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:

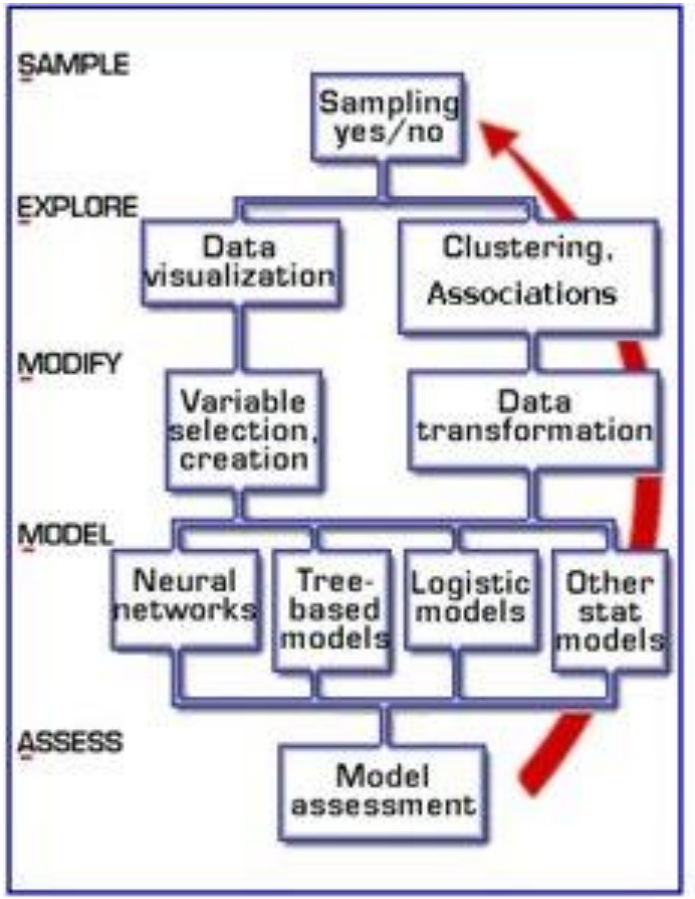
Predict future trend of stock market by analyze selected counter for short term and long-term investment

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

- Main Market
- 995 counters

- 150 Top Active counters

- Web crawling
- Pandas_datareader
- XXXX.KL
- 10 years (2010 ~ 2019)

- Hadoop
- Hive

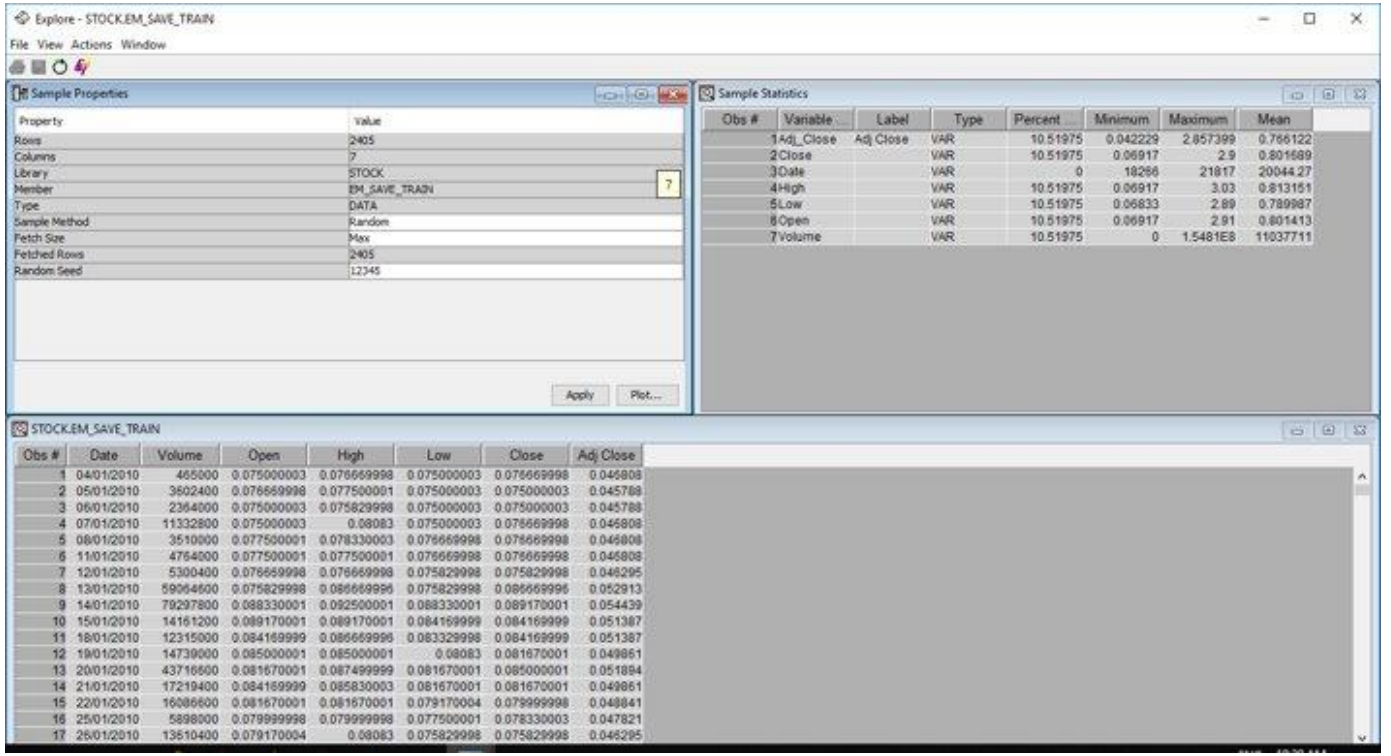
- **MY E.G. SERVICES BERHAD**
- TECHNOLOGY

