

# PHÂN TÍCH VẤN ĐỀ RỜI BỎ & ĐỀ XUẤT CHIẾN LƯỢC GIỮ CHÂN KHÁCH HÀNG TẠI NEOBANK EUROPE



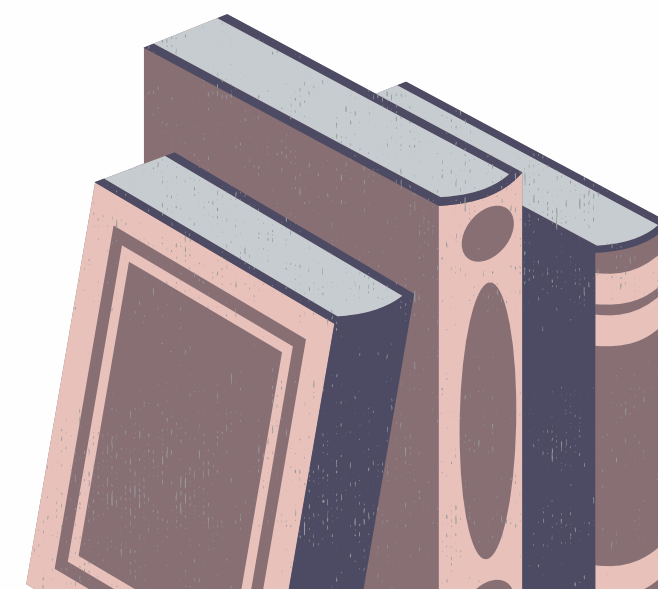


# Bối Cảnh

NeoBank Europe là một ngân hàng số có trụ sở tại Luxembourg, ra đời với tham vọng chinh phục thị trường châu Âu.

Trong chiến lược ra mắt, NeoBank đã quyết định tập trung vào ba thị trường kinh tế lớn và năng động nhất khu vực đồng Euro: Pháp, Tây Ban Nha và Đức. Họ tung ra một chiến dịch marketing đồng nhất trên cả ba quốc gia, kỳ vọng vào một mô hình kinh doanh có thể nhân rộng dễ dàng.

Tuy nhiên, sau hai năm hoạt động, dù tổng số lượng người dùng vẫn tăng, báo cáo kinh doanh gần đây đã chỉ ra một xu hướng đáng lo ngại, phá vỡ kỳ vọng ban đầu của ban lãnh đạo: hiệu suất hoạt động giữa ba thị trường cốt lõi đang có xu hướng giảm, cụ thể là tỷ lệ khách hàng rời bỏ đang khá cao, gây ảnh hưởng lớn đến lợi nhuận chung.



# Mục tiêu và kết quả kỳ vọng



- Xác định các yếu tố quan trọng ảnh hưởng đến việc khách hàng rời bỏ ngân hàng.
- Phân tích đặc điểm nhân khẩu học, hành vi và tài chính của nhóm khách hàng churn
- Đưa ra insight hỗ trợ ngân hàng xây dựng chiến lược giữ chân khách hàng.



- Xác định được nhóm khách hàng có nguy cơ rời bỏ cao (theo tuổi, hành vi, tài chính).
- Báo cáo insight: yếu tố nào ảnh hưởng mạnh nhất đến churn.
- Đề xuất chiến lược giữ chân khách hàng giá trị cao, giảm churn rate.





# Tổng quan dữ liệu

- Bộ dữ liệu được sử dụng trong dự án này là Bank Customer Churn Prediction, lấy từ nền tảng Kaggle.
- Dữ liệu bao gồm 10,000 bản ghi về khách hàng với các thông tin nhân khẩu học và tài chính

## Các biến chính

CreditScore	Geography	Gender
Age	Tenure	Balance
NumOfProducts	HasCrCard	IsActiveMember
EstimatedSalary	Exited	Complain
Satisfaction Score	Card Type	Points Earned





# Quy trình phân tích

Làm sạch và tiền xử lý dữ liệu

Phân tích mô tả (EDA)

Phân tích chuyên sâu – logistic regression để tìm  
yếu tố ảnh hưởng và dự đoán

Báo cáo tổng hợp insight và khuyến nghị chiến  
lược giữ chân khách hàng.



# Làm sạch dữ liệu

Cột CreditScore: 15 outliers. (Giới hạn: [383.00, 919.00])

Cột Age: 359 outliers. (Giới hạn: [14.00, 62.00])

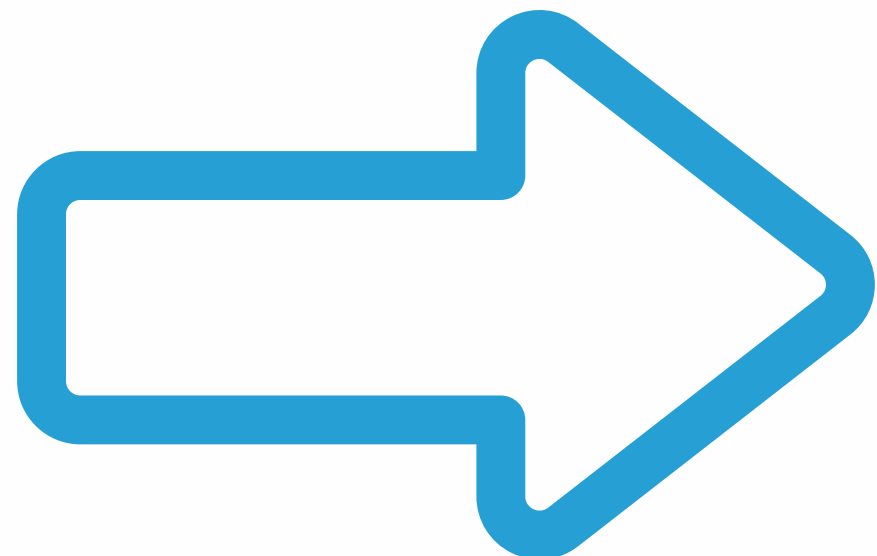
Cột Tenure: 0 outliers. (Giới hạn: [-3.00, 13.00])

Cột Balance: 0 outliers. (Giới hạn: [-191466.36, 319110.60])

Cột EstimatedSalary: 0 outliers. (Giới hạn: [-96577.10, 296967.45])

Cột Point Earned: 0 outliers. (Giới hạn: [-176.50, 1387.50])

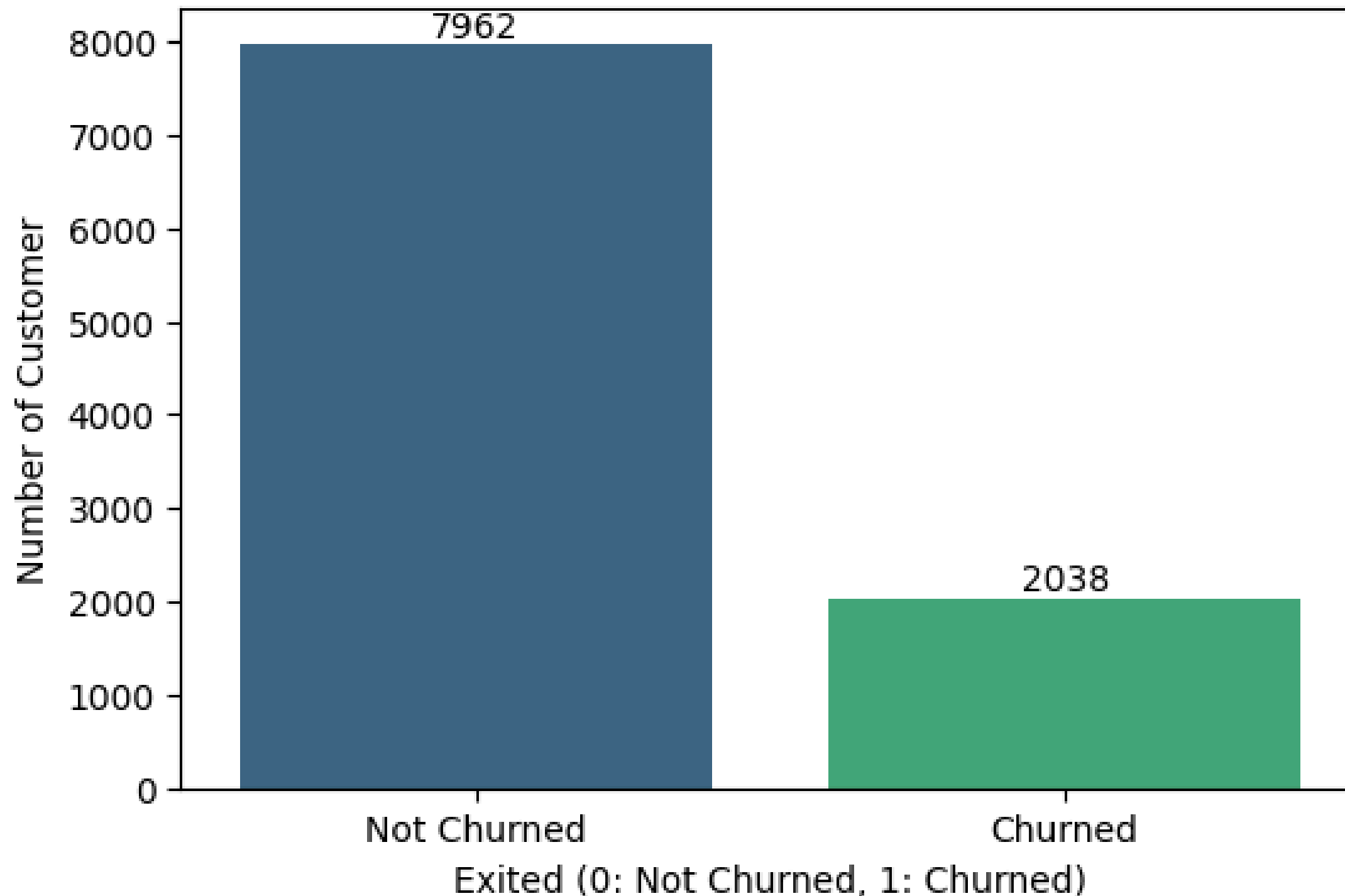
Tổng số outlier (trên tất cả các cột): 374



