



MILESTONE 6

MYEG Stock Market Analysis

ZULKANAIN BIN HASAN
WQD180031

WQD7005 Data Mining
Semester I, 2019/2020
Instructor: Profesor Dr. The Ying Wah

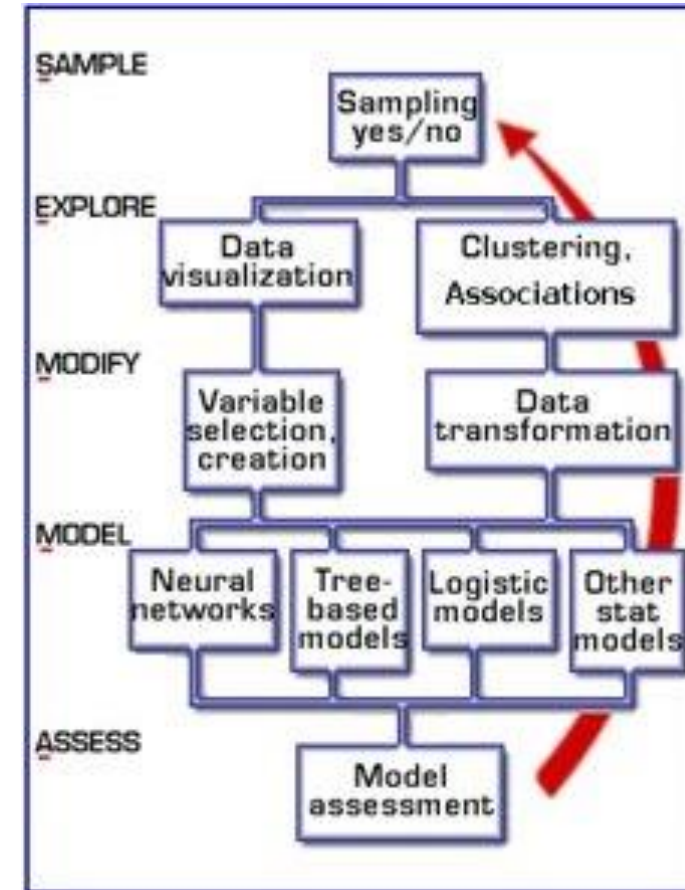
Analysis Goal:
Predict future trend
of stock market by
analyze selected
counter for short
term and long-term
investment


Analysis Data:

- Extract data from historical Bursa Malaysia Main Market
- Sample target based on high data trend
- Actual sample target approximately 10%

Methodology

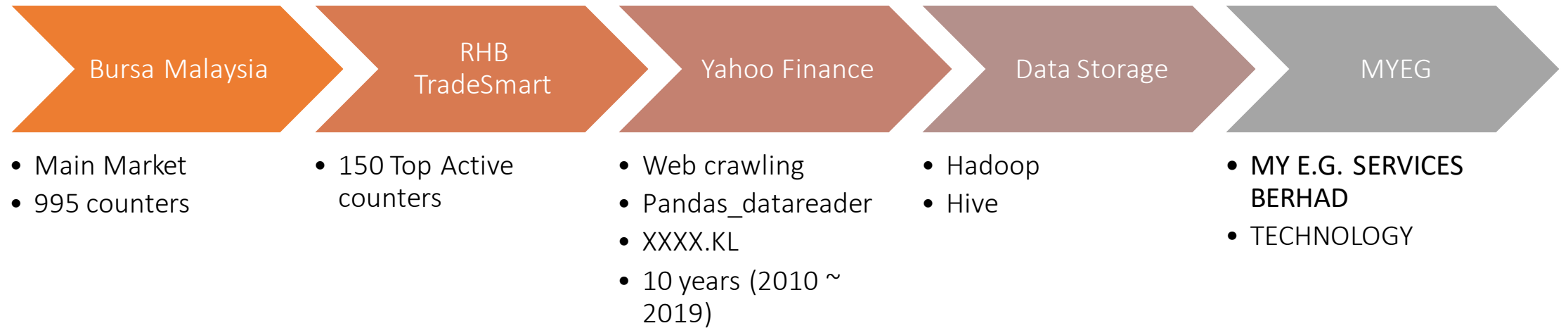
KDD	SEMMA
Pre KDD	-----
Selection	Sample
Pre processing	Explore
Transformation	Modify
Data mining	Model
Interpretation/Evaluation	Assessment
Post KDD	-----





Sample







Explore

SEMMA node:

- Cluster

1. No of variables: 7 consist of Date, Volume, Open, High, Low, Close and Adj_Close.
2. No of observations: 2405

The screenshot displays the SAS Explore window for the dataset STOCK.EM_SAVE_TRAIN. The interface includes a menu bar (File, View, Actions, Window), a toolbar, and three main panels: Sample Properties, Sample Statistics, and a data grid.