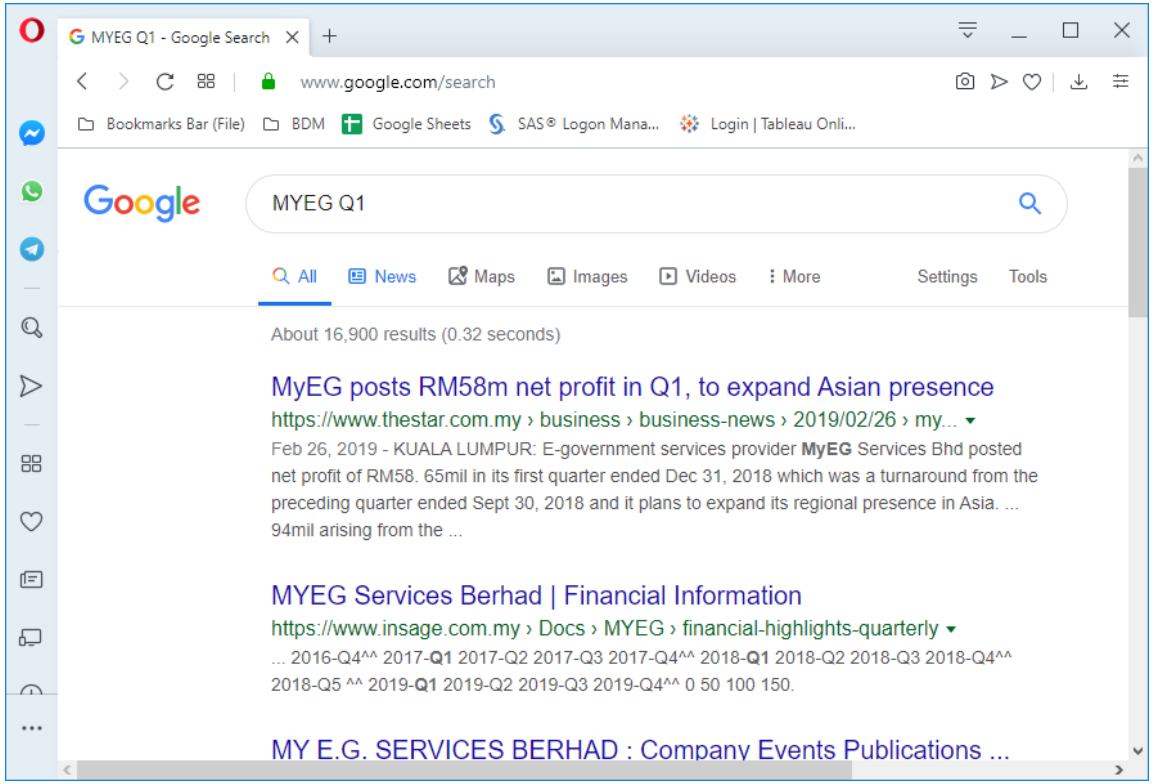
Explore

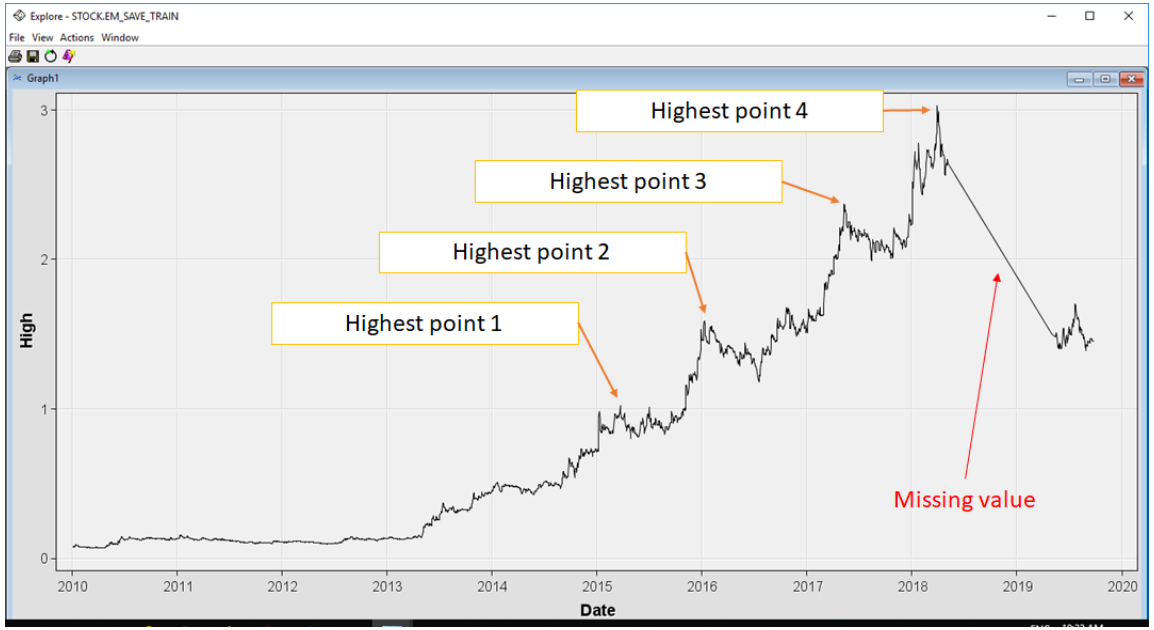
SEMMA node:

