

```
In [77]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
```

```
In [78]: sns.set()
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', -1)
```

```
In [79]: os.chdir("/Users/serrauzun/Desktop/MSDS_422_Practical")
```

```
In [80]: df = pd.read_csv('HMEQ_Loss_clean.csv')
```

```
In [81]: print(df.head())
```

```
    TARGET_BAD_FLAG  TARGET_LOSS_AMT  LOAN  MORTDUE  VALUE  YOJ  DEROG  \
0      1             641.0       1100  25860.0  39025.0  10.5  0.0
1      1            1109.0       1300  70053.0  68400.0   7.0  0.0
2      1             767.0       1500  13500.0  16700.0   4.0  0.0
3      1            1425.0       1500  65019.0  89235.5   7.0  0.0
4      0              0.0       1700  97800.0 112000.0   3.0  0.0

  DELINQ      CLAGE  NINQ  CLNO  DEBTINC  REASON_DebtCon  REASON_HomeImp  \
0  0.0    94.366667  1.0   9.0   0.0       0                  1
1  2.0   121.833333  0.0  14.0   0.0       0                  1
2  0.0   149.466667  1.0  10.0   0.0       0                  1
3  0.0   173.466667  1.0  20.0   0.0       0                  0
4  0.0   93.333333  0.0  14.0   0.0       0                  1

  REASON_Missing  JOB_Mgr  JOB_Missing  JOB_Office  JOB_Other  JOB_ProfExe  \
0      0            0           0           0           1           0
1      0            0           0           0           1           0
2      0            0           0           0           1           0
3      1            0           1           0           0           0
4      2            2           2           1           2           2
```

Separate binary columns to different df to first drop then add

```
In [82]: df_binary = df[['TARGET_BAD_FLAG', 'REASON_DebtCon', 'REASON_HomeImp', 'REASON_Missing', 'JOB_Mgr', 'JOB_Missing', 'JOB_Office', 'JOB_Other', 'JOB_ProfExe']]
```

```
In [83]: print(df_binary.head())
```

```
   TARGET_BAD_FLAG  REASON_DebtCon  REASON_HomeImp  REASON_Missing  JOB_Mgr  \
0          1             0             1             0             0
1          1             0             1             0             0
2          1             0             1             0             0
3          1             0             0             1             0
4          0             0             1             0             0

   JOB_Missing  JOB_Office  JOB_Other  JOB_ProfExe  JOB_Sales  JOB_Self
0            0           0           1           0           0           0
1            0           0           1           0           0           0
2            0           0           1           0           0           0
3            1           0           0           0           0           0
4            0           1           0           0           0           0
```

```
In [84]: x = df.copy()
x = x.drop(df_binary, axis=1)
varNames = x.columns
```

```
In [86]: print(varNames)
```

```
Index(['TARGET_LOSS_AMT', 'LOAN', 'MORTDUE', 'VALUE', 'Y0J', 'DEROG', 'DELINQ',
       'CLAGE', 'NINQ', 'CLNO', 'DEBTINC'],
      dtype='object')
```

```
In [88]: print(x.head())
```

	TARGET_LOSS_AMT	LOAN	MORTDUE	VALUE	Y0J	DEROG	DELINQ	CLAGE	\
0	641.0	1100	25860.0	39025.0	10.5	0.0	0.0	94.366667	
1	1109.0	1300	70053.0	68400.0	7.0	0.0	2.0	121.833333	
2	767.0	1500	13500.0	16700.0	4.0	0.0	0.0	149.466667	
3	1425.0	1500	65019.0	89235.5	7.0	0.0	0.0	173.466667	
4	0.0	1700	97800.0	112000.0	3.0	0.0	0.0	93.333333	

	NINQ	CLNO	DEBTINC
0	1.0	9.0	0.0
1	0.0	14.0	0.0
2	1.0	10.0	0.0
3	1.0	20.0	0.0
4	0.0	14.0	0.0

```
In [89]: print(x.head())
print(x.describe())
print( "\n\n")
```

	TARGET_LOSS_AMT	LOAN	MORTDUE	VALUE	Y0J	DEROG	DELINQ	CLAGE	\
0	641.0	1100	25860.0	39025.0	10.5	0.0	0.0	94.366667	
1	1109.0	1300	70053.0	68400.0	7.0	0.0	2.0	121.833333	
2	767.0	1500	13500.0	16700.0	4.0	0.0	0.0	149.466667	
3	1425.0	1500	65019.0	89235.5	7.0	0.0	0.0	173.466667	
4	0.0	1700	97800.0	112000.0	3.0	0.0	0.0	93.333333	

