

FINAL PROJECT:

PHÂN TÍCH DỮ LIỆU TẠI TIMA

Báo cáo Vận hành & Giải pháp Dự đoán Rủi ro

Người trình bày: HOÀNG THÁI DUY

Ngày trình bày: 27/8/2025



NỘI DUNG

I. Bối cảnh & Mục tiêu

Tại sao chúng ta cần phân tích dữ liệu ngay bây giờ?

II. Nền tảng Dữ liệu

Đảm bảo sự tin cậy cho mọi quyết định.

III. Phân tích & Kết quả

Những gì dữ liệu đang nói với chúng ta.

IV. Đề xuất & Giải pháp

Từ Insight đến Hành động.

V. Lộ trình & Thảo luận

Các bước tiếp theo.

Phần 1: Giới thiệu & Mục tiêu

Bối Cảnh: "Điểm Uốn" Trong Thị Trường FinTech



Cạnh tranh Gia tăng

Nhiều đối thủ mới, cả trong và ngoài nước, đang tạo áp lực lớn lên thị phần và lợi nhuận của TIMA.



Chi phí Thu hút Khách hàng (CAC) Tăng

Quảng cáo ngày càng đắt đỏ, khiến việc mở rộng cơ sở khách hàng trở nên kém hiệu quả hơn.



Rủi ro Nợ Xấu Tiềm ẩn

Yếu tố sống còn ảnh hưởng trực tiếp đến lợi nhuận và sự ổn định tài chính của công ty.

Câu hỏi lớn: Làm thế nào để TIMA tăng trưởng hiệu quả và bền vững?

Mục Tiêu Dự Án: Chuyển Đổi Dữ Liệu Thành Lợi Thế Cạnh Tranh

HIỂU RÕ (Insight)



Xây dựng báo cáo vận hành trực quan (Power BI Dashboard) để theo dõi sức khỏe kinh doanh theo thời gian thực cho Ban Lãnh đạo, Trưởng phòng Kinh doanh, Marketing.

DỰ BÁO (Foresight)



Xây dựng mô hình Học máy để dự đoán sớm các khoản vay có nguy cơ nợ xấu, hỗ trợ Phòng Quản lý Rủi ro, Nhân viên Thẩm định.

HÀNH ĐỘNG (Action)



Đưa ra các giải pháp cụ thể để tối ưu hóa sản phẩm, marketing và quy trình thẩm định cho tất cả các phòng ban liên quan.

PHẦN 2: TIỀN XỬ LÝ DỮ LIỆU

- Để đảm bảo độ tin cậy của phân tích, chúng ta đã thực hiện quy trình xử lý dữ liệu chặt chẽ.



TỔNG QUAN VỀ DỮ LIỆU ĐẦU VÀO & TIỀN XỬ LÝ DỮ LIỆU

Dữ liệu đầu vào gồm 2.384 bản ghi khoản vay từ 01/01/2016 đến 31/12/2019.

a. Các vấn đề dữ liệu:

- Nhiều giá trị **NaN/null** hoặc **0/lâm** ở các cột: SoTienDKVayBanDau, TienGiaiNgan, SoTienConLai, LongestOverdue.
- 3 cột trùng thông tin → **Loại bỏ**.
- Một số cột có **dữ liệu trống** hoặc **không đúng định dạng**:
- Nhiều **dữ liệu thiếu** ở trường thời gian (ThoiGianSong).

b. Xử lý dữ liệu:

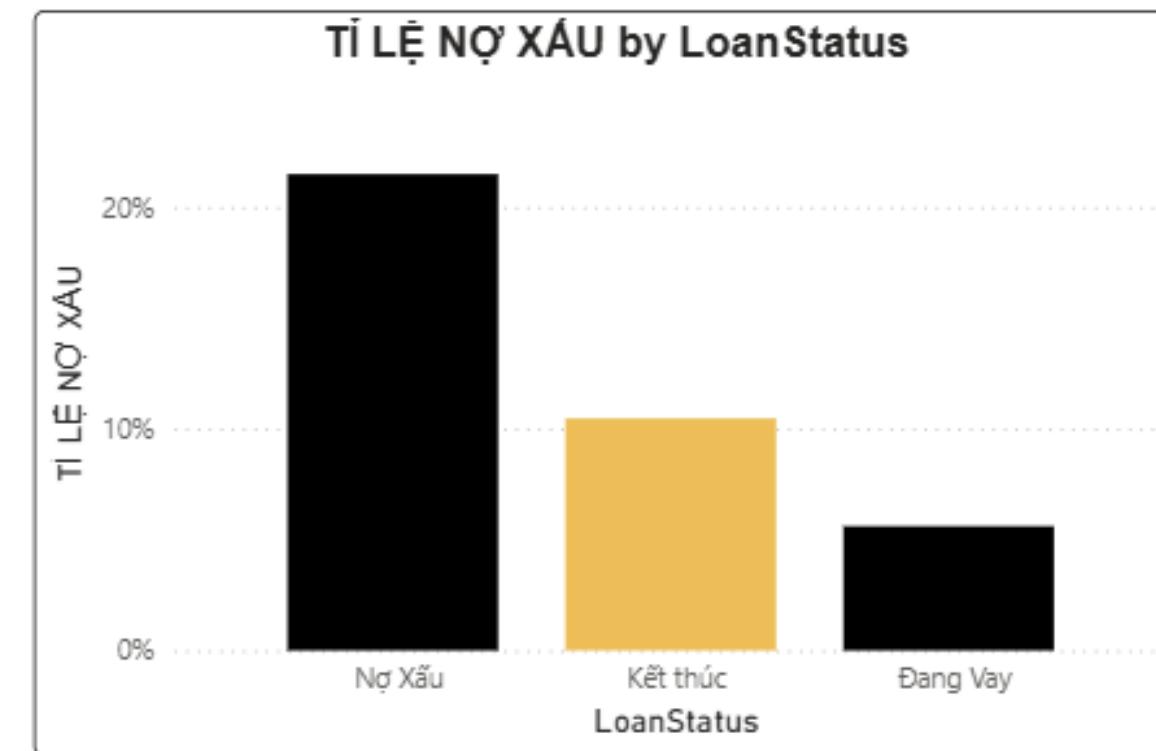
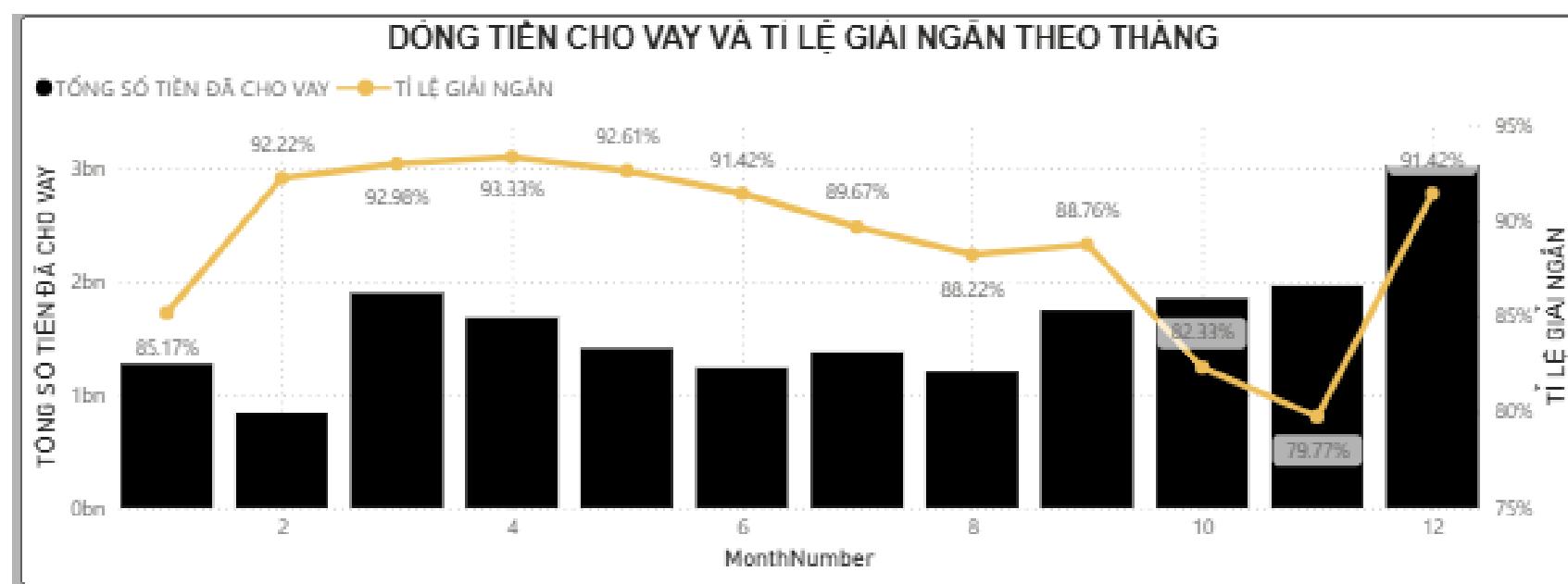
- Làm sạch & Chuẩn hóa:** Xử lý giá trị trống (N/A), đồng bộ định dạng (ngày tháng, tiền tệ), loại bỏ dữ liệu không hợp lệ. Xóa dữ liệu ngoại lai (IQR)
- Tạo thuộc tính mới (Feature Engineering):** Tuổi khách hàng, Thời hạn vay, Tỷ lệ giải ngân để tối ưu hóa phân tích.

