Data Glacier

Your Deep Learning Partner

Final Project Report

Bank Marketing (Campaign)

Group Name: Data Visionaries
Name: Abdukhakimov Asatilla
Email: tremendous54@gmail.com
Country: Uzbekistan
Company: Data Glacier
Specialization: Data Science
Github: https://github.com/asat94/Data-Glacier-Internship

2025.04.27

Agenda

PROBLEM STATEMENT

BUSINESS UNDERSTANDING

DATA EXPLORATION

EDA

MODELLING

DASHBOARD

SUMMARY

Problem Statement

- ABC Bank wants to sell its term deposit product to customers. Before launching the product, the bank aims to develop a model to understand whether a particular customer will buy their product or not, based on the customer's past interaction with the bank or other financial institutions.

Business Understanding

Objective: Use Machine Learning to predict customers likely to subscribe to term deposits.

Goal: Optimize marketing efforts, focusing resources on high-probability customers.

Benefits:

Increased conversion rates

Cost savings by minimizing wasted marketing efforts

Two Model Scenarios:

With 'Duration' Feature: Higher accuracy, but complex and not ideal for pre-call campaigns.

Without 'Duration' Feature: Simpler, more practical for operational use.

Focus: Balance technical performance and business feasibility for a transparent, effective model aligned with marketing goals.

Project Lifecycle and Deadlines

Week	Deadline	Assignment
Week 7	2025.03.19	Business Understanding
Week 8	2025.03.26	Data Understanding
Week 9	2025.04.02	Data Cleansing and Transformation
Week 10	2025.04.09	EDA
Week 11	2025.04.16	EDA Presentation and proposed modeling technique
Week 12	2025.04.23	Model Selection and Model Building/Dashboard
Week 13	2025.04.30	Final Project Report and Code

