WEI-JYE GOY
倪煒傑

應徵職位：商業數據分析師

國泰金控 數位數據暨科技發展中心

第一題

議程

1. **資料集&分析目標概述**
2. **資料集EDA**
   a. 特徵資訊——申請人&貸款相關
   b. 特徵&貸款批准狀態分析
3. **特徵工程&選擇**
4. **模型預測&評估**
5. **結果**

資料集&分析目標概述

資料集概述

- 說明：此資料集為貸款審批的金融風險資料
- 來源：OpenML Loan_Status_Classification 資料集 (ID：46434)
- 樣本量：45000
- 特徵數：14個（8個數值型，6個類別型）

分類	特徵類型	特徵名稱
申請人基本資訊	數值特徵	person_age person_income person_emp_exp
申請人基本資訊	類別特徵	person_gender person_education person_home_ownership
貸款相關資訊	數值特徵	loan_amnt loan_int_rate loan_percent_income credit_score cb_person_cred_hist_length
貸款相關資訊	類別特徵	loan_intent previous_loan_defaults_on_file

分析目標概述

- 預測貸款申請是否會被核准
- 理解影響貸款審批決策的關鍵因素
- 提供可解釋的見解，提供金融業務決策支持

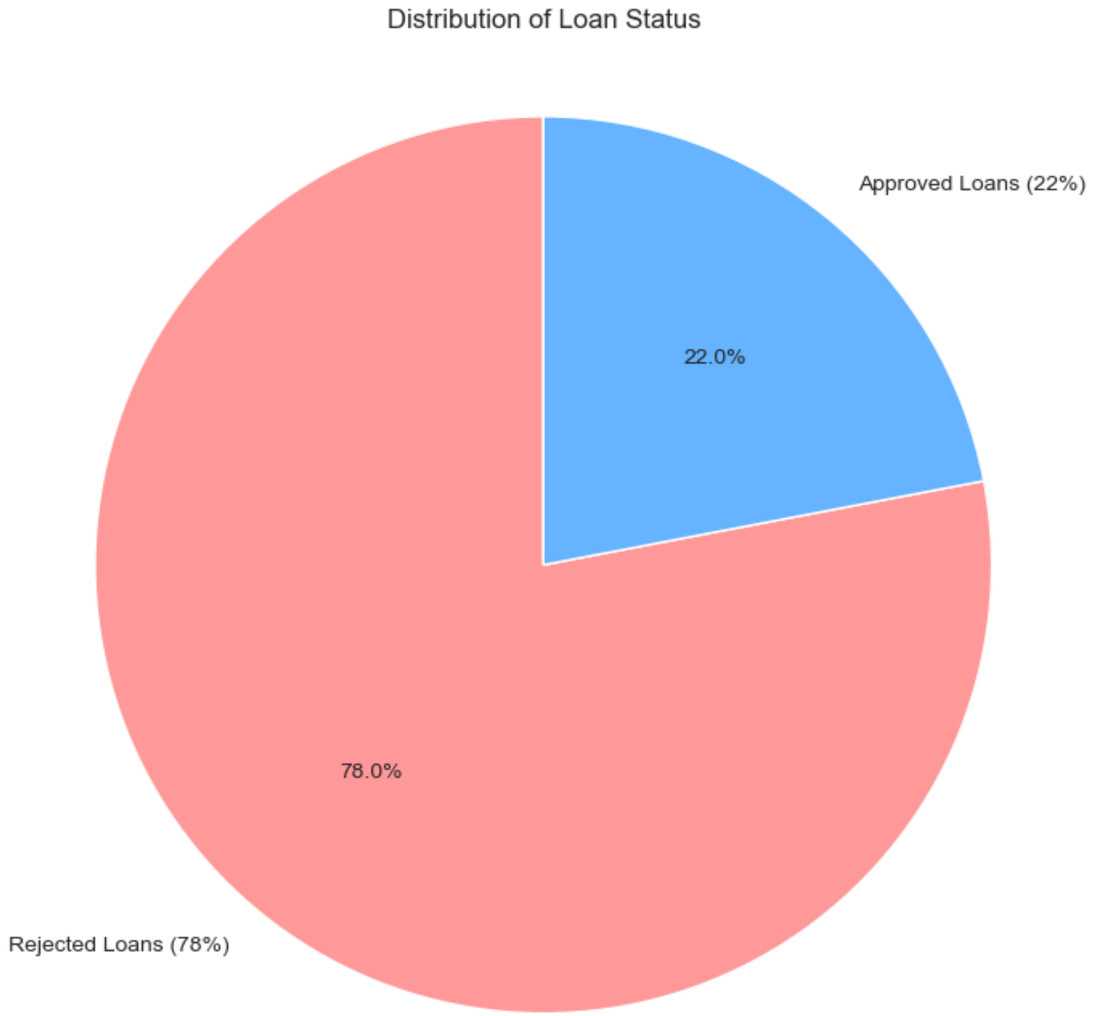
資料集EDA

特徵資訊——申請人&貸款相關

貸款批准狀態比例

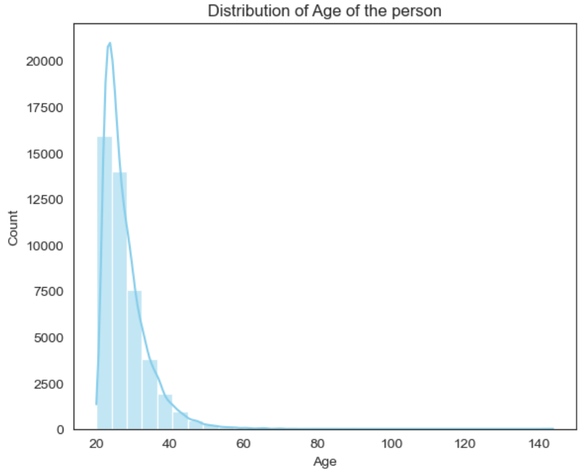
目標變量分布

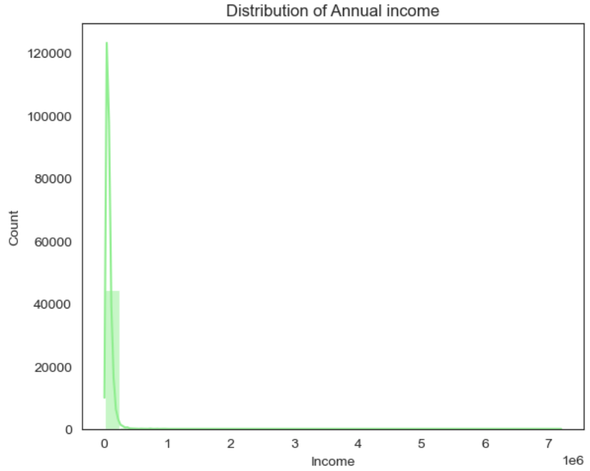
拒絕貸款約占78%，批准貸款約占22%

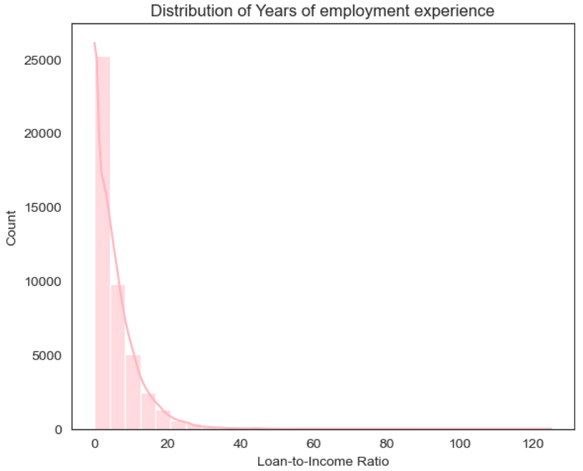


申請人基本資訊

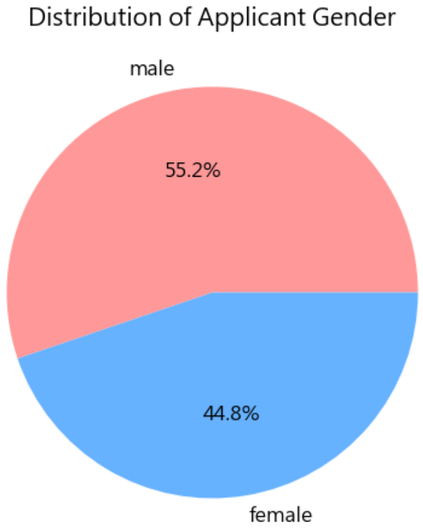
- **年齡分佈**：大多數集中在**20-40歲**，但存在極端離群值（如歲數 > 100）
- **收入分佈**：呈高度右偏，**大多數集中在低收入範圍**，部分非常高的收入為離群值。
- **工作經驗**：大多數申請人工作經驗**少於10年**，極少數極高的工作經驗值為離群點

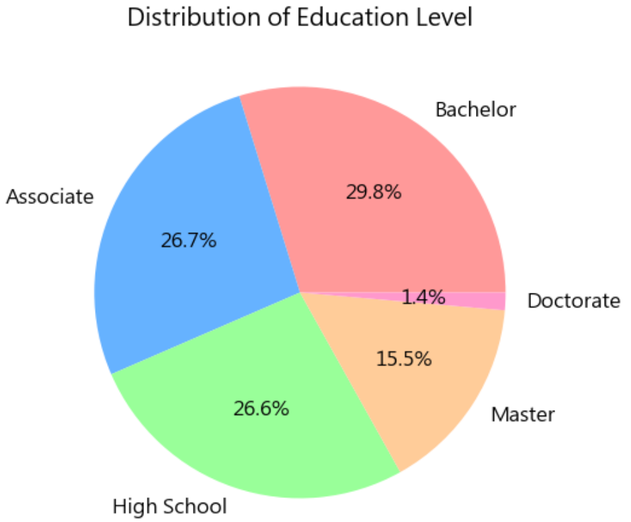


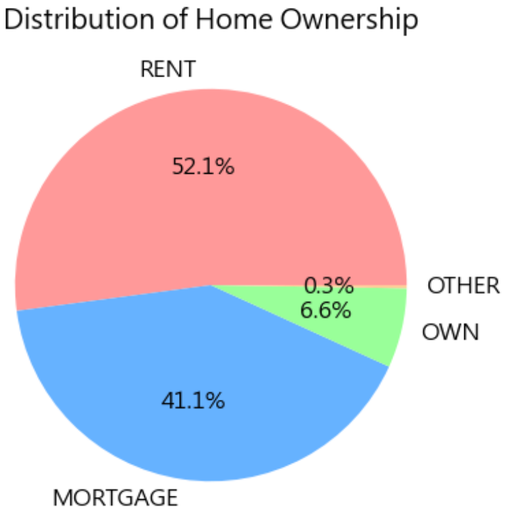




- **性別分佈**：**整體相對均衡**，但男性申請者略多
- **教育程度**：**學士學位**占比最高（29.8%），**博士學位**占比最低（1.4%）
- **住房狀況**：大多數申請人選擇**租房或自有住房**

