Overview

Dataset statistics

Number of variables	19
Number of observations	35784
Missing cells	0
Missing cells (%)	0.0%
Total size in memory	5.2 MiB
Average record size in memory	152.0 B

Variable types

Numeric	6
Categorical	13

Alerts

|	
`AccountState` has a high cardinality: 52 distinct values	**High cardinality**
`AccountZip` has a high cardinality: 3098 distinct values	**High cardinality**
`AccountNaicsName` has a high cardinality: 831 distinct values	**High cardinality**
`ServiceName` has a high cardinality: 134 distinct values	**High cardinality**
`JobName` has a high cardinality: 11876 distinct values	**High cardinality**
`# of Locations` is highly skewed ($\gamma1 = 21.46063329$)	**Skewed**
`ChargeHours` has 8313 (23.2%) zeros	**Zeros**
`ProductionRevActual` has 8313 (23.2%) zeros	**Zeros**
`Cost` has 7890 (22.0%) zeros	**Zeros**
`# of Locations` has 33003 (92.2%) zeros	**Zeros**

Reproduction

Analysis started	2022-04-13 19:22:26.818906
Analysis finished	2022-04-13 19:22:27.049133
Duration	0.23 seconds
Software version	pandas-profiling v3.1.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Clean%20Data%20V2%20Profile%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%

Variables

CalendarYear
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2020.360664
Minimum	2020
Maximum	2022
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	279.7 KiB
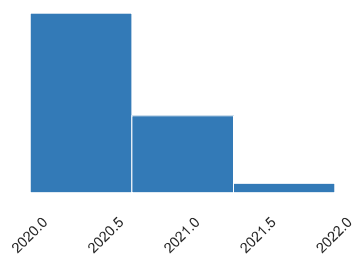
Quantile statistics

Minimum	2020
5-th percentile	2020
Q1	2020
median	2020
Q3	2021
95-th percentile	2021
Maximum	2022
Range	2
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	0.5489042578
Coefficient of variation (CV)	0.0002716862724
Kurtosis	0.4723167213

Mean	2020.360664
Median Absolute Deviation (MAD)	0
Skewness	1.208600377
Sum	72296586
Variance	0.3012958842
Monotonicity	Not monotonic
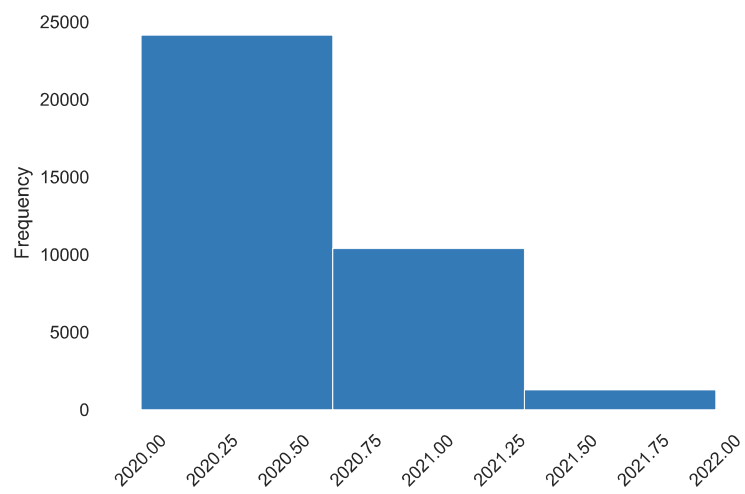
Histogram with fixed size bins (bins=3)

Value	Count	Frequency (%)
2020	24143	67.5%
2021	10376	29.0%
2022	1265	3.5%

Value	Count	Frequency (%)
2020	24143	67.5%
2021	10376	29.0%
2022	1265	3.5%

Value	Count	Frequency (%)
2022	1265	3.5%
2021	10376	29.0%
2020	24143	67.5%

OfficeName
Categorical

Distinct	41
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	2	?
Unique (%)	< 0.1%	

Sample

1st row	Wausau, WI (30)
2nd row	Wausau, WI (30)
3rd row	Wausau, WI (30)
4th row	Wausau, WI (30)
5th row	Wausau, WI (30)

Common Values

Value	Count	Frequency (%)
Chicago-Lincolnshire (57)	3272	9.1%
Milwaukee, WI (34)	2945	8.2%
Green Bay, WI (33)	2479	6.9%
Eau Claire, WI (75)	2254	6.3%
Wausau, WI (30)	2062	5.8%
Twin Cities, MN (68)	1980	5.5%
Denver West, CO (82)	1757	4.9%
Bedford, NH (11)	1568	4.4%
South Portland, ME (14)	1402	3.9%
Chicago-Naperville (47)	1324	3.7%
Other values (31)	14741	41.2%