Data Exploration

Bank-full data

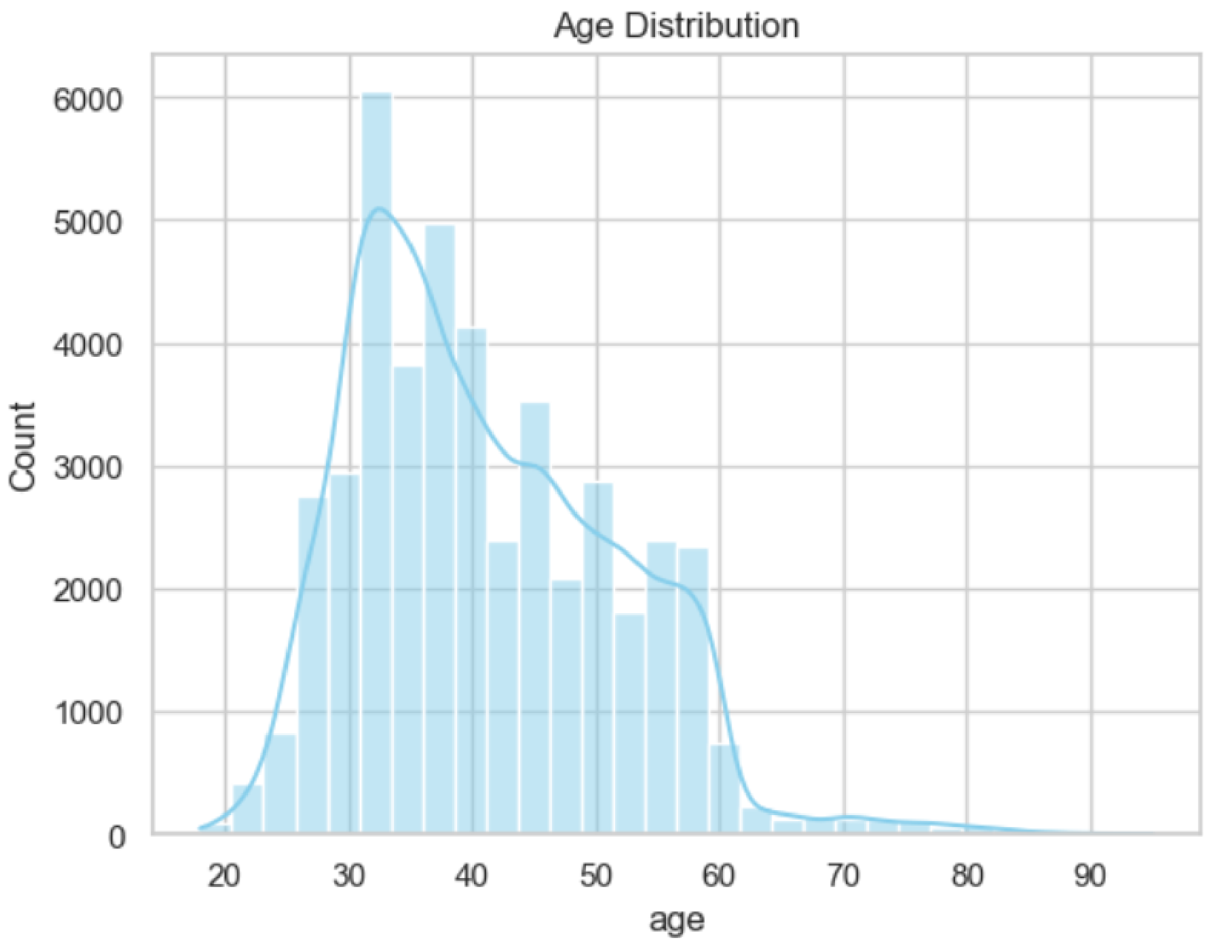
Total Data

- 45,211 rows
- 17 columns

- Data Cleaned & Formatted
- Checked Missing Values & Outliers

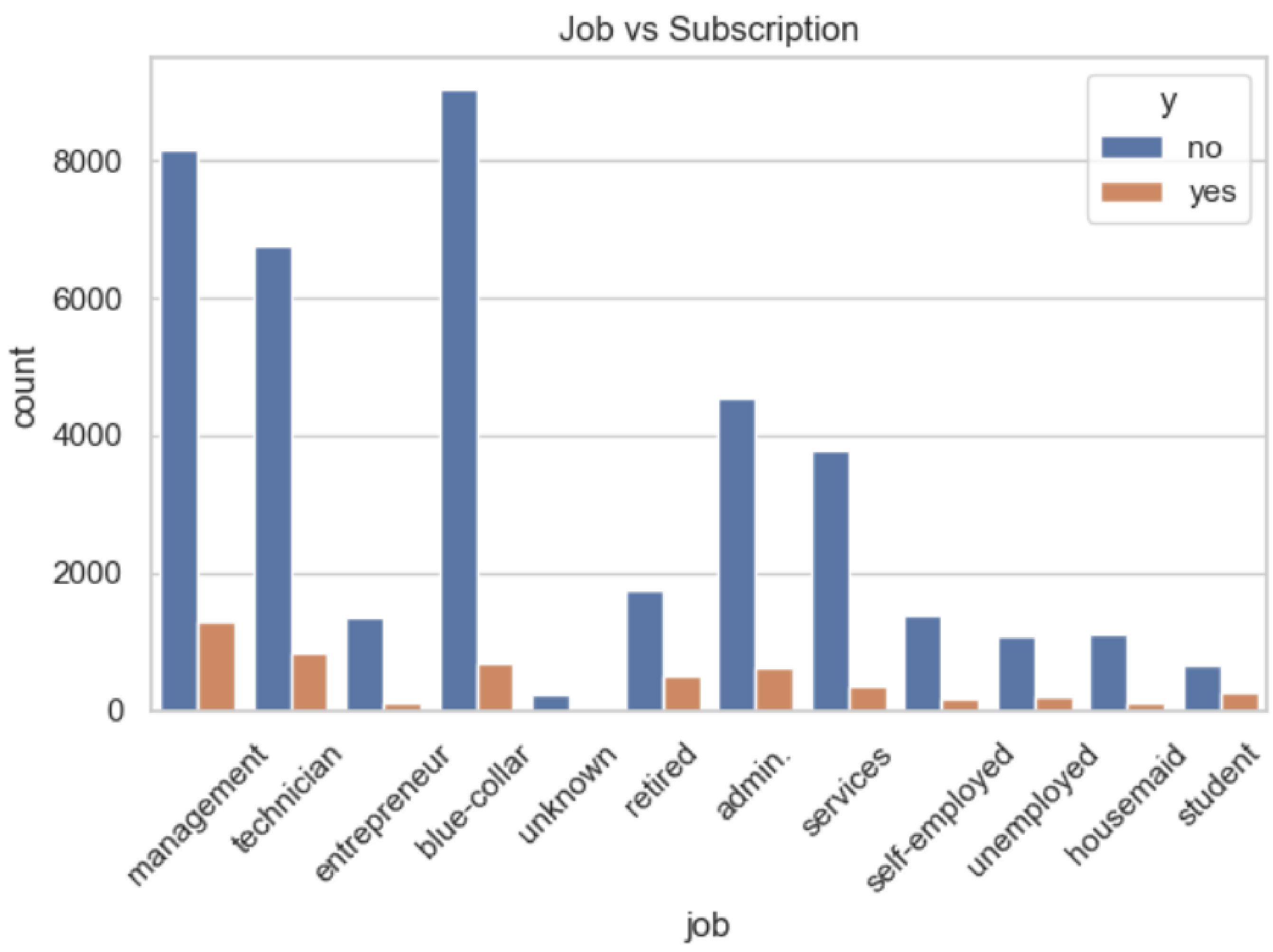
#	Column	Non-Null Count	Dtype
---	------	--------------	-----
0	age	45211 non-null	int64
1	job	45211 non-null	object
2	marital	45211 non-null	object
3	education	45211 non-null	object
4	default	45211 non-null	object
5	balance	45211 non-null	int64
6	housing	45211 non-null	object
7	loan	45211 non-null	object
8	contact	45211 non-null	object
9	day	45211 non-null	int64
10	month	45211 non-null	object
11	duration	45211 non-null	int64
12	campaign	45211 non-null	int64
13	pdays	45211 non-null	int64
14	previous	45211 non-null	int64
15	poutcome	45211 non-null	object
16	y	45211 non-null	object

Age Distribution



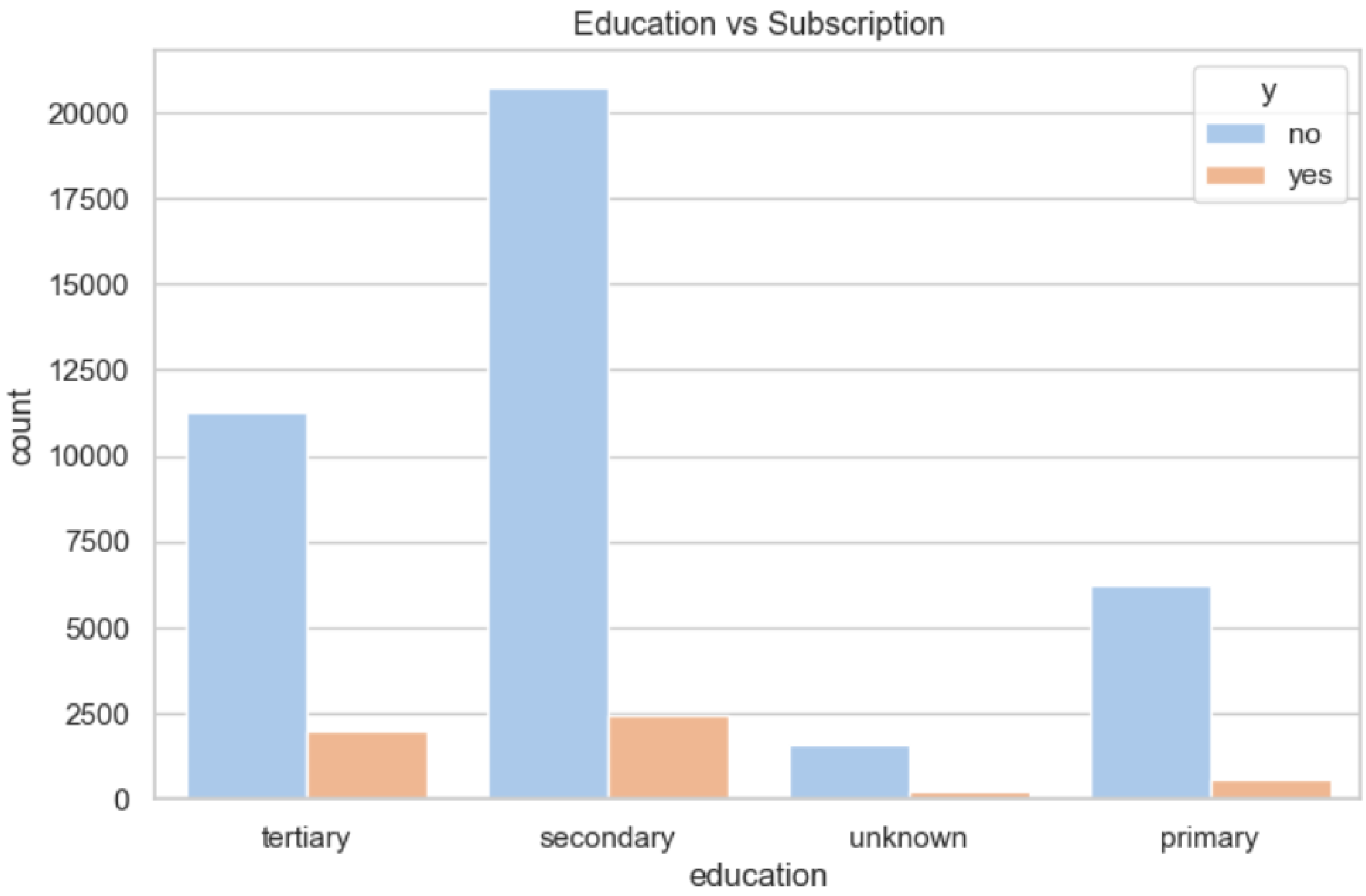
Customers are mainly between 25 to 60 years old, a key demographic for term deposits

Job vs Subscription



Management and technician roles show higher subscription rates.

