PROSPER LOAN DATA

COMMUNICATE DATA FINDINGS

Zhuo (Lydia) Fang

March 25, 2019

Prosper Loan Data Analysis

BY Zhuo (Lydia) Fang

March 25, 2019

1. Data Background

Prosper Marketplace is America' s first peer-to-peer company. From the website Prosper.com, individuals can either invest in personal loans or request borrow money. Borrowers can request personal loans from $2,000 to $35,000 per loan request. Investors can evaluate borrowers based on their credit scores, ratings, histories, occupations and loan type etc. Prosper provides the loan service, collects borrower payments and distributes payments and interests back to loan investors. Therefore, prosper also takes the responsibility to verify borrowers' identities and select personal data before funding loans. Before July 2009, Prosper provided "Credit Grades" and other credit information about its prospective lenders, since July 1, 2009, Prosper created a new model that determined "Prosper Ratings" instead.

2. Current dataset and analysis plan

The current data analysis project is one of the Udacity Data Analyst Nanodegree projects. The prosper loan data was obtained from https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv, which includes the loan information from Nov.2005 to March, 2014. Using this data set, the current project intends to conduct some exploratory analysis to investigate the loan fluctuation along with time, the distribution of borrowers 'prosper ratings, the potential influence factors on loan status etc.

3. Interested Variables

There are 81 variables in the initial dataset, while in the current project, I just selected several interested variables listed below to conduct the analysis :

- LoanStatus: The status of the loans through all years, including completed, current, defaulted, chargedoff, Past due etc.

- LoanOriginationDates: The date the loan was originated.
- BorrowerAPR: The Borrower's Annual Percentage Rate (APR) for the loan.
- BorrowerRate: The Borrower's interest rate for this loan.
- Term: The length of the loan expressed in months.
- LoanOriginalAmount: The origination amount of the loan.
- ProserPerRating (Alpha): The Prosper Rating assigned at the time the listing was created between AA - HR. Applicable for loans originated after July 2009.
- ProsperScore: A custom risk score built using historical Prosper data. The score ranges from 1-10, with 10 being the best, or lowest risk score. Applicable for loans originated after July 2009.
- ListingCategory: The category of the listing that the borrower selected when posting their listing.
- StatedMonthlyIncome: The monthly income the borrower stated at the time the listing was created.
- EmploymentStatus: The employment status of the borrower at the time they posted the listing.
- CreditScoreRangeLower: The lower value representing the range of the borrower's credit score as provided by a consumer credit rating agency.
- CreditScoreRangeUpper: The upper value representing the range of the borrower's credit score as provided by a consumer credit rating agency.
- DebtToIncomeRatio: The debt to income ratio of the borrower at the time the credit profile was pulled. This value is Null if the debt to income ratio is not available. This value is capped at 10.01 (any debt to income ratio larger than 1000% will be returned as 1001%).
- MonthlyLoanPayment: The scheduled monthly loan payment.
- TotalProsperLoans: Number of Prosper loans the borrower at the time they created this listing. This value will be null if the borrower had no prior loans.
- DelinquenciesLast7Years: Number of delinquencies in the past 7 years at the time the credit profile was pulled.

4. Data Wrangling:

Before the analysis, I did some data wrangling to make the variables more clear or more consistent with my analysis plan:

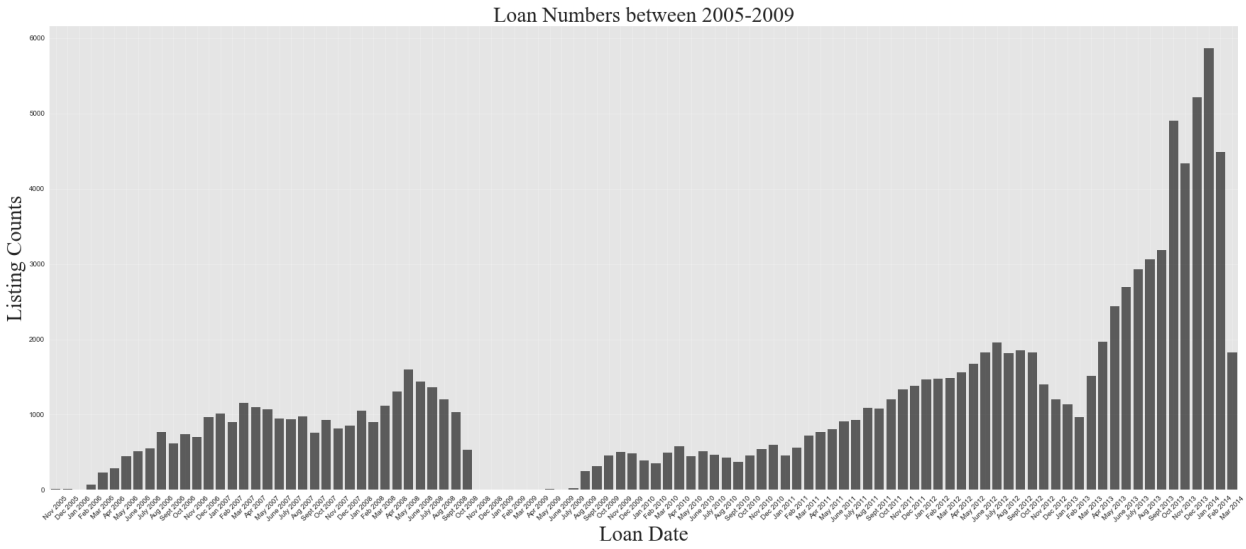
- Reduced the data size by dropping variables that would not be included in the further analyses.
- According to Prosper policy, I re-organized the LoanStatus into two categories (Normal vs High Risk), and created a new variable "LoanStatus_New".
- Categorized the loan dates into two stages (Before July 2009 vs. After July 2009), and created a new variable "Stage"
- Reorganized the LoanOraiginationDates into Month and Year, and created a new variable "LoanDate".
- The initial ListingCategory was numeric format, which was not ideal for the data analysis, I created a new variable "loantype",which was the string type representing the real loan type.
- Using the lower and upper credit score range to calculate the average credit score, and created a new variable "CreditScore".

5. Exploratory Analysis

5.1.Univariate Exploration

First, I did some simple univariate analyses to explore the general distribution of some important variables.

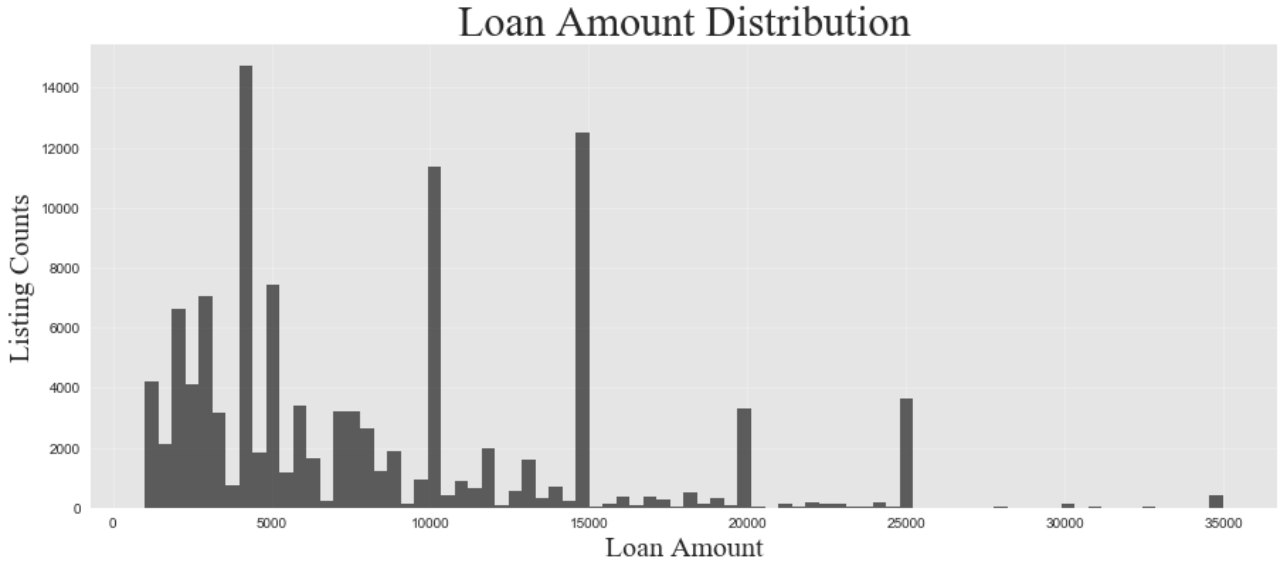
5.1.1 Do more people choose Prosper?



