

สวัสดีค่ะ สำหรับบทความนี้จะวิเคราะห์ข้อมูลเรื่อง การเสียชีวิตจากอุบัติเหตุทางถนนในประเทศไทยค่ะ โดยมีเป้าหมายคือ อยากทราบว่า มีจังหวัดไหนในประเทศไทยที่เราควรจะเฝ้าระวังหรือ ป้องกันเป็นพิเศษจากการเสียชีวิตเนื่องจากอุบัติเหตุทางถนน โดยข้อมูลที่เรานำมาใช้ในการวิเคราะห์ จะประกอบด้วย 3 ชุดข้อมูล คือ

1. ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน จากระบบบูรณาการข้อมูลการตายจากอุบัติเหตุทางถนน (3 ล้าน)

จากเว็บไซต์สำนักงานพัฒนารัฐบาลดิจิทัล <https://data.go.th/th/dataset/rtdi>

2. ข้อมูลจำนวนประชากรในแต่ละจังหวัด และแต่ละปีในประเทศไทย

จากเว็บไซต์สำนักงานสถิติแห่งชาติ

3. ข้อมูลชื่อจังหวัดภาษาไทย และภาษาอังกฤษ

จากเว็บไซต์

<https://www.enghero.com/article/%E0%B8%88%E0%B8%B1%E0%B8%87%E0%B8%AB%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B9%83%E0%B8%99%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B9%80%E0%B8%97%E0%B8%A8%E0%B9%84%E0%B8%97%E0%B8%A2%E0%B8%A0%E0%B8%B2%E0%B8%A9%E0%B8%B2%E0%B8%AD%E0%B8%B1%E0%B8%87%E0%B8%81%E0%B8%A4%E0%B8%A9>

ขั้นตอนที่ 1 : import library และเชื่อมต่อกับ google drive

```
[1] 1 import sys
    2 import pandas as pd
    3 import datetime
    4 import matplotlib.pyplot as plt

[2] 1 from google.colab import 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).
```

ขั้นตอนที่ 2 : ตรวจสอบความสมบูรณ์ของข้อมูล จัดการข้อมูล และวิเคราะห์ข้อมูล

```
[3] 1 # Accident information from https://data.go.th/th/dataset/rtdi
    2 df_accident = pd.read_excel('/content/drive/MyDrive/kh-mulphuuesiiychiwitcchaak-ubatiehtuthaangthnn-cchaakrabbuurnaakaarkh-muulkaartaaycchaak-ubatiehtuthaangthnn-3
    3
    4 df_accident_1 = df_accident[(df_accident['DEAD_YEAR'] >= 2018) & (df_accident['DEAD_YEAR'] <= 2021)][['id', 'DEAD_YEAR', 'DeadDate', 'AccProv']]
    5 df_accident_1.info()

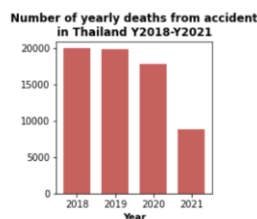
<class 'pandas.core.frame.DataFrame'>
Int64Index: 66467 entries, 3664 to 213187
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   id          66467 non-null   int64
1   DEAD_YEAR   66467 non-null   int64
2   DeadDate    66467 non-null   datetime64[ns]
3   AccProv     66466 non-null   object
dtypes: datetime64[ns](1), int64(2), object(1)
memory usage: 2.5+ MB
```

จากชุดข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนน จะทำการเลือกข้อมูลที่เกิดขึ้นในช่วงปี 2018-2021 มาใช้การวิเคราะห์ค่ะ โดยคอลัมน์ที่นำมาใช้วิเคราะห์ประกอบด้วย

- id = id ของผู้เสียชีวิต
- DEAD_YEAR = ปีที่เสียชีวิตของผู้ประสบเหตุ
- DeadDate = วันที่เสียชีวิต
- AccProv = จังหวัดที่เกิดอุบัติเหตุ

