

## *Prosper Loan Dataset*

### **1. Introduction**

This report will explore a dataset and provide a summary of its main characteristics. Subsequently, visualization of the prosper loan dataset will also be done. The 113,937 loans in this data collection have 81 characteristics, such as Borrower APR, lender yield, estimated return, and many more. However, this report will focus on will not utilize the 81 characteristics but focus on 17 of them.

### **2. Assessing the Data**

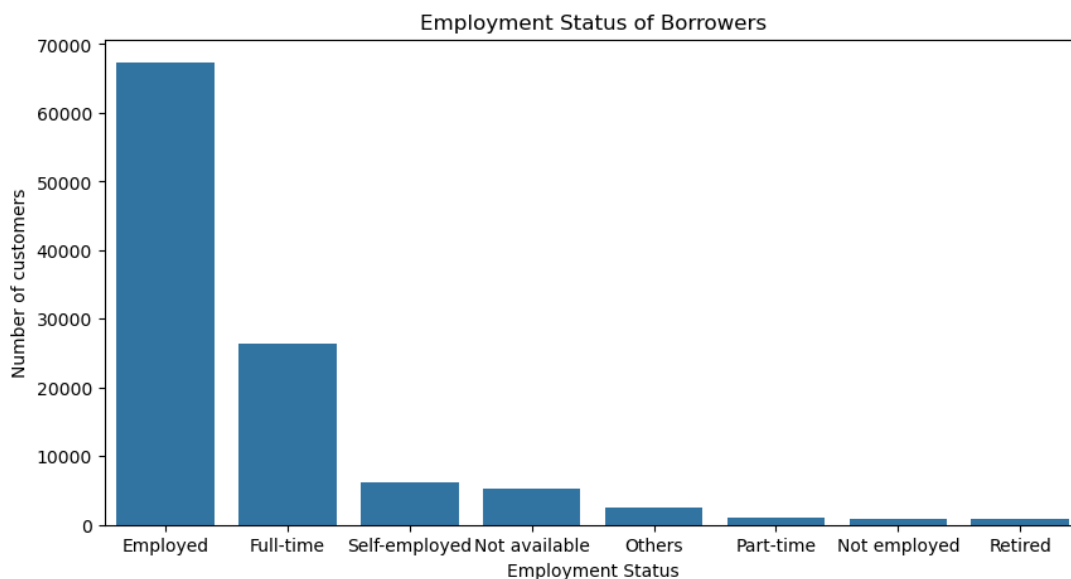
The dataset was assessed visually and programmatically to check for missing values, unique variables, and duplicated values etc. The following steps were taken to clean the dataset.

- ‘Other’ variable in employment status was renamed to ‘Others’
- Null values in prosper score and borrower APR was replaced with ‘0’.
- Prosper score column was converted from float data type to integer since the rating is between 1 – 10.
- Original loan date was also converted from object data type to date time.

