SBA Loans

By: Morgan, Deautaun, Eduardo

What is an SBA Loan?

- U.S. Small Business Administration (SBA) Loan

- A loan that business owners get as means to start their business before the opportunity making a profit.

- When you fail to repay a loan in time, you have defaulted.

Questions:

1. What factors affect how much the loan is for?

2. What factors tend to cause a loan to be defaulted on?

3. Build a model to predict which approved loan applications will default on their loan.
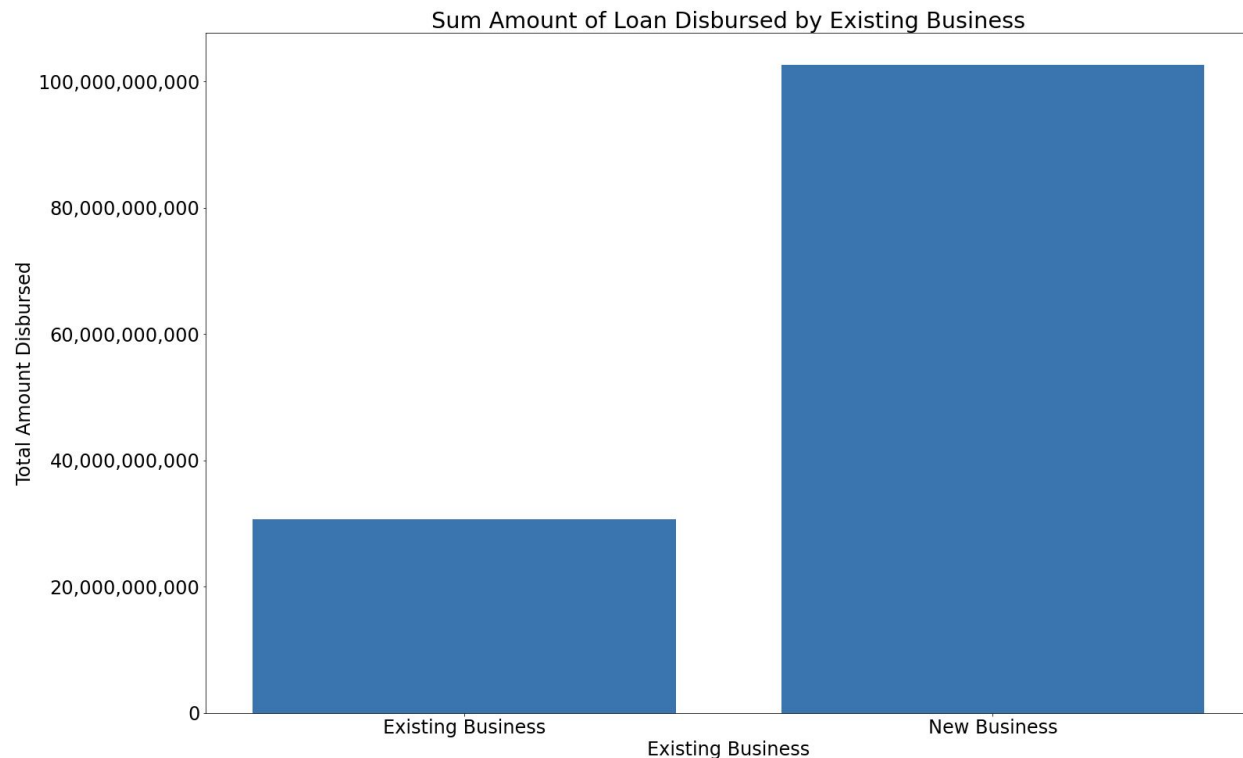
What factors affect how much the loan is for?

