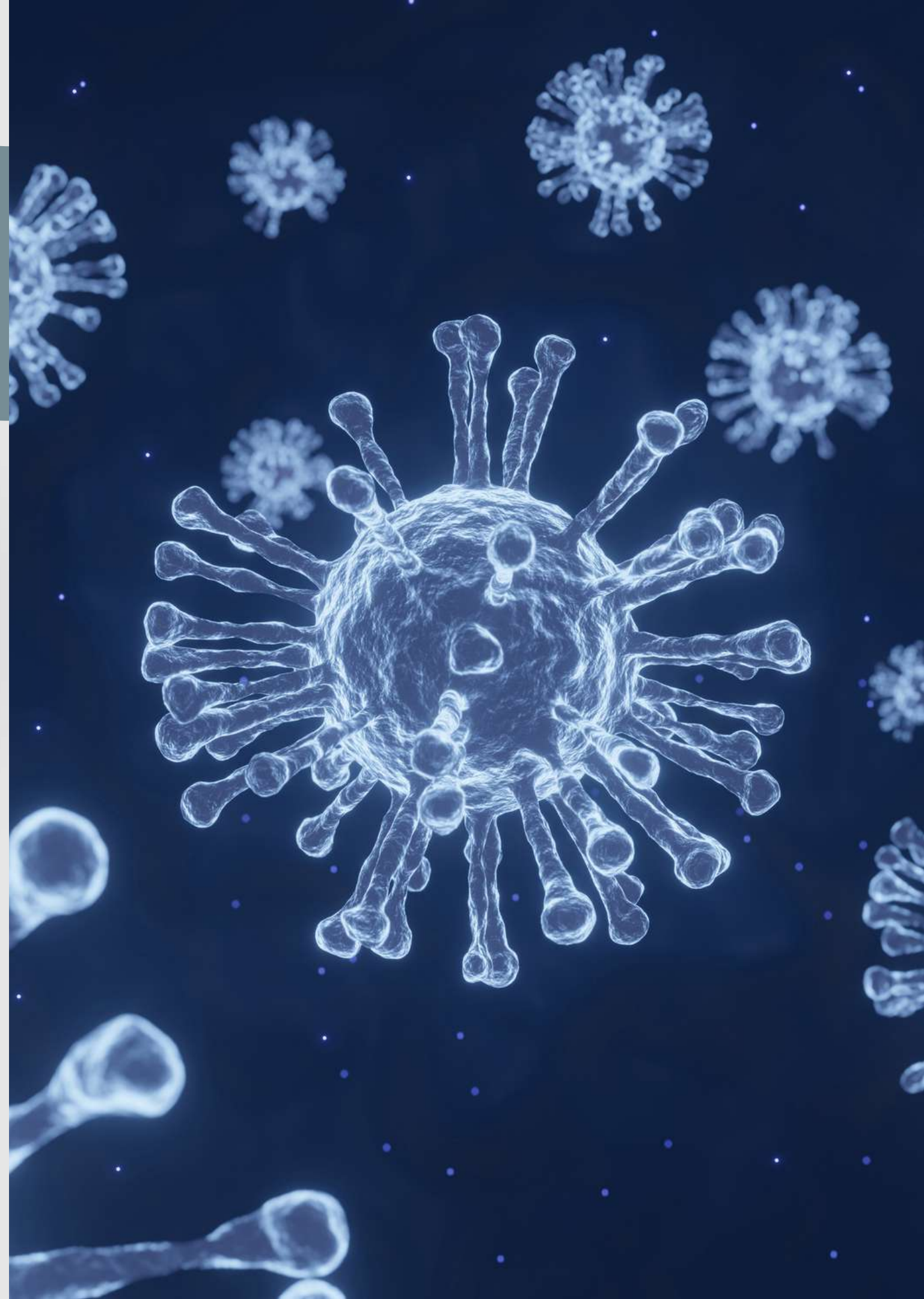


PRESENT PROJECT

กลุ่มนั้น



สมาชิก



ชญ์ญชิตา
เมธีกุลมานิต

623020516-1

หลักสูตรสารสนเทศสภิติ



นริศรา
ชัยฤทธิ์

623021001-9

หลักสูตรสภิติ



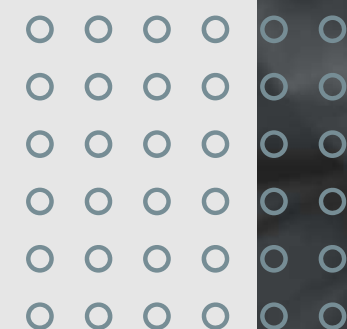
ภิญญาดา
เพ็ญสุข

623021052-2

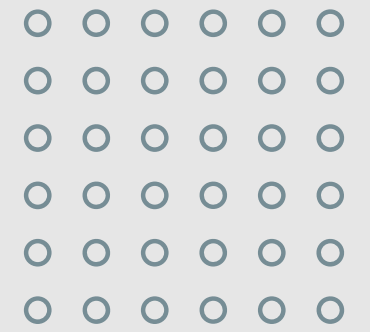
หลักสูตรสารสนเทศสภิติ

จุดมุ่งหมายที่สร้าง Dashboard

ต้องการแสดงจำนวนผู้ติดเชื้อโควิด-19 เปรียบเทียบกับจำนวนโดสวัคซีน
แต่ละประเภทที่ถูกฉีดไป(เฉพาะประชากรในประเทศไทย)จำแนกตามรายจังหวัด



ข้อมูลที่น่าสนใจทั้งหมด 3 ข้อมูล



1) รายงาน Covid-19 ประจำวัน ข้อมูลประจำประเทศไทย (ตั้งแต่วันที่ 12/01/2020 ถึง 18/03/2022)

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

2) ข้อมูลการแบ่งตามภูมิภาคของประเทศไทย

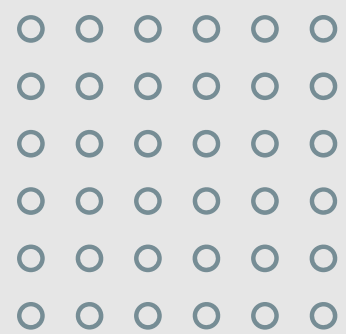
แหล่งที่มา: <https://shorturl.asia/T3Ss5>