As shown in the figure above, we can see the loan number is steadily increasing along with the time, indicating that increased number of borrowers and investors start using Prosper to make loan request or invest.

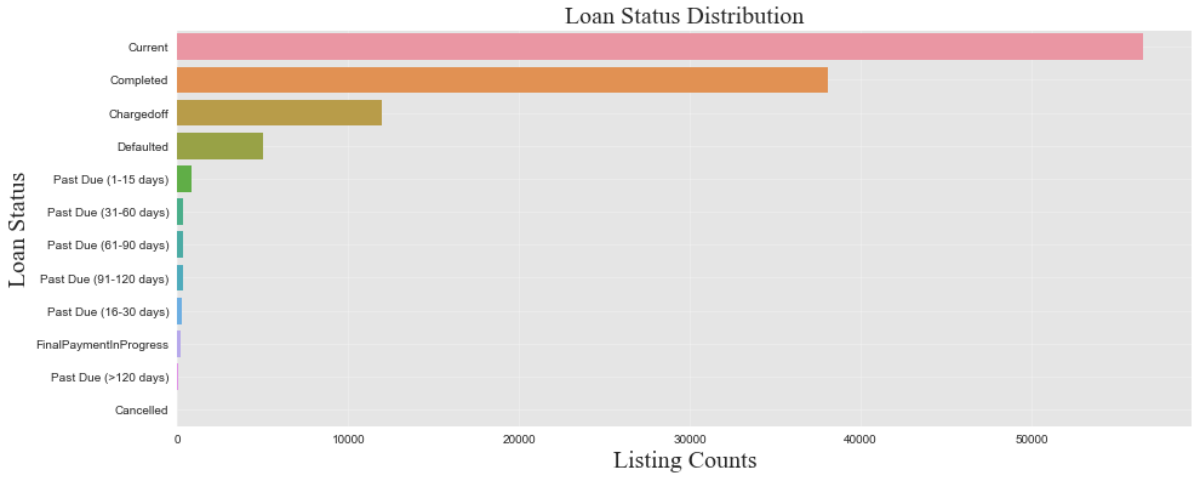
5.1.2 What is the amount of money people need?

As shown in the figure below, most of the loans are in the amount below $20,000. In fact, $5000, $1000, and $15,000 were the three most requested amounts, which were reasonable values for a personal loan.



5.1.3. How does the loan status look like?

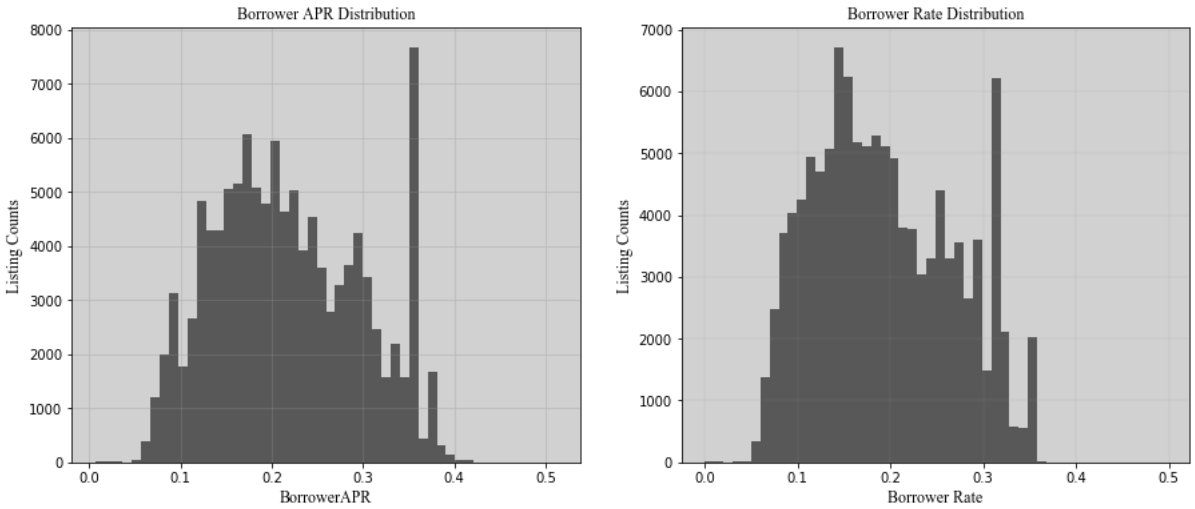
It is important to know the loan status, which can reflect whether the borrower verification and credit information work well for the online loan platform.



As shown in the figure above, we can see the majority of loans are in a normal status (Current & Completed). However, there were still a certain amount of loans were in a bad situation (Chargedoff & Defaulted), which reminded the Prosper company and investors may need more effective ways to evaluate the borrowers 'credit and payment abilities.

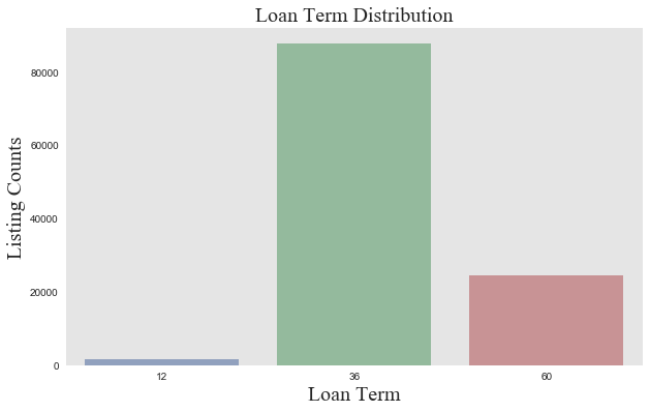
5.1.4 What is the interest rate when people borrow money?

BorrowerAPR and BorrowerRate are two important factors that affect both borrowers and investors, which can provide them more useful information to make their loan requests and invest plans in a more economical way.



As shown in the figure above, we can see that the distribution patterns of BorrowerAPR and BorrowerRate were very similar, while the BorrowerRate per loan was higher than the annual rate in general, which makes sense.

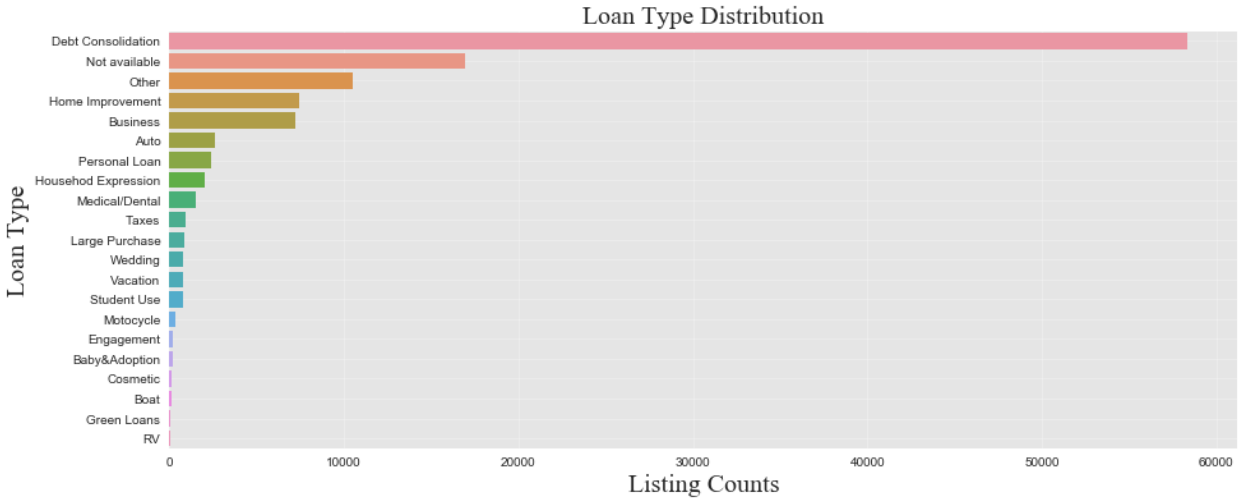
5.1.5 How long do people borrow the money?

