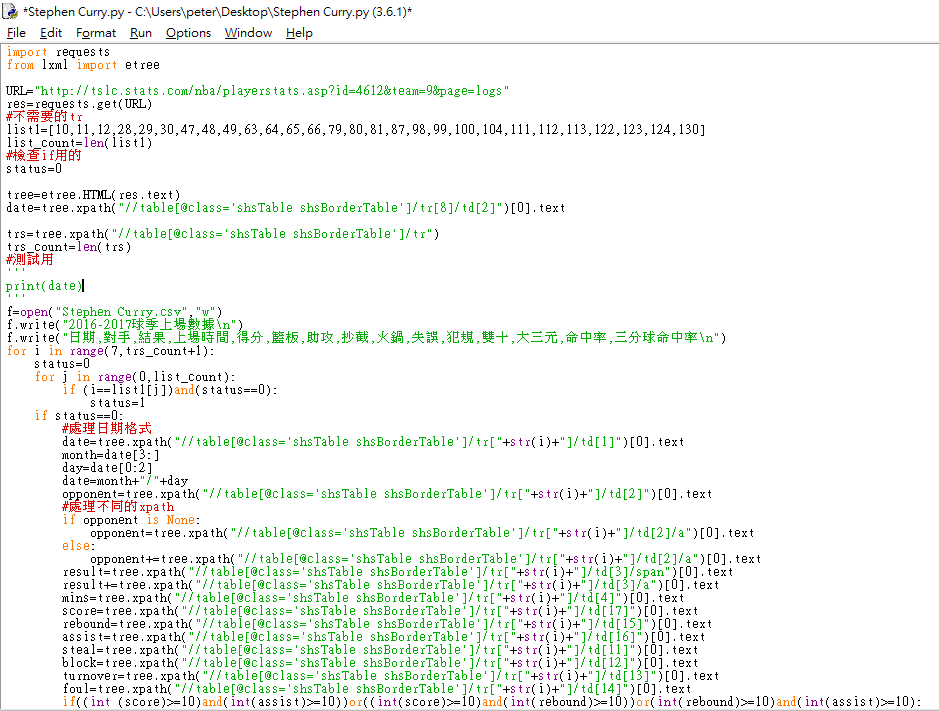
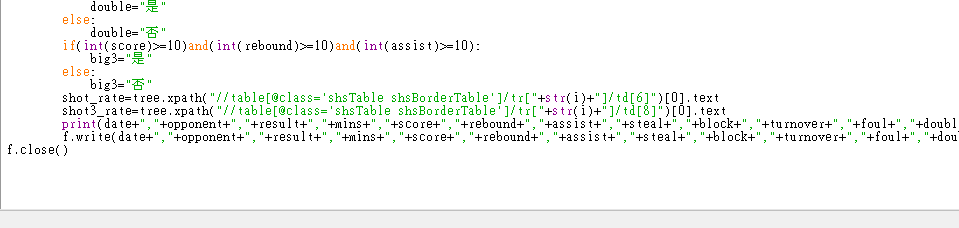
資管二甲 B0444132 張仁樵

