# Exploratory Data Analysis Report

Sala2Code

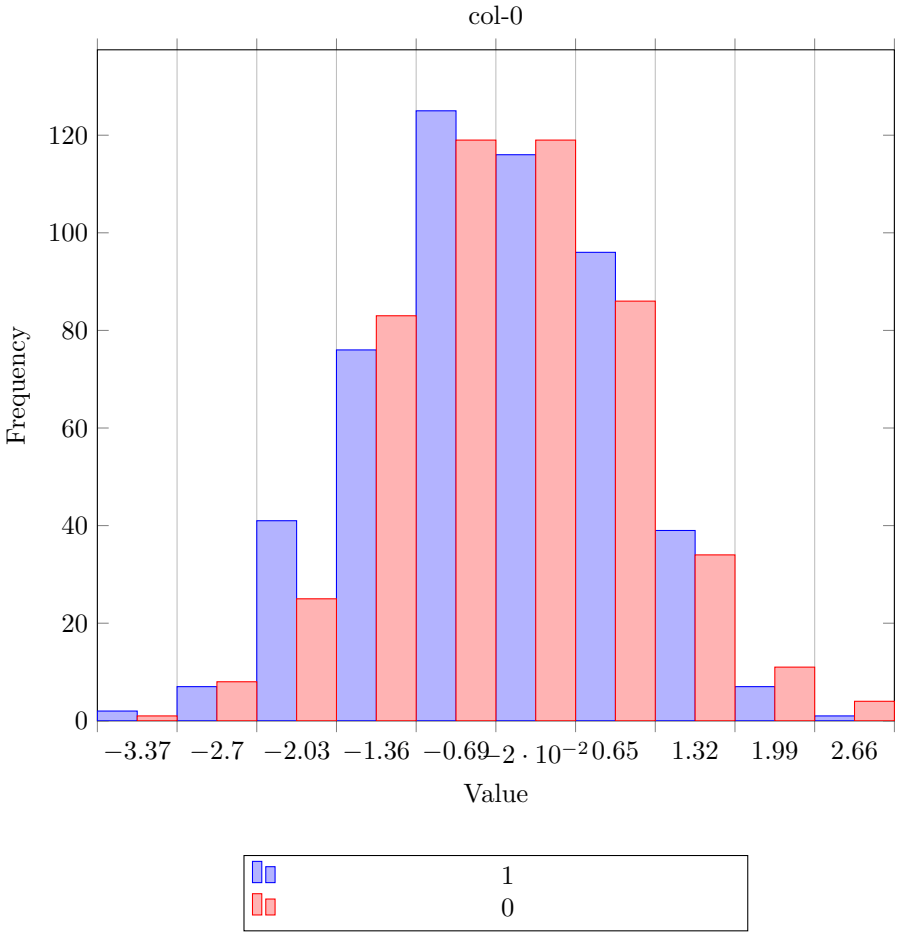
March 9, 2024

## 1 Dataset Overview

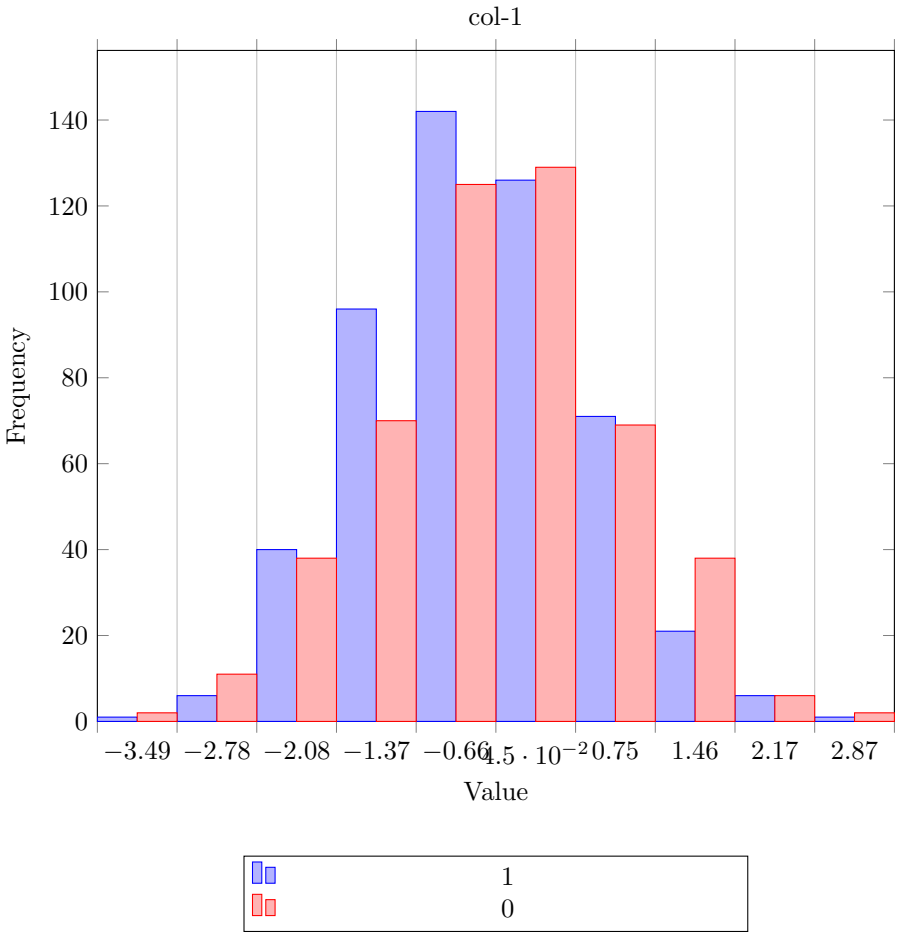
### 1.1 General Information

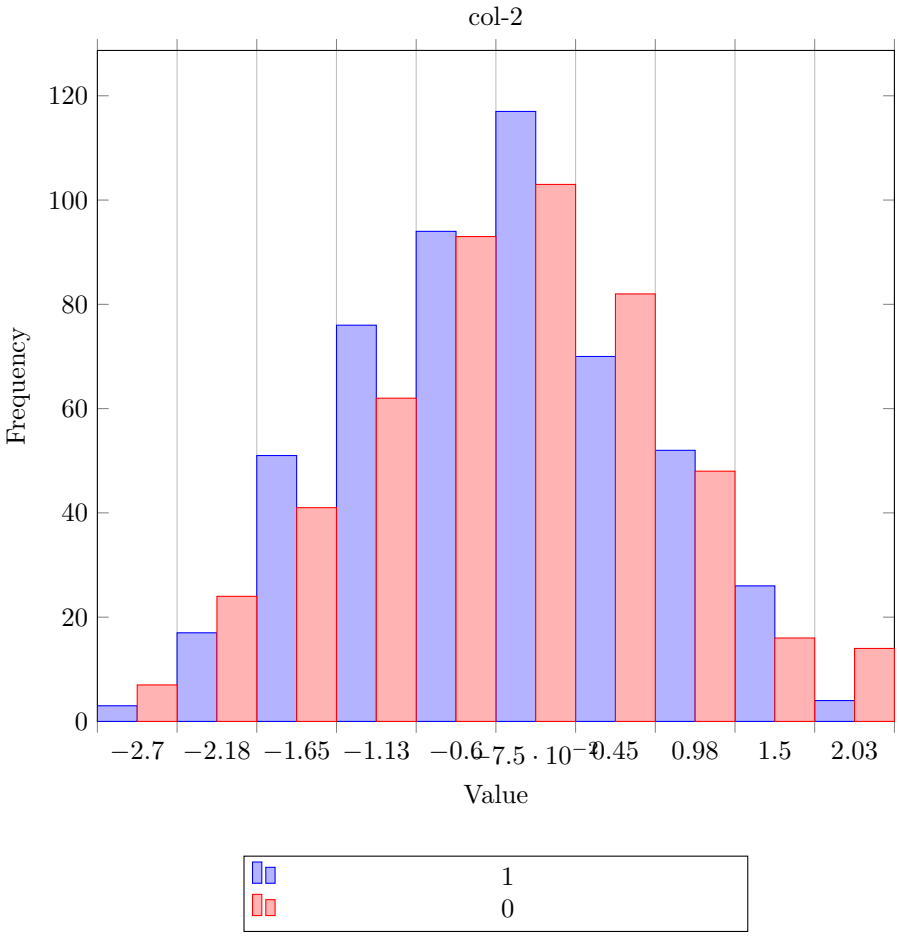
- Size: *1000* rows / *41* columns
- Target column: *col-40*

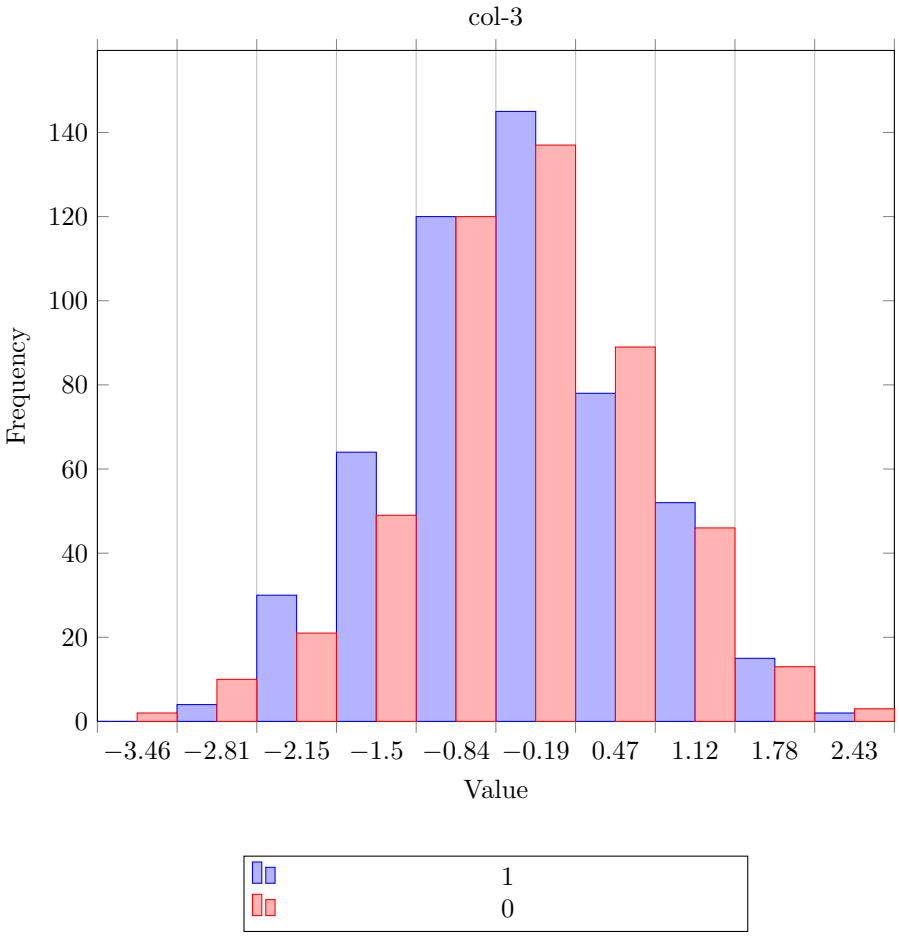
## 1.2 Continuous Columns

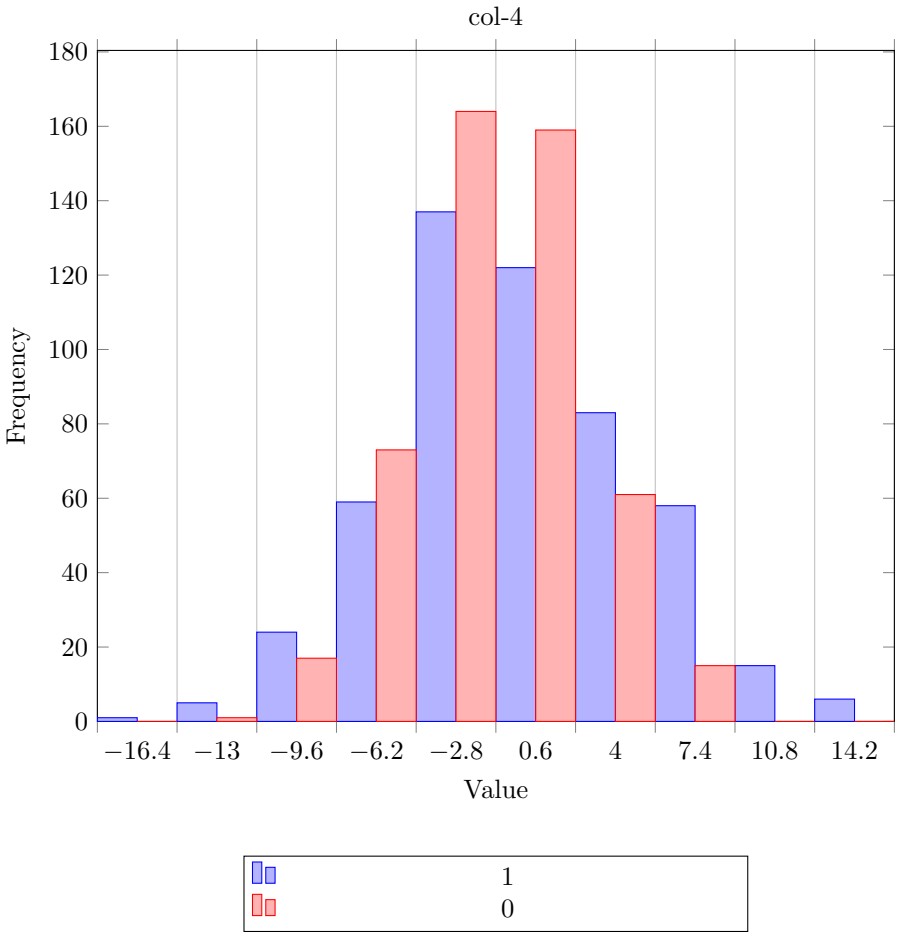


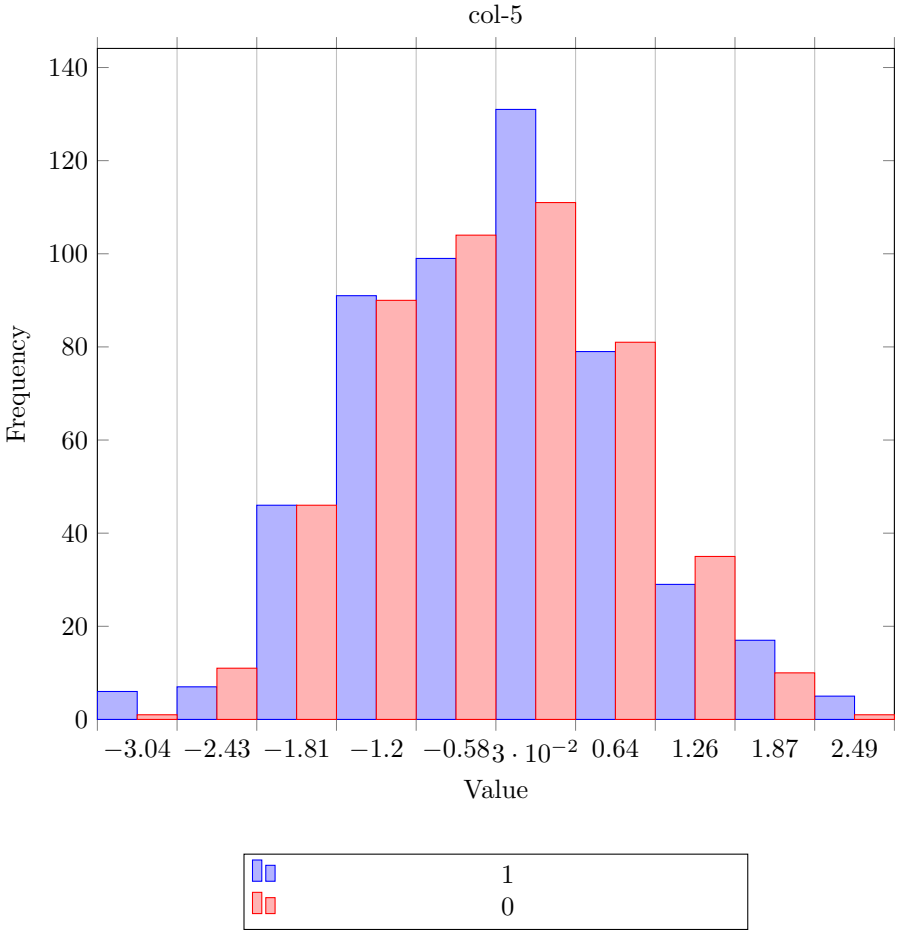

[1]Consider the first value in each column. Therefore, it is not precise.

