Project Report

Project Title:

Stock Price Analysis System Through Customized Pattern Shape

LIU, Xinzhu (EEGBM)

Supervisor

Prof. David Rossiter (CSE)

Table of Contents

1. Abstract

For individual stock investors, though have the intention to invest in stock market, lack of energy, time and resources limit their ability to go through the market detail like professionals do. Therefore, technical analysis is sometimes the preferred way to look into the market. This project is aimed to provide assistance for these individual stock investors using technical analysis by quickly identifying the most similar stock based on user's customized pattern shape. The project is achieved through a python program and this report will further explain in detail what major function this program do and how it achieves the function.

2. Introduction

2.1Background & Motivation

Roughly, there are three aspects of investment philosophies: fundamental analysis, technical analysis and quantitative analysis [1]. Fundamental analysis includes conducting industry research, company business and finance research. Institution investors usually have first-hand access to these materials. Individual investors would primarily rely on research report produced by financial institutions. Technical analysis is the research on the stock prices and prediction based on historical data. Quantitative analysis requires huge amount of data throughout stock market and time frame. This analysis is usually performed by engineering or related professionals with high level of degree. Therefore, for general individual stock investors, lack of energy, time and resources made technical analysis seems more appealing than the others. As technical analysis is essentially the research on stock patterns, predicting or describing how it moves. This project is determined to assist individual stock investors by allowing them to draw the patterns they need and quickly filter through the market to find the best match. The program starts certain given base code, which is the first block of the program, including grabbing the stock data from website, storing them into csv file and displaying them on screen.

2.2Framework & Concept

- Project Function: Find the BEST match between ANY user's desired pattern and current stock market data
- Project infrastructure: Python turtle and PIL library for user interface; Python urllib and webbrowser library for getting stock data from web; Python Numpy library for computation.
- Framework:

Generally speaking, there are four blocks inside this program, program flow shown as Figure 1 below:

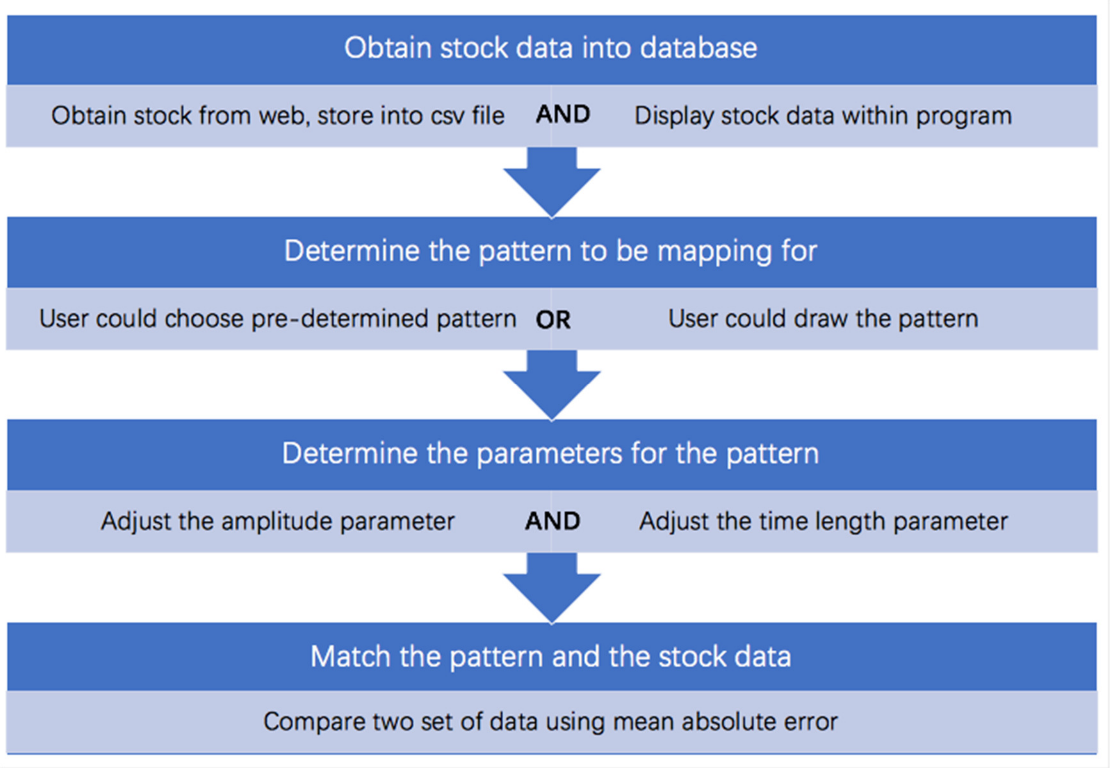

Obtain stock data into database
Obtain stock from web, store into csv file
AND
Display stock data within program
Determine the pattern to be mapping for
User could choose pre-determined pattern
OR
User could draw the pattern
Determine the parameters for the pattern
Adjust the amplitude parameter
AND
Adjust the time length parameter
Match the pattern and the stock data
Compare two set of data using mean absolute error

