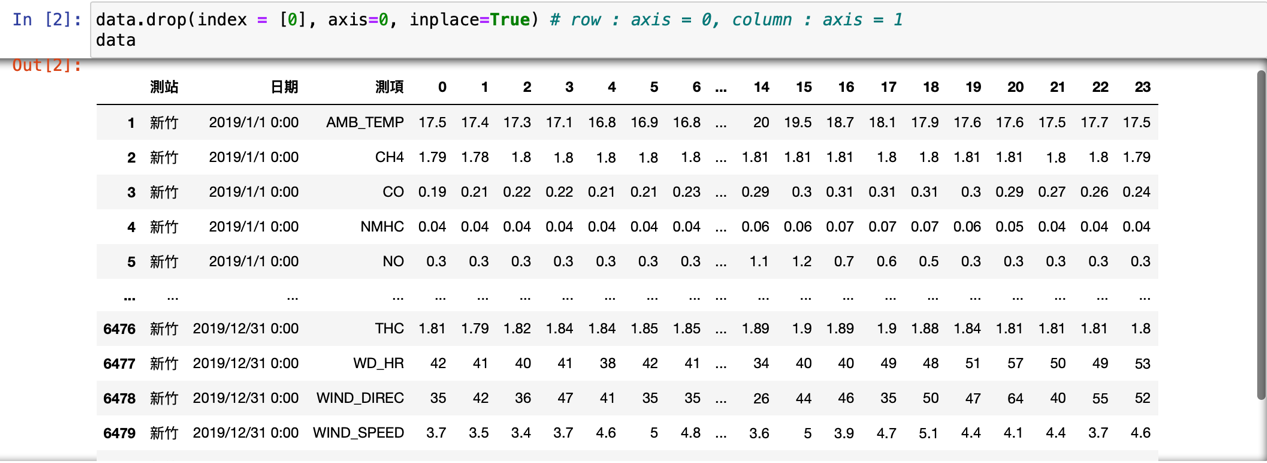
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 將未來第一個小時當預測目標，X只有PM2.5 | 將未來第六個小時當預測目標 X只有PM2.5 | 將未來第一個小時當預測目標， x取所有18種屬性 | 將未來第六個小時當預測目標 X有所有18種屬性 |
| Linear Regression | 2.6136558911044343 | 4.839609070011748 | 3.7329306223569962 | 3.7604398413161895 |
| Random Forest Regression | 3.285230352303523 | 5.560027285129604 | 3.5657181571815717 | 3.690995907230559 |