พบว่า จากข้อมูลทั้งหมด 66,467 บรรทัด ข้อมูลในแต่ละคอลัมน์ที่จะนำไปใช้วิเคราะห์มีความสมบูรณ์ครบถ้วน ไม่เป็นช่องว่าง (NULL) โดยสรุปข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนมีความสมบูรณ์อยู่ในเกณฑ์ที่ดีค่ะ สามารถนำไปใช้วิเคราะห์ต่อไปได้

```
[80] 1 df_accident_1_year = df_accident_1.groupby(['DEAD_YEAR'])['id'].count().to_frame().reset_index()
2 fig = plt.figure(figsize = (3, 3))
3 plt.bar(df_accident_1_year['DEAD_YEAR'], df_accident_1_year['id'], color = 'indianred', width = 0.7)
4 plt.xlabel("Year", weight='bold')
5 plt.title("Number of yearly deaths from accident\nin Thailand Y2018-Y2021", weight='bold')
6 plt.show()
```



ต่อมาจะดูแนวโน้มภาพรวมจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละปีค่ะ จากกราฟในปี 2018 และปี 2019 มีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนโดยเฉลี่ย 20,000 คน/ปี แต่พบว่าจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2020 และปี 2021 มีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนลดน้อยลงกว่าปี 2018 และปี 2019 มาก ซึ่งสาเหตุหลักน่าจะเกิดจากการแพร่ระบาดของโควิด-19 ที่รัฐบาลมีการออกนโยบายล็อกดาวน์ประเทศ และขอความร่วมมือประชาชนให้ออกจากบ้านเท่าที่จำเป็น จึงทำให้ข้อมูลจำนวนผู้เสียชีวิตในปี 2020 และปี 2021 ลดน้อยลงกว่าปี 2018 และปี 2019 อย่างมีนัยสำคัญ

```
[81] 1 df_accident_1['DEAD_YEAR_MONTH'] = df_accident_1['DeadDate'].dt.strftime('%Y-%m')
2 df_accident_1['DEAD_MONTH'] = df_accident_1['DeadDate'].dt.strftime('%m')
3 df_accident_1['DEAD_YEAR_MONTH'] = df_accident_1['DeadDate'].dt.strftime('%Y-%m')
4 df_accident_1_yearmonth = df_accident_1.groupby(['DEAD_YEAR_MONTH'])['id'].count()
5 df_accident_1_yearmonth = df_accident_1_yearmonth.plot(kind='bar', title='Number of monthly deaths from accident in Thailand Y2018-Y2021',
6 color = 'indianred', figsize=(18,5), fontsize=14)
```

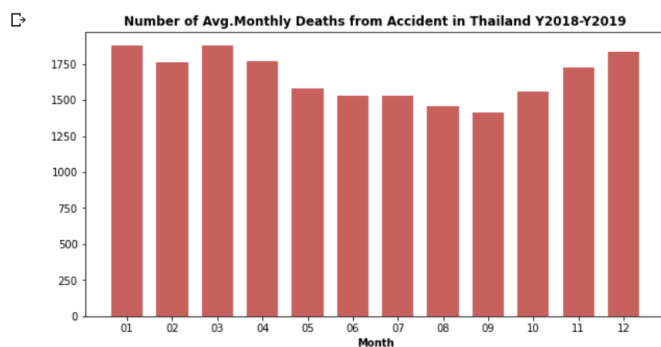


แต่เมื่อลองดูข้อมูลจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนของปี 2021 พบว่ามีจำนวนน้อยกว่าปี 2020 มากๆ จึงลองตรวจสอบความสมบูรณ์ของข้อมูล โดยพิจารณาจากวันที่ผู้เสียชีวิตจากอุบัติเหตุทางถนนล่าสุดของปี 2021 พบว่า ในปี 2021 มีข้อมูลจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนถึงแค่เดือน มิถุนายนเท่านั้น ยังไม่ครบปีเหมือนปีอื่นๆ

ในบทความนี้เราจะนำข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนของปี 2018 และ ปี 2019 มาใช้ในการวิเคราะห์ โดยมีสมมติฐานว่า ข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนของปี 2018 และปี 2019 น่าจะสะท้อนข้อมูลผู้เสียชีวิตจากอุบัติเหตุเมื่อสถานการณ์การแพร่ระบาดกลับสู่สภาวะปกติได้ดีที่สุดค่ะ

