
股價資訊爬蟲

運用 Pandas
Python 程式設計 王穎聰

連線到「台灣證券交易所」



會員專區 | 行動版 | 日本語 | English

全站搜尋



關於證交所

公司治理中心

交易資訊

上市公司

產品與服務

結算服務

市場公告

法令規章

上市公司整合資訊 (瀏覽代碼) :

代碼或名稱

搜尋

個股查詢 (瀏覽代碼) :

代碼或名稱

搜尋

上市公司條件選股 | 指標項目 :

收盤價

查詢條件 :

股價漲幅前 50 名

搜尋

四個工作

服務市場 創新商品
擴大規模 國際結盟

重要宣導 / 活動公告

盤中全面逐筆交易專區

107年度ETF短片徵選活動

107年投資講座

除權除息預告表

除權除息計畫結束表

109年03月23日將實施盤中全面逐筆交易，請點此處瀏覽詳細資訊。

107年11月23日 13:33 臺灣公司治理100指數 5,530.82 點 ▼20.29 (0.37%)



WebPro 3.0
影音傳輸網



公開資訊
觀測站

台灣證券交易所—交易資訊/盤後資訊/每日收盤行情



會員專區 | 行動版 | 日本語 | English

全站搜尋



關於證交所

公司治理中心

交易資訊

上市公司

產品與服務

結算服務

市場公告

法令規章

上市公司整合資訊 (瀏覽代碼) :

代碼

個股查詢 (瀏覽代碼) :

代碼或名稱

搜尋

上市公司條件選股 | 指標項目 :

收盤

前 50 名

搜尋

盤後資訊

每日收盤行情

每日市場成交資訊

每日第一上市外國股票成交量值

每日成交量前二十名證券

每5秒委託成交統計

各類指數日成交量值

個股日成交資訊

當日融券賣出與借券賣出成交量值

個股日收盤價及月平均價

個股月成交資訊

個股年成交資訊



109年

107年11月

點此處瀏覽詳細資訊。

30.82 點 ▼20.29 (0.37%)

重要宣導 / 活動公告

107年度ETF短片徵選活動

107年投資講座

除權除息預告表

除權除息計算結果表

Co-Location 主機共置

網路資訊商店

WebPro 3.0
影音傳輸網

盤中全面
逐筆交易
專區

股市旺
汪來富
網路圖解遊戲

公開資訊
觀測站

台灣證券交易所—分類/全部(不含..牛熊證)



臺灣證券交易所

首頁 > 交易資訊 > 盤後資訊 > 每日收盤行情

日期：民國 107 年 ▾ 11 月 ▾ 23 日 (五) ▾

分類：

大盤統計資訊

大盤統計資訊

委託及成交統計資訊

全部

全部(不含權證、牛熊證、可展延牛熊證)

封閉式基金

ETF

受益證券

認購權證(不含牛證)

認售權證(不含熊證)

牛證(不含可展延牛證)

熊證(不含可展延熊證)

可展延牛證

可展延熊證

附認股權特別股

附認股權公司債

認股權憑證

可轉換公司債

水泥工業

食品工業

塑膠工業

Q 查詢

2

※ 本資訊自民國93年2月11日起提供。

單位：元、股

107年11月23日

指數	收盤指數	指數	漲跌百分比(%)
實島股價指數	11,018.52	57.52	-0.52
發行量加權股價指數	9,667.30	47.41	-0.49
臺灣公司治理100指數	5,530.82	20.29	-0.37
臺灣50指數	7,279.80	22.57	-0.31