- Cluster

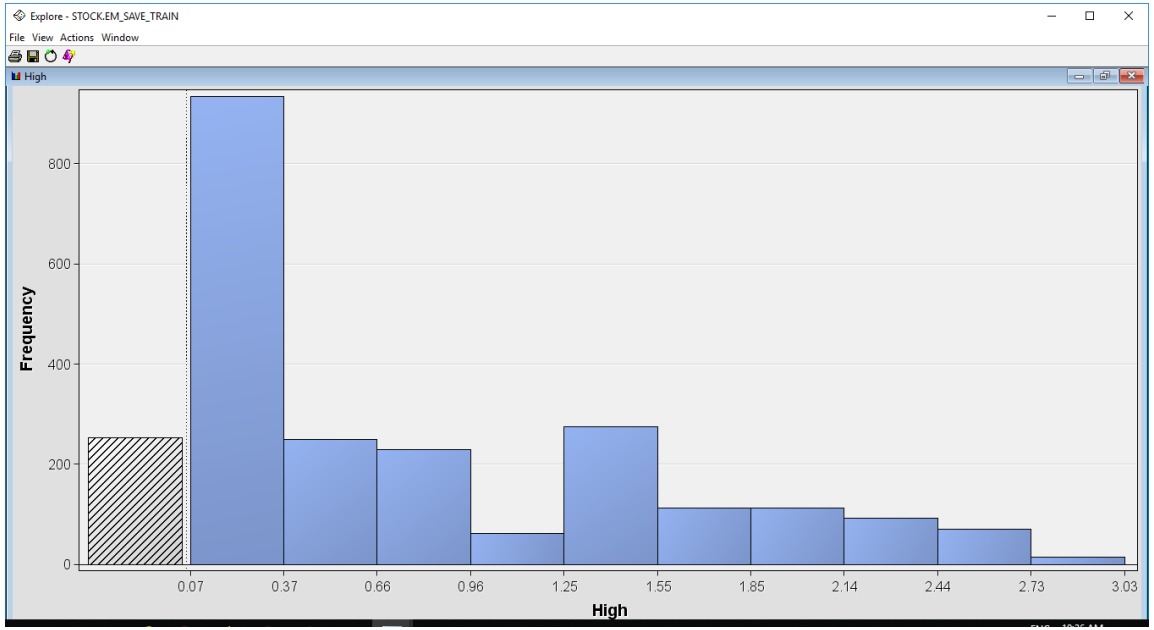
1. No of variables: 7 consist of Date, Volume, Open, High, Low, Close and Adj_Close.
2. No of observations: 2405

