Coffee Chain Sales Analysis

Business Understanding:

The "Coffee Sales Data" dataset serves as a valuable resource for evaluating the performance of a coffee chain across various states in the United States. This dataset offers insights into geographical performance variations, seasonal trends, product preferences, competitive analysis, operational efficiency and marketing expenses. By harnessing the information within this dataset, the coffee chain can make informed decisions to enhance its operations, optimize marketing strategies, and improve customer satisfaction, ultimately driving increased sales and profitability.

Data Understanding:

This dataset comprises 21 columns with a total of 1,062 records. Among the key attributes of interest within this dataset are "Product Type," "Product," "Profit," "Cost of Goods Sold," "Sales," "State," and "Target Margin." The data is aimed at uncovering and analyzing trends pertaining to profit, sales, cost of goods sold (COGS), and margin targets across various states within the United States of America. These states are often categorized into different regions, such as the East, Central, West, and North, thereby facilitating a comprehensive regional analysis of economic and sales performance.

Research Questions:

1. How do variations in the "Product type" affect "Sales" in different states?

2. What are the key drivers of "Total Expenses," and how do they vary across different "Market" segments?

3. What is the impact of "Market size" on profit within different states?

Variable Importance:

For my research question 1, it was discovered that "Sales" serves as the central variable directly impacting coffee sales performance. The specific "Product type" emerged as a crucial determinant, illustrating the profound influence of product category on sales figures. Additionally, the geographical factor, "State," played a significant role, revealing regional nuances and demographics as key contributors to variations in sales outcomes.

For my research question 2, In our analysis, we tried to uncover the key drivers of "Total Expenses" and their variations across different "Market" segments. Additionally, we tried to uncover how "Market Size" demonstrates its influence, with larger and more competitive markets requiring higher expenses compared to other market segment.

For my research question 3, we explored two distinct market types exerted their influence on the profitability of coffee sales across different states. This question aimed to uncover the

relationship between market categorizations, such as "Major" and "Small," and the resulting impact on profits.

Multicollinearity:

For Research Question 1, Correlation measures the statistical relationship between two variables and can range from -1 to 1, with 1 indicating a perfect positive correlation, 0 indicating no correlation, and -1 indicating a perfect negative correlation.

The chart suggests that the financial factors like "Margin," "Target_sales," "Cogs," and "Target_margin" have the strongest positive correlations with the key performance metric, while other factors like "Area_Code" and "Inventory_Margin" have much weaker correlations.

For Research question 2, the chart indicates that "Marketing" is the factor with the strongest positive correlation with the key performance metric, followed by "Cogs," "Target_cogs," and "Sales." The financial factors, such as "Margin," "Target_margin," "Inventory_Margin," "Target_profit," and "Profit," have positive correlations but are relatively weaker in comparison. "Area_Code" and "DifferenceBetweenActualandTarget" have the weakest correlations with the key performance metric.

For my research question3, It suggests that factors related to profitability, such as "Margin," "Target_profit," and "Target_margin," have the strongest positive correlations with better performance. Other factors, like "Sales" and "Closing the gap between actual and target performance," also have positive correlations but are slightly weaker. On the other hand, factors like "Inventory_Margin" have a negative correlation with performance, indicating that lower inventory margins are associated with better performance.

Skewed Variables:

For research question 1,Several variables, including "Cogs," "Margin," "Marketing," "Target_cogs," "Target_margin," "Target_profit," "Target_sales," and "Total_expenses," exhibit positive skewness, "DifferenceBetweenActualandTarget," "Profit," and "Area_Code" have skewness values close to zero, indicating that their distributions are nearly symmetric and "Inventory_Margin" is the only variable with strong negative skewness, indicating a left-skewed distribution.

For research question 2, Several variables such as "Cogs," "Margin," "Marketing," "Target_cogs," "Target_margin," "Target_profit," "Target_sales," and "Total_expenses." Have a positive skewness, indicating that they have right-skewed distributions. For example, "Sales" has a relatively high positive skewness value, indicating that there are significant instances of exceptionally high sales figures within the dataset, "DifferenceBetweenActualandTarget" exhibits negative skewness, implying a left-skewed distribution and Some variables, such as "Area_Code," "Profit," and "Target_profit," have skewness values close to zero.

For research question 3, Variables such as "Cogs," "Margin," "Marketing," "Target_cogs," "Target_margin," "Target_profit," "Target_sales," and "Total_expenses" all have positive skewness values, suggesting that these variables have outliers with high values that are pulling

the distribution to the right, Some variables, such as "DifferenceBetweenActualandTarget," "Profit," and "Area_Code," have skewness values close to zero and "Inventory_Margin" is the only variable with strong negative skewness, indicating a left-skewed distribution.

Data Quality:

As mentioned under the appendix, the dataset used in this analysis has already undergone a rigorous data cleaning process to ensure high data quality and reliability. Every effort has been made to address missing values, outliers, and inconsistencies, resulting in a clean and complete dataset. There were no missing values in any of the variables, and all identified outliers and anomalies have been appropriately handled. The data is highly accurate and consistent, with no known errors or inconsistencies that could impact the analysis.

Appendix:

Research Question 1:

Class Variable-Summary Statistics:

Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Market	INPUT	4	0	Central	31.64	West	31.64
TRAIN	Market_size	INPUT	2	0	Small Market	59.89	Major Market	40.11
TRAIN	Product	INPUT	13	0	Caffe Mocha	11.30	Colombian	11.30
TRAIN	Product_line	INPUT	2	0	Beans	52.54	Leaves	47.46
TRAIN	Product_type	INPUT	4	0	Espresso	27.68	Coffee	24.86
TRAIN	State	INPUT	20	0	California	6.78	Utah	6.78
TRAIN	Type	INPUT	2	0	Regular	56.50	Decaf	43.50