按下 F12 打開「檢查」



臺灣證券交易所

首頁 > 交易資訊 > 盤後資訊 > 每日收盤行情

日期：民國 107 年 ▾ 11 月 ▾ 23 日 (五) ▾ 分類：全部(不含權證、牛熊證、可展延牛熊證) ▾

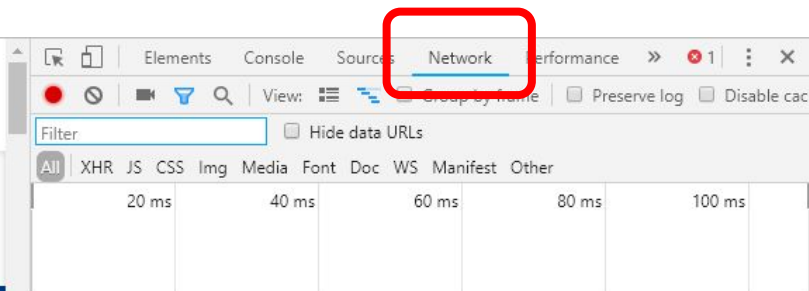
※ 本資訊自民國93年2月11日起提供。

/ HTML

單位：元、股

107年11月23日 大盤統計資訊

指數	收盤指數	漲跌(+/-)	漲跌點數	漲跌百分比(%)
實島股價指數	11,018.52	-	57.52	-0.52
發行量加權股價指數	9,667.30	-	47.41	-0.49
臺灣公司治理100指數	5,530.82	-	20.29	-0.37
臺灣50指數	7,279.80	-	22.57	-0.31
臺灣中型100指數	6,906.51	-	44.91	-0.65
臺灣資訊科技指數	9,134.70	-	22.84	-0.25
臺灣發達指數	7,577.64	-	30.66	-0.40



Recording network activity...

Perform a request or hit **Ctrl + R** to record the reload.

重新查詢看瀏覽器與伺服器的運作



臺灣證券交易所

首頁 > 交易資訊 > 盤後資訊 > 每日收盤行情

日期：民國 107 年 ▾ 11 月 ▾ 23 日 (五) ▾ 分類：全部(不含權證、牛熊證、可展延牛熊證) ▾ **1**

※ 本資訊自民國93年2月11日起提供。

列印 / HTML CSV 下載

單位：元、股

107年11月23日 大盤統計資訊

指數	收盤指數	漲跌(+/-)	漲跌點數	漲跌百分比(%)
實島股價指數	11,018.52	-	57.52	-0.52
發行量加權股價指數	9,667.30	-	47.41	-0.49
臺灣公司治理100指數	5,530.82	-	20.29	-0.37
臺灣50指數	7,279.80	-	22.57	-0.31
臺灣中型100指數	6,906.51	-	44.91	-0.65
臺灣資訊科技指數	9,134.70	-	22.84	-0.25
臺灣發達指數	7,577.64	-	30.66	-0.40

觀察

Name	Type	Initiator	Size	Time	Waterfall
MI_INDEX?respon...	xhr	main.js?...	58...	699...	
zhjson?_=154306...	xhr	main.js?...	762 B	133...	
zhjson?_=154306...	xhr	main.js?...	762 B	182...	
zhjson?_=154306...	xhr	main.js?...	762 B	189...	
zhjson?_=154306...	xhr	main.js?...	762 B	205...	
zhjson?_=154306...	xhr	main.js?...	762 B	217...	

觀察瀏覽器向伺服器提出的 Request



臺灣證券交易所

首頁 > 交易資訊 > 盤後資訊 > 每日收盤行情

日期：民國 107 年 ▾ 11 月 ▾ 23 日 (五) ▾ 分類：全部(不含權證、牛熊證、可展延牛熊證) ▾ 查詢

※ 本資訊自民國93年2月11日起提供。

列印 / HTML 下載 CSV

單位：元、股

107年11月23日 大盤統計資訊

指數	收盤指數	漲跌(+/-)	漲跌點數	漲跌百分比(%)
寶島股價指數	11,018.52	-	57.52	-0.52
發行量加權股價指數	9,667.30	-	47.41	-0.49
臺灣公司治理100指數	5,530.82	-	20.29	-0.37
臺灣50指數	7,279.80	-	22.57	-0.31
臺灣中型100指數	6,906.51	-	44.91	-0.65
臺灣資訊科技指數	9,134.70	-	22.84	-0.25
臺灣發達指數	7,577.64	-	30.66	-0.40

1

觀察

Request URL: http://www.twse.com.tw/exchangeReport/MI_INDEX?response=json&date=20181123&type=ALLBUT01-1543062598569

Request Method: GET

Status Code: 200

Remote Address: 163.29.17.152:80

Referrer Policy: no-referrer-when-downgrade

Response Headers

- Allow: GET, POST, HEAD
- Connection: keep-alive
- Content-Encoding: gzip
- Content-Type: application/json; charset=UTF-8
- Date: Sat, 24 Nov 2018 12:41:59 GMT
- Server: nginx
- Transfer-Encoding: chunked
- X-Application-Context: application:production