ต่อมาจะมาพิจารณาข้อมูลเป็นรายเดือน โดยพบว่า โดยเฉลี่ยเดือนมีนาคม เป็นเดือนที่มีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนสูงสุด และเดือนกันยายน เป็นเดือนที่มีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนโดยเฉลี่ยต่ำสุด

```
[83] 1 df_accident_2 = df_accident_1[(df_accident_1['DEAD_YEAR'] >= 2018) &
2     (df_accident_1['DEAD_YEAR'] <= 2019)][['id', 'DEAD_YEAR', 'DEAD_YEAR_MONTH', 'DEAD_MONTH', 'DeadDate', 'AccProv' ] ]
3
4 df_accident_2_month = df_accident_2.groupby(['DEAD_MONTH'])['id'].count().to_frame().reset_index()
5 df_accident_2_month['AVG_MONTHLY_ACCIDENT'] = (df_accident_2_month['id']/2).round()
6
7 fig = plt.figure(figsize = (10, 5))
8 plt.bar(df_accident_2_month['DEAD_MONTH'], df_accident_2_month['AVG_MONTHLY_ACCIDENT'], color = 'indianred', width = 0.7)
9
10 plt.xlabel("Month", weight='bold')
11 plt.title("Number of Avg.Monthly Deaths from Accident in Thailand Y2018-Y2019", weight='bold')
12 #plt.grid()
13 plt.show()
14 print(f'March : Maximum avg. monthly deaths from accident in Thailand.')
15 print(f'September : Minimum avg. monthly deaths from accident in Thailand.')
```



March : Maximum avg. monthly deaths from accident in Thailand.
September : Minimum avg. monthly deaths from accident in Thailand.

และเป็นที่สงสัยว่าทำไมเดือนมีนาคม ซึ่งไม่ได้เป็นเดือนเทศกาลและไม่ได้มีวันหยุดยาว ทำไมถึงมีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนโดยเฉลี่ย สูงกว่าเดือนอื่นๆ โดยจากการค้นหาข้อมูลเพิ่มเติมจากรายงานของกรมทางหลวง ในเดือนมีนาคม ปี 2020 ซึ่งเป็นเดือนก่อนมีการประกาศล็อกดาวน์ มีการระบุว่า ในเดือนมีนาคม ปี 2020 อุบัติเหตุส่วนใหญ่ที่เกิดขึ้นเกิดจากการขับรถด้วยความเร็วสูงกว่ากฎหมายกำหนด 75% ตัดหน้ากะทันหัน 8% หลบใน 6% อุบัติเหตุรถชนพ่วง 4% ขับรถไม่ชำนาญ ตามรูปด้านล่าง ซึ่งมีแนวโน้มสูงที่เดือนมีนาคมปี 2018 และเดือนมีนาคมปี 2019 จะมีสาเหตุของการเกิดอุบัติเหตุเหมือนปี 2020 ที่กรมทางหลวงชี้แจง นั่นก็คือ สาเหตุหลักของการเกิดอุบัติเหตุในเดือนมีนาคมมาจากการขับรถด้วยความเร็วสูงกว่ากฎหมายกำหนดค่ะ



Ref: <https://www.facebook.com/otpthailand.mot/photos/a.1342241342558148/2840605009388433/>

```
[8] 1 df_accident_1_pivot = pd.pivot_table(df_accident_1, values='id', index='AccProv', columns='DEAD_YEAR', aggfunc='count')
    2 df_accident_1_pivot
```

DEAD_YEAR	2018	2019	2020	2021
AccProv				
กระบี่	164	163	122	45
กรุงเทพมหานคร	837	893	828	399
กาญจนบุรี	307	285	269	127
กาฬสินธุ์	284	228	229	135
กำแพงเพชร	198	235	214	102
...
เพชรบุรี	199	145	155	79
เพชรบูรณ์	350	372	299	161
เลย	232	205	204	93
แพร่	163	138	110	53
แม่ฮ่องสอน	37	45	44	20

