

---

# 東吳大學 X 理律學堂

資料彙整&視覺化  
賴祐全

---

---

---

# 第一部分 資料彙整

---

---

# 套件介紹-pandas

它是python數據分析的一個套件，於2009年底開源給大家使用

Pandas 提供兩種主要的資料結構，Series 與 DataFrame。

1	a = [1,2,3,4]
2	a = pd.Series(a)
3	a

0	1
1	2
2	3
3	4

dtype: int64

1	a = [1,2,3,4]
2	a = pd.DataFrame(a)
3	a

	0
0	1
1	2
2	3
3	4

# 套件介紹-pandas

Series 是用來處理時間序列相關的資料(如感測器資料等), 主要為建立索引的一維陣列。

DataFrame 則是用來處理結構化(Table like)的資料, 有列索引(Row)與欄標籤(Column)的二維資料集, 例如關聯式資料庫、CSV等等。

# 資料彙整介紹

- 套件安裝 & 匯入
- 資料匯入 & 觀察
- 欄位處理
- 格式套用(apply函數)
- 資料索引應用(loc,iloc函數)
- 資料分組(groupby)
- 資料篩選(filter)
- 唯一值處理(unique)

# 套件安裝 & 匯入

於指令端輸入以下指令:

```
C:\¥python>pip install pandas
```

開啟jupyter後新增notebook後匯入套件:

```
1 import pandas as pd
```

# 資料匯入 & 觀察

```
1 import pandas as pd
2 data = pd.read_csv("LawTech_0727_Pandas.csv")
3 data.head()
```

使用套件中匯入csv格式檔案的資料

Unnamed: 0		title	date		judge		content	link
0	0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	109.09.30		詐欺	最高法院刑事判決 2708號\n上 訴 人 林鶴爵\n...	108年度台上字第 2708號\n上 訴 人 林鶴爵\n...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>
1	1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號\n		108年度台上字第 ...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>
2	2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號\n		108年度台上字第 ...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>
3	3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	109.01.16	違反證券交易法等罪	最高法院刑事判決 第16號\n上 訴 人 臺灣高等檢察署...		108年度台上字第 16號\n上 訴 人 臺灣高等檢察署...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>
4	4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	109.01.09	違反證券交易法等罪不服再執行羈押	最高法院刑事裁定 第46號\n抗 告 人 許金龍\n...		109年度台抗字第 46號\n抗 告 人 許金龍\n...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>

# 資料匯入 & 觀察

```
1 import pandas as pd
2 data = pd.read_csv("LawTech_0727_Pandas.csv")
3 data.head()
```

出現沒有名稱的欄位

	Unnamed: 0	title	date	judge	content	link
0	0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	109.09.30	詐欺	最高法院刑事判決 2708號\n上 訴 人 林鶴爵\n...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
1	1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號\n	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
2	2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號\n	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
3	3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	109.01.16	違反證券交易法等罪	最高法院刑事判決 第16號\n上 訴 人 臺灣高等檢察署...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
4	4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	109.01.09	違反證券交易法等罪不服再執行羈押	最高法院刑事裁定 第46號\n抗 告 人 許金龍\n...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...



# 資料匯入 & 觀察

將第一欄位當作索引(index)

```
1 data = pd.read_csv("LawTech_0727_Pandas.csv", index_col=0)
2 data.head()
```

	title	date		judge		content	link
0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	109.09.30		詐欺	最高法院刑事判決	108年度台上字第2708 號\n上 訴 人 林鶴爵\n...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決	號\n	108年度台上字第4056 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決	號\n	108年度台上字第4056 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	109.01.16	違反證券交易法等罪	最高法院刑事判決	號\n上 訴 人 臺灣高等檢察署...	108年度台上字第16	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...
4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	109.01.09	違反證券交易法等罪 不服再執行羈押	最高法院刑事裁定	號\n抗 告 人 許金龍\n...	109年度台抗字第46	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...

