Credit One Report

Questions to investigate:

1. How do you ensure that customers can/will pay their loans?

* Firstly, it is not possible to completely ensure that a customer will pay their loans. There will always be unforeseen factors that could affect a customer's ability to pay a loan.
* Taking this into account, the best way to ensure a customer's ability to pay their loans is to approve them for the correct credit rate.
* This means that finding a balance that is not too low and not too high, that way each month the customers will be within their spending budget and have the ability to pay the loan off.

1. Can we approve customers with high certainty?

* We cannot approve customers with a high level of certainty, but we can use a multitude of variables to approve customers with some certainty.
* During our analysis of the credit one data, a random forest regressor was run as the model to devise a method to find appropriate credit limits to give to customers. As well as devising a model to consider several variables that could contribute to a customers probability to default.
* Using these models we can predict with some certainty if a customer will default and the credit level they should be given, but it is not certain.

Statistically Significant Data:

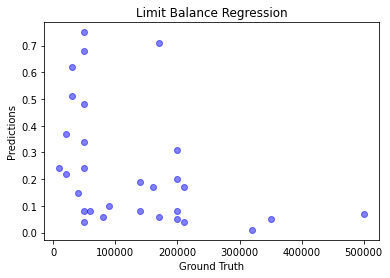
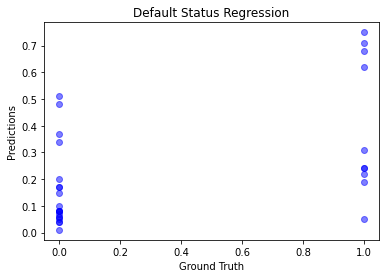
* The most important predictor of a customer's probability to default is the credit limit they are given. In order to estimate the ideal credit limit to grant to customers, the following variables were useful as they had a statistically significant correlation with the dependent variable.
* Age, monthly payment history from august (PAY\_2), amount of bill statement in May (BILL\_AMT5), amount of previous payment in april (PAY\_AMT6), education status, and default status all had a statistically significant correlation with the credit limit balance.
* Marriage status, limit balance, and payment history from September (PAY\_0) were statistically significant when correlated with default status.
* The variables stated above would have the greatest influence and should be given the most consideration when trying to determine the credit limit a customer should be given, or if they will default.

Concrete information:

* Concrete information that can be derived from the data depicts that customers who were married, had higher education status, paid bills on time and in full amounts should be given a higher credit limit as they were less likely to default.
* A customer's payment history was always a significant predictor of the limit balance they should be given as well as the likelihood of default. However, bill amount and past payment amounts were significant predictors of the credit limit a customer should be granted, but not a significant predictor of a customer's likelihood to default. This means that there are other factors outside of the data that could be influencing a customer’s probability to default.
* Limit balance is the single greatest predictor of a customer’s probability to default.
* Customers who hold undergraduate and graduate degrees defaulted more than those who didn't. Also these two classes had higher limit balances.
* Married and single customers defaulted at relatively the same rate.

Methods to uncover information:

* Correlation and collinear value tables provide the best estimate of what will affect a customer's probability to default and a given credit limit as they provide data on which variables had the greatest effect on the dependent variable.
* Linear regression allows us to analyse what a slight increase in each variable does to the dependent variables (credit limit balance and default status). Using this model we can have a multitude of different inputs that will allow us to analyse with slight certainty the credit limit a customer should be given and their likelihood to default.
* We also were given an r2 and root mean square error that allowed us to determine how good the model fits. When the dependent variable was limit balance the r2 was 61.4% and the root mean square error was 69788.99. This means that the independent variable explains 61.4% of the variance in the dependent variable. In other words the model is 61.4% accurate in explaining the difference in credit limit balances when considering the dependent variables.
* When the dependent variable was default status, the r2 was 23.4% and the root mean square error was 41.3%. So this model was less certain when trying to determine variables that affect a customer's probability to default.
* Scatterplot outputs allow us to visualise the above findings.



Recommendations:

* Consider a wide variety of variables such as marriage status, education status, previous default history and credit limit balances, and previous payment history. This data will aid you best in determining the proper limit to allow.
* Outside influences such as other loans, job status, and family status can play a role in a customer's probability to default. This isn't the only thing, macroeconomic factors such as interest rates, unemployment rates, and taxes can also play a large factor.
* When granting a customer a certain credit limit. Credit One should give them a lower credit limit than it thinks and let the customer develop a payment history at a lower credit line as this is the biggest predictor for both credit limit allowance and default status. Once Credit One deems them a trustworthy customer that pays on time, then grant them a credit limit increase. This is a potential way to protect Credit One while bringing in additional customers.