Bank Telemarketing Campaign for Selling Long Term Deposits

Dec 5, 2023
Tirthkumar Patel
tp87608n@pace.edu
Practical Data Science
MS in Data Science
Seidenberg School Of Computer Science and Information Systems, Pace University

Agenda

- Executive summary
- Project plan recap
- Data
- Exploratory data analysis
- Modeling methods
- Findings
- Recommendations and next steps

Executive summary

Problem: The success of telemarketing calls for long-term bank deposits directly impacts the bank's growth and financial stability. Increasing deposit subscriptions not only expands the bank's customer base but also enhances its capital reserves, enabling more robust lending and investment capabilities. This, in turn, contributes to the bank's profitability and competitiveness in the financial market. Therefore, the problem revolves around predicting and improving the success of these calls.

Solution: Involves analyzing customer data and communication patterns to identify key factors influencing deposit acceptance. Utilizing a classification model, we aim to analyze customer data and communication patterns deeply. By leveraging these insights, we intend to develop data-driven strategies for future marketing campaigns. The predictive power of the model allows us to strategically target potential deposit subscribers, optimizing the bank's telemarketing efforts and enhancing overall campaign success.

Project plan recap

Deliverable	Details	Due Date	Status
Data & EDA	Create final assignment deck skeleton and fill in the deck up to the end of the EDA section	10/31/23	Complete
Methods, Findings, and Recommendations	Fill in the Methods, Findings, and Recommendations sections of the deck	11/14/23	Complete
Final presentation	Send in final completed deck and present it	12/05/23	Complete

Data

Data

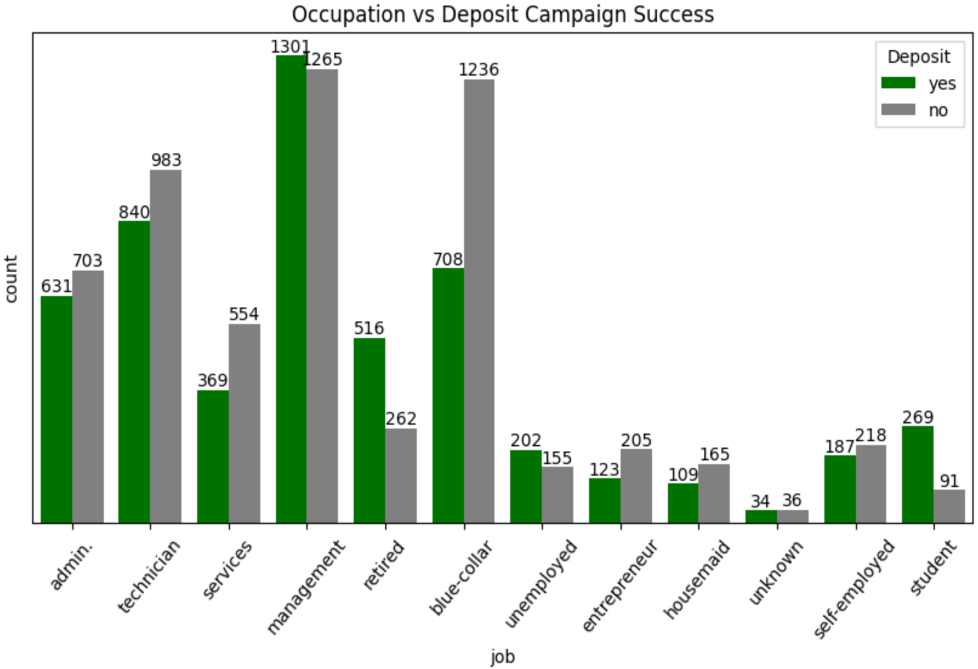
- Data details:
  - Portuguese retail bank telemarketing data for a selling bank’s long-term deposits, collected from 2008 to 2013.
  - Includes the effects of 2007-2008 global financial crisis.
  - Dataset contains a subset of the original data, with 11,162 records and 18 features, making it more manageable for analysis found on Kaggle.
  - Data source: [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014.

