

# Explanatory of Prosper Loan Data Analysis

January 2, 2019

## 1 Project Prosper Loan Data Analysis

### 2 Introduction

This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others.

This Loan Data from Prosper is last updated on 03/11/2015

```
In [4]: #import libraries
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt
% matplotlib inline
```

```
In [5]: # Read and check the loan data
df_loan = pd.read_csv('prosperLoanData.csv')
```

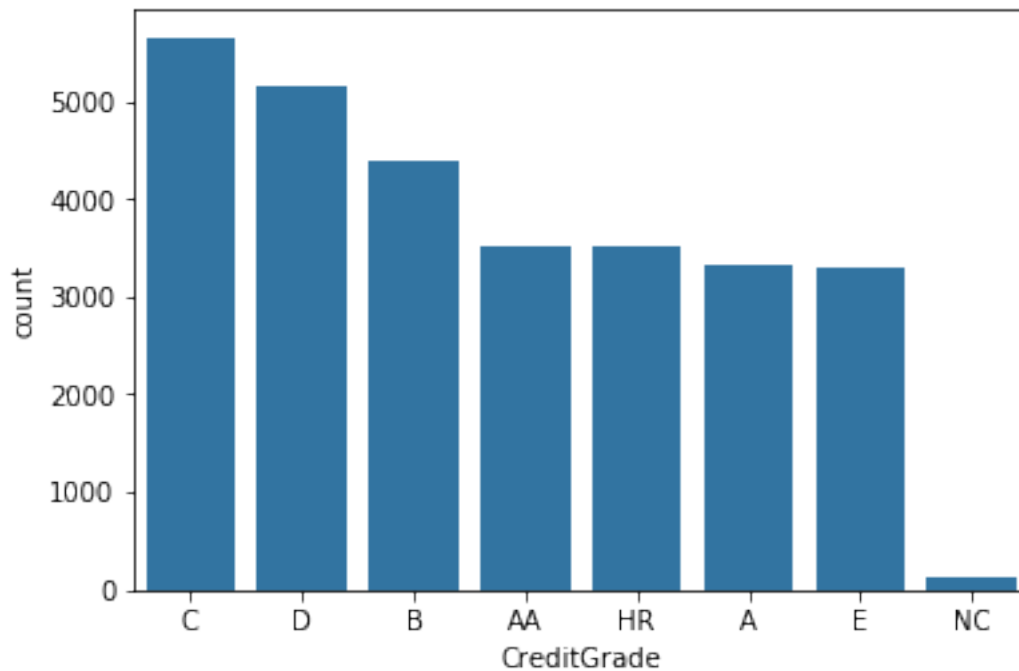
### 3 The most critical tool in a P2P lending organization is its ability to assess a borrower's creditworthiness as accurate as possible. Here, I am going to asses the tools used and to see if it is accurate in determining a person's creditworthiness mainly Credit Grade and Prosper Score

This presentation three main focuses are on the most critical part of Peer-to-Peer Lending which is CreditGrade , BorrowerRate and Lender Yield not to mention the custom built risk assesment tool called ProsperScore to assess the creditworthiness of the borrower

### 4 Research Question 1 : What are the most number of borrowers Credit Grade?

```
In [6]: # Check the univariate relationship of Credit Grade
base_color = sb.color_palette()[0]
credit_grade = df_loan['CreditGrade'].value_counts().index
sb.countplot(data = df_loan, x = 'CreditGrade', color = base_color, order = credit_gra
```