"Chúng ta đang làm việc trên một bộ dữ liệu chất lượng cao, sẵn sàng cho những phân tích sâu sắc nhất."

Phần 3: Phân tích & Kết quả

A. Phân tích Mô tả: Bức tranh toàn cảnh

KPI Tổng Quan: Sức Khỏe Kinh Doanh của TIMA



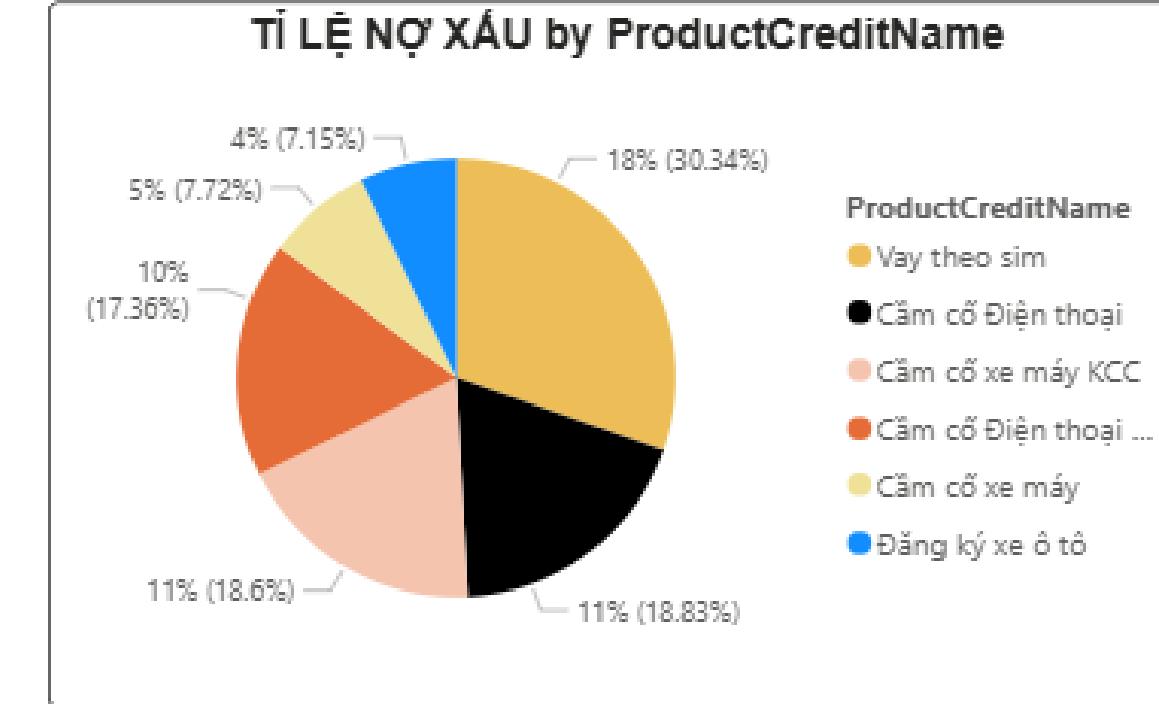
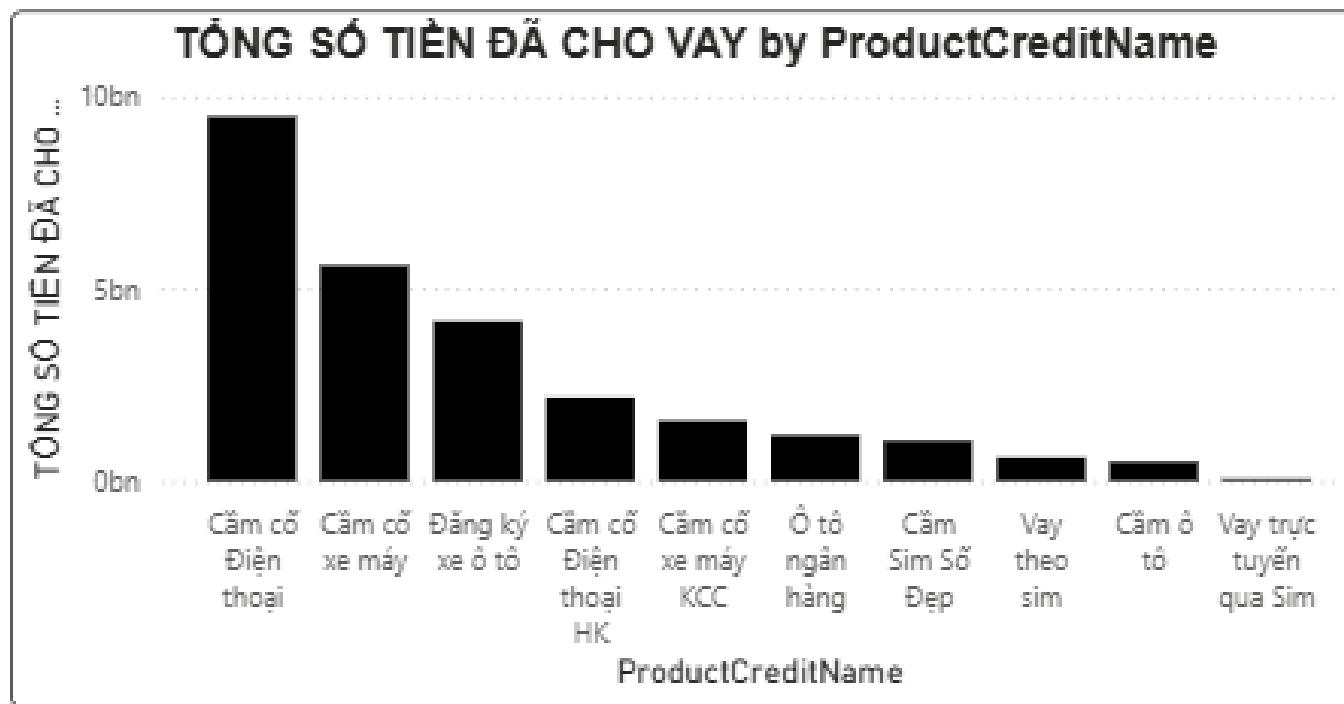
- Tiền cho vay cao nhất tháng 12, thấp nhất tháng 1.
- Tỉ lệ giải ngân >89% hầu hết các tháng, thấp nhất tháng 11 (79.77%).
- Tỉ lệ giảm dần từ tháng 3–11, tăng lại vào tháng 12.

Dòng tiền cho vay ổn định, nhưng tỉ lệ giải ngân giảm dần trong năm và chạm đáy tháng 11 trước khi phục hồi mạnh vào cuối năm.

- Hiện tại, hơn 20% tổng số khoản vay đang ở trạng thái Nợ Xấu, tương ứng với rủi ro gần 4 tỷ VND trên tổng dư nợ.

Nợ xấu chủ yếu phát sinh ở các khoản vay đã được xác định rủi ro, cần tăng cường kiểm soát và phòng ngừa ngay từ giai đoạn thẩm định.