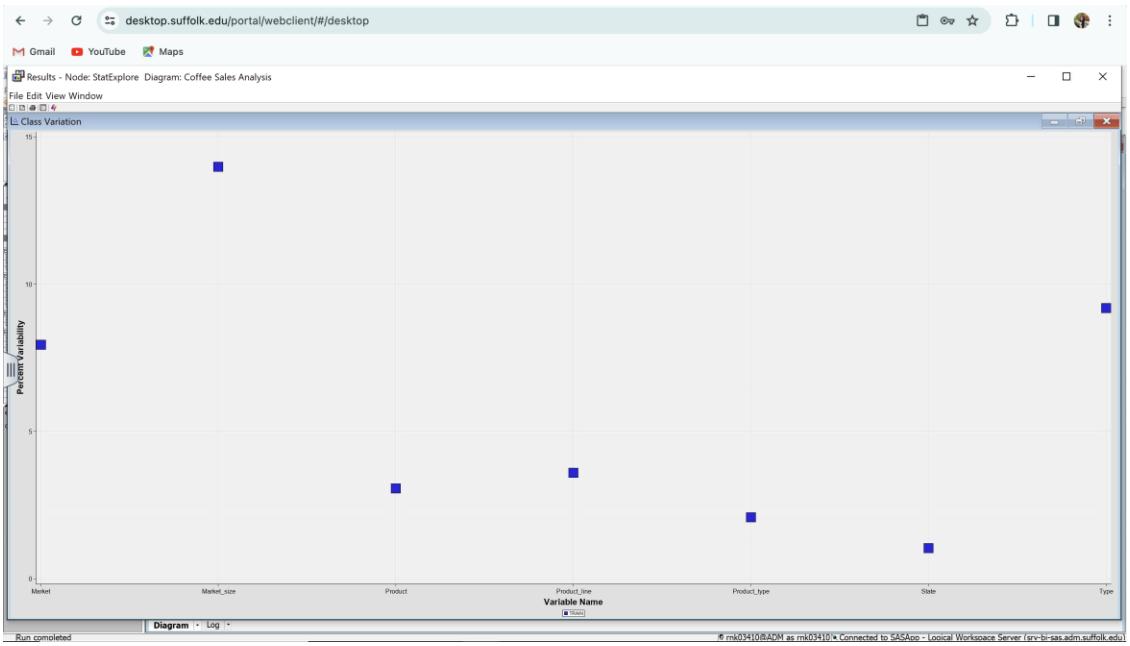
Interval Variables Summary Statistics:

Interval Variable Summary Statistics
(maximum 500 observations printed)

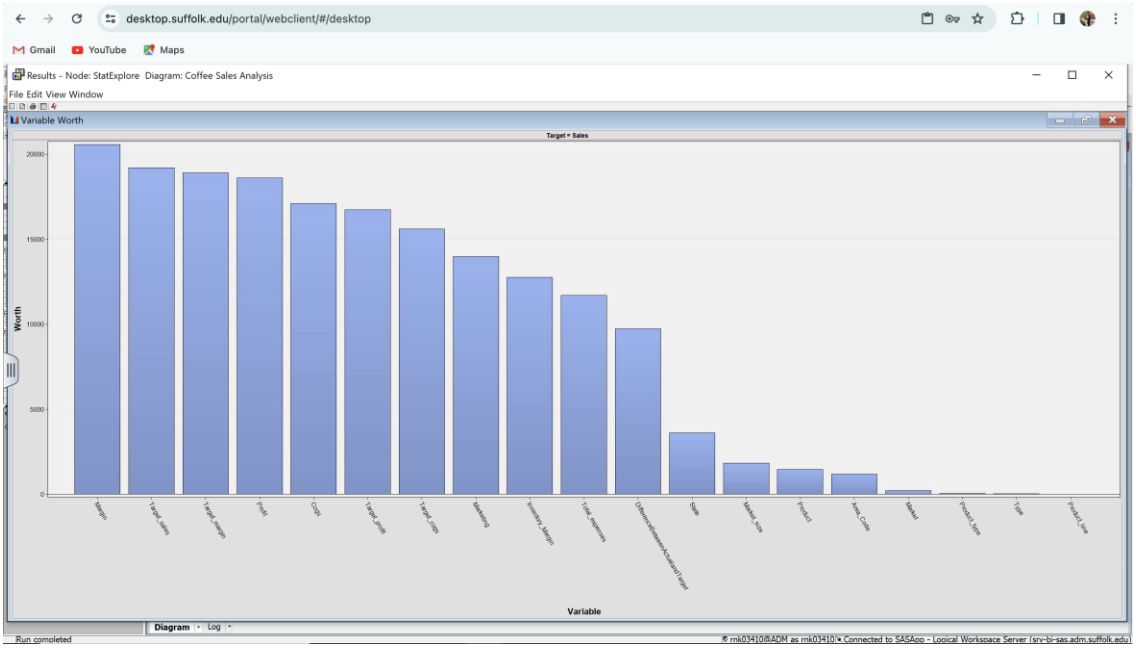
Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Profit	INPUT	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	INPUT	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479
Sales	TARGET	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726

Class Variation Plot:



Variable Worth:



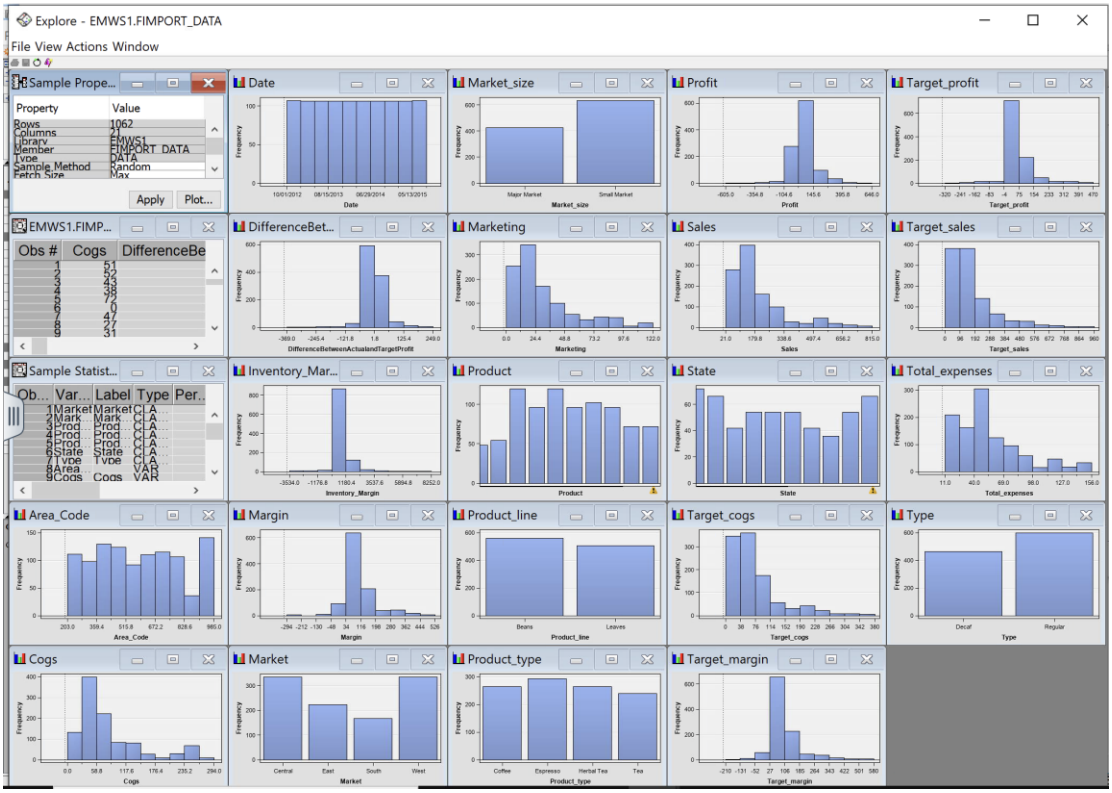
Correlation:

