109 DataScience HW3

Association Analysis

1. 程式執行

程式網址：https://colab.research.google.com/drive/1nPDeWi-guQCwr9a_aaOaaiSAo_qINzHZ?usp=sharing

從上到下全部執行（也可以只執行到 Download result 的部分）

2. 演算法簡介

▾ Download dataset into workspace

```
[1]  !gdown  --id  "1PsnRV_D0ZJ6p6do1K4wX87hYk1qZ2UkX"  --output  "data.csv"
     !gdown  --id  "1pu2RRUgOaEPB92Pq5G-zVcBwg1QOEKSm"  --output  "prediction.csv"

     Downloading...
     From: https://drive.google.com/uc?id=1PsnRV_D0ZJ6p6do1K4wX87hYk1qZ2UkX
     To: /content/data.csv
     45.6MB [00:00, 88.4MB/s]
     Downloading...
     From: https://drive.google.com/uc?id=1pu2RRUgOaEPB92Pq5G-zVcBwg1QOEKSm
     To: /content/prediction.csv
     100% 268k/268k [00:00<00:00, 35.4MB/s]
```

```
[2]  !ls

     data.csv  prediction.csv  sample_data
```

一開始先把資料集載入到 colab 的工作區中，並確認檔案確實存在。

Data cleaning

```
[3]  import pandas as pd
     import numpy as np
```

```
[4]  data = pd.read_csv("data.csv")
     data.head()
```

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

```
[5]  data_uk = data[data["Country"]=="United Kingdom"]
     data_uk.head()
```

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

在資料前處理的部分，首先先照著說明的指示把屬於英國的資料篩選並抓取出來。

```
[6]  na = pd.isna(data_uk["InvoiceNo"])
     na.to_list().count(True)

     0
```

接著根據指示，把 invoice number 為空值的資料找出並處理，不過這邊發現並沒有空值的資料，所以不處理，直接執行其他清理步驟。

```
[7]  cancelIndex = data_uk[data_uk["InvoiceNo"].str.startswith("C")].index
     data_uk_cancel = data_uk.drop(cancelIndex)
     data_uk_cancel.head()
```

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

接下來把 invoice number 開頭為 C 的資料刪除。

```
[8]  postageIndex = data_uk_cancel[data_uk_cancel["Description"] == "POSTAGE"].index
     data_uk_cancel_postage = data_uk_cancel.drop(postageIndex)
     data_uk_cancel_postage.head()
```

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

最後把郵資的商品刪除，完成資料的清理部分。

```
[9]  data.shape, data_uk.shape, data_uk_cancel.shape, data_uk_cancel_postage.shape

     ((541909, 8), (495478, 8), (487622, 8), (487570, 8))
```

從上圖可以觀察在每一步驟的清理中都有把一些不合適的資料點刪除。

除了做清理之外，因為我使用的是 mlxtend 函式庫提供的 apriori 及 association_rules 函式進行關聯性分析，所以需要把資料的格式做一些轉換才能進行分析。

Convert format

```
[10]  from mlxtend.preprocessing import TransactionEncoder

[11]  txDataDict = {}
      nan = 0

      for label, row in data_uk_cancel_postage.iterrows():
          if row["Description"] == np.nan:
              nan += 1
          elif row["InvoiceNo"] in txDataDict:
              txDataDict[row["InvoiceNo"]].append(str(row["Description"]))
          else:
              txDataDict[row["InvoiceNo"]] = [str(row["Description"])]

      txDataDict["538504"], nan
```

```
(['GREEN WIRE STANDING CANDLE HOLDER',
  'NATURAL SLATE HEART CHALKBOARD ',
  'ANTIQUE SILVER TEA GLASS ETCHED',
  'AGED GLASS SILVER T-LIGHT HOLDER',
  'FRENCH BLUE METAL DOOR SIGN 6',
  'FRENCH BLUE METAL DOOR SIGN 5',
  'FRENCH BLUE METAL DOOR SIGN 2',
  'ROTATING LEAVES T-LIGHT HOLDER',
  'ANTIQUE SILVER TEA GLASS ETCHED',
  'CHRISTMAS GINGHAM HEART',
  'WOODEN SCHOOL COLOURING SET',
  'ANTIQUE GLASS HEART DECORATION ',
  'ROTATING LEAVES T-LIGHT HOLDER',
  'HOT WATER BOTTLE BABUSHKA ',
  'CARDHOLDER HOLLY WREATH METAL',
  'WICKER WREATH LARGE',
  'HEART OF WICKER SMALL',
  'WICKER STAR '],
 0)
```

