作業1：

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

1. 你選的這組資料為何重要

Subject: Chest X-Ray Images (Pneumonia)

This is a very popular subject which AI can effectively predict and gain the good result.

1. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care

1. 蒐集而來的資料型態為何

Dataset of validated OCT and Chest X-Ray images described and analyzed in "Deep learning-based classification and referral of treatable human diseases". The OCT Images are split into a training set and a testing set of independent patients. OCT Images are labeled as (disease)-(randomized patient ID)-(image number by this patient) and split into 4 directories: CNV, DME, DRUSEN, and NORMAL.

1. 這組資料想解決的問題如何評估

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

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」)

「提升業績 & 你的假設」: 車隊載客率提高

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)

例如: Uber 有APP可以收集和分析資料

載客地點、天氣、週日周末、節日、一天時段

1. 蒐集而來的資料型態為何

載客日期、地點、天氣、週日周末、節日、一天時段

4. 你要回答的問題，其如何評估 (tips：你的假設如何驗證)