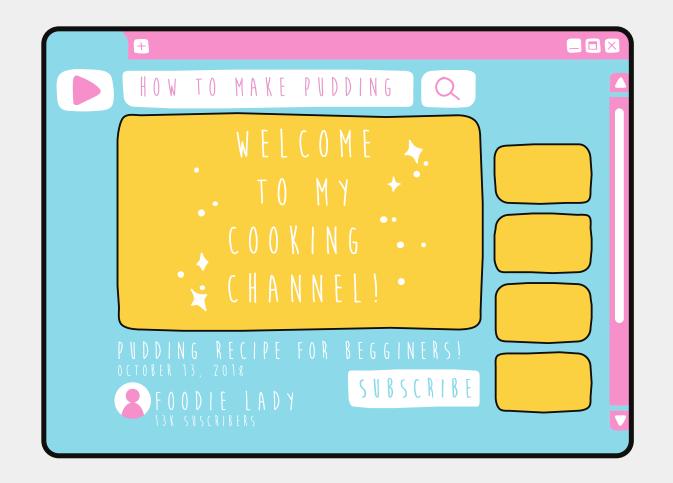


วัตถุประสงค์

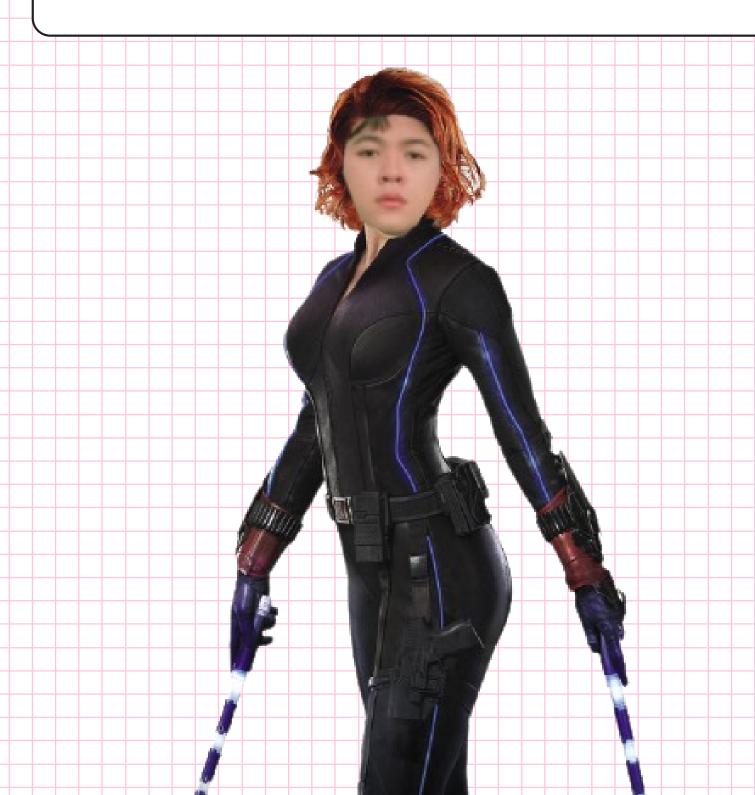




เพื่อทำนายว่าจำนวนประชากร รายได้ ค่าครองชีพ มีผลต่อการนำเข้าสินค้าทางการ เกษตรหรือไม่ โดยจะจัดกลุ่มปริมาณการนำ เข้าสินค้าให้อยู่ในกลุ่มน้อย หรือ กลุ่มมาก

