

因為考慮到產品間的階層關係，加上如果選擇比較細部的資料難以找到規則，因此使用 product id 對應到的 **product department** 來做分析

## 第一題

使用 mlxtend library

### 1. Sorted by confidence

	antecedents			
8078	(Snacks, Beverages, Household, Baking Goods, Breakfast Foods, Health and Hygiene, Snack Foods)			
2183	(Deli, Eggs, Baked Goods, Frozen Foods, Starchy Foods)			
2212	(Snacks, Eggs, Alcoholic Beverages, Frozen Foods, Baked Goods)			
2211	(Snacks, Household, Eggs, Alcoholic Beverages, Frozen Foods, Baked Goods)			
2210	(Produce, Snacks, Eggs, Alcoholic Beverages, Frozen Foods, Baked Goods)			
2208	(Eggs, Baked Goods, Frozen Foods, Starchy Foods, Health and Hygiene, Snack Foods)			
2207	(Eggs, Baked Goods, Alcoholic Beverages, Starchy Foods, Health and Hygiene, Snack Foods)			
2204	(Household, Eggs, Baked Goods, Frozen Foods, Starchy Foods, Health and Hygiene)			
2203	(Household, Eggs, Baked Goods, Alcoholic Beverages, Starchy Foods, Health and Hygiene)			
2186	(Deli, Household, Eggs, Baked Goods, Alcoholic Beverages, Starchy Foods)			
	consequents	support	confidence	lift
8078	(Produce, Frozen Foods)	0.000106	1.0	5.254164
2183	(Alcoholic Beverages, Snack Foods)	0.000106	1.0	19.744914
2212	(Household, Produce)	0.000106	1.0	5.209331
2211	(Produce)	0.000106	1.0	2.060928
2210	(Household)	0.000106	1.0	2.616187
2208	(Alcoholic Beverages)	0.000106	1.0	8.526920
2207	(Frozen Foods)	0.000106	1.0	2.653976
2204	(Alcoholic Beverages)	0.000106	1.0	8.526920
2203	(Frozen Foods)	0.000106	1.0	2.653976
2186	(Snack Foods)	0.000106	1.0	2.377874

### Visualization on rules(antecedents->consequents , sorted alphabetically)

```
rule 1 ['Baking Goods', 'Beverages', 'Breakfast Foods', 'Health and Hygiene', 'Household', 'Snack Foods', 'Snacks'] -> ['Frozen Foods', 'Produce']
rule 2 ['Baked Goods', 'Deli', 'Eggs', 'Frozen Foods', 'Starchy Foods'] -> ['Alcoholic Beverages', 'Snack Foods']
rule 3 ['Alcoholic Beverages', 'Baked Goods', 'Eggs', 'Frozen Foods', 'Snacks'] -> ['Household', 'Produce']
rule 4 ['Alcoholic Beverages', 'Baked Goods', 'Eggs', 'Frozen Foods', 'Household', 'Snacks'] -> ['Produce']
rule 5 ['Alcoholic Beverages', 'Baked Goods', 'Eggs', 'Frozen Foods', 'Produce', 'Snacks'] -> ['Household']
rule 6 ['Baked Goods', 'Eggs', 'Frozen Foods', 'Health and Hygiene', 'Snack Foods', 'Starchy Foods'] -> ['Alcoholic Beverages']
rule 7 ['Alcoholic Beverages', 'Baked Goods', 'Eggs', 'Health and Hygiene', 'Snack Foods', 'Starchy Foods'] -> ['Frozen Foods']
rule 8 ['Baked Goods', 'Eggs', 'Frozen Foods', 'Health and Hygiene', 'Household', 'Starchy Foods'] -> ['Alcoholic Beverages']
rule 9 ['Alcoholic Beverages', 'Baked Goods', 'Eggs', 'Health and Hygiene', 'Household', 'Starchy Foods'] -> ['Frozen Foods']
rule 10 ['Alcoholic Beverages', 'Baked Goods', 'Deli', 'Eggs', 'Household', 'Starchy Foods'] -> ['Snack Foods']
```

### 2. Sorted by lift

	antecedents	
8871	(Deli, Checkout, Frozen Foods, Snacks, Health and Hygiene)	
8862	(Deli, Checkout, Frozen Foods, Snacks, Produce, Snack Foods)	
8857	(Deli, Checkout, Frozen Foods, Household, Snacks, Snack Foods)	
8859	(Deli, Checkout, Frozen Foods, Snacks, Baking Goods, Produce)	
8290	(Deli, Checkout, Frozen Foods, Snacks, Health and Hygiene)	
8860	(Deli, Checkout, Frozen Foods, Snacks, Health and Hygiene, Produce)	
8869	(Deli, Checkout, Household, Snacks, Health and Hygiene, Produce)	
8868	(Deli, Checkout, Household, Baking Goods, Snacks, Snack Foods)	
8861	(Deli, Checkout, Frozen Foods, Snacks, Health and Hygiene, Snack Foods)	
8783	(Checkout, Frozen Foods, Canned Foods, Snacks, Health and Hygiene, Snack Foods)	

