

# วัตถุประสงค์



เพื่อทำนายว่าจำนวนประชากร  
รายได้ ค่าครองชีพ  
มีผลต่อการนำเข้าสินค้าทางการ  
เกษตรหรือไม่  
โดยจะจัดกลุ่มปริมาณการนำเข้า  
สินค้าให้อยู่ในกลุ่มน้อย หรือ  
กลุ่มมาก

- การนำเข้าปุ๋ยเคมี (รายประเทศ) องค์กร : ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร กรมวิชาการเกษตร จาก : <https://data.go.th>
- การนำเข้าเมล็ดพันธุ์ควบคุม (รายประเทศ) องค์กร : ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร กรมวิชาการเกษตร จาก : <https://data.go.th>
- สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง องค์กร : ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร กรมวิชาการเกษตร จาก : <https://data.go.th>
- ข้อมูลทั่วไปของประเทศไทยที่นำเข้าสินค้าทางการเกษตร จาก <https://drive.google.com/drive/folders/1PCwak7QUVizd6sfY-NvAZBRLNh5N-fhz?usp=sharing>

# แหล่งข้อมูล



## ตารางที่ 1 ตาราง ปริมาณการนำเข้าปุ๋ยเคมี



```
data_fc = pd.read_csv(os.path.join(path, 'ปริมาณการนำเข้าปุ๋ยเคมี.csv'))
data_fc #การนำเข้าไฟล์ที่มีตัวแปรคือdata_fc
```

	id	Title	Subtitle_FC	Unit	View	Freq	Y2006	Y2007	Y2008	Y2009
0	1.0	ปริมาณการนำเข้าปุ๋ยเคมีรายประเทศ	SAUDI ARABIA	เมตริกตัน	ประเทศ	รายปี	582,636.00	787,997.00	672,800.00	1,007,200.00
1	2.0	ปริมาณการนำเข้าปุ๋ยเคมีรายประเทศ	CHINA	เมตริกตัน	ประเทศ	รายปี	270,060.53	675,840.83	569,348.49	427,900.00
2	3.0	ปริมาณการนำเข้าปุ๋ยเคมีรายประเทศ	KOREA REPUBLIC OF	เมตริกตัน	ประเทศ	รายปี	426,401.90	472,745.91	287,250.00	417,400.00
3	4.0	ปริมาณการนำเข้าปุ๋ยเคมีรายประเทศ	QATAR	เมตริกตัน	ประเทศ	รายปี	268,274.00	445,726.00	460,968.00	387,500.00
4	5.0	ปริมาณการนำเข้าปุ๋ยเคมีรายประเทศ	MALAYSIA	เมตริกตัน	ประเทศ	รายปี	244,645.54	234,427.46	173,449.15	285,700.00

✓ 1 วินาที เสร็จสมบูรณ์เมื่อ 20:31

# Preprocessing



## ตารางที่ 2 ตารางปริมาณการนำเข้า เมล็ดพันธุ์



```
data_pc = pd.read_csv(os.path.join(path, 'ปริมาณการนำเข้าเมล็ดพันธุ์.csv')) #การนำเข้าไฟล์ที่มี data_pc
```

	id	Title	Subtitle_PC	Unit	View	Freq	Y2006	Y2007	Y2008
0	1	ปริมาณการนำเข้าเมล็ดพันธุ์ความ	SAUDI ARABIA	เมตริกตัน	ประเทศ	รายปี	4,910.26	6,203.20	6,274.60
1	2	ปริมาณการนำเข้าเมล็ดพันธุ์ความ	CHINA	เมตริกตัน	ประเทศ	รายปี	124,916.36	46,686.16	47,946.76
2	3	ปริมาณการนำเข้าเมล็ดพันธุ์ความ	KOREA REPUBLIC OF	เมตริกตัน	ประเทศ	รายปี	35,847	59,459.38	165,267.43
3	4	ปริมาณการนำเข้าเมล็ดพันธุ์ความ	QATAR	เมตริกตัน	ประเทศ	รายปี	47,414.77	124,915.02	580,000.00
4	5	ปริมาณการนำเข้าเมล็ดพันธุ์ความ	MALAYSIA	เมตริกตัน	ประเทศ	รายปี	37,645.03	46,467.47	23,567.58

## ตารางที่ 3 ตาราง สารป้องกันและ กำจัดโรคพืช



```
data_pd = pd.read_csv(os.path.join(path, 'สารป้องกันและกำจัดโรคพืช.csv')) #การนำเข้าไฟล์ที่มี data_pd
```

	id	Title	Subtitle_PD	Unit	View	Freq	Y2007	Y2008	Y2009
0	1	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	SAUDI ARABIA	กิโลกรัม	ประเทศ	รายปี	2,249,315.00	2,143,682.00	1,192,758.30
1	2	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	CHINA	กิโลกรัม	ประเทศ	รายปี	1,160,075.00	1,519,815.00	782,375.00
2	3	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	KOREA REPUBLIC OF	กิโลกรัม	ประเทศ	รายปี	1,875,370.00	1,724,557.00	1,332,152.90
		สารป้องกันและ							



```
[16] merged_03 = data_pd.merge(merged_data_FC_data_pc,how='left',left_on='Subtitle_PD',
merged_03
#ใช้merged_03 เพื่อเชื่อมตารางเข้าด้วยกัน
```

	id	Title	Subtitle_PD	Unit	View	Freq	Y2007	Y2008	Y2009
0	1	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	SAUDI ARABIA	กิโลกรัม	ประเทศ	รายปี	2,249,315.00	2,143,682.00	1,192,758.30
1	2	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	CHINA	กิโลกรัม	ประเทศ	รายปี	1,160,075.00	1,519,815.00	782,375.00
2	3	สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำเข้าสูง	KOREA REPUBLIC OF	กิโลกรัม	ประเทศ	รายปี	1,875,370.00	1,724,557.00	1,332,152.90