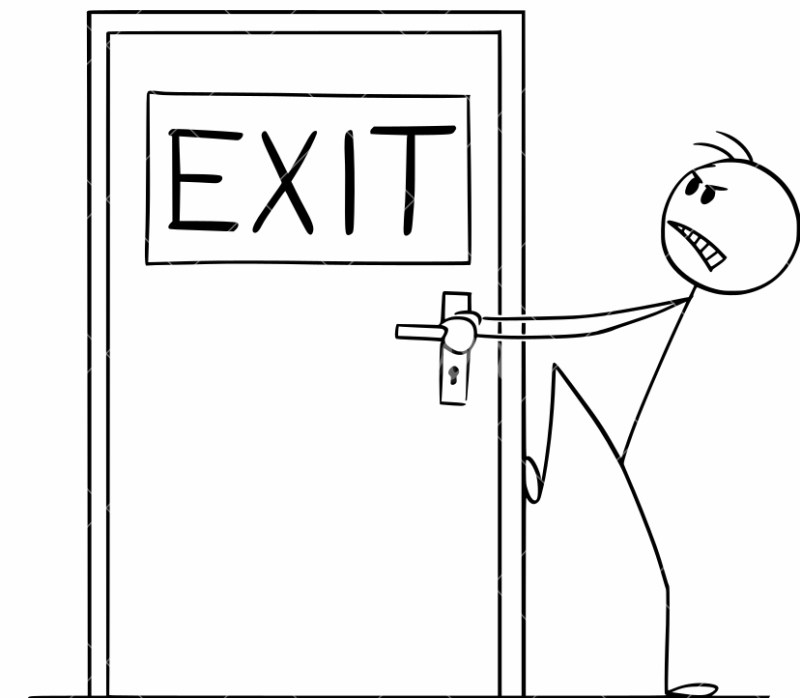
Có 374 giá trị ngoại lai, tuy nhiên các giá trị này vẫn là số ít và vẫn còn có thể chấp nhận được để tạo tính thực tế cho kết quả phân tích. Vì thế, không loại trừ các quan sát này ra khỏi mô hình.

# Phân tích khám phá

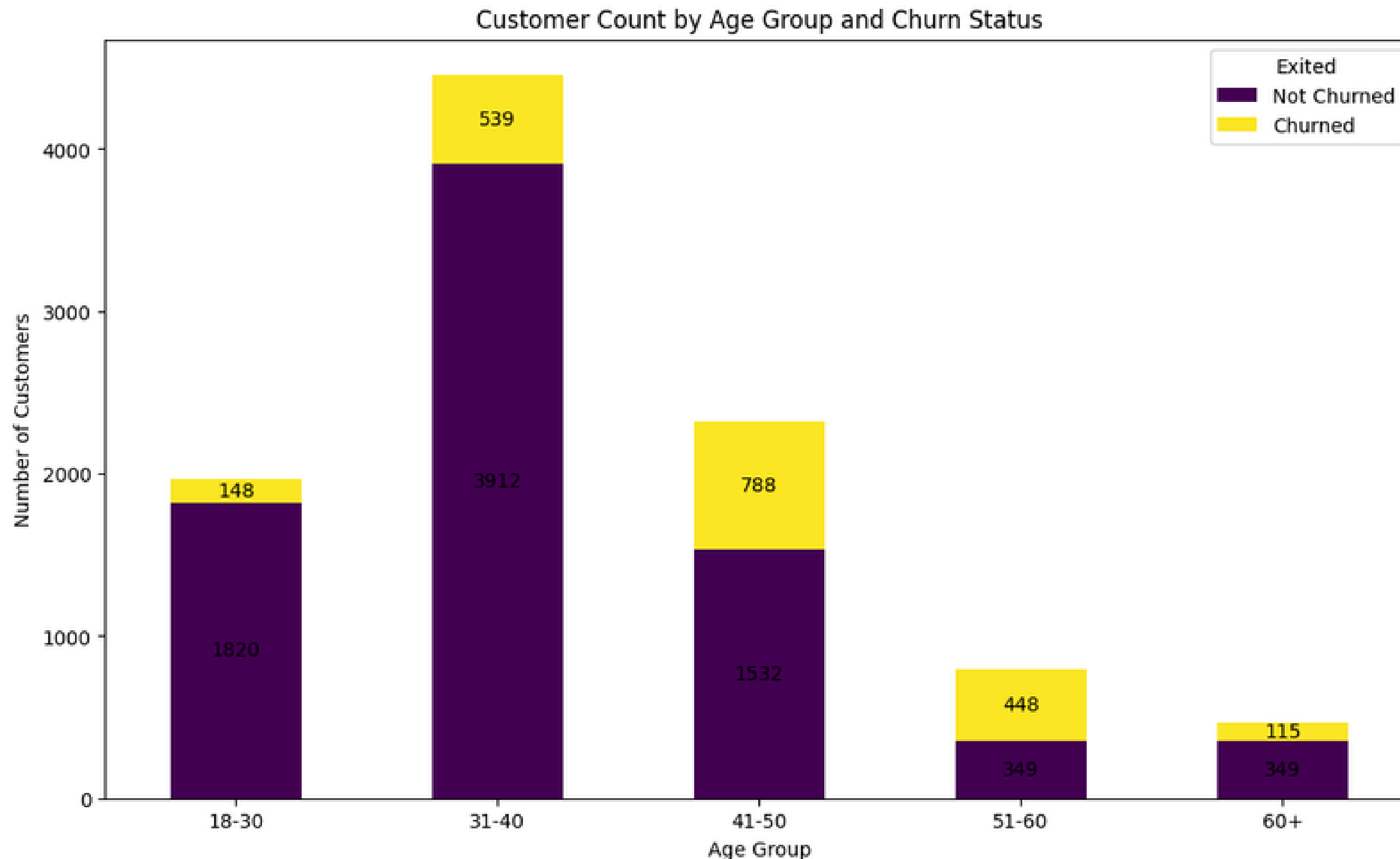
Count of Churned vs Non-Churned Customers



Số lượng khách hàng không rời đi là 7962, trong khi số lượng khách hàng rời đi là 2038, chiếm khoảng 20.4%.



# Churn theo Age group

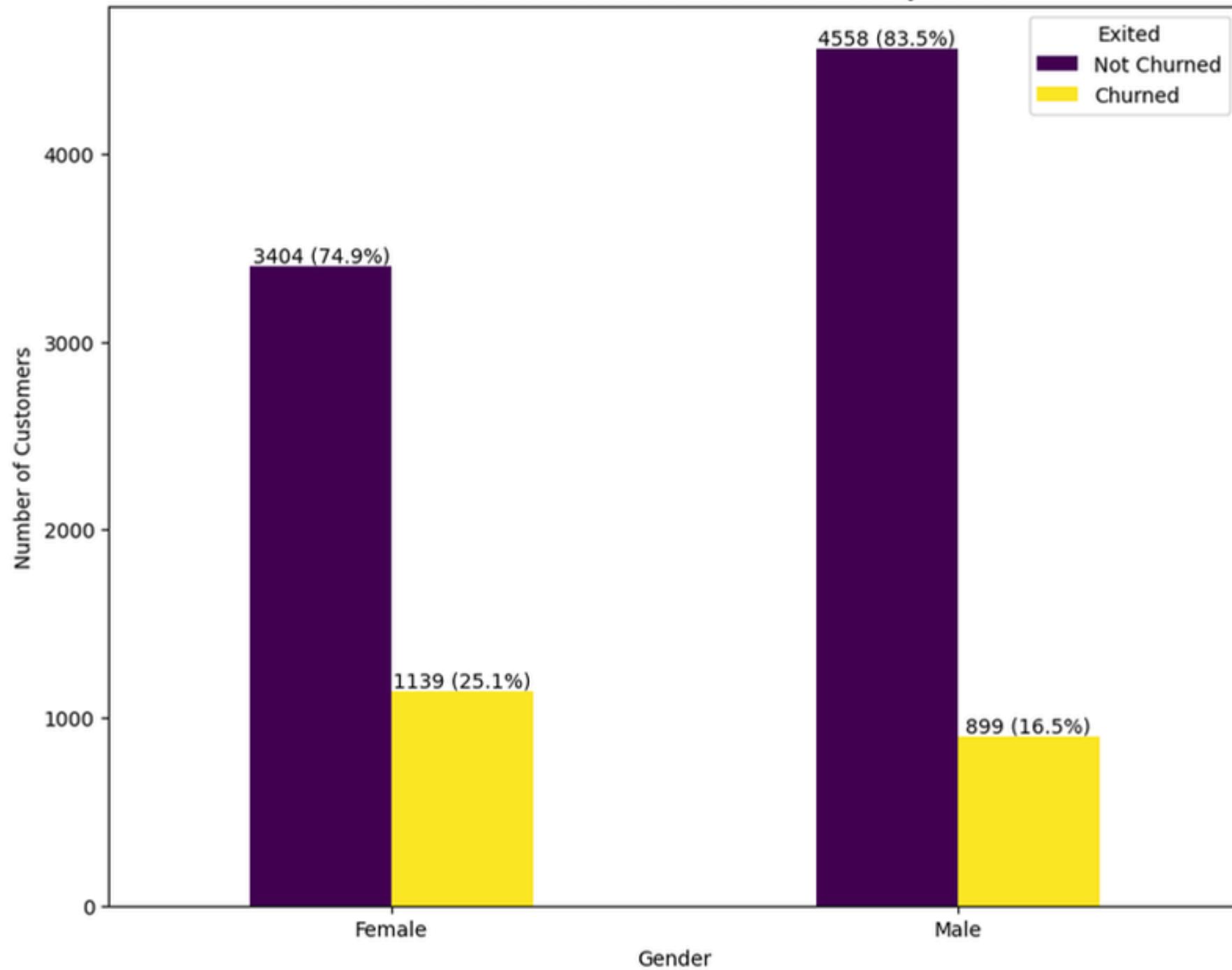


- Nhóm tuổi 31-40 có số lượng khách hàng đông nhất, cả trong nhóm churned và non-churned.
- Nhóm tuổi 41-50 có số lượng khách hàng churned cao nhất, mặc dù tổng số khách hàng ít hơn nhóm 31-40.
- Nhóm tuổi 51-60 có tỷ lệ churned trên tổng số khách hàng trong nhóm cao nhất.
- Nhóm tuổi trẻ hơn (18-30) có số lượng khách hàng churned ít nhất.