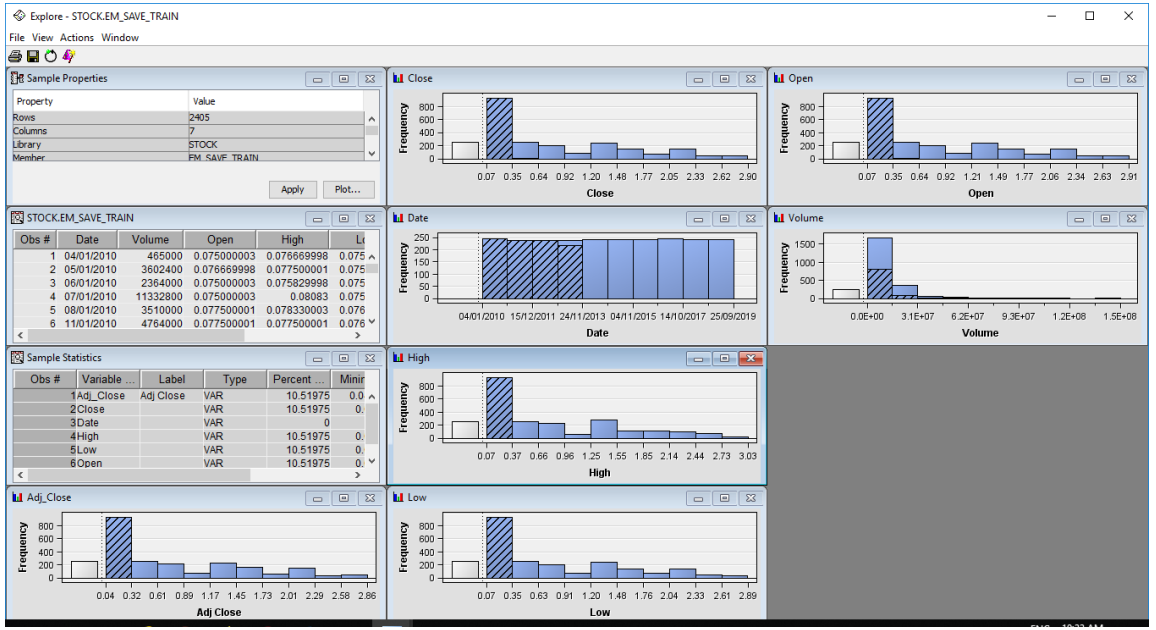






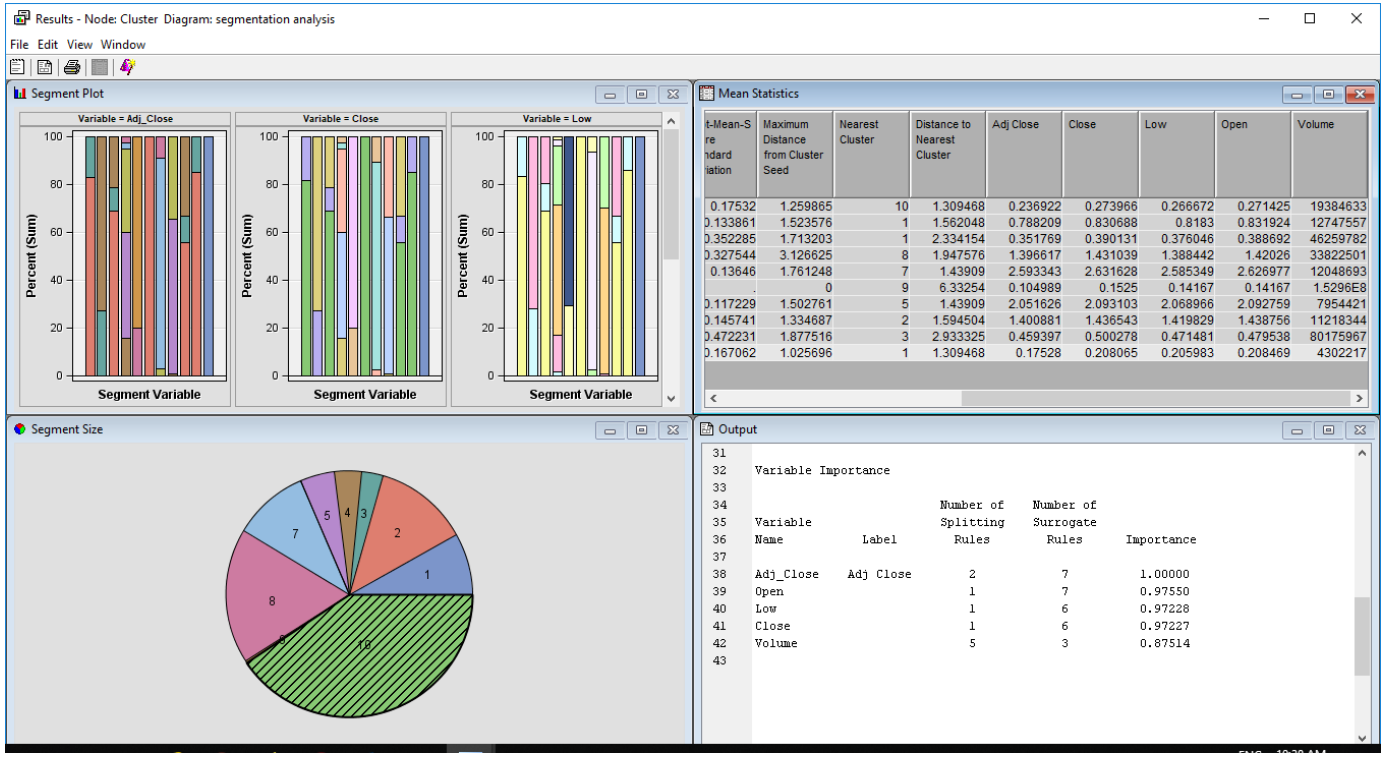
1. Line graph shows highest point for MYEG always in Q1 financial year.
2. Google search shows MYEG will release Financial Report in Q1 every year.
3. There is missing value for MYEG data from mid 2018 till mid 2019.





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.



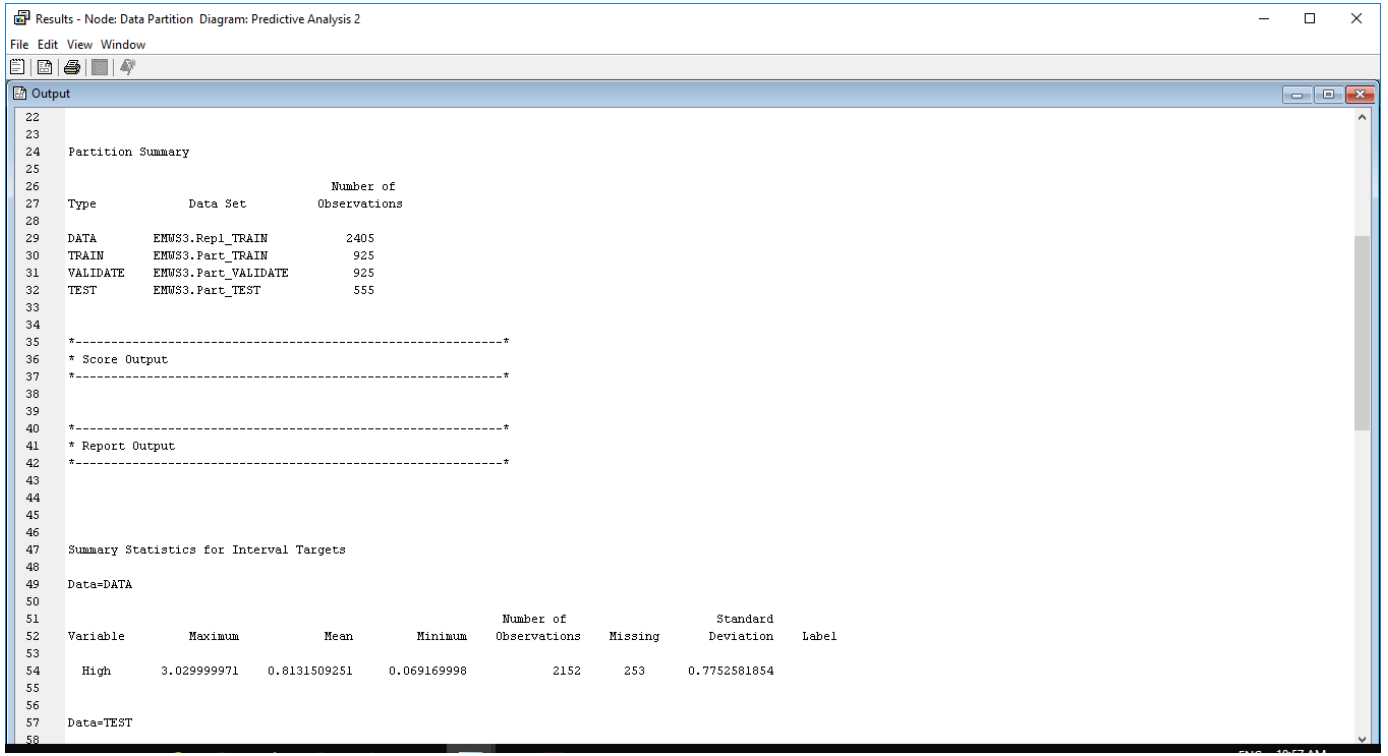
Modify

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**).

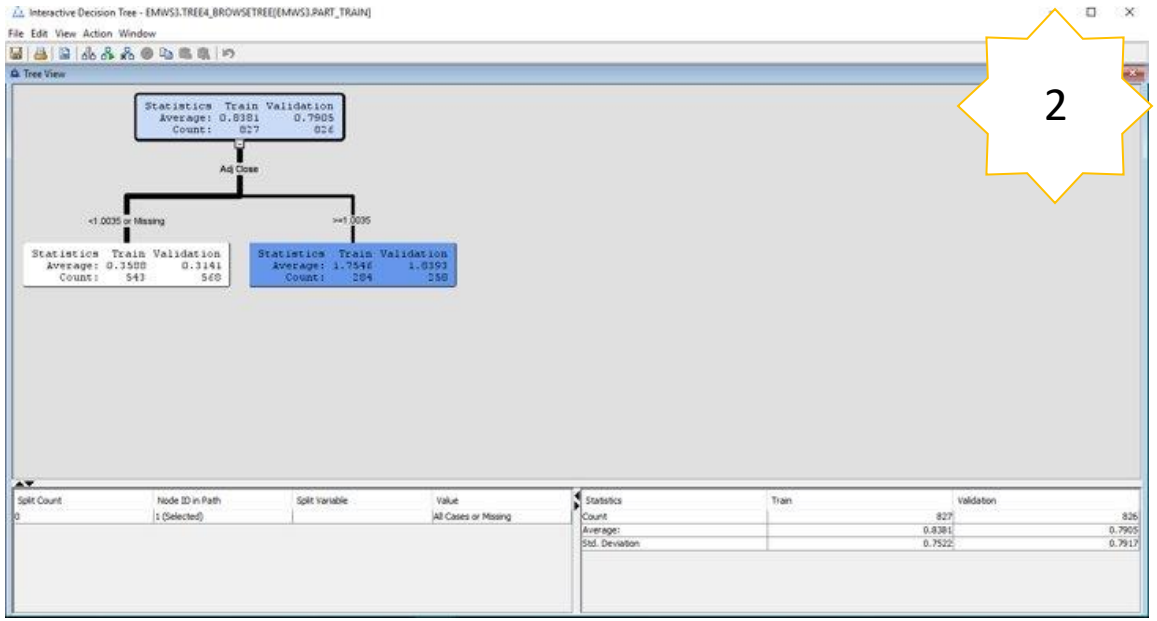


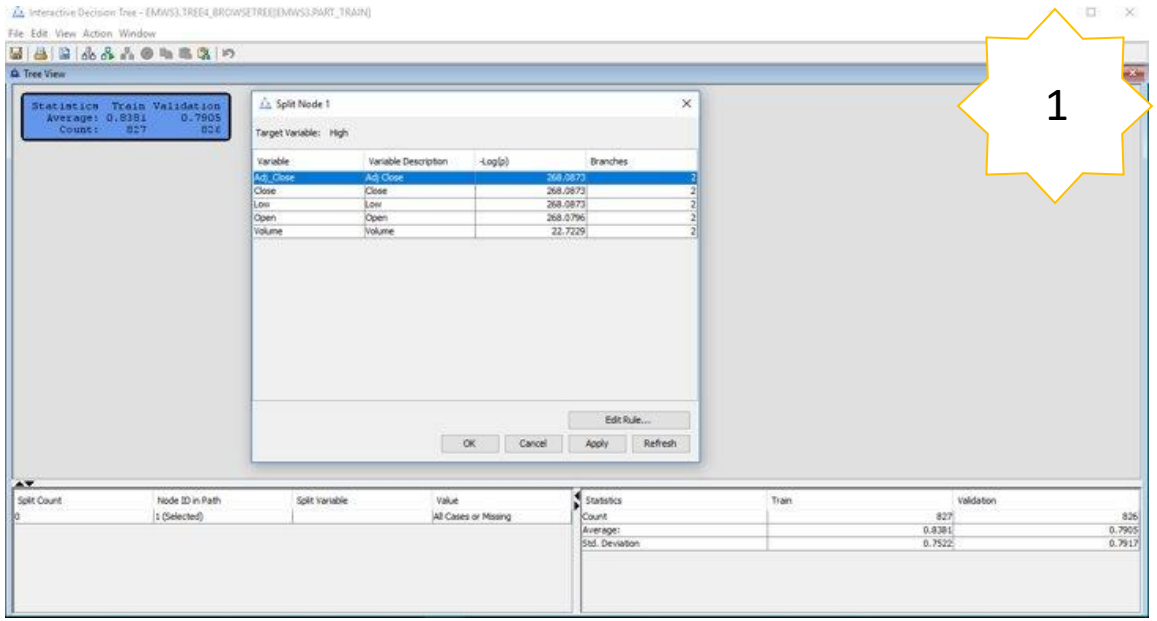

```
Partition Summary

                                    Number of
Type            Data Set          Observations

DATA        EMWS3.Repl_TRAIN          2405
TRAIN       EMWS3.Part_TRAIN           925
VALIDATE    EMWS3.Part_VALIDATE        925
TEST        EMWS3.Part_TEST            555


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


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




Summary Statistics for Interval Targets

Data=DATA

                                                          Number of                    Standard
Variable        Maximum            Mean         Minimum   Observations   Missing      Deviation    Label

  High      3.029999971    0.8131509251     0.069169998           2152      253      0.7752581854


Data=TEST
```

Model

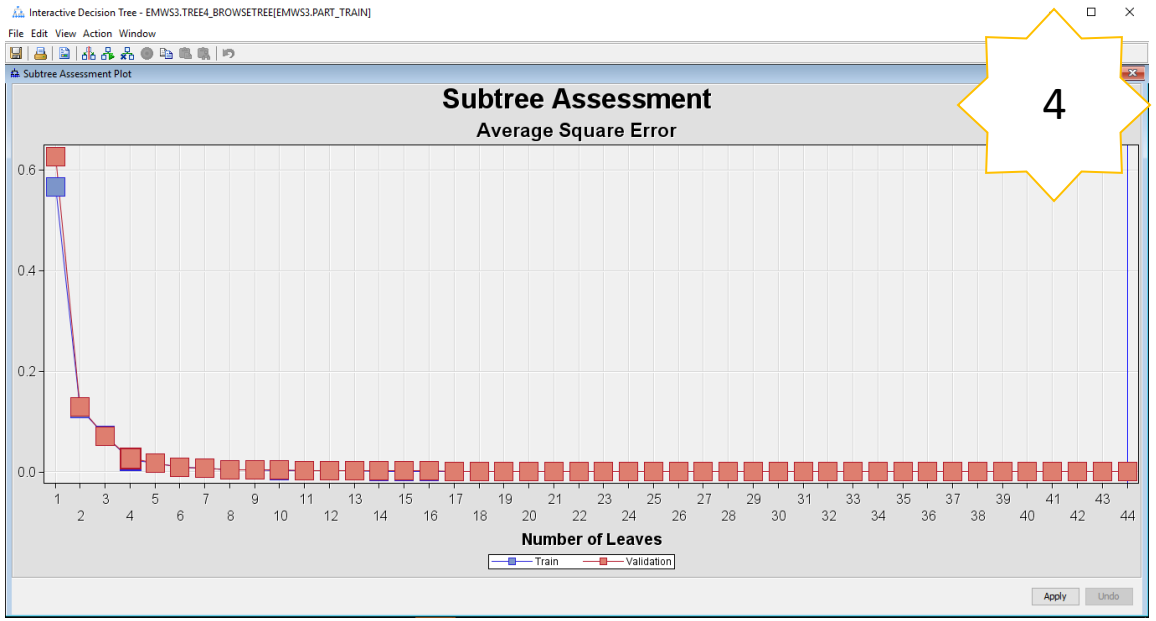
SEMMA node:

- Decision Tree

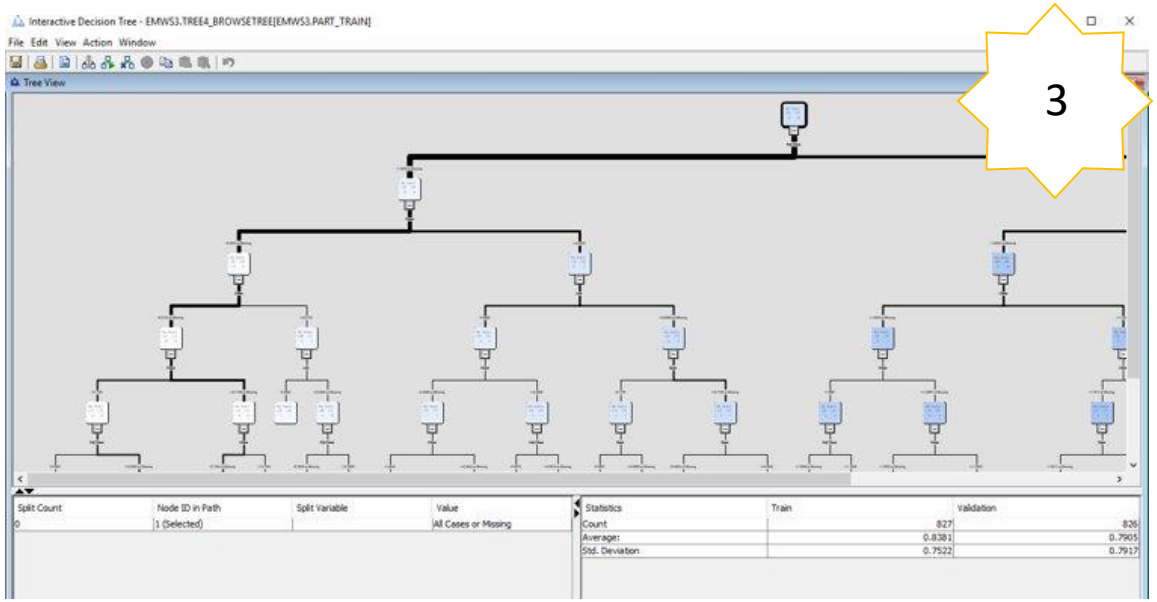




Initial decision tree result shown as per Figure 2.16. <1.0035 or Missing shows higher concentration in Split 1.



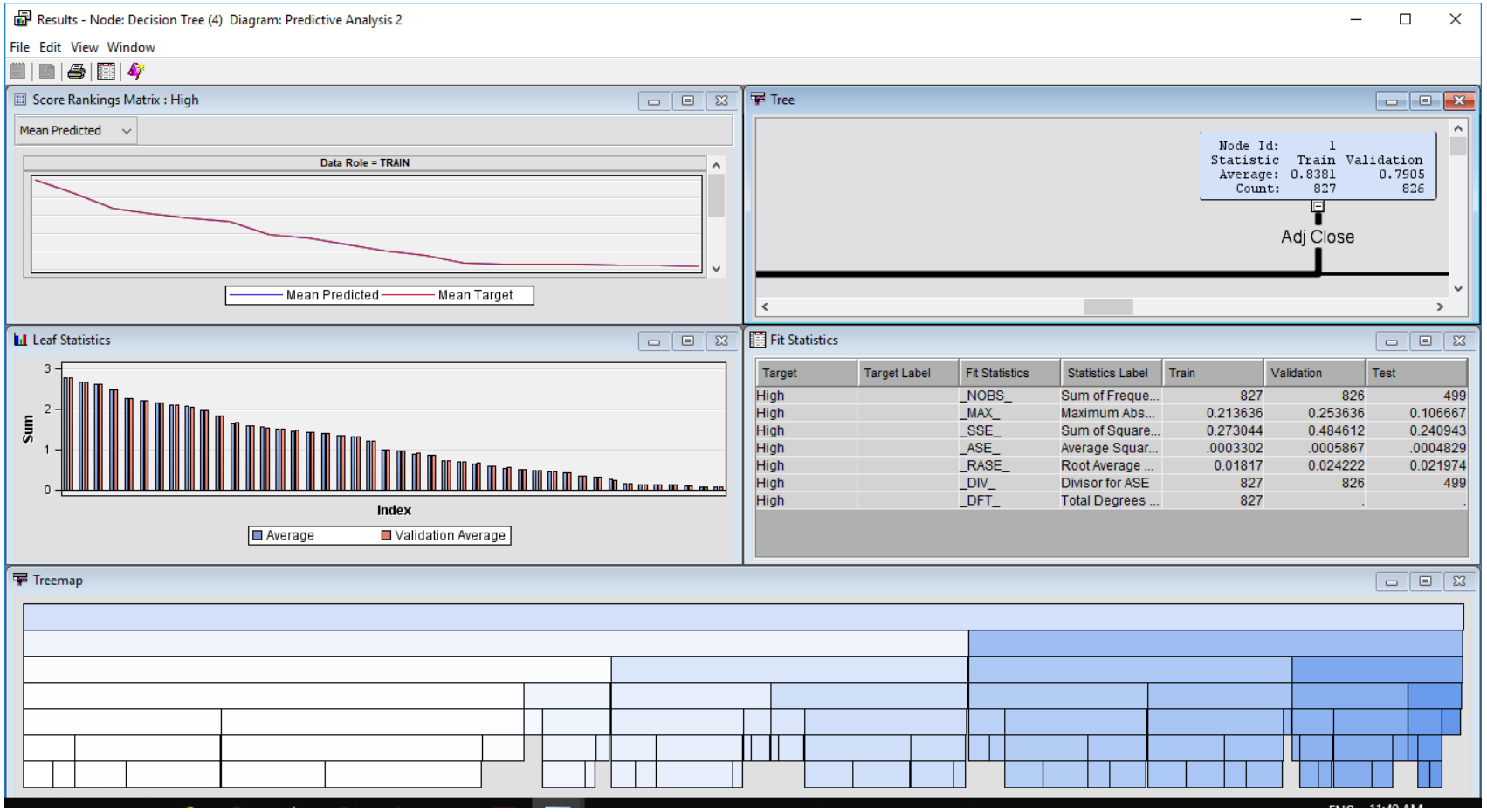
4



3

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

Full result of decision tree.

