

```
In [ ]: import pandas as pd
```

```
In [ ]: from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive

```
In [ ]: import os
```

```
path = '/content/drive/My Drive/dataviz_2021_data'
```

```
In [ ]: covid_file_path = os.path.join(path, 'pm-20-jan-2021.csv')  
print(covid_file_path)
```

/content/drive/My Drive/dataviz_2021_data/pm-20-jan-2021.csv

```
In [ ]: a = path + '/pm-18-jan-2021.csv'  
print(a)
```

/content/drive/My Drive/dataviz_2021_data/pm-18-jan-2021.csv

load data to memory

```
In [ ]: data_covid = pd.read_csv(covid_file_path)  
data_covid
```

Out[]:

	No.	announce_date	notification_date	sex	age	nationality	province_of_isolation	pro
--	-----	---------------	-------------------	-----	-----	-------------	-----------------------	-----

0	1	1/12/2020		NaN	หญิง	61.0	China	กรุงเทพมหานคร
1	2	1/17/2020		NaN	หญิง	74.0	China	กรุงเทพมหานคร
2	3	1/22/2020		NaN	หญิง	73.0	Thailand	นครปฐม
3	4	1/22/2020		NaN	ชาย	68.0	China	กรุงเทพมหานคร
4	5	1/24/2020		NaN	หญิง	66.0	China	นนทบุรี
...
12648	12649	1/20/2021	1/19/2021	หญิง	44.0	Thailand		ชลบุรี
12649	12650	1/20/2021	1/19/2021	หญิง	52.0	Thailand		ระยอง
12650	12651	1/20/2021	1/19/2021	หญิง	23.0	Thailand		ระยอง
12651	12652	1/20/2021	1/19/2021	หญิง	29.0	Thailand		ระยอง
12652	12653	1/20/2021	1/19/2021	หญิง	22.0	Burma		ตาก

12653 rows × 10 columns



In []: data_covid.head()

Out[]:

	No.	announce_date	notification_date	sex	age	nationality	province_of_isolation	province_o
--	-----	---------------	-------------------	-----	-----	-------------	-----------------------	------------

0	1	1/12/2020	NaN	หญิง	61.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

1	2	1/17/2020	NaN	หญิง	74.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

2	3	1/22/2020	NaN	หญิง	73.0	Thailand	นครปฐม	
---	---	-----------	-----	------	------	----------	--------	--

3	4	1/22/2020	NaN	ชาย	68.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	-----	------	-------	---------------	---------

4	5	1/24/2020	NaN	หญิง	66.0	China	นนทบุรี	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------	---------

In []: data_covid.shape

Out[]: (12653, 10)

ชี้ค่าในตาราง

basic

In []: data_covid['province_of_onset']

```
Out [ ]: 0      กรุงเทพมหานคร
1      กรุงเทพมหานคร
2              นครปฐม
3      กรุงเทพมหานคร
4      กรุงเทพมหานคร
...
12648      ชลบุรี
12649      ระยอง
12650      ระยอง
12651      ระยอง
12652      ดาก
Name: province_of_onset, Length: 12653, dtype: object
```

```
In [ ]: data_covid['province_of_onset'][4]
```

```
Out [ ]: 'กรุงเทพมหานคร'
```

`.iloc`

```
In [ ]: data_covid.iloc[4,7]
```

```
Out [ ]: 'กรุงเทพมหานคร'
```

Table slicing

ตัดมาเฉพาะ column ที่ต้องการ

```
In [ ]: smaller_table = data_covid[['announce_date', 'province_of_onset', 'risk']]
smaller_table
```

```
Out [ ]:
```

	announce_date	province_of_onset	risk
0	1/12/2020	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ
1	1/17/2020	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ
2	1/22/2020	นครปฐม	คนต่างชาติเดินทางมาจากต่างประเทศ
3	1/22/2020	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ
4	1/24/2020	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ
...
12648	1/20/2021	ชลบุรี	State Quarantine
12649	1/20/2021	ระยอง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้
12650	1/20/2021	ระยอง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้
12651	1/20/2021	ระยอง	Cluster ระยอง
12652	1/20/2021	ดาก	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้

12653 rows × 3 columns

ตัดเฉพาะ rows ที่ต้องการ

แบบง่าย แต่ใช้การอะไรไม่ค่อยจะได้ (มองเป็น array แล้วตัดแบบ array)

