

信用卡資料庫分析與行銷策略建議

會計三 B09702036 郭旭崴

目錄

圖目錄.....	3
表目錄.....	4
壹、前言.....	5
貳、摘要.....	6
參、資料敘述.....	7
3-1 資料來源.....	7
3-2 資料來源.....	8
3-3 小結	10
肆、資料分析.....	10
4-1 客戶靜態分析.....	10
(一)RFM 指標-五等均分法	10
(二)bobstone-自製權重法	11
(三)綜合分析	13
(四)小結	13
4-2 客戶動態分析.....	14
(一)CAI 指標(交易活躍性指標)	14

(二)CRI 指標(交易穩定度指標)	16
(三)小結	16
4-3 購物籃分析	17
(一)同時購買購物籃之分析	18
(二)下次購買購物籃之分析	19
(三)因素分析	19
(四)小結	20
4-4 客戶集群分析	21
(一)K-means 法	21
(二)各組別人口敘述統計	22
(三)F 檢定	23
(四)事後檢定	25
(四)小結	27
伍、結論	27

圖目錄

圖 1、年齡分布	9
圖 2、地區分布	9
圖 3、性別分布	9
圖 4、婚姻狀況分布	9
圖 5、職業分布	9
圖 6、RFM 和 Bob stone 比較示意圖	13
圖 7、刷卡間隔天數	15
圖 8、刷卡金額	15
圖 9、同時購買相關矩陣.....	18
圖 10、前 20 名購物籃排名推薦.....	18
圖 11、下次購買之機率矩陣.....	19
圖 12、F 檢定-若即若離型	24
圖 13、F 檢定-即將離去型	24
圖 14、F 檢定-財神爺型	24
圖 15、F 檢定-超級新星型	25
圖 16、若即若離型事後 T 檢定.....	26

表目錄

表 1、客戶資料檔示意圖.....	7
表 2、信用卡資料檔示意圖.....	7
表 3、信用卡交易檔示意圖.....	7
表 4、各組別之平均 R、F、M 以及給分.....	11
表 5、RFM 給分示意圖	11
表 6、自製權重給分機制.....	12
表 7、自製權重給分示意圖.....	12
表 8、CAI 分群結果示意圖	15
表 9、CRI 分群結果	16
表 10、CRI 示意圖	16
表 11、分群 CRI 示意圖.....	17
表 12、主成分分析	20
表 13、K-means 分群結果	21
表 14、各組名稱及人數.....	22
表 15、各組人口統計	23
表 16、各客群對應行銷策略.....	28
表 17、各項購物籃推薦品項.....	28

壹、前言：

現在這個時代，我國銀行業百家爭鳴、百花齊放。各家銀行無不希望在有限資源投入的情況下，來獲取最大商業利益。而隨著過去幾十年資訊科技的迅速發展，各個公司所掌握的客戶資料和商品資料已經能夠逐漸被完整記錄，因此過去有一些不可能發生的行銷策略正在慢慢可以被實現，例如：一對一互動式行銷。因此這份報告將以信用卡刷卡交易為例，藉用某信用卡公司資料庫既有的現行資料，來敘述分析公司針對不同客群應採取哪些不同的策略，以實現用有限資源來獲取最大利益的理想。

貳、摘要：

為了有效提供公司能夠精準掌握有價值客戶和制訂正確行銷策略，本報告以所獲得的資料進行不同面向的分析，並同時制定出不同指標以衡量不同消費行為，以下簡述各項指標和分析之意涵。

一、 RFM 指標:用來**衡量顧客靜態價值指標**，包含 R(recency)、F(frequency)以及 M(monetary)這 3 項指標，幫助我們衡量顧客對公司的總價值。

二、 CAI 指標:用來**衡量顧客的活躍性指標**，屬於動態性指標，是 RFM 指標的延伸，可以幫助我們衡量漸趨活躍型和漸趨消極型的客戶。

三、 CRI 指標:用來**衡量顧客的穩定度指標**，屬於動態性指標，可以幫助我們衡量客戶消費頻率的穩定度。

四、 購物籃分析:用來**衡量商品之間組合搭配的指標**，可以幫助我們去了解商品之間各種搭配組合的價值，以此來制定相應的購物籃組合。

五、 客戶集群分析:用來**集群相同行為客戶的方法**，可以幫助我們歸納出相同性質的顧客，以利於進行精準行銷。

參、資料敘述

3-1、資料來源

本次報告研究之資料皆來自於台灣某信用卡公司之資料庫，其中包含 3 大檔案：客戶資料檔(表 1)、信用卡資料檔(表 2)和信用卡交易檔(表 3)。

表 1、客戶資料檔示意圖

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
89	59	1953/9/21	大台北地區	學士	1:男	已婚	白領主管	2003/11/27

客戶資料檔標示著擁有該公司信用卡的 100 位客戶之基本資料，包含客戶 ID、年齡、生日、居住地、教育程度、性別婚姻狀況、職業以及辦第一張信用卡的時間。

表 2、信用卡資料檔示意圖

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
41	2027	2006/5/19	2015/8/31	500000	金卡

在信用卡資料檔中，則是表達了各客戶所擁有之信用卡 ID、信用卡開卡日、信用卡到期日、信用額度以及卡等。

表 3、信用卡交易檔示意圖

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
8	6448	15195	2010/12/1	14_旅遊	一般消費	5200	國內

在信用卡交易檔中則是包含了從 2010 年 12 月 1 日到 2012 年 11 月 30 日的每一位客戶所刷卡的資料檔，其中包含：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、刷卡金額以

