感謝助教細看完每一份腳本 以及說明手冊

我也當過教學助教 可以理解這樣的辛苦程度

(雖說我有點不知道這次程式說明要寫甚麼xd

1. 匯入套件

一張含有 文字 的圖片

自動產生的描述

就是匯入套件xd

好啦 其實這邊先匯入了後面會使用到的功能

import equests 是之後為了要使用https/http 請求

import json是因為python 不是javascript 如果用到 json data的 parse 或是 將object轉成json格式儲存可能會用到

除非你要自己刻啦 我偶而會自己,load & dump 有點慢

Import time 是我之後有讀取系統時間,所以也會用到

2.觀察網站

一張含有 文字 的圖片

自動產生的描述

現在的網頁不是1995年當時的純靜態網站,

當使用者一開讀取一個網頁時候

該網頁可能因為前後台分離 因此向後台用http要了一份使用者所求的資料

後臺給錢台的資料通常對程式其實比較友善 比較不用清除太多不必要的資訊

所以如果能用開發者模式直接找到前台fetch的資訊

那是最快的,可以省一堆字串處理

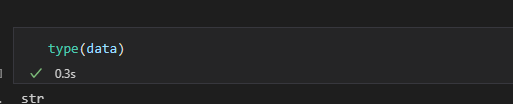
這邊就直接利用了requsets的套件,對costco後台要了一份 泡麵貨物清單的

也利用time.ctime() 直接紀錄下 爬的時間

這份程式碼的所以商品資訊都源自於一份html檔案 所以抓取時間都是相同的

一張含有 文字 的圖片

自動產生的描述



觀察一下資料並且確認一下資料型態

大概知道所求的資訊都是”products”下面的list

其實這次沒真的觀察 (因為有助教及老師細心幾乎所有的流程

通常大多自己會轉一份成json檔案

利用vscode的json外掛工具來釐清我要求的資料,路徑是甚麼

一張含有 文字, 螢幕擷取畫面, 監視器, 黑色 的圖片

自動產生的描述

感謝教授,助教協助我完成 剛剛我說的事情

這邊把str parse成 object型態

利用前面導入的函示庫 “json”

一張含有 文字 的圖片

自動產生的描述

利用迴圈查看前10筆資訊 驗證一下

Key是否是正確的

一張含有 文字, 監視器, 螢幕擷取畫面, 螢幕 的圖片

自動產生的描述

由於csv只是簡單的 逗號分割文件

所以我也沒用其他third party的轉檔工具

就用直接用逗號切一切就匯出成檔案了

Big 5是 Microsoft Excel預設的編碼格式

如果用utf8會需要處理編碼問題

可是現在大部分的文字文件都用utf8 所以我兩種都存了

不知道你們比較喜歡哪一個