

I HATE
MONDAY

HW 1

FREQUENT ITEM SET AND ASSOCIATION RULES

นำข้อมูลเข้า

```
import pandas as pd
import os

from google.colab import drive
drive.mount('/content/drive')
path = '/content/drive/MyDrive/BSC-dpdm23_file'

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

data = pd.read_csv(os.path.join(path, '/content/drive/MyDrive/bsc_dpdm25_data/reduced_marketbasket.csv'),encoding='latin-1')
```

PREPARE DATA

```
# แปลงรูปแบบเวลาใน InvoiceDate เป็น datetime
data['InvoiceDate'] = pd.to_datetime(data['InvoiceDate'])

#แปลงรูปแบบให้เป็น เดือน / วัน / ปี / ช.น. / นาท.
data['InvoiceDate'] = pd.to_datetime(data['InvoiceDate'], format='%m/%d/%Y %H:%M')

# สร้างคอลัมน์ 'Date' ดึงข้อมูลเฉพาะ
data['Date'] = data['InvoiceDate'].dt.date
data['Month'] = data['InvoiceDate'].dt.month
data['Year'] = data['InvoiceDate'].dt.year
data['Time'] = data['InvoiceDate'].dt.time

data
```

1. ช่วงเวลา CHRISTMAS & NEW YEAR (1 DEC – 31 DEC) (จัดกลุ่มตาม TRANSACTION)

เลือกเฉพาะแค่เดือน 12

```
# ดูช่วงเวลาคริสต์มาสกับปีใหม่
xmas_ny = data[data['Month']==12]
xmas_ny
```

MIN SUPPORT = 0.035

```
##กำหนดค่า min_sup = 0.035

frequentItem_XN = list(apriori(PP,min_support = 0.035 ))
for i in frequentItem_XN:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```

PAPER CHAIN KIT 50'S CHRISTMAS

YES

```
RELATIONRECORD(ITEMS=FROZENSET({"PAPER CHAIN KIT 50'S CHRISTMAS ",  
'PAPER CHAIN KIT VINTAGE CHRISTMAS'}), SUPPORT=0.047697368421052634,  
ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(),  
ITEMS_ADD=FROZENSET({"PAPER CHAIN KIT 50'S CHRISTMAS ", 'PAPER  
CHAIN KIT VINTAGE CHRISTMAS'}), CONFIDENCE=0.047697368421052634,  
LIFT=1.0)
```



PAPER CHAIN KIT VINTAGE CHRISTMAS



เลือกค่า SUPPORT ที่สูงที่สุดคือค่า SUPPORT = 0.047 และ ค่า CONFIDENCE = 0.66 หรือ 66.51%

คนที่ซื้อ PAPER CHAIN KIT VINTAGE CHRISTMAS จะมีโอกาสซื้อ PAPER CHAIN KIT 50'S CHRISTMAS อีกที่ 66.51%

1. ช่วงเวลา CHRISTMAS & NEW YEAR (1 DEC – 31 DEC) (ประจำปี + เดือน)

```
min_sup =0.3
frequentItemsets_Christmas_and_NewYear_Gb_c_m = list(apriori(transactions_Christmas_and_NewYear_Gb_c_m,min_support = min_sup))
for i in frequentItemsets_Christmas_and_NewYear_Gb_c_m:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```

```
RELATIONRECORD(ITEMS=FROZENSET({'POST', '20750'}),  
SUPPORT=0.3333333333333333, ORDERED_STATISTICS=  
[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(),  
ITEMS_ADD=FROZENSET({'POST', '20750'}),  
CONFIDENCE=0.3333333333333333,  
LIFT=1.0)ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'20750'}),  
ITEMS_ADD=FROZENSET({'POST'}), CONFIDENCE=0.9,  
LIFT=1.7357142857142858)
```

0 ດ SUPPORT = 0.33

ເນື່ອນີກາຣ້ອງ POSTAGE ມີອາກສ 33% ທີ່ RED RETROSPOT MINI CASES ຈະດູກ້ອງ
ແລະເນື່ອນີກາຣ້ອງ RED RETROSPOT MINI CASES ມີອາກສ 33% ທີ່ POSTAGE ຈະດູກ້ອງ

