Overview

Dataset statistics

Number of variables	7
Number of observations	999
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	54.8 KiB
Average record size in memory	56.1 B

Variable types

DateTime	2
Unsupported	1
Categorical	1
Numeric	3

Warnings

|	
`Is_Visit` is highly correlated with `Is_Lead`	**High correlation**
`Is_Lead` is highly correlated with `Is_Visit` and <u>1 other fields (Is_Visit, Is_Sale)</u>	**High correlation**
`Is_Sale` is highly correlated with `Is_Lead`	**High correlation**
`Is_Visit` is highly correlated with `Is_Lead`	**High correlation**
`Is_Lead` is highly correlated with `Is_Visit` and <u>1 other fields (Is_Visit, Is_Sale)</u>	**High correlation**
`Is_Sale` is highly correlated with `Is_Lead`	**High correlation**
`Is_Lead` is highly correlated with `Is_Sale`	**High correlation**
`Is_Sale` is highly correlated with `Is_Lead`	**High correlation**
`Is_Visit` is highly correlated with `Is_Lead` and <u>1 other fields (Is_Lead, Is_Sale)</u>	**High correlation**
`Is_Lead` is highly correlated with `Is_Visit` and <u>1 other fields (Is_Visit, Is_Sale)</u>	**High correlation**
`Is_Sale` is highly correlated with `Is_Visit` and <u>1 other fields (Is_Visit, Is_Lead)</u>	**High correlation**
`Sale_Date` is an unsupported type, check if it needs cleaning or further analysis	**Unsupported**
`Is_Sale` has 19 (1.9%) zeros	**Zeros**

Reproduction

|	
Analysis started	2021-09-05 23:57:16.564355
Analysis finished	2021-09-05 23:57:47.743534
Duration	31.18 seconds
Software version	pandas-profiling v3.0.0 (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%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%2

Variables

Visit_date
Date

Distinct	179
Distinct (%)	17.9%
Missing	0
Missing (%)	0.0%
Memory size	7.9 KiB

Minimum	2020-09-01 00:00:00
Maximum	2021-02-28 00:00:00

Histogram with fixed size bins (bins=50)
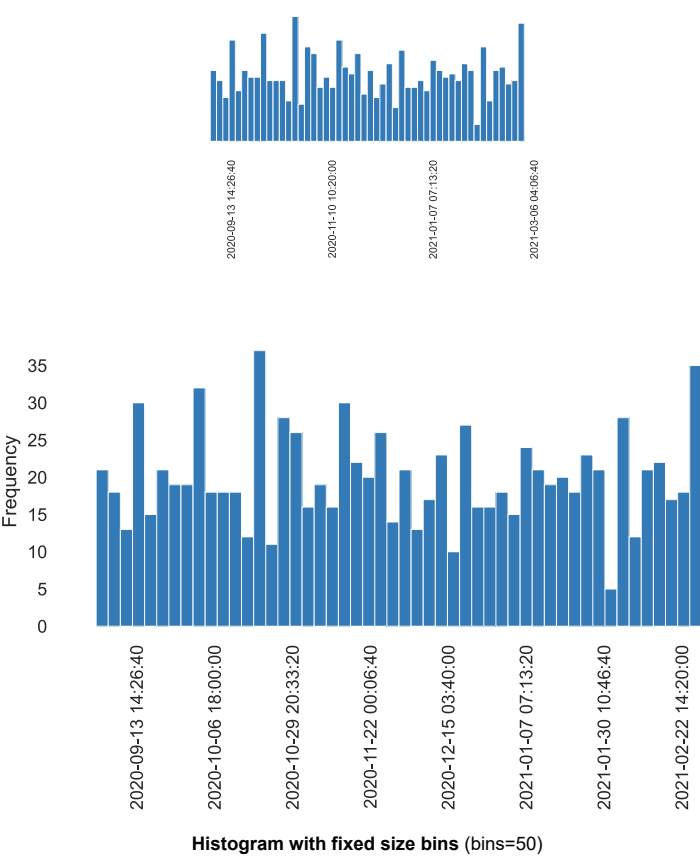
Lead_Date
Date

Distinct	185
Distinct (%)	18.5%
Missing	0
Missing (%)	0.0%
Memory size	7.9 KiB

Minimum	2020-09-01 00:00:00
Maximum	2021-03-06 00:00:00

Histogram with fixed size bins (bins=50)
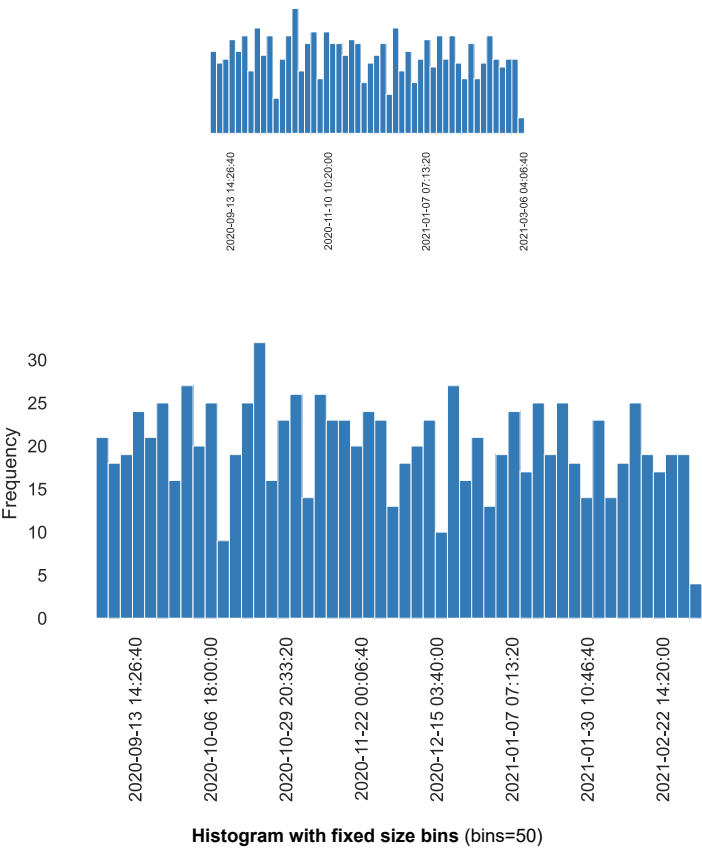
Sale_Date
Unsupported

