

HackMD Link (<https://hackmd.io/@libao3128/r1nHMrUjO>).

# Database Final Project Report: StockCare

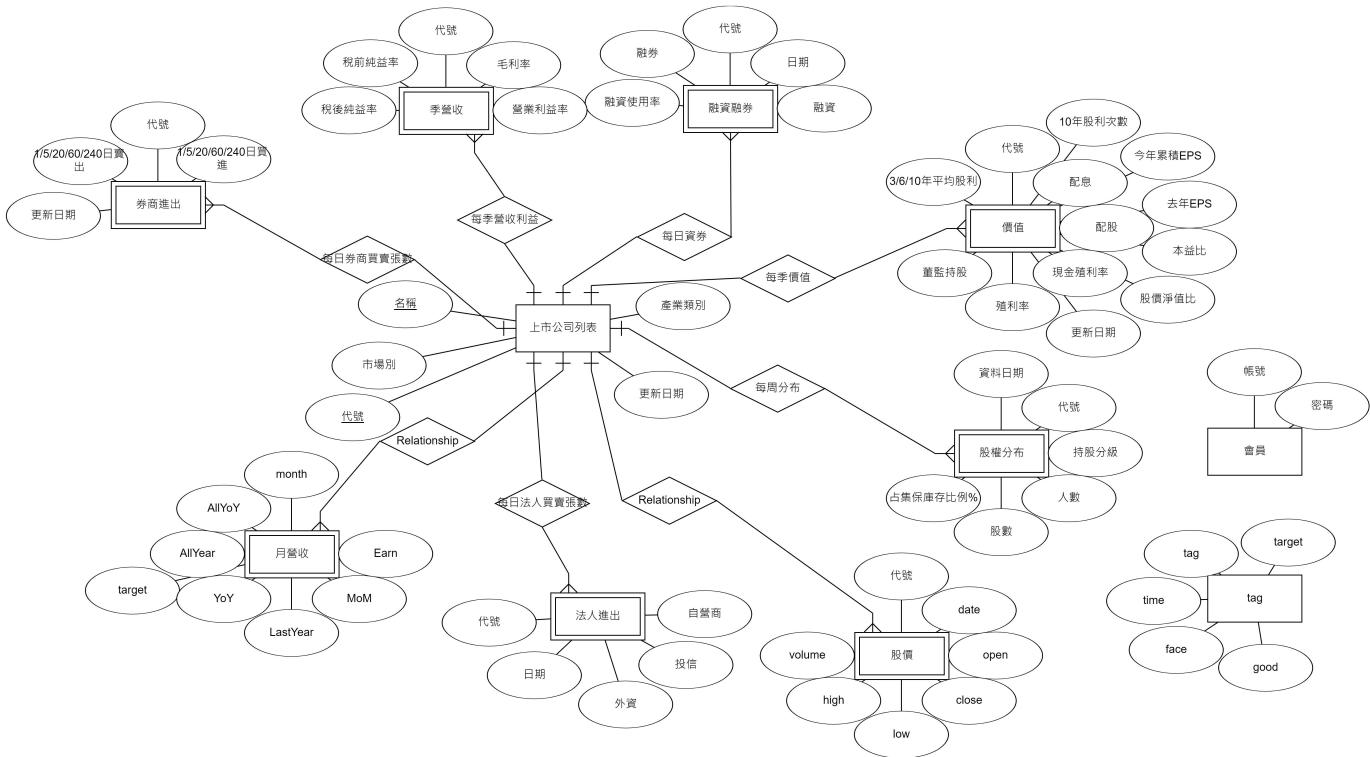
Team 6

Team Member:

0816033 簡予婕、0816072 林子鈞、0816086 黃立鈞、0816026 許傳聲

## Data

### ER model



### Table

上市公司列表

Column	type	Other
代號	INT	PK,NN,UQ
名稱	VARCHAR(8)	UQ
市場別	TEXT	
產業類別	TEXT	
更新日期	DATE	

更新頻率:每半年

來源:台灣證券交易所 (<https://www.twse.com.tw/zh/page/products/stock-code1.html>).

