

HOMEWORK 1

Frequent Item Set & Association Rules

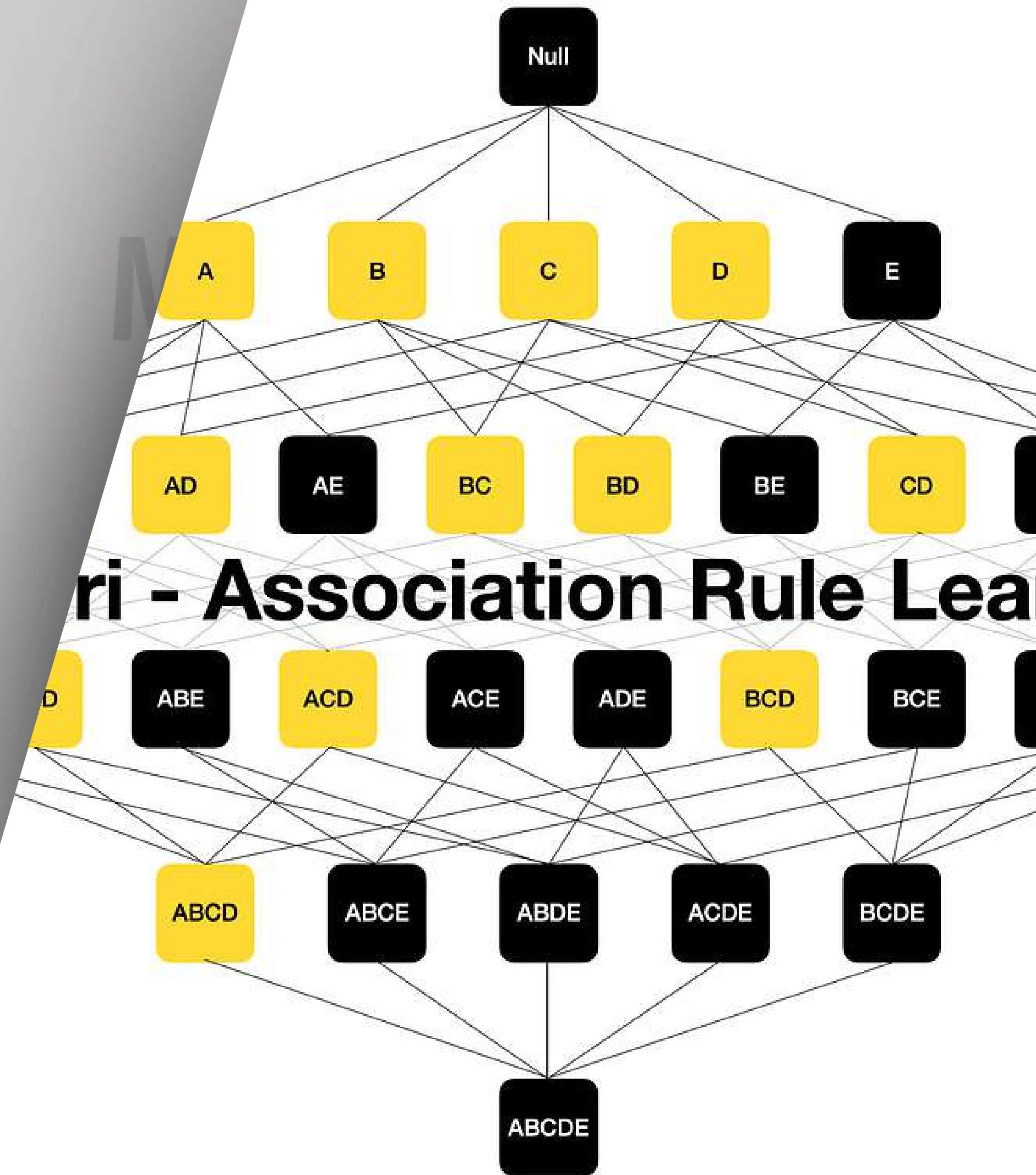
Data Set: `reduced_marketbasket`

▼ HW 1 >>> หา frequent Item Set และ Association Rules

จัดกลุ่มตาม Transaction (Invoice_No)