UNSUPPORTED

REJECTED

Missing	0
Missing (%)	0.0%
Memory size	7.9 KiB

Business_Solution
Categorical

Distinct	2
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	7.9 KiB

Length

Max length	2
Median length	2
Mean length	2
Min length	2

Characters and Unicode

Total characters	1998	
Distinct characters	4	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.
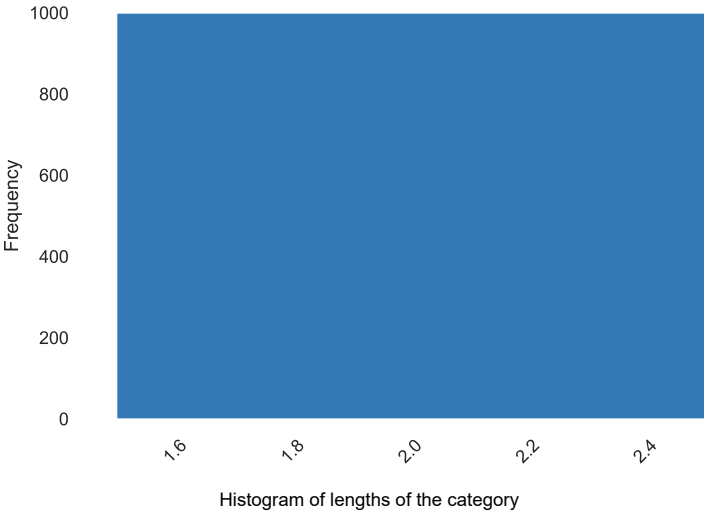
Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	DF
2nd row	OC
3rd row	OC
4th row	DF
5th row	OC

Common Values

Value	Count	Frequency (%)
OC	533	53.4%
DF	466	46.6%

Length

Histogram of lengths of the category

Pie chart

Value	Count	Frequency (%)
oc	533	53.4%
df	466	46.6%
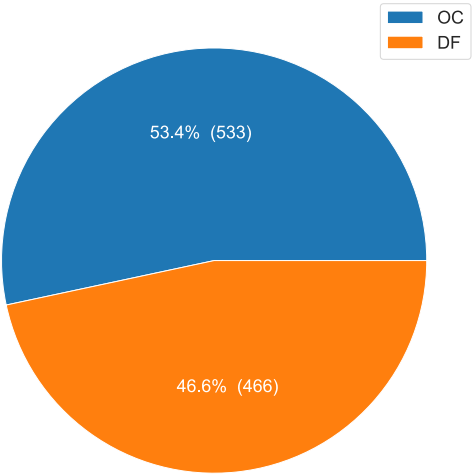
Value	Count	Frequency (%)
	533	53.4%
	466	46.6%

Most occurring characters

Value	Count	Frequency (%)
O	533	26.7%
C	533	26.7%
D	466	23.3%
F	466	23.3%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	1998	100.0%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
O	533	26.7%
C	533	26.7%
D	466	23.3%
F	466	23.3%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1998	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
O	533	26.7%
C	533	26.7%
D	466	23.3%
F	466	23.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1998	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
O	533	26.7%
C	533	26.7%
D	466	23.3%
F	466	23.3%

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

HIGH

CORRELATION (This variable has a high correlation with 1 fields: Is Lead)
HIGH CORRELATION (This variable has a high correlation with 1 fields: Is Lead)
HIGH CORRELATION (This variable has a high correlation with 2 fields: Is Lead, Is Sale)

Distinct	986
Distinct (%)	98.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	49297.31632

Minimum	25456
Maximum	77842
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	7.9 KiB
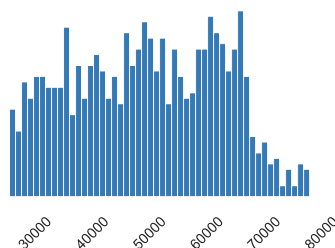
Quantile statistics

Minimum	25456
5-th percentile	28745
Q1	38421.5
median	49382
Q3	60327
95-th percentile	68530.3
Maximum	77842
Range	52386
Interquartile range (IQR)	21905.5

Descriptive statistics

Standard deviation	12894.6004
Coefficient of variation (CV)	0.2615679994
Kurtosis	-1.022914953
Mean	49297.31632
Median Absolute Deviation (MAD)	10957

Skewness	-0.01440372987
Sum	49248019
Variance	166270719.6
Monotonicity	Not monotonic

Data Profiling Report

Histogram with fixed size bins (bins=50)
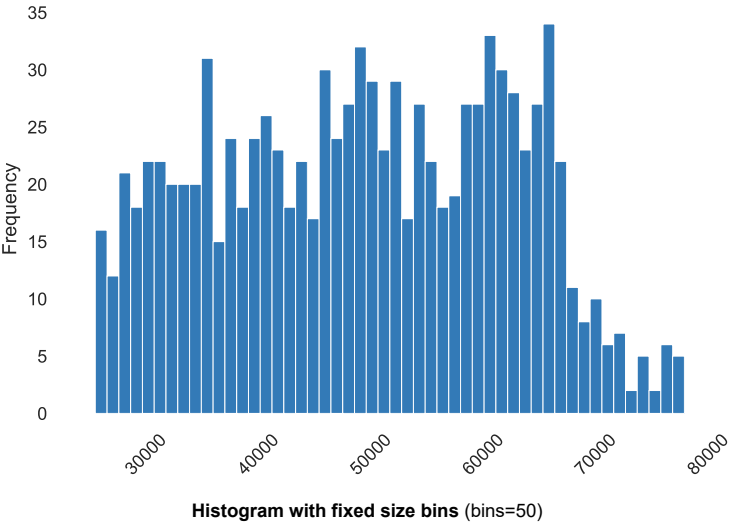
Value	Count	Frequency (%)
65815	2	0.2%
49817	2	0.2%
58938	2	0.2%
33083	2	0.2%
62403	2	0.2%
52265	2	0.2%
54012	2	0.2%
34269	2	0.2%
54194	2	0.2%
52203	2	0.2%
Other values (976)	979	98.0%

