

Python 程式教學 Chapter 24

- · 網頁爬蟲v4
- 台鐵車次資料查詢
- 使用API+POST



台鐵網站訂票系統

https://www.railway.gov.tw/tra-tip-web/tip



車種車次	出發時間▲	抵達時間	行駛時間 ▲	經由	詳細資訊	全票 ▲	優待票 ▲	訂票
☑ 莒光1	06:10	10:55	4 小時 45 分	山線	~	\$ 635	\$ 318	
員 自強103	06:30	11:19	4 小時 49 分	海線	~	\$ 824	\$ 412	
■ 復興5581	06:40	12:48	6 小時 8 分	海線	~	\$ 530	\$ 265	
員 自強105	07:00	11:54	4 小時 54 分	山線	~	\$ 824	\$ 412	
₩ 莒光507	07:13	14:06	6 小時 53 分	海線	~	\$ 635	\$ 318	
□ 莒光2	08:10	16:33	8 小時 23 分	山線	~	\$ 898	\$ 449	
員 自強113	08:28	13:04	4 小時 36 分	山線	~	\$ 824	\$ 412	4
員 自強115	08:59	13:42	4 小時 43 分	山線	~	\$ 824	\$ 412	*
員 莒光509	09:17	15:25	6 小時 8 分	海線	~	\$ 635	\$ 318	*
員 自強117	10:00	14:40	4 小時 40 分	山線	~	\$ 824	\$ 412	*
員 自強121	11:00	15:42	4 小時 42 分	山線	~	\$ 824	\$ 412	*
員 莒光513	11:18	18:03	6 小時 45 分	海線	~	\$ 635	\$ 318	4
員 自強123	11:58	16:39	4 小時 41 分	山線	~	\$ 824	\$ 412	4
員 自強125	12:59	17:41	4 小時 42 分	山線	~	\$ 824	\$ 412	*
員 莒光561	13:40	20:25	6 小時 45 分	海線	~	\$ 635	\$ 318	4