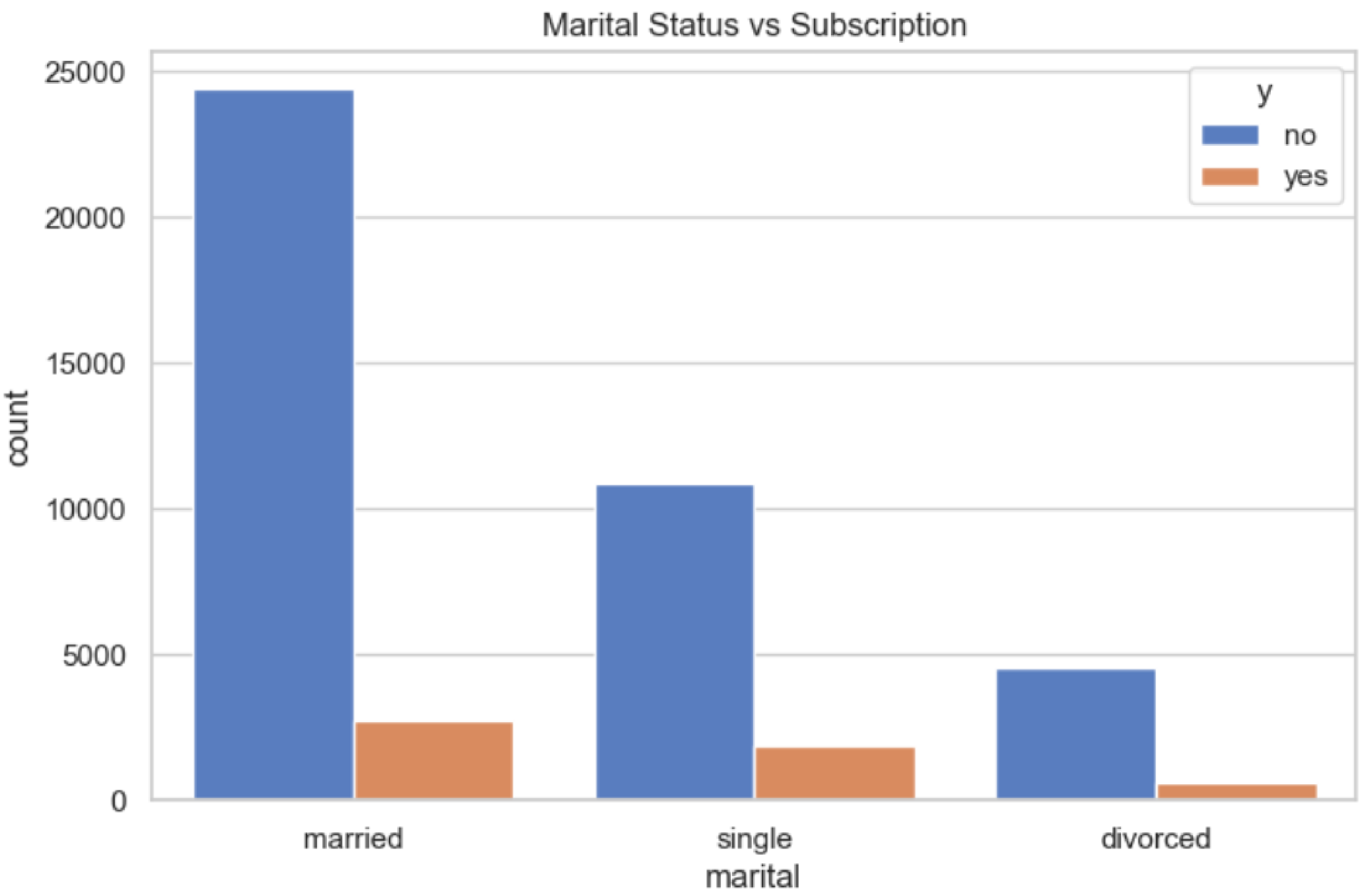
Education vs Subscription



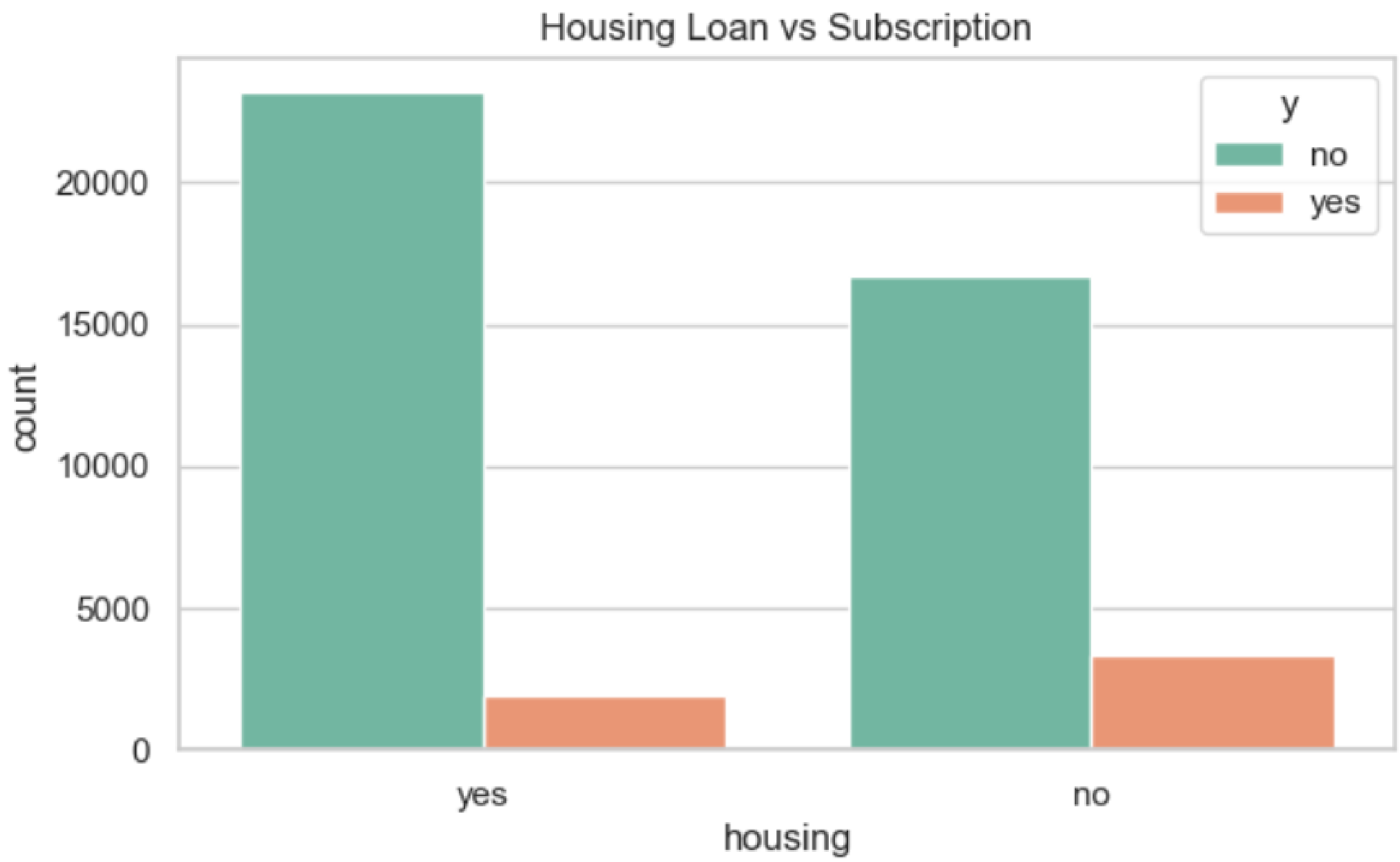
Tertiary educated customers are more likely to subscribe.