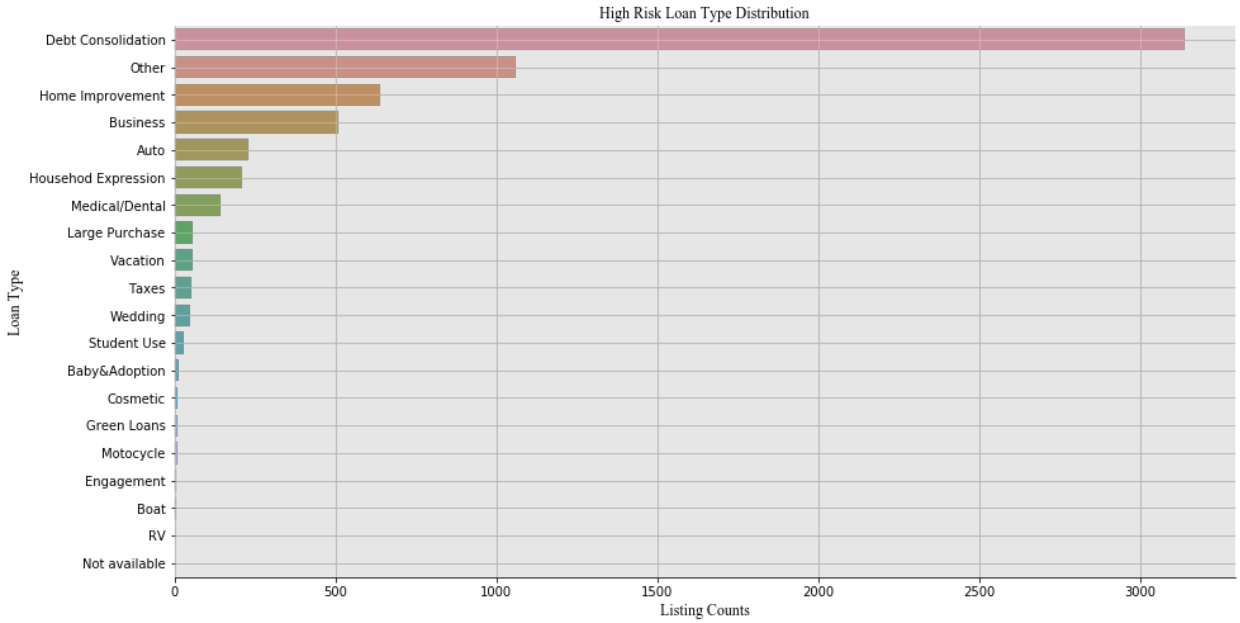


As shown in the figure above, 36 months loan term was most people' s choice, which is easy to understand. 12 months would give borrowers higher payment pressure, while 60 months results in higher interest, therefore 36 months term is the most reasonable option for most people.

5.1.6 What do people borrow money for?

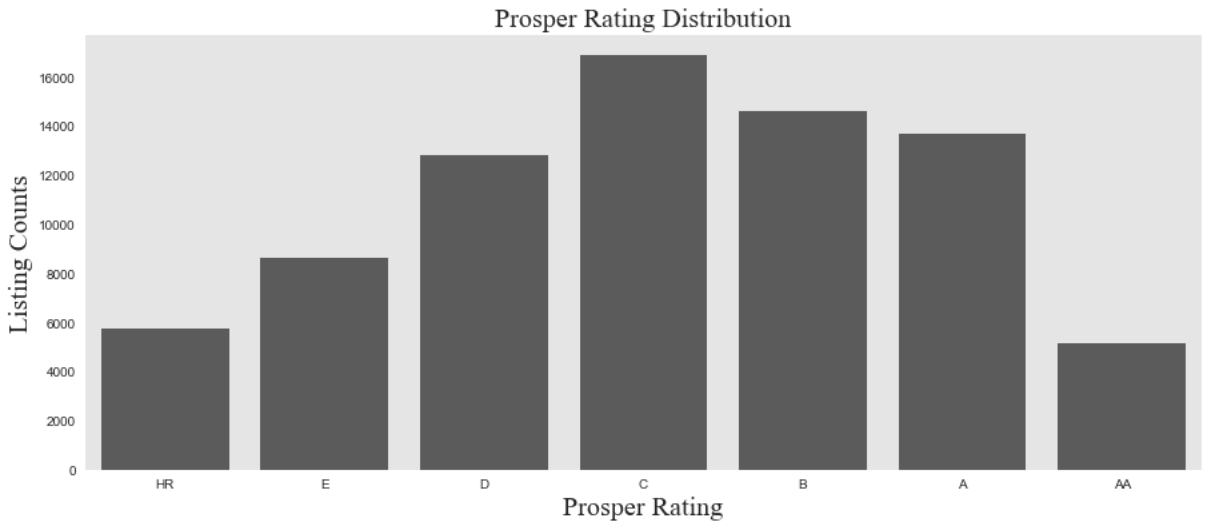
I am interested in knowing why people need loan. From the figure below, it is surprised to find that most people borrow money for debt. It seems that they are using new debt to pay the old debt? Actually, from the second figure listed below, we can see among all high risk loans, the most listed type is debt consolidation. But they were still approved for the loan request. Interesting!





5.1.7 What are borrowers' prosper ratings?

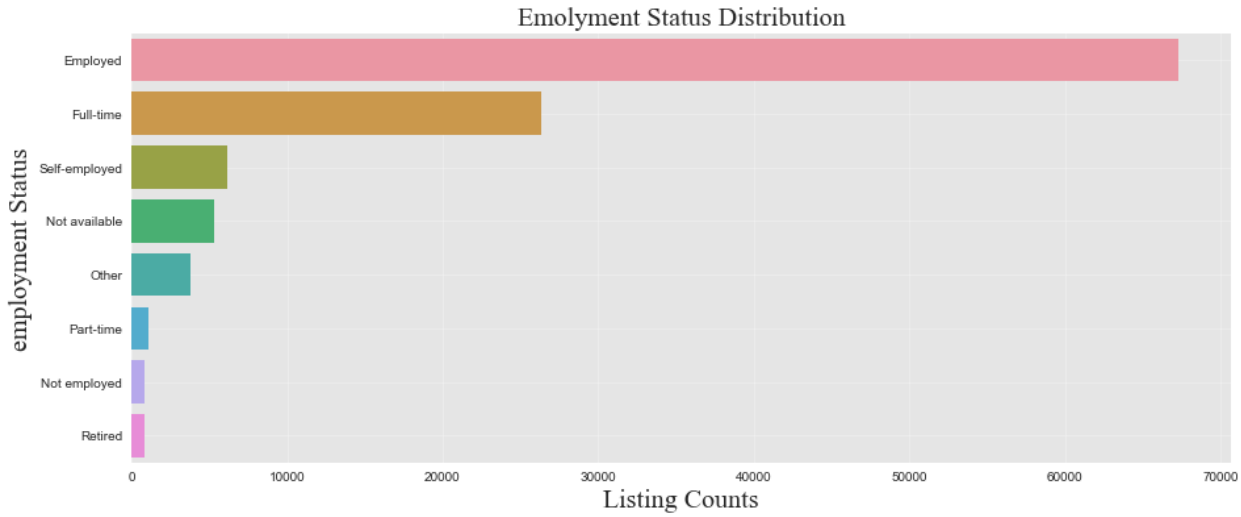
For the Prosper platform and investors, borrowers' prosper rating might be the most important factor to evaluate borrowers 'credit situation and make the decision to approve the loan or not. Therefore, it is useful to know the prosper rating distribution overall. As shown in the figure below, the prosper rating distribution is mostly like a normal distribution that most of borrowers are at the middle level (C), while the borrowers at two tails (HR & AA) are less.



