`streamlit_pandas_profiling`

`streamlit_pandas_profiling`

Overview

Dataset statistics

Number of variables	9
Number of observations	768
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	54.1 KiB
Average record size in memory	72.2 B

Variable types

Numeric	8
Categorical	1

Alerts

`Pregnancies` is highly overall correlated with `Age`	High correlation
`SkinThickness` is highly overall correlated with `Insulin`	High correlation
`Insulin` is highly overall correlated with `SkinThickness`	High correlation
`Age` is highly overall correlated with `Pregnancies`	High correlation
`Pregnancies` has 111 (14.5%) zeros	Zeros
`BloodPressure` has 35 (4.6%) zeros	Zeros
`SkinThickness` has 227 (29.6%) zeros	Zeros
`Insulin` has 374 (48.7%) zeros	Zeros
`BMI` has 11 (1.4%) zeros	Zeros

Reproduction

Analysis started	2023-08-26 12:44:07.569335
Analysis finished	2023-08-26 12:44:22.840344
Duration	15.27 seconds

|	
Duration	15.27 seconds
Software version	pandas-profiling v3.6.6 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Pandas%20Profiling%20Report%22%2C%20%22d

Variables

Select Columns

Pregnancies
Real number (R)

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Age)
ZEROS

Distinct	17
Distinct (%)	2.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.8450521

Minimum	0
Maximum	17
Zeros	111
Zeros (%)	14.5%
Negative	0
Negative (%)	0.0%
Memory size	6.1 KiB
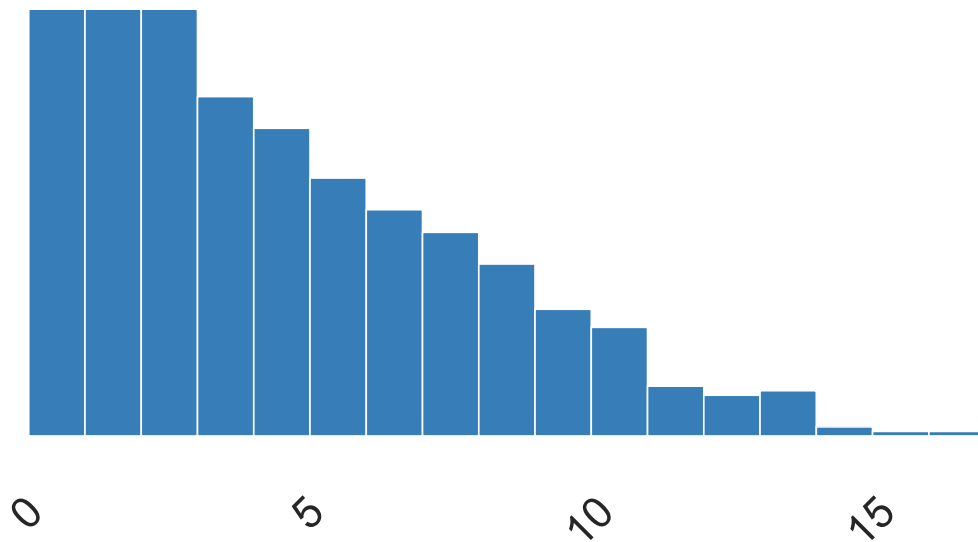
Quantile statistics

Minimum	0
5-th percentile	0
Q1	1
median	3
Q3	6
95-th percentile	10
Maximum	17
Range	17
Interquartile range (IQR)	5

Descriptive statistics

Standard deviation	3.3695781
Coefficient of variation (CV)	0.87634133
Kurtosis	0.15921978
Mean	3.8450521
Median Absolute Deviation (MAD)	2
Skewness	0.90167398
Sum	2953
Variance	11.354056
Monotonicity	Not monotonic