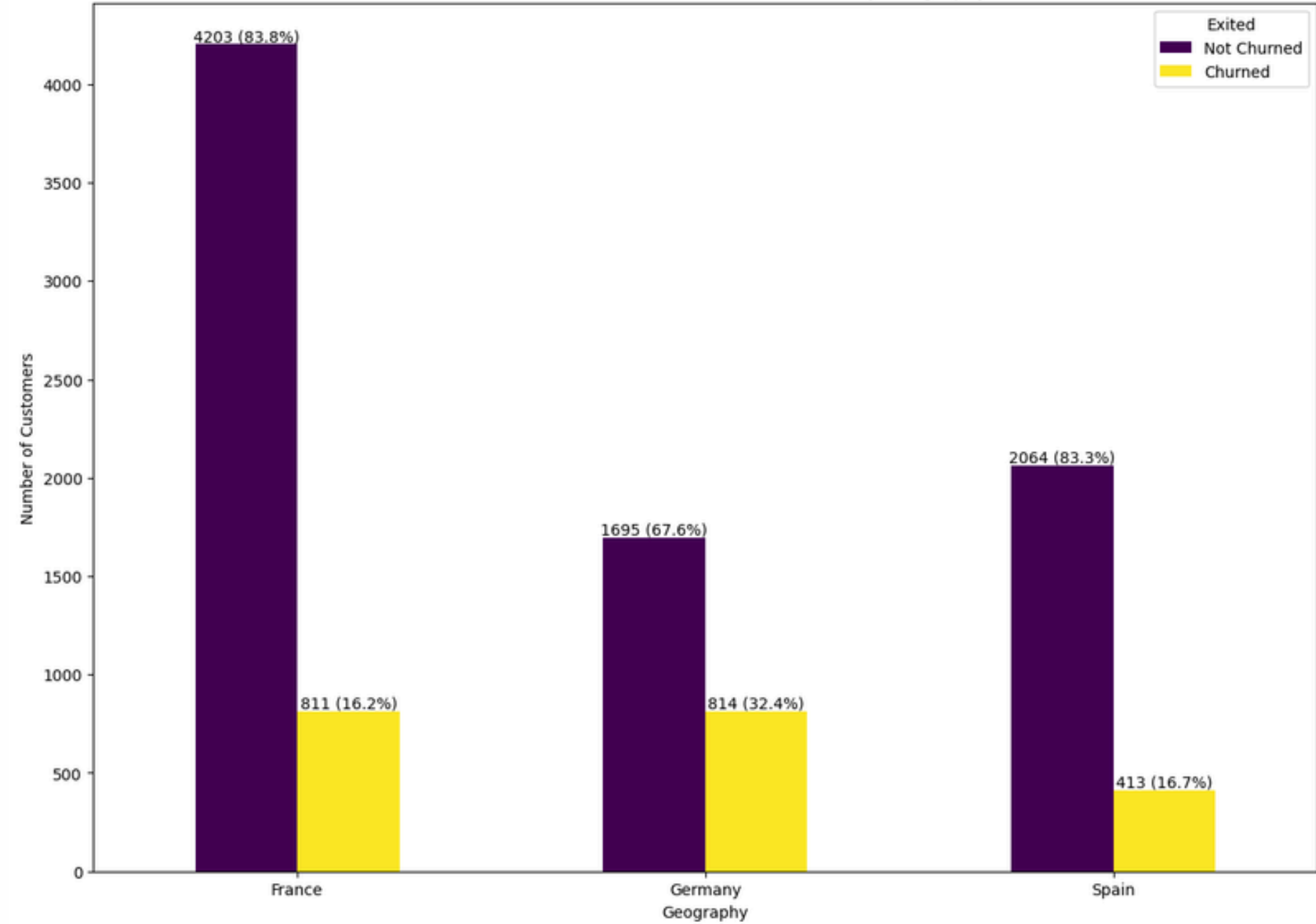


# Churn theo Gender & Geography

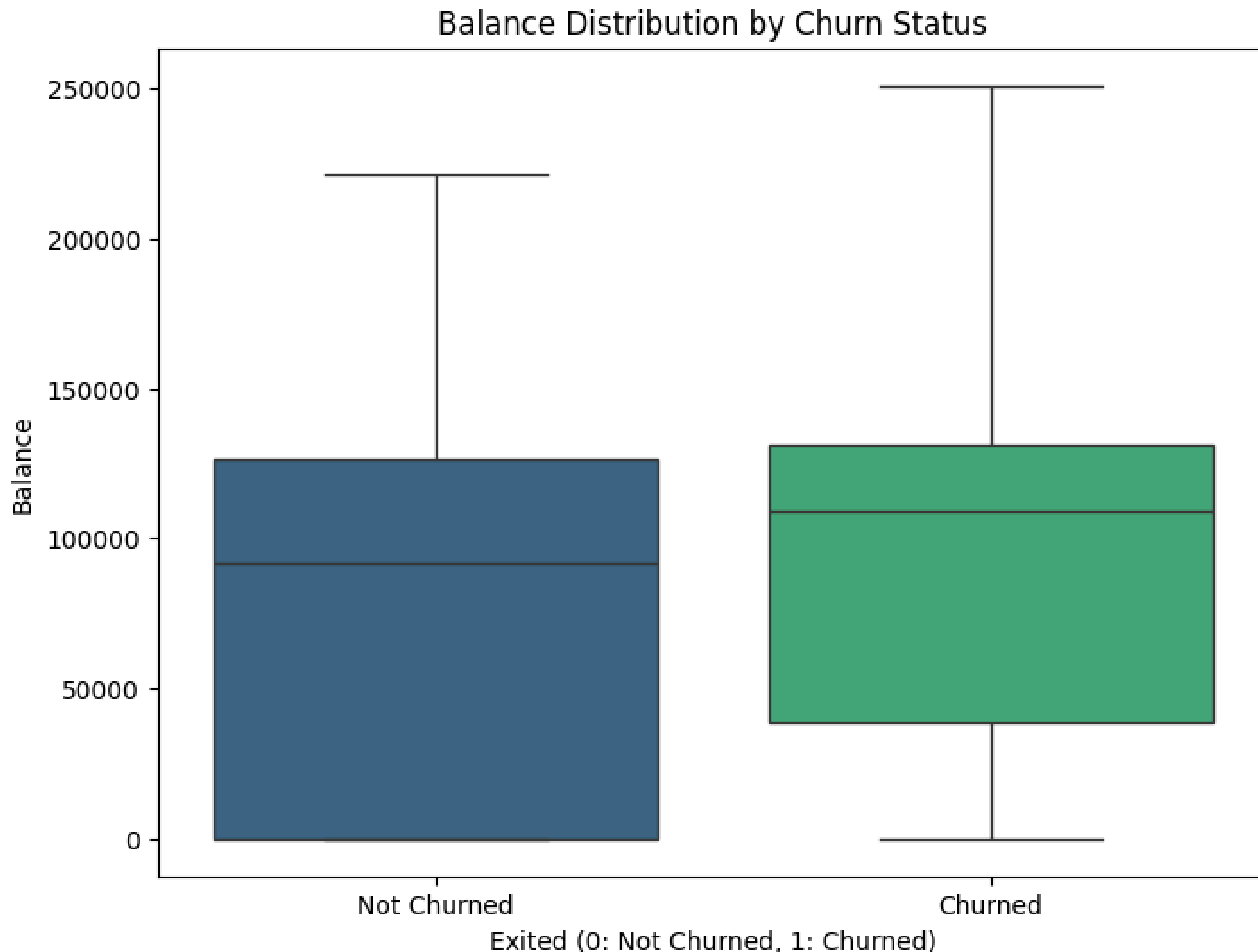
Count of Churned vs Non-Churned Customers by Gender



