

# eupoyy

Day 15 探索式數據分析

# 相關係數實作



Coding 練習日



游為翔/杜靖愷



### 知識地圖 探索式數據分析 相關係數的EDA



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

# 監督式學習 Supervised Learning

探索式 數據分析 前處理 **Exploratory** Processing Data Analysis

特徵 工程 Feature Engineering

模型 選擇 Model selection

參數調整 Fine-tuning

集成 Ensemble 非監督式學習 **Unsupervised Learning** 

> 分群 Clustering

降維 Dimension Reduction

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

#### 統計值的視覺化

相關係數 核密度函數 離散化

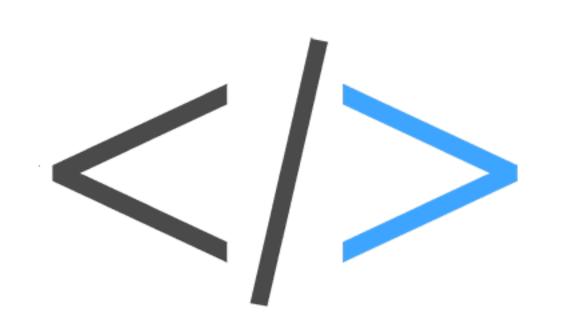
繪圖排版

常用圖形

模型體驗

### 練習重點





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



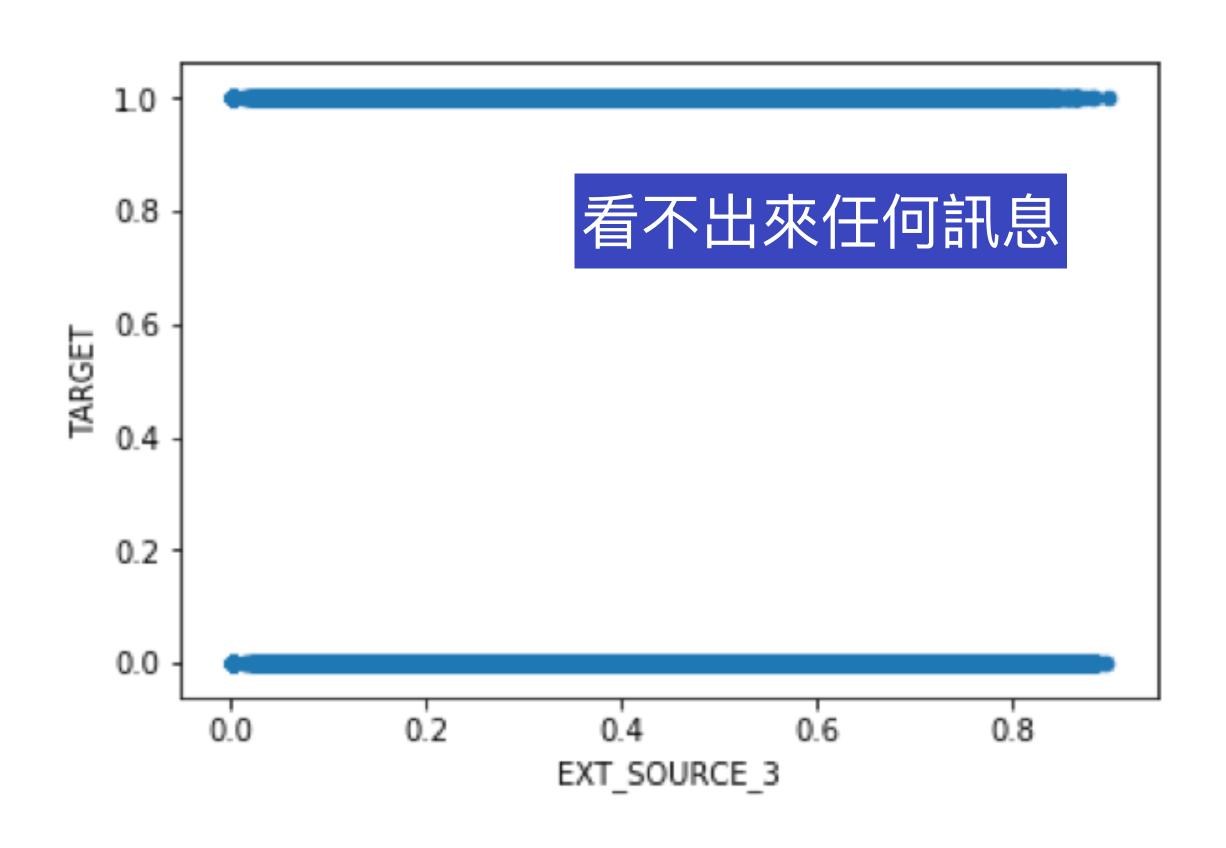