Value	Count	Frequency (%)
25456	1	0.1%
25513	1	0.1%
25699	1	0.1%
25703	1	0.1%
25753	1	0.1%
25830	1	0.1%
25868	1	0.1%
25922	1	0.1%
26006	1	0.1%
26126	1	0.1%

Value	Count	Frequency (%)
77842	1	0.1%
77545	1	0.1%
77378	1	0.1%
77322	1	0.1%
76908	1	0.1%
76611	1	0.1%
76462	1	0.1%
76214	1	0.1%
76139	1	0.1%
76075	1	0.1%

Is_Lead

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

CORRELATION (This variable has a high correlation with 2 fields: Is_Visit, Is_Sale)
HIGH CORRELATION (This variable has a high correlation with 2 fields: Is_Visit, Is_Sale)
HIGH CORRELATION (This variable has a high correlation with 1 fields: Is_Sale)
HIGH CORRELATION (This variable has a high correlation with 2 fields: Is_Visit, Is_Sale)

Distinct	886
Distinct (%)	88.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3296.526527

Minimum	845
Maximum	8126
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	7.9 KiB
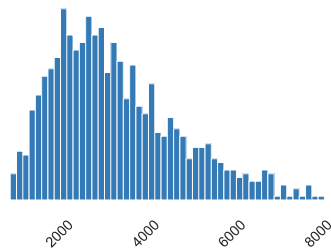
Quantile statistics

Minimum	845
5-th percentile	1398.3
Q1	2178
median	3012
Q3	4153
95-th percentile	6170.2
Maximum	8126
Range	7281
Interquartile range (IQR)	1975

Descriptive statistics

Standard deviation	1455.42151
Coefficient of variation (CV)	0.4415015316
Kurtosis	0.1882425503
Mean	3296.526527
Median Absolute Deviation (MAD)	940

Skewness	0.8015788343
Sum	3293230
Variance	2118251.773
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
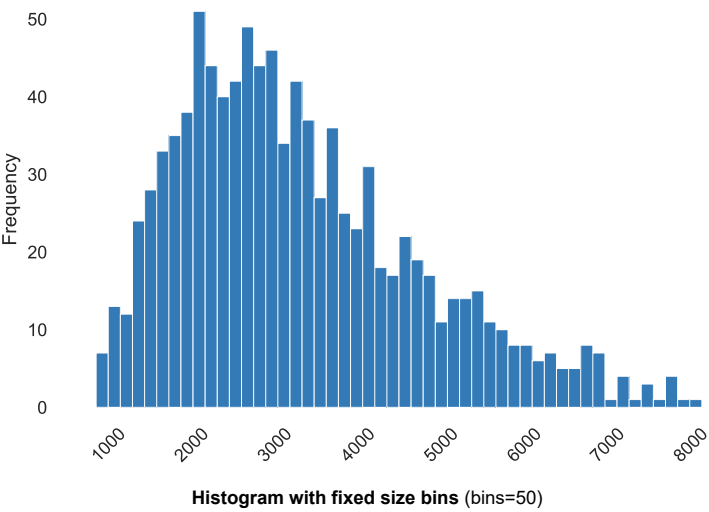
Value	Count	Frequency (%)
1886	3	0.3%
3020	3	0.3%
2388	3	0.3%
3060	3	0.3%
2770	3	0.3%
2525	3	0.3%
2689	3	0.3%
2635	3	0.3%
1940	3	0.3%
2242	3	0.3%
Other values (876)	969	97.0%

Value	Count	Frequency (%)
845	1	0.1%
855	1	0.1%
892	1	0.1%
942	1	0.1%
943	1	0.1%
950	1	0.1%
959	1	0.1%
1000	1	0.1%
1021	1	0.1%
1023	1	0.1%

Value	Count	Frequency (%)
8126	1	0.1%
7849	1	0.1%
7809	1	0.1%
7799	1	0.1%
7750	1	0.1%
7693	1	0.1%
7642	1	0.1%
7510	1	0.1%
7439	1	0.1%
7435	1	0.1%

Is_Sale

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

CORRELATION (This variable has a high correlation with 1 fields: Is Lead)
HIGH CORRELATION (This variable has a high correlation with 1 fields: Is Lead)
HIGH CORRELATION (This variable has a high correlation with 1 fields: Is Lead)
HIGH CORRELATION (This variable has a high correlation with 2 fields: Is Visit, Is Lead)
ZEROS

Distinct	115
Distinct (%)	11.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	42.06206206

Minimum	0
Maximum	133
Zeros	19
Zeros (%)	1.9%
Negative	0
Negative (%)	0.0%
Memory size	7.9 KiB
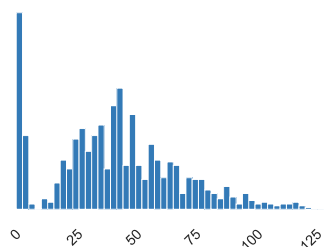
Quantile statistics

Minimum	0
5-th percentile	1
Q1	25
median	41
Q3	58
95-th percentile	90.1
Maximum	133
Range	133
Interquartile range (IQR)	33

