Overview

Dataset statistics

Number of variables	20
Number of observations	89
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	60.9 KiB
Average record size in memory	700.1 B

Variable types

Numeric	9
Categorical	11

Alerts

`Year` has constant value "2021"	**Constant**
`Product_Category` has constant value "Bikes"	**Constant**
`Sub_Category` has constant value "Mountain Bikes"	**Constant**
`Unnamed: 0` is highly overall correlated with `Sales_Order #` and 2 other fields (Sales_Order #, Day, Date)	**High correlation**
`Sales_Order #` is highly overall correlated with `Unnamed: 0` and 2 other fields (Unnamed: 0, Day, Date)	**High correlation**
`Day` is highly overall correlated with `Unnamed: 0` and 2 other fields (Unnamed: 0, Sales_Order #, Date)	**High correlation**
`Customer_Age` is highly overall correlated with `Age_Group`	**High correlation**
`Unit_Cost` is highly overall correlated with `Unit_Price` and 4 other fields (Unit_Price , Profit , Cost , Revenue, Product_Description)	**High correlation**
`Unit_Price` is highly overall correlated with `Unit_Cost` and 4 other fields (Unit_Cost , Profit , Cost , Revenue, Product_Description)	**High correlation**
`Profit` is highly overall correlated with `Unit_Cost` and 4 other fields (Unit_Cost , Unit_Price , Cost , Revenue, Order_Quantity)	**High correlation**
`Cost` is highly overall correlated with `Unit_Cost` and 4 other fields (Unit_Cost , Unit_Price , Profit, Revenue, Order_Quantity)	**High correlation**
`Revenue` is highly overall correlated with `Unit_Cost` and 4 other fields (Unit_Cost , Unit_Price , Profit , Cost, Order_Quantity)	**High correlation**
`Date` is highly overall correlated with `Unnamed: 0` and 2 other fields (Unnamed: 0, Sales_Order #, Day)	**High correlation**
`Month` is highly overall correlated with `Product_Description`	**High correlation**
`Age_Group` is highly overall correlated with `Customer_Age`	**High correlation**
`Country` is highly overall correlated with `State`	**High correlation**
`State` is highly overall correlated with `Country`	**High correlation**
`Product_Description` is highly overall correlated with `Unit_Cost` and 2 other fields (Unit_Cost , Unit_Price , Month)	**High correlation**
`Order_Quantity` is highly overall correlated with `Profit` and 2 other fields (Profit , Cost , Revenue)	**High correlation**
`Month` is highly imbalanced (91.1%)	**Imbalance**
`Unnamed: 0` is uniformly distributed	**Uniform**
`Sales_Order #` is uniformly distributed	**Uniform**
`Unnamed: 0` has unique values	**Unique**
`Unnamed: 0` has 1 (1.1%) zeros	**Zeros**
`Unit_Cost` has 1 (1.1%) zeros	**Zeros**
`Unit_Price` has 1 (1.1%) zeros	**Zeros**
`Cost` has 2 (2.2%) zeros	**Zeros**
`Revenue` has 2 (2.2%) zeros	**Zeros**

Reproduction

Analysis started	2023-09-12 08:48:00.542722
Analysis finished	2023-09-12 08:48:20.821682
Duration	20.28 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%22DATA%20ANALYSIS%20REPORT%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%2C

Variables

Select Columns ⌄

Unnamed: 0
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Sales Order #, Day, Date)   UNIFORM   UNIQUE   ZEROS

Distinct	89
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	44
Minimum	0
Maximum	88
Zeros	1
Zeros (%)	1.1%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B

Quantile statistics

Minimum	0
5-th percentile	4.4
Q1	22
median	44
Q3	66
95-th percentile	83.6
Maximum	88
Range	88
Interquartile range (IQR)	44

Descriptive statistics

Standard deviation	25.836021
Coefficient of variation (CV)	0.5871823
Kurtosis	-1.2
Mean	44
Median Absolute Deviation (MAD)	22
Skewness	0
Sum	3916
Variance	667.5

Monotonicity   Strictly increasing

Histogram with fixed size bins (bins=50)

DATA ANALYSIS REPORT
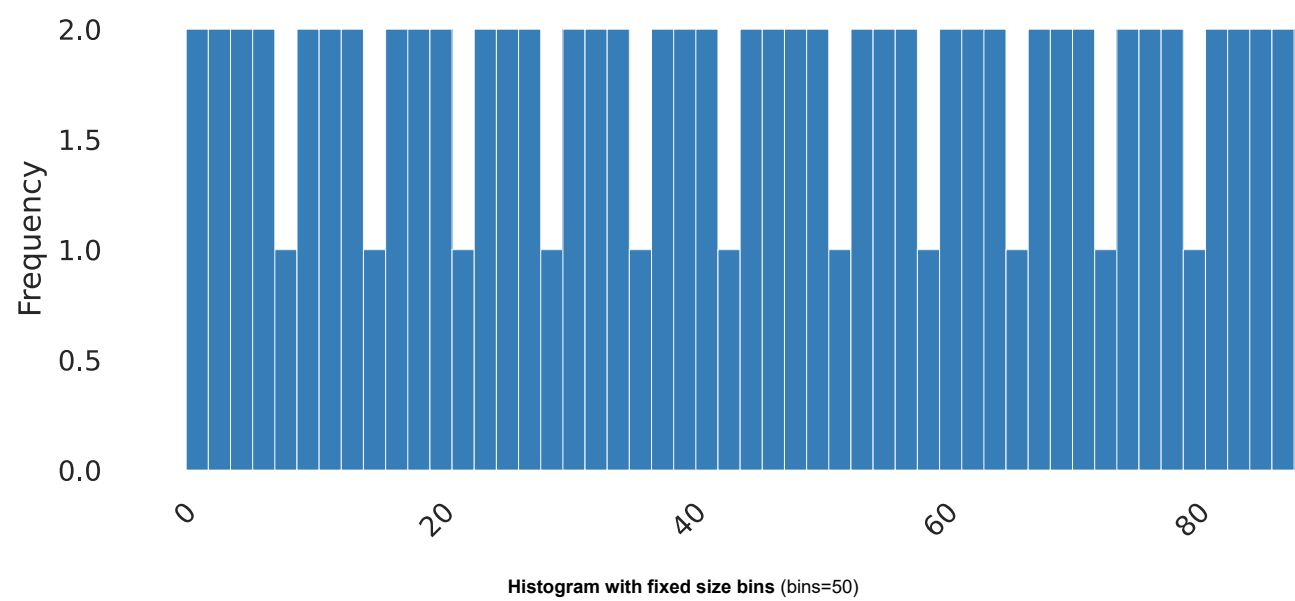
Value	Count	Frequency (%)
0	1	1.1%
45	1	1.1%
65	1	1.1%
64	1	1.1%
63	1	1.1%
62	1	1.1%
61	1	1.1%
60	1	1.1%
59	1	1.1%
58	1	1.1%
Other values (79)	79	88.8%

Value	Count	Frequency (%)
0	1	1.1%
1	1	1.1%
2	1	1.1%
3	1	1.1%
4	1	1.1%
5	1	1.1%
6	1	1.1%
7	1	1.1%
8	1	1.1%
9	1	1.1%

Value	Count	Frequency (%)
88	1	1.1%
87	1	1.1%
86	1	1.1%
85	1	1.1%
84	1	1.1%
83	1	1.1%
82	1	1.1%
81	1	1.1%
80	1	1.1%
79	1	1.1%

Sales_Order
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Unnamed: 0, Day, Date)   UNIFORM

Distinct	87
Distinct (%)	97.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	261738.07

Minimum	261695
Maximum	261782
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B

+2.617e5
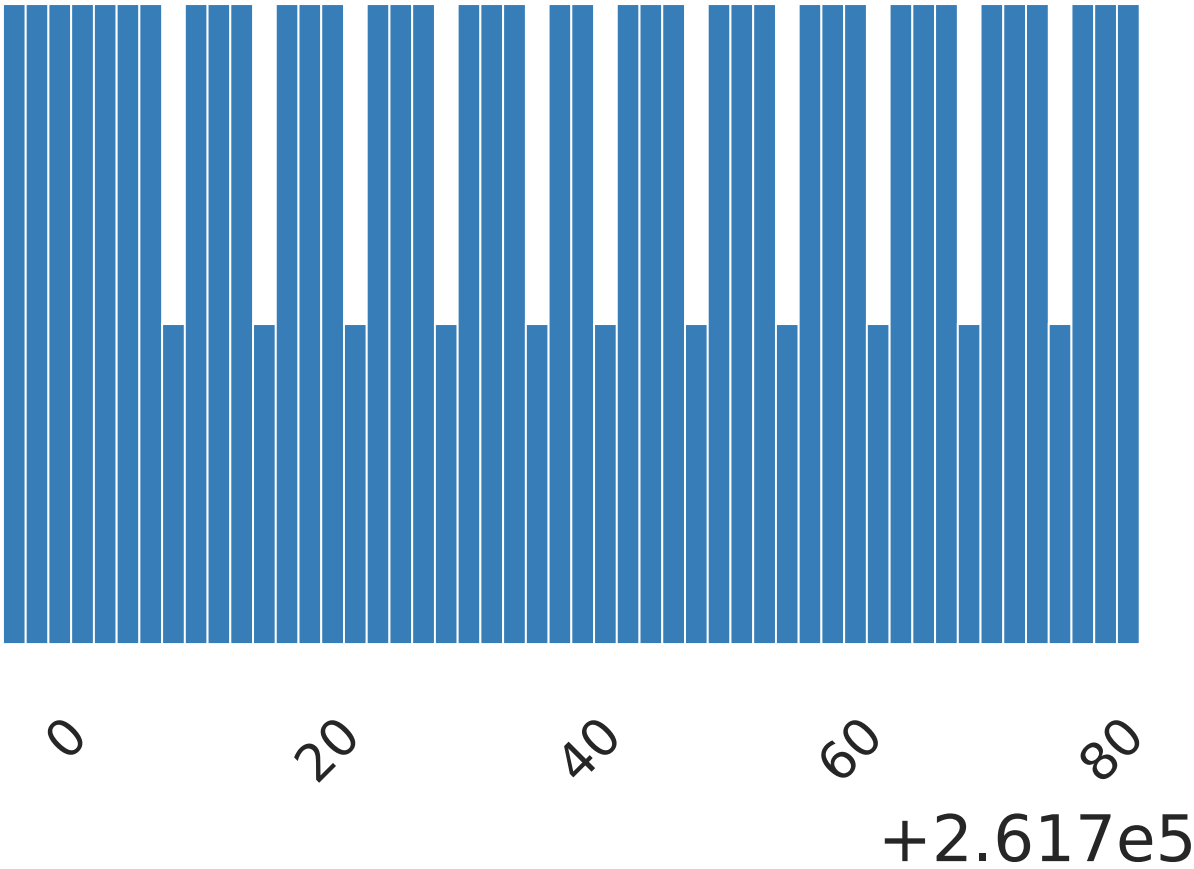
Quantile statistics

Minimum	261695
5-th percentile	261699.4
Q1	261716
median	261738
Q3	261760
95-th percentile	261777.6
Maximum	261782
Range	87
Interquartile range (IQR)	44

Descriptive statistics

Standard deviation	25.729713
Coefficient of variation (CV)	9.8303289×10^{-5}
Kurtosis	-1.2119859
Mean	261738.07
Median Absolute Deviation (MAD)	22
Skewness	0.011826881
Sum	23294688
Variance	662.01813
Monotonicity	Increasing

Histogram with fixed size bins (bins=50)
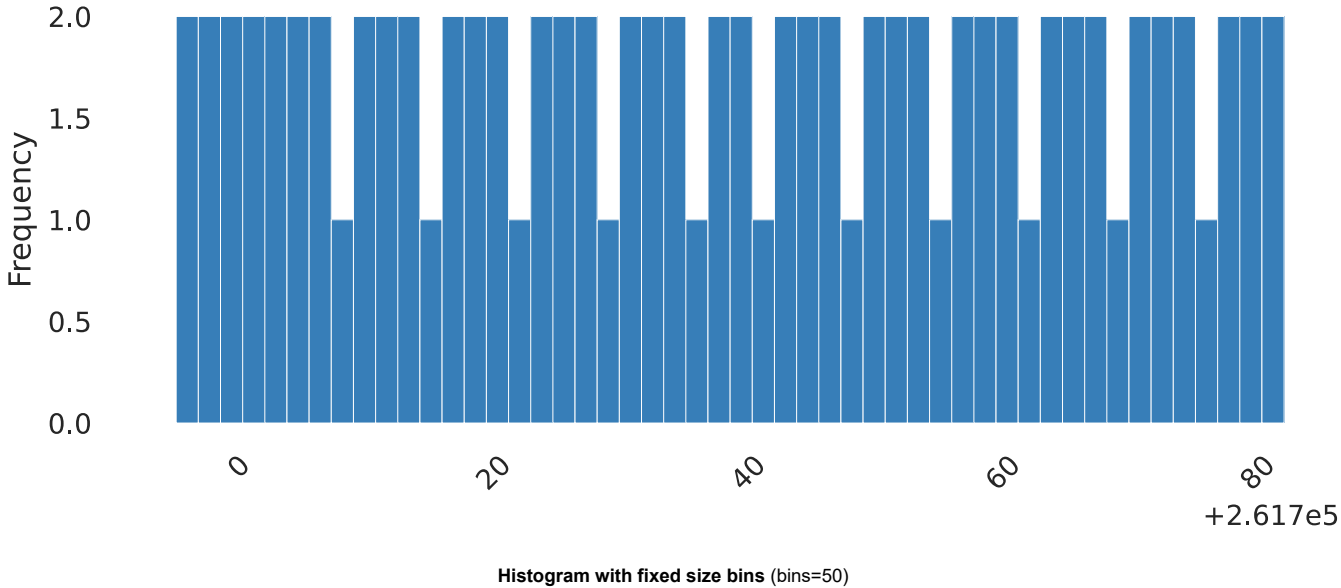
Value	Count	Frequency (%)
261695	2	2.2%
261701	2	2.2%
261751	1	1.1%
261759	1	1.1%
261758	1	1.1%
261757	1	1.1%
261756	1	1.1%
261755	1	1.1%
261754	1	1.1%
261753	1	1.1%
Other values (77)	77	86.5%

