

Group 4

REDUCED MARKETBASKET DATA





MEMBER

5

นัชชา ประทุมชาติ

18

อรัญชัย แสนเทพ

28

อริสรา ยังอยู่



HOMEWORK 3

1. วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID)
แต่ละทวิป
2. ในทั้ง 7 วัน (MONDAY TUESDAY WEDNESDAY
THURSDAY FRIDAY SATURDAY SUNDAY) ของทวิป
ยุโรปสินค้าชนิดใดขายคู่กันดีที่สุดและอยู่ในวันใด



จัดการข้อมูล

page 4

```
!pip install apyori  
from apyori import apriori
```

```
import pandas as pd  
import os  
from google.colab import drive  
drive.mount('/content/drive')  
path = '/content/drive/MyDrive/bsc_dpdm_me/reduced_marketbasket.csv'  
Data = pd.read_csv('/content/drive/MyDrive/bsc_dpdm_me/reduced_marketbasket.csv', encoding='latin-1')  
Data
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010 08:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/01/2010 08:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 08:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010 08:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 08:26	3.39	17850.0	United Kingdom

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 5

ดูว่าข้อมูลที่ให้มามีประเทศอะไรบ้างเพื่อจะมาจัดกลุ่มแต่ละทวีป
มีประเทศทั้งหมด 38 ประเทศ

```
Number of unique countries: 38
Unique Countries:
United Kingdom
France
Australia
Netherlands
Germany
Norway
EIRE
Switzerland
Spain
Poland
Portugal
Italy
Belgium
Lithuania
```

```
Japan
Iceland
Channel Islands
Denmark
Cyprus
Sweden
Austria
Israel
Finland
Bahrain
Greece
Hong Kong
Singapore
Lebanon
United Arab Emirates
Saudi Arabia
Czech Republic
Canada
Unspecified
Brazil
USA
European Community
Malta
RSA
```


1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 6

แบ่งกลุ่มประเทศตามทวีป

```
country_continent = {  
    'United Kingdom': 'Europe',  
    'France': 'Europe',  
    'Australia': 'Oceania',  
    'Netherlands': 'Europe',  
    'Germany': 'Europe',  
    'Norway': 'Europe',  
    'EIRE': 'Europe', # Ireland  
    'Switzerland': 'Europe',  
    'Spain': 'Europe',  
    'Poland': 'Europe',  
    'Portugal': 'Europe',  
    'Italy': 'Europe',  
    'Belgium': 'Europe',  
    'Lithuania': 'Europe',  
    'Japan': 'Asia',
```

```
'Iceland': 'Europe',  
'Channel Islands': 'Europe',  
'Denmark': 'Europe',  
'Cyprus': 'Europe',  
'Sweden': 'Europe',  
'Austria': 'Europe',  
'Israel': 'Asia',  
'Finland': 'Europe',  
'Bahrain': 'Asia',  
'Greece': 'Europe',  
'Singapore': 'Asia',  
'Lebanon': 'Asia',  
'United Arab Emirates': 'Asia',  
'Saudi Arabia': 'Asia',  
'Canada': 'North America',  
'Unspecified': 'Unknown',  
'Brazil': 'South America',  
'USA': 'North America',  
'European Community': 'Europe',  
'Malta': 'Europe',  
'RSA': 'South Africa'  
}  
  
# Add a 'Continent' column to the DataFrame  
Data['Continent'] = Data['Country'].map(country_continent)  
  
# Handle countries not in the mapping (fill with 'Unknown')  
Data['Continent'].fillna('Unknown', inplace=True)  
Data
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Continent
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010 08:26	2.55	17850.0	United Kingdom	Europe
1	536365	71053	WHITE METAL LANTERN	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 08:26	2.75	17850.0	United Kingdom	Europe
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 7

