

Credit One Analysis

CUSTOMER DEFAULT IDENTIFICATION REPORT

Natalie Carreno

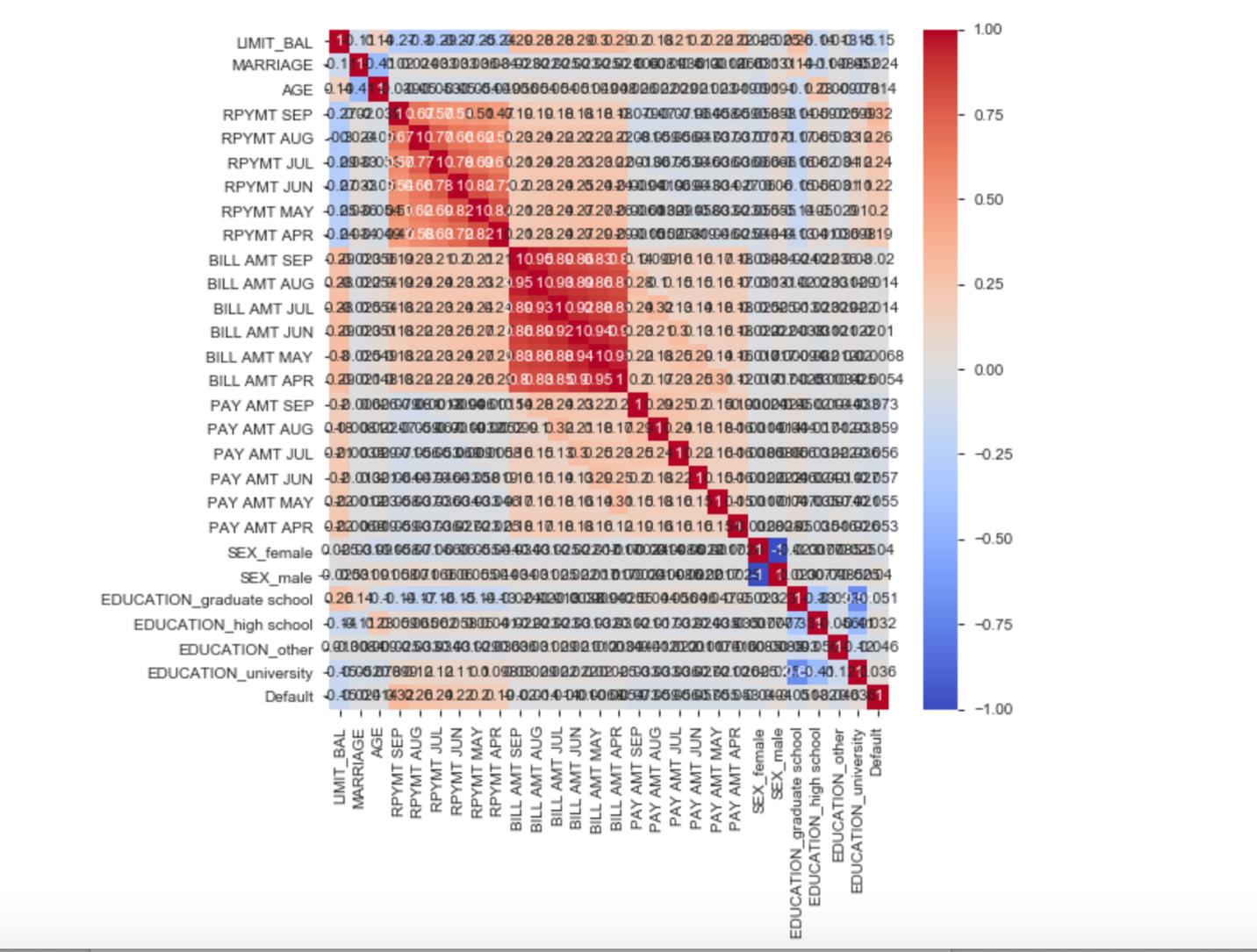
# CREDIT ONE PROBLEM DEFINITION

Credit One has seen an Increase in customers default rate. Credit One has asked to create a model that can predict the credit limit of a customer based off the data provided. This can help future loan approvals.

