

Executive Summary

Executive Summary: Customer Propensity to Subscribe for a Term Deposit

1. Project Objective:

The primary goal of this analysis was to develop a model that predicts a client's propensity to subscribe to a term deposit. By identifying clients with a higher likelihood of conversion, marketing efforts can be optimized for better efficiency and ROI.

2. Methodology:

- A comprehensive Exploratory Data Analysis (EDA) was conducted to uncover patterns and insights from the client and campaign data.
- A machine learning pipeline was built using a Random Forest Classifier to handle both numerical and categorical features.
- The model was trained on historical data and evaluated for its predictive performance.

3. Key Findings & Insights:

- Demographics: The typical client is middle-aged (~41 years), married, and holds a secondary level of education.
- Class Imbalance: There is a notable imbalance in the dataset, with only 11.7% of clients subscribing to the term deposit.
- Top Predictors: The 'duration' of the last contact is the most influential factor. Other key predictors include client 'age', 'account balance', and a 'successful outcome' in a previous campaign.
- Campaign Timing: The month of 'May' shows the highest volume of campaign activity.

4. Model Performance:

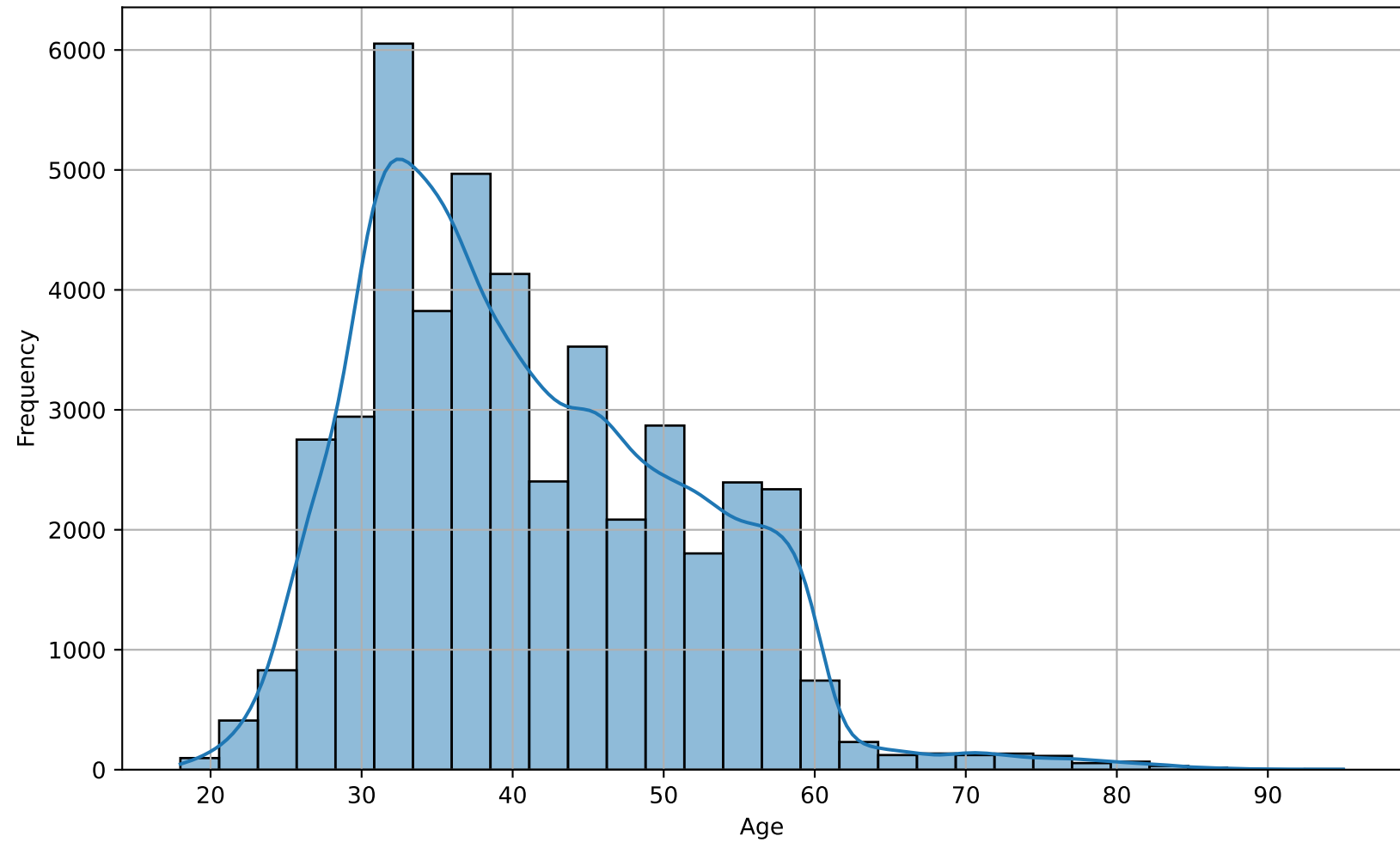
The Random Forest model demonstrated strong predictive power, achieving a high ROC AUC score of approximately 0.92. This confirms its effectiveness in distinguishing between subscribers and non-subscribers.

5. Recommendations:

- Target clients who have had successful interactions in past campaigns.
- Encourage deeper engagement during calls, as 'duration' is a key indicator of interest.
- Tailor marketing messages and timing based on key demographic features like age and job type to increase relevance and conversion rates.

Age Distribution of Clients

Age Distribution of Clients

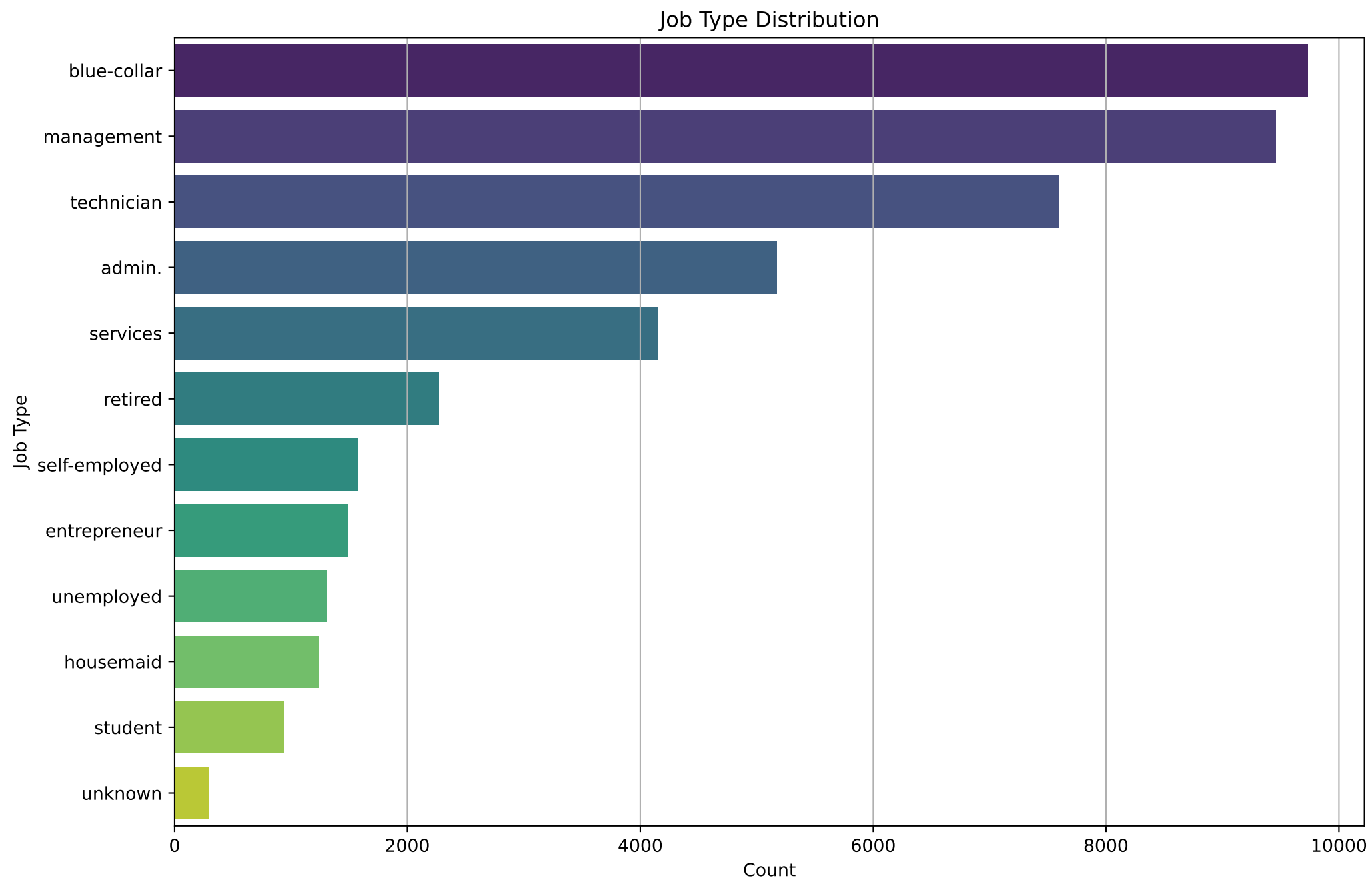


Age Summary Statistics

Age Summary Statistics:

count	45216.000000
mean	40.938186
std	10.621249
min	18.000000
25%	33.000000
50%	39.000000
75%	48.000000
max	95.000000

