

應用資料採礦探討國際線航空旅客之 線上購票行為¹

APPLICATION OF DATA MINING TO EXPLORE INTERNATIONAL AIRLINE PASSENGER ONLINE BOOKING BEHAVIOR

林祥生 Hsiang-Sheng Lin²
劉益豪 I-Hao Liu³

(96 年 8 月 1 日收稿，96 年 9 月 29 日第一次修改，96 年 12 月 3 日
第二次修改，97 年 5 月 21 日定稿)

摘 要

在網際網路日益普及下，近年來航空公司多將線上購票視為網站的必備功能，也因此與旅客之間出現新的互動關係。拜電子商務蓬勃發展之賜，航空公司透過旅客的線上購票行為，能同步獲得豐富的顧客資料及交易紀錄，而如何善用此原始資料，並透過適當的資料處理技術，提供給顧客個人化的行銷及服務，將是未來航空公司的重要課題。本研究應用資料採礦中的群集化、分類及關聯分析技術，對線上購票旅客進行探討，首先用 RFM 指標 (recency、frequency、monetary) 及艙等和平均哩程數等五項顧客價值指標對顧客進行群集化，接著以其結果建構顧客分類預測模式，最後則針對不同旅次目的進行航線之間的關聯分析。本研究成果在實地訪談航空公司相關人員後，也獲得實務上的驗證。

-
1. 本研究承行政院國家科學委員會專題研究補助 (NSC 94-2211-E-216-015)，特此致謝。
 2. 中華大學運輸科技與物流管理學系助理教授 (聯絡地址：300 新竹市五福路 2 段 707 號中華大學運輸科技與物流管理學系；電話：03-5186532；E-mail：keyman@chu.edu.tw)。
 3. 中華大學科技管理研究所運輸科技與物流管理組碩士。

關鍵詞：線上購票；資料採礦；群集化；分類；關聯規則

ABSTRACT

In recent years, more and more airlines regard on-line booking as the basic function of their homepage. Since the Internet becomes more popular day by day, the traditional travel agencies are facing the crisis of having the middleman removed. However, a new interaction has appeared between airlines and passengers. While the e-commerce develops vigorously, the airlines go through passengers' online booking behavior, namely to obtain more customer information and transaction records. Therefore, how to use this data to understand the customers, experience suitable data processing technology, and provide the customized marketing service to riders, all have become issues of future airlines. This research will explore data mining to discuss airline passengers' online booking behaviors. First, we adopt the RFM model (Recency, Frequency, Monetary), the average mileage and classes as five customer value index items to process clustering for riders. The result will make a classification and quickly distinguish customer group belongings. Again, we aim at service products classification to understand the consumers' behavior of each route. Finally, the association analysis is carried out for different trip purposes, thence appears customers' implicit connecting demands between all routes.

Key Words: Online booking; Data mining; Clustering; Classification; Association rule

一、緒 論

在傳統航空客運的銷售通路上，多由旅行社代為經手，再利用電腦訂位系統 (computer reservation system, CRS) 與航空公司相聯結，故航空業者必須支付相當的佣金給旅行社，國際航空運輸協會 (International Air Transport Association, IATA) 曾調查此一佣金支出約占總營運成本的 8% 左右，因此對航空業而言始終是一項沈重的負擔^[1]。近年來網際網路的普及，將人們帶向電子商務的時代，於是許多航空公司也紛紛投入此一領域，利用網際網路的便利性與即時性來提供網路購票的服務，並讓顧客能夠直接與航空公司進行互動，使得顧客、旅行社、航空公司三者之間的關係逐漸改變，一種「去中間化」的趨勢已悄然形成，傳統旅行社所扮演的中間商角色正面臨考驗，而航空公司將可以直接面對顧客，由顧客直接對其服務進行投票，所以航空公司之顧客關係管理亦更形重要。

航空公司屬於高成本投資，因此如何將其資產價值有效發揮在顧客身上，為其經營關鍵點。掌握與找出高價值顧客的方法，不外乎更加了解顧客，以規劃最佳的顧客服務或行銷策略，而行銷資源的運用，必須精準地花在刀口上，以期達到公司的經營目標。隨著全球網際網路的普及，透過網路訂票的航空旅客正日漸增加，而顧客每次上網交易都可留下

詳細的紀錄，如性別、年齡、旅次起迄點、搭乘日期、艙等及票價等，若能善用這些紀錄並運用適當的資料處理技術，將有助於航空業者更了解顧客的需求及偏好，以提供更符合顧客期望的優質服務。

本研究第一階段先針對航空客運業的顧客價值指標進行探討，採用過去最常被應用的 RFM 模式，由於該模式使用於各種產業時應有不同的考量，故可加入新的衍生變數或將原本的 RFM 參數加以修改，甚至增加或降低指標的權重，以期更符合不同產業之特性。第二階段則運用資料採礦 (data mining) 中的群集化、分類、關聯分析等功能，協助航空業者逐步深入了解顧客。首先運用調整過後的 RFM 模式所設定之顧客價值指標，將其指標作為群集化分析之變數，使顧客自動分成不同群組，接著運用分類功能了解不同價值顧客群的消費特性，進而預測新進顧客的價值高低，此外也應用分類功能對不同的航線進行分析，了解不同航線其顧客特性為何，最後則以關聯分析技術找出旅客對不同航線間的偏好組合，進而提出適合於不同顧客群的行銷策略。本文之研究架構可如圖 1 所示。

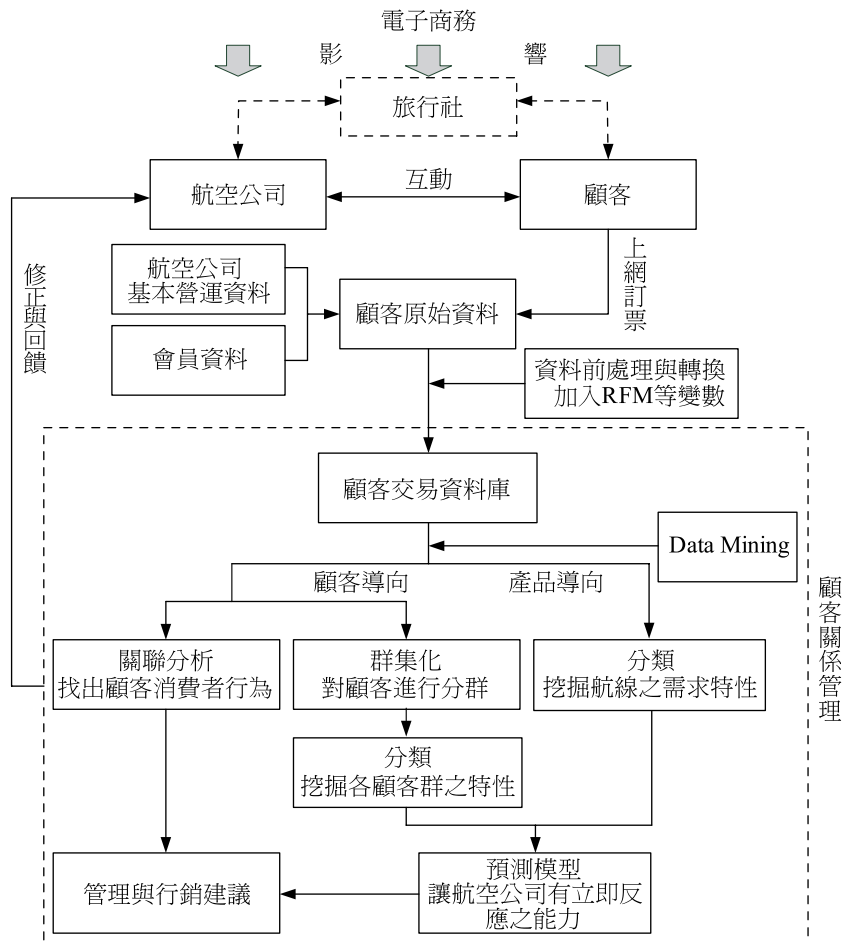


圖 1 研究架構圖

二、文獻回顧

2.1 資料採礦

2.1.1 資料採礦定義

關於資料採礦的定義，已存在許多不同版本，其中較具代表性者包括學者 Fayyad 等人^[2]將其定義為「能將資料簡化成正確的、未曾被知道的、有用的以及最容易了解之規則的一連串重要處理動作」，Berry 等人^[3,4]認為「資料採礦是針對大量的資料，利用自動化或半自動的方式進行分析，以尋找出有意義的關係或法則」，Han^[5]則定義「資料採礦是從大量資料中萃取出來的知識」，而尹相志^[6]認為「資料採礦是利用統計以及機械學習的演算法，啟發性地從大量資料中找尋隱藏具有商業價值的知識與規則，作為自動化商業策略之應用」。資料採礦在過去經常被用來分析信用卡客戶或手機用戶的消費行為，以輔助經營者從龐雜的客戶帳單資料中，挖掘出有助於未來精準行銷的決策資訊，因此美國麻省理工學院的 Technology Review 雜誌曾將此一分析技術選為改變未來世界的十大創新科技之一^[7]。

2.1.2 資料採礦功能

根據 Han 的分類^[6]，資料採礦依其模型主要可分為五大功能，其內容可概述如下：

1. 分類

乃是指將資料依據事先定義的特徵屬性進行分類處理工作，亦即根據一些已知分類資料的變數屬性數值來進行運算，在得到整個分類之相關規則後，再將未知分類的資料帶入規則，以得到最後分類結果。

2. 預測

預測是推估未來的數值以及趨勢，歷史資料可用來建立模型，以檢視近來觀察值的資料，例如由過去行銷活動所產生的反應，來預測未來新活動的回應率。

3. 群集化

是依資料本身的特性將資料進行分群的一個程序。通常在不斷的訓練與修正過程後，群集化的特性會在族群本身顯現出來，其目的在於將群組之間的差異辨識出來，並進一步對個別組內之相似樣本進行挑選。在分析之前，分析者並不知道會以何種方式或根據來進行判斷。

4. 關聯規則

或稱為關聯分析，常被應用於超級市場的銷售物品上，其基本目的是希望在多屬性或多物項組合的關聯探討中，尋找各種銷售物項彼此之間在消費者購買行為上的關聯性。

5. 順序分析

針對資料的順序關係進行採礦，例如是時間上的順序、行為上的順序等，主要是希望找出趨勢或是順序上的分析資訊。

2.1.3 資料採礦應用

盧世銘^[8]針對航空公司的會員系統，運用資料採礦技術進行顧客分群，以提出不同族群的行銷策略建議。其研究資料包括會員系統的基本資料、航空訂位系統之會員實際搭乘資料與兌獎資料、航線哩程對照表、各營業據點淡旺季對照表等，運用航段淨營收 (REV)、高艙等搭乘比例 (FCRATE)、訂位艙等代碼 (RBDNO) 及顧客價值指標 (VIDX) 等四項變數進行分群，並對於這四項指標給予權重，將顧客分為 15 群，再區分為主要核心會員、有效會員、流失會員、邊緣會員及新會員等 5 種不同特性之群集，最後分別提出其行銷策略，研究架構如圖 2 所示。該研究結果顯示，部分群集對於網路訂位及電子商務的通路偏好明顯較高，且平均淨營收屬低等級，因此間接印證網路行銷的售價敏感特性，故建議航空公司可建置網上拍賣機制，以作為促銷策略的手段。



資料來源：盧世銘^[8]

圖 2 航空公司會員群集化分析流程

王媚慧^[9]以半導體通路業者為研究對象，從過去的交易資料中，經過資料採礦後進一步分析客戶特性，並構建具有價值的產品推薦模型。該研究之目的在於探究「個別客戶群會感興趣的是哪些產品」，故先進行群集化分析，選取的變數包括公司狀況、營業收入、資本額等，試圖從群集化中找出對公司利益貢獻最大的客戶面貌，接著以關聯規則技術，挖掘客戶在不同產品之間的採購關聯性，期能提供銷售人員與決策者在行銷上的重要參考，進而提升銷售和交叉行銷的機會。

Chan^[10]研究資料採礦於銀行業顧客關係管理上之應用，並實際建構一顧客流失預測模型，首先將顧客定義為現今顧客、自願流失、非自願流失三種，再對顧客進行分類與關聯分析，在分類技術上採用決策樹、羅吉特迴歸及類神經網路等三種不同模式，以比較各

分類模式的正確性，結果發現決策樹的正確率達 81.6%、羅吉特迴歸為 80.0%、類神經網路僅 77.9%，因此決策樹為其最佳模式，而決策樹又有便於說明及解釋的優點，更容易供決策者參考。

Changchien 與 Lu^[11] 欲提供一個從資料庫中發掘關聯規則的流程，運用所得之結果進行網路行銷，來增加網路商店的競爭優勢。該研究先運用自我組織映射地圖 (self-organization map, SOM) 將資料群集化，再以模糊集合理論 (rough set theory) 找出群集內的關聯規則，並針對不同的群集進行關聯分析，以了解不同群集間的關係，最後則運用所提出的方法，對實際網路商店的資料進行研究，以達到買賣雙方滿意度的提升。

