Nick Helmers

C2T2 Lessons Learned

* Education was one of the most effective variables in terms of predicting whether a customer would default. The largest number of customers that did not default on their loan either had a university or graduate school degree.
* There were far more female customers than male customers
* The majority of customers in the dataset were between the ages of 21-35.
* A large group of customers that did not default on their loan was in the youngest age bracket between 21-35.
* With such a large number of variables, it was difficult to narrow down the variables that would help predict which factors would most accurately predict whether a customer would default.
* One takeaway from this lesson is how useful the Pandas Profiling Report can be to summarize findings from the dataset. I wasn’t aware a tool like this existed and it is always interesting and exciting when new tools are introduced that make the data science process easier.
* My conclusion from this task is that some of the visualizations generated were useful in helping to understand the data, but I don’t think we will be able to discern too much from this step until we get to the predictive analytics portion. Hopefully then we will be able to better zero in on which specific factors lead to customers defaulting so that Credit One can improve their business in the future.