



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Final Project

Báo Cáo Phân Tích Dữ Liệu Và Dự Đoán Rủi Ro Của TIMA

NGUYỄN THÀNH LINH
LỚP: BK-DA-PTHB241007

ONE LOVE. ONE FUTURE.

MỤC TIÊU CỦA PROJECT

Mục Tiêu

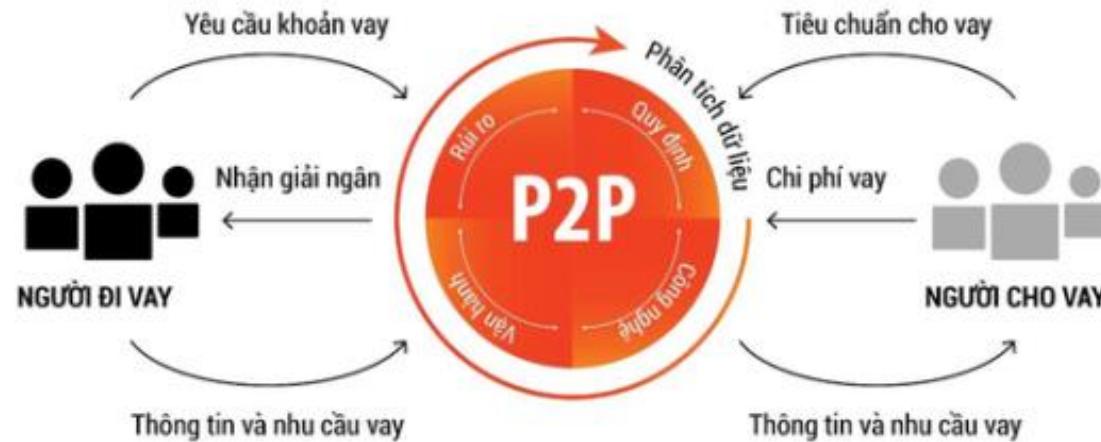
1. Nắm được tổng quan về tình hình hoạt động cho vay P2P của TIMA.
2. Đánh giá các yếu tố ảnh hưởng đến số tiền vay của khách hàng.
3. Đánh giá các yếu tố ảnh hưởng đến tỷ lệ trễ nợ của khách hàng.
4. Đưa ra các đề xuất nhằm quản lý và kiểm soát rủi ro tín dụng.

