WQD7005 Data Mining

Milestone 4

Student: Tan Kai Xuan

Student ID: WQD180036

In this milestone 4, I am going to do the interpretation data using the dataset that crawled during the group assignments. In this report, I will separate the milestone 4 into 3 parts: quantitative analysis, qualitative analysis and their comparisons.

Section 1: Quantitative analysis

For qualitative analysis, I'm going to use SAS Enterprise Miner to do some analysis on the stock dataset based on the telecommunication service providers. This sector has total 5 telco in the stock list provided by The Star, they are Axiata Group Berhad, Digi.com Berhad, Maxis Berhad, TIME DOTCOM Berhad as well as Telekom Malaysia Berhad. In the dataset, we have 21 variables and 306 observations which the data crawled from 1st March 2019 until 26th April 2019.

Table Properties

Property	Value
Table Name	STOCKDAT.TELCODATA_TRAIN
Description	
Member Type	DATA
Data Set Type	DATA
Engine	BASE
Number of Variables	21
Number of Observations	306
Created Date	April 27, 2019 12:12:50 PM SGT
Modified Date	April 27, 2019 12:12:50 PM SGT

Figure 1: Table Properties

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Chg	**Input**	**Interval**	**No**		**No**	.	.
ChgPercent	**Input**	**Interval**	**No**		**No**	.	.
Code	**Input**	**Interval**	**No**		**No**	.	.
Date	**Time ID**	**Interval**	**No**		**No**	.	.
Headline	**Text**	**Nominal**	**No**		**No**	.	.
High	**Input**	**Interval**	**No**		**No**	.	.
Last	**Input**	**Interval**	**No**		**No**	.	.
Low	**Input**	**Interval**	**No**		**No**	.	.
Name	**Input**	**Nominal**	**No**		**No**	.	.
Open	**Input**	**Interval**	**No**		**No**	.	.
ShortName	**Input**	**Nominal**	**No**		**No**	.	.
Source	**Input**	**Binary**	**No**		**No**	.	.
URL	**Text**	**Nominal**	**No**		**No**	.	.
Volume	**Input**	**Interval**	**No**		**No**	.	.
_52WeekHigh	**Input**	**Interval**	**No**		**No**	.	.
_52WeekLow	**Input**	**Interval**	**No**		**No**	.	.
compound	**Input**	**Interval**	**No**		**No**	.	.
label	**Label**	**Interval**	**No**		**No**	.	.
neg	**Input**	**Interval**	**No**		**No**	.	.
neu	**Input**	**Interval**	**No**		**No**	.	.
pos	**Input**	**Interval**	**No**		**No**	.	.

Figure 2: Metadata

Obs # ▲	Variable ...	Label	Type	Percent ...	Minimum	Maximum	Mean	Number o...	Mode Per...	Mode
1	Headline	Headline	CLASS	41.55405	.	.		.128+	1.351351	BUILDING ...
2	Name	Name	CLASS	0	.	.		.5	23.52941	AXIATA GR...
3	ShortName	ShortName	CLASS	0	.	.		.5	23.52941	AXIATA
4	Source	Source	CLASS	40.19608	.	.		.3	33.66013	THESTAR
5	URL	URL	CLASS	40.19608	.	.		.2	59.80392	HTTPS://W...
6	Chg	Chg	VAR	0	-0.18	0.45	0.00781	.	.	
7	ChgPercent	ChgPercent	VAR	0	-3.96	5.81	0.099346	.	.	
8	Code	Code	VAR	0	4863	6947	6098.35	.	.	
9	Date	Date	VAR	0	21609	21665	21639.11	.	.	
10	High	High	VAR	0	2.88	9.07	5.026209	.	.	
11	Last	Last	VAR	0	2.83	9.05	4.983954	.	.	
12	Low	Low	VAR	0	2.83	9.04	4.931699	.	.	
13	Open	Open	VAR	0	2.85	9.05	4.971209	.	.	
14	Volume	Volume	VAR	0	1439	93024	26839	.	.	
15	_52WeekHi...		VAR	0	4.88	9.07	5.94585	.	.	
16	_52WeekL...		VAR	0	2.11	7.06	4.143203	.	.	
17	compound	compound	VAR	40.19608	-0.6808	0.765	0.065121	.	.	
18	label	label	VAR	40.19608	-1	1	0.163934	.	.	
19	neg	neg	VAR	40.19608	0	0.45	0.078262	.	.	
20	neu	neu	VAR	40.19608	0.485	1	0.802126	.	.	
21	pos	pos	VAR	40.19608	0	0.515	0.119607	.	.	

