



國立臺灣大學管理學院資訊管理所

碩士論文

Department of Information Management

College of Management

National Taiwan University

Master Thesis

結合產品交易數量與 RFM 模型發展行銷策略

Integrating Product Volume of Transaction Data and RFM
Model to Develop Marketing Strategy

張家綺

Chia-Chi Chang

指導教授：曹承礎 博士

Advisor : Seng-Cho T. Chou, Ph.D.

中華民國 107 年 7 月

July 2018

國立臺灣大學管理學院資訊管理學系



碩士論文

Department of Information Management

College of Management

National Taiwan University

Master Thesis

結合產品交易數量與 RFM 模型發展行銷策略

Integrating Product Volume of Transaction Data and RFM

Model to Develop Marketing Strategy

張家綺

Chia-Chi Chang

指導教授：曹承礎 博士

Advisor: Seng-Cho Chou, Ph.D.

中華民國 107 年 07 月

July 2018

謝辭



還記得論文口試前夕的緊張與不安，以及通過論文口試的那一刻的開心與感動，我的碩士求學生涯至此畫下完美的句點。

本篇論文得以完成，真的需要感謝好多人。感謝台大資管所所有教授們的指導，您們的傾囊相授的教學、活潑生動的上課方式至今仍記憶猶新，每當學習遇到困難時，您們的細心指導與指點迷津，讓我豁然開朗，在此非常感謝您們。

特別感謝指導教授曹承礎教授，非常幸運能跟在您身邊學習，您細心指導，並不厭其煩的指出我研究中的缺失，讓我順利完成論文。也要感謝口試委員盧信銘博士、王貞雅 博士提出的寶貴意見，使我的論文內容更加完整，特此向各位老師獻上我最高的敬意。

謝謝兩年來共同努力的實驗室夥伴們，讓我在研究的旅途上不覺孤單，在課業或研究上有問題時，給予我一些建議與幫助。最重要的，感謝我的家人與男友一路以來在許多方面的支持，讓自己在生活上能夠毫無後顧之憂，您們時時鼓勵著我，並一直默默地付出，才能讓我能順利完成學業。



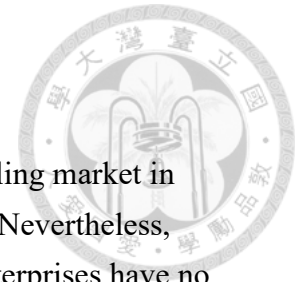
摘要

隨著零售業快速成長，全台灣零售業營業額光是去年 11 月已達 3,698 億元台幣，但現今顧客的喜好越來越多元，企業不可能完全滿足每一個顧客，而這種挑戰在零售業中卻是非常常見的。但隨著資料庫的建立以及演算法的進展，對顧客進行區隔以及分類，能使企業盡力達到滿足每位顧客的目標。顧客區隔可以讓企業了解不同顧客群的行為和喜好，根據顧客的偏好制定行銷策略，更能有效的識別最能帶來收益的關鍵顧客群。

進行顧客區隔時，顧客交易資料是預測未來顧客購買行為的最有力和最可靠的資料，但目前雖然有大量運用顧客交易資料來區隔顧客的研究，卻很少有研究將產品交易數量，也就是產品的銷量考慮在內。因此本研究在進行顧客區隔時，把產品交易數量的概念加進分析，並同時比較無產品交易數量的結果，藉此證明產品交易數量確實能使顧客區隔更加準確、有效率。最後結合加入產品交易數量的分析方式與傳統 RFM 模型將顧客分群，找出真正具有價值的顧客，作為企業在推行促銷組合的參考依據。

關鍵字：顧客區隔、交易資料、產品數量、RFM 分析、促銷組合

Abstract



As the retailing market rapidly grows, the total revenue of retailing market in Taiwan has reached 369.8 billion NT dollars in November last year. Nevertheless, customers nowadays have more various preferences than before. Enterprises have no chance to meet the taste of every individual. This kind of challenge is common among the retailing industry. With the establishment of database and growing of the computing capability, the enterprise can draw near the achievement by customer segmentation and clustering. Enterprises can plan marketing strategy according to preferences and behaviors of customers. As a result, they can identify critical customer groups and bring profit more effectively.

While processing customer segmentation, the transaction record can be the most powerful and reliable data to predict future purchasing potential. Although there are numerous research taking advantage of transaction record to perform customer segmentation, few of them take product volume into account. Accordingly, this research will include the concept product volume when performing customer segmentation. Meanwhile, this research will compare the difference between with and without product volume as an attribute to justify the idea of including product volume enables customer segmentation more accurate. At last, this research will combine customer segmentation with product volume and traditional RFM model to dig out the real valuable customer. The result can feedback to the enterprise as reference to evaluate marketing strategy.

Key words: Customer segmentation 、 Transaction data 、 Product volume 、 RFM analysis 、 Promotion 、 Product set

目錄



第 1 章	研究導論.....	7
1.1	研究背景.....	7
1.2	研究目的.....	9
第 2 章	文獻回顧.....	11
2.1	顧客區隔.....	11
2.2	RFM 模型.....	13
2.3	歐幾里得距離與曼哈頓距離.....	15
第 3 章	研究架構及方法.....	18
3.1	資料集說明.....	18
3.2	研究架構.....	21
3.3	資料前處理.....	23
3.4	研究方法.....	25
3.4.1	顧客相似度定義與計算.....	25
3.4.2	RFM 變數定義.....	31
3.4.3	疊圖分析.....	33
第 4 章	實驗結果.....	35
4.1	有數量及無數量比較.....	35
4.2	疊圖分析結果.....	36
第 5 章	結論.....	43
5.1	管理意涵.....	43
5.2	研究限制.....	43
第 6 章	文獻引用.....	45

圖目錄

圖 1	16
圖 2	16
圖 3	22
圖 4	27
圖 5	28





表目錄

表格 1.....	8
表格 2.....	18
表格 3.....	19
表格 4.....	20
表格 5.....	23
表格 6.....	24
表格 7.....	28
表格 8.....	29
表格 9.....	29
表格 10.....	30
表格 11.....	31
表格 12.....	32
表格 13.....	32
表格 14.....	35
表格 15.....	37
表格 16.....	39
表格 17.....	40
表格 18.....	41
表格 19.....	42

第1章 研究導論

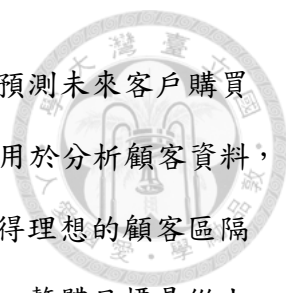


1.1 研究背景

近年來，零售業快速成長，經濟部統計處調查顯示，全台灣零售業營業額光是去年 11 月已達 3,698 億元台幣[1]，Economist Intelligence Unit[2]更預期，全球零售銷售金額會在 2018 年成長 3.4%，且亞洲仍將是零售銷售的主要成長動力，2018 年亞洲/澳大利亞地區預計將成長近 4.9%，進而達到約 10.3 萬億美元的零售額。而隨著現今顧客的喜好多樣化，企業不可能完全滿足每一個顧客。這種挑戰在零售業中是非常常見的，因為零售業提供各式各樣的產品讓顧客選擇購買[3]。然而，如何滿足顧客需求是維持生存及競爭優勢的關鍵[4]，在適當的時機滿足顧客的需求也可以幫助企業建立與顧客之間的長期關係，並提高顧客的回購意圖[5]。

一直以來，顧客關係管理持續受到許多關注，在新的商業經濟中也有良好的成果。顧客關係管理帶來許多潛在的好處，包括增加顧客保留率和忠誠度、提高獲利能力等等。因此，作為顧客關係管理的主要階段，顧客區隔在當今競爭激烈的商業領域是一個越來越緊迫的問題[6]。顧客區隔可以讓企業了解不同顧客群的行為和喜好，根據顧客的偏好制定行銷策略，更能有效的識別最能帶來收益的關鍵顧客群[7]，持續關注這些能帶來收益的顧客，針對這些高價值的顧客群集中企業資源，藉此獲得最大收益，因為顧客價值最大化是企業在激烈的競爭環境中生存的關鍵。有越來越多企業開始意識到他們的顧客資料庫是非常重要的資產[8]，運用資料庫進行顧客特徵分析，進而制定行銷策略[9]，將有限的資源分配給特定的顧客。

顧客區隔的結果在很大程度上取決於輸入變數，傳統上，這些變數可能是人口統計學、心理學、地理學、生活方式等等，其中以人口統計變數區隔為傳統上最常用的方法。然而，當顧客行為資料變得更容易取得時，像是顧客購買的東西、他們喜歡的產品類型、總支出，購買頻率或者他們是否對銷售促銷做出反應，那麼就




有可能實現更精細的顧客區隔。在所有顧客資料中，交易資料是預測未來客戶購買行為的最有力和最可靠的資料[10]。而統計或資料分析方法可以用於分析顧客資料，讓企業能為不同類型的顧客設計不同的行銷策略[11]，為了要獲得理想的顧客區隔結果，資料分析技術非常重要[12]，資料分析是一個跨學科領域，整體目標是從大量資料中找出隱藏的模式、關聯、異常或結構[13]，這些龐大的資料及往往無法透過人工方式在合理時間內達到擷取、管理、處理、並整理成為人類所能解讀的資訊，若透過資料分析技術，能迅速過濾無用資料、分類雜亂的資料，對這些進一步處理過的資料做更進一步的運用。

然而，儘管目前有大量運用顧客交易資料來區隔顧客的研究，卻很少有研究將產品交易數量，也就是產品的銷量考慮在內，根據 Lu and Furukawa [10]的研究，若不考慮產品的交易數量，可能會忽略關於顧客相似度的重要資訊。舉例來說，某公司交易資料庫內有 3 個顧客的資料，且此 3 個顧客各有 1 筆交易，公司只有生產 5 種產品如表格 1，若不考慮數量，顧客 A、顧客 B 及顧客 C 都會被認定是相似的顧客，因為他們都購買了兩個一樣的產品，分別是產品 1 及產品 2。然而，若把產品交易數量同時考慮在內的話，可以發現顧客 A 及顧客 B 都有相似的購買行為，也就是比起產品二，他們都更喜歡購買產品一，因此會認為顧客 A 及顧客 B 較為相似。而顧客 C 則是與前者相反，比起產品一，他較喜歡購買產品二，因此就顧客購買行為來看，顧客 C 與前二者較不相似。

