

**STOCK MARKET PREDICTION**

**NURHAMIDAH FARHANA BINTI ZULKARNAIN**

**WQD180058**

**WQD7005 DATA MINING**

**LECTURER: ASSOCIATE PROF. DR. TEH YING WAH**

**FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**UNIVERSITY OF MALAYA**

**2019**

1. **Introduction**

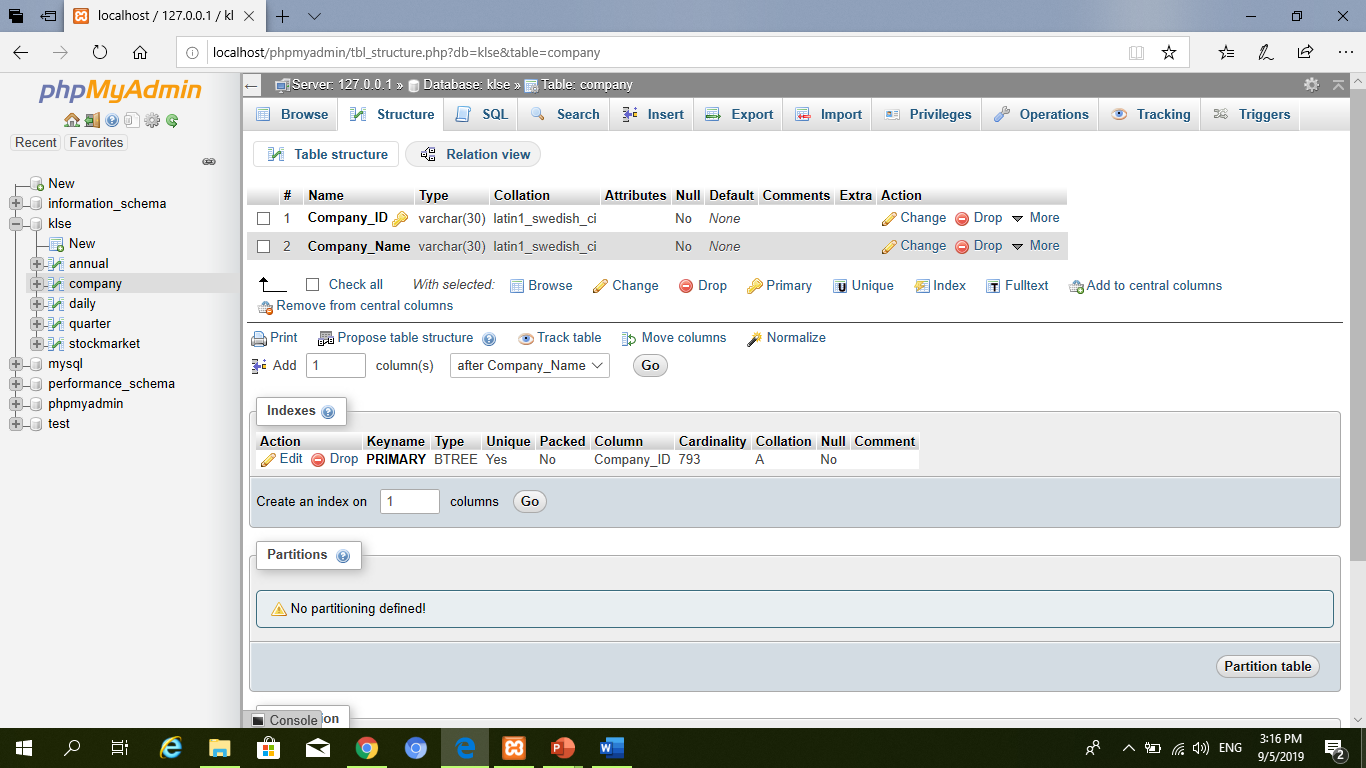
As a data scientist, there are three key skills needed, which we refer to as creating, connecting, and computing. For our Data Mining course, we are required to solve a real-world problems which is the stock market analysis by applying those three key skills. As the definition of data mining itself, we need to find the hidden information which might be valuable for the stock market data by using various techniques (Wah 2006). The stock market data will be taken from The Star Online Stock Market news in which the KLSE stock data is updated every 5 minutes. We need to crawl the stock market for several times everyday in order to observe for the pattern. There are 6 milestones to be accomplished and act as the guie to go through the analysis of the stock market as in **Figure 1**.

**Figure 1: The six milestones that need to be accomplished to analyse the stock market**

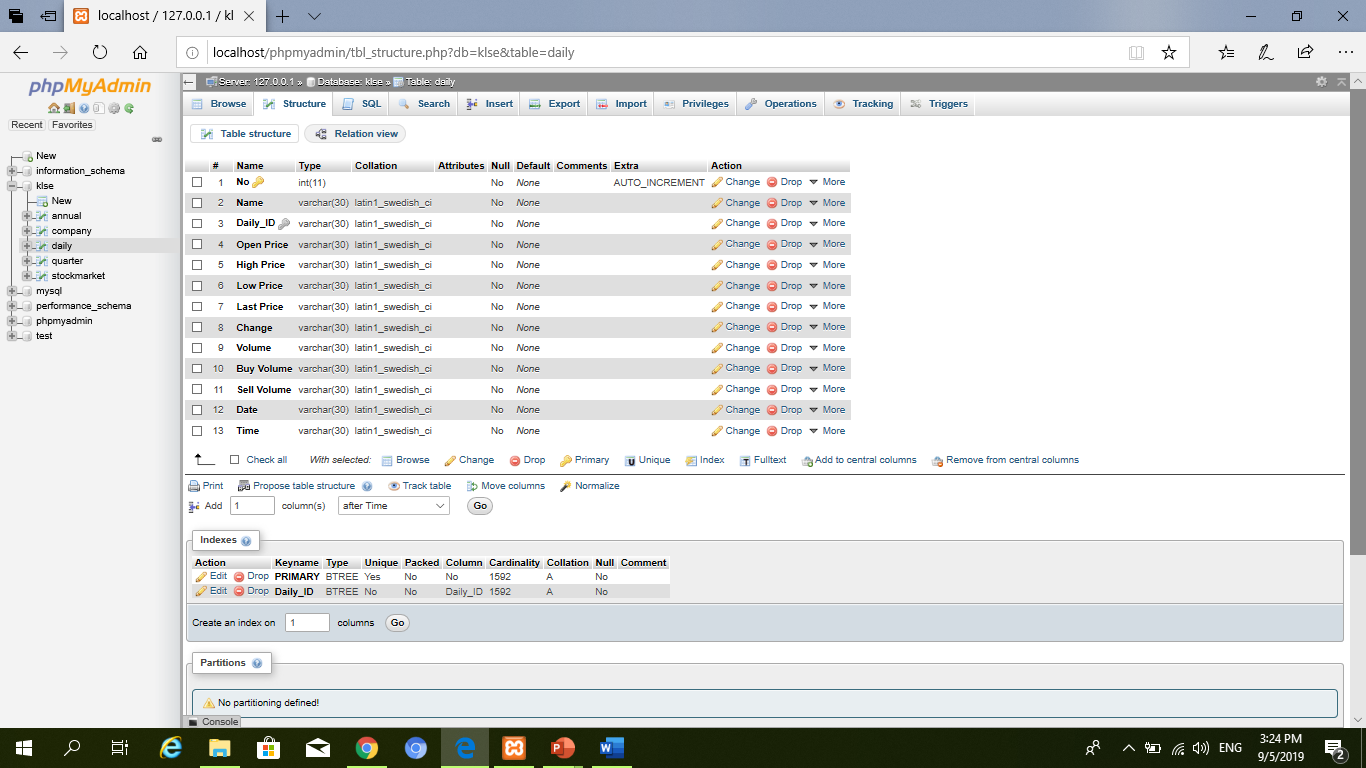
1. **Objective**
2. To crawl data related to stock market
3. To investigate the trends in stock market
4. To give recommendations around optimal actions to achieve investment objectives
5. **Methodology and Analysis**