- ของช่วงเวลา Chritmas & New Year ทั่วโลก (1 Dec - 31 Dec)
- ของประเทศในทวีป Europe
- ของประเทศในทวีป Asia
- ของช่วงเวลา ตรุษจีน ในทวีป Asia (15 Jan - 15 Feb)
- ของช่วงเวลา Chritmas & New Year ในทวีป Europe (1 Dec - 31 Dec)
- ของวันวาเลนไทน์ (14 Feb)

โดยจัดกลุ่มตามประเทศ + เดือน (ประเทศ-เดือน ถือเป็น 1 transaction)



MEMBER

นายกัตต์ธร ก้อมณี 643020513-9

นายรัชชานนท์ กิพย์พิมานพร 643020515-5

นายวุฒิชัย คำนา 643020521-0

Prepare Data



แยก เดือน/วัน/ปี
ออกจาก เวลา

	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
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/09/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/09/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/09/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/09/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/09/2011 12:50	4.95	12680.0	France

541909 rows × 8 columns



แยก เดือน/วัน/ปี ออกจากกัน



แปลงให้ค่าเดือนอยู่ในหน่วยเดียวกัน



แบ่งทวีป

Prepare Data



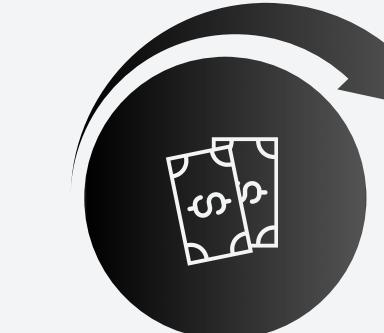
แยก เดือน/วัน/ปี
ออกจาก เวลา

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	date	day	month	year
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010 08:26	2.55	17850.0	United Kingdom	12/01/2010	1	12	2010
1	536365	71053	WHITE METAL LANTERN	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	12/01/2010	1	12	2010
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 08:26	2.75	17850.0	United Kingdom	12/01/2010	1	12	2010
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	12/01/2010	1	12	2010
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 08:26	3.39	17850.0	United Kingdom	12/01/2010	1	12	2010
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/09/2011 12:50	0.85	12680.0	France	12/09/2011	9	12	2011
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/09/2011 12:50	2.10	12680.0	France	12/09/2011	9	12	2011
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/09/2011 12:50	4.15	12680.0	France	12/09/2011	9	12	2011
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/09/2011 12:50	4.15	12680.0	France	12/09/2011	9	12	2011
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/09/2011 12:50	4.95	12680.0	France	12/09/2011	9	12	2011

541909 rows x 12 columns



แยก เดือน/วัน/ปี ออกจากกัน



แปลงให้ค่าเดือนอยู่ในหน่วยเดียวกัน



แบ่งทวีป

Function for Finding the Item

- Code สำหรับหาชื่อสินค้า จาก Column "Description"

```
def finditem(x,y) :  
    L=set(data[(data['StockCode']==x)|(data['StockCode']==y)]['Description'])  
    return L
```

```
def finditems(x, y='', z='', w=''):   
    I =set(data[(data['StockCode']==x)|(data['StockCode']==y)|(data['StockCode']==z)|(data['StockCode']==w)]['Description'])  
    return I
```





จัดกลุ่มตาม Transaction (Invoice_No)

Christmas & New Year (1 Dec - 31 Dec)

yes

RelationRecord(items=frozenset({'22910', '22086'}), support=0.047697368421052634, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22910', '22086'}), confidence=0.047697368421052634, lift=1.0), OrderedStatistic(items_base=frozenset({'22086'}), items_add=frozenset({'22910'}), confidence=0.5160142348754448, lift=7.195794834960331), OrderedStatistic(items_base=frozenset({'22910'}), items_add=frozenset({'22086'}), confidence=0.6651376146788991, lift=7.195794834960332)])

```
finditem('22910', '22086')
```