GIỚI THIỆU VỀ DOANH NGHIỆP VÀ DATASET

1.Giới thiệu về Tima



Mô hình hoạt động



KẾT NỐI BẰNG CÔNG NGHỆ



TƯ VẤN – THU HỒ SƠ



CHÁM ĐIỂM TÍN NHIỆM



THẨM ĐỊNH



KÝ HỢP ĐỒNG

1.Giới thiệu về Tima

Tổng quan về dịch vụ tài chính của Tima

Sản phẩm

Bao gồm cho vay
tín chấp và thế
chấp

Dịch vụ cho vay

Cung cấp các lựa
chọn cho vay khác
nhau

Khách hàng

Phục vụ hàng
triệu khách hàng
trên toàn quốc



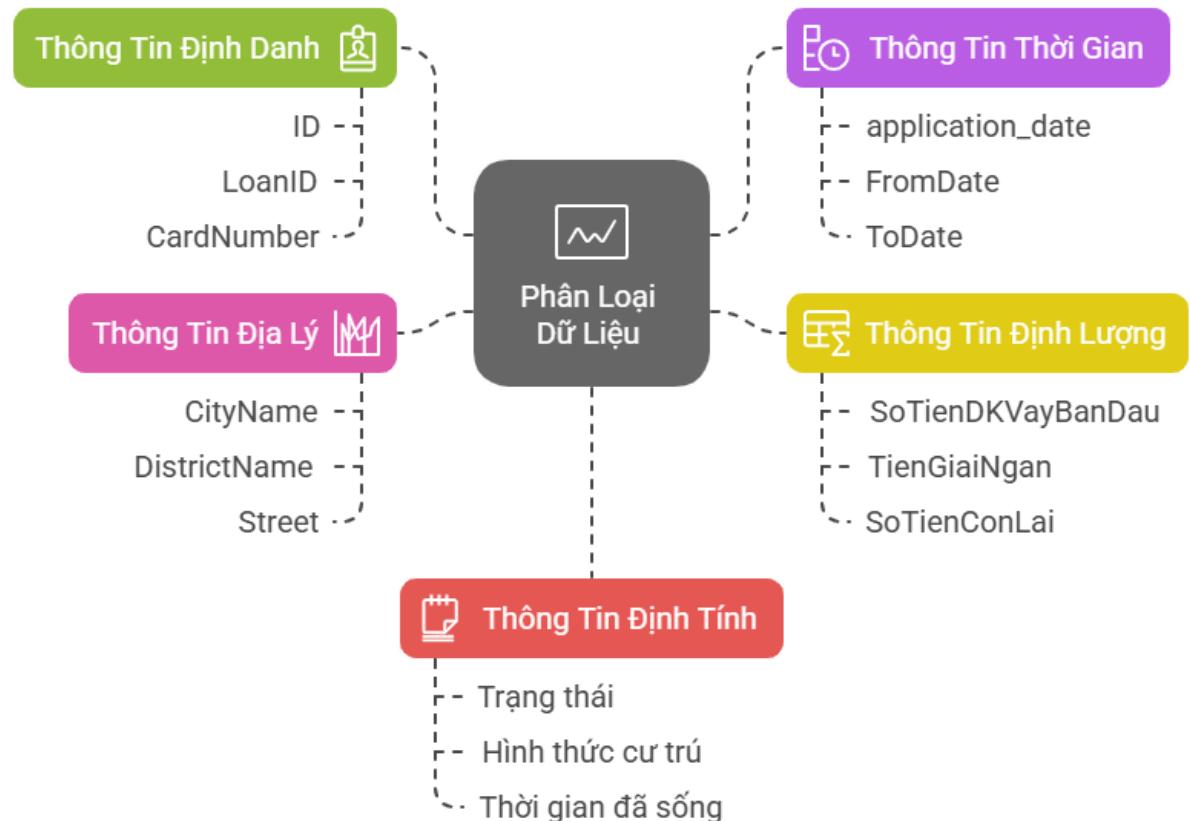
Công nghệ

Sử dụng công
nghệ hiện đại để
tối ưu hóa quy
trình

2.Tổng quan về dataset

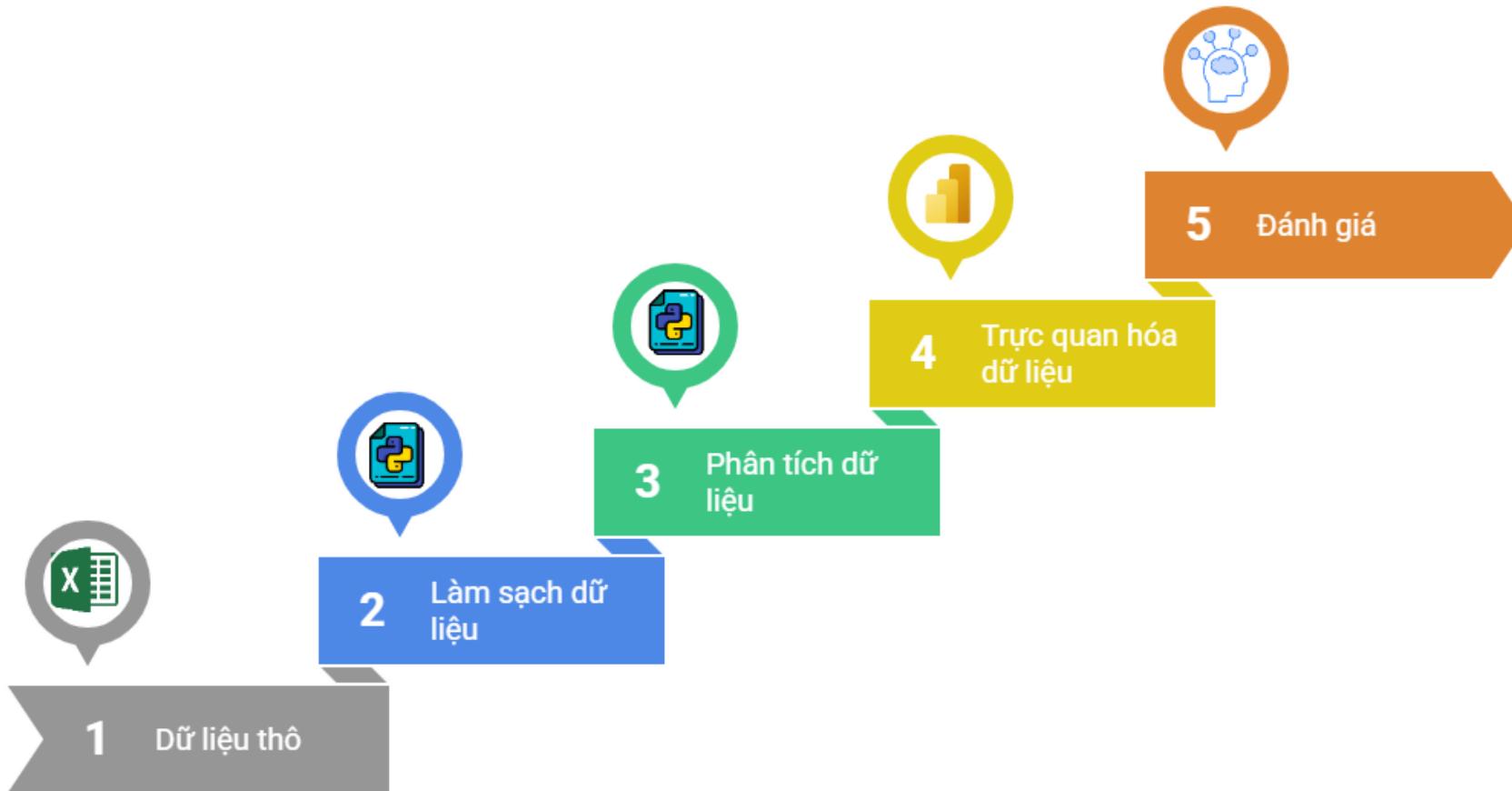
- Số lượng dòng: 2,383
- Số lượng cột: 46
- Kiểu dữ liệu: int, float, bool, datetime, object
- Link bộ dữ liệu: <https://byvn.net/zaTT>