- Assumption:
  - Provided dataset is sourced from direct marketing campaigns of a Portuguese banking institution and is comprehensive and representative of the target population. It revolves around predicting client term deposit subscriptions, aligning with the classification goal.

- Feature description:
  - Client data: numerical data (age, bank balance), categorical data (job type, marital status, education, default?, housing loan?, personal loan?)
  - Campaign data: numerical data (number of contacts this campaign, days since last contact, previous contacts for this client, duration of last contact), categorical data (communication type, month of last contact, day of the week last contact, previous campaign outcome)
  - Target: categorical data (subscribed deposit?)

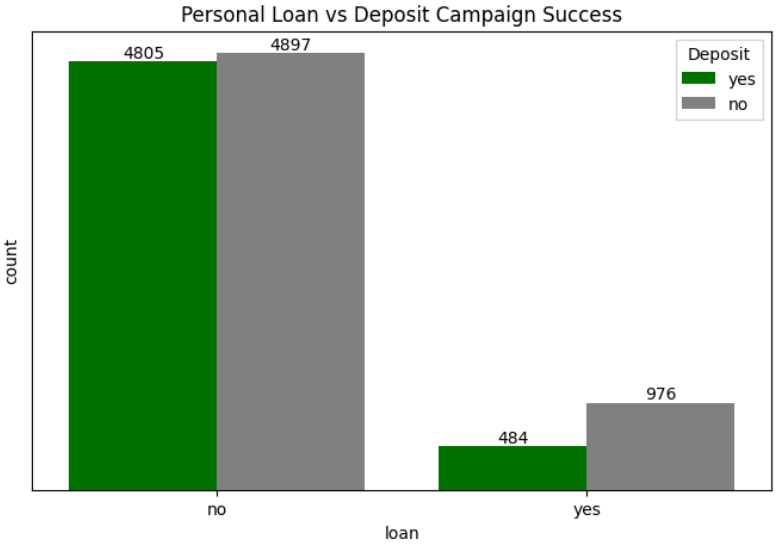
Exploratory Data Analysis

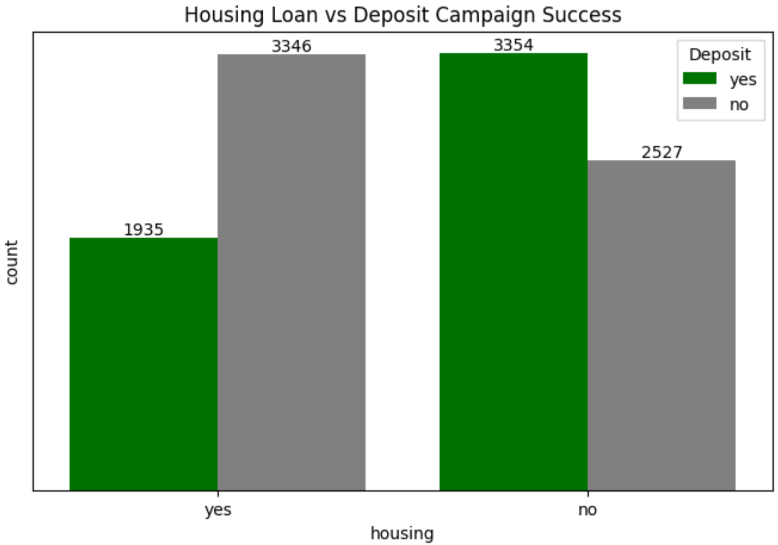
Occupation Impact on Deposit Acceptance



- Students and Retirees Say Yes, Blue Collars Say No. Deposit acceptance rate of student is the highest and blue-collar is lowest.

