



# Present Project

group

# Maniac



## สมาชิกในกลุ่ม

1

นางสาวกัญญาวิรุ์ ศรีเทียมเงิน 623020511-1 SI

2

นางสาวชลธิชา ศาลางาม 623020518-7 SI



## วัตถุประสงค์

- เพื่อเปรียบเทียบอัตราการเสียชีวิตจากสถานการณ์โควิด-19 ว่าสัมพันธ์กับจำนวนการติดเชื้อหรือไม่
- เพื่อแสดงจำนวนผู้เสียชีวิตทั้งหมดกับจำนวนผู้ติดเชื้อปัจจุบัน

## แหล่งที่มา

ข้อมูลผู้ป่วยโควิด-19 จาก Government Data of Thailand <https://data.go.th/dataset/covid-19-daily>  
ตารางภูมิภาคของประเทศไทย

<https://th.wikipedia.org/wiki/>

ยอดผู้ติดเชื้อปัจจุบันและยอดผู้เสียชีวิตสะสม

<https://ddc.moph.go.th/covid19-dashboard/>

- Import ข้อมูลผู้ป่วยโควิดของทั้ง 4 ชุต โดยตั้งแต่ระยะเวลา 12 มกราคม 2020 - 18 มีนาคม 2022

#### นำเข้าข้อมูล

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

```
In [ ]: from google.colab import drive
drive.mount('/content/drive')
path = '/content/drive/My Drive/data_viz_project'
```

Mounted at /content/drive

```
In [ ]: %cd /content/drive/MyDrive/data_viz_project
```

/content/drive/MyDrive/data\_viz\_project

- confirmed cases ชุดที่ 1

```
In [ ]: data_file_path = os.path.join(path, 'data_covid 1.csv')
data_covid1 = pd.read_csv(data_file_path)
```

- confirmed cases ชุดที่ 2

```
In [ ]: data_file_path = os.path.join(path, 'data_covid 2.csv')
data_covid2 = pd.read_csv(data_file_path)
```

- confirmed cases ชุดที่ 3

```
In [ ]: data_file_path = os.path.join(path, 'data_covid 3.csv')
data_covid3 = pd.read_csv(data_file_path)
```

- confirmed cases ชุดที่ 4

```
In [ ]: data_file_path = os.path.join(path, 'data_covid 4.csv')
data_covid4 = pd.read_csv(data_file_path)
```

- รวมข้อมูลทั้ง 4 ชุดเข้าด้วยกัน เพื่อสร้างตารางใหม่ โดยใช้คำสั่ง `pd.concat`

### รวมข้อมูลทั้ง 4 ชุดเข้าด้วยกัน

```
In [ ]: data_covid = pd.concat([data_covid1 , data_covid2 , data_covid3 , data_covid4])
data_covid
```

```
Out[ ]:
```

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	1	12/1/2020	หญิง	61.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
1	2	17/1/2020	หญิง	74.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
2	3	22/1/2020	หญิง	73.0	ปี	Thailand	นครปฐม	คนต่างชาติเดินทางมาจากต่างประเทศ	นครปฐม	เมือง
3	4	22/1/2020	ชาย	68.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
4	5	24/1/2020	หญิง	66.0	ปี	China	นนทบุรี	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
...	...	...	...	...	...	...	...	...	...	...
433548	3303165	18/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเทสสถาน/เรือนจำ	NaN	NaN
433549	3303166	18/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเทสสถาน/เรือนจำ	NaN	NaN
433550	3303167	18/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเทสสถาน/เรือนจำ	NaN	NaN
433551	3303168	18/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัตเทสสถาน/เรือนจำ	NaN	NaN
433552	3303169	18/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัตเทสสถาน/เรือนจำ	NaN	NaN

3325951 rows × 10 columns

## • เลือกคอลัมน์และสร้างตารางข้อมูลจังหวัดภูมิภาค

5

- เลือกข้อมูลที่สนใจ คือ province\_of\_inset และทำการสร้างตารางจังหวัดภูมิภาคโดยมีข้อมูลผู้ติดเชื้อปัจจุบัน คือ 18 มีนาคม 2022