2. ประเทศในทวีป EUROPE (จัดกลุ่มตาม TRANSACTION)

เลือกเฉพาะประเทศในยุโรป

```
europe_countries = data[(data['Country']!='Austria') | (data['Country']!='Belgium')| (data['Country']!='Channel Islands')  
| (data['Country']!='Cyprus') | (data['Country']!='Czech Republic') | (data['Country']!='Denmark') |  
 (data['Country']!='EIRE') | (data['Country']!='European Community')| (data['Country']!='Finland')|  
 (data['Country']!='France') | (data['Country']!='Germany')| (data['Country']!='Greece')|  
 (data['Country']!='Iceland') | (data['Country']!='Italy') | (data['Country']!='Lithuania') | (data['Country']!='Malta')  
| (data['Country']!='Netherlands') | (data['Country']!='Norway') |  
 (data['Country']!='Poland') | (data['Country']!='Portugal') | (data['Country']!='Spain') | (data['Country']!='Sweden')  
| (data['Country']!='Switzerland') | (data['Country']!='United Kingdom')]  
europe_countries #แบ่งทวีปยุโรป
```

กำหนดค่า MIN SUPPORT = 0.025

```
frequentItemsets_Europe = list(apriori(trans_Europe, min_support = 0.025))  
for i in frequentItemsets_Europe:  
    if i[0].__len__()>1:  
        print('yes')  
        print(i)
```

JUMBO BAG RED



YES

```
RELATIONRECORD(ITEMS=FROZENSET({'JUMBO BAG RED RETROSPOT',  
'JUMBO BAG PINK POLKADOT'}), SUPPORT=0.03216216216216,  
ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(),  
ITEMS_ADD=FROZENSET({'JUMBO BAG RED RETROSPOT', 'JUMBO BAG  
PINK POLKADOT'}), CONFIDENCE=0.03216216216216, LIFT=1.0],
```

JUMBO BAG PINKS



จุดที่มีค่า SUPPORT ที่มากที่สุด คือ JUMBO BAG RED RETROSPOT || และ JUMBO BAG PINK POLKADOT โดยมีค่า SUPPORT=0.67 และมี CONFIDENCE = 3.22%

เมื่อมีการซื้อ JUMBO BAG PINK จะมีโอกาสที่จะซื้อ JUMBO BAG RED อยู่ที่ 67.67%
เมื่อมีการซื้อ JUMBO BAG RED จะมีโอกาสที่จะซื้อ JUMBO BAG PINKS อยู่ที่ 39.02%

