



「你是有信用的人嗎？」  
信用卡風險預測

# 目錄

01 / 研究目的、資料介紹

02 / 資料探索

03 / 模型預測

04 / 結論

01

# 研究目的 資料介紹

# 研究目的、資料介紹

## 研究目的

希望透過這次的分析了解銀行審核信用卡申請時，會因為那些條件的不同，進而將申請人評估為高風險、低風險。

例如：

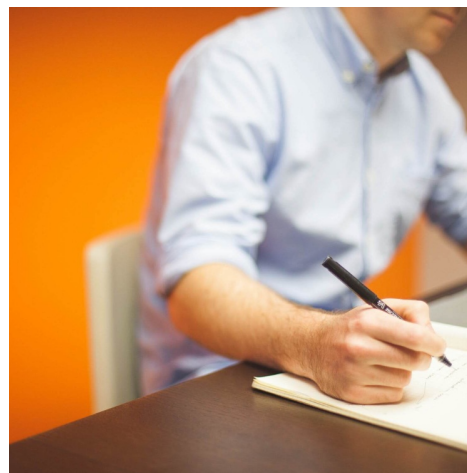
對於不同的收入類型，風險是否相同

對於不同的學歷，風險是否相同

對於不同的性別，風險是否相同

藉此進一步預測：

一位用戶申辦信用卡時，  
銀行會將其評估為高風險還是低風險？



# ● 研究目的、資料介紹

## 資料介紹 — Credit Card Classification

資料集來源 **Kaggle**

變數 **20**

樣本數目 **9709**

<b>ID</b>	客戶ID	int	<b>Account_length</b>	擁有信用卡的月數	int
<b>Gender</b>	性別	int	<b>Total_income</b>	總收入 ( 人民幣 )	num
<b>Own_car</b>	是否有車	int	<b>Age</b>	年齡	num
<b>Own_property</b>	是否有財產	int	<b>Years_employed</b>	年資	int
<b>Work_phone</b>	是否有工作手機	int	<b>Income_type</b>	收入類型	chr
<b>Phone</b>	是否有手機	chr	<b>Education_type</b>	教育程度	chr
<b>Email</b>	是否有信箱	chr	<b>Family_status</b>	婚姻狀況	chr
<b>Unemployed</b>	是否失業	chr	<b>Housing_type</b>	居住類型	chr
<b>Num_children</b>	小孩人數	Int	<b>Occupation_type</b>	工作職位	chr
<b>Num_family</b>	家庭成員人數	int	<b>Target</b>	是否為高風險	int

# ● 研究目的、資料介紹

## 資料處理

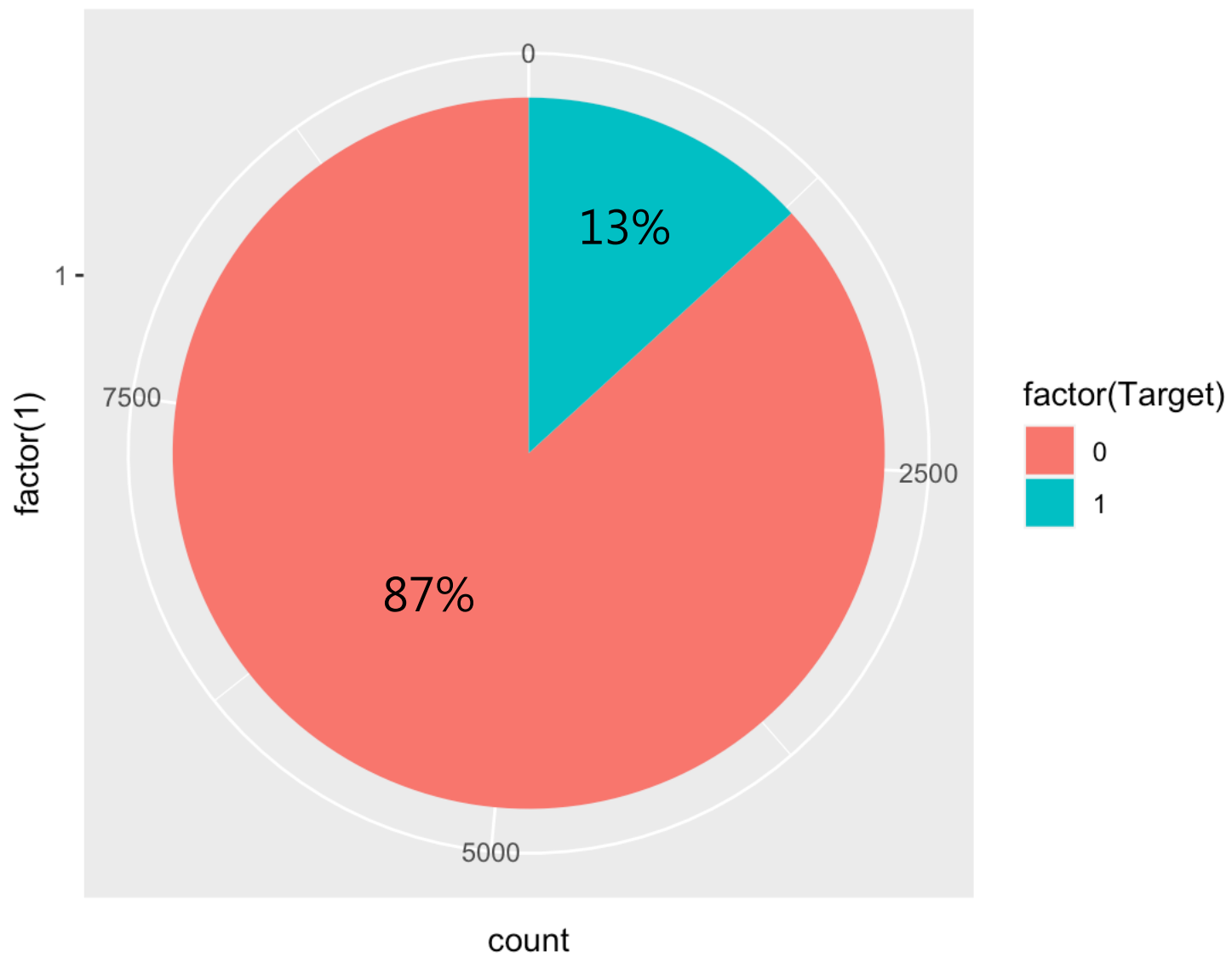
- 將部分欄位之型態轉換為factor
- 將教育程度(chr)轉換為分數(int)
- 將客戶ID刪除

ID	客戶ID	<del>int</del>	Account_length	擁有信用卡的月數	int
Gender	性別	factor	Total_income	總收入（人民幣）	num
Own_car	是否有車	factor	Age	年齡	num
Own_property	是否有財產	factor	Years_employed	年資	int
Work_phone	是否有工作手機	factor	Income_type	收入類型	factor
Phone	是否有手機	factor	Escore	教育程度	int
Email	是否有信箱	factor	Family_status	婚姻狀況	factor
Unemployed	是否失業	factor	Housing_type	居住類型	factor
Num_children	小孩人數	Int	Occupation_type	工作職位	chr
Num_family	家庭成員人數	int	Target	是否為高風險	factor

## 02 資料探索



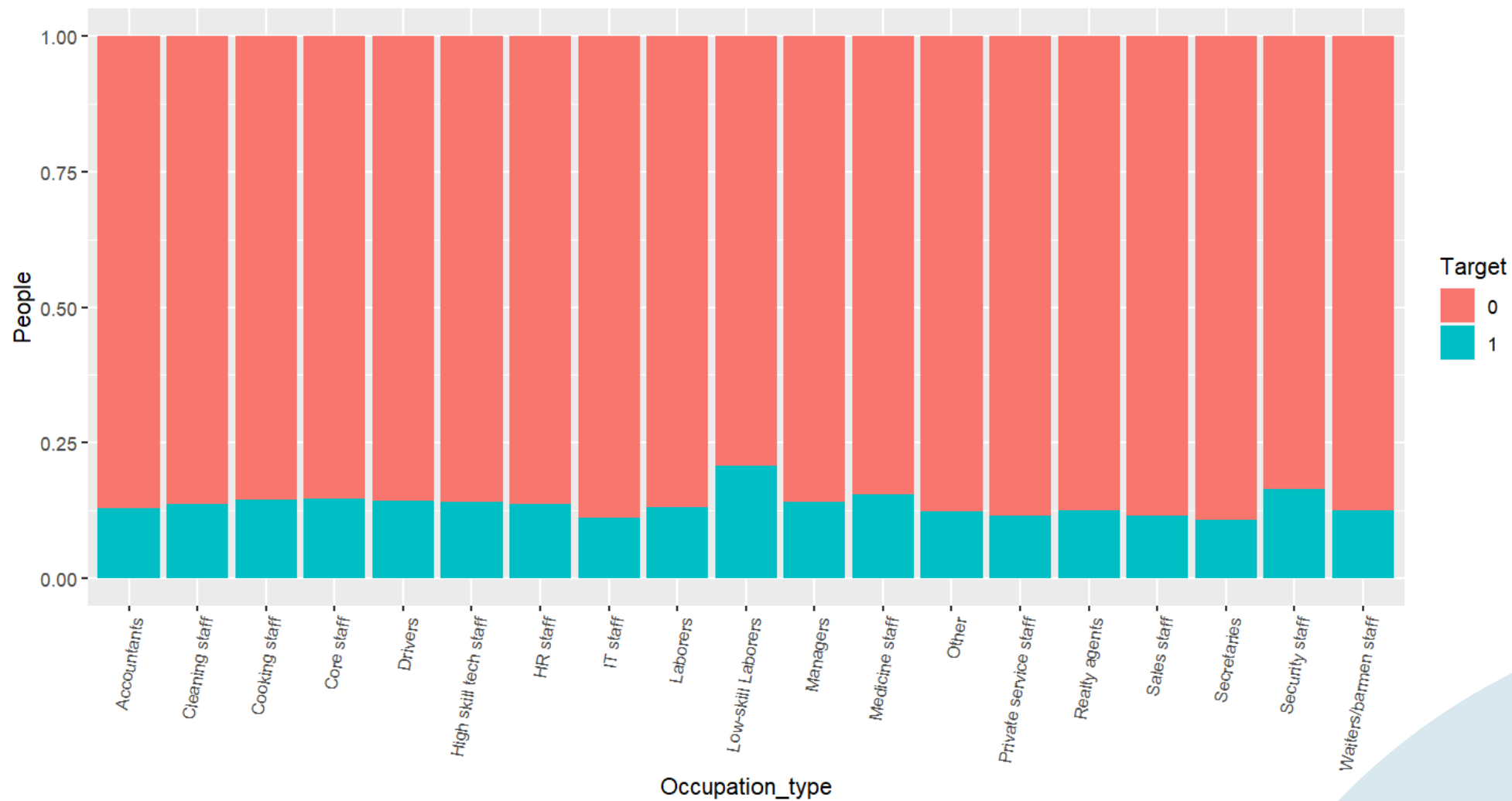
# 資料探索







# 資料探索

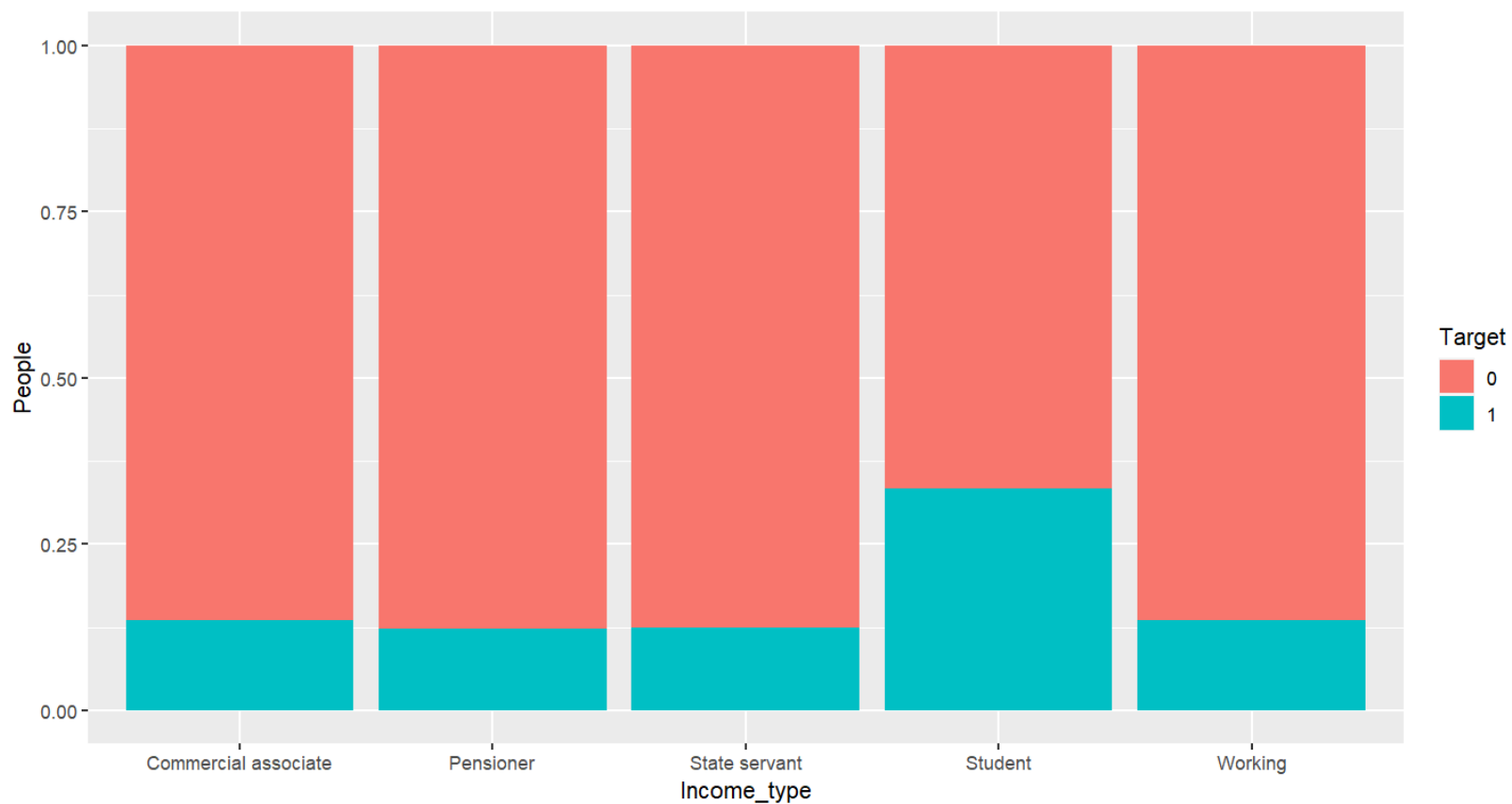


低專業技術勞工族群有較高的風險





# 資料探索

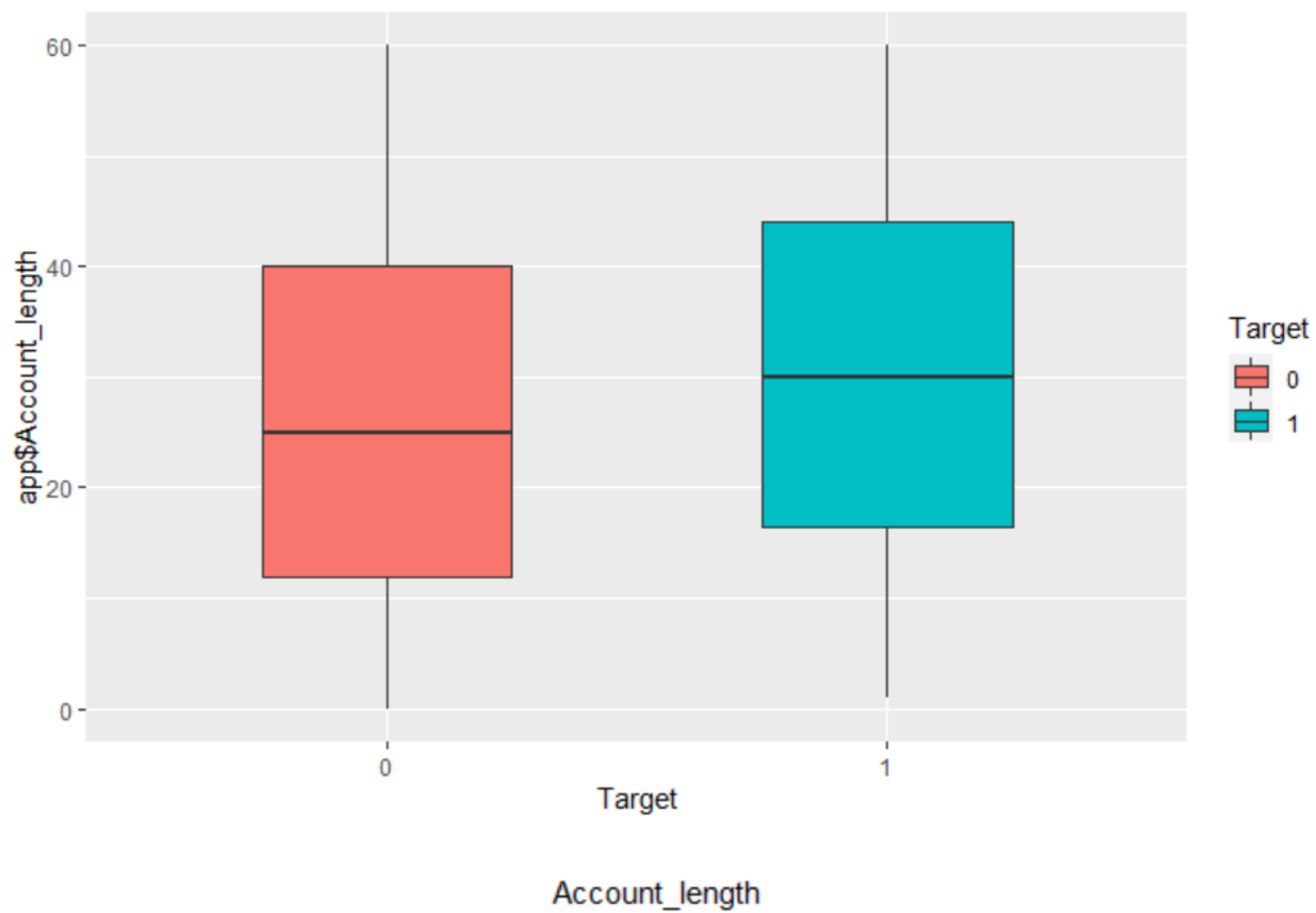


學生族群有較高的風險





## 資料探索

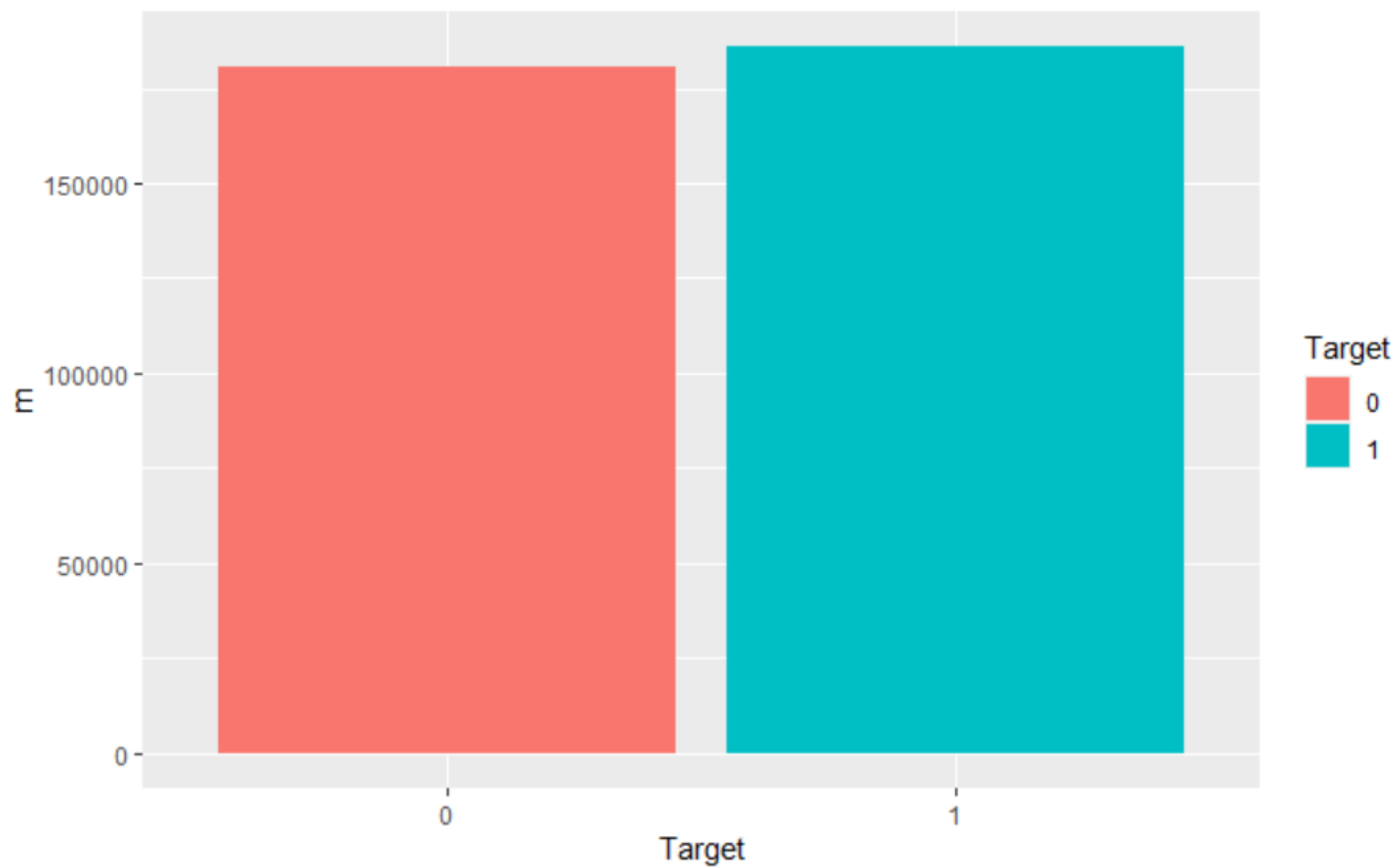


