CH1-資料分析概觀

* 商業智慧與資料探勘 (BIDM)循環
  + 商業活動 -> 資料(活動過程中,以紙張或電子媒體記錄下來) -> 探勘(將上一步驟的資料進行統整、分析和探勘作業) -> 智慧(歸納出模式與智慧，以反應商業活動的運作情形)
* 何謂資料處理鏈
  + 從收集資料，建立模型放入資料庫，汲取資料作資料倉儲，運用資料探勘產生新的見解，最終以視覺化方式將見解呈現給正確的受眾。
* 何謂資料？資料有哪些類型？什麼是資料化？
  + 所有可被記錄的事物都是資料
  + 資料類型：文字、數字、影片/圖檔(要上描述資料才能分析)
  + 資料化：現今所有現象都會被觀察與儲存。
* 何謂模式？模式有哪些類型？
  + 模式是有助於掌握現況的設計或模型，可將看似無關的事物連結在一起。
  + 模式有助於解析複雜事物，展現出更簡單易懂的趨勢。
  + 完善模式是以GAS, General(通用),Accurate(精確),Simplify(簡化)的方式來描述的模型。
  + 模式類型：時間/空間/功能
    1. 時間 => 帕金森定律，指的是工作量會一直增加到所有可用時間都用完為止。
    2. 空間 => 三維空間，例：城市規劃通常是圍繞著首都或市中心向外發展。
    3. 功能 => 80/20法則，20%的產品創造了80%的業績。
* 鑽石探勘與資料探勘之間有什麼相似處？
  + 兩者都熟悉好的模式是長什麼樣。
  + 嫻熟商業領域是十分重要的。(要有Domain Know How)
* 資料探勘有哪些不同方法？有哪些會與你目前工作相關？
  + 以電腦稽核為例，電子資料比較多的是「核心系統之帳號登入失敗紀錄、非客戶帳號嘗試登入紀錄等進行監控及分析」。**可能要的不是預測，而是了解分群及趨勢，是否合規。**
  + 有哪些方法
    1. 決策樹(分群)：將母體分類為不同類別。例：高風險或低風險患者、高價值或低價值客戶。目前最受歡迎且重要的方法，很多演算法都能用來實作決策樹。**有正確答案。**
    2. 迴歸(預測)：迴歸模型可用來推斷未來，進行預測與估計。
    3. 人工神經網路/類神經網路(預測)ANN：機器學習領域，ANN是多層非線性資料處理模型，可從過去的資料中學習並預測未來數值。要有大量過去資料，才足以訓練系統。
    4. 集群分析(分群)：分割與處理大型資料集的重要資料深勘方法。藉由分辨資料中特定的相似與不相似性，將資料集分割為特定數量的集群。**沒有正確答案。**
    5. 關聯規則探勘：尋找資料值的關聯性。常用零售業的購物籃分析，作為「推薦或搭售商品」的參考依據。例：博客來、Netflix
* 何謂儀表板？為什麼它需要被客製化？
  + 儀表板它是個通用名詞，形容一種較吸引人樣式呈現相關資料的方式。
  + 它會針對每位管理者選定的少數變數提供資訊，因此需要被客製化。
* 請繪製一張視覺圖來展示目前所在城市的氣候模式。
  + 必須同時顯示出一週內的溫度、溼度、~~風向~~以及雨量嗎？(OK)

CH2-商業智慧與應用