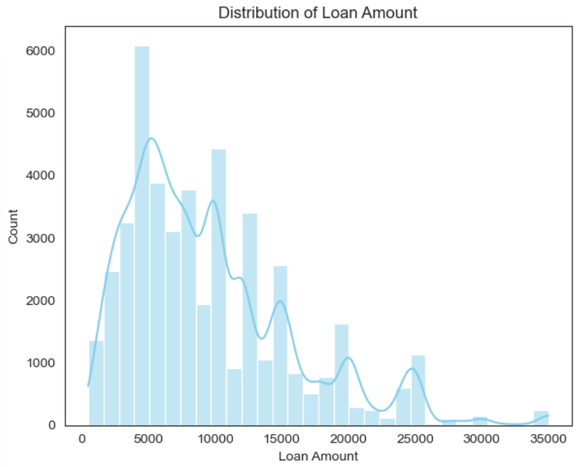


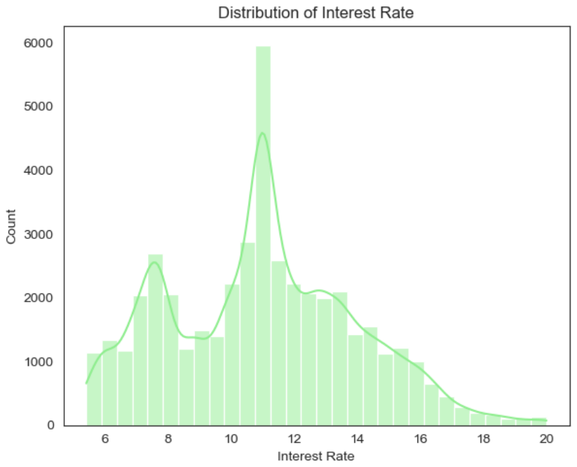


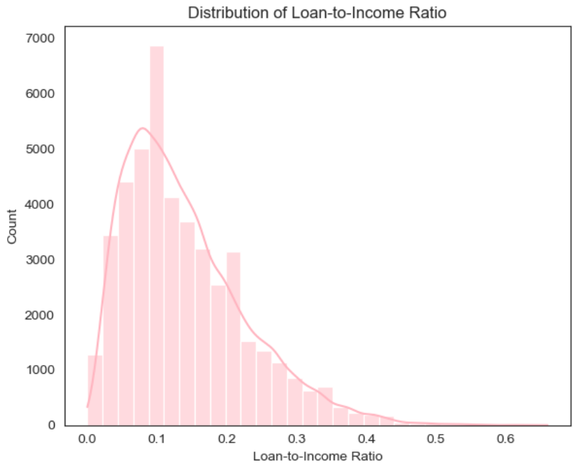


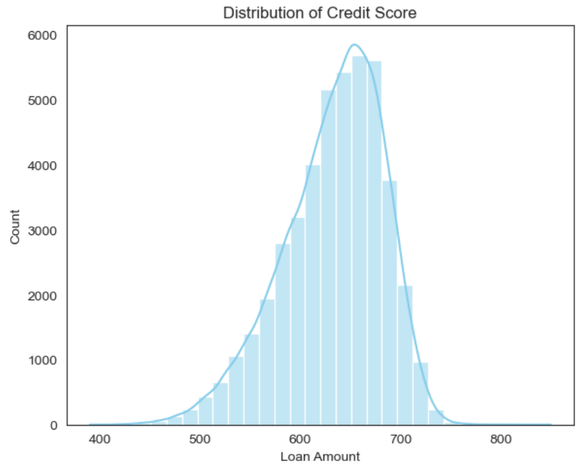
貸款相關資訊

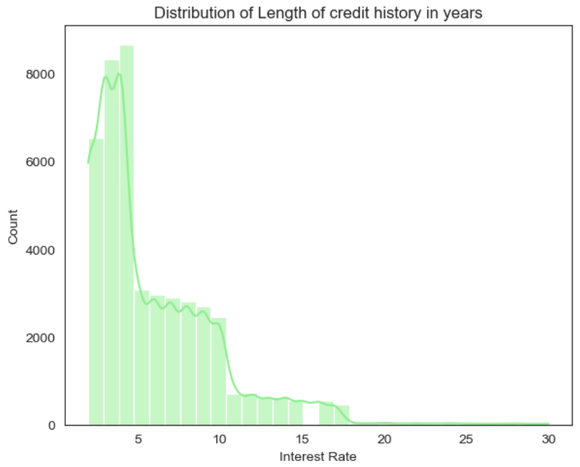
- **貸款金額**：集中在5,000至15,000之間
- **貸款利率**：主要分佈於10%至15%，**符合典型貸款利率範圍**
- **貸款佔收入比例**：大多數貸款金額僅**佔申請人收入的一小部分**，通常低於20%
- **信用歷史**：信用歷史長度**在3至5年達到峰值**，少數申請人超過10年
- **信用評分**：接近**常態分佈**，集中在600至700的中間範圍