Correlation Statistics
(maximum 500 observations printed)

Data Role=TRAIN Type=PEARSON Target=Sales

Input	Correlation
Margin	0.93943
Target_sales	0.93451
Target_margin	0.90288
Cogs	0.88847
Target_cogs	0.84654
Profit	0.79993
Target_profit	0.78596
Marketing	0.70693
Total_expenses	0.68236
DifferenceBetweenActualandTarget	0.43398
Inventory_Margin	0.20674
Area_Code	0.07694

Skewness among the variables:

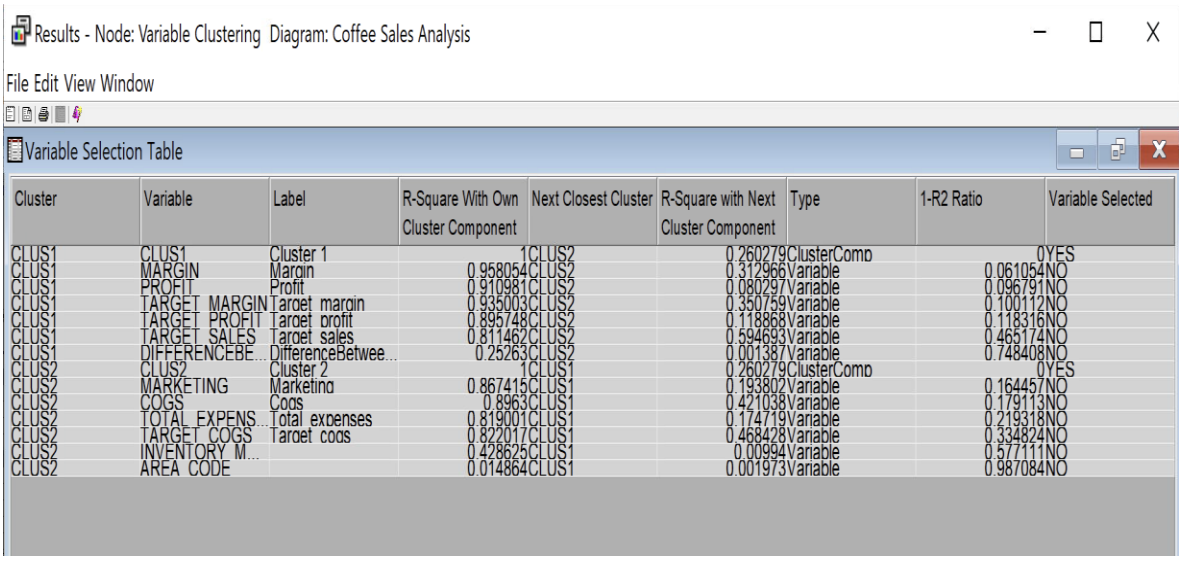


Interval Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

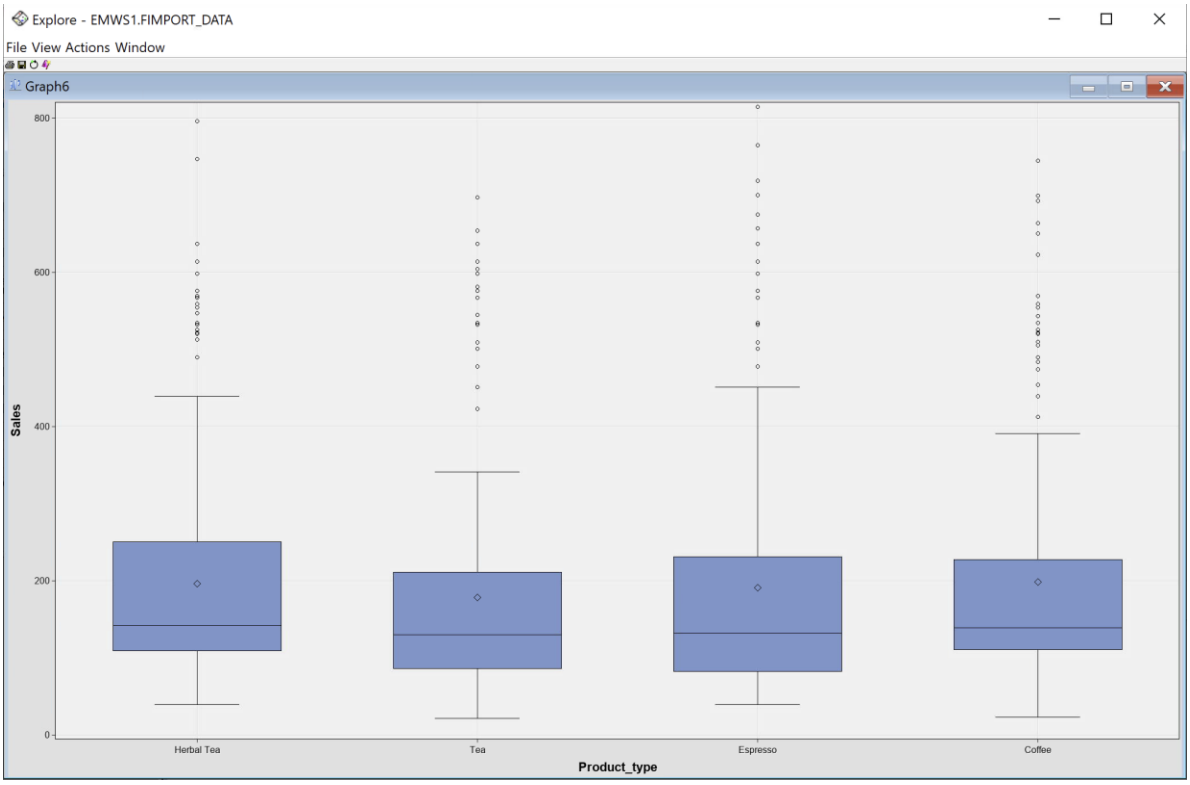
Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Profit	INPUT	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	INPUT	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479
Sales	TARGET	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726

