



# FINAL ASSIGNMENT

papon choonhaklai 6210503691



# Part I : Feature Selection and Dimensionality Reduction



# Add factors

ปัจจัยที่เพิ่ม	ระดับ จังหวัดหรือ ระดับ ประเทศ	มีข้อมูลครบทุกปี (1995-2018) หรือเป็นข้อมูลที่ทำ regression	หากทำ regression จะบอกราส่วนระหว่างข้อมูลที่มี กับ ข้อมูลที่ทำ regression
จำนวนตำแหน่ง งานว่าง (ความ ต้องการแรงงาน) ของแต่ละจังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (8:16)
จำนวนผู้ที่มาเยี่ยม เยือนในแต่ละ จังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (9:15)
ตัวนี่ราคาผู้บริโภค <sup>ทั่วไป(CPI)</sup> ที่ จำแนกเป็นราย จังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (9:15)
จำนวนคนจน (ด้าน <sup>รายจ่าย</sup> ) จำแนก เป็นรายจังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (8:16)
สถิติโรงงาน อุตสาหกรรมที่จด ทะเบียนกับ กระทรวง อุตสาหกรรม และ <sup>ได้รับอนุญาตให้</sup> ประกอบกิจการ ใหม่ (ตามพระราช บัญญัติโรงงาน พ. ศ. 2535) จำแนก ตามจังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (8:16)
จำนวนเงินให้สิน เชื่อคงเหลือของ ธนาคารออมสิน จำแนกตามจังหวัด	ระดับ จังหวัด	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (8:16)

Households and NPISHs final consumption expenditure (% of GDP)	ระดับ ประเทศ	มีข้อมูลครบทุกปี (1995-2018)	-
Central government debt, total (% of GDP)	ระดับ ประเทศ	เป็นข้อมูลที่ทำ regression	(อัตราส่วนระหว่างข้อมูลที่มี : ข้อมูลที่ทำ regression ) = (21:3)
Exports of goods and services (% of GDP)	ระดับ ประเทศ	มีข้อมูลครบทุกปี (1995-2018)	-
Broad money (% of GDP)	ระดับ ประเทศ	มีข้อมูลครบทุกปี (1995-2018)	-
Manufacturing, value added (% of GDP)	ระดับ ประเทศ	มีข้อมูลครบทุกปี (1995-2018)	-

# Chon Buri

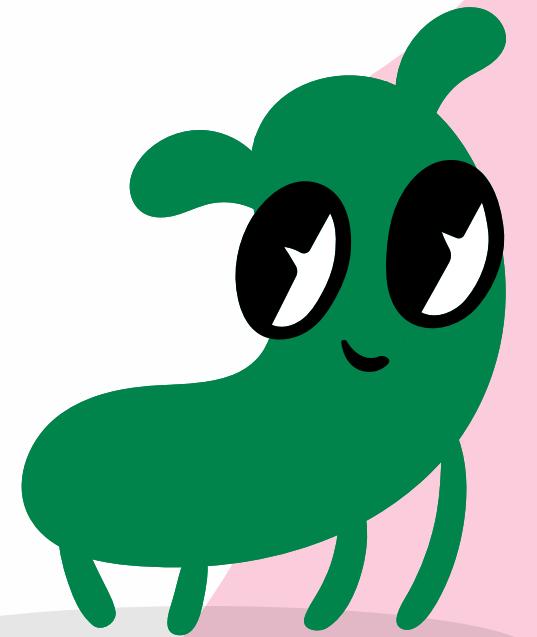


## พื้นที่

• พื้นที่	4,363 ตร.กม. (1,685 ตร.ไมล์)
อันดับพื้นที่	อันดับที่ 50

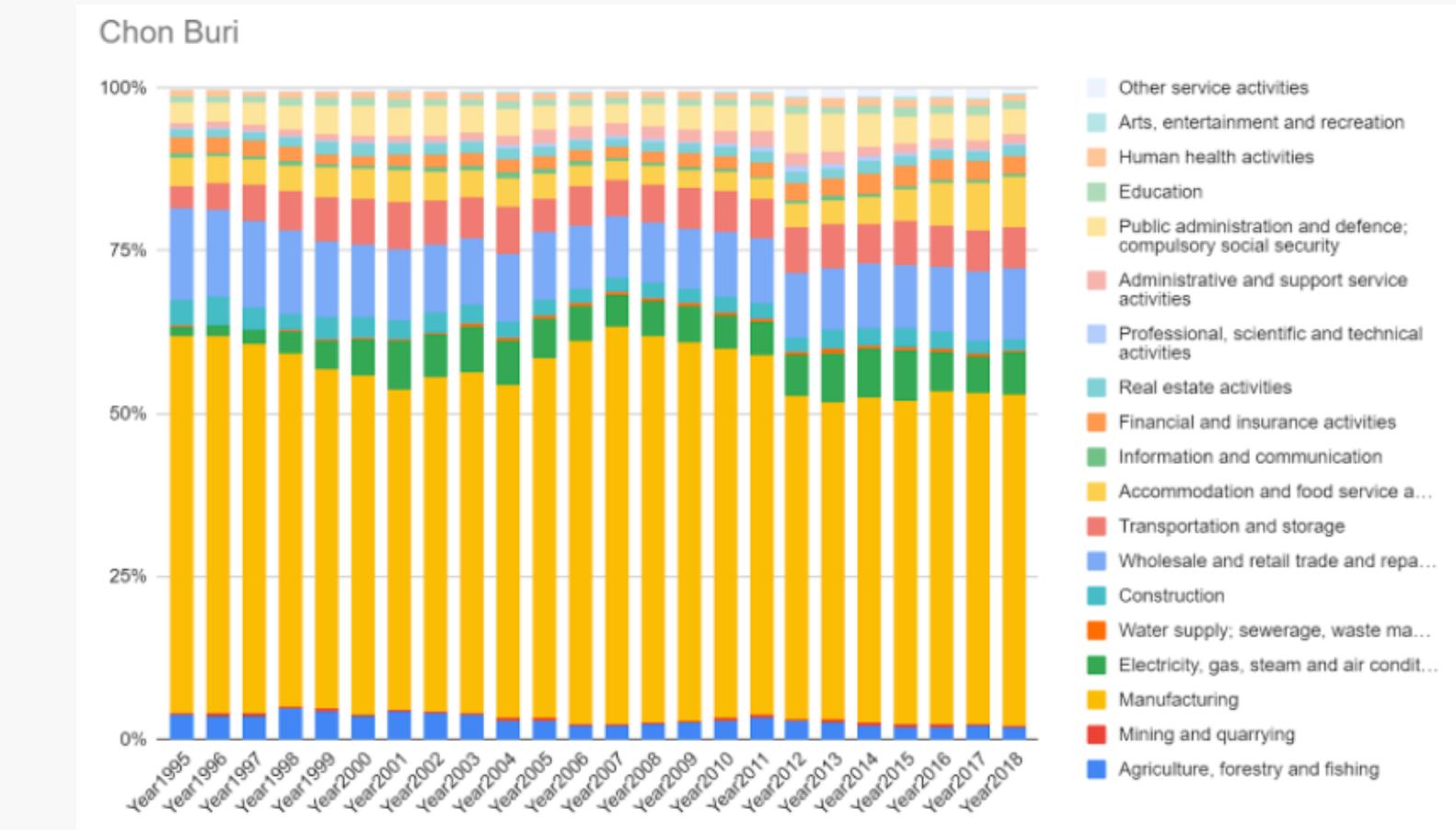
## ประชากร (พ.ศ. 2564)<sup>[1]</sup>

• พื้นที่	1,583,672 คน
• อันดับ	อันดับที่ 7
• ความหนาแน่น	362.98 คน/ตร.กม. (940.1 คน/ตร.ไมล์)
• อันดับความหนาแน่น	อันดับที่ 10

