Group 7 - MBDS April 2025
Natanel Nisenbaum
Pilar Latre
Maria Bourbon
Antonio Morfin
Emma Herrera

Term deposit subscriptions based on marketing campaigns

Strategic Impact for the Bank

Current Challenge: Phone-based marketing has low conversion rates and high operational costs.

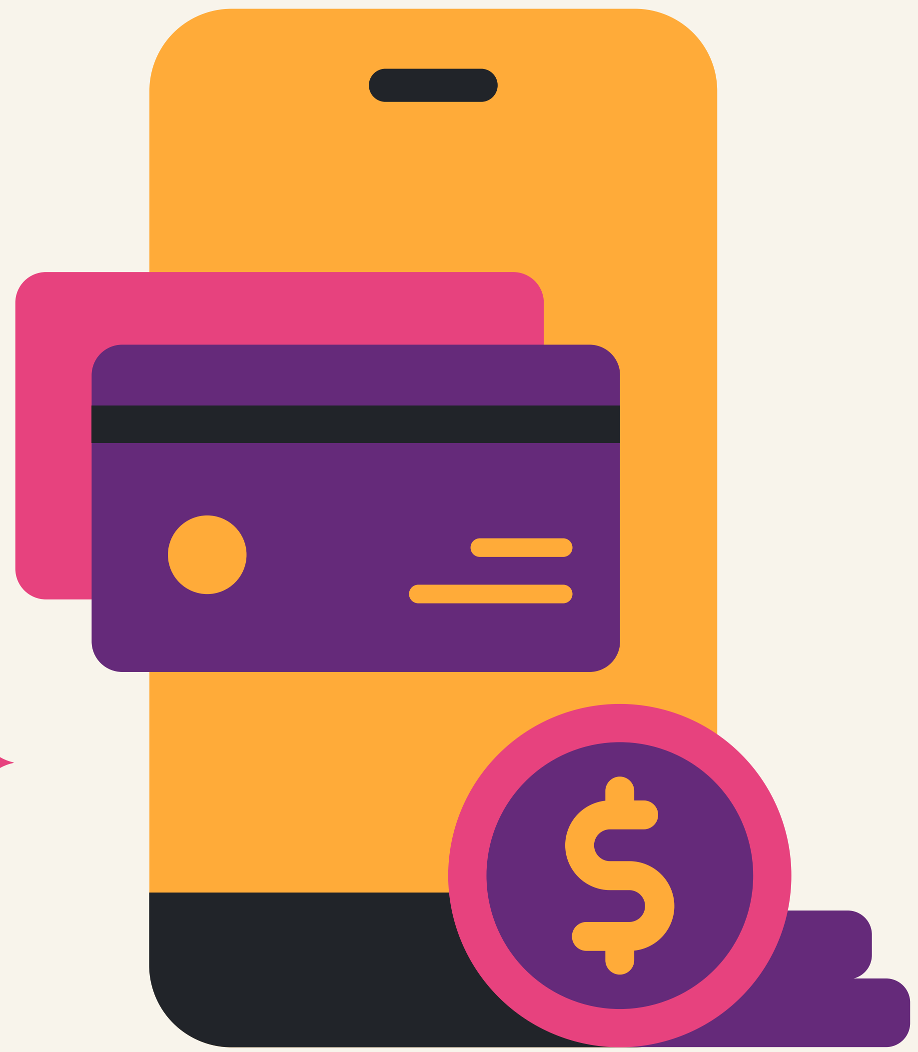
Problem: The bank often contacts the wrong clients at the wrong time.

Objective: Use predictive modeling to estimate the probability of subscription before making a call.

Expected Impact:

- Increase conversion rates by focusing on high-potential clients.
- Reduce operational costs by avoiding unpromising leads.
- Improve campaign efficiency and targeting accuracy through data-driven decisions.

Dataset Overview



Shape

- 11,162 marketing contacts
- 13 columns

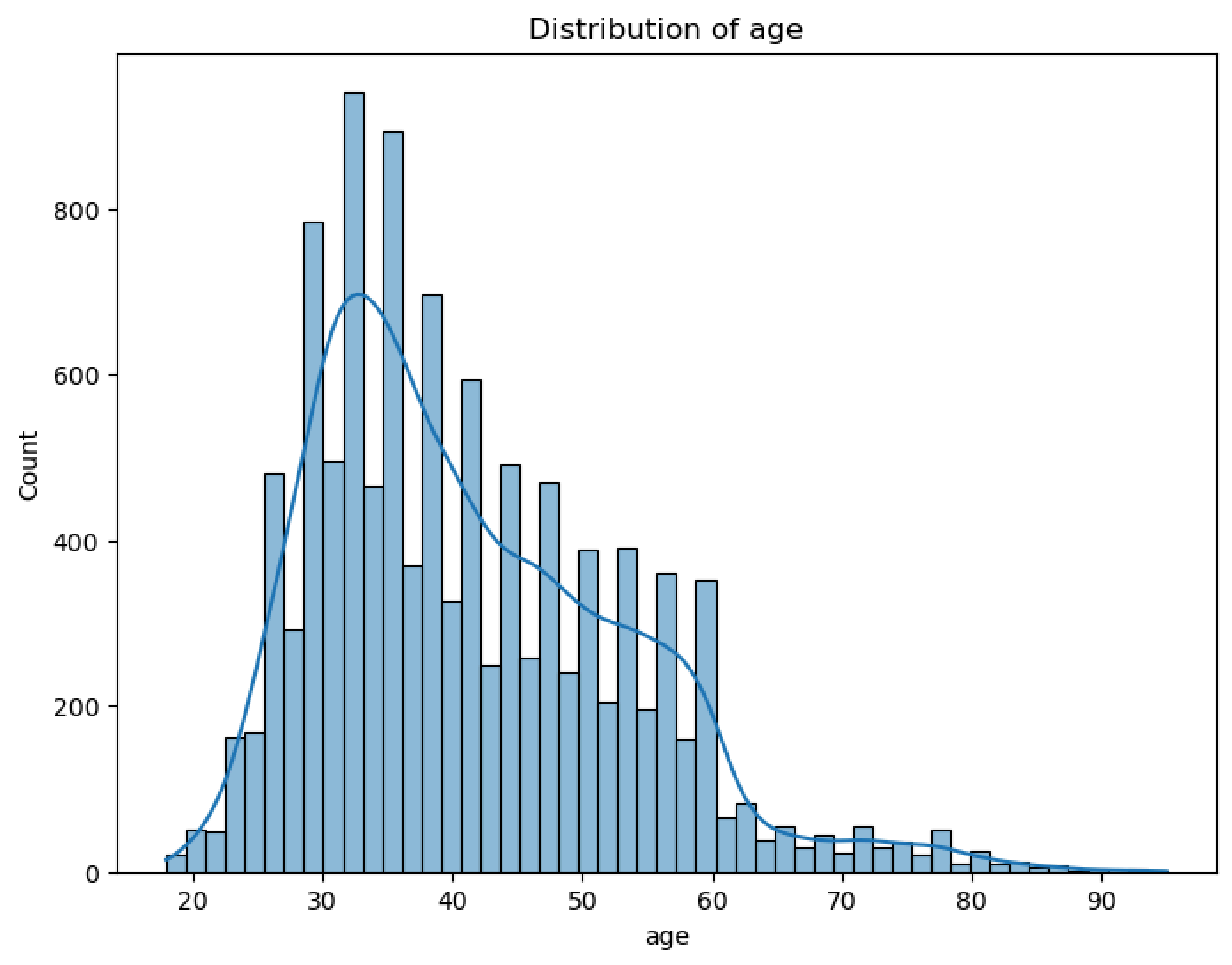
Features

- Demographics: age, job, marital, education
- Financial: balance, housing loan, personal loan
- Campaign activity: call duration, previous contacts, contact month, contact type
- Past outcomes: poutcome
- Target: deposit (1 = yes, 0 = no)

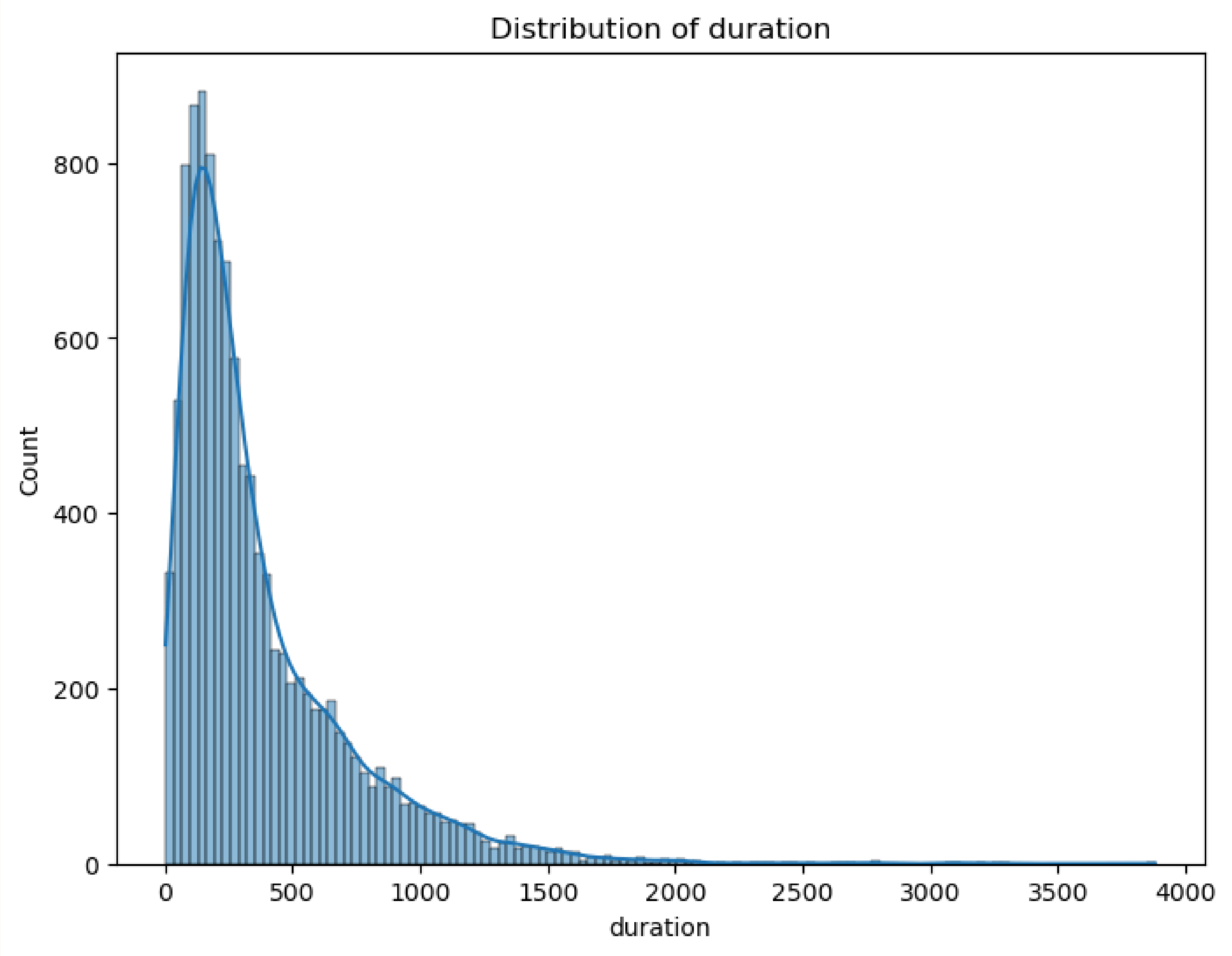
Data Cleaning & Preparation

- No missing values or duplicates
- One-Hot Encoding for all categorical features
- StandardScaler for numeric variables
- Train/Test Split: 70% training – 30% test
- Outliers kept as meaningful business signals
- Duration variable analyzed carefully (highest correlation)

Data Insights



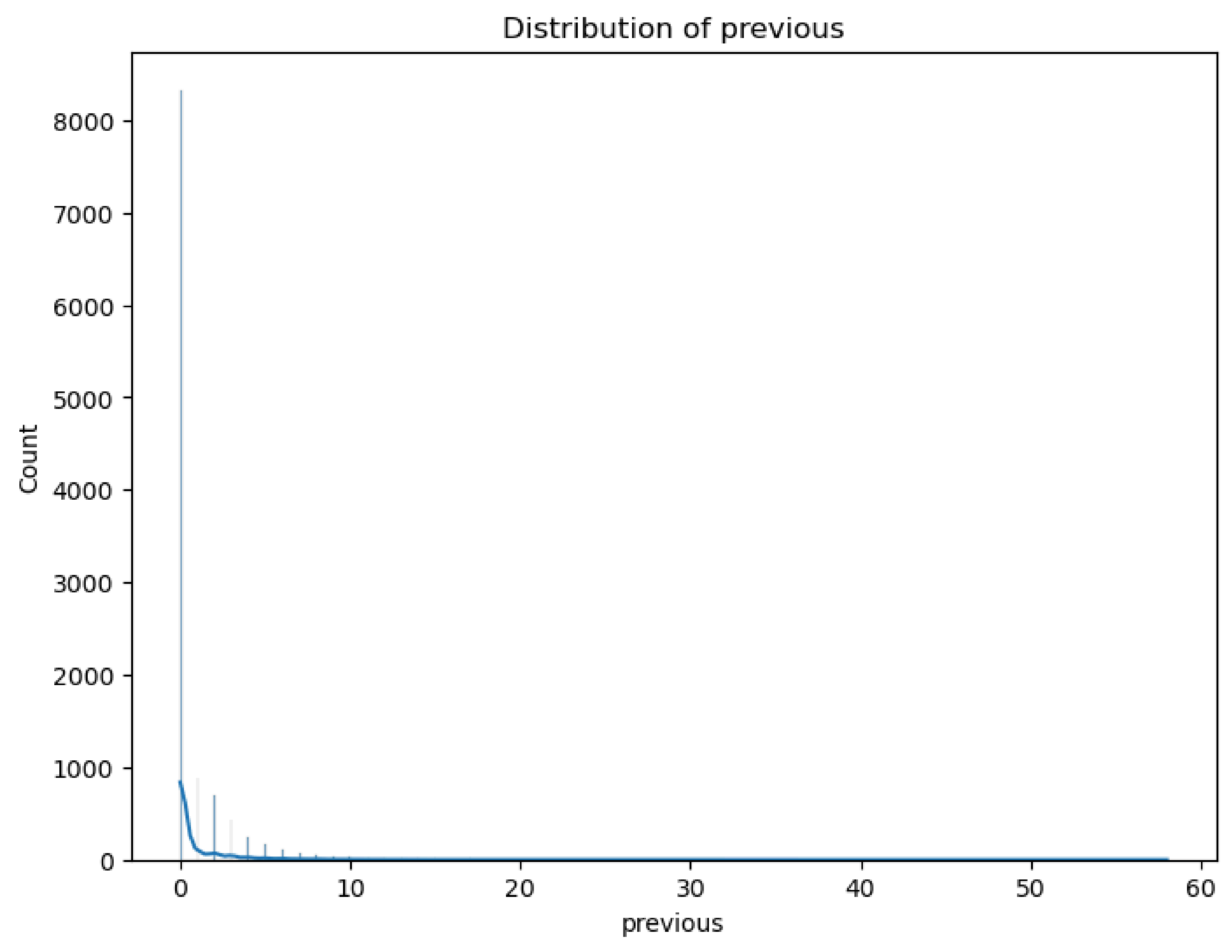
Most clients between 25–45 years old.

