

Bank Marketing Campaign

Team: DA50_G1

Members:

Vũ Văn Vững

Đỗ Trung Hiếu

Hải Luân

Thuyết trình: Đỗ Trung Hiếu

More Info

Start Slide →



TABLE OF CONTENT



1.ASK

2.Prepare

3.Analyze

4.Modeling

5.Share

6.ACT

1.ASK



- Chiến dịch tiếp thị trực tiếp của một tổ chức ngân hàng Bồ Đào Nha dựa trên các cuộc gọi điện thoại nhằm mục đích quảng bá sản phẩm tiền gửi có kỳ hạn ngân hàng. Trong quá trình này, nhân viên tiếp thị thường phải liên hệ với cùng một khách hàng nhiều lần để thuyết phục họ đăng ký sản phẩm.
- Mục tiêu chính của báo cáo này là xác định xem khách hàng có quan tâm và quyết định đăng ký tiền gửi có kỳ hạn hay không. Giúp ngân hàng dự báo được những khách hàng có tiềm năng sẽ quyết định tiền gửi có kỳ hạn.

1. Điều gì khiến khách hàng quyết định gửi có kỳ hạn tại ngân hàng ?
2. Có điểm khác biệt nào giữa khách hàng gửi kỳ có hạn hay không ?

2. Prepare

age	job	marital	education	default	housing	loan	contact	month	day_of_week	...	campaign	pdays	previous	poutcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y
56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	...	1	999	0	nonexistent	1.1	93.994	93.994	93.994	93.994	
57	services	married	high.school	unknown	no	no	telephone	may	mon	...	1	999	0	nonexistent	1.1	93.994	93.994	93.994	93.994	
37	services	married	high.school	no	yes	no	telephone	may	mon	...	1	999	0	nonexistent	1.1	93.994	93.994	93.994	93.994	
40	admin.	married	basic.6y	no	no	no	telephone	may	mon	...	1	999	0	nonexistent	1.1	93.994	93.994	93.994	93.994	

Tên cột	Kiểu dữ liệu	Mô tả
age	int64	Tuổi của khách hàng
job	object	Nghề nghiệp của khách hàng
marital	object	Tình trạng hôn nhân (ví dụ: đã kết hôn, độc thân, ly dị)
education	object	Trình độ học vấn
default	object	Có khoản nợ tín dụng không trả được không
housing	object	Có khoản vay mua nhà không
loan	object	Có khoản vay cá nhân không
contact	object	Phương thức liên lạc (ví dụ: điện thoại di động, điện thoại cố định)
month	object	Tháng của năm liên hệ cuối cùng
day_of_week	object	Ngày trong tuần liên hệ cuối cùng
duration	int64	Thời lượng của cuộc gọi cuối cùng (tính bằng giây)
campaign	int64	Số lần liên hệ với khách hàng trong chiến dịch này
pdays	int64	Số ngày kể từ lần liên hệ cuối cùng trong chiến dịch trước đó (999 có nghĩa là khách hàng chưa được liên hệ trước đó)
previous	int64	Số lần liên hệ được thực hiện trước chiến dịch này
poutcome	object	Kết quả của chiến dịch tiếp thị trước đó (ví dụ: thành công, không thành công, không tồn tại)
emp.var.rate	float64	Tỷ lệ thay đổi việc làm - chỉ số kinh tế
cons.price.idx	float64	Chỉ số giá tiêu dùng - chỉ số kinh tế
cons.conf.idx	float64	Chỉ số niềm tin của người tiêu dùng - chỉ số kinh tế
euribor3m	float64	Tỷ lệ Euribor 3 tháng - chỉ số kinh tế
nr.employed	float64	Số lượng nhân viên - chỉ số kinh tế
y	object	Biến mục tiêu (có khách hàng đăng ký tiền gửi có kỳ hạn hay không)