結果：

1.載入資料，因為資料是中文編碼，utf-8會有亂碼，所以用big5



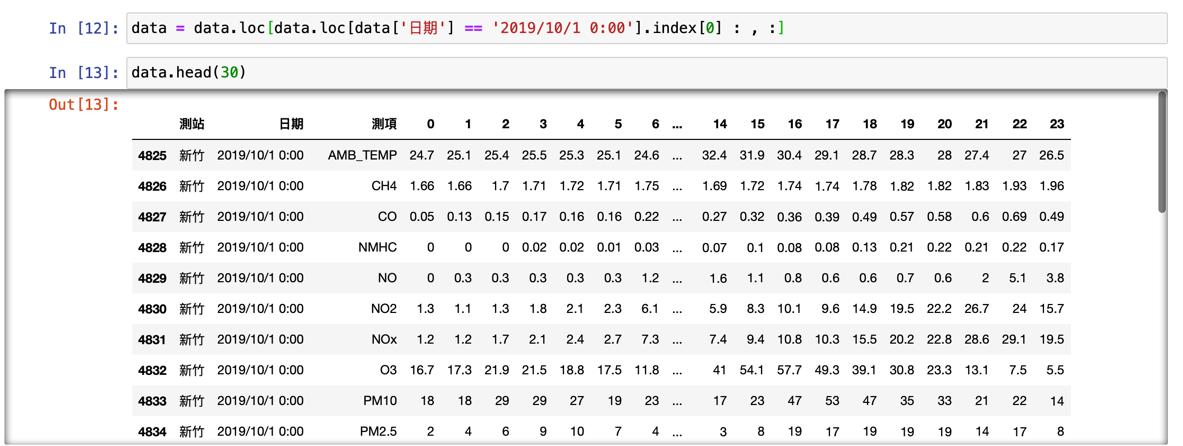
2. 清理資料



1. 處理前後空白字元



(2) 拿10月之後的資料



(3) NR表示無降雨，以0取代



(4) 把怪怪的字元拿掉，用前後兩個值的平均



掃描data的方式由左至右、由上至下

這邊是分三種狀況：

a. 如果取到了’測項’那欄，那就要回前18個列取前一天24時的資料。

b. 如果取到了超過data欄位的資料，那就要去之後18個列取下一天0時的資料。

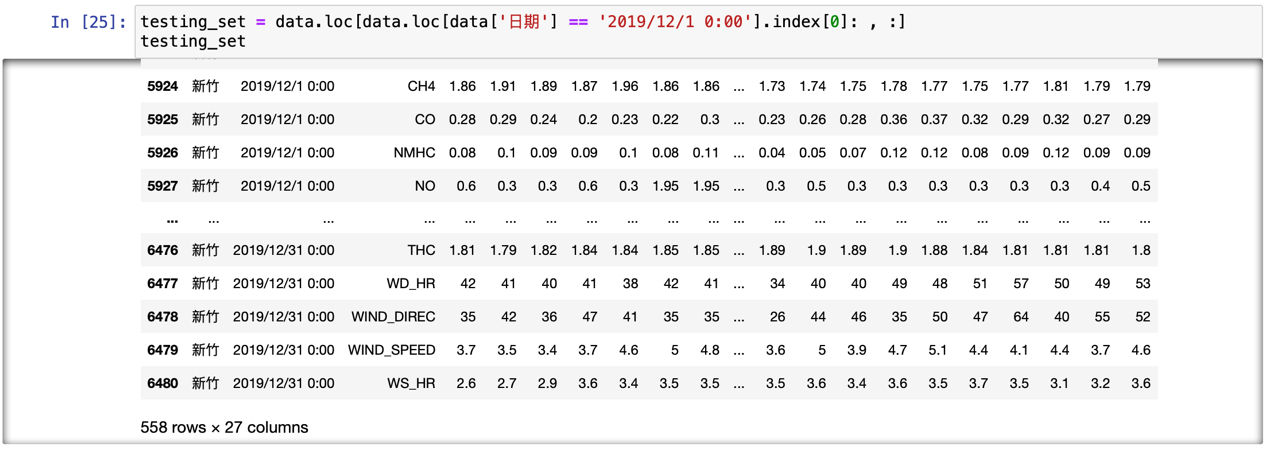
c. 剩下的狀況就是目前有異常值，但還要繼續檢查下一個是不是異常值。

(5) 將資料切割成訓練集(10.11月)以及測試集(12月)



(6) 再把每小時的資料從object transfer to float，testing\_set也做一樣的事

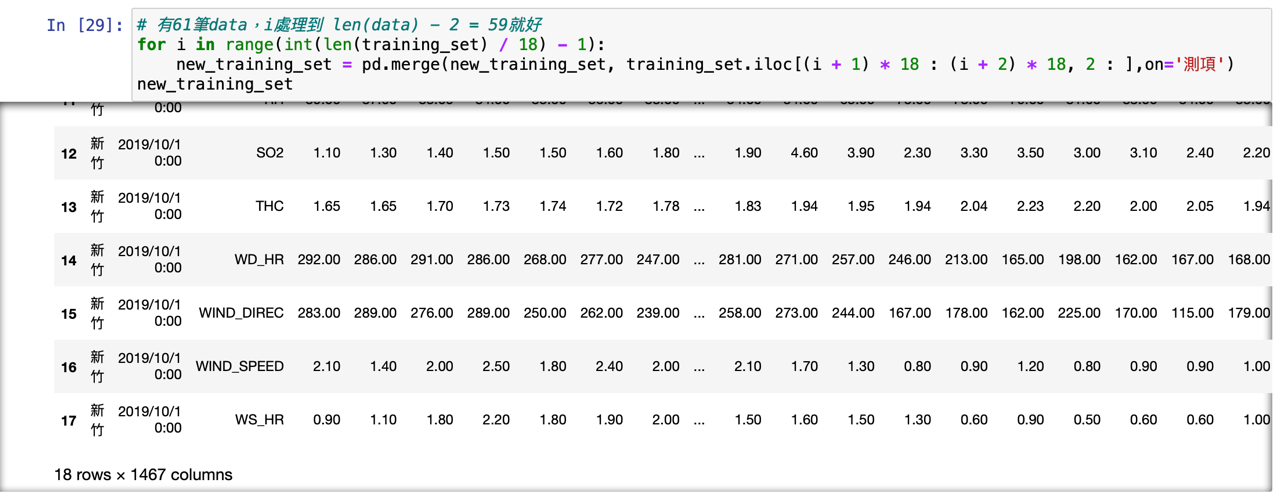




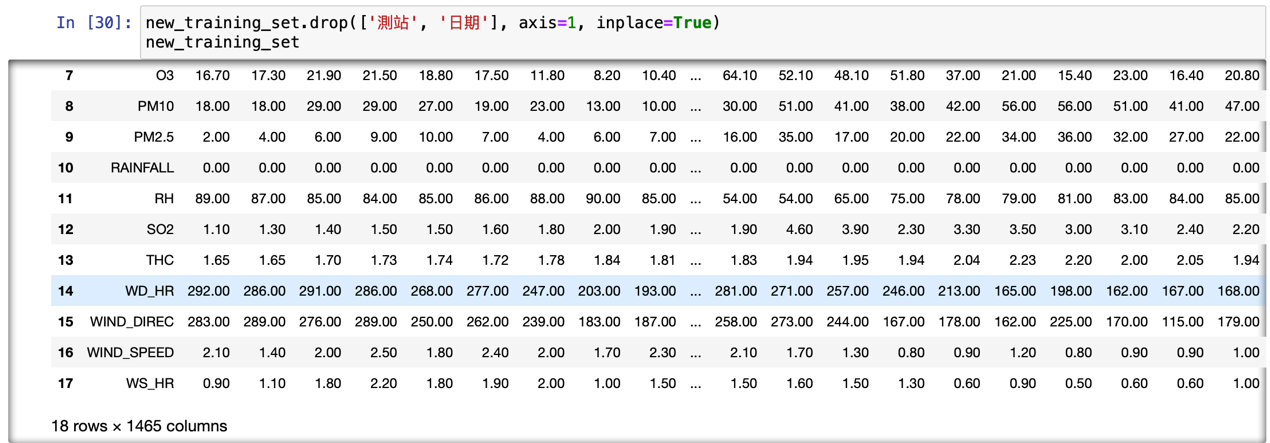


(7) 製作時序資料: 將資料形式轉換為行(row)代表18種屬性，欄(column)代表逐時數據資料 hint: 將訓練集每18行合併，轉換成維度為(18,61*24)的DataFrame(每個屬性都有61天*24小時共1464筆資料)

