Credit One
# Customer Default Identification Report

Samantha Goodman

Problem:

An increase in customer default rates is bad for Credit One since its business is approving customers for loans in the first place. This is likely to result in the loss of Credit One's business customers.

Questions to Investigate:

1.   How do you ensure that customers can/will pay their loans?
2.   Can we approve customers with high certainty?

Information found during EDA

●   Customers who were granted lower credit limits account for most of the defaults. Men who were granted less than 80,000NT in credit defaulted at a rate of almost 34%, compared to the average default rate of 22%. For this reason, raising the lowest credit offered to over 90,000NT might eliminate many of the problem customers.
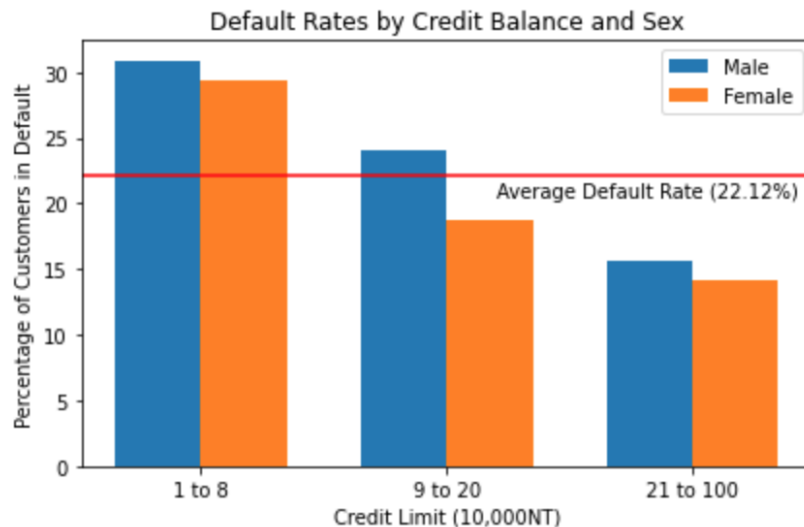
●   Since so many people with lower credit limits are defaulting, it seems like there is a problem with the decision making process whether or not to grant them a credit line. This is where I propose an alternative algorithm.

- Once a customer is two months behind in bill payments, they are 50% likely to default. Every month they are behind contributes to the chance that they will default on the loan. Information on past payments is vital to making a decision on whether to approve a customer or not.

Default Rates by Months Behind In Bill Payment
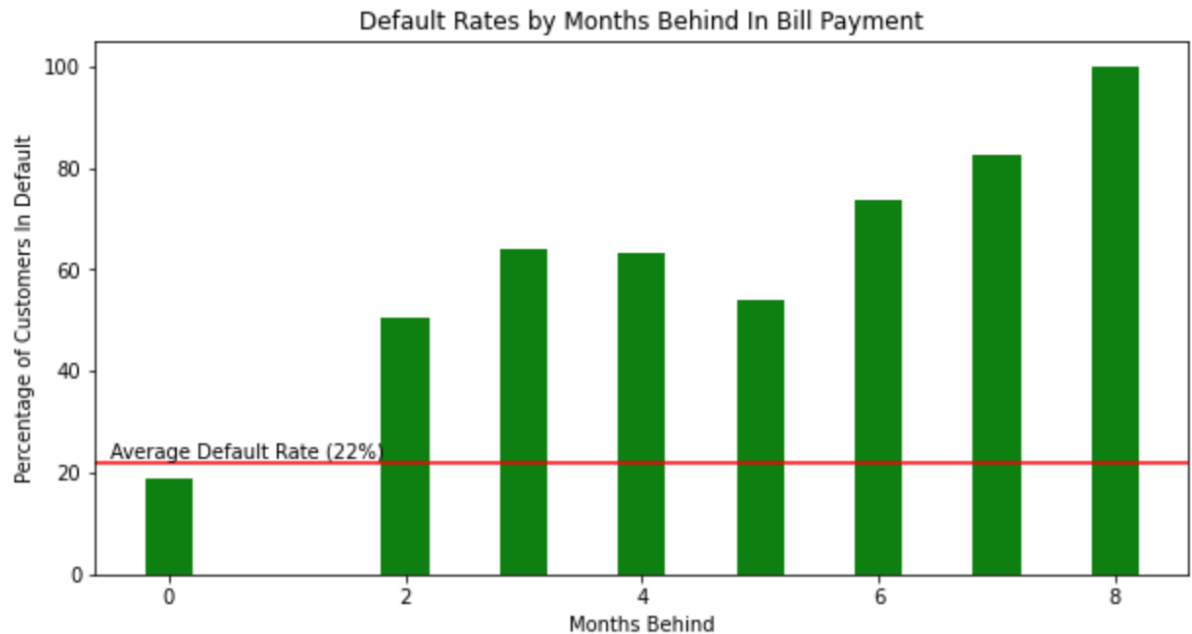
Question 1: How do you ensure that customers can/will pay their loans?

There is no way to ensure that a customer will pay their loans, other than making the right decisions at the beginning - whether to approve or not, and how high of a credit limit to grant them. For this reason I propose we examine the way in which customers are approved.

Question 2: Can we approve customers with high certainty?

Currently 22% of customers who are approved end up defaulting on their loans. My goal was to examine different machine learning algorithms, and given the data that we have on the current customers, can build a model to more accurately approve or deny potential customers credit.

First algorithm category examined: Linear Regression models

I tested several linear regression models to predict both default rates and credit limit, but without success. The best result I was able to obtain was 47% accurate at predicting credit limit, which I am not sure is particularly useful given how many customers default with the amount of credit they are currently given.

```
Random Forest Regressor 0.4664145434805736
Linear Regression 0.3581989426608166
Support Vector Regression -0.05037380099684693
```

Second algorithm category examined: Classification models

This approach yielded much better results, as these algorithms were able to take demographic data, credit limits, and payment history to accurately predict whether a customer would default on the loan.

```
Random Forest Classifier 0.8155333333333333
Decision Tree Classifier 0.7246333333333332
Gradient Boosting Classifier 0.8202333333333334
```

The best performing algorithm was the Gradient Boosting Classifier, which was able to predict customer default status with 82% accuracy.

Conclusion

From my analysis, I would recommend changing the way Credit One evaluates potential customers, and use a classification model such as Gradient Boosting to more accurately assess whether they should be given a loan or not. This algorithm was 82% accurate at predicting customer default given the data set currently available. Considering that 22% of approved customers currently end in default (78% accurate) this is an improvement.

The only caveat is that past payment data was used to build this model and thus that information (or something like it) would have to be substituted for potential customers for whom payment histories are unknown.