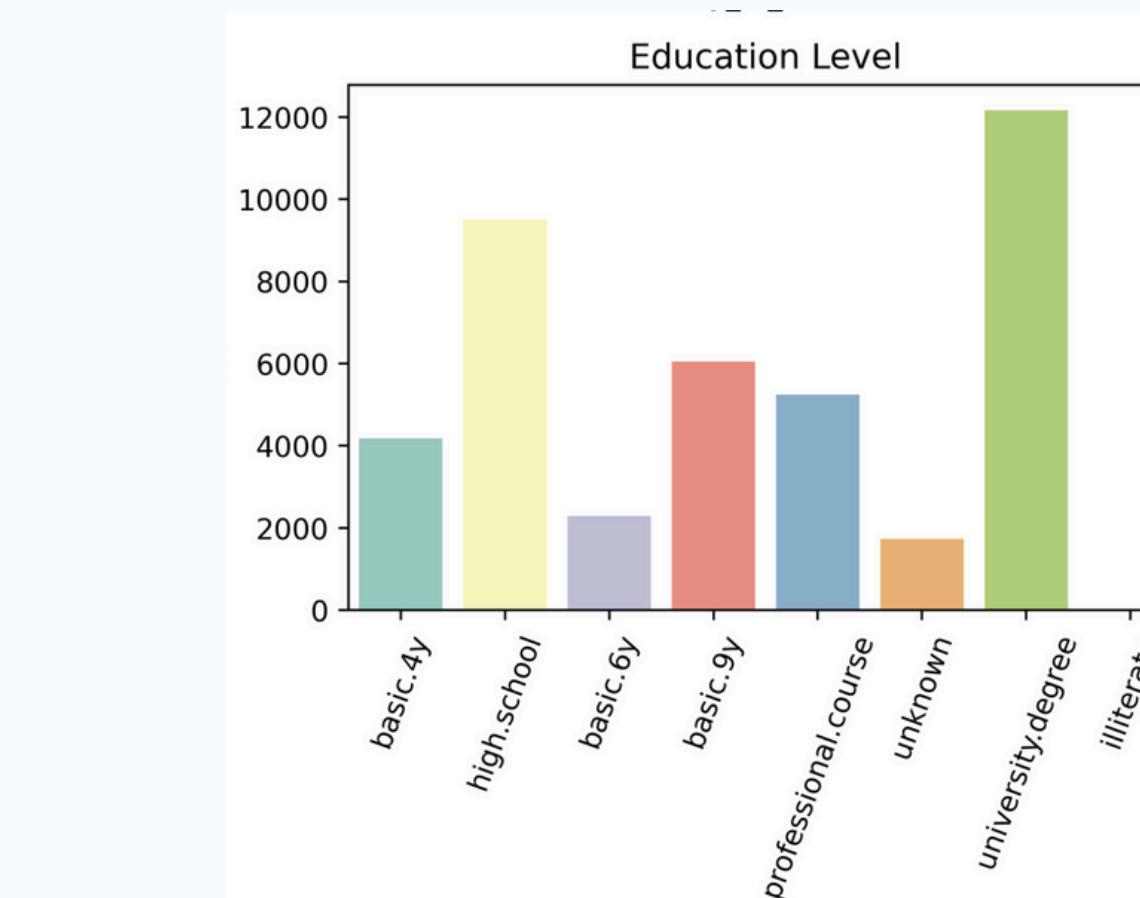
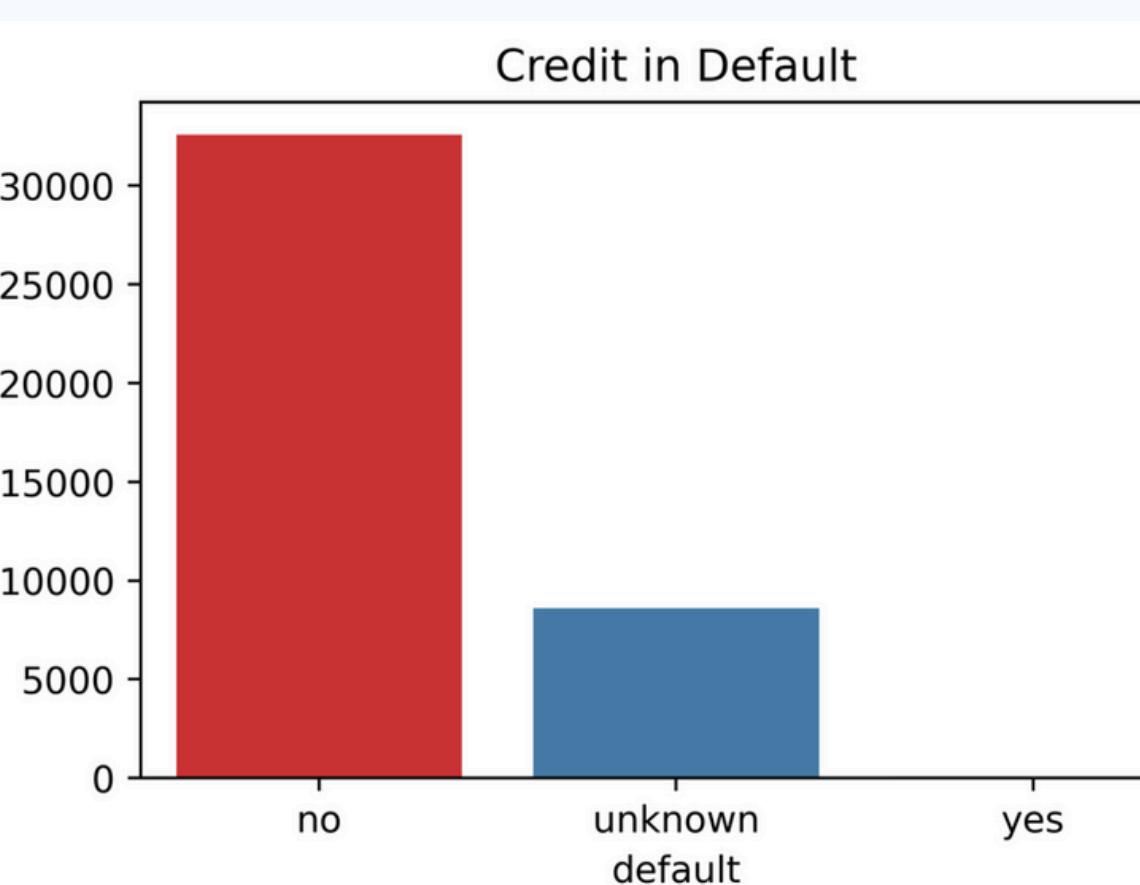
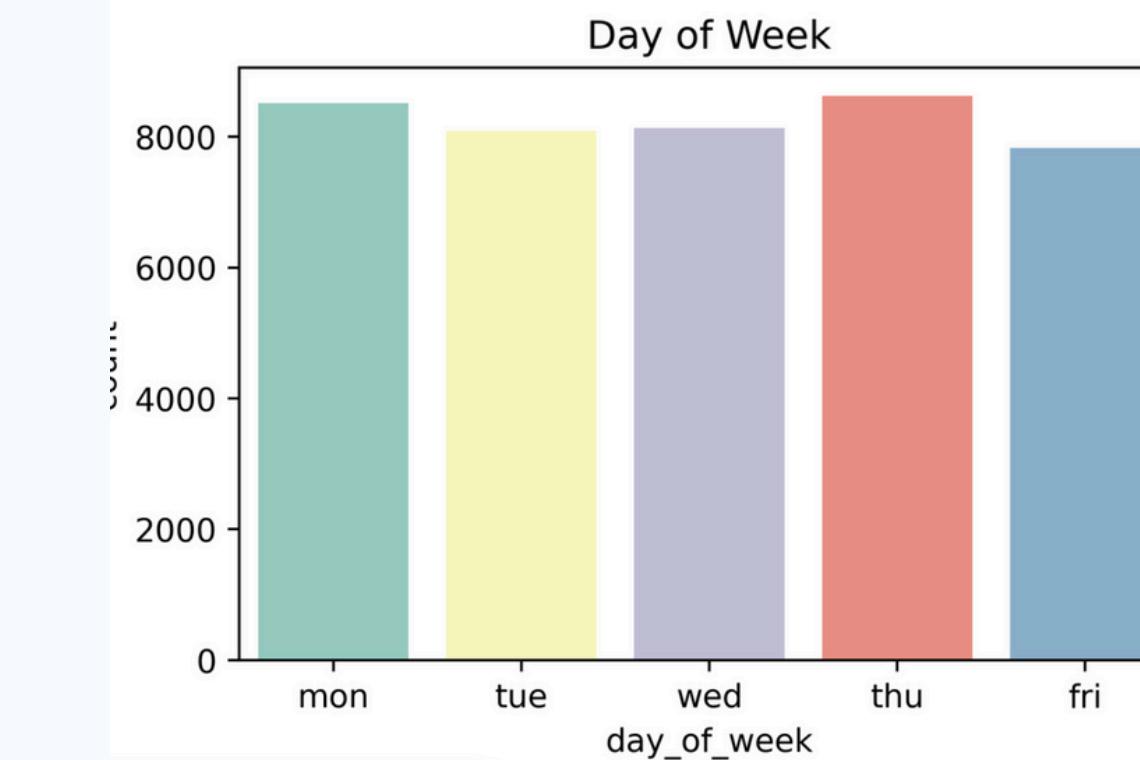
