HOME CREDIT INDONESIA SCORE CARD MODELING

By : Yoga Aprila
PBI Data Scientist - HCI
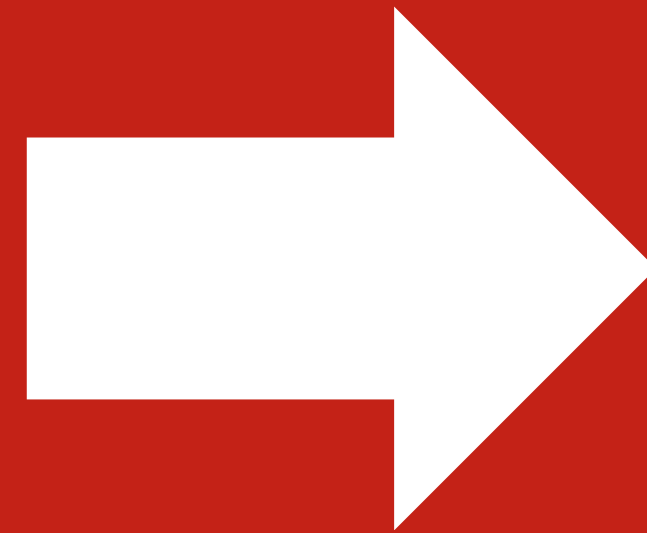
Table of Content

1. Problem Statement and Solution
2. Dataset, Goal, Objective, and Metrics
3. Data Visualization and Business Insight
4. Data Preprocessing
5. Machine Learning Implementation and Evaluation
6. Business Recommendation

