# Predicting the Success of Bank Telemarketing Campaigns

NOVA IMS
Information Management School

Enterprise Data Science Bootcamp (2025)

André Francês: 20242049
Frederico Sampaio: 20242049
Diogo Bento: 20242031
Tiago Silva:

# Table of contents

## 01

## Our Challenge

Predict the success identify the key factors that influence customers' decisions to subscribe to long-term deposits.

## 02

## Our Approach

Data cleaning, feature engineering, exploratory analysis and modeling on a multi-source dataset.

## 03

## Modeling

Built and compared multiple models (Regression, Decision Trees, Random Forests, XGBoost, CatBoost), interpreted using feature importance and SHAP/LIME.

## 04

## Outcome

Delivered insights on customer segments, drivers of campaign success, and strategic recommendations for improving targeting, timing, and telemarketing effectiveness.

01

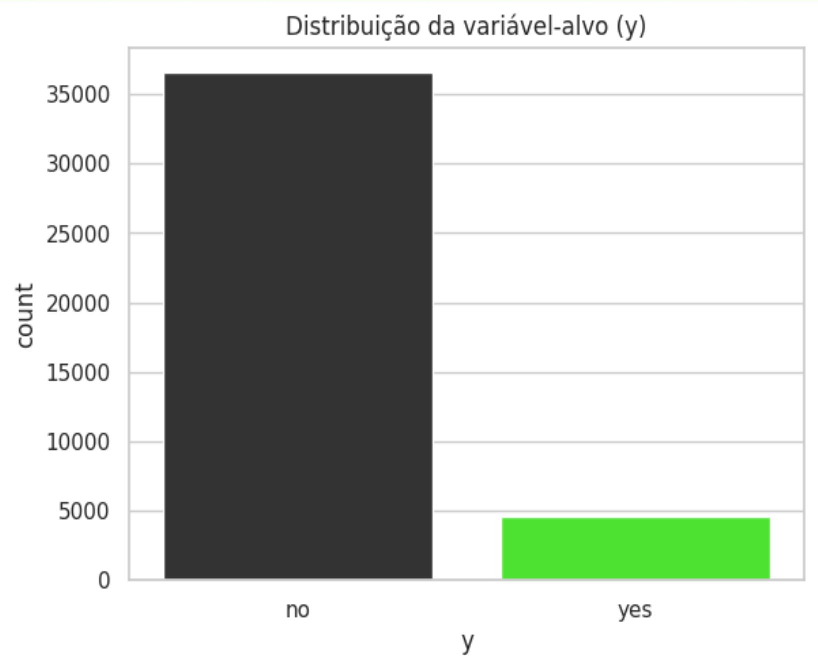
# Data exploration

EDA

# Initial Analysis

The dataset is highly imbalanced: most clients did not subscribe to the deposit.

Only a small portion responded “yes,” which highlights the challenge for predictive modeling.



# Initial Analysis

```
unknown_cols = {col: (df[col] == 'unknown').sum()
                for col in df.columns if df[col].dtype == object}

unknown_cols
```

```
{'job': np.int64(330),
 'marital': np.int64(80),
 'education': np.int64(1731),
 'default': np.int64(8597),
 'housing': np.int64(990),
 'loan': np.int64(990),
 'contact': np.int64(0),
 'month': np.int64(0),
 'day_of_week': np.int64(0),
 'poutcome': np.int64(0),
 'y': np.int64(0)}
```

Categorical values of “unknown”

```
for col in df.columns:
    print(col, "→", df[col].nunique())
```

```
age → 78
job → 12
marital → 4
education → 8
default → 3
housing → 3
loan → 3
contact → 2
month → 10
day_of_week → 5
duration → 1544
campaign → 42
pdays → 27
previous → 8
poutcome → 3
emp.var.rate → 10
cons.price.idx → 26
cons.conf.idx → 26
euribor3m → 316
nr.employed → 11
y → 2
```

Unique Values

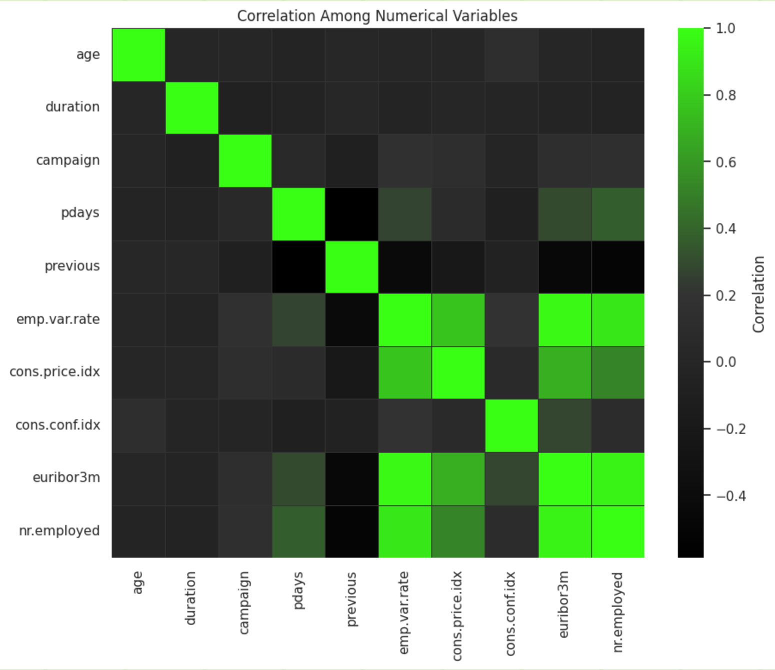
# Initial Analysis

The heatmap shows the correlation between key numerical features in the dataset.

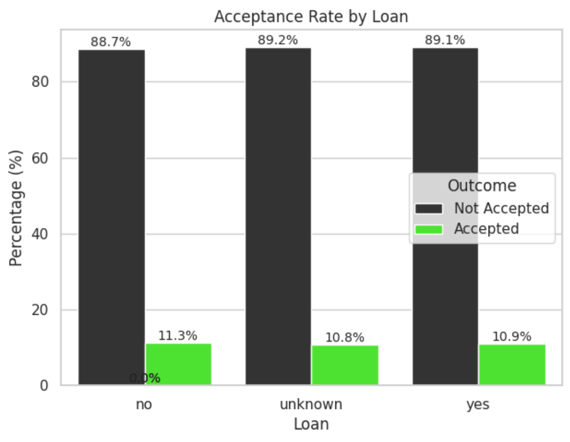
Most variables have very weak correlations with each other, indicating low multicollinearity.

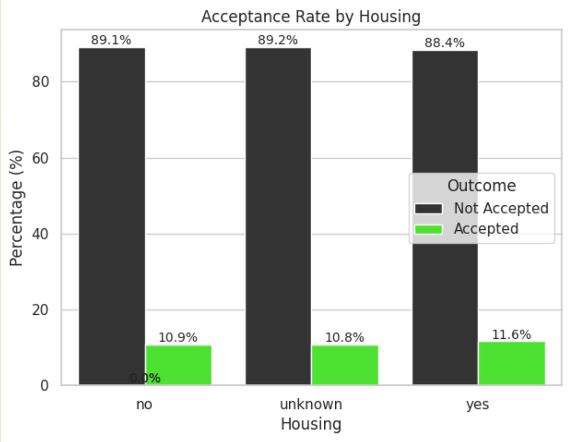
Stronger positive correlations appear among economic indicators (e.g., euribor3m, nr.employed, emp.var.rate), which tend to move together.