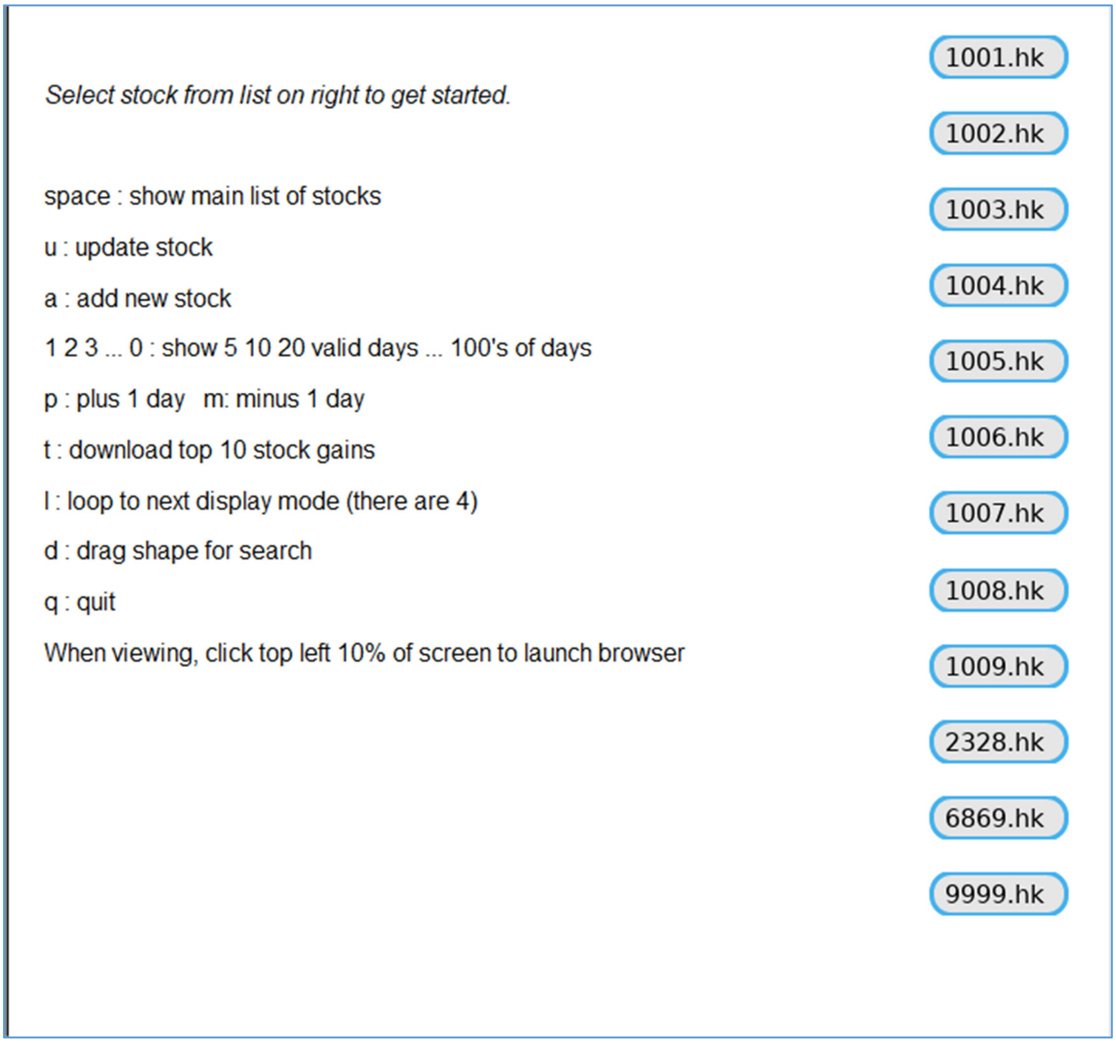
Figure 1: Program Data Flow

3. Methodology

The report would further elaborate the program methodology based on the four blocks of data flow.

3.1Obtain stock data into database

This part of the program is included in the base. This program starts with grabbing stock data from the internet and store it alongside with the program. Users could add their desired or interested stock to download and form their own watch list. At the home page which is shown as Figure 2 below, these stocks would be displayed at the right side of the screen as buttons. Users could click on the stock button to see the stock price plot, user could also change the display of the stock as well as adjusting the timeframe.



Select stock from list on right to get started.

space : show main list of stocks

u : update stock

a : add new stock

1 2 3 ... 0 : show 5 10 20 valid days ... 100's of days

p : plus 1 day   m: minus 1 day

t : download top 10 stock gains

l : loop to next display mode (there are 4)

