**教育大數據期末個人專案**

**Title:**利用Python爬蟲工具(Webdriver)抓取pchome資料並分析數據。

**ID:** 411077010 **Name:**蔡岳哲

**Abstraction:**

本專案旨在利用Python爬蟲技術（使用Webdriver）來抓取pchome網站上特定關鍵字的搜尋結果資料並分析數據。通過這個過程，我們將探討如何撰寫一個程式，並利用工具(Webdriver)進行數據的搜集、整理、分析。本專案將包含文章、程式、資料表等多個面向，並最終完成一個具有實際應用價值的數據分析程式之雛形。

**I: Preface:**

在教育領域，大數據的應用已經成為一個引人注目的議題。本專案旨在透過爬蟲技術，將網路上的數據資料收集起來，並通過數據分析提供有價值的見解。 **II: Related documents:**

本章節將探討與本專案相關的文件、研究、和先前相關專案的相關性，以建立本專案的理論基礎。

1. 使用的網頁爬蟲工具(webdriver): [chromedriver](https://chromedriver.chromium.org/)

2.Selenium教學: [Selenium教學](https://medium.com/marketingdatascience/selenium%E6%95%99%E5%AD%B8-%E4%B8%80-%E5%A6%82%E4%BD%95%E4%BD%BF%E7%94%A8webdriver-send-keys-988816ce9bed)

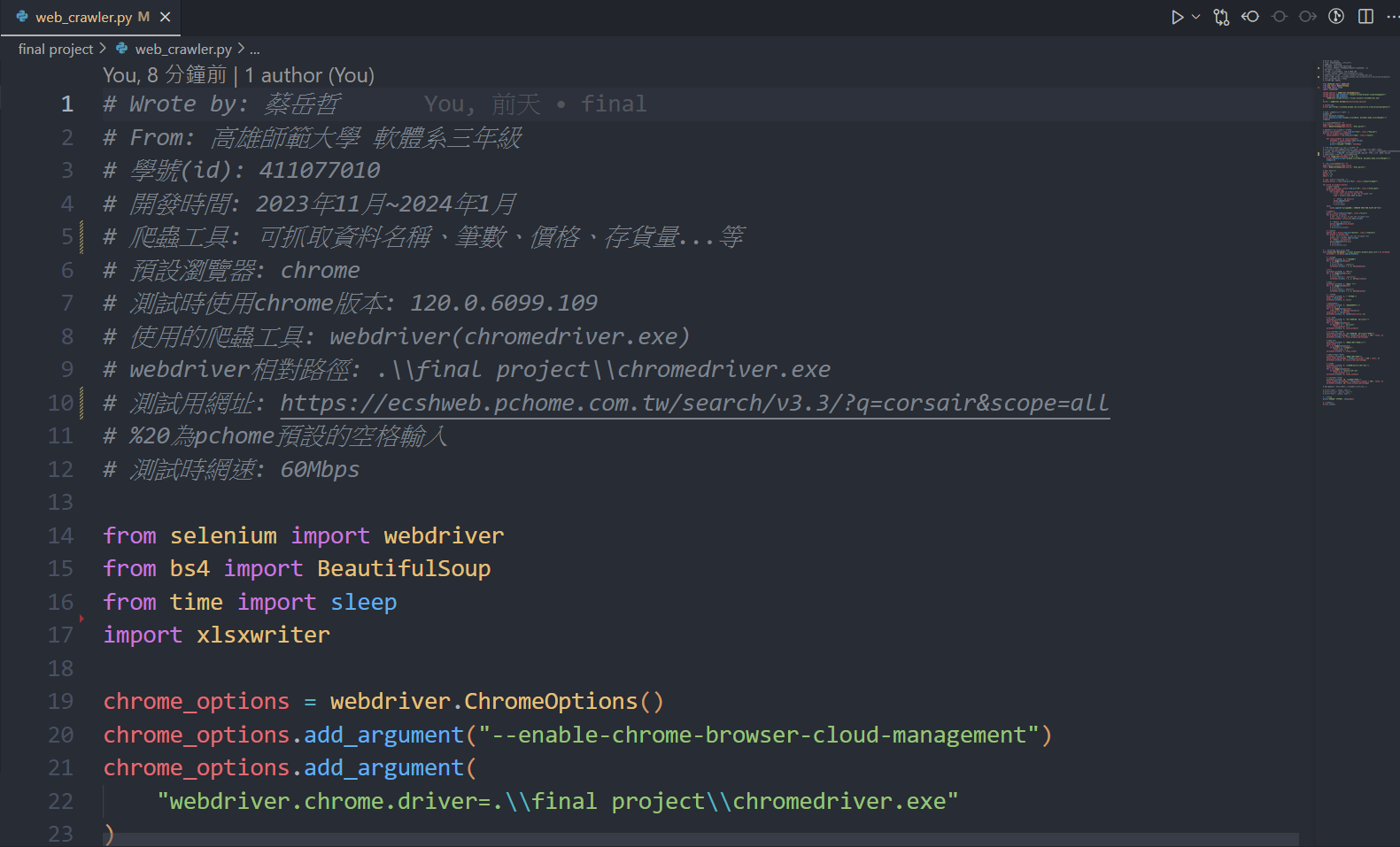
3. 預覽、測試用pchome網址:  [pchome搜尋](https://ecshweb.pchome.com.tw/search/v3.3/?q=corsair&scope=all)