Descriptive statistics

Standard deviation	27.02478177
Coefficient of variation (CV)	0.6424977865
Kurtosis	-0.02887136132
Mean	42.06206206

Median Absolute Deviation (MAD)	17
Skewness	0.4220137384
Sum	42020
Variance	730.3388298
Monotonicity	Not monotonic

Data Profiling Report

Histogram with fixed size bins (bins=50)
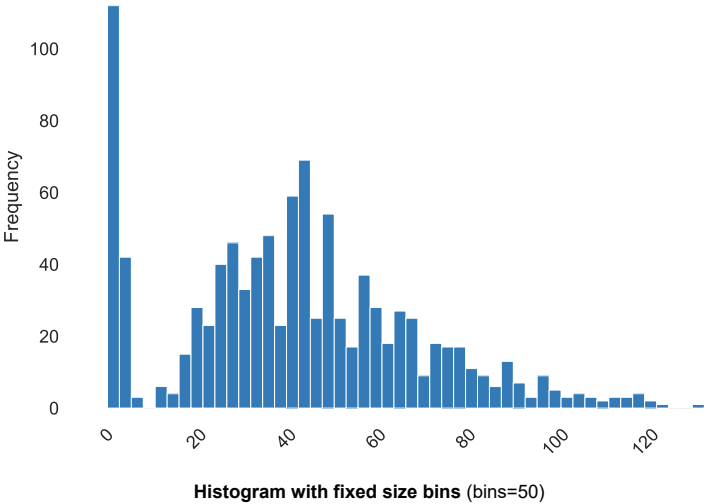
Value	Count	Frequency (%)
1	55	5.5%
2	38	3.8%
45	27	2.7%
48	26	2.6%
3	23	2.3%
40	22	2.2%
44	21	2.1%
43	21	2.1%
42	20	2.0%
0	19	1.9%
Other values (105)	727	72.8%

Data Profiling Report

Value	Count	Frequency (%)
0	19	1.9%
1	55	5.5%
2	38	3.8%
3	23	2.3%
4	13	1.3%
5	6	0.6%
6	3	0.3%
11	2	0.2%
13	4	0.4%
14	2	0.2%

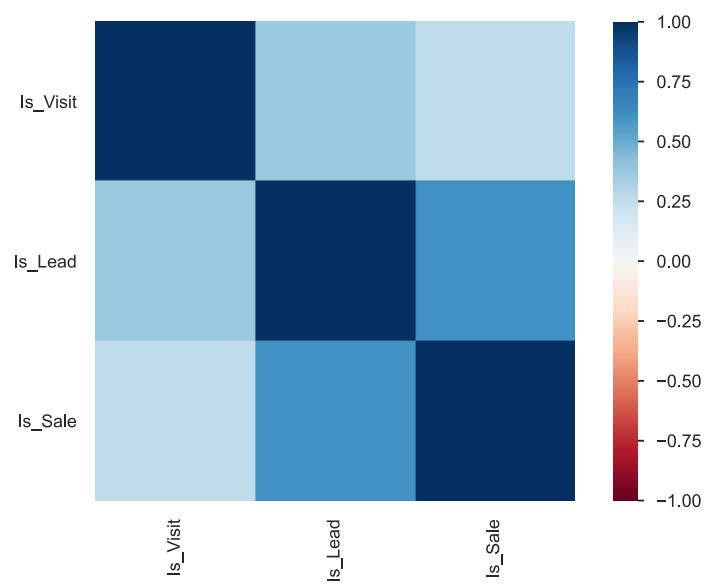| Value	Count	Frequency (%)
133	1	0.1%
125	1	0.1%
120	2	0.2%
119	1	0.1%
118	3	0.3%
116	1	0.1%
115	2	0.2%
114	2	0.2%
112	1	0.1%
111	2	0.2%

