

# ข้อมูลมูลค่าการนำเข้าสินค้าของประเทศไทยจากทั่วโลกในช่วงปี 2002 – 2022

## Dataset

ข้อมูลมูลค่าการนำเข้าสินค้าหน่วยเป็น Us Dollar Thousand มีข้อมูล 22 column และ 1259 row โดย column นั้นประกอบไปด้วย Code(รหัสสินค้า) , Product label(ชื่อสินค้า) และปี 2002 ถึง 2022 และ row นั้นประกอบไปด้วยสินค้า 1259 ประเภท (โดยรวม Total แล้ว)

## Question

- 1.การเปลี่ยนแปลงของพฤติกรรมการนำเข้าสินค้าของประเทศไทยว่าจากในอดีตจนถึงปัจจุบันยังมีการเปลี่ยนแปลงอะไรบ้าง
- 2.ประเทศไทยมีนำเข้าสินค้ากลุ่มใดเป็นมูลค่าสูงสุดและคิดเป็นสัดส่วนเท่าไร

## Process

ทำการเช็ค missing data

ทำการเปลี่ยน ข้อมูล ใน column 'Code' จากเลข 4 หลักเป็นเลข 2 หลักเพื่อจัดแบ่งข้อมูลเป็นกลุ่ม ๆ โดยใช้วิธี replace เพื่อที่จะ Group ข้อมูลที่มีรหัสสินค้ากลุ่มเดียวกันเข้าด้วยกันและเพื่อนำไปหาว่าสินค้ากลุ่มไหนมีการเปลี่ยนแปลงเพิ่มขึ้นและลดลงมากที่สุดด้วยการเทียบปี 2021 กับปี 2002 โดยใช้วิธีดังนี้

```
#Change "CODE" ปรับ code ให้เป็นเลข 2 หลักเพื่อหามูลค่านำเข้ารวมของสินค้ากลุ่มเดียวกัน
lst1 = ['01','02','03','04','05','06','07','08','09']
lst2 = [str(x) for x in range(10,100)]
lst3 = lst1+lst2

df1 = df.loc[:,['Code','Imported value in 2002','Imported value in 2021']]

import re

for i in lst3:
    df1 = df1.replace({'Code':f'{i}..'},{f'Code':i},regex = True) #เปลี่ยน code 4 หลักให้เป็น 2 หลัก

df1 = df1.sort_values(['Code'],ascending=True)
df1 = df1.drop(0).reset_index(drop=True)
display(df1)

df2 = df1.groupby('Code')[['Imported value in 2002','Imported value in 2021']].sum() #หาผลรวม column ของสินค้ากลุ่มเดียวกัน
display(df2)
```

จากนั้นทำให้พบปัญหาขึ้น

## ปัญหาที่ 1

ทำการหาการเปลี่ยนแปลงของมูลค่าการนำเข้าสินค้าแต่ละกลุ่มในปี 2021 เทียบกับปี 2002 พบว่าการที่นำข้อมูลมูลค่ามาใช้ทำให้ไม่สามารถหาการเปลี่ยนแปลงการนำเข้าได้จริง ๆ เนื่องจากมูลค่าเงินมีการเพิ่มขึ้นตามอัตราเงินเพื่อ

ผลลัพธ์ที่ได้จากการหาสินค้าที่มีมูลค่าการนำเข้าที่เปลี่ยนแปลงมากที่สุด

```
# อยากทราบว่าสินค้ากลุ่มไหนมีการเปลี่ยนแปลงมาสุดใน 20 ปี
df2['different(%)'] = df2[['Imported value in 2002','Imported value in 2021']]\
| .apply(lambda x: ((x['Imported value in 2021']-x['Imported value in 2002'])/x['Imported value in 2002'])*100,axis = 1)
High = df2.sort_values(['different(%)'],ascending=False) #ดูว่าสินค้ากลุ่มไหนเปลี่ยนแปลงเพิ่มขึ้นมากที่สุด
display(High.head(10))
Low = df2.sort_values(['different(%)'],ascending=True) #ดูว่าสินค้ากลุ่มไหนเปลี่ยนแปลงลดลงมากที่สุด
display(Low.head(5))
# เห็นว่าการเปลี่ยนแปลงเพิ่มขึ้นและลดลงตัวเลขแปลก ๆ เพราะลืมคิดไปว่ามูลค่าการนำเข้านั้นมีการเพิ่มขึ้นตามเงินเฟ้อในช่วง 20 ปีอยู่ด้วย
```