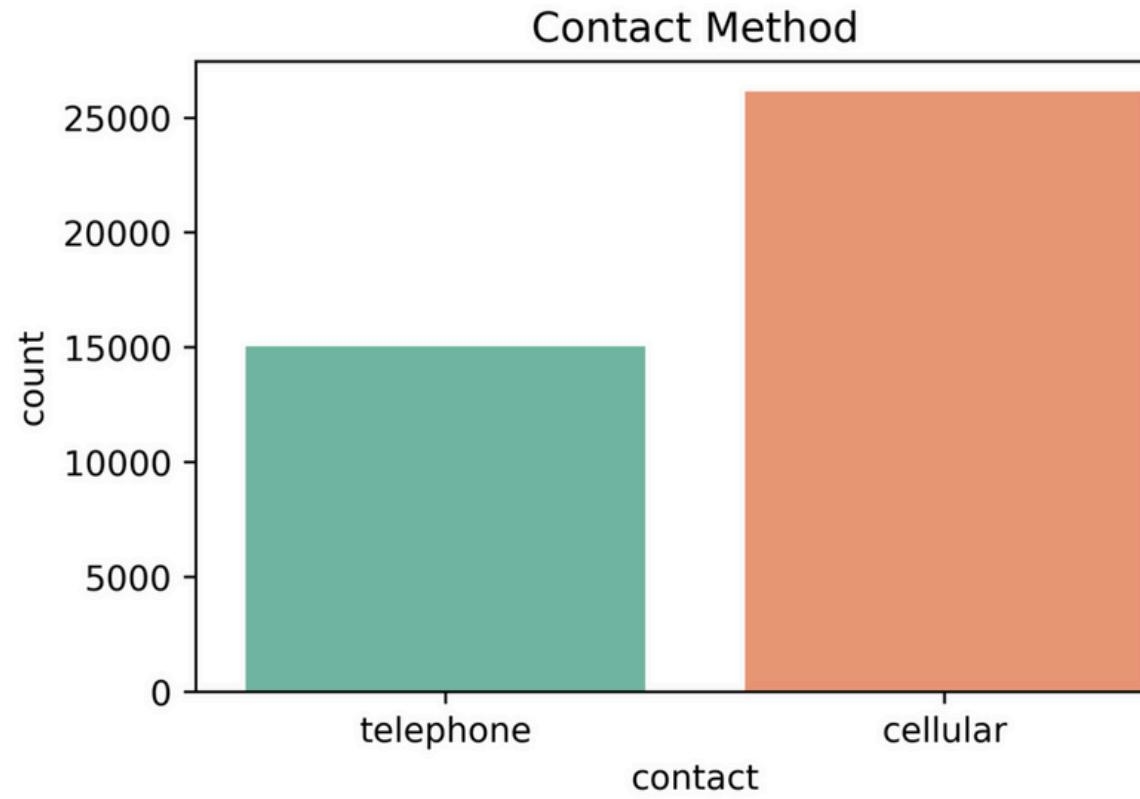
Bộ dữ liệu có 41.188 rows, 21 columns, được sắp xếp theo ngày (từ tháng 5 năm 2008 đến tháng 11 năm 2010)