Figure 3: Sample Statistics

Figure 3 shows the sample statistics of the dataset. This is the summary of the dataset which we can have a brief understanding on the dataset by knowing the minimum, maximum and mean of the variables.

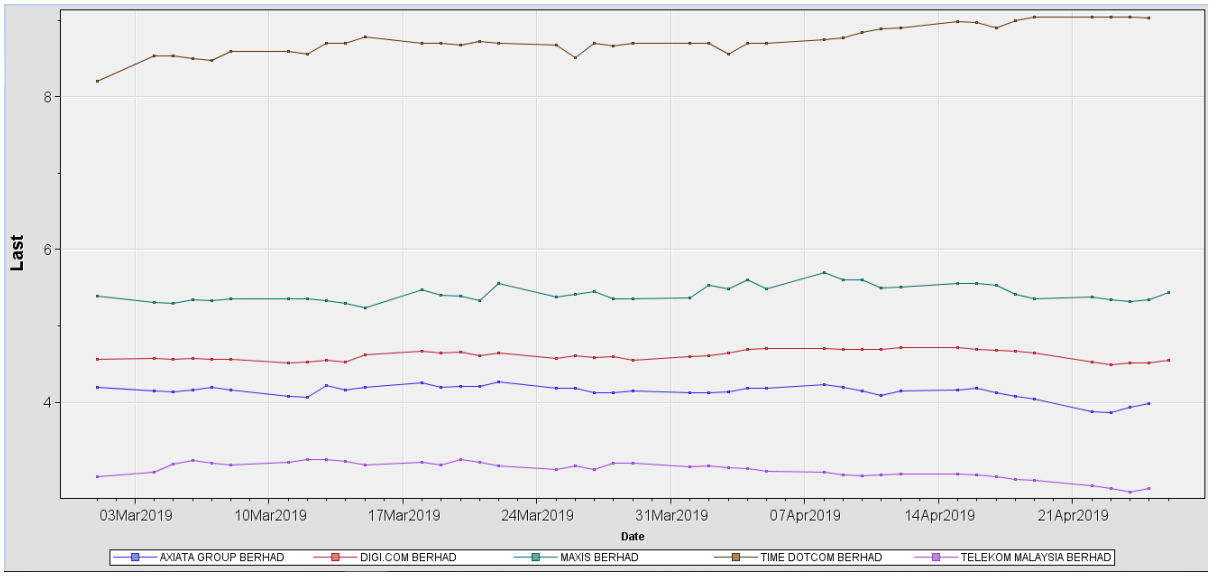

Figure 4: Last Price by Days

From Figure 4, we can know that the closing price for each stock for every single day. Each line represents to one telco in Malaysia market. From the graph above, we know that TIME DOTCOM Berhad has the highest stock price among all the telco.

