Viya-Verse-2024-06

"Data Exploration" Results

by: David Weik

Contents

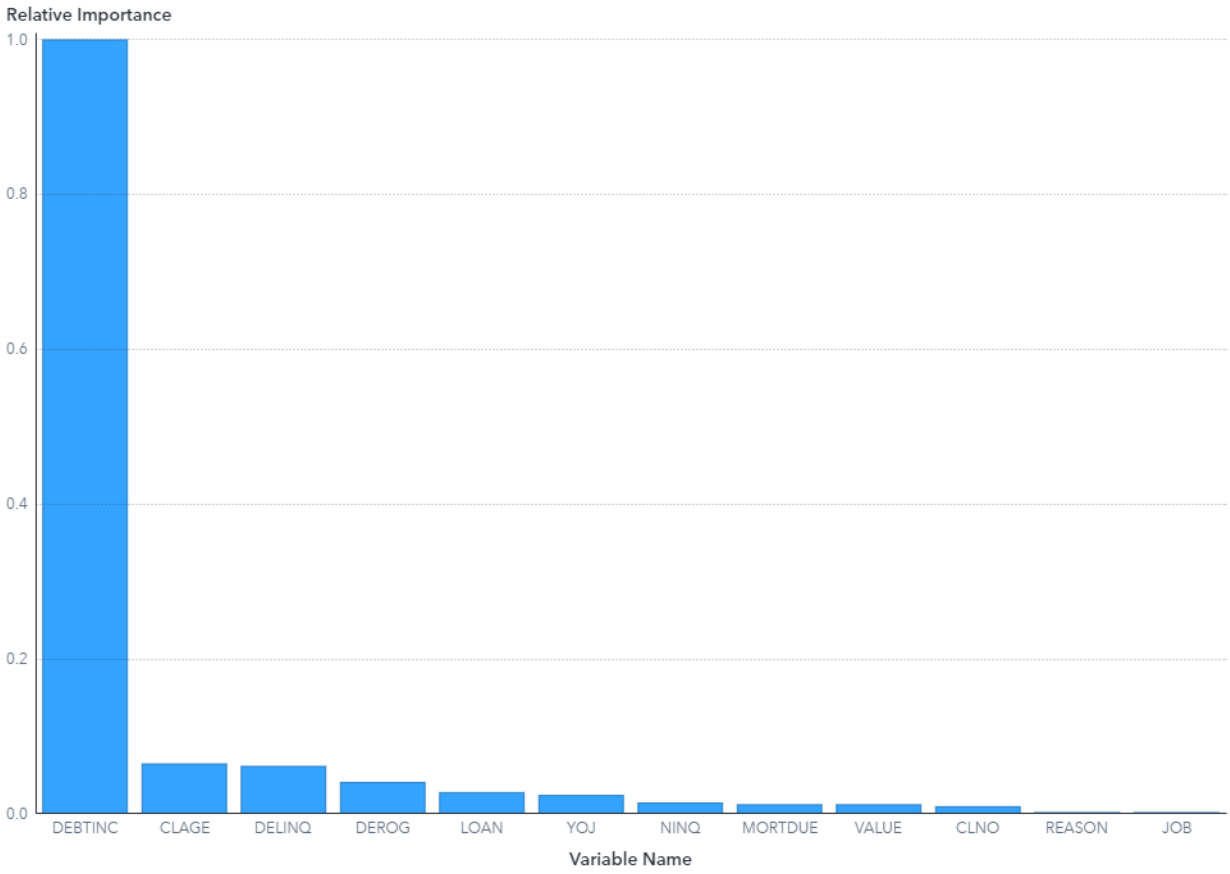
Data Partition Summary 4
Important Inputs 5
Cardinality 6
Mode Percent 7
BAD 8
DELINQ 9
DEROG 10
JOB 11
NINQ 12
REASON 13
Interval Variable Moments 14
Deviation from Normality 15
Relative Variability 16
CLAGE 17
CLNO 18
DEBTINC 19
LOAN 20
MORTDUE 21
VALUE 22
YOJ 23
Missing Values 24
Mutual Information 25
Symmetric Uncertainty 26
Entropy Reduction 27
CLAGE 28
CLNO 29
DEBTINC 30
DELINQ 31
DEROG 32
JOB 33

LOAN 34
MORTDUE 35
NINQ 36
REASON 37
VALUE 38
YOJ 39
Pearson (Absolute) 40
Table 41
Mutual Information 42
Symmetric Uncertainty 43
Entropy Reduction 44
Table 45
Properties 50
Output 51