取得方式:python爬蟲

## 法人進出

Column	type	Other
代號	INT	PK,NN
日期	DATE	PK,NN
外資	INT	
投信	INT	
自營商	INT	

更新頻率:每天

來源:康和證券法人持股 ([https://concorde.moneydj.com/z/zc/zcl/zcl\\_1101.djhtm](https://concorde.moneydj.com/z/zc/zcl/zcl_1101.djhtm)).

取得方式:python爬蟲

## 股價

Column	type	Other
代號	INT	PK,NN
date	DATE	PK,NN
open	DOUBLE	
close	DOUBLE	
low	DOUBLE	
high	DOUBLE	
volume	INT	

更新頻率:每天

來源:[yahoo finance \(https://finance.yahoo.com\)](https://finance.yahoo.com).

取得方式:python爬蟲

## 股權分布

Column	type	Other
資料日期	DATE	
代號	INT	
持股分級	INT	
人數	INT	
股數	INT	
占集保庫存數比例%	DOUBLE	

更新頻率:每周

來源:[政府資料開放平臺 \(https://data.gov.tw/dataset/11452\)](https://data.gov.tw/dataset/11452).

取得方式:手動下載CSV,並用python處理格式後匯入

## 價值

Column	type	Other
代號	DATE	
配息	DOUBLE	
配股	INT	
現金殖利率	TEXT	
殖利率	TEXT	
董監持股	TEXT	
3年平均股利	DOUBLE	
6年平均股利	DOUBLE	
10年平均股利	DOUBLE	
10年股利次數	INT	
今年累積EPS	DOUBLE	
去年EPS	DOUBLE	
本益比	DOUBLE	
股價淨值比	DOUBLE	
更新日期	TEXT	

更新頻率:每季

來源:撿股讚 (<https://stock.wespai.com/rate110>).

取得方式:利用google插件table\_capture擷取表格，再手動匯入sql

融資融券

Column	type	Other
代號	INT	PK,NN
日期	DATE	PK,NN
融資	INT	
融券	INT	
融資使用率	DOUBLE	

更新頻率:每天

來源:康和證券融資融券 ([https://concorde.moneydj.com/z/zc/zcn/zcn\\_1101.djhtm](https://concorde.moneydj.com/z/zc/zcn/zcn_1101.djhtm)).

取得方式:Python爬蟲

季營收

Column	type	Other
代號	INT	NN
毛利率	DOUBLE	
營業利益率	DOUBLE	
稅前純益率	DOUBLE	
稅後純益率	DOUBLE	
季度	TEXT	

更新頻率:每季

來源:公開資訊觀測站 (<https://mops.twse.com.tw/mops/web/index>).

取得方式:Python爬蟲

券商進出

Column	type	Other
代號	INT	PK,NN
1日買進	INT	
1日賣出	INT	
5日買進	INT	
5日賣出	INT	
20日買進	INT	
20日賣出	INT	
60日買進	INT	
60日賣出	INT	
240日買進	INT	
240日賣出	INT	
更新日期	DATE	PK,NN

更新頻率:每日

來源:康和證券主力進出 ([https://concord.moneydj.com/z/zc/zco/zco\\_1101.djhtm](https://concord.moneydj.com/z/zc/zco/zco_1101.djhtm)).

取得方式:Python爬蟲

## 月營收

Column	type	Other
month	DATE	NN,PK
target	INT	NN,PK
Earn	INT	
MoM	TEXT	
LastYear	INT	
YoY	TEXT	
AllYear	INT	
AllYoY	TEXT	

更新頻率:每月

來源:康和證券營收盈餘 ([https://concord.moneydj.com/z/zc/zce/zce\\_1101.djhtm](https://concord.moneydj.com/z/zc/zce/zce_1101.djhtm)).