- Industry

- Existing Business

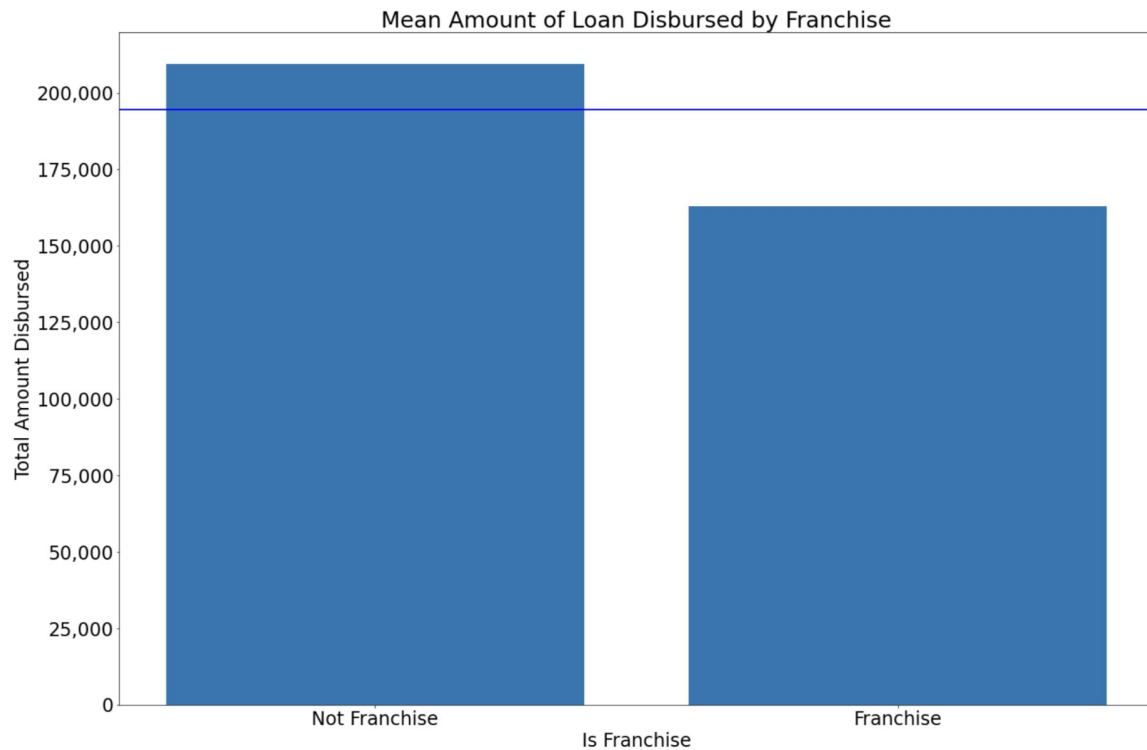- Franchise

Industries and their Average Loan Disbursed

Industry Average Loan Amount Disbursed

Sum of Loan Disbursed by Existing Business
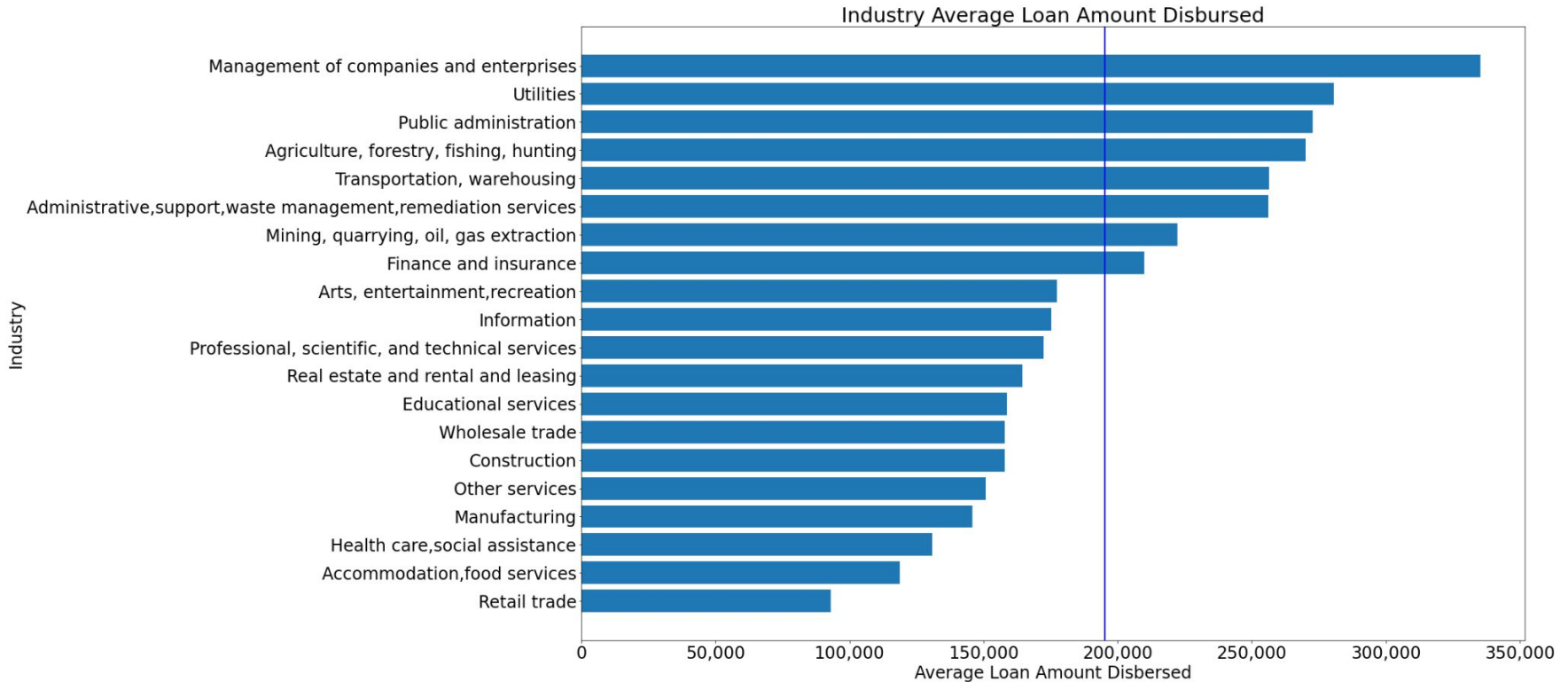