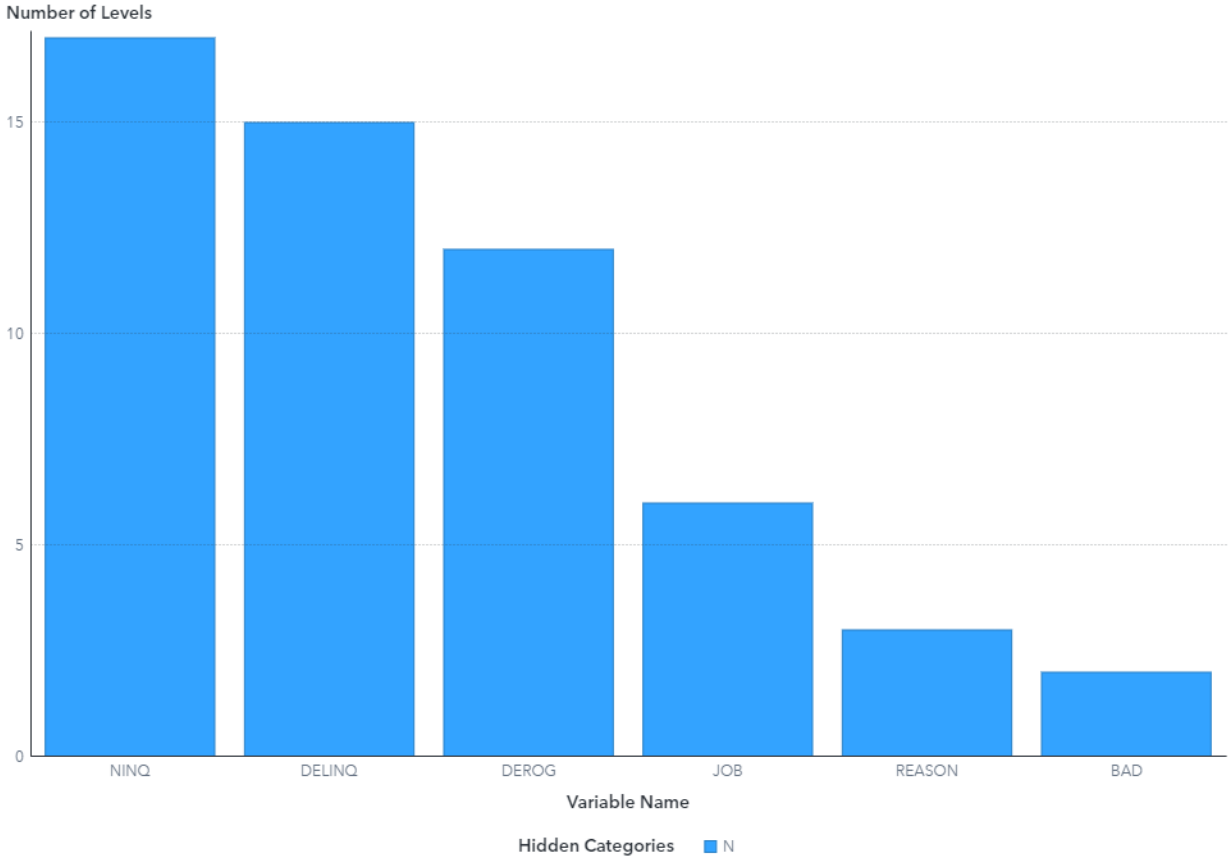
Data Partition Summary

Partition Used	Number of Observations
All data	5,960

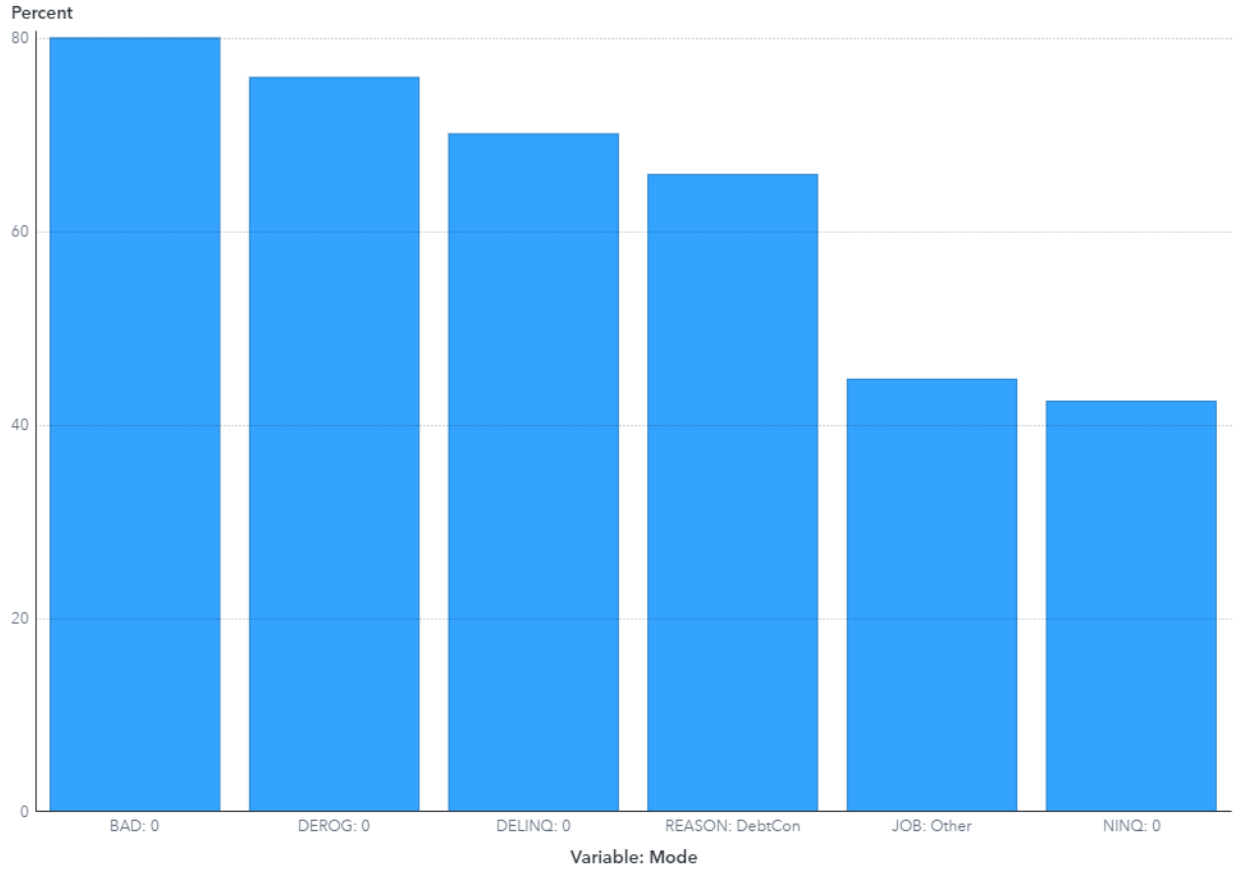
Important Inputs



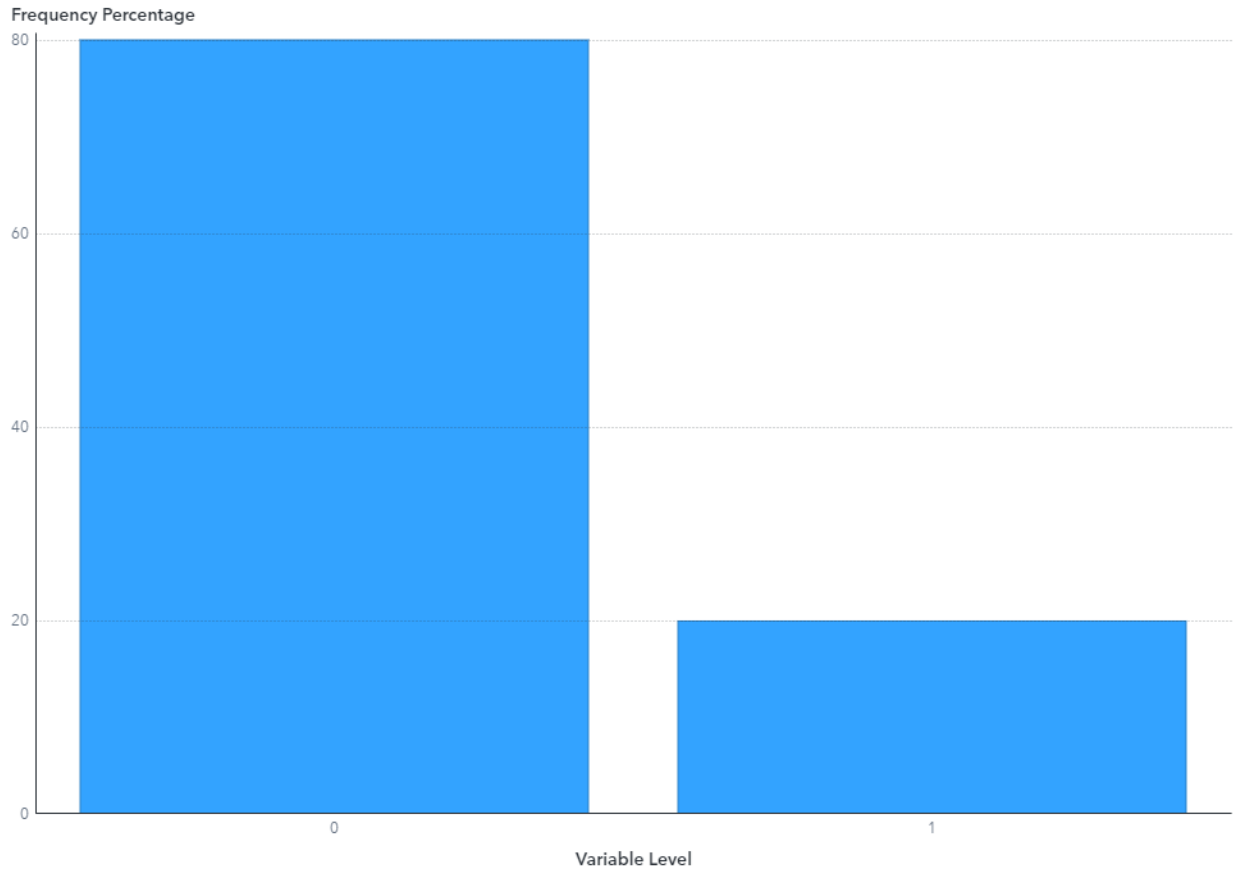
Cardinality



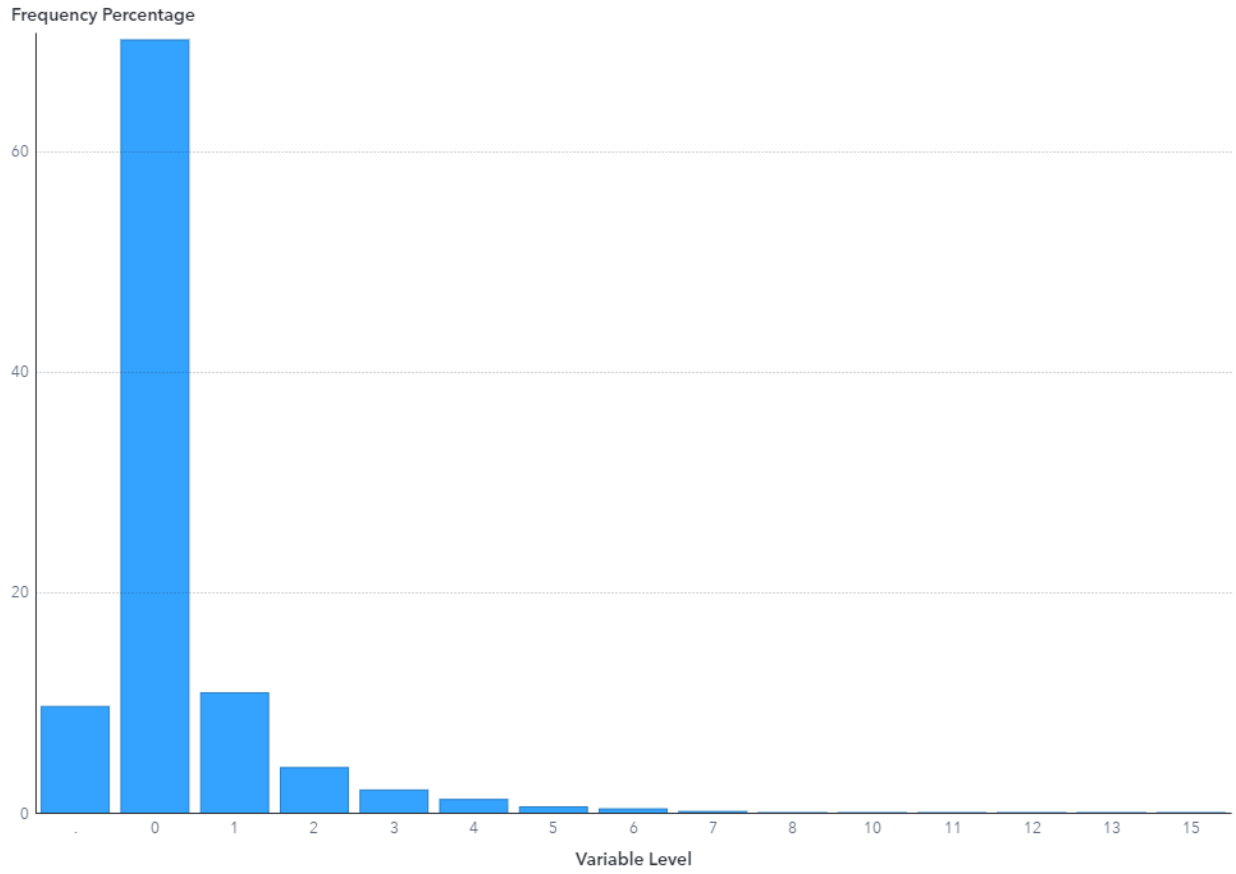
Mode Percent



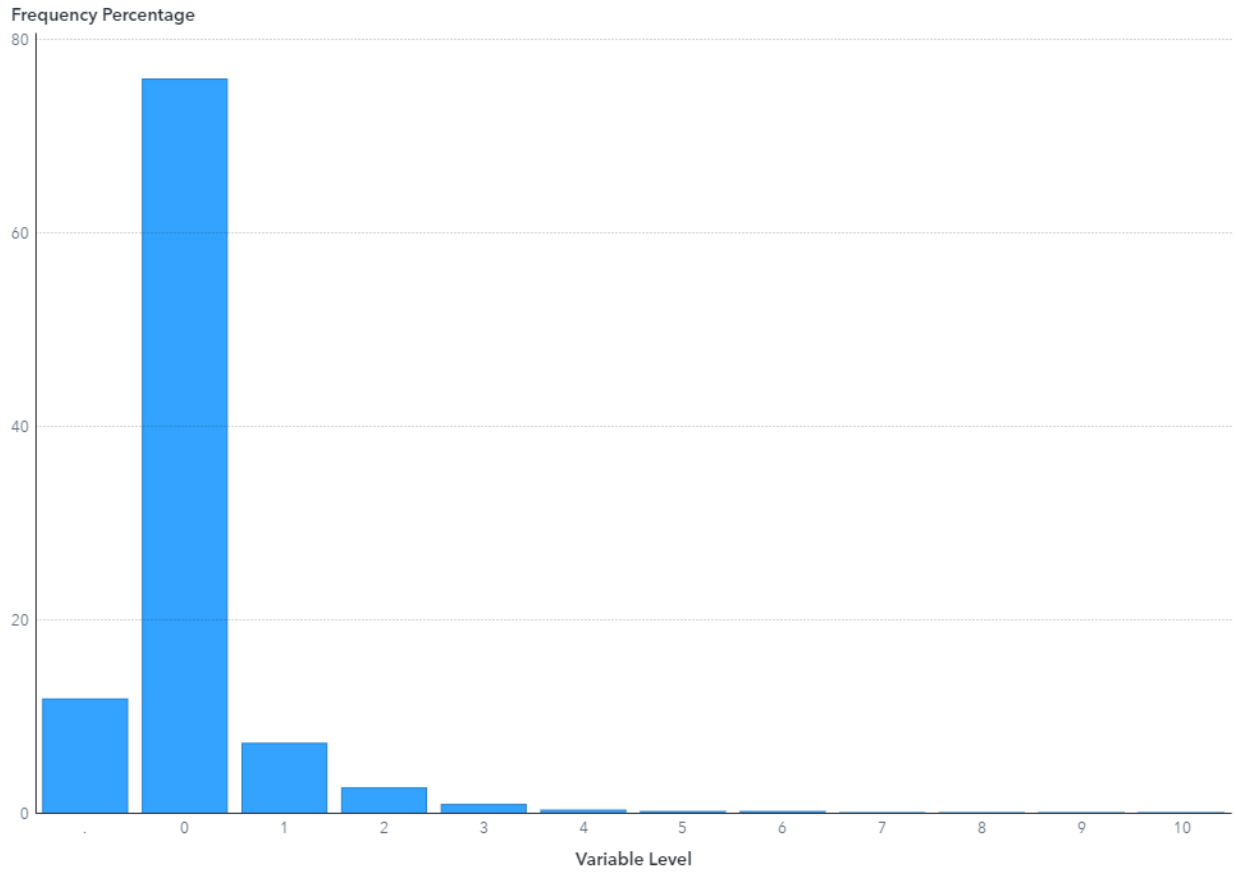
BAD



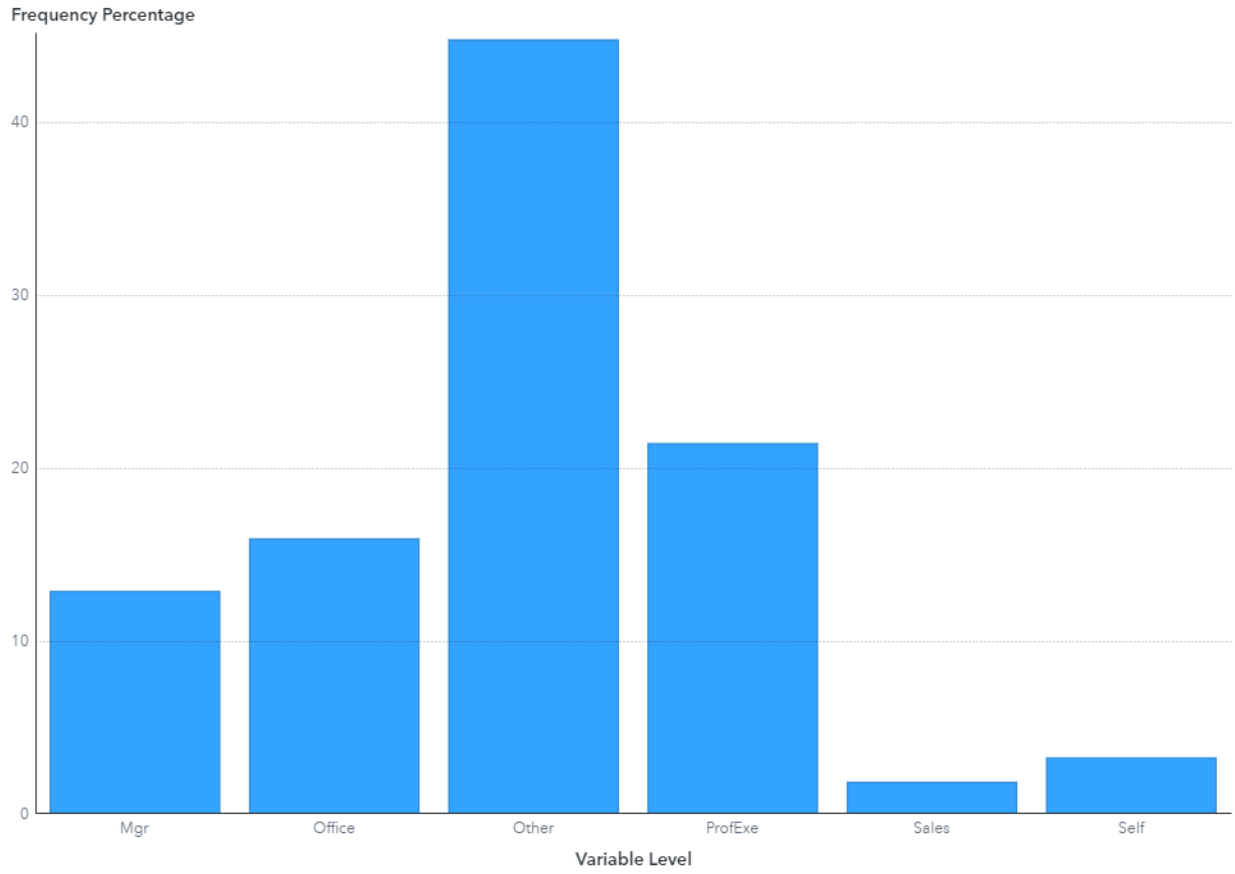
DELINQ


Frequency Percentage
60
40
20
0
.
0
1
2
3
4
5
6
7
8
10
11
12
13
15
Variable Level

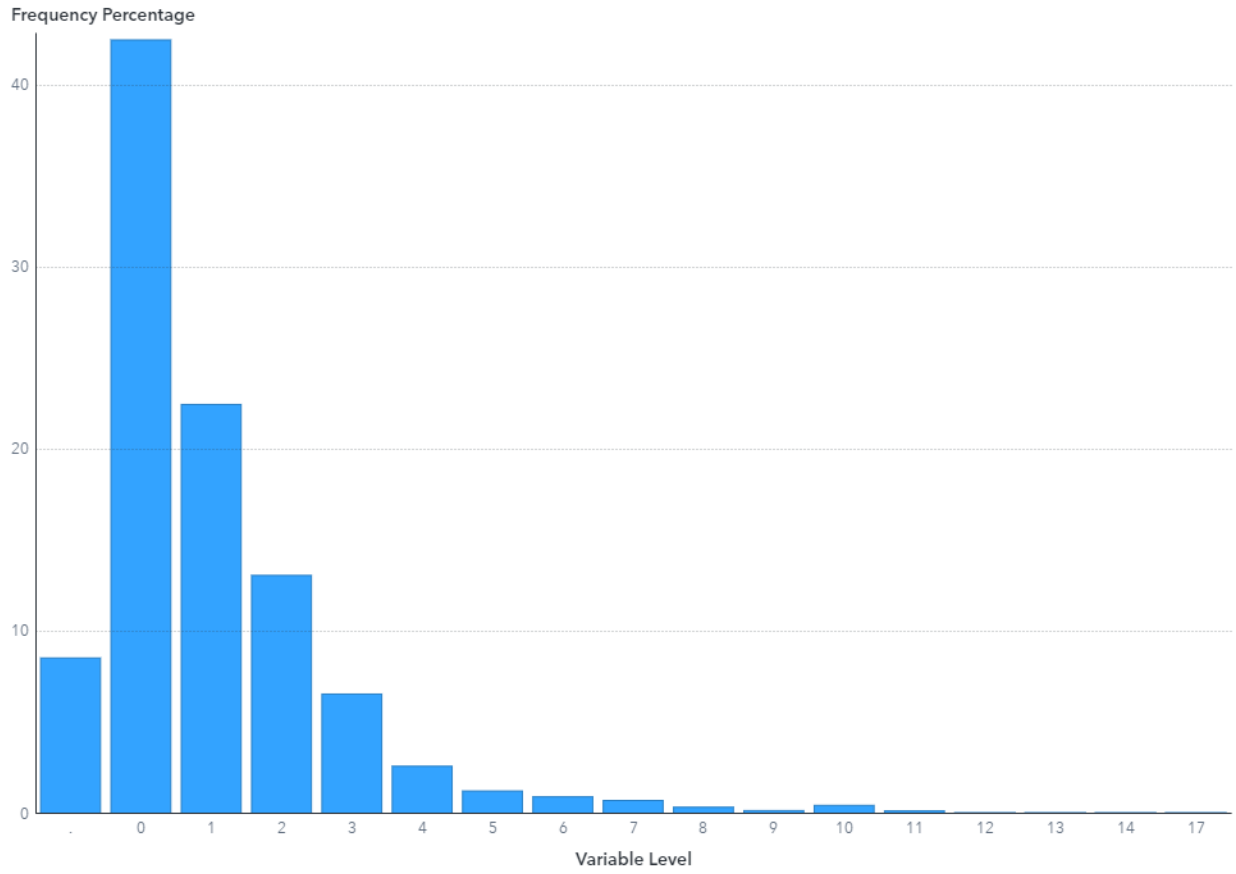
DEROG


Frequency Percentage
80
60
40
20
0
.
0
1
2
3
4
5
6
7
8
9
10
Variable Level

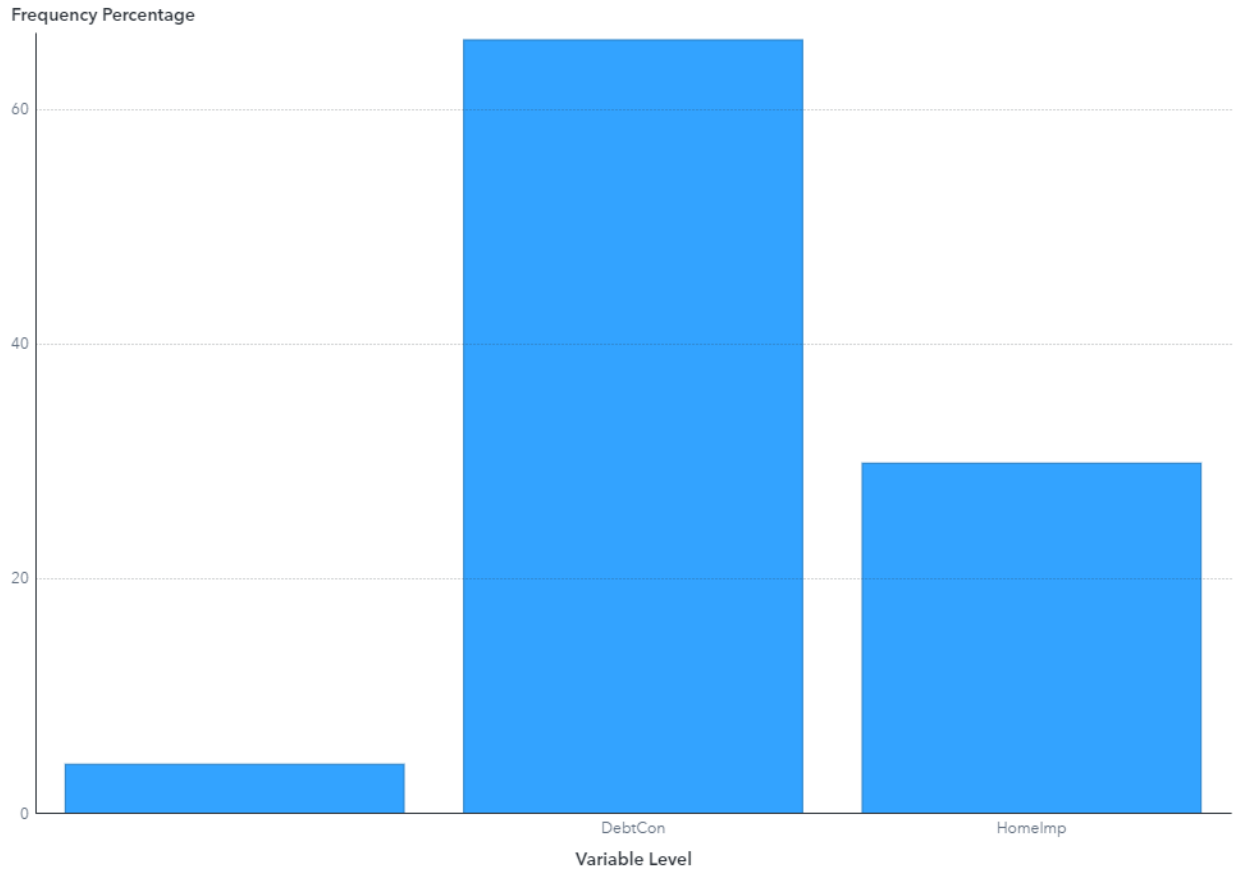
JOB


Frequency Percentage
40
30
20
10
0
Mgr
Office
Other
ProfExe
Sales
Self
Variable Level

NINQ


Frequency Percentage
40
30
20
10
0
.
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
17
Variable Level

REASON


Frequency Percentage
60
40
20
0
DebtCon
HomeImp
Variable Level

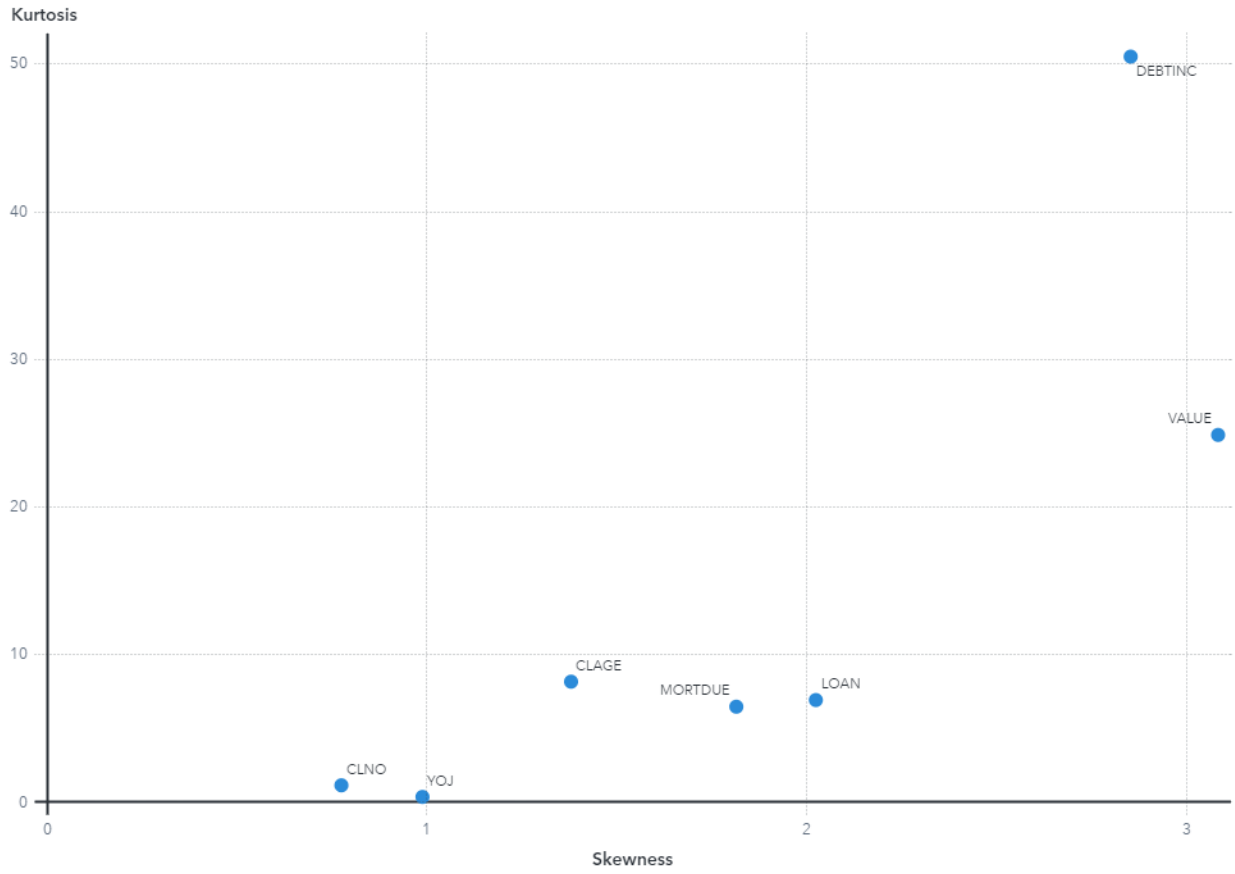
Interval Variable Moments

Variable Name	Minimum	Maximum	Mean
CLAGE	0	1,168.2336	179.7663
CLNO	0	71	21.2961
DEBTINC	0.5245	203.3121	33.7799
LOAN	1,100	89,900	18,607.9698
MORTDUE	2,063	399,550	73,760.8172
VALUE	8,000	855,909	101,776.0487
YOJ	0	41	8.9223