d : drag shape for search

q : quit

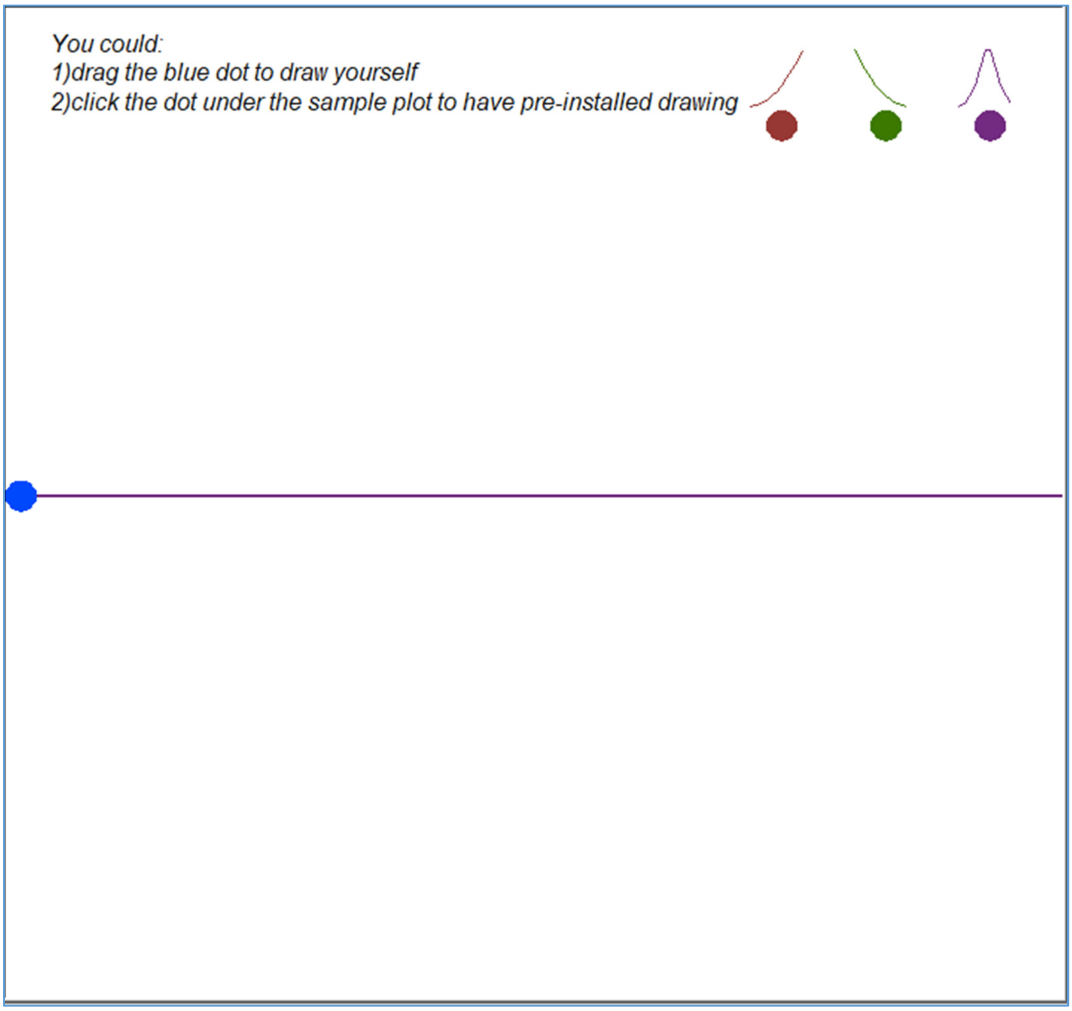
When viewing, click top left 10% of screen to launch browser

Figure 2: Program home page

3.2Determine the pattern to be mapping for

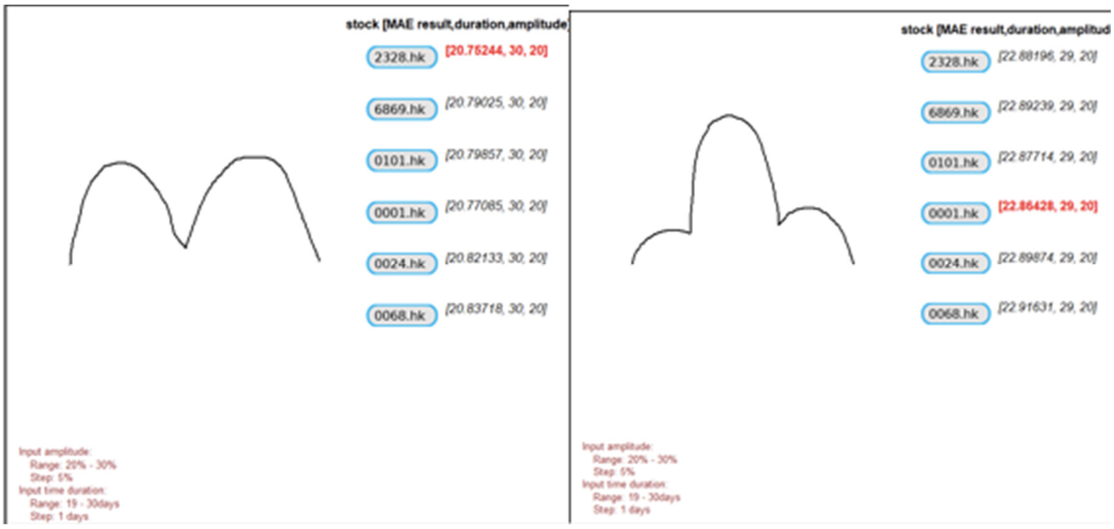
When user wish to use the main and core function of this program which is the draw and search, they first must determine a hypothetical pattern that they wish to be mapped for to be their customized pattern shape. There are two options that users could choose, display pages shown as Figure 3: 1)

Users draw the pattern using the mouse. 2) Users select desired pattern from pre-installed shapes. This search function could be used to search any customized shape users preferred using the self-drawing function. For example, the double tops, head and shoulders shown as Figure 4. After selecting the desired pattern from either of the approaches, the x and y coordinate of the sample dot that composed the desired pattern would be stored into a list (In this report, we would call it pattern list). Note the x and y coordinate only indicates the pattern shape, does not have any real-life implications.



