

การวิเคราะห์งบประมาณรายได้และรายจ่ายขององค์กรปกครองส่วนท้องถิ่นปี พ.ศ. 2562

Analysis of income and expenditure budget in 2019 (local administrative organization)





Abstract



การวิจัยนี้มีวัตถุประสงค์ เพื่อตรวจสอบว่ามีองค์กรปกครองส่วนท้องถิ่น แห่งใด มีความแตกต่างของรายได้และรายจ่ายแตกต่างกันมากจนผิดปกติหรือไม่ เพื่อดูความสัมพันธ์ระหว่างข้อมูลรายได้ขององค์กรปกครองส่วนท้องถิ่น กับ ข้อมูลรายจ่ายขององค์กรปกครองส่วนท้องถิ่น เพื่อจัดกลุ่มขององค์กรปกครองส่วนท้องถิ่นที่มีลักษณะใกล้เคียงกันเข้าด้วยกัน

การวิจัยนี้เป็นการวิจัยเชิงวินิจฉัย เริ่มด้วยการทำ Exploration & Cleansing Data จากนั้นทำการวิเคราะห์ด้วยการดูความสัมพันธ์ของข้อมูลทั้งหมด และใช้ วิธีการเรียนรู้แบบไม่มีผู้สอน โดยงานวิจัยนี้ใช้ K-Meansในการแบ่งข้อมูล จากนั้นก็ทำการวิเคราะห์ข้อมูลเชิงลึก แล้วทำ Data Visualization โดยใช้ภาษา

Python ผ่านโปรแกรม Google Colaboratory





ผลของการวิจัยแสดงให้เห็นว่า งบประมาณรายได้และรายจ่ายของ องค์กรปกครองส่วนท้องถิ่นประจำปี พ.ศ. 2562 สามารถแบ่งข้อมูลได้เป็น 2 กลุ่ม โดยกลุ่มที่ 1 มีจำนวนองค์กรปกครองส่วนท้องถิ่นทั้งหมด 6,500 แห่ง และสามารถแบ่งตามจังหวัดได้ทั้งหมด 76 จังหวัด และสามารถแบ่งตาม ภาคได้ 6 ภาค และในกลุ่มที่ 2 มีจำนวนองค์กรปกครองส่วนท้องถิ่นทั้งหมด 89 แห่ง และสามารถแบ่งตามจังหวัดได้ทั้งหมด 76 จังหวัด และสามารถแบ่ง ตามภาคได้ออกเป็น 6 ภาค

วัตถุประสงค์ของการวิจัย

01

03

02

เพื่อตรวจสอบว่ามีองค์กรปกครองส่วนท้องถิ่นแห่งใด มีความแตกต่าง ของรายได้และรายจ่ายแตกต่างกันมากจนผิดปกติหรือไม่

เพื่อดูความสัมพันธ์ของข้อมูลรายได้ขององค์กรปกครองส่วน ท้องถิ่น ข้อมูลรายจ่ายขององค์กรปกครองส่วนท้องถิ่น และ ข้อมูลระหว่างรายได้และรายจ่ายขององค์กรปกครองส่วน ท้องถิ่น

เพื่อจัดกลุ่มขององค์กรปกครองส่วนท้องถิ่นที่มีลักษณะใกล้เคียงกัน เข้าด้วยกัน และแยกกลุ่มขององค์กรปกครองส่วนท้องถิ่นที่มีลักษณะ ต่างกันออกจากกัน

ขอบเขตของการวิจัย



ขอบเขตด้านเนื้อหา: การวิจัยครั้งนี้ มุ่งเน้นศึกษาเกี่ยวกับงบประมาณรายได้ และรายจ่ายขององค์กรปกครองส่วนท้องถิ่นปี พ.ศ. 2562 ในแต่ละแห่ง

ขอบเขตด้านพื้นที่: การวิจัยครั้งนี้ มีประชากรที่ใช้ในการวิจัย คือ ประชาชนที่อยู่ ในองค์กรปกครองส่วนท้องถิ่นจำนวน 7,852 แห่ง รวมทั้งสิ้น 66,023,670 คน