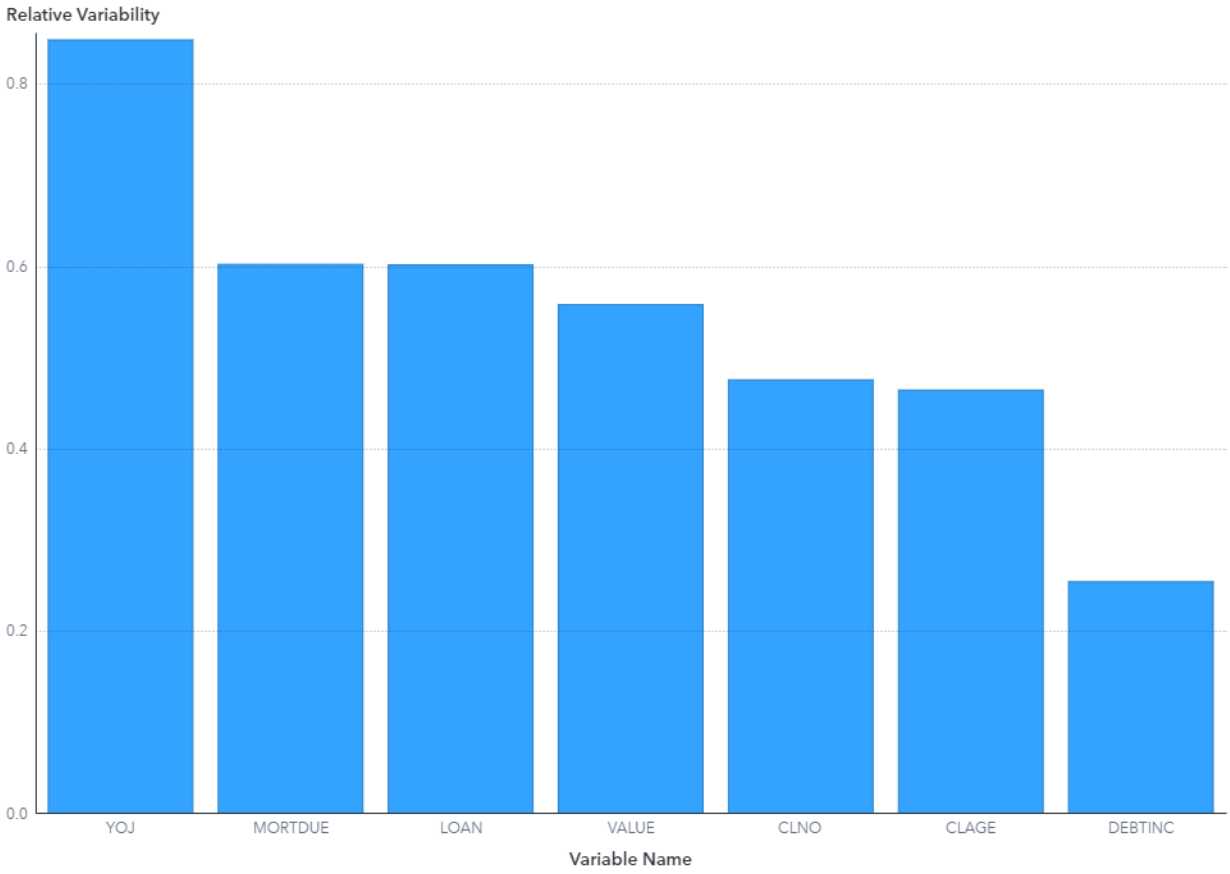
Standard Deviation	Skewness	Kurtosis	Relative Variability
83.5631	1.3795	8.1768	0.4648
10.1389	0.7751	1.1577	0.4761
8.6017	2.8524	50.5040	0.2546
11,207.4804	2.0238	6.9326	0.6023
44,457.6095	1.8145	6.4819	0.6027
56,843.9316	3.0824	24.8865	0.5585
7.5740	0.9885	0.3721	0.8489

Mean plus 2 SD	Mean minus 2 SD
346.8924	12.6402
41.5740	1.0182
50.9834	16.5764
41,022.9306	-3,806.9910
162,676.0361	-15,154.4017
215,463.9119	-11,911.8144
24.0702	-6.2257

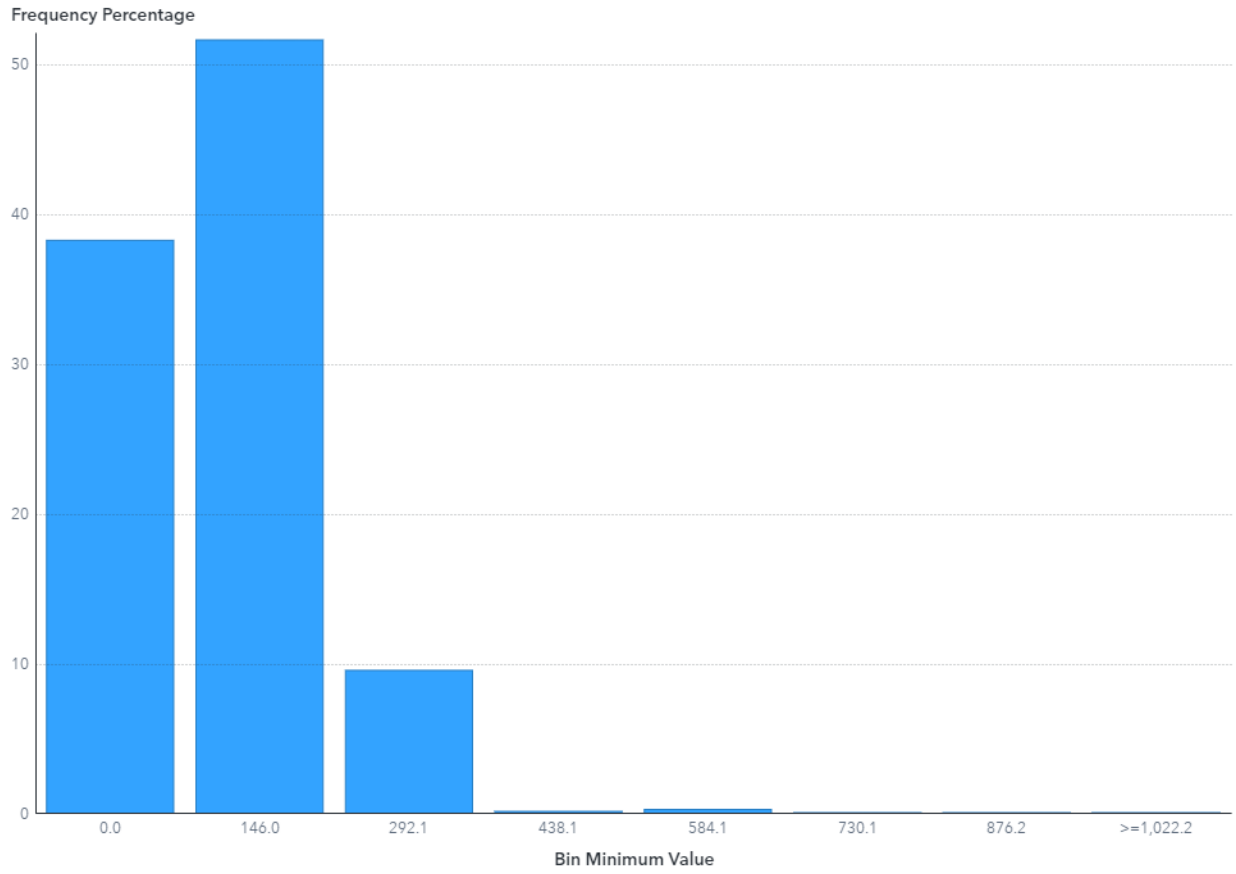
Deviation from Normality



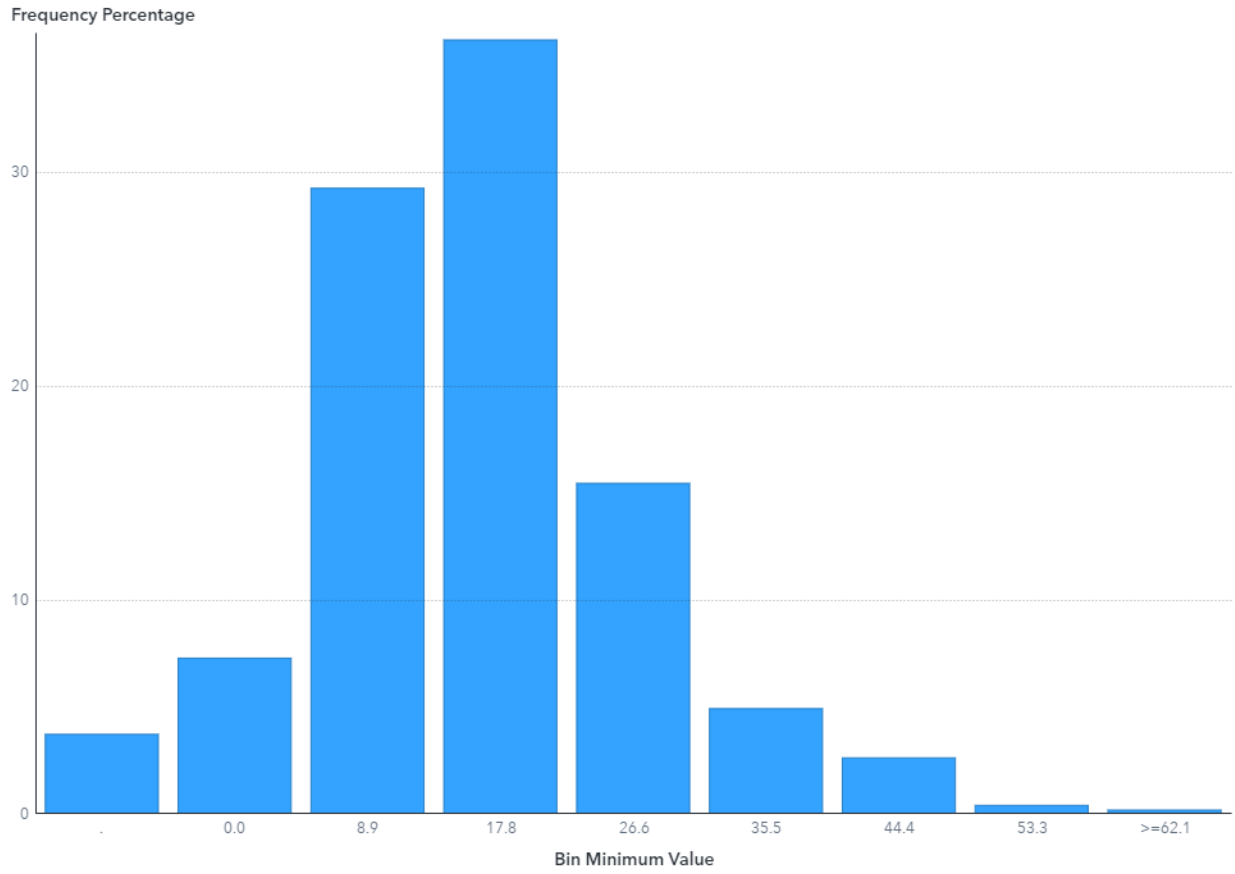
Relative Variability



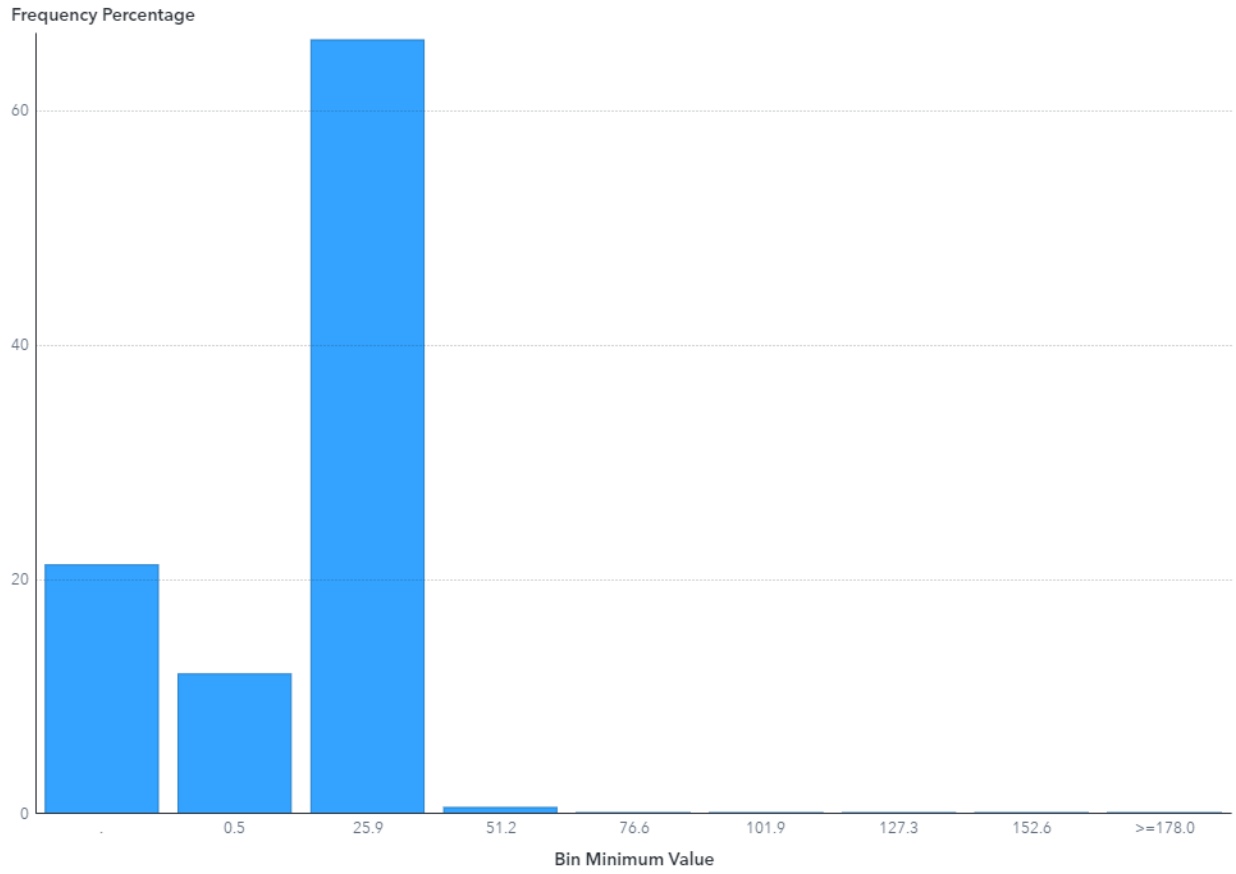
CLAGE


Frequency Percentage
50
40
30
20
10
0
0.0
146.0
292.1
438.1
584.1
730.1
876.2
>=1,022.2
Bin Minimum Value

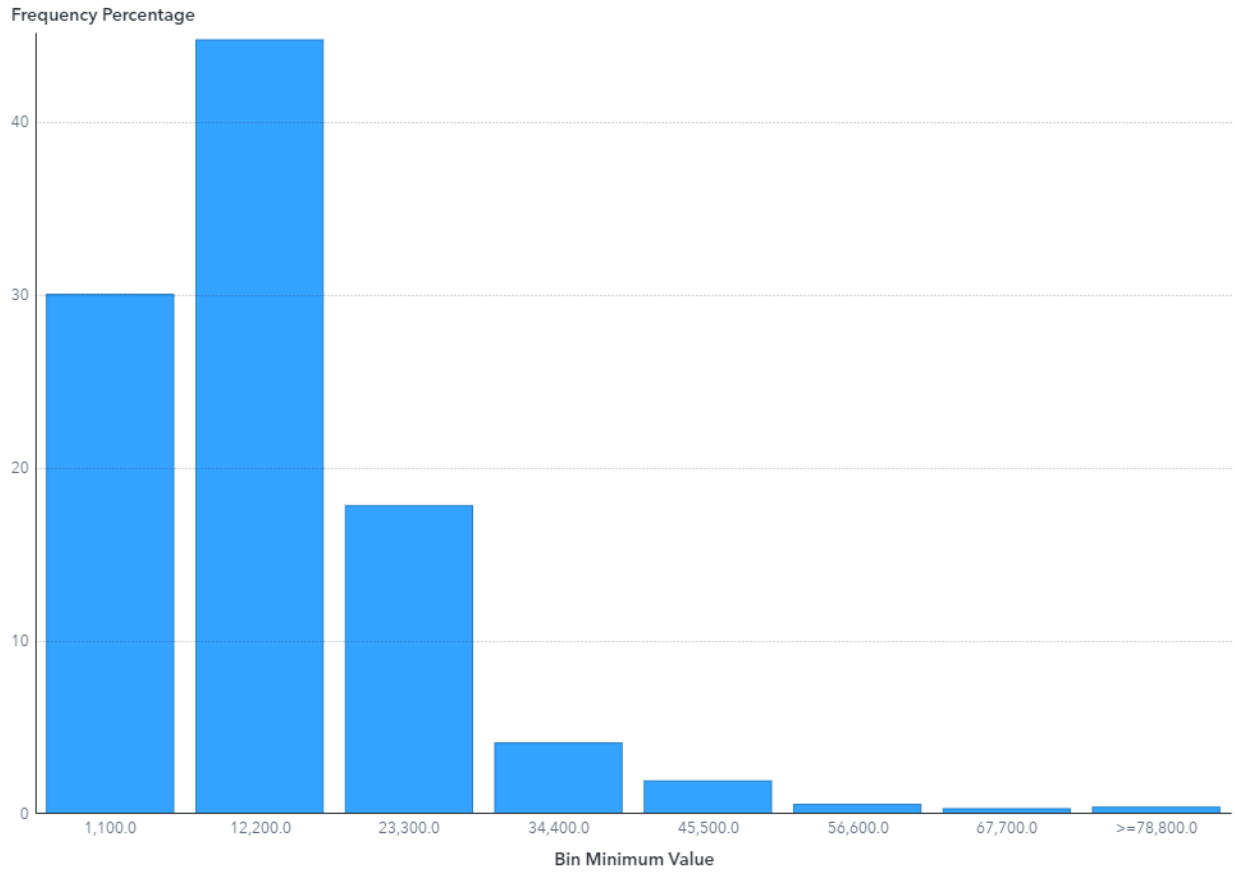
CLNO


Frequency Percentage
30
20
10
0
.
0.0
8.9
17.8
26.6
35.5
44.4
53.3
>=62.1
Bin Minimum Value