5.1.8 How many times do people loan from Prosper?

Do people borrow money repeatedly? This is an interesting question. Surprisingly, most people borrowed from Prosper more than once. The figure below is a right skewed distribution.

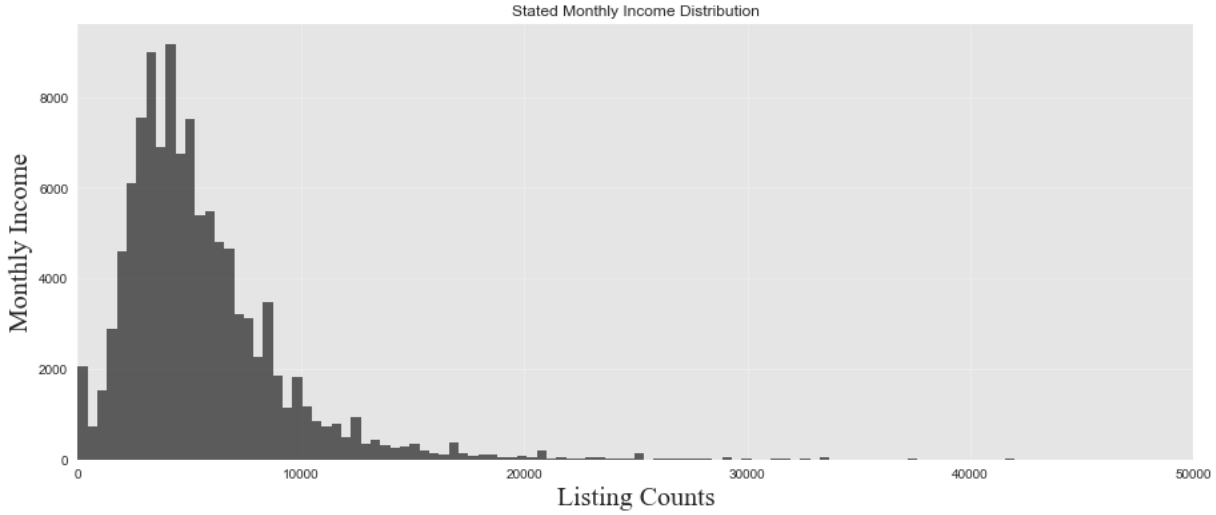
5.1.9 What is the employment status of Borrowers?



Well, as expected, the majority of borrowers are employed, especially the full-time type. Self-employed borrowers number follows after. That makes sense because, generally speaking, employed individuals have higher payment ability during the assessment.

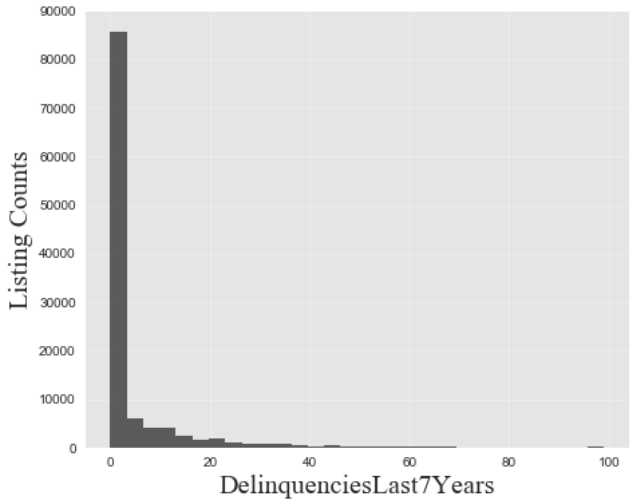
5.1.10 How much do borrowers earn?

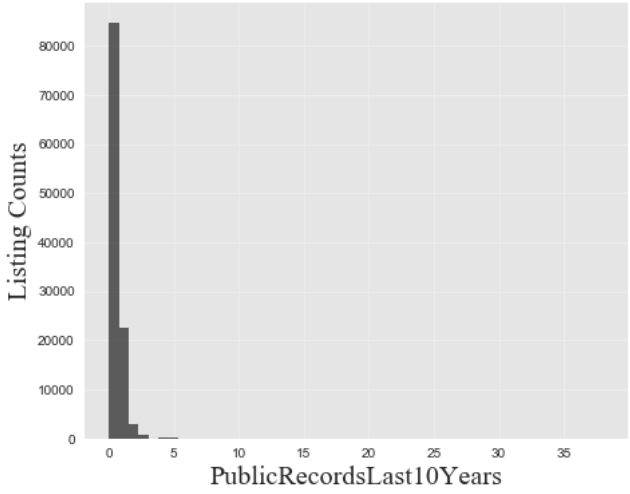
As illustrated in the figure below, the median monthly income of borrowers is around $4500 (the median income is $ 4666). Very few exceed $20,000. Well, if someone can earn more than $20,000 per month, the probability that they need a loan is relatively smaller.



5.1.11 Personal histories

Another factor influencing personal loan is borrower’ s history that associated with their credit situation, such as Delinquencies and Public Records.





As shown in the figure above, we can see that Prosper applies a strict criteria of negative personal histories when approve the loans, especially the public records. Most of borrowers should have less than 1. This suggests that if the individuals had some negative personal histories that impacted their credits, it would be difficult to get approved the loan request by Prosper.

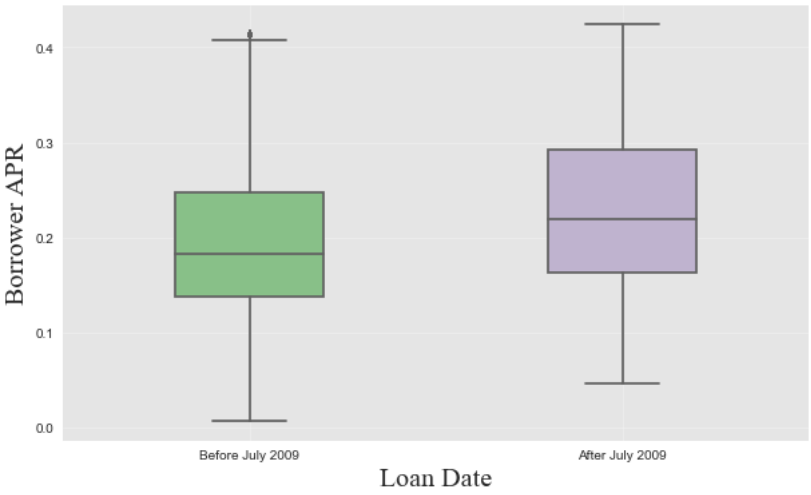
5.2 Multivariate Exploration

In this section, I would like to explore the relationships between different variables, so that we can have a better understanding of the current loan data set.

5.2.1 What changed after July 01 2009?

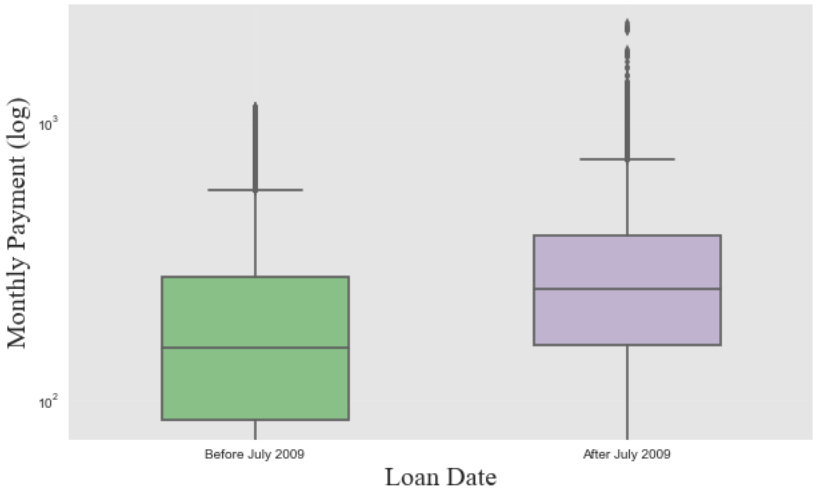
Since Prosper provided a new rating method from July 01 2009, so firstly, I am interested to know is there any difference of loan situation before and after this date.

