# Final Report: Customer Churn Analysis

## Introduction

This report summarizes the customer churn analysis performed using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. The goal of this analysis is to explore factors that contribute to customer churn in a banking dataset.

### 1. Business Understanding

* **Objective**: Identify customer behavior patterns and factors leading to churn.
* **Key Questions**:
  + What demographic factors are associated with churn?
  + How do account features influence customer retention?

### 2. Data Understanding

* **Data Source**: Customer churn dataset.
* **Data Overview**:
  + 10,000 records and 14 attributes, including demographic and account-related features (e.g., Age, Balance, Tenure).

### 3. Data Preparation

* **Data Cleaning**:
  + Removed irrelevant columns (RowNumber, CustomerId, Surname).
  + Checked for missing values and duplicates; none were found.
* **Feature Engineering**:
  + Categorical variables (Geography, Gender) converted to numerical format using Label Encoding.

### 4. Modeling

* **Models Applied**:
  + Logistic Regression
  + Decision Tree Classifier
* **Model Evaluation**:
  + Metrics used include accuracy, precision, recall, and F1-score.
  + Confusion Matrix generated for model predictions.

### 5. Evaluation

* **Findings**:
  + Demographic factors (e.g., Age, Balance) and account features significantly influence churn.
  + The logistic regression model demonstrated a balanced performance.

### 6. Deployment

* **Next Steps**:
  + Recommendations for retention strategies based on findings.
  + Continuous monitoring of customer behavior and model performance.

## Conclusion

The analysis provides insights into customer behavior, helping devise strategies to reduce churn and enhance business strategies.