Sample Properties:

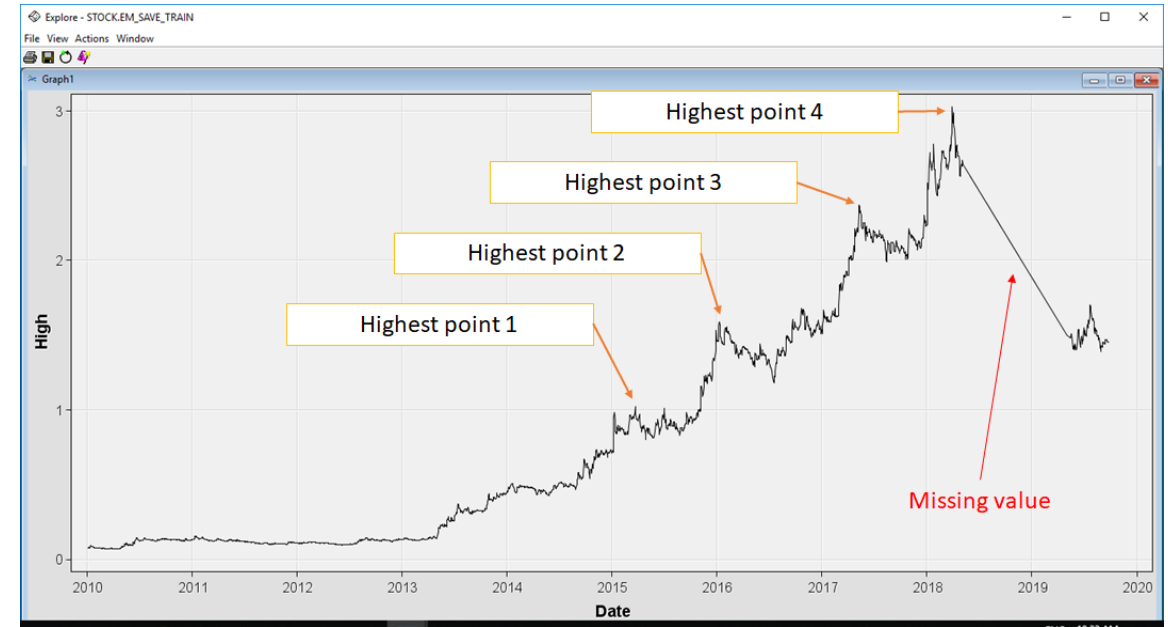
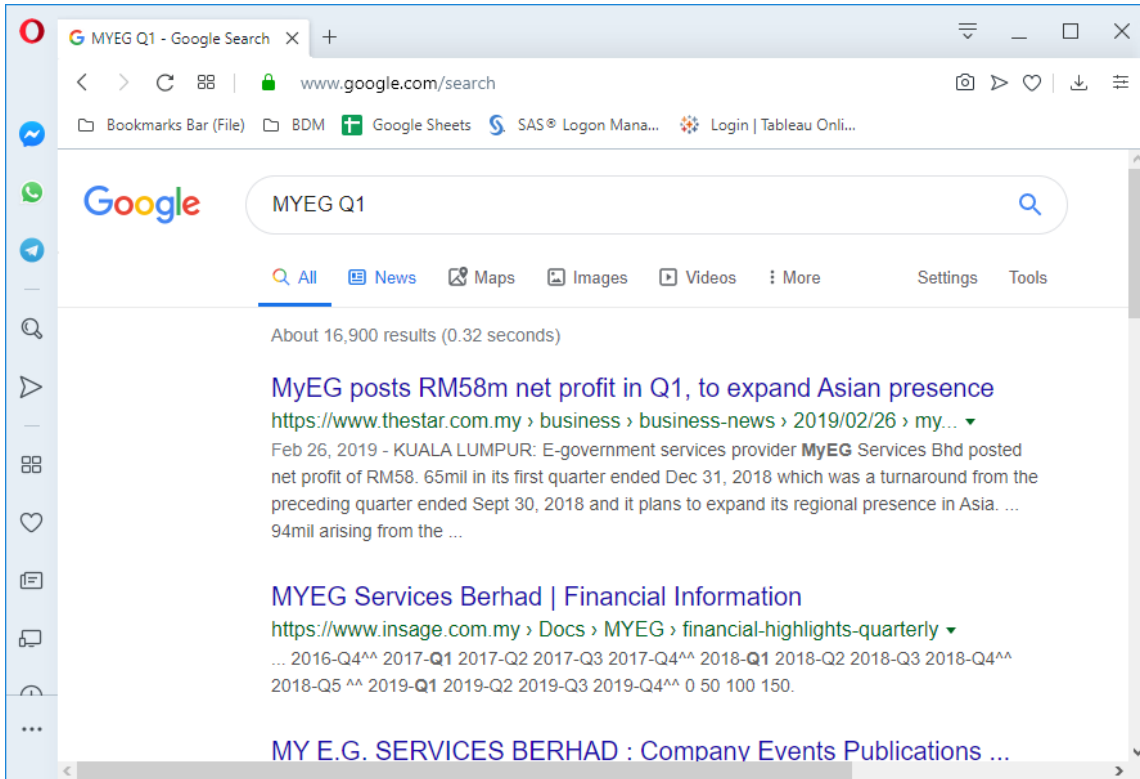
Property	Value
Rows	2405
Columns	7
Library	STOCK
Member	EM_SAVE_TRAIN
Type	DATA
Sample Method	Random
Fetch Size	Max
Fetch Rows	2405
Random Seed	12345

