**作業3: 誰和誰一國**

接下來,我們要開始進行一個「相對」而言模糊的分析,也就是集群分析,讓我們開始吧!

1. 開始之前,讓我們了解一下,集群分析是一項　監督式　非監督式　學習 （請圈選）　(10%)

2.請開啟CC GENERAL.csv. 這是一個銀行的客戶資料集。請進行必要之前處理前,並請存為ETL,也請稍加說明您做了那些處理(10%)

確認欄位的格式、事先選擇需要用到的欄位

3. 接下來,我們要使用傳說中的RFM模型。但等等,目前的資料裡並沒有這3個欄位呀,那怎麼辦呢? 請依欄位說明,合乎邏輯地說明您是如產生這3個欄位唷,(20%)

找出三個用來R:最近購買時間 Ｆ：最近購買頻率 Ｍ最近的購買金額 的欄位

R:

PURCHASES\_FREQUENCY代表購買的頻率

根據PURCHASES\_FREQUENCY 欄位進行平均處理，因為在所有欄位中並未包含日期的欄位，為了找出最近一次購買的日期，透過PURCHASES\_FREQUENCY可知，此欄位值介於0~1之間，若值越接近1，也就表示該客戶購買頻率高，近期購買的機率也較大。

F:

PURCHASES\_TRX代表購買交易的數量

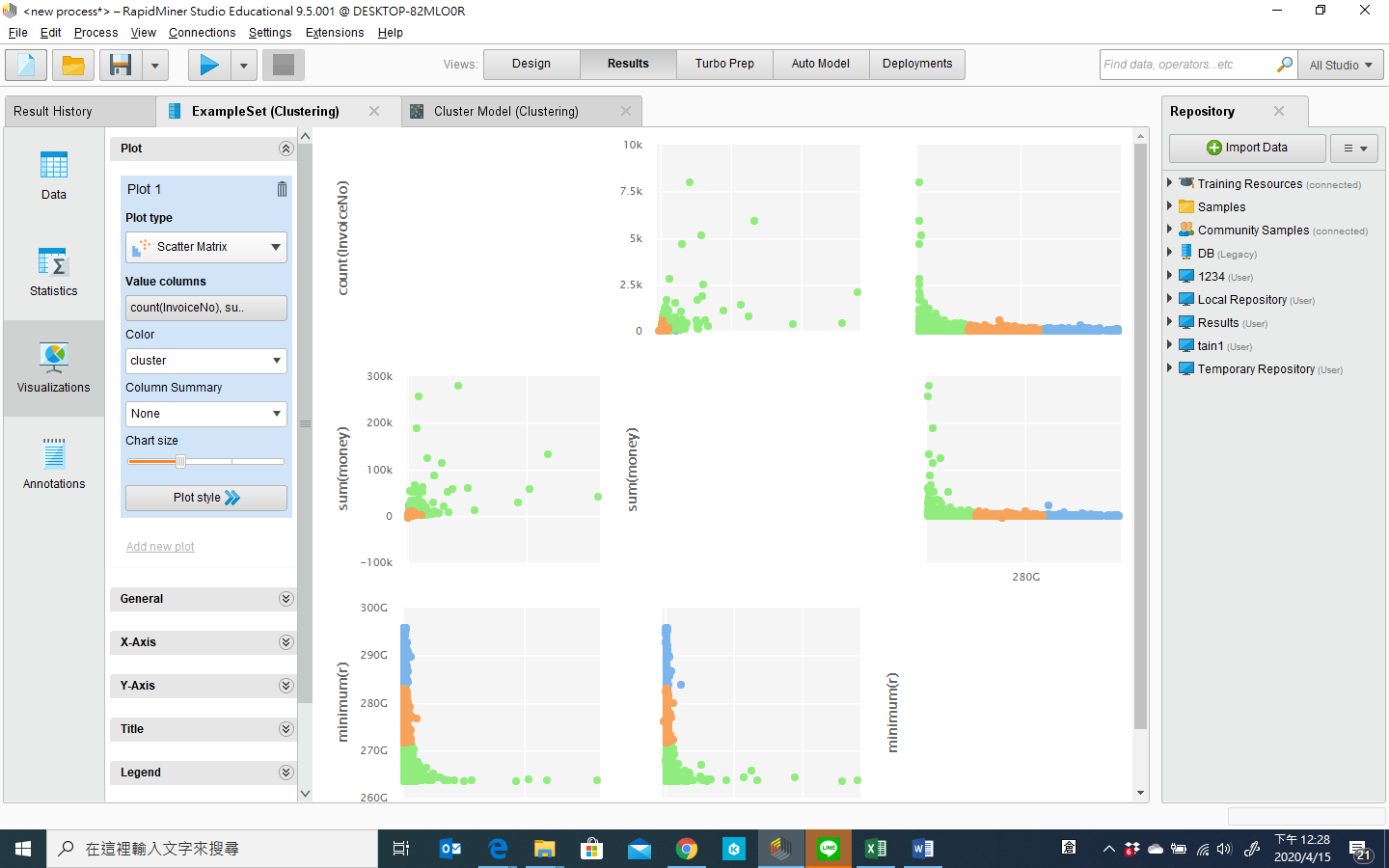
計算購買頻率可根據PURCHASES\_TRX欄位進行加總，因為購買交易的數量越多，也就表示該客戶近期的購買頻率較高。

