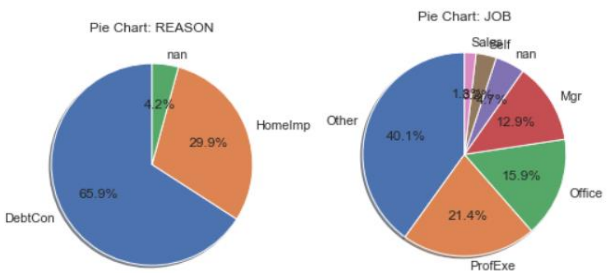


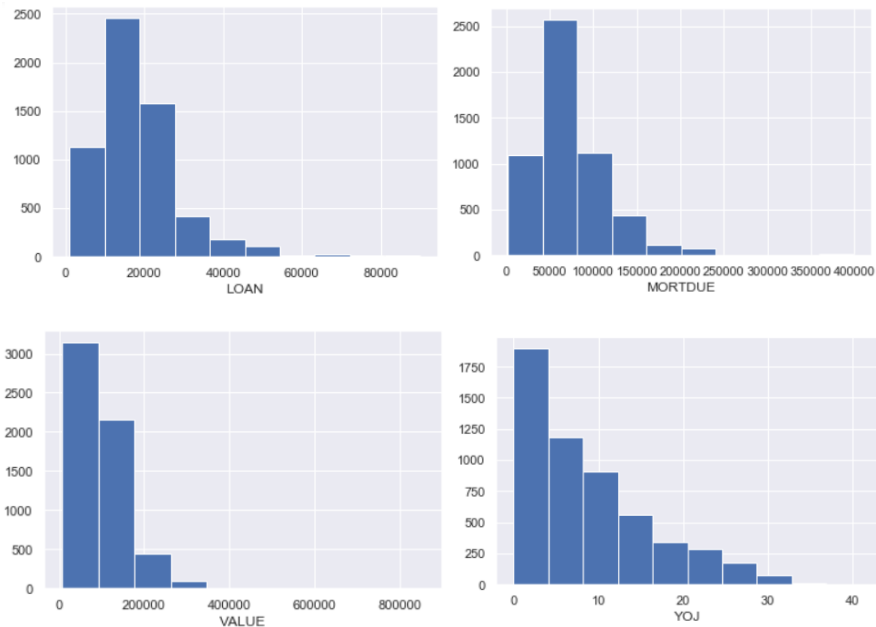
Visualizations and Analysis

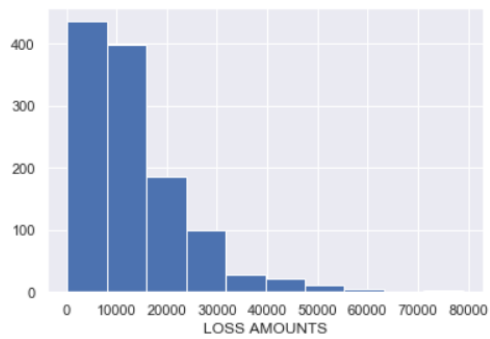
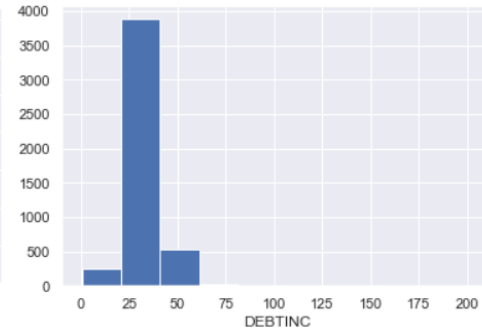
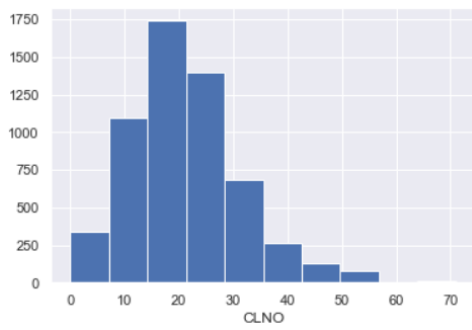
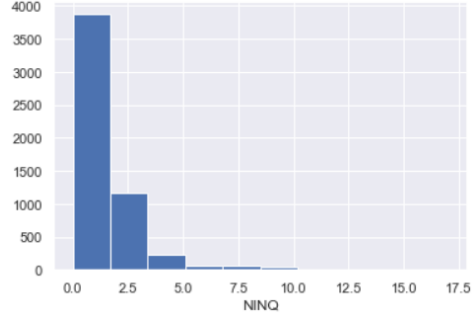
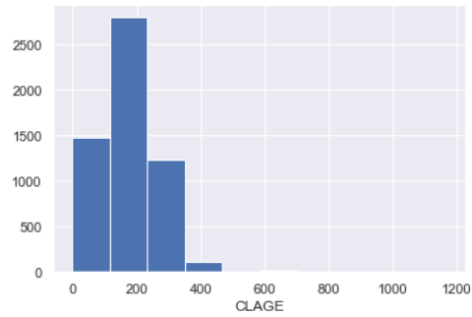
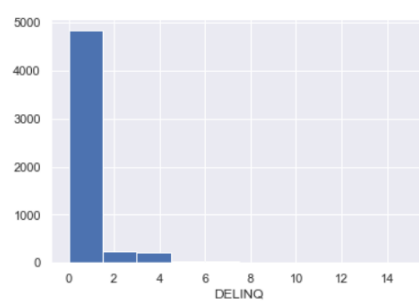
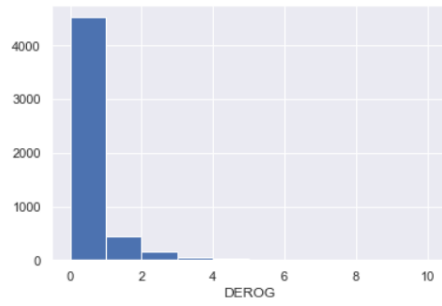
	count	mean	std	min	25%	50%	75%	max
TARGET_BAD_FLAG	5960.0	0.199497	0.399656	0.000000	0.000000	0.000000	0.000000	1.000000
TARGET_LOSS_AMT	1189.0	13414.576955	10839.455965	224.000000	5639.000000	11003.000000	17634.000000	78987.000000
LOAN	5960.0	18607.969799	11207.480417	1100.000000	11100.000000	16300.000000	23300.000000	89900.000000
MORTDUE	5442.0	73760.817200	44457.609458	2063.000000	46276.000000	65019.000000	91488.000000	399550.000000
VALUE	5848.0	101776.048741	57385.775334	8000.000000	66075.500000	89235.500000	119824.250000	855909.000000
YOJ	5445.0	8.922268	7.573982	0.000000	3.000000	7.000000	13.000000	41.000000
DEROG	5252.0	0.254570	0.846047	0.000000	0.000000	0.000000	0.000000	10.000000
DELINQ	5380.0	0.449442	1.127266	0.000000	0.000000	0.000000	0.000000	15.000000
CLAGE	5652.0	179.766275	85.810092	0.000000	115.116702	173.466667	231.562278	1168.233561
NINQ	5450.0	1.186055	1.728675	0.000000	0.000000	1.000000	2.000000	17.000000
CLNO	5738.0	21.296096	10.138933	0.000000	15.000000	20.000000	26.000000	71.000000
DEBTINC	4693.0	33.779915	8.601746	0.524499	29.140031	34.818262	39.003141	203.312149

Categorical Variable Visualizations



Numerical Variable Visualizations





Observations

- Performed preliminary statistical analysis on the dataset by using pandas describe function, we see that in our dataset we have a 20% bad rate, with an average loss amount of \$2,676. Looking at the max, we have outliers in all our numerical data fields.
- We have two categorical variables in our dataset, Reason and Job. The majority reason for the loan is Debt Consolidation which is 66%, 30% have Home Improvement as the reason for the loan and 4% where blank. Job is the second categorical variable, the largest allocation was found to be Other at 40%, followed by ProfEx at 21%, Office at 16%, Manager at 13%, Blank at 5%, Self at 4% and Sales at 1%.

Reason	Loss Rate	Avg Loss Amount
DebtCon	19%	\$3,035
HomeImp	22%	\$1,866

- Looking at the Reason variable we see that HomeImp has a loss rate 3 % more than DebtCon but with a significantly lower average loss amount. DebtCon has a 62% higher average loss amount than HomeImp.

Job	Loss Rate	Avg Loss Amount
Mgr	23%	\$3,300
Office	13%	\$1,776
Other	23%	\$2,684
ProfExe	17%	\$2,435
Sales	35%	\$5,725
Self	30%	\$6,681

- Looking at the Job variable we discover the least risky job is Office, loss rate 13% with an average loss amount of \$1,776, conversely the most risky jobs are Sales and Self with a loss rate of 35% and 30% respectively and average loss amounts of \$5,725 and \$6,681 respectively.
- Majority of the numerical variables are skewed to the right, apart from CLNO and DEBTINC which look somewhat like normal distribution.