

Exploratory Data Analysis of the HMEQ dataset

Introduction: Exploratory data analysis of past data set containing Home Equity Loans. The data set contains two target variables. The first target, TARGET_BAD_FLAG indicates whether or not the loan defaulted. If the value is set to 1, then the loan went bad and the bank lost money. If the value is set to 0, the loan was repaid.

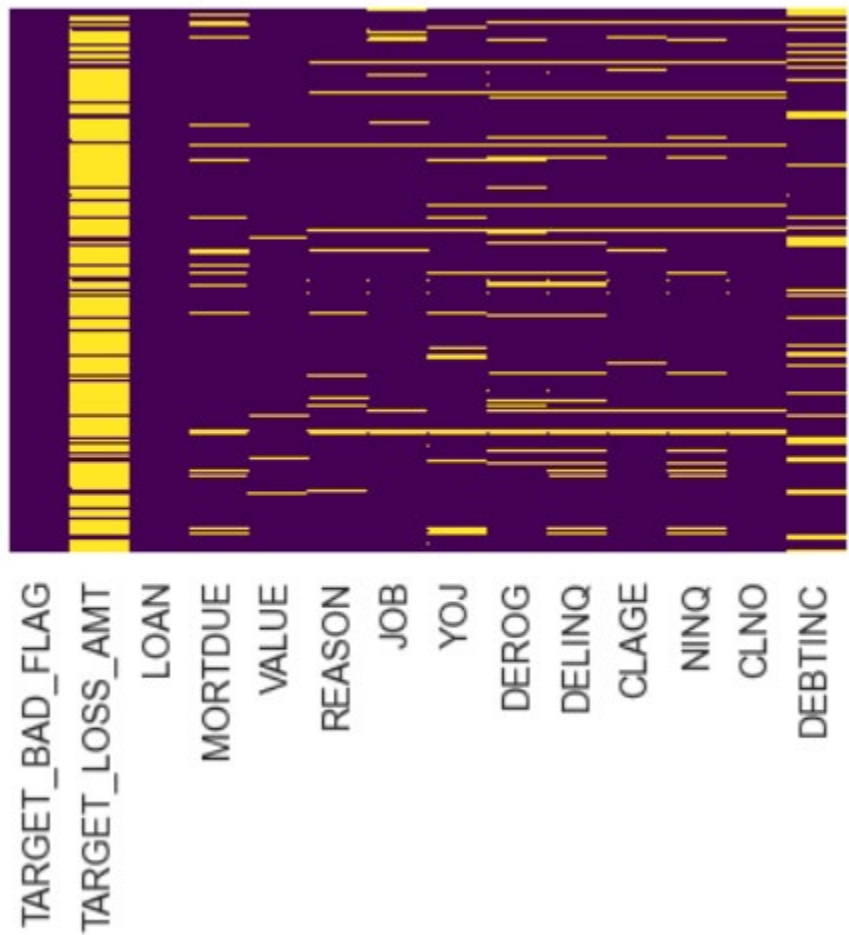
The second target, TARGET_LOSS_AMT, indicates the amount of money that was lost for loans that went bad. The remaining variables contain information about the customer at the time that the loan was issued.

The below descriptive statistics shows:

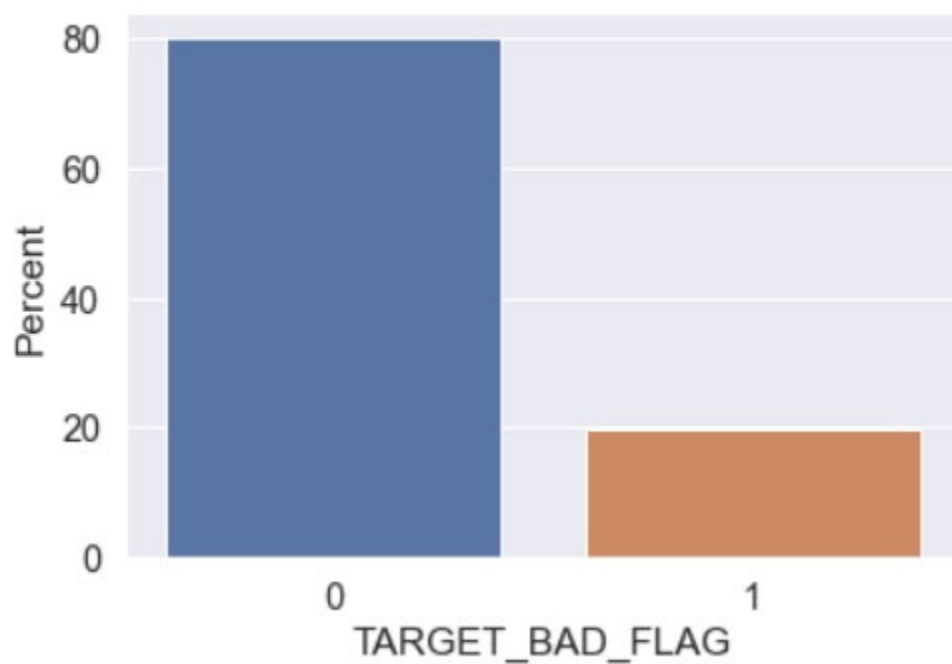
	count	mean	std	min \
TARGET_BAD_FLAG	5960.0	0.199497	0.399656	0.000000
TARGET_LOSS_AMT	1189.0	13414.576955	10839.455965	224.000000
LOAN	5960.0	18607.969799	11207.480417	1100.000000
MORTDUE	5442.0	73760.817200	44457.609458	2063.000000
VALUE	5848.0	101776.048741	57385.775334	8000.000000
YOJ	5445.0	8.922268	7.573982	0.000000
DEROG	5252.0	0.254570	0.846047	0.000000
DELINQ	5380.0	0.449442	1.127266	0.000000
CLAGE	5652.0	179.766275	85.810092	0.000000
NINQ	5450.0	1.186055	1.728675	0.000000
CLNO	5738.0	21.296096	10.138933	0.000000
DEBTINC	4693.0	33.779915	8.601746	0.524499

	25%	50%	75%	max
TARGET_BAD_FLAG	0.000000	0.000000	0.000000	1.000000
TARGET_LOSS_AMT	5639.000000	11003.000000	17634.000000	78987.000000
LOAN	11100.000000	16300.000000	23300.000000	89900.000000
MORTDUE	46276.000000	65019.000000	91488.000000	399550.000000
VALUE	66075.500000	89235.500000	119824.250000	855909.000000
YOJ	3.000000	7.000000	13.000000	41.000000
DEROG	0.000000	0.000000	0.000000	10.000000
DELINQ	0.000000	0.000000	0.000000	15.000000
CLAGE	115.116702	173.466667	231.562278	1168.233561
NINQ	0.000000	1.000000	2.000000	17.000000
CLNO	15.000000	20.000000	26.000000	71.000000
DEBTINC	29.140031	34.818262	39.003141	203.312149

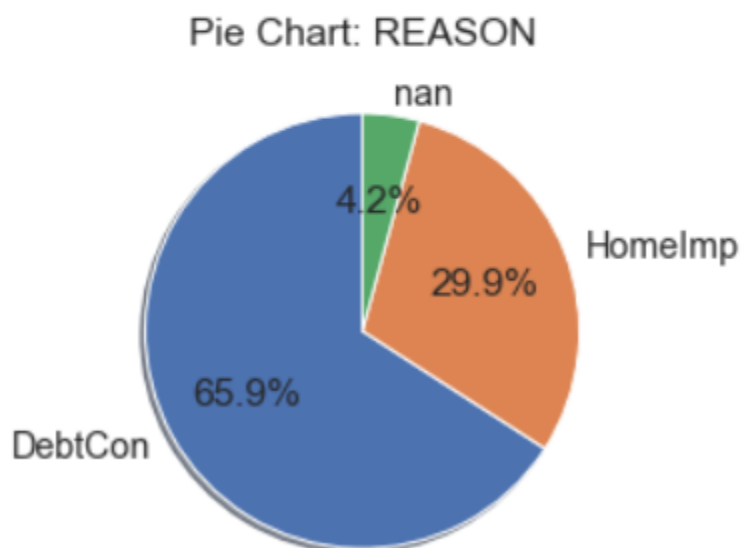
Heat Map of the Missing Values:



