

# 金融科技(finTech)\_HW3\_report

111061562\_電機系碩士班\_林泓祐

## 1. data processing

### step\_1:

利用 pandas 套件將提供的所有.csv 檔讀進來，並且合成為一個獨立  
dataFrame 物件。

### Step\_2:

根據題目要求利用 loc 過濾出所要的"商品代號"---"TX"，且到期月份  
為"202110"且交易時間為"8:45"~"13:45"的每筆資料。

### Step\_3:

將"近月價格"、"遠月價格"、"開盤集合競價"等等不需要用的資料(column)  
從 dataFrame 物件中移除。

### Step\_4:

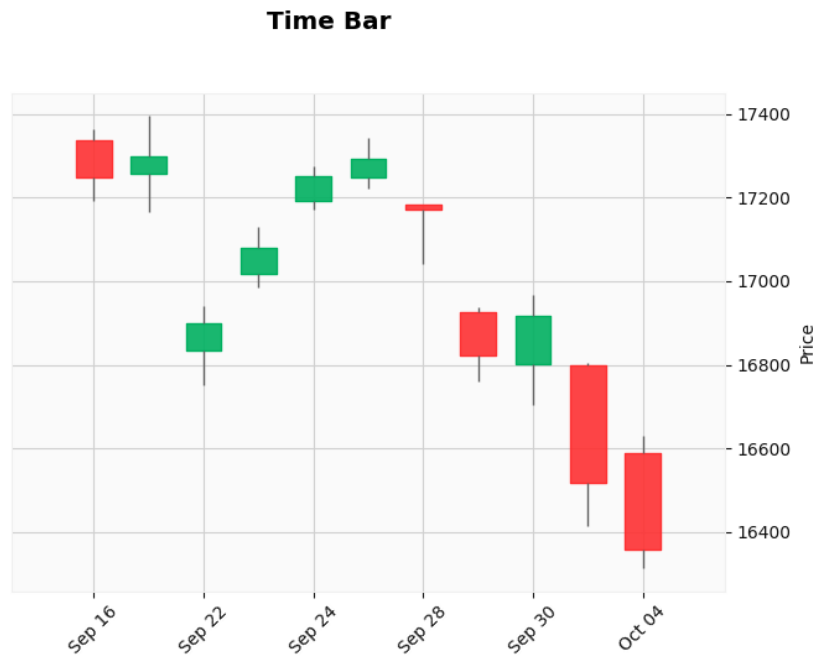
將 dataFrame 物件輸出成.csv 檔。輸出的.csv 檔如下圖所示，總共有  
1146160 筆資料，但由於資料量超過每個.csv 可以負荷的負載，所以將其分  
為兩個.csv 檔來操作，分別為 TX\_excel\_1.csv 以及 TX\_excel\_2.csv。

1	成交日期	商品代號	到期月份	成交時間	成交價格	成交數量(B+S)
2	20210916	TX	202110	84500	17337	714
3	20210916	TX	202110	84500	17338	2
4	20210916	TX	202110	84500	17337	4
5	20210916	TX	202110	84500	17340	16
6	20210916	TX	202110	84500	17341	2
7	20210916	TX	202110	84500	17342	12
8	20210916	TX	202110	84500	17343	2
9	20210916	TX	202110	84500	17345	6
10	20210916	TX	202110	84500	17344	4
11	20210916	TX	202110	84500	17344	4
12	20210916	TX	202110	84500	17345	8
13	20210916	TX	202110	84500	17344	2
14	20210916	TX	202110	84500	17341	4
15	20210916	TX	202110	84500	17345	10
16	20210916	TX	202110	84500	17347	2
17	20210916	TX	202110	84500	17347	2
18	20210916	TX	202110	84500	17344	2
19	20210916	TX	202110	84500	17347	4
20	20210916	TX	202110	84500	17348	4
21	20210916	TX	202110	84500	17344	2
22	20210916	TX	202110	84500	17343	2
23	20210916	TX	202110	84500	17342	2
24	20210916	TX	202110	84500	17342	2
25	20210916	TX	202110	84500	17342	2

