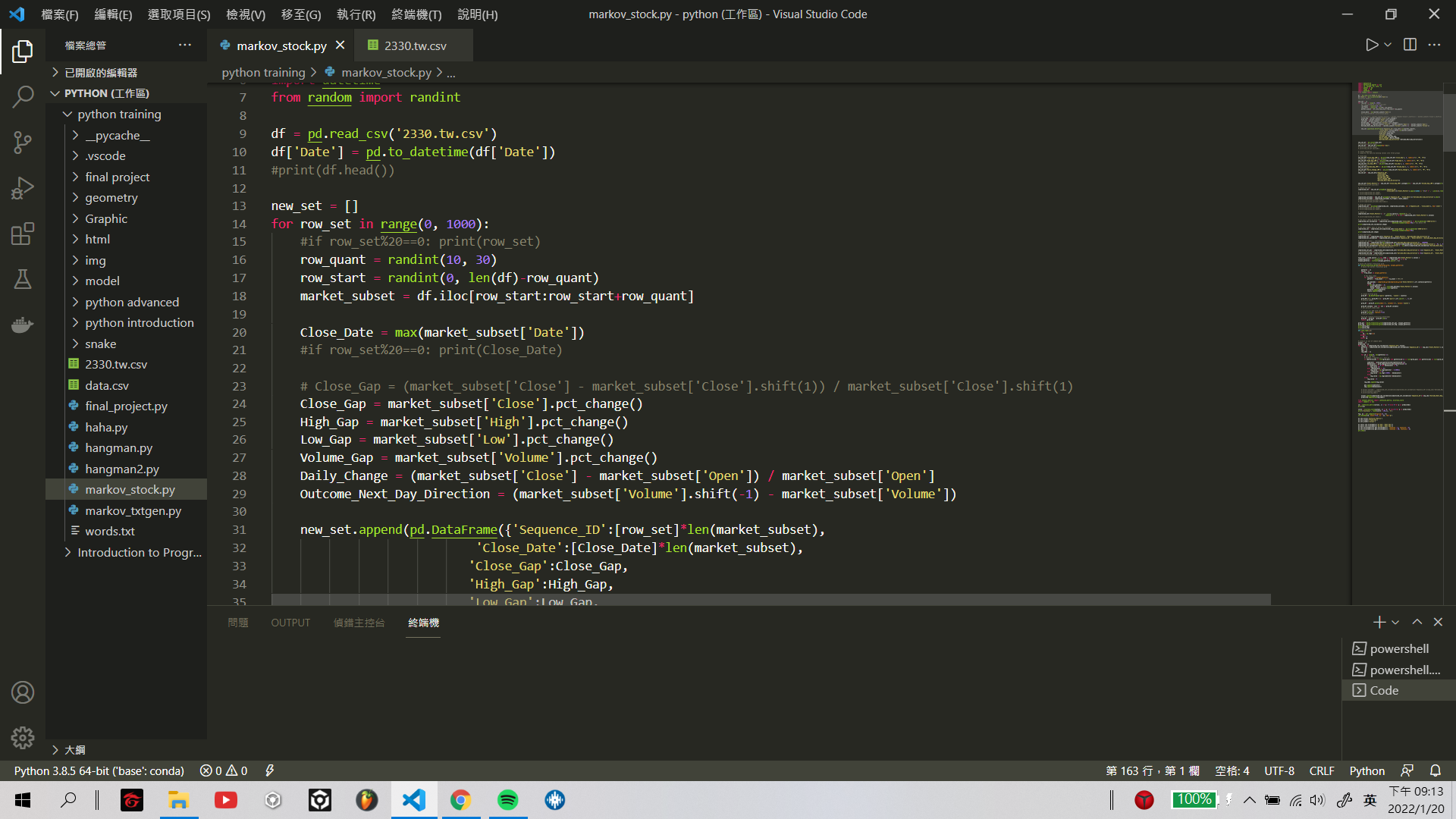
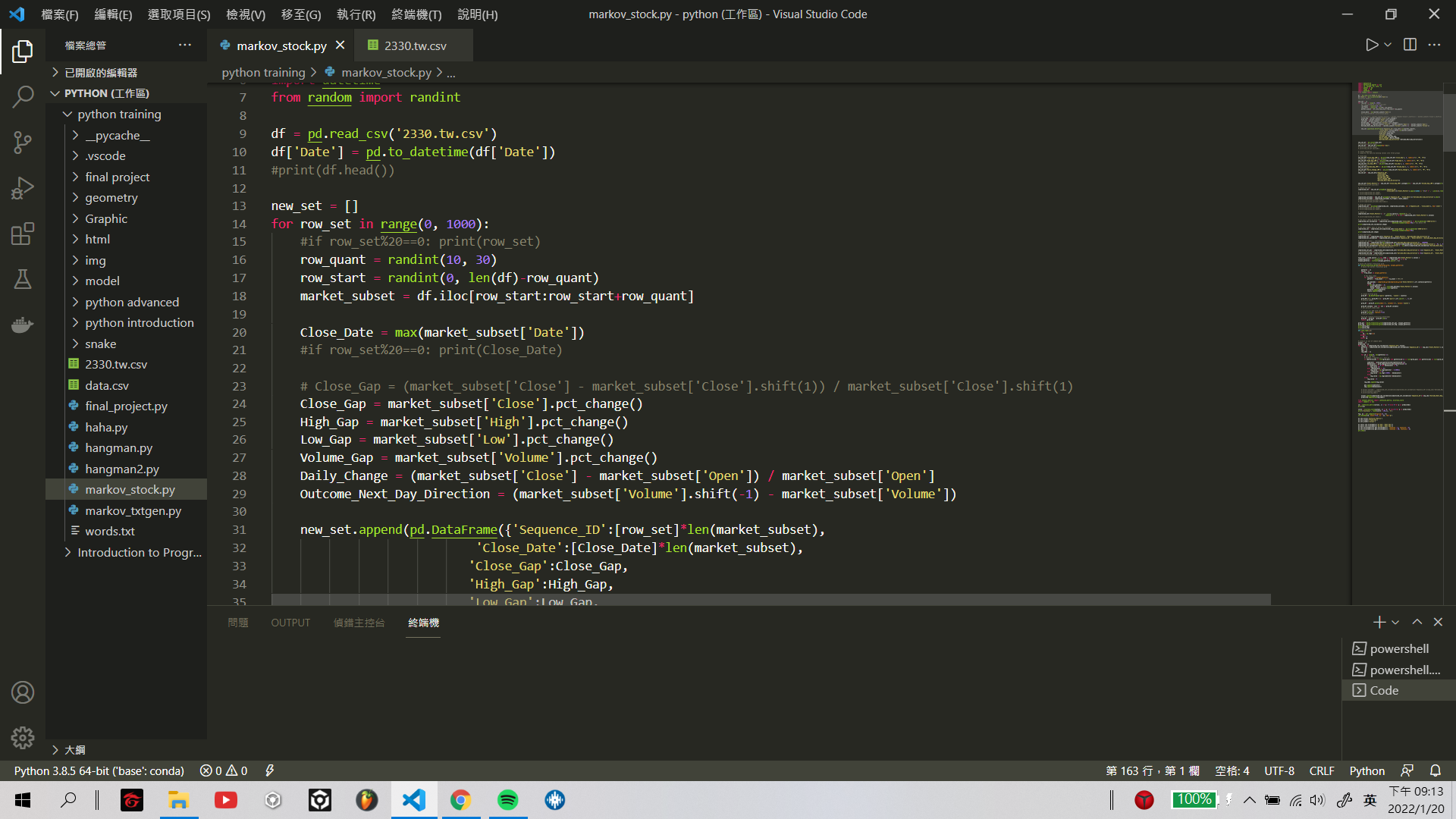
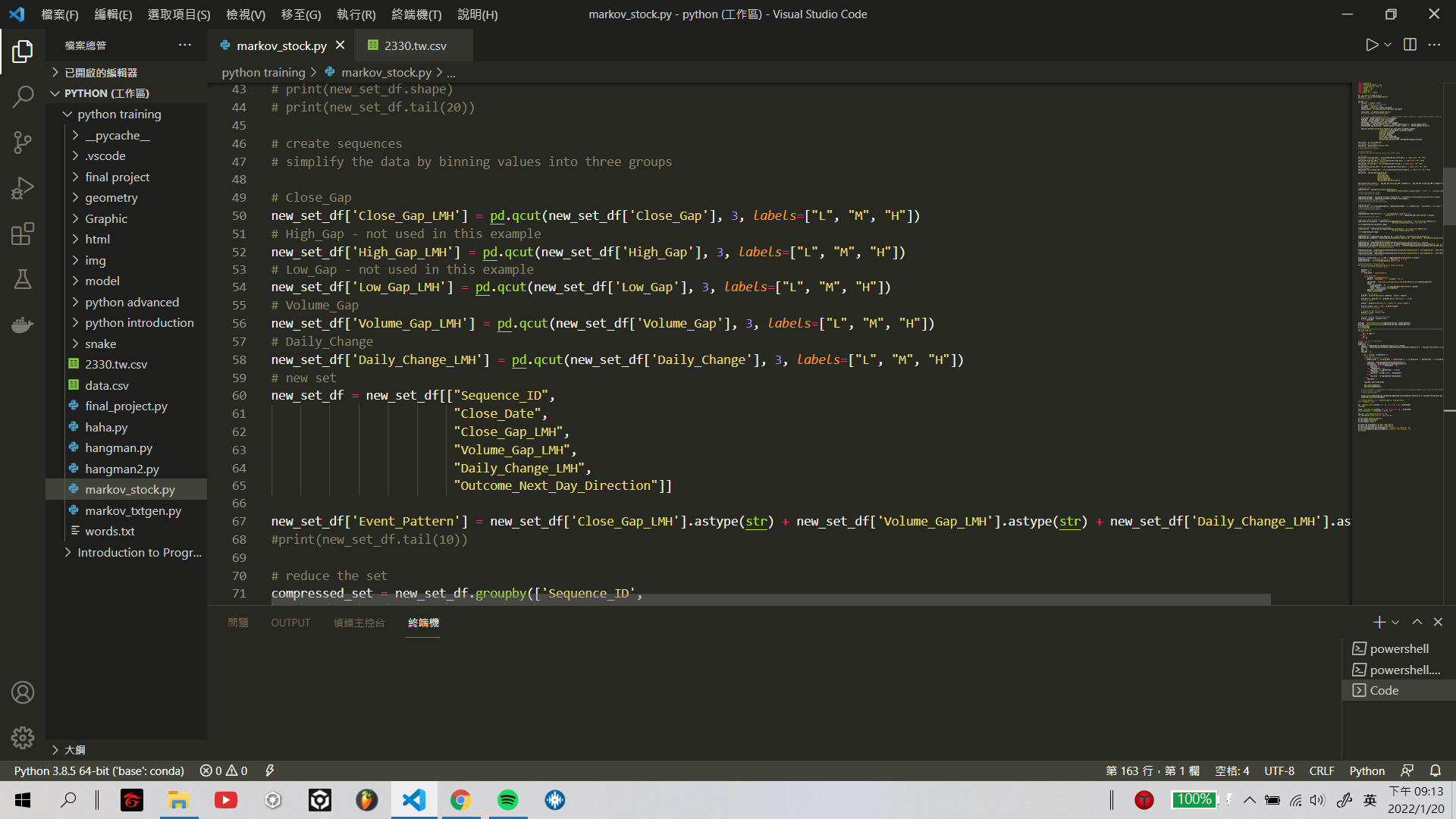