### เลือกคอลัมน์ province\_of\_onset

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

### สร้างตารางภูมิภาคโดยมีผู้ป่วยติดเชื้อปัจจุบันในตารางเดียวกัน

```
In [ ]: data_region = [['อำนาจเจริญ', 'ตะวันออกเฉียงเหนือ', '63'],
                        ['บึงกาฬ', 'ตะวันออกเฉียงเหนือ', '111'],
                        ['บุรีรัมย์', 'ตะวันออกเฉียงเหนือ', '484'],
                        ['ชัยภูมิ', 'ตะวันออกเฉียงเหนือ', '119'],
                        ['กาฬสินธุ์', 'ตะวันออกเฉียงเหนือ', '285'],
                        ['ขอนแก่น', 'ตะวันออกเฉียงเหนือ', '554'],
                        ['เลย', 'ตะวันออกเฉียงเหนือ', '139'],
                        ['มหาสารคาม', 'ตะวันออกเฉียงเหนือ', '175'],
                        ['มุกดาหาร', 'ตะวันออกเฉียงเหนือ', '21'],
                        ['นครพนม', 'ตะวันออกเฉียงเหนือ', '99'],
                        ['นครราชสีมา', 'ตะวันออกเฉียงเหนือ', '468'],
                        ['หนองบัวลำภู', 'ตะวันออกเฉียงเหนือ', '120'],
                        ['หนองคาย', 'ตะวันออกเฉียงเหนือ', '256'],
                        ['ร้อยเอ็ด', 'ตะวันออกเฉียงเหนือ', '566'],
                        ['สกลนคร', 'ตะวันออกเฉียงเหนือ', '253'],
                        ['ศรีสะเกษ', 'ตะวันออกเฉียงเหนือ', '311'],
                        ['สุรินทร์', 'ตะวันออกเฉียงเหนือ', '391'],
                        ['อุบลราชธานี', 'ตะวันออกเฉียงเหนือ', '341'],
                        ['อุดรธานี', 'ตะวันออกเฉียงเหนือ', '379'],
                        ['ยโสธร', 'ตะวันออกเฉียงเหนือ', '140'],
                        ['เขียงใหม่', 'เหนือ', '393'],
                        ['เขียงราย', 'เหนือ', '78'],
                        ['ลำปาง', 'เหนือ', '184'],
                        ['ลำพูน', 'เหนือ', '1'],
                        ['แม่ฮ่องสอน', 'เหนือ', '34'],
                        ['น่าน', 'เหนือ', '81'],
                        ['พะเยา', 'เหนือ', '73'],
                        ['แพร่', 'เหนือ', '73'],
                        ['อุดรดีด', 'เหนือ', '40'],
                        ['ตาก', 'ตะวันตก', '163'],
                        ['สุโขทัย', 'กลาง', '173'],
                        ['พิษณุโลก', 'กลาง', '188'],
                        ['พิจิตร', 'กลาง', '51'],
                        ['กำแพงเพชร', 'กลาง', '169'],
                        ['เพชรบูรณ์', 'กลาง', '101'],
                        ['นครสวรรค์', 'กลาง', '254'],
                        ['อุทัยธานี', 'กลาง', '113'],
                        ['อ่างทอง', 'กลาง', '231'],
                        ['ชัยนาท', 'กลาง', '37'],
```

- ใช้คำสั่ง pd.DataFrame เพื่อสร้างตาราง

```
In [ ]: data_region2 = pd.DataFrame.from_records(data_region, columns=header)
data_region2
```

```
Out[ ]:
```

	จังหวัด	ภาค	ผู้ติดเชื้อปัจจุบัน
0	อำนาจเจริญ	ตะวันออกเฉียงเหนือ	63
1	บึงกาฬ	ตะวันออกเฉียงเหนือ	111
2	บุรีรัมย์	ตะวันออกเฉียงเหนือ	484
3	ชัยภูมิ	ตะวันออกเฉียงเหนือ	119
4	กาฬสินธุ์	ตะวันออกเฉียงเหนือ	285
...	...	...	...
72	พังงา	ใต้	94
73	ภูเก็ต	ใต้	357
74	ระนอง	ใต้	232
75	สตูล	ใต้	294
76	ตรัง	ใต้	173

77 rows x 3 columns

- รวบรวม data\_covidupdate (ข้อมูลโควิดกับภูมิภาค) กับ data\_region2 โดยใช้คำสั่ง .merge