ลบ 'Unknow'

```
# Count the number of unique continents
num_continents = Data['Continent'].nunique()
print(f"Number of unique continents: {num_continents}")

# Get a list of unique continents
unique_continents = Data['Continent'].unique()
print("Unique Continents:")
for continent in unique_continents:
    print(continent) # Indent this line to print each conti
```

```
# Remove rows where the continent is 'Unknown'
Data_new = Data[Data['Continent'] != 'Unknown']
Data_new
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Continent
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010 08:26	2.55	17850.0	United Kingdom	Europe
1	536365	71053	WHITE METAL LANTERN	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 08:26	2.75	17850.0	United Kingdom	Europe
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	Europe

หลังจากลบแล้วข้อมูลเหลือ 99.86%

```
# Calculate the percentage of remaining data
remaining_percentage = (len(Data_new) / len(Data)) * 100
print(f"Percentage of data remaining after removing 'Unknown' continents: {remaining_percentage:.2f}%")
```

Percentage of data remaining after removing 'Unknown' continents: 99.86%

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 8

ดูข้อมูลภาพรวมแต่ละทวีป

```
# Group data by continent and customer ID
continent_customer_analysis = Data_new.groupby('Continent')['CustomerID'].agg(['nunique', 'count'])
continent_customer_analysis = continent_customer_analysis.rename(columns={'nunique': 'Number of Customers', 'count': 'Number of Purchases'})

# Calculate average quantity and average sales per invoice
average_metrics = Data_new.groupby('Continent').agg(
    Average_Quantity=('Quantity', 'mean'),
    Average_Sales=('UnitPrice', 'mean')
)

# Round average quantity to the nearest integer
average_metrics['Average_Quantity'] = average_metrics['Average_Quantity'].round().astype(int)
average_metrics['Average_Sales'] = average_metrics['Average_Sales'].round(2)
result = pd.merge(continent_customer_analysis, average_metrics, on='Continent')
result
```

Continent	Number of Customers	Number of Purchases	Average_Quantity	Average_Sales
Asia	19	977	36	26.89
Europe	4331	403787	9	4.55
North America	8	442	9	3.52
Oceania	9	1259	66	3.22
South Africa	1	58	6	4.28
South America	1	32	11	4.46

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 9

สินค้าที่นิยม 10 อันดับแรกในแต่ละทวีป

```
# Group transactions by continent and item description
continent_item_counts = Data_new.groupby(['Continent', 'Description'])['CustomerID'].count().reset_index(name='CustomerCount')