Count of Churned vs Non-Churned Customers by Geography



# Churn theo Balance

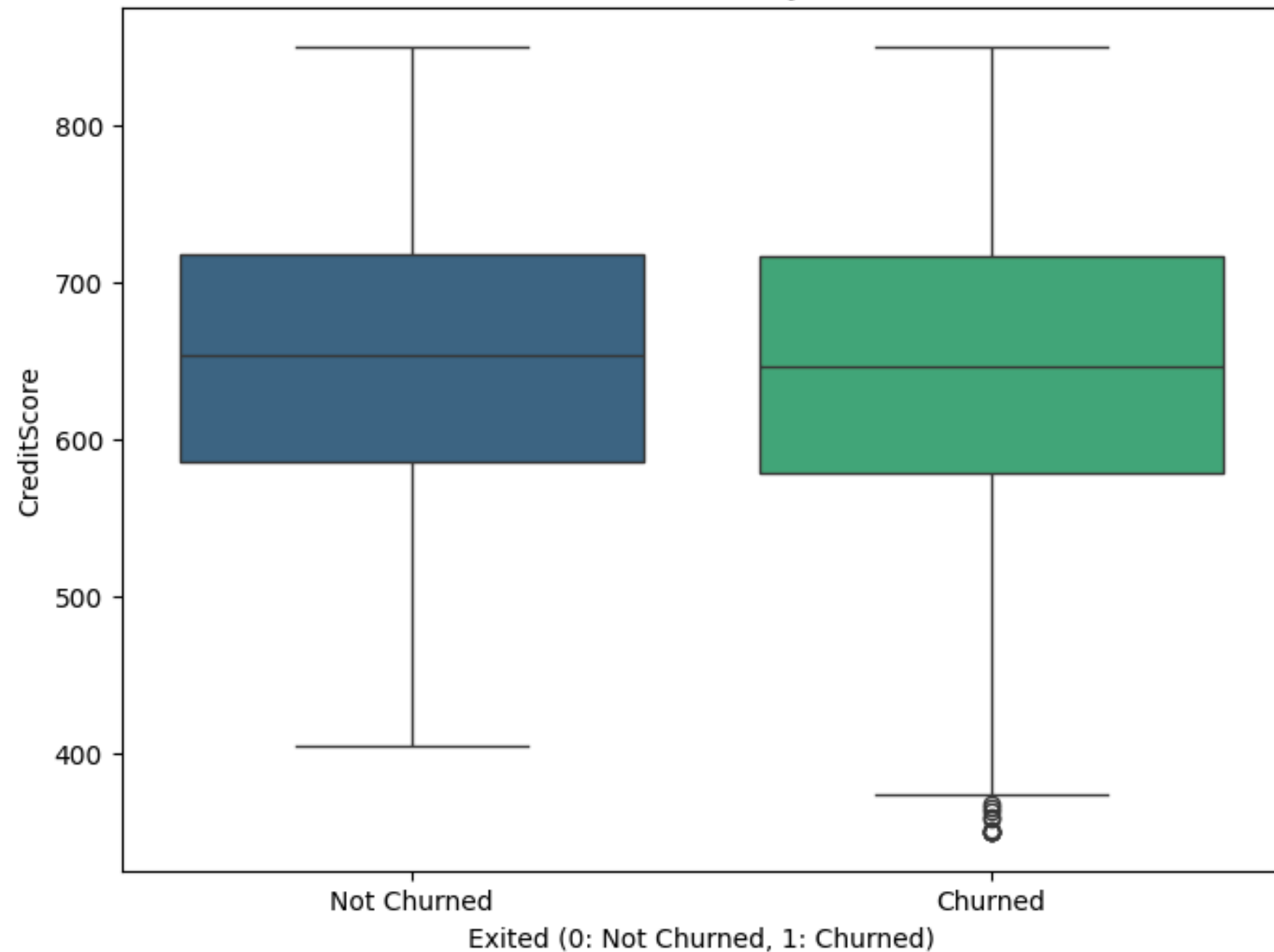


## Nhận xét:

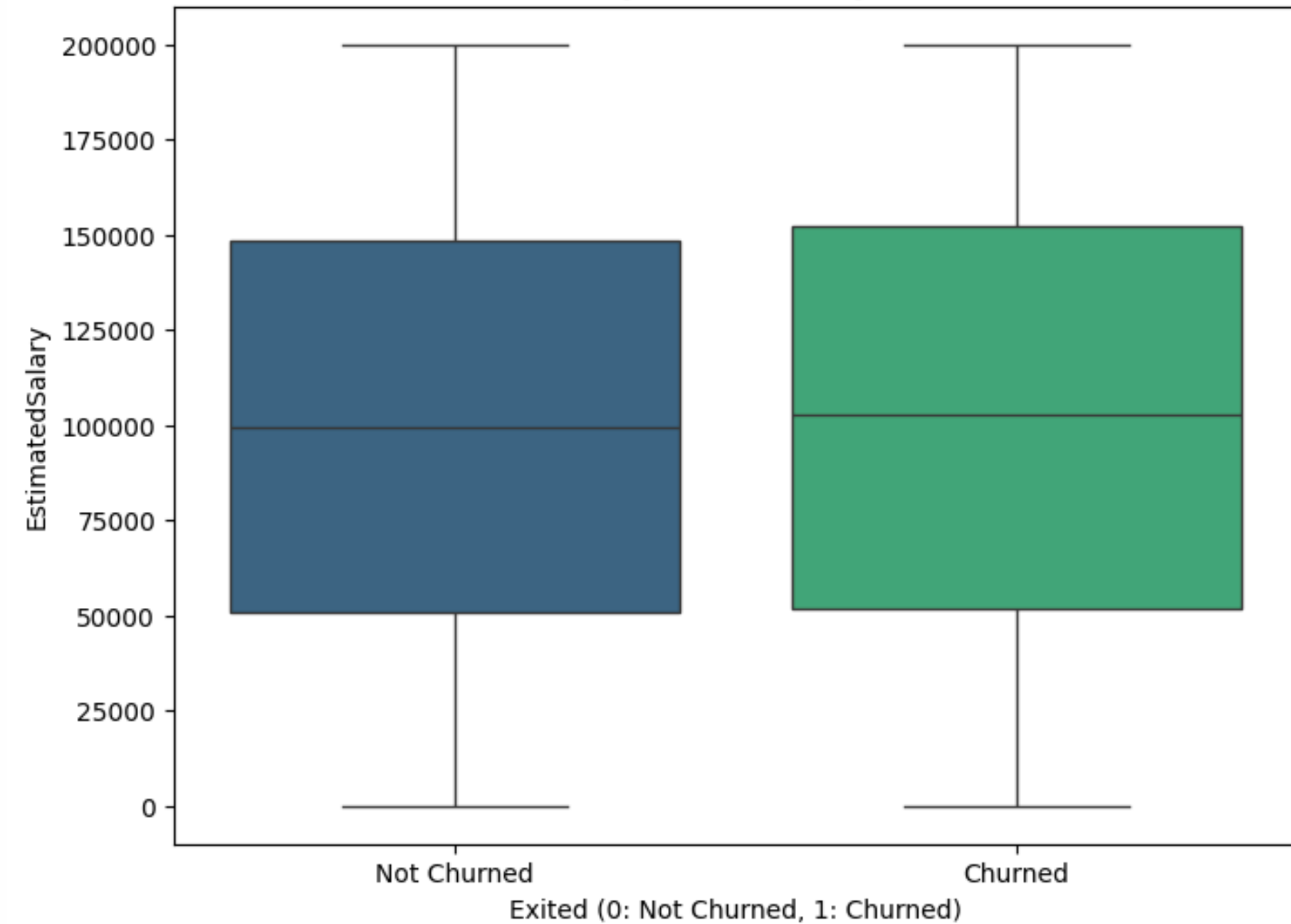
- Khách hàng churned có xu hướng có số dư tài khoản trung bình cao hơn, cho thấy khách hàng có số dư cao hơn có khả năng rời đi cao hơn.
- Với nhóm non-churned, có 1 số lượng lớn nhóm khách hàng có số dư bằng 0 hoặc rất thấp, cho thấy nhóm khách hàng này khá trung thành.

# Churn theo CreditScore và EstimatedSalary

CreditScore Distribution by Churn Status



EstimatedSalary Distribution by Churn Status



Phân bố của CreditScore và EstimatedSalary giữa hai nhóm churned và non-churned khá tương đồng, cho thấy hai biến này có thể không phải là yếu tố dự đoán churn mạnh mẽ.

