

# 大數據行銷

## —以信用卡銀行為例

吳金擇

p09323028

國立臺灣大學經濟學系研究所

本文使用 2010 年 12 月至 2012 年 11 月信用卡資料，探討信用卡客戶刷卡行為。藉由建立顧客價值指標、購物籃分析及群集分析等，逐一分析客戶資料。研究結果發現，由集群分析所建立之市場區隔，最能有效將客戶作分群，此外，以行為變數作為市場區隔，亦較傳統地理人口變數佳，建議信用卡銀行應多著重客戶刷卡行為資料，並修正「刷卡日期」之錯誤，尚能做出更精準的行銷策略。

# 目錄

1 前言.....	1
2 資料描述.....	1
2.1 客戶資料檔.....	1
2.2 信用卡資料檔.....	2
2.3 信用卡交易紀錄資料檔.....	2
3 顧客價值指標.....	2
3.1 RFM 模型.....	2
3.2 CAI.....	3
3.3 CRI.....	5
4 購物籃分析.....	5
4.1 相關係數矩陣.....	5
4.2 條件機率矩陣.....	6
4.3 因素分析.....	8
5 集群分析.....	9
5.1 華德法.....	9
5.2 K-means.....	10
5.3 列聯表檢定.....	10
5.4 單因子獨立變異數分析.....	11
6 結論.....	15
附錄.....	15

## 圖目錄

圖 2-1: 樣本客戶組成比例 .....	2
圖 3-1: CAI 累積分配圖 .....	4
圖 3-2: 顧客活躍性分布圖 .....	4
圖 3-3: CAI 監控案例 .....	4
圖 5-1: 凝聚樹狀圖 .....	9
圖 5-2: 分群散布圖 .....	10
圖 5-3: 年齡層盒鬚圖.....	12
圖 5-4: 星座屬性盒鬚圖.....	14

## 表目錄

表 2-1: 信用卡資料庫綱要 .....	1
表 3-1: Bob Stone 給分機制.....	3
表 3-2: Bob Stone RFM 模型.....	3
表 3-3: CRI 等相關估計 .....	5
表 4-1: 相關係數矩陣.....	6
表 4-2: 相關係數 Top 20 .....	6
表 4-3: 條件機率矩陣.....	7
表 4-4: 條件機率 Top 20 .....	7
表 4-5: 因素分析 .....	8
表 4-6: 因素分數 .....	8
表 5-1: 變數分類 .....	9
表 5-2: K-means 分群性質 .....	10
表 5-3: 列聯表檢定 .....	11
表 5-4: ANOVA(ln_BS_Score) .....	12
表 5-5: 年齡層事後檢定 .....	13
表 5-6: ANOVA(CAI) .....	13
表 5-7: 星座屬性事後檢定 .....	14

## 1 前言

根據金融管理委員會銀行局信用卡業務統計，我國信用卡流通卡數<sup>1</sup>年年創新高，2020 年年底流通卡數 5,012 萬張，其中中國信託商業銀行、國泰世華商業銀行、玉山商業銀行分占 14.5%、14.2%、12.7%；有效卡數<sup>2</sup>3,300 萬張，以國泰世華商業銀行、中國信託商業銀行、玉山商業銀行分占 15.6%、14.3%、13.6%，足見市場競爭之激烈。

本文藉由「資料庫行銷」(Data Base Marketing)，透過客戶基本資料、信用卡資料及信用卡交易紀錄資料，分析每位客戶對信用卡銀行的貢獻，並藉由購物籃分析與集群分析等，提供以數據為基礎的市場區隔，協助信用卡銀行做出更精準的行銷策略。

## 2 資料描述

行銷資料庫可分為「客戶資本資料檔」、「交易日期與產品特性資料檔」及「客戶交易動態資料檔」。本文資料來源為我國某信用卡銀行，自其資料庫隨機抽取 100 位客戶，將相關資料分別記錄於「客戶資料檔」、「信用卡資料檔」及「信用卡交易紀錄資料檔」。

表 2-1: 信用卡資料庫綱要

客戶資料檔	信用卡資料檔	信用卡交易紀錄資料檔
客戶 ID	信用卡 ID	交易序號
年齡	客戶 ID	信用卡 ID
生日	信用卡開卡日	客戶 ID
居住地	信用卡到期日	刷卡日期
教育程度	信用額度	刷卡產品產業分類
性別	卡等	刷卡類型
婚姻狀況		刷卡金額
職業		刷卡地點
辦第一張信用卡的時間		

### 2.1 客戶資料檔

客戶資料檔紀錄客戶的靜態基本資料，包括客戶 ID、年齡、生日、居住地、教育程度、性別、婚姻狀況、職業、辦第一張信用卡的時間。藉由客戶資料檔，

<sup>1</sup> 發卡總數減停卡總數，且卡片狀況為正常者。

<sup>2</sup> 最近六個月有消費紀錄之卡，不含 Debit 卡，只有郵購分期交易亦算有效卡，不含只有循環繳款之卡片。

可概述客戶性質，如樣本客戶性別以男性稍多、年齡層以 50 歲以上居多、婚姻狀況以已婚者為主(圖 2-1)等。

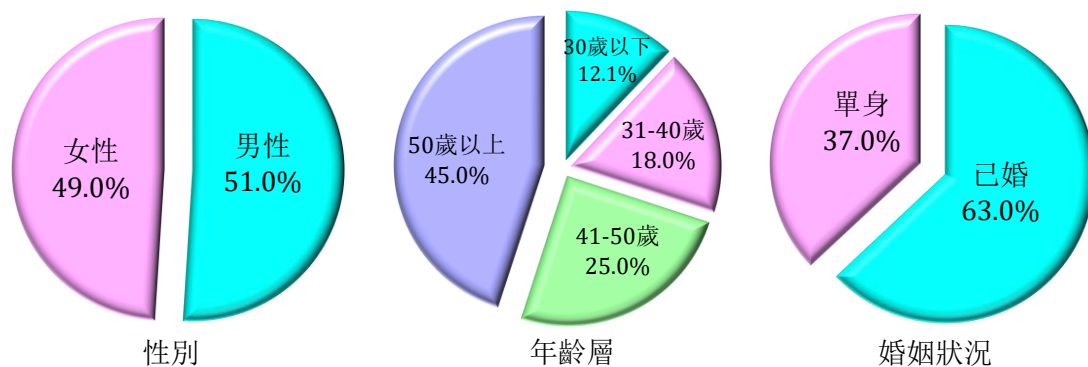


圖 2-1: 樣本客戶組成比例

## 2.2 信用卡資料檔

信用卡資料檔紀錄客戶所持信用卡相關資料，包括信用卡 ID、客戶 ID、信用卡開卡日、信用卡到期日、信用額度、卡等。藉由信用卡資料檔，可瞭解客戶持有信用卡的狀況，如樣本客戶每位平均擁有 2.4 張卡、卡等以白金卡最多以及客戶信用卡是否已到期等。

## 2.3 信用卡交易紀錄資料檔

信用卡交易紀錄資料檔紀錄客戶自 2010 年 12 月 1 日至 2012 年 11 月 30 日刷卡資料，包括交易序號、信用卡 ID、客戶 ID、刷卡日期、刷卡產品產業分類、刷卡類型、刷卡金額、刷卡地點。藉由信用卡交易紀錄資料檔，可瞭解客戶消費行為，如最常刷卡時間點、是否分期付款、是否國外消費等。

# 3 顧客價值指標

顧客價值係指客戶對企業的貢獻，常見的衡量方式有 RFM 模型、CAI 指標及 CRI 指標。

## 3.1 RFM 模型

RFM 模型由 3 項指標組成：最近購買日(Recency)、購買頻率(Frequency)及購買金額(Monetary Value)。若將 3 項指標分別依據五等均分法，即可將顧客區隔為  $5 \times 5 \times 5 = 125$  群體，惟因產業特性不同，對於指標的權重亦會不同，故本文參考 Bob Stone(1995)信用卡實證研究，將 3 項指標分數化並加總，以衡量每位顧客價值(詳細 Bob Stone 給分機制如表 3-1)。

表 3-1: Bob Stone 給分機制

指標	給分機制	權重
最近購買日	90 天內給 24 分	中
	180 天內給 12 分	
	270 天內給 6 分	
	360 天內給 3 分	
	超過 360 不給分	
購買頻率	刷卡次數 × 4 以 30 作為匯率換算	高
購買金額	平均刷卡金額(以美元計) × 10% 上限 9 分	低

以信用卡交易紀錄資料檔，將 Bob Stone RFM 模型結果列於表 3-2，可得顧客價值前 3 高者: 605、3044 及 13687，顧客價值最低 3 者: 3359、4263 及 18841。信用卡銀行可依此模型做顧客關係管理(Customer Relationship Management, CRM)，將每位客戶的顧客價值極大化。

表 3-2: Bob Stone RFM 模型

客戶 ID	R	F	M	BS_R	BS_F	BS_M	BS_Score
605	3	955	97	24	3,820	0	3,844
3044	5	661	331	24	2,644	1	2,669
13687	3	399	2,892	24	1,596	9	1,629
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3359	334	3	3,395	3	12	9	24
4263	205	3	194	6	12	1	19
18841	453	2	725	0	8	2	10

### 3.2 CAI

由於 RFM 模型未考慮顧客的動態性，故再以平均購買期間(MLE)及加權平均購買期間(WMLE)，編製顧客活躍性指標(Customer Activity Index, CAI)，提供另一顧客價值衡量方法。

$$CAI = \frac{MLE - WMLE}{MLE} \times 100\% \quad (1)$$

圖 3-1 以 CAI 配合 80/20 法則將樣本客戶區分為漸趨活躍型、穩定刷卡群

及漸趨靜止型；圖 3-2 則以 CAI 搭配購買頻率(F)區分客戶；圖 3-3 透過刷卡間隔天數監控每位客戶刷卡狀況，提早防範客戶落入靜止戶。

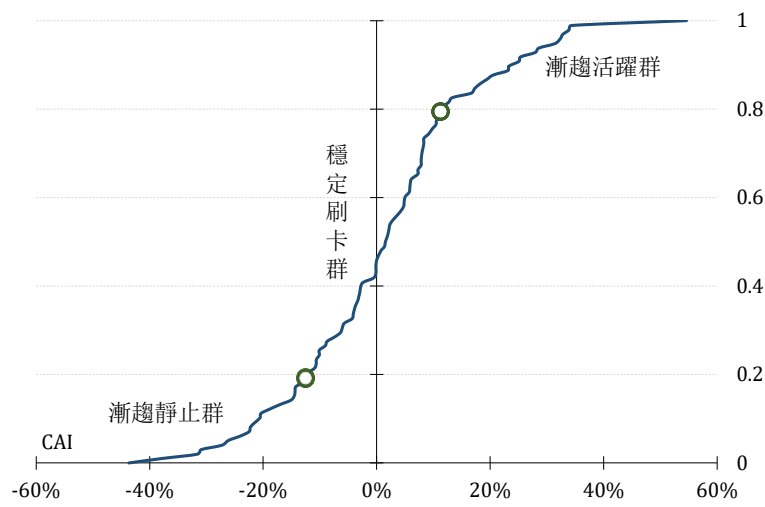


圖 3-1: CAI 累積分配圖

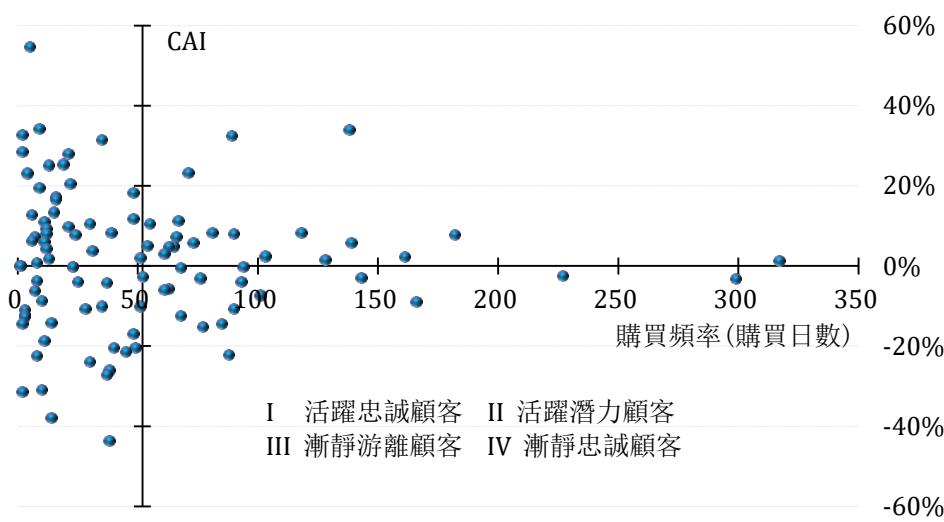


圖 3-2: 顧客活躍性分布圖

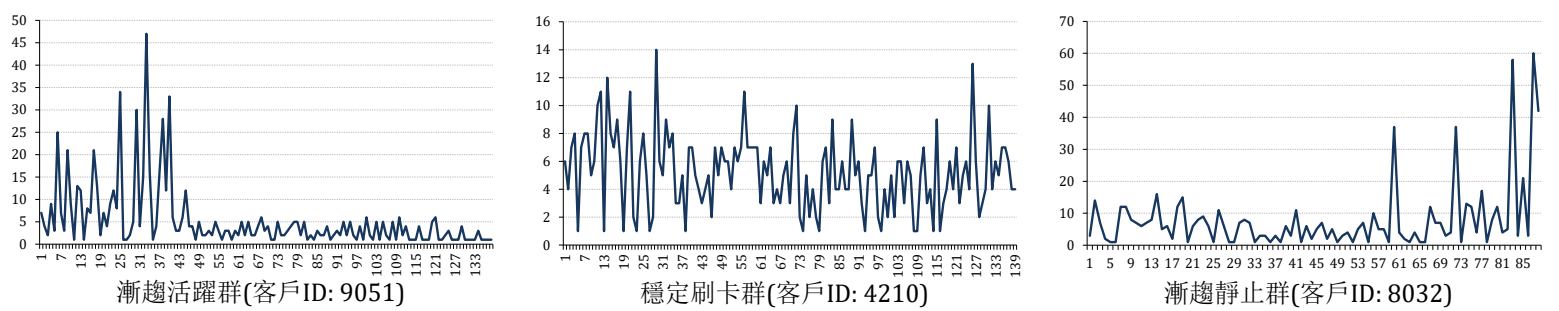


圖 3-3: CAI 監控案例

### 3.3 CRI

由於個人交易資料筆數有時較少，導致個人估計(Individual Estimate, IE)標準誤通常較大，故本文再以性別及婚姻狀況建立群體估計(Group Estimate, GE)，並以貝氏統計(Bayesian Statistics)建立貝氏估計(Bayesian Estimate, BE)，最後將三者結果結合，創造出新的顧客價值指標：交易穩定度指標(Customer Reliability Index, CRI)。當 CRI 越大，代表顧客交易資料的型態愈穩定。

$$CRI = \frac{IE - BE}{IE - GE} \times 100\% \quad (2)$$

以信用卡刷卡金額(新臺幣元)，將 CRI 等相關估計列於表 3-3，配合 80/20 法則，將位居 TOP 20%的客戶視為穩定群，位居 Bottom 20%的客戶視為高危險群，信用卡銀行即可透過相關行銷策略，維持穩定群並降低高危險群。

表 3-3: CRI 等相關估計

客戶 ID	刷卡金額_IE	刷卡金額_GE	刷卡金額_BE	CRI
9209	399	1,427	399	0.00
605	97	1,656	97	0.00
5963	394	2,077	394	0.00
⋮	⋮	⋮	⋮	⋮
4819	4,346	1,427	3,311	35.45
13802	6,059	1,427	4,374	36.36
17027	14,884	1,656	8,234	50.27

## 4 購物籃分析

由信用卡交易紀錄檔之刷卡產品產業分類欄位，可將信用卡交易分為教育、保險、捐贈等 15 類。本文藉由此資料，以下列各節方法分析客戶刷卡類別，作為信用卡刷卡促銷活動(如指定交通消費享回饋)之依據。

### 4.1 相關係數矩陣

$$r_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum(X_i - \bar{X})^2)(\sum(Y_i - \bar{Y})^2)}} \quad (3)$$

相關係數係用以衡量兩屬量變數之間的關係程度。為計算刷卡產品產業分類 15 類別的兩兩相關係數，須先針對所有客戶，依據 15 類別，分為「曾刷卡」與「不曾刷卡」，即可計算出相關係數矩陣。



表 4-1: 相關係數矩陣

	02	03	05	06	07	08	09	10	11	12	13	14	15	16	X2
02_學費/教育	1.00														
03_保險	0.36	1.00													
05_捐贈	0.16	0.08	1.00												
06_公用事業	0.06	-0.06	0.08	1.00											
07_百貨	-0.02	0.15	0.16	0.10	1.00										
08_精品	0.00	0.15	-0.01	0.04	0.29	1.00									
09_餐飲	0.09	0.16	0.07	0.00	0.35	0.32	1.00								
10_藥妝	0.13	0.27	0.21	0.00	0.32	0.41	0.40	1.00							
11_3C居家電信	0.09	0.20	0.07	0.04	0.28	0.26	0.38	0.30	1.00						
12_量販超市	0.11	0.12	0.12	0.06	0.18	0.23	0.18	0.10	0.15	1.00					
13_交通(含加值)	0.14	0.12	0.13	0.12	0.10	0.27	0.33	0.29	0.18	0.15	1.00				
14_旅遊	0.10	0.39	0.08	-0.03	0.23	0.47	0.28	0.27	0.18	0.11	0.26	1.00			
15_休閒文化	0.08	0.09	0.18	-0.02	0.38	0.37	0.32	0.32	0.26	0.12	0.32	0.40	1.00		
16_無店舖	0.19	0.19	0.25	0.14	0.26	0.26	0.23	0.37	0.30	0.09	0.35	0.18	0.42	1.00	
X2.中信錢加值	-0.05	0.08	-0.03	-0.08	0.04	-0.01	0.07	0.21	0.07	0.12	0.13	0.08	0.06	0.13	1.00

表 4-2: 相關係數 Top 20

Top	刷卡產品產業分類_列	刷卡產品產業分類_欄	相關係數
1	14_旅遊	08_精品	0.4725
2	16_無店舖	15_休閒文化	0.4182
3	10_藥妝	08_精品	0.4140
4	15_休閒文化	14_旅遊	0.4001
5	10_藥妝	09_餐飲	0.3964
6	14_旅遊	03_保險	0.3933
7	15_休閒文化	07_百貨	0.3829
8	11_3C居家電信	09_餐飲	0.3786
9	15_休閒文化	08_精品	0.3675
10	16_無店舖	10_藥妝	0.3662
11	03_保險	02_學費/教育	0.3572
12	16_無店舖	13_交通(含加值)	0.3491
13	09_餐飲	07_百貨	0.3465
14	13_交通(含加值)	09_餐飲	0.3306
15	15_休閒文化	10_藥妝	0.3217
16	10_藥妝	07_百貨	0.3206
17	15_休閒文化	09_餐飲	0.3194
18	15_休閒文化	13_交通(含加值)	0.3184
19	09_餐飲	08_精品	0.3158
20	16_無店舖	11_3C居家電信	0.3048

表 4-1 為 15 類別的相關係數矩陣，表 4-2 為相關係數前 20 高者。信用卡銀行可針對相關係數高的組合，擇一做刷卡回饋，或於客戶刷卡後，提供另一刷卡類別廣告，應皆能有效促使客戶再次刷卡。

#### 4.2 條件機率矩陣

$$P(\text{刷卡 A 類別}|\text{刷卡 B 類別}) = \frac{P(\text{刷卡 A 類別} \cap \text{刷卡 B 類別})}{P(\text{刷卡 B 類別})} \quad (4)$$

條件機率係事件 A 在事件 B 發生的條件下發生的機率。為計算刷卡產品產業分類 15 類別的兩兩條件機率，須先針對所有客戶交易紀錄，推算兩兩前後組合的次數，即可計算出條件機率矩陣。

表 4-3: 條件機率矩陣

	02	03	05	06	07	08	09	10	11	12	13	14	15	16	X2
02_學費/教育	0.07%	0.01%	0.00%	0.00%	0.00%	0.01%	0.01%	0.03%	0.04%	0.07%	0.10%	0.04%	0.00%	0.01%	0.00%
03_保險	0.04%	1.12%	0.00%	0.01%	0.09%	0.07%	0.07%	0.12%	0.16%	0.17%	0.29%	0.05%	0.03%	0.09%	0.00%
05_捐贈	0.00%	0.00%	0.12%	0.01%	0.01%	0.01%	0.01%	0.00%	0.01%	0.00%	0.16%	0.00%	0.01%	0.01%	0.00%
06_公用事業	0.00%	0.03%	0.05%	21.03%	0.05%	0.00%	0.04%	0.04%	0.16%	0.82%	0.47%	0.08%	0.03%	0.08%	0.00%
07_百貨	0.00%	0.05%	0.04%	0.04%	3.81%	0.34%	0.74%	0.23%	0.46%	1.17%	0.86%	0.13%	0.47%	0.12%	0.17%
08_精品	0.00%	0.04%	0.03%	0.01%	0.26%	0.52%	0.23%	0.27%	0.12%	0.40%	0.35%	0.16%	0.18%	0.03%	0.00%
09_餐飲	0.00%	0.04%	0.00%	0.04%	0.73%	0.14%	1.12%	0.17%	0.29%	0.42%	0.76%	0.23%	0.38%	0.03%	0.10%
10_藥妝	0.01%	0.12%	0.00%	0.01%	0.27%	0.20%	0.16%	0.85%	0.38%	0.42%	1.04%	0.13%	0.17%	0.10%	0.03%
11_3C居家電信	0.01%	0.18%	0.00%	0.21%	0.38%	0.29%	0.18%	0.39%	3.43%	0.68%	1.00%	0.14%	0.26%	0.25%	0.03%
12_量販超市	0.16%	0.16%	0.00%	0.77%	1.10%	0.37%	0.50%	0.40%	0.59%	9.22%	1.98%	0.39%	0.44%	0.21%	0.04%
13_交通(含加值)	0.05%	0.38%	0.09%	0.56%	1.17%	0.33%	0.81%	1.06%	1.02%	1.96%	11.83%	0.38%	0.64%	0.34%	0.10%
14_旅遊	0.01%	0.07%	0.00%	0.08%	0.22%	0.16%	0.16%	0.07%	0.14%	0.40%	0.40%	0.94%	0.20%	0.14%	0.00%
15_休閒文化	0.01%	0.05%	0.01%	0.01%	0.30%	0.17%	0.34%	0.14%	0.34%	0.35%	0.86%	0.20%	0.81%	0.17%	0.07%
16_無店舖	0.03%	0.08%	0.01%	0.08%	0.12%	0.04%	0.05%	0.08%	0.31%	0.12%	0.43%	0.12%	0.14%	0.76%	0.00%
X2_中信錢加值	0.00%	0.00%	0.00%	0.00%	0.13%	0.01%	0.07%	0.04%	0.01%	0.08%	0.14%	0.00%	0.04%	0.01%	0.29%

表 4-4: 條件機率 Top 20

Top	本次類別(列)	下次類別(欄)	條件機率
1	06_公用事業	06_公用事業	21.0%
2	13_交通(含加值)	13_交通(含加值)	11.8%
3	12_量販超市	12_量販超市	9.2%
4	07_百貨	07_百貨	3.8%
5	11_3C居家電信	11_3C居家電信	3.4%
6	12_量販超市	13_交通(含加值)	2.0%
7	13_交通(含加值)	12_量販超市	2.0%
8	07_百貨	12_量販超市	1.2%
9	13_交通(含加值)	07_百貨	1.2%
10	09_餐飲	09_餐飲	1.1%
11	03_保險	03_保險	1.1%
12	12_量販超市	07_百貨	1.1%
13	13_交通(含加值)	10_藥妝	1.1%
14	10_藥妝	13_交通(含加值)	1.0%
15	13_交通(含加值)	11_3C居家電信	1.0%
16	11_3C居家電信	13_交通(含加值)	1.0%
17	14_旅遊	14_旅遊	0.9%
18	15_休閒文化	13_交通(含加值)	0.9%
19	07_百貨	13_交通(含加值)	0.9%
20	10_藥妝	10_藥妝	0.8%

表 4-3 為 15 類別的條件機率矩陣，表 4-4 為條件機率前 20 高者，可發現客戶有於同一產業刷卡的現象，如一次繳完各種公共事業費或交通費用，信用卡銀行可藉由此特性，推出相關行銷活動。

### 4.3 因素分析

相關係數矩陣及條件機率矩陣僅能分析兩兩類別間的關係，無法提供「一籃子」概念，故此節以因素分析將 15 類別建構為 5 籃子，並計算出每位客戶對每個籃子的因素分數。

表 4-5: 因素分析

Baskets	分類	負荷量	特徵值 (累積百分比)
Basket 1	08_精品	0.77	3.06 (20.41%)
	15_休閒文化	0.64	
	14_旅遊	0.64	
	09_餐飲	0.62	
	07_百貨	0.60	
	10_藥妝	0.53	
	11_3C居家電信	0.49	
Basket 2	05_捐贈	0.69	1.63 (31.27%)
	16_無店舖	0.67	
	13_交通(含加值)	0.40	
Basket 3	03_保險	0.81	1.5 (41.29%)
	02_學費/教育	0.75	
Basket 4	X2.中信錢加值	0.90	1.19 (49.24%)
Basket 5	12_量販超市	0.71	1.11 (56.65%)
	06_公用事業	0.61	

以相關係數矩陣為輸入，採主成分分析及最大變異法，擷取負荷量大於 0.4 者，即可建構出 5 個購物籃(表 4-5)，如 Basket 1 由精品、休閒文化、旅遊等所組成。

表 4-6: 因素分數

客戶 ID	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
89	0.18	-0.82	-0.05	0.00	0.36
106	0.65	-1.00	0.26	0.25	-0.05
131	0.04	-0.25	1.26	-0.29	-1.93
⋮	⋮	⋮	⋮	⋮	⋮
19504	1.82	-0.64	-0.34	0.00	0.06
19639	1.48	-0.45	-0.70	-1.11	1.68
19859	1.52	0.02	1.06	0.10	-0.56

表 4-6 為因素分數表，可視為每位客戶對每個購物籃的偏好程度，以編號 89 客戶為例，可依序推薦精品、休閒文化、旅遊等類別廣告；編號 131 客戶的刷卡類別則以保險、學費為主。藉由因素分數表可有效建構每位客戶偏好的購物籃，達個別客戶產品推薦之效。

## 5 集群分析

客戶基本資料可分為地理、人口、心理及行為四類變數(表 5-1)，本文將 BS\_Score 取自然對數，確保符合常態假設後，與 CAI 做為分群變數，第一階段先以階層式集群分析法(華德法)決定區隔數，第二階段再以非階層式集群分析法(K-means)進行市場區隔，並透過列聯表檢定描述各群客戶輪廓，最後，以變異數分析辨別市場區隔的有效性。

表 5-1: 變數分類

地理	人口	心理	行為
居住地	教育程度	星座屬性	曾分期付款
	性別		曾國外消費
	婚姻狀況		最常消費週時點
	職業		最常消費月時點
	年齡層		BS_Score
			CAI

### 5.1 華德法

本文以歐式距離衡量資料距離，視每一個個體為一個集群，將各集群依序合併，合併順序視合併後集群組內變異大小而定。圖 5-1 將凝聚順序表繪製為凝聚樹狀圖，可知 3 或 4 皆為適當的分群數。

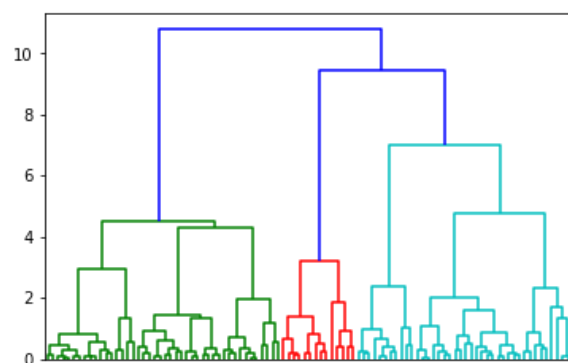


圖 5-1: 凝聚樹狀圖

## 5.2 K-means

由華德法決定分群數為 3 後，再以 K-means 將客戶分群。圖 5-2 為分群結果散布圖，表 5-2 為各群性質。由於群 0 的 BS Score 最低、CAI 最高，猜測為新戶，是最需要著重的目標客群，命名為發展群；群 1 的 BS Score 最高、CAI 次高，是信用卡銀行的主要客群，命名為主力群；群 2 的 BS Score 最低、CAI 也最低，命名為離失群。

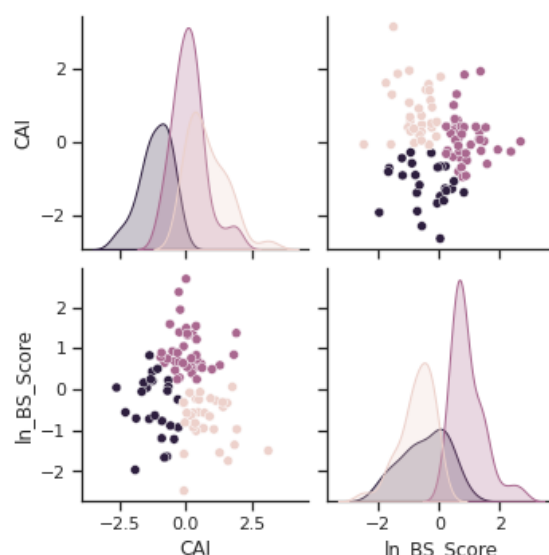


圖 5-2: 分群散布圖

表 5-2: K-means 分群性質

K-means	樣本大小	R	F	M	BS Score	CAI	命名
0	41	100.3	16.2	2,402.8	88.5	0.8	發展群
1	32	7.9	157.7	2,251.8	660.5	0.1	主力群
2	26	143.2	29.9	2,877.9	142.3	-1.2	離失群

## 5.3 列聯表檢定

列聯表檢定(卡方檢定)適用於探討兩類別變數相關性，本節將 K-means 結果視為一類別變數，分別與表 5-1 各類變數作檢定，用以描述各群客戶之輪廓。

為版面配置，表 5-3 僅呈現各類最顯著之變數，完整檢定內容請參閱附錄。由表 5-1 及表 5-3 知，地理變數方面，僅「居住地」一項，但與分群結果無顯著關係；人口變數方面，以「年齡層」的 p 值為最小；心理變數方面，僅「星座屬性」一項，但與分群結果有顯著關係；行為變數方面，以「國外消費」最為顯著，p 值僅 0.04。

表 5-3: 列聯表檢定

變數	項目	主力群 (0.41)	發展群 (0.32)	離失群 (0.26)	卡方值 (P-value)
地理變數 居住地	台中地區	0.35	0.29	0.35	1.57 (0.95)
	嘉南地區	0.33	0.44	0.22	
	大台北地區	0.44	0.31	0.25	
	高屏地區	0.43	0.33	0.24	
人口變數 年齡層	30 歲以下	0.33	0.25	0.42	7.44 (0.28)
	31-40 歲	0.61	0.22	0.17	
	41-50 歲	0.48	0.36	0.16	
	50 歲以上	0.32	0.36	0.32	
心理變數 星座屬性	土象	0.27	0.36	0.36	8.96 (0.18)*
	水象	0.32	0.41	0.27	
	火象	0.50	0.17	0.33	
	風象	0.52	0.38	0.10	
行為變數	不曾國外消費	0.35	0.38	0.27	6.56
國外消費	曾國外消費	0.64	0.14	0.23	(0.04)*

\*代表小於顯著水準 0.2

紅字代表顯著且高於該群群體比例

#### 5.4 單因子獨立變異數分析

單因子獨立變異數分析(Analysis of Variance, ANOVA)係用以檢定各群體平均值是否具顯著差異，其基本假設為母體群為常態分配、樣本間相互獨立且變異數相同。本節以  $\ln\_BS\_Score$  及 CAI 為顧客價值指標，分別與表 5-1 及 K-means 結果作檢定，以辨別該市場區隔的有效性。(為版面配置，以下僅呈現各類最顯著之變數，完整檢定內容請參閱附錄。)

##### 5.4.1 BS Score

為符合常態假設，須先將 BS Score 取自然對數，再進行檢定。表 5-4 顯示：地理變數僅「居住地」一項，其 p 值為 0.94，各居住地無顯著差異，不適合作為市場區隔變數；人口變數以「年齡層」的 p 值為最小，各年齡層具顯著差異，可作為市場區隔變數；心理變數僅「星座屬性」一項，其 p 值為 0.67，星座屬性不適合作為市場區隔變數；行為變數以「國外消費」最為顯著，p 值相當顯著；K-means 分群則是各變數中，P 值最小者，最適合作為市場區隔變數。

表 5-4: ANOVA(ln\_BS\_Score)

變數	項目	平均數	標準差	樣本大小	F 值 (P-value)
地理變數 居住地	台中地區	-0.10	0.85	17	0.13 (0.94)
	嘉南地區	0.05	0.76	9	
	大台北地區	0.07	1.01	52	
	高屏地區	-0.03	1.16	21	
人口變數 年齡層	30 歲以下	0.20	0.75	12	2.21 (0.09)*
	31-40 歲	0.36	0.98	18	
	41-50 歲	0.15	1.15	25	
	50 歲以上	-0.25	0.91	44	
心理變數 星座屬性	土象	-0.22	1.08	11	0.51 (0.67)
	水象	-0.03	0.93	37	
	火象	0.02	0.88	30	
	風象	0.22	1.20	21	
行為變數	不曾國外消費	-0.17	0.98	77	13.30
國外消費	曾國外消費	0.66	0.73	22	(0.00)*
K-means 分群	主力群	0.93	0.54	41	78.11 (0.00)*
	發展群	-0.75	0.57	32	
	離失群	-0.47	0.77	26	

\*代表小於顯著水準 0.2

由表 5-4 雖可瞭解變數項目是否具顯著差異，但若欲知差異細節，則須再作事後檢定。本文僅以年齡層為例，採 Tukey HSD 事後檢定，並輔以盒鬚圖說明，其他顯著變數之事後檢定結果列於附錄，不再贅述。

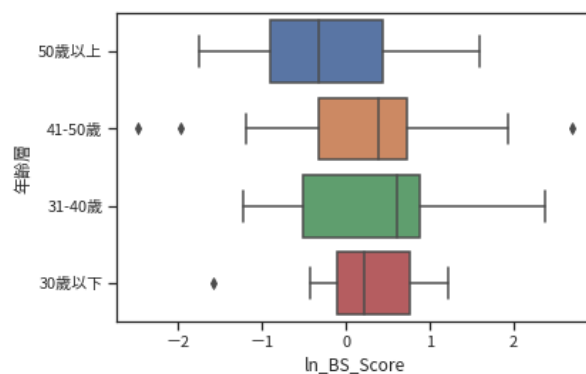


圖 5-3: 年齡層盒鬚圖

表 5-5: 年齡層事後檢定

群組 1	群組 2	群組差異	P-value	拒絕與否
30 歲以下	31-40 歲	-0.16	0.90	FALSE
30 歲以下	41-50 歲	0.05	0.90	FALSE
30 歲以下	50 歲以上	0.46	0.48	FALSE
31-40 歲	41-50 歲	0.21	0.90	FALSE
31-40 歲	50 歲以上	0.62	0.11	TRUE
41-50 歲	50 歲以上	0.41	0.34	FALSE

由圖 5-3 及表 5-5 可知，31-40 歲較 50 歲以上高 0.62 單位，以顯著水準 0.2 為標準，兩組具顯著差異，代表若以  $\ln\_BS\_Score$  為顧客價值指標，年齡層 31-40 歲的顧客價值較年齡層 50 歲以上高，是信用卡銀行最重要的年齡層。

#### 5.4.2 CAI

表 5-6: ANOVA(CAI)

變數	項目	平均數	標準差	樣本大小	F 值 (P-value)
地理變數 居住地	台中地區	-0.13	0.95	17	0.30 (0.82)
	嘉南地區	0.22	1.07	9	
	大台北地區	0.04	1.01	52	
	高屏地區	-0.08	1.03	21	
人口變數 教育程度	其他	-1.47	-	1	1.99 (0.09)*
	初中及初中以下	-0.32	1.06	4	
	高中	0.16	1.19	19	
	專科	0.45	0.85	19	
	學士	-0.23	0.93	44	
	研究所以上	0.10	0.93	12	
心理變數 星座屬性	土象	-0.29	0.62	11	2.25 (0.09)*
	水象	0.04	1.01	37	
	火象	-0.24	1.03	30	
	風象	0.43	1.01	21	
行為變數	不曾國外消費	0.08	1.03	77	2.03
國外消費	曾國外消費	-0.27	0.88	22	(0.16)*



變數	項目	平均數	標準差	樣本大小	F 值 (P-value)
K-means 分群	主力群	0.11	0.63	41	62.32 (0.00)*
	發展群	0.79	0.76	32	
	離失群	-1.15	0.60	26	

\*代表小於顯著水準 0.2

表 5-6 顯示：地理變數僅「居住地」一項，其 p 值為 0.82，各居住地無顯著差異，不適合作為市場區隔變數；人口變數以「教育程度」的 p 值為最小，教育程度不同具顯著差異，可作為市場區隔變數；心理變數僅「星座屬性」一項，其 p 值為 0.09，適合作為市場區隔變數；行為變數以「國外消費」最為顯著，p 值為 0.16；K-means 分群與表 5-4 結論相同，是各變數中，P 值最小者，最適合作為市場區隔變數。

由表 5-6 雖可瞭解變數項目是否具顯著差異，但若欲知差異細節，則須再作事後檢定。本文僅以星座屬性為例，採 Tukey HSD 事後檢定，並輔以盒鬚圖說明，其他顯著變數之事後檢定結果列於附錄，不再贅述。

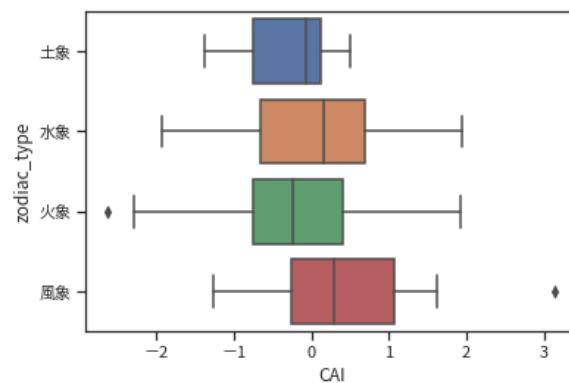


圖 5-4: 星座屬性盒鬚圖

表 5-7: 星座屬性事後檢定

群組 1	群組 2	群組差異	P-value	拒絕與否
土象	水象	-0.32	0.75	FALSE
土象	火象	-0.05	0.90	FALSE
土象	風象	-0.71	0.21	FALSE
水象	火象	0.28	0.64	FALSE
水象	風象	-0.39	0.47	FALSE
火象	風象	-0.67	0.09	TRUE

由圖 5-4 及表 5-7 可知，風象星座較火象星座高 0.67 單位，以顯著水準 0.2 為標準，兩組具顯著差異，代表若以 CAI 為顧客價值指標，風象星座的顧客價值較火象星座高，是信用卡銀行可加以著重的客群。

## 6 結論

章節 3 藉由建立顧客價值指標，提供維護顧客關係之衡量基準，如透過 CAI 監控每位客戶刷卡狀況，提早防範客戶落入靜止戶；章節 4 藉由購物籃分析，鎖定客戶需求，提供信用卡活動、廣告行銷之基準，如透過因素分數，依序提供對應產業廣告，以提高廣告效益；章節 5 比較不同變數的市場區隔效果，得出以行為變數作群集分析，優於傳統地理人口變數，最適合作為分眾行銷之依據。

章節 5 另以「最常消費週時點」分析消費者最常消費時間點，惟本信用卡資料庫之「刷卡日期」並非客戶刷卡當下的日期，造成資料庫中，週六筆數偏少，甚至無週日資料，使得分析受限。

此外，由於資料樣本大小僅 100 位客戶，本文皆為未分群分析，亦未做二因子變異數分析(Two way ANOVA)、多元迴歸分析(Multiple regression analysis)等複雜模型，故分析結果有其限制。

## 附錄

本文各項數據皆由 Python 分析而得，相關作業檔可點選以下連結查看。

1. [RFM 模型作業檔](#)
2. [CAI 作業檔](#)
3. [CRI 作業檔](#)
4. [相關係數矩陣與條件機率矩陣作業檔](#)
5. [因素分析作業檔](#)
6. [群集分析作業檔](#)