



PROJECT FINAL

ชื่อกลุ่ม SANDBOX นะจ๊ะ



รายชื่อสมาชิก

นางสาวชนัญชิตา เมธีกุลมานิต 623020516-1

นางสาวณิชากร ไชยสุวรรณ 623020521-8

นางสาวกิตติมา อุปสุข 623021040-9

นางสาวนฤมล ไชยไสภณ 623021050-6

นางสาววิญญาดา เพ็ญสุข 623021052-2



search

แหล่งที่มาของข้อมูล

- * ย่านชุมชนเก่า แหล่งที่มาของข้อมูล (องค์กร : สำนักงานนโยบายและแผนทรัพยากรธรรมชาติและสิ่งแวดล้อม)
https://data.go.th/dataset/old-community?fbclid=IwAR0TJxEAwT+BKYs853iy-LVEDE09z9wawwxii0BUfdHwwojpU0_Lvw3ciYM
- * ผู้เยี่ยมชมเยือน จำแนกตามรายจังหวัด แหล่งที่มาของข้อมูล (กระทรวงการท่องเที่ยวและกีฬา)
https://data.go.th/dataset/visitors?fbclid=IwAR0mLjK4IuJLWq99uBhG5sm6FGGhicg_iVRzQHtp8b+UjsUbFE0zmGFaRYA
- * รายได้จากการท่องเที่ยว จำแนกตามรายจังหวัด แหล่งที่มาของข้อมูล (กระทรวงการท่องเที่ยวและกีฬา)
<https://data.go.th/dataset/1559?fbclid=IwAR3-zhxvzDyap-vs3T3vdkkwxglu5ye3pwha8okkozK6MQmhyI6KNmYeU9M>
- * จำนวนประชากร จำแนกตามรายจังหวัด ปี 2561 แหล่งที่มาของข้อมูล (ระบบสถิติทางการทะเบียน)
https://stat.bora.dopa.go.th/new_stat/webPage/statByYear.php



Preprocessing



นำเข้า package ที่จะใช้ทำงาน

3

```
import pandas as pd # ทำงานกับข้อมูลลักษณะตาราง
```

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

Mounted at /content/drive

```
ls # ตรวจสอบว่า colab เชื่อมต่อ google drive
```

drive/ sample_data/

```
import os # จัดการไฟล์และ path (ที่อยู่ของไฟล์)
```

```
path = '/content/drive/My Drive/DWDM2021_DATA'
```

นำเข้าข้อมูลชุดที่ 1 ย่านชุมชนเก่า

4

ข้อมูลที่ 1 ย่านชุมชนเก่า

```
data_1 = pd.read_csv(os.path.join(path, 'old-community-area-updated-may-ใหม่ล่าสุด.csv')) # ย่านชุมชนเก่า
data_1
```

	ลำดับ ที่	ชื่อย่านชุมชนเก่า	ตำบล	อำเภอ	จังหวัด	ภาค	East	North	ลักษณะชุมชน	ลักษณะสำคัญ
0	1	นครชุม	นครชุม	เมือง กำแพงเพชร	กำแพงเพชร	กลาง	552752.0317	1822472.131	เป็นชุมชนตลาด	นครชุม เป็นชุมชนตลาดการค้า ที่มีประวัติความเป็น...
1	2	บ้านพรานกระต่าย	ถ้ากระต่าย ทอง	พรานกระต่าย	กำแพงเพชร	กลาง	562744.2062	1842633.828	เป็นชุมชน เกษตรกรรม	บ้านพรานกระต่าย เป็นชุมชน เกษตรกรรมชนบท ที่มีเร...
2	3	บ้านห้วยแม่ซ้าย (บ้านยะฟู)	แม่ยาว	เมืองเชียงราย	เชียงราย	เหนือ	576712.0936	2209887.554	เป็นชุมชน เกษตรกรรม	บ้านห้วยแม่ซ้าย เป็นชุมชน เกษตรกรรม และชุมชนชาด...
3	4	บ้านถ้าผาดอง	ท่าสุด	เมืองเชียงราย	เชียงราย	เหนือ	594024.4964	2220483.832	เป็นชุมชน เกษตรกรรม	บ้านถ้าผาดอง เป็นชุมชน เกษตรกรรม และชุมชนกลุ่มอ...
4	5	บ้านกะเหรี่ยงรวม มิตร	แม่ยาว	เมืองเชียงราย	เชียงราย	เหนือ	574379.3488	2207223.193	เป็นชุมชน เกษตรกรรม	บ้านกะเหรี่ยงรวมมิตร เป็นชุมชน เกษตรกรรม ชุมชนด...
...
608	609	ชุมชนบ้านปะอาว	ปะอาว	เมือง อุบลราชธานี	อุบลราชธานี	ตะวันออกเฉียงเหนือ	470529.6991	1697790.579	เป็นชุมชนชนบท	ชุมชนบ้านปะอาว เป็นชุมชน ชนบทตั้งอยู่บนที่ดอน พ...
609	610	ชุมชนพาณิชยกรรม เก่า	วารินชำราบ	วารินชำราบ	อุบลราชธานี	ตะวันออกเฉียงเหนือ	485281.0265	1680326.802	เป็นชุมชนย่านพาณิช ยกรรมเก่า	ชุมชนพาณิชยกรรมเก่า เป็นย่าน พาณิชยกรรมเก่าแห่ง...
610	611	ชุมชนแก่งสะพือ	พิบูลมังสาหาร	พิบูลมังสาหาร	อุบลราชธานี	ตะวันออกเฉียงเหนือ	525387.1286	1685535.918	เป็นชุมชนเมืองเก่า	ชุมชนแก่งสะพือ เป็นย่านชุมชน เก่าริมแม่น้ำมูล...
611	612	ชุมชนบ้านแคน	ดอนมดแดง	ดอนมดแดง	อุบลราชธานี	ตะวันออกเฉียงเหนือ	503277.5178	1692398.86	เป็นชุมชนชนบท	ชุมชนบ้านแคน เป็นชุมชนชนบท ตั้งอยู่บนที่ดอน พื...
612	613	ชุมชนบ้านชาด	เค็งใหญ่	ห้วยตะพาน	อำนาจเจริญ	ตะวันออกเฉียงเหนือ	420835.9213	1758462.158	เป็นชุมชนเมืองเก่า	ชุมชนบ้านชาด เป็นชุมชนเก่าตั้ง อยู่ใกล้กับแหล่ง...

613 rows × 10 columns

ตัดตารางให้เหลือเฉพาะที่ต้องการ จะได้ตาราง data_city

5

```
data_city = data_1[['ชื่อยานชุมชนเก่า', 'ตำบล', 'อำเภอ', 'จังหวัด', 'ภาค', 'East', 'North', 'ลักษณะชุมชน', 'ลักษณะสำคัญ']] # ยานชุมชนเก่า
data_city
```