Interactions
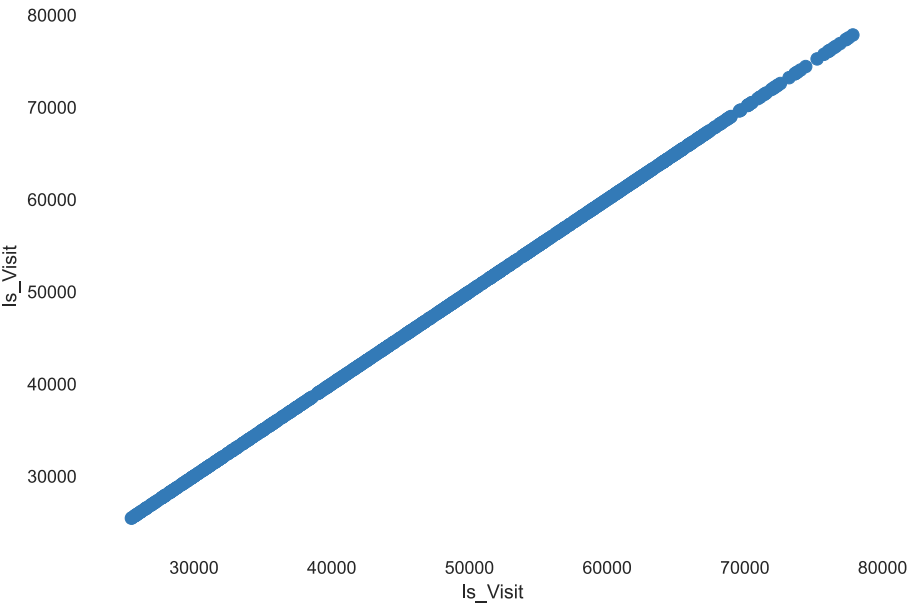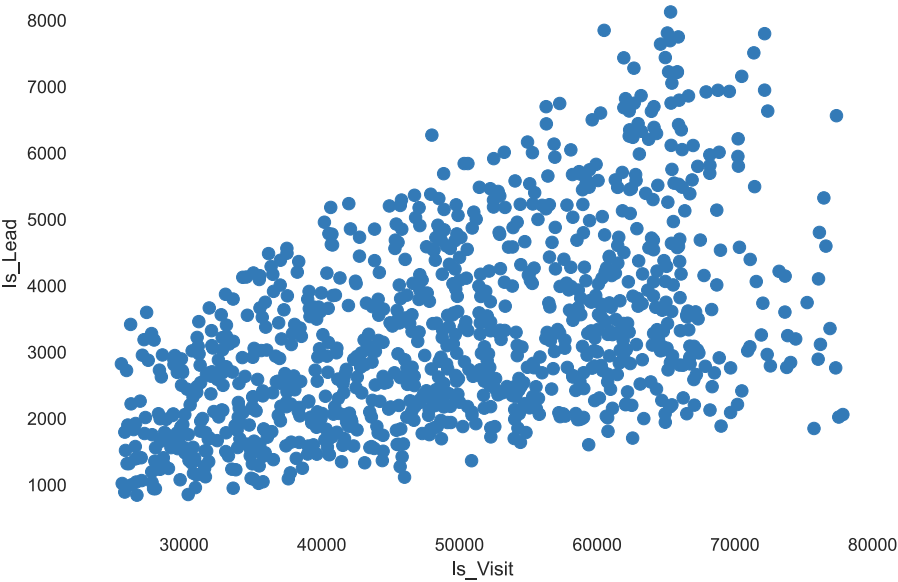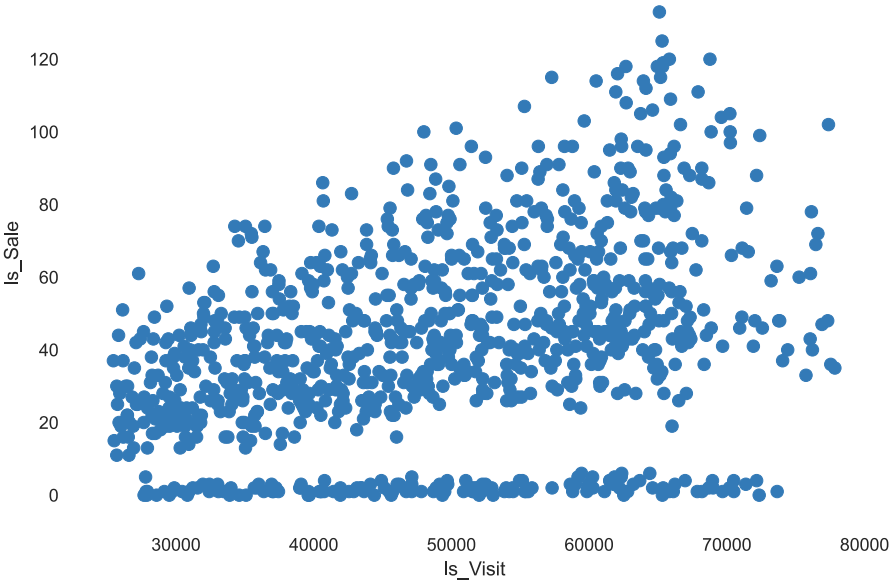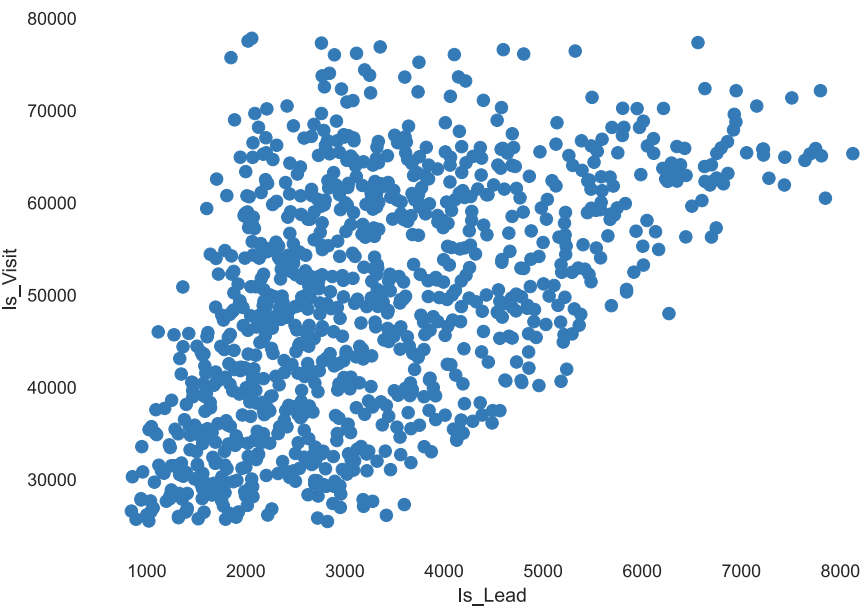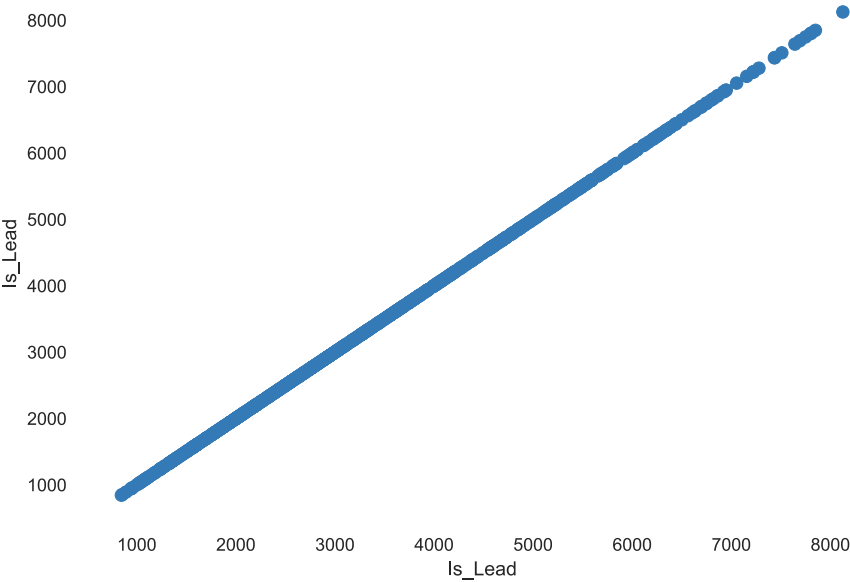