Marital Status vs Subscription

Single customers tend to have higher subscription rates than married/divorced

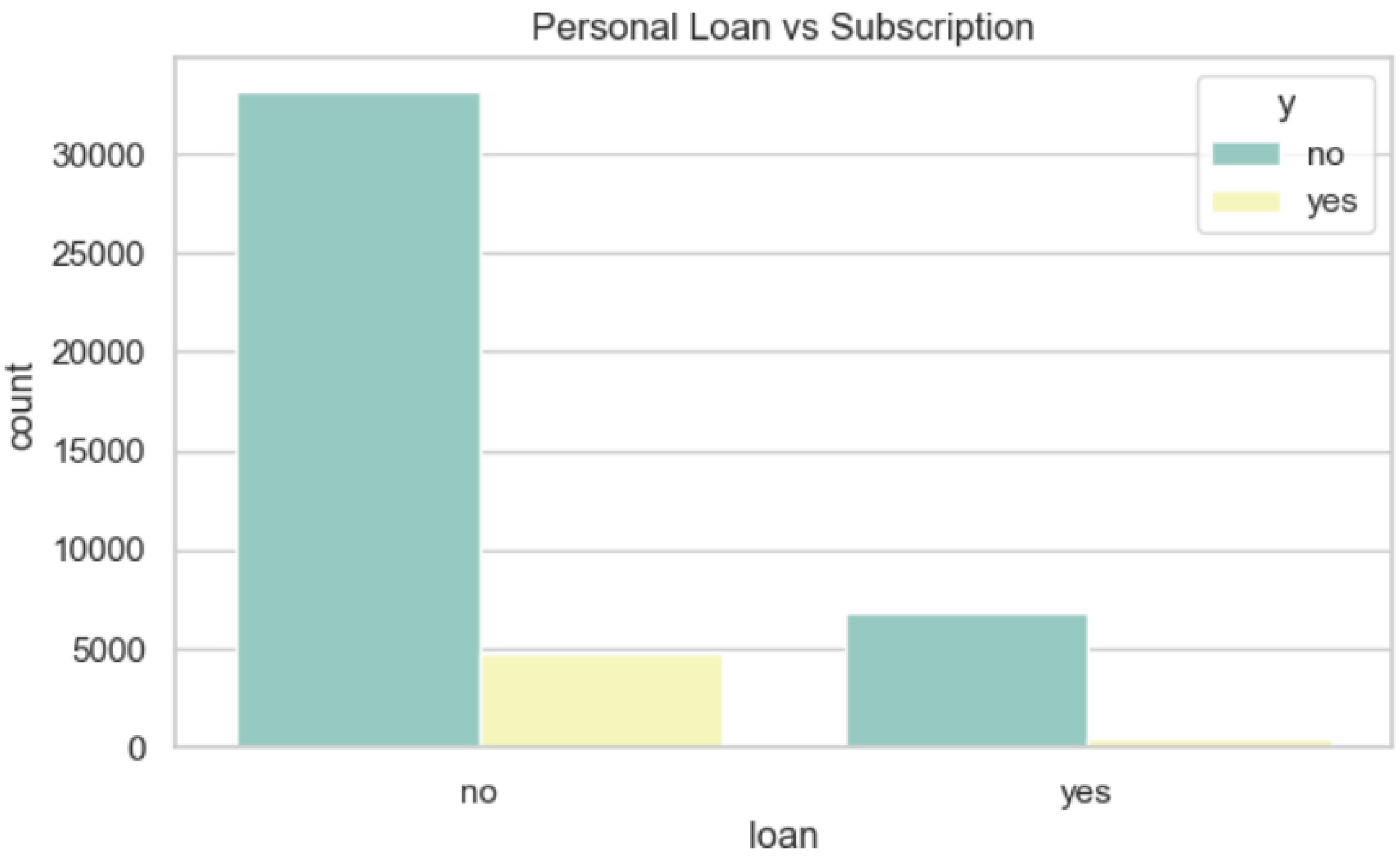


Housing Loan Status



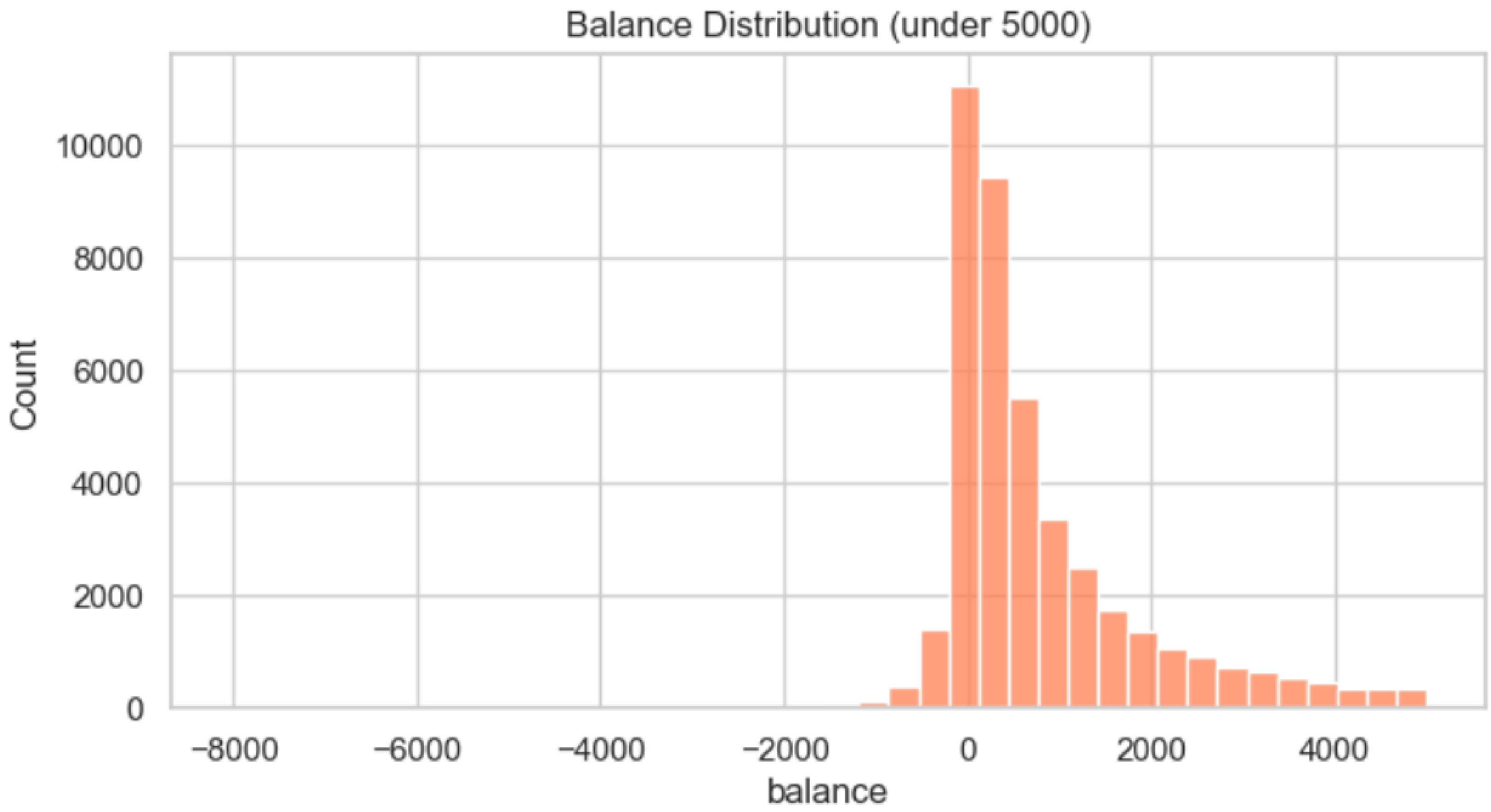
Customers without housing loans appear slightly more likely to subscribe

