

Project dashboard

กลุ่มโน้่น

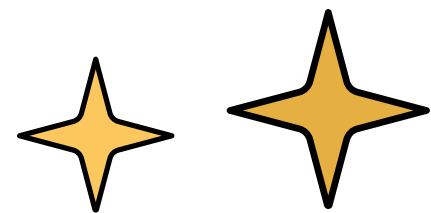


Dash
board

รายชื่อสมาชิกกลุ่ม

[51]	นางสาวทักษิณา ราชหุ่น	623021000-1	STAT
[60]	นางสาวกิตติมา อุปสุข	623021040-9	SI
[70]	นางสาวนฤมล ไสยโสภณ	623021050-6	SI





จุดมุ่งหมายที่สร้าง dashboard

- เพื่อต้องการทราบว่าผู้ติดเชืื่อนั้นเป็นเพศใดมากกว่ากัน และมีความเสี่ยงและจังหวัดใดบ้างที่ติดเชื้อมากที่สุด
- เพื่อต้องการทราบว่าในแต่ละภูมิภาคมีจำนวนผู้ติดเชื้อมีอย่างไร
- เพื่อต้องการทราบว่าในแต่ละเขตสุขภาพ มีจำนวนผู้ติดเชื้อที่อยู่ในความรับผิดชอบเป็นเท่าไร



Data ทั้ง 3 ที่ใช้ในการทำ Dashboard

- การเตรียมข้อมูลและนำเข้าข้อมูล

ข้อมูลโควิด ที่มา : <https://data.go.th/dataset/covid-19-daily>

องค์กร : กรมควบคุมโรค

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

จากวิกิพีเดีย สารานุกรมเสรี

ข้อมูลของเขตสุขภาพ ที่มา : <https://www.mhc2.go.th/newweb/information?>

องค์กร : ศูนย์สุขภาพจิต

SCAN ME



scan me



แสดงข้อมูลของผู้ติดเชื้อตั้งแต่รายแรก จนถึงรายปัจจุบัน ตั้งแต่วันที่ 12 มกราคม 2020 ถึง 12 มีนาคม 2022 โดยข้อมูลที่ต้องการนำมาใช้ คือ วันที่ของการติดเชื้อ เพศ ความเสี่ยง (ที่ติดเชื้อ) และจังหวัด

```
1 import pandas as pd # import ฟังก์ชัน pandasimport pandas as pd # import ฟังก์ชัน pandas
```

```
1 from google.colab import drive # เชื่อม drive ของเรา ถ้าเชื่อมสำเร็จจะขึ้นคำว่าMounted at /content/drive  
2 drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
1 import os # จัดการเกี่ยวกับไฟล์ต่างๆ ในไฟล์เดอร์
```

```
1 path = '/content/drive/My Drive/dataviz_2021_data' # ชื่อ data เราอยู่ไฟล์เดอร์ไหน
```



ข้อมูลโควิด ตั้งแต่วันที่ 12 มกราคม 2020 ถึง 11 สิงหาคม 2021

```
1 data_file_path1 = os.path.join(path,'confirmed-cases.csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา
2 data_covid1 = pd.read_csv(data_file_path1)
```

```
1 data_covid1
```

	No.	announce_date	Notified date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	1	12/1/2020	NaN	หญิง	61.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
1	2	17/1/2020	NaN	หญิง	74.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
2	3	22/1/2020	NaN	หญิง	73.0	ปี	Thailand	นครปฐม	คนต่างชาติเดินทางมาจากต่างประเทศ	นครปฐม	เมือง
3	4	22/1/2020	NaN	ชาย	68.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
4	5	24/1/2020	NaN	หญิง	66.0	ปี	China	นนทบุรี	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	NaN
...
816984	816985	11/8/2021	10/8/2021	หญิง	8.0	ปี	Thailand	สระแก้ว	อยู่ระหว่างการสอบสวน	สระแก้ว	ตาพระยา
816985	816986	11/8/2021	10/8/2021	ชาย	14.0	ปี	Thailand	สระแก้ว	อยู่ระหว่างการสอบสวน	สระแก้ว	ตาพระยา
816986	816987	11/8/2021	10/8/2021	ชาย	59.0	ปี	Thailand	สระแก้ว	อาชีพเสี่ยง เช่น ทำงานในสถานที่แออัด หรือทำงาน...	สระแก้ว	อรัญประเทศ
816987	816988	11/8/2021	10/8/2021	หญิง	34.0	ปี	Thailand	สระแก้ว	อื่นๆ	กรุงเทพมหานคร	บางขุนเทียน
816988	816989	11/8/2021	10/8/2021	หญิง	19.0	ปี	Thailand	สระแก้ว	อื่นๆ	พระนครศรีอยุธยา	อุทัย

