

# WQD 7005 Data Mining

Milestone 6: Provide users with Recommendation

Ng Kang Wei

WQD170068

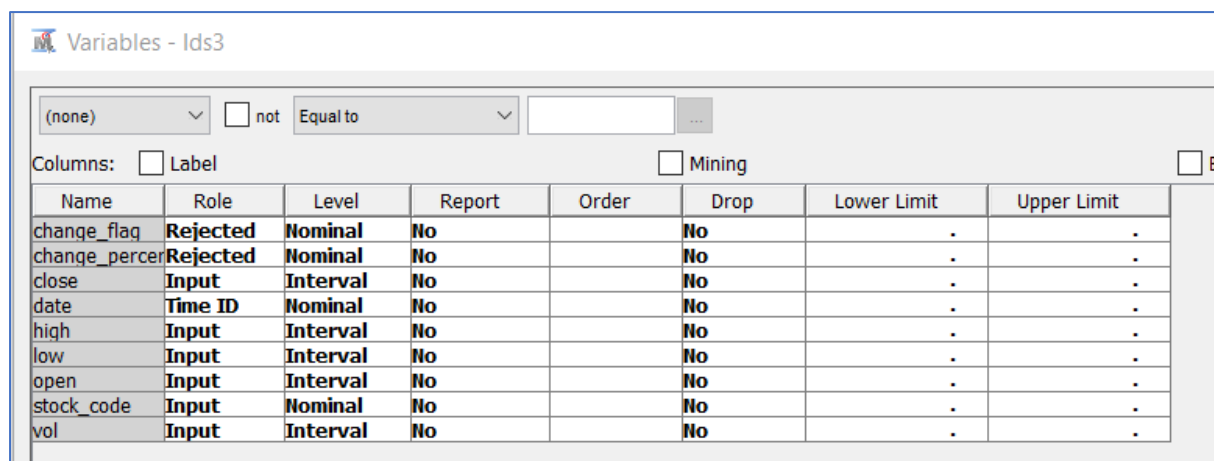
Analysis goal: Group the stocks into clusters. Then find stocks that are similar with Sime Darby Plantation Berhad.

Video URL: <https://youtu.be/yX1MHcYAiIQ>

Github URL: <https://github.com/dante825/dataMining2019>

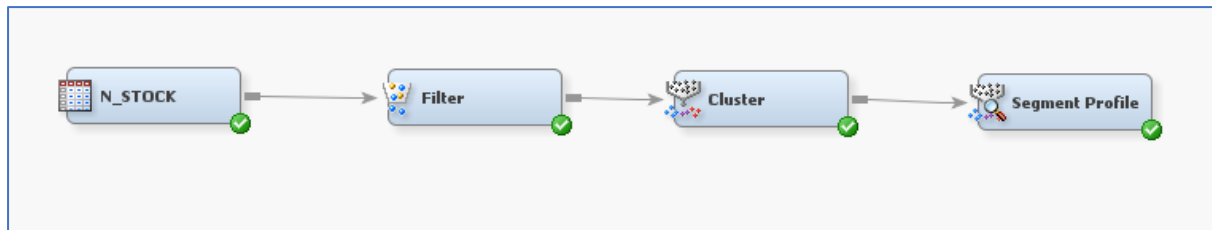
From the previous milestone, a decision tree has been constructed to indicate how should one invest in the stock Sime Darby Plantation Berhad in order to gain profit. The purpose of this milestone is finding some stocks that are similar with Sime Darby Plantation Berhad. The way to do the grouping is via clustering with SAS Enterprise Miner.

All the stocks data collected from before is included in the data. The nominal data is rejected since clustering need interval data to calculate Euclidean distance to obtain the clustering.