Job Type Distribution



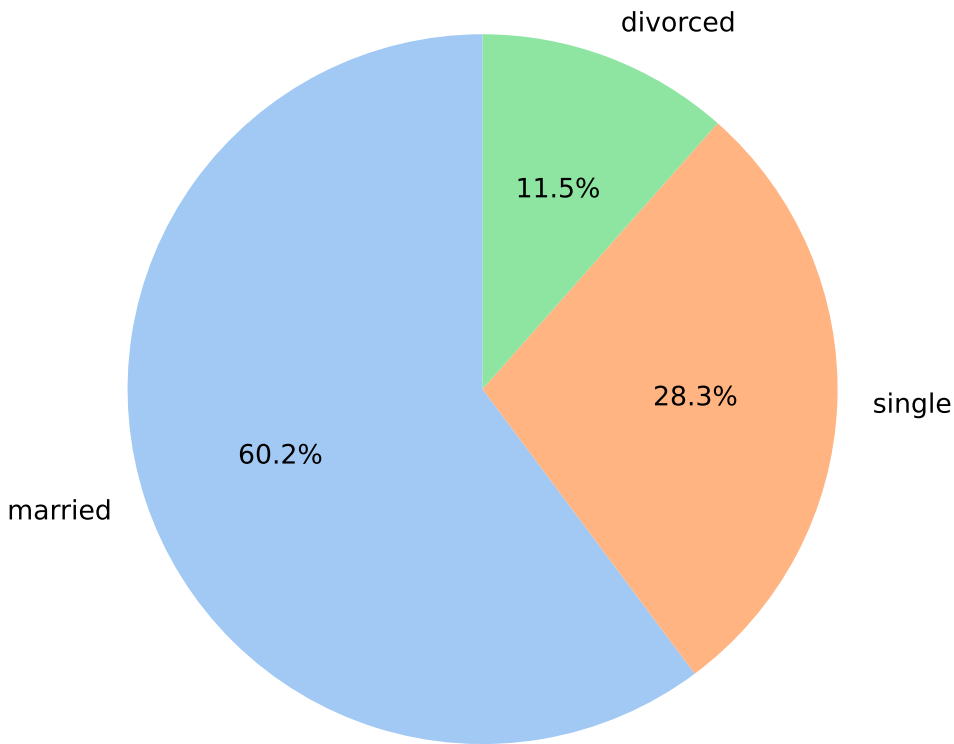
Job Type Distribution

Job Type Distribution (%):

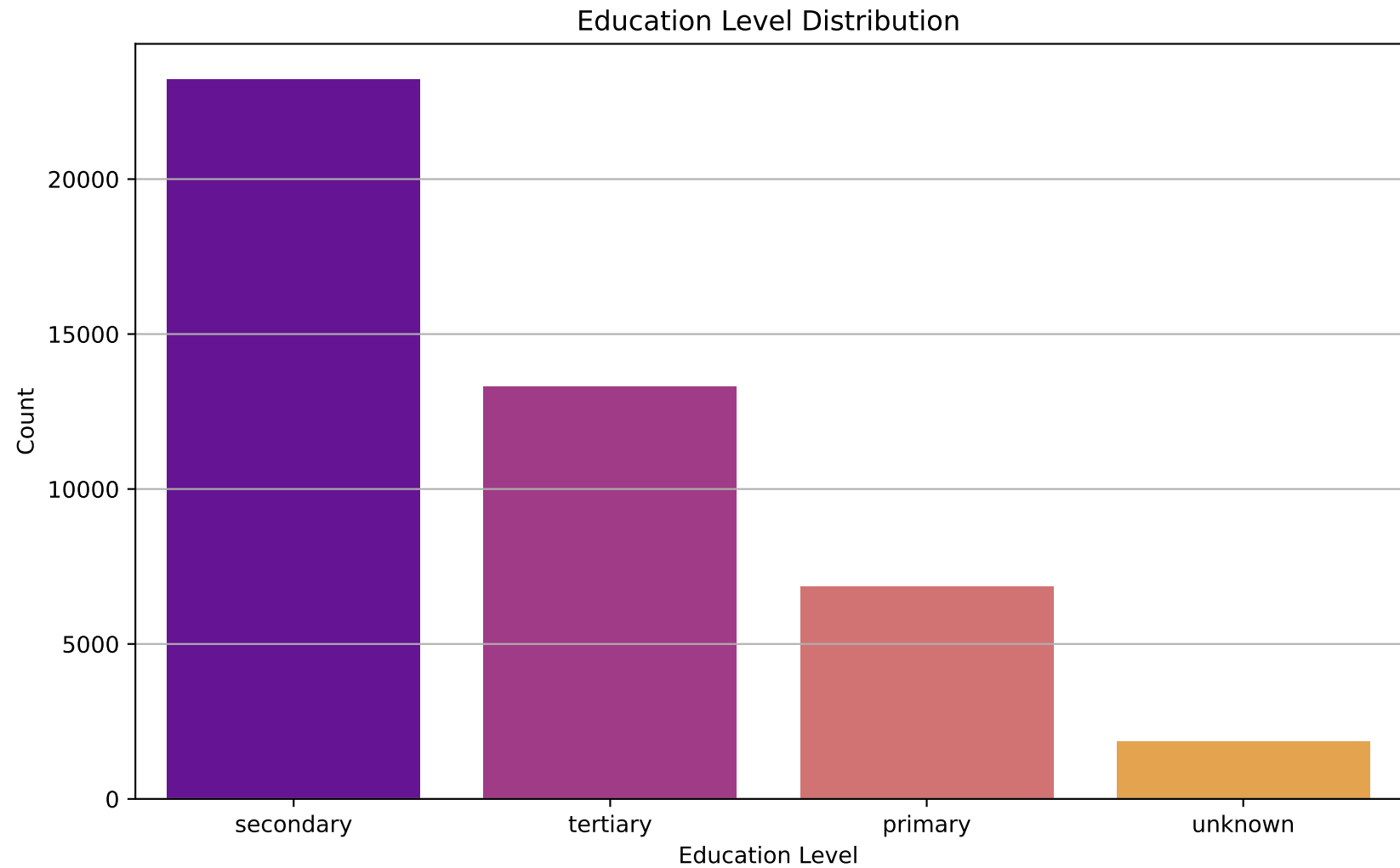
job	
blue-collar	21.523355
management	20.921798
technician	16.801575
admin.	11.436217
services	9.187013
retired	5.013712
self-employed	3.492127
entrepreneur	3.288659
unemployed	2.881723
housemaid	2.742392
student	2.074487
unknown	0.636943

Marital Status Distribution

Marital Status Distribution



Education Level Distribution



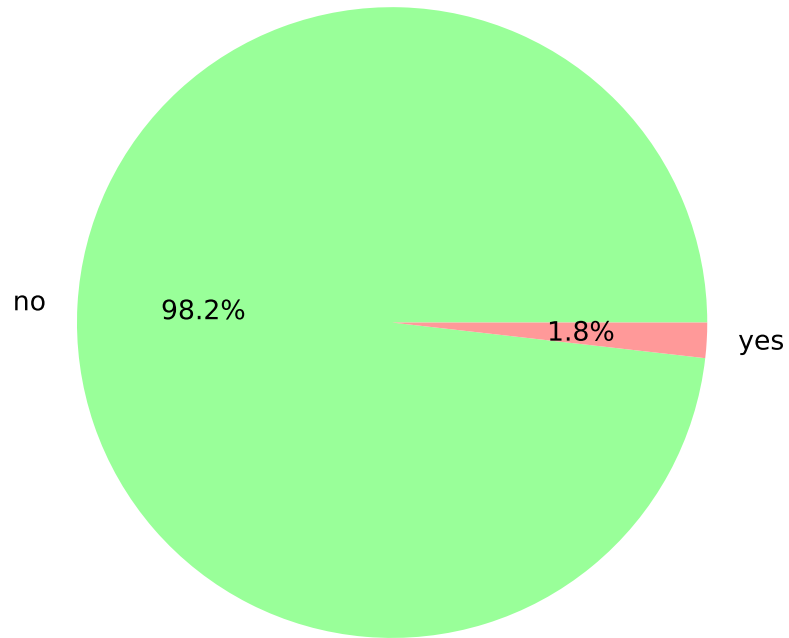
Education Level Distribution

Education Level Distribution (%):