Exploratory Data Analysis of the HMEQ dataset



```
DebtCon    3928
HomeImp    1780
NaN         252
Name: REASON, dtype: int64
['DebtCon', 'HomeImp', nan]
[3928, 1780, 252]
```



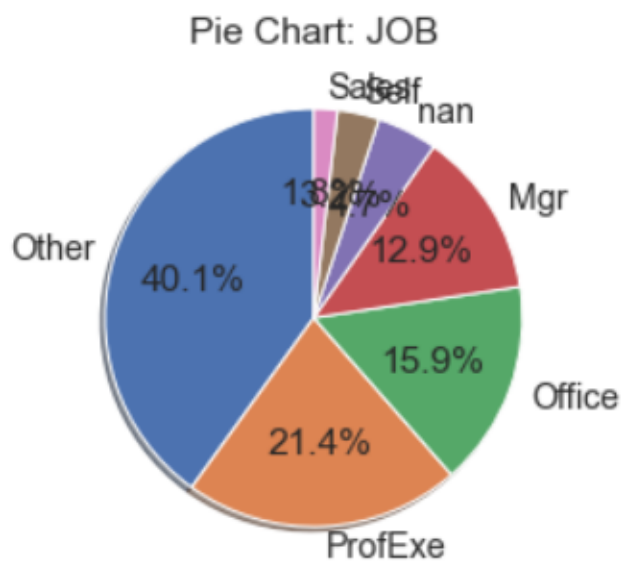
Exploratory Data Analysis of the HMEQ dataset

```
Other      2388
ProfExe    1276
Office     948
Mgr         767
NaN         279
Self        193
Sales       109
```

```
Name: JOB, dtype: int64
```

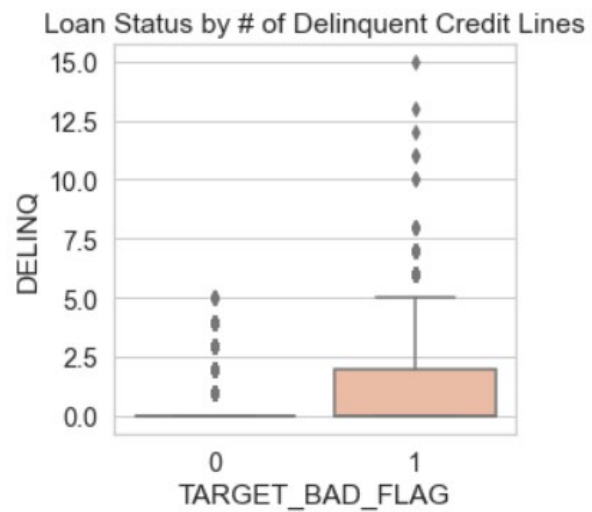
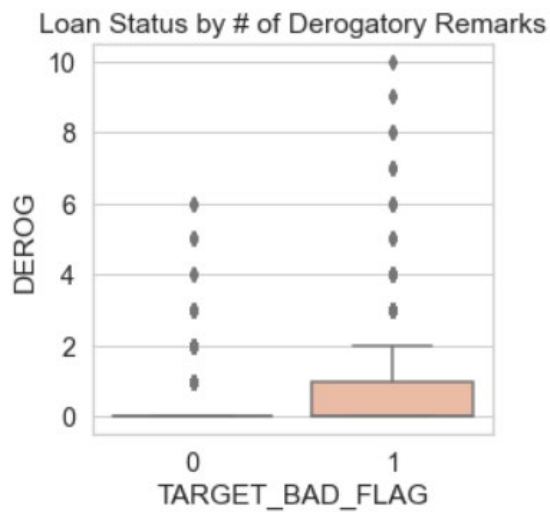
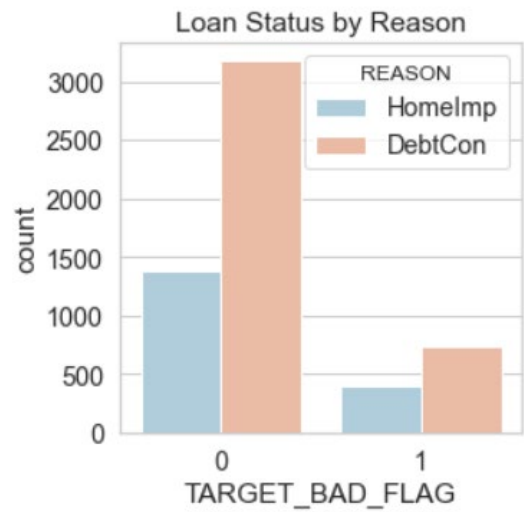
```
['Other', 'ProfExe', 'Office', 'Mgr', nan, 'Self', 'Sales']
```

```
[2388, 1276, 948, 767, 279, 193, 109]
```