Clustering:



Cluster	Variable	Label	R-Square With Own Cluster Component	Next Closest Cluster	R-Square with Next Cluster Component	Type	1-R2 Ratio	Variable Selected
CLUS1	CLUS1	Cluster 1	1	CLUS2	0.260279	ClusterComp	0	YES
CLUS1	MARGIN	Margin	0.958054	CLUS2	0.312966	Variable	0.061054	NO
CLUS1	PROFIT	Profit	0.910981	CLUS2	0.080297	Variable	0.096791	NO
CLUS1	TARGET_MARGIN	Target margin	0.935003	CLUS2	0.350759	Variable	0.100112	NO
CLUS1	TARGET_PROFIT	Target profit	0.895748	CLUS2	0.118868	Variable	0.118316	NO
CLUS1	TARGET_SALES	Target sales	0.811462	CLUS2	0.594693	Variable	0.465174	NO
CLUS1	DIFFERENCEBE...	DifferenceBetwee...	0.25263	CLUS2	0.001387	Variable	0.748408	NO
CLUS2	CLUS2	Cluster 2	1	CLUS1	0.260279	ClusterComp	0	YES
CLUS2	MARKETING	Marketing	0.867415	CLUS1	0.193802	Variable	0.164457	NO
CLUS2	COGS	Cogs	0.8963	CLUS1	0.421038	Variable	0.179113	NO
CLUS2	TOTAL_EXPENS...	Total expenses	0.819001	CLUS1	0.174719	Variable	0.219318	NO
CLUS2	TARGET_COGS	Target cogs	0.822017	CLUS1	0.468428	Variable	0.334824	NO
CLUS2	INVENTORY_M...		0.428625	CLUS1	0.00994	Variable	0.577111	NO
CLUS2	AREA_CODE		0.014864	CLUS1	0.001973	Variable	0.987084	NO

Box Plot Graph:



Research Question 2:

Class Variable-Summary Statistics:

Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Market	INPUT	4	0	Central	31.64	West	31.64
TRAIN	Market_size	INPUT	2	0	Small Market	59.89	Major Market	40.11
TRAIN	Product	INPUT	13	0	Caffe Mocha	11.30	Colombian	11.30
TRAIN	Product_line	INPUT	2	0	Beans	52.54	Leaves	47.46
TRAIN	Product_type	INPUT	4	0	Espresso	27.68	Coffee	24.86
TRAIN	State	INPUT	20	0	California	6.78	Utah	6.78
TRAIN	Type	INPUT	2	0	Regular	56.50	Decaf	43.50

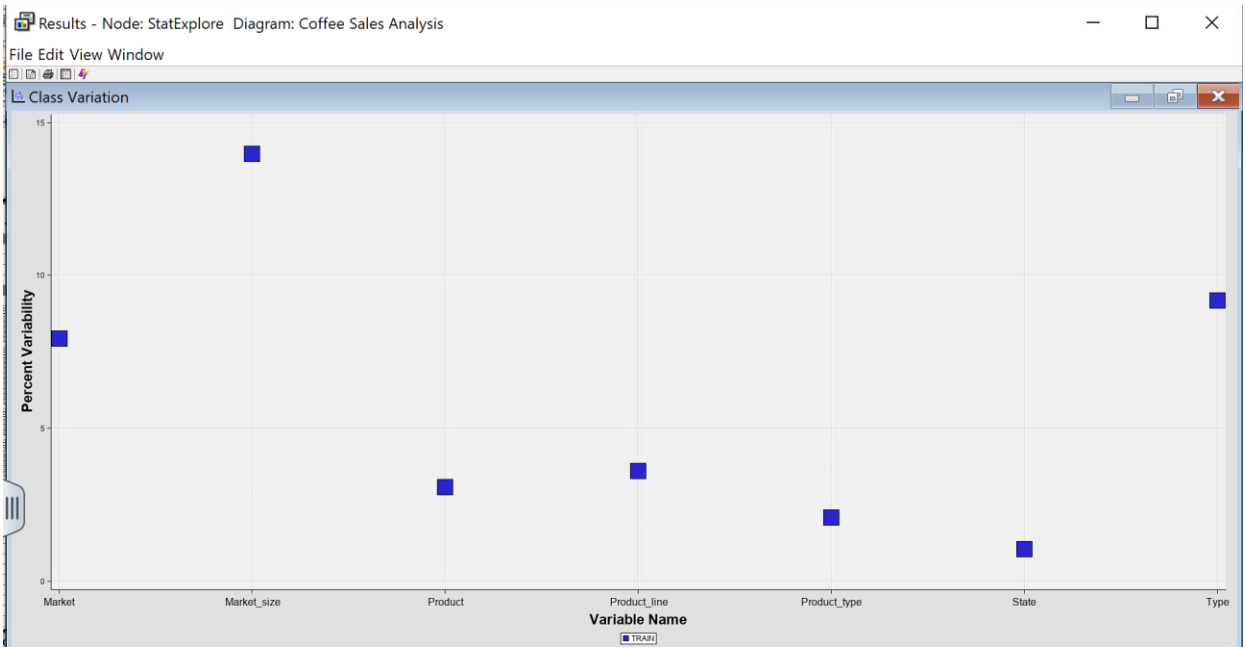
Interval Variables Summary Statistics:

Interval Variable Summary Statistics
(maximum 500 observations printed)

