天氣資料概述：

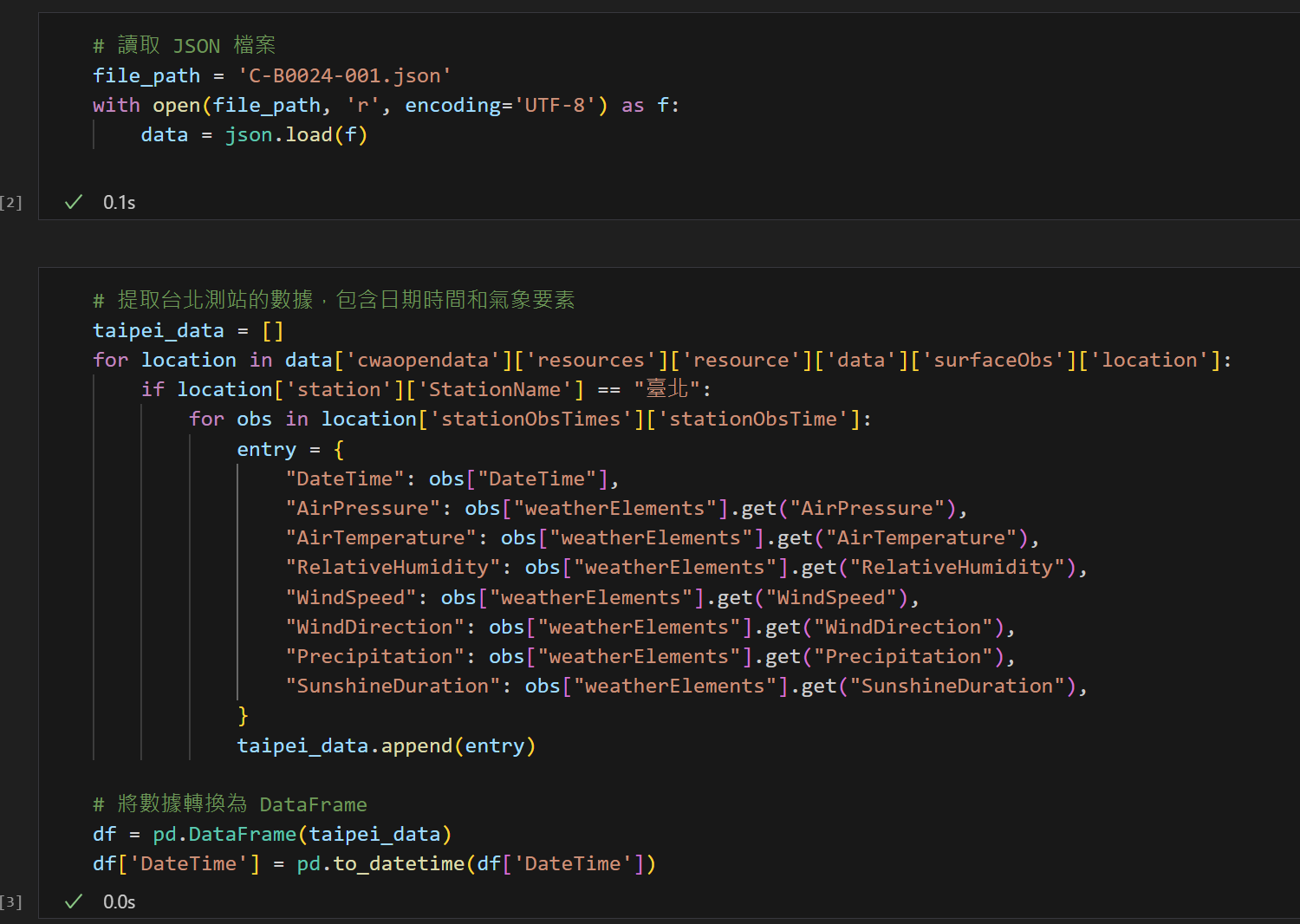
名稱：地面測站最近30天觀測資料-30天觀測資料（以小時為單位）

網址：<https://opendata.cwa.gov.tw/dataset/climate/C-B0024-001>

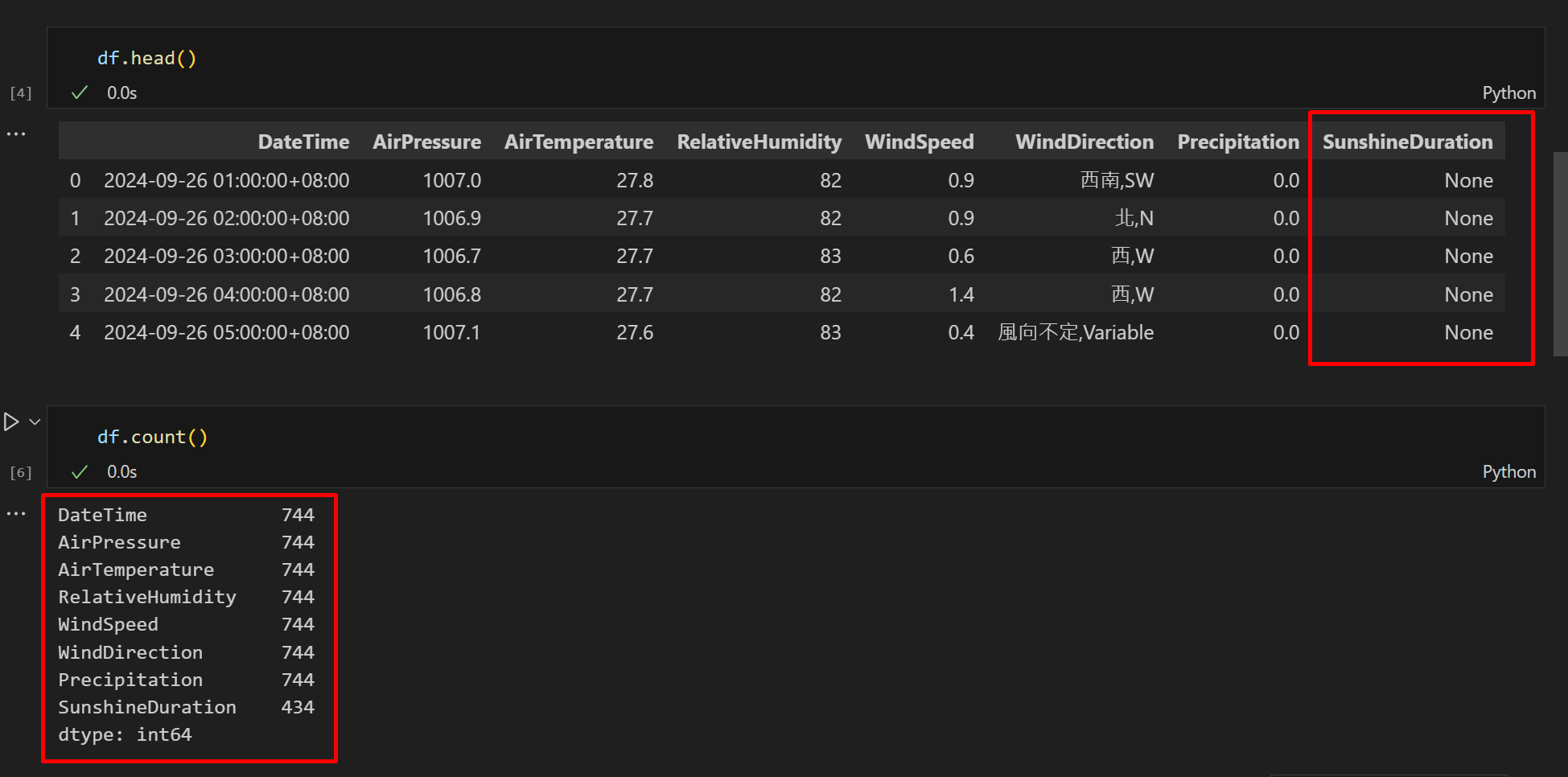