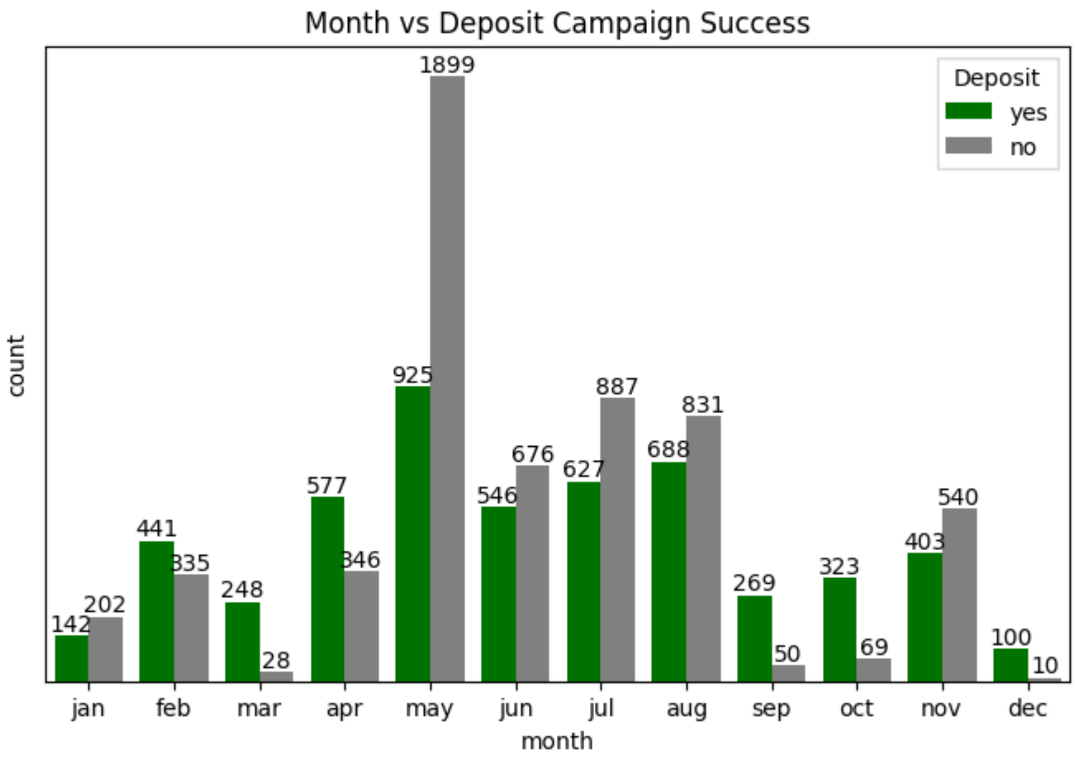
Loan Status Matters





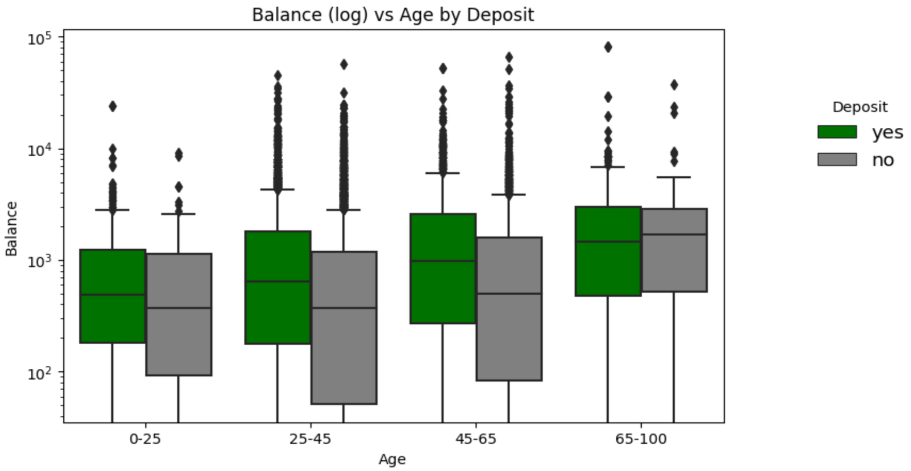
- Personal and housing loan holders are more likely to refuse deposit offer.
- Housing Loan Absence Correlates with Higher Deposit Acceptance

Deposit Acceptance monthly trend



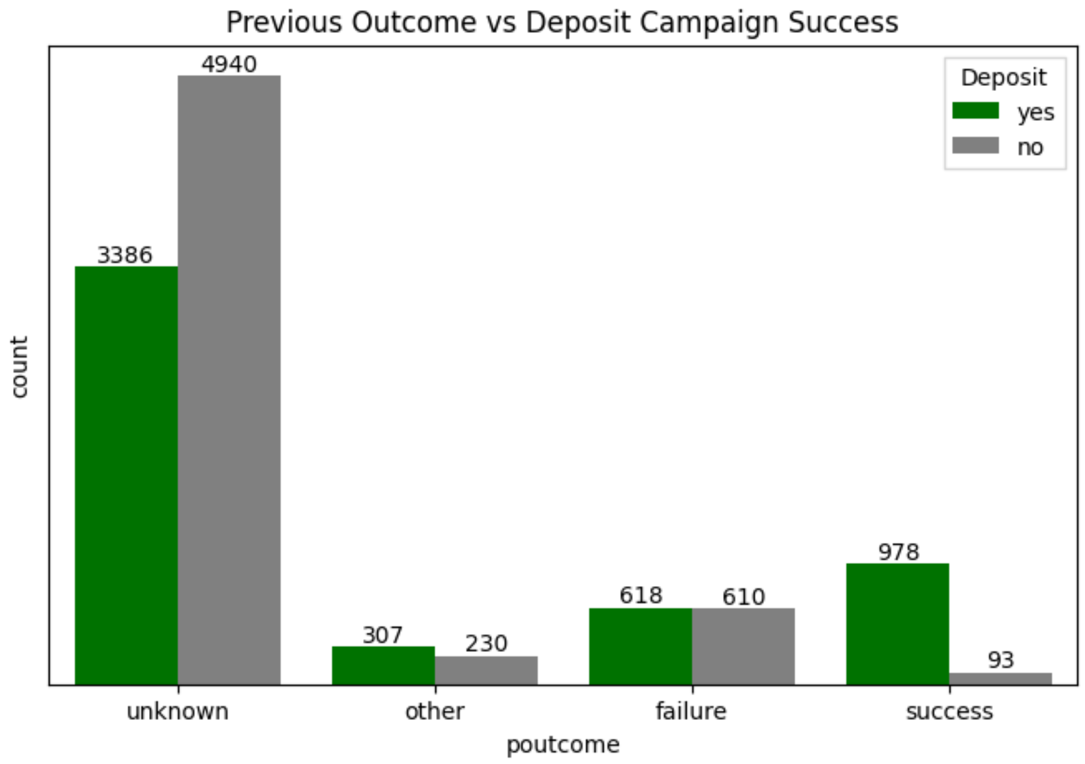
- Deposit acceptance ratio peaks in March, April, September, and October.

Age and Bank Balance Trends: Older Customers and Deposit Acceptance



- Older customers often have more money in their accounts.
- Customers who said 'yes' to the deposit offer tended to have higher account balances than those who said 'no.'
- It's worth noting that while there's a pattern, the plot doesn't provide strong statistical evidence for our analysis.

Previous Deposit Acceptance Strongly Predicts Future Acceptance



- Customers who previously accepted the deposit are significantly more likely to accept it again.

Modeling Methods

Target Variable

- The outcome variable defines our mission – predicting the success of telemarketing calls for long-term bank deposits.
- Importance: **Campaign Effectiveness**, **Decision Driver**
- The success or failure of our telemarketing efforts hinges on understanding this variable.
- Focus: All roads lead back to this variable in our data exploration and analysis.
- Business Impact: It's not just about prediction; it's about shaping future marketing campaigns.

