氣象資料開放平臺的資料擷取

- 在中央氣象局的氣象資料開放平台網頁註冊免費的帳號來獲得授權碼
 - 劉奕的作法是直接深入解析資料底層
 - 直接處理網頁上的文字
 - 不需要另外註冊帳號就可以獲得中央氣象局的資料
 - 個人覺得透過註冊帳號後取得的資料來源比較穩定,而且不會因氣象局改變資料欄位位置而需要修改程式
- 程式概括
 - 1. 下載並解壓縮CWB官方檔案
 - 2. 轉換原始檔案並擷取資料 (利用 xmltodict 第三方模組)
 - 3. 壓縮並備份資料
 - 4. 合併新舊資料 (使用強大的 pandas 第三方模組)
- 每日新增並合併前次海象監測資料(csv) http://60.249.144.60:8084

處理好的 csv 資料 (龍洞潮位站)

_													C4	A02.csv -	LibreOffice Ca			
檔案	(F) 編輯(E)	檢視(\	/) 插	入(I) 格	式(O)	羕式(Y)	工作表(S) 資料(D)	工具(T) 礼	児窗(W)	說明	(H)						
	• 🗊 • 🖞 •	-	e	Q 1	f (•	占 💪		🏹 🕭	•	•	↓† A↓	z† 🚿		Ω.	0 📮 🚍 🜉		
Libe	eration Sans	~	10 p	ot ~	В	ΣU	<u>A</u> -	r ⁸ • 🖹	ΞΞ	7	t	± + ⊥	•	6 0.0	15 .00, .00,	≧ ≣		•
A1 \vee $f_* \Sigma \cdot = dataT$						e												
	A			В	С		D	E	F		G	н	I I .		J	К	L	
1	dataTime		t	ideHeight	tideLevel	seaTe	emperature	temperature	stationPressu	re wind	Speed	windScale	windDirectio	n windDi	rectionDescription	maximumWindSpee	d maximumWindSca	ale
2	2021-04-10T23	:00:00+0	08:00	-0.1	7 退潮		20.1	19.3	10	20.1	0.8	1	1 3	306 NW			1.2	1
3	2021-04-11T00	00:00+0	00:80	-0.2	8退潮		20.1	19.4	1	.020	0.6	1	1 2	50 WSW			1.6	2
4	2021-04-11T01:00:00+08:00			-0.3	-0.35 退潮		20.1	20.6).6 1019.6		2.7	2	2 2	207 SSW		4.3		3
5	2021-04-11T02	21-04-11T02:00:00+08:00 -0.29 漲潮				20.1 20.2		1	.019	2.1	2	196 SSW				4.1	3	
б	2021-04-11T03	21-04-11T03:00:00+08:00 -0.15 漲潮				20.1 2		10	18.7	1.6	2	2 2	233 SW		3.7		3	
7	2021-04-11T04	:00:00+0	00:80	0.0	3 漲潮		20.1	. 20.8	10	18.7	1.2	1	1 2	258 WSW			1.7	2
8	2021-04-11T05	00:00+0	00:80	0.1	7 漲潮		20.2	20.7	10	18.9	2.2	2	2 2	21 SW			3.6	3
9	2021-04-11T06	:00:00+0	00:80	0.	3 漲潮		20.3	20.9	10	19.4	4.2	3	3 1	83 S			6.1	4
10	2021-04-11T07	:00:00+0	00:80	0.3	2 退潮		20.3	21.4	10	19.8	4.8	3	3 1	80 S			8	5
11	2021-04-11T08	:00:00+0	00:80	0.2	4 退潮		20.3	22.1	. 102	20.3	3.8	3	3 2	215 SW			5.4	3
12	2021-04-11T09	:00:00+0	00:80	0.1	5 退潮		20.2	22	102	20.4	2.7	2	2 2	208 SSW			4.4	3
13	2021-04-11T10	00:00+0	00:80	0.0	2 退潮		20.2	22.1	. 102	20.6	3.3	2	2 1	81 S			4.9	3
14	2021-04-11T11	:00:00+0	00:80	-0.1	3 退潮		20.2	21.9	102	20.5	2.8	2	2 1	79 S			4.1	3
15	2021-04-11T12	:00:00+0	00:80	-0.2	5 退潮		20.2	23	10	19.7	4.9	3	3 1	73 S			7.7	4
16	2021-04-11T13	:00:00+0	00:80	-0.	3 退潮		20.4	23.3	10	18.7	4.5	3	3 1	49 SSE			6.1	4
17	2021-04-11T14	:00:00+0	00:80	-0.2	7 漲潮		20.5	23.9	10	17.7	5.5	4	4 2	205 SSW			7.6	4
18	2021-04-11T15	00:00+0	00:80	-0.1	3 漲潮		20.6	23.5	10	17.4	4.5	3	3 1	74 S			6.8	4
19	2021-04-11T16	00:00+0	00:80	0.0	2 漲潮		20.7	23.8	10	17.1	5.1	3	3 1	70 S			9	5
20	2021-04-11T17	:00:00+0	00:80	0.1	7 漲潮		20.8	3 23.4	10	17.2	3.8	3	3 1	91 S			5.8	4
21	2021-04-11T18	:00:00+0	00:80	0.2	9 漲潮		20.9	23.7	10	17.4	4.2	3	3 2	202 SSW			6.7	4
22	2021-04-11T19	00:00+0	00:80	0.3	4 退潮		20.9	23.5	10	17.8	3.9	3	3 2	209 SSW			6.2	4
23	2021-04-11T20	00:00+0	00:80	0.2	7 退潮		20.8	23.2	10	17.9	4.3	3	3 2	206 SSW			6.3	4
24	2021-04-11T21	00:00+0	00:80	0.1	5 退潮		20.7	23	10	18.3	3.7	3	3 2	03 SSW			6.3	4
25	2021-04-11T22	00:00+0	00:80	0.0	2 退潮		20.7	22.8	10	18.6	3.9	3	3 2	216 SW			5.9	4
26	2021-04-11T23	00:00+0	08:00	-0.0	9 退潮		20.6	22.6	10	18.2	3.8	3	3 1	59 SSE			6.6	4
27	2021-04-12T00	00:00+0	00:80	-0.	2 退潮		20.5	22.8	10	17.8	3.9	3	3 1	75 S			6.8	4
28	2021-04-12T01	00:00+0	00:80	-0.2	9 退潮		20.5	23.2	10	17.2	4.5	3	3 1	53 SSE			7.2	4
29	2021-04-12T02	00:00+0	00:80	-0.2	7 漲潮		20.5	i 23.3	10	16.5	3.3	2	2 2	25 SW			5.4	3