You could:
1)drag the blue dot to draw yourself
2)click the dot under the sample plot to have pre-installed drawing

Figure 3: Pattern drawing starting page with no drawing taken place yet



```
stock [MAE result,duration,amplitude]
2328.hk  [20.75244, 30, 20]
6869.hk  [20.79025, 30, 20]
0101.hk  [20.79857, 30, 20]
0001.hk  [20.77085, 30, 20]
0024.hk  [20.82133, 30, 20]
0068.hk  [20.83718, 30, 20]

Input amplitude:
   Range: 20% - 30%
   Step: 5%
Input time duration:
   Range: 19 - 30days
   Step: 1 days
```

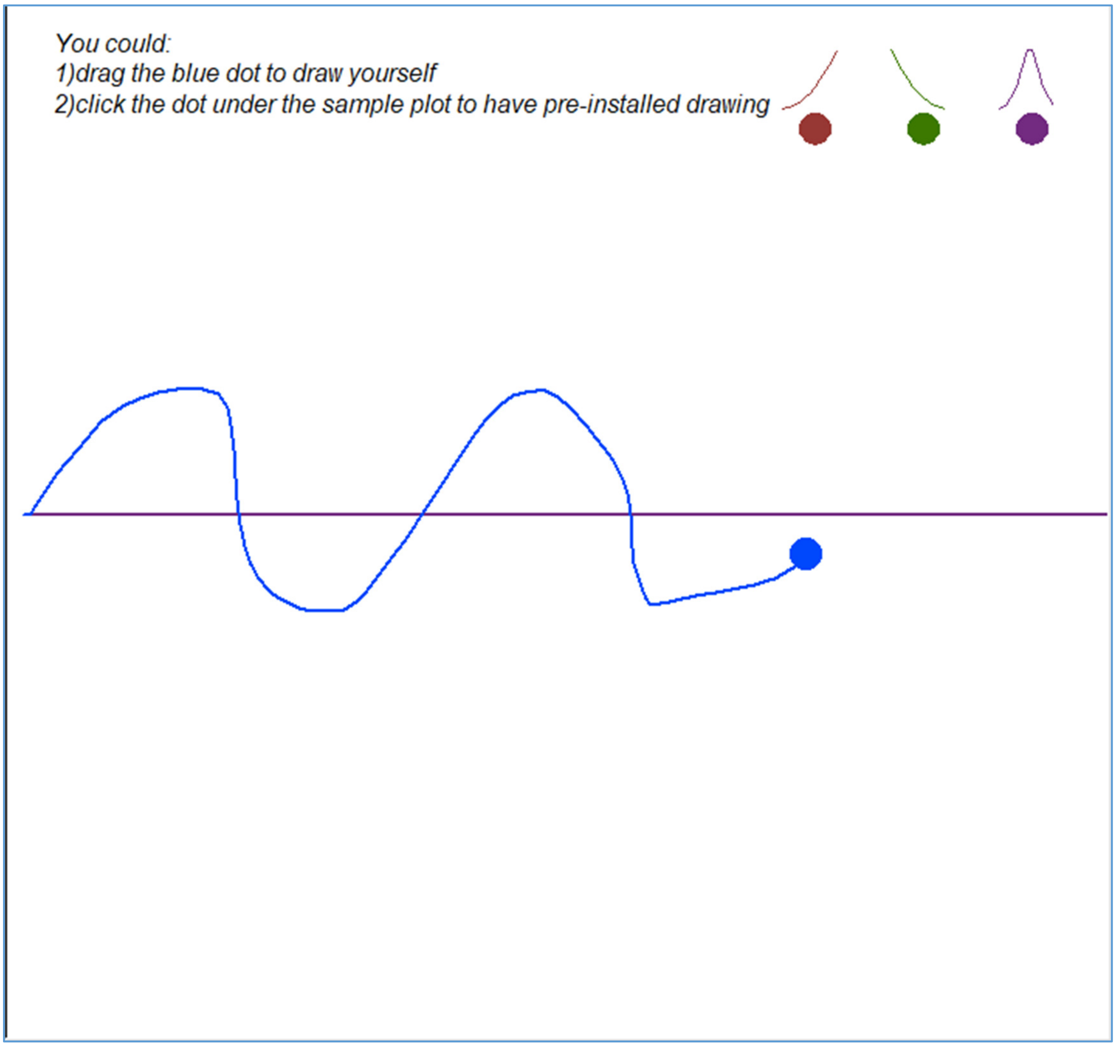
```
stock [MAE result,duration,amplitud
2328.hk  [22.88196, 29, 20]
6869.hk  [22.89239, 29, 20]
0101.hk  [22.87714, 29, 20]
0001.hk  [22.86428, 29, 20]
0024.hk  [22.89874, 29, 20]
0068.hk  [22.91631, 29, 20]

Input amplitude:
   Range: 20% - 30%
   Step: 5%
Input time duration:
   Range: 19 - 30days
   Step: 1 days
```

Figure 4: Customized pattern could be any shape user preferred (example left: double tops, example right: head and two shoulders)

3.2.1 Approach 1: User draw the pattern

Self-drawing function is achieved by turtle.ondrag(). User drag the python turtle which in this case is shown as a blue dot to draw the pattern. Upon dragging, the blue dot would be leaving a trail of blue line which is the line that user draw, giving user a clear idea what they have drawn, example shown as Figure 5. After releasing the turtle drag behavior, users then finish the processing of drawing pattern using approach 1, the x and y coordinate would be automatically stored into pattern list.