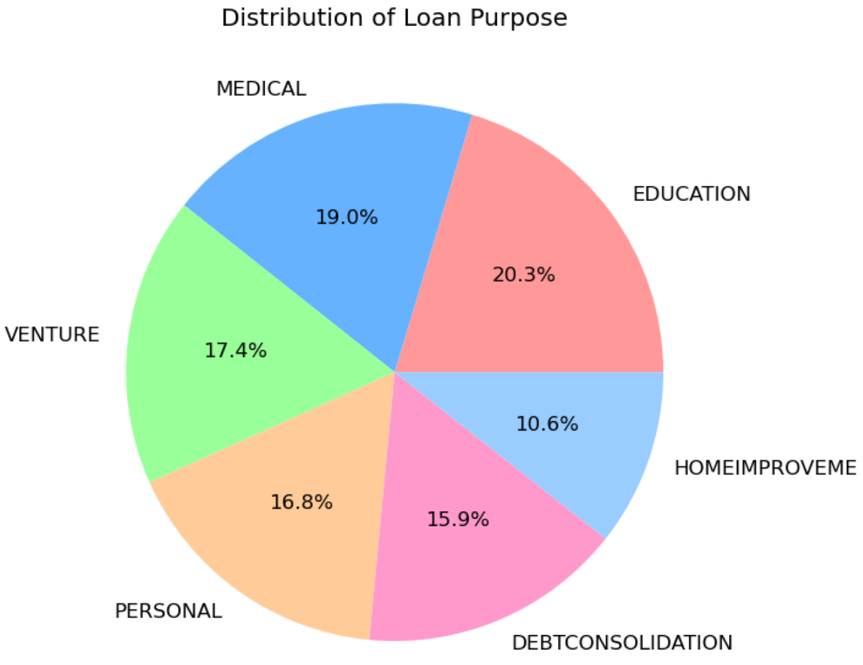


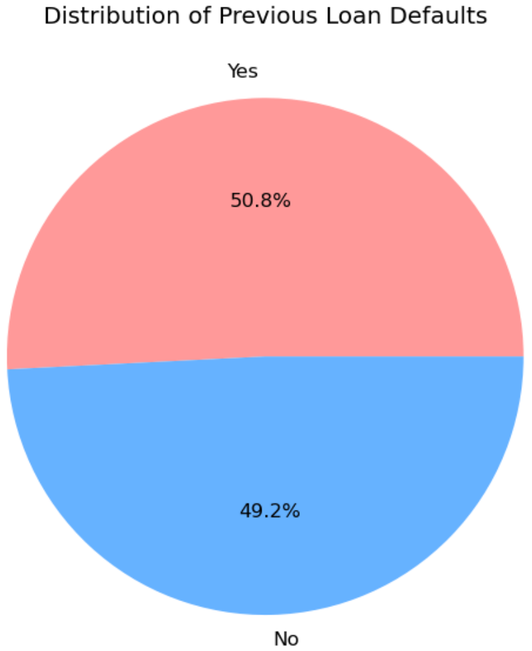




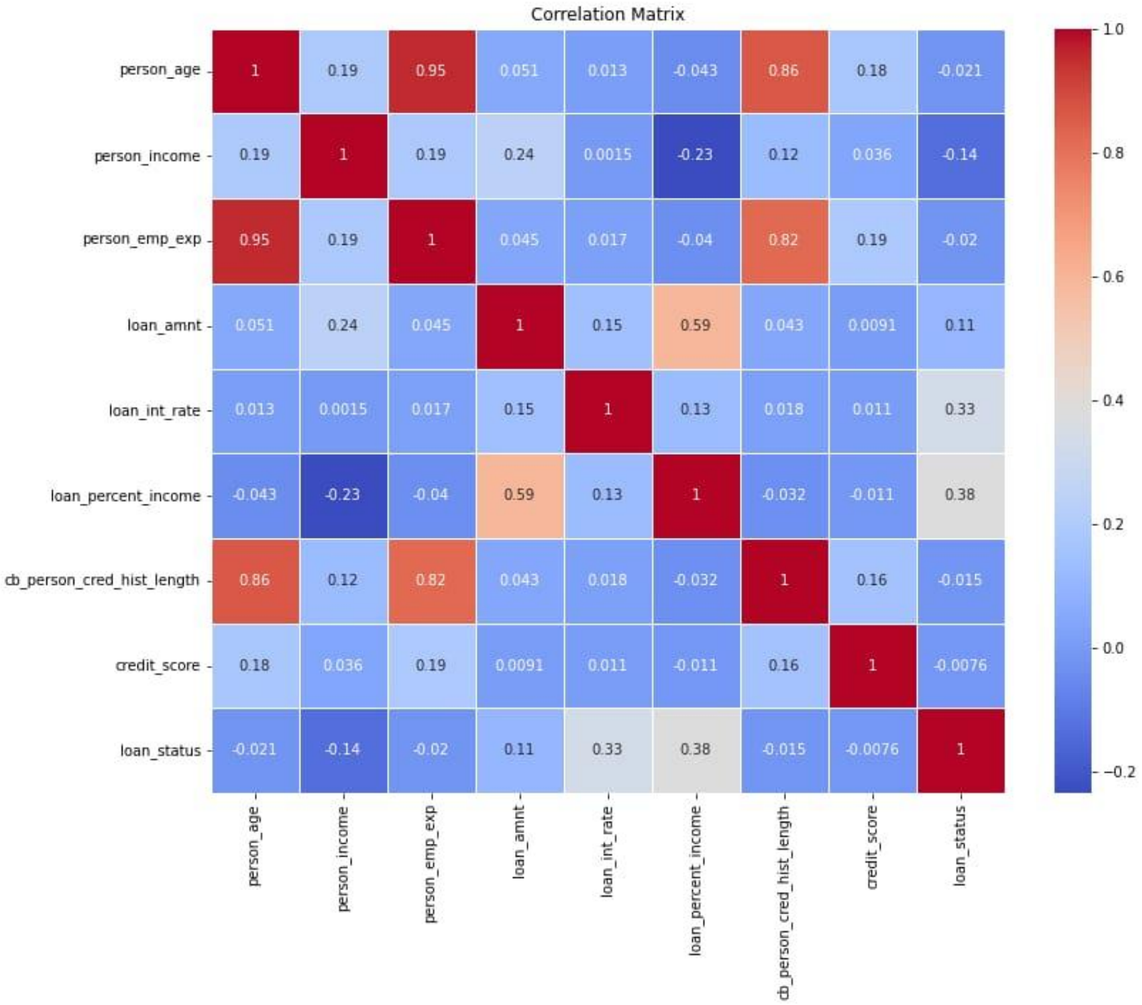
貸款相關資訊

- **貸款用途**：多樣化，包括個人消費、債務整合、醫療支出和教育等常見用途
- **貸款違約記錄**：違約記錄分布較為平均





相關係數分析

