

大數據與商業分析 X 91APP

# 期末專題 Proposal

Group 20 | 簡韻宜 李家嫻 曾以函 姜宜瑄 蔡承安 黃敏瑄

解說影片連結: <https://youtu.be/QQKvkRXU42w>

## 背景動機

在高度競爭的市場環境中，有效的顧客分群策略是企業在行銷上成功的關鍵

本專案希望透過深入的客群輪廓分析，去識別出不同消費者群體的特徵，從而針對性地優化產品與提供的服務，提升企業的市場競爭力；在行銷上，也能針對不同客群制定合宜的策略，提高轉化率與顧客忠誠度

## 預期使用方法

特徵萃取 Feature Selection

會員分群 (K-means、PCA)

購物籃分析

## Feature Selection: 從數據中挖掘可使用特徵, 以利後續分群計算

- ① 既有資料: 會員資料
- ② 自定義變數: 從主單資料中挖掘可分析特徵

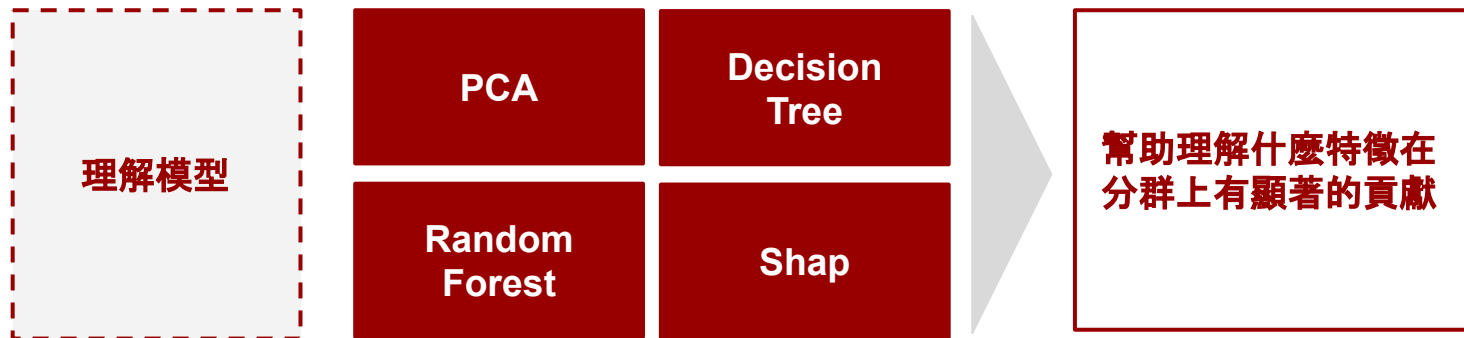


## 會員分析方法

### ① 整體輪廓分析: 使用 EDA、統計分析把整體客群的輪廓描繪出來

目的: 了解整體客群輪廓、關鍵指標有比較基準

### ② 會員分群: 使用 K-means 進行



## 會員分析目的

- ① 相似性質的客戶，利用**非監督式**機器學習方式歸類在一群，達成「**組間差異大、組內差異小**」
- ② 幫助理解相似的人有什麼**消費習慣**，日後能夠制定適合個群的**行銷手段**
- ③ 比較群間的**關鍵指標**：

除了自定義變數，可以額外計算

**LTV**  
(客戶終身價值)

=

單次消費金額

X

消費頻率

X

時間

數位支付族比例

=

數位支付次數

/

總消費次數

**>= 0.5** 則定義為  
偏好數位支付的人

偏好網購比例

=

購買通路

=

「線上」的次數

**>= 0.5** 則定義為  
偏好網購的人

## 購物籃分析: 同一個客群中, 我們可以藉由購物籃分析, 更好的推銷商品

方式: 以「商品」為主體, 用簡單的統計技巧所做出的指標, 找出商品的關聯

三個關鍵指標: Support、Confidence、Lift

目標

Rule:  $X \Rightarrow Y$

$Support = \frac{freq(X, Y)}{N}$

$Confidence = \frac{freq(X, Y)}{freq(X)}$

$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$

分析各客群的購物習慣、  
以客群特質為出發點探索商品間的關聯性

可以對各群客群的人做  
cross selling(交叉銷售)、up selling(追加銷售)

## 預期成果

先分析客群的特徵並做分類，再用購物籃分析得出客群相對應的商品種類購物習慣

應用	客群 個性化行銷	<ul style="list-style-type: none"><li>找出價格敏感度高的客群，根據不同客群的購買偏好，設計有針對性的推廣活動和優惠</li><li>依照分群結果，為每個群體提供精準的商品推薦，增加購買率</li></ul>
	商品組合促銷	<ul style="list-style-type: none"><li>針對客戶群體的購物籃商品組合種類進行促銷或調整行銷計畫(折扣週期或是折扣比率)，提升銷售轉化率</li></ul>
	市場定位與拓展	<ul style="list-style-type: none"><li>根據現有客群特徵，識別潛在新市場和新客群，制定相應的市場拓展策略</li><li>根據客戶分群結果，更加精確地分配營銷預算，將資源投放到潛力最大的群體上，提升投資回報率</li></ul>

大數據與商業分析 X 91APP

# 期末專題報告

Group 20 | 簡韻宜 李家嫻 曾以函 姜宜瑄 蔡承安 黃敏瑄



## Agenda

- ① 分析目標與架構
- ② 資料前處理
- ③ 自定義特徵
- ④ 顧客輪廓
- ⑤ 顧客分群
- ⑥ 結論

## 分析目的

透過分析會員消費紀錄與顧客行為，描繪消費者輪廓，進行更適當的顧客分群，進而提升網路商店的行銷策略，提升市場定位與拓展。

## 分析架構

本次分析先將資料集切成兩個時間段，選取出特徵值後去分析顧客整體輪廓，透過K-means進行的分群，並使用Decision Tree去了解各個特徵在分群上的貢獻，兩階段皆分群完畢後，觀察兩時間段顧客的分群標籤動向，得出顧客行動軌跡。