表格 1

	產品一	產品二	產品三	產品四	產品五
顧客 A	2	1	4	0	0
顧客 B	4	2	0	1	0
顧客 C	1	3	0	0	2




為了區隔顧客，學者提出了各種不同的資料分析方法，分群、分類、自組織映射(Self-Organizing Maps，縮寫「SOM」)及人工神經網路(Artificial Neural Network，縮寫「ANN」)等等。其中，分群方法是最常被使用的技術，也是一種區隔顧客的方法，將顧客依照相似的需求及特徵來區隔成不同的群，目標是確保同一群的所有顧客彼此相似，不同群的顧客間彼此不相似[14]。RFM 是分群當中最常被使用於顧客區隔的變數[15]，利用最近一次交易 (Recency，縮寫「R」)、交易頻率 (Frequency，縮寫「F」)及交易金額 (Monetary，縮寫「M」)將顧客的行為量化，進而找出關鍵顧客。雖然近年來 RFM 持續受到如人工神經網路等新技術所帶來的挑戰，但企業依然經常使用 RFM 來區隔顧客，因為使用其他區隔方法所得到的收益，並不能彌補他們花的成本[16]。

因此，本研究將使用交易的產品數量，結合 RFM 模型，提出一個新的促銷組合推薦模型進而優化目前企業行銷策略模式。

1.2 研究目的

企業蒐集了大量有關顧客的資料，但是他們無法發現這些資料中的潛在價值，可能使整個資料蒐集過程毫無意義[17]。目前有越來越多的企業意識到根據顧客交易資料制定商業計劃的重要性，並願意從他們的資料庫中分析出這些未開發的知識，深入了解顧客且藉此提高企業的收益。其中包括分群方法在內的資料分析技術都能使企業能夠利用這種隱藏的知識並帶來價值。

根據顧客交易資料對顧客進行區隔的最基本的假設，在於顧客的交易紀錄可以反映顧客的購買行為。我們可以很直觀的認為，如果兩個交易同時包含大多數相同的產品，則認為這兩個交易彼此相似。因此，計算兩個交易之間共同項目一直是評估兩個交易之間相似度的一般方法。然而，產品的交易數量也是交易紀錄的重要資訊，可以反映顧客的購買行為模式，非常重要且不容忽視。因此，此研究採用產品交易數量以評估交易相似度，進而進行後續顧客區隔。



本研究著重於從顧客過往的消費紀錄中，找到交易與交易、顧客與顧客之間的特徵，藉由計算彼此間的距離來衡量彼此間的相似度，並透過分群疊圖的方式，在市場中找到合適且報酬率高的方案，以改善企業產品銷售計劃的決策。

整體而言，本研究有兩個重要的研究目標，分別為以下：

- (一) 從顧客過往的消費紀錄中，計算交易與交易、顧客與顧客之間的相似度，並在計算時，把產品交易數量的概念加進分析，並同時比較無產品交易數量的結果，藉此證明產品交易數量確實能使顧客區隔更加準確、有效率。
- (二) 結合加入產品交易數量的分析方式與傳統 RFM 模型將顧客分群，找出真正具有價值的顧客，作為企業在推行促銷組合的參考依據。

本研究針對 W 超市的交易資料進行顧客區隔，在模型中增加顧客的產品交易數量用以改善顧客區隔的效率與準確度，並結合 RFM 模型，依照顧客的 RFM 值將之分群，企業可以根據分群結果解釋顧客行為，除了能找出最具價值的顧客並為其推薦適合的商品，更可以做為企業在推行促銷組合時的決策依據，優化企業的行銷效率、降低行銷成本，進而為企業提高收益。

本研究在後續其他部分安排如下，第二章將介紹所有此研究的相關文獻，包括顧客區隔的理念及應用，RFM 的基本運算概念及相關研究，第三章為本研究的架構及研究方法，其中包含了資料集的來源及說明、資料前處理。在第四章，我們將展示此研究的結果，並在最後一章敘述本研究的管理意涵及研究限制。


第2章 文獻回顧



2.1 顧客區隔

顧客的需求、行為模式和喜好大不相同，若要對所有顧客都一視同仁的提供服務，將會消耗龐大企業資源，但若要依照顧客喜好分別提出不同的應對策略，對企業又是一個非常困難且艱鉅的挑戰。顧客區隔的概念最早在 1950 年代中期由一個美國行銷專家 Wendell R. Smith 所提出，是一個依照顧客的相似度將顧客分群的技術，基本概念是，劃分在同一群的顧客具有相似的行為模式[18]。有許多研究也已經顯示出企業如何成功的應用顧客區隔來計劃行銷決策，[19]指出，為了使利潤最大化，企業應該在所有顧客中識別較有價值的客戶，而不是將所有顧客視為平等的服務所有顧客，因為有些顧客的服務成本太高，即使從長遠來看，這些顧客也沒有什麼潛力為企業帶來收益。使用顧客資訊將顧客區隔可以幫助企業識別客戶需求和不同程度的收益，根據這些顧客群的特點開發行銷策略或產品、服務。


儘管區隔的應用已經是行銷文獻和現實中非常重要的話題，但有些企業的管理者卻傾向依賴直覺和基於人口統計的傳統區隔技術，傳統的顧客區隔模型是基於顧客的人口統計、態度和心理學特徵，但他們對當今複雜的商業環境會得出相對太簡單的結果和較差的準確性。例如在航空業，以前認為把商務乘客和經濟乘客分開是一種常識。但是，由於近年來顧客所做的選擇越來越錯綜複雜，這種區隔邏輯的簡單性已經不再適用，只依靠飛行艙等作為區隔標準的航空公司可能無法恰當地制定他們的產品和行銷政策，因此，有必要重新評估這種傳統的區隔標準。T 學者[20]等人使用偏好調查問卷，分析超過 5800 名航空公司乘客的偏好資料，根據不同產品特徵的偏好來確定五個區隔市場，並使用額外的個人態度變數以及社會人口統計變數以更深入的描述這些區隔市場。



除了人口統計變數，顧客心理數據也是一種傳統的顧客區隔變數，Taeho Hong 學者[21]認為，大多研究並未試圖使用資料分析技術，並利用顧客的心理特徵來區隔顧客。這種遺漏的原因是資料分析所需要的心理數據存儲在顧客的腦海中，而不是存儲在統一格式的資料庫中，使得資料取得不易，且分析過程相對困難。他們提出了一種方法，讓線上商店能夠根據顧客的心理數據區隔他們的顧客，從而提供客製化行銷。如果把顧客按照相似消費能力分成幾個群組，那麼線上商店可以依照群組區分顧客價值，進而更專注於獲利能力高的行為，這份研究運用 k-平均演算法(k-means)和自組織映射(Self-Organizing Maps)分群，以便對顧客進行更複雜的區隔。他們將顧客區隔成五個群，並針對每個區隔的顧客都調查了購買意願，還為每個區隔市場計畫並建議了行銷策略。這個方法將有助於線上商店為現有顧客和新顧客提供客製化的行銷策略。最後還應用 k-最近鄰居方法(k-nearest neighbors)對尚未包含在資料集中的顧客進行分類，幫助行銷經理提前識別他們幾乎不知道的所有顧客，並提供差異化的行銷計劃。

然而，在傳統的顧客區隔模式逐漸被淘汰之後，近年通常是使用由各種資訊系統收集的顧客交易和行為資料（例如，交易類型、交易數量和交易歷史等等）而進行的顧客區隔。根據顧客的盈利能力為顧客提供不同的服務已經是很多公司有效的策略。FedEx 以顧客的盈利能力，將顧客分成好的、不好的及更糟的，並針對這三種顧客分別有不同的處理方式，他們試著將資源專注在好的顧客，並努力將不好的顧客提升成為好的顧客，同時也希望減少更糟的顧客存在。將顧客區隔且以不同的方式分別處理，而不是以同樣的方式向所有顧客進行行銷。同樣的情況也可以在其他公司觀察到，如美國排名第六名的銀行 First Union，他們針對不同的顧客用顏色進行區隔。綠色區隔的顧客能為公司帶來收益，而且他們能享有額外的顧客服務，而區隔紅色的顧客將會帶給銀行較多虧損，他們並不能享有任何特殊的待遇[22]。

Z 學者等人更提出一個顧客金字塔的工具，將顧客依照帶來的收益分成四個等級，分別是白金、金、鐵及鉛。白金階層內是該公司最賺錢的顧客，通常是那些對




產品使用量很大，對產品價格不會太敏感，願意投資並嘗試新產品並忠於該公司的顧客。黃金階層與白金階層不同之處在於他們的盈利水準並不高，可能是因為顧客想要擁有較好的產品折扣。黃金階層的顧客是產品類別中的重要，但可能不會忠於公司，會與多家供應商而不僅僅是某特定公司合作來降低風險。鐵階層內的顧客雖然能購買一定數量的產品，但其消費水準、忠誠度和盈利能力皆不足以享有特殊待遇。鉛階層由那些會花費公司資金、製造更多成本的顧客所組成，而且他們有時候是問題顧客，會向別人抱怨公司並且同時佔用公司資源。前兩層的顧客被視為是有價值的，後兩層的顧客對企業則不太有吸引力。企業可以利用此金字塔模型分配企業資源，開發客製化的產品及服務，從不同的顧客等級中獲得更大的收益[22]。利用顧客資訊來區隔顧客，將企業資源投入在對公司有較高潛在價值顧客，並減少投資在較無價值的顧客，將顧客的目前價值與潛在價值結合可以改善顧客金字塔模型，例如，企業可以鎖定目前價值較低但未來潛在價值較高的顧客[23]。

在所有顧客資料中，交易資料是預測未來顧客購買行為的最有力和最可靠的資料，然而，儘管目前有大量運用顧客交易資料來區隔顧客的研究，卻很少有研究將產品交易數量，也就是產品的銷量考慮在內，根據 Lu and Furukawa[10]的研究，若不考慮產品的交易數量，可能會忽略關於顧客相似度的重要資訊。但是該學者在其研究內僅僅利用一組虛擬的範例資料來說明他的理念，而不是透過現實世界的真實資料，不確定在真實資料上的表現是否會如同範例資料一樣，加入產品交易數量概念的分析確實會比較好，因此本研究將運用真實超市的資料，試圖驗證這個概念。


2.2 RFM 模型

近年來，雖然顧客交易資訊越來越容易取得，讓企業可以開發更多、更複雜的顧客區隔方法，但越是如此，方法的簡單性也越是重要。多年來，目錄業者(意旨



只有產品型錄而無門市的業者)和其他直接行銷人員已經使用 RFM 分析來區隔他們的顧客群，並優化其行銷工作的購買回應率。一次又一次地，RFM 受到像是神經網路等新技術所帶來的創新概念方法的挑戰。然而，直接行銷人員仍然依賴 RFM，因為他們發現，如果使用其他替代方法，通常不能保證實施這些方法將會為企業帶來多少成本，例如，因技術複雜度增加的成本，特別是從行銷人員手中分析之後再將其分配到程式設計師和統計人員的手中。除了因技術複雜度增加所帶來的額外成本之外，解釋方法及互相溝通的成本也同樣重要，行銷人員需要根據研究結果制定可行的策略和決策。

RFM 模型最初是在[24]被提出，用來分析和預測顧客行為。基本的 RFM 模型由最近一次交易 (Recency, 縮寫「R」)、交易頻率 (Frequency, 縮寫「F」)及交易金額 (Monetary, 縮寫「M」)所組成。此三個變數分別被分隔成五個相等的五分位數，組合成一個三位數的代碼，最後會區隔成 125 類。最近一次交易是指自上次交易以來的時間間隔，通常以天數或月份數表示，時間間隔越低，此項分數越高，分數前 20% 的部分代碼為 5，而下一個 20% 的部分代碼為 4 等等依此類推。交易頻率是指在一定時間間隔內的交易次數，最高的五分位數代碼為 5，下一個 20% 的部分代碼為 4 等等依此類推。交易金額則是在一定時間間隔內的交易金額，如同前二者一般，交易金額分數前 20% 的部分代碼為 5，而下一個 20% 的部分代碼為 4 等等依此類推。然而，這些變數的定義可以根據問題的面向而有所改變[25]，例如[16]認為，交易金額的定義應該使用平均交易金額，而不是總累計交易金額，藉此降低交易頻率和交易金額的共線性。最後，所有的顧客由 555、554、553、...、111 代碼區隔，總共有 125 個 (5×5×5) RFM 組合，最好的 RFM 代碼組合是 555，早期的研究顯示，R、F 和 M 值較大的顧客更有可能進行下一次交易[26]。而最差的 RFM 代碼組合則是 111，再根據這些不同的 RFM 代碼組合將顧客分成不同的區隔，並進一步分析他們能帶來的收益[27]。



RFM 模型有幾個優點，例如模型能被快速的實作應用[28]，使用較少的變數來獲得顧客特徵[29]，模型的結果容易被解釋等等[16]，因此近年來，RFM 模型被廣泛使用於銀行和保險業[30]、政府機構[31]、旅遊行業[32]、健康與美容業[33]等等。也有研究延伸了傳統的 RFM 模型，包括增加額外的變數，例如[34]增加了關係長度(Relationship Length)變數，並將關係長度定義為最後一筆交易與該顧客第一筆交易之間的時間間隔。在[3]研究中，除了上述的關係長度之外，學者另外增加了週期(Periodicity)變數，並將其定義為顧客連續兩次交易的時間間隔的標準差。而[35]則增加了忠誠度(Loyalty)變數，但實際上如何獲得忠誠度變數並沒有詳細說明。在[36]研究中，學者增加了另外五個變數，分別是忠誠度(Loyalty)、平均年需求(Average Annual Demand)、長期關係潛力(Long Term Relationship Potential)、年銷售收益的平均變化百分比(Average Percentage Change in Annual Sales Revenue)、年需求的平均變化百分比(Average Percentage Change in Annual Demand)，研究結果顯示提出的方法可以有效地用於企業的顧客區隔的應用。然而，除了使用額外變數來延伸傳統 RFM 模型之外，也有研究透過減少變數來改良 RFM 模型，Marcus [16]僅使用了其中兩個變數，分別是交易頻率(F)及交易金額(M)，利用波士頓矩陣(Boston Consulting Group Matrix，縮寫「BCG Matrix」)的概念，將 F、M 劃分成顧客價值矩陣(Customer Value Matrix)來區隔顧客，並針對各個區隔提供不同的因應策略。

2.3 歐幾里得距離與曼哈頓距離

顧客的交易紀錄由其購買的產品及產品數量所組成，由於假設每個產品獨立於其他產品，也就是產品之間互相獨立，因此可以將每筆交易紀錄視為一個向量，即在 n 維歐氏空間中的一個點，其中 n 是產品的總數，即把顧客交易紀錄視為具有相同維度的向量。因此，有很多幾何距離可以用來計算、比較兩筆交易紀錄之間的距離。距離的條件及定義如以下：



給定一個集合 S ，笛卡爾積 $S \times S$ 上的實值函數 $d(x, y)$ 是任意 $x, y \in S$ 的距離，

$d(x, y)$ 滿足以下條件：

- (一) $d(x, y) \geq 0$
- (二) $d(x, y) = d(y, x)$
- (三) $d(x, y) = 0 \Leftrightarrow x = y$

在 Lu and Furukawa[10] 研究內用採用的是歐幾里得距離，而歐幾里得距離與曼哈頓距離都是用於計算向量之間距離的既簡單又直覺的方法，因此本研究分別以歐幾里得距離及曼哈頓距離做為計算兩交易之間距離的方法。首先，歐幾里得距離是歐幾里得空間中兩點間「普通」（即直線）距離，定義如以下圖 1：

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

圖 1

(其中 x 代表第一筆交易， y 代表第二筆交易， n 代表資料庫內交易產品總數， $i=1, 2, \dots, n$ ，代表為第 i 個產品)

另外，曼哈頓距離為在歐幾里得空間的固定直角坐標系上兩點所形成的線段對軸產生的投影的距離總和，也可以說是兩個點上在標準坐標系上的絕對軸距之總和，命名原因是從規劃為方型建築區塊的城市（如曼哈頓）間，最短的行車路徑而來（忽略曼哈頓的單向車道以及只存在於 3、14 大道的斜向車道），任何往東三區塊、往北六區塊的的路徑一定最少要走九區塊，沒有其他捷徑。曼哈頓距離定義如下圖 2：

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|$$

圖 2

(其中 x 代表第一筆交易， y 代表第二筆交易， n 代表資料庫內交易產品總數， $i=1, 2, \dots, n$ ，代表為第 i 個產品)

本研究同時採納上述兩種距離演算法，並分別做為計算兩交易之間距離的方法，在分析結果時發現，若在本研究使用歐幾里得距離將會導致資料失真，因而在後續實驗中，僅使用較穩定且準確的曼哈頓距離演算法，詳細的比較方法與其各自的結果將在第四章說明。

第3章 研究架構及方法



3.1 資料集說明

在這項研究中的零售業者是台灣一家 W 超市，在台灣有近 260 家門市，原始數據是從公司的 POS 交易資料中擷取，交易資料中，交易時間從 2017 年 01 月 26 日至 2017 年 06 月 01 日，共得到約四個月的資料。

在資料集中，共有部門檔、產品大類檔、產品中類檔、產品小類檔、產品檔、促銷活動檔、滿額送現金卷、信用卡銷售明細、遊戲卡銷售明細、銷售明細、銷售付款明細、促銷折扣明細、交易檔、促銷活動加價購商品等等共計 24 張資料表如表格 2。且在 24 張資料表中，本研究主要使用其中「銷售明細」資料表及「信用卡銷售明細」資料表。

表格 2

表格名稱	表格內容
MF_DEPARTMENTS	部門檔
MF_PRODUCT_CATEGORIES	大類檔
MF_PRODUCT_DISSECTION	收銀大類檔
MF_PRODUCT_SECTIONS	小類檔
MF_PRODUCT_SUBCATEGORIE	中類檔
MF_PRODUCTS	產品檔
MF_PROMO_EVENT_HEADERS	促銷活動主檔
POS_CASH_COUPONS	現金券報表使用
POS_COUPON_ISSUES	滿額送現金券
POS_CREDIT_CARDS	信用卡銷售明細
POS_GAME_CARDS	遊戲卡銷售明細

POS_ITEM_CHANGES	立即換-產品
POS_ITEM_SALES	銷售明細
POS_MOBILE_PAYMENTS	行動支付記錄檔
POS_PAYMENT_MEDIAS	銷售付款明細
POS_PAYMENT_SODEXO	sodexo 禮券記錄檔
POS_PROMO_DISCOUNTS	促銷折扣明細
POS_SALES_TRANSACTIONS	銷售小計
POS_SALES_UUPON	交易檔-UUPON
POS_STAFF_DISCOUNTS	員工折扣明細
MF_PROMO_THEMES	促銷活動-主題檔
MF_PROMO_PRODUCTS	促銷活動-商品檔

在「銷售明細」中，每筆銷售記錄包含營業日期、交易時間、分店代號、收銀機台編號、收銀員代號、收銀員姓名、銷售日期、銷售時間、交易序號、產品編號等等共計 40 個屬性，所有較重要之屬性列出如表格 3。

表格 3

屬性	屬性描述	屬性	屬性描述
business_date	營業日期	category_no	產品大類
store_no	分店代號	subcategory_no	產品中類
register_no	收銀機台	normal_price	原售價
cashier_id	收銀員代號	transaction_quantity	銷售數量
cashier_name	收銀員姓名	transaction_amount	銷售金額
transaction_date	銷售日期	discount_amount	折扣金額
transaction_time	銷售時間	tax_report_date	大平台申報日期


transaction_type	交易類別	product_name	產品品名
transaction_no	交易序號	product_id	產品 ID
dissection_no	收銀大類	product_no	產品編號

在「信用卡銷售明細」中，每筆資料包含營業日期、交易時間、分店代號、收銀機台編號、收銀員代號、收銀員姓名、銷售日期、銷售時間、交易序號、信用卡別、刷卡金額、信用卡號、發票號碼等等共計 31 個屬性。所有屬性列出如表格 4。