Sum Amount of Loan Disbursed by Existing Business

New Business is only 30% of our data yet have higher Loan disbursement.

Average Loan Disbursed by Franchise

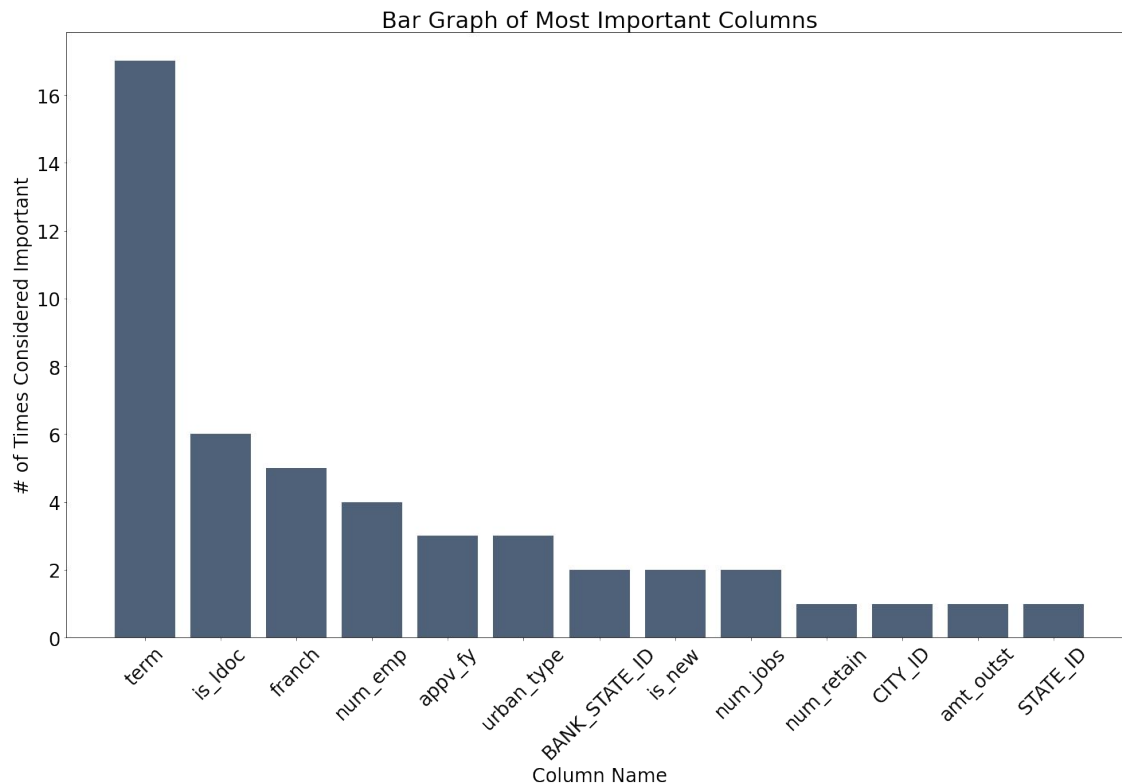Mean Amount of Loan Disbursed by Franchise