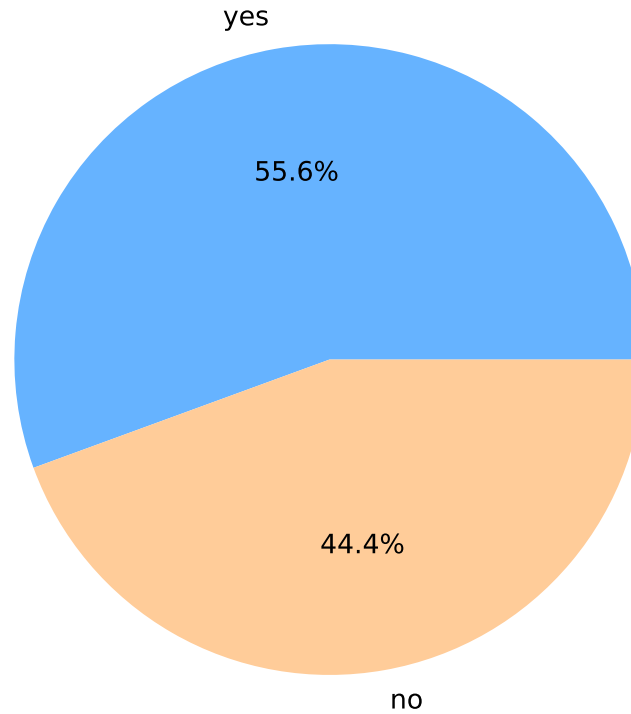
education	
secondary	51.321523
tertiary	29.418530
primary	15.152722
unknown	4.107226