### **3. Exploratory Analysis**

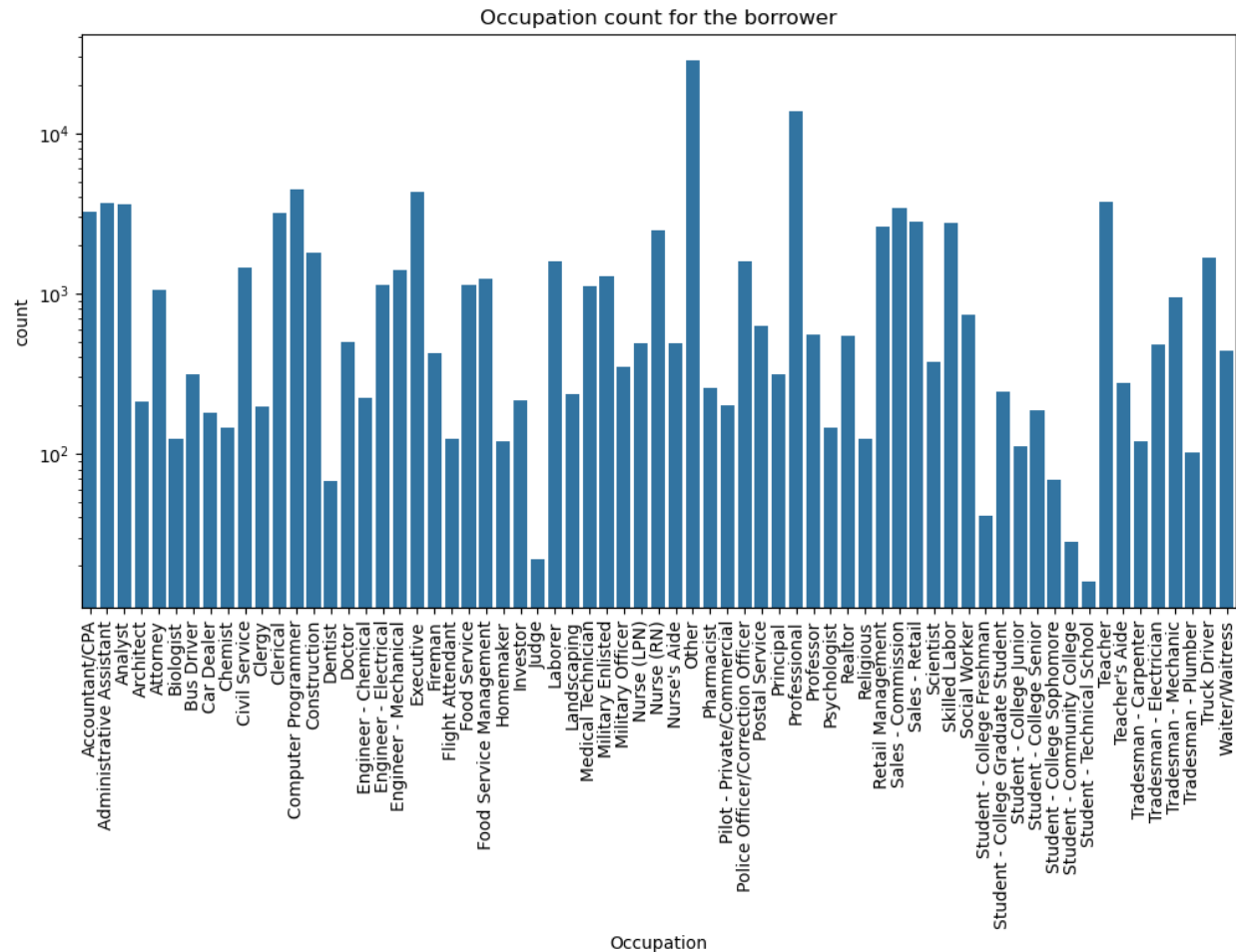
This session aims to visualize the key characteristics of the dataset to search for insights.

#### **3.1 Univariate Exploration**



## Prosper Loan Dataset

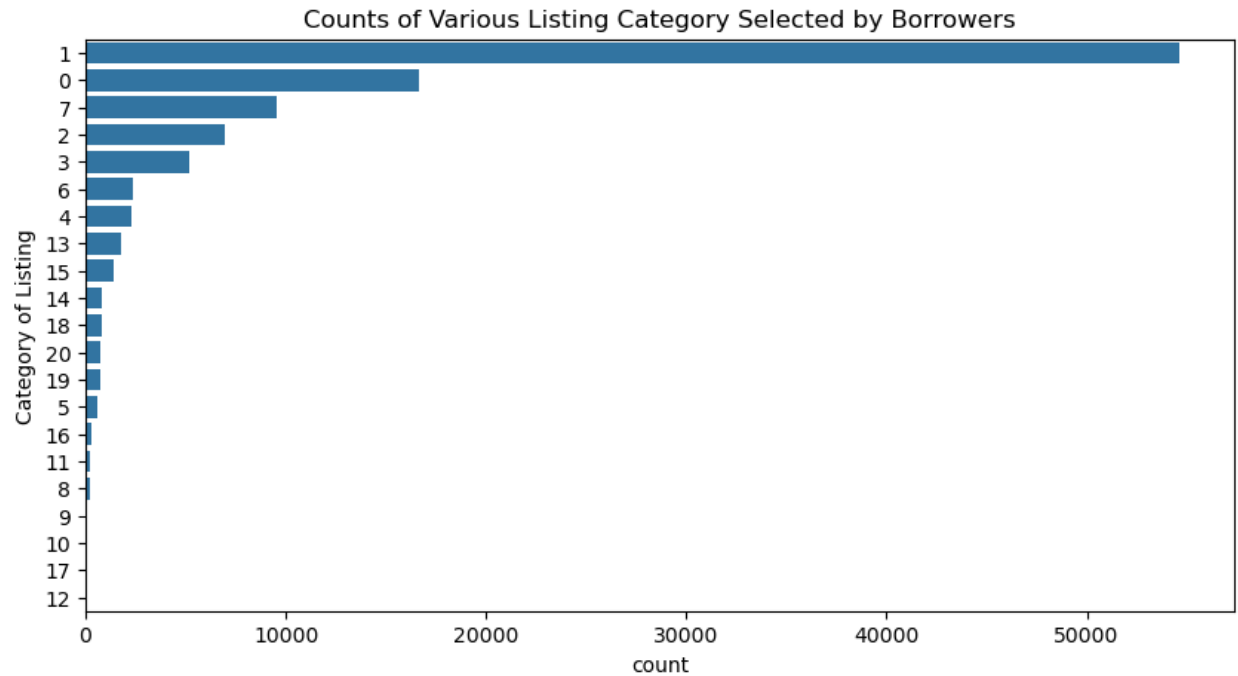
The plot above shows the different employment status categories. Most clients in the loan dataset were employed with the least being retired.