รวมตารางของ data\_covidupdate กับ data\_region2 โดยใช้คำสั่ง merge

```
In [ ]: data = data_covidupdate.merge(data_region2, how='left', left_on='province_of_onset', right_on='จังหวัด')
data
```

```
Out[ ]:
```

	province_of_onset	จังหวัด	ภาค	ผู้ติดเชื้อปัจจุบัน
0	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
1	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
2	นครปฐม	นครปฐม	กลาง	642
3	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
4	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
...	...	...	...	...
3325946	NaN	NaN	NaN	NaN
3325947	NaN	NaN	NaN	NaN
3325948	NaN	NaN	NaN	NaN
3325949	NaN	NaN	NaN	NaN
3325950	NaN	NaN	NaN	NaN

3325951 rows × 4 columns

- ทำการแก้ไขค่า missing ของข้อมูลที่ไม่ระบุค่าโดยใช้วิธี dropna() ในการแก้ไขจากค่า True ให้เป็น False

```
In [ ]: print(data.isnull().any())
print('* * 100)
data2 = data.dropna()
print(data2.isnull().any())
data2

province_of_onset      True
จังหวัด                True
ภาค                  True
ผู้ติดเชื้อปัจจุบัน      True
dtype: bool
*****
province_of_onset      False
จังหวัด                False
ภาค                  False
ผู้ติดเชื้อปัจจุบัน      False
dtype: bool
```

Out[ ]:

	province_of_onset	จังหวัด	ภาค	ผู้ติดเชื้อปัจจุบัน
0	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
1	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
2	นครปฐม	นครปฐม	กลาง	642
3	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
4	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
...	...	...	...	...
3325842	นครราชสีมา	นครราชสีมา	ตะวันออกเฉียงเหนือ	468
3325844	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
3325845	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
3325846	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	3273
3325847	กำแพงเพชร	กำแพงเพชร	กลาง	169

2920332 rows × 4 columns



- ใช้คำสั่ง groupby.size เพื่อรวมยอดผู้ติดเชื้อในแต่ละจังหวัด และใช้คำสั่ง .merge เพื่อรวมตาราง data3 กับ data\_region2

### ใช้คำสั่ง groupby.size เพื่อรวมยอดผู้ติดเชื้อแต่ละจังหวัด

```
In [ ]: data3 = data2.groupby('จังหวัด').size().reset_index(name='ยอดจำนวนผู้ติดเชื้อ')
data3
```

```
Out[ ]:
```

	จังหวัด	ยอดจำนวนผู้ติดเชื้อ
0	กระบี่	17658
1	กรุงเทพมหานคร	568574
2	กาญจนบุรี	24531
3	กาฬสินธุ์	18586
4	กำแพงเพชร	13078
...	...	...
72	เพชรบุรี	30295
73	เพชรบูรณ์	18485
74	เลย	9986
75	แพร่	5667
76	แม่ฮ่องสอน	4145

77 rows × 2 columns

### ใช้คำสั่ง .merge เพื่อรวมตาราง

```
In [ ]: data4 = data3.merge(data_region2, how='right', left_on='จังหวัด', right_on='จังหวัด')
data4
```

```
Out[ ]:
```

	จังหวัด	ยอดจำนวนผู้ติดเชื้อ	ภาค	ผู้ติดเชื้อปัจจุบัน
0	อำนาจเจริญ	2995	ตะวันออกเฉียงเหนือ	63
1	บึงกาฬ	4988	ตะวันออกเฉียงเหนือ	111
2	บุรีรัมย์	7821	ตะวันออกเฉียงเหนือ	484
3	ชัยภูมิ	5029	ตะวันออกเฉียงเหนือ	119
4	กาฬสินธุ์	18586	ตะวันออกเฉียงเหนือ	285
...	...	...	...	...
72	พังงา	11208	ใต้	94
73	ภูเก็ต	48999	ใต้	357
74	ระนอง	14783	ใต้	232
75	สตูล	14200	ใต้	294
76	ตรัง	22767	ใต้	173

77 rows × 4 columns

- ข้อมูลที่เลือกคือ ข้อมูลผู้เสียชีวิตสะสมจากโควิด-19 โดยเริ่มตั้งแต่ระยการแพร่ระบาด - 18 มีนาคม 2022 ซึ่งเป็นยอดผู้เสียชีวิตแต่ละจังหวัด