140
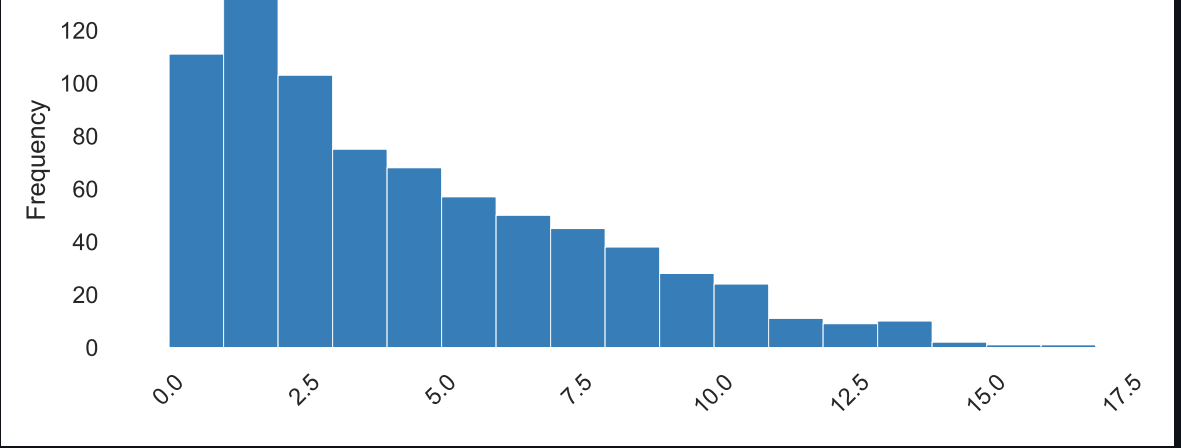
Histogram with fixed size bins (bins=17)

Value	Count	Frequency (%
1	135	17.6%
0	111	14.5%
2	103	13.4%
3	75	9.8%
4	68	8.9%
5	57	7.4%
6	50	6.5%
7	45	5.9%
8	38	4.9%
9	28	3.6%
Other values (7)	58	7.6%

Value	Count	Frequency (%
0	111	14.5%
1	135	17.6%
2	103	13.4%
3	75	9.8%
4	68	8.9%
5	57	7.4%
6	50	6.5%

7	45	5.9%
8	38	4.9%
9	28	3.6%

Value	Count	Frequency (%
17	1	0.1%
15	1	0.1%
14	2	0.3%
13	10	1.3%
12	9	1.2%
11	11	1.4%
10	24	3.1%
9	28	3.6%
8	38	4.9%
7	45	5.9%

Glucose
Real number (ℝ)

Distinct	136
Distinct (%)	17.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	120.89453

Minimum	0
Maximum	199

Zeros	5
Zeros (%)	0.7%
Negative	0
Negative (%)	0.0%
Memory size	6.1 KiB
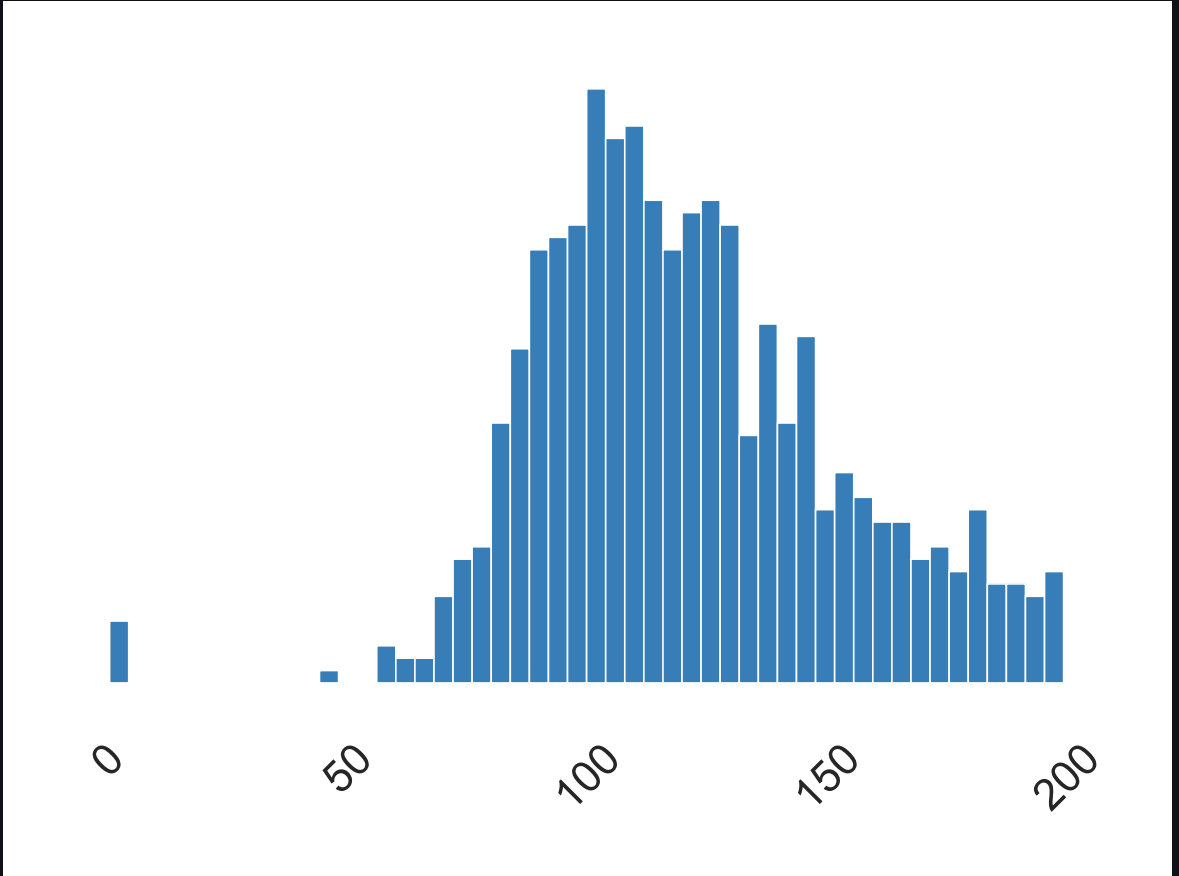
Quantile statistics