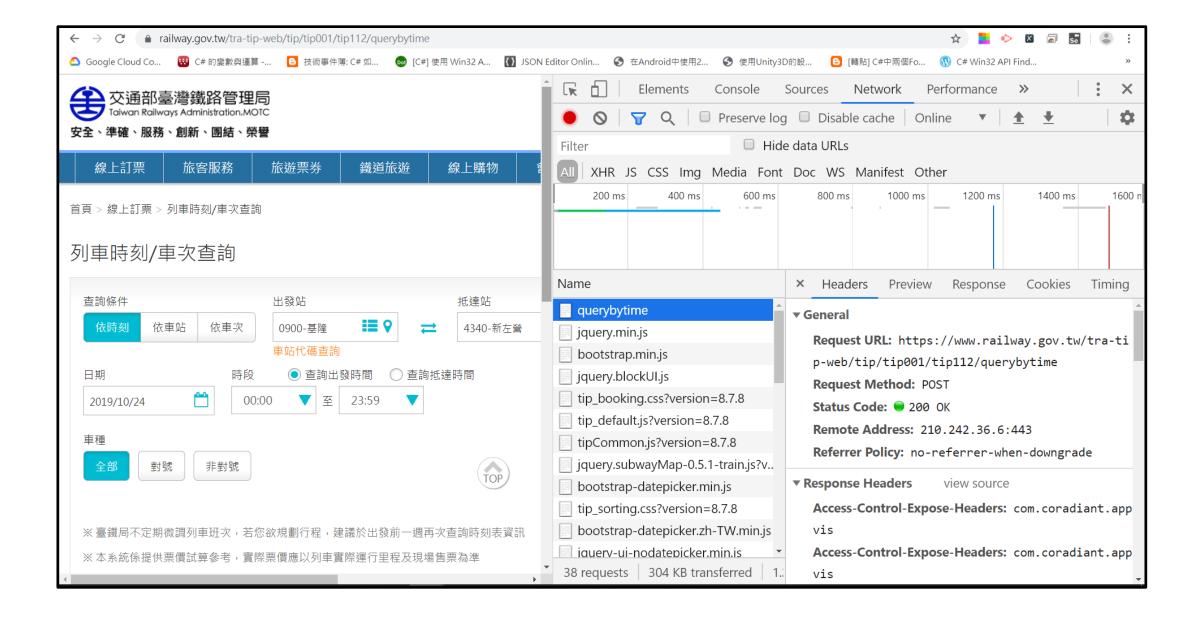
程式將透過觀察

- 取得所有站點代號
- 觀察傳送表單資料
- ■取得Crsf網站安全代碼
- 使用post發出回傳
- 收到資料後進行分析

什麼是"crsf"

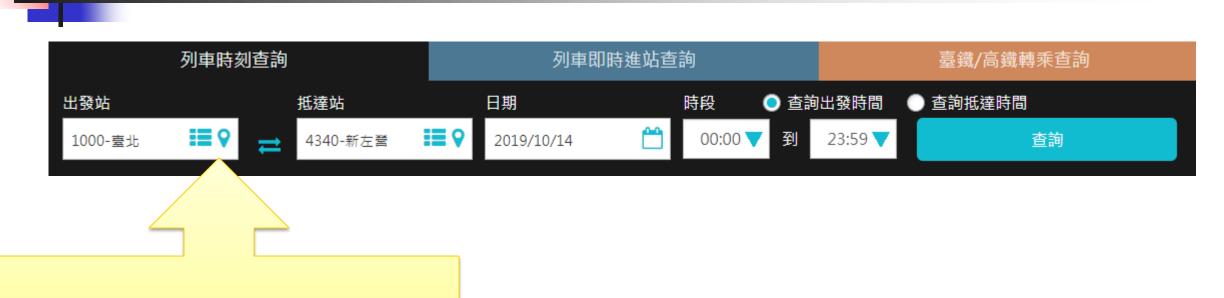
找出post用api







自動傳送表單資料



取得站點資料

■ 開啟開發人員工具(點選常用→站點名稱)





■ 站點名稱區段在ul→li→button





```
▼
 ▶ ...
 ▼<1i>>
    <button type="button" class="btn tipStation btn-</pre>
    info" style="color:#fff;" title="1020-板橋" data-
    type="startStation">板橋</button> == $0
   ▼<1i>>
    <button type="button" class="btn tipStation"</pre>
    title="1030-浮洲" data-type="startStation">浮洲
    </button>
   ▼<1i>>
    <button type="button" class="btn tipStation btn-</pre>
    info" style="color:#fff;" title="1040-樹林" data-
    type="startStation">樹林</button>
```



取得站點名稱

- 使用id="cityHot"
 - find ul
 - find_all li
 - find button取得text
 - find button取得屬性 title

```
▼<div class="line-inner hr cityHr" id="cityHot"
style="display:show"> == $0
 <div class="line-head">...</div>
 ▼<u1>
   ▶ li>...
   ▼>
      <button type="button" class="btn tipStation"
      title="0930-七堵" data-type="startStation">
      七堵</button>
    ▶ :...
   ▼>
      <button type="button" class="btn tipStation"</pre>
      title="1000-臺北" data-type="startStation">
      臺北</button>
    ▶ ...
   ▶...
```

```
import requests
from bs4 import BeautifulSoup
def getStation(url):
     global station
     resp=requests.get(url)
    if resp.status code!=200:
        return 'error'
     soup=BeautifulSoup(resp.text,'lxml')
     citys=soup.find(id='cityHot').find('ul').find all('li')
    for city in citys:
        name=city.find('button').text
        code=city.find('button')['title']
         station[name]=code
     print('取得站點資料ok')
url='https://www.railway.gov.tw/tra-tip-web/tip'
station={}
getStation(url)
print(station)
ch24 1.py
```



```
'基隆': '0900-基隆',
'七堵': '0930-七堵',
'南港': '0980-南港',
'松山': '0990-松山',
'臺北': '1000-臺北',
'萬華': '1010-萬華',
'板橋': '1020-板橋',
'樹林': '1040-樹林',
'桃園': '1080-桃園',
'中壢': '1100-中壢',
'新竹': '1210-新竹',
'竹南': '1250-竹南',
'苗栗': '3160-苗栗',
'豐原': '3230-豐原',
'臺中': '3300-臺中',
'彰化': '3360-彰化',
'員林': '3390-員林',
'斗六': '3470-斗六',
'嘉義': '4080-嘉義',
'新營': '4120-新營',
'臺南': '4220-臺南',
```