	NINQ	CLNO	DEBTINC
0	1.0	9.0	0.0
1	0.0	14.0	0.0
2	1.0	10.0	0.0
3	1.0	20.0	0.0
4	0.0	14.0	0.0

	TARGET_LOSS_AMT	LOAN	MORTDUE	VALUE	\
count	5861.000000	5861.000000	5861.000000	5861.000000	
mean	2607.528579	18294.898481	70234.565467	98396.641280	
std	7040.238782	10925.557580	36058.418382	46870.608247	

```
min    0.000000      1100.000000    2063.000000    8000.000000
25%   0.000000      11000.000000   47882.000000   66379.000000
50%   0.000000      16100.000000   65019.000000   89235.500000
75%   0.000000      22900.000000   86236.000000   117650.000000
max   78987.000000    89900.000000   228006.000000  289430.000000
```

	Y0J	DEROG	DELINQ	CLAGE	NINQ	\
count	5861.000000	5861.000000	5861.000000	5861.000000	5861.000000	
mean	8.739473	0.226241	0.403856	178.812196	1.157652	
std	7.215748	0.804001	1.081892	83.665203	1.650645	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	3.000000	0.000000	0.000000	116.818648	0.000000	
50%	7.000000	0.000000	0.000000	173.466667	1.000000	
75%	12.000000	0.000000	0.000000	226.040295	2.000000	
max	41.000000	10.000000	15.000000	1168.233561	17.000000	

	CLNO	DEBTINC
count	5861.000000	5861.000000
mean	21.069613	26.515397
std	9.774048	15.720958
min	0.000000	0.000000
25%	15.000000	20.260198
50%	20.000000	31.874217
75%	26.000000	37.882260
max	71.000000	203.312149

** NORMALIZING THE DATA **

In [90]: `theScaler = MinMaxScaler()
theScaler.fit(x)`

Out[90]: `MinMaxScaler(copy=True, feature_range=(0, 1))`

```
In [91]: x_MINMAX = theScaler.transform(x)
x_MINMAX = pd.DataFrame(x_MINMAX)
print(x_MINMAX.head())
print("\n\n")
```

	0	1	2	3	4	5	6	7	\
0	0.008115	0.000000	0.105323	0.110241	0.256098	0.0	0.000000	0.080777	
1	0.014040	0.002252	0.300917	0.214618	0.170732	0.0	0.133333	0.104289	
2	0.009710	0.004505	0.050619	0.030914	0.097561	0.0	0.000000	0.127942	
3	0.018041	0.004505	0.278637	0.288653	0.170732	0.0	0.000000	0.148486	
4	0.000000	0.006757	0.423722	0.369541	0.073171	0.0	0.000000	0.079893	

	8	9	10
0	0.058824	0.126761	0.0
1	0.000000	0.197183	0.0
2	0.058824	0.140845	0.0
3	0.058824	0.281690	0.0
4	0.000000	0.197183	0.0

```
In [92]: varNames_minmax = []
for i in varNames:
    newName = "nor_" + i
    varNames_minmax.append( newName )
print(varNames_minmax)
print( "\n\n")
```

```
['nor_TARGET_LOSS_AMT', 'nor_LOAN', 'nor_MORTDUE', 'nor_VALUE', 'nor_YOJ', 'nor_DEROG', 'nor_DELINQ', 'nor_CLAGE', 'nor_NINQ', 'nor_CLNO', 'nor_DEBTINC']
```

```
In [93]: x_MINMAX.columns = varNames_minmax
print(x_MINMAX.head())
print( "\n\n")
```

	nor_TARGET_LOSS_AMT	nor_LOAN	nor_MORTDUE	nor_VALUE	nor_Y0J	nor_DEROG	\
0	0.008115	0.000000	0.105323	0.110241	0.256098	0.0	
1	0.014040	0.002252	0.300917	0.214618	0.170732	0.0	
2	0.009710	0.004505	0.050619	0.030914	0.097561	0.0	
3	0.018041	0.004505	0.278637	0.288653	0.170732	0.0	
4	0.000000	0.006757	0.423722	0.369541	0.073171	0.0	

	nor_DELINQ	nor_CLAGE	nor_NINQ	nor_CLNO	nor_DEBTINC	
0	0.000000	0.080777	0.058824	0.126761	0.0	
1	0.133333	0.104289	0.000000	0.197183	0.0	
2	0.000000	0.127942	0.058824	0.140845	0.0	
3	0.000000	0.148486	0.058824	0.281690	0.0	
4	0.000000	0.079893	0.000000	0.197183	0.0	

```
In [94]: print(x_MINMAX.describe())
print("\n\n")
```

	nor_TARGET_LOSS_AMT	nor_LOAN	nor_MORTDUE	nor_VALUE	\
count	5861.000000	5861.000000	5861.000000	5861.000000	
mean	0.033012	0.193636	0.301720	0.321205	
std	0.089132	0.123036	0.159591	0.166544	
min	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.111486	0.202790	0.207437	
50%	0.000000	0.168919	0.278637	0.288653	
75%	0.000000	0.245495	0.372541	0.389617	
max	1.000000	1.000000	1.000000	1.000000	