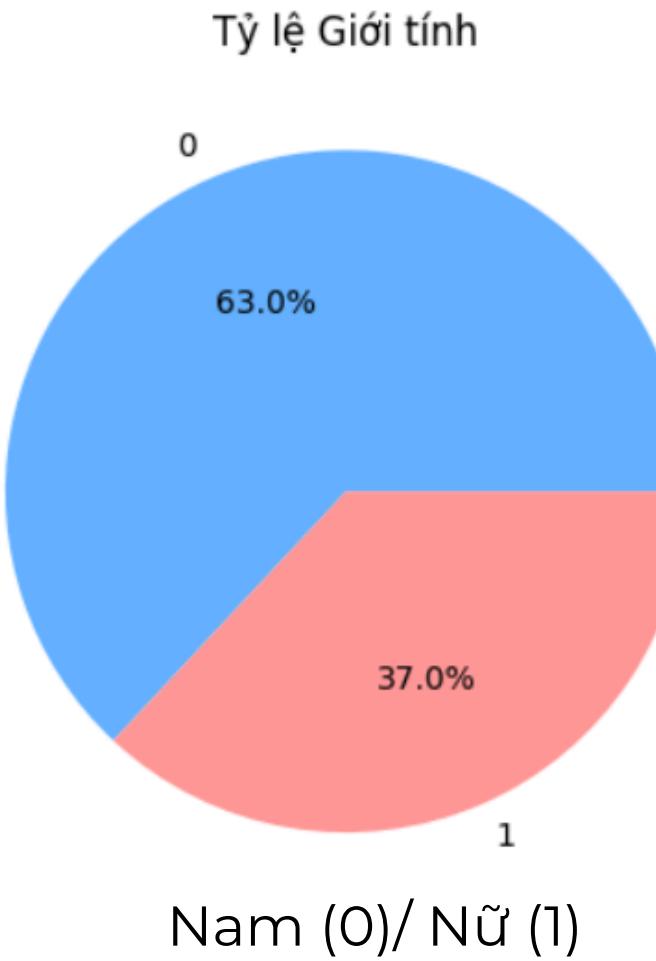
Hiệu Quả Danh Mục Sản Phẩm: Đâu là "Gà đẻ trứng vàng"?



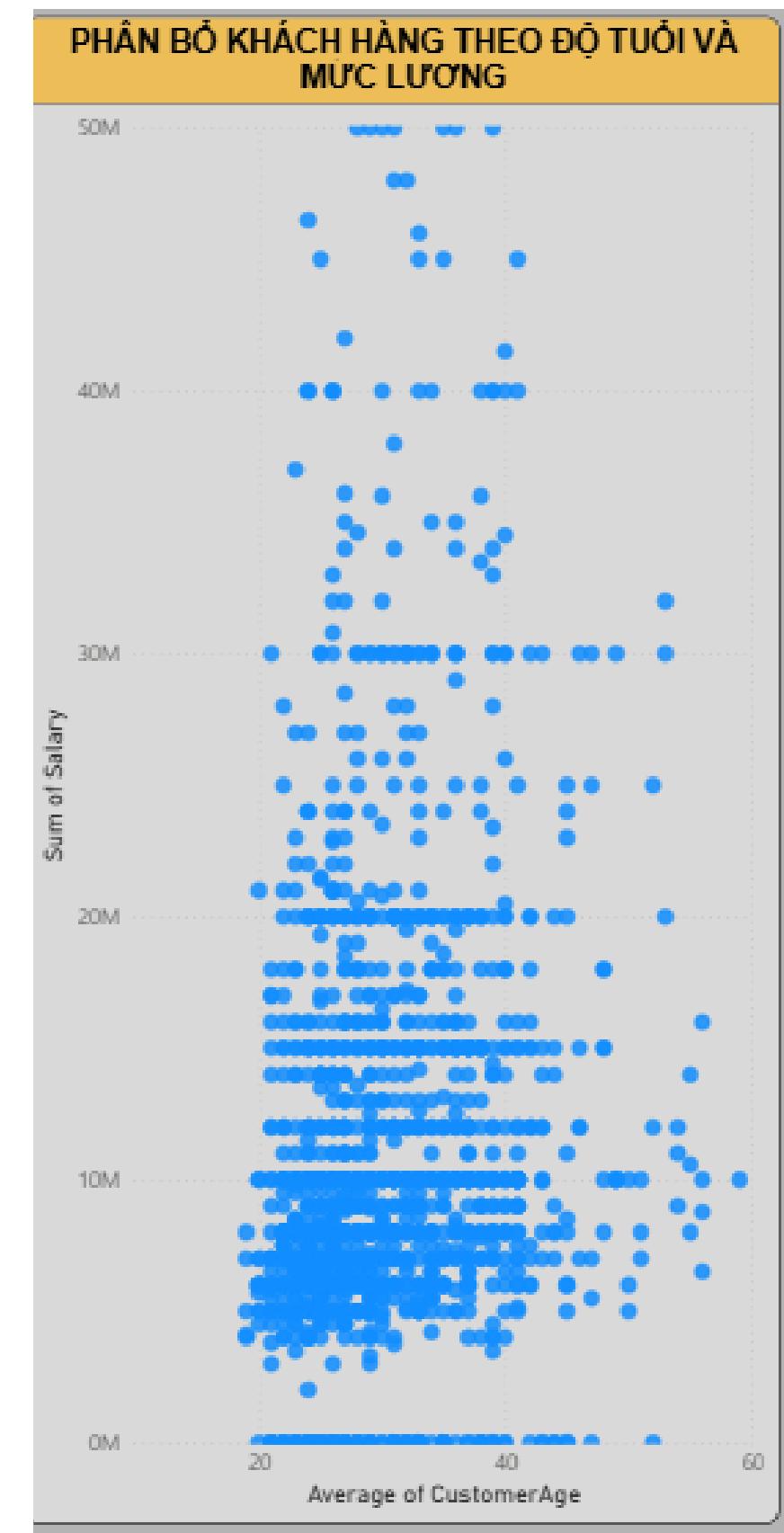
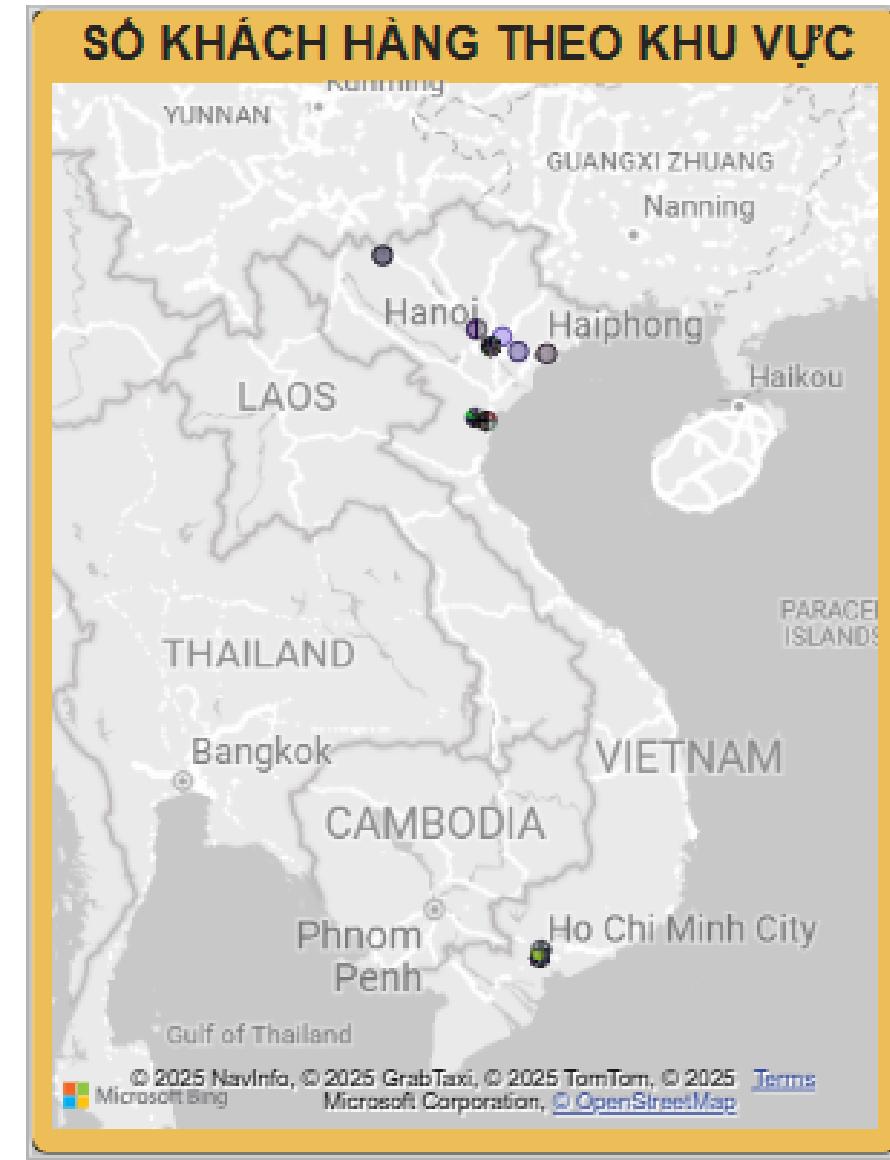
- Cầm cỗ Điện thoại là “gà đẻ trứng vàng” với doanh số cao nhất và nợ xấu ở mức chấp nhận được.
- Vay theo sim tuy nhỏ về doanh số nhưng nợ xấu cao, cần siết quản lý rủi ro.
- Báo cáo giúp xác định sản phẩm chủ lực và sản phẩm rủi ro, hỗ trợ ưu tiên nguồn lực và điều chỉnh chiến lược.