# 資料匯入 & 觀察

## 匯入其他檔案格式的方式

### 匯入資料：

`pd.read_csv(filename)`：從CSV檔案匯入資料

`pd.read_table(filename)`：從限定分隔符的文字檔案匯入資料

`pd.read_excel(filename)`：從Excel檔案匯入資料

`pd.read_sql(query, connection_object)`：從SQL表/庫匯入資料

`pd.read_json(json_string)`：從JSON格式的字串匯入資料

`pd.read_html(url)`：解析URL、字串或者HTML檔案，抽取其中的tables表格

`pd.read_clipboard()`：從你的貼上板獲取內容，並傳給`read_table()`

`pd.DataFrame(dict)`：從字典物件匯入資料，Key是列名，Value是資料

### 匯出資料：

`df.to_csv(filename)`：匯出資料到CSV檔案

`df.to_excel(filename)`：匯出資料到Excel檔案

`df.to_sql(table_name, connection_object)`：匯出資料到SQL表

`df.to_json(filename)`：以Json格式匯出資料到文字檔案

# 資料匯入 & 觀察

觀察資料的長度、基本統計信息

```
1
```

```
len(data)
```

```
44
```

```
1 data.describe()
```

	title	date	judge	content	link
count	44	44	44	44	44
unique	44	42	14	43	44
top	最高法院 107 年度 台上 字第 846 號民事判決 ( 11K )	109.04.15	違反證券 交易法	最高法院刑事判決 108年度台上字第4056號 \n ...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>
freq	1	2	16	2	1

# 欄位處理

首先觀察標題(title)的內容，可以發現中間有空白符號分隔地點，年度，判決編號，判決類型

```
1 title1 = data["title"][0]
2 print(title1)
```

最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )

# 欄位處理

使用split方法以空格分割得到一個陣列，並取出我們想要提取的內容  
(註:split只能在格式為字串的格式使用)

```
1 ts = title1.split(" ")
2 print(ts)
3 print(ts[0])
4 print(ts[1])
```

```
['最高法院', '108', '年度', '台上', '字第', '2708', '號刑事判決 ( 15K ) ']  
最高法院  
108
```

# 欄位處理

依照相同方式將資料中所有標題的地點與年份提取出來並存放於一個陣列中

```
1 location = []
2 year = []
3
4 for i in data["title"]:
5     a = i.split(" ")
6     location.append(a[0])
7
8 for i in data["title"]:
9     a = i.split(" ")
10    year.append(a[1])
```

```
1 print(location[:5])
2 print(year[:5])
```

```
['最高法院', '最高法院', '最高法院', '最高法院', '最高法院']
['108', '109', '108', '108', '109']
```

# 欄位處理

```
1 data["year"] = year
2 data["location"] = location
3 data.head()
```

將提取後的結果合併到原資料中並觀察

	title	date	judge	content	link	year	location
0	最高法院 108 年度 台上字第 2708 號刑事判決 (15K)	109.09.30	詐欺	最高法院刑事判決第2708號 上 訴 人 林鶴爵	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
1	最高法院 109 年度 台上字第 1719 號刑事判決 (29K)	109.04.15	違反證券交易法等罪	最高法院刑事判決第4056號	108年度台上字 ... https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	109	最高法院
2	最高法院 108 年度 台上字第 4056 號刑事判決 (29K)	109.04.15	違反證券交易法等罪	最高法院刑事判決第4056號	108年度台上字 ... https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
3	最高法院 108 年度 台上字第 16 號刑事判決 (37K)	109.01.16	違反證券交易法等罪	最高法院刑事判決字第16號 上 訴 人 臺灣高等檢察署...	108年度台上 ... https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
4	最高法院 109 年度 台抗字第 46 號刑事裁定 (5K)	109.01.09	違反證券交易法等罪不服再執行羈押	最高法院刑事裁定字第46號 抗 告 人 許金龍	109年度台抗 ... https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	109	最高法院

# 格式套用

使用apply函數將同一個處理方式套用到所有選取的資料中

列(row)是axis=1, 欄(column)是axis=0

大部分都是選擇列處理的方式

```
1 #apply方式套用至選取的資料
2
3 def get_year(data):
4     return data["title"].split(" ")[1]
5
6 a = data[:5].apply(get_year,axis=1)
7 print(a)
```

0 108

1 109

2 108

3 108

4 109

dtype: object



# 格式套用

第二種方式同樣會先建立一個自定義的方法，套用到資料中

```
1 def get_year(title):  
2     return title.split(" ")[1]  
3  
4 b = data[:5].apply(lambda a:get_year(a["title"]),axis=1)  
5 print(b)
```

```
0    108  
1    109  
2    108  
3    108  
4    109  
dtype: object
```