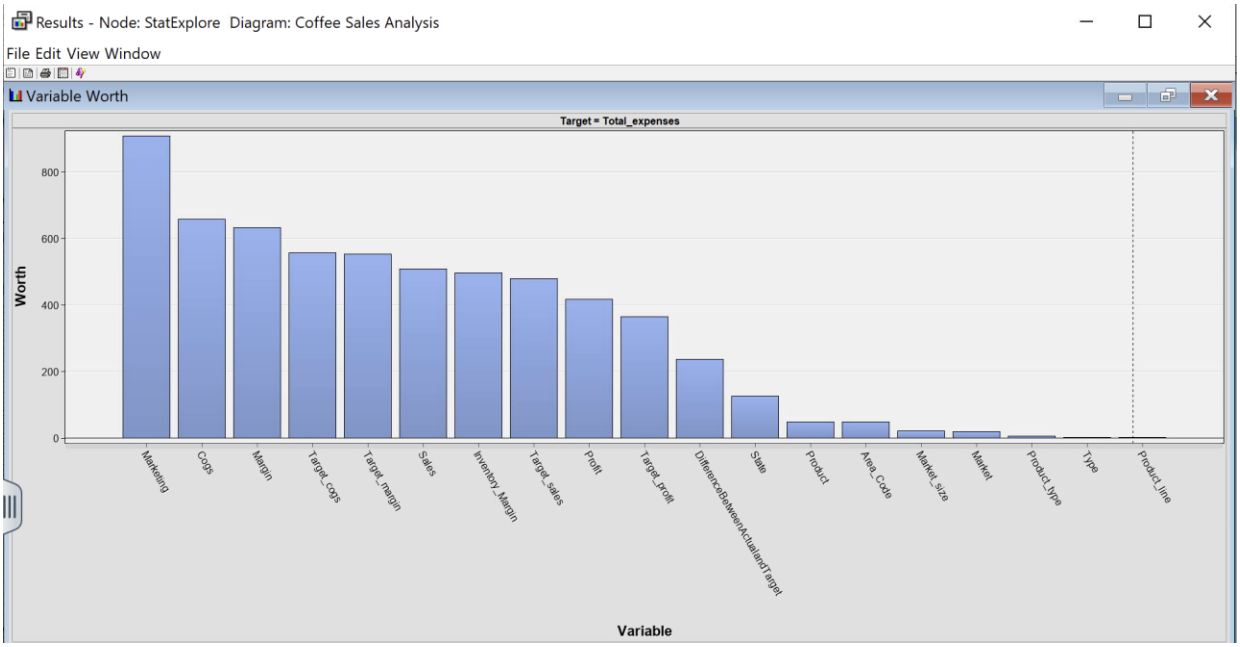
Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Profit	INPUT	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741
Sales	INPUT	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	TARGET	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479

Class Variation Plot:



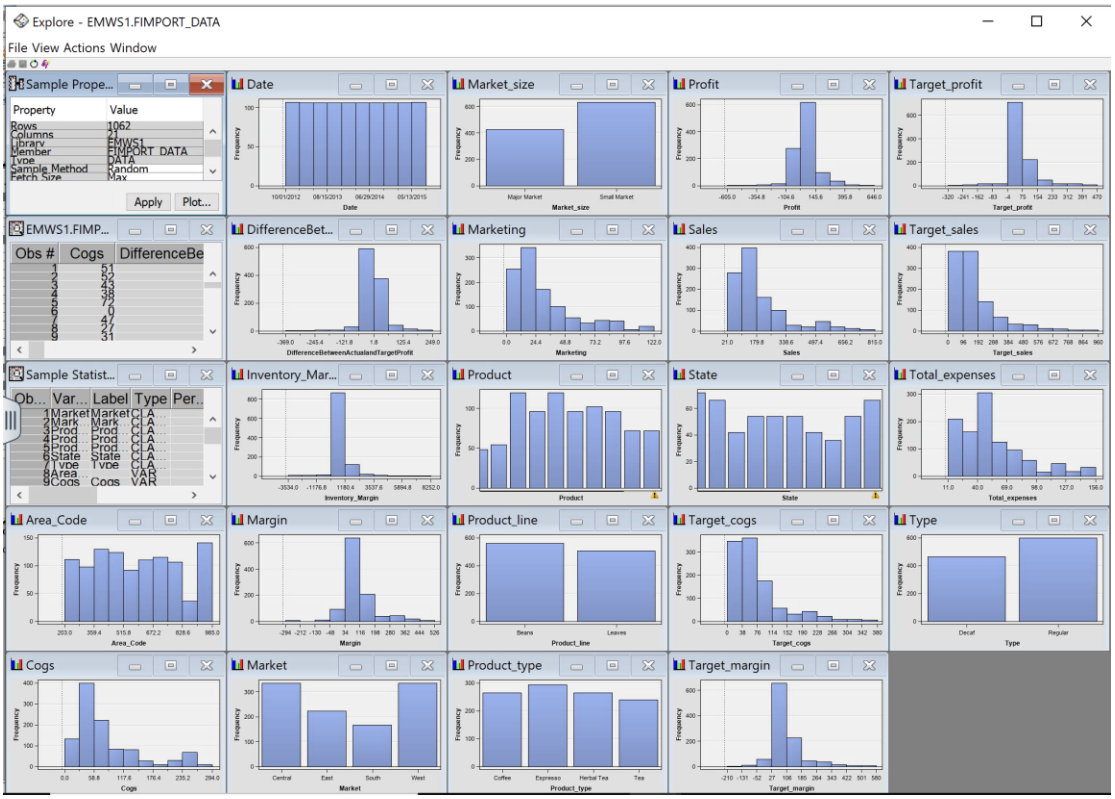
Variable Worth:



Correlation:

```
65
66    Correlation Statistics
67    (maximum 500 observations printed)
68
69    Data Role=TRAIN Type=PEARSON Target=Total_expenses
70
71    Input                                Correlation
72
73    Marketing                              0.96649
74    Cogs                                   0.78036
75    Target_cogs                            0.71981
76    Sales                                  0.68236
77    Target_sales                           0.63700
78    Margin                                 0.51454
79    Target_margin                          0.51058
80    Inventory_Margin                       0.42869
81    Target_profit                          0.22754
82    Profit                                 0.19366
83    Area_Code                              0.04424
84    DifferenceBetweenActualandTarget       0.03965
85
```

Skewness among the variables:

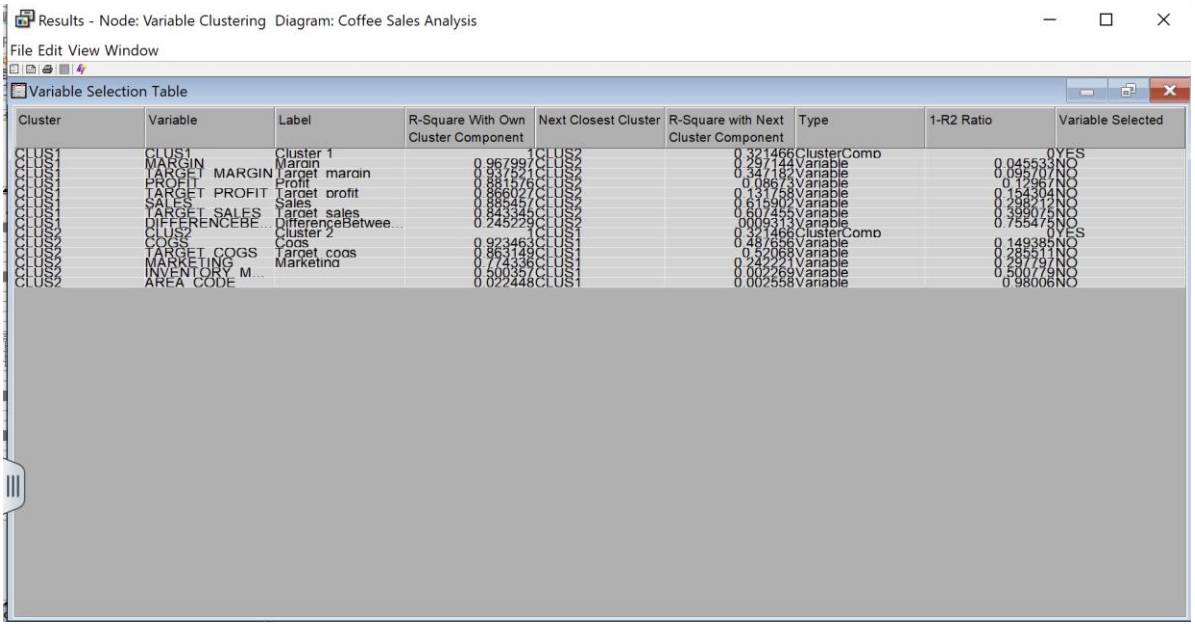


Interval Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

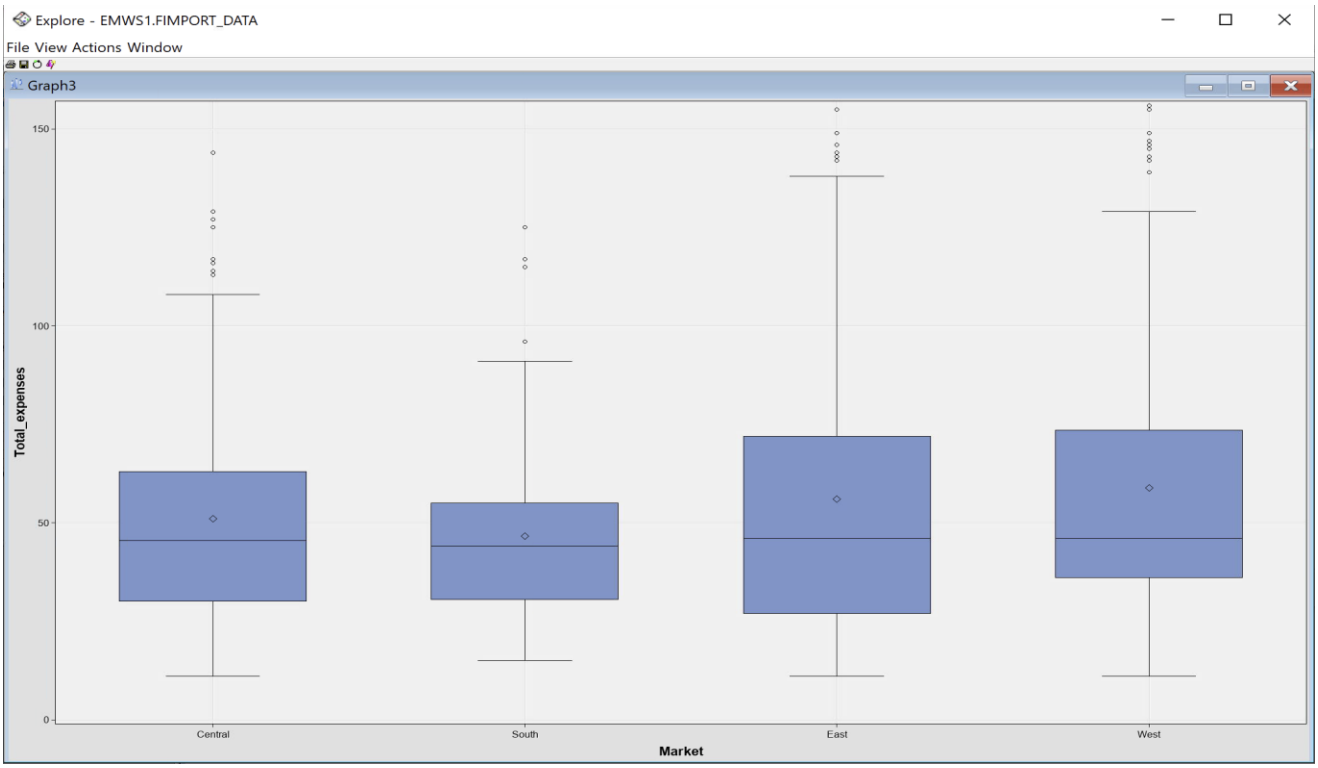
Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Profit	INPUT	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741
Sales	INPUT	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	TARGET	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479

Clustering:



Cluster	Variable	Label	R-Square With Own Cluster Component	Next Closest Cluster	R-Square with Next Cluster Component	Type	1-R2 Ratio	Variable Selected
CLUS1	CLUS1	Cluster 1	1	CLUS2	0.321466	ClusterComp		0 YES
CLUS1	MARGIN	Margin	0.967997	CLUS2	0.297144	Variable	0.045533	NO
CLUS1	TARGET_MARGIN	Target_margin	0.937521	CLUS2	0.347182	Variable	0.095707	NO
CLUS1	PROFIT	Profit	0.881576	CLUS2	0.08673	Variable	0.12967	NO
CLUS1	TARGET_PROFIT	Target_profit	0.866027	CLUS2	0.131758	Variable	0.154304	NO
CLUS1	SALES	Sales	0.885457	CLUS2	0.615902	Variable	0.298212	NO
CLUS1	TARGET_SALES	Target_sales	0.843345	CLUS2	0.607455	Variable	0.399075	NO
CLUS1	DIFFERENCEBE...	DifferenceBetwee...	0.245229	CLUS2	0.009313	Variable	0.755475	NO
CLUS2	CLUS2	Cluster 2	1	CLUS1	0.321466	ClusterComp		0 YES
CLUS2	COGS	Cogs	0.923463	CLUS1	0.487656	Variable	0.149385	NO
CLUS2	TARGET_COGS	Target_cogs	0.863149	CLUS1	0.52068	Variable	0.285511	NO
CLUS2	MARKETING	Marketing	0.774336	CLUS1	0.242221	Variable	0.297797	NO
CLUS2	INVENTORY_M...		0.500357	CLUS1	0.002269	Variable	0.500779	NO
CLUS2	AREA_CODE		0.022448	CLUS1	0.002558	Variable	0.98006	NO