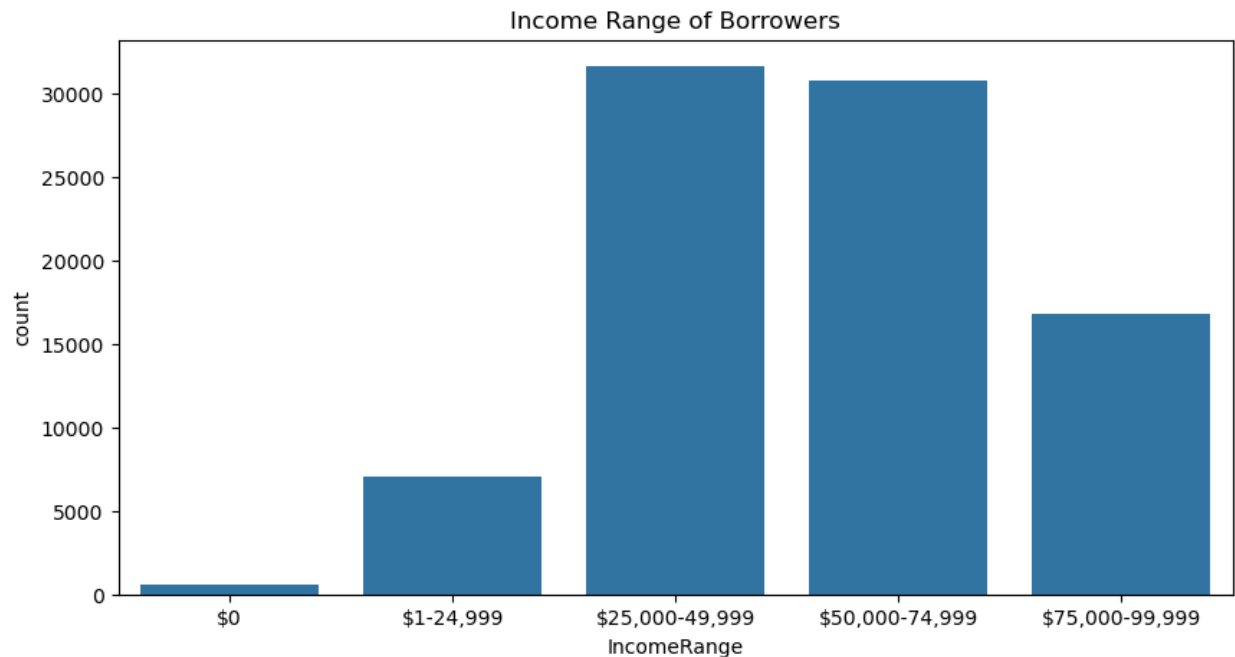
Similarly, in the different occupations selected by borrowers, most were categorized as others with professionals being in second place. Students in the technical school have the least frequency.

The plot below shows the count of listing categories selected by borrowers. The listing category ranges from (0 – 20). 0 - Not Available, 1 - Debt Consolidation, 2 - Home Improvement, 3 - Business, 4 - Personal Loan, 5 - Student Use, 6 - Auto, 7- Other, 8 – Baby & Adoption, 9 - Boat, 10 - Cosmetic Procedure, 11 - Engagement Ring, 12 - Green Loans, 13 - Household Expenses, 14 - Large Purchases, 15 - Medical/Dental, 16 - Motorcycle, 17 - RV, 18 - Taxes, 19 - Vacation, 20 - Wedding Loans. Listing category 1 had the highest frequency, indicating that most borrowers needed the loan for debt consolidation.

