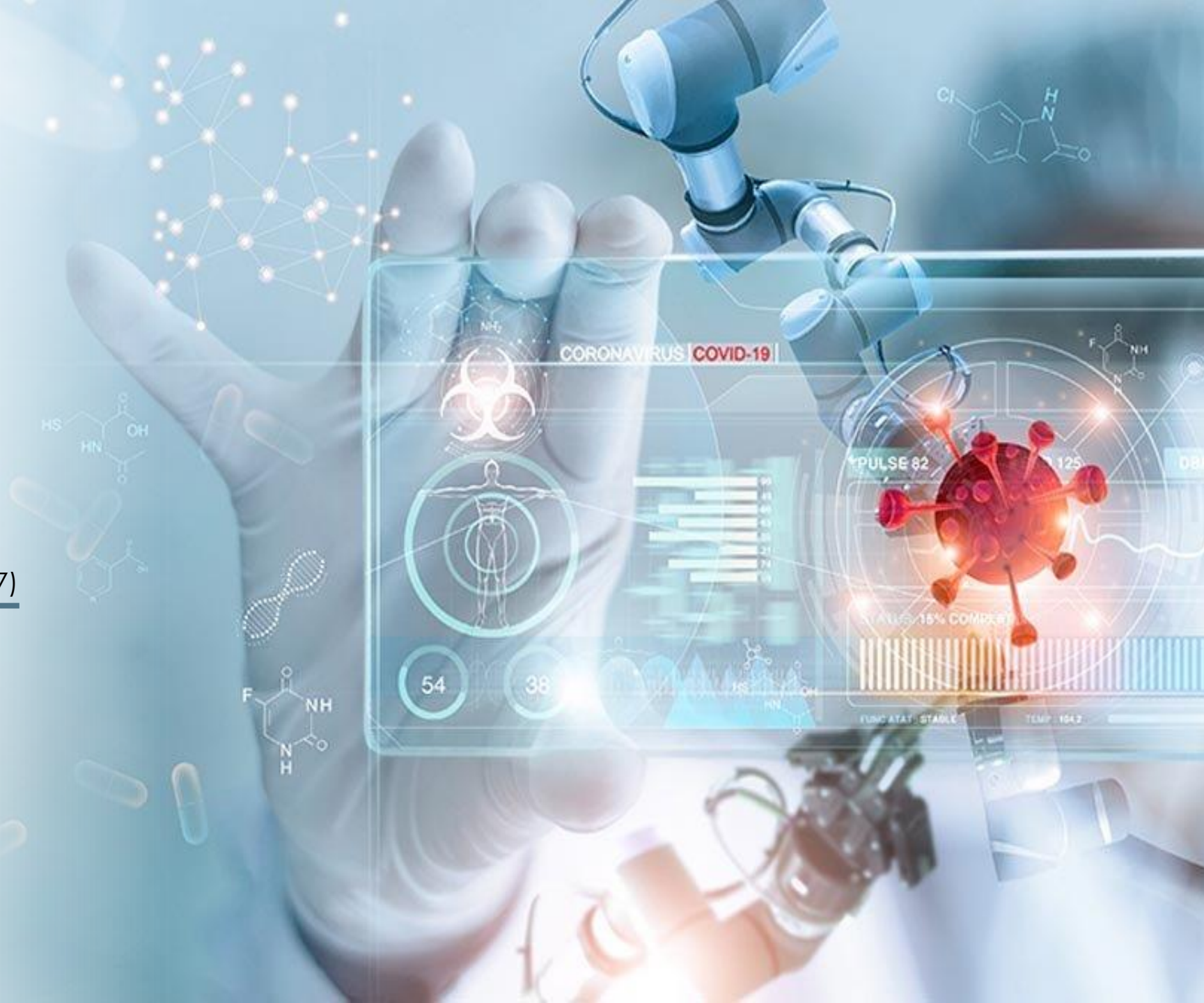


Covid Analysis

By Kantima Techaphonprasit (6410422027)



Data Source

Data Source



รายงาน COVID-19 ประจำวัน
ข้อมูลประจำประเทศไทย

 ดาวน์โหลด	รายงานจำนวนผู้ติดเชื้อ COVID-19 ประจำวันเริ่มตั้งแต่วันที่ 28 ...  819 downloads
 ดาวน์โหลด	รายงานจำนวนผู้ติดเชื้อ COVID-19 ประจำวันเริ่มตั้งแต่วันที่ 27 ...  4,587 downloads
 ดาวน์โหลด	รายงานจำนวนผู้ติดเชื้อ COVID-19 ประจำวันเริ่มตั้งแต่วันที่ 12 ...  6,397 downloads
 ดาวน์โหลด	รายงานจำนวนผู้ติดเชื้อ COVID-19 ประจำวัน  11,472 downloads
 ไปสู่ทรัพยากร	รายงาน COVID-19 แสดงค่าประจำวัน  1,892 downloads
 ไปสู่ทรัพยากร	รายงาน COVID-19 ข้อมูลสรุปตาม ช่วงเวลา (เริ่มตั้งแต่วันที่ 01/01/20)  1,727 downloads
 ดาวน์โหลด	covid-19-daily_data_dictionary  3,426 downloads

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



File in Google Drive







(press ctrl and click file photo to access file)

My Drive > Covid_Analysis ▾


Name ↓



 DB.TTF 



 Covid-cleansing_byKantima.ipynb 



 Covid-Analysis_byKantima.ipynb 

 covid_complete_cleansing.csv 

 c44_confirmed-cases-since-280265.csv 

 c33_confirmed-cases-since-271064.csv 

 c22_confirmed-cases-since-120864.csv 

 c11_confirmed-cases.csv 

https://drive.google.com/drive/folders/1A7n7vuNWr6zjVvdWMrCwOmP_4FOcX?usp=sharing

(Thai font in visualization)