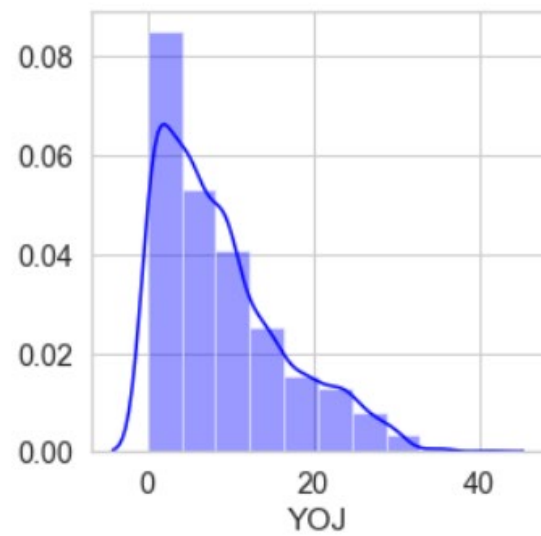
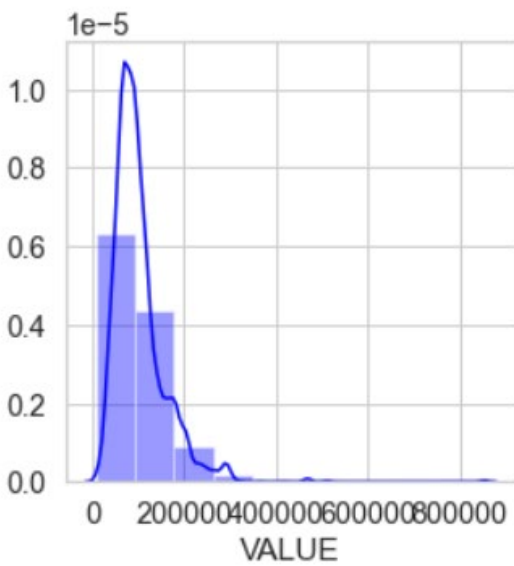
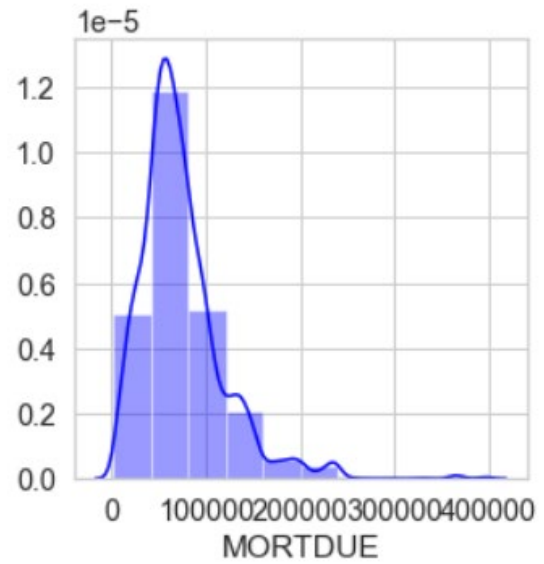
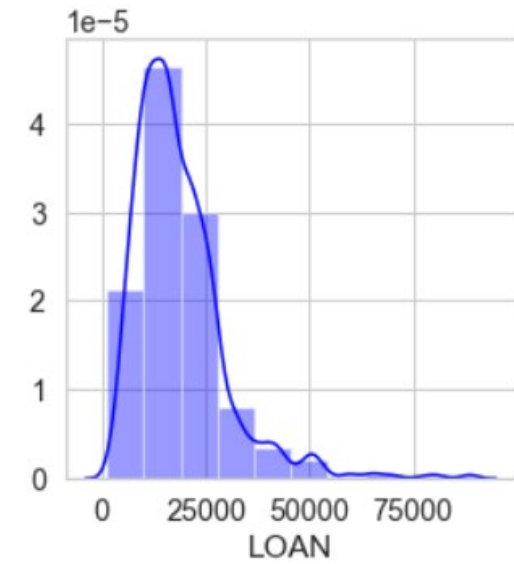