Sample Statistics:

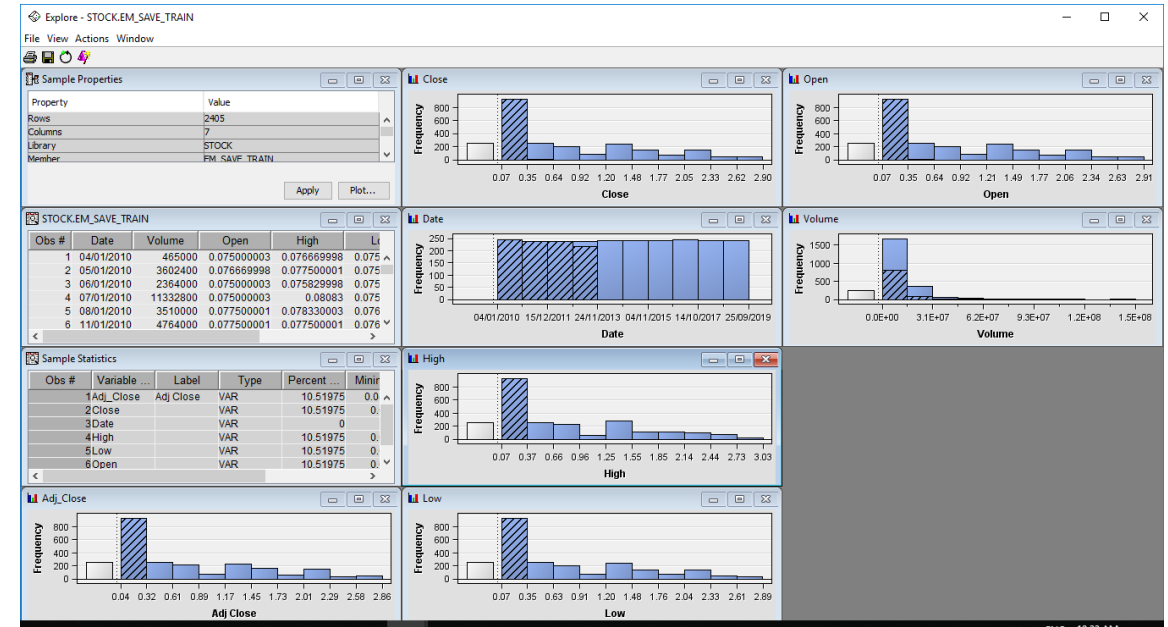
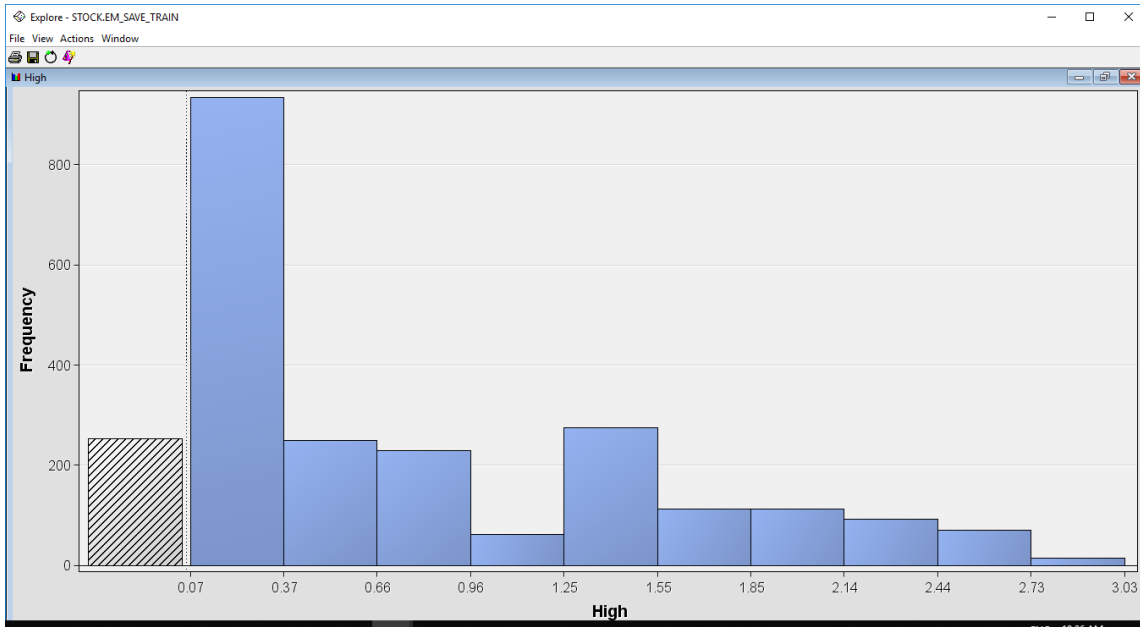
Obs #	Variable	Label	Type	Percent	Minimum	Maximum	Mean
1	Adj_Close	Adj Close	VAR	10.51975	0.042229	2.857399	0.766122
2	Close		VAR	10.51975	0.06917	2.9	0.801689
3	Date		VAR	0	18266	21817	20044.27
4	High		VAR	10.51975	0.06917	3.03	0.813151
5	Low		VAR	10.51975	0.06833	2.89	0.789987
6	Open		VAR	10.51975	0.06917	2.91	0.801413
7	Volume		VAR	10.51975	0	1.5481E8	11037711