氣象資料開放平臺的授權碼



合 → 會員資訊 > API授權碼

會

員基本資料	API授權碼
	本平臺提供透過URL下載檔案以及 RESTful API 資料攝取方法取用資料,惟因本平臺採用會員服務機制,需帶入資料項目代碼 以及有效會員之授權碼,方可取得各式開放資料。其中,資料項目代碼可至資料清單列表查詢。
	一、取得授權碼
	會員之授權碼可於下方按鈕取得
	取得授權碼
	二、更新授權碼
	一旦更新授權碼後,舊的授權碼將永久失效,並且更新授權碼後七日內無法再進行更新。

1. 下載並解壓縮 CWB 官方檔案

https://opendata.cwb.gov.tw/fileapi/v1/opendataapi/O-B0075-002?Authorization=CWB-xxxxxxx-xxxx-xxxx-xxxx-xxxx&downloadType=WEB&format=ZIP

以上為開放資料的完整網址,記得 xxx-xxx-xxx-xxx 換成真正的授權碼

2. 轉換原始檔案並擷取資料

(2) convert xml(decompressed files) into csv

```
files path = FOLDER XML + '*.xml'
                                                                   收集 xml 輸入檔案
filenames source = glob.glob(files path)
for filename source in filenames source:
                                                                   命名 csv 輸出檔名
   filename raw = os.path.basename(filename source).split('.')[0]
   filename csv = f'{FOLDER CSV}{filename raw}.csv'
   with open(filename source, mode='rb') as f:
       # get a list of 30-day records & every record is a dict with hierarchy
       my dict = xmltodict.parse(f)
                                                                               轉成 python 程式語言
       a = my dict['cwbdata']['resources']['resource']['data']['seaSurfaceObs']
                                                                               原生的資料格式並查表
       records by month = a['location']['stationObsTimes']['stationObsTime']
   # get header names from any day
   csv columns = list(flatten dict(records by month[0]).keys())
                                                                 牛成 csv 檔案需要的檔頭
   try:
       with open(filename csv, 'wt', newline='') as csvfile:
           writer = csv.DictWriter(csvfile, fieldnames=csv columns)
                                                                    產牛 csv 檔案:
           writer.writeheader()
           # records by month[0] = the latest (newest) record
                                                                    寫檔頭
           for record in records by month[::-1]:
                                                                    一行接著一行寫入(一行=一天)
               writer.writerow(flatten dict(record))
   except IOError:
       print("I/O error :", filename raw)
print('2. convert xml into csv (complete)')
```

3. 壓縮並備份資料

```
# (3) compress and backup 30-day csv files
```

```
now = time.localtime()
date = '{} {:02d}{:02d}'.format(now.tm year,now.tm mon,now.tm mday)
time = '{:02d}{:02d}'.format(now.tm hour.now.tm min.now.tm sec)
                                                                         即時生成檔案名稱
FILE BACKUP = f'cwb mmc lmonth {date} {time}.zip'
compression = zipfile.ZIP DEFLATED
                                    設定最高壓縮率
compresslevel = 9
with zipfile.ZipFile(FOLDER BACKUP + FILE BACKUP, mode='w',
               compression=compression, compresslevel=compresslevel) as myzip:
   FILES PATH = FOLDER CSV + '*.csv'
                                                                       產生 zip 壓縮檔:
   filenames current = glob.glob(FILES PATH)
                                                                       收集要被寫入的檔案 (*.csv)
   for filename current in filenames current:
                                                                       寫入不包含路徑的檔案
       filename without path = os.path.basename(filename current)
       myzip.write(filename current, arcname=filename without path)
print('3. back up csv (comlete)')
```

4. 合併新舊資料

(4) merge csv files and discard additional duplicates

```
files path = FOLDER CSV + '*.csv'
filenames new = glob.glob(files path)
                                                      收集最新的 csv 檔案
for filename new in filenames new:
   filename no path = os.path.basename(filename new)
                                                           準備舊的 csv 檔案
   filename merge = f'{FOLDER MERGE}{filename no path}'
   df new = pd.read csv(filename new)
   #df_new = pd.read csv(filename new, na values='None')
   try:
       df old = pd.read csv(filename merge)
                                                                 將新和舊的 csv 檔案
       #df old = pd.read csv(filename merge, na values='None')
                                                                 轉成 pandas 的資料結構
   except:
       df old = None
   full df = pd.concat([df old, df new])
   unique df = full df.drop duplicates(subset='dataTime')
   #unique df.to csv(filename merge) # generate additional column of indices
                                                                               根據 dataTime 欄位
   unique df.to csv(filename merge, index=False)
                                                                               來合併新和舊的紀錄
                                                                               並生成合併好的 csv
print('4. merge csv (complete)')
```