	ชื่อยานชุมชนเก่า	ตำบล	อำเภอ	จังหวัด	ภาค	East	North	ลักษณะชุมชน	ลักษณะสำคัญ
0	นครชุม	นครชุม	เมือง กำแพงเพชร	กำแพงเพชร	กลาง	552752.0317	1822472.131	เป็นชุมชนตลาด	นครชุม เป็นชุมชนตลาดการค้า ที่มี ประวัติความเป็น...
1	บ้านพรานกระต่าย	ถ้ากระต่าย ทอง	พรานกระต่าย	กำแพงเพชร	กลาง	562744.2062	1842633.828	เป็นชุมชนเกษตรกรรม	บ้านพรานกระต่าย เป็นชุมชน เกษตรกรรมชนบท ที่มีเร...
2	บ้านห้วยแม่ซ้าย (บ้านยะฟู)	แม่ยาว	เมืองเชียงราย	เชียงราย	เหนือ	576712.0936	2209887.554	เป็นชุมชนเกษตรกรรม	บ้านห้วยแม่ซ้าย เป็นชุมชน เกษตรกรรม และชุมชนชาต...
3	บ้านถ้ำผาตอง	ท่าสุด	เมืองเชียงราย	เชียงราย	เหนือ	594024.4964	2220483.832	เป็นชุมชนเกษตรกรรม	บ้านถ้ำผาตอง เป็นชุมชนเกษตรกรรม และชุมชนกลุ่มอ...
4	บ้านกะเหรี่ยงรวมมิตร	แม่ยาว	เมืองเชียงราย	เชียงราย	เหนือ	574379.3488	2207223.193	เป็นชุมชนเกษตรกรรม	บ้านกะเหรี่ยงรวมมิตร เป็นชุมชน เกษตรกรรม ชุมชนต...
...
608	ชุมชนบ้านปะอาว	ปะอาว	เมือง อุบลราชธานี	อุบลราชธานี	ตะวันออกเฉียงเหนือ	470529.6991	1697790.579	เป็นชุมชนชนบท	ชุมชนบ้านปะอาว เป็นชุมชนชนบทตั้ง อยู่บนที่ดอน พ...
609	ชุมชนพาณิชยกรรม เก่า	วารินชำราบ	วารินชำราบ	อุบลราชธานี	ตะวันออกเฉียงเหนือ	485281.0265	1680326.802	เป็นชุมชนย่านพาณิช ยกรรมเก่า	ชุมชนพาณิชยกรรมเก่า เป็นย่าน พาณิชยกรรมเก่าแห่ง...
610	ชุมชนแก่งสะพือ	พิบูลมังสาหาร	พิบูลมังสาหาร	อุบลราชธานี	ตะวันออกเฉียงเหนือ	525387.1286	1685535.918	เป็นชุมชนเมืองเก่า	ชุมชนแก่งสะพือ เป็นย่านชุมชนเก่าริม น้ำแม่น้ำมูล...
611	ชุมชนบ้านแคน	ดอนมดแดง	ดอนมดแดง	อุบลราชธานี	ตะวันออกเฉียงเหนือ	503277.5178	1692398.86	เป็นชุมชนชนบท	ชุมชนบ้านแคน เป็นชุมชนชนบทตั้ง อยู่บนที่ดอน พื...
612	ชุมชนบ้านชาด	เค็งใหญ่	ห้วยตะพาน	อำนาจเจริญ	ตะวันออกเฉียงเหนือ	420835.9213	1758462.158	เป็นชุมชนเมืองเก่า	ชุมชนบ้านชาด เป็นชุมชนเก่าตั้งอยู่ ใกล้กับแหล่ง...

613 rows × 9 columns

ตรวจสอบค่า missing ของตารางข้อมูล data_city

```
data_city.isnull().any()
```

ชื่อยานชุมชนเก่า	False
ตำบล	False
อำเภอ	False
จังหวัด	False
ภาค	False
East	False
North	False
ลักษณะชุมชน	False
ลักษณะสำคัญ	False
dtype: bool	

แสดงผลเป็น False ทั้งหมด แปลว่า ไม่มีค่า missing ในข้อมูล ย่านชุมชนเก่า

ข้อมูลี่ 2 ผู้เยี่ยมชมเยือน จำแนกตามจังหวัด

ข้อมูลี่ 2 ผู้เยี่ยมชมเยือน จำแนกตามจังหวัด

```
data_2= pd.read_csv(os.path.join(path,'visitors.csv')) # ผู้เยี่ยมชมเยือน จำแนกตามจังหวัด
data_2
```

	จังหวัด	รหัส จังหวัด	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560
0	กรุงเทพมหานคร	th-bm	36172138.0	35953546.0	35110693.0	30037911.0	38222903.0	43763002	47185031	50568902	50972772	56515597	59196331	63571
1	สมุทรปราการ	th-sp	958352.0	1262082.0	1875232.0	1292295.0	1291883.0	1489104	2107433	2204073	2350340	2780429	2900180	32189
2	นนทบุรี	th-no	830593.0	1079877.0	973721.0	1373459.0	1792164.0	2004376	1663990	1700408	1730885	2045867	2090095	25146
3	ปทุมธานี	th-pt	868449.0	1050674.0	798231.0	965318.0	1068529.0	1086433	1474292	1553170	1613073	1815389	1896549	20142
4	พระนครศรีอยุธยา	th-pa	3373929.0	3784617.0	3659402.0	3583231.0	6534638.0	4890949	6066545	6226424	6698561	6994538	7216514	76315
...
72	ตรัง	th-tg	726373.0	858638.0	993615.0	843606.0	1110021.0	1007717	1137961	1281663	1308968	1414663	1465811	15140
73	พัทลุง	th-pl	448000.0	501320.0	853674.0	483058.0	939379.0	1119171	1089992	1178007	1236012	1326468	1487188	15693
74	ปัตตานี	th-pi	160320.0	131998.0	152748.0	201401.0	184153.0	179687	214285	235606	237925	251081	257143	27273
75	ยะลา	th-yl	282463.0	269368.0	311712.0	300570.0	373269.0	335968	429044	563771	589653	622376	624879	65733
76	นราธิวาส	th-nw	336473.0	284073.0	328729.0	430301.0	547235.0	479524	489944	570973	589690	635910	646020	66573

77 rows × 16 columns

ตัดตารางข้อมูลของ data_2 เลือกเฉพาะข้อมูลปี 2555-2561

8

```
data_visitor = data_2[['จังหวัด', 'รหัสจังหวัด', '2555', '2556', '2557', '2558', '2559', '2560', '2561']] # ผู้เยี่ยมชม เดือน จำแนกตามจังหวัด
data_visitor
```

	จังหวัด	รหัสจังหวัด	2555	2556	2557	2558	2559	2560	2561
0	กรุงเทพมหานคร	th-bm	47185031	50568902	50972772	56515597	59196331	63575737	65534281
1	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639
2	นนทบุรี	th-no	1663990	1700408	1730885	2045867	2090095	2514677	2664025
3	ปทุมธานี	th-pt	1474292	1553170	1613073	1815389	1896549	2014281	2098344
4	พระนครศรีอยุธยา	th-pa	6066545	6226424	6698561	6994538	7216514	7631557	8349613
...
72	ตรัง	th-tg	1137961	1281663	1308968	1414663	1465811	1514092	1578906
73	พัทลุง	th-pl	1089992	1178007	1236012	1326468	1487188	1569333	1641841
74	ปัตตานี	th-pi	214285	235606	237925	251081	257143	272739	281988
75	ยะลา	th-yl	429044	563771	589653	622376	624879	657317	680018
76	นราธิวาส	th-nw	489944	570973	589690	635910	646020	665710	692057

77 rows × 9 columns

ทำการเปลี่ยนชื่อ column

เนื่องจากเมื่อเรานำไปเชื่อมตารางกันจะมี column ที่ซ้ำกัน

```
data_visitor.rename(columns={'2555': 'visitor2555'}, inplace=True)
data_visitor.rename(columns={'2556': 'visitor2556'}, inplace=True)
data_visitor.rename(columns={'2557': 'visitor2557'}, inplace=True)
data_visitor.rename(columns={'2558': 'visitor2558'}, inplace=True)
data_visitor.rename(columns={'2559': 'visitor2559'}, inplace=True)
data_visitor.rename(columns={'2560': 'visitor2560'}, inplace=True)
data_visitor.rename(columns={'2561': 'visitor2561'}, inplace=True)
```

