

台股上市櫃股票成交額、股價爬蟲

2021/4/4 Wesley Wei

簡介：

從一個 Excel 檔案中取得一欄股價代號，抓取固定欄位兩組的開始和結束日期，來爬這個時間範圍內的清單股票的平均成交額，和指定日期的股價並輸入回 Excel 表格。

說明：

1. 成交額，透過可選擇檔案位置的特定.xlsx 檔案，取得股票代號(A 欄)和平均成交額計算的日期區間參數(E1、F1 欄)，來爬資料和計算清單上股票日期範圍內的成交額(單位元新台幣)，並自動填入原本.xlsx 表格中 B 欄(從 B2 開始往下寫。(請參考下示意圖)
2. 股價，透過可選擇檔案位置的特定.xlsx 檔案，取得股票代號(A 欄)和日期區間(開啟日(G1 欄，結束日 H1 欄)後，將當日收盤價，存入 G 和 H 欄(G2、H2 開始往下填)。(請參考下示意圖)
3. 判斷資料來源，透過判斷 I1 欄位，判斷式上市還是上櫃，來選擇爬資料來源網站。

	A	B	C	D	E	F	G	H	I
1	股票代號	平均成交	起始股價	結束股價	2020/3/10	2020/8/20	2020/2/10	2020/6/1	上市
2	1101								
3	2354								
4	1103								
5	8401								
6	大約1500筆								
7									
8									

補充規格：

1. 爬蟲程式需用 Python 來編寫，需提供 Python 原始碼，方便後續使用者自行維護，需額外說明設定爬蟲更新時間參數在哪裡，以避免未來來源修改規格，使用者可以自己調整爬蟲更新時間。
2. 假設資料約 800 筆，平均成交額(假設 3 個月區間)、股價資料需在 2 小時內跑完。
3. 說明 1、2 功能要設計可以分開或一起執行(效能考量)。
4. 結束後需跳出完成訊息和用時多久，如有沒抓到的資料也需在原本應該填入欄位寫 null。

參考資料來源：

1. 上市價成交額來源(需注意成交額用的是千)：

https://www.twse.com.tw/zh/page/trading/exchange/STOCK_DAY.html

2. 上櫃股價成交額來源：

https://www.tpex.org.tw/web/stock/aftertrading/daily_trading_info/st43.php?l=zh-tw