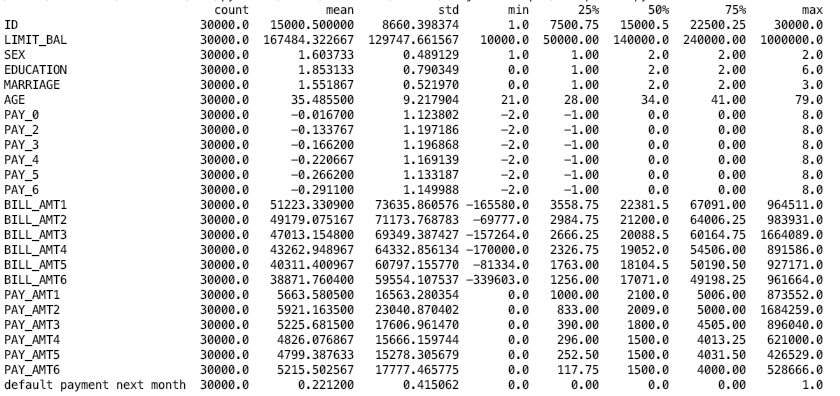
Project Proposal

Our dataset is looking at defaults on credit card payments in Taiwan. Taken from the UCI Machine Learning Repository at this [link](https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients). The dataset has 30,000 observations and includes 23 explanatory variables along with a binary target variable in the last column on whether the person defaulted on the next month’s payment (1 = Yes, 0 = No). Some brief descriptive info on the variables is presented in the table below using the pandas describe function.

Column 1 is a unique person ID, column 2 is the person’s balance limit, column 3,4,5 and 6 are the person’s sex, education level, marriage status, and age respectively. Columns 7-12 contain a history of past payments on a monthly level for the past 6 months coded based on what their payment status is like. Columns 13-18 have their monthly bill totals for the past 6 months. Columns 19-24 contain the dollar amounts they have repaid off of their bill on a monthly level for the past 6 months. After a brief EDA our plan is to test several different types of models to help predict whether or not the person is going to default on their bill. We hope to test a Neural Network, a Support Vector Machine, a Random Forest, and a Naïve Bayes Model. After looking at the misclassification rates for the 4 basic models we start we will choose 1 type of model to then optimize for performance and try to minimize the misclassification rate. The UCI repository states that the best performance with this type of dataset comes from an artificial neural network but we will test different type of models to see whether we can match neural network performance with some other type of model.