Value	Count	Frequency (%)
261695	2	2.2%
261697	1	1.1%
261698	1	1.1%
261699	1	1.1%
261700	1	1.1%
261701	2	2.2%
261702	1	1.1%
261703	1	1.1%
261704	1	1.1%
261705	1	1.1%

Value	Count	Frequency (%)
261782	1	1.1%
261781	1	1.1%
261780	1	1.1%
261779	1	1.1%
261778	1	1.1%
261777	1	1.1%
261776	1	1.1%
261775	1	1.1%
261774	1	1.1%
261773	1	1.1%

Date
Categorical

Distinct	24
Distinct (%)	27.0%
Missing	0
Missing (%)	0.0%
Memory size	5.9 KiB

Length

Max length	10
Median length	10
Mean length	10
Min length	10

Characters and Unicode

Total characters	890	
Distinct characters	11	
Distinct categories	2 (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	3	?
Unique (%)	3.4%	

Sample

1st row	2021-12-01
2nd row	2021-12-01
3rd row	2021-12-02
4th row	2021-12-02
5th row	2021-12-03

Common Values

Value	Count	Frequency (%)
2021-12-19	10	11.2%
2021-12-18	7	7.9%
2021-12-22	6	6.7%
2021-12-10	6	6.7%
2021-12-11	5	5.6%
2021-12-12	5	5.6%
2021-12-20	5	5.6%
2021-12-13	4	4.5%
2021-12-03	4	4.5%
2021-12-05	4	4.5%
Other values (14)	33	37.1%

Length
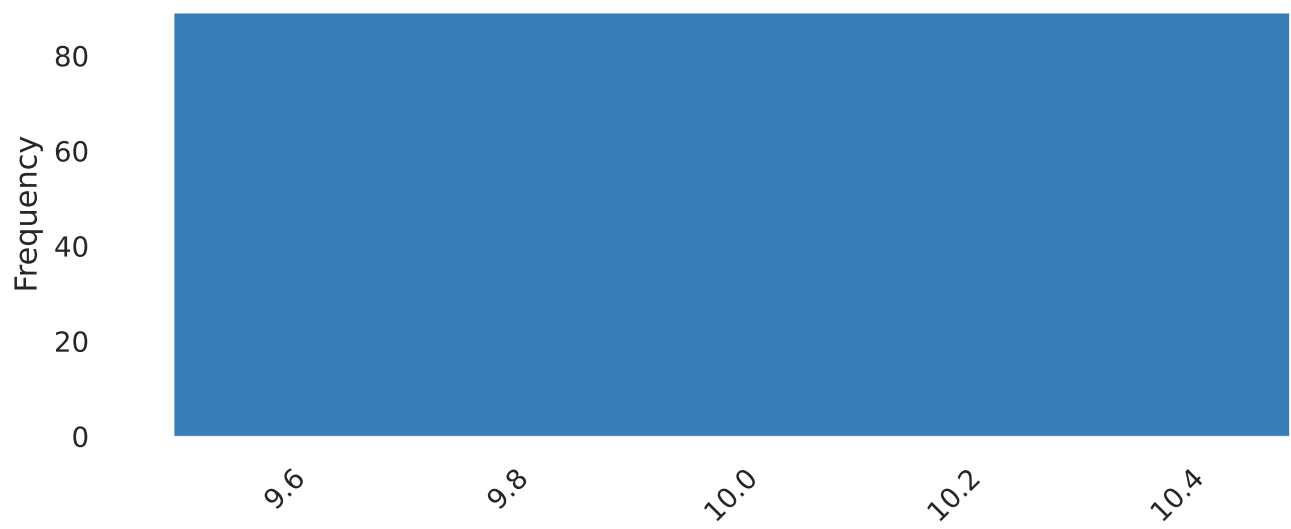
Histogram of lengths of the category

Value	Count	Frequency (%)
2021-12-19	10	11.2%
2021-12-18	7	7.9%
2021-12-22	6	6.7%
2021-12-10	6	6.7%
2021-12-11	5	5.6%
2021-12-12	5	5.6%
2021-12-20	5	5.6%
2021-12-13	4	4.5%
2021-12-03	4	4.5%
2021-12-05	4	4.5%
Other values (14)	33	37.1%

Most occurring characters

Value	Count	Frequency (%)
2	297	33.4%
1	234	26.3%
-	178	20.0%
0	125	14.0%
9	12	1.3%
8	11	1.2%
3	11	1.2%
6	7	0.8%
5	5	0.6%
7	5	0.6%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	712	80.0%
Dash Punctuation	178	20.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
2	297	41.7%
1	234	32.9%
0	125	17.6%
9	12	1.7%
8	11	1.5%
3	11	1.5%
6	7	1.0%
5	5	0.7%
7	5	0.7%
4	5	0.7%

Dash Punctuation

Value	Count	Frequency (%)
-	178	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	890	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
2	297	33.4%
1	234	26.3%
-	178	20.0%
0	125	14.0%
9	12	1.3%
8	11	1.2%
3	11	1.2%
6	7	0.8%
5	5	0.6%
7	5	0.6%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	890	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
2	297	33.4%
1	234	26.3%
-	178	20.0%
0	125	14.0%
9	12	1.3%
8	11	1.2%
3	11	1.2%
6	7	0.8%
5	5	0.6%
7	5	0.6%

Day
Real number ($\mathbb{R}$)

Distinct	24
Distinct (%)	27.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	13.52809
Minimum	1
Maximum	24
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
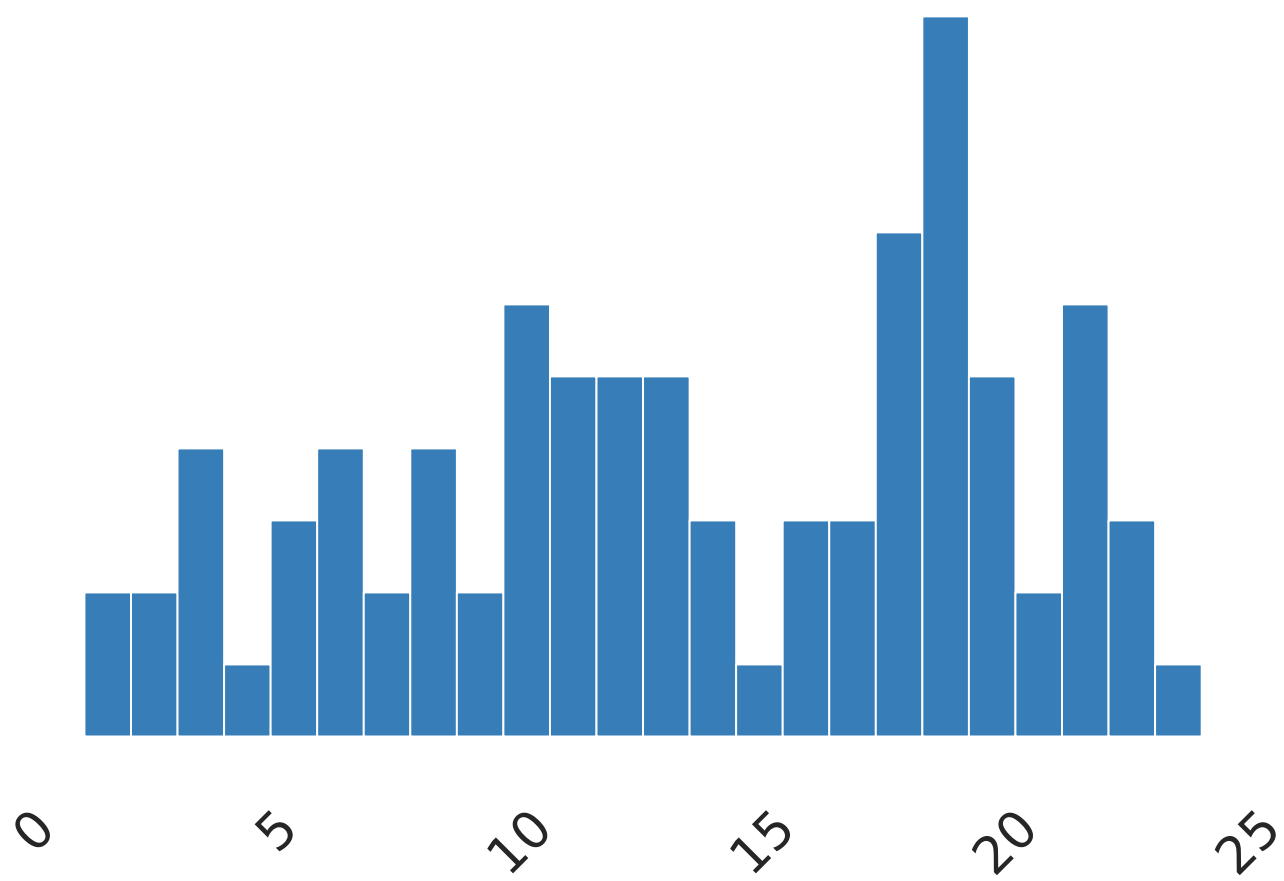
Quantile statistics

Minimum	1
5-th percentile	3
Q1	9
median	13
Q3	19
95-th percentile	22
Maximum	24
Range	23
Interquartile range (IQR)	10

Descriptive statistics

Standard deviation	6.3551874
Coefficient of variation (CV)	0.46977714
Kurtosis	-1.0581022
Mean	13.52809
Median Absolute Deviation (MAD)	6
Skewness	-0.26734215
Sum	1204
Variance	40.388407
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=24)
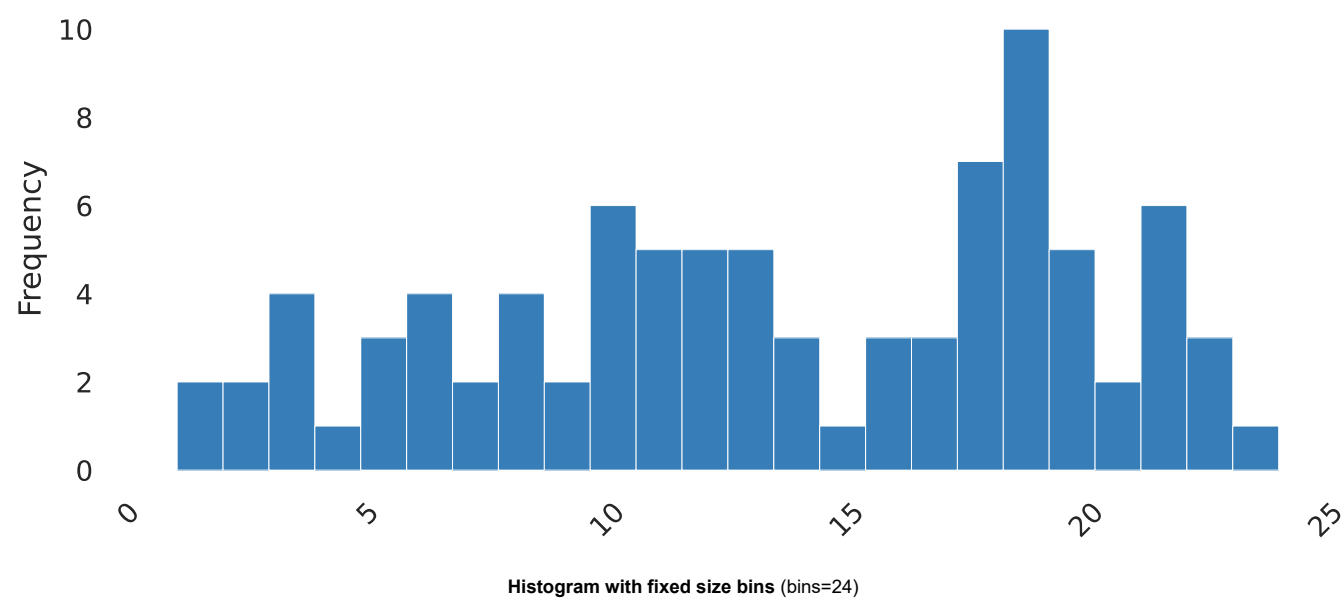
Value	Count	Frequency (%)
19	10	11.2%
18	7	7.9%
10	6	6.7%
22	6	6.7%
12	5	5.6%
20	5	5.6%
11	5	5.6%
13	5	5.6%
8	4	4.5%
6	4	4.5%
Other values (14)	32	36.0%

Value	Count	Frequency (%)
1	2	2.2%
2	2	2.2%
3	4	4.5%
4	1	1.1%
5	3	3.4%
6	4	4.5%
7	2	2.2%
8	4	4.5%
9	2	2.2%
10	6	6.7%

Value	Count	Frequency (%)
24	1	1.1%
23	3	3.4%
22	6	6.7%
21	2	2.2%
20	5	5.6%
19	10	11.2%
18	7	7.9%
17	3	3.4%
16	3	3.4%
15	1	1.1%

Month
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Product Description)   IMBALANCE

Distinct	2
Distinct (%)	2.2%
Missing	0
Missing (%)	0.0%
Memory size	5.8 KiB

Length

Max length	8
Median length	8
Mean length	7.988764
Min length	7

Characters and Unicode

Total characters	711	
Distinct characters	6	
Distinct categories	2 (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	1	?
Unique (%)	1.1%	