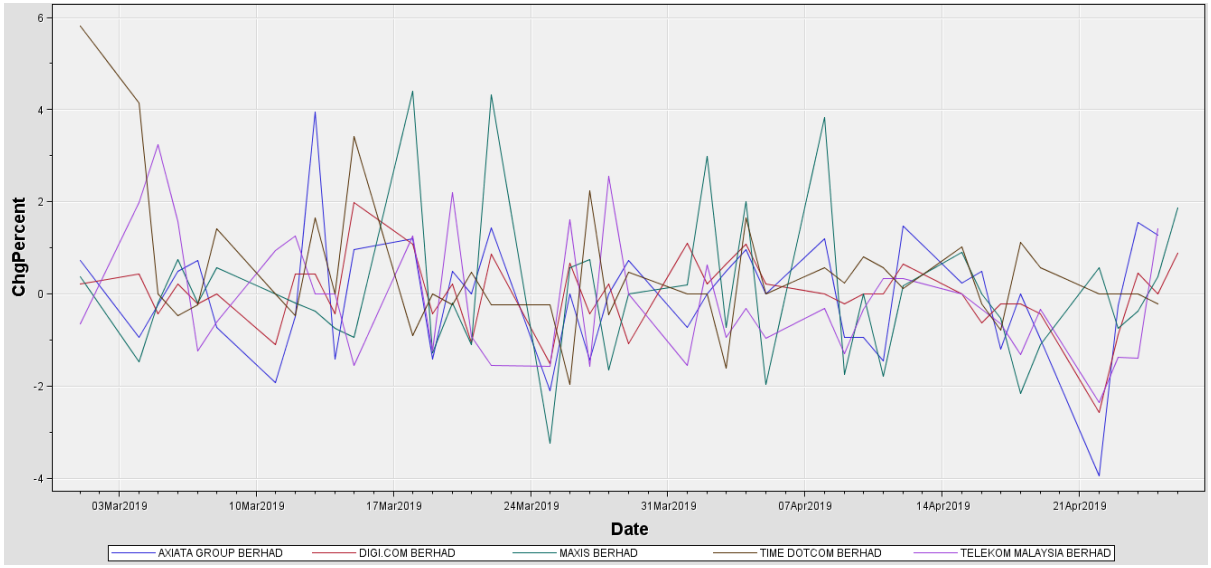

Figure 5: Change Percent per Day

Figure 5 shows the change percent of the stock price for each telco per day. If the change percent is positive, it means that the stock price has increased compared to previous day. If the

change percent is negative, the stock price has decreased as well as if the stock price is zero, it means there is no changes on the stock price.

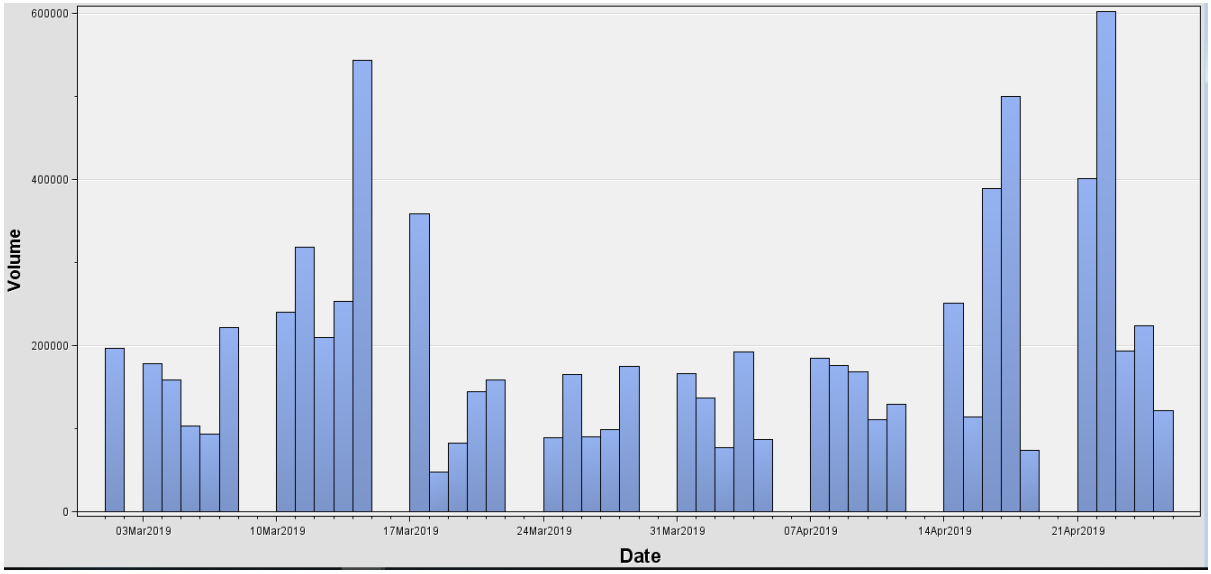

Figure 6: Volume by Days

From Figure 6, we know the total number of shares traded of telco sector from 1st March 2019 until 26th April 2019. 23rd April has the highest volume in telco market during this period.

