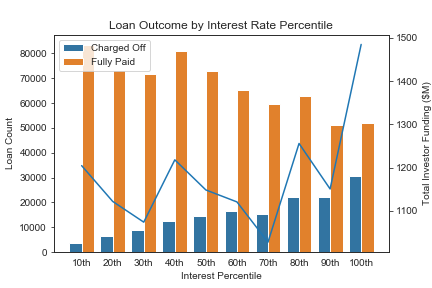
Lending Club Loan Analysis Review

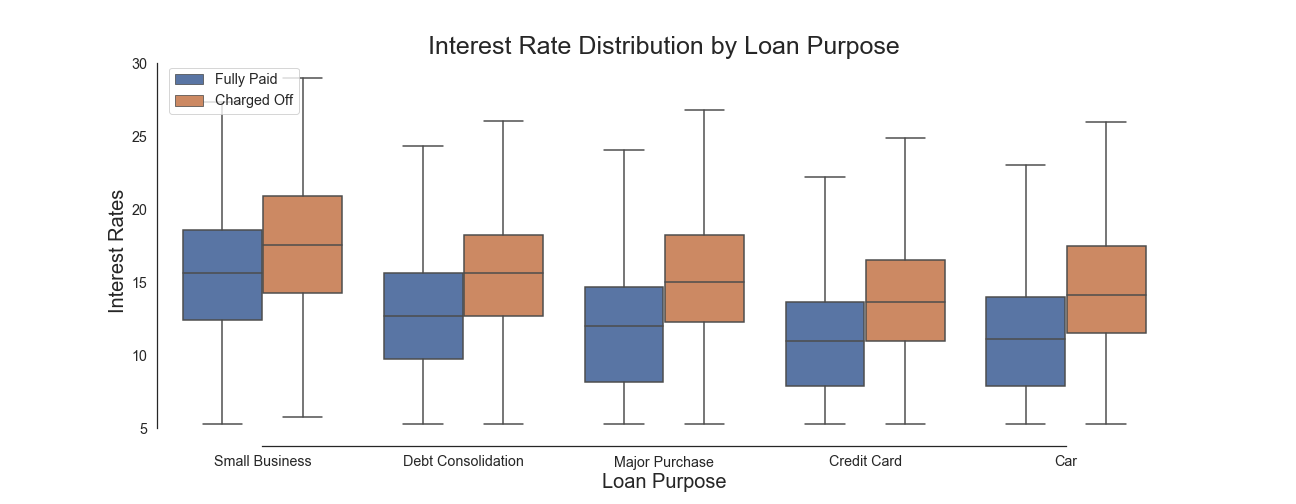
Francesco Balestra

I. Data Exploration and Evaluation

Investor appetite is heavily concentrated among loans with the highest interest rates. This demand could be behind the higher average loan sizes in high interest rate brackets.



This appetite for yield comes with significant risk, as default rates among these higher rate loans are significantly higher, as they are in fact negatively correlated. When Looking at the loans from a source perspective, there is significant spread in interest rates, and the major categories appear to be mostly highly correlated with the market.



II. Business Analysis

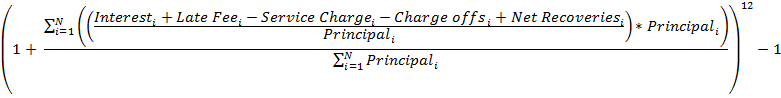
Summary

I am looking to estimate the average rate of return on 36 month loans in the LC portfolio overall and by cohort. I calculate an average return of 3.53%. This does not make investing in LC a good investment, as the average 3-yr US Treasury bills interest rate over 2007-2015 was 1.56%. Investors are not being compensated for the additional risk on these loans. Having said that, when looking at top performing loan cohorts, it is evident that there is large dispersion around the mean and with good loan selection, this could in fact be a good investment.

Process

To estimate the rate of return on the Lending Club (LC) loans, I implemented the LC suggested formula for net annualized return (see below). This formula normalizes for duration, allowing for a cross-duration comparison.

Lending Club Net Annualized Returns Formula



Major Assumptions:

* All loan entries are up to date to most recent entry date found

Results:

Average rate of return on 36 month term LC loans.

|  |  |
| --- | --- |
| Calculated Net Annualized Return | |
| Mean | Median |
| 3.53% | 5.87% |

Average rate of return on 36 month term LC loans for top cohorts by year and loan grade.

|  |  |  |  |
| --- | --- | --- | --- |
| Calculated Net Annualized Return | | | |
| Year | Grade | Mean | Median |
| 2008 | C | 1220% | 23.6% |
| 2007 | G | 111% | 29.3% |

III. Modeling

Summary

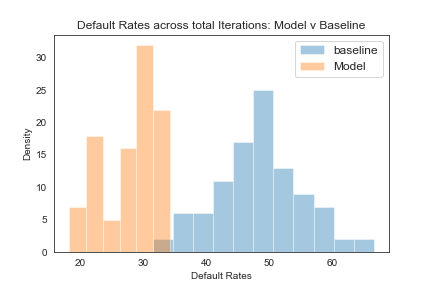
I built a random forest classification model on LC loan data to predict whether a loan would be fully repaid, given the available LC data on the loan and debtor(s). My rationale for utilizing this model was that decision trees are great at handling large feature sets and are effective in showing me which features really effect loan performance.

Model Efficacy

Initial results of the random forest model are mixed. The model has high precision in predicting loans that will default, which is important given that false positives are the most harmful to investment returns. However, it performs very poorly at predicting whether a loan will be fully paid. Overall it had a 13% accuracy boost over a baseline dummy model at 80.8%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | f1-score | precision | recall | support |
| FALSE | 0.35 | 0.56 | 0.25 | 75951 |
| TRUE | 0.89 | 0.83 | 0.95 | 300582 |
| accuracy | 0.81 | 0.81 | 0.81 | 0.81 |
| macro avg w | 0.62 | 0.70 | 0.60 | 376533 |
| avg | 0.78 | 0.78 | 0.81 | 376533 |

I ran a second, practical model evaluation on potential model loan portfolio returns versus a baseline to better understand its performance. Binning the loans by interest rate, I ran a simulation that randomly chose a number of loans to be selected from each bin and placed into a portfolio. The model chose the loans with highest predicted probability of being paid in full. The baseline randomly selected from each bin. The model had a mean portfolio default rate of 29.8% vs the 48.82% random baseline.



I believe this model was a good step for helping predict loan outcomes above the baseline. The major constraint this model has is it tends to overfit the data, which would lead me to try and utilize a few techniques to make it more general. Next, I would next try running a random forest regressor to predict net annualized returns.