Sample

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

Common Values

Value	Count	Frequency (%)
December	88	98.9%
Decmber	1	1.1%

Length
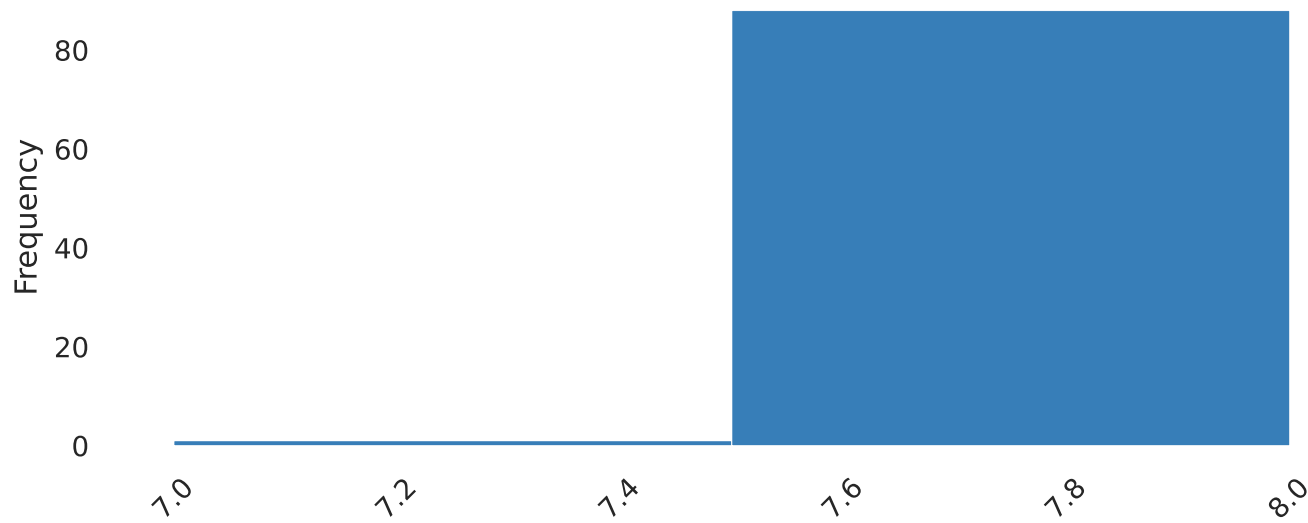
Histogram of lengths of the category

Common Values (Plot)

98.9%
(88)
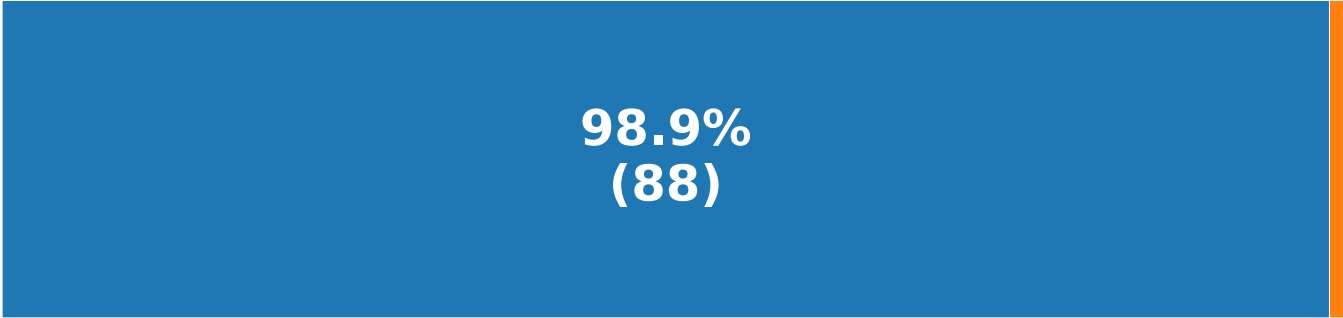
December
Decmber

Value	Count	Frequency (%)
december	88	98.9%
decmber	1	1.1%

Most occurring characters

Value	Count	Frequency (%)
e	266	37.4%
D	89	12.5%
c	89	12.5%
m	89	12.5%
b	89	12.5%
r	89	12.5%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	622	87.5%
Uppercase Letter	89	12.5%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	266	42.8%
c	89	14.3%
m	89	14.3%
b	89	14.3%
r	89	14.3%

Uppercase Letter

Value	Count	Frequency (%)
D	89	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	711	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	266	37.4%
D	89	12.5%
c	89	12.5%
m	89	12.5%
b	89	12.5%
r	89	12.5%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	711	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	266	37.4%
D	89	12.5%
c	89	12.5%
m	89	12.5%
b	89	12.5%
r	89	12.5%

Year
Categorical

Distinct	1
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Memory size	5.4 KiB

Length

Max length	4
Median length	4
Mean length	4
Min length	4

Characters and Unicode

Total characters	356	
Distinct characters	3	
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	2021
2nd row	2021
3rd row	2021
4th row	2021
5th row	2021

Common Values

Value	Count	Frequency (%)
2021	89	100.0%

Length
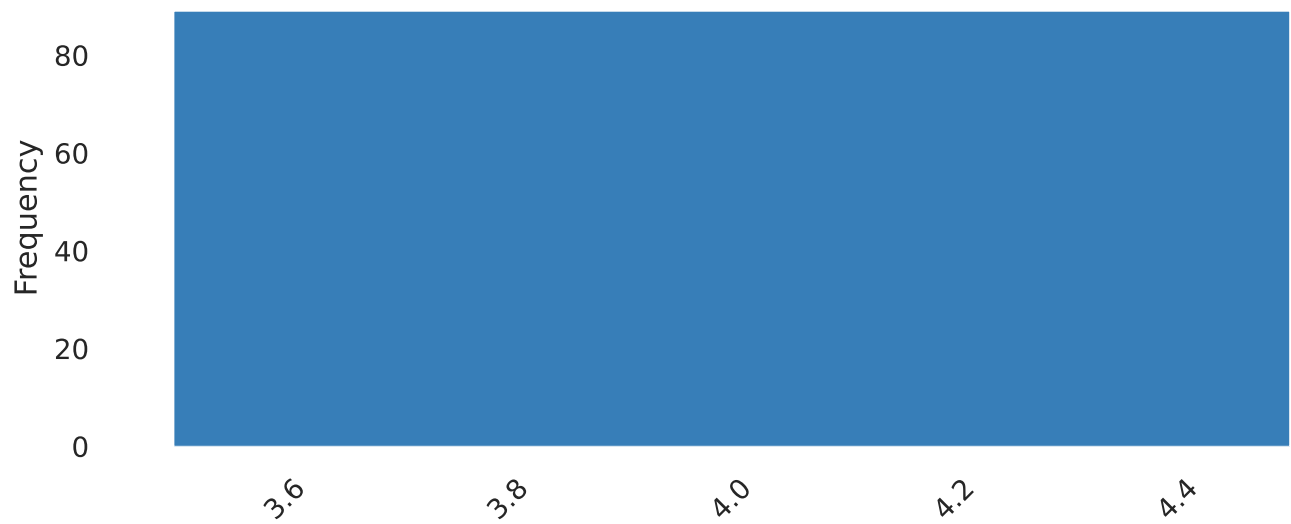
Histogram of lengths of the category

Common Values (Plot)
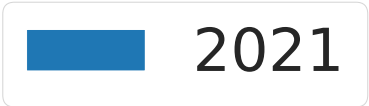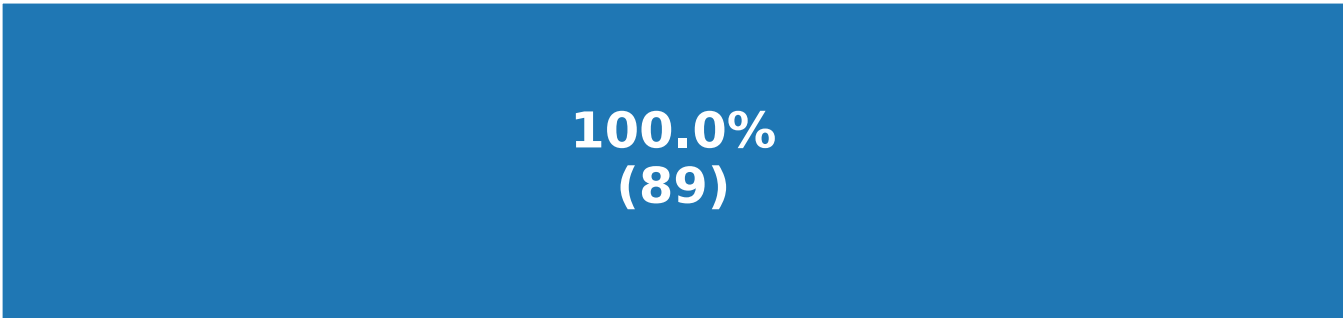
100.0%
(89)

2021

Value	Count	Frequency (%)
2021	89	100.0%

Most occurring characters

Value	Count	Frequency (%)
2	178	50.0%
0	89	25.0%
1	89	25.0%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	356	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
2	178	50.0%
0	89	25.0%
1	89	25.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	356	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
2	178	50.0%
0	89	25.0%
1	89	25.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	356	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
2	178	50.0%
0	89	25.0%
1	89	25.0%

Customer_Age
Real number (ℝ)

Distinct	28
Distinct (%)	31.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	34.202247
Minimum	17
Maximum	63
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
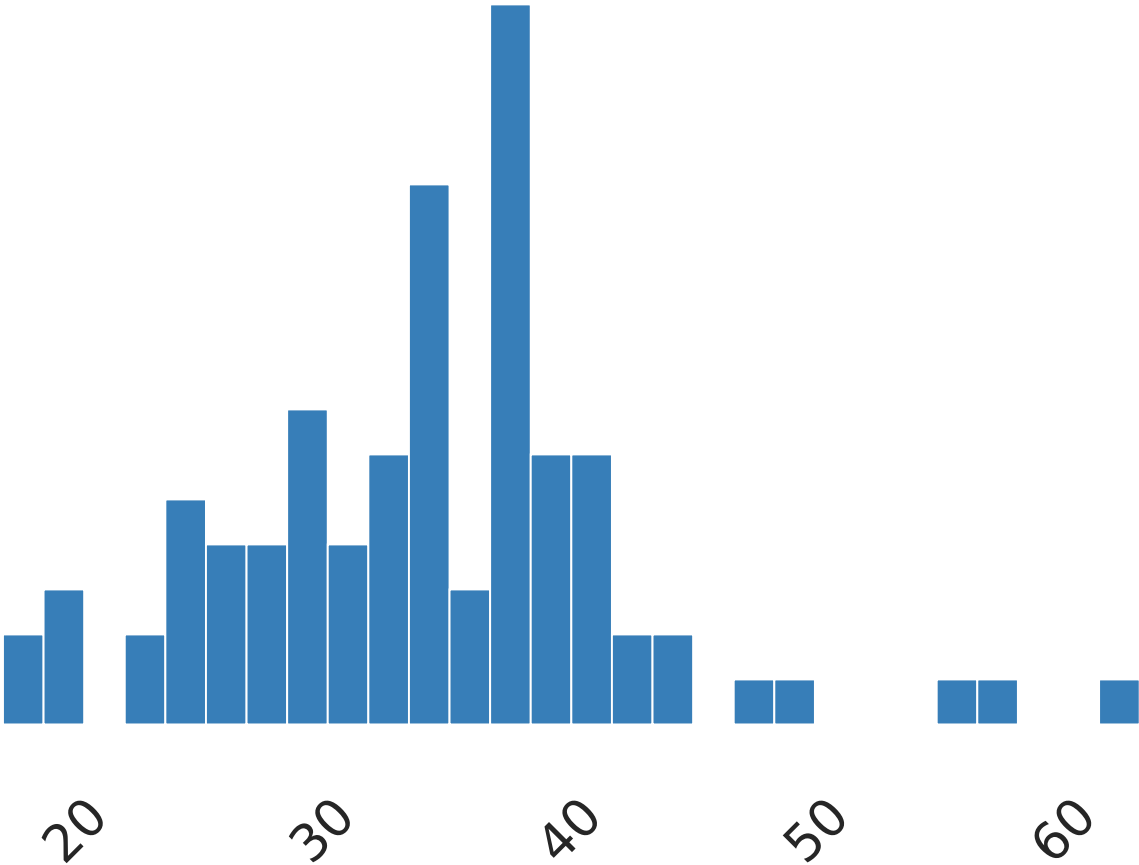
Quantile statistics

Minimum	17
5-th percentile	20.6
Q1	30
median	35
Q3	38
95-th percentile	45.8
Maximum	63
Range	46
Interquartile range (IQR)	8

