



# 致理科技大學

## 資訊管理系專題報告

### 股票分析系統

### Stock Analysis System

專題生：林廉定 (60510146)  
邵立巖 (60510145)  
趙汝佳 (60510136)  
吳芄鎡 (60510116)  
陳振琳 (60510115)  
謝品毅 (60510114)

指導教授：蘇啟鴻 博士

中華民國 108 年 06 月

致理科技大學

資訊管理系

畢業專題

股票分析系統

一〇八學年度

# 致理科技大學 專題報告審核書

本校 資訊管理 系

林廉定(60510146)、邵立巖(60510145)、

趙汝佳(60510136)、吳芃鎡(60510116)、

陳振琳(60510115)、謝品毅(60510114)

等君所提論文

股票分析系統(Stock Analysis System)

經本委員會審定通過，特此證明。

口試委員會

委員：

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

指導教授：\_\_\_\_\_

系主任：\_\_\_\_\_

中華民國 108 年 06 月

# 致理科技大學 授權書

本授權書所授權之專題報告在致理科技大學

108 學年度第 1 學期所撰寫。

**專題名稱： 股票分析系統**

本人具有著作財產權之論文或專題提要，授予致理科技大學，得重製成電子資料檔後收錄於該單位之網路，並與台灣學術網路及科技網路連線，得不限地域時間與次數以光碟或紙本重製發行。

本人具有著作財產權之論文或專題全文資料，授予教育部指定送繳之圖書館及本人畢業學校圖書館，為學術研究之目的以各種方法重製，或為上述目的再授權他人以各種方法重製，不限時間與地域，惟每人以一份為限。並可為該圖書館館藏之一。

本論文或專題因涉及專利等智慧財產權之申請，請將本論文或專題全文延至民國 年 月 日後再公開。

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。

(上述同意與不同意之欄位若未勾選, 本人同意視同授權)

同意 不同意

**學生簽名：**

(親筆正楷簽名)

**指導老師姓名：**

(親筆正楷簽名)

中華民國 年 月 日

# 摘要

專題報告名稱：股票分析系統

頁數：50

校系別：致理科技大學資訊管理系

完成時間：108 學年度第 1 學期

專題生：林廉定(60510146)、邵立巖(60510145)、趙汝佳(60510136)

吳芄鎡(60510116)、陳振琳(60510115)、謝品毅(60510114)

指導教授：蘇啟鴻 博士

關鍵詞：三大法人、Python、投資策略

股價變數與消息來源千百種，單靠個人的情報蒐集與分析能力就想掌握股市脈動常常會有力不從心的情況發生，但是少數擁有著雄厚資資金與人脈的法人就不一樣了，它們不僅有能力在第一時間獲得企業間的內幕消息，也有著利用傳播媒體散播特定資訊的資本。

本研究的目的就是要分析**三大法人**對於股票的買賣模式，從證交所提供的**三大法人**買賣超數據，觀察在法人的哪種**投資策略**下股價會有明顯的變化，並記錄起來作為我們的測試資料，當測試資料累積到一定程度時在依照歷年股價漲幅進行大幅度的投資模擬，進而找出最穩定，績效為正數的**投資策略**。

本系統使用善於處理文字與表格資料的 **Python** 語言，加上各種套件，協助系統將複雜的股市數據轉化為可讀性高的表格與圖表，並且可以進行參數的設定，調整對於獲利的期望值。

最後的測試以 2012 年的**三大法人**買賣超進行取樣，並將篩選過後的**投資策略**套用在 2018 的股市數據中，雖然最後的結果以虧損 1.4% 告終，但是鑒於 2018 年本來就屬於經濟不景氣的情況來看，可以將虧損控制在 1.4% 已經是很不錯的結果了。

# ABSTRACT

Thesis Title : Stock Analysis System

Pages : 50

University : Chihlee University of Technology

Graduate School : Department of Information Management

Date : June, 2019

Degree : Bachelor

Researcher : Lin,Lian-Ding (60510146) 、 Shao,Li-Yan (60510145)

Zhao,Ru-Jia (60510136) 、 Wu,Peng-Hong (60510116)

Chen,Zhen-Lin(60510115) 、 Xie,Pin-Yi (60510114)

Advisor : Dr. Su,Qi-Hong

Keywords : **Institutional Investors** 、 **Python** 、 **Investment Strategy**

There are thousands of stock price variables and news sources. It is often difficult to grasp the stock market pulsation by relying on personal intelligence gathering and analysis capabilities. However, a few legal persons with strong capital and contacts are not the same. They not only have The ability to obtain insider information between companies in the first place, as well as the use of the media to disseminate specific information.

The purpose of this study is to analyze **Institutional Investors** trading patterns of stocks. From **Institutional Investors** provided by the stock exchange to buy and sell super data, Observe that under the **investment strategy** of the legal person, the stock price will change significantly, and record it as us. The test data, when the test data has accumulated to a certain extent, is based on the historical stock price increase to carry out a large investment simulation, and then find the most stable, positive performance **investment strategy**.

The system uses a **Python** language that handles text and table data, plus a variety of kits to help the system convert complex stock market data into highly readable tables and charts, and parameter settings to adjust for profitable expectations.

The final test took the sampling of **Institutional Investors** transactions in 2012, and applied the selected **investment strategy** to the stock market data of 2018. Although the final result ended with a loss of 1.4%, it was originally an economic downturn in 2018. In the case of the situation, it is a very good result to be able to control the loss at 1.4%.

# 誌謝

感謝本專題的指導老師蘇老師，蘇老師淵博的專業知識，精益求精的工作風格，平易近人的人格魅力對我們影響深遠。特別感謝他給予我們研究方向，並叮嚀我們對於專題研究的注意事項以及研究過程中的小細節，讓我們的專題可以更貼近實際的股市市場，而不只是紙上談兵。

感謝班導曲老師熱心的幫助我們，曲老師鍥而不捨的教學態度，豐富的專題指導經驗，把學生們當成自己家人的包容心，感謝她不僅時時刻刻監督我們的專題進度，還不惜動用自己的課堂時間教導我們在報告上如何應對進退，給予我們時間進行討論，甚至適時的給予建議與提示，讓我們不是一昧的埋頭苦幹。

最後要感謝參與報告的同學們以及有著投資經驗的校外人士們，從蒐集資料、經驗分享、程式的測試改良、文件的製作、圖表與海報的繪製，到最後大家一起上台，正因為有大家熱情的參與並提供各方面的專業協助，這份專題才能成完整的呈現出來。



林廉定、邵立巖、趙汝佳、吳芄鎡、陳振琳、謝品毅 謹致  
致理科技大學 資訊管理 學士班  
中華民國 108 年 06 月

# 目錄

摘要 .....	i
ABSTRACT.....	ii
誌謝 .....	iii
目錄 .....	iv
<b>1. 緒論</b> .....	1
1.1 研究背景與動機 .....	1
1.2 研究目的或流程 .....	1
1.3 研究問題 .....	1
<b>2. 文獻回顧與探討</b> .....	2
2.1 機器學習 .....	2
2.2 三大法人 .....	2
2.3 數據分析 .....	4
2.4 Python .....	4
<b>3. 研究內容與方法</b> .....	10
3.1 研究背景 .....	10
3.2 研究動機與目的 .....	11
3.3 研究限制 .....	11
<b>4. 實驗結果與設計</b> .....	11
4.1 資料轉換 .....	12
4.2 規則推理 .....	12
4.3 交易模擬 .....	13
<b>5. 結論與建議</b> .....	14
參考文獻 .....	15
附錄 .....	16

# 1. 緒論

## 1.1 研究背景與動機

在家輕鬆的賺錢是許多人的夢想，投資股票市場不外乎是一種不錯的嘗試，然而股票市場卻不是依賴簡單的統計學就可以完全摸透的，經過與長輩們的討論與了解之後讓我們不禁開始思考，難道沒有一種方法可以讓我們在不熟悉市場行情的狀態下也能安心的投資股票嗎？

經過一系列的觀察後我們決定將目光放在三大法人身上，作為有著雄厚資本的外商與證券商，他們對於市場動向的掌握與直覺必定比我們還要好，這些大公司往往在暗中控制著媒體與股票市場，利用傳播媒體刻意散播對自己有利的消息，接著利用大筆資金操縱股市，進而製造大量利潤，只要可以了解這些操作手法，我們就可以搭上外商開出的順風車，獲得些許利益，比起摸透整個市場，分析他們的投資手段成功率更高。

所以我們決定以「股市分析」作為我們的專題研究主軸，從觀察並記錄三大法人的買賣數與當期股價的漲跌開始，逐步分析其中的關聯與規律，希望可以找出最安全最穩定，能夠獲利的股市投資策略。

## 1.2 研究目的或流程

本次專題是蒐集由台灣證券交易所提供的歷年各項股票數據，依靠周邊有著股市投資經驗的朋友們一同分析找出一個區間的交易數與股價漲幅的關聯性，再利用 Python 語言設計一個可以分析大量數據並進行模擬測試的系統，起初我們因為對於 Python 並不熟悉以及資料量過大導致載入歷年股市數據就花費兩個晚上，後續也遭遇到許多我們並不清楚該如何處理的問題，例如：數據分類的依據，變數的定義，邏輯運算的流程，在研究的中期也時常出現模擬結果與實際股市情形並不吻合的情況，但是經過長時間的學習與邏輯改良再加上指導老師提供的專業知識，目前系統已經可以找出三大法人買賣股票與股價漲跌幅的關聯。

## 1.3 研究問題

本研究針對的問題為：

問題一：

股價變動是否與三大法人買賣超有關聯？

問題二：

擅長資料處理的 Python 是否可以進行股票市場的分析與預測？

## 2. 文獻回顧與探討

### 2.1 機器學習

機器學習是一門多領域交叉學科，涉及概率論、統計學、逼近論、凸分析、演算法複雜度理論等多門學科。專門研究電腦怎樣模擬或實現人類的學習行為，以獲取新的知識或技能，重新組織已有的知識結構使之不斷改善自身的性能。它是人工智慧的核心，是使電腦具有智能的根本途徑。它的應用已遍及人工智慧的各個分支，如專家系統、自動推理、自然語言理解、模式識別、電腦視覺、智能機器人等領域。其中尤其典型的是專家系統中的知識獲取瓶頸問題，人們一直在努力試圖採用機器學習的方法加以克服。

學習能力是智能行為的一個非常重要的特征，但至今對學習的機理尚不清楚。人們曾對機器學習給出各種定義。H.A.Simon 認為，學習是系統所作的適應性變化，使得系統在下一次完成同樣或類似的任務時更為有效。R.s.Michalski 認為，學習是構造或修改對於所經歷事物的表示。從事專家系統研製的人們則認為學習是知識的獲取。這些觀點各有側重，第一種觀點強調學習的外部行為效果，第二種則強調學習的內部過程，而第三種主要是從知識工程的實用性角度出發的。

機器學習在人工智慧的研究中具有十分重要的地位。一個不具有學習能力的智能系統難以稱得上是一個真正的智能系統，但是以往的智能系統都普遍缺少學習的能力。例如，它們遇到錯誤時不能自我校正；不會通過經驗改善自身的性能；不會自動獲取和發現所需要的知識。它們的推理僅限於演繹而缺少歸納，因此至多只能夠證明已存在事實、定理，而不能發現新的定理、定律和規則等。隨著人工智慧的深入發展，這些局限性表現得愈加突出。正是在這種情形下，機器學習逐漸成為人工智慧研究的核心之一。它的應用已遍及人工智慧的各個分支，如專家系統、自動推理、自然語言理解、模式識別、電腦視覺、智能機器人等領域。其中尤其典型的是專家系統中的知識獲取瓶頸問題，人們一直在努力試圖採用機器學習的方法加以克服。