1 Cleansing Data (A) & Export to csv(B)

2 Import csv(B) to analyze & visualize

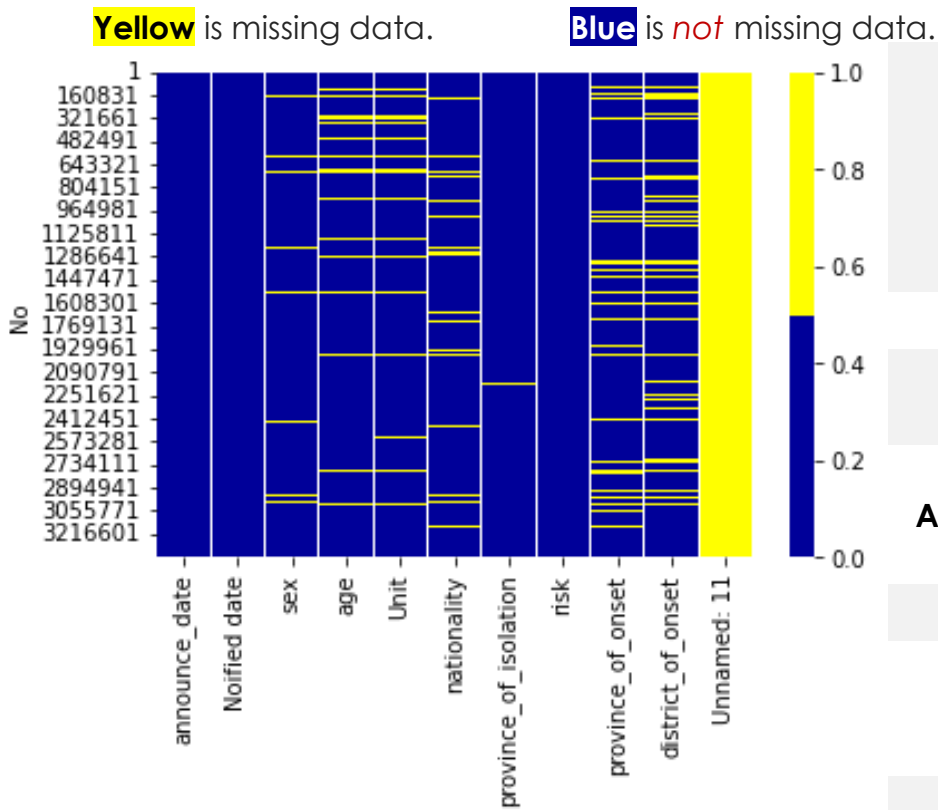
A

*ทั้ง 4 ไฟล์เป็นไฟล์
CSV ที่ถูกแปลงมาจาก
ไฟล์ excel ต้นฉบับ

Cleansing Data Logics

- Replace missing string data with “No data”
- Delete outing scope columns (district_of_onset, province_of_isolation)
- Add new columns (Age_bin, Diff_date, Diff_date_bin, Region)

Heatmap of Missing Data



```
cols = df.columns[:11]
colours = ['#000099', '#ffff00']
sns.heatmap(df[cols].isnull(), cmap=sns.color_palette(colours))
```

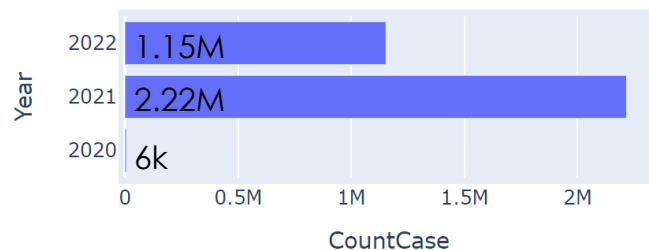
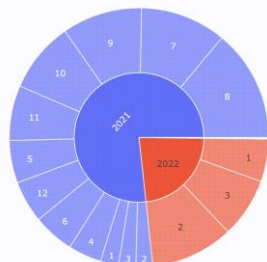
Column name	Original	After Cleansing	Cleansing Data Logic
No	unique	Set as index	
Age	(0, 44899]	(0, 120]	- Set age>120y & age<1 as null - Convert age in day unit to 1 month - Divide age in month unit by 12 to convert to year
Unit	Day, Month, Year	Year only	
Sex	Typo words	[Female, Male, No data]	
Announce_date	Type: string	Type: Datetime	
Notified_date	----- Same Announce_date -----		
Nationality	Many countries	- Keep original only Top 10 (Not include null) - Set the rest as 'Other'	- Top10 is 91% of total (3.06M cases)
Province	Typo words	Correct words	
Risk	Vary and Similar words	Grouping similar words to a group	

Overall Covid Analysis

Data Period Scope: 12 Jan 2020 - 21 Mar 2022 (2 years 3 months)

Total Case
3,377,410

No. of Case by Year, month

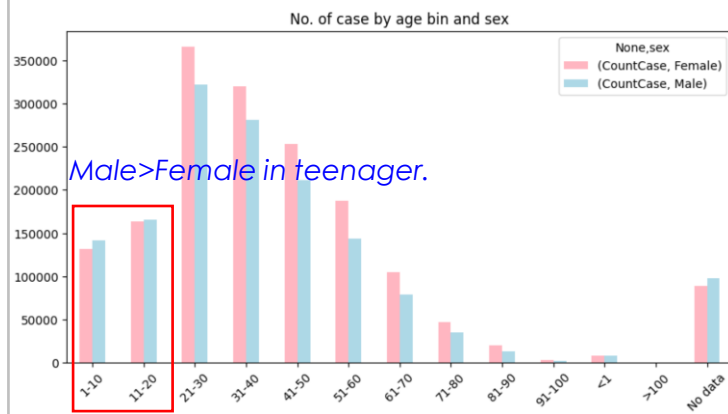
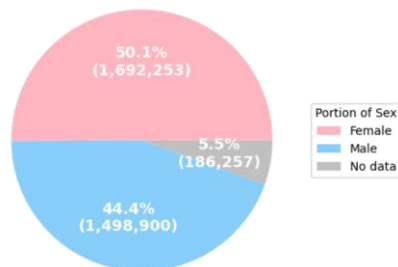
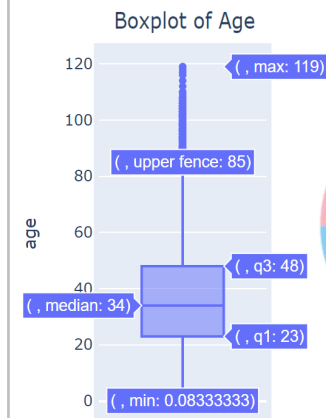


No. of New Cases over period



Age & Gender

Avg. Age = 35 years

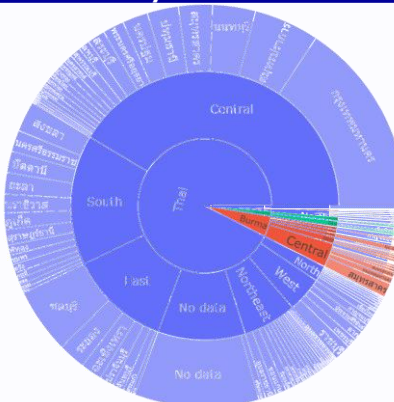


Nationality

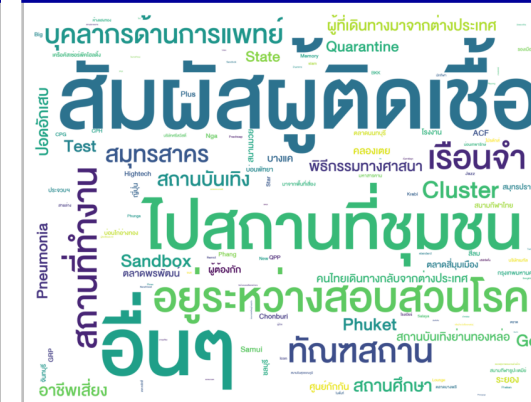
Top5

Nationality	CountCase	% of Total
Thai	2849113	84.357925
Burma	153154	4.534658
Cambodian	39995	1.184191
Other	15304	0.453128
Lao	9861	0.291969

84% is Thai

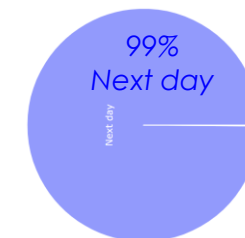


Risk



Difference Date

After case notification, case will be officially announced in next day.