# Sort by continent and transaction count in descending order to get the top 10 items for each continent
top_items_by_continent = continent_item_counts.groupby('Continent').apply(lambda x: x.nlargest(10, 'CustomerCount')).reset_index(drop=True)
top_items_by_continent
```

Continent	Description	CustomerCount
Asia	Manual	17
Asia	CHARLOTTE BAG DOLLY GIRL DESIGN	8
Asia	RED SPOTTY BISCUIT TIN	7
Asia	REGENCY CAKESTAND 3 TIER	7
Asia	PARTY BUNTING	6
Asia	LADIES & GENTLEMEN METAL SIGN	5
Asia	LUNCH BAG DOLLY GIRL DESIGN	5
Asia	PACK OF 12 TRADITIONAL CRAYONS	5
Asia	PARTY PIZZA DISH RED RETROSPOT	5
Asia	RABBIT NIGHT LIGHT	5

Europe	WHITE HANGING HEART T-LIGHT HOLDER	2065
Europe	REGENCY CAKESTAND 3 TIER	1885
Europe	JUMBO BAG RED RETROSPOT	1652
Europe	ASSORTED COLOUR BIRD ORNAMENT	1414
Europe	PARTY BUNTING	1401
Europe	LUNCH BAG RED RETROSPOT	1345
Europe	SET OF 3 CAKE TINS PANTRY DESIGN	1217
Europe	POSTAGE	1190
Europe	LUNCH BAG BLACK SKULL	1121
Europe	PACK OF 72 RETROSPOT CAKE CASES	1073

North America	CARD DOLLY GIRL	4
North America	SET OF 12 FAIRY CAKE BAKING CASES	4
North America	TEA PARTY BIRTHDAY CARD	4
North America	60 CAKE CASES DOLLY GIRL DESIGN	3
North America	BAKING SET 9 PIECE RETROSPOT	3
North America	COLOURING PENCILS BROWN TUBE	3
North America	EMBROIDERED RIBBON REEL SUSIE	3
North America	PACK OF 60 SPACEBOY CAKE CASES	3
North America	PINK HAPPY BIRTHDAY BUNTING	3
North America	PLASTERS IN TIN SPACEBOY	3

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 10

สินค้าที่นิยม 10 อันดับแรกในแต่ละทวีป

Oceania	SET OF 3 CAKE TINS PANTRY DESIGN	10
Oceania	LUNCH BAG RED RETROSPOT	9
Oceania	RED TOADSTOOL LED NIGHT LIGHT	9
Oceania	BAKING SET 9 PIECE RETROSPOT	8
Oceania	BAKING SET SPACEBOY DESIGN	8
Oceania	HANGING HEART JAR T-LIGHT HOLDER	8
Oceania	LUNCH BAG SPACEBOY DESIGN	8
Oceania	PAPER BUNTING RETROSPOT	8
Oceania	PARTY BUNTING	8
Oceania	ROSES REGENCY TEACUP AND SAUCER	8

South America	CLOCK MAGNET MUM'S KITCHEN	1
South America	COOK WITH WINE METAL SIGN	1
South America	DOLLY GIRL LUNCH BOX	1
South America	DOORMAT AIRMAIL	1
South America	EDWARDIAN PARASOL BLACK	1
South America	EDWARDIAN PARASOL NATURAL	1
South America	EDWARDIAN PARASOL PINK	1
South America	EDWARDIAN PARASOL RED	1
South America	EMERGENCY FIRST AID TIN	1
South America	FIRST AID TIN	1

South Africa	4 TRADITIONAL SPINNING TOPS	1
South Africa	ALARM CLOCK BAKELIKE RED	1
South Africa	ASSORTED BOTTLE TOP MAGNETS	1
South Africa	BAKING SET 9 PIECE RETROSPOT	1
South Africa	BELLE JARDINIERE CUSHION COVER	1
South Africa	BIRDHOUSE GARDEN MARKER	1
South Africa	CARD HOLDER LOVE BIRD SMALL	1
South Africa	CHARLOTTE BAG VINTAGE ALPHABET	1
South Africa	CHILDRENS TOY COOKING UTENSIL SET	1
South Africa	CLASSIC GLASS COOKIE JAR	1

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 11

ทวีป Europe

```
transactionsEurope = Europe_data.groupby('CustomerID')['StockCode'].apply(list)
min_sup = 0.075
frequentItemsetsEurope = list(apriori(transactionsEurope,min_support = min_sup))
for i in frequentItemsetsEurope:
    if i[0].__len__(>1:
        print('yes')
        print(i)
```

```
set(Data_new[(Data_new['StockCode']=='22910')]['Description'])
```

22910	PAPER CHAIN KIT VINTAGE CHRISTMAS
22086	PAPER CHAIN KIT 50'S CHRISTMAS

```
RelationRecord(items=frozenset({'22086', '22910'}), support=0.0750404063726622, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22086', '22910'}), confidence=0.0750404063726622, lift=1.0), OrderedStatistic(items_base=frozenset({'22086'}), items_add=frozenset({'22910'}), confidence=0.530179445350734, lift=4.916931858274152), OrderedStatistic(items_base=frozenset({'22910'}), items_add=frozenset({'22086'}), confidence=0.6959314775160599, lift=4.916931858274152)])
```


1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

ทวีป Asia