Loan and Default Status

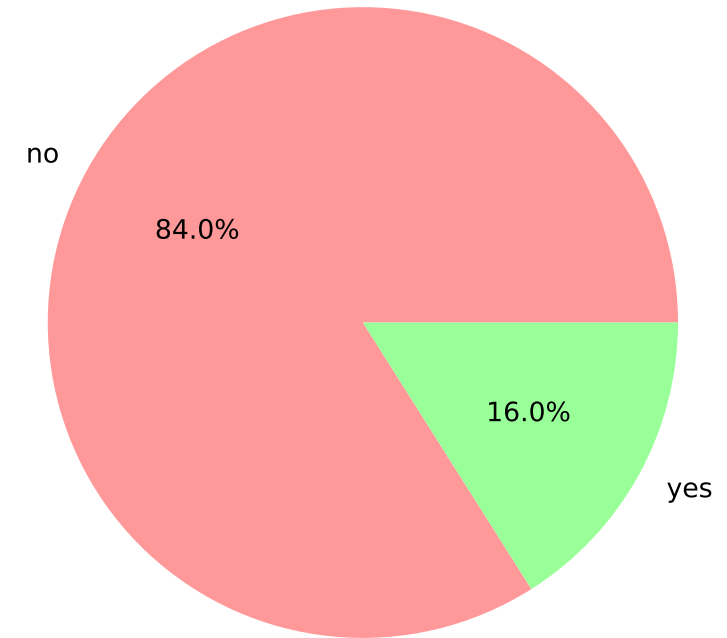
Credit in Default



Housing Loan

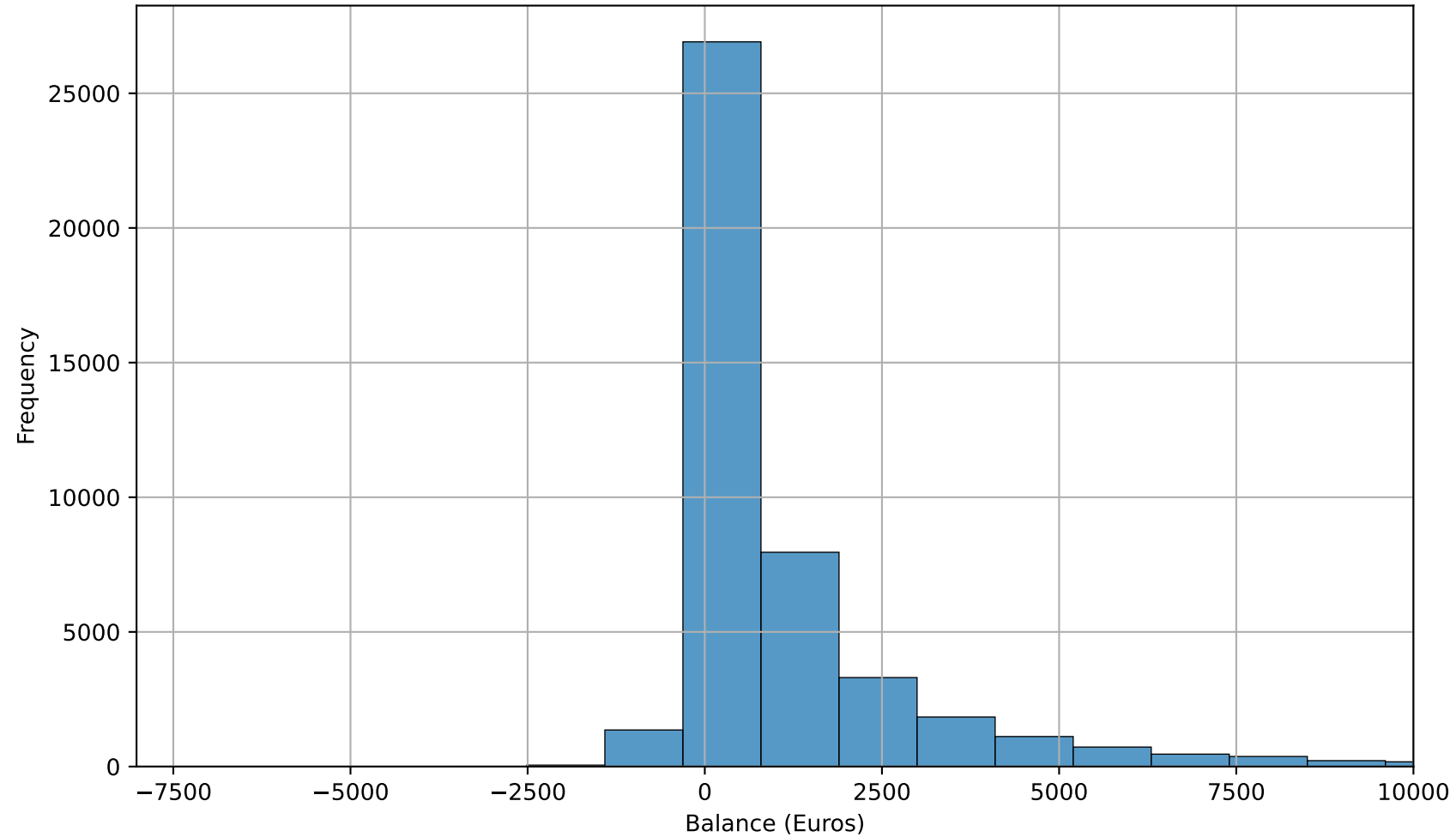


Personal Loan



Average Yearly Balance Distribution

Average Yearly Balance Distribution

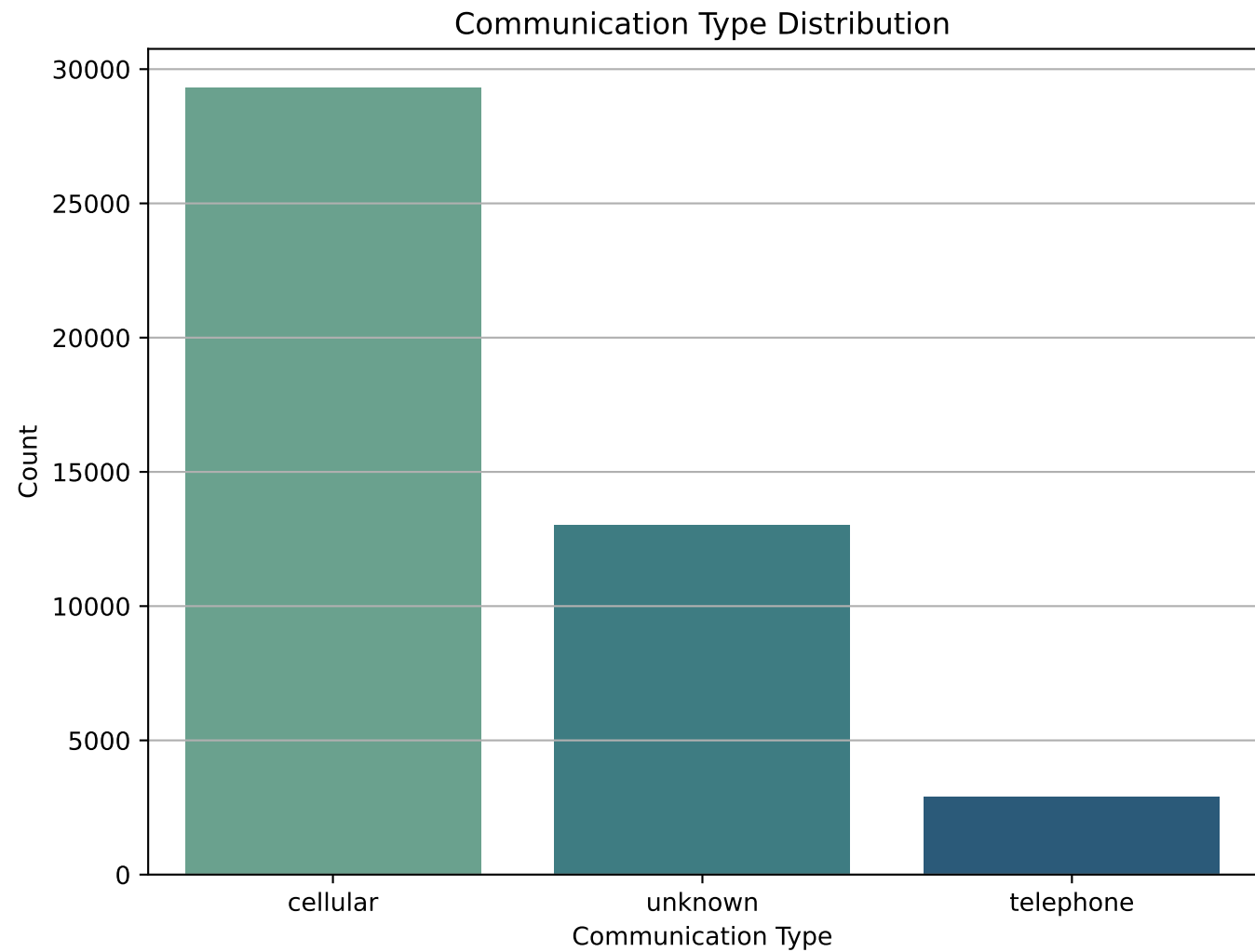


Balance Summary Statistics

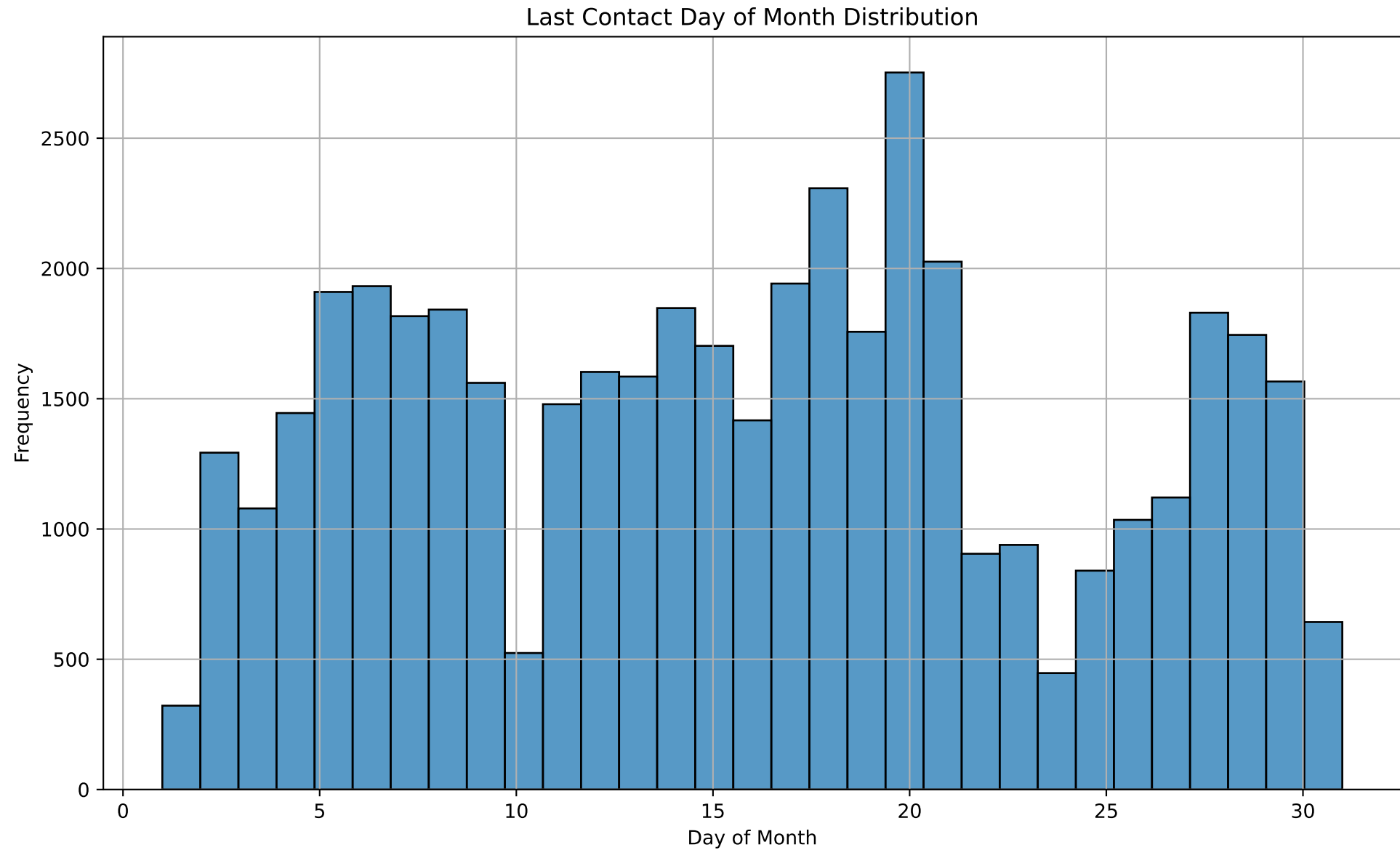
Balance Summary Statistics:

count	45216.000000
mean	1362.277844
std	3044.609674
min	-8019.000000
25%	72.000000
50%	448.500000
75%	1428.000000
max	102127.000000

