Bank Marketing Project Report

1. Introduction

Overview

This project analyzes a dataset titled **"Direct Marketing Campaign for Bank Term Deposit"**, which captures data related to a marketing campaign conducted by a financial institution. The primary medium of the campaign was through phone calls to potential customers.

Problem Statement

Identifying the key factors influencing customer decisions was necessary for effectively predicting customer subscriptions to term deposit offers and improving the efficiency of marketing strategies.

Objective

The objective of this project was to identify key factors influencing customer decisions to subscribe to a term deposit offered by the bank through comprehensive data analysis. Based on these insights, the project aimed to predict subscription likelihood and support future marketing strategies for better customer targeting and conversion rates.

2. Dataset Overview

Description

The dataset is titled **"Direct Marketing Campaign for Bank Term Deposit"**. It contains customer data collected during direct marketing campaigns conducted by the bank. The data includes:

- **Demographics**: Age, job, marital status, etc.
- **Socioeconomic Attributes**: Education level, housing loan status, personal loan status.
- **Campaign Details**: Number of contacts during the campaign, previous campaign outcomes, and duration of the last call.

Data Collection

The dataset was sourced from direct marketing calls made by the bank’s marketing team. The calls aimed to persuade customers to subscribe to a bank term deposit. Customers' responses, alongside related data, were recorded systematically to allow for further analysis.

3. Exploratory Data Analysis (EDA)

- **Data Collection & Importing**
    - Gathering and importing the dataset into the analysis environment.

- **Data Cleaning**
    - Handling missing data, duplicates, and correcting inconsistencies.
    - Removing or imputing missing values.
    - Identifying and correcting outliers or errors in the data.

- **Data Transformation**
    - Converting data types where necessary.
    - Normalizing, scaling, or encoding categorical variables.

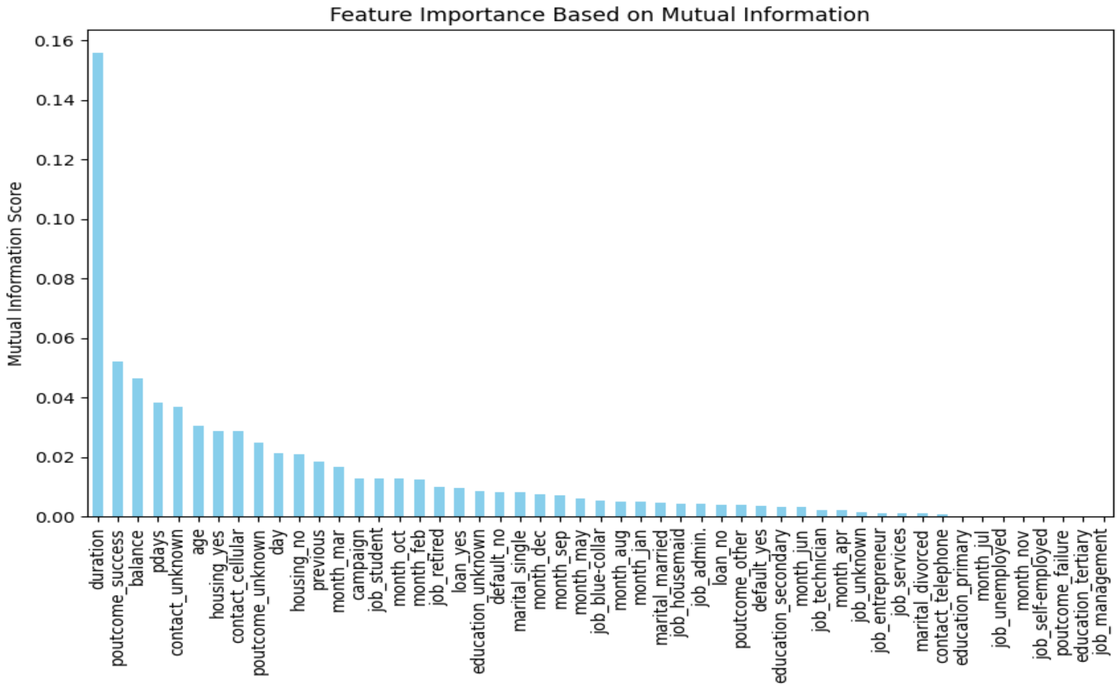
- **Descriptive Statistics**
    - Calculating measures like mean, median, mode, standard deviation, variance, etc.
    - Summarizing data distributions and checking for skewness and kurtosis.

