

Insights and Recommendations

Customer Loan Repayment

- **Repayment Rate:** 80.26% of customers have fully repaid their loan amount.

Feature Correlations

- **Loan Amount & Installment:** Highly correlated (Spearman correlation: 0.97).

Homeownership

- **Majority:** Mortgage ownership.

Loan Repayment and Loan Grade

- **Grade A:** 93% have fully repaid their loan (strong positive correlation).

Top Affordable Job Titles

- Teacher & Manager

Bank's Key Performance Metric

- **F1 Score:** Balances precision (avoiding false positives) and recall (avoiding false negatives).

Impact of Precision-Recall Gap

- **Low Recall:** Bank misses potential customers (lost opportunity cost).
- **Low Precision:** Bank approves loans to risky borrowers (increased defaults).

Loan Outcome Influencing Features

- **High Positive Impact:** Address, sub_grade, term, dti (higher values).
- **High Negative Impact:** Interest rate, annual income, total accounts, revolving credit utilization (higher values).

Geographical Location Impact

- **Yes:** Address significantly affects results.

Tradeoff Considerations

1. Balancing False Positives and False Negatives:

- **Threshold Adjustment:** Adjust the probability threshold for positive classification (e.g., from 0.5).
- **Evaluating Trade-offs:** Use the Precision-Recall curve to find the optimal balance.

2. Mitigating Non-Performing Assets (NPAs):

- **Recall Score:** Focus on maximizing recall to minimize approving defaulters (Type 2 error).

Insights - Categorical Attributes

- **Loan Term:** Higher default rates for 60-month terms.
- **Loan Grade:** Highest default rates for lower grades.
- **Removable Attributes:** Consider removing `initial_list_status`, `emp_title`, and `title` (low impact).
- **Application Type:** Higher default rates for direct pay applications.
- **Loan Purpose:** Highest default rates for small business loans.

Insights - Numerical Attributes

- **Defaulters:** Higher loan amounts, interest rates, debt-to-income ratios, open accounts, and revolving credit utilization.
- **Defaulters:** Lower annual income.

Model Performance

- **Model:** Logistic Regression (class-weighted)
- **Accuracy:** 80%
- **Negative Class:** Precision (94%), Recall (82%), F1 Score (87%)
- **Positive Class:** Precision (52%), Recall (82%), F1 Score (87%)
- **ROC AUC:** 0.90 (good class differentiation)
- **Precision-Recall AUC:** 0.78 (improvement potential)

Recommendations

- **F1 Score & Precision-Recall Trade-off:** Optimize for a balance between minimizing NPAs and maximizing opportunity cost.
- **Model Complexity:** Explore more complex models (e.g., Random Forest) for potentially better results.