2. Prepare

	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
count	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000	41188.000000
mean	40.02406	258.285010	2.567593	962.475454	0.172963	0.081886	93.575664	-40.502600	3.621291	5167.035911
std	10.42125	259.279249	2.770014	186.910907	0.494901	1.570960	0.578840	4.628198	1.734447	72.251528
min	17.000000	0.000000	1.000000	0.000000	0.000000	-3.400000	92.201000	-50.800000	0.634000	4963.600000
25%	32.000000	102.000000	1.000000	999.000000	0.000000	-1.800000	93.075000	-42.700000	1.344000	5099.100000
50%	38.000000	180.000000	2.000000	999.000000	0.000000	1.100000	93.749000	-41.800000	4.857000	5191.000000
75%	47.000000	319.000000	3.000000	999.000000	0.000000	1.400000	93.994000	-36.400000	4.961000	5228.100000
max	98.000000	4918.000000	56.000000	999.000000	7.000000	1.400000	94.767000	-26.900000	5.045000	5228.100000

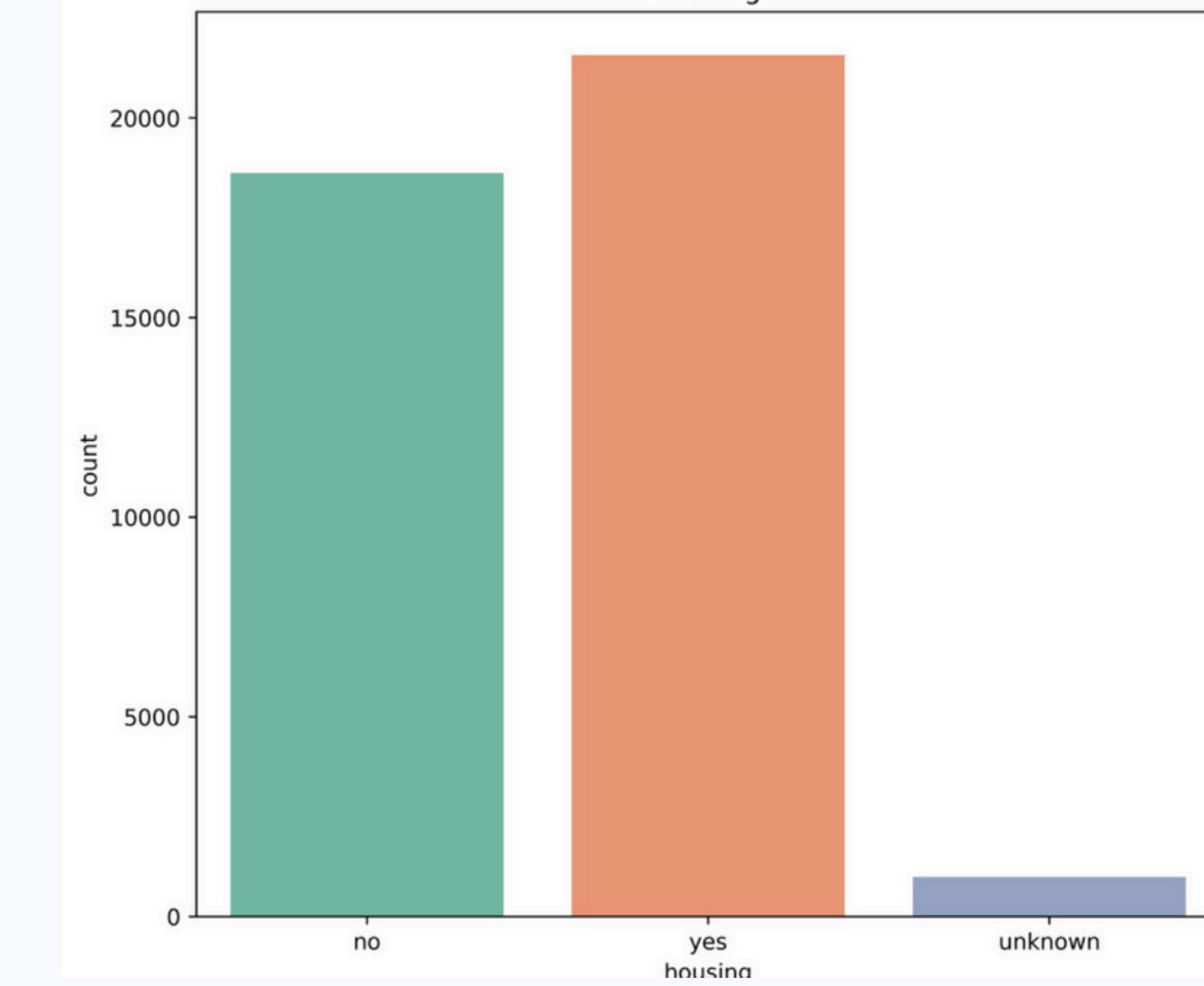
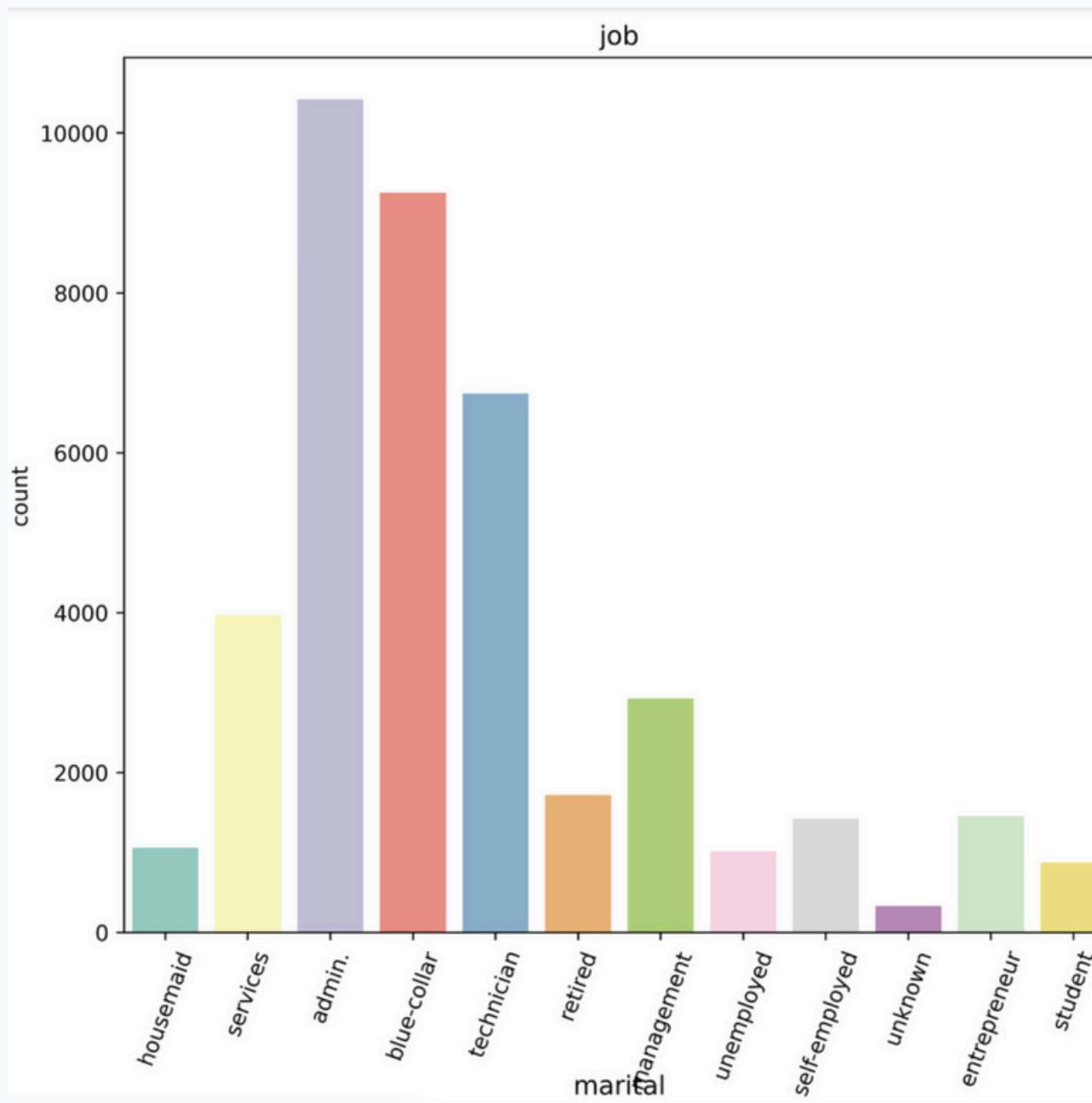
- Số lượng (count): Cho biết số lượng mẫu (khách hàng) có trong bộ dữ liệu cho mỗi biến. Trong trường hợp này, tất cả các biến đều có 41.188 mẫu, tức là có dữ liệu cho tất cả các khách hàng về các biến này.
- Trung bình (mean): Giá trị trung bình của mỗi biến. Ví dụ, tuổi trung bình của khách hàng là 40.02, thời gian trung bình của cuộc gọi là 258.29 giây.
- Độ lệch chuẩn (std): Cho biết mức độ phân tán của các giá trị xung quanh giá trị trung bình. Ví dụ, độ lệch chuẩn của tuổi là 10.42, cho thấy sự đa dạng về độ tuổi của khách hàng.
- Giá trị nhỏ nhất (min): Giá trị nhỏ nhất của mỗi biến. Ví dụ, khách hàng trẻ nhất là 17 tuổi, và có những cuộc gọi có thời lượng 0 giây.
- Các phần trăm (25%, 50%, 75%): Cho biết các giá trị tại các ngưỡng phần trăm khác nhau của phân phối. Ví dụ, 50% khách hàng có độ tuổi từ 38 trở xuống, và 75% cuộc gọi có thời lượng 319 giây trở xuống.
- Giá trị lớn nhất (max): Giá trị lớn nhất của mỗi biến. Ví dụ, khách hàng lớn tuổi nhất là 98 tuổi, và cuộc gọi dài nhất là 4918 giây.

