**## Overview of the Analysis**

The purpose of this analysis was to create and evaluate a model used for credit risk loan assessments. The goal was to determine whether the loans are classified as healthy (0) or high-risk (1). The lending data consisted of loan amount, interest rate, borrower income, debt to income, number of accounts, derogatory marks, total debt and loan status.

The stages of the machine learning process that were used for the analysis:

Data preparation: the dataset was imported and established the DataFrame.   
Feature and Label Separation: independent variables and label for dependent variable – loan\_status.   
Train Test Split: divided the dataset into training and testing sets.  
Model Importation, Installation, and Fitting: Logistic Regression model using the training data.

Prediction: fitted model to make predictions with the testing dataset.

Evaluation: evaluated the model’s predictions with the accuracy score, confusion matrix and classification report.

The method used in the analysis with the Logistic Regression model.

**## Results**

Logistic Regression Model:

precision recall f1-score support

0 1.00 0.99 1.00 18765

1 0.84 0.94 0.89 619

accuracy 0.99 19384

macro avg 0.92 0.97 0.94 19384

weighted avg 0.99 0.99 0.99 19384

**## Summary**

The model shows accuracy at 99.25%. The regression model shows near perfect precision in predicting healthy loans. For high-risk loans, the precision is slightly lower, but the recall is still high. The regression model show that it is still effective with high-risk loans.

Performance does depend on business needs. For Class 1 (high-risk loans), high recall is important to reduce defaults. I recommend the Logistic Regression model for deployment. Overall, the model is effective for identifying high-risk loans and minimizing defaults.