Minimum	0
5-th percentile	79
Q1	99
median	117
Q3	140.25
95-th percentile	181
Maximum	199
Range	199
Interquartile range (IQR)	41.25

Descriptive statistics

Standard deviation	31.972618
Coefficient of variation (CV)	0.26446703
Kurtosis	0.64077982
Mean	120.89453
Median Absolute Deviation (MAD)	20
Skewness	0.1737535
Sum	92847
Variance	1022.2483
Monotonicity	Not monotonic
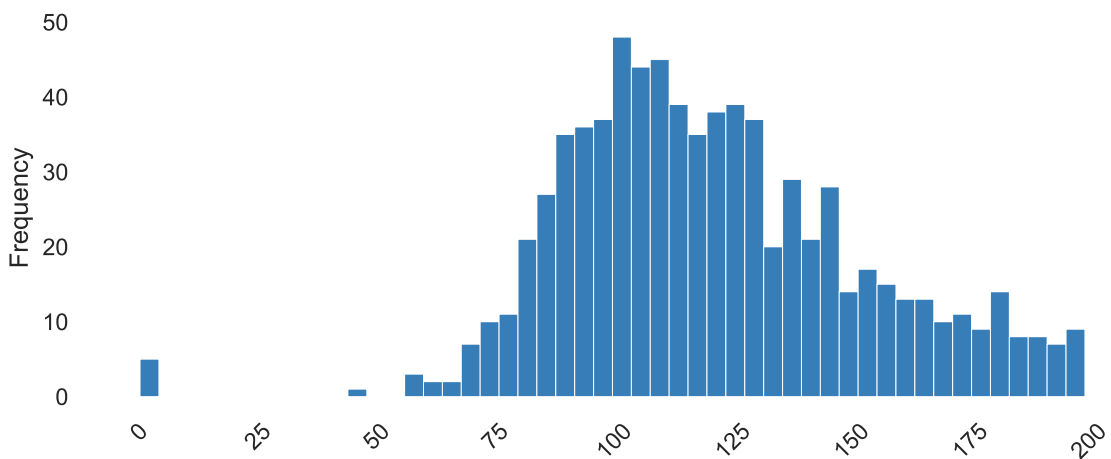
Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%
99	17	2.2%
100	17	2.2%
111	14	1.8%
129	14	1.8%
125	14	1.8%
106	14	1.8%
112	13	1.7%
108	13	1.7%
95	13	1.7%
105	13	1.7%
Other values (126)	626	81.5%

Value	Count	Frequency (%
0	5	0.7%
44	1	0.1%
56	1	0.1%
57	2	0.3%
61	1	0.1%
62	1	0.1%
65	1	0.1%
67	1	0.1%
68	3	0.4%
71	4	0.5%

Value	Count	Frequency (%
199	1	0.1%
198	1	0.1%
197	4	0.5%
196	3	0.4%
195	2	0.3%
194	3	0.4%
193	2	0.3%
191	1	0.1%
190	1	0.1%
189	4	0.5%