## Prosper Loan Dataset

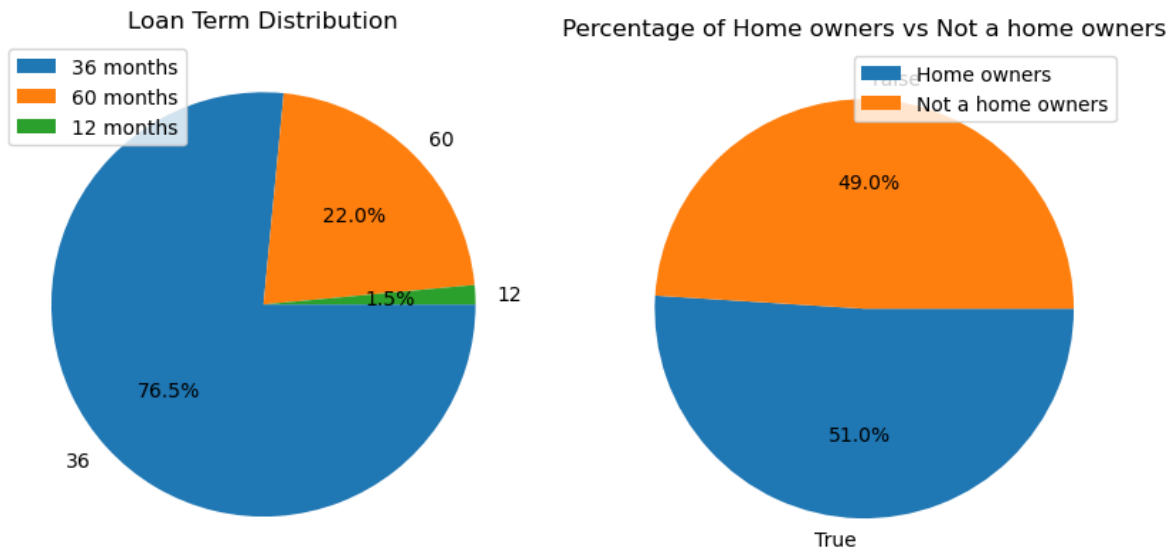


The income range of borrowers also varies significantly. Most borrowers' income ranges from \$25,000 to \$49,999, followed closely by \$50,000 to \$74,999.

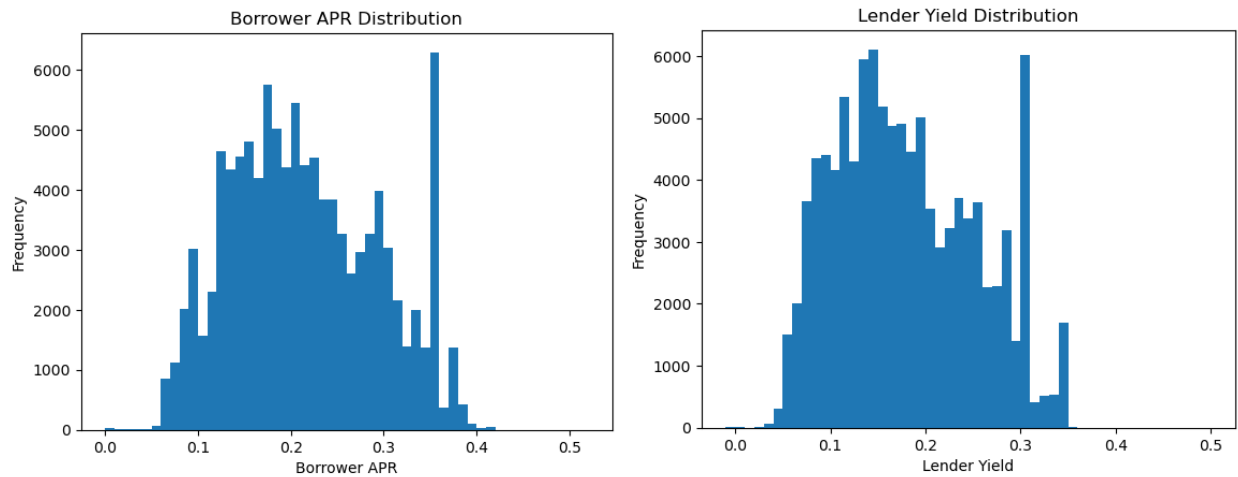


In this dataset, three loan terms are available. 12 months, 36 months, and 60 months. 76.7% of the loan term was for 36 months with 12 months being the least. Similarly, 51% of borrowers were homeowners as seen below.