จะได้ชื่อ column ดังรูป

data_visitor									
	จังหวัด	รหัสจังหวัด	visitor2555	visitor2556	visitor2557	visitor2558	visitor2559	visitor2560	visitor2561
0	กรุงเทพมหานคร	th-bm	47185031	50568902	50972772	56515597	59196331	63575737	65534281
1	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639
2	นนทบุรี	th-no	1663990	1700408	1730885	2045867	2090095	2514677	2664025
3	ปทุมธานี	th-pt	1474292	1553170	1613073	1815389	1896549	2014281	2098344
4	พระนครศรีอยุธยา	th-pa	6066545	6226424	6698561	6994538	7216514	7631557	8349613
...
72	ตรัง	th-tg	1137961	1281663	1308968	1414663	1465811	1514092	1578906
73	พัทลุง	th-pl	1089992	1178007	1236012	1326468	1487188	1569333	1641841
74	ปัตตานี	th-pi	214285	235606	237925	251081	257143	272739	281988
75	ยะลา	th-yl	429044	563771	589653	622376	624879	657317	680018
76	นราธิวาส	th-nw	489944	570973	589690	635910	646020	665710	692057

77 rows × 9 columns

ตรวจสอบค่า missing ของตารางข้อมูล data_visitor

```
data_visitor.isnull().any()
```

จังหวัด	False
รหัสจังหวัด	False
visitor2555	False
visitor2556	False
visitor2557	False
visitor2558	False
visitor2559	False
visitor2560	False
visitor2561	False
dtype:	bool

แสดงผลเป็น False ทั้งหมด แปลว่า ไม่มีค่า missing ในข้อมูล ผู้เยี่ยมชมเยือน

ข้อมูลี่ 3 รายได้จากการท่องเที่ยว จำแนกตามจังหวัด

ข้อมูลี่ 3 รายได้จากการท่องเที่ยว จำแนกตามจังหวัด

```
data_3= pd.read_csv(os.path.join(path, 'revenue.csv')) # รายได้จากการท่องเที่ยว จำแนกตามจังหวัด
data_3
```

	จังหวัด	รหัสจังหวัด	2555	2556	2557	2558	2559	2560	2561
0	กรุงเทพมหานคร	th-bm	570327.58	626535.05	634803.16	730318.31	822454.21	947946.30000	1051118.51
1	กาญจนบุรี	th-kn	11910.00	13417.00	14576.27	17659.54	20404.73	24440.77288	27138.35
2	จันทบุรี	th-ct	4214.00	4569.00	4673.31	5239.98	5630.03	7628.47000	8469.45
3	ฉะเชิงเทรา	th-cc	3571.69	3735.06	3781.62	3984.33	4240.41	4655.55000	5007.28
4	ชลบุรี	th-cb	100537.00	111109.00	105920.27	126030.24	187788.42	240610.45000	272435.77
...
72	ระนอง	th-rn	2780.00	3250.00	3272.34	3623.68	3954.75	4311.07000	4706.13
73	สงขลา	th-sg	26702.00	37276.00	41304.80	47260.67	53493.02	59831.77000	68097.42
74	สตูล	th-sa	5332.00	5905.00	6162.46	6846.72	7438.85	8168.70000	9101.47
75	สุราษฎร์ธานี	th-st	35312.00	51550.00	56952.27	65420.63	76546.39	92228.01000	104503.71
76	บึงกาฬ	th-bk	0.00	0.00	685.02	806.10	900.50	976.61000	1047.13

77 rows × 9 columns

ตัดตารางข้อมูลของ data_3 เลือกเฉพาะข้อมูลปี 2555-2561

12

```
data_income = data_3[['รหัสจังหวัด', '2555', '2556', '2557', '2558', '2559', '2560', '2561']] #รายได้จากการท่องเที่ยว จำแนกตามจังหวัด
data_income
```

	รหัสจังหวัด	2555	2556	2557	2558	2559	2560	2561
0	th-bm	570327.58	626535.05	634803.16	730318.31	822454.21	947946.30000	1051118.51
1	th-kn	11910.00	13417.00	14576.27	17659.54	20404.73	24440.77288	27138.35
2	th-ct	4214.00	4569.00	4673.31	5239.98	5630.03	7628.47000	8469.45
3	th-cc	3571.69	3735.06	3781.62	3984.33	4240.41	4655.55000	5007.28
4	th-cb	100537.00	111109.00	105920.27	126030.24	187788.42	240610.45000	272435.77
...
72	th-rn	2780.00	3250.00	3272.34	3623.68	3954.75	4311.07000	4706.13
73	th-sg	26702.00	37276.00	41304.80	47260.67	53493.02	59831.77000	68097.42
74	th-sa	5332.00	5905.00	6162.46	6846.72	7438.85	8168.70000	9101.47
75	th-st	35312.00	51550.00	56952.27	65420.63	76546.39	92228.01000	104503.71
76	th-bk	0.00	0.00	685.02	806.10	900.50	976.61000	1047.13

77 rows × 8 columns

ทำการเปลี่ยนชื่อ column

```
data_income.rename(columns={'2555': 'income2555'}, inplace=True)
data_income.rename(columns={'2556': 'income2556'}, inplace=True)
data_income.rename(columns={'2557': 'income2557'}, inplace=True)
data_income.rename(columns={'2558': 'income2558'}, inplace=True)
data_income.rename(columns={'2559': 'income2559'}, inplace=True)
data_income.rename(columns={'2560': 'income2560'}, inplace=True)
data_income.rename(columns={'2561': 'income2561'}, inplace=True)
```


จะได้ชื่อ column ดังรูป

data_income

	รหัสจังหวัด	income2555	income2556	income2557	income2558	income2559	income2560	income2561
0	th-bm	570327.58	626535.05	634803.16	730318.31	822454.21	947946.30000	1051118.51
1	th-kn	11910.00	13417.00	14576.27	17659.54	20404.73	24440.77288	27138.35
2	th-ct	4214.00	4569.00	4673.31	5239.98	5630.03	7628.47000	8469.45
3	th-cc	3571.69	3735.06	3781.62	3984.33	4240.41	4655.55000	5007.28
4	th-cb	100537.00	111109.00	105920.27	126030.24	187788.42	240610.45000	272435.77
...
72	th-rn	2780.00	3250.00	3272.34	3623.68	3954.75	4311.07000	4706.13
73	th-sg	26702.00	37276.00	41304.80	47260.67	53493.02	59831.77000	68097.42
74	th-sa	5332.00	5905.00	6162.46	6846.72	7438.85	8168.70000	9101.47
75	th-st	35312.00	51550.00	56952.27	65420.63	76546.39	92228.01000	104503.71
76	th-bk	0.00	0.00	685.02	806.10	900.50	976.61000	1047.13

77 rows × 8 columns

ตรวจสอบค่า missing ของตารางข้อมูล data_income

```
data_income.isnull().any()
```

รหัสจังหวัด	False
income2555	False
income2556	False
income2557	False
income2558	False
income2559	False
income2560	False
income2561	False
dtype:	bool

แสดงผลเป็น False ทั้งหมด แปลว่า ไม่มีค่า missing ในข้อมูล รายได้จากการท่องเที่ยว

รวม 2 ตารางเข้าด้วยกัน .merge()

เชื่อมตาราง data_visitor กับตาราง data_income เข้าด้วยกันด้วย 'รหัสจังหวัด'

```
merged_data_visitor_data_income = data_visitor.merge(data_income,how='left',left_on='รหัสจังหวัด',right_on='รหัสจังหวัด')
merged_data_visitor_data_income
```

	จังหวัด	รหัส จังหวัด	visitor2555	visitor2556	visitor2557	visitor2558	visitor2559	visitor2560	visitor2561	income2555	income2556	income2557
0	กรุงเทพมหานคร	th-bm	47185031	50568902	50972772	56515597	59196331	63575737	65534281	570327.58	626535.05	634803.1
1	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639	3451.66	3768.56	4044.50
2	นนทบุรี	th-no	1663990	1700408	1730885	2045867	2090095	2514677	2664025	2425.62	2503.18	2560.31
3	ปทุมธานี	th-pt	1474292	1553170	1613073	1815389	1896549	2014281	2098344	1979.86	2199.57	2323.12
4	พระนครศรีอยุธยา	th-pa	6066545	6226424	6698561	6994538	7216514	7631557	8349613	11569.24	12235.03	13446.66
...
72	ตรัง	th-tg	1137961	1281663	1308968	1414663	1465811	1514092	1578906	6231.00	6679.00	6806.73
73	พัทลุง	th-pl	1089992	1178007	1236012	1326468	1487188	1569333	1641841	1780.00	2146.00	2206.35
74	ปัตตานี	th-pi	214285	235606	237925	251081	257143	272739	281988	743.00	778.00	774.61
75	ยะลา	th-yl	429044	563771	589653	622376	624879	657317	680018	1759.00	2420.00	2494.90
76	นราธิวาส	th-nw	489944	570973	589690	635910	646020	665710	692057	2139.00	2208.00	2323.54