STOCK.EM_SAVE_TRAIN Data Grid:

Obs #	Date	Volume	Open	High	Low	Close	Adj Close
1	04/01/2010	465000	0.075000003	0.076669998	0.075000003	0.076669998	0.046808
2	05/01/2010	3602400	0.076669998	0.077500001	0.075000003	0.075000003	0.045788
3	06/01/2010	2364000	0.075000003	0.075829998	0.075000003	0.075000003	0.045788
4	07/01/2010	11332800	0.075000003	0.08083	0.075000003	0.076669998	0.046808
5	08/01/2010	3510000	0.077500001	0.078330003	0.076669998	0.076669998	0.046808
6	11/01/2010	4754000	0.077500001	0.077500001	0.076669998	0.076669998	0.046808
7	12/01/2010	5300400	0.076669998	0.076669998	0.075829998	0.075829998	0.046295
8	13/01/2010	59054600	0.075829998	0.086669998	0.075829998	0.086669998	0.052913
9	14/01/2010	79297800	0.088330001	0.092500001	0.088330001	0.089170001	0.054439
10	15/01/2010	14161200	0.089170001	0.089170001	0.084169999	0.084169999	0.051387
11	18/01/2010	12315000	0.084169999	0.086669998	0.083329998	0.084169999	0.051387
12	19/01/2010	14730000	0.085000001	0.085000001	0.08083	0.081670001	0.049861
13	20/01/2010	43716600	0.081670001	0.087499999	0.081670001	0.085000001	0.051884
14	21/01/2010	17219400	0.084169999	0.085830003	0.081670001	0.081670001	0.049861
15	22/01/2010	16086600	0.081670001	0.081670001	0.079170004	0.079999998	0.048841
16	25/01/2010	5898000	0.079999998	0.079999998	0.077500001	0.078330003	0.047821
17	26/01/2010	13610400	0.079170004	0.08083	0.075829998	0.075829998	0.046295



1. Line graph shows highest point for MYEG always in Q1 financial year.
2. Google search shows MYEG will release Financial Report in Q1 every year.
3. There is missing value for MYEG data from mid 2018 till mid 2019.

