作業1：

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

Recruit Restaurant Visitor Forecasting

Predict how many future visitors a restaurant will receive

<https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting>

1. 你選的這組資料為何重要

餐飲在全世界各國的生產毛額都是高佔比的產業，而開一家餐館，耗損率會是很重要的成本比例，預估食材的備料量是每個經營者每天都要面對的重要課題，若沒控制好，很有可能是餐廳經營成敗的關鍵因素。

1. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)

Recruit Holdings從Hot Pepper Gourmet美食折價網站、AirREGI POS機資訊及Restaurant Board預定資料蒐集而來。

1. 蒐集而來的資料型態為何

有三個平台各自的資料集，有餐廳的id及地理資訊，需要做交叉比對而找出餐廳的預定資料及實際客流資訊。

1. 這組資料想解決的問題如何評估

利用評價與訂位，在後續實際入店消費的關係，來做建模並預測。

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 ->提升載客數 ->降低司機上班時空車率 -> 調度司機在每個時段到熱點

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)

除了空車率是結果指標，蒐集司機在每個區域成功載客的歷史資料，做二維熱點分布及第三維度的時間

1. 蒐集而來的資料型態為何

地區資料為區域，但區域的資料須為數值型態，區域彼此的距離也是一個欄位，另有時間欄位

1. 你要回答的問題，其如何評估 (tips：你的假設如何驗證)

以空車率的指標的升降來評估方案成效

作業3：

請點選下方檢視範例依照 Day\_001\_example\_of\_metrics.ipynb 完成 Mean Squared Error 的函式

資料夾結構建立規則提醒：2nd-ML100Days > data資料夾 & homework資料夾 (ipynb檔)

(請注意data的存放位置，建議放在\*.ipynb 同一個目錄下，這樣才能在不修改code的情況下正常執行)

def MSEfun(X,Y,a,b):

temp = []

for i in range(len(X)):

num = Y[i] - a\*X[i] - b

num = pow(num,2)

temp.append(num)

all\_sum = sum(temp)

MSE = all\_sum / len(X)

return MSE