First, what are the overall loan amounts before and after July 2009? As shown in the right figure , we can see that after July 2009, the loan amount is obviously higher than that before July 2009, no matter the median amount or the Max amount. One of the potential reason is that both borrowers and investor are more comfortable with the platform and more likely to choose it.

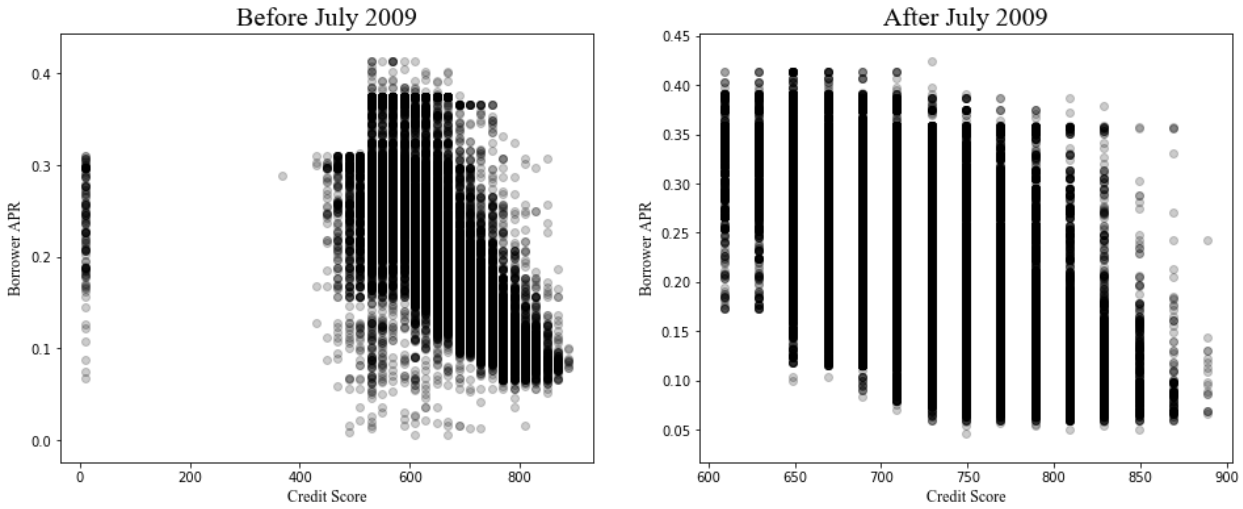


Second, how about the annual percentage rate (APR) for borrowers? As shown in the left figure, we can see that after July 2009, the APR is also higher than that before July 2009. This could be associated with the increased loan amount, or associated with the bank interests. Additionally, after July 2009, Prosper has applied a more stringent criteria in verifying the borrowers, which could also induce the APR increase.

Third, the monthly loan payment. Not surprisingly, after July 2009, borrowers' monthly payments increased as well, as shown in the figure below. Along with the increased loan amount and APR, the increased monthly payment could be expected.



Fourth, what are the relationships between borrowers' credit score and the APR?



From the figure shown above, generally, there is a negative correlation between Credit Score and APR, indicating lower credit level borrowers would face higher annual percentage rate, which is reasonable. Additionally, we can see before July 2009, there were some borrowers with very low credit levels, which might reflect that at very early stage, Prosper had applied a liberal credit criteria to attract more borrowers. When the platform becomes more mature, they started applying more stringent credit score criteria (e.g. the minimum credit score should not be less than 600).

5.2.2 How does the loan situation change along with time?

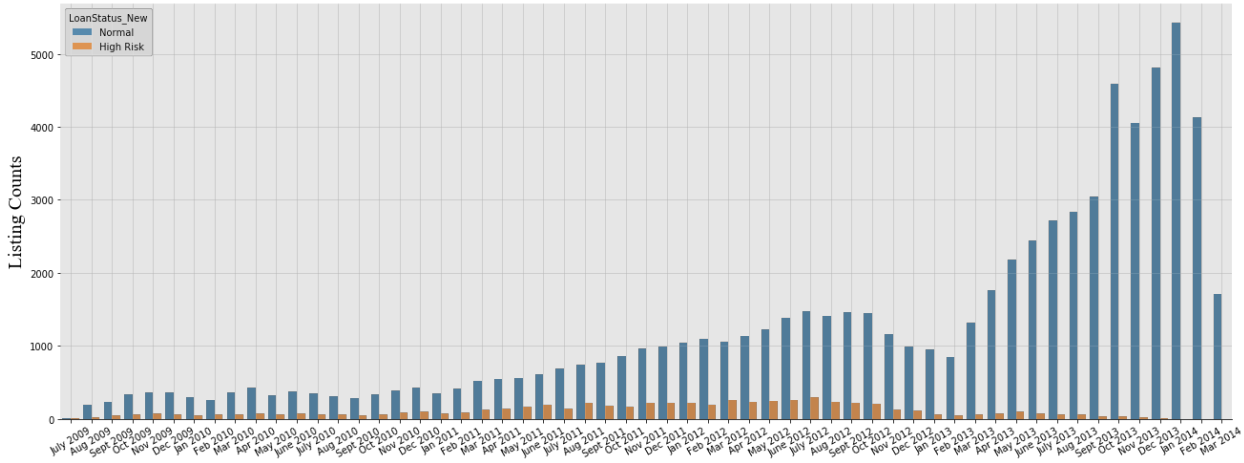
Another question that I am interested in is the loan change with time series, which can give us a better idea how the platform works.

- *The normal loan and high risk loan*

Prosper has divided the loan status into 11 categories. To make the analysis easier and the results more clear, I dived the loan status into two categories (Normal & High Risk) according to their policy:

Normal: Completed, Current, FinalPaymentInProgress, Past Due (1-15 days), Past Due (16-30 days).

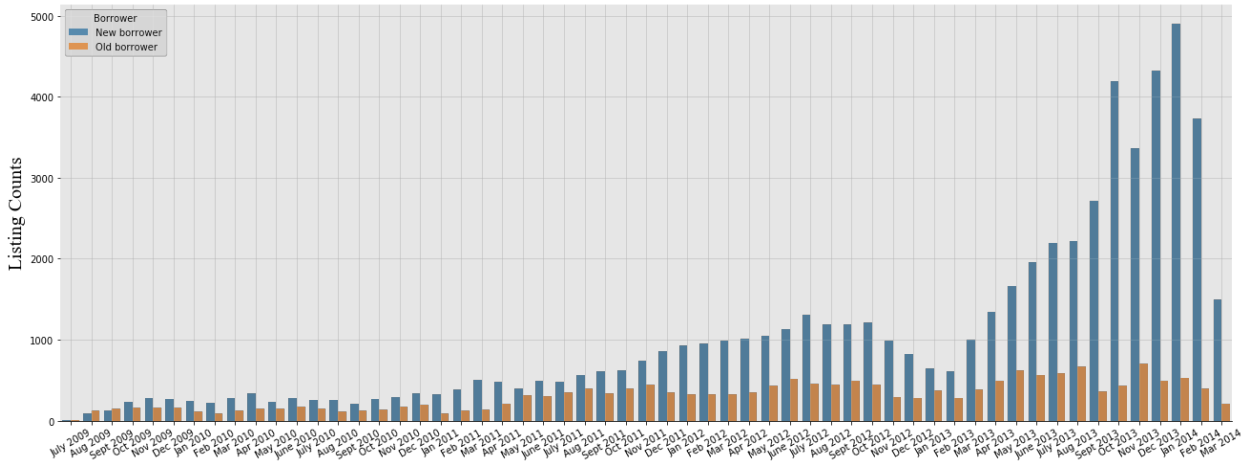
High Risk: Defaulted, Chargedoff,Past Due (31-60 days), Past Due (61-90 days), Past Due (91-120 days), Past Due (> 120 days).



