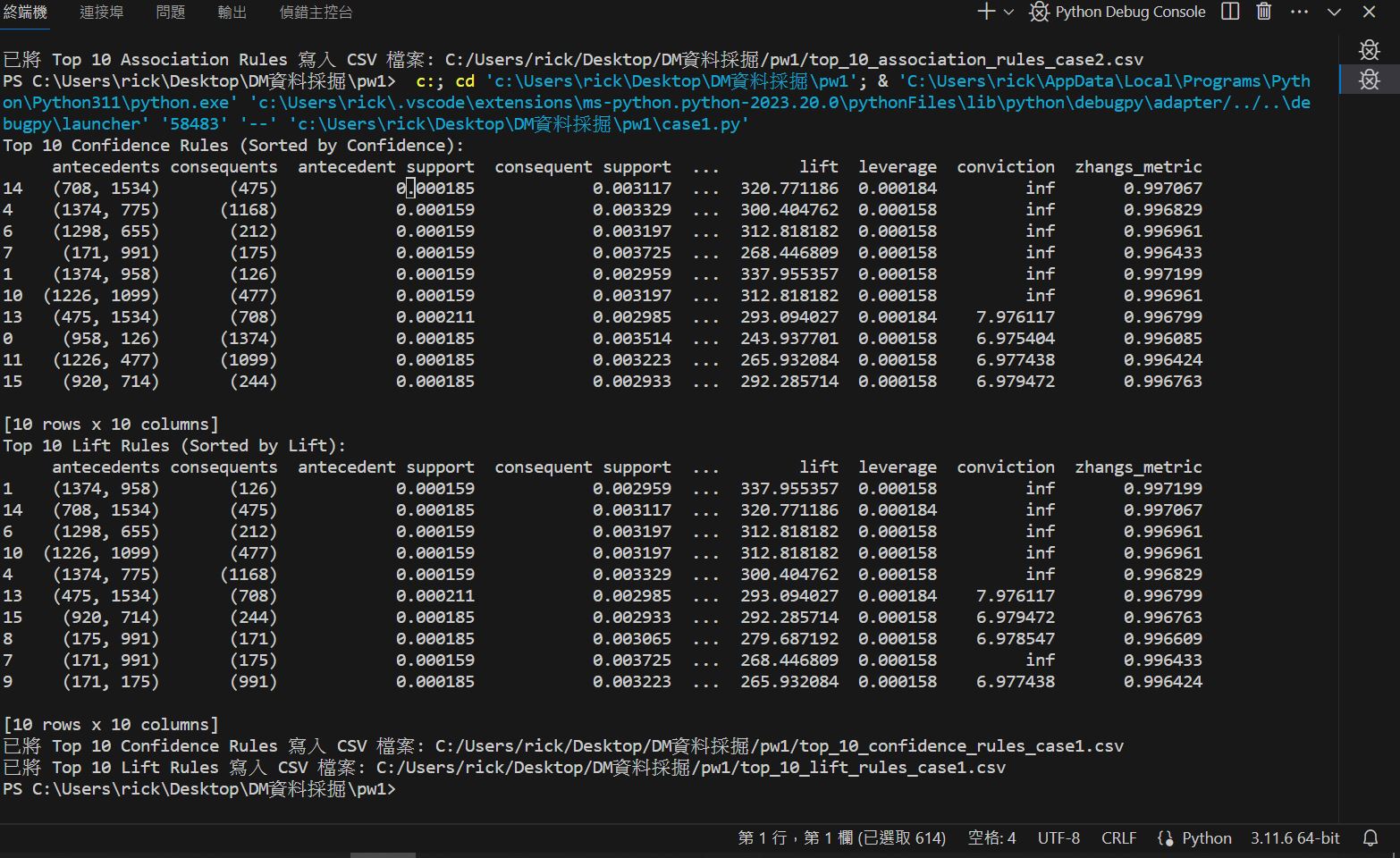
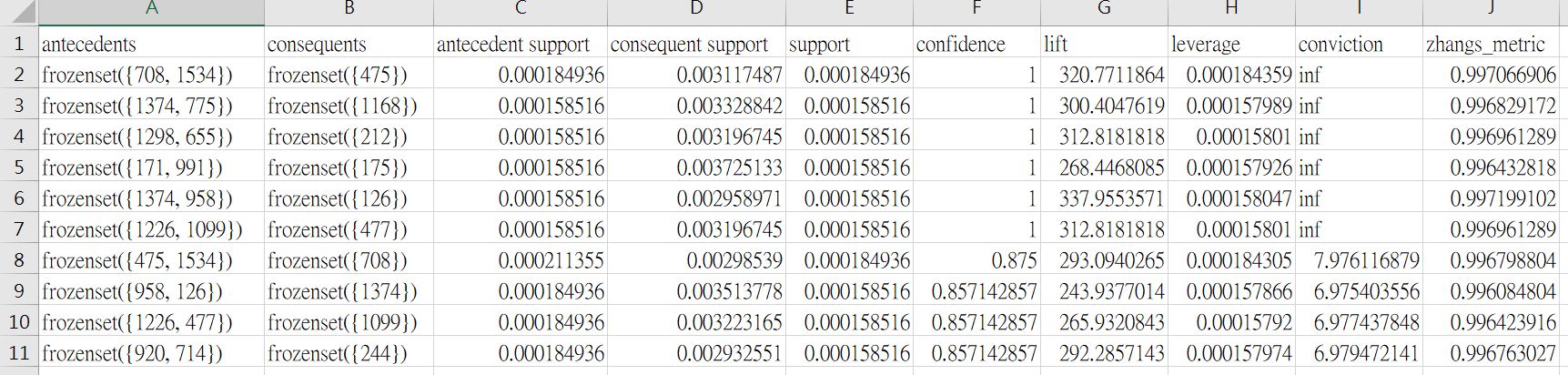
Programming Assignment 1 Deadline: 11/13