Personal Loan Status



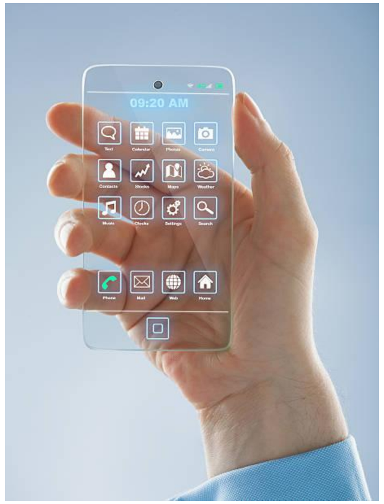
Those without personal loans show better subscription interest

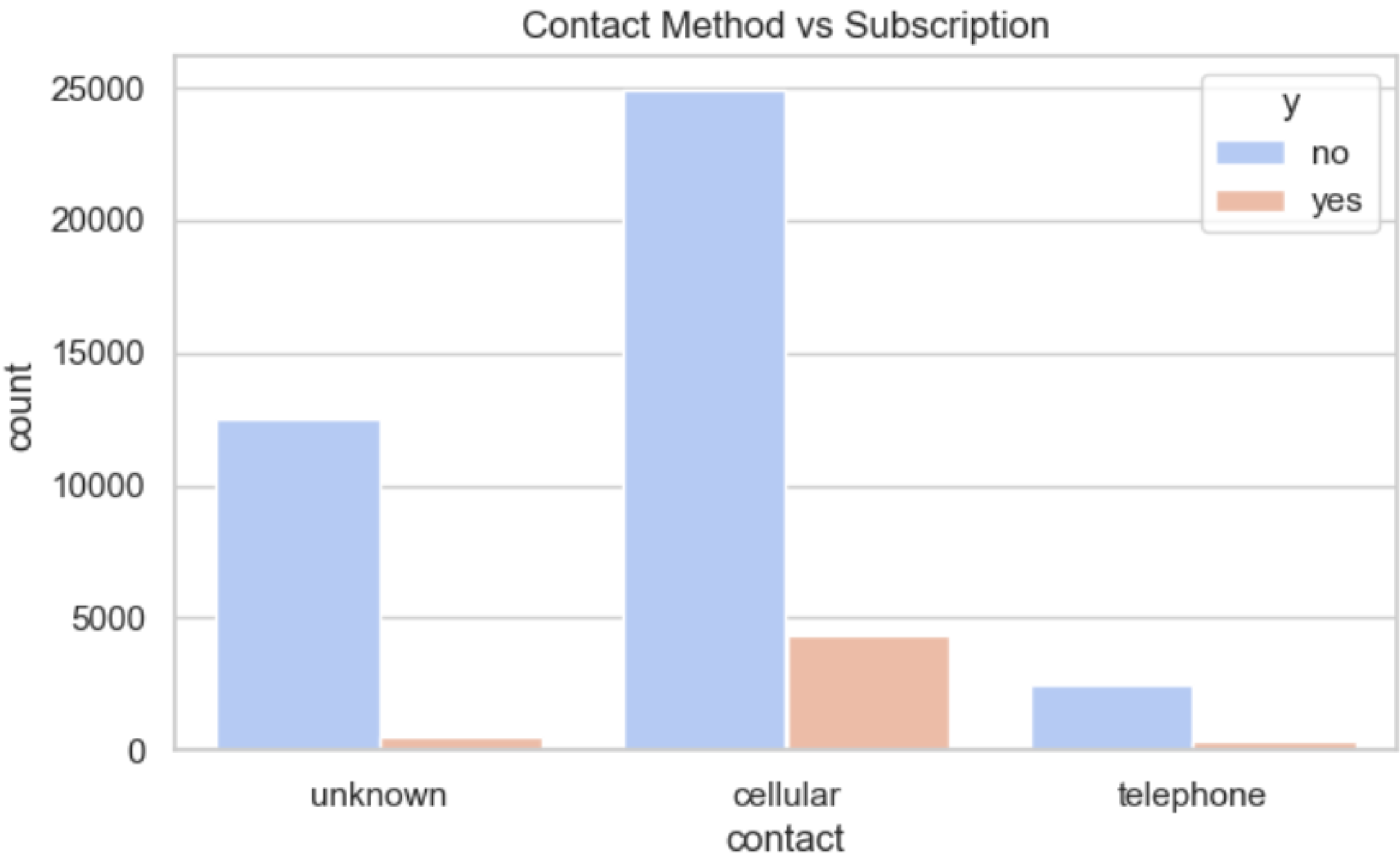
Balance Distribution (Zoomed In)