首先我先根據不同的 invoice number 將每一次的交易統整，同一個 invoice number 代表同一次的結帳。

```
[12]  txData  =  list(txDataDict.values())
      txData[:5]
```

```
[['WHITE HANGING HEART T-LIGHT HOLDER',
  'WHITE METAL LANTERN',
  'CREAM CUPID HEARTS COAT HANGER',
  'KNITTED UNION FLAG HOT WATER BOTTLE',
  'RED WOOLLY HOTTIE WHITE HEART.',
  'SET 7 BABUSHKA NESTING BOXES',
  'GLASS STAR FROSTED T-LIGHT HOLDER'],
 ['HAND WARMER UNION JACK', 'HAND WARMER RED POLKA DOT'],
 ['ASSORTED COLOUR BIRD ORNAMENT',
  "POPPY'S PLAYHOUSE BEDROOM ",
  "POPPY'S PLAYHOUSE KITCHEN",
  'FELTCRAFT PRINCESS CHARLOTTE DOLL',
  'IVORY KNITTED MUG COSY ',
  'BOX OF 6 ASSORTED COLOUR TEASPOONS',
  'BOX OF VINTAGE JIGSAW BLOCKS ',
  'BOX OF VINTAGE ALPHABET BLOCKS',
  'HOME BUILDING BLOCK WORD',
  'LOVE BUILDING BLOCK WORD',
  'RECIPE BOX WITH METAL HEART',
  'DOORMAT NEW ENGLAND'],
 ['JAM MAKING SET WITH JARS',
  'RED COAT RACK PARIS FASHION',
  'YELLOW COAT RACK PARIS FASHION',
  'BLUE COAT RACK PARIS FASHION'],
 ['BATH BUILDING BLOCK WORD']]
```

接下來把購買物品轉換成二維的 list。

```
[13]  encoder  =  TransactionEncoder()
      encode  =  encoder.fit(txData).transform(txData)
      txDataPD  =  pd.DataFrame(encode,  columns=encoder.columns_)
      txDataPD.head()
```

	4 PURPLE FLOCK DINNER CANDLES	50'S CHRISTMAS GIFT BAG LARGE	DOLLY GIRL BEAKER	I LOVE LONDON MINI BACKPACK	NINE DRAWER OFFICE TIDY	OVAL WALL MIRROR DIAMANTE	RED SPOT GIFT BAG LARGE	SET 2 TEA TOWELS I LOVE LONDON	SPA GIF
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	

5 rows × 4188 columns

最後利用 mlxtend 函式庫提供的 TransactionEncoder 方法，把資料轉換成類似
onehot encoding 的格式。

做完資料格式轉換之後，就可以開始使用模型了。

▾ Association rule

```
[14]  from mlxtend.frequent_patterns import apriori, association_rules
```

```
[15]  frequent_itemsets = apriori(txDataPD, min_support=0.01, use_colnames=True)
```

```
[16]  result = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.5)
```

我先使用 apriori 函式找出所有 support > 0.01 的交易，接著使用 association_rules 函式篩選出 confidence > 0.5 的規則，最終結果就是想要的關聯規則。

產生規則之後，接下來是透過這些規則去判斷 predict.csv 中的規則是否正確。

▾ Apply rule

```
[17]  prediction = pd.read_csv("prediction.csv")
      prediction.head()
```

	index	Association Rule antecedents	Association Rule consequents
0	0	BLUE SPOT CERAMIC DRAWER KNOB	WHITE SPOT BLUE CERAMIC DRAWER KNOB
1	1	Lighthouse Trading zero invc incorr	SILVER GLASS T-LIGHT SET, RED RETROSPOT UMBRELLA
2	2	LUNCH BAG SUKI DESIGN, LUNCH BAG APPLE DESIGN	LUNCH BAG SPACEBOY DESIGN
3	3	ELEPHANT BIRTHDAY CARD	SILVER ROCCOCO CHANDELIER
4	4	JUMBO BAG PINK VINTAGE PAISLEY	JUMBO BAG RED RETROSPOT

```
[18]  labels = []

      for label, row in prediction.iterrows():
          antecedent = set(row["Association Rule antecedents"].split(", "))
          consequent = set(row["Association Rule consequents"].split(", "))

          possibleCons = result[result["antecedents"] == antecedent]["consequents"]

          lab = 0
          for cons in possibleCons:
              if (cons == consequent):
                  lab = 1

          labels.append(lab)
```

首先讀入檔案後，在每一行都找尋檔案寫的規則是否有在關聯規則中，有的話標註 1，沒有標註 0。

Download result

```
[20] with open("result.csv", 'w') as fh:
         fh.write("index,label"+"\n")
         for i in range(len(labels)):
             fh.write(str(i)+", "+str(int(labels[i]))+"\n")
```

```
[21] from google.colab import files
     files.download('result.csv')
```