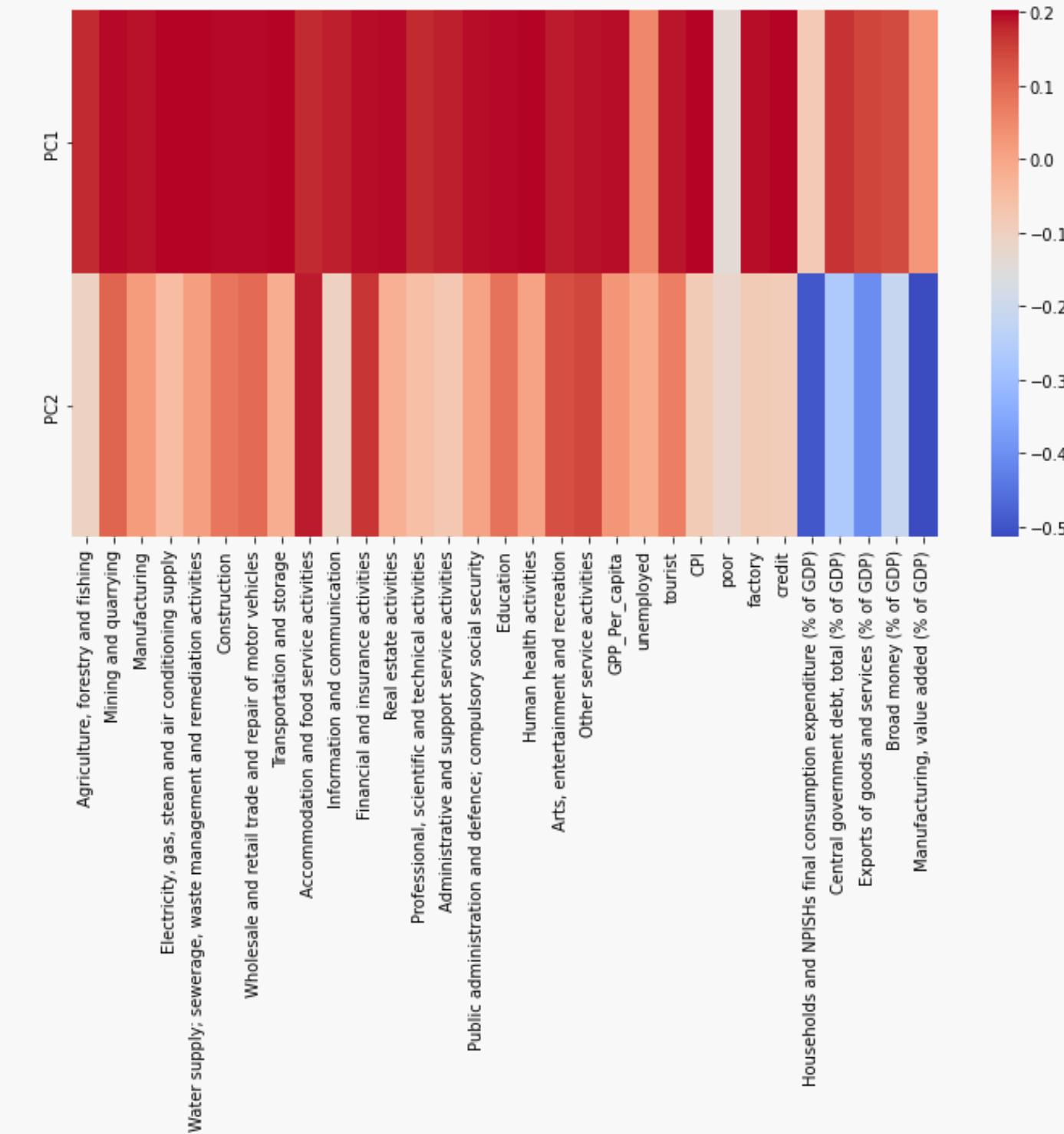
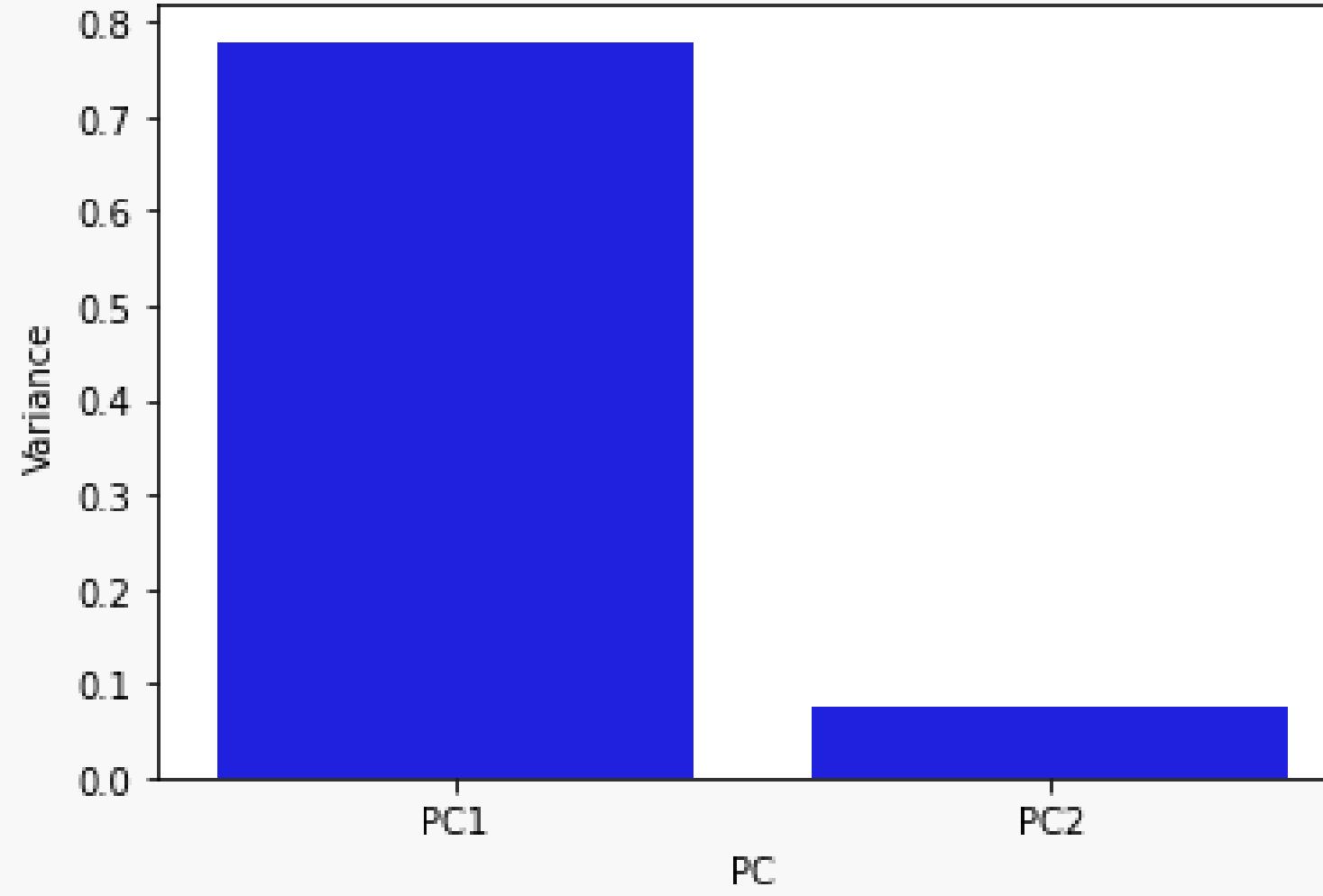


# Feature Selection

Regression			
วิธี	จำนวนปัจจัยที่ส่งผล	R-Squared	RMSE
Low Variance Filtering	21 ปัจจัย	0.6618214839453294	48288.87558969716
High Correlation Filtering	27 ปัจจัย	0.6599497191696188	48422.32684364398
Forward Feature Selection	26 ปัจจัย	0.6441596817120553	49533.80189481629
Recursive Feature Elimination	9 ปัจจัย	0.9162697094168951	24027.876169616535
Random Forest	11 ปัจจัย	0.8306469079419364	34172.01993926983



# Dimensionality Reduction (PCA)



# Prachin Buri



## พื้นที่<sup>[1]</sup>

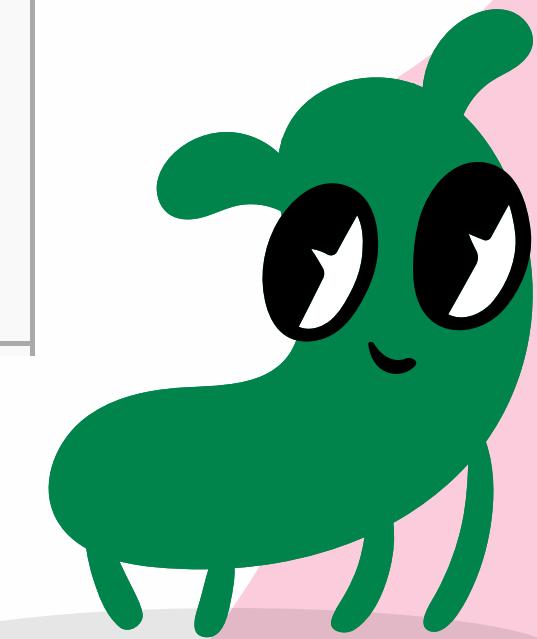
- พื้นที่ 4,762.362 ตร.กม.  
(1,838.758 ตร.ไมล์)

## อันดับพื้นที่

อันดับที่ 44

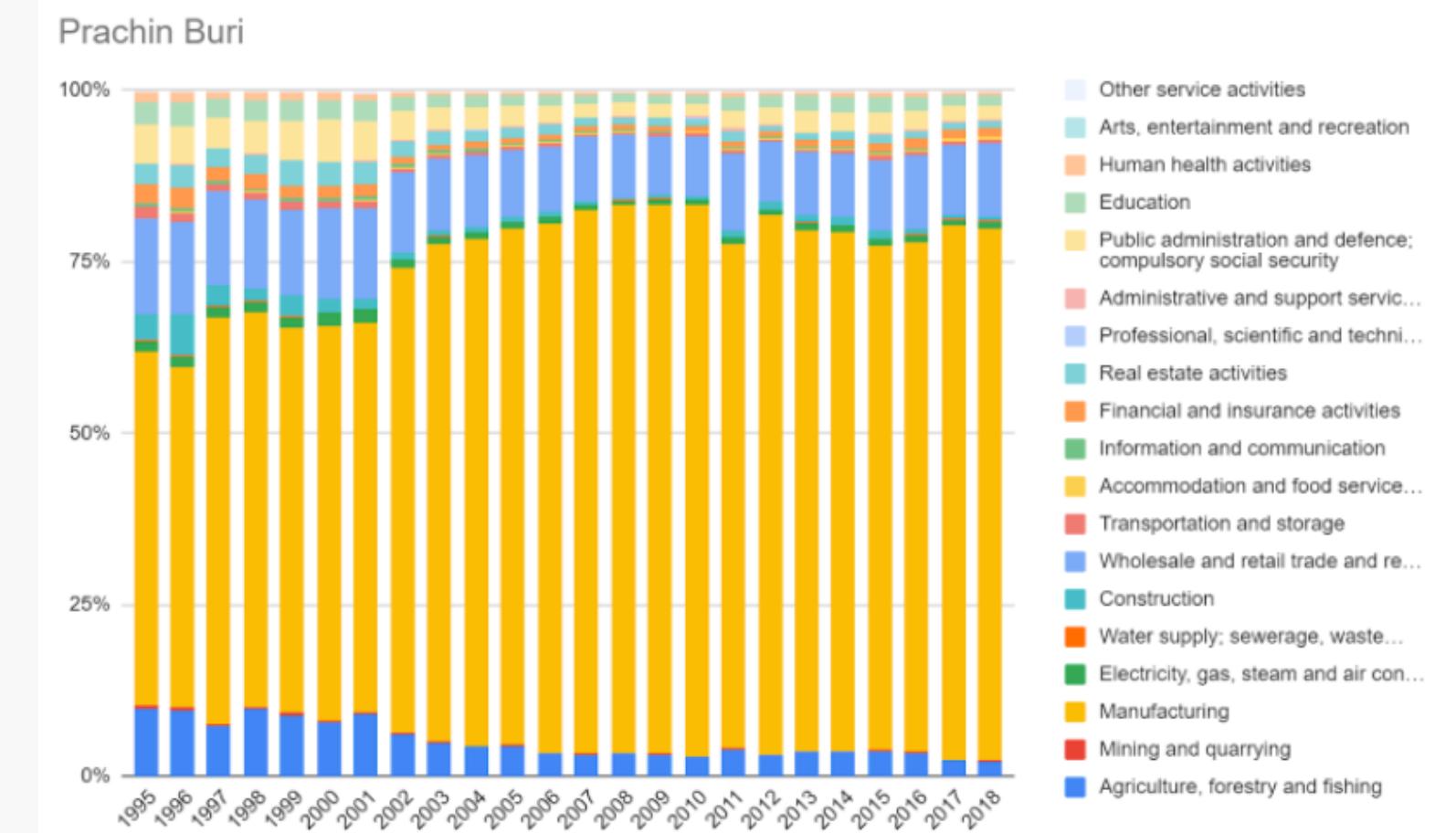
## ประชากร (พ.ศ. 2564)<sup>[2]</sup>

- พื้นที่ 495,325 คน
- อันดับ อันดับที่ 54
- ความหนาแน่น 104.01 คน/ตร.กม. (269.4 คน/ตร.ไมล์)
- อันดับความหนาแน่น อันดับที่ 48

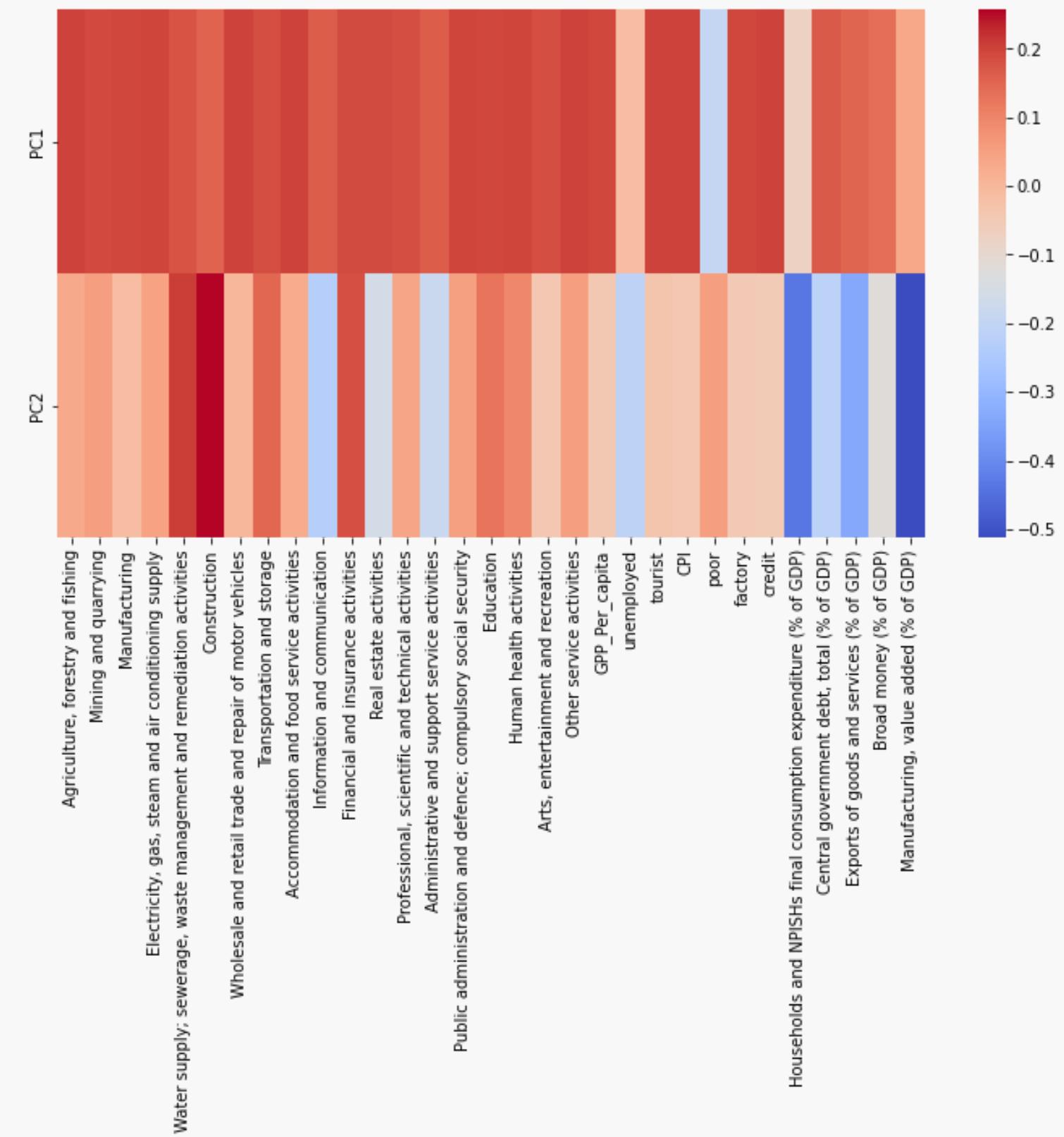
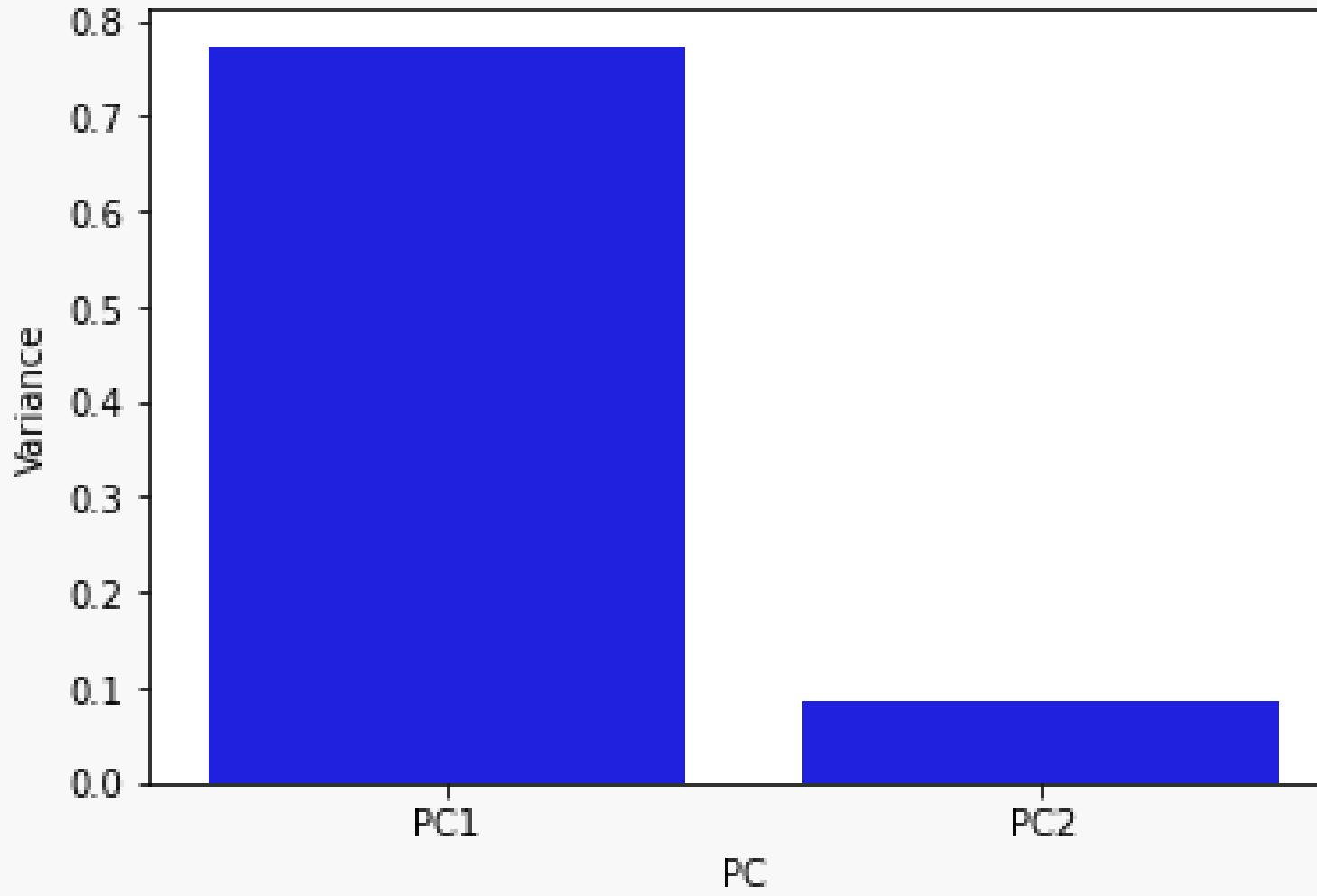


# Feature Selection

Regression			
วิธี	จำนวนปัจจัยที่ส่งผล	R-Squared	RMSE
Low Variance Filtering	8 ปัจจัย	0.9802648240008693	15189.772188098626
High Correlation Filtering	26 ปัจจัย	0.9855828819708079	12982.848389674482
Forward Feature Selection	26 ปัจจัย	0.9856297661959359	12961.72119716623
Recursive Feature Elimination	3 ปัจจัย	0.65669200621565	63353.80689810831
Random Forest	11 ปัจจัย	0.9896792267086372	10984.666542845212



# Dimensionality Reduction (PCA)



# Samut Songkhram

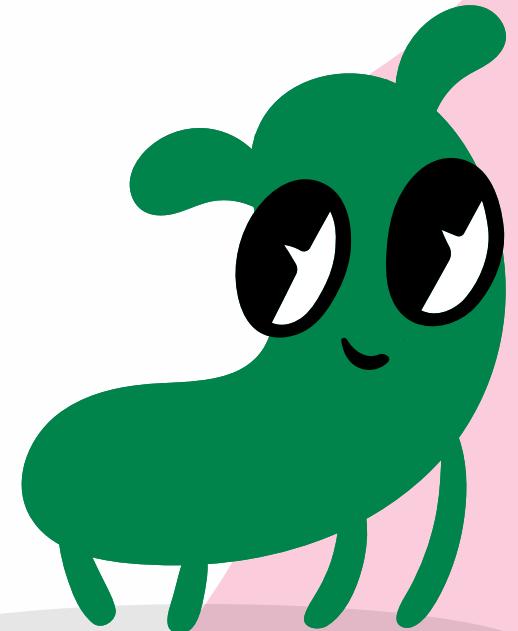


## พื้นที่<sup>[2]</sup>

- ทั้งหมด 416.707 ตร.กม.  
(160.891 ตร.ไมล์)
- อันดับพื้นที่ อันดับที่ 77

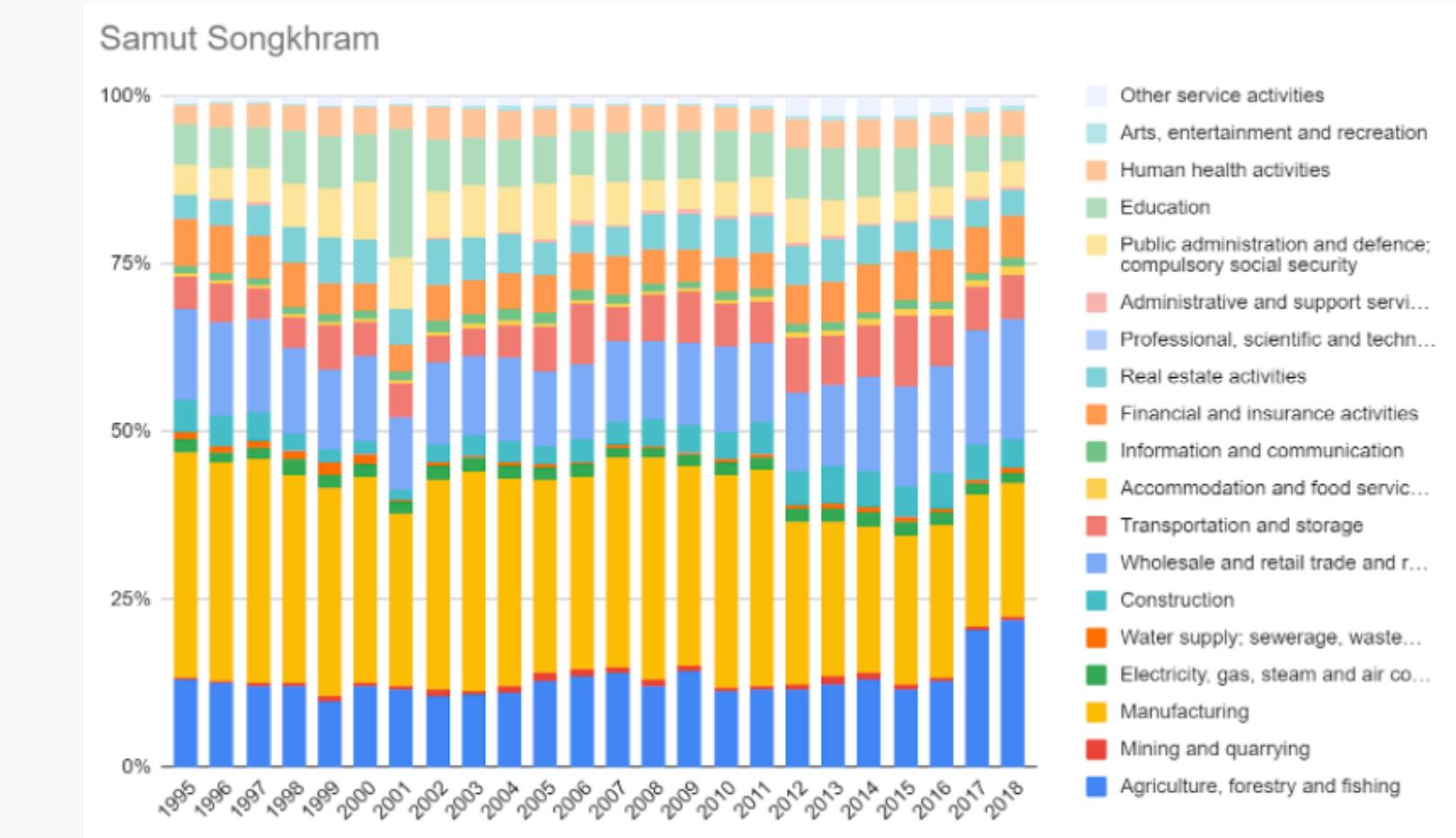
## ประชากร (พ.ศ. 2564)<sup>[3]</sup>

- ทั้งหมด 190,842 คน
- อันดับ อันดับที่ 77
- ความหนาแน่น 457.98 คน/ตร.กม.  
(1,186.2 คน/ตร.ไมล์)
- อันดับความหนาแน่น อันดับที่ 7

