

LoanAnalysis

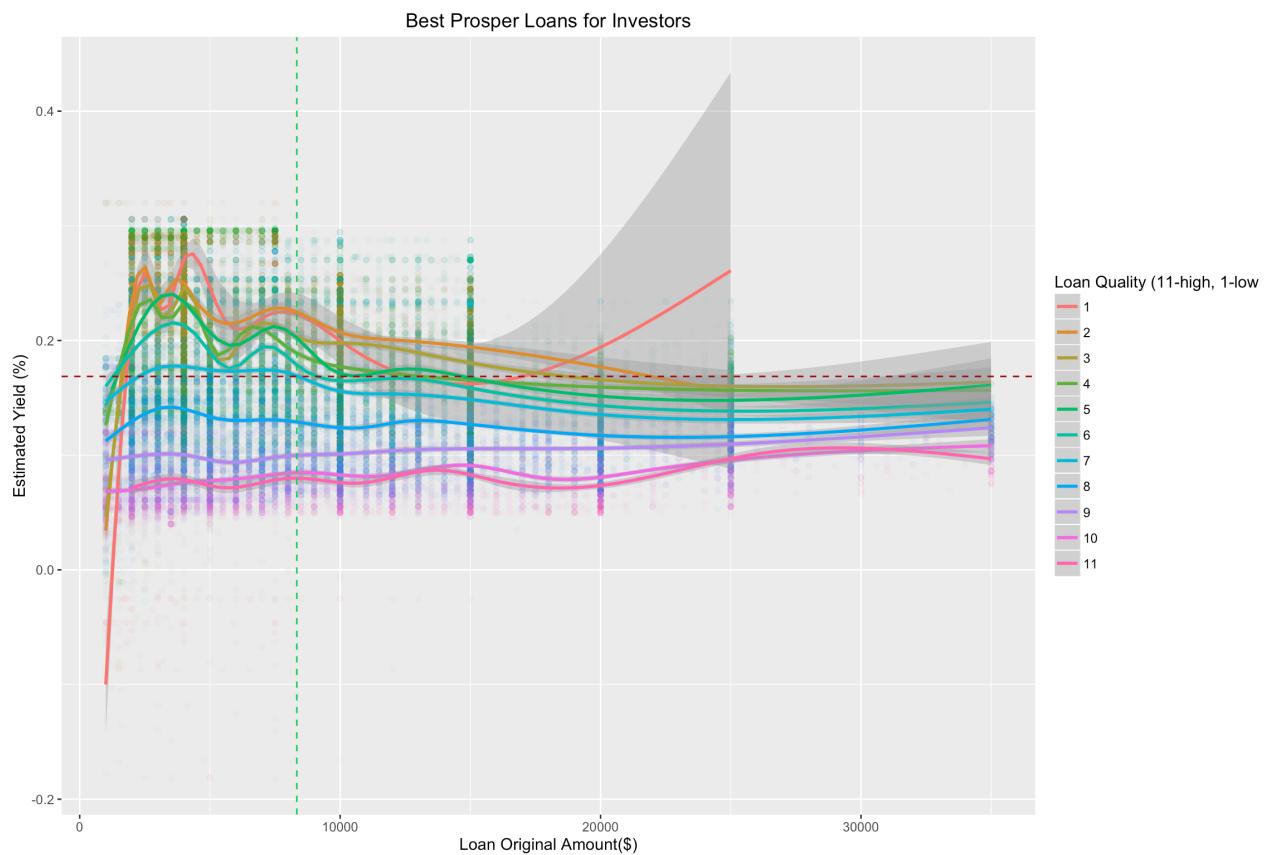
ProsperLoan Data Analysis

***### #Final Plots and Summary

Prosper Loans facilitates crowdsource funding of medium size loans. Most often these loans are used for home improvements, starting or expanding a business, attaining a degree or purchasing a car. Prosper Loans is different from a traditional bank which loans out money from it's deposits and charges interest. Instead, Prosper makes loan requests public so individuals may provide loans to borrowers and the lenders then collect the interest instead of a bank. Prosper takes flat fee on every loan that is considerably smaller than what a traditional bank will charge.

