**Infusionsoft Data Science Challenge**

Attached is data from Lending Club, containing information for their loan portfolio between 2015 and mid-2016. Lending Club provides loans for credit card refinance, debt consolidation, and various other forms of personal loans. Lending Club’s loans are peer-to-peer, meaning you as an individual investor are able to provide funding for this loan, taking on the risk and reward of a portion of the loan.

The dataset attached contains information about each loan, including funded amount, interest rate, term, installment, and status. It also contains a lot of information about the loan recipient, such as housing status, employment status, purpose of loan, location, etc. Your goal is to understand the various factors within the dataset that are signals for risky loans, and build a model to predict the probability of each loan defaulting.

R or Python is preferred, but use any language you’re comfortable with. Please send all of your code (including data import, exploratory analysis, transformations, modeling, accuracy evaluation), plus a write-up of the steps you took to solve this challenge, to [brett.nebeker@infusionsoft.com](mailto:brett.nebeker@infusionsoft.com). Include any reasoning or assumptions for each step in your write-up so we can better understand your thought process as you went through this problem. Also include any accuracy measures you used to evaluate your model(s), whether it be from a data science perspective, or from an investor perspective. Use the same email address if you have any questions.

Good luck!