

Python60 系列 中華職棒記錄查詢



</PLUS>

推動大學(程式設計)教學
MOE/Ministry of Education

推動大學程式設計教學計畫。分項六：資料分析領域與學習評量推動團隊（德明財經科大資科系江政杰老師主編）

中華職棒記錄查詢：課程簡述

▶ 問題情境：

- 手邊有一個中華職棒歷年打擊資料，我們希望撰寫程式能快速搜尋符合指定條件的紀錄、選手及達成次數。
 - 例如找出一季同時達成10全壘打10盜壘的紀錄。

▶ 本課程將撰寫一個 19 行的程式，實際課程時間 20-30 分鐘。

▶ 中華職棒全記錄查詢

<https://www.cpbl.com.tw/stats/recordall>

中華職棒球員打擊數據 - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明

狀態	姓名	年度	出數	打席	打數	打點	得分	安打	一安	二安	三安	全壘打	被三振	盜壘	打擊率	上壘率	長打率	保送	故意四壞	觸身球
退役	鷹俠	1990,90	378,322	58,58	94,52	21,3	18,58	7,0	292	0.376	0.543	41,6	5							
退役	瑞克	1990,90	381,342	45,48	110,68	33,2	7,37	7,0	322	0.38	0.491	30,1	4							
退役	洪一中	1990,90	404,365	31,52	104,78	22,4	0,17	5,0	285	0.338	0.367	27,0	4							
退役	王俊郎	1990,90	348,313	24,32	71,57	12,1	1,40	3,0	227	0.288	0.281	25,0	2							
退役	李居明	1990,88	371,335	42,43	97,70	15,4	8,34	3,0	29	0.338	0.43	27,4	0							
退役	林易增	1990,87	383,350	34,66	116,94	15,3	4,20	34,0	331	0.378	0.426	24,0	4							
退役	何良志	1990,86	330,278	15,47	67,58	5,4	0,33	12,0	241	0.342	0.288	43,0	0							
退役	林百亨	1990,86	338,310	34,31	82,63	14,0	5,28	4,0	265	0.311	0.358	21,2	0							
退役	陳金茂	1990,86	364,315	49,59	93,64	21,2	6,40	7,0	295	0.366	0.432	32,0	7							
退役	孟雷	1990,85	331,283	43,43	68,44	11,2	11,48	3,0	24	0.345	0.41	45,0	1							
退役	林振賢	1990,85	268,249	21,28	58,45	10,2	1,26	4,0	233	0.275	0.301	14,0	1							
退役	羅曼	1990,85	345,325	49,38	77,50	13,1	13,41	4,0	237	0.269	0.403	12,1	3							

中華職棒記錄查詢：課程簡述

▶ 問題情境：

- 手邊有一個中華職棒歷年打擊資料，我們希望撰寫程式能快速搜尋符合指定條件的紀錄、選手及達成次數。
 - 例如找出一季同時達成10全壘打10盜壘的紀錄。

▶ 本課程將撰寫一個 19 行的程式， 實際課程時間 20-30 分鐘。

▶ 中華職棒全記錄查詢

<https://www.cpbl.com.tw/stats/recordall>

```
共有 50 人有此成就
第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
>>>
```

中華職棒記錄查詢：學習內容

▶ 程式設計概念

- 邏輯判斷 `if ... else ...`
- 迴圈控制 `for ... in ...`
- 列表
- 字典

▶ 資料分析概念

- 搜尋與比對

中華職棒記錄查詢

完整程式展示

```
import csv
from collections import Counter

with open('中華職棒球員打擊數據.csv') as infile:
    data = list(csv.DictReader(infile))
    results = []
    for e in data:
        if int(e['全壘打'])>=10 and int(e['盜壘'])>=10:
            results.append(e)

counts = Counter([e['姓名'] for e in results])
print('共有', len(counts), '人有此成就')
for r, p in enumerate(counts.most_common(3)):
    name, times = p[0], p[1]
    print('第', r+1, '名:', name, '紀錄達成', times, '次')
    for e in results:
        if e['姓名'] == name:
            print(name, e['年度'], e['全壘打'], e['盜壘'])
```

```
共有 50 人有此成就
第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
>>>
```

中華職棒記錄查詢

完整程式展示

匯入後面會使用的函式庫

```
import csv
from collections import Counter
```

讀取打擊資料，並
檢驗是否符合條件

```
with open('中華職棒球員打擊數據.csv') as infile: # 開啟打擊數據資料檔案
    data = list(csv.DictReader(infile)) # 讀取打擊數據資料檔案
    results = [] # 存放達成記錄的球員打擊資訊
    for e in data: # 針對每一位球員
        if int(e['全壘打'])>=10 and int(e['盜壘'])>=10:
            # 如果 全壘打數>=10 且 盜壘>=10
            results.append(e) # 記錄該球員的打擊資訊
```

```
counts = Counter([e['姓名'] for e in results]) # 統計達成記錄的球員資料
print('共有', len(counts), '人有此成就') # 顯示達成記錄的球員人數
for r, p in enumerate(counts.most_common(3)): # 針對記錄次數最高的前三位球員
    name, times = p[0], p[1]
    print('第', r+1, '名:', name, '紀錄達成', times, '次') # 顯示球員記錄次數
    for e in results: # 針對所有達成記錄的球員
        if e['姓名'] == name: # 找出符合前三高次數的球員姓名資訊
            print(name, e['年度'], e['全壘打'], e['盜壘']) # 顯示此位球員每年度的全壘打與盜壘數
```

```
共有 50 人有此成就
第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
>>>
```

計算達成記錄的球員人數，
並列出前三名的詳細資訊



Python 模組

- ▶ Python 提供眾多模組讓寫程式更輕鬆。
 - 每個模組就是一個函式庫，請想像成一個工具箱。
 - 模組內包含許多函式，可直接呼叫需要的函式進行各種功能，我們就不需要自行撰寫這些複雜的程式。
- ▶ 本次程式範例使用的兩個模組，都已經內建而不需要特別安裝。

Python 模組

▶ csv 模組

- 匯入模組

```
import csv
```

- 提供讀寫處理 CSV 檔案的函式。

▶ collections 模組

- 匯入模組中的 Counter

```
from collections import Counter
```

- collections 提供多種有用的資料容器。

- 本程式只會用到 collections 模組中的 Counter 容器，利用此容器協助計算（達成記錄的）次數。

程式碼解說：資料讀取與查詢



這部份和單元三相同，看過單元三的同學可以快速跳過。

```
01 import csv
02 from collections import Counter
03
04 with open('中華職棒球員打擊數據.csv') as infile: # 開啟打擊數據資料檔案
05     data = list(csv.DictReader(infile)) # 讀取打擊數據資料檔案
06     results = [] # 存放達成記錄的球員打擊資訊
07     for e in data: # 針對每一位球員
08         if int(e['全壘打'])>=10 and int(e['盜壘'])>=10:
09             # 如果 全壘打數>=10 且 盜壘>=10
10                 results.append(e) # 記錄該球員的打擊資訊
```

- ▶ 第 01 行匯入 csv 模組，以便第 05 行可使用模組內的 DictReader() 函式。
- ▶ 第 04 行使用 open() 函式開啟「中華職棒球員打擊數據.csv」這個檔案，並以 infile 變數代表之。
 - 此處使用到 with 敘述，它可以自動關閉檔案。我們不深究細節，只要知道在哪裡填入檔名和變數名即可。
- ▶ 第 05 行使用 csv 模組內的函式 DictReader()，從 infile 代表的檔案中讀出資料；接著把資料轉換成 list 型態，最後以 data 變數代表之，以接續 07-10 行的操作。
 - 這裡我們看到一程式內可以包含讀取檔案資料、資料型態轉換和指定變數三個動作。在撰寫這樣複合的程式命令時，要注意資料流動的順序：檔案資料 → list 型態資料 → data 變數。

程式碼解說：資料讀取與查詢

```
01 import csv
02 from collections import Counter
03
04 with open('中華職棒球員打擊數據.csv') as infile: # 開啟打擊數據資料檔案
05     data = list(csv.DictReader(infile)) # 讀取打擊數據資料檔案
06     results = [] # 存放達成記錄的球員打擊資訊
07     for e in data: # 針對每一位球員
08         if int(e['全壘打'])>=10 and int(e['盜壘'])>=10:
09             # 如果 全壘打數>=10 且 盜壘>=10
10                 results.append(e) # 記錄該球員的打擊資訊
```

- ▶ 第 06 行 `result` 是個空列表，用來存放達成記錄的球員打擊資訊。
- ▶ 第 07-10 行，我們將 `data` 中的資料逐一取出，令 `e` 代表之。若 `e` 代表的資料的全壘打欄位的值大於等於 10 且盜壘欄位的值大於等於 10，我們便將這筆資料加入到 `result` 列表中。
 - `and` 是一種邏輯運算子，可以要求 `if` 判斷兩種以上的條件同時成立。
 - 如果只要任一條件成立便視為符合，則可使用 `or` 運算子。

程式概念：計數器 Counter

```
01 from collections import Counter
02
03 cnt = Counter([10, 20, 10, 20, 20, 25])
04 print(cnt[10], cnt[20], cnt[25])
05 print(cnt)
06
07 cnt = Counter(['彭政閔', '黃忠義', '張泰山', '張泰山'])
08 print(cnt['彭政閔'], cnt['張泰山'], cnt['黃忠義'])
09 print(cnt)
```

```
2 3 1
Counter({20: 3, 10: 2, 25: 1})
1 2 1
Counter({'張泰山': 2, '彭政閔': 1, '黃忠義': 1})
```

- ▶ 當我們需要計算列表中每個值出現的次數時，Counter 是一個很方便的工具。使用 Counter 前，我們必須從 collections 模組匯入它，語法如第 01 行所示。
- ▶ 第 03 行示範如何以 Counter 計算列表中各值出現的次數，我們以欲計數的列表建立一個 Counter 物件，令 cnt 代表之。
- ▶ 第 04 行示範我們可以查詢某一個特定值的出現次數，例如 cnt[10] 取出 10 出現的次數。
- ▶ 第 05 行我們則是將整個 cnt 物件列印出來，我們可以發現，它和字典的型式很接近。
- ▶ 第 03-05 行以數值列表建立計數器，第 07-09 行則以字串列表建立計數器。

程式概念：計數器 Counter

```
01 from collections import Counter
02
03 cnt = Counter(['彭政閔', '黃忠義', '張泰山', '張泰山', '黃忠義', '張泰山', '張泰山'])
04 print(cnt)
05 top1 = cnt.most_common(1)
06 print(top1)
07 top2 = cnt.most_common(2)
08 print(top2)
09 for e in top2:
10     print(e[0], e[1])
```

```
Counter({'張泰山': 4, '黃忠義': 2, '彭政閔': 1})
[('張泰山', 4)]
[('張泰山', 4), ('黃忠義', 2)]
張泰山 4
黃忠義 2
```

- ▶ 在我們建立一個 Counter 後，我們可以利用它的 most_common() 函式取出出現次數最高的資料及其出現次數。
- ▶ 第 05 行示範如何取出最高次數，第 06 行我們看到 most_common() 回傳的結果是一個列表，列表中的資料數是由 most_common() 來指定的，此處只有一筆：張泰山出現 4 次。
- ▶ 第 07 行示範取出出現次數最高的兩個（名字），第 08 行的列印結果可看出是張泰山 4 次和黃忠義 2 次。
- ▶ 第 09-10 行示範如何取出前兩高的每一筆資料。e[0] 代表出現的資料內容，e[1] 代表該資料內容的出現次數。

程式碼解說：計量與輸出結果

```
12 counts = Counter([e['姓名'] for e in results]) # 統計達成記錄的球員資料
13 print('共有', len(counts), '人有此成就') # 顯示達成記錄的球員人數
14 for r, p in enumerate(counts.most_common(3)): # 針對記錄次數最高的前三位球員
15     name, times = p[0], p[1]
16     print('第', r+1, '名:', name, '紀錄達成', times, '次') # 顯示球員記錄次數
17     for e in results: # 針對所有達成記錄的球員
18         if e['姓名'] == name: # 找出符合前三高次數的球員姓名資訊
19             print(name, e['年度'], e['全壘打'], e['盜壘']) # 顯示此位球員每年度的全壘打與盜壘數
```

- ▶ 第 12 行我們首先以列表生成式從「之前程式碼篩選出符合條件的資料 (result)」取出姓名，以這個姓名構成的新列表來建立計數器，令 counts 代表之。
- ▶ 第 13 行以 len() 函式求取 counts 的長度，也就是計數器中有多少個不同的資料。前一行我們是以球員姓名來計數，所以不同資料的筆數就代表有多少個不同 (姓名) 的球員曾有符合篩選條件的紀錄。

```
共有 50 人有此成就
第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
```

程式碼解說：計量與輸出結果

```
12 counts = Counter([e['姓名'] for e in results]) # 統計達成記錄的球員資料
13 print('共有', len(counts), '人有此成就') # 顯示達成記錄的球員人數
14 for r, p in enumerate(counts.most_common(3)): # 針對記錄次數最高的前三位球員
15     name, times = p[0], p[1]
16     print('第', r+1, '名:', name, '紀錄達成', times, '次') # 顯示球員記錄次數
17     for e in results: # 針對所有達成記錄的球員
18         if e['姓名'] == name: # 找出符合前三高次數的球員姓名資訊
19             print(name, e['年度'], e['全壘打'], e['盜壘']) # 顯示此位球員每年度的全壘打與盜壘數
```

- ▶ 第 14 行我們先對 `counts` 呼叫 `most_common(3)`，取出次數最高的前三個名字與次數，然後以 `enumerate()` 搭配 `for ... in ...` 句型依序取用。
 - 還記得嗎？使用 `enumerate()` 可以讓我們逐一取用資料之餘也能一併得到一個從 0 起算的索引值。
 - 因為我們呼叫 `most_common()` 時傳入 3，所以 `for ... in ...` 會取到 3 對 (p) 名字與出現次數，索引值 (r) 是 0, 1, 2。
 - 在我們的例子中，`most_common(3)` 得到的結果是 `[('林智勝', 8), ('彭政閔', 7), ('張泰山', 6)]`。
- ▶ `most_common()` 得到的是一或多個 (資料, 次數) 的配對，我們可以用 `[0]` 和 `[1]` 分別取得資料 (此處為姓名) 和次數，如第 15 行所示。第 16 行我們列印出該名字出現的次數。
- ▶ 第 17 行，我們逐一取出 `results` 中的每一筆紀錄 (從 `data` 中篩選出符合條件的紀錄)。
- ▶ 第 18-19 行，我們比對該紀錄的球員姓名，如果符合我們正在檢視的前幾名的姓名，便列印出該筆紀錄的姓名、年度、全壘打數和盜壘數。

```
● 共有 50 人有此成就
● 第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
● 第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
● 第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
```

中華職棒記錄查詢

完整程式展示

```
import csv
from collections import Counter

with open('中華職棒球員打擊數據.csv') as infile:
    data = list(csv.DictReader(infile))
    results = []
    for e in data:
        if int(e['全壘打'])>=10 and int(e['盜壘'])>=10:
            results.append(e)

counts = Counter([e['姓名'] for e in results])
print('共有', len(counts), '人有此成就')
for r, p in enumerate(counts.most_common(3)):
    name, times = p[0], p[1]
    print('第', r+1, '名:', name, '紀錄達成', times, '次')
    for e in results:
        if e['姓名'] == name:
            print(name, e['年度'], e['全壘打'], e['盜壘'])
```

```
>>>
===== RESTART: D:\pluto\project\程
.py =====
共有 50 人有此成就
第 1 名: 林智勝 紀錄達成 8 次
林智勝 2004 11 13
林智勝 2006 12 14
林智勝 2008 11 11
林智勝 2009 31 11
林智勝 2010 21 12
林智勝 2012 24 10
林智勝 2014 11 14
林智勝 2015 31 30
第 2 名: 彭政閔 紀錄達成 7 次
彭政閔 2002 18 16
彭政閔 2003 18 22
彭政閔 2004 18 26
彭政閔 2005 14 13
彭政閔 2007 21 12
彭政閔 2009 17 25
彭政閔 2013 10 15
第 3 名: 張泰山 紀錄達成 6 次
張泰山 1998 14 27
張泰山 1999 17 18
張泰山 2000 11 11
張泰山 2003 28 22
張泰山 2004 21 14
張泰山 2007 19 12
>>>
```

挑戰時刻

- ▶ 前面的例子是找出年度達到 10 支全壘打且 10 次盜壘的選手。
- ▶ 這次請找出年度達到 20 支全壘打且 20 次盜壘的選手。
 - 人數有點少，那麼請嘗試著把符合條件的每一位選手資料都列印出來。
- ▶ 再一題，聽說長打好手常用力揮棒，因此打擊率也會比較低。請找出年度達到 20 支全壘打且打擊率超過三成的選手。
 - 只要列出前三位次數最高的選手資料即可。