

Day 02

# 檔案存取與資料解析

## Python 下載CSV檔案與解析



出題教練：張維元



python

# 本日知識點目標

- 了解 csv 檔案格式與內容
- 能夠利用套件存取 csv 格式的檔案

# CSV 檔案格式

CSV (Comma-Separated Values, 逗點分隔值) 檔案以純文字形式儲存文字資料，每條記錄由欄位組成，欄位間利用逗號作為分隔符號。

可以使用一般的文字編輯器以原始格式開啟，也可以使用 excel 或 number 等試算表軟體以表格方式開啟

一般格式如下，第一列會記錄格式，第二列開始記錄資料

1	id,name,relation
2	1,Harry,father
3	2,Chloe,mother
4	3,Dylan,brother
5	4,Emily,sister

	C1 ▼ 1	C2 ◆	C3 ◆
1	id	name	relation
2	1	Harry	father
3	2	Chloe	mother
4	3	Dylan	brother
5	4	Emily	sister

# CSV 檔案格式優點與缺點

優點	缺點
<ul style="list-style-type: none"><li>• 結構單純</li><li>• 人機皆可讀</li><li>• 檔案小</li></ul>	<ul style="list-style-type: none"><li>• 未限定編碼(big5, utf-8 ... )</li><li>• 值內有逗點「,」可能造成欄位判斷錯誤</li><li>• 第一行不一定是欄位名稱</li><li>• 換行問題</li></ul>

# 思考流程與使用套件

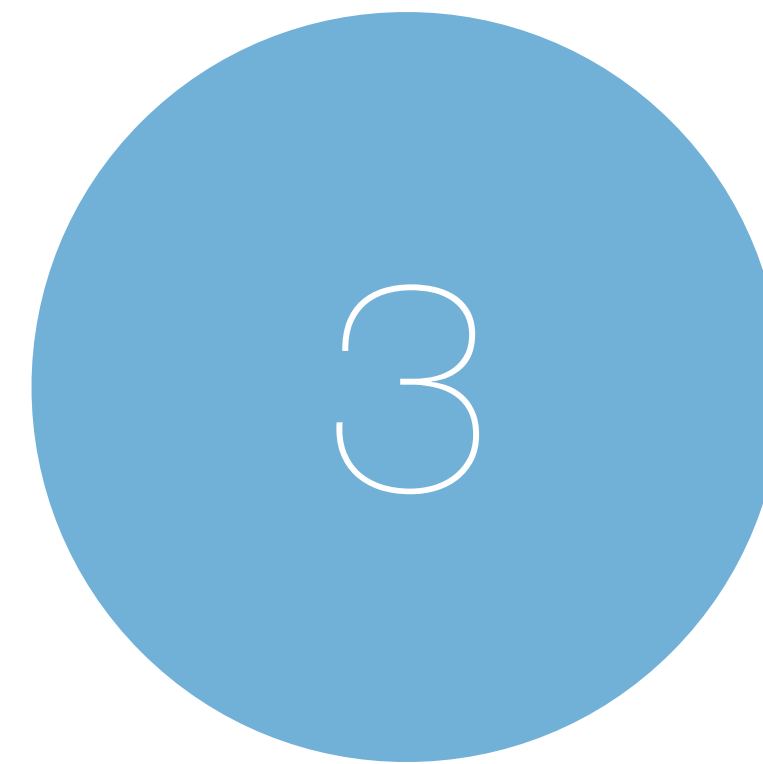
---



下載檔案



開啟檔案



解析檔案

# 一個簡單的範例

```
1
2 import csv
3
4 spamReader = csv.reader(open('eggs.csv'), delimiter=' ', quotechar='"')
5 for row in spamReader:
6     print(', '.join(row))
7
8 # Spam, Spam, Spam, Spam, Spam, Baked Beans
9 # Spam, Lovely Spam, Wonderful Spam
10
```



