

Milestone 5

Communication of Insights of Data

ZULKANAIN BIN HASAN

WQD180031

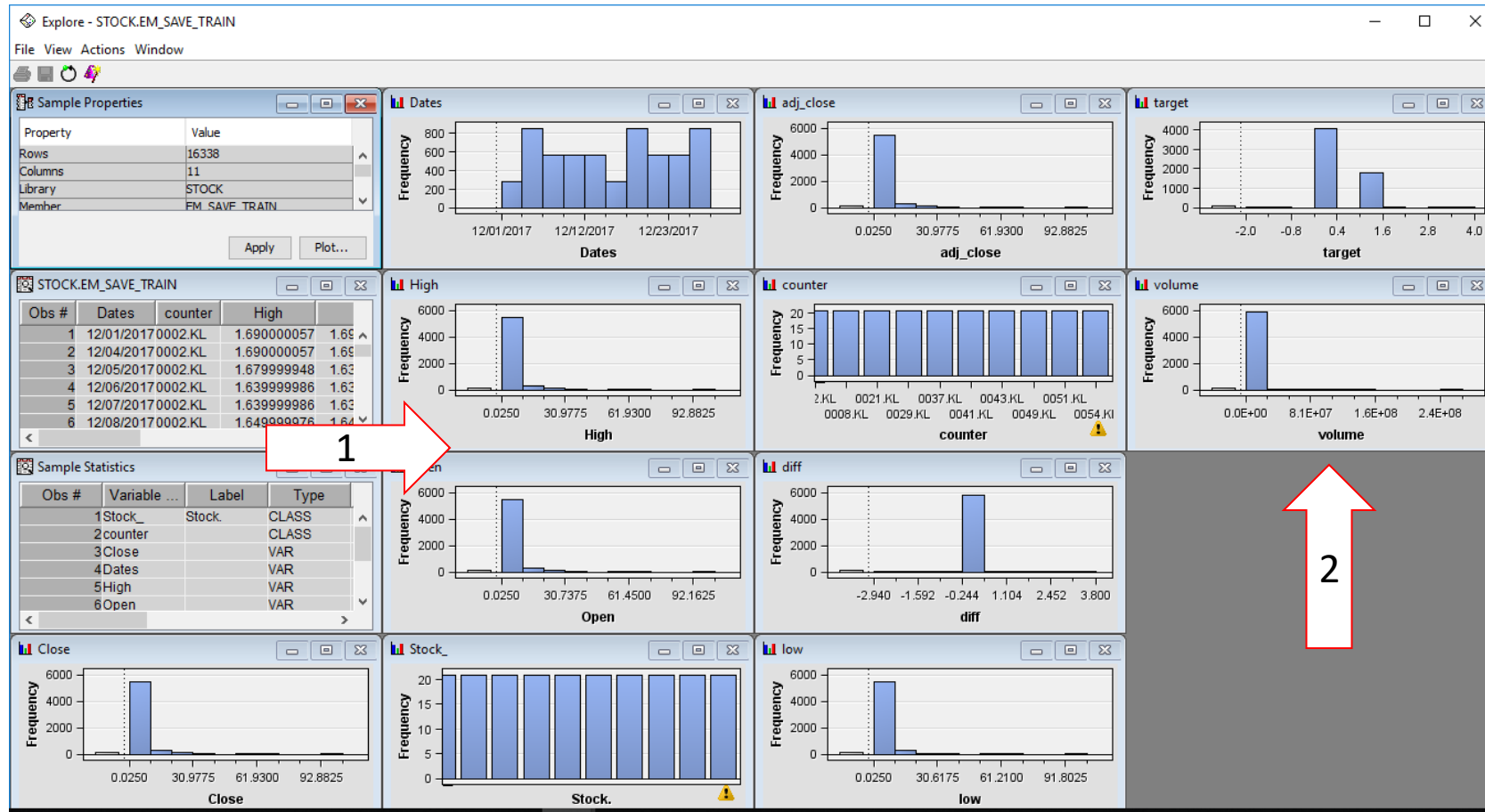
1. Create SEGMENTATION ANALYSIS diagram
2. Import data source STOCK
3. Explore STOCK variables

The screenshot displays the SAS Enterprise Miner interface. On the left, the 'Data Sources' tree shows 'stock' under 'data mining stock'. A red arrow labeled '1' points to this tree. The main workspace shows a 'segmentation analysis' diagram with a 'stock' node. A red arrow labeled '2' points to this node. Below the diagram, the 'Variables - Ids' dialog box is open, showing a table of variables. A red arrow labeled '3' points to this dialog box.

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Close	Input	Interval	No		No	.	.
Dates	Time ID	Interval	No		No	.	.
High	Input	Interval	No		No	.	.
Open	Input	Interval	No		No	.	.
Stock	Rejected	Nominal	No		No	.	.
adj_close	Input	Interval	No		No	.	.
counter	Rejected	Nominal	No		No	.	.
diff	Input	Interval	No		No	.	.
low	Input	Interval	No		No	.	.

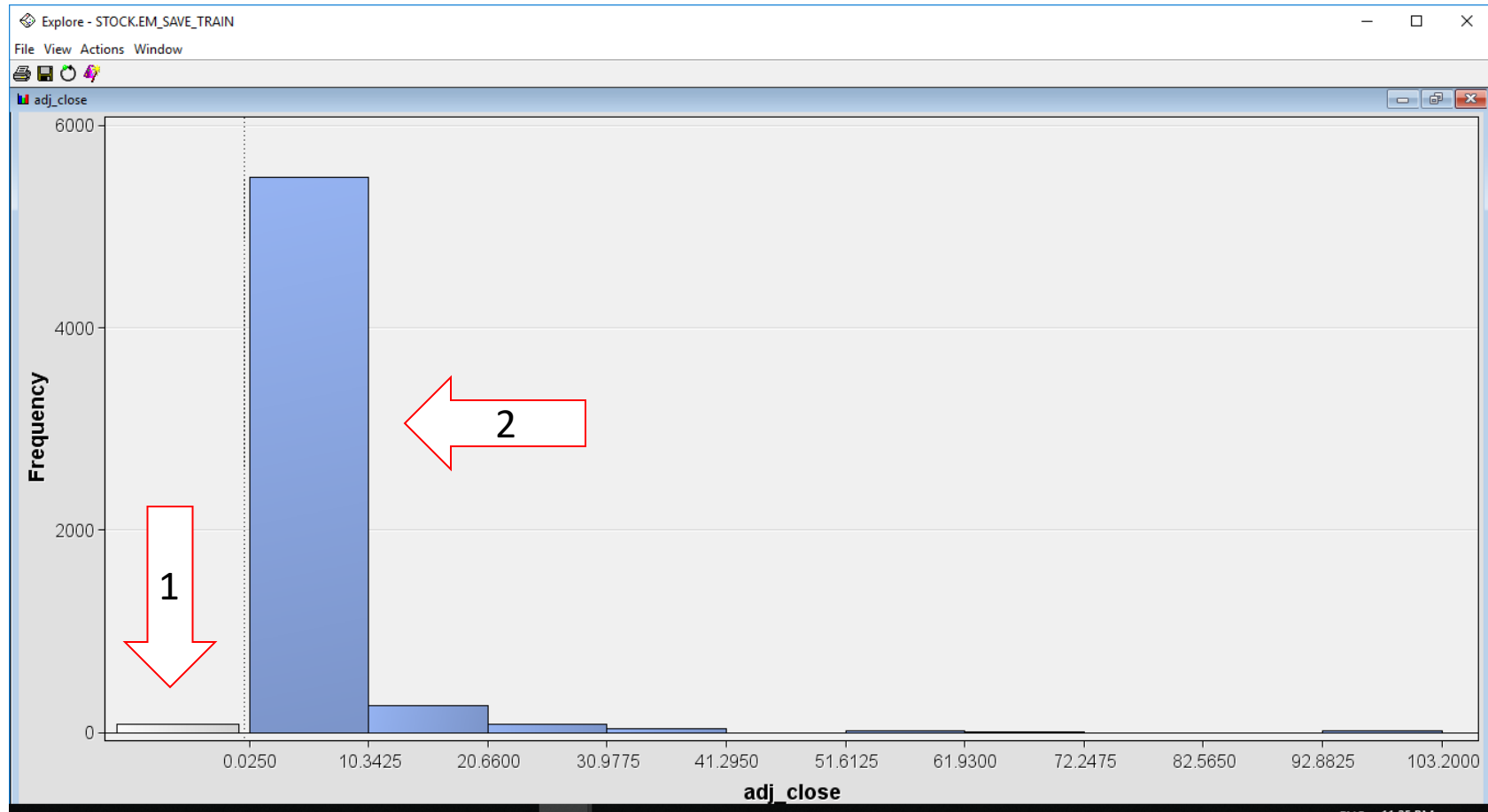
Findings:

1. Similar pattern for adj_close, High & Open
2. There is dominant value for volume

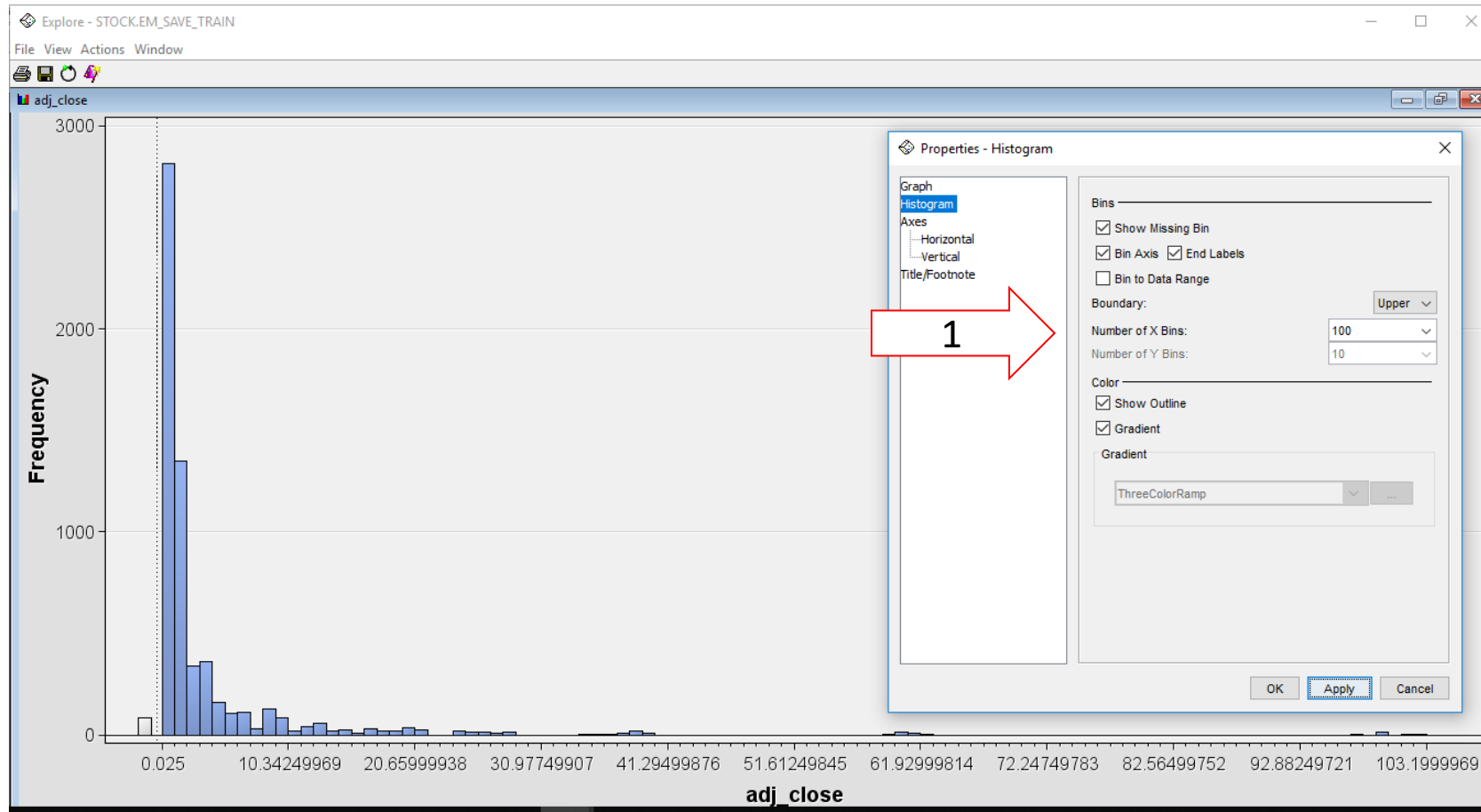


Findings:

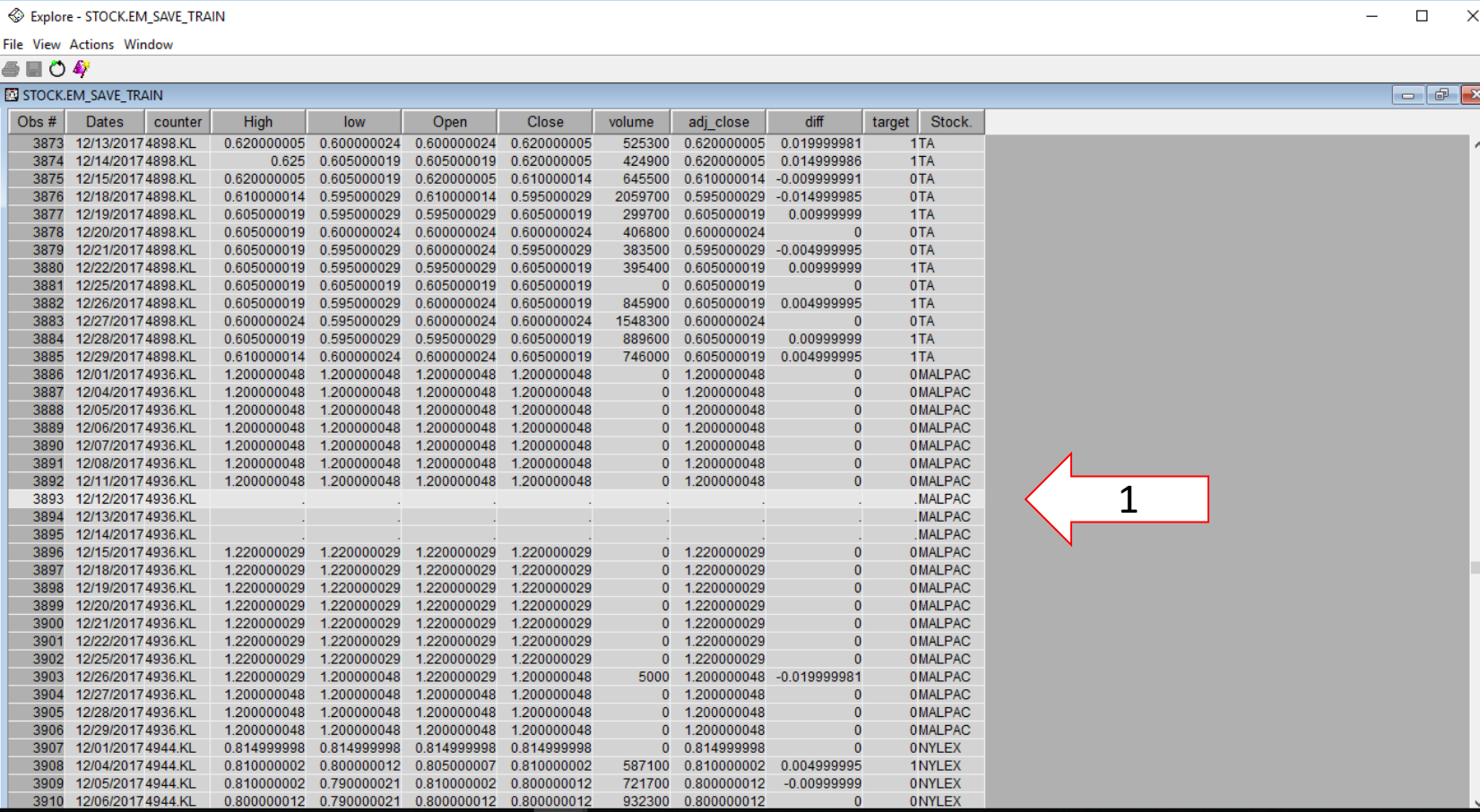
1. adj_close table shows maximum value between 0.0250 to 10.3425
2. There is missing value