AccountState
Categorical

CARDINALITY

Distinct	52
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	1	?
Unique (%)	< 0.1%	

Sample

1st row	WI
2nd row	WI
3rd row	WI
4th row	WI
5th row	WI

Common Values

Value	Count	Frequency (%)
WI	10593	29.6%
IL	6868	19.2%
CO	3082	8.6%
MT	2693	7.5%
MN	2232	6.2%
ME	1990	5.6%
NH	1466	4.1%
MO	1020	2.9%
CA	734	2.1%
PA	699	2.0%
Other values (42)	4407	12.3%

AccountZip

Categorical

CARDINALITY

Distinct	3098
Distinct (%)	8.7%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	620	?
Unique (%)	1.7%	

Sample

1st row	54474
2nd row	54474
3rd row	54474
4th row	54474
5th row	54474

Common Values

Value	Count	Frequency (%)
54307	360	1.0%
53202	350	1.0%
54701	336	0.9%
80112	325	0.9%
80111	311	0.9%
54401	245	0.7%
54115	241	0.7%
54304	224	0.6%
59601	219	0.6%
54729	200	0.6%
Other values (3088)	32973	92.1%

AccountNaicsName
Categorical

HIGH

CARDINALITY

Distinct	831
Distinct (%)	2.3%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	50	?
Unique (%)	0.1%	

Sample

1st row	Food (Health) Supplement Stores
2nd row	Food (Health) Supplement Stores
3rd row	Food (Health) Supplement Stores
4th row	Food (Health) Supplement Stores
5th row	Food (Health) Supplement Stores

Common Values

Value	Count	Frequency (%)
Lessors of Nonresidential Buildings (except Miniwarehouses)	1109	3.1%
Offices of Physicians (except Mental Health Specialists)	900	2.5%
Other Activities Related to Real Estate	894	2.5%
Lessors of Residential Buildings and Dwellings	882	2.5%
Other General Government Support	780	2.2%
Commercial Banking	775	2.2%
Limited-Service Restaurants	768	2.1%
Elementary and Secondary Schools	703	2.0%
General Medical and Surgical Hospitals	625	1.7%
Commercial and Institutional Building Construction	525	1.5%
Other values (821)	27823	77.8%

ServiceName

Categorical

HIGH

CARDINALITY

Distinct	134
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	13	?
Unique (%)	< 0.1%	

Sample

1st row	Accounting assistance - A&A
2nd row	Property Tax
3rd row	Property Tax
4th row	Accounting assistance - A&A
5th row	Other Consulting-Sales/Use

Common Values

Value	Count	Frequency (%)
COVID19 Consulting - Tax	6700	18.7%
Audit	4172	11.7%
Out of Scope	2463	6.9%
Merged Firm Caseware Conversion	1966	5.5%
Review	1358	3.8%
1065	1280	3.6%
Accounting assistance - A&A	1231	3.4%
ASC 606 Consulting	1040	2.9%
EBP - DC - Benefit Plan	1032	2.9%
1120S	978	2.7%
Other values (124)	13564	37.9%

ServiceLineName
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Audit & Accounting (10)
2nd row	Tax (20)
3rd row	Tax (20)
4th row	Audit & Accounting (10)
5th row	Tax (20)

Common Values

Value	Count	Frequency (%)
Audit & Accounting (10)	19458	54.4%
Tax (20)	16326	45.6%

IndustryName
Categorical

Distinct	9
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Mfg. Retail Distribution
2nd row	Mfg. Retail Distribution
3rd row	Mfg. Retail Distribution
4th row	Mfg. Retail Distribution
5th row	Mfg. Retail Distribution

Common Values

Value	Count	Frequency (%)
Mfg. Retail Distribution	10589	29.6%
Construction/Real Estate	7914	22.1%
Nonprofit & Government	5517	15.4%
General Office	3643	10.2%
Healthcare	3471	9.7%
Financial Services	3018	8.4%
Technology	691	1.9%
Private Client	567	1.6%
Tribal	374	1.0%

JobTaxComplexity
Categorical

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Not Applicable
2nd row	Not Applicable
3rd row	Not Applicable
4th row	Not Applicable
5th row	Not Applicable