Descriptive statistics

Standard deviation	8.0906509
Coefficient of variation (CV)	0.2365532
Kurtosis	1.8002764
Mean	34.202247
Median Absolute Deviation (MAD)	4
Skewness	0.55319265
Sum	3044
Variance	65.458631
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=28)
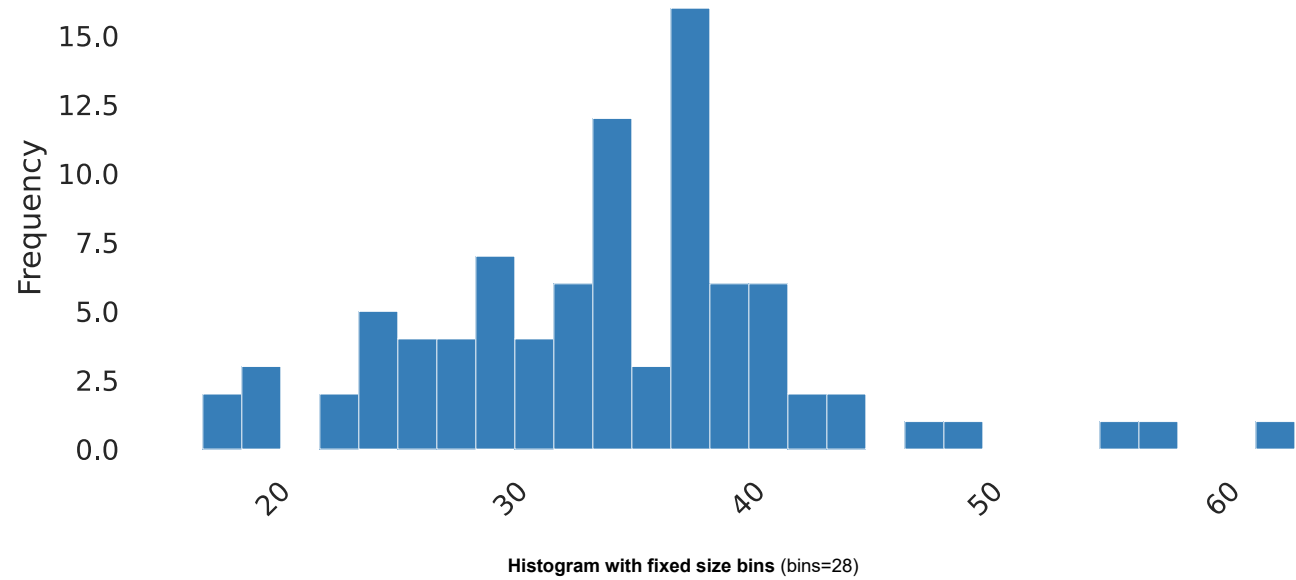
Value	Count	Frequency (%)
37	9	10.1%
35	8	9.0%
38	7	7.9%
39	6	6.7%
30	5	5.6%
31	4	4.5%
26	4	4.5%
34	4	4.5%
27	4	4.5%
41	4	4.5%
Other values (18)	34	38.2%

Value	Count	Frequency (%)
17	1	1.1%
18	1	1.1%
19	3	3.4%
23	2	2.2%
24	3	3.4%
25	2	2.2%
26	4	4.5%
27	4	4.5%
29	2	2.2%
30	5	5.6%

Value	Count	Frequency (%)
63	1	1.1%
57	1	1.1%
56	1	1.1%
49	1	1.1%
47	1	1.1%
44	2	2.2%
42	2	2.2%
41	4	4.5%
40	2	2.2%
39	6	6.7%

Age_Group
Categorical

Distinct	3
Distinct (%)	3.4%
Missing	0
Missing (%)	0.0%
Memory size	6.4 KiB

Length

Max length	20
Median length	14
Mean length	15.752809
Min length	11

Characters and Unicode

Total characters	1402	
Distinct characters	21	
Distinct categories	8 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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

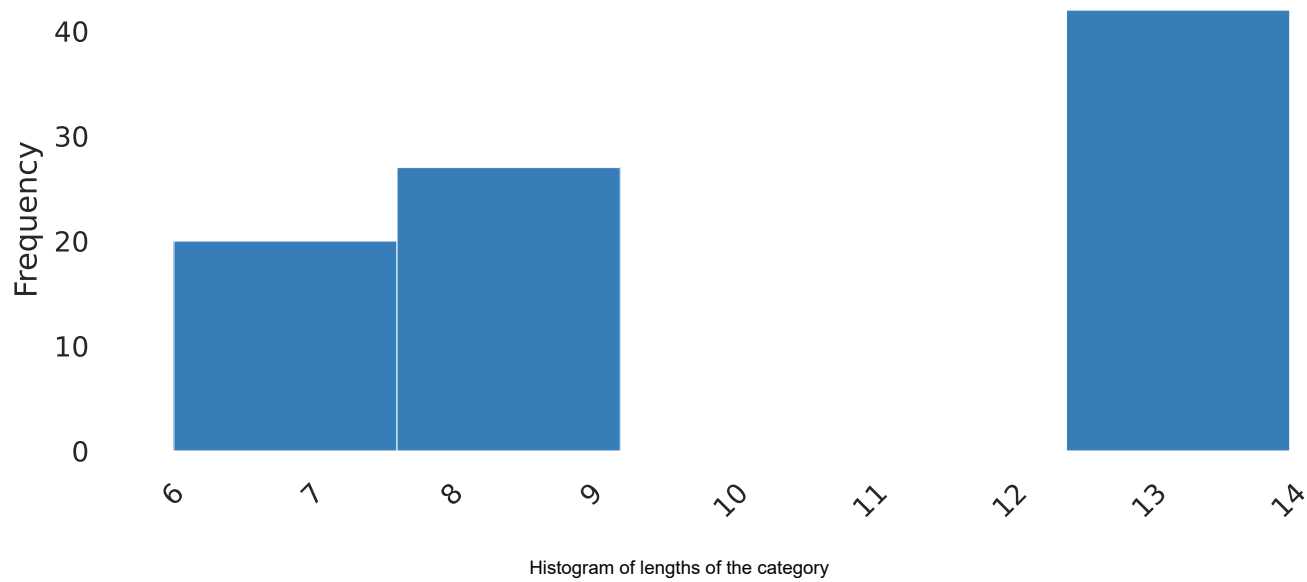|		
Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Adults (35-64)
2nd row	Adults (35-64)
3rd row	Adults (35-64)
4th row	Young Adults (25-34)
5th row	Adults (35-64)

Common Values

Value	Count	Frequency (%)
Adults (35-64)	48	53.9%
Young Adults (25-34)	31	34.8%
Youth (<25)	10	11.2%

Length
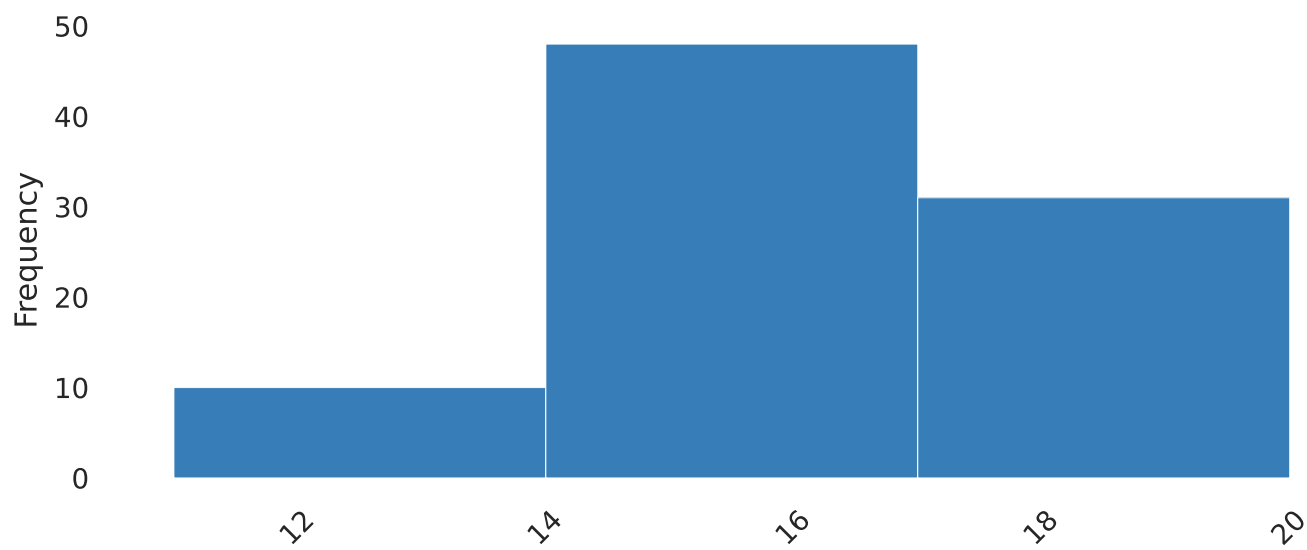
Histogram of lengths of the category

Common Values (Plot)
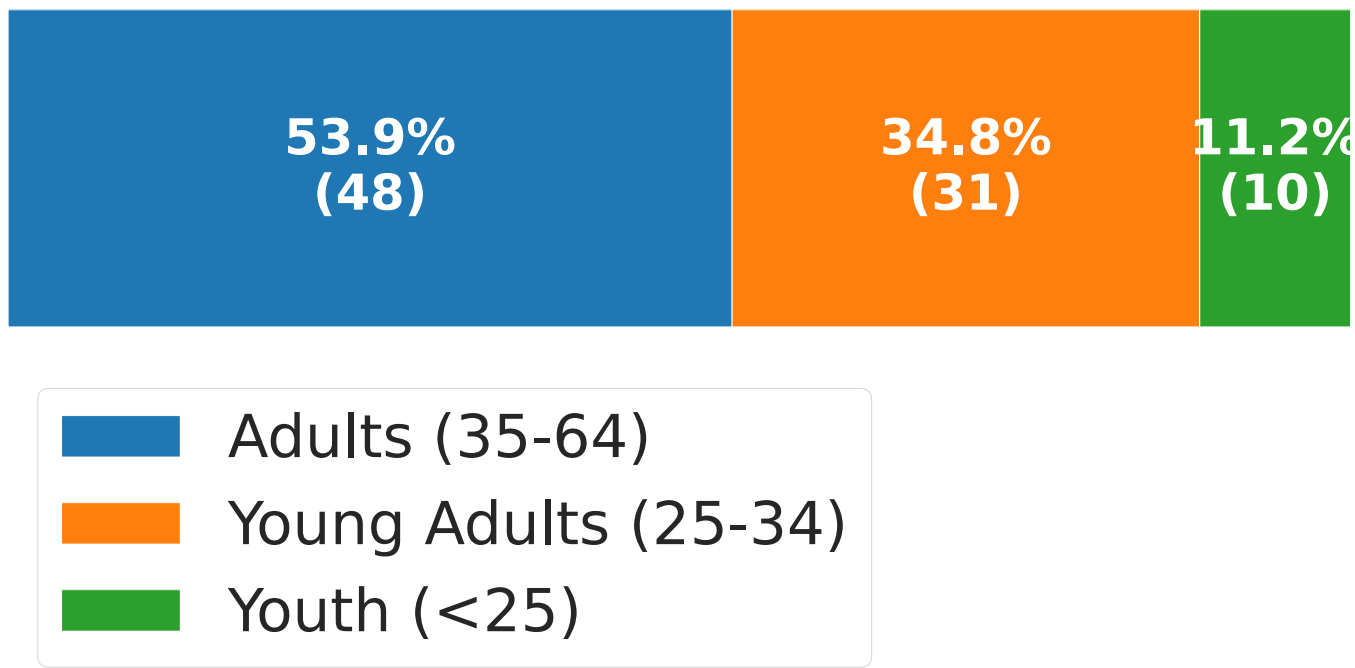
Value	Count	Frequency (%)
adults	79	37.8%
35-64	48	23.0%
young	31	14.8%
25-34	31	14.8%
youth	10	4.8%
25	10	4.8%

Most occurring characters

Value	Count	Frequency (%)
u	120	8.6%
	120	8.6%
5	89	6.3%
)	89	6.3%
t	89	6.3%
(	89	6.3%
A	79	5.6%
4	79	5.6%
d	79	5.6%
-	79	5.6%
Other values (11)	490	35.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	559	39.9%
Decimal Number	336	24.0%
Space Separator	120	8.6%
Uppercase Letter	120	8.6%
Close Punctuation	89	6.3%
Open Punctuation	89	6.3%
Dash Punctuation	79	5.6%
Math Symbol	10	0.7%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
u	120	21.5%
t	89	15.9%
d	79	14.1%
s	79	14.1%
l	79	14.1%
o	41	7.3%
n	31	5.5%
g	31	5.5%
h	10	1.8%

Decimal Number

Value	Count	Frequency (%)
5	89	26.5%
4	79	23.5%
3	79	23.5%
6	48	14.3%
2	41	12.2%

Uppercase Letter

Value	Count	Frequency (%)
A	79	65.8%
Y	41	34.2%

Space Separator

Value	Count	Frequency (%)
	120	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	89	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	89	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	79	100.0%

Math Symbol

Value	Count	Frequency (%)
<	10	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	723	51.6%
Latin	679	48.4%

Most frequent character per script

Latin

Value	Count	Frequency (%)
u	120	17.7%
t	89	13.1%
A	79	11.6%
d	79	11.6%
s	79	11.6%
l	79	11.6%
Y	41	6.0%
o	41	6.0%
n	31	4.6%
g	31	4.6%

Common

Value	Count	Frequency (%)
	120	16.6%
5	89	12.3%
)	89	12.3%
(	89	12.3%
4	79	10.9%
-	79	10.9%
3	79	10.9%
6	48	6.6%
2	41	5.7%
<	10	1.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1402	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
u	120	8.6%
	120	8.6%
5	89	6.3%
)	89	6.3%
t	89	6.3%
(	89	6.3%
A	79	5.6%
4	79	5.6%
d	79	5.6%
-	79	5.6%
Other values (11)	490	35.0%

Customer_Gender
Categorical

Distinct	2
Distinct (%)	2.2%
Missing	0
Missing (%)	0.0%
Memory size	5.2 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	89	
Distinct characters	2	
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	F
2nd row	M
3rd row	M
4th row	F
5th row	F