4. Feature Importance Analysis and Data-Driven Insights for Decision Making

Feature Importance Analysis

- **Methodology**
  To identify the most influential features for predicting term deposit subscriptions, we used **Mutual Information Classification**. This method measures the dependency between input features and the target variable, allowing us to rank features based on their predictive power.

- **Top Features Identified:**



Using the analysis, the following features were ranked as the most influential:

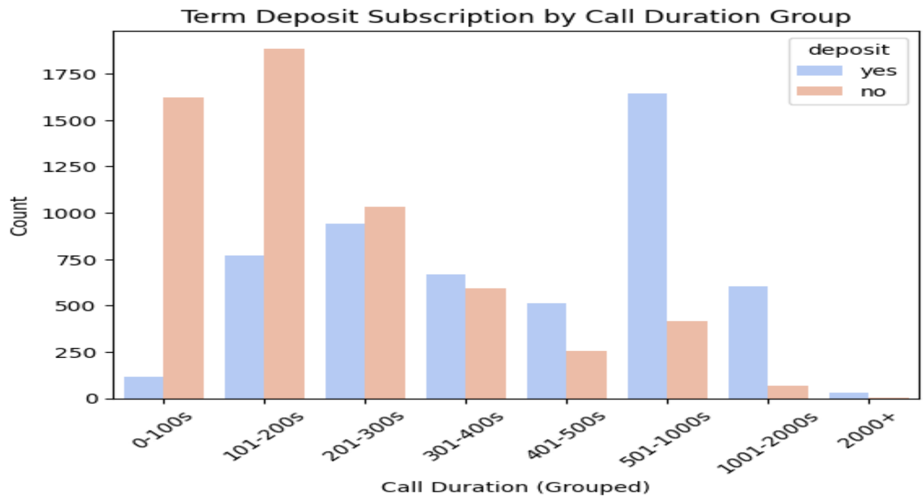
1. **Duration** of the last call (highest correlation with subscription likelihood).
2. **Previous Outcome** of prior campaigns (success or failure).
3. **Contact Count** during the campaign.
4. **Age** of the customer.
5. **Job** type and **Education** level.

In-Depth Analysis of Influential Features

For each of the top features, we conducted an in-depth analysis to understand their impact on subscription likelihood:

1. **Duration of Last Call:**

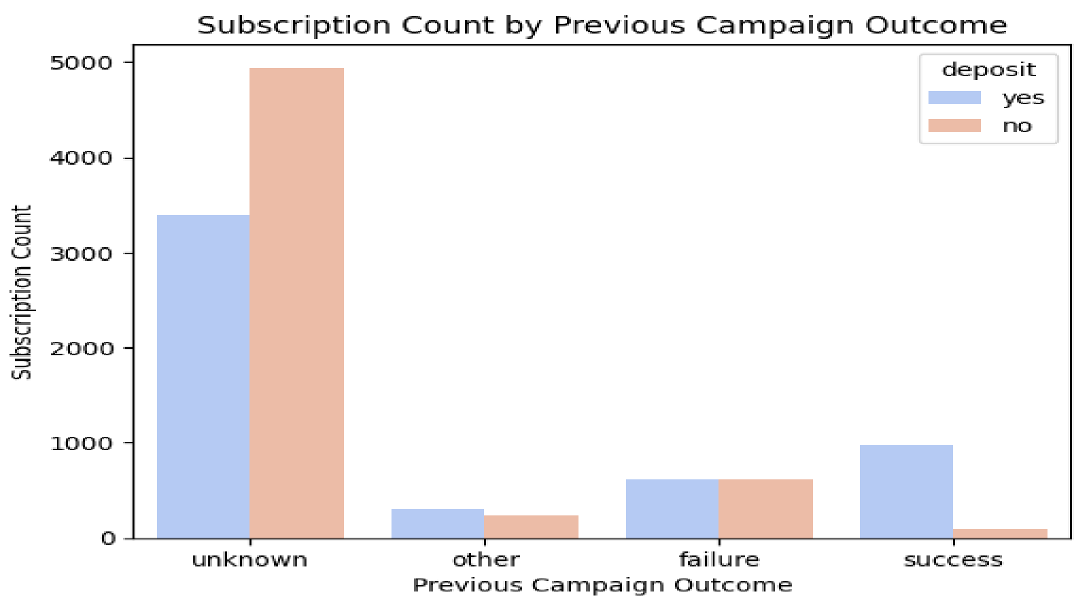
Objective: Explore how call duration impacts term deposit subscription.



