

Python 與 AI 人工智慧開發入門

第 16 堂：資料探索分析拆組

同學，歡迎你參加本課程

- ☑ 請關閉你的FB、Line等溝通工具，以免影響你上課。
- ☑ 考量頻寬，請預設關閉麥克風、攝影機，若有需要再打開。
- ☑ 隨時準備好，老師會呼叫你的名字進行互動。
- ☑ 如果有緊急事情，你必需離開線上教室，請用聊天室私訊給老師，以免老師癡癡呼喚你的名字。
- ☑ 先倒好水、上個洗手間，準備上課囉^^

課程檔案下載

The screenshot displays the Pcschoolonline website interface. The top navigation bar includes links for '巨匠電腦線上真人' (Giant Master Computer Online Real Person), '開課查詢' (Class Inquiry), '免費體驗專區' (Free Experience Area), '課程總覽' (Course Overview), '專業師資' (Professional Faculty), '學員專區' (Student Area), '講師專區' (Instructor Area), and '最新消息' (Latest News). Social media icons for 360, Facebook, and YouTube are also present. A user is logged in, indicated by '您好!' and a '登出' (Logout) button.

The main banner features the text: **程式語言 好難學?** (Programming Language is so hard to learn?), **那是因為 你還沒學過Python!** (That's because you haven't learned Python!), and **(線上老師 LIVE 直播教學 · 搶先看)** (Online Teacher LIVE Broadcast Teaching · Preview). Below the banner, it says '巨匠電腦真人課程' (Giant Master Computer Real Person Course).

A dropdown menu is open from the '學員專區' (Student Area) link, listing various options: '點數卡產品兌換' (Points Card Product Exchange), 'APCS檢測專區' (APCS Detection Special Area), '公告專區' (Announcement Special Area), '我的課表' (My Class Schedule), 'IT真人課程動位' (IT Real Person Course Positioning), '電腦分校課程動位' (Computer Branch Course Positioning), '外語真人課程動位' (Foreign Language Real Person Course Positioning), '美語分校課程動位' (American English Branch Course Positioning), '取消動位' (Cancel Positioning), '課程檔案下載' (Course Archive Download), '上課權益查詢' (Class Benefit Inquiry), '教學平台測試' (Teaching Platform Test), '學習諮詢' (Learning Consultation), '常見問題' (Common Questions), '個資維護' (Personal Information Maintenance), '忘記密碼' (Forgot Password), and '登出' (Logout). The '課程檔案下載' (Course Archive Download) option is highlighted with an orange box, and a speech bubble points to it with the text '課程檔案下載'.

ZOOM 學員操作說明

The screenshot shows the Zoom web client interface. At the top, a green banner reads '您正在觀看蘇世榮的螢幕' (You are watching Su Shihong's screen). Below this is a toolbar with icons for '游鼠' (Cursor), '文字' (Text), '筆' (Pen), '鬚鬚' (Erase), '擦除' (Erase), '格式' (Format), '撤銷' (Undo), '重做' (Redo), and '清除' (Clear). A dropdown menu is open for the '筆' icon, showing options: '原始大小' (Original size), '請求遠端控制' (Request remote control), '共同註記' (Annotate), and '退出全螢幕' (Exit full screen). An arrow points from the '共同註記' option to the '筆' icon in the toolbar.

On the left, a sidebar shows a list of participants under the heading '與會者 (15)'. The list includes '張齡月 (我)' and '婷婷'. A '解靜音' (Unmute) button is visible next to the participants. An arrow points from the '舉手' (Raise hand) button in the participant list to the '與會者' (Participants) button in the bottom toolbar.

The bottom toolbar contains several buttons: '解除靜音' (Unmute), '啟動視訊' (Start video), '邀請' (Invite), '與會者' (Participants), '共享螢幕' (Share screen), '聊天' (Chat), and '錄影' (Record). Arrows point from the numbered annotations to these buttons.

1 聊天 (Chat) - Points to the '聊天' button in the bottom toolbar.

2 共享螢幕 (Share screen) (指導演練；點評作品)
老師須先停止共享螢幕
才能請學生共享螢幕 - Points to the '共享螢幕' button in the bottom toolbar.

3 與會者/舉手 (Participants/Raise hand) - Points to the '與會者' button in the bottom toolbar.

4 解除靜音 (Unmute) - Points to the '解除靜音' button in the bottom toolbar.

5 查看選項/共同註記/筆 (連連看) (View options/Annotate/Pen) - Points to the '共同註記' option in the dropdown menu.

www.pcschoolonline.com.tw

本堂教學重點

1. 正規搜尋切割取代
2. row 的分合
3. column 分合
4. 可選擇行列分合
5. 不等量分割

預告今天作品：檔案讀取查詢

- 可從政府資料開放平台取得資料，例如：
<https://data.gov.tw/dataset/5958>
- 請載入 csv 檔案內容並設定查詢 (例如
是否可以輸入郵遞區號 列出所有的派出所?)

請輸入郵遞區號104

```
['中山分局', 'Zhongshan Precinct', '104', '臺北市中山區']  
['中山一派出所', 'Zhongshan First Police Station', '104', '臺北市中山區']  
['中山二派出所', 'Zhongshan Second Police Station', '104', '臺北市中山區']  
['圓山派出所', 'Yuanshan Police Station', '104', '臺北市中山區']  
['長春路派出所', 'Changchun Rd. Police Station', '104', '臺北市中山區']  
['長安東路派出所', 'Changan E. Rd. Police Station', '104', '臺北市中山區']
```

課程內容

正規搜尋切割取代

- 規則定義
- findall
- split
- sub

numpy陣列合併與拆解

- row 的分合
- column 分合
- 可選擇行列分合
- 不等量分割

課程內容

正規搜尋切割取代

- 規則定義
- findall
- split
- sub

numpy陣列合併與拆解

- row 的分合
- column 分合
- 可選擇行列分合
- 不等量分割

關於跳脫字元符號

- ◆ 由於正規表示式使用\符號針對特殊字元進行轉換，例如若想要查詢這個字串 `'pcschool.com'`，我們使用的正規表示式方式為 `'pcschool\\.com'`。
- ◆ Python 的字串也是使用\符號進行特殊字元轉換，上述的正規表示式於 Python 內可寫成 `'pcschool\\.com'`，如此寫法會有很多\符號出現的困擾。
- ◆ 建議可於字串前面加上 r 的前導符號就可以，所以上面的範例可改為：
`'r'python\\.com'`

findall 方法

- ◆ **findall**(搜尋字串,起始位置,結束位置)
 - ◆ 起始位置與結束位置是可選擇的參數，預設值分別為 0 與字串長度。
 - ◆ **findall** 將以 **list** 方式傳回所有可以比對的資料，若找不到可以比對的資料則會傳回一個空的 **list**。

設定跳脫與首尾

◆ 設定跳脫與首尾

符號	說明	表達式	符合條件字串
.	除了\n換行之外任意的字元	.	Abc
\	跳脫字元，使後面的一個字元改變原本的涵義	a\b	a,b
		a\\d	a\d
^	符合開頭	^ab	abc
\$	符合結尾	c\$	c

設定的符號定義

◆ 使用設定的符號定義類型

符號	說明	表達式	符合條件字串
\d	代表數字 [0-9]	a\dc	a2c
\D	代表非數字 [^\d]	a\Dd	azd
\s	代表空白字元	a\s d	a d
\S	代表不是空白字元	a\Sd	axd
\w	等於 [a-zA-Z_0-9] 數字或是英文字	a\wf	a0f
\W	等於 [^a-zA-Z_0-9] 非數字與英文字	a\Wf	a f

操作範例：搜尋數值資料

★操作檔案「re-001.py」

```
import re
s = "Tim's phone numbers are 12345-41521 and 78963-85214"
match = re.findall(r'\d', s)
if match:
    print(match)
```

計算字元數

◆ 定義字元數

符號	說明	表達式	符合條件字串
{3}	需輸入 3 個字元	[A-C]{3}	ABC BCA
{1,3}	需輸入 1 ~ 3 個字元	[A-C]{1,3}	AB B CAB
{4,}	需輸入 4 到無限多個字元	[A-C]{4,}	AABBC CCAA

操作範例：搜尋數值資料

★操作檔案「re-002.py」

```
import re
s = "Tim's phone numbers are 12345-41521 and 78963-85214"
match = re.findall(r'\d{5}', s)
if match:
    print(match)
```

依據尋找的字元範圍

◆ 針對要尋找的字元範圍

符號	說明	表達式	符合條件字串
[a-z]	表示可接受 a~z 等 26 個英文小寫字母	a[bc]d	abd acd
[^ b-e]	表示除了 b~e 之外，所有字元均接受	a[^ b-e]d	afd axd

操作範例：依據尋找的字元範圍

★操作檔案「re-003.py」

```
#ptn = r"r[au]n " #空格是否會影響？  
#ptn = r"r[au]n"  
#ptn = r"r[a-u]n"  
#ptn = r"r[a-u]{1,3}n"  
print(re.findall(ptn, "dog ran runs to cat rain rin"))
```

操作範例：依據尋找的字元範圍

★操作檔案「re-004.py」

```
import re
string1="1dog cat2 3rabbit "
print(re.findall(r"[A-Z]",string1))
print(re.findall(r"[a-c]",string1))
print(re.findall(r"[0-9]",string1))
print(re.findall(r"^[a-c]",string1))
```

計算重複數量

◆ 計算重複數量

- 注意乘號和加號都是以前一字元為目標。
- 例如 **a+b**，效果會在 **a** 身上，也就是 **a** 至少一次，**b** 只有一次。

符號	說明	表達式	符合條件字串
?	代表前面的項目重複 0 次或 1 次	abcd?	abc abcd
*	代表前面的項目重複 0 ~ 無限次	abcd*	abc abcddd
+	代表前面的項目重複 1 ~ 無限次	abcd+	abcd abcddd

操作範例 1：使用 findall 方法搜尋

★操作檔案「re-005.py」

```
import re
pattern = re.compile(r'\d+')
result1 = pattern.findall('hello 123456 789')
result2 = pattern.findall('one1two2three3four4', 0, 10)
print(result1 )
print(result2 )
```

操作範例 2：搜尋重複數量

★操作檔案「re-006.py」

```
import re
str1="abcd abcd abcb abc"
print(re.findall(r"abcd?", str1))
print(re.findall(r"abcd*", str1))
print(re.findall(r"abcd+", str1))
```

```
print(re.findall(r"[a-c]{3}", str1))
print(re.findall(r"[a-c]{1,4}", str1))
print(re.findall(r"[a-c]{4,}", str1))
```



Q . 請問執行後的結果是哪幾個？ (選擇題)

```
import re
string1="C++ Python Java C# Sqlite Django Numpy Pandas "
print(re.findall(r"[p-t]",string1))
```

(1) ['P', 'q', 't', 'p', 's']

(3) ['t', 'q', 't', 'p', 's']

(2) ['q', 'p', 's']

(4) ['P', 't', 'q', 't', 'p', 's']

split 方法

- ◆ split 方法能以比對規則針對字串進行切割後產生 list。

- ◆ split 它的使用形式如下：

split(字串,最大分割次數)

- ◆ 若不指定最大分割次數則全部分割

操作範例：使用 split 方法

```
import re
text = 'one, two...ten'
re1=re.split('[,. ]+', text)
print(re1)
re1=re.split('[,. ]+', text, maxsplit=1)
print(re1)
```

★操作檔案「re-007.py」

sub 方法

- ◆ sub 方法用於替換，它的使用形式如下

sub(取代的字串或函數,來源字串,可設定被取代的數量)

- ◆ 可設定被取代的數量若不指定時全部替換。

思考：字串取代

- ◆ `inputStr = "hello 111 world 111"`
- ◆ `replacedStr = inputStr.replace("111", "222")`
- ◆ 可是假設所有數字都想換成 222，而字串內的數值可能改變，例如
`inputStr = "hello 123 world 456"`
- ◆ 那該如何進行？

操作範例：字串替換

★操作檔案「re-008.py」

```
import re
inputStr = "hello 123 world 456"
replacedStr = re.sub("\d+", "222", inputStr)
print(replacedStr)
replacedStr = re.sub("\d+", "222 ", inputStr, count=1)
print(replacedStr)
```



Q . 請問執行後的結果是哪幾個？ (選擇題)

```
import re
inputStr = "1970,C language,1990,python"
replacedStr = re.sub("\d+", "2018", inputStr)
print(replacedStr)
```

(1) 2018,2018,2018,python

(3) 2018,C language,2018,2018

(2) 2018,C language,2018,python

(4) 2018,2018,2018,2018

課程內容

正規搜尋切割取代

- 規則定義
- findall
- split
- sub

numpy陣列合併與拆解

- row 的分合
- column 分合
- 可選擇行列分合
- 不等量分割