Overview of the Analysis

The purpose of this analysis is to develop and evaluate machine learning models for predicting credit risk, specifically distinguishing between healthy loans (Class 0) and high-risk loans (Class 1). We aim to assess the performance of two different models: one using the original dataset and the other using resampled data. This analysis will help us determine which model is better suited for the task of loan classification.

Results

Logistic Regression Model with Original Data

- **Accuracy Score**: 0.9918
- **Precision (Class 0)**: 1.00
- **Recall (Class 0)**: 0.99
- **F1-Score (Class 0)**: 1.00
- **Precision (Class 1)**: 0.85
- **Recall (Class 1)**: 0.91
- **F1-Score (Class 1)**: 0.88

Logistic Regression Model with Resampled Data

- **Accuracy Score**: 0.9938
- **Precision (Class 0)**: 1.00
- **Recall (Class 0)**: 0.99
- **F1-Score (Class 0)**: 1.00
- **Precision (Class 1)**: 0.84
- **Recall (Class 1)**: 0.99
- **F1-Score (Class 1)**: 0.91

Summary

Both machine learning models, one with the original data and the other with resampled data, performed exceptionally well in classifying loans. However, there are key differences to consider:

- The model trained on the original data achieved an accuracy score of 99.18% and had high precision and recall for both classes. It had an F1-score of 1.00 for healthy loans (Class 0) and 0.88 for high-risk loans (Class 1).
- The model trained on the resampled data achieved an even higher accuracy score of 99.38%. It also exhibited high precision and recall for both classes, with an F1-score of 1.00 for healthy loans (Class 0) and 0.91 for high-risk loans (Class 1).

Recommendation: Based on the evaluation results, we recommend using the logistic regression model trained on the resampled data for predicting loan status. While both models performed well, the resampled model slightly outperforms the original data model, especially in terms of F1-score for high-risk loans. This suggests that the resampled model is better at identifying high-risk loans, which is a critical task for a lending institution. It achieves a balance between precision and recall, making it a robust choice for loan classification.

In Summary, the resample logistic regression model is recommended for high accuracy and balanced performance in predicting credit risk analysis