✓ 1 วินาที เสร็จสมบูรณ์เมื่อ 20:31

# Preprocessing

จากนั้นรวมตารางทั้ง 3 ตาราง  
2 ครั้ง ซึ่งเราจะใช้  
Subtitle(ชื่อประเทศ) ในการ  
เชื่อมตาราง



✕ □ □

☆

# Preprocessing



ตารางที่ 4 ตาราง ข้อมูลทั่วไปนำเข้าสินค้าทางเกษตร



```
[27] path = '/content/drive/My Drive/Project'#โดยเราจะทำการระบุให้ชี้ไปที่ โฟลเดอร์ที่เราเซฟไฟล์ไว้  
[28] data_add = pd.read_csv(os.path.join(path, 'ข้อมูลทั่วไปนำเข้าสินค้าทางเกษตร.csv'))  
data_add
```

	id	country	View	population(million)	religion	capital
0	1	SAUDI ARABIA	ประเทศ	34.22	อิสลาม	ริยาด
1	2	CHINA	ประเทศ	1,400	พุทธ	ปักกิ่ง
2	3	KOREA REPUBLIC OF	ประเทศ	51	ไม่นับถือ	โซล
3	4	QATAR	ประเทศ	2	อิสลาม	โดฮา
4	5	MALAYSIA	ประเทศ	30	อิสลาม	กัวลาลัมเปอร์
5	6	RUSSIAN FEDERATION (CIS)	ประเทศ	146	คริสต์	มอสโก/มอสโก
6	7	KUWAIT	ประเทศ	4.207	อิสลาม	คูเวตซิตี
7	8	EGYPT	ประเทศ	93	อิสลาม	ไคโร
8	9	NORWAY	ประเทศ	5.39	คริสต์	ออสโล
9	10	CANADA	ประเทศ	36.28	โรมันคาทอลิก	ออตตาวา
10	11	ISRAEL	ประเทศ	9.39	ยูดาย	เยรูซาเลม
11	12	JAPAN	ประเทศ	125	พุทธ	โตเกียว
12	13	BAHRAIN	ประเทศ	0.688	อิสลาม	มานามา
13	14	PHILIPPINES	ประเทศ	109	คริสต์	มะนิลา
14	15	FINLAND	ประเทศ	5.53	คริสต์	เฮลซิงกิ

✓ 1 วินาที เสร็จสมบูรณ์เมื่อ 20:31

ตัดคอลัมน์ที่10ออก เพราะมันเป็นค่า NAN



จากนั้นทำการรวมทั้ง 4 ตาราง  
เข้าด้วยกันจะได้ดังนี้ และเช็คว่ามี  
ค่า missing แล้วทำการแก้

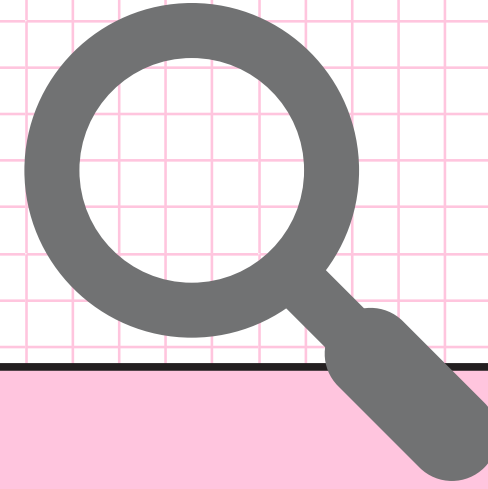
```
merged_04 = data_ADD.merge(merged_03,how='left',left_on='country',right_on='country')
merged_04
```

#add table that about coutry anothor

	id_x	country	View_x	population(million)	religion	capit
0	1	SAUDI ARABIA	ประเทศ	34.22	อิสลาม	ริย
1	2	CHINA	ประเทศ	1,400	พุทธ	ปัก
2	3	KOREA REPUBLIC OF	ประเทศ	51	ไม่นับถือ	โซล



# ปัญหา Classification



ปัญหาของเราคือ เรามีสินค้าทางการเกษตรในประเทศของเราไม่เพียงพอ เราเลยอยากรู้ว่าในปีถัดไปเราควรสั่งสินค้าจากประเทศบ้าง โดยเราจะพิจารณาจากปริมาณการนำเข้า 3 ปีย้อนหลัง คือ 2018 2019 2020 นอกจากนี้เราอยากรู้อีกว่ามีปัจจัยใดบ้างที่ส่งผลต่อการผลิตสินค้าทางการเกษตรของประเทศที่เรานำเข้า ซึ่งเรามีปัจจัยที่คาดว่าจะเกี่ยวข้องคือ จำนวนประชากร ค่าแรงโดยเฉลี่ย และค่าครองชีพโดยเฉลี่ย

