**Credit One Credit Scoring Service**

Credit One has run into a business problem with increase in the number of customers defaulting on loans.

Increase in customer default rates - This is bad for Credit One since they approve the customers for loans in the first place. If clients lose Revenue and customers it is eventually a loss of clients for Credit One

Overview of the Data: We have around 30300 records with transaction data for the past 6 months. We have 25 features which includes the biographical features like Age and Gender as well as the payment amounts and balances for the past 6 months.

Ability to predict different attributes using data modeling:

I did an initial cleanup of the dataset to eliminate some duplicated header rows and null values.

The initial analysis of data using plots and correlation matrix showed that the

* Female population tends to default less than the male population.
* Also the population of age group under 30 tends to default more.
* Anyone with a limit balance less than 100k tends to default more.
* Graduated population defaults less than the others.

We ran the data through various regression models to predict the Limit Balance which seems to be a major determinant for defaults. But the regression models didnt provide much accurate information.

The Classification models we used seem to predict the results much more accurately with Gradient Boost model being the best. The results from this model matches our initial findings based on the plots we used.

Let me know if you have any questions.