1. Changed Histogram properties to 100 number of x Bins



1. Explore TRAIN data and found there is missing value



Explore - STOCK.EM_SAVE_TRAIN

File View Actions Window

STOCK.EM_SAVE_TRAIN

Obs #	Dates	counter	High	low	Open	Close	volume	adj_close	diff	target	Stock.
3873	12/13/2017	4898.KL	0.620000005	0.600000024	0.600000024	0.620000005	525300	0.620000005	0.019999981	1TA	
3874	12/14/2017	4898.KL	0.625	0.605000019	0.605000019	0.620000005	424900	0.620000005	0.014999986	1TA	
3875	12/15/2017	4898.KL	0.620000005	0.605000019	0.620000005	0.610000014	645500	0.610000014	-0.009999991	0TA	
3876	12/18/2017	4898.KL	0.610000014	0.595000029	0.610000014	0.595000029	2059700	0.595000029	-0.014999985	0TA	
3877	12/19/2017	4898.KL	0.605000019	0.595000029	0.595000029	0.605000019	299700	0.605000019	0.009999999	1TA	
3878	12/20/2017	4898.KL	0.605000019	0.600000024	0.600000024	0.600000024	406800	0.600000024	0	0TA	
3879	12/21/2017	4898.KL	0.605000019	0.595000029	0.600000024	0.595000029	383500	0.595000029	-0.004999995	0TA	
3880	12/22/2017	4898.KL	0.605000019	0.595000029	0.595000029	0.605000019	395400	0.605000019	0.009999999	1TA	
3881	12/25/2017	4898.KL	0.605000019	0.605000019	0.605000019	0.605000019	0	0.605000019	0	0TA	
3882	12/26/2017	4898.KL	0.605000019	0.595000029	0.600000024	0.605000019	845900	0.605000019	0.004999995	1TA	
3883	12/27/2017	4898.KL	0.600000024	0.595000029	0.600000024	0.600000024	1548300	0.600000024	0	0TA	
3884	12/28/2017	4898.KL	0.605000019	0.595000029	0.595000029	0.605000019	889600	0.605000019	0.009999999	1TA	
3885	12/29/2017	4898.KL	0.610000014	0.600000024	0.600000024	0.605000019	746000	0.605000019	0.004999995	1TA	
3886	12/01/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3887	12/04/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3888	12/05/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3889	12/06/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3890	12/07/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3891	12/08/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3892	12/11/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3893	12/12/2017	4936.KL	0MALPAC
3894	12/13/2017	4936.KL	0MALPAC
3895	12/14/2017	4936.KL	0MALPAC
3896	12/15/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3897	12/18/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3898	12/19/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3899	12/20/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3900	12/21/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3901	12/22/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3902	12/25/2017	4936.KL	1.220000029	1.220000029	1.220000029	1.220000029	0	1.220000029	0	0MALPAC	
3903	12/26/2017	4936.KL	1.220000029	1.200000048	1.220000029	1.200000048	5000	1.200000048	-0.019999981	0MALPAC	
3904	12/27/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3905	12/28/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3906	12/29/2017	4936.KL	1.200000048	1.200000048	1.200000048	1.200000048	0	1.200000048	0	0MALPAC	
3907	12/01/2017	4944.KL	0.814999998	0.814999998	0.814999998	0.814999998	0	0.814999998	0	0NYLEX	
3908	12/04/2017	4944.KL	0.810000002	0.800000012	0.805000007	0.810000002	587100	0.810000002	0.004999995	1NYLEX	
3909	12/05/2017	4944.KL	0.810000002	0.790000021	0.810000002	0.800000012	721700	0.800000012	-0.009999999	0NYLEX	
3910	12/06/2017	4944.KL	0.800000012	0.790000021	0.800000012	0.800000012	932300	0.800000012	0	0NYLEX	

1. Sorted adj_close to get ascending's for clearer in order to analyze missing value

[illegible]

1. Import FILTER node and connect to data source STOCK in diagram
2. Change Default Filtering Method for Interval Variables to User-Specified Limits

The screenshot displays the SAS Enterprise Miner interface. On the left, the 'Data Sources' tree shows 'stock' under 'data mining stock'. The 'Diagram' pane shows a flow from 'stock' to 'Filter'. The 'Properties' pane on the left shows the configuration for the 'Filter' node. The 'Interval Variables' section is expanded, and the 'Default Filtering Method' is set to 'User-Specified Limits'. A status bar at the bottom indicates 'Diagram segmentation analysis opened'.

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	Yes
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	Yes
Tuning Parameters	
Score	
Create Score Code	Yes
Update Measurement Level	No
Status	

Diagram segmentation analysis opened

1. Open Interactive Interval Filter by select the Interval Variables ellipsis

The screenshot displays the SAS Enterprise Miner interface. On the left, the 'data mining stock' project tree is visible, with 'segmentation analysis' selected. Below the tree, the 'Property' pane shows various settings for the selected task. A red arrow labeled '1' points to the 'Interval Variables' section in this pane. The 'Interactive Interval Filter' dialog box is open in the center, displaying a table of variables and their filtering methods. The dialog box has a title bar and a close button. The table lists variables such as 'Close', 'Dates', 'High', 'Open', 'adj_close', 'diff', 'low', 'target', and 'volume'. The 'Filtering Method' column shows 'Default' for all variables. The 'Keep Missing Values' column shows 'Default' for all variables. The 'Filter Lower Limit' and 'Filter Upper Limit' columns show '.' for all variables. The dialog box also has a 'Generate Summary' button and 'OK' and 'Cancel' buttons.

Enterprise Miner - data mining stock

File Edit View Actions Options Window Help

data mining stock

- Data Sources
 - stock
- Diagrams
 - predictive analysis
 - Predictive Analysis 2
 - segmentation analysis
- Model Packages

Property Value

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	Yes
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	Yes
Tuning Parameters	
Score	
Create Score Code	Yes
Update Measurement Level	No
Status	

Default Filtering Method

Default filtering method for interval variables.

Diagram segmentation analysis opened

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
Close	No	Default	Default	.	.
Dates	No	Default	Default	.	.
High	No	Default	Default	.	.
Open	No	Default	Default	.	.
adj_close	No	Default	Default	.	.
diff	No	Default	Default	.	.
low	No	Default	Default	.	.
target	No	Default	Default	.	.
volume	No	Default	Default	.	.

Generate Summary

OK Cancel

Diagram Log

wqd180031@siswa.um.edu.my as u35818613 Connected to SASApp - Logical Workspace Server (odaws01-apse1.oda.sas.com)

1. Change adj_close Filter Lower Limit to 0.1 and run the FILTER node

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
Close	No	Default	Default	.	.
Dates	No	Default	Default	.	.
High	No	Default	Default	.	.
Open	No	Default	Default	.	.
adj_close	No	Default	Default	0.1	.
diff	No	Default	Default	.	.
low	No	Default	Default	.	.
target	No	Default	Default	.	.
volume	No	Default	Default	.	.

< >

Generate Summary OK Cancel

Finding:

1. 577 observations are Excluded

Results - Node: Filter Diagram: segmentation analysis

File Edit View Window

Limits for Interval Variables

Variable	Role	Minimum	Maximum	Filter Method	Keep Missing Values	Label
adj_close	INPUT		0.1	.MANUAL	Y	

Output

```
37
38
39
40
41 Number Of Observations
42
43 Data
44 Role    Filtered  Excluded  DATA
45
46 TRAIN   15761     577      16338
47
48
49
50 Statistics for Original and FILTERED Data
51 (maximum 500 observations printed)
52
53 Data Role=TRAIN Variable=adj_close
```

← 1

1. Import CLUSTER node and connect to FILTER node
2. Open CLUSTER node Variables
3. Select Close, High and adj_close variables

The screenshot displays the SAS Enterprise Miner interface. On the left, a tree view shows the project structure with 'segmentation analysis' selected. Below it, the 'General' and 'Train' property windows are visible. The main workspace shows a workflow diagram with three nodes: 'stock', 'Filter', and 'Cluster', connected in sequence. A red arrow labeled '1' points to the 'Cluster' node. A second red arrow labeled '2' points to the 'Train' property window. A third red arrow labeled '3' points to the 'Variables - Clus' dialog box, which is open and shows a table of variables.