最後的部分是儲存結果到 result.csv 之中，並進行下載。

3. 關聯性規則觀察

antecedents	consequents	antecec	consequ	support	confidence
frozenset({'BEADED CRYSTAL HEART PINK ON STICK'})	frozenset({'DOTCOM POSTAGE'})	0.0103	0.0353	0.010009	0.975728
frozenset({'SUKI SHOULDER BAG', 'JAM MAKING SET PRINTED'})	frozenset({'DOTCOM POSTAGE'})	0.0108	0.0353	0.010407	0.963134
frozenset({'REGENCY TEA PLATE ROSES ', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0118	0.0167	0.011104	0.944915
frozenset({'WOODEN HEART CHRISTMAS SCANDINAVIAN', 'WOODEN TREE CHRISTMAS SCANDINAVIAN'})	frozenset({'WOODEN STAR CHRISTMAS SCANDINAVIAN'})	0.0116	0.0239	0.010905	0.943966
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER ROSEMARY'})	0.0115	0.0115	0.010656	0.930435
frozenset({'WOODLAND CHARLOTTE BAG', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0124	0.0451	0.011553	0.928
frozenset({'HERB MARKER ROSEMARY'})	frozenset({'HERB MARKER THYME'})	0.0115	0.0115	0.010656	0.926407
frozenset({'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN '})	frozenset({'REGENCY TEA PLATE ROSES '})	0.012	0.019	0.011104	0.925311
frozenset({'CHARLOTTE BAG SUKI DESIGN', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0129	0.0451	0.011851	0.915385
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER ROSEMARY'})	0.0114	0.0115	0.010308	0.907895
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER THYME'})	0.0114	0.0115	0.010308	0.907895
frozenset({'REGENCY CAKESTAND 3 TIER', 'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0145	0.0465	0.013146	0.907216
frozenset({'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0132	0.0167	0.012001	0.906015
frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0272	0.0465	0.024549	0.90293
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER PARSLEY'})	0.0115	0.0114	0.010308	0.9
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER MINT'})	0.0114	0.0116	0.010208	0.899123

antecedents	consequents	antecec	consequ	support	confidence
frozenset({'BEADED CRYSTAL HEART PINK ON STICK'})	frozenset({'DOTCOM POSTAGE'})	0.0103	0.0353	0.010009	0.975728
frozenset({'SUKI SHOULDER BAG', 'JAM MAKING SET PRINTED'})	frozenset({'DOTCOM POSTAGE'})	0.0108	0.0353	0.010407	0.963134
frozenset({'REGENCY TEA PLATE ROSES ', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0118	0.0167	0.011104	0.944915
frozenset({'WOODEN HEART CHRISTMAS SCANDINAVIAN', 'WOODEN TREE CHRISTMAS SCANDINAVIAN'})	frozenset({'WOODEN STAR CHRISTMAS SCANDINAVIAN'})	0.0116	0.0239	0.010905	0.943966
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER ROSEMARY'})	0.0115	0.0115	0.010656	0.930435
frozenset({'WOODLAND CHARLOTTE BAG', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0124	0.0451	0.011553	0.928
frozenset({'HERB MARKER ROSEMARY'})	frozenset({'HERB MARKER THYME'})	0.0115	0.0115	0.010656	0.926407
frozenset({'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN '})	frozenset({'REGENCY TEA PLATE ROSES '})	0.012	0.019	0.011104	0.925311
frozenset({'CHARLOTTE BAG SUKI DESIGN', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0129	0.0451	0.011851	0.915385
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER ROSEMARY'})	0.0114	0.0115	0.010308	0.907895
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER THYME'})	0.0114	0.0115	0.010308	0.907895
frozenset({'REGENCY CAKESTAND 3 TIER', 'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0145	0.0465	0.013146	0.907216
frozenset({'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0132	0.0167	0.012001	0.906015
frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0272	0.0465	0.024549	0.90293
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER PARSLEY'})	0.0115	0.0114	0.010308	0.9
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER MINT'})	0.0114	0.0116	0.010208	0.899123

上面是關聯性規則中 confidence 大約超過 0.9 的商品組合，其中可以發現有很多商品組合會造成有關連的原因是那一類商品都算是同一類商品，只是顏色或是款式不同而已，舉例來說第 3 行是買「玫瑰款式茶杯盤」和「粉色茶杯盤」就會買「綠色茶杯盤」、第 4 行則是跟聖誕節裝飾有關、第 5 行跟植物的標示牌相關等，紅色框是我把 antecedents 和 consequence 是同款商品的部分（相同文字）標示出來的部分。

雖然大多數有相關聯的原因是因為是同一類型商品的關係，不過仍有一些商品組合滿有意思的，值得探討。

antecedents	consequents	anteced	consequ	support	confidence
frozenset({'BEADED CRYSTAL HEART PINK ON STICK'})	frozenset({'DOTCOM POSTAGE'})	0.0103	0.0353	0.010009	0.975728
frozenset({'SUKI SHOULDER BAG', 'JAM MAKING SET PRINTED'})	frozenset({'DOTCOM POSTAGE'})	0.0108	0.0353	0.010407	0.963134
frozenset({'REGENCY TEA PLATE ROSES ', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0118	0.0167	0.011104	0.944915
frozenset({'WOODEN HEART CHRISTMAS SCANDINAVIAN', 'WOODEN TREE CHRISTMAS SCANDINAVIAN'})	frozenset({'WOODEN STAR CHRISTMAS SCANDINAVIAN'})	0.0116	0.0239	0.010905	0.943966
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER ROSEMARY'})	0.0115	0.0115	0.010656	0.930435
frozenset({'WOODLAND CHARLOTTE BAG', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0124	0.0451	0.011553	0.928
frozenset({'HERB MARKER ROSEMARY'})	frozenset({'HERB MARKER THYME'})	0.0115	0.0115	0.010656	0.926407
frozenset({'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN '})	frozenset({'REGENCY TEA PLATE ROSES '})	0.012	0.019	0.011104	0.925311
frozenset({'CHARLOTTE BAG SUKI DESIGN', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0129	0.0451	0.011851	0.915385
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER ROSEMARY'})	0.0114	0.0115	0.010308	0.907895
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER THYME'})	0.0114	0.0115	0.010308	0.907895
frozenset({'REGENCY CAKESTAND 3 TIER', 'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0145	0.0465	0.013146	0.907216
frozenset({'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0132	0.0167	0.012001	0.906015
frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0272	0.0465	0.024549	0.90293
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER PARSLEY'})	0.0115	0.0114	0.010308	0.9
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER MINT'})	0.0114	0.0116	0.010208	0.899123

上方所列的這個組合是有買「玫瑰款式茶杯組」和「粉色茶杯組」和「3 層蛋糕架」同時也會買「綠色茶杯組」，推測原因是買這樣組合的顧客可能會和朋友喝下午茶，所以會需要買較多的茶杯，但是可能覺得粉色和玫瑰款色調太相近了所以選擇不一樣的顏色。

antecedents	consequents	anteced	consequ	support	confidence
frozenset({'BEADED CRYSTAL HEART PINK ON STICK'})	frozenset({'DOTCOM POSTAGE'})	0.0103	0.0353	0.010009	0.975728
frozenset({'SUKI SHOULDER BAG', 'JAM MAKING SET PRINTED'})	frozenset({'DOTCOM POSTAGE'})	0.0108	0.0353	0.010407	0.963134
frozenset({'REGENCY TEA PLATE ROSES ', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0118	0.0167	0.011104	0.944915
frozenset({'WOODEN HEART CHRISTMAS SCANDINAVIAN', 'WOODEN TREE CHRISTMAS SCANDINAVIAN'})	frozenset({'WOODEN STAR CHRISTMAS SCANDINAVIAN'})	0.0116	0.0239	0.010905	0.943966
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER ROSEMARY'})	0.0115	0.0115	0.010656	0.930435
frozenset({'WOODLAND CHARLOTTE BAG', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0124	0.0451	0.011553	0.928
frozenset({'HERB MARKER ROSEMARY'})	frozenset({'HERB MARKER THYME'})	0.0115	0.0115	0.010656	0.926407
frozenset({'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN '})	frozenset({'REGENCY TEA PLATE ROSES '})	0.012	0.019	0.011104	0.925311
frozenset({'CHARLOTTE BAG SUKI DESIGN', 'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG PINK POLKADOT'})	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	0.0129	0.0451	0.011851	0.915385
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER ROSEMARY'})	0.0114	0.0115	0.010308	0.907895
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER THYME'})	0.0114	0.0115	0.010308	0.907895
frozenset({'REGENCY CAKESTAND 3 TIER', 'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0145	0.0465	0.013146	0.907216
frozenset({'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN '})	0.0132	0.0167	0.012001	0.906015
frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0272	0.0465	0.024549	0.90293
frozenset({'HERB MARKER THYME'})	frozenset({'HERB MARKER PARSLEY'})	0.0115	0.0114	0.010308	0.9
frozenset({'HERB MARKER PARSLEY'})	frozenset({'HERB MARKER MINT'})	0.0114	0.0116	0.010208	0.899123

不過這 2 個組合我就不太確定是為什麼了，可能需要有相關領域知識才能知道這 2 個組合的關聯。

除了以上商品組合之間的關聯，我還發現大部分有較高關聯的組合都是跟茶杯、花、包包相關，結合最一開始的觀察推測可能是因為這類型商品最能吸引消費者購

買近似款式的商品，如果有需要大量增加商品種類的需求可能可以優先考慮這一類
型的產品。