Danh mục sản phẩm đang phụ thuộc nhiều vào **Cầm cỗ Điện thoại** như nguồn lợi nhuận chính, trong khi một số sản phẩm nhỏ như **Vay theo sim** tiềm ẩn rủi ro nợ xấu cao, cần được kiểm soát chặt để bảo vệ hiệu quả tổng thể

Chân Dung Khách Hàng: Ai Đang Vay Tiền của Chúng Ta?



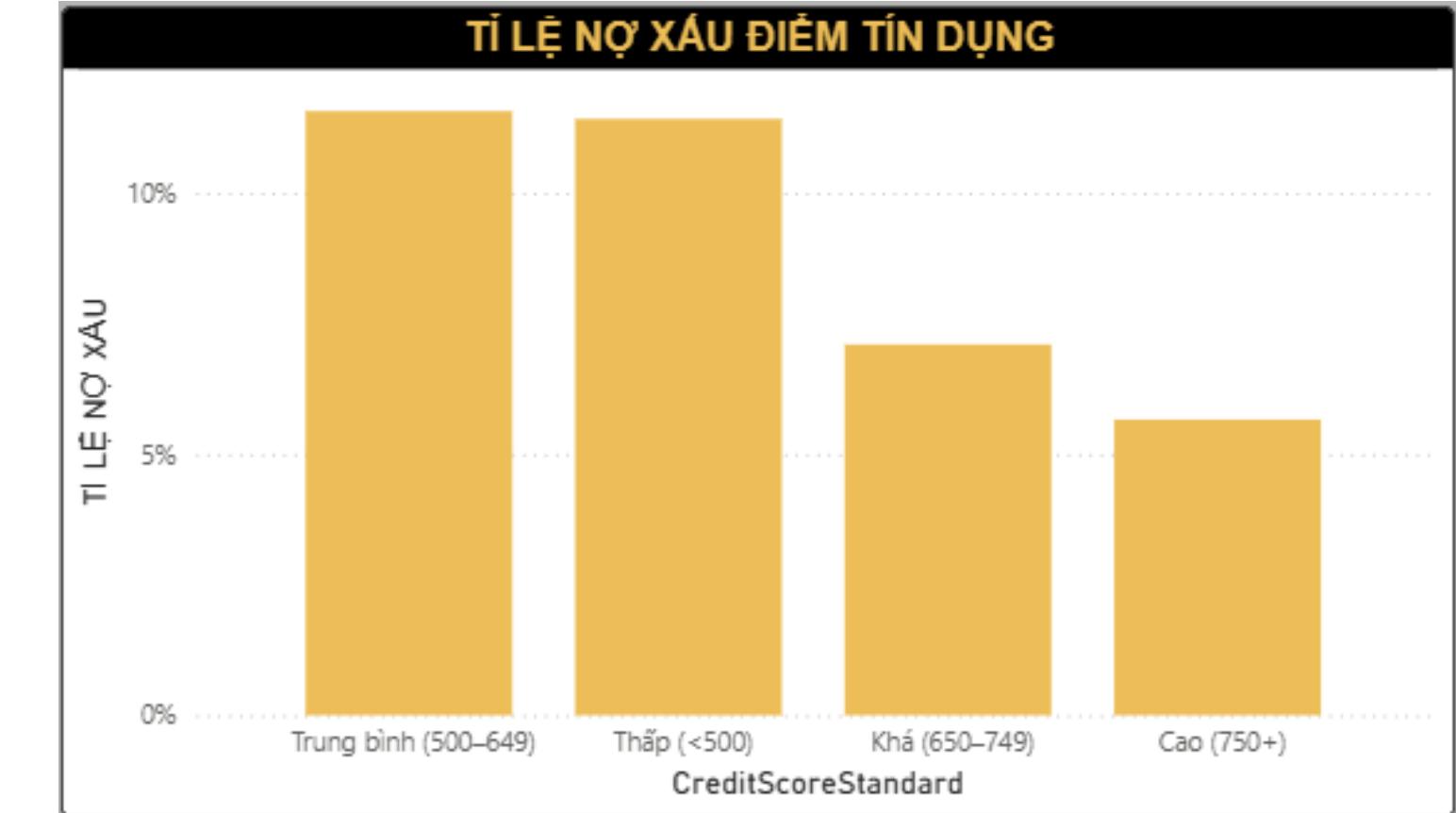
DistrictName	# UniqueCustomerCount
Hoàng Mai	156
Đống Đa	117
Long Biên	115
Nam Từ Liêm	110
Hà Đông	108
Hai Bà Trưng	104
Thanh Xuân	103
Bắc Từ Liêm	98
Cầu Giấy	86
Ba Đình	73



Phân tích từ các biểu đồ ta có thể thấy phần lớn khách hàng thuộc **giới tính Nam**, nằm trong **độ tuổi từ 20 - 40** có **mức thu nhập tương đối thấp** từ 0 - 20tr và chủ yếu ở các Trung Tâm Kinh Tế lớn như **Hà Nội** hay **TpHCM** (cụ thể khách hàng ở Hà Nội tập trung phần lớn ở các **khu vực gần trung tâm, nội thành**: Hoàng Mai, Đống Đa,...)

Ảnh Hưởng Của Điểm Tín Dụng

Số KH Theo Vùng & Điểm Tín Dụng					
CityName	Cao (750+)	Khá (650–749)	Thấp (<500)	Trung bình (500–649)	Total
Bắc Ninh		1			1
Hà Nội	40	399	202	774	1415
Hải Dương			1		1
Hải Phòng			1		1
Ho Chi Minh	13	78	10	120	221
Hòa Bình			2	2	4
Lào Cai				1	1
Phú Thọ		1	2	6	9
Thanh Hóa				4	4
Vĩnh Phúc			1	1	2
Total	53	479	219	908	1659