主要欄位：觀測時間,測站氣壓,溫度,相對濕度,風速,風向,降水量,日照時數

目的：想要做Youbike數量預測，希望可以有天氣資料，讓預測模型更精準

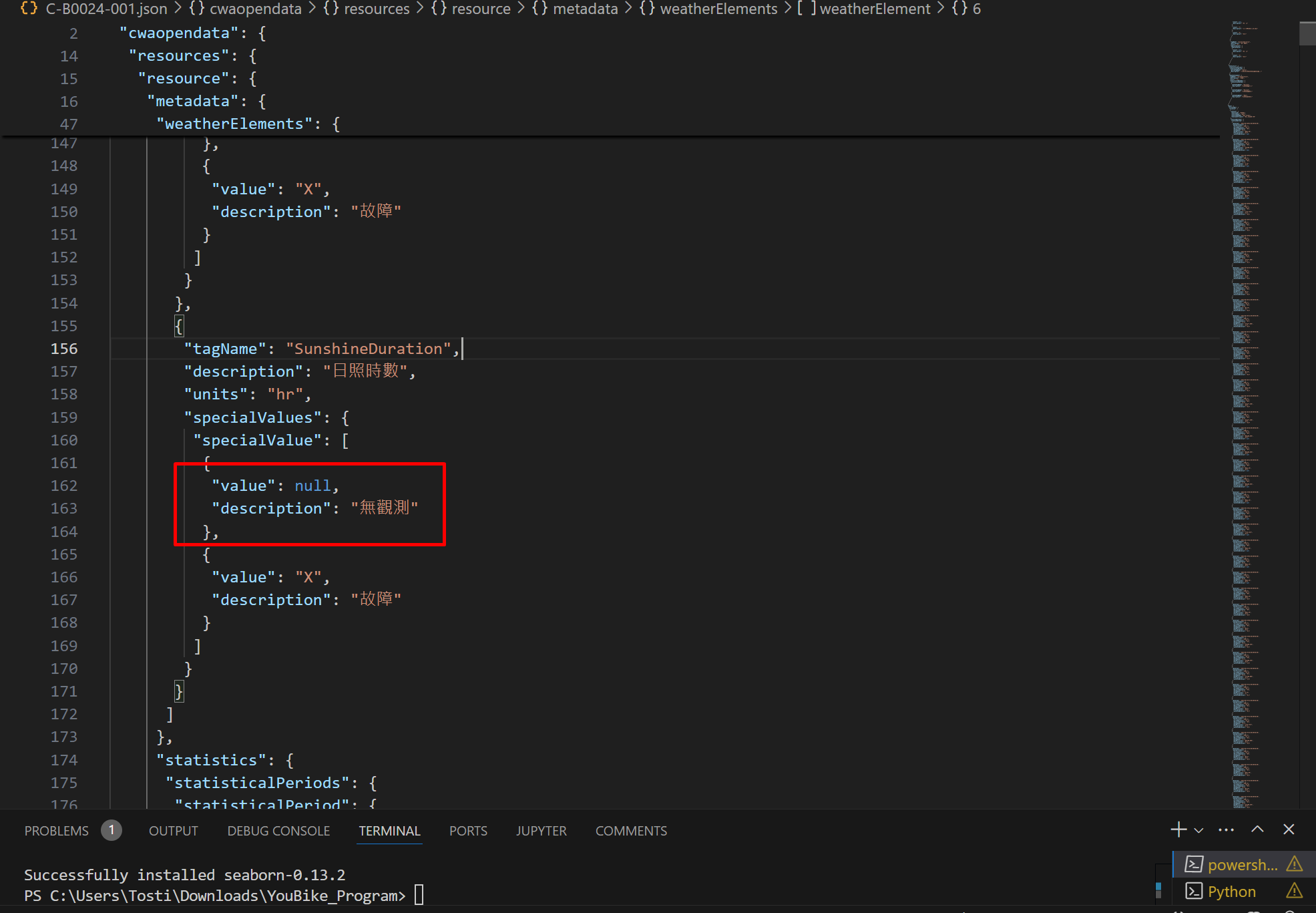
讀取檔案並將資料轉換為DataFrame（只取臺北市測站的觀測資料）



資料觀察，發現日照時數欄位有許多空值

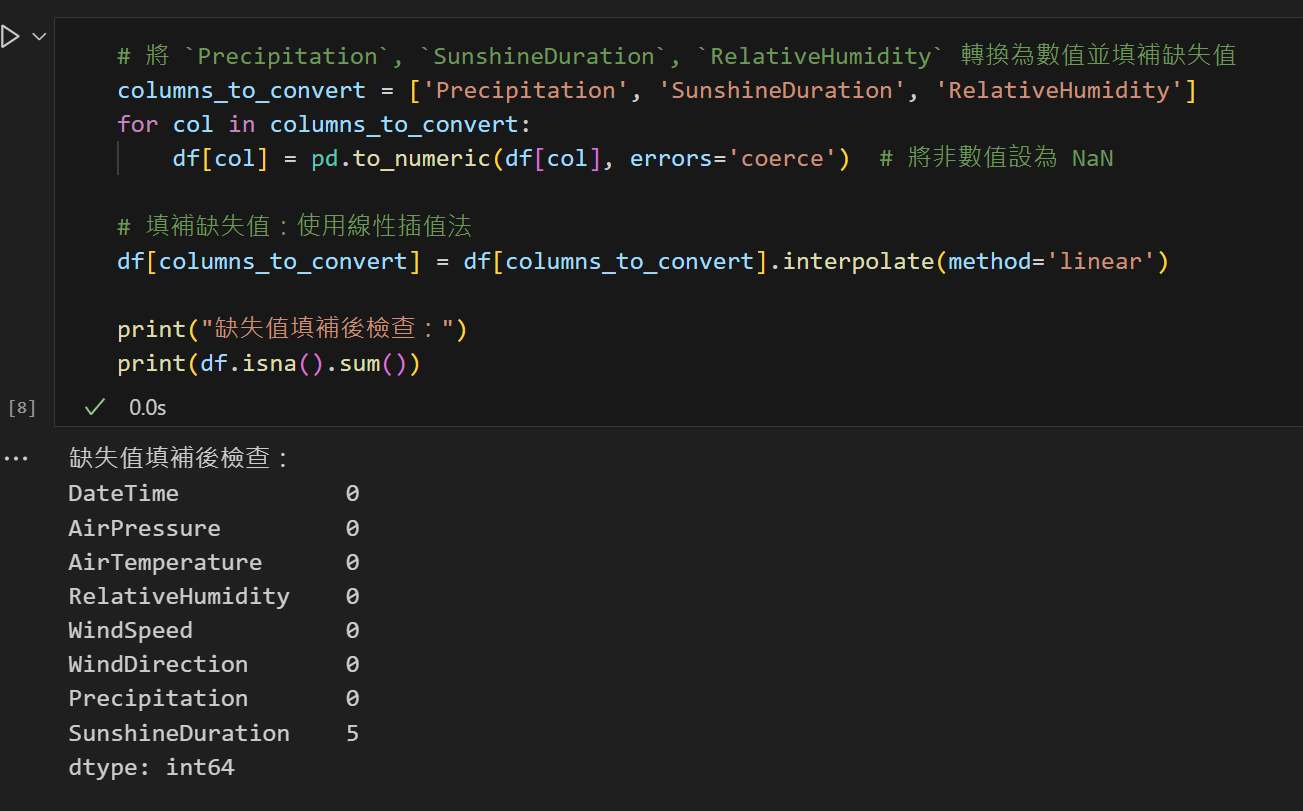


調查資料，發現是觀測測站故障，因此無觀測



對缺失值做填補處理，感覺補零不太恰當，問ChatGPT有何填補方法

他建議連續的資料可以使用線性插值法填補空缺，但不知為何沒有填補完全



ChatGPT：可能是因為這些缺失值位於數據的開頭或結尾，導致線性插值無法對這些位置進行填補。線性插值只會填補內部的缺失值，而不會處理位於開頭或結尾的 NaN 值。