77 rows x 4 columns

จากนั้นเราจะนำข้อมูลผู้เสียชีวิตจากอุบัติเหตุทางถนนมาทำการ pivot และนับจำนวนผู้เสียชีวิตในแต่ละปีและแต่ละจังหวัด จัดรูปแบบออกมาเป็นข้อมูลที่พร้อมนำไปใช้วิเคราะห์ต่อไปดังรูปด้านบนค่ะ

โดยจากเป้าหมายของการวิเคราะห์คือ อยากทราบว่า จังหวัดไหนในประเทศไทยที่เราควรจะเฝ้าระวังหรือป้องกันเป็นพิเศษจากการเสียชีวิตเนื่องจากอุบัติเหตุทางถนน เพื่อลดจำนวนการเสียชีวิตจากอุบัติเหตุทางถนนลง แต่เนื่องจากแต่ละจังหวัดมีจำนวนประชากรมาก/น้อยแตกต่างกัน ซึ่งอาจส่งผลให้จังหวัดที่มีจำนวนประชากรมาก มีแนวโน้มที่จะมีจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนมากกว่าจังหวัดที่มีจำนวนประชากรน้อย เราสร้างตัวแปรใหม่ โดยการนำข้อมูลจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในปี 2019 ซึ่งเป็นข้อมูลล่าสุดที่ใช้ในการวิเคราะห์มาหารกับจำนวนประชากรในปี 2019 ของแต่ละจังหวัด ซึ่งจะได้ออกมาเป็นสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัด และนำมาใช้ในการเปรียบเทียบกันได้

สัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัดปี 2019 (%)

$$= \frac{\text{จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัดปี 2019}}{\text{จำนวนประชากรในแต่ละจังหวัดปี 2019}} \times 100$$

ต่อมา มาดูในส่วนของข้อมูลจำนวนประชากรแต่ละจังหวัดของประเทศไทยที่ได้จากเว็บไซต์สำนักงานสถิติแห่งชาติค่ะ ที่เราจะนำมาหารเพื่อให้ได้สัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัดปี 2019

เริ่มต้นจากการตรวจสอบความสมบูรณ์ของข้อมูลจำนวนประชากรแต่ละจังหวัดในแต่ละปี พบว่าจากข้อมูลทั้งหมด 2,132 บรรทัด ข้อมูลในแต่ละคอลัมน์ที่จะนำไปใช้วิเคราะห์มีความสมบูรณ์ครบถ้วน ไม่เป็นช่องว่าง (NULL) ข้อมูลมีความสมบูรณ์อยู่ในเกณฑ์ที่ดีค่ะ สามารถนำไปใช้วิเคราะห์ต่อไปได้

```
[9] 1 # Yearly Thai Population in each province from http://statbbi.nso.go.th/staticreport/page/sector/th/01.aspx
    2 df_population = pd.read_excel('/content/drive/MyDrive/sector_01_11102_TH.xlsx', sheet_name = 'ข้อมูล')
    3 df_population.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2132 entries, 0 to 2131
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   ภาค                  2132 non-null  object
1   จังหวัด                2132 non-null  object
2   กลุ่มอายุ (ปี)         2132 non-null  object
3   2555                  2132 non-null  int64
4   2556                  2132 non-null  int64
5   2557                  2132 non-null  int64
6   2558                  2132 non-null  int64
7   2559                  2132 non-null  int64
8   2560                  2132 non-null  int64
9   2561                  2132 non-null  int64
10  2562                  2132 non-null  int64
11  2563                  2132 non-null  int64
12  2564                  2132 non-null  int64
dtypes: int64(10), object(3)
memory usage: 216.7+ KB
```

