

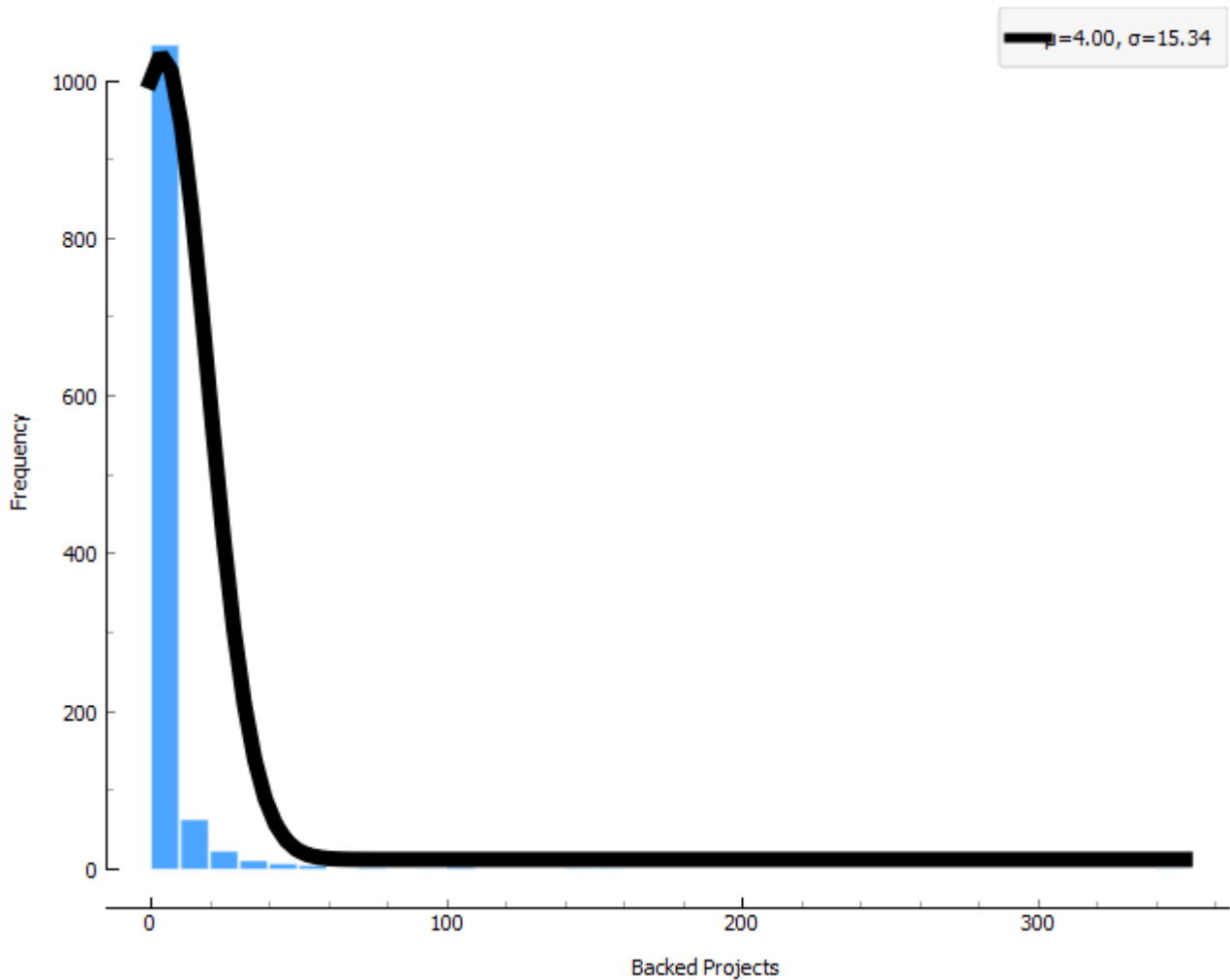
Constructed features

Backed_Projects_log: log(Backed_Projects +1) (numeric)
Goal_log : log(Goal +1) (numeric)
Pledge_Levels_log: log(Pledge_Levels+1) (numeric)
Min_Pledge_Tiers_log: log(Min_Pledge_Tiers+1) (numeric)
Max_Pledge_Tiers_log: log(Max_Pledge_Tiers+1) (numeric)

We have five variables with extreme values, I checked the websites of these projects on kickstarter and made sure that these values really could be and they are correct, so I could not delete them and used the log function.

