Milestone 3

Processing of data (Group)

Co-Variance of the stocks.

To implement this milestone, the covariance of the stocks is calculated to find the stocks which are moving up and down together. To do so, the steps below have been done:

1. Calculate the percentage of stock changes

Changes % = (closing – open) /open X 100

Add the virtual column into KLSE table, upd\_klse

Alter table upd\_klse

Add `v\_updownpercent` **double**(**18**,**8**) GENERATED ALWAYS AS ((((`lastDone` - `open`) / `open`) \* **100**)) VIRTUAL

Create script of upd\_klse table:  
  
CREATE TABLE `upd\_klse` (

`thedate` **date** DEFAULT NULL,

`thetime` char(**20**) DEFAULT NULL,

`comp\_name` **text**,

`comp\_code` char(**50**) DEFAULT NULL,

`stock\_code` char(**50**) DEFAULT NULL,

`open` **double**(**18**,**8**) DEFAULT NULL,

`low` **double**(**18**,**8**) DEFAULT NULL,

`high` **double**(**18**,**8**) DEFAULT NULL,

`lastDone` **double**(**18**,**8**) DEFAULT NULL,

`chg` **double**(**18**,**8**) DEFAULT NULL,

`chgPercent` **double**(**18**,**8**) DEFAULT NULL,

`vol` **double**(**18**,**8**) DEFAULT NULL,

`buy` **double**(**18**,**8**) DEFAULT NULL,

`buyval` **double**(**18**,**8**) DEFAULT NULL,

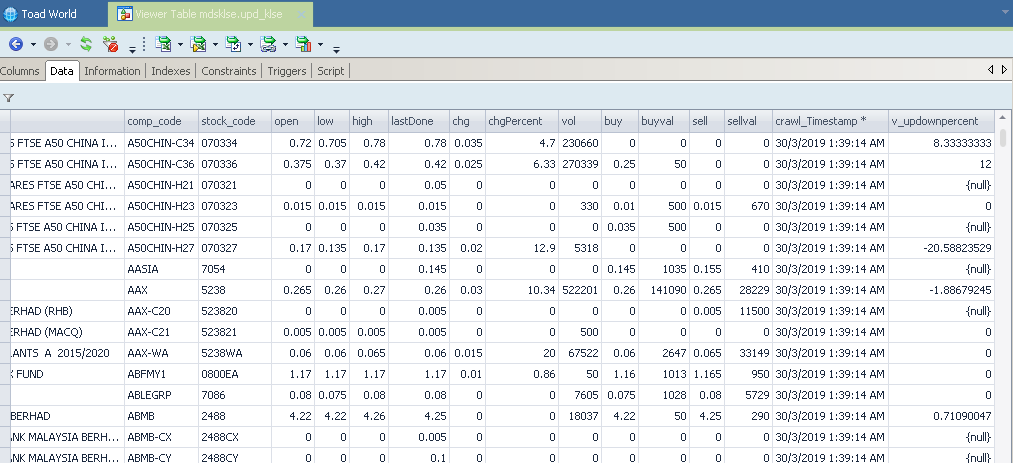
`sell` **double**(**18**,**8**) DEFAULT NULL,

`sellval` **double**(**18**,**8**) DEFAULT NULL,

`crawl\_Timestamp` **timestamp** NOT NULL DEFAULT CURRENT\_TIMESTAMP ON UPDATE CURRENT\_TIMESTAMP,

`v\_updownpercent` **double**(**18**,**8**) GENERATED ALWAYS AS ((((`lastDone` - `open`) / `open`) \* **100**)) VIRTUAL

) ENGINE=InnoDB DEFAULT CHARSET=utf8;

For the v\_updownpercent values will be null if ‘open’ is missing value.  
  
The data in mysql:  


1. Only the columns below are selected for finding covariant stocks:

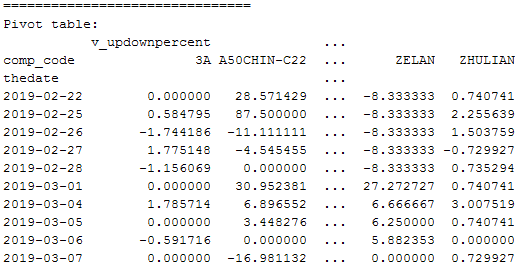
* thedate
* comp\_code
* v\_updownpercent

The query criterial is only v\_updownpercent is not null value is selected because the null of v\_updownpercent is caused by missing values of ‘open’ column.

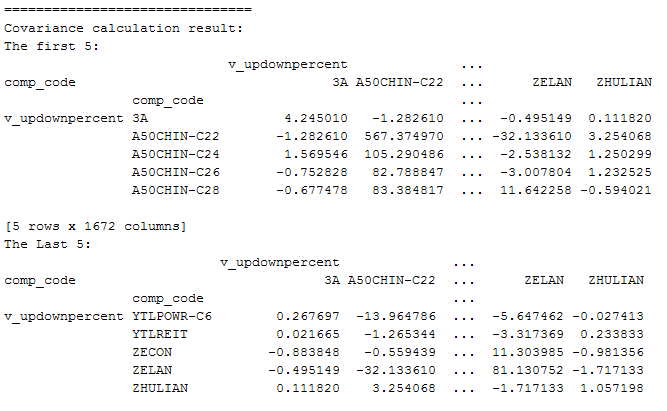
The query:

**" select thedate, comp\_code, v\_updownpercent "** \  
 **" from upd\_klse where v\_updownpercent is not null "** \  
 **" group by thedate, comp\_code, v\_updownpercent "** \  
 **" order by comp\_code, thedate"**

1. Load the data into pandas library to manage the dataset by dataframe.

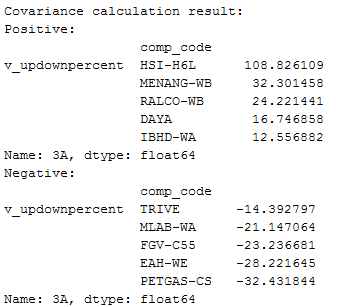


1. Calculate the Covariant

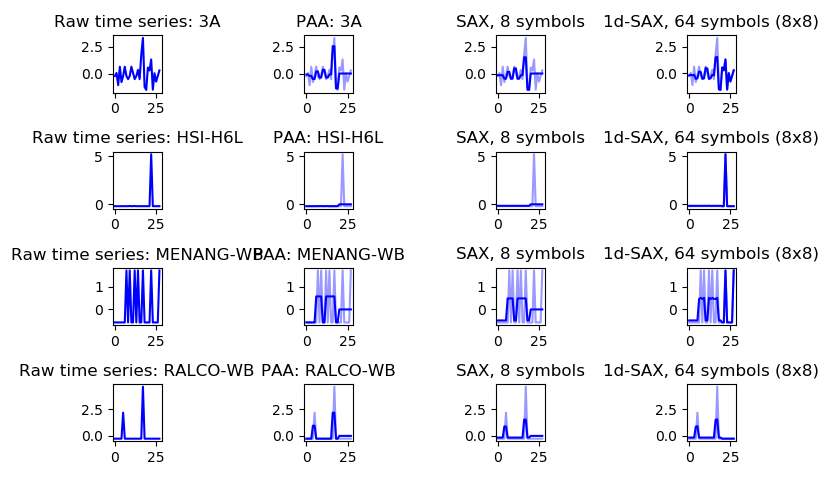


1. To find the covariant stocks with a stock, 3A

With sorted the covariant value:



1. With the higher positive covariant stock with ‘3A’, the stocks, ‘HSL-K6L’, ‘MENAG-WB’ and ‘RALCO-WB’ are selected to view their stock movement with PAA, SAX, 1s-SAX transform:



Output for python program:  
  
===============================

Pivot table:

v\_updownpercent ...

comp\_code 3A A50CHIN-C22 ... ZELAN ZHULIAN

thedate ...

2019-02-22 0.000000 28.571429 ... -8.333333 0.740741

2019-02-25 0.584795 87.500000 ... -8.333333 2.255639

2019-02-26 -1.744186 -11.111111 ... -8.333333 1.503759

2019-02-27 1.775148 -4.545455 ... -8.333333 -0.729927

2019-02-28 -1.156069 0.000000 ... -8.333333 0.735294

2019-03-01 0.000000 30.952381 ... 27.272727 0.740741

2019-03-04 1.785714 6.896552 ... 6.666667 3.007519

2019-03-05 0.000000 3.448276 ... 6.250000 0.740741

2019-03-06 -0.591716 0.000000 ... 5.882353 0.000000

2019-03-07 0.000000 -16.981132 ... 0.000000 0.729927

[10 rows x 1672 columns]

===============================

Covariance calculation result:

The first 5:

v\_updownpercent ...

comp\_code 3A A50CHIN-C22 ... ZELAN ZHULIAN

comp\_code ...

v\_updownpercent 3A 4.245010 -1.282610 ... -0.495149 0.111820

A50CHIN-C22 -1.282610 567.374970 ... -32.133610 3.254068

A50CHIN-C24 1.569546 105.290486 ... -2.538132 1.250299

A50CHIN-C26 -0.752828 82.788847 ... -3.007804 1.232525

A50CHIN-C28 -0.677478 83.384817 ... 11.642258 -0.594021

[5 rows x 1672 columns]

The Last 5:

v\_updownpercent ...

comp\_code 3A A50CHIN-C22 ... ZELAN ZHULIAN

comp\_code ...

v\_updownpercent YTLPOWR-C6 0.267697 -13.964786 ... -5.647462 -0.027413

YTLREIT 0.021665 -1.265344 ... -3.317369 0.233833

ZECON -0.883848 -0.559439 ... 11.303985 -0.981356

ZELAN -0.495149 -32.133610 ... 81.130752 -1.717133

ZHULIAN 0.111820 3.254068 ... -1.717133 1.057198

[5 rows x 1672 columns]

===============================

===============================

Covariance calculation result:

Positive:

comp\_code

v\_updownpercent HSI-H6L 108.826109

MENANG-WB 32.301458

RALCO-WB 24.221441

DAYA 16.746858

IBHD-WA 12.556882

Name: 3A, dtype: float64

Negative:

comp\_code

v\_updownpercent TRIVE -14.392797

MLAB-WA -21.147064

FGV-C55 -23.236681

EAH-WE -28.221645

PETGAS-CS -32.431844

Name: 3A, dtype: float64

===============================

Positive cov: ['3A', 'HSI-H6L', 'MENANG-WB', 'RALCO-WB']

Num Stock: 4

========================

Process finished with exit code 0