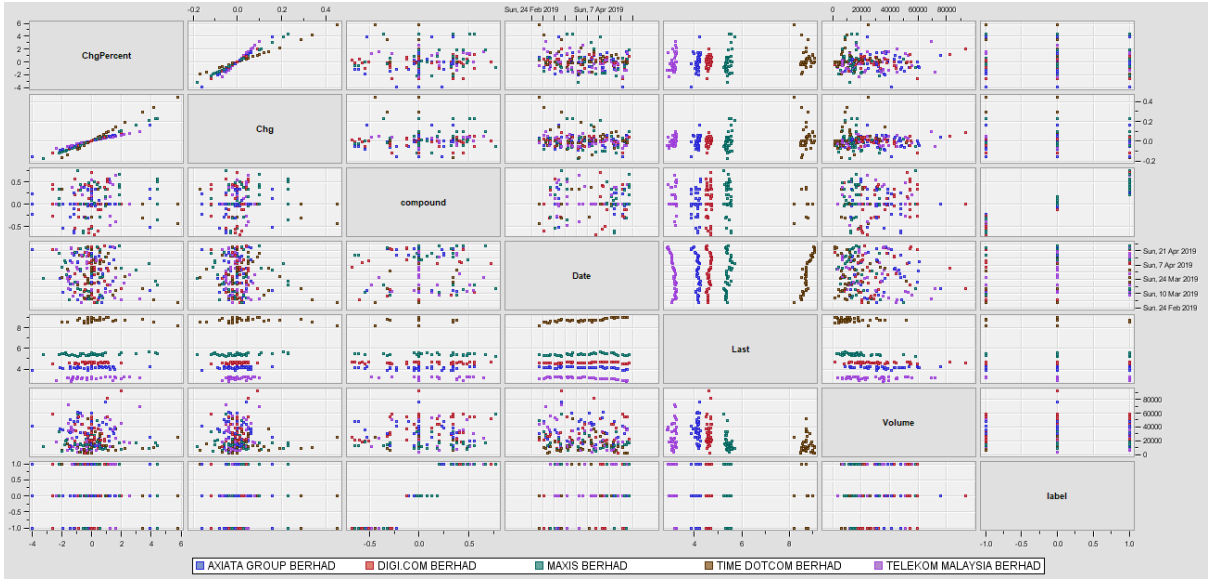

Figure 7: Correlation Matrix

Figure 7 shows the correlation matrix with selected variables: Date, Last, Change Percent, Change, Volume, Compound and Label. Compound and Label are based on the sentiment analysis of the news headline that we've crawled mainly from The Edge and The Star.

Section 2: Qualitative analysis

In qualitative analysis, I am going to apply sentiment analysis on the news headline on telecommunication service providers that we crawled during the group assignments. Sentiment analysis combines the power of natural language processing and text analysis to classify response as 'positive', 'negative' or 'neutral'. NLTK's built in Vader Sentiment Analyzer will rank the headline as positive, negative or neutral using a lexicon of positive or negative words. There are total 4 columns from the sentiment scoring: Positive, Negative, Neutral and compound. Compound is a single number that scores the sentiment which ranges from -1 to 1. I will consider the compound value greater than 0.2 as positive and less than -0.2 as negative.

```
Positive headlines:

['klci erases gains in line with cautious regional markets',
 'time dotcom fy2018 net profit soars to rm289 million',
 'putrajaya isnt supposed to buy loyalty with pay rise says dr m',
 'buying loyalty not the way says mahathir',
 'klci gains 0.26 as select blue chips lift']

Negative headlines:

['mixed results for telco sector amid moderating regulatory pressures',
 'politics sent malaysia stocks up now u-turns as doubts emerge',
 'telekom malaysia cut to neutral at public investment bank',
 'klci drifts lower in line with regional pause',
 'klci drifts lower in line with regional pause']
```