檢視查詢按鈕(Elements)

- 檢視查詢按鈕
- 取得傳送網址
- 取得csrf

▼<form id="queryForm" action="/tra-tip-web/tip/tip001/tip112/ querybytime" method="post"> == \$0

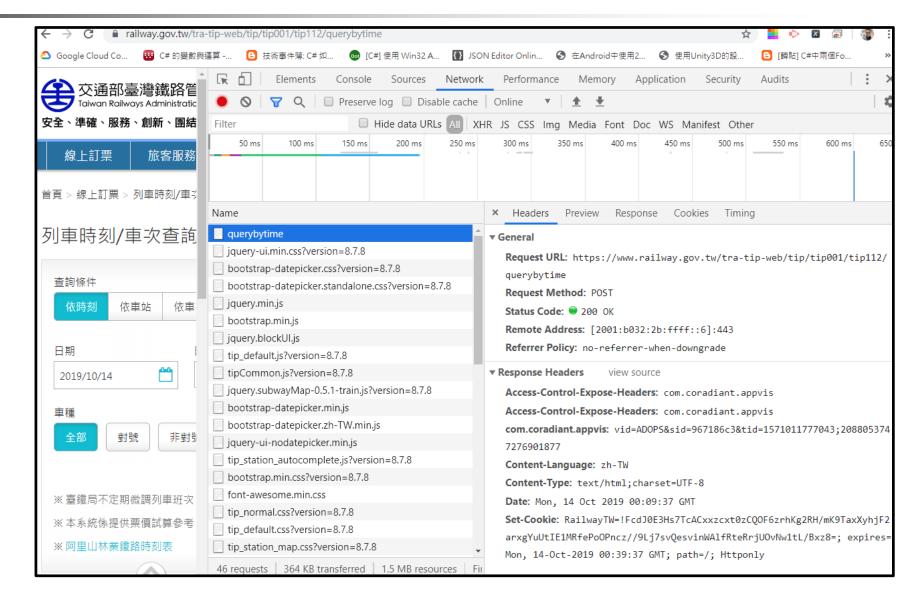
```
▼<form id="queryForm" action="/tra-tip-web/tip/tip001/tip112/
querybytime" method="post">
   <input type="hidden" name="_csrf" value="2d5ce3c7-ead7-469b-8f86-</pre>
   0fdae60980fb">
   <input type="hidden" name="trainTypeList" value="ALL">
   <input type="hidden" name="transfer" value="ONE">
  <div class="zone site-wrapper">...</div>
  <div class="zone exchange-margin">...</div>
  ▼<div class="zone site-wrapper">
    ▶ <div class="form title">...</div>
    ▼<div class="input-wrapper">
       <input id="endStation" type="text" name="endStation" data-</pre>
       autocomplete="station" placeholder="抵達站" title="抵達站"
       class="ui-autocomplete-input" autocomplete="off">
     ▶<button type="button" href="#" title="文字站點杳詢" class="icon
     icon-list" data-type="endStation">...</button>
     ▶<button type="button" href="#" title="圖片站點查詢" class="icon
     icon-map" data-type="endStation">...</button>
     </div>
   </div>
  ▶ <div class="zone date-wrapper">...</div>
  <div class="zone time-wrapper">...</div>
  ▼<div class="zone v-bottom">
     <input class="btn btn-basic btn-lg" type="submit" value="查詢"</pre>
     title="杳詢"> == $0
```