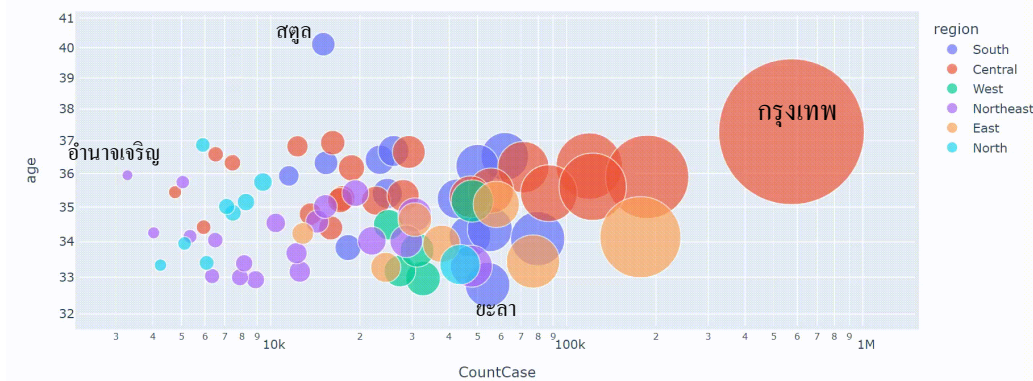
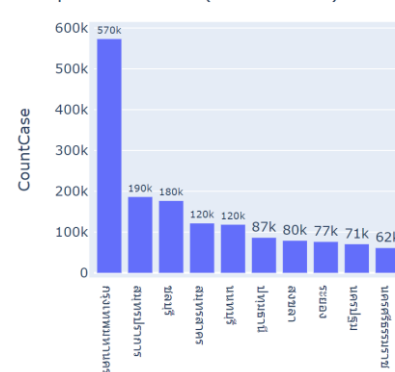


Area & Average Age

Top10 Province

37% is in BKK

Top 10 Provinces (No. of Cases)



Statistics Q&A

1) วันที่ประกาศยอดผู้ติดเชื้อสูงสุด คือวันที่เท่าไร? และมีจำนวนกี่คน

วันที่ 2022-03-18

27,071 เคส

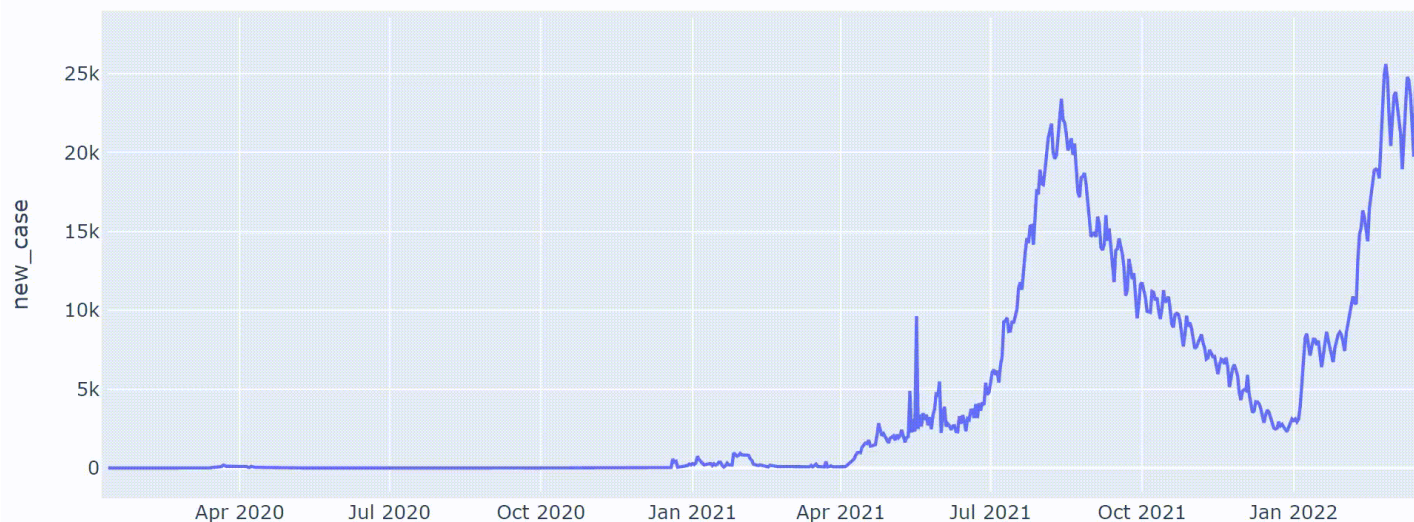
```
1 maxday = df[['announce_date', 'CountCase']].groupby('announce_date').sum()
2 print(f'วันที่ประกาศยอดผู้ติดเชื้อสูงสุด คือวันที่ 2022-03-18 จำนวน 27,071 คน')
3 maxday.sort_values(by = 'CountCase', ascending = False).head(1)
```

☞ วันที่ประกาศยอดผู้ติดเชื้อสูงสุด คือวันที่ 2022-03-18 จำนวน 27,071 คน

CountCase

announce_date
2022-03-18
27071

No. of New Case over period



```
import plotly.express as px
fig = px.line(timeseall, x='Announce_date', y=["CountCase"],
              title = 'Daily New Case')

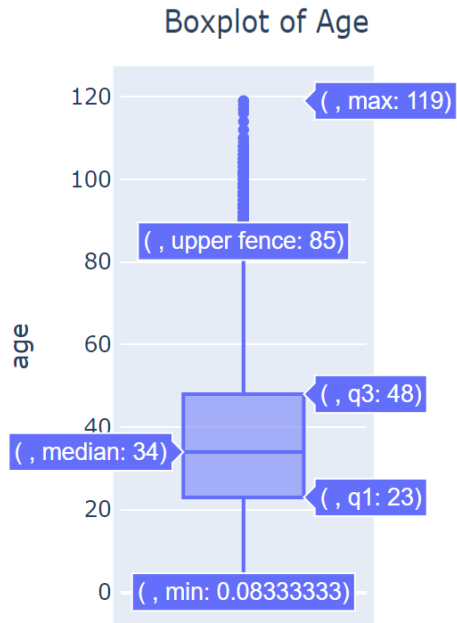
fig.update_layout(
    autosize=False,
    width=800,
    height=300,
    margin=dict(l=50,r=20,b=20,t=50,pad=4))
fig.show()
```