Common Values

Value	Count	Frequency (%)
F	50	56.2%
M	39	43.8%

Length
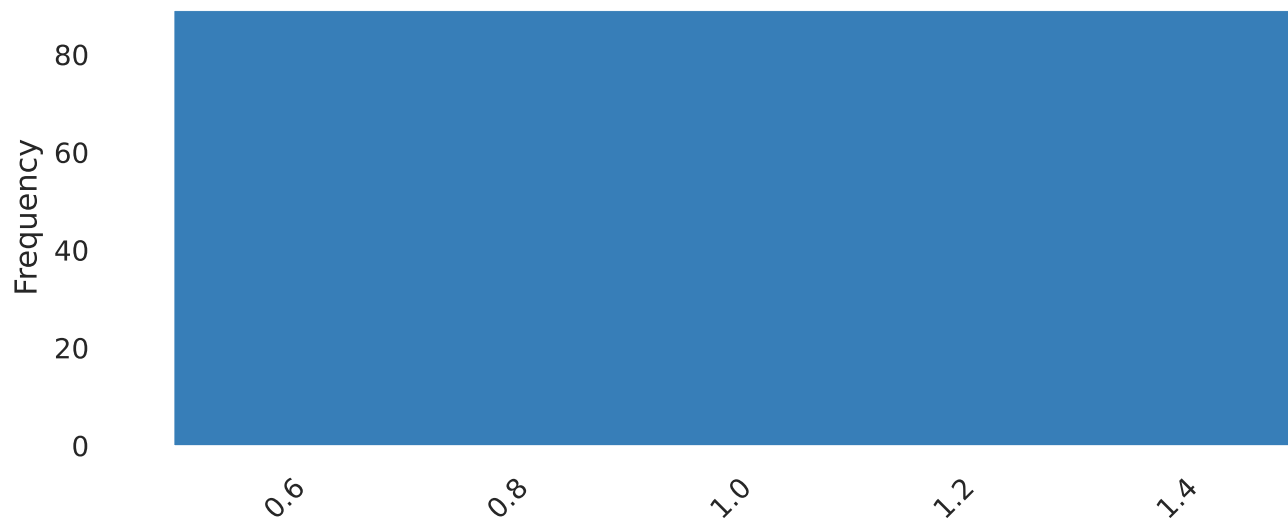
Histogram of lengths of the category

Common Values (Plot)

56.2%
(50)

43.8%
(39)

F
M

Value	Count	Frequency (%)
f	50	56.2%
m	39	43.8%

Most occurring characters

Value	Count	Frequency (%)
F	50	56.2%
M	39	43.8%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	89	100.0%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
F	50	56.2%
M	39	43.8%

Most occurring scripts

Value	Count	Frequency (%)
Latin	89	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
F	50	56.2%
M	39	43.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	89	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
F	50	56.2%
M	39	43.8%

Country
Categorical

Distinct	9
Distinct (%)	10.1%
Missing	0
Missing (%)	0.0%
Memory size	6.0 KiB

Length

Max length	14
Median length	13
Mean length	10.41573
Min length	6

Characters and Unicode

Total characters	927	
Distinct characters	23	
Distinct categories	3 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	3	?
Unique (%)	3.4%	

Sample

1st row	United States
2nd row	United Kingdom
3rd row	United States
4th row	Australia
5th row	United States

Common Values

Value	Count	Frequency (%)
United States	30	33.7%
Australia	27	30.3%
United Kingdom	9	10.1%
France	8	9.0%
Germany	6	6.7%
Canada	6	6.7%
United States	1	1.1%
United States	1	1.1%
United States	1	1.1%

Length

Histogram of lengths of the category

Common Values (Plot)

**33.7%
(30)**

**30.3%
(27)**

**10.1%
(9)**

**9.0%
(8)**
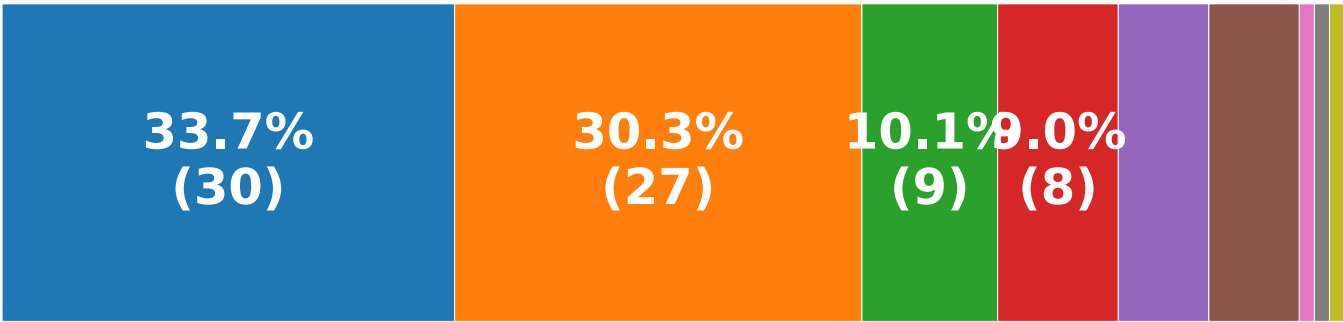
- United States
- Australia
- United Kingdom
- France
- Germany
- Canada
- United States
- United States
- United States

Value	Count	Frequency (%)
united	42	32.1%
states	33	25.2%
australia	27	20.6%
kingdom	9	6.9%
france	8	6.1%
germany	6	4.6%
canada	6	4.6%

Most occurring characters

Value	Count	Frequency (%)
t	135	14.6%
a	119	12.8%
e	89	9.6%
i	78	8.4%
n	71	7.7%
s	60	6.5%
d	57	6.1%
	45	4.9%
U	42	4.5%
r	41	4.4%
Other values (13)	190	20.5%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	751	81.0%
Uppercase Letter	131	14.1%
Space Separator	45	4.9%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
t	135	18.0%
a	119	15.8%
e	89	11.9%
i	78	10.4%
n	71	9.5%
s	60	8.0%
d	57	7.6%
r	41	5.5%
l	27	3.6%
u	27	3.6%
Other values (5)	47	6.3%

Uppercase Letter

Value	Count	Frequency (%)
U	42	32.1%
S	33	25.2%
A	27	20.6%
K	9	6.9%
F	8	6.1%
G	6	4.6%
C	6	4.6%

Space Separator

Value	Count	Frequency (%)
	45	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	882	95.1%

Value	Count	Frequency (%)
Common	45	4.9%

Most frequent character per script

Latin

Value	Count	Frequency (%)
t	135	15.3%
a	119	13.5%
e	89	10.1%
i	78	8.8%
n	71	8.0%
s	60	6.8%
d	57	6.5%
U	42	4.8%
r	41	4.6%
S	33	3.7%
Other values (12)	157	17.8%

Common

Value	Count	Frequency (%)
	45	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	927	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
t	135	14.6%
a	119	12.8%
e	89	9.6%
i	78	8.4%
n	71	7.7%
s	60	6.5%
d	57	6.1%
	45	4.9%
U	42	4.5%
r	41	4.4%
Other values (13)	190	20.5%

State
Categorical

Distinct	17
Distinct (%)	19.1%
Missing	0
Missing (%)	0.0%
Memory size	6.0 KiB

Length

Max length	19
Median length	17
Mean length	10.94382
Min length	4

Characters and Unicode

Total characters	974	
Distinct characters	36	
Distinct categories	6 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	6	?
Unique (%)	6.7%	

Sample

1st row	California
2nd row	England
3rd row	California
4th row	New South Wales
5th row	California

Common Values

Value	Count	Frequency (%)
California	20	22.5%
New South Wales	14	15.7%
Washington	9	10.1%
England	9	10.1%
British Columbia	6	6.7%
Victoria	6	6.7%
Queensland	6	6.7%
Oregon	4	4.5%
Seine (Paris)	4	4.5%
Nordrhein-Westfalen	3	3.4%
Other values (7)	8	9.0%

Length

Histogram of lengths of the category
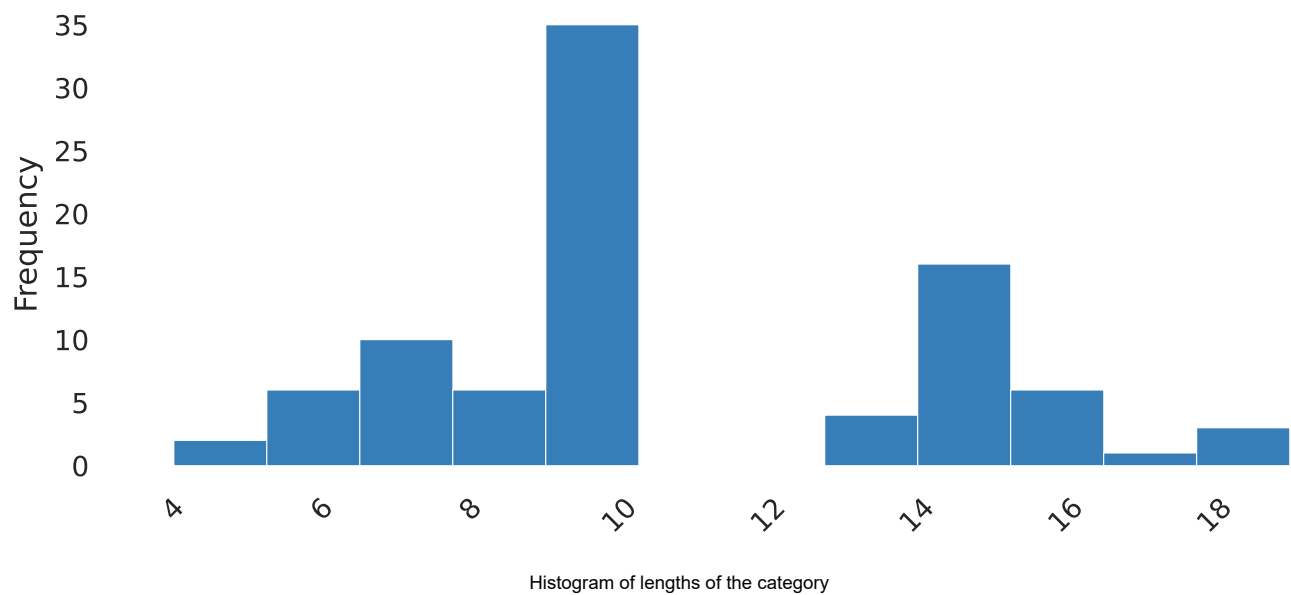
Value	Count	Frequency (%)
california	20	15.2%
south	15	11.4%
wales	14	10.6%
new	14	10.6%
washington	9	6.8%
england	9	6.8%
victoria	6	4.5%
queensland	6	4.5%
seine	6	4.5%
columbia	6	4.5%
Other values (13)	27	20.5%

Most occurring characters

Value	Count	Frequency (%)
a	102	10.5%
i	95	9.8%
n	89	9.1%
e	73	7.5%
o	65	6.7%
l	59	6.1%
r	50	5.1%
s	48	4.9%
	43	4.4%
t	42	4.3%
Other values (26)	308	31.6%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	786	80.7%
Uppercase Letter	134	13.8%
Space Separator	43	4.4%
Open Punctuation	4	0.4%
Close Punctuation	4	0.4%
Dash Punctuation	3	0.3%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	102	13.0%
i	95	12.1%
n	89	11.3%
e	73	9.3%
o	65	8.3%
l	59	7.5%
r	50	6.4%
s	48	6.1%
t	42	5.3%
h	33	4.2%
Other values (8)	130	16.5%

Uppercase Letter

Value	Count	Frequency (%)
C	26	19.4%
W	26	19.4%
S	23	17.2%
N	18	13.4%
E	9	6.7%
Q	6	4.5%
V	6	4.5%
B	6	4.5%
O	4	3.0%
P	4	3.0%
Other values (4)	6	4.5%

Space Separator

Value	Count	Frequency (%)
	43	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	4	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	4	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	3	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	920	94.5%
Common	54	5.5%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	102	11.1%
i	95	10.3%
n	89	9.7%
e	73	7.9%
o	65	7.1%
l	59	6.4%
r	50	5.4%
s	48	5.2%
t	42	4.6%
h	33	3.6%
Other values (22)	264	28.7%

Common

Value	Count	Frequency (%)
	43	79.6%
(	4	7.4%
)	4	7.4%
-	3	5.6%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	974	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
a	102	10.5%
i	95	9.8%
n	89	9.1%
e	73	7.5%
o	65	6.7%
l	59	6.1%
r	50	5.1%
s	48	4.9%
	43	4.4%
t	42	4.3%
Other values (26)	308	31.6%

Product_Category
Categorical

Distinct	1
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Memory size	5.5 KiB

Length

Max length	5
Median length	5
Mean length	5
Min length	5

Characters and Unicode

Total characters	445	
Distinct characters	5	
Distinct categories	2 (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	Bikes
2nd row	Bikes
3rd row	Bikes
4th row	Bikes
5th row	Bikes

Common Values

Value	Count	Frequency (%)
Bikes	89	100.0%

Length
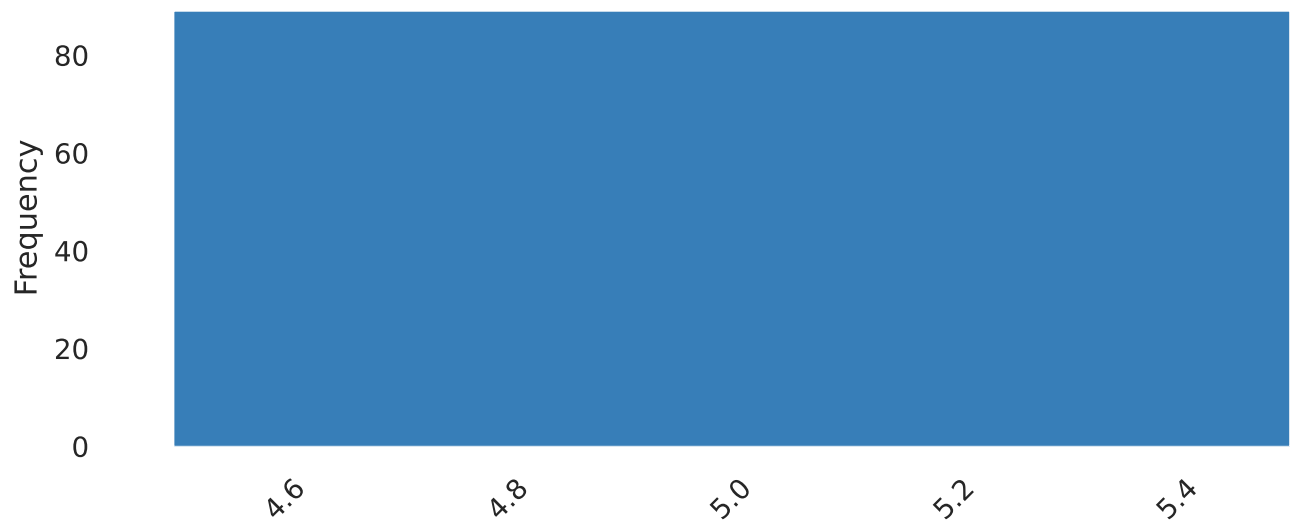
Histogram of lengths of the category

