亞洲大學

資訊工程學系

資訊挖掘期中報告

盤後分析(股票)

資訊工程4A-資電組

組員：黃聖捷(104021034)

邱昱霖(104021035)

林清益(104021016)

林俊仕(104021018)

授課老師：王經篤教授

中華民國 108 年 5 月 23 日

# 完成項目

綠色:完成 紅色:未完成

* 下載資料?
* 資料讀取?
* 資料整理?
* 資料觀察與分析?
* 資料預測?
* Python(Numpy?Pandas?Matplotlib?Scipy?)
* Bonus

Youtube影片連結:[影片在這](https://www.youtube.com/watch?v=G7noTrL2CbQ)

動機

　　為統計、分析股市走向，也盡可能做出預測，所以統計與分析過去股市漲跌，討論漲跌背後原因之可能性。

執行過程與結果

**下載資料?**

在抓取資料的時候主要使用兩個套件:

* Twstock
* Pandas-DataReader

Twstock是開源的套件，在[github](https://github.com/mlouielu/twstock)上面找的到，它的資料來源主要來自於『證券交易所(TWSE)』與『證券櫃檯買賣中心(TPEX)』，這個套件抓取非常的方便好用，只不過TWSE有request limit,每五秒鐘3個request，超過的話就會被ban，需要注意一下。

Pandas-DataReader也是方便好用的套件，不同的是我們抓取資料的時候來源是『yahoo』。



上圖是pandas-datareader抓取資料後可以依照參數名稱取得的相關資料

Ex:df\_2330[‘Open’]可取得開盤價格

**資料讀取?**

資料讀取面向主要是讀取csv檔案，所以使用的套件就是:

* Csv
* Numpy
* pandas

通過讀取csv檔取得資料，並轉換成我們要的資料格式並放入到panda.DataFrame裡面，方便在畫圖時使用。

**資料整理?**

資料整理不再於對csv檔案或是用套件抓取資料的處理，通常都是已經經過整理或是進行讀取之後再把格式轉換成我們需要的即可，那資料整理是在整理什麼呢?我們這組的題目是『股

票盤後分析，所以大多的資料整理都會在對於股票、股市、或是分析圖表上面的整理，先來說說對於此次我們觀察的目標—電腦四大品牌:

* 技嘉
* 宏碁
* 微星
* 華碩

對於這些品牌我們可能都很熟悉，所以找了這四大品牌的股票做觀察與分析，所以就製作了他們的圖表去做觀察。

再來就是對於股票漲跌現象的新聞、時事整理，或是一些現在的國際趨勢(中美貿易戰等等)與各公司的消息，並探討它們造成的影響。

**資料觀察與分析?**

對於我們觀察分析目標之四大品牌，我們找了他們近半年(2018年12月至今)的股市漲跌資訊，並製作成K線圖，先附上我們其中一張。



圖表中有蠟燭(K棒)、均線、KD值與交易量，我們就透過觀察這些圖表來進行討論與分析。除了製程圖表之外，我們總嘗試著做了下面這些事情:

* 將2018年12月開始至今的資料製成K線圖
* 觀察趨勢走向，嘗試理解當時的背景與時事
* 解釋趨勢漲跌因素

**→將2018年12月開始至今的資料製成K線圖**

利用pandas-datareader套件抓取股票資訊再製成K線圖，主要抓取的四大家2018年12月至今的資料，先附上K線圖表，之後會解釋與教學如何看這張圖表與該看什麼

2377-微星

2376-技嘉

2357-華碩(ASUS)



2353-宏碁(ACER)

**→如何看圖表**

先看最上面這張(蠟燭):

一根一根紅色綠色的我們稱之為蠟燭，由開盤價、收盤價、最高價、最低價來做計算，先說說顏色差別，蠟燭會有分兩種顏色，分別為紅色與綠色，紅色的代表意義是今日的收盤價>開盤價，也就是今天整個幅度是漲的，綠色反之收盤價<開盤價，我們在分細項一點的話，可以分成兩個方面看，可以發現蠟燭有分**粗**與**細**的部分，粗的部分可以看開盤價與收盤價，細的部分可以看最高價與最低價。

紅色蠟燭

最高價



最低價

開盤價

收盤價

綠色蠟燭



收盤價

開盤價

打個例子來看，我們可以看看華碩的最後一根蠟燭是綠色的，代就收盤價低於開盤價，整個走向是往下跌的，而細細的那跟就可以看當天的最高價與最低價(綠色蠟燭粗體下方邊是當日的收盤價、粗體上方式開盤價)

※小技巧:當蠟燭長度很長，就代表當日的波動很大(最高與最低差距大)，可能是背後有什麼故事喔!

→均線

在蠟燭下方你會看到有兩條線，一條藍色的一條橘色的，這兩條線我們稱之為『均線』，均線又分為月均線、季均線等等(依照計算天數分別)，圖片中藍色的月均線

→KD值

→觀察現象與解釋

討論與問題

附上討論時的心智圖

主要討論了有幾項:

* 股票公司選擇
* 如何觀看K線圖
* 觀察分析K線圖跟背景時事

心得

昱霖:這次的報告很有趣的，由觀察數據來推斷實際公車的使用情況，分析人們思路與影響著他們的事情，再去判斷人們所下的決定，而人們所下的決定與選擇就會影響著數據，影響交通情況，很享受這個過程，可以更了解人的思維，了解了人的思維就可以去做事情的應對，改善現況。但是由於路段資料少，搭乘人數也少，無法提供更有針對性與改善的建議。

俊仕: Data mining，將要探討的資料進行蒐集，再進行整理，最後在對資料進行分析與發現其中的問題。整個過程真的可以發現一些我們平時看不到的現象。不過這其中最辛苦困難的部分還是在資料的蒐集，必須要蒐集到非常大量的資料量才能看的其中的差異，所以這是得花上極大量的時間與精力的。資料挖掘這門課上到期中的時間，對我來說真的還是挺有趣的。

聖捷: 在製作報告的過程中，收集並且整理資料所需要的時間並沒有一開始想像的那麼的快速，而資料也需要從多方面去考慮會造成的原因，像是各個時段、票卷、月份等等才有辦法去推敲出一些訊息。當瞭解到有許多方面的原因之後，可能還有更延伸的事物會影響最後的結果，像是票卷更新等等，隨著資料量越來越多才覺得資料挖掘不是一件簡單的事情。

清益: 我認為做分析報告與資料搜集還有猜測分析時，其實資料的正確性是第一重要，雖然擁有數據中心的資料，可是觀看路線後發現的其實跟我們自己切身的感覺是不太相同的。比如說覺得應該要很多人搭乘的時間班次相對的看不出來差異，這時候我對正確性就抱有疑問了，所以我認為在做這種分析的時候對於資料的正確與分析是佔有一大部分重要的程度的。

工作分配表

|  |  |
| --- | --- |
| 姓名 | 貢獻 |
| 黃聖捷 | 資料整理與集合、coding |
| 林俊仕 | 資料整理與集合、PPT製作 |
| 林清益 | 影片製作、想關資料查詳 |
| 邱昱霖 | Word檔案製作、coding |

每個人都有參與討論，都有提供意見與想法，並對問題提出可能的解決方案並執行(查找相關資料等等)，再由清益提供更多更詳細的資料後供給PPT製作人員。