先匯入csv檔，這裡是台積電2019和2020年的股票資料。並且把時間轉成可以比較大小的格式。



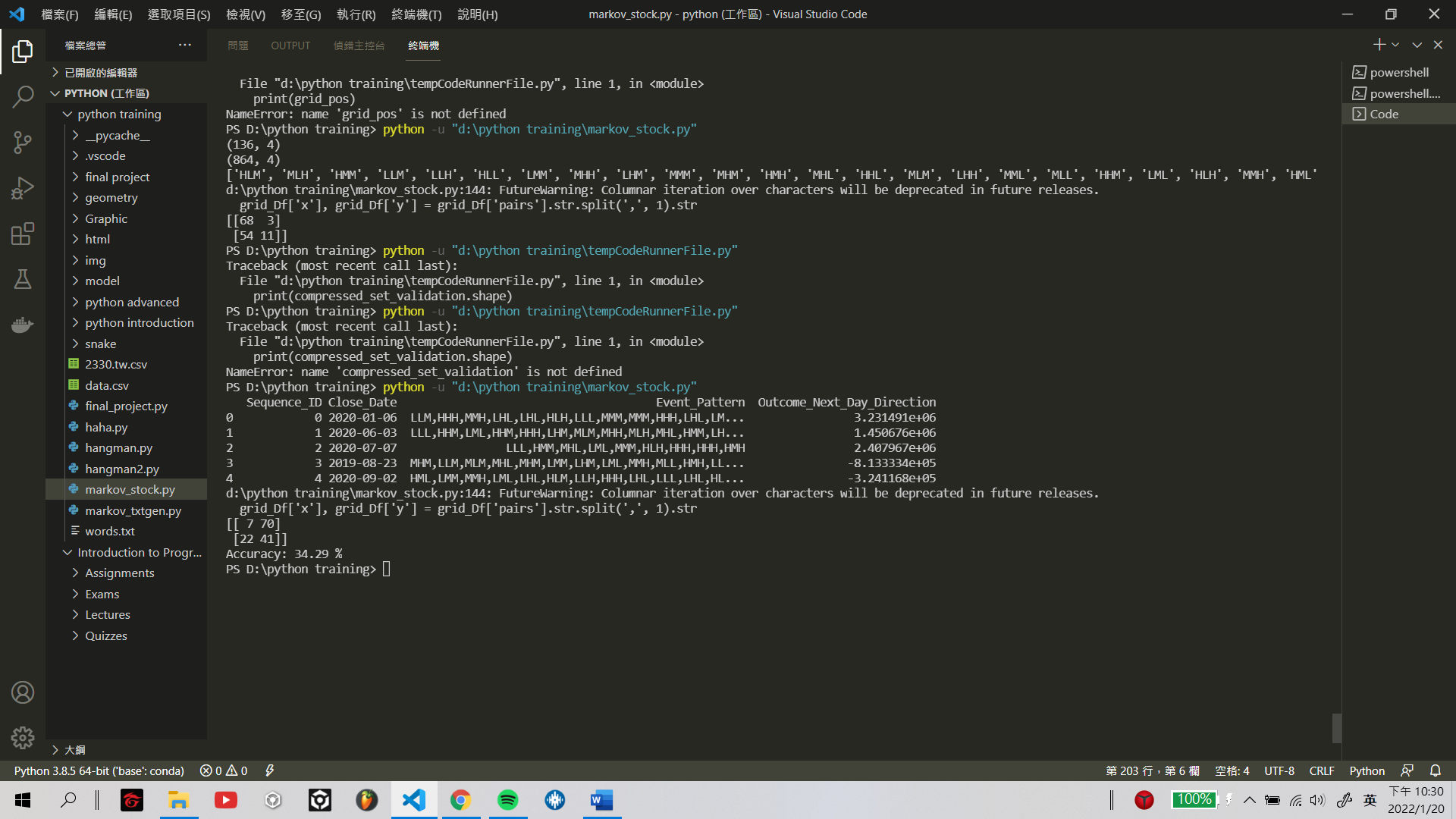
取1000筆資料（PPT內取10000筆），每一筆資料會隨機取10至30天的數據做為一個數據串，並且以其中最大的日期作為代號。