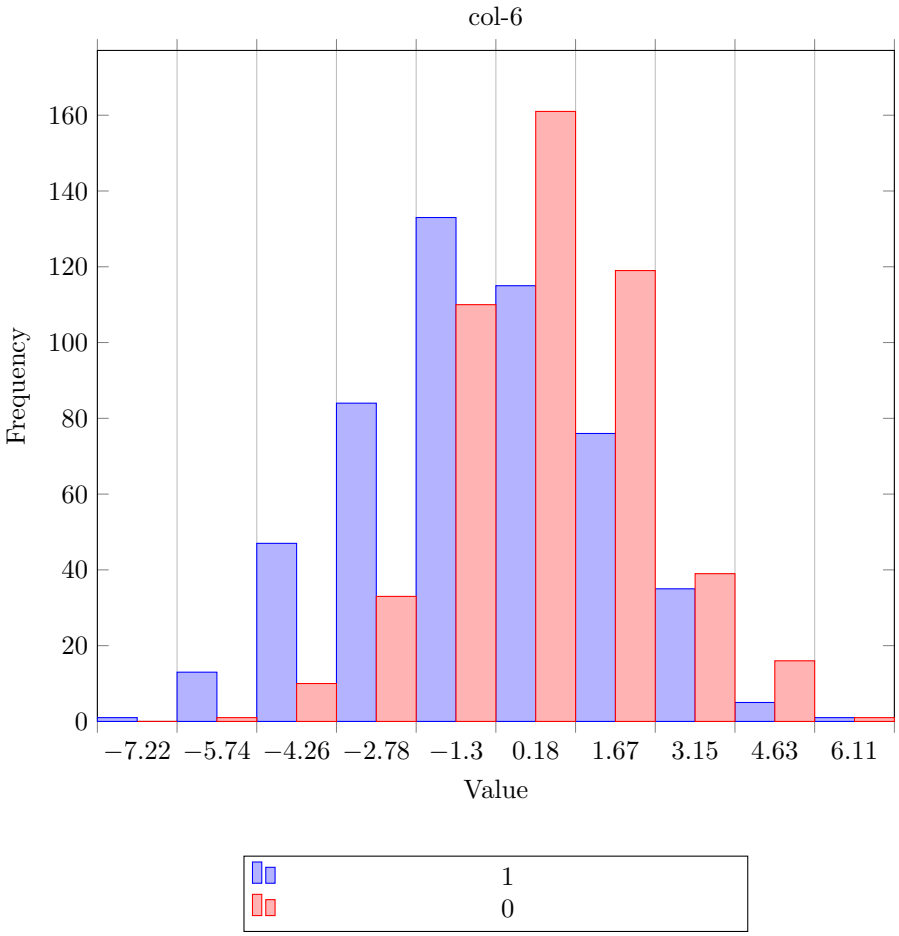

col-1
Frequency
0
20
40
60
80
100
120
140
−3.49
−2.78
−2.08
−1.37
−0.66
4.5 · 10−2
0.75
1.46
2.17
2.87
Value
1
0

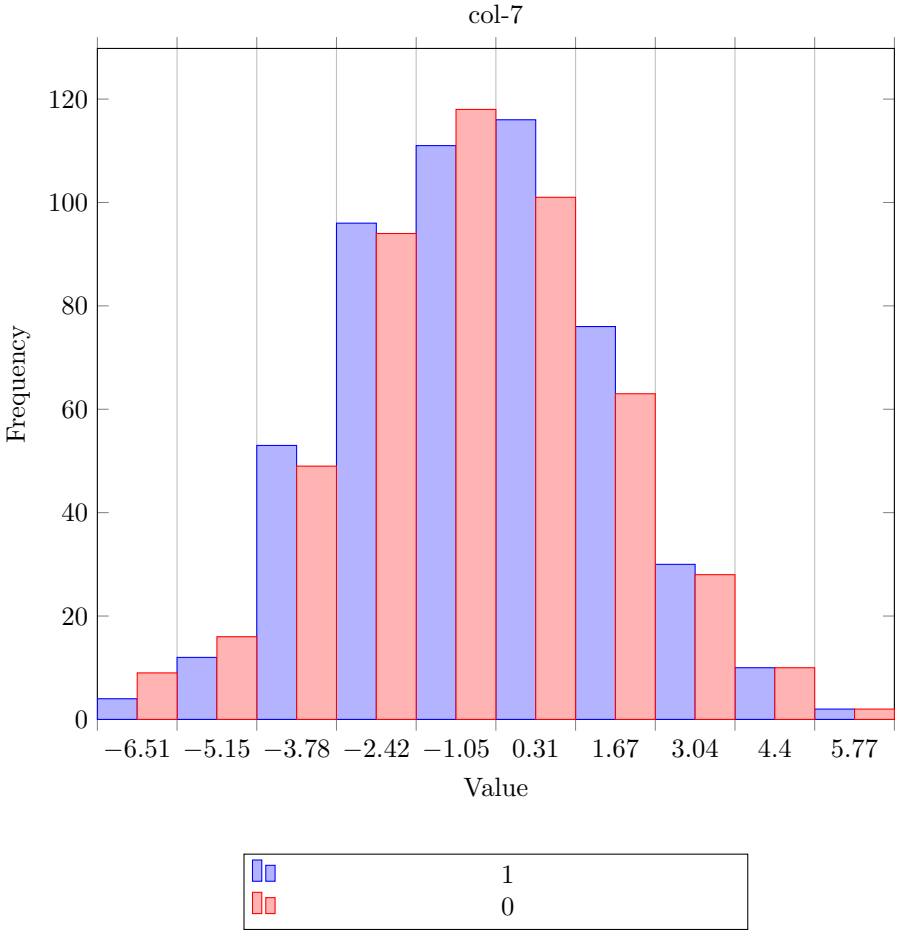

col-2
Frequency
120
100
80
60
40
20
0
−2.7
−2.18
−1.65
−1.13
−0.6
−7.5 · 10−4
0.45
0.98
1.5
2.03
Value
1
0

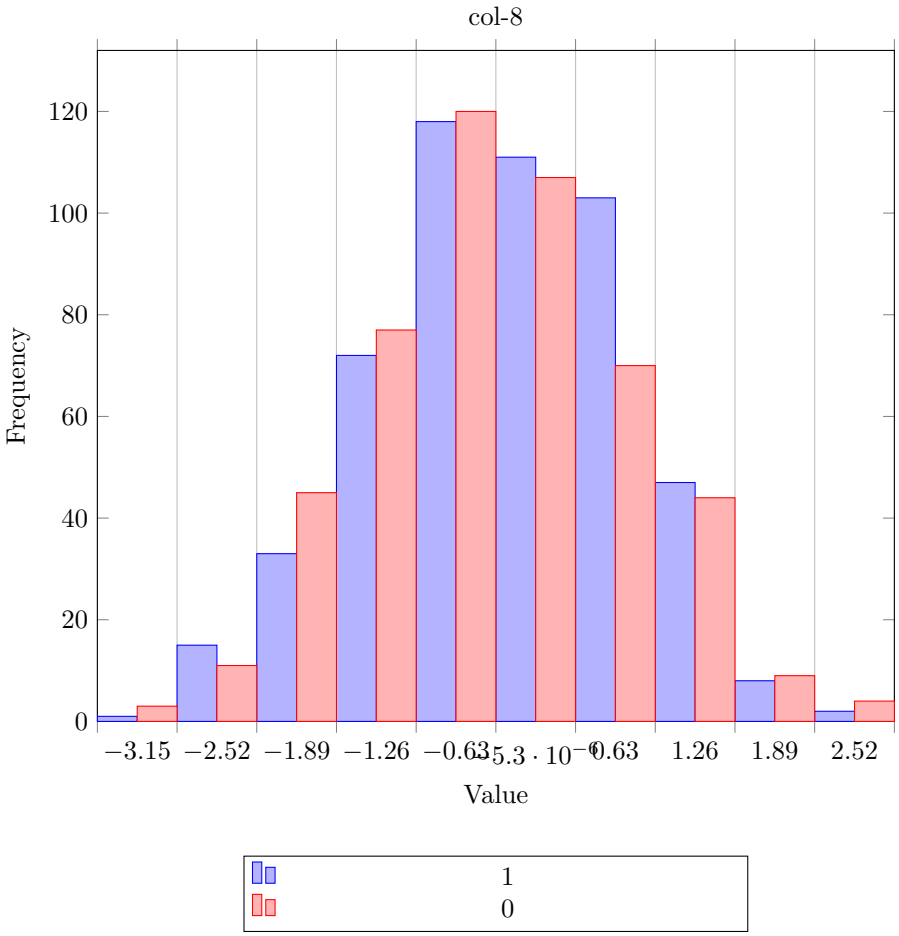

col-3
Frequency
0
20
40
60
80
100
120
140
−3.46
−2.81
−2.15
−1.5
−0.84
−0.19
0.47
1.12
1.78
2.43
Value
1
0

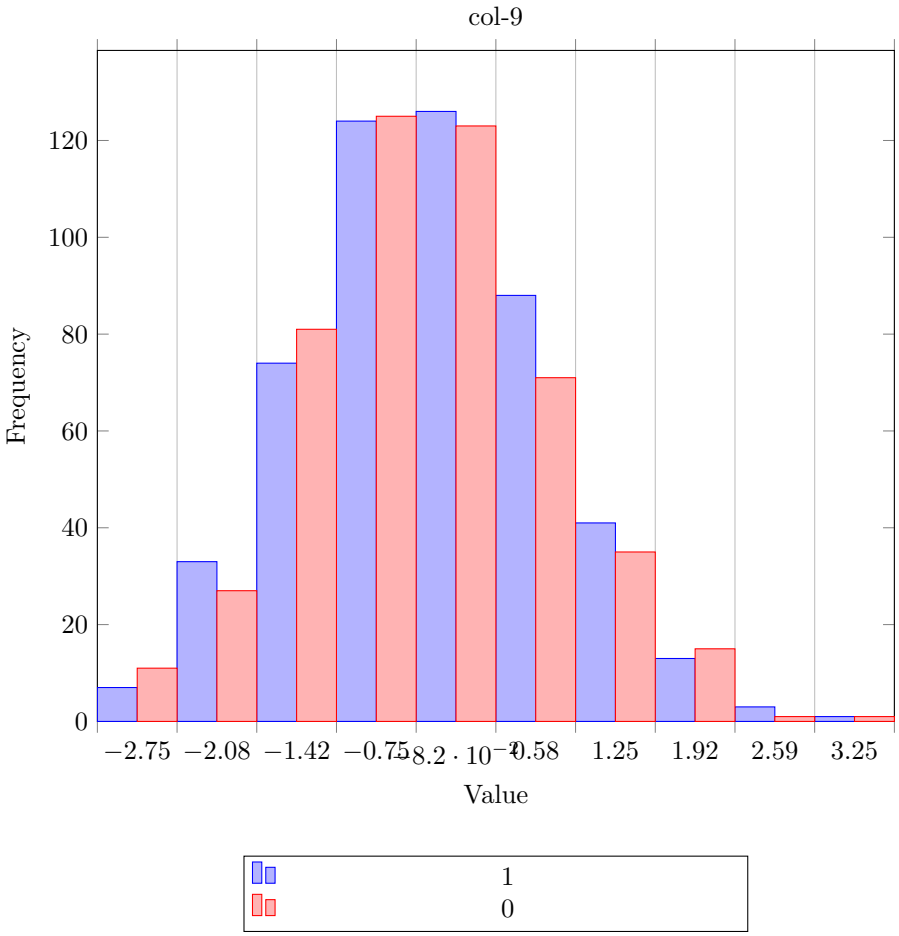

col-4
180
160
140
120
100
80
60
40
20
0
Frequency
−16.4
−13
−9.6
−6.2
−2.8
0.6
4
7.4
10.8
14.2
Value
1
0

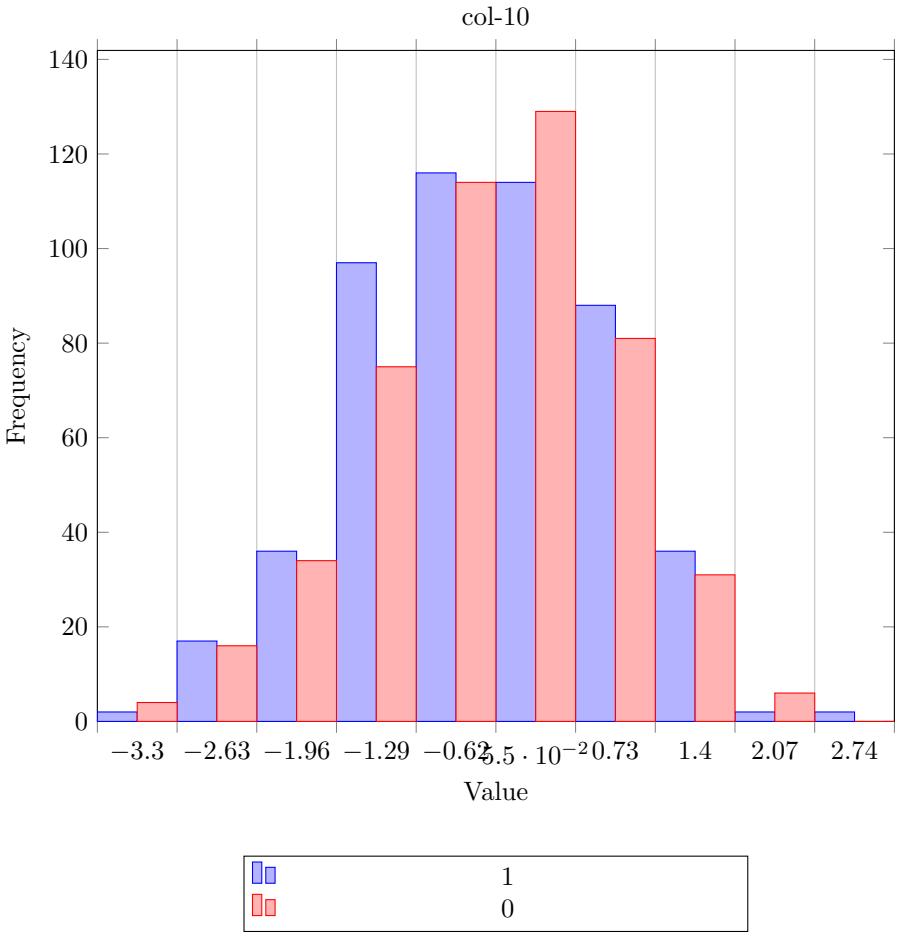

col-5
Frequency
140
120
100
80
60
40
20
0
−3.04
−2.43
−1.81
−1.2
−0.58
3 . 10−2
0.64
1.26
1.87
2.49
Value
1
0