	Imported value in 2002	Imported value in 2021	different(%)
Code			
02	5033	202842	3930.240413
07	25934	1011568	3800.547544
43	9685	361409	3631.636551
86	9067	260613	2774.302415
06	2877	72306	2413.242961
09	17986	448232	2392.116090
65	1896	46403	2347.415612
08	61712	1264458	1948.966165
97	391	7804	1895.907928
93	8930	166012	1759.036954

	Imported value in 2002	Imported value in 2021	different(%)
Code			
99	542675	0	-100.000000
51	121572	74053	-39.087125
50	9773	6228	-36.273406
52	752858	567642	-24.601718
53	18367	15340	-16.480645

## แก้ปัญหาที่ 1

ทำการแปลงข้อมูลใหม่ให้เทียบเป็นสัดส่วนจากการนำเข้าทั้งหมดแทน

จากนั้นจึงกลับมาทำการแปลงข้อมูลจาก US dollar thousand ให้อยู่ในรูปของสัดส่วนเมื่อเทียบกับทั้งหมด โดยวิธีการดังนี้

```
#ทำการปรับข้อมูลให้อยู่ในรูปของ ratio เพื่อง่ายต่อการเปรียบเทียบ
df_to_ratio1 = df

df_to_ratio1 = df_to_ratio1.sort_values(['Code'],ascending=True)
df_to_ratio1 = df_to_ratio1.drop(0).reset_index(drop=True)
df_to_ratio1_drop = df_to_ratio1.drop(['Code','Product label'],axis = 1)

df_to_ratio2 = df_to_ratio1_drop.apply(lambda x: x/sum(x)*100, axis=0)

df_ratio = pd.concat( [df_to_ratio1, df_to_ratio2], ignore_index=True,axis=1 )

lst1 = [x for x in range(2,22)]

for i in lst1:
    df_ratio = df_ratio.drop(i,axis=1)

lst2 = [str(x) for x in range(2002,2022)]
df_ratio.columns = ['Code','Product label']+lst2
display(df_ratio)
```

จากนั้นจึงทำการหาสัดส่วนการนำเข้าของสินค้ากลุ่มเดียวกันและหาการเปลี่ยนแปลงของปี 2021 เทียบกับปี 2002 ในส่วนนี้จะเลือกมาแค่ข้อมูล 3 column คือ Code , ปี 2002 และ ปี 2021 จากนั้นทำการหาผลรวมของสินค้าที่อยู่กลุ่มเดียวกัน และทำการ drop ข้อมูลในส่วน of commodities not elsewhere specific ออกและเมื่อได้ข้อมูลเสร็จนำมาหาการเปลี่ยนแปลงระหว่างปี 2021 กับปี 2002 โดยเลือกกลุ่มที่มีการเปลี่ยนแปลงเพิ่มขึ้นสูงสุด 3 กลุ่มและลดลงสูงสุด 3 กลุ่ม หลังจากนั้นนำ Data frame ที่มีชื่อของกลุ่มนั้นมา join เพื่อจะรู้ว่า Code นั้น ๆ อยู่กลุ่มอะไร