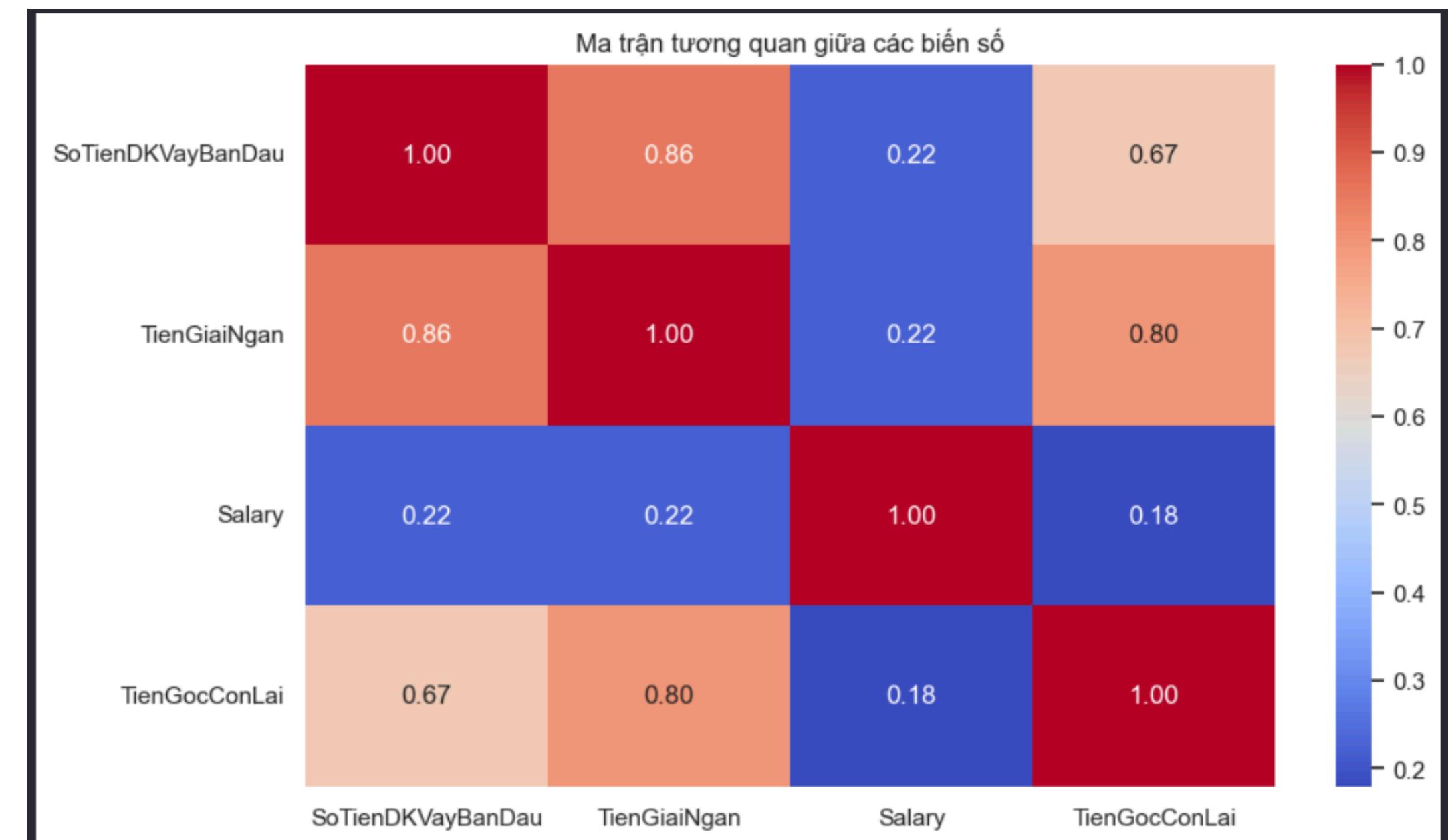


Thiết lập phân khúc duyệt vay theo điểm tín dụng.
Kiểm soát chặt nhóm < 500: xét thêm điều kiện hỗ trợ.
Ưu tiên tái tiếp cận nhóm > 700 để mở rộng bền vững.

B. Phân tích Chuẩn Đoán: Tương quan & Mối quan hệ

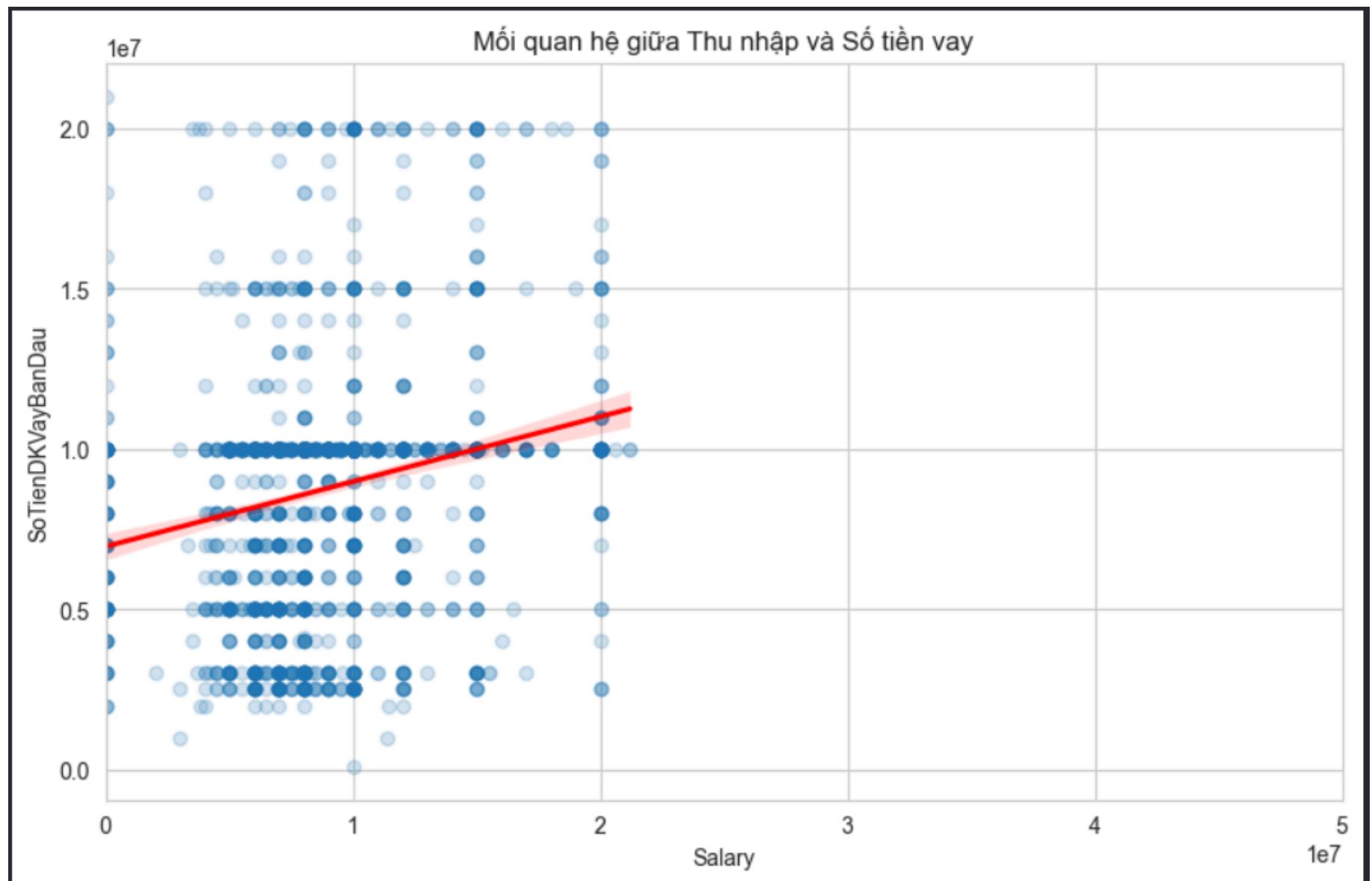
Tổng quan về sự tương quan

- Các biến về khoản vay (SoTien..., TienGiaiNgan...) tương quan chặt chẽ với nhau -> Dữ liệu logic.
- Điểm nổi bật: Salary (Lương) có tương quan rất yếu với số tiền vay.
- Suy ra: Lương không phải là yếu tố chính quyết định quy mô khoản vay.

