## > 環境安裝及套件安裝:Anaconda

網址下載 Anaconda: <a href="https://www.anaconda.com/distribution/">https://www.anaconda.com/distribution/</a>

## **Anaconda 2019.10 for Windows Installer**

# Python 3.7 version Download 64-Bit Graphical Installer (462 MB) 32-Bit Graphical Installer (410 MB)



圖 1. Anaconda 下載頁面

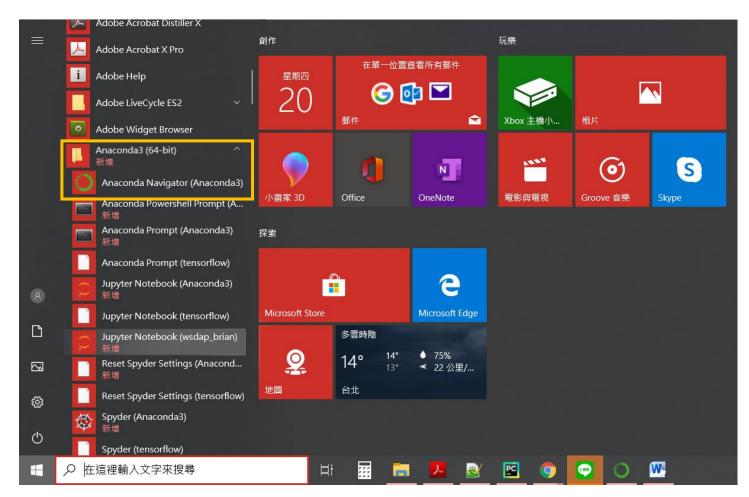


圖 2. 啟動 Anaconda Navigator

<u>F</u>ile <u>H</u>elp

# **ANACONDA** NAVIGATOR

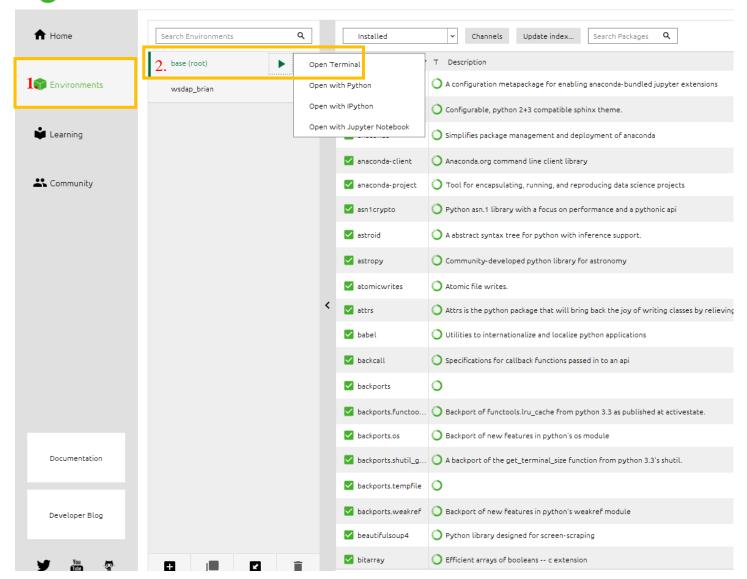


圖 3. 開啟命令提示字元 (Terminal)



圖 4. 確認 Python 版本

通常會在一台電腦上同時開發多個專案,不同專案需要安裝不同的套件,而不想把所有套件都混裝在全域環境中,因此會需要為專案建置一個虛擬環境,虛擬環境的建置可以透過 Anaconda Navigator 完成。首先切換到 Anaconda Navigator 左側的 Environments ,在下點擊 Create ,會看到一個視窗要求輸入虛擬環境的名稱,輸入 wsdap\_brian ,同時也會在視窗看到虛擬環境的安裝目錄 (Location) ,如圖 5. 。



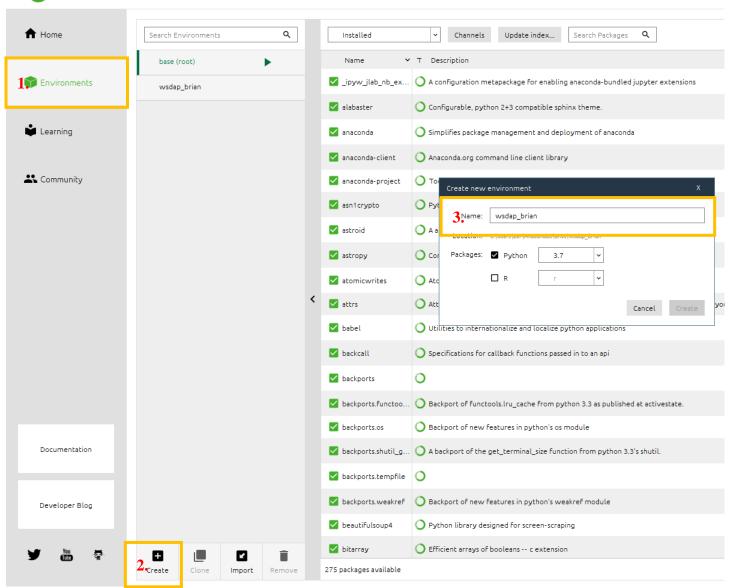


圖 5. 建立虚擬環境

虛擬環境建置完成後,就可以在 Environments 內看到。若想在該環境內安裝套件,只要先點擊環境 名稱,並且在右邊的搜尋欄位搜尋並安裝即可,只需要三個套件:

- requests
- beautifulsoup4
- html5lib

在搜尋欄搜尋套件名稱,勾選該套件,點擊下方 Apply 安裝,如圖 6. ,就可以依序安裝好上述三個套件。接著,開啟這個虛擬環境的 Terminal ,會看到開啟的命列前面多了虛擬環境的名稱 (wsdap\_brian) ,代表正在這個虛擬環境中,也可以使用指令 python—c "import requests" 測試 requests 套件是否已經安裝好完成,如圖 7. ,沒有錯誤訊息就代表安裝完成。

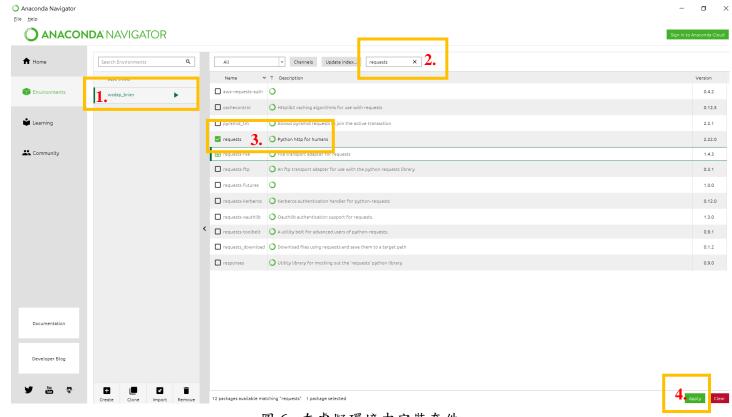


圖 6. 在虛擬環境中安裝套件

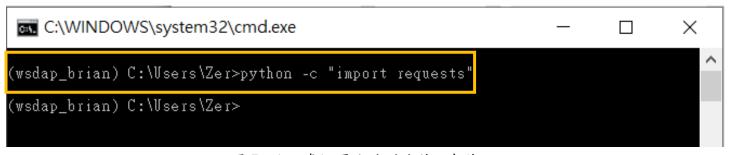


圖 7. 進入虛擬環境並測試引入套件

有時候想安裝的套件並未收錄在 Anaconda 內,此時直接搜尋會找不到該套件,這種情況的安裝方式是透過命令列 (pip install) ,例如想安裝 googlemaps 這個未收錄套件,方式是在命令輸入 pip install googlemaps ,如圖 8.。

```
C:\WINDOWS\system32\cmd.exe

(wsdap_brian) C:\Users\Zer>python -c "import requests"

(wsdap_brian) C:\Users\Zer>pip install googlemaps
Requirement already satisfied: googlemaps in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (4.2.0)
Requirement already satisfied: requests<3.0,>=2.20.0 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from googlemaps) (2.22.0)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from requests<3.0,>=2.20.0->googlemaps) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from requests<3.0,>=2.20.0->googlemaps) (2019.11.28)
Requirement already satisfied: idna<2.9,>=2.5 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from requests<3.0,>=2.20.0->googlemaps) (2.8)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from requests<3.0,>=2.20.0->googlemaps) (2.8)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\zer\anaconda3\envs\wsdap_brian\lib\site-packages (from requests<3.0,>=2.20.0->googlemaps) (2.8)
(wsdap_brian) C:\Users\Zer>

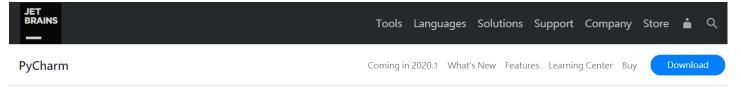
(wsdap_brian) C:\Users\Zer>
```

圖 8. 使用 pip install 安裝 Anaconda 未收錄套件

# ▶ 使用 IDE: PyCharm

官網下載並安裝 Community 版 (社群版) 的 Pycharm :

https://www.jetbrains.com/pycharm/download/#section=windows





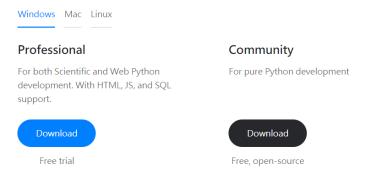
Version: 2019.3.3 Build: 193.6494.30 7 February 2020

System requirements

Installation Instructions

Other versions

# **Download PyCharm**





Get the Toolbox App to download PyCharm and its future updates with ease

圖 9. PyCharm 下載頁面

若建立專案時忘記連結 Anaconda Navigator 虛擬環境,或之後想要改用另外一個虛擬環境,可以在專案開啟後,選擇左上角的 File 選單,進入 Settings 選項,在 Project Interpreter 依樣可以連結虛擬環境,如圖 10. 、圖 11. 、圖 12. 。

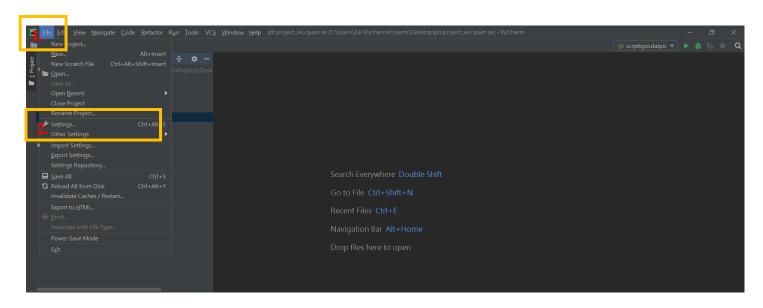


圖 10. 修改專案的虛擬環境

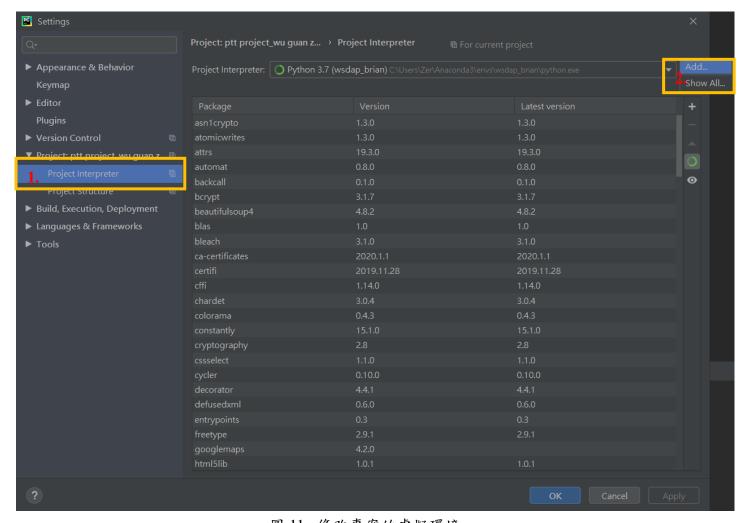


圖 11. 修改專案的虛擬環境

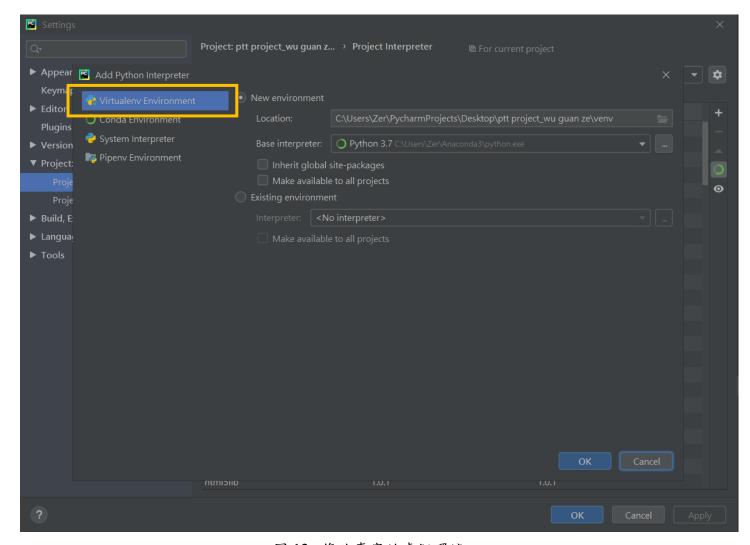


圖 12. 修改專案的虛擬環境

# ▶ 使用 Jupyter Notebook

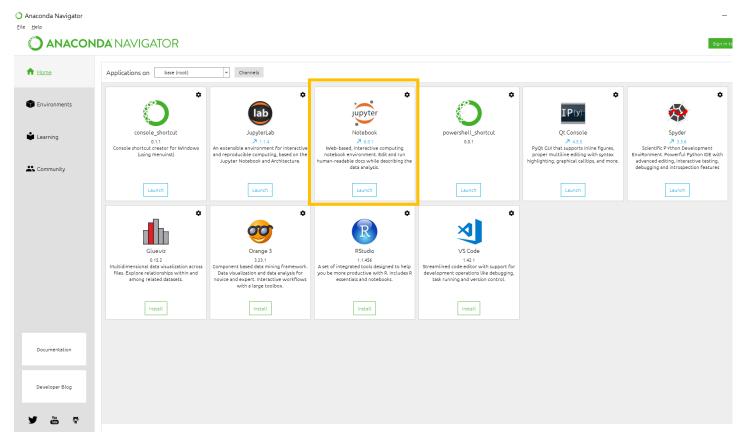


圖 13. 從 Anaconda Navigator 啟動全域環境的 Jupyter Notebook

首先在 Anaconda Navigator 內切換到 Enviroments 列表,選擇建立的虛擬環境 (此處名稱為 wsdap\_brian),先在這個環境下安裝 Jupyter Notebook。在右方套件列表的下拉式選單中選擇 All ,並搜尋 jupyter ,點選 jupyter 套件,並點擊下方的 Apply 安裝,如圖 14.。

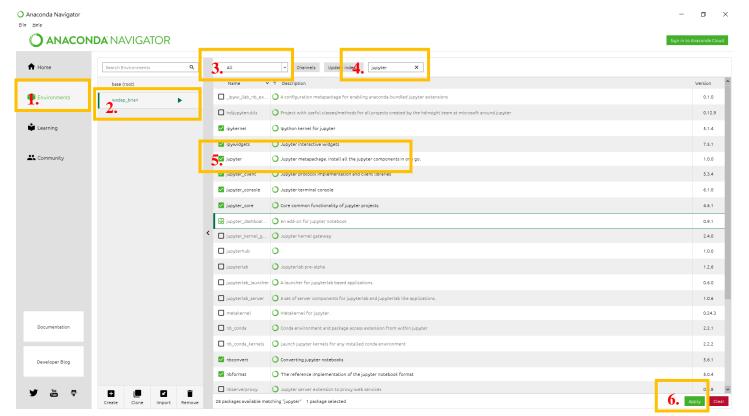


圖 14. 為虛擬環境安裝 Jupyter Notebook

安裝完成之後,就可以啟東該虛擬環境的 Jupyter Notebook 。 點選虛擬環境右邊的綠色箭頭,並點擊 Open with Jupyter Notebook , 如圖 15. 。

# **ANACONDA** NAVIGATOR

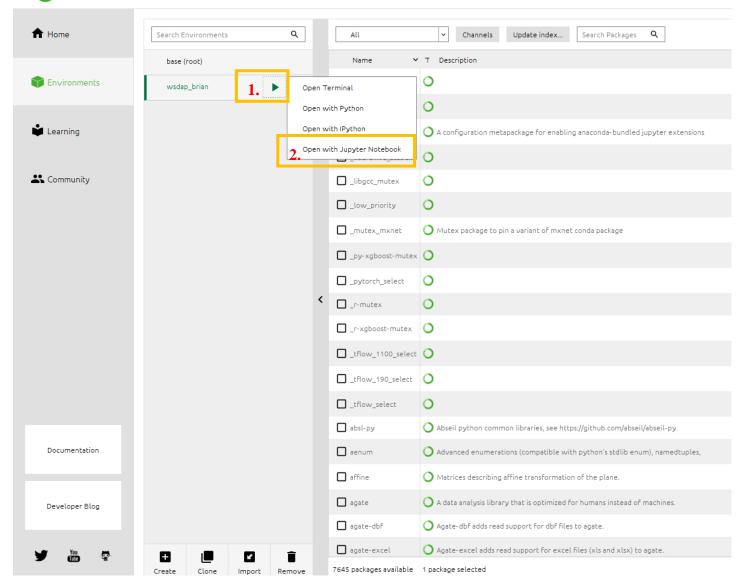


圖 15. 啟動虛擬環境的 Jupyter Notebook

Jupyter Notebook 是一個以瀏覽器為主的開發介面,所以啟動後會開啟瀏覽器,且預設連結到 localhost:8888 的網址,如圖 16.。

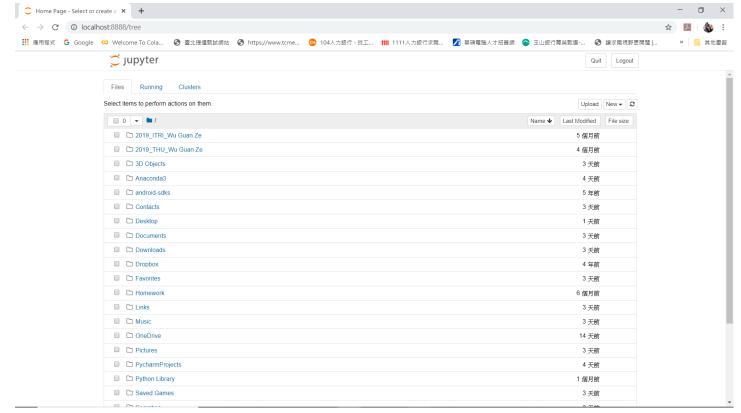


圖 16. Jupyter Notebook 啟動畫面

## ▶ 檔案名稱: scrpttgosdat.py (PyCharm)

抓取 ptt gossiping 單一分頁的看板、作者、標題、時間,(如圖 17.~ 圖 21.)。

圖 17. scrpttgosdat.py 程式碼

```
ms = gossiping_divs.find_all("div", ['article-metaline', 'article-metaline-right', 'push']) # 扣掉 看板作者、標題、時間、推文

for m in ms:
    m.extract()

gossiping_divs = gossiping.find('div', id='main-content')
print(gossiping_divs.text)
```

圖 18. scrpttgosdat.py 程式碼

```
Run:

scrpttgosdat ×

C:\Users\Zer\Anaconda3\envs\wsdap_brian\python.exe "C:\Users\Zer\PycharmProjects\Desktop\ptt project_wu guan ze\scrpttgosdat.py"

看板 aby123 (批踢踢小新手)

作者 Gossiping

標題 [問卦] 到底為哈要隨滿實情啊

時間 Tue Feb 4 17:58:35 2020
```

圖 19. scrpttgosdat.py 執行結果

圖 20. scrpttgosdat.py 執行結果

圖 21. scrpttgosdat.py 執行結果

# ▶ 檔案名稱: scrpttgosdatpic.py (PyCharm) 、 scrpttgosdatpic.json 、

#### scrpttgosdatpic.csv , scrpttgosdatpicture

ptt gossiping 板的當頁面的所有貼文,抓取文章標題、文章作者、文章連結、貼文時間、推文數量,並把5筆資料存入 scrpttgosdatpic.json 、scrpttgosdatpic.csv 檔案 (如圖 22. ~ 圖 28.);之後再進入ptt gossiping 板的當頁面的第一則貼文,抓取分頁的看板、作者、標題、時間、文章內容,以及檢查有沒有圖片,如果有圖片就下載並儲存 scrpttgosdatpicture 的資料夾裡 (如圖 29. ~ 圖 36.)。

```
Eile Edit View Navigate Code Refactor Run Tools VCS Window Help ptt project_wu guan ze [C:\Users\Zer\PycharmProjects\Desktop\ptt project_wu guan ze]
ptt project_wu guan ze > 🛵 scrpttgosdatpic.py
  🖧 scrpttgosdat.py 🗴 🐔 scrpttgosdatpic.py 🗴 🚦 scrpttgosdatpic.csv 🗴 🥻 scrpttgosdatpic.json
         # ptt gossiping 板的當頁面的所有貼文,抓取相關資料#
         ptt_url_01 = 'https://www.ptt.cc/bbs/Gossiping/index.html'
         ptt_html_01 = requests.get(ptt_url_01, cookies={'over18':'1'})
         ptt Soup 01 = BeautifulSoup(ptt html 01.text, 'html5lib')
                               # 本頁面文章
         articles = []
         pttdivs = ptt_Soup_01.find_all('div', 'r-ent')
                                                                             # 文章標顥
                                                                              # 文章作者
                                                                              # 文章連結
                                                                              # 貼文時間
                                                                              # 推文數量
```

圖 22. scrpttgosdatpic.py 程式碼

```
# 寫入 JSON 檔案
with open('scrpttgosdatpic.json', 'w', encoding='utf-8') as fnObj:
    json.dump(articles, fnObj, ensure_ascii=_False, indent=2)

# 寫入 CSV 檔案

with open('scrpttgosdatpic.csv', 'w', newline='', encoding='utf-8-sig') as csvFile:

fields = ['title', 'author', 'href', 'ptime', 'push_num'] # 定義欄位

dictWriter = csv.DictWriter(csvFile, fieldnames=fields) # 將 dictionary 寫入 CSV 檔案

dictWriter.writeheader() # 寫入標題

for inputArticle in articles: # 寫入資料

dictWriter.writerow(inputArticle)
```

圖 23. scrpttgosdatpic.py 程式碼

```
🎒 scrpttgosdatpic.json - 記事本
                                                                                                                \times
檔案(\underline{F}) 編輯(\underline{E}) 格式(\underline{O}) 檢視(\underline{V}) 說明
   "title": "[爆卦] 山東監獄武漢肺炎207例新增",
"author": "jamshana",
    "href": "/bbs/Gossiping/M.1582251784.A.981.html",
    "ptime": " 2/21", "push_num": "19"
    "href": "/bbs/Gossiping/M.1582251820.A.848.html", "ptime": " 2/21", "push_num": "9"
    "title": "[問卦] 美國CDC 日本1級警示台灣為什麼不提升?", "author": "lavernel",
    "href": "/bbs/Gossiping/M.1582251836.A.75C.html", "ptime": " 2/21", "push_num": "8"
    "title": "[新聞] 香港防暴警察首度確診 發燒仍參加警聚與", "author": "VladGuerrero",
    "href": "/bbs/Gossiping/M.1582251882.A.A97.html",
"ptime": " 2/21",
"push_num": "8"
    "title": "Re: [問卦] 不停航嗎?",
"author": "oooo",
"href": "/bbs/Gossiping/M.1582251891.A.9EB.html",
"ptime": " 2/21",
"push_num": ""
    "title": "Re: [問卦] 武統台灣已成大陸社會共識!",
"author": "vaiking0120",
    "href": "/bbs/Gossiping/M.1582251931.A.933.html",
"ptime": " 2/21",
"push_num": "3"
```

圖 24. 存儲資料到 scrpttgosdatpic.json 的結果

```
🎒 scrpttgosdatpic.json - 記事本
                                                                                                                                                             \times
檔案(\underline{F}) 編輯(\underline{E}) 格式(\underline{O}) 檢視(\underline{V}) 說明
     "title": "[新聞] 誠念「九字真言」治癒疫病的神奇故事",
"author": "mykey8585",
"href": "/bbs/Gossiping/M.1582251937.A.15D.html",
"ptime": " 2/21",
"push_num": "9"
     "title": "[問卦] 與中國口徑一致的中共同路人484很明顯了",
"author": "hyoerious",
"href": "/bbs/Gossiping/M.1582251973.A.682.html",
"ptime": " 2/21",
"push_num": "1"
      "href": "/bbs/Gossiping/M.1582251986.A.051.html", "ptime": " 2/21", "push_num": ""
     "title": "Re: [問卦] 中國投票高達85%支持武統台灣要怕嗎",
"author": "Nagasumi",
"href": "/bbs/Gossiping/M.1582251999.A.BF6.html",
"ptime": " 2/21",
"push_num": ""
      "title": "[問卦] 該檢舉我們公司女主管嗎?", "author": "b8301013", " 1500051000 A 701
      author": "08301013",
"href": "/bbs/Gossiping/M.1582251999.A.701.html",
"ptime": " 2/21",
"push_num": "6"
     "title": "Re: [問卦] 碳佐麻里 vs 屋馬",
"author": "tachikoma",
"href": "/bbs/Gossiping/M.1582252006.A.C3D.html",
"ptime": " 2/21",
"push_num": "1"
```

圖 25. 存储資料到 scrpttgosdatpic.json 的結果

```
🧻 scrpttgosdatpic.json - 記事本
                                                                                                                                                                         Χ
檔案(\underline{F}) 編輯(\underline{E}) 格式(\underline{O}) 檢視(\underline{V}) 說明
      "title": "[問卦] 像馬英九這樣從來沒有失敗的人生,爽嗎??",
"author": "vega0501",
"href": "/bbs/Gossiping/M.1582252018.A.D32.html",
       "ptime": "2/21",
"push_num": "8"
  },
{
      "title": "[新聞]令人髮指!他硬上同居人女兒 還先問是否處",
"author": "SakuraYui",
"href": "/bbs/Gossiping/M.1582252064.A.D50.html",
"ptime": " 2/21",
"push_num": "6"
  },
  {
     "title": "Re: [問卦] 房仲這職業有什麼屁用嗎?",
"author": "doraBBO",
"href": "/bbs/Gossiping/M.1582252077.A.OCO.html",
"ptime": " 2/21",
"push_num": ""
      "title": "[爆卦] 好可怕的關聯",
"author": "buttl106",
"href": "/bbs/Gossiping/M.1582252118.A.488.html",
"ptime": " 2/21",
"push_num": "5"
      "title": "[問卦] 兩個黃種人在火車大聲用英文..",
"author": "neutron",
"href": "/bbs/Gossiping/M.1582252134.A.C7B.html",
"ptime": " 2/21",
"push_num": "1"
      "title": "[新聞] 新國會開議 高虹安力推「愛的迫降」",
"author": "linhu8883324",
      "href": "/bbs/Gossiping/M.1582252141.A.A4F.htm1", "ptime": " 2/21", "push_num": "2"
  },
```

圖 26. 存儲資料到 scrpttgosdatpic.json 的結果

```
🧻 scrpttgosdatpic.json - 記事本
                                                                                                                                                                                   ×
檔案(\underline{F}) 編輯(\underline{E}) 格式(\underline{O}) 檢視(\underline{V}) 說明
   {
       "title": "Re: [新聞] 新北失聯男「管永安」找到了! 藏身淡水橋",
"author": "ryohgi",
"href": "/bbs/Gossiping/M.1582252169.A.4DE.html",
"ptime": " 2/21",
"push_num": "2"
      "title": "[新聞] 大S狂掃口罩被罵翻 噤聲1月現身「一刀剪",
"author": "Rostislav",
"href": "/bbs/Gossiping/M.1582252193.A.AC2.html",
"ptime": " 2/21",
"push_num": ""
       "title": "[公告] 關於板上疫情相關問卦&八卦板板規",
"author": "arsonlolita",
"href": "/bbs/Gossiping/M.1580443754.A.30D.html",
"ptime": " 1/31",
"push_num": "5"
       "title": "[公告] 武漢肺炎討論專區&二月份置底閒聊文",
"author": "Bignana",
"href": "/bbs/Gossiping/M.1580672851.A.909.html",
"ptime": " 2/03",
"push_num": "X3"
       "title": "[徵求] 行車記錄器 國道北上17.7K 晚上18:3", "author": "berger200350",
       "href": "/bbs/Gossiping/M.1581497232.A.D82.html",
"ptime": " 2/12",
"push_num": "8"
       "title": "[徵求] 行車記錄器 台中市向上路6段往中沙路",
"author": "xodd123456",
"href": "/bbs/Gossiping/M.1581499831.A.700.html",
"ptime": " 2/12",
"push_num": "23"
   }
```

圖 27. 存儲資料到 scrpttgosdatpic.json 的結果

	<b>3</b> 17	, 3	latpic.csv - Microsoft Excel (產品啟動失敗)		
檔案 常用 插入 版面配置 公式 資料 校閱 檢視 負載測試 Acrobat 小組					
	G26 <b>▼</b> (* fx	D	0	D	Г
1	A	В	C	D	Е
1	title	author	href	ptime	push_num
2	[爆卦] 山東監獄武漢肺炎207例新增	iamshana	/bbs/Gossiping/M.1582251784.A.981.html	2/21	19
3	[新聞] 再有2軍人確診 南韓軍隊明起禁官兵休假外	runa2	/bbs/Gossiping/M.1582251820.A.848.html	2/21	Ç
4	[問卦] 美國CDC 日本1級警示台灣為什麼不提升?	laverne1	/bbs/Gossiping/M.1582251836.A.75C.html	2/21	8
5	[新聞] 香港防暴警察首度確診 發燒仍參加警聚與	VladGuerrero	/bbs/Gossiping/M.1582251882.A.A97.html	2/21	8
6	Re: [問卦] 不停航嗎?	0000	/bbs/Gossiping/M.1582251891.A.9EB.html	2/21	
7	Re: [問卦] 武統台灣已成大陸社會共識!	vaiking0120	/bbs/Gossiping/M.1582251931.A.933.html	2/21	3
8	[新聞] 誠念「九字真言」治癒疫病的神奇故事	mykey8585	/bbs/Gossiping/M.1582251937.A.15D.html	2/21	Ç
9	[問卦] 與中國口徑一致的中共同路人484很明顯了	hyoerious	/bbs/Gossiping/M.1582251973.A.682.html	2/21	]
10	Re: [新聞] 政院拍板 遇重大建設 國土計畫法可隨時	drea	/bbs/Gossiping/M.1582251986.A.051.html	2/21	
11	Re: [問卦] 中國投票高達85%支持武統台灣要怕嗎	Nagasumi	/bbs/Gossiping/M.1582251999.A.BF6.html	2/21	
12	[問卦] 該檢舉我們公司女主管嗎?	b8301013	/bbs/Gossiping/M.1582251999.A.701.html	2/21	(
13	Re: [問卦] 碳佐麻里 vs 屋馬	tachikoma	/bbs/Gossiping/M.1582252006.A.C3D.html	2/21	1
14	[問卦] 像馬英九這樣從來沒有失敗的人生,爽嗎??	vega0501	/bbs/Gossiping/M.1582252018.A.D32.html	2/21	8
15	[新聞]令人髮指!他硬上同居人女兒 還先問是否處	SakuraYui	/bbs/Gossiping/M.1582252064.A.D50.html	2/21	(
16	Re: [問卦] 房仲這職業有什麼屁用嗎?	doraBBO	/bbs/Gossiping/M.1582252077.A.0C0.html	2/21	
17	[爆卦] 好可怕的關聯	butt1106	/bbs/Gossiping/M.1582252118.A.488.html	2/21	4
18	[問卦] 兩個黃種人在火車大聲用英文	neutron	/bbs/Gossiping/M.1582252134.A.C7B.html	2/21	
19	[新聞] 新國會開議 高虹安力推「愛的追降」	linhu8883324	/bbs/Gossiping/M.1582252141.A.A4F.html	2/21	2
20	Re: [新聞] 新北失聯男「管永安」找到了! 藏身淡水橋	ryohgi	/bbs/Gossiping/M.1582252169.A.4DE.html	2/21	
21	[新聞] 大S狂掃口罩被罵翻 噤聲1月現身「一刀剪	Rostislav	/bbs/Gossiping/M.1582252193.A.AC2.html	2/21	•
22	[公告] 關於板上疫情相關問卦&八卦板板規	arsonlolita	/bbs/Gossiping/M.1580443754.A.30D.html	1/31	
23	[公告] 武漢肺炎討論專區&二月份置底閒聊文	Bignana	/bbs/Gossiping/M.1580672851.A.909.html	2/03	X3
24	[徵求] 行車記錄器 國道北上17.7K 晚上18:3	berger200350	/bbs/Gossiping/M.1581497232.A.D82.html	2/12	113
25	[徵求] 行車記錄器 台中市向上路6段往中沙路	xodd123456	/bbs/Gossiping/M.1581499831.A.700.html	2/12	23