816989 rows × 11 columns



ข้อมูลโควิด ตั้งแต่วันที่ 12 สิงหาคม 2021 ถึง 25 ตุลาคม 2021

```
1 data_file_path2 = os.path.join(path,'confirmed-cases-since-120864.csv')
2 data_covid2 = pd.read_csv(data_file_path2)
```

```
1 data_covid2
```

	No.	announce_date	Notified date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	816990	12/8/2021	11/8/2021	ชาย	7.0	ปี	Thailand	เชียงราย	อื่นๆ	NaN	NaN
1	816991	12/8/2021	11/8/2021	ชาย	1.0	ปี	Thailand	เชียงราย	อื่นๆ	NaN	NaN
2	816992	12/8/2021	11/8/2021	ชาย	35.0	ปี	Thailand	เชียงราย	อื่นๆ	NaN	NaN
3	816993	12/8/2021	11/8/2021	หญิง	33.0	ปี	Thailand	เชียงราย	อื่นๆ	NaN	NaN
4	816994	12/8/2021	11/8/2021	หญิง	14.0	ปี	Thailand	เชียงราย	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	NaN	NaN
...
1042163	1859153	25/10/2021	24/10/2021	ชาย	45.0	ปี	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1042164	1859154	25/10/2021	24/10/2021	ชาย	31.0	ปี	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1042165	1859155	25/10/2021	24/10/2021	ชาย	31.0	ปี	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1042166	1859156	25/10/2021	24/10/2021	ชาย	24.0	ปี	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1042167	1859157	25/10/2021	24/10/2021	ชาย	25.0	ปี	NaN	เชียงใหม่	หัตถสถาน/เรือนจำ	NaN	NaN

1042168 rows × 11 columns



ข้อมูลโควิด ตั้งแต่วันที่ 26 ตุลาคม 2021 ถึง 27 กุมภาพันธ์ 2022

```
1 data_file_path3 = os.path.join(path,'confirmed-cases-since-271064.csv')
2 data_covid3 = pd.read_csv(data_file_path3)
```

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:2882: DtypeWarning: Columns (4) have mixed types.Specify dtype option on import or set low_memory=False.
exec(code_obj, self.user_global_ns, self.user_ns)

```
1 data_covid3
```

	No.	announce_date	Notified date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	1859158	26/10/2021	25/10/2021	ชาย	20.0	ปี	Thailand	สาป่าง	อื่นๆ	สาป่าง	เมืองสาป่าง
1	1859159	26/10/2021	25/10/2021	หญิง	42.0	ปี	Thailand	สาป่าง	อื่นๆ	สาป่าง	เมืองสาป่าง
2	1859160	26/10/2021	25/10/2021	หญิง	33.0	ปี	Thailand	สาป่าง	ไปสถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	สาป่าง	ห้างฉัตร
3	1859161	26/10/2021	25/10/2021	หญิง	52.0	ปี	Thailand	สาป่าง	ไปสถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	สาป่าง	ห้างฉัตร
4	1859162	26/10/2021	25/10/2021	หญิง	84.0	ปี	Thailand	สาป่าง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	สาป่าง	เมืองสาป่าง
...
1010454	2869612	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1010455	2869613	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
1010456	2869614	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	หัตถสถาน/เรือนจำ	NaN	NaN
1010457	2869615	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	หัตถสถาน/เรือนจำ	NaN	NaN
1010458	2869616	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	หัตถสถาน/เรือนจำ	NaN	NaN