3.Analyze



3.Analyze

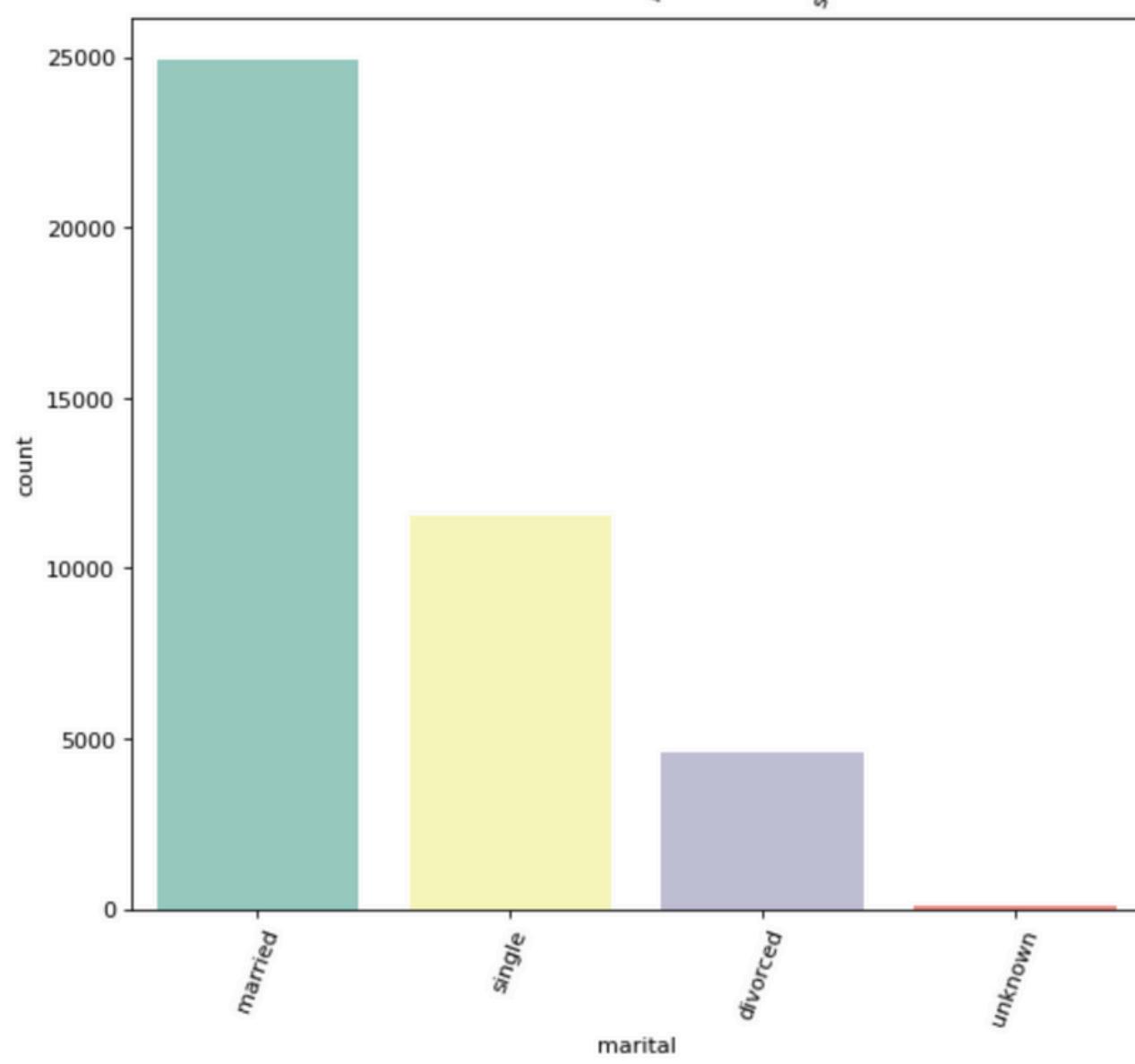
- Tình trạng nhà ở: Phần lớn khách hàng có nhà ở, điều này có thể liên quan đến khả năng tài chính và ổn định của họ



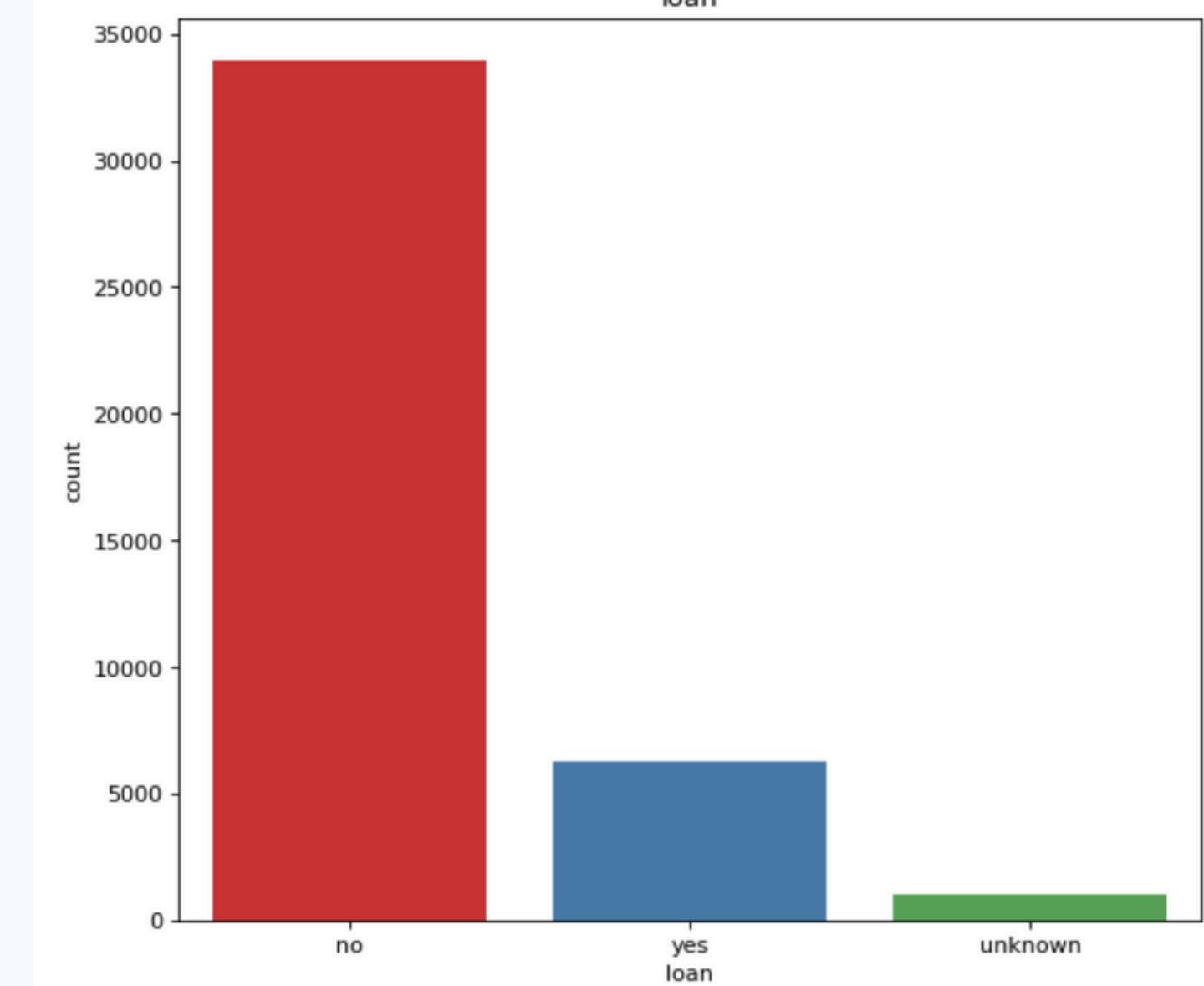
- Nghề nghiệp: Sự đa dạng về nghề nghiệp, với các nhóm nghề nghiệp phổ biến như "blue-collar", "management" và "technician". Điều này có thể cung cấp thông tin về đối tượng khách hàng chính của chiến dịch tiếp thị.

