Executive Summary

Objective

Maximize revenue from a direct marketing campaign by targeting the top 15% of clients most likely to purchase financial products: credit cards, consumer loans, and mutual funds. A data-driven approach was implemented to achieve this goal.

Approach

The framework involved several key steps:

1. **Exploratory Data Analysis (EDA)**:
    - Identified key patterns, correlations, and outliers in client demographics, transactions, and product data.
    - Ensured data quality by addressing missing values and inconsistencies.
2. **Feature Engineering**:
    - Applied class balancing techniques (e.g., class weights) to address dataset imbalances.
    - Selected key features based on their correlation with purchase and revenue, and scaled them as we are selecting between tree or regression models.
    - Outliers were not removed as Random Forests are robust to them.
3. **Model Development and Selection**:
    - Classifier to estimate likelihood of purchase and Regressors to predict revenue, training only on the positive class.
    - Baseline models included Logistic Regression and Random Forests for classification, and Random Forest Regressors for revenue prediction.
    - Hyperparameter optimization with Optuna identified the best-performing models.
    - Models were trained on the entire training data after obtaining the best parameters, simulating production.
    - Model Performance:
        - **Sale_CC Random Forest**: ROC-AUC = 0.825
        - **Revenue_CC Random Forest**: MSE = 929.853
        - **Sale_CL Logistic Regression**: ROC-AUC = 0.646
        - **Revenue_CL Random Forest**: MSE = 77.976
        - **Sale_MF Random Forest**: ROC-AUC = 0.981
        - **Revenue_MF Random Forest**: MSE = 350.486
4. **Optimization**:
    - Expected revenue for each client-product combination was calculated as:
      $\text{Expected Revenue} = P(Y=1 \mid X) * E[R \mid X, Y=1]$
      Where:
        - $P(Y=1 \mid X)$: Probability of purchase from the classifier model.
        - $E[R \mid X, Y=1]$: Predicted revenue from the regression model, focusing on positive cases.
    - Clients were assigned to products with the highest expected revenue, and the top 15% were targeted.

Why This Approach? This structured approach leverages Random Forest's ability to handle non-linear relationships and feature interactions while preventing overfitting through ensemble learning. Combining classification and regression models balances purchase likelihood with revenue potential, while focusing regressors on positive cases eliminates noise from non-purchasers, maximizing prediction accuracy for high-value clients.

Key Insights

Credit Cards

- Higher savings and spending activity increased ownership, while debts reduced it.
- Cashless transactions drove revenue, but older clients and those with high savings contributed less.

Mutual Funds

- Wealthy clients with existing accounts were more likely to buy.
- Disposable income strongly influenced revenue; older clients contributed less.

Consumer Loans

- Younger clients and those with higher overdrafts required loans.
- Older clients generated higher loan revenue; high credit card balances reduced loan dependency.

Results

- **90 clients** (15% of the dataset) were selected:
  - **56 clients** for Credit Cards
  - **21 clients** for Consumer Loans
  - **13 clients** for Mutual Funds
- Total expected revenue: **$1,050.43**
  - Likely to be overestimated due to high MSE from Revenue_CC random forest.

Recommendations

- Retrain models periodically to adapt to evolving client behavior.
- Incorporate calibration methods (e.g., isotonic regression) for better probability predictions.
- Leverage real-time data for improved targeting.
- Explore ensemble methods to enhance model accuracy.

Disclaimer

This analysis is based on historical data and predictive modeling techniques. While every effort has been made to ensure accuracy, results may vary due to changes in client behavior, market conditions, or data quality. This framework is intended to provide strategic insights for decision-makers. Implementation should be carried out by relevant stakeholders, with periodic validations to ensure its continued effectiveness and alignment with business goals.

This framework demonstrates the potential of data-driven strategies for direct marketing, enabling scalable revenue maximization through strategic client selection and product allocation.