Figure 8: Outputs of positive and negative headlines

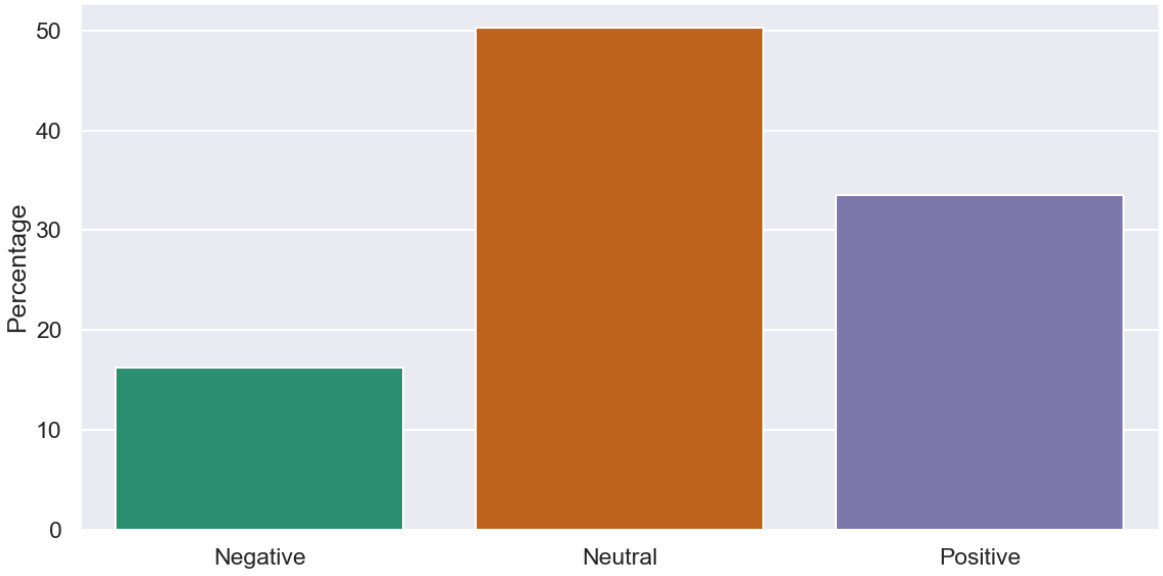

Figure 9: Summary of the headlines

Section 3: Comparison of qualitative and quantitative analysis

In this section, I am going to show the comparison of qualitative (sum of compound) and quantitative (last price of the day) using SAS Enterprise Miner.

Stock prices move up and down every minute due to fluctuations in supply and demand. If more people want to buy a particular stock, its market price will increase. Conversely, if more people want to sell a stock, its price will drop. This relationship between supply and demand is tied into the type of news reports that are issued at any particular moment.