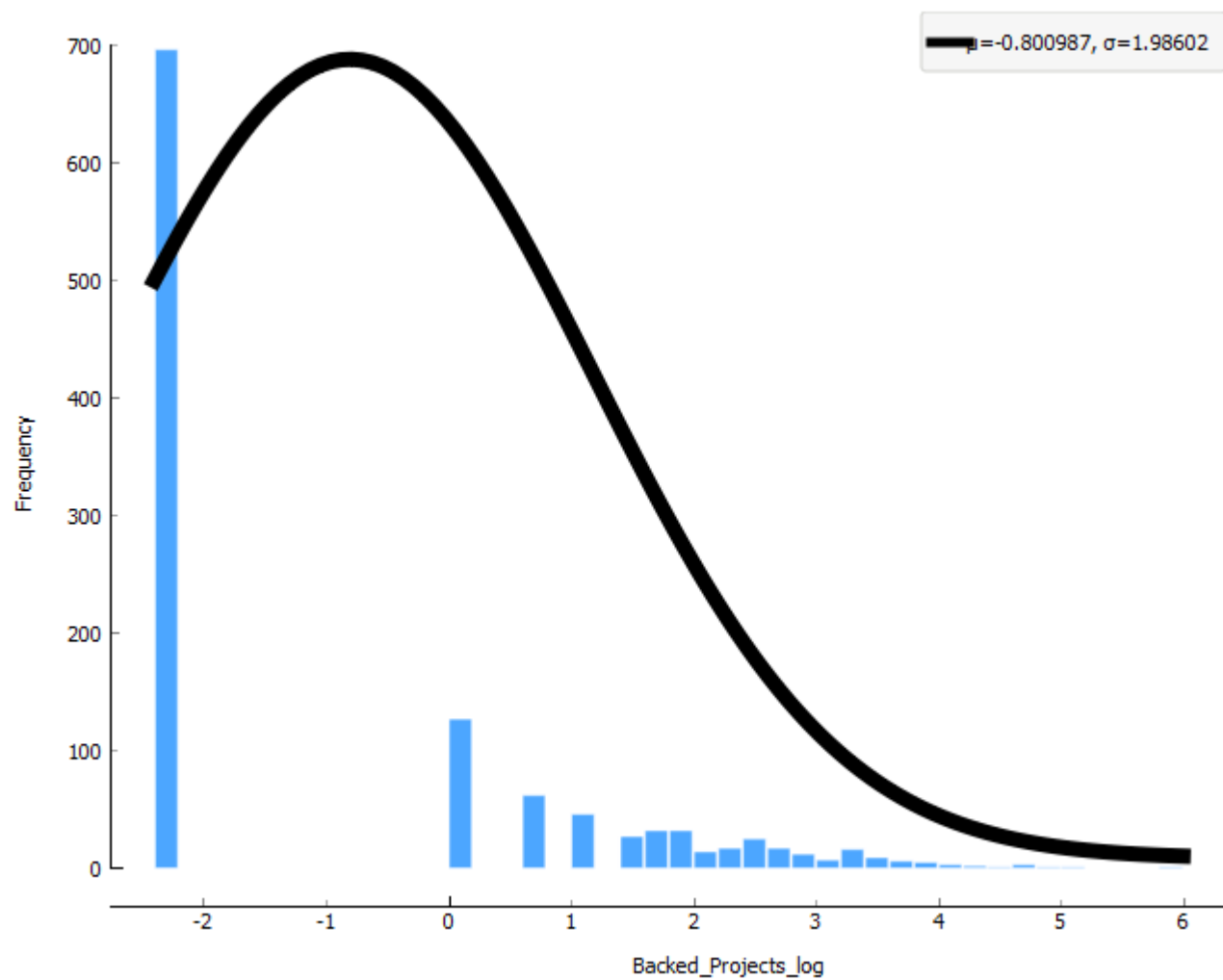
Distributions

Fri Oct 02 20, 23:34:39



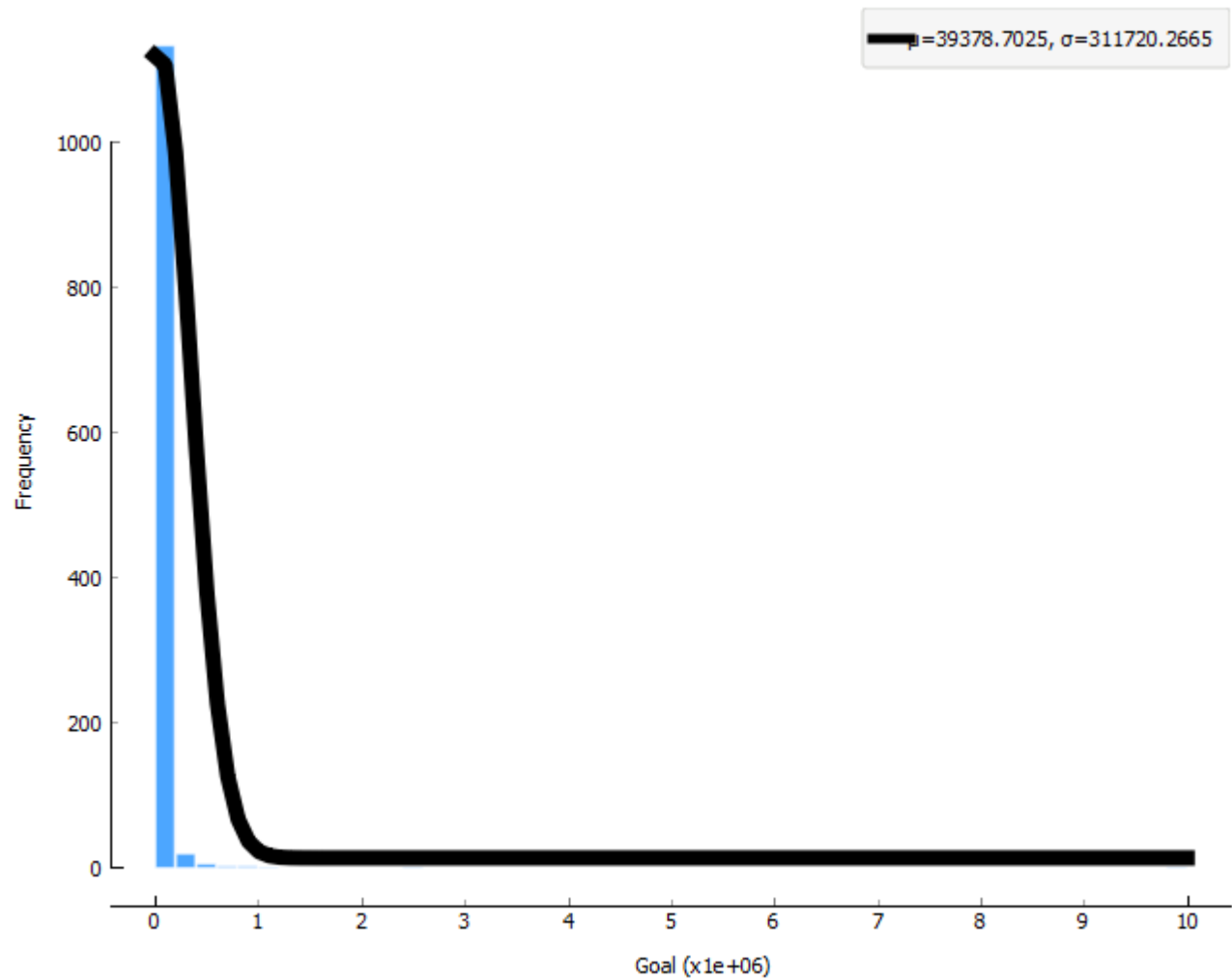
Distribution of 'Backed Projects'

before



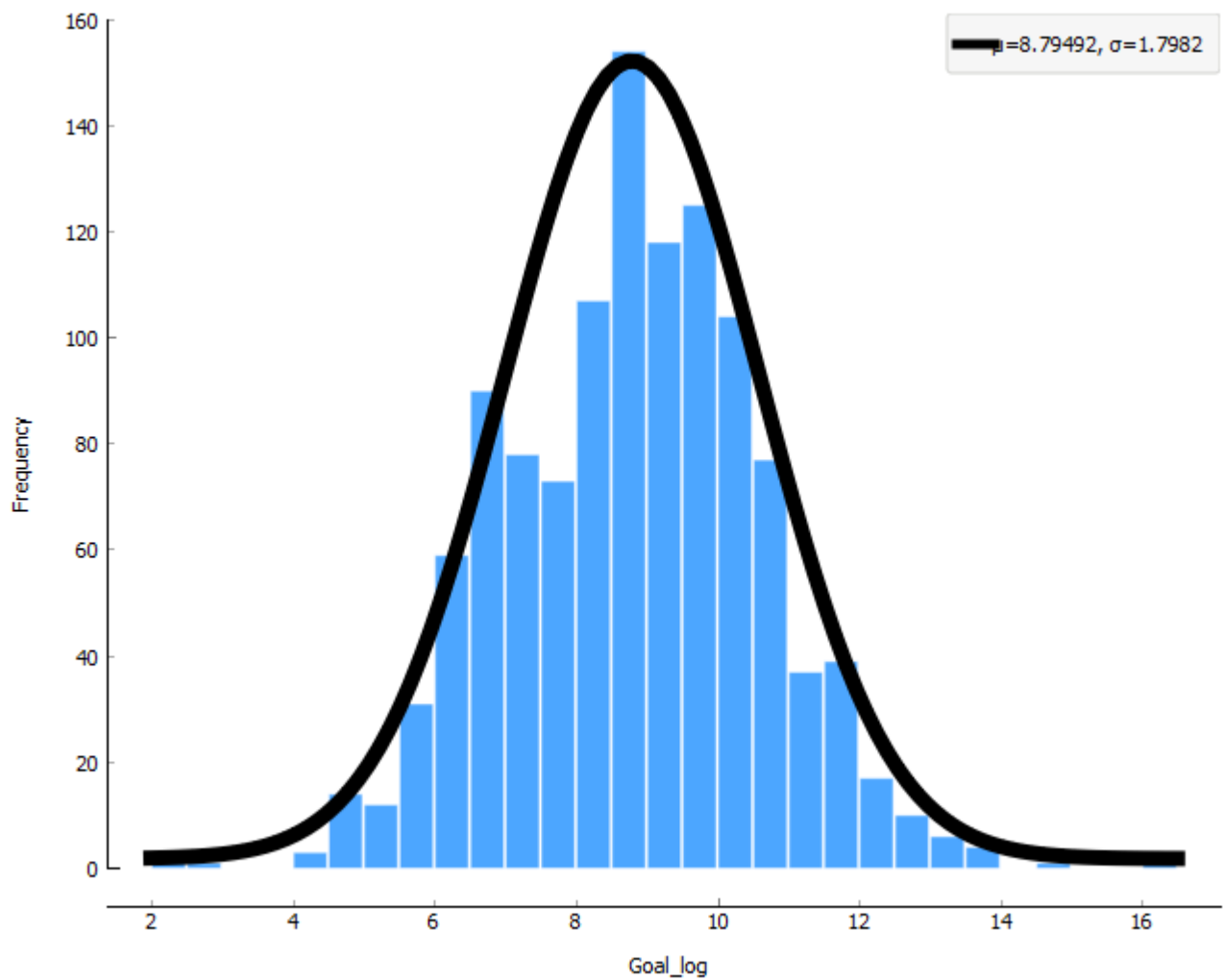
Distribution of 'Backed_Projects_log'

after.
it still not normal distribution, but seems better.



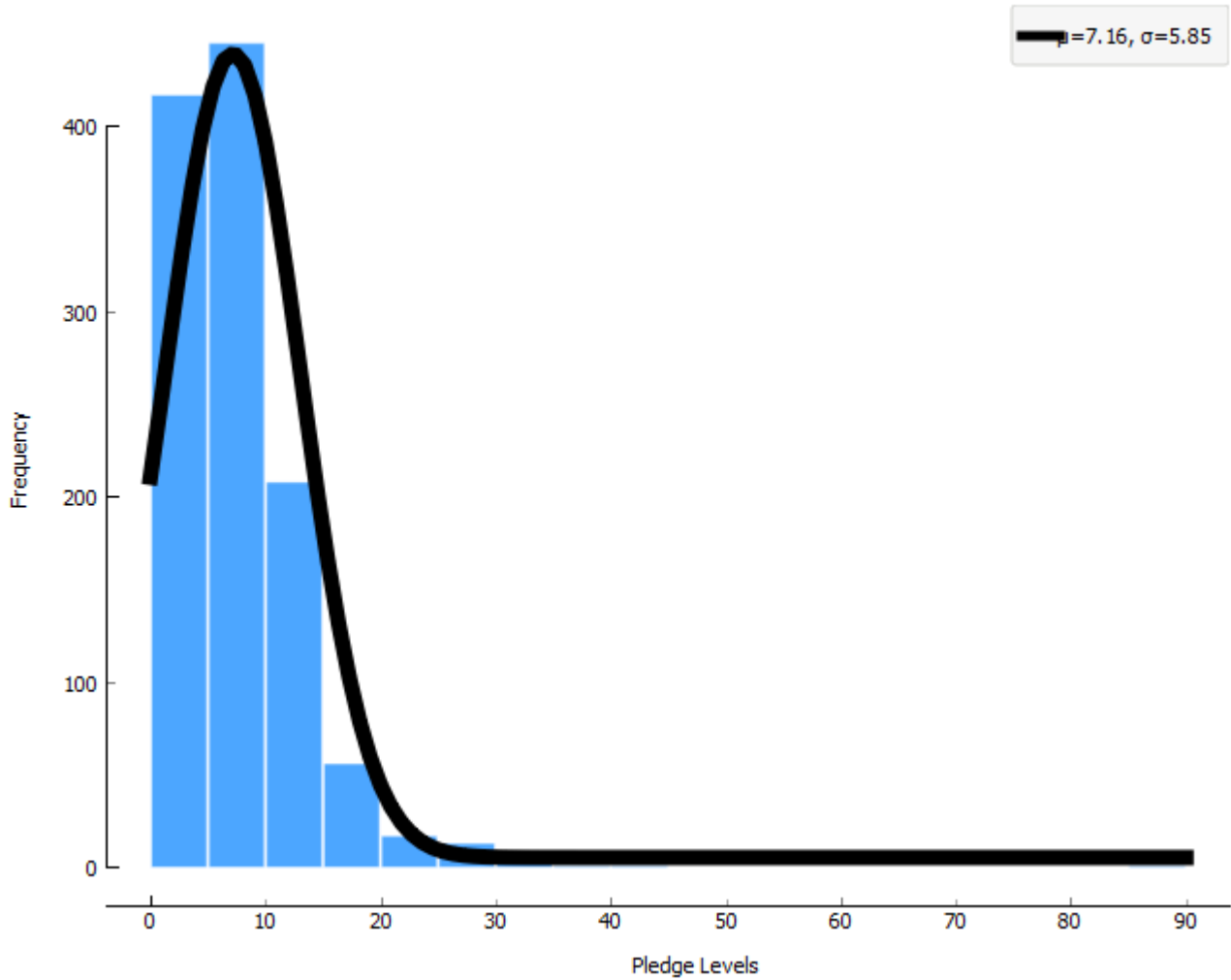
Distribution of 'Goal'

before



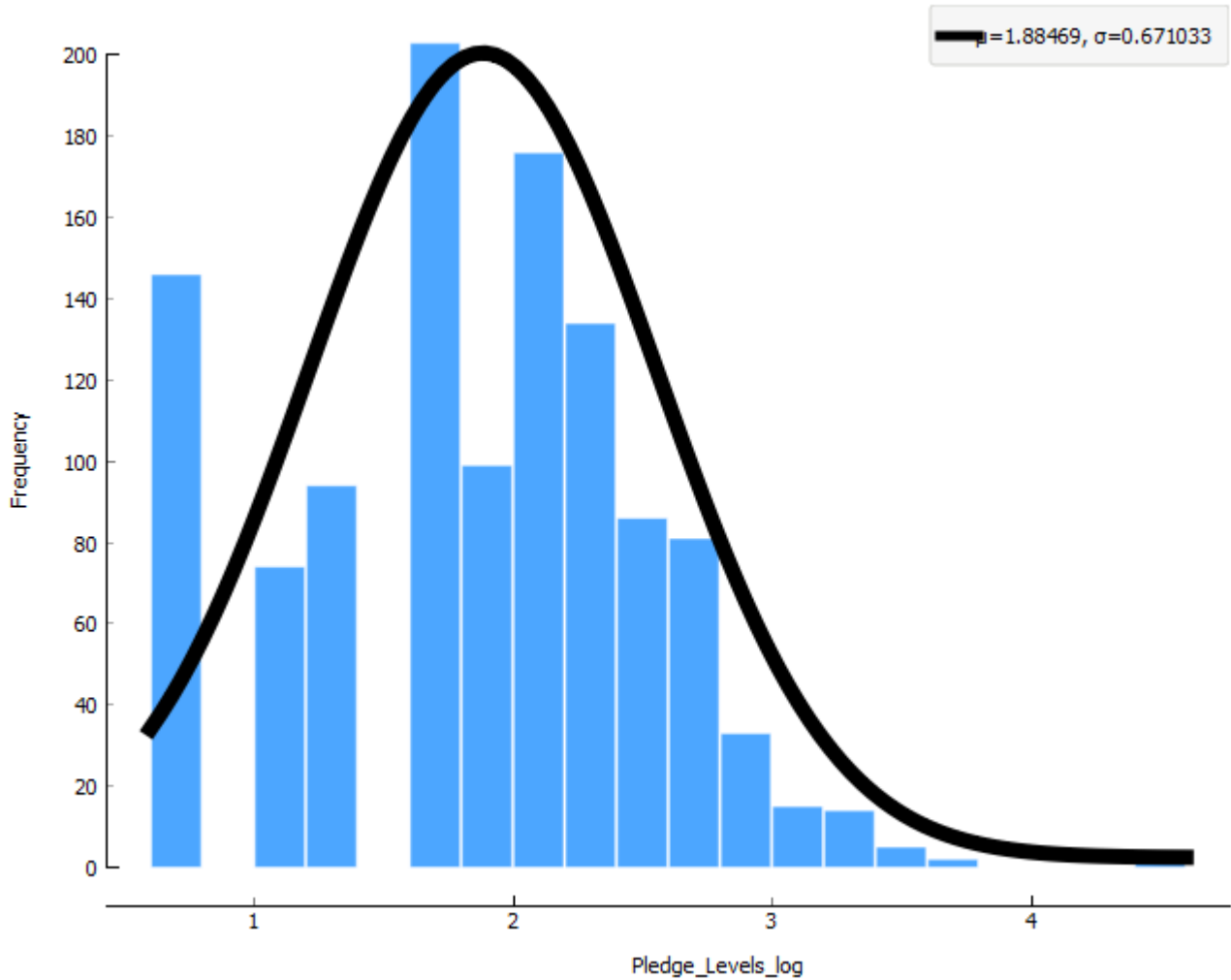
Distribution of 'Goal_log '

