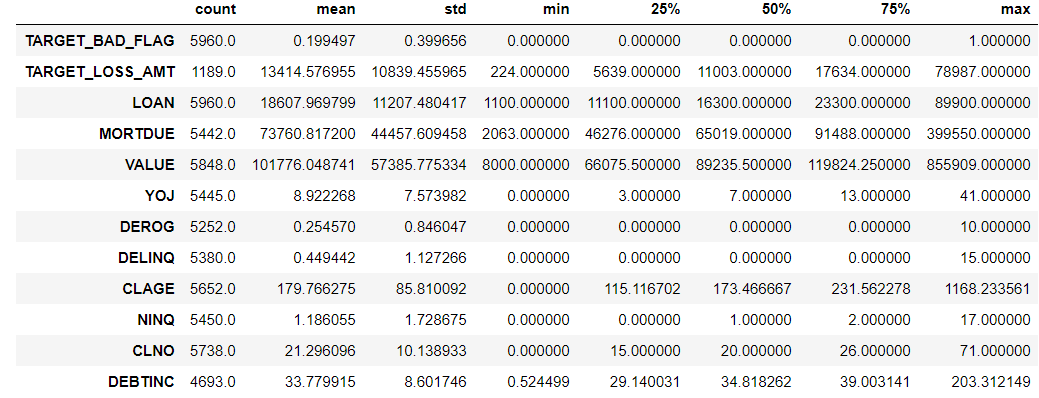
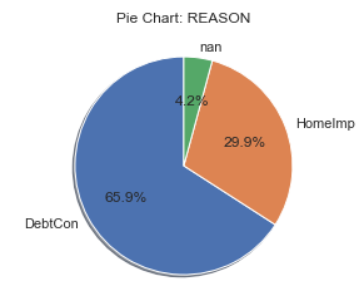
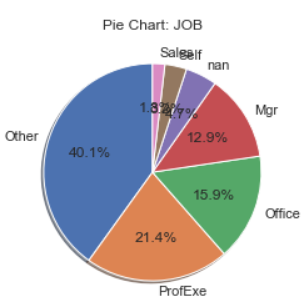
MSDS-422-SEC56\_Assignment 1 Daniel Zaremba

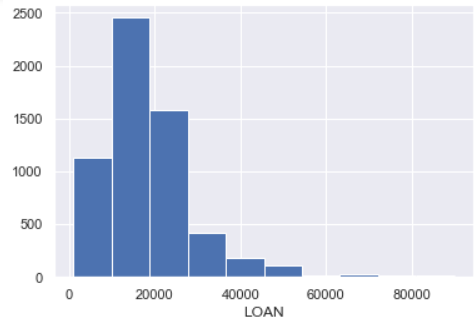
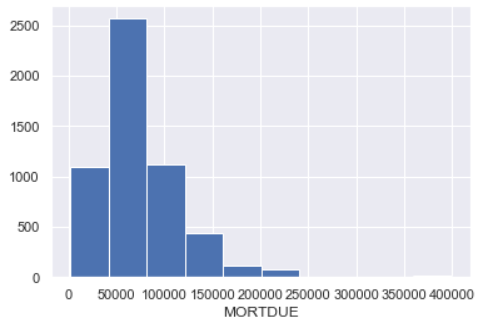
**Visualizations and Analysis**

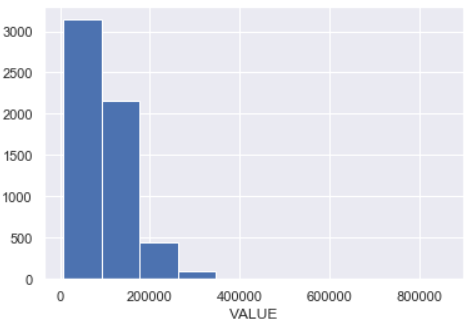
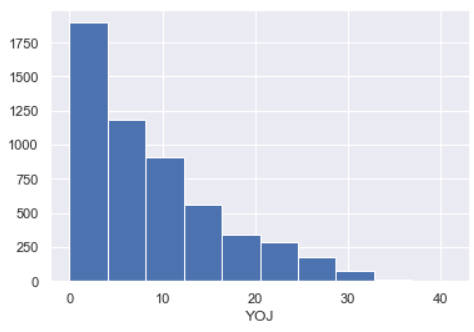