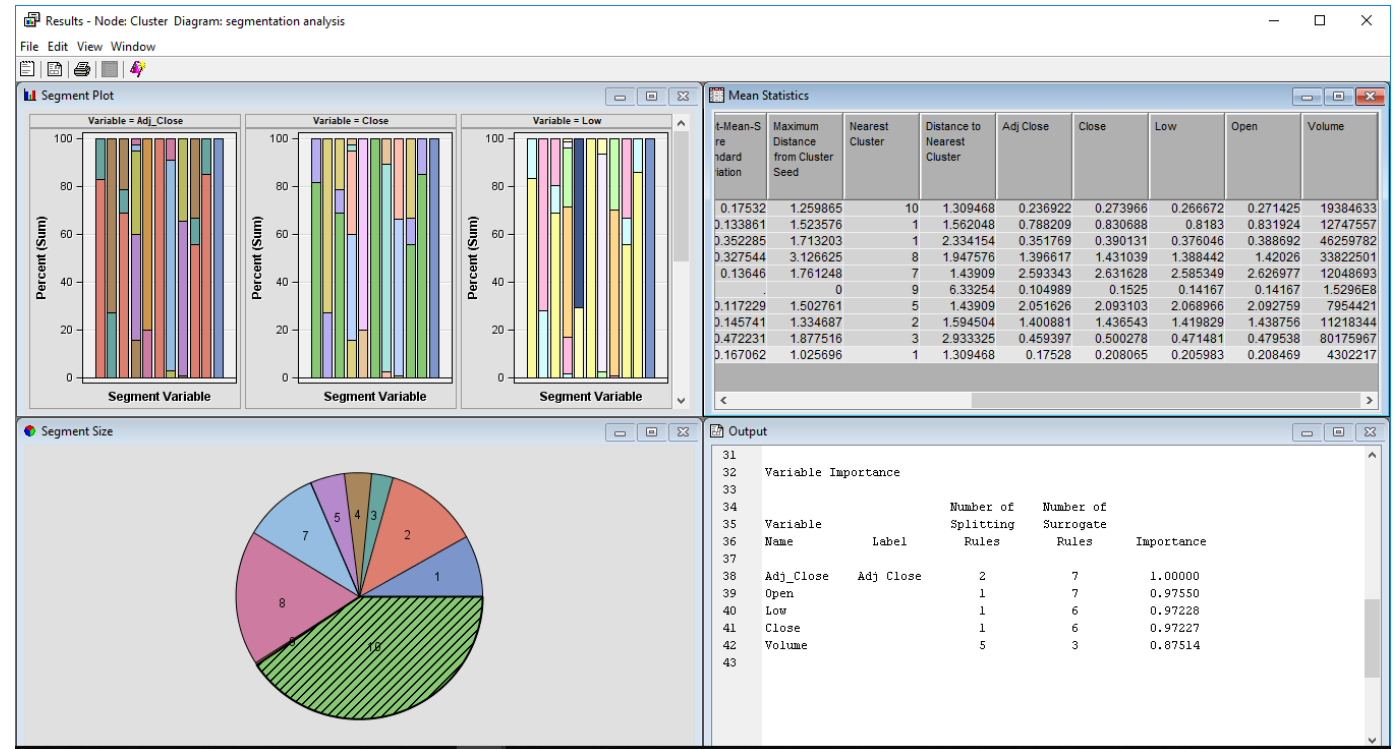


1. Histogram shows the highest point (High variable) for MYEG is in range of 0.07 to 0.37 with 935 times
2. This range tabulated from year 2010 to 2013.
3. Year 2019, MYEG stock is at mid-point which is 1.25 to 1.55.

1. Cluster 10 is the biggest cluster with the lowest Volume (4302217).

2. Cluster 9 is the smallest cluster however it consist of the biggest Volume (80175967).

3. Adj_Close is the importance variable with 1.0000 points followed by Open, Low, Close and Volume.





Modify

SEMMA node:

- Replacement

1. Setting done to High variable by using Replacement node (Limit Method = User Specified and Replacement Lower Limit = 1).
 2. Setting done to High variable Data Set Allocation by using Data Partition node (Training = 50 and Validation = 50).
-

Results - Node: Data Partition Diagram: Predictive Analysis 2

File Edit View Window

Output

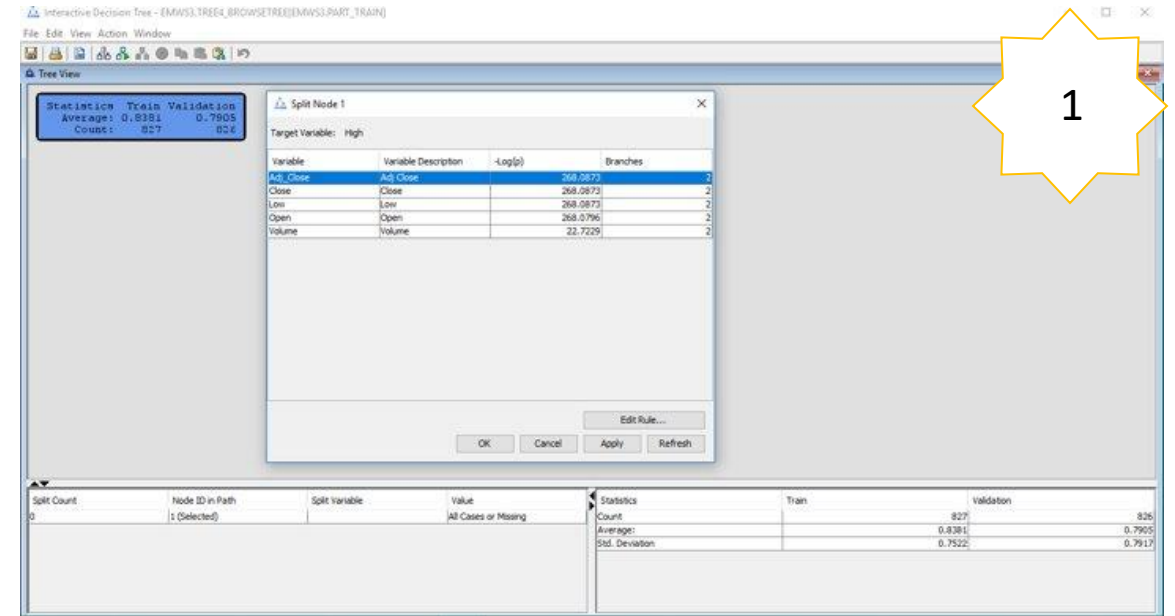
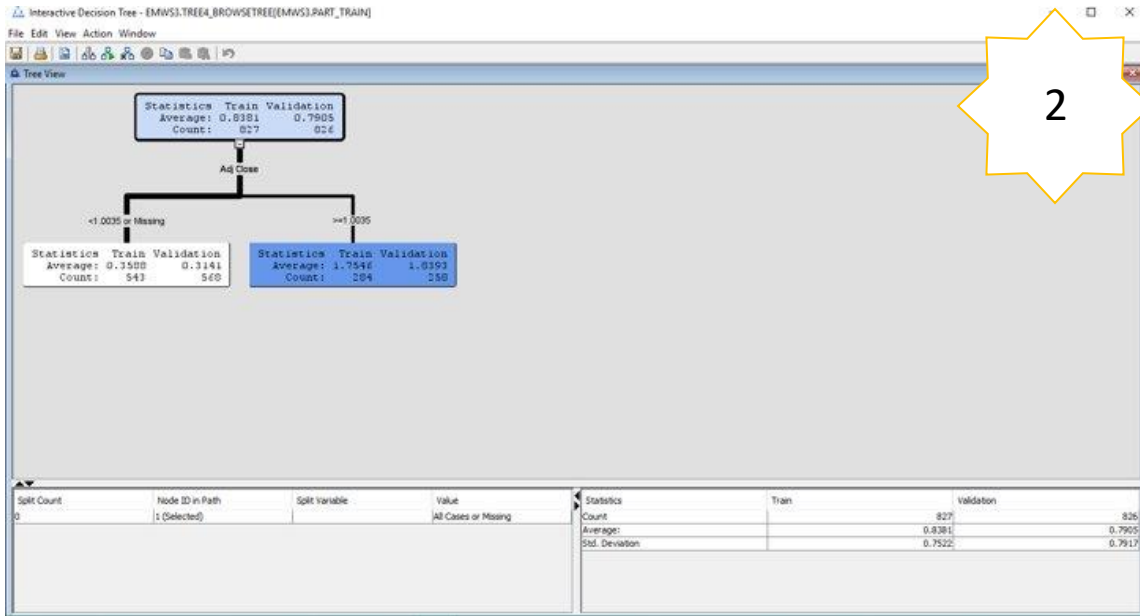
```
22
23
24 Partition Summary
25
26
27 Type          Data Set          Number of
28                               Observations
29 DATA          EMWS3.Repl_TRAIN    2405
30 TRAIN          EMWS3.Part_TRAIN    925
31 VALIDATE       EMWS3.Part_VALIDATE  925
32 TEST           EMWS3.Part_TEST     555
33
34
35 *-----*
36 * Score Output
37 *-----*
38
39
40 *-----*
41 * Report Output
42 *-----*
43
44
45
46
47 Summary Statistics for Interval Targets
48
49 Data=DATA
50
51
52 Variable      Maximum      Mean      Minimum      Number of
53                               Observations      Missing      Standard
54                               Deviation      Label
55
56 High          3.029999971  0.8131509251  0.069169998      2152      253      0.7752581854
57
58 Data=TEST
```