โดยที่กล่าวมามีวิธีทำดังนี้

```
#หาสัดส่วนการนำเข้าของสินค้ากลุ่มเดียวกันและหาการเปลี่ยนแปลงของปี 2021 เทียบกับ ปี 2002
df_diff = df_ratio_2digit.loc[:,['Code','2002','2021']]
df_diff = df_diff.sort_values(['Code'],ascending=True)
df_diff = df_diff.drop(0).reset_index(drop=True)
df_diff = df_diff.groupby('Code')[['2002','2021']].sum() #หาผลรวม column ของสินค้ากลุ่มเดียวกัน
display(df_diff)
#ทำการ drop ข้อมูลในส่วน of Commodities not elsewhere specific ออก
df_diff = df_diff.drop(df_diff.index[len(df_diff)-1])

# อยากทราบว่าสินค้ากลุ่มไหนมีการเปลี่ยนแปลงมาสุดใน 20 ปี
df_diff['different(%)'] = df_diff[['2002','2021']]\
|.apply(lambda x: ((x['2021']-x['2002'])/x['2002'])*100,axis = 1)
increase = df_diff.sort_values(['different(%)'],ascending=False) #ดูว่าสินค้ากลุ่มไหนเปลี่ยนแปลงเพิ่มขึ้นมากที่สุด
increase = increase.loc[:, 'different(%)']
decrease = df_diff.sort_values(['different(%)'],ascending=True) #ดูว่าสินค้ากลุ่มไหนเปลี่ยนแปลงลดลงมากที่สุด
decrease = decrease.loc[:, 'different(%)']

#ทำการ join ข้อมูลเพื่อให้รู้ว่า สินค้าเพิ่มขึ้นและลดลง อย่างละ 3 กลุ่ม มีอะไรบ้าง
df_diff_increase_product = pd.merge(df_group_name,increase.head(3),left_on='Code',right_on='Code',how='right')
display(df_diff_increase_product)

df_diff_decrease_product = pd.merge(df_group_name,decrease.head(3),left_on='Code',right_on='Code',how='right')
display(df_diff_decrease_product)

#เก็บค่าไว้เป็น array
code_increase = df_diff_increase_product.loc[ : , 'Code'] #เก็บ code ของกลุ่มที่เพิ่มขึ้น
code_increase_label = df_diff_increase_product.loc[ : , 'Product label'] #เก็บ product label ของกลุ่มที่เพิ่มขึ้น

code_decrease = df_diff_decrease_product.loc[ : , 'Code'] #เก็บ code ของกลุ่มที่ลดลง
code_decrease_label = df_diff_decrease_product.loc[ : , 'Product label'] #เก็บ product label ของกลุ่มที่ลดลง
```