ดังนั้น เราจึงได้ทำการวิเคราะห์ Data mining เพื่อทำนายว่าจำนวนประชากร รายได้ ค่าครองชีพ มีผลต่อการนำเข้าสินค้าทางการเกษตรหรือไม่ โดยจะจัดกลุ่มปริมาณการนำเข้าสินค้าให้อยู่ในกลุ่มน้อย หรือ กลุ่มมาก



# Classification

- 1.สร้างตารางที่เราจะนำไปจัดหมวดหมู่
- 2.ตรวจสอบขนาด
- 3.ตรวจสอบค่า missing ค่า NAN
- 4.Drop.NA
- 5.Cluster
- 6.นำตาราง Clusterไปรวมกับข้อมูลเก่าแล้วตัดแค่ข้อมูลที่เราต้องการ
- 7.ตรวจสอบ Type ของข้อมูลแล้วนำไปทำโมเดล
- 8.ตรวจสอบ Type อีกครั้ง

```
Data = Datafloat[['total','population(million)','living expenses(USD)','avg income(USD)']]  
Data#เลือกเฉพาะคอลัมที่ต้องการได้ดังนี้
```

	total	population(million)	living expenses(USD)	avg income(USD)
0	1	34.220	952.0	4754.0
1	1	1400.000	752.0	884.0
2	0	51.000	1154.0	2738.0
3	1	2.000	1856.0	4684.0
4	1	30.000	652.0	882.0
5	0	146.000	424.0	891.0
6	0	4.207	1326.0	3024.0
7	1	93.000	449.0	250.0
8	1	5.390	2074.0	6521.0
9	1	36.280	1701.0	3620.0
10	0	9.390	1774.0	3589.0
11	0	125.000	1370.0	3465.0

# เลือก model

จากการทำทั้ง 3 โมเดลการทำนายพบว่า จำนวนประชากร รายได้ ค่าครองชีพ มีผลต่อการนำเข้าสินค้าทางการเกษตรว่าแต่ละประเทศจัดอยู่ในกลุ่มมากหรือน้อยได้ ซึ่งได้ค่าความแม่นยำทั้ง 3 ค่า คือ KNN 62.50% , Decision Tree 58.34%, Random Forest 62.50% เราจึงตัดสินใจเลือก Random Forest โมเดล เนื่องจาก Random Forest ใช้ได้ทั้งกับปัญหา classification และ regression

```
cross_val_score(skRF, X_train, y_train, cv=12).mean()
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_
% (min_groups, self.n_splits)), UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_va
estimator.fit(X_train, y_train, **fit_params)
0.625
```

✕ □ —

```
from sklearn.ensemble import RandomForestClassifier
skRF = RandomForestClassifier(random_state=1)
skRF.fit(X_train,y_train)
skRF.score(X,y)
y_predrf = skRF.predict(X_test)
y_predrf
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel
This is separate from the ipykernel package so
array([0, 1, 1, 1, 1, 1])
```

<

>

# การวัดผล

เราจะวัดผลความแม่นยำโดยใช้ `cross_val_score` เพราะเราจะได้ไม่ต้องแบ่งกลุ่ม data เอง แบบ `accuracy_score` `cross-val` จะแบ่งและวัดให้เลย ซึ่งสรุปการวัดผลได้ดังนี้

จากการทำทั้งสามโมเดลการทำนายพบว่า จำนวนประชากร รายได้ ค่าครองชีพ มีผลต่อการนำเข้าสู่สินค้าทางการเกษตรว่าแต่ละประเทศจัดอยู่ในกลุ่มมากหรือน้อยได้ ที่มีความความแม่นยำมากที่สุดคือ RF โดยมีความแม่นยำ 62.50% เท่ากัน

```
[ ] cross_val_score(KNN_1, X_train, y_train, cv=12).mean()

/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:198: UserWarning:
  % (min_groups, self.n_splits)), UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:220: UserWarning:
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:240: UserWarning:
  estimator.fit(X_train, y_train, **fit_params)
```

```
[ ] cross_val_score(myTree, X_train, y_train, cv=12).mean()

/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:198: UserWarning:
  % (min_groups, self.n_splits)), UserWarning)
0.5833333333333333
```

```
▶ cross_val_score(skRF, X_train, y_train, cv=12).mean()

/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:198: UserWarning:
  % (min_groups, self.n_splits)), UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:220: UserWarning:
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:240: UserWarning:
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selection.py:260: UserWarning:
  estimator.fit(X_train, y_train, **fit_params)
```

