



# 隨堂練習8

(爬ppt 省錢版)

Peter的好朋友，Leo 的生日快到了，Peter 知道 Leo 最近很想要換耳機，想買給他當生日禮物。不過買完耳機後 Peter 的錢包將會見底，所以他想從日常三餐下手省錢。請大家從「PTT 省錢版」中找到有用的省錢快報給 Peter 參考！

目標:

- (1) 利用 Beautiful soup 套件爬取 PTT 省錢版
- (2) 轉為 Data Frame 格式

## 輸入格式

無

## 輸出格式

0

- 0 "[情報] Red Bull Taiwan LINE免費領紅牛"
- 1 "[情報] 淘寶官方Line抽獎"
- 2 "[情報] 福勝亭自由時報漲價了335"
- 3 "[情報] Line官方使用經驗調查 填完拿10P"
- 4 "[情報] APPLE超品日 iPhone 16e"
- 5 "[情報] Aesop洗髮露500ml 限時\$999"
- 6 "[公告] 省錢板板規(113.9.10修訂)"
- 7 "[情報] 全台捐血贈品"
- 8 "[公告] 違規禁言公告 暨 板規討論公告"
- 9 "[集中] 贈送集中文"
- 10 "[集中] 未滿1元集點抽獎資訊集中文"

## Hint

請參考這個檔案進行修改(記得要複製回自己的雲端窩~)

也可以參考這個網路教材窩 (

<https://youtu.be/9Z9xKWfNo7k?si=B2F9qrOKSLkN80uI>)

<https://colab.research.google.com/drive/1wJYGTNbtVSwrBqCM8su8jBAbq8J255FL?usp=sharing>

## Answer&步驟分解

[https://colab.research.google.com/drive/1kiVqZY1jC8EgiLIncxnIRJuLa\\_vIW-pl?usp=sharing](https://colab.research.google.com/drive/1kiVqZY1jC8EgiLIncxnIRJuLa_vIW-pl?usp=sharing)(重點是從新竹版改成省錢版)

```

# 抓取 PTT 省錢版的網頁原始碼(HTML)
import urllib.request as req
import json

url= "https://www.ptt.cc/bbs/Lifeismoney/index.html" # 可以換成想要抓的PTT版之
# 建立一個 Request 物件，附加 Request Headers 的資訊(看起來像人類)
request=req.Request(url, headers={
    "User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
})
with req.urlopen(request) as response:
    data=response.read().decode("utf-8")
# 解析原始碼，取得每篇文章的標題
import bs4
root=bs4.BeautifulSoup(data, "html.parser")
titles=root.find_all("div", class_="title") # 尋找所有 class="title" 的 div 標籤
for title in titles:
    if title.a !=None: # 如果標題包含 a 標籤(仍存在於網頁上)，就印出來
        finish = title.a.string
        json_dict = json.dumps(finish,ensure_ascii=False).encode('utf8')
        decode = json_dict.decode()
        print(decode)

# 轉為csv檔
temp = []
for title in titles:
    if title.a !=None: # 如果標題包含 a 標籤(仍存在於網頁上)，就印出來
        finish = title.a.string
        json_dict = json.dumps(finish,ensure_ascii=False).encode('utf8')
        decode = json_dict.decode()
        temp.append(decode)
        print(decode)
# 轉換成 DataFrame
import pandas as pd
df = pd.DataFrame(temp)

```