#### 請於官網檢視範例參考範例程式碼與HW檔

- 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。

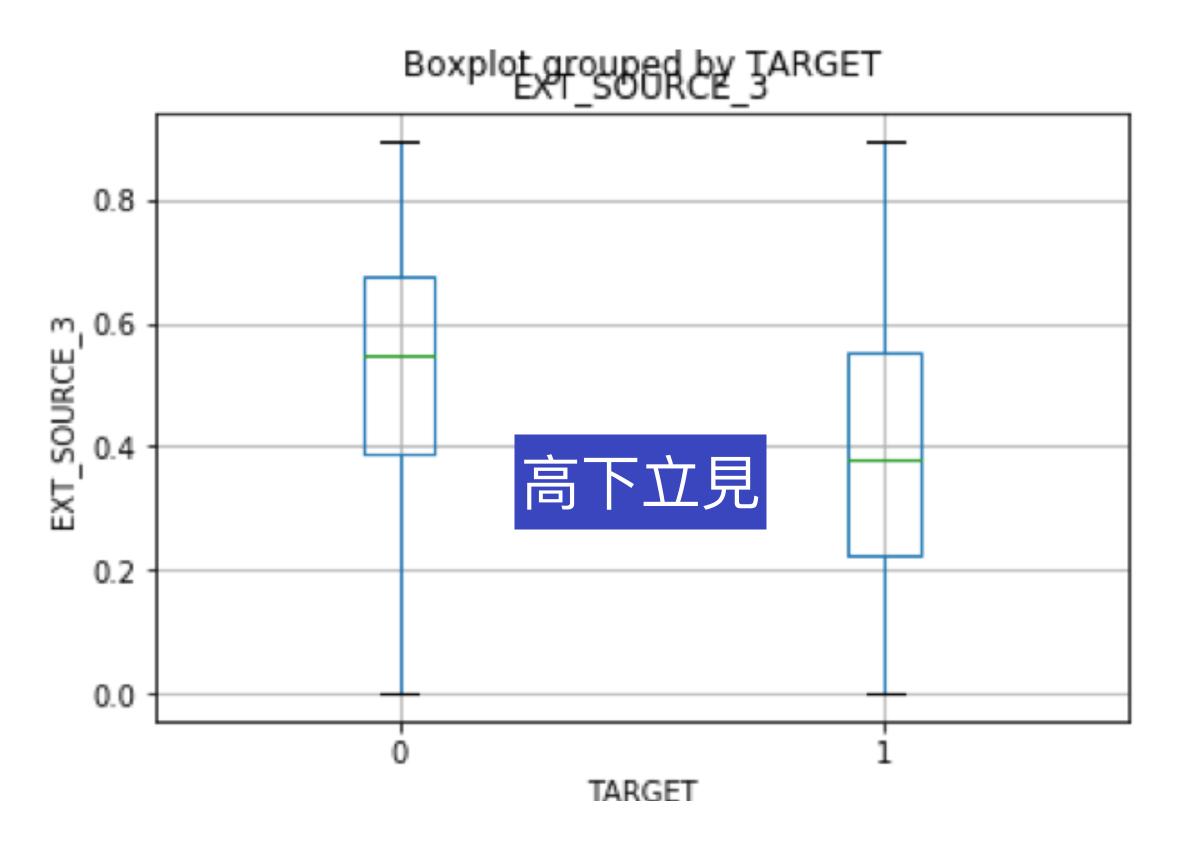
# Tips: 遇到 y 的本質不是連續數值時



#### 直接以原始數值繪圖



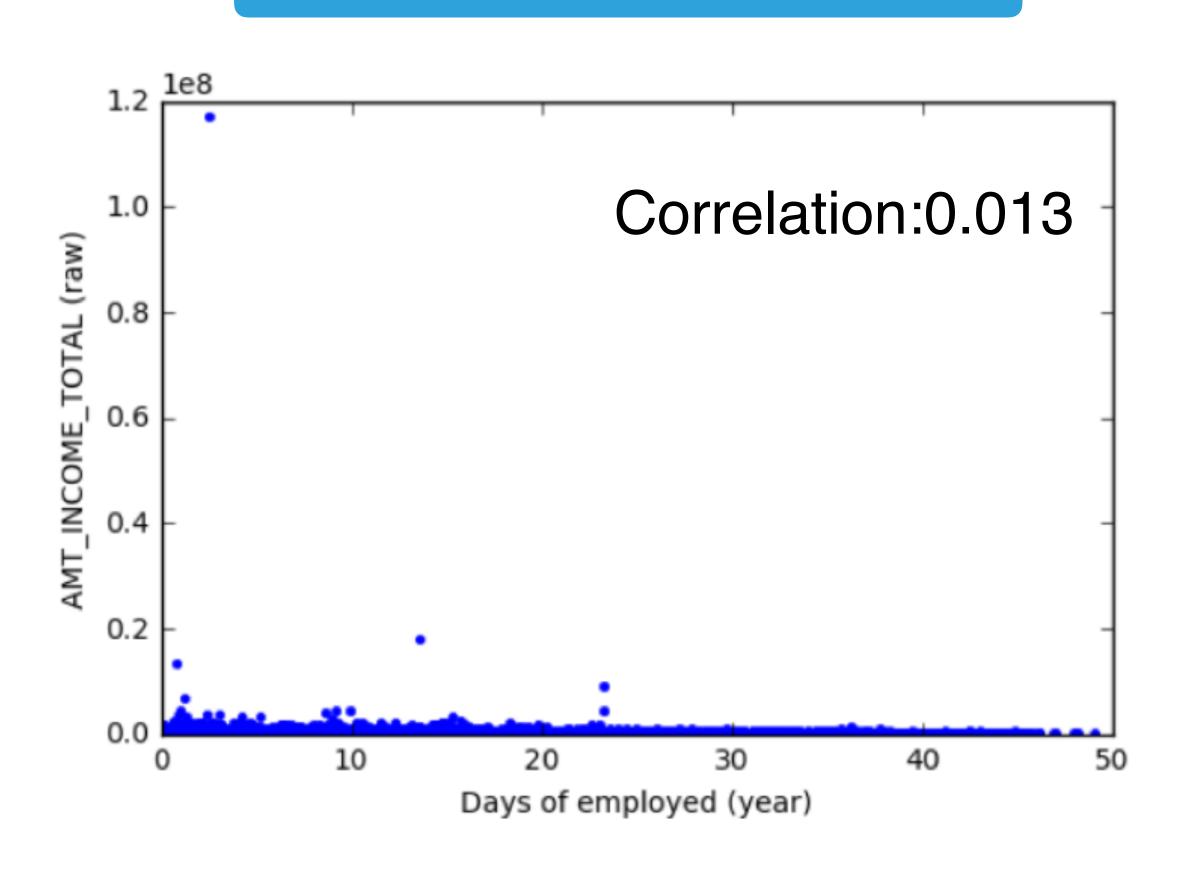
#### 可以換一個角度來看



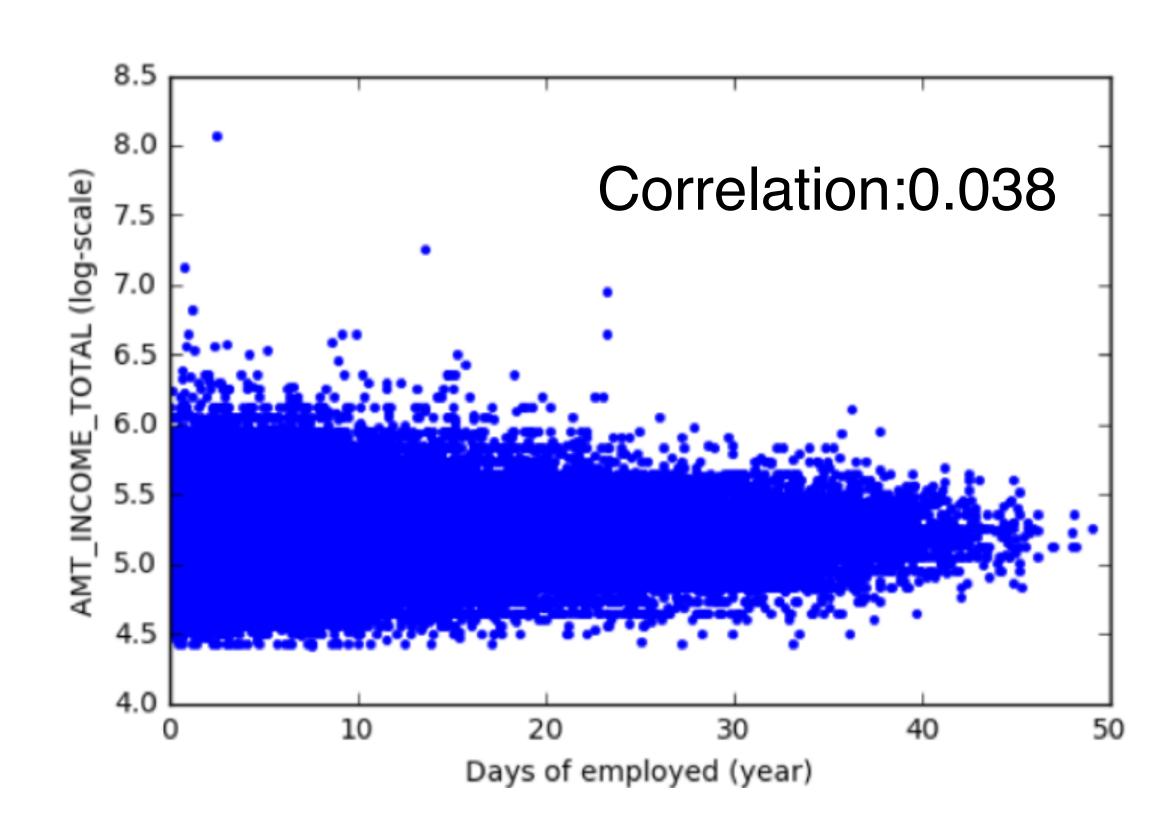
# Tips: 檢視不同數值範圍的變數



#### 直接以原始數值繪圖



#### 將Y軸轉換(log-scale)





請跳出PDF至官網Sample Code&作業 開始解題