1010459 rows × 11 columns



ข้อมูลโควิด ตั้งแต่วันที่ 28 กุมภาพันธ์ 2022 ถึง 12 มีนาคม 2022

```
1 data_file_path4 = os.path.join(path, 'confirmed-cases-since-280265.csv')
2 data_covid4 = pd.read_csv(data_file_path4)
```

```
1 data_covid4
```

	No.	announce_date	Notified date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	2869617	28/2/2022	27/2/2022	ชาย	41.0	ปี	Thailand	พะเยา สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้		พะเยา	เมืองพะเยา
1	2869618	28/2/2022	27/2/2022	หญิง	45.0	ปี	Thailand	พะเยา สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้		พะเยา	เมืองพะเยา
2	2869619	28/2/2022	27/2/2022	หญิง	40.0	ปี	Thailand	พะเยา สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้		พะเยา	เมืองพะเยา
3	2869620	28/2/2022	27/2/2022	ชาย	44.0	ปี	Thailand	พะเยา สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้		พะเยา	เมืองพะเยา
4	2869621	28/2/2022	27/2/2022	ชาย	11.0	ปี	Thailand	พะเยา สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้		พะเยา	เมืองพะเยา
...
291620	3161237	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291621	3161238	12/3/2022	11/3/2022	ชาย	NaN	NaN	Cambodian	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291622	3161239	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291623	3161240	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291624	3161241	12/3/2022	11/3/2022	NaN	NaN	NaN	NaN	เชียงใหม่	หัตถสถาน/เรือนจำ	NaN	NaN

291625 rows × 11 columns



ข้อมูลของภูมิภาค

• ใช้วิธี List (Row Oriented)

```
1 data_region= pd.DataFrame.from_records(data,columns=headers) # สร้าง dataframe
2 data_region
```

	จังหวัด	ภูมิศาสตร์	การเมือง	อุดมศึกษา	การท่องเที่ยว
0	อำนาจเจริญ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ
1	บึงกาฬ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ
2	บุรีรัมย์	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ
3	ชัยภูมิ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ
4	กาฬสินธุ์	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ	ตะวันออกเฉียงเหนือ
...
72	พังงา	ใต้	ใต้	ใต้ฝั่งตะวันตก/ใต้ฝั่งอันดามัน	ใต้
73	ภูเก็ต	ใต้	ใต้	ใต้ฝั่งตะวันตก/ใต้ฝั่งอันดามัน	ใต้
74	ระนอง	ใต้	ใต้	ใต้ฝั่งตะวันตก/ใต้ฝั่งอันดามัน	ใต้
75	สตูล	ใต้	ใต้	ใต้ฝั่งตะวันตก/ใต้ฝั่งอันดามัน	ใต้
76	ตรัง	ใต้	ใต้	ใต้ฝั่งตะวันตก/ใต้ฝั่งอันดามัน	ใต้

77 rows × 5 columns



ข้อมูลเขตสุขภาพ

- ใช้วิธี Dictionary (column oriented)

```
1 data_mentalhealth= pd.DataFrame.from_dict(data_mhc) # สร้างตาราง pandas จาก data_mhc
2 data_mentalhealth #แสดงข้อมูลในตารางทั้งหมด
```

	จังหวัด	เขตสุขภาพ
0	กรุงเทพมหานคร	เขตสุขภาพส่วนกลาง
1	เชียงใหม่	เขตสุขภาพที่ 1
2	เชียงราย	เขตสุขภาพที่ 1
3	แพร่	เขตสุขภาพที่ 1
4	น่าน	เขตสุขภาพที่ 1
...
72	ตรัง	เขตสุขภาพที่ 12
73	พัทลุง	เขตสุขภาพที่ 12
74	ปัตตานี	เขตสุขภาพที่ 12
75	ยะลา	เขตสุขภาพที่ 12
76	นราธิวาส	เขตสุขภาพที่ 12

77 rows × 2 columns



รวมข้อมูลโควิด ทั้ง 4 ไฟล์ให้อยู่ในตารางเดียวกัน

```
1 data_covid = pd.concat([data_covid1,data_covid2,data_covid3,data_covid4])
2 data_covid
```

