# Credit One – Data Analysis Report – Richard Valades

**Analysis Summary:**

Our task for this EDA was to perform a thorough data analysis of the Credit One data. We began with a statistical analysis of the data that we preprocessed. It showed an average age of 35 among customers, with the largest proportion having a marital status of single. The largest proportion of customers were college educated or higher, with a significant segment of customers having a graduate school level of education.

The highest proportion of limit balances were on the lower end of the range with an average of $140,000 and a high of $240,000. One outlier was present in the $1M range. Billing and payment history was provided, and correlation analysis showed a minimal correlation between payment amount and payment history. There were no other meaningful correlations from any of the other features.

Loan default rates were highest among the university education level customers, with a significantly lower proportion of graduate level customers. High School education level customers were also present among default customers but at lower rates. Default rates for married versus single customers were similar with single customers having a slightly greater proportion of default rates than married customers. Divorced customers had very low rates of default.

Women comprised a larger proportion of total customers, as well as higher rates of default. The distribution of customer ages tends to skew to a younger customer base with the largest proportion being in the 25-40 range. Default rates based on age followed a similar pattern.

**Lessons Learned:**

While the billing and payment data was useful, it was very difficult to discern any real pattern in the data. Analysis of those data points was not revealing and what we were able to discover about the data was unremarkable. In of itself, there were no real indicators that seemed predictive of which customers would default on their loans. In the current form, the data does not lend itself to determining any definitive insights. For modeling and machine learning, it will be necessary to process this data further for the algorithms to identify any patterns that might be predictive.

**Recommendations:**

In this EDA, we evaluated the data without feature selection or any modification of the dataset beyond the preprocessing stage. While the payment and billing histories may be helpful, applying them in the current file structure along with the demographic information may not be the best approach. Consolidating the billing and payment data into “summary” statistics would make the data easier to work with at the EDA stage and might be more revealing.