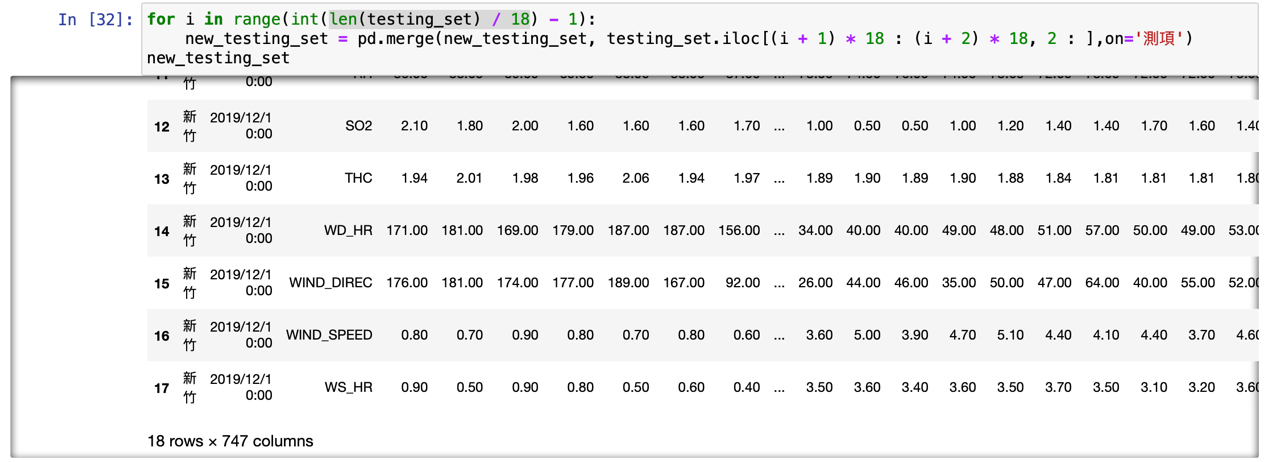
Training\_set

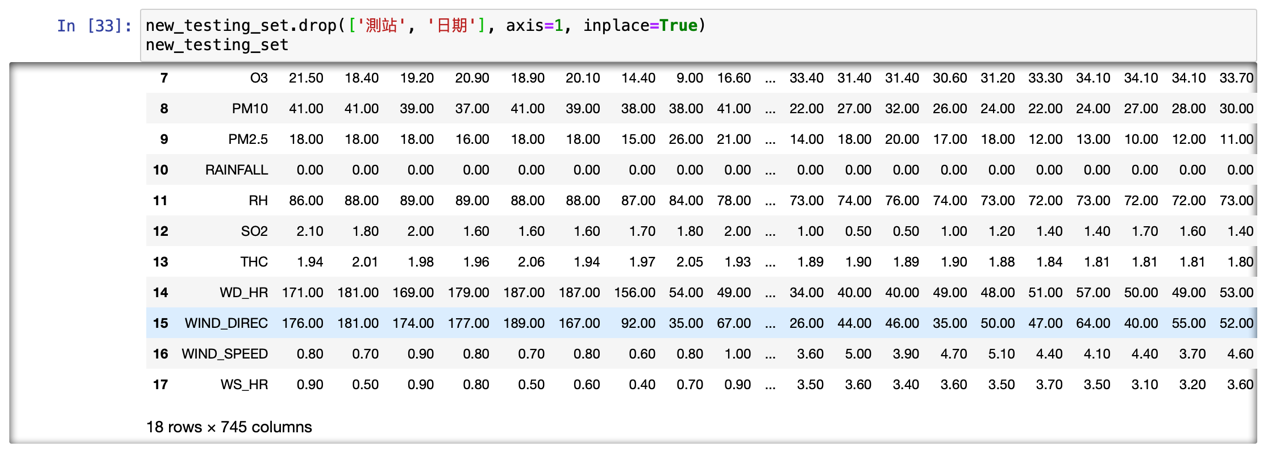


再把不想留的column拿掉



Testing\_set

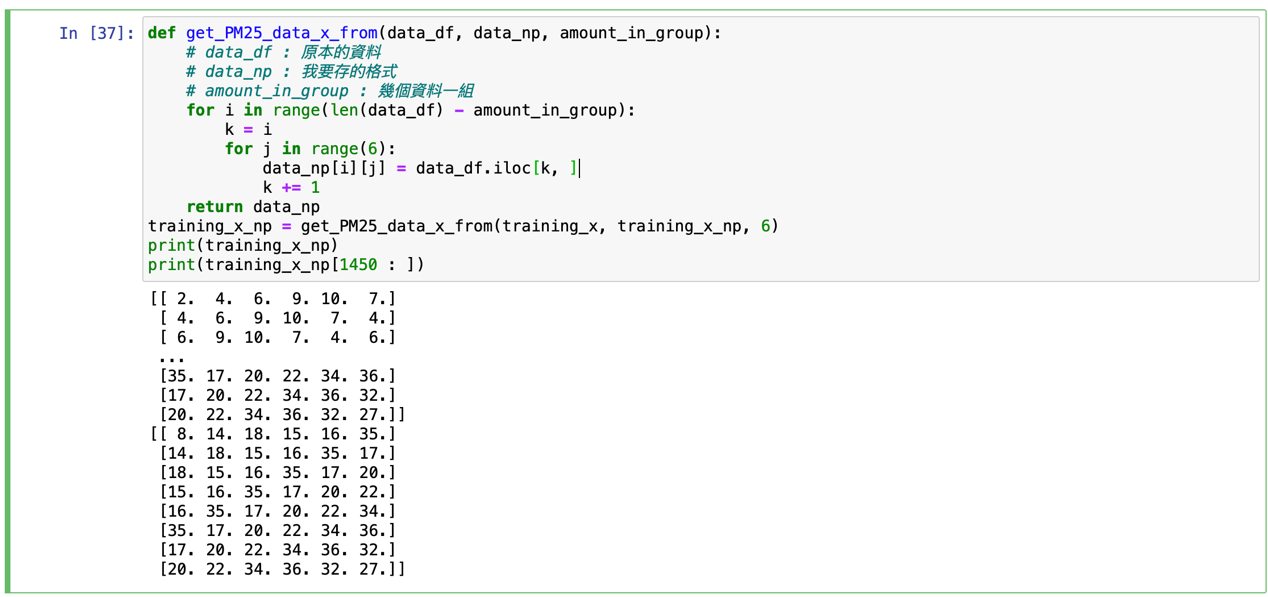


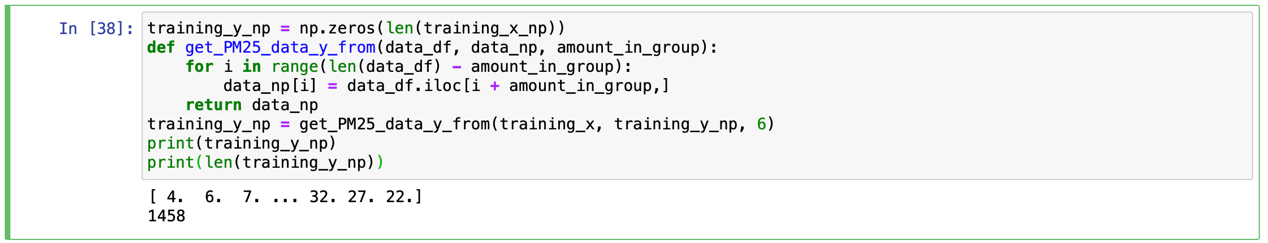


3. 將未來第一個小時當預測目標，X只有PM2.5 (e.g. X[0]會有6個特徵，即第0~5小時的PM2.5數值)

