

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)。

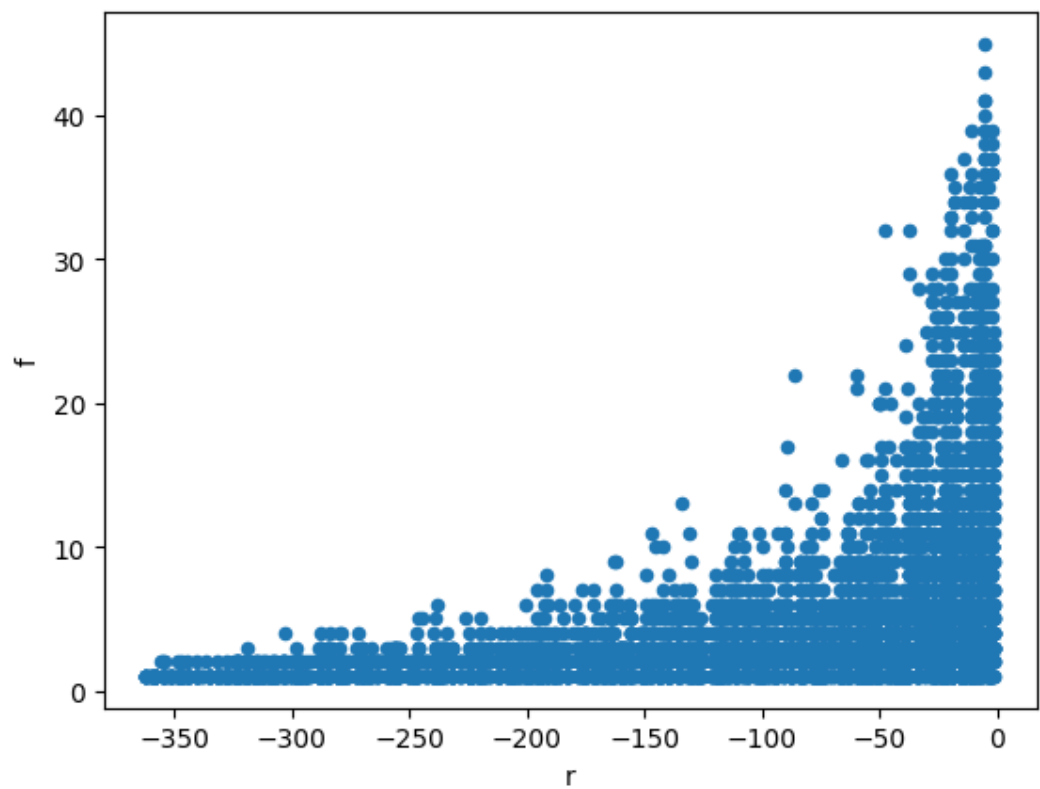
第二題：請針對你所設計的 RFM 計算方式，根據顧客在 1 ~ 10 月的購物行為資料，進行 RFM 值轉換。