## Prosper Loan Dataset

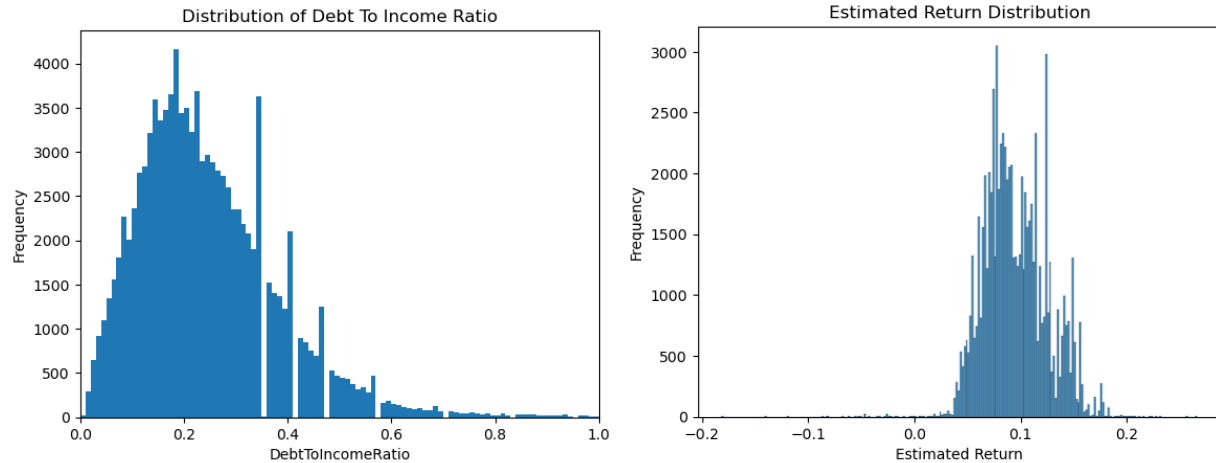


Borrower APR is very important in any loan or credit facility. This indicates the Borrower's Annual Percentage Rate (APR) for the loan. The histogram below shows the distribution of borrower APR in the dataset. The plot shows a normal distribution.

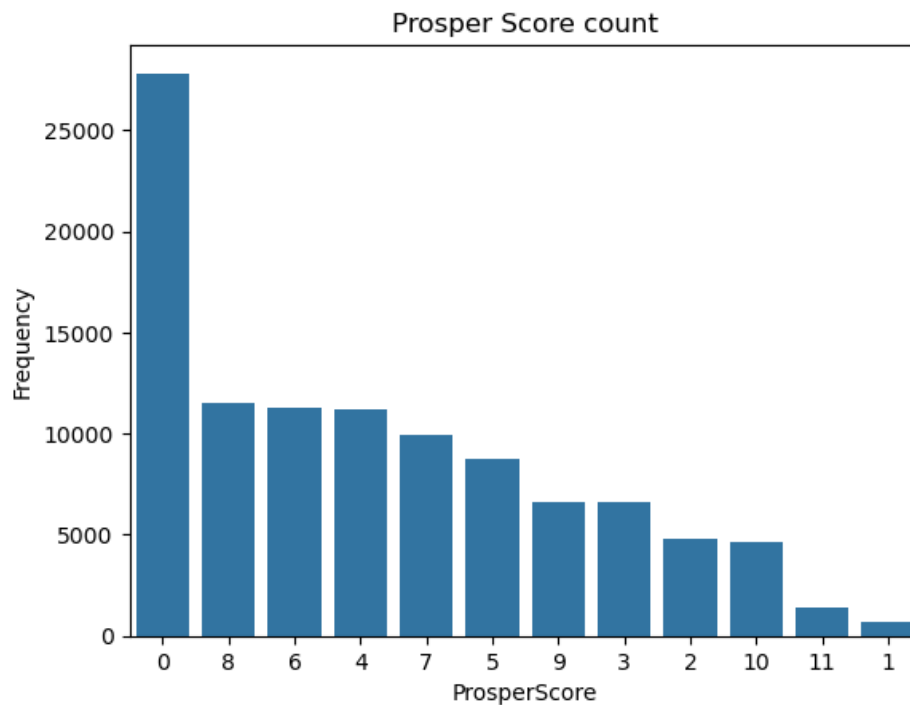


Similarly, the distribution of lender yield was also visualized, indicating a normal distribution as well.

## *Prosper Loan Dataset*



However, the distribution of the debt-to-income ratio is negatively skewed. The estimated return for the loan shows a normal distribution.