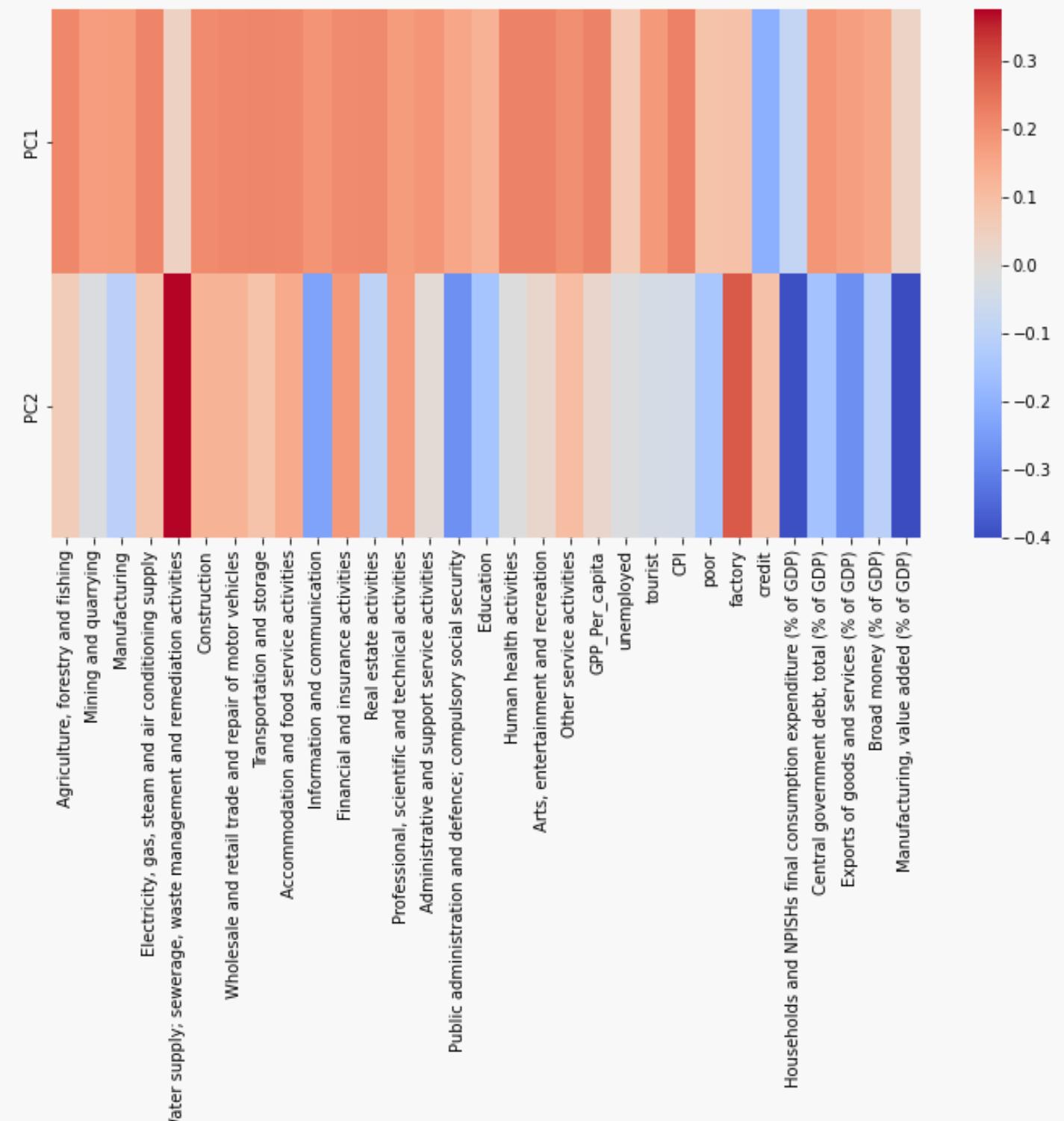
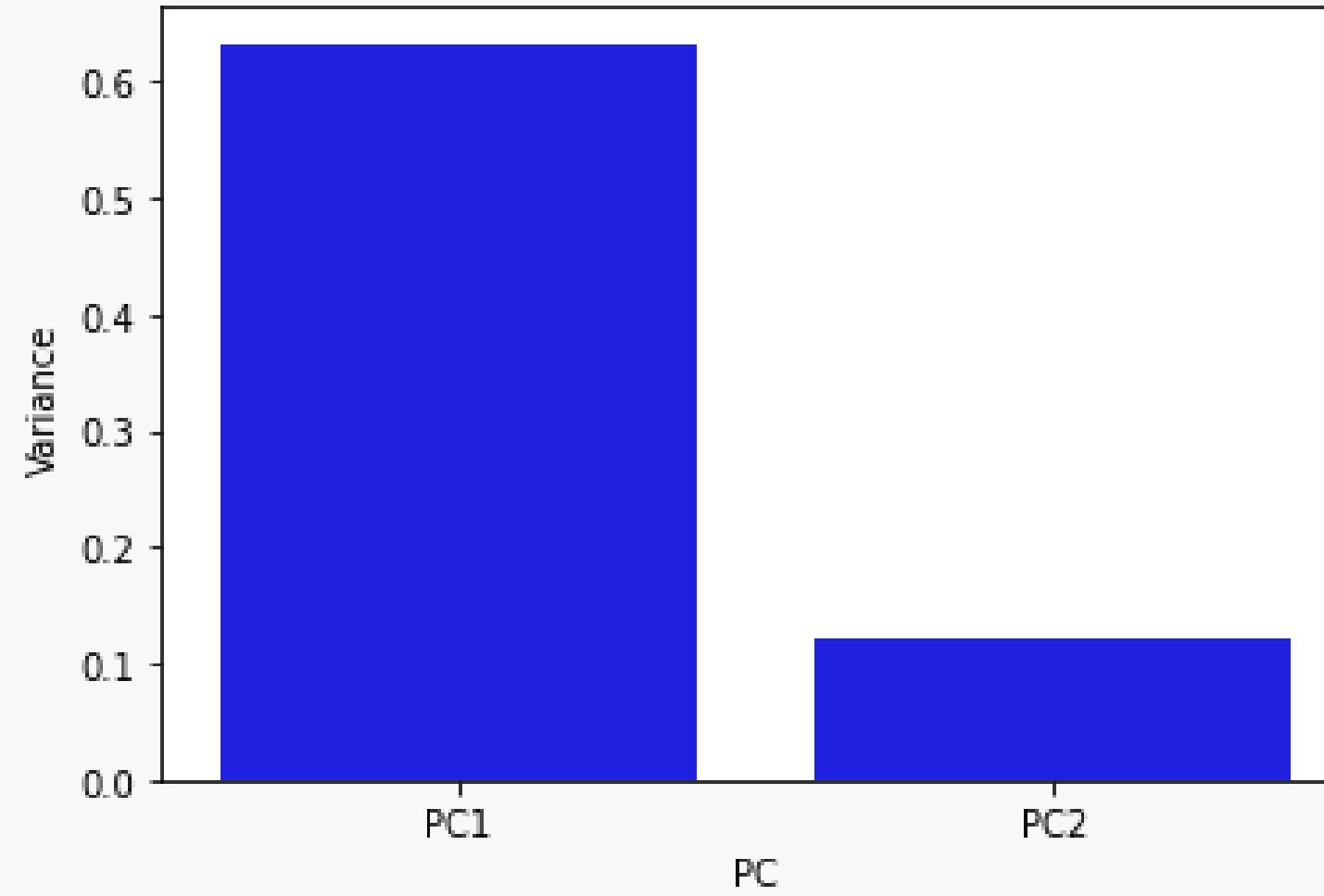


# Feature Selection

Regression			
วิธี	จำนวนปัจจัยที่ส่งผล	R-Squared	RMSE
Low Variance Filtering	9 ปัจจัย	0.8975816176853071	6281.3901774446995
High Correlation Filtering	27 ปัจจัย	0.9833323308050322	2533.9862343352684
Forward Feature Selection	24 ปัจจัย	0.97621239669641	3027.2072773646305
Recursive Feature Elimination	9 ปัจจัย	0.9111585294705711	5850.253489823966
Random Forest	5 ปัจจัย	0.9412506793212084	4757.38435137459



