**Problem Statement**

I plan to analyze the loan dataset from Lending Club from the year 2007-2015. I plan to identify to things: 1. Identify the characteristics/variables that would lead to a high likelihood of default on the loans and 2. Identify the characteristics/variables that would lead to high likelihood of paying the loan off early. Then after the characteristics are identified I would like to forecast the potential losses that would be incurred on Lending Club from people that default as well as the people that pay their loan off early.

**Hypothesis/Assumptions**

I believe variables that lead to default and the variables that lead to paying early will be inversely proportional to each other. Borrowers that meet Some variables that I plan to explore include, but aren’t limited to:

* High income vs low income
* Source of income verified vs unverified
* Home owners vs non-home owners
* Employment length (we’ll say more than 2 years at current job vs less than 2 years)
* Last\_fic\_range\_high/last\_fic\_range\_low (credit score)
* Length in the credit bureau we’ll say more than 2 years vs less than 2 years

Once the variables are identified that lead to default/prepayment I will find borrowers in which all variables are present use them predict prepayment or default and project how much could be lost by Lending Club.

**Success Metrics**

This will be considered successful if I am able to identify and separate 3 to 5 variables that help identify borrowers that will potenitally default or prepay a loan. Once identified I will take the amounts of the loans to determine the potential losses that could be incurred on Lending Club.

**Limitations of the Dataset**

Some limitations of the dataset are that it does not include marriage stats. I was hoping to compare married vs unmarried borrowers. Also, this only takes into account the borrowers information at the time of the application. There are more many factors that could occur to borrowers during the life of the loan could lead to default or early payment, including but not limited to work promotion or loss of employment, inheritance of large sum of money, medical expenses, etc.