Statistics Q&A

2) อายุเฉลี่ยของผู้ที่ติดโควิดเป็นเท่าใด

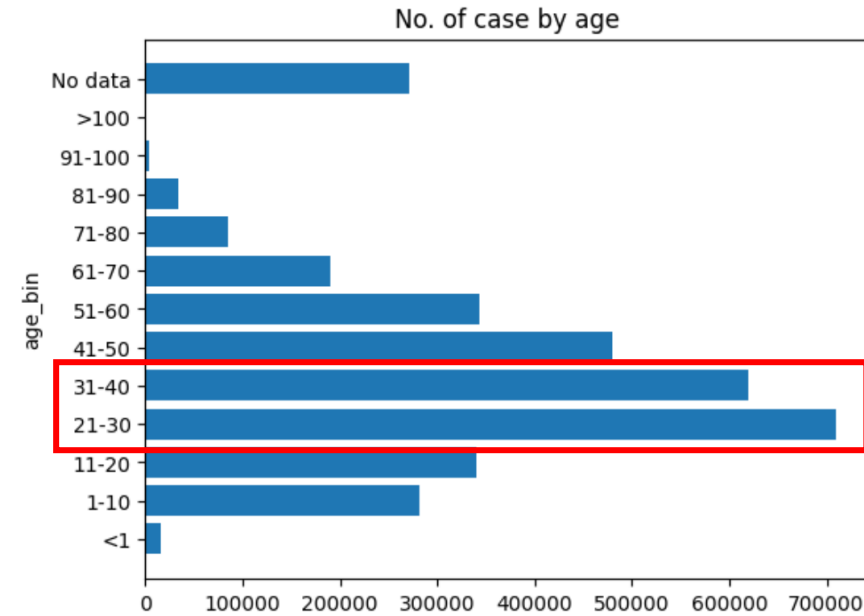
= 35 ปี



```
1 df.age.agg(['mean','min','max'])
```

```
mean    35.387867
min      0.083333
max    119.000000
Name: age, dtype: float64
```

```
import plotly.express as px
fig = px.box(df, y="age", title = 'Boxplot of Age')
fig.update_layout(
    autosize=False,
    width=250,
    height=500,
    margin=dict(l=50,r=50,b=100,t=100,pad=4))
fig.show()
```



ส่วนใหญ่เป็น
วัยทำงาน

```
import matplotlib.pyplot as plt
import numpy as np
```

```
plt.rcParams()
fig, ax = plt.subplots()
```

```
y_pos = ['<1','1-10','11-20','21-30','31-40','41-50','51-60','61-70','71-80','81-90','91-100','>100','No data']
performance = [15983, 282398, 340657, 709084, 619773, 479938, 343601, 190561, 85475, 33693, 4773, 165, 271309]
```

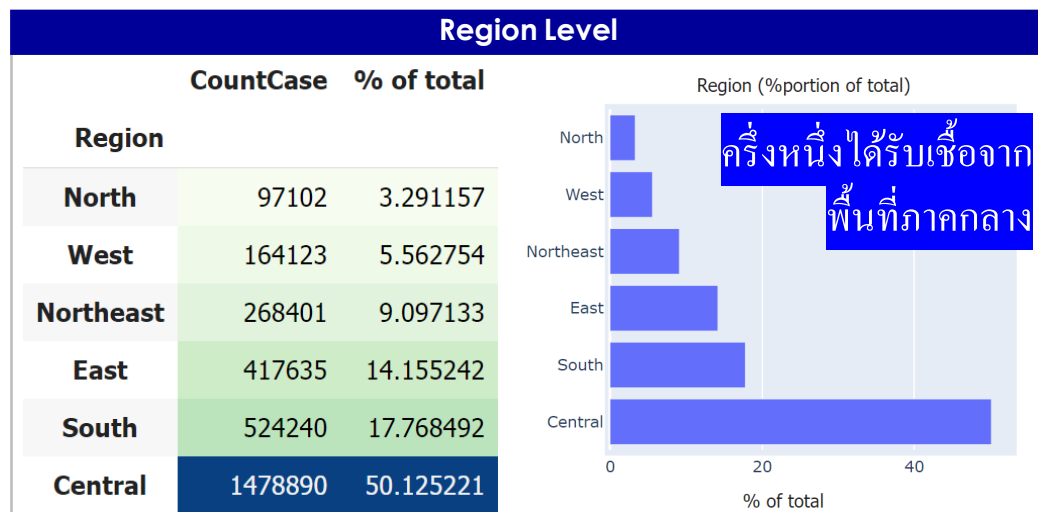
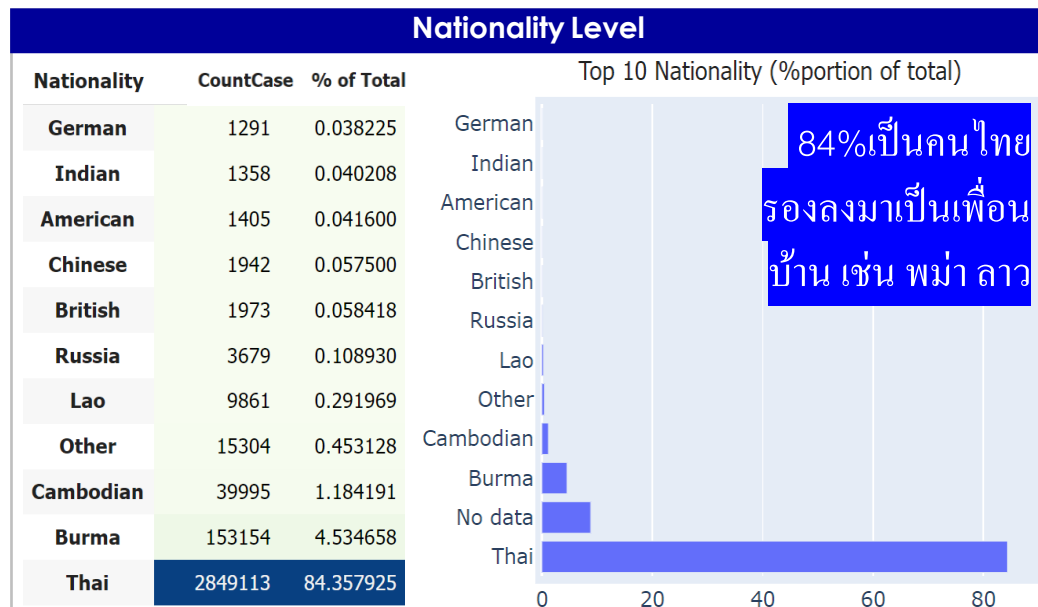
```
ax.barh(y_pos, performance, align='center')
ax.set_ylabel('age_bin')
ax.set_title('No. of case by age')
```

```
plt.show()
```