In my 3 plots I will examine this dataset from each of the three stakeholders in the Prosper Loans model: 1) Borrowers seeking a loan, 2) Investors looking to profit from backing a loan, and 3) the Prosper Loans company overseeing the operation. In my plots I will investigate how each party can maximize the Prosper Loan process. How can borrowers get the best borrowing rates? Which loans should investors back to get the highest yield in return? For which type of loans is Prosper Loans having the most success and growing business?

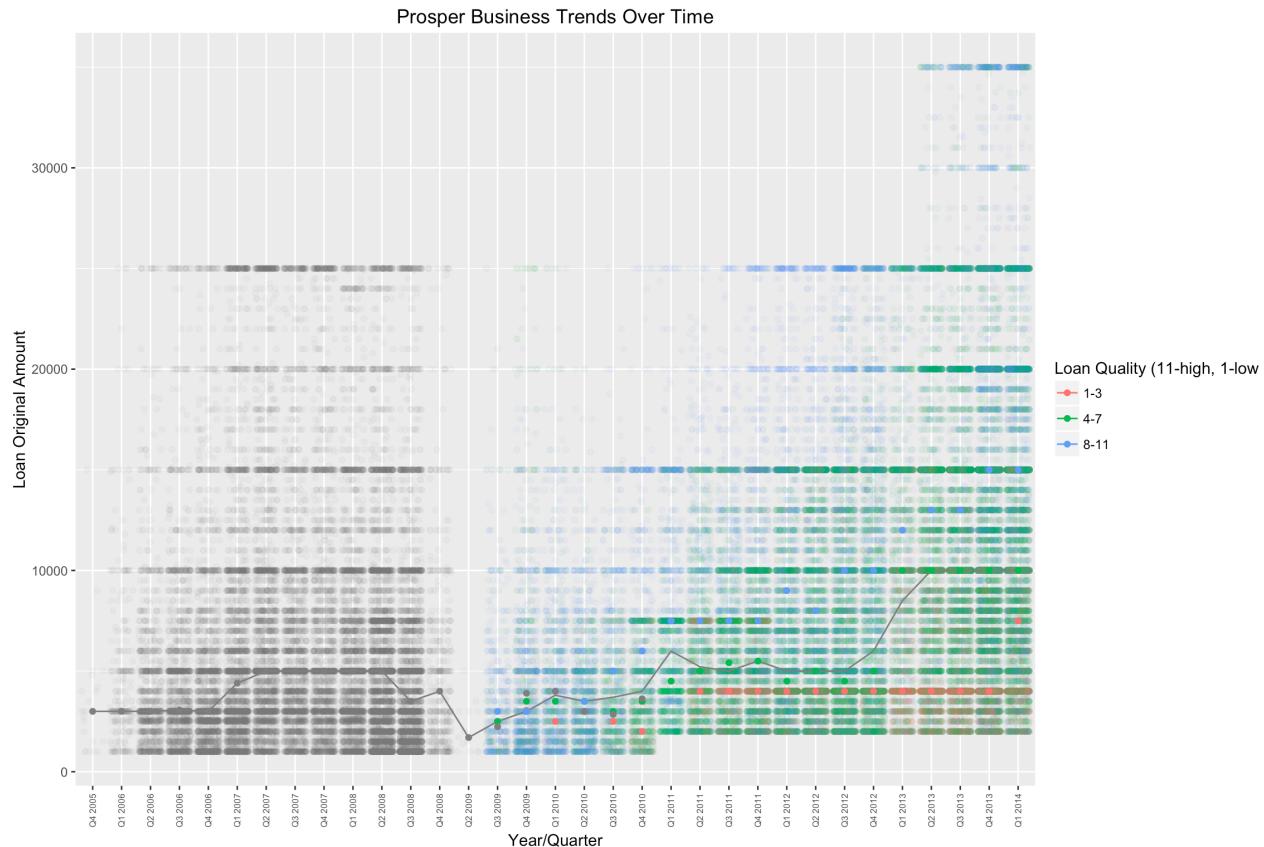
Final Plot 1 - Prosper Loans investors best strategy for high yields Prosper gives each loan application a quality rating between 1 and 11 (Prosper Score) based on the credit score and finances of the prospective borrower. In plot 1 I show the estimated yield for loans of each score of all amounts. The plot shows that the lender can receive a higher yield in return for his or her investment by backing a loan with a lower score. The best loan for investors to back in order to receive maximum return has Original Amount between \$2000-6000 and ProsperScore between 2 and 6. Loans with higher quality scores return less yield because the borrower has excellent credit and will be offered a low borrowing rate. Loans with lower quality score sometimes have negative yield because of frequent defaults by borrowers with bad credit history. The mean yield for all loans is 16.9%, but lenders who back \$2000 - \$6000 loans with quality score in range 2-6 will receive estimated yields upwards of 25%.



```
## Source: local data frame [12 x 2]
##
##    ProsperScore      avg
##             (dbl)    (dbl)
## 1           11 0.08060713
## 2           10 0.08263573
## 3            9 0.10153552
## 4            8 0.12835065
## 5            7 0.15980594
## 6            6 0.17894558
## 7            5 0.19658030
## 8            4 0.19668490
## 9            3 0.21641180
## 10           2 0.23689066
## 11           1 0.23815458
## 12          NA      NA
```