In Milestone 1, we are required to crawl the data. We are considering to crawl 796 companies listed in Main Market Bursa Malaysia. The list of the companies can be obtained from this link: http://www.bursamalaysia.com/market/listed-companies/list-of-companies/main-market/. The list of companies were save in .csv file as “KLSE.csv” in which contains the name and symbol of each company. By using Spyder (Python 3.7), the stock ticker’s symbol from “KLSE.csv” is used to get data from The Star Online Stock Market news. As a group of four student, we divided the task to crawl the data by quarter in which a person will crawl in the morning, another person will crawl in the noon, the other person will crawl in the evening and the last person will crawl at night. It took about 15-20 minutes to complete one crawl. The data was crawled from 7 March 2019 until 20 March 2019 with excluded of Saturday and Sunday as the stock market is closed for business during weekend. All crawled data were stored in shared Google Drive to enable everyone to access it.

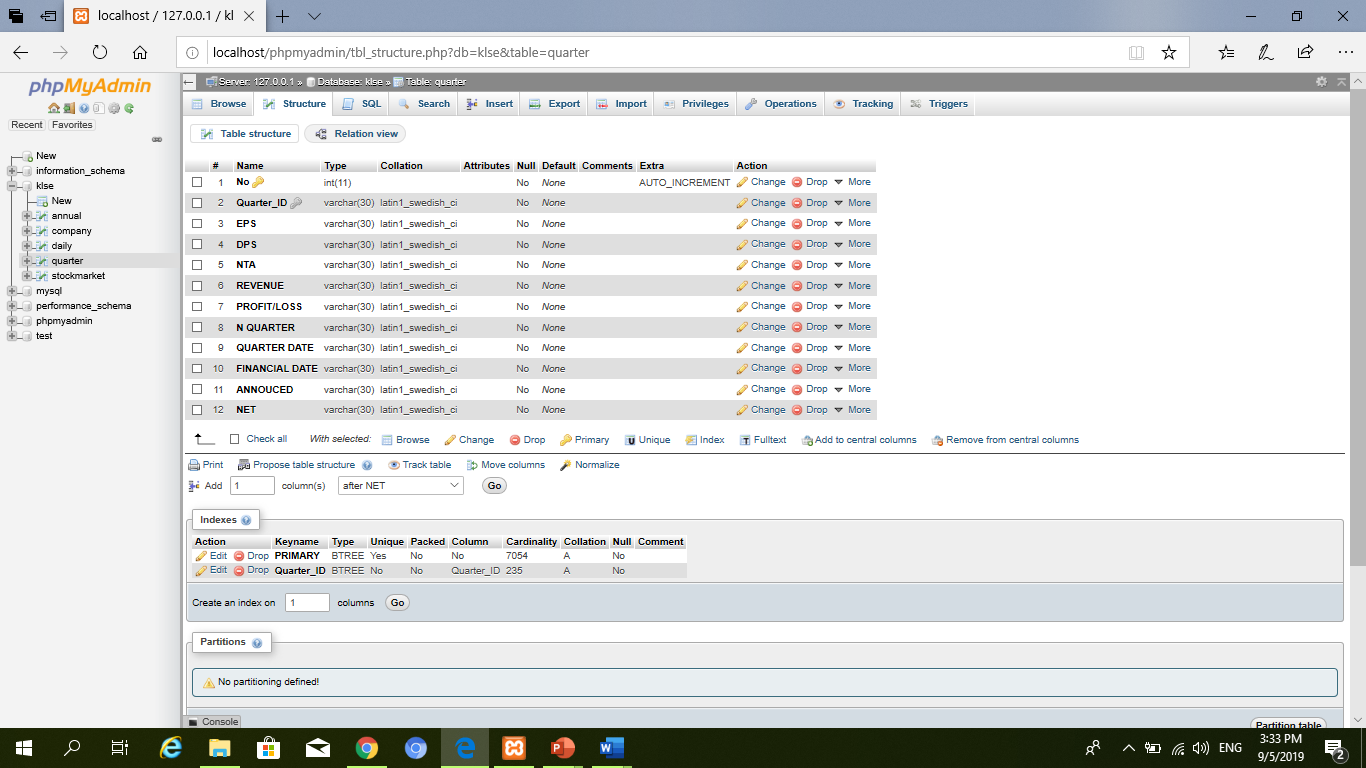
In Milestone 2, more data were crawled to gain as many information about the company listed in the stock market. The annual report and quarterly report for each company was crawled to gain more insights. All data were then stored in MySQL through XAMPP (see **Figure 2, Figure 3, Figure 4, Figure 5** and **Figure 6**) and combined by using Star Schema in order to see the connection between each crawled data (**Figure 7**). The fact table named as klse\_stockmarket was drill down into four dimension tables which are klse\_company, klse\_daily, klse\_quarter and klse\_annual with four foreign keys which are Company\_ID, Daily\_ID, Quarter\_ID and Annual\_ID respectively.



