Determination of Factors influencing Interest Rates apart from Fico Score

Introduction:

The Fico score is mostly used to calculate credit worthiness of customer in US. It is mainly calculated based on the factors such as payment history, credit utilization, length of credit history etc. Based on the score, the banks will make lending decisions. The higher the score, the better chances of getting loan at a better rate.

Though Fico score is the main factor in deciding the interest rate of other factors also influence it. In the analysis described below, we will see that both Amount Requested and Loan Length plays an important role.

Methods;

Data Collection

For our analysis we used a sample of 2,500 peer-to-peer loans issued through the Lending Club (https://www.lendingclub.com/home.action). The data were downloaded on Saturday 16 Feb 2013 using the R programming language.

Exploratory Analysis
After some cleaning, simple linear regression was performed for several variables against interest rate. By comparing the resulting coefficients of determination some variables appeared to fit significantly better: loan length and amount requested.

Statistical Modeling
Multivariate linear regression model was applied using loan length and amount requested variables, as they were revealed in our exploratory analysis. Coefficients were estimated with ordinary least squares and standard errors.

Reproducibility
Performed with R programming language [1], the analysis was structured using "ProjectTemplate"[2] package and it has been published on a Github repository[3]. The graphs can be found at assignment1/graphs folder and assignment1/src has source files. Assignment1/munging has code to clean up loans data.

Results:
An exploratory analysis determines great differences on interest rate for 36 and 60 months loan length (1). Even though the amount requested slightly affects interest rate, there appears to be a trend.

A separated linear regression (1) on interest rate (%) vs. FICO range (score) for both 36 and 60 months loan's length shows the clear trend of higher interest rate on 60 month. Adding the amount requested (2) to the model, as a relation with loan length, improves the coefficient of

determination, and that drives us to think it's a better way to model it. So, taking into account potential confounders, our final regression model was:

$IR = b_0 + b_1 (FR) + b_2 (AR) + b3(LL) + e$

where b_0 is an intercept term, b_1 represents a change of interest associated with each higher value for FICO range and b2 it's b_2 the coefficient representing the change in amount requested (AR) impacts the interest rate and b3 for loan length (LL). The error term e represents all sources of unmeasured and un-modeled random variation.

We observed a statistically significant (P = 2e-16) association for this model, with a coefficient of determination (r-squared) of 0.7868.

Conclusions:
Our analysis suggests that there is a significant association between interest rates and the amount of money the applicant would pay each month, for applicants with same FICO score.
Our model is best fitted between 700-820 FICO Ranges, regarding the original sample lacks of enough data outside this limits.

As conclusion, as FICO score itself reflects some of the rest of variables, it's reasonable to think the lender doesn't takes gives such importance to other components again on calculation of interest rate.

1. R-project.org
2. www.projecttemplate.net
3. www.github.com/mmutham/coursera-dataanalysis