

Milestone 1

Acquisition of Data

WQD7005 – DATA MINING

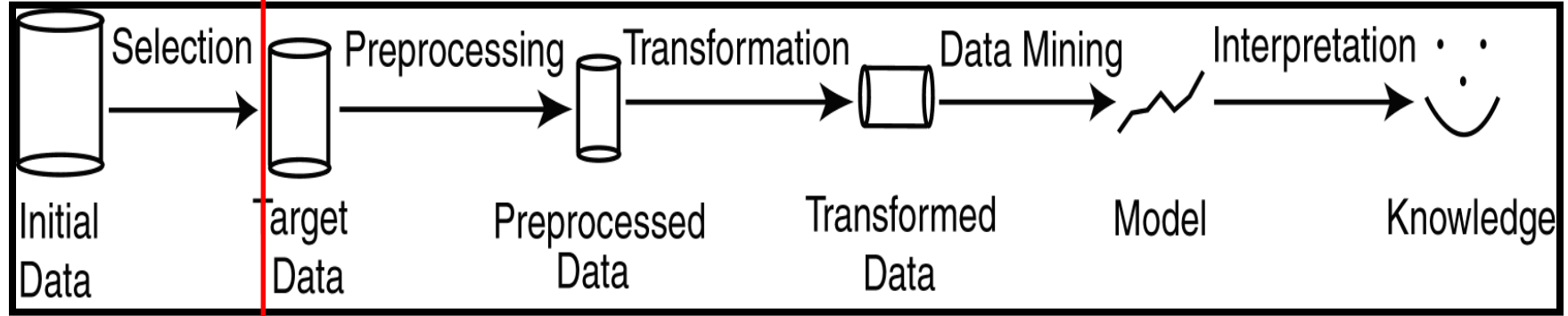
Zulkanain Bin Hasan
WQD180031

Group Member:
WQD170072 Nur Athirah Bt. Razak

Knowledge Discovery in Databases (KDD): process of finding useful information and patterns in data:

1. Selection (Pre-Mining 1): Obtain data from various sources

Milestone 1



Desktop/Link | X | assignment-1- X | pystock-crawl X | RHB

< > C | RHB Investment Bank Bhd [MY] www.rhbtradesn

Bookmarks Ba... BDM Google Sheets SAS® Logon...

RHB TRADESMART

Control . Convenience . Rewards

Focus data:

1. KLSE Main Market
2. Most active counter based on Top Volume as of 25/09/2019
3. Using RHB Trade Smart platform to identify 150 most active counters

CI: 1,592.33 (-0.6 / -0.038%) ▲ 349 ▼ 427 ▼ 425 T.Vol(Shares) 1,975,613,300 Turnover(RM) 1,572,286,938

MyPage Templates Portfolio Alert Orders Market Info News Subscriptions

Calculator Price Alert Matched Rejected Message

Trading 1 Scoreboard Indices Idx Components Stk Stock Monitor

Quotes

BME Normal Code/Name MAIN Top Volume 1 2 3 4 5

Exch	M	C	N	Sts	Name	Curr	Ref	Open	High	Low	B/Qty	Buy	Sell	S/Qty	Last	L.Vol	Chg	%Chg	Volume	R/D	TOCP
BME	N	01		A	GPACKET-WB	MYR	0.260	0.260	0.270	0.260	131,339,400	0.265	0.270	18,484,400	0.265	10,000	0.005	1.92	88,792,500	R	0.265
BME	N	5:	*	A	ARMADA	MYR	0.340	0.340	0.340	0.325	18,002,800	0.330	0.335	5,280,300	0.330	25,000	-0.010	-2.94	71,194,000	R	0.330
BME	N	5:		A	SAPNRG °	MYR	0.285	0.290	0.295	0.285	17,287,000	0.290	0.295	9,126,200	0.290	81,300	0.005	1.75	61,757,400	R	0.290
BME	N	5:	*	A	VELESTO °	MYR	0.310	0.315	0.325	0.310	2,737,500	0.320	0.325	767,900	0.325	5,000	0.015	4.84	32,103,900	R	0.325
BME	N	7:	*	A	KNM °	MYR	0.400	0.405	0.410	0.395	4,270,300	0.395	0.400	1,698,500	0.395	2,500	-0.005	-1.25	30,899,600	R	0.395
BME	N	01	*	A	MTRONIC °	MYR	0.050	0.050	0.055	0.045	9,696,600	0.050	0.055	11,458,100	0.050	44,900	0.000	0.00	29,622,800	R	0.050

Market Queue - [BME] GPACKET-WB (0082WB)