- Observations showed that longer call durations had a significantly higher conversion rate.
- Customers with call durations exceeding 5 minutes were more likely to subscribe.
- **Recommendation:** Encourage call center agents to focus on engaging conversations with potential customers.

2. **Previous Campaign Outcome:**

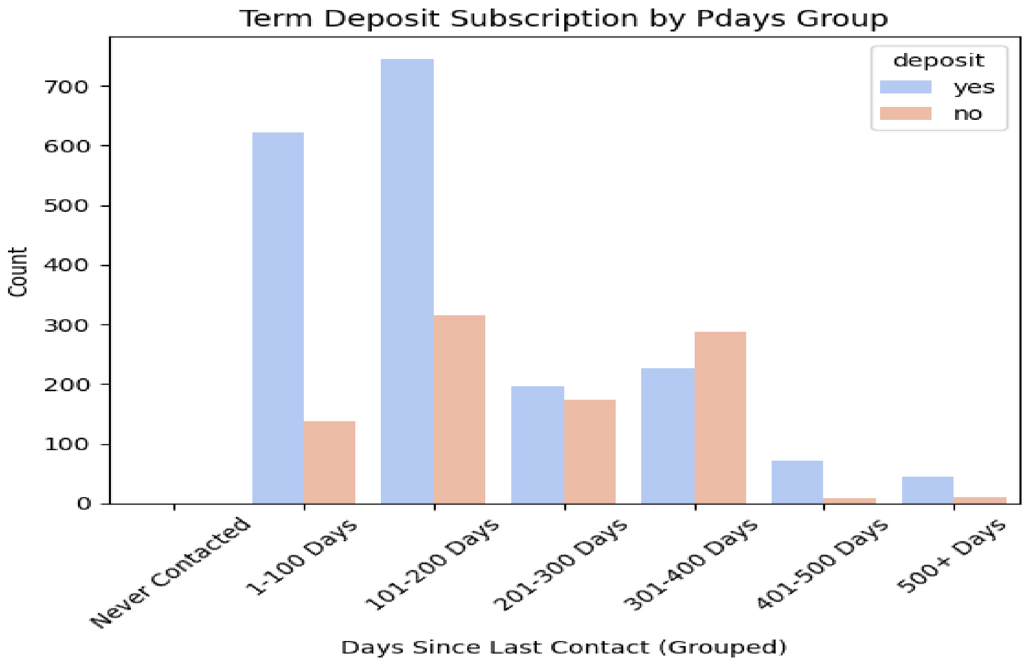
Objective: Analyze the impact of successful outcomes in previous campaigns.



- Observation showed Customers with **successful outcomes** were more likely to subscribe, while **failed or unknown outcomes** reduced subscription likelihood.
- **Recommendation**: Focus on re-targeting customers with prior success and re-engage failed outcomes with tailored offers.
    - Improve data tracking to reduce the "unknown" category.

3. **Contact Count During Campaign:**

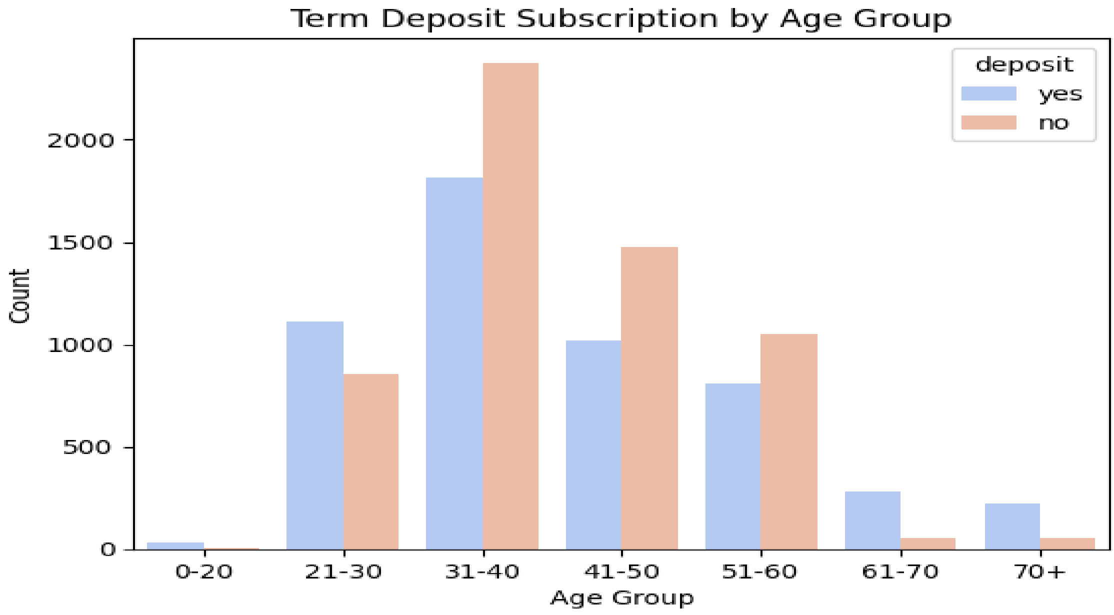
   Objective: Examine how the recency of contact impacts subscription.