此次作業主要目的在讓同學學習運用 Python 由 Foodmart Database 超級市場的顧客及其交易資料中，探勘 Frequent Itemsets、Association Rules、Quantitative Association Rules。

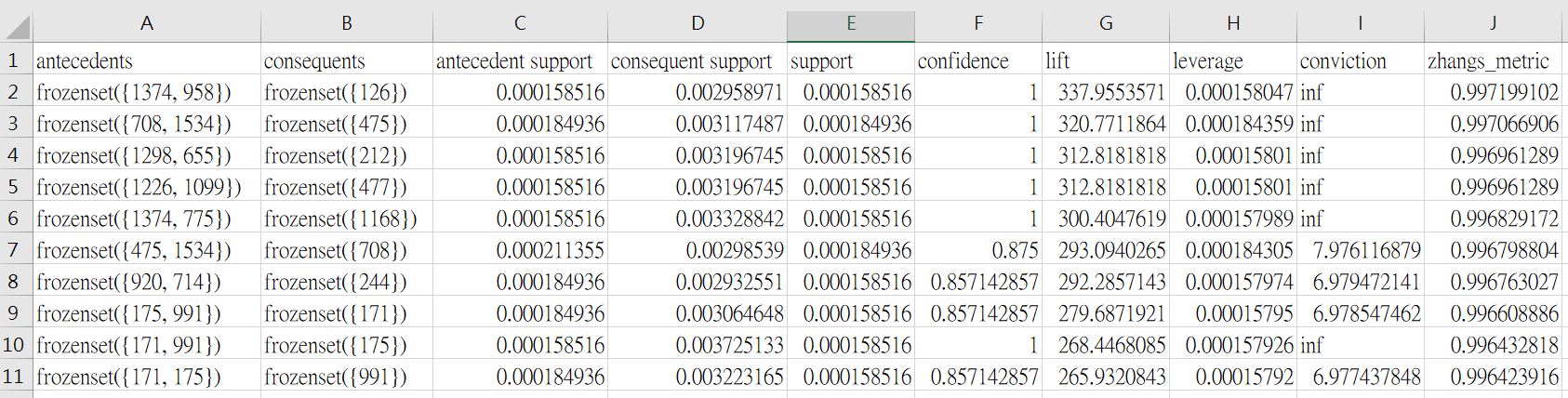
1. 從 Foodmart Data 的交易資料中，探勘符合 Minimum Support = 0.00015 且 Minimum Confidence = 0.8 的 Association Rules，並列出Confidence 最高的前 10 條 Rules 以及 lift 最高的前 10 條，並比較這兩者的異同。



