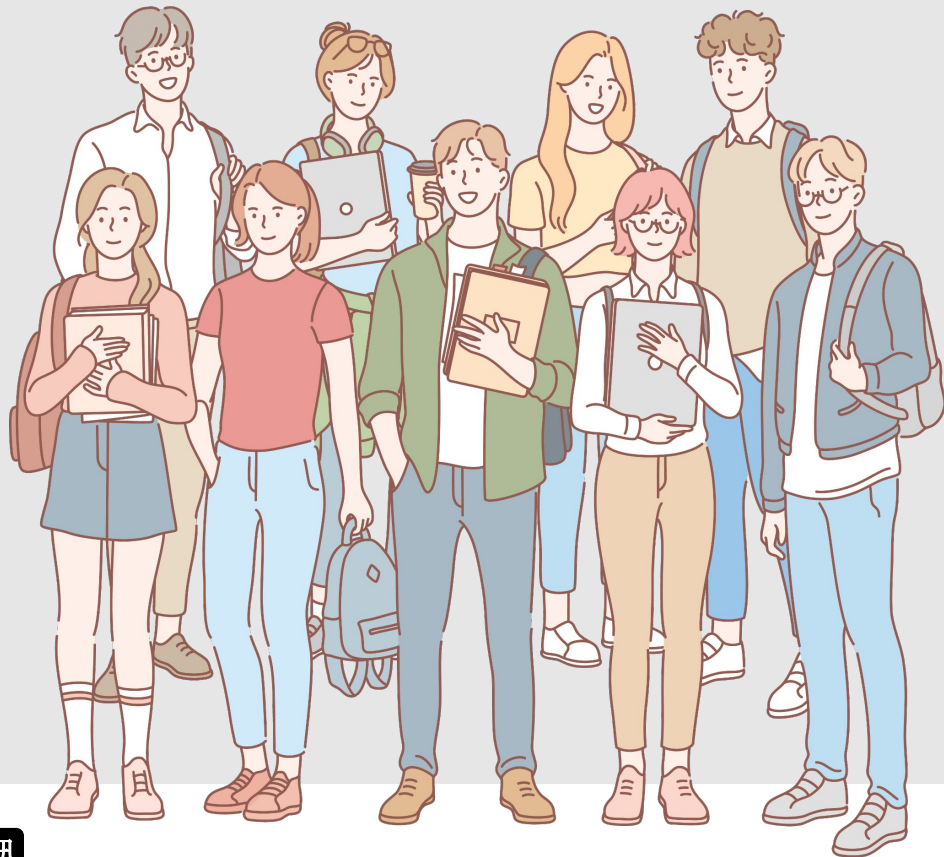


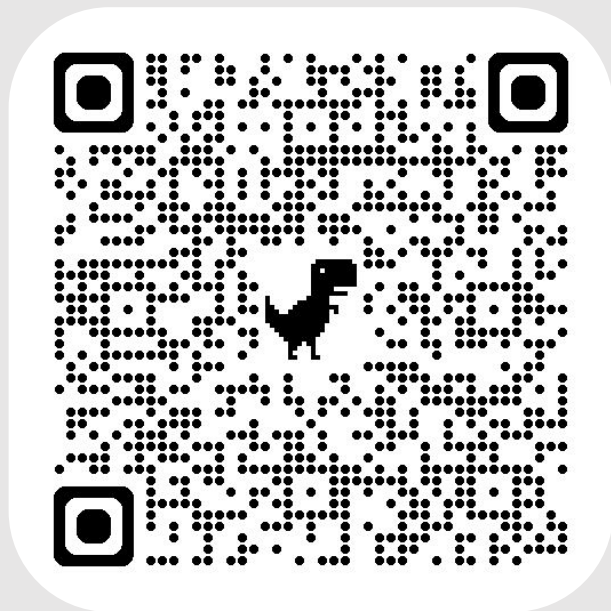
在教育上的探討 Python大數據

國立台北科技大學



黃信溢 茶米老師

茶 米 老 師 教 室 。 沃 克 資 訊



<https://reurl.cc/RqbkDz>

茶 米 老 師 教 室 。 黃 信 溢

範例下載 會議連結



茶米老師教室
Facebook粉絲專頁



茶米老師教室
LINE 官方帳號



茶米老師教室
YouTube頻道



黃信溢 茶米老師教室

edreamer@gmail.com

<http://www.facebook.com/edreamertw>

<http://www.youtube.com/edreamertw>

資訊教育初學特訓班

Facebook社團





<https://lin.ee/TqMnsTT>



茶米老師教室

[illegible]

關於Pandas

Pandas(<https://pandas.pydata.org>) 是一個為了解決數據分析任務而建立的模組，其命名的來源就與強調的功能很有關係：**面板 (Panel)**、**資料 (Data)** 和 **分析 (Analysis)**。也有人將名稱與 Pandas 使用的三種資料結構：**Panel**、**DataFrame**、**Series** 相互呼應。



Pandas的三種資料結構

- **Series**: 一維的資料結構, 與List串列相似。
- **DataFrame**: 二維的表格型資料結構, 可視為Series的容器。
- **Panel**: 三維的陣列資料結構, 可視為DataFrame的容器。

Series

Panel

	A	B	C	D	E	F	G	H	I
1	日期	成交股數	成交金額	開盤價	最高價	最低價	收盤價	漲跌價差	成交筆數
2	111/08/01	24,991,291	12,569,771,761	506	508	500	504	-5	26,792
3	111/08/02	42,669,591	20,973,293,337	494	496	488.5	492	-12	63,879
4	111/08/03	29,838,832	14,823,224,632	494	501	493	501	9	25,570
5	111/08/04	26,589,086	13,279,624,282	499	503	495	500	-1	27,173
6	111/08/05	35,052,642	17,966,410,242	509	516	507	516	16	49,928
7	111/08/08	20,568,971	10,531,710,250	510	515	509	512	-4	18,131