表格 4

屬性	屬性描述	屬性	屬性描述
business_date	營業日期	media_type2	信用卡別
store_no	分店代號	approval_code	授權碼
register_no	收銀機台	credit_amount	刷卡金額
cashier_id	收銀員代號	credit_card_no	信用卡號
cashier_name	收銀員姓名	invoice_no	發票號碼 10 碼
transaction_date	銷售日期	invoice_type	發票類型
transaction_time	銷售時間	invoice_amount	發票金額/ 折讓發票金額
transaction_type	交易類別	create_date	建檔日期
transaction_no	交易序號	create_user_no	建檔人員
media_type	付款類別	file_name	檔案名稱

然而，由於該 W 超市在資料取得期間尚未建置顧客會員系統，因此不能透過顧客姓名、會員編號等等來分辨顧客，所以，為了能從龐大交易資料中分辨每一筆交易資料中的顧客，得知特定顧客的交易紀錄，本研究將結合該資料集內的兩張資



料表，分別是「銷售明細」資料表及「信用卡銷售明細」資料表，利用顧客消費時所使用的信用卡卡號來辨別顧客，而非使用會員編號等等。本研究利用此兩張資料表中相對應的屬性當作 KEY，找到在「銷售明細」資料表中的某交易資料在「信用卡銷售明細」資料表中實際消費使用的信用卡卡號，進而代替會員編號來辨別每一位顧客，而這些相對應的屬性為營業日期、分店代號、收銀機台、交易序號，必須透過此 4 個屬性，方能代表每一筆交易。

最後，此資料集內共有 260 家分店的資料，本研究使用其中交易筆數最多的師大門市資料，在 2017 年 01 月 26 日至 2017 年 06 月 01 日合計約 4 個月內，總交易筆數為 634894 筆，然而，因 W 超市無會員系統，故無法計算這些交易資料其中的顧客總人數。為了辨別顧客，此研究僅採納結帳方式為信用卡之交易，在合併「銷售明細」資料表及「信用卡銷售明細」資料表後，交易總筆數減少為約 13 萬筆，顧客總數為 3304 名，交易產品種類總共近 10000 種商品。

3.2 研究架構

本研究著重於從顧客過往的消費紀錄中，找到交易與交易、顧客與顧客之間的特徵，藉由計算彼此間的距離來衡量彼此間的相似度，並透過分群疊圖的方式，在市場中找到合適且報酬率高的方案，以改善企業產品銷售計劃的決策。然而，過往大部分的文獻及研究中，往往省略了交易紀錄中產品交易數量的概念，因此本研究將採納產品交易數量的概念，分析台灣 W 超市的顧客交易資料，目的在於發掘顧客潛力，並且替企業提供理想促銷方案的篩選對策。

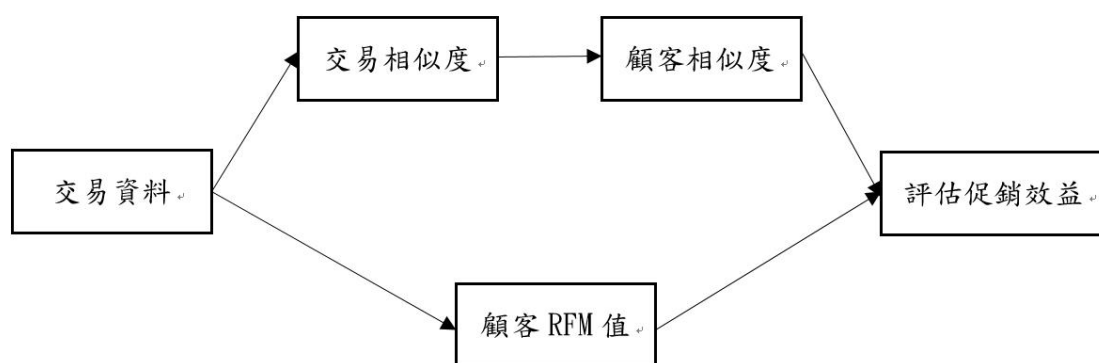
此研究方法分三步驟進行如圖 3，第一步驟為計算顧客間的相似度。先將每一筆交易轉換成向量，而向量的維度、方向、投影量皆由每一筆交易內容所組成，而顧客的每個交易行為由其每次交易紀錄所組成的向量所定義。將每筆交易轉為向量之後，接著計算每一筆交易向量間的曼哈頓距離，最後透過顧客間每筆交易之相

似度計算出顧客間的相似度，計算方法為，將顧客間每筆交易向量距離全部加總後取平均，得到的數值便是兩個顧客在我們定義之下的交易相似度。而在評估相似度時，本研究使用了兩種方法進行比較，其一為有參考各個產品交易數量的結果，另一則無參考產品交易數量。在這個階段，本研究會定義一個「虛擬顧客」來篩選出企業欲行銷的目標產品組合，將此虛擬顧客加進原有的顧客群之後，重新計算顧客相似度。

第二步，利用 RFM 分析法，透過分析顧客過往的交易紀錄，區分出顧客的最近一次交易 (Recency, 縮寫「R」)、交易頻率 (Frequency, 縮寫「F」) 及交易金額 (Monetary, 縮寫「M」)，並依照顧客總體區間來評分，由這三個屬性分別區分出五個等級，滿足各個屬性的前五分之一的顧客分數為 5，後五分之一的分數為 1，以此類推。再將顧客的 RFM 值相加，便得到的是此顧客最終的 RFM 值，也是此研究用來評估顧客價值的依據。

最後一步，由上兩步得到的結果，我們將欲評估的行銷組合視為一名虛擬顧客，且此虛擬顧客僅有一筆交易，加入顧客交易資料中，重複操作第一步中的交易相似度作法，並取與虛擬顧客一定距離以內的顧客來加總其 RFM 值，平均下來的結果可以視為此虛擬顧客，也就是欲評估的行銷組合所吸引的顧客與其對於公司的顧客價值，此值越高，意味著此行銷組合能為公司帶來越大的成效。

反覆進行上面的步驟，企業就能在眾多產品組合中，濾出有效並能導出高價值的顧客需求的理想行銷組合。





3.3 資料前處理

本研究使用其中交易筆數最多的師大門市資料，分店代號為 36，在 2017 年 01 月 26 日至 2017 年 06 月 01 日合計 4 個月內，總交易筆數為 634894 筆。

本研究使用的資料合併了「銷售明細」資料表及「信用卡銷售明細」資料表，利用此兩張資料表中相對應的屬性，分別是營業日期、分店代號、收銀機台、交易序號，當作串接兩張資料表的 KEY，對應在「銷售明細」資料表中的某筆交易資料在「信用卡銷售明細」資料表中實際消費使用的信用卡卡號，進而代替會員編號來辨別每一位顧客，必須透過此 4 個屬性，方能代表每一筆交易。而根據研究目的，本研究在 61 個不同的屬性中選擇其中 8 個屬性進行分析，分別是具有 KEY 作用的營業日期、分店代號、收銀機台編號、交易序號，及信用卡卡號、產品編號、產品品名及產品交易數量。在接下來的篇幅中，統稱此兩張資料表形成的新資料表稱為總表。

總表內的交易紀錄，每一筆交易紀錄代表著每筆交易內的每項產品，如某顧客在當次交易購買了 2 項產品，則會產生 2 筆交易紀錄如表格 5，其中每筆交易紀錄的營業日期、分店代號、收銀機台、交易序號、信用卡卡號都相同，在產品編號、產品品名及產品交易數量則紀錄該交易內該產品的資訊。

表格 5

營業日期	分店代號	收銀機台編號	交易序號	信用卡卡號	產品編號	產品品名	產品交易數量
02-2 月- 17	36	2	9064	5181540*****	681015	白蘿蔔	2

02-2 月- 17	36	2	9064	5181540*****	151050	白吐司	1
------------------	----	---	------	--------------	--------	-----	---

總表內共計總交易筆數為 130674 筆，顧客總數為 3304 名，交易產品種類總共近 10000 種產品，本節將總表資料分別針對交易、顧客、產品進行四步驟資料前處理。

首先，總表內的交易紀錄如同上述，每一筆交易紀錄代表著每筆交易內的每項產品，若某顧客在當次交易購買了 2 項產品，則會有 2 筆交易紀錄，本研究先將總表的交易紀錄，透過相同的營業日期、分店代號、收銀機台及交易序號，整理成每一筆交易紀錄為當次交易內的所有交易產品紀錄，例如，某顧客在當次交易購買了 2 項產品，則只會有一筆交易紀錄，記錄著此兩項交易產品的資訊如表格 6，經過整理後，交易紀錄從 130674 筆降為 31783 筆。

表格 6

營業日期	分店代號	收銀機台編號	交易序號	信用卡卡號	產品編號	產品品名	產品交易數量
02-2 月-17	36	2	9064	5181540*****	681015, 151050	白蘿蔔, 白吐司	2,1

第二步，根據資料顯示，顧客在四個月內的平均交易筆數為 9.6 筆，因此本研究只採用在 2017 年 01 月 26 日至 2017 年 06 月 01 日期間交易筆數超過 10 次的顧客，經過篩選之後，顧客總數從原本的 3304 名，降為 2617 名，而交易筆數從 31783 筆降為 25119 筆。

第三，由於本研究目的是希望提供企業在產品組合決策上之考量，產品組合為兩項(含)以上的產品組合而成，因此，本研究只採納每一筆交易紀錄中，交易產品

種類大於 1 的產品，也就是不採計只購買一種產品的交易，此步驟將交易紀錄筆數從原本的 25119 筆，降為 19390 筆，顧客由 2617 名降為 685 名。

最後，在總表內，所有顧客曾經購買的產品種類數為 9555 種，若全部的產品都納入分析的話，在計算過程將會耗費大量時間，且不能專注在對本研究目標有實際效用的產品們，因此我們認為應該針對產品進行篩選。

與傳統的篩選方法不同，本研究將以產品交易總數量為篩選條件，而非交易內該產品購買比例。我們認為，企業會更在意那些在總銷售量中佔有較大比例的產品，而不是那些僅僅被購買數次的產品，如此一來，將能為公司決策者在決策時帶來更多幫助，只考慮那些產品交易總數較大且貢獻巨額利潤的產品，可以幫助公司更有效率的增加利潤。因此，原先產品種類數為 9555 種，在篩選交易總次數大於 20 的條件之後，將降為 1295 種。然而，由於篩選產品種類之後，會造成顧客交易紀錄內無任何產品(所有該顧客購買的產品的交易總次數皆小於 20)，因此，再次刪除交易內無任何產品的交易紀錄之後，交易紀錄由原先的 19390 筆降為 18273 筆。

綜合上述四個前處理步驟之後，總表的資料為交易紀錄為 18273 筆，顧客為 685 名，產品種類為 1295 種。以下所有分析都將使用此表。

3.4 研究方法

3.4.1 顧客相似度定義與計算

本研究中計算顧客間相似度時，將採用產品的購買比例，而不是實際購買數量。而此作法的原因在於購買行為的解讀。舉例來說，顧客 A 在某一次的消費紀錄中，購買了蘋果 1 顆與香蕉 3 根，顧客 B 購買了蘋果 2 顆與香蕉 6 根，雖然兩顧客間的交易有著數量上的差異，但仔細觀察顧客 B 的產品購買量皆為顧客 A 的兩倍，表示這兩個顧客有著類似的購買行為，可能只是因為家庭人數而帶來的差別。

對於每一筆交易，我們將每一個產品都視為一個獨立的維度，維度之間的值並不會彼此互相影響。而對於每個交易向量在該產品維度的投影量，本研究在此節將會分兩方面來探討，釐清有無考慮產品交易數量對於計算出的結果是否有影響。

- (一)將任一產品在該維度的長度定義為其購買數目除以總購買產品數目，得出的結果就相當於此產品在此次交易中佔總消費數的比例，而所有產品比例加總為 1。
- (二)去除產品交易數量的影響，該產品不論被購買了多少個，都只視為 1，每一個維度的投影長度為 1 除以購買產品種類數，交易中每一個產品種類不論其數量，都對結果產生同樣貢獻。

上述這兩種方式的差別在於是否能夠有效區分對於類似產品裡，顧客不同的偏好。在這裡舉兩個替代性需求的產品為例，如牛奶與優酪乳，人們容易將這兩樣產品彼此替代，因而彼此具有競爭關係。在不考慮消費數量的情況下，如果兩筆交易中皆含有這兩樣產品，我們將無法區分顧客對這兩樣產品的偏好。而在加進產品交易數量這個屬性時，我們可以輕易判斷顧客不同的偏好，且有利於資料演算的進行，進而推斷這個顧客在這兩項替代性高的產品下，他較為偏好何者。這個額外的資訊將大大提高公司對於顧客消費行為的理解與評斷，甚至可以作進一步的消費行為規劃。

本研究使用交易之間的距離作為交易相似度，顧客之間的距離作為顧客相似度。在定義出每筆交易的向量之後，我們接著要計算向量與向量之間的距離，最後再利用交易之間的相似度計算顧客間的相似度如圖 4，若顧客 A 有三筆交易紀錄 A1、A2、A3，而虛擬顧客有 1 筆交易紀錄，則顧客 A 與虛擬顧客之顧客距離為顧客 A 及虛擬顧客這 3 段交易距離的平均。

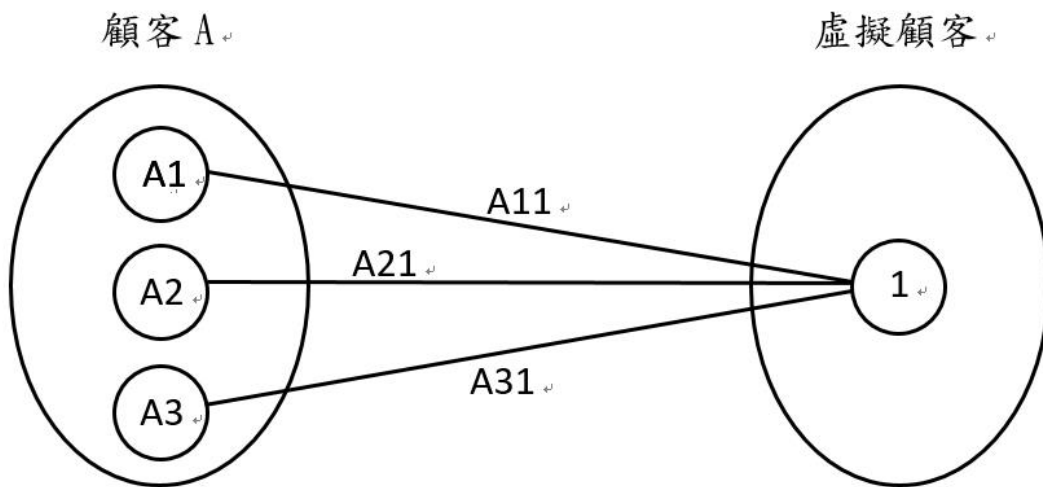


圖 4

在此次研究中的距離計算方法，我們採用曼哈頓距離，而非一般常見的歐幾里得距離。如同本研究第二章提到，曼哈頓距離的定義為每次的移動只能往一個獨立的維度移動，所以當計算兩個不同向量的距離時，我們必須沿著每個產品的維度移動，每次移動距離為同一座標軸相減的絕對值，而總移動距離才是兩筆交易之間的距離。曼哈頓距離可以不論路徑、不論移動次數，都在兩個點之間得到相同距離。如圖 5，在二維情況下，實線距離為曼哈頓距離。

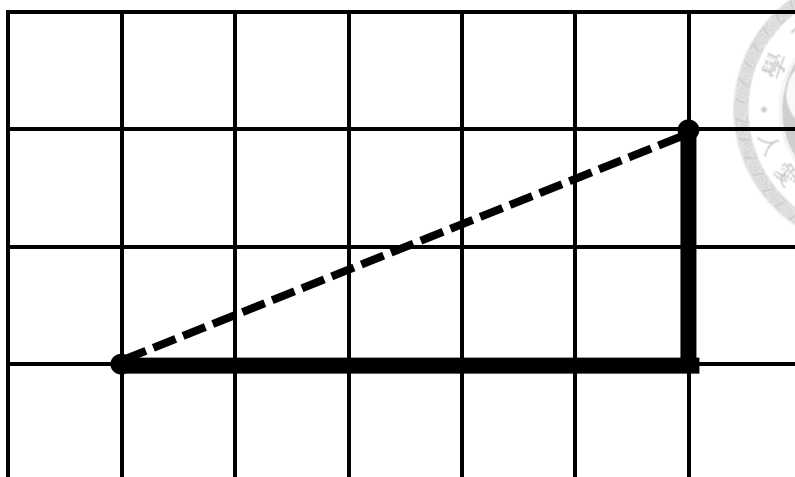
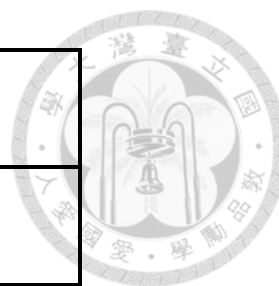


圖 5

而歐幾里得距離是將每個維度的差值取平方相加後，再取根號就可以得到兩個點的直線距離。而歐幾里得距離在本次實驗中會因為平方運算而造成資料失真。對於兩筆交易資料而言，若與相同的虛擬交易組合比較，購買越多品項的交易組合，歐幾里得距離就有一個比較低的起始點。相對於所有購買產品數較多的交易紀錄，交易之間距離就會有一個平移的趨勢，因此交易距離便無法真正反映出交易之間彼此的相似度，而是相當於利用一個受到處理而標準不一的資料進行比較。如圖 5，在二維情況下，虛線距離為歐幾里得距離。

以下我們以表格 7 當作例子來解釋兩種距離計算方法所產生的差異。假設顧客 A、B、C，分別具有如表格 7 的交易紀錄，且顧客 B、C 與顧客 A 所消費的產品均無重疊。

表格 7

	產品一	產品二	產品三	產品四	產品五	產品六	產品七
顧客 A	1	1	0	0	0	0	0
顧客 B	0	0	1	1	0	0	0
顧客 C	0	0	0	0	1	1	1

這些交易紀錄經由本篇研究先前提到的購買產品量占總消費數量比例的運算後，將結果列在表格 8。

表格 8

	產品一	產品二	產品三	產品四	產品五	產品六	產品七
顧客 A	0.5	0.5	0	0	0	0	0
顧客 B	0	0	0.5	0.5	0	0	0
顧客 C	0	0	0	0	0.33	0.33	0.33

表格 9 及表格 10 分別是將交易紀錄取歐幾里得距離以及曼哈頓距離之後的結果。由於顧客 B、C 與顧客 A 所消費的產品均無重複，我們希望的結果是顧客 B 與 A 的距離要跟顧客 C 與 A 的距離一樣，因為他們之間購買的產品都不一樣。根據表格 9，利用歐幾里得距離方法算出的顧客距離，顧客 B 與顧客 A 的距離和顧客 C 與顧客 A 的距離會是不同的，而且顧客 A 距離顧客 C 會較顧客 B 接近。這是因為歐幾里得距離本身運算造成的資料失真。歐幾里得距離運算上會先取平方個別相加後再取根號值。當一位顧客的購買產品種類越多的時候，其各個產品所佔的比例就會下降，也就是根號內的每一項都會是比較小的數，而比較小的數字經由平方相加在開根號後，得到的值會同樣也是會偏小的。這樣下來，即使像我們舉的例子中的購買產品種類完全不同的幾位顧客，也會因為某些顧客購買的種類較多而得到相對較小的距離，甚至，在購買數量較多且無重疊的交易距離可能比有重疊但購買數量較少的距離更近。

表格 9

	顧客 A	顧客 B	顧客 C
顧客 A	0	1	0.847
顧客 B	1	0	0.847
顧客 C	0.847	0.847	0



相對的，對於曼哈頓距離，就不存在這樣的問題，若交易紀錄中的產品品項完全無重疊時，經由曼哈頓演算法得到的距離會完全一模一樣。不會使我們的演算法在計算距離時因為距離本身的定義而失真，見表格 10。

表格 10

	顧客 A	顧客 B	顧客 C
顧客 A	0	2	2
顧客 B	2	0	2
顧客 C	2	2	0

另外，由於此研究方法可透過交易紀錄的資料庫，同時定義出消費者的消費習慣，更可以找出其他類似行為的消費者族群。因此，我們可以定義一個「虛擬顧客」進而來篩選出我們欲行銷的目標產品組合。在此研究中，我們制訂一個虛擬顧客，該虛擬顧客具有下列特性來幫助演算法從顧客消費紀錄中選取我們有興趣的目標顧客以及產品組合，該虛擬顧客的定義如下：

- (一)顧客只擁有這唯一 1 筆交易，且只具有目標產品組合的消費紀錄，無其他額外產品消費，避免被其他產品影響。
- (二)產品交易數量吻合目標產品組合設定比例，本研究中導入了產品交易數量概念，幫助我們評估顧客的偏好。

在這次研究中，因為資料集的資料主要由交易紀錄、地點、時間、卡號等所組成，並未其他個人資料的屬性，如果應用到更廣泛的資料集，則在虛擬顧客的定義可以依照需求增加更多有興趣的屬性，增加除了交易相似性之外更多的資訊，讓分析結果更豐富。



3.4.2 RFM 變數定義

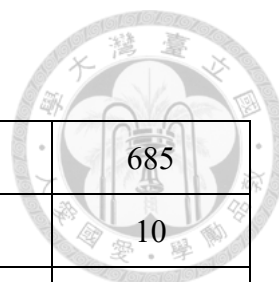
本研究使用的 RFM 模型由三個變數所組成，分別是最近一次交易 (Recency，縮寫「R」)、交易頻率 (Frequency，縮寫「F」)及交易金額 (Monetary，縮寫「M」)。本研究將上述三個變數分別按照降冪排序，將之各分隔成 5 個相等的 5 分位數，第一個五分位數最好，得分為 5，第五個五分位數最差，得分為 1 分。本研究使用的資料集經資料前處理過後，顧客總數為 685 名，因此各五分位數都會包含 137 名顧客。將所有顧客的 R、F、M 分數都算出來之後，將可以得到每個顧客的 RFM 總分。

首先針對最近一次交易，也就是變數 R 進行處理。最近一次交易是指自上次交易距離資料蒐集截止日 2017/06/01 的時間間隔，通常以天數或月份數表示，最近一次交易的時間距離資料截止日越近，則此項分數越高，將日期由後至前排序前 20%，也就是時間間隔小於排序第 137 顧客的 2017/05/31，分數為 5，而下一個 20% 區隔的分數為 4 等等依此類推，在此將各五分位數的分隔點，也就是第 137,274, 411, 548, 685 個顧客列於表格 11：

表格 11

顧客排序	137	274	411	548	685
日期	2017/05/31	2017/05/30	2017/05/28	2017/05/25	2017/05/17
分數	5	4	3	2	1

接著針對交易頻率，也就是變數 F 進行處理。交易頻率是指在一定時間間隔內的交易次數，交易次數越高，則此項分數越高，將所有顧客的交易次數由大至小排序之後，交易次數前 20%，也就是大於等於排序第 137 個顧客的交易次數 58 次，分數為 5，而下一個 20% 區隔的分數為 4 等等依此類推，在此將各五分位數的分隔點，也就是第 137,274, 411, 548, 685 個顧客列於表格 12：



表格 12

顧客排序	137	274	411	548	685
交易次數	58	33	19	12	10
分數	5	4	3	2	1

最後針對交易金額，也就是變數 M 進行處理。交易金額則是在一定時間間隔內的交易金額，交易金額越高，則此項分數越高。本研究認為，交易金額與交易頻率之間有共線性存在，導致交易頻率越高，交易金額也會隨之越來越高，本研究針對交易金額的定義應為平均交易金額，而不是總累計交易金額，藉此降低交易頻率和交易金額的共線性。如同前二者一般，交易金額分數前 20%，也就是大於等於排序第 137 個顧客的交易金額，分數為 5，而下一個 20% 的區隔分數為 4 等等依此類推。在此將各五分位數的分隔點，也就是第 137, 274, 411, 548, 685 個顧客列於表格 13：

表格 13

顧客排序	137	274	411	548	685
平均交易金額	1903.9000	596.4667	494.1538	419.5556	341.0714
分數	5	4	3	2	1

經過上述步驟得到所有顧客的 R、F、M 分數之後，將可以得到每個顧客的 RFM 總分，後續將透過 RFM 總分評估顧客價值，進而做為企業在推行促銷組合時的決策依據，優化企業的行銷效率、降低行銷成本，進而為企業提高收益。



3.4.3 疊圖分析

經由上述的步驟，我們可以得到兩組跟顧客交易紀錄有關的資料。由第一步得到與虛擬顧客的距離，可以彙整消費習慣與我們設定的產品組合相似或是滿足的顧客。而第二步得到的 RFM 分數，則能藉由顧客過往歷史交易紀錄，歸納出顧客對於公司的價值以及其消費潛力。

在這一步，我們會統整上面兩步的結果來正式判斷產品促銷組合的價值與適當性。首先，公司將欲推銷之銷售產品組合提出，並假設為一虛擬顧客，擁有一筆與此促銷產品組合相同的交易紀錄，按照上述第一步提到的作法，將此虛擬顧客與其他顧客的距離經由計算其曼哈頓距離分群，這個距離也就是我們認為的顧客交易習慣相似度。當有了距離這個指標之後，就可以知道當推出這個產品組合做促銷後，那些與之相似的顧客是否有其他的消費習慣，或是這個產品組合促銷後，這些相似顧客的購買潛力是否值得公司推出這個促銷專案。

關於後者評估消費潛力的部份，此研究採用的就是稍早提到的 RFM 分析法，在研究方法第二步驟時，已經將所有顧客根據過往的消費紀錄各自評分，若以 RFM 佔相同比重來評斷顧客價值，那將每一位顧客的 RFM 值相加，就可以得知這位顧客最終對公司的營利潛力。取距離虛擬顧客一定範圍內的顧客，當作是符合我們定義中與虛擬顧客有相似消費行為的顧客(本研究取的相似消費顧客為所有顧客的前 1/5)，此定義可以由決策者參考顧客組成或是店家規劃來制定。而取得這些相似顧客的人選後，我們將這些顧客的 RFM 值取總平均，就可以得到，若實行這次的促銷計畫，可以吸引到多少消費潛力的顧客。而當這個總平均值大於我們預設的目標值後，我們便認定這個促銷組合足以替公司帶來足夠的效益。換句話說，我們的做法相當於將第一步得到的相似顧客群與 RFM 的高價值顧客群進行疊圖，重疊部份越多表示這個促銷產品的評價越好，越容易成功，可以為公司帶來更多的利潤。反之，若是預設的虛擬產品組合得到一個比較偏低的目標顧客 RFM 值，代表也許這組欲促銷的產品組合並不是那麼的吸引人，並不足以替公司帶來好的促銷結果，如此可

以在促銷產品組合的挑選上做一些調整或者甚至重新規劃，這是一個可以過濾促銷策略的快速篩選機制。

本實驗第三步同時也探討有無產品交易數量所帶來的影響，用如上述一樣的方法來驗證我們在第一步驟中所提到的，未將產品交易數量納入時，無法分辨顧客偏好產品的弊病。同樣的手法，我們架構一個虛擬顧客，其虛擬交易紀錄中，購買的產品屬於兩種(或多種)具有競爭性或具有替代性的產品，但在實驗中會改變這兩種(或多種)產品的數量比例，來模擬顧客在挑選這些替代性商品的時候，會產生的偏好。最後藉由結果，可以驗證是否如我們所預期，當兩樣替代性產品在顧客購買行為中出現對某一產品的偏好時，未考慮產品交易數量的演算法將無法從運算結果得到足夠的資訊。而相反的，若納入產品交易數量，可以明確的分辨出具有特定偏好的顧客族群，進而獲得更進一步的資訊，或是在最後邁入促銷計畫階段時，能有更充足且精準的顧客行為預測。

第4章 實驗結果



4.1 有數量及無數量比較

在上一章提到，為了區分有納入產品交易數量與無納入產品交易數量的演算法對運算結果的差別，本研究挑了兩樣產品進行實驗，確實驗證上述所提到的，在納入產品交易數量的情況下，可以有效區隔顧客的偏好，並從交易資料中離藕出重要的顧客資訊。如表格 14，本次實驗挑選的兩樣產品分別為牛奶與豆漿，選這兩種產品是因為這兩種產品同為重要的蛋白質飲品來源，也具有沖泡的功能，兩者之間具有明顯的替代及競爭關係。為了區分產品交易數量是否會影響結果，我們同時跑了兩次分析，第一次牛奶購買數量是豆漿的兩倍，另一次牛奶是豆漿的二分之一。經由這組虛擬產品組合需求的定義，如果納入產品交易數量進行距離計算，可以辨別出對於牛奶或是豆漿分別有偏好的顧客，但反之，未納入產品交易數量的計算，在這組虛擬交易組合下，會得到一模一樣的結果如表格 14：

表格 14

	虛擬顧客		R	F	M	RFM
	牛奶	豆漿				
有產品交易數量	2	1	2.67	2.64	2.79	8.10
有產品交易數量	1	2	2.78	2.68	2.84	8.30
無產品交易數量	2	1	2.74	2.62	2.76	8.12
無產品交易數量	1	2	2.74	2.62	2.76	8.12

經由上述實驗可以發現，這次模擬的結果符合先前的預期。首先，在未考慮產品交易數量的屬性下，對於牛奶和豆漿這兩種分別有偏好的顧客將視為同一種消費行為，所以不論這兩種產品的數量組合如何變化，運算得到的顧客群都會是完全

同樣的一群人，這點經由個別 R、F、M 值以及總體平均 RFM 值可以驗證。因此，若不考慮交易產品的數量，則有部份的消費資訊、交易習慣會被忽略。

相反的，由表格 14 可以看到，如果對同一資料集進行運算，在納入產品購買數量的情況下，可以得到截然不同的結果。不只是單一 R、F、M 值在同樣產品組合產生的結果不同，考慮數量的演算法更可以有效區分出消費者不同的消費偏好。以同樣表格 14 的結果可以看出，在牛奶與豆漿的產品組合中，調動兩樣產品的比例，可以區別出顧客對於兩種產品的偏好。舉例來說，當牛奶的比例被調高時，這時演算法圈選到與虛擬顧客距離較相近的顧客，可以解讀為在預設的這兩樣產品中，被圈選到的顧客群對於牛奶有較大的偏好。

除此之外，這次的模擬結果不只幫助我們藉由交易紀錄得到顧客偏好，能更精確的替公司決定出理想的促銷產品組合。如果未加入產品交易數量作為一重要屬性，並不能在牛奶與豆漿之間分辨何者的貢獻比較有價值，比較受歡迎的項目。而如果把產品交易數量作為一個重要屬性進行分析，能明確的分辨出何者的貢獻比較大，並且能從眾多產品組合中離羈出貢獻價值較大的產品。由此次 RFM 結果來看，於 W 超市師大分店消費的顧客群中，豆漿類產品對牛奶類產品比例高的時候，可以得到比較大的 RFM 值，這也就代表，當超市想要對牛奶豆漿產品納入促銷組合時，調整豆漿銷售比例高一點的組合將更能吸引潛在價值比較高的消費者，更能增加公司促銷產品的成效。

4.2 疊圖分析結果

為了更進一步演練此研究所提到的方法是否能在真正的交易資料中進行，先將初始資料經過前處理，再利用電腦程式替我們廣泛且隨機的挑選具有潛力的產品組合，其中程式基本流程如以下：

(一)篩選所有產品中被消費次數大於 100 次的產品成為我們選取產品的資料來



源，避免被過於龐大的產品種類影響計算速度及效率，且鮮少被購買的產品對分析較無意義。

- (二)從資料來源隨機挑選 2-5 樣產品，各產品的數量則從 1-3 中隨機產生，大範圍地替企業找出各種可能的組合，並將此組合定義為虛擬顧客。
- (三)利用此虛擬顧客與原有的顧客交易紀錄進行演算法計算以及疊圖分析，得到與虛擬顧客距離較近的前 1/5 的顧客之平均 RFM 值。
- (四)定義 RFM 值的標準以過濾出有潛力的促銷商品組合。

對公司行銷策略而言，如果能鎖定價值高的顧客，將會對促銷的產生事半功倍的效果，因此，若最終虛擬產品組合經過疊圖後能得到一個比我們預想標準還要高的 RFM 數值，也就是本研究所認定的理想產品組合，是企業可以納入候選名單的有利產品組合。由程式隨機演算法的結果列出 20 組於表格 15：

表格 15

組別	產品種類數	產品數量分布	RFM 平均	R 平均	F 平均	M 平均
1	2	3, 2	10.82	3.61	3.89	3.32
2	2	2, 3	9.8	3.29	3.5	3.01
3	4	3, 3, 2, 1	10.7	3.36	3.89	3.45
4	3	3, 1, 2	9.67	3.22	3.39	3.06
5	3	3, 1, 2	9.22	3.18	3.08	2.96
6	2	3, 3	9.05	2.79	3.02	3.24
7	4	3, 1, 1, 2	9.28	2.89	2.76	3.63
8	2	2, 2	10.44	3.39	3.54	3.51
9	5	2, 1, 3, 1, 2	8.91	2.88	3.04	2.99
10	4	1, 1, 2, 1	8.82	2.99	2.75	3.08

11	2	3, 1	9.16	3.02	3.16	2.98
12	2	3, 1	10.12	3.27	3.51	3.34
13	2	2, 3	10.02	3.24	3.57	3.21
14	5	2, 3, 2, 3, 2	8.76	2.8	2.79	3.17
15	5	2, 2, 3, 3, 1	9.22	2.98	2.95	3.29
16	5	3, 3, 3, 2, 3	8.83	2.85	2.87	3.11
17	5	3, 1, 3, 3, 3	8.41	2.62	2.55	3.24
18	2	3, 3	9.02	2.86	2.92	3.24
19	5	2, 2, 2, 3, 2	8.88	3.04	2.58	3.26
20	2	2, 1	8.89	3.01	2.93	2.95

在表格 15 中，「產品種類數」表示我們所挑選出來的商品組合中，共含有幾種產品。「購買數量分布」則是表示這幾樣產品分別購買了多少數量，這些數量的資訊，比起未考慮數量的情況下，更能幫助區分顧客的偏好。「RFM 平均」是距離虛擬顧客前 1/5 群顧客其各別 R、F、M 值加總後的總平均。本研究將 R、F、M 各別的權重訂為相同，但如果有特殊需求，或特殊評估目的，也可以將 R、F、M 值分別加權後再相加，以得出符合目的的結果。最後，R 平均、F 平均、M 平均表示同 RFM 平均的定義，為那一群顧客中各別 R、F、M 值的平均。

這些隨機的產品組合，的確可以吸引到不同的顧客群，而在這些顧客群中，每個人的平均 RFM 數值，或稱顧客價值並不相同。如前面所提，本研究希望由演算法得到的結果是，一個好的產品組合，必須是離虛擬顧客交易越近的顧客群，若將其投影到其對應的 RFM 值時，大部分將會落在 RFM 值較高的區域。

因此，雖然可以利用上述隨機選取的方式自動將產品排列組合，並經由本研究的距離方法得到與虛擬產品組合距離比較近的顧客及其 RFM 值，但並不是每一個



產品組合都是值得企業作為促銷產品組合，可以依照當下的行銷計畫，預估銷售規模等等決定過濾顧客價值的下限，而評估標準可以由促銷預定計畫而有不同的標準，像是取 RFM 加總後的總平均值，或是因為特殊目的而取個別的 R、F、M 值篩選，而不同的行銷策略，會需要不同篩選標準以及決定性的參數。接下來將舉出 4 種利用此研究提出之演算法所推出的促銷方案。

(一) 注重回購率的行銷規劃：

在部分行銷規劃中，企業希望能提升顧客的回購率。因此，對於最近才來店消費的顧客會比較有興趣。舉例來說，許多市場上的促銷行為，會是在消費後提供有使用期限的折價卷或是限時回店禮，希望顧客除了滿意這次的購物經驗之外能在短時間內再次光顧。而這種重視短時間能夠回店消費的促銷規劃，可以套用在此研究 RFM 模型的 R 值，藉由提高 R 的權重重新分析，或是以比較極端的情況下直接對 R 值進行篩選。以表格 15 為例，若將 R 平均值的過濾標準訂為 3.3，也就是在目標產品組合中，其經由演算後得到的顧客 R 平均值，必須大於 3.3 才能被視為是較為理想的產品組合，能促進回購率。表格 16 為篩選之後的結果：

表格 16

組別	產品種類數	產品數量分布	RFM 平均	R 平均	F 平均	M 平均
1	2	3, 2	10.82	3.61	3.89	3.32
3	4	3, 3, 2, 1	10.7	3.36	3.89	3.45
8	2	2, 2	10.03	3.39	3.54	3.1

只有上面這 4 組產品組合符合這次促銷的標準。這樣的方法除了可以得到產品組合外，我們可以把這個標準值回推到當初資料設定值，R 值等於 3.3 大約是距離上一次來店 3 天內的顧客，而這個值可以做為企業評估促銷回饋值的參考標準。例如，企業可以給顧客一張 3 天內回店消費的優惠卷，回饋這些顧客，並刺激這些



對企業而言價值高的顧客，以激發其回購動機，還能吸引到一些 R 值較低的顧客，增加回店消費的動力，進而達到增加回購率的目的與妥善的行銷規劃。

(二)強調來店頻率的行銷規劃：

有些企業特別重視顧客在一定期間內來店的頻率，就像現在許多商店，都會推集點優惠，或者是網站連續登入即享有禮品回饋，這都是希望在期間內提升顧客消費的次數。為了迎合這種行銷規劃，可以套用在此研究 RFM 模型並加重 F 值權重，或是比照上一個例子，直接對 F 平均值訂定標準後過濾。若將 F 平均值的過濾值訂為 3.5，篩選之後的結果如表格 17：

表格 17

組別	產品種類數	產品數量分布	RFM 平均	R 平均	F 平均	M 平均
1	2	3, 2	10.82	3.61	3.89	3.32
2	2	2, 3	9.8	3.29	3.5	3.01
3	4	3, 3, 2, 1	10.7	3.36	3.89	3.45
8	2	2, 2	10.03	3.39	3.54	3.1
12	2	3, 1	10.12	3.27	3.51	3.34
13	2	2, 3	10.02	3.24	3.57	3.21

上表 6 組產品組合符合這次促銷的標準。若如同上例回推初始資料，得到 F 值為 3.5 的標準大約是在資料蒐集期間來店頻率為 26 次的顧客。舉例來說，促銷方案可以訂為凡在期間內當日來店都可以獲得一點集點點數，若是期間內滿 25 點，即可兌換贈品。這個數目經由本研究演算法制定，不只可以回饋到原本價值高的顧客，更可以促進其他 F 值較低顧客的消費動力，進而提高企業利益。



(三)注重單一消費金額的促銷計劃：

在許多情況下，促銷計畫可能會以總價折抵，或是目標價格回饋方式，來達到增進顧客消費額的結果。這種行銷方式會希望顧客單一消費金額高，在此研究的RFM模型中，就是M平均值。比照前兩個例子的做法如表格 18 將 M 平均值以 3.2 進行篩選：

表格 18

組別	產品種類數	產品數量分布	RFM 平均	R 平均	F 平均	M 平均
1	2	3, 2	10.82	3.61	3.89	3.32
3	4	3, 3, 2, 1	10.7	3.36	3.89	3.45
6	2	3, 3	9.05	2.79	3.02	3.24
7	4	3, 1, 1, 2	9.28	2.89	2.76	3.63
12	2	3, 1	10.12	3.27	3.51	3.34
13	2	2, 3	10.02	3.24	3.57	3.21
15	5	2, 2, 3, 3, 1	9.22	2.98	2.95	3.29
17	5	3, 1, 3, 3, 3	8.41	2.62	2.55	3.24
18	2	3, 3	9.02	2.86	2.92	3.24
19	5	2, 2, 2, 3, 2	8.88	3.04	2.58	3.26

以 M 平均值 3.2 作為過濾標準時，回推資料庫得到的平均消費金額大約為 514 元，因此，在促銷規劃時，企業能制定在顧客消費金額 510 元以上才能得到小禮物等回饋。這個金額回饋到原本的高價值顧客，更能吸引到原本 M 平均值較低的顧客，提升其消費的動機與衝動。因此，這個標準值如果經由本研究的演算法制定，將有一定的功效和說服力。



(四)重視總體的促銷計劃：

促銷行為有時並不特別強調單一的顧客消費行為，而是強調總體平均的貢獻度。一般對產品組合的促銷行為，例如折扣等，都是屬於比較單純考慮整體貢獻的促銷方式。對本研究而言，RFM 平均值將 R 平均值、F 平均值、M 平均值不做各別加權，視為相等比重取整體平均進行篩選，概念等同平均考慮整體消費者貢獻性質的方法。將 RFM 平均值以 10.0 作為過濾標準產生的結果如表格 19：

表格 19

組別	產品種類數	產品數量分布	RFM 平均	R 平均	F 平均	M 平均
1	2	3, 2	10.82	3.61	3.89	3.32
3	4	3, 3, 2, 1	10.7	3.36	3.89	3.45
8	2	2, 2	10.03	3.39	3.54	3.1
12	2	3, 1	10.12	3.27	3.51	3.34
13	2	2, 3	10.02	3.24	3.57	3.21

RFM 平均值是基本的評估整體顧客價值的方法，適用比較無差別無限制的促銷計劃，使企業能以顧客總價值評估整體促銷方案。

第5章 結論



5.1 管理意涵

本研究提出，在計算顧客相似度並進行顧客區隔時，應將產品交易數量的概念加進分析，同時實驗有產品交易數量及無產品交易數量的分析，並比較有無產品交易數量是否會造成結果不同。本研究實驗證明，產品交易數量對分析顧客而言非常重要，並確實能區分顧客不同的偏好，進而使顧客區隔更加準確、有效率，對企業而言，也能使企業在制定行銷方案時能有更全面的依據。

本研究使用真實世界資料進行分析時，觀察到歐幾里得距離所造成的實驗失真，並即時使用曼哈頓距離作為計算交易相似性與顧客相似性之改善，並於研究方法中詳細列出兩種距離計算方法的差異。

本研究將欲行銷之產品組合視為一虛擬顧客，加入產品交易數量的分析方式，找出與虛擬顧客最相似的顧客群，並結合傳統 RFM 模型計算顧客價值，找出真正具有價值的顧客，作為企業在推行促銷組合的參考依據，並在最後依照企業行銷的不同目的分別舉出四種行銷方案。

5.2 研究限制

隨著時間推進，交易資料會逐漸增加、豐富，若能在制定行銷策略的同時，即時分析當時資料庫的所有顧客，將會使分析更加全面且準確。

截至實驗結束為止，該超市尚未建設會員系統，因此資料庫內缺乏顧客其他的個人資訊，因此本研究只能就其購買行為分析，尚無法涵蓋顧客的基本資料分析，如果能應用到資料屬性更廣泛的資料集，則在虛擬顧客的定義上，可以依照企業個

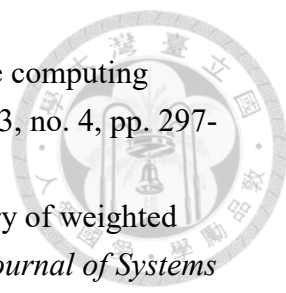
別需求增加更多分析屬性，如顧客性別、年齡等等，獲得除了交易相似性之外更多的資料，讓實驗結果更加獲益。

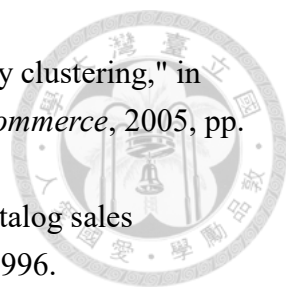
本研究雖證明產品交易數量確實能改善分析結果，分辨顧客不同的喜好，並在最後結合 RFM 模型為顧客評分，作為企業推行促銷組合的參考依據，在實驗截止前並未於該超市中實際實行，因此未來可以依據實驗結果向企業推薦最佳產品組合，希望能優化企業行銷效率、降低行銷成本，進而為企業提高收益。

第6章 文獻引用



- [1] 經濟部統計處. (2017). Available:
https://www.moea.gov.tw/MNS/dos/bulletin/Bulletin.aspx?kind=8&html=1&menu_id=6727&bull_id=4878
- [2] EIU. (2015). *Forecast for global retail sales growth from 2008 to 2018*. Available:
https://www.pwccn.com/home/webmedia/635593364676310538/rc_outlook_2015_16.pdf
- [3] S. Peker, S. Peker, A. Kocyigit, A. Kocyigit, P. E. Eren, and P. E. Eren, "LRFMP model for customer segmentation in the grocery retail industry: a case study," *Marketing Intelligence & Planning*, vol. 35, no. 4, pp. 544-559, 2017.
- [4] D. Van den Poel and B. Lariviere, "Customer attrition analysis for financial services using proportional hazard models," *European journal of operational research*, vol. 157, no. 1, pp. 196-217, 2004.
- [5] H. H. Chang and P. W. Ku, "Implementation of relationship quality for CRM performance: Acquisition of BPR and organisational learning," *Total Quality Management*, vol. 20, no. 3, pp. 327-348, 2009.
- [6] M. Namvar, M. R. Gholamian, and S. KhakAbi, "A two phase clustering method for intelligent customer segmentation," in *Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on*, 2010, pp. 215-219: IEEE.
- [7] S. Dibb, "Market segmentation: strategies for success," *Marketing Intelligence & Planning*, vol. 16, no. 7, pp. 394-406, 1998.
- [8] A. D. Athanassopoulos, "Customer satisfaction cues to support market segmentation and explain switching behavior," *Journal of business research*, vol. 47, no. 3, pp. 191-207, 2000.
- [9] J. Kim, E. Suh, and H. Hwang, "A model for evaluating the effectiveness of CRM using the balanced scorecard," *Journal of interactive Marketing*, vol. 17, no. 2, pp. 5-19, 2003.
- [10] K. Lu and T. Furukawa, "A Framework for Segmenting Customers Based on Probability Density of Transaction Data," in *Advanced Applied Informatics (IIAIAI), 2012 IIAI International Conference on*, 2012, pp. 273-278: IEEE.
- [11] Y. Zhang, J. Jiao, and Y. Ma, "Market segmentation for product family positioning based on fuzzy clustering," *Journal of Engineering Design*, vol. 18, no. 3, pp. 227-241, 2007.

- 
- [12] L. Zeng, L. Li, and L. Duan, "Business intelligence in enterprise computing environment," *Information Technology and Management*, vol. 13, no. 4, pp. 297-310, 2012.
- [13] Y.-H. Hu, T. C.-K. Huang, and Y.-H. Kao, "Knowledge discovery of weighted RFM sequential patterns from customer sequence databases," *Journal of Systems and Software*, vol. 86, no. 3, pp. 779-788, 2013.
- [14] M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge, "Knowledge management and data mining for marketing," *Decision support systems*, vol. 31, no. 1, pp. 127-137, 2001.
- [15] G. Punj and D. W. Stewart, "Cluster analysis in marketing research: Review and suggestions for application," *Journal of marketing research*, pp. 134-148, 1983.
- [16] C. Marcus, "A practical yet meaningful approach to customer segmentation," *Journal of consumer marketing*, vol. 15, no. 5, pp. 494-504, 1998.
- [17] A. Ansari and A. Riasi, "Customer clustering using a combination of fuzzy c-means and genetic algorithms," *International Journal of Business and Management*, vol. 11, no. 7, p. 59, 2016.
- [18] H.-C. Chang and H.-P. Tsai, "Group RFM analysis as a novel framework to discover better customer consumption behavior," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14499-14513, 2011.
- [19] A. Wilson, V. A. Zeithaml, M. J. Bitner, and D. D. Gremler, *Services marketing: Integrating customer focus across the firm*. McGraw Hill, 2012.
- [20] T. Teichert, E. Shehu, and I. von Wartburg, "Customer segmentation revisited: The case of the airline industry," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 227-242, 2008.
- [21] T. Hong and E. Kim, "Segmenting customers in online stores based on factors that affect the customer's intention to purchase," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2127-2131, 2012.
- [22] V. A. Zeithaml, R. T. Rust, and K. N. Lemon, "The customer pyramid: creating and serving profitable customers," *California Management Review*, vol. 43, no. 4, pp. 118-142, 2001.
- [23] P. C. Verhoef and B. Donkers, "Predicting customer potential value an application in the insurance industry," *Decision support systems*, vol. 32, no. 2, pp. 189-199, 2001.
- [24] A. M. Hughes, "Boosting response with RFM," *Marketing Tools*, pp. 4-8, 1996.
- [25] J. R. Miglautsch, "Thoughts on RFM scoring," *Journal of Database Marketing & Customer Strategy Management*, vol. 8, no. 1, pp. 67-72, 2000.

- 
- [26] J. Wu and Z. Lin, "Research on customer segmentation model by clustering," in *Proceedings of the 7th international conference on Electronic commerce*, 2005, pp. 316-318: ACM.
- [27] G. R. Bitran and S. V. Mondschein, "Mailing decisions in the catalog sales industry," *Management science*, vol. 42, no. 9, pp. 1364-1381, 1996.
- [28] R. Kahan, "Using database marketing techniques to enhance your one-to-one marketing initiatives," *Journal of Consumer Marketing*, vol. 15, no. 5, pp. 491-493, 1998.
- [29] U. Kaymak, "Fuzzy target selection using RFM variables," in *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, 2001, vol. 2, pp. 1038-1043: IEEE.
- [30] N.-C. Hsieh, "An integrated data mining and behavioral scoring model for analyzing bank customers," *Expert systems with applications*, vol. 27, no. 4, pp. 623-633, 2004.
- [31] S. F. King, "Citizens as customers: Exploring the future of CRM in UK local government," *Government Information Quarterly*, vol. 24, no. 1, pp. 47-63, 2007.
- [32] S. H. Ha and S. C. Park, "Application of data mining tools to hotel data mart on the Intranet for database marketing," *Expert Systems with Applications*, vol. 15, no. 1, pp. 1-31, 1998.
- [33] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," *Procedia Computer Science*, vol. 3, pp. 57-63, 2011.
- [34] D.-C. Li, W.-L. Dai, and W.-T. Tseng, "A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7186-7191, 2011.
- [35] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259-5264, 2010.
- [36] H. Güçdemir and H. Selim, "Integrating multi-criteria decision making and clustering for business customer segmentation," *Industrial Management & Data Systems*, vol. 115, no. 6, pp. 1022-1040, 2015.