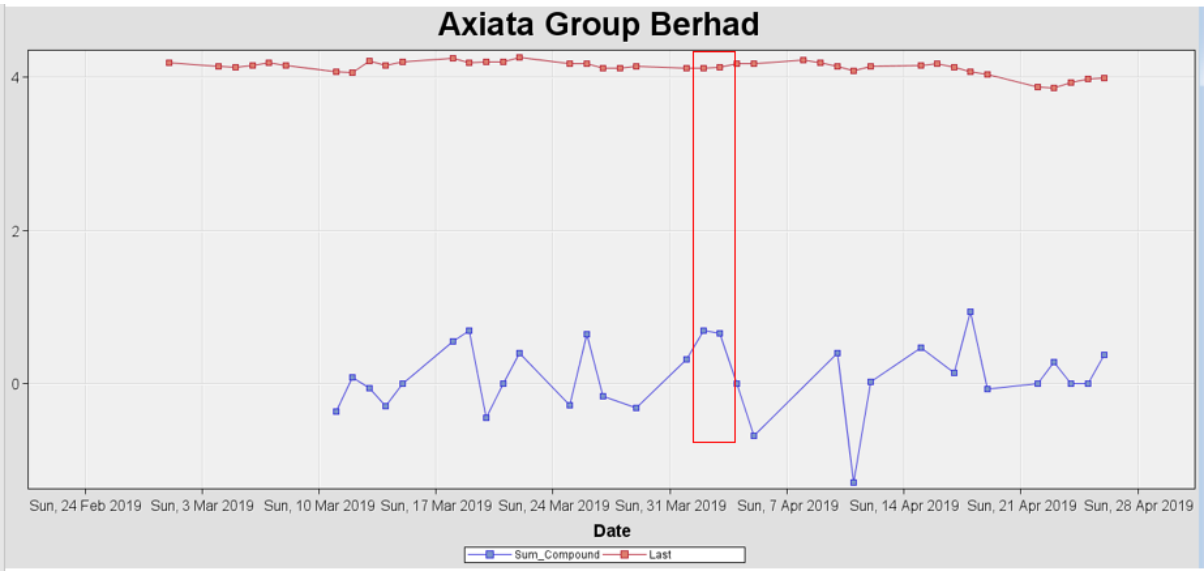

Figure 10: Axiata Group Berhad

Figure 10 shows different variables in one graph. X-axis represents Date, Y-axis represents two variable that we are going to interpret: blue colour line for Sum_Compound and red colour line shows the Last price of the day. From Figure 10, we can see that the sum of compound is closely related to the last price of the day. In the red box pointed in Figure 10, we know that when the sum of compound is higher compared to previous day, the last price will tend to increase. However, when the sum of compound dropped, the last price will be dropped eventually.

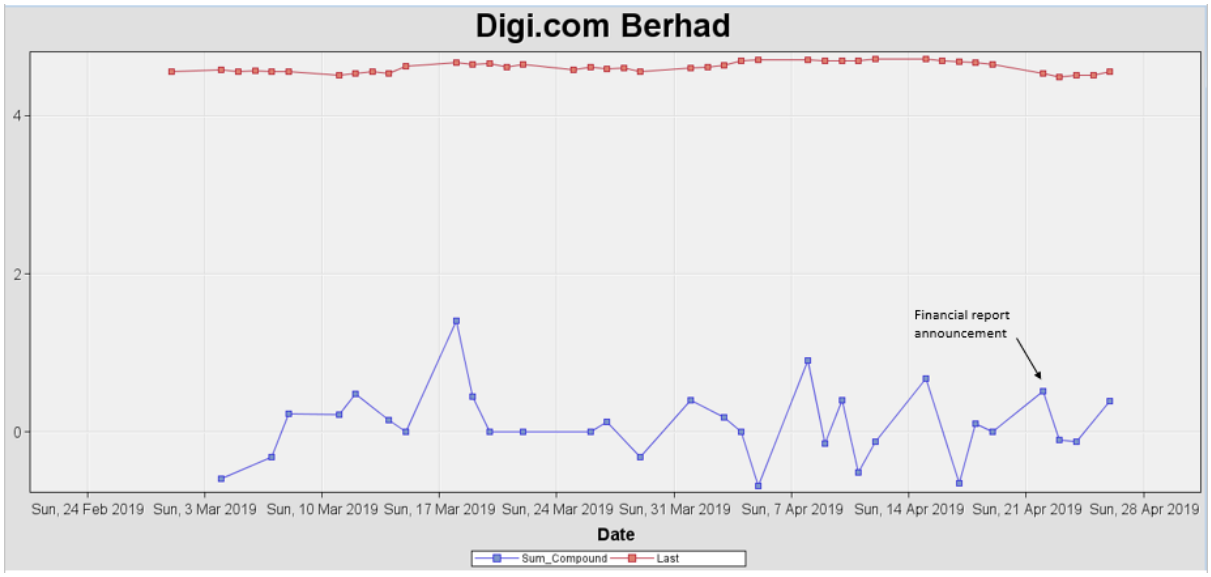

Figure 11: Digi.com Berhad

Figure 11 included the financial report announcement for Digi.com Berhad on 22nd April 2019. We can see that the sum of compound is increased compared to the previous day. Despite of the last price is the lowest, the last price is increasing for the next two days. Hence, the financial announcement is one of the factors that affect the last price of the stocks.