又學到一課啦！

於是向前後填補，這時欄位確實填補完全了



依據以下的標準來推估天氣狀況：

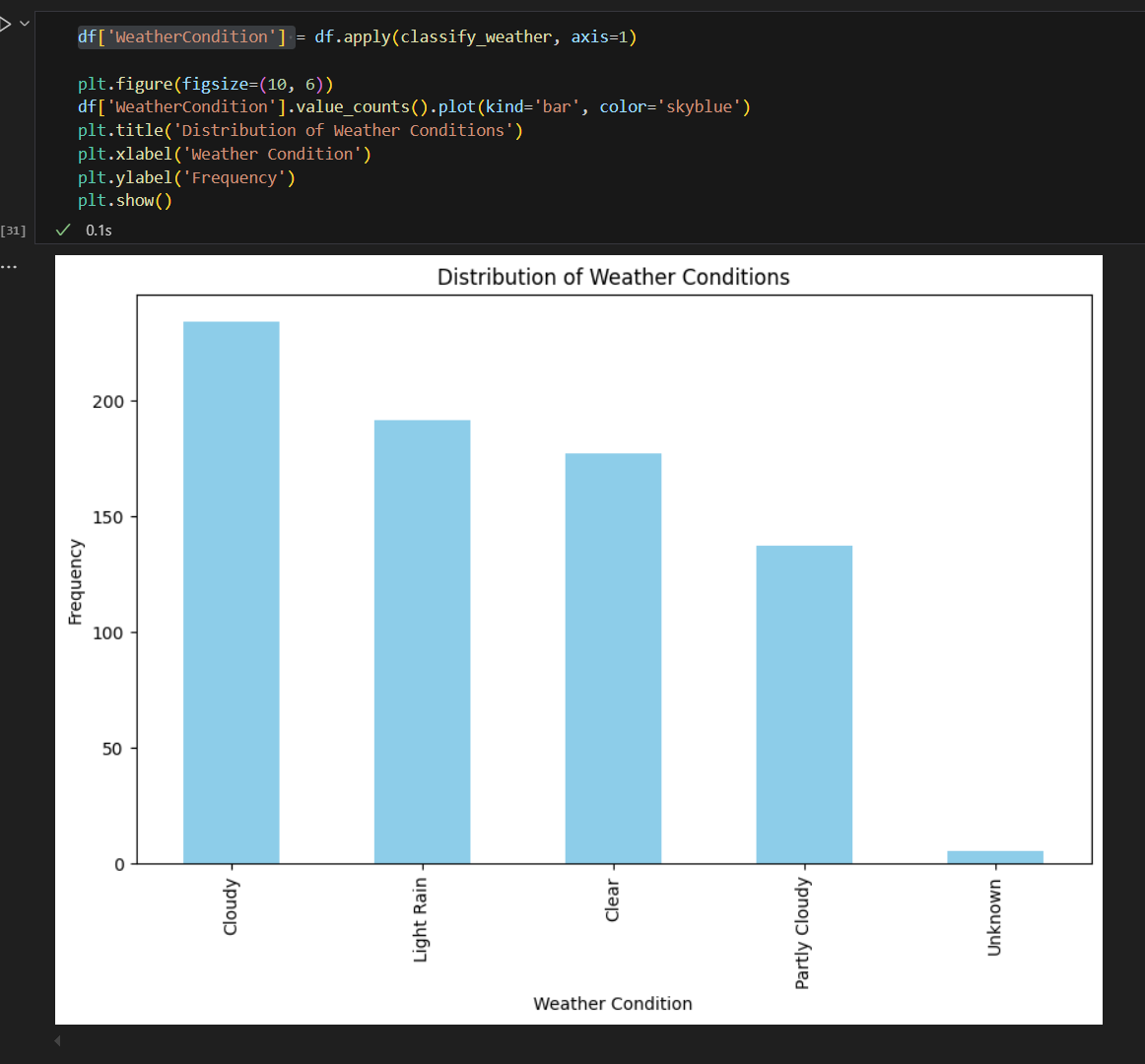
晴天 (Clear)：當降水量為 0 且有日照時數。