BQty	Buy	Sell	SQty
131,339,400	0.265	0.270	18,484,400
58,322,600	0.260	0.275	22,671,300
400,000	0.255	0.280	5,353,000

Order Pad - [BME] GPACKET-WB (0082WB) Lot Size 100

BME GPACKET-WB Select Account No. BUY SELL Enter Trading PIN Keep PIN

Order Type Validity Qty (x1) Price (MYR) 0082WB - GPACKET-WB (Active) Lot Size 100

Limit Day

Stop Price This order will be placed upon meeting this criteria. BME Code/Name If Price Is Greater Than

SUBMIT

Identify Stock Name and Stock Code for all 150 counters

RHB Investment Bank Bhd [MY] www.rhbtradesmart.com/web/index.html#/trading1

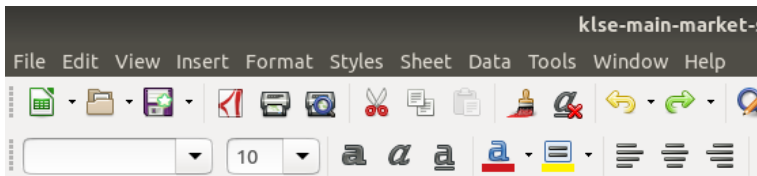
Bookmarks Ba... BDM Google Sheets SAS® Logon... Login | Tablea...

Quotes

BME Normal Code/Name MAIN Top Volume 1 2 3 4 5

Exch	M	C	I	Sts	Name	Curr	Ref	Open	High	Low	B/Qty	Buy	Sell	S/Qty	Last	L.Vol	Chg	%Chg	Volume	R/D	TOCP
BME	N	01		A	GPACKET-WB	MYR	0.260	0.260	0.270	0.260	131,339,400	0.265	0.270	18,484,400	0.265	10,000	0.005	1.92	88,792,500	R	0.265
BME	N	51		A	ARMADA	MYR	0.340	0.340	0.340	0.325	18,002,800	0.330	0.335	5,280,300	0.330	25,000	-0.010	-2.94	71,194,000	R	0.330
BME	N	51		A	SAPNRG °	MYR	0.285	0.290	0.295	0.285	17,287,000	0.290	0.295	9,126,200	0.290	81,300	0.005	1.75	61,757,400	R	0.290
BME	N	51		A	VELESTO °	MYR	0.310	0.315	0.325	0.310	2,737,500	0.320	0.325	767,900	0.325	5,000	0.015	4.84	32,103,900	R	0.325
BME	N	71		A	KNM °	MYR	0.400	0.405	0.410	0.395	4,270,300	0.395	0.400	1,698,500	0.395	2,500	-0.005	-1.25	30,899,600	R	0.395
BME	N	01		A	MTRONIC °	MYR	0.050	0.050	0.055	0.045	9,696,600	0.050	0.055	11,458,100	0.050	44,900	0.000	0.00	29,622,800	R	0.050
BME	N	71		A	KOMARK °	MYR	0.340	0.340	0.345	0.330	2,160,000	0.325	0.330	21,900	0.330	3,100	-0.010	-2.94	27,664,100	R	0.330
BME	N	41		A	JAKS °	MYR	0.740	0.745	0.785	0.745	140,900	0.775	0.780	554,200	0.775	19,100	0.035	4.73	23,706,500	R	0.775
BME	N	51		A	SAPNRG-WA °	MYR	0.125	0.120	0.130	0.120	8,208,300	0.125	0.130	12,557,700	0.125	5,000	0.000	0.00	18,808,200	R	0.125
BME	N	71		A	YONGTAI °	MYR	0.200	0.205	0.210	0.200	60,800	0.205	0.210	3,454,500	0.210	12,000	0.010	5.00	15,745,700	R	0.210
BME	N	51		A	AAX °	MYR	0.165	0.170	0.170	0.160	15,288,200	0.160	0.165	7,781,400	0.165	10,000	0.000	0.00	15,585,700	R	0.165
BME	N	71		A	SCOMIES °	MYR	0.120	0.120	0.125	0.105	1,311,900	0.110	0.115	2,237,600	0.115	2,000	-0.005	-4.17	14,862,000	R	0.115
BME	N	31		A	KBUNAI-WC	MYR	0.015	0.025	0.030	0.025	202,321,900	0.025	0.030	98,631,700	0.030	10,000	0.015	100.00	14,622,500	R	0.030
BME	N	51		A	VELESTO-WA °	MYR	0.115	0.120	0.125	0.120	2,600,000	0.125	0.130	3,426,300	0.125	5,000	0.010	8.70	13,992,700	R	0.125
BME	N	11		AC	CIMB	MYR	5.040	5.050	5.060	5.020	190,300	5.030	5.040	84,600	5.040	2,000	0.000	0.00	13,806,700	R	5.040
BME	N	51		A	IKHMAS °	MYR	0.125	0.125	0.130	0.115	266,000	0.120	0.125	1,724,100	0.125	20,000	0.000	0.00	13,154,500	R	0.125
BME	N	71		A	KNM-WB °	MYR	0.175	0.175	0.185	0.170	960,500	0.175	0.180	2,243,300	0.180	73,200	0.005	2.86	12,509,000	R	0.180
BME	N	51		A	AYREIT °	MYR	1.820	1.820	1.840	1.820	1,601,100	1.820	1.840	33,400	1.840	10,000	0.020	1.10	11,850,000	R	1.840