after.
we have almost normal distribution



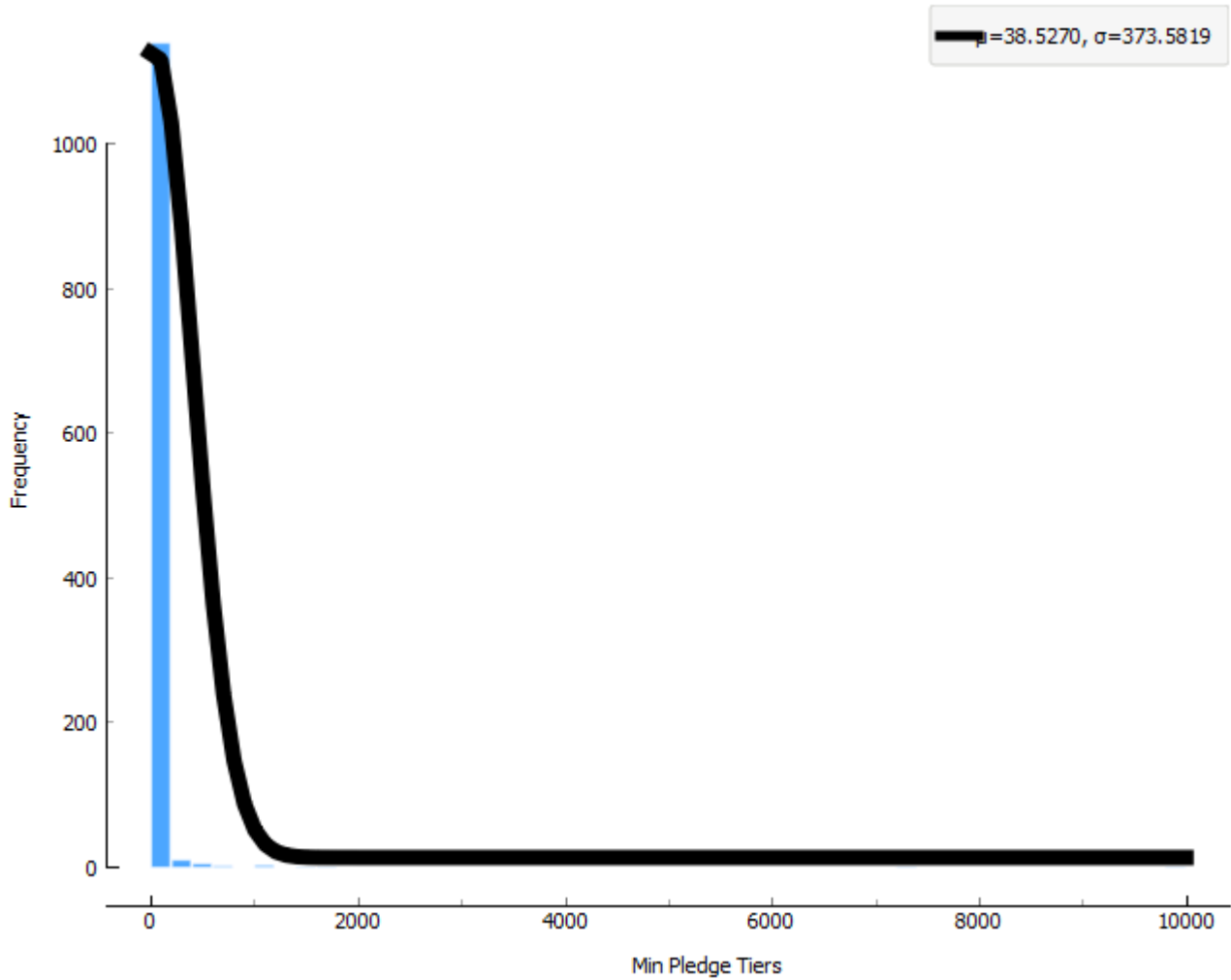
Distribution of 'Pledge Levels'