02



ขอบเขตด้านระยะเวลา: การวิจัยครั้งนี้ ทำการวิจัยในช่วงเดือนกันยายนจนถึง เดือนตุลาคม พ.ศ. 2563

الأرا والخرب وبالات الأوالكي تنبي أنها أرجا وألازا الطفر ووبالات المالكي تما

ขั้นตอนการดำเนินงาน

- **Ingest and Exploration Data Preparation and Cleansing Data** 03 **Normalization** 04 **Correlation** 05 **K-Means**
 - 106 Insight and Data Visualization

01

Ingest and Exploration Data

Income

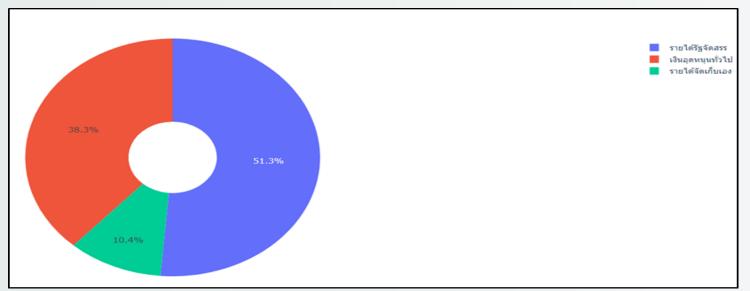
| | ลำดับ | ชื่อท้องถิ่น | 5N | อำเภอ | จังหวัด | พื้นที่ (ตร.กม.) | ประชากร ณ30กย61 | ภาษีโรงเรือน และที่ดิน | ภาษีบำรุงท้อง ที่ | ภาษีป่าย | อากรการ ฆ่าสัตว์ | อากร รังนก อีแอ่น | ภาษีบำรุง อบจ. จาก สถานค้าปลีก ยาสูบ | ภาษีบ่ารุง อบจ. จาก สถานค้าปลีก น้ำมัน | ค่า ธรรมเนียม บำรุง อบจ. จากผู้เข้าพัก โรงแรม | ค่าธรรมเนียม | ทรัพย์สิน | สาธารณูปโภค | បើលលេងីល |
|---|-------|----------------------|-----|-------------|---------|---------------------|--------------------|---------------------------|----------------------|--------------|---------------------|-------------------------|---|---|---|--------------|--------------|-------------|--------------|
| 0 | 1.0 | กรุงเทพมหานคร | NaN | NaN | NaN | 1568.74 | 5677054.0 | 1.508583e+10 | 1.417550e+08 | 9.306081e+08 | 859812.0 | 0.00 | 0.00 | 2.227340e+08 | NaN | 1.138100e+09 | 1.102230e+09 | 41809382.46 | 1.091146e+09 |
| 1 | 2.0 | เมืองพัทยา | NaN | บางละมุง | ชลบุรี | 208.10 | 119130.0 | 4.587382e+08 | 3.444746e+06 | 6.376105e+07 | 0.0 | 0.00 | 0.00 | 0.000000e+00 | 0.00 | 1.296367e+08 | 5.367728e+07 | 1849453.31 | 1.555355e+08 |
| 2 | 3.0 | อบจ.กระบี่ | 1.0 | เมืองกระบี่ | กระบี่ | NaN | NaN | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.0 | 81276.11 | 20570588.16 | 5.496926e+06 | 32315881.31 | 8.129370e+06 | 5.895315e+06 | 0.00 | 1.234181e+07 |
| 3 | 4.0 | ทม.กระบี่ | 3.0 | เมืองกระบี่ | กระบี่ | 19.00 | 32306.0 | 4.634454e+07 | 1.140400e+06 | 1.055219e+07 | 63830.0 | 81276.11 | 0.00 | 0.000000e+00 | 0.00 | 2.318395e+07 | 2.378127e+07 | 5504205.89 | 5.828320e+06 |
| 4 | 5.0 | ทต.เกาะลันตา ใหญ่ | 4.0 | เกาะลันตา | กระบี่ | 0.81 | 1184.0 | 2.418063e+05 | 2.490740e+03 | 4.003200e+04 | 0.0 | 81276.11 | 0.00 | 0.000000e+00 | 0.00 | 1.994830e+05 | 6.322416e+05 | 457535.00 | 2.610000e+03 |