	No.	announce_date	Notified date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset	district_of_onset
0	1	12/1/2020	NaN	หญิง	61.0	ปี	China	กรุงเทพมหานคร คนต่างชาติเดินทางมาจากต่างประเทศ		กรุงเทพมหานคร	NaN
1	2	17/1/2020	NaN	หญิง	74.0	ปี	China	กรุงเทพมหานคร คนต่างชาติเดินทางมาจากต่างประเทศ		กรุงเทพมหานคร	NaN
2	3	22/1/2020	NaN	หญิง	73.0	ปี	Thailand	นครปฐม คนต่างชาติเดินทางมาจากต่างประเทศ		นครปฐม	เมือง
3	4	22/1/2020	NaN	ชาย	68.0	ปี	China	กรุงเทพมหานคร คนต่างชาติเดินทางมาจากต่างประเทศ		กรุงเทพมหานคร	NaN
4	5	24/1/2020	NaN	หญิง	66.0	ปี	China	นนทบุรี คนต่างชาติเดินทางมาจากต่างประเทศ		กรุงเทพมหานคร	NaN
...
291620	3161237	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291621	3161238	12/3/2022	11/3/2022	ชาย	NaN	NaN	Cambodian	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291622	3161239	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291623	3161240	12/3/2022	11/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	หัตถสถาน/เรือนจำ	NaN	NaN
291624	3161241	12/3/2022	11/3/2022	NaN	NaN	NaN	NaN	เชียงใหม่	หัตถสถาน/เรือนจำ	NaN	NaN

3184023 rows × 11 columns



PREPROCESSING (COLAB NOTEBOOK)

จัดการค่า Missing และ
ตรวจสอบค่า Missing

```
1 data_covid.isnull().any() # ดูว่ามีค่า missing หรือไม่
```

No.	False
announce_date	False
Notified date	True
sex	True
age	True
Unit	True
nationality	True
province_of_isolation	True
risk	True
province_of_onset	True
district_of_onset	True
dtype:	bool



PREPROCESSING (COLAB NOTEBOOK)

แทนค่า ไม่ระบุ ในค่า missing

```
1 set(data_covid['sex']) # ดูว่า sex เก็บค่าอะไรไว้บ้าง
```

```
{nan,  
'\xa0หญิง',  
'ช',  
'ชาย',  
'ญ',  
'นาย',  
'ร.ด.ท.',  
'หญิง',  
'หญิง',  
'หญิง',  
'หญิง',  
'ทท'}
```

```
1 data_covid['sex'].mode() # หาค่า mode ของ sex
```

```
0 หญิง  
dtype: object
```

```
1 data_covid[['sex']] = data_covid[['sex']].fillna(value={'sex':'ไม่ระบุ'}) # แทนค่า mode ในค่า missing  
2 data_covid[['sex']]
```

	sex
0	หญิง
1	หญิง
2	หญิง
3	ชาย
4	หญิง
...	...
291620	หญิง
291621	ชาย
291622	หญิง
291623	หญิง
291624	ไม่ระบุ

3184023 rows × 1 columns

PREPROCESSING (COLAB NOTEBOOK)

แก้ข้อมูลที่ผิดให้อยู่ในรูปแบบเดียวกัน

```
1 # แก้ข้อมูลที่ผิดให้อยู่ในรูปแบบเดียวกัน
2 data_covid['sex'].replace({'\xa0หญิง':'หญิง',
3                             'ช':'ชาย',
4                             'ญ':'หญิง',
5                             'นาย':'ชาย',
6                             'ร.ด.ท.':'ชาย',
7                             'หญิง':'หญิง',
8                             'หญิง':'หญิง',
9                             'หญิง':'หญิง',
10                            'พ':'หญิง'}, inplace=True)
```



PREPROCESSING (COLAB NOTEBOOK)

จัดการค่า Missing ของ risk โดยจะแทนว่า ไม่ระบุ

```
1 data_covid[['risk']] = data_covid[['risk']].fillna(value={'risk': 'ไม่ระบุ'}) # แทนค่า 'ไม่ระบุ' ในค่า missing
2 data_covid[['risk']]
```

	risk
0	คนต่างชาติเดินทางมาจากต่างประเทศ
1	คนต่างชาติเดินทางมาจากต่างประเทศ
2	คนต่างชาติเดินทางมาจากต่างประเทศ
3	คนต่างชาติเดินทางมาจากต่างประเทศ
4	คนต่างชาติเดินทางมาจากต่างประเทศ
...	...
291620	หัตถสถาน/เรือนจำ
291621	หัตถสถาน/เรือนจำ
291622	หัตถสถาน/เรือนจำ
291623	หัตถสถาน/เรือนจำ
291624	หัตถสถาน/เรือนจำ