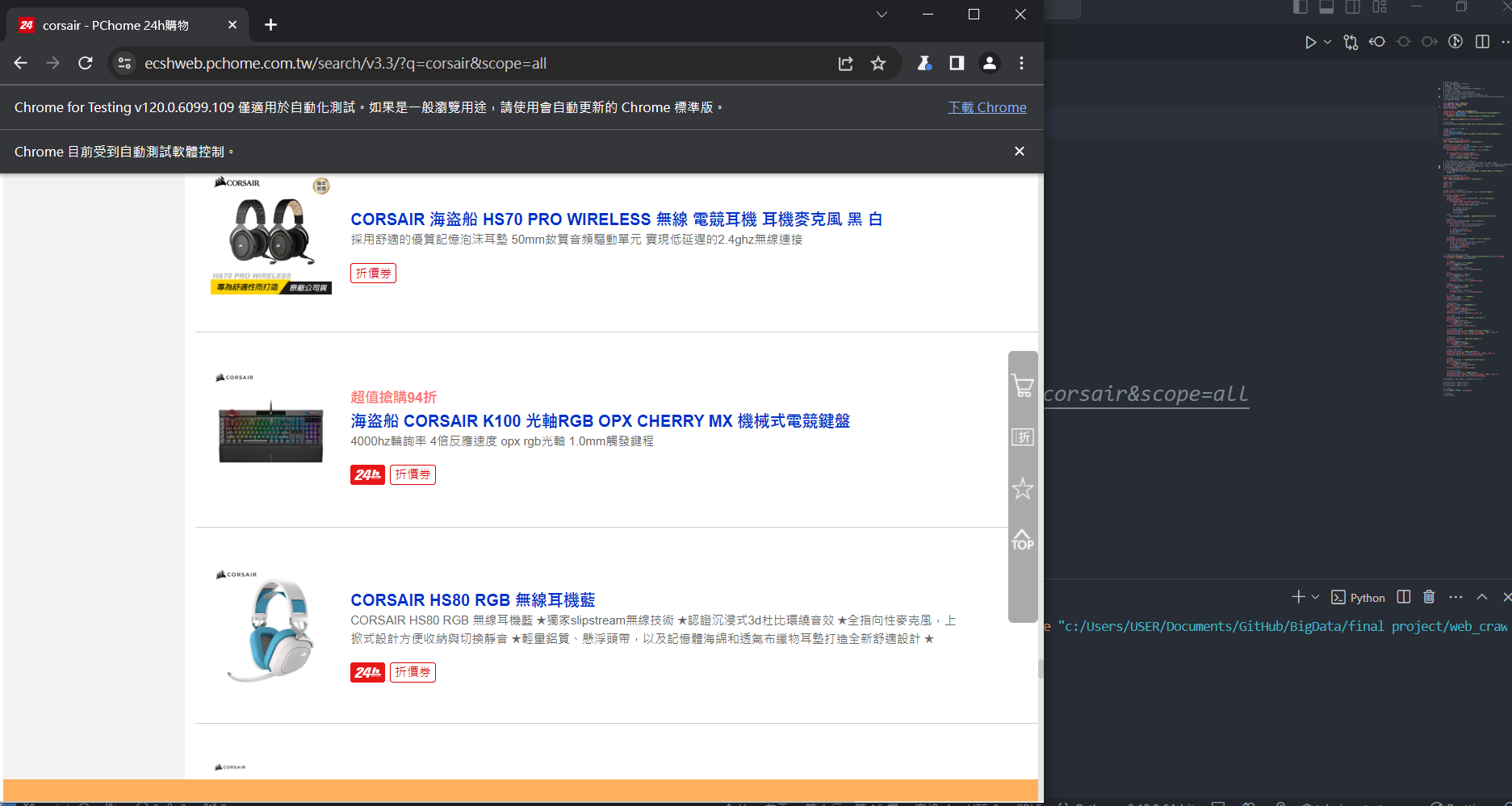
4. beautiful-soup教學:[beautiful-soup教學](https://steam.oxxostudio.tw/category/python/spider/beautiful-soup.html)

5. XlsxWriter教學: [XlsxWriter教學](https://pypi.org/project/XlsxWriter/I)

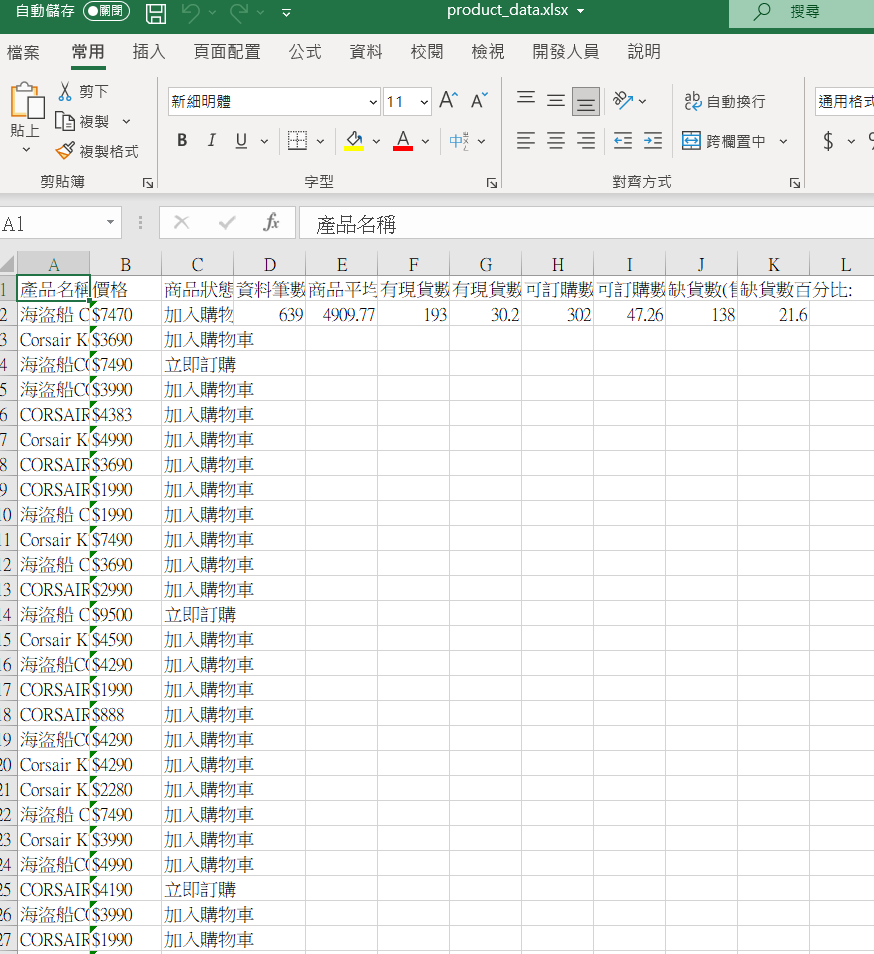
**III: Step**

**1.Article:**

(1) run code:****

(2) crawling web:

(3) data analysis by code(demo part of code): 

(4) convert into xlsx: 

**2.Program:**

使用Python爬蟲技術，利用Webdriver，從定義爬取的目標網站抓取所需文字與資料，透過程式編寫，實現自動化的數據收集與分析。

**3.Hardware:**

考慮到爬蟲可能需要長時間運行，以下將討論適用的配置，確保順利執行整個數據搜集過程。

*# 預設瀏覽器: chrome*

*# 測試時使用chrome版本: 120.0.6099.109*

*# 使用的爬蟲工具: webdriver(chromedriver.exe)*

*# webdriver相對路徑: .\\final project\\chromedriver.exe*

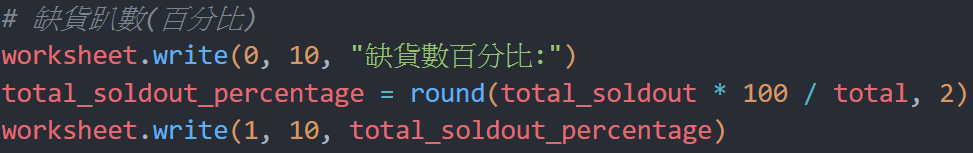
*# 測試用網址:* [*https://ecshweb.pchome.com.tw/search/v3.3/?q=corsair&scope=all*](https://ecshweb.pchome.com.tw/search/v3.3/?q=corsair&scope=all)

*# 測試時網速: 60Mbps*

**IV: System establishment(Data analyse):**

使用簡單的加法將資料總數得出來，利用除法將資料百分比分析出來。

以下為其中一個範例:

**  
V: Conclusion:**

最終，我們將總結整個專案，回顧所取得的成果、遇到的挑戰，可能的擴展方向和改進方法，以提升系統的效能和應用價值。(1)取得的成果:

可以將前端文件除去css、js等html渲染，並獲取指定的name、class中的資料。將獲得的資料分析並自動寫入到excel中，以便使用者直接了解數據。

(2)遇到的挑戰:

若網速不夠快，爬蟲的速度便會大大降低，所需時間會變多很多。若網站搜尋到的筆數太多，也會造成爬蟲時間過長，並且抓取到的資料到最後也可能與使用者所搜尋的無關。

(3)可能的擴展方向和改進方法:

可以利用其他工具，或是再次進行資料分類，將過於不符的資料去除，以得到更有用的總資料。

利用其他資料結構、其他工具來爬取資料，或許可以提升效能。

**##報告:**[**報告影片**](https://youtu.be/1wx68w_zjEs)