Phân loại theo loại dữ liệu

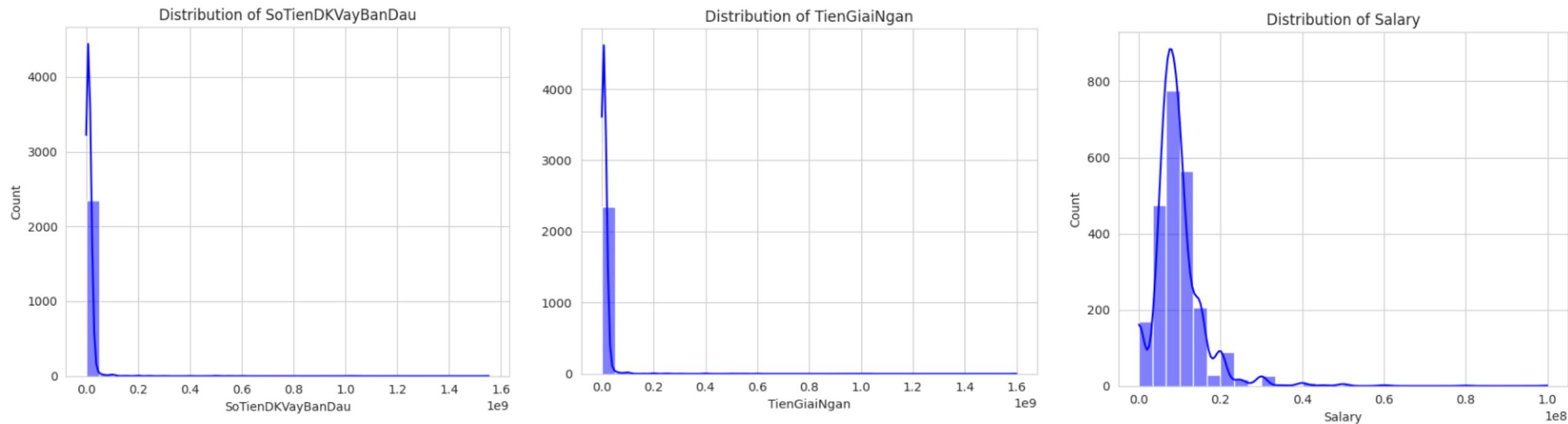


QUY TRÌNH TIỀN PHÂN TÍCH DỮ LIỆU

1. Các công cụ và kĩ thuật phân tích đã sử dụng



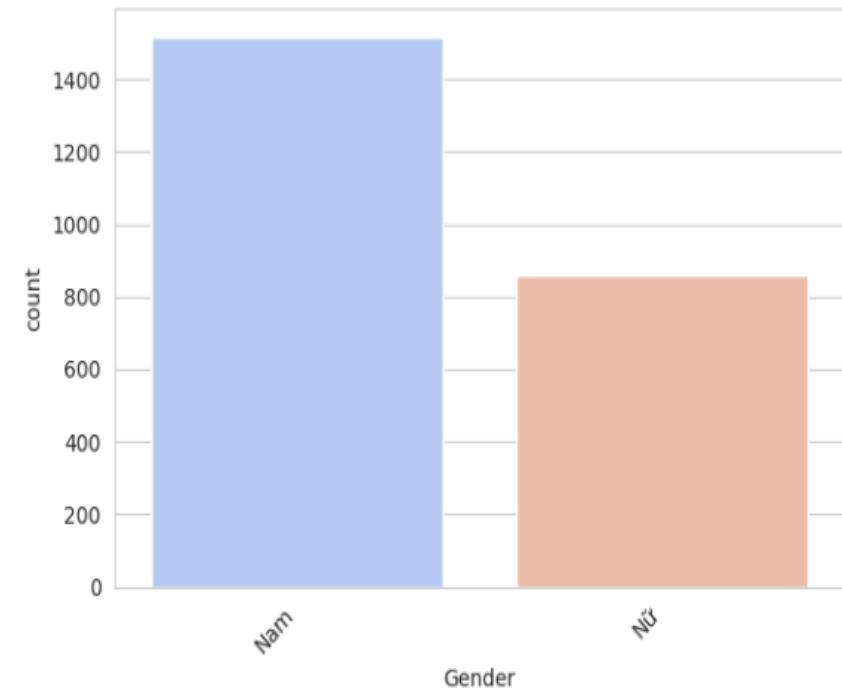
2. Khám phá dữ liệu (EDA)



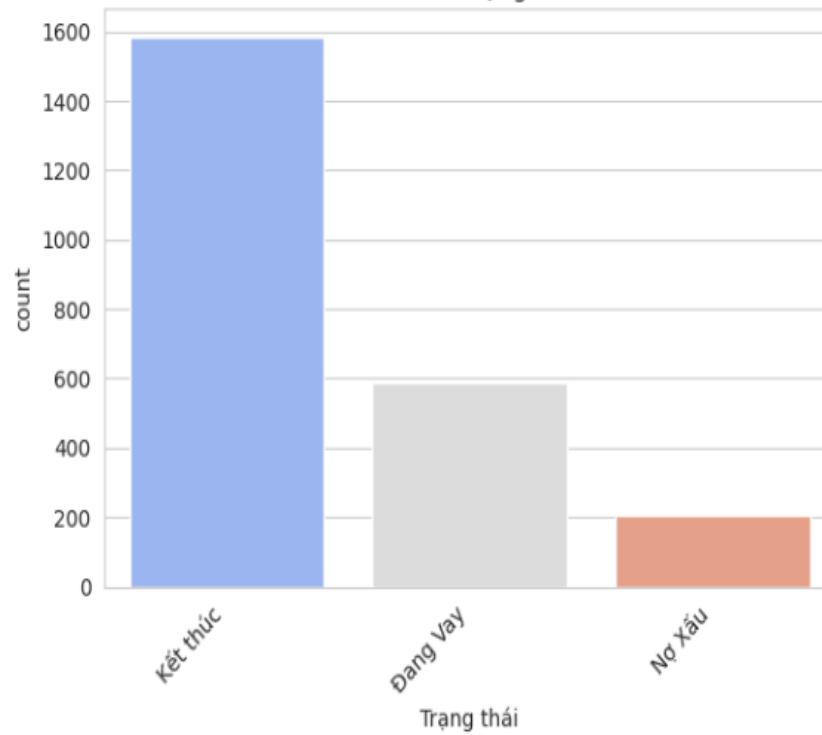
- **Histogram** và **KDE Plot** để khám phá phân phối biến số liên tục (**SoTienDKVayBanDau**, **TienGiaiNgan**, **Salary**)