77 rows × 16 columns

ตรวจสอบค่า missing

```
merged_data_visitor_data_income.isnull().any() # ตรวจสอบค่า missing
```

```
จังหวัด      False
รหัสจังหวัด  False
visitor2555   False
visitor2556   False
visitor2557   False
visitor2558   False
visitor2559   False
visitor2560   False
visitor2561   False
income2555    False
income2556    False
income2557    False
income2558    False
income2559    False
income2560    False
income2561    False
dtype: bool
```

แสดงผลเป็น false ทั้งหมด หมายความว่า ไม่มีค่า missing ในข้อมูล จากการรวมตารางของ รายได้ และ ผู้เยี่ยมชมเยือนเข้าด้วยกัน

17

[illegible]

ตรวจสอบค่า missing

18

```
data_mix.isnull().any() # ตรวจสอบค่า missing
```

จังหวัด	False
รหัสจังหวัด	False
visitor2555	False
visitor2556	False
visitor2557	False
visitor2558	False
visitor2559	False
visitor2560	False
visitor2561	False
income2555	False
income2556	False
income2557	False
income2558	False
income2559	False
income2560	False
income2561	False
ชื่อยานชุมชนเก่า	True
ตำบล	True
อำเภอ	True
ภาค	True
East	True
North	True
ลักษณะชุมชน	True
ลักษณะสำคัญ	True
dtype:	bool

ในการแสดงผล พบว่ามีข้อมูล ชื่อยานชุมชนเก่า, ตำบล, อำเภอ, ภาค, East, North, ลักษณะชุมชน, ลักษณะสำคัญ
ที่ยังมีค่า missing อยู่ จึงต้องทำการจัดการค่า missing ก่อน

จัดการค่า missing ด้วยวิธี handling missing
(แทนด้วย class ใหม่ (unknow)) โดยใช้คำสั่ง .fillna()

```
data_mix[['ชื่อยานชุมชนเก่า']] = data_mix[['ชื่อยานชุมชนเก่า']].fillna('ไม่มีข้อมูล')
data_mix[['ตำบล']] = data_mix[['ตำบล']].fillna('ไม่มีข้อมูล')
data_mix[['อำเภอ']] = data_mix[['อำเภอ']].fillna('ไม่มีข้อมูล')
data_mix[['ภาค']] = data_mix[['ภาค']].fillna('ไม่มีข้อมูล')
data_mix[['East']] = data_mix[['East']].fillna('ไม่มีข้อมูล')
data_mix[['North']] = data_mix[['North']].fillna('ไม่มีข้อมูล')
data_mix[['ลักษณะชุมชน']] = data_mix[['ลักษณะชุมชน']].fillna('ไม่มีข้อมูล')
data_mix[['ลักษณะสำคัญ']] = data_mix[['ลักษณะสำคัญ']].fillna('ไม่มีข้อมูล')
```

ตรวจสอบค่า missing อีกรอบ

```
data_mix.isnull().any() # ตรวจสอบอีกรอบ
```

จังหวัด	False
รหัสจังหวัด	False
visitor2555	False
visitor2556	False
visitor2557	False
visitor2558	False
visitor2559	False
visitor2560	False
visitor2561	False
income2555	False
income2556	False
income2557	False
income2558	False
income2559	False
income2560	False
income2561	False
ชื่อยานชุมนุมชนเก่า	False
ตำบล	False
อำเภอ	False
ภาค	False
East	False
North	False
ลักษณะชุมชน	False
ลักษณะสำคัญ	False
dtype:	bool

แสดงผลเป็น False แปลว่าไม่มีค่า missing แล้ว สามารถนำข้อมูลไปใช้ได้

ข้อมูลี่ 4 จำนวนประชากร จำแนกตามรายจังหวัด ปี 2561

ข้อมูลี่ 4 จำนวนประชากร จำแนกตามรายจังหวัด ปี 2561

```
data_pop = pd.read_csv(os.path.join(path, 'stat_c61.csv'))
data_pop
```

	รหัสจังหวัด	Male	Female	Total	House
0	th-bm	2679453	2997195	5676648	2959524
1	th-sp	634679	691929	1326608	675382
2	th-no	581620	664675	1246295	689035
3	th-pt	543652	602440	1146092	603834
4	th-pa	393570	423871	817441	322991
...
72	th-tg	314781	328335	643116	229122
73	th-pl	255830	269214	525044	192087
74	th-pi	355424	362653	718077	188289
75	th-yl	264928	267398	532326	164531
76	th-nw	397100	405374	802474	212873

77 rows × 5 columns

เชื่อม data_pop เข้ากับ data_mix เข้ากันด้วย 'รหัสจังหวัด' จะได้ data_all 22

```
data_all = data_mix.merge(data_pop,how='left',left_on='รหัสจังหวัด',right_on='รหัสจังหวัด')
data_all
```

	จังหวัด	รหัส จังหวัด	visitor2555	visitor2556	visitor2557	visitor2558	visitor2559	visitor2560	visitor2561	income2555	income2556	income2557	income2558
0	กรุงเทพมหานคร	th-bm	47185031	50568902	50972772	56515597	59196331	63575737	65534281	570327.58	626535.05	634803.16	7303
1	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639	3451.66	3768.56	4044.50	4901
2	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639	3451.66	3768.56	4044.50	4901
3	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639	3451.66	3768.56	4044.50	4901
4	สมุทรปราการ	th-sp	2107433	2204073	2350340	2780429	2900180	3218934	3373639	3451.66	3768.56	4044.50	4901
...
609	นราธิวาส	th-nw	489944	570973	589690	635910	646020	665710	692057	2139.00	2208.00	2323.54	2524

ตรวจสอบค่า missing อีกรอบ

23

```
data_all.isnull().any() # ตรวจสอบอีกรอบ
```

```
จังหวัด                False
รหัสจังหวัด            False
visitor2555            False
visitor2556            False
visitor2557            False
visitor2558            False
visitor2559            False
visitor2560            False
visitor2561            False
income2555             False
income2556             False
income2557             False
income2558             False
income2559             False
income2560             False
income2561             False
ชื่อยานชุมนุมชนเก่า     False
ตำบล                 False
อำเภอ                 False
ภาค                  False
East                  False
North                 False
ลักษณะชุมชน          False
ลักษณะสำคัญ          False
Male                  False
Female                False
Total                 False
House                 False
dtype: bool
```

แสดงผลเป็น False แปลว่าไม่มีค่า missing แล้ว สามารถนำข้อมูลไปใช้ได้

classification



ทำนุภาพลักษณะของชุมชนเก่า โดยใช้ข้อมูลจำนวนครัวเรือน
ข้อมูลจำนวนประชากรแต่ละจังหวัด และ รายได้จากการ
ท่องเที่ยวของแต่ละจังหวัด

```
data_cut = data_all[['House', 'Total', 'income2561', 'ลักษณะชุมชน']]
data_cut
```

	House	Total	income2561	ลักษณะชุมชน
0	2959524	5676648	1051118.51	ไม่มีข้อมูล
1	675382	1326608	6656.34	เป็นชุมชนตลาด
2	675382	1326608	6656.34	เป็นชุมชนตลาด
3	675382	1326608	6656.34	เป็นชุมชนตลาด
4	675382	1326608	6656.34	เป็นชุมชนประมง
...
609	212873	802474	3137.67	เป็นชุมชนตลาด
610	212873	802474	3137.67	เป็นชุมชนเกษตรกรรม
611	212873	802474	3137.67	เป็นชุมชนเกษตรกรรม
612	212873	802474	3137.67	เป็นชุมชนตลาด
613	212873	802474	3137.67	เป็นชุมชนตลาด

