[作業1] 請上 Kaggle, 在Competitions或Dataset中找一組競賽或資料並寫下：

網頁：<https://www.kaggle.com/unitednations/global-commodity-trade-statistics#commodity_trade_statistics_data.csv>

1. 你選的這組資料為何重要：

可知全球進出口貿易現況，並從過去資料推測未來貿易走勢。但此份資料不包含各國對各國數據，若有此相關數據也可找出某產品的主力市場與競爭對手國。

1. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)：

聯合國公布的相關數據

1. 蒐集而來的資料型態為何：

為結構化資料，包含10個欄位，如下表

|  |  |  |
| --- | --- | --- |
| 欄位名稱 | 資料型態 | 內容 |
| country\_or\_area | 文字 | 207個國家與歐盟 |
| year | 數字 | 1988年-2016年 |
| comm\_code | 文字 | 6碼的HSCODE商品帶碼(csv上面有些是5碼須補0，故應屬文字型態資料) |
| commodity | 文字 | 產品名稱 |
| flow | 文字 | 進口、出口、復進口、復出口 |
| trade\_usd | 數字 | 商品金額(美元) |
| weight\_kg | 數字 | 商品重量(公斤) |
| quantity\_name | 文字 | 商品單位(個數、公尺、公斤、公升) |
| quantity | 數字 | 商品單位量質 |
| category | 文字 | 6碼產品分類(分成12大類) |

1. 這組資料想解決的問題如何評估

可利用時間序列來推測未來走勢

[作業2] 想像你經營一個自由載客車隊，你希望能透過數據分析以提升業績，請你思考並描述你如何規劃整體的分析/解決方案：

1. 核心問題為何 (tips：如何定義 「提升業績 & 你的假設」)

如何增加每日的載客數、減少空車時間與增加客人回客率

1. 資料從何而來 (tips：哪些資料可能會對你想問的問題產生影響 & 資料如何蒐集)

內部資料：從歷史資料得知那些地區與時段是搭乘熱點

外部資料：演唱會或大型室內或室外場合舉行的日期、時間與民眾討論熱度，精準推估要派多少台車，或者是否優優惠方案吸引乘客搭乘

1. 蒐集而來的資料型態為何

內部資料：結構化資料

外部資料：結構化資料或非結構化資料(如網路輿情聲量)

1. 你要回答的問題，其如何評估 (tips：你的假設如何驗證)

可利用回歸模型解決問題，比較利用數據分析前後「是否增加每日的載客數」、「是否減少空車時間」與「是否增加客人回客率」。並可利用RMSE來推估此模型分析情形。