- การนำเข้าปุ๋ยเคมี (รายประเทศ) องค์กร : ศูนย์ เทคโนโลยีสารสนเทศและการสื่อสาร กรม วิชาการเกษตร จาก : https://data.go.th
- การนำเข้าเมล็ดพันธุ์ควบคุม (ราย ประเทศ)องค์กร : ศูนย์เทคโนโลยีสารสนเทศและ การสื่อสาร กรมวิชาการเกษตรจาก : https://data.go.th
- สารป้องกันและกำจัดโรคพืชที่มีปริมาณการนำ เข้าสูง องค์กร : ศูนย์เทคโนโลยีสารสนเทศและ การสื่อสาร กรมวิชาการเกษตร จาก : https://data.go.th
- ข้อมูลทั่วไปของประเทศที่นำเข้าสินค้าทางการ เกษตร จาก https://drive.google.com/drive/folders /1PCwak7QUVizd6sfY-NvAZBRLNh5Nfhz?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	¥2006	¥2007	¥2008	3
0	1.0	ปริมาณ การนำ เข้าปุ๋ย เคมี ราย ประเทศ	SAUDI ARABIA	เมตริกตัน	ประเทศ	รายปี	582,636.00	787,997.00	672,800.00	1,007,2
1	2.0	ปริมาณ การนำ เข้าปุ๋ย เคมี ราย ประเทศ	CHINA	เมตริกตัน	ประเทศ	รายปี	270,060.53	675,840.83	569,348.49	427,9
2	3.0	ปริมาณ การนำ เข้าปุ๋ย เคมี ราย ประเทศ	KOREA REPUBLIC OF	เมตริกตัน	ประเทศ	รายปี	426,401.90	472,745.91	287,250.00	417,4
3	4.0	ปริมาณ การนำ เข้าปุ๋ย เคมี ราย ประเทศ	QATAR	เมตริกตัน	ประเทศ	รายปี	268,274.00	445,726.00	460,968.00	387,5
4	5.0	ปริมาณ การนำ เข้าปุ๋ย	MALAYSIA	เมตริกตัน	ประเทศ	รายปี	244,645.54	234,427.46	173,449.15	285,7

🗸 1 วินาที เสร็จสมบูรณ์เมื่อ 20:31

000

Preprocessing



์ ตารางที่ 2 ตารางปริมาณการนำเข้า_เ เมล็ดพันธ์ data_pc = pd.read_csv(os.path.join(path,'ปริมาณการนำเข้าเมล็ดพันธุ์.csv')) #การนำเข้าไท data_pc id Title Subtitle_PC Unit View Freq **Y2006** Y2007 Y2008 ปริมาณ การนำ SAUDI ARABIA เมตริกตัน ประเทศ รายปี 4,910.26 6,203.20 6,274.60 พันธุ์ ควบคุม ปริมาณ การนำ CHINA เมตริกตัน ประเทศ รายปี 124,916.36 47,946.76 46,686.16 เมล็ด พันธ์ ควบคุม ปริมาณ การนำ เมตริกตัน ประเทศ รายปี 35,847 59,459.38 165,267.43 เมล็ด พันธุ์ ควบคุม ปริมาณ การนำ QATAR เมตริกตัน ประเทศ รายปี 47,414.77 124,915.02 580,000.00 เมล็ด พันธ์ ควบคุม ปริมาณ การนำ MALAYSIA เมตริกตัน ประเทศ รายปี 37,645.03 46,467.47 23,567.58 เมล็ด พันธุ์ ควบคม

ตารางที่ 3 ตาราง สารป้องกันและ กำจัดโรคพืช data_pd = pd.read_csv(os.path.join(path,'สารป้องกันและกำจัดโรคพีช.csv')) #การนำเข้าไฟล์ที่มี data pd 8 Subtitle_PD Unit View Freq id Title **Y2007 Y2008 Y2009** สาร ป้องกัน และ กำจัด SAUDI ARABIA กิโลกรัม ประเทศ รายปี โรคพืช 2,249,315.00 2,143,682.00 1,192,758.30 ปริมาณ การนำ เข้าสง สาร ป้องกัน และ กำจัด CHINA กิโลกรัม ประเทศ รายปี โรคพืช 1,160,075.00 1,519,815.00 782,375.00 ปริมาณ การนำ เข้าสูง สาร ป้องกัน และ กำจัด กิโลกรัม ประเทศ รายปี โรคพืช 1,875,370.00 1,724,557.00 1,332,152.90 ปริมาณ การนำ เข้าสูง สาร ป้องกัน

1 วินาที เสร็จสมบูรณ์เมื่อ 20:31

me me	erg lům	ed_ erg	03	ata_pd.merge(เ			C_data	_pc,how='left	',left_on='S	ubtitle_PD',
		id	Title	a-14111						
				Subtitle_PD	Unit	View	Freq	¥2007	¥2008	¥2009
0)	1	สาร ป้องกัน และ กำจัด โรคพุช ที่มี ปริมาณ การนำ เข้าสูง	SAUDI ARABIA	กิ โลกรัม	ประเทศ	รายปี	2,249,315.00	2,143,682.00	1,192,758.30
1	ı	2	สาร ป้องกัน และ กำจัด โรคพุช ที่มี ปริมาณ การนำ เข้าสูง	CHINA	กิ โลกรัม	ประเทศ	รายปี	1,160,075.00	1,519,815.00	782,375.00
2	2	3	สาร ป้องกัน และ กำจัด โรคพืช ที่มี ปริมาณ การนำ เข้าสูง	KOREA REPUBLIC OF	กิ โลกรัม	ประเทศ	รายปี	1,875,370.00	1,724,557.00	1,332,152.90
			ุ สาร							

Preprocessing

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



Preprocessing



ตารางที่ 4 ตาราง ข้อมูลทั่วไปนำเข้าสินค้าทางเกษตร 🔲 🗖 🔀

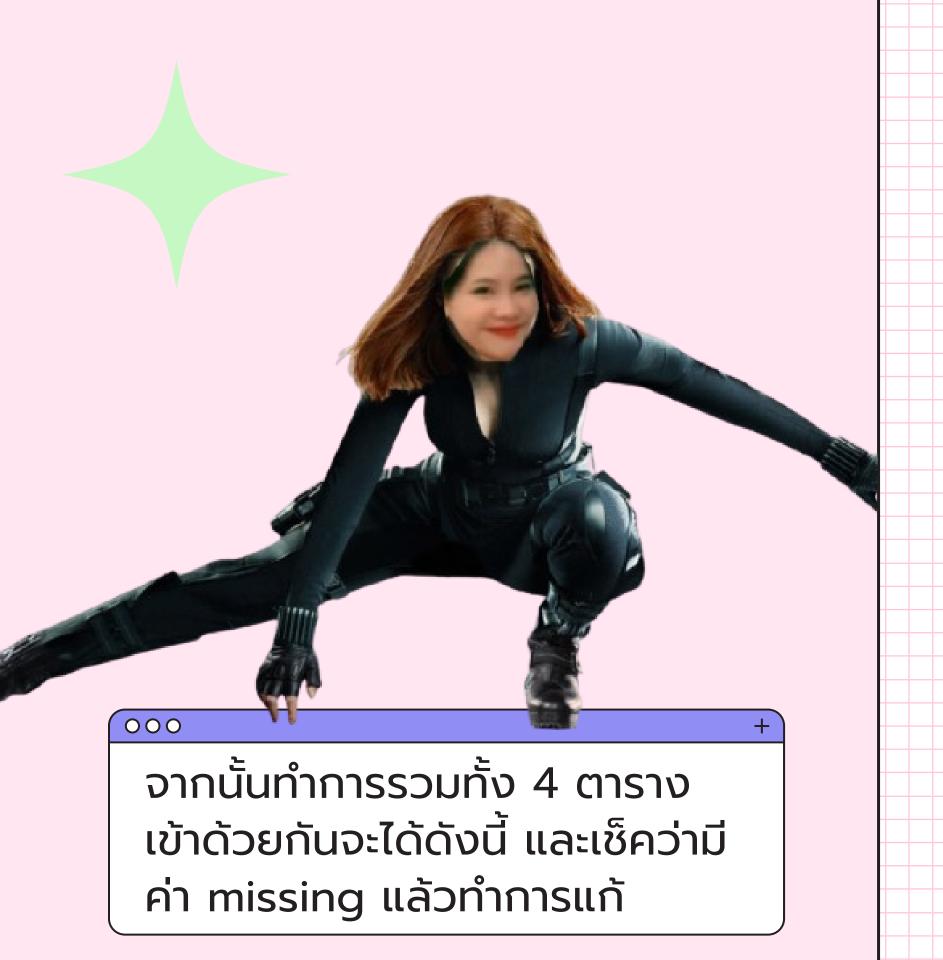
[27] path = '/content/drive/My Drive/Project'#โดยเราจะทำการระบุให้ซีไปที โฟนเดอร์ทีเราเซฟไฟล์ไ

[28] data_add = pd.read_csv(os.path.join(path,'ข้อมูลทั่วไปนำเข้าสินค้าทางเกษตร.csv')) data_add

	iđ	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	มินสค์/ มอสโก
6	7	KUWAIT	ประเทศ	4.207	อิสลาม	คูเวตชิตี
7	8	EGYPT	ประเทศ	93	อิสลาม	ไคโร
8	9	NORWAY	ประเทศ	5.39	คริสต์	ออสโล
9	10	CANADA	ประเทศ	36.28	โรมันคาธอลิก	ออตตาวา
10	11	ISRAEL	ประเทศ	9.39	ยูดาห์	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

ตัดคอลัมน์ที่10ออก เพราะมันเป็นค่า NAN

🗸 1 วินาที เสร็จสมบูรณ์เมื่อ 20:31



000 merged_04 = data_ADD.merge(merged_03,how='left',left_on='country merged_04 #add table that about coutry anothor country View_x population(million) religion id_x capit SAUDI ARABIA ประเทศ อิสลาม ไม่นับถือ

ปัญหา Classification

000

ปัญหาของเราคือ เรามีสินค้าทางการเกษตรในประเทศของเราไม่ เพียงพอ เราเลยอยากรู้ว่าในปีถัดไปเราควรสั่งสินค้าจากประเทศบ้าง โดยเราจะพิจารณาจากปริมาณการนำเข้า 3 ปีย้อนหลัง คือ 2018 2019 2020 นอกจากนี้เราอยากรู้อีกว่ามีปัจจัยใดบ้างที่ส่งผลต่อการผลิต สินค้าทางการเกษตรของประเทศที่เรานำเข้า ซึ่งเรามีปัจจัยที่คาดว่าจะ เกี่ยวข้องคือ จำนวนประชากร ค่าแรงโดยเฉลี่ย และค่าครองชีพโดย เฉลี่ย

ดังนั้น เราจึงได้ทำการวิเคราะห์ Data mining เพื่อทำนายว่าจำนวน ประชากร รายได้ ค่าครองชีพ มีผลต่อการนำเข้าสินค้าทางการเกษตร หรือไม่ โดยจะจัดกลุ่มปริมาณการนำเข้าสินค้าให้อยู่ในกลุ่มน้อย หรือ กลุ่มมาก



Classification

- 1.สร้างตารางที่เราจะนำไปจัดหมวดหมู่
- 2.ตรวจสอบขนาด
- 3.ตรวจสอบค่า missing ค่า NAN
- 4.Drop.NA
- 5.Cluster
- 6.นำตาราง Clusterไปรวมกับข้อมูลเก่า แล้วตัดแค่ข้อมูลที่เราต้องการ
- 7.ตรวจสอบ Type ของข้อมูลแล้วนำไปทำ โมเดล
- 8.ตรวจสอบ Type อีกครั้ง

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

9		total	population(million)	<pre>living expenses(USD)</pre>	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 มีผล ต่อการนำเข้าสินค้าทางการเกษตรว่าแต่ละประเทศจัดอยู่ในกลุ่มมากหรือน้อยได้ ซึ่งได้ค่าความแม่นยำทั้ง 3 ค่า คือ KNN 62.50%, Desition 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)
/usr/local/lib/python3.7/dist-packages/sklearn/model selection/ va
  estimator.fit(X train, y train, **fit params)
0.625
```

```
\times \Box -
from sklearn.ensemble import RandomForestClassif
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