3) ข้อมูลการฉีดวัคซีนรายจังหวัด แยกตามชนิดวัคซีน

แหล่งที่มา: <https://shorturl.asia/ijnIg>





โหลดข้อมูลผู้ติดเชื้อโควิด-19 เข้า COLAB



ติดตั้ง Package ที่ต้องใช้ในการโหลดข้อมูลจาก Drive เข้ามาใน Colab

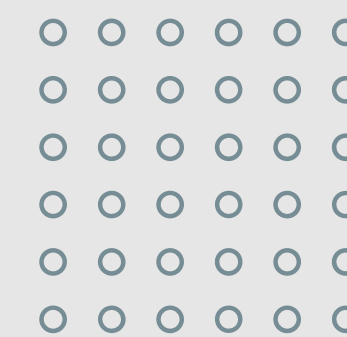
```
[1] 1 import pandas as pd # import ฟังก์ชัน pandas
```

```
[2] 1 from google.colab import drive # package เชื่อม google drive  
2 drive.mount('/content/drive')
```

Mounted at /content/drive

```
[3] 1 import os # จัดการเกี่ยวกับไฟล์ต่างๆ ในโฟลเดอร์
```

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

```
[ ] 1 data_file_path = os.path.join(path, 'confirmed(1).csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ
    2 datacovid1 = pd.read_csv(data_file_path)
    3 datacovid1

/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)
```

	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
...
839766	839767	12/8/2021	11/8/2021	ชาย	52.0	ปี	NaN	กาญจนบุรี	ทัตสถาน/เรือนจำ	กาญจนบุรี	NaN
839767	839768	12/8/2021	11/8/2021	ชาย	34.0	ปี	Thailand	กาญจนบุรี	ทัตสถาน/เรือนจำ	กาญจนบุรี	NaN
839768	839769	12/8/2021	11/8/2021	ชาย	31.0	ปี	Thailand	กาญจนบุรี	ทัตสถาน/เรือนจำ	กาญจนบุรี	NaN
839769	839770	12/8/2021	11/8/2021	ชาย	83.0	ปี	Thailand	กาญจนบุรี	ทัตสถาน/เรือนจำ	กาญจนบุรี	NaN
839770	839771	12/8/2021	11/8/2021	ชาย	47.0	ปี	Thailand	กาญจนบุรี	ทัตสถาน/เรือนจำ	กาญจนบุรี	NaN

839771 rows x 11 columns

ข้อมูลผู้ติดเชื้อโควิด-19 วันที่ 12/01/2020 ถึง 12/08/2021



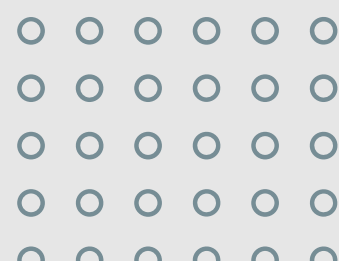
ข้อมูลผู้ติดเชื้อโควิด-19 วันที่ 12/08/2021 ถึง 25/10/2021



```
[ ] 1 data_file_path = os.path.join(path, 'confirmed(2).csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ
    2 datacovid2 = pd.read_csv(data_file_path)
    3 datacovid2
```

	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	ทัตสถาน/เรือนจำ	NaN	NaN
1042164	1859154	25/10/2021	24/10/2021	ชาย	31.0	ปี	NaN	กรุงเทพมหานคร	ทัตสถาน/เรือนจำ	NaN	NaN
1042165	1859155	25/10/2021	24/10/2021	ชาย	31.0	ปี	NaN	กรุงเทพมหานคร	ทัตสถาน/เรือนจำ	NaN	NaN
1042166	1859156	25/10/2021	24/10/2021	ชาย	24.0	ปี	NaN	กรุงเทพมหานคร	ทัตสถาน/เรือนจำ	NaN	NaN
1042167	1859157	25/10/2021	24/10/2021	ชาย	25.0	ปี	NaN	เชียงใหม่	ทัตสถาน/เรือนจำ	NaN	NaN

1042168 rows x 11 columns



ข้อมูลผู้ติดเชื้อโควิด-19

วันที่ 26/10/2021 ถึง 27/02/2022



```
[ ] 1 data_file_path = os.path.join(path, 'confirmed(3).csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ
    2 datacovid3 = pd.read_csv(data_file_path)
    3 datacovid3
```

/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)

	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	ลำปาง	อื่นๆ	ลำปาง	เมืองลำปาง
1	1859159	26/10/2021	25/10/2021	หญิง	42.0	ปี	Thailand	ลำปาง	อื่นๆ	ลำปาง	เมืองลำปาง
2	1859160	26/10/2021	25/10/2021	หญิง	33.0	ปี	Thailand	ลำปาง	ไปสถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	ลำปาง	ห้างจัตดร
3	1859161	26/10/2021	25/10/2021	หญิง	52.0	ปี	Thailand	ลำปาง	ไปสถานที่ชุมชน เช่น ตลาดนัด สถานที่ท่องเที่ยว	ลำปาง	ห้างจัตดร
4	1859162	26/10/2021	25/10/2021	หญิง	84.0	ปี	Thailand	ลำปาง	สัมผัสใกล้ชิดกับผู้ป่วยยืนยันรายก่อนหน้านี้	ลำปาง	เมืองลำปาง
...
1010454	2869612	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	กรุงเทพมหานคร	ทัณฑสถาน/เรือนจำ	NaN	NaN
1010455	2869613	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	กรุงเทพมหานคร	ทัณฑสถาน/เรือนจำ	NaN	NaN
1010456	2869614	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	ทัณฑสถาน/เรือนจำ	NaN	NaN
1010457	2869615	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	ทัณฑสถาน/เรือนจำ	NaN	NaN
1010458	2869616	27/2/2022	26/2/2022	ชาย	NaN	NaN	NaN	ปัตตานี	ทัณฑสถาน/เรือนจำ	NaN	NaN

1010459 rows x 11 columns

```
[ ] 1 data_file_path = os.path.join(path, 'confirmed(4).csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ
    2 datacovid4 = pd.read_csv(data_file_path)
    3 datacovid4
```

	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
...
433548	3303165	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัณฑสถาน/เรือนจำ	NaN	NaN
433549	3303166	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัณฑสถาน/เรือนจำ	NaN	NaN
433550	3303167	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัณฑสถาน/เรือนจำ	NaN	NaN
433551	3303168	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัณฑสถาน/เรือนจำ	NaN	NaN
433552	3303169	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัณฑสถาน/เรือนจำ	NaN	NaN

433553 rows x 11 columns

ข้อมูลผู้ติดเชื้อโควิด-19

วันที่ 28/02/2022 ถึง 18/03/2022



ต่อตารางผู้ติดเชื้อโควิด-19 ทั้ง 4 ข้อมูลไว้ในตารางเดียว โดยใช้คำสั่ง .concat

```
[ ] 1 datamix = pd.concat([datacovid1,datacovid2,datacovid3,datacovid4]) # รวมตาราง Data Covid ทั้ง 4 ข้อมูล โดยใช้คำสั่ง .concat()  
2 datamix
```