	nor_YOJ	nor_DEROG	nor_DELINQ	nor_CLAGE	nor_NINQ	\
count	5861.000000	5861.000000	5861.000000	5861.000000	5861.000000	
mean	0.213158	0.022624	0.026924	0.153062	0.068097	
std	0.175994	0.080400	0.072126	0.071617	0.097097	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.073171	0.000000	0.000000	0.099996	0.000000	
50%	0.170732	0.000000	0.000000	0.148486	0.058824	
75%	0.292683	0.000000	0.000000	0.193489	0.117647	
max	1.000000	1.000000	1.000000	1.000000	1.000000	

```
In [95]: x_new = pd.concat([x, x_MINMAX, df_binary], axis=1)
```

```
In [96]: x_new.head()
```

Out[96]:

	TARGET_LOSS_AMT	LOAN	MORTDUE	VALUE	YOJ	DEROG	DELINQ	CLAGE	NINQ	CLNO	DEBTINC	nor_TARGET
0	641.0	1100	25860.0	39025.0	10.5	0.0	0.0	94.366667	1.0	9.0	0.0	0.0
1	1109.0	1300	70053.0	68400.0	7.0	0.0	2.0	121.833333	0.0	14.0	0.0	0.0
2	767.0	1500	13500.0	16700.0	4.0	0.0	0.0	149.466667	1.0	10.0	0.0	0.0
3	1425.0	1500	65019.0	89235.5	7.0	0.0	0.0	173.466667	1.0	20.0	0.0	0.0
4	0.0	1700	97800.0	112000.0	3.0	0.0	0.0	93.333333	0.0	14.0	0.0	0.0

** STANDARDIZING THE DATA **

```
In [97]: theScaler = StandardScaler()  
theScaler.fit(x)
```

Out[97]: StandardScaler(copy=True, with_mean=True, with_std=True)

```
In [98]: y_std = theScaler.transform(x)
y_std = pd.DataFrame(y_std)
print(y_std.head())
print( "\n\n" )
```

	0	1	2	3	4	5	6	\
0	-0.279351	-1.573958	-1.230735	-1.266822	0.244005	-0.281418	-0.373319	
1	-0.212870	-1.555650	-0.005036	-0.640043	-0.241087	-0.281418	1.475452	
2	-0.261452	-1.537343	-1.573541	-1.743174	-0.656880	-0.281418	-0.373319	
3	-0.167981	-1.537343	-0.144654	-0.195473	-0.241087	-0.281418	-0.373319	
4	-0.370407	-1.519036	0.764531	0.290257	-0.795477	-0.281418	-0.373319	

	7	8	9	10	
0	-1.009413	-0.095518	-1.234969	-1.686771	
1	-0.681092	-0.701393	-0.723366	-1.686771	
2	-0.350779	-0.095518	-1.132648	-1.686771	
3	-0.063897	-0.095518	-0.109443	-1.686771	
4	-1.021765	-0.701393	-0.723366	-1.686771	

```
In [99]: varNames_std = []
for i in varNames:
    newName = "std_" + i
    varNames_std.append( newName )
```

```
In [100]: y_std.columns = varNames_std
print(y_std.head())
print("\n\n")
```

	std_TARGET_LOSS_AMT	std_LOAN	std_MORTDUE	std_VALUE	std_Y0J	std_DEROG	\
0	-0.279351	-1.573958	-1.230735	-1.266822	0.244005	-0.281418	
1	-0.212870	-1.555650	-0.005036	-0.640043	-0.241087	-0.281418	
2	-0.261452	-1.537343	-1.573541	-1.743174	-0.656880	-0.281418	
3	-0.167981	-1.537343	-0.144654	-0.195473	-0.241087	-0.281418	
4	-0.370407	-1.519036	0.764531	0.290257	-0.795477	-0.281418	

	std_DELINQ	std_CLAGE	std_NINQ	std_CLNO	std_DEBTINC	
0	-0.373319	-1.009413	-0.095518	-1.234969	-1.686771	
1	1.475452	-0.681092	-0.701393	-0.723366	-1.686771	
2	-0.373319	-0.350779	-0.095518	-1.132648	-1.686771	
3	-0.373319	-0.063897	-0.095518	-0.109443	-1.686771	
4	-0.373319	-1.021765	-0.701393	-0.723366	-1.686771	

```
In [101]: print(y_std.describe())
print("\n\n")
```