Features: Strategic Selection

- **Focused Insights:** Each feature selected serves a strategic purpose, aligning with our project objectives.
- **Informed Decisions:** Chosen features are key indicators influencing our outcome variable.

Key Groups:

1. Consumer: Features are categorized into groups, such as demographic and interaction patterns.
2. Actionable Data: Emphasis on features that provide actionable insights for campaign improvement.

Selection Criteria:

1. Relevance: Features were chosen based on their relevance to telemarketing success.
2. Data Availability: Prioritized features with complete and reliable data for accurate modeling.

Logistic Regression

Why Logistic Regression?

- It’s a classification model that will predict if a person will subscribe long term deposit, based on demographic and previous campaign data.

Example:

- If a person has a history of subscribing, likes talking on the phone, and is older, model predicts a higher chance of subscribing.

Why this Logistic Regression?

- Transparent and straightforward, Logistic Regression helps us see the clear influence of each feature on the prediction.

Curious?

- Dive Deeper: [1 Logistic Regression]

Findings

Unveiling Strategic Insights

Model Performance:

- Subscription Predictions: The model accurately predicts customer subscription outcomes with 81% accuracy.
- Success Stories: Successfully identifying 833 subscribers.
- Learning Opportunities: Recognizing 194 instances of misjudged subscriptions.
- Subscribed: 81% of predicted subscriptions genuinely resulted in a “yes”.
- Not Subscribed: 83% of predicted non-subscriptions were indeed “no”.

Key Findings:

1. Duration: The duration of the call is the most influential factor, positively impacting subscription likelihood significantly.
2. Previous Campaign Success: Clients with a successful outcome in the previous campaign are more likely to subscribe.
3. Month of March and December: Strongly correlates with increased subscription likelihood.

Recommendations & Next Steps

Recommendations

1. Duration Matters: **Optimize call duration**
    - Actionable Insight: Implement measures to encourage and facilitate longer customer interactions during telemarketing calls.
    - Agent Training, Script Optimization, and Incentive: Equip agents with communication skills that foster customer engagement, refine call scripts to encourage interaction and effectiveness, and introduce bonuses tied to call duration and selling product (deposits)
2. Seasonal Patterns: **Leverage Successful Months**
    - Actionable Insight: Strategically plan and intensify marketing efforts during months with historically high subscription rates.
3. Past Success Indicator: **Leverage Previous Success**
    - Actionable Insight: Prioritize engagement with customers who have a positive history of subscription acceptance, tailoring marketing strategies to their preferences.

Next Steps

1. Advance modeling exploration and addressing misjudged predictions:
    - Opportunity: Investigate advance modeling techniques to address incorrect predictions by the current model.
    - Recommendation: Explore state of the art machine learning and deep learning models to build a more sophisticated models for improving accuracy.

2. Data expansion for comprehensive analysis:
    - Opportunity: Enrich the dataset by collecting additional relevant data points.
    - Recommendation: Expand the project scope to include a broader range of features for a more comprehensive understanding of customer behavior.

Appendix

Additional Information

- Git Repo: GitHub

Logistic Regression

- Logistic Regression is a statistical model that employs the logistic function to estimate the probability of a binary outcome, making it well-suited for classification tasks; in our context, predicting whether a client will subscribe to a long-term deposit.
- Mathematical Representation:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

- Key Components:
  - p: Probability of the positive class (subscription)
  - β_0: Intercept
  - $\beta_1, \beta_2, \ldots, \beta_n$: Coefficients for each feature
  - $X_1, X_2, \ldots, X_n$: Values of corresponding features
- Applied level:
  - Features used: Client data (age, bank balance, job type, etc.) and previous campaign data (duration, contact type, etc.).
- Practical Insight:
  - The model reveals the impact of each feature on the likelihood of subscription.