**Capstone Project-1 Proposal- Loan Risk Analysis**

**Jun Yin**

* What is the problem you want to solve?
* Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn’t have done otherwise?
* What data are you using? How will you acquire the data?
* Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.
* What are your deliverables? Typically, this includes code, a paper, or a slide deck.

These days, most people are living with loans through their entire lives. Whole bunches of different loan products are becoming available to cover every aspect of people’s daily life. Depending on individual’s situation, the question whether the loan should be approved is the problem resolved here.

The banks and credit unions will be the client interested in this problem, since they do not want to lose the potential customers, in the meantime, default risks is the last thing they would like to take.

The raw data used in this project will be downloaded from Quandl, Kaggle Competitions or US Government Open Data. “Loan risk” are going to be used as key words for raw data searching.

The first step to deal with the raw data should be data cleaning. Find out if there is any missing data will be the first operation, and following by filling in reasonable estimations for missing data. According to multiple available information, such as credit score, annual income, total debt, etc, the model will give a percentage of the individual is capable to repay loans. A perfect fitting model would be delivered to help predict if the loan application approved or not.