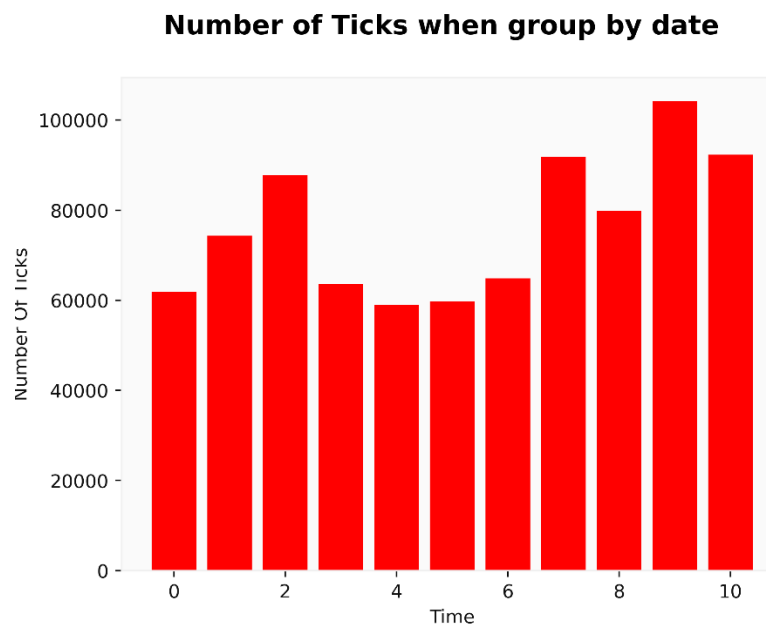
## 2. Time Bar

### Solution:

針對“成交日期”做 groupby()，也就是以每天為單位合併資料，所以共有 11 筆資料(11 天的資料)，以每天的第一筆“成交價格”為開盤價，最後一筆為收盤價。接著將計算後的資料整理成 timebar\_result=["Date", "Volume", "Open", "High", "Low", "Close"]，並將每天的 timebar\_result 轉為 DataFrame 的格式輸出(time\_bar)。最後使用 mplfinance 套件，把 time\_bar 化成 k 線圖，綠色代表收高，紅色代表收低。



下圖為在 group\_by\_date 下，每天的 tick 數量(number of ticks per date)。



### 3. Tick Bar

#### Solution:

依題意要求，將每 10000 筆資料合併為一個 group，因為總資料量為

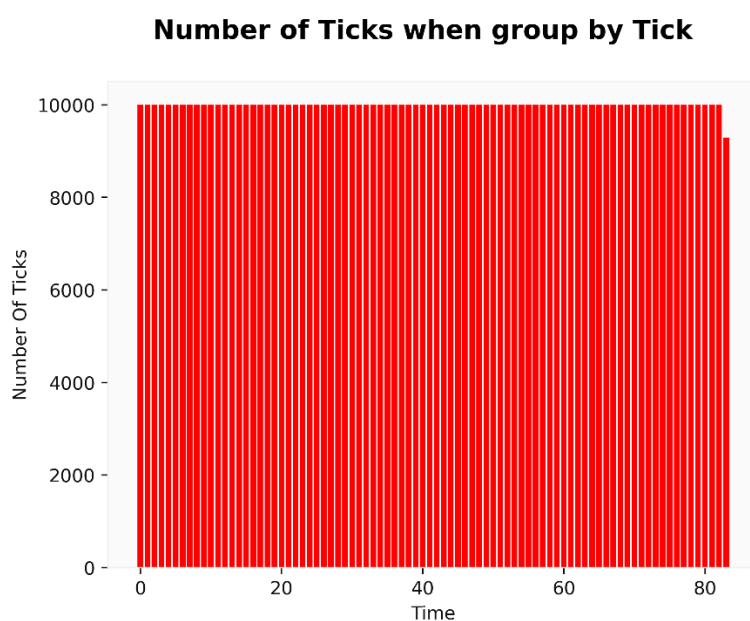
839299 比所以合併之後，總共會有 84 筆資料( $839299 // 10000 + 1$ )。開