นำเข้าข้อมูลที่เลือกเองโดยข้อมูลคือ ผู้เสียชีวิตจากโควิด-19

```
In [ ]: data_file_path = os.path.join(path, 'ผู้เสียชีวิต.xlsx')  
data5 = pd.read_excel(data_file_path)
```

```
In [ ]: data5
```

```
Out[ ]:
```

	No.	จังหวัด	ผู้เสียชีวิตสะสม	หน่วย
0	1	กรุงเทพมหานคร	7378	คน
1	2	นนทบุรี	433	คน
2	3	ปทุมธานี	907	คน
3	4	สมุทรปราการ	1597	คน
4	5	สมุทรสาคร	957	คน
...	...	...	...	...
72	73	อุดรธานี	202	คน
73	74	อุดรดิตถ์	88	คน
74	75	อุทัยธานี	44	คน
75	76	อุบลราชธานี	275	คน
76	77	อ่างทอง	163	คน

77 rows × 4 columns

- เลือกข้อมูลที่ต้องการ คือ จังหวัดและผู้เสียชีวิตสะสม หลังจากนั้นใช้ .merge เพื่อรวมตารางโดยใช้ 'จังหวัด' เป็นตัวเชื่อม

### เลือกคอลัมน์ จังหวัด และ ผู้เสียชีวิตสะสม

```
In [ ]: data5updated = data5[['จังหวัด', 'ผู้เสียชีวิตสะสม']]
data5updated
```

```
Out[ ]:
```

	จังหวัด	ผู้เสียชีวิตสะสม
0	กรุงเทพมหานคร	7378
1	นนทบุรี	433
2	ปทุมธานี	907
3	สมุทรปราการ	1597
4	สมุทรสาคร	957
...	...	...
72	อุดรธานี	202
73	อุดรดิษฐ์	88
74	อุทัยธานี	44
75	อุบลราชธานี	275
76	อ่างทอง	163

77 rows × 2 columns

### รวมตารางโดยใช้คำสั่ง .merge เพื่อใช้จังหวัดเป็นตัวเชื่อมทั้ง 2 ตาราง

```
In [ ]: data6 = data4.merge(data5updated, how='right', left_on='จังหวัด', right_on='จังหวัด')
data6
```

```
Out[ ]:
```

	จังหวัด	ยอดจำนวนผู้ติดเชื้อ	ภาค	ผู้ติดเชื้อปัจจุบัน	ผู้เสียชีวิตสะสม
0	กรุงเทพมหานคร	568574.0	กลาง	3273	7378
1	นนทบุรี	118032.0	กลาง	510	433
2	ปทุมธานี	86307.0	กลาง	667	907
3	สมุทรปราการ	185942.0	กลาง	976	1597
4	สมุทรสาคร	122128.0	กลาง	942	957
...	...	...	...	...	...
72	อุดรธานี	14264.0	ตะวันออกเฉียงเหนือ	379	202
73	อุดรดิษฐ์	9193.0	เหนือ	40	88
74	อุทัยธานี	5768.0	กลาง	113	44
75	อุบลราชธานี	27896.0	ตะวันออกเฉียงเหนือ	341	275
76	อ่างทอง	16726.0	กลาง	231	163

77 rows × 5 columns

- ตรวจสอบค่า missing อีกรอบ แล้วแก้ไขค่า missing โดยการ dropna() หลังจากนั้นก็จะได้ข้อมูลที่ต้องการเพื่อนำไปโหลดข้อมูลทำ Dashboard ต่อไป

### ตรวจสอบ ค่า missing อีกรอบ

```
In [ ]: print(data6.isnull().any())
print('*** 100)

จังหวัด                False
ยอดจำนวนผู้ติดเชื้อ    True
ภาค                   True
ผู้ติดเชื้อปัจจุบัน     True
ผู้เสียชีวิตสะสม        False
dtype: bool
*****
```

### แก้ไขค่า missing โดยการ dropna()

```
In [ ]: data6_dropped = data6.dropna()

In [ ]: print(data6_dropped.isnull().any())

จังหวัด                False
ยอดจำนวนผู้ติดเชื้อ    False
ภาค                   False
ผู้ติดเชื้อปัจจุบัน     False
ผู้เสียชีวิตสะสม        False
dtype: bool
```

### เลือกข้อมูลเพื่อเตรียมทำ dashboard

```
In [ ]: data6_dropped.to_csv('Dataacovid-19.csv', encoding='utf-8')
```

# Dashboard