	consequents	support	confidence	lift
8871	(Beverages, Household, Baking Goods, Produce, Snack Foods)	0.000106	1.0	133.278169
8862	(Beverages, Household, Health and Hygiene, Baking Goods)	0.000106	1.0	128.744898
8857	(Beverages, Health and Hygiene, Produce, Baking Goods)	0.000106	1.0	95.103015
8859	(Beverages, Household, Health and Hygiene, Snack Foods)	0.000106	1.0	88.644028
8290	(Beverages, Household, Baking Goods, Snack Foods)	0.000106	1.0	79.518908
8860	(Beverages, Household, Baking Goods, Snack Foods)	0.000106	1.0	79.518908
8869	(Beverages, Baking Goods, Frozen Foods, Snack Foods)	0.000106	1.0	78.529046
8868	(Beverages, Health and Hygiene, Produce, Frozen Foods)	0.000106	1.0	75.550898
8861	(Beverages, Household, Produce, Baking Goods)	0.000106	1.0	65.486159
8783	(Beverages, Household, Produce, Baking Goods)	0.000106	1.0	65.486159

## Rules:

```
rule 1 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce', 'Snack Foods']
rule 2 ['Checkout', 'Deli', 'Frozen Foods', 'Produce', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Health and Hygiene', 'Household']
rule 3 ['Checkout', 'Deli', 'Frozen Foods', 'Household', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Health and Hygiene', 'Produce']
rule 4 ['Baking Goods', 'Checkout', 'Deli', 'Frozen Foods', 'Produce', 'Snacks'] -> ['Beverages', 'Health and Hygiene', 'Household', 'Snack Foods']
rule 5 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Snack Foods']
rule 6 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Produce', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Snack Foods']
rule 7 ['Checkout', 'Deli', 'Health and Hygiene', 'Household', 'Produce', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Frozen Foods', 'Snack Foods']
rule 8 ['Baking Goods', 'Checkout', 'Deli', 'Household', 'Snack Foods', 'Snacks'] -> ['Beverages', 'Frozen Foods', 'Health and Hygiene', 'Produce']
rule 9 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce']
rule 10 ['Canned Foods', 'Checkout', 'Frozen Foods', 'Health and Hygiene', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce']
```

### 3. 比較

甲、Sort by confidence 的結果中，沒有品項是每一條 rule 都出現的，只有 alcoholic beverages, baked goods, eggs 是最常一起出現的三個品項

乙、Sort by lift 的結果中 checkout, Deli, beverages, baking goods 在前 10 條 rule 中都會共同出現

丙、在前 10 筆 rule 中，觀察到這 10 條 rule 的 support 和 confidence 都相同，只有 lift 有高度差異，因此前 10 條 rule 之間的比較採用 sort by lift 比較能找出 10 條 rule 中較好的關聯規則，且使用 sort by lift 比較能找到 10 條 rule 中的共同特色。

## 第二題

### Product using FP Growth

#### 1. Sorted by confidence

	antecedents
6694	(Canned Foods, Snack Foods, Deli, Household, Health and Hygiene, Produce, Carousel)
7745	(Deli, Household, Health and Hygiene, Snacks, Produce, Checkout)
7730	(Deli, Household, Baking Goods, Health and Hygiene, Frozen Foods, Snacks, Checkout)
7741	(Deli, Baking Goods, Health and Hygiene, Frozen Foods, Snacks, Checkout)
7749	(Snack Foods, Deli, Household, Health and Hygiene, Baking Goods, Beverages, Snacks, Checkout)
7748	(Snack Foods, Deli, Household, Health and Hygiene, Baking Goods, Beverages, Frozen Foods, Checkout)
7747	(Deli, Health and Hygiene, Frozen Foods, Snacks, Checkout)
7746	(Deli, Health and Hygiene, Frozen Foods, Snacks, Produce, Checkout)
7744	(Deli, Household, Health and Hygiene, Frozen Foods, Snacks, Checkout)
7736	(Snack Foods, Deli, Household, Health and Hygiene, Snacks, Checkout)

	consequents	support	confidence	lift
6694	(Dairy)	0.000106	1.0	6694 3.802974
7745	(Frozen Foods, Baking Goods, Snack Foods)	0.000106	1.0	7745 17.386771
7730	(Produce, Snack Foods)	0.000106	1.0	7730 4.771936
7741	(Produce, Household, Snack Foods)	0.000106	1.0	7741 11.529394
7749	(Frozen Foods)	0.000106	1.0	7749 2.653976
7748	(Snacks)	0.000106	1.0	7748 8.486771
7747	(Household, Produce, Baking Goods, Snack Foods)	0.000106	1.0	7747 33.231782
7746	(Household, Baking Goods, Snack Foods)	0.000106	1.0	7746 18.267857
7744	(Produce, Baking Goods, Snack Foods)	0.000106	1.0	7744 13.556948
7736	(Produce, Frozen Foods, Baking Goods)	0.000106	1.0	7736 14.580508