## 預期目的

### 客群 個性化行銷

- 找出價格敏感度高的客群，根據不同客群的購買偏好，設計有針對性的推廣活動和優惠
- 依照分群結果，為各群體提供精準的商品推薦，增加購買率

### 市場定位與拓展

- 根據現有客群特徵，識別潛在新市場和新客群，制定相應的市場拓展策略
- 根據客戶分群結果，更加精確地分配行銷預算，將資源投放到潛力最大的群體上，提升投資回報率

## Agenda

- ① 分析目標與架構
- ② **資料前處理**
- ③ 自定義特徵
- ④ 顧客輪廓
- ⑤ 顧客分群
- ⑥ 結論

## 主要使用資料: Order data & Member data

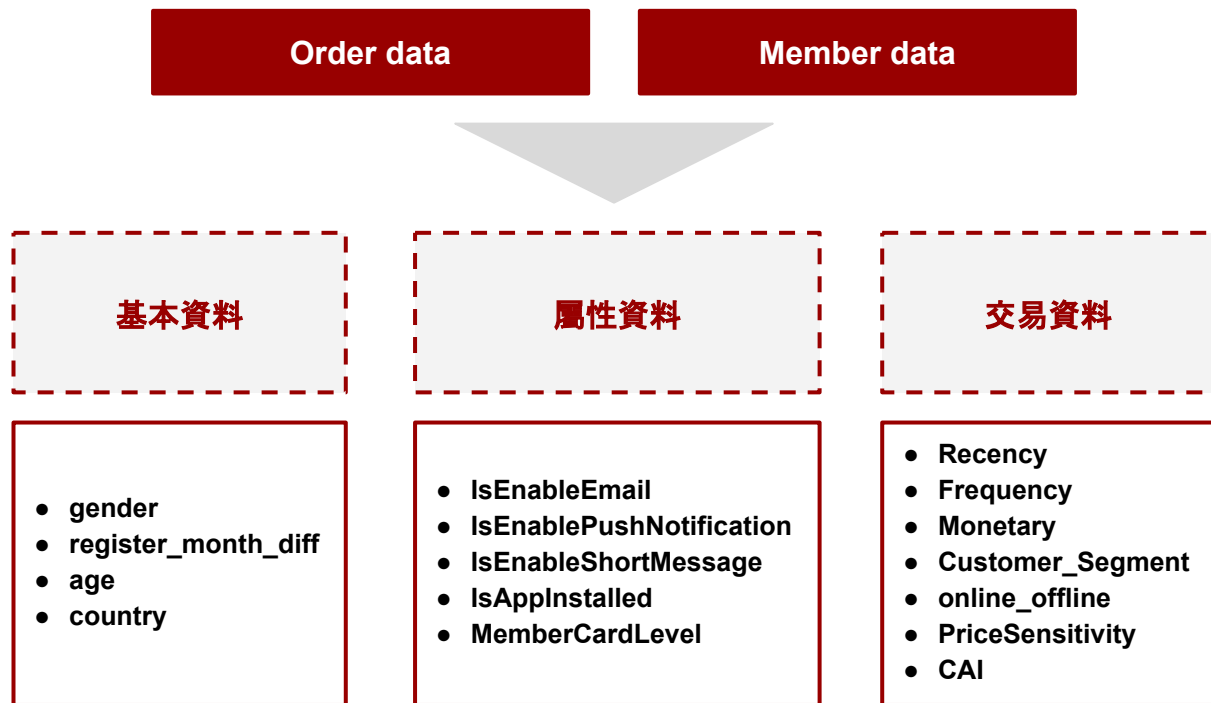
在分群貼標籤之前先進行分割數據集、資料的處理以及新自定義欄位的計算



## Agenda

- ① 分析目標與架構
- ② 資料前處理
- ③ 自定義特徵
- ④ 顧客輪廓
- ⑤ 顧客分群
- ⑥ 結論

## 利用合併過後的資料整理自定義特徵, 利用特徵進行顧客分群



## 利用Order data和Member data的資料整理自定義特徵, 利用特徵進行顧客分群

資料種類	Column Name	意義	備註
基本資料	gender	會員性別	挑選會員資料中可能造成顧客消費行為差異的特徵進行分析
	register_month_diff	註冊會員距離2024-03的月數	
	age	會員年齡	
	country	會員所在國家	
屬性資料	IsEnableEmail IsEnablePushNotification IsEnableShortMessage	是否開啟Email/App/ 簡訊通知	資訊接收率與精準行銷
	IsAppInstalled	是否下載App	客戶互動與體驗
	MemberCardLevel	會員卡等級	識別客戶價值與忠誠度

## 利用Order data和Member data的資料整理自定義特徵, 利用特徵進行顧客分群

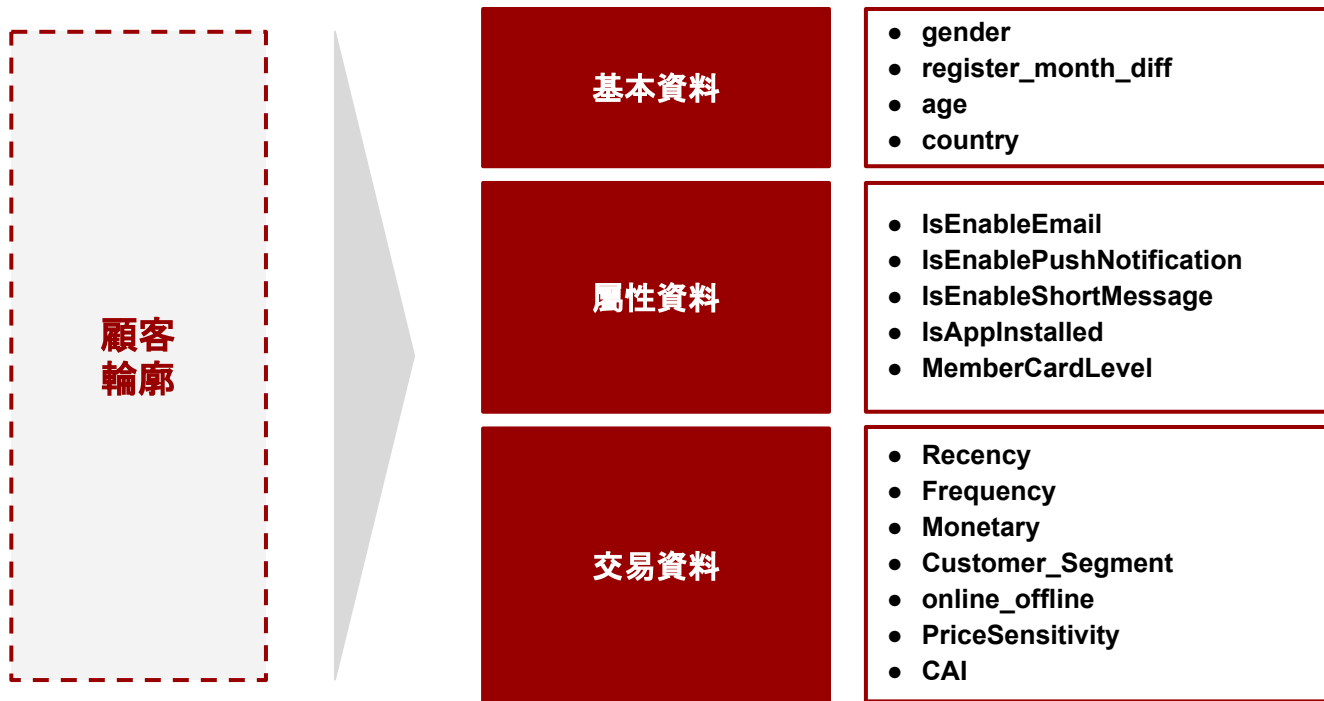
資料種類	Column Name	意義
交易資料	Recency	每個會員最後一次購買距離現在的天數。
	Frequency	每個會員的購買次數。
	Monetary	每個會員的總消費金額。
	Customer_Segment	將 R、F、M使用五等分法將顧客劃分為五個等級。 分數合併成一個 RFM 分數, 然後根據 RFM 分數進行ABCDE的分級。
	online_offline	線上及線下的購買習慣
	PriceSensitivity	計算總折扣使用金額和總消費金額, 再計算折扣使用金額佔總消費金額的比例。
	CAI	顧客的購買行為活躍度( $\text{average\_days} - \text{weighted\_average\_days}$ ) / $\text{average\_days} * 100$ )



## Agenda

- ① 分析目標與架構
- ② 資料前處理
- ③ 自定義特徵
- ④ **顧客輪廓**
- ⑤ 顧客分群
- ⑥ 結論

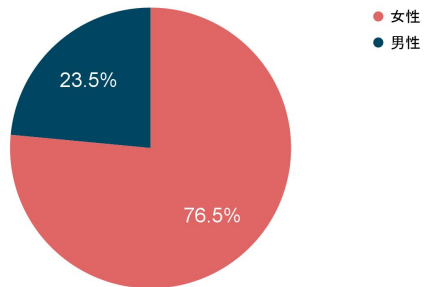
## 顧客輪廓



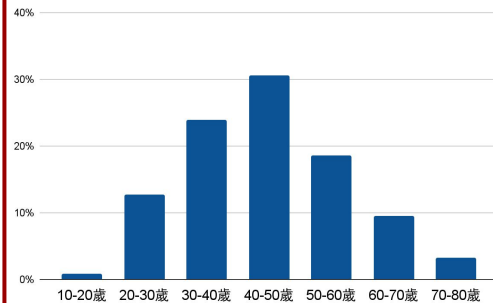
## 顧客輪廓概述

### 基本資料

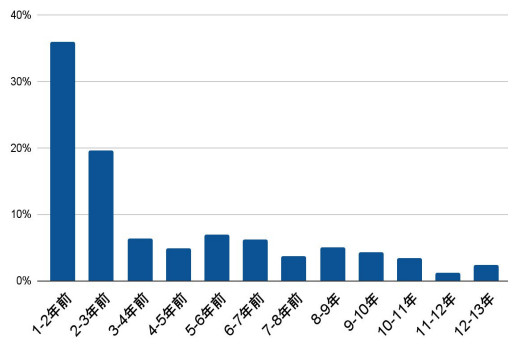
性別比



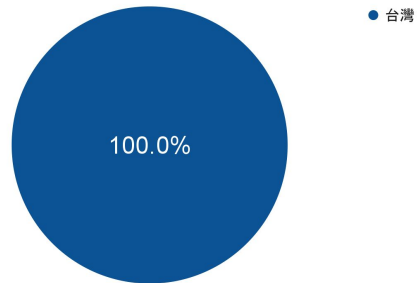
年齡分佈



註冊會員時間



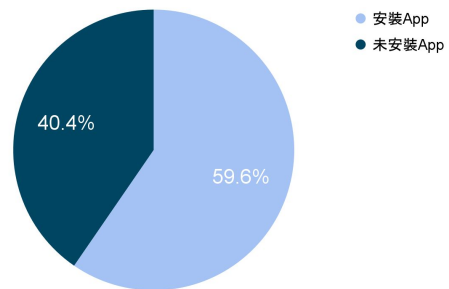
國籍



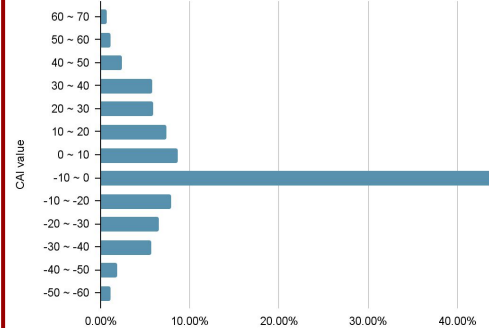
## 顧客輪廓概述

### 屬性資料

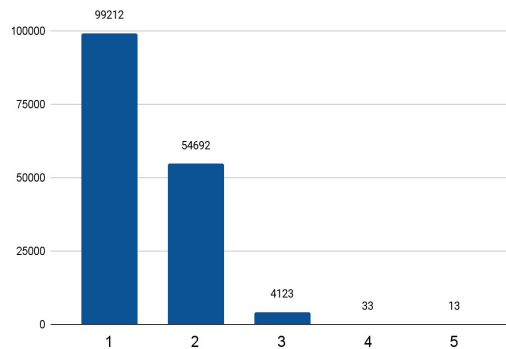
#### 是否安裝 App



#### CAI



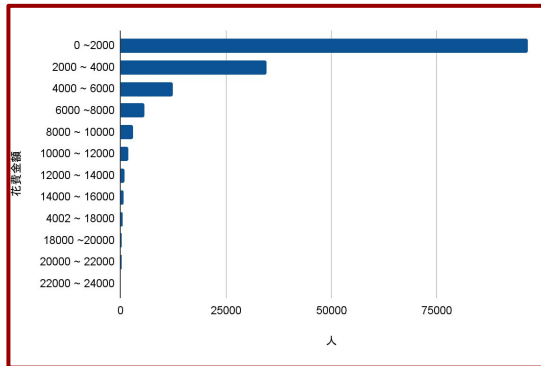
#### 會員卡等級



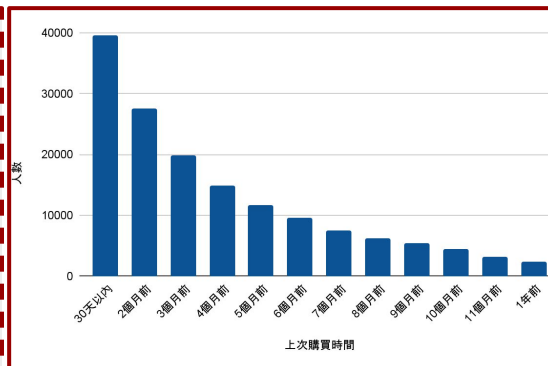
## 顧客輪廓概述

### 交易資料

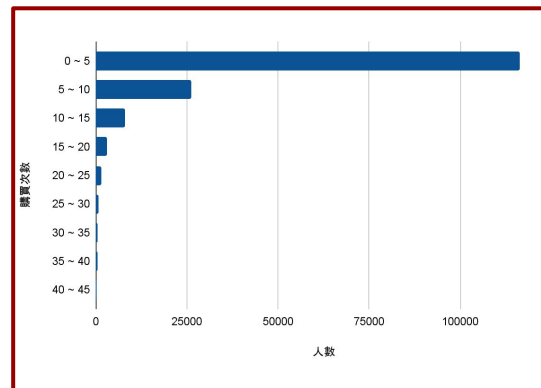
#### 總消費金額



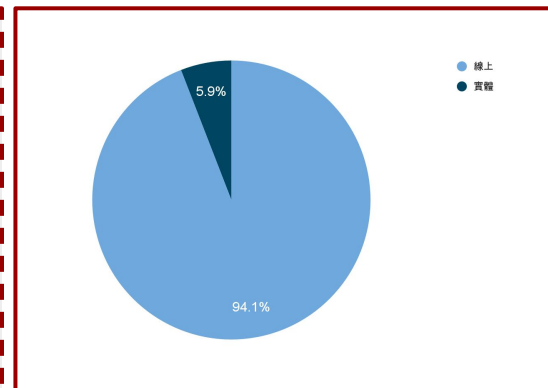
#### Recency



#### 總消費次數



#### 消費途徑



## Agenda

- ① 分析目標與架構
- ② 資料前處理
- ③ 自定義特徵
- ④ 顧客輪廓
- ⑤ **顧客分群**
- ⑥ 結論

## 第一階段分群標籤的特徵指標

分群	群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
人數佔比	5%	1%	6%	40%	48%
Recency	182	110	37	100	109
Frequency	8	7	23	4	4
Monetary	2703	2787	13515	1898	2025
最近一次消費距離時間長		常消費、總金額高 「高價值VIP客群」		群4與群5的特徵 趨近相似	

## 第一階段分群標籤的特徵指標

分群	群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
CAI	8.46	2.89	0.61	-0.08	0.13
線上購物	1	0.19	0.05	0	0
價格敏感度	0.4	0.32	0.35	0.22	0.27
APP下載	0.95	0.98	0.81	0	1
活躍度、價格敏感度最高			群4與群5的特徵趨近相似		



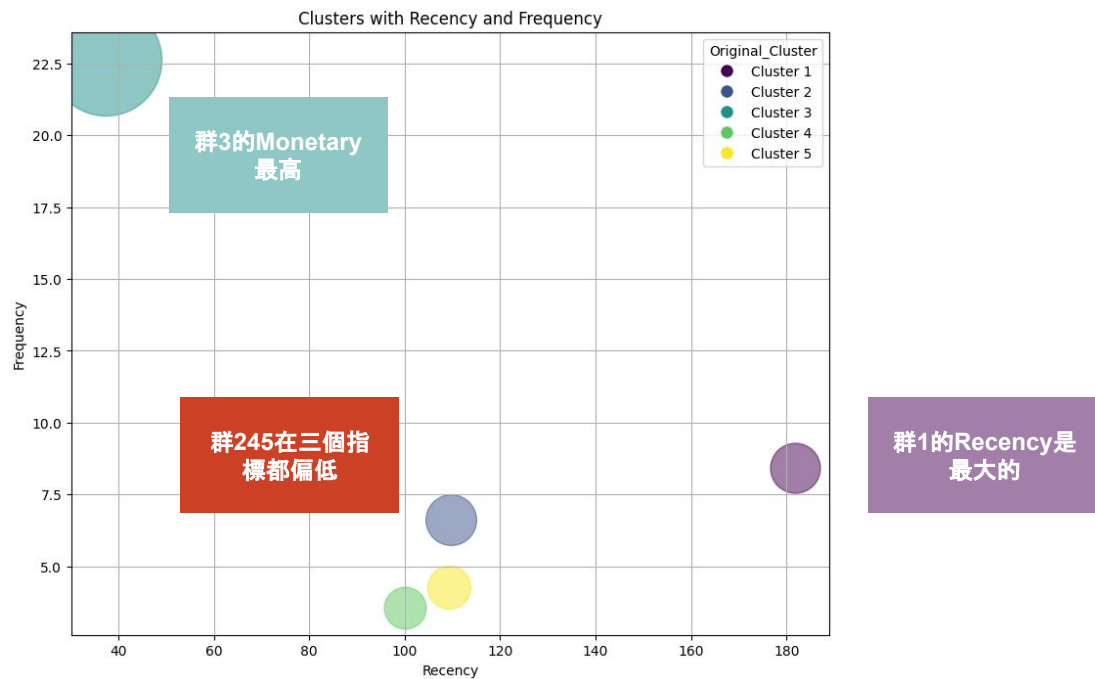
## 第一階段分群標籤的 Customer\_segement 佔比

分群		群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
Customer segement	A級客戶	0.15	0.23	0.66	0.19	0.2
	B級客戶	0	0.03	0	0.06	0.04
	D級客戶	0	0.01	0	0.02	0.02
	E級客戶	0.84	0.73	0.34	0.72	0.75
	人數佔比	5%	1%	6%	40%	48%

群3穩定高消費族群的A級客戶比例最高，但其他四個客群都是E級客戶的比例最高

## 第一階段客戶分群RFM

縱軸:購買次數 F | 橫軸:最近一次消費距離天數 R | 圓圈大小:消費金額 M



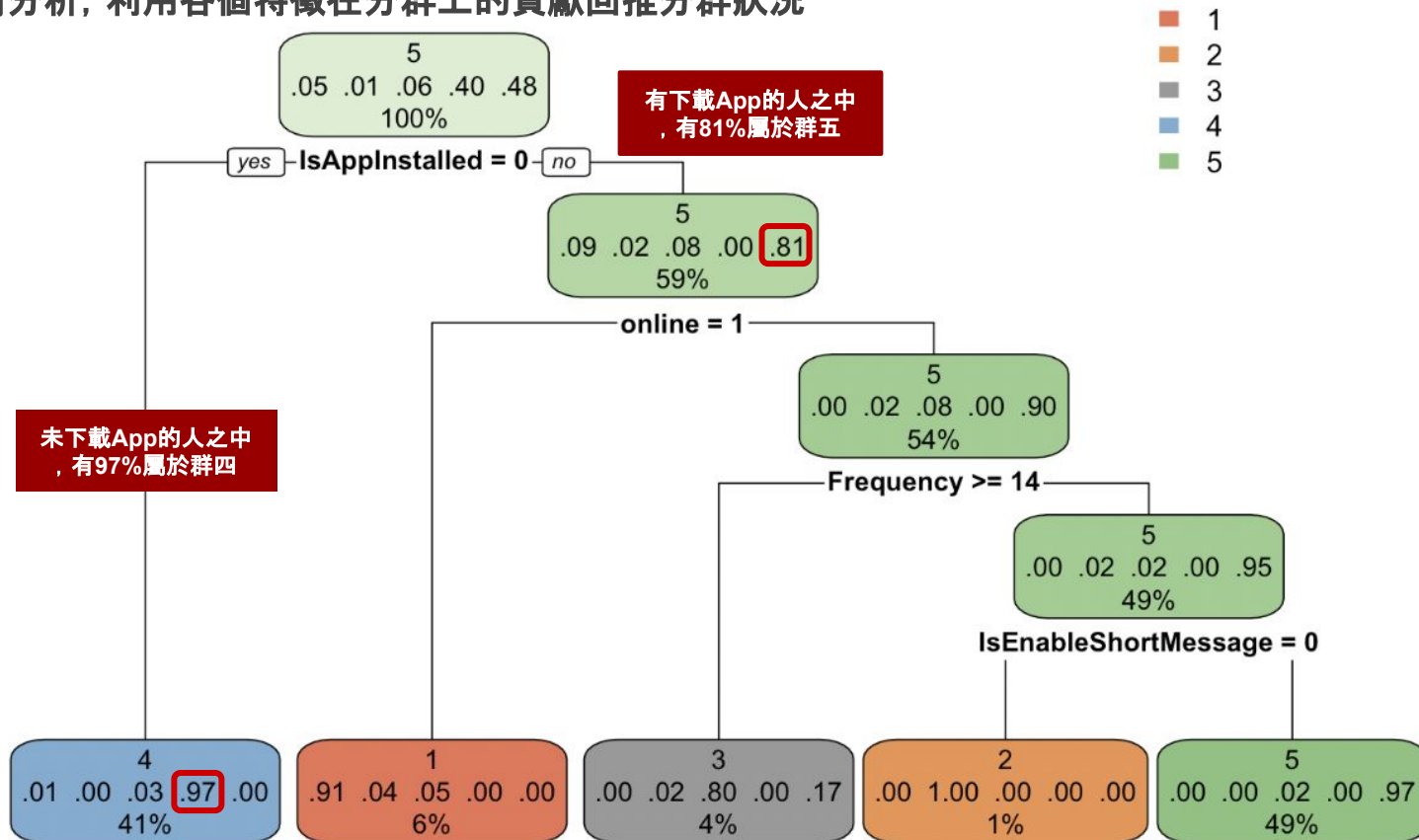
## 由分群數據中可發現不同群體的消費習性，並制定對應策略

分群	群一、線上客群(5%)	群二、實體消費族群(1%)	群三、穩定高消費族群(6%)
消費特徵	<ul style="list-style-type: none"><li>1. 完全線上消費</li><li>2. 折扣佔比高，對價格敏感</li><li>3. 於過去某時間段頻繁購買</li></ul>	<ul style="list-style-type: none"><li>1. 偏好實體店消費</li><li>2. 不喜歡線上消費與推播</li><li>3. 上次消費是三到四個月前</li></ul>	<ul style="list-style-type: none"><li>1. 主要為實體購買</li><li>2. 高消費且頻繁消費</li><li>3. 會員卡級別高</li><li>4. 距上次消費為一個月</li></ul>
顧客輪廓	線上折扣活動會購買的族群	高度依賴實體店的客群	店裡高消費常客

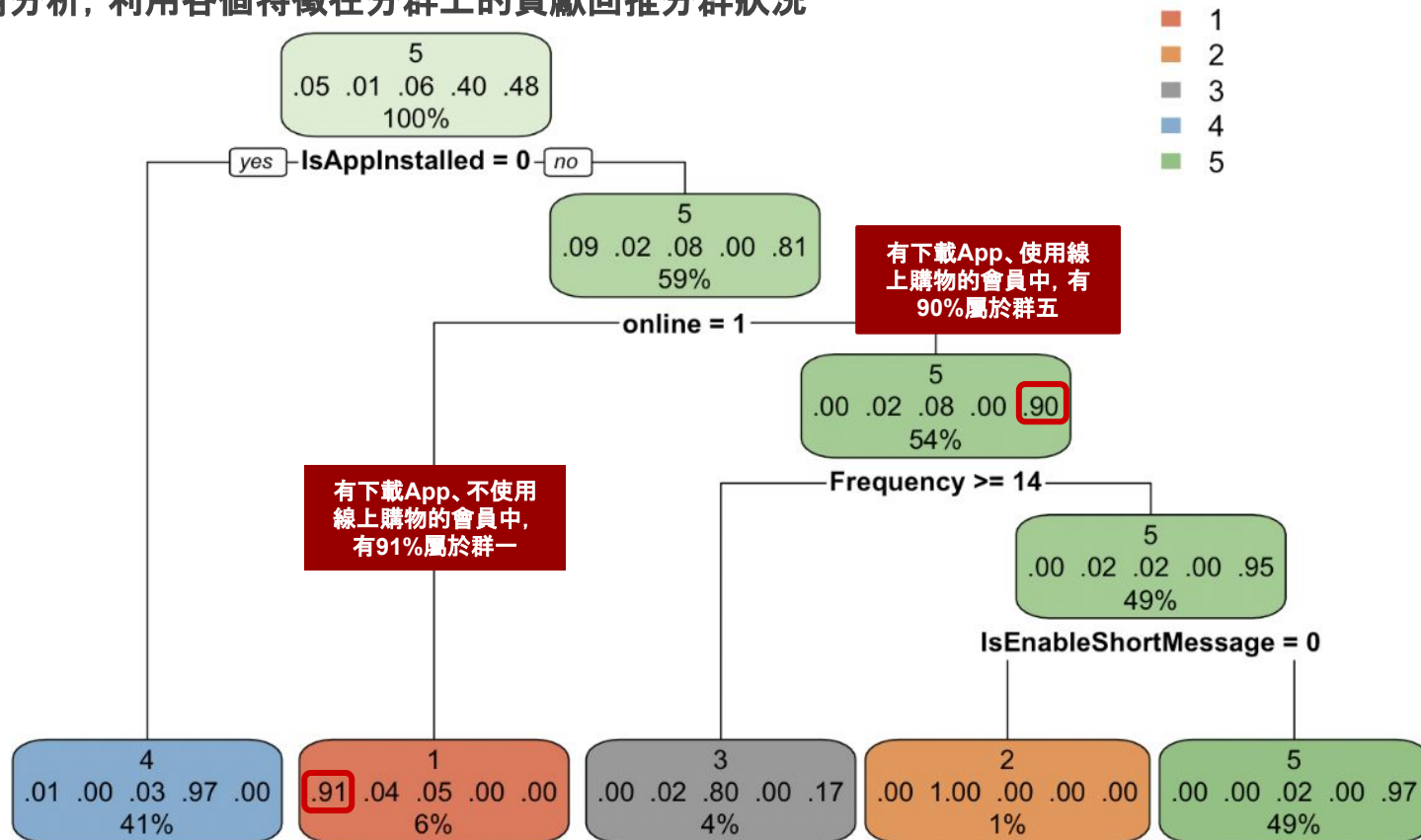
## 由分群數據中可發現不同群體的消費習性，並制定對應策略

分群	群四、逐漸流失群 (40%)	群五、路過看看族群 (48%)
消費特徵	<ul style="list-style-type: none"><li>1. 完全實體消費、沒有下載App</li><li>2. CAI值為負數，此族群可能漸漸流失</li><li>3. 對品牌忠誠度低</li></ul>	<ul style="list-style-type: none"><li>1. 與群四幾乎相同</li><li>2. 近期消費有比較活絡</li><li>3. CAI &gt; 0, 但有潛在流失風險</li></ul>
顧客輪廓	恰巧消費幾次的客群	恰好消費幾次的小資族群

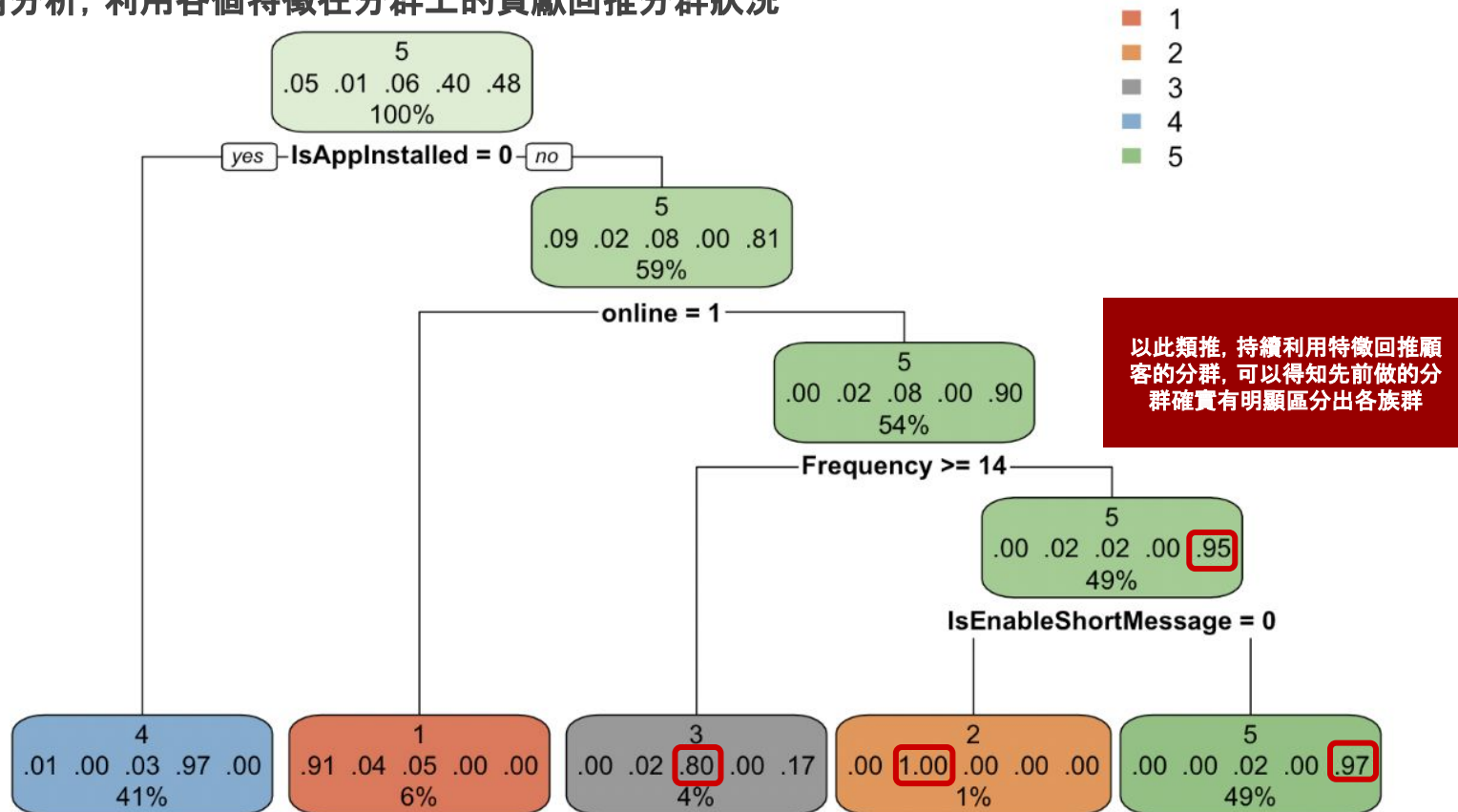
## 決策樹分析, 利用各個特徵在分群上的貢獻回推分群狀況



## 決策樹分析, 利用各個特徵在分群上的貢獻回推分群狀況

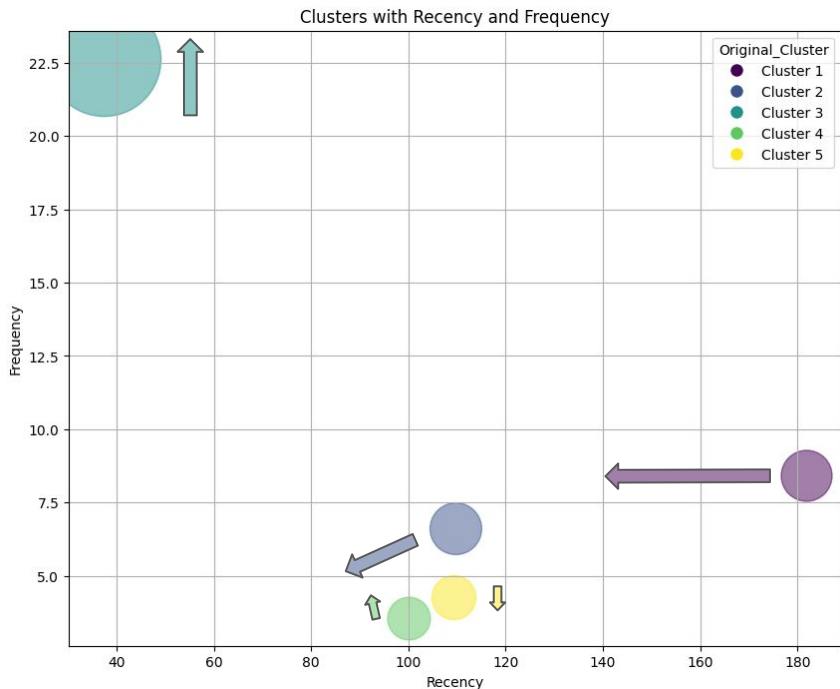


## 決策樹分析，利用各個特徵在分群上的貢獻回推分群狀況

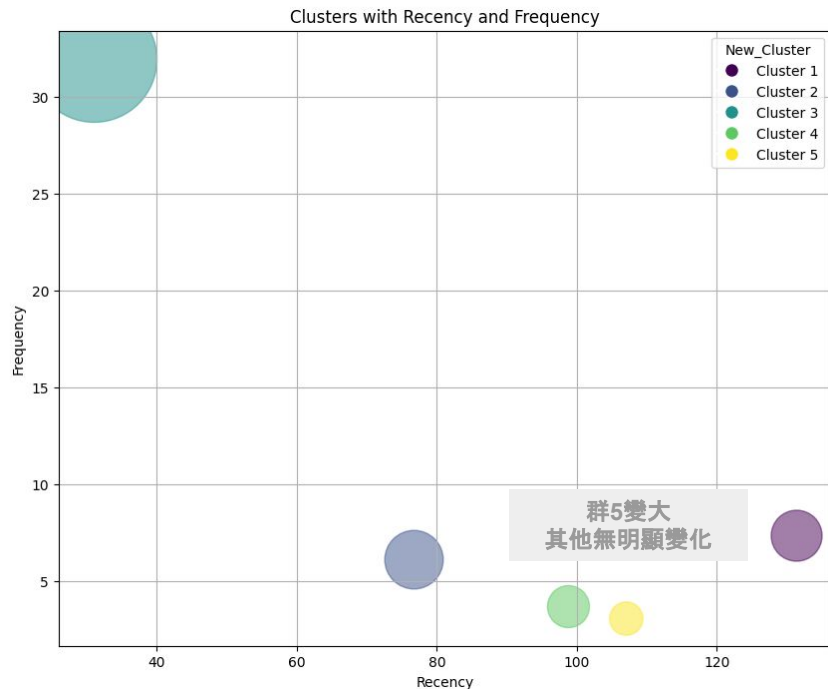


## 兩階段客戶分群的變化

縱軸：購買次數 F | 橫軸：最近一次消費距離天數 R | 圓圈大小：總消費金額 M



第一階段2022-02~2023-02



第二階段2023-02~2024-02



## 第二階段新分群標籤的特徵指標

分群	群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
人數佔比	6.8%	23.7%	2.3%	21%	46.2%
Recency	131.29	76.60	30.99	98.67	107.03
Frequency	7.39	6.13	31.96	3.72	3.11
Monetary	2703.29	3569.06	16450.26	1839.22	1158.59
最近一次消費距離時間長		消費金額次多	常消費、總金額高 「高價值VIP客群」	群4與群5的特徵 趨近相似	

## 第二階段新分群標籤的特徵指標

分群	群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
CAI	5.46	1.33	3.76	0.41	0.26
線上購物	0.45	0.00	0.14	0.00	0.00
價格敏感度	0.36	0.28	0.38	0.23	0.25
APP下載	0.94	0.56	0.81	0.43	0.40
活躍度顯著高 價格敏感度高		此四項特徵不明顯		活躍度高 價格敏感度高	
				群4與群5的特徵 趨近相似	

## 第二階段新分群標籤的 Customer\_segement 佔比

分群		群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
Customer segement	A級客戶	0.20	0.36	0.73	0.20	0.14
	B級客戶	0.01	0.01	0.00	0.04	0.06
	D級客戶	0.01	0.01	0.00	0.02	0.02
	E級客戶	0.78	0.61	0.27	0.74	0.78
	人數佔比	6.8%	23.7%	2.3%	21%	46.2%

群3穩定高消費族群的A級客戶比例最高，但其他四個客群都是E級客戶的比例最高

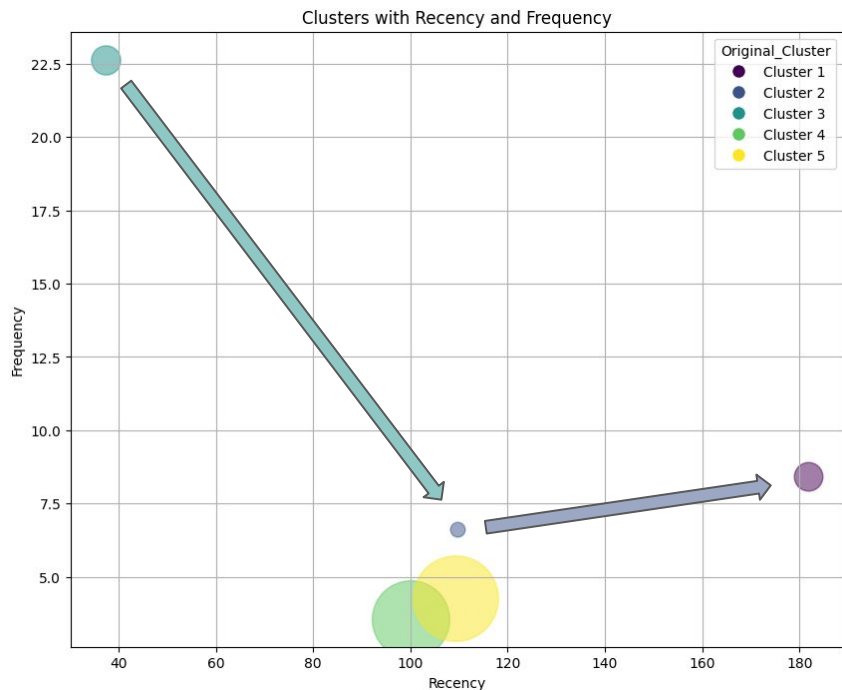
## 兩階段客戶分群的轉移矩陣

		第二階段				
第一階段	分群	群1 線上客群	群2 實體消費族群	群3 穩定高消費族群	群4 逐漸流失群	群5 路過看看族群
	群1	0.54	0.18	0.07	0.06	0.15
	群2	0.97	0.00	0.03	0.00	0.00
	群3	0.12	0.55	0.27	0.06	0.00
	群4	0.01	0.40	0.01	0.24	0.35
	群5	0.09	0.41	0.02	0.19	0.30

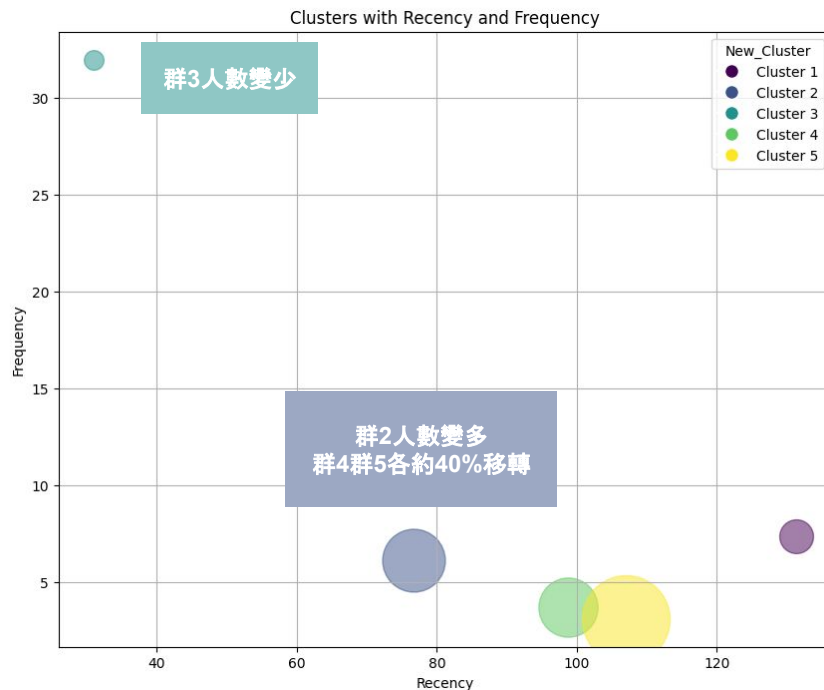
- 有**97%**於第一階段貼上**群2**標籤的顧客第二階段轉變成**群1**
- 有**55%**於第一階段貼上**群3**標籤的顧客第二階段轉變成**群2**
- 重要的高頻率且高金額消費的群3(穩定高消費實體購物群)於第二階段僅留下原先**27%**的顧客

## 兩階段客戶分群的變化

縱軸:購買次數F | 橫軸:最近一次消費距離天數R | 圓圈大小:人數



第一階段2022-02~2023-02



第二階段2023-02~2024-02

## Agenda

- ① 分析目標與架構
- ② 資料前處理
- ③ 自定義特徵
- ④ 顧客輪廓
- ⑤ 顧客分群
- ⑥ 結論

