

作業 1：

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

Competitions-Google Landmark Recognition Challenge

1. 你選的這組資料為何重要
可使遊客旅遊時對當地建築、文化有更深入地了解, 並且可節省時間在查詢建築資訊上。
2. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)
Google Landmark Recognition Challenge and Google Landmark Retrieval Challenge 計畫提供。
Google 的使用者上傳的公開圖片
3. 蒐集而來的資料型態為何
.jpg
4. 這組資料想解決的問題如何評估
用 Global Average Precision (GAP)評估。每張圖片透過預測會有一個對應之地標及信心水準, 然後在依照 GAP 的評估公式計算出平均準確率。

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」)
假設可透過改善車隊行駛路線, 使司機可更快速接送乘客、減少空車時間比例與增加載客率。
2. 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)
 1. 事前事後車隊行駛路線資料、空車時間
 2. 乘客搭乘資訊(上車地點、搭乘時間)資料來源: 車隊派遣紀錄資料檔
3. 蒐集而來的資料型態為何
文字檔、excel
4. 你要回答的問題, 其如何評估 (tips: 你的假設如何驗證)
先分析過去車隊行駛路線與最佳路線之差異, 針對此處改善, 之後在其他條件不變下, 比較路線時間改善前後之差異與平均載客時間/平均行駛時間是否有顯著關聯。