ทำการ plot กราฟในส่วนของกลุ่มสินค้าที่มีการเปลี่ยนแปลงเพิ่มขึ้นและลดลงเพื่อดูว่าในกลุ่มนั้นมีสินค้าอะไรบ้างที่มีการเปลี่ยนแปลงอย่างเห็นได้ชัด โดยมีวิธีการดังนี้

```
#ทำการ plot กราฟ ในส่วน of กลุ่มสินค้าที่มีการเปลี่ยนแปลงเพิ่มขึ้นเพื่อดูว่าในกลุ่มสินค้านั้นมีสินค้าอะไรบ้างที่มีการเปลี่ยนแปลงอย่างเห็นได้ชัด
code_increase #ตัวแปรที่เก็บค่า code ของกลุ่มสินค้าที่เพิ่มขึ้น
code_increase_label #ตัวแปรที่เก็บค่า product label ของกลุ่มที่เพิ่มขึ้น
for i,k in zip(code_increase,code_increase_label):
    print(f'กลุ่มสินค้า(Code):{i}')
    df_for_plot1 = df_value.loc[df_value['Code'].str.contains(f'^{i}', case=False, regex=True)] #'^01' จะใช้ loop ในการ plot
    mean = int(df_for_plot1[df_for_plot1.columns[-1]].mean())#หาค่า mean ของ column สุดท้ายเนื่องจากเป็นปีล่าสุด
    #filter value ที่มากกว่า mean เท่านั้นเนื่องจากกราฟเอะเกินไปและสินค้าบางชนิดในได้มีมูลค่าที่สูงพอที่จะน่าสนใจ
    df_for_plot1 = df_for_plot1[ df_for_plot1[df_for_plot1.columns[-1]] >= mean ]
    display(df_for_plot1)
    df_for_plot2 = df_for_plot1.loc[:,df_for_plot1.columns != 'Product label']
    df_for_plot2.set_index('Code',inplace= True)
    df_for_plot2 = df_for_plot2.transpose()
    df_for_plot2.index.names = ['Year']

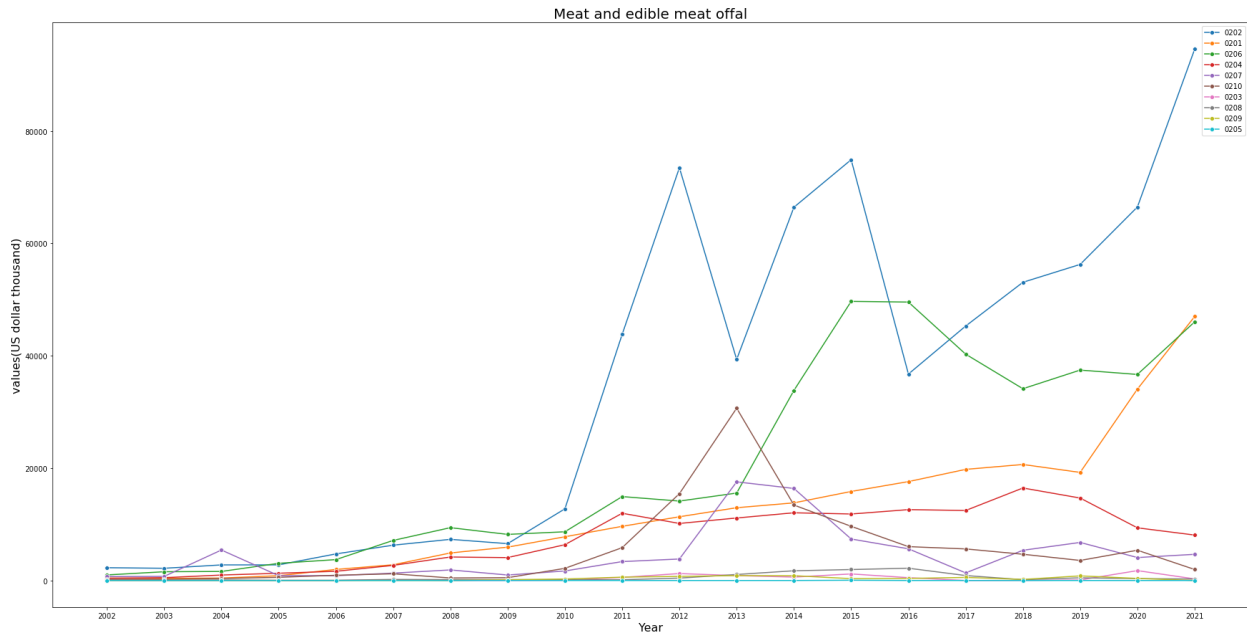
    code4_increase = df_for_plot1.loc[ : , 'Code']

    fig, ax = plt.subplots(figsize=(30, 15))
    plt.title(k, fontsize = 20)
    for j in code4_increase:
        sns.lineplot(data = df_for_plot2, x = 'Year', y = j, marker = "o", label = j)
    plt.xlabel("Year", size=16)
    plt.ylabel("values(US dollar thousand)", size=16)
    plt.legend(loc='best')
    plt.show()
```

โดยในขั้นตอนนี้พบปัญหา

## ปัญหาที่ 2

กราฟของสินค้าต่าง ๆ ที่อยู่ในกลุ่มมันเยอะเกินไปและสินค้าบางชนิดไม่ได้มีมูลค่าสูงพอที่จะน่าสนใจ



## แก้ปัญหาที่ 2

โดยการกรองข้อมูลในปีล่าสุดของสินค้าในกลุ่มเดียวกันว่า ถ้าในปีล่าสุดมีสินค้าชนิดไหนมีค่าน้อยกว่าค่าเฉลี่ยของกลุ่มจะไม่นำมาพิจารณา โดยมีวิธีแก้ปัญหาดังนี้

```
mean = int(df_for_plot1[df_for_plot1.columns[-1]].mean())#หาค่า mean ของ column สุดท้ายเนื่องจากเป็นปีล่าสุด
#filter value ที่มากกว่า mean เท่านั้นเนื่องจากกราฟเยอะเกินไปและสินค้าบางชนิดไม่ได้มีมูลค่าที่สูงพอที่จะน่าสนใจ
df_for_plot1 = df_for_plot1[ df_for_plot1[df_for_plot1.columns[-1]] >= mean ]
```

