employment

Feature

-Borrower's credit grade : ProsperRating (numeric), CreditScoreRangeLower

-Borrower's financial related data : IncomeRange, StatedMonthlyIncome, DebtToIncomeRatio, CurrentDelinquencies, AmountDelinquent, PublicRecordsLast12Months, BankcardUtilization,

-Borrower's income/asset related data :EmploymentStatus, IsBorrowerHomeowner

-Loan information : LoanStatus, LoanOriginalAmount, BorrowerAPR

-Others :ListingCategory (numeric)

|  |  |  |
| --- | --- | --- |
| -LoanStatus |  | Just to check to have an idea of the data set. |
| -ProsperRating (numeric), |  | Slightly left skew? |
| -EmploymentStatus | Combined category | Employed a lot |
| -IsBorrowerHomeowner |  | 50-50 |
| -IncomeRange | Changed data na and unemployed | Many people with high income |
| ListingCategory (numeric) |  |  |
| -CreditScoreRangeLower | Compared with upper. And changed 0 into nan | Uni modal, but there are several sudden increase on lower side. Higher side has smooth. |
| -StatedMonthlyIncome | Log scale | Unimodal, |
| -DebtToIncomeRatio | Changed some into 0 | Quite wide spread |
| -CurrentDelinquencies |  | Mostly 0 |
| -AmountDelinquent | logscale | Mostly |
| -PublicRecordsLast12Months |  | Mostly 0 |
| -BankcardUtilization | log | Funny shape a lot have 0 |
| LoanOriginalAmount | Loig | Kinda uni/bimodal?.... |
| BorrowerAPR |  | Funny peak around 0.36. what is it.. -> check closer with proseper rating |

Credit Score Range (Lower) : contained a lot of “0”. Given the nature of the score, I treated it as N/A.

Employment status: I included “Full-time” and “Part-time” in the employed category and treated “Not Available” as N/A.

Income range: I included “Not employed” in \$0 and treated “Not displayed” as N/A.

Stated monthly income: as it has a long-tailed distribution, I plotted it in the log-scale. It showed the roughly unimodal distribution.

Debt to income ratio: some outlier had a monthly income less than \$1. I changed this outlier’s debt to income ratio into N/A and also replaced ”0” with N/A, considering the nature of the feature.

Amount of delinquent: it has a long-tailed distribution, I plotted it in the log-scale. It showed the roughly unimodal distribution.

Bank card utilization: it has a long-tailed distribution with most of the data at the value of 0. I plotted in the log-scale, and it looks truncated unimodal distribution.

Loan original amount: As it has a long tailed distribution, I plotted in the log-scale. Although the curve is not smooth, it is roughly looked bimodal.

Questions: APR what tf is the 0.35 and 0.35 ? 3 peak 0.18, 0.3, 0.36 peak.

plt.xticks([1e3, 2e3, 5e3, 1e4, 2e4], ['1k','2k', '5k', '10k', '20k'])