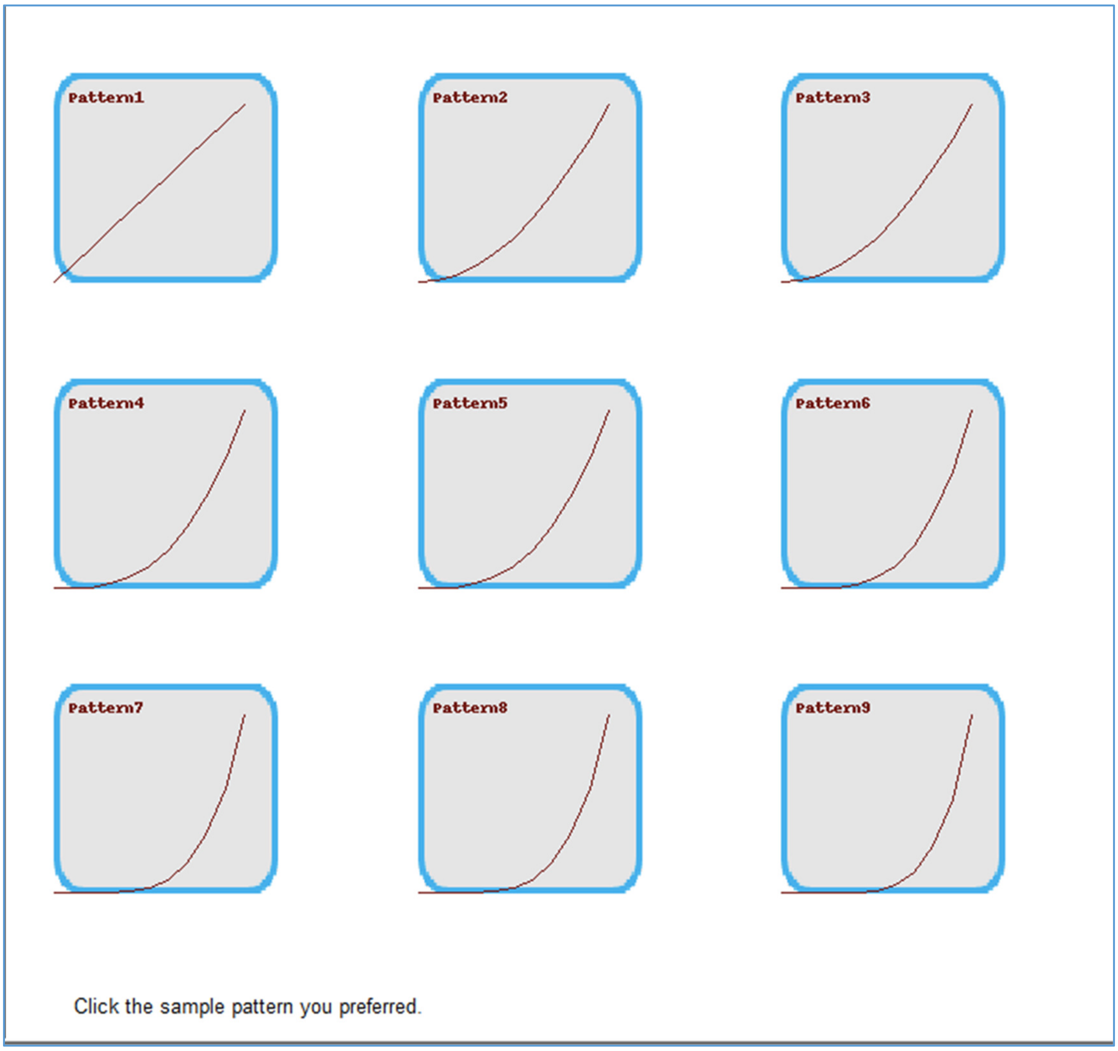


You could:
1)drag the blue dot to draw yourself
2)click the dot under the sample plot to have pre-installed drawing

Figure 5: Example drawing page showing the trail of the pattern drawn by user

3.2.2 Approach 2: User select from pre-installed pattern shapes

There are three sets of pre-installed pattern insider this program. 1) Upward slope 2) Downward slope 3) Pump and dump/pyramid shape. At the top of the pattern page, three example patterns representing the three sets of pre-installed pattern are shown. After selecting one of them, a new page with 9 pre-installed patterns of that shape would be shown, differentiating with each other slightly in shape, shown as the Figure 6 below. By clicking on one preferred shape out of these 9 shapes buttons, users then finish the process of selecting pattern using the approach 2, the x and y coordinate would be automatically stored into pattern list.