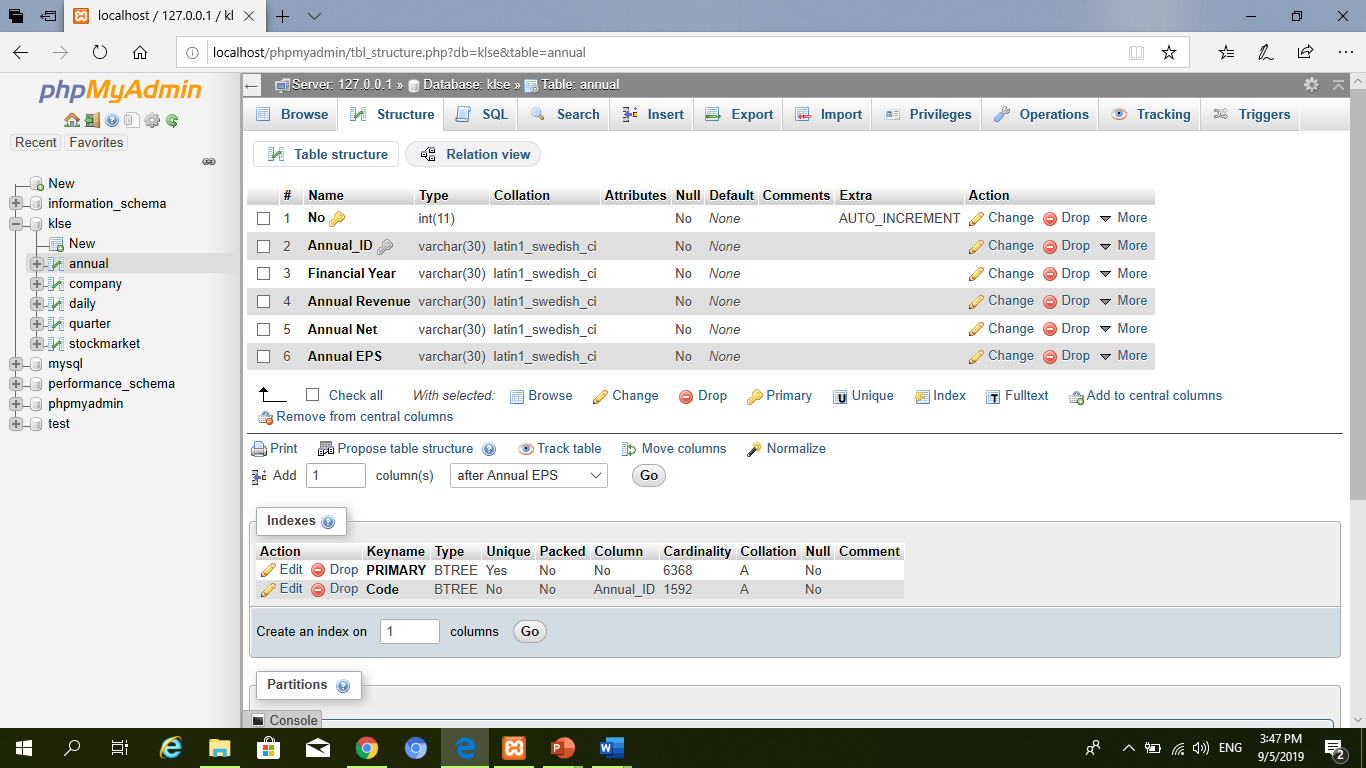
**Figure 2: Table structure of company details (klse\_company)**



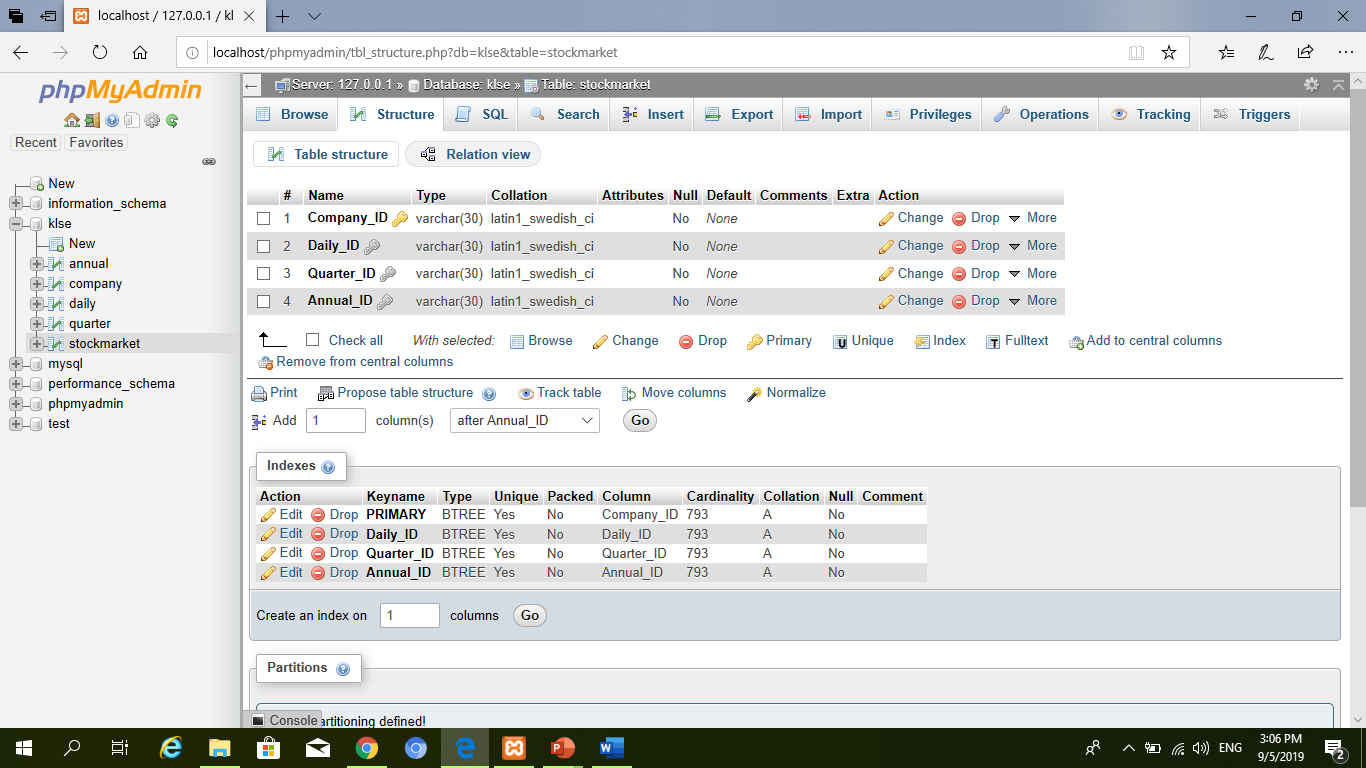
**Figure 3: Table structure for daily crawled data (klse\_daily)**