Top 10 Confidence Rules



Top 10 Lift Rules



Confidence最高的Rules告訴我們當某些項目（前項antecedents）出現時，另一些項目（後項consequents）出現的可能性有多高，而Lift最高的Rules則告訴我們這些項目（前項和後項）之間的關係有多強。兩組Rules的前後項有些相似，但排序不同。

|  |  |
| --- | --- |
| Confidence 最高的前 10 條：  (708, 1534) => (475)  (1374, 775) => (1168)  (1298, 655) => (212)  (171, 991) => (175)  (1374, 958) => (126)  (1226, 1099) => (477)  (475, 1534) => (708)  (958, 126) => (1374)  (1226, 477) => (1099)  (920, 714) => (244) | Lift 最高的前 10 條：  (1374, 958) => (126)  (708, 1534) => (475)  (1298, 655) => (212)  (1226, 1099) => (477)  (1374, 775) => (1168)  (475, 1534) => (708)  (920, 714) => (244)  (175, 991) => (171)  (171, 991) => (175)  (171, 175) => (991) |

Confidence和Lift都是評估項目之間關係的指標，但它們的計算方式不同。

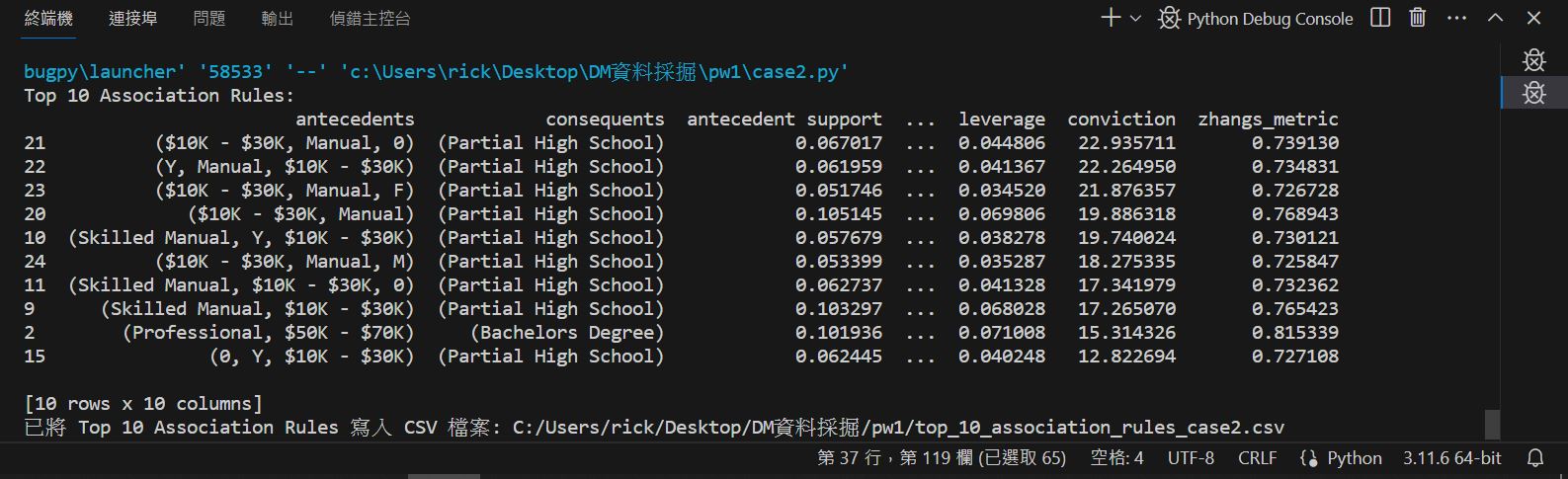
Confidence是規則的信心度，表示在前項出現的情況下，後項出現的可能性有多高。