2. Khám phá dữ liệu (EDA)

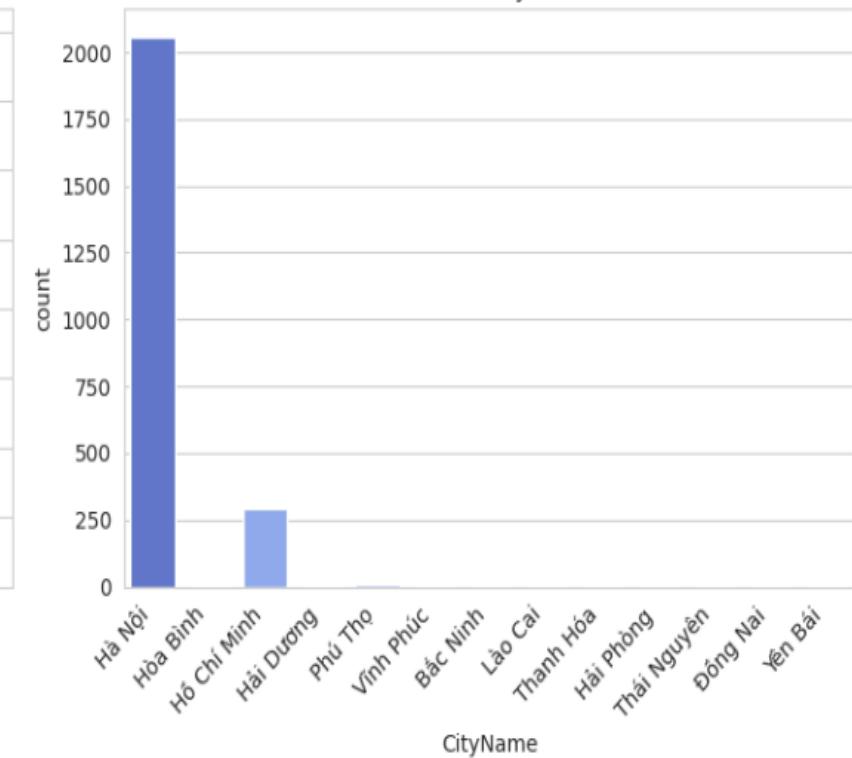
Phân bổ Gender



Phân bổ Trạng thái

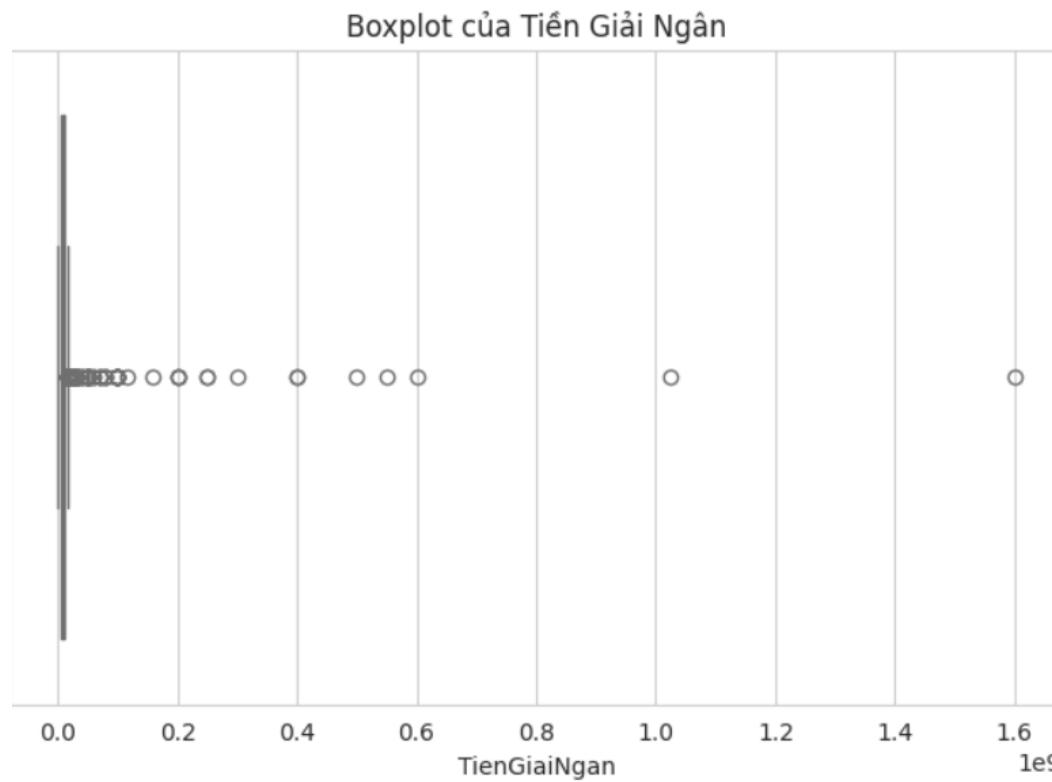


Phân bổ CityName



- Count plot để kiểm tra sự phân bổ tần suất của các nhóm phân loại (**Gender, Trạng thái, CityName**)

2. Khám phá dữ liệu (EDA)



- **Boxplot** cho các biến số liên tục để kiểm tra các ngoại lai và sự phân phối của **TienGiaiNgan**.

KẾT QUẢ PHÂN TÍCH

1. Phân tích tổng quan

- Tổng số lượng khách hàng của TIMA là **1735 khách hàng**.
- Các khách hàng đã thực hiện **1867 khoản vay** với tổng số tiền là **25 tỷ đồng**.

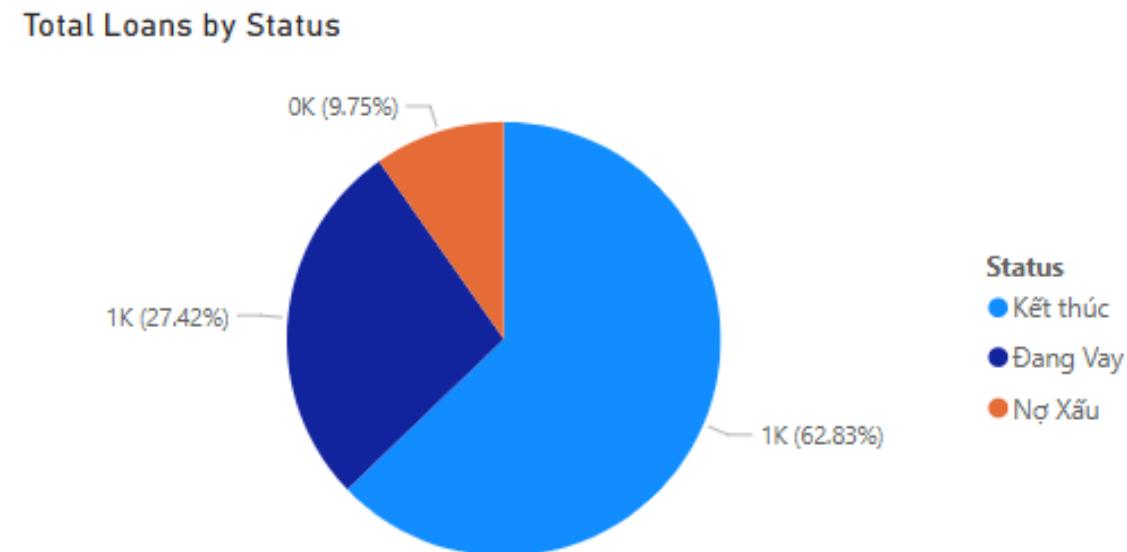
1735
Total Customers

1867
Total Loans

25bn
Total Loan Amount

Tình trạng khoản vay

- 62.83% “Kết thúc”** → Tín hiệu tích cực, khách hàng trả nợ tốt.
- 27.42% “Đang vay”** → Duy trì nguồn thu, cần theo dõi sát để phòng rủi ro.
- 9.75% “Nợ xấu”** → Tỷ lệ cao, đáng lo ngại