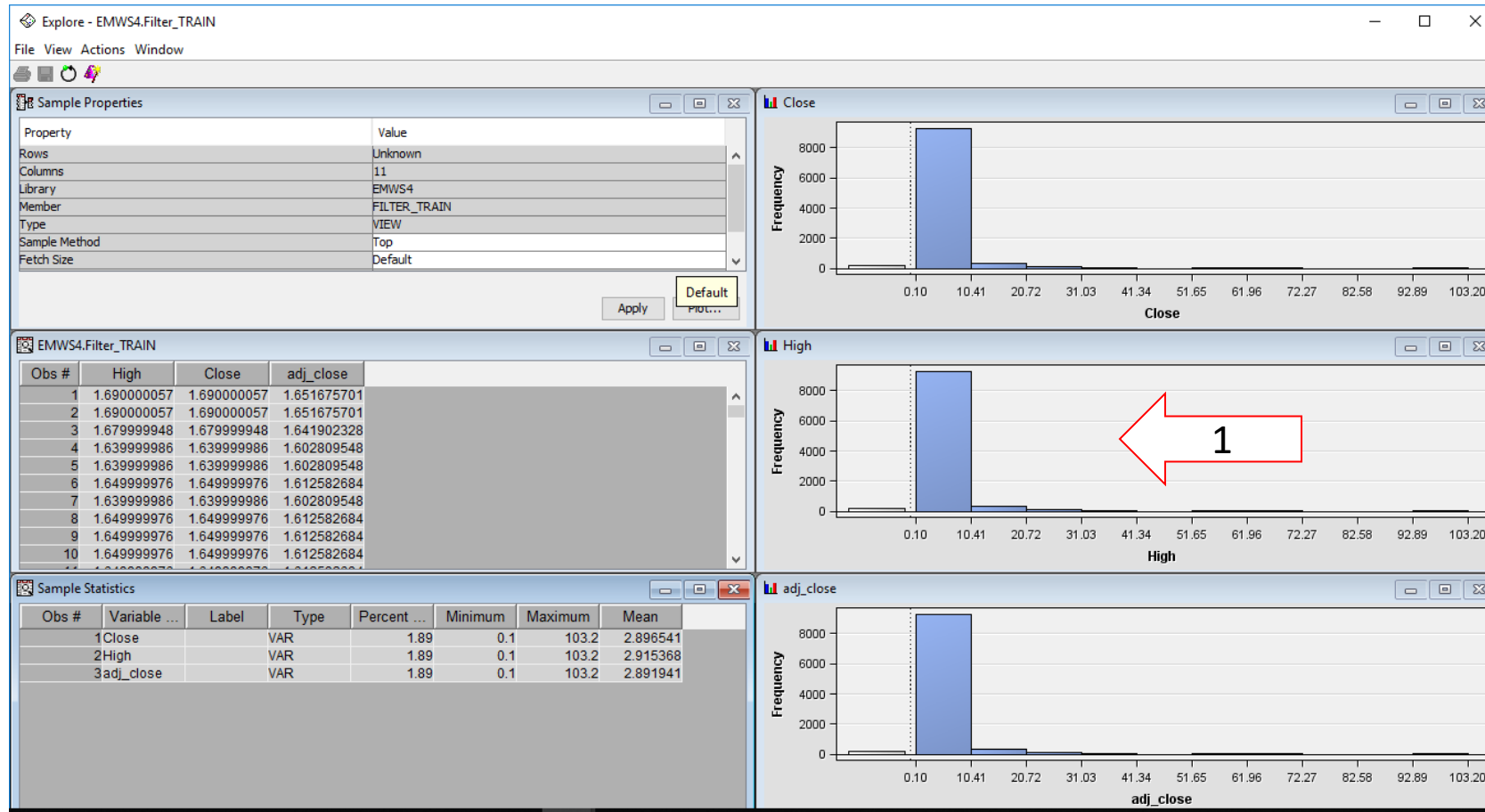
Variables - Clus

Name	Use	Report	Role	Level
Close	Default	No	Input	Interval
High	Default	No	Input	Interval
Open	Default	No	Input	Interval
Stock_	No	No	Rejected	Nominal
adj_close	Default	No	Input	Interval
counter	No	No	Rejected	Nominal
diff	Default	No	Input	Interval
low	Default	No	Input	Interval
volume	Default	No	Input	Interval

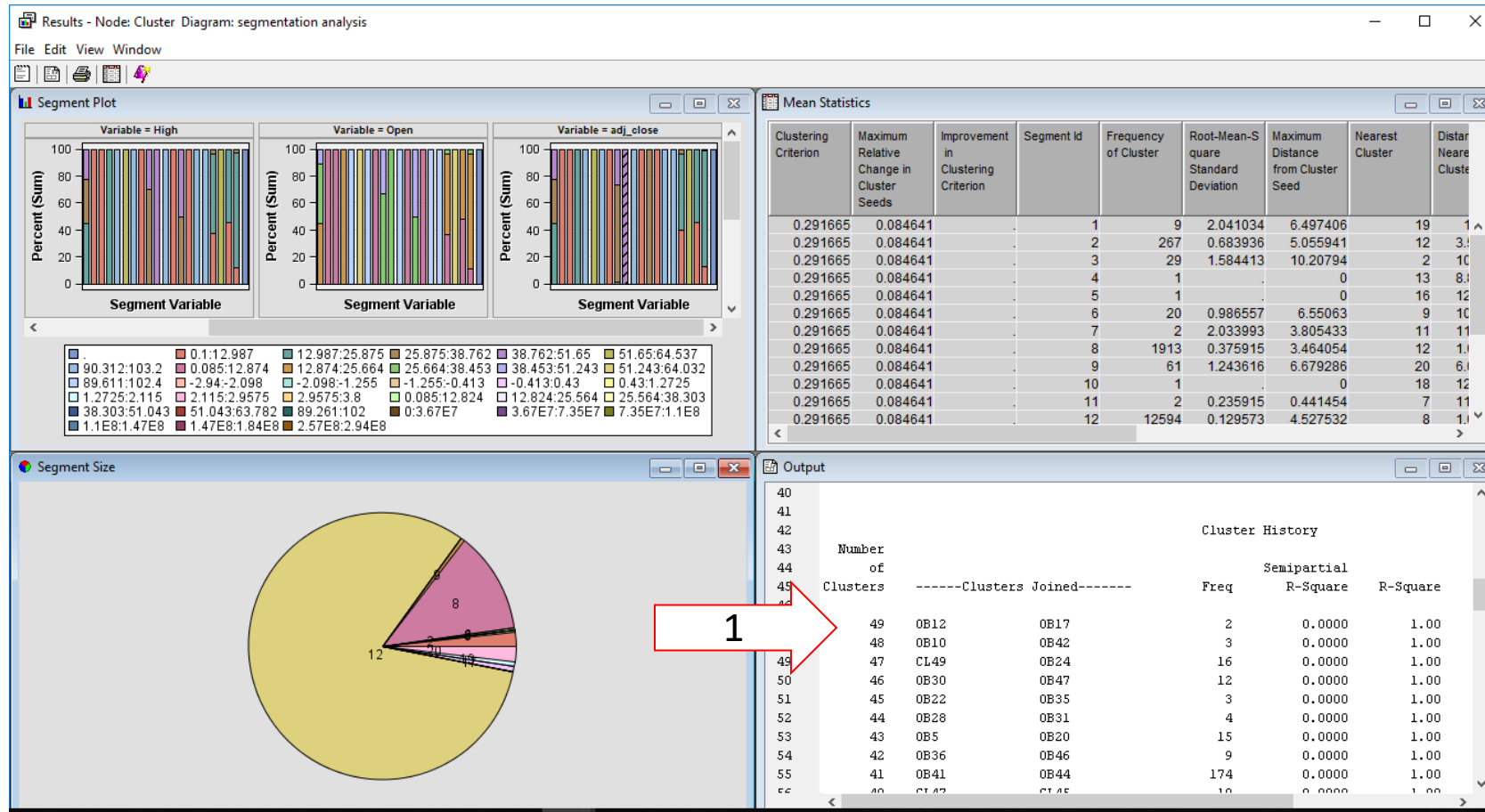
Run completed

wqd180031@siswa.um.edu.my as u35818613 Connected to SASApp - Logical Workspace Server (odaws01-apse1.oda.sas.com)

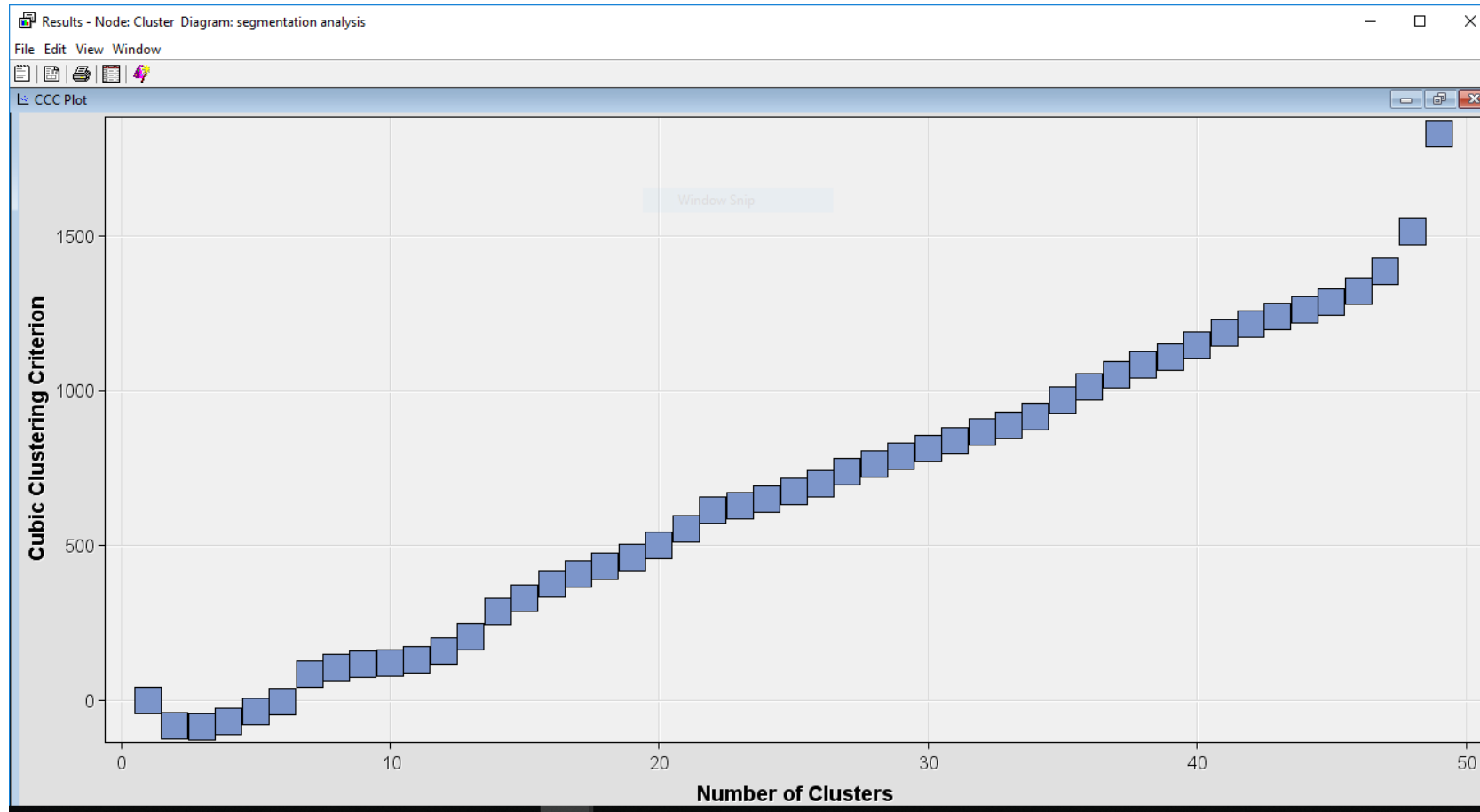
1. Explore Close, High and adj_close and it shows similar pattern



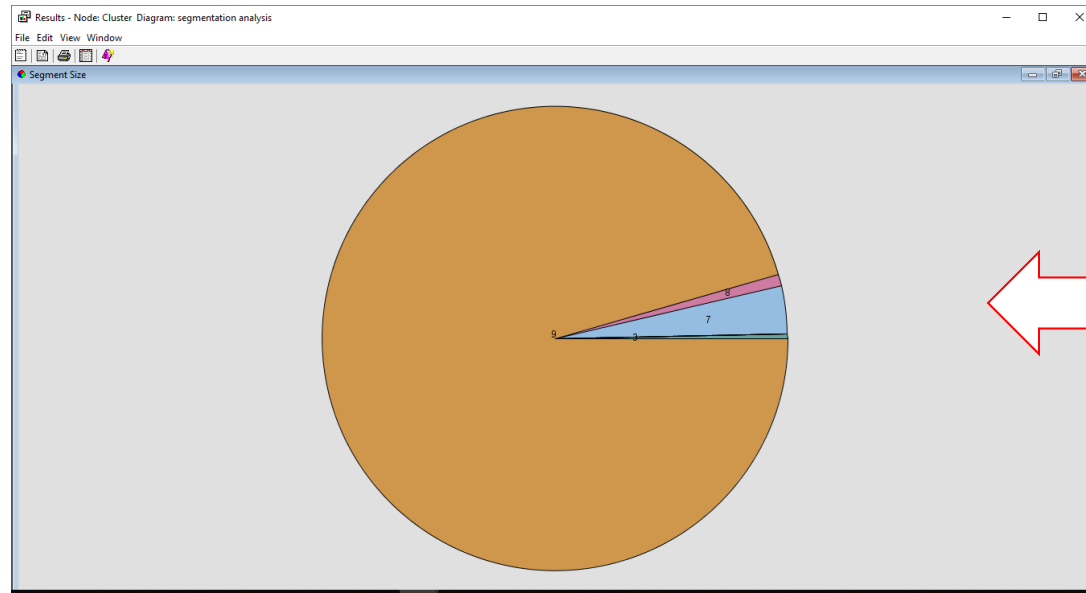
1. Run CLUSTER node and found there is 49 cluster have been created



1. View Summary Statistics in CCC Plot, it shows clearly all 49 clusters



1. Change Specification Method to User Specified with default 10
2. Re-run the CLUSTER node to create maximum 10 Clusters.



Results - Node: Cluster Diagram: segmentation analysis

File Edit View Window

Mean Statistics

Clustering Criterion	Maximum Relative Change in Cluster Sizes	Improvement in Clustering Criterion	Segment #	Frequency of Cluster	Proportion of Sample Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster	Close	High	Open	W_Like	W_T	W_F	Waste
0.45941	0.007682		1	2	0.235919	8.447454	8	11.52648	102	102.45	99.8	102	2.450002	99.43	126402
0.45941	0.007682		2	2	1.180284	2.207628	8	38.87419	0.81	0.7175	0.6775	0.81	-0.0875	0.5825	2.811888
0.45941	0.007682		3	35	3.850112	10.39984	7	22.76818	77.18343	77.51543	77.17429	77.18343	0.899143	76.97543	98008.67
0.45941	0.007682		4	2	3.303877	6.181015	3	27.38688	99.73	102.1	102.1	99.73	-2.37	99.4	131550
0.45941	0.007682		5	5	2.871988	8.685818	7	17.4214	24.232	24.378	25.792	24.232	-1.56	24.10	3245798
0.45941	0.007682		6	2	2.633983	3.805433	1	15.38648	102.8	102.8	99.25	102.8	3.450201	98.75	3.78856
0.45941	0.007682		7	494	1.319558	8.722183	9	8.920428	19.19814	19.29919	19.1851	19.19321	0.030008	19.04575	1701684
0.45941	0.007682		8	120	1.832516	16.84898	9	7.147998	0.846888	0.867839	0.837881	0.846888	0.002876	0.917889	44438639
0.45941	0.007682		9	1498	0.291473	7.445411	7	8.920428	1.54851	1.556881	1.543825	1.543258	0.002183	1.532208	888837.3
0.45941	0.007682		10	55	3.958232	18.28653	7	14.85074	29.33087	29.38897	27.868	29.32964	1.362887	27.858	1686147

1. Select Exported Data
2. Select TRAIN and Explore

The screenshot displays the SAS Enterprise Miner interface. On the left, the 'data mining stock' tree shows a workflow with 'segmentation analysis' selected. The 'Property' pane on the left lists various settings for the 'Train' node, including 'Internal Standardization' set to 'Standardization', 'Number of Clusters' set to 'User Specify', and 'Clustering Method' set to 'Ward'. A red arrow labeled '1' points to the 'Exported Data' property in the 'Train' node.

The main workspace shows a 'segmentation analysis' diagram. A dialog box titled 'Exported Data - Cluster' is open, displaying a table of data sources and their roles. A red arrow labeled '2' points to the 'TRAIN' row in this table. Below the table, a progress bar indicates 'Loading EMWS4.Clus_TRAIN ...' with 'Rows: 2,000 Columns: 14'. The dialog box has buttons for 'Browse...', 'Explore...', 'Properties...', and 'OK'.