Box Plot graph:



Research Question 3:

Class Variable-Summary Statistics:

Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Market	INPUT	4	0	Central	31.64	West	31.64
TRAIN	Market_size	INPUT	2	0	Small Market	59.89	Major Market	40.11
TRAIN	Product	INPUT	13	0	Caffe Mocha	11.30	Colombian	11.30
TRAIN	Product_line	INPUT	2	0	Beans	52.54	Leaves	47.46
TRAIN	Product_type	INPUT	4	0	Espresso	27.68	Coffee	24.86
TRAIN	State	INPUT	20	0	California	6.78	Utah	6.78
TRAIN	Type	INPUT	2	0	Regular	56.50	Decaf	43.50

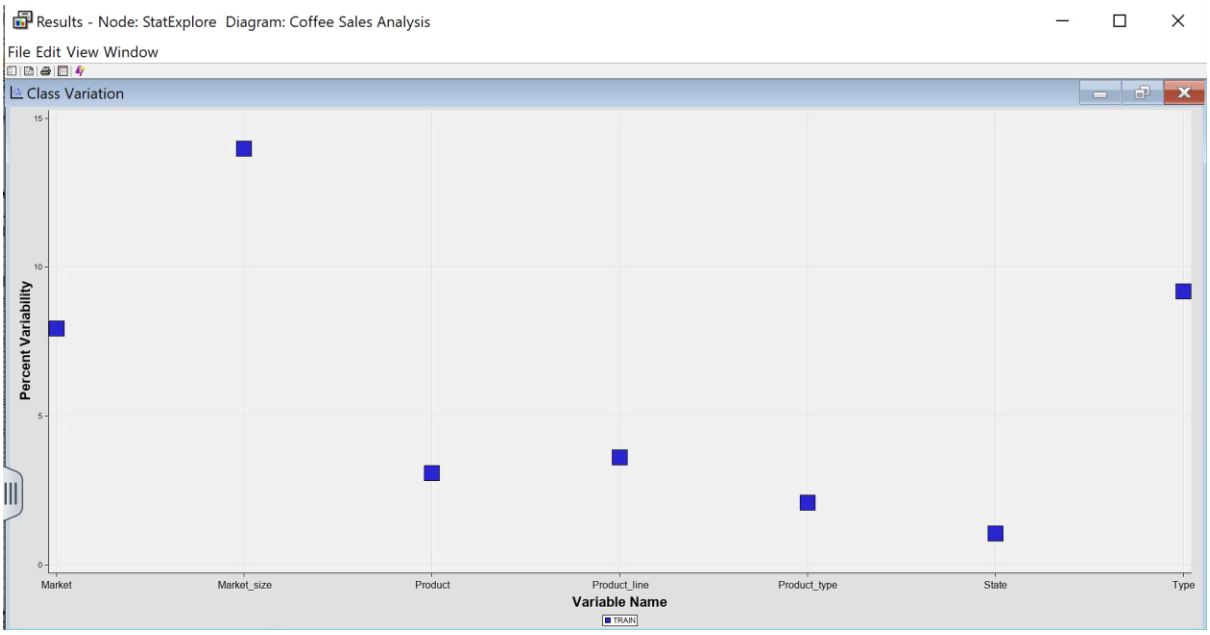
Interval Variables Summary Statistics:

Interval Variable Summary Statistics
(maximum 500 observations printed)

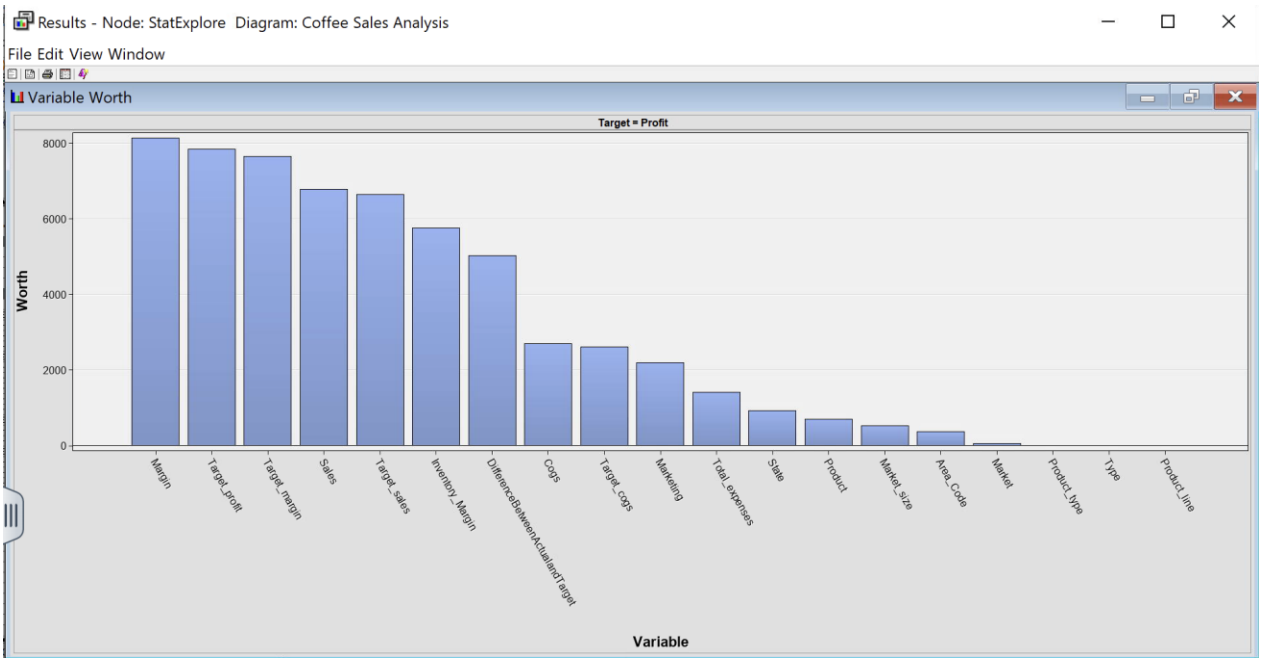
Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Sales	INPUT	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	INPUT	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479
Profit	TARGET	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741