# Churn theo Number of Products & Active Member

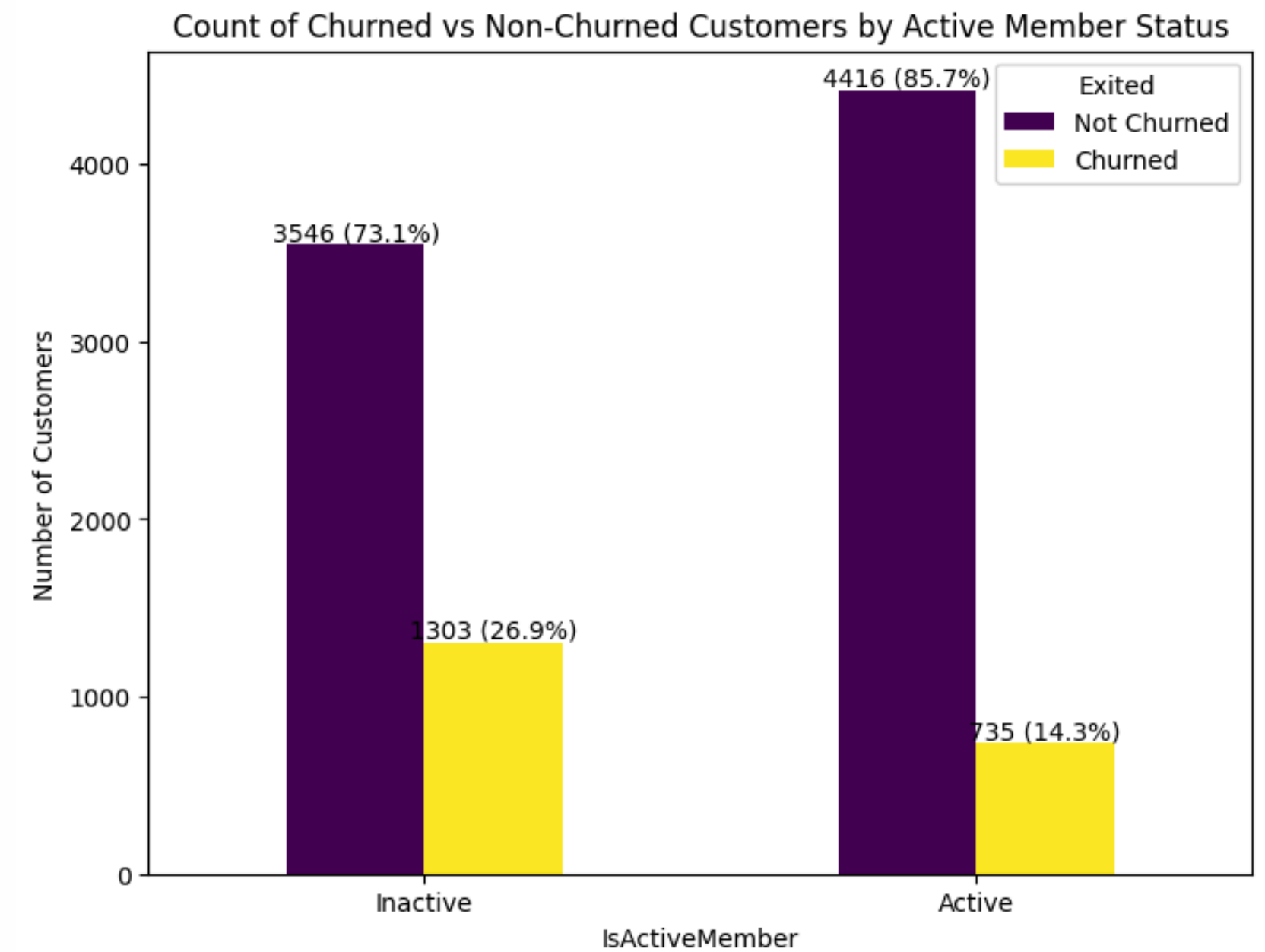
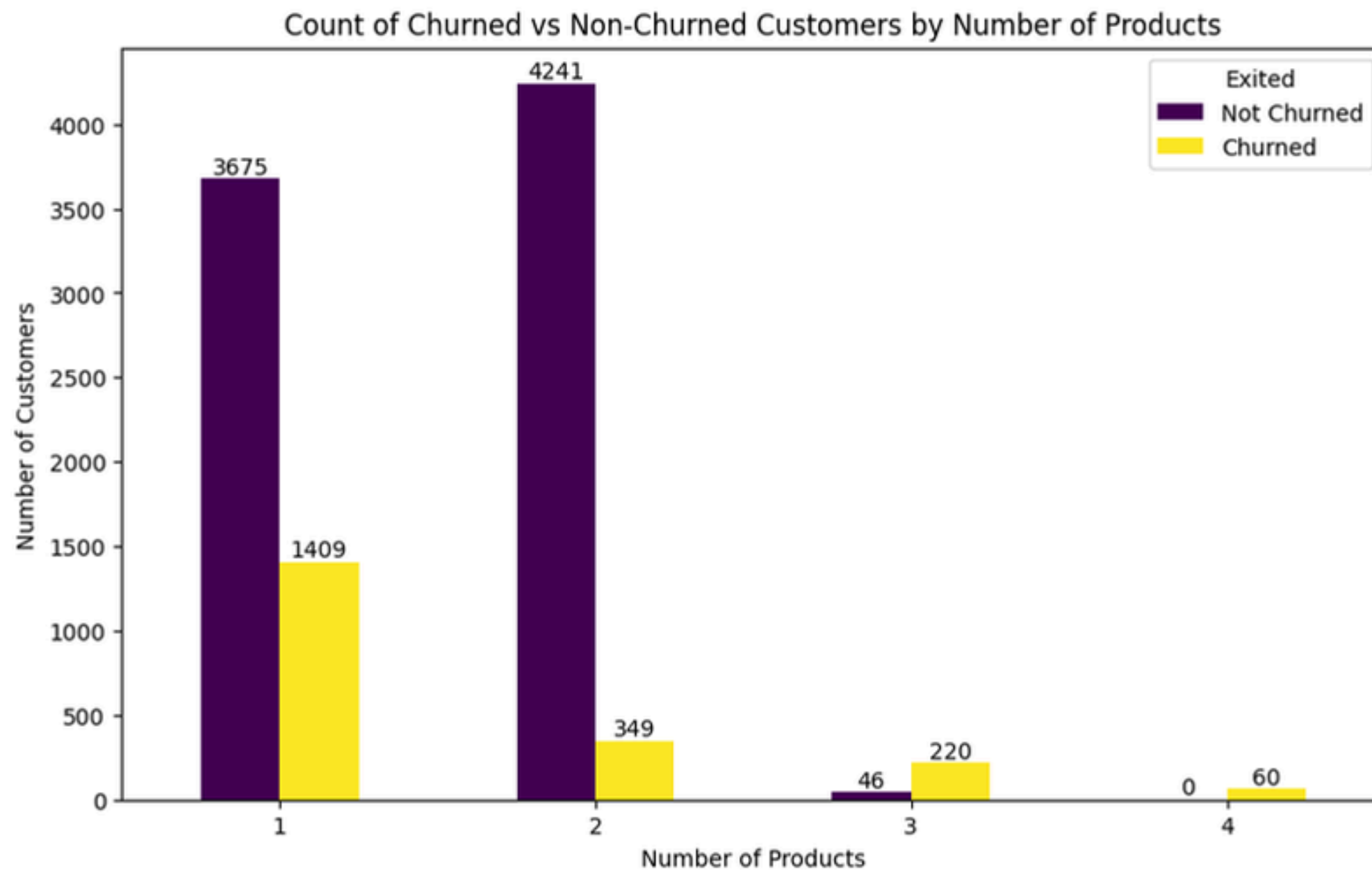
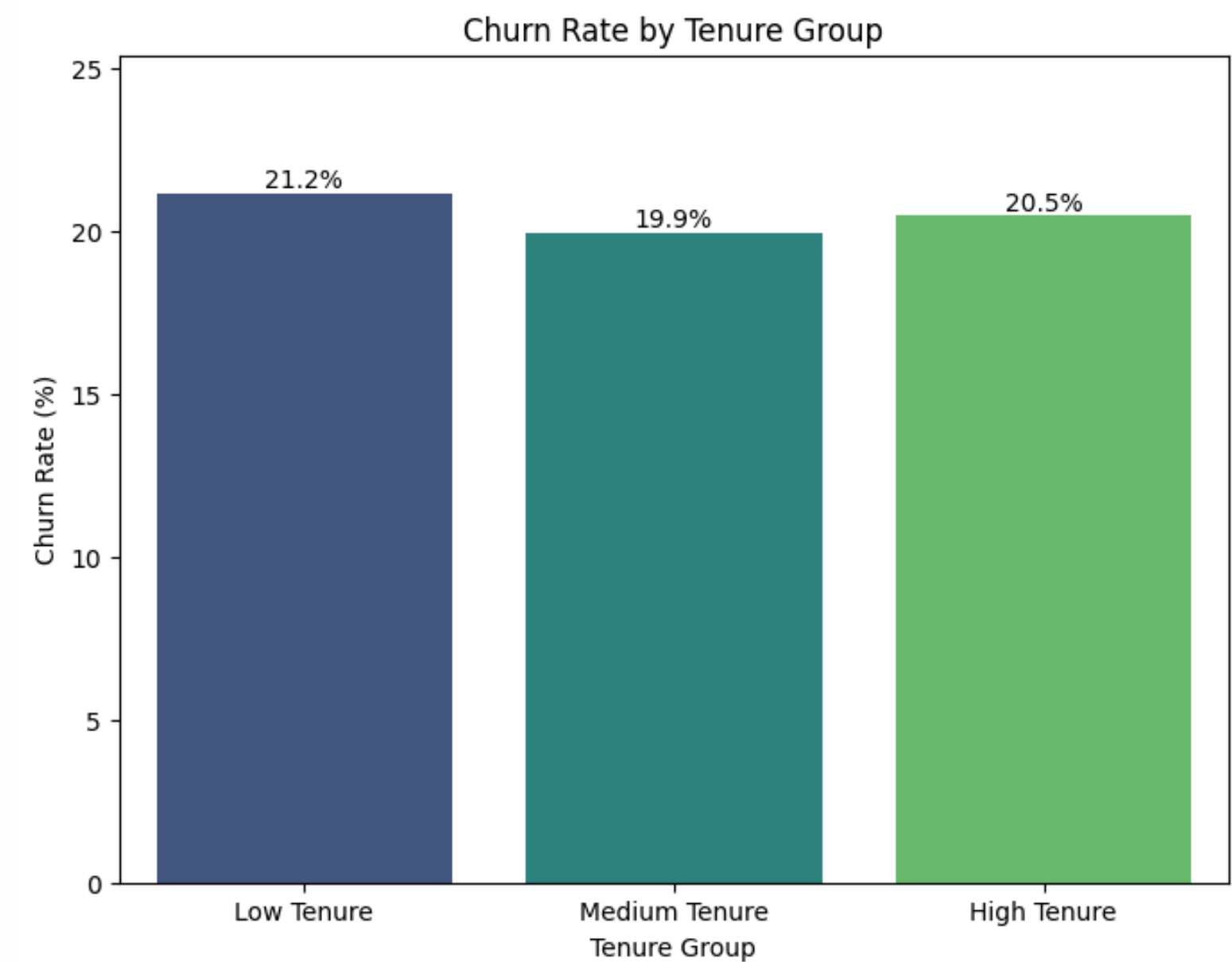
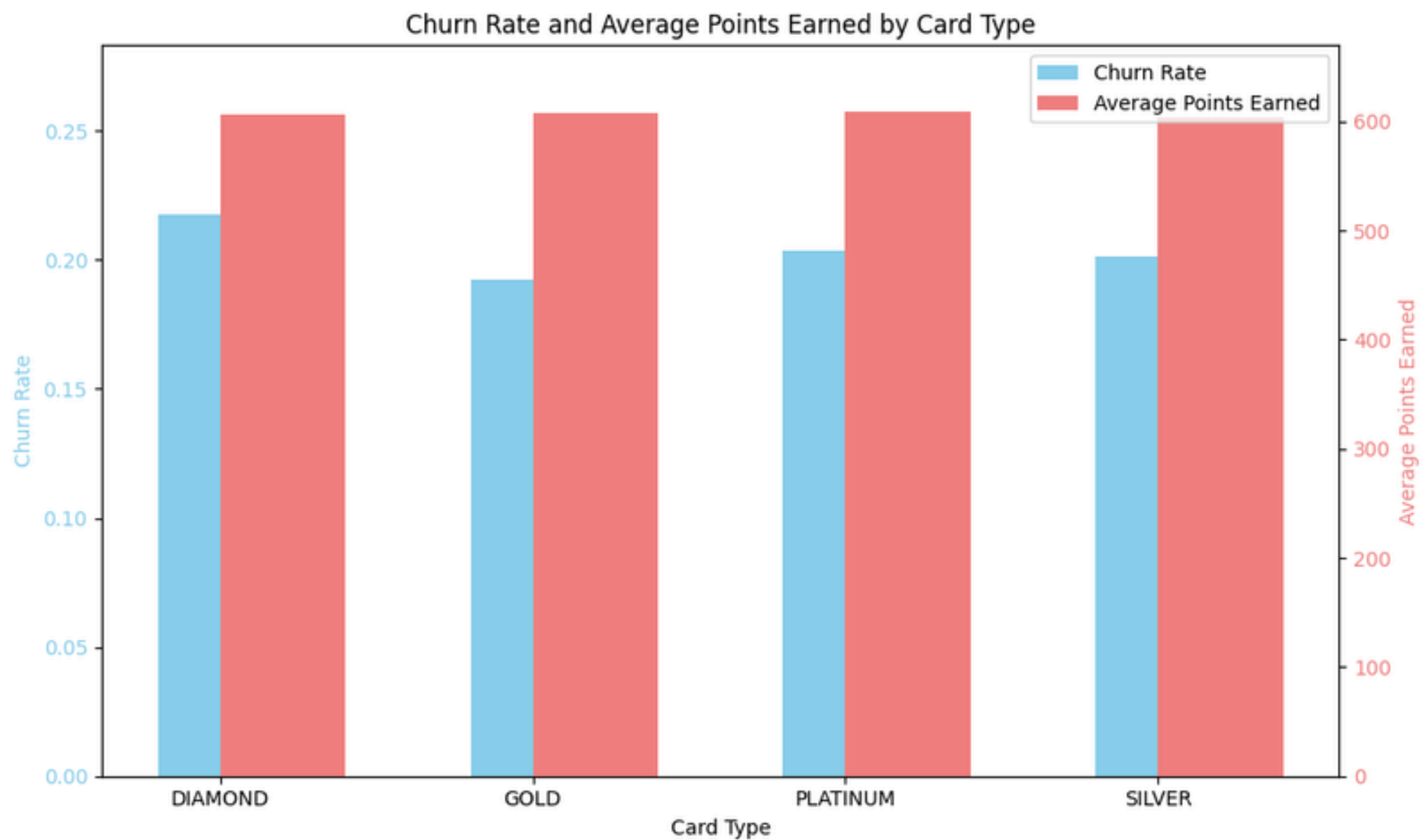