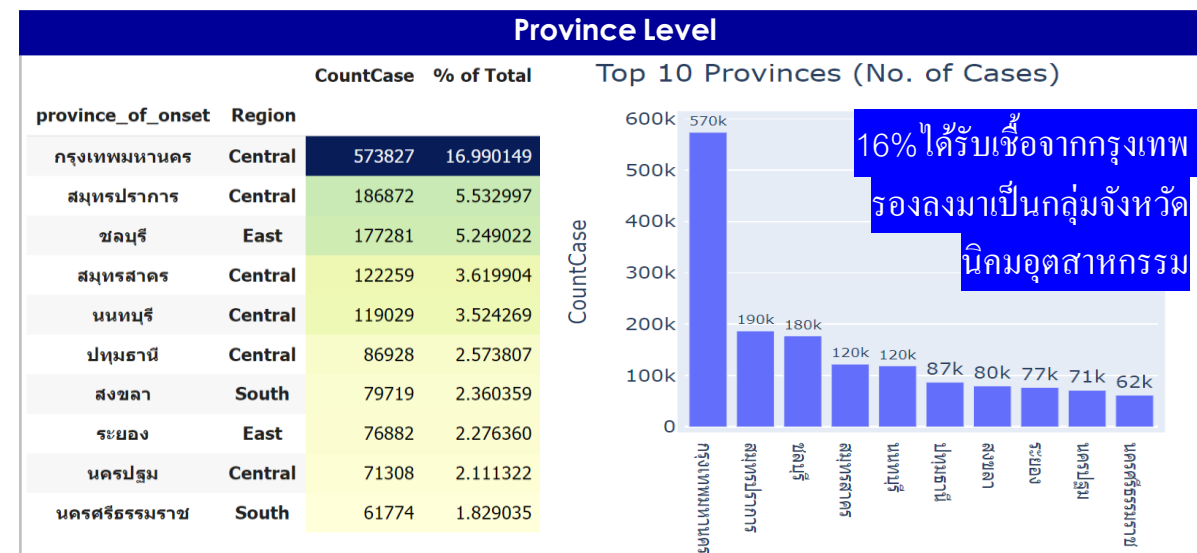
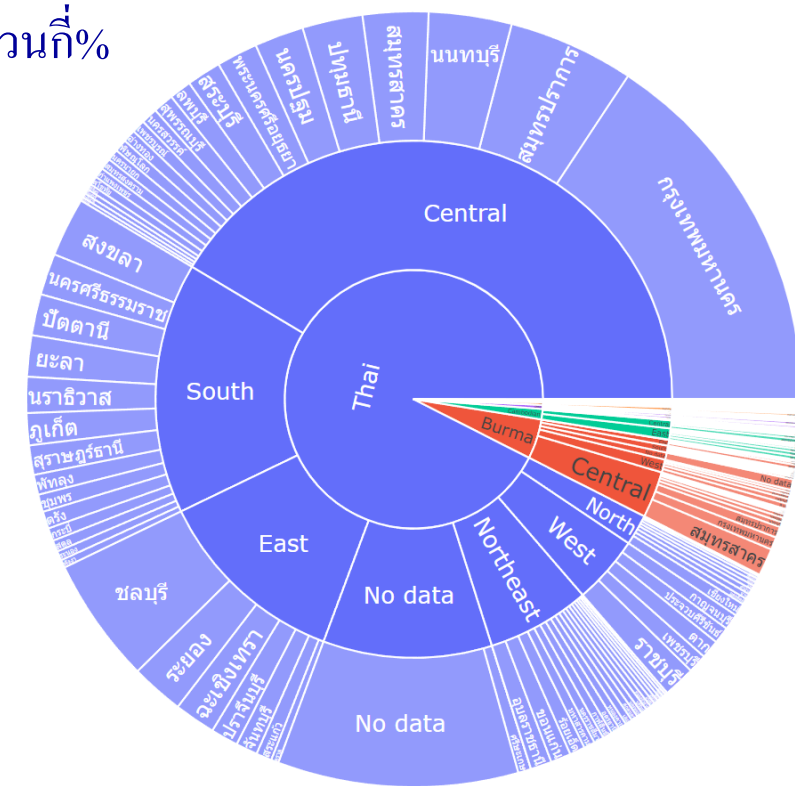


Statistics Q&A

- 3) ผู้ป่วยโควิดมีสัญชาติใดมากที่สุด และคิดเป็นสัดส่วนกี่%
- 4) ผู้ป่วยโควิดตรวจพบเชื้อจากพื้นที่ไหนมากที่สุด และคิดเป็นสัดส่วนกี่%

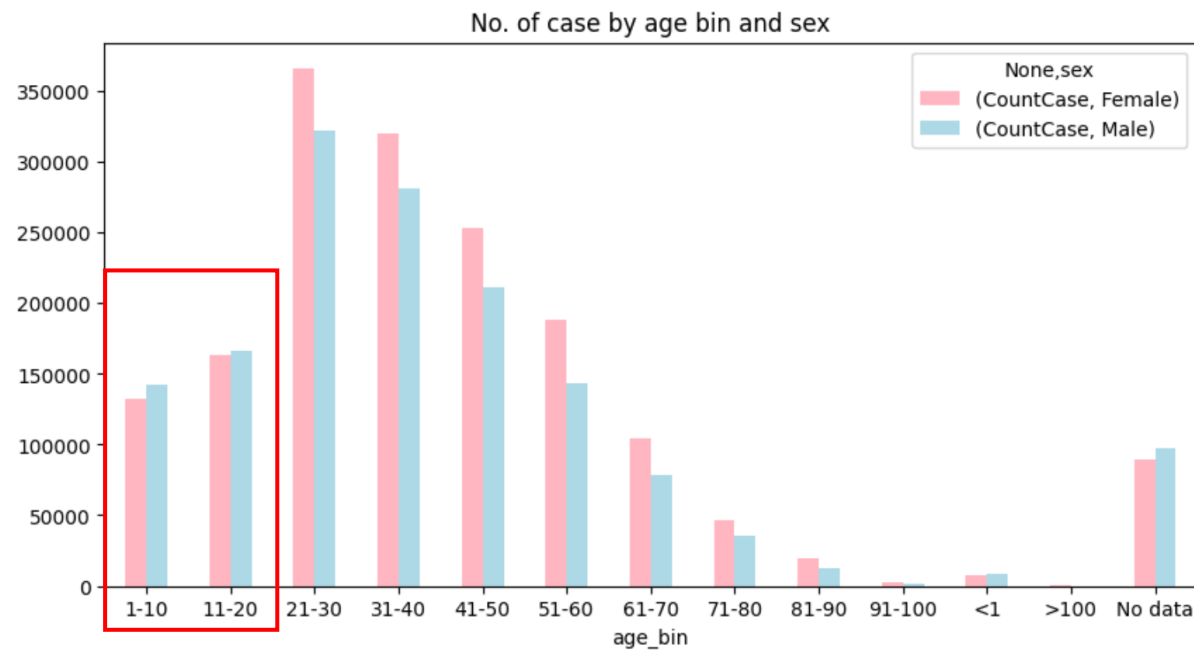
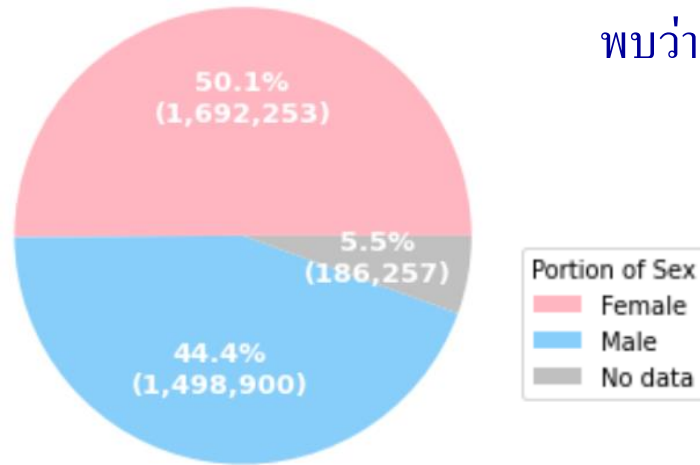


