# Module 12 Report Template

## Overview of the Analysis

In this section, describe the analysis you completed for the machine learning models used in this Challenge. This might include:

\* Explain the purpose of the analysis.

ANS: This analysis is to determine the accuracy of the model that will help determine how worthy potential borrowers are from a peer to peer lending services company.

\* Explain what financial information the data was on, and what you needed to predict.

ANS: The financial information is data from the lending activity from a peer to peer lending services company. I need to predict build a regression model that will predict the credit worthiness of borrowers. To help determine which of them are high-risk or low-risk.

\* Provide basic information about the variables you were trying to predict (e.g., `value\_counts`).

ANS: Precision: This is the ratio of true predictions to the total number of positive predictions.

Recall: This is the sensitivity or true positive rate

F1-Score: This is the harmonic mean of the precision and recall, providing a single metric that balances them both.

Support: This is the total number of actual occurrences of each class in the dataset

Accuracy: this measures the overall correctness of the model’s predictions.

Macro Average: This is the mean of precision and it treats all classes equally and independently.

Weighted Average: This is the mean of precision, recall and F1-Score. It accounts for each class imbalance by giving more importance to classes with more instances

\* Describe the stages of the machine learning process you went through as part of this analysis.

ANS:

1. Read the data I was given and split them into Training and testing data sets.

2. Create labels X and Y.

3. Split the data into training and testing datasets by using the train\_test\_split method.

4. Fit the logistic regression model by using the training data with a random state of 1 and then making a prediction using the testing data.

5. Generating a confusion matrix to measure the performance of the model.

\* Briefly touch on any methods you used (e.g., `LogisticRegression`, or any other algorithms).

ANS: The Logistic Regression Model is uses to estimate the probability of an event occurring. I used the Logistic Regression Model to predict the healthy loans in the data set with perfect or near-perfect scores in precision, recall and F1-Scores

## Results

Using bulleted lists, describe the accuracy scores and the precision and recall scores of all machine learning models.

\* Machine Learning Model 1:

\* Description of Model 1 Accuracy, Precision, and Recall scores.

Model Performance for class 0 (Healthy Loan)

• Precision 1.00: The model is extremely precise in predicting healthy loans (class 0). Almost all predicted healthy loans are indeed healthy.

• Recall 0.99: The model is also highly effective in identifying actual healthy loans. It correctly identifies 99% of healthy loans.

• F1-Score 1.00 : The F1-score, which is the harmonic mean of precision and recall, is perfect, indicating a balance between precision and recall for healthy loans.

• Support 18765: The number of actual healthy loans in the dataset is 18765.

Model Performance for Class 1 (High-Risk Loan)

• Precision 0.85: The model's precision for predicting high-risk loans (class 1) is 0.85. This means that 85% of loans predicted to be high-risk are actually high-risk.

• Recall 0.91: The model successfully identifies 91% of the actual high-risk loans.

• F1-Score 0.88: The F1-score for high-risk loans is 0.88, indicating a good balance between precision and recall, though not as perfect as for class 0.

• Support 619: The number of actual high-risk loans in the dataset is 619.

## Summary

Summarize the results of the machine learning models, and include a recommendation on the model to use, if any. For example:

\* Which one seems to perform best? How do you know it performs best?

ANS: The healthy loans (Class 0) perform best with high precision, recall and an F1-Score all near 1.00.

\* Does performance depend on the problem we are trying to solve? (For example, is it more important to predict the `1`'s, or predict the `0`'s? )

ANS: Yes. The performance depends on the problem we are trying to solve. When it is crucial to predict high-risk loans, then 1 must be predicted to ensure that high risk loans are identified. If predicting healthy loan is equally crucial, then there must be a balance between the precision and the recall for class 0.

If you do not recommend any of the models, please justify your reasoning.

Recommendations: Based on the metrics above, the logistics regression model performs well with high accuracy, precision and recall for both classes