ต่อมาได้ทำการหาว่าสินค้ากลุ่มใดมีส่วนการนำเข้ามากที่สุด 5 อันดับแรก โดยเฉลี่ย 5 ปีย้อนหลังและสินค้านั้นคิดเป็นมูลค่าการนำเข้าเท่าไร ในขั้นตอนนี้พบปัญหาคือ

## ปัญหาที่ 3

มีการเขียน code เดิมซ้ำ ๆ บ่อย ในส่วนของการปรับค่าของ 'code' ให้อยู่ในรูปของเลข 2 หลัก

## แก้ปัญหาที่ 3

ย้อนกลับไปแก้ไขและสร้างตัวแปรที่เก็บ data frame ของมูลค่าและสัดส่วนที่เก็บค่าของ 'code' เป็นเลข 2 หลักเพิ่มขึ้นมาเนื่องจากง่ายต่อการนำไปใช้ต่อไม่ต้องมาเขียนซ้ำ ๆ โดยมีวิธีการแก้ดังนี้

```
#เปลี่ยน code ใน data frame มูลค่าและสัดส่วน ให้เป็นเลข 2 หลัก
lst1 = ['01','02','03','04','05','06','07','08','09']
lst2 = [str(x) for x in range(10,100)]
lst3 = lst1+lst2
df_value_2digit = df_value.sort_values(['Code'],ascending=True)

for i in lst3:
    df_value_2digit = df_value_2digit.replace({'Code':f'^{i}\\.\\.'},{'Code':i},regex = True) #เปลี่ยน code 4 หลักให้เป็น 2 หลัก
display(df_value_2digit)

df_ratio_2digit = df_ratio.sort_values(['Code'],ascending=True)
for i in lst3:
    df_ratio_2digit = df_ratio_2digit.replace({'Code':f'^{i}\\.\\.'},{'Code':i},regex = True)
display(df_ratio_2digit)
```

หลังจากสร้างตัวแปรที่เก็บจึงได้ไปทำการแก้ไขในส่วนอื่น ๆ ที่ทำไปแล้วด้วย และจากนั้นจึงเริ่มมาหาว่าสินค้ากลุ่มใดมีส่วนการนำเข้ามากที่สุด 5 อันดับแรก โดยเฉลี่ย 5 ปีย้อนหลังและสินค้านั้นคิดเป็นมูลค่าการนำเข้าเท่าไร โดยมีวิธีการหาดังนี้

```
#สินค้ากลุ่มใดมีส่วนในการนำเข้ามากที่สุด 5 อันดับแรก โดยเฉลี่ย 5 ปีย้อนหลัง
#สินค้ากลุ่มนั้นคิดเป็นมูลค่าการนำเข้าเท่าไร
df_ratio_2digit
df_ratio_5year = df_ratio_2digit.loc[:,['Code','Product label','2017','2018','2019','2020','2021']]
#ทำการ drop ข้อมูลในส่วนของ Commodities not elsewhere specific ออก
df_ratio_5year = df_ratio_5year.drop(df_ratio_5year.index[len(df_ratio_5year)-1])

#หาค่าเฉลี่ยของสัดส่วนมูลค่าการนำเข้า 5 ปีย้อนหลัง
df_ratio_5year = df_ratio_5year.sort_values(['Code'],ascending=True)
df_ratio_5year = df_ratio_5year.groupby('Code')[['2017','2018','2019','2020','2021']].sum() #หาผลรวม column ของสินค้ากลุ่มเดียวกัน
df_ratio_5year['5 years mean(ratio)'] = df_ratio_5year.agg(['mean'], axis=1 ).sort_values(['mean'],ascending=False)
df_ratio_5year_mean = df_ratio_5year.loc[:,['5 years mean(ratio)']]

#ทำการ join ข้อมูลเพื่อให้รู้ค่า สินค้าที่มีสัดส่วนเฉลี่ยมีอะไรบ้าง
df_ratio_5year_mean_product = pd.merge(df_group_name,df_ratio_5year_mean,left_on='Code',right_on='Code',how='right')
df_ratio_5year_mean_product = df_ratio_5year_mean_product.sort_values(['5 years mean(ratio)'],ascending=False) #sort value มากไปน้อย

#ทำการหามูลค่าการนำเข้าสินค้า 5 อันดับแรก โดยเฉลี่ย 5 ปีย้อนหลัง
#ทำการ group dataframe df_value_2digit ด้วย Code
df_value_5year = df_value_2digit.groupby('Code')[['2017','2018','2019','2020','2021']].sum()
df_value_5year['5 years mean(US dollar thousand)'] = df_value_5year.agg(['mean'], axis=1 ).sort_values(['mean'],ascending=False)
df_value_5year_mean = df_value_5year.loc[:,['5 years mean(US dollar thousand)']]

#ทำการรวม dataframe df_ratio_5year_mean_product_top5 เข้ากับ df_value_5year_mean เพื่อดูมูลค่าเฉลี่ยของสินค้านั้น
df_ratio_value_5year_mean = pd.merge(df_ratio_5year_mean_product,df_value_5year_mean,left_on='Code',right_on='Code',how='left')
df_ratio_value_5year_mean_top5 = df_ratio_value_5year_mean.head(5) #เลือกกลุ่มที่มีมูลค่าการนำเข้าสูงสุด 5 อันดับ
display(df_ratio_value_5year_mean_top5)

#เก็บค่าตัวแปร
code_high = df_ratio_value_5year_mean_top5.loc[:, 'Code'] #เก็บ code ของกลุ่มที่เพิ่มขึ้น
code_high_label = df_ratio_value_5year_mean_top5.loc[:, 'Product label'] #เก็บ product label ของกลุ่มที่เพิ่มขึ้น
```

