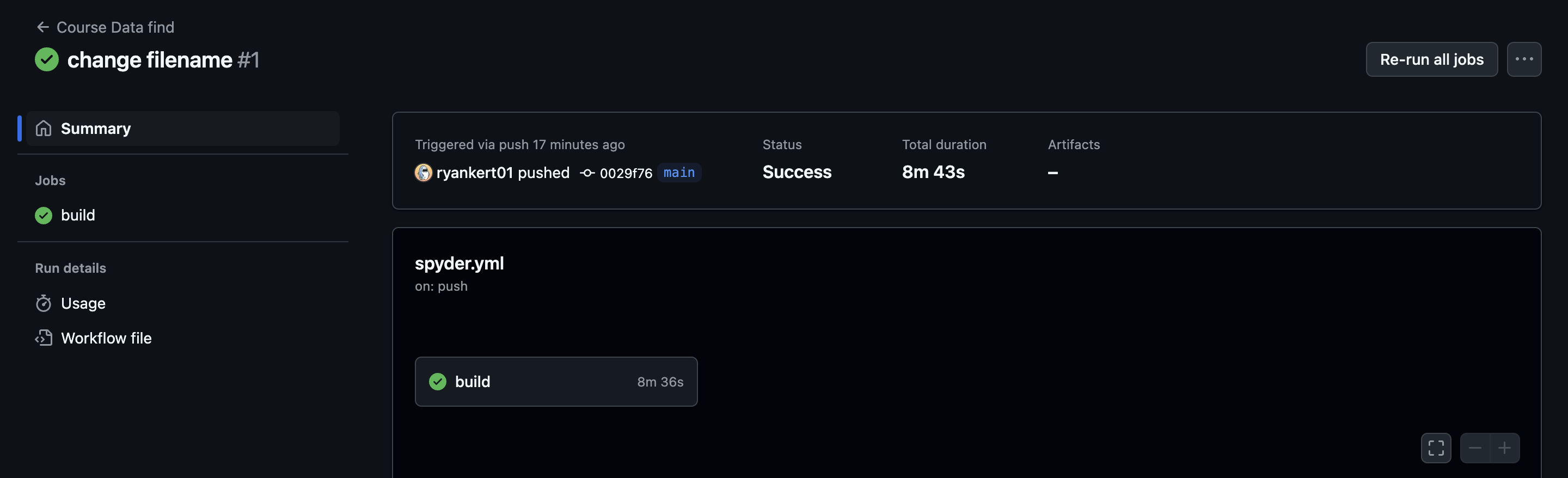
# OSS Midtern - 期中專題報告：網路爬蟲 （1103339黃咸誠)

# 1. 為何選擇該網頁、爬蟲程式概略說明、爬蟲JSON或CSV輸出結果說明

Github Link: <https://github.com/ryankert01/midterm-oss>

Colab Link: <https://colab.research.google.com/drive/1bzfU_tZLvTnLFP_jfDFKd3YfR5EgP4g4?usp=sharing>

## Github Action

Github Action 會執行完兩個爬蟲，再將結果存在gh-pages ，避免main branch太過臃腫。

## api.py - 新北市A1交通事故資料

***給助教的話***：執行時需要用portal帳密為環境變數，例如：

ACCOUNT\_TOKEN=portal帳號 ACCESS\_TOKEN=portal密碼 python static.py

選擇新北市政府開放資料平台中的102年度A1類交通事故資料，主要是因為A1類交通事故為死亡案件，具備重要的交通安全分析價值。此資料透過API提供，結構清楚，適合自動化爬取，並可應用於地圖視覺化與交通熱點分析。程式會自動分頁，將所有資料擷取下來，並轉存為api.csv，包含事故時間、地點、死亡人數、受傷人數與肇事車種等資訊。