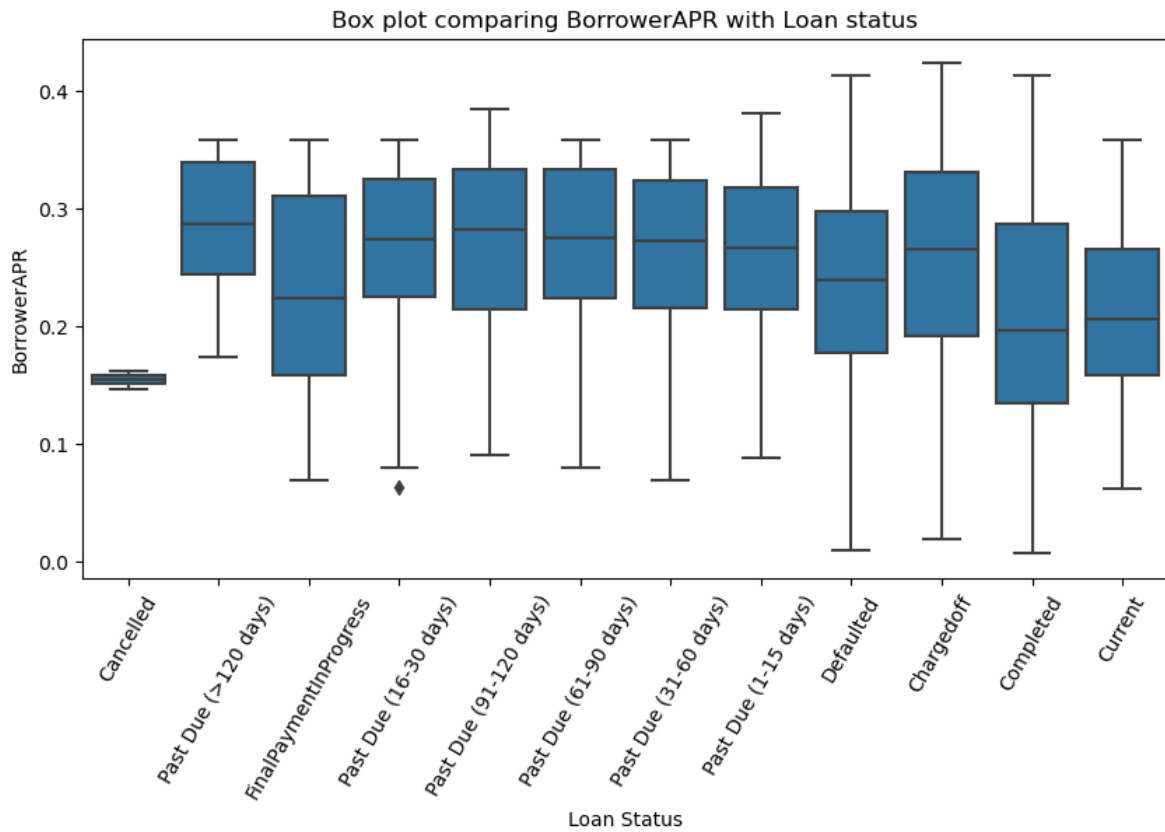
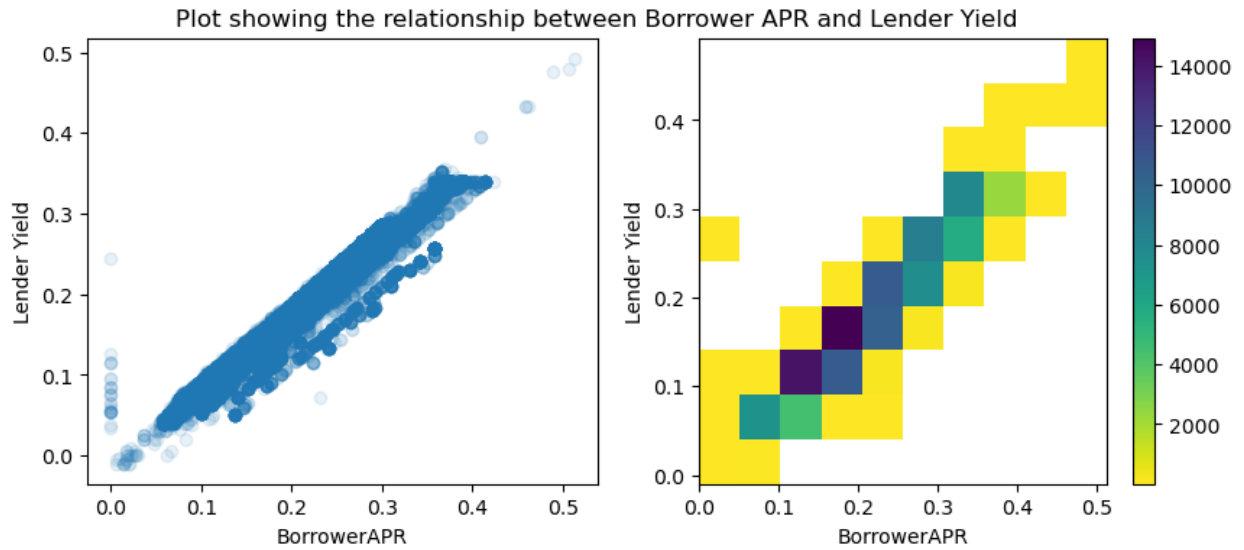


The above plot shows a custom risk score created from Prosper data in the past. The score is on a scale from 1 to 11, with 11 being the best or lowest risk score. Loans that were started after July 2009 are eligible. From the plot above, a lot of prosper score was recorded as 0. This is because no value was recorded before July 2009.

