

1. 從 Foodmart Data 的交易資料中，找出符合 Minimum Support = 0.00015 且 Minimum Confidence = 0.8 的 Association Rules，並分別列出 Confidence 最高的前 10 條 Rules 以及 Lift 最高的前 10 條 Rules，並比較這兩者的異同。

▼ 表一 Transaction Confidence 最高的前 10 條

	antecedents	consequents	support	confidence	lift
1	{'Imagine Low Fat French Fries', 'CDR Hot Chocolate'}	{'Quick Extra Lean Hamburger'}	0.000185	1	320.2203
2	{'High Top Summer Squash', 'Even Better Sharp Cheddar Cheese'}	{'High Top New Potatoes'}	0.000159	1	267.9858
3	{'Booker Low Fat String Cheese', 'Bravo Fancy Canned Anchovies'}	{'High Top Oranges'}	0.000159	1	312.281
4	{'Plato French Roast Coffee', 'High Quality Scissors'}	{'Dollar Monthly Sports Magazine'}	0.000159	1	312.281
5	{'Hilltop 200 MG Acetaminophen', 'Just Right Canned Yams'}	{'Faux Products HCL Nasal Spray'}	0.000159	1	337.375
6	{'Hilltop 200 MG Acetaminophen', 'Cormorant Scented Toilet Tissue'}	{'Horatio No Salt Popcorn'}	0.000159	1	299.8889
7	{'Quick Extra	{'Imagine	0.000185	0.875	292.5907

	Lean Hamburger', 'CDR Hot Chocolate'}	Low Fat French Fries'}			
8	{ 'High Top New Potatoes', 'Even Better Sharp Cheddar Cheese' }	{ 'High Top Summer Squash' }	0.000159	0.857143	279.2069
9	{ 'High Top New Potatoes', 'High Top Summer Squash' }	{ 'Even Better Sharp Cheddar Cheese' }	0.000159	0.857143	265.4754
10	{ 'BBB Best Tomato Sauce', 'Imagine Frozen Cheese Pizza' }	{ 'Best Choice Apple Fruit Roll' }	0.000159	0.857143	291.7838

▼ 表二 Transaction Lift 最高的前 10 條

	antecedents	consequents	support	confidence	lift
1	{ 'Hilltop 200 MG Acetaminophen', 'Just Right Canned Yams' }	{ 'Faux Products HCL Nasal Spray' }	0.000159	1	337.375
2	{ 'Imagine Low Fat French Fries', 'CDR Hot Chocolate' }	{ 'Quick Extra Lean Hamburger' }	0.000185	1	320.2203
3	{ 'Booker Low Fat String Cheese', 'Bravo Fancy Canned Anchovies' }	{ 'High Top Oranges' }	0.000159	1	312.281
4	{ 'Plato French Roast Coffee', 'High Quality	{ 'Dollar Monthly Sports	0.000159	1	312.281

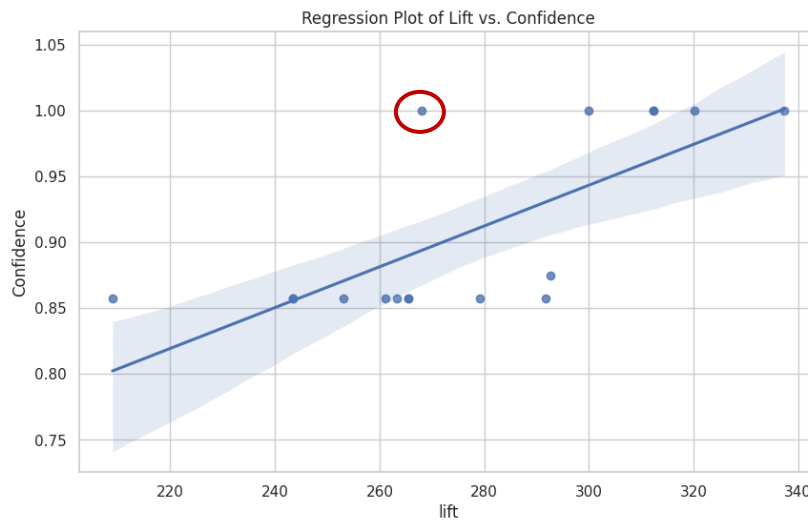
	Scissors'}	Magazine'}			
5	{'Hilltop 200 MG Acetaminophen', 'Cormorant Scented Toilet Tissue'}	{'Horatio No Salt Popcorn'}	0.000159	1	299.8889
6	{'Quick Extra Lean Hamburger', 'CDR Hot Chocolate'}	{'Imagine Low Fat French Fries'}	0.000185	0.875	292.5907
7	{'BBB Best Tomato Sauce', 'Imagine Frozen Cheese Pizza'}	{'Best Choice Apple Fruit Roll'}	0.000159	0.857143	291.7838
8	{'High Top New Potatoes', 'Even Better Sharp Cheddar Cheese'}	{'High Top Summer Squash'}	0.000159	0.857143	279.2069
9	{'High Top Summer Squash', 'Even Better Sharp Cheddar Cheese'}	{'High Top New Potatoes'}	0.000159	1	267.9858
10	{'High Top New Potatoes', 'High Top Summer Squash'}	{'Even Better Sharp Cheddar Cheese'}	0.000159	0.857143	265.4754

結果分析

整體而言，lift 和 confidence 的前十名排序大致相同，通常 lift 越高，confidence 也會越高，表示購買前項且購買後項的機率很高，該規則具有參考價值。

不過，其中有一個較特別的關聯規則：

('High Top Summer Squash', 'Even Better Sharp Cheddar Cheese') → ('High Top New Potatoes')，相對於其他 confidence 同為 1 的規則，其 lift 較低，如下圖所示：



發生這種情況可能是因為 High Top New Potatoes 本身在整體交易中的出現頻率較高。雖然每當購買 High Top Summer Squash 和 Even Better Sharp Cheddar Cheese 時，High Top New Potatoes 也一定會被購買，但根據公式：

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B)}{P(B)}$$

當 High Top New Potatoes 本身的出現頻率很高時，confidence 也高時，lift 有機會也會跟著降低。

2. 有時候我們有興趣的資料不只有產品間的資訊，也會想要由 User Profile 探索顧客的基本資料。在給定 Minimum Support = 0.05 且 Minimum Confidence = 0.9 的條件下，探勘 Foodmart 顧客基本資料的屬性 {customer_state_province, yearly_income, gender, total_children, num_children_at_home, education, occupation, homeowner} 間的 Association Rules，並列出 10 條。

▼ 表三 Customer Lift 最高的前 10 條

	antecedents	consequents	support	confidence	lift
1	(\$50K - \$70K, Professional)	(Bachelors Degree)	0.096975	0.951336	3.734511
2	(homeowner_yes, \$50K - \$70K, Professional)	(Bachelors Degree)	0.053691	0.945205	3.710446
3	(Professional, \$50K - \$70K, num_children_at_home_0)	(Bachelors Degree)	0.062543	0.942815	3.701063
4	(\$10K - \$30K, Manual, num_children_at_home_0)	(Partial High School)	0.064974	0.969521	3.221605
5	(\$10K - \$30K, Manual, homeowner_yes)	(Partial High School)	0.060014	0.968603	3.218554
6	(\$10K - \$30K, Manual, F)	(Partial High School)	0.050092	0.968045	3.216701
7	(\$10K - \$30K, Manual)	(Partial High School)	0.101449	0.964847	3.206075
8	(Skilled Manual, \$10K - \$30K, homeowner_yes)	(Partial High School)	0.055637	0.964587	3.205209
9	(\$10K - \$30K, Manual, M)	(Partial High School)	0.051357	0.961749	3.195778
10	(Skilled Manual, \$10K - \$30K, num_children_at_home_0)	(Partial High School)	0.060208	0.95969	3.188937

資料前處理

由於原數據中的 `total_children` 和 `num_children_at_home` 為數字，為避免演算法將其視為相同的值，因此在進行關聯規則分析之前，將兩個數值特徵加上前綴標註。例如：當 `total_children` 值為 0，以 `total_children_0` 命名。

關聯規則結果

透過 **Apriori** 算法，統整為以下具代表性的結果：

1. (\$50K - \$70K, Professional) -> (Bachelors Degree)

當顧客的年收入在 5-7 萬之間，且職業為 **Professional** 時，97% 的顧客擁有大學教育程度。該規則的 **Lift=3.73**，顯示收入、職業與教育程度之間具有強烈的正向關聯。

2. (\$10K - \$30K, num_children_at_home_0, Manual) -> (Partial High School)

當顧客年收入介於 1~3 萬、家裡無子女，且從事 **Manual** 工作時，教育程度有 96.7% 未完成高中學業。該規則的 **support=0.065**，**lift=3.22**，顯示年收入和職業類型對教育程度有非常高的關聯，具有參考價值。

3. (\$10K - \$30K, Manual, homeowner_yes) -> (Partial High School) & (Skilled Manual, \$10K - \$30K, homeowner_yes) -> (Partial High School)

從事 **Manual** 或 **Skilled manual** 工作、年收入介於 1~3 萬且擁有房地產的顧客皆約有 96% 未完成高中學業。

4. (\$10K - \$30K, F, Manual) -> (Partial High School) & (\$10K - \$30K, M, Manual) -> (Partial High School)

當年收入介於 1~3 萬之間且從事 **Manual** 工作的顧客中，女性 (**confidence = 0.968**) 未完成高中學業的機率略高於男性 (**confidence = 0.961**) 。這組合中的男女比例相對接近，兩者皆佔全體顧客的約 5% 。

結果分析與討論

1. 收入和教育背景的關聯

年收入較低的顧客(1~3 萬)通常擁有較低的學歷背景(**Partial High School**)，而年收入較高的顧客(5-7 萬) 擁有較高的學歷背景(**Bachelors Degree**) 。這表示低收入群體可能因為教育程度較低而限制他們的年收入，也可能是因為收入低，導致教育機會不足。

2. 年收入與房地產的關聯

從第五筆與第八筆資料發現了一個驚人的結果，當顧客年收入介於 1~3 萬，且從事 **Manual** 或 **Skilled Manual** 工作時，約有 11% 的顧客擁有房地產。可推知儘管年收入較低，這些顧客群體或許對房產有的強烈需求，因此透過貸款或其他金融工具購買房地產。

透過顧客的關聯式分析，我們能更清楚顧客的特徵。這些洞察能幫助企業更精確地進行市場細分，針對不同顧客群體提供量身訂做的產品與服務，以更有效地滿足不同顧客的需求。

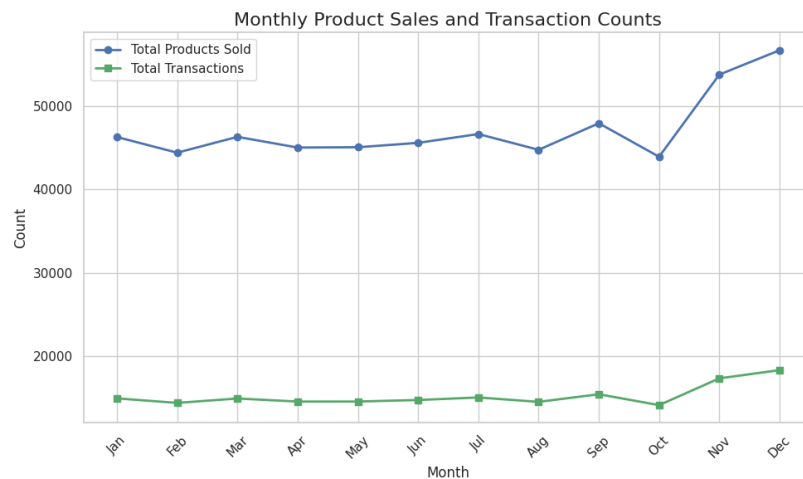
3. 在美國由於聖誕節，12 月是購物的旺季。請探索分析比較 12 月與 1~11 月的顧客購物行為。 有哪些相似的地方，有哪些差異的地方？

全年銷售量與交易次數

▼ 表四 1~12 月總銷售量、交易次數

month	total_products_sold	total_transactions
1	46313	14936
2	44431	14392
3	46334	14924
4	45049	14552
5	45085	14556
6	45611	14741
7	46671	15044
8	44777	14517
9	47964	15432
10	43945	14125
11	53807	17339
12	56729	18325

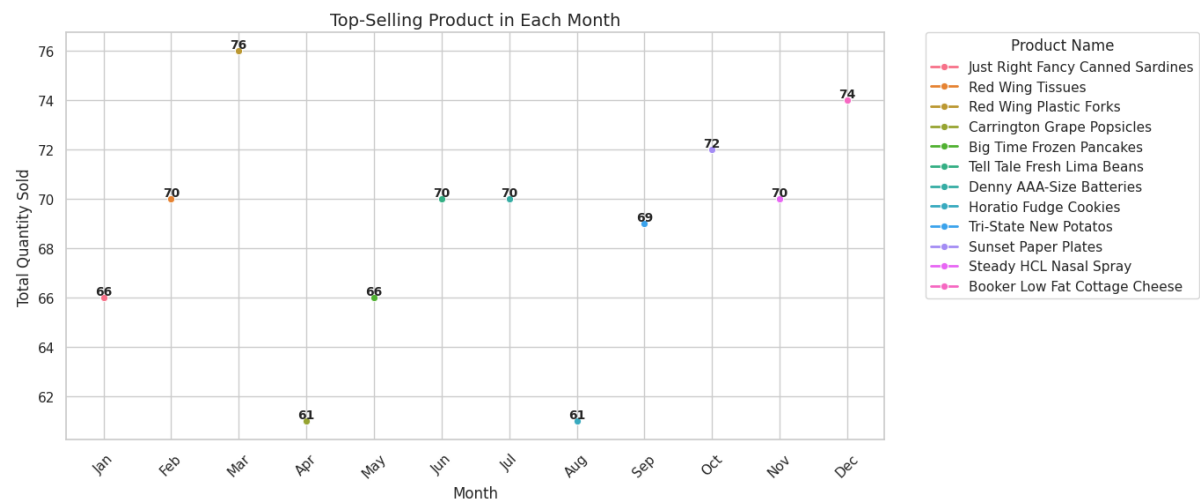
▼ 圖一 1~12 月總銷售量、交易次數折線圖



從表四和圖一中可以看出，十二月份的總銷售量和總交易次數顯著高於其他十一個月。十月份的銷售和交易量處於全年最低點，隨後在十一月份逐步回升，最終在十二月達到最高點，顯示出明顯的增長。

月銷售量冠軍

▼ 圖二 1~12 月 銷售量冠軍



從圖二可以看出，1988 年每月銷售量最高的產品數量相近，而且單項產品銷售量最高的並不在 12 月，其排名只在第二的「Booker Low Fat Cottage Cheese」，月銷售量為 74 件。其中 3 月份的銷售冠軍為「Red Wing Plastic Forks」，銷售量達 76 件，成為月銷售量最高的產品。整體來看，雖然每月的銷售量冠軍銷售數量大致相當，但 4 月和 8 月的銷售冠軍皆僅售出 61 件，顯示出這兩個月的銷售量明顯較低。

6 月與 12 月銷售量差距最大的產品分析

▼ 表五 6 月與 12 月銷量差距最大的前十名產品

	product_id	product_name	quantity_december	quantity_june	ratio
1069	559	Fast Dried Dates	37	5	7.400000
933	436	Steady 200 MG Acetominifen	43	6	7.166667
1288	756	Cormorant Plastic Spoons	50	7	7.142857
517	1464	Amigo Lox	50	7	7.142857
1125	609	Landslide Pepper	54	8	6.750000
1364	824	Ebony Peaches	50	8	6.250000
1059	55	Club Large Curd Cottage Cheese	45	8	5.625000
239	1213	Sphinx White Bread	42	8	5.250000
673	201	High Top Plums	31	6	5.166667
898	404	Big Time Frozen Mushroom Pizza	56	11	5.090909

本表在比較顧客在一般月份（以 6 月為例）和購物旺季（12 月）購買產品的差異。從表五可以看出，12 月的產品銷量相比 6 月有顯著增長，這可能與聖誕節，還有年末促銷有關。許多與慶祝活動或聚餐相關的食品，如“Fast Dried Dates”、“Sphinx White Bread”、“Big Time Frozen Mushroom Pizza”、“Ebony Peaches”、“High Top Plums”和“Amigo Lox”等，需求量在 12 月大幅上升。此外，天氣變冷也可能影響顧客的購物行為，因為寒冷的 12 月容易引發感冒，導致“Steady 200 MG Acetominifen”等感冒藥品的銷量增加。