Jasmina Trbakovic

Project 2 Task2 Lessons Learned

1. How do you ensure that customers can/will pay their loans? Can we do this?

Creditors cannot ensure customers will pay for their loan, but there are various attributes to consider to determine if customers will repay as their income and employment status. Credit One did not provide that information for their customers.

1. Did you learn anything of potential business value from this analysis?

CreditOne has a total credit limit balance of 1 million within its 30,000 customers. Customers with Graduate and University level education seem to have a higher volume of non-default payments. It would seem that the higher education the less of chance that they will default on their loans. Highest number of people are found in age group 25 to 40 with highest density around 30 years old. Customers within the age group of 27 to 37 also have the largest count for non-default payments. Female customers have the largest number of non-default payments. Both single and married have the highest number of non-default payments, but also default payments comparing to divorced and others. 6636 customers out of 30,000 will default next month.

People who had history of paying previous payments ahead of time are less likely to default next month.

1. What are the main lessons you've learned from this experience?

Data had to be downloaded from an external server. I had trouble connecting to a server several times which prolonged my work even more as I had to run all cells multiple times when some cells would not run for no reason. Finally, I just used a .csv dataset instead of connecting via server. Besides MySQL server, data had to be cleaned, transformed and discretized multiple times. Some data had to be converted from text to numbers variables in order to work with it more easily. I also had difficulty converting sex/education/default to an integer without using get dummies code which created several variables. This was very frustrating as it took me a long time to figure it out. First, I had to convert default dependent variable using Label Encoder as it is meant to be used on dependent variables more. Then, the Ordinal Encoder was used for Sex and Education features as it makes EDA easier.

1. What recommendations would you give to the Guido regarding your findings?

In conclusion, our study reveals that single females within the age group of 27-37 with university and higher education level are likely to default on payments.