多雲 (Partly Cloudy)：當降水量為 0，日照時數為 0，相對濕度小於 80%。

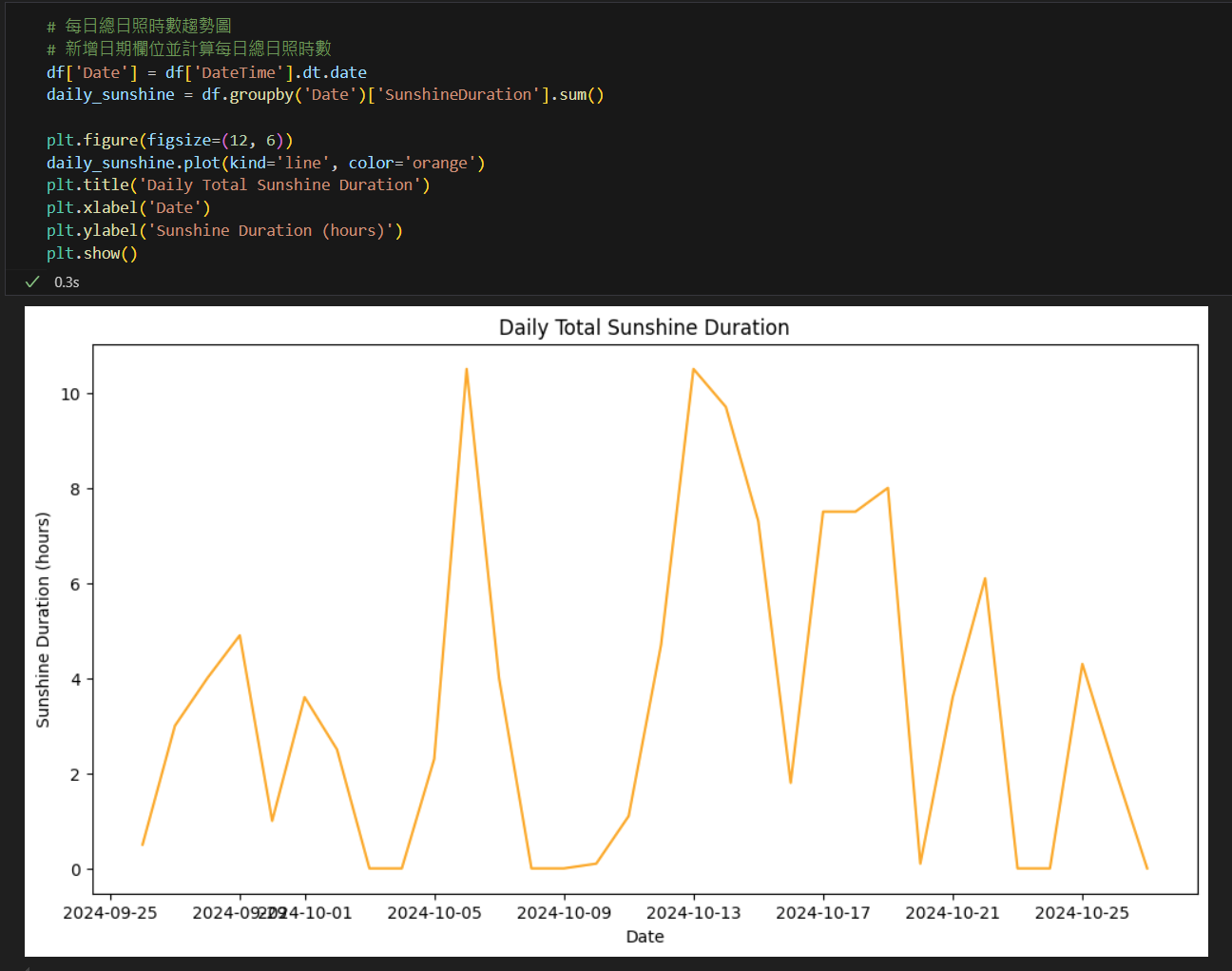
陰天 (Cloudy)：當降水量為 0，日照時數為 0，且相對濕度超過 80%。