Stock Code Stock Name



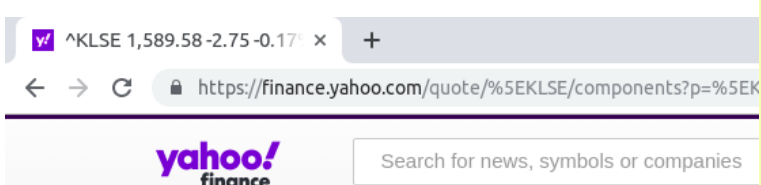
1. Copy all data from RHB Trade Smart platform and transfer to Libre Office
2. Delete others info in order to keep only Stock Code

	A	B	C	D	E	F	G	H	I	J	K
1	0082WB		A	GPACKET-WB	MYR	0.26	0.26	0.27	0.26	131339400	
2	5210*		A	ARMADA	MYR	0.34	0.34	0.34	0.325	18002800	
3	5218		A	SAPNRG °	MYR	0.285	0.29	0.295	0.285	17287000	
4	5243*		A	VELESTO °	MYR	0.31	0.315	0.325	0.31	2737500	
5	7164*		A	KNM °	MYR	0.4	0.405	0.41	0.395	4270300	
6	43*		A	MTRONIC °	MYR	0.05	0.05	0.055	0.045	9696600	
7	7017*		A	KOMARK °	MYR	0.34	0.34	0.345	0.33	2160000	
8	4723*		A	JAKS °	MYR	0.74	0.745	0.785	0.745	140900	
9	5218WA		A	SAPNRG-WA °	MYR	0.125	0.12	0.13	0.12	8208300	
10	7066*		A	YONGTAI °	MYR	0.2	0.205	0.21	0.2	60800	
11	5238*		A	AAX °	MYR	0.165	0.17	0.17	0.16	15288200	
12	7045*		A	SCOMIES °	MYR	0.12	0.12	0.125	0.105	1311900	
13	3115WC		A	KBUNAI-WC	MYR	0.015	0.025	0.03	0.025	202321900	
14	5243WA		A	VELESTO-WA °	MYR	0.115	0.12	0.125	0.12	2600000	
15	1023*		AC	CIMB	MYR	5.04	5.05	5.06	5.02	190300	
16	5268*		A	IKHMAS °	MYR	0.125	0.125	0.13	0.115	266000	
17	7164WB		A	KNM-WB °	MYR	0.175	0.175	0.185	0.17	960500	
18	5106*		A	AXREIT °	MYR	1.82	1.82	1.84	1.8	1681100	

Sheet1

Find All ☐ Formatted Display ☐ Match Case

Sheet 1 of 1 | 1048576 rows, 28 columns selected | Default | English (USA) | Average: 1108904.91117605; Sum: 2324264693.825 | 140%

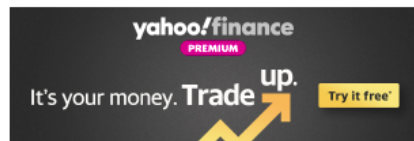
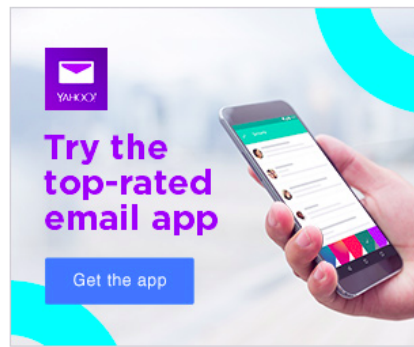


1. Identify Stock Code for KLSE Components at Yahoo Finance website
2. Found that all KLSE Components Code save as 'xxxx.KL' in Yahoo Finance website