8	111/08/09	24,370,709	12,372,442,661	507	511	504	510	-2	25,433

台積電(2330)8月份 台積電(2330)9月份 台積電(2330)10月份

DataFrame

Pandas 的模組使用

Pandas使用前必須先匯入模組，語法如下：

```
import pandas as pd
```

Pandas常用的資料結構有：

- **Series**：一維的資料陣列，index(索引)為資料儲存順序。
- **DataFrame**：二維的資料陣列，由 index(索引、列) 及 columns(行、欄) 組合而成。

Series 一維陣列

宣告語法

```
se = pd.Series(串列)
```

```
# 建立Series
```

```
price = [100, 50, 120, 80, 30]
```

```
se = pd.Series(price)
```

```
print(se)
```

```
# 求所有值
```

```
print(se.values)
```

```
# 求所有索引
```

```
print(se.index)
```

```
# 自訂索引
```

```
fruits = ['Apple', 'Banana', 'Cherry', 'Orange', 'Tomato']
```

```
prices = [100, 50, 120, 80, 30]
```

```
se2 = pd.Series(prices, index=fruits)
```

```
# 取得數值資料的統計資訊
```

```
se2.describe()
```

```
# 統計:加總、最小值、最大值、平均值、中位數
```

```
se2.sum()
```

```
se2.min()
```

```
se2.max()
```

```
se2.mean()
```

```
se2.median()
```

```
# 取值 loc(用索引名稱),iloc(用索引值)
```

```
se2.loc['Apple']
```

```
se2.iloc[0]
```

DataFrame 二維陣列

定義 DataFrame 資料

```
import pandas as pd
```

```
df = pd.DataFrame({
```

```
    "姓名":["林小明", "陳聰明", "黃美麗", "張小  
娟", "廖小誠"],
```

```
    "國文":[65,92,78,83,70],
```

```
    "英文":[90,72,76,93,56],
```

```
    "數學":[81,85,91,89,77],
```

```
    "社會":[79,53,47,94,80]
```

```
})
```