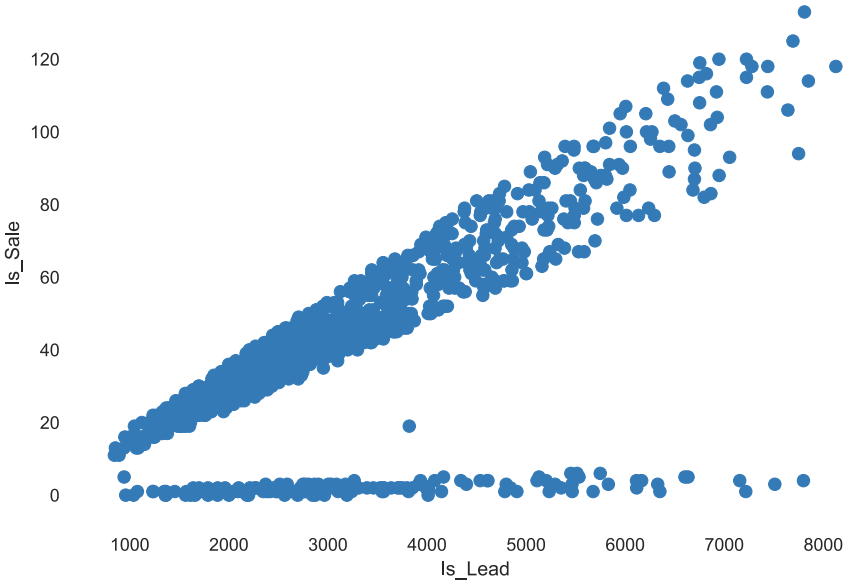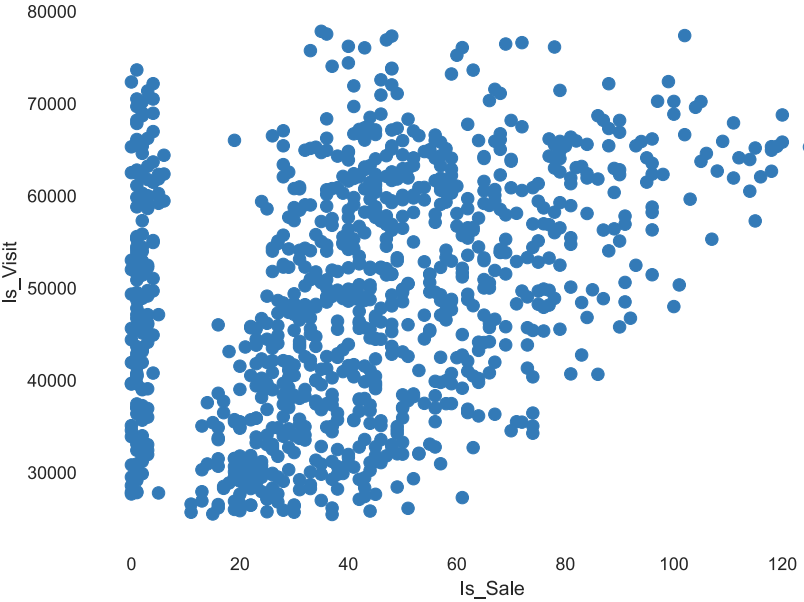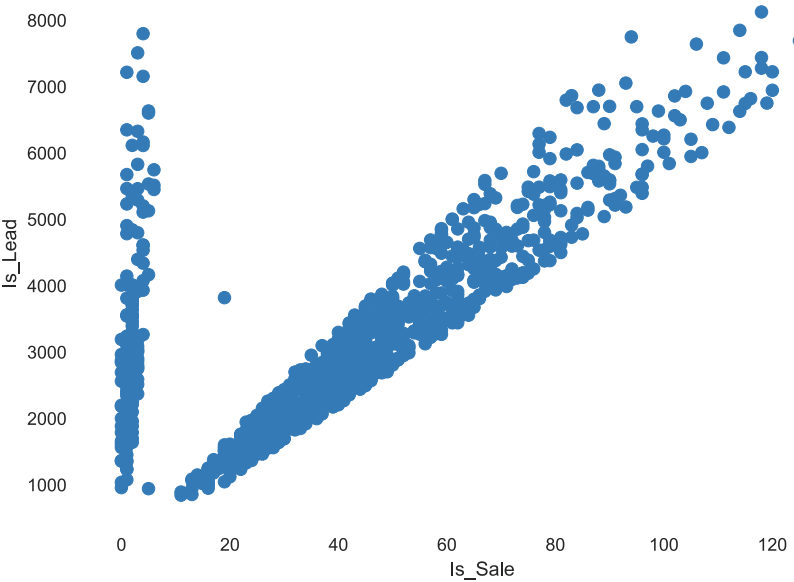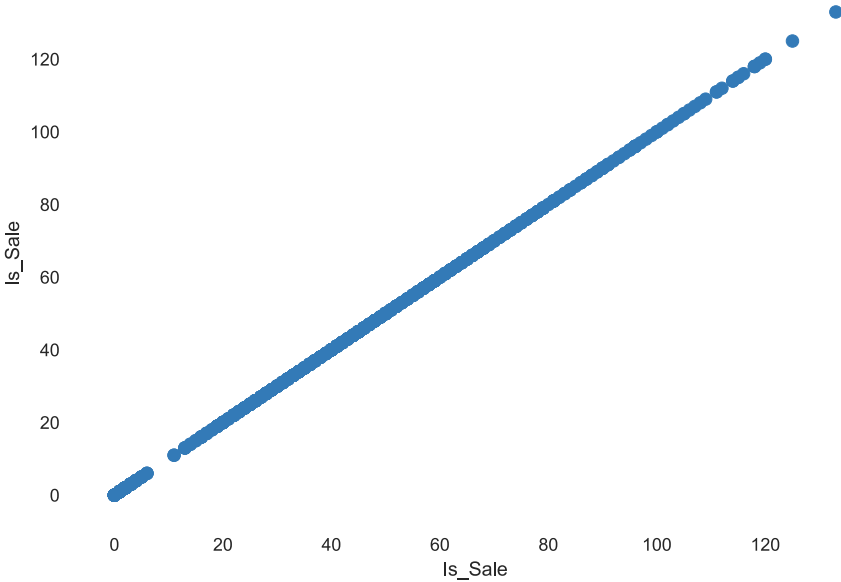
Correlations
