Raul Barrantes Pampillo
Cenfotec
Learned Lessons Report
Data Science with Python
Task 3: Build and Evaluate Models Customer Default Identification Report

Problem to solve

An increase in customer default rates is bad for Credit One since its business is approving customers for loans in the first place. This is likely to result in the loss of Credit One's business customers.

Questions to Investigate

How do you ensure that customers can/will pay their loans?

As we work with data in hand, we can use the advantages actually we have in hand as data, and prediction methods we are starting to implement now. In fact we cannot assure in a 100% rate all clients approved will not have a default payment. So, based on this, we will select the costumers with the common attributes with a positive approval rate.

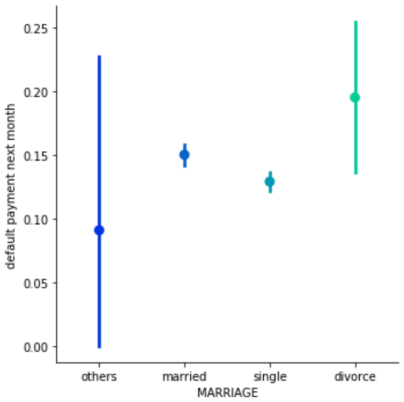
Can we approve customers with high certainty?

Yeah, we can have a high certainty rate if we predict customer default rate based on the attributes with the highest correlation. In addition to this, it is important to update and train the model with relevant data, which can help to make better predictions and make certain assumptions.

Factors that can affect default payment status

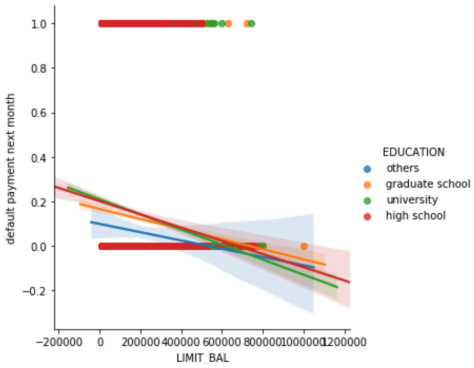
Marriage

In divorced people there is a higher rate to default payment. As with people married and single there is a minor rate, while with other marriage status there is a wider range.


0.25
0.20
0.15
0.10
0.05
0.00
default payment next month
others
married
single
divorce
MARRIAGE

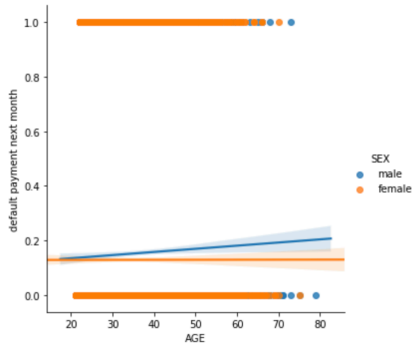
Education

Here we can see correlation between education, the limit balance, and default payment. We can see there is a biggest range of default with high school as highest education than in other levels (graduate or university)



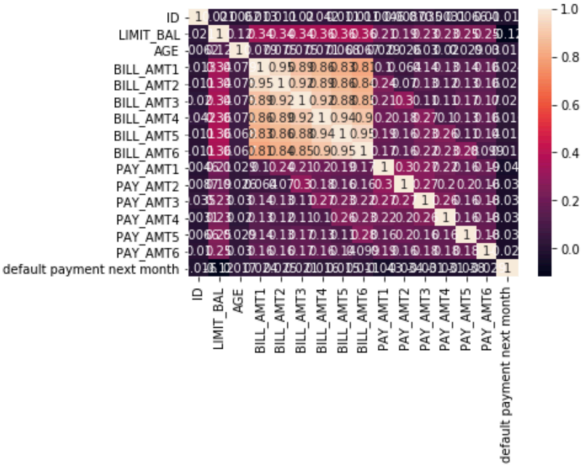
Gender and Age

Another factor that can have influence is the gender; but as we can see on the plot bellow there no high correlation between gender and default credit rate, this is due the rate in female and male is the same. In relation to the Age, there is no so much difference between ages.



Correlation between numeric attributes

As we can see, between the numeric attributes (Related to payment and bills), the correlation between the default payment attribute is the same.



Relevant conclusions

Based on this; there are the following conclusions, to predict which customer can have a default payment:

- Using numeric and historic attributes/elements is relevant to predict future outcomes.
- Gender and age is not an attribute that might influence the outcome.
- If we have more training data we can make more specific and exact prediction.
- Even though age is not a factor, it is important to have information of a wide range of age in order to know which are the overall habits for people.