# 練習1

提取判決編號以及判決類別並新增到資料中(難度:中等3分)

	title	date	judge	content	link	year	location	number	judge_type
0	最高法院 108 年度 台上 字第 2708 號 刑事判決 ( 15K )	109.09.30	詐欺	最高法院刑事判決 度台上字第2708號\n上 訴 人 林鶴爵\n...	108年 <a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>	108	最高法 院	2708	刑事判決
1	最高法院 109 年度 台上 字第 1719 號 刑事判決 ( 29K )	109.04.15	違反證券 交易法等 罪	最高法院刑事判決 度台上字第4056號\n	108年 <a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>	109	最高法 院	1719	刑事判決
2	最高法院 108 年度 台上 字第 4056 號 刑事判決 ( 29K )	109.04.15	違反證券 交易法等 罪	最高法院刑事判決 度台上字第4056號\n	108年 <a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx?ro=...</a>	108	最高法 院	4056	刑事判決

# 資料索引應用

這裡介紹兩個方式以索引提取部分資料，分別是loc以及iloc

loc是以標籤為主，逗號前後分別是行標籤(橫向)以及列標籤(直向)

```
: 1 data.loc[:5,"title"]  
  
: 0 最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )  
1 最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )  
2 最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )  
3 最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )  
4 最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )  
5 最高法院 107 年度 台上 字第 846 號民事判決 ( 11K )  
Name: title, dtype: object
```

# 資料索引應用

loc方法同時取出兩個欄位的資料

```
1 data.loc[:5,["title","judge"]]
```

	title	judge
0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	詐欺
1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	違反證券交易法等罪
2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	違反證券交易法等罪
3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	違反證券交易法等罪
4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	違反證券交易法等罪不服再執行羈押
5	最高法院 107 年度 台上 字第 846 號民事判決 ( 11K )	請求損害賠償

# 資料索引應用

`iloc`是以索引為主，逗號前後分別是行索引(橫向)以及列索引(直向)

```
1 data.iloc[:5,0]
```

```
0    最高法院 108 年度 台上 字第 2708 號刑事判決 (15K)
1    最高法院 109 年度 台上 字第 1719 號刑事判決 (29K)
2    最高法院 108 年度 台上 字第 4056 號刑事判決 (29K)
3          最高法院 108 年度 台上 字第 16 號刑事判決 (37K)
4          最高法院 109 年度 台抗 字第 46 號刑事裁定 (5K)
Name: title, dtype: object
```

# 資料索引應用

iloc方法同時取出兩個欄位的資料

```
1 data.iloc[:5,[0,2]]
```

	title	judge
0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	詐欺
1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	違反證券交易法等罪
2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	違反證券交易法等罪
3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	違反證券交易法等罪
4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	違反證券交易法等罪不服再執行羈押

# 資料索引應用

比較loc與iloc不同點：

```
1 #loc 先行標籤(橫) 後列標籤(直)  
2 data.loc[:5,"title"]
```

```
0 最高法院 108 年度 台上 字第 2708 號刑事判決 (15K)  
1 最高法院 109 年度 台上 字第 1719 號刑事判決 (29K)  
2 最高法院 108 年度 台上 字第 4056 號刑事判決 (29K)  
3 最高法院 108 年度 台上 字第 16 號刑事判決 (37K)  
4 最高法院 109 年度 台抗 字第 46 號刑事裁定 (5K)  
5 最高法院 107 年度 台上 字第 846 號民事判決 (11K)
```

Name: title, dtype: object

```
1 #iloc 取行索引和列索引(從0開始)  
2 data.iloc[:5,0]
```

```
0 最高法院 108 年度 台上 字第 2708 號刑事判決 (15K)  
1 最高法院 109 年度 台上 字第 1719 號刑事判決 (29K)  
2 最高法院 108 年度 台上 字第 4056 號刑事判決 (29K)  
3 最高法院 108 年度 台上 字第 16 號刑事判決 (37K)  
4 最高法院 109 年度 台抗 字第 46 號刑事裁定 (5K)
```

Name: title, dtype: object



## 練習2

以loc及iloc方式取出第15~20筆資料且只有呈現他的日期(date)及地點(location)

(難度:簡單1分)

	date	location
15	105.08.23	最高法院
16	105.08.04	最高法院
17	104.12.17	最高法院
18	104.11.19	最高法院
19	104.10.22	最高法院
20	104.10.08	最高法院