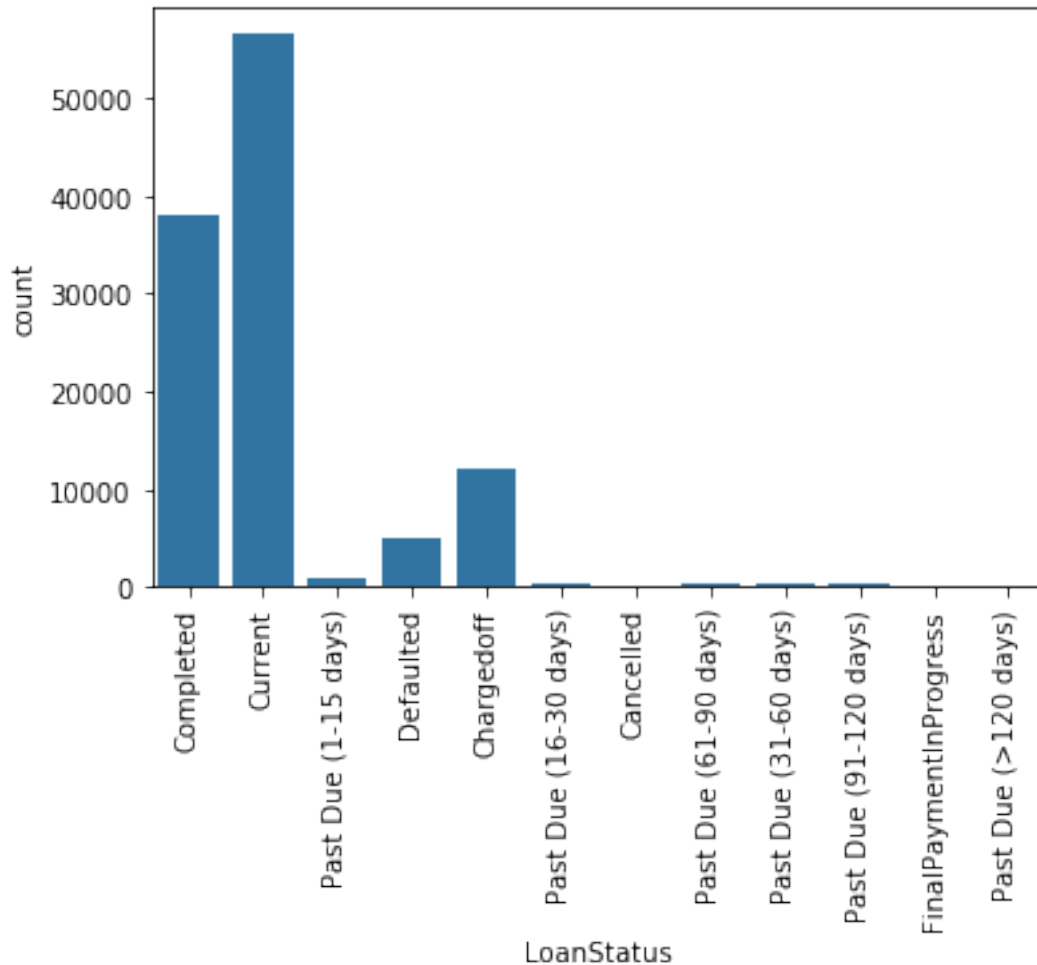
Out [6]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1a21fd4ba8>



Here, we found out that the most number of borrowers CreditGrade is C following by D

## 5 Research Question 2 : Since there are so much low Credit Grade such as C and D , does it lead to a higher amount of delinquency?

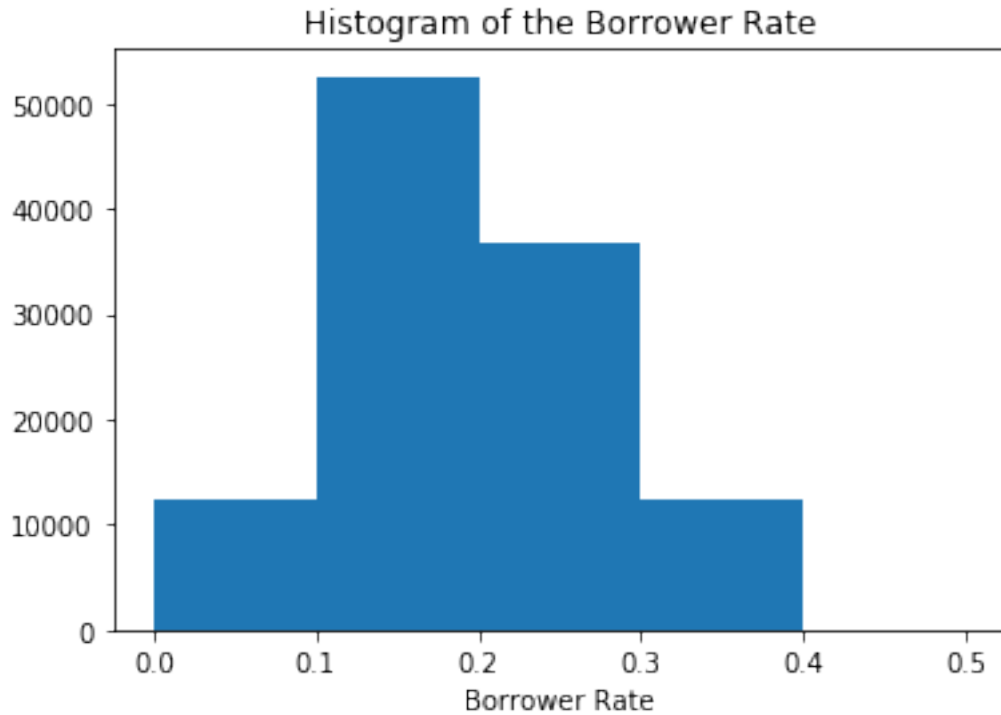
```
In [7]: # Check the univariate relationship of Loan Status
base_color = sb.color_palette()[0]
sb.countplot(data = df_loan, x = 'LoanStatus', color = base_color)
plt.xticks(rotation = 90);
```



