

EXPLANATORY DATA VISUALIZATION

Data Exploration-Loan dataset from Prosper

By Carolynne Kemunto Machuka

Introduction

Data set

The data is a loan data set that has been provided by Udacity. It contains 113,937 loans with 81 variables on each loan. The key column of interest for this analysis include original loan amount, borrower rate (or interest rate), monthly loan repayment, bank credit utilization, loan status, income range, employment status, lender's yield, investors and maturity term.

Assessing and cleaning data

The data set had numerous variables. Relevant columns for this exploration were extracted from this data to a new data frame hereby referred as loan data clean. The new data frame was assessed. Duplicated rows and those with null values were dropped. Data type of status of loan was converted to categorical and listing creation date column converted to date format. The clean data set was then stored and retrieved for exploratory analysis as tidy loan data.

Summary of findings.

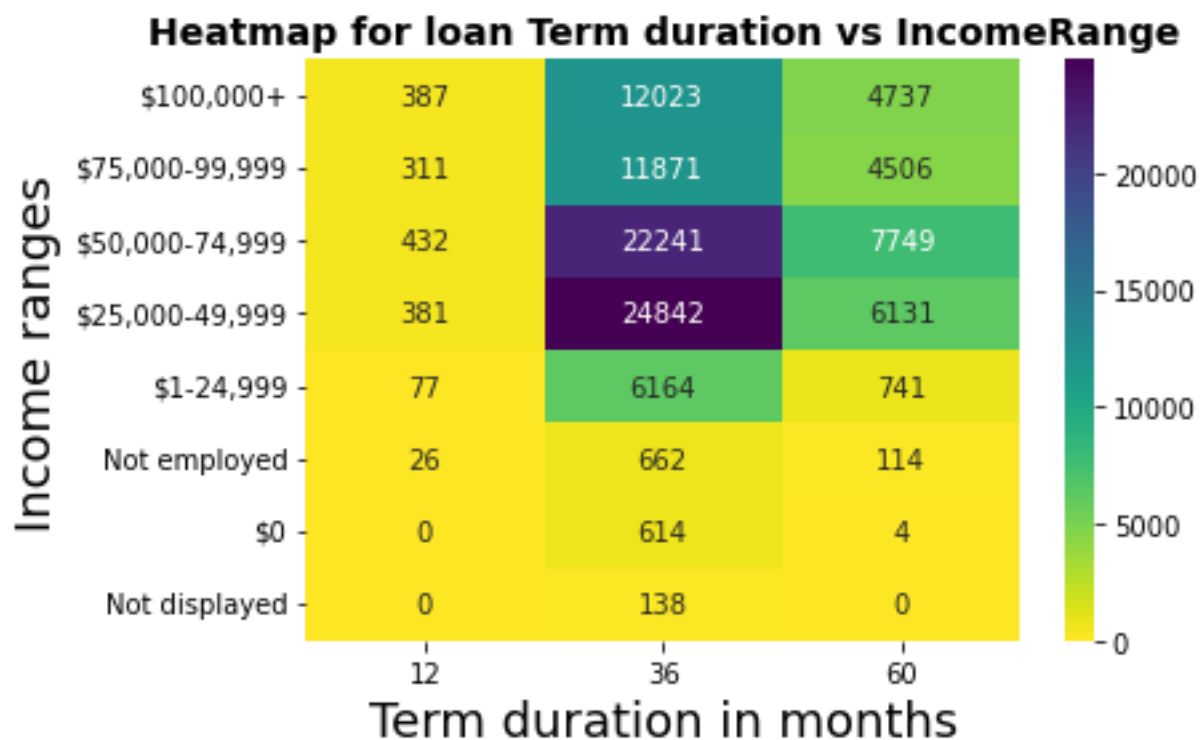
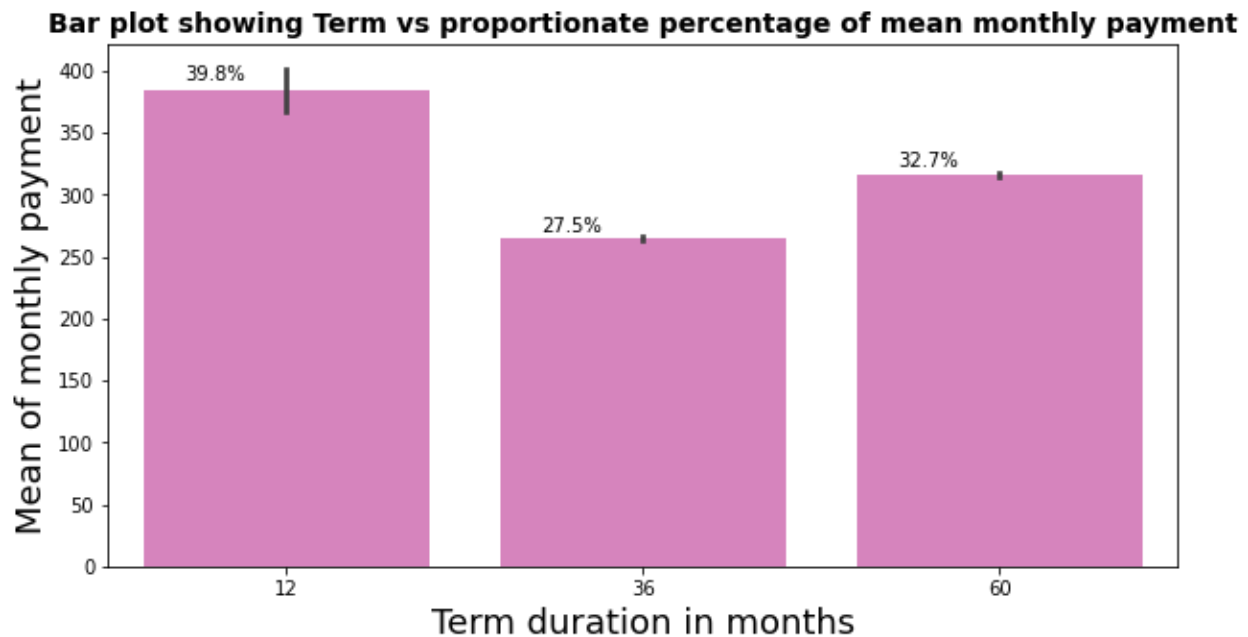
Factors affecting loan term maturity.