- Moderate contact frequency (3–5 calls) yielded the highest subscription rate.
- Excessive contact attempts led to reduced effectiveness and potential customer fatigue.
- **Recommendation**: Develop a strategy to follow up with customers sooner after previous contact. Limit the number of follow-up calls to avoid customer irritation.

4. **Age of Customers:**

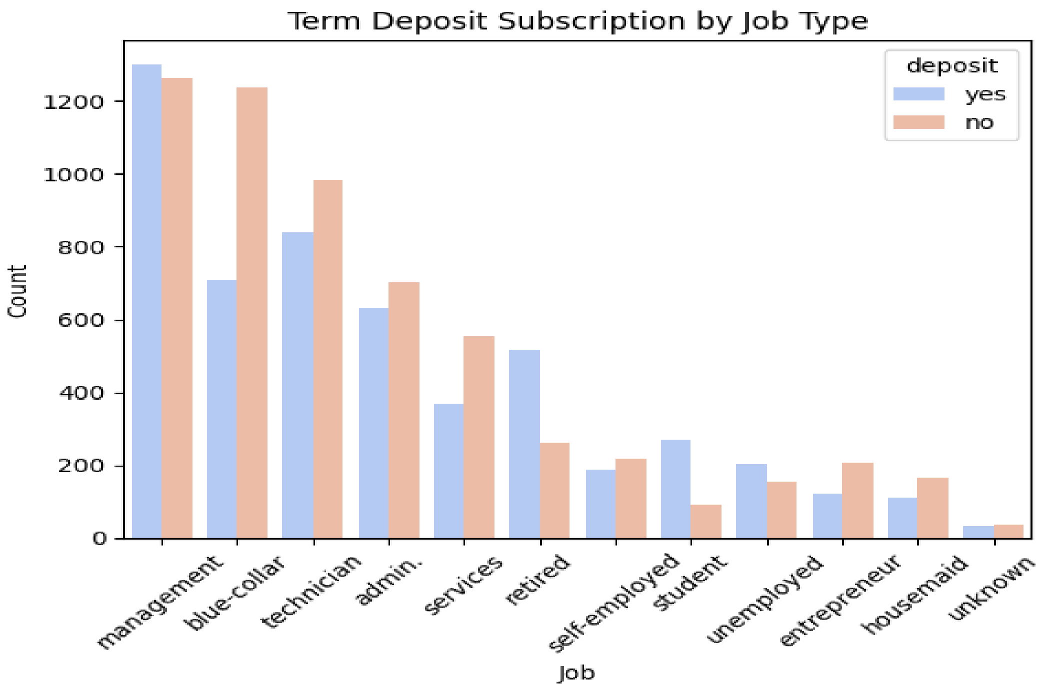
Objective: Analyze how age influences term deposit subscriptions.



- Subscription rates were higher among middle-aged customers (30–50 years).
- Younger and older customers showed lower engagement.
- **Recommendation:** If specific age groups show higher subscription rates, tailor marketing efforts (e.g., personalized offers or language) for those demographics.

5. **Job Type and Education Level:**

Objective: Examine how occupation impacts subscription likelihood.

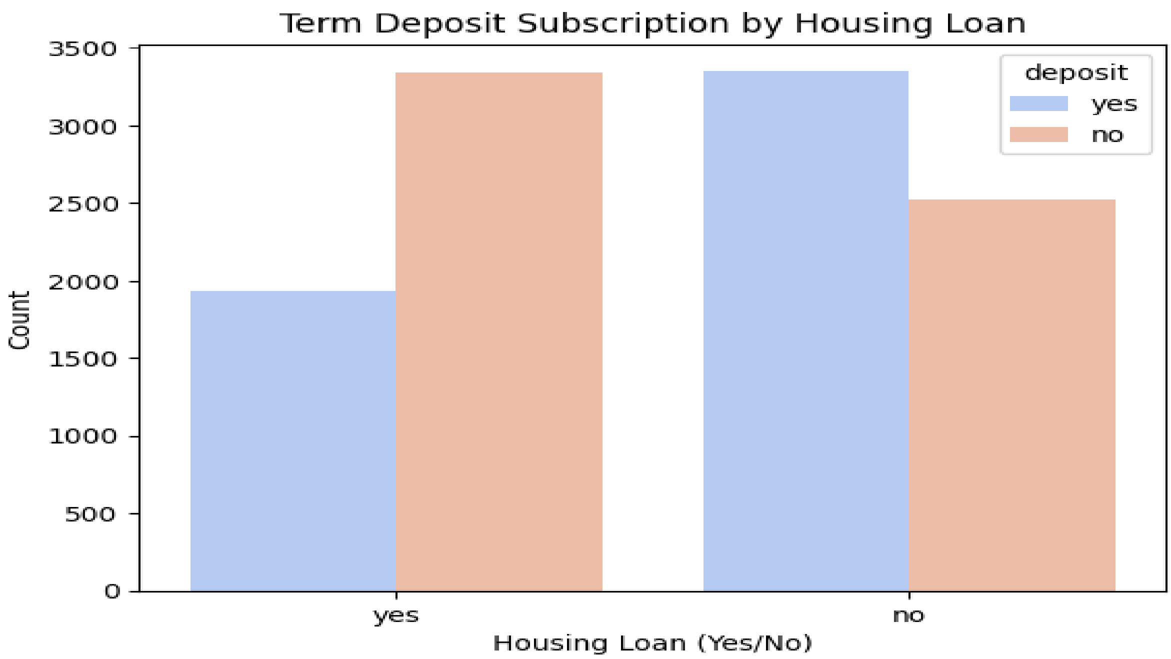


- Customers in **management**, **technician**, and **retired** job categories have higher subscription rates (blue bars are more prominent).
- **Blue-collar** and **admin** jobs show a higher number of "no" responses (orange bars dominate).

Recommendation: Prioritize campaigns for **management**, **retired**, and **technician** job types, as they are more likely to subscribe.
Tailor campaigns based on job profiles. For instance, retirees might be interested in secure savings, while students may respond to short-term savings plans.

6. Housing Loan Status:

Objective: Study how having a housing loan impacts subscription likelihood.



- Customers without a housing loan were more likely to subscribe to a term deposit compared to those with an active housing loan.
- This may indicate that customers with existing financial commitments are less inclined to take on additional products.

Recommendation: Tailor campaigns to customers without active housing loans as they exhibit higher interest in term deposit products.

For customers with housing loans, promote smaller deposit amounts or flexible options to address their financial constraints.

Data-Driven Strategies

Based on the analysis, the following strategies were recommended to improve subscription rates:

- Focus campaigns on customers with successful prior outcomes and longer call durations.
- Limit the number of follow-ups to reduce customer fatigue while maintaining engagement.
- Develop age-specific marketing strategies, especially targeting middle-aged customers.
- Personalize offers for professionals and highly educated individuals to improve relevance.
- Train call center agents on effective conversational strategies to optimize call duration.
- Focus marketing efforts on customers without active housing loans to maximize conversion rates.
- Provide flexible or tailored term deposit options to customers with housing loans to make offers more appealing.

Model Building And Preprocessing

- **Feature Selection**: I selected the important features based on their mutual information score greater than 0.
- **Data Splitting**: The dataset was split into 80% training and 20% test sets.
- **Feature Scaling**: I scaled the features using StandardScaler to ensure equal contribution of features, as Logistic Regression is sensitive to feature scaling.
- **Model Training**: I trained a Logistic Regression model on the scaled training data.