Businesses that are not a franchise make more than the average loan disbursed.

What factors tend to cause a loan to be defaulted on?

Determining Important Factors

Term appeared in every trial telling us it is the most important factor in producing a high accuracy model.

Bar Graph of Most Important Columns

Deeper Dive: Term

Borrowers who didn't default have the highest frequency

People who don't default usually have loan terms of 60, 80, 120, 180, 240 and 300 months.

Term Histogram for Default / No Default by Correctness
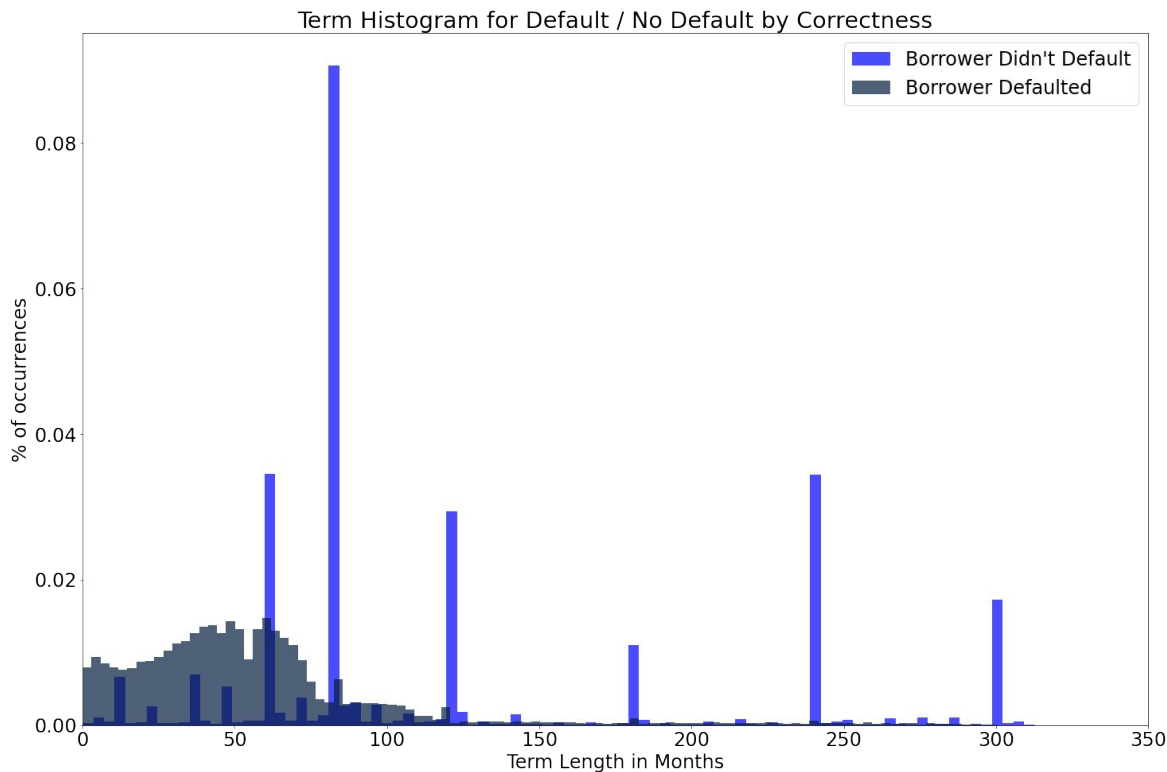
What if we only looked at the year of approval?