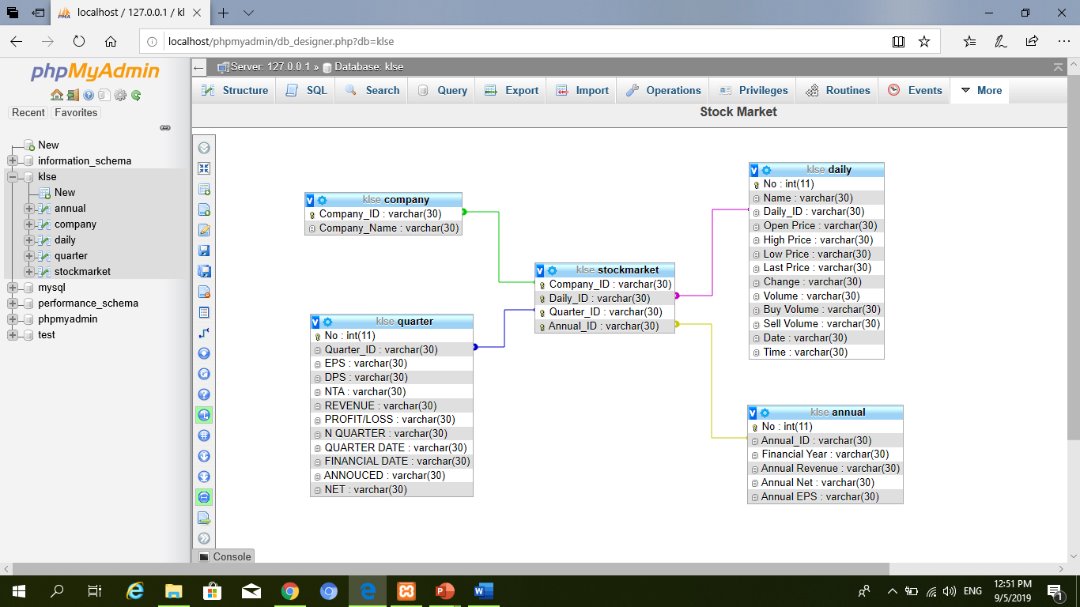
**Figure 4: Table structure for Quarterly Report (klse\_quarter)**



**Figure 5: Table structure for Annual Report (klse\_annual)**



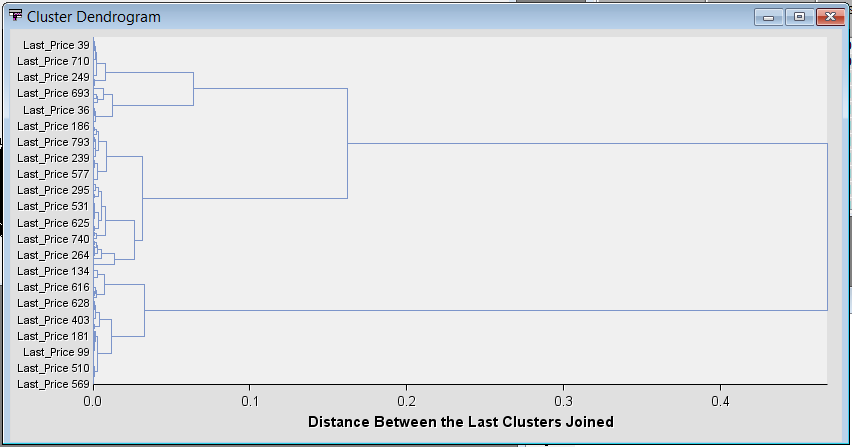
**Figure 6: Table structure of stock market which is the combination of all crawled data (klse\_stockmartket)**



**Figure 7: The Star Schema for Stock Market**

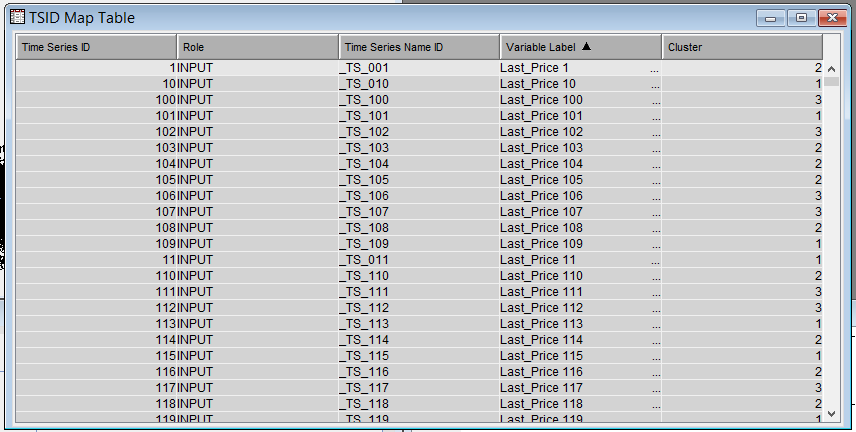
In Milestone 3, all data collected were processed. Data scientists spend most of their time exploring, cleaning and preparing their data for modelling. This helps them to build accurate models and check assumptions required for fitting models. Data pre-processing involves several phase which are data cleaning, data integration, data reduction and data transformation. Some of the steps taken for data pre-processing in this project including handling missing values, adding new column, deleting unwanted column and duplicate data, and also changing the date format. All steps for data pre-processing were did in Python 3.7.

In Milestone 4, SAS Enterprise Miner which is the SAS solution for data mining was used to further processing the data. SAS Enterprise Miner helps in streamlining the data mining process by creating a highly accurate predictive and descriptive models. Enterprise Miner’s works in a process flow diagram which means that manual coding is unnecessary and it also reduces the model development time for both business analysts and statisticians. However, user can still integrate their code and build new nodes for redistribution as the system is customizable and extensible. For the interpretation of data, the data was analysed by using clustering model in TS Similarity. **Figure 8** shows the dendrogram for the data based on hierarchical clustering.



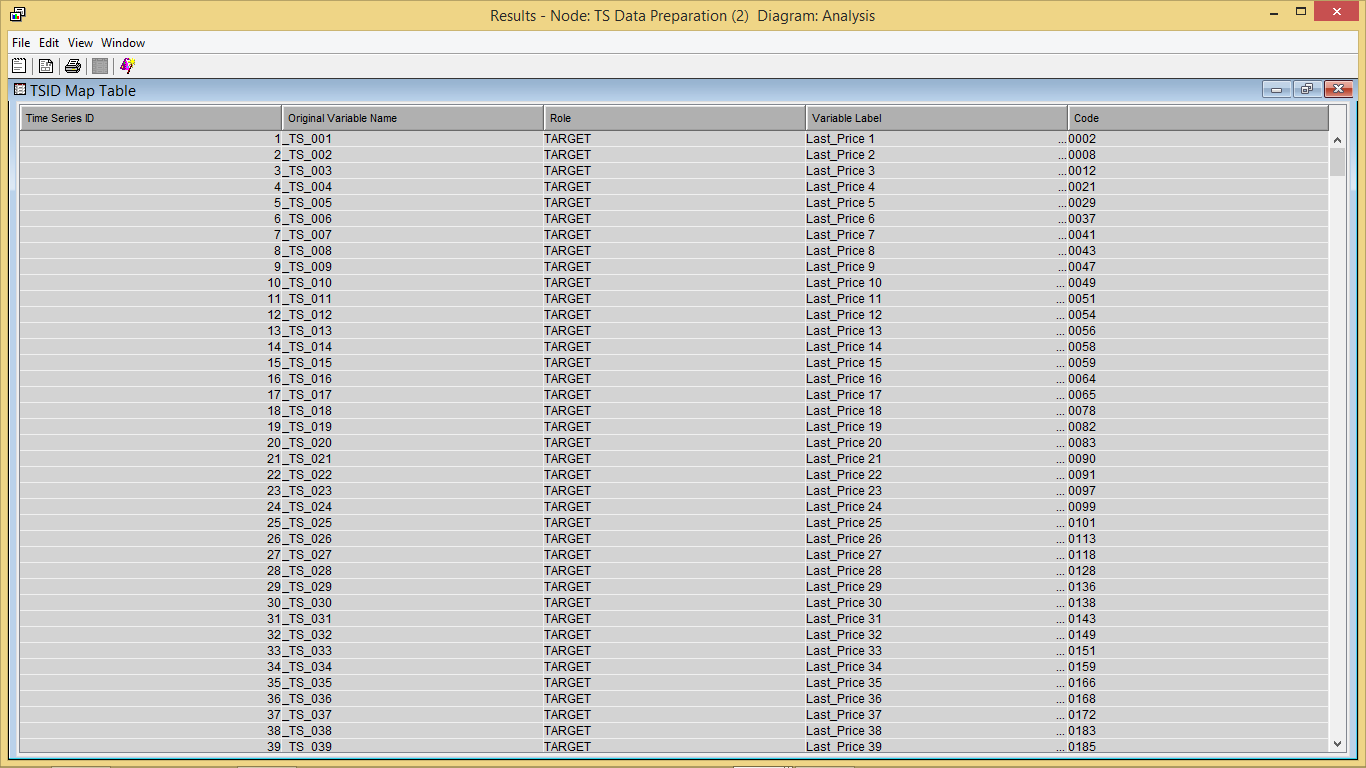
**Figure 8: Cluster dendrogram based on the column Last\_Price**

