# Value Analysis of Prospective Loan Clients

Prepared for:

Client XX, Company YY

By:

Justin Lear

UVA Department of Statistics

Advised by: Professor XX

MM/DD/YYY

## Executive Summary

Client XX of Company YY presented an issue that his company is facing in determining which potential loan clients are worth pursuing. He provided data on a one-month snapshot of the company's loan portfolio with various financial variables given for each customer. He is interested in determining a method to differentiate "good" vs. "bad" loan clients in an effort to expand the company's client base. To distinguish clients that are worth pursuing, it is recommended that expected total loan return value be calculated for each customer. Through probability theory and lasso regression, an estimate of the loan's total revenue can be calculated, while accounting for loan amount, repayment probability, and risk, allowing for objective comparisons across clients. With this metric, Company YY can quantitatively determine what separates a *good* loan from a *bad* loan.

## Introduction

### Objectives

The main objective of this report is to create a model that will determine a *good* versus *bad* loan, while taking riskiness, profitability, and social responsibility into account. The secondary objective is to provide a value proposal containing information on the type of client to focus on.

## Approach

For the scope of this problem, it was decided that the company's optimal avenue of pursuit is to maximize returns, while minimizing risk. To do this, client value is weighed according to their expected total return on a loan. This metric is able to quantify how much revenue the client is expected to be worth, allowing Company YY to make objective decisions on whether a loan client is worth pursuing. The formula for this value is summarized below.

$$Expected\ Total\ Return = Total\ Payment + Payment\ Probability\ \times [(Term \times Installment) - Total\ Payment]$$

This method takes into account the amount that the client has already paid, the total potential value of the loan, and the probability that the client fully pays it off. The 'Total Payment', 'Term' and 'Installment' values are all provided in the dataset, however 'Payment Probability' still needs to be calculated. Prior to this being done, a risk variable was added to the dataset to be regressed upon. It measures the inherent risk of a client, given their proportion of principal paid compared to the funded amount of the loan with respect to how long they have been paying the loan. This value is measured as the residual distance of a regression equation, where large positive values indicate less risky customers, as they tend to pay more than the minimum payment on a monthly basis, and negative numbers indicate more risky clients who typically pay less than the minimum monthly payment. The plots of these regression equations are displayed in Figures 1 – 3 in the *Appendix*. To calculate the probability of a client repaying their loan, we take a subset of the data where clients either fully paid off the loan or were charged off/defaulted. For accurate inferences later on, the subset needs to be representative of the entire dataset; plots comparing the two groups are displayed in Table 1 and Figures 4 and 5 in the *Appendix*. Lasso regression was performed on the subset data to determine a classification model that returns the probability that a client fully pays off their loan given certain variables in the dataset. This model can then be applied to the entire dataset, using the same predictor variables as in the subset, leading to a new variable that indicates the probability of a client fully repaying their loan.

According to standard probability theory, the expected value of an event is calculated by multiplying the probability that an event happens by the outcome itself. This means that when accounting for how much of a loan has already been paid back, we can calculate how much of the remaining loan is expected to be paid back using the previously mentioned probability. If the formula above is followed, a value representing the expected total return is calculated for each client, providing the company with an indicative variable that they can base their decisions upon.

## Results

Upon conducting the analysis of the provided dataset and creating a model, it was determined that the clients that have the highest expected total loan value have a higher probability of fully repaying their loans, pay more than the minimum payment every month, have been paying back the current loan for several months, and have relatively large funded loan amounts. As mentioned in the previous section, lasso regression was used to calculate the probability that a client fully repays their loan; one advantage of lasso regression is that it emphasizes the most important predictor variables in a model, while reducing the insignificant predictors to have no impact, leading to a list of variables with the most predictive power. The most important variables in predicting the probability of repayment are: the outstanding principle to be paid back, how long the client has been paying the loan at the current month, and the risk variable described in the *Approach* section.

For the given dataset, the expected total loan values expressed values between $34.14 to $59,805.00. The code used for this analysis is provided in the *Code Appendix* section so the results can be reproduced and verified. The expected total loan value can be used to provide a financial assessment of whether a client is a viable customer and how much revenue they should generate for the company. Ultimately, this tool can be used to answer the question of interest and provide a quantitative answer to classify a loan as *good* or *bad*.