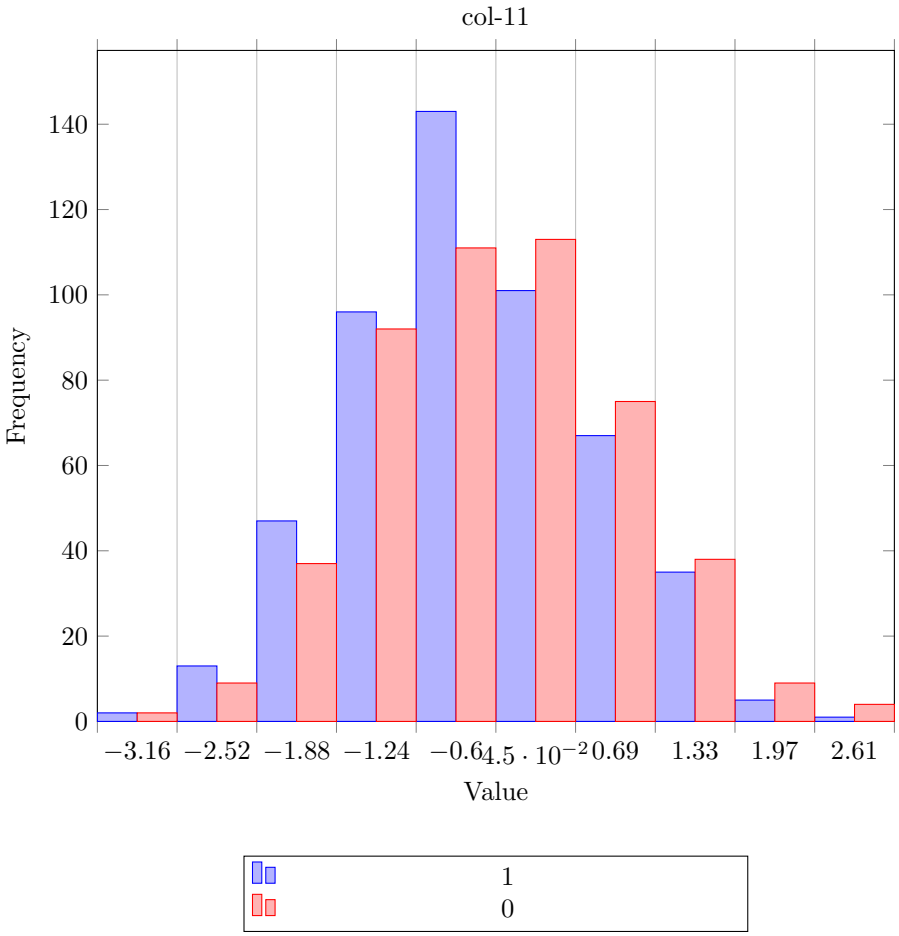

col-6
Frequency
0
20
40
60
80
100
120
140
160
−7.22
−5.74
−4.26
−2.78
−1.3
0.18
1.67
3.15
4.63
6.11
Value
1
0

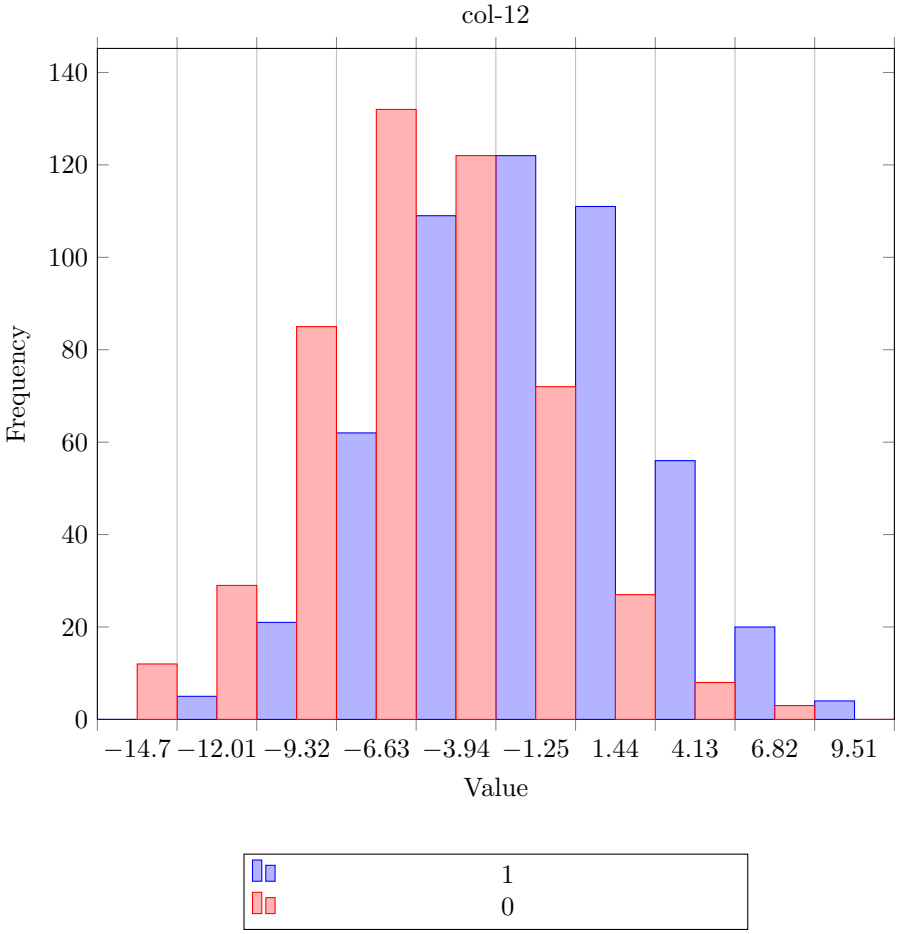

col-7
Frequency
120
100
80
60
40
20
0
−6.51
−5.15
−3.78
−2.42
−1.05
0.31
1.67
3.04
4.4
5.77
Value
1
0

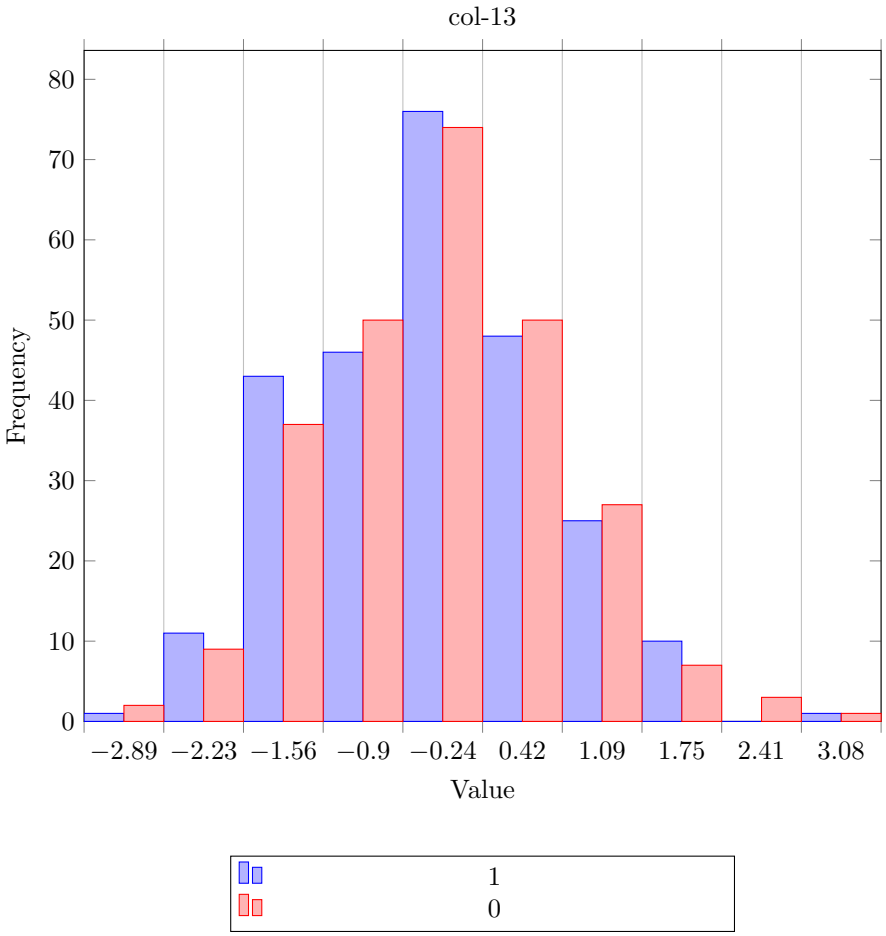

col-8
Frequency
120
100
80
60
40
20
0
−3.15
−2.52
−1.89
−1.26
−0.63
5.3 · 10−9
0.63
1.26
1.89
2.52
Value
1
0

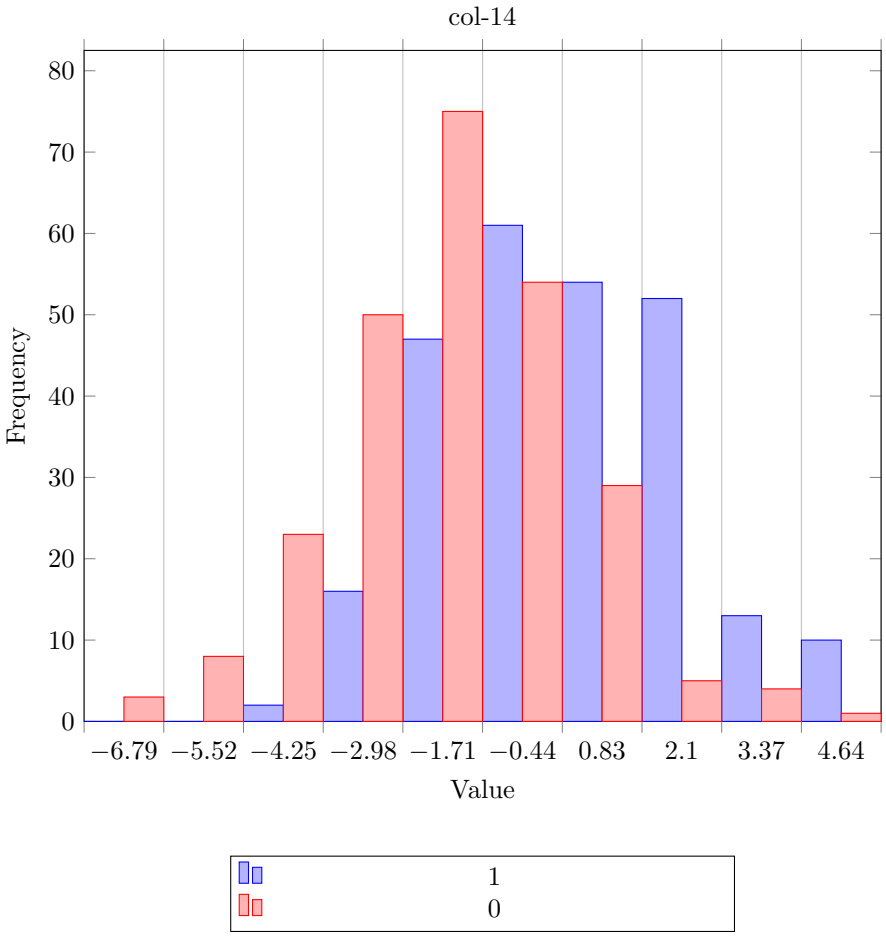

col-9
Frequency
0
20
40
60
80
100
120
−2.75
−2.08
−1.42
−0.75
−8.2 · 10−4
0.58
1.25
1.92
2.59
3.25
Value
1
0

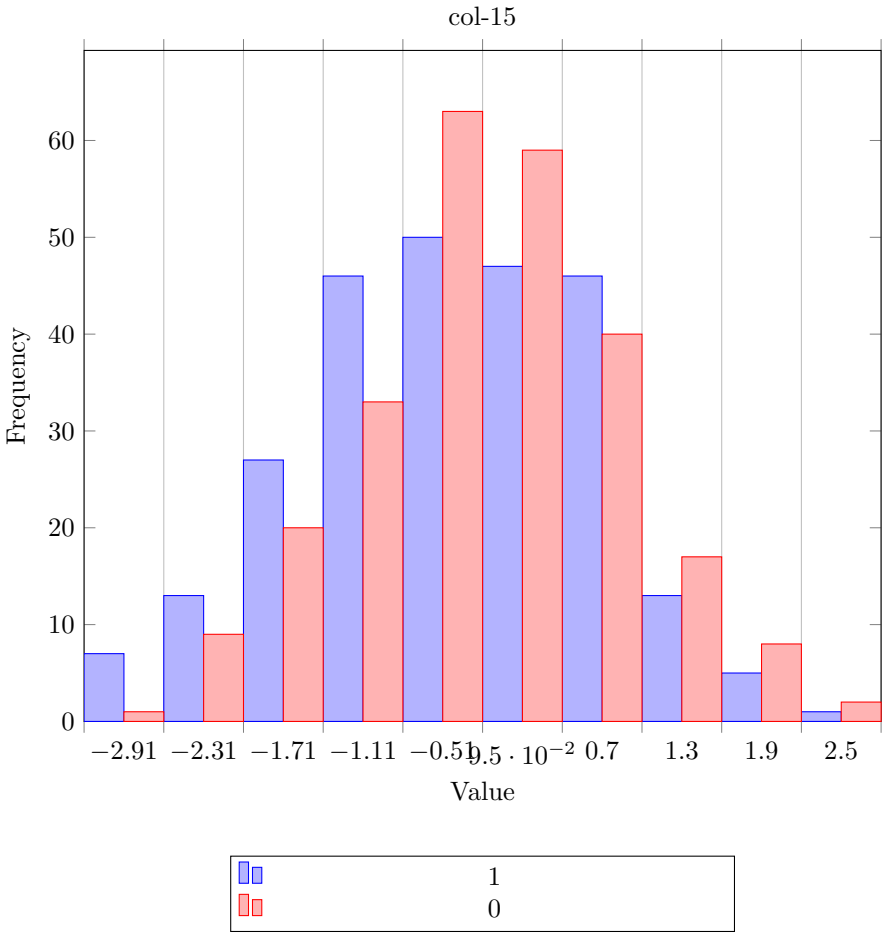

col-10
140
120
100
80
60
40
20
0
Frequency
−3.3
−2.63
−1.96
−1.29
−0.62
$5.5 \cdot 10^{-2}$
0.73
1.4
2.07
2.74
Value
1
0