Common Values

Value	Count	Frequency (%)
Not Applicable	31330	87.6%
Intermediate	1737	4.9%
Basic	1662	4.6%
Complex	1055	2.9%

JobName
Categorical

CARDINALITY

Distinct	11876
Distinct (%)	33.2%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	8590	?
Unique (%)	24.0%	

Sample

1st row	Accounting assistance - A&A-2013
2nd row	Property Tax-260068
3rd row	Property Tax-260068
4th row	Special Accounting Project
5th row	Other Consulting-Sales/Use-260068

Common Values

Value	Count	Frequency (%)
Covid19 Consulting - Tax	6010	16.8%
Merged Firm Caseware Conversion	1827	5.1%
2020 Audit	635	1.8%
2019 Audit	581	1.6%
Fast-Virtual Outsourced Accounting	361	1.0%
Estimates, Planning, Projections	326	0.9%
Fast-Special Projects	313	0.9%
Other tax consulting	289	0.8%
2020 1065	286	0.8%
Audit	261	0.7%
Other values (11866)	24895	69.6%

ChargeHours
Real number (ℝ)

Distinct	4903
Distinct (%)	13.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	20.19998826

Minimum	-27
Maximum	2145.75
Zeros	8313
Zeros (%)	23.2%
Negative	8
Negative (%)	< 0.1%
Memory size	279.7 KiB
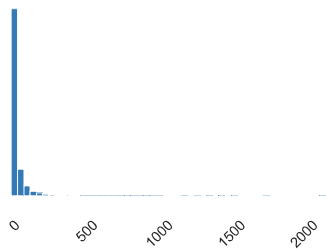
Quantile statistics

Minimum	-27
5-th percentile	0
Q1	0.1
median	1.5
Q3	11.05
95-th percentile	108.6425
Maximum	2145.75
Range	2172.75
Interquartile range (IQR)	10.95

Descriptive statistics

Standard deviation	60.29123225
Coefficient of variation (CV)	2.984716202
Kurtosis	141.8223325
Mean	20.19998826
Median Absolute Deviation (MAD)	1.5
Skewness	8.549774492
Sum	722836.38

| **Variance** | 3635.032686 |
| **Monotonicity** | Not monotonic |
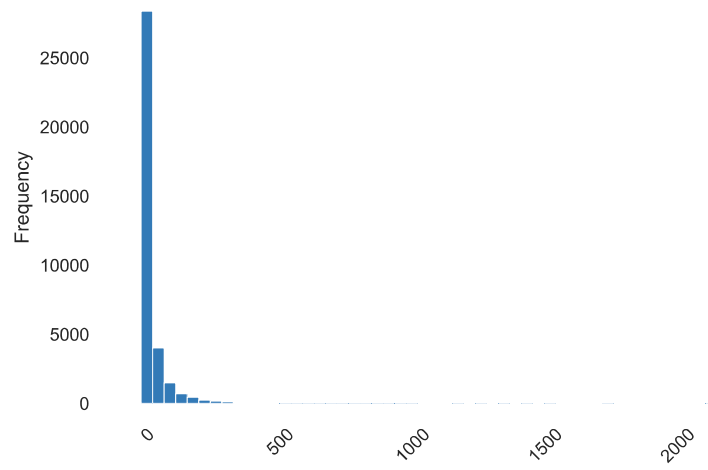
Histogram with fixed size bins (bins=50)

| Value	Count	Frequency (%)
0	8313	23.2%
0.5	1275	3.6%
0.2	901	2.5%
0.3	867	2.4%
1	834	2.3%
0.1	614	1.7%
0.4	597	1.7%
1.5	527	1.5%
0.25	436	1.2%
2	386	1.1%
Other values (4893)	21034	58.8%

Value	Count	Frequency (%)
-27	1	< 0.1%
-10.9	1	< 0.1%
-10.4	1	< 0.1%
-5.6	1	< 0.1%
-3.5	1	< 0.1%

Value	Count	Frequency (%)
2145.75	1	< 0.1%
1725.15	1	< 0.1%
1515.17	1	< 0.1%
1407.65	1	< 0.1%
1360.35	1	< 0.1%

ProductionRevActual

Real number ($\mathbb{R}$)

Distinct	15675
Distinct (%)	43.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4167.171501

|	
Minimum	-10395
Maximum	364657.75
Zeros	8313
Zeros (%)	23.2%
Negative	6
Negative (%)	< 0.1%
Memory size	279.7 KiB
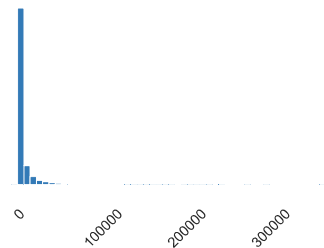
Quantile statistics