Problem and Solution Statement

PROBLEM

Ensuring that customers who can repay their loans are not rejected to increase profitability and Identifying customers who are likely to face repayment issues to minimize loans granted to unqualified customers and reduce potential losses.
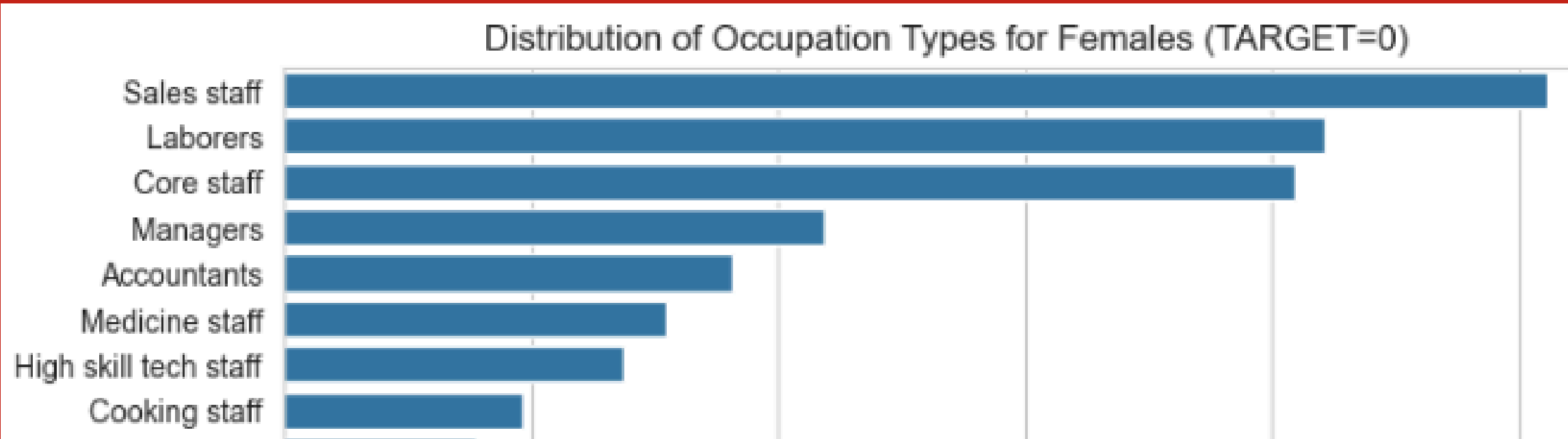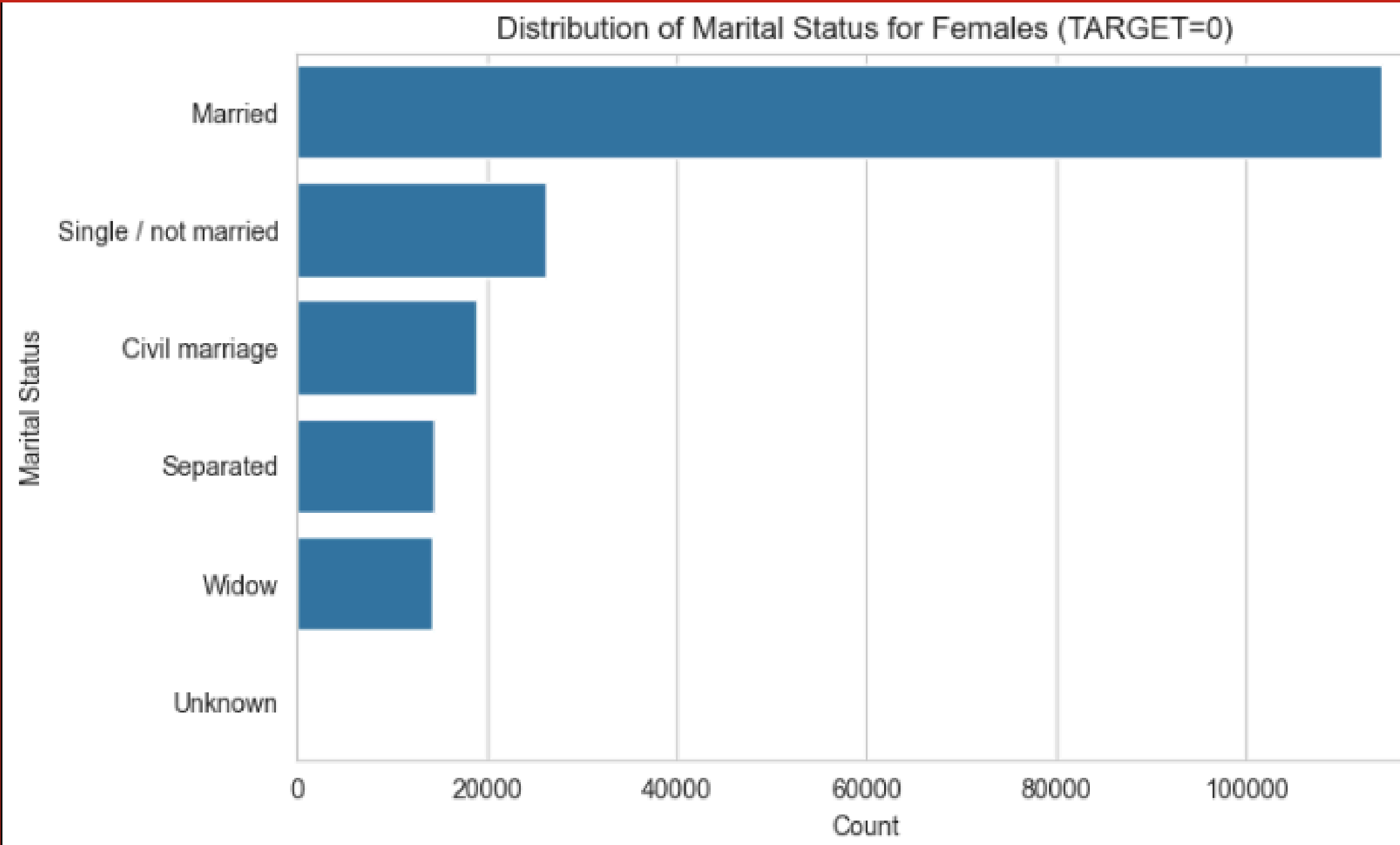
SOLUTION