小雨 (Light Rain)：當降水量介於 0.1mm 和 10mm 之間。

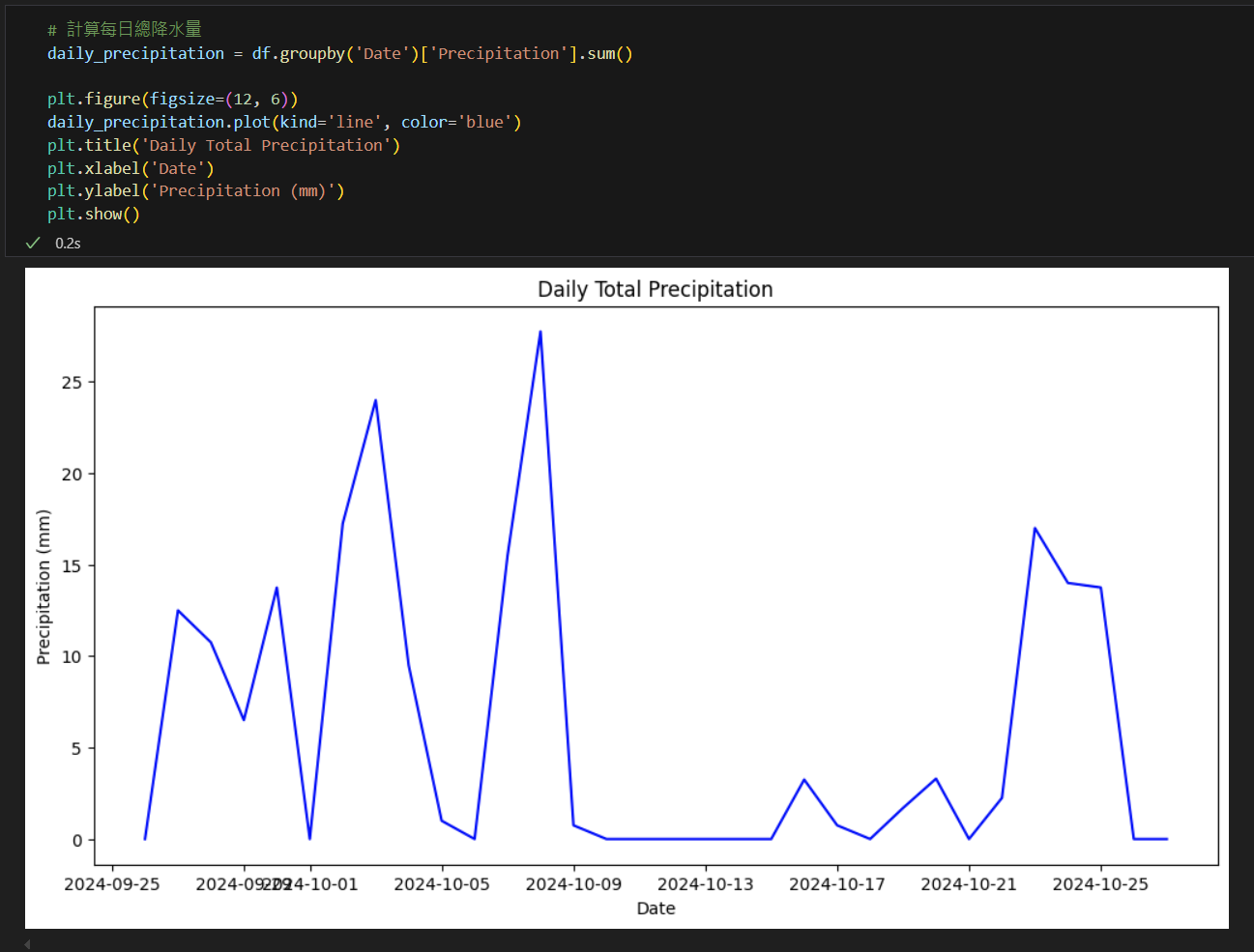
大雨 (Heavy Rain)：當降水量超過 10mm。

簡單資料視覺化一下  


來看一下每日總日照時數趨勢



每日總降雨量趨勢圖



將兩趨勢疊合看一下關聯性



這時突然觀察到一個現象，雖然降雨量與日照時數看得出關聯，但為何日照時數的變化幅度會那麼小呢？明明這一個月內也有許多大晴天的日子？