3.Analyze

- Khoản vay: Phần lớn khách hàng không có khoản vay, điều này có thể ảnh hưởng đến khả năng chi tiêu và đầu tư của họ.

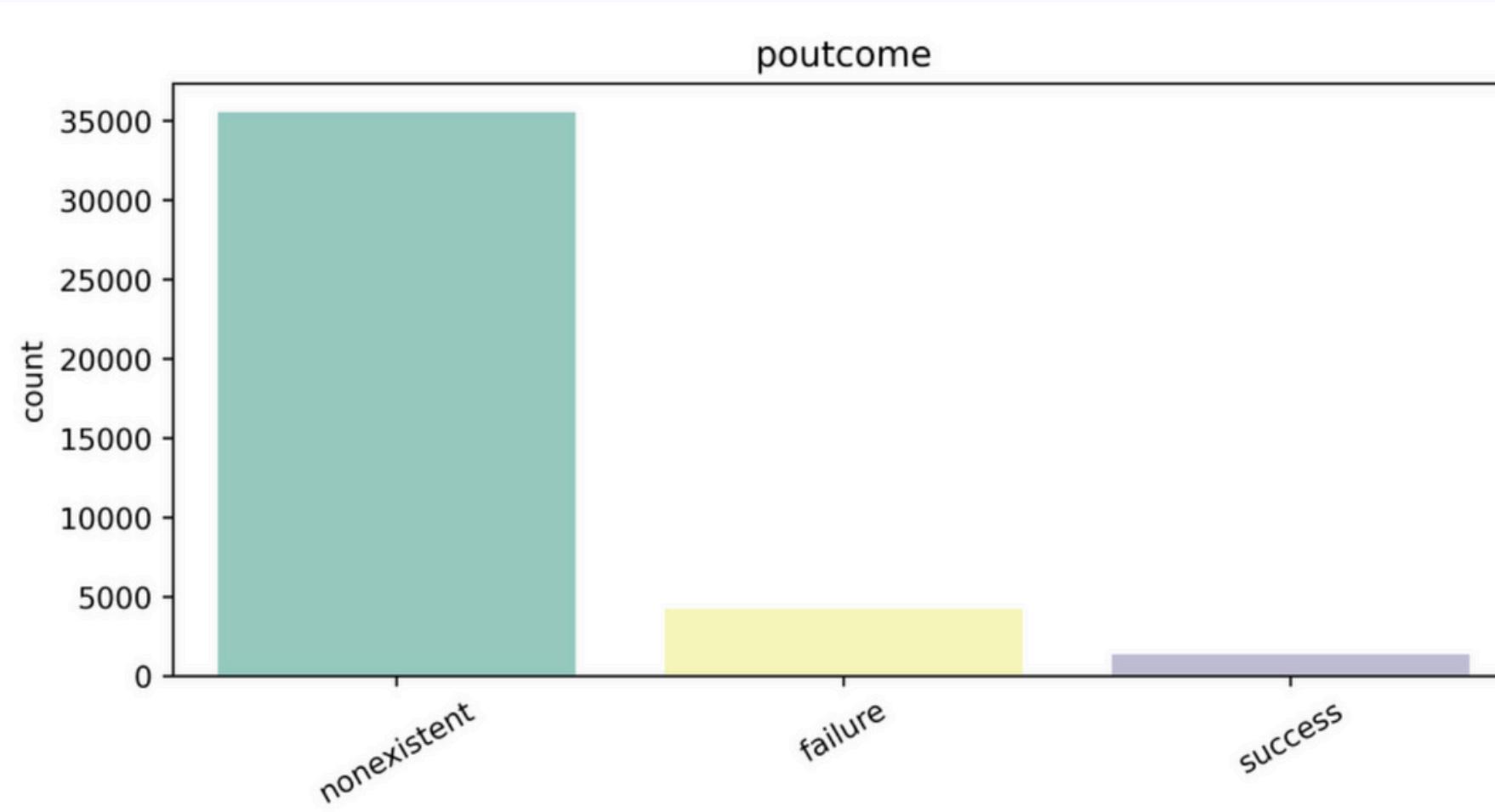
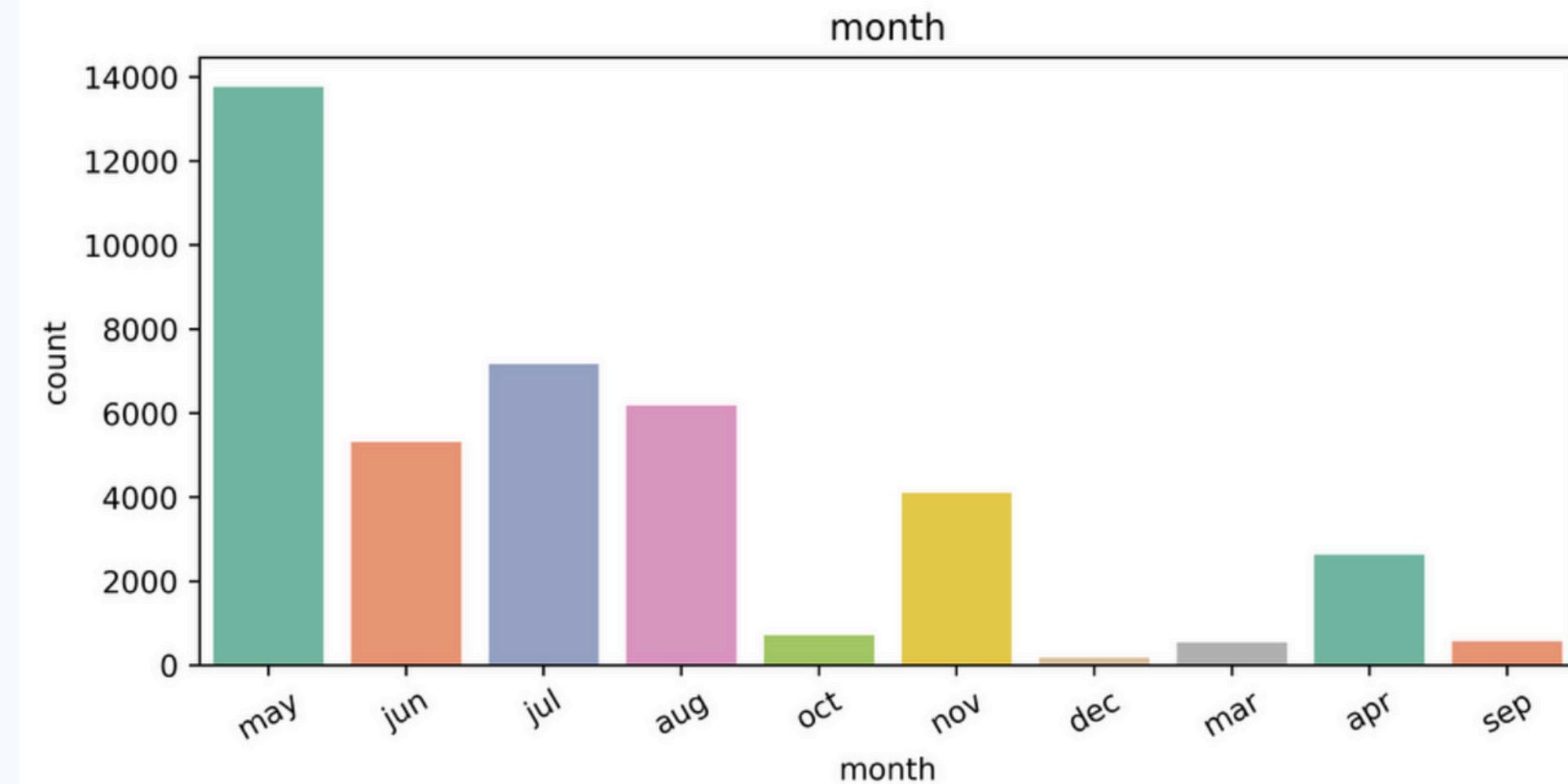