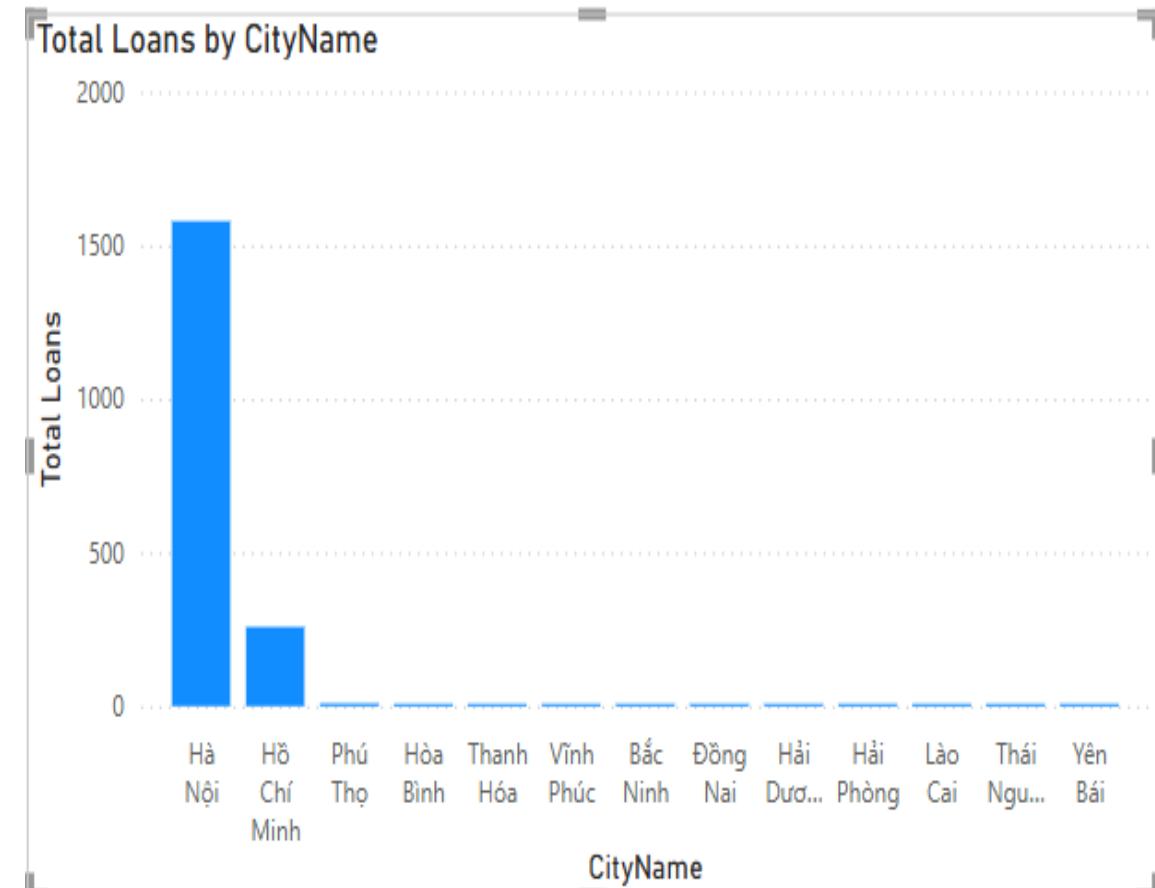


1. Phân tích tổng quan

Tổng số khoản vay theo tỉnh thành:

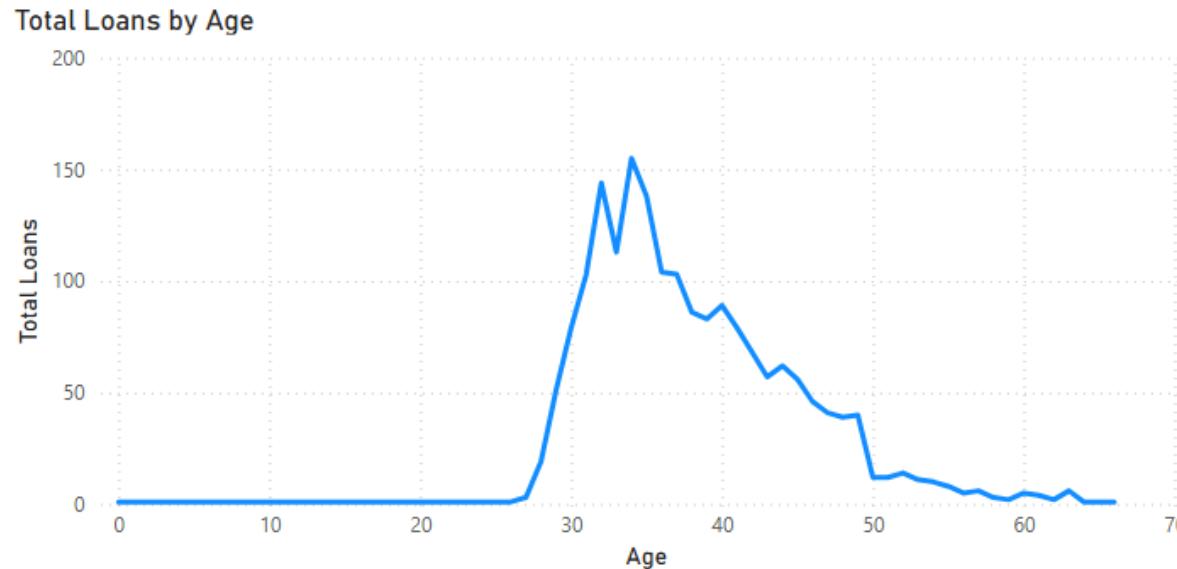
- Các khoản vay chiếm đa số ở 2 thành phố lớn là **Hà Nội** và **TP. Hồ Chí Minh** với lần lượt là **1579** và **258** khoản vay

➡ Tập trung tại Hà Nội & TP.HCM **do dân cư đông đúc, mức chi tiêu cao** và **dễ tiếp cận** dịch vụ tài chính.



1. Phân tích tổng quan

Phân tích độ tuổi khách hàng



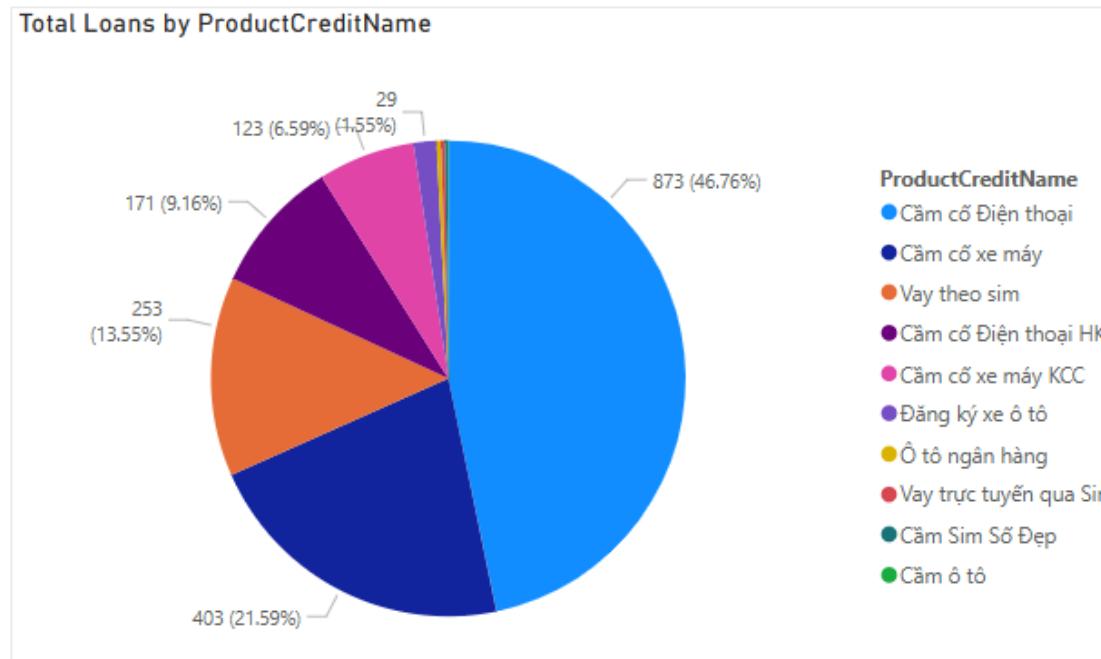
Age	Total Loans
34	155
32	144
35	138
33	113
36	104
31	103
37	103
40	89
38	86
39	83
30	79
41	79
42	68
Total	1867