Here, we found out that the number of Loan that is still on-going is the highest while Completed loan is the second highest. This insight allows us to understand that the high amount of low CreditGrade doesn't make the delinquency rate much higher

## 6 Research Question 3 : What is the highest number of BorrowerRate?

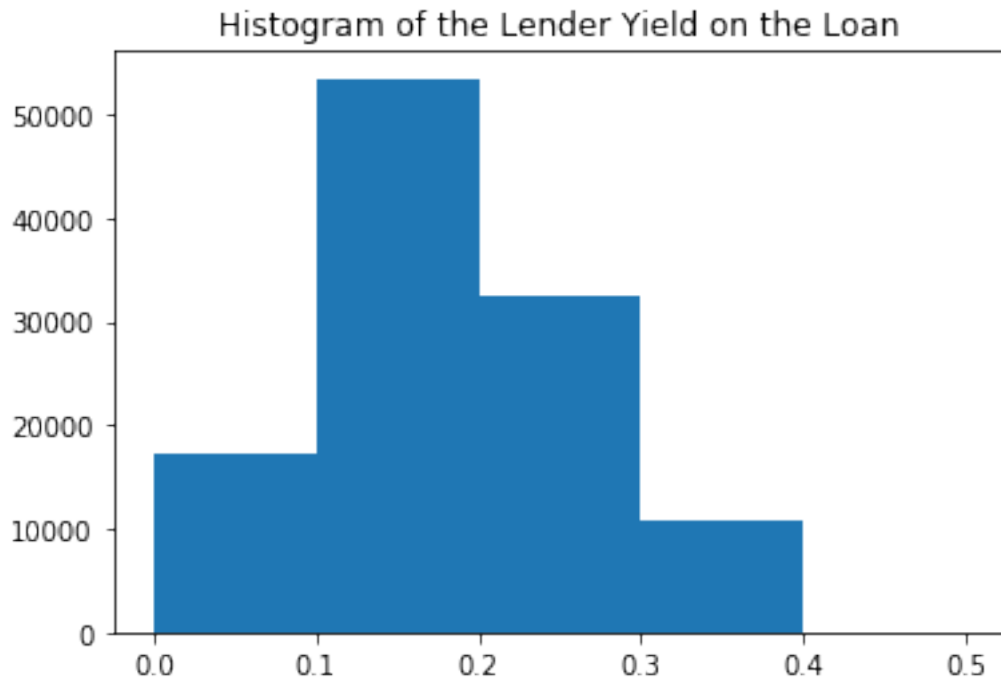
```
In [8]: # Check the univariate relationship of Borrower rate
bin_edges = np.arange(0, df_loan['BorrowerRate'].max()+ 0.1, 0.1)
plt.title('Histogram of the Borrower Rate')
plt.xlabel('Borrower Rate')
plt.hist(data = df_loan, x = 'BorrowerRate', bins = bin_edges);
```



Here, we have found that between 0.1 and 0.2 is the highest borrower rate while the range between 0.0 and 0.1 and the range between 0.3 and 0.4 has the lowest borrower rate

**7 Research Question 4 : Since the highest number of Borrower Rate is between 0.1 and 0.2, does the highest number of Lender Yield is between 0.1 and 0.2?**

