

0523 Lecture

- Review
 - 還記得嗎? List, array, dict, Series, DataFrame
- Today
 - 練習造一個 DataFrame
 - Aggregation (聚集)
 - Groupby (分群再聚集, split-apply-combine)
 - Pivot table (樞紐分析)
- 資料集
 - 政府資料開放平台
 - 鐵達尼號資料集
- 5/30 : 確定分組名單
- 6/20, 6/27: 期末分組上台報告

Review

- 上課程式example:
 - <https://github.com/alabtsai>
- for loop, while, if , string, int, float, function
- List
 - ex: t=[11, 33, 'dd']
- Array (記得 import numpy)
 - ex: t=[11 33 55]
- Dictionary
 - Key-value pair
 - ex: { 'a':2, 'b':5, 'cc':[33,66,77.2] }
- Series and DataFrame (記得 import pandas)
 - [FGU-Class/Lecture_0516.pdf](#)

Review : 抓取 json 資料

```
import json
import requests
res=requests.get("https://od.cdc.gov.tw/eic/NHI_EnteroviralInfection.json")
result=json.loads(res.text)
import pandas as pd
df=pd.DataFrame(result)
```

result is a list of dicts !

The first element in result is result[0]

```
{'健保就診總人次': '105',
'就診類別': '住院',
'年': '2008',
'年齡別': '0-2',
'縣市': '台中市',
'腸病毒健保就診人次': '0',
'週': '14'}
```

This is a string ! Not int !

Review : 抓取 json 資料

- df.info() 用 info() 去初步了解 DataFrame
- <class 'pandas.core.frame.DataFrame'>
- RangeIndex: 112173 entries, 0 to 112172
- Data columns (total 7 columns):
- 健保就診總人次 112173 non-null object
- 就診類別 112173 non-null object
- 年 112173 non-null object
- 年齡別 112173 non-null object
- 縣市 112173 non-null object
- 腸病毒健保就診人次 112173 non-null **object** 並非整數type
- 週 112173 non-null object

分析:

- 每年台灣腸病毒健保就診人次多少?

```
In [20]: df1=pandas.read_csv("NHI_EnteroviralInfection.csv")
```

```
In [21]: df1[:10]
```


Out[21]:

| | 年 | 週 | 就診類別 | 年齡別 | 縣市 | 腸病毒健保就診人次 | 健保就診總人次 |
|---|------|----|------|-----|-----|-----------|---------|
| 0 | 2008 | 14 | 住院 | 0-2 | 台中市 | 0 | 105 |
| 1 | 2008 | 14 | 住院 | 0-2 | 台北市 | 2 | 151 |
| 2 | 2008 | 14 | 住院 | 0-2 | 台東縣 | 0 | 14 |
| 3 | 2008 | 14 | 住院 | 0-2 | 台南市 | 0 | 20 |
| 4 | 2008 | 14 | 住院 | 0-2 | 宜蘭縣 | 0 | 44 |
| 5 | 2008 | 14 | 住院 | 0-2 | 花蓮縣 | 0 | 17 |
| 6 | 2008 | 14 | 住院 | 0-2 | 金門縣 | 0 | 1 |
| 7 | 2008 | 14 | 住院 | 0-2 | 屏東縣 | 0 | 19 |
| 8 | 2008 | 14 | 住院 | 0-2 | 苗栗縣 | 0 | 1 |
| 9 | 2008 | 14 | 住院 | 0-2 | 桃園市 | 0 | 141 |

分析:

- 每年台灣腸病毒健保就診人次多少?

```
df.groupby('年')['腸病毒健保就診人次'].sum()
```



| 年 | |
|------|--------|
| 2008 | 360000 |
| 2009 | 270178 |
| 2010 | 855397 |
| 2011 | 409542 |
| 2012 | 483404 |
| 2013 | 578879 |
| 2014 | 503170 |
| 2015 | 535773 |
| 2016 | 590212 |
| 2017 | 455533 |
| 2018 | 90211 |

分析:

- 不同縣市每年腸病毒就診類別(住院,門診)的總人次是多少?

```
In [20]: df1=pandas.read_csv("NHI_EnteroviralInfection.csv")
```

```
In [21]: df1[:10]
```

