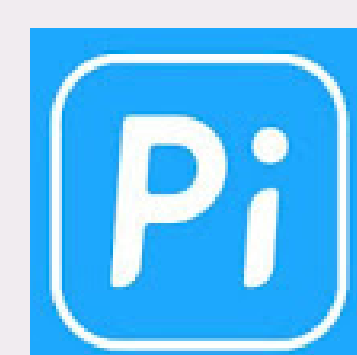
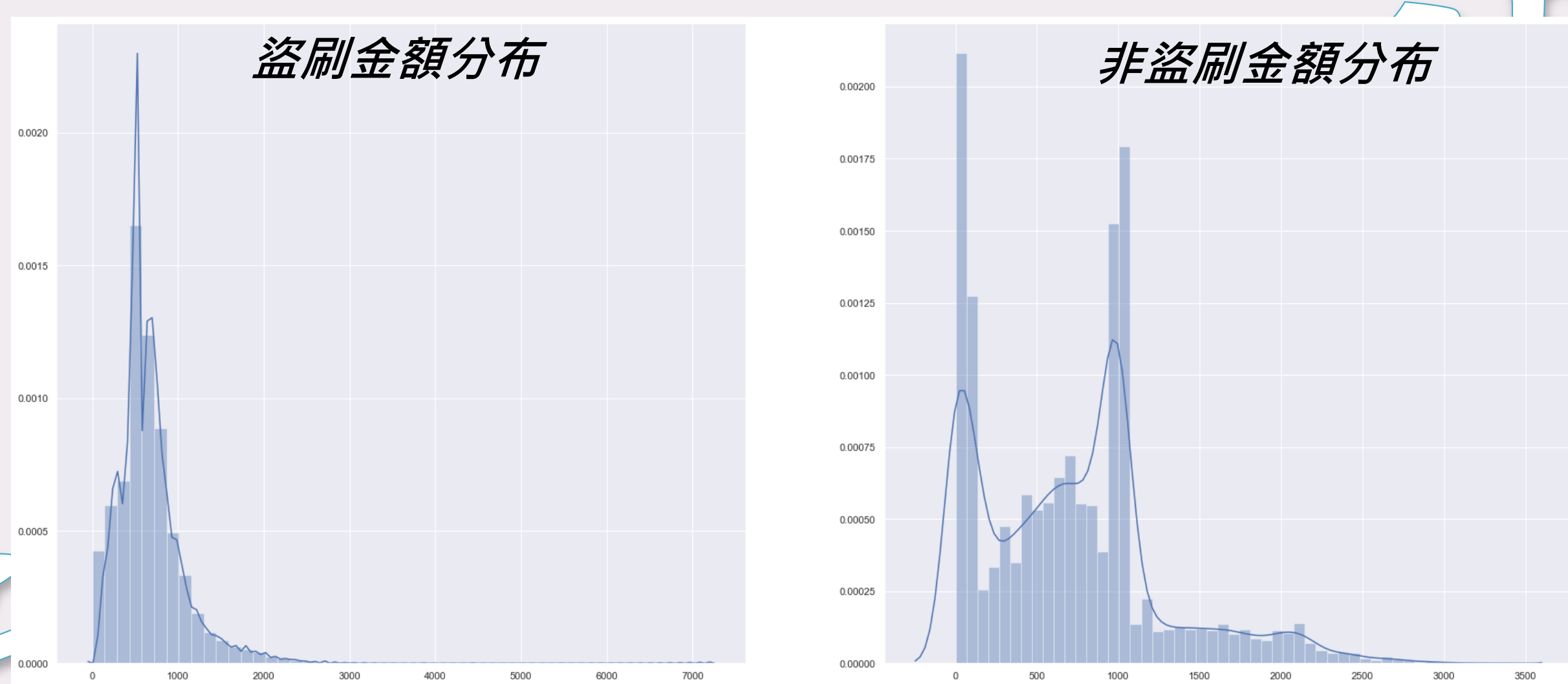


## Assume

# 從觀察資料對比賽做了哪些假設? 消費面切入

- 經常性消費居多 - 消費金額集中在NTD3000內
- 視卡號為最小顆粒度 - 即使一人多卡，也會因活動模式不同而產生不同的消費模式
- 分析玉山信用卡商品類別 - 利用歷史消費偏好(mcc, 國外消費, 網購消費)作為卡片類別的判斷依據，依此對每張卡號進行客戶貼標



