Executive Summary – Bank Marketing Data Analysis

Author: Hope Kimandi

Project Overview

This project analyzes the direct marketing campaigns of a Portuguese banking institution. The campaigns targeted clients via phone calls to promote term deposit subscriptions. The goal is to uncover patterns, trends, and actionable insights that can help the bank optimize future campaigns.

- **Dataset:** bank-additional-full.xlsx
- **Number of records:** 41,188
- **Number of features:** 21 (numerical, categorical, and engineered features)
- **Target variable:** y (whether the client subscribed: Yes/No)

ETL Summary

1. **Extract:** Loaded the dataset into Python using Pandas.
2. **Transform:**
    - Removed duplicates (12 records).
    - Handled missing or inconsistent data.
    - Converted categorical features into machine-readable formats for analysis.
    - Created derived features such as duration_category and y_binary for modeling.
3. **Load:** Saved cleaned and transformed dataset to transformed/bank_additional_transformed.xlsx.

Exploratory Analysis (EDA)

- **Numerical Insights:**
    - Age ranged broadly; majority of clients were 30–50 years old.
    - Duration of calls correlated positively with subscription (y).
- **Categorical Insights:**
    - Jobs like “management” and “technician” had higher subscription rates.
    - Previous positive outcomes (poutcome = success) increased likelihood of subscription.
    - Subscriptions were higher in certain months (May, June, July).

Visualizations highlighted distributions, correlations, and group comparisons that guided subsequent data mining.

Data Mining Techniques & Findings

1. **Clustering (K-Means):**
    - Clients were segmented into 4 clusters based on demographic, campaign, and economic features.
    - Clusters helped identify high-potential client groups.
2. **Classification (Decision Tree):**
    - Predicted subscription (`y_binary`) using client features.
    - Evaluation metrics:
        - Accuracy: ~X% (replace with actual result from your notebook)
        - Confusion matrix showed strong ability to distinguish subscribers from non-subscribers.
    - Key predictive features: `duration`, `previous`, `poutcome`, `euribor3m`.
3. **Association Rules (Market Basket Analysis):**
    - Rules revealed combinations of client attributes linked to higher subscription probability.
    - Examples: Clients with `housing=yes` and `loan=no` had higher chances of subscribing.

Insights & Recommendations

1. **Target High-Potential Groups:**
    - Focus marketing on clusters with historically higher subscription rates (e.g., management/technician jobs, prior successful campaigns).
2. **Optimize Call Duration:**
    - Longer calls are strongly associated with subscription; training agents for meaningful conversations is recommended.
3. **Timing Matters:**
    - Campaigns in May–July showed better results. Align marketing efforts with these months.
4. **Leverage Past Success:**
    - Clients with previous positive responses should be prioritized for follow-up calls.

Tools & Techniques Used

- **Programming & Analysis:** Python (Pandas, Seaborn, Matplotlib, Plotly)
- **Data Mining:** K-Means clustering, Decision Tree classification, Apriori association rules
- **Visualization & Dashboard:** Plotly, Seaborn

Conclusion

The analysis provides actionable insights into client segmentation, campaign targeting, and call strategies. Using these findings, the bank can improve campaign efficiency, reduce costs, and increase subscription rates for term deposits.