## Prosper Loan Dataset

### 3.2 Bivariate Exploration

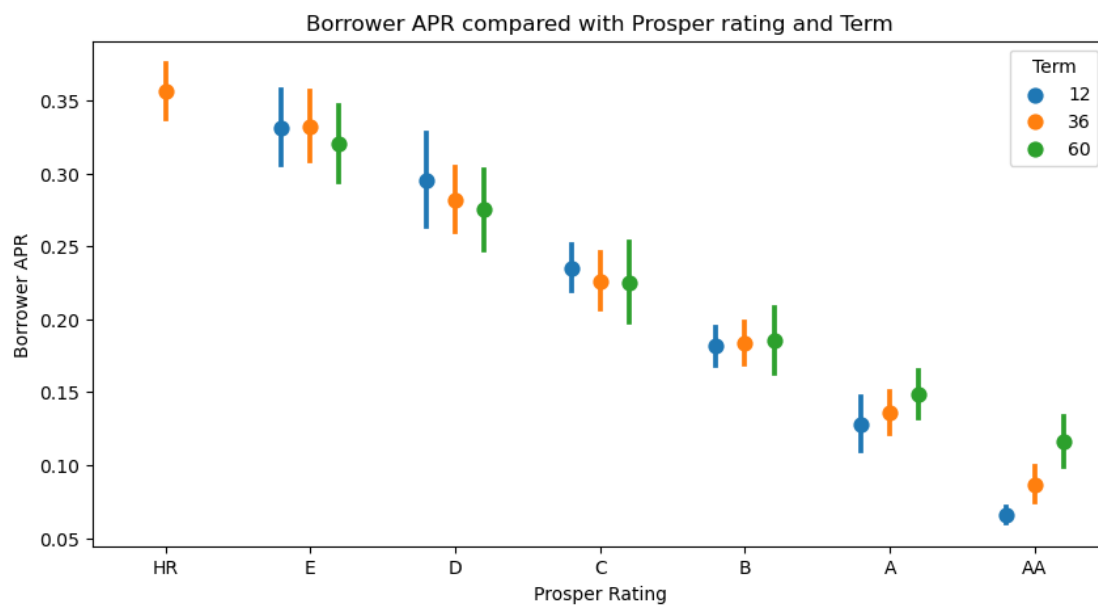
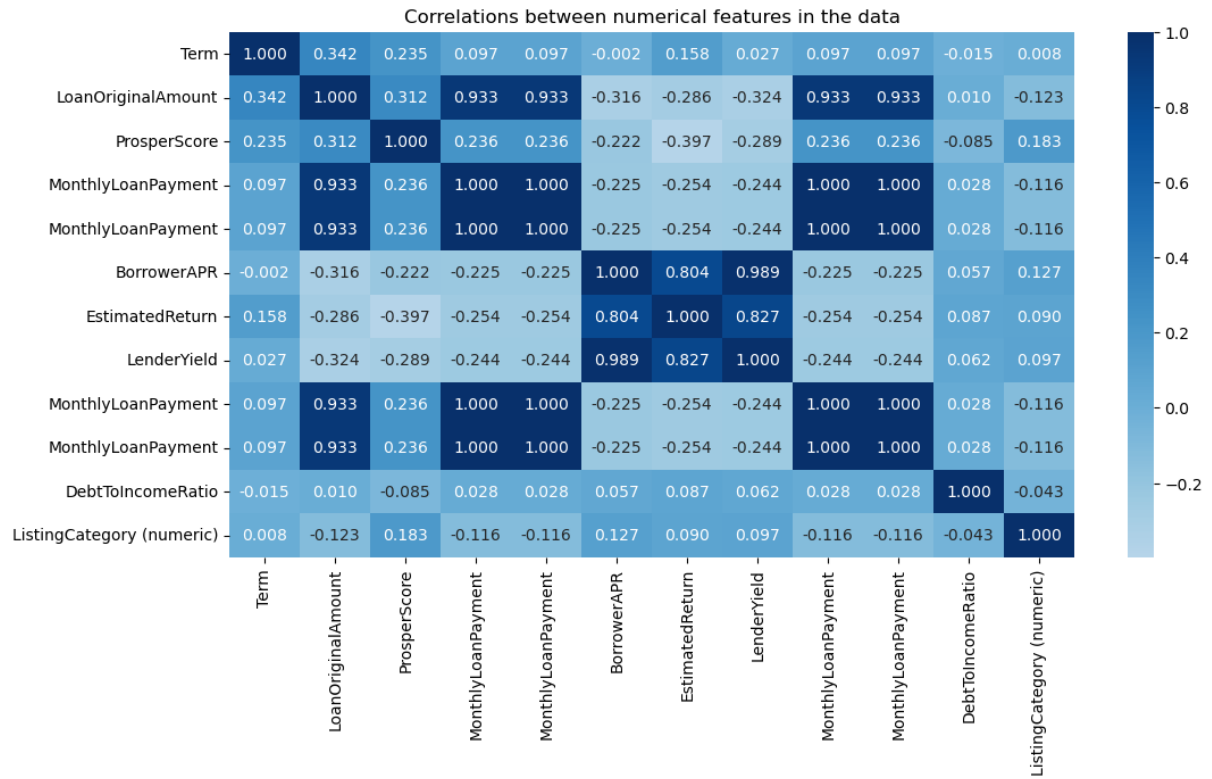
The scatter plot and heat map show a positive correlation between the borrowerAPR and lender yield. This indicates that as the APR increases, the lender yield also increases.