The term of loan is in 3 categories: 12,36 and 60 months.

- 36 months attracts the largest number of borrowers because it has the lowest monthly payment.
- 12-month loan term offers the lowest interest rate but borrowers shy away from it due to high monthly payment deduction during this duration. Borrowers seem to focus more on the monthly payment rather than lower interest rates.
- Heavy consumers of loans are middle income earners earning between \$25,000 and \$74,999, interestingly followed by high income

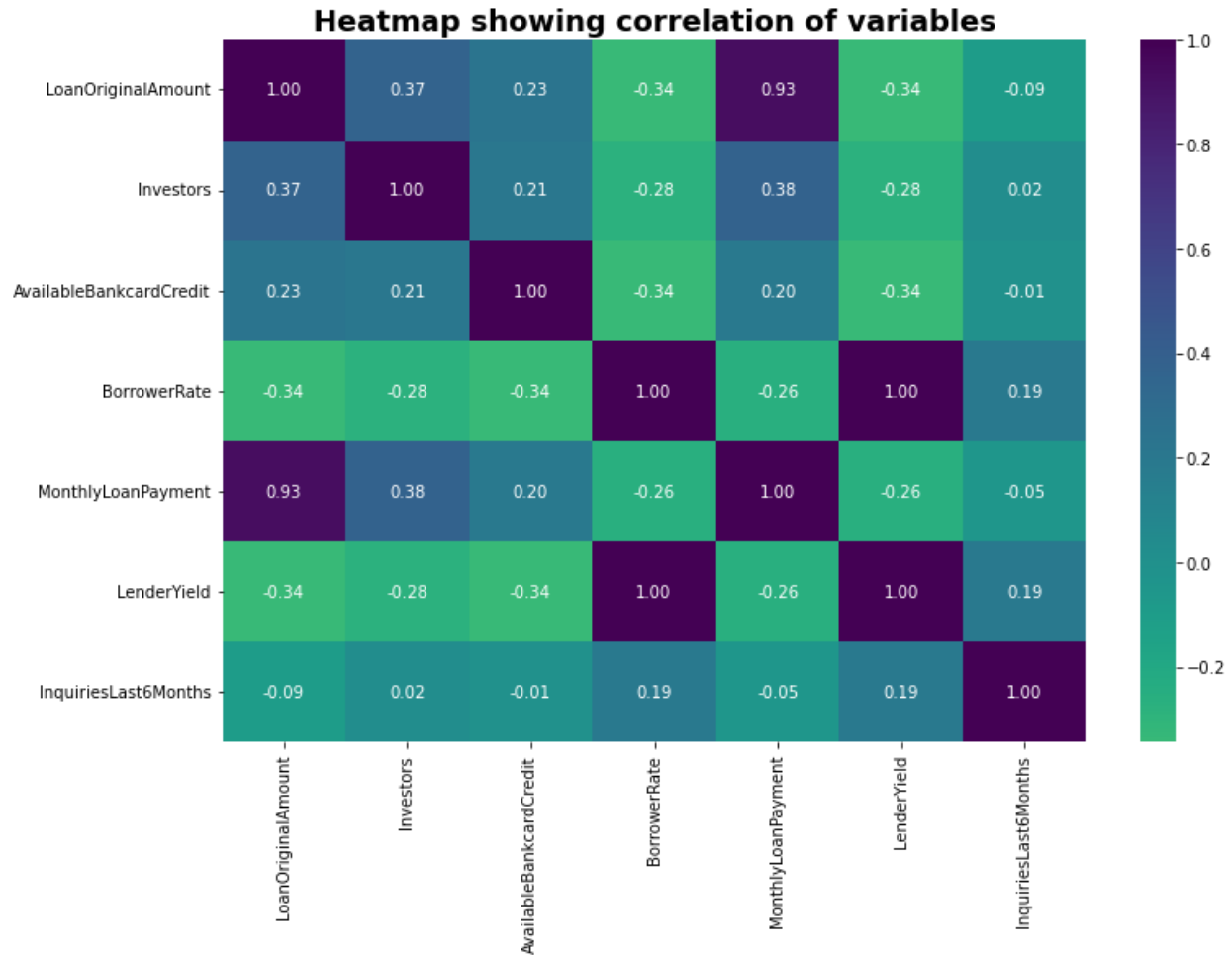
earners earning \$75,000+. Subsequently, they receive the largest loan amount and pay the largest monthly payment.

- The lion's share of the principal loan amount is taken by borrowers taking it for 60 months. Majority of them earn \$ 25,000+.



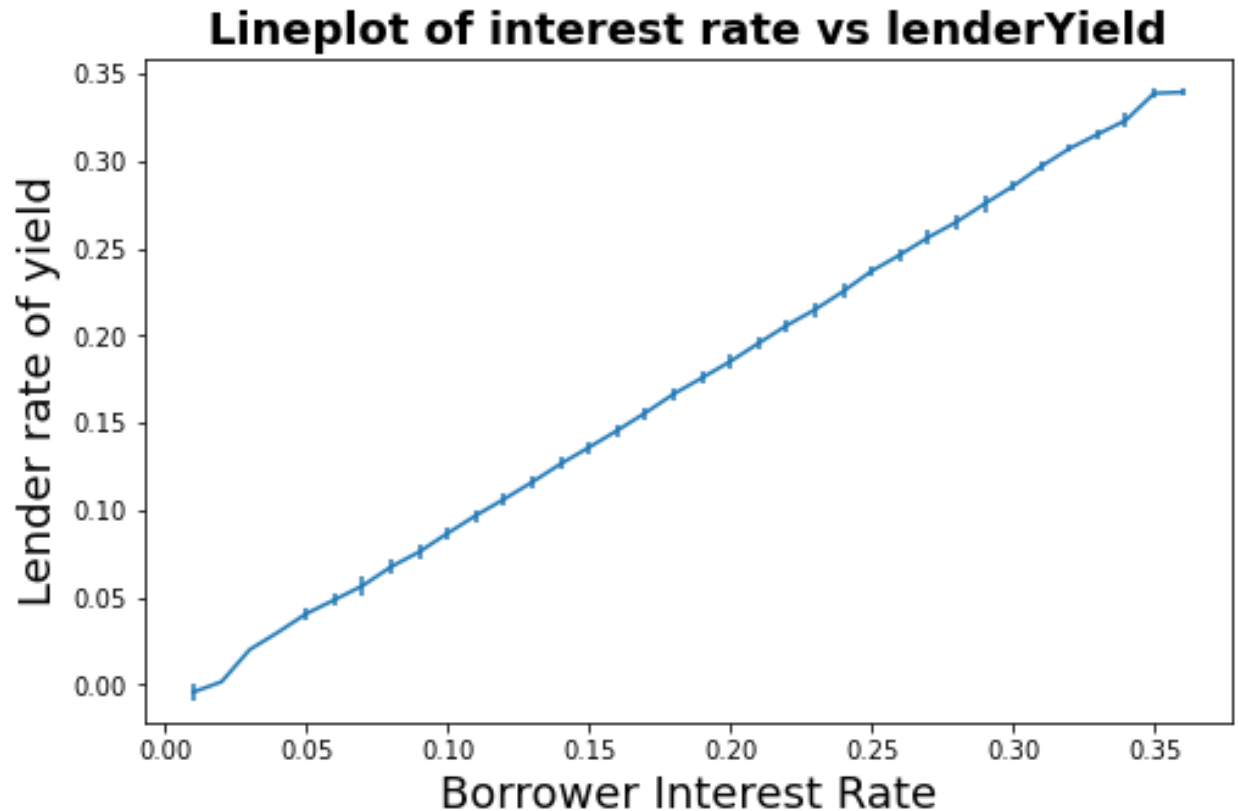
Factors affecting the original loan amount.