Port	Table	Role	Data Exists
TRAIN	EMWS4.Clus_TRAIN	Train	Yes
VALIDATE	EMWS4.Clus_VALIDATE	Validate	No
TEST	EMWS4.Clus_TEST	Test	No
CLUSSTAT	EMWS4.Clus_OUTSTAT	Cluster Statistics	Yes
CLUSMEAN			Yes
VARMAP			Yes

Run completed

Connected to SASApp - Logical Workspace Server (odaws01-apse1.oda.sas.com)

1. Create 3D Scatter Plot

Explore - EMWS4.Clus_TRAIN

File View Actions Window

Sample Properties

Property	Value
Rows	Unknown
Columns	14
Library	EMWS4
Member	CLUS_TRAIN
Type	VIEW
Sample Method	Top
Fetch Size	Default
Fetch Rows	2000
Random Seed	12345

Sample Statistics

Obs #	Variable ...	Label	Type	Percent ...	Minimum	Maximum	Mean	Number of
1	Stock_	Stock	CLASS	0				.97
2	_SEGMENT...	Segment D...	CLASS	0				.6
3	counter		CLASS	0				.97
4	Close		VAR	0	0.1	28.4	3.274267	
				0	21154	21182	21168.76	
				0	0.040539	16.94558	0.976547	
				0	0.1	28.4	3.292488	
				0	0.095	28.4	3.267448	
				0	2	10	8.8255	
				0	0.1	28.4	3.268558	
				0	-0.82	0.959999	0.006819	
				0	0.095	28.3	3.246469	
				0	0	1	0.322	
				0	0	2.6847E8	2395962	

Select a Chart Type

Scatter

Scatter plot in 3 dimensions.

1

Cancel < Back Next > Finish

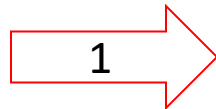
EMWS4.Clus_TRAIN

Obs #	Dates	counter	High	low
1	12/01/2017 0002.KL	1.690000057	1.690000057	
2	12/04/2017 0002.KL	1.690000057	1.690000057	
3	12/05/2017 0002.KL	1.679999948	1.639999986	
4	12/06/2017 0002.KL	1.639999986	1.639999986	
5	12/07/2017 0002.KL	1.639999986	1.639999986	
6	12/08/2017 0002.KL	1.649999976	1.649999976	
7	12/11/2017 0002.KL	1.639999986	1.639999986	
8	12/12/2017 0002.KL	1.649999976	1.649999976	
9	12/13/2017 0002.KL	1.649999976	1.649999976	1.649999976
10	12/14/2017 0002.KL	1.649999976	1.649999976	1.649999976
11	12/15/2017 0002.KL	1.649999976	1.649999976	1.649999976
12	12/18/2017 0002.KL	1.75	1.659999967	1.700000048
13	12/19/2017 0002.KL	1.700000048	1.700000048	1.700000048
14	12/20/2017 0002.KL	1.720000029	1.700000048	1.720000029
15	12/21/2017 0002.KL	1.720000029	1.720000029	1.720000029
16	12/22/2017 0002.KL	1.769999981	1.720000029	1.769999981
17	12/25/2017 0002.KL	1.769999981	1.769999981	1.769999981

distance Segment Description

0.157836	Cluster9
0.157836	Cluster9
0.337346	Cluster9
0.150645	Cluster9
0.152285	Cluster9
0.152429	Cluster9
0.151118	Cluster9
0.152429	Cluster9
0.153212	Cluster9
0.152585	Cluster9
0.15152	Cluster9
0.435558	Cluster9
0.15921	Cluster9
0.246932	Cluster9
0.162206	Cluster9
0.54185	Cluster9
0.171029	Cluster9

1. Change SEGMENT Role to Color, adj_close to X, Stock to Y and volume to Z



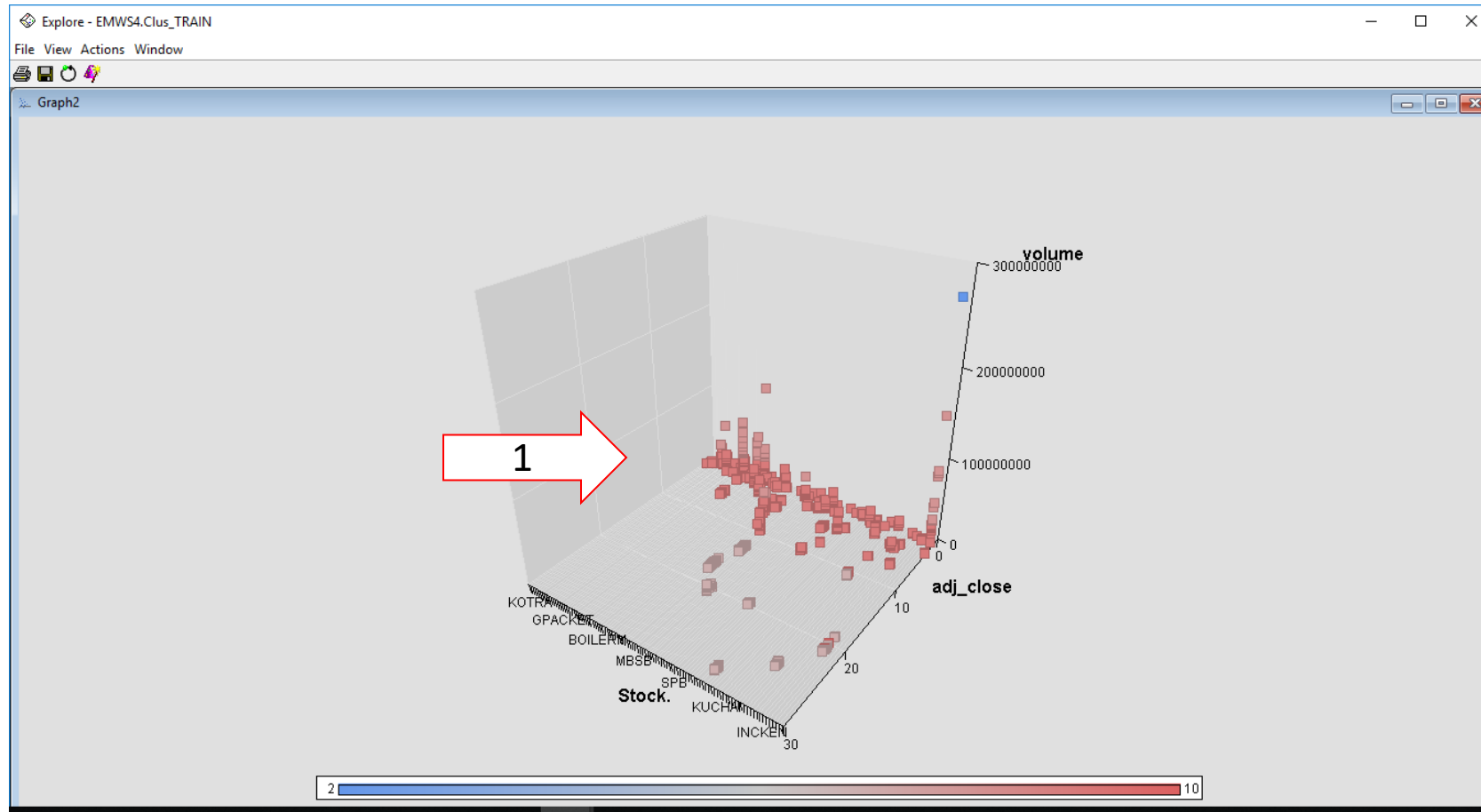
Select Chart Roles ×

▲ Variable	Role	Type	Description	Format
SEGMENT	Color	Numeric	Segment Id	
_SEGMENT_LABEL_		Character	Segment Description	
adj_close	X	Numeric	adj_close	BEST12.
Close		Numeric	Close	BEST12.
counter		Character	counter	\$7.
Dates		Numeric	Dates	MMDDYY10.
diff		Numeric	diff	BEST12.
Distance		Numeric	Distance	
High		Numeric	High	BEST12.
low		Numeric	low	BEST12.
Open		Numeric	Open	BEST12.
Stock_	Y	Character	Stock.	
target		Numeric	target	BEST12.
volume	Z	Numeric	volume	BEST12.