## Rules:

```
rule 1 ['Canned Foods', 'Carousel', 'Deli', 'Health and Hygiene', 'Household', 'Produce', 'Snack Foods'] -> ['Dairy']
rule 2 ['Checkout', 'Deli', 'Health and Hygiene', 'Household', 'Produce', 'Snacks'] -> ['Baking Goods', 'Frozen Foods', 'Snack Foods']
rule 3 ['Baking Goods', 'Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Household', 'Snacks'] -> ['Produce', 'Snack Foods']
rule 4 ['Baking Goods', 'Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Household', 'Produce', 'Snack Foods']
rule 5 ['Baking Goods', 'Beverages', 'Checkout', 'Deli', 'Health and Hygiene', 'Household', 'Snack Foods', 'Snacks'] -> ['Frozen Foods']
rule 6 ['Baking Goods', 'Beverages', 'Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Household', 'Snack Foods'] -> ['Snacks']
rule 7 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Household', 'Produce', 'Snack Foods']
rule 8 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Produce', 'Snacks'] -> ['Baking Goods', 'Household', 'Snack Foods']
rule 9 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Household', 'Snacks'] -> ['Baking Goods', 'Produce', 'Snack Foods']
rule 10 ['Checkout', 'Deli', 'Health and Hygiene', 'Household', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Frozen Foods', 'Produce']
```

## 2. Sorted by lift

	antecedents \		consequents	support \		confidence	lift
7900	(Snacks, Frozen Foods, Health and Hygiene, Checkout, Deli)		(Beverages, Snack Foods, Produce, Baking Goods, Household)	0.000106	7900	1.0	133.278169
7888	(Snacks, Frozen Foods, Snack Foods, Produce, Checkout, Deli)		(Beverages, Health and Hygiene, Baking Goods, Household)	0.000106	7888	1.0	128.744898
7889	(Snacks, Frozen Foods, Snack Foods, Checkout, Deli, Household)		(Beverages, Produce, Health and Hygiene, Baking Goods)	0.000106	7889	1.0	95.103015
7890	(Snacks, Frozen Foods, Produce, Checkout, Baking Goods, Deli)		(Beverages, Household, Health and Hygiene, Snack Foods)	0.000106	7890	1.0	88.644028
7783	(Snacks, Frozen Foods, Health and Hygiene, Checkout, Deli)		(Beverages, Household, Baking Goods, Snack Foods)	0.000106	7783	1.0	79.518908
7882	(Snacks, Frozen Foods, Health and Hygiene, Produce, Checkout, Deli)		(Beverages, Frozen Foods, Baking Goods, Snack Foods)	0.000106	7882	1.0	79.518908
7895	(Snacks, Health and Hygiene, Produce, Checkout, Deli, Household)		(Beverages, Produce, Household, Baking Goods)	0.000106	7895	1.0	78.529046
7899	(Snacks, Snack Foods, Checkout, Baking Goods, Deli, Household)		(Beverages, Frozen Foods, Produce, Health and Hygiene)	0.000106	7899	1.0	75.550898
7681	(Snacks, Frozen Foods, Health and Hygiene, Checkout, Deli)		(Beverages, Produce, Household, Baking Goods)	0.000106	7681	1.0	65.486159
7881	(Snacks, Frozen Foods, Health and Hygiene, Snack Foods, Checkout, Deli)		(Beverages, Produce, Household, Baking Goods)	0.000106	7881	1.0	65.486159

Rules:

```
rule 1 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce', 'Snack Foods']
rule 2 ['Checkout', 'Deli', 'Frozen Foods', 'Produce', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Health and Hygiene', 'Household']
rule 3 ['Checkout', 'Deli', 'Frozen Foods', 'Household', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Health and Hygiene', 'Produce']
rule 4 ['Baking Goods', 'Checkout', 'Deli', 'Frozen Foods', 'Produce', 'Snacks'] -> ['Beverages', 'Health and Hygiene', 'Household', 'Snack Foods']
rule 5 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Snack Foods']
rule 6 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Produce', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Snack Foods']
rule 7 ['Checkout', 'Deli', 'Health and Hygiene', 'Household', 'Produce', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Frozen Foods', 'Snack Foods']
rule 8 ['Baking Goods', 'Checkout', 'Deli', 'Household', 'Snack Foods', 'Snacks'] -> ['Beverages', 'Frozen Foods', 'Health and Hygiene', 'Produce']
rule 9 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce']
rule 10 ['Checkout', 'Deli', 'Frozen Foods', 'Health and Hygiene', 'Snack Foods', 'Snacks'] -> ['Baking Goods', 'Beverages', 'Household', 'Produce']
```

## 3. 比較

- 甲、Sort by confidence 的 10 條規則中，發現 **baking goods** 經常和 **checkout** 一起出現
- 乙、Sort by list 的 10 條規則中，發現 **checkout** 經常和 **deli** 一起出現

## 第三題

Apriori on user profile

### Data preprocessing

- 資料格式是**純數字**的欄位，都會在數字後面補上欄位單位 e.g. total children, home children etc. 因為 rule 的 antecedents 和 consequents 都是不具順序性的集合，需要做事前編碼才能在跑出規則後解讀是屬於哪個欄位的資料。
- 生日**只留下**年份與月份**，並且把出生年份轉換為**年齡**(用 1998 年扣)，再把不同年齡層區分成 young(0-29), middle(30-59), old(>= 60)
- 婚姻狀態**的'M'改成 married，因為性別的 M 代表男性，如果沒有事前修改，當集合中出現一個 M 會難以解讀是代表男性還是代表已婚
- 其他格式不變