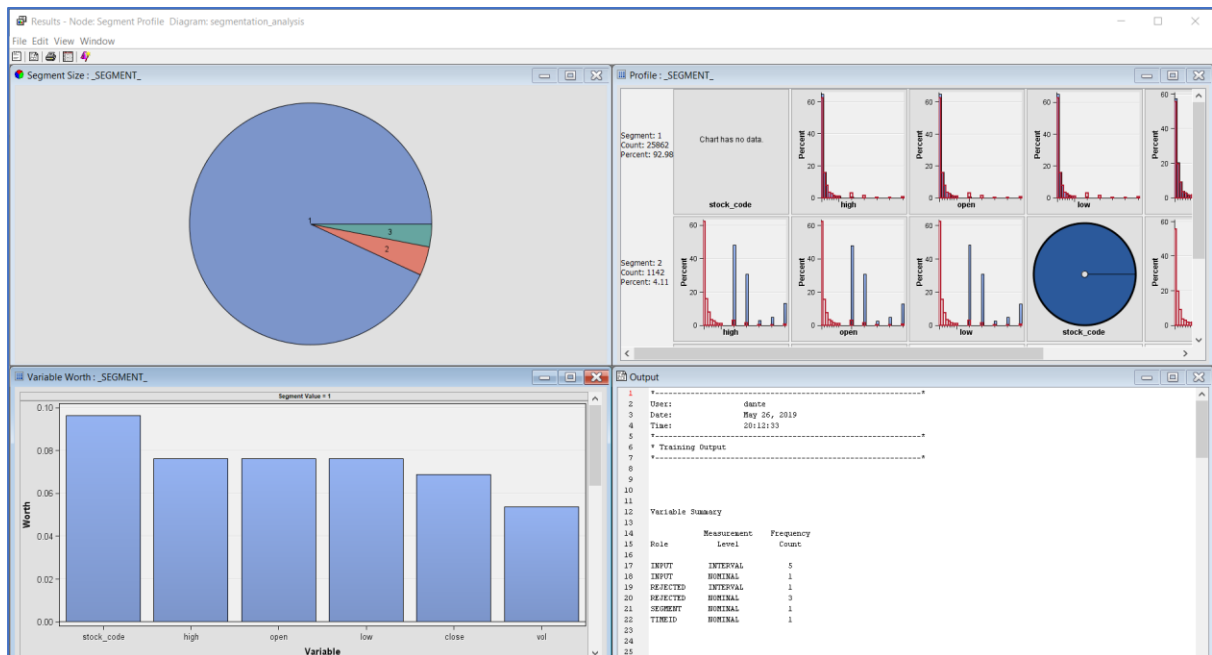


| Name           | Role     | Level    | Report | Order | Drop | Lower Limit | Upper Limit |
|----------------|----------|----------|--------|-------|------|-------------|-------------|
| change_flag    | Rejected | Nominal  | No     |       | No   | .           | .           |
| change_percent | Rejected | Nominal  | No     |       | No   | .           | .           |
| close          | Input    | Interval | No     |       | No   | .           | .           |
| date           | Time ID  | Nominal  | No     |       | No   | .           | .           |
| high           | Input    | Interval | No     |       | No   | .           | .           |
| low            | Input    | Interval | No     |       | No   | .           | .           |
| open           | Input    | Interval | No     |       | No   | .           | .           |
| stock_code     | Input    | Nominal  | No     |       | No   | .           | .           |
| vol            | Input    | Interval | No     |       | No   | .           | .           |

The clustering process in SAS is shown below:



The result of the clustering is shown below:



SAS is using K-means algorithm to group the stocks. The K or number of clusters is determined automatically by SAS. The result of the clustering show that there are 3 clusters among the stocks. The cluster membership count is as follows:

| Frequencies: _SEGMENT_ |               |                 |                            |
|------------------------|---------------|-----------------|----------------------------|
| Segment Variable       | Segment Value | Frequency Count | Percent of Total Frequency |
| _SEGMENT_              | 1             | 25862           | 92.9753                    |
| _SEGMENT_              | 2             | 1142            | 4.1056                     |
| _SEGMENT_              | 3             | 812             | 2.9192                     |

The first cluster has the most members, 92.9753 % of the stocks and the 2<sup>nd</sup> cluster and 3<sup>rd</sup> cluster has 4.1056% and 2.9192% of the stocks respectively.

The most important variables used to segment the stocks are high and open, except in segment 3. The most important variable in segment 3 is volume.

Variable: \_SEGMENT\_ Segment: 1 Count: 25862  
Decision Tree Importance Profiles

| Variable   | Worth    | Rank |
|------------|----------|------|
| stock_code | 0.096425 | 1    |
| high       | 0.076229 | 2    |
| open       | 0.076192 | 3    |
| low        | 0.076168 | 4    |
| close      | 0.068896 | 5    |
| vol        | 0.053737 | 6    |

Variable: \_SEGMENT\_ Segment: 2 Count: 1142  
Decision Tree Importance Profiles

| Variable   | Worth    | Rank |
|------------|----------|------|
| high       | 0.077012 | 1    |
| open       | 0.076744 | 2    |
| low        | 0.076651 | 3    |
| stock_code | 0.074736 | 4    |
| close      | 0.069482 | 5    |
| vol        | 0.001187 | 6    |

Variable: \_SEGMENT\_ Segment: 3 Count: 812  
Decision Tree Importance Profiles

| Variable   | Worth    | Rank |
|------------|----------|------|
| vol        | 0.055374 | 1    |
| stock_code | 0.023641 | 2    |
| low        | 0.001003 | 3    |
| open       | 0.001001 | 4    |
| high       | 0.000992 | 5    |
| close      | 0.000468 | 6    |

The objective of this analysis is to find stocks in the same group with Sime Darby Plantation Berhad. From the result of the clustering (shown below), Sime Darby Plantation Berhad (5285) is in cluster 2.

