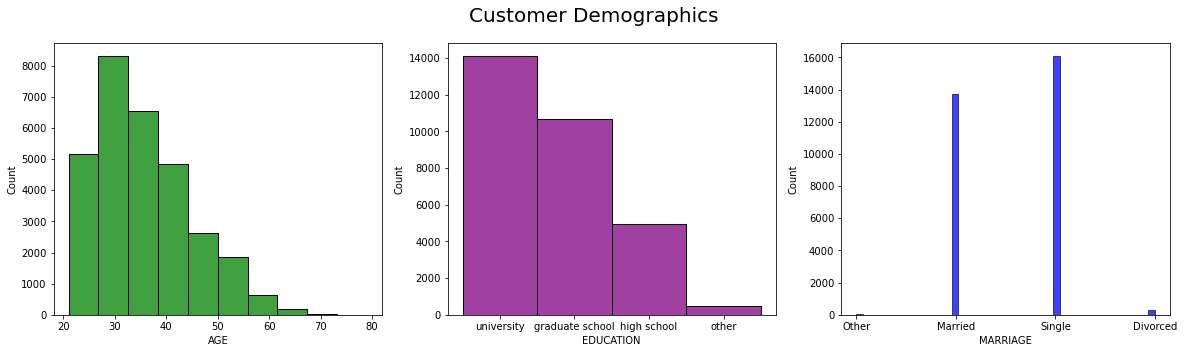
Tom Snyder

UT Data Analytics

June 16, 2021

Building the Models

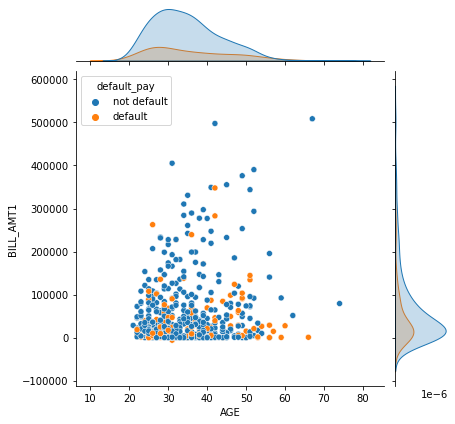
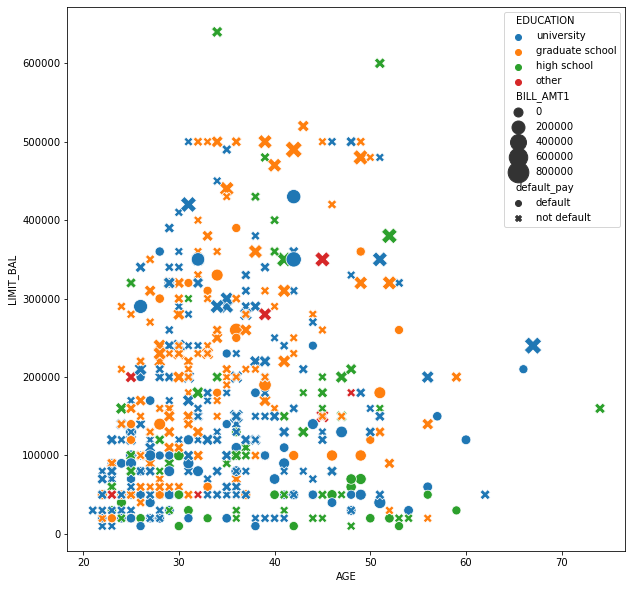
Credit One Customer Default Data Analysis

After the data was cleaned and processed, I began with exploratory data analysis to examine the customer demographics. This helped to gain an overall view of the customers and identify if they skew in one direction or another.

Patterns noted about customers:

* Tend to be younger (most are <40 yrs old)
* Biggest group is university educated
* There are slightly more singles than married

Using multidimensional plotting, we can compare more of the demographics against those that default and those that didn’t to identify patterns. In the following plot, we see a holistic picture of customer demographics against default status. The default pattern is not obvious here, though it can be seen that those with graduate school education tend to have higher balance limits. Also, it can be seen that those with higher balance limits have higher bill amounts, though one would expect this to be so.

In the next plot, we compare balance limit and age to identify a pattern of defaults. We can see that customers will default regardless of age or balance limit. However, defaults tend to be consistent across age, whereas they are less prevalent for customers with higher balance limits. Even more so, older customers with very low balance limits tend to default more often.

The last plot we’ll look at here is a heat map that shows correlation between each feature of the data set. Some obvious correlations exist, such as those with an outstanding balance for 3 months will tend to have an outstanding balance for 4 months, and those with a higher balance limit will have higher payment amounts.

However, when it comes to building a model to predict default payment, my attempts were met with little success. While I couldn’t produce an accurate model that predicted a binary answer to defaulting, I had slightly more success using a regression model that could be used to predict the likelihood of defaulting (as a result of 0 to 1). This model could be used to anticipate how much risk CreditOne is leveraging when taking on certain clients and hedge that with less risky clients. While the model may not be accurate on a micro level, it can still provide some insight at the macro level.