- Nhóm tuổi **30–40** chiếm tỷ trọng lớn nhất. Sau tuổi 40, số lượng vay giảm dần rõ rệt. Người dưới 25 và trên 50 gần như không vay.

💡 Insight rút ra: Khách hàng vay vốn tập trung chủ yếu ở nhóm 30–40 tuổi – đây là **độ tuổi ổn định công việc, có nhu cầu cao về tài chính** (mua nhà, nuôi con, đầu tư...).

1. Phân tích tổng quan

Số lượng khoản vay theo từng loại sản phẩm tín dụng:



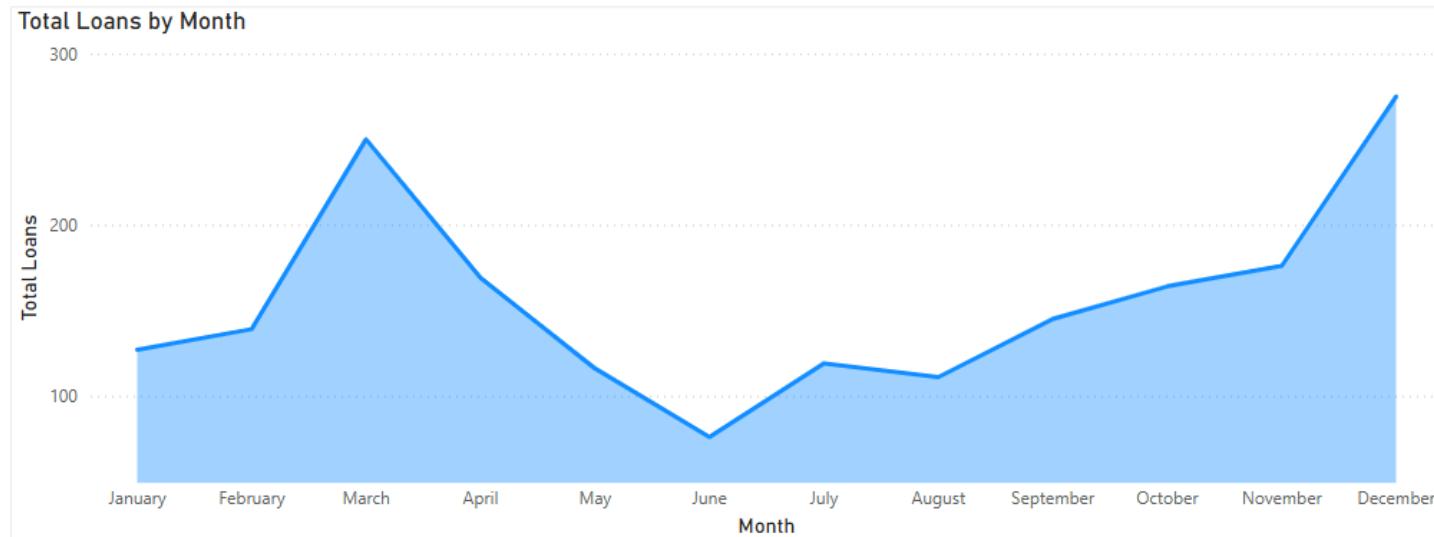
• **Điện thoại** chiếm **gần 47% tổng số khoản vay**, đây là sản phẩm phổ biến nhất.

• Nguyên nhân:

- Người dùng chủ yếu thuộc phân khúc thu nhập trung bình/thấp.
- Điện thoại là tài sản dễ tiếp cận, dễ định giá, dễ thanh lý → phù hợp để cầm cố.
- Quy trình vay đơn giản, phê duyệt nhanh.

1. Phân tích tổng quan

Phân tích thời điểm vay



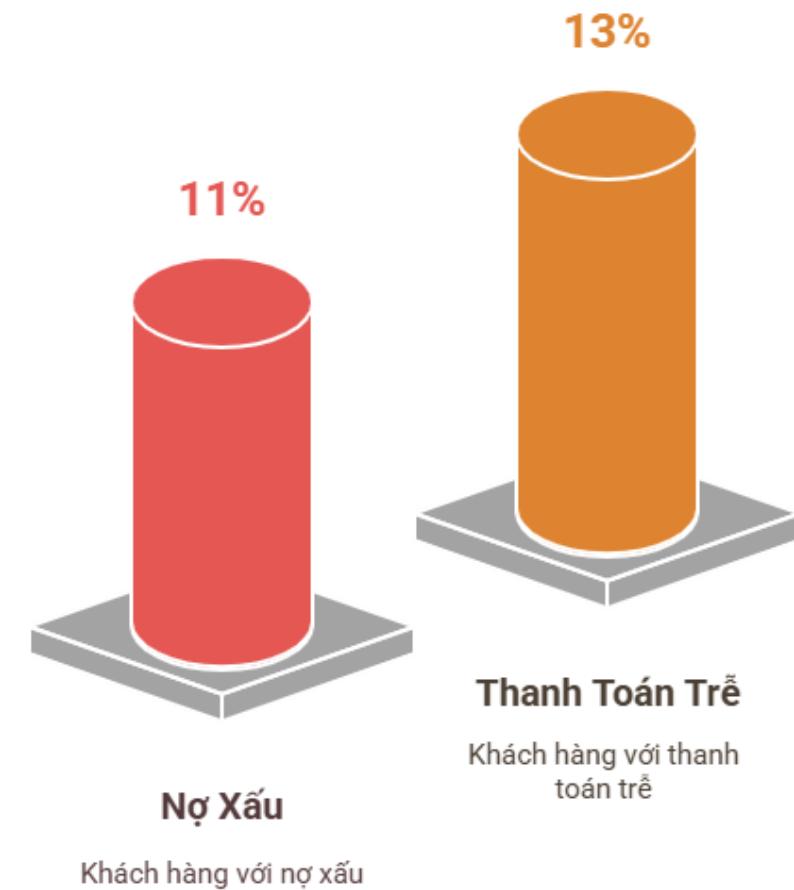
- Số lượng khoản vay tăng mạnh vào **tháng 3** và **tháng 12**, chiếm tỷ lệ cao nhất trong năm.
➡ Đây là thời điểm người dân có nhu cầu **nhu cầu đầu tư, kinh doanh** và **chuẩn bị chi tiêu lớn cho dịp lễ Tết**.

