**Suggestions for the Bank Marketing Team to Increase Term Deposit Subscriptions**

**Key Findings from Data Analysis**

* **Previous engagement is a strong predictor**: Customers who were contacted in past campaigns and had a successful outcome are more likely to subscribe again.
* **Economic conditions affect subscription likelihood**: Employment rates and financial stability indicators correlate with customer decisions.
* **The timing of marketing calls matters**: Some months perform significantly better than others in converting customers.
* **Excessive calls reduce conversion rates**: Too many follow-ups result in lower chances of subscription.
* **Customer demographics influence success**: Factors like job type, education, and financial history impact a customer's decision.

**Actionable Strategies for the Bank Marketing Team**

**1. Retarget Engaged Customers from Previous Campaigns**

📊 **Insight:** Customers who responded positively in past campaigns have a higher likelihood of subscribing again.

✅ **Strategy:**

* Prioritize customers who had a **successful outcome** in previous campaigns.
* Use **personalized follow-ups** for those who showed interest but did not subscribe previously.

**2. Optimize Contact Strategy Based on Past Campaign Performance**

📊 **Insight:** The outcome of previous contacts (poutcome) is a strong predictor of future behaviour.

✅ **Strategy:**

* **Prioritize customers** with a history of positive responses.
* Adjust messaging and approach for customers who previously declined.

**3. Target High-Performing Months for Marketing Efforts**

📊 **Insight:** Subscription rates vary significantly across different months.

✅ **Strategy:**

* **Increase call volume** in months with historically high success rates.
* For low-performing months, introduce **limited-time promotions** to attract more customers.

**4. Limit the Number of Calls per Customer**

📊 **Insight:** Excessive calls (campaign feature) negatively impact subscription rates.

✅ **Strategy:**

* Reduce the number of follow-up calls and focus on **quality interactions**.
* Use **customer response patterns** to determine optimal call frequency.

**5. Segment Customers Based on Demographics**

📊 **Insight:** Certain job roles and education levels have higher subscription rates.

✅ **Strategy:**

* Personalize marketing messages based on **job type and education level**.
* Provide financial education to customers with lower awareness of term deposits.

**6. Address Customer Concerns About Loans and Financial Stability**

📊 **Insight:** Customers with **multiple loans** are less likely to subscribe.

✅ **Strategy:**

* Educate customers on the **advantages of term deposits over risky investments**.
* Provide incentives for financially stable customers.

**7. Use SMS and Email for Follow-ups**

📊 **Insight:** Some customers prefer **non-intrusive communication** methods.

✅ **Strategy:**

* If a customer is unresponsive to phone calls, follow up via **SMS or email**.
* Provide easy **online subscription options** to increase convenience.

**4. Conclusion & Final Recommendations**

By implementing the above strategies, the bank can:

* Improve conversion rates while **reducing customer irritation from excessive calls**.
* Optimize marketing spending by focusing on **high-potential customers**.
* Align marketing efforts with **economic trends and customer preferences**.

**Challenges Faced and Solutions Implemented**

**Project: PRCP-1000 - Portuguese Bank Marketing Campaign**

**1. Data Cleaning Challenges**

**Challenge: Missing and inconsistent values**

* **Issue:** The dataset contained missing values in the **pdays** column (replaced with 999 for "not contacted") and some categorical columns (e.g., **default** had 'NA' values).
* **Solution:**
  + **pdays**: Replaced 999 with NaN and then applied **MinMax scaling** for normalization.
  + **default column**: Replaced 'NA' values with 'unknown' to maintain consistency.
  + Used data.isnull().sum() to check for missing values and fillna() for appropriate imputation.

**2. Handling Class Imbalance**

**Challenge: The dataset was highly imbalanced (more "No" than "Yes" in target variable y).**

* **Issue:** The imbalance could cause the model to favor the majority class, leading to poor recall for "Yes" cases.
* **Solution:**
  + Applied **class weighting** (class\_weight='balanced') in models like Logistic Regression, Decision Tree, and Random Forest.
  + Applied **compute\_sample\_weight()** in model training for fair weighting.

**3. Encoding Categorical Variables**

**Challenge: Categorical data needed proper encoding for ML models.**

* **Issue:**
  + **High-cardinality categorical features** like **job** and **education** required careful encoding.
  + Some models perform better with different encoding techniques.
* **Solution:**
  + **One-Hot Encoding** for low-cardinality features: marital, poutcome, contact
  + **Label Encoding** for ordinal features: education, month
  + **Binary Encoding** for binary categorical features: default, housing, loan.
  + **Frequency Encoding** for features: job, day\_of\_week.
  + Used ColumnTransformer to apply encoding efficiently.

**4. Feature Scaling**

**Challenge: Numerical features had different ranges.**

* **Issue:**
  + Features like **pdays** had a large range difference compared to **age** or **campaign**, which could bias models.
* **Solution:**
  + **MinMaxScaler** for **pdays** and **previous** (to keep values between 0 and 1).
  + **StandardScaler** for **age, campaign, cons\_price\_idx, cons\_conf\_idx, euribor3m** to normalize feature distribution.

**5. Model Performance & Overfitting**

**Challenge: Certain models overfitted the training data.**

* **Issue:**
  + **Decision Tree & Random Forest** showed very high accuracy on training but lower accuracy on test data, indicating overfitting.
* **Solution:**
  + Used **GridSearchCV** for hyperparameter tuning.
  + **Gradient Boosting & XGBoost** provided better generalization due to built-in regularization.

**6. SVM Model Running Too Slow**

**Challenge: SVM took too long for training.**

* **Issue:**
  + The dataset was large, and SVM with **RBF kernel** was computationally expensive.
* **Solution:**
  + Reduced feature set using **feature importance analysis**.
  + Used **linear kernel** for faster computation.
  + Considered **reducing the dataset size** for SVM training.