觀察傳送表單資料(Network)

■ A11→

querybytime



■ 觀察querybytime → Headers



▼ General

Request URL: https://www.railway.gov.tw/tra-tip-web/tip/tip001/tip112/querybytime Request Method: POST Status Code: 200 OK 需傳送的網址(post) Remote Address: [2001:b032:2b:ffff::6]:443 Referrer Policy: no-referrer-when-downgrade ▼ Form Data view URL encoded view source csrf: 2d5ce3c7-ead7-469b-8f86-0fdae60980fb

trainTypeList: ALL

transfer: ONE

startStation: 1000-臺北

endStation: 4340-新左營

rideDate: 2019/10/14

startOrEndTime: true

startTime: 00:00

endTime: 23:59

需傳送的表單資料

■取得傳送網址(post)

python

■ 取得crsf code

```
resp=requests.get(url)
if resp.status_code==200:
    soup=BeautifulSoup(resp.text,'lxml')
    request_url=soup.find(id='queryForm')['action']
    crsf_code=soup.find(id='queryForm').find('input',{'name':'_csrf'})['value']

print(request_url,crsf_code,sep='\n')
```

/tra-tip-web/tip/tip001/tip112/querybytime 906a9867-7b79-4d84-8322-028c9fa9bf2e

```
<input type="hidden" name="_csrf" value="2d5ce3c7-ead7-469b-8f86-
0fdae60980fb">
```

```
import requests
from bs4 import BeautifulSoup
def getStation(url):
    global station
    resp=requests.get(url)
    if resp.status code!=200:
        return 'error'
    soup=BeautifulSoup(resp.text,'lxml')
    citys=soup.find(id='cityHot').find('ul').find all('li')
    for city in citys:
        name=city.find('button').text
        code=city.find('button')['title']
        station[name]=code
    print('取得站點資料ok')
    request url=soup.find(id='queryForm')['action']
    crsf_code=soup.find(id='queryForm').find('input',{'name':'_csrf'})['value']
    #crsf code=soup.find(id='queryForm').find('input')['value']
    return request_url,crsf_code
url='https://www.railway.gov.tw/tra-tip-web/tip'
station={}
request_url,crsf_code=getStation(url)
print(station)
print(request url,crsf code)
```



取得站點資料ok

```
{'基降': '0900-基降', '七堵': '0930-七堵',
'南港': '0980-南港', '松山': '0990-松山',
'臺北': '1000-臺北', '萬華': '1010-萬華',
'板橋': '1020-板橋', '樹林': '1040-樹林',
'桃園': '1080-桃園', '中壢': '1100-中壢',
'新竹': '1210-新竹', '竹南': '1250-竹南',
'苗栗': '3160-苗栗', '豐原': '3230-豐原',
'臺中': '3300-臺中', '彰化': '3360-彰化',
'昌林': '3390-昌林', '斗六': '3470-斗六',
'嘉義': '4080-嘉義', '新愛': '4120-新愛',
'臺南': '4220-臺南', '岡山': '4310-岡山',
'新左營': '4340-新左營', '高雄': '4400-高
雄', '屏東': '5000-屏東', '潮州': '5050-潮
州', '臺東': '6000-臺東', '玉里': '6110-玉
里', '花蓮': '7000-花蓮', '蘇澳新': '7130-
蘇澳新','宜蘭':'7190-宜蘭','瑞芳':
'7360-瑞芳'}
/tra-tip-web/tip/tip001/tip112/
querybytime 2a1d0bfc-bd0d-465f-a3ea-
fb3a17297af1
```