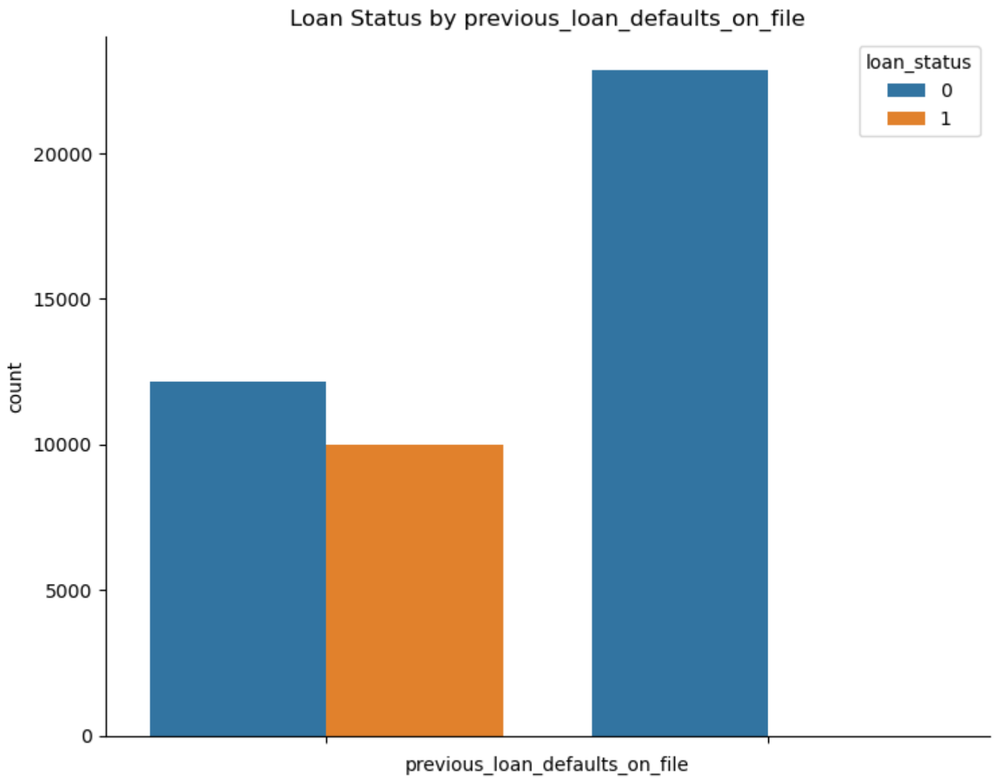


- **強相關關係**
  - 年齡與工作經驗(0.95)
  - 年齡與信用歷史長度(0.86)
- **中等相關關係**
  - 貸款金額與收入比率(0.59)
  - 貸款狀態與利率(0.33)
  - 貸款狀態與收入比率(0.38)
- **負相關關係**
  - 個人收入與貸款收入比率(-0.23)
  - 個人收入與貸款狀態(-0.14)