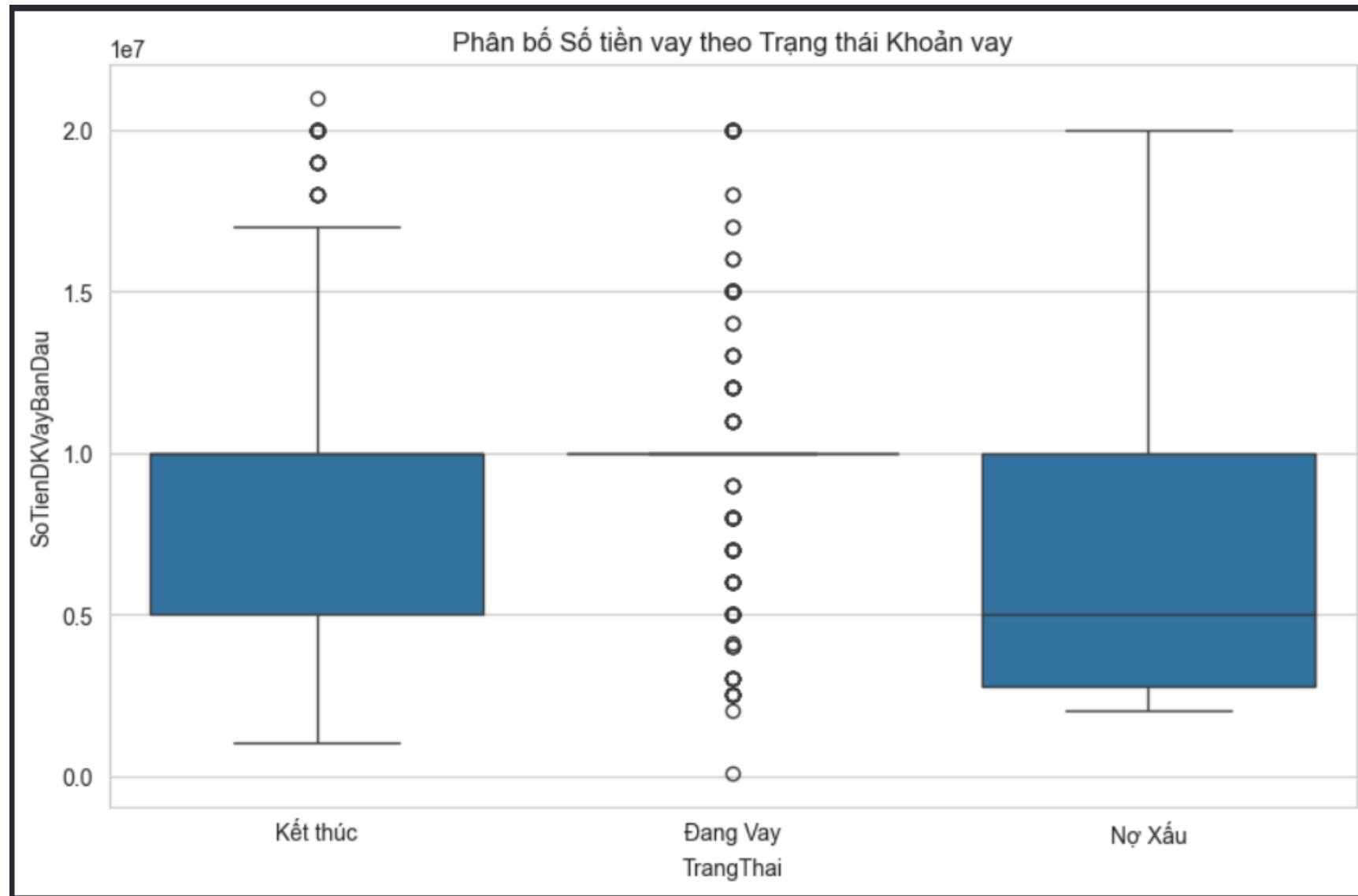


Mối quan hệ giữa Đặc điểm khách hàng và Hành vi vay

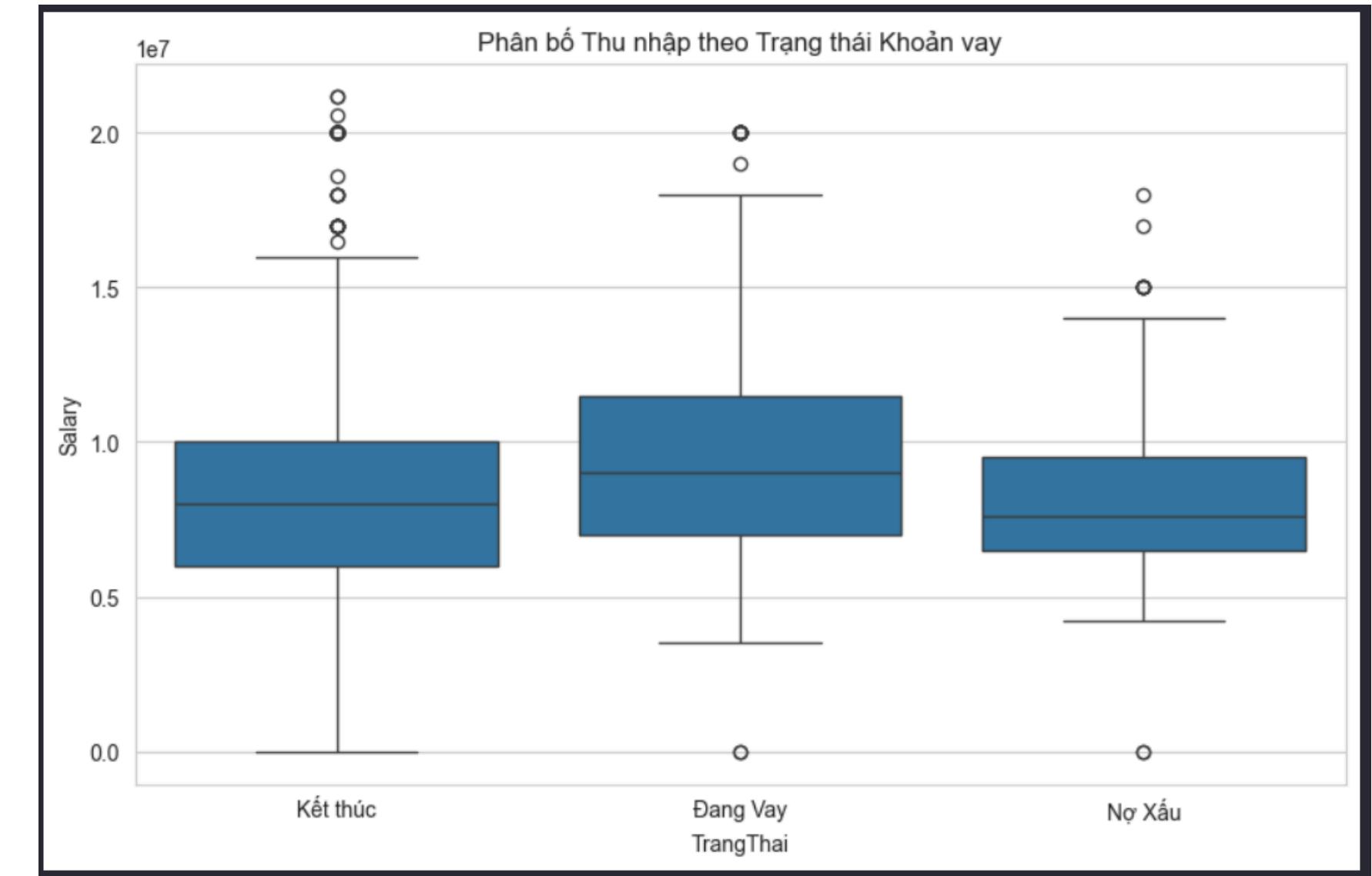
- Kết quả: Hệ số tương quan là +0.22.
- Nhận định: Có một mối tương quan dương, ở mức độ trung bình. Những người có thu nhập cao hơn có xu hướng vay số tiền lớn hơn. Đây là một dấu hiệu tích cực, cho thấy việc thẩm định có xem xét đến khả năng chi trả.



Mỗi quan hệ liên quan đến Rủi ro (Nợ xấu)



Ảnh hưởng của "Số tiền vay" đến "Nợ xấu"



Ảnh hưởng của "Thu nhập" theo "Tình trạng nợ xấu"

Chẩn đoán Nguyên nhân Gốc rẽ

- Từ các phân tích trên, chúng ta có thể chẩn đoán các nguyên nhân gốc rẽ dẫn đến nợ xấu tại TIMA:

Hồ sơ Tài chính Yếu

Khả năng trả nợ dễ bị ảnh hưởng bởi các biến động tài chính cá nhân.

Sản phẩm có Rủi ro Nội tại

Một số tài sản đảm bảo kém an toàn, dẫn đến tỷ lệ nợ xấu vượt trội.

Thẩm định chưa Tối ưu

Các khoản vay quá lớn so với thu nhập của khách hàng làm tăng gánh nặng trả nợ

Yếu tố Vùng miền

Có sự khác biệt về rủi ro giữa các khu vực địa lý

Phần 4: ĐỀ XUẤT VÀ GIẢI PHÁP

GIỚI THIỆU MÔ HÌNH DỰ ĐOÁN RỦI RO VỐN

Vấn Đề Của Quy Trình Hiện Tại

Thẩm định dựa trên kinh nghiệm:

- Thiếu nhất quán:** Quyết định có thể khác nhau giữa các chuyên viên.
- Bỏ sót rủi ro tiềm ẩn:** Khó nhận diện các mẫu hành vi phức tạp.
- Tốc độ chậm:** Cần thời gian để rà soát thủ công từng hồ sơ.
- Tính "Phản ứng":** Thường chỉ phát hiện nợ xấu khi đã quá muộn.

Giải Pháp Bằng Học Máy

Thẩm định dựa trên dữ liệu:

- Nhất quán & Khách quan:** Áp dụng một tiêu chuẩn duy nhất cho mọi hồ sơ.
- Dự báo sớm:** "Nhìn thấu" các hồ sơ có nguy cơ cao ngay từ đầu.
- Tốc độ tức thì:** Trả kết quả trong vài giây.
- Tính "Chủ động":** Cảnh báo trước để có hành động phòng ngừa.

