Lessons Learned

There were several lessons learned while completing the exploratory data analysis for the Credit One data that could have potential business value:

* The limit balance for the data goes up to a potential $1,000,000, but most of the limit balances for the customers are under $300,000. So, when looking at the probability to default, a lower limit balance may correspond to a higher probability to default. Customers with higher limit balances tend to pay off their debt.
* Limit balance and education/marriage have a correlation that is statistically significant. Education and marriage could be used as a good predictor of a customer’s probability to default.
* The length of time a customer takes to repay a bill is the strongest predictor of a customer’s likeliness to default on a loan. This has the highest correlation with default status.
* Customers who attended a university had the highest rate of defaults.
* Married and single customers were almost equally as likely to default on a loan.
* Customers who waited longer to pay a bill typically paid less of the bill.
* Customers who paid on time or before the required payment date had a higher limit balance.

Main Lessons I learned from this experience:

* What we come up with in our heads of a typical customer who would default on a loan is not always accurate of the actual people who have defaulted on loans.
* Comparing 2 variables to see their relation to one another is a beneficial way to perform exploratory data analysis and will lead to findings that could have business value.
* Seaborn is the best package to use to explore and illustrate data.
* Having miniature hypothesis when going into an EDA gives you a plan to work with and a method of exploring large amounts of data. This allows you to make insights into the data that may appear difficult or hidden if you perform random analysis on the data.
* Summarizing the data using pandas profiling gives a quick and accurate summary of main data points and trends in the dataset.

Recommendations:

* If you want to try and predict the likelihood of a customer defaulting, it is best to look at their past payment history, timeframe in which they paid it, and amount of the bill paid. This is the biggest indicator and has the most relation with default status.
* Education and marriage are good factors to use to evaluate a customer’s credit worthiness, but it is not an end all be all. Other variables are more important.
* The limit balance covers a wide range, so by itself it is not a good indicator of the probability of a customer to default. But when used in combination with other variables to analyze a customer’s probability of default it is useful.