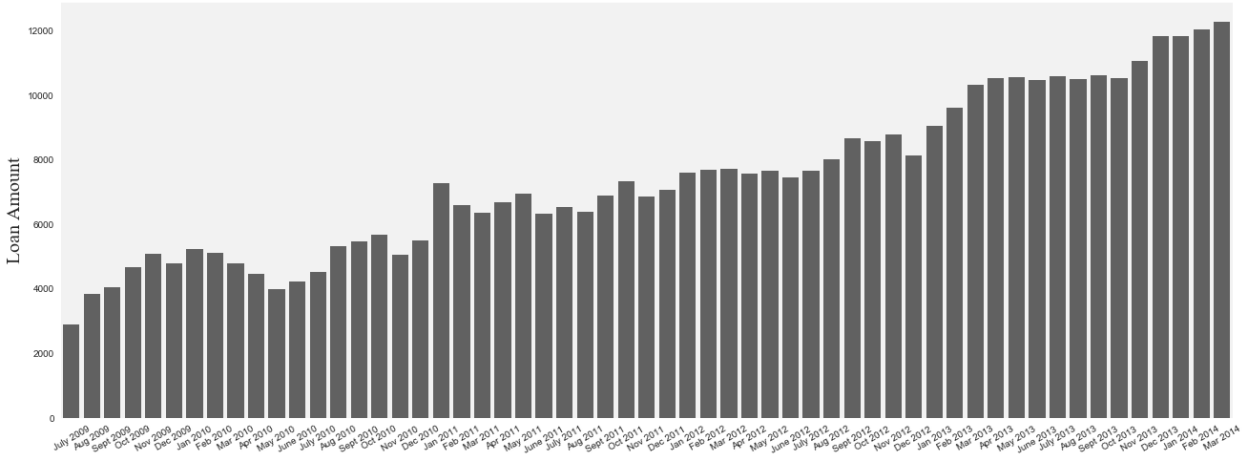
As shown in the figure above, although there was slight fluctuation, the number of Normal loans are increased steadily along with the time, while the number of High Risk loan was decreased with the time. This indicates that after years of operation and experience, Prosper has made some progress in risk management, and decreased the investment failure.

- *New customers and repeat customers*

As we explored in the previous section, most of customers have borrowed money from Prosper more than once. As we can see in the figure below. Along with the time, the number of new customers is increased steadily, while the number of old customers remains stable with slight fluctuation between April 2011 to March 2014.

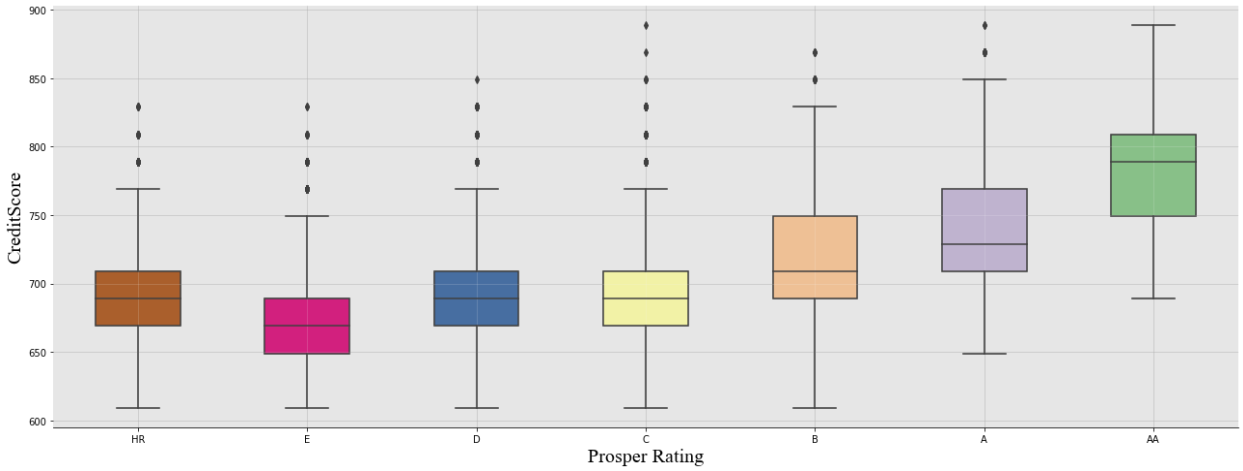


- Loan Amount along with time



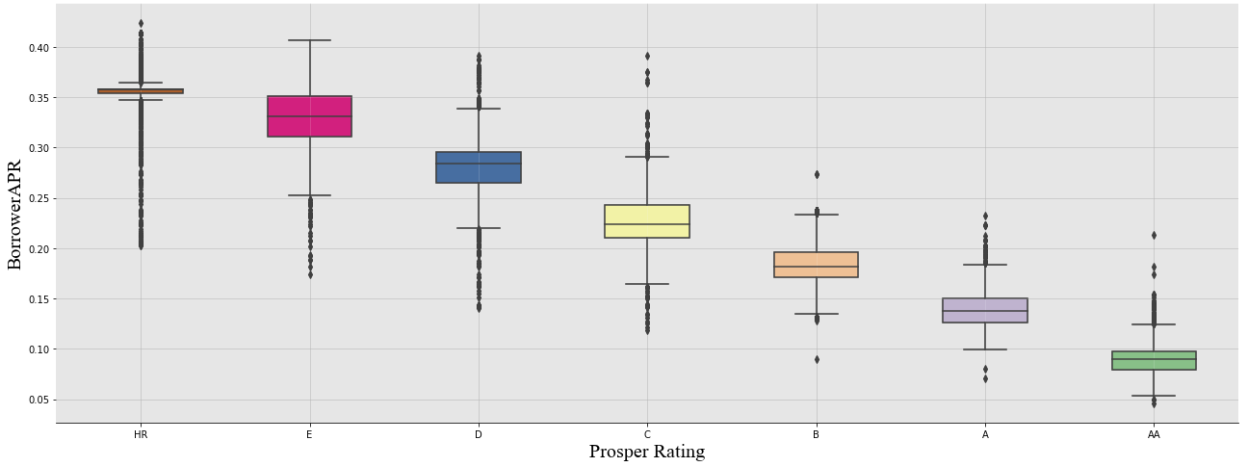
As expected, the average loan amount has been increased steadily along with time.

5.2.3 Relationship between Prosper Rating and Credit Score



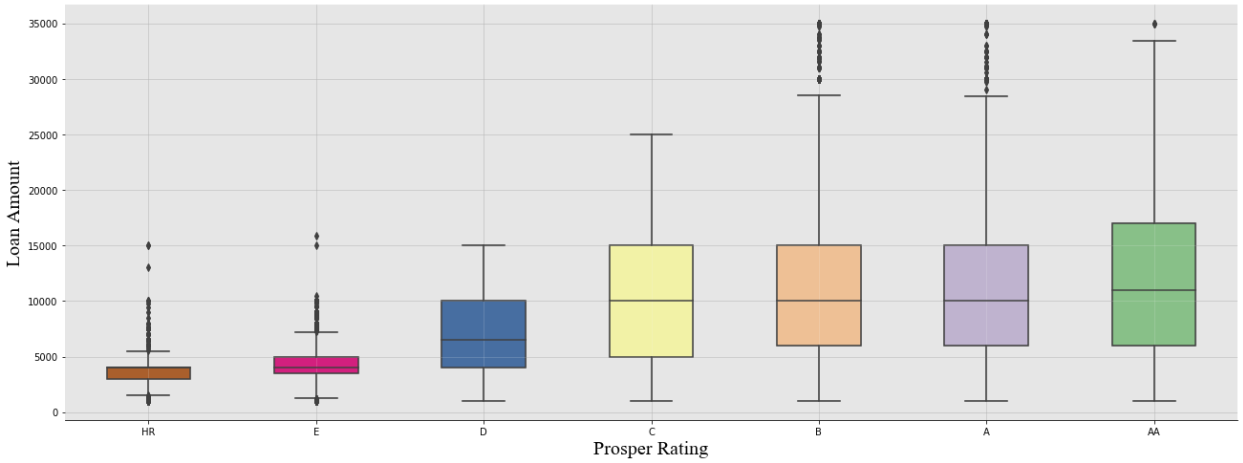
As shown in the box plot above, we can see that the higher prosper rating level is associated with higher credit score (e.g., C, B, A ,AA), but it is not an absolutely linear relationship in the lower prosper rating level (e.g., HR, E D).