2. ประเทศไทยในทวีป EUROPE (ประเทศไทย + เดือน)

```
transactions_europe = Europe_data.groupby(['Country', 'month'])['StockCode'].apply(list)
min_sup = 0.4
frequentItemsets_europe= list(apriori(transactions_europe,min_support = min_sup))
for i in frequentItemsets_europe:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```

GREEN REGENCY TEACUP AND SAUCER



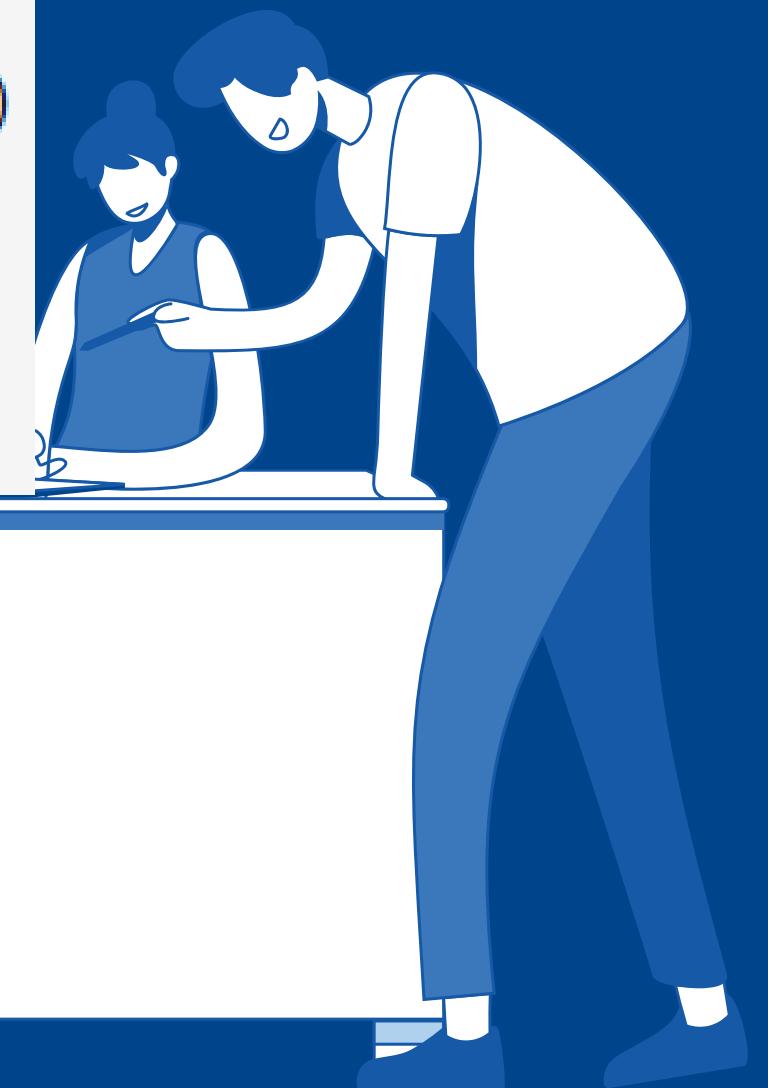
ROSES REGENCY TEACUP AND SAUCER



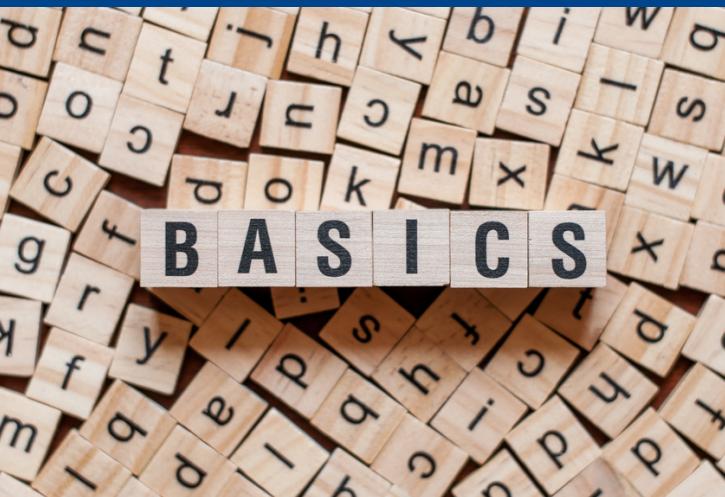
จากผลนี้ค่า SUPPORT = 0.032 และ CONFIDENCE = 0.41
เมื่อนำการซื้อ GREEN REGENCY TEACUP AND SAUCER
จะนำไปสืบ ROSES REGENCY TEACUP AND SAUCER อยู่ที่ 41%

3. ช่วงเวลา ตรุษจีนในทวีป ASIA (15 JAN – 15 FEB) (จัดกลุ่มตาม TRANSACTION)