แต่เนื่องจากจังหวัดในประเทศไทยมีทั้งหมด 77 จังหวัด แต่ข้อมูลมีมากถึง 2,132 บรรทัด จึงลองดูลักษณะข้อมูลเพิ่มเติม พบว่า ข้อมูลมีการแบ่งจำนวนประชากรในแต่ละจังหวัด ออกเป็นช่วงอายุต่างๆ จึงทำให้ข้อมูลมีจำนวนบรรทัดมากกว่าจำนวนจังหวัด และมีบรรทัดที่สรุปรวมทุกช่วงอายุของจำนวนนั้นๆ ด้วย ซึ่งเราจะใช้บรรทัดที่เค้ารวบรวมมาให้ของแต่ละจังหวัดนี้แทนจำนวนประชากรในจังหวัดนั้น โดยใช้เงื่อนไข (df_population['กลุ่มอายุ (ปี)'] == 'รวม') อีกทั้งยังมีการรวมข้อมูลเป็นรายภาค และรวมเป็นทั่วราชอาณาจักรด้วย เราจะกรองเอาข้อมูลในคอลัมน์ "จังหวัด" ที่เป็น 'ทั่วราชอาณาจักร', 'ภาคกลาง', 'ภาคเหนือ', 'ภาคตะวันออกเฉียงเหนือ' และ 'ภาคใต้' ออก เพื่อความสะดวกในการนำไป join กับข้อมูลจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนเพื่อคำนวณหาสัดส่วนจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัดต่อไป

```
[10] 1 df_population_1 = df_population[(df_population['กลุ่มอายุ (ปี)'] == 'รวม') &
2 (df_population['จังหวัด'] != 'ทั่วราชอาณาจักร') &
3 (df_population['จังหวัด'] != 'ภาคกลาง') &
4 (df_population['จังหวัด'] != 'ภาคเหนือ') &
5 (df_population['จังหวัด'] != 'ภาคตะวันออกเฉียงเหนือ') &
6 (df_population['จังหวัด'] != 'ภาคใต้')][['จังหวัด', 2561, 2562, 2563, 2564]]
7 df_population_1.rename({'จังหวัด': 'PROVINCE_TH', 2561: 'POPULATION_Y2018', 2562: 'POPULATION_Y2019',
8 2563: 'POPULATION_Y2020', 2564: 'POPULATION_Y2021'}, axis=1, inplace=True)
9 df_population_1
```

	PROVINCE_TH	POPULATION_Y2018	POPULATION_Y2019	POPULATION_Y2020	POPULATION_Y2021
26	กรุงเทพมหานคร	5676648	5666264	5588222	5527994
78	สมุทรปราการ	1326608	1344875	1351479	1356449
104	นนทบุรี	1246295	1265387	1276745	1288637
130	ปทุมธานี	1146092	1163604	1176412	1190060
156	พระนครศรีอยุธยา	817441	820188	819088	820512
...
2002	ตรัง	643116	643164	640574	639788
2028	พัทลุง	525044	524865	523077	522541
2054	ปัตตานี	718077	725104	726015	729581
2080	ยะลา	532326	536330	538602	542314
2106	นราธิวาส	802474	808020	804429	809660

77 rows x 5 columns

และอีกตัวแปรหนึ่งที่เราจะสร้างคือ อัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนโดยเทียบปี 2018 กับ ปี 2019 ว่าแต่ละจังหวัดมีการเติบโตของจำนวนผู้เสียชีวิตเป็นอย่างไร โดยคำนวณจาก

อัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนน (%)

$$= \frac{(\text{จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนปี 2019} - \text{จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนปี 2018})}{\text{จำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนปี 2018}} \times 100$$

