The Lending Club Corporation: Predicting and Profiling Customers Who Will Miss a Payment

Sinem Buber Singh, Liang Zhao, and Dan Silitonga

The Graduate Center

sbuber@gc.cuny.edu, ch00226855@gmail.com, dannysilitonga@gmail.com

March 25, 2015

Introduction to Peer to Peer Lending and the Data

- The Lending Club Corporation works as a third party intermediary connecting investors and borrowers. Borrowers borrow money mainly to re-finance their credit card debts, as a more favorable interest rates may be obtained. Investors lend money to gain an attractive rate of return of their investments, assuming that their borrowers do not default on the loans.
- The entire dataset is about 1GB to 3GB.
- Data clean up is required.

Goals

- To build a model to predict those who will be late in their payments as well as those who are charged with a very high interest rate.
- To be able to create a profile of these groups of people. Who are these people? What characteristics do they have in common?
- The long term goal, which may be beyond the scope of this course, is to quantify the role of high interest loans on the financial well being of NYC residents, especially among the City's most vulnerable population. How much do these loans play a role in putting or keeping NYC residents in poverty?

Data Clean Up Highlights

- The column interest rate has a percent sign symbol. This would create issue for plotting and building a regression model. The percent symbol needs to be stripped and the variable converted into a float.
- The column loan length has the word month in each record. The same procedure is required.
- The FICO score column includes a range and should be converted into 2 columns of low and high scores.

Data Exploration Highlights

- The loan amount does not seem normally distributed. The center of the distribution is skewed, with a large tail on the right of the distribution.
- The amount requested also does not seem normally distributed. Both loan amount and amount share the same look and feel. One difference is that some loans are rejected by the investors. Consequently, the loan amount is set to 0.
- There is a linear relationship between interest rates and FICO Score and Monthly Income

Linear Model

There is a linear relationship between interest rates and FICO Score and Monthly Income, with an R square of 65 percent.