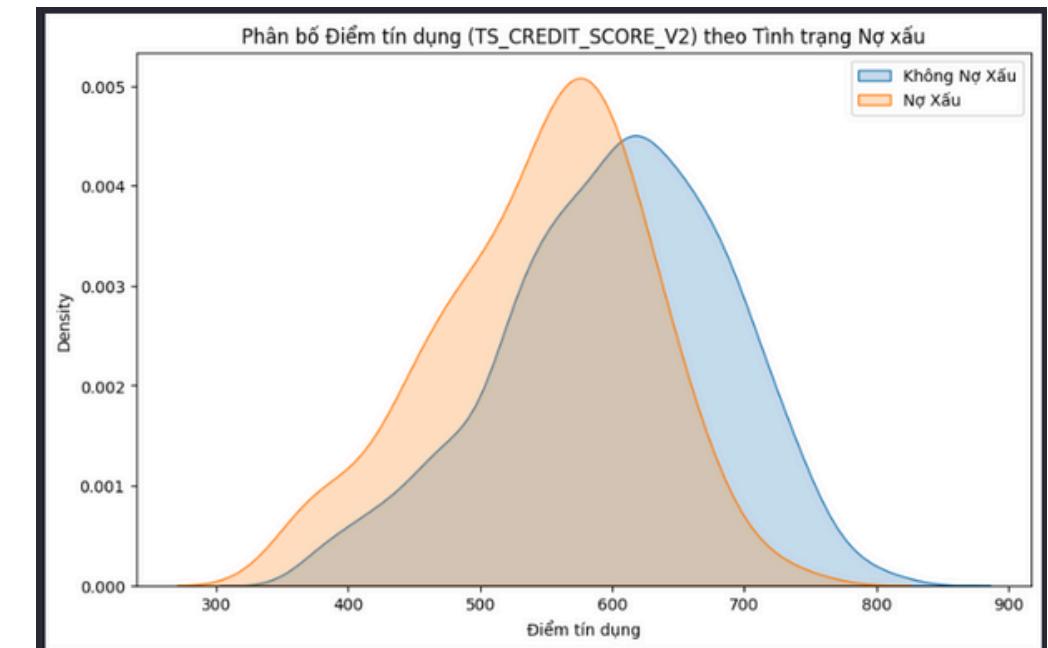
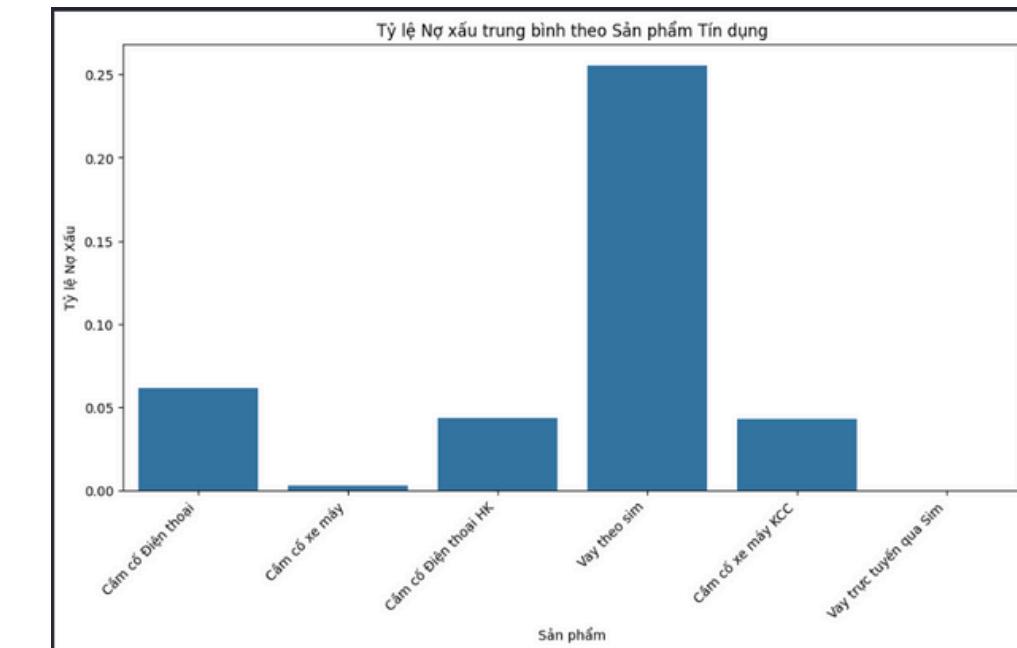
ĐÁNH GIÁ MÔ HÌNH VÀ BỘ DỮ LIỆU

MÔ HÌNH

- **Mô hình được chọn: Logistic Regression.**
- **Lý do lựa chọn:**
 - **Tính hiệu quả:** Là một mô hình mạnh mẽ và phổ biến cho bài toán phân loại nhị phân.
 - **Tốc độ nhanh:** Huấn luyện và dự đoán nhanh chóng, phù hợp để tích hợp vào hệ thống thời gian thực.
 - **Dễ diễn giải:** Chúng ta có thể phân tích các hệ số của mô hình để hiểu được yếu tố nào đang ảnh hưởng nhiều nhất đến dự đoán rủi ro. Đây là điểm cộng rất lớn khi trình bày với các bên nghiệp vụ.

BỘ DỮ LIỆU

- Tập huấn luyện (train): **1555 mẫu**
- Tập kiểm tra (test): **389 mẫu**
- Tỷ lệ nợ xấu trong tập huấn luyện: **6.75%**
- Tỷ lệ nợ xấu trong tập kiểm tra: **6.68%**



Hiệu Quả Mô Hình: "Độ Chính Xác Đến 80%"

Mô hình dự đoán rủi ro đã chứng minh hiệu quả vượt trội trong việc xác định cả khách hàng tốt.

99%

Xác định Khách hàng Tốt

Mô hình chính xác trong việc nhận diện 99% các khách hàng có khả năng thanh toán (True Negative).

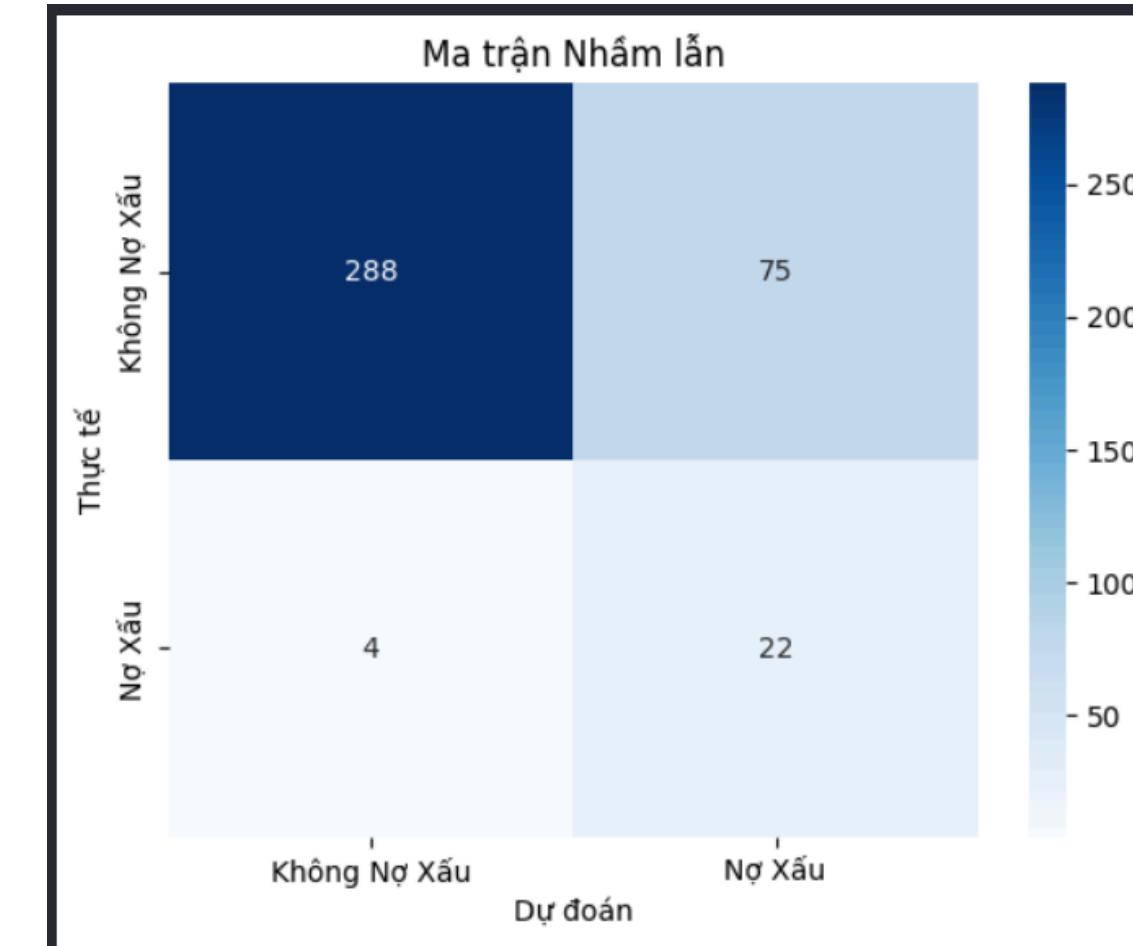
87.8%

Khả năng phân biệt tổng thể (AUC) xuất sắc

Mô hình đáng tin cậy trong việc phân loại khách hàng.