The pie chart shows how the categorical values are split in the dataset.

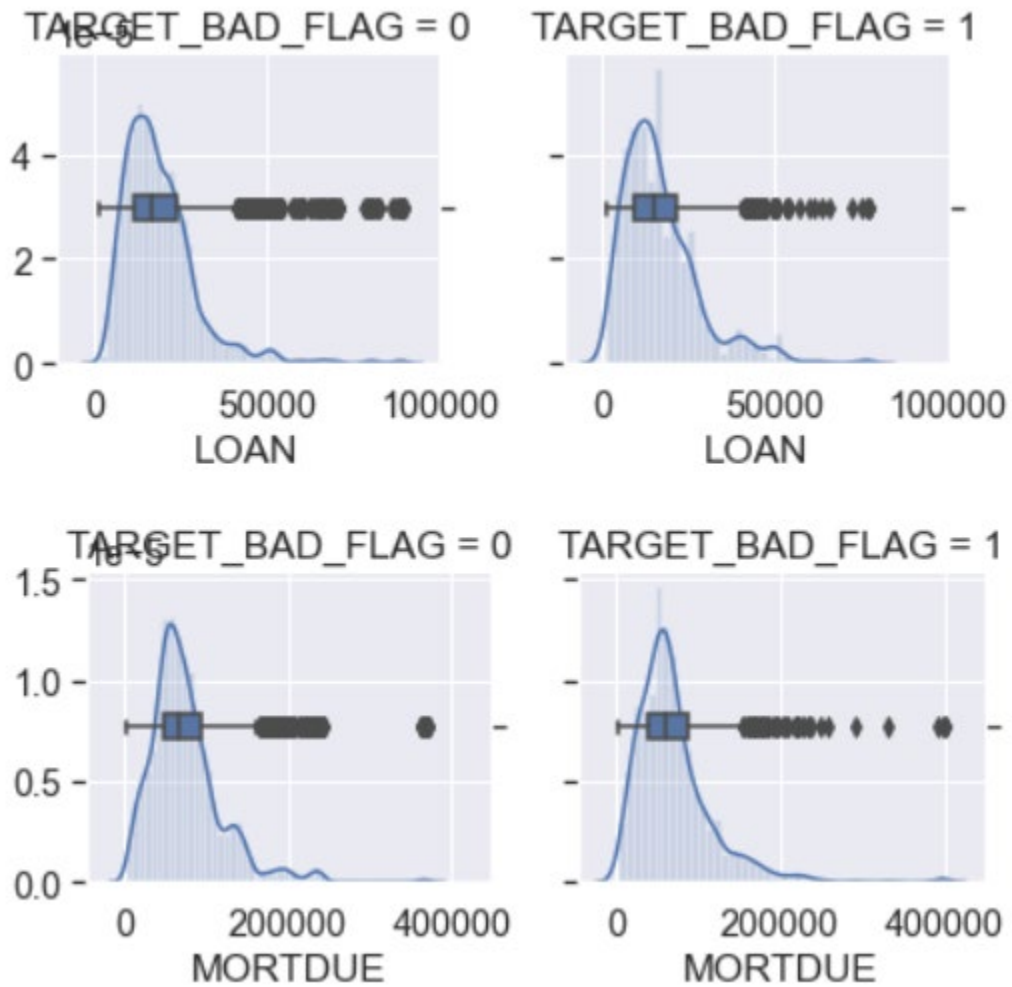
Exploratory Data Analysis of the HMEQ dataset