複製貼上 Request 的網址來查看

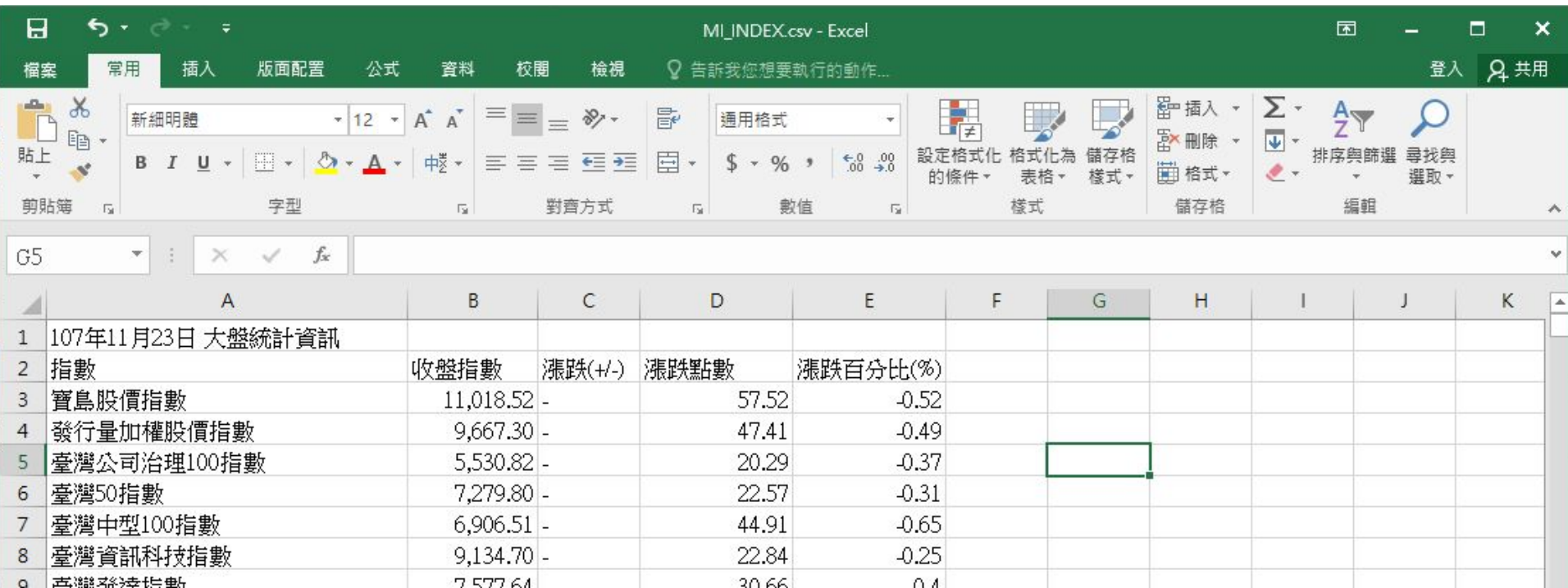
Browser window showing the TWSE website (www.twse.com.tw/exchangeReport/...) with a JSON response displayed in the console.

JSON Response:

```
{
  "data4": [
    [
      ["上漲(漲停)", "2,055(4)", "213(3)"],
      ["下跌(跌停)", "4,788(39)", "581(0)"],
      ["持平", "513", "106"],
      ["未成交", "9,316", "11"],
      ["無比價", "1,168", "5"]
    ],
    "fields1": [
      "指數", "收盤指數", "漲跌(+/-)", "漲跌點數", "漲跌百分比(%)"
    ],
    "subtitle4": "漲跌證券數合計",
    "notes4": [
      "\"漲跌價差\"為當日收盤價與前一日收盤價比較。",
      "\"無比價\"含前一日無收盤價、當日除權、除息、新上市、恢復交易者。",
      "\"外幣成交值係以本公司當日下午3時30分公告匯率換算後加入成交金額。<br>公告匯率請參考本公司首頁>交易資訊>雙幣ETF專區>代號對應及每日公告匯率。"
    ],
    "data5": [
      ["0050", "元大台灣50", "4,041,215", "2,078", "304,440,003", "75.70", "75.85", "75.05", "75.35"]
    ]
  ]
}
```


將網址上的 json 改為 csv

會下載 csv 檔, 會用 Excel 開啟觀看結果



The screenshot shows the Microsoft Excel interface with the file 'MI_INDEX.csv' open. The ribbon includes '檔案', '常用', '插入', '版面配置', '公式', '資料', '校閱', and '檢視'. The '常用' ribbon is active, showing options for font, alignment, numbers, styles, and editing. The data is organized in a table with columns A through K. The first column (A) lists various market indices, and the subsequent columns (B-E) provide numerical data for each index, including closing values, changes, and percentages. The row for '臺灣公司治理100指數' is highlighted.