5.2.4 *Relationship between Prosper Rating and Borrower APR*



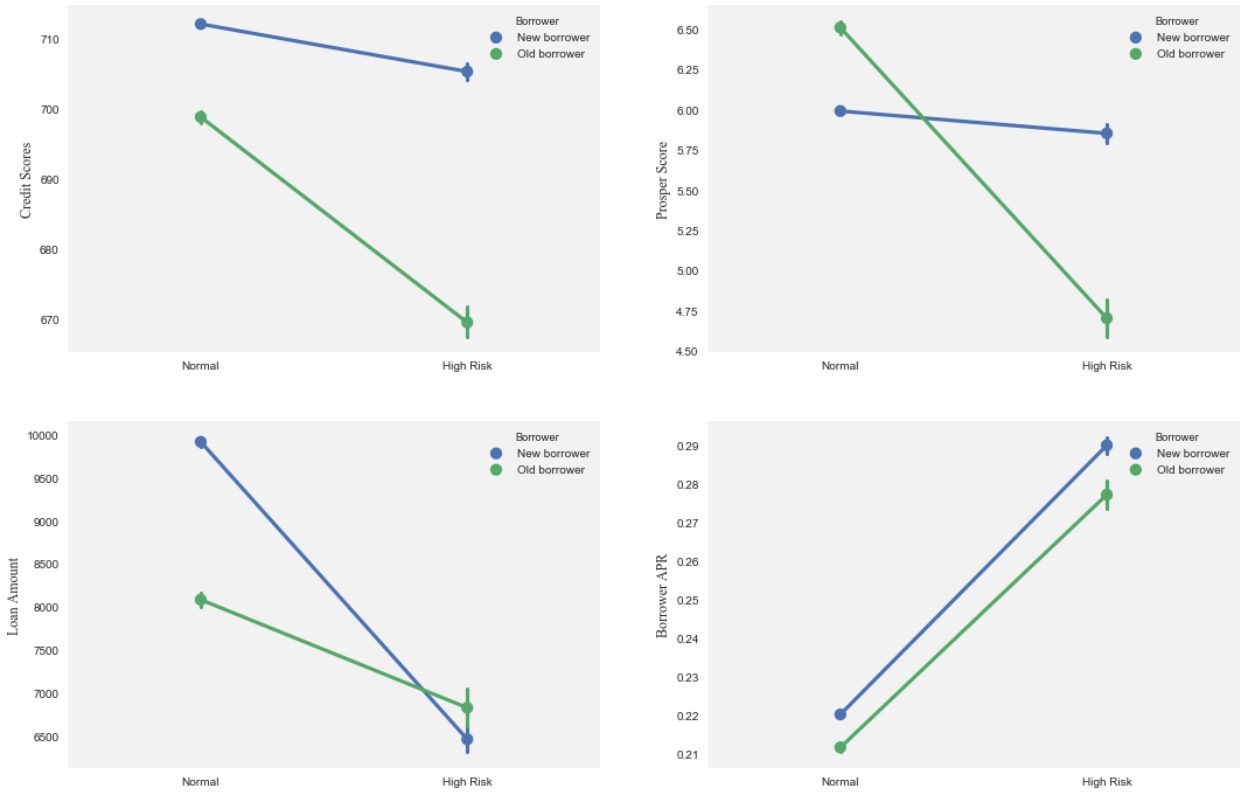
Well, the above figure shows a clear negative relationship between Borrower APR and Prosper Rating level. Lower Prosper Rating is associated with higher APR. Make sense!

5.2.5 Relationship between Prosper Rating and Loan Amount



According to the above figure, it seems that when the borrower' s Prosper Rating level reaches a certain level (re; B-AA) their borrow abilities will not differ a lot.

5.2.6 Interaction effects between borrower status (Old vs. New) and Loan Status (Normal vs. High)



As shown in the figure above, we can see there are clear interaction effects between borrower status and loan status on several variables:

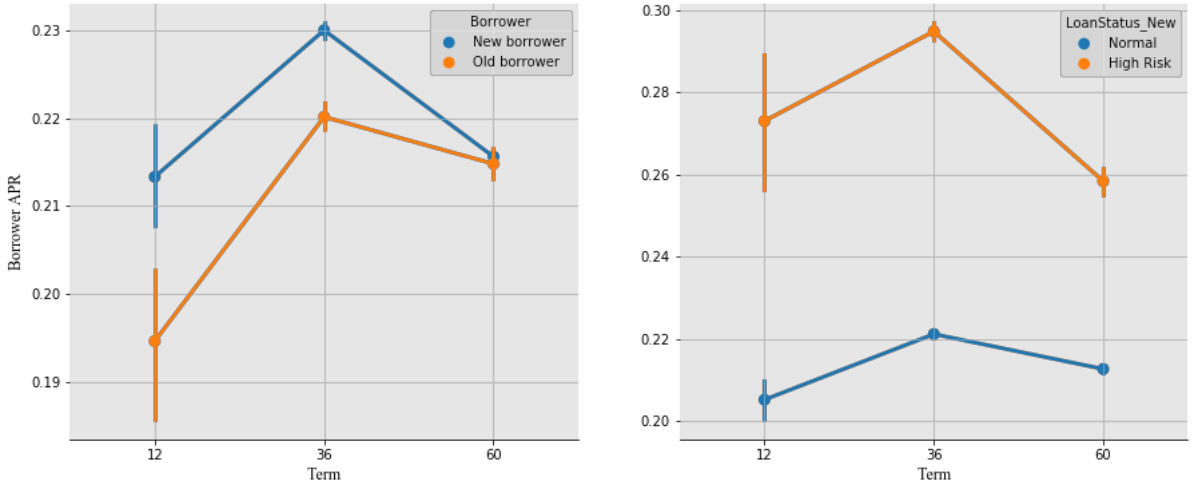
- Credit Score: Overall, the normal loan borrowers have higher credit score than the High risk loan borrowers, and new borrowers have higher credit score than old borrowers. However, we can see the credit score difference between old and new borrowers is larger for those high risk loans. There is an interaction effect trend.

- Prosper Score: Similarly, for both old borrowers and new borrowers, the normal loan borrowers have higher Prosper score than the High risk loan borrowers. But there is a clear interaction effect. Specifically, for those normal loans, the new borrowers have lower prosper score than the old borrowers, while it is opposite for the high risk loans. This is correct. Because for the normal loans, new borrowers have no loan history with Prosper, so there is no Prosper credit information. However, for those high risk loans, the repeated borrowers have Prosper histories, but negative (e.g. risk loan). These negative loan histories will impact old borrowers' prosper scores.

- Loan Amount: The interaction pattern for loan amount is pretty similar with that for Prosper Score, and the possible reason could be the same that new customers have no history, but the negative history would impact more.

- Borrower APR: The APR for new Borrowers is always higher than old borrowers regardless of the loan status.

5.2.7 Interaction effects between Loan Term and borrower status as well as loan status on the borrower APR.



- Term & Borrower Status: From the figure shown above (left), we can see that overall, the new borrowers have higher APR than old borrowers, but the difference is decreasing when the loan term is increasing. Especially if the loan term is 60 months, the APR for new and old borrowers are almost the same.