Minimum	-10395
5-th percentile	0
Q1	2.05
median	336.315
Q3	2506.6125
95-th percentile	22123.2775
Maximum	364657.75
Range	375052.75
Interquartile range (IQR)	2504.5625

Descriptive statistics

Standard deviation	11950.92951
Coefficient of variation (CV)	2.86787561
Kurtosis	96.6934473
Mean	4167.171501
Median Absolute Deviation (MAD)	336.315
Skewness	7.371533614
Sum	149118065

| **Variance**	142824716.2
Monotonicity	Not monotonic
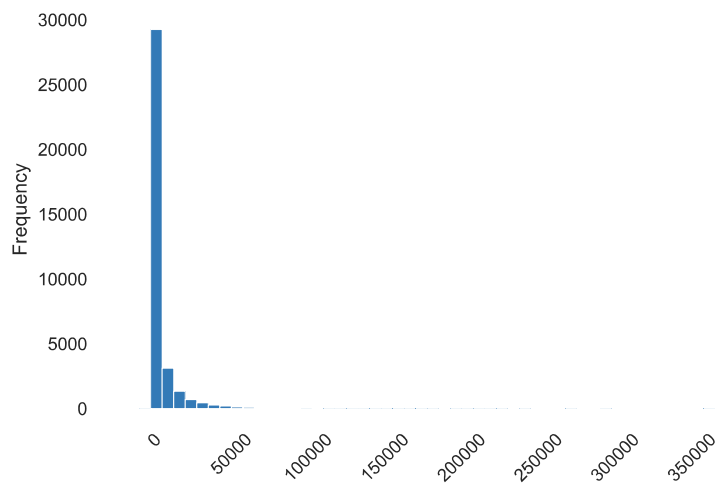
Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	8313	23.2%
105	71	0.2%
65	57	0.2%
120	54	0.2%
130	53	0.1%
78	52	0.1%
0.38	49	0.1%
60	48	0.1%
80	48	0.1%
140	47	0.1%
Other values (15665)	26992	75.4%

Value	Count	Frequency (%)
-10395	1	< 0.1%
-1120	1	< 0.1%
-192.5	1	< 0.1%
-68.9	1	< 0.1%
-60.5	1	< 0.1%

Value	Count	Frequency (%)
364657.75	1	< 0.1%
294195.8	1	< 0.1%
269566.75	1	< 0.1%
240389.05	1	< 0.1%
225730.75	1	< 0.1%

Cost
Real number (ℝ)

Distinct	21290
Distinct (%)	59.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	934.2453535

Minimum	-4500
Maximum	88564.87
Zeros	7890
Zeros (%)	22.0%
Negative	259
Negative (%)	0.7%
Memory size	279.7 KiB
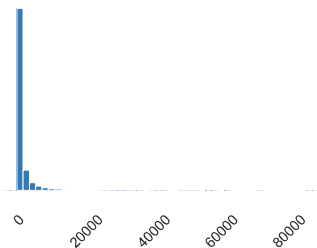
Quantile statistics

Minimum	-4500
5-th percentile	0
Q1	6.31
median	81.86
Q3	563.265
95-th percentile	4918.7055
Maximum	88564.87
Range	93064.87
Interquartile range (IQR)	556.955

Descriptive statistics

Standard deviation	2732.895026
Coefficient of variation (CV)	2.925243369
Kurtosis	134.179345
Mean	934.2453535
Median Absolute Deviation (MAD)	81.86
Skewness	8.401460999
Sum	33431035.73

Variance	7468715.222
Monotonicity	Not monotonic
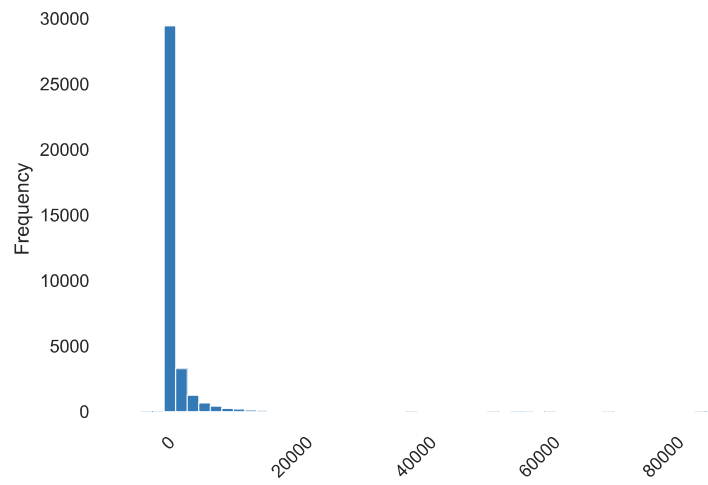
Histogram with fixed size bins (bins=50)