	A	B	C	D	E	F	G	H	I	J	K
1	107年11月23日 大盤統計資訊										
2	指數	收盤指數	漲跌(+/-)	漲跌點數	漲跌百分比(%)						
3	寶島股價指數	11,018.52	-	57.52	-0.52						
4	發行量加權股價指數	9,667.30	-	47.41	-0.49						
5	臺灣公司治理100指數	5,530.82	-	20.29	-0.37						
6	臺灣50指數	7,279.80	-	22.57	-0.31						
7	臺灣中型100指數	6,906.51	-	44.91	-0.65						
8	臺灣資訊科技指數	9,134.70	-	22.84	-0.25						
9	臺灣發達指數	7,577.64	-	30.66	0.4						

開始撰寫程式1

```
import requests
response = requests.get("http://www.twse.com.tw/exchangeReport/
                        MI_INDEX?response=csv&date=20181123&type=ALLBUT0999
                        &_=1543066495514")
print( response.text[:1000] )  #暫時列印 1000 個字元
```

印出結果:

```
"107年11月23日 大盤統計資訊"
"指數","收盤指數","漲跌(+/-)","漲跌點數","漲跌百分比(%)",
"寶島股價指數","11,018.52","-","57.52","-0.52",
"發行量加權股價指數","9,667.30","-","47.41","-0.49",
....
```

撰寫程式2 (注意: 非正確結果)

```
import pandas as pd
from io import StringIO
df = pd.read_csv( StringIO(response.text) )
```

出現錯誤結果: **Traceback (most recent call last):**

pandas.errors.ParserError: Error tokenizing data. C error:
 Expected 6 fields in line 163, saw 17

因為在轉換成 DataFrame 時, 分析原先只有 6 個欄位, 但分析到中間, 卻出現了 17 欄, 不知該如何處理, 因此產生 Error

撰寫程式2 (正確應先過濾資料)

```
newlines = []  
lines = response.text.split("\n")  
for line in lines:  
    if len( line.split(',') ) == 17:    #有17項資料的才是我們需要的  
        newlines.append( line )  
print( newlines[:10] )
```

```
["證券代號","證券名稱","成交股數","成交筆數","成交金額","開盤價","最高價","  
最低價","收盤價","漲跌(+/-)","漲跌價差","最後揭示買價","最後揭示買量","最後揭  
示賣價","最後揭示賣量","本益比",\r',  
'"0050","元大台灣50","4,041,215","2,078","304,440,003","75.70","75.85",  
"75.05","75.35","-", "0.25","75.30","168","75.35","93","0.00",\r', ....
```

撰寫程式3 (再重組成字串, 產生 DF)

```
import pandas as pd
from io import StringIO
df = pd.read_csv( StringIO("\n".join( newlines )) )
print( df.head() ) #印出 DataFrame 的前五筆
```

	證券代號	證券名稱	成交股數	...	最後揭示賣量	本益比	Unnamed: 16
0	"0050"	元大台灣50	4,041,215	...	93 0.00	NaN	
1	"0051"	元大中型100	6,934	...	10 0.00	NaN	
2	"0052"	富邦科技	100,000	...	1 0.00	NaN	
3	"0053"	元大電子	1,000	...	20 0.00	NaN	
4	"0054"	元大台灣50	11,000	...	6 0.00	NaN	

[5 rows x 17 columns]

有「等號」, 該去除

撰寫程式3 (修改去除等號)

```
import pandas as pd
from io import StringIO
df = pd.read_csv( StringIO("\n".join( newlines ).replace("=", "")) )
print( df.head() )
```

	證券代號	證券名稱	成交股數	...	最後揭示賣量	本益比	Unnamed: 16
0	0050	元大台灣50	4,041,215	...	93 0.00	NaN	
1	0051	元大中型100	6,934	...	10 0.00	NaN	
2	0052	富邦科技	100,000	...	1 0.00	NaN	
3	0053	元大電子	1,000	...	20 0.00	NaN	
4	0054	元大台商50	11,000	...	6 0.00	NaN	