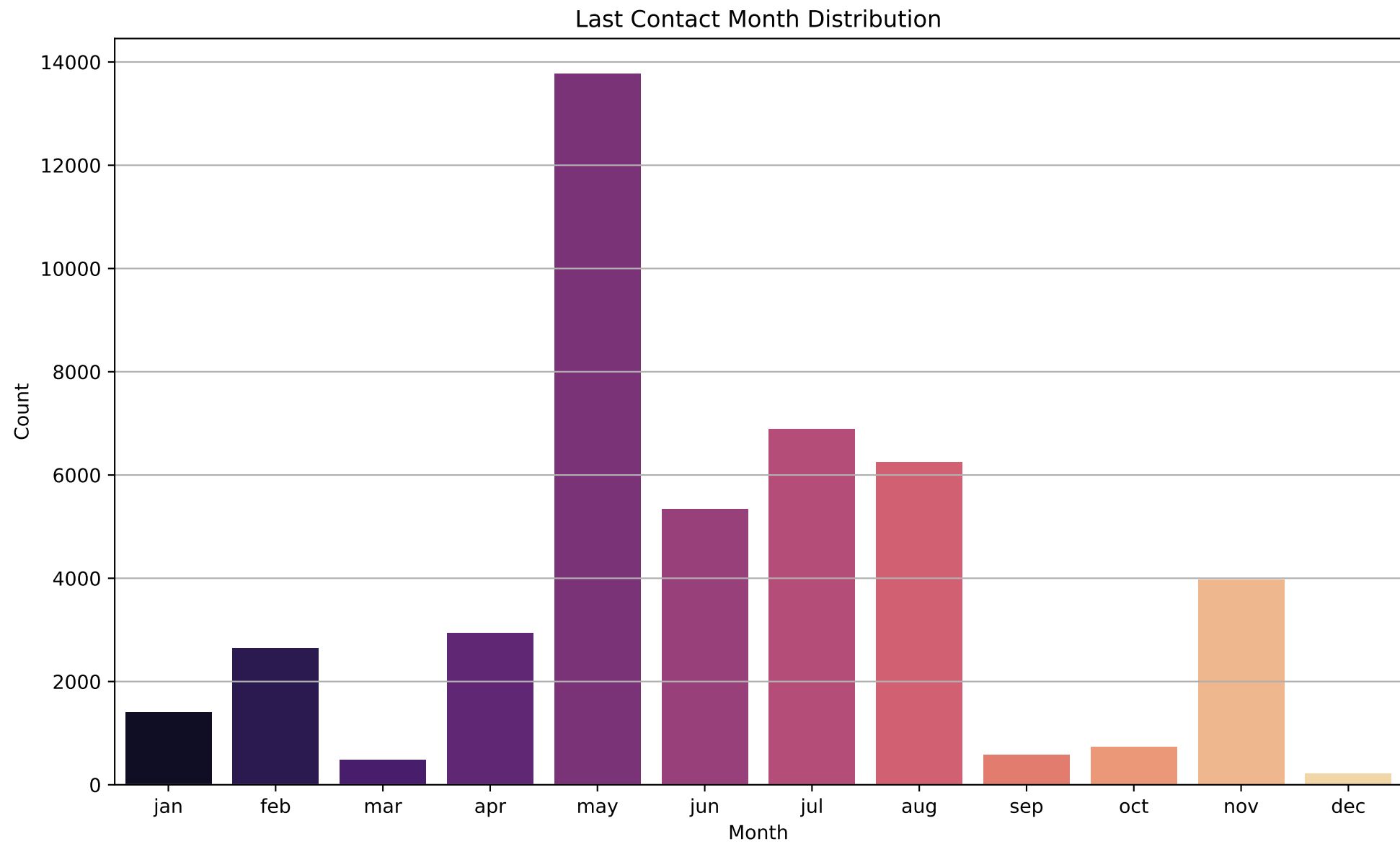
Communication Type Distribution



Last Contact Day of Month Distribution

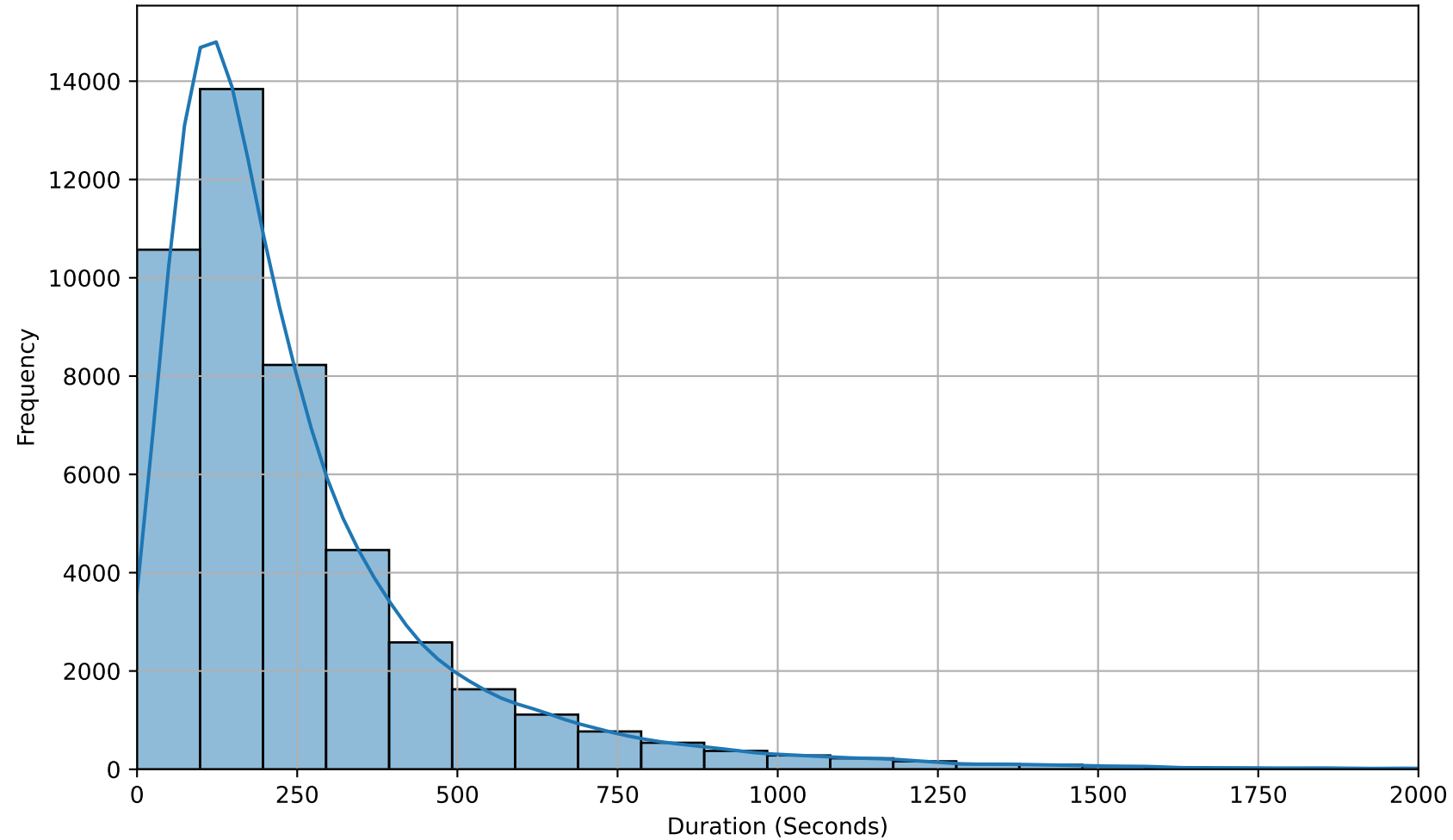


Last Contact Month Distribution



Last Contact Duration Distribution

Last Contact Duration Distribution



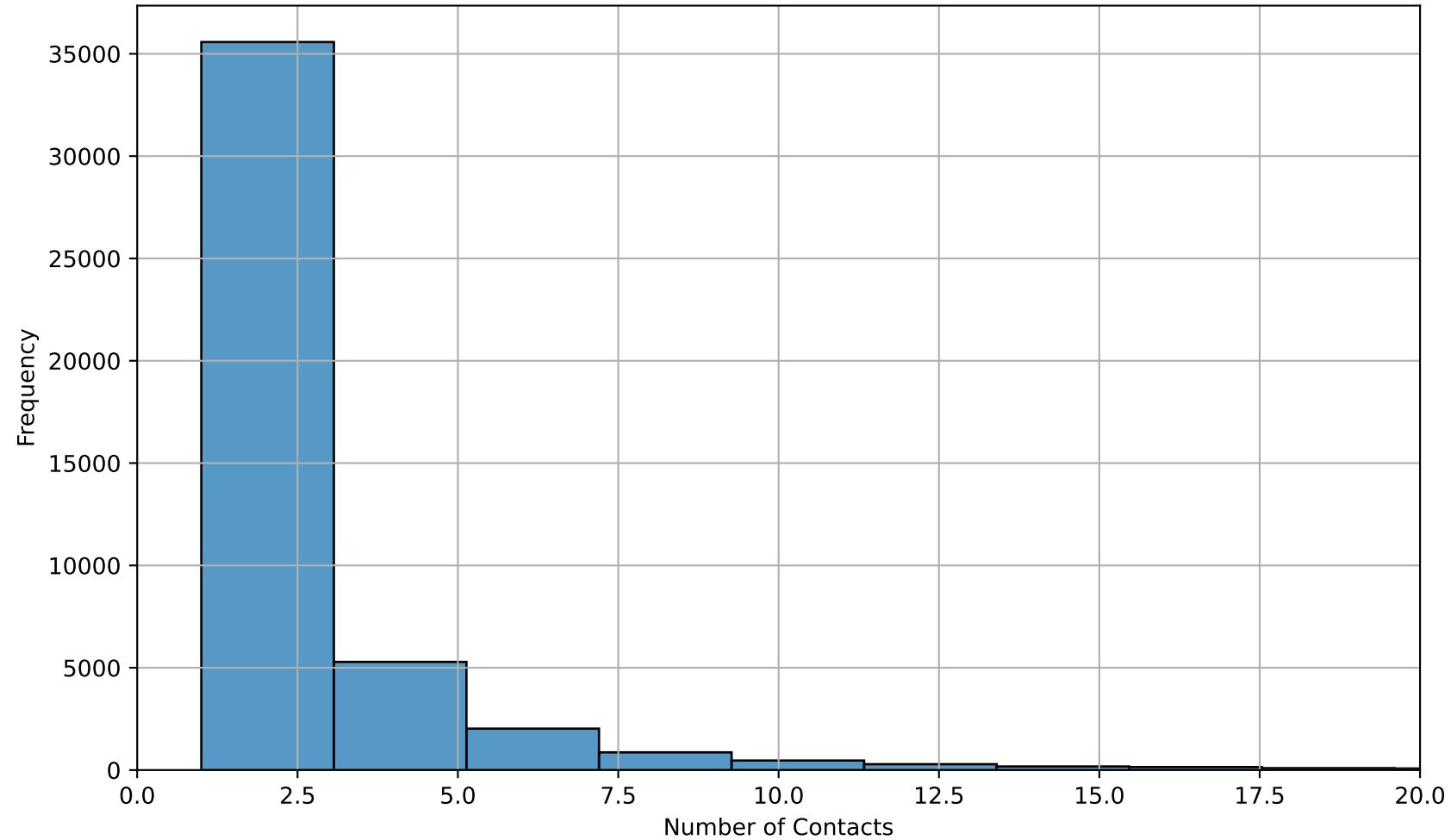
Duration Summary Statistics

Duration Summary Statistics:

count	45216.000000
mean	258.166202
std	257.515482