## วัดผลเปรียบเทียบความแม่นยำ

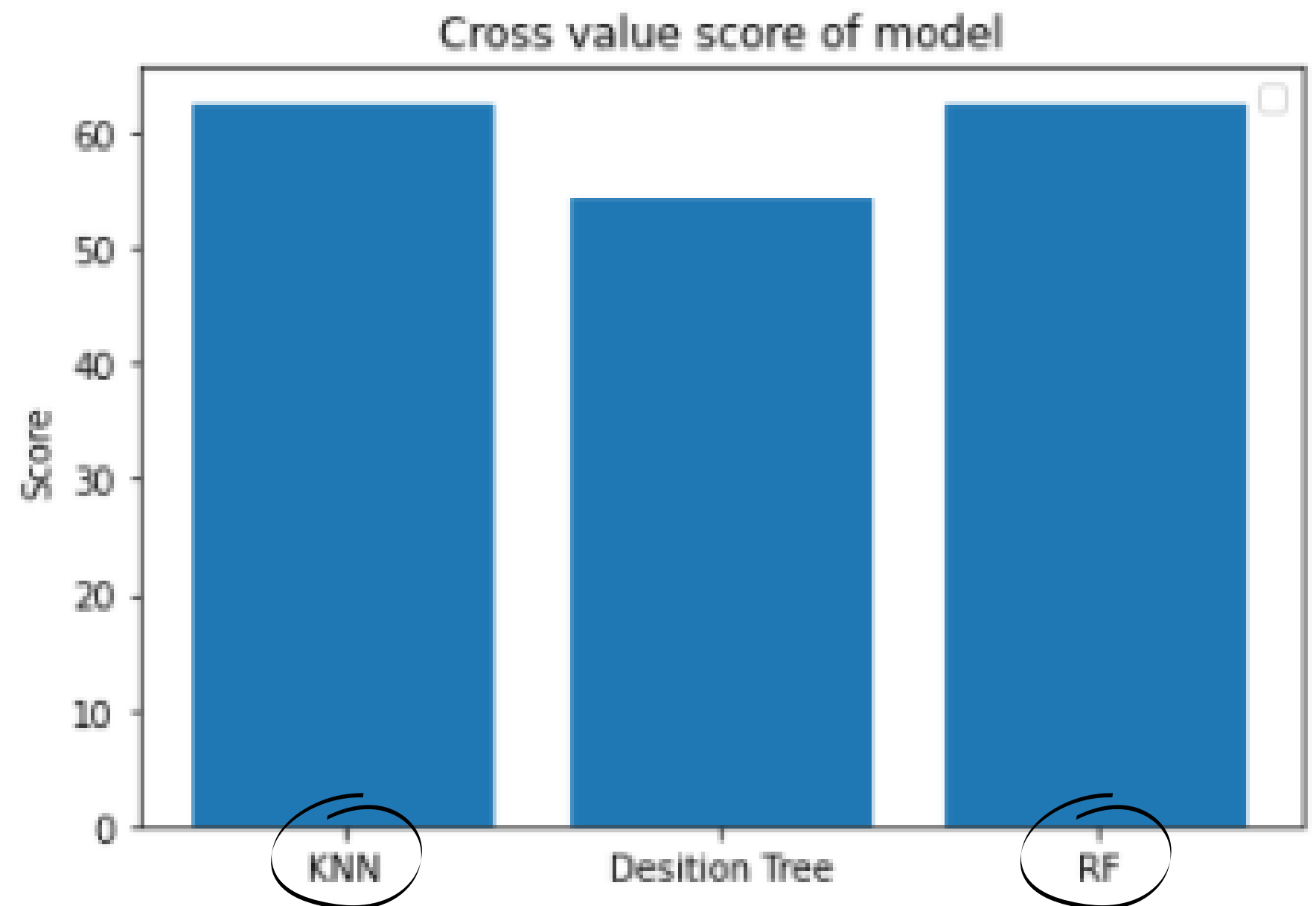
เราจะเห็นว่าโมเดล KNN  
และ Random Forest มี  
ความแม่นยำเท่ากัน



`width = 0.35` #ขนาดของแท่ง

```
plt.bar([0,1,2],cross_val_score)
plt.ylabel('Score')
plt.title('Cross value score of model')
plt.xticks([0,1,2],labels)
plt.legend();
plt.show()
```

No handles with labels found to put in legend.