**爬蟲輸出結果**為api.csv表格資料，包含date、location、dead person、hurt person、field3等欄位。

一張含有 文字, 收據, 螢幕擷取畫面, 數字 的圖片

AI 產生的內容可能不正確。

## static.py - 元智大學課程資料

選擇元智大學選課系統進行資料擷取，主要原因為該系統需要登入，適合練習登入驗證、表單模擬提交等爬蟲技術。資料包含所有系所與學期的完整課程內容，程式透過遍歷學期與系所，擷取課程名稱、時間、授課老師、英語授課標記等資訊，最終整理為static.json，方便進行熱門課程統計、時間衝突檢查與英語授課比例分析。

**爬蟲輸出結果**為static.json多層巢狀資料，依學期與系所儲存課程詳細內容。一張含有 文字, 螢幕擷取畫面 的圖片

AI 產生的內容可能不正確。

# 2. 交通事故熱點地圖應用規劃

## 目標

利用新北市102年度A1類交通事故資料，透過地理座標轉換與地圖視覺化，製作交通事故熱點地圖。此地圖可協助民眾了解高風險路段，並提供政府作為交通改善參考。

## 方法

資料清理location欄位，利用Google Maps API或OpenStreetMap Nominatim進行Geocoding，取得經緯度座標。再透過Folium或Leaflet進行視覺化，標記所有事故點，並依事故次數或死亡人數顯示熱區。

## 預期成果

製作出包含新北市102年度所有A1類交通事故的互動式地圖，支援點擊查看細節、事故熱區顯示與互動查詢，提升資料可視性與交通安全意識。

### 遇到的問題與解法

1. **地址格式不一致或不完整，影響地理編碼**：  
   在資料中，有些地點欄位存在格式不統一或資訊不足的情況，這會導致無法成功進行地理編碼（Geocoding）。解法是先針對原始資料進行前處理，統一地址的格式，例如移除多餘文字、補齊新北市或各行政區名稱。此外，將所有地址資料進行標準化，最後再批次進行座標轉換，以提高轉換成功率。
2. **部分地址無法成功轉換為座標**：  
   即使經過前處理，仍有部分特殊或模糊的地址無法被自動解析。解法是嘗試使用不同的 Geocoding 服務，例如 Google Maps API 與 OpenStreetMap Nominatim API 互相搭配使用。如果自動轉換仍失敗，則可透過人工查詢或比對其他資料來源，確認正確座標，並將這些特殊案例記錄於失敗清單中，以方便後續人工修正或特別處理。
3. **Google Maps API 有使用次數或費用限制**：  
   Google Maps API 每日免費使用次數有限，超過後需要支付額外費用。解法是優先使用免費的 OpenStreetMap Nominatim API 進行座標轉換，僅在必要時才使用 Google Maps API。此外，透過程式紀錄已成功轉換的地址與座標，將結果存檔，避免重複查詢，降低 API 使用量與成本。
4. **多筆事故發生在相同座標，導致視覺化時資料重疊**：  
   有些交通事故發生在同一個地點，若直接將所有資料標記在地圖上，容易造成重疊無法點擊或辨識。解法是採用「聚合標記」（Cluster Marker）技術，當地圖縮小或事故點數量多時，自動將附近的標記合併成一個群組，使用者放大地圖後可查看個別事故點。另一種方法是利用「熱區圖」（Heatmap）方式呈現，透過顏色深淺來顯示事故的密集程度。
5. **資料視覺化效果不足或操作不便**：  
   如果僅以靜態地圖顯示，資料的互動性與使用者體驗較差。解法是導入 Folium 或 Google Maps 的互動式地圖功能，讓使用者可以自由縮放地圖、拖曳瀏覽，並可點擊每個事故點查看詳細資料，例如發生時間、死傷人數與事故車種。此外，也可以增加篩選或搜尋功能，方便使用者依地區或時間查詢事故資訊，提升操作的便利性與地圖的實用價值。

## 參考資料

1. 新北市政府開放資料：<https://data.ntpc.gov.tw/>
2. Google Maps API：<https://developers.google.com/maps/documentation/geocoding/start>
3. OpenStreetMap Nominatim：<https://nominatim.openstreetmap.org/>
4. Folium：<https://python-visualization.github.io/folium/>
5. Leaflet：<https://leafletjs.com/>