



# SPSS資料採礦

## --什麼是資料採礦

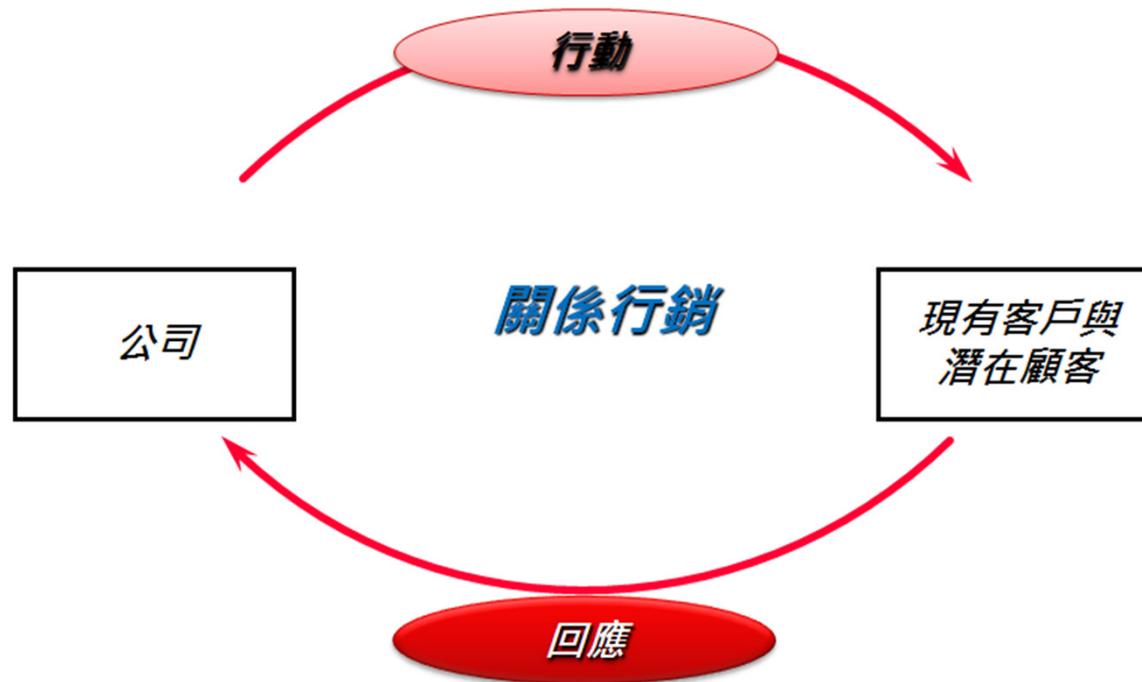


## Agenda

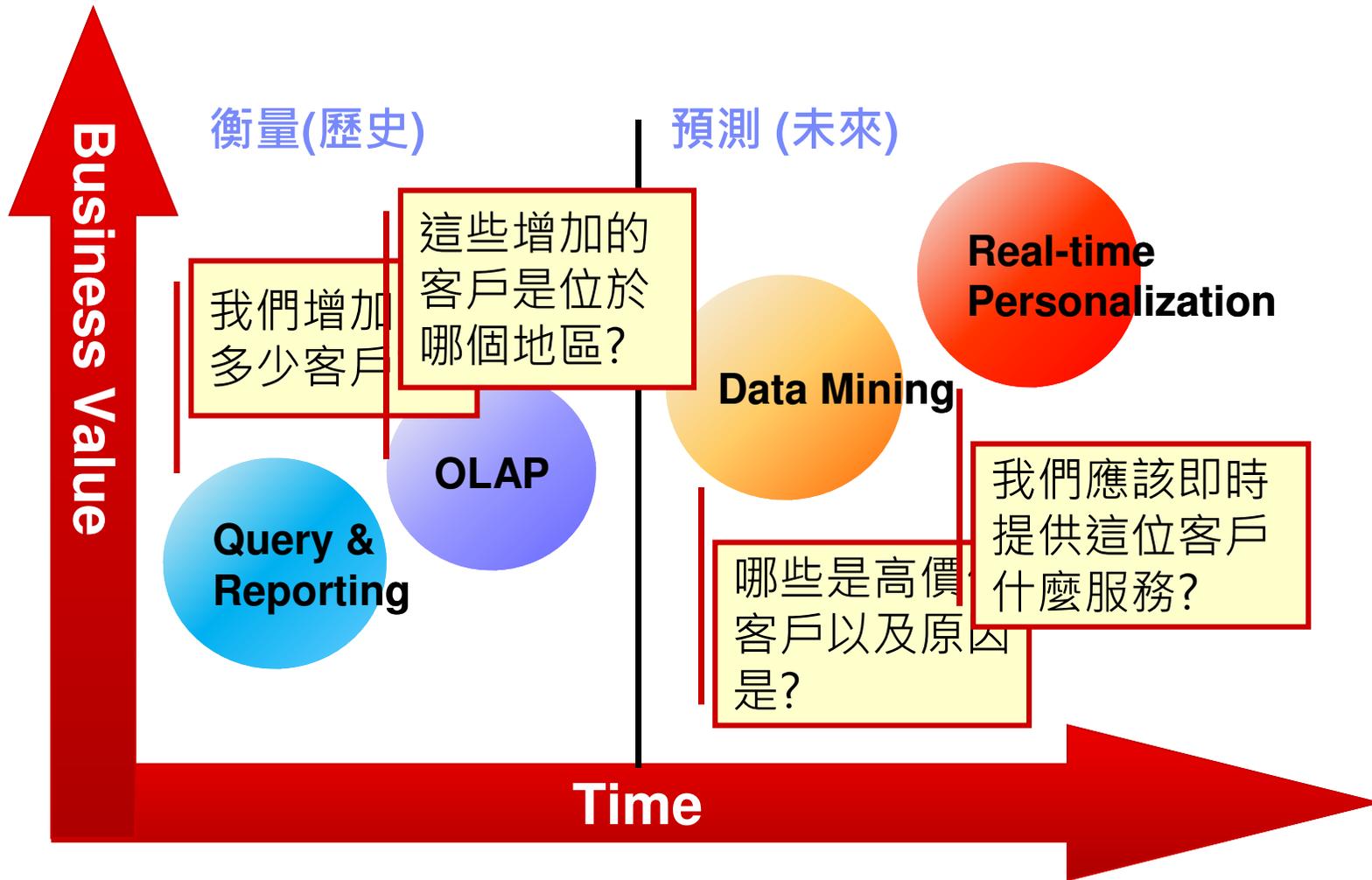
- 何謂資料採礦
- 資料採礦的主要功能有哪些
- 資料採礦與分析型顧客關係管理

## 分析型顧客關係管理

- 在既有客戶群中創造最大價值
- 從客戶的回應過程中建立**學習**關係，以改善銷售循環



## Data Mining 的角色定位?



# 何謂Data Mining ?

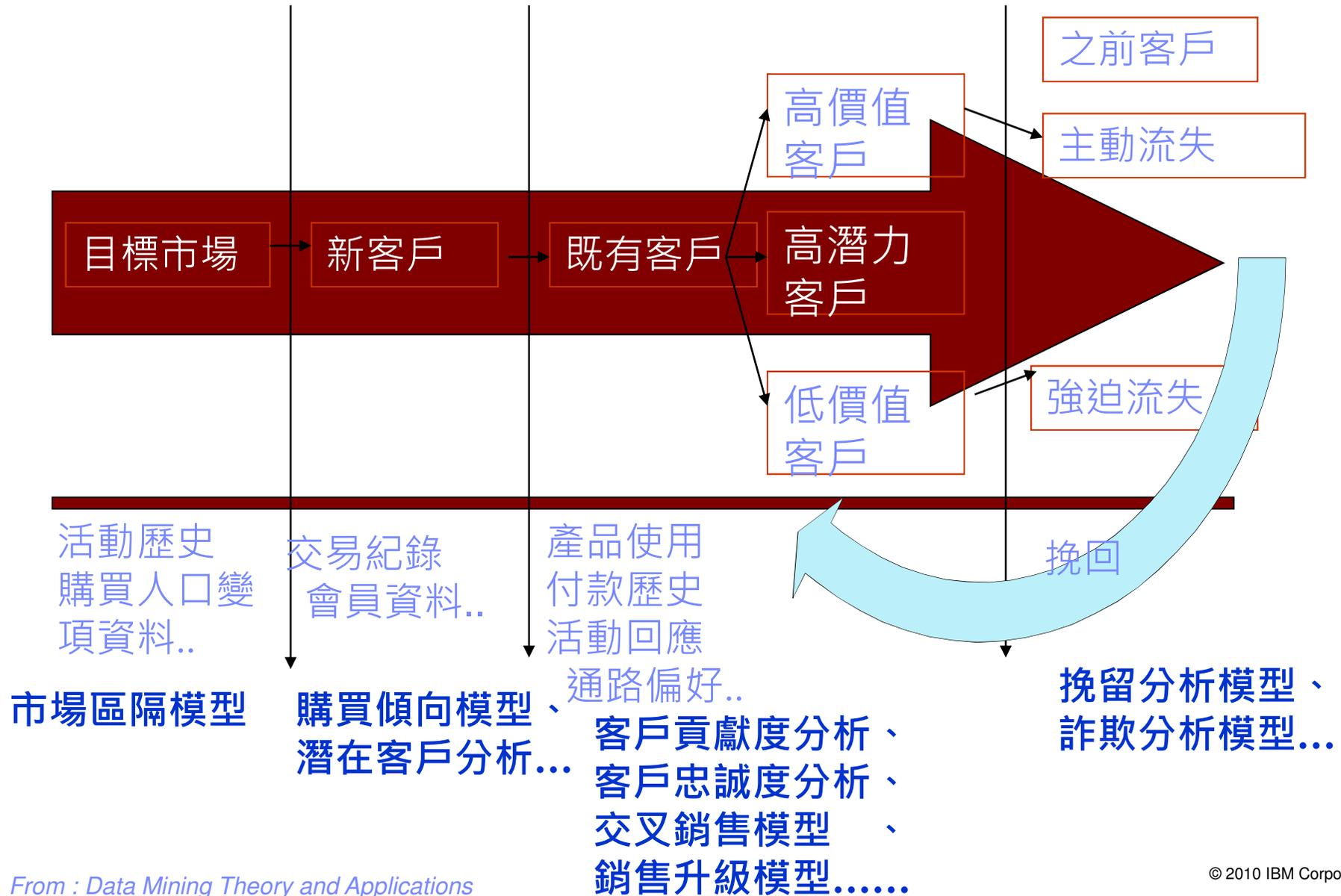
The screenshot displays three overlapping windows from the IBM Business Analytics software interface:

- Top Left Window:** A 3D scatter plot titled "Built Model: 3 sel" showing data points in a 3D space with axes labeled "Engine size", "Price in thousands", and "Price in thousands". Below the plot is a description: "This chart is a lower-dimensional projection of the data and contains a total of 1,952 segments." Buttons for "View: Predictor Space" and "Reset" are visible.
- Top Middle Window:** A "Bayesian Network" diagram showing a network of nodes representing variables. The nodes include "debt\_equity", "branch", "age\_youngest\_child", "household\_debt\_to\_equity\_ratio", "non\_worker\_percentage", "pension\_p", "white\_collar\_percentage", "bad\_payment", "call\_center\_contact", "gender", "marital", "loan\_accounts", "gold\_card", "number\_transactions", "months\_current", "income", "rfm\_score", "numb", and "average#balance". The "response" node is highlighted in red. Buttons for "View: Basic" and "Reset" are visible.
- Top Right Window:** A table titled "response[1] (1)" showing segment rules. The table has columns for "id", "Segment Rules", "Score", "Cover (n)", "Freque...", and "Probab...".

id	Segment Rules	Score	Cover (n)	Freque...	Probab...
	All segments including Rem...		13,504	1,952	14.45%
1	months_customer months_customer = "0"	Excluded	1,747	0	0.00%
2	rfm_score rfm_score <= 0.000	Excluded	6,003	0	0.00%
3	rfm_score, income rfm_score > 12.333 and 1 income > 52213.000		555	456	82.16%
4	income income > 55267.000	1	643	551	85.69%
5	number_transactions, rfr number_transactions > ;1 rfm_score > 12.333		533	206	38.65%
6	age age > 30 and age <= 34	1	505	111	21.98%
7	rfm_score	1	1,220	252	20.74%

**Data Mining** 是一個從複雜資料中挖掘出被隱藏且具有價值之型式與關係(hidden pattern and relationship)的過程

# 資料採礦與分析型CRM



## 資料採礦的功能

- 分類(Classification)
- 預測(Prediction)
- 關聯分組(Affinity Group)
- 群集化(Clustering)
- 敘述(Description)



# SPSS資料採礦

## --資料採礦能幫我們什麼？



# 分類

- 輸出變數為類別變數
- 流失客戶預測、購買回應預測都是典型的分類問題

The left screenshot shows a decision tree structure with the following nodes and branches:

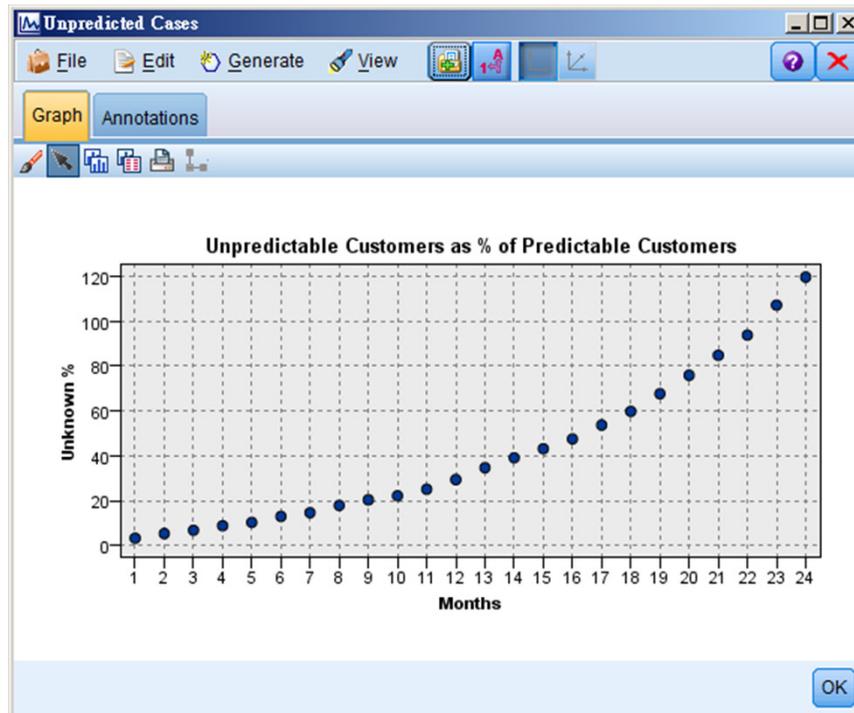
- income <= 55,648 [Mode: 0]
  - number\_transactions <= 12 [Mode: 0] => 0
  - number\_transactions > 12 [Mode: 1]
    - branch = Catburg [Mode: 1]
      - gold\_card = 1 [Mode: 0] => 0
      - gold\_card = 0 [Mode: 1]
        - number\_products <= 5 [Mode: 1] => 1
        - number\_products > 5 [Mode: 0] => 0
    - branch = Foxton [Mode: 0]
      - gender = F [Mode: 1] => 1
      - gender = M [Mode: 0]
        - months\_current\_account <= 14 [Mode: 1] => 1
        - months\_current\_account > 14 [Mode: 0] => 0
    - branch = Kingsville [Mode: 1]
      - white\_collar\_percentage <= 25 [Mode: 0] => 0
      - white\_collar\_percentage > 25 [Mode: 1] => 1
    - branch = Princeton [Mode: 1]
      - loan\_accounts <= 0 [Mode: 0] => 0
      - loan\_accounts > 0 [Mode: 1] => 1
    - branch = Queensbury [Mode: 0]
      - marital in ["M" "U"] [Mode: 0] => 0
      - marital in ["S"] [Mode: 1] => 1
    - branch = Ravensville [Mode: 0]
      - income <= 38,681 [Mode: 1] => 1
      - income > 38,681 [Mode: 0] => 0
  - income > 55,648 [Mode: 1] => 1

The right screenshot shows a detailed view of a node with the following table:

Category	%	n
0.000	91.785	11552
1.000	8.215	1034
<b>Total</b>	<b>93.202</b>	<b>12586</b>

## 預測

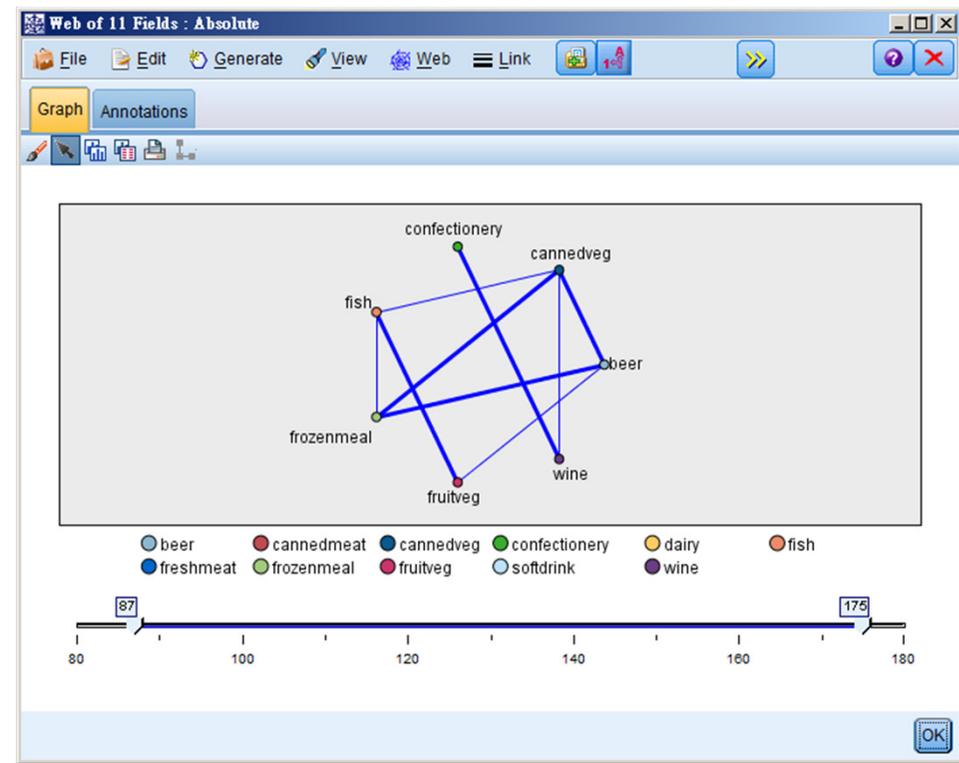
- 輸出變數為連續數值
- 購買力分析、顧客終身價值分析、銷售量業績預測、金融商品價格波動等都是典型預測問題
- 通常與「分類」結合，先根據分類決定是否購買，再根據預測推估購買金額



## 關聯分組

- 找出產品間購買相關性
  - Association rule 找出哪些產品會一起被購買
  - Sequence rule 找出有時間因果的產品關聯順序

Consequent	Antecedent	Support %	Confidence %
frozenmeal	beer cannedveg	16.7	87.425
cannedveg	beer frozenmeal	17.0	85.882
beer	frozenmeal cannedveg	17.3	84.393



## 關聯分組

- **Wal-mart:**週四晚上，尿布與啤酒
- **Safe-Way:**下架的藍乳酪可以利用關聯規則制定促銷組合以及交叉銷售策略

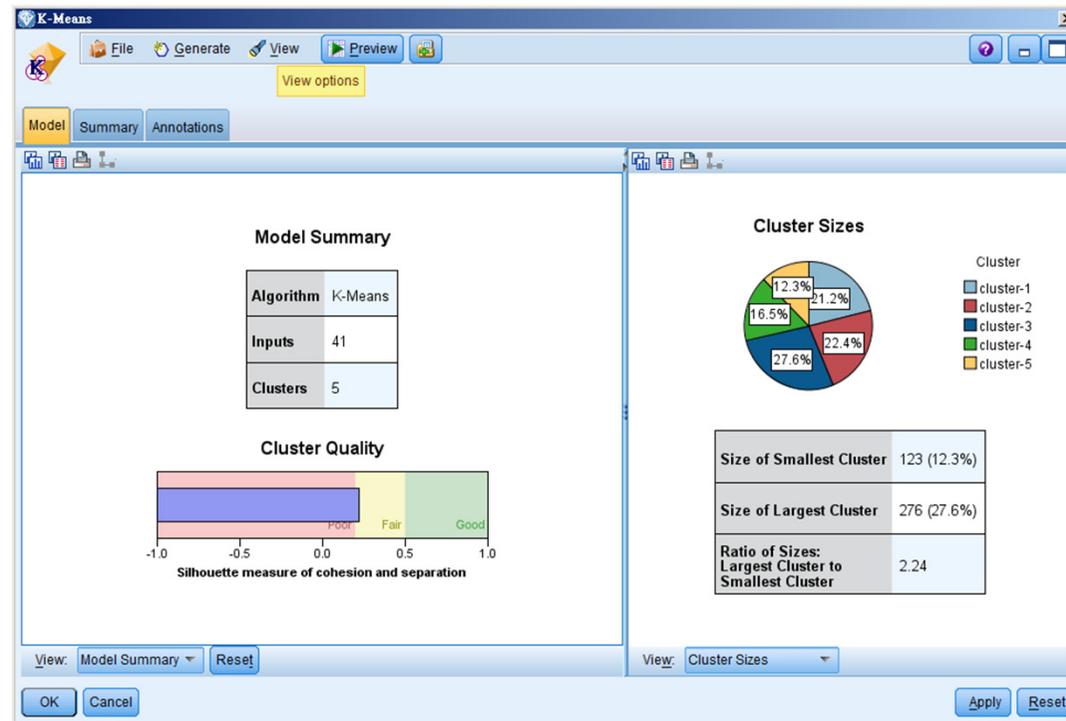


&



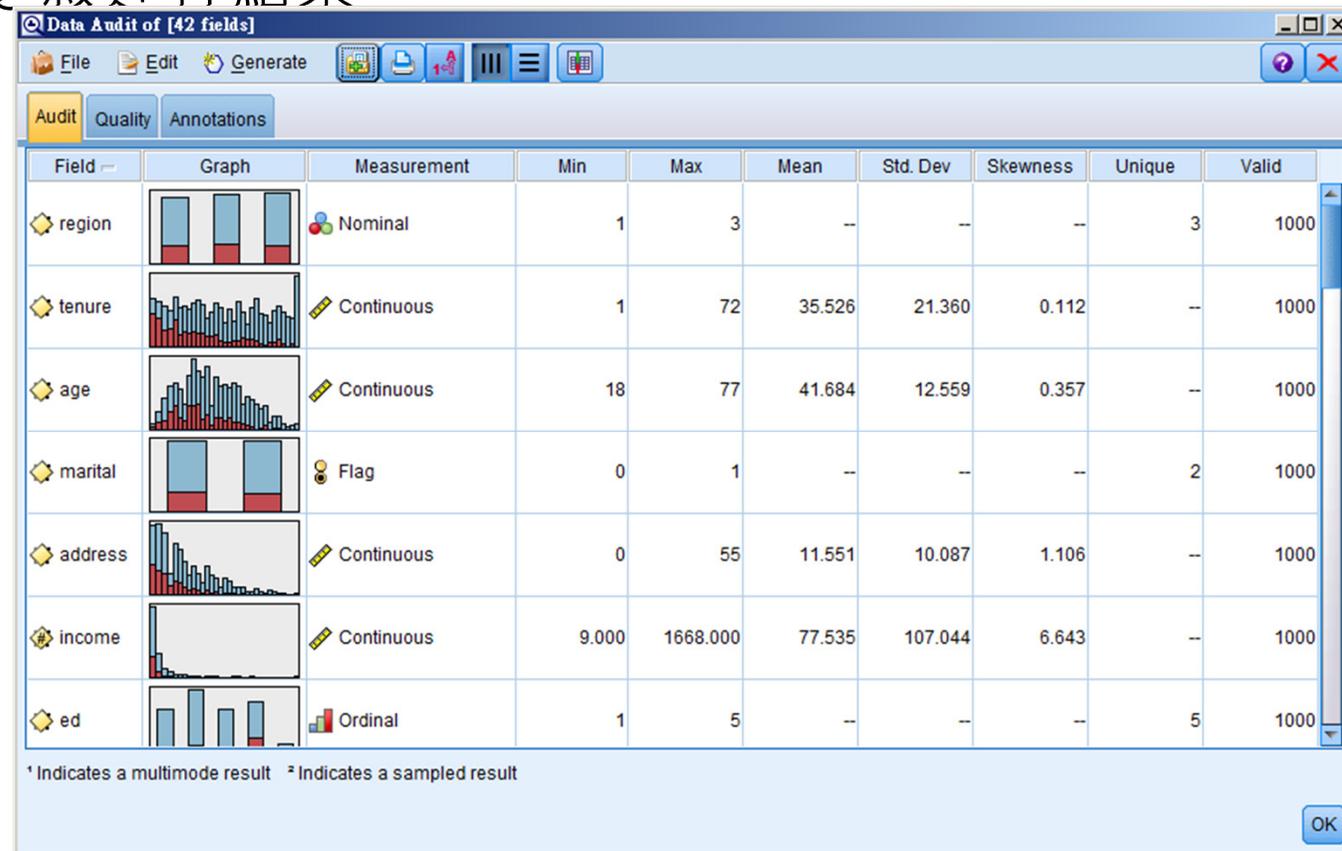
## 群集化

- 透過資料相近性，將資料自動區隔為數個有意義的群集
- 找尋資料的內部結構
- 沒有既定的輸出變數，與「分類」不同



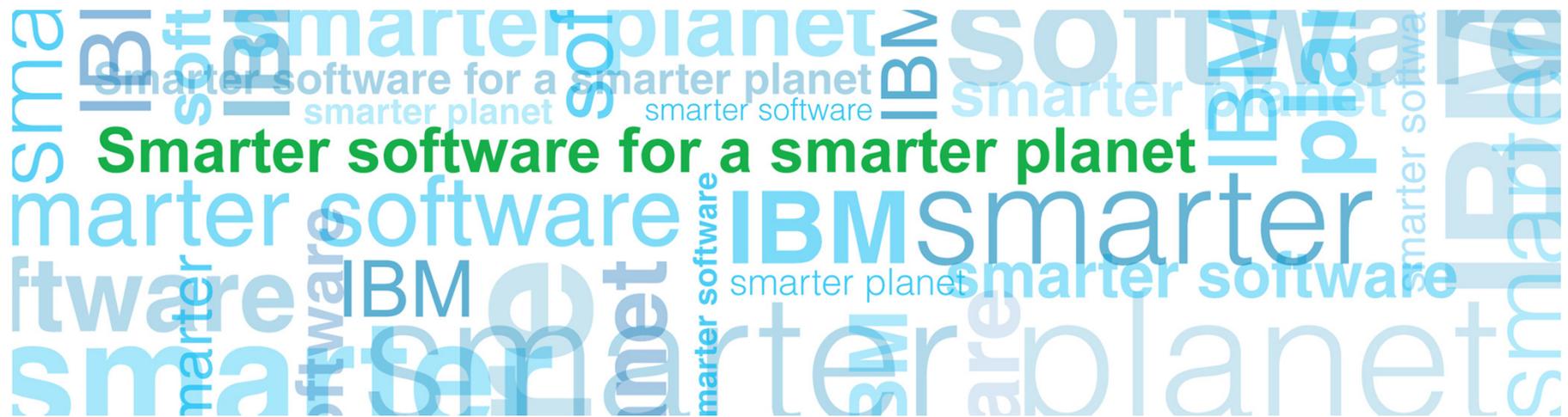
## 敘述

- 有時「人」就是最好的資料採礦工具
- 透過資料採礦程式將資料清理、視覺化，就能產生有意義之敘述性結果





# SPSS資料採礦 --電話行銷案例



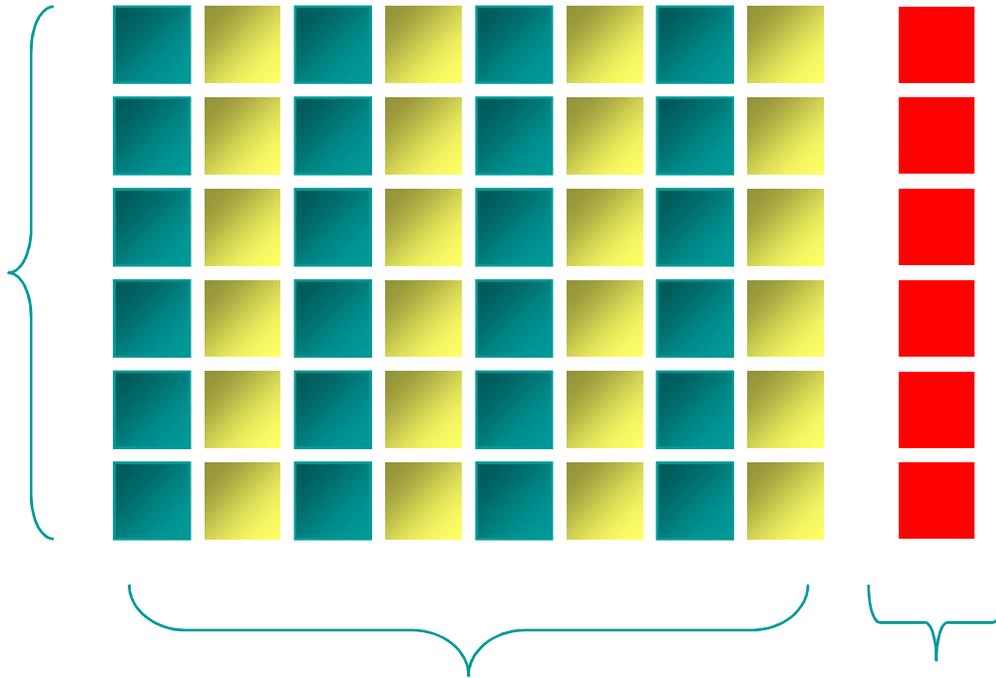
# 電話行銷



Copyright © 2000 United Feature Syndicate, Inc.  
Redistribution in whole or in part prohibited

# 資料取樣:什麼是預測?

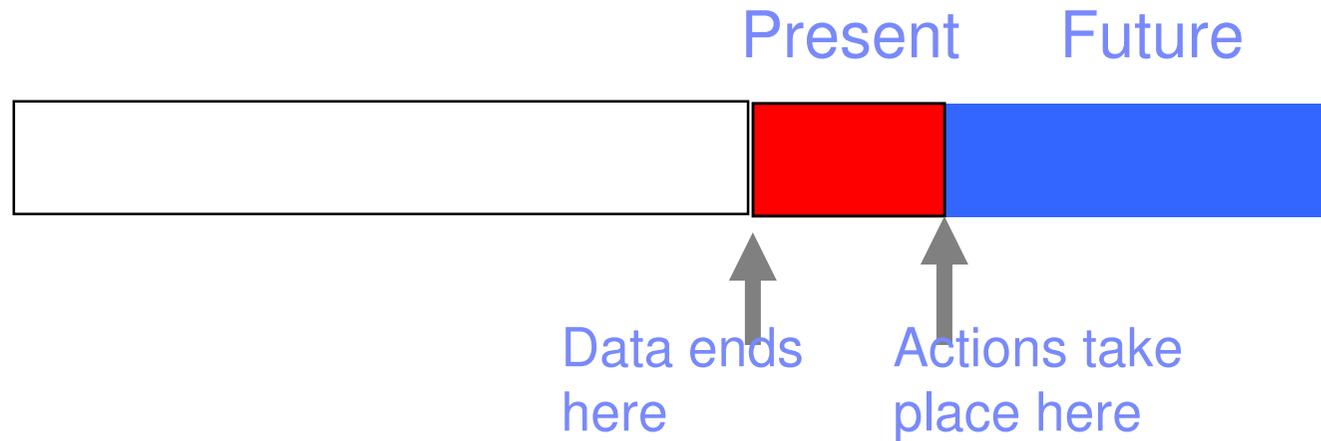
案例



輸入變數(自變數)

輸出/目標變數(依變數)

## 資料採礦是關於時間的

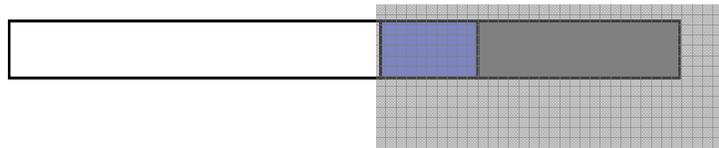


- 建立模型 使用歷史資料--結果是已知的
- 應用 (或是 評分) 模型 使用現有的資料
- 執行結果 使用未來的資料(下週或是下個月)--結果是未知的

## Profiling 與 Prediction

### PROFILING

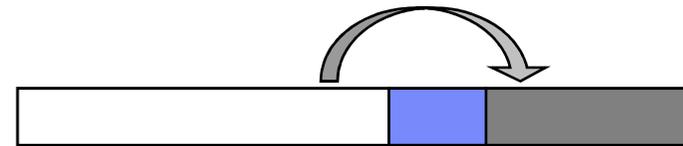
會買保險的是什麼樣的人



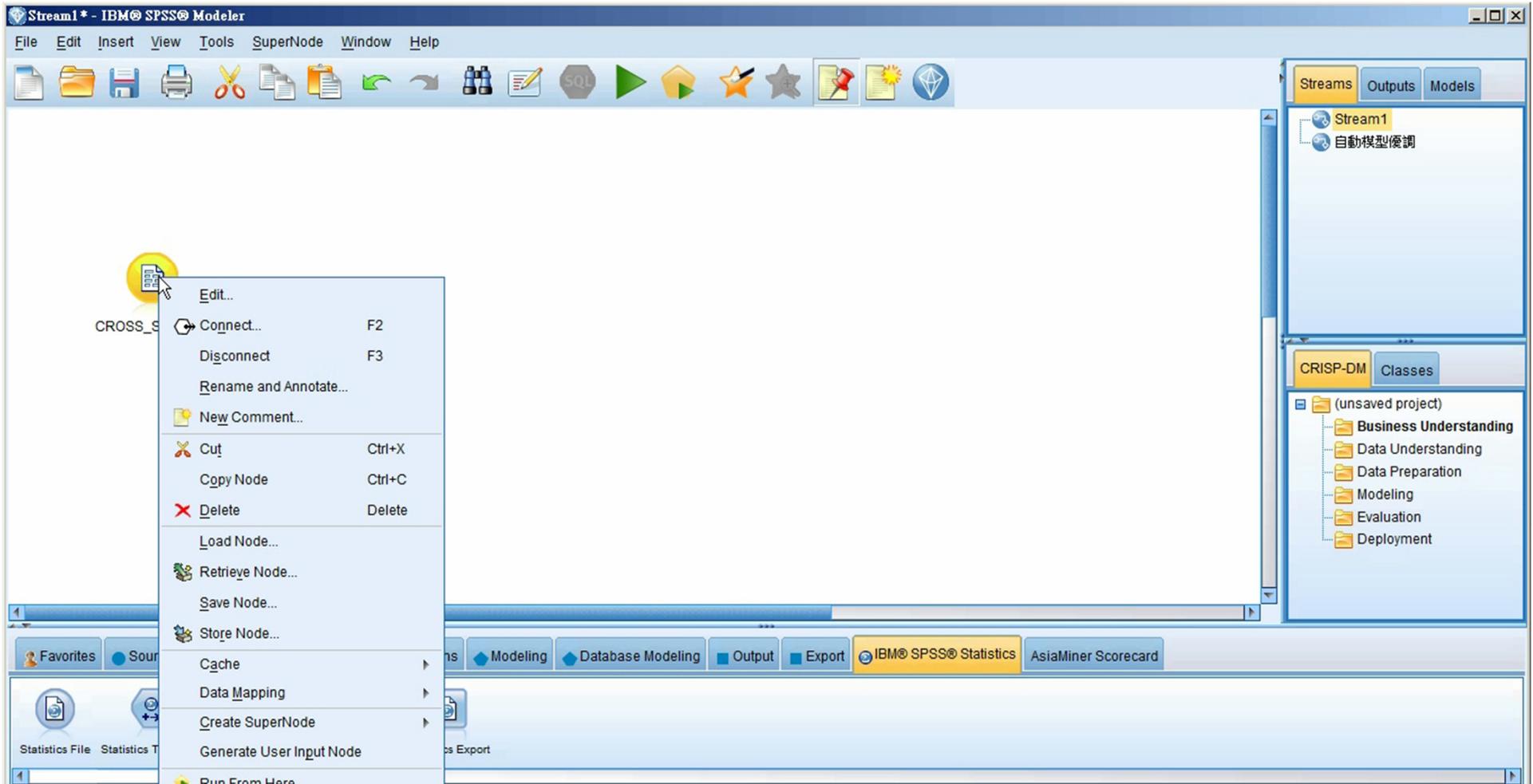
使用過去的資料來描述過去  
，用來提供描述未來的參考  
，輸入變數可以是與輸出變數同時產出

### PREDICTION

建立一個能夠預測會買保險客戶模型

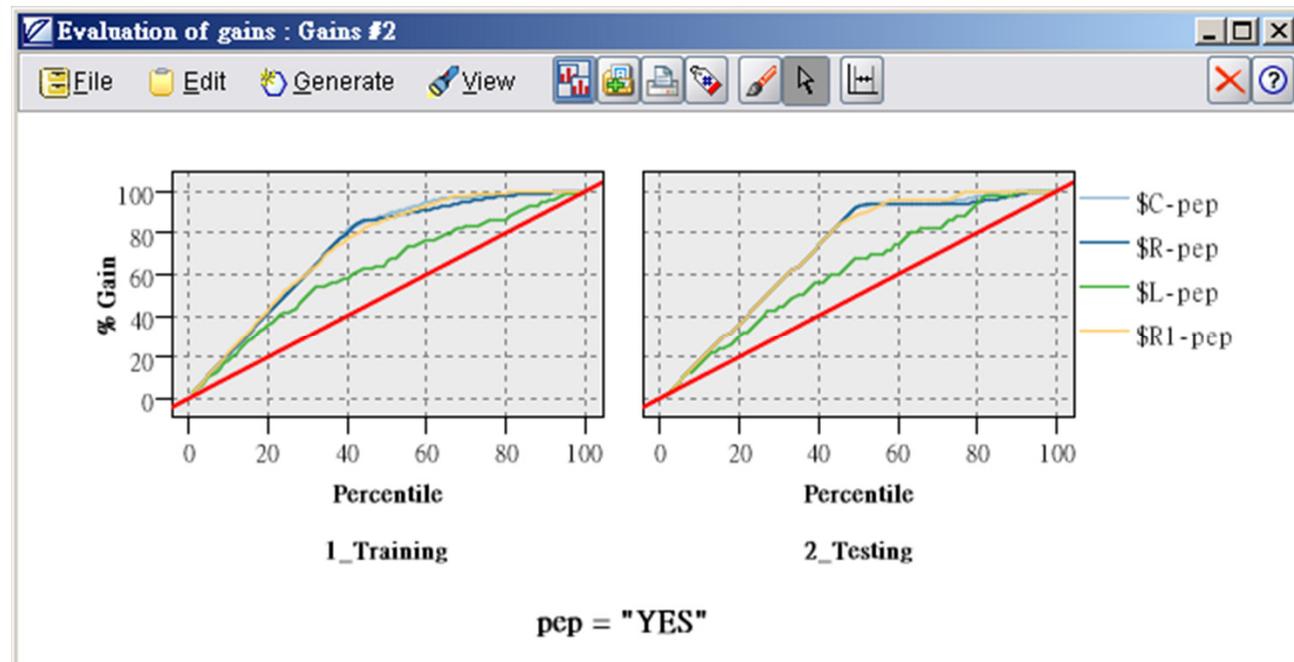


使用歷史資料預測未來，  
使用之預測變數必須比輸出變數較早出現



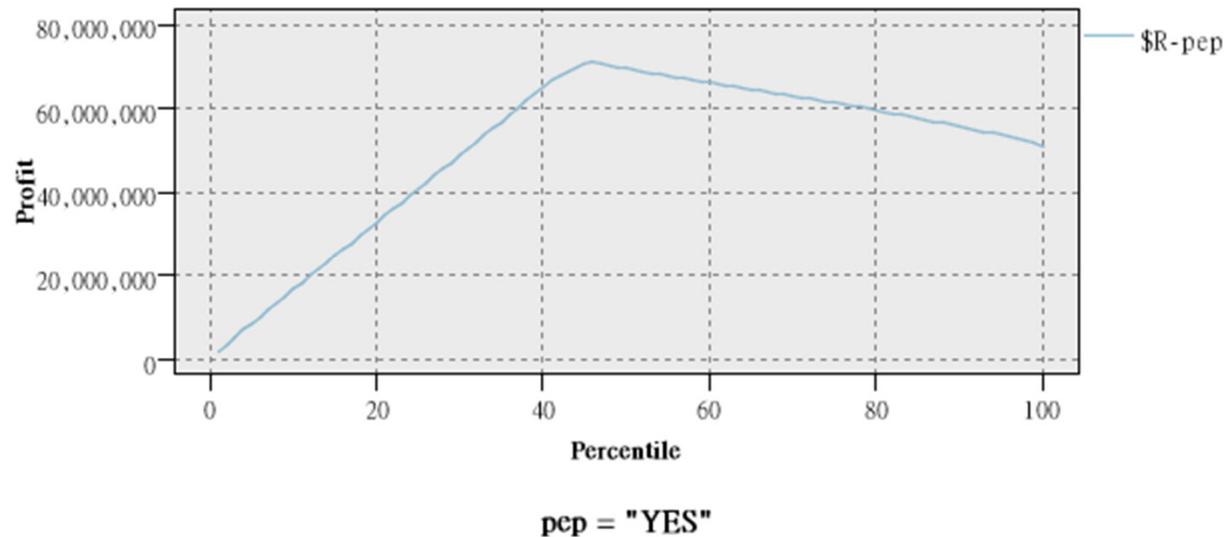
# Gains Chart

- 用以比較模型之間績效，以篩選出最佳預測模型
- 提供訓練組與測試組比較，以檢視模型穩定性



# Profit Chart

- 輸入單位接觸成本以及單位成交收益，即可計算樣本預估獲利，以及預估最佳獲利點
- 資料採礦效益不僅是**收益提升**，同時可將**節省人力**再投入行銷投資中





# SPSS資料採礦 --電信業顧客流失



## 企業問題

- 競爭白熱化的市場
- 2005/10開始啟動NP(Number Portable)
- 電信業需要投入大量成本在客戶挽留，需要透過資料採礦鎖定高流失客戶

## 顧客流失預警常遇到的問題



透過資料採礦預測客戶何時會流失!雖然名單主卻但是名單拿到手上時卻來不及進行挽留行銷



模型內大多數是低用量行為變數，難以有效挽留重要的高價值客戶!!



高流失客戶幾乎都是撈出合約剛期滿的客戶，需要更精準的挽留目標!!



Retention活動成效很好但是Churn Rate卻沒有明顯下降?

# 時窗(Time Window)

Jan Feb Mar Apr May Jun July Aug Sep



資料處理日

挽留活動最晚出手日

申辦競爭對手門號

合約到期

顧客停機