และคิดเป็นสัดส่วนกี่%



Insight Q&A

- ภาพรวมแล้ว จำนวนผู้ป่วยหญิงมากกว่าผู้ชาย แต่เมื่อนำอายุมา cross กันพบว่าในวัยเด็ก หรือวัยรุ่น (≤ 20 ปี) จะพบผู้ป่วยที่เป็นผู้ชายมากกว่า



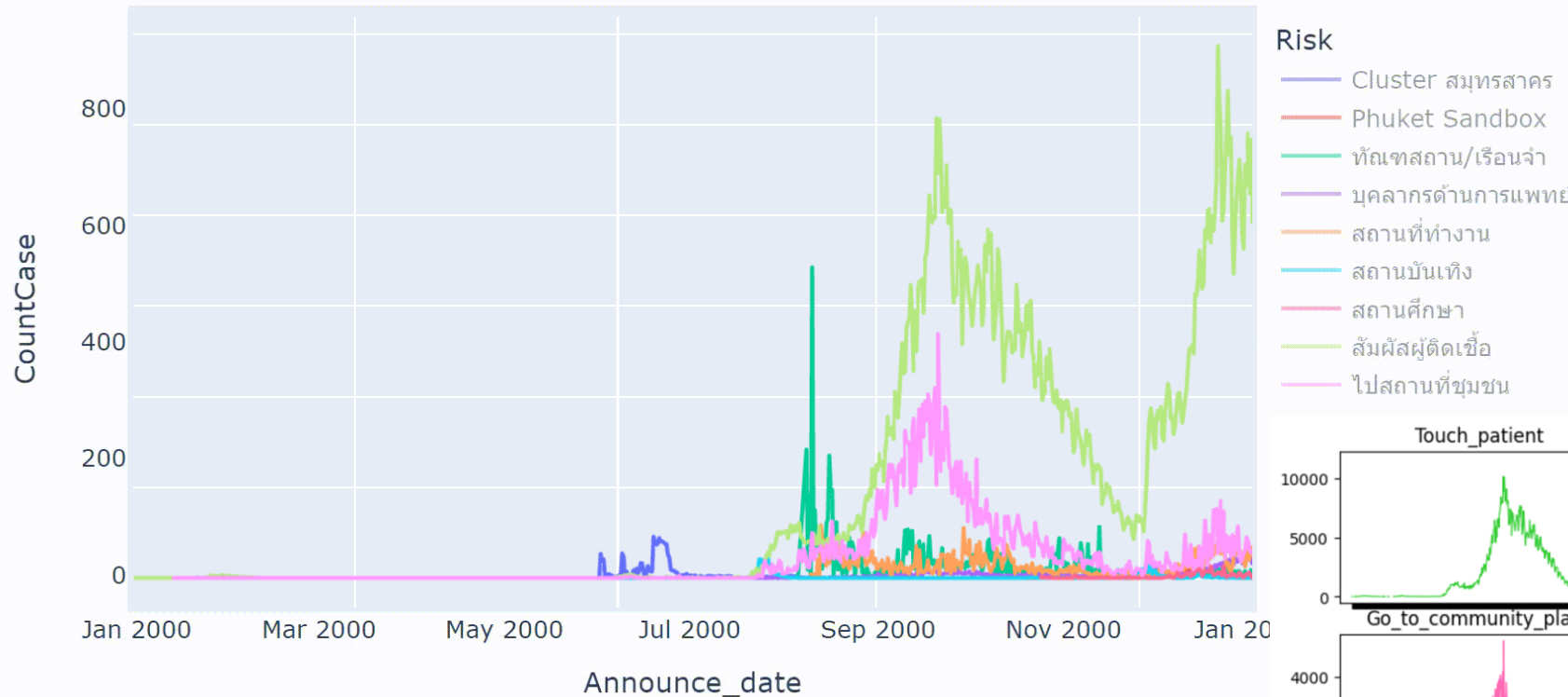
CountCase			CountCase		
age_bin	sex		age_bin	sex	
1-10	Female	131800	61-70	Female	104448
	Male	142026		Male	78341
11-20	Female	163652	71-80	Female	46516
	Male	165762		Male	35087
21-30	Female	365877	81-90	Female	19476
	Male	322044		Male	12654
31-40	Female	320005	91-100	Female	3035
	Male	280977		Male	1544
41-50	Female	252981	<1	Female	7599
	Male	211167		Male	8189
51-60	Female	187804	>100	Female	117
	Male	143651		Male	40

	Count	Case	% of Total
risk			
สัมผัสผู้ติดเชื้อ	1419	789	42.03
ไปสถานที่ชุมชน	335	394	9.93
สถานที่ทำงาน	93	466	2.76
ทัศนสถาน/เรือนจำ	93	388	2.76
บุคลากรด้านการแพทย์	30	214	0.89
Cluster สมุทรสาคร	17	900	0.52
สถานบันเทิง	8	953	0.26
Phuket Sandbox	6	474	0.19
Test and Go	6	021	0.17
สถานศึกษา	5	893	0.17

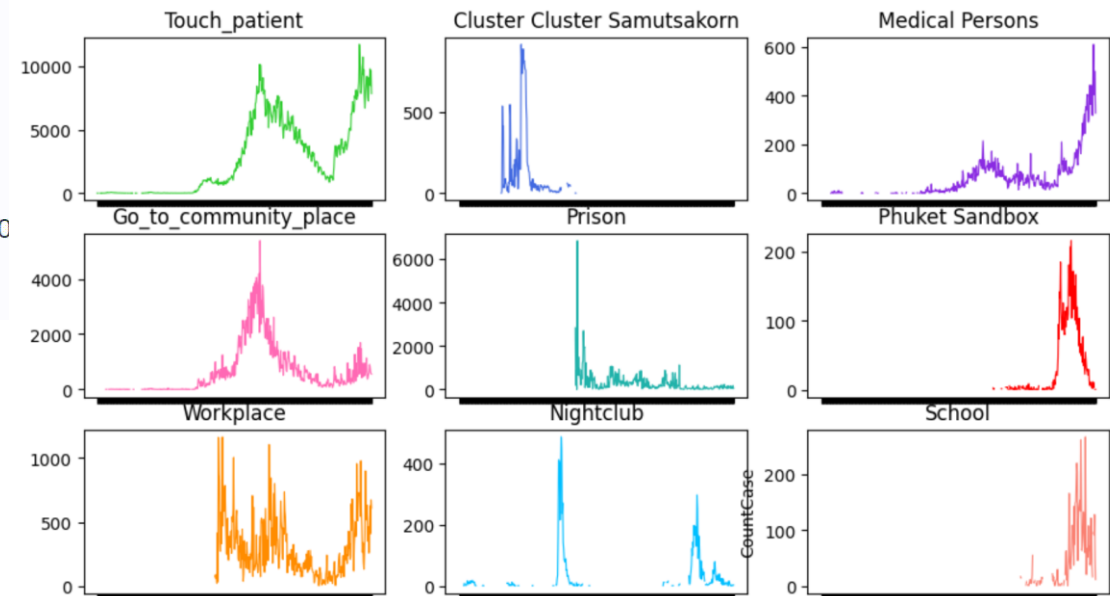
	Count	Case	% of Total
risk			
สัมผัสผู้ติดเชื้อ	1419	789	42.037804
ไปสถานที่ชุมชน	3353	394	9.930509
สถานที่ทำงาน	934	466	2.767387
ทัศนสถาน/เรือนจำ	933	388	2.765077
บุคลากรด้านการแพทย์	302	14	0.894591
Cluster สมุทรสาคร	179	900	0.529992
สถาบันเทিং	89	53	0.265085
Phuket Sandbox	64	74	0.191685
Test and Go	60	21	0.178273
สถานศึกษา	58	93	0.174483