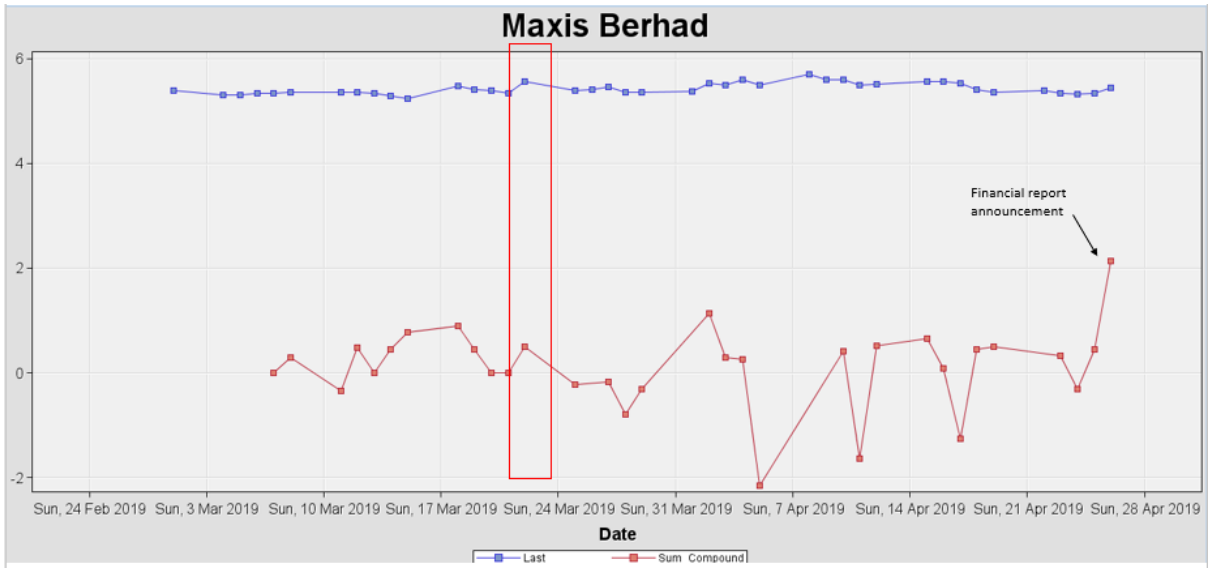

Figure 12: Maxis Berhad

In Figure 12, we can see a very obvious changes of the sum of compound which affect the last price of the day in Maxis Berhad. The financial report announcement that day, the sum of compound is the highest, and thus the price of the Maxis Berhad increasing.