取得方式:Python爬蟲

## Tag

Column	type	Other
target	INT	
tag	VARCHAR(20)	
time	INT	1:長線,0:短線
good	INT	1:多,0:空
face	INT	1:籌碼面,2:技術面,3:基本面

更新頻率:每日

這個table利用我們寫好的query檔紀錄了每天個股的所有指標tag，並用good與face來記錄對應的指標應該呈現在哪個面向(face)，長短線(time),多空(good)的格子中。

## 會員

Column	type	Other
帳號	TEXT	
密碼	TEXT	

更新頻率:無

這個table紀錄了我們曾經註冊過的會員資料，以方便日後增加新功能使用。

爬蟲用的Python code請見github中相對應table檔名的.py檔

## Database

我們使用mySQL作為我們的Database，並使用mySQL workbench作為管理工具，所有有關股票的資料都是利用python中的beautiful soup套件與request指令從網路上爬蟲取得或是手動下載之後再匯入Database。

而會員資料則是由客戶端創立帳號與密碼，並可以隨時更改密碼。

我們用Node.js & Express.js 建立Web API，用require express和bodyParser取得user input，啟動web service，連線資料庫，傳遞sql語法取得JSON檔，解析JSON並將結果更改至HTML。

## Application

我們的網頁透過蒐集各種有關上市公司的資訊，並用簡易明瞭的表格分別從籌碼面、基本面、技術面呈現各上市公司的多空狀態，讓使用者可以藉此輔助找到狀態優良的個股與掌握自己持股的狀態與健康。

在首頁的部分，我們可以直接輸入上市公司的股票代碼，藉此連結到該公司的健檢頁面，如果在這裡輸入了非上市公司的股票代碼則會跳出錯誤的訊息，同時在搜尋列中也加入無法輸入非數字以外字元的防呆設計。

在首頁的右上角有會員的註冊與登入，按下註冊後就會跳到會員註冊的頁面，只要輸入電子郵件與密碼就可以建立屬於自己的帳號，接著在點登入頁面就可以用註冊過的帳號密碼登入網頁，未來預計會新增功能，讓使用者可以調整自己的自選股票清單，方便日後檢查自己清單中的個股。

進入個股健檢的網站後，我們設計了三大面向:籌碼、技術、基本面的表格，並用多空與長線短線作為分類依據，直接明瞭的呈現該公司目前的狀態，透過點擊右上角的小標我們可以快速地跳到自己在意的面向，如此一來便可以在很短的時間蒐集多間公司每天的狀態，加速投資者每天做功課的時間。

我們建立的指標分為三大面向:

### 籌碼面

在籌碼分析中，我們蒐集每天個股的盤後籌碼資訊例如法人買賣超、券商分點、融資融券的買賣的增減，並以此排序、篩選找出優勢或弱勢的个股，除此之外，我們也統計長時間內的籌碼變化，例如股權分布或是籌碼集中度，以此掌握長時間的多空狀態，以下是我們建立的指標。

**法人包含:外資、投信、自營商**

**籌碼集中度:( 區間買超前15名的買超張數合計 - 區間賣超前15名的賣超張數合計 ) / 區間成交量**

面向	指標	篩選條件
法人	法人連續買/賣天數	統計每天法人的進出， 並篩選出法人連續3/5/10日買超或賣超的个股
法人	法人由賣轉賣或由買轉賣	篩選出過去幾天法人連續買/賣超數日， 並在當日做的相反的操作代表法人動向的改變
法人	近一周/月/季內法人買/ 賣超前100/50/10名	將一段時間內個公司的法人進出加總後 排序並找出排名前100/50/10的公司
法人	土洋合作買/賣超	當日投信、自營商、外資同時買超、賣超的个股， 代表法人對公司的看法相同
融資融券	近一周/ 月融資融券劇烈變化	當融資融券的變化張數 >500x天數時我們稱為劇烈變化， 可以用來觀察散戶對个股的看法
融資融券	融資使用率過高	當一個股融資使用率過高代表多數散戶看好該公司， 容易引來主力的收割
主力動向	近一周/月/季/ 年籌碼集中或分散	透過統整每日券商分點的張數與累積成交量 來了解主力大戶的動向。當籌碼集中度>10% 稱為集中<10%則稱為分散
股權分布	大股東持有比>70%	透過證交所每週公布的股權分布表， 找出持有千張以上的大戶持有的比例大於70%的个股
股權分布	大股東持股比增加/減少	透過前後一個星期大股東持股比例的變化大於1% 找出大股東短期之內持股的調整， 影響短線多空的發展