# Which attributes in the data can we deem to be statistically significant to the problem at hand?

After analyzing the data provided by Credit One, there is not a significant trend that can predict why a customer tends to default. Out of the 30,000 customers only 6,636 will default next month. Customers demographics such as age, marital status, education and gender were not helpful to predict this. The distribution of each attribute was pretty even for default and not default customers.

As seen on the correlation heatmap below, there is not a strong correlation of default with any other attribute. The strongest is 0.32 with repayment status in September 2005.

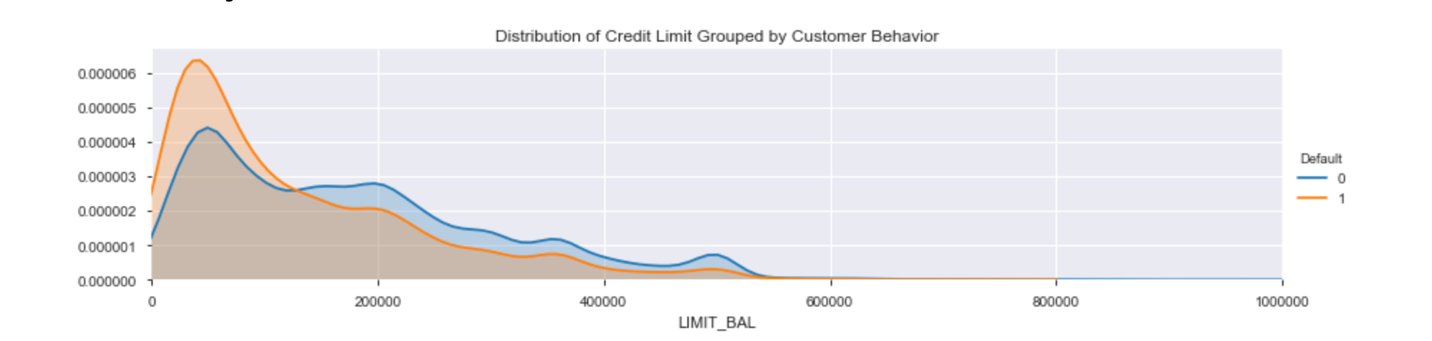


3 different classification models were created to try to predict why a customer will default and the accuracy wasn’t above 65%. Therefore, there’s not a defined pattern among the data that was helpful to predict if a customer will default. Gradient Boosting Classifier had the highest accuracy of 65% in terms of the default rate.