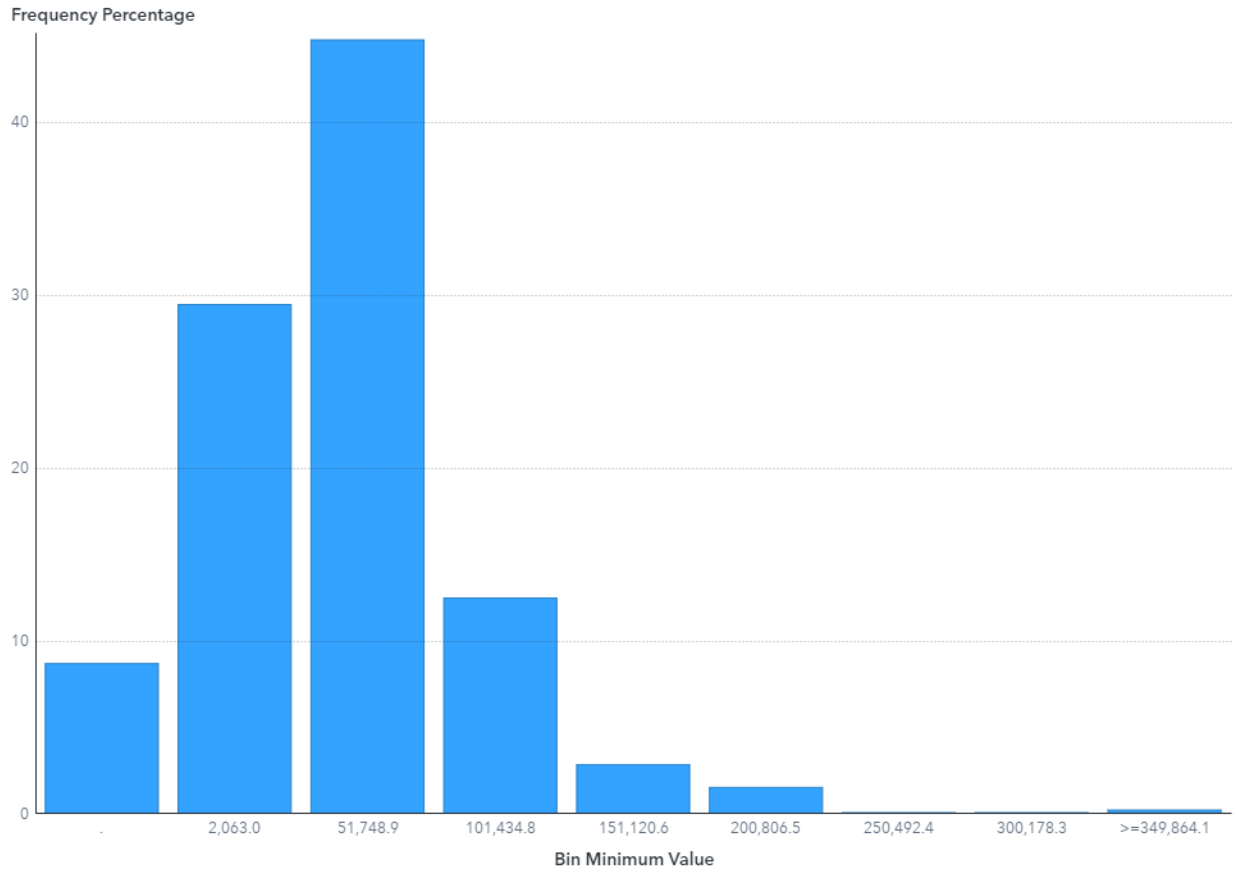
DEBTINC


Frequency Percentage
60
40
20
0
.
0.5
25.9
51.2
76.6
101.9
127.3
152.6
>=178.0
Bin Minimum Value

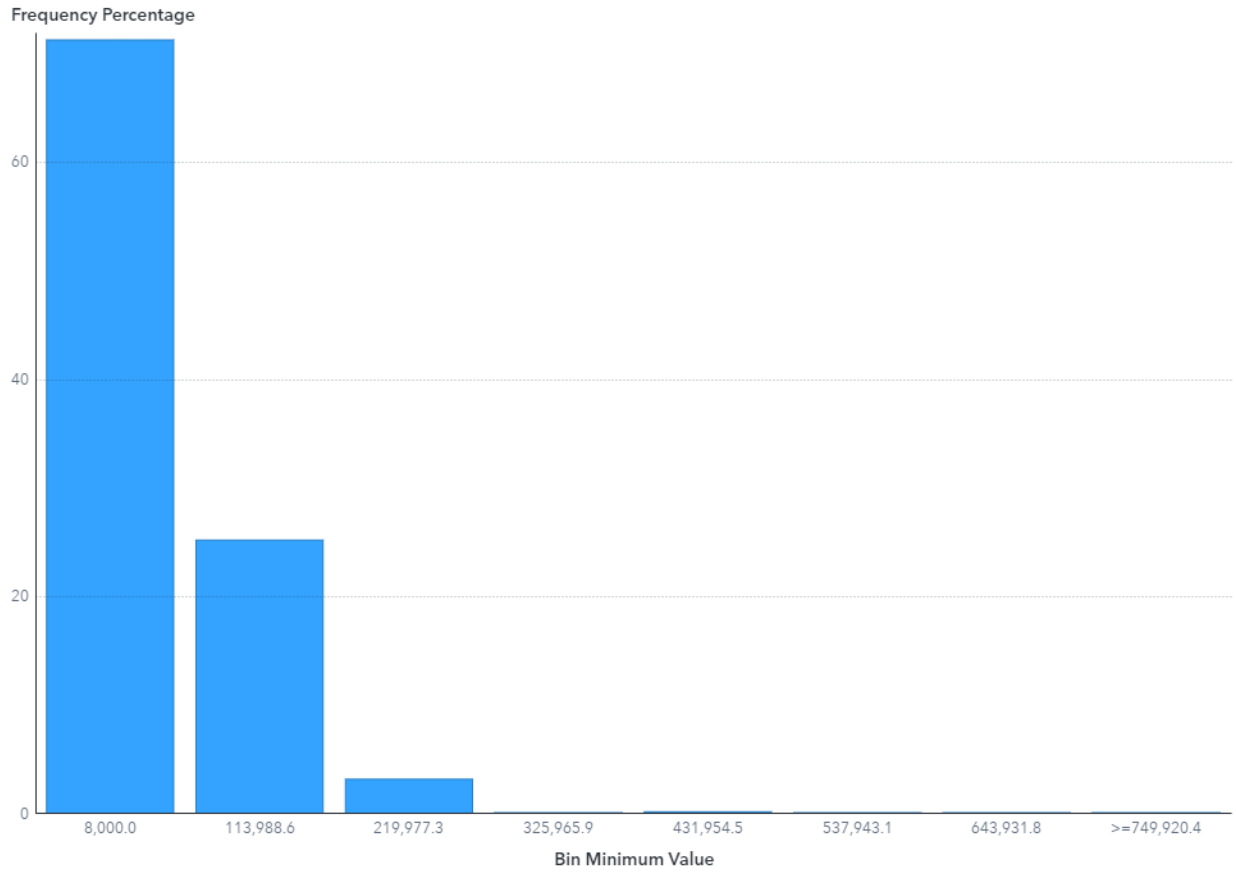
LOAN


Frequency Percentage
40
30
20
10
0
1,100.0
12,200.0
23,300.0
34,400.0
45,500.0
56,600.0
67,700.0
>=78,800.0
Bin Minimum Value

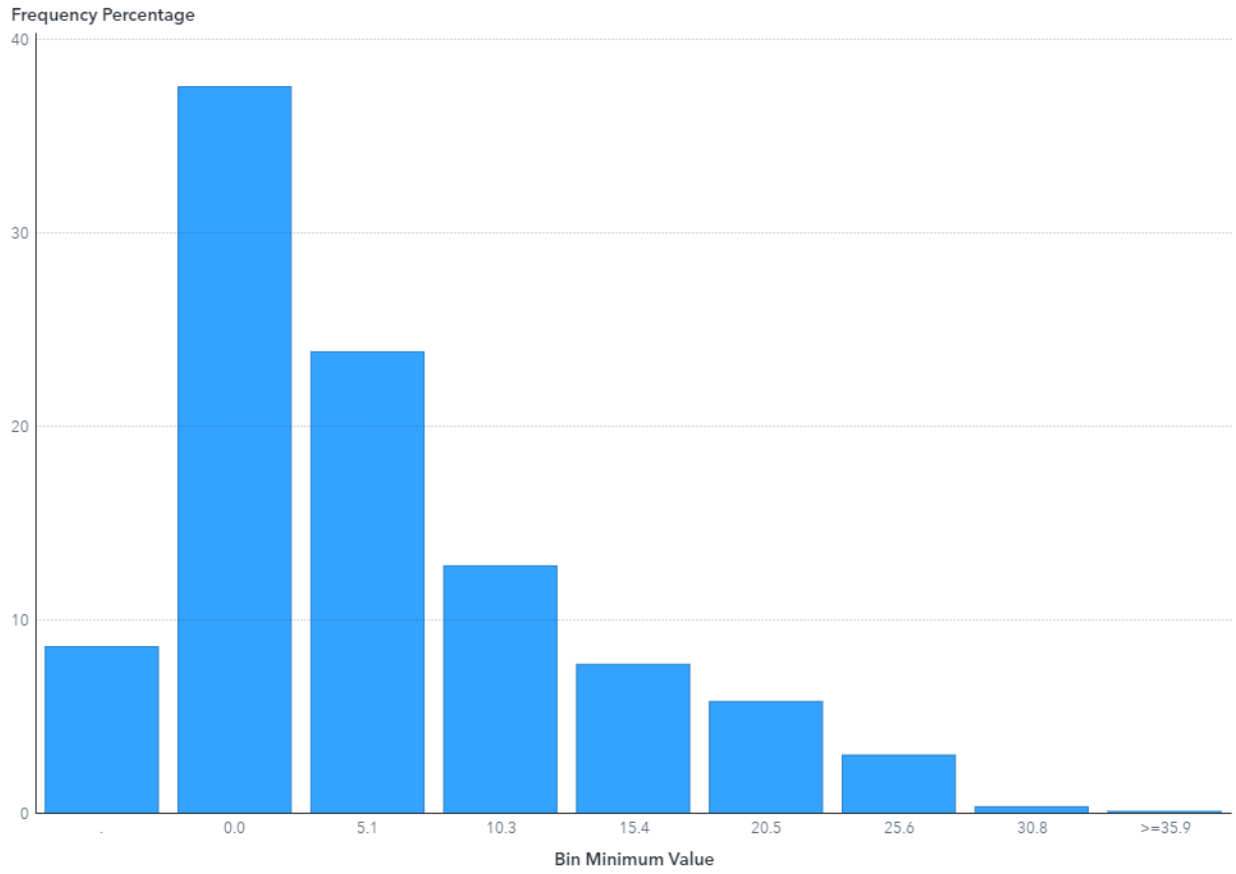
MORTDUE


Frequency Percentage
40
30
20
10
0
.
2,063.0
51,748.9
101,434.8
151,120.6
200,806.5
250,492.4
300,178.3
>=349,864.1
Bin Minimum Value

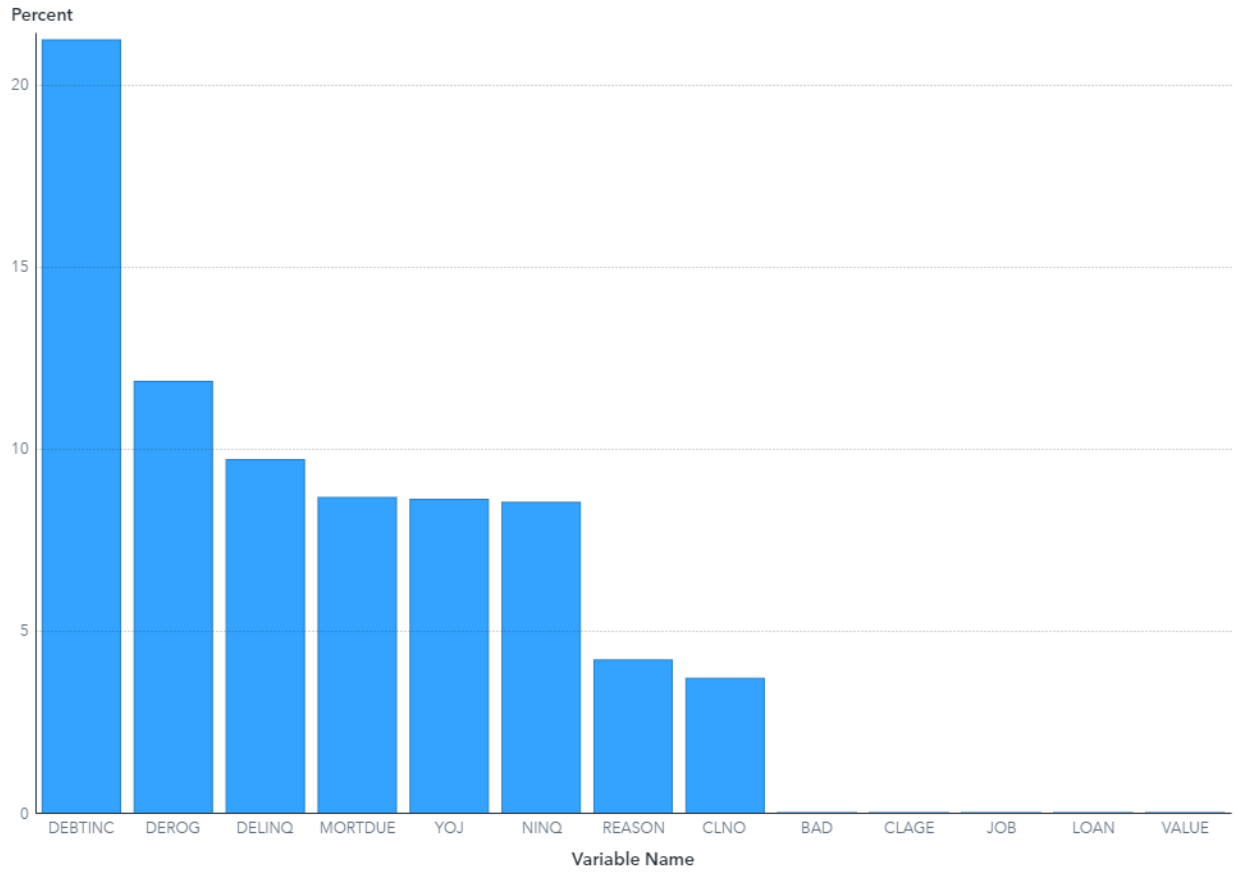
VALUE


Frequency Percentage
60
40
20
0
8,000.0
113,988.6
219,977.3
325,965.9
431,954.5
537,943.1
643,931.8
>=749,920.4
Bin Minimum Value

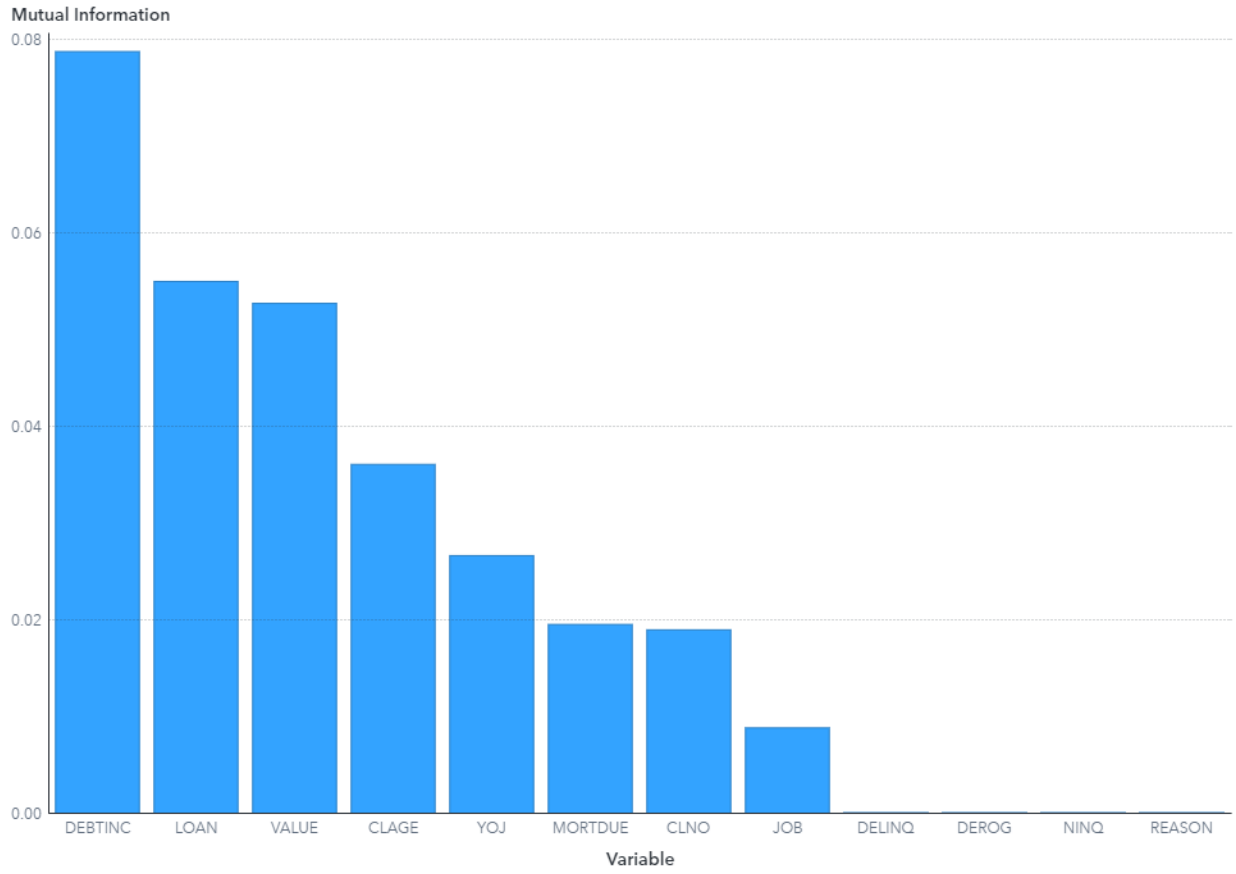
YOJ


Frequency Percentage
40
30
20
10
0
.
0.0
5.1
10.3
15.4
20.5
25.6
30.8
>=35.9
Bin Minimum Value

