

# D15 : EDA from Correlation

[簡報閱讀](#)[範例與作業](#)[問題討論](#)

## 相關係數實作 Coding 練習日



## 知識地圖

### 機器學習概論 Introduction of Machine Learning

#### 監督式學習 Supervised Learning



#### 非監督式學習 Unsupervised Learning



### 探索式數據分析 Exploratory Data Analysis (EDA)

#### 統計值的視覺化

相關係數	繪圖排版
核密度函數	常用圖形
離散化	模型體驗

相關係數實作 Coding 練習日

## 練習重點

練習重點



Day10 相關係數實作



**Tips:** 遇到  $y$  的本質不是連續數值時



**Tips:** 檢視不同數值範圍的變數



解題時間



- 可以用相關係數來迅速找到和預測目標最有線性關係的變數
- 相關係數通常搭配散佈圖來一起了解預測目標與變數的關係

## Day10 相關係數實作

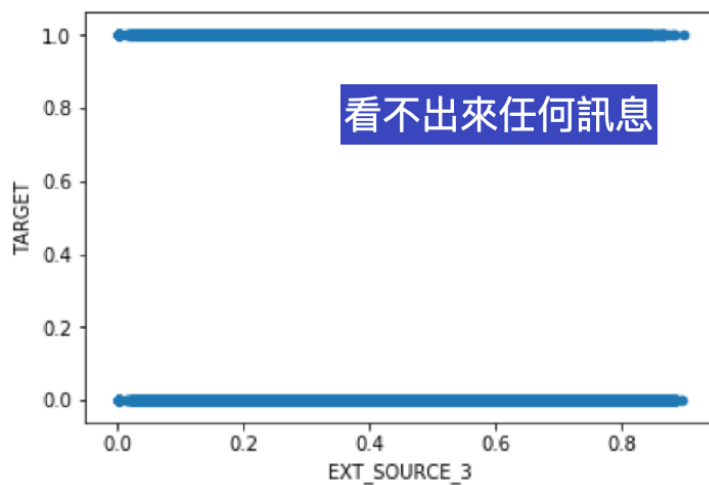
請於官網檢視範例參考範例程式碼與HW檔

- 列出目標 (TARGET) 與所有欄位之間相關係數，數值最大以及最小各 15 個
- 通過相關係數的結果觀察有興趣的欄位與 TARGET 或其他欄位的相關係數，並嘗試找出有趣的訊息

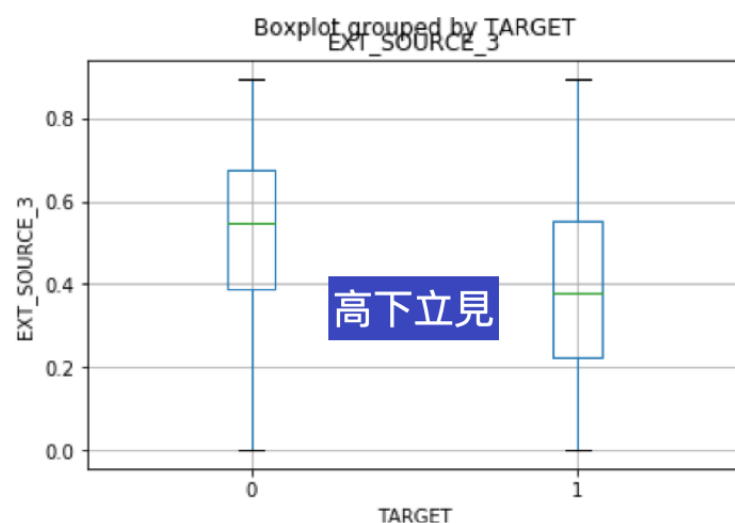
最好的方式當然是畫圖，舉例來說，我們知道 EXT\_SOURCE\_3 這個欄位和 TARGET 之間的相關係數是 -0.178919 (在已經這個資料集已經是最負的了!)，那我們可以 EXT\_SOURCE\_3 為  $x$  軸，TARGET 為  $y$  軸，把資料給畫出來。

**Tips:** 遇到  $y$  的本質不是連續數值

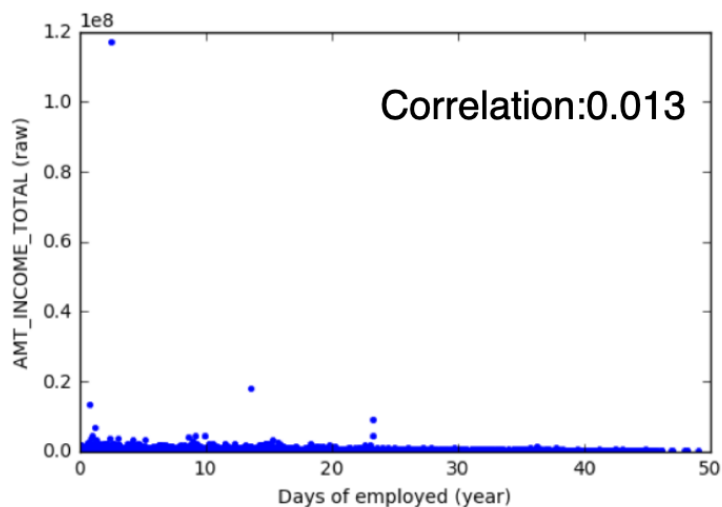




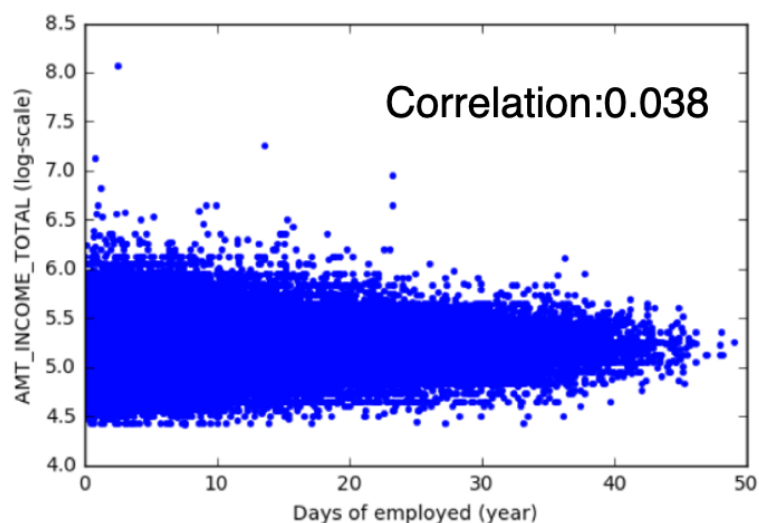
可以換一個角度來看



**Tips:** 檢視不同數值範圍的變數



將 Y 軸轉換 (log-scale)



## 解題時間



Sample Code & 作業  
開始解題



[下一步：閱讀範例與完成作業](#)