```
min_sup = 0.3
frequentItemsets_Asia= list(apriori(transactionsAsia,min_support = min_sup))
for i in frequentItemsets_Asia:
    if i[0].__len__()>3:
        print('yes')
        print(i)
```



BATH BUILDING BLOCK WORD



SET 3 PAPER VINTAGE CHICK PAPER EGG



มีหลายค่าที่มีการซื้อของซ้ำกัน

SUPPORT: 0.3333 (33.33%)

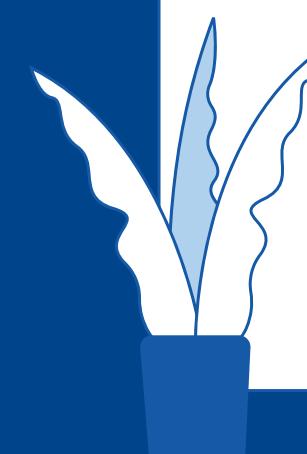
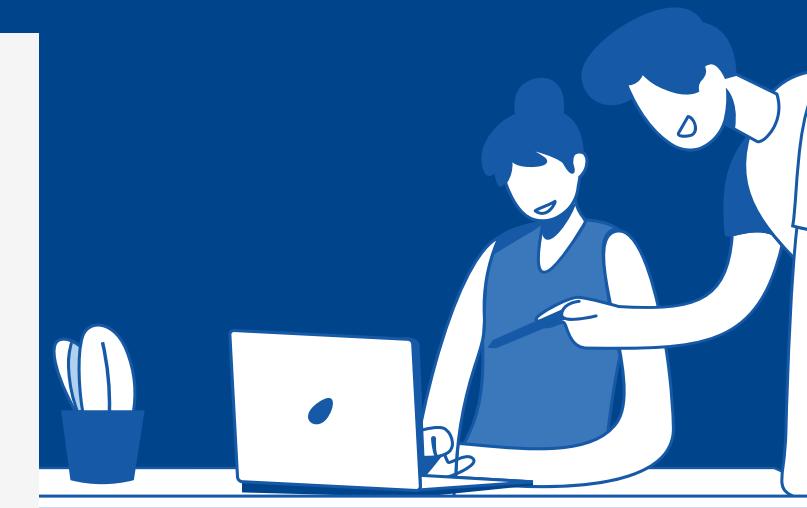
RULE 1: (NONE → BATH BUILDING BLOCK WORD, SET 3 PAPER VINTAGE CHICK PAPER EGG, JAM
MAKING SET WITH JARS : CONFIDENCE=0.3333

RULE 2: (BATH BUILDING BLOCK WORD → SET 3 PAPER VINTAGE CHICK PAPER EGG, JAM MAKING SET
WITH JARS : CONFIDENCE=1.0000

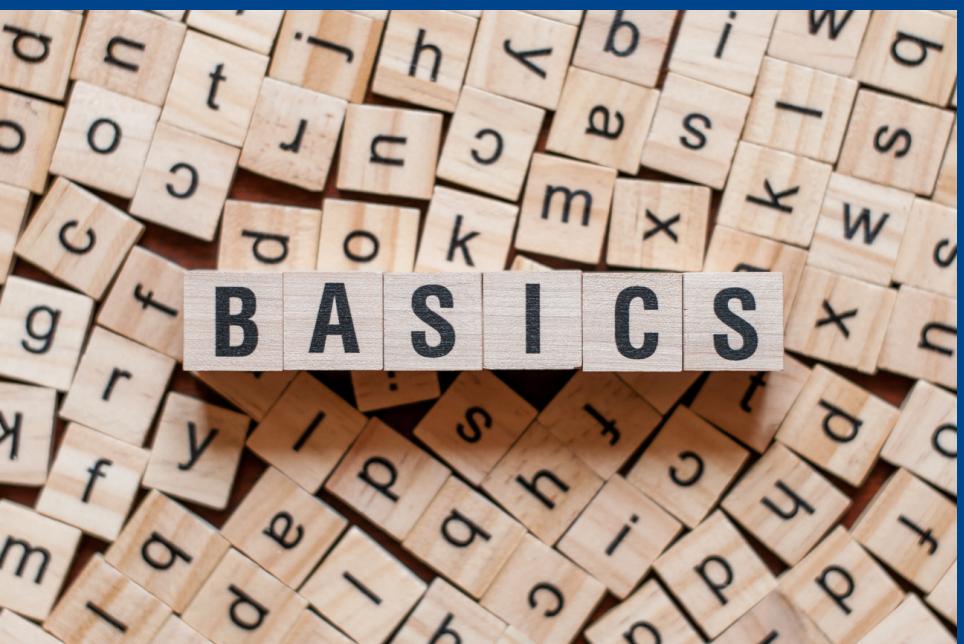
มีค่า SUPPORT = 0.3 และ CONFIDENCE = 1
เมื่อมีการซื้อ BATH BUILDING BLOCK WORD
, มีโอกาสที่จะซื้อ SET 3 PAPER VINTAGE CHICK PAPER EGG อยู่ที่ 100%

3. ช่วงเวลา ครุชิ้น ในทวีป ASIA (15 JAN – 15 FEB) (ประเทศ + เดือน)