## 技術面

在技術面中，我們透過分析股價、均線的相對關係來篩選出強勢與弱勢的技術型態，以此來判斷目前个股的多空狀況。

以下是我們建立的指標:

**上關價:昨低 + (昨高-昨低) x 1.382**



下關價: 昨高 - (昨高 - 昨低) x 1.382

**n日均線:** 過去n天的收盤價的平均，我們指標採用的均線組合為投資人常用的**5日/10日/20日/60日**又名**周線/雙周線/月線/季線**

面向	指標	條件
股價	創新高/新低	透過統計過去一段時間的股價， 找出當日股價突破或跌破過去一周/雙周/一月/一季/ 一年 新高或新低的個股
股價	突破上/下關價	上下關價是投資人常參考的短線指標， 用來判斷當天股價相對於昨天的強弱。
均線、 股價	股價突破或跌破均線	當昨日的股價低於當日均線而隔日高於同期均線時， 我們稱之為突破均線， 反之則稱為跌破
均線、 股價	股價突破或跌破糾結均線	當所有不同期的均價範圍界在2%， 我們稱為均線糾結， 代表該股價在一段時間內沒有大幅變動， 而當今日股價大於糾結均線區3%以上稱為突破， 代表股價脫離了長時間的盤整區， 反之則為跌破
均線	均線多/空頭排列	當短期均線價格大於長期均線價格， 即: 周均價 > 雙周均價 > 月均價 > 季均價 稱為多頭排列， 反之稱為空頭排列
均線	均線多頭/空頭	計算個股今日與昨日同期均線的漲幅或跌幅並排序， 找出 漲幅最高的50名與跌幅最高的50名稱為均線多頭與空頭

例外處理: 當一間公司上市天數不足均線需要的天數時，則將該公司排除在所有需要均線做為參考的指標之外，因為這時均線的型態不足以做為參考依據。例如: A公司只上市15天，則他的20日均價與60日均價相同，則無法以此作為突破或跌破月均線獲季均線的依據，同時也無法透過均線的排列狀況判斷多空，因此在計算均線時須排除任何上市天數不足的公司。