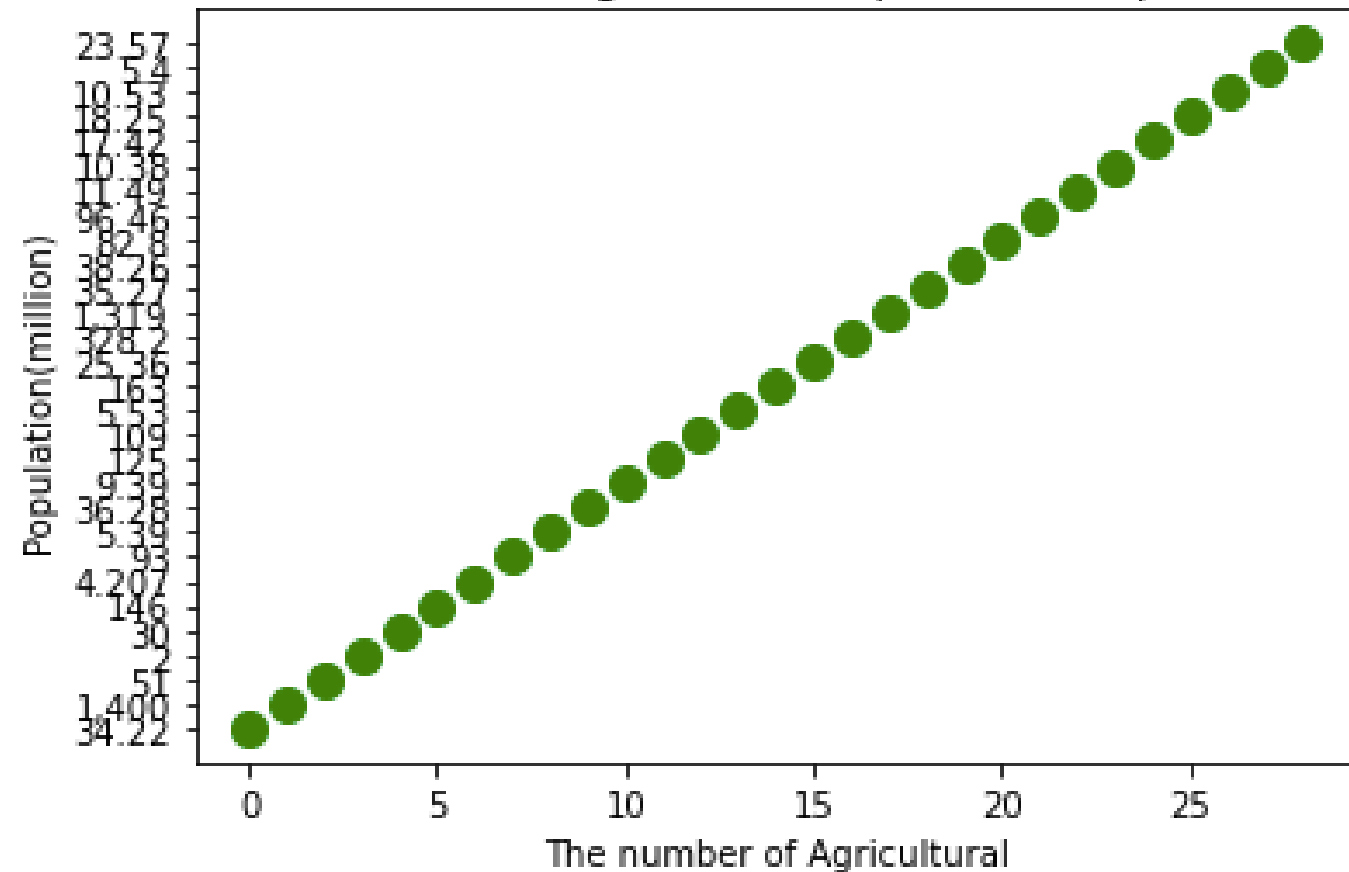
## จำนวนการนำเข้าสินค้ากับจำนวนประชากร



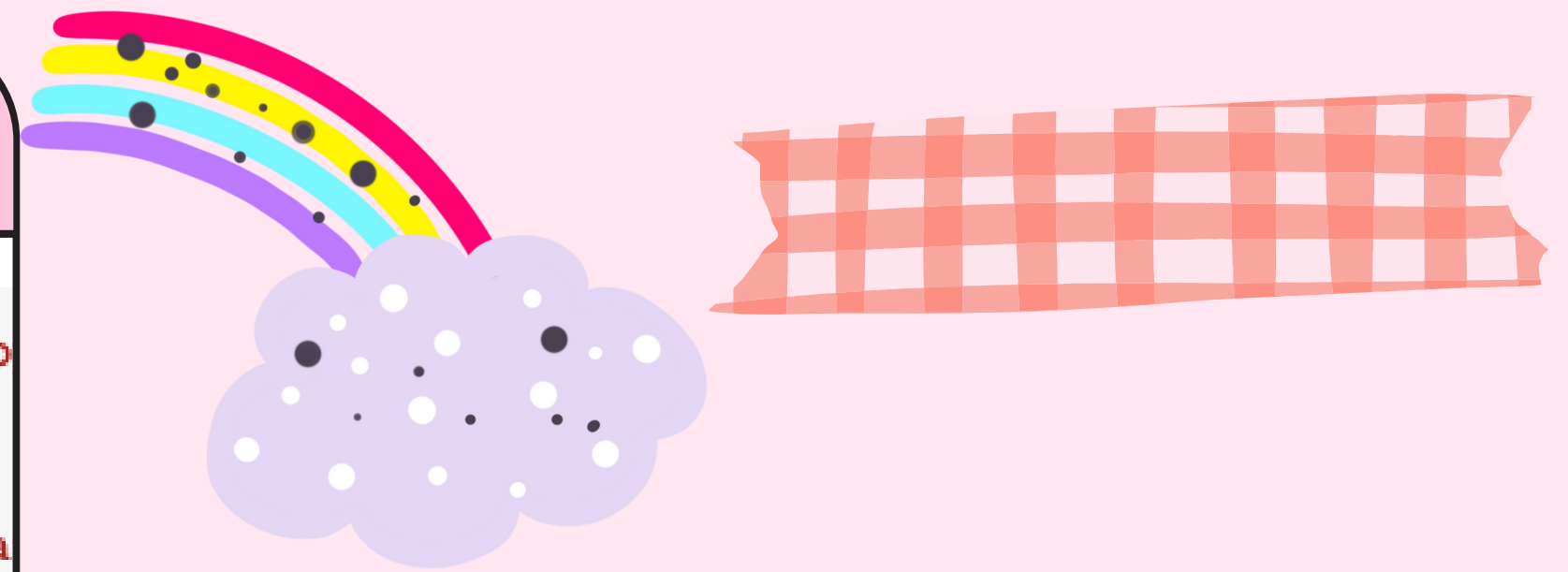
```
plt.scatter(range(len(data_new1['total'])), data_new1['popu  
plt.xlabel('The number of Agricultural')  
plt.ylabel('Population(million)')  
plt.title('The number of Agricultural imports and Popula  
plt.show()
```