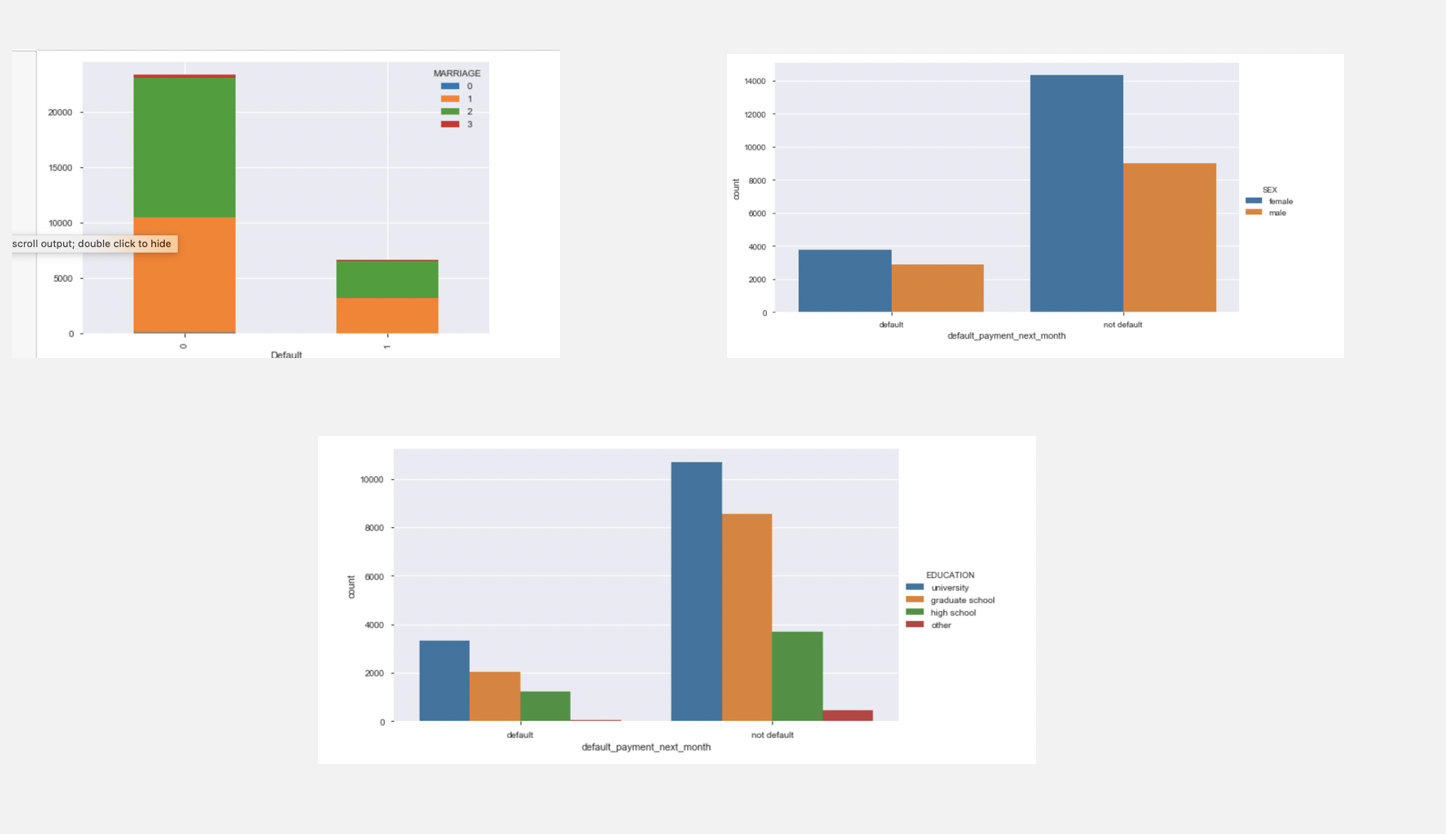
Credit Limit was used as the dependent variable to see if it could be predicted. As seen on the correlation heatmap, there is not a single attribute with a strong correlation. The strongest was 0.28 with the bill amount in September 2005. 3 regression models were built, but they were not very successful either.

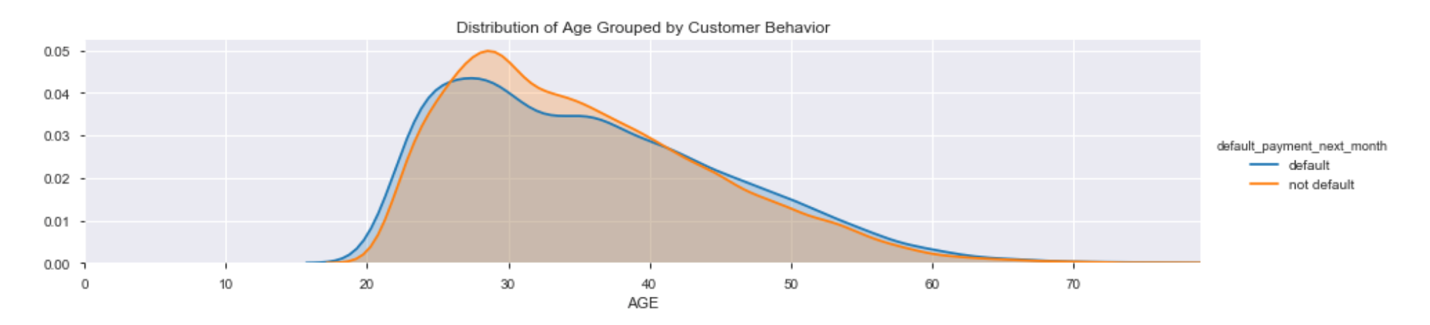
# What concrete information can we derive from the data we have?

Credit limit is distributed evenly among default/not default customers. Customers with a lower credit limit tend to default slightly more as seen on the graph below.



As mentioned before, demographics in the data were not helpful to be able to predict if a customer will default or not. See graphs below for marital, gender, education and age graphs.





If financial data as income, credit score, income/debt ratio was incorporated to this dataset maybe we could have better predictions. Risk could be evaluated and assessed to be able to determine who could be approved for a loan or not.