614 rows × 4 columns

เลือกข้อมูล column ที่ต้องการจะใช้ ได้แก่

- House = จำนวนครัวเรือน
- Total = จำนวนประชากร จำแนกตามรายจังหวัด ปี 2561
- income2561 = รายได้จากการท่องเที่ยว จำแนกตามรายจังหวัด ปี 2561
- ลักษณะชุมชน = แสดงรายละเอียดลักษณะสำคัญของชุมชน

Decision tree



ขั้นตอนนี้ คือ การกำหนดตัวแปร x, y เพื่อทำ decision tree
 โดยกำหนด x คือ House, Total, income2561 และ y คือ ลักษณะชุมชน (สิ่งที่เราจะทำนาย)

```
#กำหนด X และ y เพื่อทำ decision tree
X = data_cut[['House', 'Total', 'income2561']]
y = data_cut['ลักษณะชุมชน']
```

X

	House	Total	income2561
0	2959524	5676648	1051118.51
1	675382	1326608	6656.34
2	675382	1326608	6656.34
3	675382	1326608	6656.34
4	675382	1326608	6656.34
...
609	212873	802474	3137.67
610	212873	802474	3137.67
611	212873	802474	3137.67
612	212873	802474	3137.67
613	212873	802474	3137.67

614 rows × 3 columns

Train - model

Train Model

```
from sklearn.model_selection import train_test_split
```

Train-Test

#แบ่งข้อมูลออกเป็น 2 ส่วน ด้วยการสุ่ม

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=8)
```

X.shape

(614, 3)

X_train.shape

(307, 3)

X_test.shape

(307, 3)

Train - Validation

Train - Validation

#แบ่งข้อมูลออกเป็น 2 ส่วน ด้วยการสุ่ม

```
X_train2, X_val, y_train2, y_val = train_test_split(X_train, y_train, test_size=0.5, random_state=8)
```

#import

```
from sklearn.tree import DecisionTreeClassifier
```

#import accuracy_score จาก sklearn.metrics

```
from sklearn.metrics import accuracy_score
```


Decision tree 1 ใช้เกณฑ์ max_depth=5,random_state=13

Decision tree 1 ใช้เกณฑ์ max_depth=5,random_state=13

```
#define  
mytree = DecisionTreeClassifier(max_depth=5,random_state=13)  
#train  
mytree.fit(X_train2,y_train2)  
#test  
mytree_result = mytree.predict(X_val)  
#accuracy_score  
accuracy_score(y_val, mytree_result)
```

0.43506493506493504

Decision tree 2 ใช้เกณฑ์ splitter='random',random_state=13

Decision tree 2 ใช้เกณฑ์ splitter='random',random_state=13

```
#define  
mytree2 = DecisionTreeClassifier(splitter='random',random_state=13)  
#train  
mytree2.fit(X_train2,y_train2)  
#test  
mytree_result2 = mytree2.predict(X_val)  
#accuracy_score  
accuracy_score(y_val, mytree_result2)
```

0.487012987012987

Decision tree 3 ใช้เกณฑ์ criterion='entropy', splitter='random', random_state=13

Decision tree 3 ใช้เกณฑ์ criterion='entropy', splitter='random', random_state=13

```
#define  
mytree3 = DecisionTreeClassifier(criterion='entropy', splitter='random', random_state=13)  
#train  
mytree3.fit(X_train2, y_train2)  
#test  
mytree_result3 = mytree3.predict(X_val)  
#accuracy_score  
accuracy_score(y_val, mytree_result3)
```

0.5

Decision tree 4 ใช้เกณฑ์

criterion='gini',max_leaf_nodes=50,splitter='random',random_state=13

Decision tree 4 ใช้เกณฑ์ criterion='gini',max_leaf_nodes=50,splitter='random',random_state=13

```
#define
mytree4 = DecisionTreeClassifier(criterion='gini',max_leaf_nodes=50,splitter='random',random_state=13)
#train
mytree4.fit(X_train2,y_train2)
#test
mytree_result4 = mytree4.predict(X_val)
#accuracy_score
accuracy_score(y_val, mytree_result4)
```

0.5064935064935064

KNN



Import

```
#import  
from sklearn.neighbors import KNeighborsClassifier
```

KNN1

ใช้เกณฑ์ $n_neighbors=3$, $weights='uniform'$

KNN1

ใช้เกณฑ์ $n_neighbors=3$, $weights='uniform'$

```
# Define  
neigh1 = KNeighborsClassifier(n_neighbors=3, weights='uniform') #ถามเพื่อนบ้านที่ใกล้สุด 3 คน และเชื่อทุกคนเท่ากัน  
# Train  
neigh1.fit(X_train2,y_train2)  
# Test  
knn1_result = neigh1.predict(X_val)  
accuracy_score(y_val, knn1_result)
```

0.5064935064935064

KNN2

ใช้เกณฑ์ `n_neighbors=10, weights='distance'`

KNN2

ใช้เกณฑ์ `n_neighbors=10, weights='distance'`

```
# Define
neigh2 = KNeighborsClassifier(n_neighbors=10, weights='distance')#ถามเพื่อนบ้านที่ใกล้สุด 10 คน และเชื่อคนที่ใกล้มากกว่าไกล
# Train
neigh2.fit(X_train2,y_train2)
# Test
knn2_result = neigh2.predict(X_val)
accuracy_score(y_val, knn2_result)
```

0.5

KKN3

ใช้เกณฑ์ $n_neighbors=1$

KKN3

ใช้เกณฑ์ $n_neighbors=1$

```
# Define
neigh3 = KNeighborsClassifier(n_neighbors=1)#เชื่อเพื่อนบ้านที่ใกล้ที่สุดคนเดียว
# Train
neigh3.fit(X_train2,y_train2)
# Test
knn3_result = neigh3.predict(X_val)
accuracy_score(y_val, knn3_result)
```

0.461038961038961

Neural Network



Import

```
from sklearn.neural_network import MLPClassifier
```

Train-Test

ANN1

ANN1

```
# Define
perceptron1 = MLPClassifier(random_state=13,max_iter=2000,hidden_layer_sizes=1,learning_rate_init=0.05)
# Train
perceptron1.fit(X_train2,y_train2)
# Test
ann1_result = perceptron1.predict(X_val)
accuracy_score(y_val, ann1_result)
```

0.2792207792207792

Train-Test

ANN2

ANN2

```
# Define
perceptron2 = MLPClassifier(random_state=13,max_iter=2000,hidden_layer_sizes=10,learning_rate_init=100)
# Train
perceptron2.fit(X_train2,y_train2)
# Test
ann2_result = perceptron2.predict(X_val)
accuracy_score(y_val, ann2_result)
```

0.37662337662337664

Train-Test

ANN3

ANN3

```
# Define
perceptron3 = MLPClassifier(random_state=13,max_iter=20000,hidden_layer_sizes=10,learning_rate_init=0.01)
# Train
perceptron3.fit(X_train2,y_train2)
# Test
ann3_result = perceptron3.predict(X_val)
accuracy_score(y_val, ann3_result)
```

0.43506493506493504

Retrain & Evaluate

จากค่า accuracy ของ Decision tree 4 และ KNN1 มีค่าเท่ากันคือ 0.5064935064935064
จึงทำการ Train ทั้งสองตัวแบบเพื่อหาตัวแบบที่เหมาะสมที่สุด

เลือก Decision tree 4 แล้ว Train ใหม่ด้วย Training

