

Python and Data

Week 1

Keynotes

<https://pydata.taipa.dev>

為甚麼是 **Python**?

- 易學
- 容易找到參考資料
- 豐富的 Library
- 但是並非非 Python 不可 (JavaScript / Ruby)

運行 **Python** 的方法

- 使用 `python(.exe) script.py` 執行 (Windows, macOS, Linux...) - Command line script, GUI
- 在伺服器上執行 (通常使用 Framework 如 django) - 建立網站
- 在 iOS / iPad 上執行 (Pyto, Pythonista) - 可執行 Shortcut, Widget
- 在 IoT 設備上執行 (使用 micropython) - 記錄感應器資料或執行動作
- 在 SaaS 上執行 (如 replit.com) - 編寫 Bot
- 在任何地方執行...

課程目標

- 透過使用 Python 解決生活上的問題
- 輸入－處理－輸出 (Input-Process-Output, and repeat)
- 輸入(讀取及解釋)資料
 - 通過讀取格式化的資料檔案 (CSV, XML, JSON)
 - 通過 Application Programming Interface (API) 取得資料
 - 通過 Scraping 抓取原始資料
 - 自動化的方法瀏覽網頁取得資料

- 處理(計算及整合)資料
 - 解譯原始資料 (Parsing)
 - 組合、計算結果 (Panda, Numpy)
 - 處理文字資料
- 輸出(整理及可視化)資料
 - 繪畫圖表 (Plot)
 - 將資料輸出至檔案 (TXT, DOCX, XLSX 等格式)
- 自動化測試及驗證結果
- 處理例外情況

課堂型式

- 透過範例方式講解
- Weekly Challenge

Data — 資料

來源：

- 人手記錄 (如：統計局、平台)
- 自動化方式記錄 (如：感應器記錄溫度、濕度等)
- 通過抓取網上資料 (Scraping)
- 開放 API

注意使用授權、私隱以及道德問題，如有疑問應徵詢資料持有人的同意。

資料的重要性

- 我們可以通過資料得知一個人（或一群人）的喜好、行為甚至健康狀況
- 各行各業都需要數據，如零售、教育、公共事業、衛生領域
- 大數據 — 當數據的量足夠大，所產生的效果更加大

(不太好)的例子

- Google - 關建字 \$\$\$\$\$

可以通過人搜尋的關建字得知最多人搜尋的產品，它可以將資料提供給零售商，零售商可以在目標產品的搜尋結果賣廣告。

- Facebook - 個人資料及喜好 \$\$\$\$\$\$\$

因為需要登入，它可以為每個人建立一個檔案，以記錄每個人的行為，分析他們的喜好。

- 智能手錶 - 大量健康資訊

智能設備如手錶，所獲得的資料更加個人化，甚至可以獲得當事人不知道的一些健康資料。

但是，這些健康資料亦可以作為科研的數據。

所以很多產品和服務都是免費，因為由你提供的數據價值遠大於他們提供服務的成本

Visual Studio Code & Jupyter Notebook