Final Plot 2 - Prosper Loan business and lending trends over time Like many financial institutions during the 2008-2009 US housing market and financial crisis, Prosper significantly scaled back business. In the first quarter of the 2009, when the economy reached a low point, Prosper facilitated 0 loans for the entire period. Prosper initially started facilitating loans in 2005 and saw business grow steadily before falling off quickly in 2008. Since the second quarter of 2009, Prosper has steadily

grown and surpassed pre- crisis business marks. Plot 3 shows all loans facilitated by Prosper. In 2009 Prosper began assigning the quality grade Prosper Score (1-11) to all loan applications so this data is visible by plot coloring after 2009 but not before. Prosper is increasing business volume in recent years (shown by the density of points) and increasing the median amount of loans (shown by the black line connecting the median loan value for each year/quarter). From the colored plots, we see that Prosper's largest increases in loan volume come from loans with medium prosper score in range 4-7. This plot also shows us that propser raised the minimum loan limit from \$1000 to \$2000 at the end of 2010, and raised the maximum limit from \$2500 to \$3500 starting the second quarter of 2013.



```
## 
## Q4 2005 Q1 2006 Q2 2006 Q3 2006 Q4 2006 Q1 2007 Q2 2007 Q3 2007 Q4 2007 Q1 2008 Q2 2008 Q3 2008 Q4 2008 Q1 2009 Q2 2009 Q3 2009 Q4 2009 Q1 2010 Q2 2010 Q3 2010 Q4 2010 Q1 2011 Q2 2011 Q3 2011 Q4 2011 Q1 2012 Q2 2012 Q3 2012 Q4 2012 Q1 2013 Q2 2013 Q3 2013 Q4 2013 Q1 2014
##    22      315     1254     1934     2403     3079     3118     2671     2592
##   Q1 2008 Q2 2008 Q3 2008 Q4 2008 Q1 2009 Q2 2009 Q3 2009 Q4 2009 Q1 2010 Q2 2010
##   3074     4344     3602      532       13      585     1449     1243     1539
##   Q3 2010 Q4 2010 Q1 2011 Q2 2011 Q3 2011 Q4 2011 Q1 2012 Q2 2012 Q3 2012
##   1270     1600     1744     2478     3093     3913     4435     5061     5632
##   Q4 2012 Q1 2013 Q2 2013 Q3 2013 Q4 2013 Q1 2014
##   4425     3616     7099     9180    14450    12172
```

Final Plot 3 - ProsperLoan best strategy for borrowers While lenders may select what type of loan or loans to invest in with Prosper in order to get a higher or lower return, borrowers do not have this flexibility. The rates for borrowers are determined by their credit score and personal finances.