- Tình trạng hôn nhân: Phần lớn khách hàng đã kết hôn, điều này có thể ảnh hưởng đến quyết định tài chính và sự ổn định tài chính của họ.



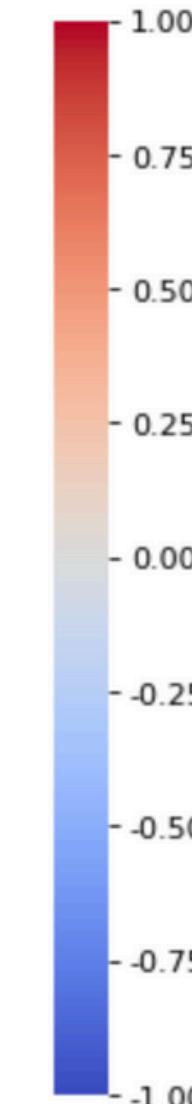
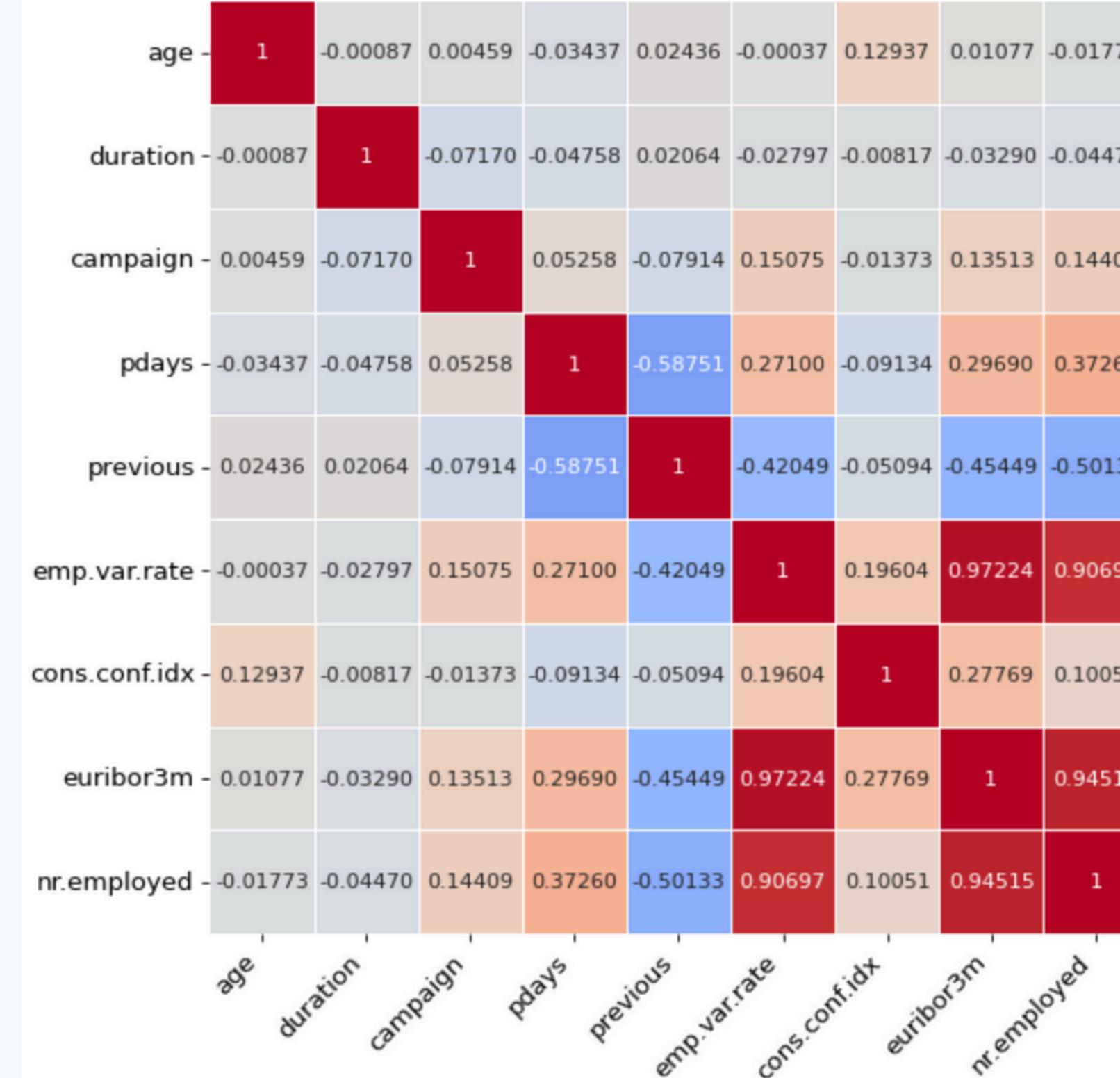
3.Analyze

- Tháng 5: Tháng này có số lượng liên lạc cao nhất, điều này có thể gợi ý rằng chiến dịch tiếp thị chính được thực hiện vào tháng này. Ngân hàng có thể đã tập trung nhiều nguồn lực vào tháng này để tối đa hóa số lượng liên lạc.



- Kết quả của chiến dịch trước: Phần lớn khách hàng không có kết quả từ chiến dịch trước đó, có thể là do họ chưa từng được liên lạc trước đó hoặc thông tin không được ghi nhận. Một số nhỏ đã thất bại hoặc thành công, điều này có thể ảnh hưởng đến chiến lược tiếp thị hiện tại.

3.Analyze



- **Tương quan mạnh:** Tập trung vào các cặp biến có tương quan mạnh, như giữa emp.var.rate, euribor3m và nr.employed. Những tương quan này có thể ảnh hưởng đến quyết định của khách hàng về việc gửi tiền tiết kiệm kỳ hạn.
- **Tương quan yếu:** Đa số các biến còn lại không có tương quan mạnh với nhau, cho thấy chúng có thể mang lại thông tin độc lập cho mô hình dự đoán.
- **Chiến lược:** Ngân hàng có thể tập trung vào các yếu tố kinh tế như tỷ lệ biến động việc làm và lãi suất Euribor để dự đoán hành vi của khách hàng.