3184023 rows × 1 columns

```
1 set(data_covid['risk'])
```

```
{' Samui plus',
'10.อื่นๆ',
'2.สัมผัสผู้ติดเชื้อ',
'8.สำรวจกลุ่มเสี่ยง (survey)',
'9.ขอตรวจหาเชื้อเอง',
'ACF รongเมือง',
'ACF สนามกีฬาประเทมิย',
'ACF สนามกีฬาไทย-ญี่ปุ่น',
'ACF สีสม',
'ACF เคหะหลักสี่',
'BKK Sandbox',
'Chonburi Sandbox',
'Cluster Icon siam',
"Cluster Memory 90's กรุงเทพมหานคร",
'Cluster New Jazz กรุงเทพมหานคร',
'Cluster The Lounge Salaya',
'Cluster คลองเตย',
'Cluster งานมอเตอร์โชว์',
'Cluster จนท. สนามบินสุวรรณภูมิ',
'Cluster จันทบุรี',
'Cluster จันทบุรี (ชาวถิ่น)',
'Cluster จันทบุรี (ชาวถิ่น)',
'Cluster ชลบุรี',
...}
```



PREPROCESSING (COLAB NOTEBOOK)

แก้ข้อมูลที่ผิดให้อยู่ในรูปแบบเดียวกัน

```
1 # แก้ข้อมูลที่ผิดให้อยู่ในรูปแบบเดียวกัน
2 data_covid['risk'].replace({' Samui plus':'Samui Plus',
3                             '10.อื่นๆ':'อื่นๆ',
4                             '2.สัมผัสผู้ติดเชื้อ':'สัมผัสผู้ติดเชื้อ',
5                             '8.สำรวจกลุ่มเสี่ยง (survey)':'กลุ่มเสี่ยง',
6                             '9.ขอตรวจหาเชื้อเอง':'ขอตรวจหาเชื้อเอง',
7                             'Cluster จันทบุรี (ชาวกัมพูชา)':'Cluster จันทบุรี (ชาวกัมพูชา)',
8                             'Cluster ตราด ':'Cluster ตราด',
9                             'Karbi Sandbox':'Kabi Sandbox',
10                            'Kabi Sandbox ':'Kabi Sandbox',
11                            'Kabi sandbox ':'Kabi Sandbox',
12                            'Phang nga Sandbox':'Phang Nga Sandbox',
13                            'Phangnga Sandbox':'Phang Nga Sandbox',
14                            'Phare Sandbox':'Phrae Sandbox',
15                            'Phunga Sandbox':'Phung Nga Sandbox',
16                            'SHUI':'อื่นๆ',
17                            'Samui plus':'Samui Plus',
18                            'cluster ชลบุรี(อยู่ระหว่างการสอบสวน)':'cluster ชลบุรี',
19                            'prachuap Sandbox':'Prachuap Sandbox',
20                            'กลับมาจาก จ.กรุงเทพมหานคร':'ผู้เดินทางมาจากพื้นที่เสี่ยง',
21                            'คนไทยมาจากพื้นที่เสี่ยง':'ผู้เดินทางมาจากพื้นที่เสี่ยง',
22                            'คนต่างชาติเดินทางมาจากต่างประเทศ':'ผู้ที่เดินทางมาจากต่างประเทศ',
23                            'คนไทยเดินทางกลับจากต่างประเทศ':'ผู้ที่เดินทางมาจากต่างประเทศ',
24                            'คนไทยเดินทางมาจากต่างประเทศ':'ผู้ที่เดินทางมาจากต่างประเทศ',
25                            'ติดเชื้อในพื้นที่ (สัมผัส ผป.ยืนยัน)':'พื้นที่เสี่ยง',
26                            'บุคลากรทางการแพทย์':'บุคลากรด้านการแพทย์และสาธารณสุข',
```