### 2.2 三大法人

何謂法人？何謂自然人？

在法律的眼裡有兩種人，一種是『自然人』、另外一種就是『法人』，像你我這種人生父母養的血肉之軀，就是『自然人』。

像公司、營利組織，是法律上賦予權力的一個團體，就是『法人』。

簡單來說，法人並非“人”，而是一個“團體”的代稱，但同樣具有權利能力的社會組織。

所謂三大法人？

其實就是『外資』、『投信』、『自營商』。

只看字面不容易了解，但看財經字典又太複雜，所以我們個別簡單介紹：

1. 「外資」：泛指台灣以外的資金的外國機構投資者，從國外把錢換成台幣後，在台灣投資，政府的中央銀行會看緊那些人限制他們的投資金額，把投資的錢送進來和送出國也都需要政府同意。

2. 「投信」：投資信託股份有限公司的簡稱，就是大眾口中的基金公司，募集投資人的錢代為操作。
3. 「自營商」：經過政府認可的證券公司，用自己的錢和證交所直接買賣。

### 三大法人的影響力！

你可能常常聽到叔叔阿姨們在買賣股票，可是你可能不知道，

散戶們的交易在市場上其實只是冰山的一角！

就很像在去菜市場買菜，一般人只買個一兩把青菜，老闆說：我說多少就是多少。

但如果有個人一去就買個一萬把青菜，老闆就會說：你說多少就多少！

現在大家能理解報章媒體成天報導三大法人動向的原因了吧！

#### 1. 直接影響力

三大法人的資金雄厚，只要開始買股票就可以把股價拉得很高；

賣股票的時候，可能一次就把股價打得很低。

這是三大法人對股市的直接影響。

#### 2. 間接影響力

間接影響是三大法人買進某檔個股，就會吸引本土大戶、散戶投資人買進

所以三大法人的買賣、持股狀況，成了籌碼分析的重要指標。

三大法人的交易資訊哪裡有？

現在已經是資訊便利的時代，只要到台灣證券交易所，

每天第一手的三大法人買賣資訊都在裡面。



108年05月31日 三大法人買賣註日報

證券代號	證券名稱	外資買進 張數(千張)	外資買進 金額(萬元)	外資賣出 張數(千張)	外資賣出 金額(萬元)	外資買賣 淨張數(千張)	外資買賣 淨金額(萬元)	自資買進 張數(千張)	自資買進 金額(萬元)	自資賣出 張數(千張)	自資賣出 金額(萬元)	自資買賣 淨張數(千張)	自資買賣 淨金額(萬元)	三大法人 買賣淨張數	三大法人 買賣淨金額
800771	富邦VIX	75,817,000	2,526,000	73,291,000	0	0	0	0	0	0	0	0	0	0	0
2888	聯華金	49,080,000	12,659,275	36,400,725	0	0	0	600,000	404,062	196,938	1,322,000	260,000	138,000	122,000	1,724,000
00637L	元大高深300 ETF	33,364,000	11,967,000	21,377,000	0	0	0	299,000	0	299,000	-8,626,000	121,000	100,000	21,000	25,974,000
2887	台新金	17,290,300	7,201,106	10,079,106	0	0	0	62,000	16,000	47,000	-44,000	0	44,000	-44,000	0
2891	中隆金	24,769,000	16,416,126	6,353,674	0	0	0	99,000	79,000	21,000	84,000	2,000	16,000	-14,000	100,000
3481	群創	32,676,000	24,831,347	7,843,653	0	0	0	105,067	-105,067	133,000	0	58,000	-85,000	523,000	336,000
2886	茂豐金	19,284,000	13,491,650	5,792,350	0	0	0	839,000	17,000	822,000	1,234,000	1,197,000	12,000	1,185,000	87,000
038413	聯群資訊 ETF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
00673R	元大50 ETF	1,548,000	5,000	1,543,000	0	0	0	0	0	0	0	0	0	0	0
2824	聯益金	12,820,000	6,871,168	6,626,832	0	0	0	25,000	-25,000	38,000	4,000	4,000	0	49,000	7,000
2324	仁寶	11,690,000	6,107,406	5,582,594	0	0	0	336,000	0	336,000	304,000	0	0	334,000	25,000
03211	元大光安 ETF	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3037	群興	10,839,000	5,625,000	5,214,000	0	0	0	368,000	160,000	238,000	306,000	593,000	252,000	341,000	73,000
1216	統一	11,636,000	6,102,736	5,532,264	0	0	0	11,000	19,000	-8,000	-77,000	58,000	88,000	-30,000	5,000
038696	聯群資訊 ETF	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(圖片來源：台灣證交所)

資料來源：CMoney

<https://www.cmoney.tw/notes/note-detail.aspx?nid=6234>  
108/6/2

### 2.3 數據分析

#### 數據分析的概念

數據分析是指通過建立審計分析模型對數據進行核對、檢查、復算、判斷等操作，將被審計單位數據的現實狀態與理想狀態進行比較，從而發現審計線索，搜集審計證據的過程。

#### 數據分析的目的與意義

數據分析的目的是把隱沒在一大批看來雜亂無章的數據中的信息集中、萃取和提煉出來，以找出所研究對象的內在規律。在實用中，數據分析可幫助人們作出判斷，以便採取適當行動。數據分析是組織有目的地收集數據、分析數據，使之成為信息的過程。這一過程是質量管理體系的支持過程。在產品的整個壽命周期，包括從市場調研到售後服務和最終處置的各個過程都需要適當運用數據分析過程，以提升有效性。例如J.開普勒通過分析行星角位置的觀測數據，找出了行星運動規律。又如，一個企業的領導人要通過市場調查，分析所得數據以判定市場動向，從而制定合適的生產及銷售計劃。因此數據分析有極廣泛的應用範圍。

### 2.4 Python

## 1 Python 是什麼？

各位好，今天我們要來介紹 Python 這一套功能強大、直譯式並且物件導向的程式語言。你或許聽過，或許沒聽過，但現在有機會來認識它了。

功能強大的直譯式程式語言不只有 Python，為什麼我們單單要來介紹它呢？Python 有幾項特點：

Python 具有許多物件導向的特性，然而並不要求一定得用物件導向的方式撰寫。撇開物件導向的議題不談，Python 通常可以靠比較短的程式碼完成比較多的功能，或者可以寫得比較清楚。

Python 自 1990 年由 Guido van Rossum 在荷蘭的 CWI 開始發展以來，從 0.9 進步到今天的 2.4.2，不但累積了相當完整的標準程式庫 (模組)，更有無以計數的非標準模組，而且絕大部分都是開放原始碼的。單以內建的模組來講，從簡單的數學運算、字串處理、網際網路協定連線、網際網路資料處理、各種壓縮格式，以及 POSIX 與主要作業系統的支援功能等等，含括的範圍非常地廣泛。

各種主要的作業系統都支援 Python。Python 程式常常不需要修改，便可以同時在 Linux 與 Windows 平台上執行，即使撰寫 GUI 程式 (透過 PyGTK, wxPython 等 binding) 也是一樣。所撰寫的 Python 程式透過標準的 distutils (模組) 進行包裝後，用標準的方式即可安裝於各種平台；在 Windows 下更可以自動產生方便的可執行 installer。

Python 算是執行效率不錯的直譯式語言，但畢竟比不上 C 和 Fortran。然而只要我們想，大可以 C/C++ 或 Fortran 撰寫高效率的模組；這些模組的使用方式，與內建模組以及用 Python 撰寫的模組完全一樣。最好的是，撰寫的方法並不困難。

Python 的出現讓計算機編程語言不再是生僻的專業技能，而是常人都能學習和使用的萬金油。

《經濟學人 (經濟學)》近日對的 Python 的一篇專題報導，揭秘了這一把計算機思維帶入尋常百姓家的神奇編程語言。

誕生於 1989 年 12 月，Python 最初是荷蘭計算機科學家 Guido van Rossum

為自己設立的一個的聖誕項目。

由於當時的其他編程語言都有各種缺點，這位科學家就想創建自己的編程語言。

而他建立編程語言的原則也很簡單：

首先，作為編程語言它應該易於閱讀，一個板塊內的每行命令只要採用縮進格式書寫即可，而不用被層層的花括號括起來；

其次，它應該讓用戶可以創建自己的專用編碼包，然後可以將其提供給其他人，使其作為新程序編寫的基礎；

第三，他想要一個「簡短，獨特且略帶神秘」的名字。

Python 是完全物件導向的語言。函式、模組、數字、字串都是物件。並且完全支援繼承、重載、衍生、多重繼承，有益於增強原始碼的複用性。Python 支援重載運算子，因此 Python 也支援泛型設計。相對於 Lisp 這種傳統的函數式程式語言，Python 對函式語言程式設計只提供了有限的支援。有兩個標準庫 (functools, itertools) 提供了與 Haskell 和 Standard ML 中類似的函數式程式設計工具。

雖然 Python 可能被粗略地分類為「手稿語言」，但實際上一些大規模軟體開發計劃例如 Zope、Mnet 及 BitTorrent，Google 也廣泛地使用它。Python 的支持者較喜歡稱它為一種高階動態程式語言，原因是「手稿語言」泛指僅作簡單程式設計任務的語言，如 shell script、VBScript 等只能處理簡單任務的程式語言，並不能與 Python 相提並論。

Python 本身被設計為可擴充的。並非所有的特性和功能都整合到語言核心。Python 提供了豐富的 API 和工具，以便程式設計師能夠輕鬆地使用 C、C++、Cython 來編寫擴充模組。Python 編譯器本身也可以被整合到其它需要手稿語言的程式內。因此，有很多人把 Python 作為一種「膠水語言」使用。使用 Python 將其他語言編寫的程式進行整合和封裝。在 Google 內部的很多專案，例如 Google 應用服務引擎使用 C++ 編寫效能要求極高的部分，然後用 Python 或 Java/Go 調用相應的模組。[6]《Python 技術手冊》的作者馬特利 (Alex Martelli) 說：「這很難講，不過，2004 年，Python 已在 Google 內部使用，Google 招募許多 Python 高手，但在這之前就已決定使用 Python。他們的目的是儘量使用 Python，在不得已時改用 C++；在操控硬體的場合使用 C++，在快速開發

時候使用 Python。」[7]

Python 的設計哲學是「優雅」、「明確」、「簡單」。Python 開發者的哲學是「用一種方法，最好是只有一種方法來做一件事」，也因此它和擁有明顯個人風格的其他語言很不一樣。在設計 Python 語言時，如果面臨多種選擇，Python 開發者一般會拒絕花俏的語法，而選擇明確沒有或者很少有歧義的語法。這些準則被稱為「Python 格言」。在 Python 解釋器內運行 `import this` 可以獲得完整的列表。