組成表單資料

▼ Form Data view source view URL encoded

_csrf: 2d5ce3c7-ead7-469b-8f86-0fdae60980fb

trainTypeList: ALL

transfer: ONE

startStation: 1000-臺北

endStation: 4340-新左營

rideDate: 2019/10/14

startOrEndTime: true

startTime: 00:00

endTime: 23:59

```
form_data={
    '_csrf': 296ad198-7acb-499b-96ba-d8a163920b3c
    'trainTypeList': 'ALL'
    'transfer': 'ONE'
    'startStation': 1000-臺北
    'endStation': 4340-新左營
    'rideDate': 2019/10/14
    'startOrEndTime': 'true'
    'startTime': 00:00'
    'endTime': 23:59
```



設定時間/起訖站點

import time

```
today=time.strftime('%Y/%m/%d')
sTime='00:00'
eTime='23:59'
start_station=station['臺北']
end_station=station['新左營']
```

```
form data={
    ' csrf': crsf code,
    'trainTypeList': 'ALL',
    'transfer': 'ONE',
    'startStation': start station,
    'endStation': end station,
    'rideDate': today,
    'startOrEndTime': 'true',
    'startTime': sTime,
    'endTime': eTime
```



進行表單傳送

- 使用requests. post
- 組成url

resp=requests.post('https://www.railway.gov.tw'+request_url,data=form_data)



傳送表單



進行資料擷取

車種車次	出發時間▲	抵達時間	行駛時間 ▲	經由	詳細資訊	全票 ▲	優待票 ▲
☑ 莒光1	06:10	10:55	4 小時 45 分	山線	~	\$ 635	\$ 318
⊋ 自強103	06:30	11:19	4 小時 49 分	海線	~	\$ 824	\$ 412
⊋ 復興5581	06:40	12:48	6 小時 8 分	海線	~	\$ 530	\$ 265
⊋ 自強105	07:00	11:54	4 小時 54 分	山線	~	\$ 824	\$ 412
⊋ 莒光507	07:13	14:06	6 小時 53 分	海線	~	\$ 635	\$ 318
⊋ 莒光2	08:10	16:33	8 小時 23 分	山線	~	\$ 898	\$ 449
⊋ 自強113	08:28	13:04	4 小時 36 分	山線	~	\$ 824	\$ 412
⊋ 自強115	08:59	13:42	4 小時 43 分	山線	~	\$ 824	\$ 412
⊋ 莒光509	09:17	15:25	6 小時 8 分	海線	~	\$ 635	\$ 318
⊋ 自強117	10:00	14:40	4 小時 40 分	山線	~	\$ 824	\$ 412
⊋ 自強121	11:00	15:42	4 小時 42 分	山線	~	\$ 824	TOP
⊋ 莒光513	11:18	18:03	6 小時 45 分	海線	~	\$ 635	\$ 318
☑ 自強123	11:58	16:39	4 小時 41 分	山線	~	\$ 824	\$ 412

■ 資料區段在tr class="trip-column"



■ 詳細資料在td

```
▼ == $0
  ▼
   ▼>
     ▼<span class="icon-fa icon-train chukuang">
       ::before
      </span>
      <a href="/tra-tip-web/tip/tip001/tip112/querybytrainno?rideDate=2019/10/</pre>
      14&trainNo=1" target="_blank" class="links" title="列車時刻表(另開新視窗)"> 営光1
      </a>
    06:10
  10:55
  4 小時 45 分
   td>山線
```

```
soup=BeautifulSoup(resp.text,'lxml')
trs=soup.find('table').find('tbody').find_all('tr',class_='trip-column')

for tr in trs:
    tds=tr.find_all('td')
    print(tds[0].text.strip(),tds[1].text,tds[2].text,tds[3].text)
```