# 資料分組

groupby函數介紹, 選擇一個欄位並觀察該欄位分布狀況以及個數

```
1 #groupby 資料分組
2 judge1 = data.groupby("judge")
```

```
1 judge1.size()
```

```
judge
偽造文書等罪          1
偽造有價證券          1
損害賠償              2
詐欺                  1
請求侵權行為損害賠償          1
請求侵權行為損害賠償等      1
請求損害賠償          4
違反商業會計法        1
違反證券交易法       16
違反證券交易法等罪     12
違反證券交易法等罪不服再執行羈押    1
違反證券交易法等罪聲請具保停止羈押    1
違反證券交易法等罪聲請再審及停止執行    1
違反證券交易法等罪限制出境出海      1
dtype: int64
```

# 資料分組

使用get\_group方法取出其中一項分類資料

```
1 judge1.get_group("違反證券交易法")
```

	title	date	judge	content	link	year	location
9	最高法院 106 年度 台上 字第 278 號刑事判決 ( 12K )	106.09.28	違反證券 交易法	最高法院刑事判決 第278號\n上 訴 人 臺灣高等法院檢...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	106	最高法 院
11	最高法院 106 年度 台上 字第 86 號刑事判決 ( 33K )	106.05.31	違反證券 交易法	最高法院刑事判決 六號\n上 訴 人 葉禾庠 ( 原名葉健和 ) ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	106	最高法 院
12	最高法院 106 年度 台上 字第 1115 號刑事判決 ( 13K )	106.05.04	違反證券 交易法	最高法院刑事判決 五號\n上 訴 人 黃文賢\n選任辯護人...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	106	最高法 院
15	最高法院 105 年度 台上 字第 2144 號刑事判決 ( 20K )	105.08.23	違反證券 交易法	最高法院刑事判決 四號\n上 訴 人 台灣高等法院檢察署檢...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	105	最高法 院
16	最高法院 105 年度 台上 字第 1948 號刑事判決 ( 14K )	105.08.04	違反證券 交易法	最高法院刑事判決 八號\n上 訴 人 郭麗美\n選任辯護人...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	105	最高法 院

## 練習3

取出年份為104年且判決為"違反證券交易法等罪"的資料

(難度:中等3分)

	title	date	judge	content	link	year	location	number	judge_type
17	最高法院 104 年度 台上 字第 3871 號刑 事判決 ( 26K )	104.12.17	違反證券 交易法等 罪	最高法院刑事判決 一○四年 度台上字第三八七一號\n上 訴 人 台 灣高等法院檢察署檢...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	104	最高法 院	3871	刑事判決
19	最高法院 104 年度 台上 字第 3215 號刑 事判決 ( 49K )	104.10.22	違反證券 交易法等 罪	最高法院刑事判決 一○四年 度台上字第三二一五號\n上 訴 人 台 灣高等法院台南分院...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	104	最高法 院	3215	刑事判決
26	最高法院 104 年度 台上 字第 512 號刑 事判決 ( 7K )	104.02.26	違反證券 交易法等 罪	最高法院刑事判決 一○四 年度台上字第五一二號\n上 訴 人 台 灣高等法院檢察署檢...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	104	最高法 院	512	刑事判決

# 資料篩選

使用套件中filter函數，篩選部分資料(提取標題、日期、判決三個欄位)

```
1 new1 = data.filter(["title", "date", "judge"])
2 new1.head()
```

	title	date	judge
0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	109.09.30	詐欺
1	最高法院 109 年度 台上 字第 1719 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪
2	最高法院 108 年度 台上 字第 4056 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪
3	最高法院 108 年度 台上 字第 16 號刑事判決 ( 37K )	109.01.16	違反證券交易法等罪
4	最高法院 109 年度 台抗 字第 46 號刑事裁定 ( 5K )	109.01.09	違反證券交易法等罪不服再執行羈押

# 資料篩選

```
1 new2=data[data["title"].str.contains("台上")]
2 print(len(new2))
3 new2.head()
```

篩選資料中"標題"欄位有包含"台上"兩個字的資料

39

	title	date	judge	content		link	year	location
0	最高法院 108 年度 台上 字第 2708 號刑事判決 (15K)	109.09.30	詐欺	最高法院刑事判決 2708號	上 訴 人 林鶴爵	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
1	最高法院 109 年度 台上 字第 1719 號刑事判決 (29K)	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號	108年度台上字第 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	109	最高法院
2	最高法院 108 年度 台上 字第 4056 號刑事判決 (29K)	109.04.15	違反證券交易法等罪	最高法院刑事判決 4056號	108年度台上字第 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
3	最高法院 108 年度 台上 字第 16 號刑事判決 (37K)	109.01.16	違反證券交易法等罪	最高法院刑事判決 第16號	上 訴 人 臺灣高等檢察署...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院
5	最高法院 107 年度 台上 字第 846 號民事判決 (11K)	108.03.27	請求損害賠償	最高法院民事判決 846號	上 訴 人 財團法人證...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	107	最高法院



## 練習4

篩選出日期為下半年度的資料(6~12月), 且只有顯示標題(title)、日期(date)、內容(content) (難度:困難5分)

	title	date		content
0	最高法院 108 年度 台上 字第 2708 號刑事判決 ( 15K )	109.09.30	最高法院刑事判決	108年度台上字第2708號\n上 訴 人 林鶴爵\n...
7	最高法院 107 年度 台上 字第 2937 號刑事判決 ( 9K )	107.10.11	最高法院刑事判決	107年度台上字第2937號\n上 訴 人 臺灣高等檢察...
9	最高法院 106 年度 台上 字第 278 號刑事判決 ( 12K )	106.09.28	最高法院刑事判決	106年度台上字第278號\n上 訴 人 臺灣高等法院檢...
10	最高法院 106 年度 台上 字第 349 號民事判決 ( 17K )	106.07.26	最高法院民事判決	106年度台上字第349號\n上 訴 人 財團法人證券投...
13	最高法院 105 年度 台抗 字第 1030 號刑事裁定 ( 3K )	105.12.29	最高法院刑事裁定	一○五年度台抗字第一○三○號\n抗 告 人 余信昌\n上列抗告人...
14	最高法院 105 年度 台上 字第 3278 號刑事判決 ( 14K )	105.12.14	最高法院刑事判決	一○五年度台上字第三二七八號\n上 訴 人 翁輝傑\n上列上訴人...

# 唯一值處理

使用unique方法，取出某欄位中的所有不重複的唯一值資料

```
1 data["year"].unique()
```

```
array(['108', '109', '107', '106', '105', '104', '103', '102', '101',  
      '100', '99', '98', '96', '85'], dtype=object)
```

# 唯一值處理

使用unique方法，取出某欄位中的所有不重複的唯一值資料

使用變數儲存資料

```
1 j = data["judge"].unique()  
2 print(j)
```

```
['詐欺' '違反證券交易法等罪' '違反證券交易法等罪不服再執行羈押' '請求損害賠償' '違反證券交易法等罪聲請具保停止羈押'  
'請求侵權行為損害賠償' '違反證券交易法' '違反證券交易法等罪限制出境出海' '請求侵權行為損害賠償等'  
'違反證券交易法等罪聲請再審及停止執行' '違反商業會計法' '損害賠償' '偽造文書等罪' '偽造有價證券']
```

觀察資料格式

```
1 print(type(j))
```

```
<class 'numpy.ndarray'>
```



# 唯一值處理

取出後是陣列的格式，  
可以用索引(index)的方式取出其中的資料

## 觀察資料格式

```
1 print(type(j))  
  
<class 'numpy.ndarray'>
```

## 列出第一筆資料

```
1 print(j[0])  
2 print(j[1])
```

詐欺  
違反證券交易法等罪

## 練習5

利用unique方法找出判決為"損害賠償"資料 (難度:中等3分)

	title	date	judge	content	link	year	location	number	judge_type
5	最高法院 107 年度 台上 字第 846 號民 事判決 ( 11K )	108.03.27	請求損害 賠償	最高法院民事判決 107年度台上字第846號n上 訴 人 財 團法人證...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	107	最高法 院	846	民事判決
8	最高法院 106 年度 台上 字第 1247 號 民事判決 ( 19K )	107.03.14	請求侵權 行為損害 賠償	最高法院民事判決 106年度台上字第1247號n上 訴 人 財團 法人證券...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	106	最高法 院	1247	民事判決
10	最高法院 106 年度 台上 字第 349 號民 事判決 ( 17K )	106.07.26	請求損害 賠償	最高法院民事判決 106年度台上字第349號n上 訴 人 財 團法人證券投...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	106	最高法 院	349	民事判決
21	最高法院 104 年度 台上 字第 1894 號 民事判決 ( 22K )	104.10.07	請求損害 賠償	最高法院民事判決 104年度台上第一八九四號n上 訴 人 財團法 人證券投資人及...	<a href="https://law.judicial.gov.tw/FJUD/data.aspx?ro=...">https://law.judicial.gov.tw/FJUD/data.aspx? ro=...</a>	104	最高法 院	1894	民事判決