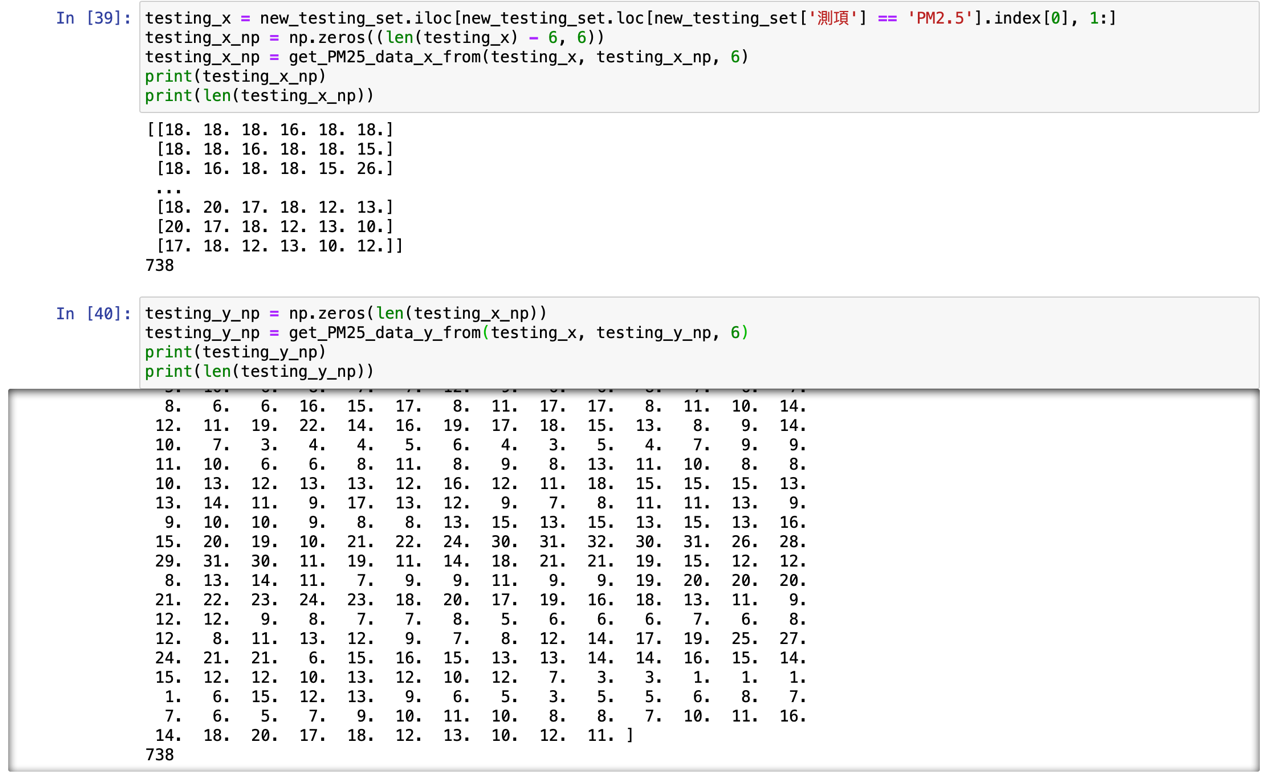
(1)這裡整理training\_set的x, y 資料



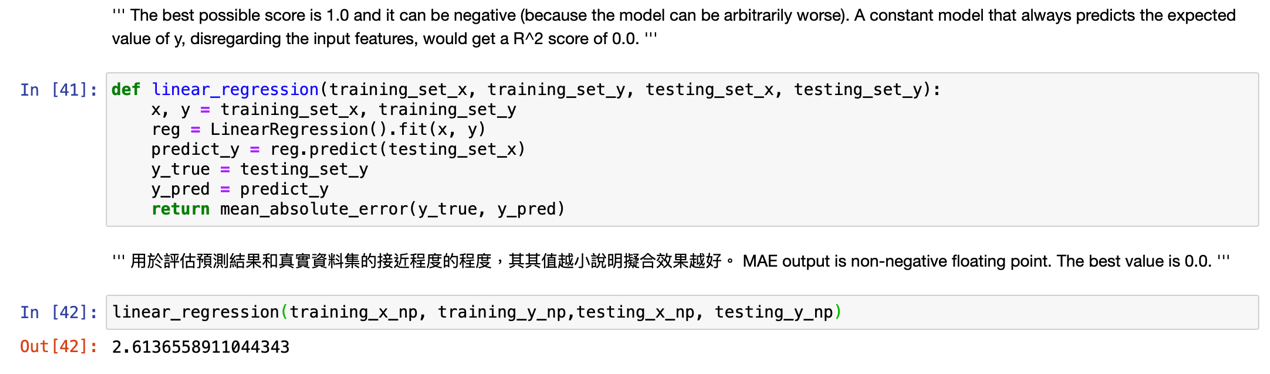




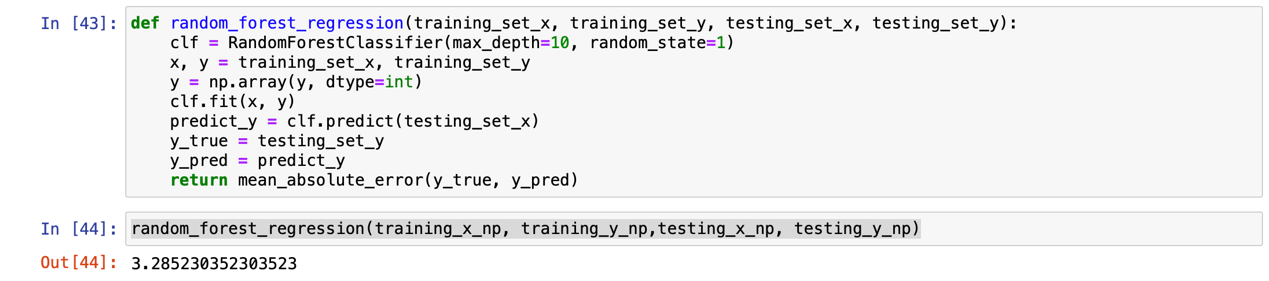
(2) 這裡整理testing\_set的x, y 資料