Surprisingly, there are not many factors that a prospective borrower may adjust on his or her application (such as loan amount or payback period) to get favorable rates. A borrower looking for the best rates on a loan ought to compare Prosper with other lending institutions. In this plot I show the strong relationship between credit score and borrower APR. The correlation coefficient for these features is (-0.6682872). I also use coloring to show median values for large and small loans to show that larger loans tend to be charged lower borrowerAPR. Like with many lending institutions, borrowers with the best credit history will be offered the best borrowing rates.

```

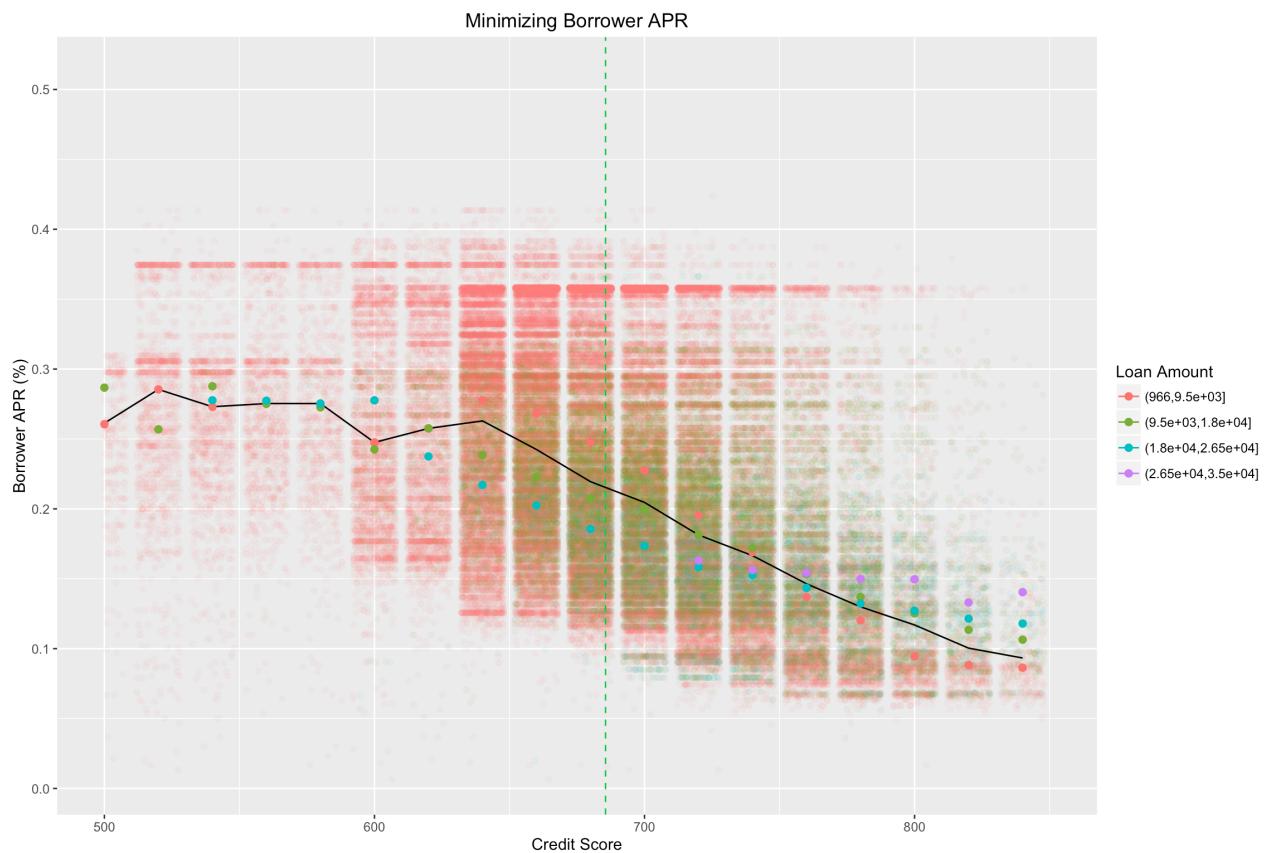
##   EstimatedEffectiveYield BorrowerState          Occupation
##   Min.    :-0.183        CA      :14717  Other           :28617
##   1st Qu.: 0.116        TX      : 6842  Professional     :13628
##   Median  : 0.162        NY      : 6729  Computer Programmer : 4478
##   Mean    : 0.169        FL      : 6720  Executive       : 4311
##   3rd Qu.: 0.224        IL      : 5921  Teacher          : 3759
##   Max.    : 0.320        : 5515  Administrative Assistant: 3688
##   NA's    :29084         (Other):67493 (Other)          :55456
##   EmploymentStatusDuration BorrowerAPR          ProsperScore
##   Min.    : 0.00          Min.   :0.00653  Min.   : 1.00
##   1st Qu.: 26.00          1st Qu.:0.15629  1st Qu.: 4.00
##   Median  : 67.00          Median :0.20976  Median : 6.00
##   Mean    : 96.07          Mean   :0.21883  Mean   : 5.95
##   3rd Qu.:137.00          3rd Qu.:0.28381  3rd Qu.: 8.00
##   Max.    :755.00          Max.   :0.51229  Max.   :11.00
##   NA's    :7625            NA's   :25      NA's   :29084
##   CreditScoreRangeLower BorrowerAPR.1          LoanOriginalAmount
##   Min.    : 0.0            Min.   :0.00653  Min.   : 1000
##   1st Qu.:660.0            1st Qu.:0.15629  1st Qu.: 4000
##   Median  :680.0            Median :0.20976  Median : 6500
##   Mean    :685.6            Mean   :0.21883  Mean   : 8337
##   3rd Qu.:720.0            3rd Qu.:0.28381  3rd Qu.:12000
##   Max.    :880.0            Max.   :0.51229  Max.   :35000
##   NA's    :591              NA's   :25
##   AvailableBankcardCredit
##   Min.    : 0
##   1st Qu.: 880
##   Median  : 4100
##   Mean    : 11210
##   3rd Qu.:13180
##   Max.    :646285
##   NA's    :7544