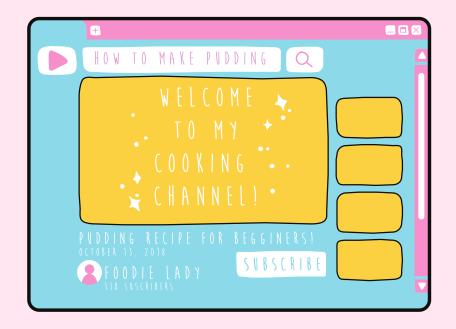
เราจะวัดผลความแม่นยำโดยใช้ cross_val_score เพราะเราจะ ได้ไม่ต้องแบ่งกลุ่ม data เอง IIUU accuracy_scrore cross-val จะแบ่งและวัดให้เลย ซึ่งสรุปการวัดผลได้ดังนี้ จากการทำทั้งสามโมเดลการ ทำนายพบว่า จำนวนประชากร รายได้ ค่าครองชีพ มีผลต่อ การนำเข้าสินค้าทางการเกษตร ว่าแต่ละประเทศจัดอยู่ในกลุ่ม มากหรือน้อยได้ ที่มีความความ แม่นยำมากที่สุดคือ RF โดยมี ความแม่นยำ 62.50% เท่ากัน

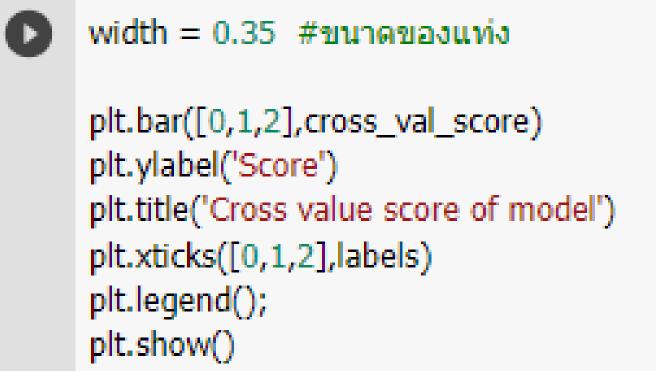
```
cross_val_score(KNN_1, X_train, y_train, cv=12).mean()
/usr/local/lib/python3.7/dist-packages/sklearn/model selectic
  % (min groups, self.n splits)), UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/model_selectic
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.7/dist-packages/sklearn/model selectic
cross_val_score(myTree, X_train, y_train, cv=12).mean()
/usr/local/lib/python3.7/dist-packages/sklearn/model sel
  % (min_groups, self.n_splits)), UserWarning)
0.58333333333333334
```

- cross_val_score(skRF, X_train, y_train, cv=12).mean()
- /usr/local/lib/python3.7/dist-packages/sklearn/model_se
 % (min_groups, self.n_splits)), UserWarning)
 /usr/local/lib/python3.7/dist-packages/sklearn/model_se
 estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.7/dist-packages/sklearn/model_se
 estimator.fit(X_train, y_train, **fit_params)