The number of Agricultural imports and Population



```
[113] plt.plot(range(len(data_new1['total'])), data_new1['popul  
plt.xlabel('The number of Agricultural')
```



จากกราฟจะสามารถสรุปผลได้ว่า  
จำนวนประชากรเป็นอีกปัจจัยหนึ่งที่มี  
ผลต่อการผลิตสินค้าทางการเกษตร  
ถ้าประเทศที่มีประชากรจำนวนมากจะ  
มีความสามารถผลิตสินค้าทางการ  
เกษตรได้มากเช่นกัน หรือกล่าวได้ว่า  
ปริมาณการนำเข้าสินค้าทางการ  
เกษตรกับจำนวนประชากรมีความ  
สัมพันธ์ไปในทิศทางเดียวกัน

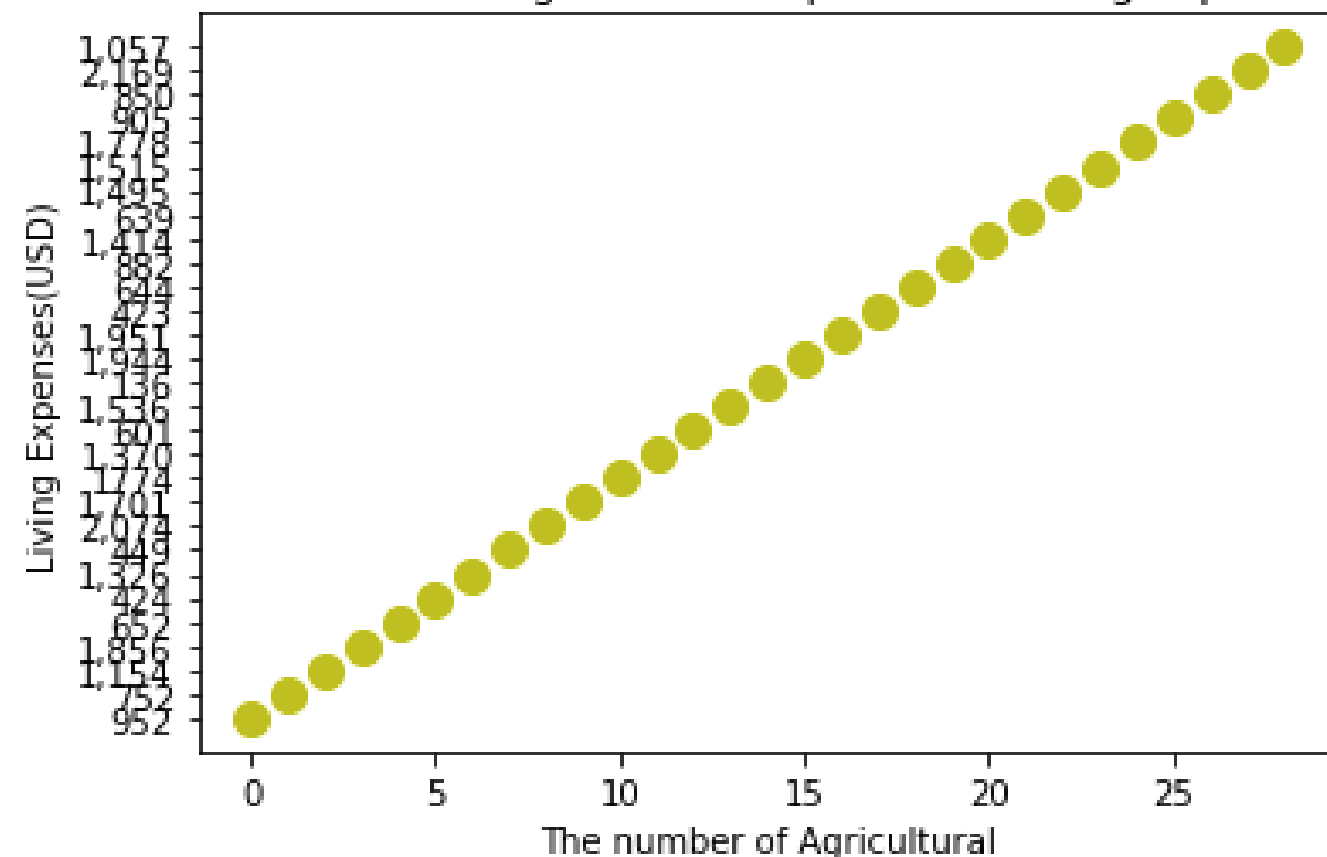
## จำนวนการนำเข้าสินค้าทางเกษตรกับ จำนวนค่าครองชีพ



```
plt.scatter(range(len(data_new1['total'])),data_new1['living ex  
plt.ylabel('Living Expenses(USD)')  
plt.xlabel('The number of Agricultural')  
plt.title('The number of Agricultural imports and Living Exp  
plt.show()
```