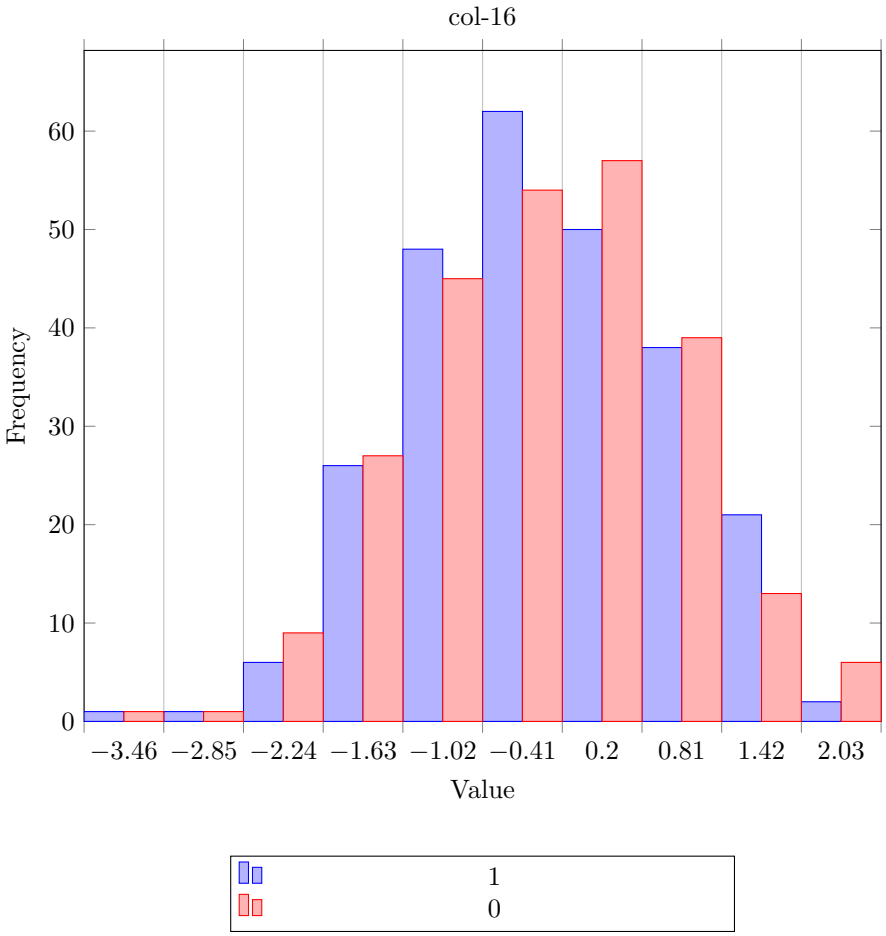

col-11
Frequency
0
20
40
60
80
100
120
140
−3.16
−2.52
−1.88
−1.24
−0.6
4.5 · 10−2
0.69
1.33
1.97
2.61
Value
1
0

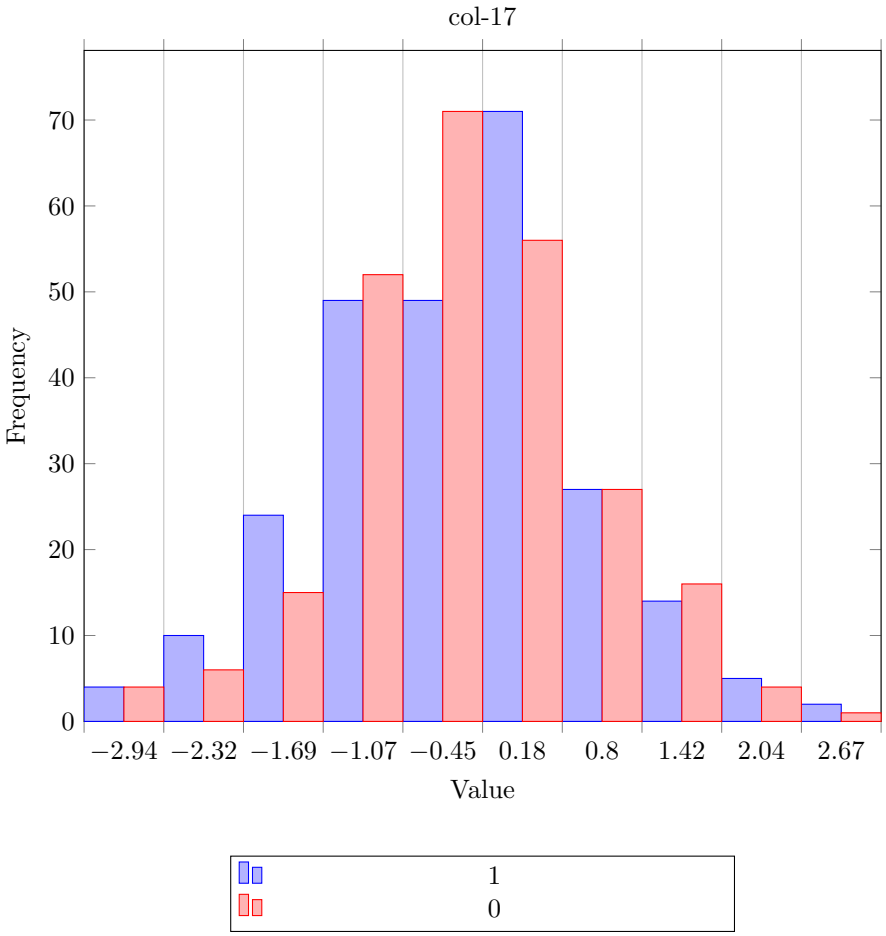

col-12
Frequency
140
120
100
80
60
40
20
0
−14.7
−12.01
−9.32
−6.63
−3.94
−1.25
1.44
4.13
6.82
9.51
Value
1
0

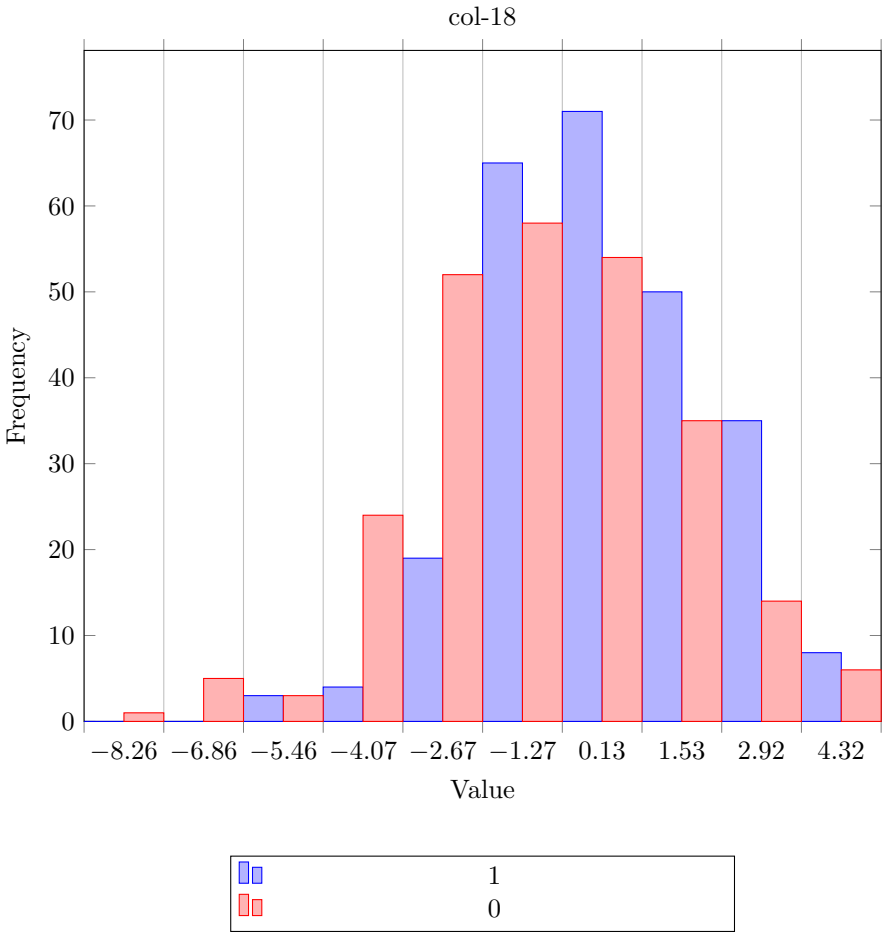

col-13
Frequency
0
10
20
30
40
50
60
70
80
−2.89
−2.23
−1.56
−0.9
−0.24
0.42
1.09
1.75
2.41
3.08
Value
1
0

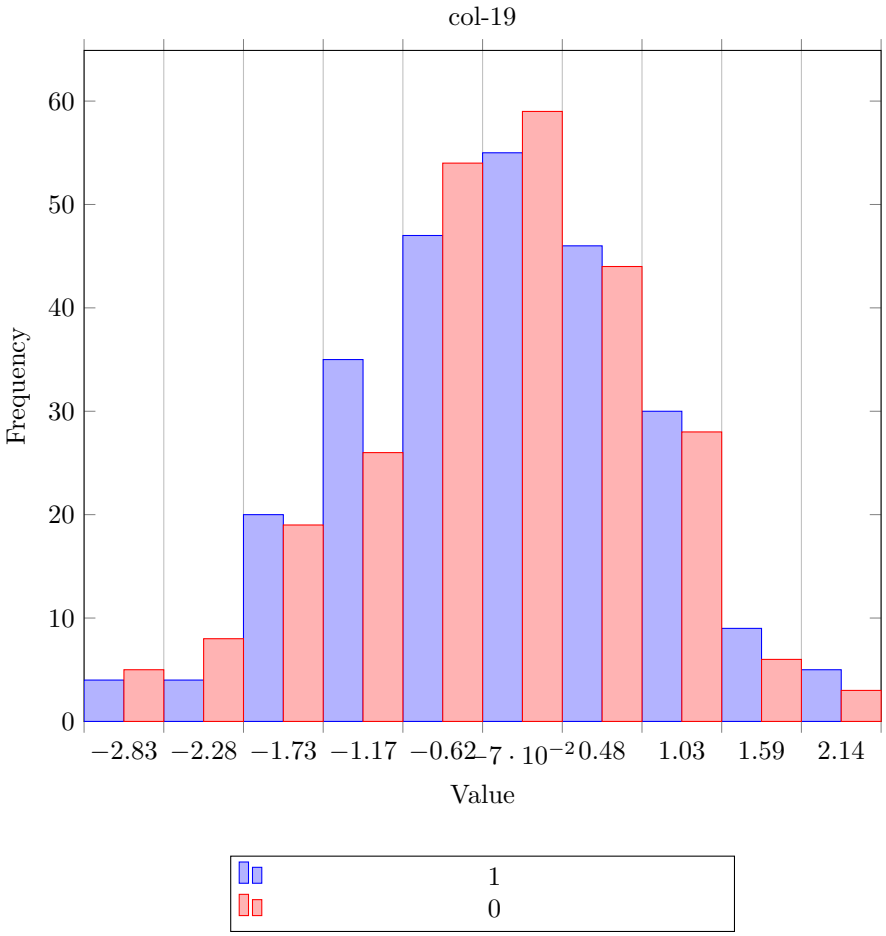

col-14
Frequency
Value
80
70
60
50
40
30
20
10
0
−6.79
−5.52
−4.25
−2.98
−1.71
−0.44
0.83
2.1
3.37
4.64
1
0

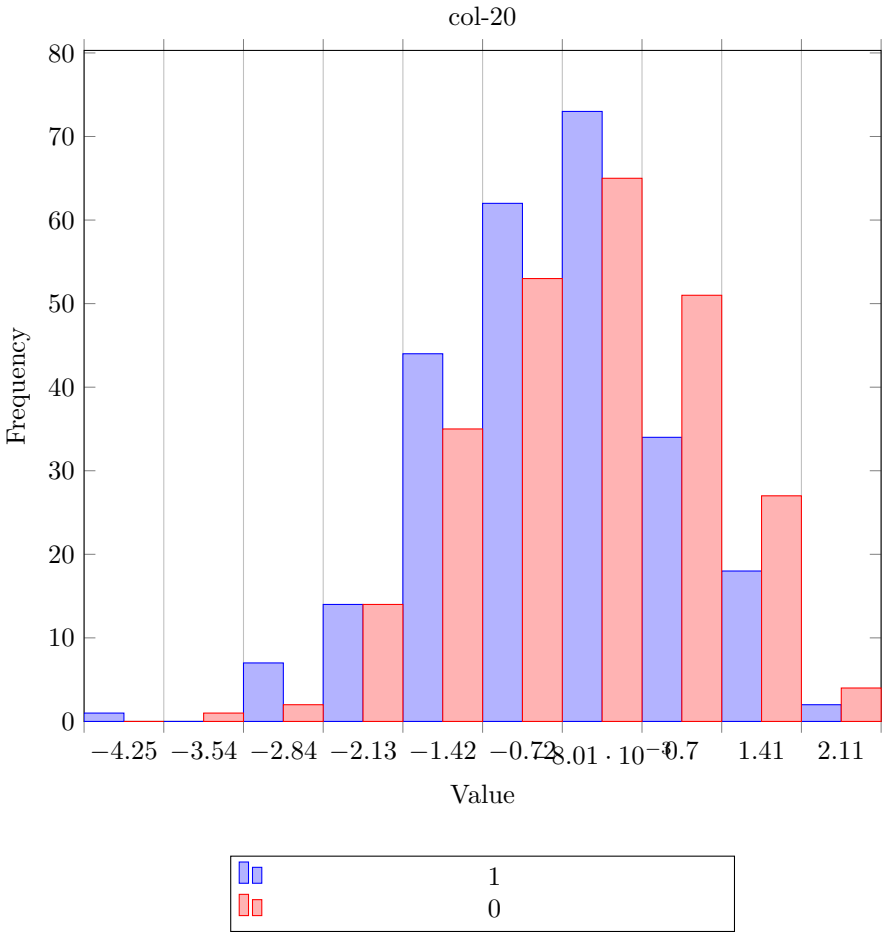

col-15
Frequency
0
10
20
30
40
50
60
−2.91
−2.31
−1.71
−1.11
−0.51
9.5 · 10−2
0.7
1.3
1.9
2.5
Value
1
0

