

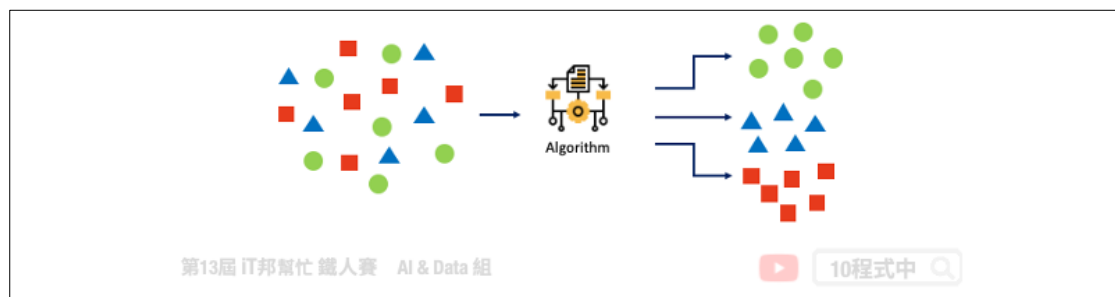
作業 5: 誰和誰一國

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) (5%)

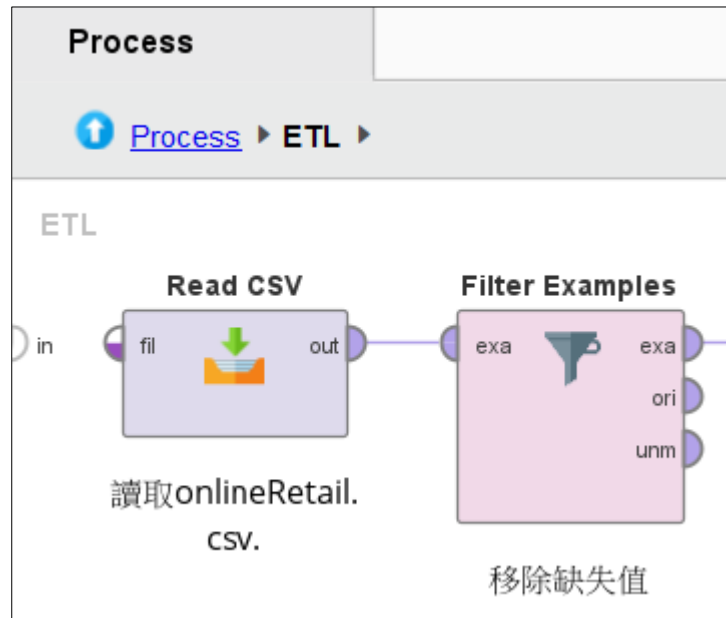
Ans: 非監督式

集群分析 (Ex. K-means) 根據物以類聚的原理,目標是根據特徵把資料樣本分為 K 群。在訓練模型時僅須對機器提供輸入的特徵,並利用分群演算法自動從這些特徵中找出鄰近的集群中心作為該類別。



2.請開啟 onlineRetail.csv. 這是一個交易記錄的資料集。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%)



{ ETL }

參考這篇資料 <https://www.kaggle.com/code/astakalaswapnil/online-retail-clustering>，我將有缺失值的資料全部移除。

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? 請產生這 3 個欄位唷,說明如下:

→依顧客的角度(CustomerID)

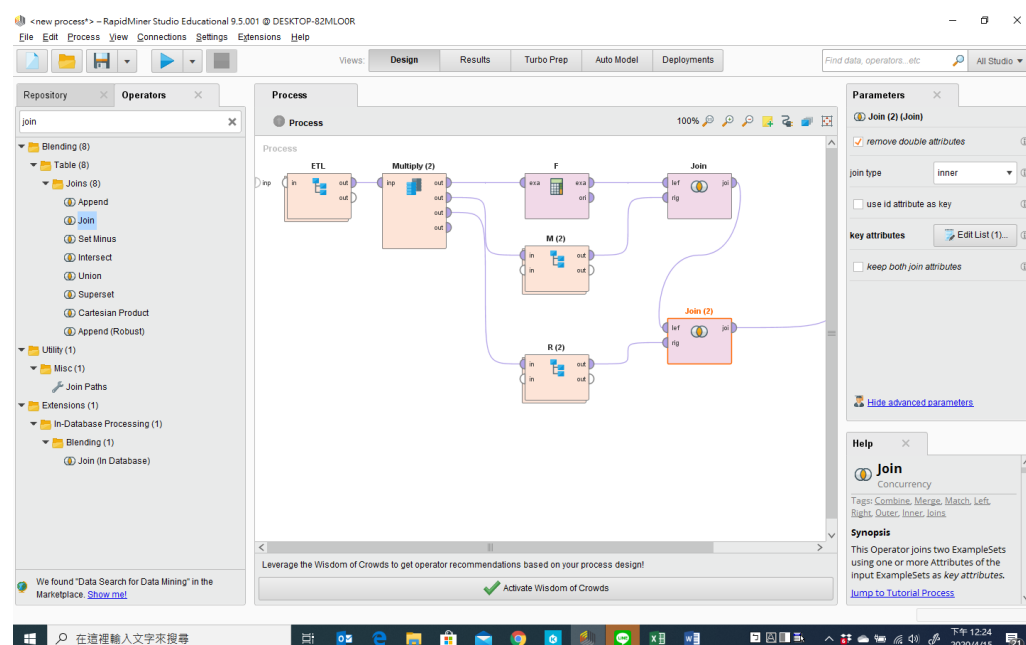
R: 距離今天的消費日期 (在這份作業我是用今天日期減去 Invoice Date)

F: 總共消費的次數

M: 總共消費的金額

第 3 題,我們是從消費者的角度出發唷! 請分別將 R/F/M 之資料產制方式存入

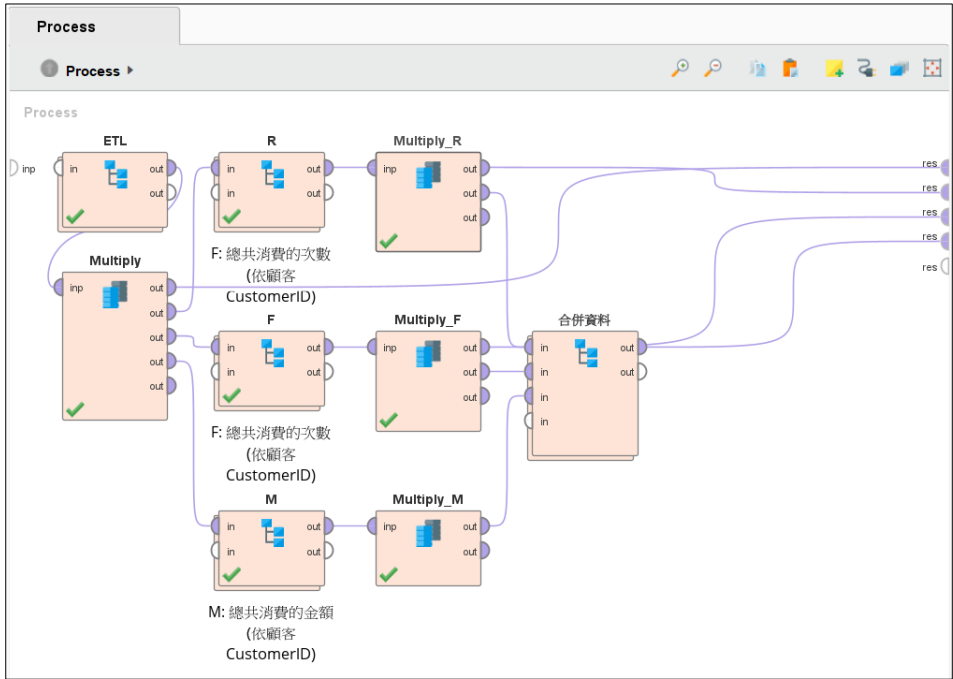
Subprocess. (20%) 如下圖所示



※ R 不建議直接在 Excel 做完，因為計算方式或許不同。

(每一個軟體都有強項)

Ans :



[↑ 圖片更正] R 為距離今天的消費日期

CustomerID	Recency	Frequency	Money
12346	4494	2	0
12347	4170	182	4310
12348	4243	31	1797.240
12349	4187	73	1757.550
12350	4478	17	334.400
12352	4204	95	1545.410
12353	4372	4	89
12354	4400	58	1079.400
12355	4382	13	459.400
12356	4191	59	2811.430
12357	4201	131	6207.670
12358	4170	19	1168.060
12359	4176	254	6245.530
12360	4220	129	2662.060
12361	4455	10	189.900
12362	4171	274	5154.580
12363	4278	23	552.000
12364	4176	85	1313.100

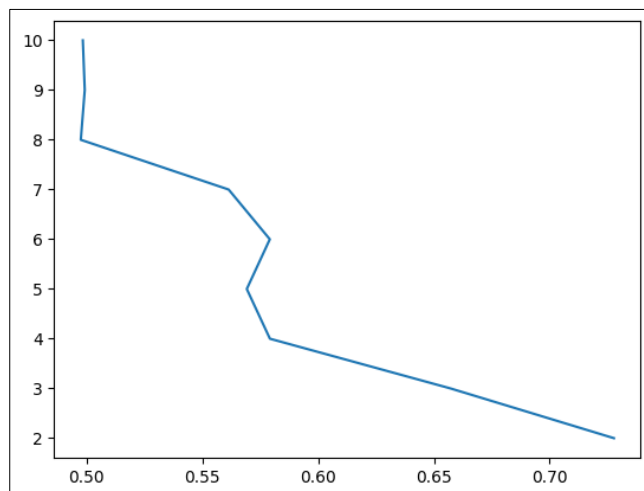
4. 最後,請對上述整併後的資料進行 Cluster 分析,並請說明相關結果,請將結果存為學號_1.rmp(15%)。並請匯製一張您覺得最能說明相關分析結果之分群說明圖,加以說明(5%)。



↑ 此為示意圖，答案不一定要完全一樣。

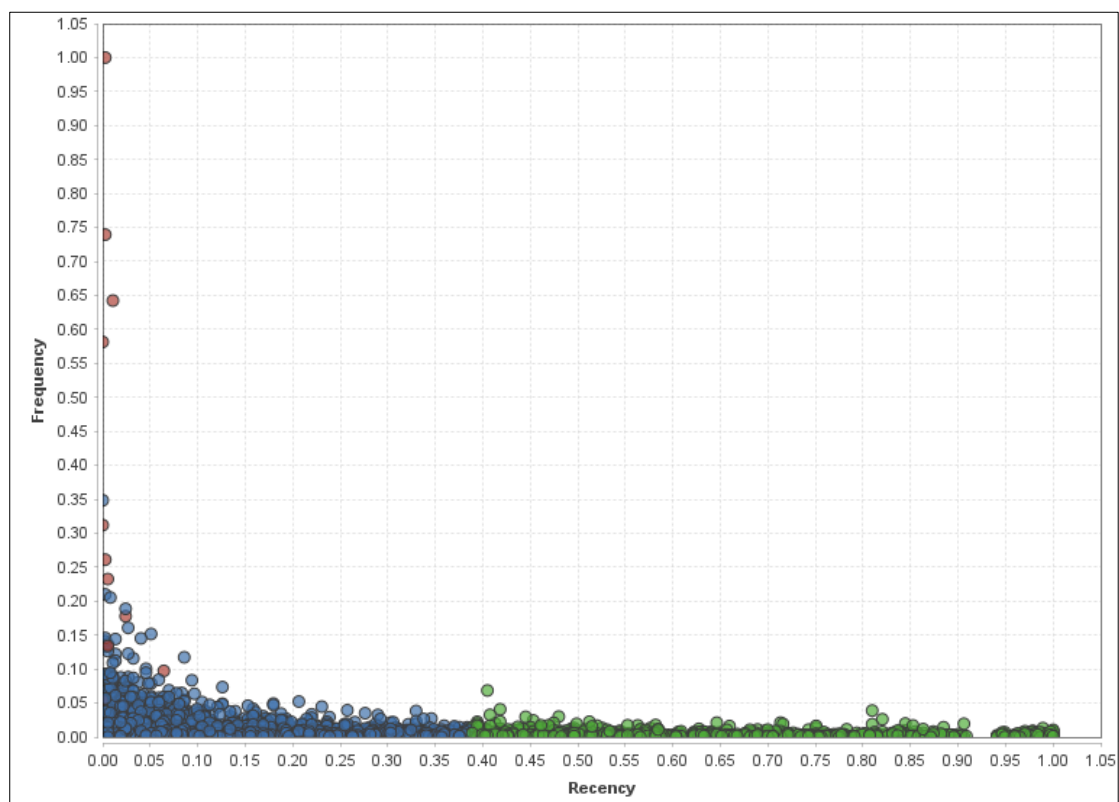
Ans：

根據下方折線圖 (X 軸為 Silhouette Score ; Y 軸為 K 值) · Silhouette Score 算出來的值越大，代表這個 k 的選擇越好，所以選擇分為 2 群或 3 群較合理，Silhouette Score 分數分別為 0.727727 及 0.657201。



k值	Silhouette Score分數
2	0.727727
3	0.657201
4	0.579198
5	0.569207
6	0.579151
7	0.561374
8	0.497542
9	0.499250
10	0.498418

※ 因為分為兩類，說明上感覺有點太過簡化，所以我選擇分為三類。



Cluster	R	F	客戶類型
Cluster0 藍色	近	低	發展客戶
Cluster1 綠色	遠	低	挽留客戶 或 流失客戶
Cluster2 紅色	近	高	重要價值客戶 或 一般價值客戶

- Recency：最近一次消費。顧客上次消費時間愈近，對公司或品牌肯定比較有印象，加上如果顧客的消費或服務體驗良好，很可能會再次進行消費
- Frequency：消費頻率，特定時間內的消費頻率越高，反映顧客對品牌的忠誠度越高，顧客價值越高。

➤ Cluster0 藍色

- A. 行為特徵：近期購買過、購買頻率低。
- B. 可能是**新客戶**，或是新的批發商/企業採購者。
- C. 營銷策略：提供會員積分服務，介紹新產品/功能，給與一定程度的優惠來**提高留存率**。

➤ Cluster1 綠色

- A. 行為特徵：近期沒有購買 且 購買頻率低。
- B. 可能是**即將流失或已流失**的客戶。
- C. 營銷策略：若過去消費紀錄客單價高，則通過短信、郵件、電話等介紹最新產品/功能/升級服務促銷折扣等，避免流失；若消費較低，則可以暫時放棄此部分用戶。

➤ Cluster2 紅色

- A. 行為特徵：近期購買過、購買頻率高。
- B. 對店家來說這部分的客戶**最有價值**。
- C. 營銷策略：若是消費高的客群，應鼓勵升級為 VIP，提供個性化服務，店家應投入更多的資源在這部份客戶，維持這些會員的忠誠度；
若是消費較低的客人，則列為潛力股，提供社群服務，介紹新產品/功能促進消費。

※ 藉由分析各類型顧客的特徵及行為來制定後續的經營策略，例如：提早驚覺即將流失的顧客、放棄沉睡顧客或是將行銷費用重點花在 VIP 顧客上。

5. 接下來,我們要改變分析角度,變成是對**產品面項**使用傳說中的 RFM 模型,但
等等,目前的資料裡還是沒有這 3 個欄位呀,所以請產生這 3 個欄位唷,說明如下:

→依產品的角度(StockCode)

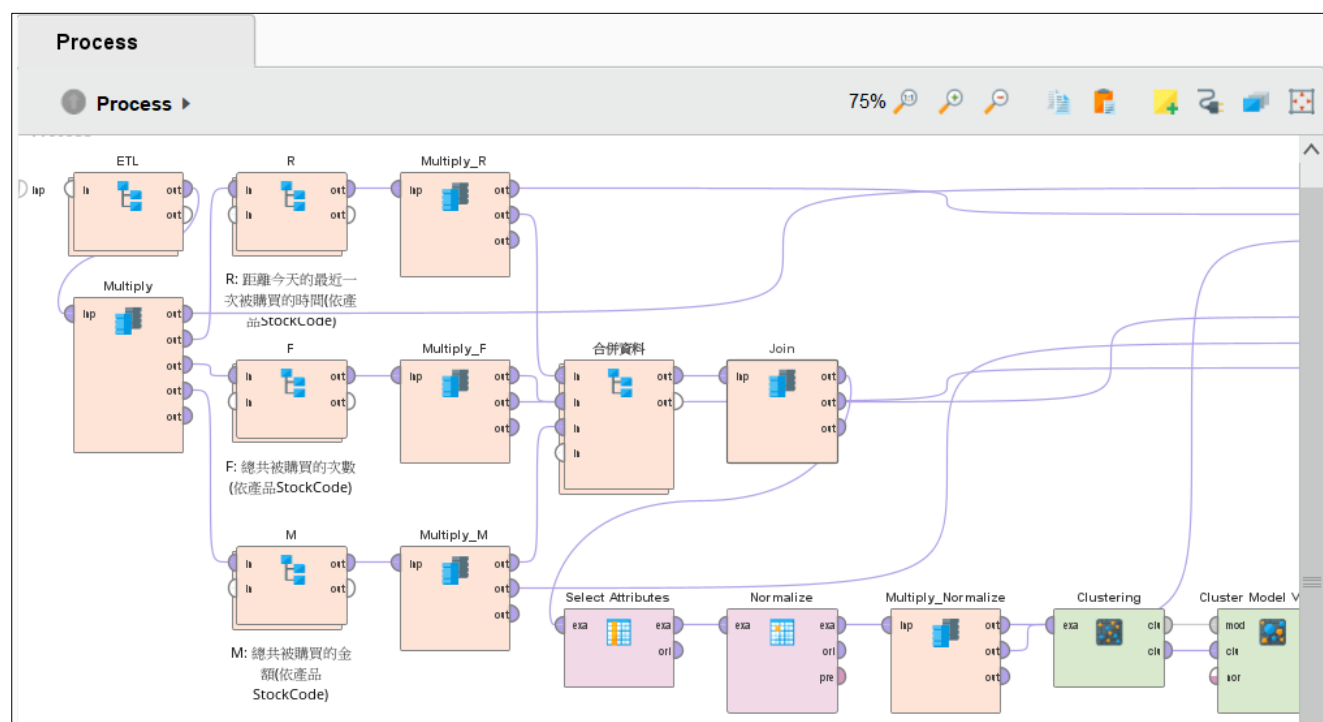
R: 距離今天的最近一次被購買的時間 (在這份作業我是用今天日期減去
Invoice Date)

F: 總共被購買的次數

M: 總共被購買的金額

第 5 題,我們是以**產品的角度**出發唷! 請分別將 R/F/M 之資料產制方式存入

Subprocess. (20%)

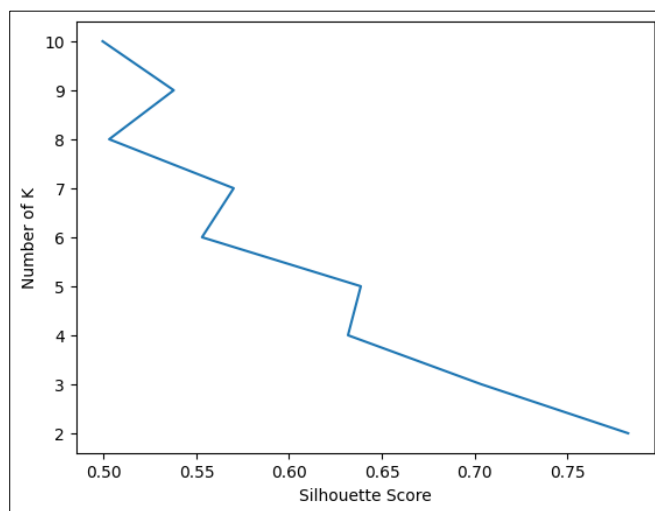


StockCode	Recency	Frequency	Money
10002	4405	49	699.550
10080	4188	21	114.410
10120	4175	30	40.530
10123C	4423	3	3.250
10124A	4203	5	6.720
10124G	4203	4	7.140
10125	4170	64	930.300
10133	4263	125	1139.410
10135	4172	122	1785.440
11001	4174	69	1715.950
15030	4173	13	41.470
15034	4181	95	588.770
15036	4177	329	15538.760
15039	4171	74	1111.050
15044A	4174	80	1245.330
15044B	4242	49	862.700

6 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_2.rmp(15%)。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%)。

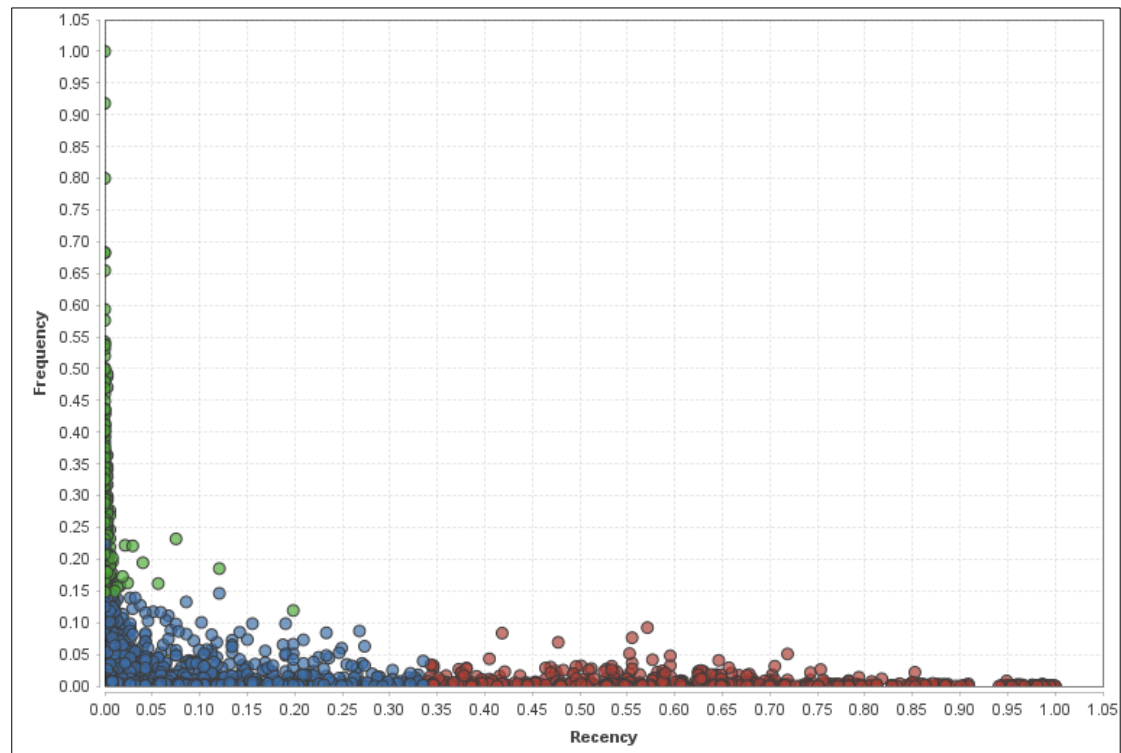
Ans :

依據下方折線圖，選擇分為 2 群或 3 群較合理，Silhouette Score 分數分別為 0.782883 及 0.703647。



K值	Silhouette Score分數
2	0.782883
3	0.703647
4	0.631757
5	0.638673
6	0.552996
7	0.570104
8	0.502924
9	0.537757
10	0.499374

※ 因為分為兩類，說明上感覺有點太過簡化，所以我選擇分為三類。



當面對龐大的產品線時，若在每個產品上都花費相同的行銷成本，可能會是一筆很大的費用。因此透過 Cluster 來有效檢視，一方面可以幫助企業制定出最適合的策略，也可以將投資報酬率極大化，把時間、金錢與心力花在重點上。

Cluster	R	F	推測產品類型
Cluster0 藍色	近	低	新上市產品
Cluster1 綠色	近	高	暢銷產品
Cluster2 紅色	遠	低	不暢銷產品

- Recency：產品最近一次售出 / 產品最近是否有銷售（數字越大，代表該產品最後一次被購買的日期離今天較遠；反之，數字越小，代表該產品最後一次被購買的日期離今天較近）。

- Frequency：產品銷售頻率，特定時間內產品銷售頻率越高，對公司越有價值。

➤ Cluster0 藍色

- 行為特徵：產品近期有售出，但是頻率並不高。
- 推測為**新上市**的產品，近期有客戶購買，但尚未被廣大市場所接受；
也因為剛推出，所以頻率不高。
- 店家應當問卷調查有購買的客人，請顧客給予意見回饋後，討論後續銷售策略，將新產品發展為主要價值商品。

➤ Cluster1 綠色

- 行為特徵：產品經常被銷售，且近期也有顧客購買。
- 這是店家的**主力商品**。
- 店家應優先將資源放在這部分，大力宣傳這類型的商品，為公司帶來更多營收。

➤ Cluster2 紅色

- 行為特徵：產品難以被銷售 (被顧客購買頻率較低)，且近期也沒有人購買。
- 這項產品是店家的**失敗作**，應當改良，或直接放棄 (趕緊賠本銷售、清空存貨)。

[Reference]

- 使用 K-Means 將 RFM 原則應用於客戶聚類

<https://kknews.cc/zh-tw/code/k2qk5gb.html>

- Online Retail Data Set

<https://www.kaggle.com/datasets/vijayuv/onlineretail?datasetId=3466&sortBy=voteCount>

- Online Retail (Clustering)

<https://www.kaggle.com/code/astakalaswapnil/online-retail-clustering>

- Python 数据分析案例-使用 RFM 模型与基于 RFM 的 K-Means 聚类算法
实现电商用户价值分层

<https://blog.csdn.net/maiuida123/article/details/117072743>

- 善用 RFM 分析，標籤化顧客分群！分眾進攻 RFM 模型攻略

<https://www.marketersgo.com/marketing-strategies-tactics/202301/rfm/>

- 【會員經營】解析 RFM 顧客分群法 找出誰是你的 VIP 客戶！

<https://inboundmarketing.com.tw/rfm-customer-analysis/>