In order to know the exact number of cluster that the data belongs to, we observe the TSID Map Table to identify the number of clusters. Based on the **Figure 9** below, it shows that the data was divided into three clusters.

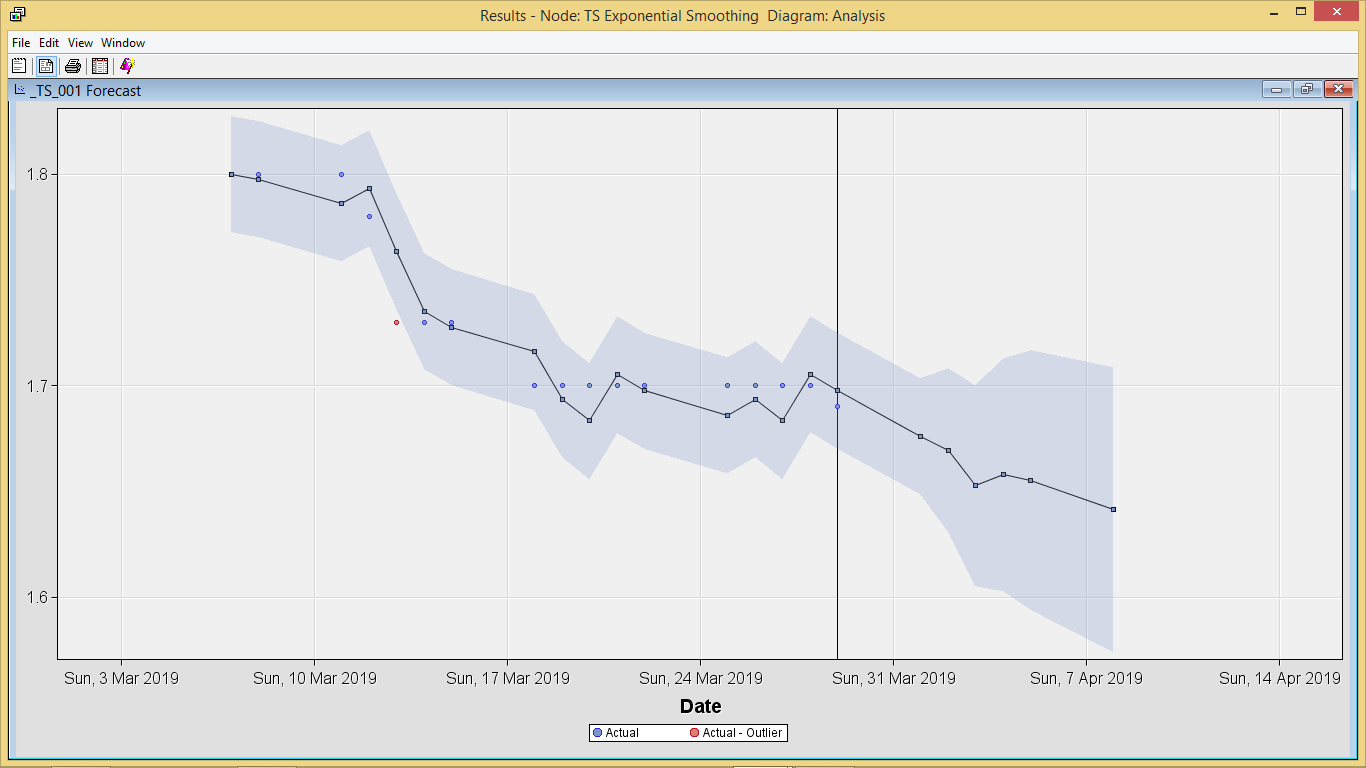


**Figure 9: TSID Map table shows the number of clusters for each TS ID**

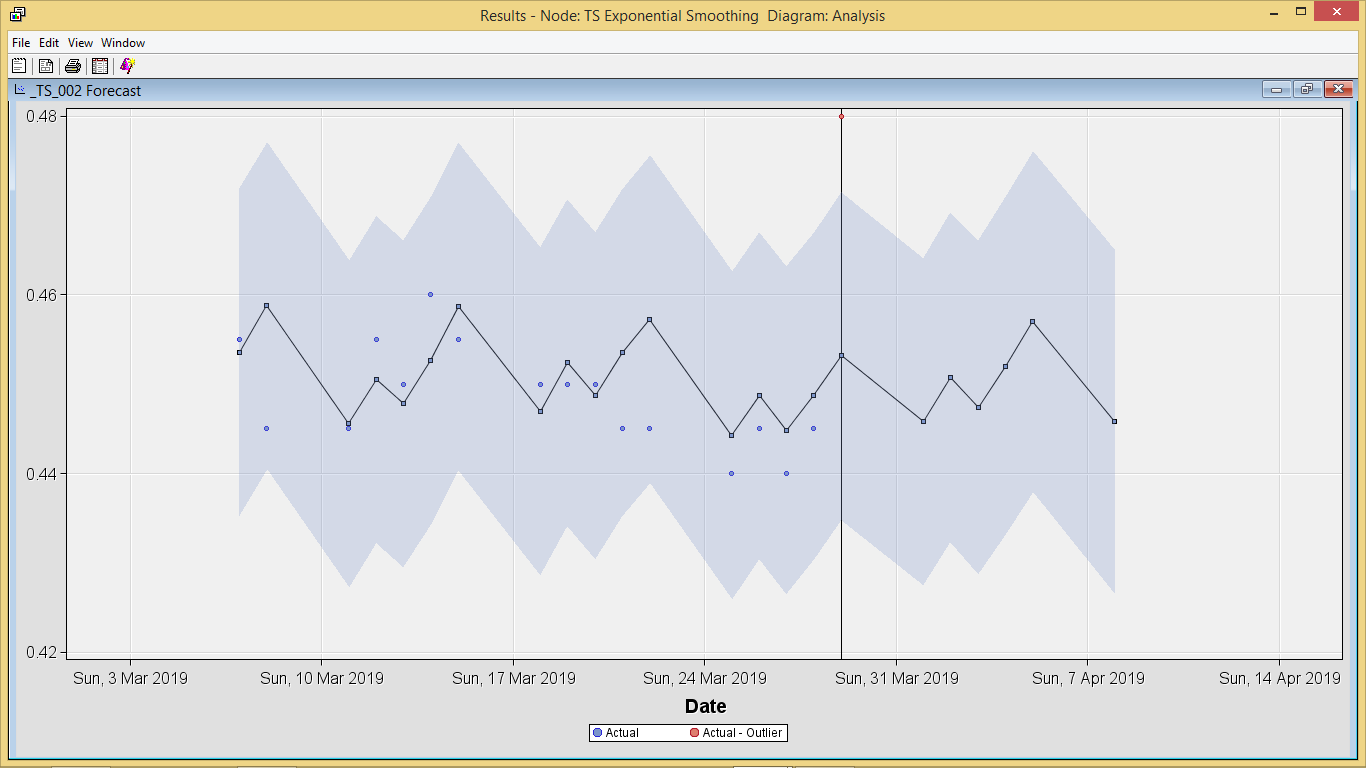
In milestone 5, we are using the Time Series (TS) node. It is a method of investigating time series data which belongs to the Modify category of the SAS SEMMA (sample, explore, modify, model, assess) data mining process. The TS node enables us to understand the trends and seasonal variation in large amounts of time series and transactional data. The analysis that can be performed by TS node are time series analysis, forecasting and also transactional data. For