PREPROCESSING (COLAB NOTEBOOK)

จัดการค่า Missing ของ province_of_onset

โดยแทนค่า ไม่ระบุ ในค่า missing

```
1 data_covid[['province_of_onset']] = data_covid[['province_of_onset']].fillna(value={'province_of_onset': 'ไม่ระบุ'}) # แทนค่า mode ในค่า missing
2 data_covid[['province_of_onset']]
```

	province_of_onset
0	กรุงเทพมหานคร
1	กรุงเทพมหานคร
2	นครปฐม
3	กรุงเทพมหานคร
4	กรุงเทพมหานคร
...	...
291620	ไม่ระบุ
291621	ไม่ระบุ
291622	ไม่ระบุ
291623	ไม่ระบุ
291624	ไม่ระบุ

3184023 rows × 1 columns



PREPROCESSING (COLAB NOTEBOOK)

แก้ข้อมูลที่ผิดให้อยู่ในรูปแบบเดียวกัน

```
1 set(data_covid['province_of_onset']) # ดูว่าเก็บค่าอะไรไว้บ้าง
```

```
{'\tประจวบคีรีขันธ์',  
'USA',  
'\xa0กาญจนบุรี',  
'\xa0ชลบุรี',  
'\xa0ตราด',  
'\xa0นครปฐม',  
'\xa0สมุทรสงคราม',  
'\xa0สมุทรสาคร',  
'\xa0อุทัยธานี',  
'กทม',  
'กทม.',  
'กบินทร์บุรี',  
'กระบี่',  
'กระบี่',  
'กรุงเทพ',  
'กรุงเทพมหานคร',  
'กรุงเทพมหานคร.',  
'กรุงเทพมหานคร. / พระนครศรีอยุธยา',  
'กรุงเทพมหานคร. / ศรีสะเกษ',  
'กรุงเทพฯ',  
'กะทู้',  
'กัมพูชา',  
'กาญจนบุรี',  
'กาญจนบุรี',  
'กาญจนบุรี',  
'กาฬสินธุ์',
```

```
1 data_covid['province_of_onset'].replace({'กาญจนบุรี':'กาญจนบุรี',  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
        'ลำพูน':'ลำพูน',  
        'พัทยา':'ชลบุรี',  
        'อุดรธานี':'อุดรธานี',  
        'เพชรบูรณ์':'เพชรบูรณ์',  
        'เพชรบูรณ์':'เพชรบูรณ์',  
        'ประจวบคีรีขันธ์ ':'ประจวบคีรีขันธ์',  
        '\xa0สมุทรสาคร':'สมุทรสาคร',  
        'อุดรธานี':'อุดรธานี',  
        'กำแพงเพชร':'กำแพงเพชร',  
        'หนองบัวลำภู':'หนองบัวลำภู',  
        'อำนาจเจริญ':'อำนาจเจริญ'}, inplace=True)
```



PREPROCESSING (COLAB NOTEBOOK)

ตัดเฉพาะ column ที่ต้องการนำมาใช้

```
data_clean = data_covid[['announce_date', 'sex', 'risk', 'province_of_onset']]
data_clean
```

	announce_date	sex	risk	province_of_onset
0	12/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
1	17/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
2	22/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	นครปฐม
3	22/1/2020	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
4	24/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร

ตรวจสอบค่า missing อีกครั้ง พบว่าเป็น False แสดงว่าไม่มีค่า missing

```
1 data_clean.isnull().any() # เช็คค่า missing
```

```
announce_date    False
sex               False
risk              False
province_of_onset False
dtype: bool
```



PREPROCESSING (COLAB NOTEBOOK)

เลือกภูมิภาคที่ต้องการจะนำไปใช้

ในที่นี้กลุ่มของพวกเราเลือกใช้การแบ่งภูมิภาคแบบการเมือง

```
data_policy = data_region[['จังหวัด','การเมือง']]
data_policy
```

	จังหวัด	การเมือง
0	อำนาจเจริญ	ตะวันออกเฉียงเหนือ
1	บึงกาฬ	ตะวันออกเฉียงเหนือ
2	บุรีรัมย์	ตะวันออกเฉียงเหนือ
3	ชัยภูมิ	ตะวันออกเฉียงเหนือ
4	กาฬสินธุ์	ตะวันออกเฉียงเหนือ

