Report on Bank Marketing Campaign Analysis

# Introduction

The purpose of this analysis was to evaluate a bank marketing dataset, identify patterns influencing customer subscription to term deposits, and develop predictive models to optimize future marketing campaigns. The work followed a structured approach, encompassing data cleaning, exploratory data analysis (EDA), feature engineering, and modeling to derive actionable insights and improve campaign outcomes.

# Data Preprocessing

## Steps Undertaken

1. Missing Values and Duplicates:  
- No missing values were identified.  
- Duplicate records were removed to ensure data integrity.  
2. Data Type Corrections:  
- Variables were correctly formatted to match their intended types (e.g., numerical, categorical).  
3. Unique Values Validation:  
- Ensured all categorical variables contained meaningful and interpretable categories.

# Exploratory Data Analysis (EDA)

## Key Insights

1. Demographic Characteristics:  
- Most customers are aged 30–45.  
- Dominant job roles include management, technical, and blue-collar jobs.  
- Secondary and tertiary education levels are predominant.  
2. Financial Profile:  
- Majority of accounts hold balances below €20,000, with many clustered around €0.  
3. Engagement Patterns:  
- Peak activity occurs mid-month (days 15–20), with notable engagement on days 5–10 and 25–30.  
- Campaign efficiency declines after 3–5 contact attempts.  
- Most customers are first-time contacts, with 0 previous interactions in the dataset.  
4. Target Variable (`y`):  
- Approximately 12% of customers subscribed to term deposits.  
- Subscription likelihood increases with longer call durations and efficient timing.

# Correlation Analysis

## Key Findings

Positive Correlations:  
- Call duration (0.39) strongly correlates with subscription likelihood.  
- Successful prior campaign outcomes (0.27) positively impact current subscriptions.  
Negative Correlations:  
- Previous contact attempts (-0.093) and past campaign failures (-0.078) negatively correlate with subscription success.  
- Days since last contact (-0.86) indicates quicker follow-ups for successful clients.

# Modeling

## Models Evaluated

Algorithms Used:  
- Logistic Regression  
- Support Vector Classifier (SVC)  
- Random Forest  
- Decision Tree  
- XGBoost  
- K-Nearest Neighbors (KNN)

Preprocessing for Models:  
- Yeo-Johnson transformation was applied to normalize features.   
- OneHotEncoder was applied to transform the categorical features.   
- Both balanced (SMOTE oversampling and random undersampling) and unbalanced datasets were tested.

# Model Performance

1. XGBoost:  
- AUC: 0.93  
- Strongest overall performance, excelling in handling imbalanced data.  
2. SVC and Random Forest:  
- AUC: 0.92  
- Both models performed well with balanced data, achieving high True Positive Rates (89% and 93%, respectively), These were the chosen models since they had the highest TPR  
3. Logistic Regression:  
- AUC: 0.90  
- Reliable baseline model with robust prediction accuracy.  
4. KNN and Decision Tree:  
- Lower performance, with AUCs of 0.88 and 0.79, respectively.

# Insights and Recommendations

## Key Insights

1. Customer Characteristics:  
- Younger and retired customers show higher conversion rates.  
- Customers with secondary and tertiary education exhibit strong engagement.  
2. Optimal Campaign Strategies:  
- Focus on mid-month and end-of-month days, particularly in March, September, October, and December.  
- Limit contact attempts to 3–5 and prioritize longer call durations (500–1500 seconds).  
3. Contact Methods:  
- Cellular communication is the most effective channel.

## Recommendations

1. Targeted Engagement:  
- Focus campaigns on high-conversion segments: students, retired individuals, and those with secondary or tertiary education.  
2. Efficient Contact Strategy:  
- Ensure longer call durations while limiting contact attempts to avoid customer fatigue.  
3. Model Deployment:  
- Implement SVC or Random Forest models for campaign predictions, leveraging their superior performance metrics.  
4. Seasonal Adjustments:  
- Optimize campaigns during high-conversion months, using insights from previous outcomes to prioritize re-engagement.

# Conclusion

The analysis revealed that customer demographics, engagement patterns, and call durations significantly influence term deposit subscriptions. The implementation of predictive models such as SVC and Random Forest can greatly improve campaign efficiency and customer targeting. By integrating these findings into future strategies, the bank can enhance conversion rates and maximize marketing ROI.