min	0.000000
25%	103.000000
50%	180.000000
75%	319.000000
max	4918.000000

Number of Contacts During Campaign

Number of Contacts During Campaign



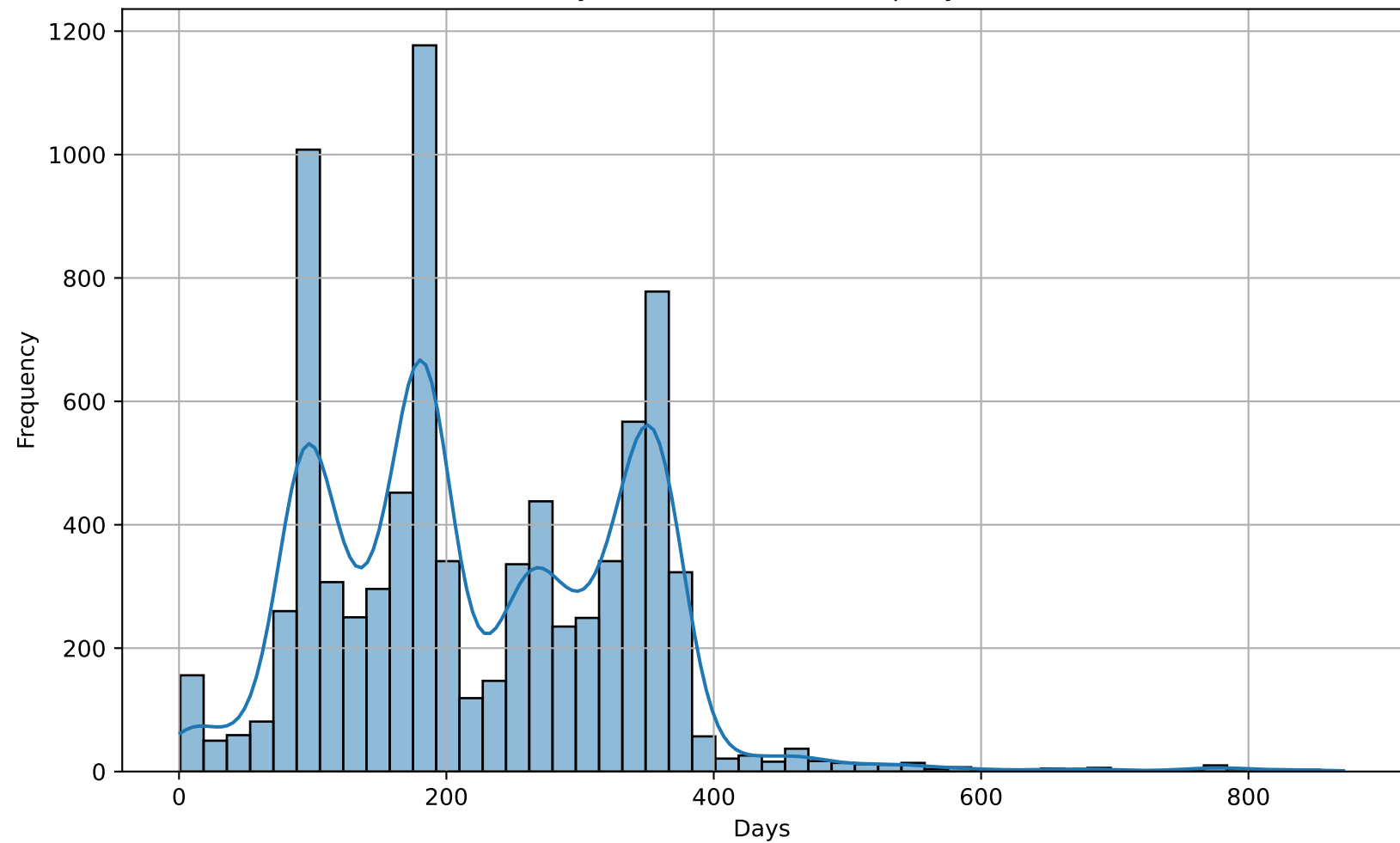
Campaign Contacts Summary

Campaign Contacts Summary:

count	45216.000000
mean	2.763668
std	3.097896
min	1.000000
25%	1.000000
50%	2.000000
75%	3.000000
max	63.000000

Days Since Last Contact (pdays)

Days Since Last Contact (pdays)