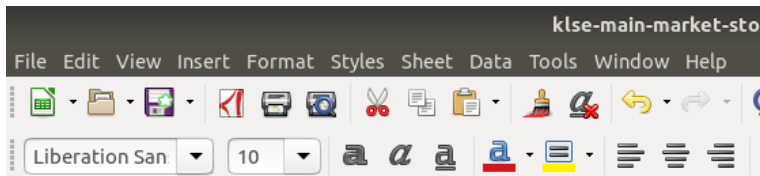
Currency in MYR

Top 30 Components

Symbol	Company Name	Last Price	Change	% Change	Volume
6947.KL	DiGi.Com Berhad	4.6900	0.0000	0.00%	2,428,000
5225.KL	IHH Healthcare Berhad	5.70	0.00	0.00%	1,429,100
3034.KL	Hap Seng Consolidated Berhad	9.85	0.00	0.00%	169,300
7113.KL	Top Glove Corporation Berhad	4.7300	0.0000	0.00%	4,735,200
4065.KL	PPB Group Berhad	18.22	0.00	0.00%	1,457,600
1082.KL	Hong Leong Financial Group Berhad	16.12	0.02	+0.12%	662,500
4707.KL	Nestlé (Malaysia) Berhad	145.50	0.20	+0.14%	211,300
8869.KL	Press Metal Bhd	4.8900	-0.0100	-0.20%	766,100
5681.KL	Petronas Dagangan Bhd	23.48	0.06	+0.26%	502,600
3816.KL	MISC	7.72	-0.02	-0.26%	4,903,800



Waiting for apx.moatads.com...



1. Rename the original Stock Code as per Stock Code in Yahoo Finance website
2. Save this file as 'klse-main-market-stocks-rhbtradesmart-topvolume-code.csv'

	A	B	C	D	E	F	G	H	I	J	K
1	0037.KL										
2	0043.KL										
3	0056.KL										
4	0065.KL										
5	0078.KL										
6	0078WB.KL										
7	0082.KL										
8	0082WB.KL										
9	0090.KL										
10	0099.KL										
11	0128.KL										
12	0138.KL										
13	0166.KL										
14	0172.KL										
15	0172WA.KL										
16	0183.KL										
17	0185.KL										
18	0192.KL										

Sheet1

4715

Find All

Formatted Display

Match Case

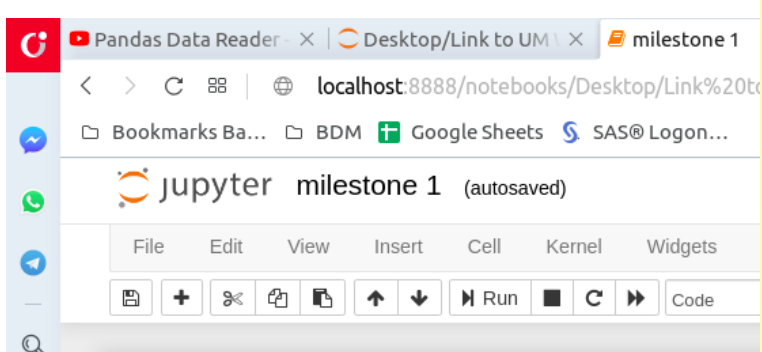
Sheet 1 of 1

Default

English (USA)

Average ; Sum: 0

140%



Python Script:

1. Import pandas and pandas_datareader packages in order to crawl data from Yahoo Finance website
2. Set date to crawl data (3 months) e.g. start = '2019-06-01', end = '2019-09-25'
3. Called previous 'klse-main-market-stocks-rhbtradesmart-topvolume-code.csv' in order to crawl data for all 150 counters

```
In [1]: import os  
os.getcwd()
```

```
Out[1]: '/mnt/70887E57887E1BB2/Users/zulkanh/Downloads/UM WQD - SARJANA SAINS DATA/SEMESTER I, 2019 2020/WQD7005 Perlombongan Data (Data Mining)/assignment/milestone 1'
```

```
In [2]: import numpy as np  
import pandas as pd  
import pandas_datareader.data as web
```

```
In [3]: start = '2019-06-01'  
end = '2019-09-25'
```

```
In [4]: stocks_query = pd.read_csv('klse-main-market-stocks-rhbtradesmart-topvolume-code.csv', index_col=[0])
```

```
In [5]: stocks = web.DataReader(stocks_query, data_source='yahoo')
```

```
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s  
ymbol: '0078WB.KL', replacing with NaN.  
  warnings.warn(msg.format(sym), SymbolWarning)  
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s  
ymbol: '0002WB.KL', replacing with NaN.
```


Google Chrome browser tabs: Pandas Data Reader, Desktop/Link to UM, milestone 1

Address bar: localhost:8888/notebooks/Desktop/Link%20to

Bookmarks: Bookmarks Ba..., BDM, Google Sheets, SAS@ Logon...