2. Phân tích Rủi ro tín dụng

1. Tỷ lệ nợ xấu

- Tỷ lệ này cho thấy có khoảng **11%** số khách hàng của TIMA có **nợ xấu** (HasBadDebt = 1).
- Đây là nhóm có **rủi ro tín dụng cao** và có khả năng không trả được nợ

Tỷ lệ Khách hàng của TIMA với Các Vấn đề Nợ



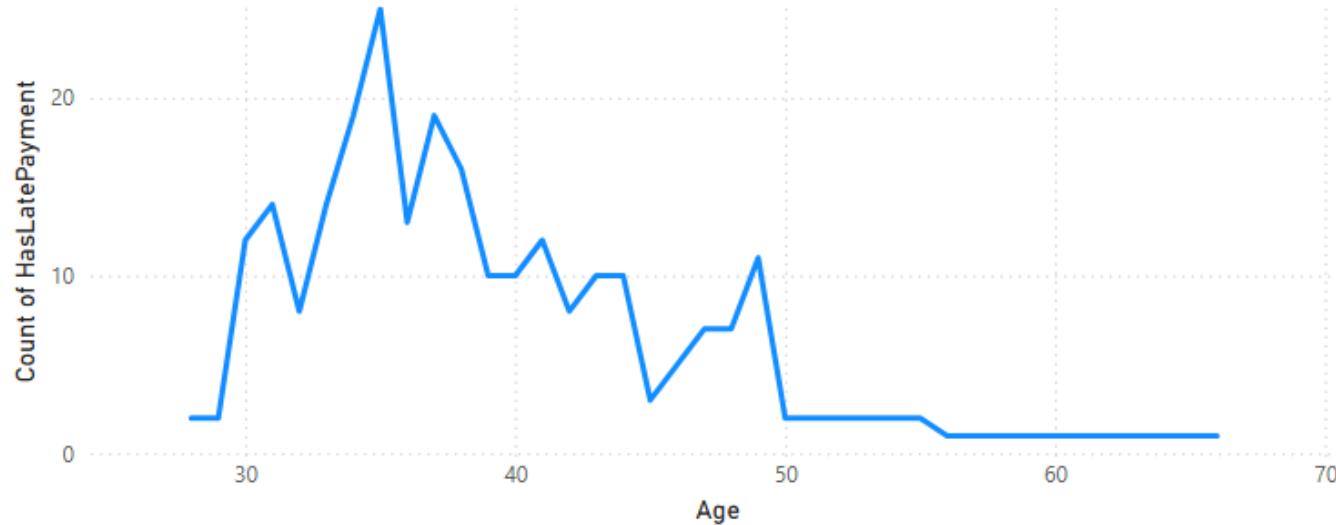
2. Tỷ lệ khách hàng có nợ trễ hạn

- Có khoảng **13%** khách hàng của TIMA có **nợ trễ hạn** (HasLatePayment = 1).
- Mặc dù trễ hạn không đồng nghĩa với mất khả năng thanh toán, nhưng đây là nhóm có nguy cơ trở thành nợ xấu trong tương lai.

2. Phân tích Rủi ro tín dụng

Phân tích yếu tố ảnh hưởng đến nợ trễ hạn

Count of HasLatePayment by Age



CityName	Late Payment Rate
Hòa Bình	0.33
Thanh Hóa	0.25
Phú Thọ	0.22
Hồ Chí Minh	0.16
Hà Nội	0.13
Total	0.13

- Theo biểu đồ ta thấy **nhóm tuổi từ 25–35** có tỷ lệ **trễ nợ** cao hơn hẳn so với nhóm còn lại, có thể do chưa có tích lũy tài chính vững.
- Tỷ lệ trễ hạn có sự chênh lệch rõ ở các tỉnh như **Hòa Bình, Thanh Hóa, Phú Thọ** cho thấy rủi ro tín dụng cần được kiểm soát chặt tại khu vực này.

2. Phân tích Rủi ro tín dụng

Phân tích yếu tố ảnh hưởng đến nợ trễ hạn

Nhân viên chính thức chiếm phần lớn các khoản vay trễ hạn do số lượng vay cao

Dù các nhóm như **nhân viên không có hợp đồng, công nhân, hay lao động tự do** có số lượng vay ít hơn, nhưng đặc điểm công việc thiếu ổn định, không có bảo hiểm hoặc hợp đồng ràng buộc khiến **nguy cơ trễ hạn cao hơn nếu xét theo tỷ lệ**

Count of HasLatePayment by JobName



MÔ HÌNH DỰ ĐOÁN

1.Xây dựng mô hình

Dự báo khả năng trả nợ đúng hạn của khách hàng (Loan Default Prediction)

- **Biến mục tiêu:** `Loan_Status` (Trạng thái khoản vay: "Trả đúng hạn" hoặc "Trễ hạn").
- **Cách tính:**
 - **Trả đúng hạn(`Loan_Status=0`):** Nếu số tiền còn lại (`SoTienConLai`) = 0 và ngày đáo hạn (`ToDate`) đã qua, hoặc khách hàng không có thanh toán trễ (`HasLatePayment=0`).
 - **Trễ hạn(`Loan_Status=1`) :** Nếu số tiền còn lại (`SoTienConLai`) > 0 và ngày đáo hạn (`ToDate`) đã qua, hoặc nếu có thanh toán trễ (`HasLatePayment=1`).
- **Biến độc lập:** gồm các thông tin đặc trưng về khách hàng và khoản vay, có thể ảnh hưởng đến khả năng trả nợ:
 - Thông tin khách hàng: Nghề nghiệp (`JobName`), tỉnh thành (`CityName`), độ tuổi, giới tính...
 - Lịch sử tín dụng: Có nợ xấu (`HasBadDebt`), có trễ hạn trước đó (`HasLatePayment`), số khoản vay đã từng có.
 - Thông tin khoản vay: Số tiền vay, số tiền còn lại (`SoTienConLai`), thời gian vay, ngày đáo hạn (`ToDate`), v.v.