## Prosper Loan Dataset

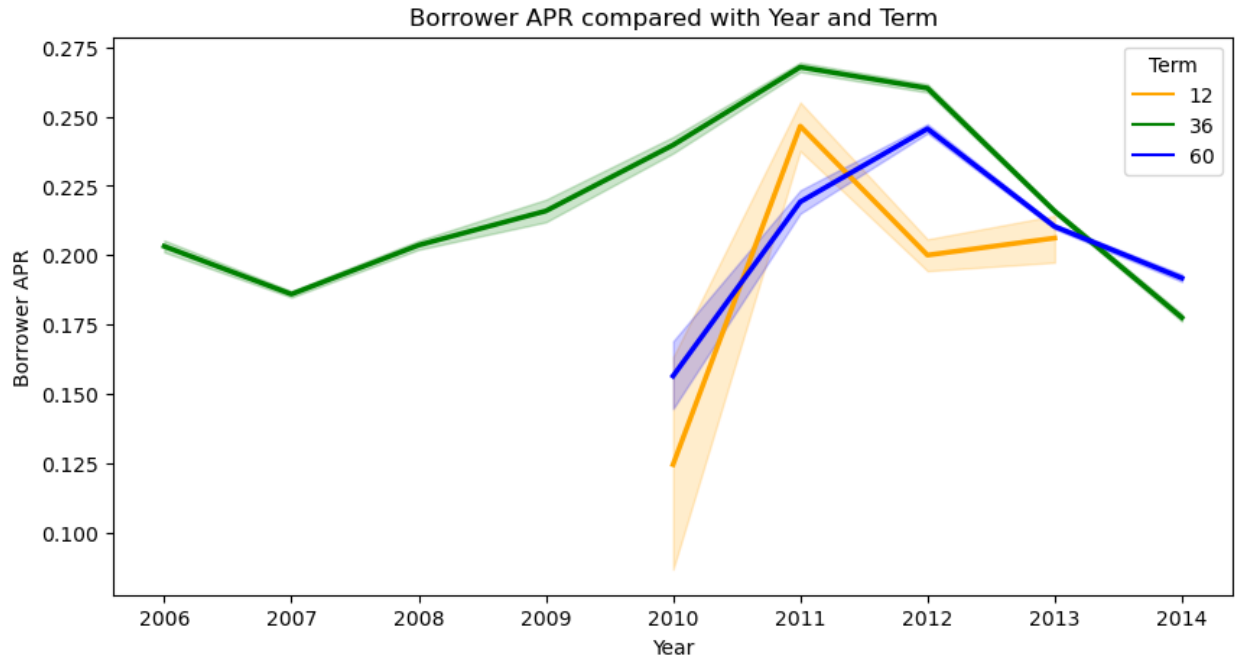
### 3.3 Multivariate Exploration

The correlation matrix below was done to compare the relation between numerical variables in the dataset. The co-efficient varies from +1 through 0 to -1. It can be deduced that lender yield and borrower APR have a high positive correlation.



### *Prosper Loan Dataset*

Prosper rating is the rating assigned at the time the listing was assigned. When compared with the borrower APR across the term, it can be deduced from the plot above that as the borrower APR reduces, the rating also reduces.



The line plot shows the distribution of borrower APR across the year for each term. Borrower APR was highest in 2011 for 12 and 36-month terms. and in 2012, 60 months loan term had the highest APR.