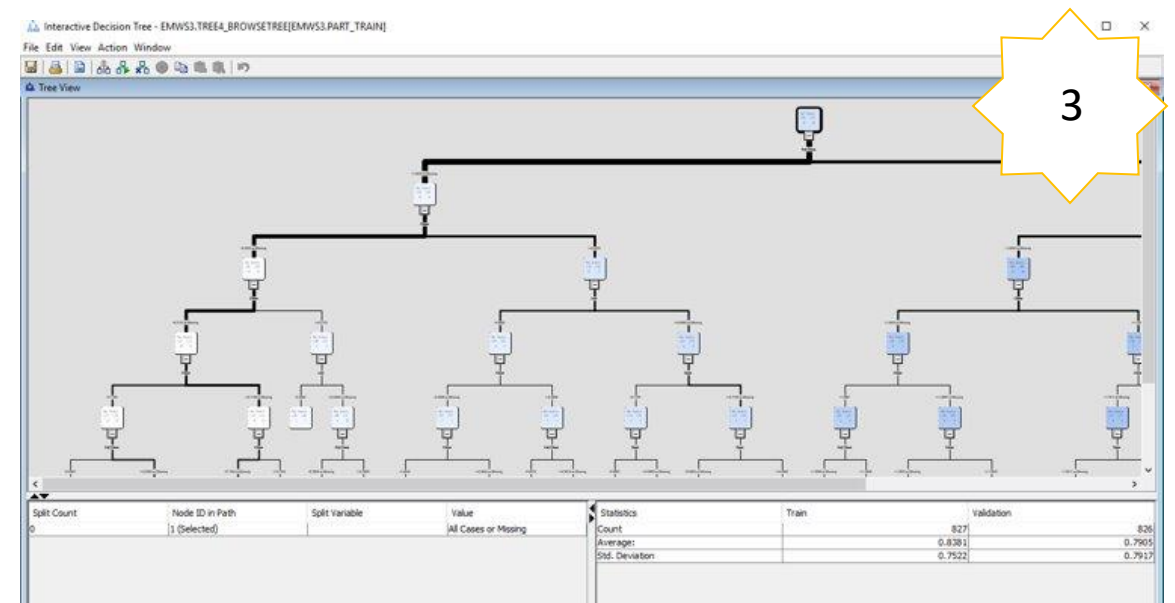
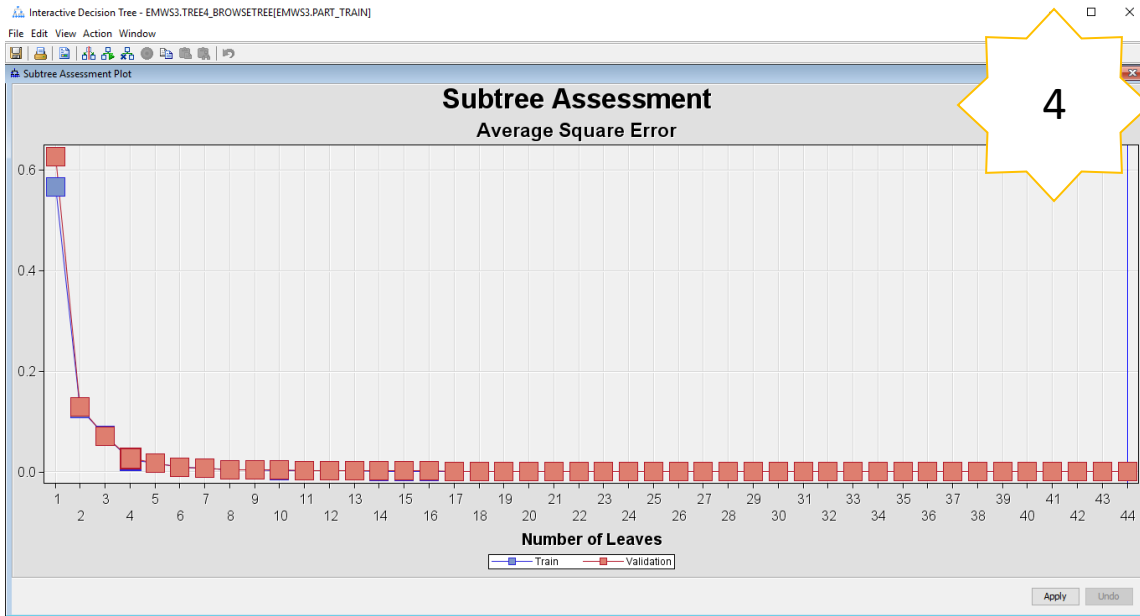
Model