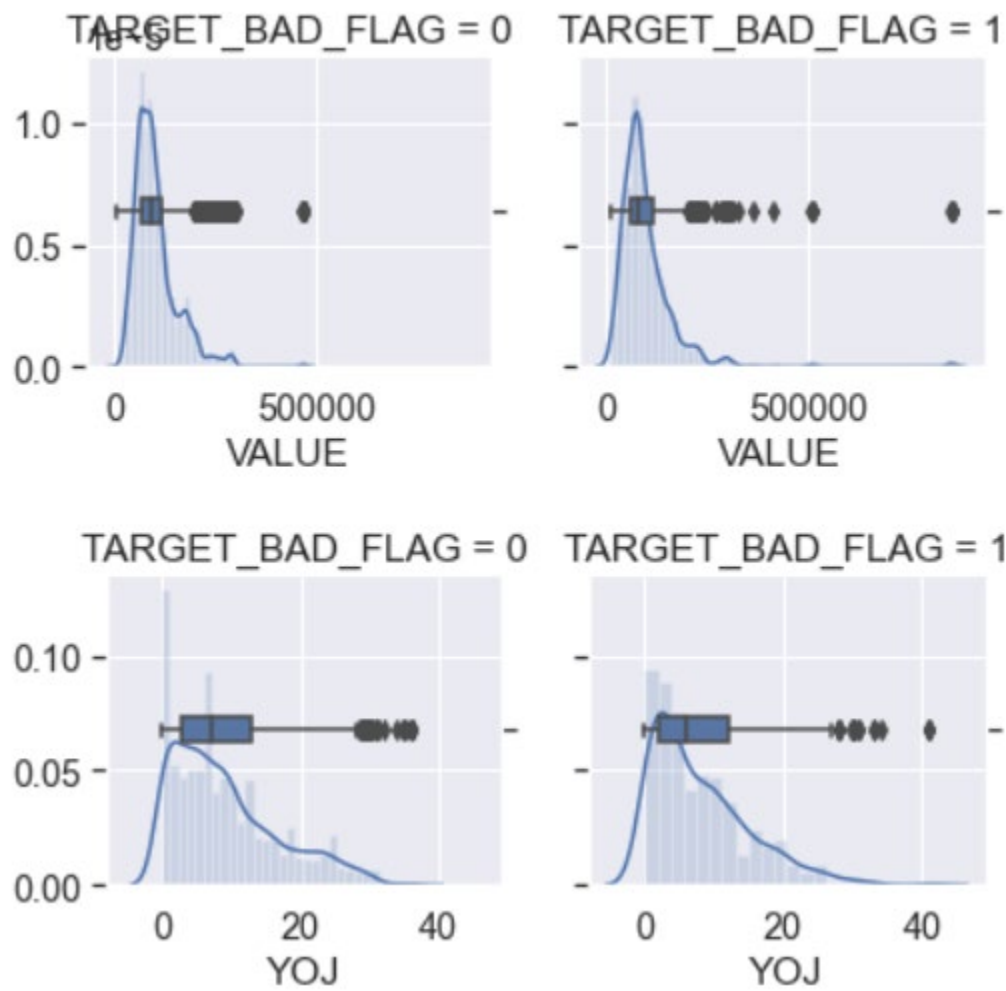
Exploratory Data Analysis of the HMEQ dataset