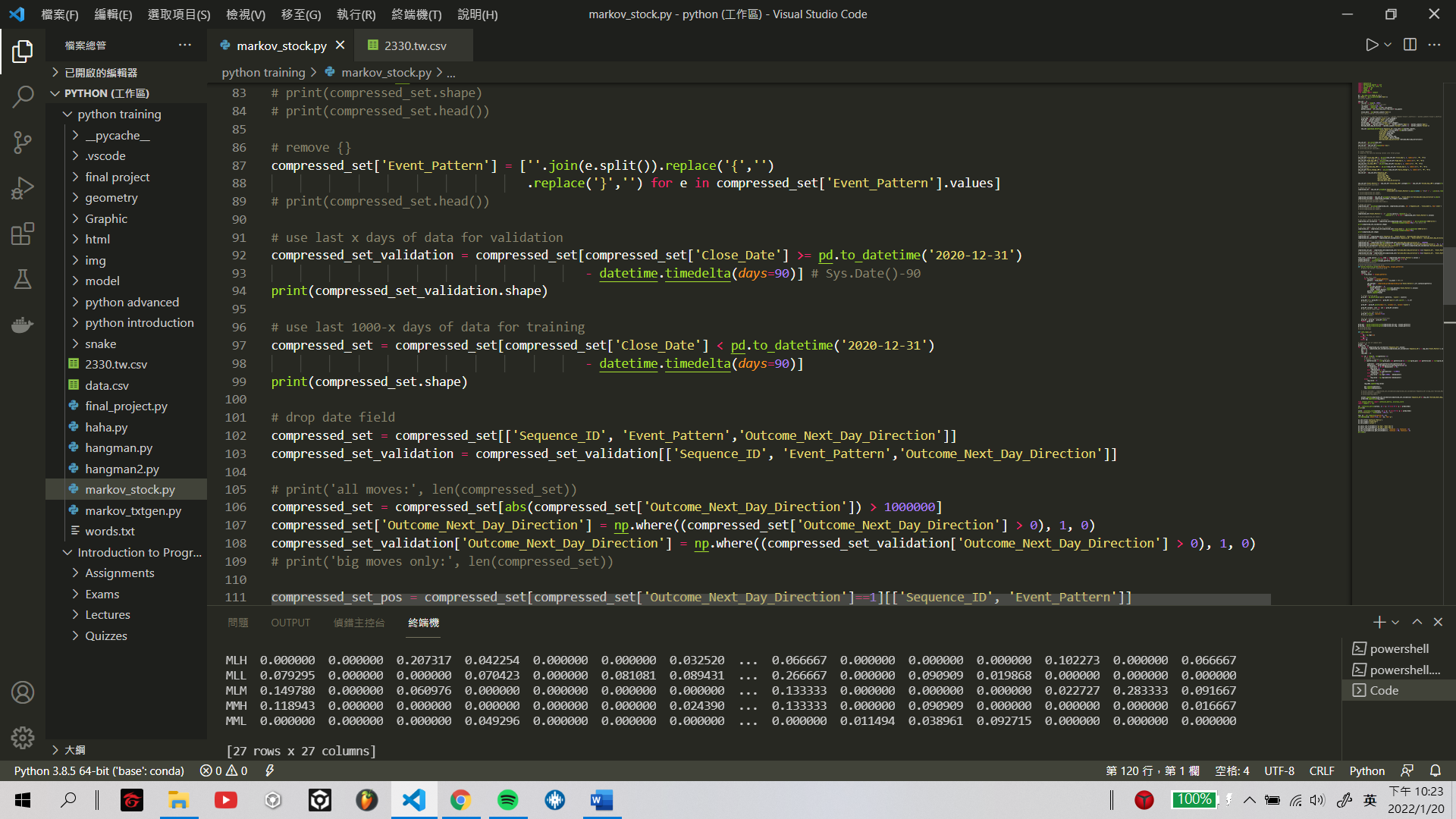
這裡把每一個數字替換掉，Close\_Gap把收盤價換成與前一天收盤價的變動百分比，以此類推。Daily\_Change則是當天(收盤價 – 開盤價)/開盤價，Outcome\_Next