7852 rows 42 columns

01

Ingest and Exploration Data





Ingest and Exploration Data

Spending

| 1 to 5 of 5 entries Filter | | | | | | | | | | | |
|----------------------------|---------|-----------|-------|---------------|---------------|-------------|------------|--|--|--|--|
| index | จังหวัด | อำเภอ | ปรเภท | อปห. | งบ | หมวด | จ่ายจริง | | | | |
| 0 | กระบี่ | เกาะลันตา | ทต. | เกาะลันตาใหญ่ | งบกลาง | งบกลาง | 2070762.86 | | | | |
| 1 | กระบี่ | เกาะลันตา | ทต. | เกาะลันตาใหญ่ | งบเงินอุดหนุน | เงินอุดหนุน | 1146000.0 | | | | |
| 2 | กระบี่ | เกาะลันตา | ทต. | เกาะลันตาใหญ่ | งบดำเนินงาน | ค่าใช้สอย | 3061358.37 | | | | |
| 3 | กระบี่ | เกาะลันตา | ทต. | เกาะลันตาใหญ่ | งบดำเนินงาน | ค่าตอบแทน | 201950.0 | | | | |
| 4 | กระบี่ | เกาะลันตา | ทต. | เกาะลันตาใหญ่ | งบดำเนินงาน | ค่าวัสดุ | 821572.18 | | | | |
| Show 25 v per page | | | | | | | | | | | |

82310 rows 7 columns

Income



ทำการเลือกคอลัมน์ที่ต้องการ แล้วจัดกลุ่มตาม "ชื่อท้องถิ่น" และเลือกมาเฉพาะแถวที่ คอลัมน์ "ประชากร ณ30กย61" ใม่เป็น 0

```
1 income_ana = (
2   income_select[['ชื่อท้องถิ่น','ประชากร ณ30กย61','รายได้จัดเก็บเอง','รายได้รัฐจัดสรร','เงินอุดหนุนทั่วไป', 'รายได้รวม']]
3   .groupby(['ชื่อท้องถิ่น']).sum()
4   .reset_index()
5   .sort_values(by='รายได้รวม', ascending=True)
6 )
1 income_whole = income_ana[income_ana['ประชากร ณ30กย61'] != 0]
```

Spending



ทำการสร้างคอลัมน์ "ชื่อท้องถิ่น" ขึ้นมา เพื่อเตรียมไว้สำหรับการ Join กับ ตารางของ Income

```
1 ### Spending ###
2 spending['ชื่อท้องถิ่น'] = spending['ปรเภท'] + spending['อปท.']
```

Spending



ทำ Pivot Table

```
1 spending_pivot = spending.pivot_table(index='ชื่อท้องถิ่น', columns='หมวด', values='จ่ายจริง').groupby('ชื่อท้องถิ่น').sum()

1 spending_pivot.shape

(6706, 11)
```

Spending



ทำการเลือกคอลัมน์ที่ต้องการ แล้วจัดกลุ่มตาม "ชื่อท้องถิ่น"

```
1 spending_real = (
2 spending[['ชื่อท้องถิ่น', 'จ่ายจริง']]
3 .groupby('ชื่อท้องถิ่น').sum()
4 .reset_index()
5 .sort_values(by='จ่ายจริง', ascending=True)
6 )
```

Spending



ทำการ Join ระหว่างตารางแรกที่เอาข้อมูลเข้ามา กับตารางที่ทำ Pivot Table มาแล้ว และเลือกมาเฉพาะแถวที่คอลัมน์ "จ่ายจริง" ไม่เป็น 0

```
1 spending_whole = pd.merge(spending_pivot,

2 spending_real[['ชื่อท้องถิ่น', 'จ่ายจริง']],

3 on = 'ชื่อท้องถิ่น',

4 how = 'left')
```

```
1 spending_whole = spending_whole[spending_whole['จ่ายจริง'] != 0]
```