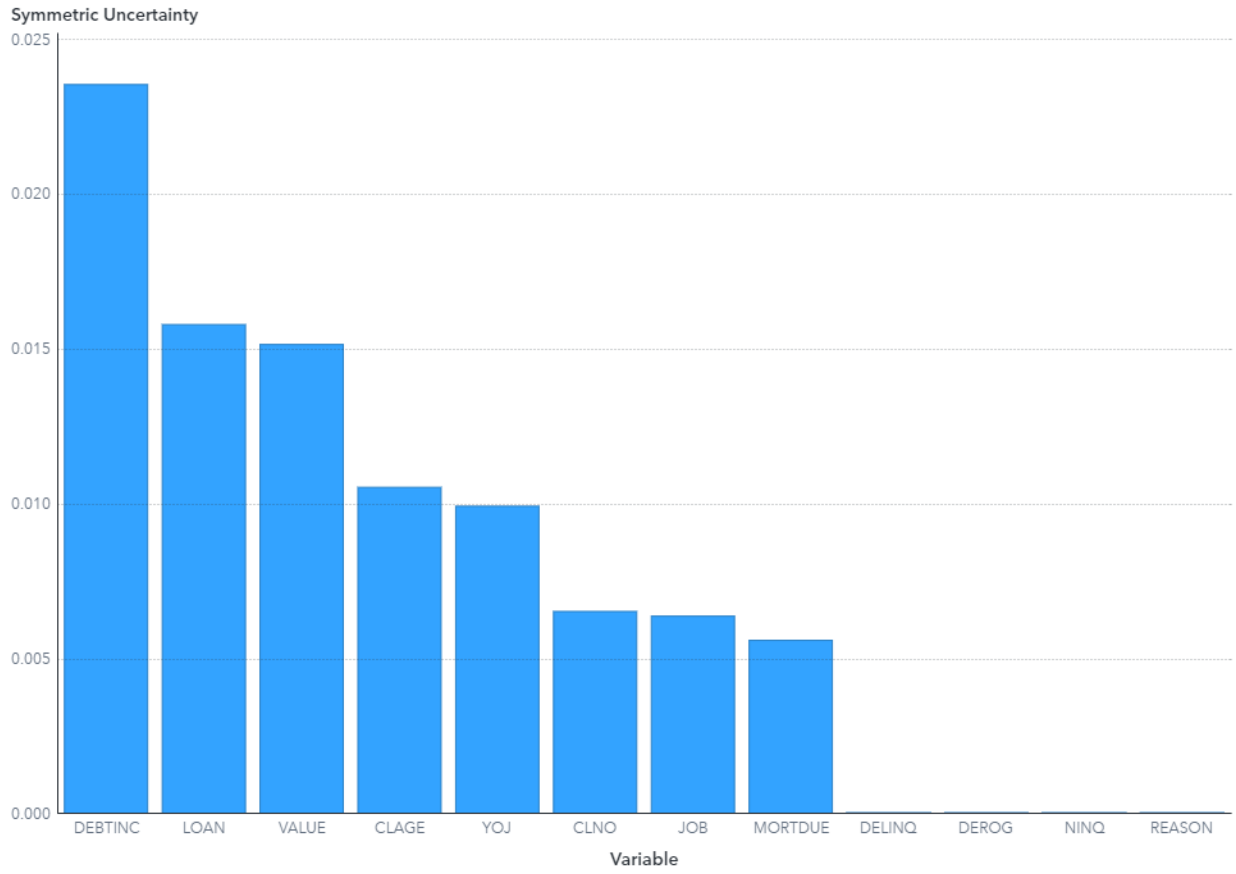
Missing Values



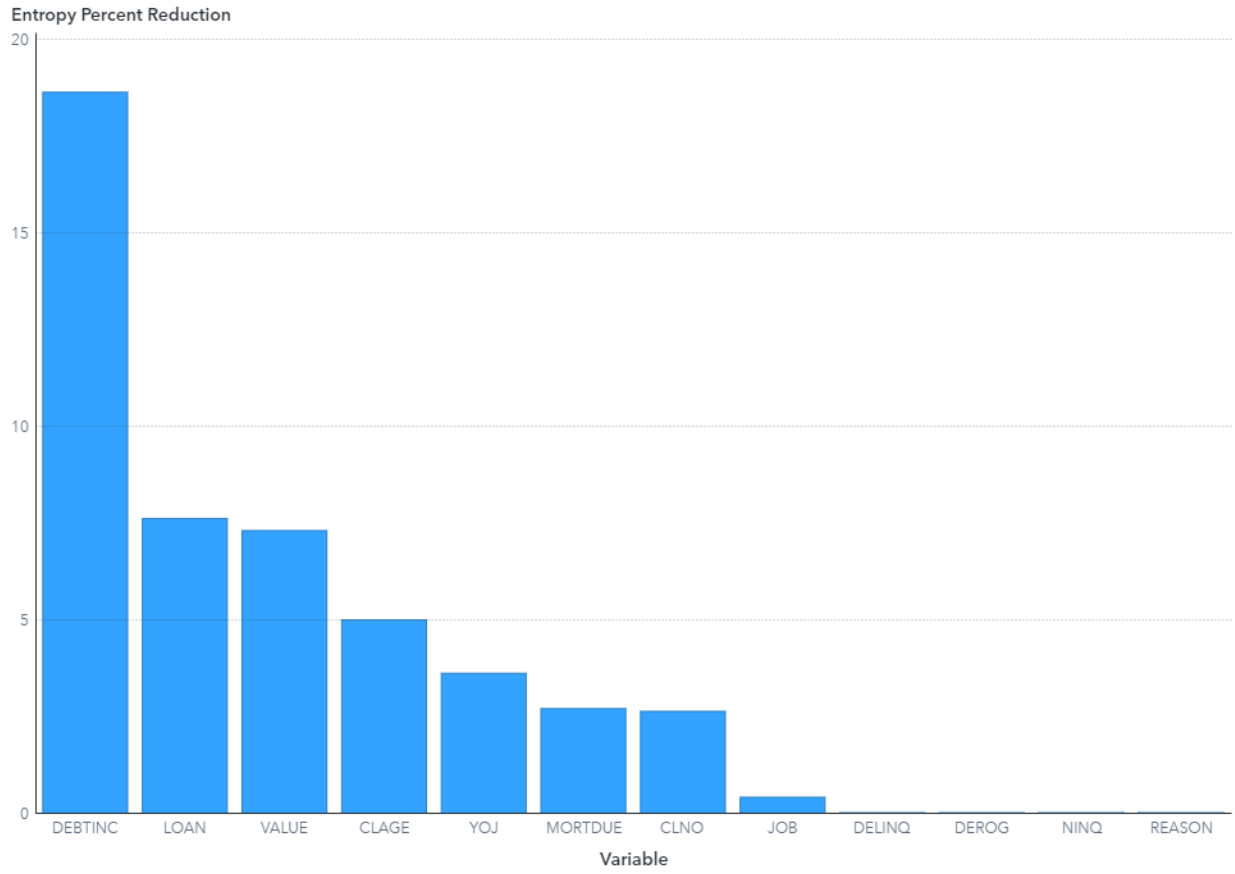
Mutual Information



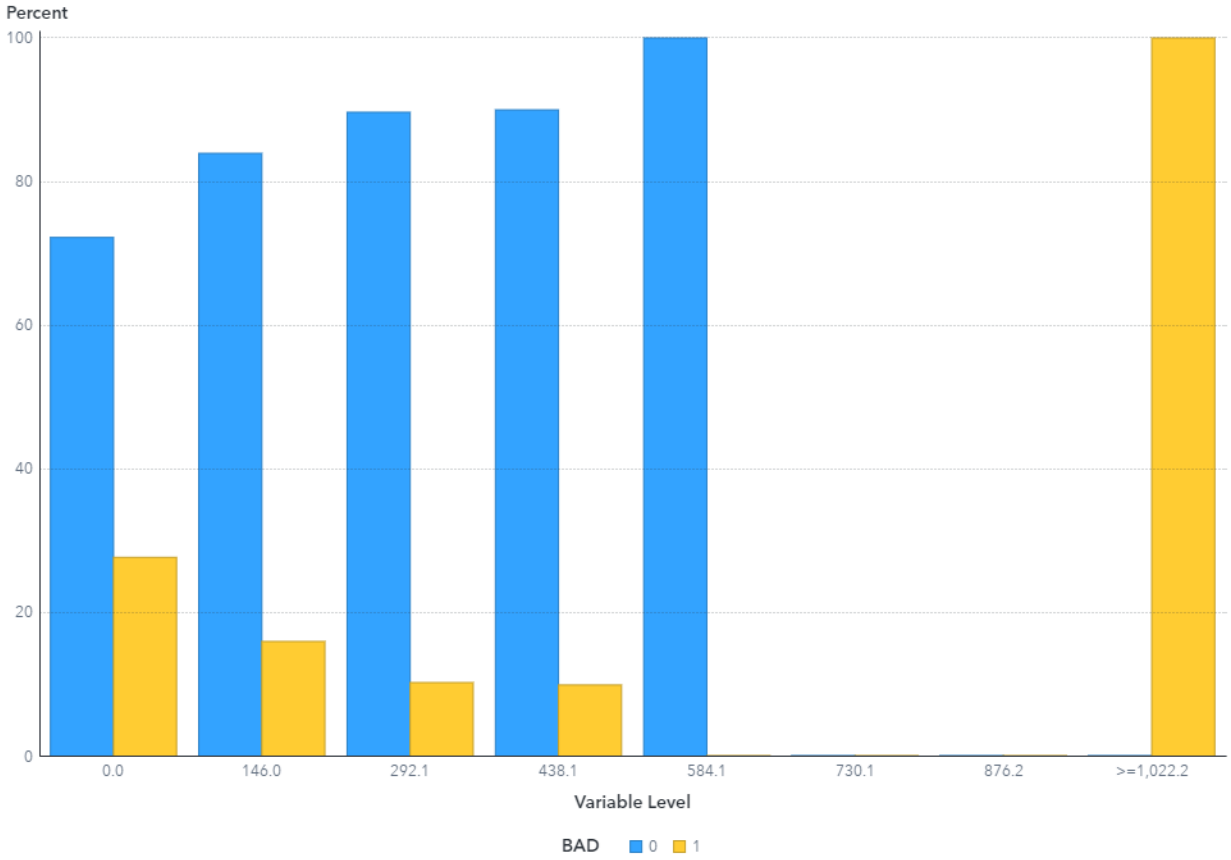
Symmetric Uncertainty



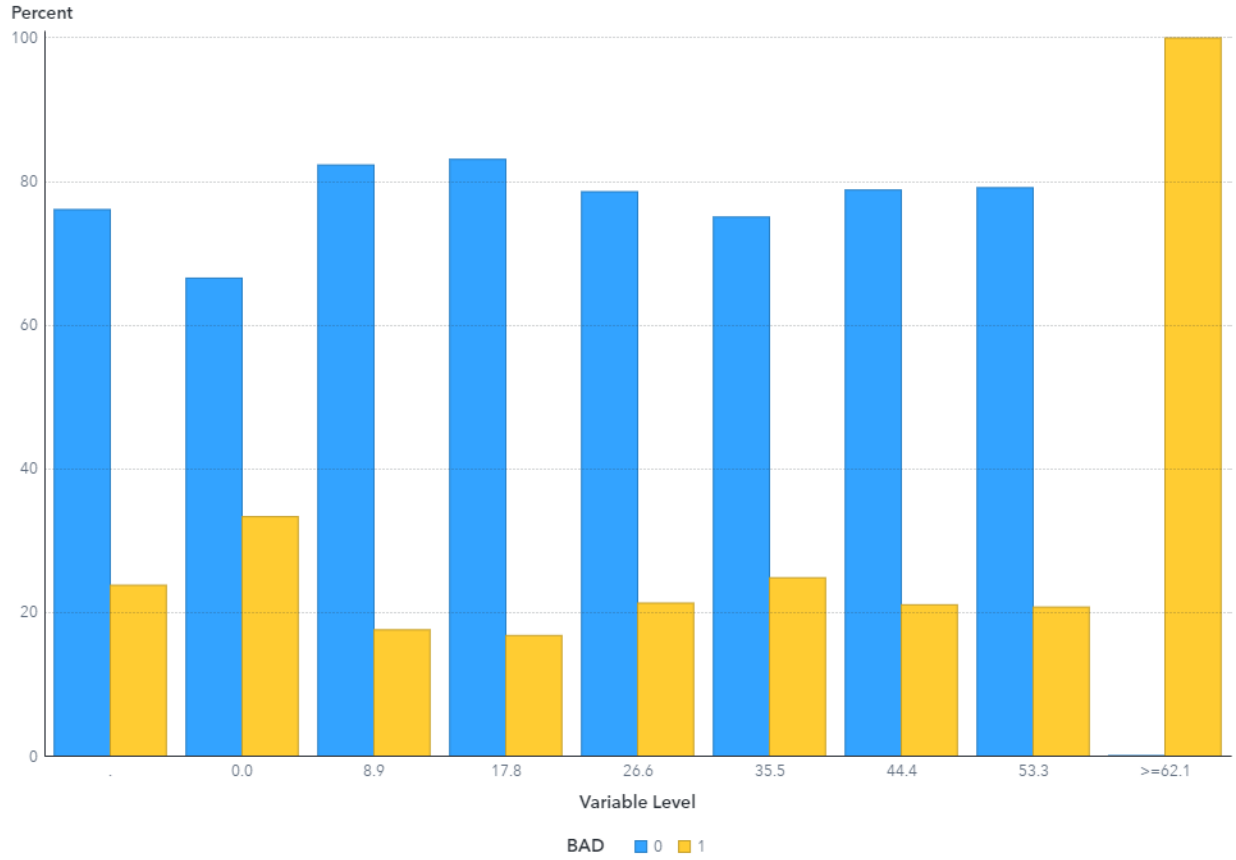
Entropy Reduction



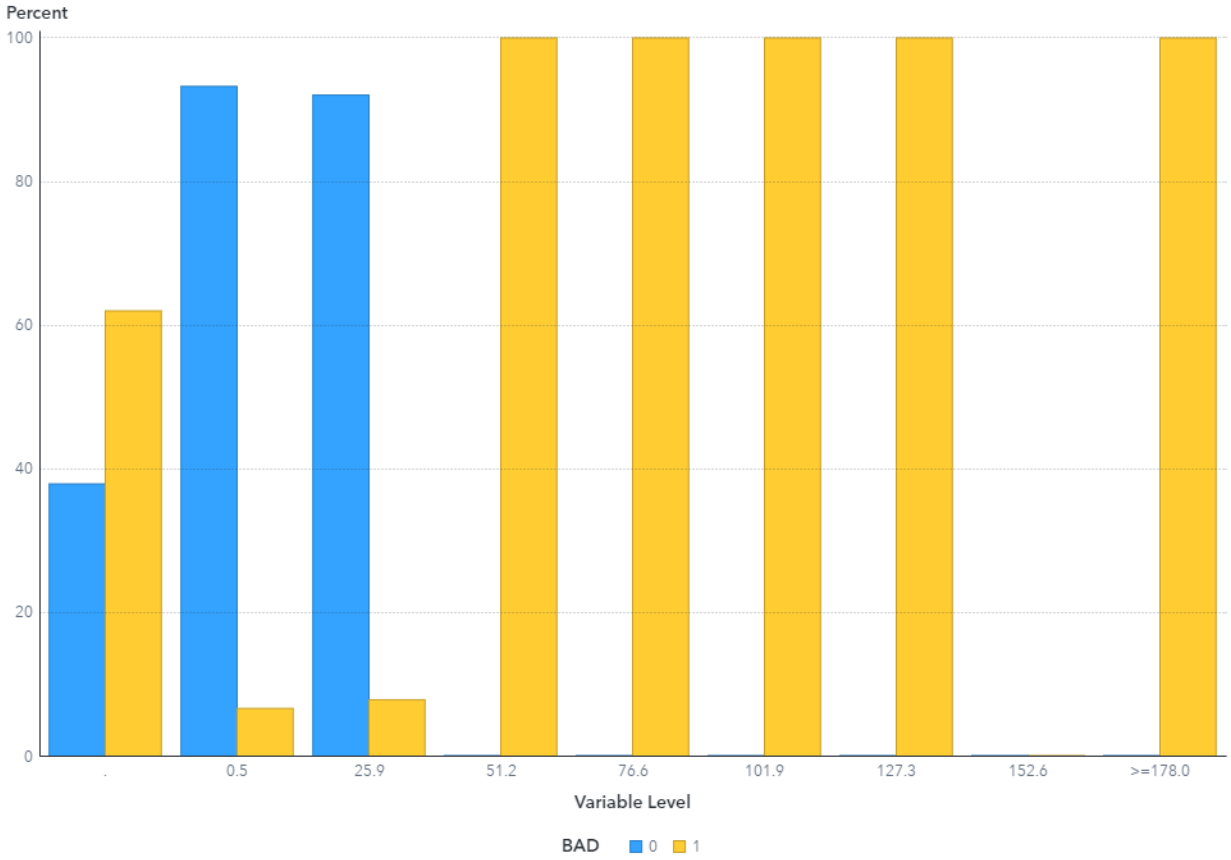
CLAGE


Percent
100
80
60
40
20
0
0.0
146.0
292.1
438.1
584.1
730.1
876.2
>=1,022.2
Variable Level
BAD
0
1