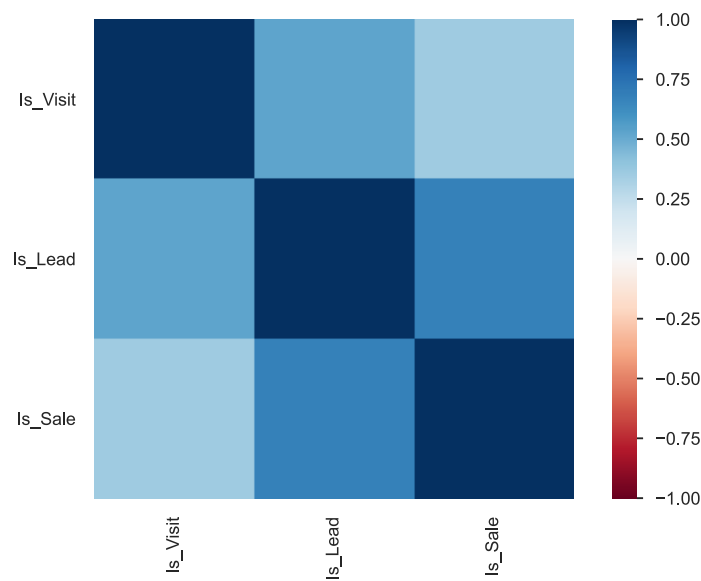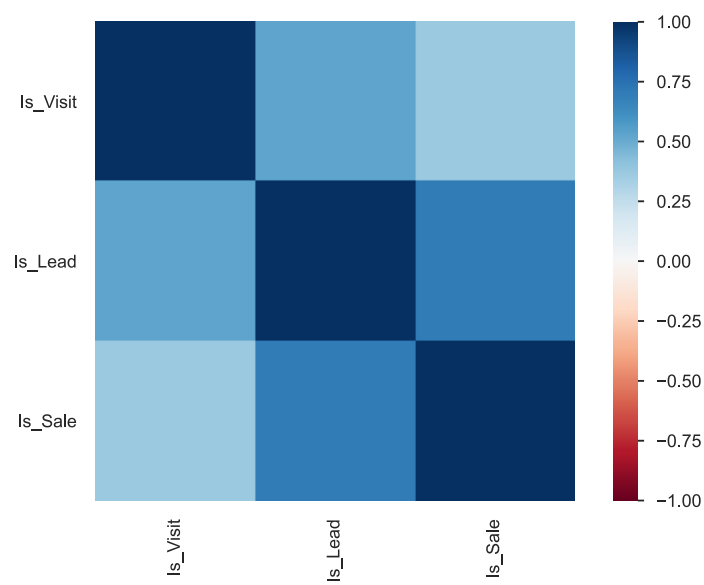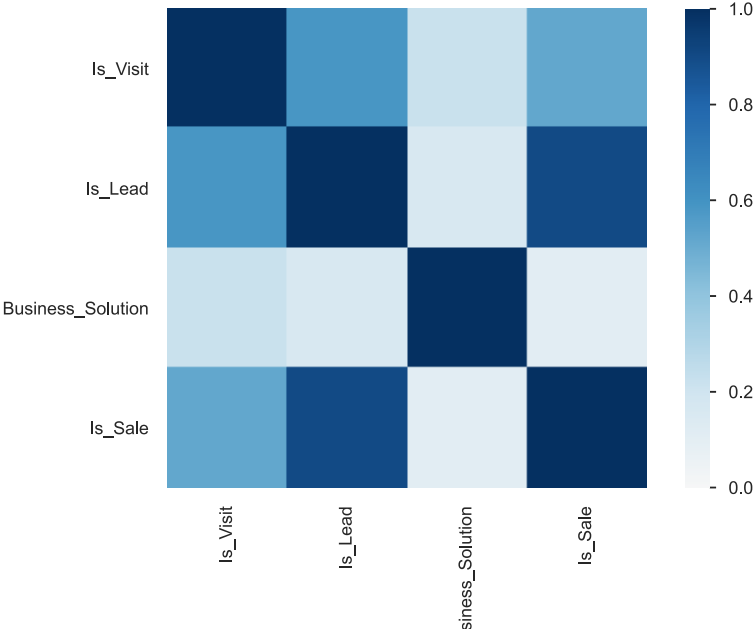
Missing values
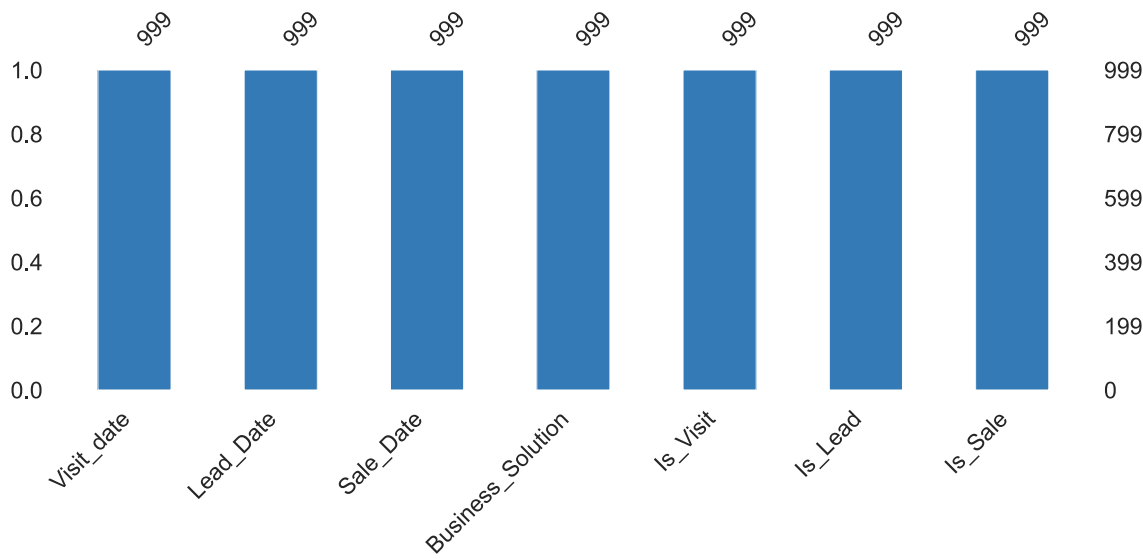
A simple visualization of nullity by column.