\_Day\_Direction則是交易量的變動。

把所有的變動量分級成Low, Medium, High，以便表示。

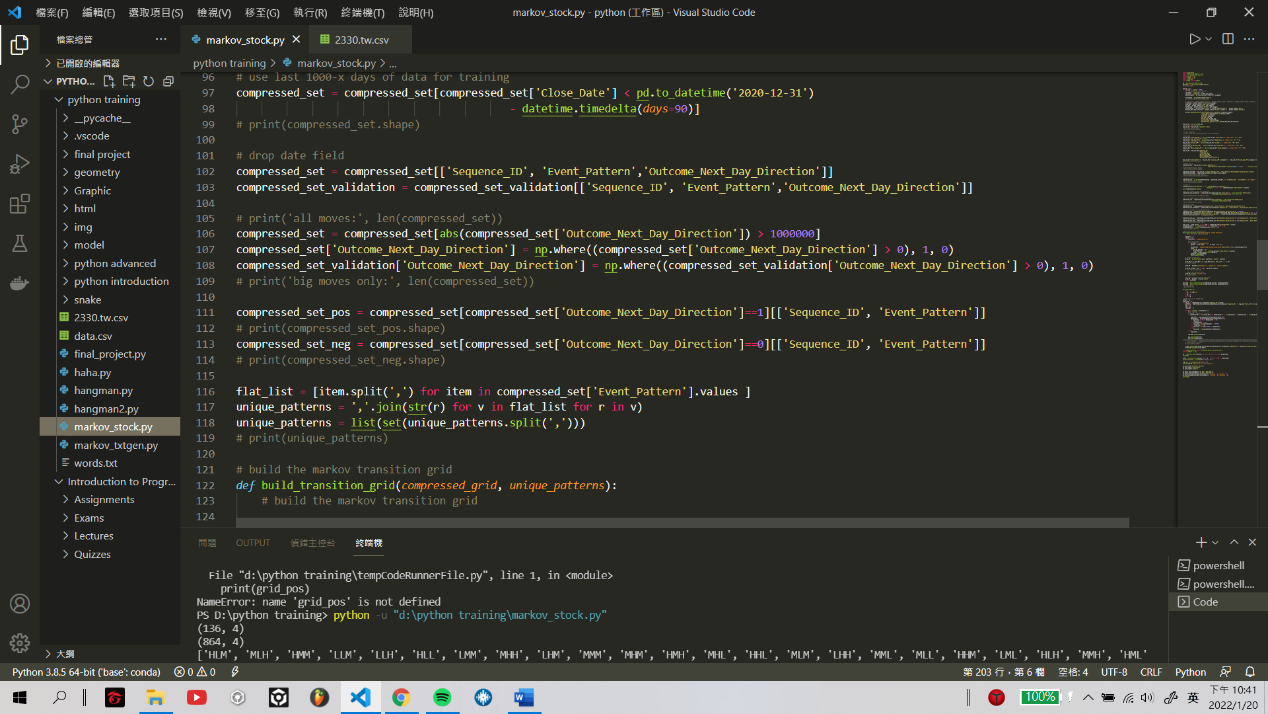
我們不會使用High\_Gap和Low\_Gap，因為這樣便會有3的5次方大。



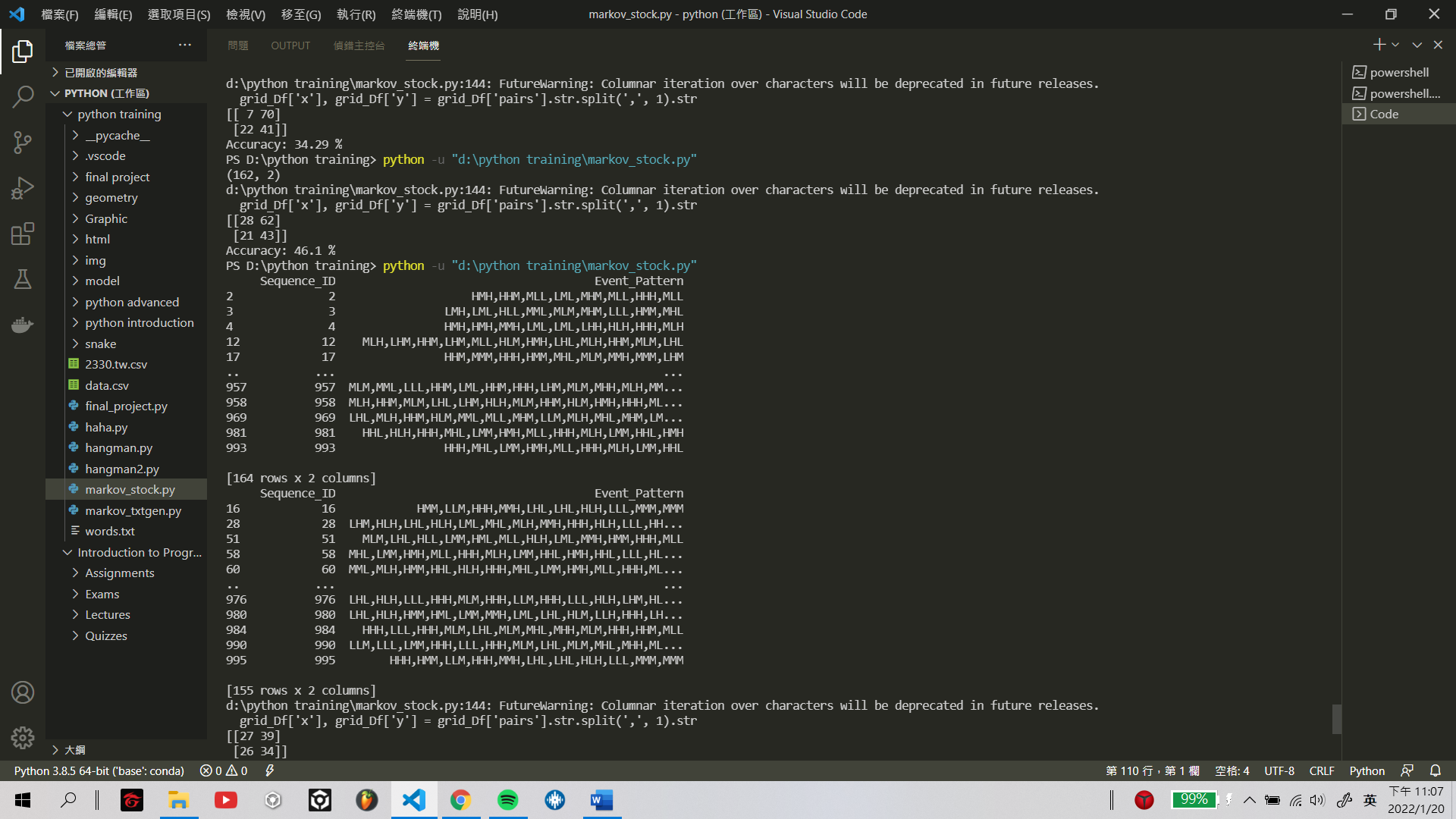
最後會將資料整理成這樣。LLM表示著低Close\_Gap，低Volum\_Gap，中Daily\_Change。以此類推。



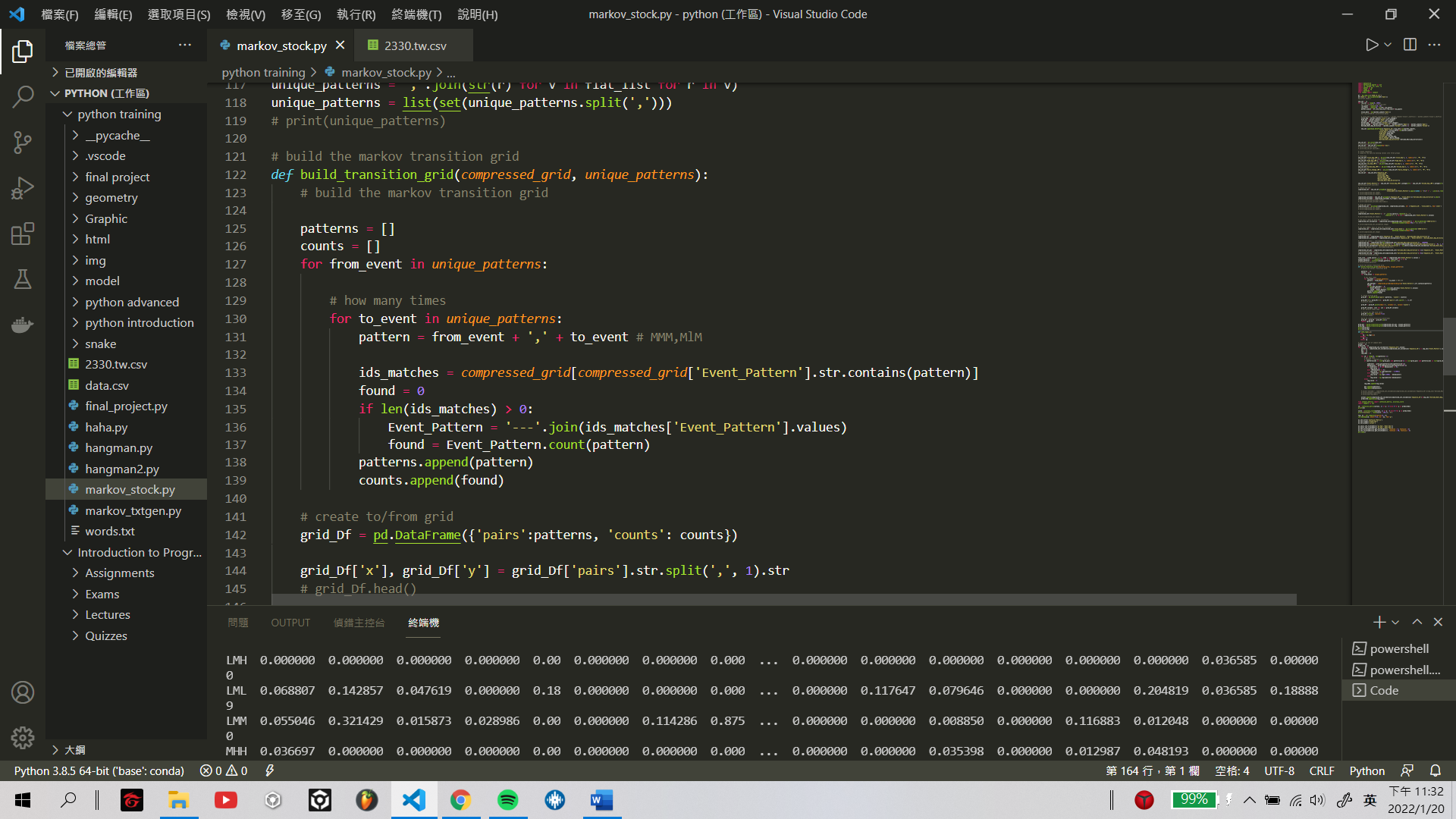
將資料分成兩筆，一筆為拿來寫出馬可夫矩陣的training，一則為當作基準的Validation。

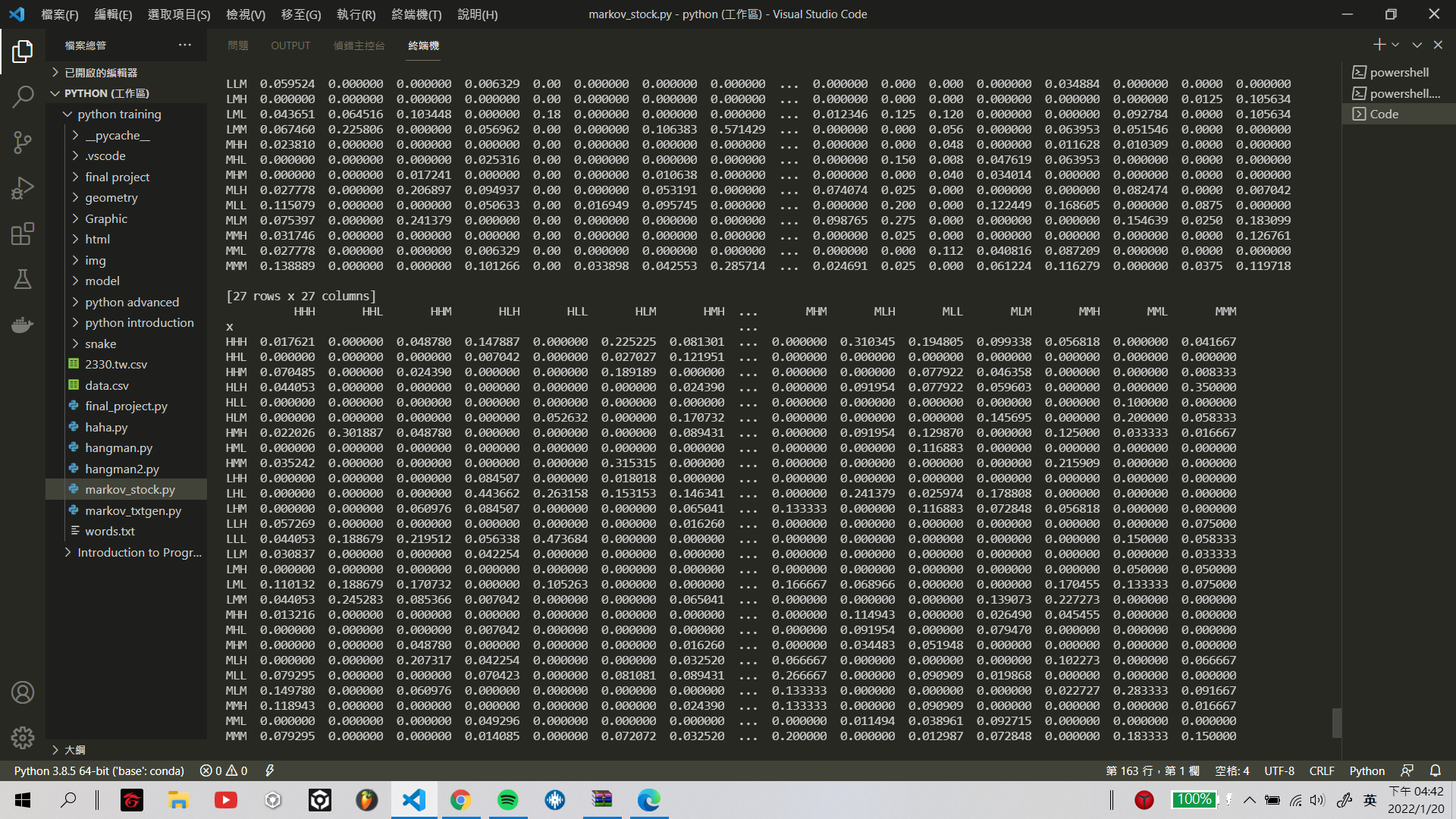


將交易量的變動小於1000000的資料串刪除，並且將交易量變