# Mid-term Project

# Lending club – Qi Huang

* **The reason why you are interested in this project**

The lending club project is related to loan pricing based on individual borrowing information. As one of the major business in commercial bank, personal banking contributes significantly to revenues, and revenues directly relate to interest rates. Usually, interest rates depend both on intrinsic features of loans issued and on client’s credit information. For example, long-term loan tends to have higher interest rates compared to short-term loans, and clients who have lower credit scores tend to borrow loans with higher expenses. I am personally interested in this project since I am targeting to be a quantitative analyst in the finance industry. And majority of quantitative analyst opportunities are in the commercial banks conducting asset pricing. Going through this project would provide me “real-life” experience for future career.

* **The dataset you are going to use for your project**

The dataset I am going to use is the lending data between 2012-2015, from Lending club website. Target outcome would be the interest rate, and the selected variables would be the lending amount, lending period, credit score and state. Meanwhile, considering the different tax effect on the interest rate at state level, I might find out tax data from government website to strengthen my model.

* **The question you are going to answer with this project**

This project is trying to investigate the relationship between the interest rate of each loan and the requested loan amount. While each loan amount requested is considered at individual level, all the other variables such as the credit rating of clients, the state where the clients borrow the money would be treated as group level effect. The groups would be distinguished by credit rating as well as if the client has a mortgage or not.

* **Potential problem that you foresee with this project**

Since the dataset is pretty large and contains lots of information, clean the data and select preferred variables would take some time. Also, doing EDA to investigate the relationship between interest rate and other variables are important and challenging as well.

* **The timeline that you will work on this project**

11.9-11.15 data cleaning

11.16-11.22 modeling

11.23-11.27 cross-validation

* **Future direction that you might take this project to**

After completing the pricing part with interest rate, I will direct this project to risk analytics, trying to figure out if the late payment has some relationship with interest rate level, credit rating and other variables. Try out simulation to predict the future late payment or charge-offs.