	std_TARGET_LOSS_AMT	std_LOAN	std_MORTDUE	std_VALUE	\	
count	5.861000e+03	5.861000e+03	5.861000e+03	5.861000e+03		
mean	7.420537e-15	-4.178879e-15	3.785480e-16	3.050651e-16		
std	1.000085e+00	1.000085e+00	1.000085e+00	1.000085e+00		
min	-3.704066e-01	-1.573958e+00	-1.890748e+00	-1.928807e+00		
25%	-3.704066e-01	-6.677482e-01	-6.199516e-01	-6.831653e-01		
50%	-3.704066e-01	-2.009129e-01	-1.446544e-01	-1.954727e-01		
75%	-3.704066e-01	4.215340e-01	4.438020e-01	4.108119e-01		
max	1.084991e+01	6.554467e+00	4.375813e+00	4.076108e+00		
	std_Y0J	std_DEROG	std_DELINQ	std_CLAGE	std_NINQ	\
count	5.861000e+03	5.861000e+03	5.861000e+03	5.861000e+03	5.861000e+03	
mean	3.540553e-16	-4.045561e-16	1.258430e-15	-1.902211e-16	6.565110e-16	
std	1.000085e+00	1.000085e+00	1.000085e+00	1.000085e+00	1.000085e+00	
min	-1.211270e+00	-2.814184e-01	-3.733187e-01	-2.137417e+00	-7.013930e-01	
25%	-7.954771e-01	-2.814184e-01	-3.733187e-01	-7.410349e-01	-7.013930e-01	
50%	-2.410867e-01	-2.814184e-01	-3.733187e-01	-6.389736e-02	-9.551764e-02	
75%	4.519012e-01	-2.814184e-01	-3.733187e-01	5.645373e-01	5.103578e-01	
max	4.471232e+00	1.215744e+01	1.349246e+01	1.182697e+01	9.598489e+00	
	std_CLNO	std_DEBTINC				
count	5.861000e+03	5.861000e+03				
mean	4.589402e-16	-1.952447e-15				
std	1.000085e+00	1.000085e+00				
min	-2.155853e+00	-1.686771e+00				
25%	-6.210457e-01	-3.979231e-01				
50%	-1.094433e-01	3.409002e-01				
75%	5.044796e-01	7.231005e-01				
max	5.108901e+00	1.124689e+01				

```
In [102]: y_new = pd.concat([x, y_std, df_binary], axis=1)
print(y_new.head())
```

	TARGET_LOSS_AMT	LOAN	MORTDUE	VALUE	Y0J	DEROG	DELINQ	CLAGE	\
0	641.0	1100	25860.0	39025.0	10.5	0.0	0.0	94.366667	
1	1109.0	1300	70053.0	68400.0	7.0	0.0	2.0	121.833333	
2	767.0	1500	13500.0	16700.0	4.0	0.0	0.0	149.466667	
3	1425.0	1500	65019.0	89235.5	7.0	0.0	0.0	173.466667	
4	0.0	1700	97800.0	112000.0	3.0	0.0	0.0	93.333333	
	NINQ	CLNO	DEBTINC	std_TARGET_LOSS_AMT	std_LOAN	std_MORTDUE	std_VALUE		\
0	1.0	9.0	0.0	-0.279351	-1.573958	-1.230735	-1.266822		
1	0.0	14.0	0.0	-0.212870	-1.555650	-0.005036	-0.640043		
2	1.0	10.0	0.0	-0.261452	-1.537343	-1.573541	-1.743174		
3	1.0	20.0	0.0	-0.167981	-1.537343	-0.144654	-0.195473		
4	0.0	14.0	0.0	-0.370407	-1.519036	0.764531	0.290257		
	std_Y0J	std_DEROG	std_DELINQ	std_CLAGE	std_NINQ	std_CLNO		\	
0	0.244005	-0.281418	-0.373319	-1.009413	-0.095518	-1.234969			
1	-0.241087	-0.281418	1.475452	-0.681092	-0.701393	-0.723366			
2	-0.656880	-0.281418	-0.373319	-0.350779	-0.095518	-1.132648			
3	-0.241087	-0.281418	-0.373319	-0.063897	-0.095518	-0.109443			
4	-0.795477	-0.281418	-0.373319	-1.021765	-0.701393	-0.723366			
	std_DEBTINC	TARGET_BAD_FLAG	REASON_DebtCon	REASON_HomeImp				\	
0	-1.686771	1	0	1					
1	-1.686771	1	0	1					
2	-1.686771	1	0	1					
3	-1.686771	1	0	0					
4	-1.686771	0	0	1					
	REASON_Missing	JOB_Mgr	JOB_Missing	JOB_Office	JOB_Other	JOB_ProfExe		\	
0	0	0	0	0	1	0			
1	0	0	0	0	1	0			
2	0	0	0	0	1	0			
3	1	0	1	0	0	0			
4	0	0	0	1	0	0			
	JOB_Sales	JOB_Self							
0	0	0							

1	0	0
2	0	0
3	0	0
4	0	0