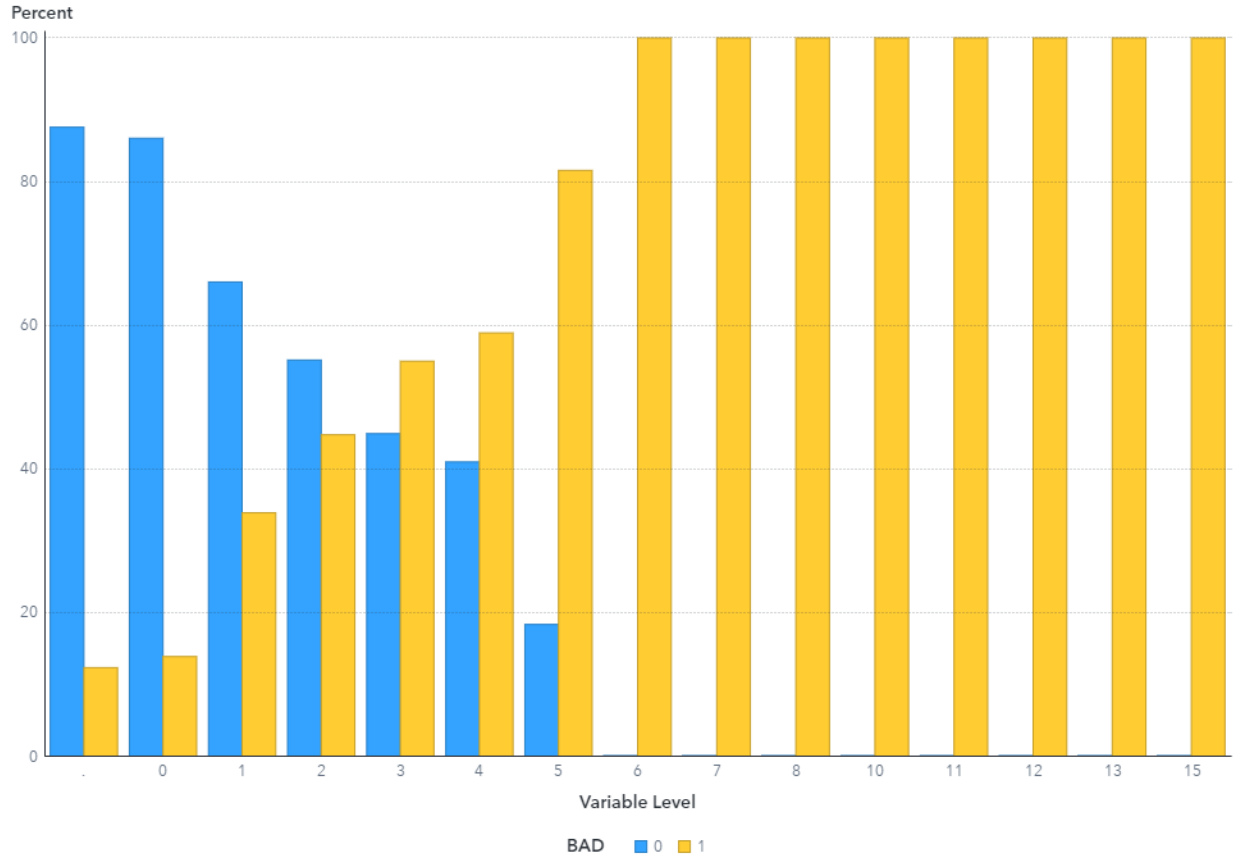
CLNO


Percent
100
80
60
40
20
0
.
0.0
8.9
17.8
26.6
35.5
44.4
53.3
>=62.1
Variable Level
BAD
0
1

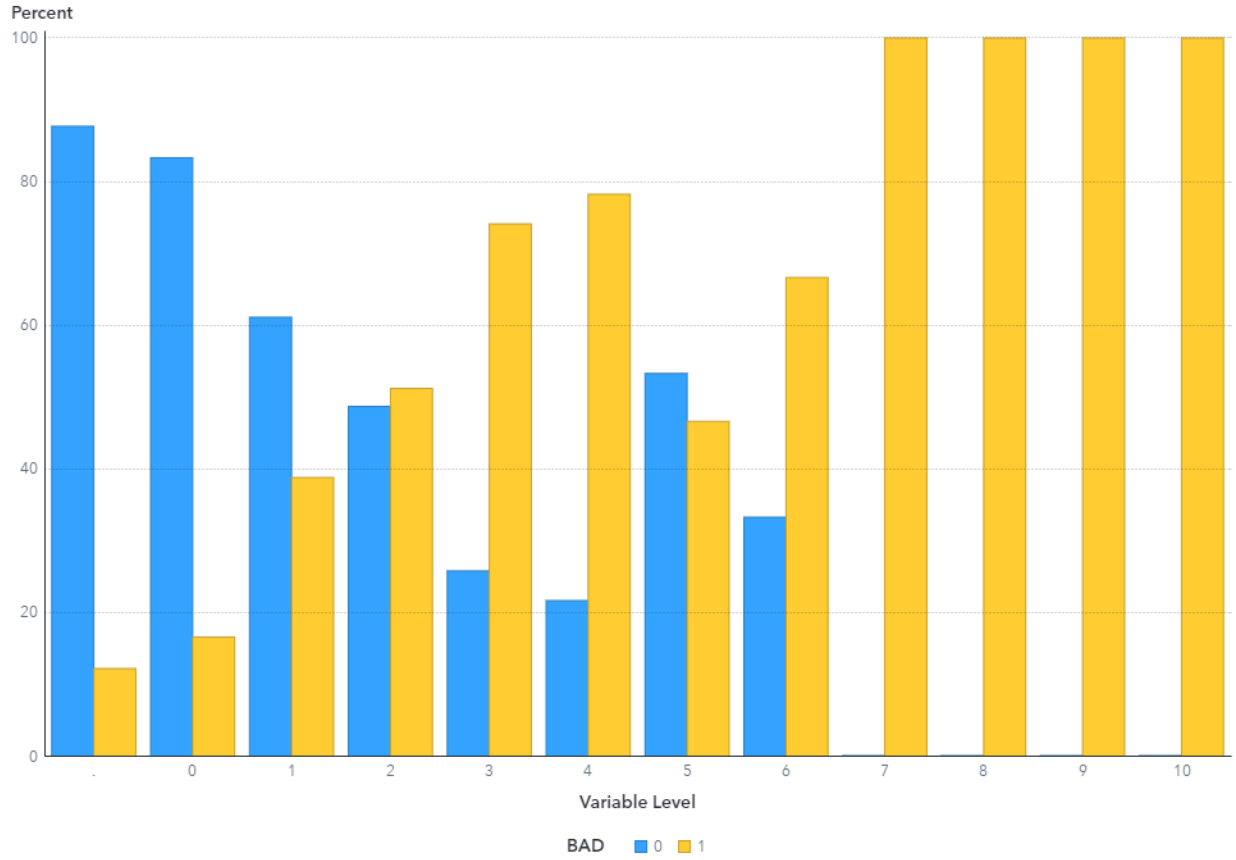
DEBTINC


Percent
100
80
60
40
20
0
.
0.5
25.9
51.2
76.6
101.9
127.3
152.6
>=178.0
Variable Level
BAD
0
1

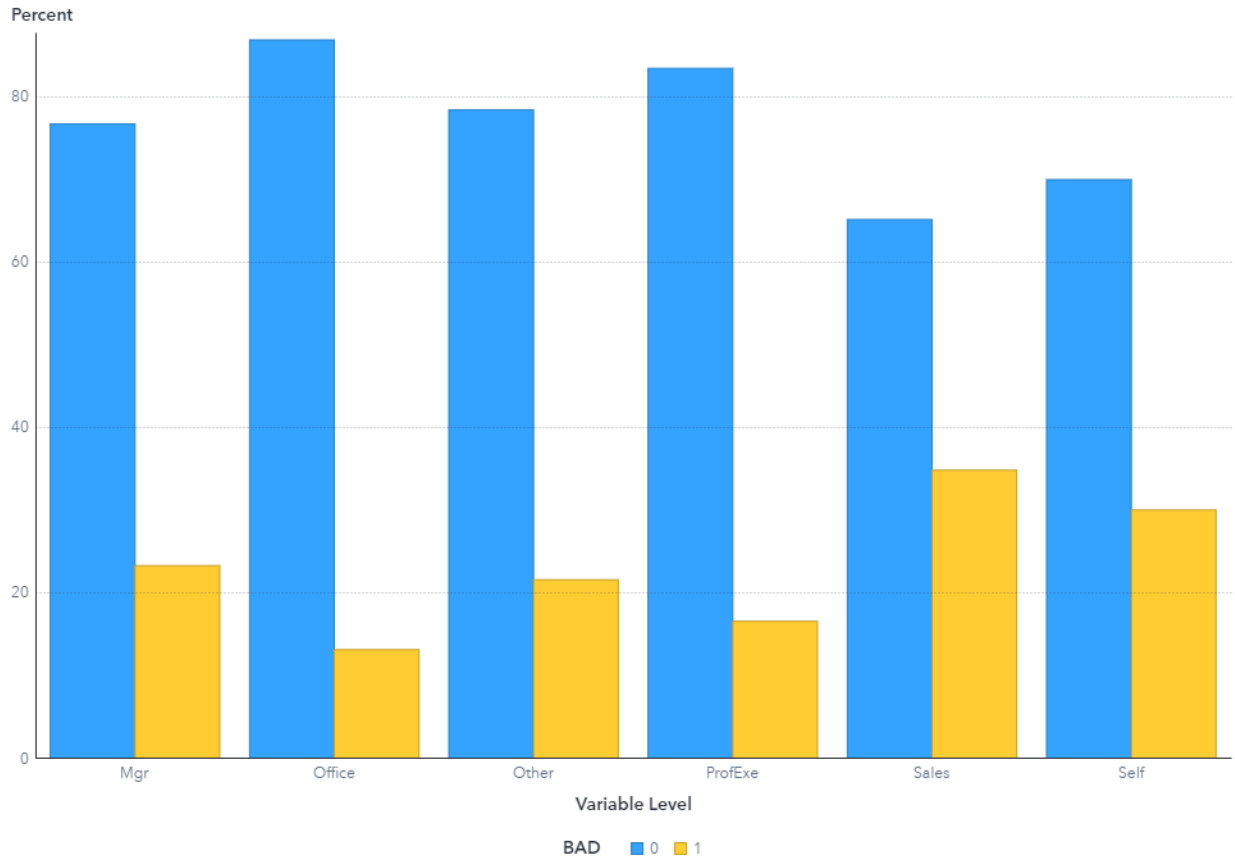
DELINQ


Percent
100
80
60
40
20
0
.
0
1
2
3
4
5
6
7
8
10
11
12
13
15
Variable Level
BAD
0
1