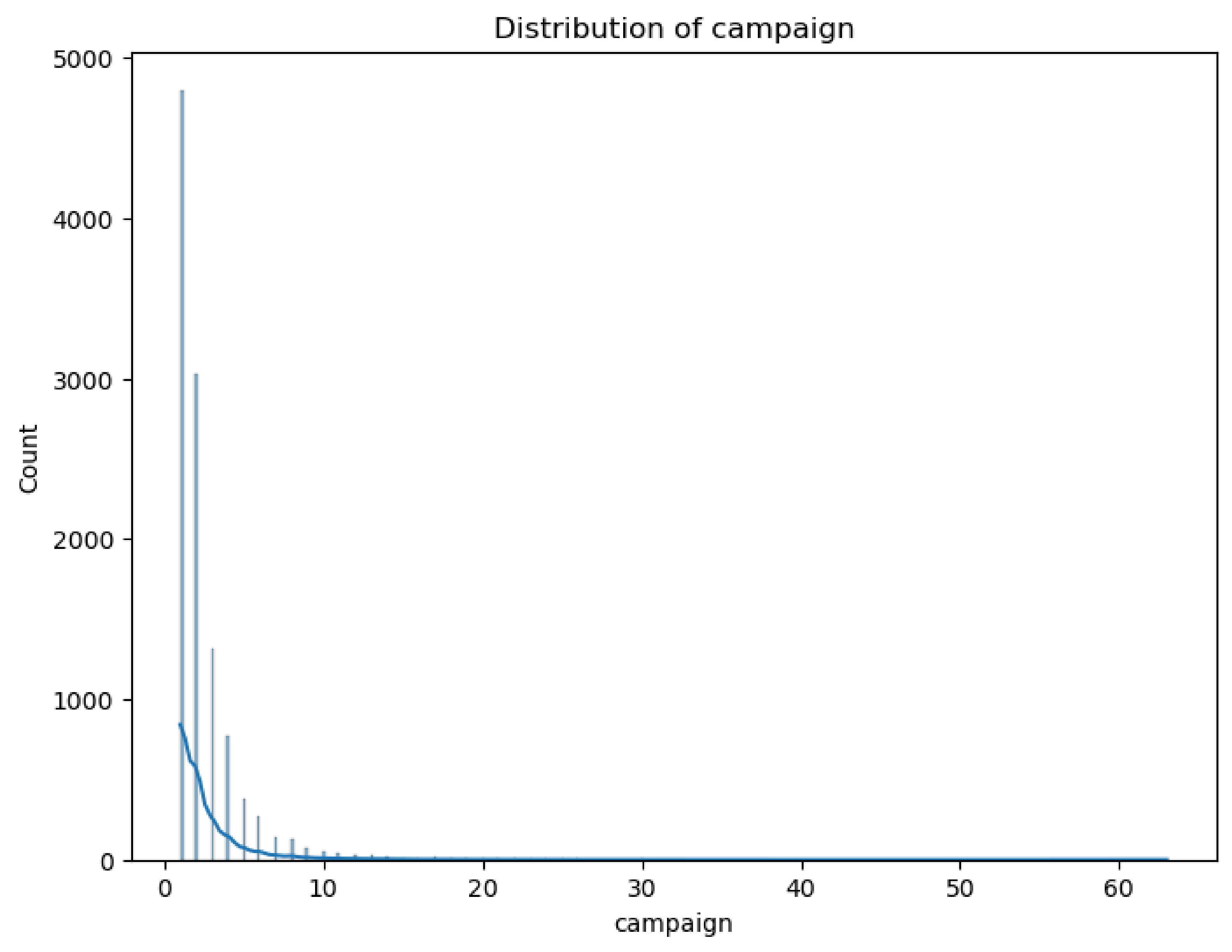


Strong positive relationship with deposit success.

longer calls = higher conversion.

Data Insights





Clients contacted repeatedly or previously successful (poutcome_success) are more likely to subscribe.

Call Duration = Conversion

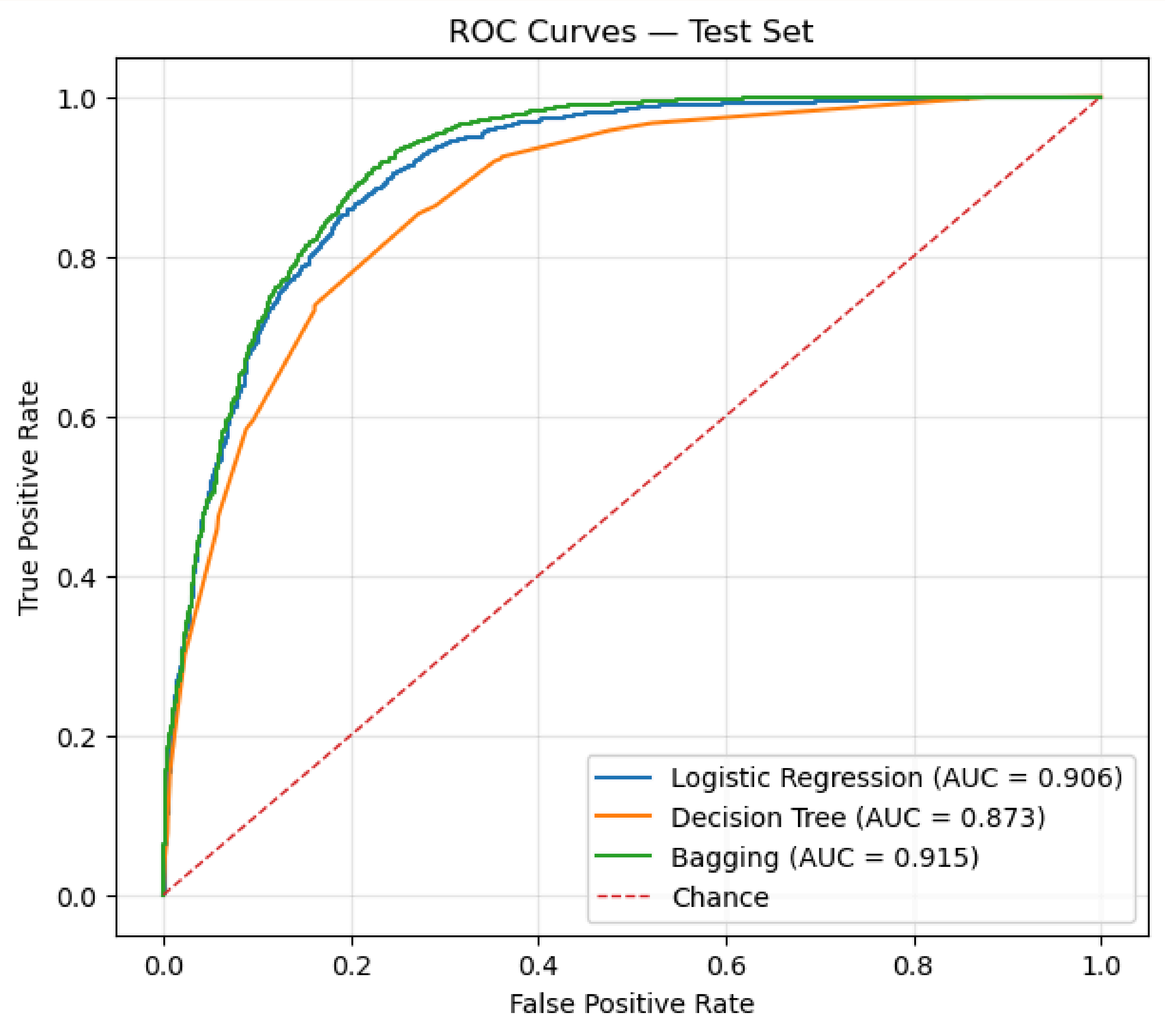
- Very short calls almost never convert
- Long calls have significantly higher subscription rates

- Duration is not known before calling → but helps train the model
- Other features compensate for this at scoring time

Models Performance

Model	Accuracy	ROC-AUC	Observations
Logistic Regression	0.82	0.90	Strong baseline, interpretable
Decision Tree	0.80	0.87	Non-linear interactions
Bagging Classifier	**0.84**	**0.915**	Best generalization & stability

Models Performance



- The Bagging model achieved the highest AUC (0.915), confirming its superior ability to separate subscribers from non-subscribers across all thresholds.
- Logistic Regression (AUC = 0.906) performed very competitively.
- The Decision Tree (AUC = 0.873) shows slightly weaker performance, reflecting its higher variance and limited depth compared to the ensemble.

Best Model: Bagging Classifier

- **Accuracy**: 0.836
- **Precision**: 0.809
- **Recall**: 0.861
- **F1-Score**: 0.834
- **ROC-AUC**: 0.915

- **Low overfitting:** train/test difference < 0.03
- Stable across repeated runs

Top Predictors of Deposit Loan Subscription

Duration

Poutcome_success

Past positive outcomes

Contact type (unknown vs cellular)

Housing loan (yes)

Indicates financial stability and cross-sell potential

Previous contacts & client balance

Low-importance but useful variables

Job, marital status, month, education

Recommendations for Managers

- Prioritize previously engaged clients
- Focus on longer and meaningful call

- Avoid contacting clients too frequently
- Use the model for lead scoring

- Target clients with strong financial indicators
- Reallocate agent time based on predicted probability

Deployment Strategy for the Bank

- Integrate model into CRM
- Generate probability score for every client
- Prioritize calls by predicted likelihood
- Combine with segmentation rules
- Monitor performance monthly
- Update model with new campaign data

Conclusion

Machine learning can boost campaign efficiency and optimize targeting.

Behavioral factors like call duration and previous outcomes drive predictions more than demographics.

This model enables predictive lead scoring, helping the bank increase conversions and reduce costs.