[5 rows x 17 columns]

這個索引值沒有意義，改所引為「證券代號」

撰寫程式3 (修改索引值)

```
import pandas as pd
from io import StringIO
df = pd.read_csv( StringIO("\n".join( newlines ).replace("=", "")) )
df = df.set_index("證券代號")
print( df.head() )
```

證券名稱	成交股數	成交筆數	...	最後揭示賣量	本益比	Unnamed: 16
證券代號						
0050 元大台灣50	4,041,215	2,078	...	93 0.00	NaN	
0051 元大中型100	6,934	9	...	10 0.00	NaN	
0052 富邦科技	100,000	5	...	1 0.00	NaN	
0053 元大電子	1,000	1	...	20 0.00	NaN	
0054 元大台商50	11,000	9	...	6 0.00	NaN	

[5 rows x 16 columns]

這是字串，有逗點，需去除，才能轉數字處理

撰寫程式4 (去除數字字串上的逗點)

```
df = df.astype( str )  
df = df.apply( lambda s: s.str.replace(",", "") )  
print( df.head( ) )
```

證券名稱	成交股數	成交筆數	...	最後揭示賣量	本益比	Unnamed: 16
證券代號	...					
0050 元大台灣50	4041215	2078	...	93 0.00	nan	
0051 元大中型100	6934	9	...	10 0.00	nan	
0052 富邦科技	100000	5	...	1 0.00	nan	
0053 元大電子	1000	1	...	20 0.00	nan	
0054 元大台商50	11000	9	...	6 0.00	nan	

[5 rows x 16 columns]

撰寫程式5 (轉換字串為數字,去除非數字)

```
df = df.apply(lambda s: pd.to_numeric(s, errors="coerce"))
df = df[ df.columns[ df.isnull().sum() != len( df ) ] ]
print( df.head() )
```

證券代號	成交股數	成交筆數	成交金額	開盤價 ...	最後揭示買量	最後揭示賣價	最後揭示賣量	本益比
0050	4041215	2078	304440003	75.70 ...	168	75.35	93	0.0
0051	6934	9	198225	28.61 ...	2	28.53	10	0.0
0052	100000	5	4753290	47.68 ...	1	47.67	1	0.0
0053	1000	1	30330	30.33 ...	20	30.24	20	0.0
0054	11000	9	213170	19.57 ...	20	19.34	6	0.0

[5 rows x 13 columns]

撰寫程式6 (選取收盤價較開盤價盈餘 5%)

```
print( df[ df["收盤價"] / df["開盤價"] > 1.05 ] )
```


撰寫程式7 (寫入與讀取 CSV 檔)

```
df.to_csv( "daily_price.csv", encoding="utf_8_sig" )  
df = pd.read_csv( "daily_price.csv", index_col=["證券代號"] )  
print( df.head() )
```

撰寫程式8 (寫入與讀取 SQLite)

```
import sqlite3
conn = sqlite3.connect("test.sqlite3")

df.to_sql( "daily_price", conn, if_exists="replace" )
df = pd.read_sql( "select * from daily_price", conn, index_col=["證券代號"] )
print( df.head() )
```

下載安裝可開啟 SQLite 檔案的應用程式

← → ↻ https://sqlitebrowser.org



將書籤放置在書籤列上，即可快速前往各個網頁。

DB Browser for SQLite

The Official home of the DB Browser for
SQLite



// News

2018-10-10 - The first Windows and mac alpha builds for our next major release [are ready](#). Please [try them out](#), and [report any weirdness](#).

2018-08-09 - We've just started testing a new download server cluster. If anything seems weird with our downloads, please report it [here](#).

2018-06-08 - We've just created a Patreon account. Please [become a Patron](#) of DB Browser for SQLite! :)



Download 32-bit
Windows .exe



Download 64-bit
Windows .exe

整理

1. `pd.to_numeric(series)` 將 series 轉型成數字
2. `df.apply(func)` 將 dataframe 中的每個 series 都透過 func 處理
3. `lambda x: y` 製作一個不具名的 function, 傳入 x 而輸出 y
4. `df.set_index(col_name)` 設定某個欄位當做 index