Case Study: Predicting Customer Response to Bank Marketing Campaigns

1. Business Case Narrative

A mid-sized European retail bank runs marketing campaigns to encourage customers to open long-term deposit accounts. Campaigns are conducted via phone calls, emails, and in-branch outreach.

Despite considerable investment, only a small percentage of customers respond positively. The bank's analytics team has been tasked with improving campaign efficiency by identifying which customers are most likely to subscribe.

Your role:
- As a data analytics consultant, you must develop predictive models, extract insights, and recommend an optimal targeting strategy.

2. Data Description (from UCI Repository)

Dataset Source: UCI Bank Marketing Dataset
(https://archive.ics.uci.edu/ml/datasets/bank+marketing)

- Number of records: ~41,000
- Target variable: y (binary: 'yes' = subscribed to term deposit, 'no' = did not)

Feature groups:
- Demographics: age, job, marital, education
- Financial info: balance, loan, housing
- Contact details: contact, day, month, duration (of last call)
- Campaign history: campaign (number of contacts in current campaign), pdays, previous, poutcome
- Macro environment: emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m, nr.employed

3. Learning Objectives

By the end of the case, you should be able to:
1. Perform data exploration on imbalanced datasets.
2. Apply predictive modeling techniques (logistic regression, decision trees, ensemble methods, etc.).
3. Use resampling or cost-sensitive methods for imbalance.
4. Evaluate models using business-relevant metrics (accuracy, sensitivity, specificity, etc.).

5. Interpret feature importance and generate actionable business insights.
6. Translate findings into marketing strategy recommendations.

4. To Do's

Part A: Data Understanding & Preparation

- Summarize dataset characteristics.
- Explore class balance: what proportion of customers subscribed?
- Identify most relevant variables — which are categorical vs numerical?
- Perform feature engineering (e.g., binning age groups, creating call frequency ratios, etc.).
- Are there features that should be excluded, transformed, etc.?

Part B: Modeling

- Train at least two predictive models:
  1. Logistic Regression
  2. LDA
- Handle class imbalance appropriately.
- Compare models using AUC, accuracy, sensitivity-specificity tradeoffs, etc.

Part C: Insights

- What are the top predictors of subscription?
- Create customer personas/segments most likely to respond.
- Are there surprising or counterintuitive features?

Part D: Strategy & Recommendations

- If the bank can only contact 20% of its customers, who should they target?
- What is the expected lift in conversions compared to random targeting?
- Discuss ethical considerations: could targeting reinforce bias (e.g., excluding older customers)?

6. Deliverables

- Write the case-study report

- Submit case-study report. One per group. Mention names of group members. .

- Upload the .R script or .Rmd files for your analysis.