Report



ทำการ Join ตารางระหว่าง Income กับ Spending และทำการลบ Missing Value ทั้งตาราง

```
1 report_whole = pd.merge(spending_whole,
2 income_whole[['ชื่อท้องถิ่น','ประชากร ณ30กย61','รายได้จัดเก็บเอง','รายได้รัฐจัดสรร','เงินอุดหนุนทั่วไป', 'รายได้รวม']],
3 on = 'ชื่อท้องถิ่น',
4 how = 'left')

1 report_whole = report_whole.dropna()
```

6598 rows 18 columns —

Report



ทำการสร้างข้อมูล Income กับ Spending แบบต่อหัวประชากรขึ้นมา

```
1 report_ana = report_whole.copy()
1 report_ana_whole = report_ana.drop(columns=['ประชากร ณ30กย61'])
 1 report_ana_whole.shape
(6598, 17)
1 IncomePerPeople = report_ana_whole['รายได้รวม']/report_whole['ประชากร ณ30กย61']
 2 report ana whole['IncomePerPeople'] = IncomePerPeople
 1 spendingPerPeople = report_ana_whole['จ่ายจริง']/report_whole['ประชากร ณ30กย61']
 2 report ana whole['SpendingPerPeople'] = spendingPerPeople
```

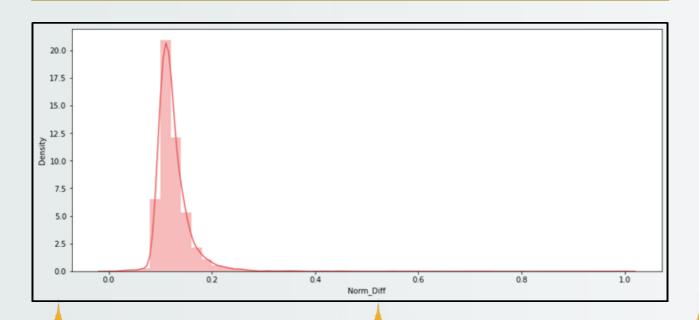
Report



ทำการหาผลต่างของข้อมูล Income กับ Spending แบบต่อหัวประชากร

- 1 diffBtweenIncomeSpending = report_ana_whole['IncomePerPeople'] report_ana_whole['SpendingPerPeople']
- 2 report_ana_whole['Diff'] = diffBtweenIncomeSpending

ทำ Normalization ของค่าผลต่างระหว่าง Income กับ Spending แบบต่อหัวประชากร โดยใช้สูตรต่อไปนี้ : norm_x = (x - min_x) / (max_x - min_x)



Normalization

ไม่ต้องการค่า Norm_Diff ที่มีค่ามากกว่า 0.05 เลยทำการเลือกค่าที่คอลัมน์ "Diff" ที่น้อยกว่า 20000

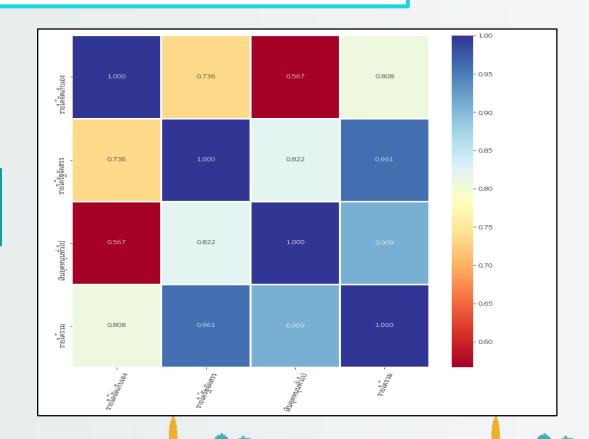
```
1 report_ana_diff = report_ana_whole.loc[report_ana_whole['Diff'] < 20000]</pre>
```

