

Prosper Loan Data - Default Analysis

Investigation Overview

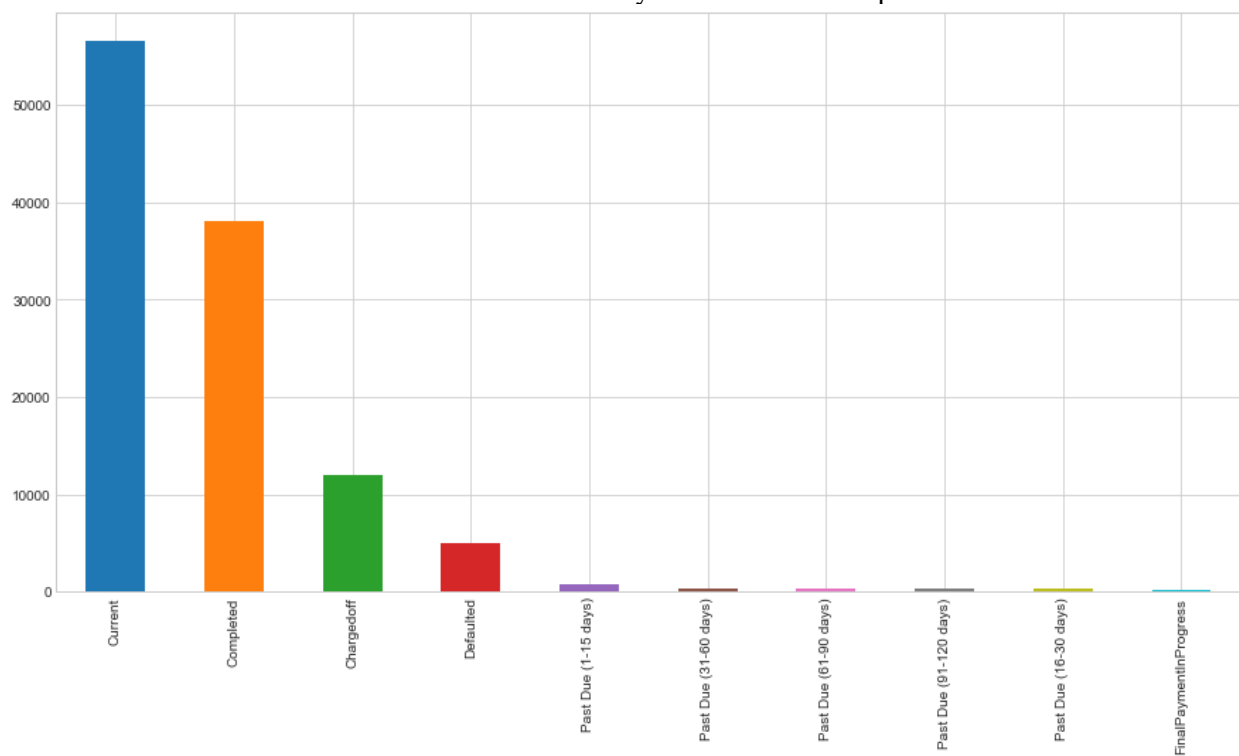
This notebook will document my efforts to investigate an interesting dataset from the Prosper peer-to-peer lending platform. There are many questions one could ask of this data, but in this exercise, I will focus on questions related to this project rubric.

Dataset Overview

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.

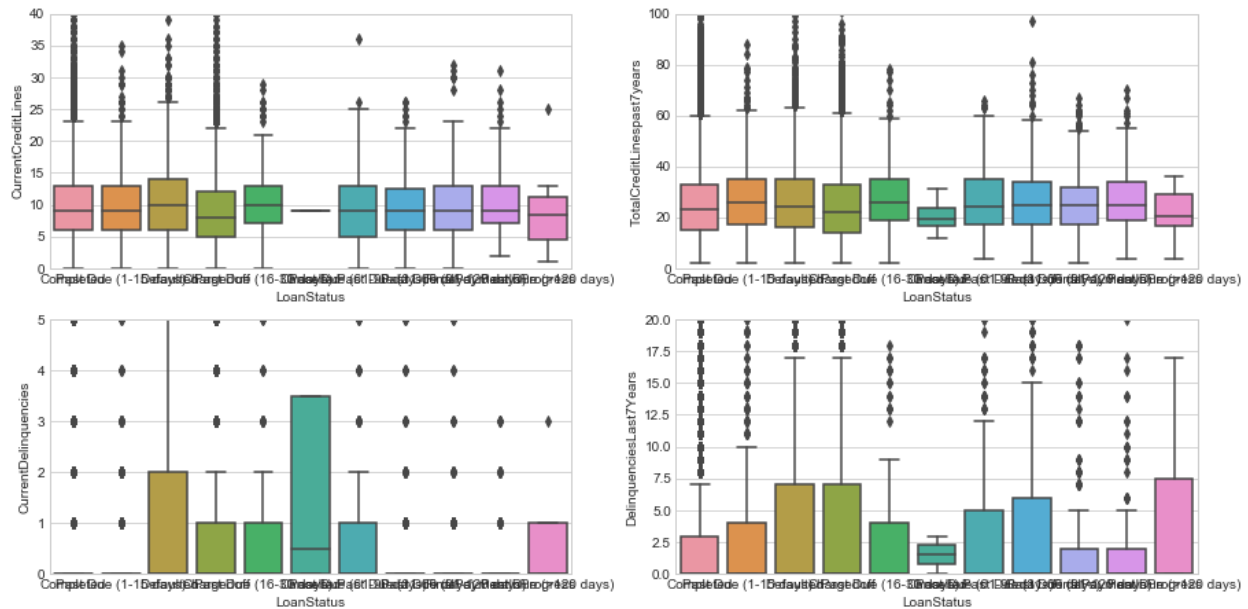
Distribution of Data

The distribution of the loan amount seems to be mostly in current and completed.



Exploring credit history

Here we can see the number of credit lines doesn't seem to have a particularly significant relationship with default, but delinquencies do. Among loans that were completed, it is very rare for the debtor to have any current delinquencies.



showing an increase in price. Within each level of clarity, we also observe an increase in price based on the color grade (J is lowest, D is highest).

Exploring loan characteristics:

The distribution of the loan amount seems to be consistent between both classes.