# Dimensionality Reduction (PCA)



# Key Factors for each province

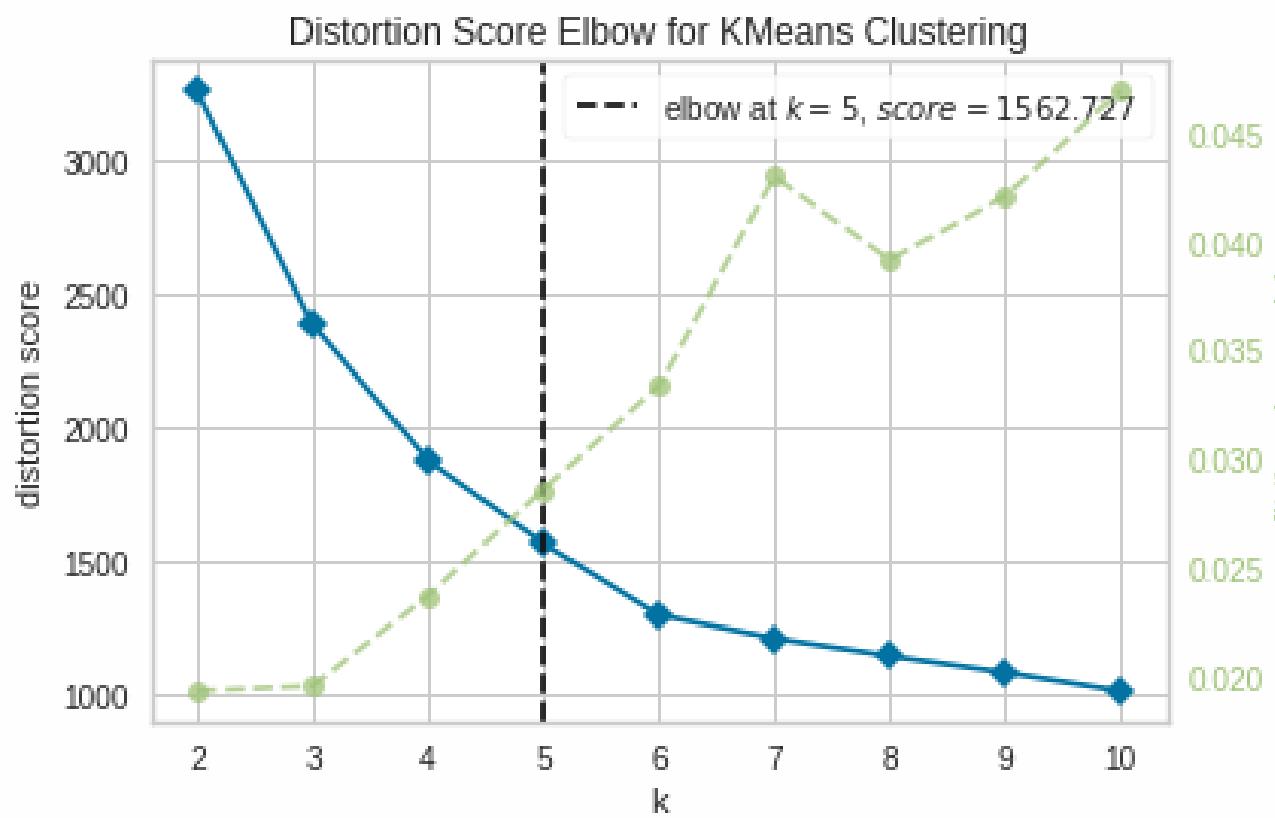
จังหวัดชลบุรี	จังหวัดปราจีนบุรี	จังหวัดสมุทรสงคราม
Manufacturing	Manufacturing	Manufacturing
Wholesale and retail trade and repair of motor vehicles	Wholesale and retail trade and repair of motor vehicles	Wholesale and retail trade and repair of motor vehicles
	Agriculture, forestry and fishing	Agriculture, forestry and fishing
Transportation and storage		
CPI(ตัวชนีราคาผู้บริโภคทั่วไป)	CPI(ตัวชนีราคาผู้บริโภคทั่วไป)	CPI(ตัวชนีราคาผู้บริโภคทั่วไป)
tourist(จำนวนผู้เยี่ยมเยือน )	tourist(จำนวนผู้เยี่ยมเยือน )	tourist(จำนวนผู้เยี่ยมเยือน )
Credit(จำนวนเงินให้สินเชื่อ)	Credit(จำนวนเงินให้สินเชื่อ)	
factory(จำนวนโรงงาน)		
Exports of goods and services (% of GDP)	Exports of goods and services (% of GDP)	Exports of goods and services (% of GDP)
Central government debt, total (% of GDP)	Central government debt, total (% of GDP)	
		Broad money (% of GDP)



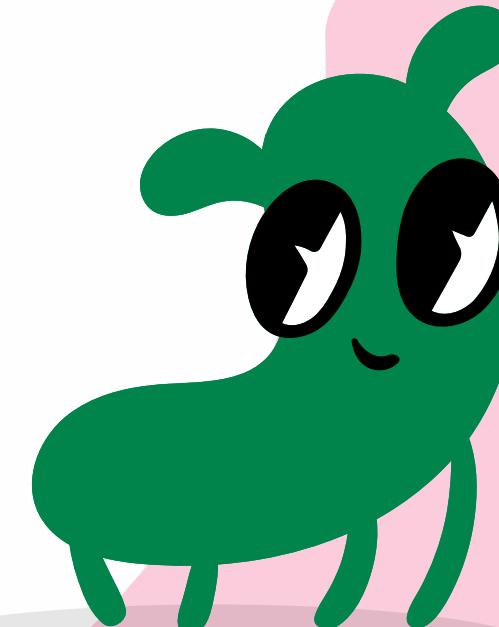
# Part 2 : Clustering



# K - Mean (All Factors)



Cluster_id	จำนวนข้อมูลที่อยู่ใน cluster นั้นๆ
0	4
1	1
2	24
3	24
4	24



# จังหวัดที่ได้รับมอบหมาย

- ชลบุรี อยู่ในกลุ่ม 5 โดยมี 24 จังหวัดอยู่ในกลุ่มนี้
- ปราจีนบุรี อยู่ในกลุ่ม 3 โดยมี 24 จังหวัดอยู่ในกลุ่มนี้
- สมุทรสงคราม อยู่ในกลุ่ม 3 โดยมี 24 จังหวัดอยู่ในกลุ่มนี้



# Factor Selection

Agriculture, forestry and fishing

Manufacturing

Wholesale and retail trade and repair of motor vehicles

Transportation and storage'