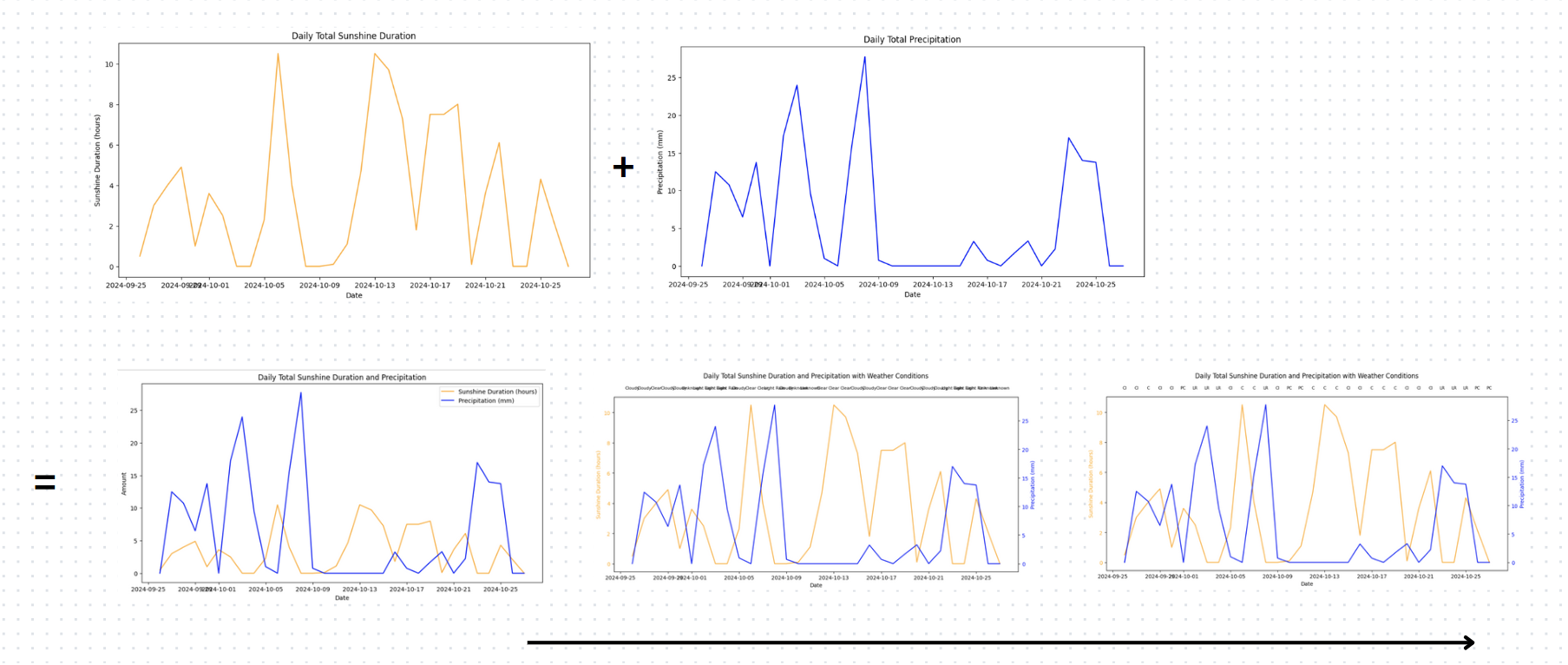
這時才發現單位不同共用縱軸的影響求助了ChatGPT，學了如何畫雙縱軸圖：



上方標籤文字過長遮擋到彼此，將標籤簡寫一下：



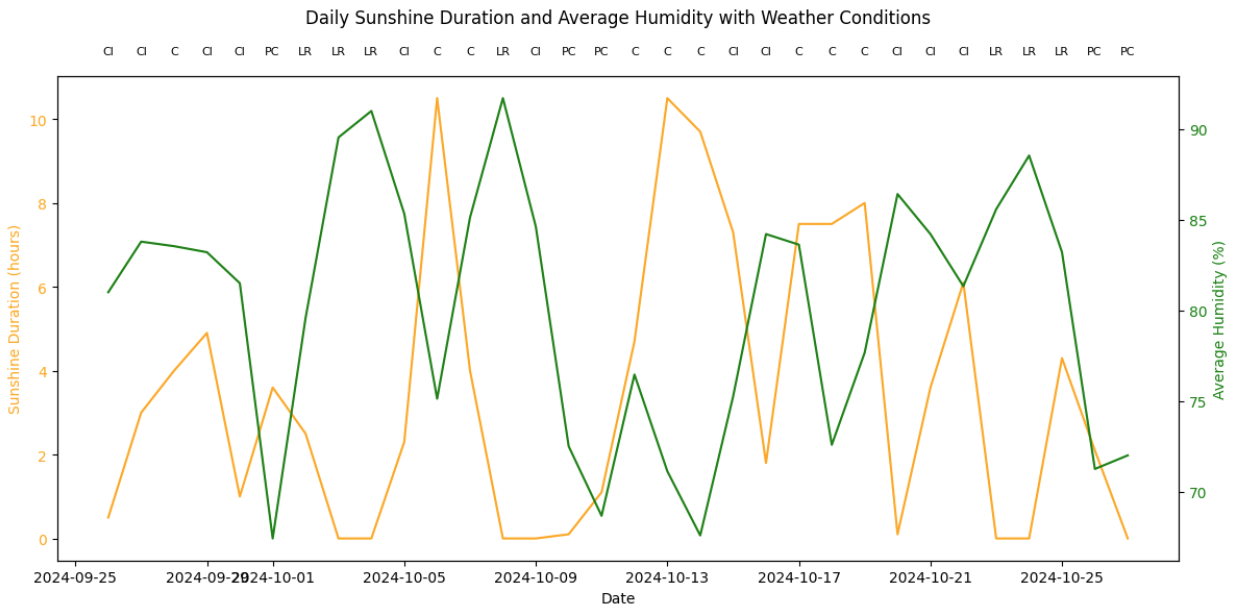
日照、降水量趨勢圖過程總結：

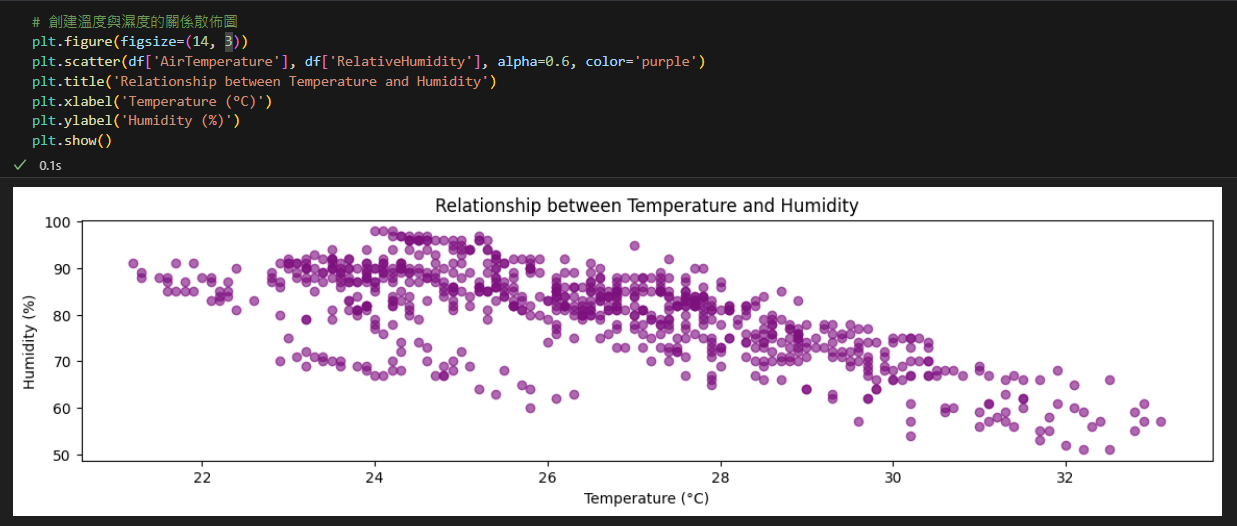


濕度與溫度的關聯性：

程式碼類似上面展示的範例，就不多做贅述了。

可以看出濕度與日照時數的關聯性



溫度與濕度散佈圖，大致能看出溫度越高濕度越低。  


結語

最後想透過YouBike資料，探討天氣狀況與腳踏車借還狀況的關聯性，但可惜Youbike的資料我們才開始儲存沒幾天，沒有能跟這三十天天氣資料對應的Youbike資料，因此只能先做天氣部分的資料清洗及簡單統計及視覺化的程式碼，等到Youbike資料收集三十天後再加入分析關聯性，若關聯性夠高的話，在進一步加入預測模型，提升預測腳踏車樹功能的精準度及成效。

\*現在正在動態將Youbike資料收集進Database