Most customers have balances under 5000, indicating a middle-income audience

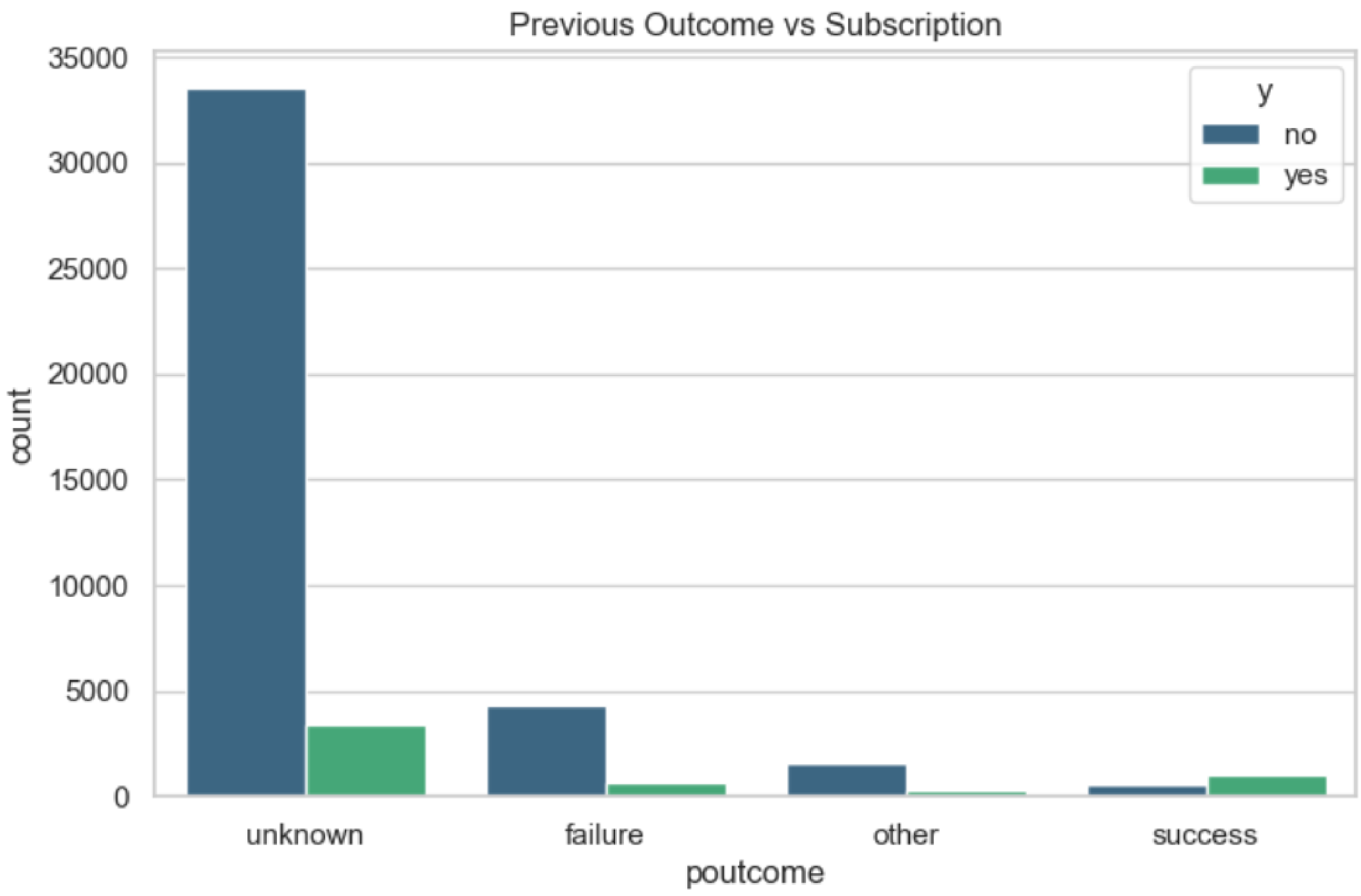
Contact Method vs Subscription





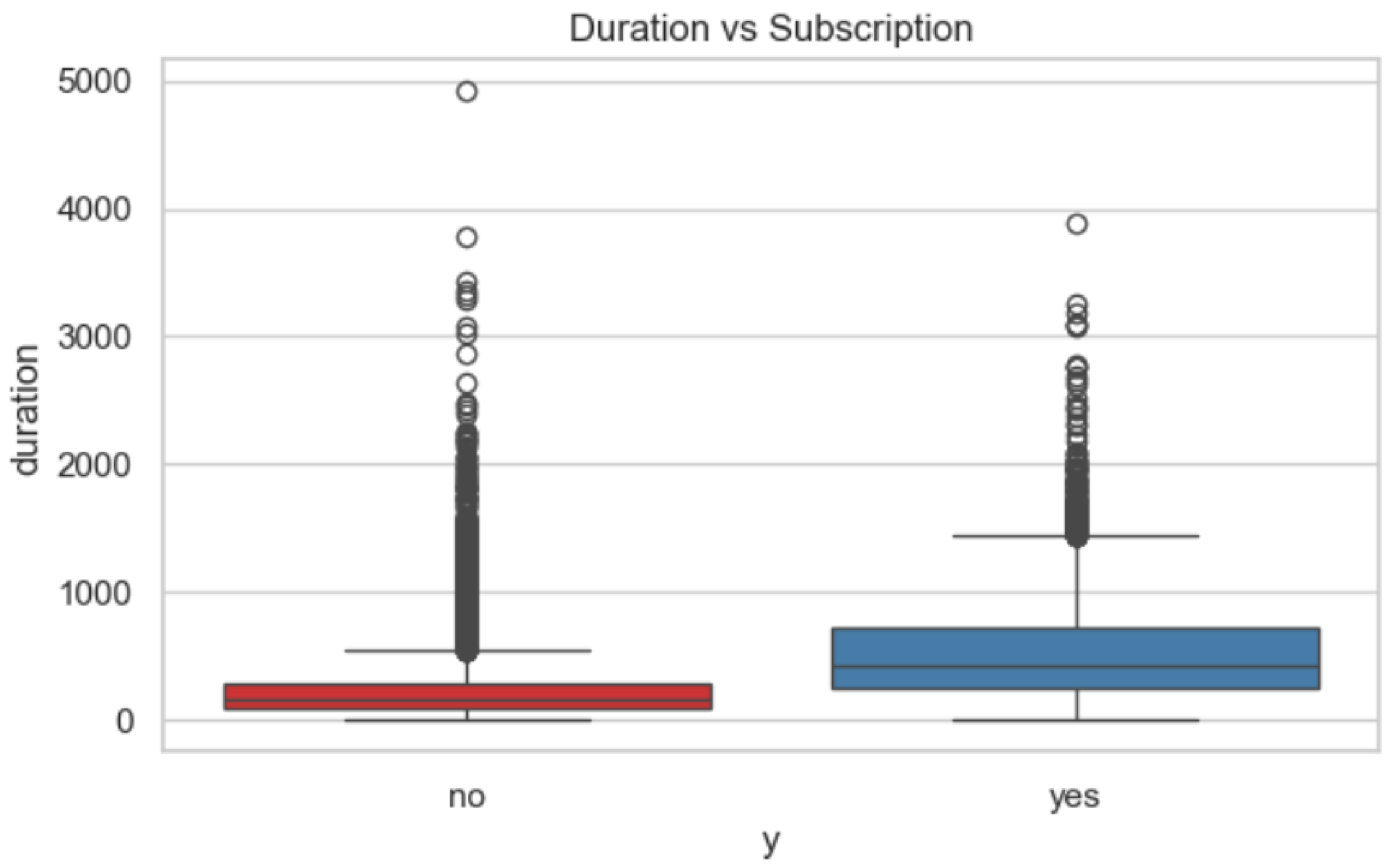
Customers contacted via cellular show better subscription rates

Previous outcome vs Subscription



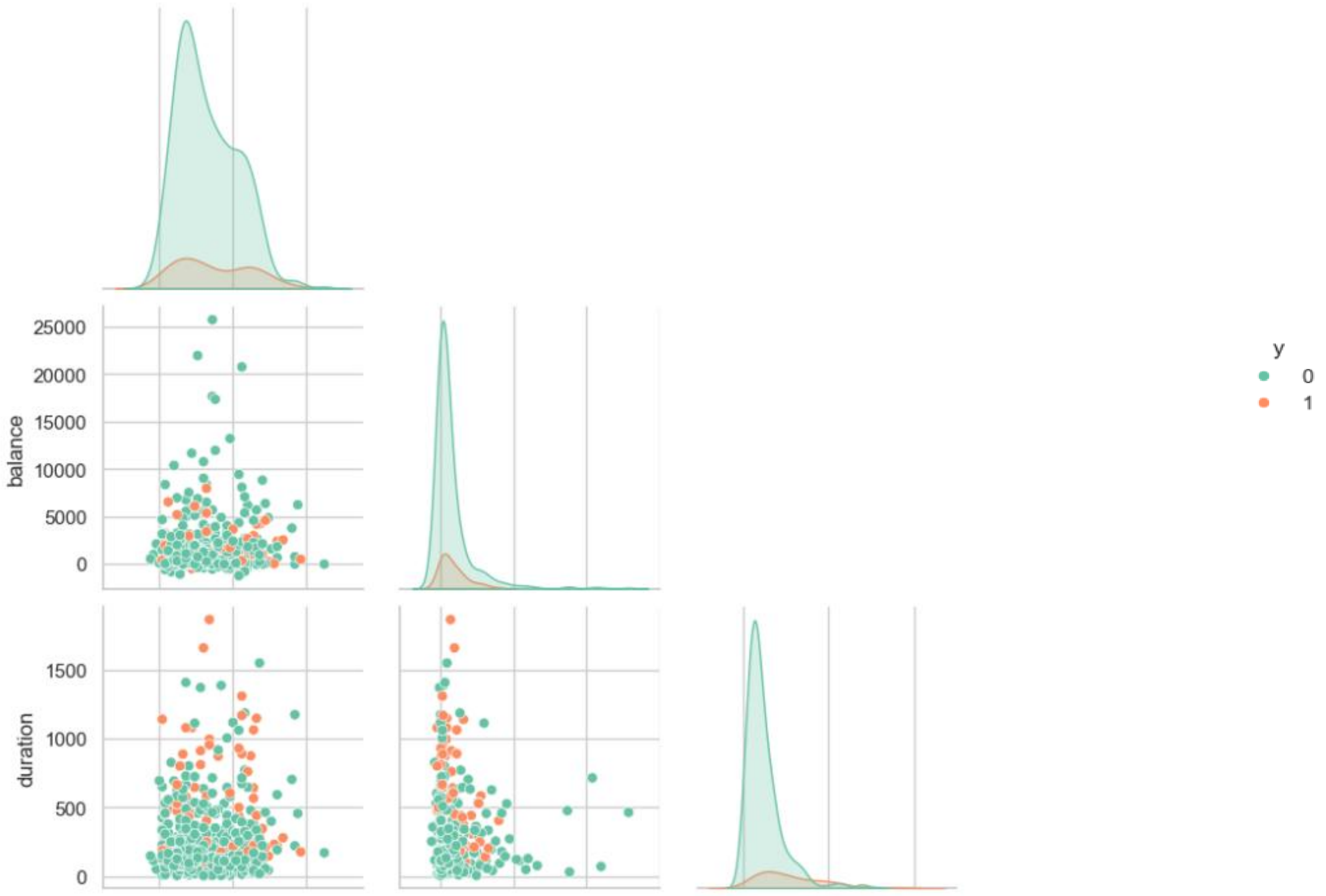
Successful outcomes in prior campaigns drastically improve subscription chances