Common Values (Plot)

100.0%
(89)

Bikes

Value	Count	Frequency (%)
bikes	89	100.0%

Most occurring characters

Value	Count	Frequency (%)
B	89	20.0%
i	89	20.0%
k	89	20.0%
e	89	20.0%
s	89	20.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	356	80.0%
Uppercase Letter	89	20.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	89	25.0%
k	89	25.0%
e	89	25.0%
s	89	25.0%

Uppercase Letter

Value	Count	Frequency (%)
B	89	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	445	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
B	89	20.0%
i	89	20.0%
k	89	20.0%
e	89	20.0%
s	89	20.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	445	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
B	89	20.0%
i	89	20.0%
k	89	20.0%
e	89	20.0%
s	89	20.0%

Sub_Category
Categorical

Distinct	1
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Memory size	6.3 KiB

Length

Max length	14
Median length	14
Mean length	14
Min length	14

Characters and Unicode

Total characters	1246	
Distinct characters	12	
Distinct categories	3 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	Mountain Bikes
2nd row	Mountain Bikes
3rd row	Mountain Bikes
4th row	Mountain Bikes
5th row	Mountain Bikes

Common Values

Value	Count	Frequency (%)
Mountain Bikes	89	100.0%

Length

Histogram of lengths of the category

Common Values (Plot)

100.0%
(89)
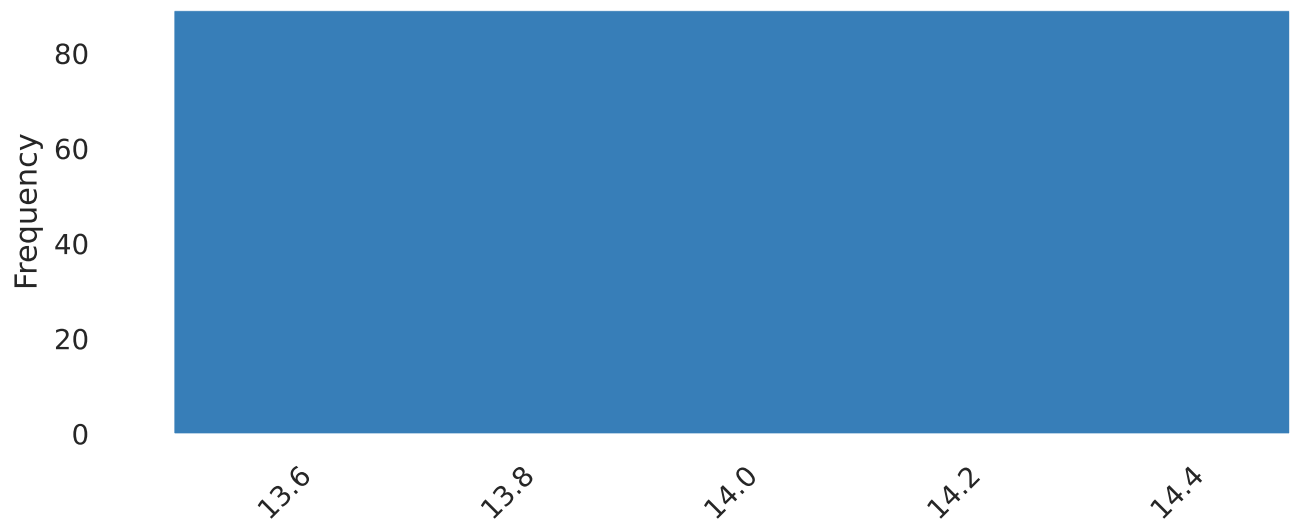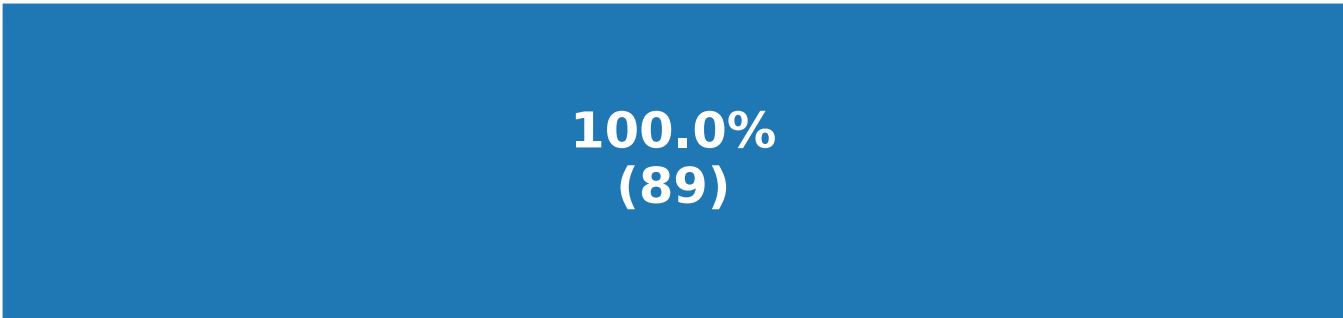
Mountain Bikes

Value	Count	Frequency (%)
mountain	89	50.0%
bikes	89	50.0%

Most occurring characters

Value	Count	Frequency (%)
n	178	14.3%
i	178	14.3%
M	89	7.1%
o	89	7.1%
u	89	7.1%
t	89	7.1%
a	89	7.1%
	89	7.1%
B	89	7.1%
k	89	7.1%
Other values (2)	178	14.3%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	979	78.6%
Uppercase Letter	178	14.3%
Space Separator	89	7.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
n	178	18.2%
i	178	18.2%
o	89	9.1%
u	89	9.1%
t	89	9.1%
a	89	9.1%
k	89	9.1%
e	89	9.1%
s	89	9.1%

Uppercase Letter

Value	Count	Frequency (%)
M	89	50.0%
B	89	50.0%

Space Separator

Value	Count	Frequency (%)
	89	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1157	92.9%
Common	89	7.1%

Most frequent character per script

Latin

Value	Count	Frequency (%)
n	178	15.4%
i	178	15.4%
M	89	7.7%
o	89	7.7%
u	89	7.7%
t	89	7.7%
a	89	7.7%
B	89	7.7%
k	89	7.7%
e	89	7.7%

Common

Value	Count	Frequency (%)
	89	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1246	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
n	178	14.3%
i	178	14.3%
M	89	7.1%
o	89	7.1%
u	89	7.1%
t	89	7.1%
a	89	7.1%
	89	7.1%
B	89	7.1%
k	89	7.1%
Other values (2)	178	14.3%

Product_Description

Categorical

Distinct	19
Distinct (%)	21.3%
Missing	0
Missing (%)	0.0%
Memory size	7.0 KiB

Length

Max length	25
Median length	23
Mean length	22.595506
Min length	7

Characters and Unicode

Total characters	2011	
Distinct characters	29	
Distinct categories	6 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	6	?
Unique (%)	6.7%	

Sample

1st row	Mountain-200 Black, 46
2nd row	Mountain-200 Silver, 42
3rd row	Mountain-400-W Silver, 46
4th row	Mountain-400-W Silver, 42
5th row	Mountain-200 Black, 46

Common Values

Value	Count	Frequency (%)
Mountain-200 Black, 46	15	16.9%
Mountain-200 Silver, 38	14	15.7%
Mountain-200 Black, 38	13	14.6%
Mountain-200 Silver, 42	9	10.1%
Mountain-200 Black, 42	7	7.9%
Mountain-400-W Silver, 46	6	6.7%
Mountain-200 Silver, 46	4	4.5%
Mountain-400-W Silver, 42	4	4.5%
Mountain-500 Silver, 42	3	3.4%
Mountain-400-W Silver, 38	2	2.2%
Other values (9)	12	13.5%

Length

Histogram of lengths of the category
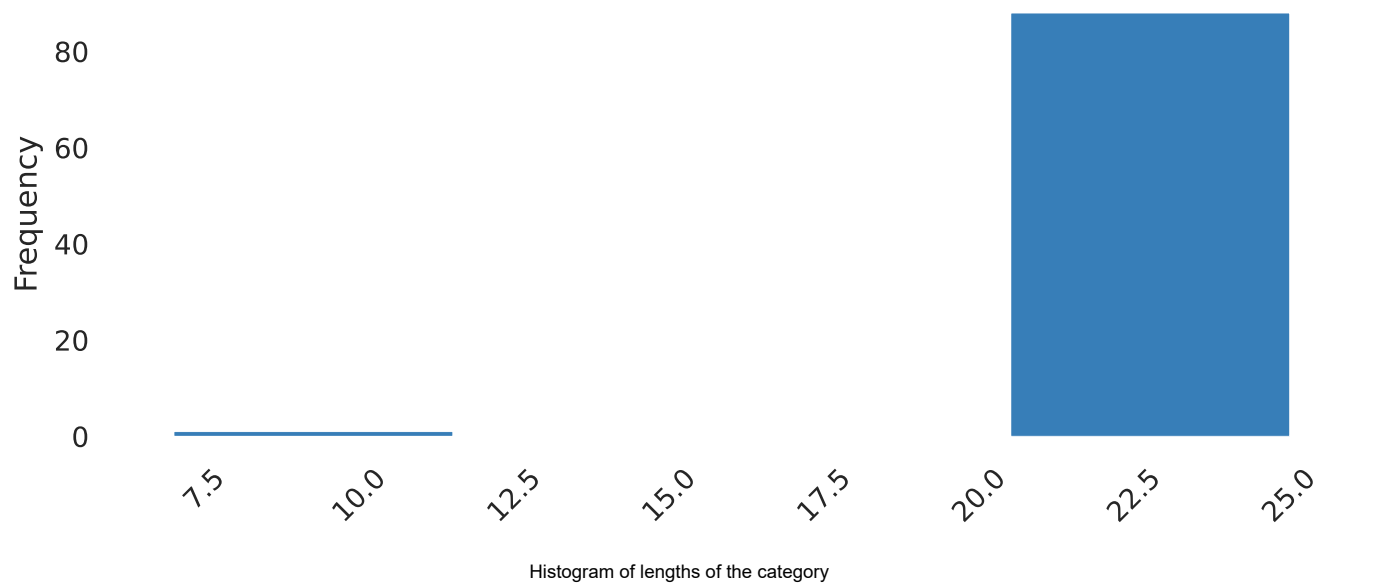
Value	Count	Frequency (%)
mountain-200	62	23.4%
black	44	16.6%
silver	44	16.6%
38	31	11.7%
46	25	9.4%
42	25	9.4%
mountain-400-w	12	4.5%
mountain-500	10	3.8%
mountain-100	4	1.5%
40	3	1.1%
Other values (4)	5	1.9%

Most occurring characters

Value	Count	Frequency (%)
n	179	8.9%
0	179	8.9%
	176	8.8%
a	132	6.6%
i	132	6.6%
-	100	5.0%
o	89	4.4%
M	88	4.4%
u	88	4.4%
t	88	4.4%
Other values (19)	760	37.8%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1018	50.6%
Decimal Number	440	21.9%
Uppercase Letter	189	9.4%
Space Separator	176	8.8%
Dash Punctuation	100	5.0%
Other Punctuation	88	4.4%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
n	179	17.6%
a	132	13.0%
i	132	13.0%
o	89	8.7%
u	88	8.6%
t	88	8.6%
l	88	8.6%
k	45	4.4%
c	44	4.3%
v	44	4.3%
Other values (3)	89	8.7%

Decimal Number

Value	Count	Frequency (%)
0	179	40.7%
2	88	20.0%
4	70	15.9%
8	32	7.3%
3	31	7.0%
6	25	5.7%
5	11	2.5%
1	4	0.9%

Uppercase Letter

Value	Count	Frequency (%)
M	88	46.6%
B	44	23.3%

Value	Count	Frequency (%)
S	44	23.3%
W	12	6.3%
U	1	0.5%

Space Separator

Value	Count	Frequency (%)
	176	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	100	100.0%

Other Punctuation

Value	Count	Frequency (%)
,	88	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1207	60.0%
Common	804	40.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
n	179	14.8%
a	132	10.9%
i	132	10.9%
o	89	7.4%
M	88	7.3%
u	88	7.3%
t	88	7.3%
l	88	7.3%
k	45	3.7%
c	44	3.6%
Other values (8)	234	19.4%

Common

Value	Count	Frequency (%)
0	179	22.3%
	176	21.9%
-	100	12.4%
2	88	10.9%
,	88	10.9%
4	70	8.7%
8	32	4.0%
3	31	3.9%
6	25	3.1%
5	11	1.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	2011	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
n	179	8.9%
0	179	8.9%
	176	8.8%
a	132	6.6%
i	132	6.6%
-	100	5.0%
o	89	4.4%
M	88	4.4%
u	88	4.4%
t	88	4.4%
Other values (19)	760	37.8%

Order_Quantity
Categorical

Distinct	4
Distinct (%)	4.5%
Missing	0
Missing (%)	0.0%
Memory size	5.2 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	89	
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	4
2nd row	1
3rd row	2
4th row	1
5th row	2