Các yếu tố dự báo quan trọng nhất:

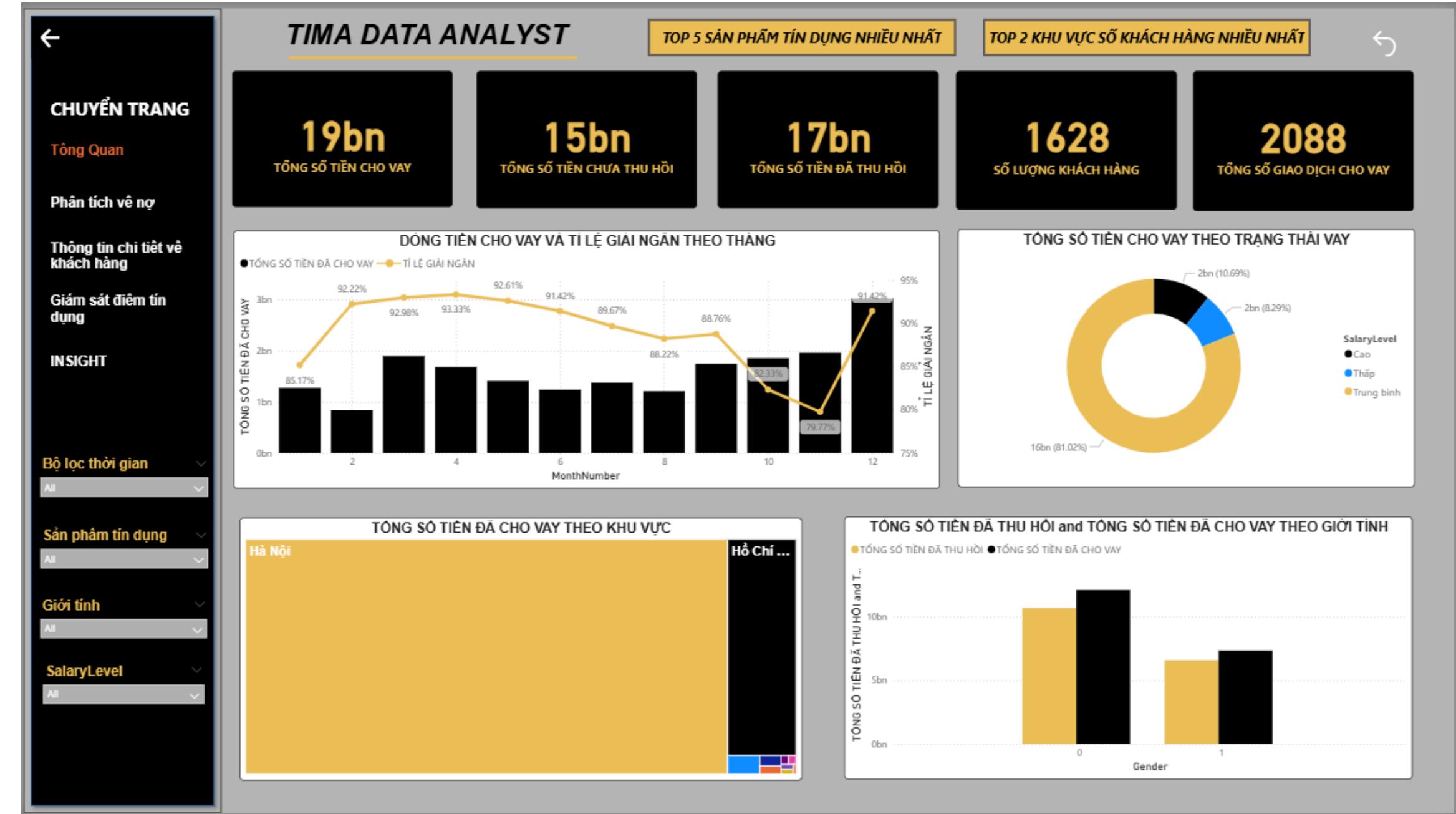
- TS_CREDIT_SCORE_V2:** Điểm tín dụng nội bộ của khách hàng.
- LongestOverdue:** Số ngày quá hạn cao nhất trong lịch sử tín dụng.
- ProductCreditName:** Đặc biệt là sản phẩm "Vay theo sim" có rủi ro cao.
- HasBadDebt:** Lịch sử có nợ xấu trước đó.



GIẢI PHÁP: CÔNG CỤ VÀ CHIẾN LƯỢC

Power BI Dashboard: Trung tâm Chỉ huy Điều hành

- Dành cho **Ban Lãnh đạo & Quản lý**: Theo dõi **Real-time KPIs** (dòng tiền, nợ xấu). **So sánh** hiệu quả theo Vùng/Sản phẩm/Team.
- Dành cho **Marketing & Sales**: Click vào "**Thông tin chi tiết về khách hàng**" để xem chân dung chi tiết, đo lường **hiệu quả chiến dịch**.
- Dành cho bộ phận **Quản lý rủi ro**: Click vào "**Giám sát điểm tín dụng**" để xem **đối tượng** cần được quan tâm.



Tích hợp Mô hình Dự đoán vào Hệ thống CRM

Mô hình này sẽ được tích hợp vào quy trình nghiệp vụ như thế nào?

- Ứng dụng thực tế:** Khi có một hồ sơ vay mới, các thông tin của khách hàng sẽ được đưa vào mô hình. Mô hình sẽ trả về một "Điểm Rủi Ro" (dưới dạng xác suất từ 0 đến 1) và một nhãn dự đoán (Rủi ro cao / Rủi ro thấp).

- Kế hoạch hành động (Luồng xử lý):**

a. **Luồng Xanh (Điểm rủi ro < 0.4):** Hồ sơ an toàn. Ưu tiên duyệt nhanh để nâng cao trải nghiệm khách hàng.

b. **Luồng Vàng (0.4 <= Điểm rủi ro < 0.7):** Cần xem xét. Chuyển cho nhân viên thẩm định theo quy trình chuẩn.

c. **Luồng Đỏ (Điểm rủi ro >= 0.7):** Rủi ro cao. Hệ thống tự động gắn cờ đỏ, gửi cảnh báo cho chuyên viên thẩm định cao cấp để kiểm tra sâu, xem xét giảm hạn mức hoặc từ chối.

--- Đánh giá Mô hình Dự báo ---					
Báo cáo phân loại:	precision	recall	f1-score	support	
0	0.99	0.80	0.88	363	
1	0.24	0.88	0.37	26	
accuracy			0.80	389	
macro avg	0.61	0.84	0.63	389	
weighted avg	0.94	0.80	0.85	389	
Điểm ROC AUC: 0.8828					

# SoTienDKVayBanDau	# Salary	# TS_CREDIT_SCORE_V2	# Actual_IsBadDebt	# Predicted_Probability	⋮ Recommendation	...
2500000.0	6000000	500.0	1	0.8435450271160719	Luồng Đỏ: Rủi ro cao - Thẩm định Kỹ / Đề xuất phương án thay thế	
10000000.0	5000000	592.0	0	0.31454459377089034	Luồng Xanh: Ưu tiên Duyệt nhanh	
10000000.0	12000000	675.0	0	0.02987875487549924	Luồng Xanh: Ưu tiên Duyệt nhanh	
10000000.0	10000000	685.0	0	0.20053941952928844	Luồng Xanh: Ưu tiên Duyệt nhanh	
5000000.0	10000000	702.0	0	0.004459733667524883	Luồng Xanh: Ưu tiên Duyệt nhanh	

ĐỀ XUẤT: Chiến lược theo Phòng ban

• Phòng Quản lý Rủi ro & Sản phẩm:

- Ngay lập tức: **Rà soát** lại chính sách của sản phẩm "Vay theo Sim". Cân nhắc **tăng yêu cầu về thẩm định** (ví dụ: xác minh nơi ở, thu nhập) hoặc **điều chỉnh lại biểu phí rủi ro** cho sản phẩm này.
- Sử dụng "**Điểm Rủi Ro**" từ mô hình để xây dựng ma trận Lãi suất - Rủi ro, đảm bảo lãi suất **tương xứng** với mức độ rủi ro của từng khách hàng.

• Phòng Marketing:

- Sử dụng chân dung "**khách hàng ít rủi ro**" (thu nhập ổn định, điểm tín dụng cao, vay các sản phẩm an toàn) để **tối ưu hóa target** quảng cáo, **giảm chi phí** cho các tệp khách hàng không hiệu quả.
- **Hạn chế** các chiến dịch quảng cáo **tập trung** vào các quận/huyện đang có **tỷ lệ nợ xấu cao**.

• Phòng Kinh doanh & Sales (Tư vấn & Thẩm định):

- Đào tạo nhân viên sử dụng Power BI Dashboard để **nắm bắt sản phẩm hiệu quả** và "**Điểm Rủi Ro**" như một công cụ hỗ trợ.
- Với các **hồ sơ "Luồng Đỏ"**, nhân viên tư vấn có thể chủ động **đề xuất** khách hàng giảm số tiền vay hoặc chuyển sang một sản phẩm khác phù hợp hơn với khả năng tài chính của họ, **tăng tỷ lệ duyệt thành công** và **giảm rủi ro**.

Q&A

Thảo luận & Các bước tiếp theo

KẾT THÚC

