

## Python 資料科學實務 Homework #4

- 繳交日期：2023/05/28，午夜 12 點前
- 繳交方式：
  - 本次作業請直接上傳至本課程 Moodle 數位教學平台(請先自行登入系統)
  - 繳交期限一到不再接收任何檔案。
  - 上傳前請先建立一個檔案夾，檔案夾名稱改為你(妳)的學號(例如 s00001)，並將所有檔案放入此檔案夾內，並將檔案夾以 ZIP 方式壓縮後上傳。
  - 在作業繳交截止時間以前，可以重複繳交。
- 作業規範：請務必自己練習寫作，可相互討論，但不可抄襲！
- 作業題目：擷取網頁資料練習
  - 請以老師所給的上課程式範例網路頁面資料擷取為基礎，延伸撰寫一個可以擷取任意指定相關物件頁面的程式。
  - 作業主題：
    - 到【[聯合新聞網](https://udn.com/news/breaknews/1)】的新聞【[即時列表](https://udn.com/news/breaknews/1)】網頁(<https://udn.com/news/breaknews/1>)上，程式會自動蒐集擷取所有在【精選】、【要聞】、【娛樂】、【財經】、【文教】分項主題新聞網頁內的新聞標題。



- 本作業以 **Jupyter Notebook** 文件來完成。
- 續上述 (1) 中所提及的要求，要將所有搜尋到的即時新聞結果，以 **Pandas DataFrame** 的方式整理收集後呈現在 Notebook 上，每一個主題單獨以一個表

格來呈現。每一個主題的即時新聞網頁網址(實際是部分網址字串)，請以**正規表達式 (regular expression)** 的技巧擷取出來，以防日後報社更改每一個即時新聞主題所對應的網址。

```
><script>...</script>
><div data-options="{ 'autoplay': true, 'autoPlayTimer': 10000}" class="udn-slider " style="background-image: none;">...</div>
▼<section class="thumb-news cate-list context-box" id="breaknews">
  ><header class="context-box__header context-box__header--start">...</header> flex
  ▼<section class="cate-list subheader"> flex
    <a href="/news/breaknews/1/0#breaknews" class="active">精選</a> == $0
    <a href="/news/breaknews/1/99#breaknews" class>不分類</a>
    <a href="/news/breaknews/1/1#breaknews" class>要聞</a>
    <a href="/news/breaknews/1/8#breaknews" class>娛樂</a>
    <a href="/news/breaknews/1/7#breaknews" class>社會</a>
    <a href="/news/breaknews/1/3#breaknews" class>地方</a>
    <a href="/news/breaknews/1/5#breaknews" class>國際</a>
    <a href="/news/breaknews/1/4#breaknews" class>兩岸</a>
    <a href="/news/breaknews/1/6#breaknews" class>財經</a>
    <a href="/news/breaknews/1/11#breaknews" class>股市</a>
    <a href="/news/breaknews/1/7#breaknews" class>運動</a>
    <a href="/news/breaknews/1/9#breaknews" class>生活</a>
    <a href="/news/breaknews/1/13#breaknews" class>數位</a>
    <a href="/news/breaknews/1/12#breaknews" class>文教</a>
  </section>
  ><div class="context-box__content story-list__holder story-list__holder--full">...</div> flex
  ><div class="btn-holder btn-holder--center thumb-news__button thumb-news__button--nb">...</div> flex
```

作業配分： 程式執行正確性與完整性，參照上述 (B) 項說明，合計 100 分

注意事項：

- i. 以上只要未完成或缺少的部份由本課程助教逕行給予各位同學適當的成績！