The number of Agricultural imports and Living Expenses



```
[115] plt.plot(range(len(data_new1['total'])),data_new1['living ex  
plt.ylabel('Living Expenses(USD)')  
plt.xlabel('The number of Agricultural')  
plt.title('The number of Agricultural imports and Living F
```

จากกราฟจะสามารถสรุปผลได้ว่า  
ค่าครองชีพเป็นอีกปัจจัยหนึ่งที่มีผล  
ต่อการผลิตสินค้าทางการเกษตร  
ถ้าประเทศที่มีค่าครองชีพสูงจะมี  
ความสามารถผลิตสินค้าทางการ  
เกษตรได้สูงเช่นกัน หรือกล่าวได้  
ว่า ปริมาณการนำเข้าสินค้าทางการ  
เกษตรกับค่าครองชีพมีความ  
สัมพันธ์ไปในทิศทางเดียวกัน



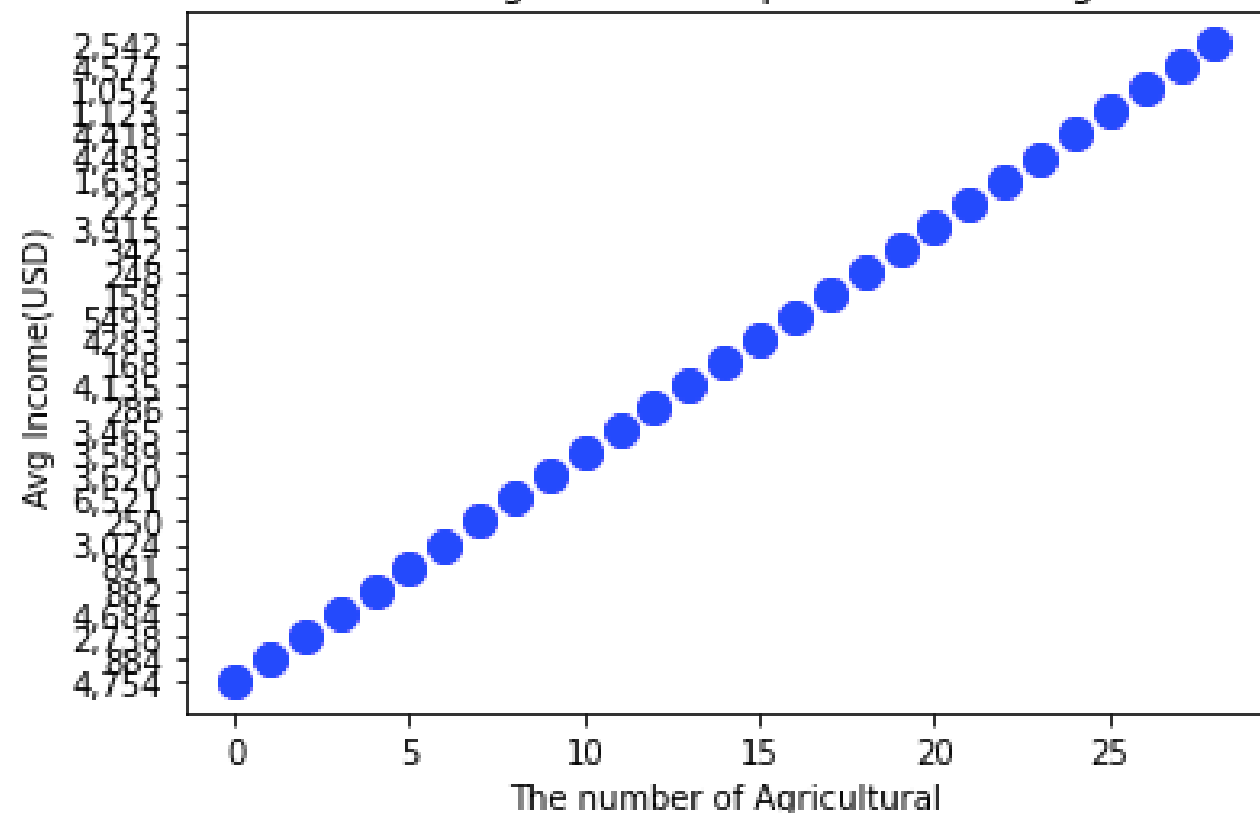
## จำนวนการนำเข้าสินค้าทางเกษตร กับรายได้เฉลี่ย