高風險族群持有信用卡月數較長





## 資料探索

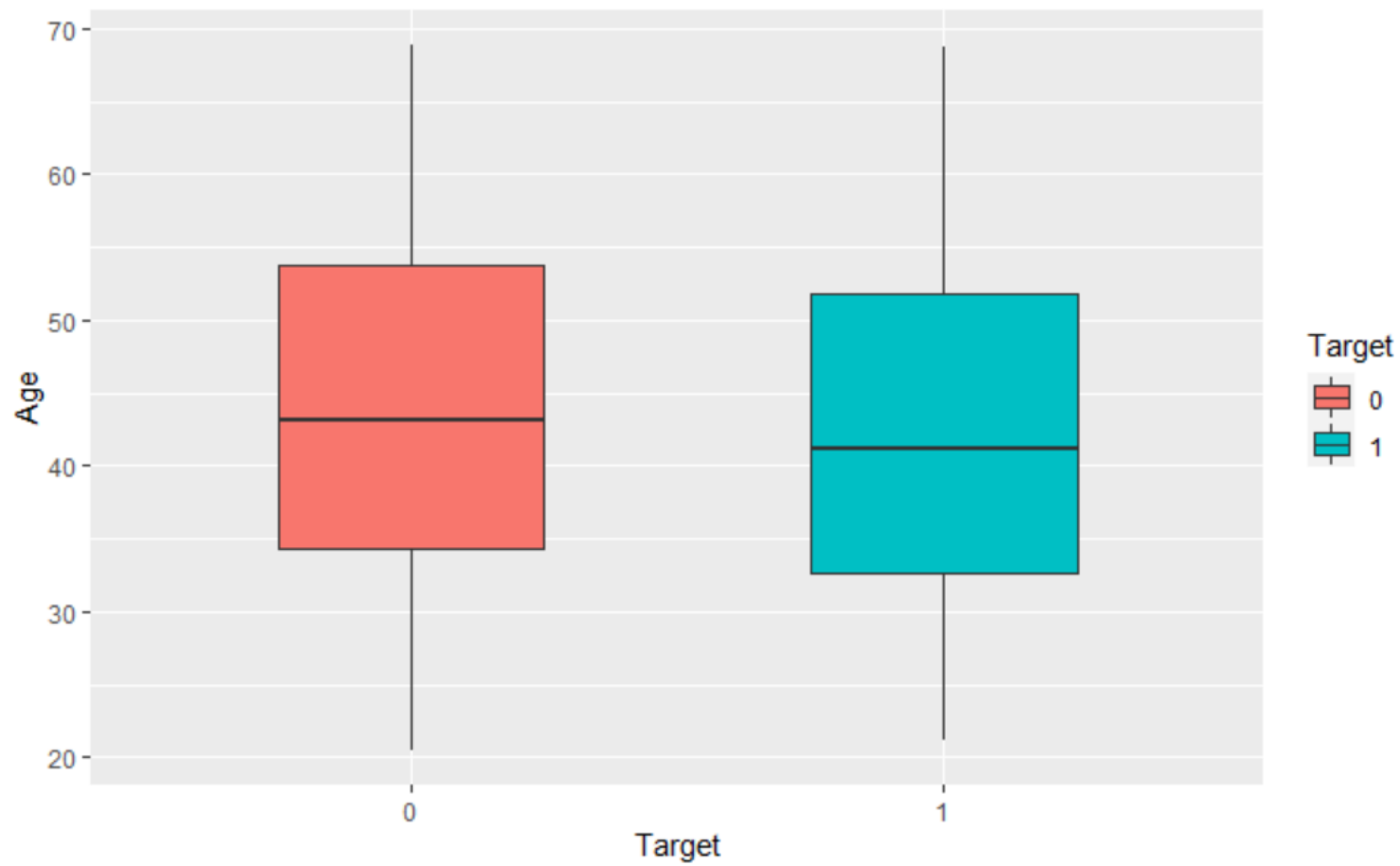


高風險族群平均收入較高





## 資料探索

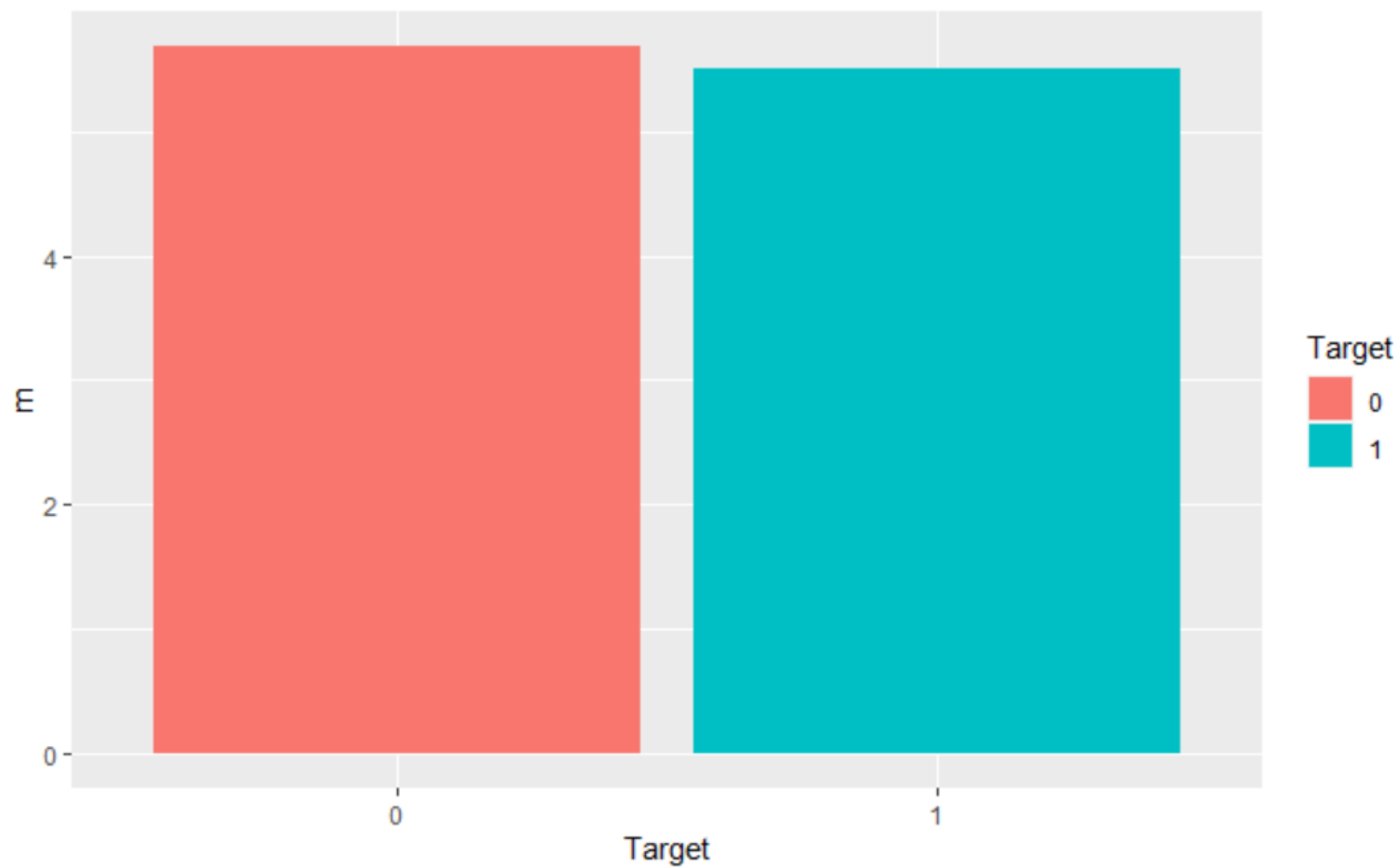


高風險族群較年輕





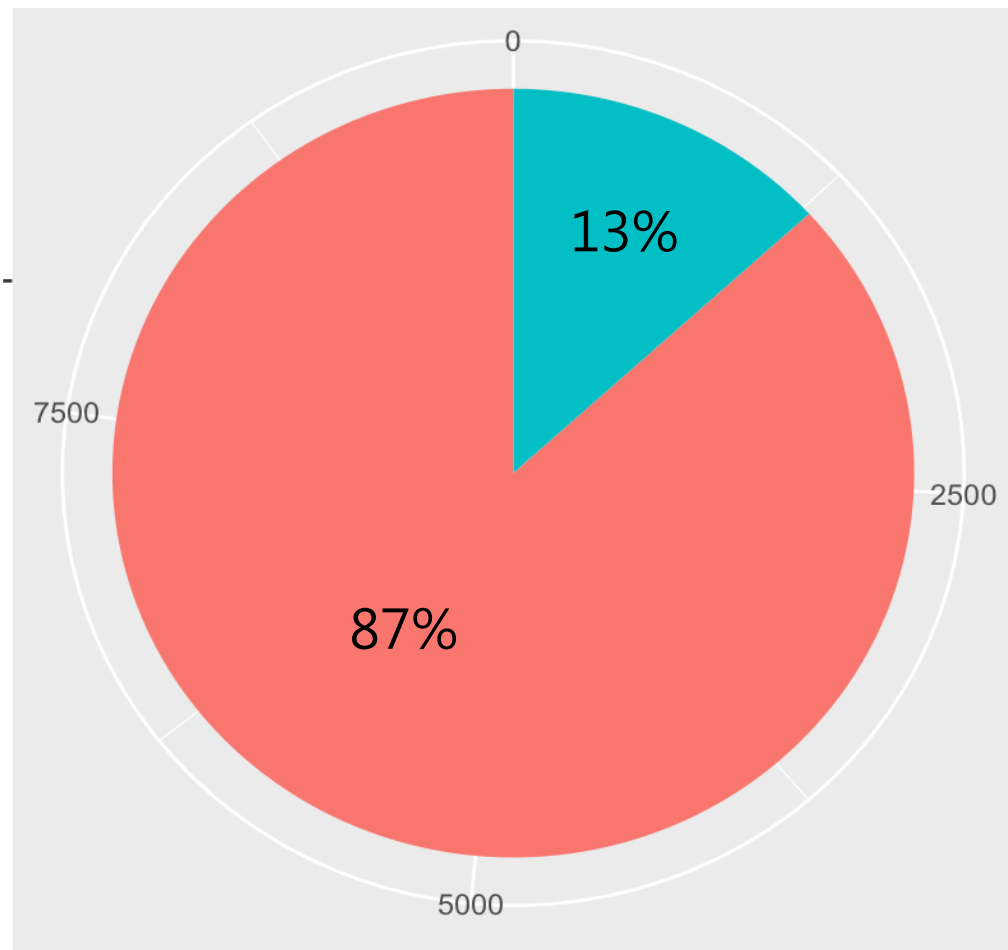
## 資料探索



高風險族群工作年資較短

# 03 模型預測

# ● 模型預測



7

	TR
總筆數	6796
0	5898
1	898

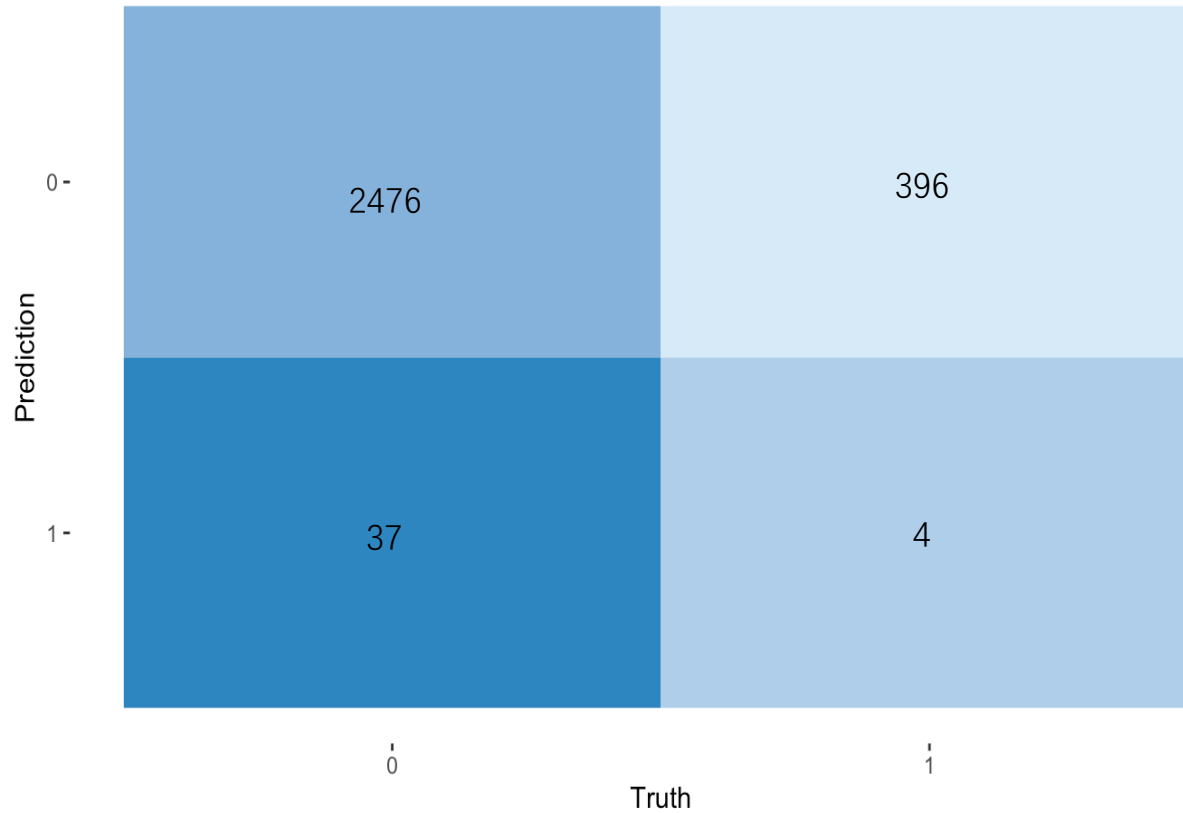
3

	TS
總筆數	2913
0	2528
1	385

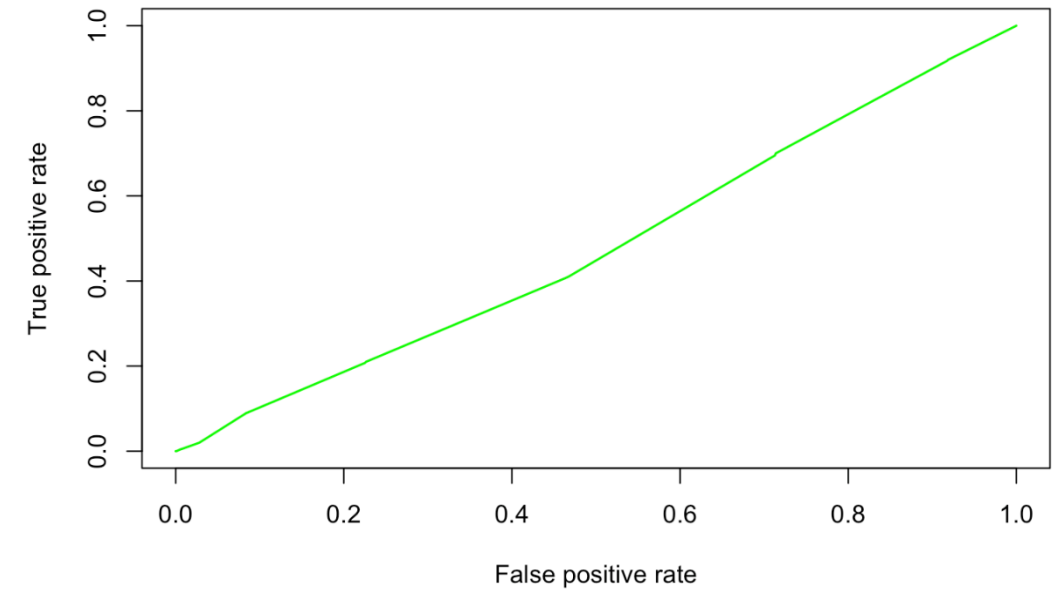




# 模型預測\_KNN



accuracy : 0.8514



AUC : 0.5638679

Recall : 0.0976

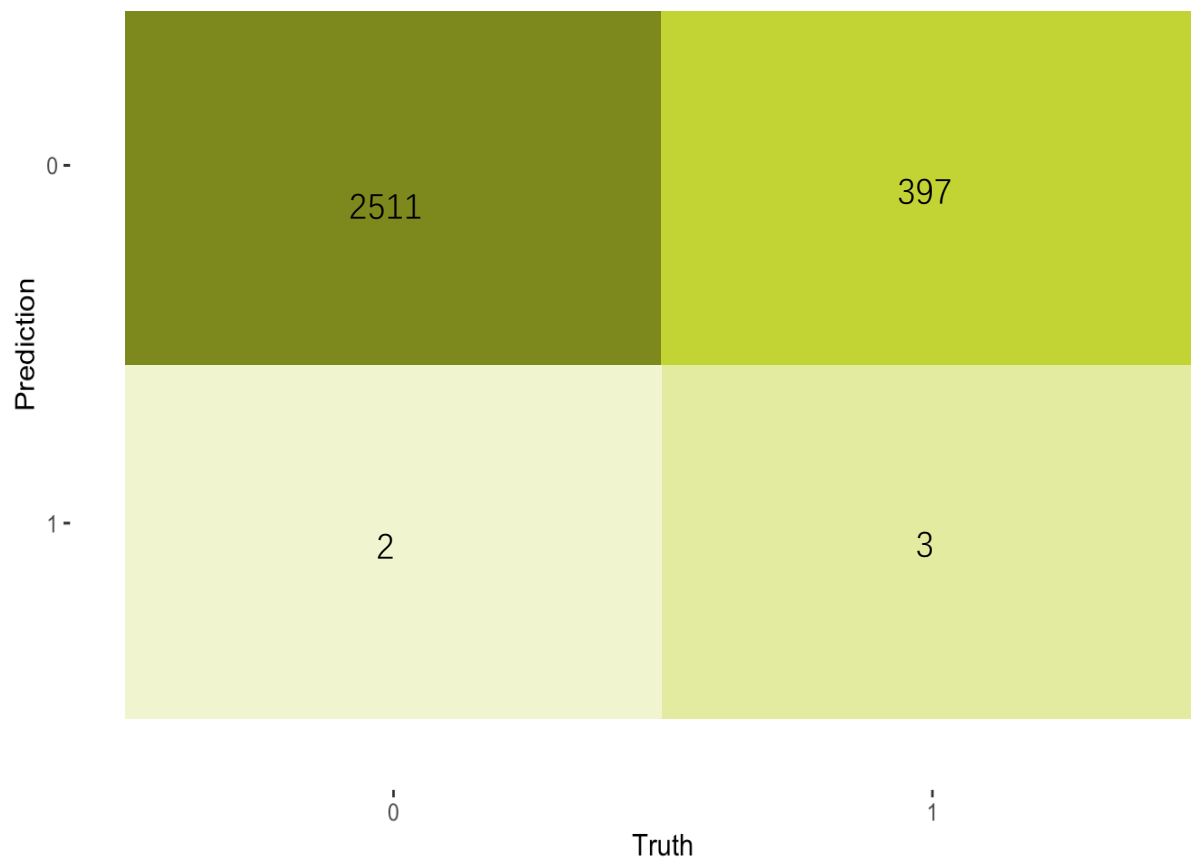
Precision : 0.01

F1-score : 0.01814

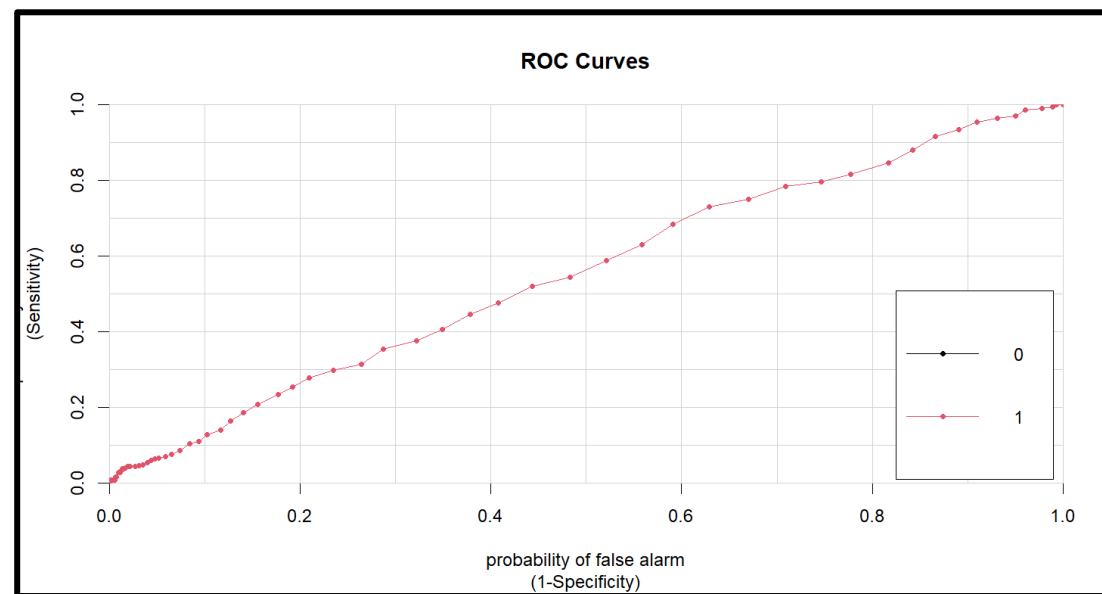




# 模型預測\_隨機森林



accuracy : 0.863



AUC : 0.5523

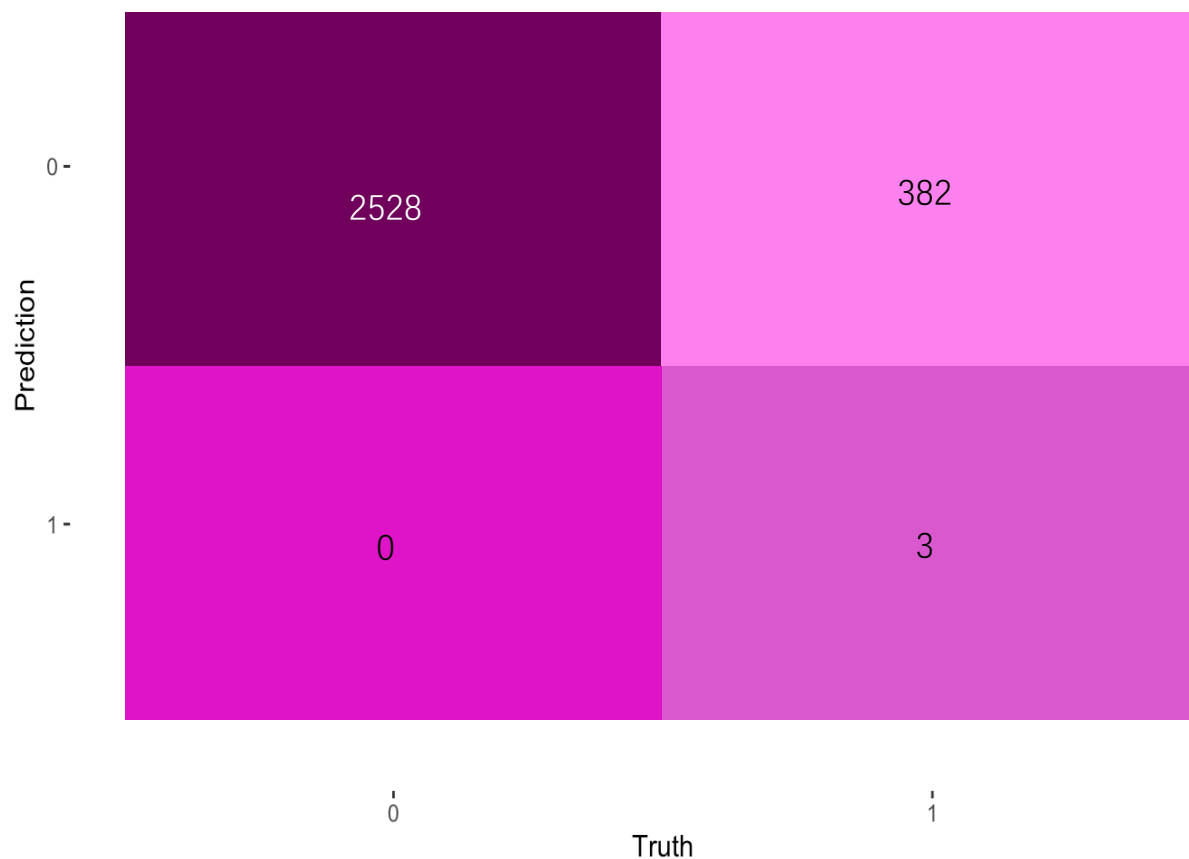
Recall : 0.6

Precision : 0.008

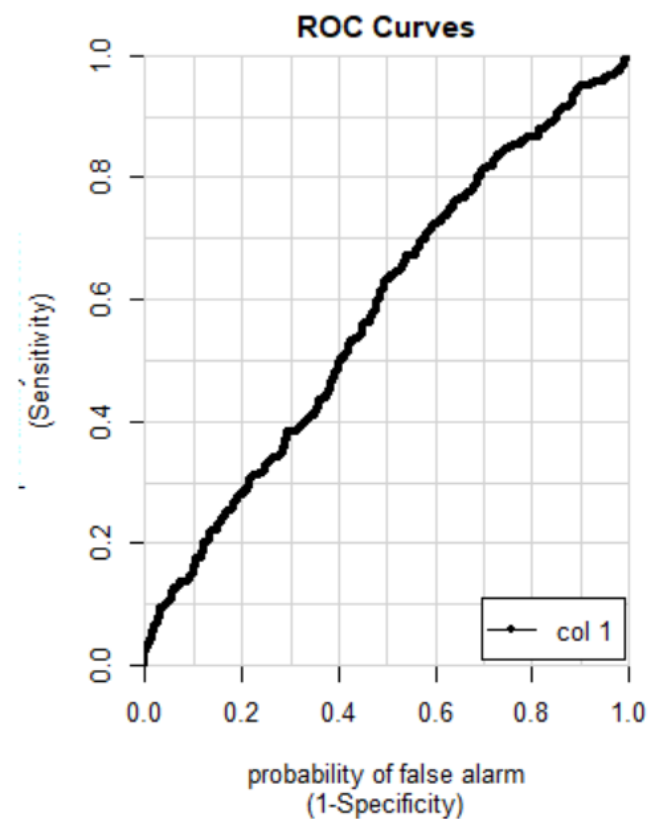
F1-score : 0.016



## 模型預測\_邏輯式回歸



accuracy : 0.869



AUC : 0.5774669

Recall : 0.008

Precision : 1.000

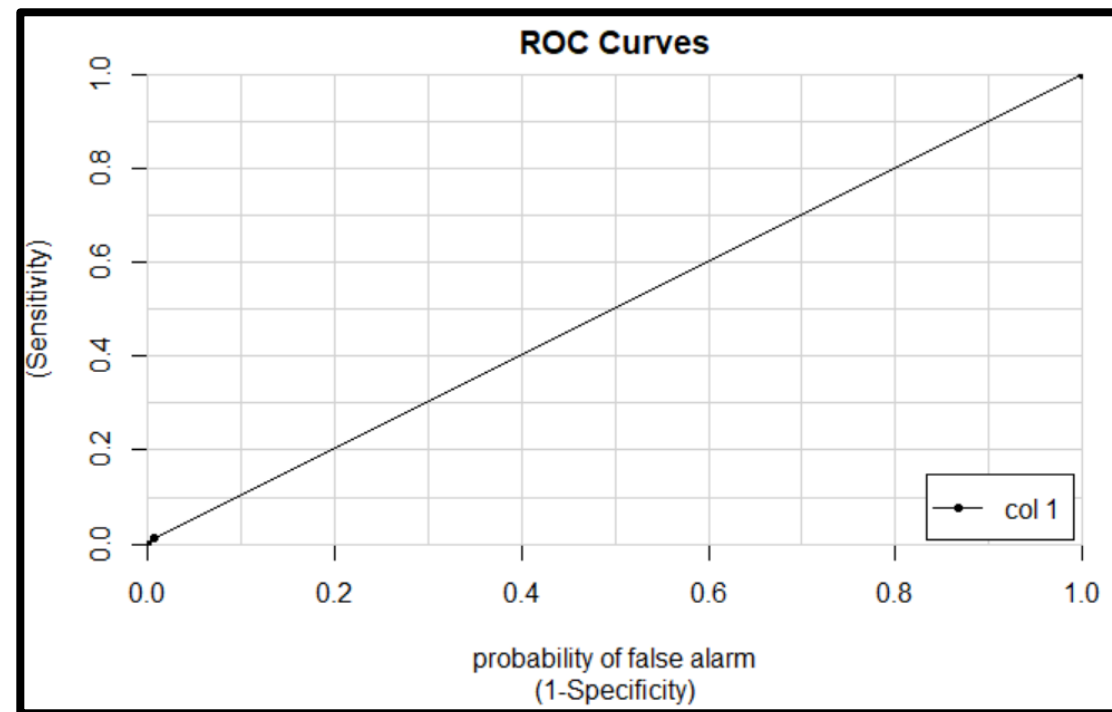
F1-score : 0.015



## ● 模型預測\_XGBoost

Acture	Predict	
	0	1
0	2511	17
1	380	5