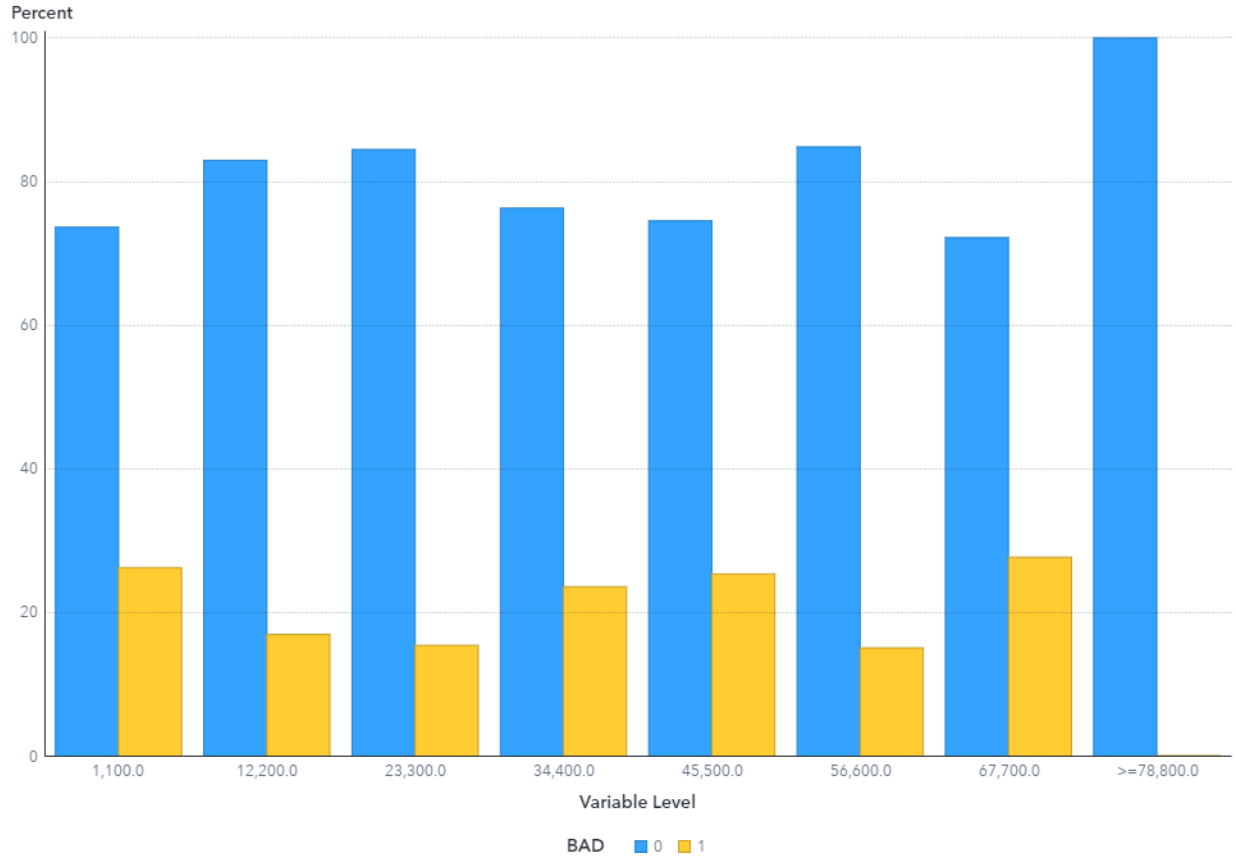
DEROG



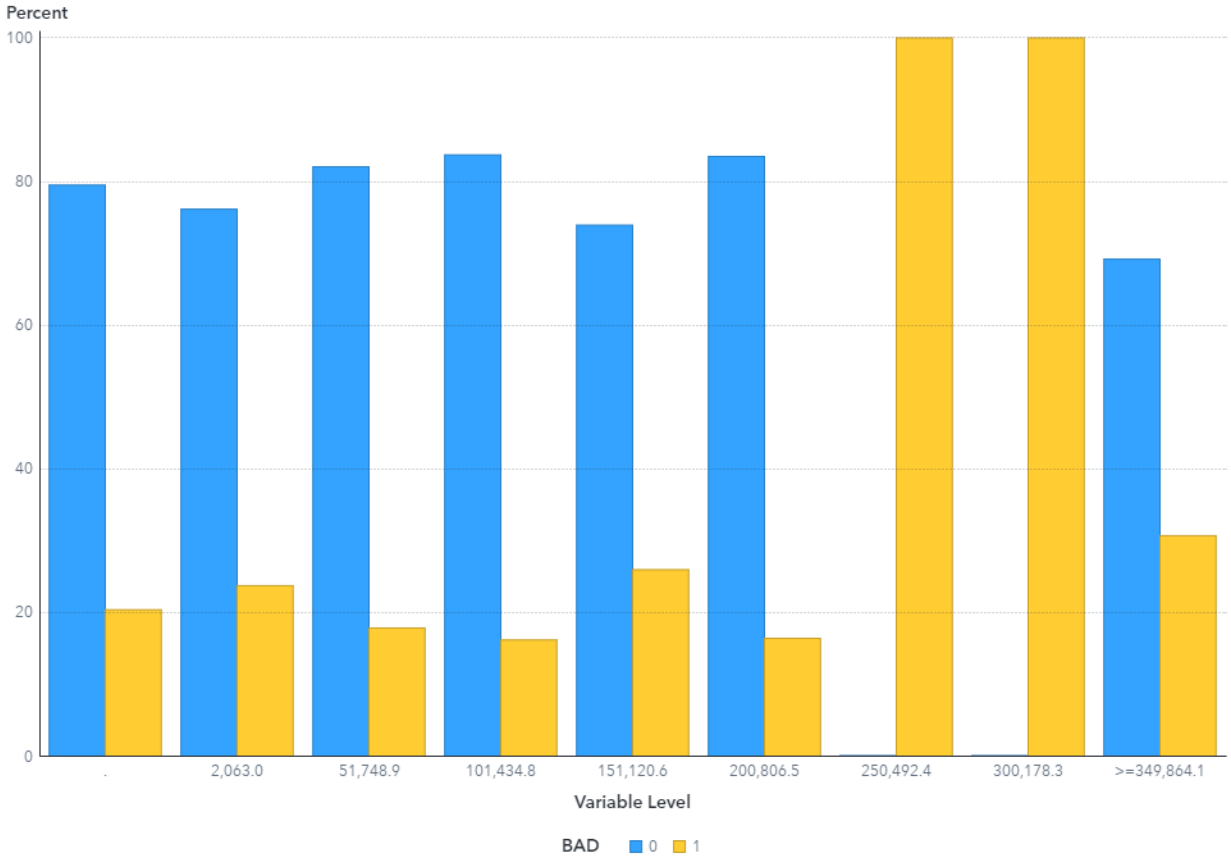
JOB


Percent
80
60
40
20
0
Mgr
Office
Other
ProfExe
Sales
Self
Variable Level
BAD
0
1

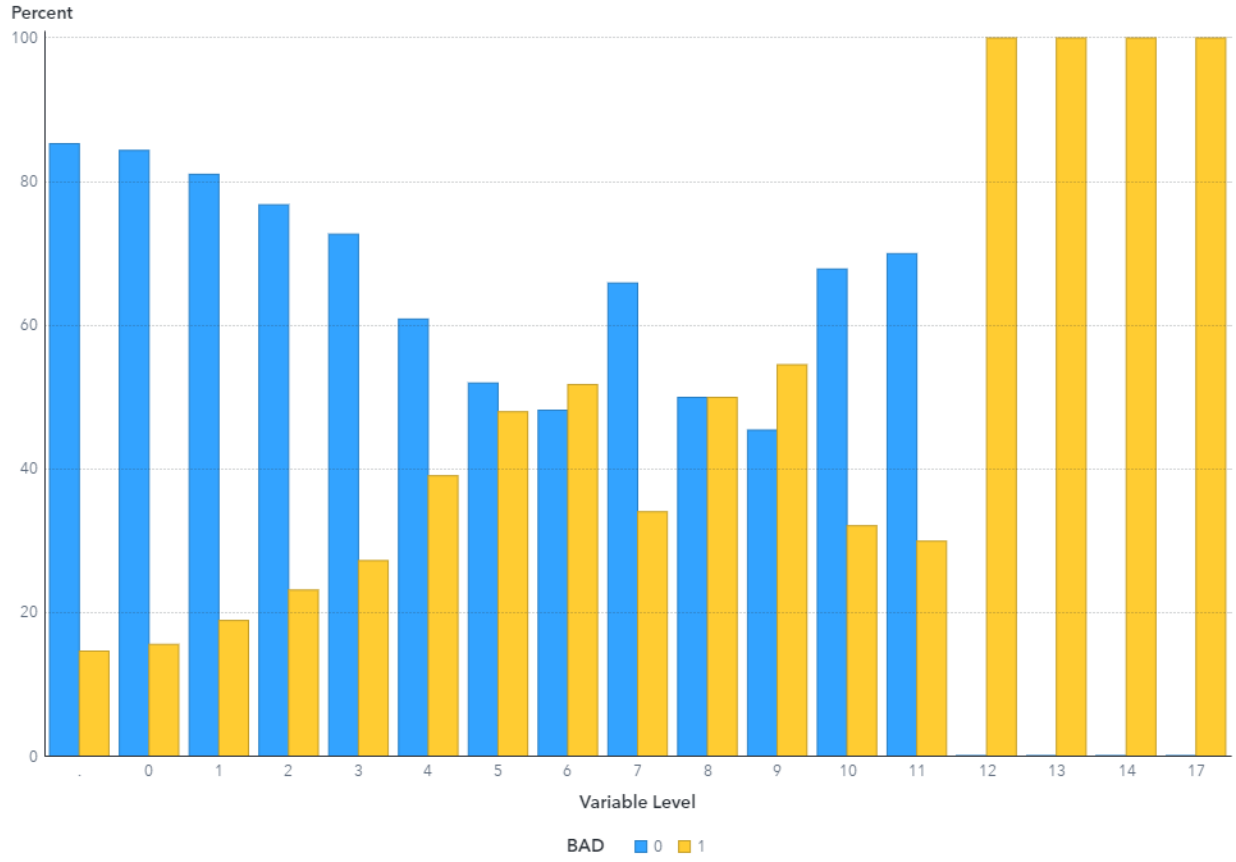
LOAN


Percent
100
80
60
40
20
0
1,100.0
12,200.0
23,300.0
34,400.0
45,500.0
56,600.0
67,700.0
>=78,800.0
Variable Level
BAD
0
1

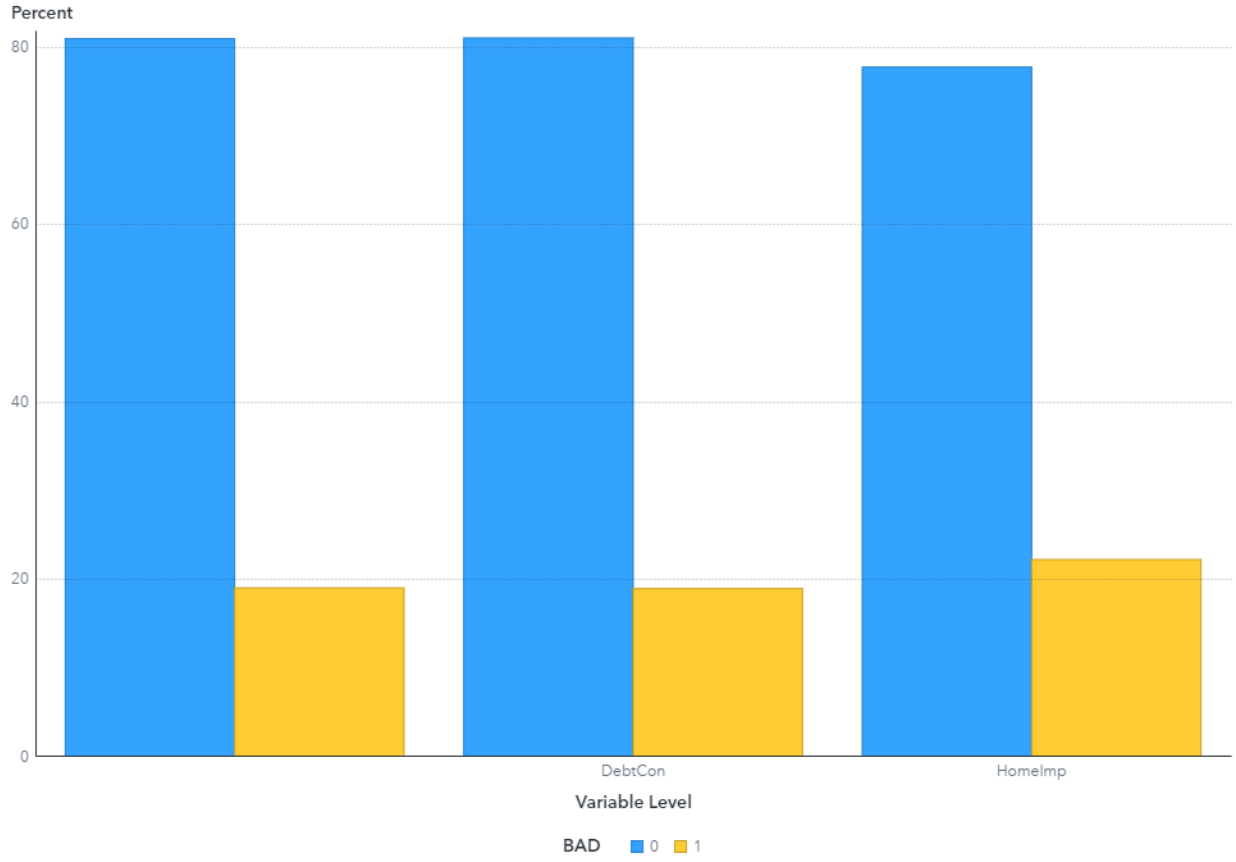
MORTDUE


Percent
100
80
60
40
20
0
.
2,063.0
51,748.9
101,434.8
151,120.6
200,806.5
250,492.4
300,178.3
>=349,864.1
Variable Level
BAD
0
1

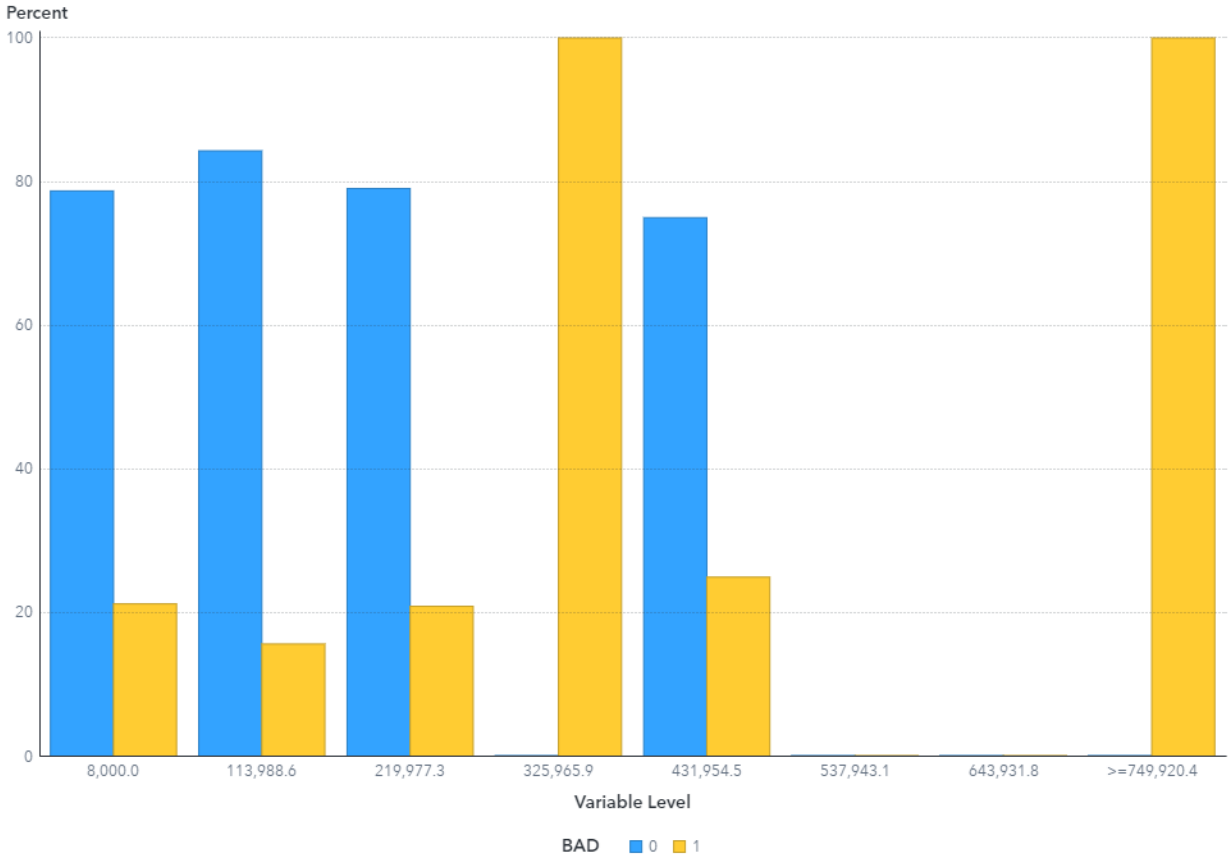
NINQ


Percent
100
80
60
40
20
0
. 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 17
Variable Level
BAD 0 1

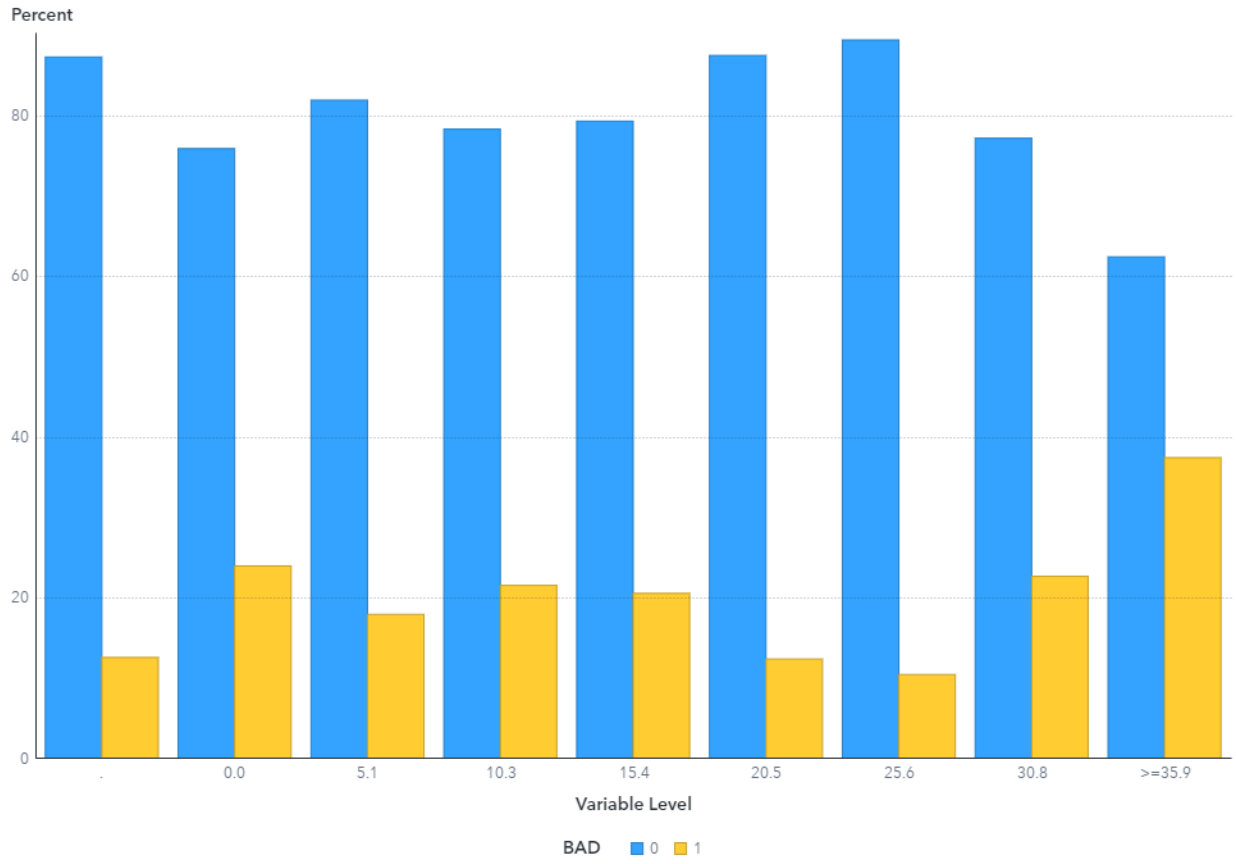
REASON



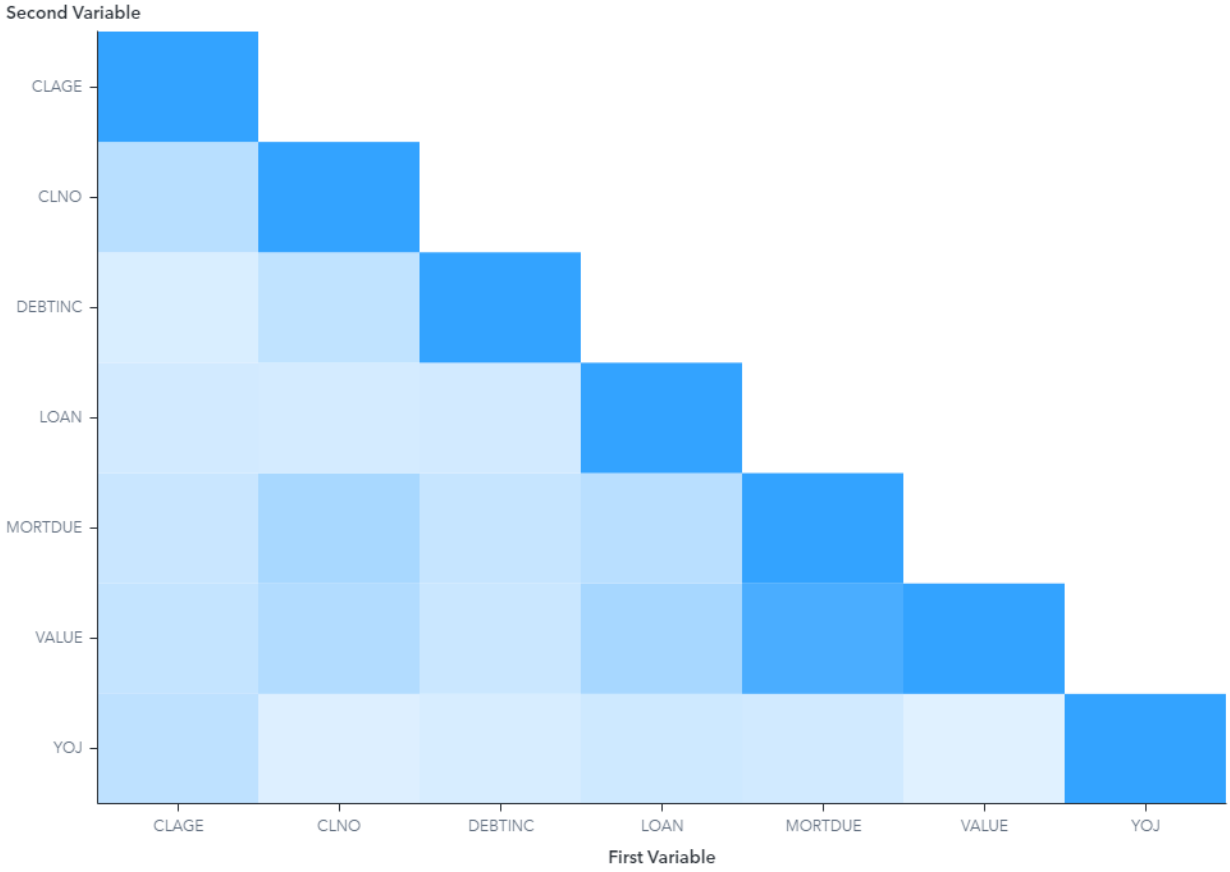
VALUE


Percent
100
80
60
40
20
0
8,000.0
113,988.6
219,977.3
325,965.9
431,954.5
537,943.1
643,931.8
>=749,920.4
Variable Level
BAD
0
1

YOJ


Percent
80
60
40
20
0
.
0.0
5.1
10.3
15.4
20.5
25.6
30.8
>=35.9
Variable Level
BAD
0
1

