

◆ 一共有 4 個 csv 檔案，分別為

- train\_feature.csv
- train\_label.csv
- test\_feature.csv
- sample\_submission.csv

train\_feature.csv 為用來訓練資料的特徵，空白的儲存格即為缺失值，可能由於蒐集資料中的各種原因沒有記錄到，所以用與不用取決於自己，如果要使用通常會以眾數或平均數補充值。

train\_label.csv 的資料為 train\_feature.csv 資料中對應列的房價。

test\_feature.csv 的資料即為需要預測的新資料，我們並不知道這份資料的真實房價為何，我們需要利用歷史資訊來訓練模型，幫助我們判斷房價。

sample\_submission.csv 為要求輸出的格式，Index 對應於 test\_feature.csv，需要輸出指定形狀的房價資料。

◆ 注意事項:

1. 當讀取 train\_feature.csv 以及 test\_feature.csv 時，若使用 read\_csv 請在函式要包含 ,encoding='big5')，才可順利讀取資料。
2. 輸出必須是與 sample\_submission.csv 相同格式。

◆ Kaggle Grading Scheme :

Your RMSE A	Points
$A \leq \mu - 2\sigma$	100
$A \leq \mu - 1/2\sigma$	$\geq 95$
$A \leq \mu + \sigma$	$\geq 85$
$A \leq \mu + 2\sigma$	$\geq 80$
$A \leq 20000$ (min baseline)	$\geq 70$

◆ 期末分數計算:

- Kaggle 分數排名占期末 70%
- PPT 占 30% (資料前處理、視覺化、模型選擇與介紹、結果比較...等)