

機器學習觀念與應用 作業二
112703003 資訊二 黃柏淵

第一題：請針對 R、F、M 的意義，分別定義計算公式，需說明定義。

- Recency = 當前日期 - 客戶最近一次購買的日期

```
df_trans["trans_days"] = [(date(int(d.split("/")[2]),int(d.split("/")[0]),int(d.split("/")[1]))-date(1998,12,31)).days for d in df_trans["transaction_date"]]
```

為 df_trans DataFrame 新增一個欄位 "trans_days"，計算每筆交易的日期與固定參考日期 1998 年 12 月 31 日 之間的天數差異。

- Frequency = 客戶在特定時間範圍內的購買次數。

```
f_value = df_trans.groupby(["customer_id"])["transaction_date"].nunique()
```

把交易數據按客戶 ID 分組，然後使用 nunique() 計算每個客戶的獨立交易日期數量。

- Monetary = 客戶在特定時間範圍內的總消費金額。

```
# look up the price of the product by its product_id
def lookup_price(product_id, df=df_prod):
    return df_prod[df_prod["product_id"]==product_id]["product_retail_price"].values[0]

0.0s

df_trans["price"] = [lookup_price(pid) for pid in df_trans["product_id"]]
df_trans["amount"] = df_trans["quantity"]*df_trans["price"]
m_value = df_trans.groupby(["customer_id"])['amount'].apply(sum)
```

先為每筆交易添加產品價格，再計算每筆交易的金額 (quantity * price)，按客戶分組，計算每位客戶的總消費金額 (M)。

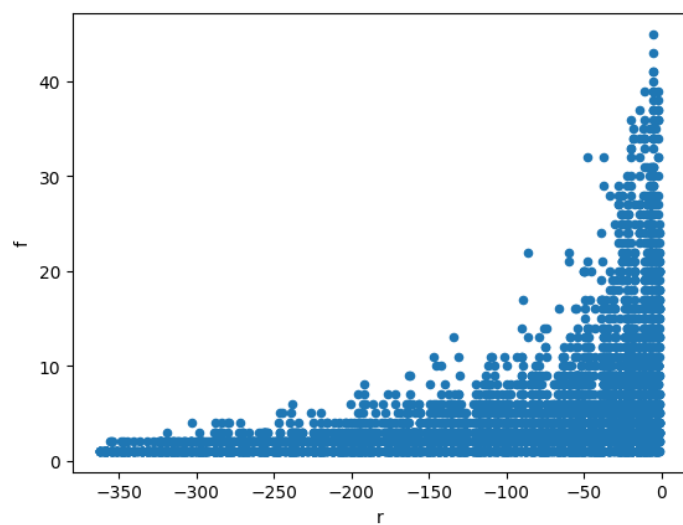
第二題：請針對你所設計的 RFM 計算方式，根據顧客在 1 ~ 10 月的購物行為資料，進行 RFM 值轉換。

- i. 根據所有顧客的 RFM 值分佈，提供各別的最小值和最大值

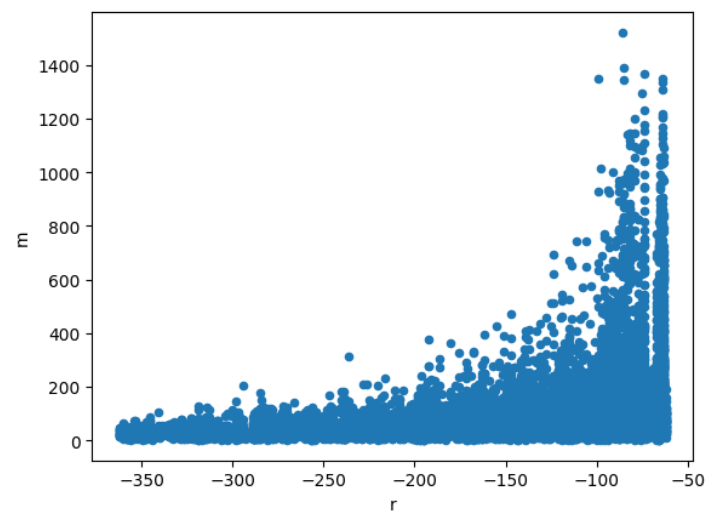
	Metric	Min	Max
0	Recency (R)	-362.00	-62.00
1	Frequency (F)	1.00	37.00
2	Monetary (M)	1.42	1521.42

ii. 分別畫出 RF、RM、FM 二維散布圖。

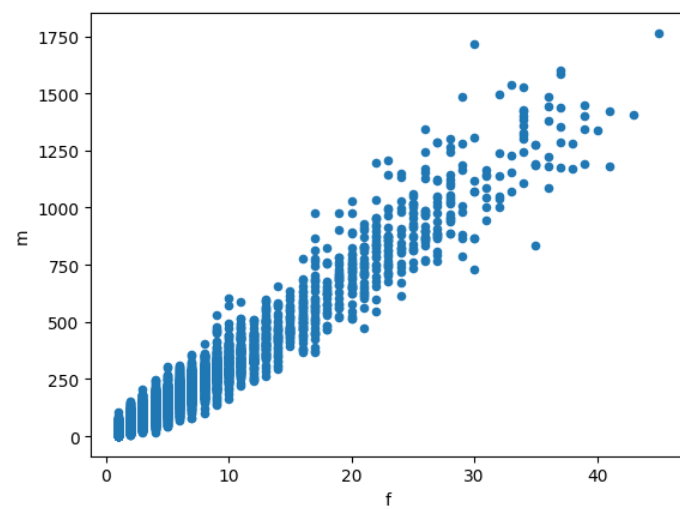
(1)RF



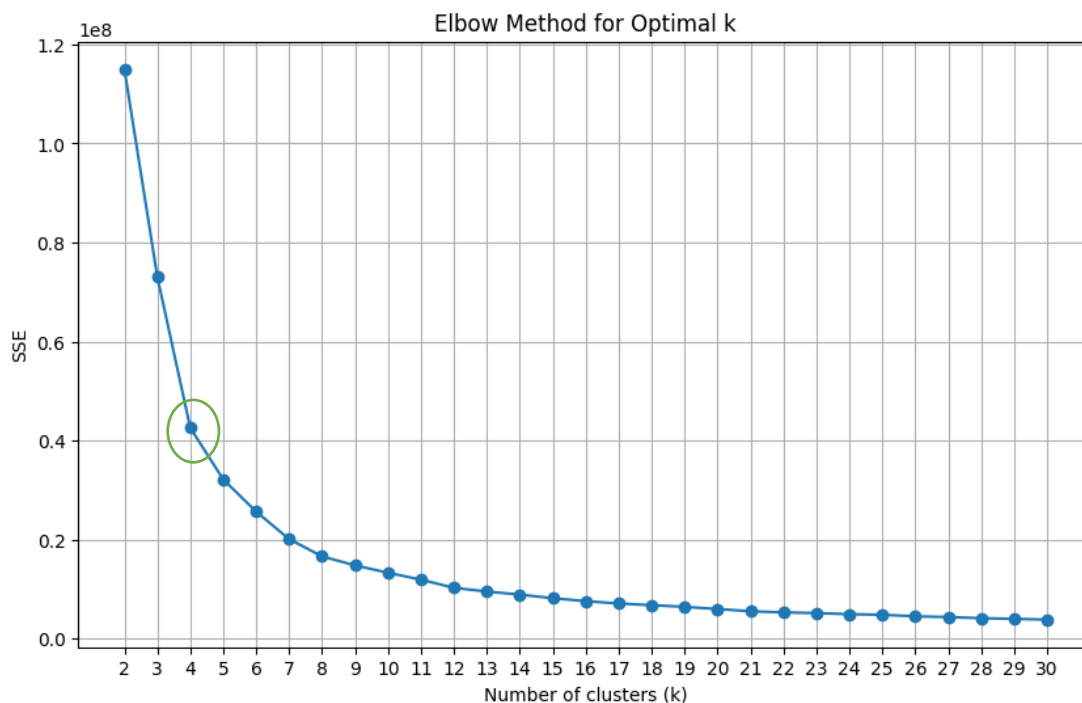
(2)RM



(3)FM



第三題：承第二題，根據轉換後資料，每一位顧客包含三個屬性資料(即 R 值, F 值, M 值)，請根據此資料，透過 K-means 分群演算法，將顧客分群，並觀察不同參數 K 對分群結果 SSE 的影響，請設置 K = 2 ~ 30，利用 Elbow 方法找出 K 應該設置多少？



K 應該設置為 4。

第四題：承第三題，使用 Elbow 方法選定一個 K 值，再執行 K-means 分群演算法，請列出每一群的中心點，例如中心點 C1: R=70、F=25、M=800。K 個中心點的列表請根據中心點的 M 數值降冪排序。

	F	M	R
2	21.387755	802.550510	-75.853741
1	8.887454	295.708958	-94.205720
3	2.803859	79.768588	-116.478605
0	1.470393	40.908039	-263.306490

第五題：承第四題，請你從分群結果挑選出其中一群顧客，為他們設計專屬的行銷活動，以增加客戶留存或增加營收。

i. 此群中心點的 R、F、M 值分別是多少？(群組 2)

	F	M	R
2	21.387755	802.550510	-75.853741

ii. 此群的顧客總人數？

```
Group 2 count: 294
```

iii. 請針對這群顧客的 RFM 值分佈，分別說明最小值和最大值的範圍。

```
R range: [-124, -63]
F range: [10, 37]
M range: [549.7900000000001, 1521.4199999999998]
```

iv. 請說明你的具體行銷活動內容與目的。

群組 2 特性分析：

- Recency (-86 到 -1)：這群顧客的最近一次購買距今約 2 到 4 個月，平均為約 94 天前，屬於中度活躍用戶
- Frequency (14 到 45)：顧客的購買頻率明顯偏高，最低有 10 次、最高達 37 次，表示他們屬於忠誠度高的客群，可能是固定或常回購的顧客。
- Monetary (643.80 到 1763.58)：消費金額區間分布廣，平均值可能接近中高範圍，最低金額也不低，這說明這群人是具備穩定消費能力的顧客，其中部分人甚至有高價值客戶潛力。
- 產品偏好：

```
=== Cluster 2 產品偏好 ===
product_name
Carlson Chocolate Milk      44
High Top New Potatos        44
Super Chunky Peanut Butter  43
Shady Lake Ravioli          43
Carrington Turkey TV Dinner 43
..
Plato Sesame Oil            14
Colony White Bread          14
Pleasant Canned Tuna in Oil 13
Washington Berry Juice      11
CDR Apple Preserves         7
```

行銷方案：

1. 限時優惠券

- 目標：召回 60 天以上未回購的高價值顧客，提高再購率。
- 對象篩選：Cluster 2 中 Recency ≤ -60 的顧客（即兩個月未購）。
- 優惠類型：商品 8 折券、滿 \$500 折 \$100、限時 48 小時使用。

2. 產品組合包推薦

- 目標：提升客單價，滿足顧客一次性大量購買的需求。
- 組合包設計：
 - 懶人料理包：TV Dinner + Peanut Butter + Chocolate Milk
 - 家庭儲糧包：New Potatoes + Sesame Oil + Tuna in Oil
- 售價策略：原價 700 元 → 組合價 599 元（搭配滿額贈，如送果汁）。

3. 會員活動／積分回饋

- 目標：提高活躍顧客黏著度、延長顧客生命週期。
- 對象篩選：Frequency ≥ 20 的 Cluster 2 顧客。
- 活動設計：
 - 單月消費滿三次 → 點數 2 倍送。
 - 累積滿 1000 點可兌換指定商品（例如：超市禮券、熱銷組合）。

註明：部分程式碼以 Grok 及 ChatGPT 輔助使用