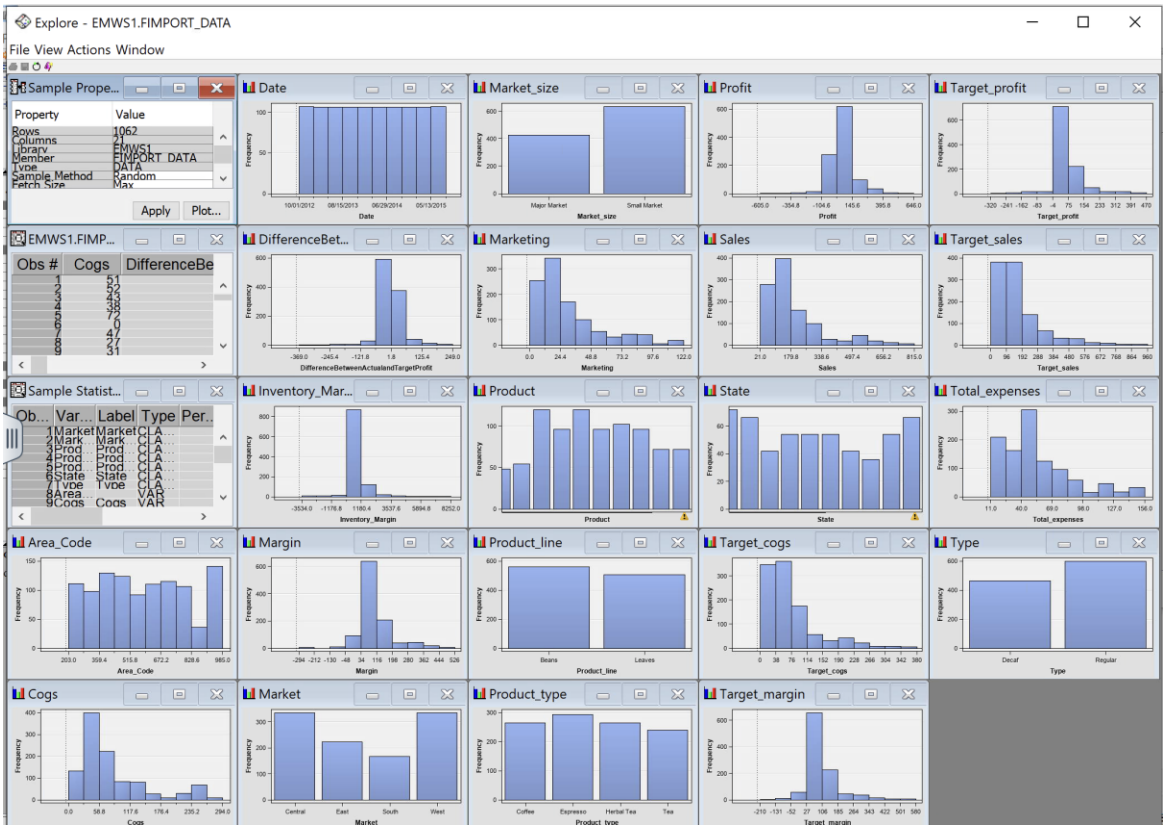
Class Variation Plot:



Variable Worth:



Skewness among the variables:



Interval Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Area_Code	INPUT	587.0301	225.2992	1062	0	203	573	985	0.07884	-0.99255
Cogs	INPUT	82.39925	64.82429	1062	0	0	57	294	1.536993	1.609019
DifferenceBetweenActualandTarget	INPUT	0.387006	44.33118	1062	0	-369	-3	249	-0.15646	11.51824
Inventory_Margin	INPUT	815.1751	916.1564	1062	0	-3534	659	8252	2.604671	21.74568
Margin	INPUT	102.4237	91.2867	1062	0	-294	73	526	1.222219	4.295716
Marketing	INPUT	30.43315	25.96345	1062	0	0	22	122	1.487507	1.637996
Sales	INPUT	191.0499	148.2703	1062	0	21	133	815	1.75649	2.76726
Target_cogs	INPUT	71.67608	65.70158	1062	0	0	50	380	1.896974	3.918186
Target_margin	INPUT	96.81733	89.46718	1062	0	-210	70	580	1.887134	5.965454
Target_profit	INPUT	60.16949	77.82487	1062	0	-320	40	470	1.083739	7.347542
Target_sales	INPUT	168.4934	145.9552	1062	0	0	120	960	2.177533	5.638313
Total_expenses	INPUT	53.83616	31.70353	1062	0	11	46	156	1.188945	1.063479
Profit	TARGET	60.5565	100.5166	1062	0	-605	39	646	0.282312	8.973741

Clustering:

Results - Node: Variable Clustering Diagram: Coffee Sales Analysis

File Edit View Window

Variable Selection Table

Cluster	Variable	Label	R-Square With Own Cluster Component	Next Closest Cluster	R-Square with Next Cluster Component	Type	1-R2 Ratio	Variable Selected
CLUS1	CLUS1	Cluster 1	1	CLUS2	0.380542	ClusterComp	0	YES
CLUS1	MARGIN	Margin	0.963305	CLUS2	0.312966	Variable	0.053411	NO
CLUS1	TARGET_MARGIN	Target margin	0.956698	CLUS2	0.350759	Variable	0.06696	NO
CLUS1	TARGET_PROFIT	Target profit	0.850966	CLUS2	0.118868	Variable	0.169139	NO
CLUS1	SALES	Sales	0.913544	CLUS2	0.619681	Variable	0.227324	NO
CLUS1	TARGET_SALES	Target sales	0.890393	CLUS2	0.594693	Variable	0.27043	NO
CLUS1	DIFFERENCEBE...	DifferenceBetwee...	0.195805	CLUS2	0.001387	Variable	0.805312	NO
CLUS2	CLUS2	Cluster 2	1	CLUS1	0.380542	ClusterComp	0	YES
CLUS2	MARKETING	Marketing	0.867415	CLUS1	0.294155	Variable	0.187839	NO
CLUS2	COGS	Cogs	0.8963	CLUS1	0.542555	Variable	0.226693	NO
CLUS2	TOTAL_EXPENS...	Total expenses	0.819001	CLUS1	0.271351	Variable	0.248404	NO
CLUS2	TARGET_COGS	Target cogs	0.822017	CLUS1	0.58415	Variable	0.427998	NO
CLUS2	INVENTORY_M...		0.428625	CLUS1	2.96E-7	Variable	0.571375	NO
CLUS2	AREA_CODE		0.014864	CLUS1	0.003012	Variable	0.988112	NO

Box Plot Graph: