# Task #09 作業 01【簡答題】資料探勘與關聯分析

1. 可以試著簡單解釋「統計學」、「機器學習」和「資料探勘」這三者的差異。

統計學著重於利用數學統計進行數據的推斷和解釋，強調數據的基本特徵和變異性。

機器學習更注重模型的訓練和預測，強調自動化的學習和性能改進。

資料探勘則專注於從大量數據中發現隱藏的模式和知識，結合了統計學和機器學習的技術來處理和分析數據。

2. 資料探勘主要專注在「Correlation」的關係，可以舉個例子分享「Correlation」與「Causation」有所差異。

例如直播課程中提到的冰淇淋銷售及防曬乳銷售的關係，兩者存在著正相關correlation關係(天氣炎熱大家都會去買冰淇淋或防曬乳，但不代表冰淇淋銷售增加會導致防曬乳的銷售也增加causation)。

3. 資料探勘是一種資料驅動（Data Driven）的方式，請問什麼是「資料驅動」？

資料驅動的方式強調以數據為基礎來進行決策和行動，從而提高決策的準確性和有效性。資料探勘作為一種資料驅動的方法，透過深入挖掘數據中的價值，協助做出更明智、更有根據的決策。(數據收集和分析、客觀性、實證支持)

4. 在關聯分析方法中，有 Apriori 和 Association Rules 兩種算法，請問他們的差異是什麼？

Apriori主要在生成frequent item set,觀察物品出現的機率。

Association Rules主要利用條件機率(當A出現的時候B也出現的機率)來觀察不同物品的關聯性，進而產生關聯規則。

在多人的購買清單中，Apriori通常會顯示物品在不同組合下出現的機率。

而Association Rules會表示當已有A物品時，B物品也會出現的機率是多少。

5. 可以舉一個適合使用關聯法進行分析的案例，包含資料集的使用情境。

就如同課程所教學的超市購物籃物品購買分析狀況，一家大型超市希望透過分析顧客的購物籃數據來了解哪些商品經常一起購買，以便於優化商品擺放、提升銷售和制定促銷策略。

透過Association Rules的觀察，可以了解哪些物品具有關聯性，顧客較有機會一起購買，因此賣場可藉此安排規劃一些促銷活動等等增加購買，也可以提高顧客滿意度。