1 report_ana_diff.shape

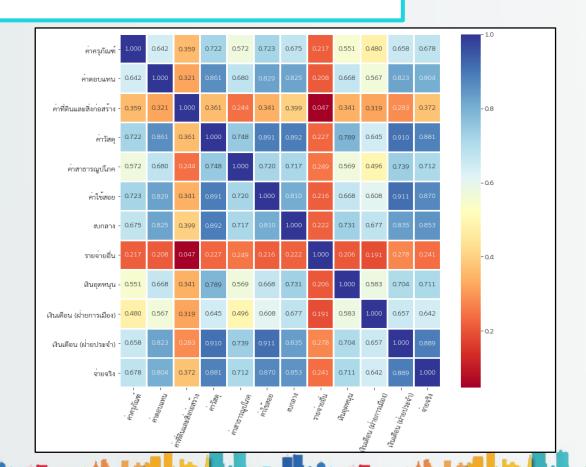
(6589, 20)

6589 rows 20 columns

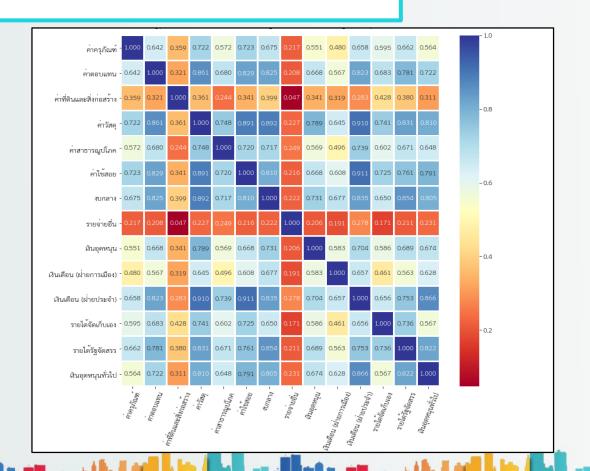
ความสัมพันธ์ระหว่างข้อมูล รายได้ กับ รายได้



ความสัมพันธ์ระหว่างข้อมูล รายจ่าย กับ รายจ่าย



ความสัมพันธ์ระหว่างข้อมูล รายได้ กับ รายจ่าย



Silhouette Score



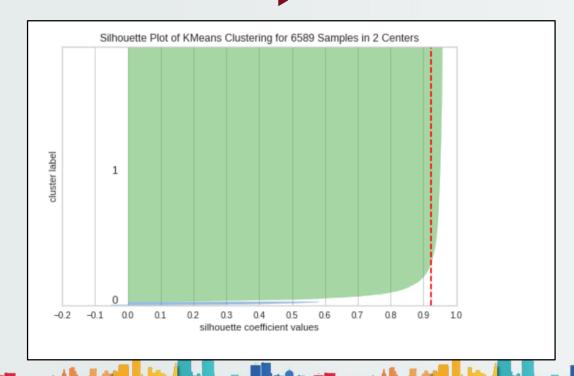
ทำการทดสอบเพื่อหาค่า k ที่เหมาะสม

```
1 z = report ana diff[['ต่าตรุภัณฑ์', 'ต่าตอบแทน', 'ต่าที่ดินและสิ่งก่อสร้าง', 'ต่ารัสดุ', 'ต่าสาธารณปโภค', 'ต่าใช้สอย', 'งบกลาง', 'รายจ้อน', 'เงินอุดหนุน', 'เงินต็อน (ฝ่ายการเมือง)', 'เงินเดือน (ฝ่ายประจำ)', 'จ่ายจริง', 'รายได้รัดเก็บเอง', 'รายได้รัฐจัดสรร', 'เงินอุดห
 1 x = report_ana_diff[['เงินอุดหนุนทั่วไป','เงินเดือน (ฝ่ายประจำ)']]
 1 result_sill_score = sil_score(z, 2, 10)
 2 print(f'scores = {result_sill_score}')
 3 print(f'optimal number of clusters = {max(result_sill_score)[1]}')
scores = [[0.9124, 2], [0.8368, 3], [0.7391, 4], [0.6573, 5], [0.5708, 6], [0.5709, 7], [0.5708, 8], [0.5329, 9], [0.4683, 10]]
optimal number of clusters = 2
 1 result sill score = sil score(x, 2, 10)
 2 print(f'scores = {result sill score}')
 3 print(f'optimal number of clusters = {max(result_sill_score)[1]}')
scores = [[0.9219, 2], [0.8447, 3], [0.7218, 4], [0.5757, 5], [0.5712, 6], [0.5052, 7], [0.5107, 8], [0.498, 9], [0.4633, 10]]
optimal number of clusters = 2
```

Silhouette Score



 $\vec{\hat{\eta}} k = 2$



พบได้ว่าค่า Silhouette ของ K-Means จากข้อมูลตัวอย่างทั้งหมด 6,589 ตัวอย่าง มีค่าสูงสุดคือ 0.9219 สรุปได้ว่าค่า k = 2 มีความเหมาะสมกับข้อมูล ตัวอย่างมากที่สุด

Cluster

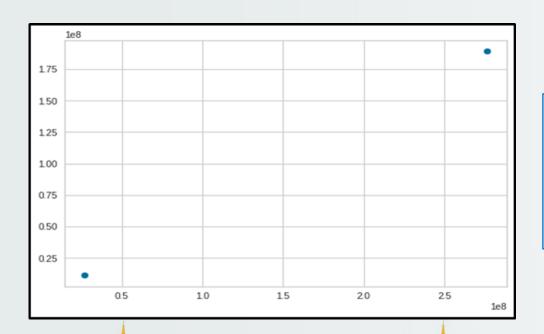


จากค่าของจุดศูนย์กลางของทั้ง 2 Cluster ทำให้ทราบว่า Cluster 1 มีค่ามากกว่า Cluster 0 มาก และหลังจากนั้นทำการเปลี่ยนชื่อ จาก Cluster 0 เป็น Cluster 2

Cluster

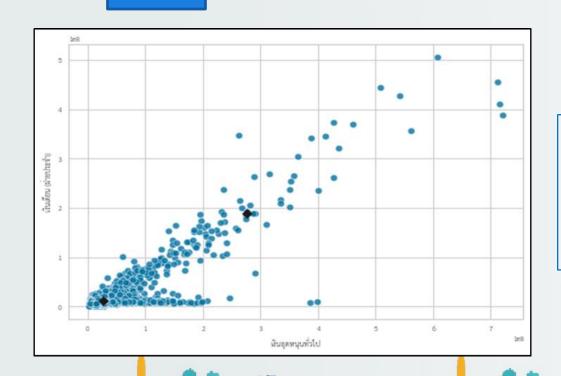


Fitted จุดกึ่งกลางของแต่ละ Cluster



พบได้ว่า Cluster ของข้อมูลตัวอย่างทั้งหมด 6,589 ตัวอย่าง ซึ่งคือ Cluster 2 และ 1 นี้ สรุปได้ว่า จุดศูนย์กลางของข้อมูลในแต่ละ Cluster มีการแบ่งแยกกันอย่างชัดเจน

Fitted Data



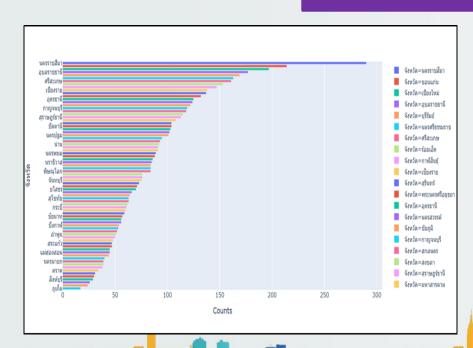
เป็นแผนภาพการกระจายที่แสดงถึงความสัมพันธ์ ระหว่าง ตัวแปร "เงินอุดหนุนทั่วไป" และตัวแปร "เงินเดือน (ฝ่ายประจำ)" โดยทั้งสองตัวแปรมีการเปลี่ยนแปลง ไปในทิศทางเดียวกัน 06

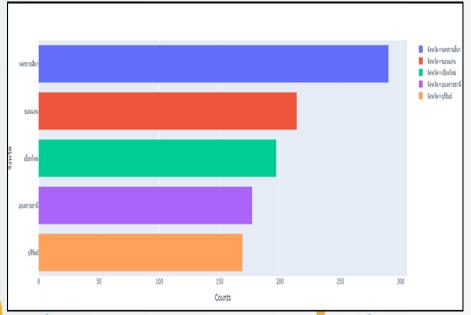
Insight and Data Visualization

Counts

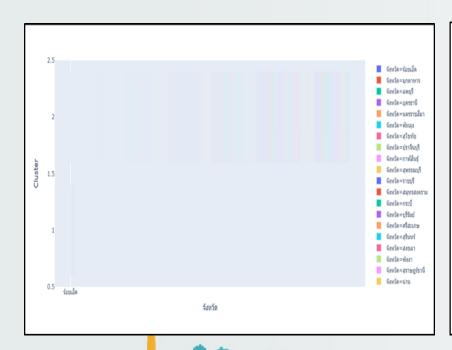


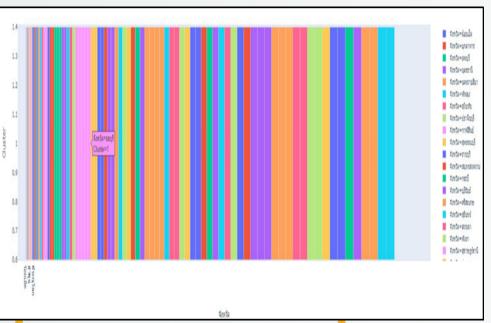
ดูข้อมูลแบบรวมทุกองค์กรปกครองส่วนท้องถิ่นในทุก ๆ จังหวัดของประเทศไทย กับ 5 จังหวัดแรกที่มีจำนวนองค์กรปกครองส่วนท้องถิ่นมากที่สุด