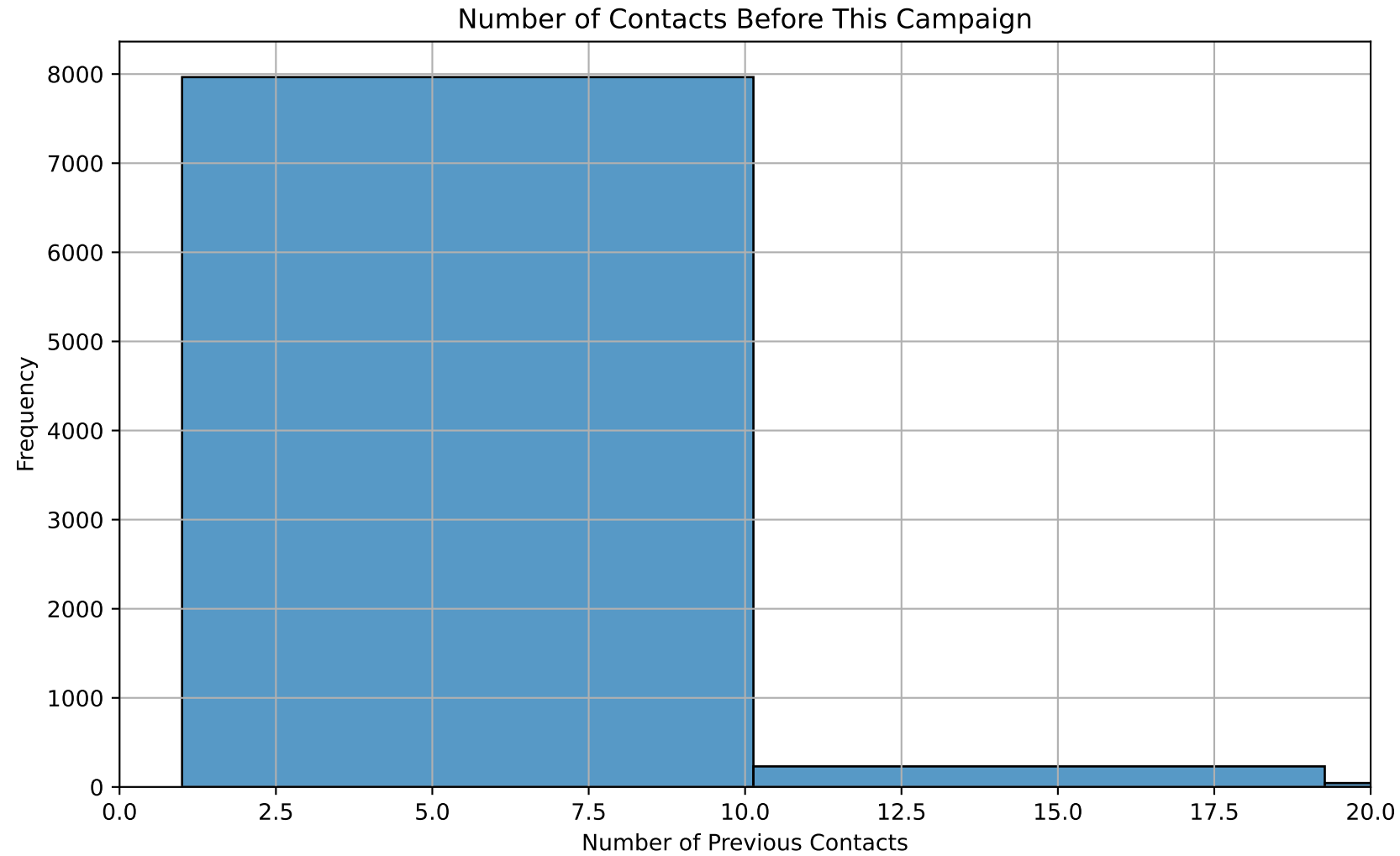


Pdays Summary

Pdays Summary (excluding clients not previously contacted):

count	8260.000000
mean	224.545884
std	115.342569
min	1.000000
25%	133.000000
50%	194.000000
75%	327.000000
max	871.000000

Number of Contacts Before This Campaign

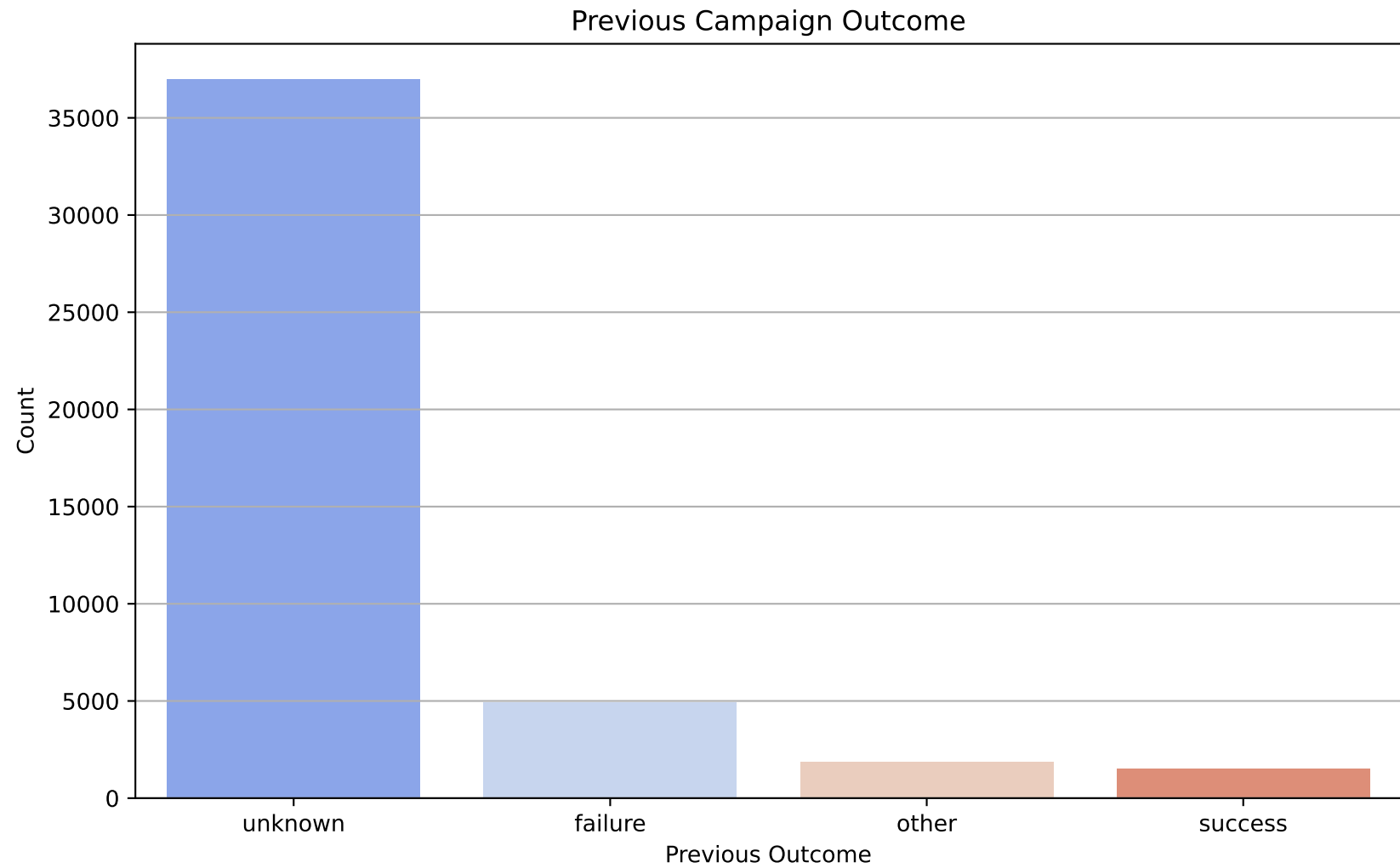


Previous Contacts Summary

Previous Contacts Summary (for clients with previous contact):

count	8260.000000
mean	3.178571
std	4.560415
min	1.000000
25%	1.000000
50%	2.000000
75%	4.000000
max	275.000000

Previous Campaign Outcome



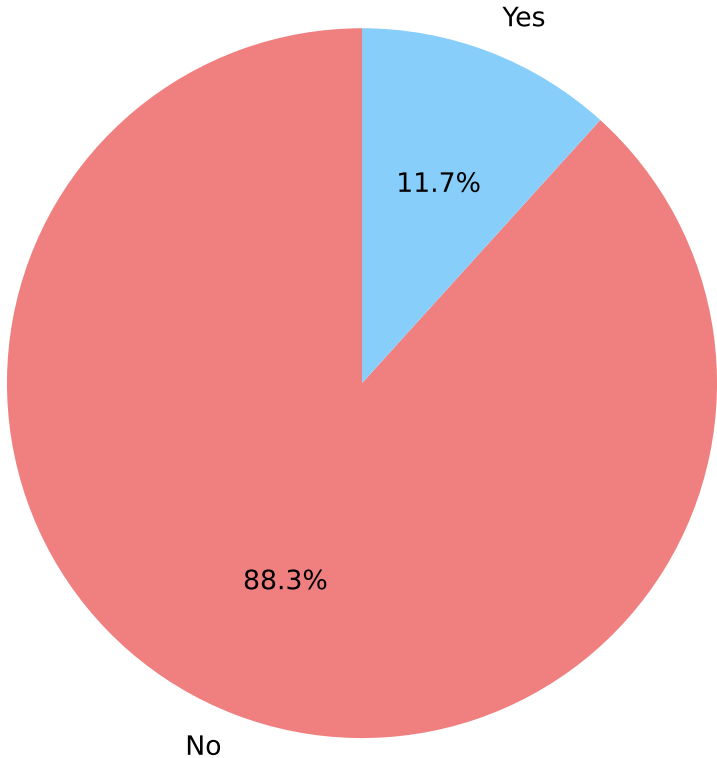
Previous Campaign Outcome

Previous Campaign Outcome (%):