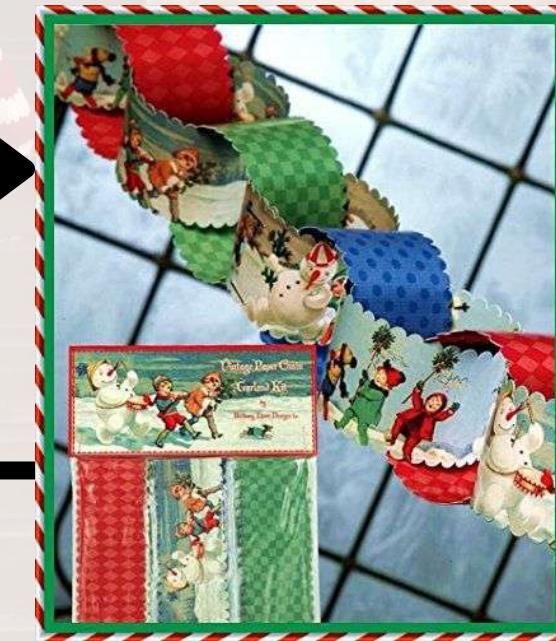
```
["PAPER CHAIN KIT 50'S CHRISTMAS ", "PAPER CHAIN KIT VINTAGE CHRISTMAS"]
```



Christmas & New Year (1 Dec - 31 Dec)



confidence=66.51%



confidence=51.60%

Europe

yes

RelationRecord(items=frozenset({'22386', '85099B'}), support=0.03233453810578679, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22386', '85099B'}), confidence=0.03233453810578679, lift=1.0), OrderedStatistic(items_base=frozenset({'22386'}), items_add=frozenset({'85099B'}), confidence=0.6775244299674267, lift=8.203944050584402), OrderedStatistic(items_base=frozenset({'85099B'}), items_add=frozenset({'22386'}), confidence=0.39152941176470585, lift=8.203944050584402)])

```
finditem('22386', '85099B')
```

```
{'JUMBO BAG PINK POLKADOT', 'JUMBO BAG RED RETROSPOT'}
```



Europe



confidence=67.75%



confidence=39.15%

Chinese New Year in Asia (15 Jan - 15 Feb)