| Segment Variable | Segment Value    | Variable   | Value  | Frequency Count | Percent  | Group Index |
|------------------|------------------|------------|--------|-----------------|----------|-------------|
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5568   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5568   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5533   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5533   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5517   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5517   | 4               | 0.015467 | 1           |
| SEGMENT          | 2                | stock code | 5517   | 24              | 2.101576 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5436   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5436   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5401WA | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5401WA | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5401   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5401   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5386WE | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5386WE | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5386   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5386   | 27              | 0.1044   | 1           |
| SEGMENT          | 3                | stock code | 5386   | 1               | 0.123153 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5371   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5371   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5355   | 25              | 0.089876 | 0           |
| SEGMENT          | 1                | stock code | 5355   | 25              | 0.090667 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5347   | 28              | 0.100661 | 0           |
| SEGMENT          | 2                | stock code | 5347   | 28              | 2.451839 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5289   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5289   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5288   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5288   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5286   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5286   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5285   | 28              | 0.100661 | 0           |
| SEGMENT          | 2                | stock code | 5285   | 27              | 2.364273 | 1           |
| SEGMENT          | 3                | stock code | 5285   | 1               | 0.123153 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5284   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5284   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5283WA | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5283WA | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5283   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5283   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5281   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5281   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5280   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5280   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5279   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5279   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5278   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5278   | 28              | 0.108267 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5277   | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5277   | 25              | 0.090667 | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5277   | 3               | 0.369458 | 1           |
| SEGMENT          | 3                | stock code | 5277   | 0               | 0.100661 | 0           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5276WA | 28              | 0.100661 | 0           |
| SEGMENT          | 1                | stock code | 5276WA | 27              | 0.1044   | 1           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5276WA | 1               | 0.123153 | 1           |
| SEGMENT          | 3                | stock code | 5276WA | 28              | 0.100661 | 0           |
| OVERALL          | EMWS2 Clus TRAIN | stock code | 5276   | 28              | 0.1044   | 1           |
| SEGMENT          | 1                | stock code | 5276   | 27              | 0.1044   | 1           |

Some of the other stocks that are in cluster 2 are shown as follows:

| Segment Variable | Segment Value | Variable   | Value  | Frequency Count | Percent  | Group Index |
|------------------|---------------|------------|--------|-----------------|----------|-------------|
| SEGMENT          | 2             | stock code | 0097   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 1023   | 15              | 1.313485 | 1           |
| SEGMENT          | 2             | stock code | 1066   | 23              | 2.014011 | 1           |
| SEGMENT          | 2             | stock code | 1062   | 19              | 1.663748 | 1           |
| SEGMENT          | 2             | stock code | 1155   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 1163   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 1783   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 1818   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 1869   | 22              | 1.926445 | 1           |
| SEGMENT          | 2             | stock code | 1929   | 17              | 1.488616 | 1           |
| SEGMENT          | 2             | stock code | 2291   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 2593   | 28              | 2.770708 | 1           |
| SEGMENT          | 2             | stock code | 2658   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 3034   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 3042   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 3182   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 3301   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 3816   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 3807   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 4006   | 28              | 2.770708 | 1           |
| SEGMENT          | 2             | stock code | 4065   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 4324   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 4448   | 8               | 0.709525 | 1           |
| SEGMENT          | 2             | stock code | 4588   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 4731   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 5014   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 5031   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 5095   | 6               | 0.525394 | 1           |
| SEGMENT          | 2             | stock code | 5139   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 5168   | 11              | 0.963222 | 1           |
| SEGMENT          | 2             | stock code | 5183   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 5225   | 26              | 2.276708 | 1           |
| SEGMENT          | 2             | stock code | 5235SS | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 5289   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 5347   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 5517   | 24              | 2.101576 | 1           |
| SEGMENT          | 2             | stock code | 6012   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 6033   | 27              | 2.364273 | 1           |
| SEGMENT          | 2             | stock code | 6139   | 2               | 0.175131 | 1           |
| SEGMENT          | 2             | stock code | 6351   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 7084   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 7090   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 7765   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 8621   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 8334   | 28              | 2.451839 | 1           |
| SEGMENT          | 2             | stock code | 8622   | 27              | 2.364273 | 1           |
| SEGMENT          | 1             | stock code | 0002   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0008   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0012   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0021   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0029   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0029WB | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0037   | 28              | 0.108267 | 1           |
| SEGMENT          | 1             | stock code | 0041   | 11              | 0.042533 | 1           |
| SEGMENT          | 1             | stock code | 0043   | 17              | 0.065734 | 1           |

To list a few examples:

- CIMB Group Holdings Berhad (1023)
- RHB Bank Berhad (1066)
- Genting Plantation Berhad (2291)
- Hap Seng Consolidated Berhad (3034)
- QL Resources Berhad (7084)

These stocks are in the same cluster with Sime Darby Plantation Berhad. The decision tree in milestone 5 should be able to apply on these stocks as well. I recommend investors who are interested in the stock of Sime Darby Plantation Berhad to look into these stocks as well.