this project, I only set the forecast until the next 6 days. Based on the forecasting, it shows that some companies are going down, some companies are going up and some other companies did not show any drastic changes. As per **Figure 10** below, it shows the description for each company (Code) which is assigned to a new TS ID. **Figure 11, Figure 12** and **Figure 13** shows the forecasting for three selected companies which is TS\_001, TS\_002 and TS\_003 respectively. All three companies have different trend of forecasting. TS\_001 have the trend of moving downwards, meanwhile TS\_002 did not shows any drastic changes neither moving downwards nor upwards. However, TS\_003 shows the trend of moving upwards.



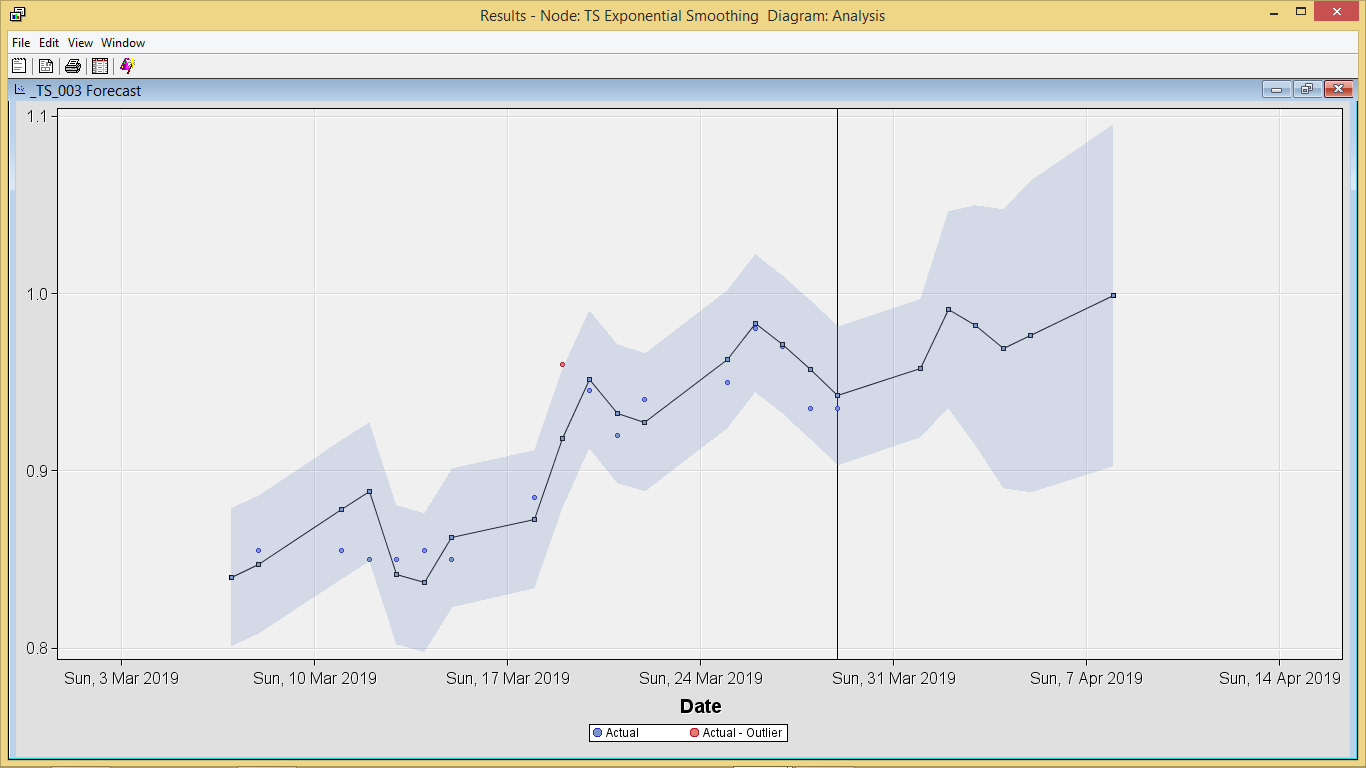
**Figure 10: Description for each company (Code) which is assign to a new TS ID**