及刷卡地點。

3-2 資料狀況概述

根據我們所掌握的上述資料，我們可以簡單的先從幾個不同的方面去敘述這一份客戶資料。

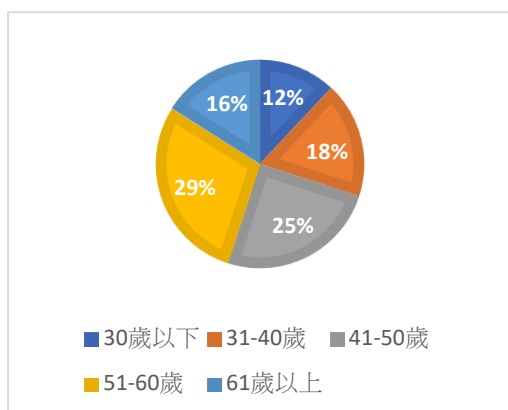


圖 1、年齡分布

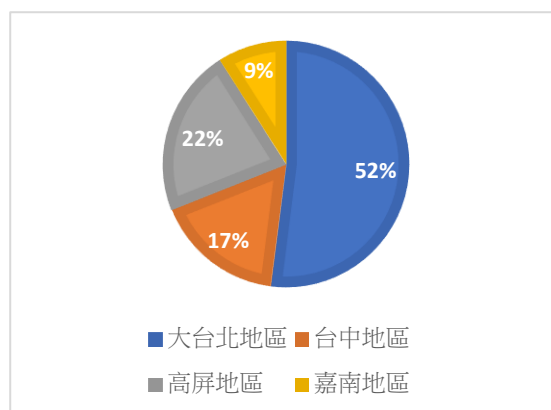


圖 2、地區分布

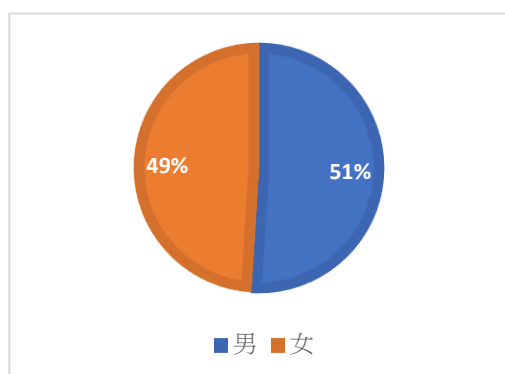


圖 3、性別分布

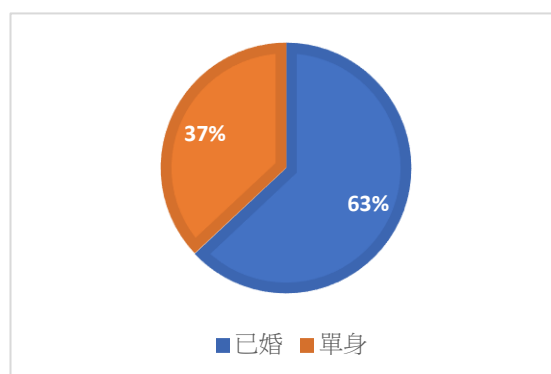


圖 4、婚姻狀況分布

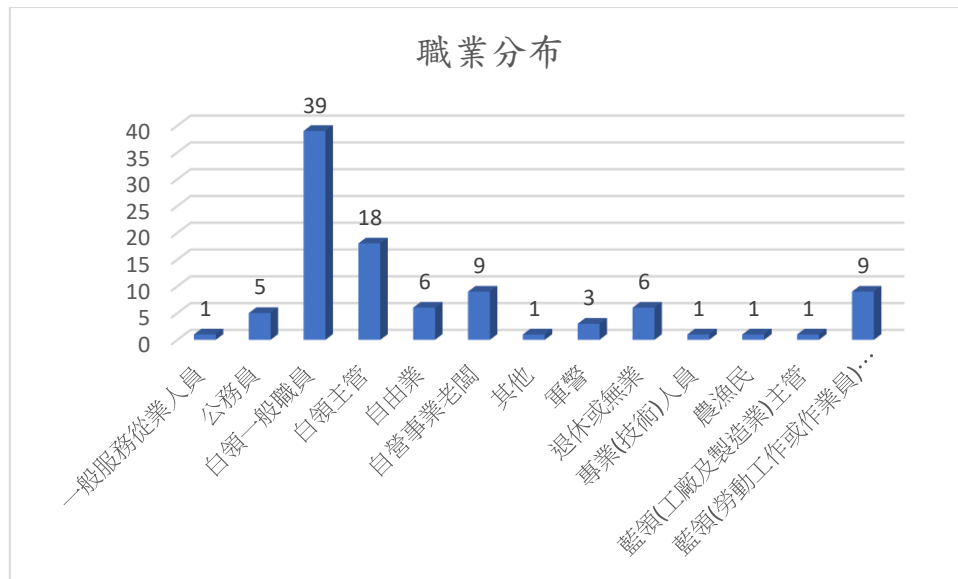


圖 5、職業分布

從上面各個圖表中，我們可以獲得以下資訊：

- 1、 就年齡分布來說，這 100 位客戶是以 51-60 歲為最大宗，佔整體的 29%，接著第二大宗為 41-50 歲，佔整體比例 25%。
- 2、 就地區分布來說，一共有 52%的人來自於大台北地區，超過了半數之多。
- 3、 就性別分布來說，男女比例差不多各佔一半（51%:49%）。
- 4、 就婚姻狀況來說，已婚者以 63%佔據大多數比例，單身者僅佔據 37%。
- 5、 就職業分布來說，白領一般職員以 39%的比例遙遙領先其他職業為第一名，第二名則是以白領主管佔據 18%的比例佔據。

3-3 小結

藉由簡單的以一般敘述性人口統計資料，我們可以先初步的認識整份資料的客戶是擁有哪些特質的人佔據多數，以增進我們對整個資料的掌握程度。其中最能看出差別的特徵值分別是地區分布和婚姻狀況，因此我們可以敘述這些客戶為大多來自於大台北地區或是大多是已婚人士(這裡不能用交集的方式來敘述)。

肆、資料分析

4-1、客戶靜態分析

在經歷上述簡單的資料認識和探索以後，我們僅僅只是初步認識認識這群人，但這些人口統計特徵並不能幫助我們去準確地分辨出每一位顧客的價值，因此我們需要有其他指標去評估每一位客戶的價值。

(一)RFM 指標—五等均分法

RFM 指標的 RFM 分別代表了最近購買期間(Recency)、購買頻率(Frequency)以及購買金額(Monetary)。RFM 的給分機制為排序前 20%的客戶得到 5 分，次 20%的客戶獲得 4 分，其餘以此類推(表 4、表 5)。

表 4、各組別之平均 R、F、M 以及給分

組別	平均最近購買期間(天)	平均頻率(次)	平均消費金額(元)	給分
1	2.4	242.55	478446.45	5
2	4.15	82.1	148296.8	4
3	12.85	40.8	75690.85	3
4	45.3	16.9	29720.25	2
5	304.3	5.85	9204.45	1

表 5、RFM 給分示意圖

客戶 ID	R-score	F-score	M-score
89	3	4	4
106	3	4	3
131	1	2	3
179	1	1	1
261	1	2	2
605	5	5	3
773	3	3	3
861	2	4	5
867	3	2	2
920	3	5	5
1250	2	2	3

(二) bob stone—自製權重法

由於 RFM 指標的給分方式是平等的給予 R、F、M 三個指標一樣的權重，然而這並非針對所有產業都適用，有些產業比起購買金額來說，可能更看重的是消費頻率，反之亦然。所以我們需要針對要探討的公司來重新給予每一個指標新的權重(表 6、表 7)。

表 6、自製權重給分機制

指標	給分機制	權重
R	30 天內:24 分 30-60 天:18 分 60-90 天:12 分 90-120 天:6 分 120 天以上:0 分	中
F	購買頻率 * 2 最大值:50 分	高
M	平均金額*0.01(手續費)*0.01 最大值:9 分	低

表 7、自製權重給分示意圖

客戶 ID	R-score	F-score	M-score
89	24	50	9
106	24	50	9
131	0	32	6
179	0	8	0
261	0	26	1
605	24	50	9
773	24	50	5
861	18	50	9
867	24	30	4
920	24	50	9
1250	18	48	9

上述自製給分機制理由為:信用卡刷卡是各家銀行都會有的服務，因此在各家銀行競爭尤為激烈。若要評價顧客價值，應該以頻率為優先量，這代表著忠誠度，但不應該使得權重大到讓另外 2 個指標評估毫無意義，因此設定上限 50。而 Recency 應該作為次要標準，來判斷是否顧客近期不再刷卡，忠誠度流失。而最後，金額的高低與否雖然也會影響銀行利潤，但銀行的主要利潤

來源並非來源於此，因此更重要的是顧客對銀行的忠誠度。

(三)綜合評估

原始資料檔				五等均分法					調整後Bob Stone				
客戶ID	Recency	Frequenc	Money	R	F	M	RFMscore	rank	Rscore	Fscore	Mscore	RFMscore	rank
6118	1	250	968004	5	5	5	15	1	24	50	9	83	1
9051	1	205	265085	5	5	5	15	1	24	50	9	83	1
13687	3	399	1153973	5	5	5	15	1	24	50	9	83	1
15195	3	209	711772	5	5	5	15	1	24	50	9	83	1
17586	1	237	226459	5	5	5	15	1	24	50	9	83	1
4210	4	196	337601	4	5	5	14	6	24	50	9	83	1
5287	4	264	914311	4	5	5	14	6	24	50	9	83	1
6143	1	64	300735	5	4	5	14	6	24	50	9	83	1
6256	4	139	243291	4	5	5	14	6	24	50	9	83	1
15353	1	115	154864	5	5	4	14	6	24	50	9	83	1
17190	5	192	397379	4	5	5	14	6	24	50	9	83	1

圖 6、RFM 和 Bob stone 比較示意圖

如同圖 6 表示一樣，在五等均分法中的前幾位優質客戶到了 bob stone 法中評價時，排名全部被歸類到第一名。其理由是，我們有根據產業特性去修改了給分權重，因此對我們的產業(信用卡)來說，其實前幾名的客戶價值都是最高的，沒必要再去細分。而這樣的做法也更利於我們去找到更有價值的客戶。

(四)小結

比起一般的五等均分法，各家產業應該依照自己產業的特性去自製評分標準，這樣才會使得分數或是指標較為客製化且合理。

此處例子因為是信用卡交易資料，因此以 F 為權重最高，R 次之，M 最低。

4-2、客戶動態分析

經過上述的靜態性分析以後，我們對於顧客的交易頻率、金額還有近期消費時間有了初步簡單的認知和評分。然而，這個樣子依然還不夠我們去完整的了解一個客戶。因此本章節決定要從其他角度——也就是從時間的角度，去重新認知每一位客戶。

(一) CAI 指標(交易活躍性指標)

活躍性指標(Customer Activity Index, CAI)是一種結合購買期間的加權平均和算術平均的新指標。

$$\text{計算公式: } \text{CAI} = (\text{MLE} - \text{WMLE}) / \text{MLE} * 100\%$$

其中，MLE 代表購買期間的算術平均數，WMLE 代表購買期間的加權平均數。

若購買期間的算術平均大致和加權平均相等，則 $\text{CAI} = 0$ ，代表客戶購買行為較為規律。若算術平均數明顯大於加權平均數，即 $\text{CAI} > 0$ ，則代表客戶購買行為越來越活躍。而若算術平均數明顯小於加權平均數時，即 $\text{CAI} < 0$ ，則代表客戶購買行為越來越沉寂。

再計算完每一位客戶的 CAI 指數以後，將其簡單分類為 3 個

群體，前 20%代表漸趨活躍型，後面 20%代表漸趨靜止型，不屬於兩者的則代表屬於中間型，其分群結果如表 8:

表 8、CAI 分群結果示意圖

CAI 群別	客戶人屬	人數比例	消費日平均刷卡金額	平均刷卡間隔天數
漸趨活躍型	20	20%	58753.10	39.79029395
穩定刷卡型	60	60%	123561.55	25.10575074
漸趨靜止型	20	20%	82415.37	48.06005298

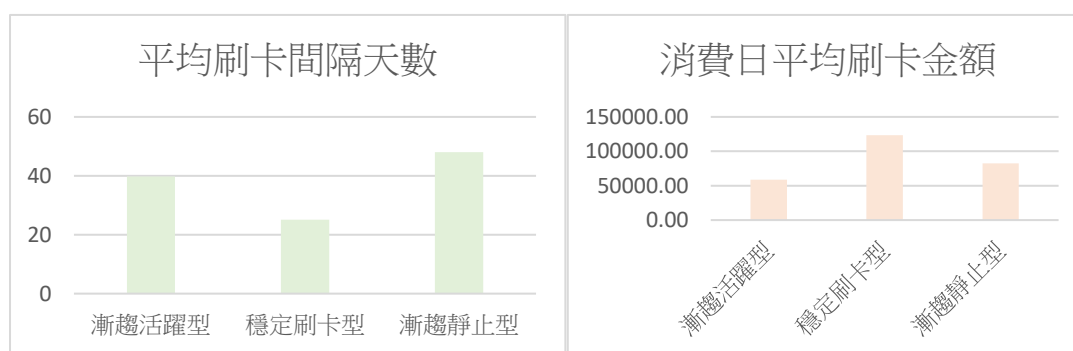


圖 7、刷卡間隔天數

圖 8、刷卡金額

根據上述發現，漸趨活躍型的人他們的刷卡金額低於另外 2 者，可見這個群體應該大部分是屬於刷卡多次但花費比較小的人，而在間格天數的部分，雖然是逐漸活躍，但平均上來說，還是不及穩定刷卡型的人，因此針對這一組漸趨活躍型的人，應該要設計行銷策略讓其成為穩定客戶。

而站在漸趨靜止型的部份，這群人不但越來越不活躍，且刷卡天數還很長，因此算是最容易流失的客戶，縱使金額花費上不算低(第 2 名)，但其忠誠度卻是最低的，因此若要執行行銷策略，應該要設計如何慢慢地提升其忠誠度，使其慢慢活躍起來。

(二) CRI 指標(交易穩定度指標)

交易穩定度指標(Customer Reliability Index, CRI)，是一種拿來衡量客戶交易是否已經形成一個穩定型態的指標。其以貝氏統計為基礎，假設群體呈現常態分配，輔以個人估計和群體估計來校正、加權，以此來形成新的指標。其意涵在於，若 CRI 指數越高，代表其個人交易穩定度越高；若 CRI 指數越低，代表其個人交易穩定度還不穩定。

這次研究簡單以性別(男、女)和婚姻狀況(已婚、未婚)來簡單區分組，一共分出 4 組，其分群結果如下表 9:

分群	平均值 - 刷卡金額
1(男，單身)	2076.706154
2(男，已婚)	1656.062536
3(女，單身)	1426.744349
4(女，已婚)	2688.464823

表 9、CRI 分群結果

依照上述的分群，我們可以建構出每一位客戶的 CRI 指標，且由大到小排列，其結果如下表 10:

客戶 ID	分群	個人平均值刷卡金額	CRI
17027	2	14883.52941	50.27214615
13802	3	6058.6	36.36372101
4819	3	4346.076923	35.44688385
13196	1	10339.84615	33.20968231
5425	1	7600.081967	17.91444229
131	2	4347.375	17.49772157
7293	2	4646.4	14.75147977
8831	4	5572	11.14073519

客戶 ID	分群	個人平均值刷卡金額	CRI
1314	2	7756.794872	10.90456552
16848	3	4335	10.79222427

表 10、CRI 示意圖

分群	計數	CRI 平均
1(男，單身)	18	356.22%
2(男，已婚)	33	434.18%
3(女，單身)	19	469.27%
4(女，已婚)	29	133.43%
總計	99	338.64%

表 11、分群 CRI 示意圖

在把每一位客戶的 CRI 計算出來以後，我們可以依照一開始的分類標準把牠們取平均，就可以得到上述的表 11。表 11 告訴我們，在這 4 組分群裡面，以(女，已婚)這組消費最為穩定，(女，單身)這組消費最不穩定。

(三) 小結

本小節在原有的靜態分析基礎上，引入了動態分析的方法，包含了 CAI 指標以即 CRI 指標，幫助公司區分出漸趨活躍和漸趨靜止型客戶以即最為穩定、不穩定的群組，因此，公司應當針對不同類別客戶採取相對應行銷策略。

4-3、購物籃分析

上述 4-1、4-2 小節皆是針對顧客的特質去做簡單的分析，然而我

們除了可以針對顧客分析以外，商品也是我們不能夠忽視的一環，因此本小節會針對各種被購買的商品進行分析，以此來幫助公司製作出在不同情境下所專屬的購物籃，以利各種行銷活動。

(一) 同時購買購物籃之分析：

在給定 100 位客戶的交易資料，若該客戶買過該商品則標示為 1，沒有買過該商品則標示為 0，藉此可以產生產品購買相關係數矩陣，而藉由觀察這個相關係數矩陣，我們便可以觀察出推薦的商品組合，其結果如圖 9、圖 10。

	學費/教育	保險	捐贈	公用事業	百貨	精品	餐飲	藥妝	3C居家電信	量販超市	交通(含加值)	旅遊	休閒文化	無店舖	中信錢加值
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

圖 9、同時購買相關矩陣

排名	購物籃組合		相關係數
1	精品	14_旅遊	0.472538418
2	休閒文化	16_無店舖	0.418192287
3	精品	10_藥妝	0.413965785
4	旅遊	15_休閒文化	0.400085075
5	餐飲	10_藥妝	0.396376631
6	保險	14_旅遊	0.393311125
7	百貨	15_休閒文化	0.382903601
8	餐飲	11_3C居家電信	0.37858451
9	精品	15_休閒文化	0.367510164
10	藥妝	16_無店舖	0.366244494
11	學費/教育	03_保險	0.357189735
12	交通(含加值)	16_無店舖	0.349132766
13	百貨	09_餐飲	0.346533006
14	餐飲	13_交通(含加值)	0.330595609
15	藥妝	15_休閒文化	0.321718045
16	百貨	10_藥妝	0.320648207
17	餐飲	15_休閒文化	0.319439288
18	交通(含加值)	15_休閒文化	0.318397019
19	精品	09_餐飲	0.315810566
20	3C居家電信	16_無店舖	0.304786945

圖 10、前 20 名購物籃排名推薦

根據圖 9 的圖表，我們可以知道每一個商品與其他商品組合的效果。把其中相關係數最高的組合列出來產生出圖 10，可以看出最高的商品組合是精品和旅遊。因此最適合拿來做商品籃的組合就是精品和旅遊。

(二) 下次購買購物籃之分析：

除了同時購買的組合以外，我們也可以關注這次買了某一種商品，下一次買了另一種商品可能性，以此來建立起商品購物籃組合。依此產生之機率矩陣如圖 10。

	02_學費/教育	03_保險	05_捐贈	06_公用	07_百貨	08_精品	09_餐飲	10_藥妝	11_3C居	12_量販	13_交通	14_旅遊	15_休閒	16_無店舖	X2_中信	X2_中信錢加值
02_學費/教育		0.04%	0.02%	0.00%	0.00%	0.02%	0.00%	0.02%	0.02%	0.06%	0.10%	0.12%	0.02%	0.00%	0.00%	0.40%
03_保險			0.06%	1.10%	0.00%	0.02%	0.13%	0.10%	0.08%	0.13%	0.21%	0.23%	0.37%	0.08%	0.02%	2.62%
05_捐贈				0.00%	0.00%	0.17%	0.02%	0.02%	0.02%	0.00%	0.02%	0.23%	0.00%	0.02%	0.02%	0.54%
06_公用事業					0.00%	0.04%	0.06%	12.88%	0.06%	0.02%	0.02%	0.04%	0.17%	0.75%	0.46%	14.67%
07_百貨						0.00%	0.08%	0.04%	0.02%	2.00%	0.23%	0.75%	0.31%	0.42%	1.39%	7.32%
08_精品							0.00%	0.06%	0.04%	0.02%	0.21%	0.39%	0.27%	0.10%	0.39%	2.50%
09_餐飲								0.00%	0.06%	0.00%	0.02%	0.54%	0.17%	1.06%	0.19%	4.87%
10_藥妝									0.02%	0.10%	0.00%	0.02%	0.31%	0.21%	0.21%	4.41%
11_3C居家電信										0.04%	0.29%	0.02%	0.29%	0.42%	0.29%	7.97%
12_量販超市											0.15%	0.23%	0.00%	0.67%	1.46%	17.67%
13_交通(含加值)												0.06%	0.46%	0.15%	0.60%	25.68%
14_旅遊													0.06%	0.08%	0.00%	3.52%
15_休閒文化														0.08%	0.02%	4.31%
16_無店舖															0.04%	2.43%
X2_中信錢加值																1.08%

圖 11、下次購買之機率矩陣

根據圖 11 所示，這次消費交通相關產品下一次也消費交通的機率是最高的，而這次消費公用事業下次也是消費公用事業的機率是次高的。因此在未來如果有相關行銷計畫時，也可以活用此圖表。

(三) 因素分析：

前述之購物籃分析主要皆是各商品間 1 對 1 間的關係，且都

是只能討論 2 個商品組成 1 個購物籃的效力，若要討論到多個商品組合成 1 個購物籃，就必須要使用到主成分分析，其結果如表 12。

	1	2	3	4	5
08_精品	0.772	-0.089	0.073	-0.093	0.049
15_休閒文化	0.641	0.378	-0.038	-0.019	-0.181
14_旅遊	0.639	-0.061	0.398	-0.059	-0.215
09_餐飲	0.62	0.102	0.06	0.154	0.154
07_百貨	0.6	0.175	-0.111	-0.037	0.18
10_藥妝	0.532	0.337	0.158	0.305	-0.049
11_3C 居家電信	0.491	0.132	0.087	0.133	0.228
05_捐贈	-0.017	0.695	0.111	-0.079	0.048
16_無店舖	0.351	0.67	0.071	0.136	0.017
13_交通(含加值)	0.365	0.405	0.088	0.2	0.095
03_保險	0.221	0.005	0.806	0.095	-0.037
02_學費/教育	-0.105	0.316	0.747	-0.09	0.156
X2.中信錢加值	0.019	0.051	-0.025	0.903	0.012
12_量販超市	0.243	-0.129	0.19	0.216	0.713
06_公用事業	-0.016	0.297	-0.144	-0.315	0.611

表 12、主成分分析

根據上述分析，我們可以得知顧客主要會將信用卡用於 5 種情況(黃底為同一購物籃之產品)，1. 精品旅遊休閒類、2. 捐款或是小額花費、3. 保險或是教育、4. 加值和 5. 量販超市購買及公用事業花費。因此若有相關行銷策略需要搭配可以參考此主成分分析之結果。

(四) 小結：

本小節之目的是在於幫助公司依照過去顧客之消費習慣去訂

定商品的購物籃和行銷策略，因此公司可以依照自己當時所需之情況，自行調整商品購物籃矩陣，依此達成行銷目的，而最後之主成分分析則是可以用一次搜尋多種以上的商品放入同一個購物籃的可能性。因此建議公司在訂定相關行銷策略時，可以多利用此方法。

4-4、客戶集群分析

在上面的幾個小節中，我們運用了各種方法去衡量一群客戶，包含了RFM, CAI, CRI 指標等，因此這一個小節，我們要試圖利用這些指標幫我們的客戶分群，並且試圖從中觀察出一些客戶的特徵，以利公司進行行銷。

(一)K-means 法：

K-means 法的精神是讓每一個分出來的組別差異最大，組間差異最小，而藉由這個方法，能使我們能夠有效地去區別每一個群組。而在此分群中，我們採用的變數包含 Recency, Frequency, Monetary, CAI, CRI。其分群結果如下：

	集群			
	1	2	3	4
R 指標	4	5	5	5
F 指標	171	2644	3820	896
M 指標	9	9	9	9
CRI 指標	3.85531	0.04154	0.00076	0.3323
CAI 指標	0.74845	-3.3242	1.36844	4.30353

表 13、K-means 分群結果

組別	若即若離型 (分群 1)	即將離去型 (分群 2)	財神爺型 (分群 3)	超級新星型 (分群 4)
人數	86 人	1 人	1 人	11 人
人數百分比	86.87%	1.01%	1.01%	11.11%

表 14、各組名稱及人數

根據表 13 各組分群結果，我們依照各組特性幫各組命名，其結果呈現如表 14。而其理由分別為：

若即若離型(分群 1):這組的特色是行為較不穩定(CRI 很大)，有略微活躍的趨勢(CAI>0)。雖然有略微活躍的趨勢，但近期也有一段時間沒有使用信用卡了(R=4)，而且其消費頻率極低(F 較小)。

即將離去型(分群 2):這組的特色是行為蠻穩定，但是已經有一大段時間沒有使用信用卡(CAI 極小)。雖然說近期有使用信用卡(R=5)，但相對於以往(F 高)，頻率沒有那麼高了。

財神爺型(分群 3):這組的行為也是很穩定的，且有越來越活躍的趨勢。除此之外，其使用頻率最高，是指標上各方面都是頂尖的群體。

超級新星型(分群 4):這組的行為是近期突然變得很活躍，雖然還沒有到很穩定，但是是充滿爆發力的群體。

(二)各組別人口敘述統計：

在幫助客戶分完群組以後，我們嘗試要用一般性的人口敘述

統計去描繪出各組別的特徵，而下表 15 為各組特徵：

	組別	若即若離型 (分群 1)	即將離去型 (分組 2)	財神爺型 (分群 3)	超級新星型 (分群 4)	卡方值
性別	男性	84.30%	2.00%	2.00%	11.80%	0.572
	女性	89.60%	0.00%	0.00%	10.40%	
年齡	21-30	91.70%	0.00%	0.00%	8.30%	0.809
	31-40	83.30%	5.60%	0.00%	11.10%	
	41-50	84.00%	0.00%	4.00%	12.00%	
	51-60	89.30%	0.00%	0.00%	10.70%	
	60 以上	87.50%	0.00%	0.00%	12.50%	
教育程度	其他	100.00%	0.00%	0.00%	0.00%	0.96
	高中	84.20%	0.00%	5.30%	10.50%	
	專科	89.50%	0.00%	0.00%	10.50%	
	學士	88.60%	2.30%	0.00%	9.10%	
	研究所以上	83.30%	0.00%	0.00%	16.70%	
	初中及初中 以下	75.00%	0.00%	0.00%	25.00%	
婚姻狀況	單身	89.20%	0.00%	0.00%	10.80%	0.745
	已婚	85.50%	1.60%	1.60%	11.30%	
居住地	大台北地區	86.50%	1.90%	1.90%	9.60%	0.929
	台中地區	94.10%	0.00%	0.00%	5.90%	
	嘉南地區	88.90%	0.00%	0.00%	11.10%	
	高屏地區	81.00%	0.00%	0.00%	19.00%	

表 15、各組人口統計

(三)F 檢定

通過上述檢定我們能夠觀察到各組別人口的狀況，然而我們更好奇的是，是否這樣區分後，有哪一些人真得更有價值，更能夠從這群客戶中獲利，因此我們需要更進一步地去使用 F 檢定來驗證我們的猜測。這裡採用的衡量指標分別是：

1. LN(獲利指標 1)：LN(刷卡總金額)

2. LN(獲利指標 2): LN(刷卡平均金額)

3. LN(獲利指標 3): LN(RFM 指標加總)

而下圖分別代表各組的 F 檢定結果，紅色字體代表有顯著性

差異($P \leq 0.2$)，可以用來說明此分類有助於我們辨別重要客群：

人口統計區隔之獲利指標與F檢定-若即若離型(86人)													
區隔	組別	LN獲利指標1				LN獲利指標2				LN獲利指標3			
		平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)
性別	男性	10.8957	1.3662	43	0.893(0.347)	7.5744	0.94539	43	1.183(0.28)	4.8514	0.9506	43	0.1(0.753)
	女性	10.6026	1.50691	43		7.3659	0.8289	43		4.7839	1.03123	43	
年齡	21-30	10.7159	0.9421	11	0.381(0.822)	7.0283	0.79653	11	1.105(0.36)	5.2157	0.92335	11	1.966(0.108)
	31-40	10.9933	1.19321	15		7.3495	0.89227	15		5.1573	0.9578	15	
	41-50	10.9331	1.78607	21		7.5825	0.97984	21		4.8697	1.11893	21	
	51-60	10.6322	1.47816	25		7.5072	0.86519	25		4.662	0.90514	25	
	60以上	10.4467	1.4498	14		7.7118	0.84075	14		4.3408	0.84684	14	
教育程度	其他	11.5218	-	1	0.3(0.912)	7.8842	-	1	0.158(0.977)	5.1059	-	1	0.346(0.883)
	高中	10.6298	1.62883	16		7.4674	0.86491	16		4.7128	1.01841	16	
	專科	10.8307	1.26838	17		7.4471	0.8292	17		4.92	0.99004	17	
	學士	10.8371	1.41827	39		7.5245	0.98933	39		4.855	1.01887	39	
	研究所以上	10.6139	1.82379	10		7.2851	0.87405	10		4.8315	1.06702	10	
	初中及初中以下	9.9736	0.48814	3		7.3876	0.43621	3		4.1698	0.1166	3	
婚姻狀況	單身	10.9005	1.08181	33	0.59(0.444)	7.4037	0.85013	33	0.296(0.588)	5.0342	0.98645	33	2.629(0.109)
	已婚	10.655	1.62278	53		7.5115	0.91952	53		4.6828	0.97137	53	
居住地	大台北地區	10.8552	1.47394	45	0.34(0.797)	7.5576	0.88712	45	0.451(0.717)	4.8317	0.95584	45	0.16(0.923)
	台中地區	10.7858	1.15005	16		7.4924	0.732	16		4.8494	0.96823	16	
	嘉南地區	10.737	0.97438	8		7.2841	0.77374	8		4.9576	0.77149	8	
	高屏地區	10.4397	1.792	17		7.3051	1.10063	17		4.6846	1.21972	17	

圖 12、F 檢定-若即若離型

人口統計區隔之獲利指標與F檢定-即將離去型(1人)													
區隔	組別	LN獲利指標1				LN獲利指標2				LN獲利指標3			
		平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)
性別	男性	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-
年齡	31-40	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-
教育程度	學士	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-
婚姻狀況	已婚	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-
居住地	大台北地區	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-
職業	白領一般職員	12.2965	-	1	-	5.8028	-	1	-	7.8853	-	1	-

圖 13、F 檢定-即將離去型

人口統計區隔之獲利指標與F檢定-財神爺型(1人)													
區隔	組別	LN獲利指標1				LN獲利指標2				LN獲利指標3			
		平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)
性別	男性	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-
年齡	41-50	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-
教育程度	高中	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-
婚姻狀況	已婚	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-
居住地	大台北地區	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-
職業	藍領(勞動工作或作業員)一般職員	11.4324	-	1	-	4.5706	-	1	-	8.2517	-	1	-

圖 14、F 檢定-財神爺型

人口統計區隔之獲利指標與F檢定-超級新星型(11人)													
區隔	組別	LN獲利指標1				LN獲利指標2				LN獲利指標3			
		平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)	平均數	標準差	樣本數	F值(顯著性)
性別	男性	13.0265	0.59336	6	0.148(0.709)	7.6258	0.51884	6	0.117(0.74)	6.8028	0.1356	6	0.099(0.761)
	女性	12.8625	0.82072	5		7.5164	0.53868	5		6.7499	0.38947	5	
年齡	21-30	11.9882	-	1	0.707(0.616)	6.8465	-	1	0.639(0.654)	6.5482	-	1	0.703(0.618)
	31-40	13.0567	1.02719	2		7.5619	0.98943	2		6.8953	0.03722	2	
	41-50	13.3073	0.74974	3		7.7559	0.5184	3		6.9518	0.37439	3	
	51-60	12.7733	0.10459	3		7.4937	0.12583	3		6.6836	0.02254	3	
	60以上	13.064	0.93615	2		7.8088	0.48256	2		6.6605	0.44532	2	
教育程度	高中	13.1069	0.87548	2	2.47(0.155)	7.6574	0.69663	2	1.032(0.462)	6.8508	0.1762	2	4.498(0.051)
	專科	12.3429	0.50164	2		7.1204	0.38734	2		6.6275	0.11218	2	
	學士	12.857	0.4752	4		7.5205	0.5249	4		6.7396	0.09333	4	
	研究所以上	13.8709	0.12426	2		8.1156	0.20631	2		7.1528	0.32692	2	
婚姻狀況	初中及初中以下	12.402	-	1	1.331(0.278)	7.4675	-	1	2.143(0.155)	6.3456	-	1	0.012(0.914)
	單身	12.6473	0.78531	4		7.2838	0.66809	4		6.7663	0.16718	4	
	已婚	13.126	0.59087	7		7.7431	0.3325	7		6.7859	0.32185	7	
居住地	大台北地區	13.1386	0.85853	5	0.189(0.901)	7.6752	0.62998	5	0.116(0.948)	6.8651	0.31879	5	0.26(0.852)
	台中地區	12.8926	-	1		7.6352	-	1		6.6619	-	1	
	嘉南地區	12.6976	-	1		7.3943	-	1		6.7069	-	1	
	高屏地區	12.797	0.64323	4		7.4828	0.52644	4		6.718	0.27629	4	

圖 15、F 檢定-超級新星型

(四) 事後檢定：

在區分出用哪種人口統計可以幫助我們有效區隔出有價值客群以後，我們更在意的是，是哪一個群組有價值，因此我們需要做事後檢定來幫助我們掌握資料。而以下資料以若即若離型(分群1)為範例：

在圖 11 中我們可以知道用年齡和婚姻狀況都可以幫助我們區分出客群，但其中年齡的區間有 2 個以上，我們因此需要找到其中到底是哪一個區間有顯著差異，下圖為結果呈現，其中紅色字體代表有顯著差異(p 值<0.2)，可以幫助我們去確實區分差異：

ANOVA事後T檢定(若即若離型)-年齡					
應變數	(I) 年齡	(J) 年齡	平均值差異 (I-J)	標準誤	顯著性
LN獲利指標1	21-30	31-40	-0.2774	0.57908	0.633
		41-50	-0.21724	0.54295	0.69
		51-60	0.08367	0.52781	0.874
		60以上	0.26922	0.58776	0.648
		21-30	0.2774	0.57908	0.633
	31-40	41-50	0.06016	0.49316	0.903
		51-60	0.36108	0.47644	0.451
		60以上	0.54662	0.5421	0.316
		21-30	0.21724	0.54295	0.69
	41-50	31-40	-0.06016	0.49316	0.903
		51-60	0.30091	0.43181	0.488
		60以上	0.48646	0.50333	0.337
		21-30	-0.08367	0.52781	0.874
	51-60	31-40	-0.36108	0.47644	0.451
		41-50	-0.30091	0.43181	0.488
		60以上	0.18555	0.48696	0.704
		21-30	-0.26922	0.58776	0.648
	60以上	31-40	-0.54662	0.5421	0.316
		41-50	-0.48646	0.50333	0.337
		51-60	-0.18555	0.48696	0.704
		21-30	-0.3212	0.35243	0.365
LN獲利指標2	21-30	41-50	-0.55424	0.33044	0.097
		51-60	-0.47892	0.32123	0.14
		60以上	-0.68356	0.35771	0.06
		21-30	0.3212	0.35243	0.365
	31-40	41-50	-0.23304	0.30014	0.44
		51-60	-0.15771	0.28996	0.588
		60以上	-0.36235	0.32993	0.275
		21-30	0.55424	0.33044	0.097
	41-50	31-40	0.23304	0.30014	0.44
		51-60	0.07533	0.2628	0.775
		60以上	-0.12931	0.30633	0.674
		21-30	0.47892	0.32123	0.14
	51-60	31-40	0.15771	0.28996	0.588
		41-50	-0.07533	0.2628	0.775
		60以上	-0.20464	0.29636	0.492
		21-30	0.68356	0.35771	0.06
	60以上	31-40	0.36235	0.32993	0.275
		41-50	0.12931	0.30633	0.674
		51-60	0.20464	0.29636	0.492
		31-40	0.05836	0.38298	0.879
LN獲利指標3	21-30	41-50	0.34599	0.35909	0.338
		51-60	0.5537	0.34907	0.117
		60以上	.87490*	0.38872	0.027
		21-30	-0.05836	0.38298	0.879
	31-40	41-50	0.28764	0.32616	0.38
		51-60	0.49534	0.3151	0.12
		60以上	.81654*	0.35852	0.025
		21-30	-0.34599	0.35909	0.338
	41-50	31-40	-0.28764	0.32616	0.38
		51-60	0.20771	0.28558	0.469
		60以上	0.5289	0.33288	0.116
		21-30	-0.5537	0.34907	0.117
	51-60	31-40	-0.49534	0.3151	0.12
		41-50	-0.20771	0.28558	0.469
		60以上	0.32119	0.32205	0.322
		21-30	-.87490*	0.38872	0.027
	60以上	31-40	-.81654*	0.35852	0.025
		41-50	-0.5289	0.33288	0.116
		51-60	-0.32119	0.32205	0.322

圖 16、若即若離型事後 T 檢定

觀察圖 16 的 LN(獲利指標 3)的紅色字體，可以觀察到 21-30 歲的群體和 31-40 歲的群體都和 51-60 歲和 60 歲以上的群體有顯著差異，然後 31-40 歲的群體則是和 60 歲以上的有顯著差異。

搭配圖 12，我們知道在若即若離型這一個分群，若我們要以年齡作為區分的話，我們應該把重心放在經營 21-30 歲和 31-40 歲這 2 個群體上。

(五) 小結：

在這一個章節，我們利用先前所做的數據指標，來幫每一位客戶分到不同群組，爾後我們利用所掌握的人口統計資料來分析是否有哪些敘述是能夠被使用，這樣的過程是可以被運用在任何的行銷策略上，且能夠被數據佐證的。

以上面例子為例，我們得出結論，在若即若離型的顧客中，若以年齡來分類，應該以 21-30 歲和 31-40 歲為重要對象。

伍、結論：

在此次報告中共有 100 為信用卡客戶，在經過各種指標分析後，有效樣本為 99 人，分為 4 群：若即若離型、即將離去型、財神爺型和超級新星型，而各組別皆有不同特徵和行銷策略，以下做簡單的整理匯總：

分群	特色	人數	有價值顧客類性	行銷決策
若即若離型	行為還不穩定，有略顯活躍趨勢，但已經有一段時間沒刷卡了	86	20-40 歲、單身	探明原因，為何已經有一段時間沒有刷卡，並針對問題進行相對應決策
即將離去型	行為穩定但已經很久沒有刷卡	1	-	需探明原因，是否已經完全沒有價值了，若尚有挽留機會，應該嘗試看看
財神爺型	行為穩定且越來越活躍，且各項指標皆很高	1	-	VIP 客戶，可以在不會影響客戶價值的情況下，轉移部份資源到其他群組客戶，或是可以提供專屬服務
超級新星型	行為尚未穩定，不過卻很活躍	86	已婚、學歷研究所以上	要設計行銷策略，使得顧客能夠逐漸穩定下來，如：消費券、會員累積點數…等

表 16、各客群對應行銷策略

除了有針對客群的研究外，還有針對商品形成購物籃的相關研究，下表 17 以主成分分析產生之購物籃為例，建議公司在制定產品相關策略時，可以有相關依據。

分群	購物籃品項
1	精品、休閒文化、旅遊、餐飲、百貨、藥妝、3C 居家電信
2	捐贈、無店鋪、交通(含加值)
3	保險、學費/教育
4	中信錢加值
5	量販超市、公用事業

表 17、各項購物籃推薦品項

總結整份研究之結論，是希望藉由這份報告之研究，在不論是顧客行銷層面或是商品配置層面，都能夠降低公司之營運行銷成本，以達到最佳的行銷成果，以幫助公司成長。