- i. 根據所有顧客的 RFM 值分佈，提供各別的最小值和最大值

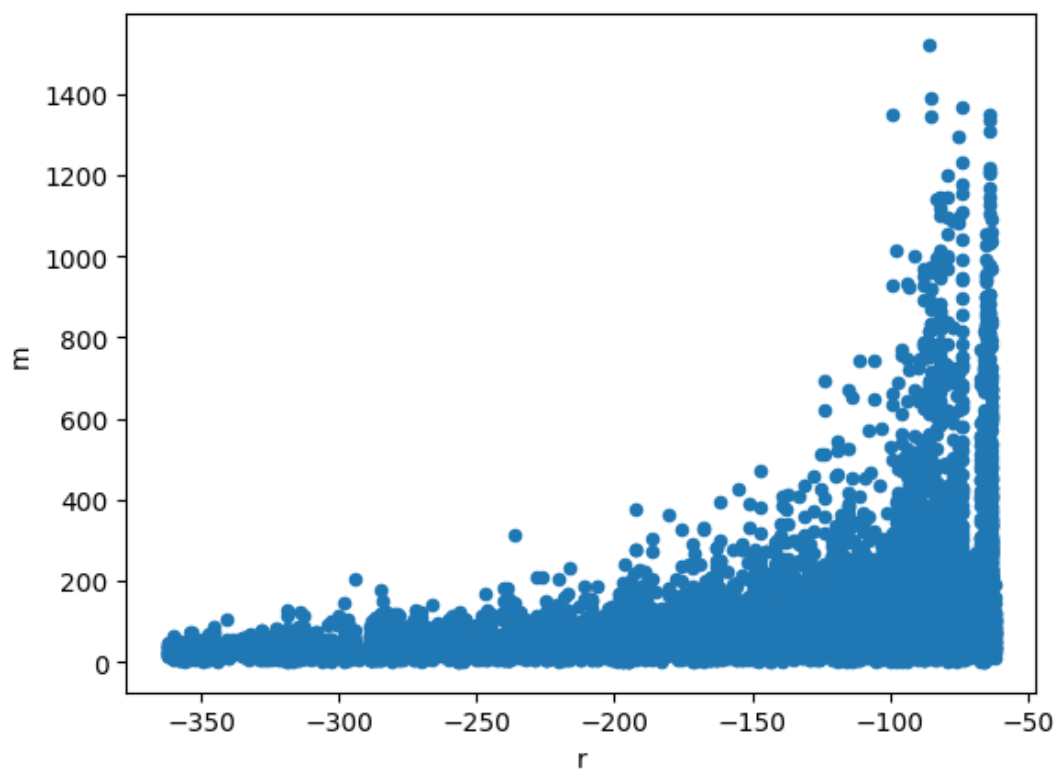
	Metric	Min	Max
0	Recency (R)	-362.00	-62.00
1	Frequency (F)	1.00	37.00
2	Monetary (M)	1.42	1521.42

ii. 分別畫出 RF、RM、FM 二維散布圖。

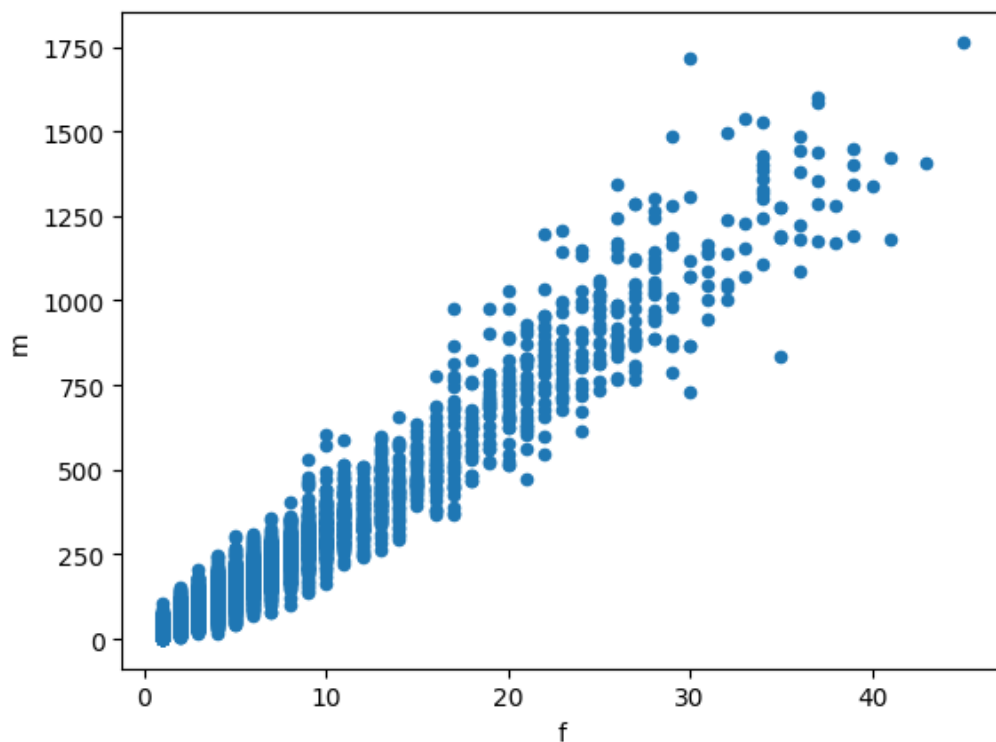
(1)RF



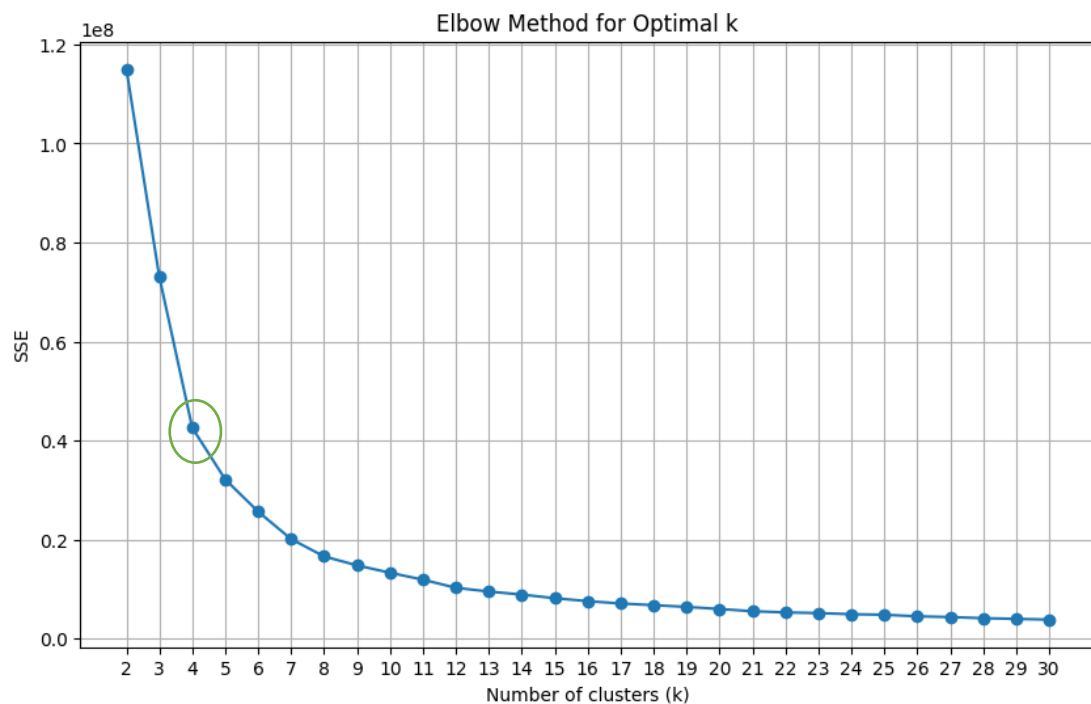
(2)RM



(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 $K = 2 \sim 30$ ，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]
```

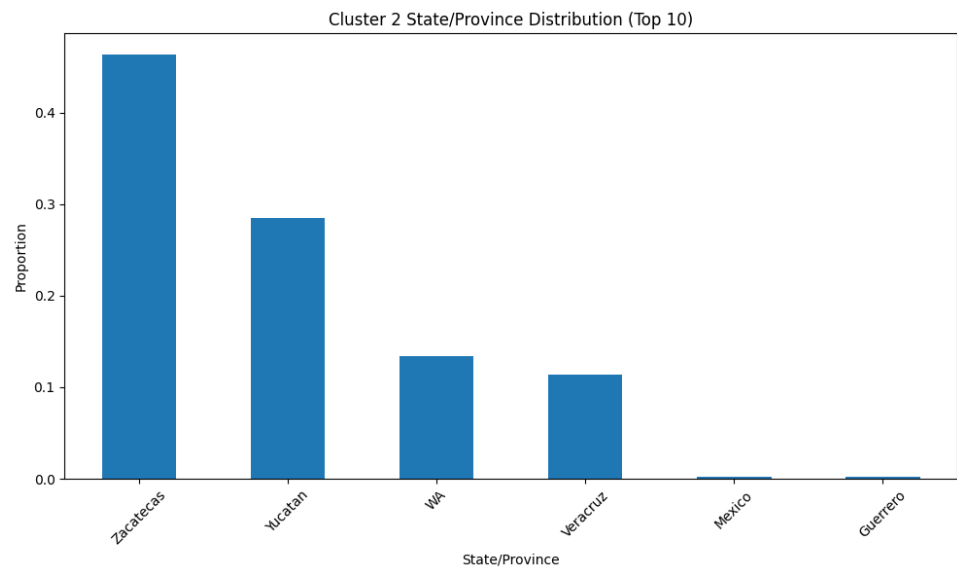
- iv. 請說明你的具體行銷活動內容與目的。

群組 2 特性分析：

- Recency：離上次消費已 2 至 4 個月，正在流失中。
- Frequency：顧客的購買頻率明顯偏高，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：偏中高，代表對平台的消費意願強、忠誠度高。
- 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
```

- 地區分布：



主要集中在墨西哥 Zacatecas、Yucatan，具地域集中性。

- 收入分析：

Cluster 2 Income Distribution (Jan-Oct):

yearly_income	
\$30K - \$50K	0.36
\$10K - \$30K	0.19
\$50K - \$70K	0.16
\$70K - \$90K	0.13
\$130K - \$150K	0.07
\$90K - \$110K	0.05
\$110K - \$130K	0.03
\$150K +	0.02

這裡發現一個有趣的現象，Monetary 分佈廣，平均金額偏中高，代表他們「有花錢」，但收入分布顯示多數集中於 \$10K - \$70K 以下，屬於中低收入，代表他們不是高收入，但願意在特定商品上花錢，雖然可支配所得有限，但會在特定商品上願意花錢。

行銷方案：

1. 多地聯合配送優惠
 - 目標：促進社交分享與交叉下單。
 - 「Zacatecas x Yucatan 跨州團購免運」。
 - 「好友一起買免運／折扣」的活動，提供指定州之間共享優惠碼。
2. 地區性廣告或推播（如 Facebook/Google）
 - 目標：高點擊、高轉換的地區定向投放。
 - 設定地理位置定向的廣告（如只推播給在 Zacatecas 的人），並搭配常見地區用語，增加親切感。

3. 地區限定優惠 × 心理認同
 - 「Zacatecas 限時熱銷 Top3 組合」。
 - 在顧客比例高的州，推出對應地區的「專屬熱門商品組合」，在文案需要強調地區標籤。
4. 分層價值回饋制
 - 目標：針對 Monetary 前 30% 推「會員升等制」。
 - 累積消費滿 \$100 變 VIP → 下次 95 折。
 - 消費滿 \$200 可兌換高價值商品（如家庭料理包）。
5. 生活解決型組合包
 - 幫顧客解決「吃什麼」的日常問題，用「省事 + 高 CP 值」定位取代「低價」。
 - 組合包範例：TV Dinner + Tuna in Oil + Peanut Butter + Milk。
6. 回購階梯獎勵制度
 - 買越多，回饋越多。
 - 第 2 次：滿 \$20 送 \$2 折扣券。
 - 第 3 次：免費送 1 個小商品（果汁等）。
 - 第 5 次：升級 VIP（享受永久 95 折）。
 - 不要一次給大折扣，而是讓顧客感受到「越買越划算」，才有動力繼續買。

註明：部分程式碼以 Grok 及 ChatGPT 輔助使用