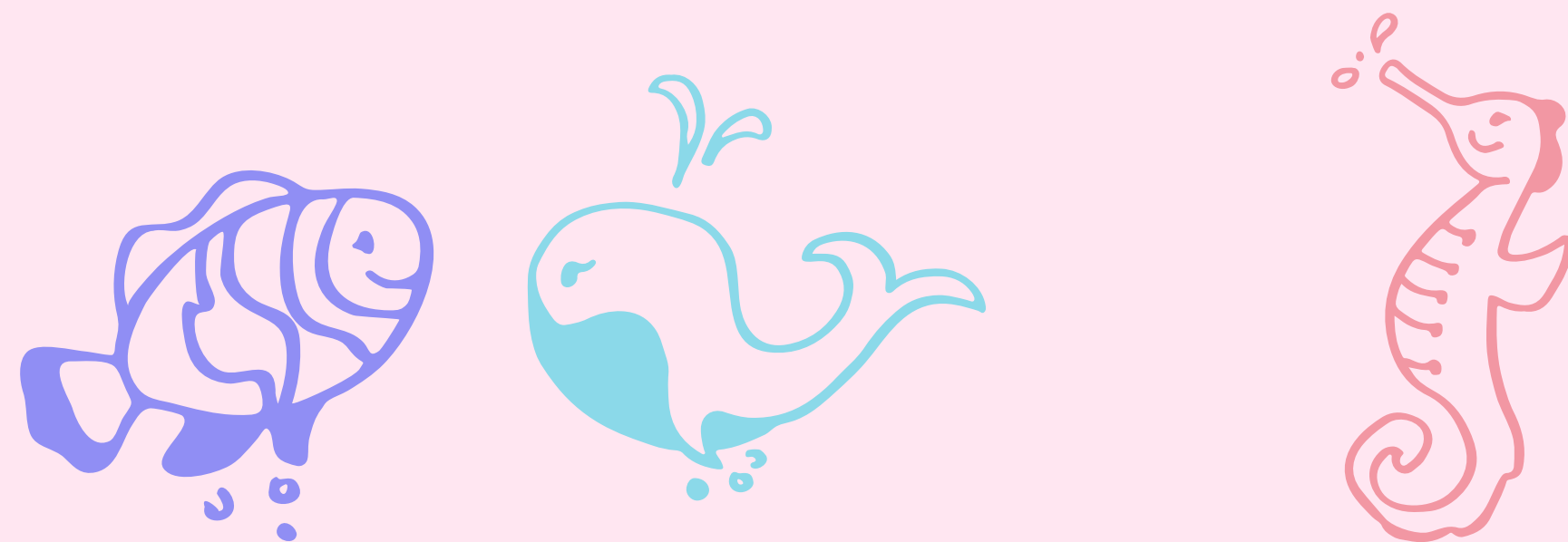
```
plt.scatter(range(len(data_new1['total'])),data_new1['avg income(USD)'])
plt.ylabel('Avg Income(USD)')
plt.xlabel('The number of Agricultural')
plt.title('The number of Agricultural imports and Average Income')
plt.show()
```



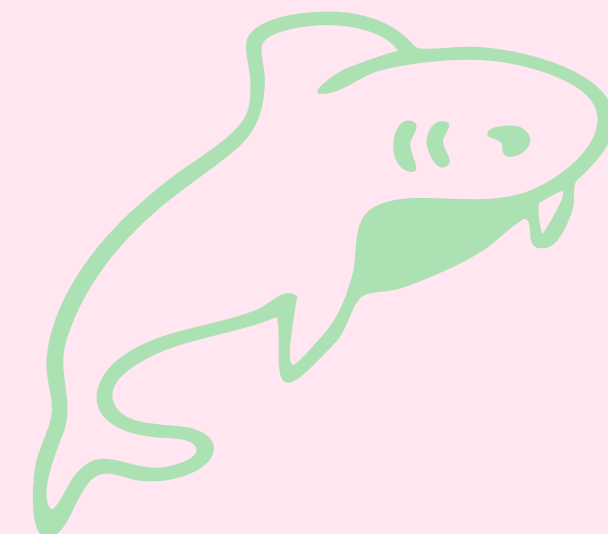
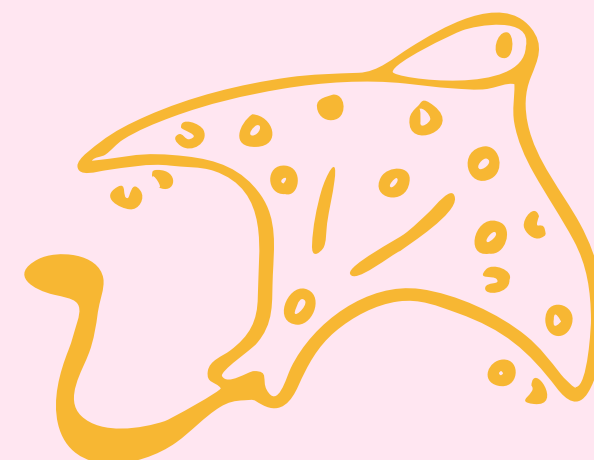
The number of Agricultural imports and Average Income



```
[117] plt.plot(range(len(data_new1['total'])),data_new1['avg income(USD)'])
plt.ylabel('Avg Income(USD)')
plt.xlabel('The number of Agricultural')
plt.title('The number of Agricultural imports and Average Income')
```



จากกราฟจะสามารถสรุปผลได้ว่า รายได้เฉลี่ยเป็นอีกปัจจัยหนึ่งที่มีผลต่อการผลิตสินค้าทางการเกษตร ถ้าประเทศที่มีรายได้เฉลี่ยสูงจะมีความสามารถผลิตสินค้าทางการเกษตรได้สูงเช่นกัน หรือกล่าวได้ว่า ปริมาณการนำเข้าสินค้าทางการเกษตรกับรายได้เฉลี่ยมีความสัมพันธ์ไปในทิศทางเดียวกัน





MARVEL STUDIOS

# BLACK WIDOW

สิ้นสุดภารกิจ นาทาชา