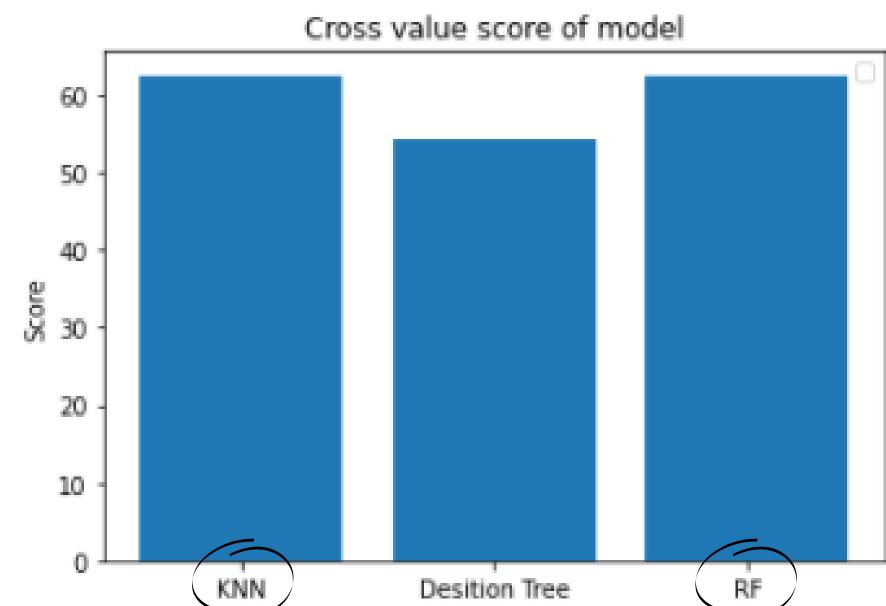


เราจะเห็นว่าโมเดล KNN และ Random Forest มี ความแม่นยำเท่ากัน



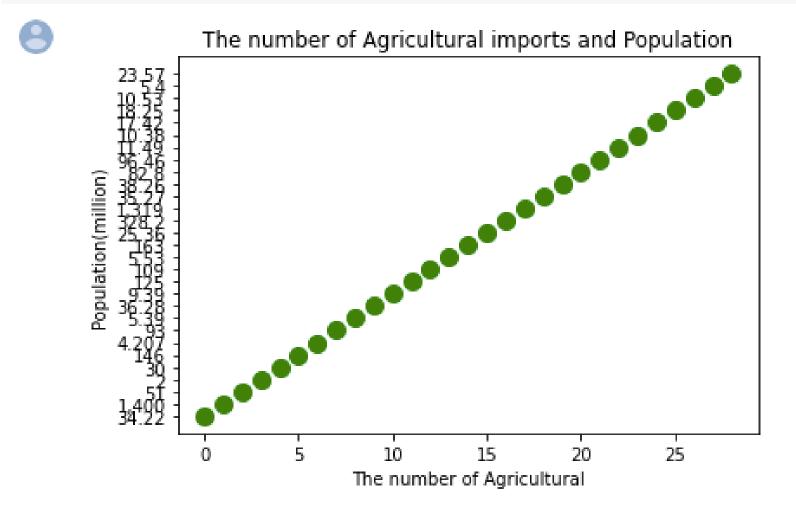


No handles with labels found to put in legend.