YES

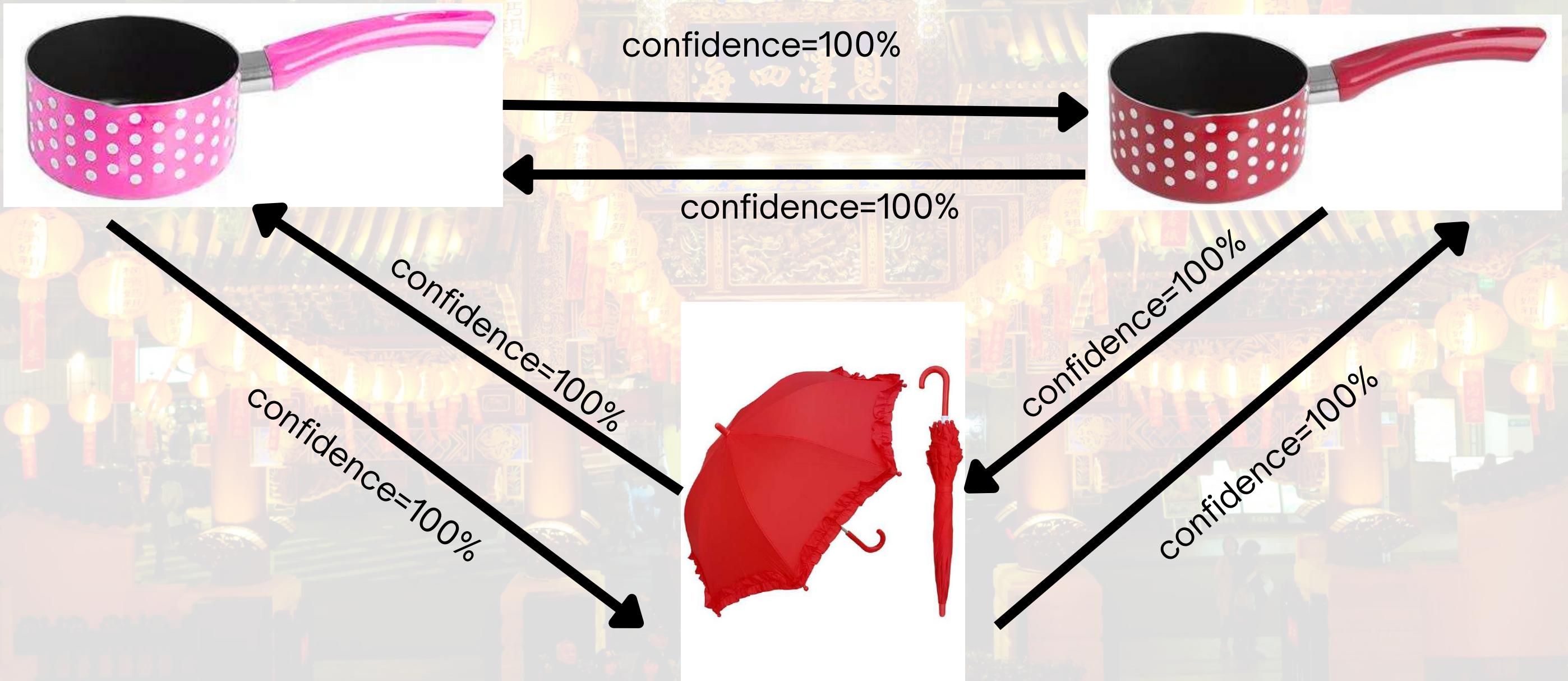
RelationRecord(items=frozenset({'22202', '20682', '22551', '22203'}), support=0.3333333333333333, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22202', '20682', '22551', '22203'}), confidence=0.3333333333333333, lift=1.0), OrderedStatistic(items_base=frozenset({'20682'}), items_add=frozenset({'22202', '22551', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202'}), items_add=frozenset({'20682', '22551', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22203'}), items_add=frozenset({'22202', '20682', '22551'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22551'}), items_add=frozenset({'22202', '20682', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '20682'}), items_add=frozenset({'22551', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'20682', '22203'}), items_add=frozenset({'22202', '22551'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'20682', '22551'}), items_add=frozenset({'22202', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '22203'}), items_add=frozenset({'20682', '22551'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '22551'}), items_add=frozenset({'20682', '22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22551', '22203'}), items_add=frozenset({'22202', '20682'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '20682', '22203'}), items_add=frozenset({'22551'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '20682', '22551'}), items_add=frozenset({'22203'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'20682', '22551', '22203'}), items_add=frozenset({'22202'}), confidence=1.0, lift=3.0), OrderedStatistic(items_base=frozenset({'22202', '22551', '22203'}), items_add=frozenset({'20682'}), confidence=1.0, lift=3.0)])

```
finditems('20682', '22551', '22203', '22202')
```

```
{'MILK PAN PINK POLKADOT',
 'MILK PAN RED RETROSPOT',
 'PLASTERS IN TIN SPACEBOY',
 'RED RETROSPOT CHILDRENS UMBRELLA',
 'nan'}
```



Chinese New Year in Asia (15 Jan - 15 Feb)



Christmas & New Year in Europe (1 Dec - 31 Dec)

YES

RelationRecord(items=frozenset({'22910', '22086'}), support=0.04790221341261976, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22910', '22086'}), confidence=0.04790221341261976, lift=1.0), OrderedStatistic(items_base=frozenset({'22086'}), items_add=frozenset({'22910'}), confidence=0.5160142348754448, lift=7.165023343889778), OrderedStatistic(items_base=frozenset({'22910'}), items_add=frozenset({'22086'}), confidence=0.6651376146788991, lift=7.165023343889778)])

```
finditem('22086', '22910')
{"PAPER CHAIN KIT 50'S CHRISTMAS ", "PAPER CHAIN KIT VINTAGE CHRISTMAS"}
```



Christmas & New Year in Europe (1 Dec - 31 Dec)



confidence=66.51%



confidence=51.60%

Valentine's Day (14 Feb)

YES

RelationRecord(items=frozenset({'22969', '22720'}), support=0.09090909090909091, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'22720', '22969'}), confidence=0.09090909090909091, lift=1.0), OrderedStatistic(items_base=frozenset({'22720'}), items_add=frozenset({'22969'}), confidence=0.625, lift=6.875), OrderedStatistic(items_base=frozenset({'22969'}), items_add=frozenset({'22720'}), confidence=1.0, lift=6.875)])

```
finditems('22720', '22969')
```

```
{'HOMEMADE JAM SCENTED CANDLES', 'SET OF 3 CAKE TINS PANTRY DESIGN'}
```