```
#import
from sklearn.tree import DecisionTreeClassifier
#define
mytree_4final = DecisionTreeClassifier(criterion='gini',max_leaf_nodes=50,splitter='random',random_state=13)
#train
mytree_4final.fit(X_train,y_train)
#test-evaluate
mytree_4_finalresult = mytree_4final.predict(X_test)
dectree_ac = accuracy_score(y_test, mytree_4_finalresult)
dectree_ac
```

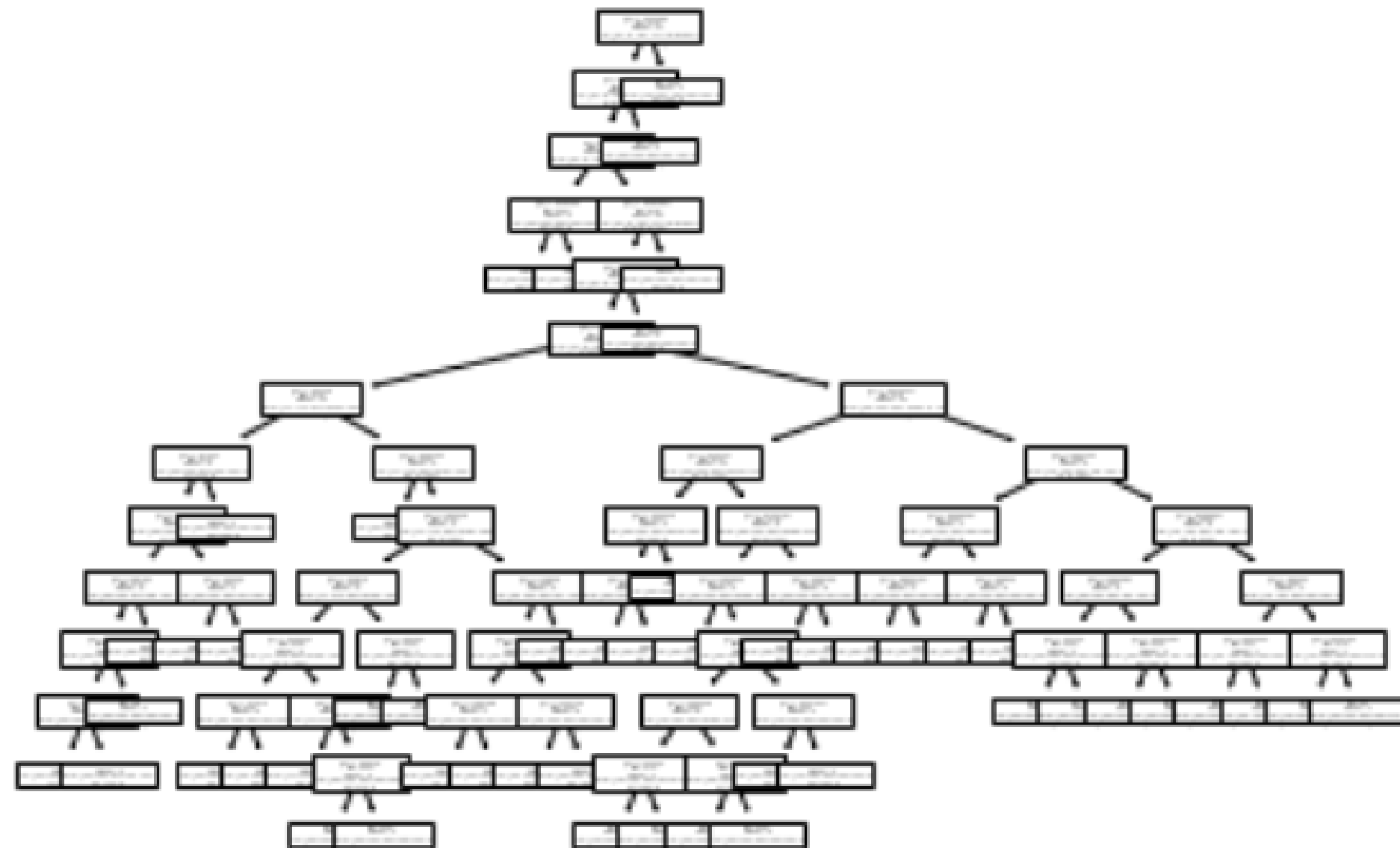
0.5374592833876222

Plot tree

44

```
from sklearn.tree import plot_tree
```

```
plot_tree(mytree_4final);
```



เลือก KNN1 มา Train ใหม่ อีกครั้ง

เลือก KNN1 มา Train ใหม่ อีกครั้ง

```
: # Define
neigh1final = KNeighborsClassifier(n_neighbors=3, weights='uniform') #ถามเพื่อนบ้านที่ใกล้สุด 3 คน และเชื่อทุกคนเท่ากัน
# Train
neigh1final.fit(X_train,y_train)
# Test
knn1final_result = neigh1final.predict(X_test)
knn_ac = accuracy_score(y_test, knn1final_result)
knn_ac
: 0.5211726384364821
```

สรุป จากการ Train ของทั้งสองตัวแบบ พบว่า ค่า accuracy ของ Decision tree 4 = 0.5374592833876222
และ ค่า accuracy ของ KNN1 = 0.5211726384364821

ดังนั้น เมื่อนำสองค่ามาเทียบกันพบว่า ค่า accuracy ของ KNN มีค่าน้อยกว่า Decision tree จึงเลือก **ตัวแบบ Decision tree** เพราะมีความเหมาะสมที่สุด

```
from matplotlib import pyplot as plt
```

```
labels = ['Decision Tree', 'KNN']
```

```
accuracy_Y = [dectree_ac,knn_ac]
```

```
#plot
```

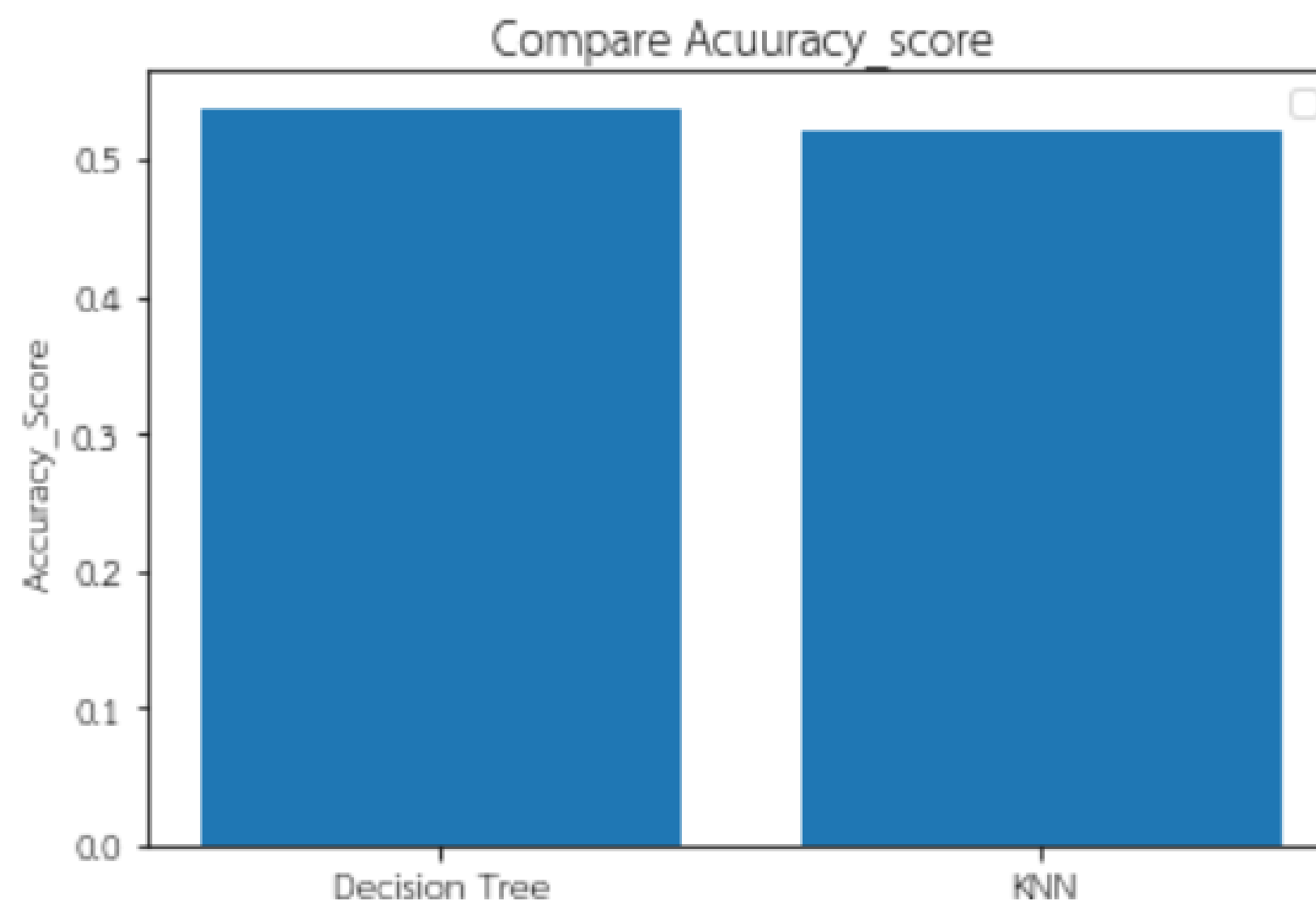
```
plt.bar(labels,accuracy_Y)
```