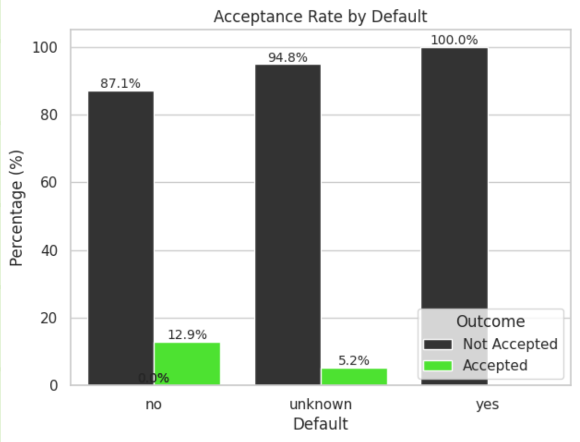
Customer-related variables like age, duration, and campaign show minimal correlation with economic variables, suggesting they contribute independently to the model.



# Financial Background



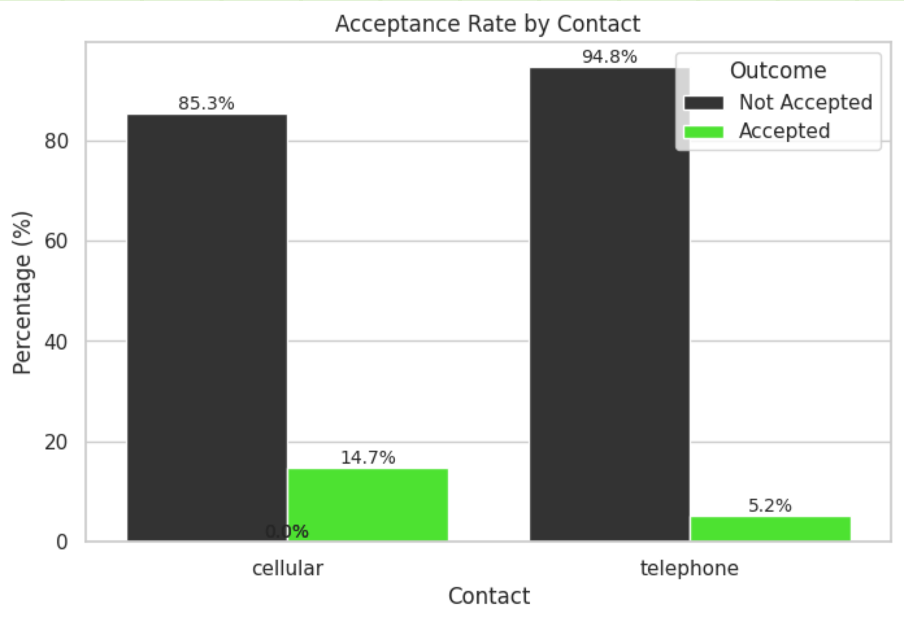




These graphs show that acceptance rates stay low across loan, housing and default categories, with only small differences between groups.

Overall, these variables have minimal impact on whether a customer is accepted or not.

# Contact Information

We can also see that “cellular” or “telephone” aren’t determining factors for acceptance, although there is around a 10% difference.

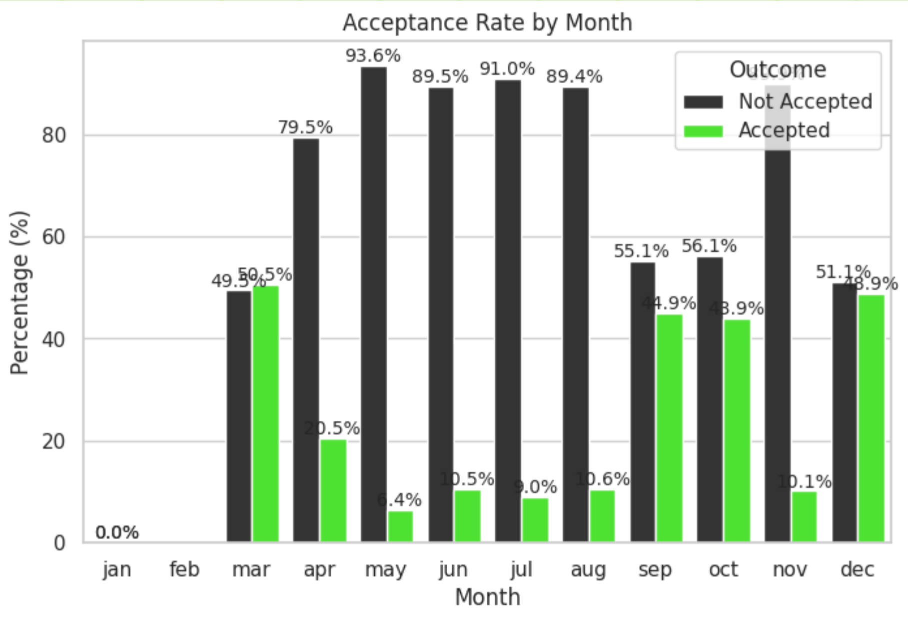


# Contact Information

This chart shows clear seasonal variation in acceptance rates.

Early and late months of the year—especially March and September to December—have noticeably higher acceptance levels.

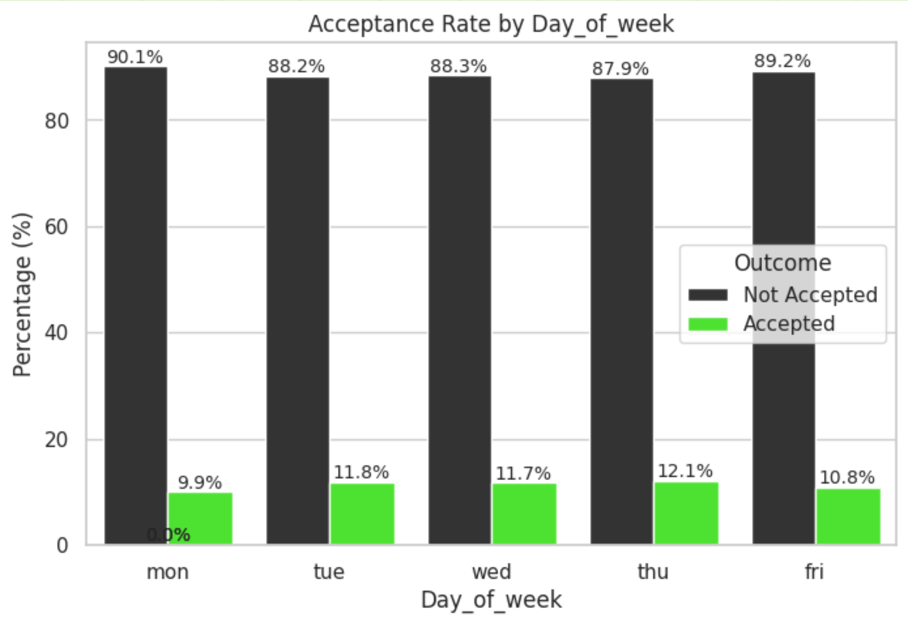
In contrast, the middle months (April to August) show sharply lower acceptance rates, indicating that campaign outcomes are strongly influenced by the time of year.



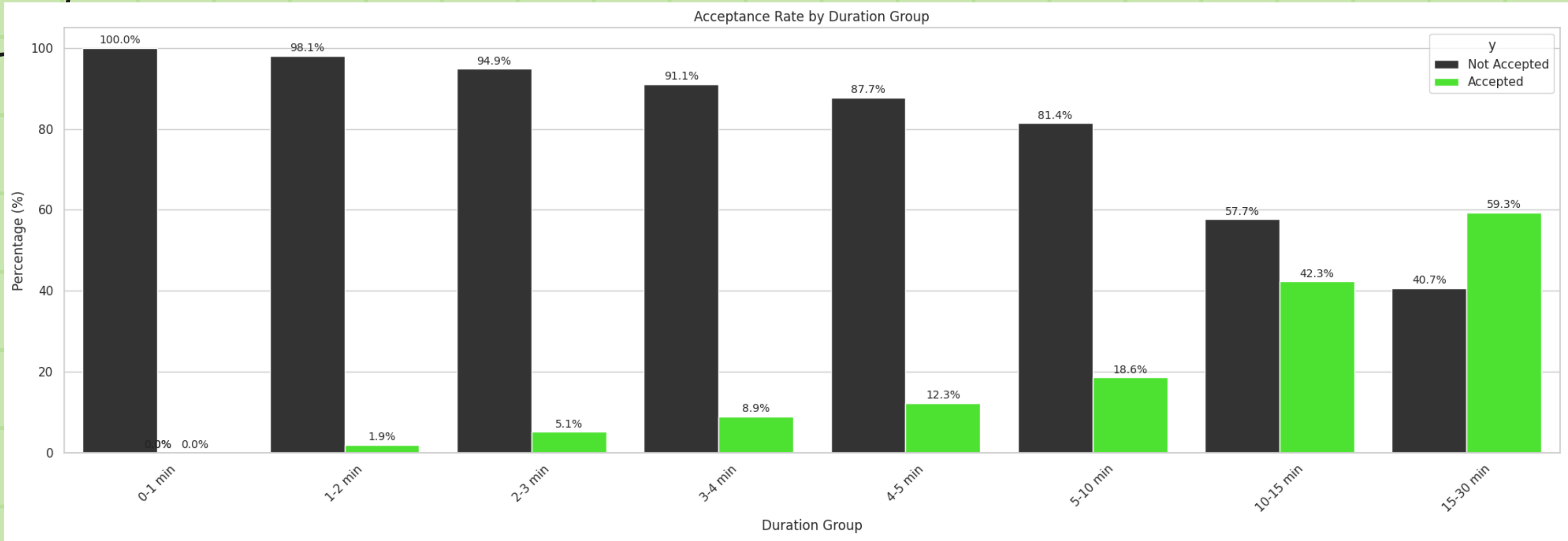
# Contact Information

However, the same can't be said about the day of the week.

This chart shows that throughout the week there is no real change in the acceptance rates.



# Contact Information




As for the duration, there is a clear increase in acceptance rate the longer the duration is, to the point of there being higher change of acceptance in the 15 to 30 minute groups.

# Economic Features

Most indicators—such as employment variation rate, consumer price index, and consumer confidence—show relatively small ranges and low variability, meaning the economic environment was fairly stable during the observed period.

The euribor3m and number of employees are consistent with narrow interquartile ranges.

Overall, the economic variables do not display extreme fluctuations, suggesting that the dataset reflects a steady economic period rather than one with major volatility.

Descriptive statistics (economic variables):

| | emp.var.rate | cons.price.idx | cons.conf.idx | euribor3m | nr.employed |
|---|---|---|---|---|---|
| count | 41188.000000 | 41188.000000 | 41188.000000 | 41188.000000 | 41188.000000 |
| mean | 0.081886 | 93.575664 | -40.502600 | 3.621291 | 5167.035911 |
| std | 1.570960 | 0.578840 | 4.628198 | 1.734447 | 72.251528 |
| min | -3.400000 | 92.201000 | -50.800000 | 0.634000 | 4963.600000 |
| 25% | -1.800000 | 93.075000 | -42.700000 | 1.344000 | 5099.100000 |
| 50% | 1.100000 | 93.749000 | -41.800000 | 4.857000 | 5191.000000 |
| 75% | 1.400000 | 93.994000 | -36.400000 | 4.961000 | 5228.100000 |
| max | 1.400000 | 94.767000 | -26.900000 | 5.045000 | 5228.100000 |

# Economic Features

This table compares the overall average values of the economic indicators with the averages specifically for customers who accepted the offer (y = yes).

The results show that accepted customers tend to come from slightly weaker economic conditions—lower employment variation rate, lower consumer confidence, lower Euribor rate, and fewer employees.

These small shifts suggest that customers are more likely to accept the offer during periods of mild economic downturn or uncertainty.

| Variable | Overall Mean | Mean (y = yes) |
| --- | --- | --- |
| Employment Variation Rate | 0.080000 | -1.230000 |
| Consumer Price Index | 93.580000 | 93.350000 |
| Consumer Confidence Index | -40.500000 | -39.790000 |
| Euribor 3 Month Rate | 3.620000 | 2.120000 |
| Number of Employees | 5167.040000 | 5095.120000 |

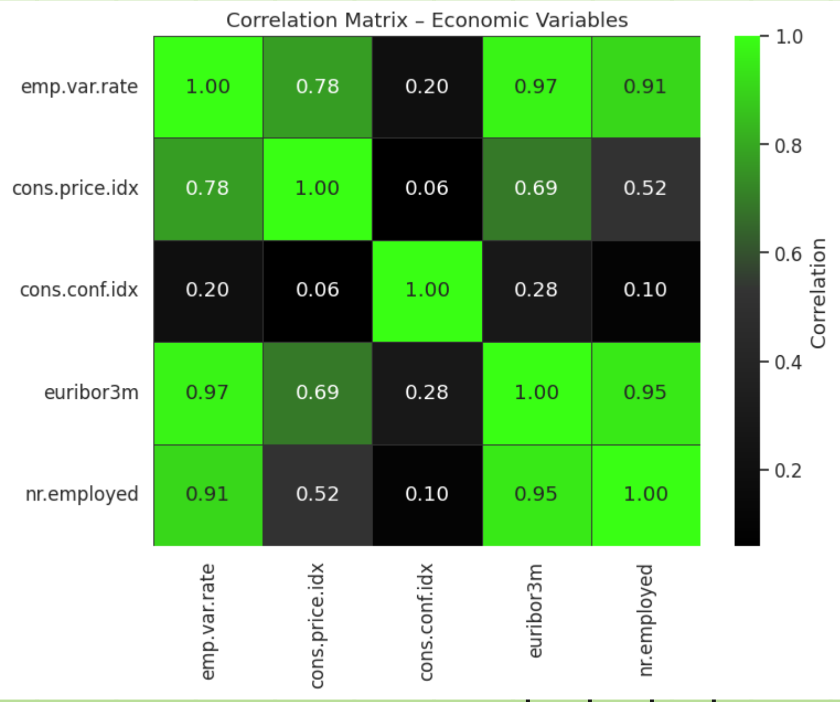
# Economic Features

This correlation matrix shows strong relationships among economic variables.

Employment variation rate, euribor rate, and number of employees are highly correlated with each other, indicating they tend to move together in the economy.

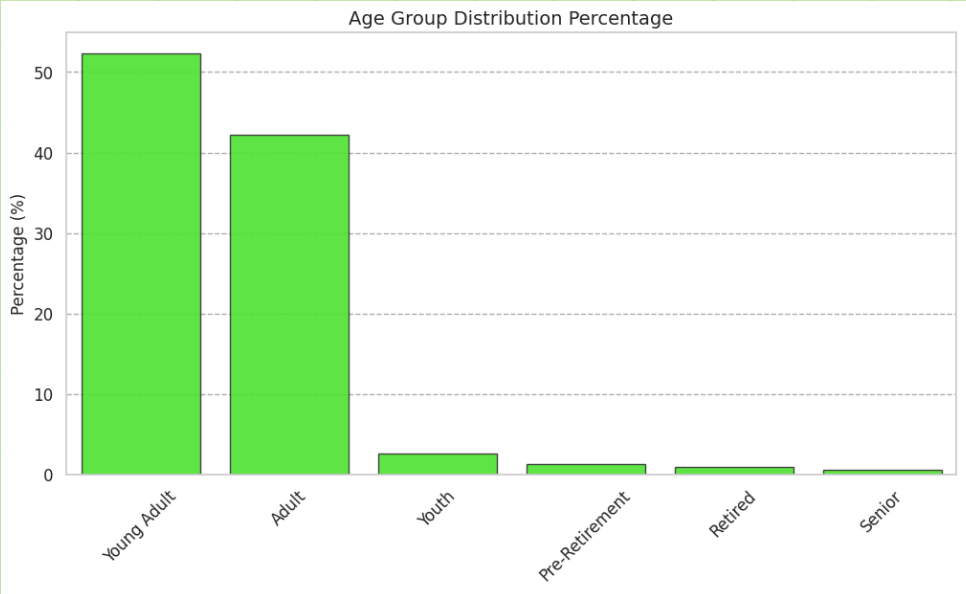
Consumer price index also shows moderate correlations with these variables.

In contrast, consumer confidence has very weak correlations with the others, suggesting it behaves independently and captures a different aspect of economic conditions.



# Demographic Variables

| | Age | Age Group |
|---|---|---|
| 0 | <= 24 | Youth |
| 1 | 25 – 39 | Young Adult |
| 2 | 40 – 59 | Adult |
| 3 | 60 – 64 | Pre-Retirement |
| 4 | 65 – 74 | Retired |
| 5 | >= 75 | Senior |

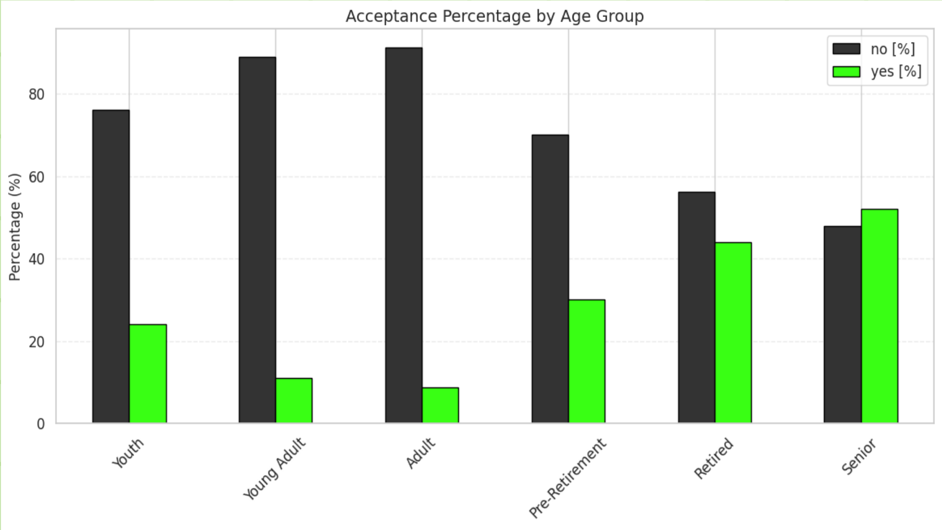



We grouped users into age groups to simplify interpretation. Most users fall between 25 and 59 years old, making young and middle-aged adults the dominant segments.

# Demographic Variables

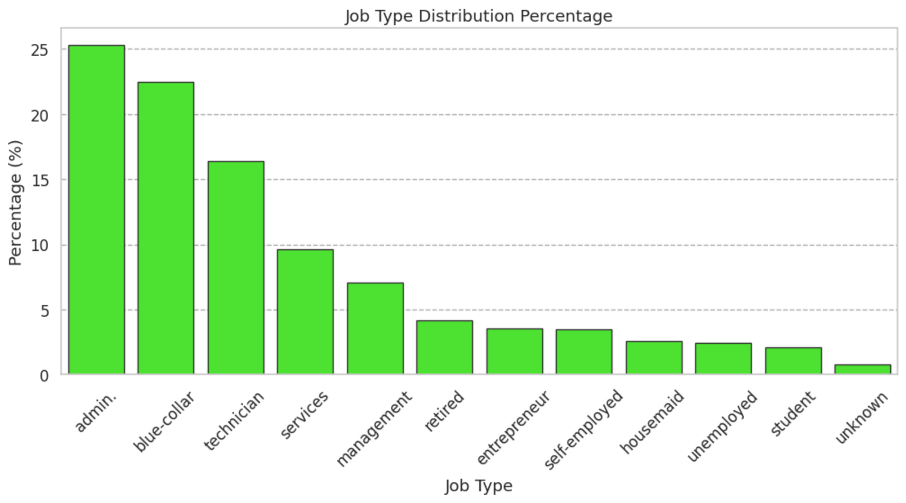
Despite “Young Adults” and “Adults” making up the largest share of the dataset, they have the lowest acceptance rates.

In contrast, older groups, representing less than 10% of users, display much higher acceptance, highlighting a strong age-related difference in acceptance.



# Demographic Variables

| job | |
|---|---|
| admin. | 10422 |
| blue-collar | 9254 |
| technician | 6743 |
| services | 3969 |
| management | 2924 |
| retired | 1720 |
| entrepreneur | 1456 |
| self-employed | 1421 |
| housemaid | 1060 |
| unemployed | 1014 |
| student | 875 |
| unknown | 330 |



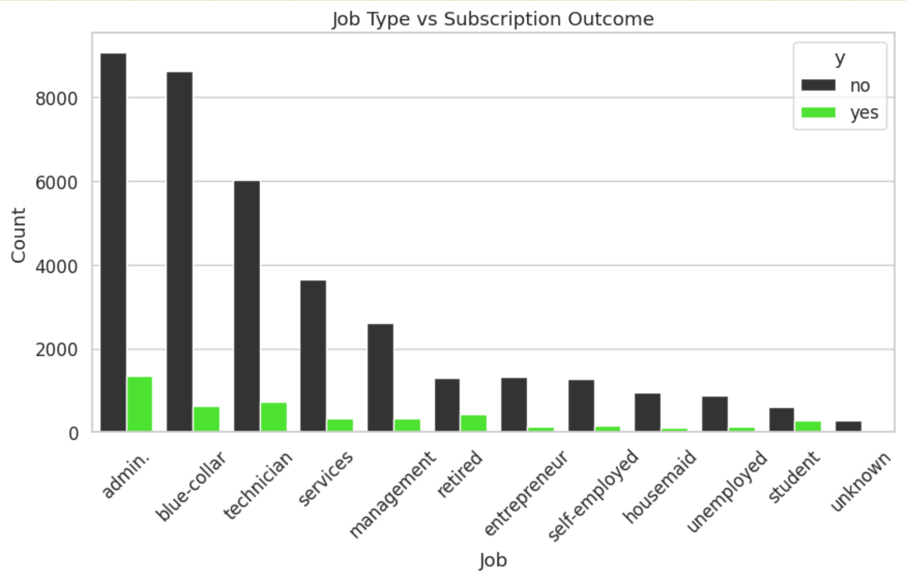

Most users work in administrative (25.3%), blue-collar (22.5%), or technician (16.4%) roles — together, these three make up over 64% of the dataset.

Services (9.6%) and management (7.1%) are the next largest groups, while retired, entrepreneur, self-employed, housemaid, unemployed, and student each account for 4% or less.

# Demographic Variables

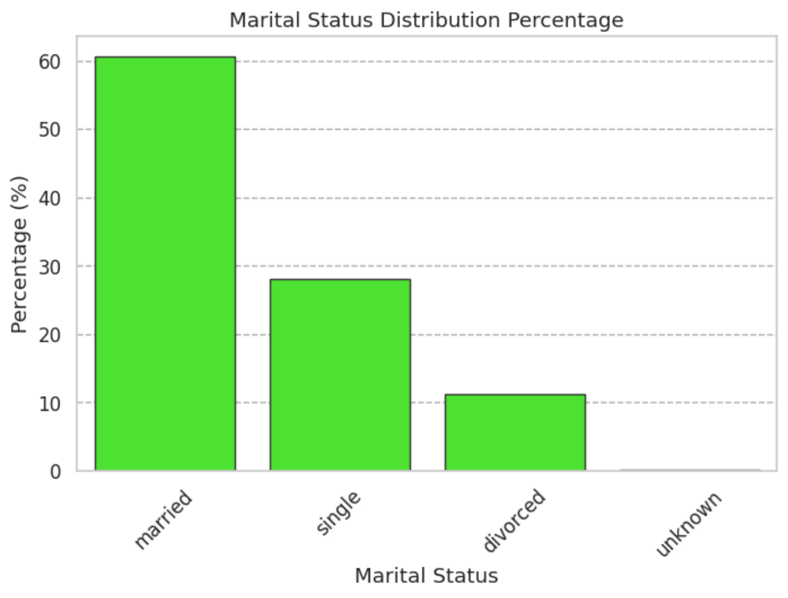
This chart shows that the largest job groups—admin., blue-collar, and technician—also produce the highest number of subscriptions simply because they dominate the dataset.

However, smaller groups like retired and students show relatively higher acceptance counts compared to their size, indicating they are more responsive than the major jobs.



# Demographic Variables

| marital | |
|---|---|
| married | 24928 |
| single | 11568 |
| divorced | 4612 |
| unknown | 80 |



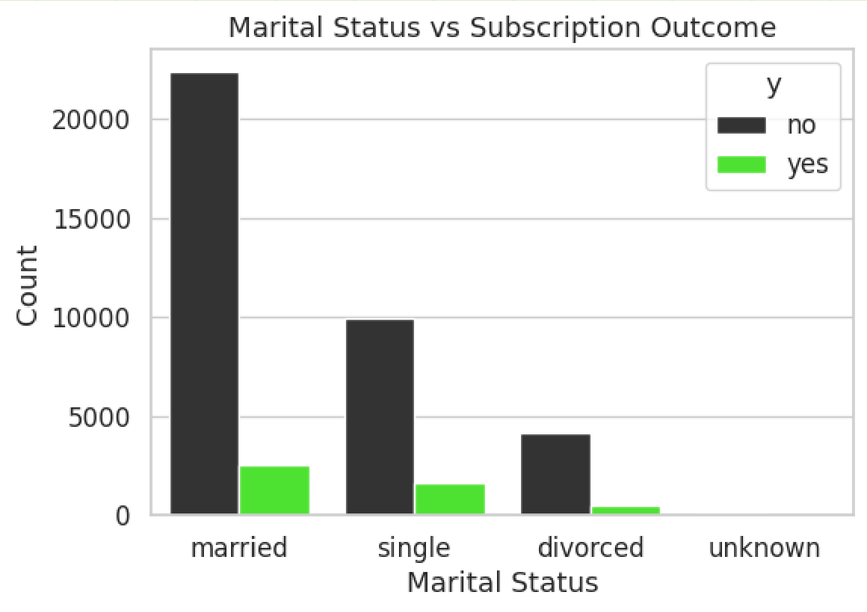

The dataset is dominated by married individuals, who make up about 60% of all users—far more than singles and divorced clients.

# Demographic Variables

Married individuals make up the largest portion of the dataset and therefore account for most of both the ‘yes’ and ‘no’ outcomes.

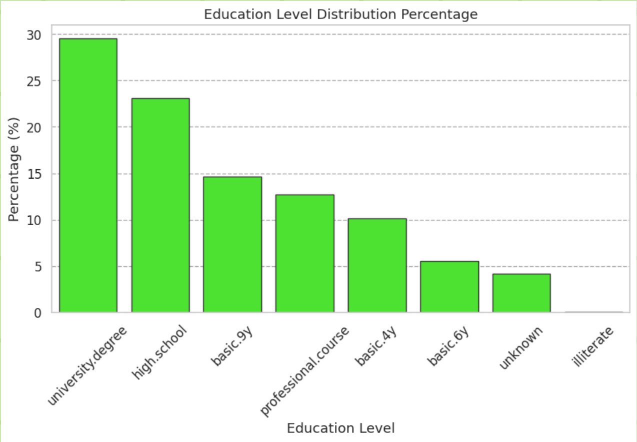
Although the single and divorced groups are smaller in size, they show a higher ratio of ‘yes’ to ‘no’ outcomes compared to married individuals.

The ‘unknown’ category is minimal and has almost no impact on the results.



# Demographic Variables

| education | |
|---|---|
| university.degree | 12168 |
| high.school | 9515 |
| basic.9y | 6045 |
| professional.course | 5243 |
| basic.4y | 4176 |
| basic.6y | 2292 |
| unknown | 1731 |
| illiterate | 18 |




The dataset is dominated by users with higher education levels: *university degree* and *high school* together make up the largest portion of the sample.

Mid-level categories such as *basic.9y*, *professional.course*, and *basic.4y* follow with moderate representation.

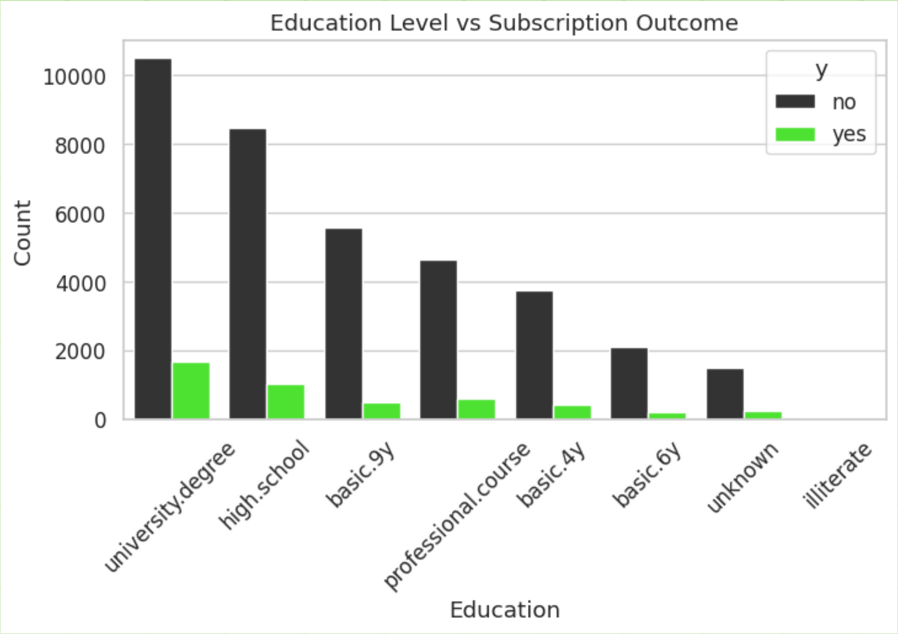
Lower education levels, including *basic.6y*, *unknown*, and especially *illiterate*, account for only a small fraction of the dataset.

# Demographic Variables

Users with university degrees and high-school education make up the largest groups and also generate the highest number of subscriptions overall.

However, while higher-educated groups show more “yes” responses in absolute terms, their acceptance rates are not dramatically higher than the other categories.

Lower-educated groups have less subscriptions mainly because they represent a much smaller portion of the dataset.



02

# Data Pre-Processing

Missing Values, Outliers, Encoding

# Handling Missing Values

1. Replaced all 'unknown' entries with actual missing values (NaN) to make them easier to detect.
2. Categorical features had missing values filled using the most frequent category (mode).
3. Numerical features had missing values filled using the median, which avoids distortion from extreme values.

# Handling Outliers

- Applied Winsorization to numerical columns:
    1. Values below the 1st percentile were set to the 1st percentile.
    2. Values above the 99th percentile were set to the 99th percentile.
- This keeps the dataset stable while preserving overall distribution patterns.

# Handling Outliers

Different encoding strategies were used based on the type of variable:

- One-hot encoding for multi-category features (job, marital status, contact type, poutcome, duration group).
- Binary encoding for yes/no variables (loan, housing, target variable y).
- Derived features:
  - **was_contacted_before**: indicates whether the client had any previous contact (based on pdays).
- Ordinal encoding for naturally ordered variables: Age groups (Youth → Senior) Education levels (Illiterate → University)

# Feature Engineering

1. Created duration groups and encoded them both as ordinal and one-hot to support different model types.

2. Created midpoint-based numerical duration (duration_minutes) to represent ranges with approximate call lengths.

03

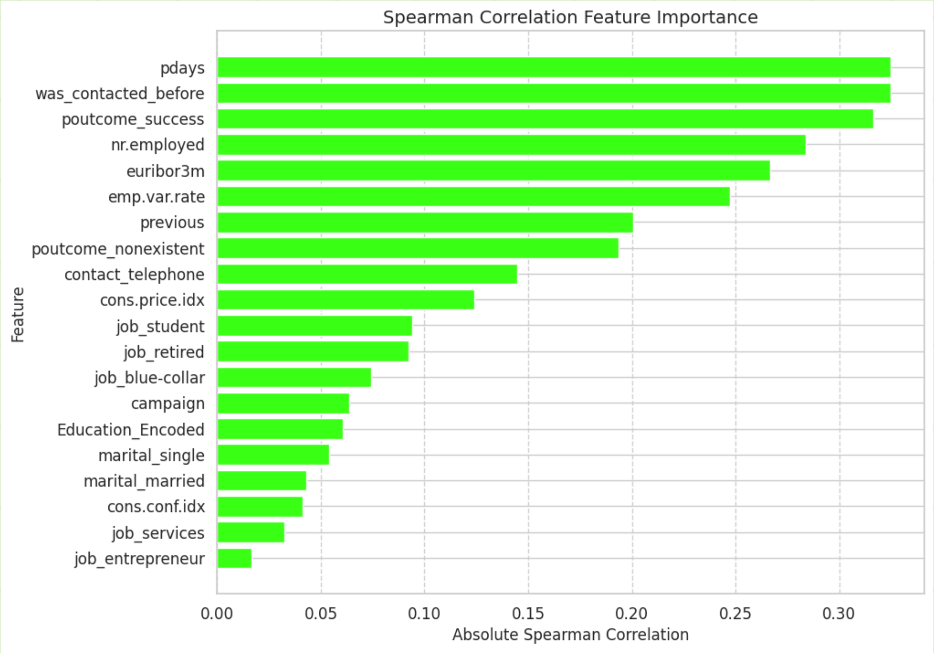
# Feature Selection

Spearman, Chi-square, Lasso...

# Spearman Correlation

The most influential features are previous contact indicators like pdays and was_contacted_before, followed by past campaign outcomes.

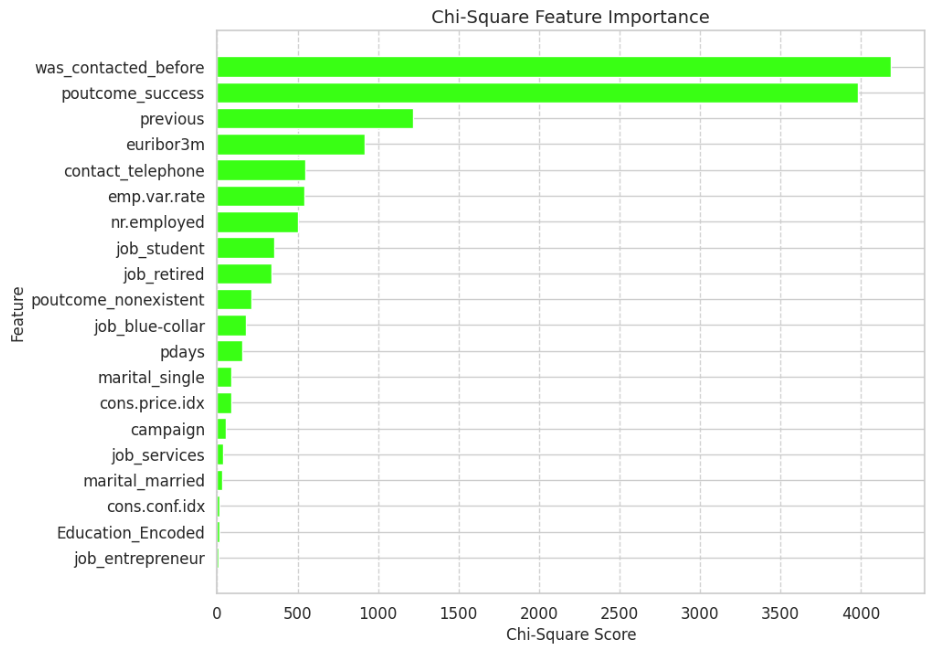
Economic indicators show moderate importance, while demographic factors such as job, education, and marital status have relatively lower impact.



# Chi-Square

The strongest predictors are previous contact behavior, with *was_contacted_before* and *poutcome_success* dominating by a large margin.

Other contact-related and economic variables show moderate importance, while demographic features contribute far less.



# RFE

The RFE-selected features showed that both past customer interactions and broader economic conditions are highly influential in predicting subscription.

Contact-related variables such as *pdays*, whether the client was previously contacted, and past campaign outcomes (*poutcome_success* and *poutcome_nonexistent*) stand out as major drivers.

Alongside these, macroeconomic indicators—*emp.var.rate*, *euribor3m*, *cons.price.idx*, *cons.conf.idx*, and *nr.employed*—also play a significant role, suggesting that customers' decisions are shaped not only by direct marketing history but also by overall economic sentiment.

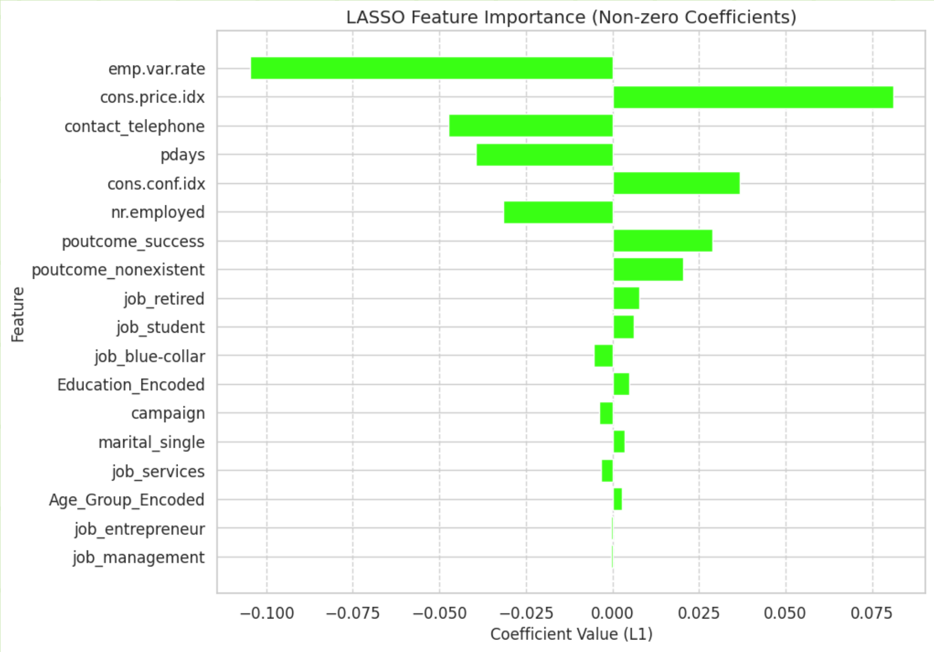
| Selected Features (RFE) |
|---:|
| pdays |
| emp.var.rate |
| cons.price.idx |
| cons.conf.idx |
| euribor3m |
| nr.employed |
| contact_telephone |
| poutcome_nonexistent |
| poutcome_success |
| was_contacted_before |

# LASSO

LASSO highlights a small set of influential features, mainly economic indicators and contact history.

*cons.price.idx* and *emp.var.rate* show the strongest effects, while *contact_telephone*, *pdays*, and *nr.employed* also play meaningful roles.

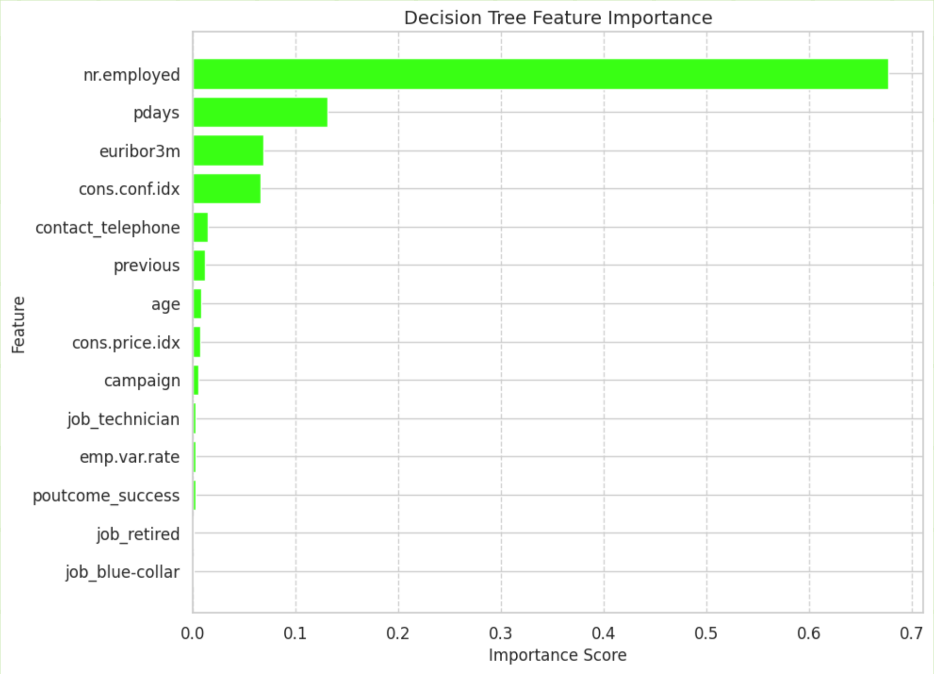
Past campaign outcomes contribute as well, but with smaller coefficients.



# Decision Tree

The decision tree relies mostly on *nr.employed*, which dominates the importance scores.

*pdays*, *euribor3m*, and *cons.conf.idx* contribute moderately, while all other features have only minimal influence.



# Features Dropped

Feature selection methods consistently agreed on removing low impact variables.

The features we dropped were **age, housing, loan, job_housemaid, job_management, job_self-employed, job_technician, job_unemployed, and Age_Group_Encoded**.

The remaining features—mainly contact history, campaign outcomes, and key economic indicators—were consistently strong and retained for modeling.

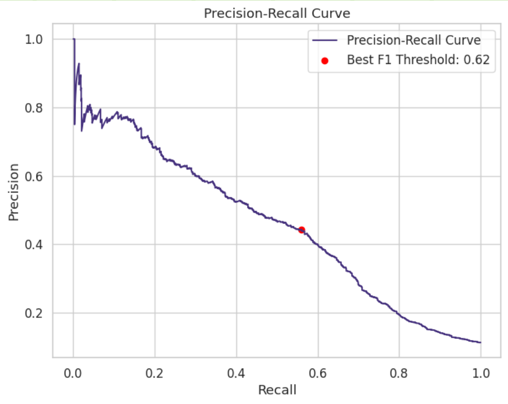
| Feature | Spearman | Chi-Square | RFE | LASSO | Decision Tree | Decision |
|---|---|---|---|---|---|---|
| age | | | | | YES | REMOVE |
| housing | | | | | | REMOVE |
| loan | | | | | | REMOVE |
| campaign | X | X | | YES | YES | KEEP |
| pdays | X | X | YES | YES | YES | KEEP |
| previous | X | X | | | YES | KEEP |
| emp.var.rate | X | X | YES | YES | YES | KEEP |
| cons.price.idx | X | X | YES | YES | YES | KEEP |
| cons.conf.idx | X | X | YES | YES | YES | KEEP |
| euribor3m | X | X | YES | | YES | KEEP |
| nr.employed | X | X | YES | YES | YES | KEEP |
| job_blue-collar | X | X | | YES | YES | KEEP |
| job_entrepreneur | X | X | | YES | | KEEP |
| job_housemaid | | | | | | REMOVE |
| job_management | | | | YES | | REMOVE |
| job_retired | X | X | | YES | YES | KEEP |
| job_self-employed | | | | | | REMOVE |
| job_services | X | X | | YES | | KEEP |
| job_student | X | X | | YES | | KEEP |
| job_technician | | | | | YES | REMOVE |
| job_unemployed | | | | | | REMOVE |
| marital_married | X | X | | | | TRY WITH/WITHOUT |
| marital_single | X | X | | YES | | KEEP |
| contact_telephone | X | X | YES | YES | YES | KEEP |
| poutcome_nonexistent | X | X | YES | YES | | KEEP |
| poutcome_success | X | X | YES | YES | YES | KEEP |
| was_contacted_before | X | X | YES | | | KEEP |
| Age_Group_Encoded | | | | YES | | REMOVE |
| Education_Encoded | X | X | | YES | | KEEP |

04

# Modeling

Regression, Random Forest, XGBoost, ...

# Logistic Regression



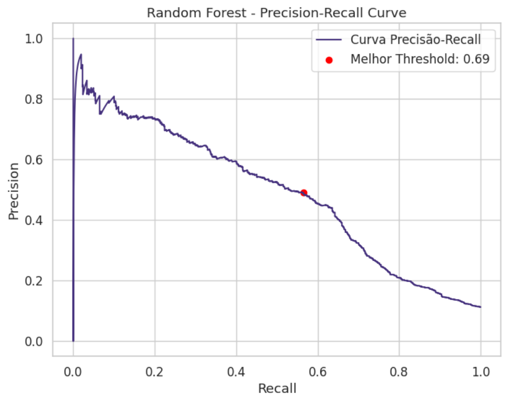

```
Classification Report (threshold adjusted):
              precision    recall  f1-score   support

           0       0.94      0.91      0.93      7310
           1       0.44      0.56      0.50       928

    accuracy                           0.87      8238
   macro avg       0.69      0.74      0.71      8238
weighted avg       0.89      0.87      0.88      8238

Confusion Matrix:
 [[6657  653]
 [ 408  520]]
```

The logistic regression model reached **87% accuracy** after adjusting the decision threshold. For the positive class (subscribers), the model achieved **0.56 recall**, meaning it correctly identified 56% of potential subscribers, with a **precision of 0.44**.

The confusion matrix reflects this balance, with **520 true positives** and **408 false negatives**, showing that the model captures more positive cases while still controlling false alarms.

The precision-recall curve highlights the chosen threshold (~0.62) as a good compromise between detecting more subscribers and keeping precision at a reasonable level.