Valentine's Day (14 Feb)



confidence=100%



confidence=62.5%

จัดกลุ่มตาม Transaction (ประทุม + เดือน)



จัดกลุ่มตาม Transaction (ประเทค+เดือน)

Christmas & New Year (1 Dec - 31 Dec)

yes

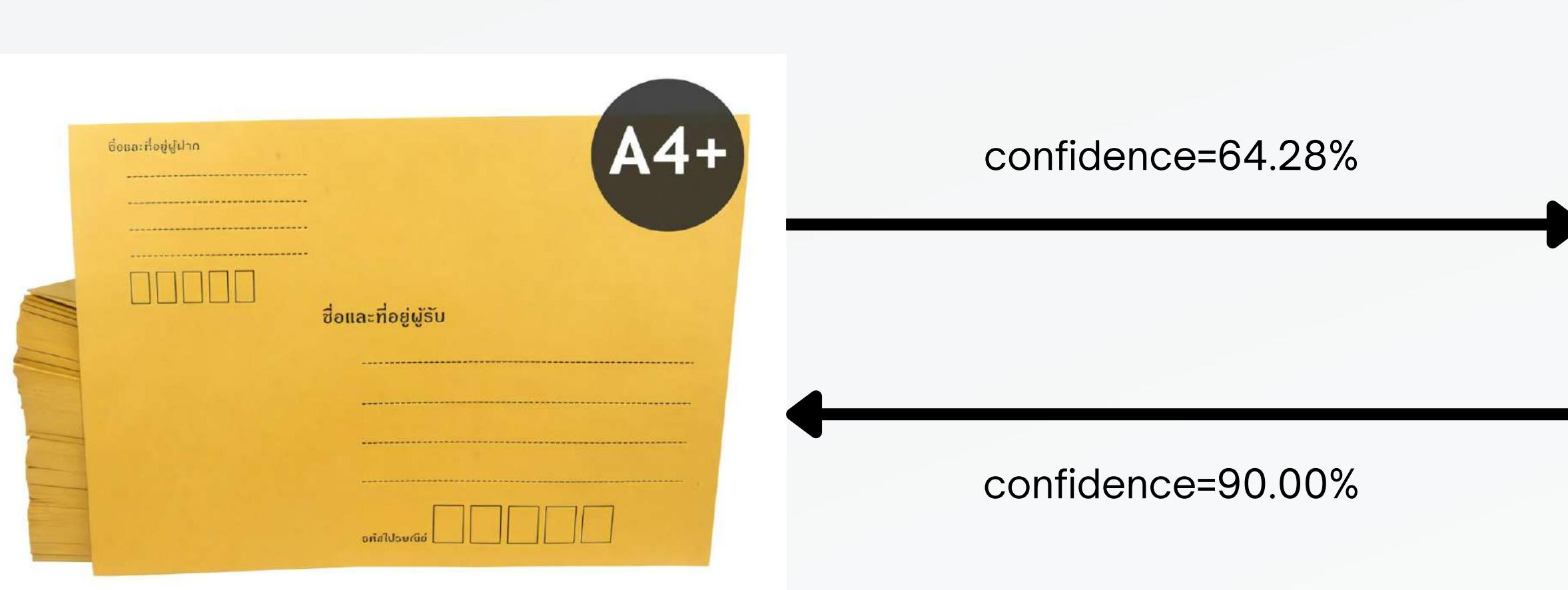
```
RelationRecord(items=frozenset({'POST', '20750'}), support=0.3333333333333333,  
ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'POST',  
'20750'}), confidence=0.3333333333333333, lift=1.0),  
OrderedStatistic(items_base=frozenset({'20750'}), items_add=frozenset({'POST'}),  
confidence=0.9, lift=1.7357142857142858),  
OrderedStatistic(items_base=frozenset({'POST'}), items_add=frozenset({'20750'}),  
confidence=0.6428571428571429, lift=1.735714285714286)])
```

```
[56] finditem('POST', '20750')  
  
{'POSTAGE', 'RED RETROSPOT MINI CASES', nan}
```