	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
...
433548	3303165	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเตสสถาน/เรือนจำ	NaN	NaN
433549	3303166	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเตสสถาน/เรือนจำ	NaN	NaN
433550	3303167	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	กรุงเทพมหานคร	ทัตเตสสถาน/เรือนจำ	NaN	NaN
433551	3303168	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัตเตสสถาน/เรือนจำ	NaN	NaN
433552	3303169	18/3/2022	17/3/2022	หญิง	NaN	NaN	NaN	สงขลา	ทัตเตสสถาน/เรือนจำ	NaN	NaN

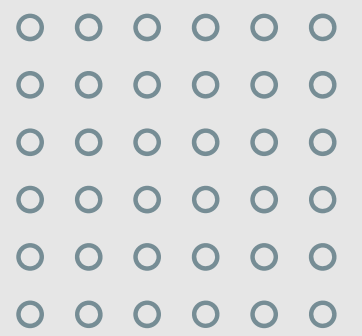
3325951 rows × 11 columns

```
[ ] 1 datamix.isnull().any() # เช็คค่า missing
```

```
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
```

ตรวจสอบค่า missing ทุกคอลัมน์ในตาราง datamix
จะเห็นว่ามียังคอลัมน์ No. และ announce_date ที่แสดง
เป็น False เนื่องจากไม่มีค่า missing

แทนค่า missing คอลัมน์ sex ด้วย class ใหม่เป็น 'ไม่ระบุ'



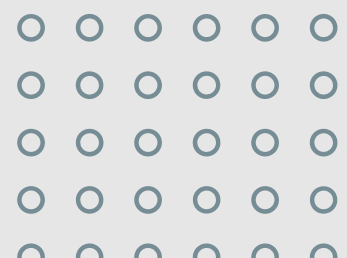
```
[ ] 1 datamix[['sex']] = datamix[['sex']].fillna('ไม่ระบุ')

[ ] 1 set(datamix['sex']) # ดูว่ามีกลุ่มข้อมูลใดบ้าง

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

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

แทนที่ข้อมูลที่เขียนผิดให้อยู่ในกลุ่ม ไม่ระบุ



```
[ ] 1 set(datamix['sex'])

{'ชาย', 'หญิง', 'ไม่ระบุ'}
```

```
[ ] 1 datamix.isnull().any()

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

ดังนั้นข้อมูลจะมีเพียง 3 กลุ่ม
จากนั้นเช็ค missing

แทนค่า missing คอลัมน์ province_of_onset ด้วย class ใหม่เป็น 'ไม่ระบุ'

```
[ ] 1 datamix[['province_of_onset']] = datamix[['province_of_onset']].fillna('ไม่ระบุ')
```

```
[ ] 1 len(set(datamix['province_of_onset'])) # ดูว่าข้อมูลใน province_of_onset มีกี่ค่า  
2 # จะเห็นได้ว่ามี 437 ค่า แสดงว่าข้อมูลมีค่าผิดปกติ เนื่องจาก ประเทศไทยมี 77 จังหวัด
```

437

```
[ ] 1 set(datamix['province_of_onset'])
```

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

ข้อมูลจังหวัดมีทั้งหมด 437 กลุ่ม แสดงว่าข้อมูลมีค่าผิดปกติ
จะเห็นได้ว่าการกรอกข้อมูลผิดอยู่นหลายข้อมูล



```
[ ] 1 datamix['province_of_onset'].replace({'\tประจวบคีรีขันธ์': 'ประจวบคีรีขันธ์',
2                                     'USA': 'ไม่ระบุ',
3                                     '\xa0กาญจนบุรี': 'กาญจนบุรี',
4                                     '\xa0ชลบุรี': 'ชลบุรี',
5                                     '\xa0ตราด': 'ตราด',
6                                     '\xa0นครปฐม': 'นครปฐม',
7                                     '\xa0สมุทรสงคราม': 'สมุทรสงคราม',
8                                     '\xa0สมุทรสาคร': 'สมุทรสาคร',
9                                     '\xa0อุทัยธานี': 'อุทัยธานี',
10                                    'กทม': 'กรุงเทพมหานคร',
11                                    'กทม.': 'กรุงเทพมหานคร',
12                                    'กบินทร์บุรี': 'ปราจีนบุรี',
13                                    'กระบี่': 'กระบี่',
14                                    'กระบี่': 'กระบี่',
15                                    'กรุงเทพ': 'กรุงเทพมหานคร',
16                                    'กรุงเทพมหานคร': 'กรุงเทพมหานคร',
17                                    'กรุงเทพมหานคร.': 'กรุงเทพมหานคร',
18                                    'กรุงเทพมหานคร. / พระนครศรีอยุธยา': 'กรุงเทพมหานคร',
19                                    'กรุงเทพมหานคร. / ศรีสะเกษ': 'กรุงเทพมหานคร',
20                                    'กรุงเทพฯ': 'กรุงเทพมหานคร',
21                                    'กะทู้': 'ไม่ระบุ',
22                                    'กัมพูชา': 'ไม่ระบุ',
```

○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○

ทำการแทนที่ข้อมูลที่ผิดปกติ
ให้กลุ่มในกลุ่ม ไม่ระบุ

```
[ ] 1 len(set(datamix['province_of_onset']))
```

78

เมื่อแทนที่ข้อมูลเรียบร้อยแล้ว
เช็คกลุ่มข้อมูลได้ 78 กลุ่ม
แสดงว่าไม่มีค่าผิดปกติ

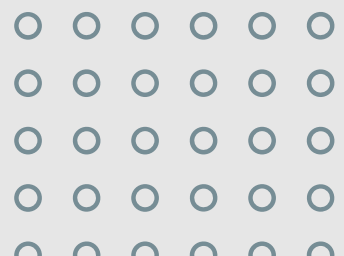
เลือกคอลัมน์ข้อมูลผู้ติดเชื้อโควิด-19 ที่จะนำไปสร้าง DashBoard



```
[ ] 1 datamix = datamix[['No.', 'announce_date', 'sex', 'province_of_onset']]  
    2 datamix
```

	No.	announce_date	sex	province_of_onset
0	1	12/1/2020	หญิง	กรุงเทพมหานคร
1	2	17/1/2020	หญิง	กรุงเทพมหานคร
2	3	22/1/2020	หญิง	นครปฐม
3	4	22/1/2020	ชาย	กรุงเทพมหานคร
4	5	24/1/2020	หญิง	กรุงเทพมหานคร
...
433548	3303165	18/3/2022	หญิง	ไม่ระบุ
433549	3303166	18/3/2022	หญิง	ไม่ระบุ
433550	3303167	18/3/2022	หญิง	ไม่ระบุ
433551	3303168	18/3/2022	หญิง	ไม่ระบุ
433552	3303169	18/3/2022	หญิง	ไม่ระบุ