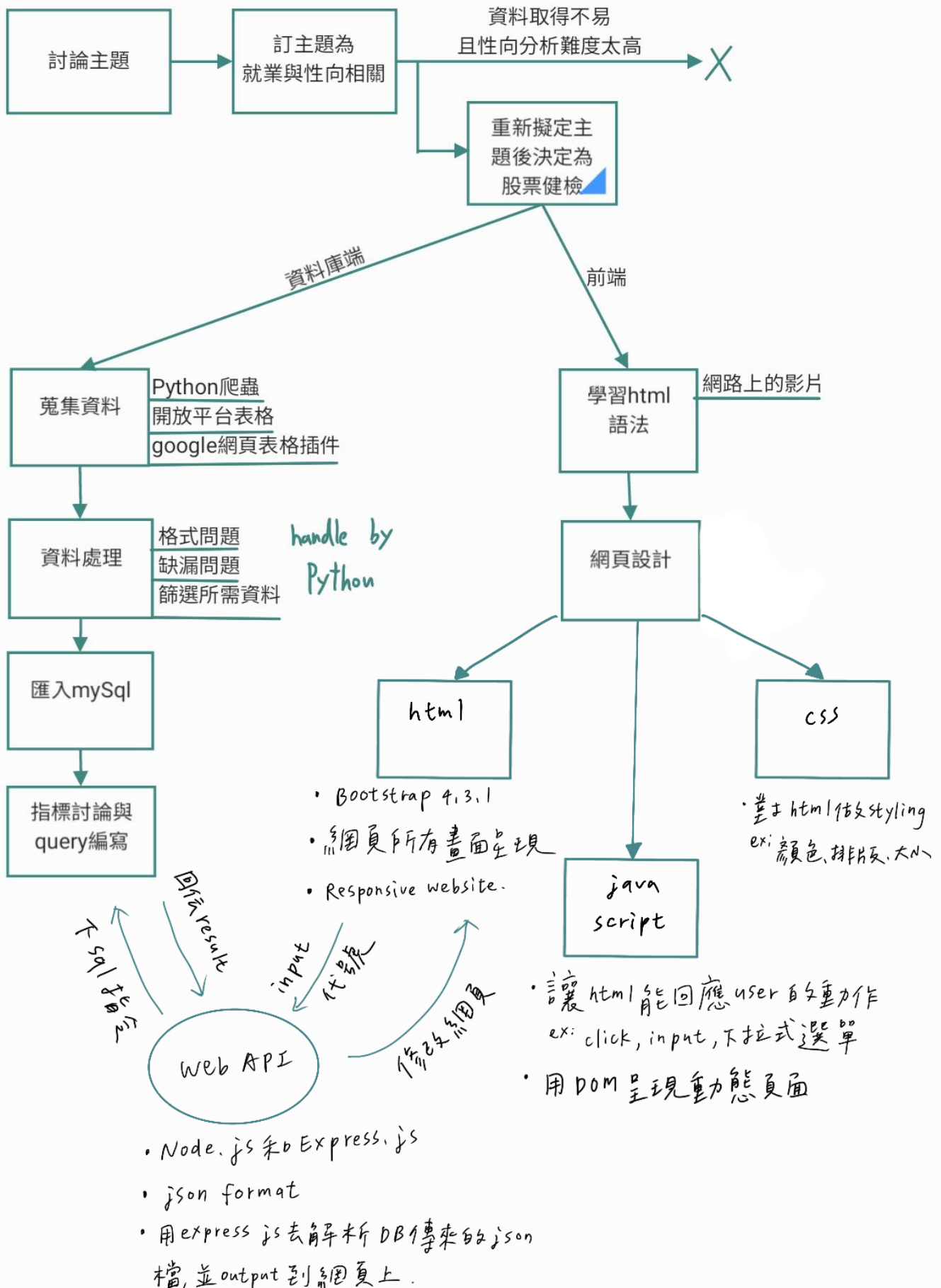
## 基本面

面向	指標	條件
獲利	近12月營收同業前>20%	統計過去一年個公司的營業收入加總，並依照公司產業類別排名取出前百分之20，作為各個產業的龍頭股
獲利	近4/8/12季毛利率>50%	統計過去3年內各公司的平均毛利率，毛利率高代表該公司盈利能力強
獲利	近三年平均毛利率同業前20%	統計過去3年內各公司的平均毛利率，並以產業類別排名，以消冗不同產業之間毛利率之間的先天差異，方便比
獲利	近3年營業利率在同業前10%	統計過去3年各公司的營業利益率，並以產業類別排名，以篩選不同產業之間營業利益率的佼佼者
獲利	月營收月成長(MoM)>20%	統計各公司上個月與當月營收的成長幅度，找出當月營收大增的公司
獲利	月營收年成長(YoY)>20%	統計各公司去年同期與當月營收的成長幅度，找出當月營收相較於去年同期大增的公司
價值	過去3/6/10年平均股利>3.5	統計過去數年的股利分配狀況，藉此推算出長期營運良好的公司
價值	近10年股利分配率>80%	統計過去數年的股利分配狀況，找出長年有分配股利的公司
價值	股利>3	從今年已經公布股利分配的公司中找出高股利的公司
價值	現金殖利率>5%	從今年已經公布股利分配的公司找出高現金殖利率的公司
價值	股價淨值比<1.2	利用股價與淨值的比例找出股價相對於淨值在低水位公司
價值	股價淨值比<1.2	利用股價與淨值的比例找出股價相對於淨值已經在高位的公司
價值	本益比<15	利用股價與EPS的比例找出股價相對於利益在相對低位的公司

詳細的query請見github中的query.sql

## Others

# Progress of project



# Problem

1. 利用python網站爬蟲時，因為經常需要跳轉不同的網頁，有些有request limit的網站如:證交所，會限制一段時間內的request次數，如果超過ip會被ban掉，一開始沒有注意到ip就被ban了。