Out [21]:

| | 年 | 週 | 就診類別 | 年齡別 | 縣市 | 腸病毒健保就診人次 | 健保就診總人次 |
|---|------|----|------|-----|-----|-----------|---------|
| 0 | 2008 | 14 | 住院 | 0-2 | 台中市 | 0 | 105 |
| 1 | 2008 | 14 | 住院 | 0-2 | 台北市 | 2 | 151 |
| 2 | 2008 | 14 | 住院 | 0-2 | 台東縣 | 0 | 14 |
| 3 | 2008 | 14 | 住院 | 0-2 | 台南市 | 0 | 20 |
| 4 | 2008 | 14 | 住院 | 0-2 | 宜蘭縣 | 0 | 44 |
| 5 | 2008 | 14 | 住院 | 0-2 | 花蓮縣 | 0 | 17 |
| 6 | 2008 | 14 | 住院 | 0-2 | 金門縣 | 0 | 1 |
| 7 | 2008 | 14 | 住院 | 0-2 | 屏東縣 | 0 | 19 |
| 8 | 2008 | 14 | 住院 | 0-2 | 苗栗縣 | 0 | 1 |
| 9 | 2008 | 14 | 住院 | 0-2 | 桃園市 | 0 | 141 |

Pivot table (樞紐分析)

- `DataFrame.pivot_table(values, index, columns, aggfunc)`
- 不同縣市每年腸病毒就診類別(住院,門診)的總人次是多少?
- Column :不同縣市
- 每年就診類別(住院,門診): `index=['年', '就診類別']`
- 總人次: `aggfunc='sum'`
- Values: `values='腸病毒健保就診人次'`

Pivot table (樞紐分析)

```
dff1=df.pivot_table(values='腸病毒健保就診人次',index=['年'],columns='縣市',aggfunc='sum')
```

```
In [18]: dff1=df.pivot_table(values='腸病毒健保就診人次',index=['年'],columns='縣市',aggfunc='sum')
#dff1['宜蘭縣']
```

```
In [17]: dff1
```

Out[17]:

| 縣市 | 南投縣 | 台中市 | 台北市 | 台南市 | 台東縣 | 嘉義市 | 嘉義縣 | 基隆市 | 宜蘭縣 | 屏東縣 | ... | 新竹市 | 新竹縣 | 桃園市 | 澎湖縣 | 花蓮縣 | 苗栗縣 | 連江縣 | 金門縣 | 雲林縣 | 高雄市 |
|------|-------|--------|-------|-------|------|-------|------|-------|-------|-------|-----|-------|-------|-------|------|------|-------|-----|------|-------|-------|
| 年 | | | | | | | | | | | | | | | | | | | | | |
| 2008 | 10070 | 55893 | 27644 | 41518 | 2940 | 8285 | 4046 | 4838 | 7301 | 8328 | ... | 8022 | 10343 | 35016 | 1526 | 2665 | 10263 | 82 | 1741 | 14903 | 35102 |
| 2009 | 6798 | 47000 | 18645 | 25028 | 2865 | 6965 | 3185 | 4353 | 7339 | 7894 | ... | 5783 | 7810 | 25310 | 2211 | 3289 | 7719 | 90 | 1475 | 11105 | 24356 |
| 2010 | 23820 | 139973 | 71167 | 72514 | 4589 | 19743 | 9300 | 13845 | 15705 | 19837 | ... | 23591 | 26961 | 86691 | 2582 | 6537 | 19306 | 143 | 3123 | 30872 | 80403 |
| 2011 | 8923 | 56156 | 35338 | 36270 | 2434 | 10290 | 5518 | 6787 | 7550 | 9593 | ... | 10802 | 11285 | 46764 | 1119 | 3658 | 8641 | 20 | 1091 | 19556 | 35403 |
| 2012 | 8896 | 69643 | 45551 | 39939 | 2407 | 8394 | 4054 | 8654 | 10368 | 11873 | ... | 13996 | 14115 | 50126 | 1037 | 4223 | 12465 | 27 | 930 | 15872 | 40008 |
| 2013 | 15951 | 92019 | 46780 | 51551 | 3986 | 10599 | 5562 | 7625 | 11486 | 15471 | ... | 14698 | 14295 | 53771 | 2435 | 6225 | 12173 | 260 | 3730 | 22513 | 54794 |
| 2014 | 12582 | 74662 | 46438 | 36303 | 2721 | 10651 | 6354 | 8199 | 11816 | 12271 | ... | 14776 | 13737 | 53009 | 2318 | 3485 | 9710 | 77 | 2948 | 18784 | 43369 |
| 2015 | 11099 | 81918 | 47871 | 39942 | 2849 | 9871 | 6601 | 9259 | 11529 | 14713 | ... | 17839 | 15029 | 52724 | 946 | 5792 | 10727 | 70 | 1676 | 17767 | 48594 |
| 2016 | 14219 | 98011 | 49906 | 51576 | 3470 | 8734 | 6251 | 8948 | 11781 | 13441 | ... | 17708 | 14039 | 57241 | 2717 | 5543 | 13219 | 153 | 2517 | 17566 | 55025 |
| 2017 | 9200 | 66735 | 45752 | 39227 | 2004 | 6275 | 4517 | 7517 | 10005 | 9067 | ... | 12650 | 10992 | 50620 | 944 | 4425 | 9400 | 66 | 4420 | 12702 | 44417 |