```

```

##   1-3    4-7   8-11  NA's
## 14400  45283 25170 29084

```



Correlation ProsperScore to APR

```
## [1] -0.6682872
```

Correlation Credit Score to ProsperScore

```
## [1] 0.369603
```

Correlation Prosper Score to BankCard Credit

```
## [1] 0.318558
```

Correlation Prosper Score to Loan Amount

```
## [1] 0.2662933
```

***### ##Reflection

While investigating Prosper Loans and answering my general questions with calculations and plots, I found several areas that would be interesting for future investigation.

While looking at Prosper from the perspective of an investor/lender, I noticed that the data is often

makes assumptions about events that will happen regarding the loans in the future, but these events are uncertain. For example, for completed loans or loans made in the less recent past, we have data available about whether the loan is paid off successfully or if it is past due or in default. For defaulted loans we can calculate how much the investor lost. It appears that investors with Prosper can lose money. I assume there is some degree of protection against default or fraud but determining the precise policies was beyond the scope of my project. While we can see the repurcusions of defaults for investors in past loans, for currently active loans we are not given an assment of the risk the loan will default. I assume risk can be calculated based on the prosper score or credit score. This information can be used to better determine the best loans for investors to back. This would be an excellent topic to investigate in a future project.

For my investigation into how prospective borrowers may get the best deal on a loan from Prosper, the dat indicated that credit score was the primary determining factor behind the rate offered to the borrower (the correlation coefficient was -0.6682872). However there are likely other factors at play as well in determining what rates are offered to borrowers. I used a ggpairs chart with features in Prosper data to try to determine other relationships and I see that loan amount and available credit are also factors, while borrower state of residence and occupation were not. There is guesswork and intuition involved in this approach, so I would interested in using a more precis approach that can narrow down and rank the features. I think using machine learning principal component analysis would be an excellent for this question. There is a separate but related question I would like to answer regarding how Prosper evaluates loan applicants. While most institutions us credit score, Prosper assigns each loan a rating called Prosper Score. These metrics are very similar in that they both evaluate the applicants credit history and personal finances, but I would be interested in investigating to see how they are different. What factors does Prosper consider in evaluating applicants that credit agencies do not? Do people with certain credit history benefit more from using Prosper versus a tradional bank?

Finally, for the plot showing the trends of the Prosper business over time, where I identified that prosper is seeing growth in medium-score loans, I would be interested in discussing my assessment with a person fron the company to see if it is accurate. Did prosper set out thinking this niche would be their ideal market, or did this success develop over time. I would not be surprised to learn that Prosper initially targeted smaller loans with it's service, because median loan amount has steadily risen since 2009 and Prosper raised minimum loan threshold and maximum loan threshold in 2011 and 2013 respectivley. In studying the overall dynamics of Prosper I would be interested to learn more about the demographics of the borrowers, but I agree that this non-financial and credit related information should be shielded to prevent bias in lending to borrowers.

'''