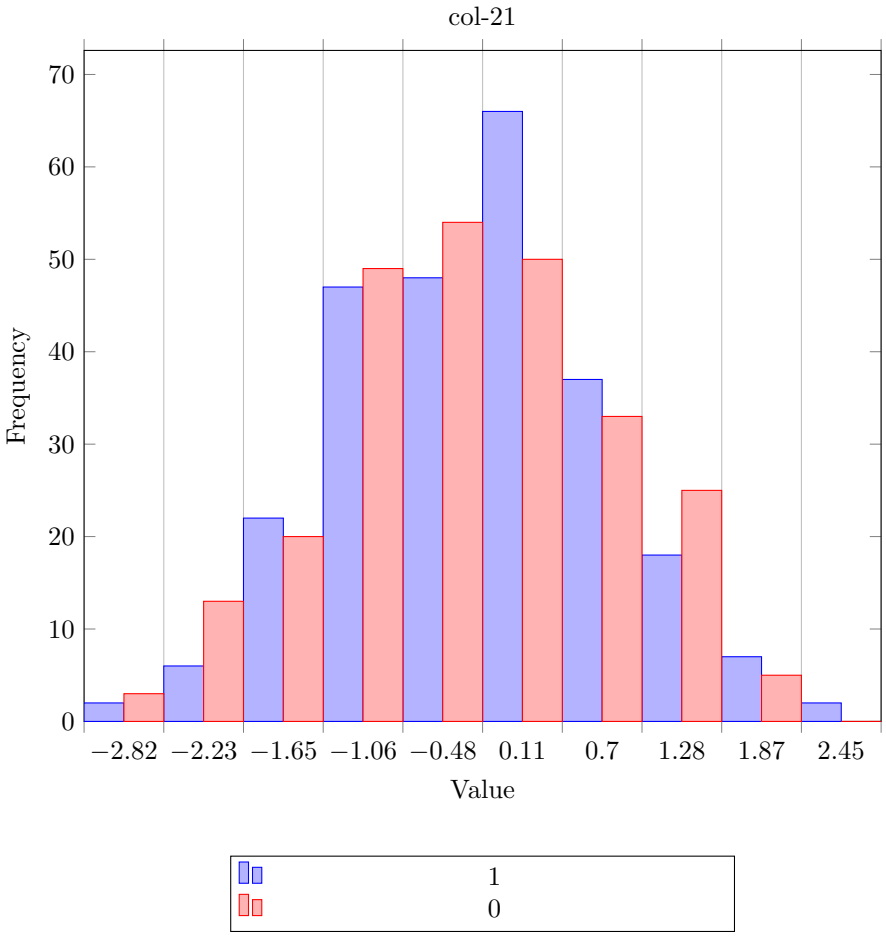

col-16
Frequency
0
10
20
30
40
50
60
−3.46
−2.85
−2.24
−1.63
−1.02
−0.41
0.2
0.81
1.42
2.03
Value
1
0

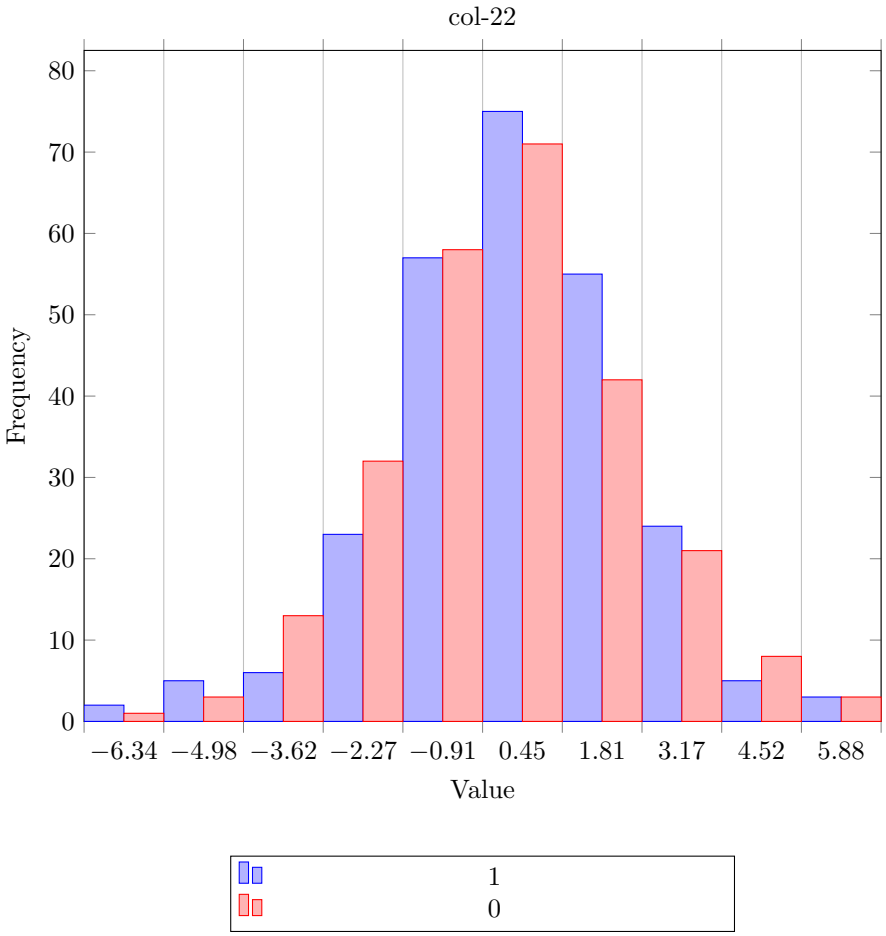

col-17
Frequency
70
60
50
40
30
20
10
0
−2.94
−2.32
−1.69
−1.07
−0.45
0.18
0.8
1.42
2.04
2.67
Value
1
0

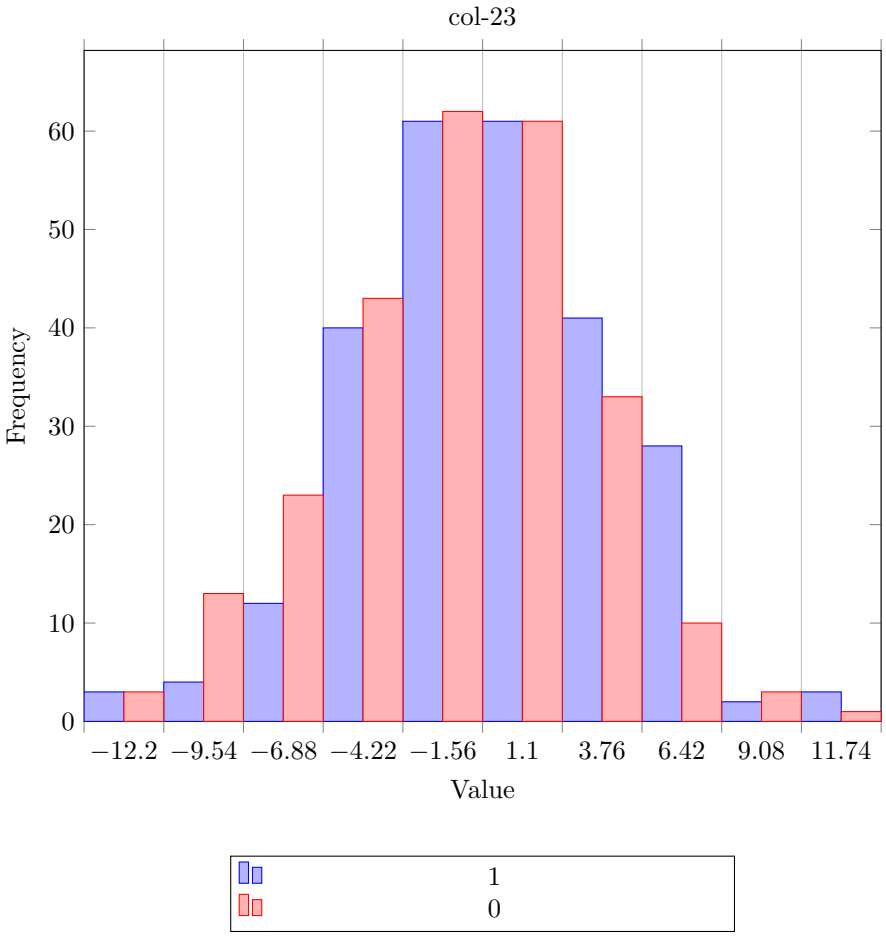

col-18
Frequency
70
60
50
40
30
20
10
0
−8.26
−6.86
−5.46
−4.07
−2.67
−1.27
0.13
1.53
2.92
4.32
Value
1
0

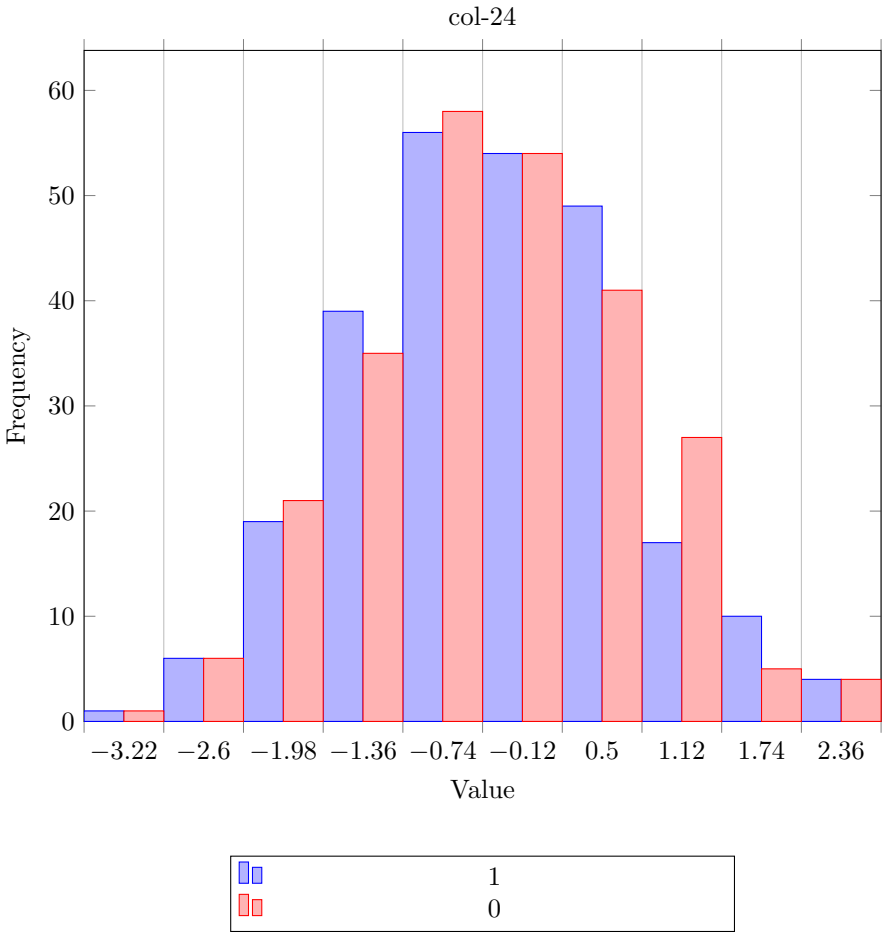

col-19
Frequency
0
10
20
30
40
50
60
−2.83
−2.28
−1.73
−1.17
−0.62
−7 · 10−2
0.48
1.03
1.59
2.14
Value
1
0

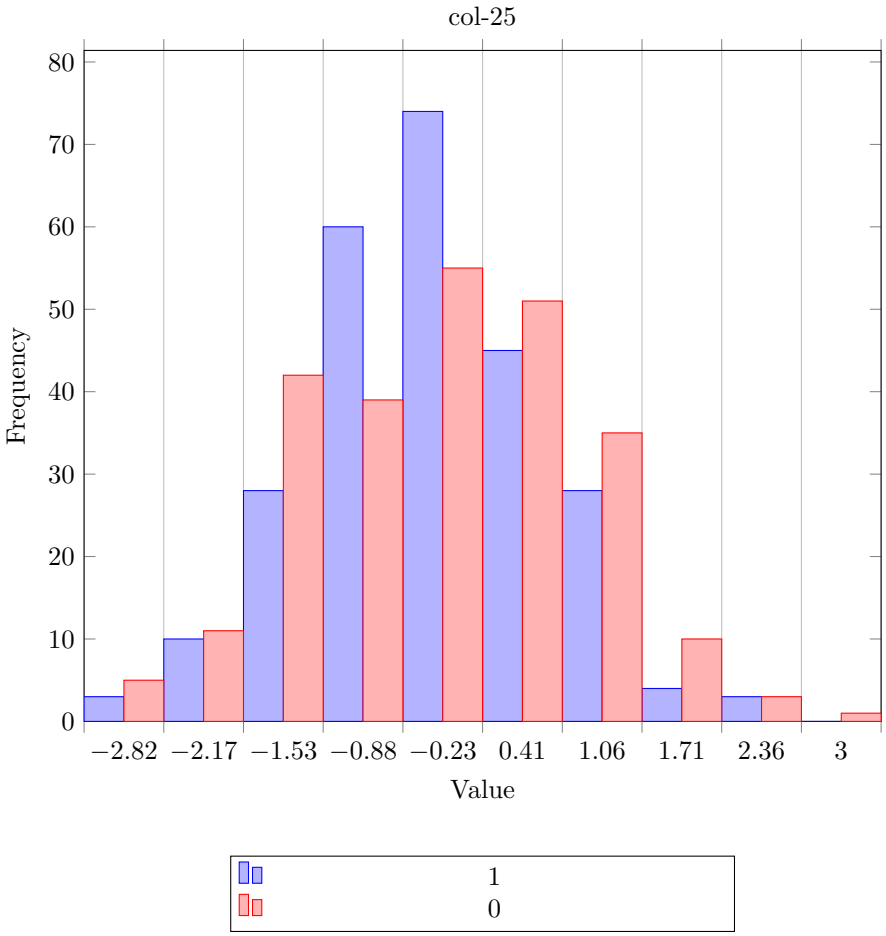

col-20
Frequency
Value
0
10
20
30
40
50
60
70
80
−4.25
−3.54
−2.84
−2.13
−1.42
−0.72
$8.01 \cdot 10^{-3}$
0.7
1.41
2.11
1
0

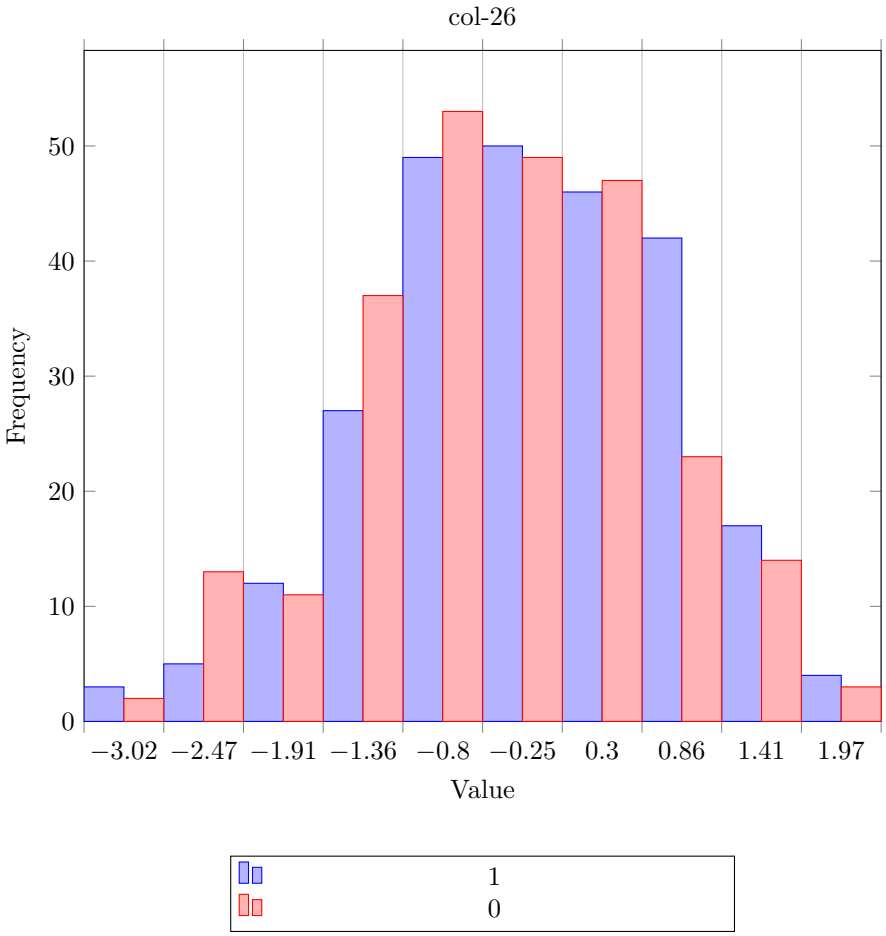

col-21
Frequency
70
60
50
40
30
20
10
0
−2.82
−2.23
−1.65
−1.06
−0.48
0.11
0.7
1.28
1.87
2.45
Value
1
0