---

---

## 第二部分

# 資料視覺化

---

---

The slide features a minimalist design with a light blue background. It is framed by two thick teal horizontal bars at the top and bottom, each accompanied by a thinner light blue line. Two short, dark grey horizontal bars are positioned symmetrically on the left and right sides, below the main title area.

# 套件介紹-plotly

plotly 能輕鬆生成各種圖表的框架，快速繪製賞心悅目的圖表，還能幫圖表做互動效果、部署到瀏覽器。

在 python 中最廣為人知的圖表繪製套件是 Matplotlib，但是它相當複雜，若想畫比較漂亮的圖需要花很多時間和精力。除非想畫很高階複雜的圖表，否則使用 plotly 就足夠了。

# 資料視覺化步驟

- 套件安裝 & 匯入
- 套件使用說明
- 資料處理
- 視覺化套用

# 套件安裝 & 匯入

於指令端輸入以下指令:

```
C:\¥python>pip install plotly
```

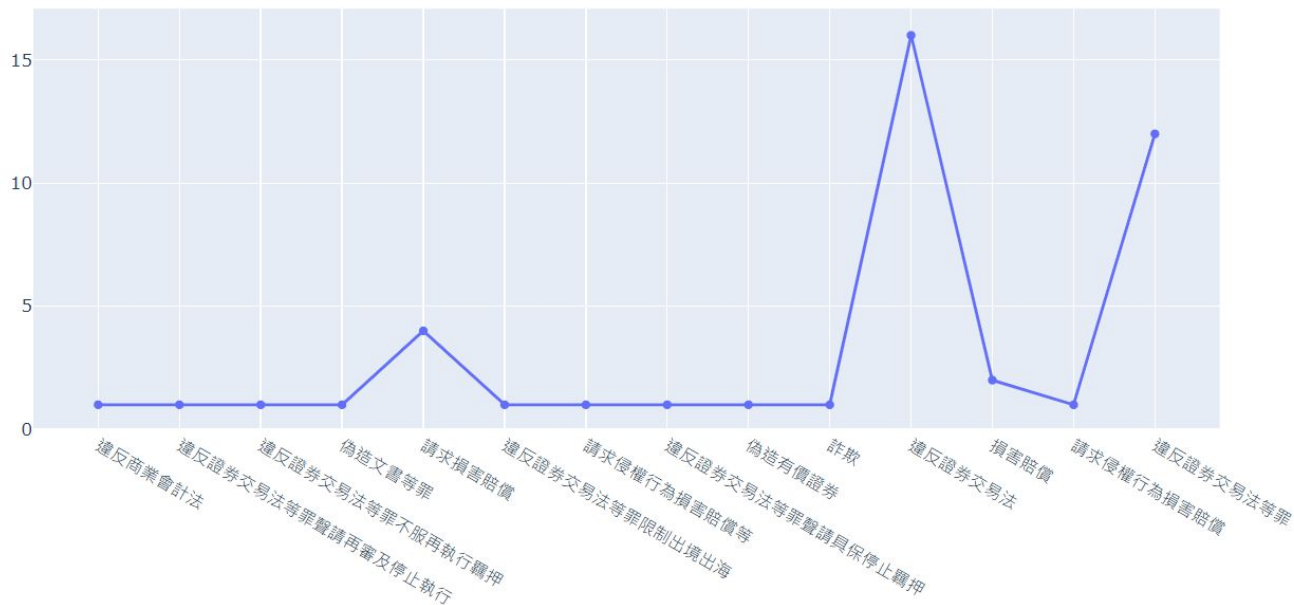
開啟jupyter後新增notebook後匯入套件:

```
import plotly
import plotly.graph_objs as go
```