```
RelationRecord(items=frozenset({'22630', '23240', '22629', '22423'}), support=0.15789473684210525, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22630', '23240', '22629', '22423'}), confidence=0.15789473684210525, lift=1.0), OrderedStatistic(items_base=frozenset({'22423'}), items_add=frozenset({'23240', '22630', '22629'}), confidence=0.5, lift=3.166666666666667), OrderedStatistic(items_base=frozenset({'22629'}), items_add=frozenset({'23240', '22630', '22423'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'22630'}), items_add=frozenset({'23240', '22629', '22423'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240'}), items_add=frozenset({'22630', '22629', '22423'}), confidence=0.6, lift=3.8000000000000003), OrderedStatistic(items_base=frozenset({'22629', '22423'}), items_add=frozenset({'23240', '22630'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'22630', '22423'}), items_add=frozenset({'23240', '22629'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240', '22423'}), items_add=frozenset({'22630', '22629'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'22629', '22630'}), items_add=frozenset({'23240', '22423'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240', '22629'}), items_add=frozenset({'22630', '22423'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240', '22630'}), items_add=frozenset({'22629', '22423'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'22630', '22629', '22423'}), items_add=frozenset({'23240'}), confidence=1.0, lift=3.8000000000000003), OrderedStatistic(items_base=frozenset({'23240', '22629', '22423'}), items_add=frozenset({'22630'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240', '22630', '22423'}), items_add=frozenset({'22629'}), confidence=1.0, lift=6.333333333333334), OrderedStatistic(items_base=frozenset({'23240', '22629', '22630'}), items_add=frozenset({'22423'}), confidence=1.0, lift=3.166666666666667))] [ ]
set(Data_new[(Data_new['StockCode']=='22630')]['Description'])
```

22630	DOLLY GIRL LUNCH BOX	22423	REGENCY CAKESTAND 3 TIER, damages, faulty
22629	SPACEBOY LUNCH BOX	23240	SET OF 4 KNICK KNACK TINS DOILEY

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 13

ทวีป Oceania

```
RelationRecord(items=frozenset({'22138', '22617'}), support=0.4444444444444444, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22138', '22617'}), confidence=0.4444444444444444, lift=1.0), OrderedStatistic(items_base=frozenset({'22138'}), items_add=frozenset({'22617'}), confidence=1.0, lift=2.25), OrderedStatistic(items_base=frozenset({'22617'}), items_add=frozenset({'22138'}), confidence=1.0, lift=2.25)])
```

22617	BAKING SET SPACEBOY DESIGN, mouldy, thrown away.
22138	BAKING SET 9 PIECE RETROSPOT

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 14

ทวีป North America

```
RelationRecord(items=frozenset({'22138', '22619', '22551',  
'22649', '21232', '23159', '22951', '16161P', '22417'}),  
support=0.25, ordered_statistics=  
[OrderedStatistic(items_base=frozenset(),  
items_add=frozenset({'22138', '22619', '22551', '22649',  
'21232', '23159', '22951', '16161P', '22417'}),  
confidence=0.25, lift=1.0),  
OrderedStatistic(items_base=frozenset({'16161P'}),  
items_add=frozenset({'22138', '22619', '22551', '22649',  
'21232', '23159', '22951', '22417'}), confidence=1.0,  
lift=4.0),.....,
```

22417	PACK OF 60 SPACEBOY CAKE CASES
22649	STRAWBERRY FAIRY CAKE TEAPOT
22551	PLASTERS IN TIN SPACEBOY
16161P	WRAP ENGLISH ROSE
22138	BAKING SET 9 PIECE RETROSPOT
23159	SET OF 5 PANCAKE DAY MAGNETS
22619	SET OF 6 SOLDIER SKITTLES
22951	60 CAKE CASES DOLLY GIRL DESIGN
21232	STRAWBERRY CERAMIC TRINKET BOX, STRAWBERRY CERAMIC TRINKET POT

1.วิเคราะห์พฤติกรรมผู้บริโภค (GROUPBY CUSTOMER ID) แต่ละทวีป

page 15

South America

StockCode	PurchaseCount
15056BL	1
15056N	1
23179	1
23178	1
23054	1
23053	1
23052	1
23051	1
23050	1
23049	1
22993	1
22722	1
22699	1
22698	1

South Africa

StockCode	PurchaseCount
20676	1
23316	1
22915	1
22960	1
23081	1
23169	1
23201	1
23203	1
23205	1
23209	1
23240	1
23245	1
23284	1
23298	1

เนื่องจากทวีป South America และ South Africa
มีลูกค้าแต่ละคนเดียว แม้จะมาซื้อสินค้าหลายครั้ง แต่ก็
ไม่มีสินค้าใดซ้ำกันเลย จึงไม่สามารถสรุปพฤติกรรมผู้
บริโภคของสองทวีปนี้ได้

2. ในทั้ง 7 ;วัน (MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY SUNDAY) ของทวีปยุโรปของชนิดใดขายคู่กันดีที่สุดและอยู่ในวันใด

page 16