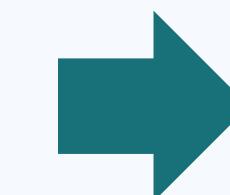
4. Modeling

Model Performance Metrics

Model	Accuracy	Precision	Precision	Recall	Recall	F1-score	F1-score
		(0)	(1)	(0)	(1)	(0)	(1)
Logistic Regression	0.91	0.93	0.65	0.97	0.40	0.95	0.50
Random Forest	0.92	0.94	0.68	0.97	0.54	0.96	0.60
AdaBoost	0.91	0.93	0.66	0.97	0.41	0.95	0.50

Confusion Matrix

Actual \ Predicted	0 (No)	1 (Yes)
0 (No)	10,689	300
1 (Yes)	630	738



Report for Random Forest

Class	Precision	Recall	F1-score	Support
No	0.94	0.97	0.96	10,989
Yes	0.68	0.54	0.60	1,368
Accuracy	0.92			
Macro avg	0.81	0.75	0.78	12,357
Weighted avg	0.91	0.92	0.92	12,357

ROC Curves

- Logistic Regression: AUC = 0.94
- Random Forest: AUC = 0.96
- AdaBoost: AUC = 0.94

Đánh giá mô hình

- **Mô hình Logistic Regression:**

- Độ chính xác cao (0.91) và F1-score tốt cho lớp "Không" (0.95).
- Precision và Recall cho lớp "Có" (lớp quan trọng) thấp (0.65 và 0.40).
- AUC = 0.94.

- **Mô hình Random Forest:**

- Độ chính xác cao nhất (0.92) và F1-score tốt nhất cho cả hai lớp (0.96 cho lớp "Không" và 0.60 cho lớp "Có").
- Precision và Recall cho lớp "Có" tốt hơn các mô hình khác (0.68 và 0.54).
- AUC = 0.96.
- Khuyến nghị: Dựa vào các chỉ số tổng thể và hiệu suất cao của Recall và F1-score, Random Forest là mô hình tốt nhất để sử dụng cho bài toán này.

- **Mô hình AdaBoost:**

- Độ chính xác cao (0.91) và F1-score tốt cho lớp "Không" (0.95).
- Precision và Recall cho lớp "Có" tốt hơn Logistic Regression nhưng vẫn thấp hơn Random Forest (0.66 và 0.41).
- AUC = 0.94.



- Random Forest là mô hình hiệu quả nhất cho bài toán dự đoán khách hàng có đăng ký gửi tiền tiết kiệm kỳ hạn hay không. Với độ chính xác cao nhất, cùng với Recall và F1-score tốt cho lớp "Có" (lớp quan trọng), Random Forest giúp ngân hàng dự đoán chính xác hơn và đưa ra các quyết định kinh doanh hiệu quả.

- Khuyến nghị Sử dụng: Áp dụng mô hình Random Forest trong các chiến dịch tiếp thị để tăng tỷ lệ thành công của các cuộc gọi tiếp thị và tối ưu hóa nguồn lực tiếp thị của ngân hàng.

5. Share

1. Điều gì khiến khách hàng quyết định gửi có kỳ hạn tại ngân hàng ?

Các yếu tố quyết định bao gồm tuổi tác, trình độ học vấn, thời gian liên hệ, nghề nghiệp và kết quả chiến dịch trước đó. Những khách hàng tham gia chiến dịch trước đó thường sẽ tham gia tiếp chiến dịch mới của ngân hàng, thói quen hoặc có được trải nghiệm tốt từ ngân hàng.

2. Có điểm khác biệt nào giữa khách hàng gửi kỳ hạn hay không ?

Có sự khác biệt rõ rệt về độ tuổi, trình độ học vấn, tình trạng hôn nhân, loại công việc và thời gian liên hệ. Những khách hàng có xu hướng gửi tiền kỳ hạn thường có các đặc điểm ổn định và thu nhập cao hơn.

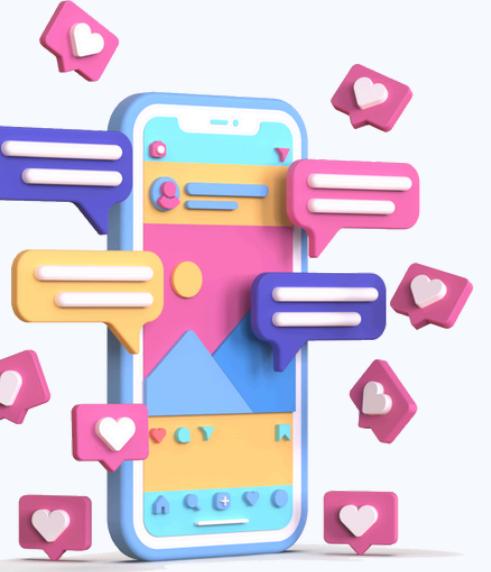
6.ACT



Tập trung vào nhóm khách hàng tiềm năng
Phân tích đặc điểm và phân loại khách hàng để tùy chỉnh nội dung tiếp thị.



Tận dụng công nghệ và phân tích dữ liệu chiến dịch và tự động hóa tiếp thị.



Tăng cường chất lượng cuộc gọi tiếp thị
Đào tạo nhân viên và tối ưu hóa thời gian liên hệ.

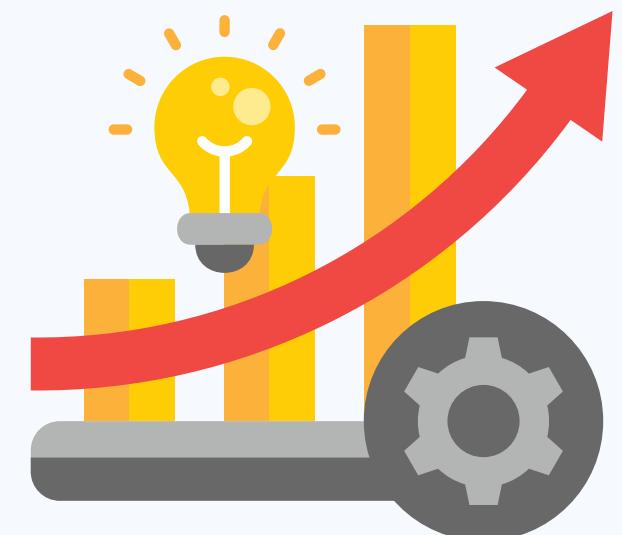


Cung cấp khuyến mãi và chương trình chăm sóc khách hàng đặc biệt.



Cải thiện nội dung chiến dịch

Tùy chỉnh nội dung và sử dụng bằng chứng thực tế.



Đo lường và cải thiện liên tục

Theo dõi kết quả và điều chỉnh chiến lược tiếp thị.

THANK YOU

Back To Home →