before



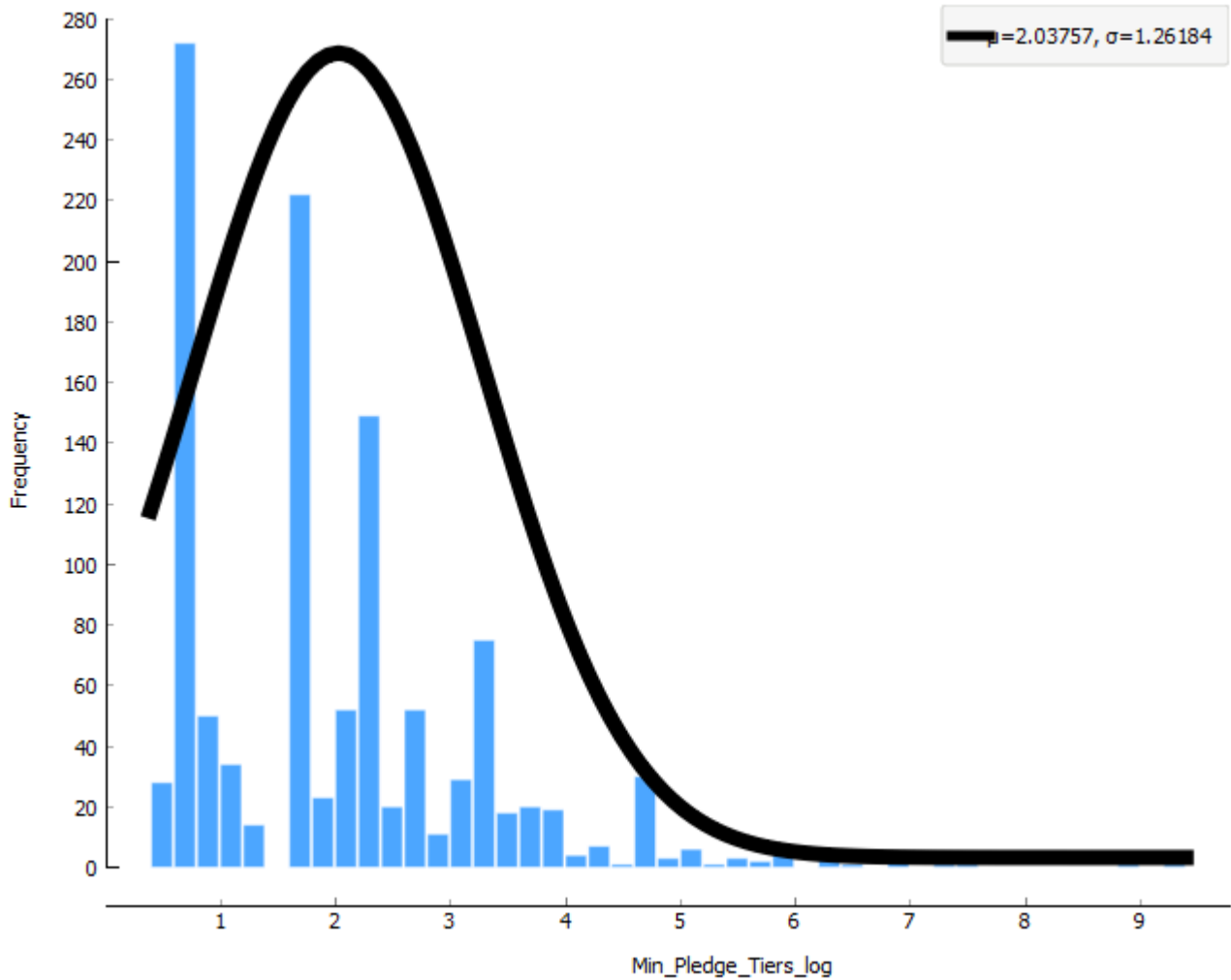
Distribution of 'Pledge_Levels_log'

after, much better



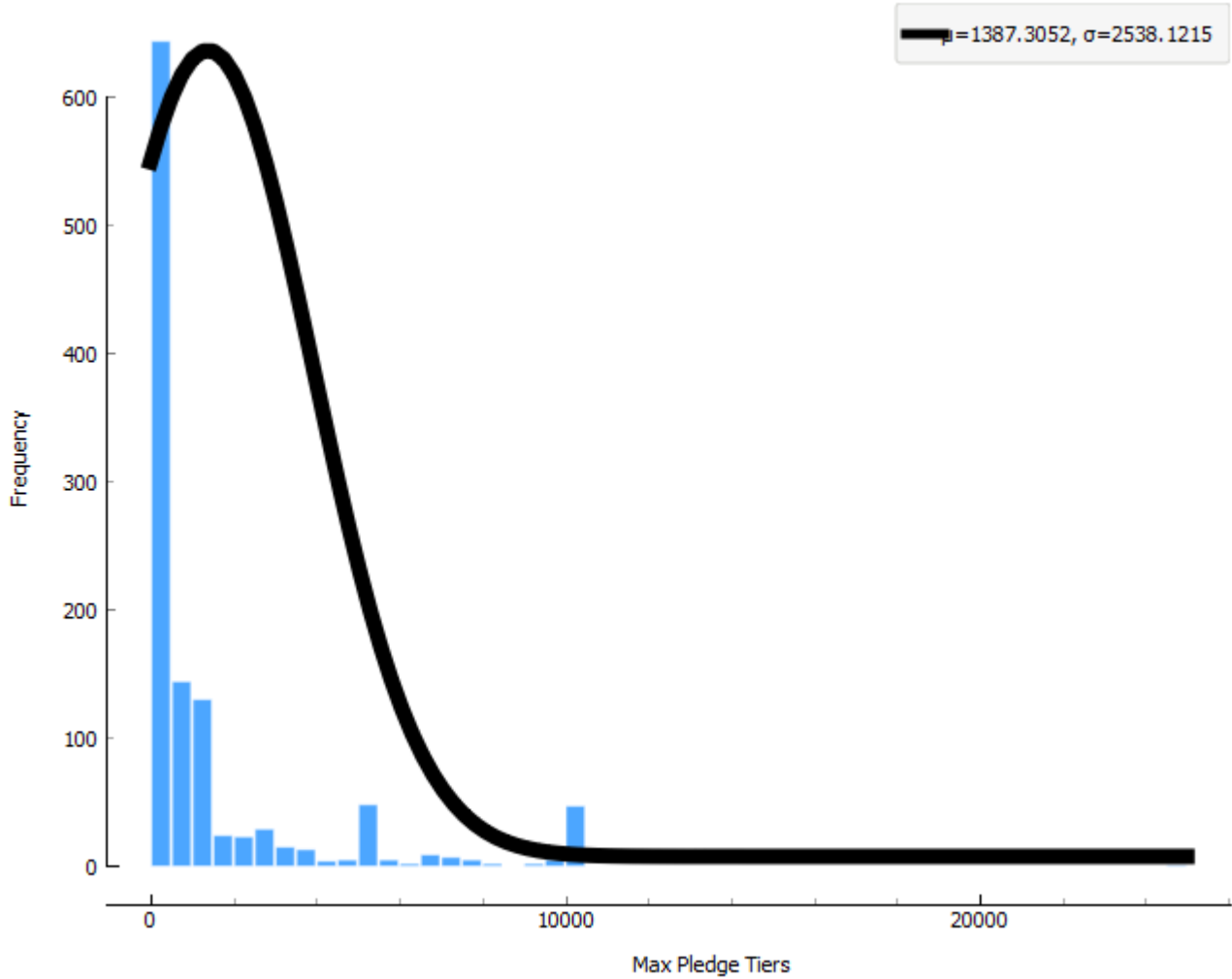
Distribution of 'Min Pledge Tiers'

before



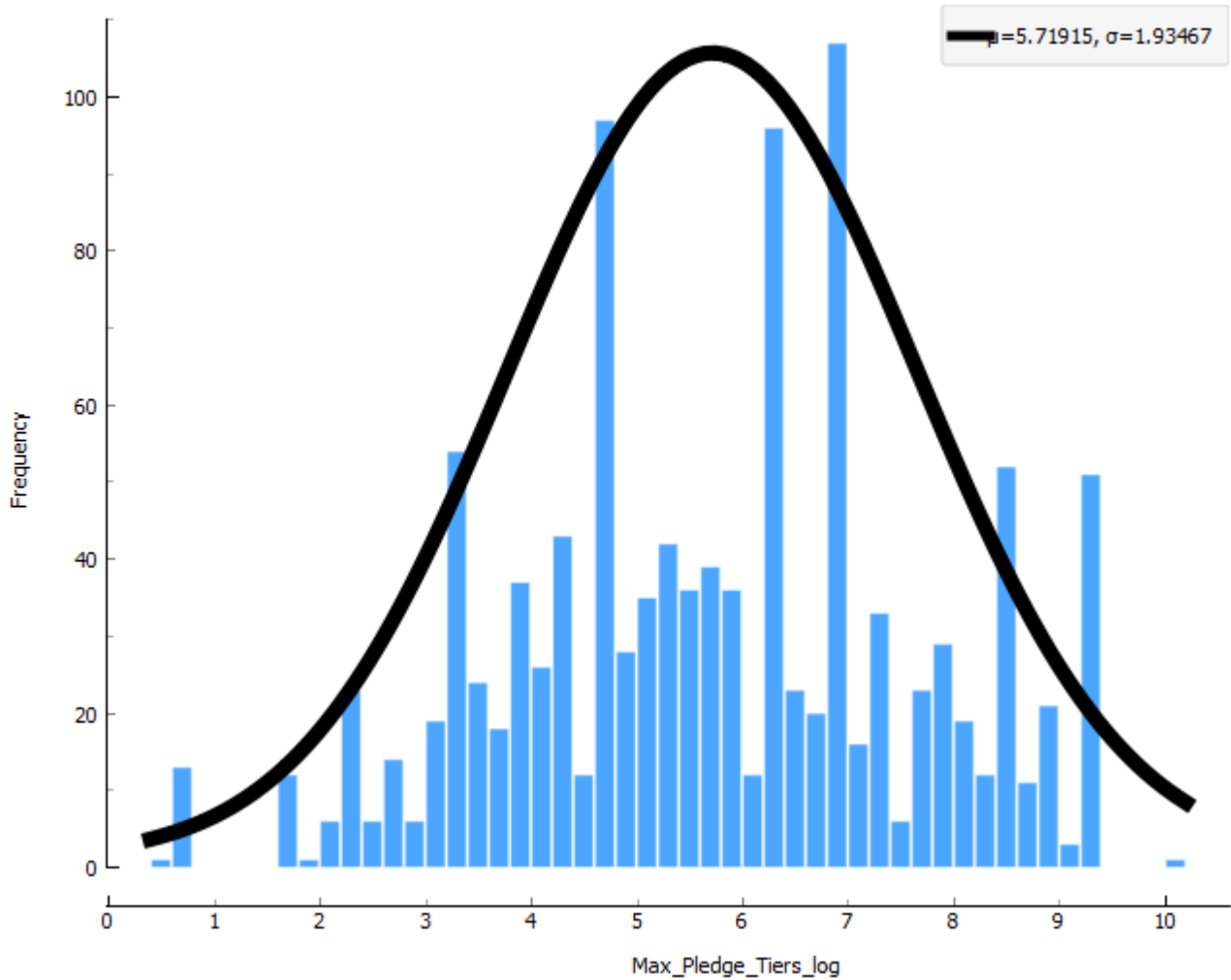
Distribution of 'Min_Pledge_Tiers_log'