Chart 1: Số lượng khách hàng rời đi tập trung chủ yếu ở nhóm có 1 sản phẩm (1409). Tuy nhiên, nhóm có 3 hoặc 4 sản phẩm có tỷ lệ churn cực kỳ cao (gần 100%), mặc dù số lượng khách hàng trong nhóm này ít.

Chart 2: Số lượng khách hàng không hoạt động rời đi (1303) nhiều hơn đáng kể so với khách hàng hoạt động rời đi (735)

# Churn theo Number of Products & Active Member

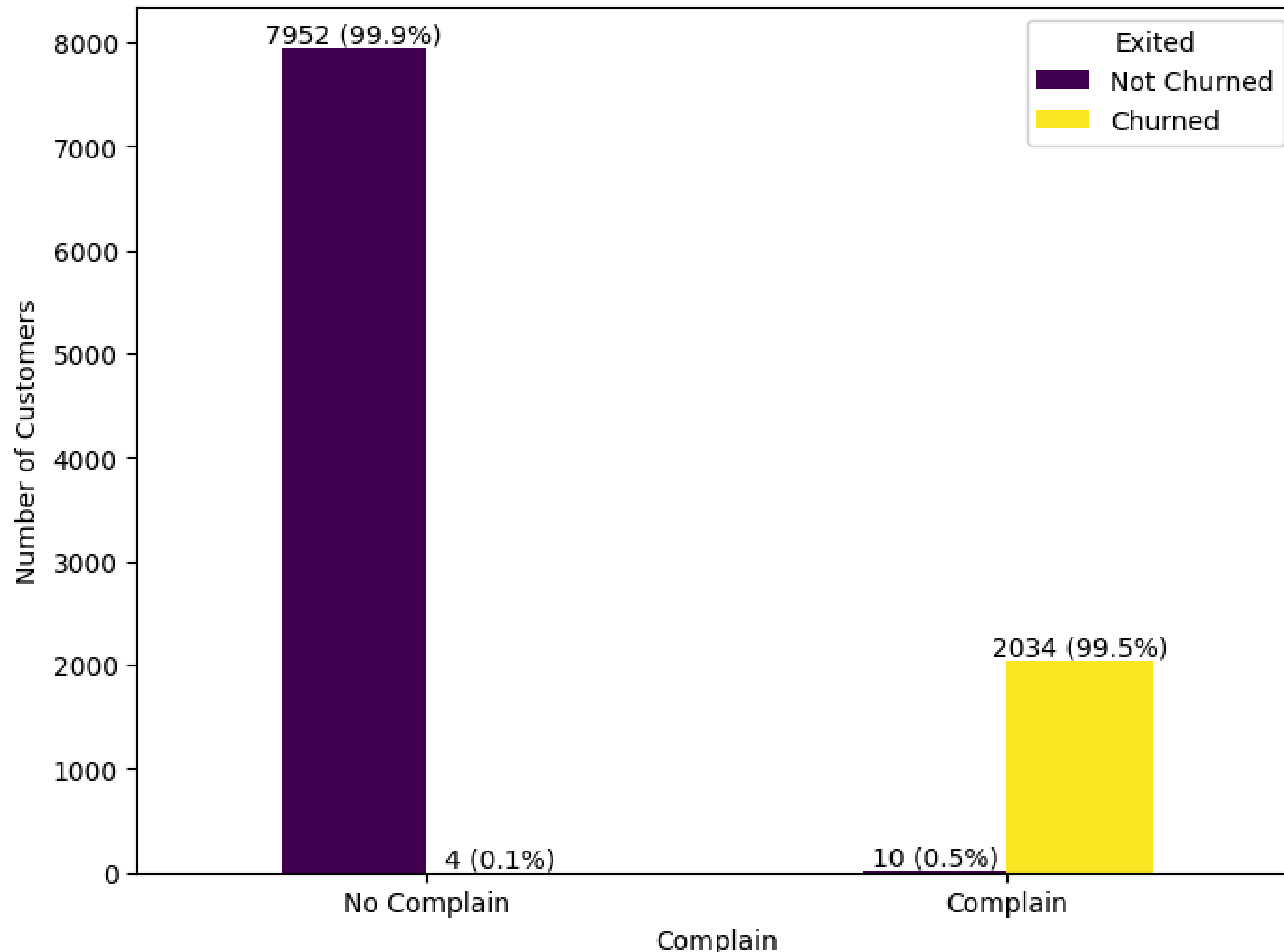


Điểm trung bình tích lũy cũng tương đồng giữa các loại thẻ  
Không có mối liên hệ mạnh mẽ giữa card type và Churn rate  
Tỷ lệ churn khá đồng đều giữa các nhóm thời gian gắn bó



# Churn theo Complain

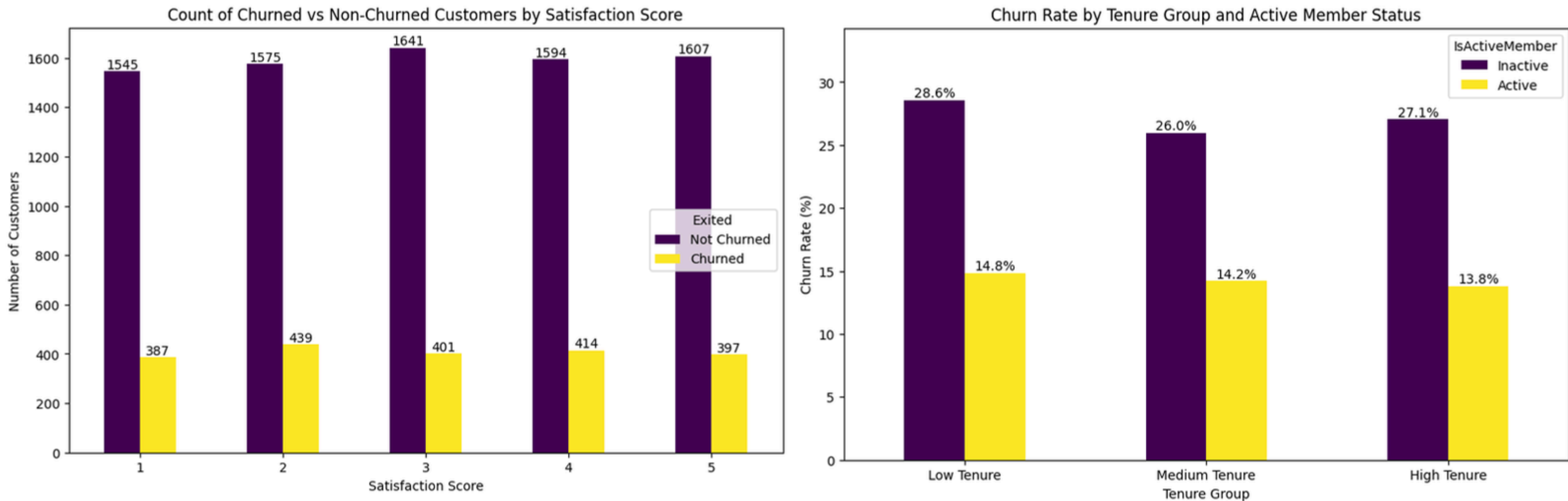
Count of Churned vs Non-Churned Customers by Complain Status