Python 的設計哲學強調程式碼的可讀性和簡潔的語法

### 應用範圍

#### 網路應用程式

Python 經常被用於 Web 開發。比如，通過 `mod_wsgi` 模組，Apache 可以運行用 Python 編寫的 Web 程式。使用 Python 語言編寫的 Gunicorn 作為 Web 伺服器，也能夠執行 Python 語言編寫的 Web 程式。Python 定義了 WSGI 標準應用介面來協調 Http 伺服器與基於 Python 的 Web 程式之間的溝通。一些 Web 框架，如 Django、Pyramid、TurboGears、Tornado、web2py、Zope、Flask 等，可以讓程式設計師輕鬆地開發和管理複雜的 Web 程式。

Python 對於各種網路協定的支援很完善，因此經常被用於編寫伺服器軟體、網路爬蟲。第三方函式庫 Twisted 支援非同步線上編寫程式和多數標準的網路協定（包含用戶端和伺服器），並且提供了多種工具，被廣泛用於編寫高性能的伺服器軟體。另有 `gevent` 這個流行的第三方庫，同樣能夠支援高性能並行的網路開發。

#### GUI 開發

Python 本身包含的 Tkinter 庫能夠支援簡單的 GUI 開發。但是越來越多的 Python 程式設計師選擇 wxPython 或者 PyQt 等 GUI 套件來開發跨平台的桌面軟體。使用它們開發的桌面軟體執行速度快，與用戶的桌面環境相契合。通過 PyInstaller 還能將程式釋出為獨立的安裝程式包。

#### 作業系統

在很多作業系統裡，Python 是標準的系統元件。大多數 Linux 發行版和 Mac OS X 都整合了 Python，可以在終端機下直接執行 Python。有一些 Linux 發行版的安裝器使用 Python 語言編寫，比如 Ubuntu 的 Ubiquity 安裝器、Red Hat Linux 和 Fedora 的 Anaconda 安裝器。在 RPM 系列 Linux 發行版中，有一些系統元件就是用 Python 編寫的。Gentoo Linux 使用 Python 來編寫它的 Portage 軟體包管理系統。Python 標準庫包含了多個調用作業系統功能的函式庫。通過

pywin32 這個第三方軟體包，Python 能夠存取 Windows 的 COM 服務及其它 Windows API。使用 IronPython，Python 程式能夠直接調用 .Net Framework。

### 科學計算

NumPy、SciPy、Matplotlib 可以讓 Python 程式設計師編寫科學計算程式。有些公司會使用 Scons 代替 make 構建 C++ 程式。

很多遊戲使用 C++ 編寫圖形顯示等高效能模組，而使用 Python 或者 Lua 編寫遊戲的邏輯、伺服器。相較於 Python，Lua 的功能更簡單、體積更小；而 Python 則支援更多的特性和資料類型。很多遊戲，如 EVE Online 使用 Python 來處理遊戲中繁多的邏輯。

YouTube、Google、Yahoo!、NASA 都在內部大量地使用 Python。OLPC 的作業系統 Sugar 項目的大多數軟體都是使用 Python 編寫。

### 使用 Python 編寫的著名應用

Youtube - 影片社群網站

Reddit - 社交分享網站

Dropbox - 檔案分享服務

豆瓣網 - 圖書、唱片、電影等文化產品的資料資料庫網站

Django - 鼓勵快速開發的 Web 應用框架

Pylons - Web 應用框架

Zope - 應用伺服器

Plone - 內容管理系統

Instagram - 是一款免費提供線上圖片及影片分享的社交應用軟體,使用 Django 作為後台

TurboGears - 另一個 Web 應用快速開發框架

Twisted - Python 的網路應用程式框架

Fabric - 用於管理成百上千台 Linux 主機的程式庫

Python Wikipedia Robot Framework - MediaWiki 的機器人程式

MoinMoinWiki - Python 寫成的 Wiki 程式

Trac - 使用 Python 編寫的 BUG 管理系統

Mailman - 使用 Python 編寫的郵寄清單軟體

Mezzanine - 基於 Django 編寫的內容管理系統

Flask - Python 微 Web 框架

Webpy - Python 微 Web 框架

Bottle - Python 微 Web 框架

EVE - 網路遊戲 EVE 大量使用 Python 進行開發

Blender - 使用 Python 作為建模工具與 GUI 語言的開源 3D 繪圖軟體

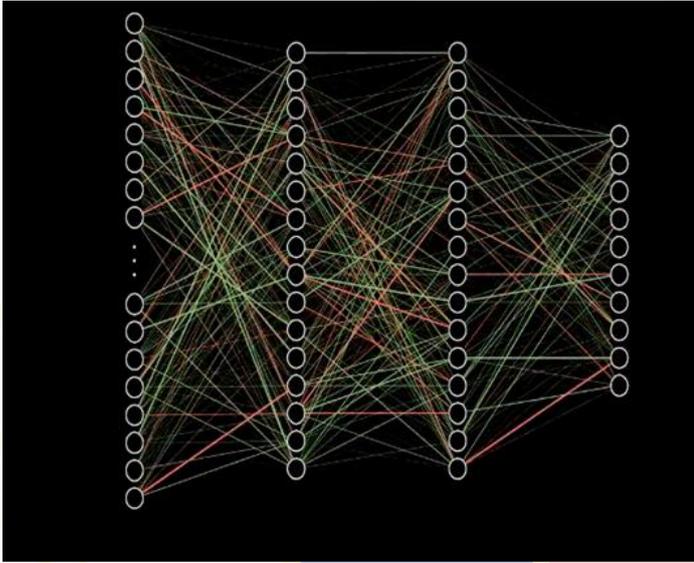
Inkscape - 一個開源的 SVG 向量圖形編輯器。

知乎 - 一個問答網站

果殼 - 一個泛科技主題網站

Odoo-仍在持續發展壯大且最受歡迎的 ERP 軟體

圖片來源：TechOrange 科技報橘)



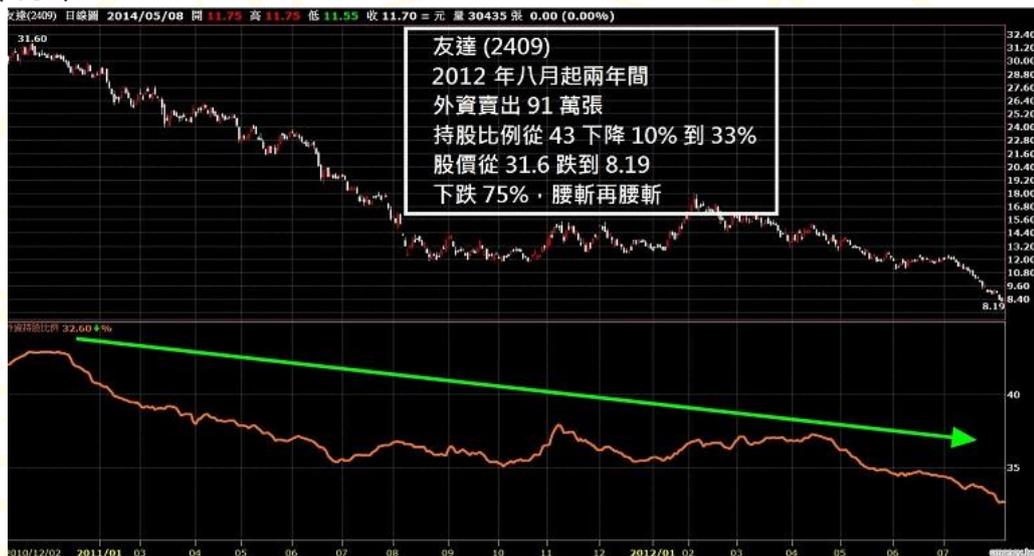
## 3.研究內容與方法

### 3.1 研究背景

股市中，三大法人掌握著股票漲跌的關鍵，當他們開始買股、股票可能就要漲，當他們開始賣、股票可能就要跌，而觀察他們的動向則可以讓投資者事先抓住時機。三大法人的資金雄厚，只要開始買股票就可以把股價拉得很高；賣股票的時候，可能一次就把股價打得很低。這是三大法人對股市的直接影響。下面提供一個例子，可以看出法人出手，會對股價有重大影響：

友達 (2409) 2012 年八月起遭外資大砍，

股價跌到 8 元



三大法人的錢非常多，只要開始買股票就可以把股價拉得很高，賣股票的時候可能一次就把股價打得很低，這是三大法人對股市的直接影響除此之外，間接影響也會常常出現例如：外資買進某檔個股就會吸引本土大戶、散戶投資人買進所以三大法人的買賣、持股狀況，成了籌碼分析的重要指標。

上市	持股比率 (%)	近20日成交比重 (%)	近20日買賣超
外資	38.05%	53.01%	643.70億
投信	1.40%	3.11%	-7.09億
自營商	0.17%	16.11%	77.27億
合計	<b>39.62%</b>	<b>72.23%</b>	713.88億
上櫃	持股比率 (%)	近20日成交比重 (%)	近20日買賣超
外資	18.85%	14.76%	39.54億
投信	1.87%	3.71%	6.25億
自營商	0.66%	9.70%	16.57億
合計	<b>21.38%</b>	<b>28.17%</b>	62.36億

證交所每天都會統計公布台股大盤(所有股票)合計的三大法人買賣金額，以及公布個股個別的三大法人買賣金額。

### 3.2 研究動機與目的

依據上述關於三大法人與股價的關聯，我們覺得這對於還不熟悉股市交易的頭資者來說會是一個很好的出發點，如果三大法人的投資策略與股價有著直接的關係的話，投資者就算不懂怎麼分析 K 線或是 SMA 線也能跟著三大法人的動向獲利，所以我們此研究主要是藉由紀錄證交所每天所統計的三大法人買賣金額，以及後續股價漲跌幅度，將其整理成一種投資公式，並投入到股票市場中進行模擬，驗證只依據三大法人的買賣金額是否可以穩定的獲利。

### 3.3 研究限制

資訊化系統設備在資訊化的時代，撰寫程式時會遇到系統出錯，導致程式無法執行，部分的程式語言皆預老師指導，也增加了我們的一些難度。

時間、人力及財力之受限本專題研究因是在就學期間，時間並不算太過充裕，以至於本研究計畫，無法有更詳盡且完整的研究，在財力方面由於是學生，所以導致財力上有些許不足。

## 4. 實驗結果與設計



我們將證交所提供的 2012 年至 2017 年所有股市交易資料經由 Python 的套件轉換為表格，其中包含股票代碼、交易量、交易日期、開盤價、最高價、最低價、收盤價

	ID	Date	Volume	Open	High	Low	Close	+/-	InvestTrade
0	A0050	2012-05-10	7132826	51.6	52.0	51.55	51.70	0.000000	-3650000
1	A0050	2012-05-11	29045919	51.6	51.6	51.05	51.15	-1.063830	-23573000
2	A0050	2012-05-14	10200944	51.3	51.3	50.95	51.20	0.097752	-4481000
3	A0050	2012-05-15	21664396	50.9	51.2	50.55	51.20	0.000000	-11092000
4	A0050	2012-05-16	32978827	51.0	51.1	50.10	50.20	-1.953125	-27283000