**Figure 11: Forecasting for TS\_001**

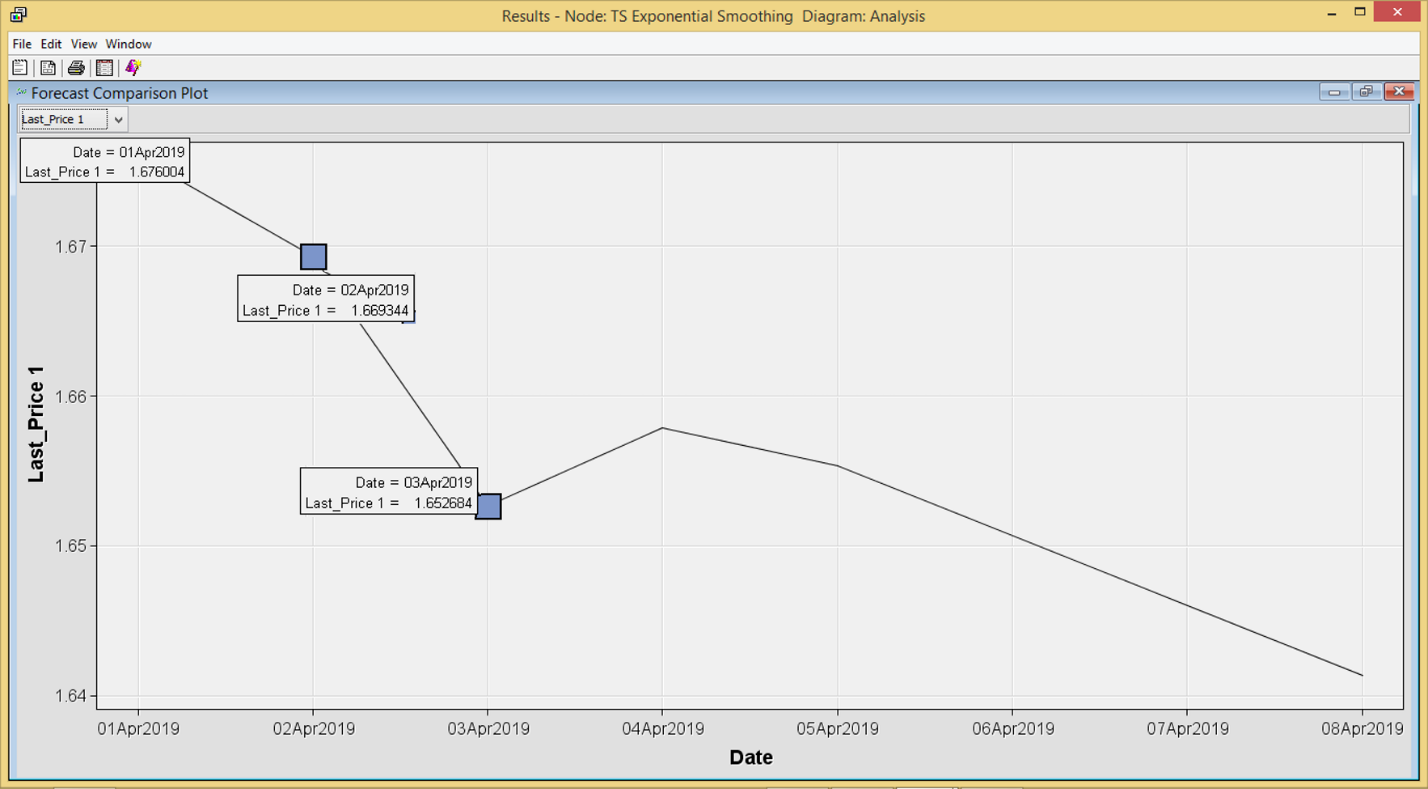


**Figure 12: Forecasting for TS\_002**

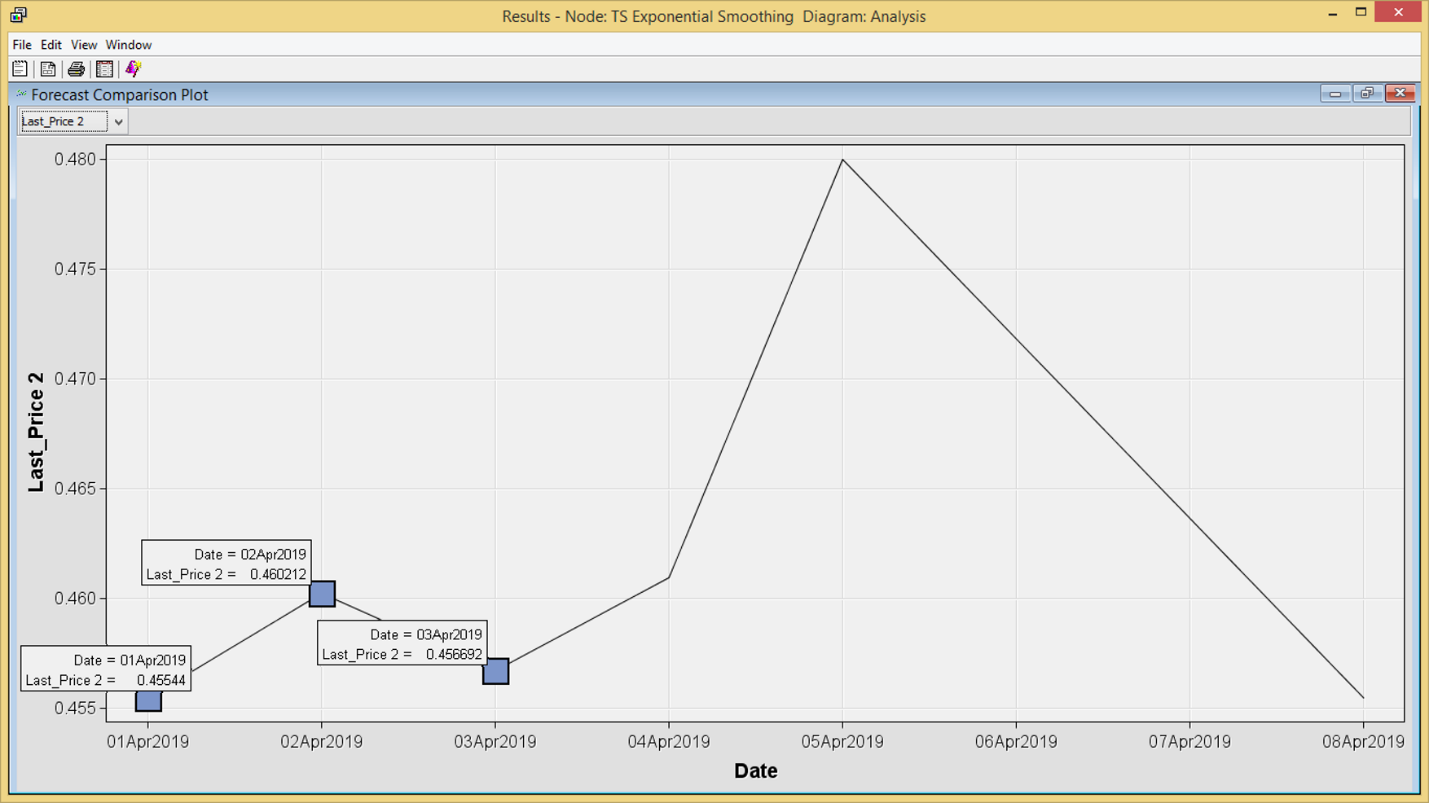


**Figure 13: Forecasting for TS\_003**

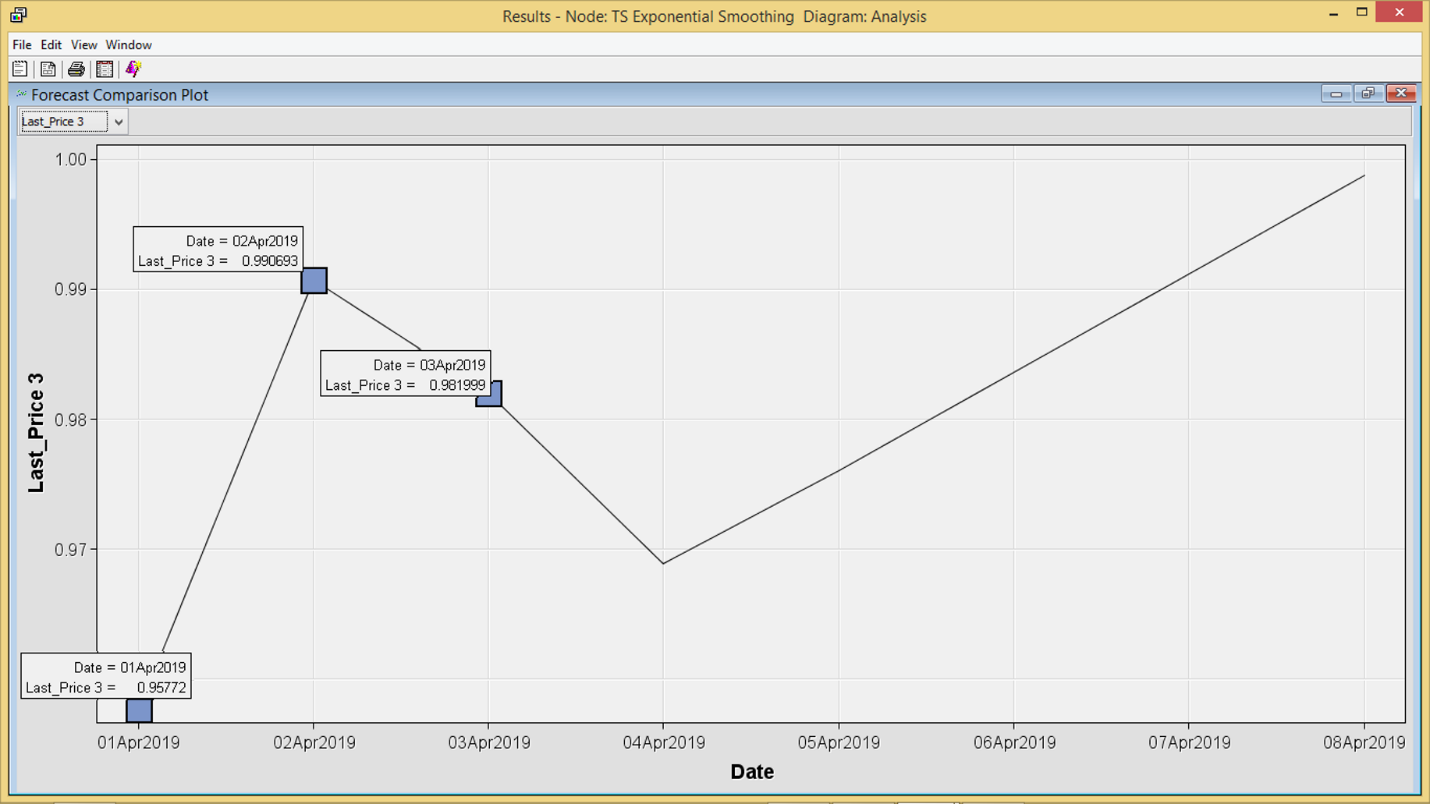
In Milestone 6, we further analysed the time series data by comparing the forecasting from TS in SAS with the real value of the Last Price for 1st April until 3rd April 2019. The Last price is the price at which the last transaction occurred. This information is valuable for investor to know how much are people willing to buy and willing to sell and also the most recent transaction occurred. The predicted value is shown in the **Figure 14, Figure 15** and **Figure 16**. The actual value is shown in the **Table 1** below with the comparison to the forecasting value.



**Figure 14: Predicted value for TS\_001**

****

**Figure 15: Predicted value for TS\_002**



**Figure 16: Predicted value for TS\_003**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Company ID** | **Company name** | **Actual Last Price** | **Predict Last Price** | **Difference between actual and predicted price** |
| 1/4/2019 | TS\_001 | KOTRA INDUSTRIES BHD | 1.69 | 1.6760 | 0.014 |
| TS\_002 | WILLOWGLEN MSC BHD | 0.455 | 0.4554 | 0.0004 |
| TS\_003 | THREE-A RESOURCES BHD | 0.93 | 0.9577 | -0.0277 |
| 2/4/2019 | TS\_001 | KOTRA INDUSTRIES BHD | 1.66 | 1.6693 | -0.0093 |
| TS\_002 | WILLOWGLEN MSC BHD | 0.445 | 0.4602 | -0.0152 |
| TS\_003 | THREE-A RESOURCES BHD | 0.93 | 0.9907 | -0.0607 |
| 3/4/2019 | TS\_001 | KOTRA INDUSTRIES BHD | 1.63 | 1.6527 | -0.0227 |
| TS\_002 | WILLOWGLEN MSC BHD | 0.475 | 0.4567 | 0.0183 |
| TS\_003 | THREE-A RESOURCES BHD | 0.94 | 0.9820 | -0.0420 |