# 套件功能說明



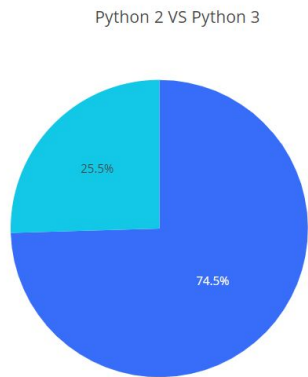
判決數量視覺化



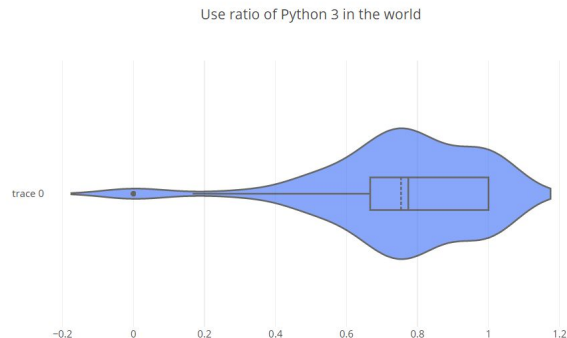
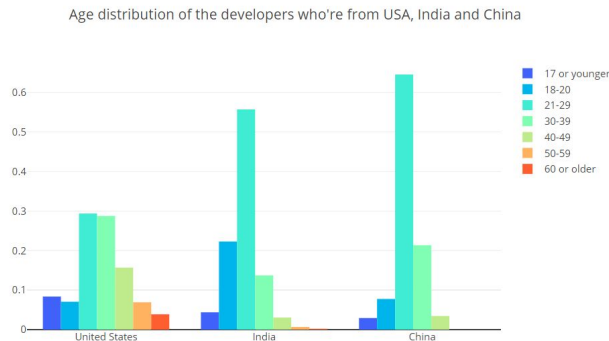


# 套件功能說明

除了上述的折線圖之外，**plotly**也提供其他類型的統計圖表可以使用，如點圖、圓餅圖、長條圖，琴狀圖等等



■ Python 3  
■ Python 2



<https://www.itread01.com/content/1542564499.html>

# 資料處理

```
1 location = []
2 year = []
3 for i in data["title"]:
4     a = i.split(" ")
5     location.append(a[0])
6     year.append(a[1])
7
8 data["year"] = year
9 data["location"] = location
10 data.head()
```

與第一部分相同方式提取年份與地點

	title	date	judge	content	link	year	location	
0	最高法院 108 年度 台上字第 2708 號刑事判決 ( 15K )	109.09.30	詐欺	最高法院刑事判決第2708號\n上 訴 人 林鶴爵\n...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院	
1	最高法院 109 年度 台上字第 1719 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決第4056號\n	108年度台上字 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	109	最高法院
2	最高法院 108 年度 台上字第 4056 號刑事判決 ( 29K )	109.04.15	違反證券交易法等罪	最高法院刑事判決第4056號\n	108年度台上字 ...	https://law.judicial.gov.tw/FJUD/data.aspx?ro=...	108	最高法院

# 資料處理

首先取出判決(judge)這個欄位，轉換成陣列格式，作為後續視覺化呈現用

```
1 j = list(set(data["judge"]))  
2 print(j)
```

```
['違反商業會計法', '違反證券交易法等罪聲請再審及停止執行', '違反證券交易法等罪不服再執行羈押', '偽造文書等罪', '請求損害賠償', '違反證券  
交易法等罪限制出境出海', '請求侵權行為損害賠償等', '違反證券交易法等罪聲請具保停止羈押', '偽造有價證券', '詐欺', '違反證券交易法', '損害  
賠償', '請求侵權行為損害賠償', '違反證券交易法等罪']
```

# 資料處理

計算判決欄位中所有類別出現的次數

```
1 judge = list(data["judge"])
2 judge_count= []
3
4 for i in j:
5     print(i,":",judge.count(i))
6     judge_count.append(judge.count(i))
```

違反證券交易法等罪聲請具保停止羈押 : 1  
請求損害賠償 : 4  
違反證券交易法 : 16  
違反證券交易法等罪聲請再審及停止執行 : 1  
違反商業會計法 : 1  
偽造有價證券 : 1  
詐欺 : 1  
損害賠償 : 2  
請求侵權行為損害賠償等 : 1  
違反證券交易法等罪不服再執行羈押 : 1  
違反證券交易法等罪限制出境出海 : 1  
偽造文書等罪 : 1  
請求侵權行為損害賠償 : 1  
違反證券交易法等罪 : 12

