Lending Club Loan Data

**Proposal**:

I’ve decided to analyze Lending Club’s loan data available on Kaggle which has all loans issued through 2007-2015. Since I work for a Top 10 US Bank and credit-card issuer, I see this as a good opportunity to get my hands dirty with credit related decisioning models.

Below are high-level hypothesis going into the project:

* From a credit decisioning perspective (before the loan is given out), I would like to analyze the effect of the following on systematic on-time loan repayments:
  + Debt to income ratios.
  + Credit factors including fico ranges, age of credit history, months since last delinquency
  + Utilization rates on past revolving balances.
* From an ongoing monitoring perspective, I’d like to analyze the same factors mentioned above in addition to:
  + Monthly payments vs. committed monthly payments.
  + Increases or decreases in household income.
  + Increases or decreases in utilization rates on revolving balances.

**Goals and Success Metrics:**

Success would depend on the model’s ability to accurately predict the two outcomes mentioned above with a certain degree of confidence. In the real world, I would take my learnings and directly apply it to a lending product I’ll be working on over the next 6 months.

**Risks or limitations:**

The dataset is less than 1M rows. Ideally, we would have a lot more information to build a well-trained model. In addition, there could data inaccuracies for income related information because these figures are self-reported for the most part.