莒光1 06:10 10:55 4 小時 45 分 自強103 06:30 11:19 4 小時 49 分 復興5581 06:40 12:48 6 小時 8 分 自強105 07:00 11:54 4 小時 54 分 莒光507 07:13 14:06 6 小時 53 分 莒光2 08:10 16:33 8 小時 23 分 自強113 08:28 13:04 4 小時 36 分 自強115 08:59 13:42 4 小時 43 分 莒光509 09:17 15:25 6 小時 8 分 自強117 10:00 14:40 4 小時 40 分 自強121 11:00 15:42 4 小時 42 分 莒光513 11:18 18:03 6 小時 45 分 自強123 11:58 16:39 4 小時 41 分 自強125 12:59 17:41 4 小時 42 分 莒光561 13:40 20:25 6 小時 45 分 自強129 14:00 18:32 4 小時 32 分 自強135 15:00 19:41 4 小時 41 分 自強175 16:00 20:44 4 小時 44 分 莒光521 16:04 22:48 6 小時 44 分 自強139 17:00 21:34 4 小時 34 分 自強141 17:30 22:14 4 小時 44 分 自強145 18:10 23:09 4 小時 59 分 自強149 19:00 23:39 4 小時 39 分 自強181 19:30 00:04 4 小時 34 分



程式練習(找尋是否還能購買車票)

■ 在最後一個td(如果為空則為無法訂票)

4 小時 40 分	山線	~	\$ 824	\$ 412	<i><</i>
4 小時 42 分	山線	~	\$ 824	\$ 412	<i><</i>
6 小時 45 分	海線	~	\$ 635	\$ 318	<i><</i>
4 小時 41 分	山線	~	\$ 824	\$ 412	4
4 小時 42 分	山線	~	\$ 824	\$ 412	4
6 小時 45 分	海線	~	\$ 635	\$ 318	OP P
4 小時 32 分	山線	~	\$ 824	\$ 412	ØP Ø
4 小時 41 分	山線	~	\$ 824	\$ 412	<i><</i>
4 小時 44 分	山線	~	\$ 824	\$ 412	<i>\(\phi \)</i>

```
$ 824
 $ 412
▼
 ▼<form action="/tra-tip-web/tip/tip001/tip117/tra2traUrl" method="post" target=
 " blank">
     <input type="hidden" name=" csrf" value="7a66f92b-b5a3-4430-adde-</pre>
     011b68c1cb6e">
     <input type="hidden" name="queryDate" value="2019/10/14">
                                                <input type="hidden"</pre>
    th:name="transferList[__${trainStat.index}__].trainNo"
    th:value="${train.trainNo}"> -->
     <input type="hidden" name="transferList[0].startStation.stationCode" value=</pre>
     "1000">
     <input type="hidden" name="transferList[0].endStation.stationCode" value=</pre>
     "4340">
     <input type="hidden" name="departureTime" value="2019-10-14 10:00:00.0">
     <input type="hidden" name="transferList[0].trainType.code" value="3">
     <input type="hidden" name="transferList[0].trainType.isReserve" value="true">
     <input type="hidden" name="transferList[0].trainNo" value="117">
   ▼<button type="submit" class="icon icon-ticket" title="訂票">
       ::before == $0
       "訂票"
     </button>
```

```
for tr in trs:
    tds=tr.find_all('td')
    if tds[-1].text.strip()!='':
        print(tds[0].text.strip(),tds[1].text,tds[2].text,tds[3].text,
        tds[-1].text.strip())
```



```
自強121 11:00 15:42 4 小時 42 分 訂票 莒光513 11:18 18:03 6 小時 45 分 訂票 自強123 11:58 16:39 4 小時 41 分 訂票 自強125 12:59 17:41 4 小時 42 分 訂票 莒光561 13:40 20:25 6 小時 45 分 訂票 自強129 14:00 18:32 4 小時 32 分 訂票 自強135 15:00 19:41 4 小時 44 分 訂票 自強175 16:00 20:44 4 小時 44 分 訂票 自強139 17:00 21:34 4 小時 34 分 訂票 自強141 17:30 22:14 4 小時 59 分 訂票 自強145 18:10 23:09 4 小時 59 分 訂票 自強149 19:00 23:39 4 小時 34 分 訂票 自強149 19:00 23:39 4 小時 34 分 訂票 自強141 17:30 00:04 4 小時 34 分 訂票
```



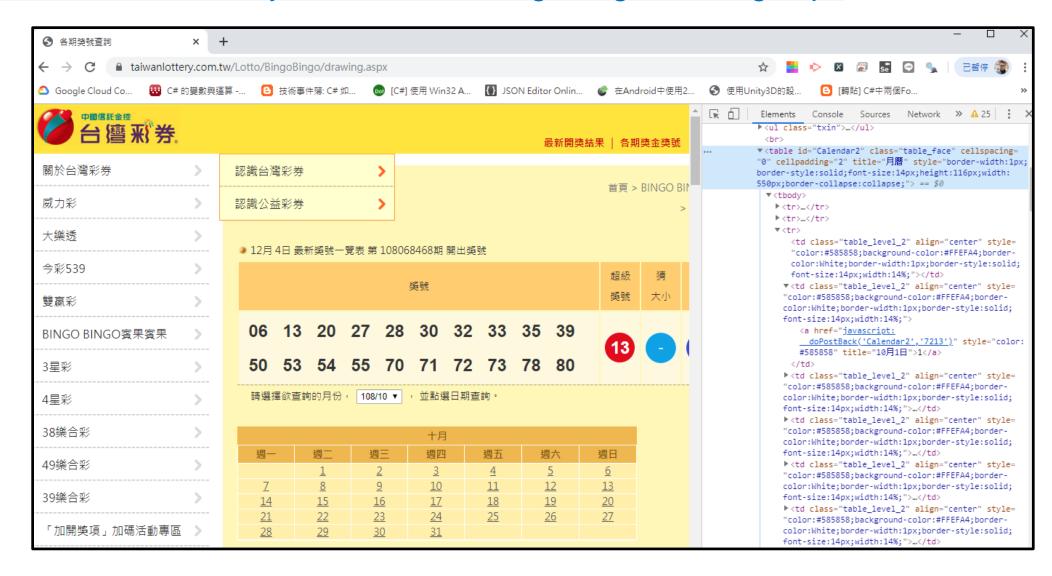
後續運用

■ 指定車次/日期,完成半/自動訂票

■ 程式練習(bingobingo開獎號碼爬找)



https://www.taiwanlottery.com.tw/Lotto/BingoBingo/drawing.aspx



• 依照月/日進行查詢

十二月 38 集台彩 星期三 星期四 星期五 星期六 星期日 遊戲介紹 1 各期獎號與開獎結果 6 7 8 開獎時間 11 13 14 15 18 20 21 22 冷熱門球號 26 27 28 29



使用api+post 切換月跟日

12月16日 獎號一覽表 依大小順序排序 ▼ 顯示當日所有期數

獎號一覽表僅顯示當日最新96期之獎號,如欲顯示當日所有期數之獎號,請點選上方『顯示當日所有期數』按鈕進 行查詢。

爬找每天的開獎 號碼

期別	獎號	超級獎號	猜大小
102071248	04 07 10 28 30 31 32 33 34 44 50 56 57 61 65 66 70 78 79 80	70	-
102071247	07 08 11 12 19 24 25 34 39 43 44 46 47 55 57 60 67 71 79 80	60	-
102071246	02 04 07 10 12 23 27 28 32 38 40 46 48 54 58 61 62 67 70 80	70	-
102071247	07 08 11 12 19 24 25 34 39 43 44 46 47 55 57 60 67 71 79 80	60	-
102071246	02 04 07 10 12 23 27 28 32 38 40 46 48 54 58 61 62 67 70 80	60	-
102071248	04 07 10 28 30 31 32 33 34 44 50 56 57 61 65 66 70 78 79 80	70	-
102071248	04 07 10 28 30 31 32 33 34 44 50 56 57 61 65 66 70 78 79 80	70	-
102071248	04 07 10 28 30 31 32 33 34 44 50 56 57 61 65 66 70 78 79 80	70	-
102071246	02 04 07 10 12 23 27 28 32 38 40 46 48 54 58 61 62 67 70 80	60	-
102071246	02 04 07 10 12 23 27 28 32 38 40 46 48 54 58 61 62 67 70 80	60	-