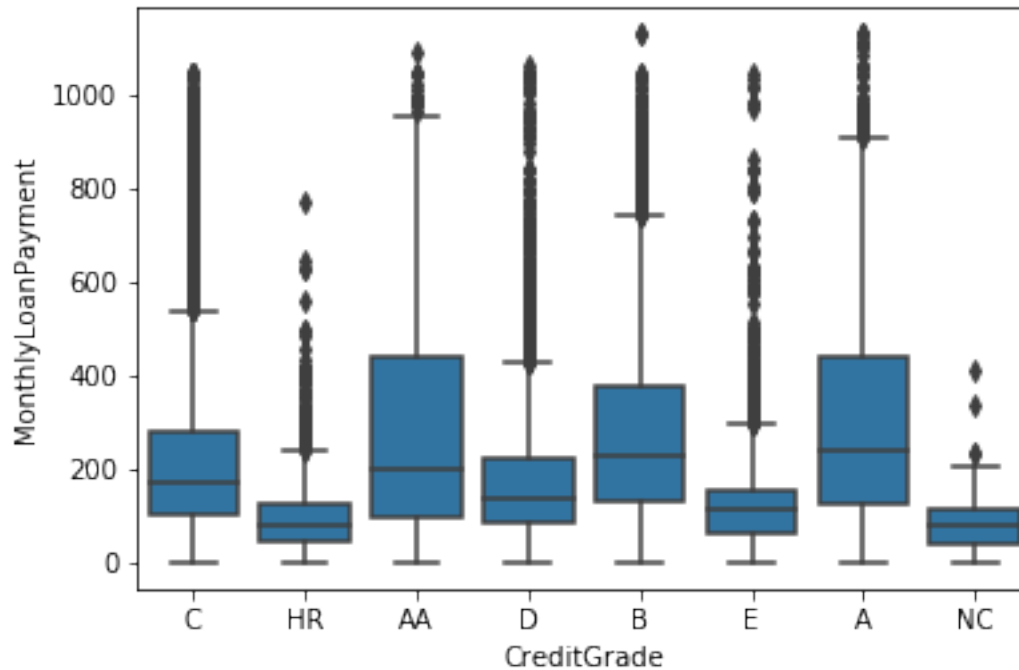
```
In [9]: # Check the univariate relationship of Lender Yield on Loan
bin_edges = np.arange(0, df_loan['LenderYield'].max()+ 0.1, 0.1)
plt.title('Histogram of the Lender Yield on the Loan')
plt.hist(data = df_loan, x = 'LenderYield', bins = bin_edges);
```



This confirms our hypothesis because the highest number of Lender Yield is between 0.1 and 0.2.

## 8 Research Question 5 : Is the Credit Grade really accurate? Does higher Credit Grade leads to higher Monthly Loan Payment? As for Higher Credit Grade we mean from Grade AA to B

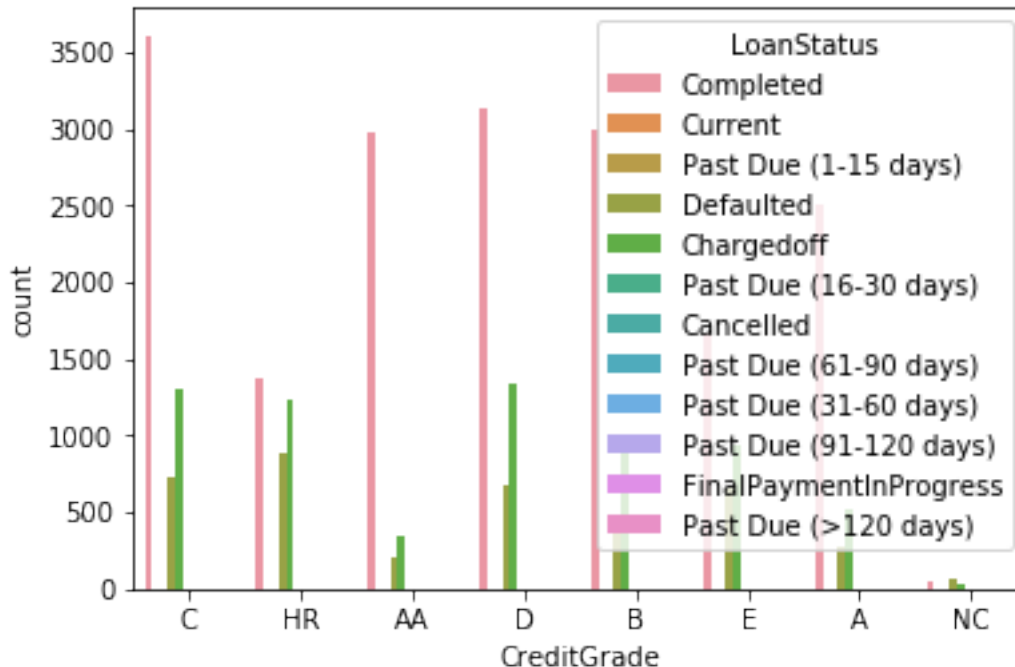
```
In [10]: # Check the Bivariate Relationship between CreditGrade and MonthlyLoan Payment using
         sb.boxplot(data = df_loan, x = 'CreditGrade', y = 'MonthlyLoanPayment', color = base_
```



Here, we can discover clearly that High Credit Grade from Double A (AA) to B has among the highest monthly loan payment which is equal or above USD 200 which allows us to draw insight that the Credit Grade has been used in checking each borrower's credit grade accurately