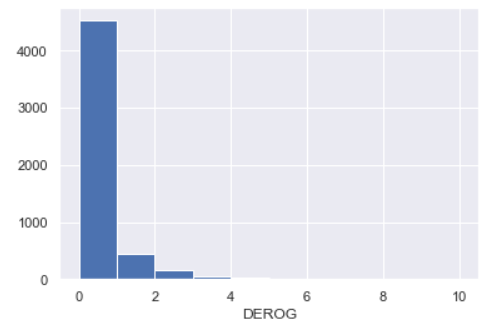
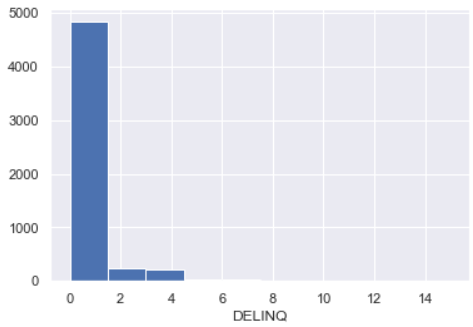
**Categorical Variable Visualizations**

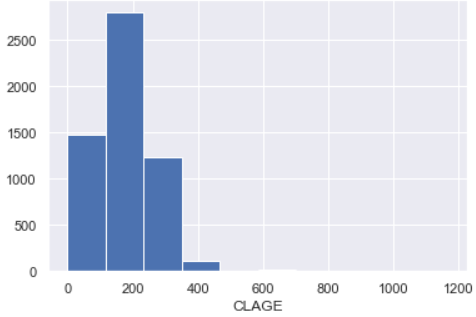
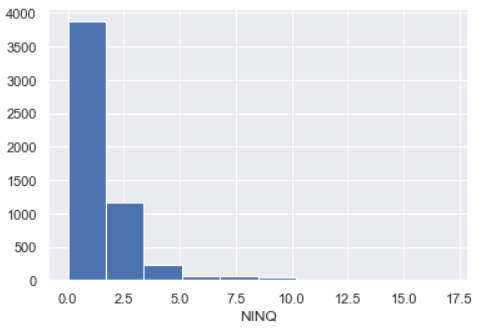
 

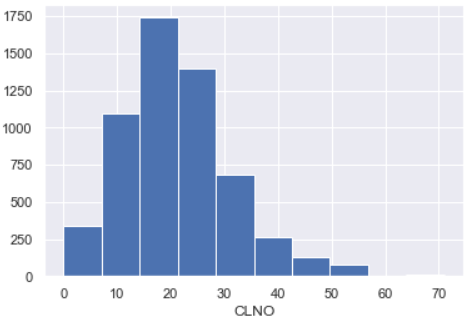
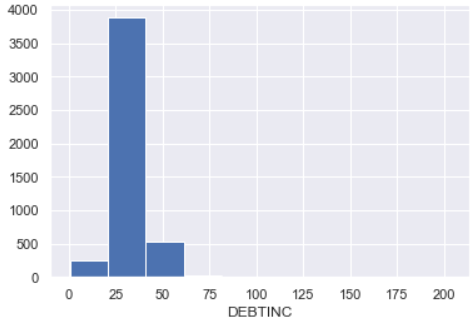
**Numerical Variable Visualizations**

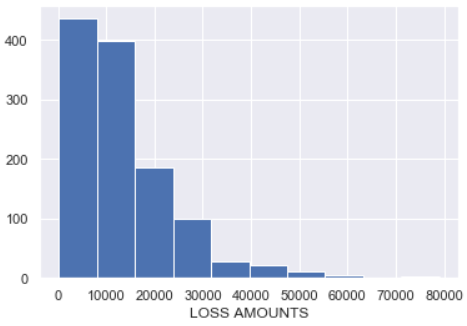
 



**Observations**

* 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 Consildation 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.