Pivot table (樞紐分析)

`df.pivot_table(values='腸病毒健保就診人次', index=['年','就診類別'], columns='縣市',aggfunc='sum')`

```
In [19]: df.pivot_table(values='腸病毒健保就診人次',index=['年','就診類別'],columns='縣市',aggfunc='sum')
```

Out[19]:

| 年 | 就診類別 | 縣市 | 南投縣 | 台中市 | 台北市 | 台南市 | 台東縣 | 嘉義市 | 嘉義縣 | 基隆市 | 宜蘭縣 | 屏東縣 | ... | 新竹市 | 新竹縣 | 桃園市 | 澎湖縣 | 花蓮縣 | 苗栗縣 | 連江縣 | 金門縣 | 雲林縣 | 高雄市 |
|------|------|----|-------|--------|-------|-------|------|-------|------|-------|-------|-------|-----|-------|-------|-------|------|------|-------|-----|------|-------|-------|
| | | | | | | | | | | | | | | | | | | | | | | | |
| 2008 | 住院 | | 14 | 765 | 739 | 62 | 6 | 11 | 30 | 78 | 13 | 36 | ... | 158 | 31 | 152 | 0 | 60 | 2 | 1 | 47 | 10 | 291 |
| | 門診 | | 10056 | 55128 | 26905 | 41456 | 2934 | 8274 | 4016 | 4760 | 7288 | 8292 | ... | 7864 | 10312 | 34864 | 1526 | 2605 | 10261 | 81 | 1694 | 14893 | 34811 |
| 2009 | 住院 | | 153 | 986 | 820 | 115 | 59 | 186 | 237 | 124 | 297 | 136 | ... | 243 | 29 | 728 | 0 | 130 | 18 | 0 | 17 | 8 | 620 |
| | 門診 | | 6645 | 46014 | 17825 | 24913 | 2806 | 6779 | 2948 | 4229 | 7042 | 7758 | ... | 5540 | 7781 | 24582 | 2211 | 3159 | 7701 | 90 | 1458 | 11097 | 23736 |
| 2010 | 住院 | | 352 | 2327 | 2305 | 251 | 68 | 408 | 588 | 388 | 1279 | 228 | ... | 545 | 220 | 1838 | 1 | 122 | 202 | 1 | 25 | 54 | 1557 |
| | 門診 | | 23468 | 137646 | 68862 | 72263 | 4521 | 19335 | 8712 | 13457 | 14426 | 19609 | ... | 23046 | 26741 | 84853 | 2581 | 6415 | 19104 | 142 | 3098 | 30818 | 78846 |
| 2011 | 住院 | | 165 | 891 | 1110 | 142 | 63 | 189 | 339 | 207 | 783 | 138 | ... | 277 | 96 | 1188 | 2 | 106 | 24 | 0 | 30 | 61 | 698 |
| | 門診 | | 8758 | 55265 | 34228 | 36128 | 2371 | 10101 | 5179 | 6580 | 6767 | 9455 | ... | 10525 | 11189 | 45576 | 1117 | 3552 | 8617 | 20 | 1061 | 19495 | 34705 |
| 2012 | 住院 | | 128 | 1193 | 1256 | 212 | 68 | 149 | 264 | 206 | 583 | 199 | ... | 327 | 93 | 1152 | 1 | 101 | 101 | 0 | 25 | 20 | 783 |
| | 門診 | | 8768 | 68450 | 44295 | 39727 | 2339 | 8245 | 3790 | 8448 | 9785 | 11674 | ... | 13669 | 14022 | 48974 | 1036 | 4122 | 12364 | 27 | 905 | 15852 | 39225 |