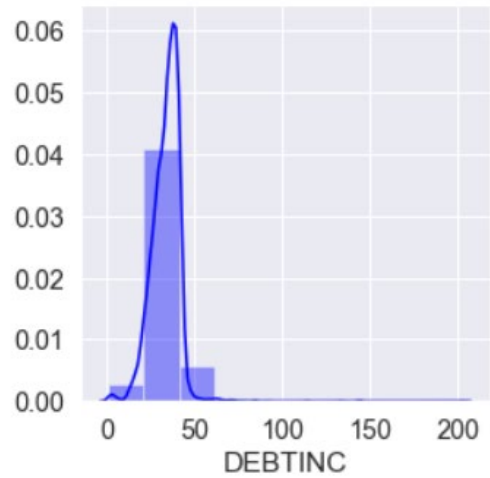
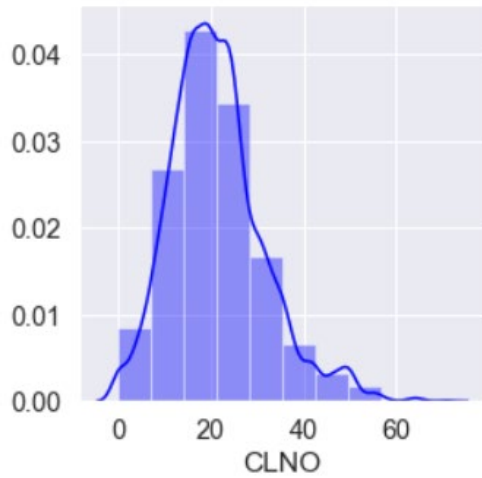
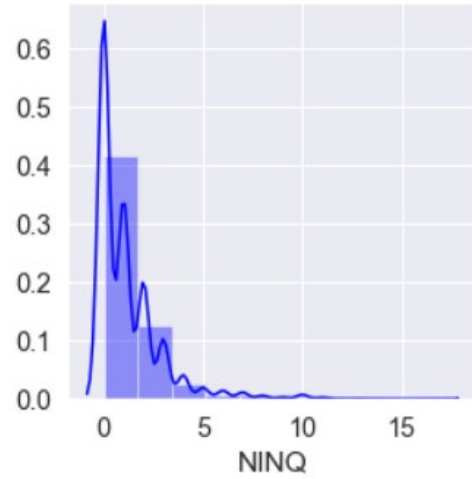
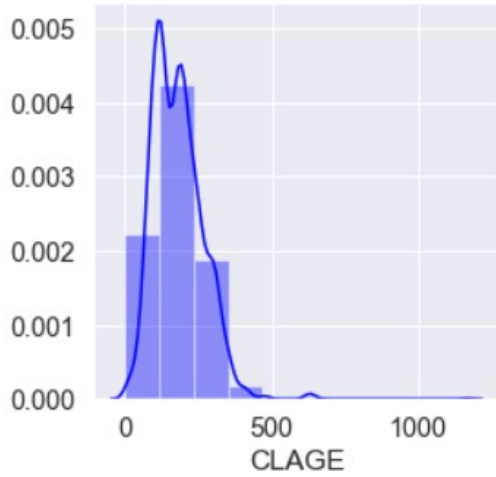
Exploratory Data Analysis of the HMEQ dataset