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Christmas & New Year (1 Dec - 31 Dec)



จัดกลุ่มตาม Transaction (ประเทค+เดือน)

Europe

yes

```
RelationRecord(items=frozense({'22326', 'POST'}), support=0.41150442477876104,  
ordered_statistics=[OrderedStatistic(items_base=frozense(), items_add=frozense({'22326', 'POST'}),  
confidence=0.41150442477876104, lift=1.0),  
OrderedStatistic(items_base=frozense({'22326'}), items_add=frozense({'POST'}), confidence=0.9207920792079208,  
lift=1.2688964018353055),  
OrderedStatistic(items_base=frozense({'POST'}), items_add=frozense({'22326'}), confidence=0.5670731707317073,  
lift=1.2688964018353055)])
```

```
finditem('POST', '22326')
```

```
{'POSTAGE', 'ROUND SNACK BOXES SET OF4 WOODLAND ', nan}
```



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Europe



confidence=92.08%



confidence=56.67%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)

yes

RelationRecord(items=frozenset({'20682', '22202'}), support=0.4, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'20682', '22202'}), confidence=0.4, lift=1.0), OrderedStatistic(items_base=frozenset({'20682'}), items_add=frozenset({'22202'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'22202'}), items_add=frozenset({'20682'}), confidence=1.0, lift=2.5)])

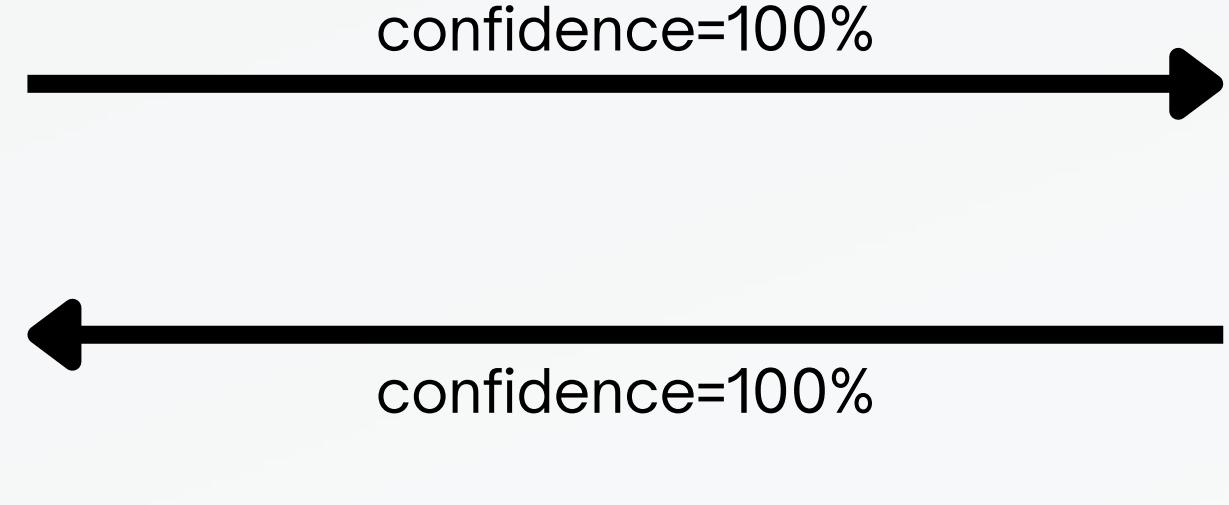
02

```
finditem('20682', '22202')
{'MILK PAN PINK POLKADOT', 'RED RETROSPOT CHILDRENS UMBRELLA'}
```