資料集EDA

特徵&貸款批准狀態分析

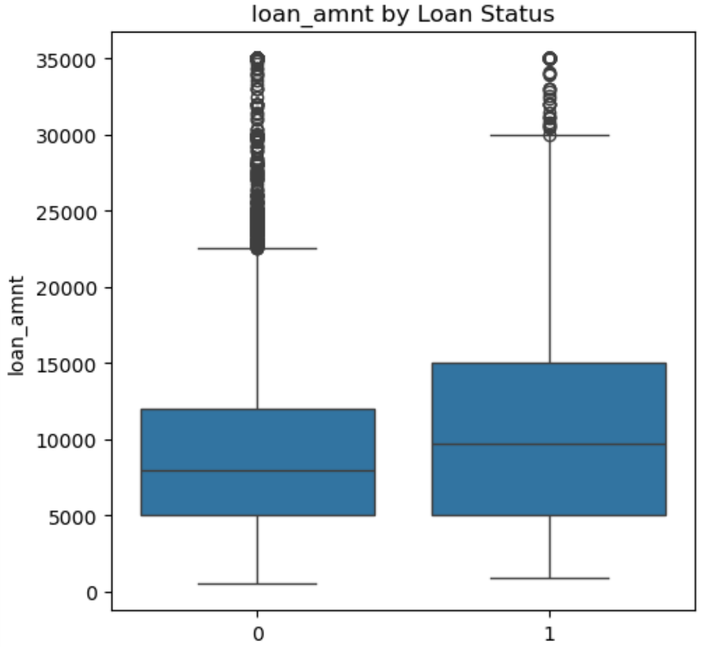
貸款違約記錄與貸款批准狀態

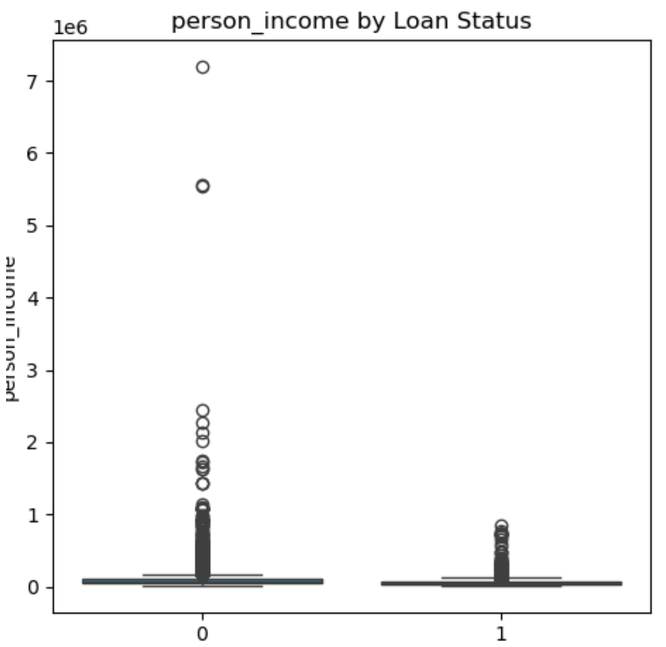


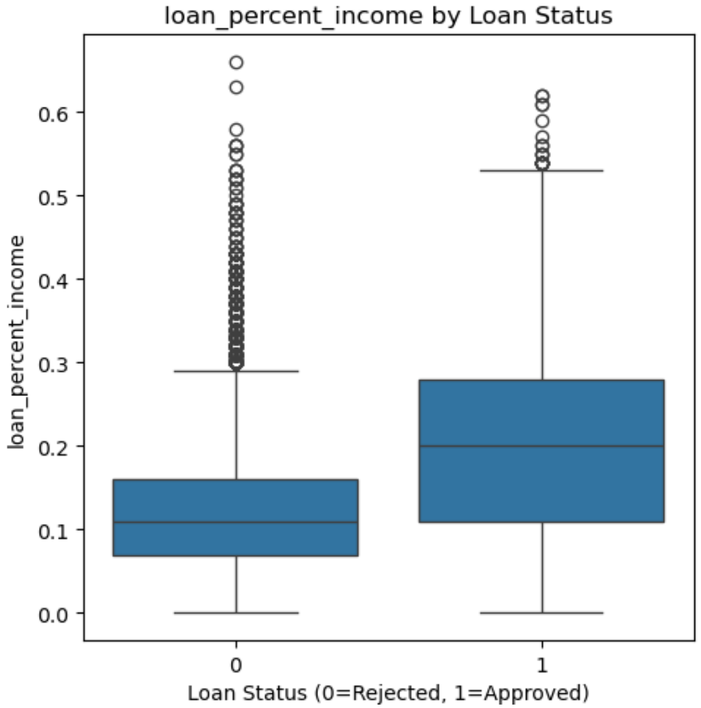
- 有過去違約記錄的申請人，拒絕率顯著高於無違約記錄者。
- 此特徵對貸款審批有強烈影響，因為過去違約表明潛在風險。

貸款&收入與貸款批准狀態

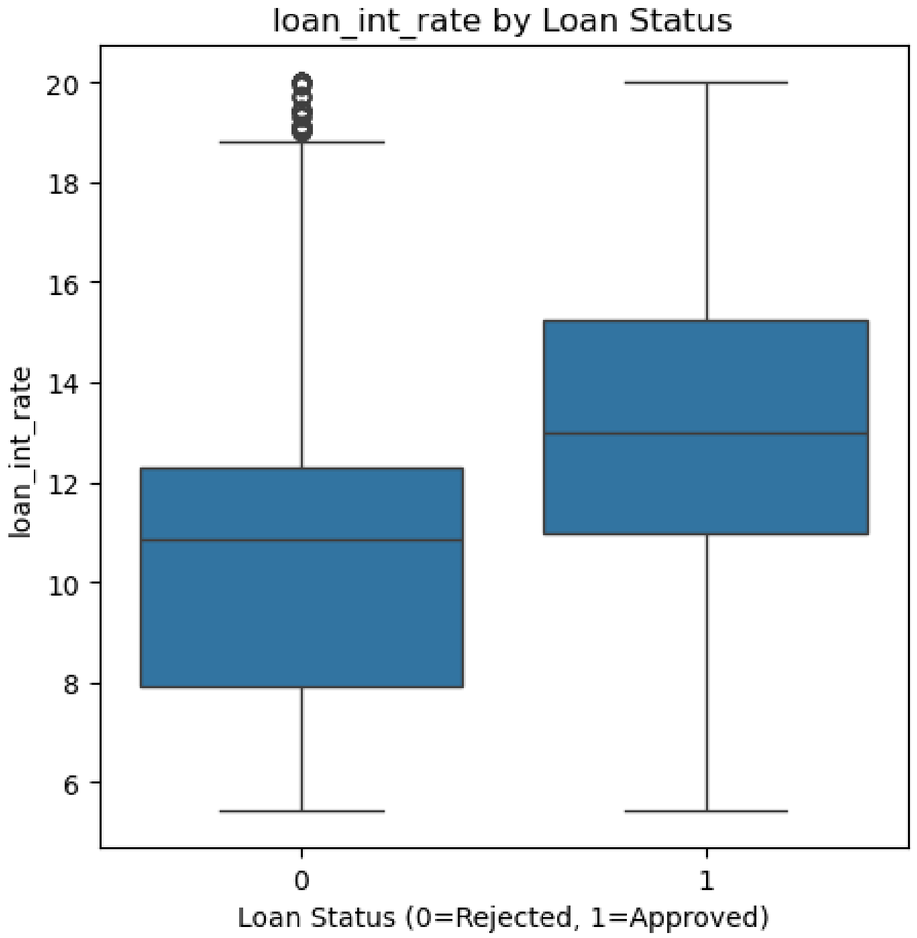
- 高收入並不一定更容易獲得貸款批准，**收入不是決定性因素**
- **獲批貸款**的申請者通有**較高的貸款收入比**，銀行願意向**客戶提供更大額度的貸款**
- **被拒絕的貸款**申請普遍具有**較低的貸款收入比**，銀行傾向於**拒絕過低貸款金額的申請**
- 兩組數據都存在異常值，表明有些申請者要求的貸款金額遠超過其收入水平，無論是否獲批都可能存在風險。







貸款利率與貸款批准狀態



- 獲批貸款的較高利率分布表明，銀行可能採用**風險定價策略**，對風險較高但仍可接受的客戶收取更高利率，而不是直接拒絕。
- 這反映了金融機構在風險管理和收益之間尋求平衡的策略。

特徵工程&選擇

特徵工程

Binary Encoding

性別、貸款違約記錄

Ordinal Encoding

教育程度

One-Hot Encoding

房屋擁有狀況&貸款用途

這些方法確保特徵數據轉換為模型能有效利用的數值格式，
同時保留原有的結構，有助於提升預測準確性和模型表現。

特徵選擇

使用F檢驗特徵選擇(SelectKBest)進行特徵評估

計算每個特徵與目標變量之間的相關性

選擇F-score最高的前15個特徵

- 降低模型複雜度
- 減少過擬合風險
- 提高模型效能
- 增加模型可解釋性

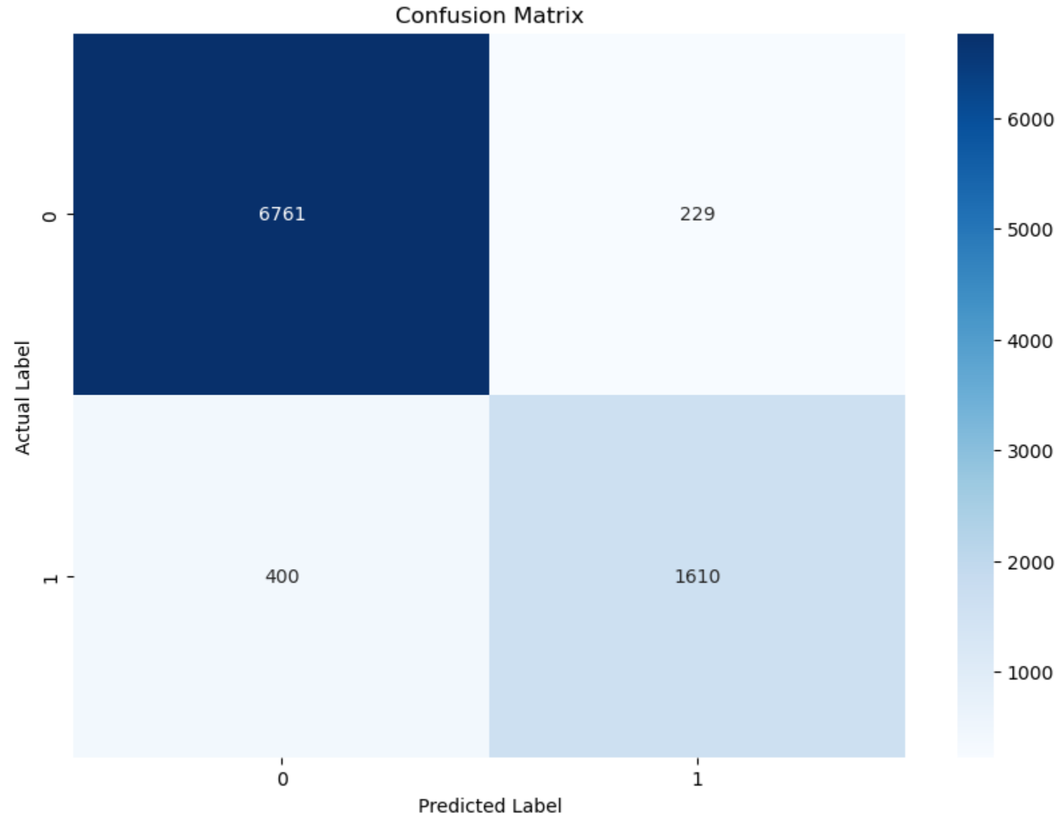
特徵選擇

特徵	重要性分數
previous_loan_defaults_on_file_No	18824.73
previous_loan_defaults_on_file_Yes	18824.73
loan_percent_income	7824.79
loan_int_rate	5574.45
person_home_ownership_RENT	3135.77
person_home_ownership_MORTGAGE	2148.03
person_income	845.53
loan_amnt	528.21
person_home_ownership_OWN	398.28
loan_intent_VENTURE	335.22
loan_intent_DEBTCONSOLIDATION	320.76
loan_intent_MEDICAL	192.08
loan_intent_EDUCATION	185.10
loan_intent_HOMEIMPROVEMENT	51.58
loan_intent_PERSONAL	22.77

模型預測&評估

模型預測&評估

- 模型使用：Random Forest
- **n_estimators = 100，random_state = 42**



	precision	recall	f1-score	support
0	0.94	0.97	0.96	6990
1	0.88	0.80	0.84	2010
accuracy			0.93	9000
macro avg	0.91	0.88	0.90	9000
weighted avg	0.93	0.93	0.93	9000

模型預測&評估

優點

- 模型在**預測拒絕貸款方面**表現優異
- 整體**準確率達到93%**，顯示模型具有良好的預測能力
- 誤報率較低，有助於**降低銀行風險**

不足

- 對批准貸款的預測能力**略低於拒絕貸款**
- 存在一定的假陰性(400例)，可能導致**損失潛在的優質客戶**

拒絕貸款預測：

- Precision：94%
- Recall：97%
- F1分數：0.96

批准貸款預測：

- Precision ： 88%
- Recall ： 80%
- F1分數：0.84

結果討論

結果討論

- 作為自動化審批的初步篩選工具
- 對於模型預測為**拒絕的案例**，可以較為**確信其風險較高**
- 對於模型預測為**批准的案例**，建議進行**人工複核**

第二題

GenAI 技術之理解

- GenAI 為一個充滿海量數據訓練後的**大腦**，其理解和生成的邏輯**與人類其實是一樣**的。
  如：人類透過**閱讀和觀察來學習**，通過**消化**來獲得**知識與技能**

- GenAI 應用之廣泛 ，能夠一定程度上**輔助**人類的生產力，GenAI為人類之無所不能的助手，然而GenAI模型的內部運作往往如同**黑箱**，難以解釋其決策過程。對需要高透明度和可信度的應用場景是一個重大挑戰。