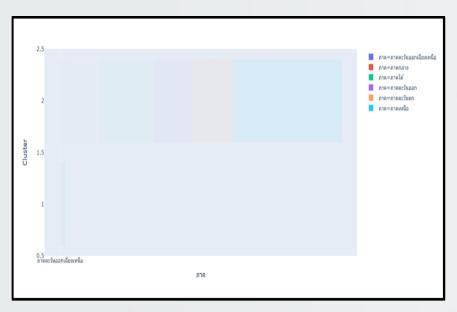


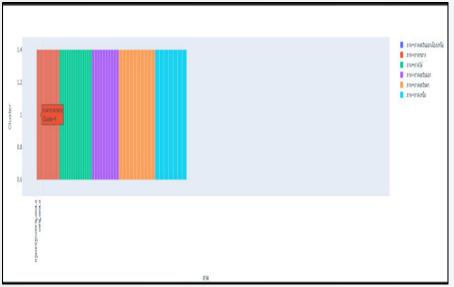
การแบ่งข้อมูลตามจังหวัด



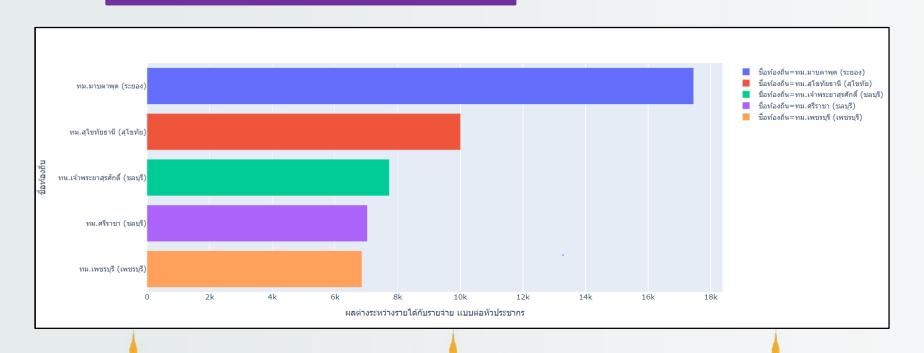


การแบ่งข้อมูลตามภาค

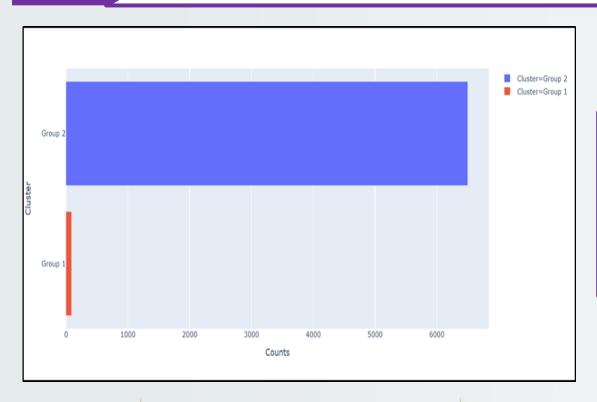




ข้อมูลขององค์กรปกครองส่วนท้องถิ่นที่มีความผิดปกติ







เปรียบเทียบตัวแปร "Cluster" กับ "Counts" สรุปได้ว่าใน Cluster 2 มีจำนวนองค์กร ปกครองส่วนท้องถิ่นทั้งหมด 6,500 แห่ง และใน Cluster 1 มีจำนวนองค์กรปกครอง ส่วนท้องถิ่นทั้งหมด 89 แห่ง ดังนั้นข้อมูล ส่วนใหญ่อยู่ใน Cluster 2

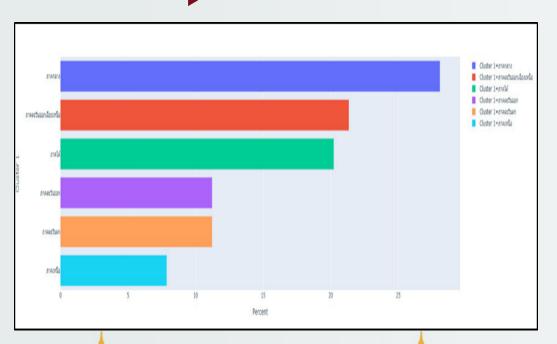
06

Insight and Data Visualization

Cluster 1



คิดเป็น เปอร์เซ็นต์ แบบแบ่งตามภาค



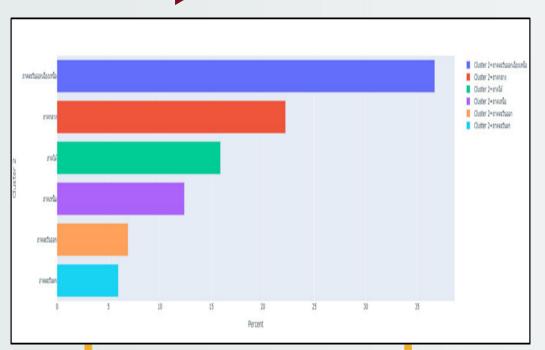
จากภาพ มีทั้งหมด 6 ภาค โดยคิดเป็น ภาคกลาง 28.09% ภาคตะวันออกเฉียงเหนือ 21.35% ภาคใต้ 20.23% ภาคตะวันออก 11.23% ภาคตะวันตก 11.23% ภาคเหนือ 7.87% สรุปได้ว่าเป็น ภาคกลาง มีจำนวนจังหวัด มากที่สุด ใน Cluster 1 06

Insight and Data Visualization

Cluster 2



คิดเป็น เปอร์เซ็นต์ แบบแบ่งตามภาค



จากภาพ มีทั้งหมด 6 ภาค โดยคิดเป็น ภาคตะวันออกเฉียงเหนือ 36.7% ภาคกลาง 22.2% ภาคใต้ 15.88% ภาคเหนือ 12.37% ภาคตะวันออก 6.9% ภาคตะวันตก 5.95% สรุปได้ว่าเป็น ภาคตะวันออกเฉียงเหนือ มีจำนวนจังหวัดมากที่สุด ใน Cluster 2