Insight Q&A

No. of case by Top 9 Risk type



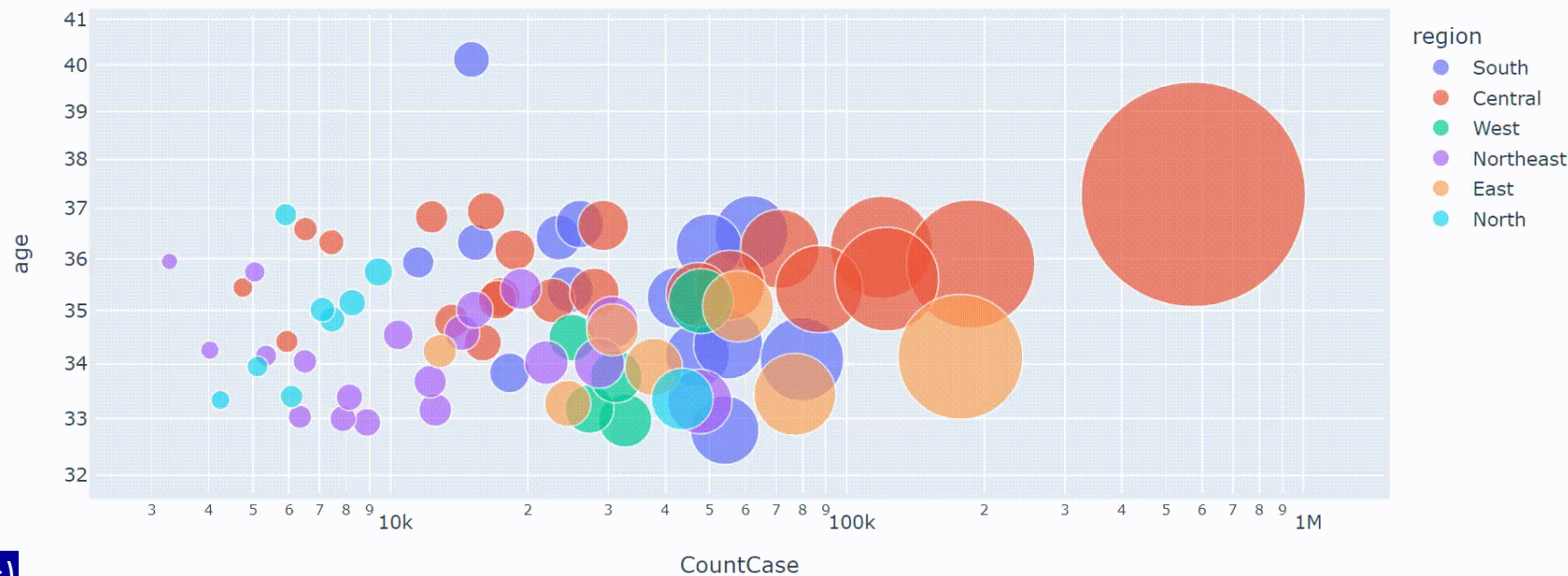
Top 9 Risk types over time



Insight Q&A



No. of case VS Average Age by province



Top10 Provinces (No. of cases)

province_of_onset	Region	CountCase	age mean
กรุงเทพมหานคร	Central	573827	37.283593
สมุทรปราการ	Central	186872	35.899374
ชลบุรี	East	177281	34.127394
สมุทรสาคร	Central	122259	35.600319
นนทบุรี	Central	119029	36.223747
ปทุมธานี	Central	86928	35.410048
สงขลา	South	79719	34.085571
ระยอง	East	76882	33.438874
นครปฐม	Central	71308	36.192064
นครศรีธรรมราช	South	61774	36.519186

Max-Min: No. of cases by Province

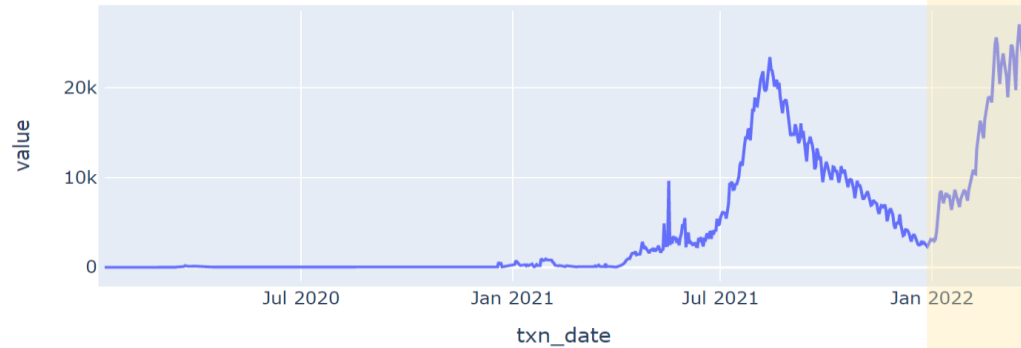
- กรุงเทพฯ: มีจำนวนผู้ติดเชื้อสูงสุด 573,827 คน
- อำนาจเจริญ: มีจำนวนผู้ติดเชื้อน้อยที่สุด 3,282 คน

Max-Min: Average Age by Province

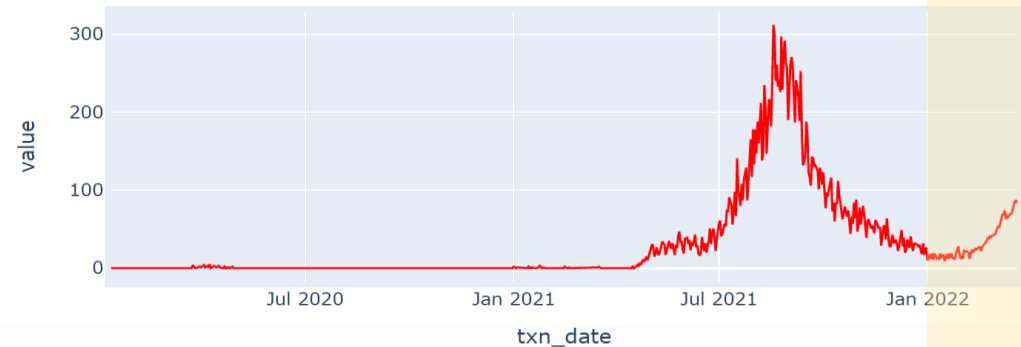
- สตูล: มีอายุเฉลี่ยของผู้ป่วยสูงที่สุด ที่ 40 ปี
- ยะลา: มีอายุเฉลี่ยของผู้ป่วยน้อยที่สุด ที่ 32 ปี