```
transactionsAsia = ad_Asia_data.groupby(['Country', 'month'])['StockCode'].apply(list)
min_sup = 0.4
frequentItemsets_A= list(apriori(transactionsAsia,min_support = min_sup))
for i in frequentItemsets_Asia:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```



BATH BUILDING BLOCK WORD



LADIES & GENTLEMEN METAL SIGN



มีหลายค่าที่มีการซื้อของช้ากัน

นิค่า SUPPORT = 0.33 และ มี CONFIDENCE = 0.33
มีการซื้อ BATH BUILDING BLOCK WORD จะมีโอกาสที่จะซื้อ LADIES & GENTLEMEN METAL SIGN อยู่ที่ 33%

4. ของช่วงเวลา CHRISTMAS & NEW YEAR ในทวีป EUROPE (1 DEC – 31 DEC)

```
min_sup =0.08
frequentItemsetsEU= list(apriori(transactionsEU_C_N,min_support = min_sup))
for i in frequentItemsetsEU:
    if i[0].__len__()>3:
        print('yes')
        print(i)
frequentItemsetsEU
```



PAPER CHAIN KIT 50'S CHRISTMAS



```
RELATIONRECORD(ITEMS=FROZENSET({'PAPER CHAIN KIT VINTAGE  
CHRISTMAS', "PAPER CHAIN KIT 50'S CHRISTMAS "}),  
SUPPORT=0.047697368421052634, ORDERED_STATISTICS=  
[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(),  
ITEMS_ADD=FROZENSET({'PAPER CHAIN KIT VINTAGE CHRISTMAS',  
"PAPER CHAIN KIT 50'S CHRISTMAS "}),  
CONFIDENCE=0.047697368421052634, LIFT=1.0)
```

PAPER CHAIN KIT VINTAGE CHRISTMAS



จุดที่มีค่า SUPPORT ที่มากที่สุด คือ PAPER CHAIN KIT 50'S CHRISTMAS และ PAPER CHAIN KIT VINTAGE CHRISTMAS จาก DATASET โดยมีค่า SUPPORT=0.047 และมี CONFIDENCE (ค่าสนับสนุน)= 4.7%

เมื่อมีการซื้อ PAPER CHAIN KIT 50'S CHRISTMAS จะมีการซื้อ PAPER CHAIN KIT VINTAGE CHRISTMAS ด้วย อยู่ที่ 51.6%
เมื่อมีการซื้อ PAPER CHAIN KIT VINTAGE CHRISTMAS จะมีการซื้อ PAPER CHAIN KIT 50'S CHRISTMAS ด้วย อยู่ที่ 66.51%

4. ช่วงเวลา CHRISTMAS & NEW YEAR ในทวีป EUROPE (1 DEC – 31 DEC) (ประเทศ + เดือน)

```
transactions_Christmas_NewYear_Eu_CM = data_Christmas_NewYear_Eu.groupby(['Country', 'month'])['StockCode'].apply(list)
min_sup = 0.4
frequentItemsets_Christmas_NewYear_Eu_CM= list(apriori(transactions_Christmas_NewYear_Eu_CM,min_support = min_sup))
for i in frequentItemsets_Christmas_NewYear_Eu_CM:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```



RED RETROSPOT MINI CASES



POSTAGE(ค่าส่งของทางไปรษณีย์)



SUPPORT=0.4090 ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(['POST']),
ITEMS_ADD=FROZENSET(['20750']), CONFIDENCE=0.6428, LIFT=1.4142)])

0 ณ SUPPORT = 0.4 នូវ គោលការណែនាំ
' នឹងការចេញផ្លូវ POSTAGE
និងការកំណត់ថា RED RETROSPOT MINI CASES ធម្មតា 64%

5. วันวาเลนไทน์ (14 FEB) (จัดกลุ่มตาม TRANSACTION)

```
transactions_valen = data_valen14.groupby('InvoiceNo')['StockCode'].apply(list)
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
and should_run_async(code)

min_sup = 0.09
frequentItemsets_valen= list(apriori(transactions_valen,min_support = min_sup))
for i in frequentItemsets_valen:
    if i[0].__len__(>1):
        print('yes')
        print(i)
```

SET OF 3 CAKE TINS PANTRY DESIGN



HOMEMADE JAM SCENTED CANDLES