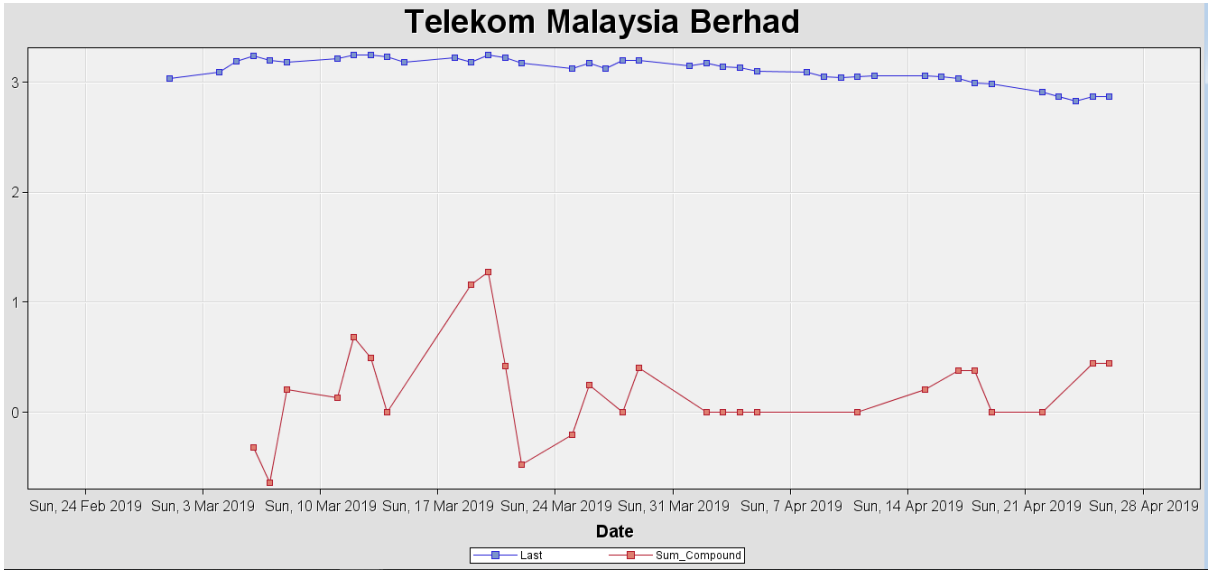

Figure 13: Telekom Malaysia Berhad

In Figure 13, we know that the sum of compound affects the stock price and cause to increase or fluctuate accordingly.

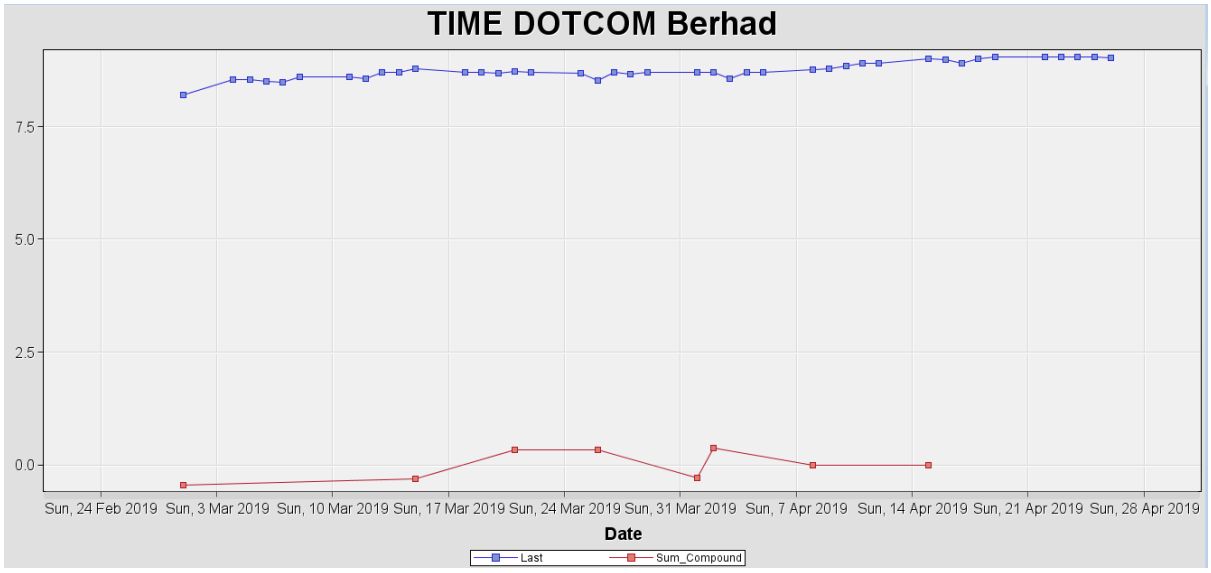

Figure 14: TIME DOTCOM Berhad

In Figure 14, there is no any news after 14th April, thus, the price of TIME DOTCOM Berhad is kept consistently for the next few days.

In conclusion, from Figure 10 to Figure 14 for different telco company, we know that the news is closely related to the price of the day. Negative news will cause the individual to sell the stocks, and the price will fall on that particular day or a few days. Positive news will normally cause individuals to buy the stocks and caused the increase in stock price.

Video links

Belows are two video links, one with the full presentation, another one with sentiment analysis python codes presentation.

Full presentation

https://drive.google.com/file/d/1GFZOTsvGVe2Sl98_GUq5iP2GzLpY1ta_/view?usp=sharing

Sentiment Analysis Python Codes presentation

https://drive.google.com/file/d/1SLeT-dPud8JalFdqARr2TlM5yL9G-HTd/view?usp=sharing