Lift是規則的提升度，表示前項和後項之間的相關性有多強，相對於它們在不同條件下的隨機出現機率，即規則的信心度與結論出現機率的比值。

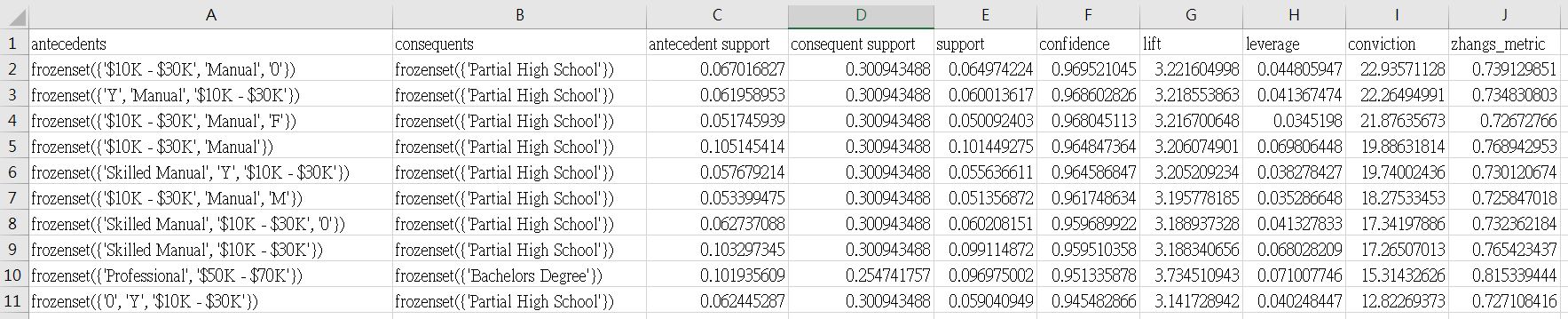
這兩組Rules可能有不同的排序方式，但都提供了關於項目之間關聯性的有用訊息。

1. 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 探勘顧客的基本資料。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。



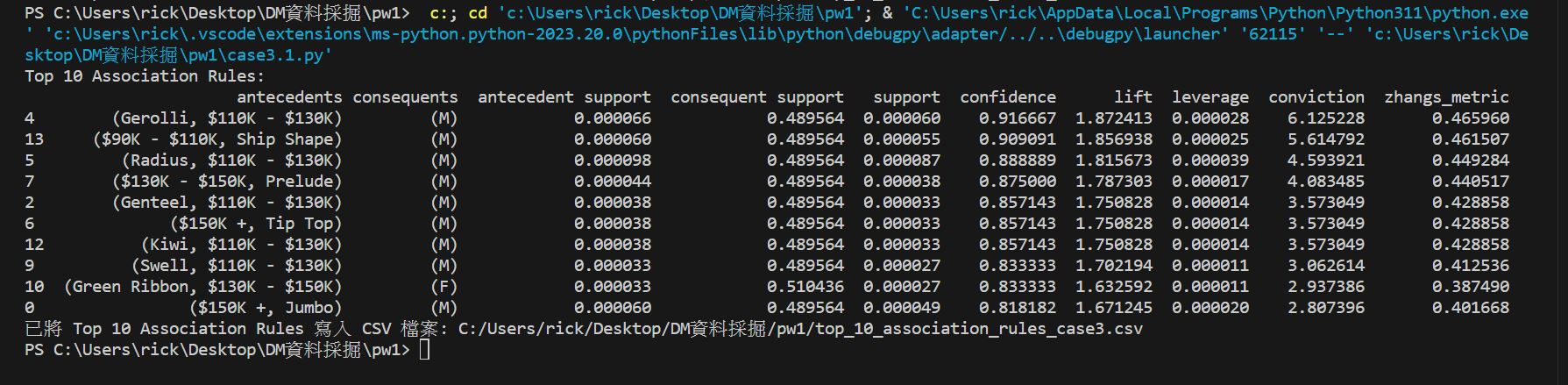
Top 10 Association Rules



1. 請探勘 Foodmart Data 中，顧客背景資料與其交易資料之間的關係(Quantitative Association Rules)。例如 80% 女性顧客常買保養品。請自行設定 Minimum Support、Minimum Confidence， 找出 10 條你覺得有意義的 Rules。請說明你的作法及相關參數設定。

selected\_columns = ['gender', 'yearly\_income','product\_brand'] (選擇一)

min\_support = 0.000025、min\_confidence = 0.8



1.在收入$110K-$130K的男性有91.6%的機率會購買Gerolli(漢堡)

Gerolli Seasoned Hamburger 1.34元

Gerolli Extra Lean Hamburger 3.38元

2.在收入$90K-$110K的男性有90.9%的機率會購買Ship Shape(漢堡)

Ship Shape Seasoned Hamburger 1.75元

Ship Shape Extra Lean Hamburger 2.97元

3.在收入$110K-$130K的男性有81.8%的機率會購買Radius(穀物)

Radius Oatmeal麥片2.53元

Radius Corn Puffs玉米澱粉1.38元

Radius Wheat Puffs小麥粉2.64元

Radius Grits粗玉米澱粉0.86元

4.在收入$130K-$150K的男性有87.5%的機率會購買Prelude(太陽眼鏡)

Prelude Rosy Sunglasses 2.3元

5.在收入$110K-$130K的男性有85.7%的機率會購買Genteel(漢堡)

Genteel Seasoned Hamburger 1.15元

Genteel Extra Lean Hamburger 1.9元

6.在收入$150K以上的男性有85.7%的機率會購買Tip Top(海鮮)

Tip Top Lox燻鮭魚0.37元

Tip Top Scallops扇貝3.6元

7.在收入$110K-$130K的男性有85.7%的機率會購買Kiwi(海鮮)

Kiwi Lox燻鮭魚2.25元

Kiwi Scallops扇貝2.38元

8.在收入$110K-$130K的男性有83.3%的機率會購買Swell(水果罐頭)

Swell Canned Mixed Fruit 2.34元

Swell Canned Peaches 1.63元

9.在收入$130K-$150K的女性有83.3%的機率會購買Green Ribbon(水果罐頭)

Green Ribbon Canned Mixed Fruit 1.31元

Green Ribbon Canned Peaches 3.11元

10.在收入$150K以上的男性有81.8%的機率會購買Jumbo(蛋)

Jumbo Small Brown Eggs 1.42元

Jumbo Large Brown Eggs 2.68元

Jumbo Small Eggs 0.75元

Jumbo Large Eggs 3.84元

Jumbo Egg Substitute 2.97元

由上述可推測：

收入$90K-$110K的男性很大的機率會購買Ship Shape漢堡

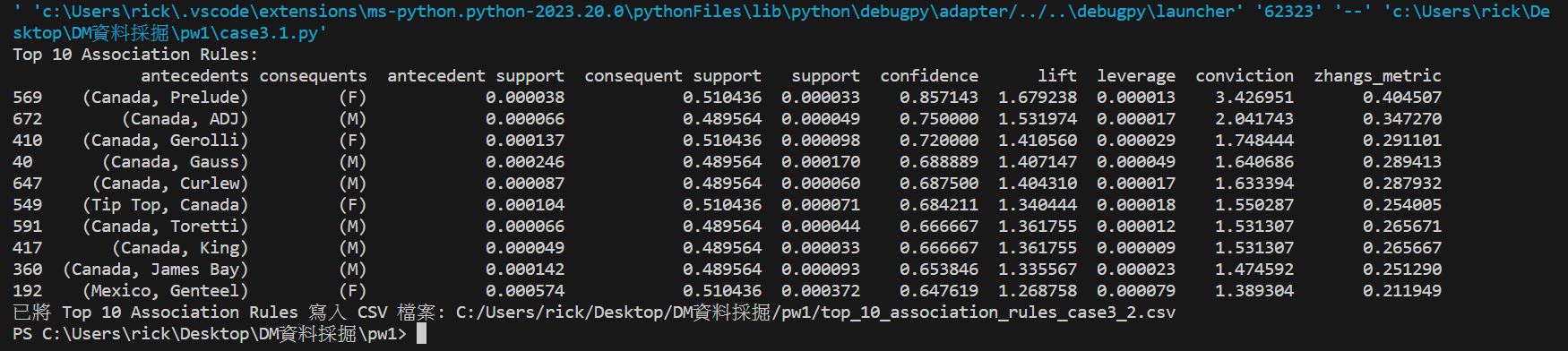
收入提升到$110K-$130K的男性，購買健康的穀物、海鮮、水果罐頭的機率較高

當收入$150K以上的男性購買Prelude太陽眼鏡、Jumbo蛋的機率較高

可以推測收入提升後，生活品質的提升

再來跟改selected\_columns = ['gender','customer\_country','product\_brand'] (選擇二)

min\_support = 0.000025、min\_confidence = 0.8

11.加拿大籍的女性有85.7%的機率會購買Prelude(太陽眼鏡)

12.加拿大籍的男性有75%的機率會購買ADJ(太陽眼鏡)

可以推測同為加拿大籍，女性偏好購買Prelude品牌的太陽眼鏡、男性則是ADJ品牌

13.加拿大籍的女性有72%的機率會購買Gerolli(漢堡)

14.加拿大籍的男性有68.8%的機率會購買Gauss(雜誌)

15.加拿大籍的男性有68.7%的機率會購買Curlew(海鮮)

16.加拿大籍的女性有68.4%的機率會購買Tip Top(海鮮)

可以推測同為加拿大籍，男性偏好購買Curlew的海鮮、女性則是Tip Top的海鮮

17.加拿大籍的男性有66.6%的機率會購買Toretti(太陽眼鏡)

18.加拿大籍的男性有66.6%的機率會購買King(太陽眼鏡)

可以推測加拿大籍的男性很大的機率喜歡購買太陽眼鏡，如ADJ、Toretti、King品牌

19.加拿大籍的男性有65.3%的機率會購買James Bay(五金行：眼鏡螺絲起子、地圖)

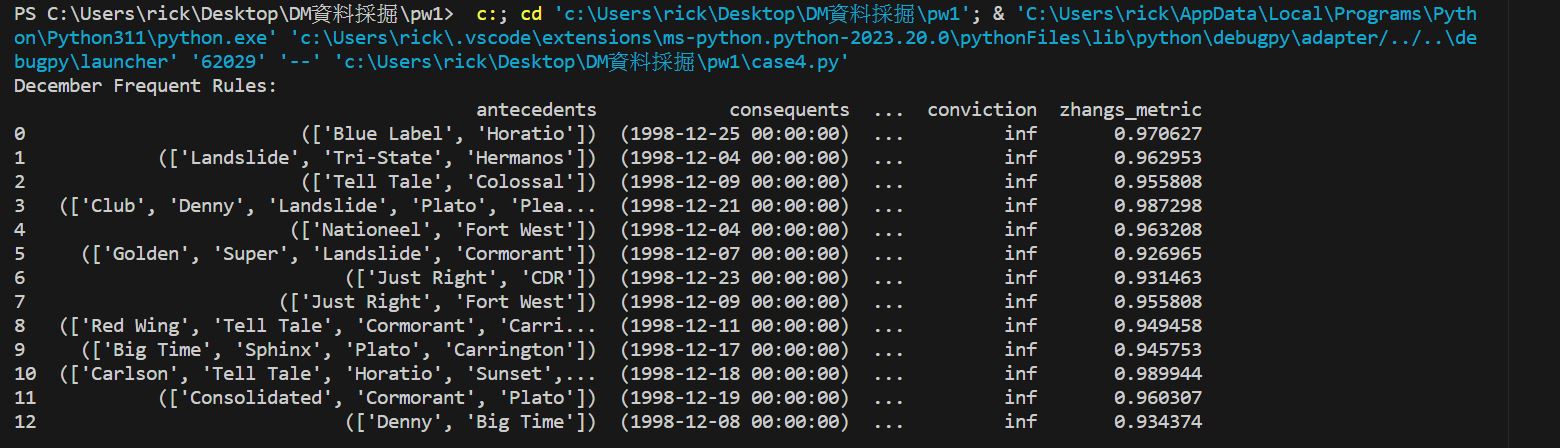
20.墨西哥籍的男性有64.7%的機率會購買Genteel(漢堡)

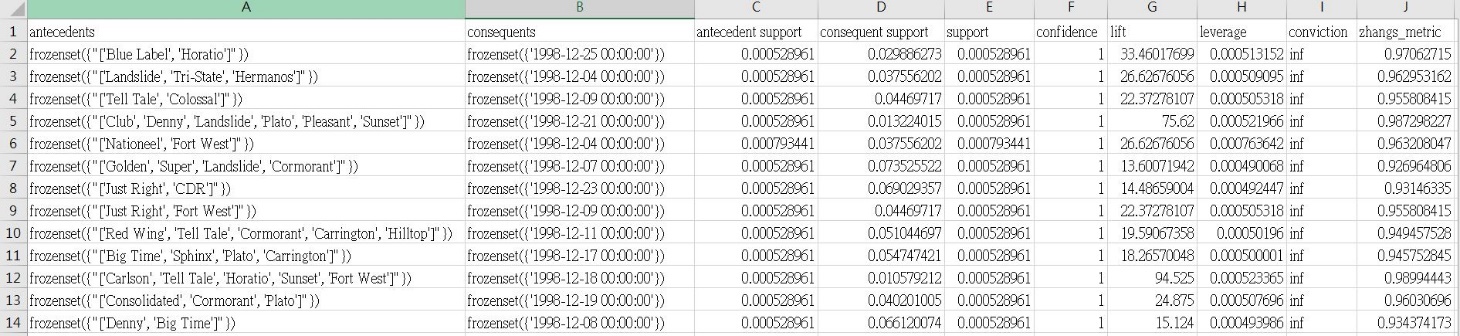
那我們結合(選擇一)'gender', 'yearly\_income','product\_brand'以及(選擇二)'customer\_country'

可以推測有55.44%的機率是墨西哥籍男性收入是在$110K-130K並且喜歡購買Genteel漢堡

1. 在美國由於聖誕節，12月是購物的旺季。請探勘分析比較 12 月與 1 ~ 11

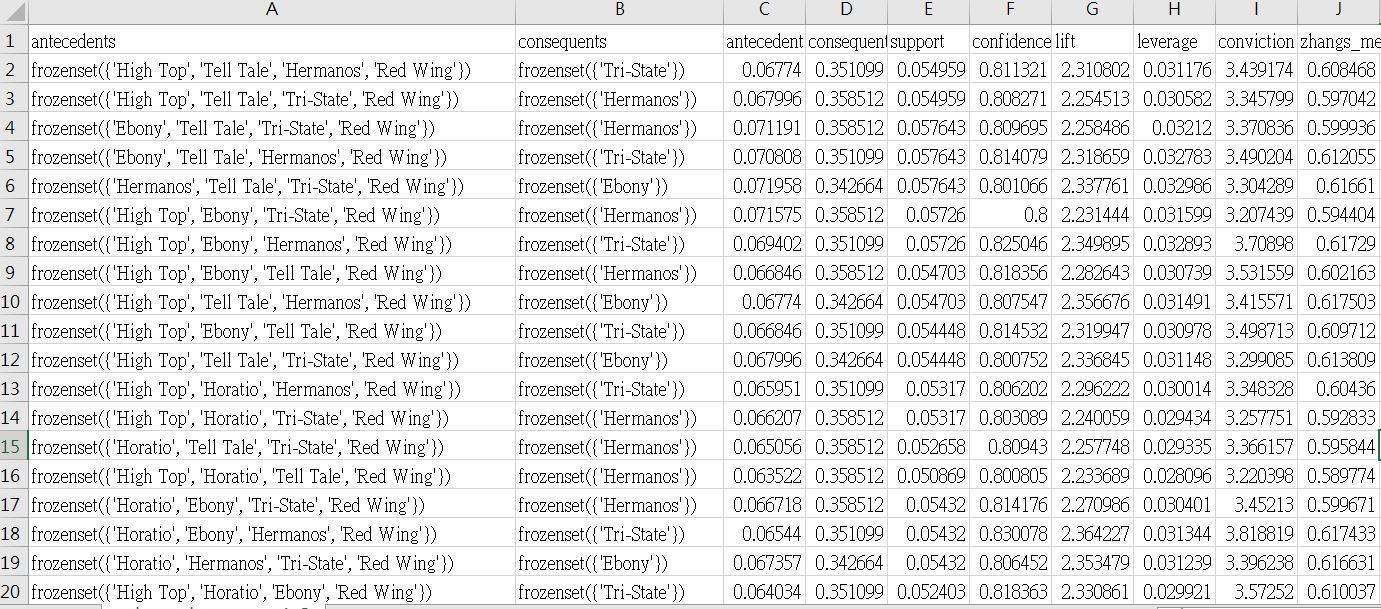
月的顧客購物行為。 有哪些相似的地方，有哪些差異的地方?







由12月份的資料可推測，Plato(烘焙坊)、Fort West(零食)、Cormorant(餐具)為當月購買機率最大的品牌，其中Plato它家的Apple Butter為必買商品。



1.購買High Top、Tell Tale、Hermanos、Reg Wing的產品，有很大的機率會一同購買Tri-State的商品

2.購買High Top、Tell Tale、Tri-State、Reg Wing的產品，有很大的機率會一同購買Hermanos的商品

3.購買Ebony、Tell Tale、Tri-State、Reg Wing的產品，有很大的機率會一同購買Hermanos的商品

4.購買Ebony、Tell Tale、Hermanos、Reg Wing的產品，有很大的機率會一同購買Tri-State的商品

5. 購買High Top、Horatio、Hermanos、Reg Wing的產品，有很大的機率會一同購買Ebony的商品

可以推測1月到11月最常購買的商品是這五個品牌High Top、Tell Tale、Reg Wing、Tri-State、Hermanos，再來是Ebony品牌，接下來是Horatio品牌。而這幾個品牌分別是High Top食材店、Tell Tale食材店、Reg Wing器具行、Tri-State食材店、Hermanos食材店、Ebony食材店、Horatio零食店。

可以得知1月到11月平時主要都是去購買日常食材、器具，再來才是零食。而12月購物的旺季，反而是買些果醬、零食、餐具為主。

相似的地方是，美國人真的很喜歡吃零食，雖然日常所需為優先，但很大的機率是會購買零食。而到了12月開始買糖、巧克力、果醬、奶油等等，或許是為了聖誕節的到來而準備，做些甜點、塔派，然後放縱吃零食，還需要增添餐具去擺放、使用食物。