## 使用 apriori 演算法

### Sorted by confidence

	antecedents \
324786	(0 homeChildren, \$50K - \$70K, Bachelors Degree, N, 2 carNums, Guerrero)
75943	(4 carNums, Management, \$150K +, BC)
76117	(5 totalChildren, \$150K +, Management, F, WA)
76225	(Graduate Degree, \$150K +, OR, Y, M)
76224	(Partial College, WA, \$150K +)
76223	(Partial College, Professional, WA, \$150K +)
76222	(Partial College, F, WA, \$150K +)
76221	(Y, Partial College, WA, \$150K +)
76220	(\$150K +, Partial College, F, WA, Professional)
76219	(\$150K +, Partial College, Y, WA, Professional)

	consequents	support	confidence	lift
324786	(F, Professional)	0.000195	1.0	6.094250
75943	(Y, High School Degree)	0.000195	1.0	5.530393
76117	(5 homeChildren)	0.000195	1.0	52.454082
76225	(Professional)	0.000195	1.0	3.039917
76224	(Y, F, Professional)	0.000195	1.0	9.414835
76223	(Y, F)	0.000195	1.0	3.317522
76222	(Y, Professional)	0.000195	1.0	4.766342
76221	(F, Professional)	0.000195	1.0	6.094250
76220	(Y)	0.000195	1.0	1.661710
76219	(F)	0.000195	1.0	2.017069

### Rules:

```
rule 1 ['$50K - $70K', '0 homeChildren', '2 carNums', 'Bachelors Degree', 'Guerrero', 'N'] -> ['F', 'Professional']
rule 2 ['$150K +', '4 carNums', 'BC', 'Management'] -> ['High School Degree', 'Y']
rule 3 ['$150K +', '5 totalChildren', 'F', 'Management', 'WA'] -> ['5 homeChildren']
rule 4 ['$150K +', 'Graduate Degree', 'M', 'OR', 'Y'] -> ['Professional']
rule 5 ['$150K +', 'Partial College', 'WA'] -> ['F', 'Professional', 'Y']
rule 6 ['$150K +', 'Partial College', 'Professional', 'WA'] -> ['F', 'Y']
rule 7 ['$150K +', 'F', 'Partial College', 'WA'] -> ['Professional', 'Y']
rule 8 ['$150K +', 'Partial College', 'WA', 'Y'] -> ['F', 'Professional']
rule 9 ['$150K +', 'F', 'Partial College', 'Professional', 'WA'] -> ['Y']
rule 10 ['$150K +', 'Partial College', 'Professional', 'WA', 'Y'] -> ['F']
```

Note: 'Y'代表此顧客是 house owner; 'N'代表此顧客不是 house owner

解讀：職業是 professional 的顧客中，大多是 partial college 的學歷

### Sorted by lift

	antecedents \
277567	(Y, 0 carNums, 2 totalChildren, Bachelors Degree)
151158	(M, Y, 0 carNums, Bachelors Degree)
277587	(WA, \$10K - \$30K, Management)
148475	(Y, 0 carNums, 2 totalChildren, Bachelors Degree)
277570	(M, Y, 0 carNums, Bachelors Degree)
148489	(WA, \$10K - \$30K, Management)
277582	(\$10K - \$30K, WA, 2 totalChildren, Management)
151167	(WA, \$10K - \$30K, Management)
277584	(M, WA, \$10K - \$30K, Management)
277538	(Y, 0 carNums, 2 totalChildren, Bachelors Degree, M)

		consequents	support	\	confidence	lift
277567	(M, WA, \$10K - \$30K, Management)		0.000195	277567	1.0	5140.5
151158	(WA, \$10K - \$30K, Management)		0.000195	151158	1.0	5140.5
277587	(Y, 0 carNums, 2 totalChildren, Bachelors Degree, M)		0.000195	277587	1.0	5140.5
148475	(WA, \$10K - \$30K, Management)		0.000195	148475	1.0	5140.5
277570	(\$10K - \$30K, WA, 2 totalChildren, Management)		0.000195	277570	1.0	5140.5
148489	(Y, 0 carNums, 2 totalChildren, Bachelors Degree)		0.000195	148489	1.0	5140.5
277582	(M, Y, 0 carNums, Bachelors Degree)		0.000195	277582	1.0	5140.5
151167	(M, Y, 0 carNums, Bachelors Degree)		0.000195	151167	1.0	5140.5
277584	(Y, 0 carNums, 2 totalChildren, Bachelors Degree)		0.000195	277584	1.0	5140.5
277538	(WA, \$10K - \$30K, Management)		0.000195	277538	1.0	5140.5

Rules:

```
rule 1 ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'Y'] -> ['$10K - $30K', 'M', 'Management', 'WA']
rule 2 ['0 carNums', 'Bachelors Degree', 'M', 'Y'] -> ['$10K - $30K', 'Management', 'WA']
rule 3 ['$10K - $30K', 'Management', 'WA'] -> ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'M', 'Y']
rule 4 ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'Y'] -> ['$10K - $30K', 'Management', 'WA']
rule 5 ['0 carNums', 'Bachelors Degree', 'M', 'Y'] -> ['$10K - $30K', '2 totalChildren', 'Management', 'WA']
rule 6 ['$10K - $30K', 'Management', 'WA'] -> ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'Y']
rule 7 ['$10K - $30K', '2 totalChildren', 'Management', 'WA'] -> ['0 carNums', 'Bachelors Degree', 'M', 'Y']
rule 8 ['$10K - $30K', 'Management', 'WA'] -> ['0 carNums', 'Bachelors Degree', 'M', 'Y']
rule 9 ['$10K - $30K', 'M', 'Management', 'WA'] -> ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'Y']
rule 10 ['0 carNums', '2 totalChildren', 'Bachelors Degree', 'M', 'Y'] -> ['$10K - $30K', 'Management', 'WA']
```

解讀

收入 10-30k 的顧客都大多都沒有車子(0 carNums)

Bachelors degree 的顧客 occupation 大多是 management

## 第四題

參數預設設定: min\_support=0.0001; min\_confidence = 0.9; sorted by lift(因為 confidence 大多數都等於 1.0)，如果跑出來的 rule 數過多，就會增加 min\_support，反之就會減少 min\_support

印出格式：antecedents(印出的部分)都是交易中的產品類別，consequent 都是顧客屬性，因為 terminal 大小的關係只有印出 antecedents，待會會在下面篇幅說明所屬的 consequents

### 1. Gender

```
['F' 'M']
86 (Frozen Foods, Alcoholic Beverages, Produce, Baked Goods, Breakfast Foods)
891 (Frozen Foods, Alcoholic Beverages, Produce, Household, Deli, Canned Foods, Baked Goods)
396 (Alcoholic Beverages, Produce, Baking Goods, Deli, Beverages, Starchy Foods)
919 (Snack Foods, Alcoholic Beverages, Household, Baking Goods, Deli, Canned Foods, Starchy Foods)
368 (Alcoholic Beverages, Household, Baking Goods, Deli, Baked Goods, Health and Hygiene)
...
1254 (Dairy, Snacks, Snack Foods, Frozen Foods, Produce, Household, Deli, Canned Foods)
26 (Beverages, Canned Products, Baking Goods, Eggs)
600 (Dairy, Produce, Household, Periodicals, Baked Goods, Health and Hygiene)
506 (Snack Foods, Alcoholic Beverages, Household, Deli, Canned Foods, Starchy Foods)
324 (Canned Products, Eggs, Produce, Household, Health and Hygiene)
Name: antecedents, Length: 105, dtype: object
-----
179 (Frozen Foods, Canned Foods, Starchy Foods, Baked Goods, Health and Hygiene)
597 (Snack Foods, Frozen Foods, Canned Foods, Starchy Foods, Baked Goods, Health and Hygiene)
596 (Frozen Foods, Household, Canned Foods, Starchy Foods, Baked Goods, Health and Hygiene)
4 (Canned Products, Household, Alcoholic Beverages, Baked Goods)
Name: antecedents, dtype: object
-----
```

(虛線上是女生(consequent = 'F'), 虛線下是男生(consequent='M'))

甲、在分析中發現，男生所購買的 itemSet pattern 較為固定，女生購買的 itemSet 較為多樣化，而 canned foods 經常會和 baked goods 一起購買

## 2. marital\_status

```
355 (Household, Snacks, Frozen Foods, Checkout, Deli)
143 (Frozen Foods, Checkout, Health and Hygiene, Produce, Alcoholic Beverages)
973 (Household, Beverages, Baking Goods, Frozen Foods, Checkout, Deli, Produce)
21 (Frozen Foods, Baked Goods, Checkout, Breakfast Foods)
357 (Snacks, Frozen Foods, Checkout, Deli, Produce)
791 (Household, Frozen Foods, Snacks, Checkout, Deli, Produce)
631 (Snack Foods, Baking Goods, Snacks, Checkout, Canned Foods, Health and Hygiene)
23 (Frozen Foods, Snacks, Baked Goods, Checkout)
227 (Baking Goods, Frozen Foods, Canned Foods, Checkout, Periodicals)
172 (Household, Beverages, Eggs, Snacks, Baked Goods)
419 (Household, Baking Goods, Frozen Foods, Checkout, Produce, Alcoholic Beverages)
129 (Frozen Foods, Canned Foods, Checkout, Health and Hygiene, Alcoholic Beverages)
709 (Household, Beverages, Frozen Foods, Checkout, Canned Foods, Dairy)
110 (Baking Goods, Frozen Foods, Checkout, Dairy, Alcoholic Beverages)
285 (Beverages, Frozen Foods, Meat, Canned Foods, Dairy)
651 (Household, Starchy Foods, Baking Goods, Canned Products, Dairy, Produce)
205 (Starchy Foods, Eggs, Health and Hygiene, Produce, Baked Goods)
358 (Household, Snacks, Checkout, Deli, Produce)
192 (Household, Frozen Foods, Checkout, Dairy, Baked Goods)
182 (Frozen Foods, Canned Foods, Checkout, Dairy, Baked Goods)
500 (Snack Foods, Frozen Foods, Eggs, Deli, Dairy, Alcoholic Beverages)
Name: antecedents, dtype: object
-----
772 (Household, Snack Foods, Starchy Foods, Frozen Foods, Breakfast Foods, Dairy)
725 (Household, Beverages, Snack Foods, Seafood, Canned Foods, Produce)
324 (Household, Starchy Foods, Frozen Foods, Breakfast Foods, Dairy)
58 (Snacks, Deli, Canned Foods, Canned Products)
325 (Starchy Foods, Frozen Foods, Breakfast Foods, Dairy, Produce)
312 (Starchy Foods, Frozen Foods, Canned Foods, Breakfast Foods, Dairy)
564 (Household, Canned Foods, Breakfast Foods, Dairy, Health and Hygiene, Baked Goods)
320 (Frozen Foods, Canned Foods, Breakfast Foods, Periodicals, Produce)
326 (Snack Foods, Starchy Foods, Frozen Foods, Breakfast Foods, Dairy)
135 (Household, Snack Foods, Seafood, Canned Foods, Alcoholic Beverages)
```

(虛線上是已婚(consequents = 'married'), 虛線下是未婚(consequents = 'S'))

甲、在分析中發現單身(consequent = 'S')的人購買的產品 pattern 較為固定，starchy foods、frozen foods 和 breakfast foods 經常一起購買

乙、在已婚的 itemSets 中，checkout, dairy 和 produce(生鮮產品)常一起出現

## 3. total\_children: 反映家庭狀況，

在分析 total children 的 itemSets 時，min\_support 原先設成 0.0001 的情況下，除了 total children = 0 的 itemSets 為空，其他都有資料，後來有再把 min\_support 降低為 0.00005 去分析沒有小孩顧客的 frequent itemSets，但後來因考量到沒有小孩的顧客還可以往下區分成已婚和未婚的情況，所以把 min\_support 設成 0.0001 去分析不同小孩數量的購買 pattern

甲、從 5-total children(consequents = '5 total children')的家庭購買 pattern 來看，beverages 經常和 deli(熟食)一起出現

```

16038      (Snack Foods, Baked Goods, Household, Produce, Eggs, Health and Hygiene, Beverages, Deli)
15650      (Baked Goods, Household, Baking Goods, Produce, Eggs, Health and Hygiene, Beverages, Deli)
11654      (Baked Goods, Household, Baking Goods, Produce, Eggs, Beverages, Deli)
15742      (Snack Foods, Baked Goods, Household, Baking Goods, Produce, Eggs, Beverages, Deli)
11620      (Baked Goods, Household, Baking Goods, Eggs, Health and Hygiene, Beverages, Deli)
15677      (Snack Foods, Baked Goods, Household, Baking Goods, Eggs, Health and Hygiene, Beverages, Deli)
17478      (Snack Foods, Baked Goods, Household, Baking Goods, Produce, Eggs, Health and Hygiene, Beverages, Deli)
12117      (Frozen Foods, Dairy, Baking Goods, Starchy Foods, Health and Hygiene, Beverages, Deli)
Name: antecedents, dtype: object

```

乙、在 1-total children itemSets(consequent = '1 total children')中，可看出  
eggs 經常和 periodicals 一起出現

```

4454      (Produce, Dairy, Household, Starchy Foods, Periodicals, Eggs)
9945      (Produce, Dairy, Baking Goods, Household, Starchy Foods, Periodicals, Eggs)
4320      (Produce, Baking Goods, Household, Starchy Foods, Periodicals, Eggs)
959      (Produce, Starchy Foods, Periodicals, Eggs, Alcoholic Beverages)
14987      (Frozen Foods, Dairy, Produce, Baking Goods, Household, Health and Hygiene, Snacks, Alcoholic Beverages)
Name: antecedents, dtype: object

```

4. member\_card：反映會員忠誠度，可透過關聯分析找出不同忠誠度會員喜愛購買的產品差異

甲、當 min\_support 增加為 0.0002 and min\_confidence = 0.9，只有 **bronze card** 有資料，但其他會員等級沒有資料。後來發現原因是 10282 位會員中，有 5703 位是銅卡會員。而銅卡會員的消費行為中，發現會買 **Deli(罐頭肉品)**的銅卡顧客，通常會一起購買 **canned foods 以及 frozen foods**

```

[frozenset({'Baked Goods', 'Produce', 'Alcoholic Beverages', 'Baking Goods', 'Health and Hygiene', 'Dairy'})
frozenset({'Alcoholic Beverages', 'Seafood', 'Baked Goods', 'Canned Foods'})
frozenset({'Deli', 'Canned Foods', 'Frozen Foods', 'Produce', 'Household', 'Breakfast Foods'})
frozenset({'Deli', 'Canned Foods', 'Frozen Foods', 'Household', 'Breakfast Foods', 'Health and Hygiene'})
frozenset({'Carousel', 'Household', 'Canned Products'})
frozenset({'Beverages', 'Frozen Foods', 'Canned Foods', 'Alcoholic Beverages', 'Baking Goods', 'Health and Hygiene'})
frozenset({'Deli', 'Canned Foods', 'Alcoholic Beverages', 'Baking Goods', 'Health and Hygiene', 'Dairy'})
frozenset({'Deli', 'Frozen Foods', 'Canned Foods', 'Alcoholic Beverages', 'Baking Goods', 'Health and Hygiene'})
frozenset({'Periodicals', 'Meat', 'Snack Foods', 'Health and Hygiene'})
frozenset({'Breakfast Foods', 'Produce', 'Meat', 'Canned Foods'})

```

乙、後來把 min\_support 放寬設成 0.0001，發現 **normal card** 的購物 pattern：normal card 的顧客經常 **beverages, canned products 以及 checkout(雜物 e.g. 螺絲起子、開罐器)**一起購買

```

[frozenset({'Household', 'Beverages', 'Canned Products', 'Checkout'})
frozenset({'Canned Products', 'Snack Foods', 'Beverages', 'Checkout'})
frozenset({'Canned Products', 'Beverages', 'Frozen Foods', 'Checkout'})
frozenset({'Checkout', 'Beverages', 'Canned Products'})]

```

5. age：由 birthdate 切分而來，再用 1998 - birth\_date 的 year(出生年份)得到顧客年齡，看不同年齡層是否有不同偏好的產品，再進一步切分成年輕(0-29)、中年(30-59)、老年(>=60)三個年齡層

甲、min\_support 設成 0.0002，依照 lift 排序 pattern 的前 11 筆中，發現 **中年顧客中(consequent='middle')**，**starchy foods 經常和 baking goods 一起購買**

乙、此外，**household(廚具)**經常和 **starchy foods(e.g. pasta)** 出現



```
[frozenset({'Starchy Foods', 'Beverages', 'Deli', 'Snacks', 'Snack Foods', 'Baking Goods'})
frozenset({'Starchy Foods', 'Beverages', 'Health and Hygiene', 'Deli', 'Snacks'})
frozenset({'Starchy Foods', 'Health and Hygiene', 'Deli', 'Snacks', 'Snack Foods', 'Baking Goods'})
frozenset({'Alcoholic Beverages', 'Checkout', 'Periodicals'})
frozenset({'Alcoholic Beverages', 'Dairy', 'Snack Foods', 'Eggs', 'Baked Goods'})
frozenset({'Starchy Foods', 'Produce', 'Beverages', 'Deli', 'Snacks', 'Baking Goods'})
frozenset({'Produce', 'Frozen Foods', 'Periodicals', 'Health and Hygiene', 'Deli', 'Canned Foods'})
frozenset({'Starchy Foods', 'Frozen Foods', 'Seafood', 'Household', 'Canned Foods'})
frozenset({'Frozen Foods', 'Meat', 'Snack Foods', 'Canned Foods', 'Dairy', 'Baking Goods'})
frozenset({'Dairy', 'Frozen Foods', 'Snack Foods', 'Eggs', 'Baked Goods', 'Baking Goods'})
frozenset({'Starchy Foods', 'Household', 'Deli', 'Snacks', 'Snack Foods', 'Baking Goods'})]
```

丙、因為年輕顧客 pattern 很多，min\_support 設成 0.1 的情況下，把 pattern 使用 lift 排序，發現年輕年齡層(consequent = 'young')breakfast foods 和 canned products 常一起出現，推測是因為 breakfast 是一些澱粉類食物 (e.g. pancake, waffle, etc.)，canned products 是一些湯罐頭，兩個搭配當早餐吃

```
[frozenset({'Baking Goods', 'Beverages', 'Frozen Foods', 'Canned Products', 'Breakfast Foods'})
frozenset({'Baking Goods', 'Carousel', 'Canned Foods', 'Alcoholic Beverages', 'Household', 'Health and Hygiene'})
frozenset({'Household', 'Deli', 'Seafood', 'Health and Hygiene', 'Snack Foods', 'Snacks'})
frozenset({'Dairy', 'Baking Goods', 'Carousel', 'Canned Foods', 'Alcoholic Beverages'})
frozenset({'Baking Goods', 'Beverages', 'Produce', 'Canned Products', 'Breakfast Foods'})
frozenset({'Baking Goods', 'Beverages', 'Household', 'Canned Products', 'Breakfast Foods'})
frozenset({'Baking Goods', 'Carousel', 'Canned Foods', 'Alcoholic Beverages', 'Health and Hygiene'})]
```

## 第五題

比較 12 月與 1~11 月的顧客購物行為。有哪些相似的地方，有哪些差異  
以下使用apriori演算法，並且依照lift進行排序(因為confidence幾乎都是1.0)  
12月顧客購物行為

	antecedents	lift
1998551	(Deli, Checkout, Household, Canned Products)	3781.0
2389152	(Frozen Foods, Snack Foods, Canned Products, Checkout, Produce, Health and Hygiene, Seafood)	3781.0
2389050	(Frozen Foods, Household, Beverages, Canned Products, Checkout, Produce, Snack Foods)	3781.0
2389141	(Frozen Foods, Baking Goods, Canned Products, Breakfast Foods, Checkout, Snack Foods, Health and Hygiene)	3781.0
2389142	(Frozen Foods, Baking Goods, Canned Products, Breakfast Foods, Checkout, Snack Foods, Seafood)	3781.0
2389143	(Frozen Foods, Baking Goods, Canned Products, Breakfast Foods, Snack Foods, Health and Hygiene, Seafood)	3781.0
2389144	(Frozen Foods, Snack Foods, Canned Products, Breakfast Foods, Checkout, Produce, Health and Hygiene)	3781.0
2389145	(Frozen Foods, Snack Foods, Canned Products, Breakfast Foods, Checkout, Produce, Seafood)	3781.0
2389146	(Frozen Foods, Snack Foods, Canned Products, Breakfast Foods, Produce, Health and Hygiene, Seafood)	3781.0
2389148	(Frozen Foods, Snack Foods, Baking Goods, Canned Products, Checkout, Produce, Health and Hygiene)	3781.0

12月顧客購物行為加上長度計算

	antecedents	len
1998551	(Canned Products, Beverages, Deli, Checkout)	11
2389141	(Beverages, Baking Goods, Snack Foods, Seafood, Canned Products, Snacks, Household)	14
2388956	(Health and Hygiene, Beverages, Snack Foods, Checkout, Canned Products, Baked Goods, Household)	14
2388950	(Health and Hygiene, Beverages, Snack Foods, Checkout, Seafood, Baked Goods, Household)	14
2388951	(Health and Hygiene, Beverages, Snack Foods, Checkout, Seafood, Baked Goods, Frozen Foods)	14
2388952	(Health and Hygiene, Beverages, Snack Foods, Checkout, Seafood, Snacks, Household)	14
2388953	(Health and Hygiene, Beverages, Snack Foods, Checkout, Seafood, Snacks, Frozen Foods)	14
2388954	(Health and Hygiene, Beverages, Snack Foods, Checkout, Seafood, Household, Frozen Foods)	14
2388955	(Health and Hygiene, Beverages, Snack Foods, Checkout, Canned Products, Baked Goods, Snacks)	14
2389047	(Health and Hygiene, Beverages, Deli, Canned Products, Snacks, Household, Frozen Foods)	14

1-11月顧客購物行為



		antecedents	...	lift
8329	(Frozen Foods, Snacks, Household, Baking Goods, Checkout, Snack Foods)	...	104.509202	
8330	(Frozen Foods, Snacks, Household, Baking Goods, Checkout, Health and Hygiene)	...	71.127349	
8323	(Canned Foods, Frozen Foods, Snacks, Checkout, Snack Foods, Health and Hygiene)	...	65.143403	
7881	(Frozen Foods, Snacks, Household, Baking Goods, Checkout, Snack Foods)	...	58.439108	
8313	(Frozen Foods, Snacks, Household, Baking Goods, Checkout, Produce, Snack Foods)	...	58.439108	
8324	(Canned Foods, Frozen Foods, Snacks, Household, Baking Goods, Checkout)	...	56.973244	
8212	(Deli, Snacks, Alcoholic Beverages, Household, Snack Foods, Starchy Foods)	...	52.094801	
4930	(Deli, Frozen Foods, Baked Goods, Eggs, Starchy Foods)	...	46.480218	
8328	(Canned Foods, Snacks, Beverages, Checkout, Produce, Health and Hygiene)	...	45.426667	
8333	(Snacks, Baking Goods, Beverages, Checkout, Produce, Health and Hygiene)	...	44.477807	

### 1-11月顧客購物行為加上長度計算

		antecedents	...	len
8332	(Baking Goods, Snack Foods, Checkout, Snacks, Household, Frozen Foods)	...	10	
8330	(Health and Hygiene, Baking Goods, Checkout, Snacks, Household, Frozen Foods)	...	10	
8331	(Health and Hygiene, Snack Foods, Checkout, Snacks, Frozen Foods, Canned Foods)	...	10	
7883	(Baking Goods, Snack Foods, Checkout, Snacks, Household, Frozen Foods)	...	9	
8305	(Produce, Baking Goods, Snack Foods, Checkout, Snacks, Household, Frozen Foods)	...	10	
8333	(Baking Goods, Checkout, Snacks, Household, Frozen Foods, Canned Foods)	...	10	
8213	(Snack Foods, Deli, Alcoholic Beverages, Snacks, Household, Starchy Foods)	...	10	
4931	(Deli, Baked Goods, Frozen Foods, Eggs, Starchy Foods)	...	8	
8324	(Produce, Health and Hygiene, Beverages, Checkout, Snacks, Canned Foods)	...	10	
8323	(Produce, Health and Hygiene, Beverages, Baking Goods, Checkout, Snacks)	...	10	

### 12月與非12月購物行為比較

1. 12月時，顧客一次性購買的數量較多(len欄位)
2. Seafood在12月時，購買次數大幅提升，相較於1-11月，12月時seafood和其他日常用品共同出現的次數上升
3. Lift在12月時較高且標準差小，推測是因為大家在聖誕節時因為要慶祝而購買的商品類似