The Lending Club is an online, peer to peer lending platform. They aggregate requests for loans and investors in order to provide an alternative venue to the more common fundraising platforms. Investors participating in the platform face credit risk, the possibility that the loan they made will default. The lending club attempts to provide clarity by assigning a grade to loans that rates their likelihood of default and assigns interest rates based on said loans. I am exploring the possibility that it may be possible to predict the rate of default for the platform’s loans with greater accuracy, given the data they already collect, than the current rating system.

* + What important fields and information does the data set have?
    - There are several fields of importance, I will focus on loan amount, home ownership status, annual income, fico score, and loan status
  + What kind of cleaning and wrangling did you need to do?
    - I have to create some calculated values, simplify some features, and remove entries with missing data.
  + Are there other datasets you can find, use and combine with, to answer the questions that matter? No, I have no way of finding information on these individuals from a different source since the data is anonimzied.
* Any preliminary exploration you’ve performed and your initial findings. Test the hypotheses one at a time. Often, the data story emerges as a result of a sequence of testing hypothesis e.g. You first tested if X was true, and because it wasn't, you tried Y, which turned out to be true.
  + I performed some basic analysis using inferential statistics and found some trends related to the applicant’s location.
* Based on these findings, what approach are you going to take? I am going to use supervised machine learning.