```
plt.ylabel('Accuracy_Score')
```

```
plt.title('Compare Acuuracy_score')
```

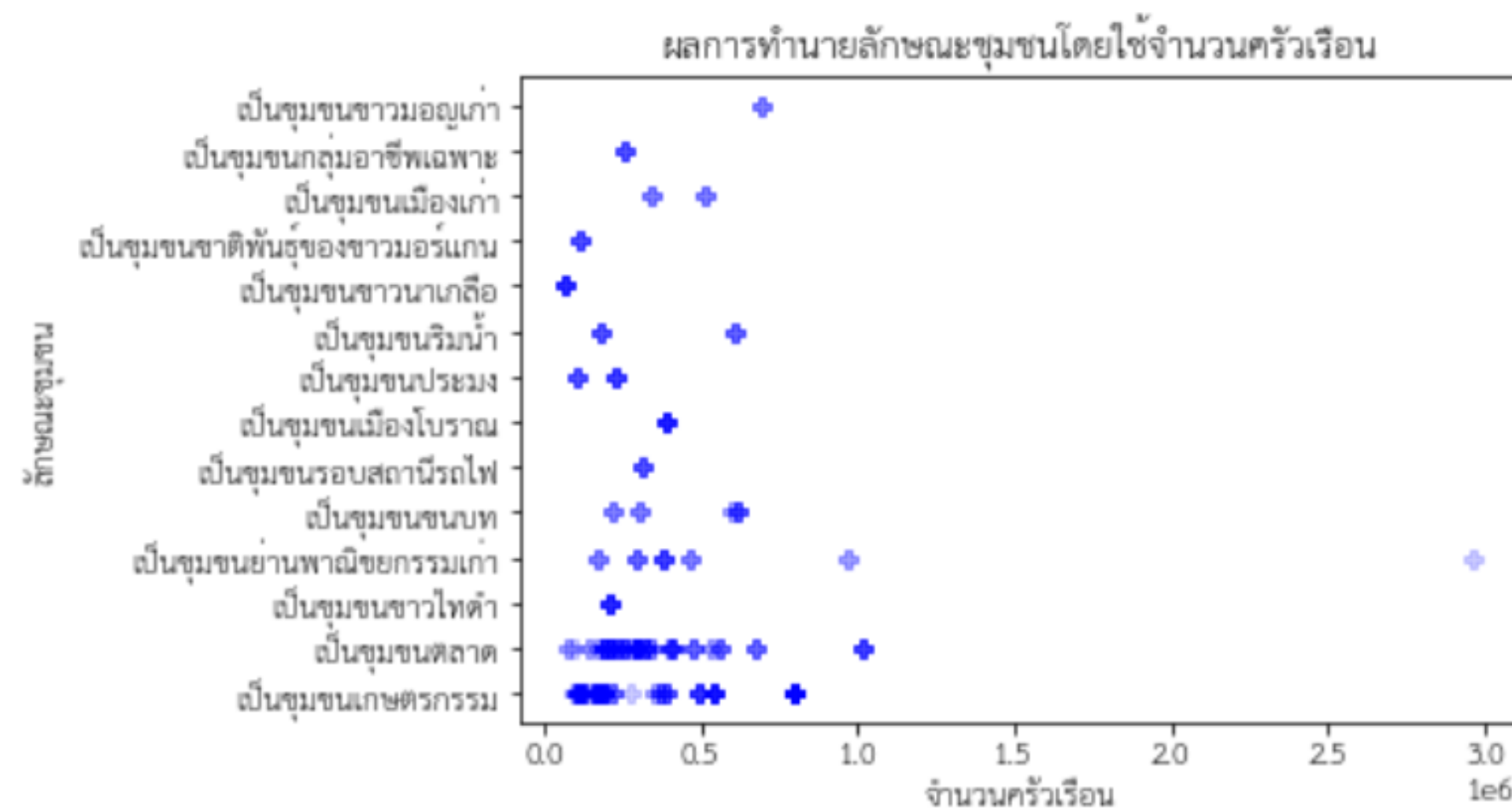
```
plt.legend();
```

No handles with labels found to put in legend.



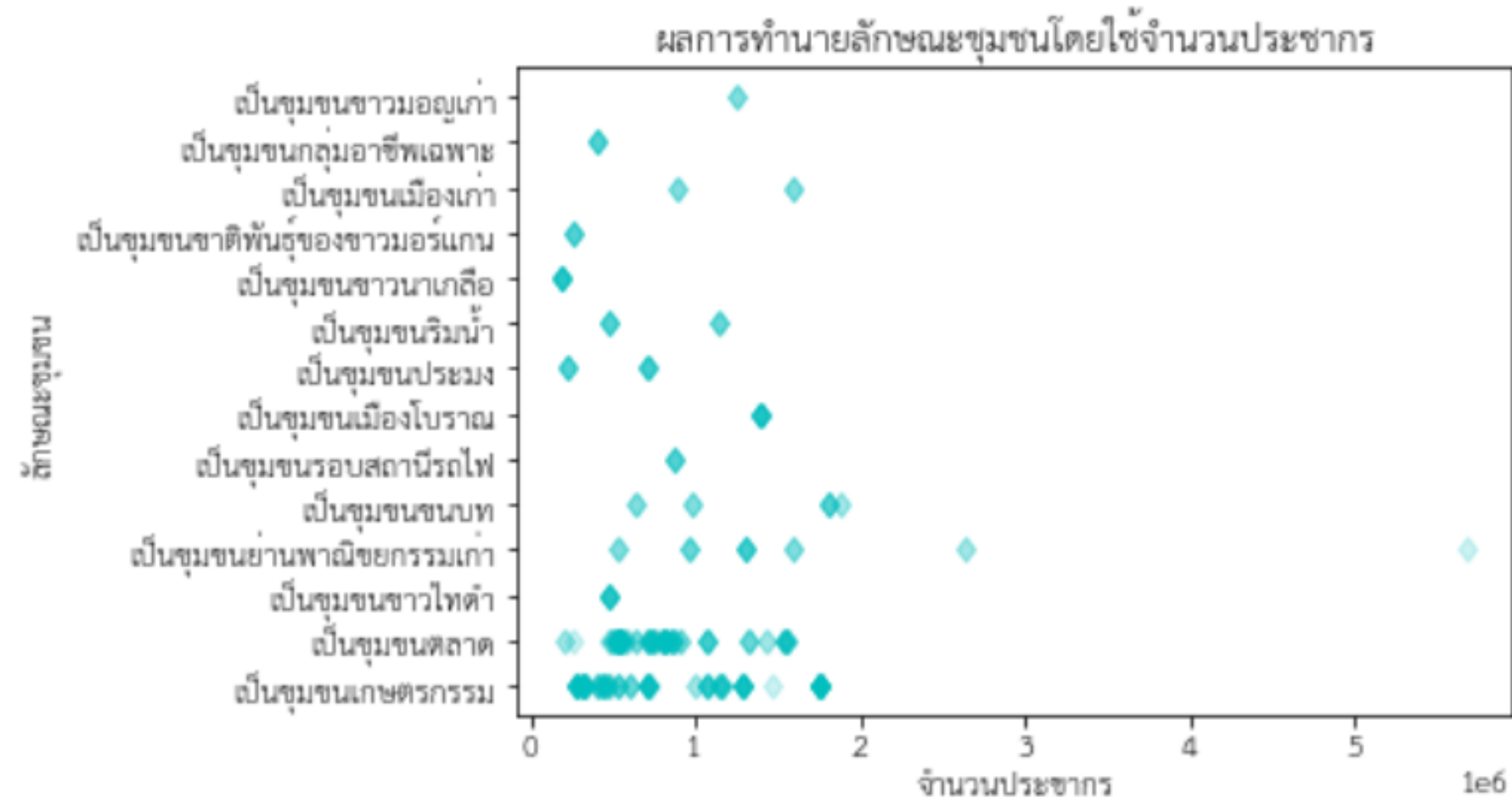
```
from matplotlib import pyplot as plt # เรียกใช้อัลกอริทึมที่เราต้องการ
```