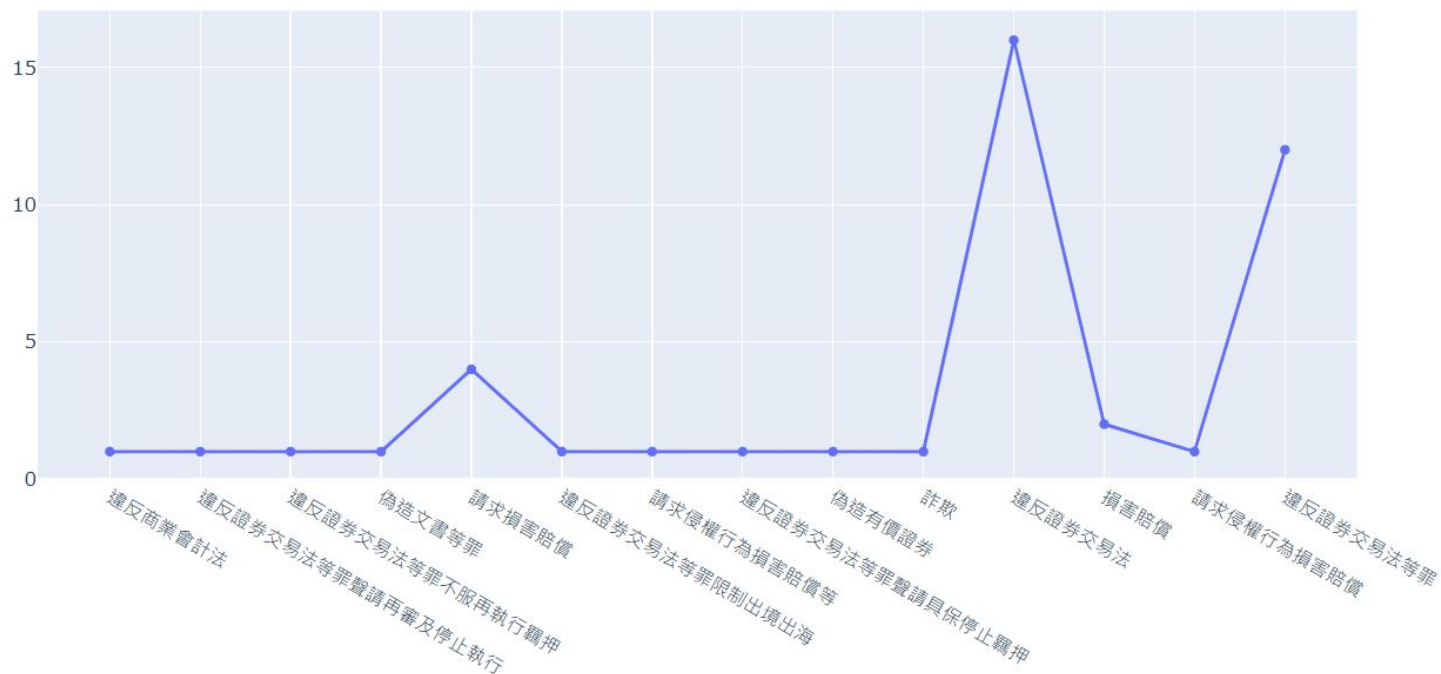
# 視覺化套用

視覺化將判決類型與次數以折線圖(Scatter plot)呈現

```
1 plotly.offline.init_notebook_mode(connected=True)
2
3 plotly.offline.iplot({
4     "data": [go.Scatter(x=j, y=judge_count)],
5     "layout": go.Layout(title="判決數量視覺化")
6 })
7
```

# 視覺化套用

判決數量視覺化



# 圖像設定

x,y: x,y軸資料

show\_link: 是否呈現連結

link\_text: 連結名稱

image: 檔案下載格式，可放 None , 'jpeg', 'png', 'webp', 'svg'

filename: 檔案名稱

image\_height,image\_width: 圖檔的長度與寬度(單位:px)

```
1 plotly.offline.init_notebook_mode(connected=True)
2
3 plotly.offline.iplot([{'x': j, 'y': judge_count}], show_link=True, link_text='Go to plot.ly',
4                      image=None, filename='my_plotly_plot',
5                      image_height = 300,
6                      image_width = 500)
```