3325951 rows × 4 columns



นับจำนวนผู้ติดเชื้อของ Data Covid แต่ละจังหวัด โดยนำไปไว้ในตารางชื่อ 'infected'

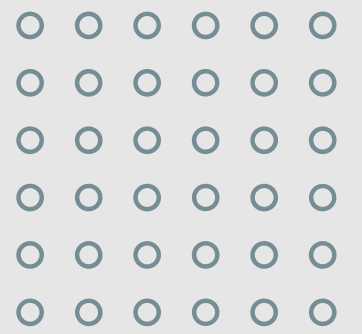
```
[ ] 1 cv_province = datamix.groupby('province_of_onset').size().reset_index(name = 'infected')
     2 cv_province
```

	province_of_onset	infected
0	กระบี่	17667
1	กรุงเทพมหานคร	572034
2	กาญจนบุรี	24534
3	กาฬสินธุ์	18596
4	กำแพงเพชร	13106
...
73	เพชรบูรณ์	18497
74	เลย	9990
75	แพร่	5668
76	แม่ฮ่องสอน	4145
77	ไทรบุรี	400450

78 rows × 2 columns

นับจำนวนเพศของ Data Covid แต่ละจังหวัด

```
[ ] 1 # นับจำนวนของเพศหญิงในคอลัมน์ sex ที่อยู่ในกลุ่มจังหวัดเดียวกัน(province_of_onset) มาใส่คอลัมน์ที่ชื่อว่า female
2 cv_female = datamix[datamix['sex']=='หญิง'].groupby('province_of_onset').size().reset_index(name='female')
3 # นับจำนวนของเพศชายในคอลัมน์ sex ที่อยู่ในกลุ่มจังหวัดเดียวกัน(province_of_onset) มาใส่คอลัมน์ที่ชื่อว่า male
4 cv_male = datamix[datamix['sex']=='ชาย'].groupby('province_of_onset').size().reset_index(name='male')
5 # นับจำนวนของเพศ 'ไม่ระบุ' ในคอลัมน์ sex ที่อยู่ในกลุ่มจังหวัดเดียวกัน(province_of_onset) มาใส่คอลัมน์ที่ชื่อว่า other
6 cv_other = datamix[datamix['sex']=='ไม่ระบุ'].groupby('province_of_onset').size().reset_index(name='other')
```



[] 1 cv_female

	province_of_onset	female
0	กระบี่	9430
1	กรุงเทพมหานคร	266381
2	กาญจนบุรี	13294
3	กาฬสินธุ์	9807
4	กำแพงเพชร	6985
...
73	เพชรบูรณ์	8526
74	เลย	5520
75	แพร่	2926
76	แม่ฮ่องสอน	2217
77	ไม่ระบุ	162814

78 rows × 2 columns

[] 1 cv_male

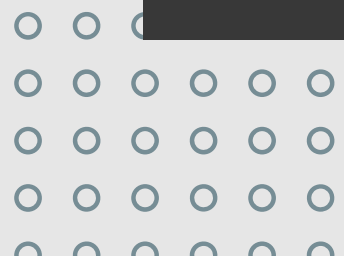
	province_of_onset	male
0	กระบี่	7273
1	กรุงเทพมหานคร	250966
2	กาญจนบุรี	10756
3	กาฬสินธุ์	7815
4	กำแพงเพชร	5881
...
73	เพชรบูรณ์	7924
74	เลย	4299
75	แพร่	2498
76	แม่ฮ่องสอน	1813
77	ไม่ระบุ	171943

78 rows × 2 columns

[] 1 cv_other

	province_of_onset	other
0	กระบี่	964
1	กรุงเทพมหานคร	54687
2	กาญจนบุรี	484
3	กาฬสินธุ์	974
4	กำแพงเพชร	240
...
73	เพชรบูรณ์	2047
74	เลย	171
75	แพร่	244
76	แม่ฮ่องสอน	115
77	ไม่ระบุ	65693

78 rows × 2 columns



ต่อตารางเพศหญิง-ชาย โดยใช้ province_of_onset เป็นคีย์เชื่อม

คำสั่ง .merge()

```
[ ] 1 # ทำการต่อตารางเพศหญิง-ชาย โดยใช้ province_of_onset เป็นคีย์เชื่อม  
2 cv_sex1 = cv_female.merge(cv_male,how='left',left_on='province_of_onset',right_on='province_of_onset')  
3 cv_sex1
```

	province_of_onset	female	male
0	กระบี่	9430	7273
1	กรุงเทพมหานคร	266381	250966
2	กาญจนบุรี	13294	10756
3	กาฬสินธุ์	9807	7815
4	กำแพงเพชร	6985	5881
...
73	เพชรบูรณ์	8526	7924
74	เลย	5520	4299
75	แพร่	2926	2498
76	แม่ฮ่องสอน	2217	1813
77	ไม่ระบุ	162814	171943

78 rows × 3 columns

นำตารางที่ได้เมื่อสักครู่มาต่อกับเพศ ไม่ระบุ โดยใช้ province_of_onset เป็นคีย์เชื่อม คำสั่ง .merge()

```
[ ] 1 # นำตารางเพศที่ 'ไม่ระบุ (cv_other)' มาต่อกับตารางที่ต่อแล้ว(cv_sex1) โดยใช้ province_of_onset เป็นคีย์เชื่อม  
2 cv_sex2 = cv_sex1.merge(cv_other,how='left',left_on='province_of_onset',right_on='province_of_onset')  
3 cv_sex2
```

	province_of_onset	female	male	other
0	กระบี่	9430	7273	964
1	กรุงเทพมหานคร	266381	250966	54687
2	กาญจนบุรี	13294	10756	484
3	กาฬสินธุ์	9807	7815	974
4	กำแพงเพชร	6985	5881	240
...
73	เพชรบูรณ์	8526	7924	2047
74	เลย	5520	4299	171
75	แพร่	2926	2498	244
76	แม่ฮ่องสอน	2217	1813	115
77	ไม่ระบุ	162814	171943	65693

78 rows × 4 columns

ต่อตารางผู้ติดเชื้อ กับ เพศ โดยใช้ province_of_onset เป็นคีย์เชื่อม

คำสั่ง .merge()

```
[ ] 1 cv_covid = cv_province.merge(cv_sex2,how='left',left_on='province_of_onset',right_on='province_of_onset')
2 cv_covid
```

	province_of_onset	infected	female	male	other
0	กระบี่	17667	9430	7273	964
1	กรุงเทพมหานคร	572034	266381	250966	54687
2	กาญจนบุรี	24534	13294	10756	484
3	กาฬสินธุ์	18596	9807	7815	974
4	กำแพงเพชร	13106	6985	5881	240
...
73	เพชรบูรณ์	18497	8526	7924	2047
74	เลย	9990	5520	4299	171
75	แพร่	5668	2926	2498	244
76	แม่ฮ่องสอน	4145	2217	1813	115
77	ไม่ระบุ	400450	162814	171943	65693

