

To begin the exploratory data analysis for assignment one, the initial step was to truly understand the data set. Luckily we have been provided a data dictionary that provides an in-depth description of each field within our home equity loan data set. The two target fields that we will leverage as our points of reference are, TARGET\_BAD\_FLAG and TARGET\_LOSS\_AMT. TARGET\_BAD\_FLAG represents whether a loan defaulted (an individual did not pay back their loan) and TARGET\_LOSS\_AMT represents the amount of money a bank has lost due to a default.

Initial findings; it was observed that of the 5,960 loans, 1,189 defaulted, resulting in the banks involved losing a total of \$15,949,932. The average loss resulting from a loan is approximately \$13.4k while the median is ~\$11k. In regards to the different reasons for individuals inquiring on debt, it was observed that there is a higher probability for individuals to use home equity loans for home improvement, compared to debt consolidation; however, the loss of money was higher for those defaulted loans that were categorized as debt consolidation. Finally, it was observed that individuals who work in sales had the highest proportion of defaulted loans, with those who were self-employed following behind; however, those who were self-employed resulted in the largest amount of money lost from the defaulted loans.