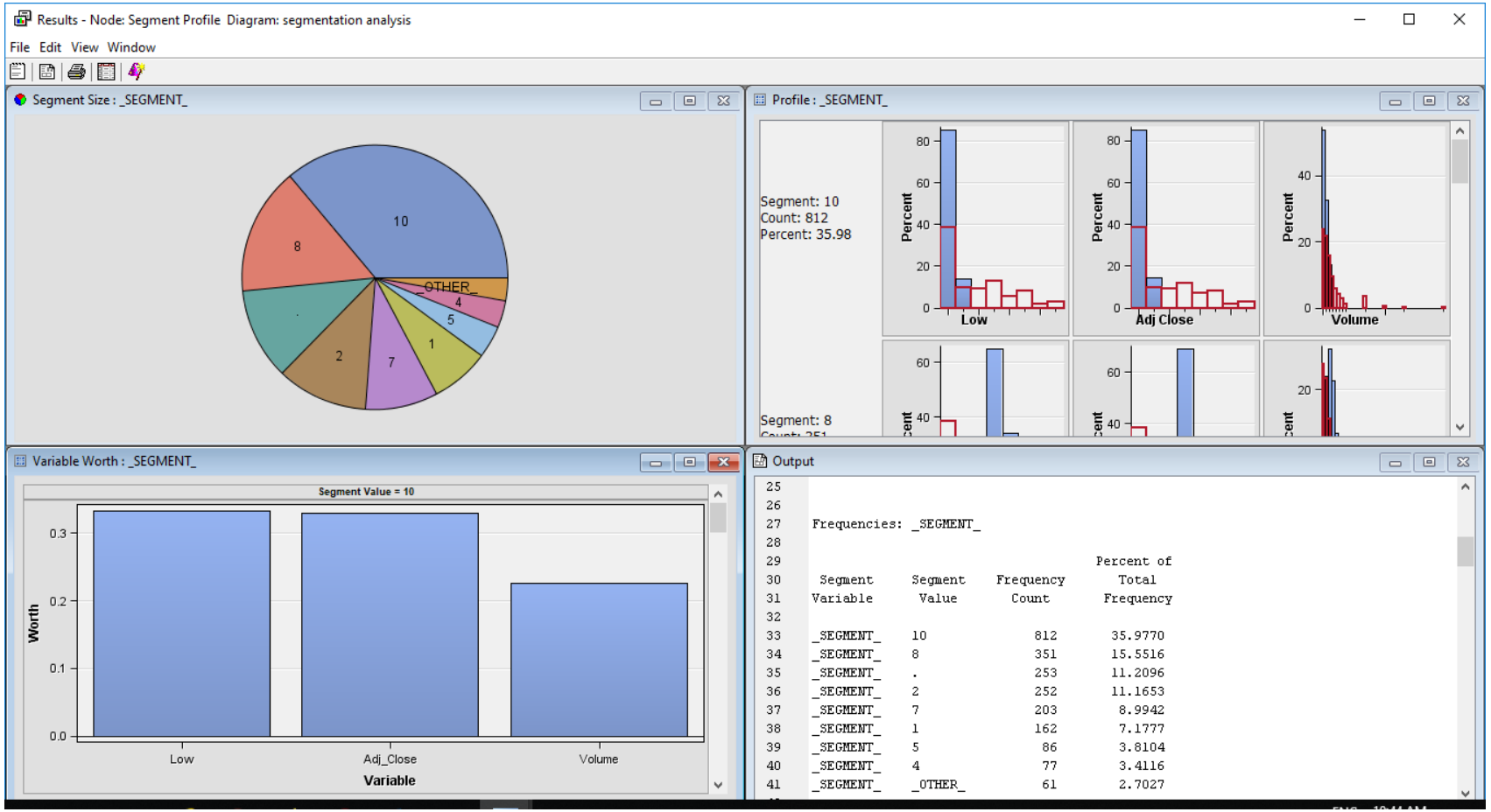


Assess

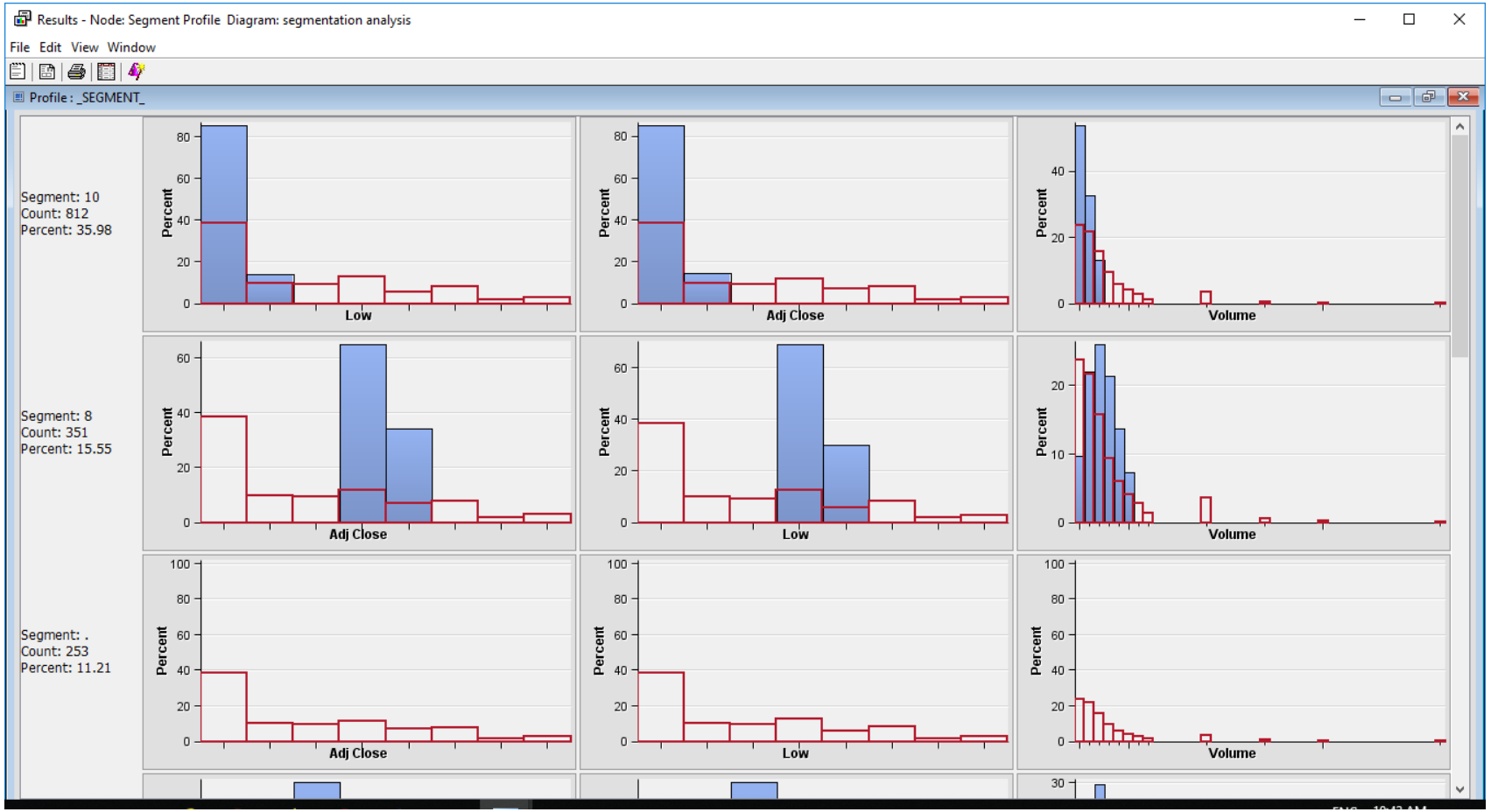
SEMMA node:

- 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.