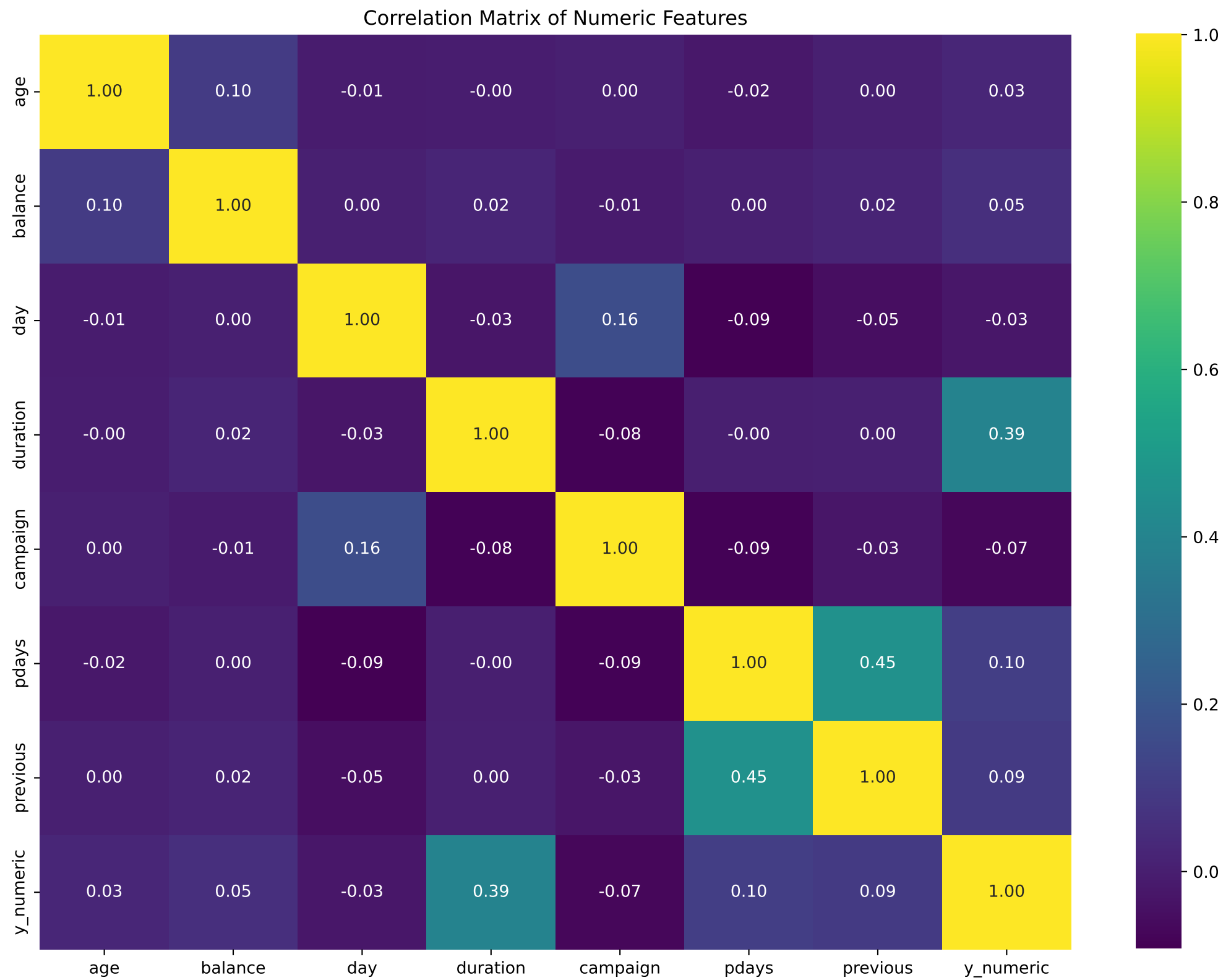
poutcome	
unknown	81.743188
failure	10.841295
other	4.069356
success	3.346161

Subscription to Term Deposit

Subscription to Term Deposit (Target Variable)



Correlation Matrix



Correlation Matrix Details

Correlation with Target Variable (y_numeric):

y_numeric	1.000000
duration	0.394387
pdays	0.103699
previous	0.093576
balance	0.052821
age	0.025648
day	-0.028307
campaign	-0.073294

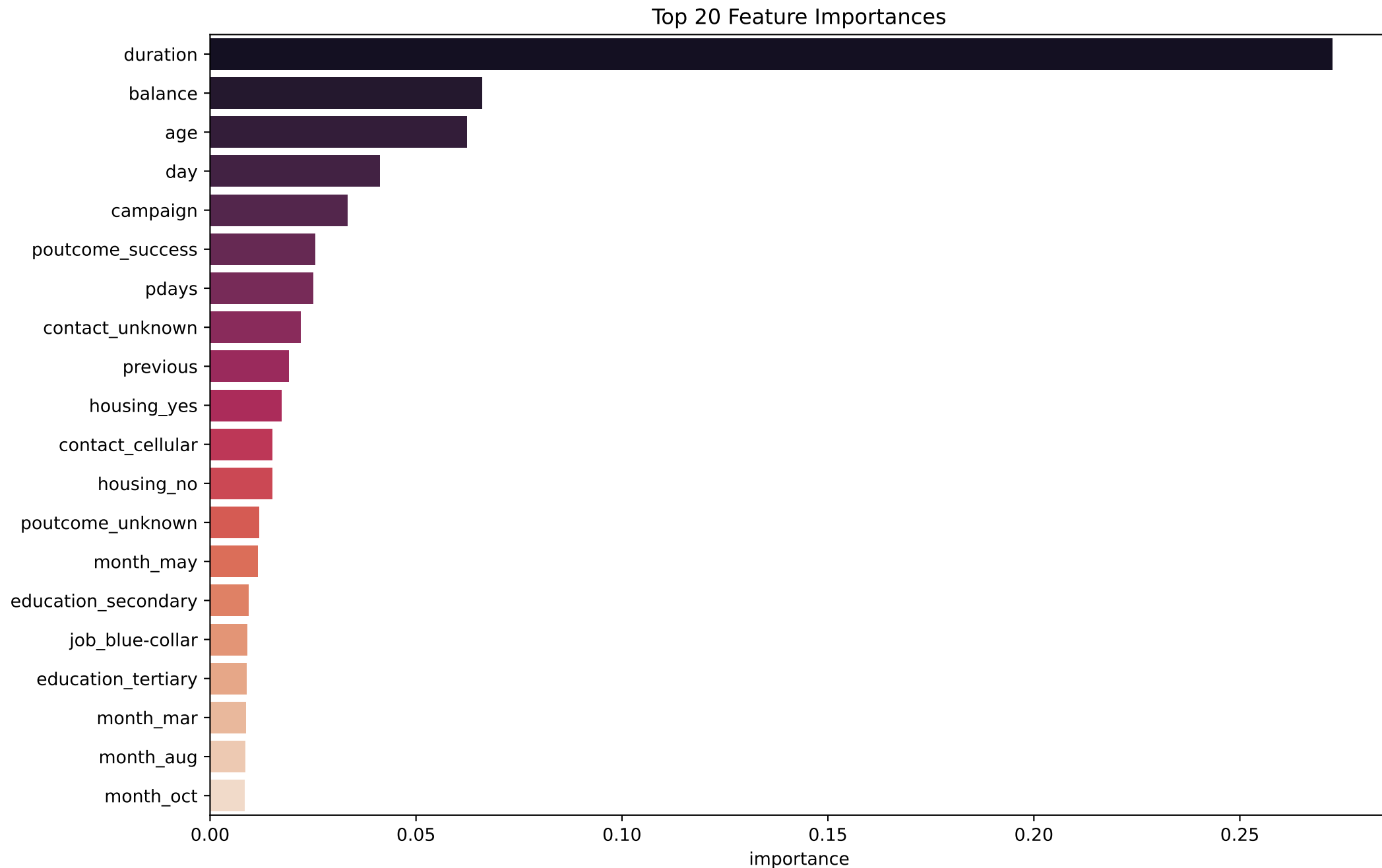
Model Evaluation Metrics

Classification Report:

	precision	recall	f1-score	support
0	0.91	0.99	0.95	7985
1	0.72	0.28	0.41	1059
accuracy			0.90	9044
macro avg	0.81	0.63	0.68	9044
weighted avg	0.89	0.90	0.88	9044

ROC AUC Score: 0.9245

Top 20 Feature Importances



Feature Importances

Top 20 Most Important Features:

	feature	importance
3	duration	0.272455
1	balance	0.065988
0	age	0.062305
2	day	0.041230
4	campaign	0.033371
372	poutcome_success	0.025561
5	pdays	0.025030
40	contact_unknown	0.022059
6	previous	0.019121
35	housing_yes	0.017344
38	contact_cellular	0.015113
34	housing_no	0.015066
373	poutcome_unknown	0.011948
49	month_may	0.011664
28	education_secondary	0.009310
8	job_blue-collar	0.009005
29	education_tertiary	0.008947
48	month_mar	0.008761
42	month_aug	0.008517
51	month_oct	0.008330