# 更多的參數使用

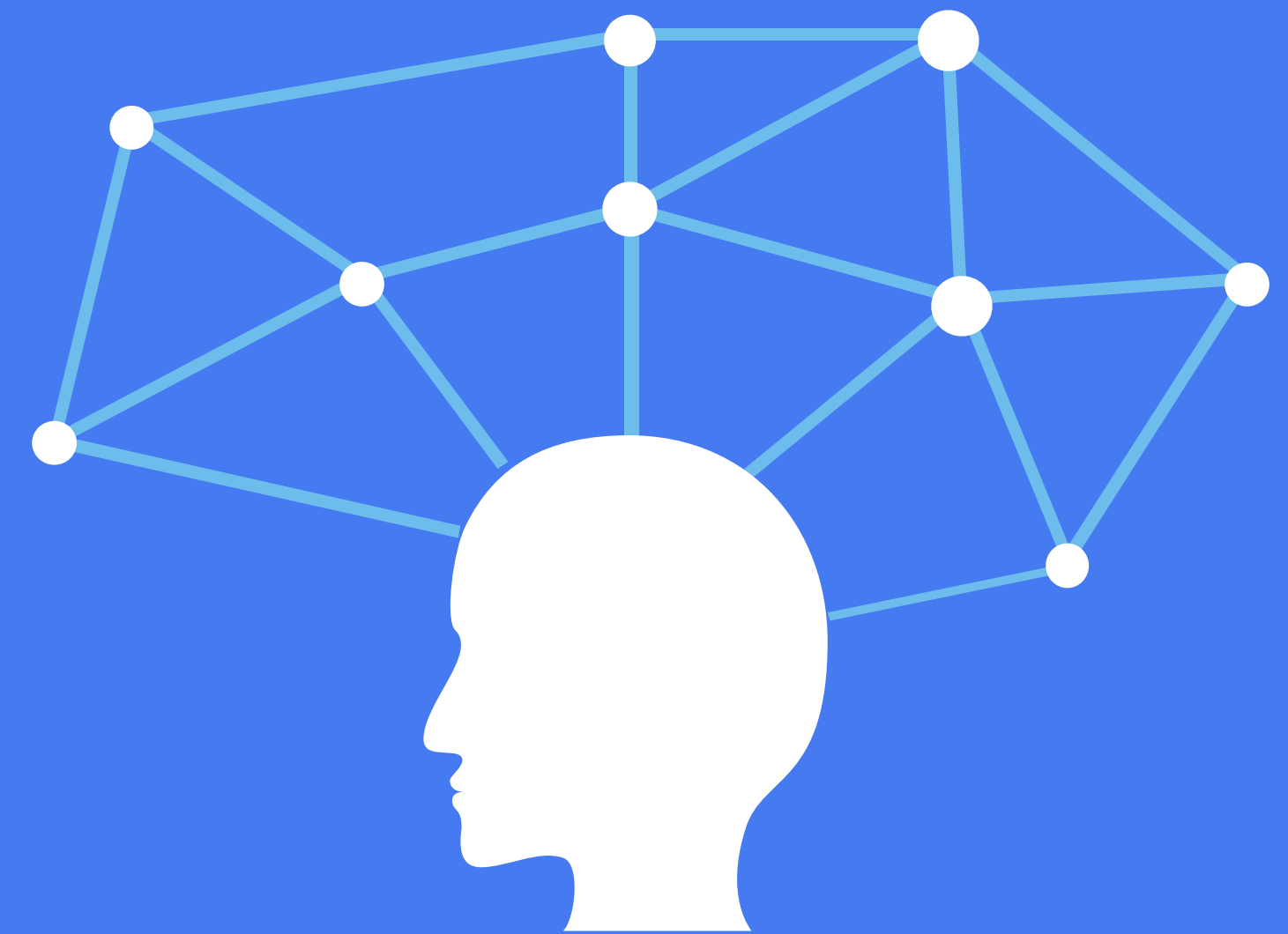
`CSV.reader(csvfile, dialect='excel', **fmtparams)`

- 返回值：csv object
- 參數表：
  - csvfile: 檔案位置
  - dialect: 用什麼方式定義 CSV 格式，預設採用逗號（excel 風格），常見的設定有：
    - delimiter：分隔符號，預設是逗號
    - lineterminator：換行符號，預設是 `\r\n`
    - 更多可以參考：<https://docs.python.org/3.0/library/csv.html#csv-fmt-params>

fmtparams: 有更多參數可以複寫，一般不會動到

# 重要知識點複習

- 了解 csv 檔案格式與內容
- 利用套件存取 csv 格式的檔案







## CSV File Reading and Writing

- CSV 函式庫的官方文件，上面有完整的參數列表可以參考。



## Reading and Writing CSV Files in Python

- 除了基本的使用外，該文章提供其他種存取 CSV 格式的做法。

# 解題時間

## LET'S CRACK IT

請跳出 PDF 至官網 Sample Code & 作業

開始解題