## Nhận xét:

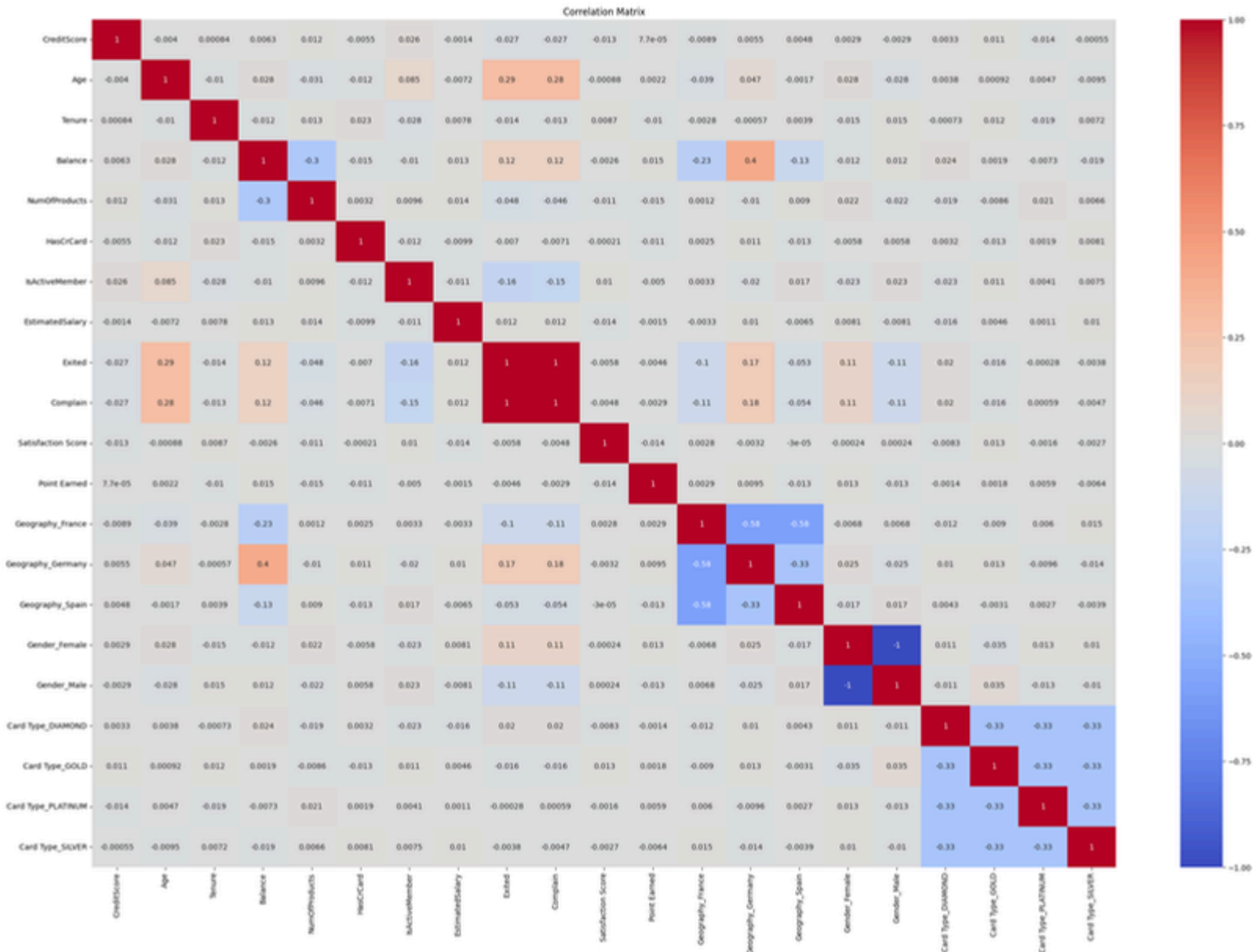
- Biểu đồ này cho thấy một mối quan hệ cực kỳ mạnh mẽ: Gần như tất cả khách hàng có khiếu nại (Complain = 1) đều rời đi (2034 churned vs 10 non-churned).
- Ngược lại, trong nhóm không khiếu nại (Complain = 0), chỉ có rất ít khách hàng rời đi (4 churned vs 7952 non-churned), cho thấy biến 'Complain' là một yếu tố dự báo churn cực kỳ mạnh.

# Churn theo Satisfaction score & Churn theo Tenure group và Active member



- Khách hàng không hoạt động có tỷ lệ churn cao hơn đáng kể ở tất cả các nhóm thời gian gắn bó
- Đối với cả nhóm khách hàng hoạt động và không hoạt động, tỷ lệ churn đều khá đồng đều giữa các nhóm tenure

# Hồi quy Logistic



## Ma trận tương quan

- Biến 'Exited' có tương quan dương mạnh nhất với biến 'Complain'. Điều này có nghĩa là khách hàng có khiếu nại thì khả năng rời đi rất cao.
- Biến 'Exited' có tương quan dương với 'Age'. Điều này cho thấy khách hàng lớn tuổi hơn có xu hướng rời đi nhiều hơn.
- Biến 'Exited' có tương quan âm nhỏ với 'IsActiveMember'. Khách hàng hoạt động tích cực có xu hướng ít rời đi hơn.
- Các biến khác như CreditScore, Tenure, HasCrCard, EstimatedSalary, Card Type, Point Earned và các biến địa lý/giới tính có tương quan tương đối yếu hoặc rất yếu với 'Exited'.



# Hồi quy Logistic

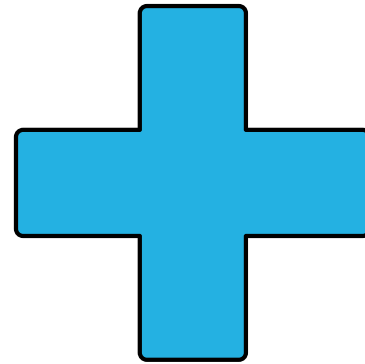
## Logit Regression Results

=====						
Dep. Variable:	Exited	No. Observations:	8000			
Model:	Logit	Df Residuals:	7982			
Method:	MLE	Df Model:	17			
Date:	Fri, 10 Oct 2025	Pseudo R-squ.:	0.9848			
Time:	07:52:00	Log-Likelihood:	-61.410			
converged:	True	LL-Null:	-4044.5			
Covariance Type:	nonrobust	LLR p-value:	0.000			
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
const	-7.6923	1.335	-5.763	0.000	-10.309	-5.076
CreditScore	0.1055	0.313	0.337	0.736	-0.508	0.719
Age	1.1768	0.322	3.659	0.000	0.546	1.807
Tenure	-0.0811	0.309	-0.262	0.793	-0.688	0.525
Balance	0.0166	0.389	0.043	0.966	-0.746	0.779
NumOfProducts	-0.1091	0.216	-0.506	0.613	-0.532	0.313
HasCrCard	-0.0778	0.678	-0.115	0.909	-1.406	1.250
IsActiveMember	-1.7947	0.721	-2.491	0.013	-3.207	-0.382
EstimatedSalary	0.0500	0.300	0.167	0.868	-0.537	0.637
Complain	14.8266	1.190	12.460	0.000	12.494	17.159
Satisfaction Score	-0.3438	0.318	-1.082	0.279	-0.967	0.279
Point Earned	-0.5942	0.342	-1.740	0.082	-1.264	0.075
Geography_France	0.0202	0.905	0.022	0.982	-1.753	1.793
Geography_Germany	-0.7690	0.912	-0.844	0.399	-2.556	1.018
Gender_Male	0.2866	0.641	0.447	0.655	-0.970	1.544
Card Type_DIAMOND	-0.4461	0.898	-0.497	0.619	-2.206	1.314
Card Type_GOLD	-0.6953	0.913	-0.762	0.446	-2.485	1.094
Card Type_PLATINUM	-0.7042	0.912	-0.772	0.440	-2.491	1.083
=====						

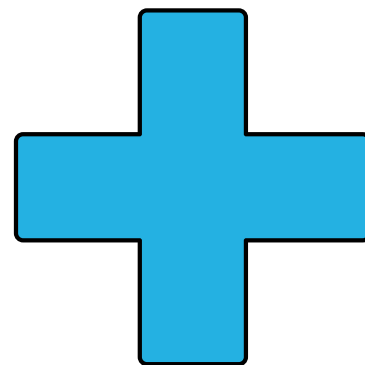
- Age: Hệ số dương (1.1768) và P-value rất nhỏ (0.000). Điều này có nghĩa là khi tuổi tăng lên (sau khi chuẩn hóa), khả năng khách hàng rời đi tăng lên đáng kể.
- IsActiveMember: Hệ số âm (-1.7947) và P-value nhỏ (0.013). Điều này có nghĩa là khách hàng là thành viên hoạt động tích cực thì khả năng rời đi giảm đi đáng kể.
- Complain: Hệ số dương rất lớn (14.8266) và P-value rất nhỏ (0.000). Điều này cho thấy khách hàng có khiếu nại thì khả năng rời đi là cực kỳ cao. Đây là yếu tố dự báo mạnh nhất trong mô hình.