Jupyter interface: milestone 1 (autosaved)

Menu bar: File, Edit, View, Insert, Cell, Kernel, Widgets, Help

Toolbar: Save, New, Copy, Paste, Undo, Redo, Run, Stop, Refresh, Code

Code editor content:

```
In [4]: stocks_query = pd.read_csv('KSE-Mid-Market-Stocks-Financials-mkt-cp-volume-code.csv', index_col=0)

In [5]: stocks = web.DataReader(stocks_query, data_source='yahoo')

/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '0078WB.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '0082WB.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '0172WA.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '3115WC.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '4456WD.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '5156WD.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
ymbol: '5199WC.KL', replacing with NaN.
  warnings.warn(msg.format(sym), SymbolWarning)
/home/zulkanh/anaconda3/lib/python3.7/site-packages/pandas_datareader/base.py:230: SymbolWarning: Failed to read s
```

1. Found there is some error during crawl data for Warrant counters e.g. 0078WB.KL, 0082WB.KL and 0172WA.KL
2. Recheck at Yahoo Finance website and identify that all Warrant counters are not available in Yahoo Finance Website

Pandas Data Reader - X Desktop/Link to UM X milestone 1

localhost:8888/notebooks/Desktop/Link%20to

Bookmarks Ba... BDM Google Sheets SAS® Logon...

jupyter milestone 1 (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

In [6]: `stocks.head()`

Out[6]:

Attributes	High	...	Adj Close													
Symbols	0043.KL	0056.KL	0065.KL	0078.KL	0078WB.KL	0082.KL	0082WB.KL	0090.KL	0099.KL	0128.KL	...	7277.KL	8206.KL	8419.KL	8524.KL	86...
Date																
2010-01-04	NaN	0.11932	0.160	0.03125	NaN	1.28	NaN	0.42333	0.33333	0.20422	...	0.60871	0.08454	NaN	NaN	1.2
2010-01-05	NaN	0.11932	0.170	0.03250	NaN	1.35	NaN	0.30667	0.32576	0.21273	...	0.60871	0.08660	NaN	NaN	1.3
2010-01-06	NaN	0.11932	0.185	0.03531	NaN	1.35	NaN	0.30333	0.32197	0.20848	...	0.61325	0.08454	NaN	NaN	1.2
2010-01-07	NaN	0.12102	0.185	0.03438	NaN	1.45	NaN	0.42333	0.31818	0.20848	...	0.60871	0.08248	NaN	NaN	1.3
2010-01-08	NaN	0.12273	0.200	0.03750	NaN	1.44	NaN	0.42333	0.31818	0.20422	...	0.60871	0.08248	NaN	NaN	1.3

5 rows x 894 columns

In [7]: `stocks.tail()`

Out[7]:

1. Check `stock.head()` and found that historical data crawl from Yahoo Finance give results from January 2010 not as per start date setting
2. This activity managed to crawl 894 columns of data for all 150 counters

<

>

◀

Link to UM WQD...NA SAINS DATA

SEMESTER I, 2019 2020

Recent

Home

Desktop

Documents

Downloads

Music

Pictures

Videos

Trash

Other Locations

Name

milestone 1.ipynb

23.9 KB

Document

10:00 PM

milestone 1.pdf

16.9 kB

Document

10:01 PM

pandas-datareader.pdf

288.5 kB

Document

8:42 PM

ts-web-guide.pdf

17.1 MB

Document

12:30 AM

klse-main-market-stock-rhbtradesmart-topvolume-data.csv

23.3 MB

Text

9:59 PM

klse-main-market-stocks-code-rhbtradesmart-topvolume.csv

16.5 kB

Text

12:44 AM

klse-main-market-stocks-rhbtradesmart-topvolume.csv

16.5 kB

Text

12:45 AM

klse-main-market-stocks-rhbtradesmart-topvolume-code.csv

1.2 kB

Text

1:01 AM

milestone 1.py

596 bytes

Text

10:00 PM

1. Save crawl data to 'klse-main-market-stocks-rhbtradesmart-topvolume-data.csv'
2. Prepare data for next milestone

"klse-main-market-stock-rhbtradesmart-topvolume-data.csv" selected (23.3 MB)