| Value	Count	Frequency (%)
0	7890	22.0%
14.02	36	0.1%
6.31	35	0.1%
2.69	29	0.1%
5.7	25	0.1%
3.27	25	0.1%
8.03	24	0.1%
4.04	20	0.1%
10.57	17	< 0.1%
7.29	17	< 0.1%
Other values (21280)	27666	77.3%

Value	Count	Frequency (%)
-4500	1	< 0.1%
-2367.78	1	< 0.1%
-275.28	1	< 0.1%
-173.37	1	< 0.1%
-91.13	1	< 0.1%

Value	Count	Frequency (%)
88564.87	1	< 0.1%
85873.19	1	< 0.1%
70267.88	1	< 0.1%
61756.14	1	< 0.1%
58110.19	1	< 0.1%

NumAssociatesOnJob

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	32
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.532584395

Minimum	1
Maximum	39
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	279.7 KiB
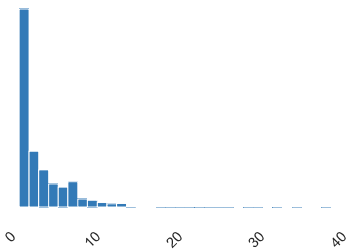
Quantile statistics

Minimum	1
5-th percentile	1
Q1	2
median	2
Q3	5
95-th percentile	10
Maximum	39
Range	38
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	2.883837582
Coefficient of variation (CV)	0.8163534851
Kurtosis	5.978743776
Mean	3.532584395
Median Absolute Deviation (MAD)	1
Skewness	1.973035507
Sum	126410

Variance	8.316519202
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=32)
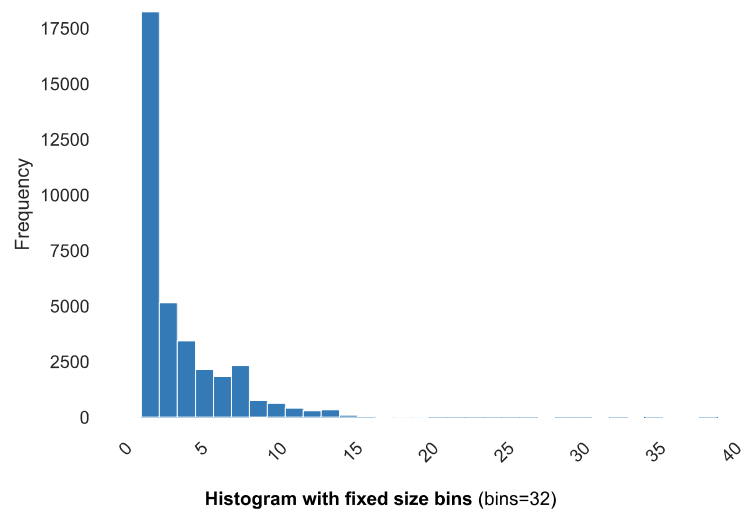
Value	Count	Frequency (%)
2	10511	29.4%
1	7729	21.6%
3	5155	14.4%
4	3431	9.6%
5	2149	6.0%
6	1836	5.1%
7	1316	3.7%
8	1014	2.8%
9	761	2.1%
10	622	1.7%
Other values (22)	1260	3.5%

Value	Count	Frequency (%)
1	7729	21.6%
2	10511	29.4%
3	5155	14.4%
4	3431	9.6%
5	2149	6.0%

| Value	Count	Frequency (%)
39	1	< 0.1%
38	1	< 0.1%
35	1	< 0.1%
32	1	< 0.1%
30	1	< 0.1%