tourist

CPI

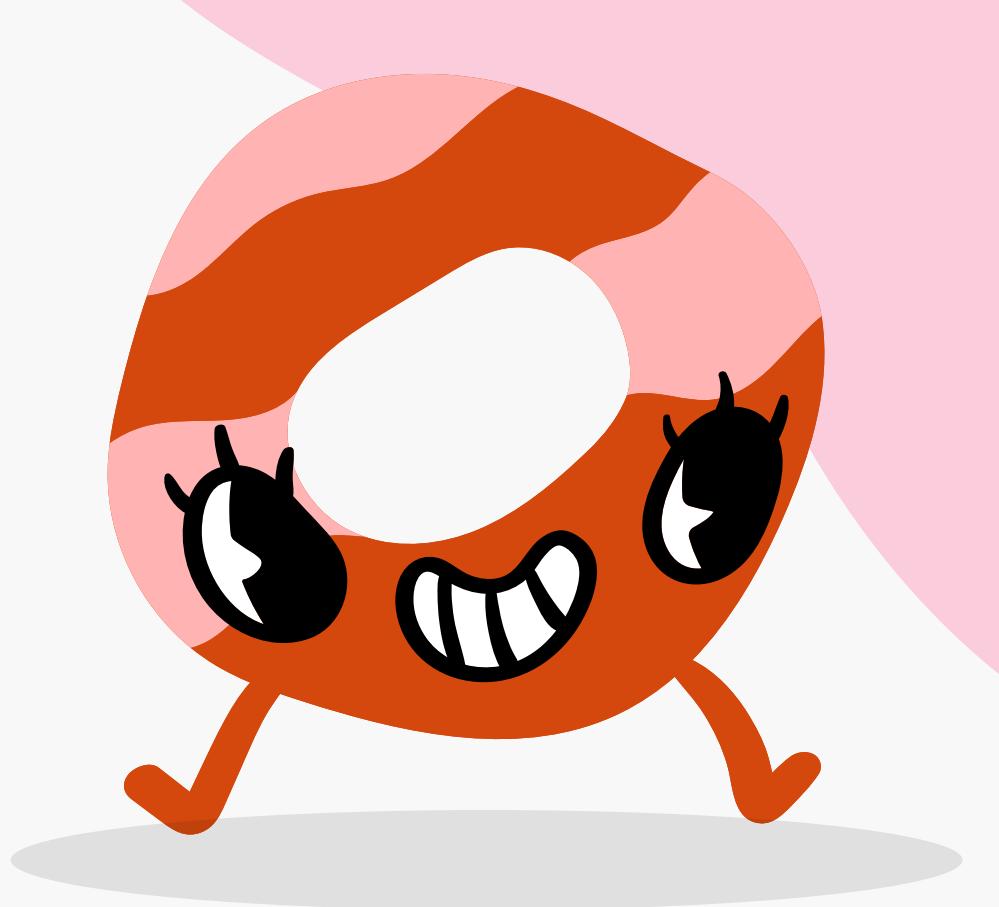
factory

credit

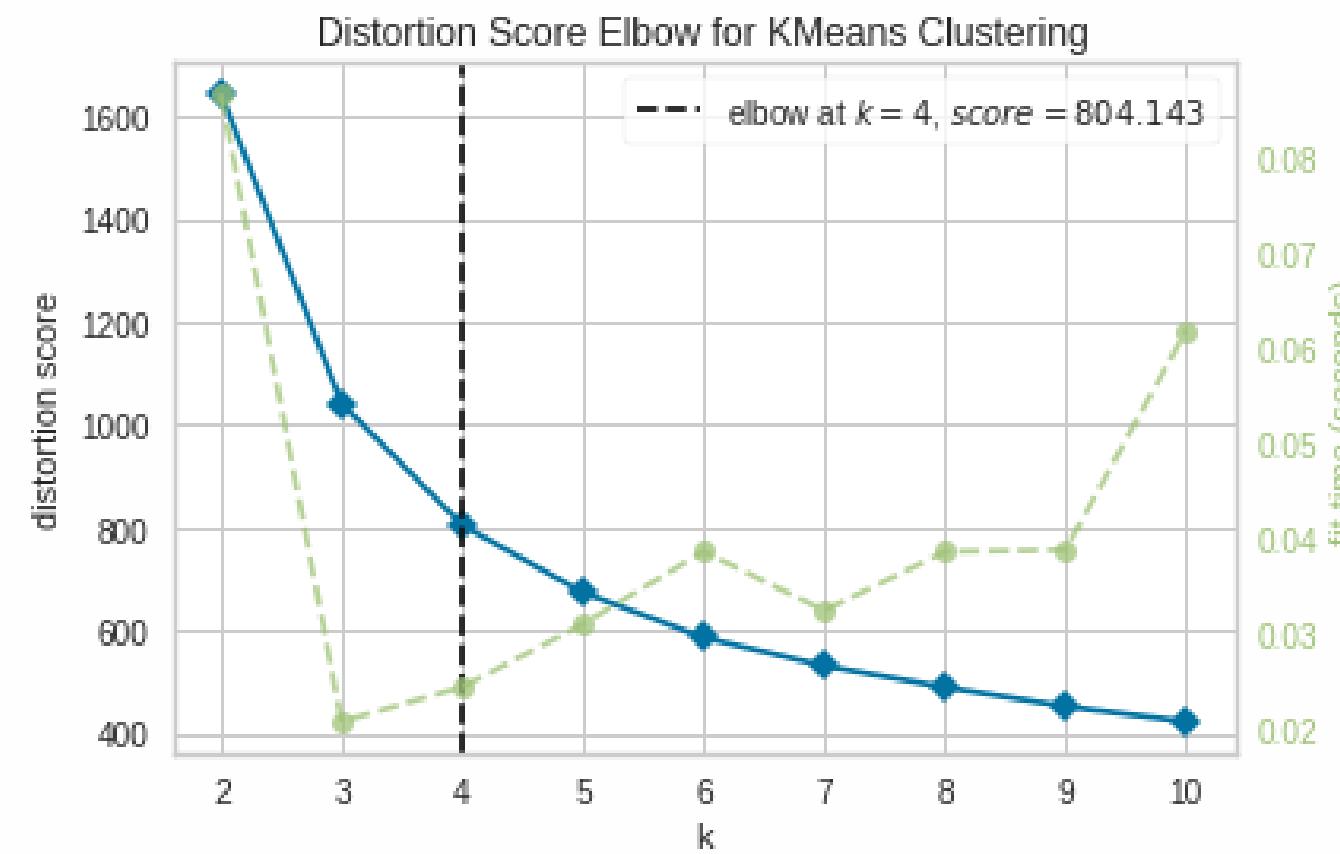
Central government debt, total (% of GDP)

Exports of goods and services (% of GDP)

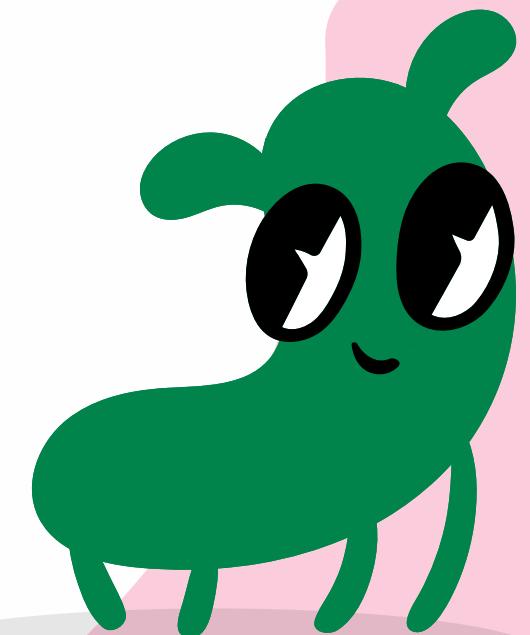
Broad money (% of GDP)



# K - Mean (Some Factors)



Cluster_id	จำนวนข้อมูลที่อยู่ใน cluster นั้นๆ
0	24
1	47
2	1
3	5



# จังหวัดที่ได้รับมอบหมาย

- ชลบุรี อยู่ในกลุ่ม 4 โดยมี 5 จังหวัดอยู่ในกลุ่มนี้
- ปราจีนบุรี อยู่ในกลุ่ม 2 โดยมี 47 จังหวัดอยู่ในกลุ่มนี้
- สมุทรสงคราม อยู่ในกลุ่ม 2 โดยมี 47 จังหวัดอยู่ในกลุ่มนี้



# ALL VS SOME



## Group 5 (All)

ปัจจัย	ค่าเฉลี่ย	อยู่ในช่วง
Agriculture, forestry and fishing	13128.83	กลาง
Mining and quarrying	66586.67	สูง
Manufacturing	376209.3	สูง
Electricity, gas, steam and air conditioning supply	38665.08	สูง
Water supply; sewerage, waste management and remediation activities	3060.417	กลาง
Construction	11101	กลาง
Wholesale and retail trade and repair of motor vehicles	84664.25	กลาง
Transportation and storage	64056.83	กลาง
Accommodation and food service activities	21026.17	กลาง
Information and communication'	3432.75	กลาง
Financial and insurance activities	18747.92	กลาง
Real estate activities	10493.08	กลาง
Professional, scientific and technical activities	8782.833	กลาง

## Group 4 (some)

ปัจจัย	ค่าเฉลี่ย	อยู่ในช่วง
Agriculture, forestry and fishing	11991.73333	กลาง
Manufacturing	344729	กลาง
Wholesale and retail trade and repair of motor vehicles	78746.66667	กลาง
Transportation and storage	53368.86667	กลาง

ปัจจัย	ค่าเฉลี่ย	อยู่ในช่วง
tourist(จำนวนผู้เยี่ยมเยือน )	6628357.333	กลาง
CPI(ดัชนีราคาผู้บริโภคทั่วไป)	101.2933333	กลาง
factory(จำนวนโรงงาน)	258.2666667	สูง
Credit(จำนวนเงินให้สินเชื่อ)	21170.64533	กลาง

ปัจจัย	ค่าเฉลี่ย	อยู่ในช่วง
Central government debt, total (% of GDP)	39.40042048	กลาง
Exports of goods and services (% of GDP)	66.32115258	กลาง
Broad money (% of GDP)	124.1777993	กลาง

# Part 3 : Decision Tree



# Class 3 Classification : 32 country

	country	hdi	y_predicted
8	Australia	3	3
9	Austria	3	3
16	Belgium	3	3
31	Canada	3	3
43	Cyprus	3	3
44	Czech Republic	3	3
46	Denmark	3	3
55	Estonia	3	3
59	Finland	3	3
60	France	3	3
64	Germany	3	3
66	Greece	3	3
76	Iceland	3	3
81	Ireland	3	3
82	Israel	3	3
83	Italy	3	3
85	Japan	3	3
90	Korea, Rep.	3	3
100	Lithuania	3	3
101	Luxembourg	3	3
107	Malta	3	3
122	New Zealand	3	3
127	Norway	3	3
137	Portugal	3	3
152	Singapore	0	3
153	Slovak Republic	3	3
154	Slovenia	2	3
158	Spain	0	3
163	Switzerland	3	3
178	United Kingdom	3	3
179	United States	3	3
180	Uruguay	3	3



- **population** อยู่ในช่วง **4,464,702** คน ถึง **39,057,280** คน
- **GDP** อยู่ในช่วง **87,800** ล้าน ถึง **1.4** ล้านล้าน
- **Education primary** อยู่ในช่วง **100.23** ถึง **103.3**
- **Education secondary** อยู่ในช่วง **102.5** ถึง **120**
- **Education tertiary** อยู่ในช่วง **63.7** ถึง **82.72**
- **mortality infant rate** อยู่ในช่วง **2.25** ถึง **3.7**
- **basic drinking water** อยู่ในช่วง **99.65** ถึง **100**
- **forest land** อยู่ในช่วง **18.36** ถึง **37.74**
- **broadband** อยู่ในช่วง **29.32** ถึง **39.32**
- **mobile** อยู่ในช่วง **116.83** ถึง **135.66**
- **birth rate** อยู่ในช่วง **9.175** ถึง **11.425**