78 rows × 5 columns

เช็คค่า missing

```
[ ] 1 cv_covid.isnull().any()

province_of_onset    False
infected              False
female               False
male                 False
other                False
dtype: bool
```


สร้างตารางจังหวัดแบ่งตามภูมิภาคต่างๆ



[illegible]

```
[ ] 1 provincemix = pd.DataFrame.from_dict(data_province) # สร้างตาราง pandas จาก data_province
    2 provincemix # แสดงข้อมูลในตารางทั้งหมด
```

	จังหวัด	6 ภูมิภาค (ภูมิศาสตร์)	4 ภูมิภาค (การเมือง)	6 ภูมิภาค (อุดมศึกษา)	5 ภูมิภาค (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

ตารางที่ได้จากการสร้างเมื่อสักครู่นี้



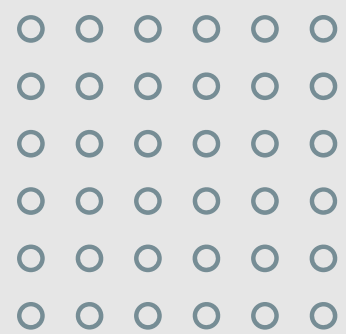
```
1 pv_province = provincemix[['จังหวัด', '6 ภูมิภาค (ภูมิศาสตร์)']]
2 pv_province
```

	จังหวัด	6 ภูมิภาค (ภูมิศาสตร์)
0	อำนาจเจริญ	ตะวันออกเฉียงเหนือ
1	บึงกาฬ	ตะวันออกเฉียงเหนือ
2	บุรีรัมย์	ตะวันออกเฉียงเหนือ
3	ชัยภูมิ	ตะวันออกเฉียงเหนือ
4	กาฬสินธุ์	ตะวันออกเฉียงเหนือ
...
72	พังงา	ใต้
73	ภูเก็ต	ใต้
74	ระนอง	ใต้
75	สตูล	ใต้
76	ตรัง	ใต้

77 rows × 2 columns

เลือกภูมิภาคที่ต้องการนำไปสร้าง DashBoard
 ในที่นี้เลือก 6 ภูมิภาคแบ่งตามภูมิศาสตร์





โหลดข้อมูลจำนวนโดสวัคซีน แบ่งตามประเภทวัคซีน



ข้อมูลการฉีดวัคซีนแบ่งตามประเภทวัคซีน รายจังหวัด

```
[ ] 1 data_file_path = os.path.join(path, 'covid_typevac.csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ  
2 datavac = pd.read_csv(data_file_path)  
3 datavac
```

Unnamed: 0	total_doses	province	AstraZeneca	Sinovac	Sinopharm	Pfizer	Johnson & Johnson	date
0	0	7883107	กรุงเทพมหานคร	4802964	2076263	903434	98177	1744 2021-08-20T12:12
1	1	962259	สมุทรปราการ	517274	281818	142539	20609	0 2021-08-20T12:12
2	2	837661	นนทบุรี	377568	403579	49283	7231	0 2021-08-20T12:12
3	3	729276	ปทุมธานี	307075	194572	209470	18159	0 2021-08-20T12:12
4	4	251774	พระนครศรีอยุธยา	89642	104000	55663	2469	0 2021-08-20T12:12
...
72	72	203270	ตรัง	41972	92912	65438	2948	0 2021-08-20T12:12
73	73	117203	พัทลุง	35107	79258	0	2838	0 2021-08-20T12:12
74	74	220018	ปัตตานี	118115	93386	1167	7350	0 2021-08-20T12:12
75	75	212161	ยะลา	104093	84271	20868	2929	0 2021-08-20T12:12
76	76	215845	นราธิวาส	95468	106417	1324	12636	0 2021-08-20T12:12

77 rows × 9 columns

ตัดตารางให้เหลือเฉพาะตารางที่ต้องการ โดยใช้คำสั่ง `.iloc[]`

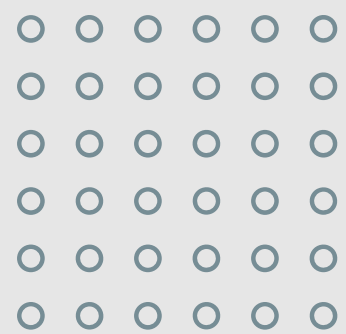
```
[ ] 1 datavac = datavac.iloc[:,1:8] # ตัดตารางที่ต้องการใช้  
2 datavac
```

	total_doses	province	AstraZeneca	Sinovac	Sinopharm	Pfizer	Johnson & Johnson
0	7883107	กรุงเทพมหานคร	4802964	2076263	903434	98177	1744
1	962259	สมุทรปราการ	517274	281818	142539	20609	0
2	837661	นนทบุรี	377568	403579	49283	7231	0
3	729276	ปทุมธานี	307075	194572	209470	18159	0
4	251774	พระนครศรีอยุธยา	89642	104000	55663	2469	0
...
72	203270	ตรัง	41972	92912	65438	2948	0
73	117203	พัทลุง	35107	79258	0	2838	0
74	220018	ปัตตานี	118115	93386	1167	7350	0
75	212161	ยะลา	104093	84271	20868	2929	0
76	215845	นราธิวาส	95468	106417	1324	12636	0

77 rows x 7 columns

เช็คค่า missing

```
[ ] 1 datavac.isnull().any()  
  
total_doses      False  
province         False  
AstraZeneca      False  
Sinovac          False  
Sinopharm        False  
Pfizer           False  
Johnson & Johnson False  
dtype: bool
```



ขั้นตอนต่อไปจะเป็นการรวมตารางทั้ง 3



ตาราง dt_cvpv คือรวมตารางโควิด กับ ภูมิภาค ใช้ จังหวัด เป็นคีย์เชื่อม

คำสั่ง `.merge()`

```
[ ] 1 dt_cvpv = cv_covid.merge(pv_province,how='left',left_on='province_of_onset',right_on='จังหวัด')
2 dt_cvpv
```

	province_of_onset	infected	female	male	other	จังหวัด	6 ภูมิภาค (ภูมิศาสตร์)
0	กระบี่	17667	9430	7273	964	กระบี่	ใต้
1	กรุงเทพมหานคร	572034	266381	250966	54687	กรุงเทพมหานคร	กลาง
2	กาญจนบุรี	24534	13294	10756	484	กาญจนบุรี	ตะวันตก
3	กาฬสินธุ์	18596	9807	7815	974	กาฬสินธุ์	ตะวันออกเฉียงเหนือ
4	กำแพงเพชร	13106	6985	5881	240	กำแพงเพชร	กลาง
...
73	เพชรบูรณ์	18497	8526	7924	2047	เพชรบูรณ์	กลาง
74	เลย	9990	5520	4299	171	เลย	ตะวันออกเฉียงเหนือ
75	แพร่	5668	2926	2498	244	แพร่	เหนือ
76	แม่ฮ่องสอน	4145	2217	1813	115	แม่ฮ่องสอน	เหนือ
77	ไม่ระบุ	400450	162814	171943	65693	NaN	NaN