Common Values

Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Length
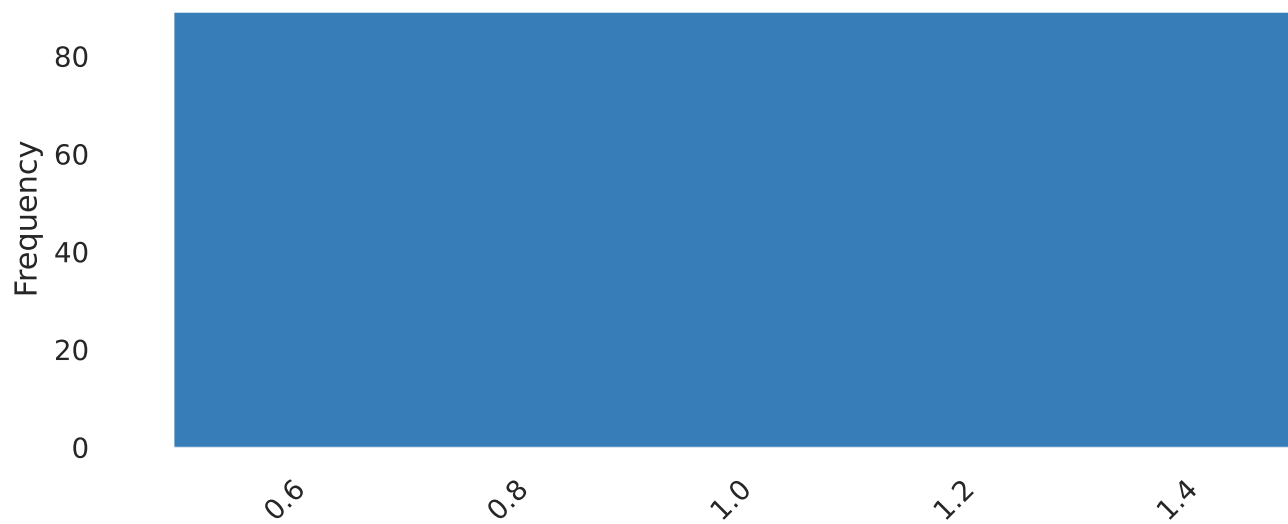
Histogram of lengths of the category

Common Values (Plot)
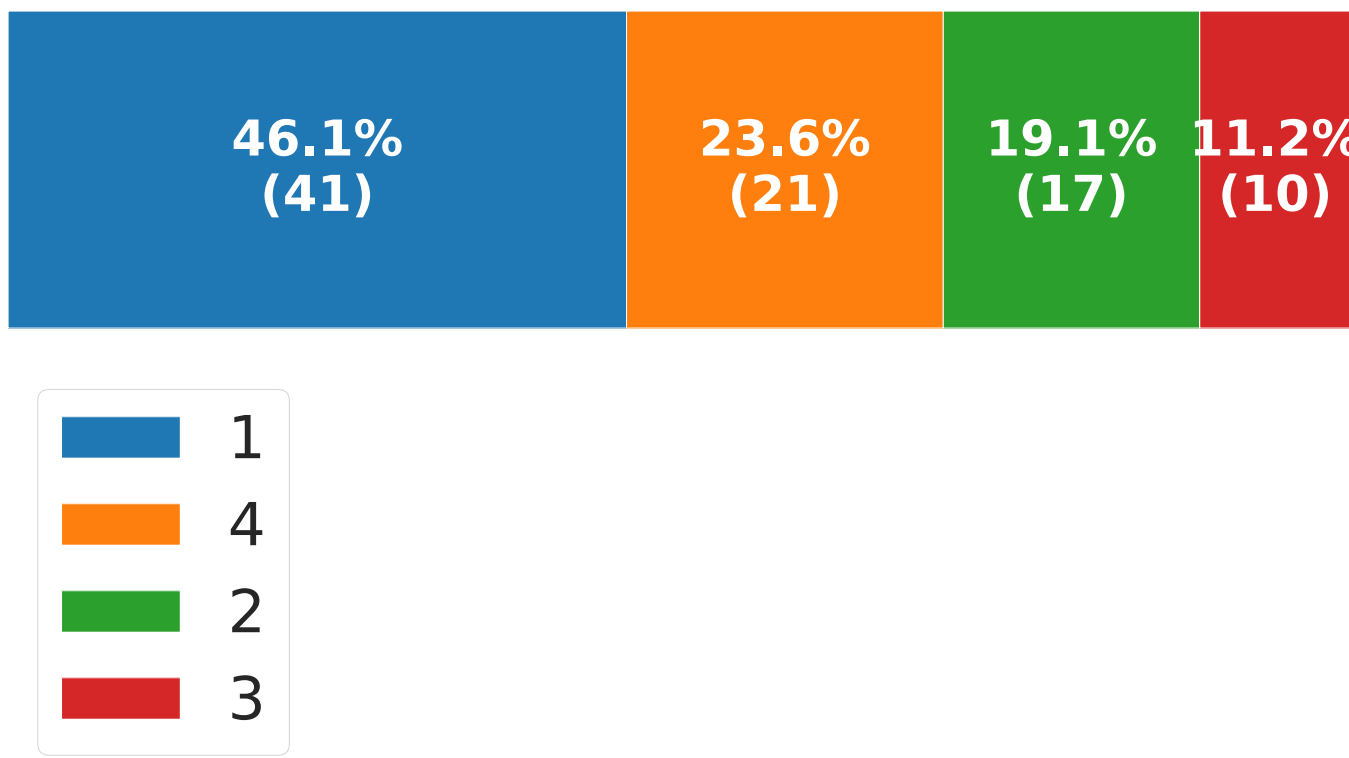
Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Most occurring characters

Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	89	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Most occurring scripts

Value	Count	Frequency (%)
Common	89	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	89	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
1	41	46.1%
4	21	23.6%
2	17	19.1%
3	10	11.2%

Unit_Cost
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 5 fields: Unit Price , Profit , Cost , Revenue, Product Description)   ZEROS

Distinct	8
Distinct (%)	9.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1052.2472

Minimum	0
Maximum	1912
Zeros	1
Zeros (%)	1.1%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
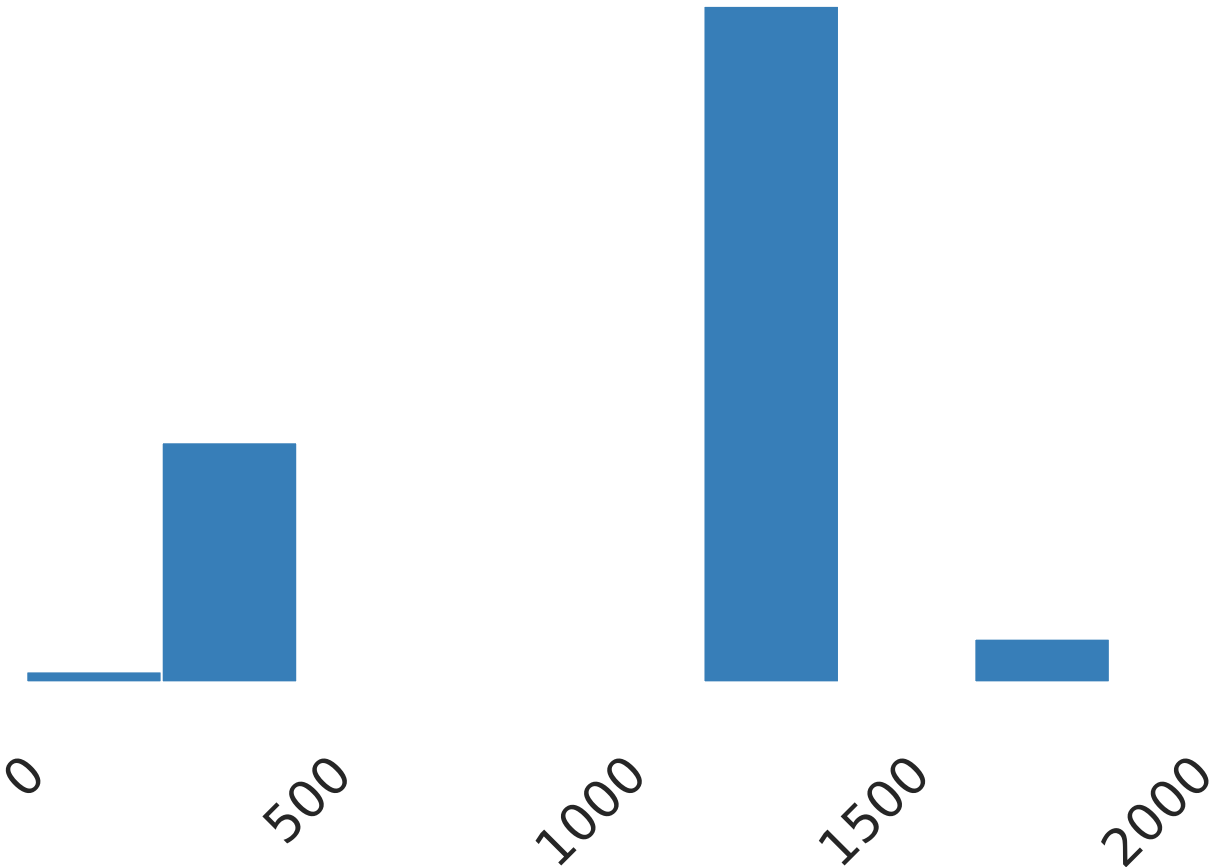
Quantile statistics

Minimum	0
5-th percentile	295
Q1	420
median	1252
Q3	1266
95-th percentile	1266
Maximum	1912
Range	1912
Interquartile range (IQR)	846

Descriptive statistics

Standard deviation	440.499
Coefficient of variation (CV)	0.41862692
Kurtosis	-0.34666154
Mean	1052.2472
Median Absolute Deviation (MAD)	14
Skewness	-0.74246303
Sum	93650
Variance	194039.37
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=8)
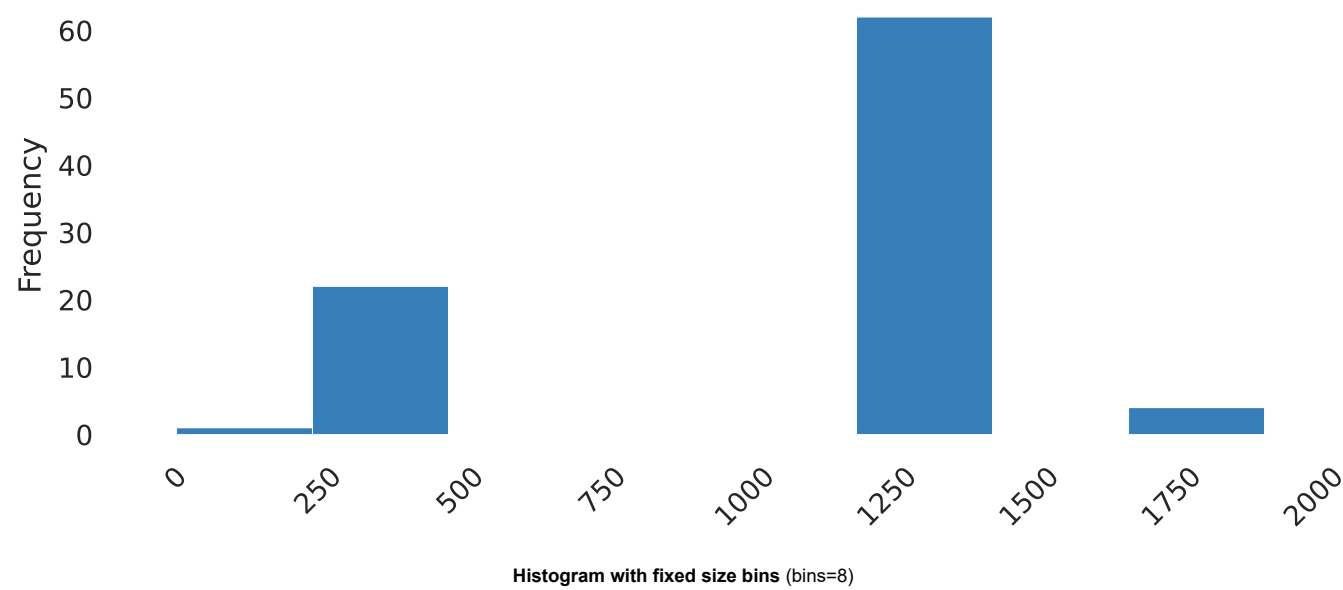
Value	Count	Frequency (%)
1252	35	39.3%
1266	27	30.3%
420	12	13.5%
295	6	6.7%
308	4	4.5%
1898	3	3.4%
0	1	1.1%
1912	1	1.1%

Value	Count	Frequency (%)
0	1	1.1%
295	6	6.7%
308	4	4.5%
420	12	13.5%
1252	35	39.3%
1266	27	30.3%
1898	3	3.4%
1912	1	1.1%

DATA ANALYSIS REPORT

Value	Count	Frequency (%)
1912	1	1.1%
1898	3	3.4%
1266	27	30.3%
1252	35	39.3%
420	12	13.5%
308	4	4.5%
295	6	6.7%
0	1	1.1%

Unit_Price
Real number ($\mathbb{R}$)

HIGH CORRELATION (This variable has a high overall correlation with 5 fields: Unit Cost , Profit , Cost , Revenue, Product Description)   ZEROS

Distinct	8
Distinct (%)	9.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1940.9438

