# Credit One Preliminary Report

Credit One has seen an increase in the number of customers defaulting on loans they have secured from different business partners. Credit One is the credit scoring service and may lose business if this problem is not solved quickly. Clients depend on Credit One to make accurate recommendations on extending credit, and as a result of this recent up-tick in defaults, clients are losing revenue and customers, which could lead to the eventual loss of clients and revenue for Credit One.

Credit One learned some lessons from a similar problem last year:

1. We cannot control customer spending habits
2. We cannot always go from what we find in our analysis to the underlying "why"
3. We must focus on the problem(s) we can solve: What attributes in the data can we deem to be statistically significant to the problem at hand?
4. What concrete information can we derive from the data we have?
5. What proven methods can we use to uncover more information and why?

Credit One is seeking a better way to understand:

* How much credit to allow someone to use
* At the very least, whether someone should even be extended credit

The data science team has been enlisted to design and implement a creative, empirically sound solution. This presents some lessons learned from conducting Exploratory Data Analytics (EDA) before we begin the modeling portion of the effort.

## Key take-aways and Lessons Learned

We have gained a deeper understanding of the data provided, and some of the problems with the data. These lessons identified the benefit of conducting a thorough EDA. For example, the information provided had redundant records, and also extraneous rows of headers that all had to be removed.

Initially, none of the features were numeric, and had to be converted to numeric. Following that conversion, 3 of the 24 features provided were still not numeric and needed to be changed in order to proceed with EDA and modeling. The features that needed to be converted are:

* Sex
* Education
* Default payment next month

After cleaning up the provided data, we ended up with records on 30,000 customers, and 24 features on each customer.

A dive into understanding of the relationship between the different features proved that we have to just accept the data. According to the data dictionary, the features relating to repayment status should show when payments were made. However, trying to track the logic through the payment status fields did not produce the expected results. As a result, we decided that the data had to be accepted and we moved on.

The EDA did not provide any surprising or significant insights into the problem at hand. The correlation and covariance matrices did not reveal any significant findings. A thorough investigation into the interrelationships between features also did not reveal any surprising insight. Such events occur in data science and are not unexpected.

# Recommendations for Credit One

With EDA completed, we will now progress to data modeling, and use linear regression models were appropriate, and also classification models. We will also seek to develop a creative, empirically sound model to reduce the number of customers defaulting on their loans by better predicting:

* How much credit to extend
* And/or, whether to approve an applicant at all

We will provide the results of the modeling effort in two weeks.