一共有4個csv檔案，分別為

train\_feature.csv

train\_label.csv

test\_feature.csv

sample\_submission.csv

這四個資料是用於預測該地點房價的數值。

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。

具體流程可參考: <https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/649776/>

https://ithelp.ithome.com.tw/articles/10235449

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。