	姓名	國文	英文	數學	社會
0	林小明	65	90	81	79
1	陳聰明	92	72	85	53
2	黃美麗	78	76	91	47
3	張小娟	83	93	89	94
4	廖小誠	70	56	77	80

顯示 index, columns

```
print(df.index)
```

```
print(df.columns)
```

取得所有值

```
print(df.values)
```

取得第一筆(列)資料

```
print(df.values[0])
```

取得欄資料(單欄、多欄)

```
print(df['國文'])
```

```
print(df[['姓名','英文']])
```

取得指定欄列的值

```
print(df.values[0][1])
```

```
print(df.iloc[0, 1])
```

```
print(df.loc[0, '國文'])
```

DataFrame 二維陣列

重新指定索引欄

```
df1 = df.set_index('姓名')
```

	國文	英文	數學	社會
姓名				
林小明	65	90	81	79
陳聰明	92	72	85	53
黃美麗	78	76	91	47
張小娟	83	93	89	94
廖小誠	70	56	77	80

取得指定欄列的值

```
df1.index
```

```
df1['國文']
```

```
df1[['國文', '英文']]
```

用索引號

```
df1.iloc[0]
```

用列名

```
df1.loc['林小明']
```

取得最前或最後幾筆資料

.head(n) .tail(n) n 預設為5

```
df.head(2)
```

```
df.tail(2)
```

篩選

```
df[df["姓名"] == "陳聰明"]
```

```
df[df["英文"] > 80]
```

依值排序

df.sort_values([欄名], ascending=布林值))

```
df.sort_values(['國文'], ascending=False)
```

DataFrame 二維陣列

新增資料

```
data = {'姓名': '李小英', '國文': 89, '英文': 65, '數學': 87, '社會': 90}
```

```
df = df.append(data,  
ignore_index=True)
```

修改資料

```
df.loc[0, '姓名'] = '王小明'
```

```
df.loc[0, ['國文', '英文']] = [75, 80]
```

```
df.iloc[0, 1:5] = [70, 85, 78, 88]
```

刪除

```
# df.drop(列名或欄名 [, axis=1])
```

```
# 如果刪除是欄, 要加 axis=1
```

刪除一行

```
df.drop(0, inplace=True)
```

刪除多列

```
df.drop([3,4], inplace=True)
```

刪除一欄

```
df.drop('社會', axis=1, inplace=True)
```

刪除多欄

```
df.drop(['英文', '社會'], axis=1, inplace=True)
```



用Pandas讀取檔案

Pandas可以讀取本地或是網路上「表格式」的資料，如CSV、Excel、json、html
讀取資料，方法如下：

- `read_csv()`
- `read_excel()`
- `read_json()`
- `read_html()`

以 `read_csv()` 為例，語法如下：

`pd.read_csv(來源 [, 工作表(索引或名稱), names=自訂表頭串列,
header=當表頭的列數, index_col=當索引的欄, encoding=編碼])`

用Pandas儲存檔案

儲存資料

- `to_csv(檔名, encoding="編碼")`
- `to_excel(檔名, encoding="編碼")`
- `to_json(檔名, encoding="編碼")`
- `to_html(檔名, encoding="編碼")`



Pandas專題實作

Python大數據



專題資料來源網站

- AQI即時空氣品質
<https://data.gov.tw/dataset/40448>
- 中央氣象局地震預報中心
<https://scweb.cwa.gov.tw/zh-tw/earthquake/data>
- 教育部統計處(學校基本統計資訊)
https://depart.moe.edu.tw/ed4500/News_Content.aspx?n=5A930C32CC6C3818&sms=91B3AAE8C6388B96&s=8AF80DB14ADF7370

感謝您的參與

黃信溢 文淵閣工作室

<http://www.facebook.com/edreamertw>

<http://www.youtube.com/edreamertw>

david@e-happy.com.tw

