

# 信用卡違約預防策略

Default of Credit Card Clients Dataset

台灣信用卡客戶為例

羅嘉承

---

資料來源:<https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset>



# Outline

---

01.商業痛點

02.執行策略

03.資料介紹

04.模型建立

05.方案建議





商業痛點

客戶信用卡違約率  
高達22%



# 客戶時常無法如期還款，或者發生不良債務，這將對銀行的資產和業務產生負面影響

## 遇到的問題

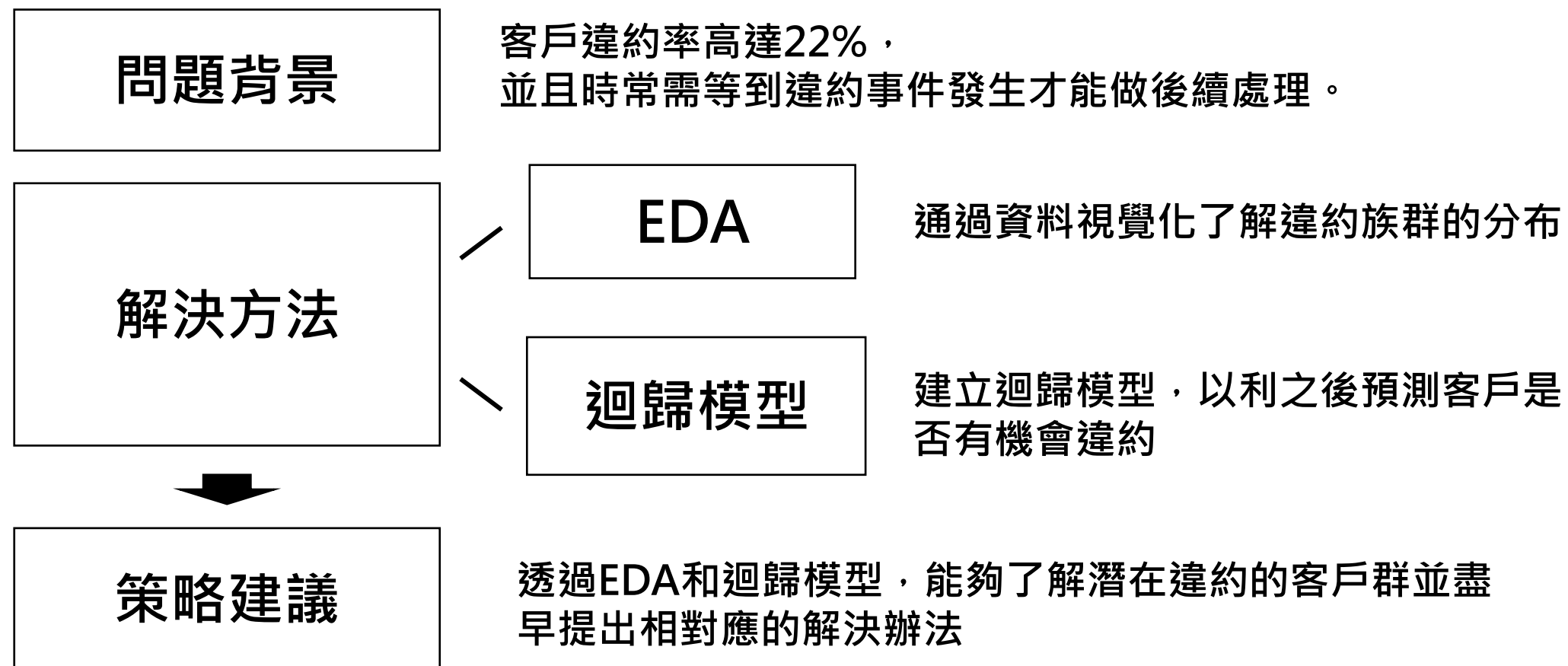
- 不了解違約率高的客群，無法提前注意。
- 往往必須等到違約發生才開始有動作。



## 需求

- 希望能夠更加深入了解客戶的相關訊息及消費模式，已發現潛在違約的可能性。
- 及時發現潛在違約，以利後續的方案處理。

# 執行策略：分析輪廓特徵並輔以迴歸模型驗證





# 資料介紹



# 30,000筆資料 包含客戶資訊以及信用卡資訊

客戶資訊

年齡

婚姻

教育

性別

信用卡資訊

信用  
額度

帳單  
金額

付款  
金額

還款  
狀態

(4~9月的資料，以月為單位)