目標:抓取Curry 2016-2017球季的數據

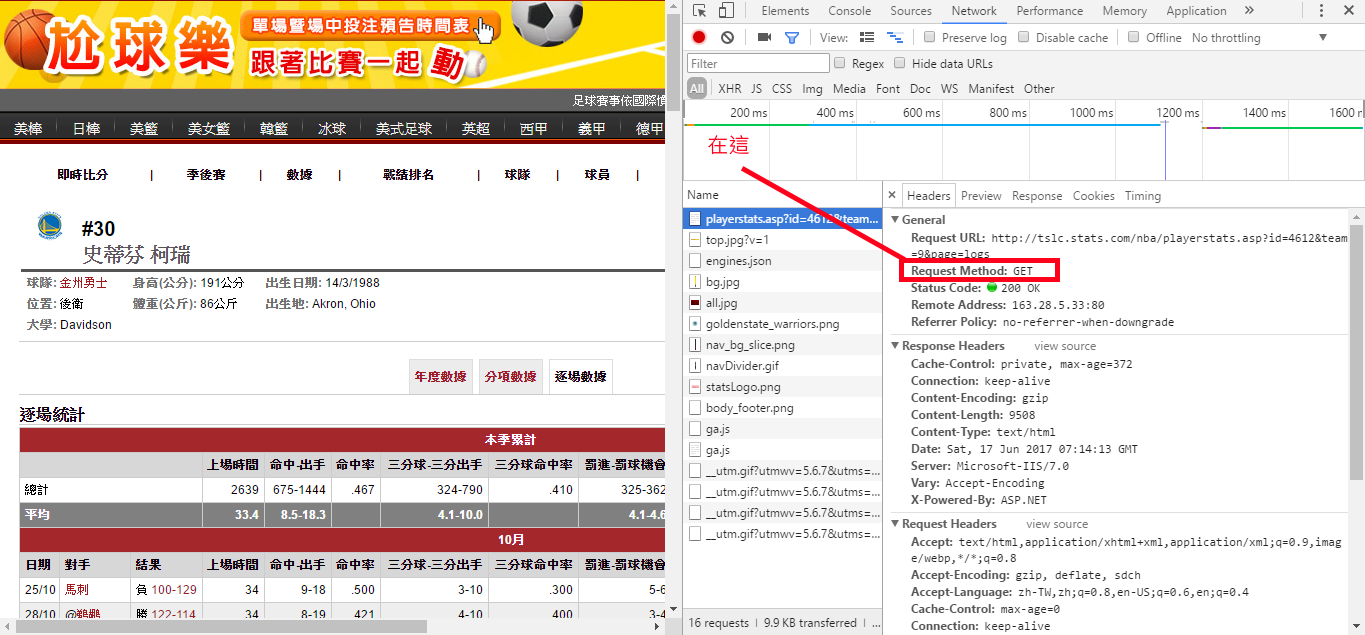




1.首先利用import語法，抓requests和etree套件。

2.設URL變數為所要抓取資料的網址。

3.開啟網站的開發人員工具(按F12)中的network看是get還post。



4.使用requests中的get()取得網頁，並放入res變數中。

5.設list1陣列為不需要的行數，設list\_count為list1的長度，設status為0(後面for迴圈中的if會用到)。

6.使用etree中的HTML()取得網頁的html的內容，並放入tree變數中。

7. 設date變數進行x-path語法的測試看能不能抓到資料(若為None代表不是文字型態的東西，但路徑是對的)。

8. 設trs變數為x-path中的table的tr，並利用trs\_count表示為trs的長度(即有幾行)。

9.開啟一個名叫Stephen Curry的.csv檔(用Write模式，無論有沒有重複的檔案都會自動覆寫過去)。

10.用write()語法先寫入第一行為2016-2017球季上場數據，以及第二行為日期,對手,結果,上場時間,得分,助攻,抄截,火鍋,失誤,犯規,雙十,大三元,命中率,三分球命中率( ,表示Excel中換到右邊的一格)。

11.接著使用for迴圈(從7開始)抓取每一行所需的資料，並利用和一個小的for迴圈裡面包含一個if判斷式去判斷不需要的行數。

12.通過第二個if的驗證後，代表為我們需要的資料可以進行處理。

13.首先因為網站的日期格式為 日/月，但Excel的格式為 月/日，所以利用字串擷取進行處理，變為我們所需的格式。

14.接著因為對手會分是主場還是客場(會多加@)，所以要if判斷式進行不同x-path的處理。

15.另外因為我想知道Curry是否有雙十或大三元，所以多寫了兩個if判斷式進行判斷。

16.在for迴圈中最後使用write()語法把每一次抓取到的資料放入到csv檔中。

17.最後用close()關閉.csv檔。

執行結果如下(.csv檔會和.py檔在同一個目錄下):