after, also looks better



Distribution of 'Max Pledge Tiers'

before



Distribution of 'Max_Pledge_Tiers_log'

after, looks way better.

Settings

- Categorical features:** Divide by number of values
- Numeric features:** Leave them as they are
- Class:** Leave it as it is

transforming categorical variable with 7 values to bin count.


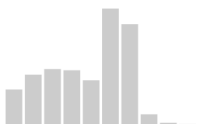


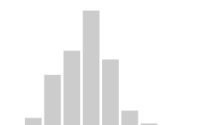




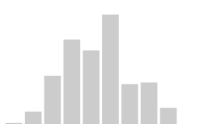

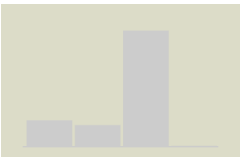
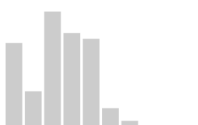
Data instances: 1163
Features: 1
Meta attributes: None

	Type
1	0.166667
2	0.166667
3	0.166667
4	0.166667
5	0.166667
6	0.166667
7	0.166667
8	0.166667
9	0.166667
10	0.166667
11	0.166667
12	0.166667
13	0.166667
14	0.166667
15	0.166667
16	0.166667
17	0.166667
18	0.166667
19	0.166667
20	0.166667
21	0.166667
22	0.166667
23	0.166667
24	0.166667
25	0.166667
26	0.166667
27	0.166667
28	0.166667
29	0.166667
30	0.166667
31	0.166667
32	0.166667
33	0.166667
34	0.166667
35	0.166667
36	0.166667
37	0.166667
38	0.166667
39	0.166667
40	0.166667
41	0.166667
42	0.166667
43	0.166667
44	0.166667
45	0.166667
46	0.166667

47	0.166667
48	0.166667
49	0.166667
50	0.166667
51	0.166667
52	0.166667
53	0.166667
54	0.166667
55	0.166667
56	0.166667
57	0.166667
58	0.166667
59	0.166667
60	0.166667
61	0.166667
62	0.166667
63	0.166667
64	0.166667
65	0.166667
66	0.166667
67	0.166667
68	0.166667
69	0.166667
70	0.166667
71	0.166667
72	0.166667
73	0.166667
74	0.166667
75	0.166667
76	0.166667
77	0.166667
78	0.166667
79	0.166667
80	0.166667
81	0.166667
82	0.166667
83	0.166667
84	0.166667
85	0.166667
86	0.166667
87	0.166667
88	0.166667
89	0.166667
90	0.166667
91	0.166667
92	0.166667
93	0.166667
94	0.166667
95	0.166667
96	0.166667
97	0.166667
98	0.166667
99	0.166667

100 0.166667
+ 1063 more

each value gets a number, same values gets the same number.

Name		Distribution	Center	Dispersion	Min.	Max.	Missing
N	Backed_Projects_log		-0.800987	-2.47946	-2.30259	5.85249	0 (0%)
	Creator Desc Len		360.90	0.46	0	757	0 (0%)
	Duration		33.66	0.34	4	62	0 (0%)
C	Funded		no	0.677			0 (0%)
N	Goal_log		8.79492	0.204458	2.3979	16.1181	0 (0%)
	Has FB=0		0.484953	1.03056	0	1	0 (0%)
	Has FB=1		0.515047	0.970345	0	1	0 (0%)
N	Has Video		0.66	0.73	0	1	0 (0%)
	Images		7.87	1.56	0	95	0 (0%)
	Max_Pledge_Tiers_log		5.71915	0.33828	0.565314	10.1267	0 (0%)
N	Min_Pledge_Tiers_log		2.03757	0.619284	0.530628	9.21044	0 (0%)
C	Month		Mar	0.822			0 (0%)
N	Pledge_Levels_log		1.88469	0.356044	0.693147	4.49981	0 (0%)
	Previous Projects		0.82	2.95	0	29	0 (0%)

N

Proj Desc Len

4095.01

0.91

59

28293

0 (0%)

C	Selected		No	0			0 (0%)
C	Selected (1)		No	0			0 (0%)
C	Selected (2)		No	0			0 (0%)

N

Title Len

36.73

0.49

1

82

0 (0%)

N

Type

0.522499

0.670182

0

1

0 (0%)

N

Videos

0.82

1.19

0

13

0 (0%)

in the end of clinsing

after we treated exteme values, we don't have mising values,

clinsing step is over.