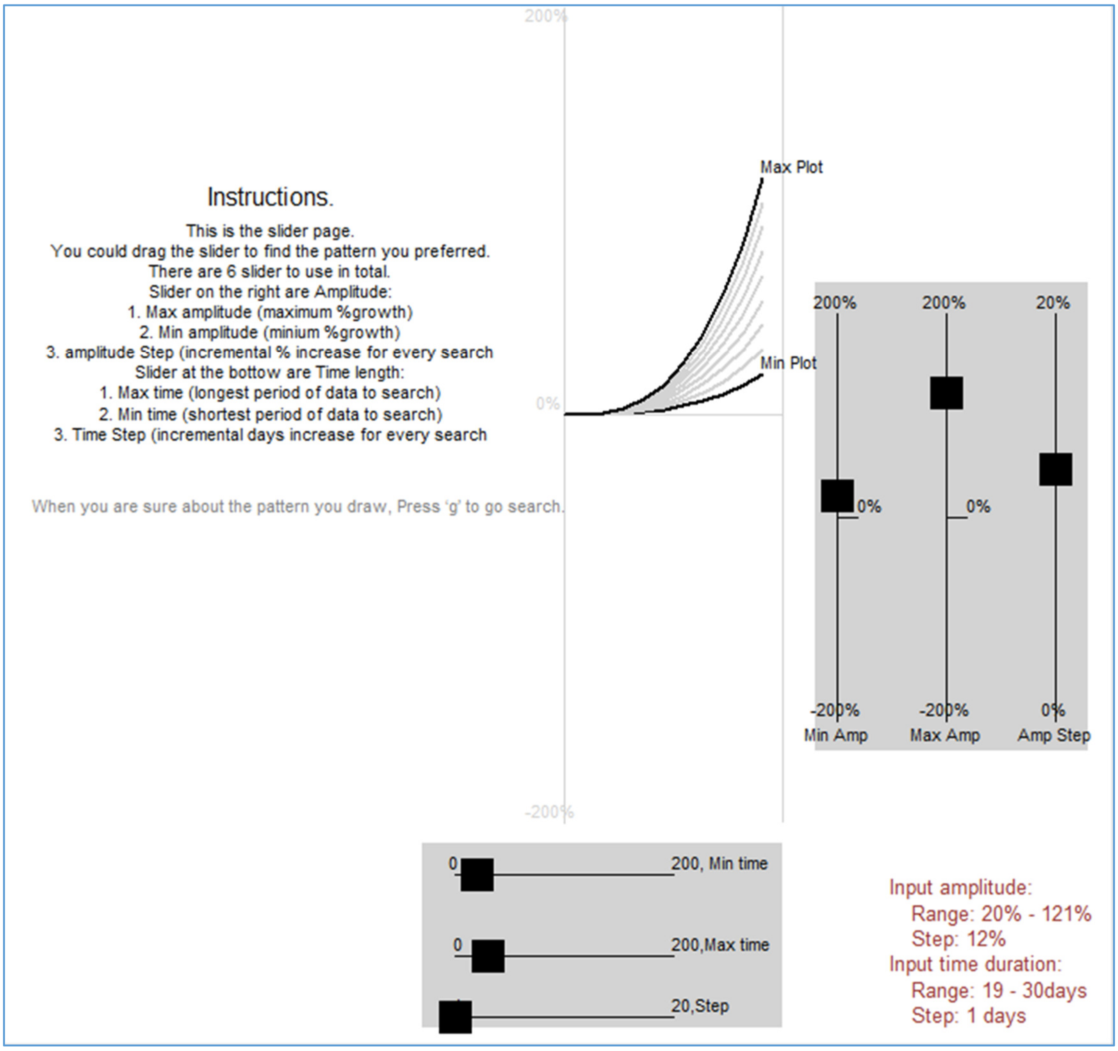
Click the sample pattern you preferred.

Figure 6: Pre-installed multiple patterns selection page

3.3Determine the parameters for the pattern

After selecting the desired pattern, users have so far only selected the shape of the pattern. There are certain parameters needed to be adjusted before going into the matching process. In this stage, a slider page would be shown whether using approach 1 or 2 when selecting patterns, shown in the Figure

7 below.



Instructions.

This is the slider page.
You could drag the slider to find the pattern you preferred.
There are 6 slider to use in total.
Slider on the right are Amplitude:
1. Max amplitude (maximum %growth)
2. Min amplitude (minium %growth)
3. amplitude Step (incremental % increase for every search
Slider at the bottow are Time length:
1. Max time (longest period of data to search)
2. Min time (shortest period of data to search)
3. Time Step (incremental days increase for every search

When you are sure about the pattern you draw, Press 'g' to go search.

Input amplitude:
Range: 20% - 121%
Step: 12%
Input time duration:
Range: 19 - 30days
Step: 1 days

Figure 7: Parameter slider page

3.3.1 Parameters that needed adjustment

The reason we need these parameters is for two reasons:

1) Transform one line of pattern into one set of lines

2) Adjust the pattern shape so that the x and y coordinates have real life implications, such as percentage growth and time.

Based on the objective stated above, we determined the parameter that needed adjustment, which are mainly two categories: amplitude/percentage growth and time length.

