***How do you ensure that customers can/will pay their loans? Can we do this?*** Creditors cannot ensure that customers will pay their loan, however, several attributes should be considered to determine their ability to repay. The customer’s current income, assets, and employment status are key components to determine the loan risk. At this time CreditOne has not provided these key attributes for their customer base.

***What attributes in the data can we deem to be statistically significant to the problem at hand?*** At this point, we are able to identify that fields **PAY\_1** to **PAY\_6** have the strongest positive correlation in our data as seen in our *Correlation figure*. One can assume that this is based on the fact that customers are likely either always pay or when their payments go to default, they may continue on this patter for some time.

***Did you learn anything about potential business value from this analysis? What concrete information can we derive from the data we have?*** We can derive that CreditOne has a total credit line of over +5 billion within its 30,000 customers. Customer’s in Graduate School and University tend to have credit approved with greater frequency than other customers with lower education, consequently, these two education level groups also seem to have a higher volume of defaulted accounts (*see Figure 1*). In the same manner, customer’s within the age groups of 20-29 and 30-39 have the most approved credit lines. In this study, it is evident that those two age groups also have a larger count of defaulted accounts (see *Figure 2*). 78% of all customers have **no** default payments within the identified 6 months which means that 22% of the approved customers do **have** defaulted payments (*see Figure 3*). The approved line of credit for those customers with defaulted payments is 17.2%, this translates to $863 million (*see Figure 4*). There is no relevant relationship between those customers with defaulted payments and their marital status (*see Figure 5*). Additionally, Female customers have the largest number of default payments by 57% while Males have 43% (*see Figure 6*). As seen in Figure 7, customers with a credit line less than 100k have more defaulting payments compared to other lines of credit.

***What are the main lessons you've learned from this experience?*** This exercise gave us a greater opportunity to take a deeper dive into the various aspects of data preprocessing. Not only did we have to download from an external server but clean, transform, and discretize our data multiple times. In addition to converting data elements from word text to numbers variables in order to easily manipulate our data. One key lesson for me was packages and updates. The first week that I started working on this I struggled to get panda-profiling to work but succeeded after some time. When I started working on this again, I started getting all sorts of errors again along with notifications about other system upgrades. After upgrading, I later started getting other errors. It was pretty frustrating but part of the work.

***What recommendations would you give to the Guido regarding your findings?*** In summary, our study reveals that married females ages 30-39 with university as their highest level of education are likely to default payments. In a similar manner, single males ages 20-29 with university as their highest level of education are likely to default payments.