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



จัดกลุ่มตาม Transaction (ประเทค+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)

yes

```
RelationRecord(items=frozenset({'82552', '82551', '85150'}), support=0.4,  
ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'82552', '82551', '85150'}), confidence=0.4, lift=1.0),  
OrderedStatistic(items_base=frozenset({'82551'}), items_add=frozenset({'82552', '85150'}), confidence=1.0, lift=2.5),  
OrderedStatistic(items_base=frozenset({'82552'}), items_add=frozenset({'82551', '85150'}), confidence=1.0, lift=2.5),  
OrderedStatistic(items_base=frozenset({'85150'}), items_add=frozenset({'82552', '82551'}), confidence=0.6666666666666667,  
lift=1.6666666666666667), OrderedStatistic(items_base=frozenset({'82552', '82551'}), items_add=frozenset({'85150'}), confidence=1.0,  
lift=1.6666666666666667), OrderedStatistic(items_base=frozenset({'82551', '85150'}), items_add=frozenset({'82552'}), confidence=1.0, lift=2.5),  
OrderedStatistic(items_base=frozenset({'82552', '85150'}), items_add=frozenset({'82551'}), confidence=1.0, lift=2.5)])
```

03

```
finditems(['82552', '82551', '85150'])
```

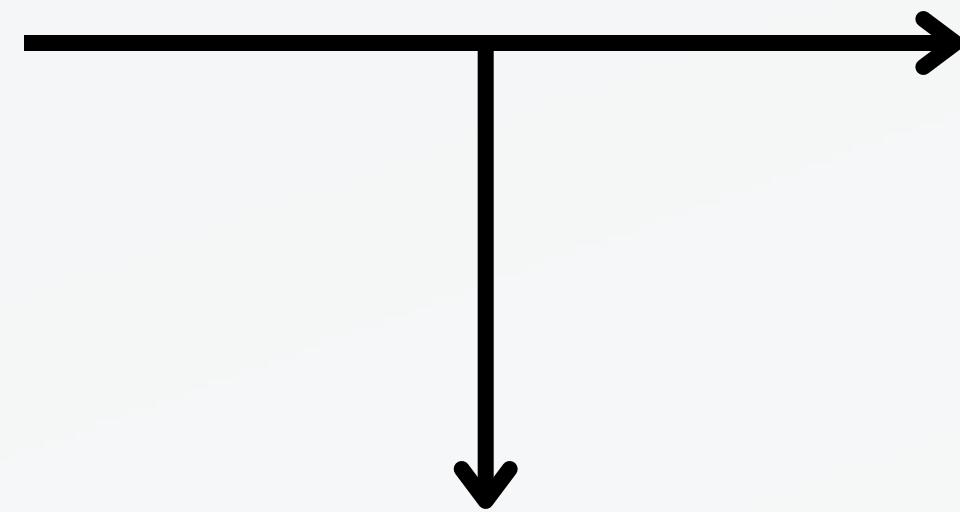
```
{'LADIES & GENTLEMEN METAL SIGN',  
'LAUNDRY 15C METAL SIGN',  
'WASHROOM METAL SIGN',
```

จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%

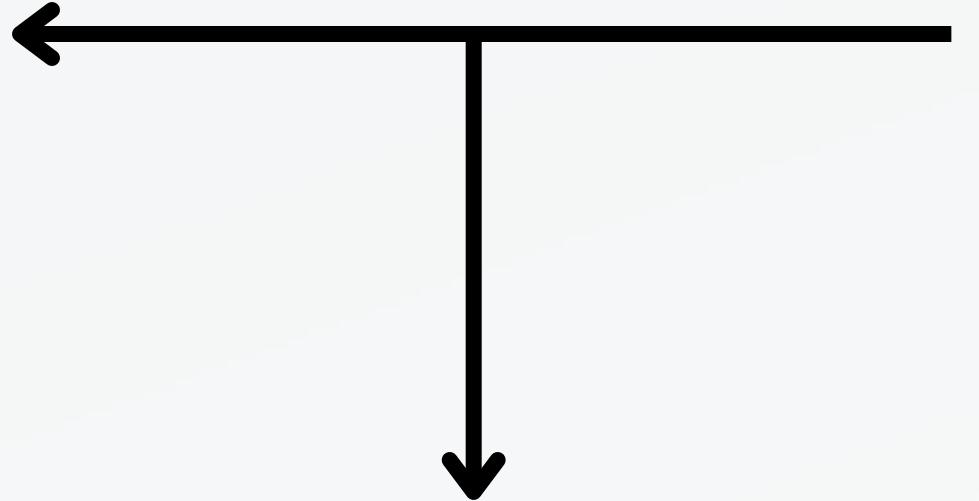


จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=66.67%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)



confidence=100%



จัดกลุ่มตาม Transaction (ประเทศ+เดือน)