Visit_date Lead_Date Sale_Date Business_Solution Is_Visit Is_Lead Is_Sale

1

999
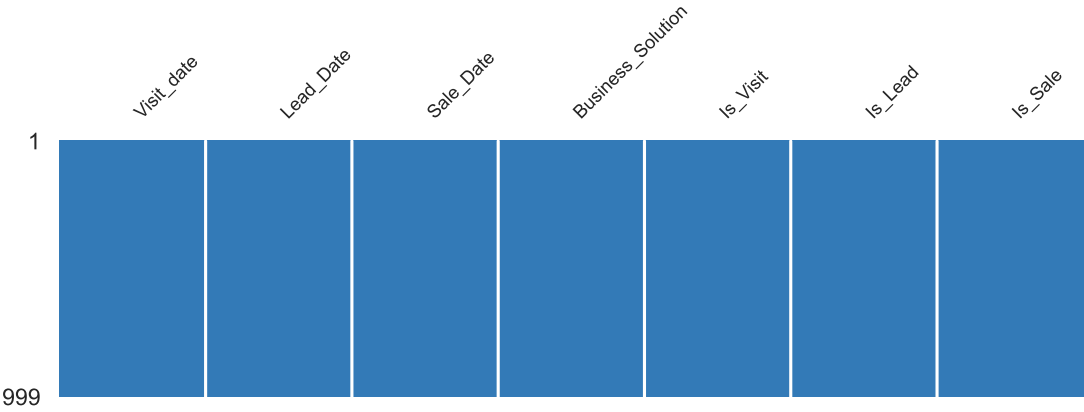
Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

Sample

First rows

	Visit_date	Lead_Date	Sale_Date	Business_Solution	Is_Visit	Is_Lead	Is_Sale
0	2020-12-16	2020-12-16	2021-01-07 00:00:00	DF	62323	3303	45
1	2021-01-12	2021-01-12	2021-02-07 00:00:00	OC	29822	2501	35
2	2021-01-28	2021-01-28	2021-02-19 00:00:00	OC	50566	4549	66
3	2021-02-28	2021-02-28	2021-03-15 00:00:00	DF	53024	3192	0
4	2020-11-04	2020-11-04	2020-11-19 00:00:00	OC	30593	3020	41
5	2021-02-11	2021-02-11	2021-02-29	OC	52574	2453	2
6	2021-02-28	2021-02-28	2021-03-12 00:00:00	OC	51044	5110	4
7	2020-10-26	2020-10-26	2020-11-21 00:00:00	OC	64453	3187	50
8	2020-11-13	2020-11-13	2020-12-11 00:00:00	DF	43309	3113	39
9	2020-09-07	2020-09-07	2020-10-02 00:00:00	DF	34603	3556	49

Last rows

	Visit_date	Lead_Date	Sale_Date	Business_Solution	Is_Visit	Is_Lead	Is_Sale
989	2021-02-23	2021-02-23	2021-03-10 00:00:00	OC	47123	4167	5
990	2020-12-27	2020-12-27	2021-01-20 00:00:00	OC	66503	3559	43
991	2020-10-08	2020-10-08	2020-10-28 00:00:00	OC	33032	1502	25
992	2020-11-17	2020-11-17	2020-12-03 00:00:00	OC	50342	3372	44
993	2020-11-17	2020-11-17	2020-12-29 00:00:00	DF	61198	3279	58
994	2020-12-27	2020-12-27	2021-02-11 00:00:00	DF	61939	6684	84
995	2020-10-20	2020-10-20	2020-11-18 00:00:00	OC	59268	2955	48
996	2020-09-17	2020-09-17	2020-10-08 00:00:00	DF	49676	4680	75
997	2020-10-07	2020-10-07	2020-11-11 00:00:00	OC	49679	4253	72
998	2020-09-26	2020-09-26	2020-10-22 00:00:00	DF	45347	4688	76

Report generated with pandas-profiling (https://github.com/pandas-profiling/pandas-profiling).