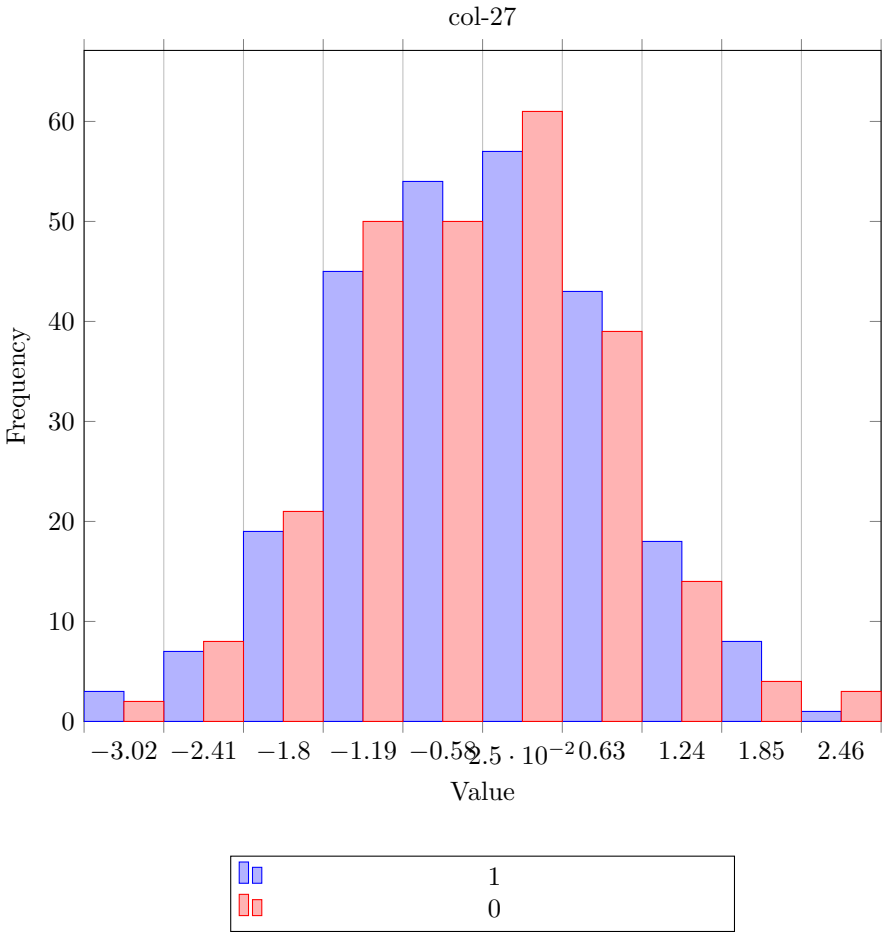

col-22
Frequency
80
70
60
50
40
30
20
10
0
−6.34
−4.98
−3.62
−2.27
−0.91
0.45
1.81
3.17
4.52
5.88
Value
1
0

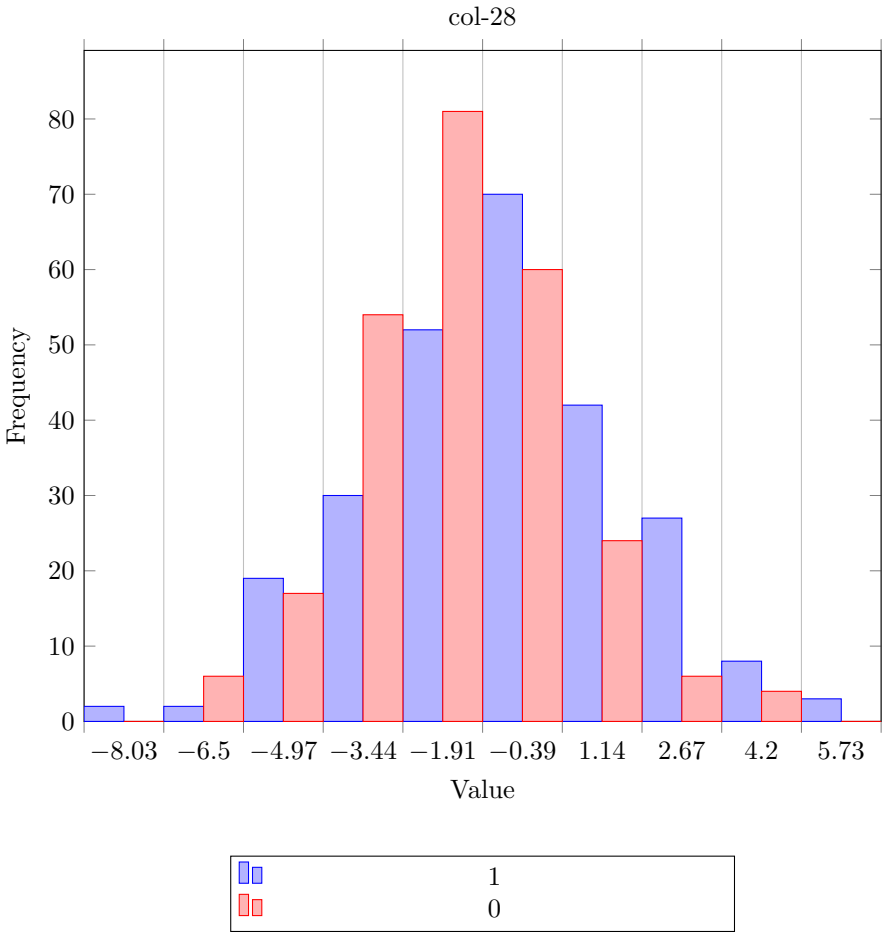

col-23
Frequency
0
10
20
30
40
50
60
−12.2
−9.54
−6.88
−4.22
−1.56
1.1
3.76
6.42
9.08
11.74
Value
1
0

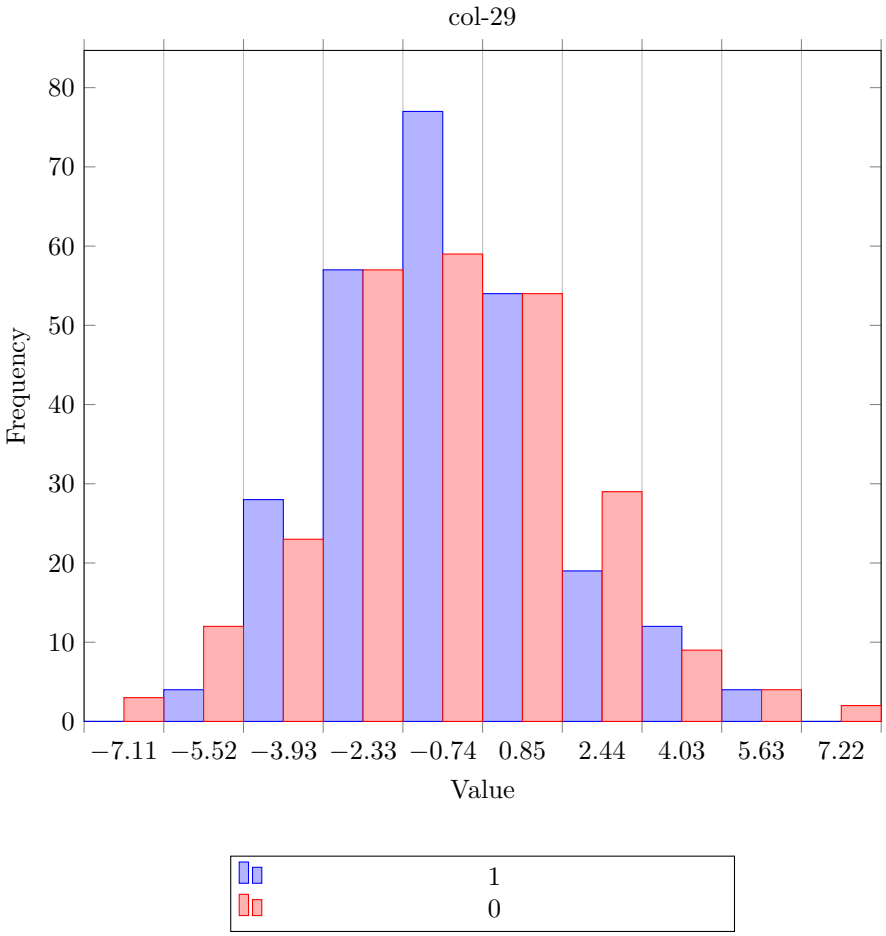

col-24
Frequency
0
10
20
30
40
50
60
−3.22
−2.6
−1.98
−1.36
−0.74
−0.12
0.5
1.12
1.74
2.36
Value
1
0

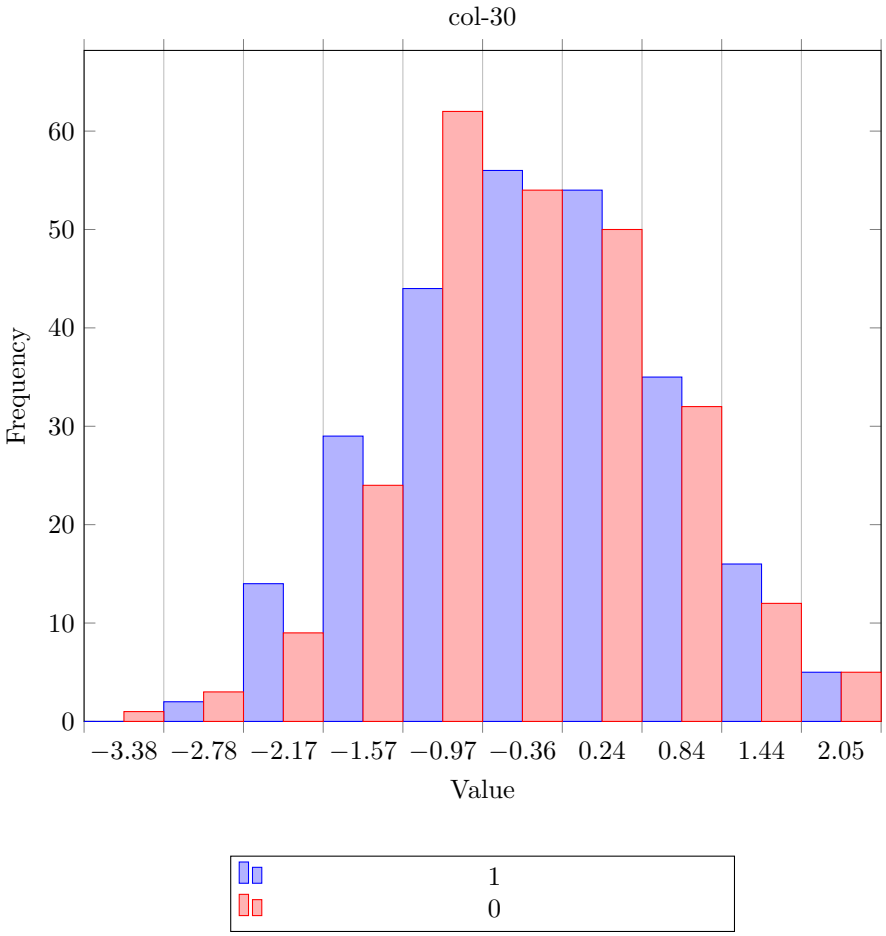

col-25
Frequency
80
70
60
50
40
30
20
10
0
−2.82
−2.17
−1.53
−0.88
−0.23
0.41
1.06
1.71
2.36
3
Value
1
0

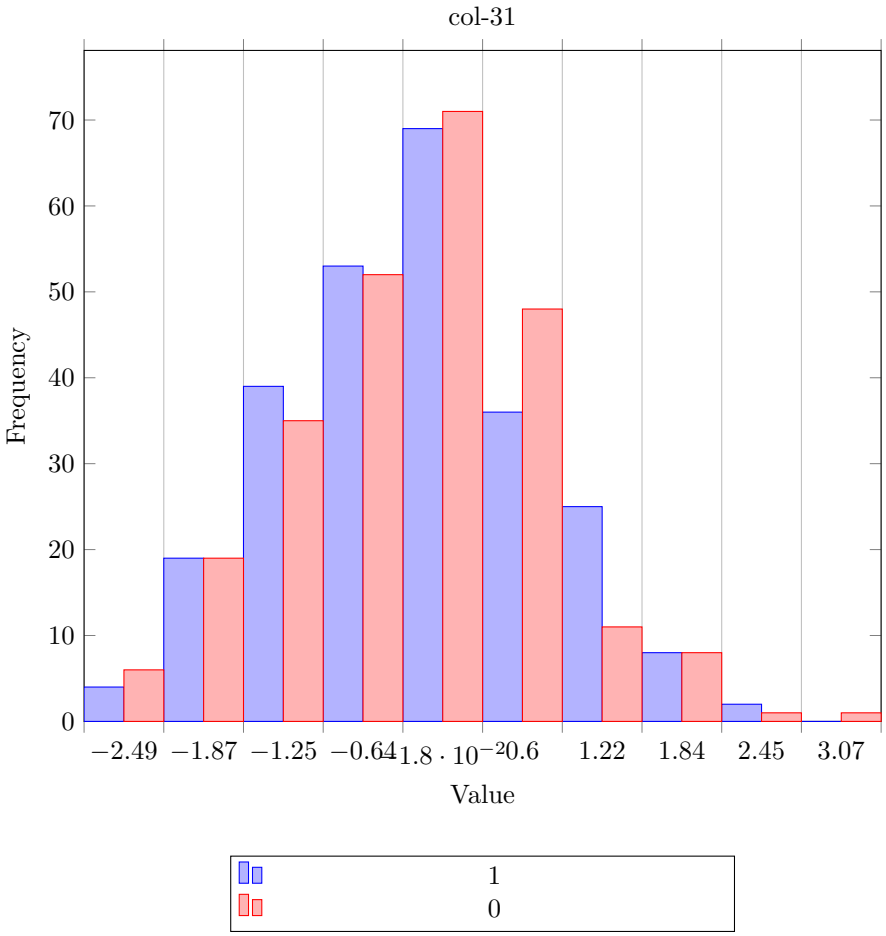

col-26
Frequency
0
10
20
30
40
50
−3.02
−2.47
−1.91
−1.36
−0.8
−0.25
0.3
0.86
1.41
1.97
Value
1
0