78 rows × 7 columns

นำตาราง dt_cvpc มารวมกับตารางวัคซีน ใช้ province เป็นคีย์เชื่อม

คำสั่ง .merge()

```
[ ] 1 dt_DashBoard = dt_cvpc.merge(datavac,how='left',left_on='province_of_onset',right_on='province')
2 dt_DashBoard
```

	province_of_onset	infected	female	male	other	จังหวัด	6 ภูมิภาค (ภูมิศาสตร์)	total_doses	province	AstraZeneca	Sinovac	Sinopharm	Pfizer	Johnson & Johnson
0	กระบี่	17667	9430	7273	964	กระบี่	ใต้	182330.0	กระบี่	72829.0	104287.0	3131.0	2083.0	0.0
1	กรุงเทพมหานคร	572034	266381	250966	54687	กรุงเทพมหานคร	กลาง	7883107.0	กรุงเทพมหานคร	4802964.0	2076263.0	903434.0	98177.0	1744.0
2	กาญจนบุรี	24534	13294	10756	484	กาญจนบุรี	ตะวันตก	142407.0	กาญจนบุรี	40496.0	72925.0	25484.0	3502.0	0.0
3	กาฬสินธุ์	18596	9807	7815	974	กาฬสินธุ์	ตะวันออกเฉียงเหนือ	131875.0	กาฬสินธุ์	38048.0	87741.0	2185.0	3901.0	0.0
4	กำแพงเพชร	13106	6985	5881	240	กำแพงเพชร	กลาง	108806.0	กำแพงเพชร	28989.0	75773.0	668.0	3376.0	0.0
...
73	เพชรบูรณ์	18497	8526	7924	2047	เพชรบูรณ์	กลาง	214327.0	เพชรบูรณ์	70067.0	137029.0	4169.0	3062.0	0.0
74	เลย	9990	5520	4299	171	เลย	ตะวันออกเฉียงเหนือ	141055.0	เลย	44257.0	92207.0	1948.0	2643.0	0.0
75	แพร่	5668	2926	2498	244	แพร่	เหนือ	79724.0	แพร่	24923.0	49739.0	2675.0	2387.0	0.0
76	แม่ฮ่องสอน	4145	2217	1813	115	แม่ฮ่องสอน	เหนือ	41539.0	แม่ฮ่องสอน	12377.0	27508.0	245.0	1409.0	0.0
77	ไม่ระบุ	400450	162814	171943	65693	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

78 rows × 14 columns

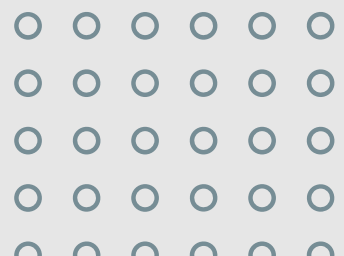
ตัดแถวที่ 78 ออก (จังหวัดไม่ระบุ) ใช้คำสั่ง .iloc[]



```
[ ] 1 # ตัดแถว ไม่ระบุ ออก
2 dt_DashBoard = dt_DashBoard.iloc[:77,:]
3 dt_DashBoard
```

	province_of_onset	infected	female	male	other	จังหวัด	6 ภูมิภาค (ภูมิศาสตร์)	total_doses	province	AstraZeneca	Sinovac	Sinopharm	Pfizer	Johnson & Johnson
0	กระบี่	17667	9430	7273	964	กระบี่	ใต้	182330.0	กระบี่	72829.0	104287.0	3131.0	2083.0	0.0
1	กรุงเทพมหานคร	572034	266381	250966	54687	กรุงเทพมหานคร	กลาง	7883107.0	กรุงเทพมหานคร	4802964.0	2076263.0	903434.0	98177.0	1744.0
2	กาญจนบุรี	24534	13294	10756	484	กาญจนบุรี	ตะวันตก	142407.0	กาญจนบุรี	40496.0	72925.0	25484.0	3502.0	0.0
3	กาฬสินธุ์	18596	9807	7815	974	กาฬสินธุ์	ตะวันออกเฉียงเหนือ	131875.0	กาฬสินธุ์	38048.0	87741.0	2185.0	3901.0	0.0
4	กำแพงเพชร	13106	6985	5881	240	กำแพงเพชร	กลาง	108806.0	กำแพงเพชร	28989.0	75773.0	668.0	3376.0	0.0
...
72	เพชรบุรี	30298	15776	13796	726	เพชรบุรี	ตะวันตก	189120.0	เพชรบุรี	77906.0	97959.0	10431.0	2824.0	0.0
73	เพชรบูรณ์	18497	8526	7924	2047	เพชรบูรณ์	กลาง	214327.0	เพชรบูรณ์	70067.0	137029.0	4169.0	3062.0	0.0
74	เลย	9990	5520	4299	171	เลย	ตะวันออกเฉียงเหนือ	141055.0	เลย	44257.0	92207.0	1948.0	2643.0	0.0
75	แพร่	5668	2926	2498	244	แพร่	เหนือ	79724.0	แพร่	24923.0	49739.0	2675.0	2387.0	0.0
76	แม่ฮ่องสอน	4145	2217	1813	115	แม่ฮ่องสอน	เหนือ	41539.0	แม่ฮ่องสอน	12377.0	27508.0	245.0	1409.0	0.0

77 rows × 14 columns



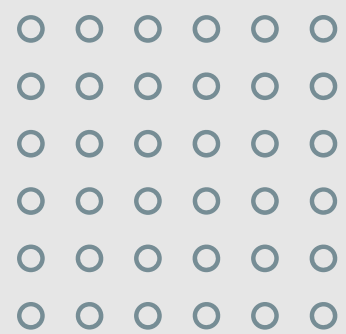
เช็คค่า missing ของตารางที่รวมกันเรียบร้อยแล้ว ก่อนนำไปสร้าง DashBoard

```
[ ] 1 dt_DashBoard.isnull().any()

province_of_onset      False
infected               False
female                False
male                  False
other                  False
จังหวัด                False
6 ภูมิภาค (ภูมิศาสตร์) False
total_doses            False
province               False
AstraZeneca            False
Sinovac                False
Sinopharm              False
Pfizer                 False
Johnson & Johnson     False
dtype: bool
```

