



2023.03.15 會議記錄

updates in 2023.03.15

+ 會議資訊 Information

- 本次會議
- 截止日期 Deadline
- 下次會議

+ 會議內容 Contents

- 珮如 的 想法
- 文友 的 想法
- 仲元 的 想法

+ 工作追蹤 Work Tracking

- 珮如 的 工作
 - 文友 的 工作
 - 仲元 的 工作
-

+ 會議資訊 Information

- 本次會議

- 12:00 → 開會
 - 會議主題 → 追蹤工作進度, 思考新工作
-

- 截止日期 Deadline

- 仲元 做出甘特圖 → 2023.03.17
-

- 下次會議

- 下次開會日期 →
 - 若 仲元 做出甘特圖 → 2023.03.17 12:00

- 若 沒有做出 → 2023.03.19 12:00

+ 會議內容 Contents

- 追蹤至今工作進度
- 提出目前還有什麼具體想法

▼ 珮如 的想法

- 大盤資料已爬取完，說要測均線
- 週三, 週四去醫院
- 要想一下怎麼簡化專題

▼ 文友 的想法

1. 爬取每日買賣超總彙表

```
File "D:\Anaconda\lib\site-packages\pandas\io\nparsers\c_parser_wrapper.py", line 75, in __init__
    self._reader = parsers.TextReader(src, **kwds)

File "pandas\_libs\parsers.pyx", line 551, in pandas._libs.parsers.TextReader.__cinit__

EmptyDataError: No columns to parse from file
```

- EmptyDataError: No columns to parse from file → Dataframe找不到欄位
 1. 被網站方監測到正在爬蟲，短暫關閉連線
Solution → 重新執行程式
 2. 被網站方監測到正在爬蟲，封鎖IP
Solution → 使用VPN
 3. 爬取的當天是沒有資料的

Solution → 跳過這天

- 針對證交所反爬蟲的狀況，使用fake_useragent, requests.Session(), time.sleep()和VPN處理這個問題
- 有待改善的部分
 - 撰寫監控，即時訊息
 - 利用工作排程器自動化這部分內容

2. 爬取個股日成交資訊

資料日期：民國 112 年 02 月 股票代碼 (瀏覽)：2409

查詢

※ 本資訊自民國99年1月4日起開始提供

列印 / HTML CSV 下載

2409 友達
2409R 友達甲
2409S 友達乙
2409T 友達丙

112年02月 2409 友達 各日成交資訊

單位：元、股

日期	成交股數	成交金額	開盤價	最高價	最低價	收盤價	漲跌價差	成交筆數
112/02/01	36,369,661	602,329,121	16.75	16.80	16.35	16.65	-0.10	10,101
112/02/02	24,915,605	418,318,564	16.70	16.85	16.60	16.80	+0.15	9,066
112/02/03	25,595,656	424,734,731	16.85	16.85	16.50	16.50	-0.30	8,927
112/02/06	19,160,886	317,273,915	16.45	16.70	16.45	16.50	0.00	5,325
112/02/07	58,502,292	989,817,386	16.65	17.10	16.65	17.00	+0.50	14,692
112/02/08	112,562,833	1,977,349,221	17.10	17.90	16.90	17.70	+0.70	26,456
112/02/09	67,980,943	1,192,860,796	17.35	17.85	17.10	17.65	-0.05	16,607
112/02/10	55,683,620	964,108,033	17.60	17.65	17.15	17.15	-0.50	15,110
112/02/13	57,844,947	965,870,880	17.10	17.25	16.50	16.55	-0.60	22,241
112/02/14	55,227,544	941,201,075	16.75	17.25	16.70	17.15	+0.60	13,180
112/02/15	57,644,731	998,965,050	17.00	17.60	16.90	17.25	+0.10	15,391
112/02/16	80,995,009	1,442,444,208	17.50	18.00	17.40	17.95	+0.70	23,829
112/02/17	78,446,062	1,420,716,918	17.80	18.35	17.70	18.20	+0.25	19,885
112/02/20	118,317,474	2,212,111,202	18.50	18.95	18.20	18.85	+0.65	32,562
112/02/21	55,262,439	1,030,213,555	18.60	18.90	18.45	18.60	-0.25	15,877
112/02/22	54,477,818	1,013,496,545	18.40	18.75	18.35	18.75	+0.15	12,701
112/02/23	74,074,175	1,358,678,288	18.70	18.75	18.15	18.30	-0.45	20,250
112/02/24	87,641,154	1,632,293,242	18.15	18.80	18.15	18.60	+0.30	16,209

- 查詢網址方式是股票代碼和以資料日期為格式，每個報表以月為產量
 - 大量requests會耗時間 ← 目前糾結在這
-

▼ 仲元 的 想法

- 詢問專案Dead Line
-

+ 工作追蹤 Work Tracking

▼ 珮如 的 工作

- 決定好簡化專案想法
 - 做Baseline Model研究
-

▼ 文友 的 工作

- 整理資料集
 - 建立AWS環境
 - 等 珮如 決定好簡化想法後，再決定要怎麼整合投信
-

▼ 仲元 的 工作

- 週五前完成甘特圖
- 資料抓取，不要用爬蟲，抓各股
個股日成交資訊TWSE 臺灣證券交易所