Insight Q&A

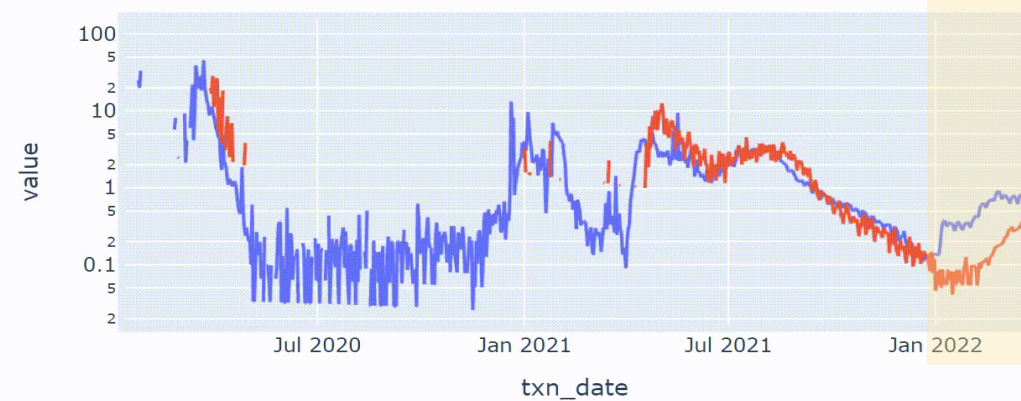
Daily New Case



Daily New Death



Growth rate

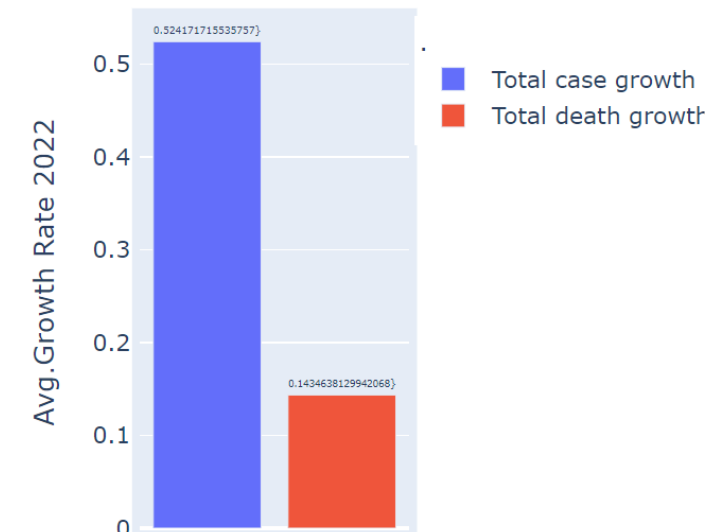


- ในปี 2022 ทั้งอัตราการเติบโตของจำนวนผู้ติดเชื้อ (New Case Growth rate) และอัตราการเติบโตของจำนวนผู้ตาย (Death Growth rate) เพิ่มสูงขึ้น แต่อัตราการเติบโตของจำนวนผู้ตาย เพิ่มสูงขึ้นในอัตราที่ต่ำกว่า

variable

— Total case growth rate
— Total death growth rate

0.524171715535757 0.1434638129942068



Summary

No. of case

- ทั้งหมด 3.37M เคส (2020-01-12 ถึง 2022-03-21)

Age & Gender

- Avg. age = 35 ปี Median = 34 ปี
- Min-Max = 1 เดือน – 199 ปี
- ภาพรวมผู้หญิงมากกว่าผู้ชายในทุกระดับอายุ
- ยกเว้น เด็กหรือวัยรุ่น (<=20 ปี) จะมีผู้ป่วยชายมากกว่าหญิง

Difference Date

- 99.9% เมื่อได้รับแจ้งเคสการติดเชื้อแล้ว
เจ้าหน้าที่จะประกาศจำนวนผู้ติดเชื้อในวันถัดไป

Nationality

- 84% เป็นคนไทย
- รองลงมาเป็นประเทศเพื่อนบ้าน เช่น พม่า กัมพูชา ลาว

Region

- ครึ่งหนึ่งของทั้งหมด ตรวจพบเชื้อในภาคกลาง รองลงมาเป็นภาคใต้ และภาคเหนือมีจำนวนผู้ป่วยน้อยที่สุด

Province

- 17% ตรวจพบเชื้อที่กรุงเทพฯ อีก 13% รองลงมา พบในกลุ่มจังหวัดนิคมอุตสาหกรรม (เช่น ชลบุรี,สมุทรปราการ) รองลงมาอีก 10% เป็นกลุ่มจังหวัดปริมณฑล (เช่น นนทบุรี,ปทุมธานี)

Timeline

- ปี 2020 New case วันละหลักสิบ
- ต้นปี 2021 New case วันละหลักร้อย
- หลังสงกรานต์ ปี 2021 เริ่มติดเชื้อวันละหลักพัน
- ก.ค. ปี 2021 New case ถึงหลักหมื่นต้นๆ
- ส.ค. ปี 2021 Peak สูงถึงหลัก 20k
- ปี 2022 เป็นต้นมาเริ่มกลับมาระบาดอีกครั้ง สูงสุด 27k ในเดือนมี.ค. 2021

Risk

- 47% ได้รับเชื้อจากการสัมผัสผู้ติดเชื้อ รองลงมาเป็นการไปสถานที่ชุมชน

Timeline & Risk

- เริ่มติดโควิดจำนวนมาก จาก Cluster สมุทรสาคร ในช่วงปลายปี 2021
- ต่อมาเดือนพ.ค. 2022 ในวันเดียว กลุ่มเรือนจำตรวจพบเชื้อเกือบ 7 พันคน
- Cluster สถานบันเทิง จะเกิดขึ้นหลังเทศกาล เช่นสงกรานต์ ปีใหม่
- พ.ค. กลางปี 2021 กลับมาระบาดอีกครั้ง-ลอกใหญ่ ผู้คนเริ่มออกมาใช้ชีวิตตามปกติ จึงติดเชื้อจากสถานที่ทำงาน สถานที่ชุมชน หรือคนใกล้ชิด

Timeline & Growth

- ก่อนหน้าปี 2022 อัตราการเติบโตของยอดผู้ป่วย ใกล้เคียงกับอัตราการเติบโตของการตาย
- ปี 2022 เป็นต้นมา อัตราการเติบโตของการเสียชีวิต เพิ่มขึ้นน้อยกว่าอัตราการเติบโตของยอดผู้ป่วย