BloodPressure

Real number (ℝ)

Distinct	47
Distinct (%)	6.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	69.105469

Minimum	0
Maximum	122
Zeros	35
Zeros (%)	4.6%
Negative	0
Negative (%)	0.0%
Memory size	6.1 KiB
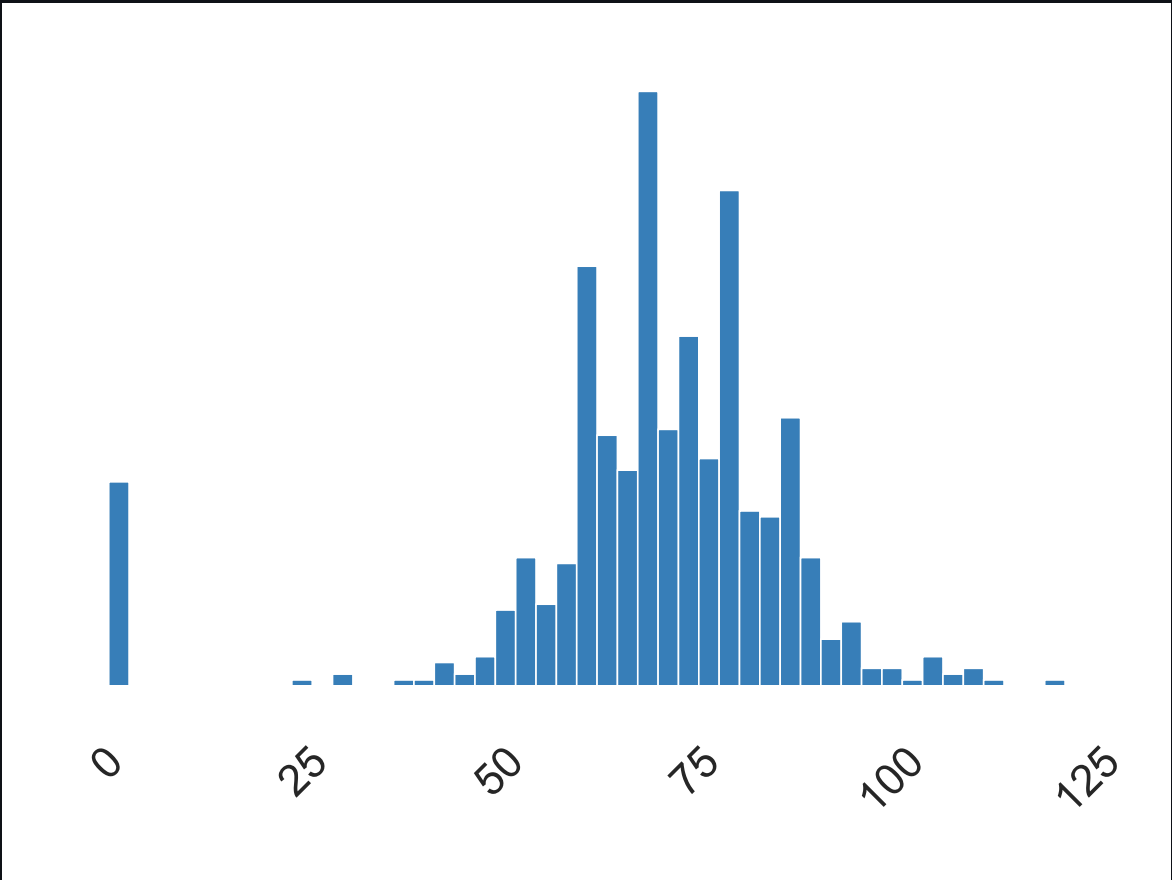
Quantile statistics

Minimum	0

5-th percentile	38.7
Q1	62
median	72
Q3	80
95-th percentile	90
Maximum	122
Range	122
Interquartile range (IQR)	18