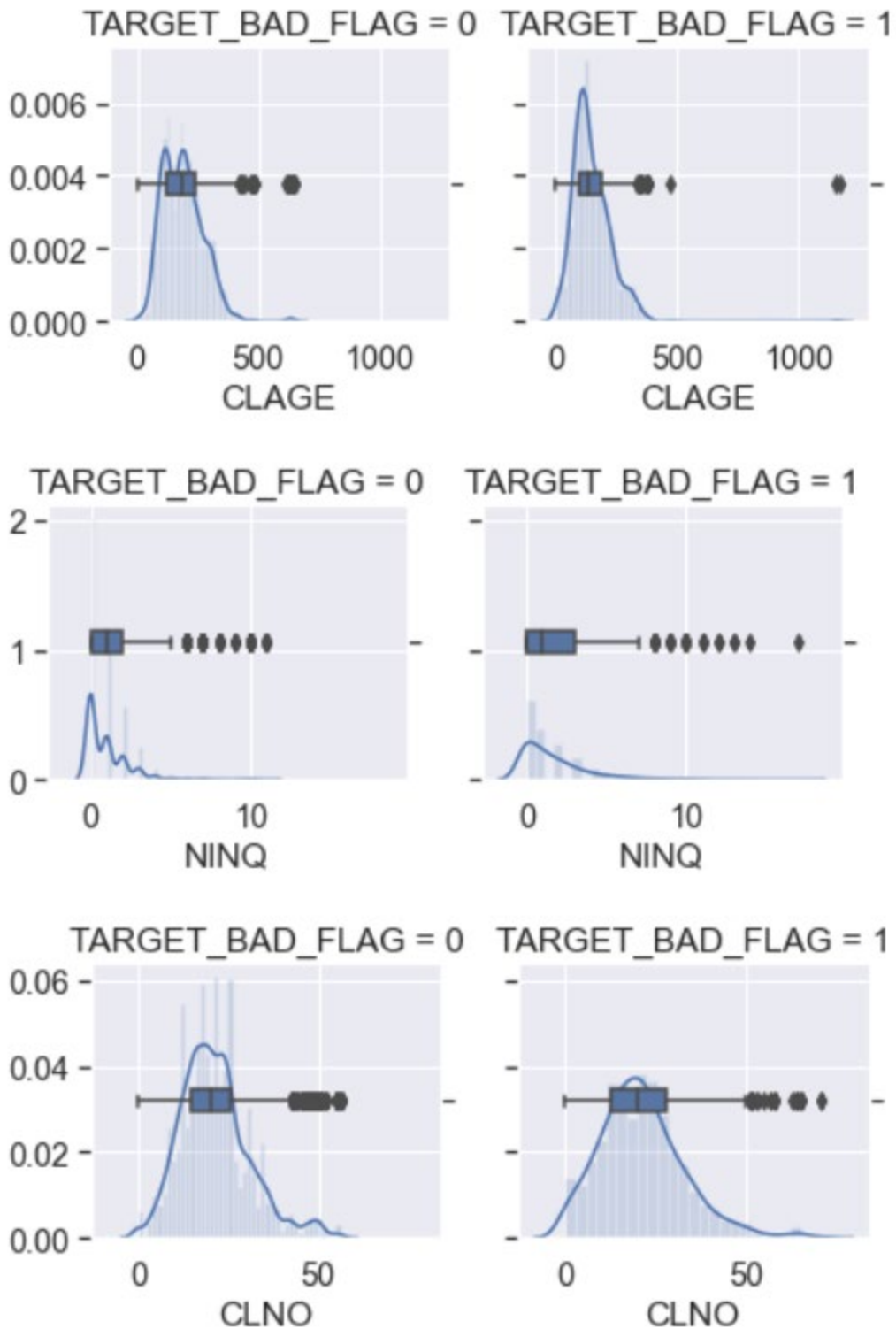


Exploratory Data Analysis of the HMEQ dataset



Exploratory Data Analysis of the HMEQ dataset





Exploratory Data Analysis of the HMEQ dataset