```
#ผลการทำนายลักษณะชุมชนโดยใช้จำนวนครัวเรือน
plt.plot(X_test['House'],knn1final_result,'Pb',alpha =0.2)
plt.xlabel('จำนวนครัวเรือน')
plt.ylabel('ลักษณะชุมชน')
plt.title('ผลการทำนายลักษณะชุมชนโดยใช้จำนวนครัวเรือน')
plt.show();
```



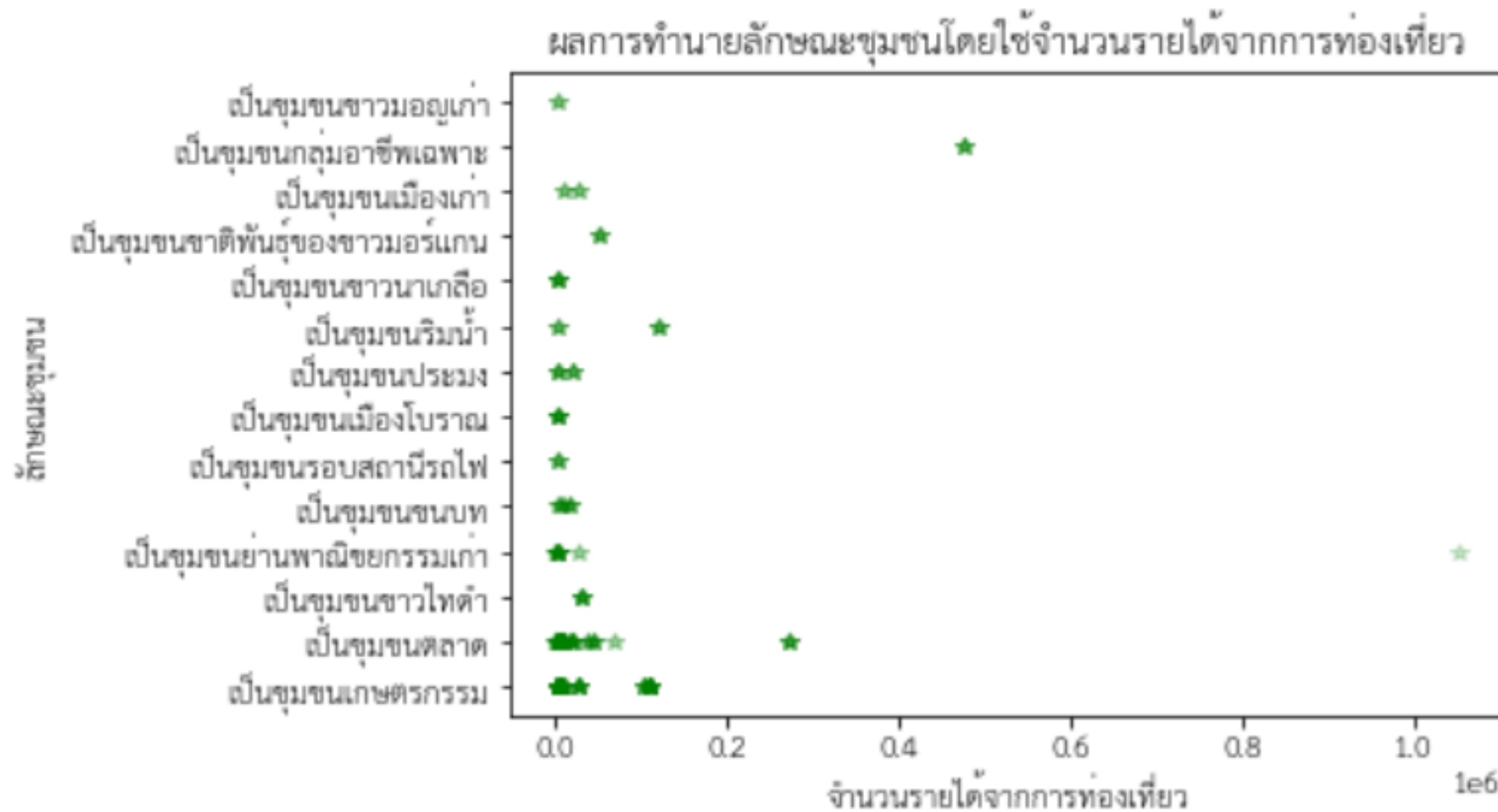
สรุป จากผลการทำนายลักษณะชุมชนโดยใช้จำนวนครัวเรือน พบว่า จังหวัดที่มีจำนวนครัวเรือนมาก ส่วนใหญ่จะเป็นชุมชนย่านพาณิชยกรรมเก่า, เป็นชุมชนตลาด, เป็นชุมชนเกษตรกรรม


```
#ผลการทำนายลักษณะชุมชนโดยใช้จำนวนประชากร
plt.plot(X_test['Total'],knn1final_result,'dc',alpha =0.2)
plt.xlabel('จำนวนประชากร')
plt.ylabel('ลักษณะชุมชน')
plt.title('ผลการทำนายลักษณะชุมชนโดยใช้จำนวนประชากร')
plt.show();
```



สรุป จากผลการทำนายลักษณะชุมชนโดยใช้จำนวนประชากร พบว่า จังหวัดที่มีจำนวนประชากรมาก ส่วนใหญ่จะเป็นชุมชนย่านพาณิชย์กรรมเก่า, เป็นชุมชนตลาด, เป็นชุมชนเกษตรกรรม

```
#ผลการทำนายลักษณะชุมชนโดยใช้จำนวนรายได้จากการท่องเที่ยว
plt.plot(X_test['income2561'],knn1final_result,'*g',alpha =0.2)
plt.xlabel('จำนวนรายได้จากการท่องเที่ยว')
plt.ylabel('ลักษณะชุมชน')
plt.title('ผลการทำนายลักษณะชุมชนโดยใช้จำนวนรายได้จากการท่องเที่ยว')
plt.show();
```



สรุป จากผลการทำนายลักษณะชุมชนโดยใช้จำนวนรายได้จากการท่องเที่ยว พบว่า จังหวัดที่มีรายได้จากการท่องเที่ยวมาก ส่วนใหญ่จะเป็นชุมชนตลาด, เป็นชุมชนเกษตรกรรม



THANK YOU