(3) 做Linear Regression

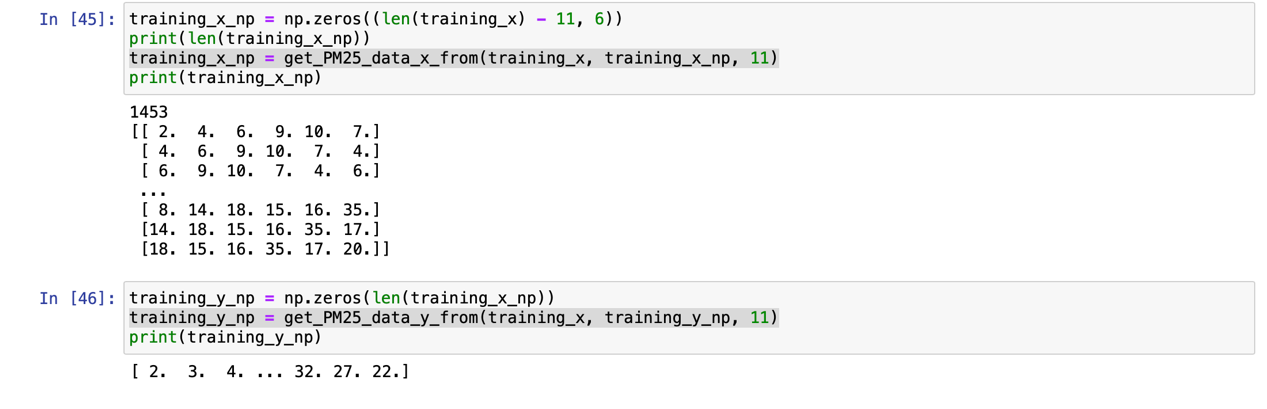


(4) 做Random Forest Regression

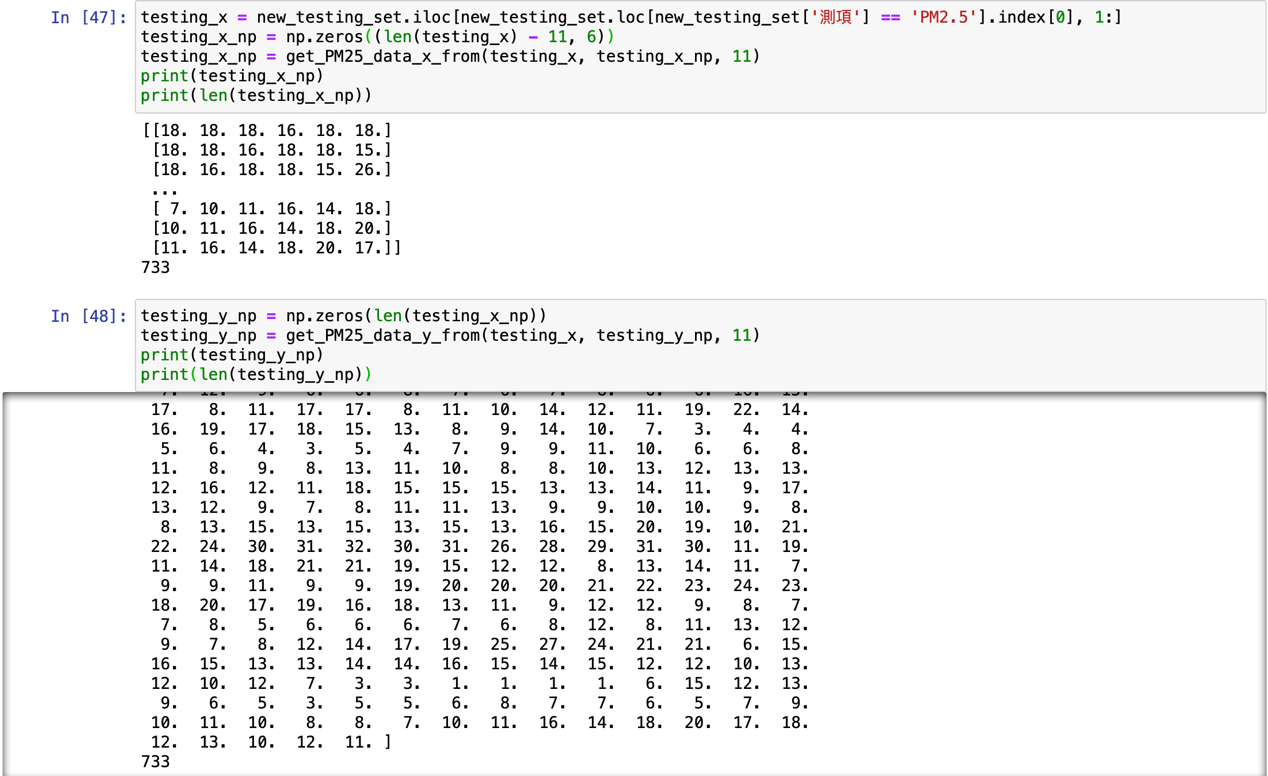


4. 將未來第六個小時當預測目標 X只有PM2.5 (e.g. X[0]會有6個特徵，即第0~5小時的PM2.5數值)

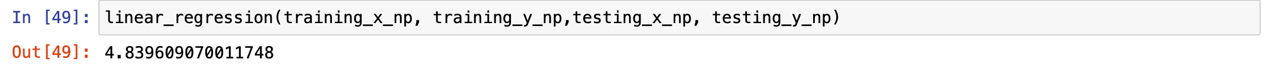
(1) 這裡整理training\_set的x, y 資料



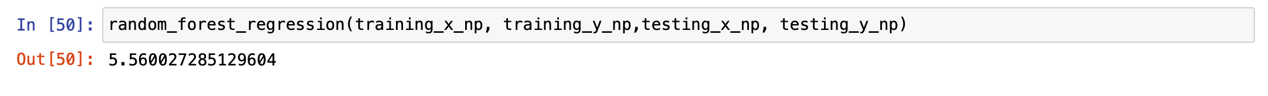
1. 這裡整理testing\_set的x, y 資料



(3) 做Linear Regression

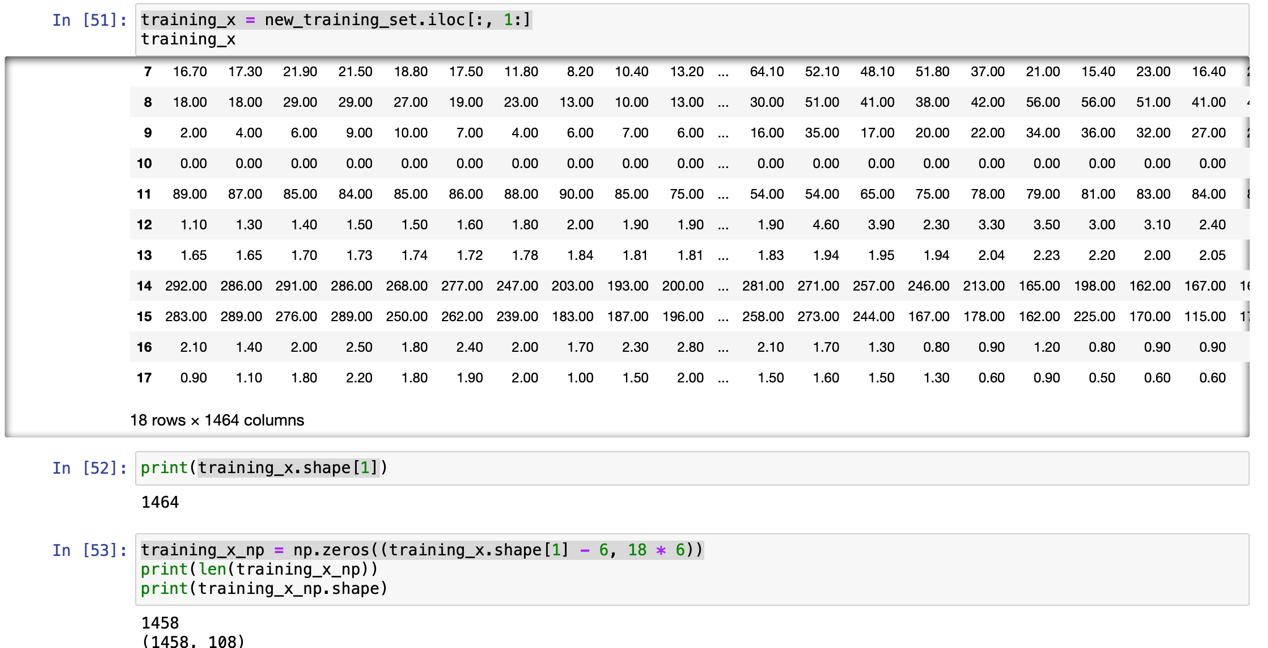


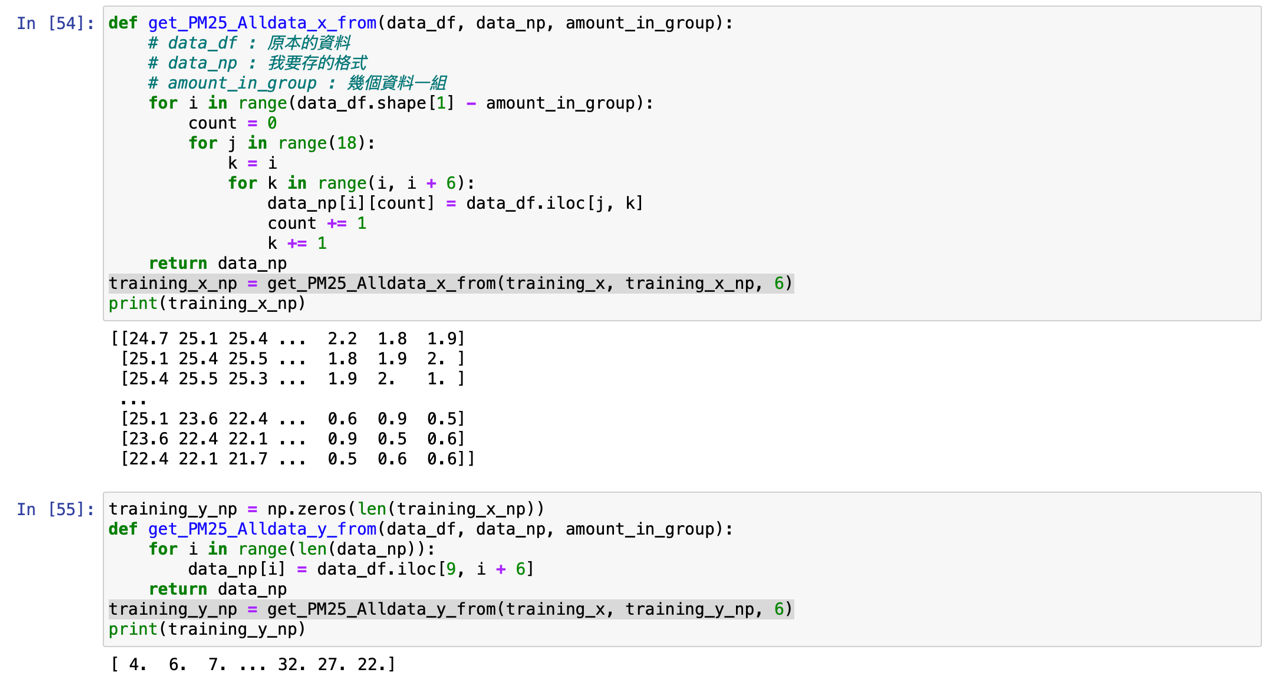
(4) 做Random Forest Regression



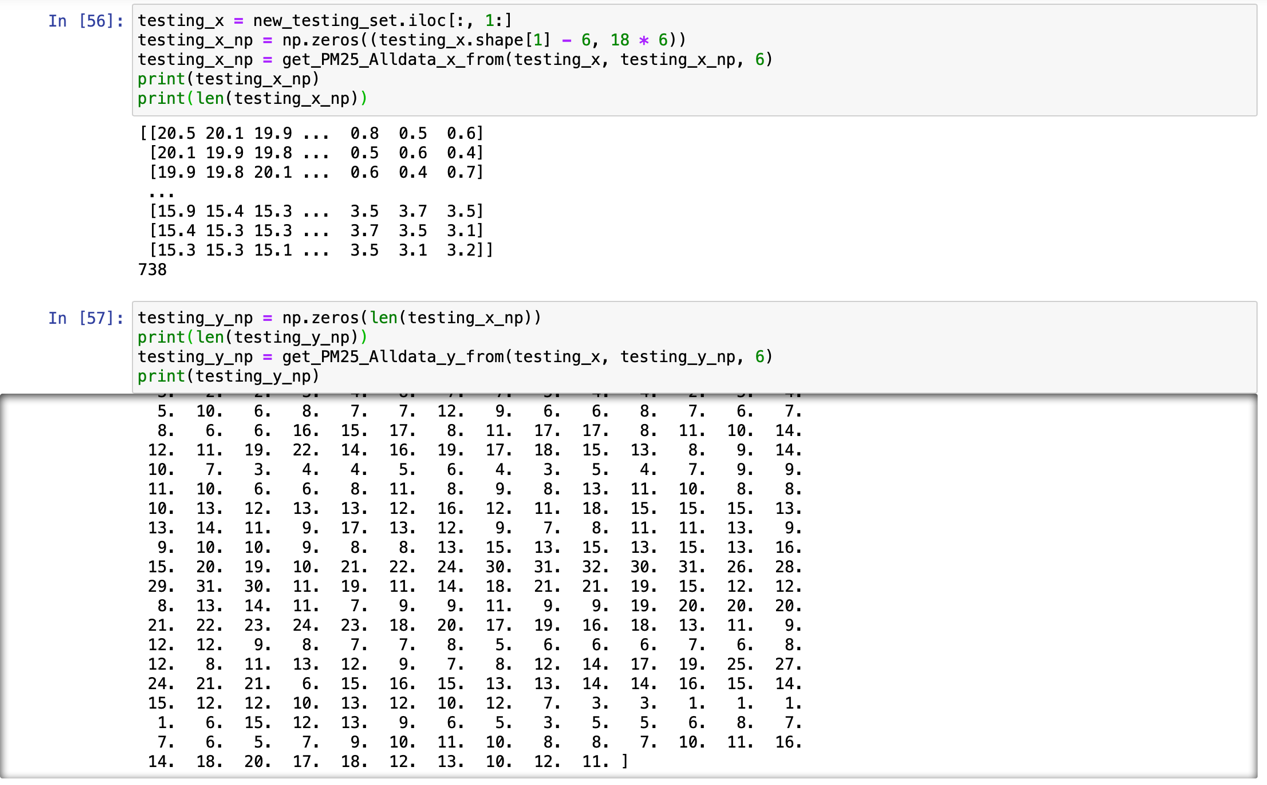
5. 將未來第一個小時當預測目標， x取所有18種屬性 (e.g. X[0]會有18\*6個特徵，即第0~5小時的所有18種屬性數值)

(1) 這裡整理training\_set的x, y 資料

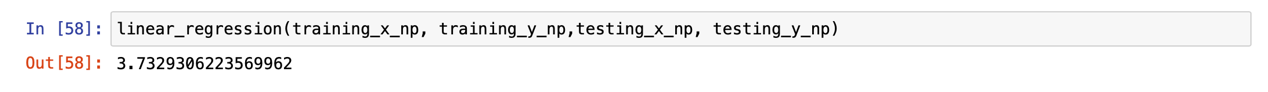




(2) 再處理testing set的x, y



(3) 做Linear Regression



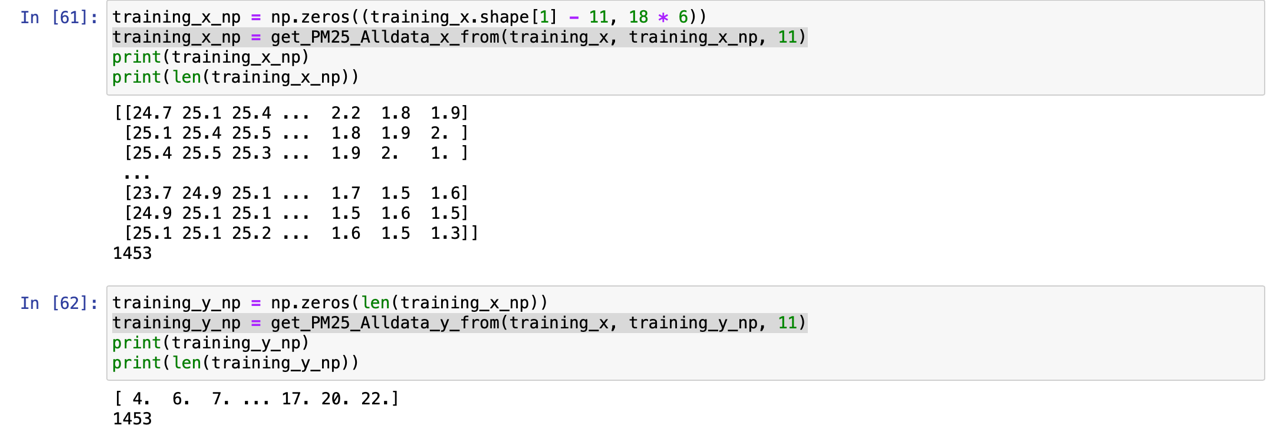
(4) 做Random Forest Regression



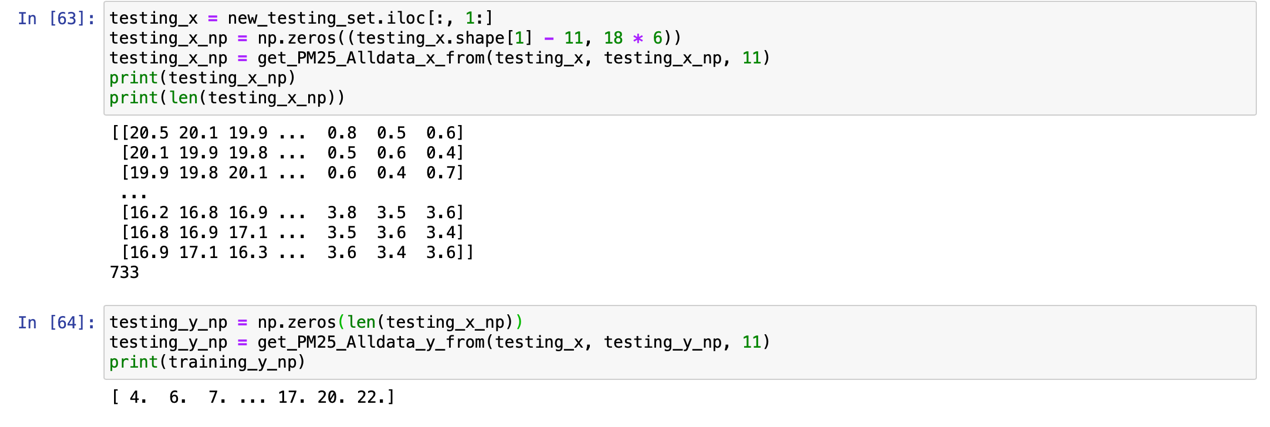
6. 將未來第六個小時當預測目標 X有所有18種屬性 (e.g. X[0]會有18\*6個特徵，即第0~5小時的所有18種屬性數值)

(1) 這裡整理training\_set的x, y 資料

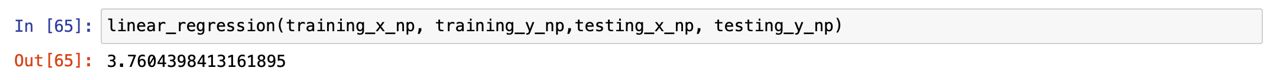




(2)再處理testing set的x, y



(3)做Linear Regression



(4)做Random Forest Regression