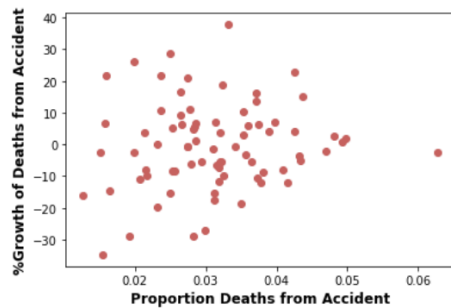
```
[17] 1 df_final = pd.merge(df_accident_1_pivot, df_population_1, left_on='AccProv', right_on='PROVINCE_TH')
2 df_final.rename({'2018': 'ACCIDENT_Y2018', '2019': 'ACCIDENT_Y2019'}, axis=1, inplace=True)
3 df_final = pd.merge(df_final, df_mapping, left_on='PROVINCE_TH', right_on='PROVINCE_TH')
4 df_final = df_final[['PROVINCE_TH', 'ACCIDENT_Y2018', 'ACCIDENT_Y2019',
5 'POPULATION_Y2018', 'POPULATION_Y2019', 'PROVINCE_ENG']]
6 df_final['PERCENTAGE_GROWTH'] = (df_final['ACCIDENT_Y2019'] - df_final['ACCIDENT_Y2018']) * 100 / df_final['ACCIDENT_Y2018']
7 df_final['ACCIDENT_PROPORTION_Y2019'] = df_final['ACCIDENT_Y2019'] * 100 / df_final['POPULATION_Y2019']
8 df_final
```

	PROVINCE_TH	ACCIDENT_Y2018	ACCIDENT_Y2019	POPULATION_Y2018	POPULATION_Y2019	PROVINCE_ENG	PERCENTAGE_GROWTH	ACCIDENT_PROPORTION_Y2019
0	กระบี่	164	163	473738	476739	Krabi	-0.609756	0.034191
1	กรุงเทพมหานคร	837	893	5676648	5666264	Krung Thep Maha Nakhon (Bangkok)	6.690562	0.015760
2	กาญจนบุรี	307	285	893151	895525	Kanchanaburi	-7.166124	0.031825
3	กาฬสินธุ์	284	228	985346	983418	Kalasin	-19.718310	0.023184
4	กำแพงเพชร	198	235	727807	725867	Kamphaeng Phet	18.686869	0.032375
...
72	เพชรบุรี	199	145	484294	485191	Phetchaburi	-27.135678	0.029885

จากนั้นเอาข้อมูลสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนน และอัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนในแต่ละจังหวัดมา plot scatter โดย

แกน x แทนสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนน (%)

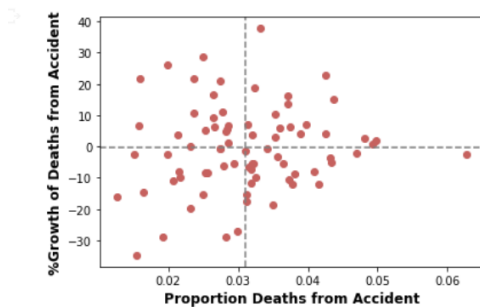
แกน y แทนอัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนน (%)



<Figure size 1300x600 with 0 Axes>
Mean of Accident proportion Y2019: 0.03
Mean of %Growth: -0.49

ซึ่งเราพบว่าข้อมูลสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนนมีค่ามากกว่า 0 เราจึงทำการสร้างแกนอ้างอิงขึ้นมาใหม่ โดยทำการ plot กราฟเส้นตามแนวแกน y ด้วยค่าเฉลี่ยของสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนน และ plot กราฟเส้นตามแนวแกน x ด้วยค่าเฉลี่ยของอัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนน ซึ่งจะช่วยให้แบ่งข้อมูลออกเป็น 4 ส่วน