```
# Convert 'InvoiceDate' to datetime objects
europe_data['InvoiceDate'] = pd.to_datetime(europe_data['InvoiceDate'])

# Extract the day of the week (0=Monday, 6=Sunday)
europe_data['DayOfWeek'] = europe_data['InvoiceDate'].dt.dayofweek

# Create a dictionary to map dayofweek to day names
day_names = {
    0: 'Monday', 1: 'Tuesday', 2: 'Wednesday', 3: 'Thursday',
    4: 'Friday', 5: 'Saturday', 6: 'Sunday'
}

# Map dayofweek to day names in the dataframe
europe_data['DayName'] = europe_data['DayOfWeek'].map(day_names)
```

StockCode	
DayName	
Friday	[[84991, 21212, 22938, 21124, 84457, 22857, 21...
Monday	[[23480, 21265, 21636, 22372, 22375, 22371, 22...
Sunday	[[23077, 23078, POST], [22064, 21232, 22066, 2...
Thursday	[[22376, 22374, 22371, 22375, 20665, 23076, 21...
Tuesday	[[23166, 23166], [85116, 22375, 71477, 22492, ...
Wednesday	[[84625A, 84625C, 85116, 20719, 22375, 22376, ...

```
# Group by DayName and aggregate the items
day_items = customer_day_items.groupby('DayName')['StockCode'].apply(list)
```


ของชนิดใดขายคู่กันดีที่สุดในวันใด

page 17

```
# Apply apriori
if transactions :
    frequent_itemsets = list(apriori(transactions, min_support=0.07)) # Adjust min_support as needed

print(f"\nAssociation Rules for {day}:")
for itemset in frequent_itemsets:
    if len(itemset[0]) > 1 : # Print rules with more than one item
        print(itemset) # Indented this line to print the itemset when the condition is True
```

```
RelationRecord(items=frozenset({'82482', '82494L'}), support=0.07079646017699115, ordered_statistics=[
OrderedStatistic(items_base=frozenset(), items_add=frozenset({'82482', '82494L'}), confidence=0.07079646017699115, lift=1.0),
OrderedStatistic(items_base=frozenset({'82482'}), items_add=frozenset({'82494L'}), confidence=0.7272727272727273,
lift=7.9298245614035086), OrderedStatistic(items_base=frozenset({'82494L'}), items_add=frozenset({'82482'}),
confidence=0.7719298245614035, lift=7.9298245614035086)])
```


ของชนิดใดขายคู่กันดีที่สุดและอยู่ในวันใด

82494L	WOODEN FRAME ANTIQUE WHITE , cracked, crushed ctn , nan
82482	WOODEN PICTURE FRAME WHITE FINISH



THANK YOU
THANK YOU!