Duration vs Subscription

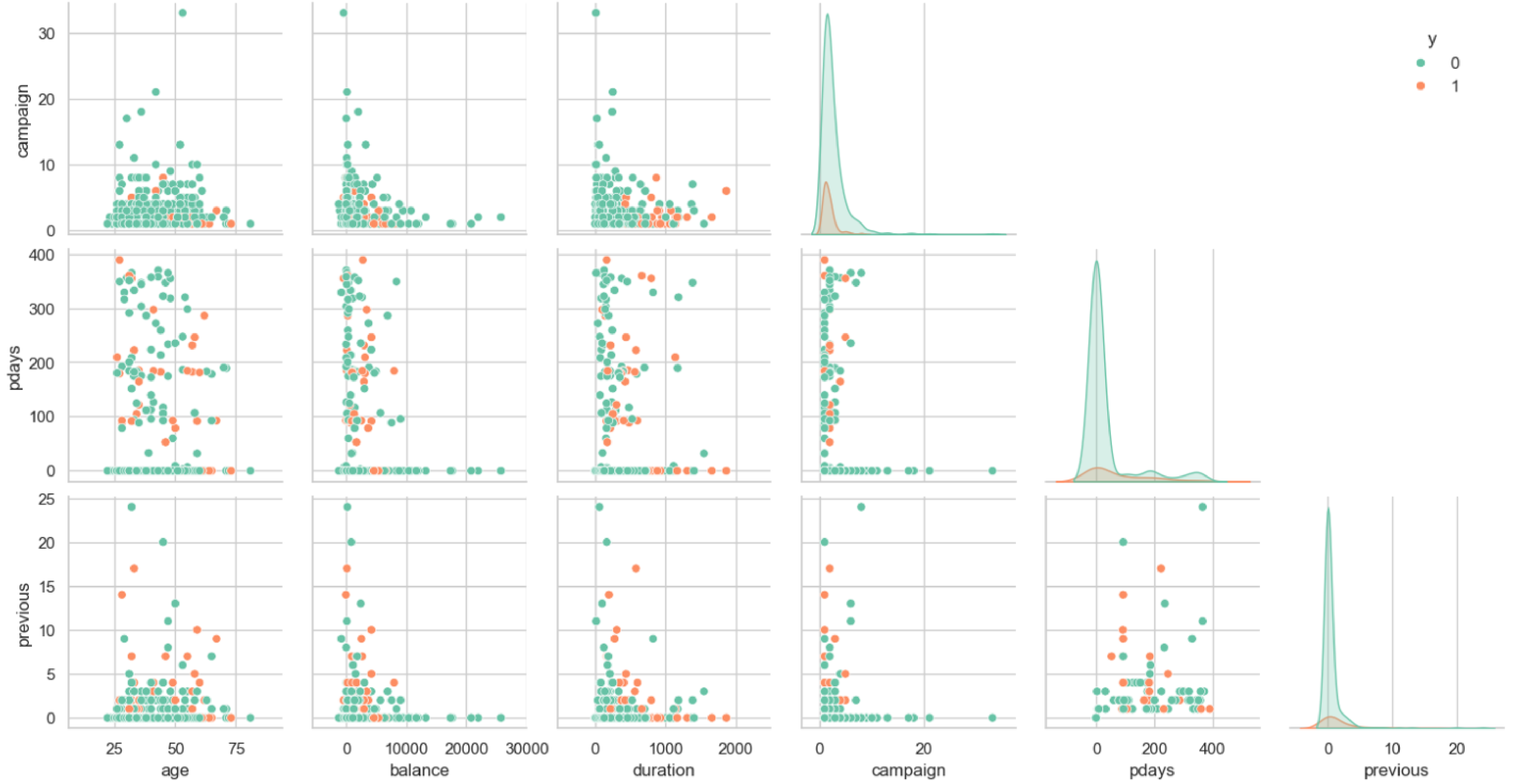


<u>Longer call durations</u> often lead to a 'yes' decision — indicating interest builds over time.

Pair Plot (Scatter Matrix)



Pair Plot (Scatter Matrix)



Modelling

=== LR with duration ===

	precision	recall	f1-score	support
0	0.91	0.98	0.94	7952
1	0.64	0.30	0.41	1091
accuracy			0.90	9043
macro avg	0.78	0.64	0.68	9043
weighted avg	0.88	0.90	0.88	9043

=== RF with duration ===

	precision	recall	f1-score	support
0	0.92	0.97	0.94	7952
1	0.63	0.41	0.49	1091
accuracy			0.90	9043
macro avg	0.78	0.69	0.72	9043
weighted avg	0.89	0.90	0.89	9043