# Random Forest



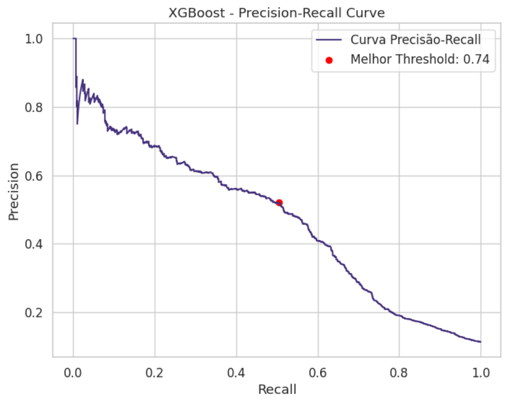

```
ROC-AUC: 0.8126128443558658

Classification Report (threshold adjusted):
              precision    recall  f1-score   support

           0       0.94      0.93      0.93      7310
           1       0.49      0.56      0.53       928

    accuracy                           0.88      8238
   macro avg       0.72      0.75      0.73      8238
weighted avg       0.89      0.88      0.89      8238

Confusion Matrix:
 [[6766  544]
 [ 404  524]]
```

The Random Forest model achieved **88% accuracy** after adjusting the decision threshold. For the positive class (subscribers), it reached a recall of **0.56**, meaning it correctly identified 56% of potential subscribers — with a **precision of 0.49**.

The confusion matrix shows 524 true positives and 404 false negatives, indicating a stronger ability to detect subscribers while keeping errors controlled.

The precision–recall curve highlights the chosen threshold (~0.69) as a good trade-off between identifying more subscribers and maintaining reasonable precision.

# XGBoost




```
ROC-AUC: 0.799

Classification Report (threshold ajustado):
              precision    recall  f1-score   support

           0       0.94      0.94      0.94      7310
           1       0.52      0.50      0.51       928

    accuracy                           0.89      8238
   macro avg       0.73      0.72      0.73      8238
weighted avg       0.89      0.89      0.89      8238

Confusion Matrix:
 [[6880  430]
 [ 460  468]]
```

The XGBoost model achieved **89%** accuracy after optimizing the decision threshold.

For the positive class (subscribers), the model reached a **recall of 0.50**, correctly identifying half of the potential subscribers, with a **precision of 0.52**.

The confusion matrix shows 468 true positives and 460 false negatives, reflecting a balanced but still challenging trade-off when predicting minority-class cases.

The precision–recall curve highlights the selected threshold (~0.74) as a point that maximizes the minority-class F1-score, maintaining acceptable precision.

# CatBoost

```
Confusion Matrix:
[[6769  541]
 [ 400  528]]

Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.93      0.94      7310
           1       0.49      0.57      0.53       928

    accuracy                           0.89      8238
   macro avg       0.72      0.75      0.73      8238
weighted avg       0.89      0.89      0.89      8238
```

The CatBoost model achieved **89% accuracy** after adjusting the decision threshold. For the positive class (subscribers), it reached a **recall of 0.57**, correctly identifying 57% of potential subscribers, with a **precision of 0.49**.

The confusion matrix shows **528 true positives** and **400 false negatives**, indicating that the model captures more subscriber cases while keeping errors at a reasonable level.

The performance of the majority class (0) remains strong, with **0.94 precision**, **0.93 recall**, and **0.94 F1-score**, showing consistent reliability in identifying non-subscribers.

# Let's Compare

| Metric / Class | Logistic Regression | Random Forest | XGBoost | CatBoost |
|---|---|---|---|---|
| ROC-AUC | 0.792 | **0.813** | 0.799 | **0.799** |
| Accuracy | 0.87 | 0.88 | **0.89** | **0.89** |
| Class 0 Precision | 0.94 | 0.94 | **0.94** | **0.94** |
| Class 0 Recall | 0.91 | **0.93** | **0.94** | 0.93 |
| Class 0 F1 | 0.93 | 0.93 | **0.94** | **0.94** |
| Class 1 Precision | 0.44 | **0.49** | **0.52** | 0.49 |
| Class 1 Recall | 0.56 | 0.56 | 0.50 | **0.57** |
| Class 1 F1 | 0.50 | **0.53** | 0.51 | **0.53** |

The four models show similar performance overall, but each has different strengths.

**Random Forest** achieves the highest ROC-AUC, indicating the best discriminative power.

**XGBoost** and **CatBoost** reach the highest overall accuracy.

For the minority class, **XGBoost** provides the highest precision, while **CatBoost** achieves the best recall, making it more effective at identifying actual subscribers.

**Random Forest** offers the best balance between precision and recall, reflected in the highest F1-score for class 1.

# 05

# Conclusion Analysis

Model metrics and analysis of results

# Model Metrics

```
=== Logistic Regression (Train) ===
ROC-AUC: 0.783
Accuracy: 0.781
Class 0 -> Precision: 0.948, Recall: 0.797, F1: 0.866
Class 1 -> Precision: 0.291, Recall: 0.656, F1: 0.403


=== Random Forest (Train) ===
ROC-AUC: 0.832
Accuracy: 0.849
Class 0 -> Precision: 0.949, Recall: 0.877, F1: 0.912
Class 1 -> Precision: 0.394, Recall: 0.628, F1: 0.484


=== XGBoost (Train) ===
ROC-AUC: 0.879
Accuracy: 0.859
Class 0 -> Precision: 0.959, Recall: 0.878, F1: 0.917
Class 1 -> Precision: 0.424, Recall: 0.706, F1: 0.530


=== CatBoost (Train) ===
ROC-AUC: 0.834
Accuracy: 0.913
Class 0 -> Precision: 0.920, Recall: 0.988, F1: 0.953
Class 1 -> Precision: 0.778, Recall: 0.324, F1: 0.458
```

Train

```
=== Logistic Regression - Test (Train) ===
ROC-AUC: 0.792
Accuracy: 0.788
Class 0 -> Precision: 0.953, Recall: 0.801, F1: 0.870
Class 1 -> Precision: 0.305, Recall: 0.689, F1: 0.423


=== Random Forest - Test (Train) ===
ROC-AUC: 0.813
Accuracy: 0.854
Class 0 -> Precision: 0.952, Recall: 0.881, F1: 0.915
Class 1 -> Precision: 0.408, Recall: 0.647, F1: 0.500


=== XGBoost - Test (Train) ===
ROC-AUC: 0.799
Accuracy: 0.842
Class 0 -> Precision: 0.949, Recall: 0.868, F1: 0.907
Class 1 -> Precision: 0.379, Recall: 0.635, F1: 0.475


=== CatBoost - Test (Train) ===
ROC-AUC: 0.814
Accuracy: 0.902
Class 0 -> Precision: 0.913, Recall: 0.982, F1: 0.947
Class 1 -> Precision: 0.656, Recall: 0.265, F1: 0.378
```

Test

# Model Metrics

Across all models, we observe consistent patterns between training and test results, indicating stable generalization.

• **Logistic Regression** shows the weakest performance, especially for the minority class 1, with low precision and moderate recall. It identifies many positives and false alarms.
• **Random Forest** improves recall and overall F1 for class 1 compared to Logistic Regression, achieving a stronger balance between precision and recall. It generalizes well, with very similar train and test results.
• **XGBoost** delivers the best recall for class 1 among the tree-based models during training, and remains competitive on the test set. It tends to capture more true positives but at the cost of more false positives.
• **CatBoost** achieves the highest accuracy and strongest class 0 performance, but its recall for class 1 drops significantly on the test set—indicating that it becomes more conservative in predicting the positive class.

**Overall:**

Random Forest and XGBoost offer the most balanced performance for detecting class 1, while CatBoost excels in overall accuracy but struggles to identify positive cases.

# Four Conclusions

01

## Best Performance (AUC and Accuracy)

CatBoost with Accuracy = 0.902 and AUC = 0.814

02

## Identify New Clients (Class 1 Recall)

Logistic Regression with Class 1 Recall = 0.689

03

## Best Precision (Class 1 Precision)

CatBoost with Class 1 Precision = 0.656

04

## Generalization (Train vs Test)

All models are consistent and don’t overfit