2. Kết quả của mô hình

Tổng quan

- **Độ chính xác (Accuracy):** 93.9%

- Mô hình hoạt động tốt trên cả hai lớp:

- **Lớp 0 (Không trễ nợ):** F1-score = 0.95
- **Lớp 1 (Có trễ nợ):** F1-score = 0.93

- **Weighted avg** và **Macro avg** ở chỉ số F1-score đều đạt 94%

→ Mô hình đạt F1-score cao cả ở **macro avg** và **weighted avg**, cho thấy hiệu quả đồng đều giữa hai nhóm khách hàng (trả **đúng hạn** và **trễ hạn**), đồng thời duy trì hiệu suất cao trên toàn bộ dữ liệu thực tế.

	Accuracy: 0.939203354297694				
	precision	recall	f1-score	support	
0	0.91	0.98	0.95	261	
1	0.97	0.89	0.93	216	
accuracy				0.94	477
macro avg	0.94	0.93	0.94	477	
weighted avg	0.94	0.94	0.94	477	

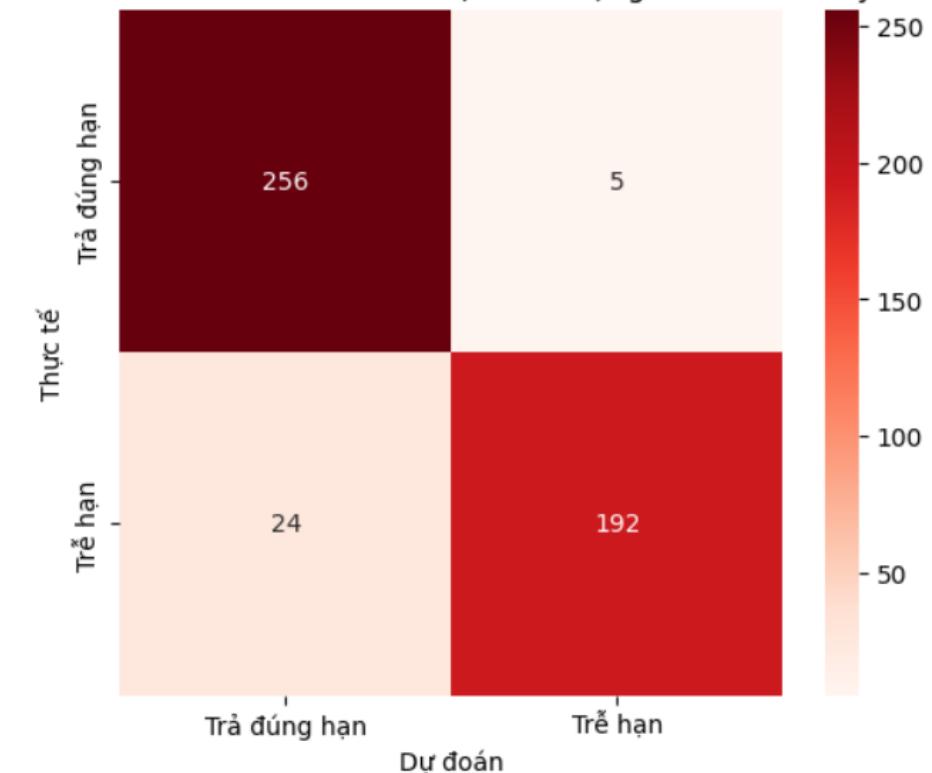
2. Kết quả của mô hình

Hiệu quả mô hình qua Confusion Matrix

- 256 khách hàng thực sự trả đúng hạn và mô hình dự đoán đúng.
- 192 khách hàng thực sự trễ hạn và được mô hình phát hiện chính xác.
- 24 khách hàng trễ hạn nhưng bị dự đoán nhầm là trả đúng hạn.
- 5 khách hàng thực sự trả đúng hạn nhưng bị báo nhầm là trễ hạn.

→ Mức sai lệch **chấp nhận được**, đặc biệt trong bối cảnh dữ liệu thực tế.

Biểu đồ Confusion Matrix – Dự đoán trạng thái khoản vay



❖ Qua kết quả ta thấy mô hình có **hiệu suất cao, ổn định và cân bằng**, phù hợp để ứng dụng vào bài toán **dự đoán rủi ro trễ hạn** trong cho vay khách hàng.

KẾT LUẬN VÀ ĐỀ XUẤT

Kết luận và đề xuất

Khách hàng

- Phát triển sản phẩm theo phân khúc khách hàng mục tiêu: Với nhóm tuổi 30–40, có thể gắn các gói vay với tiện ích đời sống: vay tiêu dùng, cải tạo nhà, đầu tư nhỏ.
- Không ưu tiên nhóm **< 25 & > 50 tuổi** → Cầu thấp, rủi ro cao, chỉ mở rộng khi có dữ liệu bổ sung.
- Mở rộng chiều sâu tại **Hà Nội & TP.HCM** → Tăng độ phủ, hợp tác đối tác, ưu đãi cạnh tranh.

Khoản vay:

- Với các sản phẩm hiện chiếm tỷ lệ thấp như: **cầm ô tô, vay trực tuyến qua sim...** → cần khảo sát thị trường và A/B Testing để xác định tiềm năng thực tế trước khi đầu tư mở rộng.
- Tập trung vào 3 sản phẩm chiếm >80% tổng số khoản vay: **Cầm cố điện thoại, xe máy, vay theo sim.**
→ Tăng trải nghiệm số hóa, rút ngắn thời gian phê duyệt, kết hợp ưu đãi theo mùa vụ để giữ chân khách hàng.

Kết luận và đề xuất

Rủi ro tín dụng

- Rà soát lại hồ sơ nhóm đã có nợ xấu để cập nhật chính sách phê duyệt và thu hồi nợ hiệu quả hơn.
- Bổ sung yếu tố hành vi (lịch sử trễ hạn, mức độ trả nợ đúng hạn) để tăng độ chính xác.
- Gọi nhắc nợ sớm, hỗ trợ khách hàng khó khăn tạm thời để tránh phát sinh nợ xấu.

Mô hình dự đoán

- Thử nghiệm thêm các mô hình khác:
Logistic Regression, Gradient Boosting, XGBoost, LightGBM...
→ So sánh hiệu suất để chọn mô hình tốt nhất cho bài toán.
- Tuning tham số (Hyperparameter tuning):
Dùng Grid Search / Random Search để tối ưu kết quả.

A large, faint watermark of the HUST logo is visible across the entire slide. The logo consists of the letters "HUST" in a white, bold, sans-serif font, with a stylized orange and red dotted arrow graphic to its right.

HUST

THANK YOU !