Descriptive statistics

Standard deviation	19.355807
Coefficient of variation (CV)	0.28009082
Kurtosis	5.1801566
Mean	69.105469
Median Absolute Deviation (MAD)	8
Skewness	-1.843608
Sum	53073
Variance	374.64727
Monotonicity	Not monotonic
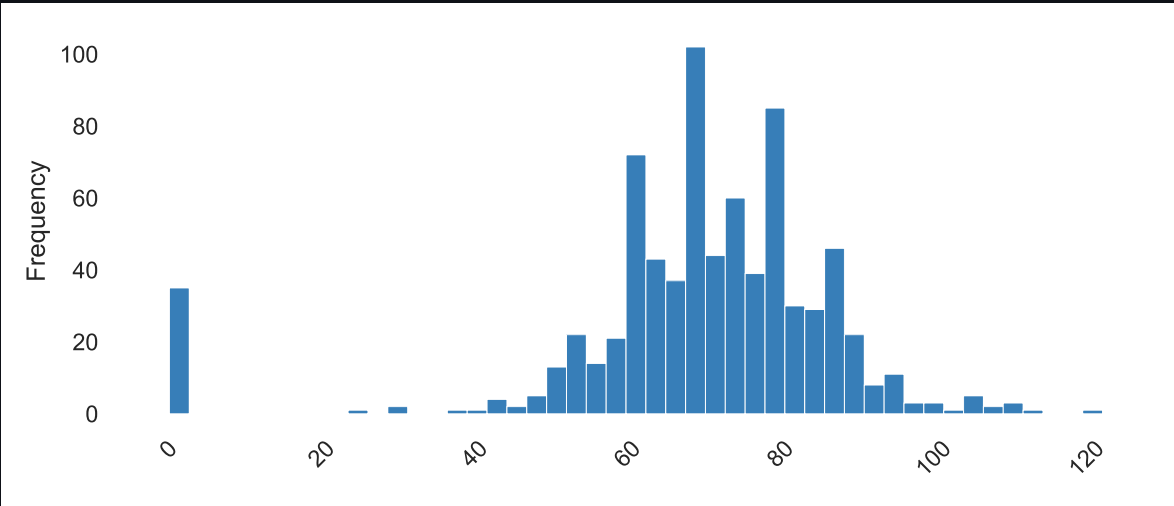
Histogram with fixed size bins (bins=47)

Value	Count	Frequency (%)
70	57	7.4%
74	52	6.8%
78	45	5.9%
68	45	5.9%

Value	Count	Frequency (%)
72	44	5.7%
64	43	5.6%
80	40	5.2%
76	39	5.1%
60	37	4.8%
0	35	4.6%
Other values (37)	331	43.1%

Value	Count	Frequency (%
0	35	4.6%
24	1	0.1%
30	2	0.3%
38	1	0.1%
40	1	0.1%
44	4	0.5%
46	2	0.3%
48	5	0.7%
50	13	1.7%
52	11	1.4%

Value	Count	Frequency (%
122	1	0.1%
114	1	0.1%
110	3	0.4%
108	2	0.3%
106	3	0.4%
104	2	0.3%
102	1	0.1%

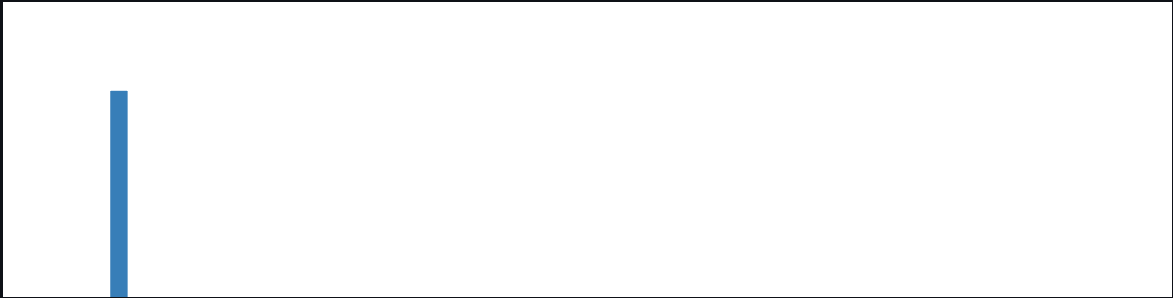|		
100	3	0.4%
98	3	0.4%
96	4	0.5%

SkinThickness
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Insulin) ZEROS

Distinct	51
Distinct (%)	6.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	20.536458

Minimum	0
Maximum	99
Zeros	227
Zeros (%)	29.6%
Negative	0
Negative (%)	0.0%
Memory size	6.1 KiB

https://30days.streamlit.app/?challenge=Day14

Made with Streamlit