PREPROCESSING (COLAB NOTEBOOK)

รวมตารางทุกตาราง ได้แก่ ข้อมูลโควิด ภูมิภาค และเขตสุขภาพ เป็นตารางเดียวกัน
ด้วยวิธีการ merge

โดยจะรวมตารางโควิด กับ ภูมิภาคก่อน

```
data_covpol = data_clean.merge(data_policy,how='left',left_on='province_of_onset',right_on='จังหวัด') # รวมตาราง data_clean  
data_covpol
```

	announce_date	sex	risk	province_of_onset	จังหวัด	การเมือง
0	12/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง
1	17/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง
2	22/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	นครปฐม	นครปฐม	กลาง
3	22/1/2020	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง
4	24/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง

และจากนั้นนำมารวมตาราง กับ เขตสุขภาพ

```
data = data_covpol.merge(data_mentalhealth,how='left',left_on='จังหวัด',right_on='จังหวัด') # รวมตาราง data  
data
```

	announce_date	sex	risk	province_of_onset	จังหวัด	การเมือง	เขตสุขภาพ
0	12/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
1	17/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2	22/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	นครปฐม	นครปฐม	กลาง	เขตสุขภาพที่ 5
3	22/1/2020	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
4	24/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง

PREPROCESSING (COLAB NOTEBOOK)

เปลี่ยนชื่อคอลัมน์จาก การเมือง เป็น ภูมิภาค

```
data.rename(columns={'การเมือง': 'ภูมิภาค'}, inplace=True)
```

data

	announce_date	sex	risk	province_of_onset	จังหวัด	ภูมิภาค	เขตสุขภาพ
0	12/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
1	17/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2	22/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	นครปฐม	นครปฐม	กลาง	เขตสุขภาพที่ 5
3	22/1/2020	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
4	24/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง

PREPROCESSING (COLAB NOTEBOOK)

ตรวจสอบว่ามีค่า missing ทั้งหมดกี่ตัว

```
1 data.isnull().sum() # ดูว่ามีค่า missing กี่ตัว
```

```
announce_date      0
sex                 0
risk                0
province_of_onset   0
จังหวัด             503545
ภูมิภาค             503545
เขตสุขภาพ           517754
dtype: int64
```



PREPROCESSING (COLAB NOTEBOOK)

เนื่องจากค่า missing ที่เกิดขึ้นนั้น มาจากข้อมูลของ data['province_of_onset'] ว่า ไม่
ระบุ ทำให้ไม่สามารถนำมาเชื่อมกับตารางเขตสุขภาพของ data_mentalhealth['จังหวัด']
ได้ จึงเกิดค่า NaN ขึ้น ดังนั้นจะทำการ drop ค่าทิ้งไป

```
data_project = data.dropna() # ลบค่า missing  
data_project
```

	announce_date	sex	risk	province_of_onset	จังหวัด	ภูมิภาค	เขตสุขภาพ
0	12/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
1	17/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2	22/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	นครปฐม	นครปฐม	กลาง	เขตสุขภาพที่ 5
3	22/1/2020	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
4	24/1/2020	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง

PREPROCESSING (COLAB NOTEBOOK)

เนื่องจากไฟล์ของ data_project นั้นมีข้อมูลมากเกินไป ทำให้ไม่สามารถนำข้อมูลไปใช้ในการทำ Dashboard ได้ ดังนั้น จึงตัดข้อมูลเหลือเพียงวันที่ที่ต้องการดู

```
data_project['announce_date'] = pd.to_datetime(data_project['announce_date'],format="%d/%m/%Y") # เปลี่ยน format ของข้อมูลวันที่ให้อยู่ในรูปแบบเดียวกัน
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
"""Entry point for launching an IPython kernel.
```

```
data_project[['announce_date']]
```

	announce_date
0	2020-01-12
1	2020-01-17
2	2020-01-22
3	2020-01-22
4	2020-01-24



PREPROCESSING (COLAB NOTEBOOK)

วันที่ 1 ธันวาคม 2020 ถึง วันที่ 31 มกราคม 2021 เป็นระยะเวลา 2 เดือน
(ต้องการทราบว่าในช่วงเฉลิมฉลองส่งท้ายปี 2020 มีผู้ติดเชื้อเป็นอย่างไรบ้าง)

```
1 data_11220_31121 = data_project[(data_project['announce_date']>='2020-12-01') & (data_project['announce_date']<='2021-01-31')]  
2 data_11220_31121
```

	announce_date	sex	risk	province_of_onset	จังหวัด	ภูมิภาค	เขตสุขภาพ
4008	2020-12-01	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	สมุทรปราการ	สมุทรปราการ	กลาง	เขตสุขภาพที่ 6
4009	2020-12-01	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	สมุทรปราการ	สมุทรปราการ	กลาง	เขตสุขภาพที่ 6
4010	2020-12-01	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
4011	2020-12-01	หญิง	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
4012	2020-12-01	ชาย	ผู้ที่เดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
...
17994	2021-01-31	ชาย	Cluster มหาสารคาม	มหาสารคาม	มหาสารคาม	ตะวันออกเฉียงเหนือ	เขตสุขภาพที่ 7
17995	2021-01-31	หญิง	Cluster มหาสารคาม	ราชบุรี	ราชบุรี	กลาง	เขตสุขภาพที่ 5
17996	2021-01-31	หญิง	Cluster สมุทรสาคร	สมุทรสงคราม	สมุทรสงคราม	กลาง	เขตสุขภาพที่ 5
17997	2021-01-31	ชาย	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	นครปฐม	นครปฐม	กลาง	เขตสุขภาพที่ 5
17998	2021-01-31	หญิง	การค้นหาผู้ป่วยเชิงรุกและค้นหาผู้ติดเชื้อในชุมชน	สมุทรสงคราม	สมุทรสงคราม	กลาง	เขตสุขภาพที่ 5

3919 rows × 7 columns

PREPROCESSING (COLAB NOTEBOOK)

วันที่ 1 ธันวาคม 2021 ถึง วันที่ 31 มกราคม 2022 เป็นระยะเวลา 2 เดือน
(ต้องการทราบว่าในช่วงเฉลิมฉลองส่งท้ายปี 2021 มีผู้ติดเชื้อเป็นอย่างไรบ้าง)

```
1 data_11221_31122 = data_project[(data_project['announce_date']>='2021-12-01') & (data_project['announce_date']<='2022-01-31')]
2 data_11221_31122
```


	announce_date	sex	risk	province_of_onset	จังหวัด	ภูมิภาค	เขตสุขภาพ
2153688	2021-12-01	หญิง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	เชียงราย	เชียงราย	เหนือ	เขตสุขภาพที่ 1
2153689	2021-12-01	หญิง	ไม่ระบุ	เชียงราย	เชียงราย	เหนือ	เขตสุขภาพที่ 1
2153690	2021-12-01	ชาย	ไม่ระบุ	เชียงราย	เชียงราย	เหนือ	เขตสุขภาพที่ 1
2153691	2021-12-01	หญิง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	เชียงราย	เชียงราย	เหนือ	เขตสุขภาพที่ 1
2153692	2021-12-01	หญิง	ไม่ระบุ	เชียงราย	เชียงราย	เหนือ	เขตสุขภาพที่ 1
...
2480793	2022-01-31	ชาย	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2480794	2022-01-31	หญิง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2480795	2022-01-31	หญิง	ไม่ระบุ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง	เขตสุขภาพส่วนกลาง
2480796	2022-01-31	ชาย	ไปสถานที่ชุมชน	นนทบุรี	นนทบุรี	กลาง	เขตสุขภาพที่ 4
2480797	2022-01-31	ชาย	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้า	ปทุมธานี	ปทุมธานี	กลาง	เขตสุขภาพที่ 4

288696 rows × 7 columns



PREPROCESSING (COLAB NOTEBOOK)

Save file ออกมา เพื่อนำไปใช้ในการทำ Dashboad



```
1 data_11221_31122.to_csv('data_11221_31122.csv', encoding='utf-8')
```

```
1 data_11220_31121.to_csv('data_11220_31121.csv', encoding='utf-8')
```



Dashboard



**THANK
YOU!**