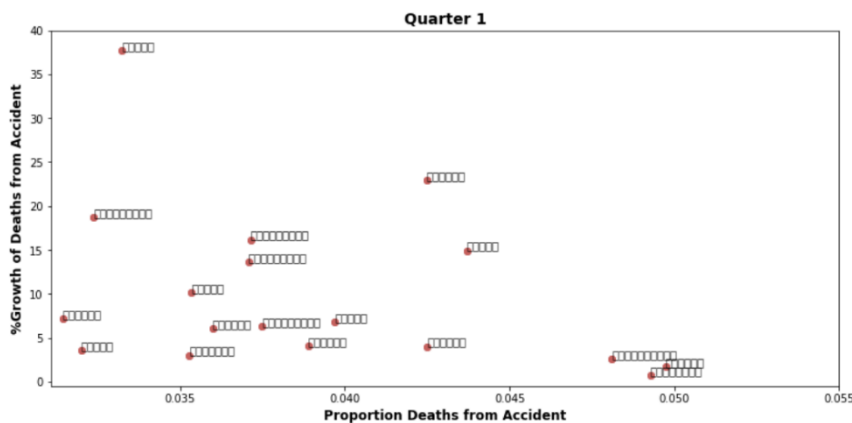
```
[108] 1 ACCIDENT_PROPORTION_Y2019_MEAN = df_final['ACCIDENT_PROPORTION_Y2019'].mean()
2 PERCENTAGE_GROWTH_MEAN = df_final['PERCENTAGE_GROWTH'].mean()
3 plt.scatter(df_final['ACCIDENT_PROPORTION_Y2019'], df_final['PERCENTAGE_GROWTH'], color = 'indianred')
4 plt.xlabel('Proportion Deaths from Accident', fontsize=12, weight='bold')
5 plt.ylabel('%Growth of Deaths from Accident', fontsize=12, weight='bold')
6 plt.axvline(ACCIDENT_PROPORTION_Y2019_MEAN, c='gray', ls='--')
7 plt.axhline(PERCENTAGE_GROWTH_MEAN, c='gray', ls='--')
8 plt.figure(figsize=(13, 6), dpi = 100)
9 plt.show()
10
11 print(f'Mean of Accident proportion Y2019: {ACCIDENT_PROPORTION_Y2019_MEAN:.2f}')
12 print(f'Mean of %Growth: {PERCENTAGE_GROWTH_MEAN:.2f}')
```



<Figure size 1300x600 with 0 Axes>
Mean of Accident proportion Y2019: 0.03
Mean of %Growth: -0.49

โดยข้อมูลจังหวัดที่อยู่ใน Quarter 1 หรือตำแหน่งในส่วนบนขวา ถือว่าเป็นจังหวัดที่คิดว่าควรมีการเฝ้าระวัง เนื่องจากเป็นจังหวัดที่มีสัดส่วนผู้เสียชีวิตจากอุบัติเหตุทางถนน และอัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนสูงกว่าค่าเฉลี่ย และจากกราฟ scatter เราไม่ทราบค่า แต่ละจุดคือจังหวัดใด จึงอยากสร้าง label ชื่อจังหวัดเพื่อความสะดวกในการตีความ แต่เนื่องจากใช้ชื่อภาษาไทย พอแสดงในกราฟแล้วไม่สามารถอ่านชื่อจังหวัดได้ ดังรูปด้านล่าง

```
[109] 1 df_final_q1 = df_final[(df_final['ACCIDENT_PROPORTION_Y2019'] >= ACCIDENT_PROPORTION_Y2019_MEAN) &
2 (df_final['PERCENTAGE_GROWTH'] >= PERCENTAGE_GROWTH_MEAN)]
3 plt.scatter(df_final_q1['ACCIDENT_PROPORTION_Y2019'], df_final_q1['PERCENTAGE_GROWTH'], color = 'indianred')
4
5 x = df_final_q1['ACCIDENT_PROPORTION_Y2019'].tolist()
6 y = df_final_q1['PERCENTAGE_GROWTH'].tolist()
7 text = df_final_q1['PROVINCE_TH'].tolist()
8
9 plt.scatter(x, y, color = 'indianred')
10
11 # Loop for annotation of all points
12 for i in range(len(x)):
13     plt.annotate(text[i], (x[i], y[i] + 0.2))
14
15 # adjusting the scale of the axes
16 plt.xlim((ACCIDENT_PROPORTION_Y2019_MEAN, 0.055))
17 plt.ylim((PERCENTAGE_GROWTH_MEAN, 40))
18 plt.title('Quarter 1', fontsize=14, weight='bold')
19 plt.xlabel('Proportion Deaths from Accident', fontsize=12, weight='bold')
20 plt.ylabel('%Growth of Deaths from Accident', fontsize=12, weight='bold')
21 plt.gcf().set_size_inches(13, 6)
22 plt.show()
```