因此要在金融業結合GenAI之應用，我們必須考量倫理和法律的挑戰，
1. GenAI訓練過程可能包含**敏感資料和個資**，金融業必須考量如何**保護用戶隱私的同時有效利用**
2. GenAI能夠創造出高度逼真的內容，可能被用於**製造或者傳播假資訊**，這部分包含到**責任歸屬**以及對**社會倫理的挑戰**

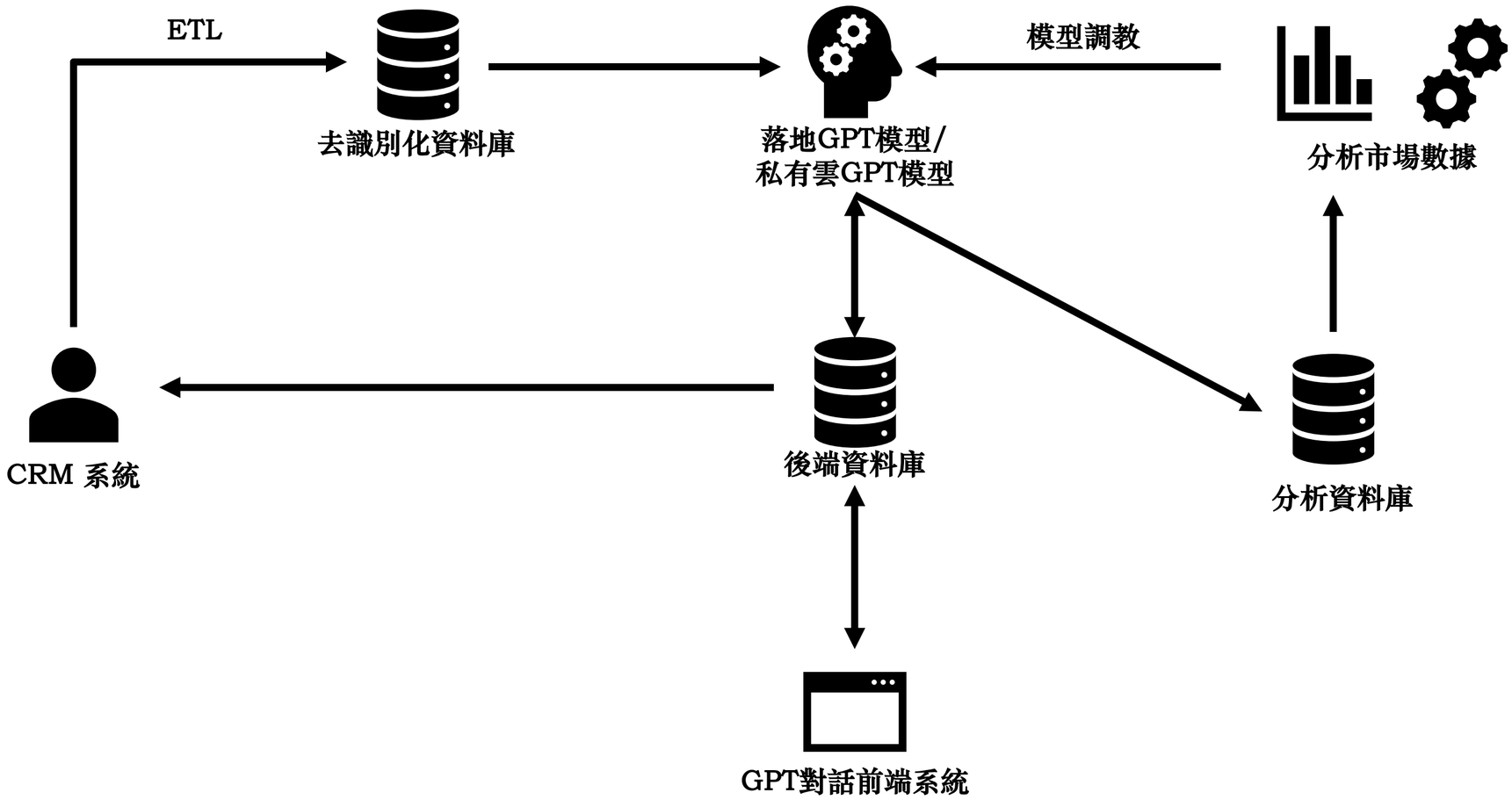
GenAI x 金融應用—個性化金融顧問服務

- **技術應用**：客戶財務數據結合GPT模型

- **實施方法**：
  - **整合與去識別化客戶數據**，將數據進行個性化分析，透過GPT模型**生成金融建議**
  - 開發基於GPT的對話系統，串接**模型產出之建議**，讓客戶能夠透過對話方式匹配適合產品
  - 將模型結果導入CRM系統，使得**理專能夠瞭解客戶匹配之商品**

- **效益**：
  - **降低人力成本**，提高客戶服務效率
  - 提供24/7全天的**個性化金融顧問服務**

GenAI x 金融應用—個性化金融顧問服務



GenAI x 金融應用—智能風險評估與貸款審批系統

- **技術應用**：結合金融數據、地理空間數據和開放數據源

- **實施方法**：
  - 使用GAN模型**生成虛擬的地理環境數據**，豐富訓練數據集
  - GIS系統提供精確的地理空間數據，包括**地形、交通、人口密度**等
  - 訓練Transformer模型，**學習地理位置特徵與貸款風險之間的關係**
  - 將GIS系統與GenAI模型整合，**實現實時風險評估**
  - 該系統可以更準確地評估不動產價值和潛在風險，提高貸款審批的效率和準確性

- **效益**：
  - 準確地**評估不動產價值和潛在風險**
  - 提高貸款審批的**效率和準確性**