Chinese New Year in Asia (15 Jan - 15 Feb)

04

```
finditem('21439', '22620', '21216', '20725')
```

yes

```
RelationRecord(items=frozenset({'21439', '22620', '21216', '20725'}), support=0.4, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'21439', '22620', '21216', '20725'}), confidence=0.4, lift=1.0), OrderedStatistic(items_base=frozenset({'20725'}), items_add=frozenset({'21439', '22620', '21216'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21216'}), items_add=frozenset({'21439', '22620', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439'}), items_add=frozenset({'22620', '21216', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'22620'}), items_add=frozenset({'21439', '21216', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21216', '20725'}), items_add=frozenset({'21439', '22620'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '20725'}), items_add=frozenset({'22620', '21216'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'22620', '20725'}), items_add=frozenset({'21439', '21216'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '21216'}), items_add=frozenset({'22620', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'22620', '21216'}), items_add=frozenset({'21439', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '22620'}), items_add=frozenset({'21216', '20725'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '21216', '20725'}), items_add=frozenset({'22620'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'22620', '21216', '20725'}), items_add=frozenset({'21439'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '22620', '20725'}), items_add=frozenset({'21216'}), confidence=1.0, lift=2.5), OrderedStatistic(items_base=frozenset({'21439', '22620', '21216'}), items_add=frozenset({'20725'}), confidence=1.0, lift=2.5)])
```

```
{ '4 TRADITIONAL SPINNING TOPS',  
'BASKET OF TOADSTOOLS',  
'LUNCH BAG RED RETROSPOT',  
'LUNCH BAG RED SPOTTY',  
'SET 3 RETROSPOT TEA, COFFEE, SUGAR' }
```



จัดกลุ่มตาม Transaction (ประเทศ+เดือน)

Christmas & New Year in Europe (1 Dec - 31 Dec)

yes

RelationRecord(items=frozenset({'POST', '20750'}), support=0.4090909090909091, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'POST', '20750'}), confidence=0.4090909090909091, lift=1.0), OrderedStatistic(items_base=frozenset({'20750'}), items_add=frozenset({'POST'}), confidence=0.9000000000000001, lift=1.4142857142857146), OrderedStatistic(items_base=frozenset({'POST'}), items_add=frozenset({'20750'}), confidence=0.6428571428571429, lift=1.4142857142857144)])

```
finditem('POST', '20750')
{'POSTAGE', 'RED RETROSPOT MINI CASES', nan}
```

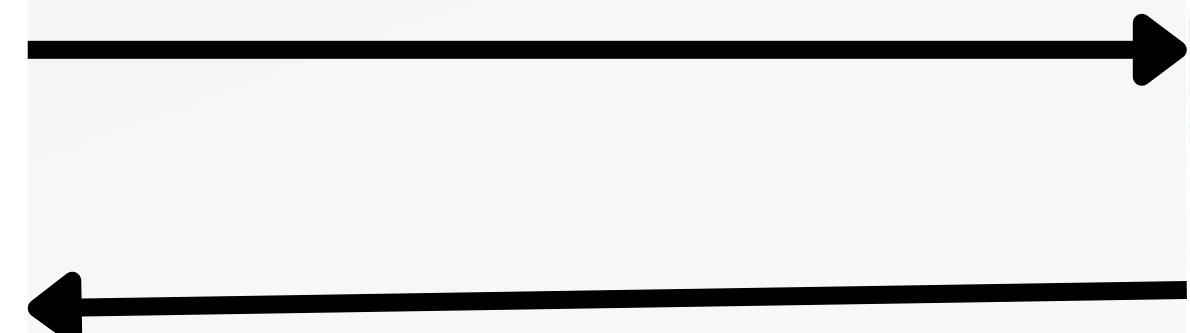


จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Christmas & New Year in Europe (1 Dec - 31 Dec)



confidence=90.00%



confidence=64.29%



จัดกลุ่มตาม Transaction (ประเภท+เดือน)

Valentine's Day (14 Feb)