## 9 Research Question 6 : Here we look at the Completed Loan Status and Defaulted Rate to determine the accuracy of Credit Grade.

```
In [11]: # Check the Bivariate Relationship between CreditGrade and LoanStatus
sb.countplot(data = df_loan, x = 'CreditGrade', hue = 'LoanStatus');
```



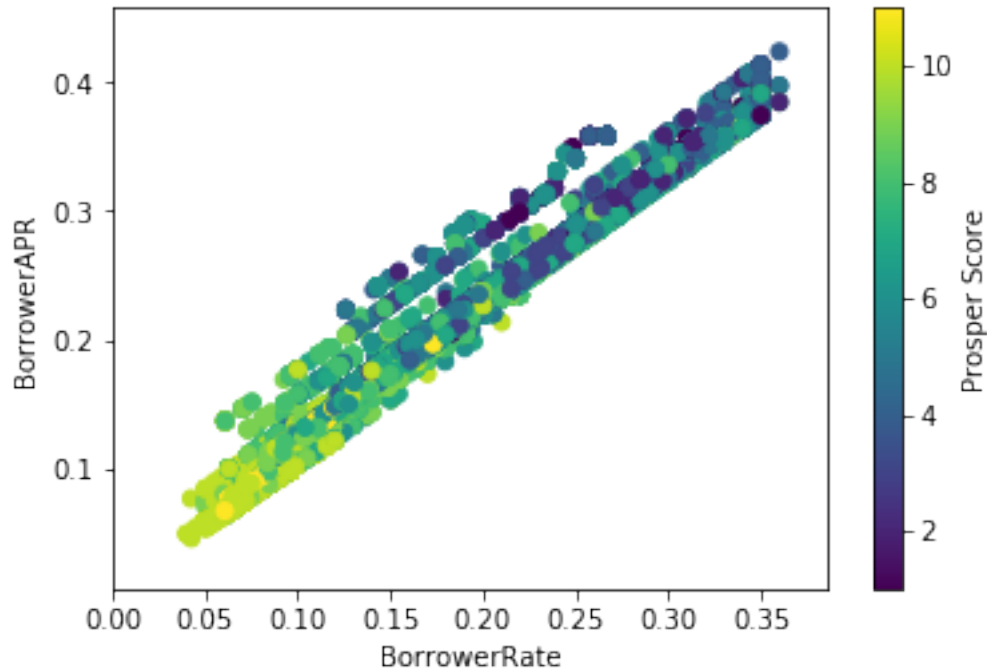
If we look at Double A Credit Grade , it has a higher ratio of Completed Loan Status to Defaulted Rate while C has a low ratio of Completed Loan Status to Defaulted Rate

By Looking at both the Completed Loan Status and Defaulted Rate we can determine even further that the Credit Grade is truly accurate

**10 Research Question 7 : Now we know the Credit Grade is accurate and is a tool that is used by the organization in determining the person's creditworthiness. Now we need to understand does the ProsperScore, the custom built risk assesment system is being used in determing borrower's rate?**

**10.1 From a theoretical standpoint, if the higher ProsperScore leads to lower Borrower Rate and Borrower Annual Percentage Rate that means the Prosper Score is being used alongside the Credit Grade in determing a person's creditworthiness.**

```
In [12]: # Check the Multivariate Relationship between BorrowerRate and BorrowerAPR
plt.scatter(data = df_loan, x = 'BorrowerRate', y = 'BorrowerAPR',
            c = 'ProsperScore')
plt.colorbar(label = 'Prosper Score')
plt.xlim(df_loan['BorrowerRate'].min())
plt.ylim(df_loan['BorrowerAPR'].min())
plt.xlabel('BorrowerRate')
plt.ylabel('BorrowerAPR');
```



This confirm our hypothesis

Here, we can see that both BorrowerRate and BorrowerAPR has a directly proportional relationship in terms of prosper score. With a low BorrowerRate and low BorrowerAPR has a high ProsperScore while high borrowerRate and high BorrowerAPR has a low ProsperScore

In [ ]: *# Conclusion*

Both tools which is Credit Grade and Prosper Score are accurate and is able to assess's borrower Creditworthiness accurately which can lead to the p2p lending organization to have a high confidence level in using its tool mainly Credit Grade and Prosper Score to determine a person's creditworthiness in a bigger market population.

## 10.2 Business Insight

Since the most important assest of a P2P lending Organization is its ability in using its tool to determine a borrower's creditworthiness as accurate as possible. The organization would be more confident to market its organization as a great investment for investor to invest in hence leading to more borrower and higher market capitilization and boost revenue growth

In [ ]: