Tom Snyder

Credit One

May 28, 2021

Lessons Learned Report

Exploratory Data Analysis on Customer Default Payments Data

Exploring this data began with cleaning, interpreting the label encoding, and converting the data elements to manipulatable formats (strings to numeric). Once this was completed, I analyzed the data using conventional EDA techniques including: finding correlation and covariance, univariate exploration, bivariate and multivariate exploration, and visualization.

One of the lessons I learned is that cleaning data is an iterative process, depending on how the data will be analyzed. I ran into issues applying data analysis techniques because the format or labels of the data weren’t what I expected. This led me to add more cleansing to the data as I went so that I could solve the problems I was working on.

Another important lesson I learned is that making my exploration process easy for someone else to follow and understand is important. As I explored the data and tested many different aspects, I grouped similar exploration tasks together and eliminated code that ultimately went unused. This process of editing my notebook made me look at the big picture of how I was applying statistical and data analysis concepts to my exploration process.

## Recommendations

My recommendations are to apply machine learning to this data set to see if usable models can be built that will accurately predict a customer’s default risk. The factors included in the data set show some visible correlation, but it is worthwhile to use modeling to discover correlations that may not be visible at first glance.