**Table 1: Comparison between actual and predicted value of Last Price for three companies**

Based on the **Table 1** above, we could see that some prediction is nearly the same and some are totally different. Stock market prices are not something that we could spot easily as it does not move in a straight lines. There are common trends of stock market which are uptrend and downtrend. Uptrend is classified as a series of higher highs and higher lows as what we can see from **Figure 13**, while a downtrend consists of lower lows and lower highs as per **Figure 11**.

1. **Conclusion**

In conclusion, there is no evidence of perfectly correct in predicting market behavior. In this project, we are only did a small part of prediction but there are lots of theory that analyst are using in predicting the stock market. El-Wassal (2013) concluded that there are three principles concerning the stock market development. Firstly, the stock market development is difficult, complex, multi-faceted, and long-term process. Secondly, stock market development is part of a process of development of a financial system and thirdly, the development of a stock market is primarily a private sector activity but it actually need the supporting role of the government for a market to develop. However, the key of success in stock trading is to buy and sell stocks at the right time for the right price.

1. **References**

El-Wassal, K. A. (2013). *The Development of Stock Markets: In Search of a Theory*. International Journal of Economics and Financial Issues, 3 (3), 606-624.

SAS Enterprise Miner – Time Series Node. Retrieved from <http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_intro_sect050.htm>

Wah, T.Y. (2006). *Introduction to Data Mining*. Kuala Lumpur: University of Malaya Press.

1. **Presentation link**

<https://prezi.com/p/ysi6udeat_df/stock-market-prediction/>