# 類別資料介紹

## 婚姻

- 已婚 : 1
- 單身 : 2
- 其他 : 3

## 教育

- 研究所 : 1
- 大學 : 2
- 高中 : 3
- 其他 : 4

## 性別

- 男性 : 1
- 女性 : 2

## 還款狀態

- 按時還款 : -1
- 一個月遲繳 : 1
- 兩個月遲繳 : 2
- .
- .
- .
- 九個月及以上 : 9

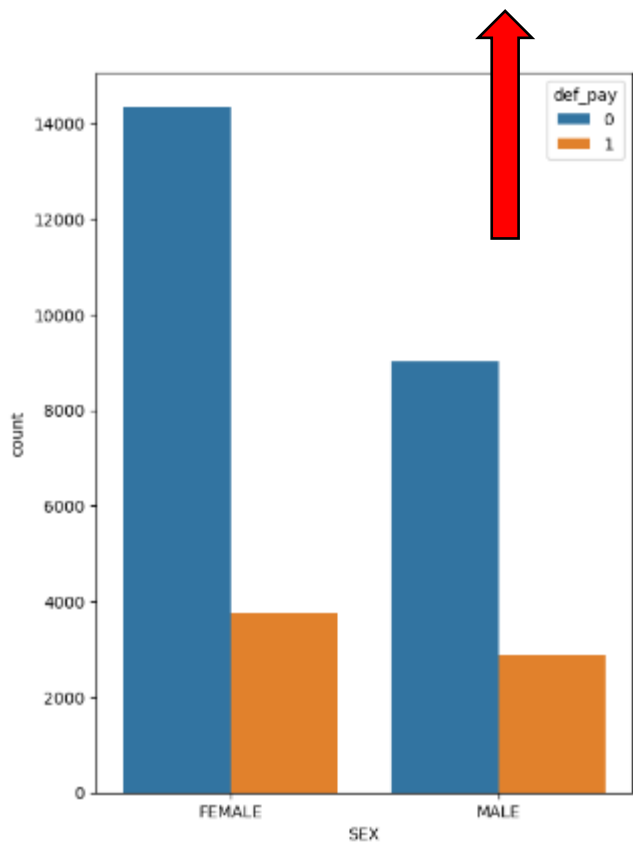




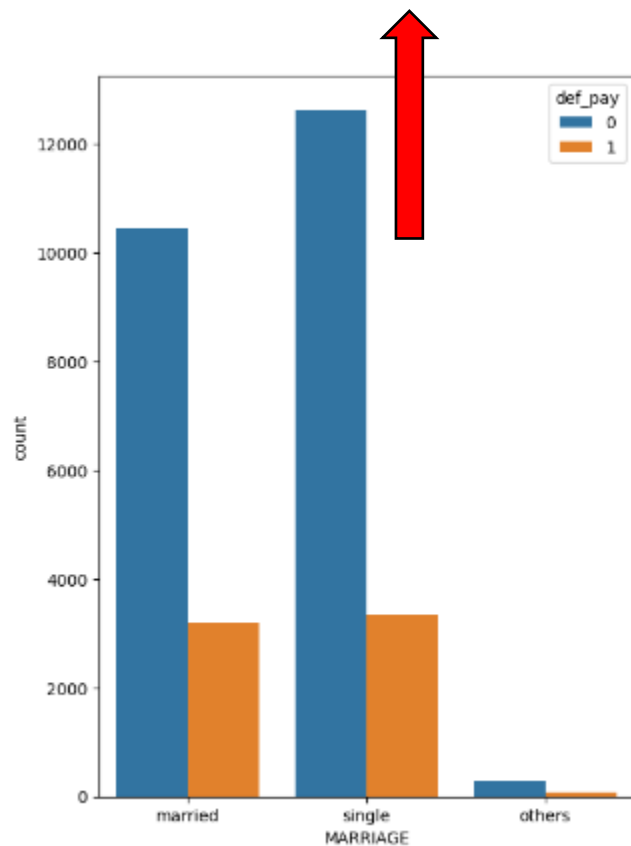
# 怎樣的客戶更有 違約的可能性



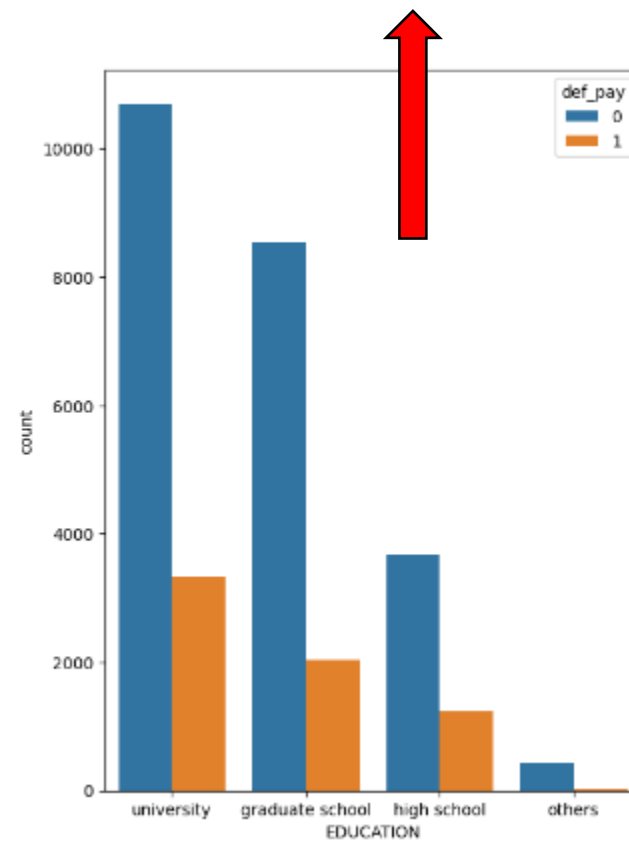
男性違約率比女性多 4%



已婚違約率比單身多 3%

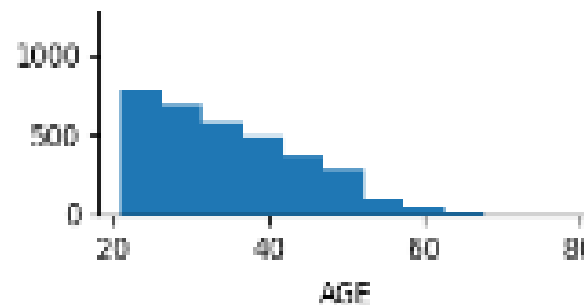


違約者中受過高等教育的比率較高

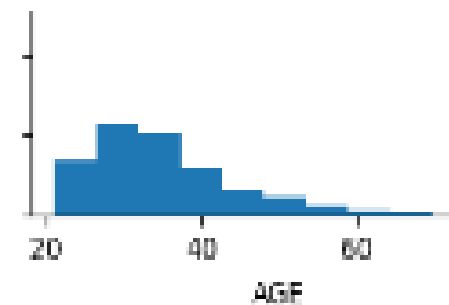


# 結合年齡資訊

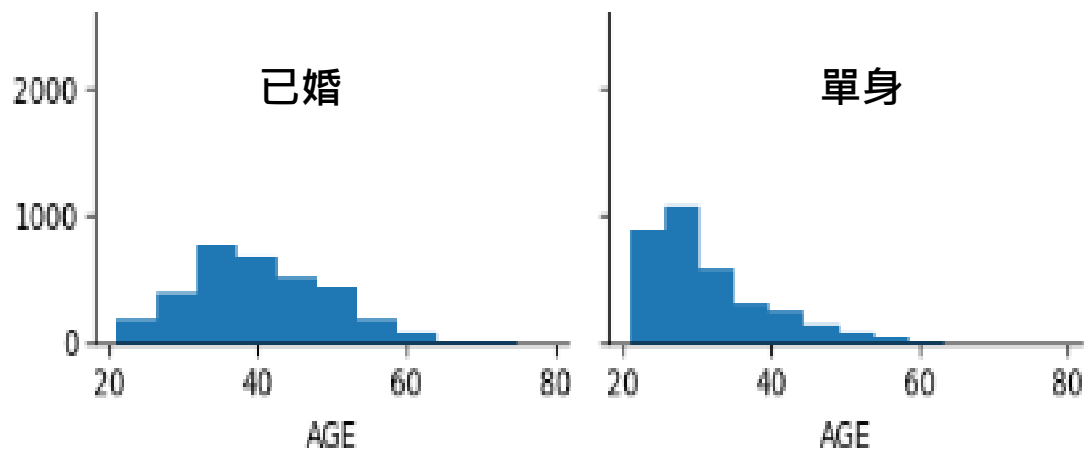
研究所



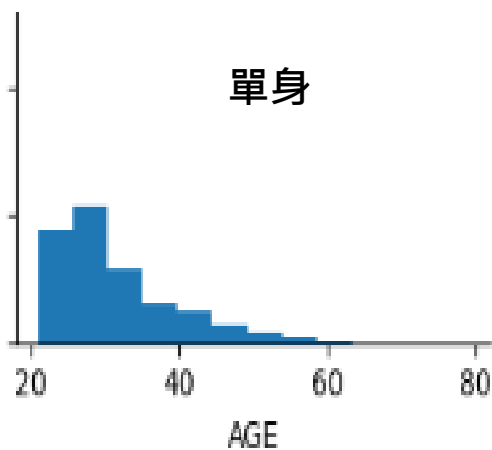
大學



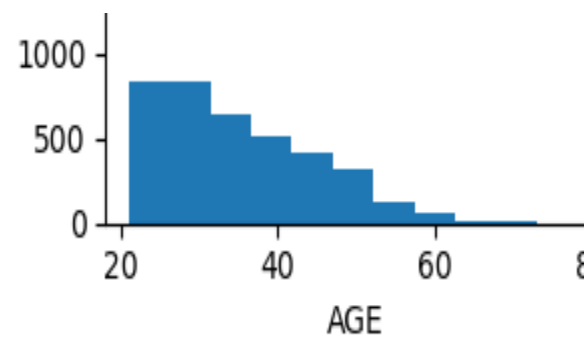
已婚



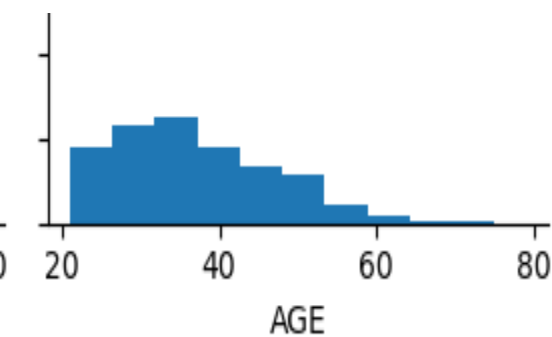
單身



女性



男性



## 小結

擁有以下特徵輪廓：  
年齡介於20~40歲、男性、已婚、  
教育程度較高(大學以上)者  
更傾向會信用卡違約





# 迴歸模型建立



選用解釋性高之模型，了解各變數對於目標之影響

## 預測模型建立

羅吉斯迴歸 Logistic Regression

解釋性高，篩選重要特徵對目標變數正負影響

決策樹 Decision Tree

可視化預測過程，篩選重要特徵

## 資料清理

- 未知類別處理
- 缺失值處理

## 特徵工程

- 資料集平衡

## 模型選取

- 決策樹
- 羅吉斯回歸

# 資料清理

## 未知類別處理 缺失值處理

處理資料時發現到有與特徵說明不符的情況，以婚姻為例：  
已婚 = 1、單身 = 2、其他 = 3，  
卻出現有資料標籤為0的情況，  
可以視為未知類別或是該資料有  
缺失值，因此以“其他”代替。

# 資料集平衡

## SMOTE

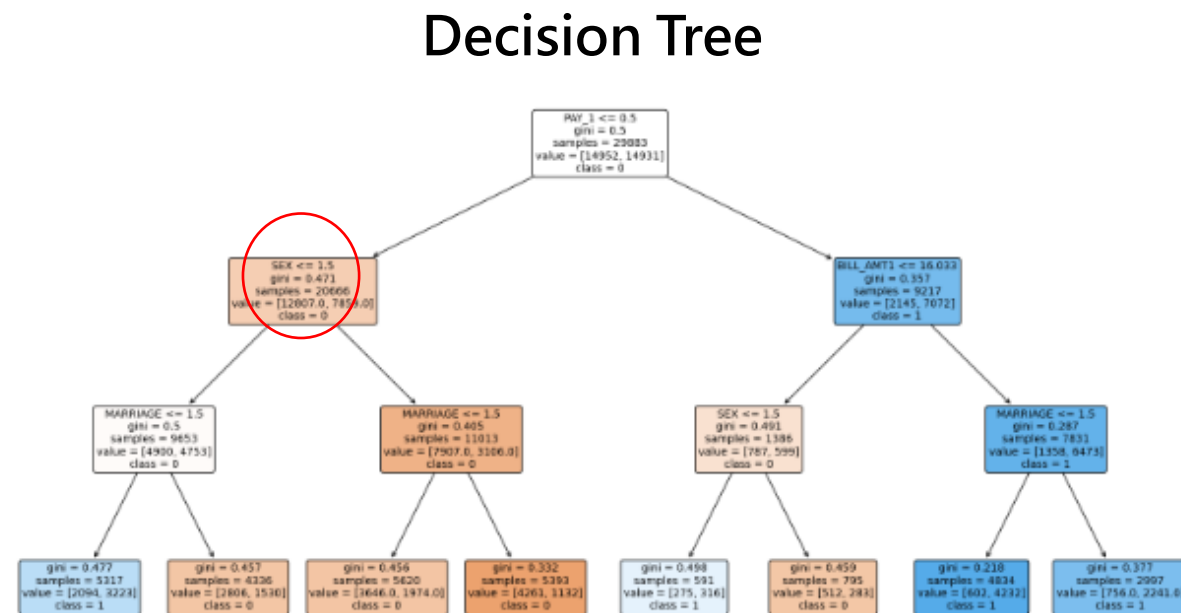
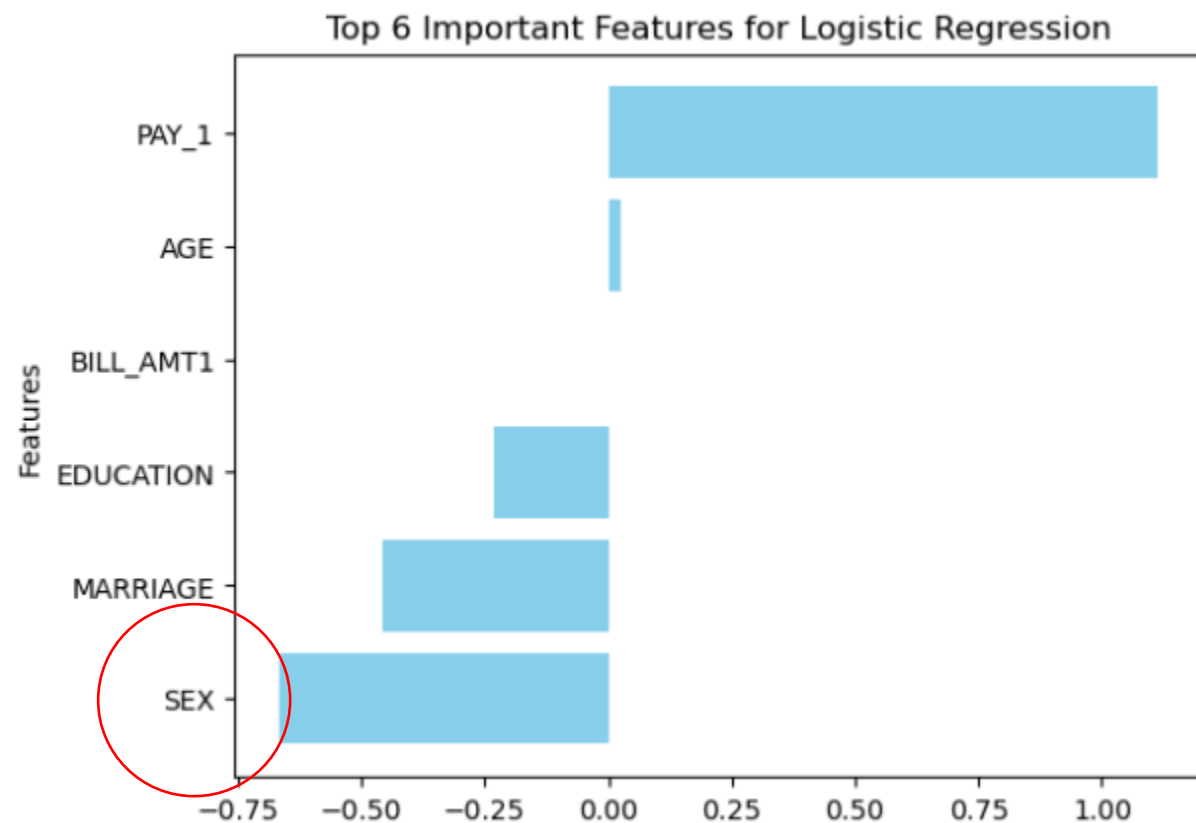
- 由於違約比率是22%，資料集呈現明顯的不平衡情況。
- 透過SMOTE生成合成樣本，使得少數類別樣本數量增加。
- 違約數：非違約數 = 1：1



# Decision Tree達到更好的Recall和Accuracy

	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.74	0.67	0.70	0.72
Decision Tree	0.74	0.73	0.73	0.73

可以看到Logistic Regression和Decision Tree都是性別影響較大



商業問題 執行策略 資料介紹 客戶輪廓 模型建立 模型結果 方案建議



# 方案建議



## 充分了解客戶輪廓



- 性別對於預測是否違約具有相當的影響力
- 針對不同的客群可以提供定制化的借還款計畫
- 透過模型可以預測全新的用戶未來可能違約與否



# Thanks -