- Most borrowers prefer low loan amount of between \$ 1000 to \$15,000 because such amounts attract low monthly payment.
- Higher loan amount has lower yield rate and thus lower interest rate. However, they attract high monthly payment thus shying away borrowers who forego this low interest rates incentives it provides to have a little more in their pockets after their monthly payment deduction.
- Loan amount has a strong correlation of 0.93 with monthly payment implying that as principal amount increases the monthly payment will proportionately increase and vice versa. However, loan amount seems to have a weak positive correlation with number of investors and available bank card credit.



Factors affecting the borrower's rate of interest.

- Lender's yield has a perfect positive correlation of 1 with interest rate hence is the key factor affecting interest rate. High lender's yield results to an equal high rate of interest for borrowers.
- High monthly payment attracts low interest rates of borrowing. Interest rate has a weak negative correlation with original loan amount.



Factors affecting the Loan outcome.

- Employed and Full time personnel are leading in completed loans. Interestingly, they are also leading in defaulted loans, probably due to large number of borrowers in this category. All the employment status have defaulters, though the defaulters' proportion of 0.0316 is relatively small
- Inquiries are made more by those with low monthly loan payment. However, inquiries for those with completed and defaulted loans spikes especially for those with high monthly payment. This can be attributed to the nature of these two status.

Key Insights.

- ❖ Lender's yield has a perfect positive correlation of 1 with interest rate hence is the key factor affecting interest rate. High lender's yield results to an equal high rate of interest for borrowers.
- ❖ 36 months attracts the largest number of borrowers because it has the lowest monthly payment. 12-month loan term offers the lowest interest rate but borrowers shy away from it due to high monthly payment deduction during this duration. Borrowers seem to focus more on the monthly payment rather than lower interest rates
- ❖ Heavy consumers of loans are middle income earners earning between \$25,000 and \$74,999, interestingly followed by high income earners earning \$75,000+. Subsequently, middle income earners, receive the largest loan amount and pay the largest monthly payment.
- ❖ Higher loan amount has lower yield rate and thus lower interest rate. However, they attract high monthly payment thus shying away borrowers who forego this low interest rates incentives it provides to have a little more in their pockets after their monthly payment deduction.
- ❖ Loan amount has a strong correlation of 0.93 with monthly payment implying that as principal amount increases the monthly payment will proportionately increase and vice versa. However, loan amount seems to have a weak positive correlation with number of investors and available bank card credit.
- ❖ Employed and Full time personnel are leading in completed loans. Interestingly, they are also leading in defaulted loans, probably due to large number of borrowers in this category. All the employment status have defaulters, though their proportion of 0.0316 is relatively small