

410921208 楊右宇

```
from selenium import webdriver
from time import sleep
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
import matplotlib.pyplot as plt
import matplotlib.image as img
import requests
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
import re

url = 'https://www.cdc.gov.tw/Bulletin/List/MmgtpeidAR5Ooi4-fgHzQ'
html = requests.get(url)
sp = BeautifulSoup(html.text, 'html.parser')

news_all = sp.find_all(class_="content-boxes-v3")

news = []

for i in news_all:
    if "COVID-19 確定病例" in i.text:
        news += [i]

# print(news)
Date = []
Case_Increase = []
Local = []
Overseas = []
herfs = []

for i in range(3):
    data = re.findall(r'[\d+,]+' , news[i].text)
    herf = news[i].find('a')
```

```

        h = 'https://www.cdc.gov.tw' + herf.attrs['href']
#         print(herf)
        herfs += [h]
#         print(data)
        Date += ['/'.join(data[0:3])]
        Case_Increase += [data[3].replace(',', ' ')]
        Local += [data[5].replace(',', ' ')]
        Overseas += [data[6].replace(',', ' ')]

# print(Date)
# print(Local)
# print(Case_Increase)
# print(Overseas)
# print(herfs)

df = pd.DataFrame({'Date':Date,
                   'Case Increase':Case_Increase,
                   'Local':Local,
                   'Overseas':Overseas,
                   'Press Link':herfs})
# df.index = df['Date']

from datetime import datetime
from datetime import date

now = datetime.now()
current_time = now.strftime("%H:%M")

today = date.today()

# dd/mm/YY
d1 = today.strftime("%Y/%m/%d")

df.loc[3] = ['資料取得時間：', d1, current_time, "", ""]
# df.loc[3] = '資料取得時間：'

```

df

```
In [58]: from selenium import webdriver
from time import sleep
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
import matplotlib.pyplot as plt
import matplotlib.image as img
import requests
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
import re

url = 'https://www.cdc.gov.tw/Bulletin/List/HmgtpeidAR50oai4-fgHzQ'
html = requests.get(url)
sp = BeautifulSoup(html.text, 'html.parser')

news_all = sp.find_all(class_="content-boxes-v3")
news = []

for i in news_all:
    if "COVID-19確定病例" in i.text:
        news += [i]

# print(news)
Date = []
Case_Increase = []
Local = []
Overseas = []
herfs = []

for i in range(3):
    data = re.findall(r'[\d+]', news[i].text)
    herf = news[i].find('a')
    h = 'https://www.cdc.gov.tw' + herf.attrs['href']
    # print(herf)
    herfs += [h]
    # print(data)
    Date += ['/'.join(data[0:3])]
    Case_Increase += [data[3].replace(',','')]
    Local += [data[5].replace(',','')]
    Overseas += [data[6].replace(',','')]
```

```
Case_Increase += [data[3].replace(',','')]
Local += [data[5].replace(',','')]
Overseas += [data[6].replace(',','')]

# print(Date)
# print(Local)
# print(Case_Increase)
# print(Overseas)
# print(herfs)

df = pd.DataFrame({'Date':Date,
                  'Case_Increase':Case_Increase,
                  'Local':Local,
                  'Overseas':Overseas,
                  'Press_Link':herfs})
# df.index = df['Date']

from datetime import datetime
from datetime import date

now = datetime.now()
current_time = now.strftime("%H:%M")

today = date.today()

# dd/mm/YY
d1 = today.strftime("%Y/%m/%d")

df.loc[3] = ['資料取得時間:', d1, current_time, "", ""]
# df.loc[3] = '資料取得時間:'
df
```

Out[58]:

	Date	Case Increase	Local	Overseas	Press Link
0	2022/5/11	57216	57188	28	<a href="https://www.cdc.gov.tw/Bulletin/Detail/3aEsJh8...">https://www.cdc.gov.tw/Bulletin/Detail/3aEsJh8...</a>
1	2022/5/10	50828	50780	48	<a href="https://www.cdc.gov.tw/Bulletin/Detail/ObYs7SL...">https://www.cdc.gov.tw/Bulletin/Detail/ObYs7SL...</a>
2	2022/5/9	40304	40263	41	<a href="https://www.cdc.gov.tw/Bulletin/Detail/G_e_Y3...">https://www.cdc.gov.tw/Bulletin/Detail/G_e_Y3...</a>
3	資料取得時間:	2022/05/12	12:54		