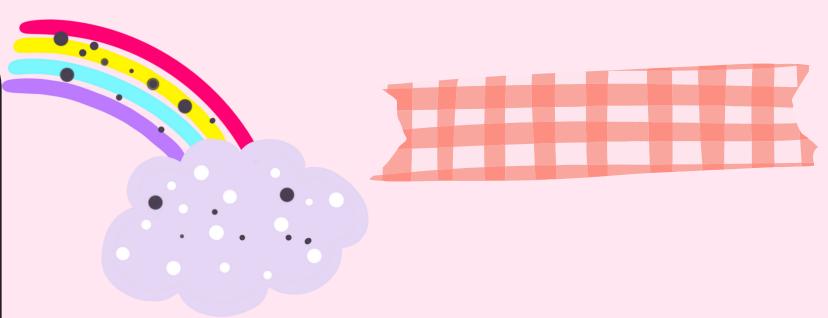


จำนวนการนำเข้าสินค้ากับจำนวนประชากร 🔲 🗖 🛭

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



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

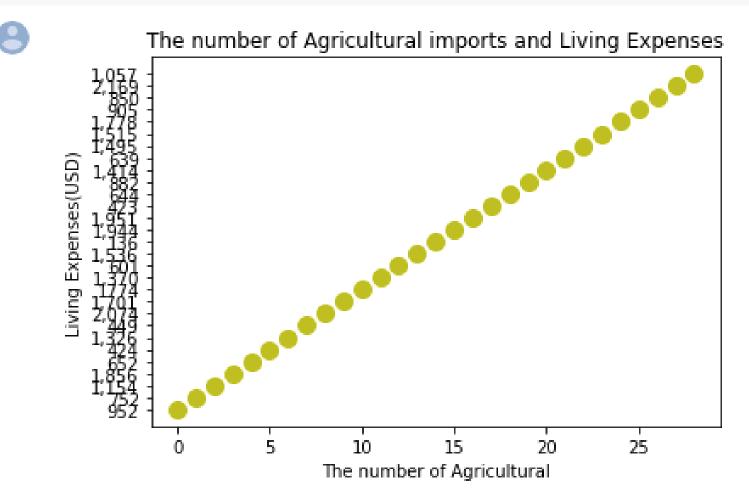


จากกราฟจะสามารถสรุปผลได้ว่า จำนวนประชากรเป็นอีกปัจจัยนึงที่มี ผลต่อการผลิตสินค้าทางการเกษตร ถ้าประเทศที่มีประชากรจำนวนมากจะ มีความสามารถผลิตสินค้าทางการ เกษตรได้มากเช่นกัน หรือกล่าวได้ว่า ปริมาณการนำเข้าสินค้าทางการ เกษตรกับจำนวนประชากรมีความ สัมพันธ์ไปในทิศทางเดียวกัน

จำนวนการนำเข้าสินค้าทางเกษตรกับ จำนวนค่าครองชีพ



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



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

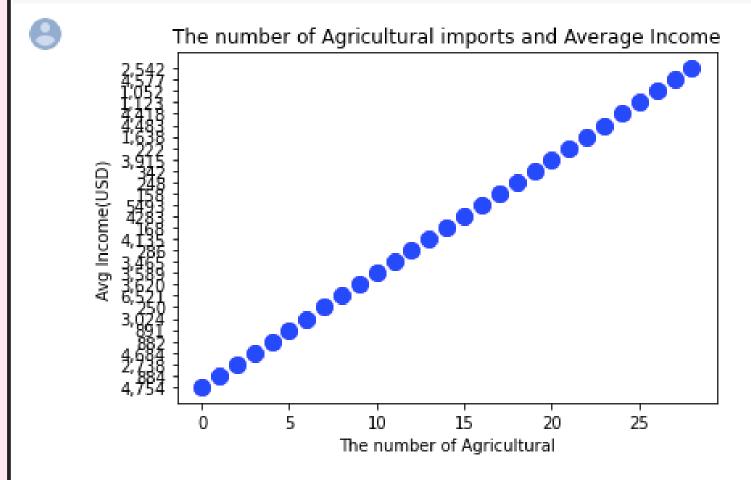


จากกราฟจะสามารถสรุปผลได้ว่า ค่าครองชีพเป็นอีกปัจจัยนึงที่มีผล ต่อการผลิตสินค้าทางการเกษตร ถ้าประเทศที่มีค่าครองชีพสูงจะมี ความสามารถผลิตสินค้าทางการ เกษตรได้สูงเช่นกัน หรือกล่าวได้ ว่า ปริมาณการนำเข้าสินค้าทางการ เกษตรกับค่าครองชีพมีความ สัมพันธ์ไปในทิศทางเดียวกัน

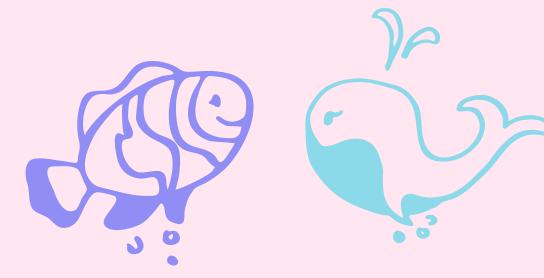
จำนวนการนำเข้าสินค้าทางเกษตร กับรายได้เฉลี่ย



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



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





จากกราฟจะสามารถสรุปผลได้ว่า รายได้เฉลี่ยเป็นอีกปัจจัยนึงที่มีผลต่อการ ผลิตสินค้าทางการเกษตร ถ้าประเทศที่มี รายได้เฉลี่ยสูงจะมีความสามารถผลิตสินค้า ทางการเกษตรได้สูงเช่นกัน หรือกล่าวได้ ว่า ปริมาณการนำเข้าสินค้าทางการเกษตร กับรายได้เฉลี่ยมีความสัมพันธ์ไปในทิศทาง เดียวกัน

