Executive Summary

In the loans data, the unit of observation is each loan. Each record in the dataset represents a loan

The response variable is PRSM.

| Name | Description |
| --- | --- |
| *PRSM* | PRSM score of the loan, which is used to identifying potentially risk loans.Based on our correlation analysis, we select out 5 predictors described below. |

Predictors

| Name | Description |
| --- | --- |
| *FICO* | Fair Isaac Credit Score range from 300 to 850. |
| *TotalAmtOwed* | Fair Isaac Credit Score range from 300 to 850. |
| *WomanOwned* | An indicator of whether the business is owned by a woman |
| *CorpStructuret* | which records whether business is structured as a soleproprietorship, corporation, limited liability corporation (LLC), or a partnership |
| *NAICS* | The North American Industry Classification System provides a 5- or 6-digit code that classifies different industries. |

To introduce a base line, we will select the average of numeric variables and most frequent of categorical variables as our typical instance.

| FICO | TotalAmtOwed | WomanOwned | CorpStructure | NAICS444240 |
| --- | --- | --- | --- | --- |
| 697 | 232420 | 0 | Sole | False |

We create a MLM model to predict the PRSM, and the predicted PRSM on this typical borrower is

fit lwr upr  
0.5651608 0.3628741 0.7674474

In our model, the estimate of CorpStuctureSole is negative, which means a sole CorpStructure will decrease the PRSM.

The Adjusted R-square is 0.8079, which means the model could explain 80.79% of the variance is explained by the predictor variables in our model.