# Dự báo

Độ chính xác



Báo cáo phân loại



Ma trận nhầm lẫn



# Dự báo

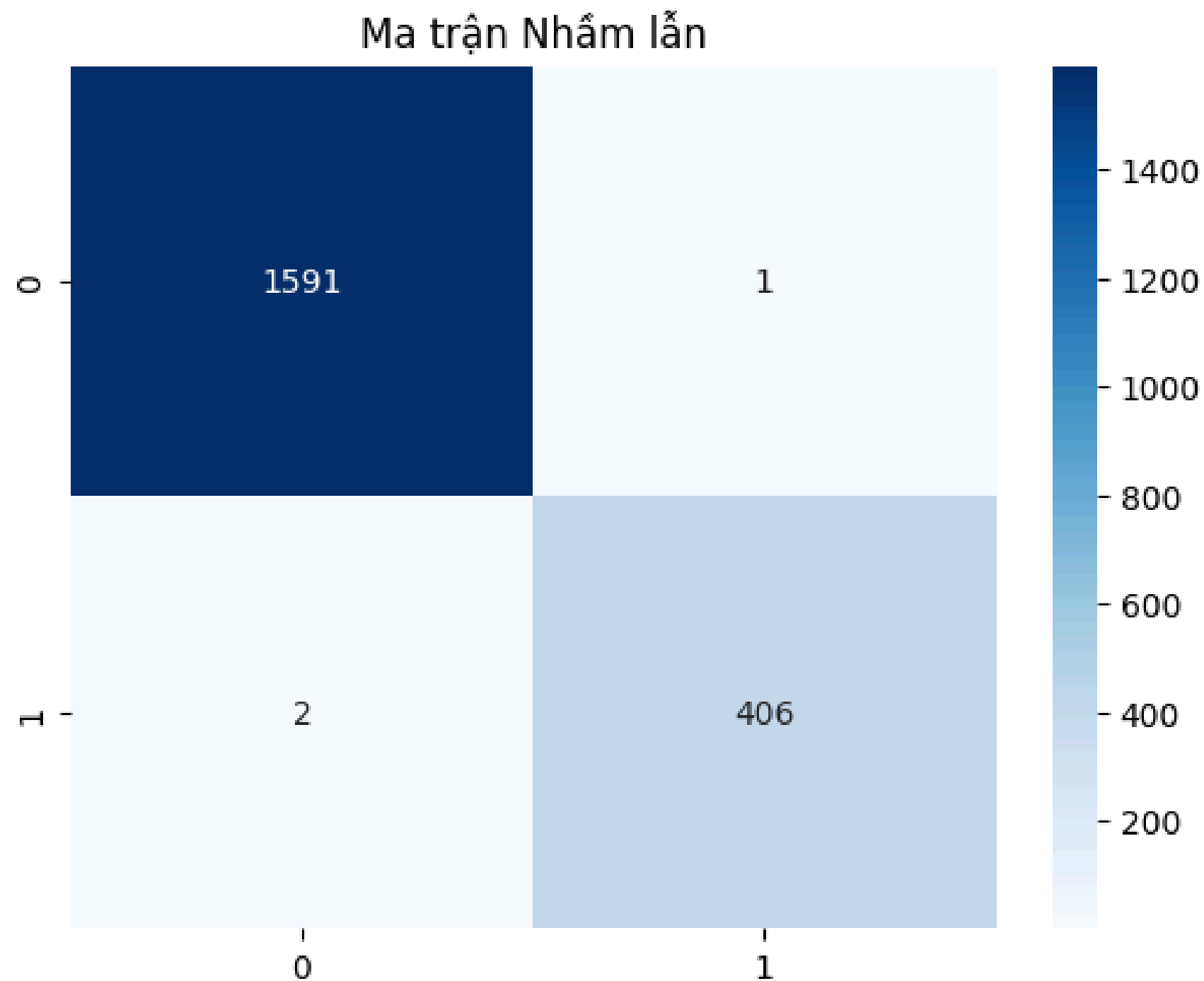
1. Độ chính xác (Accuracy): 0.9985

2. Báo cáo phân loại:

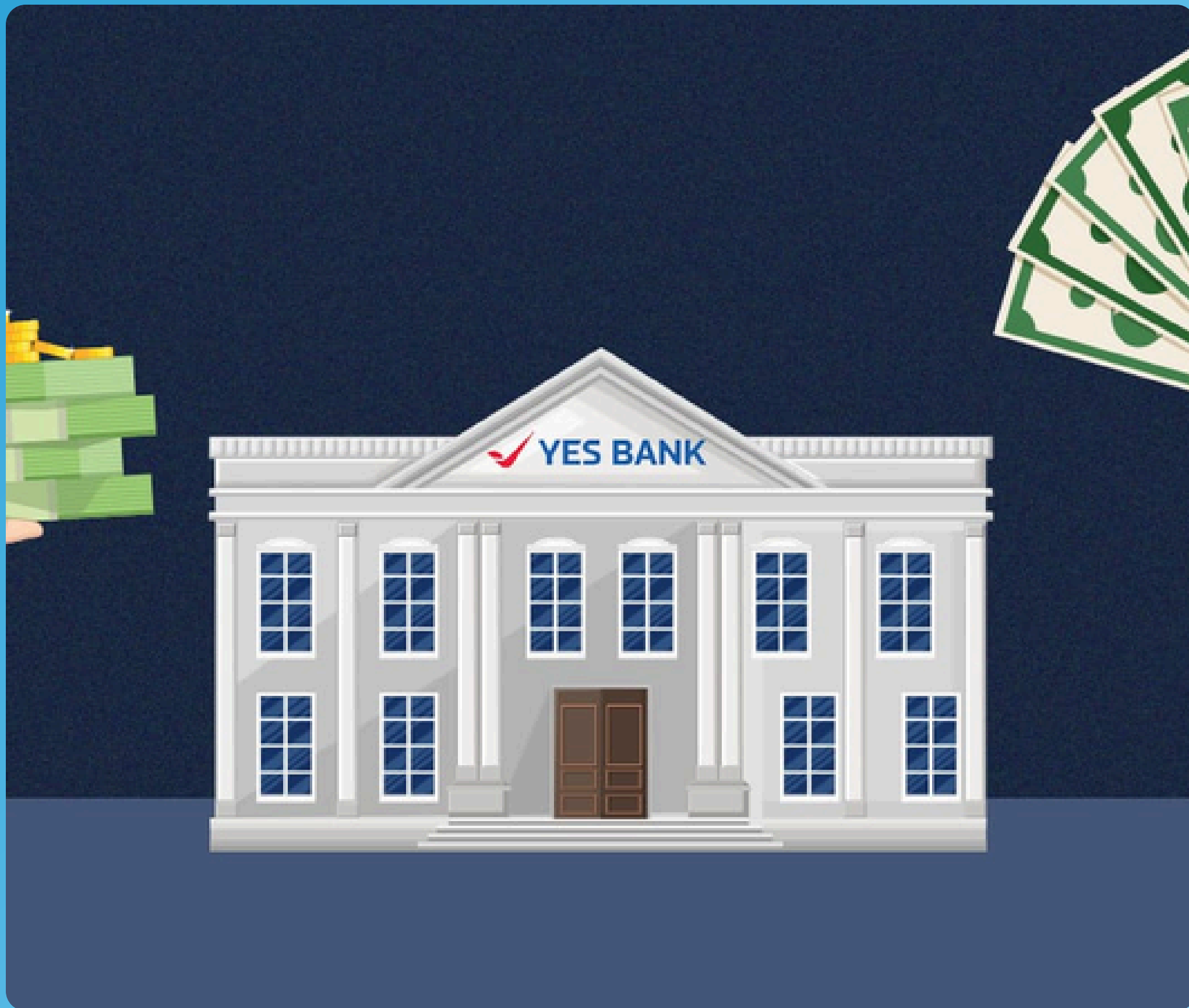
Báo cáo phân loại				
	Precision	Recall	f1-score	support
Non-churn	1.00	1.00	1.00	1592
Churn	1.00	1.00	1.00	408

# Dự báo

## 3. Ma trận nhầm lẫn



- True Negatives (TN): 1591. Mô hình dự đoán đúng 1591 khách hàng Non-Churned là Non-Churned.
- False Positives (FP): 1. Mô hình dự đoán sai chỉ 1 khách hàng Non-Churned là Churned (Lỗi Loại I).
- False Negatives (FN): 2. Mô hình dự đoán sai chỉ 2 khách hàng Churned là Non-Churned (Lỗi Loại II).
- True Positives (TP): 406. Mô hình dự đoán đúng 406 khách hàng Churned là Churned.



# Tổng kết và đề xuất

# Những phát hiện chính

- Tỷ lệ rời bỏ: Khoảng 20.4% khách hàng đã rời bỏ ngân hàng.

- Yếu tố ảnh hưởng mạnh nhất (từ EDA và Mô hình):

Khiếu nại (Complain): Khách hàng có khiếu nại gần như chắc chắn sẽ rời đi (tỷ lệ churn gần 100%). Đây là yếu tố dự báo churn cực kỳ mạnh mẽ.

- Các yếu tố có ảnh hưởng đáng kể (từ EDA và Mô hình):

Tuổi (Age): Khách hàng lớn tuổi hơn có xu hướng rời đi cao hơn, đặc biệt nhóm 41-60 tuổi có tỷ lệ/số lượng churn cao.

Trạng thái hoạt động : Khách hàng không hoạt động có tỷ lệ churn cao hơn đáng kể so với khách hàng hoạt động, bất kể thời gian gần bó hay số dư.

- Các yếu tố có ảnh hưởng vừa phải hoặc chỉ trong một số nhóm nhất định (từ EDA):

+ Giới tính (Gender): Phụ nữ có tỷ lệ churn cao hơn nam giới.

+ Địa lý (Geography): Khách hàng ở Đức có tỷ lệ churn cao hơn đáng kể so với Pháp và Tây Ban Nha.

+ Số dư tài khoản (Balance): Khách hàng churned có xu hướng có số dư trung bình cao hơn; nhóm non-churned có nhiều khách hàng với số dư bằng 0.

+ Số lượng sản phẩm (NumOfProducts): Khách hàng có 3 hoặc 4 sản phẩm có tỷ lệ churn cực kỳ cao, mặc dù số lượng ít.

# Đề xuất

Đối với khách hàng có khiếu nại	Thiết lập quy trình xử lý khiếu nại ưu tiên và nhanh chóng
	Theo dõi sát sao sau khi giải quyết khiếu nại
	Phân tích nguyên nhân gốc rễ của khiếu nại
Đối với khách hàng không hoạt động	<ul style="list-style-type: none"><li>Chiến dịch kích hoạt lại: Triển khai các chương trình khuyến mãi, ưu đãi hoặc liên hệ cá nhân để khuyến khích khách hàng sử dụng lại các dịch vụ của ngân hàng.</li></ul>
	<ul style="list-style-type: none"><li>Cá nhân hóa đề xuất dịch vụ: Dựa trên lịch sử giao dịch hoặc thông tin nhân khẩu học để gợi ý các sản phẩm/dịch vụ phù hợp với nhu cầu tiềm năng của khách hàng không hoạt động.</li></ul>
	<ul style="list-style-type: none"><li>Đơn giản hóa quy trình sử dụng: Kiểm tra và cải thiện trải nghiệm người dùng trên các kênh để giảm thiểu rào cản cho việc tái sử dụng.</li></ul>



# Đề xuất

Đối với Khách hàng lớn tuổi	Cung cấp dịch vụ hỗ trợ cá nhân hóa: Phân công nhân viên hỗ trợ riêng hoặc thiết lập kênh ưu tiên cho nhóm khách hàng này.
	Giáo dục về công nghệ và dịch vụ
	Phát triển hoặc điều chỉnh các sản phẩm tài chính phù hợp hơn với nhu cầu và giai đoạn cuộc đời của nhóm tuổi này (ví dụ: sản phẩm hưu trí, đầu tư an toàn).
Đối với Khách hàng ở Đức	Nghiên cứu sâu hơn về lý do churn tại Đức: Tìm hiểu các yếu tố văn hóa, cạnh tranh hoặc đặc thù thị trường tại Đức có thể gây ra tỷ lệ churn cao hơn.
	Điều chỉnh chiến lược sản phẩm và marketing: Phát triển các sản phẩm hoặc chiến dịch marketing phù hợp hơn với khách hàng tại Đức.
	Tăng cường sự hiện diện và hỗ trợ địa phương: Đảm bảo dịch vụ khách hàng và các kênh hỗ trợ tại Đức đáp ứng tốt nhu cầu của khách hàng.



**THANK YOU!**



A graphic with a solid blue background. The words "THANK YOU!" are written in a bold, white, sans-serif font, slanted upwards from left to right. The text has a subtle drop shadow. Above the word "THANK" is a white decorative swirl. Below the word "YOU!" is another white decorative swirl. The background is sparsely decorated with small white dots of varying sizes, resembling stars or confetti.