MACHINE LEARNING

Dataset, Goal, Objective, Metric

Dataset	Goal	Objective	Metric
The dataset contains 307,511 customers who have taken loans, with 122 features such as ID, gender, income, etc.	A machine learning model will predict credit scores to identify clients' repayment capability and potential payment issues.	• Define Problem, Goals, Objective, and Metrics. • Perform EDA • Data Preprocessing • Modeling and Evaluation • Business Recommendation	Recall, Precision, and Balanced Accuracy

Data Visualization and Business Insight

Top 10 Occupations with Highest Loan Acceptance Probability
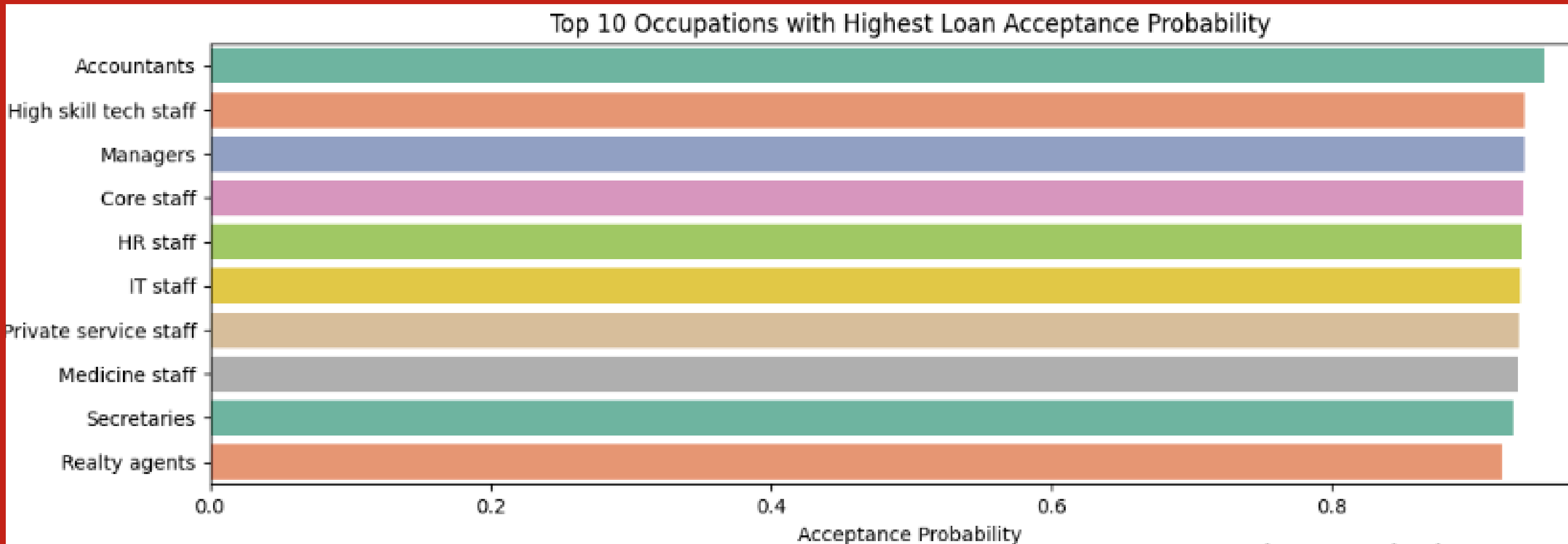
We can focus on occupations with high loan approval probabilities from the visualization and implement targeted campaigns to maximize profitability.

Distribution of Marital Status for Females (TARGET=0)

Distribution of Occupation Types for Females (TARGET=0)

The distribution of female customers with no payment issues is dominated by married women working as sales staff, laborers, and core staff. Targeted campaigns for these characteristics can help increase profits.

Data Preprocessing

1 Null Value
Drop columns with more than 20% missing values and impute the remaining missing data using the median and mode.

2 Unnecessary Column
Remove unrepresentative columns like ID and FLAG_DOCUMENT.

3 Anomalies Value
Replace anomalies (e.g., XNA) in Gender and Organization Type columns.

Data Preprocessing

4 **Feature Encoding**
Apply label encoding for ordinal categories and one-hot encoding for nominal categories.

5 **Class Imbalance**
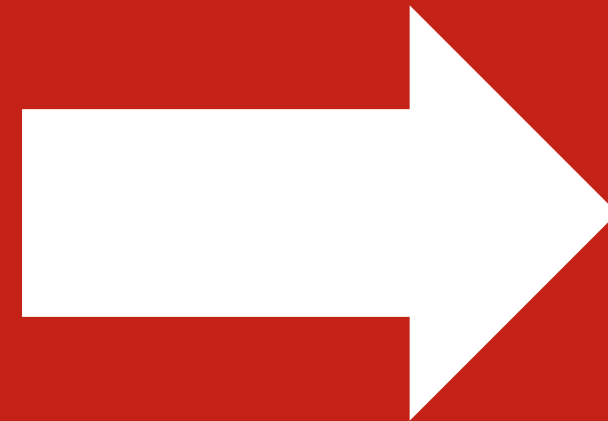Handle class imbalance using SMOTE, setting the minority class to 70% of the majority class.

6 **Outliers**
Address outliers using capping without discarding data

Machine Learning Implementation and Modeling

LOGISTIC REGRESSION

→

Balanced Accuracy : 0.93
Precision: 0.99
Recall: 0.87

Model is highly accurate in predicting customers who default (high precision) but could improve in detecting all defaulters (recall). The high balanced accuracy indicates strong overall performance, especially with imbalanced classes.

Business Recommendation

Insight	Recommendation	Action	Impact
Top Occupations with High Loan Acceptance Probability	Focus on job segments with high acceptance probabilities such as *Accountants*, *High Skill Tech Staff*, and *Managers*.	Create targeted marketing campaigns for these jobs on Social Media.	Increase loan application approvals
Characteristics of Female Borrowers with Smooth Payment	Focus on female clients, especially married women, and dominant occupations (*Sales Staff, Laborers, Core Staff, Managers*).	Provide rewards such as cashback or lower interest rates for women with smooth payments.	Boost loyalty among female clients.
Business Impact: Model Performance	Improve recall of the model to better identify high-risk customers.	Add external data (third-party credit history).- Implement a pre-approval system powered by machine learning.	Reduce losses due to defaults.
Feature Importance of for Risk Customer (TARGET=1)	Leverage important features like CNT_FAM_MEMBERS, AMT_CREDIT, and AMT_ANNUITY to enhance risk evaluation.	Design credit limits based on family size (CNT_FAM_MEMBERS) and repayment capacity (AMT_ANNUITY)	Decrease high-risk customers
Features Important for Low-Risk Customers (TARGET=0)	Target married customers and those in civil marriages with tailored loan offers based on family size (CNT_CHILDREN).	Provide family-oriented loan packages, such as education loans or home improvement loans.	Attract low-risk customers.