จึงนำข้อมูลชื่อจังหวัดภาษาอังกฤษมา join กับไฟล์ข้อมูล เพื่อให้ได้ label ชื่อจังหวัดภาษาอังกฤษ แต่เมื่อ import ไฟล์ข้อมูลชื่อจังหวัด เข้าใน python เกิด error ขึ้นว่า encoding utf-8 ไม่สามารถ encode ได้

```
1 df_mapping = pd.read_csv('/content/drive/MyDrive/province_mapping.csv')
2 df_mapping.info()

UnicodeDecodeError                                Traceback (most recent call last)
<ipython-input-25-56077f0188cc> in <module>
----> 1 df_mapping = pd.read_csv('/content/drive/MyDrive/province_mapping.csv')
      2 df_mapping.info()

~
~
~
9 frames
/usr/local/lib/python3.7/dist-packages/pandas/_libs/parsers.pyx in pandas._libs.parsers.raise_parser_error()

UnicodeDecodeError: 'utf-8' codec can't decode byte 0xa8 in position 0: invalid start byte
```

จึงได้เปลี่ยนไปใช้ encoding = "TIS-620" แทน และชื่อจังหวัดมีทั้งหมด 77 จังหวัด ครบถ้วน สามารถนำไปใช้วิเคราะห์ต่อได้

```
[11] 1 # Mapping TH province to ENG province from https://www.enghero.com/article/%E0%B8%88%E0%B8%B1%E0%B8%87%E0%B8%AB%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B9%83%E0%B8%99%E0%B8%9B
2 df_mapping = pd.read_csv('/content/drive/MyDrive/province_mapping.csv', encoding = "TIS-620")
3 df_mapping.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   จังหวัดภาษาไทย       77 non-null     object
1   จังหวัดภาษาอังกฤษ     77 non-null     object
dtypes: object(2)
memory usage: 1.3+ KB
```

โดยจากกราฟพบว่า ใน Quarter 1 มีจังหวัดทั้งหมด 18 จังหวัด ซึ่งเป็นจังหวัดที่มีสัดส่วนผู้เสียชีวิตปี 2019 และอัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนสูงกว่าค่าเฉลี่ย

โดยมี 3 จังหวัด คือ พะเยา, ภูเก็ต, ลำพูน ซึ่งเป็นจังหวัดที่มีสัดส่วนผู้เสียชีวิตในจังหวัดปี 2019 และ อัตราการเติบโตของจำนวนผู้เสียชีวิตจากอุบัติเหตุทางถนนสูงกว่าจังหวัดอื่นๆในประเทศไทย และคิดว่า 3 จังหวัดนี้ น่าจะเป็นจังหวัดที่เราควรจะต้องเฝ้าระวังมากขึ้นหรือป้องกันเป็นพิเศษ เพื่อที่จะลดจากการเสียชีวิตเนื่องจากอุบัติเหตุทางถนนค่ะ

```
[110] 1 df_final_q1 = df_final[(df_final['ACCIDENT_PROPORTION_Y2019'] >= ACCIDENT_PROPORTION_Y2019_MEAN) &
0s      (df_final['PERCENTAGE_GROWTH'] >= PERCENTAGE_GROWTH_MEAN)]
2
3
4 x = df_final_q1['ACCIDENT_PROPORTION_Y2019'].tolist()
5 y = df_final_q1['PERCENTAGE_GROWTH'].tolist()
6 text = df_final_q1['PROVINCE_ENG'].tolist()
7
8 plt.scatter(x, y, color = 'indianred')
9 plt.title('Quarter 1', fontsize=14, weight='bold')
10 plt.xlabel('Proportion Deaths from Accident', fontsize=14, weight='bold')
11 plt.ylabel('%Growth of Deaths from Accident', fontsize=14, weight='bold')
12
13 # Loop for annotation of all points
14 for i in range(len(x)):
15     plt.annotate(text[i], (x[i], y[i] + 0.2), fontsize=11)
16
17 # adjusting the scale of the axes
18 plt.xlim((ACCIDENT_PROPORTION_Y2019_MEAN, 0.055))
19 plt.ylim((PERCENTAGE_GROWTH_MEAN, 40))
20 plt.gcf().set_size_inches(13, 6)
21 plt.show()
```