เมื่อหาค่าเสร็จนำที่ได้นำไปสร้างกราฟเส้นเพื่อดูทิศทางหรือแนวโน้มของสินค้ากลุ่มนั้นในช่วง 10 ปี

```
#plot กราฟเพื่อดูเทรนของสินค้า 5 กลุ่มนี้
for i,k in zip(code_high,code_high_label):
    print(f'กลุ่มสินค้า(Code):{i}')
    df_for_plot1 = df_value_2digit.loc[df_value_2digit['Code'].str.contains(f'^{i}', case=False, regex=True)] #'^01' จะใช้ loop ในการ plot
    df_for_plot1 = df_for_plot1.loc[:,['Code','2012','2013','2014','2015','2016','2017','2018','2019','2020','2021']]
    df_for_plot1 = df_for_plot1.groupby('Code').sum()

    df_for_plot2 = df_for_plot1.loc[:,df_for_plot1.columns != 'Product label']
    df_for_plot2 = df_for_plot2.transpose()
    df_for_plot2.index.names = ['Year']

    fig, ax = plt.subplots(figsize=(20, 10))
    plt.title(k, fontsize = 20)
    sns.lineplot(data = df_for_plot2, x = 'Year', y = i, marker = "o", label = j)
    plt.xlabel("Year", size=16)
    plt.ylabel("values(US dollar thousand)", size=16)
    plt.legend(loc='best')
    plt.show()
```

ผลที่ได้จากการทำหุคข้อมูลนี้ :

### 1.การเปลี่ยนแปลงของพฤติกรรมกรนำเข้าสินค้าของประเทศไทยว่าจากในอดีตจนถึงปัจจุบันยังมีการเปลี่ยนแปลงอะไรบ้าง

พบว่ากลุ่มสินค้าที่มีการเปลี่ยนแปลงเพิ่มขึ้นมากที่สุด 3 อันดับแรก ได้แก่

อันดับที่ 1 02: Meat and edible meat offal(เนื้อสัตว์และเครื่องในสัตว์ที่กินได้) มีการเพิ่มขึ้นของสัดส่วนอยู่ที่ 868% โดยสินค้าที่มีมูลค่าสูงสุดในกลุ่มนี้ได้แก่ เนื้อวัวแช่แข็งเนื้อวัวสดหรือแช่เย็น เครื่องในที่กินได้ของวัว หมู แกะ แพะม้า ลาทั้งแบบ สด แช่เย็นและแช่แข็ง เป็นต้น

อันดับที่ 2 07: Edible vegetables and certain roots and tubers(ผักที่กินได้และผักที่เป็นรากและหัว) มีการเพิ่มขึ้นของสัดส่วนอยู่ที่ 837% โดยสินค้าที่มีมูลค่าสูงสุดในกลุ่มนี้ได้แก่

1. รากและหัวของมันสำปะหลัง เผือก ถั่วลิสงไม่ แก่นตะวัน มันเทศและผักชนิดที่เกี่ยวข้องที่มีปริมาณแป้งหรืออินนูลินสูง

2. ผักอื่น ๆ ที่สดหรือแช่เย็นยกเว้นพวกมันฝรั่ง มะเขือเทศ แครอท หัวผักกาด เป็นต้น

3. แครอท ผักกาด บิทรูทสำหรับสลัด salsify ขึ้นฉ่าย หัวไชเท้าและรากที่รับประทานได้ ทั้งสดและแช่แข็ง

อันดับที่ 3 43: Fur skins and artificial fur มีการเพิ่มขึ้นของสัดส่วนอยู่ที่ 796% โดยการเพิ่มขึ้นของสินค้ากลุ่มนี้เกิดจากเพิ่มขึ้นของหนังเฟอร์ดิบ ในส่วนหัว หาง อุ้งเท้าและชิ้นส่วนอื่น ๆ

พบว่ากลุ่มสินค้าที่มีการเปลี่ยนแปลงลดลงมากที่สุด 3 อันดับแรก ได้แก่

อันดับที่ 1 51: Wool, fine or coarse animal hair; horsehair yarn and woven fabric มีการลดลงของสัดส่วนอยู่ที่ 85% โดยสินค้าที่มีมูลค่าสูงสุดในกลุ่มนี้ได้แก่ ขนสัตว์ เส้นด้าย เป็นต้น

อันดับที่ 2 50: Silk (ผ้าไหม) มีการลดลงของสัดส่วนอยู่ที่ 84% โดยสินค้าที่มีมูลค่าสูงสุดในกลุ่มนี้ได้แก่ ผ้าทอที่ทำด้วยไหมหรือเศษไหม เศษไหม เส้นด้ายไหม เป็นต้น

อันดับที่ 3 52: Cotton(ฝ้าย) มีการลดลงของสัดส่วนอยู่ที่ 81% โดยสินค้าที่มีมูลค่าสูงสุดในกลุ่มนี้ได้แก่ ฝ้าย ฝ้ายทอที่ทำจากฝ้าย ที่มีฝ้ายมากกว่า 85% เป็นต้น

จึงสรุปได้ว่าในปัจจุบันประเทศไทยหรือคนไทยมีการเปลี่ยนแปลงพฤติกรรมในการนำเข้าสินค้าที่เปลี่ยนไป

1.พบว่าคนไทยมีการนำเข้าเนื้อเพิ่มมากขึ้นเมื่อคิดเป็นสัดส่วน บ่งบอกถึงคนไทยนิยมรับประทานเนื้อเพิ่มมากขึ้นในแง่มุมมองเศรษฐกิจถ้าหากรัฐบาลสามารถส่งเสริมหรือสนับสนุนให้เนื้อของไทยมีเกรดที่ดีเท่าต่างประเทศได้น่าจะช่วยให้สามารถลดการนำเข้าสินค้าจากต่างประเทศได้ซึ่งเป็นผลดีต่อระบบเศรษฐกิจด้วย

2.พบว่าคนไทยมีการนำเข้าผักและผักที่เป็นรากและหัวเพิ่มมากขึ้นเมื่อคิดเป็นสัดส่วน บ่งบอกถึงพฤติกรรมกรนิยมบริโภคผักในกลุ่มนี้

3.พบว่าคนไทยมีการนำเข้า Fur skins and artificial fur เพิ่มมากขึ้นเมื่อคิดเป็นสัดส่วน บ่งบอกถึงพฤติกรรมกรนิยมใช้วัสดุต่าง ๆ เป็นเครื่องหนังมากยิ่งขึ้น

4.พบว่าคนไทยมีการนำเข้าสินค้าพวก ขนสัตว์ ผ้าไหมและฝ้าย ลดลงเมื่อคิดเป็นสัดส่วน ซึ่งอาจจะเกี่ยวข้องกับข้อที่ 3 คนไทยอาจจะมีการย้ายไปใช้สินค้าในกลุ่มนั้นมากขึ้น

## 2.ประเทศไทยมีนำเข้าสินค้ากลุ่มใดเป็นมูลค่าสูงสุดและคิดเป็นสัดส่วนเท่าไร

พบว่ากลุ่มที่มีสัดส่วนและมูลค่าการนำเข้าเฉลี่ย 5 ปีย้อนหลังสูงสุด คือ

อันดับ 1 Electrical machinery and equipment หรือ เครื่องจักรไฟฟ้า อุปกรณ์ไฟฟ้าและส่วนประกอบที่เกี่ยวข้อง โดยมีสัดส่วนการนำเข้าเฉลี่ยอยู่ที่ 19% เมื่อเทียบกับทั้งหมดและคิดเป็นมูลค่า 45,572,908,400 US dollar สินค้ากลุ่มนี้มีทิศทางเป็นแนวโน้มขาขึ้นอย่างชัดเจนในช่วง 10 ปีที่ผ่านมา

อันดับ 2 Mineral fuels, mineral oils and products of their distillation หรือ เชื้อเพลิง น้ำมันและผลิตภัณฑ์จากการกลั่น โดยมีสัดส่วนการนำเข้าเฉลี่ยอยู่ที่ 15% เมื่อเทียบกับทั้งหมดและคิดเป็นมูลค่า 36,435,056,800 US dollar

อันดับ 3 Machinery, mechanical appliances, nuclear reactors, boilers หรือเครื่องจักร อุปกรณ์ เครื่องปฏิกรณ์นิวเคลียร์และบอยเลอร์ โดยมีสัดส่วนการนำเข้าเฉลี่ยอยู่ที่ 11.99% เมื่อเทียบกับทั้งหมดและคิดเป็นมูลค่า 28,541,446,600 US dollar

อันดับ 4 Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal, and articles หรือไข่มุกธรรมชาติหรือไข่มุกเลี้ยง หินอัญมณีหรือกึ่งหินอัญมณี โลหะมีค่า โลหะที่หุ้มด้วยโลหะมีค่าและของที่ทำด้วยของดังกล่าว โดยมีสัดส่วนการนำเข้าเฉลี่ยอยู่ที่ 5.35% เมื่อเทียบกับทั้งหมดและคิดเป็นมูลค่า 12,805,070,600 US dollar

อันดับ 5 Iron and steel หรือเหล็กและเหล็กกล้า โดยมีสัดส่วนการนำเข้าเฉลี่ยอยู่ที่ 4.98% เมื่อเทียบกับทั้งหมดและคิดเป็นมูลค่า 11,970,993,600 US dollar สินค้ากลุ่มนี้มีการนำเข้าเพิ่มขึ้นอย่างมากในปี 2021 เมื่อเทียบกับ 2020 เนื่องมาจากในปีที่ผ่านมาเหล็กและเหล็กกล้ามีราคาที่สูงขึ้นตามราคาตลาดโลก

สิ่งที่ควรปรับปรุงหรือเพิ่มเติมในการทำครั้งนี้ :

1. ข้อมูลที่นำมาวิเคราะห์นั้นเป็น time series data และนำมาเพียงแค่เรื่องเดียวจึงทำให้การวิเคราะห์ทำได้ยากคิดว่าถ้าจะให้ข้อมูลที่เพิ่มกว่าควรจะมีการนำข้อมูลอื่น ๆ มาด้วย เช่น ข้อมูลการนำเข้าสินค้าของประเทศไทยเมื่อเทียบกับประเทศหรือภูมิภาคนั้น ๆ เพื่อสามารถระบุข้อมูลได้เพิ่มขึ้นมาสินค้าต่าง ๆ ที่นำเข้ามาจากที่ไหนและคิดว่าข้อมูลที่น่ามาเหมาะกับการนำไปวิเคราะห์แบบ predictive มากกว่า

2. สินค้าบางกลุ่มมีความผันผวนตามราคาตลาดโลกหรือมีความผันผวนเกี่ยวกับฤดู (Seasonality) เข้ามาด้วยทำให้ไม่สามารถดูทิศทางของสินค้าบางกลุ่มได้