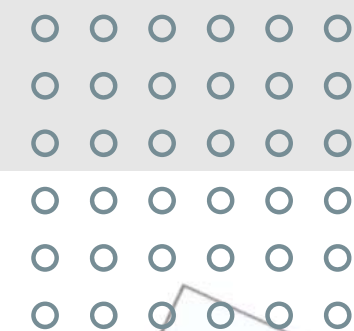
save ข้อมูลเป็นไฟล์ .csv เพื่อที่จะนำข้อมูลเข้า google data studio

```
[ ] 1 dt_DashBoard.to_csv('dt_DashBoard.csv', encoding='utf-8')
```



ตัวอย่าง DASHBOARD ของกลุ่มนั้น

ดูต่อได้ที่ Github



รายงานข้อมูลผู้ติดเชื้อ Covid-19 และข้อมูลจำนวนโดสวัคซีนแต่ละประเภทที่ถูกฉีดไป จำแนกตามรายจังหวัด

จำนวนผู้ติดเชื้อ
68.4K

จำนวนโดสวัคซีน
353.3K



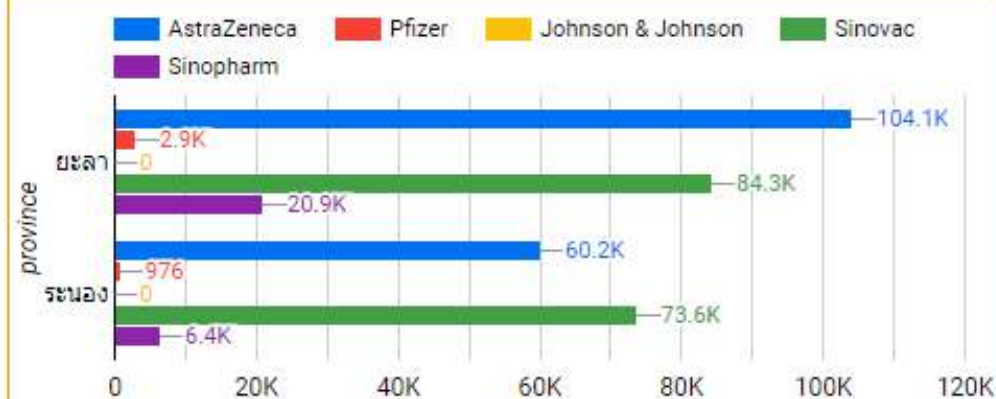
จังหวัด: ยะลา... (2) ▾

ภูมิภาค: ไต (1) ▾

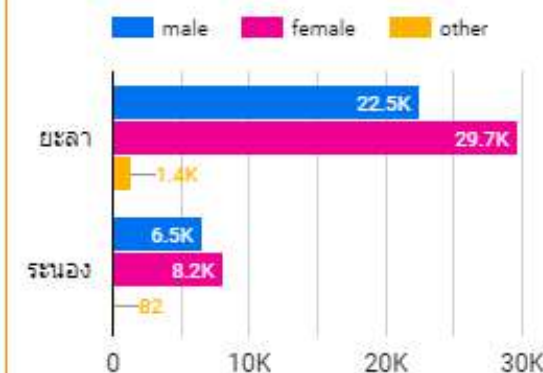
ร้อยละของผู้ติดเชื้อแบ่งตามภูมิภาค



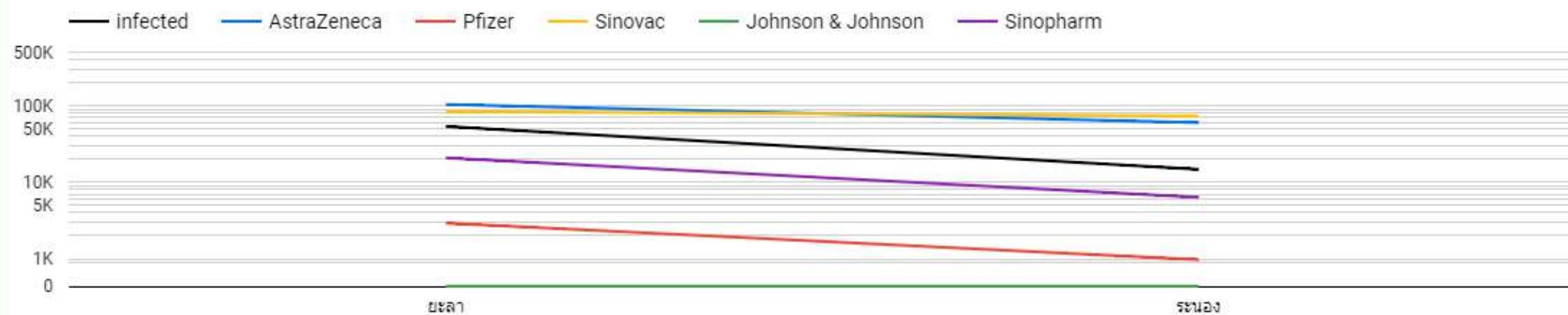
จำนวนโดสวัคซีนที่ถูกฉีดไป (รายจังหวัด)



จำนวนผู้ติดเชื้อ (แบ่งตามเพศ)



กราฟเส้นเปรียบเทียบระหว่างจำนวนวัคซีนแต่ละประเภท(โดส)กับจำนวนผู้ติดเชื้อ (รายจังหวัด)



เปรียบเทียบ 2 จังหวัด ได้แก่ ยะลา กับ ระนอง
แผนภูมิวงกลม: 2 จังหวัดนี้อยู่ภาคใต้
กราฟแท่ง(ที่1): วัคซีนแต่ละประเภทที่ได้รับ

- ยะลาได้รับ AstraZeneca มากที่สุด รองลงมา Sinovac, Sinopharm, Pfizer และ Johnson & Johnson ตามลำดับ
- ระนองได้รับ Sinovac มากที่สุด รองลงมา AstraZeneca, Sinopharm, Pfizer และ Johnson & Johnson ตามลำดับ

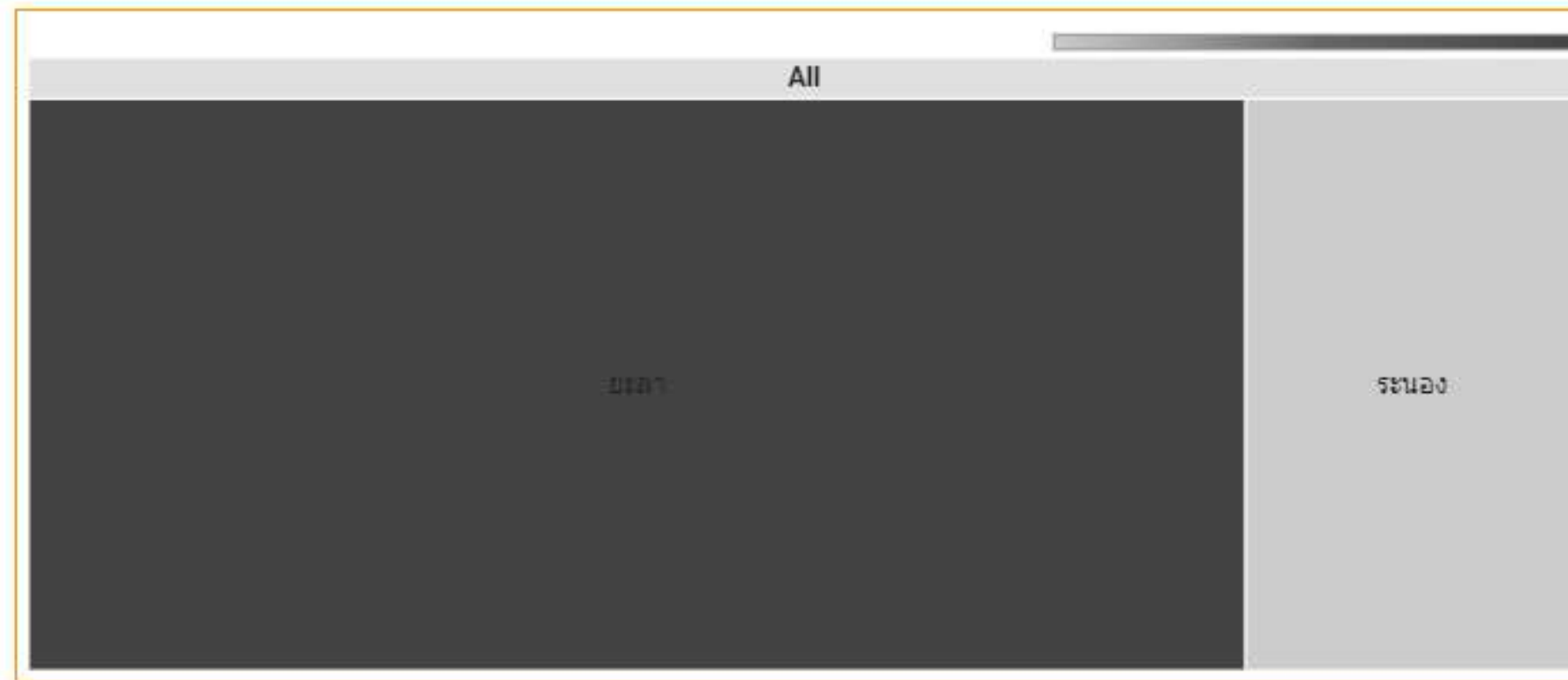
กราฟแท่ง(ที่2): เพศของผู้ติดเชื้อโควิด-19

- ยะลา มีผู้ติดเชื้อโควิดเป็นเพศหญิงมากที่สุด (ประมาณ 29,700 คน)
- ระนอง มีผู้ติดเชื้อโควิดเป็นเพศหญิงมากที่สุด (ประมาณ 8,200 คน)

กราฟเส้น: เปรียบเทียบจำนวนผู้ติดเชื้อกับวัคซีน

ตารางแสดงจำนวนโดสวัคซีนที่ถูกฉีดไปเปรียบเทียบกับจำนวนผู้ติดเชื้อ

จังหวัด	AstraZ...	Pfizer	Johnson & J...	Sinovac	Sinoph...	total_doses	infected ▾
1. ยะลา	104,093	2,929	0	84,271	20,868	212,161	53,653
2. หนอง	60,181	976	0	73,584	6,372	141,113	14,783



Tree map แสดงจำนวนโดสวัคซีนที่ถูกฉีดไปเปรียบเทียบกับจำนวนผู้ติดเชื้อ



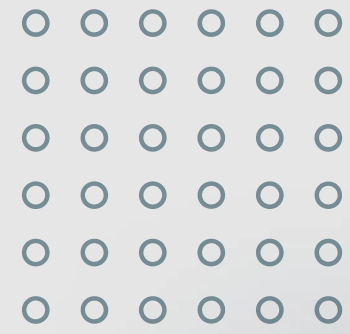
แผนที่แสดงจำนวนผู้ติดเชื้อ ถ้าแสดงเป็นสีแดง แสดงว่ามีผู้ติดเชื้อมาก

เนื่องจากข้อมูลวัคซีนที่ได้มา วัคซีนแต่ละประเภทถูกแบ่งเป็นจำนวนโดสวัคซีนตามรายจังหวัดไว้เรียบร้อยแล้ว
จึงไม่สามารถเปรียบเทียบได้ว่า 1 คนฉีดวัคซีนกี่โดส อะไรบ้าง กลุ่มนั้นทำได้เพียงแค่แสดงจำนวนโดสเปรียบเทียบกับ
กับจำนวนคนติดโควิด-19

```
[ ] 1 data_file_path = os.path.join(path, 'covid_typevac.csv') # เอาที่อยู่ของไฟล์เชื่อมกับที่อยู่ของไฟล์เรา โดยกำหนดชื่อไฟล์ที่เราต้องการ
2 datavac = pd.read_csv(data_file_path)
3 datavac
```

Unnamed: 0	total_doses	province	AstraZeneca	Sinovac	Sinopharm	Pfizer	Johnson & Johnson	date
0	0	7883107	กรุงเทพมหานคร	4802964	2076263	903434	98177	1744 2021-08-20T12:12
1	1	962259	สมุทรปราการ	517274	281818	142539	20609	0 2021-08-20T12:12
2	2	837661	นนทบุรี	377568	403579	49283	7231	0 2021-08-20T12:12
3	3	729276	ปทุมธานี	307075	194572	209470	18159	0 2021-08-20T12:12
4	4	251774	พระนครศรีอยุธยา	89642	104000	55663	2469	0 2021-08-20T12:12
...
72	72	203270	ตรัง	41972	92912	65438	2948	0 2021-08-20T12:12
73	73	117203	พัทลุง	35107	79258	0	2838	0 2021-08-20T12:12
74	74	220018	ปัตตานี	118115	93386	1167	7350	0 2021-08-20T12:12
75	75	212161	ยะลา	104093	84271	20868	2929	0 2021-08-20T12:12
76	76	215845	นราธิวาส	95468	106417	1324	12636	0 2021-08-20T12:12

77 rows × 9 columns



THANK YOU
FOR
WATCHING