Amplitude/percentage growth is referred to stock prices, when we have a base day with a certain base price, all latter day prices would be shown as a percentage growth to that base price. So, the pattern data should be adjusted to the desired percentage growth/amplitude. There are three parameters regarding amplitude: 1) **amplitude max**, referring to the line with highest amplitude/growth percentage of the set of lines 2) **amplitude min**, referring to the line with lowest amplitude/growth percentages of the set of lines 3) **amplitude step**, referring to from highest amplitude and lowest amplitude, what should be the lines amplitude in between should have. For example, when amplitude max, amplitude min, amplitude step is 30,20,5 accordingly, the set of lines should have amplitude of 20,25,30, three lines in total.

Time length is referred to how many days of stock data should the matching process go through. Noted because of this is a program that aim to provide insight on real life investment decision, investment opportunity only applies at the current moment. Thus, there is no point if we found an investment opportunity in the past. Time length then is referred to how many days counting from today should be searching for. Same with amplitude, there are three parameters regarding time length: 1) **time length max**, referring to the longest days to be searching for 2) **time length min**, referring to the shortest days to be searching for 3) **time length step**, referring to from longest days to shortest days, what should the days searching for in between be. If set as default, the time length step would then be 1 days as conventional. For example, when time length max, time length min, time length step is 15,9,2 accordingly, pattern will be comparing with each stock data that contains 9, 11, 13, 15 days of stock price, meaning for each pattern with a certain amplitude, it will be comparing with each stock on watch list 4 times.

3.3.2 Using slider to adjust parameters

As could be seen from the picture shown in Figure 7, the parameters are adjusted using the slider. 6 parameters are displayed as 6 sliders which is also 6 python turtles. Users could drag the slider to the value they preferred. 6 python turtle position would then be recorded. For each time user drag, exact value of the 6 parameters would be shown at the corner of the page for clear message. After adjusting all 6 parameters, user could press certain keyboard to start the matching process.

3.4Match the pattern and the stock data

At the matching stage, pattern would go through three loops for matching.

First loop is to iterate the pattern with different amplitude, inside the first loop, the program goes into second loop, which is to go through the stock data on watch list which then turns to the third loop, time length which will determine how many days of stock data should be searching for each stock. Inside this loop, the program would perform the sampling, amplitude re-arranging about the pattern data, retrieve the desired number of days of the stock data and finally compare these two prepared sets of data using the mean absolute error approach.

3.4.1 Sample pattern data when needed

Inside the loop, sampling of the pattern is required because the data point quantity that formed the pattern shape may very much likely different from the data point quantity since we would search through different time length and the data point quantity of the pattern shape is fixed. Thus, in order to compare between these two points, these two sets need to have the same data point quantity to continue. To prevent any information loss from the real-life stock data as each of them refers to something concrete in the real world, it would be better to down-sampling or to fill data point using the pattern list. The method used to achieve this goal is called cubic interpolation [2]. Cubic interpolation works like this, when given a continuous set of real points, cubic interpolation gives the most probable point to add within that range using high degree equations. In this program, four data points are used to find the fifth point, for each given point, we could raise a third-degree equation with four unknowns as follows:

$$y = a_0 + a_1x^1 + a_2x^2 + a_3x^3$$

With four unknown variables and four equations, unknown variables could then be solved and input the desired x, y could then be found, hence the fifth point. For graphic example, refer to Figure 8 down below where case is given four exact points from f(x) and asked to give a fifth point within rage.