```
In [ ]: data_covid.iloc[1:5,:]
```

```
Out[ ]:   No.  announce_date  notification_date  sex  age  nationality  province_of_isolation  province_o
```

1	2	1/17/2020	NaN	หญิง	74.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

2	3	1/22/2020	NaN	หญิง	73.0	Thailand	นครปฐม	
---	---	-----------	-----	------	------	----------	--------	--

3	4	1/22/2020	NaN	ชาย	68.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	-----	------	-------	---------------	---------

4	5	1/24/2020	NaN	หญิง	66.0	China	นนทบุรี	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------	---------

แบบ advanced แต่ใช้ประโยชน์ได้เยอะกว่า (ตัดแบบใช้ logic query)

```
In [ ]: data_covid[data_covid['province_of_onset']=='ขอนแก่น']
```

Out[]:

	No.	announce_date	notification_date	sex	age	nationality	province_of_isolation	pro
180	181	3/18/2020	3/15/2020	ชาย	33.0	Thailand	ขอนแก่น	
462	463	3/22/2020	3/21/2020	หญิง	36.0	Thailand	ขอนแก่น	
1466	1467	3/30/2020	3/26/2020	ชาย	19.0	Thailand	ขอนแก่น	
1970	1971	4/3/2020	3/31/2020	หญิง	70.0	Thailand	ขอนแก่น	
2637	2638	4/15/2020	4/14/2020	หญิง	63.0	Thailand	ขอนแก่น	
2673	2674	4/17/2020	4/16/2020	ชาย	68.0	Thailand	ขอนแก่น	
5948	5949	12/26/2020	12/25/2020	หญิง	32.0	Thailand	ขอนแก่น	
6082	6083	12/27/2020	12/26/2020	หญิง	36.0	Thailand	ขอนแก่น	
9333	9334	1/7/2021	1/6/2021	ชาย	17.0	Thailand	ขอนแก่น	
9334	9335	1/7/2021	1/6/2021	ชาย	20.0	Thailand	ขอนแก่น	
10610	10611	1/12/2021	1/11/2021	ชาย	17.0	Thailand	ขอนแก่น	
11517	11518	1/16/2021	1/15/2021	ชาย	12.0	Thailand	ขอนแก่น	
11697	11698	1/17/2021	1/16/2021	หญิง	17.0	Thailand	ขอนแก่น	
11698	11699	1/17/2021	1/16/2021	หญิง	37.0	Thailand	ขอนแก่น	

No. announce_date notification_date sex age nationality province_of_isolation prov

การชี้ค่าในตารางด้วย data_covid[x] ค่า *x* สามารถมีค่าที่ใส่เข้าไปได้ 2 แบบ

1. ชื่อ column
2. ใส่ rows ที่ต้องการ (ใส่ list ที่มี *ขนาดเท่ากับจำนวนแถว* แล้วบอกว่าเอา แถวไหนบ้างโดยใช้ logical expression (True / False))

```
In [ ]: eight_rows_covid = data_covid.iloc[:,8,:] # ตารางตัวอย่าง
eight_rows_covid
```

Out[]:

No.	announce_date	notification_date	sex	age	nationality	province_of_isolation	province_o
-----	---------------	-------------------	-----	-----	-------------	-----------------------	------------

0	1	1/12/2020	NaN	หญิง	61.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

1	2	1/17/2020	NaN	หญิง	74.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

2	3	1/22/2020	NaN	หญิง	73.0	Thailand	นครปฐม	
---	---	-----------	-----	------	------	----------	--------	--

3	4	1/22/2020	NaN	ชาย	68.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	-----	------	-------	---------------	---------

4	5	1/24/2020	NaN	หญิง	66.0	China	นนทบุรี	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------	---------

5	6	1/25/2020	NaN	หญิง	33.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

6	7	1/26/2020	NaN	หญิง	57.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

7	8	1/26/2020	NaN	หญิง	73.0	China	ประจวบคีรีขันธ์	ประจวบ
---	---	-----------	-----	------	------	-------	-----------------	--------


```
In [ ]: eight_rows_covid[[True,False,True,True,False,True,False,False]] #2. ใส่ rows ที่
```

```
Out[ ]: No. announce_date notification_date sex age nationality province_of_isolation province_o
```

0	1	1/12/2020	NaN	หญิง	61.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

2	3	1/22/2020	NaN	หญิง	73.0	Thailand	นครปฐม	
---	---	-----------	-----	------	------	----------	--------	--

3	4	1/22/2020	NaN	ชาย	68.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	-----	------	-------	---------------	---------

5	6	1/25/2020	NaN	หญิง	33.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

```
In [ ]: eight_rows_covid['province_of_onset'] == 'กรุงเทพมหานคร' #สร้าง List ของ logical exp
```

```
Out[ ]: 0    True
        1    True
        2   False
        3    True
        4    True
        5    True
        6    True
        7   False
        Name: province_of_onset, dtype: bool
```

```
In [ ]: eight_rows_covid[eight_rows_covid['province_of_onset'] == 'กรุงเทพมหานคร'] #เราเอา 2
```

Out[]:

No.	announce_date	notification_date	sex	age	nationality	province_of_isolation	province_o
-----	---------------	-------------------	-----	-----	-------------	-----------------------	------------

0	1	1/12/2020	NaN	หญิง	61.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

1	2	1/17/2020	NaN	หญิง	74.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

3	4	1/22/2020	NaN	ชาย	68.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	-----	------	-------	---------------	---------

4	5	1/24/2020	NaN	หญิง	66.0	China	นนทบุรี	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------	---------

5	6	1/25/2020	NaN	หญิง	33.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

6	7	1/26/2020	NaN	หญิง	57.0	China	กรุงเทพมหานคร	กรุงเทพ
---	---	-----------	-----	------	------	-------	---------------	---------

HW (ให้ใช้เฉพาะที่อาจารย์สอนไปแล้วในวิชานี้)

- คำนวณ อายุเฉลี่ย ของผู้หญิง และผู้ชาย ของข้อมูลทั้งหมด
- คำนวณ อายุเฉลี่ย ของผู้หญิง และผู้ชาย ของผู้ป่วยในจังหวัดขอนแก่น

- หาจำนวนผู้ป่วยที่เป็นคน "คนต่างชาติเดินทางมาจากต่างประเทศ"

In []: `this_data = data_covid[['sex', 'age', 'province_of_onset']]`

In []: `this_data`

```
Out[ ]:
```

	sex	age	province_of_onset
0	หญิง	61.0	กรุงเทพมหานคร
1	หญิง	74.0	กรุงเทพมหานคร
2	หญิง	73.0	นครปฐม
3	ชาย	68.0	กรุงเทพมหานคร
4	หญิง	66.0	กรุงเทพมหานคร
...
12648	หญิง	44.0	ชลบุรี
12649	หญิง	52.0	ระยอง
12650	หญิง	23.0	ระยอง
12651	หญิง	29.0	ระยอง
12652	หญิง	22.0	ตาก

12653 rows × 3 columns

```
In [ ]: female = this_data[this_data['sex']=='หญิง']
female['age']
```

```
Out[ ]:
```

0	61.0
1	74.0
2	73.0
4	66.0
5	33.0
...	...
12648	44.0
12649	52.0
12650	23.0
12651	29.0
12652	22.0

Name: age, Length: 5332, dtype: float64

```
In [ ]: sum = 0
N = 0
for a in female['age']:
    if a > 0:
        sum += a # sum = sum + a
        N += 1

print(f'อายุเฉลี่ย ของ ผู้ป่วยหญิง {sum/N}')
```

อายุเฉลี่ย ของ ผู้ป่วยหญิง 36.63955450022614

```
In [ ]: female_KK = female[female['province_of_onset']=='ขอนแก่น']

sum = 0
N = 0
for a in female_KK['age']:
    if a > 0:
        sum += a # sum = sum + a
        N += 1

print(f'อายุเฉลี่ย ของ ผู้ป่วยหญิง ในจังหวัดขอนแก่น {sum/N}')
```