Assignment 1

1. Acquisition of data (Group)

* Limited to the following area:

KLSE stock data (Web crawler)  
https://www.thestar.com.my/business/marketwatch/stocks/?qcounter=SAPNRG

Target:

To create a program to crawl the KLSE data for each stock code, such as SAPNRG

Tasks have been done:

1. Get the List of the stock code

Challenging:  
Due to the stock code listed by alphabet screen has been designed by requiring the browser to load the HTML source code and enable its javascript feature to render the screen, so that the missing content can be shown. If there is no javascript rendering feature, some content of the page will looks like missing in the page. Python coding unable to render the missing content.

Solution:

A program for accessing stock list page by alphabet is written in java and applying ui4j library. Then, all the stock code is gathered and saved into a file, compCodeList.txt.

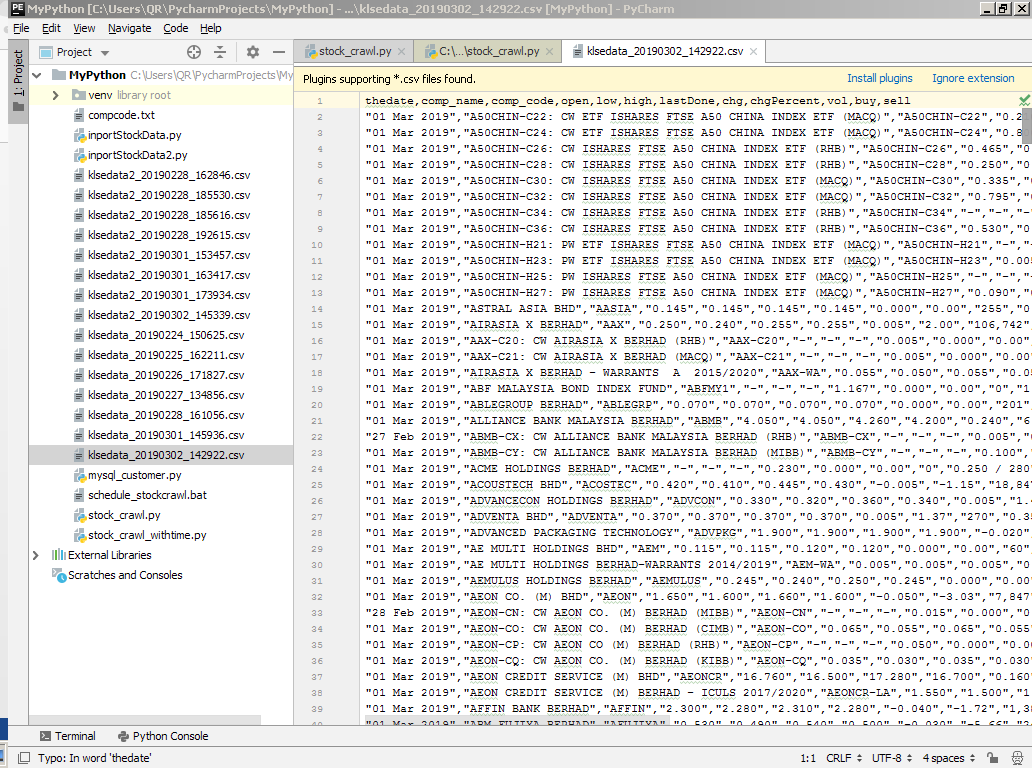
Result:

Total of the stock code is 1864.

1. By the list of stock code, get the KLSE data (information of the stock) after stock market close.

Solution:

Python program, stock\_crawl.py, is created to download by accessing the URL as below by looping the every stock code in the file, compCodeList.txt. [https://www.thestar.com.my/business/marketwatch/stocks/?qcounter=<stockcode](https://www.thestar.com.my/business/marketwatch/stocks/?qcounter=%3cstockcode)>

CSV files to be generated for keeping all the stock info with the column name:  
thedate,comp\_name,comp\_code,open,high,low,lastDone,chg,chgPercent,vol,buy,sell   
  


1. Keep the stock data every day in mysql database

Solution:  
Create the database and table by the scripts:

-- create a database  
create database MDSKLSE;

-- access MDSKLSE database  
use MDSKLSE;

-- Create a table KLSE2  
create table KLSE2 (  
thedate **date**,

comp\_name **text**,

comp\_code char(**50**),

open char(**10**),

low char(**10**),

high char(**10**),

lastDone char(**10**),

chg char(**10**),

chgPercent char(**10**),

vol char(**10**),

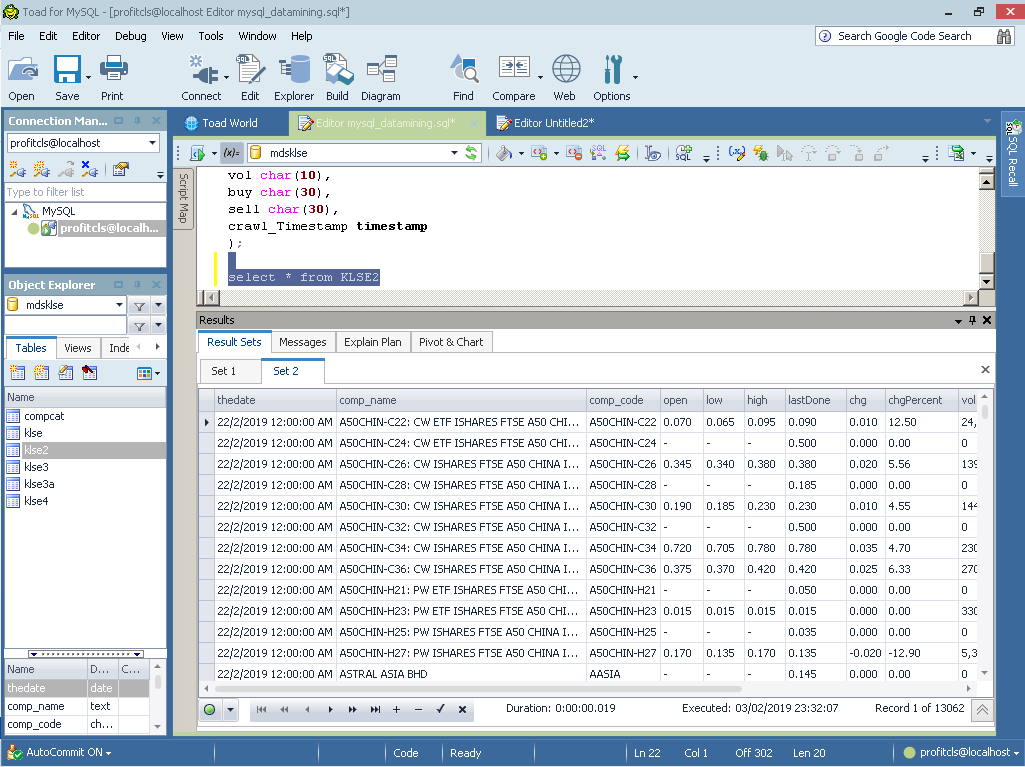
buy char(**30**),

sell char(**30**),

crawl\_Timestamp **timestamp**

);

When the stock data for all the stock code have been downloaded, the stock will be inserted into the table KLSE2 by the python script, stock\_crawl.py.



1. Repeat the process 2 every day to query the data

Solution:  
Run the download stock info every day after the stock market has closed. And, the stock data is inserted into mysql and generate the csv file of the stock market closing data.