Liao 與 Chen^[12] 利用資料採礦技術為愛買量販景美店建立資料庫，並運用關聯分析來找出可交叉銷售之商品，作為電子型錄設計及行銷的參考。該研究在資料庫的設計上，分為概念性資料庫、邏輯資料庫設計、實體資料庫設計等三部分，再運用關聯規則對量販商品進行購買偏好分析，最後提出產品導向與自有品牌導向等二種電子型錄策略，以提供個案公司參考。

有關前述文獻所研究之產業別及使用的資料採礦技術，可彙整比較如表 1 所示。

表 1 相關文獻所採用的資料採礦技術比較表

文 獻	產 業 別	群集化	關聯分析	分 類
Lu, 2001	網路商店	◎	◎	—
Koh 與 Chan, 2002	銀 行 業	◎	◎	◎
Liao, 2004	零 售 業	—	◎	—
王媚慧, 2004	半導體業	◎	◎	—
盧世銘, 2004	航空公司	◎	—	—

資料來源：本研究整理

2.2 顧客價值

2.2.1 顧客價值的涵義

有關顧客價值的涵義，Arthur^[13] 認為「所謂顧客終身價值 (customer lifetime value)，是指在未來一段時間之內，企業或廠商可以從個別顧客獲得之利潤的淨現值。」，而在資訊科技發達的今日，企業其實可依據資料庫中顧客的購買紀錄，算出每位顧客可能貢獻於企業的終身價值。透過顧客終身價值的計算，企業除了可以預測未來的營收情況外，還能確認出顧客價值的高低，以分配不同的企業資源於不同價值的顧客身上，而進一步採取差異化的行銷策略。關於顧客價值的運算，有許多學者提出不同意見，而現今較常被應用的模式，正是 Arthur^[13] 所提出的 RFM 模式。

2.2.2 RFM 模式定義

RFM 模式來自於「最近購買時間 (recency)」、「購買頻率 (frequency)」、「購買金額 (monetary amount)」三項指標的組合，係評量顧客忠誠度與顧客貢獻度時，最常使用的一個評估方法。根據 Arthur 的研究^[13]，RFM 通常以平均數來代表，利用這三項指標可用來衡量企業與顧客的關係，辨識每位顧客的潛在價值，據以擬出顧客關係策略。企業常利用 RFM 指標對顧客評分，其目的在量化顧客的消費行為，使其符合科學行銷公式應用。RFM 模式可說明如下：

1. 最近購買時間 (recency)

為測量時間的一個量度，根據顧客上次消費的日期，計算由最後一次購買起算至現在之時間。購買某產品之時間距現在愈近，隱含著該顧客再次惠顧的強度愈高，反之若離現在愈遠，該顧客之再購意願將隨之降低，表示其購買行為可能已改變或轉換至他處消費。

2. 購買頻率 (frequency)

測量一個時段內(如一個月、一個星期等)顧客所購買的總次數，或是衡量顧客在此時段內與公司之互動程度，購買頻率愈高，代表顧客與公司互動程度愈高。在購買紀錄中，若購買某產品之頻率愈高，也表示該顧客對此產品有愈高的熱衷程度，持續購買之動機亦較強。

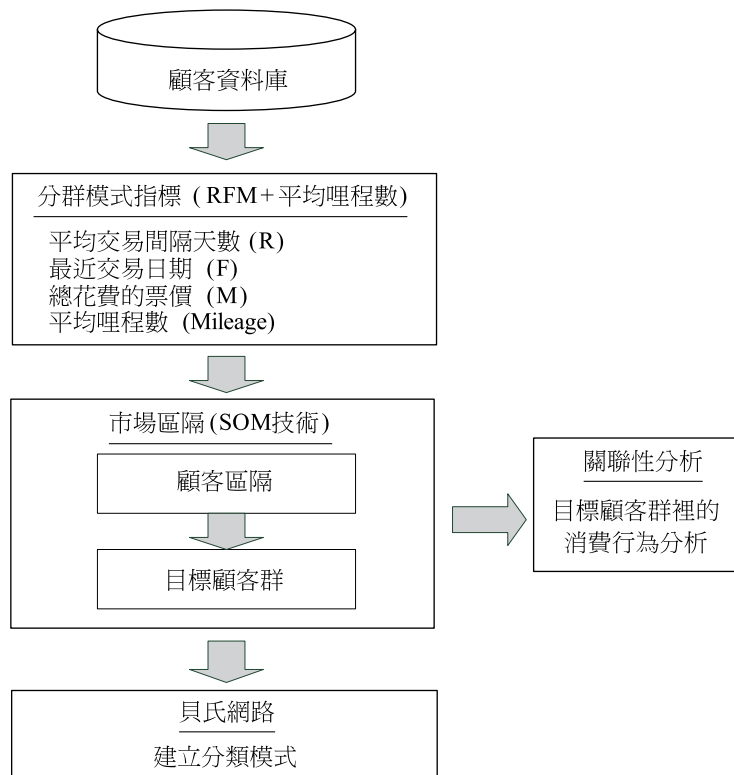
3. 購買金額 (monetary amount)

係指在某一時段內購買的總金額。顧客購買某產品之總金額，既是對公司的實質金錢貢獻，亦代表對此產品之興趣指標。顧客對某產品消費愈多的金額，不但代表該顧客擁有愈高的消費能力(如頭等艙)，也展現其對該產品的需求特性。

2.2.3 RFM 模式應用

薛文佑^[14]針對某航空公司的全體交易資料進行研究，根據其消費者資料庫，運用資料採礦探索顧客消費之習性，試圖找出該公司顧客的價值所在。該研究首先利用 RFM 模式所改良的最近交易日期、消費次數及平均購買金額，外加平均哩程數的指標並給予權重值，作為群集化的變數依據，並以自組映射網路進行群集化分析；接著針對所選出的群集，利用關聯分析來得到航線間隱藏的規則所在；最後再利用貝氏網路 (Bayesian network) 模式構建分類預測模型，一旦日後出現新進顧客，便能快速研判該顧客的群集屬性，以利決策者迅速掌握顧客特性，進而提供適切的行銷方案。其研究成果的預測率約為 40%，研究架構如圖 3 所示。

謝文雄^[15]研究對象為某五金製造業客戶，運用資料採礦來了解顧客之消費行為，在顧客分析的討論上，則是將原來 RFM 模式之 FM 改為平均次數與平均金額，作為群集化分析之輸入變數。經由群集化可找出具有高忠誠度及高利益的目標顧客，接著運用決策樹分析了解其消費特性類型。最後運用關聯分析發現某些規則，如購買風輪架及水泥車的顧客，通常也會再購買四輪車。



資料來源：薛文佑^[14]。

圖 3 航空公司 RFM 模式應用研究流程

蔡永恆^[16]也是應用資料採礦技術中的群集化、分類及關聯分析，研究銀行顧客的消費行為，群集化部分使用 SOM 分析，而在指標選擇上將原始的 RFM 模式加以修改，R 變數修改為最近一個月來的次數，F 變數改為每月平均次數，M 變數則是平均每次交易金額，再加上 T 變數（存摺餘額）以進行群集化。此外採用決策樹建立分類模型，根據所分析的結果，其預測率高達 95%，顯示出此四項指標擁有良好的解釋能力。

Hsieh^[17]欲了解銀行的客戶行為，提出一個整合資料採礦與行為評分的分析模式，運用銀行客戶的最近交易日期 (recency, R)、交易頻率 (frequency, F)、交易金額 (monetary, M) 及償還能力 (repayment ability, RA) 作為群集化變數，採用 SOM 方法將顧客分為 16 群，並將顧客重新定義為循環型 (revolver users)、處理型 (transactor users)、便利型 (convenience users) 等三類顧客。接著進行顧客行為變數的敏感度分析，選出 10 個最有代表性的行為變數進行關聯分析，找出行為變數間的規則。

有關上述文獻對於 RFM 模式的修改或新增變數，可整理比較如表 2 所示。

表 2 相關文獻對 RFM 模式修改對照表

文 獻	研究對象	修改變數	增加權重	增加新變數
蔡永恆，2000	銀行業	◎	—	◎
謝文雄，2002	五金業	◎	—	—
薛文佺，2004	航空客運業	◎	◎	◎
Hsieh，2004	銀行業	◎	—	◎

資料來源：本研究整理。

2.3 航空公司實務課題

2.3.1 航空公司線上售票現況

隨著電子商務的普及，「提供線上售票服務」已成為航空業者網站的基本功能，根據 Forrester Research 在 2005 年訪問六家亞太區最主要航空公司的結果顯示，業界普遍認為網路訂票在未來五年內將成長三倍。而亞太區航空協會也發現，其協會會員公司的客運營收中，15.7% 來自線上售票，全日空 (ANA) 更早已超過此數 (2004 年達 22%)。在各航空公司開始對線上售票寄予厚望之際，預計在 2010 年以前，線上售票就可能占航空客運總營收的 42%。2005 年 5 月起，南非航空取消機票佣金制度，同時調降票價 4%，今後顧客向旅行社購買機票，旅行社將收取專業服務費用，而顧客也漸漸適應這種支付服務費及諮詢費的嶄新商業模式。此外，2004 年底首航臺灣的捷星航空，係近年來從澳洲進軍亞太地區的低成本航空業者 (low cost carrier, LCC)，該公司也要求旅客透過網路訂票並線上刷卡，凡是由旅行社經手者，必須另付手續費^[18]。

從前述案例可知，線上售票已是航空業者在電子商務時代的必然趨勢，而本研究實地拜訪航空公司相關主管後也發現，在線上訂票旅客與日俱增之下，業者已開始注意客戶資料及交易紀錄的潛在價值，擬積極規劃更完整的顧客資料庫及線上交易紀錄檔，進一步利用資料採礦的分析工具，挖掘過去無法掌握的商業智慧及產業情報，透過對旅客需求的了解，提供更客製化的精緻服務，以達到顧客滿意度的提升。

2.3.2 航空公司線上售票功能與特性

在線上交易機制日漸成熟與普及的環境下，各航空業者多已提供線上購票的功能，並鼓勵消費者直接線上訂位及付費，而各公司的行銷手法、目的、對象皆有所不同，因此在網站上所提供的資訊及購票流程也略有差異，購票所需填寫的資料亦因會員或非會員而異。本研究針對國人較熟悉的三家國際航空公司 (國泰、長榮、華航) 進行比較，彙整各家業者提供的資訊及要求旅客填寫的資料，如表 3 所示。

由表 3 可發現，長榮在行程資訊方面提供了多航段的選擇，給想要自助旅行的旅客更完整的行程安排，不必分次進行購票動作，最多可選擇 6 個目的地；華航則提供了中停或

直飛的選項，以利有時間壓力的旅客便於選擇。在機票資訊方面，三家航空公司皆提供相似的內容。在非會員訂票所需填寫的內容上，長榮要求務必填寫護照號碼及出生日期，代表該公司可能對非會員的資料格外重視，想進一步將其深化為會員，而三家業者都要求旅客填寫電子郵件信箱，以便透過 E-mail 提供更多相關的促銷及行程資訊；在會員資料填寫部分，必填項目三家航空公司近乎相同，但國泰的語言選擇添加了語音選項，期望給旅客更客製化的貼心服務。

表 3 各航空公司線上售票的資訊提供及填寫資料

國泰航空 (CATHAY PACIFIC)		長榮航空 (EVA AIR)		中華航空 (CHINA AIRLINES)	
一、行程資訊 ➢出發地、目的地 ➢單程 或 來回 ➢出發日、回程日 ➢艙等 ➢座位數 (成人、兒童) 二、機票資訊 ➢票價 ➢機型 ➢詳細時間		一、行程資訊 ➢出發地、目的地 ➢單程 或 來回 或 多航段 ➢出發日、回程日 ➢艙等 ➢座位數 (成人、兒童) 二、機票資訊 ➢票價 ➢機型 ➢詳細時間		一、行程資訊 ➢出發地、目的地 ➢單程 或 來回 ➢出發日、回程日 ➢艙等 ➢座位數 (成人、兒童) ➢中停 或 直飛 二、機票資訊 ➢票價 ➢機型 ➢詳細時間	
非會員	會員 (馬可孛羅會)	非會員	會員 (貴賓聯誼會)	非會員	會員 (華夏哩程酬賓)
➢性別* ➢姓名* ➢聯絡電話* ➢E-mail* ➢喜好餐點	➢性別* ➢姓名* ➢國籍* ➢出生日期* ➢護照號碼* ➢語言* (書面、語音) ➢座位喜好 ➢餐點喜好 ➢個人網頁	➢性別* ➢姓名* ➢聯絡電話* ➢E-mail* ➢喜好餐點 ➢護照號碼* ➢出生日期* ➢聯絡地址	➢性別* ➢姓名* ➢國籍* ➢出生日期* ➢護照號碼* ➢語言* ➢E-mail* ➢聯絡地址* ➢聯絡電話* ➢工作職稱 ➢任職公司 ➢經常艙等	➢性別* ➢姓名* ➢聯絡電話* ➢E-mail* ➢喜好餐點	➢性別* ➢姓名* ➢國籍* ➢出生日期* ➢護照號碼* ➢語言* ➢E-mail* ➢聯絡地址 ➢聯絡電話 ➢工作職稱 ➢任職公司 ➢職業別 ➢業務別 ➢餐點喜好 ➢座位喜好
*代表必填之資料					

資料來源：本研究整理。

在非必要填寫的部分，華航及長榮甚至詢問旅客的任職單位、職稱等個人化問題，試圖透過這些資料的掌握，進一步加強其顧客關係管理。此外，旅客若加入國泰馬可孛羅會員，將可擁有個人網頁、購票紀錄查詢、個人行程等客製化的服務。而在座位方面，國泰及長榮提供線上劃位的功能，旅客可自行選擇座位，就像線上訂電影票一樣，而華航只能選擇靠窗或走道的偏好，且不保證最後的座位安排如旅客所願。

三、分析模式構建

本研究所構建的分析模式，係採用 SAS 統計軟體的 Enterprise Miner 資料採礦技術，分別進行群集化、分類及關聯分析的探討。在完成業界訪談的實務了解後，3.1 節先探討評量顧客價值的 RFM 模式，並提出本研究不同於傳統模式的修正版，以作為進行群集化及分類的資料基礎；3.2 節對群集化功能中的 K-質心法 (K-means) 演算法進行說明，其分群結果係下階段分類研究的重要依據；3.3 節介紹分類功能中的卡方自動相互檢測 (chi-squared automatic interaction detection, CHAID)，3.4 節則解釋關聯分析功能的 Apriori 演算法，而分類及關聯分析的成果，可提供航空公司制定行銷策略的決策資訊。因此，本研究所使用的分析工具及其關係，可示意如圖 4 所示。

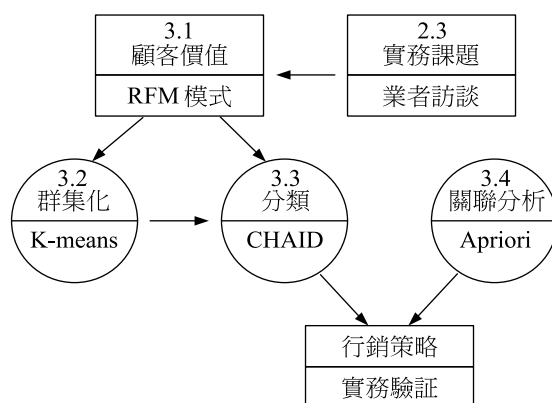


圖 4 本研究所使用的分析工具及其關係示意圖

3.1 旅客價值分析子模式 — RFM 模式

本研究首先對 Arthur^[13]所提出的傳統 RFM 模式，進行變數上的修改與增加，以期更符合航空客運業之特性。

1. 最近購買日 (recency)

在本研究中就是指旅客上次線上訂票至今的天數。若顧客最近一次購買日愈接近分析

時點，往往代表顧客對於行銷活動的刺激反應會較大，但此一現象也因產業特性而有所不同。機票對於一般人而言，其花費金額較高，故在最近購買日的權重處理也必須特別考量。

2. 購買頻率 (frequency)

當顧客購買次數愈多時，企業會認定其忠誠度及顧客價值是相對較高的。由於顧客有老顧客與新進顧客之分，如果只用購買次數來衡量顧客價值，對於新進顧客將產生價值判斷上的偏誤，進而錯失具有高潛力的新顧客，因此本研究以一定時間內的購買次數來作為衡量指標。

3. 消費金額 (monetary)

當顧客的購買金額愈高時，企業會認為其顧客價值愈高，對企業貢獻也愈多。可是如果單以一段時間內的消費總金額來看，則累積交易次數較少的顧客，其價值有被低估的可能，故本研究以「平均購買金額」作為顧客之消費金額。

此外在航空客運業上，旅客相同的消費金額可能有二種不同的解釋，如花費同為 35,000 元的二位旅客，可能前者為搭乘「臺北—東京」的頭等艙，而後者搭乘「臺北—洛杉磯」的經濟艙，此時單就消費金額來說，二位旅客的價值相同，但如果進一步探討其價值貢獻的話，航空公司會比較偏好第一位顧客。因此，本研究除了原有的 RFM 指標之外，擬再加入艙等及哩程數的指標進行探討，茲解釋如下。

4. 艙等 (class)

一般而言，航空公司將艙等分為頭等艙、商務艙及經濟艙，本研究給予每個艙等不同的分數，加總後再除以消費次數，可得到平均艙等分數，然後進行顧客價值的分析。各艙等的給分標準為經濟艙 1 分、商務艙 3 分、頭等艙 5 分，而將旅客歷次搭乘艙等的分數加總後除以搭乘次數，即為平均艙等分數。

5. 平均哩程數 (average mileage)

哩程數也是航空業應該特別考量的顧客價值指標，而只考慮累積總哩程數，對於新進顧客的價值判斷不公平，因此本研究將以平均哩程數作為指標進行探討。

本研究將傳統 RFM 模式加以修改，加上艙等及平均哩程數後，共使用五項指標進行顧客價值分析，各指標的定義及給分標準分別如表 4 及表 5 所示，在五項指標上，都是以相同時段內的交易資料來進行分析，時間為一年。由於透過 SAS 資料採礦的群集化分析，會將顧客分為多個群集，在進行分析及後續分類功能時，過多群組將造成解釋上的困難，因此對所得之群集進行重組（廣義化），經過廣義化後的資料將更有意義，也更容易解讀^[5]。前人研究如謝文雄^[15]、蔡永恆^[16]等，皆將群集化後的結果進行二構面之重組，以期獲得更容易解釋的結果，而本研究將依忠誠度及利益二個構面，對群集化後的結果進行重組。在忠誠度的評量上，是以 R、F 值作為判斷依據，若此二項指標評分皆大於整體平均值，則定義為高忠誠度顧客；若其中有一項指標低於整體平均值，便定義為低忠誠度顧客。而在利益構面的評量上，係以 M、C、Mi 三項指標作為判斷標準，若此三項指標評分有二者以上高於整體平均值，便視為高利益顧客；若有二項以上低於整體平均值，則定義

為低利益顧客。表 6 定義顧客分群、構面定義及判斷準則，高於平均數者以 ↑ 表示，反之則為 ↓。

表 4 本研究所使用 R、F、M、C、Mi 指標的定義

指 標	定 義
最近購買日 (Recency)	離分析時點最近一次上網訂票之天數
購買頻率 (Frequency)	購買次數／單位時間
平均消費金額 (Average Monetary)	消費總金額／消費總次數
艙等 (Class)	艙等總分數／消費總次數
平均哩程數 (Average Mileage)	總哩程數／消費總次數

表 5 各項顧客價值指標的給分標準

指標構面	評分	評分標準
最近購買日 (Recency)	5 分 4 分 3 分 2 分 1 分	0 ~ 40 天 41 ~ 70 天 71 ~ 130 天 131 ~ 180 天 180 天以上
購買頻率 (Frequency)	5 分 4 分 3 分 2 分 1 分	8 次以上 7 ~ 8 次 5 ~ 6 次 4 次 3 次以下
平均消費金額 (Average Monetary)	5 分 4 分 3 分 2 分 1 分	$30000 < M$ $14500 < M \leq 30000$ $12000 < M \leq 14500$ $9000 < M \leq 12000$ $M \leq 9000$
平均艙等分數 (Average Class)	5 分 4 分 3 分 2 分 1 分	$3.5 < C$ $2.5 < C \leq 3.5$ $2 < C \leq 2.5$ $1 < C \leq 2$ $C \leq 1$
平均哩程數 (Average Mileage)	5 分 4 分 3 分 2 分 1 分	$5000 < Mi$ $1700 < Mi \leq 5000$ $1150 < Mi \leq 1700$ $900 < Mi \leq 1150$ $Mi < 900$

表 6 顧客分群及判斷準則

顧客分群	顧客群組 1	顧客群組 2	顧客群組 3	顧客群組 4
構面定義	忠誠度↑ 利益↑	忠誠度↑ 利益↓	忠誠度↓ 利益↑	忠誠度↓ 利益↓
R/F/M/C/Mi	↑↑↑↑↑ ↑↑↑↑↓ ↑↑↑↑↑ ↑↑↑↓↑ ↑↑↓↑↑	↑↑↓↓↓ ↑↑↓↑↑ ↑↑↓↑↓ ↑↑↓↑↓ ↑↑↑↓↓	↑↓↑↑↑ ↓↑↑↑↓ ↓↓↑↑↑ 等	↑↓↓↓↓ ↓↑↑↓↓ ↓↓↑↑↓ 等
判斷準則	忠誠度高： R、F 值皆高於整 體平均值	忠誠度高： R、F 值皆高於整 體平均值	忠誠度低： R、F 值有一項以 上低於整體平均	忠誠度低： R、F 值有一項以 上低於整體平均
	利益高： M、C、Mi 三者間 有二項以上高於整 體平均值	利益低： M、C、Mi 三者間 有二項以上低於整 體平均值	利益高： M、C、Mi 三者間 有二項以上高於整 體平均值	利益低： M、C、Mi 三者間 有二項以上低於整 體平均值

3.2 旅客群集化分析子模式 — K-means 法

K-means 法是由 Mac Queen^[19] 於 1967 年所提出，是最早發展且最具代表性的群集化計算方式，其他的分群方法也大多以 K-means 法為基礎而加以改良或修正。K-means 法的演算方法為，先找出 K 個隨機的群集中心點，再根據各個資料點與這些群集中心點的距離，來決定哪些資料點適於合為同一群集，其距離係採歐幾里得距離 (Euclidean distance) 公式進行運算，公式如下：

$$d(I_i, I_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

物件 I_i 座標點為 $(x_{i1}, x_{i2}, \dots, x_{in})$ ；

物件 I_j 座標點為 $(x_{j1}, x_{j2}, \dots, x_{jn})$ ；

$d(I_i, I_j)$ 為 I_i 與 I_j 之距離。

隨後再根據這些群集所得之群集中心點，重新利用新的群集中心點來修正群集的結果，並反覆執行這些步驟，直到達成設定的終止條件為止，而終止條件的設定通常為一判斷函數，其公式及符號說明如下。

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - u_i|^2$$

k = 群集數目； C_i = 第 i 個群集； p = 資料集； u_i = 第 i 群集的質心。

傳統 K-means 法的缺點為必須先設定群集數目，但在群集分析前，往往無法客觀預設合理的群集個數。所幸本研究採用的 SAS Enterprise Miner 軟體，係改良後的 K-means 法，毋需事先輸入群集數，而由軟體自行運算至最佳群集數為止。K-means 法的演算步驟可說明如下，概念示意圖則如圖 5 所示。

步驟 0：隨機選擇開始的 K 個群集分割，例如設定分割成為 2 個群集。

步驟 1：將資料分配給離質心最近的群集中心，並且產生一個新的分割。如選取 P1 及 P2 二點作為資料的質心，將離 P1 及 P2 最近的其他資料分成同一群 (如 C1 及 C2)。

步驟 2：根據每一群集的資料計算新的群集中心。如將 C1 及 C2 中的資料，重新計算其質心點，出現新的質心點 P3 及 P4。

步驟 3：將離 P3 及 P4 最近的其他資料分成同一群 (如 C3 及 C4)。

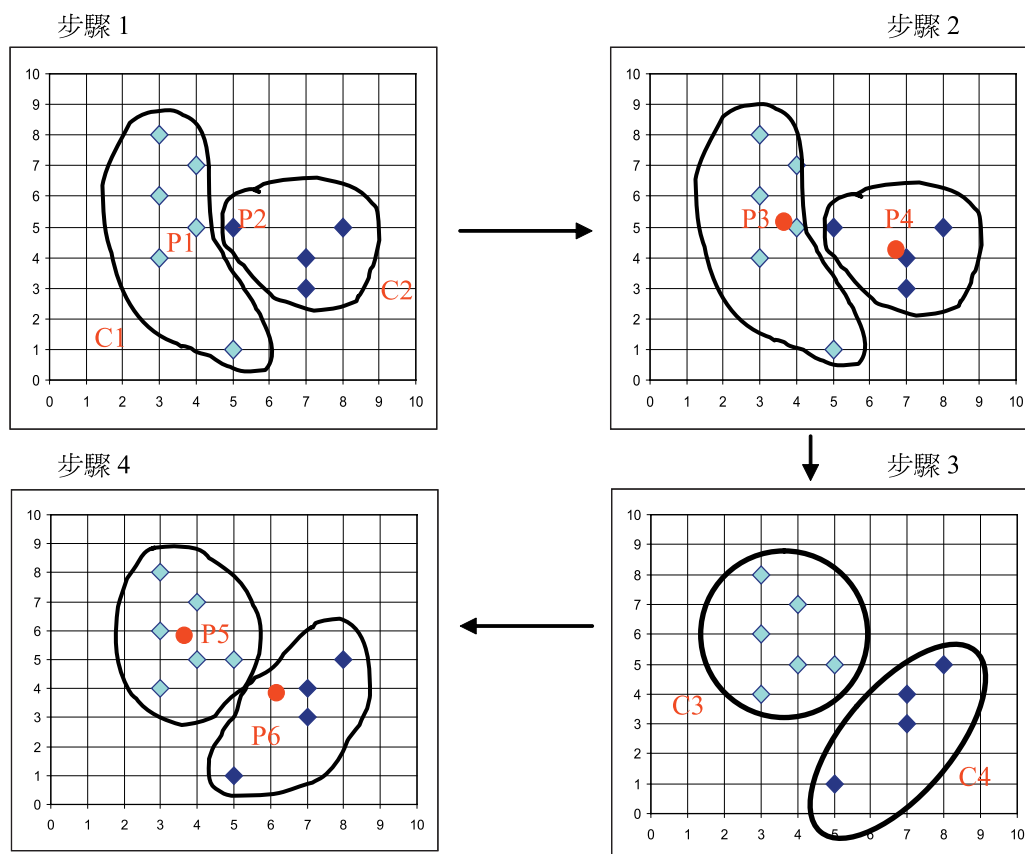


圖 5 K-means 法演算概念圖^[3]

步驟 4：再重新計算 C3 及 C4 的質心 (P5 及 P6)。重複執行步驟 3 與步驟 4 直到找出最佳解，即為無法再分割。

3.3 旅客分類分析子模式 — CHAID 法

CHAID 是由 Kass^[20]於 1980 年發展出來的一種非常有效率之統計區隔方法，可透過統計檢定來計算所有可能潛在預測變數的預測值，而 CHAID 也會自動將數值合併起來當它們對於目標變數而言是齊一／相似的，而其他的數值依然維持現狀。CHAID 在生長成決策樹時，會選擇最佳的預測變數作為第一層的節點，如此每個根節點都會根據選擇的變數，繼續分割直到整個決策樹完全成長為止。而統計測量是依據目標變數 (target variable) 的差異而有不同的計算方式，當目標變數是連續型變數 (continuous variable) 時，使用的統計檢定方法為 F 檢定，若目標變數是類目變數 (categorical variable) 時，則改用卡方檢定。卡方值之計算步驟及公式如下：

$$\chi^2 = \sum_{i \in \text{categories}} \frac{(t_0 - e_0[i])^2}{e_0[i]} + \frac{(t_1[i] - e_1[i])^2}{e_1[i]}$$

步驟 1：對每一個分類屬性均做交叉表檢定，就分類屬性 A 做 N 個屬性值與 M 個分類目標的 M×N 交叉表檢定。

步驟 2：就各個分類屬性進行排列組合，並計算卡方值及 P 值。若 P 值大於事先定義的顯著水準，則將兩個屬性值合併，並重複做排列組合及交叉表檢定，直到 P 值均小於顯著水準或分類屬性經合併後，僅剩下兩個組合。

步驟 3：從合併後的分類屬性中，選 P 值最小者，依該分類屬性在步驟 2 的合併結果對資料進行分割。

步驟 4：分割後的資料分別重複以上步驟，直到節點中包含的資料數小於事先定義的數值或 P 值大於顯著水準為止，決策樹就不再繼續成長。

3.4 航線關聯分析子模式 — Apriori 演算法

關聯法則 (association rule) 主要在探討龐大資料庫中某些項目彼此之間的關聯性，故經常被用來研究商品搭配銷售的行銷價值，應用在航空公司的課題上，則可挖掘各類型旅客對不同航線的偏好，進而發現原先未知的消費者行為或航線選擇特性，作為精準行銷的決策依據。關聯法則的基本定義有以下三項：

1. 項目集 (item set)

亦即商品項目所構成的集合。一個集合代表某些特定商品的組合，以航空客運業而言，其項目內容可包含航線起迄點 (空間特性) 及班機起降時段 (時間特性) 等。

2. 支持度 (support level)

係指某項目集在所有顧客交易中出現的機率。例如在 1,000 筆的超市消費紀錄中，有 300 筆交易被發現同時含有 {牛奶, 麵包} 的購買品項，則該項目集的支持度為 $300/1000 = 30\%$ 。支持度愈高，表示這種組合過去被購買的頻率愈高，就機率公式來解釋如下：

$$\text{Support}(A \Rightarrow B) = P(A \cap B)$$

3. 信賴度 (confidence level)

代表被實驗的因子在主要因子中所占有的條件比例，如前例中購買牛奶者共有 600 筆消費紀錄，而其中又有 300 筆同時購買麵包，則其信賴度為 $300/600 = 50\%$ 。就機率公式來解釋如下：

$$\text{Confidence}(A \Rightarrow B) = P(B | A)$$

圖 6 即為 Apriori 演算法的計算概念流程，其演算步驟可舉簡例示範如下：

- 步驟 1：將顧客的消費紀錄列表，其中 ID 代表顧客編號，ITEM 則是產品代號，因此從第一張表可看出 ID 編號為 1 的顧客購買了代號 2、5、7 這三項產品。接著暫定信賴度為 70% 且支持度為 40%，亦即在 5 位顧客中，某項目集的購買者不得少於 2 人。
- 步驟 2：先設定項目集的元素從一個開始，並計算各項目集被購買的次數，如項目集 {1} 被購買了 1 次，項目集 {2} 被購買了 3 次 (第二張表)。
- 步驟 3：剔除未滿足所設定之支持度的項目集，故項目集 {1} 被刪除 (第三張表)。
- 步驟 4：針對剩下的項目集，將其擴充為任意 2 個產品的組合，如 {2,3}、{2,4} 等。接著計算各項目集出現交易的次數，如 ID 編號 1 及 4 均包含項目集 {2,5}，故其次數為 2 (第四張表)。
- 步驟 5：剔除未滿足支持度的項目集，於是只剩下 {2,5} 及 {2,7} 達到支持度的標準 (第五張表)。
- 步驟 6：再進行 3 個元素的項目集測試，結果發現僅有的 {2,5,7} 只剩 1 位顧客，故停止運算。
- 步驟 7：在項目集 {2,5} 的情況下，顧客有可能先買產品 2 再買產品 5 (2→5) 或先買產品 5 再買產品 2 (5→2)，所以必須進行信賴度測試。由於在購買產品 2 的條件下 (ID 為 1、3、4) 也會購買產品 5 (ID 為 1、4) 的信賴度為 $2/3$ ，未達所設定的 70% 信賴度，故予以刪除。同理可算出 5→2、2→7、7→2 的信賴度，最後滿足信賴度標準的為 5→2 及 7→2。
- 步驟 8：若前述步驟所得之關聯分析結果，無法發現較隱藏之規則，可重新調整其支持度及信賴度標準，再重複進行步驟 1~7，以達到較佳之分析結果。

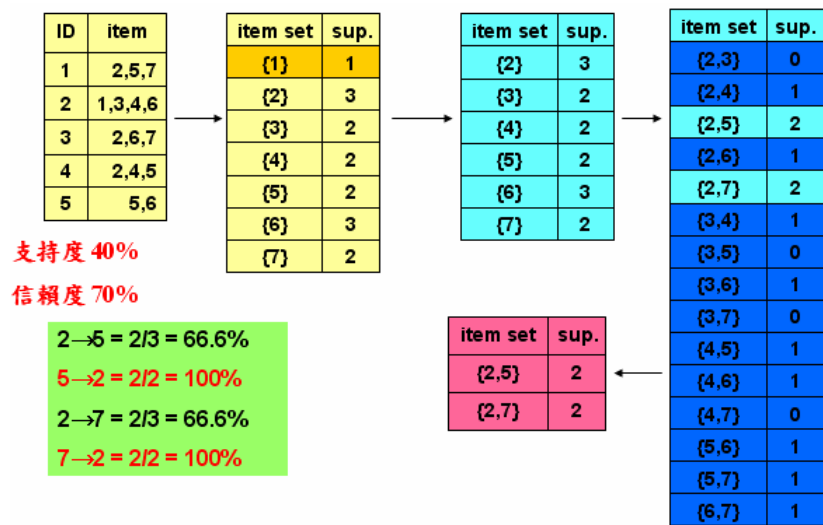


圖 6 Apriori 演算法概念圖

四、模式應用及分析

由於航空公司交易紀錄與顧客資料均屬業者重要的商業機密，難以取得完整的實際資料，故本研究改以模擬資料進行分析，並依據分析的複雜度陸續解除限制。首先解除旅客的購票張數限制，允許一次購買多張票，接著取消方向限制，即旅客可購買來回票，最後則解除會員資格限制，讓非會員旅客也可線上購票，因此共可分為四種購票型態，如圖 7 所示。

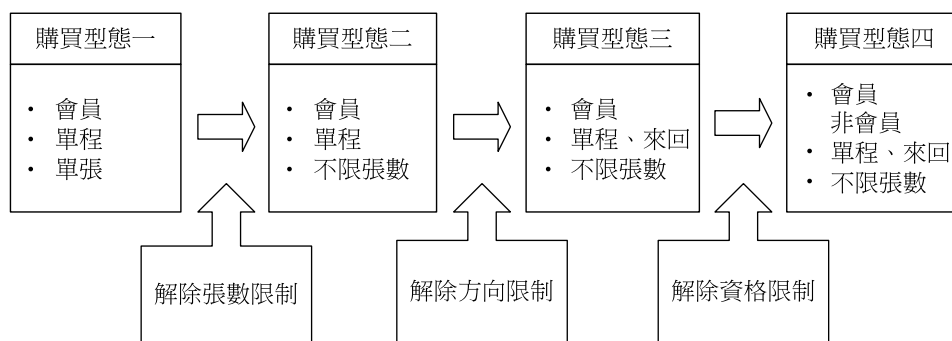


圖 7 本研究對購票型態解除限制的程序圖

4.1 資料模擬

有關本研究所需資料之模擬準則及處理流程，可如圖 8 所示。以下 4.1 節將說明本研究如何參考觀光年報、民航局資料及航空公司年報等進行資料模擬，並在航空公司現有線上購票欄位外，新增本研究所需之資料欄位，4.2 節則說明這些交易紀錄如何進行資料處理及轉換。

4.1.1 模擬準則

雖然本研究無法取得真實完整的航空旅客交易紀錄與會員資料，但在實地拜訪某國籍航空公司有關部門深入了解後，發現根據交通部觀光局、民航局及該航空公司的相關統計資料，可自行模擬航空旅客的會員資料及線上交易紀錄，其「仿真」程度並不影響本研究之進行。因此，本研究係以下列準則進行所需資料庫之構建，其模擬概念可如圖 9 所示。

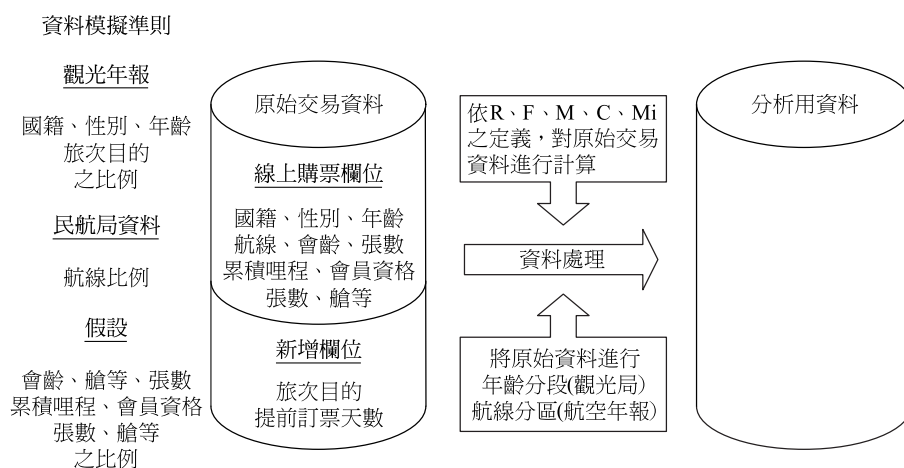


圖 8 資料模擬及處理流程

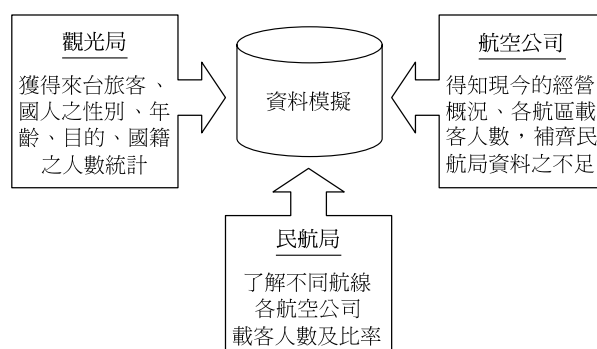


圖 9 旅客資料模擬概念圖

1. 觀光局

本研究首先依據交通部觀光局的「2004 年觀光年報」及與臺灣觀光相關之統計資料，掌握來臺旅客的國籍、性別、目的、年齡及職業分布等，並從來臺旅客年齡與目的之交叉分析，了解外籍旅客的消費特性，再加上國人出國目的地、性別、年齡、目的之人數統計與比例，可作為本研究構建旅客基本資料庫的樣本基礎。

2. 民航局

交通部民航局提供關於民航運輸的統計數據，其中「臺灣地區各機場國際航線班機載客率」的資料，可幫助本研究了解臺灣各航空公司所有航線的出、入境旅客人數，作為本研究模擬不同航線旅客需求量及其特性之依據。

3. 航空公司

各航空公司每年均公布「年度報告」，由其內容可了解目前的經營概況，並得知不同航區占總旅客人數之分布比例。由於民航局在某些航線載客資料上未必盡列，需要航空公司的年度報告加以補強，因此本研究模擬之旅客量，係以民航局資料為基準，再用航空公司年報補其不足。

在上述準則的模擬下，本研究應用 Microsoft Office 之 Access 軟體作成資料庫，共可分為顧客資料表、交易資料表、及航線資料表等三部分，如圖 10 所示。

1. 顧客資料表

包括顧客編號 (ID)、國籍 (Nationality)、性別 (Sex)、年齡 (Age)、會員資格 (Club)、會齡 (Club_year)、累積哩程 (Acc_mile) 等。

2. 交易資料表

包括顧客編號 (ID)、購買總張數 (Ticket)、成人張數 (Adult)、兒童張數 (Child)、消費總金額 (Money)、搭乘艙等 (Class)、旅次目的 (Purpose)、出發日期 (Dep_Time)、航線編號 (Line) 等。

3. 航線資料表

包括航線編號 (Line)、起點 (Departure)、訖點 (Arrival)、哩程數 (Mile) 等。

由於會員資格及會齡、艙等、購票張數、累積哩程的真實資料無法取得，本研究係在航空業者的指導及協助下進行資料假設。在會齡的設定上，較常搭乘頭等艙及商務艙的顧客，會齡的範圍在 5~10 年間，而經濟艙顧客則設定為 2~3 年；由於上網訂票日與實際出發日之相隔天數，可能為重要之消費者特性，故本研究加入「提前訂票天數 (Dif_Day)」變數進行探討，並假設商務旅客的提前訂票天數為 1~15 天，觀光與探親旅客分別為 15~40 天及 20~40 天，而其他旅次目的則為 10~40 天。各項資料欄位所代表的意義，可如表 7 所示。

4.1.2 資料處理

在進行群集化及分類之前，必須先將旅客線上購票紀錄進行資料處理及轉換，故本研究進行以下之資料處理：

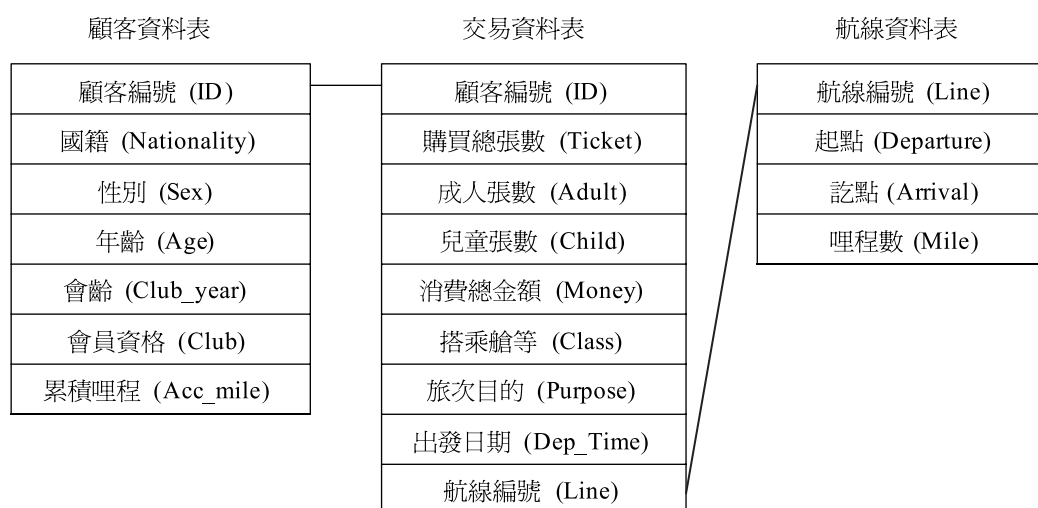


圖 10 本研究所構建之資料結構及其欄位

表 7 各資料欄位所代表的意義說明

欄位名稱	代表意義
顧客代碼 (ID)	購票顧客之編號
購買總張數 (Ticket)	顧客購買張數 (包含成人、兒童)
成人張數 (Adult)	購買成人票張數
兒童張數 (Child)	購買兒童票張數
消費總金額 (Money)	此次購票之總金額
搭乘艙等 (Class)	顧客所搭乘之艙等
旅次目的 (Purpose)	顧客出國之旅次目的
出發日期 (Dep_Time)	班機出發之日期
航線編號 (Line)	搭乘航線之編號
國籍 (Nationality)	顧客之國籍
性別 (Sex)	顧客之性別
年齡 (Age)	顧客之年齡
會員資格 (Club)	顧客是否為會員
會齡 (Club_year)	顧客成為航空公司會員之年數
提前訂票天數 (Dif_Day)	出發日期與訂票日期相距之天數
累積哩程 (Acc_mile)	會員所搭乘之總哩程數

1. 顧客價值

分別計算顧客距離上次訂票的天數、消費頻率、平均消費金額、平均艙等分數及平均哩程數，並依據評分標準給予分數，求出每位顧客之 R、F、M、C、Mi 指標分數。

2. 年齡分段

將顧客之年齡進行分段化 (segmentation)，其年齡分段範圍以交通部觀光局之觀光年報為基準，例如編號 1 代表 10~19 歲、2 代表 20~29 歲，以此類推。將顧客年齡分段有利於進行分類分析之解釋能力。

3. 航線分區

將所有航線依航空公司年報中各航區營運報表之分類方式進行分區，如代碼 1 為香港，編號 2 代表東北亞，共分為六大航區。在分類模式中，將針對不同航區進行分析，以了解其市場特徵。

4.2 線上購票旅客之群集化分析

航空公司以往對顧客屬性並無特別的區隔方式，造成無法對不同顧客群進行更客製化的行銷服務，且大多以特定期間內搭乘的總哩程數及高價艙等搭乘次數，作為顧客價值高低的標準，因此在顧客價值判斷上稍嫌不足。本節探討在四種不同的購票型態下，分別對各價值指標給予不同的評分標準，以便觀察其群集化分析結果的異同。本研究採用 SAS 資料採礦功能的 Enterprise Miner 進行群集化分析，將欲判斷的指標 (R、F、M、C、Mi) 作為輸入變數，軟體將自行判斷最適的群集結果，因此在不同型態下所判斷出之群集數也不盡相同。

4.2.1 型態一：現有會員限訂購一張單程票

購票型態一的主要限制在於顧客必須是會員，有詳細的個人資料 (如出生日期、性別、國籍)，每次上網只能訂購一張機票，且限購單程，各價值指標的評分標準如表 5 所示。

透過 Enterprise Miner 的群集化技術，型態一共可得到 10 個顧客群集，而為了方便後續分析、解釋及提供行銷建議，本研究將依忠誠度及利益二個構面，對群集化結果進行重組，在忠誠度的評量上是以 R、F 值作為判斷依據，而利益構面則採用 M、C、Mi 等三項指標，其判斷標準已如表 6 所示。因此原來的 10 個顧客群集，可廣義化為較明確的 4 群顧客，如表 8 所示。

4.2.2 型態二：現有會員可訂購多張單程票

型態二解除購票張數的限制，每位顧客可一次購買多張單程票，於是造成平均消費金額有上升的趨勢，故必須調整對平均消費金額的評分標準。透過群集化技術，可得到 15 個顧客群集，同樣運用忠誠度及利益二個構面，對群集化後之結果進行重組，所得之結果如表 9 所示。

表 8 型態一顧客群集化之結果

群集	次數	R	F	M	C	Mi	R/F/M/C/Mi	顧客群組
1	50	2.42	3.68	1.82	2.46	1.76	↓↑↓↑↓	4
2	12	3.67	4.67	4.17	3.33	3.67	↑↑↑↑↑	1
3	13	4.54	2.31	4.69	4.00	4.00	↑↓↑↑↑	3
4	64	2.19	1.70	1.50	1.64	1.42	↓↓↓↓↓	4
5	62	4.65	3.92	1.81	2.18	1.73	↑↑↓↑↓	2
6	29	2.21	2.59	2.79	1.48	3.48	↓↓↑↑↑	3
7	13	1.85	2.38	4.85	4.00	3.85	↓↓↑↑↑	3
8	61	4.33	1.85	1.51	2.10	1.43	↑↓↓↓↓	4
9	24	4.42	2.25	3.08	1.42	3.88	↑↓↑↑↑	3
10	7	3.52	3.81	3.44	1.71	3.43	↑↑↑↑↑	1

表 9 型態二顧客群集化之結果

群集	次數	R	F	M	C	Mi	R/F/M/C/Mi	顧客群組
1	16	3.94	2.13	1.19	1.13	1.19	↑↓↓↓↓	4
2	23	4.17	2.43	3.39	1.83	1.87	↑↓↑↓↓	3
3	17	4.06	2.35	3.88	1.18	4.76	↑↓↑↑↑	3
4	27	1.93	3.04	4.70	4.56	4.59	↓↑↑↑↑	3
5	27	4.19	4.81	2.89	1.85	3.85	↑↑↓↑↑	2
6	14	4.43	2.29	4.57	4.71	4.71	↑↓↑↑↑	3
7	20	1.60	1.75	1.80	3.70	1.80	↓↓↓↓↓	4
8	33	3.24	4.12	2.18	3.33	1.91	↑↑↓↑↓	2
9	28	1.57	1.61	1.43	1.07	1.89	↓↓↓↓↓	4
10	28	1.61	3.96	2.21	1.79	2.71	↓↑↓↑↓	4
11	18	2.39	2.50	3.67	3.28	1.61	↓↓↑↑↓	3
12	22	3.64	2.77	2.23	2.41	3.41	↑↓↓↑↑	4
13	32	1.69	2.38	3.97	1.50	4.47	↓↓↑↑↑	3
14	11	4.73	1.73	2.09	3.45	1.36	↑↓↑↑↓	4
15	19	4.53	4.53	3.74	3.58	3.32	↑↑↑↑↑	1

4.2.3 型態三：現有會員可訂購多張來回票

型態三解除方向限制後，每位顧客可選擇購買來回票，也造成平均消費金額上升的現象，而來回票哩程是單程票的雙倍，故平均哩程數也會提高，因此這兩項指標亦應給予新

的評分標準。型態三的群集化分析共可得到 15 組顧客群，經過重組後的分群結果如表 10 所示。

表 10 型態三顧客群集化之結果

群集	次數	R	F	M	C	Mi	R/F/M/C/Mi	顧客群組
1	24	2.04	2.79	1.33	4.13	1.58	↓↓↓↑↓	4
2	7	1.29	4.71	4.43	2.43	4.14	↓↑↑↓↑	3
3	15	4.33	1.93	3.20	3.87	4.33	↑↓↑↑↑	3
4	25	4.48	2.12	1.52	1.40	1.64	↑↓↓↓↓	4
5	26	2.15	2.19	4.54	3.85	4.31	↓↓↑↑↑	3
6	30	2.17	3.93	2.43	3.40	3.23	↓↑↓↑↑	3
7	11	1.27	1.55	3.18	3.45	2.18	↓↓↑↑↓	3
8	9	4.44	2.22	3.67	3.44	1.78	↑↓↑↑↓	3
9	48	1.81	2.31	1.56	1.21	1.96	↓↓↓↓↓	4
10	25	3.96	3.84	4.20	2.60	4.56	↑↑↑↑↑	1
11	29	4.17	4.31	1.79	1.90	2.97	↑↑↓↑↑	2
12	29	1.76	2.14	4.21	1.21	4.38	↓↓↑↑↑	3
13	19	4.74	3.53	1.37	4.00	1.37	↑↑↓↑↓	2
14	9	4.00	1.33	3.56	1.22	4.67	↑↓↑↑↑	3
15	29	3.59	4.17	3.10	2.38	1.97	↑↑↑↓↓	2

購票型態四係解除會員資格的限制，會員或非會員均可訂票，所造成的影響為會齡及累積哩程等欄位。在群集化分析上也採用最近購買日 (R)、購買頻率 (F)、平均消費金額 (M)、平均艙等分數 (C)、平均哩程數 (Mi) 等，由於這五項指標皆與有無會員資格限制無關，故本研究發現其群集化結果與購買型態三雷同。

4.2.4 各購票型態群集化結果之比較分析

經由前節分析可知，購票型態二解除張數限制後，「平均消費金額」普遍上升，而型態三再開放旅客購買來回票後，「平均消費金額」及「平均哩程數」均會增加，因此三種購票型態的指標評分標準必須有所不同，其差異可比較如表 11 所示。

本研究將不同購票型態下各群組之「平均年齡」、「最近購買日」、「頻率」等九大特性進行比較，結果如表 12 所示。經過比較可發現在三種型態下，群組 1 顧客之「平均年齡」、「頻率」、「平均艙等分數」、「會齡」多為最高，且「平均提前訂票天數」較短，顯示此群組之顧客忠誠度較高，且會員較資深，行程規劃天數也較短，為航空公司最

有價值之顧客。而群組 4 正好相反，忠誠度較低，顧客較年輕，購票頻率不高，行程規劃天數較長，是對業者價值貢獻最低的顧客。從表 12 發現，透過 R、F、M、C、Mi 等五項指標對顧客的群集化分析，可有效達到顧客區隔之目的，讓航空公司得以研擬客製化的行銷策略。

表 11 不同型態評分差異比較

指標	評分	型態一	型態二	型態三
平均消費金額 (average monetary)	5 分	$30000 < M$	$45000 < M$	$60000 < M$
	4 分	$14500 < M \leq 30000$	$25000 < M \leq 45000$	$30000 < M \leq 60000$
	3 分	$12000 < M \leq 14500$	$15000 < M \leq 25000$	$20000 < M \leq 30000$
	2 分	$9000 < M \leq 12000$	$10000 < M \leq 15000$	$13000 < M \leq 20000$
	1 分	$M \leq 9000$	$M \leq 10000$	$M \leq 13000$
平均哩程數 (average mileage)	5 分	$5000 < Mi$	$5000 < Mi$	$6500 < Mi$
	4 分	$1700 < Mi \leq 5000$	$1700 < Mi \leq 5000$	$2000 < Mi \leq 6500$
	3 分	$1150 < Mi \leq 1700$	$1150 < Mi \leq 1700$	$1500 < Mi \leq 2000$
	2 分	$900 < Mi \leq 1150$	$900 < Mi \leq 1150$	$1000 < Mi \leq 1500$
	1 分	$Mi < 900$	$Mi < 900$	$Mi < 1000$

表 12 不同購買型態群組特性分析比較表

群組	平均年齡	最近購買日	頻率	平均金額	平均艙等	平均哩程	平均提前訂票天數	會齡	累積哩程
型態一									
1	48.3	67.1	10.6	32813.3	2.3	5026.6	22.7	4.4	22727.6
2	46.1	42.1	8.7	12056.1	1.8	1329.9	21.9	3.6	4736.7
3	42.0	120.8	4.9	33632.7	2.1	5693.9	25.7	4.1	24209.6
4	43.3	129.7	4.9	11007.9	1.8	1078.0	24.5	3.6	3854.2
型態二									
1	49.9	36.8	9.7	37938.5	2.5	2318.9	15.9	4.8	10952.6
2	47.9	69.4	9.3	17756.3	1.9	1735.6	22.6	3.7	5998.2
3	42.5	110.9	4.9	43256.4	2.1	3999.0	24.3	4.1	17331.3
4	42.2	134.2	4.8	12128.7	1.6	1150.3	26.1	3.3	3753.6
型態三									
1	48.2	54.6	8.0	61384.2	1.9	7330.2	21.8	3.8	28246.2
2	46.2	52.4	8.4	19510.4	1.9	1578.7	23.2	3.7	5481.2
3	42.7	132.5	5.1	45289.7	2.1	4694.4	24.0	4.2	18980.1
4	42.3	131.0	4.7	13537.3	1.6	1220.2	25.8	3.1	3713.2

若運用忠誠度及利益二構面，對四群顧客進行定位，可區隔其客群差異如圖 11 所示。其中群組 1 是高忠誠度、高利益之核心顧客，群組 2 為高忠誠度、低利益之有效顧客，群組 3 則是低忠誠度、高利益之游離顧客，而群組 4 代表低忠誠度、低利益之邊緣顧客。

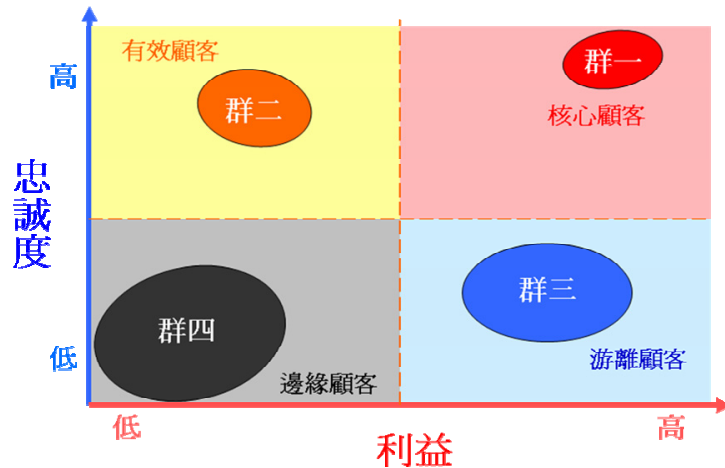


圖 11 各群組的旅客特性比較

4.3 線上購票旅客及航線之分類分析

本研究運用資料採礦的 CHAID 決策樹技術，分別對航空公司旅客及航區（航線）進行分類分析。在顧客導向方面，係以前節所區隔出的四種顧客群組為基礎，運用分類技術自動篩選有效的預測變數（如年齡區隔、平均哩程等），進而建立新進顧客的分類預測模型，以期日後有新顧客出現時，便能迅速判斷該顧客屬於何種群組，以利決策者精準掌握消費者的特性及需求，顧客導向的分類模型可示範如圖 12 所示。在產品導向方面，也可運用分類技術對航區（航線）進行旅客特性的歸納，經由分類變數的過濾，構建航區（航線）的分類模型，凡是符合某一屬性組合（如國籍、年齡區隔及提前訂票天數等）的消費者，即可預判其最可能屬於某航線的旅客，產品導向的分類模型可示範如圖 13 所示。

4.3.1 顧客導向分類模型

1. 型態一：現有會員限訂購一張單程票

針對型態一的購票情境，決策樹共選取五項變數，第一層的預測變數為「平均哩程」，第二層包括「平均哩程」、「平均金額」及「平均艙等分數」，最後一層則是「平均提前訂票天數」及「年齡區隔」，該預測模型共產生 13 個分類規則，詳如表 13 所示。其中，規則 1 表示若旅客的平均搭乘哩程數小於 1192.48 哩，且金額小於 9605.55 元，則判斷該旅客屬於群組 4 的正確率可達 92.8%（其他規則依此類推）。表 14 顯示購票型態一的分類結

果，如群組 1 實際共有 19 位顧客，但預測模型將其中 5 位誤判為群組 3，而群組 3 及群組 4 的分類最接近實際結果，故整體分類正確率約為 83%。有關購票型態一的完整決策樹，可如圖 14 所示。

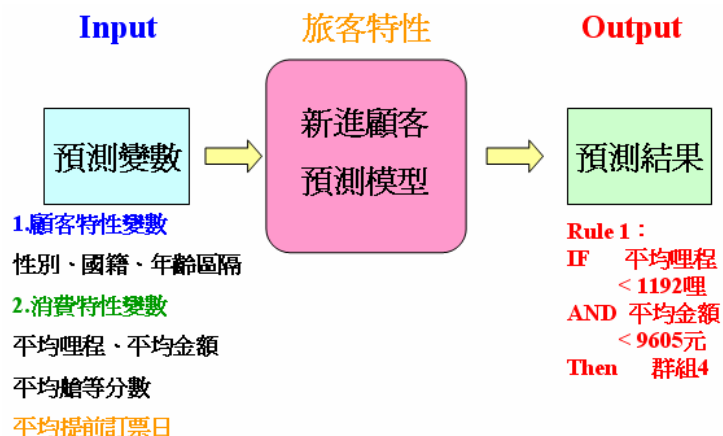


圖 12 顧客導向的分類模型示範

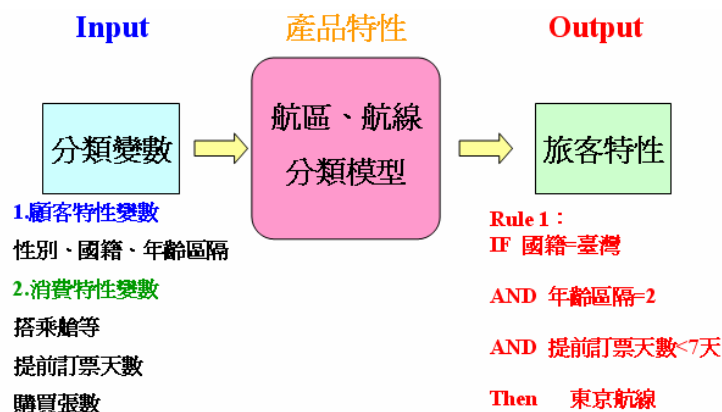


圖 13 產品導向的分類模型示範

2. 型態二：現有會員可訂購多張單程票

分類決策樹在型態二的購票情境下共使用四項變數，第一層為「平均金額」(不同於型態一的「平均哩程」)，第二層變數包含「平均哩程」及「平均提前訂票天數」，第三層則為「平均哩程」、「平均提前訂票天數」及「年齡區隔」，共產生 17 個分類規則，其整體分類正確率約為 79%。

表 13 購票型態一的決策樹分類規則

規則編號	規 則	群組	正確率
1	IF 平均哩程 < 1192.48 AND 平均金額 < 9605.55	4	92.8%
2	IF 平均哩程 < 1192.48 AND 9605.55 < 平均金額 < 13945.00 AND 平均提前訂票日 < 12	2	71.4%
3	IF 平均哩程 < 1192.48 AND 9605.55 < 平均金額 < 13945.00 AND 12 < 平均提前訂票日 < 33	4	71.0%
4	IF 平均哩程 < 1192.48 AND 9605.55 < 平均金額 < 13945.00 AND 33 ≤ 平均提前訂票日	4	100.0%
5	IF 平均哩程 < 1192.48 AND 13945.00 ≤ 平均金額	4	6.3%
6	IF 1192.48 < 平均哩程 < 1334.21	4	66.7%
7	IF 1334.21 < 平均哩程 < 1997.00	2	60.5%
8	IF 1997.00 ≤ 平均哩程 < 2984.40	4	81.0%
9	IF 2984.40 ≤ 平均哩程 AND 平均艙等分數 < 2.04 AND 年齡區隔 = 1、2、3	3	94.4%
10	IF 2984.40 ≤ 平均哩程 AND 平均艙等分數 < 2.04 AND 年齡區隔 = 4	3	100.0%
11	IF 2984.40 ≤ 平均哩程 AND 平均艙等分數 < 2.04 AND 年齡區隔 = 5、6	1	100.0%
12	IF 2984.40 ≤ 平均哩程 AND 2.04 < 平均艙等分數 < 2.83	1	100.0%
13	IF 2984.40 ≤ 平均哩程 AND 2.83 ≤ 平均艙等分數	3	87.5%

表 14 購票型態一的分類結果及預測正確率

實際	預 測			
	群 1	群 2	群 3	群 4
群 1	14	0	5	0
群 2	0	31	0	31
群 3	0	1	76	2
群 4	0	18	0	157
分類正確率：83%				

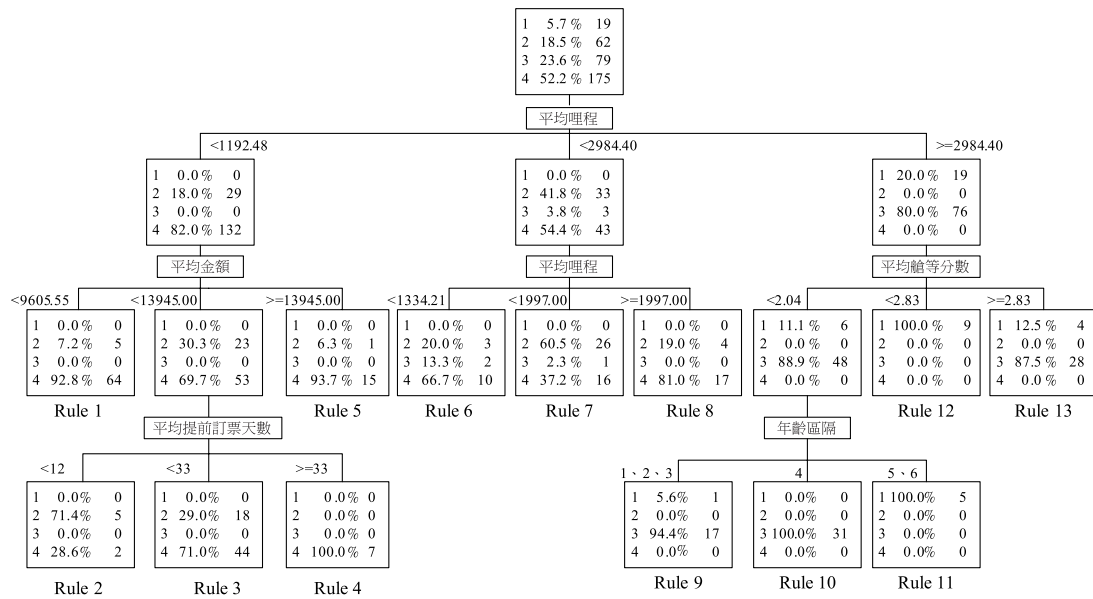


圖 14 購票型態一的決策樹分類結果

3. 型態三：現有會員可訂購多張來回票

決策樹僅包含三項變數，第一層選取「平均哩程」(其變數範圍明顯大於型態一)，第二、三層皆為「平均金額」及「平均艙等分數」，共產生 15 個分類規則，整體分類正確率 71%則低於前述二種購票情境。

4. 型態四：購票方式完全無限制

型態四解除會員限制，故增加「會員資格」變數進行分類分析，共選入 4 項變數，第一層為「平均哩程」，第二層為「平均艙等分數」，第三層則包含「平均金額」及「會員資格」，共產生 14 個分類規則，整體分類正確率達 73%，較型態三佳。

表 15 整理經由前述分析可知，分類決策樹在各種購票型態下所選取的變數項互有異同，而分類正確率依序為 83%、79%、71%、73%，因此可發現隨著購買型態愈來愈複雜，模型的分類正確率也逐漸降低，而型態四在解除會員限制下，加入「會員資格」變數，故其正確率有小幅提升。由此可見，若航空公司在網頁設計時，加入更多旅客消費特性(如出國天數、旅費預算等)的填寫，將可能有助於分類結果準確性之提升。

4.3.2 產品(航區及航線)導向分類模型

1. 不同航區之分類模型

本節先從航空公司較大範圍的飛航區域，運用決策樹進行分類分析，以挖掘各航區的消費者特性。各航區的劃分係根據航空業者的年度營運報表為準，並以前節購票型態四的

表 15 各型態分類結果

購買型態	預測變數	正確率
型態一	✓ 平均哩程 ✓ 平均金額 ✓ 平均艙等分數 ✓ 平均提前訂票天數 ✓ 年齡區隔	83%
型態二	✓ 平均哩程 ✓ 平均金額 ✓ 平均提前訂票天數 ✓ 年齡區隔	79%
型態三	✓ 平均哩程 ✓ 平均金額 ✓ 平均艙等分數	71%
型態四	✓ 平均哩程 ✓ 平均金額 ✓ 平均艙等分數 ✓ 會員資格	73%

情境下進行分析。決策樹共產生三層架構並使用四項變數，第一層為「國籍」，第二層選取「艙等」、「提前訂票天數」及「國籍」，第三層變數為「性別」、「提前訂票天數」、「國籍」，該預測模型共產生 16 個分類規則，如表 16 所示。

由表 16 可知，偏好往返航區 3 (東南亞) 的旅客預測規則共有三個 (如圖 15 所示)，其中規則 1 為臺灣籍 (國籍變數) 搭乘經濟艙 (艙等變數) 的男性 (性別變數)，但其判斷能力欠佳，正確率僅 33.6%，而規則 5 及規則 15 分別有 82.1% 及 95.3% 以上，故有較高的參考價值。表 17 顯示各航區的分類結果，其整體分類正確率約為 53%，其中東南亞及美加航區呈現較佳的預測能力，澳洲航區則因樣本數較少而無法有效掌握消費者的特性。不同航區的完整決策樹，詳如圖 16 所示。

2. 主要航線之分類模型

除了航區分類模型外，本研究也針對某國籍航空公司載客量前五大的主力航線進行分析，分別為香港、東京、曼谷、洛杉磯及溫哥華航線。結果決策樹共選取五項分類變數，第一層為「國籍」，第二層包含「年齡區隔」及「國籍」，第三層選取「艙等」及「提前訂票天數」，共產生 15 個分類規則，其整體分類正確率約為 61%，優於前述航區別的 53%。由於地理範圍較大的航區，其消費者的同質性原本就不如單一航線的聚焦性，因此本研究認為以該公司高度仰賴前五大市場的營運特性而言，今後應著重於主力航線別的分類分析，再輔以航區別的預測結果，將可有效判斷各目標市場的旅客面貌，使業者更有機會進行精準行銷。

表 16 航區別決策樹分類規則

規則編號	規 則	航區	正確率
1	IF 國籍=1 AND 艙等=1 AND 性別=1	3	30.0%
2	IF 國籍=1 AND 艙等=1 AND 性別=2	4	33.6%
3	IF 國籍=1 AND 艙等=2 AND 提前訂票天數<17	1	39.6%
4	IF 國籍=1 AND 艙等=2 AND $17 < \text{提前訂票天數} < 38$	4	33.5%
5	IF 國籍=1 AND 艙等=2 AND $38 \leq \text{提前訂票天數}$	3	82.1%
6	IF 國籍=1 AND 艙等=3	4	86.7%
7	IF 國籍=2 AND 提前訂票天數 < 12	2	88.0%
8	IF 國籍=2 AND $12 < \text{提前訂票天數} < 15$	1	100.0%
9	IF 國籍=2 AND $15 \leq \text{提前訂票天數}$	2	96.0%
10	IF 國籍=3 AND 提前訂票天數 < 2	1	60.0%
11	IF 國籍=3 AND $2 < \text{提前訂票天數} < 4$	2	100.0%
12	IF 國籍=3 AND $4 \leq \text{提前訂票天數}$	1	86.4%
13	IF 國籍=4	4	100.0%
14	IF 國籍=5	2	82.0%
15	IF 國籍=6	3	95.3%
16	IF 國籍=7	5	50.0%

表 17 不同航區分類結果

實際	預測					
	香港	東北亞	東南亞	美、加	歐洲	澳洲
香港	246	23	90	108	0	0
東北亞	123	233	95	157	0	0
東南亞	24	0	238	91	0	0
美、加	24	0	84	331	0	0
歐洲	30	0	36	38	8	0
澳洲	0	0	17	0	8	0
分類正確率：53%						

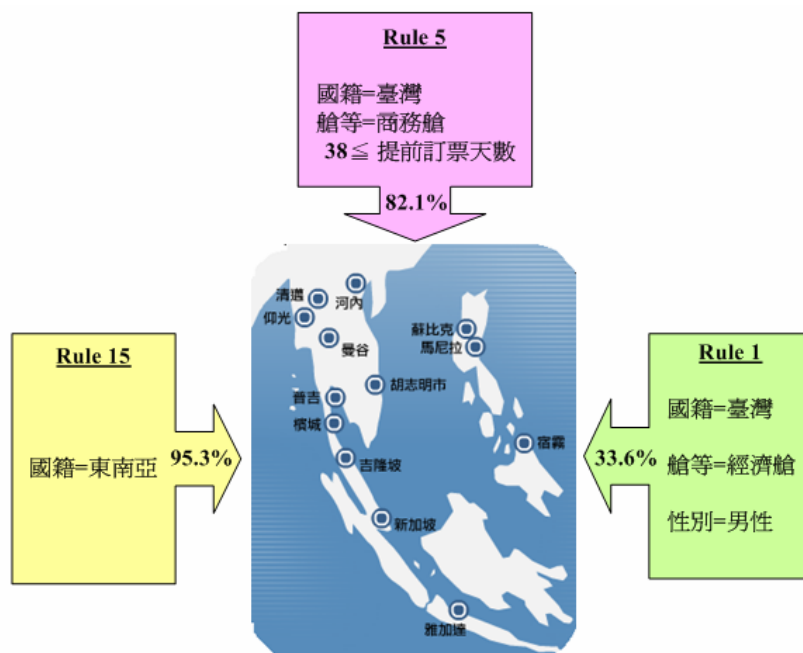


圖 15 前往東南亞 (航區 3) 之消費者特性

4.4 線上購票旅客之關聯分析

關聯分析最常被用來了解超市商品的搭售可能性，因此若將航空公司各航線也視為超市中的商品，則利用關聯規則分析來探索所有顧客上網的購票紀錄，也可能找出旅客在不同航線之間的消費關聯偏好。而旅次目的是航空旅客最重要的需求特性之一，故本研究除了進行不同購票型態的關聯分析外，也將針對各種旅次目的（商務、觀光、探親及其他），比較其關聯分析結果的差異。

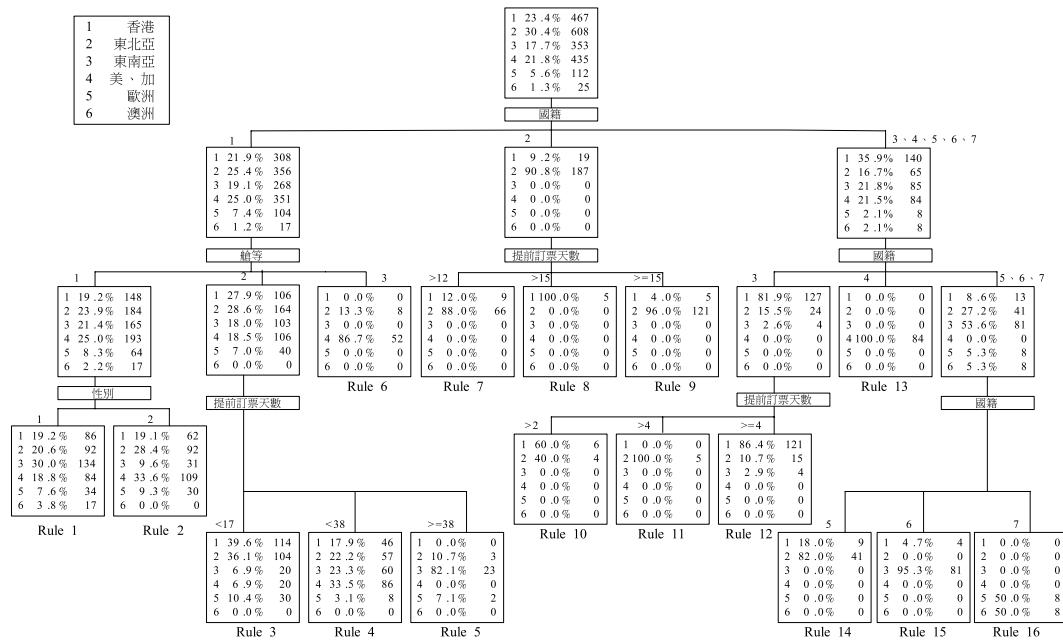


圖 16 航區別決策樹分類結果

本研究以最大規則長度 3 (即最多可達三條航線間的關聯性) 為分析範圍, 在多次測試參數差異後, 決定採最小支持數目 5%、最小信賴度 50% 為參數設定值。以下示範購票型態一之下, 全體交易紀錄的關聯分析結果。

表 18 顯示兩條航線之間的關聯分析, 結果發現多數規則屬於起迄點往返的關聯性, 如規則編號 1 為購買「香港—臺北 (25)」機票之旅客也會購買「臺北—香港 (13)」, 這雖屬顯而易見的事實。但規則編號 5 指出「臺北—東京 (1)」航線與「臺北—香港 (13)」航線之間, 可能存在被業者低估的消費關聯性, 而規則編號 24 也發現「臺北—名古屋 (3)」與「臺北—東京(1)」的高度關聯, 如圖 17 所示。至於三條航線之間的關聯規則, 詳如表 19 所示, 其中規則編號 4 代表搭乘「臺北—名古屋 (3)」的旅客也傾向搭乘「名古屋—臺北 (9)」及「臺北—東京 (1)」。本研究認為, 利用資料採礦的關聯規則分析, 可有效挖掘隱而未見的消費者特性, 進而協助業者提供更貼近顧客需求的旅運服務。

本研究接著針對不同購票型態下的各種旅次目的進行關聯分析, 發現在不同型態所得之結果差異不大, 如型態一與型態二僅在全體關聯分析之結果有異, 而型態二與型態三在全體及觀光旅次的關聯分析有所不同。型態一不同旅次目的別之關聯分析結果可比較如表 20 所示, 其中就整體交易紀錄而言, 以臺北、香港、東京三點之間的航線關聯性最高, 商務旅次為臺北、東京、名古屋的組合, 觀光旅次則對臺北、東京、琉球三地有最相似的消費偏好。從前述關聯分析可看出, 不同旅次目的之顧客對於航線的偏好, 也有類似超市商品搭售的行銷潛力, 因此航空業者應善用關聯法則的分析工具, 找出不同旅客族群的潛在需求及航線偏好組合, 將有限的行銷資源花在刀口上, 以客製化的行銷套裝方案, 提供更符合旅客需要的優質服務。

表 18 購票型態一之下的兩航線關聯規則表

編號	支持度	信賴度	規則	編號	支持度	信賴度	規則
1	37.31	100.00	25 ==> 13	20	8.06	100.00	5 ==> 11
2	37.31	76.69	13 ==> 25	21	8.06	69.23	9 ==> 7
3	32.84	98.21	7 ==> 1	22	8.06	69.23	9 ==> 1
4	32.84	96.49	1 ==> 7	23	8.06	69.23	3 ==> 7
5	20.00	58.77	1 ==> 13	24	8.06	69.23	3 ==> 1
6	18.81	56.25	7 ==> 13	25	6.57	100.00	29 ==> 17
7	13.13	100.00	28 ==> 16	26	6.57	100.00	17 ==> 29
8	13.13	95.65	16 ==> 28	27	5.67	79.17	50 ==> 41
9	12.54	100.00	12 ==> 6	28	5.67	86.36	41 ==> 50
10	12.54	100.00	6 ==> 12	29	5.37	100.00	53 ==> 44
11	11.94	95.24	51 ==> 42	30	5.37	100.00	44 ==> 53
12	11.94	88.89	42 ==> 51	31	5.37	90.00	52 ==> 43
13	11.64	100.00	9 ==> 3	32	5.37	100.00	43 ==> 52
14	11.64	100.00	3 ==> 9	33	5.37	75.00	50 ==> 51
15	10.75	87.80	34 ==> 22	34	5.37	75.00	50 ==> 42
16	10.75	100.00	22 ==> 34	35	5.37	81.82	29 ==> 28
17	9.55	100.00	47 ==> 38	36	5.37	81.82	29 ==> 16
18	9.55	86.49	38 ==> 47	37	5.37	81.82	17 ==> 28
19	8.06	100.00	11 ==> 5	38	5.37	81.82	17 ==> 16