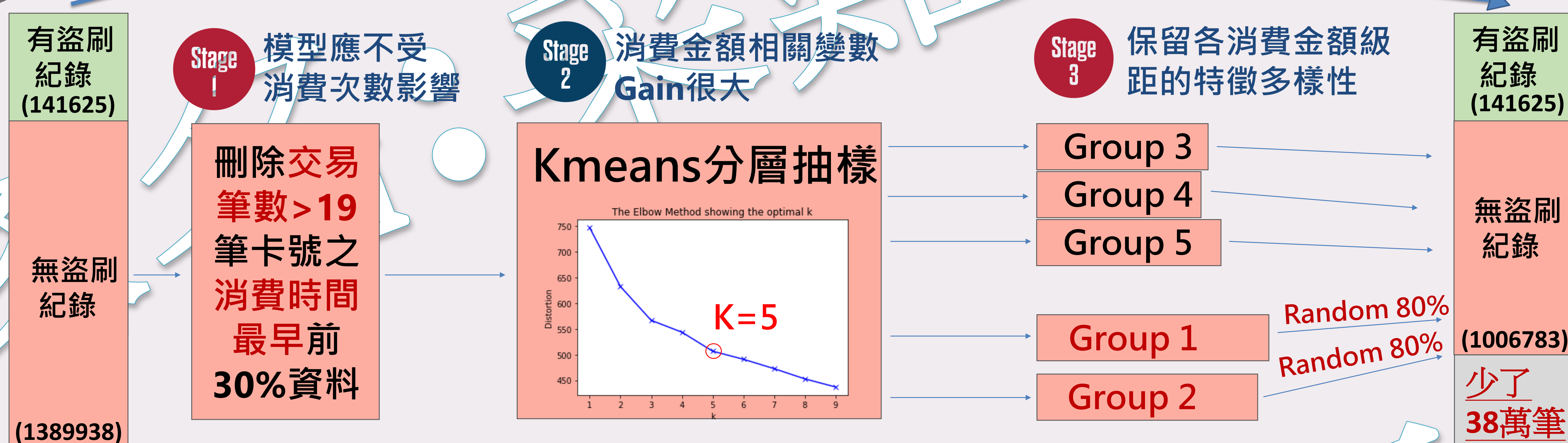
## Diff

# 基於信用卡特性做了哪些特徵? 抓出異常差異

- 特定通路任務有加碼，幹嘛去刷其他通路?  
- 統計每張卡號在此筆消費以前，過去消費紀錄之特店代號眾數類別
- 每張卡的消費模式會隨意變動?  
- 計算本筆消費金額與過去平均消費金額之差異
- 消費分數差異!  
- 統計常用消費組合並依Baseline Model之特徵重要度給予權重

## How

# 如何處理不平衡資料? 對無盜刷紀錄資料做Undersampling



## Model

# 編碼, stacking & RIPPER演算法觀察分裂規則

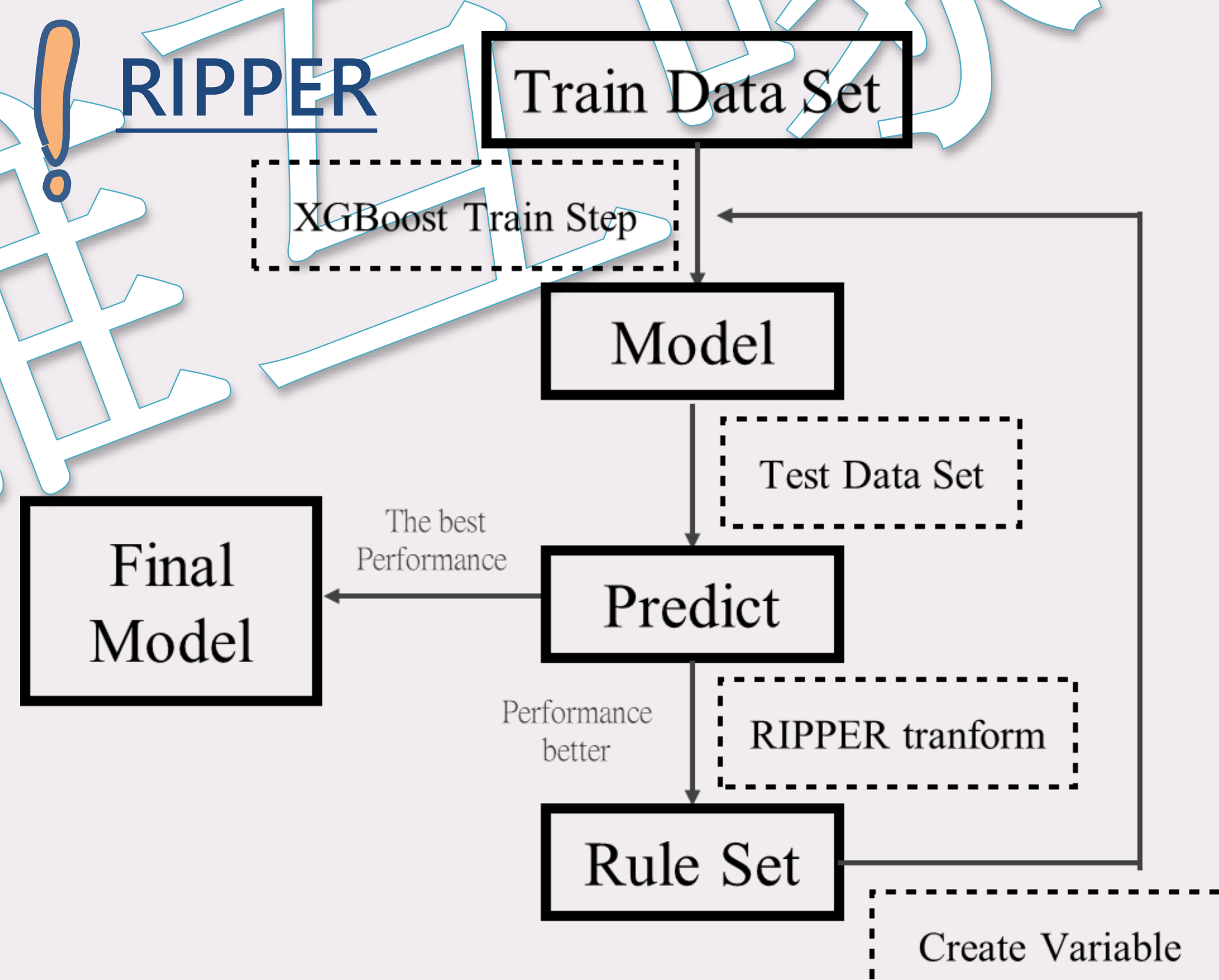
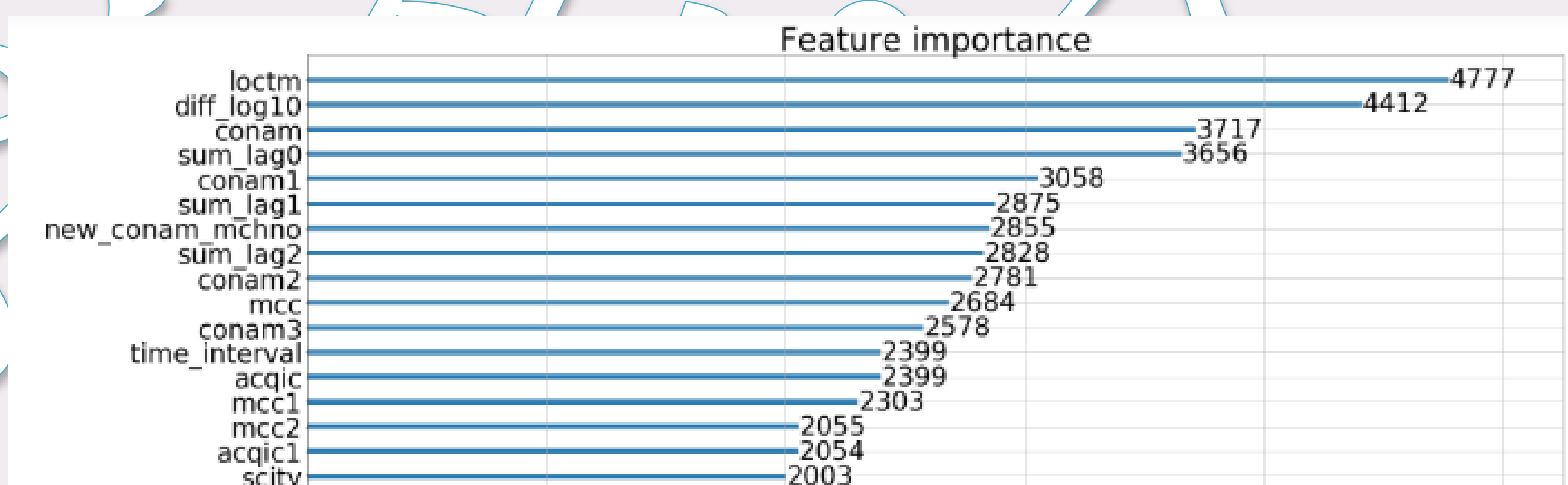
! Target Encoding & 一對一函數轉換對類別變數進行編碼

RIPPER

Train Data Set

! StratifiedKFold + GridSearch  
決定XGBoost最適參數

! Stacking XGBoost & LightGBM



F1 Private Score = 0.668