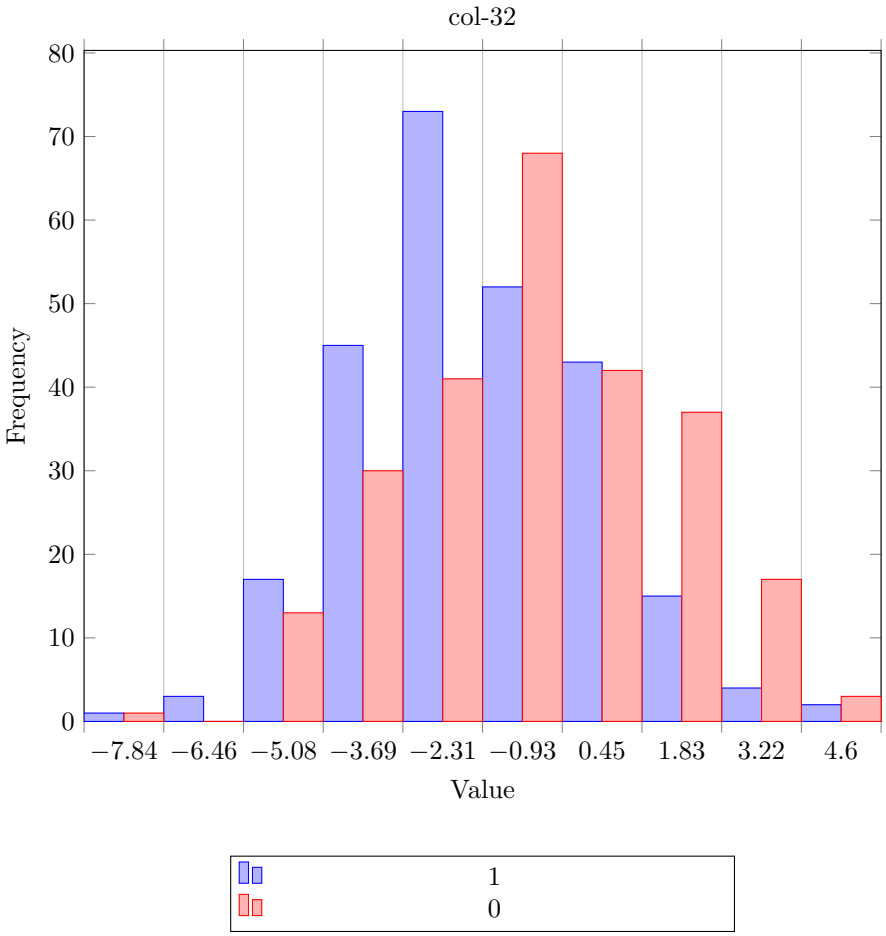

col-27
Frequency
0
10
20
30
40
50
60
−3.02
−2.41
−1.8
−1.19
−0.58
2.5 · 10−2
0.63
1.24
1.85
2.46
Value
1
0

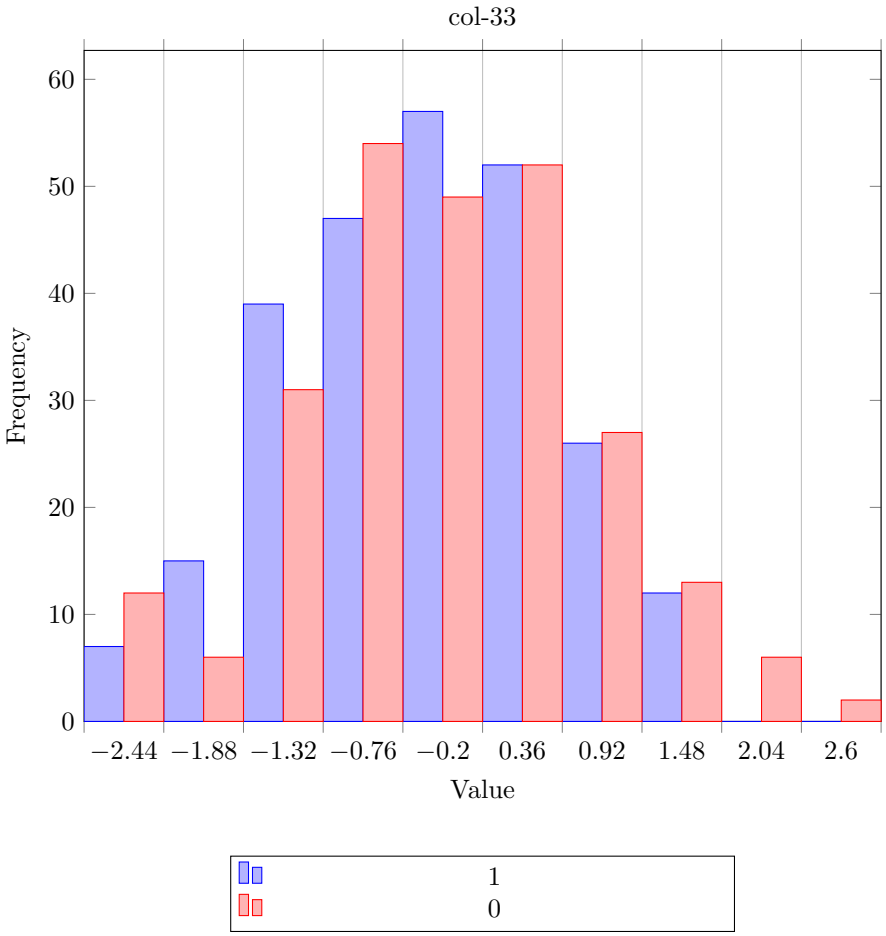

col-28
Frequency
0
10
20
30
40
50
60
70
80
−8.03
−6.5
−4.97
−3.44
−1.91
−0.39
1.14
2.67
4.2
5.73
Value
1
0

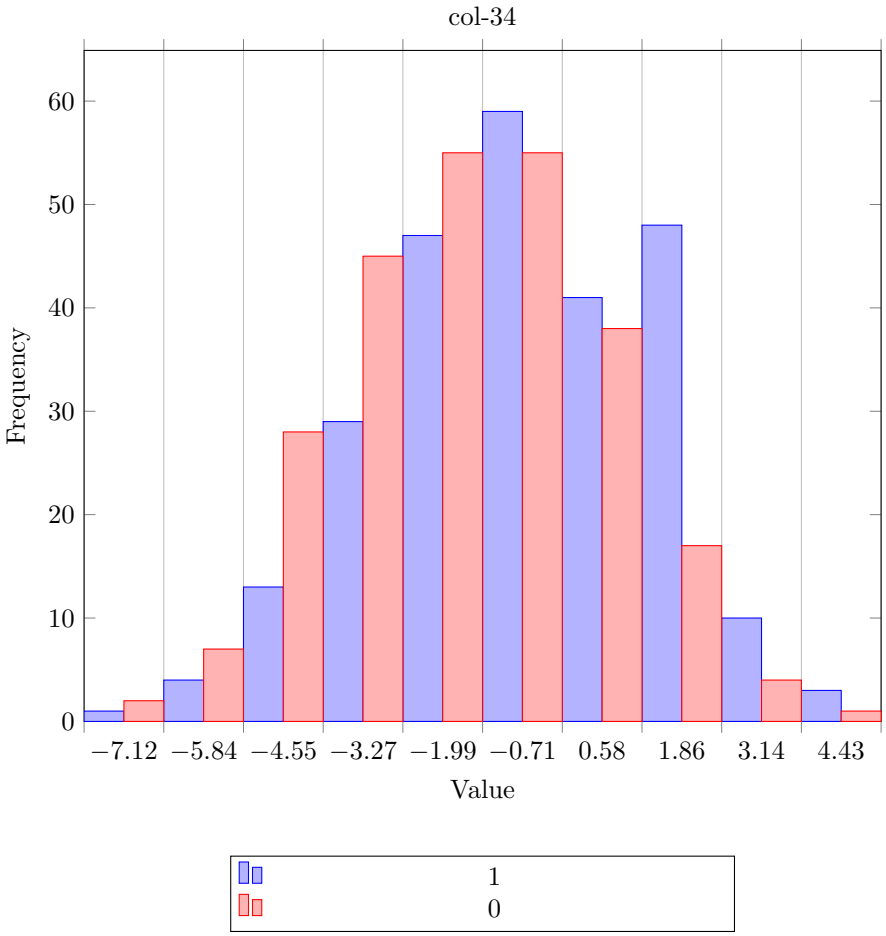

col-29
Frequency
0
10
20
30
40
50
60
70
80
−7.11
−5.52
−3.93
−2.33
−0.74
0.85
2.44
4.03
5.63
7.22
Value
1
0

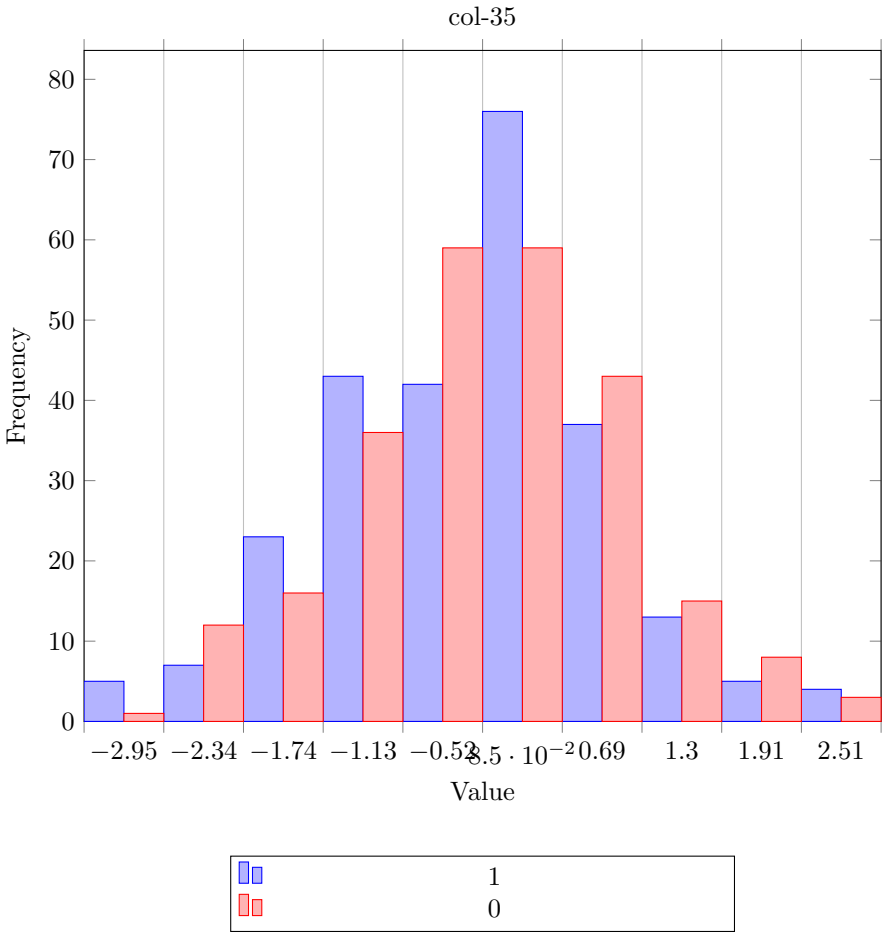

col-30
Frequency
0
10
20
30
40
50
60
−3.38
−2.78
−2.17
−1.57
−0.97
−0.36
0.24
0.84
1.44
2.05
Value
1
0

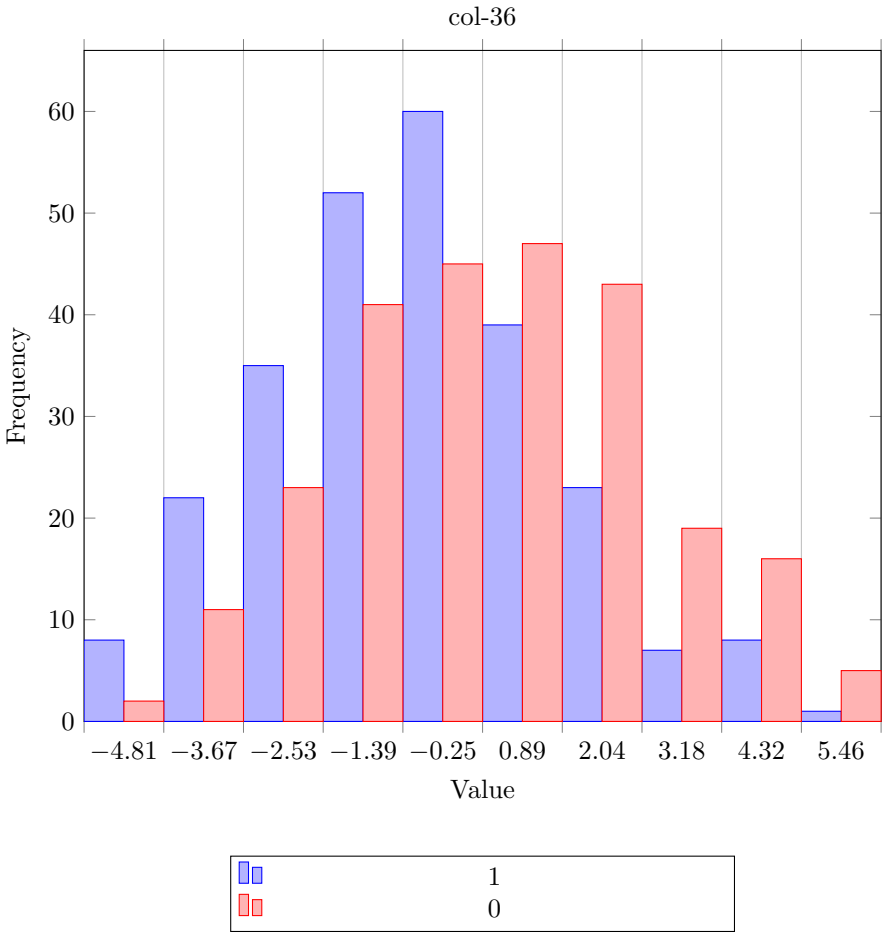

col-31
Frequency
70
60
50
40
30
20
10
0
−2.49
−1.87
−1.25
−0.64
1.8 · 10−20.6
1.22
1.84
2.45
3.07
Value
1
0

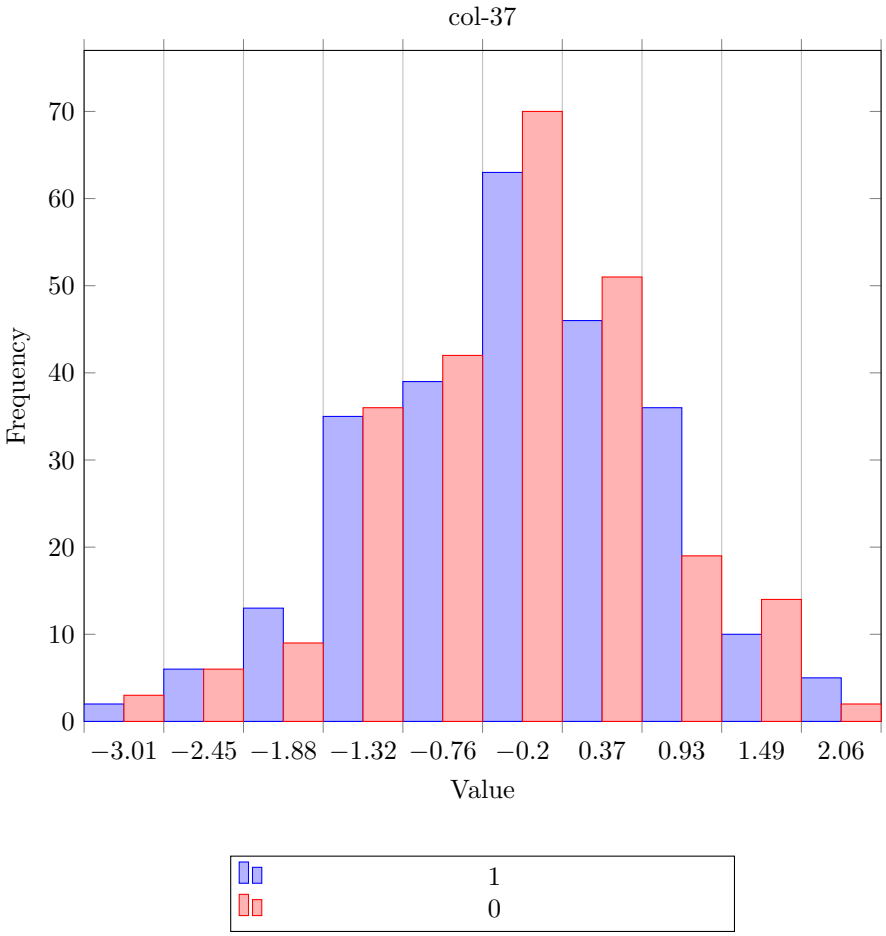

col-32
Frequency
80
70
60
50
40
30
20
10
0
−7.84
−6.46
−5.08
−3.69
−2.31
−0.93
0.45
1.83
3.22
4.6
Value
1
0

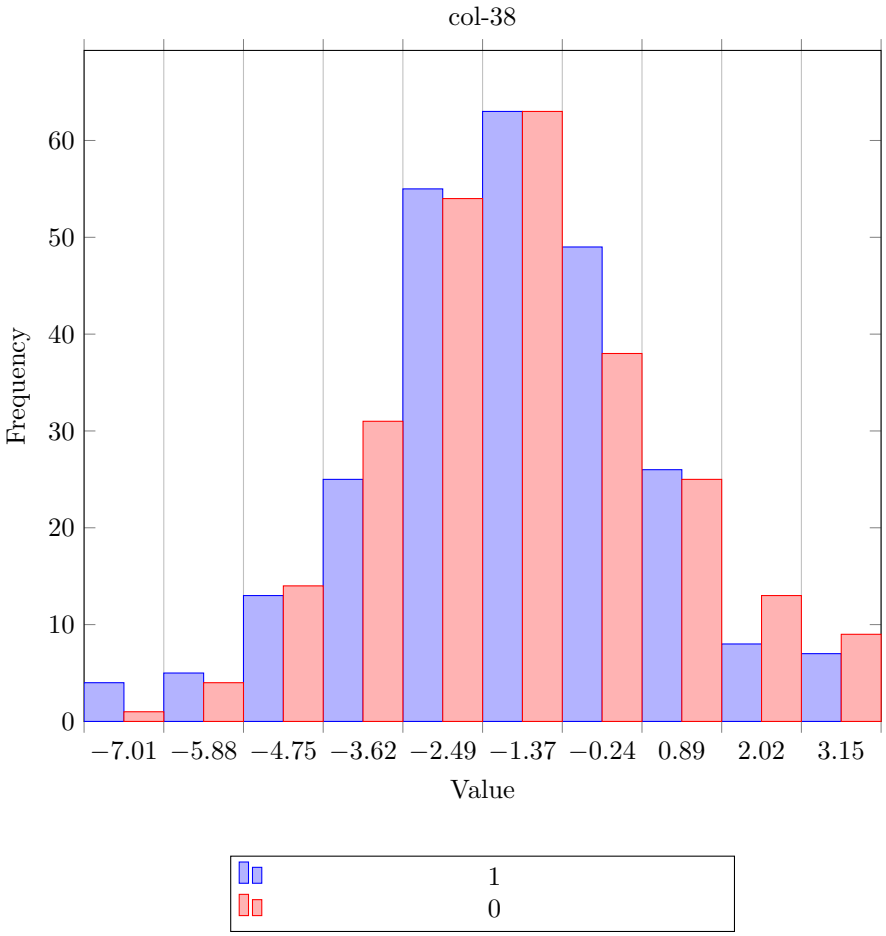

col-33
Frequency
60
50
40
30
20
10
0
−2.44
−1.88
−1.32
−0.76
−0.2
0.36
0.92
1.48
2.04
2.6
Value
1
0

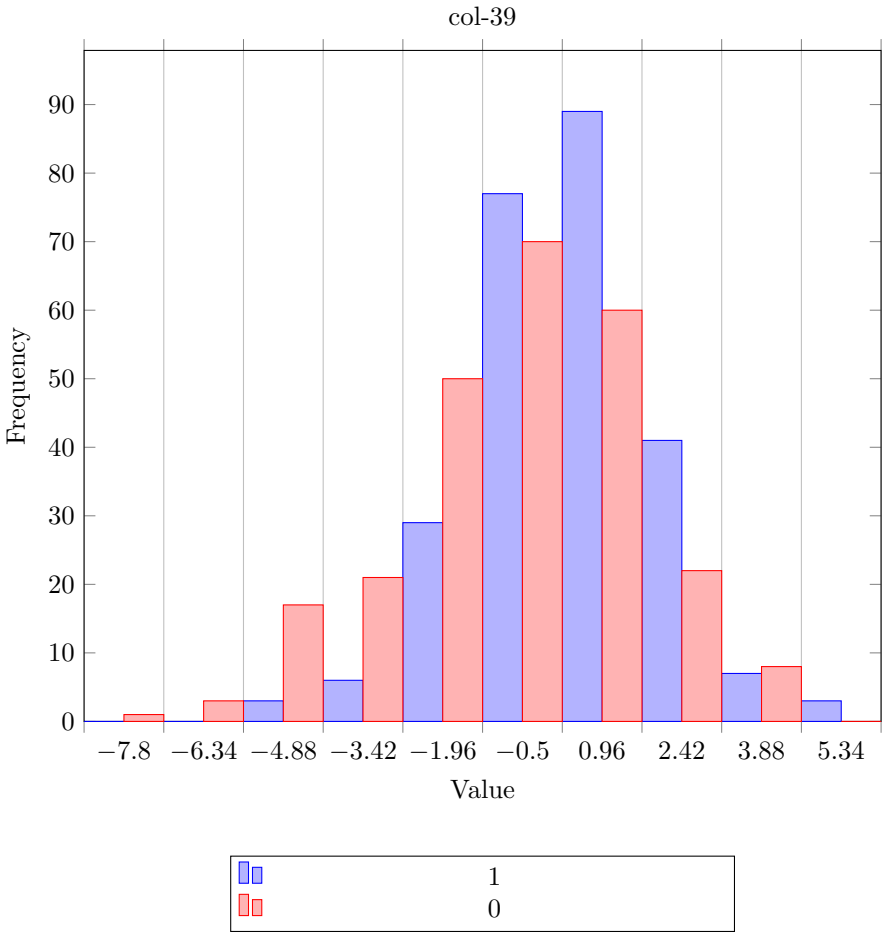

col-34
Frequency
60
50
40
30
20
10
0
−7.12
−5.84
−4.55
−3.27
−1.99
−0.71
0.58
1.86
3.14
4.43
Value
1
0


col-35
Frequency
0
10
20
30
40
50
60
70
80
−2.95
−2.34
−1.74
−1.13
−0.52
8.5 · 10−2
0.69
1.3
1.91
2.51
Value
1
0

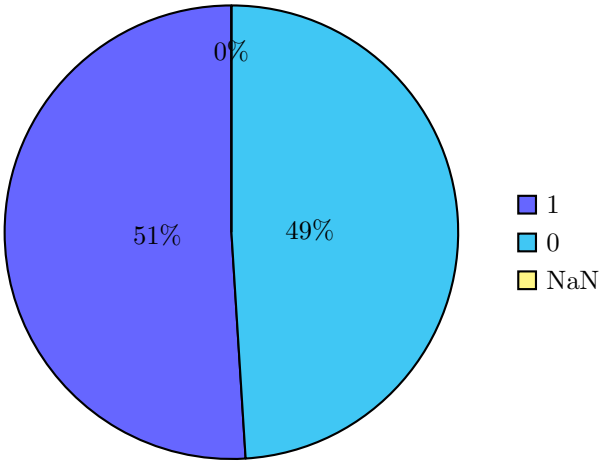

col-36
Frequency
0
10
20
30
40
50
60
−4.81
−3.67
−2.53
−1.39
−0.25
0.89
2.04
3.18
4.32
5.46
Value
1
0


col-37
Frequency
0
10
20
30
40
50
60
70
−3.01
−2.45
−1.88
−1.32
−0.76
−0.2
0.37
0.93
1.49
2.06
Value
1
0


col-38
Frequency
0
10
20
30
40
50
60
−7.01
−5.88
−4.75
−3.62
−2.49
−1.37
−0.24
0.89
2.02
3.15
Value
1
0


col-39
Frequency
0
10
20
30
40
50
60
70
80
90
−7.8
−6.34
−4.88
−3.42
−1.96
−0.5
0.96
2.42
3.88
5.34
Value
1
0

### 1.3 Categorical Columns




**col-40**

Table 1: Dataset Columns Overview

| Column Name | Data Type [1] | NULL |
|---|---|---|
| col-0 | float | 0 |
| col-1 | int | 0 |
| col-2 | int | 0 |
| col-3 | int | 0 |
| col-4 | int | 0 |
| col-5 | float | 0 |
| col-6 | int | 0 |
| col-7 | int | 0 |
| col-8 | float | 0 |
| col-9 | float | 0 |
| col-10 | int | 0 |
| col-11 | int | 0 |
| col-12 | float | 0 |
| col-13 | int | 479 |
| col-14 | int | 493 |
| col-15 | float | 493 |
| col-16 | float | 493 |
| col-17 | float | 493 |
| col-18 | int | 493 |
| col-19 | int | 493 |
| col-20 | int | 493 |
| col-21 | float | 493 |
| col-22 | float | 493 |
| col-23 | int | 493 |
| col-24 | float | 493 |
| col-25 | float | 493 |
| col-26 | float | 493 |
| col-27 | float | 493 |
| col-28 | int | 493 |
| col-29 | int | 493 |
| col-30 | float | 493 |
| col-31 | float | 493 |
| col-32 | int | 493 |
| col-33 | float | 493 |
| col-34 | int | 493 |
| col-35 | float | 493 |
| col-36 | int | 493 |
| col-37 | int | 493 |
| col-38 | int | 493 |
| col-39 | float | 493 |
| col-40 | int | 0 |

*Note:* Blue represents continuous variables, Red represents categorical variables.

| Column Name | Mean | Median | Min | Max | Variance | Std | Outliers |
|---|---|---|---|---|---|---|---|
| col-0 | 0.0256 | 0.0279 | -3.37 | 3.33 | 1.02 | 1.01 | 4 |
| col-1 | -0.0245 | -0.0332 | -3.49 | 3.58 | 1.03 | 1.02 | 7 |
| col-2 | -0.0241 | 0.00815 | -2.7 | 2.55 | 0.958 | 0.979 | 0 |
| col-3 | -0.00227 | 0.00233 | -3.46 | 3.09 | 0.941 | 0.97 | 7 |
| col-4 | 1.09 | 0.863 | -16.4 | 17.6 | 20.6 | 4.54 | 15 |
| col-5 | -0.00625 | 0.027 | -3.04 | 3.1 | 0.977 | 0.989 | 3 |
| col-6 | 0.497 | 0.582 | -7.22 | 7.59 | 4.48 | 2.12 | 7 |
| col-7 | -0.0379 | 0.0188 | -6.51 | 7.13 | 4.98 | 2.23 | 8 |
| col-8 | 0.0264 | 0.0221 | -3.15 | 3.15 | 1 | 1 | 2 |
| col-9 | -0.0036 | -0.0361 | -2.75 | 3.92 | 1.03 | 1.01 | 6 |
| col-10 | -0.0162 | 0.0195 | -3.3 | 3.41 | 1.05 | 1.03 | 6 |
| col-11 | -0.038 | -0.076 | -3.16 | 3.25 | 0.971 | 0.985 | 5 |
| col-12 | -1.99 | -2.1 | -14.7 | 12.2 | 20.5 | 4.53 | 4 |
| col-13 | 0.0178 | 0.0172 | -2.89 | 3.74 | 1 | 1 | 3 |
| col-14 | 0.00846 | -0.0454 | -6.79 | 5.91 | 4.48 | 2.12 | 4 |
| col-15 | -0.0159 | -0.0145 | -2.91 | 3.1 | 1.07 | 1.04 | 4 |
| col-16 | 0.0601 | 0.0787 | -3.46 | 2.64 | 0.947 | 0.973 | 2 |
| col-17 | 0.0105 | 0.0555 | -2.94 | 3.29 | 1.01 | 1.01 | 10 |
| col-18 | 0.344 | 0.265 | -8.26 | 5.72 | 4.71 | 2.17 | 7 |
| col-19 | 0.0384 | 0.0877 | -2.83 | 2.69 | 0.957 | 0.978 | 5 |
| col-20 | 0.0569 | 0.0696 | -4.25 | 2.82 | 1.04 | 1.02 | 2 |
| col-21 | 0.0386 | 0.0706 | -2.82 | 3.04 | 0.97 | 0.985 | 3 |
| col-22 | 0.942 | 1.02 | -6.34 | 7.24 | 4.21 | 2.05 | 12 |
| col-23 | 0.781 | 0.86 | -12.2 | 14.4 | 18 | 4.24 | 9 |
| col-24 | -0.0503 | -0.0542 | -3.22 | 2.98 | 1.07 | 1.03 | 4 |
| col-25 | 0.0203 | -0.00423 | -2.82 | 3.65 | 1.05 | 1.02 | 2 |
| col-26 | -0.0309 | -0.0291 | -3.02 | 2.52 | 1.03 | 1.02 | 4 |
| col-27 | -0.0416 | -0.000326 | -3.02 | 3.07 | 1.01 | 1.01 | 4 |
| col-28 | -0.39 | -0.496 | -8.03 | 7.26 | 5.29 | 2.3 | 7 |
| col-29 | 0.135 | 0.0306 | -7.11 | 8.81 | 5.7 | 2.39 | 8 |
| col-30 | 0.0159 | 0.0348 | -3.38 | 2.65 | 1 | 1 | 1 |
| col-31 | 0.0755 | 0.106 | -2.49 | 3.69 | 0.93 | 0.965 | 2 |
| col-32 | -0.53 | -0.55 | -7.84 | 5.98 | 4.87 | 2.21 | 3 |
| col-33 | 0.0235 | 0.000173 | -2.44 | 3.16 | 0.971 | 0.986 | 1 |
| col-34 | -0.478 | -0.491 | -7.12 | 5.71 | 4.62 | 2.15 | 3 |
| col-35 | 0.0765 | 0.138 | -2.95 | 3.12 | 1.03 | 1.02 | 8 |
| col-36 | 0.469 | 0.361 | -4.81 | 6.6 | 4.84 | 2.2 | 2 |
| col-37 | 0.0251 | 0.0913 | -3.01 | 2.62 | 0.927 | 0.963 | 5 |
| col-38 | -0.887 | -0.84 | -7.01 | 4.28 | 3.92 | 1.98 | 8 |
| col-39 | 0.564 | 0.757 | -7.8 | 6.8 | 4.13 | 2.03 | 16 |

Table 2: Data types and null value counts for each column.