accuracy : 0.864



AUC : 0.5031312

Recall : 0.013

Precision : 0.227

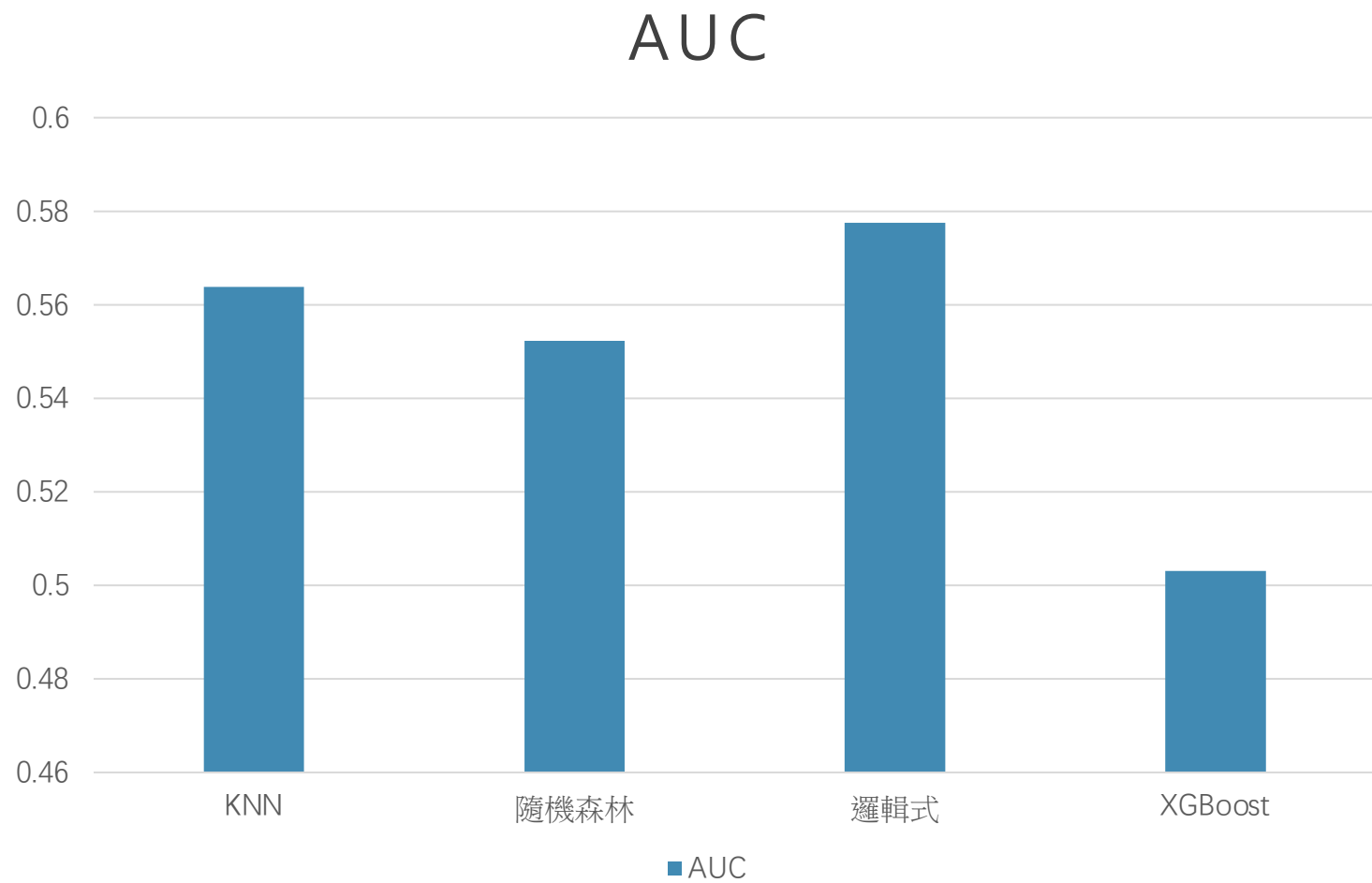
F1-score : 0.025

## ● 模型預測\_共同重要變數

KNN	隨機森林	邏輯式回歸	XGBoost
Age	Age	Num_children	Age
Years_employed	Years_employed	Num_family	Years_employed
Account_length	Account_length	Account_length	Account_length



## 模型預測\_準確度比較





04

# 結論

## 結論



邏輯式回歸在Accuracy跟AUC表現較好



年齡較低、有信用卡的月數較久、工作年資短，  
較有可能成為高風險族群



資料中的變數大多與結果沒有關係、原始變數  
無法衍生其他變數協助預測，導致模型的表現  
不好





**Thanks For Listening**