圖 28. 儲存資料到 scrpttgosdatpic.csv 的結果

圖 29. scrpttgosdatpic.py 程式碼

```
ms = gossiping_divs.find_all("div", ['article-metaline', 'article-metaline-right', 'push']) # 扣掉 看板作者、標題、時間、推文

for m in ms:
    m.extract()

gossiping_divs = gossiping_soup.find('div', id='main-content')
print(gossiping_divs.text)

photos = [] # 圖片網址
url_photos = gossiping_divs.find_all('a') # 找專所有圖片

for photo in url_photos:
    href_photo = photo['href']
    if href_photo.startswith('https://i.imgur'): # 判斷圖片網址
    photos.append(href_photo)

for photo in photos: # 列印圖片網址
    print(photo)
```

圖 30. scrpttgosdatpic.py 程式碼

圖 31. scrpttgosdatpic.py 程式碼

```
Run: scrpttgosdatpic ×

C:\Users\Zer\Anaconda3\envs\wsdap_brian\python.exe "C:\Users\Zer\PycharmProjects\Desktop\ptt project_wu guan ze\scrpttgosdatpic.py"

看板 iamshana (CANINE力集中在一些重要S)

作者 Gossiping

標題 [爆卦] 山東監獄武漢肺炎207例新增

時間 Fri Feb 21 10:23:02 2020
```

圖 32. scrpttgosdatpic.py 執行結果

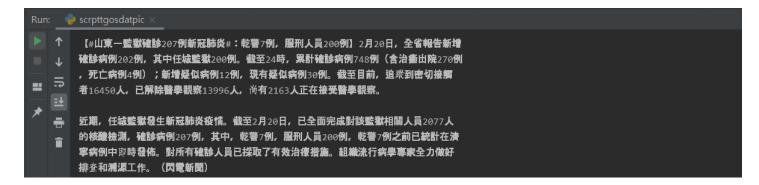


圖 33. scrpttgosdatpic.py 執行結果



圖 34. scrpttgosdatpic.py 執行結果



圖 35. scrpttgosdatpic.py 執行結果

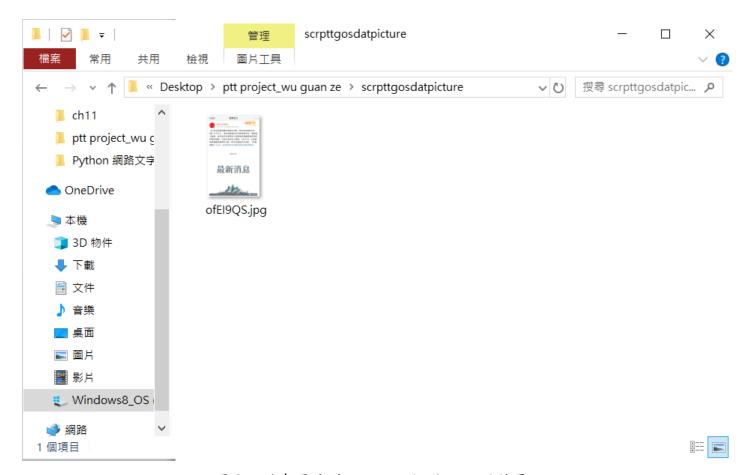


圖 36. 儲存圖片到 scrpttgosdatpicture 的結果