巨量資料與統計分析Fall 2017

第一次作業（2017/10/11繳交）

1. (a)仿造第一章最後的巨量資料應用，各組依照學號最後兩位數字總和除以6之餘數，例如：學號尾數11及14加總為25，除以6後餘1，選擇「1：商管學院」。各組搜尋相關研究及報導，選擇兩個巨量資料在該學院（或領域）的可能應用，討論及說明巨量分析和傳統方法在應用上有何相同、相異之處，並說明使用巨量資料具有哪些優勢。

(b)延續(a)的討論，說明在各領域發展巨量資料需要哪些配套措施，如：設備、人力、制度、技術等，本身的限制及適用時機是什麼，以及可能會遭遇的阻力及困難？

註：可參考問題解決流程的四個步驟，逐步說明每一階段的配合要件，尤其著重於整體價值與實務考量（及阻力）。例如：比特幣(Bitcoin)和區塊鏈(Block Chain)是金融科技(Fintech)的重要發展方向之一，但近日中國政府宣布將於9月30日關閉中國主要的虛擬貨幣交易所（停止新用戶註冊及交易業務），並於10月31日前逐步停止與人民幣的兌換交易業務。即便中國有全球最大的比特幣礦場，為什麼中國政府還是將比特幣視為非法貨幣？

1. 如何彰顯大數據的價值，不能只依賴統計方法或數量分析工具，也需要應用領域的專業知識、資料特性等，因此資料科學家需要瞭解應用領域的價值、與人溝通的能力等。以Walmart的資料挖掘為例，尿布、啤酒兩者的關連，可作為業者發展如：商品定價、賣場動線、物流倉儲等之商機。
2. 仿造便利商店配合101跨年煙火的人群，舉例說明臺灣百貨零售業（包括好事多、家樂福等的大賣場）還有哪些項目的分析可為客戶帶來商機（開源、節流），並詳細說明如何將分析結果化為實際應用。
3. 除了百貨零售業外，臺灣還有哪些行業可應用大數據分析。
4. 藉由上網者使用搜尋引擎的習性，Google比美國疾管局更早發現冬季流感流行，然而Google公司使用上網搜尋紀錄的成功經驗似乎只有一次。
5. 請上網搜尋相關文獻（例如：*Detecting influenza epidemics using search engine query data*），整理、提出各組的看法，指出大數據的使用限制；
6. 除了Google Trends外，國際間也有不少大數據應用的案例，請蒐集至少各一個成功及失敗的範例，並說明成功的關鍵因素。
7. 雖然巨量資料帶來許多新的可能，但過與不及，都會造成不同程度的問題。請仿造上課講義中提到的「大數據的黑暗面」，各組延續第一題的應用範例，說明發展巨量資料需要考量哪些技術面及非技術面的潛在問題。