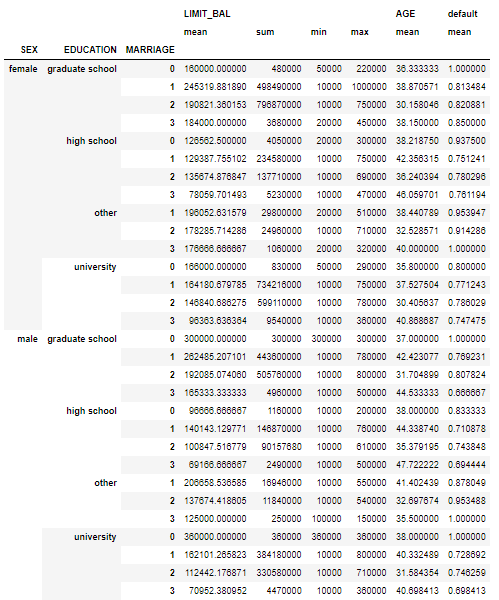
EDA: Credit One Analysis

In my analysis of the variables, I learned that this is a classification problem and not a regression problem. The fundamental question of will someone default or not results in a binary function.

The exploratory data analysis aimed to identify patterns that could indicate if a client has difficulty making their payments, which would cause the financial institution to lose money. In the EDA, it was identified from this dataset that only .7788 of the 30,000 creditors were making their payments. Of the 30,000 creditors that defaulted, an estimated amount of $730,891,447 was a loss. The average customer still owed around $110,000 after defaulting.

The Demographic data yielding import insights and can be summarized by the table below:



Lessons learned:

Explore every aspect of the data. Understand that machine learning will take place in n-dimensional space and that the small insights that you derive from analysis in 2d and 3d space will give insight into the fundamental problem at hand.

Recommendations:

Use classification problem learning algorithms such as logistic regression, support vector classification, K-Nearest Neighbor, and Random Forest algorithms to explore the best predictors.