盤價以及收盤價分別為每筆資料中的第一以及最後的“成交價格”，最大及最

小分別為每筆資料中最高以及最低的“成交價格”。接著與 Timebar 相同，將每筆資料彙總成 `Tick_result = ["Date", "Volume", "Open", "High", "Low", "Close"]`，最後輸出成 `DataFrame(tick_bar)` 用 `mplfinance` 畫出，如下圖所示，可以看出大趨勢與 Time bar 相同。



tick 則是除了最後一筆資料是尾數之外，其餘皆為 10000



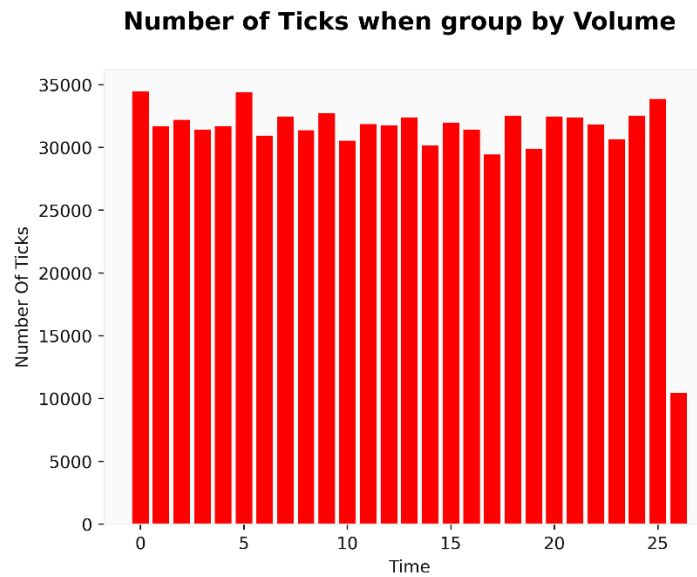
## 4. Volume Bar

## Solution:

依題意要求，將每 100,000 個“成交數量”。計算累積成交數量的方式，是利用 `cumsum()` 將“成交數量”從 20210915 的第一筆資料開始做累加至 20211004 的最後一筆資料，可以得知總資料的“成交數量”為 2633818，將  $2633818 // 100,000 = 26$ ，可得在 `groupby` 之後，會有 27(0~26)筆資料，最後像 Time bar 一樣，將每筆資料彙總成 `volume_result = ["Date", "Volume", "Open", "High", "Low", "Close"]`，最後輸出成 `DataFrame(volume_bar)` 用 `mplfinance` 畫出，如下圖所示，可以看出大趨勢與 Time bar、Tick bar 相同。



下圖為 `groupby "volume"` 時，每筆資料的 number of ticks。



## 5. Dollar Bar

### Solution:

依題意要求要將每 100,000,000 的交易金額作為 grouping 的依據，計算的方式是原本每筆資料的“成交價格” \* “交易數量”獲得“每筆資料的交易金額”，接著一樣用 cumsum()將每筆資料的交易金額做累加，可以得知最後得總交易金額為  $4.46 \times 10^{10}$ ，以  $1^9$  為 group 的話可以合併成 45 筆資料 (0~44)，最後像 Time bar 一樣，將每筆資料彙總成 volume\_result = ["Date", "Volume", "Open", "High", "Low", "Close"]，最後輸出成 DataFrame(Dollar\_bar)用 mplfinance 畫出，如下圖所示，可以看出大趨勢與 Time bar、Tick bar、Volume bar 相同。

**Dollar Bar**



下圖為 groupby“交易金額”時，每筆資料的 number of ticks。

**Number of Ticks when group by Dollar**