### 4.1 資料轉換

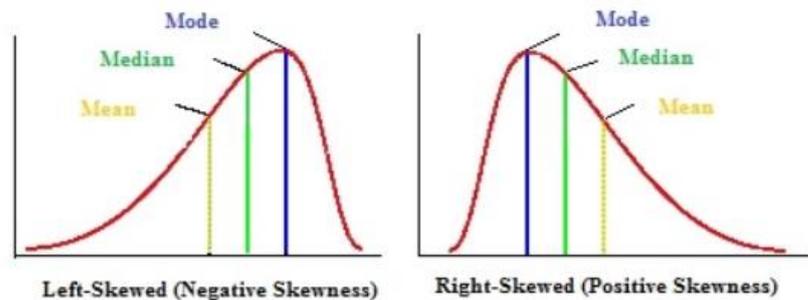
- 計算每日漲跌幅%。
- 每一筆交易資料(每一天的每一檔股票交易資料，含法人買賣)增加五個欄位，前一日到前五日 (D-1, D-2, D-3, D-4, D-5)三大法人買賣超佔總交易量的比例。
- 每一筆資料再增加五個欄位，未來五天 (D, D+1, D+2, D+3, D+4)的股價漲跌幅%。
- 將三大法人買賣佔總交易量的比例 D-1, D-2, ... D-5 轉換為標籤 (Labels)，(大買，中買，小買，沒買，小賣，中賣，大賣) ==> (3, 2, 1, 0, -1, -2, -3)。
- 計算 D 日開始後五日股價累計漲跌幅最大值，也就是最樂觀的結果，我們稱為 rewards。下圖就是將資料整理出來後產生的表格。

ID	Date	Volume	Open	High	Low	Close	+/-	InvestTrade	D-1	D-2	D-3	D-4	D-5	D	D+1	D+2	D+3	D+4	D-1_Close
0	A0050	7132826	51.60	52.00	51.55	51.70	0.000000	-3650000	-3	-3	-3	-3	0.000000	-1.063830	0.097752	0.000000	-1.953125	51.70	
1	A0050	29045919	51.60	51.60	51.05	51.15	-1.063830	-23573000	-3	-3	-3	-3	-1.063830	0.097752	0.000000	-1.953125	1.195219	51.70	
2	A0050	10200944	51.30	51.30	50.95	51.20	0.097752	-4481000	-3	-3	-3	-3	0.097752	0.000000	-1.953125	1.195219	-2.224409	51.15	
3	A0050	21664396	50.90	51.20	50.55	51.20	0.000000	-11092000	-3	-3	-3	-3	0.000000	-1.953125	1.195219	-2.224409	0.563721	51.20	
4	A0050	32978827	51.00	51.10	50.10	50.20	-1.953125	-27283000	-3	-3	-3	-3	-1.953125	1.195219	-2.224409	0.563721	0.900901	51.20	
5	A0050	7940204	50.25	50.85	50.25	50.80	1.195219	-2575000	-3	-3	-3	-3	1.195219	-2.224409	0.563721	0.900901	-1.805556	50.20	
6	A0050	37830294	50.05	50.25	49.62	49.67	-2.224409	-29659000	-2	-3	-3	-3	-2.224409	0.563721	0.900901	-1.805556	-2.222267	50.80	

### 4.2 規則推理

- 我們將所有交易資料(2012-05-02~2017-12-31 之間的資料)共，利用七個標籤分組 (groupby, 類似 SQL 的 Group by 作業)，理論上有  $7 * 7 * 7 * 7 * 7 = 16807$  個分組，實際上有些分組沒有資料。

- 計算各分組的敘述統計量。如果某一個分組(某一種 pattern 出現的情況)，reward 的平均值夠大，例如 3%，表示這種 pattern 出現的條件下，後面五天，最樂觀的漲幅平均值是 3%。當然我們還要考慮 minimum, std (standard deviation, 標準差)，甚至可以考慮偏態 (skewness)。
- 簡單來說，下圖(X 軸是漲跌幅%，Y 軸是發生次數 Frequency)，所以這個圖就是 Histogram)左邊的圖偏態是正值，所以向右偏，右圖的偏態是負值，所以向左偏。我們想要找的規則是，在那些條件下：
  - \*\* Skewness 正值越大越好(越往右偏)
  - \*\* Mean 越大越好(整個圖形往右平移)
  - \*\* std 越小越好(發散程度不大，較規律)
  - \*\* min 越大越好(風險較低)



### 4.3 交易模擬

- 依照前述方法找出一組規則之後，我們用來模擬 2018 年整年的交易(規則導出時沒有採用 2018 年的資料)。
- 模擬結果可以看出這些規則的績效。

在模擬過程中，我們會制定期望值，當規則觸發時電腦會依照上一天的收盤價來出價購買股票，如果買進則等到此股價漲到期望金額或是跌到最大限度容忍值時賣出，不斷循環操作，並將過程都記錄下來。如下圖所示。

```

Date: 2018-01-02
觸發規則: 7, on date= 2018-01-02 ,stock= A1474 , and rule mean= 2.52
觸發規則: 0, on date= 2018-01-02 ,stock= A2368 , and rule mean= 2.74
觸發規則: 7, on date= 2018-01-02 ,stock= A2538 , and rule mean= 2.52
觸發規則: 6, on date= 2018-01-02 ,stock= A2883 , and rule mean= 2.53
觸發規則: 7, on date= 2018-01-02 ,stock= A3164 , and rule mean= 2.52
*買進: A3164 , 出價 22.655 >= 當日最低價 22.65
觸發規則: 9, on date= 2018-01-02 ,stock= A3189 , and rule mean= 2.51
觸發規則: 10, on date= 2018-01-02 ,stock= A6214 , and rule mean= 2.5
觸發規則: 8, on date= 2018-01-02 ,stock= A6525 , and rule mean= 2.51
庫存= 1
Date: 2018-01-03
觸發規則: 0, on date= 2018-01-03 ,stock= A2388 , and rule mean= 2.74
觸發規則: 7, on date= 2018-01-03 ,stock= A2637 , and rule mean= 2.52
觸發規則: 7, on date= 2018-01-03 ,stock= A3056 , and rule mean= 2.52
觸發規則: 3, on date= 2018-01-03 ,stock= A3416 , and rule mean= 2.6
觸發規則: 2, on date= 2018-01-03 ,stock= A4725 , and rule mean= 2.6
觸發規則: 0, on date= 2018-01-03 ,stock= A6189 , and rule mean= 2.74
庫存= 1
Date: 2018-01-04

```

- 2018 年的模擬結果：

規則觸發次數= 560  
買到次數= 127  
買到且賣掉 126  
買到沒賣掉(期末持股)= 1  
賺次數= 77  
虧次數= 49  
庫存市值績效(已扣除手續費與證交稅)= 0.36 %  
已實現績效= 34.35 %  
整體績效(包含庫存盈虧)= 34.71 %

- 2019 年 1 月到 3 月的模擬結果

規則觸發次數= 73  
買到次數= 12  
買到且賣掉 11  
買到沒賣掉(期末持股)= 1  
賺次數= 10  
虧次數= 1  
庫存市值績效(已扣除手續費與證交稅)= 1.83 %  
已實現績效= 36.4 %  
整體績效(包含庫存盈虧)= 38.23 %

- 依照上述的模擬結果可以確定三大法人購買股票與股價確實是有關連的，而我們也可以利用這項資訊進行股市投資。

## 5.結論與建議

在研究股票分析當中，我們在撰寫程式的時候遇到了一些問題，我們也嘗試自行構想出解決方法，或是詢問老師以及上網搜尋資料，來解決我們的問題，也從中學習到我們在課堂中所學習不到的思想和知識，也希望在未來發展過程當中，能加入更多我們現在所不知道的變數，由於本專題在時間、人力及財 都有所受限下，時間並不算太過充裕，以至於本研究計畫，無法有更詳盡且完 整個研究，以及財力方面由於是學生，導致財力上有些許不足，所以希望在未來我們可以往不同的方向去做發展，也可以讓此股票分析達到更廣泛、穩定的應用，甚至可以套用在其他分析中。

## 參考文獻

- [1] Python: <http://www.python.org/>
- [2] (1, 2) Python 下載頁: <http://www.python.org/download/>
- [3] PEP 289 Generator Expressions in Python 2.4:  
<http://docs.python.org/whatsnew/node4.html>
- [4] PEP 318 Decorators for Functions and Methods in Python 2.4:  
<http://docs.python.org/whatsnew/node6.html>
- [5] Python 2.4 新增項目: <http://docs.python.org/whatsnew/whatsnew24.html>
- [6] PIL: <http://www.pythonware.com/products/pil/>
- [7] 葉平的 Python 教材:  
<http://www.ossacc.org/Download/TM/OssaccTM/Python/>
- [8] Python Tutorial: <http://docs.python.org/tut/tut.html>
- [9] Python Language Reference: <http://docs.python.org/ref/ref.html>
- [10] Dive into Python: <http://diveintopython.org/>
- [11] SciPy 的效能比較:  
<http://www.scipy.org/documentation/weave/weaveperformance.html>
- [12] Python web programming: <http://wiki.python.org/moin/WebProgramming>
- [12]TechOrange 科技報橘 108/6/2



畢業專題 系統操作手冊  
附錄

【專題執行計畫表】

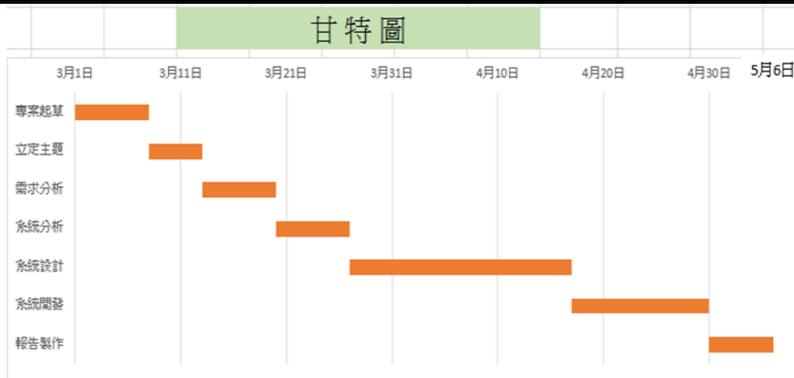
組名	清一色		
組員	班 級	學 號	姓 名
	夜資三 A	60510114	謝品毅
	夜資三 A	60510115	陳振琳
	夜資三 A	60510116	吳芄鈺
	夜資三 A	60510136	趙汝佳
	夜資三 A	60510145	邵立嚴
	夜資三 A	60510146	林廉定
選定合作單位	名稱	英福達	
	負責人		聯絡人
	電話		電話
	地址		
	業務描述	利用國際間的時差，從美國的收盤價去推估台灣股市的開盤價。	
專題名稱	股票分析系統		
專題資訊系統功能描述	<p>每天早上 5:30 到國外的財經網站抓取美股收盤價，因為時差的關係美國股市同一天的收盤一定比台灣股市同一天開盤早，又因美股與台股有所關聯，系統可以藉此在台股開盤前推測當天的漲跌情況。</p>		
指導老師 簽名		日期	2019 年 02 月 17 日
備註			

**【專題工作分配表】**

組名	清一色	填寫人					
專題名稱	股票分析系統	填寫日期	年月日				
主要編號	主要工作項目	主要成員姓名			機動成員姓名		
		邵立巖同學	林廉定同學	陳振琳同學	謝品毅同學	趙汝佳同學	吳芄鉉同學
	專案起草	V	V	V	V	V	V
	立定主題	V	V		V		V
	需求分析	V	V			V	V
	系統分析	V	V				V
	系統設計	V	V	V	V		
	系統開發	V					
	報告製作		V	V	V	V	
請打✓ 請打✓							