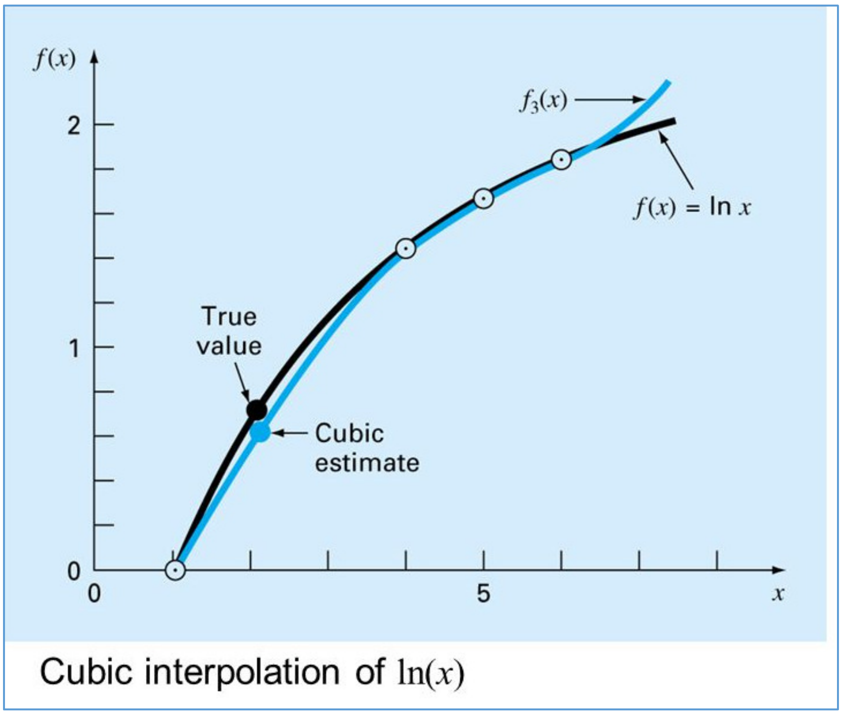


Figure 8: Cubic interpolation example: ln(x) estimation

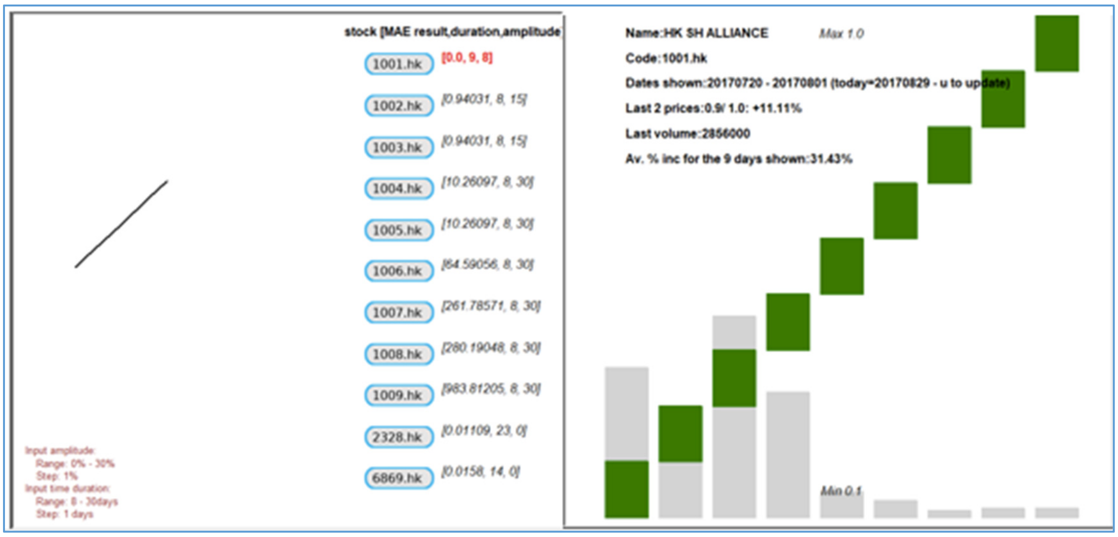
3.4.2 Mean Absolute Error comparison

When the two sets of data are both adjusted, at the core of these three loops is the Mean Absolute Error (MAE) method [3] to determine the likeliness about the pattern and the stock data. For this mean absolute error, the equation is as follows:

$$\mathrm{MAE} = \frac{\sum_{required\ \ time\ \ length}|stock\ \ amplitude - pattern\ \ amplitude|}{Time\ \ legth}$$

The formula indicates since both sets of data is in percentage form, the comparison could then be conducted using this relative term. Therefore, for the best stock that matches the pattern is the one that deviates the least from the pattern, which is exactly with the smallest mean absolute error.

For the result display, program displays each stock with its best MAE value corresponding with the amplitude and the time length that resulted in that MAE value. On top of that, the program would also highlight the best MAE value among all stocks on watch list as shown at the left part of Figure 9. For example, in Figure 9, user wants to search for an upward slope straight line which could be recapped from the left of the result page. As could be seen from the right-hand side of the display list, of all stocks, it is stock1001.hk that has the best MAE result, when clicking on the stock button, it could also be seen that the stock price of stock1001.hk is approximately an upward slope straight line. (Note: stock1001.hk here is a fake data list for demonstrating purpose.)



```
stock [MAE result,duration,amplitude]
1001.hk  [0.0, 9, 8]
1002.hk  [0.94031, 8, 15]
1003.hk  [0.94031, 8, 15]
1004.hk  [10.26097, 8, 30]
1005.hk  [10.26097, 8, 30]
1006.hk  [64.59056, 8, 30]
1007.hk  [261.78571, 8, 30]
1008.hk  [280.19048, 8, 30]
1009.hk  [983.81205, 8, 30]
2328.hk  [0.01109, 23, 0]
6869.hk  [0.0158, 14, 0]

Input amplitude:
   Range: 0% - 30%
   Step: 1%
Input time duration:
   Range: 8 - 30days
   Step: 1 days

Name:HK SH ALLIANCE
Code:1001.hk
Dates shown:20170720 - 20170801 (today=20170829 - u to update)
Last 2 prices:0.9/ 1.0: +11.11%
Last volume:2856000
Av. % inc for the 9 days shown:31.43%
```

Figure 9: Program result demonstration

4. Conclusion

For individual stock investors, though they have the intention to invest in stock market, lack of energy, time and resources limit their ability to go through the market detail like professionals do. Therefore, technical analysis is sometimes the preferred way to look into the market. This program is desired to help individual stock investors to quickly search through the stock market and find the stock with the stock pattern that they desired using python. The convenience that this program is limited and may varied upon different users. It does not provide a new solution for investment but it will greatly shorten the time needed to browse the market for someone with the need. Also, it provides another quick scan for stock investors that may have this need, they now have another approach to utilize.

5. References

1. Feng, Y., & Palomar, D. (2016). *A Signal Processing Perspective on Financial Engineering* (pp. 5-6). United States: Now Publishers. Inc.
2. *Cubic interpolation - Paulinternet.nl*. (2017). *Paulinternet.nl*. from http://www.paulinternet.nl/?page=bicubic Retrieved on 28 August 2017
3. *MAE and RMSE — Which Metric is Better?* (2017). *Human in a Machine World*. Retrieved 28 August 2017, from https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d