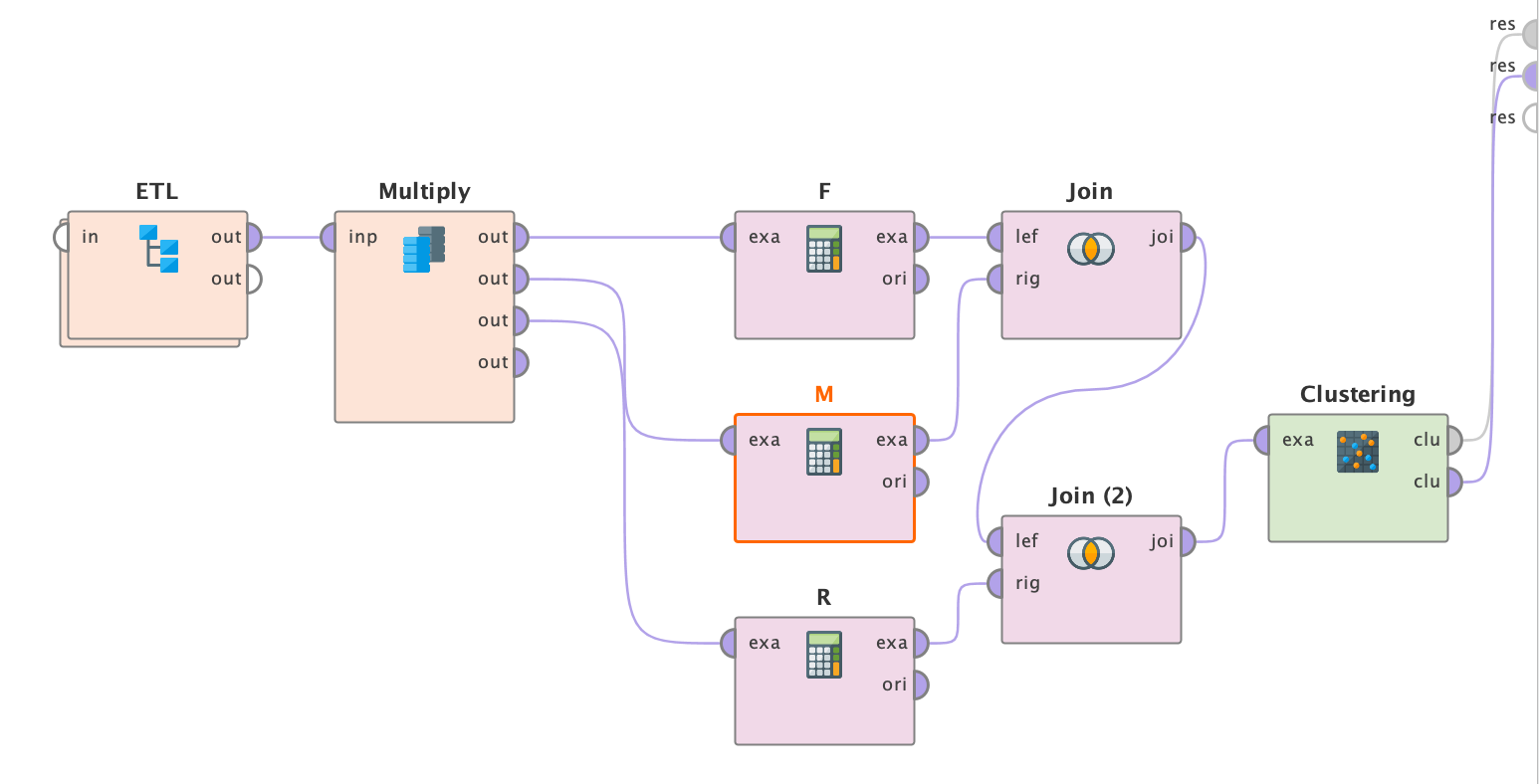
M:

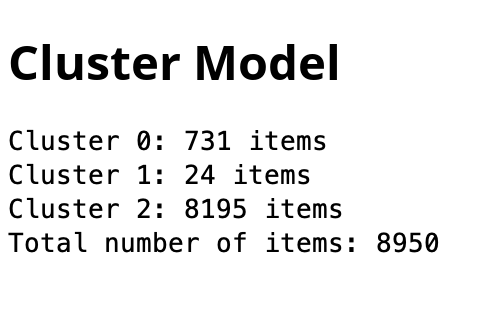
PURCHASES 代表從賬戶購買的金額

計算購買金額可根據PURCHASES欄位進行加總，可得知該客戶的購買金額多少

4. 最後,請對上述整併後的資料進行Cluster分析,並請說明相關結果,請將結果存為學號\_1.rmp(20%)。並請匯製一張您覺得最能說明相關分析結果之分群說明圖,加以說明(10%)。





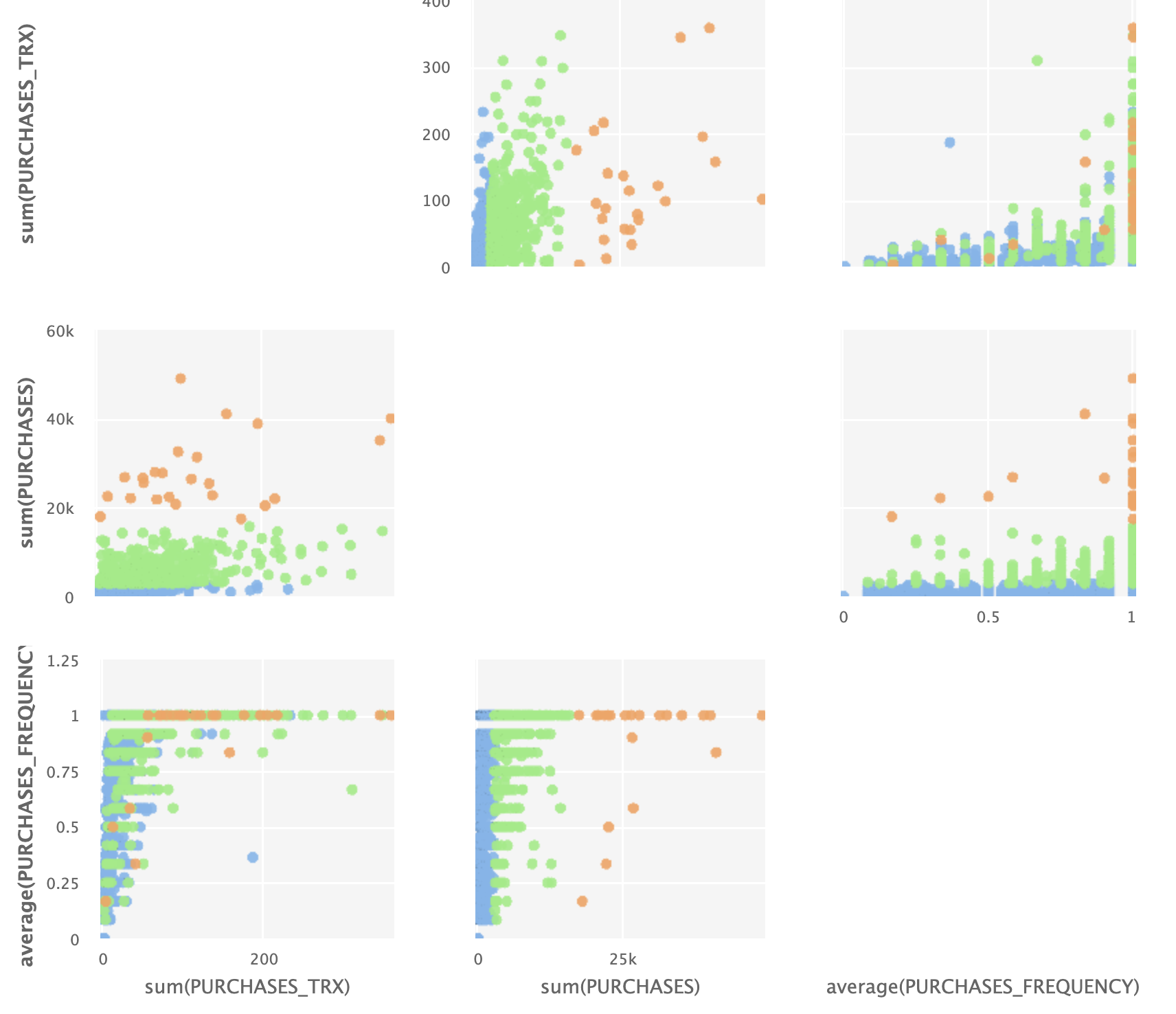


最終分群結果分為三群：

cluster0(綠色)：可代表中產階級

cluster1(橘色)：可代表富貴階級

cluster2(藍色)：可代表一般百姓



5. 最後,請對上述整併後的資料進行Cluster分析,並請說明相關結果,請說明您給予銀行的建議及策略。(30%)

從結果可知，富貴階級的購買金額皆最大，此外，購買頻率並未因為購買金額較大而減少。

中產階級的購買金額並未向富貴階級那麼高，但購買頻率相較一般百姓較高。

一般百姓可明顯發現當購買金額越大時，交易次數變得很少，反而金額越小次數越多。購買頻率方面，在三個階級上，一般百姓的使用也較其他兩者低。

給銀行的建議：

因為一般百姓所賺的錢本來就相對較少，因此，如果想要增進一般百姓對帳戶的使用率的話，可以用提供高回饋的方式，讓一般百姓覺得使用此帳戶購買東西跟其他家銀行比起來更能花更少的錢買到想要的東西。

中高階級的客戶，購買頻率本來就較高，而金額也相對較大，因此，可提供貸款優惠，促進他們透過跟銀行貸款來消費更高額的商品。