**【GANTT 圖】**

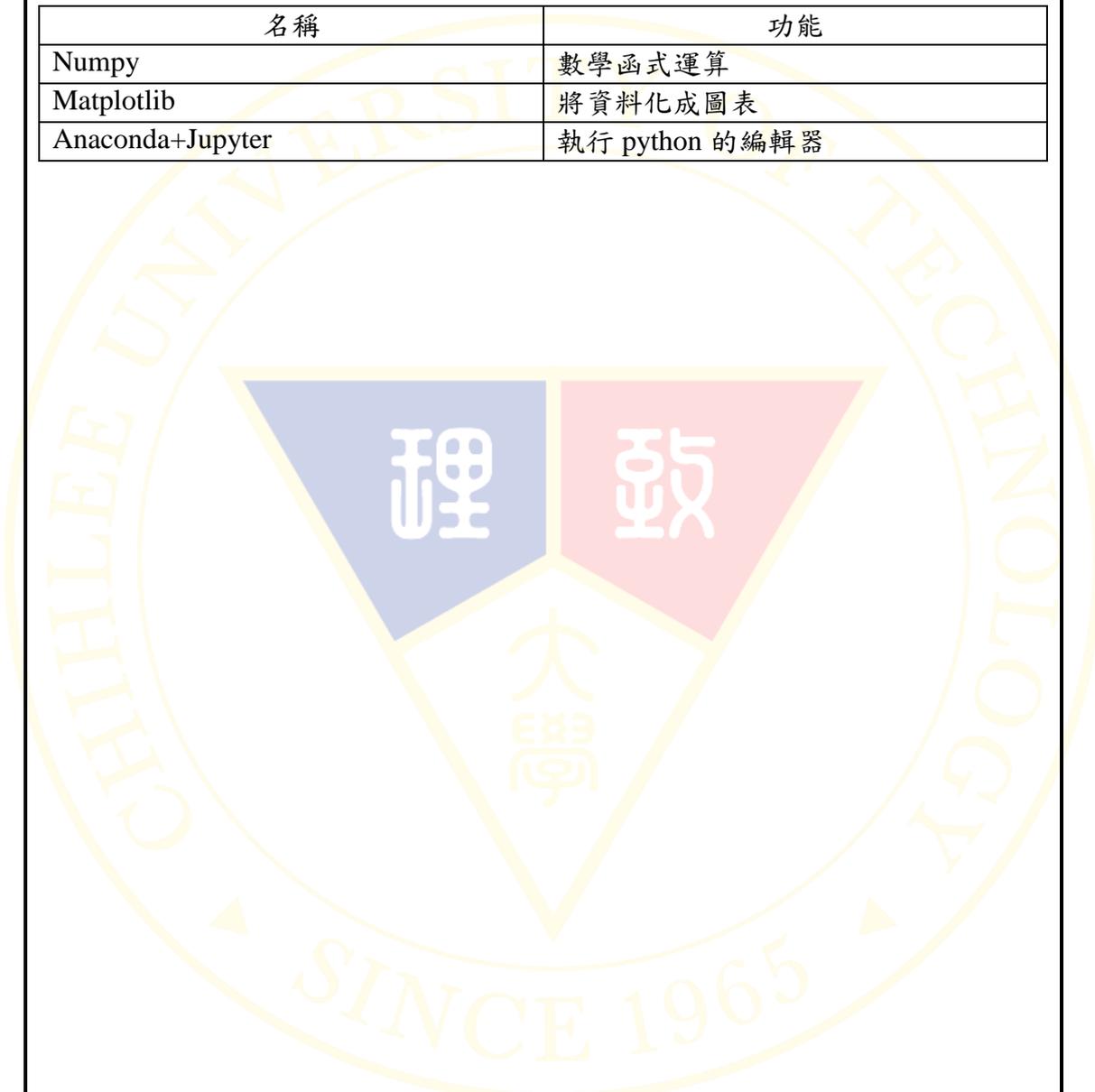
組名	清一色	填寫人	
專題名稱	股票分析系統	填寫日期	年月日



### 【開發工具清單】

組名	清一色	填寫人	
專題名稱	股票分析系統	填寫日期	年月日

名稱	功能
Numpy	數學函式運算
Matplotlib	將資料化成圖表
Anaconda+Jupyter	執行 python 的編輯器



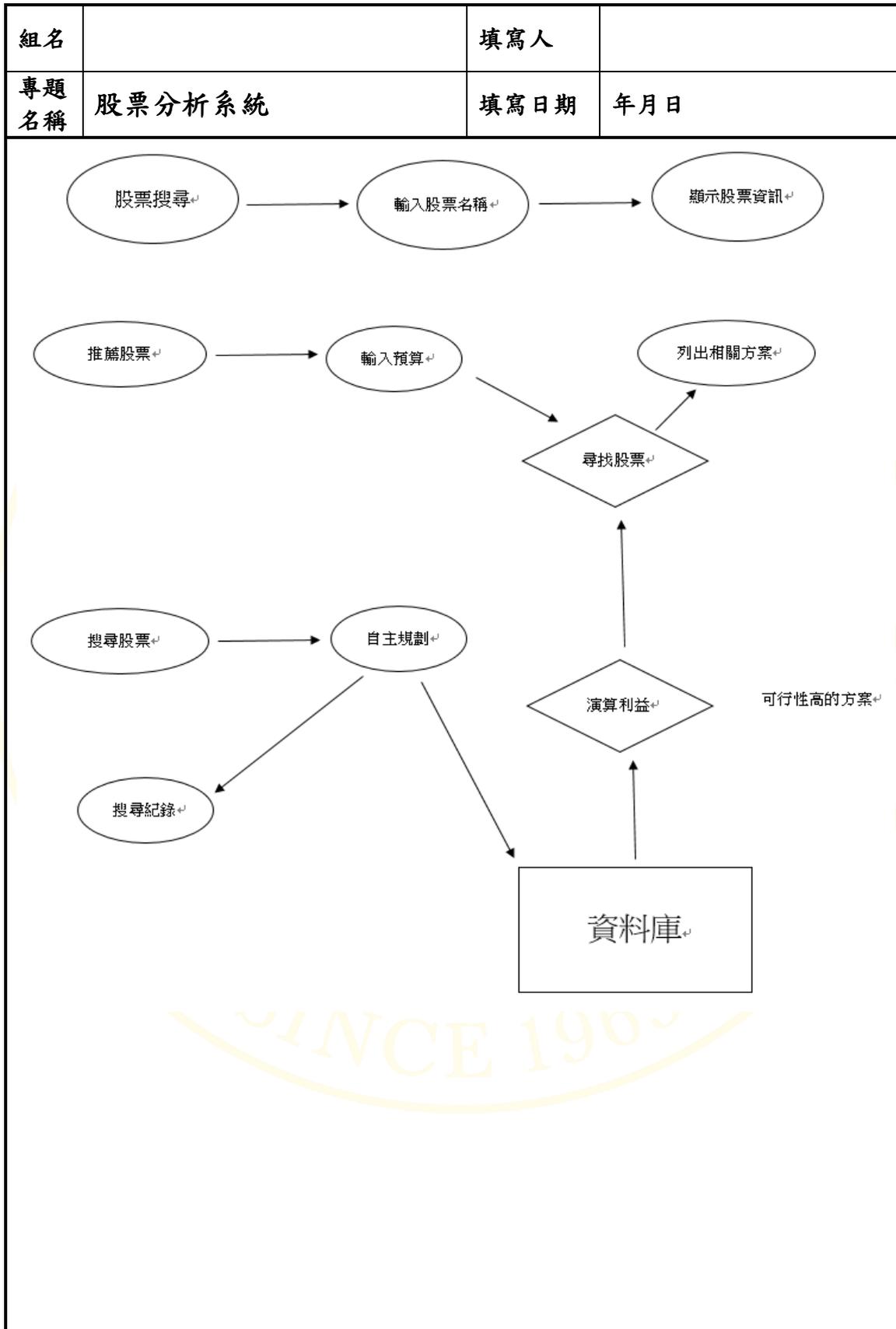
【需求訪談計畫表】

組名	清一色	填寫人	
專題名稱	股票分析系統	填寫日期	年月日
<p>基本資料</p> <p>1.您的性別？ <input type="checkbox"/>男 <input type="checkbox"/>女</p> <p>2.您的年齡？ <input type="checkbox"/>25歲以下 <input type="checkbox"/>26歲~35歲 <input type="checkbox"/>36歲~45歲 <input type="checkbox"/>46歲~55歲 <input type="checkbox"/>56歲以上</p> <p>3.教育程度 <input type="checkbox"/>高中職以下 <input type="checkbox"/>大學/專科 <input type="checkbox"/>研究所以上</p> <p>4.婚姻狀態 <input type="checkbox"/>已婚 <input type="checkbox"/>未婚</p> <p>5.目前職業類別？ <input type="checkbox"/>農林漁牧礦業 <input type="checkbox"/>軍公教 <input type="checkbox"/>工業 <input type="checkbox"/>金融業 <input type="checkbox"/>學生 <input type="checkbox"/>服務業 <input type="checkbox"/>資訊相關產業 <input type="checkbox"/>製造業 <input type="checkbox"/>其他：</p> <p>簡答題</p> <p>1.為什麼想使用我們的股票分析系統？</p> <p>2.有甚麼需要改進的地方？</p> <p>3.其他建議</p>			

【需求訪談紀錄表】

組名	清一色	填寫人	吳芃鎰
專題名稱	股票分析系統	填寫日期	年月日
<p>1.為什麼想使用我們的股票分析系統？          透過美股和台股之間的時差來進行股票投資，能夠更加的準確，且可以匯入歷年股市資料來進行簡易的股市交易模擬，確認我們投資的精準度。</p> <p>2.客戶的需求          現代人忙碌，能用的時間少之又少，能夠有個分析歷年股票漲幅的系統，可以節省很多時間。</p> <p>3.其他          關於介面上的部份，可以設計得更加整齊，使用的時候能更加流暢。</p>			

【使用個案圖】



【藍圖】

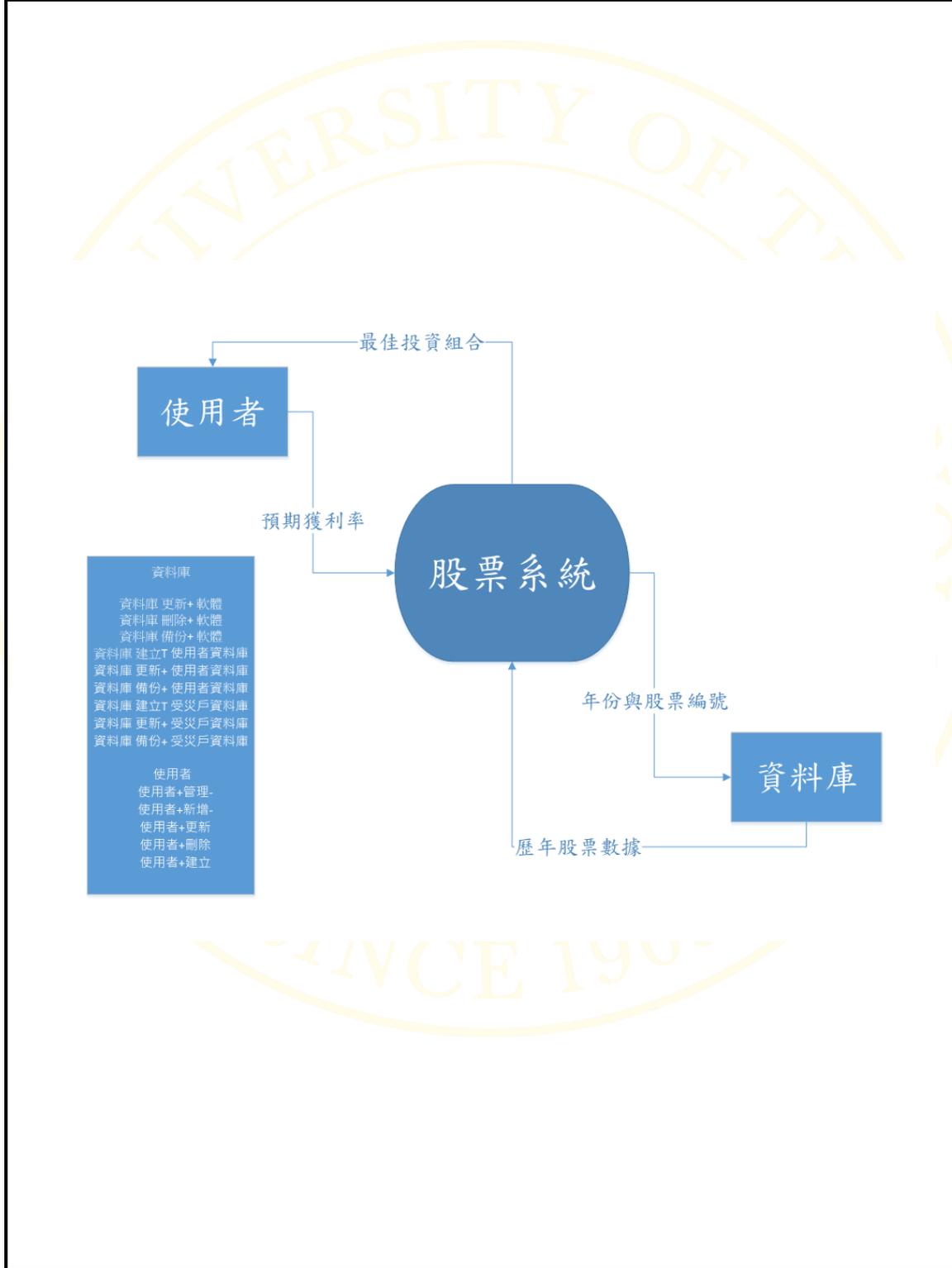
組名		填寫人								
專題名稱	股票分析系統	填寫日期	年月日							
count>100 (規則出現次數)		mean $\geq$ 2.5 (平均值)								
		std $\leq$ 3.8 (標準差)								
股票代碼	日期	成交量	開盤價	最高價	最低價	收盤價	漲跌	前一天漲跌	當天漲跌	後一天漲跌
A0050	20120510	7132826	51.6	52.0	51.55	51.70	0	-3	0	-1.06
A0050	20120511	29045919	51.6	51.6	51.05	51.15	-1.06	-3	-1.06	0.09
A0050	20120514	10200944	51.3	51.3	50.95	51.20	0.097	-3	0.09	0

### 【資料詞彙】

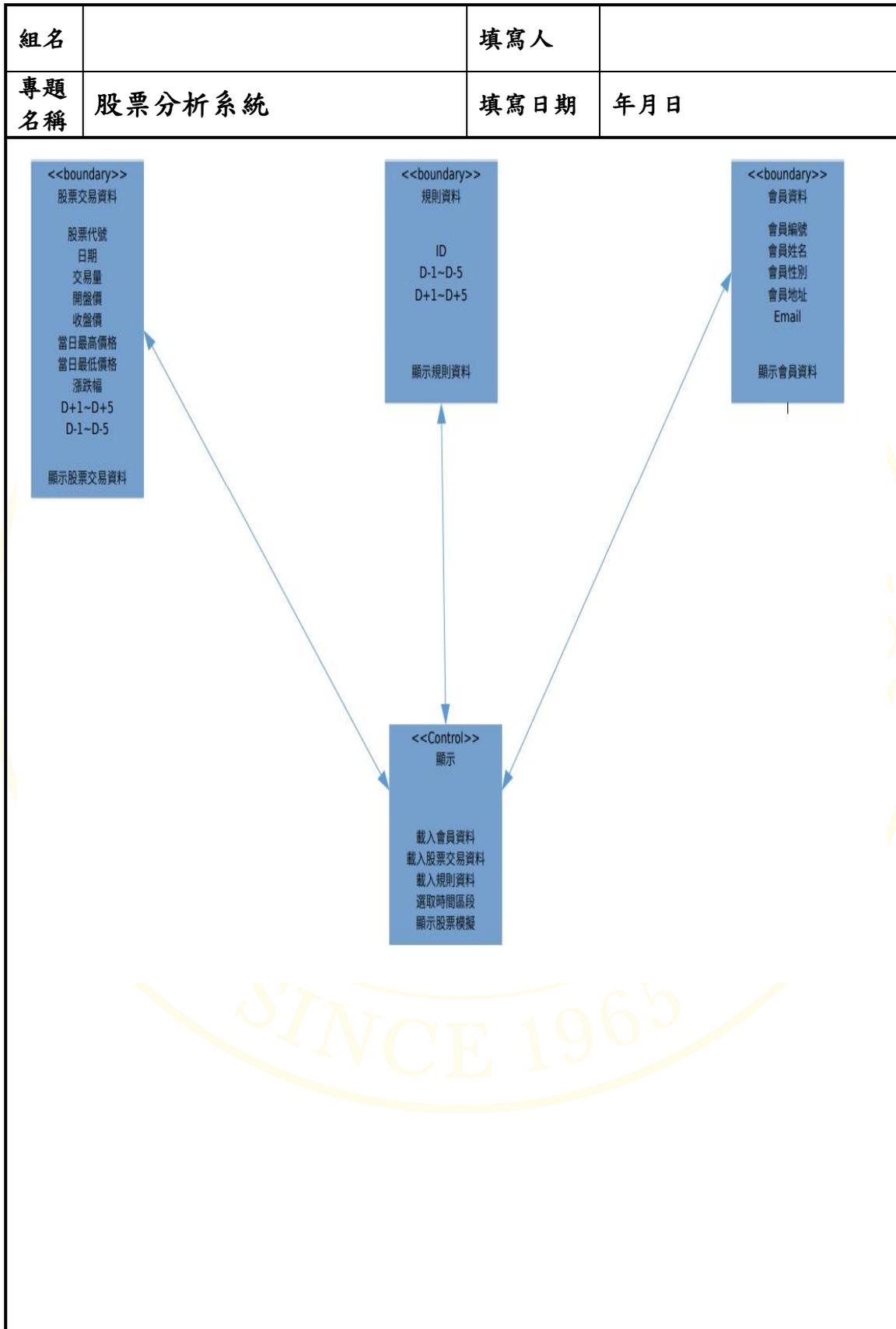
組名	填寫人
專題名稱	填寫日期 年月日
<p>股票分析系統</p>	
<p>Youtube - 影片社群網站</p> <p>Reddit - 社交分享網站</p> <p>Dropbox - 檔案分享服務</p> <p>豆瓣網 - 圖書、唱片、電影等文化產品的資料資料庫網站</p> <p>Django - 鼓勵快速開發的 Web 應用框架</p> <p>Pylons - Web 應用框架</p> <p>Zope - 應用伺服器</p> <p>Plone - 內容管理系統</p> <p>Instagram - 是一款免費提供線上圖片及影片分享的社交應用軟體,使用 Django 作為後台</p> <p>TurboGears - 另一個 Web 應用快速開發框架</p> <p>Twisted - Python 的網路應用程式框架</p> <p>Fabric - 用於管理成百上千台 Linux 主機的程式庫</p> <p>Python Wikipedia Robot Framework - MediaWiki 的機器人程式</p> <p>MoinMoinWiki - Python 寫成的 Wiki 程式</p> <p>Trac - 使用 Python 編寫的 BUG 管理系統</p> <p>Mailman - 使用 Python 編寫的郵寄清單軟體</p> <p>Mezzanine - 基於 Django 編寫的內容管理系統</p> <p>Flask - Python 微 Web 框架</p> <p>Webpy - Python 微 Web 框架</p> <p>Bottle - Python 微 Web 框架</p> <p>EVE - 網路遊戲 EVE 大量使用 Python 進行開發</p> <p>Blender - 使用 Python 作為建模工具與 GUI 語言的開源 3D 繪圖軟體</p> <p>Inkscape - 一個開源的 SVG 向量圖形編輯器。</p> <p>知乎 - 一個問答網站</p> <p>果殼 - 一個泛科技主題網站</p> <p>Odo - 仍在持續發展壯大且最受歡迎的 ERP 軟體</p>	

【活動圖】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日



【類別圖】



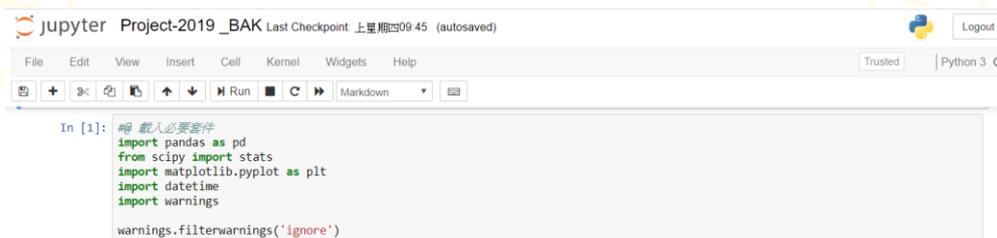
## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號		畫面名稱	
------	--	------	--

系統畫面



The screenshot shows a Jupyter Notebook window titled "Project-2019\_BAK Last Checkpoint. 上星期四09:45 (autosaved)". The code in the cell is as follows:

```
In [1]: #載入必要套件
import pandas as pd
from scipy import stats
import matplotlib.pyplot as plt
import datetime
import warnings
warnings.filterwarnings('ignore')
```

操作說明

點選 RUN 按鈕執行

## 【使用者操作手冊】

<b>組名</b>		<b>填寫人</b>	
<b>專題名稱</b>	股票分析系統	<b>填寫日期</b>	年月日

本頁如不敷使用，請自行複製新頁面。

<b>畫面編號</b>	<b>畫面名稱</b>
-------------	-------------

系統畫面

```

In [7]: # 過濾掉交易量少於五百萬元的交易資料
df1 = df.loc[(df['Volume'] * df['close'] >= 5000000)] # 848855
print(len(df1))

# 刪除 NaN (Not A Number) 資料，這些資料有問題
df2 = df1.dropna() # 刪除所有有 NaN 的 rows
print(len(df2)) # 848844

# 刪除 2018-01-01 以後的資料，換檔規則不能用 2018 年資料，而是用在模擬
df3 = df2.loc[df2['Date'] < "2018-01-01"]

print(len(df3)) # 710452 2
df3.to_csv("D:/ML/filtered20120502_20171231.csv")
    
```

操作說明

- 將螢光處 1 所標記的過濾規則修改為自己所期望的值；  
例如：如果想要過濾掉交易量低於 40000 筆以下的股票，則將數字改為 40000
- 新增並儲存一個空白的 CSV 檔(以下稱為檔案 1)，接著將螢光處 2 的路徑改為其檔案 1 的路徑

## 【使用者操作手冊】

<b>組名</b>		<b>填寫人</b>	
<b>專題名稱</b>	<b>股票分析系統</b>	<b>填寫日期</b>	<b>年月日</b>

本頁如不敷使用，請自行複製新頁面。

<b>畫面編號</b>	<b>畫面名稱</b>
-------------	-------------

系統畫面

```

jupyter Project-2019_BAK Last Checkpoint: 上星期四09:45 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 C
In [10]: # 將法人買賣超轉換為 3, 2, 1, 0, -1, -2, -3 的 Labels
def factorize(x):
    global L1, L2
    if x <= -L2:
        return -3
    elif x <= -L1:
        return -2
    elif x < 0:
        return -1
    elif x == 0:
        return 0
    elif x >= L2:
        return 3
    elif x >= L1:
        return 2
    else:
        return 1

# =====
df = pd.read_csv("D:/ML/filtered20120502_20171231.csv")
# 將法人買賣超比例區分為 7 個 Levels, (3, 2, 1, 0, -1, -2, -3)
# 這裡可以著調整參數
# 注意: 嘗試過 L1 = 15, L2 = 50, 結果 D-1,...D-5 出現 3 的機率非常低, 不合理, 調整 L2 = 40
L1 = 15
L2 = 40

df['D-1'] = df['D-1'].apply(factorize)
df['D-2'] = df['D-2'].apply(factorize)
df['D-3'] = df['D-3'].apply(factorize)
df['D-4'] = df['D-4'].apply(factorize)
df['D-5'] = df['D-5'].apply(factorize)

jupyter Project-2019_BAK Last Checkpoint: 上星期四09:45 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 C
elif x == 0:
    return 0
elif x >= L2:
    return 3
elif x >= L1:
    return 2
else:
    return 1

# =====
df = pd.read_csv("D:/ML/filtered20120502_20171231.csv")
# 將法人買賣超比例區分為 7 個 Levels, (3, 2, 1, 0, -1, -2, -3)
# 這裡可以著調整參數
# 注意: 嘗試過 L1 = 15, L2 = 50, 結果 D-1,...D-5 出現 3 的機率非常低, 不合理, 調整 L2 = 40
L1 = 15
L2 = 40

df['D-1'] = df['D-1'].apply(factorize)
df['D-2'] = df['D-2'].apply(factorize)
df['D-3'] = df['D-3'].apply(factorize)
df['D-4'] = df['D-4'].apply(factorize)
df['D-5'] = df['D-5'].apply(factorize)
df.to_csv("D:/ML/labeled_15_40_20120502_20171231.csv")
    
```

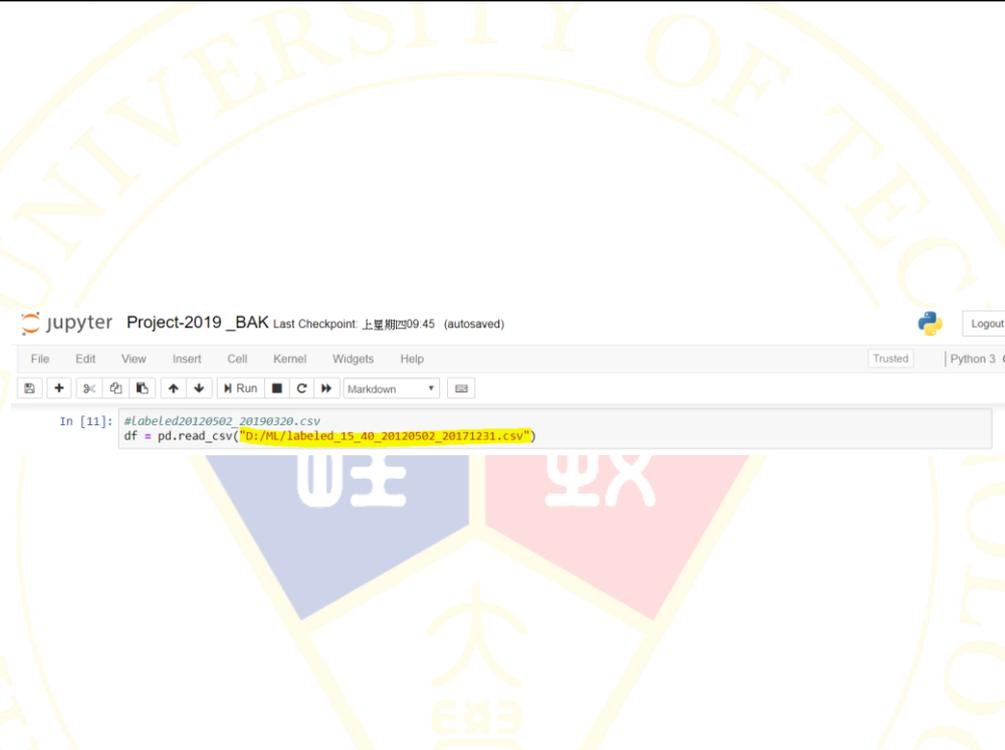
操作說明

- 將螢光處 1 的路徑替換為檔案 1 的路徑
- 新增並儲存一個空白的 CSV 檔(以下稱為檔案 2)，接著將螢光處 2 的路徑改為其檔案 2 的路徑
- 執行 RUN

## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
系統畫面	 <p>The screenshot shows a Jupyter Notebook window titled 'Project-2019_BAK'. The code cell contains the following Python code: <code>#Labeled20120502_20190320.csv</code> and <code>df = pd.read_csv('D:/ML/labeled_15_40_20120502_20171231.csv')</code>. The file path in the second line is highlighted in yellow.</p>
操作說明	<ul style="list-style-type: none"> <li>● 將螢光處的路徑替換為<u>檔案 2</u>的路徑</li> <li>● 執行 RUN</li> </ul>

## 【使用者操作手冊】

<b>組名</b>		<b>填寫人</b>	
<b>專題名稱</b>	股票分析系統	<b>填寫日期</b>	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
------	------

系統畫面

```

In [13]: # 計算 rewards 找出規則
df = pd.read_csv("LabelEd20120502_20190320.csv")

df['y'] = df['D']
df['y+1'] = df['D'] + df['D+1']
df['y+2'] = df['D'] + df['D+1'] + df['D+2']
df['y+3'] = df['D'] + df['D+1'] + df['D+2'] + df['D+3']
df['y+4'] = df['D'] + df['D+1'] + df['D+2'] + df['D+3'] + df['D+4']

df['Y'] = df[['y', 'y+1', 'y+2', 'y+3', 'y+4']].max(axis=1)

# group by 作業後分別計算滾動統計量 describe
gup = df.groupby([df['D-1'], df['D-2'], df['D-3'], df['D-4'], df['D-5']])
df_gp_y = gup['Y'].describe()

# 將滾動統計量依照 mean 和 std 排序，mean 大的排前面，std 小的排前面
df_opt_rules = df_gp_y.sort_values(by=['mean', 'std'], ascending=[0,1])

# 將規則存檔
df_gp_y.to_csv("D:/ML/Model1_Max_Unsorted_20120502_20171231.csv")
df_opt_rules.to_csv("D:/ML/Model1_Max_Sorted_20120502_20171231.csv")
    
```

操作說明

- 新增並儲存一個空白的 CSV 檔(以下稱為檔案 3)，接著將螢光處 1 的路徑改為其檔案 3 的路徑
- 新增並儲存一個空白的 CSV 檔(以下稱為檔案 4)，接著將螢光處 2 的路徑改為其檔案 4 的路徑

## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
------	------

系統畫面



```

jupyter Project-2019_BAK Last Checkpoint: 上星期四09:45 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
In [14]: # 選擇 golden rules, 這種 pattern 至少要出現過 50 次, reward 平均值 >= 2.5
df_opt_rules
df_o = df_opt_rules.loc[(df_opt_rules["count"]>100.0) & (df_opt_rules["mean"]>=2.5)&(df_opt_rules["std"]<=3.8)]
print(len(df_o))
# 秀出來看看
df_o.iloc[0:30]
df_o.to_csv("D:/ML/GoldenRules.csv")
    
```

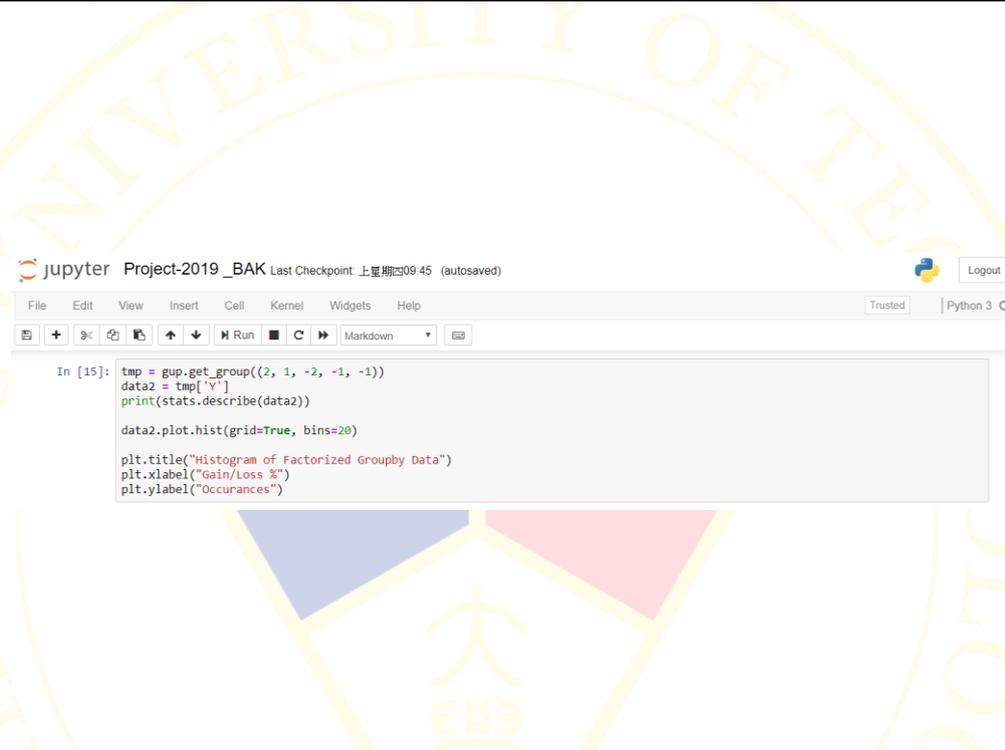
操作說明

- 新增並儲存一個空白的 CSV 檔(以下稱為檔案 5)，接著將螢光處的路徑改為其檔案 5 的路徑

## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
系統畫面	 <p>The screenshot shows a Jupyter Notebook titled 'Project-2019_BAK'. The code in the cell is as follows:</p> <pre>In [15]: tmp = gup.get_group((2, 1, -2, -1, -1)) data2 = tmp['Y'] print(stats.describe(data2))  data2.plot.hist(grid=True, bins=20)  plt.title("Histogram of Factorized Groupby Data") plt.xlabel("Gain/Loss %") plt.ylabel("Occurances")</pre>
操作說明	<ul style="list-style-type: none"> <li>● 執行 RUN</li> </ul>

## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
系統畫面	 <p>The screenshot shows a Jupyter Notebook window titled 'Project-2019_BAK'. The code in the cell is as follows:</p> <pre>In [16]: tmp = gup.get_group((3, 3, 3, 2, 1)) data2 = tmp['Y'] print(stats.describe(data2)) data2.plot.hist(grid=True, bins=20)  plt.title("Histogram of Factorized Groupby Data") plt.xlabel("Gain/Loss %") plt.ylabel("occurances")</pre>
操作說明	<ul style="list-style-type: none"> <li>● 執行 RUN</li> </ul>

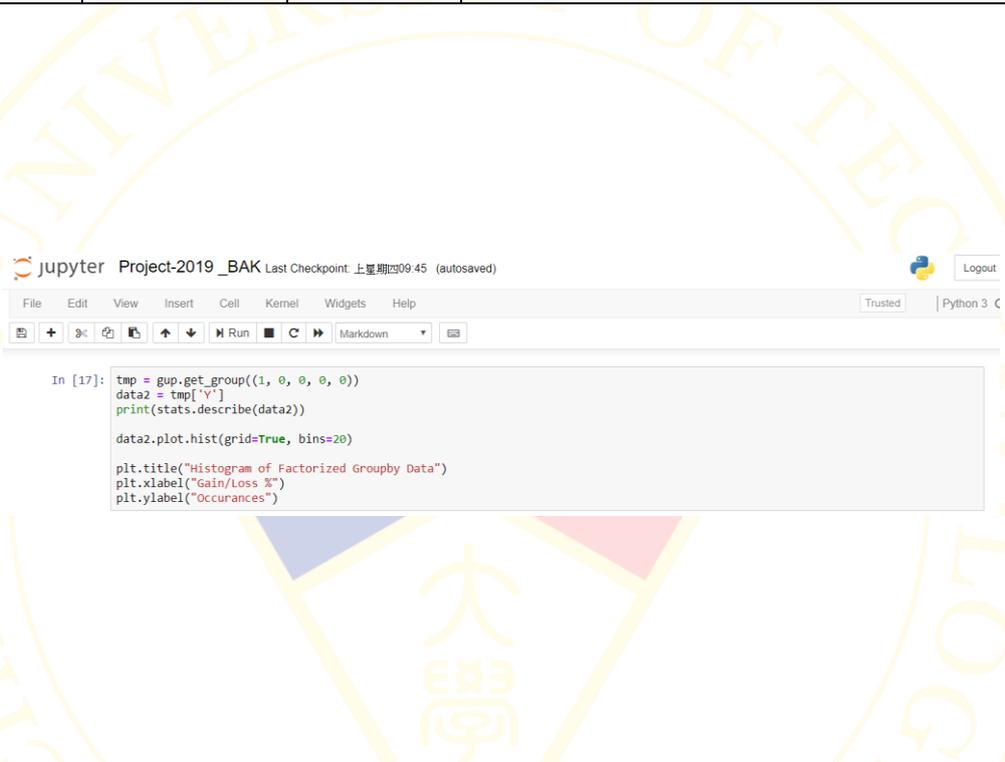
## 【使用者操作手冊】

組名		填寫人	
專題名稱	股票分析系統	填寫日期	年月日

本頁如不敷使用，請自行複製新頁面。

畫面編號	畫面名稱
------	------

系統畫面



操作說明

- 執行 RUN

## 【使用者操作手冊】

<b>組名</b>		<b>填寫人</b>	
<b>專題名稱</b>	<b>股票分析系統</b>	<b>填寫日期</b>	<b>年月日</b>

本頁如不敷使用，請自行複製新頁面。

<b>畫面編號</b>	<b>畫面名稱</b>
-------------	-------------

系統畫面

```

In [18]: satisfy = 1.01
stopLoss = 0.97

dealerRate = 0.000513 # 手續費折扣，買賣都要付
tax = 0.003 # 證交稅，只有賣出才付

df = pd.read_csv("labeled20120902_20190320.csv")

#df=df.iloc[6:len(df)-7] # 去掉 2017 和 2019 的資料
df = df.loc[(df['Date']>='2019-01-01')&(df['Date']<='2019-6-30')]

# Load all necessary data
#df=pd.read_csv("c:/PythonData/Model1_Factorized_WithD.csv") # 交易資料，包含分組標籤
d=df['Date'].drop_duplicates() # trasaction dates
#d=d.iloc[6:len(d)-7] # 去除 2017 和 2019 的多餘資料
s=df['ID'].drop_duplicates() # stocks
gr=pd.read_csv("GoldenRules.csv") # rules Model1_GoodRules.csv

result=pd.DataFrame(columns=["ID","Date","GRule#","買價","是否買到","滿足價","停損價","BRule#","賣價","是否賣掉","賣出日期","費用"])

for x in d:
    # 'ID','Date','Volume','Open','High','Low','Close','InvestTrade','D-1_Close','D-1','D-2','D-3','D-4','D-5'
    t=df.loc[(df['Date']==str(x))]
    print("Date:", x)

    for y in range(0,len(t)):
        xData = t.iloc[y,:]

        for z in range(0,len(gr)):
            rData = gr.iloc[z,:]
            if ((rData["D-1"]==xData["D-1"]) and (rData["D-2"]==xData["D-2"]) and (rData["D-3"]==xData["D-3"]) and (rData["D-4"]
            rule = "/" + str(rData["D-1"]) + "." + str(rData["D-2"]) + "." + str(rData["D-3"]) + "." + str(rData["D-4"]) +

```

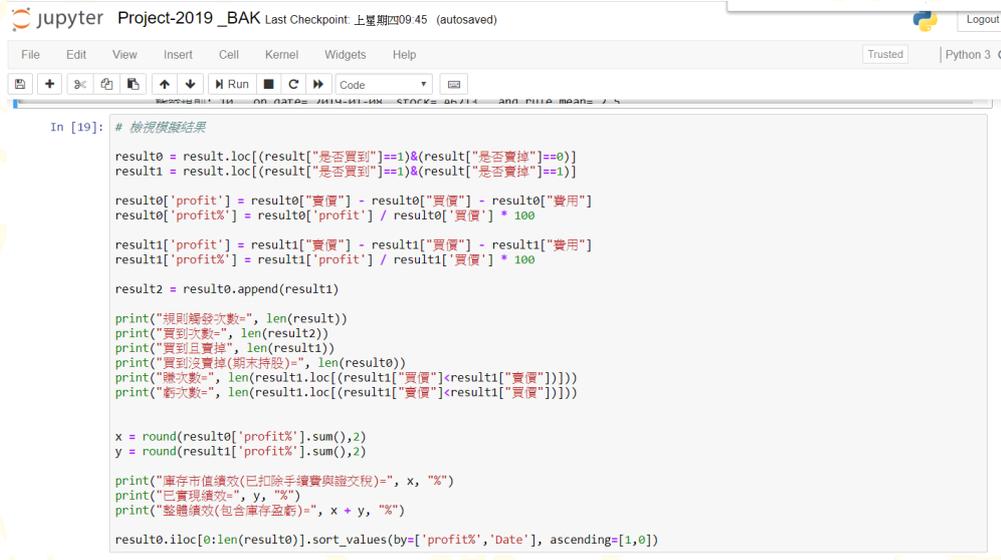
操作說明

- 找出本機測試用的 SOURCE 檔案路徑，並將螢光處 1 的路徑替換為測試用 SOURCE 檔路徑
  - 將螢光處 2 的路徑替換為檔案 5 的路徑
  - 執行 RUN

## 【使用者操作手冊】

<b>組名</b>		<b>填寫人</b>	
<b>專題名稱</b>	<b>股票分析系統</b>	<b>填寫日期</b>	<b>年月日</b>

本頁如不敷使用，請自行複製新頁面。

<b>畫面編號</b>	<b>畫面名稱</b>
系統畫面	
操作說明	<ul style="list-style-type: none"> <li>● 執行 RUN 檢視模擬結果</li> </ul>

**【會議記錄】**

專題名稱	股票分析系統					
會議編號	001	召集人兼主席	林廉定	紀錄者	趙汝佳	
討論主題	網站整體設計			會議時間	108/05/10	
				會議地點	501 電腦教室	
上次會議	決議事項		執行狀況			
	第一次會議		第一次會議			
本次會議	本週工作進度		本週工作內容		負責人員	
	1.討論網站內容大綱概要 2.討論網站外觀草圖		架設基本網站框架		林廉定 邵立嚴	
本次會議內容	每位人員提出對於網站草稿設計的想法 負責人員提出執行上的困難，成員共同討論					
決議事項（與主席裁示）						
無						
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名	
下次會議	召集人	林廉定	紀錄者	趙汝佳	時間	108/05/10
					地點	501 電腦教室
預定討論主題	依照專案進度擬定					
指導老師意見	無					

**【會議記錄】**

專題名稱	股票分析系統					
會議編號	002	召集人兼主席	林廉定	紀錄者	趙汝佳	
討論主題	網站功能與其他事項討論			會議時間	108/05/17	
				會議地點	501 電腦教室	
上次會議	決議事項			執行狀況		
	1.討論網站內容大綱概要 2.討論網站外觀草圖			進度約七成		
本次會議	本週工作進度		本週工作內容		負責人員	
	1.網站功能概論 2.網站外觀最後定案 3.開發工具清單討論 4.訪談和問卷內容討論		1.網站功能及外觀 2.工具清單擬定 3.問卷製作		所有人員	
本次會議內容	主要討論工作內容及全組分工問題					
決議事項 (與主席裁示)						
由組長分配全組工作						
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名	
下次會議	召集人	林廉定	紀錄者	趙汝佳	時間	108/05/17
					地點	501 電腦教室
預定討論主題	依照專案進度擬定					
指導老師意見	無					

### 【會議記錄】

專題名稱	股票分析系統					
會議編號	003	召集人兼主席	林廉定	紀錄者	趙汝佳	
討論主題	網站測試及細節探討			會議時間	108/05/24	
				會議地點	501 電腦教室	
上次會議	決議事項		執行狀況			
	1. 網站功能概論 2. 網站外觀最後定案 3. 開發工具清單討論 4. 訪談和問卷內容討論		已完成			
本次會議	本週工作進度		本週工作內容		負責人員	
	1. 測試網站及 Bug 回報 2. 專案內容細部討論		1. 使用個案圖繪製 2. 活動圖繪製 3. 使用者操作手冊 4. 測試計劃書撰寫		所有人員	
本次會議內容	網站測試沒有問題，專案內容進度仍須努力。					
決議事項 (與主席裁示)						
無						
請簽名	請簽名	請簽名	請簽名	請簽名	請簽名	
下次會議	召集人	林廉定	紀錄者	趙汝佳	時間	108/05/24
					地點	501 電腦教室
預定討論主題	暫無					
指導老師意見	無					

## 符號（公式）說明

Symbol	Meaning
$\Theta$	Debye's constant or characteristic temperature
$\Omega$	efficiency; number of molecules
$\Psi$	availability of a closed system
$\Delta$	internal energy (change) of reaction
$\Phi$	availability of a closed system
$I$	specific irreversibility
$\Lambda$	critical state
$M$	Joule-Thomson coefficient
$N$	stoichiometric coefficient (number of moles in chemical equation)
$\varepsilon$	cutoff ratio