```
SUPPORT=0.0909 ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'22969'}),  
ITEMS_ADD=FROZENSET({'22720'}), CONFIDENCE=0.4 , LIFT=6.875)])
```

0.01 SUPPORT = 0.09 และ ค่า CONFIDENCE = 0.4
เมื่อมีการซื้อ HOMEMADE JAM SCENTED CANDLES
จะการซื้อ SET OF 3 CAKE TINS PANTRY DESIGN 40%

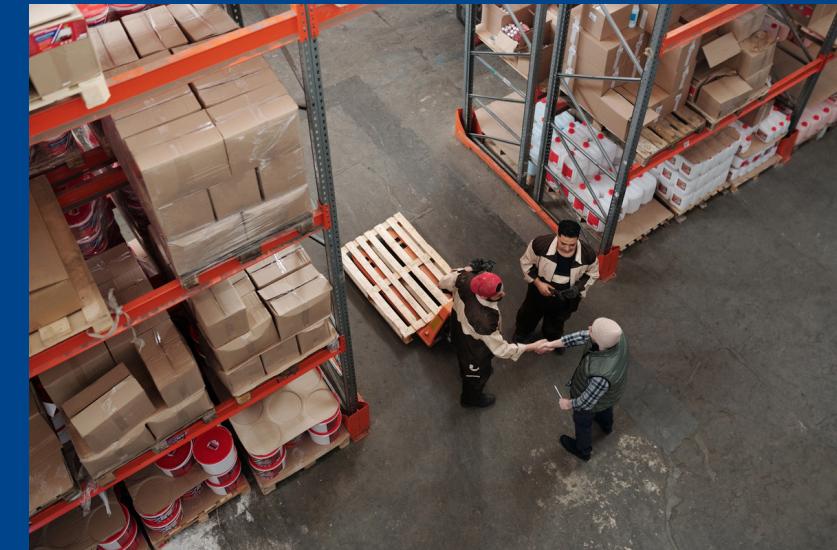
5. วันวาเลนไทน์ (14 FEB) (ประเทศไทย + เดือน)

```
transactions_valen = data_valen.groupby(['Country', 'month'])['StockCode'].apply(list)
min_sup = 0.4
frequentItemsets_valen= list(apriori(transactions_valen,min_support = min_sup))
for i in frequentItemsets_valen:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```

REGENCY CAKESTAND 3 TIER



DOLLY GIRL LUNCH BOX



SUPPORT=0.416 ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'22630'}),
ITEMS_ADD=FROZENSET({'22423'}), CONFIDENCE=0.9090, LIFT=1.8181)]

0 ດ ຮ SUPPORT = 0.41 ແລະ ຄ່າ CONFIDENCE = 0.4
' ເນື້ອນີກາຮູ້ຈຳກັດ REGENCY CAKESTAND 3 TIER
ຈະນົວກາສທີຈະຈຳກັດ DOLLY GIRL LUNCH BOX ດ້ວຍອຢູ່ທີ່ 90%

6. จัดกลุ่มตามประเทศ + เดือน โดยไม่นิ่ง่อนไข

```
transactions_C_M = data_test.groupby(['Country', 'month'])['StockCode'].apply(list)
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning
  and should_run_async(code)

min_sup = 0.3
frequentItemsets_C_M= list(apriori(transactions_C_M,min_support = min_sup))
for i in frequentItemsets_C_M:
    if i[0].__len__()>1:
        print('yes')
        print(i)
```

POSTAGE(ค่าส่งของทางไปรษณีย์)



ROUND SNACK BOXES SET OF4 WOODLAND



SUPPORT=0.3209 CONFIDENCE=0.8407
ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET(['22326']), ITEMS_ADD=FROZENSET(['POST'])

09 SUPPORT = 0.3209 และ ค่า CONFIDENCE = 0.84
เนื่องจากการซื้อ ROUND SNACK BOXES SET OF4 WOODLAND จะมีโอกาสซื้อแบบ POSTAGE อยู่ 84%

THANK YOU