=== LR w/o duration ===

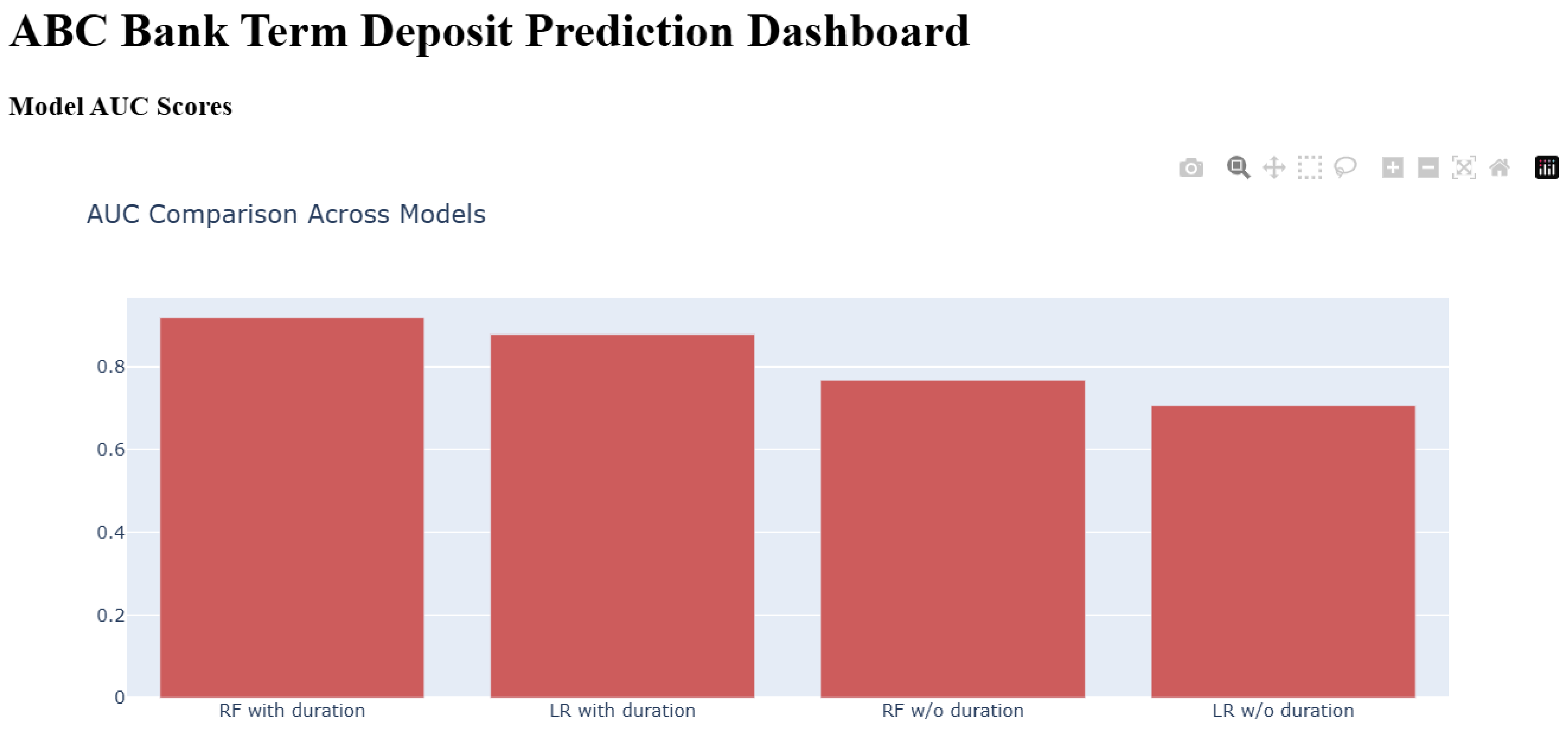
	precision	recall	f1-score	support
0	0.90	0.99	0.94	7952
1	0.70	0.16	0.26	1091
accuracy			0.89	9043
macro avg	0.80	0.57	0.60	9043
weighted avg	0.87	0.89	0.86	9043

=== RF w/o duration ===

	precision	recall	f1-score	support
0	0.90	0.99	0.94	7952
1	0.65	0.20	0.31	1091
accuracy			0.89	9043
macro avg	0.78	0.59	0.63	9043
weighted avg	0.87	0.89	0.86	9043

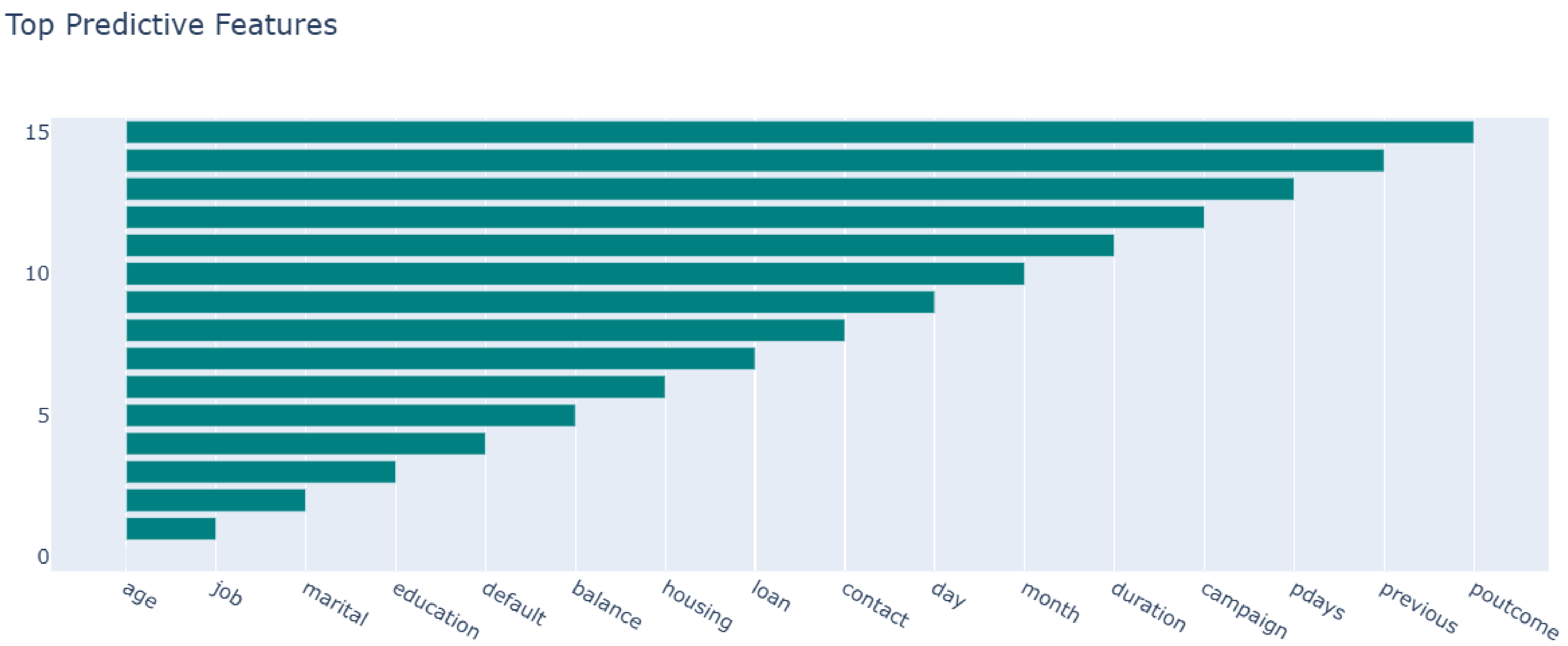
- Including the **duration** variable significantly improves performance across all models.
- **Random Forest with duration** achieves the best overall performance

ABC Bank Term Deposit Prediction Dashboard



Random Forest with Duration

Random Forest (with Duration) - Feature Importances



Select Model for ROC Curve

Select Model for ROC Curve

RF with duration

Summary

Objective: Predict customer subscription using past interaction data.

Key Findings:

- *Higher subscription rates among management/technician roles and tertiary-educated customers.*
- *Longer call duration and positive past outcomes strongly linked to subscriptions.*
- *Subscribed customers often have higher balances.*

Model Performance:

Random Forest with 'duration' achieved top ROC-AUC (0.918).

Conclusion:

ABC Bank can enhance targeting and maximize marketing ROI using the Random Forest model with key customer insights.

THANK YOU

Data Glacier

Your Deep Learning Partner