動為正的設為1，交易量變動為負的設為0。

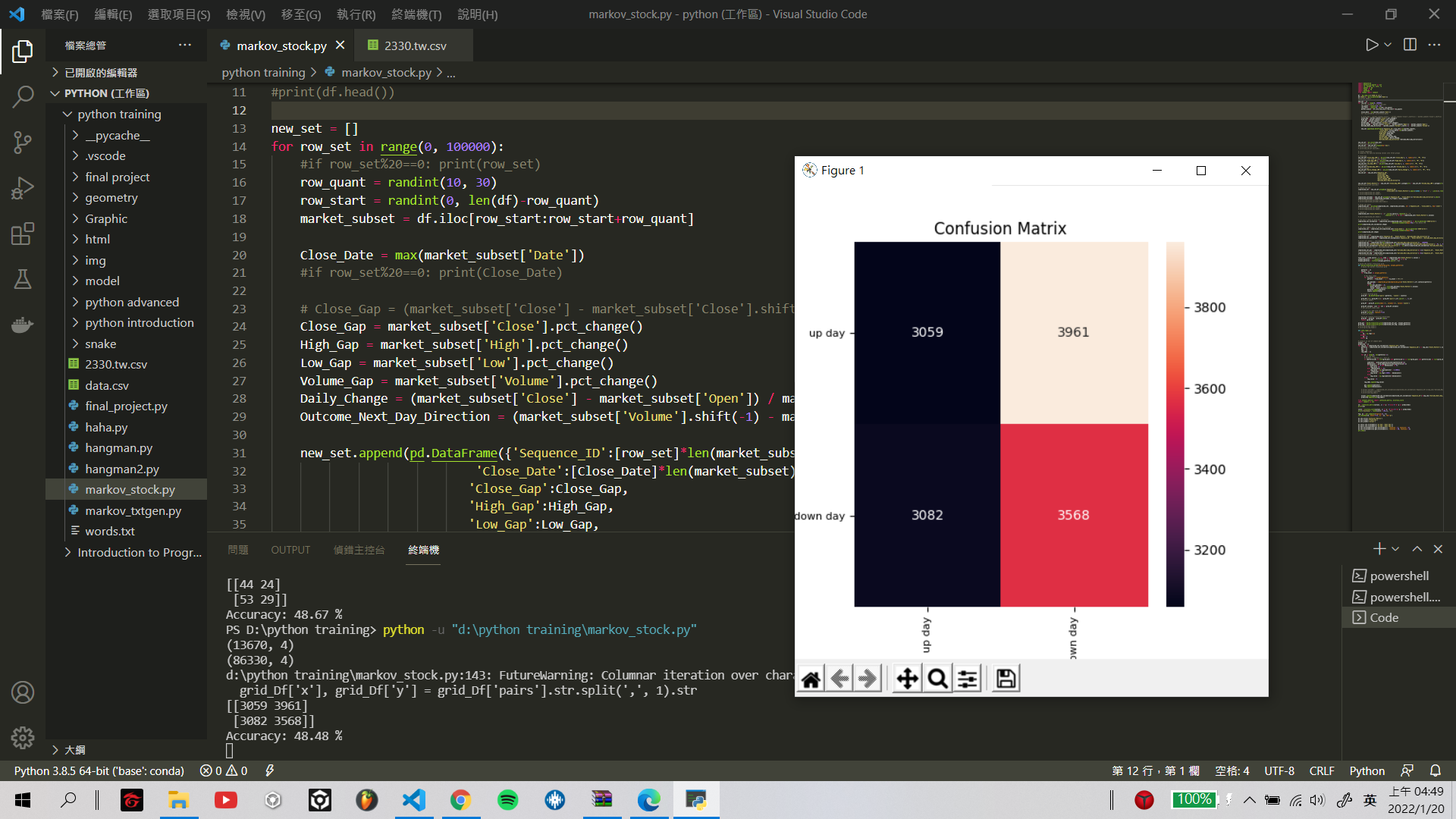


最後就會分成這兩類，上面是交易變動量為正，下面則是為負。



這裡的from\_event 跟 to \_event代表著上一個狀態到下一個狀態，例如MMM到MLM。

這是最後做出的馬可夫矩陣。上面是交易量為正的結果，底下則是交易量為負的結果。



最後用validation和training兩個data比對出來的預測準確率和其confusion matrix，橫軸是預測資料預測出的交易量上升下降，縱軸則是實際上的。