# ASSIGNMENT 2

**Assigned Aspect :** ด้านการศาสนา

วัฒนธรรมและนักนากการ

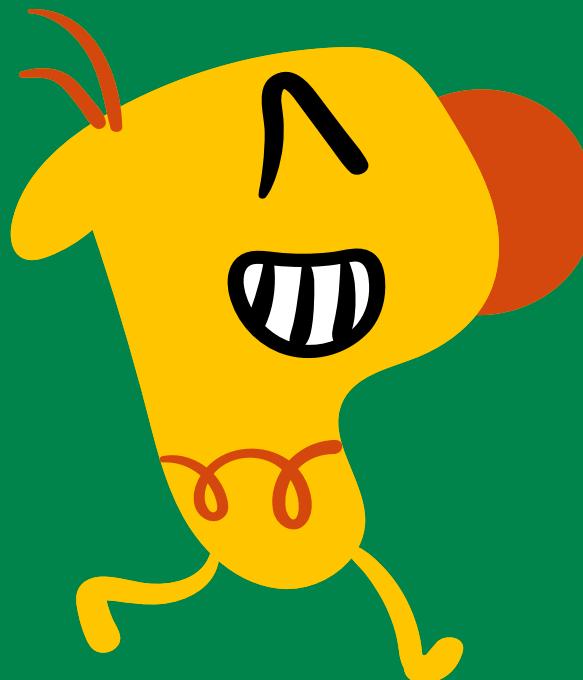
**First Ministry :** กระทรวงศึกษาธิการ



# Clustering Part



# กระทรวง ศึกษาธิการ



# ผลลัพธ์ของวิธี K-mean ในแต่ละปี

ปี 2560

กลุ่ม 2

	กระทรวง-งบรายจ่าย	งบบุคลากร	งบดำเนินงาน	งบลงทุน	งบเงินอุดหนุน	งบรายจ่ายอื่น	รวม	cluster_id
17	กระทรวงศึกษาธิการ	258020.8	36824.8	33810.3	175509.7	9244.4	513410.0	1

ปี 2561

กลุ่ม 6

	กระทรวง-งบรายจ่าย	งบบุคลากร	งบดำเนินงาน	งบลงทุน	งบเงินอุดหนุน	งบรายจ่ายอื่น	รวม	cluster_id
17	กระทรวงศึกษาธิการ	250754.5	33148.8	34287.6	180797.9	8958.9	507947.7	5

ปี 2562

กลุ่ม 5

	กระทรวง-งบรายจ่าย	งบบุคลากร	งบดำเนินงาน	งบลงทุน	งบเงินอุดหนุน	งบรายจ่ายอื่น	รวม	cluster_id
17	กระทรวงศึกษาธิการ	237718.1	28233.1	32548.7	179311.0	9835.5	487646.4	4



# ผลลัพธ์ของวิธี K-mean ในแต่ละปี

ปี 2563

กลุ่ม 3

	กระทรวง-งบรายจ่าย	งบบุคลากร	งบดำเนินงาน	งบลงทุน	งบเงินอุดหนุน	งบรายจ่ายอื่น	รวม	cluster_id
1	กระทรวงกลาโหม	92346.6	24777.8	15433.6	5162.5	95632.9	233353.4	2
17	กระทรวงศึกษาธิการ	230196.3	19847.4	22527.3	89316.7	6772.6	368660.3	2
18	กระทรวงสาธารณสุข	99248.1	14860.6	17465.5	6945.7	209.9	138729.8	2

ปี 2564

กลุ่ม 2

	กระทรวง-งบรายจ่าย	งบบุคลากร	งบดำเนินงาน	งบลงทุน	งบเงินอุดหนุน	งบรายจ่ายอื่น	รวม	cluster_id
1	กระทรวงกลาโหม	91027.4	24102.9	14426.1	4794.4	80179.8	214530.6	1
17	กระทรวงศึกษาธิการ	223199.6	20029.5	19437.0	88628.9	5154.7	356449.7	1
18	กระทรวงสาธารณสุข	107548.0	14411.4	17025.5	6637.5	179.9	145802.3	1



# Regression Part

**Assigned Aspect :** ด้านการศาสนา วัฒนธรรม  
และนันทารากาศ



# Part I.I

	กระบวนการทางภายใน	กระบวนการศึกษาอิเล็กทรอนิกส์	กระบวนการทางภาษาอังกฤษ	กระบวนการห้องเรียนชั้นเดียวและสิ่งแวดล้อม	กระบวนการวัดและประเมินผล
0	102793.9181	190262.8362	4620.8538	14883.2082	2239.6428
1	139800.1791	203752.3867	4658.0021	16150.1673	2480.8597
2	160643.4755	225592.2688	4829.7984	16237.9477	2937.3920
3	179115.7526	282254.3414	5826.1381	17515.4592	4281.0641
4	190930.1453	301085.1582	6152.7260	18819.2856	4371.0958
5	195269.3223	332298.6160	6416.5642	20603.2571	4914.6578
6	187998.7071	346713.0933	6251.6728	20108.5233	4347.8166
7	230781.2686	391131.8796	6899.2261	22936.7930	5119.5841
8	285255.0000	420490.0000	6592.7000	25585.0000	5468.0000
9	308835.0000	460411.7000	7430.3000	30657.8000	5957.1000
10	333145.2000	482788.6000	9171.6000	31487.3000	6624.7000
11	340171.6000	501326.1000	7247.2000	30245.4000	7047.4000
12	341820.8000	517076.7000	7192.6000	35877.9000	7742.3000
13	332808.5000	513410.0000	7034.7000	33721.0000	7386.9000
14	354303.6000	507947.7000	7073.7000	33644.7000	8102.3000
15	371801.7000	487646.4000	6889.2000	30692.6000	8115.3000
16	353007.4000	368660.3000	7553.1000	30370.0000	8569.7000

พิจารณากระทรวงที่มี  
ความสัมพันธ์กับ  
ด้านการศาสนา วัฒนธรรม  
และนับนาทีการ สูงสุด 5  
อันดับแรก



# Part 1.2

พิจารณาเฉพาะ 5 กระบวนการ ได้แก่

- กระบวนการการท่องเที่ยวและกีฬา
- กระบวนการวัฒนธรรม
- กระบวนการศึกษาธิการ
- กระบวนการพัฒนาสังคมและความมั่นคงของมนุษย์
- กระบวนการการต่างประเทศ



# Part 1.1 VS Part 1.2

OLS Regression Results						
Dep. Variable:	การศึกษา วัฒนธรรม และนิยมทางการ	R-squared:	0.980			
Model:	OLS	Adj. R-squared:	0.977			
Method:	Least Squares	F-statistic:	339.2			
Date:	Fri, 04 Mar 2022	Prob (F-statistic):	1.38e-12			
Time:	13:58:39	Log-Likelihood:	-136.94			
No. Observations:	17	AIC:	279.9			
Df Residuals:	14	BIC:	282.4			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
กราฟร่วงมหาดไทย	0.0346	0.004	8.202	0.000	0.026	0.044
กราฟร่วงพาณิชย์	2.2780	0.326	6.992	0.000	1.579	2.977
const	-7152.1137	1393.509	-5.132	0.000	-1.01e+04	-4163.335
Omnibus:	0.218	Durbin-Watson:	1.273			
Prob(Omnibus):	0.897	Jarque-Bera (JB):	0.410			
Skew:	-0.112	Prob(JB):	0.815			
Kurtosis:	2.273	Cond. No.	1.87e+06			

OLS Regression Results						
Dep. Variable:	การศึกษา วัฒนธรรม และนิยมทางการ	R-squared:	0.987			
Model:	OLS	Adj. R-squared:	0.984			
Method:	Least Squares	F-statistic:	327.6			
Date:	Fri, 04 Mar 2022	Prob (F-statistic):	1.71e-12			
Time:	13:59:34	Log-Likelihood:	-133.22			
No. Observations:	17	AIC:	274.4			
Df Residuals:	13	BIC:	277.8			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
กราฟร่วงการท่องเที่ยวและกีฬา	0.7572	0.065	11.687	0.000	0.617	0.897
กราฟร่วงวัฒนธรรม	1.2668	0.264	4.792	0.000	0.696	1.838
กราฟร่วงการต่างประเทศ	1.0600	0.491	2.158	0.050	-0.001	2.121
const	-3247.5256	2385.548	-1.361	0.197	-8401.188	1906.137
Omnibus:	0.312	Durbin-Watson:	1.122			
Prob(Omnibus):	0.855	Jarque-Bera (JB):	0.465			
Skew:	-0.040	Prob(JB):	0.793			
Kurtosis:	2.194	Cond. No.	1.66e+05			

Part 1.1

Part 1.2

THANK YOU!