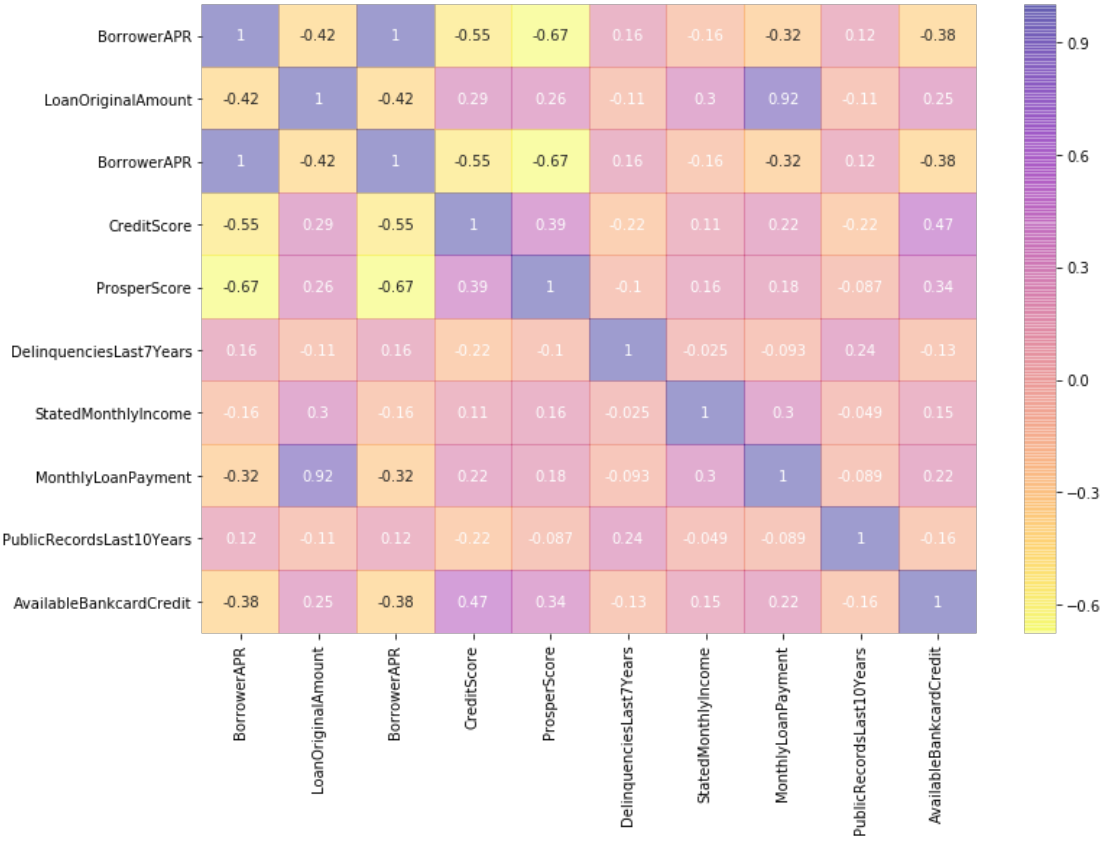
- Term & Loan Status: From the figure shown above (right), we can see that for those high risk loans, the APR is much more higher than that of normal loans, regardless of the loan term.

- Overall, 36 months term loan has the highest APR.

5.3 Correlational exploration

At the end, I would like to conduct some correlational exploration to investigate the correlational relationship between factors.

5.3.1 Correlational matrix

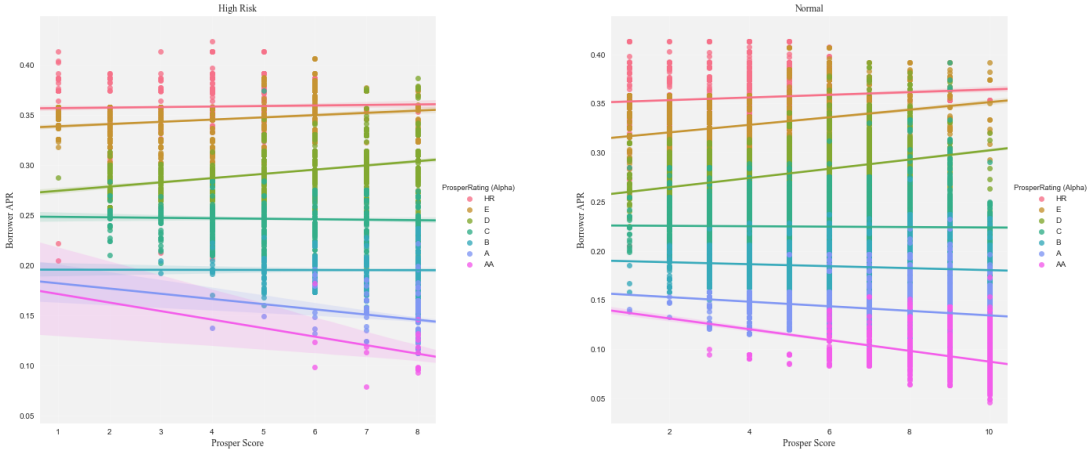


As illustrated in the figure above, the strongest positive correlation is between LoanOriginalAmount and MonthlyLoanPayment, and the strongest negative correlation is between prosper score and Borrower APR.

5.3.2 Does prosper rating level modulate the relationship between different factor?

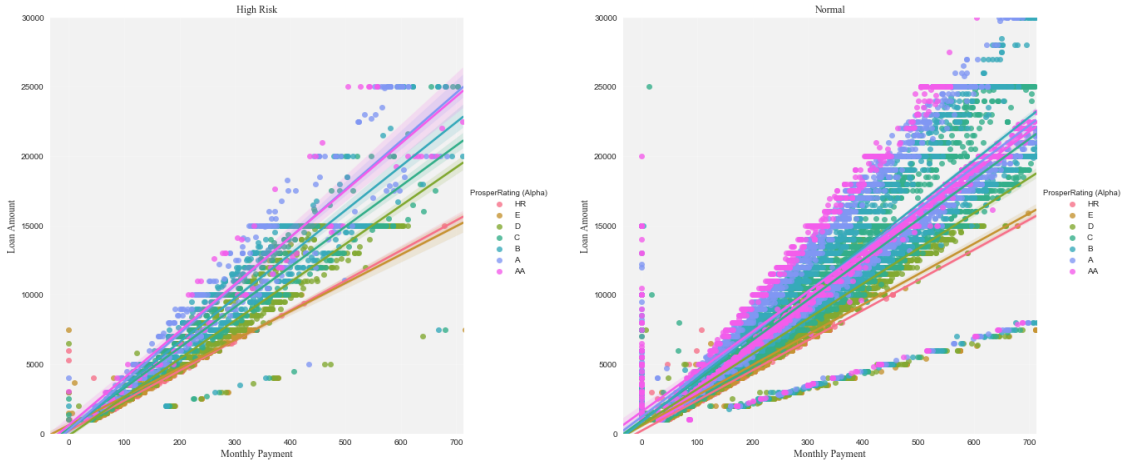
Finally, I would like to investigate the relationship between the prosper score and Borrower APR and the relationship between the monthly payment and loan amount in the normal loans and high risk loans. But I also want to see if the prosper rating level will modulate the correlation pattern, so I made scatter plots for different prosper ratings.

- Correlation between Prosper Score and Borrower APR



As shown in the figure above, we can see the negative correlation between prosper score and borrower APR for both normal and high risk loans. But we can see, compared to the high risk loans, in the normal loans, a larger number of borrowers have higher prosper rating and lower APR at the bottom of the scatter plot.

- Correlation between Monthly payment and Loan amount



What we can see form the above figure is that the positive correlation patterns between monthly payment and loan amount are similar for both normal and high risk loans.

6. Summary

- The customer number of Prosper and the approved loan amount are increasing steadily along with time.
- After years of operation and experience, the failure loan number has been decreased.
- There is a certain number of repeated customers borrowed money from Prosper several times.
- Most people borrow money for debt consolidation, and this type of loan also has the largest number in the high risk loans.
- Employed individuals with monthly income around $ 4500 are the major population who borrow money from Prosper.
- Borrowers' credit situation (i.e. credit score & Prosper score) is the most important factors that influencing lots of loan features, such as APR, Rate, loan amounts, loan status etc.
- Borrowers' past personal negative history (e.g. Delinquencies and Public Records) will impact their loan application.

- Repeated borrowers and new borrowers have different APR level and loan amount, which depends on their past credit history.
- Although with the highest APR, most of borrowers still select the 36 Term loan.
- Not surprisingly, if you borrow more, you need pay more every month.

7. Challenge and limitation

- The biggest challenge of the current project for me is that I am not major in finance or related field. Therefore, I found it is difficult to define the research question. I have tried my best to explore the data, there might be some questions not interesting, but I am glad to try the new topic and I would like to do more in future.

- The current project is exploratory, so I did not do any statistical analyses. We can grasp the patterns or relationships through the visualization, but we cannot make a solid conclusion.

- All the analyses are exploratory, no causal direction can be determined.