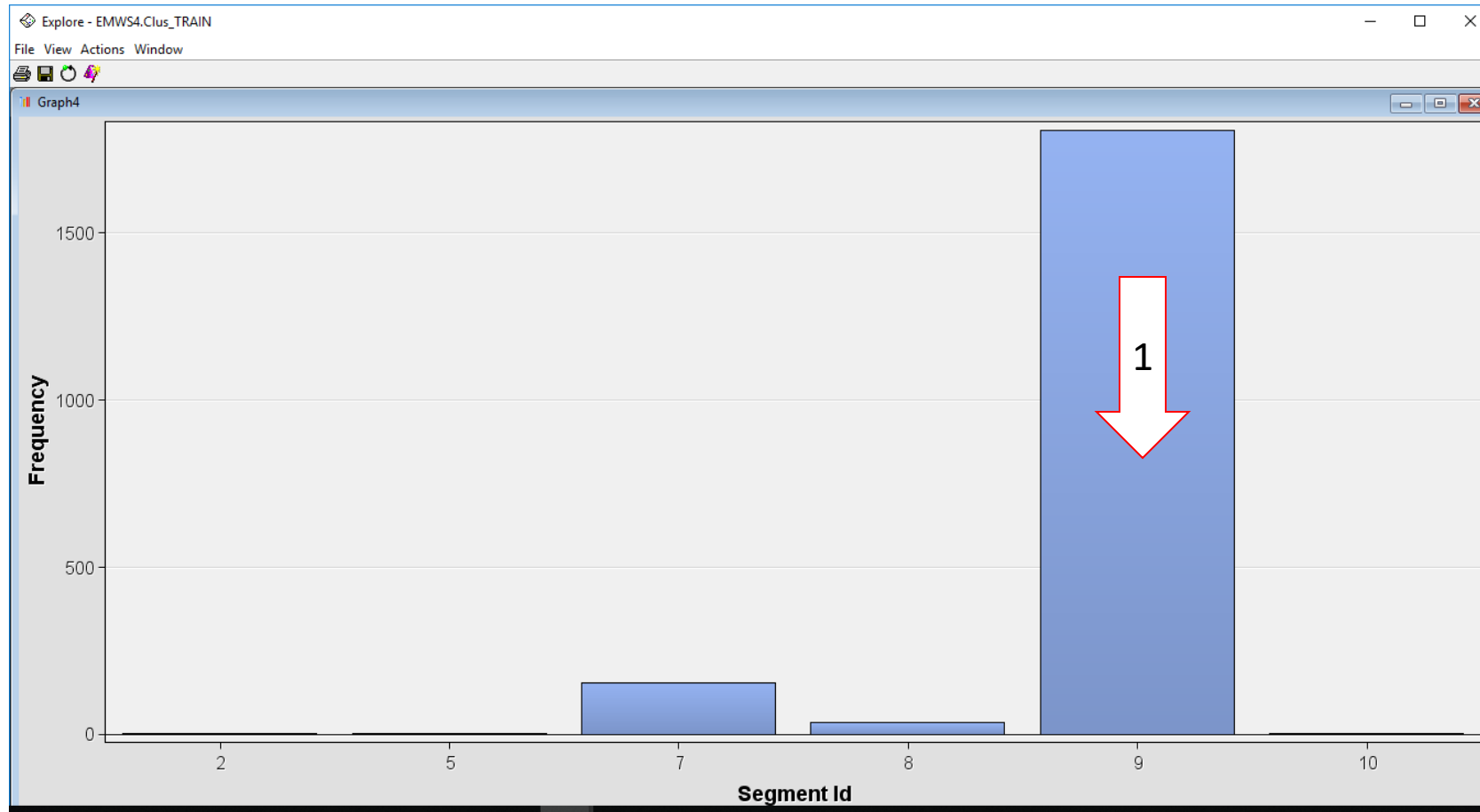
☐ Allow multiple role assignments

Finding:

1. Most of the data plotting at 1 area

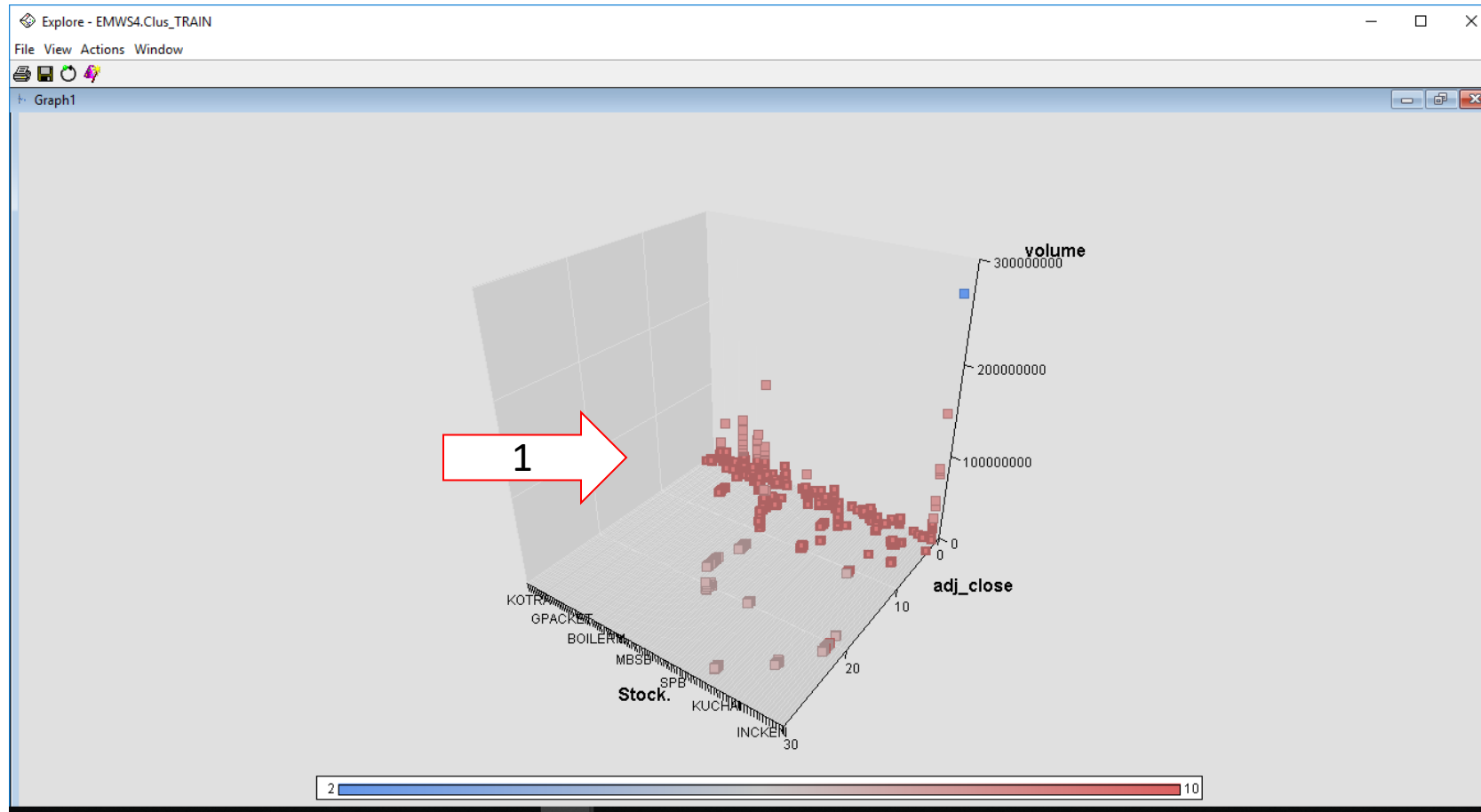


1. Select Segmen_id 9 of TRAIN for further analysis



Finding:

1. Bold RED color represent Segment_id 9



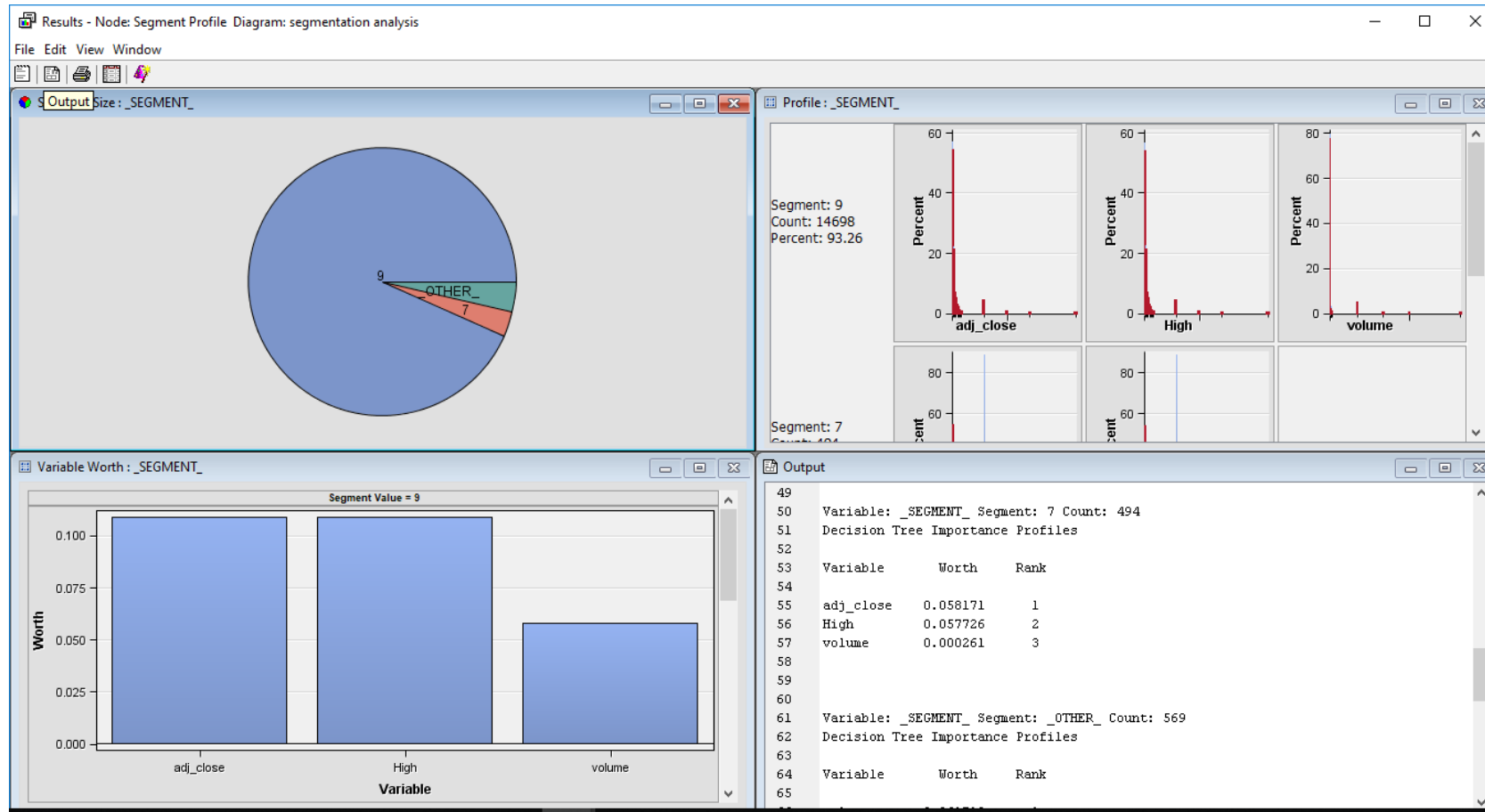
1. Import SEGMENT PROFILE node to diagram and connect to CLUSTER node
2. Select SEGMENT PROFILE node Variables
3. Change it Variables properties

The screenshot displays the SAS Enterprise Miner interface. On the left, a tree view shows the project structure, with 'segmentation analysis' selected. Below it, the 'Property' pane shows the 'General' tab for the 'Prof' node. A red arrow labeled '2' points to this pane. The main workspace shows a workflow diagram with nodes: 'stock', 'Filter', 'Cluster', and 'Segment Profile'. A red arrow labeled '1' points to the 'Segment Profile' node. A 'Variables - Prof' dialog box is open, showing a table of variables. A red arrow labeled '3' points to this dialog box.

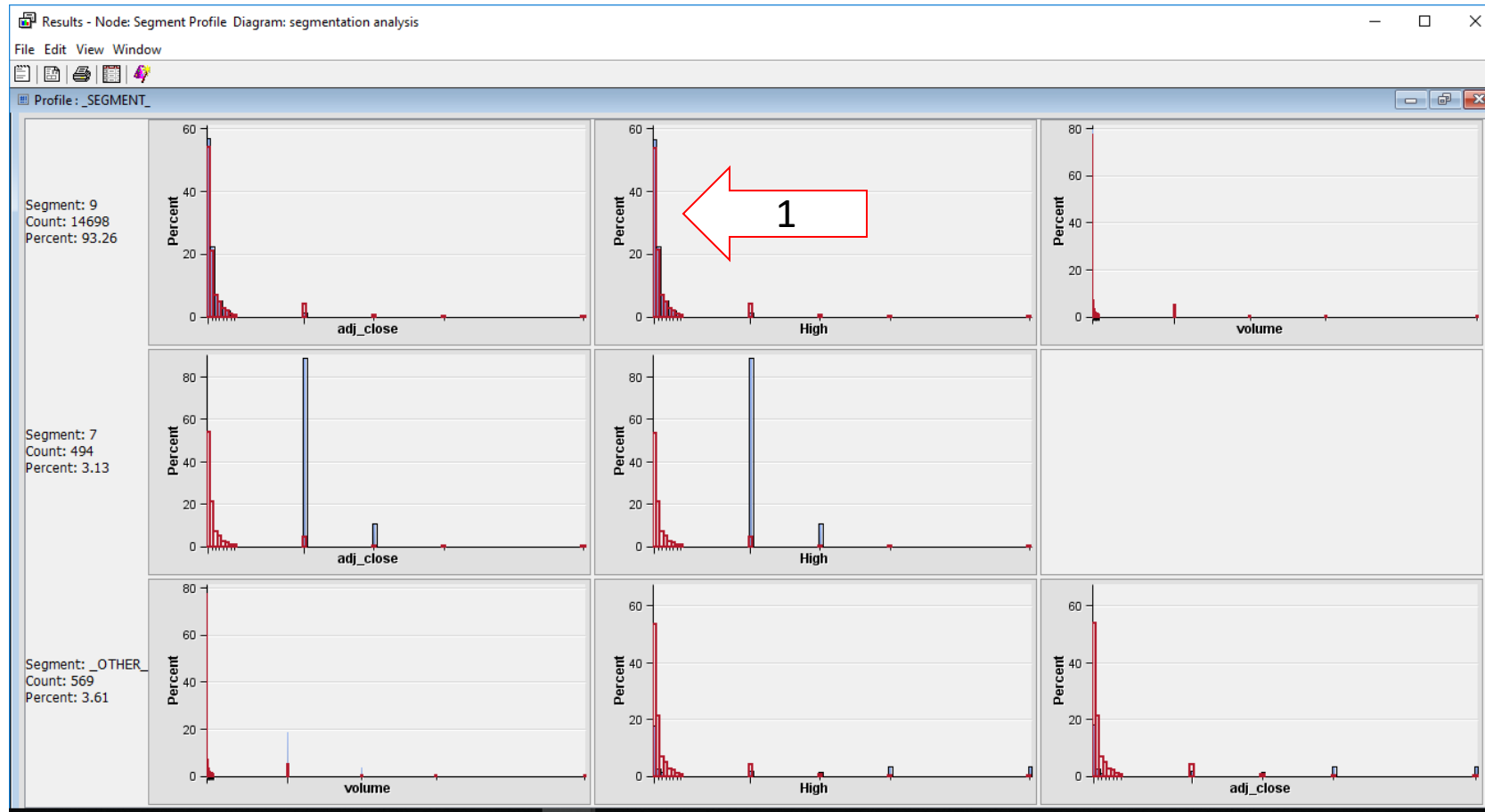
Variables - Prof

Name	Use	Report	Role	Level
Close	No	No	Input	Interval
Dates	No	No	Time ID	Interval
Distance	Default	No	Rejected	Interval
High	Default	No	Input	Interval
Open	No	No	Input	Interval
Stock_	Default	No	Rejected	Nominal
SEGMENT	Default	No	Segment	Nominal
_SEGMENT_LABEL	Default	No	Rejected	Nominal
adj_close	Default	No	Input	Interval

1. Run the SEGMENT PROFILE node and analyze the results



1. Maximize the Profile window and found features of each segment become apparent



1. Maximize the SEGMENT window and found adj_close and High show similar character compared to volume