**solution:** 雖然後來限制了request的頻率，但證交所的限制頻率真的太低了，如果要爬完所有資料會花很多時間，因此需要大量跳轉網頁的資料例如:股價、籌碼資料就從沒有限制的康和證券與yahoo finance取得。

2. 當需要爬取的資料較多的時候，有些網站會因為長時間的連線而暫時中斷，導致爬到一半的時候就因斷線而中止程式，導致要全部重來花上不少時間。

**solution:** 再爬取資料的過程中就要同時轉成csv檔儲存紀錄，避免因斷線導致資料遺失，但在紀錄的同時也要注意不可以太過頻繁導致一直開關檔案降低效率。

3. 透過開放平台直接下載的表格經常會因為格式或特殊字元的問題導致匯入database的錯誤。

**solution:** 利用python先把檔案讀進去處理exception後再重新輸出與匯入。

4. 股價的table資料量太大，導致編寫有關技術面的query時會有效率上的問題。

**solution:** 在寫技術面的query時要盡量避免讓整個股價的table做Cartesian product，以免花太多時間，同時要注意其他會影響效率的因素。

5. 資料庫端與網頁端連接時使用json檔做資料傳遞，其中有用到中文的部分因為database裡TEXT型態編碼的關係，輸出json檔時皆為utf-8編碼。

**solution:** 將需要輸出成json檔的column轉變成VARCHAR()就可以正常輸出中文了。

6. 在做HTML網頁的時候發現網頁非常簡陋。

**solution:** 雖然老師說網頁呈現不佔分，但是我認為一個好的網頁，有良善的使用者介面是重要的一件事，所以我去學了CSS styling，可以用CSS去美化整個網頁的排版跟配色。

7. 我發現可以使用Bootstrap來建立一個RWD，但是剛開始都無法使用，而且就算照著網路上的教學去做成果都不一樣。

**solution:** 我原本下載的Bootstrap版本是5.0.1，但是上網查了之後發現目前這個版本在網路上的適用性不高，且影片中教學使用的版本是4.3.1，兩個版本的指令有些就不一樣，所以才無法使用。因此我選擇將自己的版本改成4.3.1，然後就順利的完成我想要的設計。

8. 在寫JavaScript的時候遇到很多困難及error。

**solution:** 因為HTML、CSS和JavaScript的語法都不盡相同，所以我很常會出現搞混用法然後整個網頁就會crash掉。而且同一種功能又有很多種寫法，例如我想找一個

class="token"的item，我可以用document.getElementsByClassName("token")或是我可以用document.querySelector(".token")，但是缺點就是只會回傳第一個，所以還要決定要用這個還是querySelectorAll，就造成很多不便。

所以我去學了jQuery，就可以把整個替換成\$(".token")，而且在選取的時候他會選取全部的物件，也不必擔心未來做刪減時會產生錯誤。

## Contribution

---

### 0816072 林子鈞

將code上傳至Github並做版本控制

利用python爬table: 法人進出

協助完成API功能，利用Express.js幫忙解析及輸出JSON檔

### 0816033 簡仔婕

前端所有呈現的部分包括html, css, js以及與express js的撰寫

1. HTML：負責網頁的架構 ex: button, text, input bar...
2. CSS：負責美化HTML的網頁內容 ex: 背景顏色, font, padding...
3. JavaScript & jQuery：利用DOM做動態網頁呈現，以及下拉式選單、與使用者互動的功能
4. Bootstrap (4.3.1)：主要負責Responsive Web Design(RWD)的呈現，讓使用者可以在不同的device上面都可以順利瀏覽網站
5. Node.js & Express.js (建立Web API)：用require express和bodyParser取得user input，啟動web service，連線資料庫，傳遞sql語法取得JSON檔，解析JSON並將結果更改至HTML。

presentation的時候負責網頁demo與講解

### 0816026 許傳聲

製作ER Diagram(for presentation)

編寫Query:基本面

製作Spotlight Video

### 0816086 黃立鈞

利用python爬table:上市公司列表、月營收、券商進出、季營收、股價、融資融券。

利用google插建取得table:價值

下載table並exception handel:股權分布

將table匯入mySQL workbench，並設定需要的資料型態、primary key等。

編寫Query:籌碼面、技術面、基本面。

製作presentation的ppt。

撰寫Final Report:Data、Database、Other part。

製作ER Diagram(for report)

## Repository

---

Link:Team9 (<https://github.com/jason10191019/Team9>).