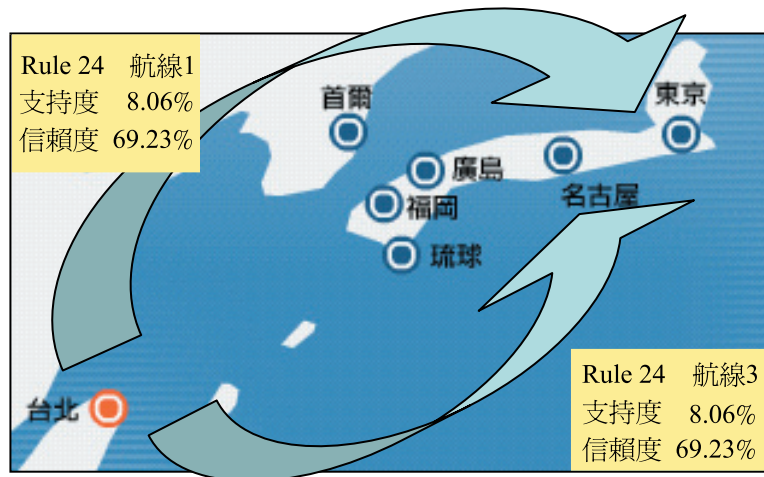


圖 17 購票型態一之下的兩航線關聯規則示意圖

表 19 購票型態一之下的三航線關聯規則表

編號	支持度	信賴度	規則	編號	支持度	信賴度	規則
1	18.81	56.25	7 ==> 13 & 1	18	8.06	100.00	9 & 7 ==> 1
2	18.81	55.26	1 ==> 13 & 7	19	8.06	100.00	9 & 1 ==> 7
3	18.81	100.00	13 & 7 ==> 1	20	8.06	69.23	9 ==> 3 & 1
4	18.81	94.03	13 & 1 ==> 7	21	8.06	69.23	3 ==> 9 & 1
5	18.81	57.27	7 & 1 ==> 13	22	8.06	69.23	9 & 3 ==> 1
6	14.03	100.00	25 & 1 ==> 13	23	8.06	100.00	9 & 1 ==> 3
7	14.03	70.15	13 & 1 ==> 25	24	8.06	100.00	3 & 1 ==> 9
8	12.84	100.00	25 & 7 ==> 13	25	8.06	69.23	3 ==> 7 & 1
9	12.84	68.25	13 & 7 ==> 25	26	8.06	100.00	7 & 3 ==> 1
10	12.84	100.00	25 & 7 ==> 1	27	8.06	100.00	3 & 1 ==> 7
11	12.84	91.49	25 & 1 ==> 7	28	5.67	100.00	12 & 7 ==> 6
12	8.06	69.23	9 ==> 7 & 3	29	5.67	100.00	7 & 6 ==> 12
13	8.06	69.23	3 ==> 9 & 7	30	5.67	100.00	12 & 7 ==> 1
14	8.06	100.00	9 & 7 ==> 3	31	5.67	100.00	12 & 1 ==> 7
15	8.06	69.23	9 & 3 ==> 7	32	5.67	100.00	12 & 1 ==> 6
16	8.06	100.00	7 & 3 ==> 9	33	5.67	100.00	6 & 1 ==> 12
17	8.06	69.23	9 ==> 7 & 1	34	5.67	100.00	7 & 6 ==> 1

表 20 購票型態一之下旅次目的別的關聯分析比較

旅次目的	最高關聯組	次高關聯組	第三關聯組
整體	臺北、香港、東京	臺北、東京、名古屋	臺北、東京、首爾
商務	臺北、東京、名古屋	臺北、香港、東京	臺北、洛杉磯、舊金山
觀光	臺北、東京、琉球	臺北、東京、首爾	臺北、琉球、首爾
探親	臺北、香港、福岡	臺北、香港、雅加達	—
其他	臺北、洛杉磯、舊金山	臺北、胡志明、曼谷	臺北、香港、曼谷

4.5 實務驗證及行銷策略研擬

為確保研究成果的實務貢獻，本研究曾多次與航空公司行銷、業務及電子商務等相關部門主管進行訪談，除了確認本研究的方向是否合乎業界現狀需求外，更重要的是驗證研究方法及成果在實務上的實用價值。首先在群集化部分，業者目前仍多以搭乘哩程及次

數，作為判斷顧客價值之簡略標準，較無法有效區隔出價值顧客，而本研究採用五項指標來評分，並應用群集化技術將顧客分群，可讓航空公司明確分辨各族群旅客的需求特性及對公司的貢獻度。

接著本研究分別構建顧客及產品導向的分類模式，顧客導向之分類成果，能有效幫助業者儘速掌握新進顧客的需求特性及消費潛力，而產品導向可讓航空公司更了解各航區及航線的旅客面貌。目前業者多憑從業人員的個人經驗，對顧客需求及航線發展潛力進行判斷，既無標準化之判斷準則，亦缺乏科學數據的佐證，因此本研究所提出的分析模式，可望填補航空公司現階段分析工具的不足。

至於關聯分析部分，業界對於航線關聯性的了解，也多仰賴資深人員的主觀判斷，較不容易發現隱藏之規則，且無法累積相關之商業智慧。而本研究針對不同旅次目的進行關聯分析，證實了旅客對航線的關聯偏好，會因為旅次目的而有所不同，且隱藏在龐大交易資料庫中更多的航線關聯性，正是航空公司最寶貴的產業情報，因此業者應大力挖掘並善用資料採礦技術在航空產業的分析功能。本研究所構建的分析模式，可與業界實務做法比較如表 21 所示。

本研究根據顧客群集化的結果，共將航空旅客區隔成四群，經由參考過去航空行銷策略之相關文獻及本研究過程所發現之特性，再與航空公司行銷部門進行討論，嘗試提出不同顧客群之行銷策略建議如表 22 所示。

表 21 本研究分析模式與業界實務做法之比較

分析模式	分析方式及主要結果	實務做法
群集化	運用 R、F、M、C、Mi 等五項指標判斷顧客價值並進行群集化，可將顧客明確區隔為四群貢獻度不同的旅客族群。	多使用頻率 (F) 及哩程 (Mi) 來粗分顧客族群，區隔效果較不明顯，對顧客價值的判斷也較不客觀。
分類	顧客導向： 有助於航空公司儘速掌握新進顧客的需求特性及消費潛力。 產品導向： 對航區及航線進行分析，分析結果可使業者更了解各航線的旅客面貌。	多憑個人經驗對顧客需求及航線發展潛力進行判斷，既無標準化之判斷準則，亦缺乏科學數據的佐證。
關聯分析	對不同旅次目的顧客所搭乘的航線進行關聯分析，可發現航線之間的關聯性，以提供產品組合銷售之可能性。	航線關聯性多由航空公司主觀判斷，較不容易發現隱藏之規則，且無法累積相關之商業智慧。

表 22 不同顧客群組之特性及行銷策略

顧客群組	特性	行銷策略	行銷方案
核心顧客 (群組 1)	✓ 搭乘頻率高 ✓ 艙等、金額、哩程高 ✓ 提前訂票天數短 ✓ 年齡、會齡高	精緻化服務	代為安排飯店住宿及優質租車完善的專車接送或專人服務
有效顧客 (群組 2)	✓ 搭乘頻率較高 ✓ 艙等、金額、哩程較低 ✓ 提前訂票天數較短 ✓ 年齡高、會齡低	價格優惠	座位艙等升級之價格優惠中、長程航線之價格優惠
游離顧客 (群組 3)	✓ 搭乘頻率較低 ✓ 艙等、金額、哩程較高 ✓ 提前訂票天數較長 ✓ 年齡較低、會齡較高	差異化策略	進一步了解其搭乘需求，提供令其感受到差異化的服務，搶在同業之前吸引其惠顧
邊緣顧客 (群組 4)	✓ 搭乘頻率低 ✓ 艙等、金額、哩程低 ✓ 提前訂票天數長 ✓ 年齡低、會齡低	非優先行銷	不列入航空公司的目標顧客

五、結論與建議

5.1 結論

1. 透過本研究所提出之最近購買日 (R)、購買頻率 (F)、平均消費金額 (M)、平均艙等分數 (C) 及平均哩程數 (Mi) 等五項指標，對航空公司線上購票旅客進行顧客價值分析，可幫助業者明確分辨顧客對公司業績貢獻的高低，相較於目前業界多僅以哩程及頻率來作粗略判斷，更具客觀性。
2. 運用資料採礦的群集化分析技術，可協助航空公司瞄準目標顧客，進行更客製化的行銷。如本研究將旅客分為四群，其中群組 1 為最具價值之核心顧客 (忠誠度高、利益高)，其搭乘艙等較高，消費能力強，較渴望獨一無二的優越服務，因此在行銷策略上，業者應提供專車接送、專人服務等高規格的待遇。
3. 本研究構建顧客導向的分類模型，可預判新進顧客之群組屬性，所得之分類正確率都在 70% 以上 (先前薛文佶^[14]研究之正確率為 40%)，因此可更有效預測新開發顧客之潛在價值，有助於航空公司及早因應與適時掌握商機。
4. 針對顧客旅次目的之不同，本研究以關聯規則進行旅客搭乘偏好及航線關聯性的分析，結果證實在不同旅次目的下，其航線關聯性將有所差異。而關聯規則也確實能挖掘航線之間隱藏的關聯性，提供航空公司最寶貴的商業智慧，且有利於業者進行更多航線的組

合銷售。

5. 經由與航空公司的訪談，得知業界已開始體認資料採礦技術在挖掘旅客需求上的分析能力，並將視為公司未來鎖定之重點工作。在網際網路普及與資訊科技成熟的大環境下，航空公司唯有利用科學化的分析工具，有效地挖掘旅客需求並提供更客製化的服務，才能在競爭激烈的業界生存。

5.2 建議

1. 為了更深入了解旅客的消費習性及潛在需求，建議航空公司今後應在購票網頁的設計上，加入更多與行為變數相關的填項，將有助於呈現更完整的消費者面貌。例如本研究在進行不同旅次目的之航線關聯分析時，就發現其所得之結果也有所差異，因此建議業者將旅次目的作為線上訂票之優先填項，以期掌握更精準的旅客需求。
2. 本研究運用資料採礦對航空旅客線上購票交易紀錄進行分析，所採用之演算法包含群集化的 K-means 法、分類模式的 CHAID 法及關聯規則的 Apriori 法，而資料採礦之演算工具眾多，後續研究可針對其他方法進行航空業的探討，如運用類神經網路構建分類模式，或可呈現不同的分析結果。
3. 本研究僅針對線上購票的顧客進行分析，其結果並不代表航空公司全部旅客之消費行為，若後續研究獲得更完整的營運資料，可嘗試利用資料採礦技術，分別對傳統通路（旅行社）與線上購票的顧客進行研究，並比較其異同，屆時應可提供航空公司更有價值的產業情報。

參考文獻

1. 林祥生、劉益豪，「透過航空公司線上購票之顧客區隔研究」，2006 年國際學術研討會管理組論文集，銘傳大學，民國九十五年，頁 625-644。
2. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, Vol. 17, No. 3, 1996, pp. 37-54.
3. Berry, M. J. A. and Linoff, G., *Data Mining Technique for Marketing, Sale, and Customer Support*, Wiley Computer, N. J., 1997.
4. Berry, M. J. A. and Linoff, G., *Mastering Data Mining: The Art and Science of Customer Relationship Management, Sale, and Customer Support*, Wiley Computer, N. J., 1999.
5. Han, J. and Kamber, M., *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publisher Inc, San Francisco, 2000.
6. 尹相志，**SQL 2000 Analysis Service 資料採礦服務**，維科圖書公司，臺北縣，民國九十二年。
7. MIT, *The Technology Review Ten*, MIT Technology, Boston, 2001.

8. 盧世銘，「資料採礦技術之商業應用研究－以航空公司會員系統為例」，政治大學經營管理研究所碩士論文，民國九十三年。
9. 王媚慧，「應用資料探勘於顧客的行為分析－以半導體業為例」，淡江大學資訊工程研究所碩士論文，民國九十三年。
10. Chan, K. H. C. and Gerry, K. L., "Data Mining and Customer Relationship Marketing in the Banking Industry", *Singapore Management Review* 2002, Vol. 24, 2002, pp. 1-27.
11. Changchien, S. W. and Lu, T. C., "Mining Association Rules Procedure to Support On-Line Recommendation by Customers and Products Fragmentation", *Expert Systems with Applications*, Vol. 20, 2001, pp. 325-335.
12. Liaoa, S. H. and Chen, Y. J., "Mining Customer Knowledge for Electronic Catalog Marketing", *Expert Systems with Applications*, Vol. 27, 2004, pp. 521-532.
13. Arthur, H., *Strategic Database Marketing*, Probus Publishing, Chicago, 1994.
14. 薛文佑，「航空業顧客價值之辨識暨消費資料庫之規則探勘」，臺北科技大學商業自動化與管理研究所碩士論文，民國九十三年。
15. 謝文雄，「以資料探勘探討顧客消費之行為」，雲林科技大學資訊管理研究所碩士論文，民國九十一年。
16. 蔡永恆，「應用資料挖掘技術研究銀行顧客消費行為」，靜宜大學資訊管理研究所碩士論文，民國八十八年。
17. Hsieh, N. C., "An Integrated Data Mining and Behavioral Scoring Model for Analyzing Bank Customers", *Expert Systems with Applications*, Vol. 27, 2004, pp. 623-633.
18. 劉益豪，「應用資料採礦探討航空旅客之線上購票行為」，中華大學科技管理研究所碩士論文，民國九十五年。
19. Mac Queen, J. B., "Some Methods for Classification and Analysis of Multivariate Observation", Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, UC Berkeley, 1967, pp. 281-297.
20. Kass, G. V., "An Exploratory Technique for Investigating Large Quantities of Categorical Data", *Applied Statistics*, Vol. 29, 1980, pp. 119-127.

