**Problem Statement**

**Background**

Money lending is the world’s oldest profession, and thus began the concept of a Lending club. While investing in a Lending club, we want all the lendees to pay back the amount with proper interest to make it profitable. This case based on the ability to predict people with the ability to pay back the money.

**Problem Description**

Due to lack of thorough background checks and standard models to analyze the potential lendees, today there are a lot of people who are unable to pay off their debt and thus their debt keeps on increasing and the lending club won’t be a success. If this problem is ignored, it will lead to more of people with increasing debt and loss of investment.

**Available Data**

There are two data sets, RejectStats.csv and LoanStats.csv, one of these files contain complete lending club loan data for all loans issued through the time period of 2007-11, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information and the other one contains the list and details of all loan applications that did not meet Lending Club's credit underwriting policy.

**Assignment Goal**

The first objective is to analyze all the parameters in the dataset, deal with the missing variables and determine the most important ones based on their contribution and correlation with others to get an appropriate number of parameters for the models.

Build classification models that will generate a flag whether to give a loan or not. Using logistic regression and neural network models and then deciding upon the best model.

Once decided to give a loan, build models to decide what interest rate to give. This can be done by one prediction model considering all lendees or clustering the lendees using clustering algorithms and then using the prediction algorithms based on the clusters. Then calculate the accuracy measures and try to improve the models.