SEMMA node:

- Decision Tree

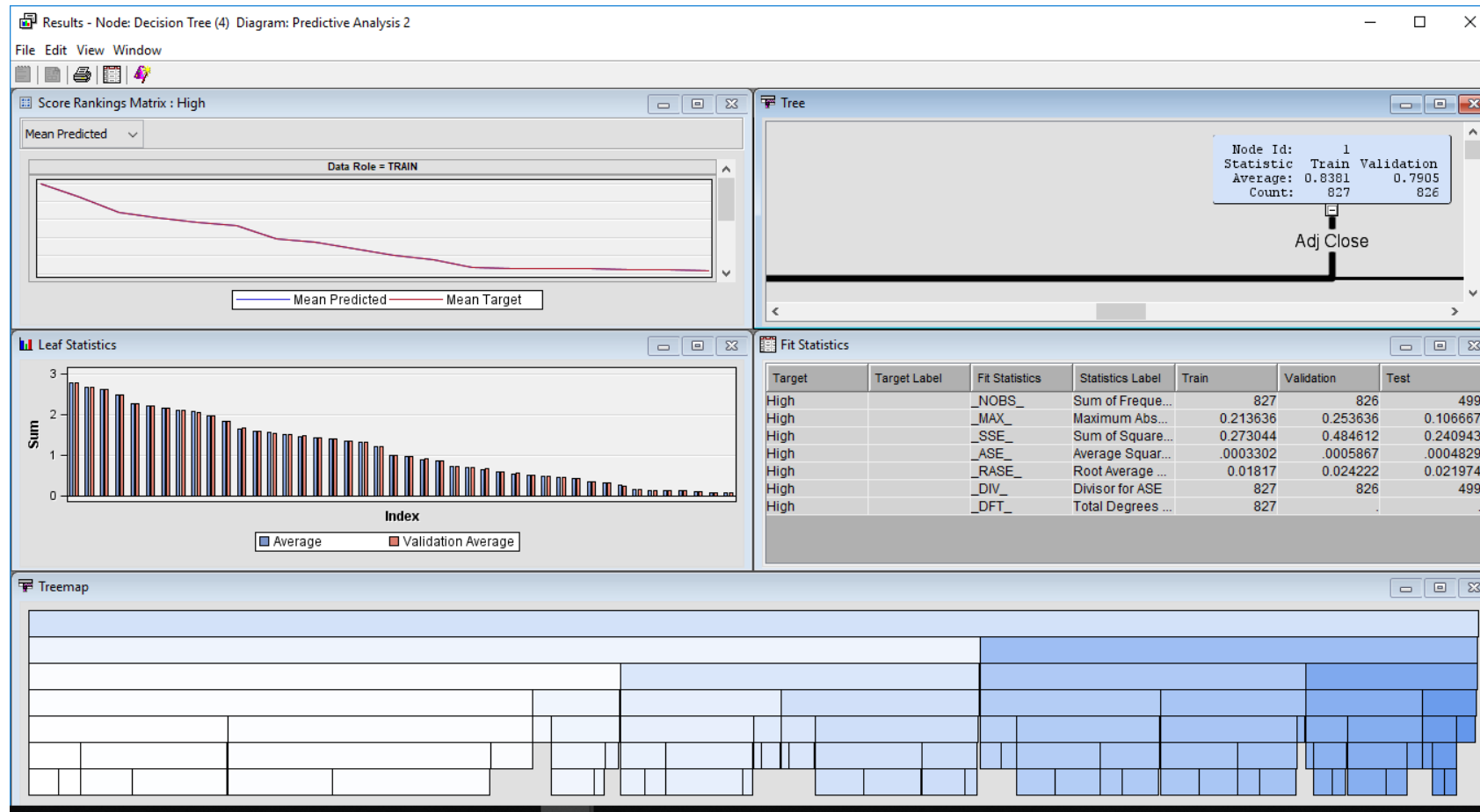


Initial decision tree result shown as per Figure 2.16. <1.0035 or Missing shows higher concentration in Split 1.