# of Locations
Real number (ℝ≥0)

ZEROS

Distinct	27
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.2459199642

Minimum	0
Maximum	91
Zeros	33003
Zeros (%)	92.2%
Negative	0
Negative (%)	0.0%
Memory size	279.7 KiB
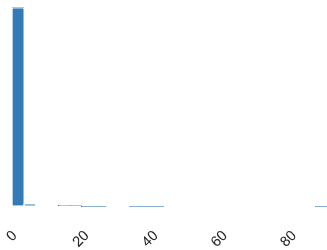
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	1
Maximum	91
Range	91
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	1.766786769
Coefficient of variation (CV)	7.18439747
Kurtosis	760.1325277
Mean	0.2459199642
Median Absolute Deviation (MAD)	0
Skewness	21.46063329

Sum	8800
Variance	3.121535487
Monotonicity	Not monotonic
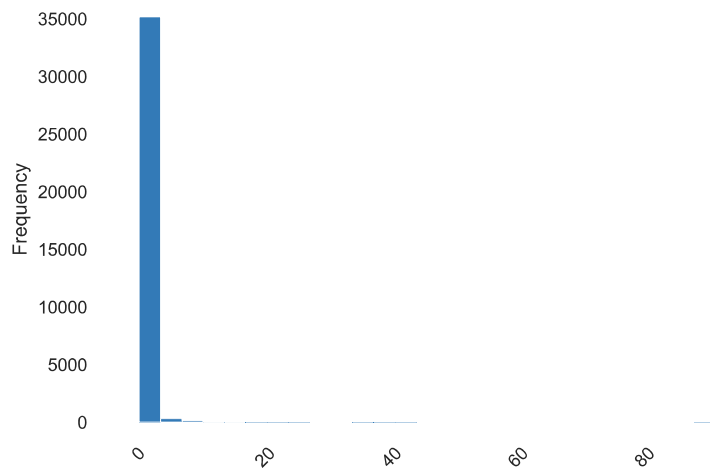
Histogram with fixed size bins (bins=27)

Value	Count	Frequency (%)
0	33003	92.2%
1	1504	4.2%
2	435	1.2%
3	239	0.7%
4	145	0.4%
5	97	0.3%
6	85	0.2%
7	54	0.2%
10	41	0.1%
8	23	0.1%
Other values (17)	158	0.4%

Value	Count	Frequency (%)
0	33003	92.2%
1	1504	4.2%
2	435	1.2%
3	239	0.7%
4	145	0.4%

Value	Count	Frequency (%)
91	3	< 0.1%
43	2	< 0.1%
40	10	< 0.1%
37	10	< 0.1%
26	4	< 0.1%

Annual Sales
Categorical

Distinct	17
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	5M - 9.9M
2nd row	5M - 9.9M
3rd row	5M - 9.9M
4th row	5M - 9.9M
5th row	5M - 9.9M

Common Values

Value	Count	Frequency (%)
Unknown	10578	29.6%
5M - 9.9M	7714	21.6%
10M - 24.9M	4171	11.7%
50M - 99M	3284	9.2%
1M - 2.4M	2062	5.8%
2.5M - 4.9M	1769	4.9%
$500K - $999K	1484	4.1%
25M - 49M	1308	3.7%
$200K - $499K	1088	3.0%
100M - 249M	903	2.5%
Other values (7)	1423	4.0%

Assets
Categorical

Distinct	17
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Unknown
2nd row	Unknown
3rd row	Unknown
4th row	Unknown
5th row	Unknown

Common Values

Value	Count	Frequency (%)
Unknown	33313	93.1%
100M - 249M	465	1.3%
250M - 499M	362	1.0%
50M - 99M	300	0.8%
1M - 2.4M	183	0.5%
2.5M - 4.9M	162	0.5%
500M - 749M	149	0.4%
$500K - $999K	137	0.4%
10M - 24.9M	132	0.4%
25M - 49M	124	0.3%
Other values (7)	457	1.3%

Employees
Categorical

Distinct	12
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Unknown
2nd row	Unknown
3rd row	Unknown
4th row	Unknown
5th row	Unknown

Common Values

Value	Count	Frequency (%)
Unknown	16198	45.3%
1 - 9	6675	18.7%
10 - 24	3476	9.7%
25 - 49	2820	7.9%
100 - 249	2626	7.3%
50 - 99	2434	6.8%
250 - 499	829	2.3%
500 - 999	419	1.2%
1,000 - 2,499	201	0.6%
2,500 - 4,999	80	0.2%
Other values (2)	26	0.1%

SEC/Publicly Traded
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	279.7 KiB

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	No
2nd row	No
3rd row	No
4th row	No
5th row	No

Common Values

Value	Count	Frequency (%)
No	35596	99.5%
Yes	188	0.5%