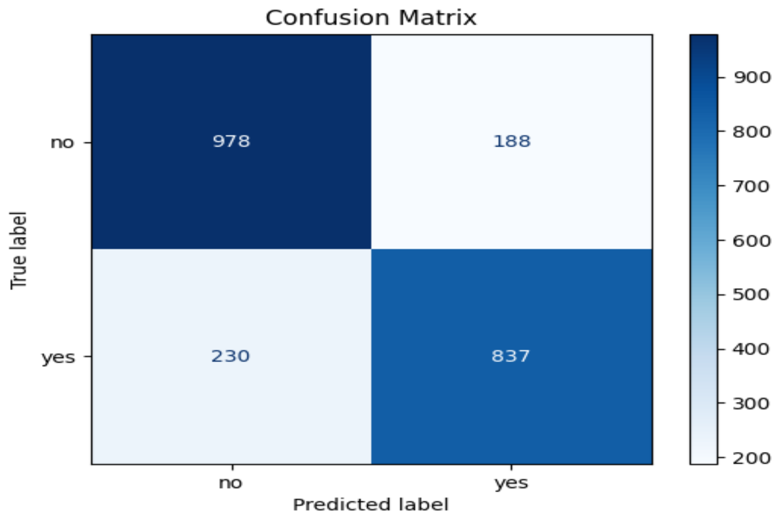
Model Evaluation

```
# Evaluate the model (optional: include additional metrics)
train_score = model.score(X_train, y_train)
test_score = model.score(X_test, y_test)

print(f"Training Accuracy: {train_score:.2f}")
print(f"Test Accuracy: {test_score:.2f}")
```

```
Training Accuracy: 0.83
Test Accuracy: 0.81
```

The model achieves a **training accuracy of 83%** and a **test accuracy of 81%**, indicating it has generalized well to unseen data. The minimal 2% gap suggests the model effectively captures the patterns in the dataset without overfitting. This performance is appropriate for predicting the likelihood of a customer **making a deposit** based on the given features, making it a reliable tool for the task



The confusion matrix reflects the **model's performance** in predicting whether a customer will make a deposit:

1. **True Negatives (978)**:
   - The model correctly predicted 978 customers as **"no"**, meaning they will not make a deposit.
2. **False Positives (188)**:
   - The model incorrectly predicted **188 customers** as "yes" (likely to deposit), but in reality, they did not.
3. **False Negatives (230)**:
   - The model missed **230 customers** who were actual depositors but were predicted as "no."
4. **True Positives (837)**:
   - The model successfully predicted **837 customers** as "yes," indicating they would make a deposit.

```
Classification Report:
              precision    recall  f1-score   support

          no       0.81      0.84      0.82      1166
         yes       0.82      0.78      0.80      1067

    accuracy                           0.81      2233
   macro avg       0.81      0.81      0.81      2233
weighted avg       0.81      0.81      0.81      2233
```

- The model correctly predicts most cases (**1815 out of 2233**).
- **Recall for "yes" (78.4%)**: Captures most depositors but misses about **21.6%**.
- **Precision for "yes" (81.6%)**: Out of predicted depositors, 81.6% are accurate.
- **False Negatives (230)**: Missing depositors could lead to lost marketing opportunities.
- **False Positives (188)**: Targeting non-depositors wastes resources.

Conclusion for the Project

Project Conclusion

This project focused on predicting whether customers would subscribe to a deposit product, empowering the bank to optimize its marketing strategies. By leveraging **data-driven insights**, the model achieved an **81% accuracy**, with a solid balance between **precision (81.6%)** and **recall (78.4%)**, ensuring effective targeting of potential depositors while minimizing resource wastage on non-depositors.

Key highlights include:

- **Feature Selection**: Mutual Information reduced noise and improved model efficiency by retaining only impactful features.
- **Model Performance**: The logistic regression model provided explainability and consistent results, making it ideal for business use cases.
- **Business Value**: The model streamlines marketing efforts, reduces campaign costs, and increases customer conversions.

Impact

• **Reduced Marketing Costs**: By accurately targeting potential depositors, the bank can reduce resources spent on ineffective campaigns.
• **Increased Revenue**: High recall ensures the bank identifies most customers likely to make deposits, leading to better conversion rates.
• **Customer Engagement**: Data-driven predictions allow the bank to tailor personalized communication, improving customer experience and trust.