Based on 44 leaf-trees generated, 40 leaf-trees is a lower misclassification rate as shown in Subtree Assessment.

Full result of decision tree.



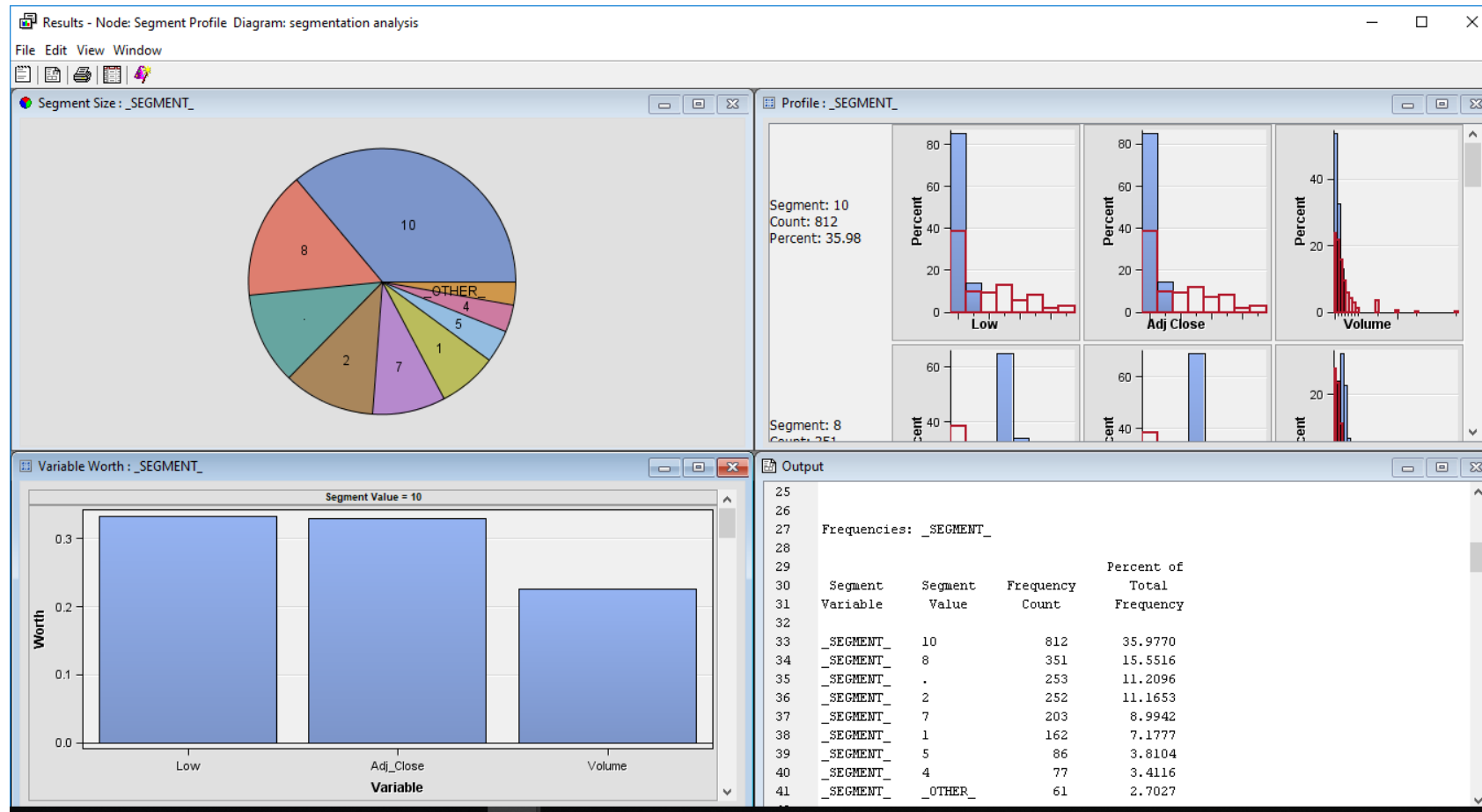


Assess

SEMMA node:

- Segment Profile

- Segment 10 has the highest frequency which is 812.
- Followed by segment 8 with 351 frequency.



- Segment 10, Low = min, Adj_Close = min, Volume = min.
- Segment 8, Adj_Close = center, Low = center, Volume = min.



Conclusion

- MYEG stock shows good uptrend from year 2013 to 2018.
- Q1 every year is a good period to gain higher profit in investment (positive sentiment due to company Financial Report).
- The best time to invest in MYEG stock is in Q3 (lowest point every year).
- Even though current value of MYEG stock is at mid-point which is 1.25 to 1.55 it still shows a good sign to improved based similar position as per mid 2017.