## Conclusions and Recommendations

For this report, the concept of a *good* loan is analyzed strictly from a financial perspective, with the goal of maximizing returns for the Company YY, while minimizing risk. To provide an assessment of what a *good* loan is, an expected total loan return value can be calculated for each client that uses probability theory to estimate the total loan value over the course of the loan's lifetime given the current snapshot of client information. From here, a threshold of expected total loan value can be used to differentiate *good* loan clients versus *bad* loan clients. According to the model, *good* loan clients are more likely to pay back the full amount, have a low risk value, have a long payment history on the current loan, and have relatively large loan funded amounts. In the future, the model can be improved upon by incorporating personal information such as age, gender, and education level. The expected total loan value can quantify how much a client is worth to the company and allows for quantitative oriented decision making in regards to expanding their client base.

# Appendix

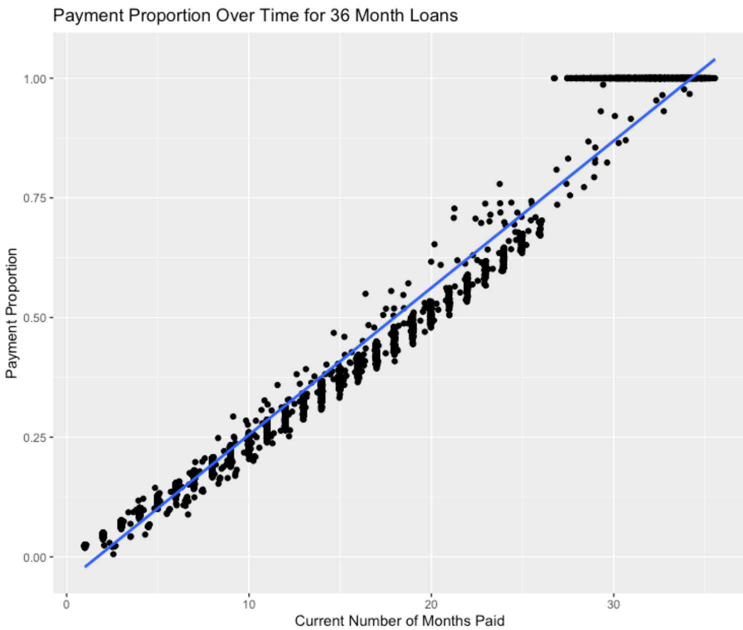


**Figure 1: 36 Month Risk Regression Plot**

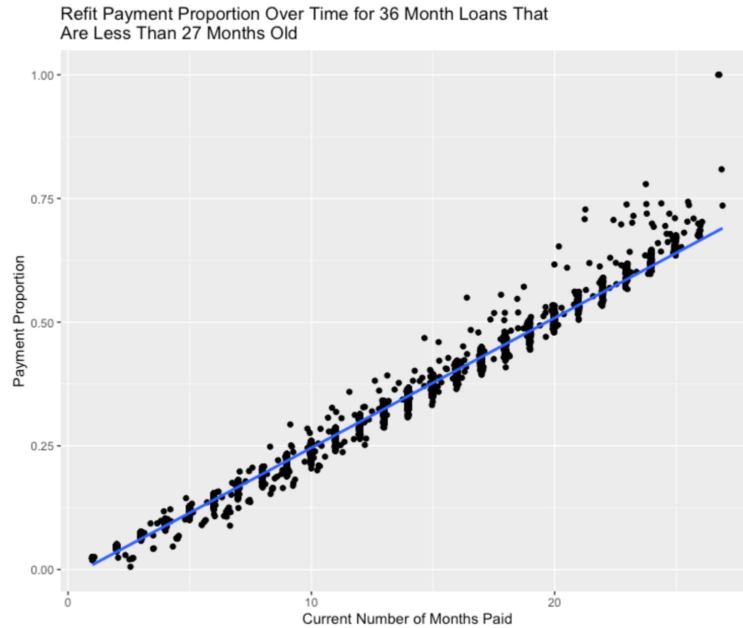


**Figure 2: 36 Month Risk Regression Plot for Less Than 27 Months**

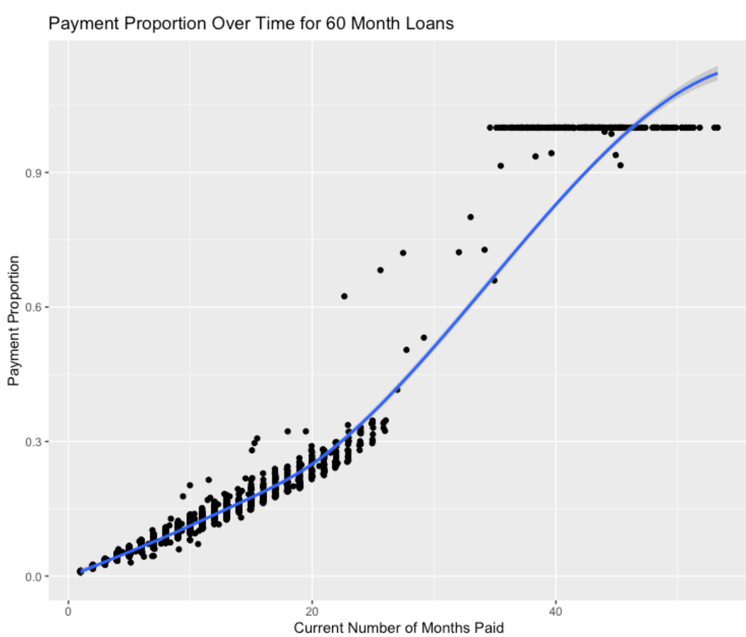


**Figure 3: 60 Month Risk Regression Plot**

| | Loan Purpose | | |
|---|---|---|---|
| | Credit Card | Debt Consolidation | Other |
| **Full Dataset** | 23.24% | 59.48% | 17.28% |
| **Subset** | 18.06% | 59.75% | 22.19% |

**Table 1: Proportion of Loan Purposes in Full Dataset vs Subgroup**

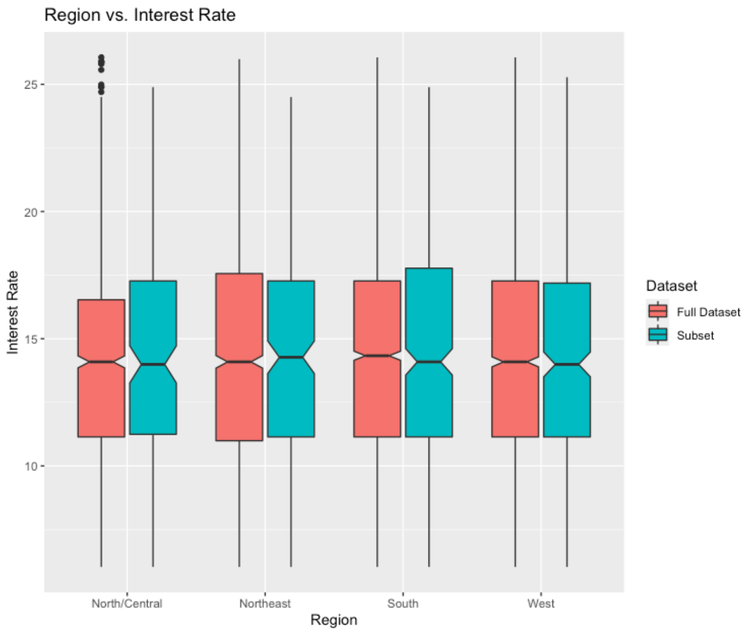


**Figure 4: Comparing Interest Rates Across Regions in Full Dataset and Subset**

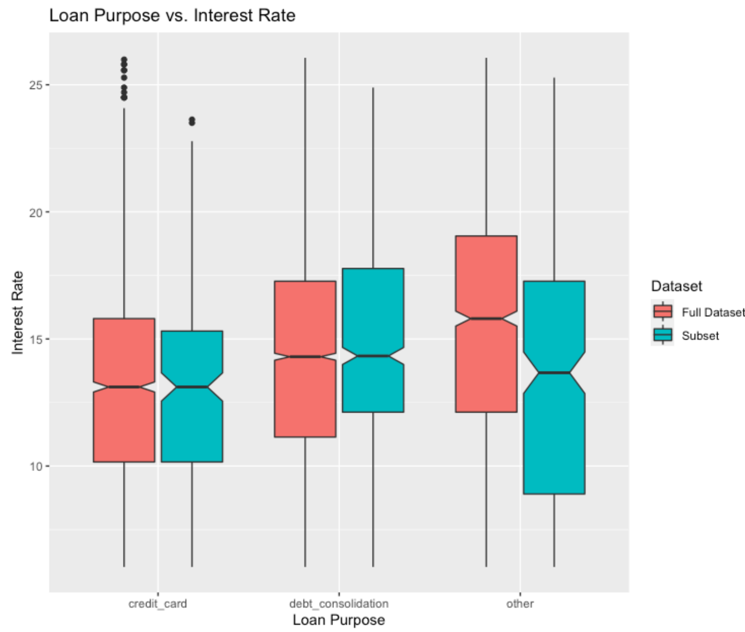


**Figure 5: Comparing Interest Rates Across Purposes in Full Dataset and Subset**