Fiscal year of approval seemed to be the second best "predictor"

2008 recession

Histogram of appv_fy by Default Status
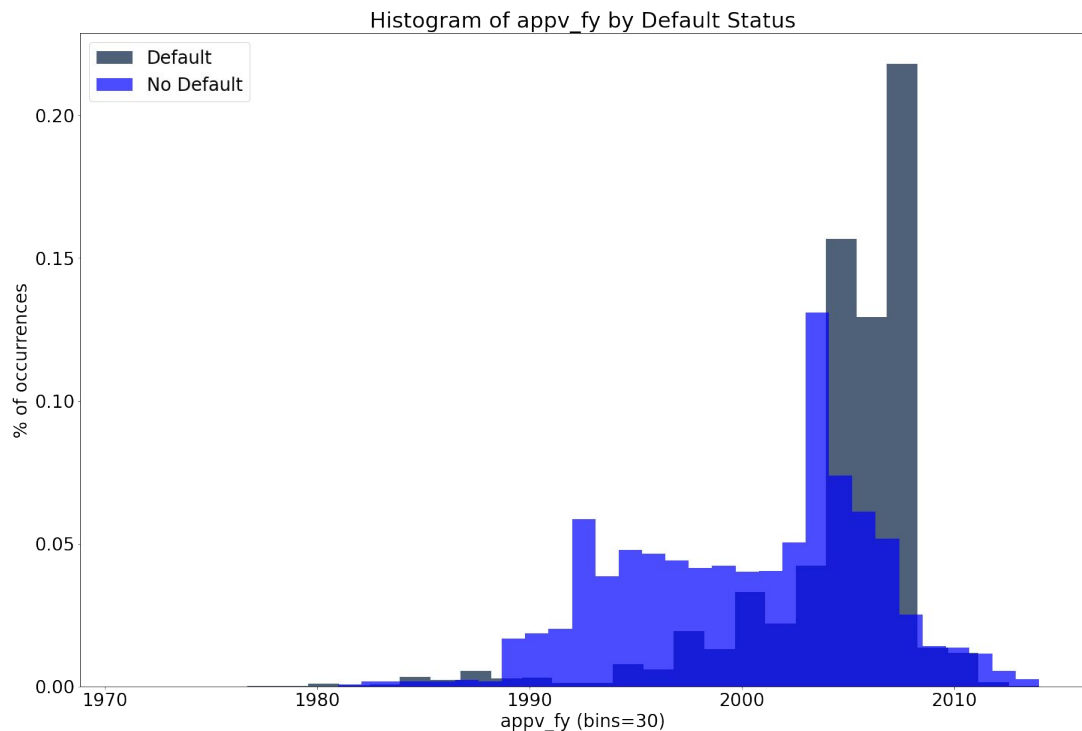
Model to predict which approved loan applications will default

Prediction Model Summary

- Used scikit-learn logistic regression to predict if a borrower defaulted or not.

- Only 18.16% of all borrowers defaulted, therefore data was balanced by sampling from non-defaulters.

- Using 2 out of 27 columns: loan term, and fiscal year of approval.

Prediction Model

- Precision: Out of the loans we predicted to have defaulted, 77% of them were actually defaulted.
- Recall: From all the defaulted loans, 83% of them were predicted correctly.
- Similarly, for the "False" defaulted, see that out of the loans that *actually did not* default, our model predicted 75% of them correctly.
- Accuracy: Out of the loans that *actually did* default, we predicted 79% of them would default.

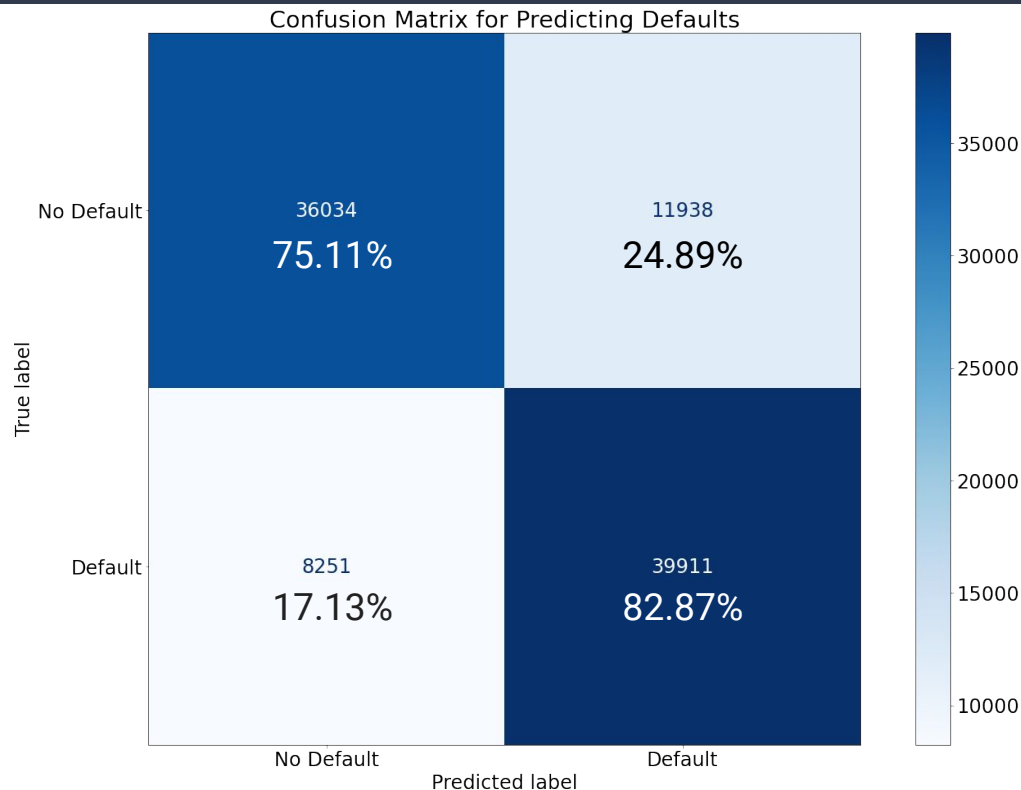| default	precision	recall
False	0.81	0.75
True	0.77	0.83
accuracy		0.79

Confusion Matrix

This model is better at predicting defaults than non-defaults

Confusion Matrix for Predicting Defaults

There is a better model.

Term Histogram for Default / No Default by Correctness
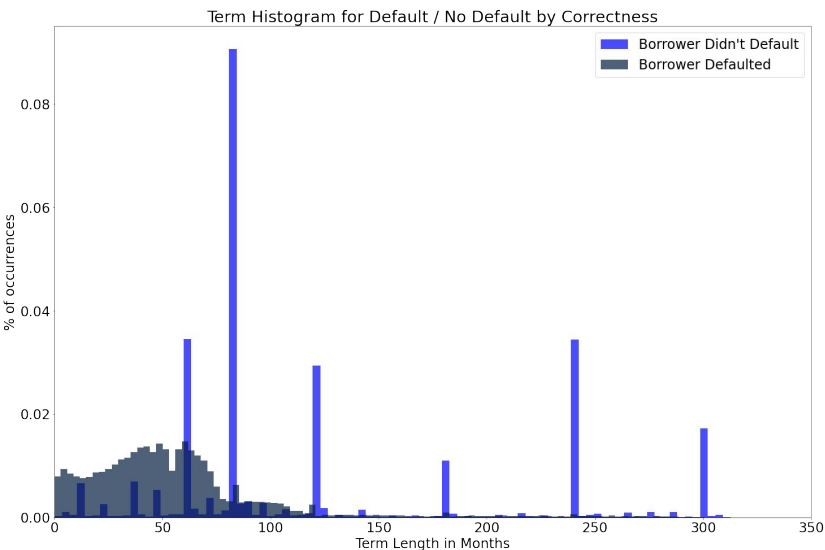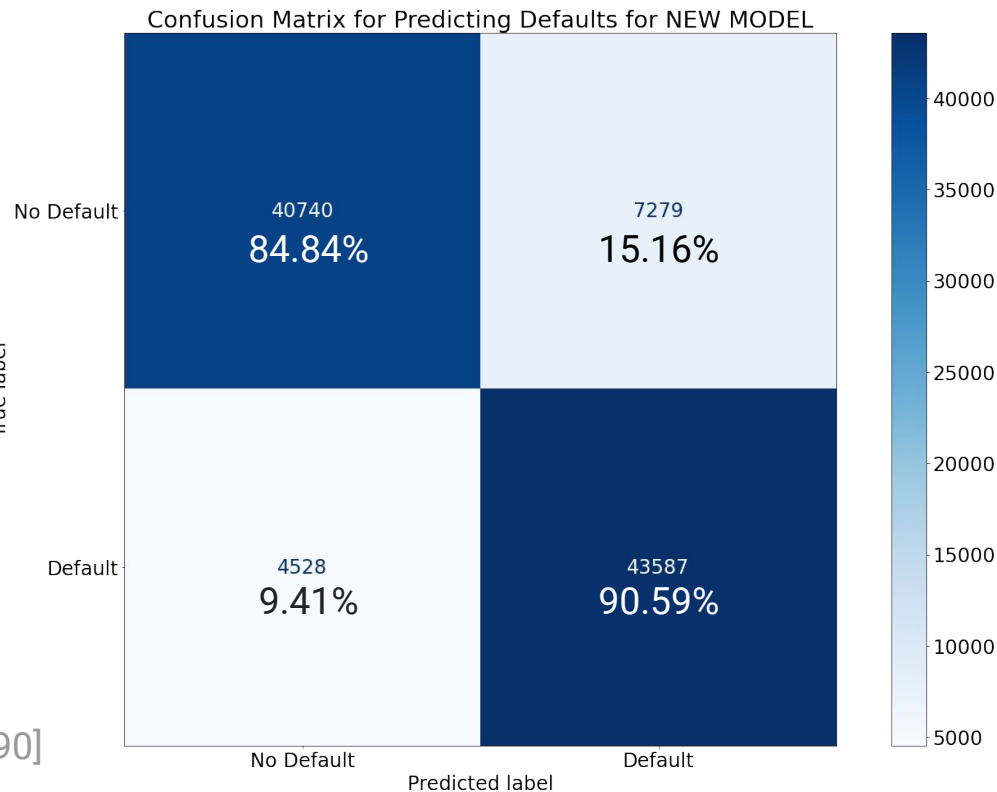
Previous model: 79% accuracy avg.
New model: 88% accuracy avg.

[84, 60, 240, 120, 300, 180, 36, 12, 48, 72, 24, 96, 90]

Confusion Matrix for Predicting Defaults for NEW MODEL

Findings

- Term and fiscal year of approval are the two factors that gave a high model accuracy.
- We can predict if a loan will default with 88% accuracy using a logistic regression model.
- Non-Franchise Businesses make higher than the average loan disbursement.
- New Businesses get higher loans.
- Management and Utilities Industries on average get higher loans.