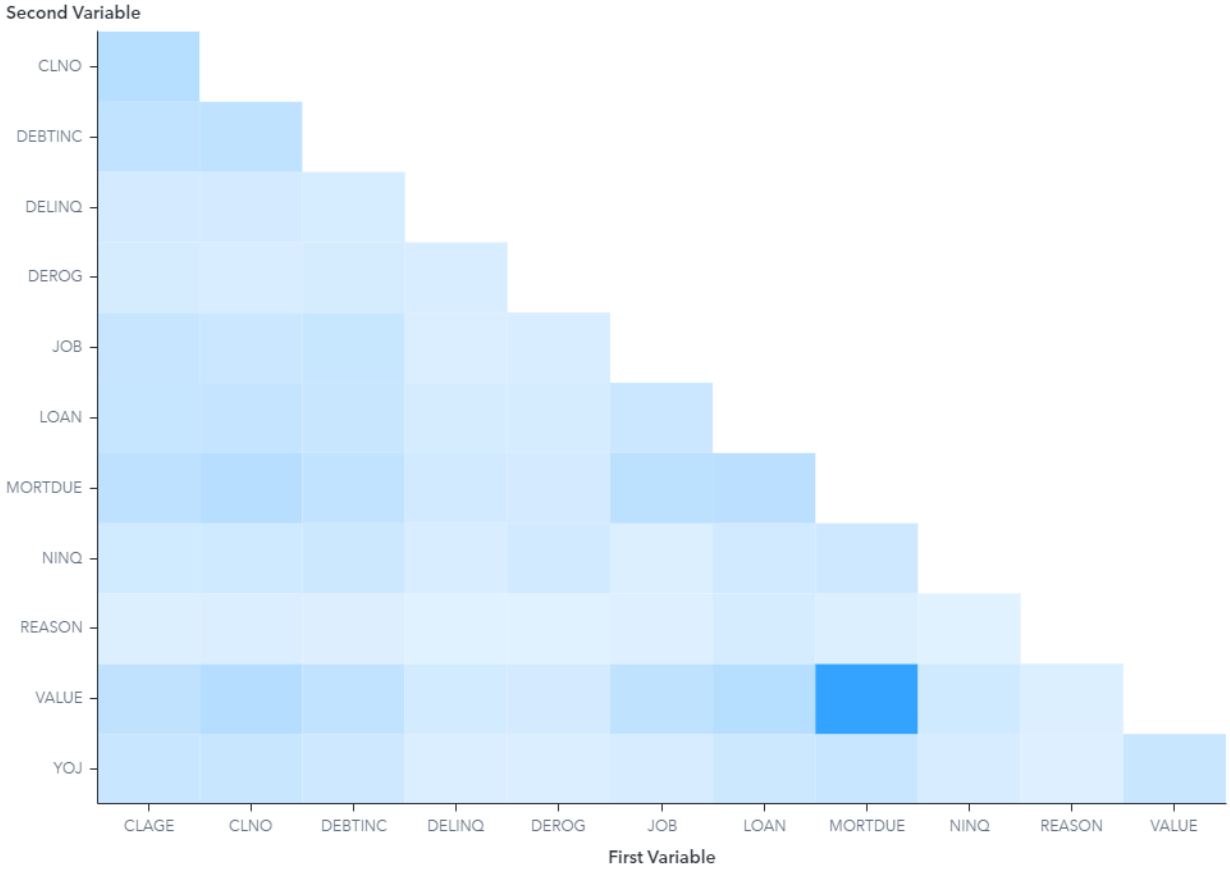
Pearson (Absolute)



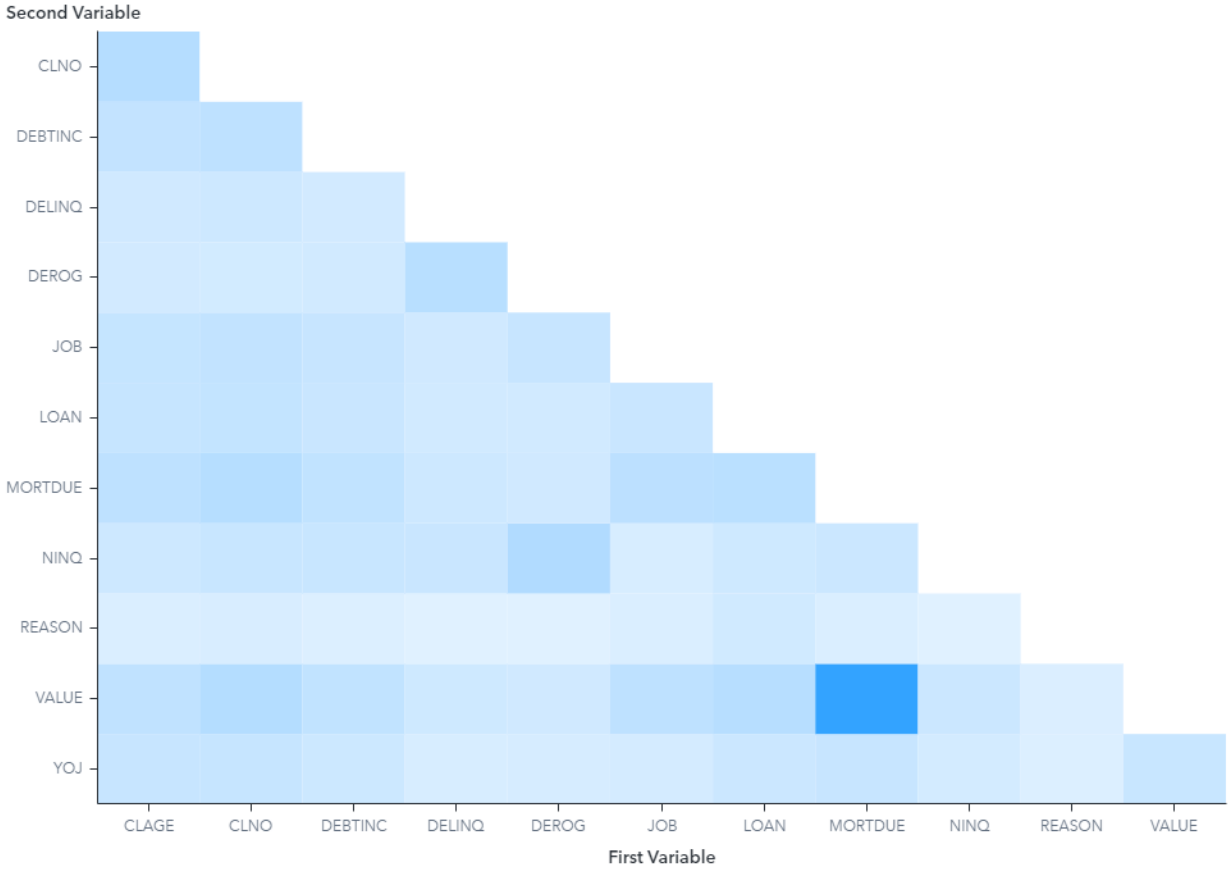
Table

First Variable	Second Variable	Pearson (Absolute)	Sign
MORTDUE	VALUE	0.8635	+
LOAN	VALUE	0.3319	+
CLNO	MORTDUE	0.3236	+
CLNO	VALUE	0.2663	+
CLAGE	CLNO	0.2316	+
LOAN	MORTDUE	0.2286	+
CLAGE	YOJ	0.1992	+
CLNO	DEBTINC	0.1855	+
CLAGE	VALUE	0.1679	+
DEBTINC	MORTDUE	0.1549	+
CLAGE	MORTDUE	0.1384	+
DEBTINC	VALUE	0.1313	+
LOAN	YOJ	0.1057	+
MORTDUE	YOJ	0.0885	-
CLAGE	LOAN	0.0869	+
DEBTINC	LOAN	0.0847	+
CLNO	LOAN	0.0726	+
DEBTINC	YOJ	0.0559	-
CLAGE	DEBTINC	0.0429	-
CLNO	YOJ	0.0248	+
VALUE	YOJ	0.0076	+

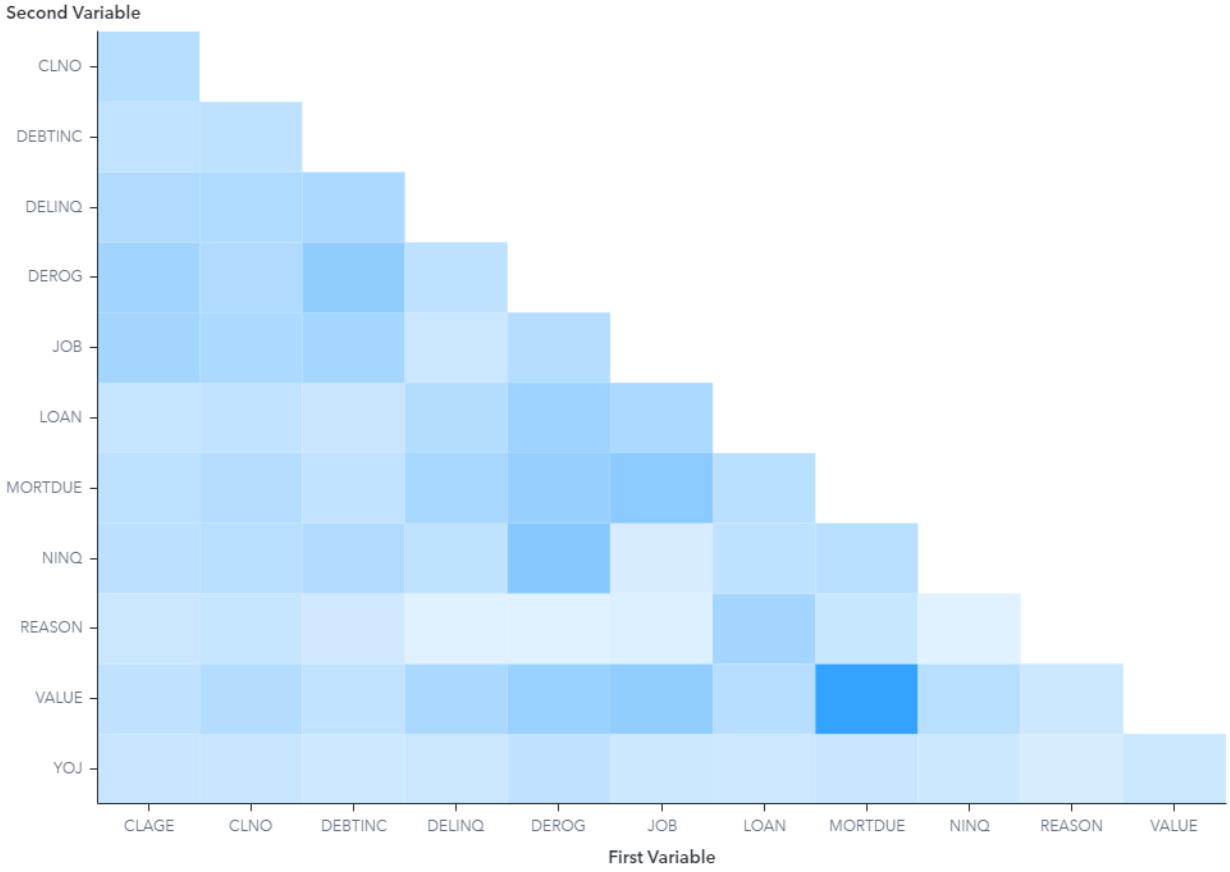
Mutual Information



Symmetric Uncertainty



Entropy Reduction



Table

First Variable	Second Variable	Mutual Information	Symmetric Uncertainty
MORTDUE	VALUE	1.5367	0.3079
CLNO	VALUE	0.3760	0.0767
CLAGE	CLNO	0.3712	0.0758
LOAN	VALUE	0.3658	0.0720
CLNO	MORTDUE	0.3630	0.0741
LOAN	MORTDUE	0.3360	0.0661
JOB	MORTDUE	0.3112	0.0640
CLAGE	MORTDUE	0.3033	0.0601
CLNO	DEBTINC	0.2950	0.0602
JOB	VALUE	0.2879	0.0597
CLAGE	VALUE	0.2801	0.0556
DEBTINC	MORTDUE	0.2646	0.0540
DEBTINC	VALUE	0.2639	0.0538
CLAGE	DEBTINC	0.2593	0.0515
CLNO	LOAN	0.2471	0.0504
CLAGE	LOAN	0.2288	0.0454
CLAGE	JOB	0.2205	0.0463
DEBTINC	JOB	0.2133	0.0440
CLAGE	YOJ	0.2058	0.0439
MORTDUE	YOJ	0.2057	0.0441
CLNO	YOJ	0.2026	0.0447
VALUE	YOJ	0.2001	0.0424
DEBTINC	LOAN	0.1979	0.0404
CLNO	JOB	0.1918	0.0527
JOB	LOAN	0.1892	0.0401
DEBTINC	NINQ	0.1725	0.0425
LOAN	YOJ	0.1706	0.0362

First Variable	Second Variable	Mutual Information	Symmetric Uncertainty
DEBTINC	YOJ	0.1586	0.0347
MORTDUE	NINQ	0.1538	0.0372
NINQ	VALUE	0.1494	0.0363
CLNO	NINQ	0.1486	0.0410
CLAGE	NINQ	0.1365	0.0334
LOAN	NINQ	0.1281	0.0311
DEROG	NINQ	0.1244	0.0833
DELINQ	MORTDUE	0.1219	0.0326
DELINQ	VALUE	0.1154	0.0310
DEROG	MORTDUE	0.1045	0.0283
CLNO	DELINQ	0.1034	0.0327
CLAGE	DELINQ	0.1029	0.0278
DEROG	VALUE	0.1019	0.0276
LOAN	REASON	0.0969	0.0271
DEROG	LOAN	0.0960	0.0263
DELINQ	LOAN	0.0952	0.0257
CLAGE	DEROG	0.0911	0.0249
DEBTINC	DEROG	0.0910	0.0252
DEBTINC	DELINQ	0.0883	0.0245
NINQ	YOJ	0.0771	0.0229
JOB	YOJ	0.0725	0.0213
CLNO	DEROG	0.0695	0.0235
DELINQ	NINQ	0.0681	0.0404
DELINQ	DEROG	0.0672	0.0702
DEROG	JOB	0.0645	0.0435
DELINQ	JOB	0.0462	0.0277
DELINQ	YOJ	0.0457	0.0156
DEROG	YOJ	0.0445	0.0163

First Variable	Second Variable	Mutual Information	Symmetric Uncertainty	Entropy Percent Reduction
CLNO	REASON	0.0409	0.0137	30.7348
MORTDUE	REASON	0.0377	0.0106	7.8920
CLAGE	REASON	0.0333	0.0094	7.3859
JOB	NINQ	0.0314	0.0148	7.2007
REASON	VALUE	0.0307	0.0086	7.6461
DEBTINC	REASON	0.0247	0.0069	6.6151
REASON	YOJ	0.0155	0.0056	14.8259
JOB	REASON	0.0153	0.0103	6.0630
NINQ	REASON	0	0	6.2029
DEROG	REASON	0	0	13.9494
DELINQ	REASON	0	0	5.6151
				5.3964
				5.3779
				5.1996
				5.1874

Entropy Percent Reduction
4.5864
10.6864
10.3611
4.1075
4.1165
4.2526
3.9383
4.0324
9.1983
9.1675
8.4998
3.3602
3.2323
7.0529
6.9143
6.8791
6.3154
5.9292
15.7359
9.9586
9.5327
12.8390
8.5382
8.4968
12.3327
10.8183
11.6230
7.8674

Entropy Percent Reduction
11.0258
14.1200
9.1490
3.5406
3.4435
8.4167
5.7059
6.0029
7.8029
3.8189
3.7324
5.5397
4.5941
4.3340
3.7140
1.5073
3.4265
2.7913
1.7354
0.7330
0
0
0

Properties

Property Name	Property Value
classLevelsCutoff	32
clusterIntervalVars	false
codeLocation	mlearning
dataMiningVersion	V2024.06
domModeCutoffPct	80
flatCutoff	-1
fullDatasetReconstitution	false
minNumVarsToCluster	20
missingCutoffPct	25
numIntervalBins	8
numVariableThreshold	50
peakyCutoff	2
rareModeCutoffPct	20
selectionCriterion	IMPORTANCE
skewedCutoff	1
sourcePartition	ALL
templateRevision	1
tsnePerplexity	30
tsneProjection	false

Output

The SAS System

The TREESPLIT Procedure

Model Information	
Split Criterion	IGR
Pruning Method	None
Max Branches per Node	2
Max Tree Depth	10
Tree Depth Before Pruning	10
Tree Depth After Pruning	10
Number of Leaves Before Pruning	43
Number of Leaves After Pruning	43

	Training
Number of Observations Read	5960
Number of Observations Used	5960

Missing Values

Variable Name	Number of Missing Values	Percentage Missing
BAD	0	0.0000
CLAGE	0	0.0000
CLNO	222	3.7248
DEBTINC	1267	21.2584
DELINQ	580	9.7315
DEROG	708	11.8792
JOB	0	0.0000
LOAN	0	0.0000
MORTDUE	518	8.6913
NINQ	510	8.5570
REASON	252	4.2282
VALUE	0	0.0000
YOJ	515	8.6409

The SAS System

The TREESPLIT Procedure

Fit Statistics for Selected Tree		
	Number of Leaves	Misclassification Rate
Training	43	0.1117

Variable Importance			
	Training		
Variable	Importance	Relative Importance	Count
DEBTINC	711.43	1.0000	6
CLAGE	45.8828	0.0645	4
DELINQ	43.6508	0.0614	6
DEROG	28.9457	0.0407	6
LOAN	19.6009	0.0276	4
YOJ	16.9115	0.0238	2
NINQ	10.0068	0.0141	3
MORTDUE	8.3586	0.0117	4
VALUE	8.3414	0.0117	4
CLNO	6.6283	0.0093	2
REASON	1.2000	0.0017	1

Missing Values

Variable Name	Number of Missing Values	Percentage Missing
BAD	0	0.0000
CLAGE	0	0.0000
CLNO	222	3.7248
DEBTINC	1267	21.2584
DELINQ	580	9.7315
DEROG	708	11.8792
JOB	0	0.0000
LOAN	0	0.0000
MORTDUE	518	8.6913
NINQ	510	8.5570
REASON	252	4.2282
VALUE	0	0.0000
YOJ	515	8.6409