Prosper loan data

This data set contains 113,937 loans with 81 variables on each loan.

After exploring this dataset I've selected the variables which will help me on this investigation.
They are 19 variables:

ListingCreationDate
Term
LoanStatus
ClosedDate
BorrowerAPR
BorrowerRate
LenderYield
ProsperRating (Alpha)
ProsperScore
ListingCategory
BorrowerState
EmploymentStatus
IsBorrowerHomeowner
IncomeRange
StatedMonthlyIncome
LoanCurrentDaysDelinquent
LoanOriginalAmount
MonthlyLoanPayment
LP_ServiceFees , Investors)

Adding two new columns

Listing year – closing year

Cleaning my data through :

- removing the outlier scores from ProsperScore
- creating orderd categories for income range and alphabetic prosperRating
- changing listing category column name
- changing dtypes to 'category' dtype for all categories columns
- changing dtype of the date columns from object to datetime

Creating functions to avoid repetitive code:

- creating a function for sizing the figure (figure_size)
- creating a function to order my column due to value counts (ordermycol)
- creating a function to plot horizontal countplots(h_count)
- creating a function to plot vertical countplots(v_count)

Univariant Visualizations:

- ***we can notice here that most of our borrowers are between employed and full-time employment state***
- ***CA state has the highest count of borrowers***
- ***the most common income range is between 25k and 49,999 USD***
- ***the common reason for borrowing is Debt consolidation***
- ***that most of loans are the longest-term loans***
- ***through log scale distribution for Loan Original Amount we have normal distribution with highest streak between 4k and 10k $***
- ***through log scale distribution for monthly loan payment is slightly left skewed distribution with high peak in between 100 - 500 $***
- ***most loans have investors number between [1-10] for the rest loans high peak starts for about (35 - 250) investors.***

Bivariate Visualizations:

- ***the highest Borrower rate is for the cosmetic prodcedures***
- ***we find that the completed and current mean is less than any status***
- ***the Not employed have higher IQR thant the empolyed status***

Multivariate Visualizations:

- ***the higher prosper score the higher chance that the borrower can get higher amount of loan < ($20,000)***

- ***A positive correlation between Lender Yield and Borrower APR***
- ***A negative correlation between Lender Yield and Prosper Score***
- ***the borrower rate increased for some years in case of non homeowners, however the peak was between (2010 - 2011) and it decreased after 2011. As regards of homeowners, there wasn't much difference till 2009 but after 2009, the graph went up till 2011 and started falling off from then.***