropy and Gini yielded the best results; subsequently, (MISC/GINI)(MISC/ENT) exhibited the lowest validation misclassification rate. Additionally, applying the decision tree after imputation improves the prediction in this instance. Below are the context and results of this decision tree:



Output

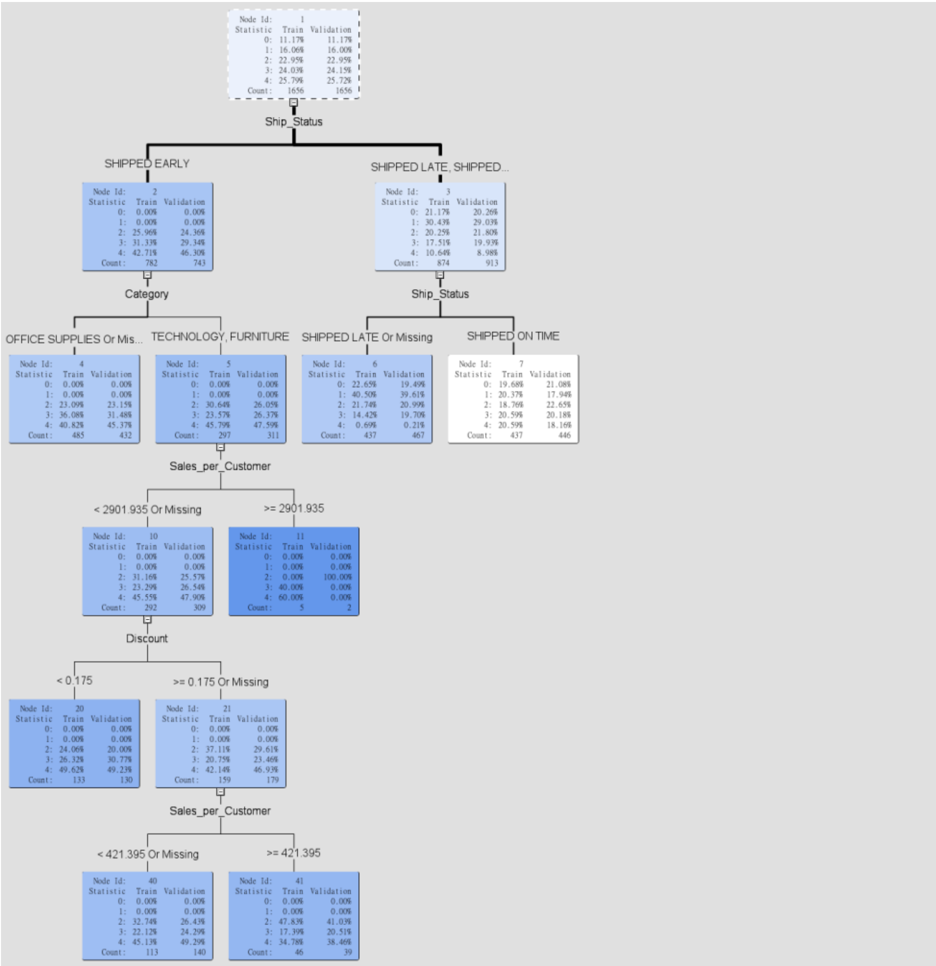
```
*------------------------------------------------------------*
User:                tkleung86
Date:                November 29, 2023
Time:                21:37:49
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary


          Measurement    Frequency
 Role        Level          Count

TARGET      ORDINAL           1
```

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error
Y	MdlComp2	Tree5	Decision Tree (MISC/GINI)	Satisfaction	Satisfaction	0.625604	1656	0.633454	0.993135	1164.088	0.14059	0.374954	8280	6624	1656	0.625604	
	MdlComp	Tree2	Decision Tree (MISC/ENT)	Satisfaction	Satisfaction	0.625604	1656	0.633454	0.993135	1164.373	0.140625	0.375	8280	6624	1656	0.625604	

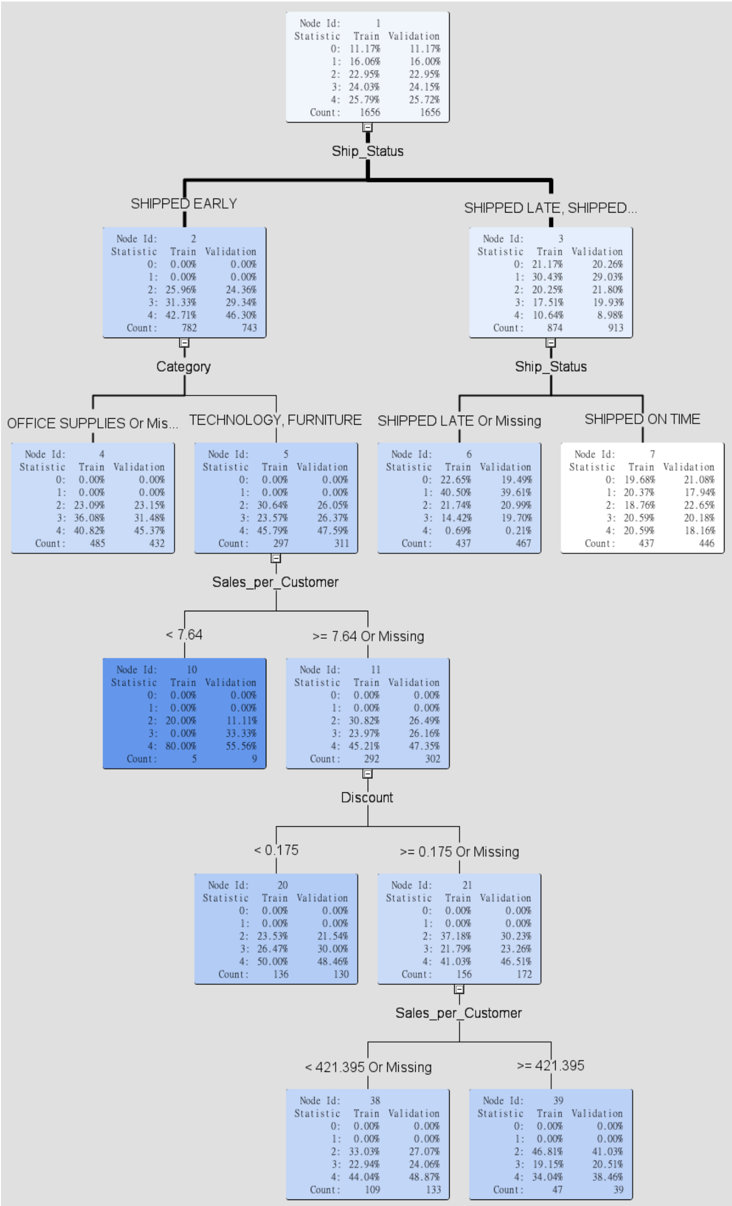
Results:

(MISC/GINI):



Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Ship_Status		2	1.0000	1.0000	1.0000
Category	Category	1	0.1858	0.0000	0.0000
Sales_per_Customer		2	0.1217	0.1078	0.8854
Discount	Discount	1	0.1210	0.0679	0.5608

(MISC/ENT):



Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Ship_Status		2	1.0000	1.0000	1.0000
Category	Category	1	0.1858	0.0000	0.0000
Discount	Discount	1	0.1285	0.0341	0.2651
Sales_per_Customer		2	0.1233	0.1025	0.8311

Logistic Regression

We will use Logistic Regression to iteratively remove independent variables which are not important to the target variable from a model to simplify it and improve its interpretability and predictive performance. The resulting model will only include the predictors that are statistically significant and contribute meaningfully to the predict the target variable(satisfaction). It has 3 selection methods, which are forward, backward, and stepwise.

Selection Options

Property	Value
Sequential Order	No
Entry Significance Level	1.0
Stay Significance Level	0.5
Start Variable Number	0
Stop Variable Number	0
Force Candidate Effects	0
Hierarchy Effects	Class
Moving Effect Rule	None
Maximum Number of Steps	20

Results:

Those 3 selection modeling are the same misclassification in these three models.

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Y	Reg	Reg	Forward Re...	Satisfaction	Satisfaction	0.619565
	Reg3	Reg3	Stepwise R...	Satisfaction	Satisfaction	0.619565
	Reg2	Reg2	Backward ...	Satisfaction	Satisfaction	0.619565

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	4	1	-1.0572	0.0562	354.17	<.0001		0.347
Intercept	3	1	-0.00725	0.0491	0.02	0.8828		0.993
Intercept	2	1	0.9828	0.0552	316.96	<.0001		2.672
Intercept	1	1	2.0733	0.0780	706.42	<.0001		7.951

All model comparison

After comparing the Average Squared error of the two final models, which are Decision tree (Gini), Logistic regression (Forwardward). The Decision tree (MISC/GINI) performs the best, which is the lowest Average Squared error.

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error
Y	MdlComp3	Tree5	Decision Tree (MISC/GINI)	Satisfaction	Satisfaction	0.141784
	MdlComp4	Reg	Forward Regression	Satisfaction	Satisfaction	0.144864