Minimum	0
Maximum	3400
Zeros	1
Zeros (%)	1.1%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
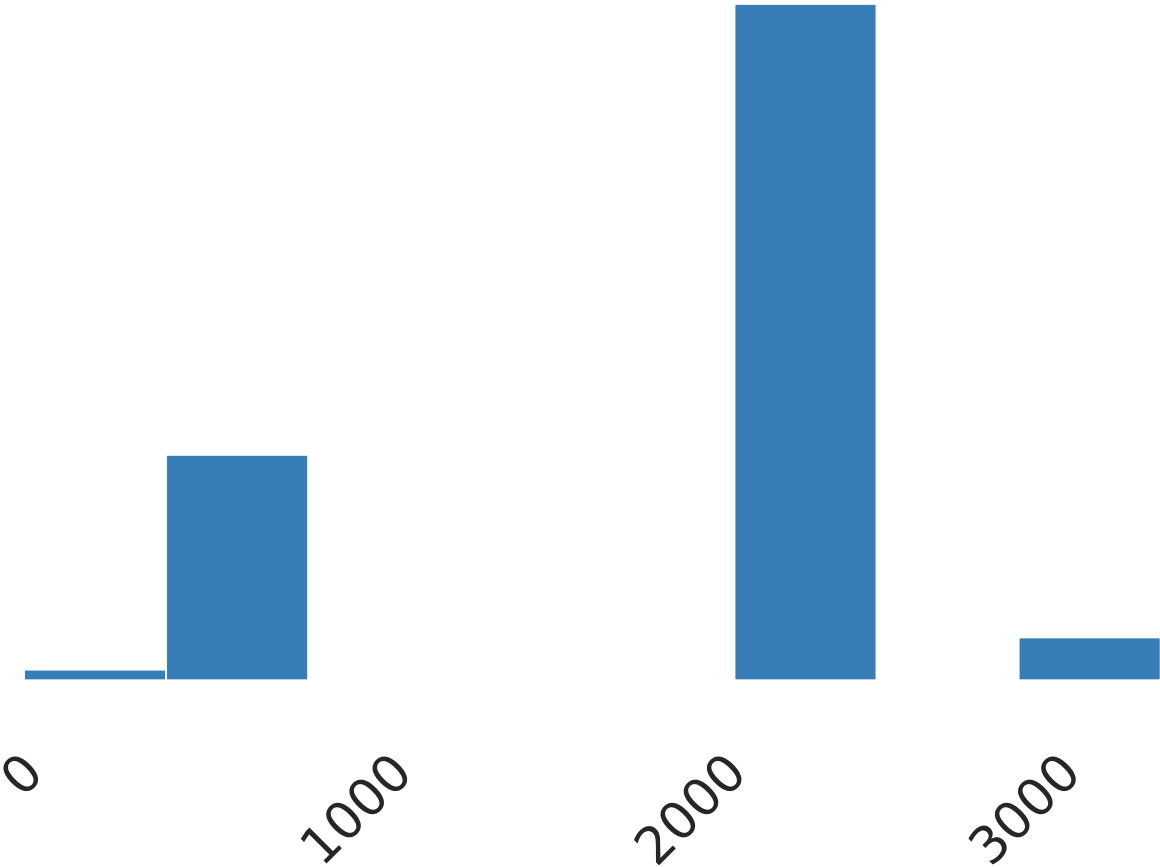
Quantile statistics

Minimum	0
5-th percentile	540
Q1	2295
median	2295
Q3	2320
95-th percentile	2320
Maximum	3400
Range	3400
Interquartile range (IQR)	25

Descriptive statistics

Standard deviation	789.95574
Coefficient of variation (CV)	0.40699568
Kurtosis	-0.27044962
Mean	1940.9438
Median Absolute Deviation (MAD)	25
Skewness	-0.87048122
Sum	172744
Variance	624030.08
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=8)
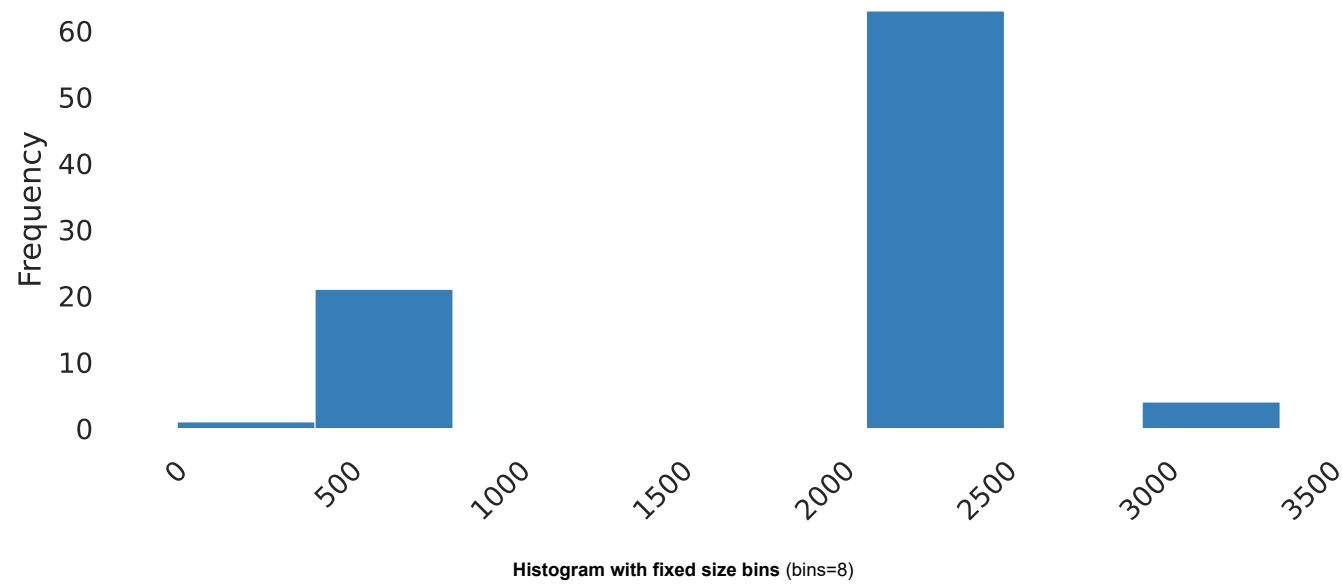
Value	Count	Frequency (%)
2295	36	40.4%
2320	27	30.3%
769	11	12.4%
540	6	6.7%
565	4	4.5%
3375	3	3.4%
0	1	1.1%
3400	1	1.1%

Value	Count	Frequency (%)
0	1	1.1%
540	6	6.7%
565	4	4.5%
769	11	12.4%
2295	36	40.4%
2320	27	30.3%
3375	3	3.4%
3400	1	1.1%

Value	Count	Frequency (%)
3400	1	1.1%
3375	3	3.4%
2320	27	30.3%
2295	36	40.4%
769	11	12.4%
565	4	4.5%
540	6	6.7%
0	1	1.1%

Profit
Real number ($\mathbb{R}$)

Distinct	21
Distinct (%)	23.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1848.8876
Minimum	245
Maximum	5908
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
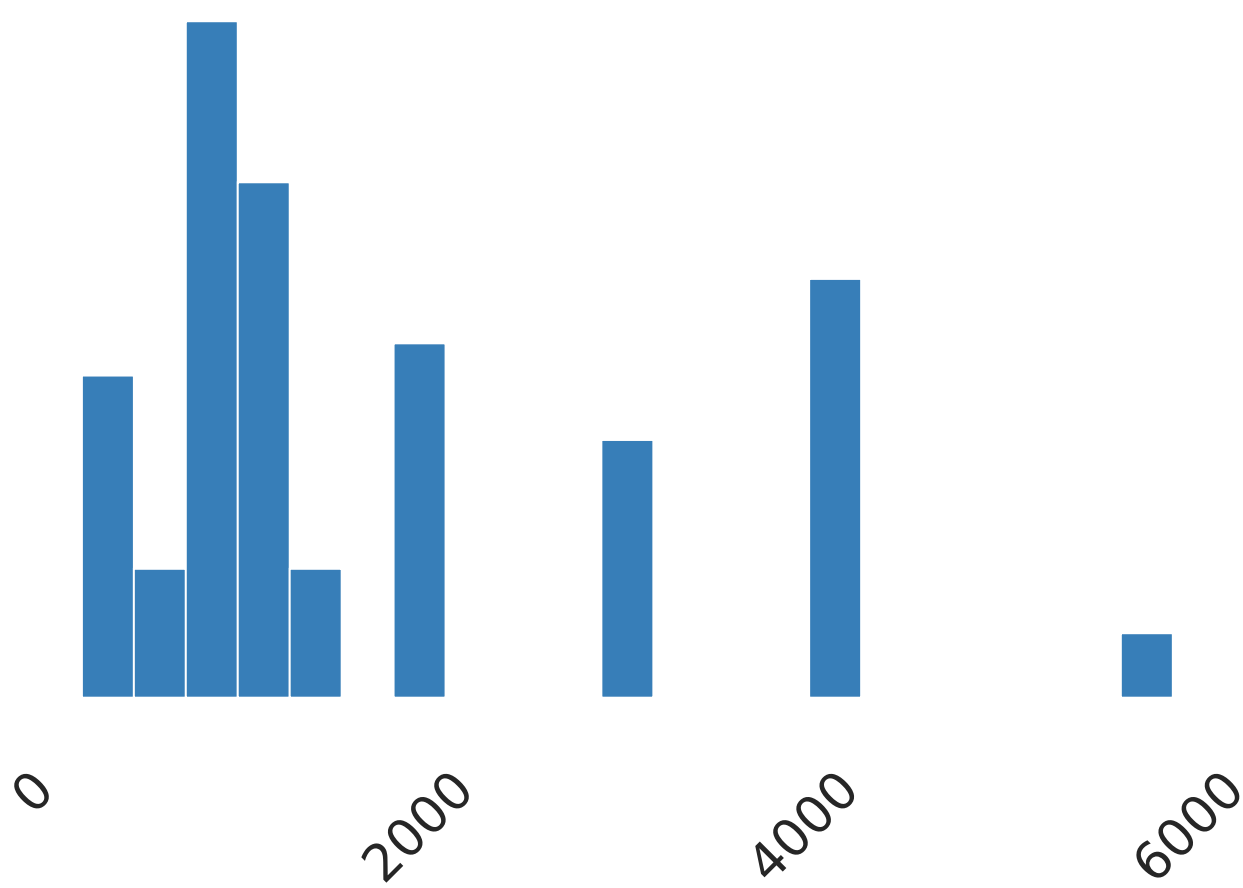
Quantile statistics

Minimum	245
5-th percentile	293.8
Q1	1043
median	1054
Q3	2954
95-th percentile	4216
Maximum	5908
Range	5663
Interquartile range (IQR)	1911

Descriptive statistics

Standard deviation	1394.4012
Coefficient of variation (CV)	0.75418383
Kurtosis	0.2122731
Mean	1848.8876
Median Absolute Deviation (MAD)	434
Skewness	1.0915287
Sum	164551
Variance	1944354.6
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=21)
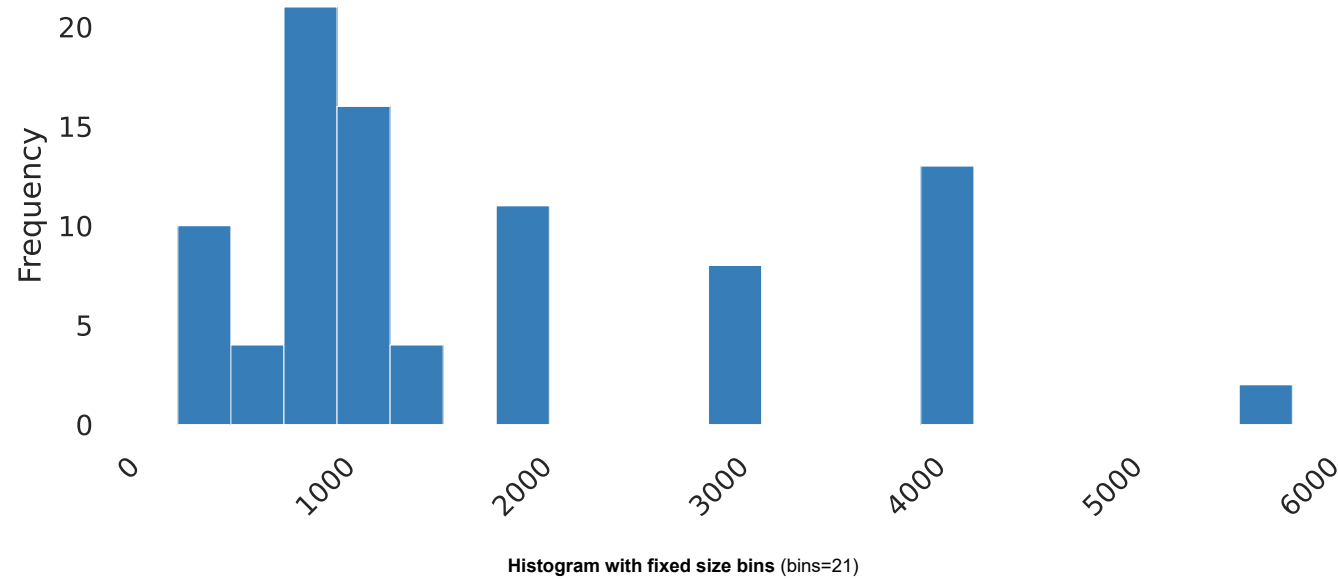
Value	Count	Frequency (%)
1043	16	18.0%
1054	16	18.0%
4172	9	10.1%
2086	8	9.0%
3162	4	4.5%
4216	4	4.5%
349	4	4.5%
1396	3	3.4%
2108	3	3.4%
245	3	3.4%
Other values (11)	19	21.3%

Value	Count	Frequency (%)
245	3	3.4%
257	2	2.2%
349	4	4.5%
490	1	1.1%
698	3	3.4%
735	1	1.1%
980	1	1.1%
1028	2	2.2%
1043	16	18.0%
1047	2	2.2%

Value	Count	Frequency (%)
5908	2	2.2%
4216	4	4.5%
4172	9	10.1%
3162	4	4.5%
3129	3	3.4%
2954	1	1.1%
2108	3	3.4%
2086	8	9.0%
1488	1	1.1%
1396	3	3.4%

Cost
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 5 fields: Unit Cost , Unit Price , Profit , Revenue, Order Quantity)   ZEROS

Distinct	22
Distinct (%)	24.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2204.236
Minimum	0
Maximum	7592
Zeros	2
Zeros (%)	2.2%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
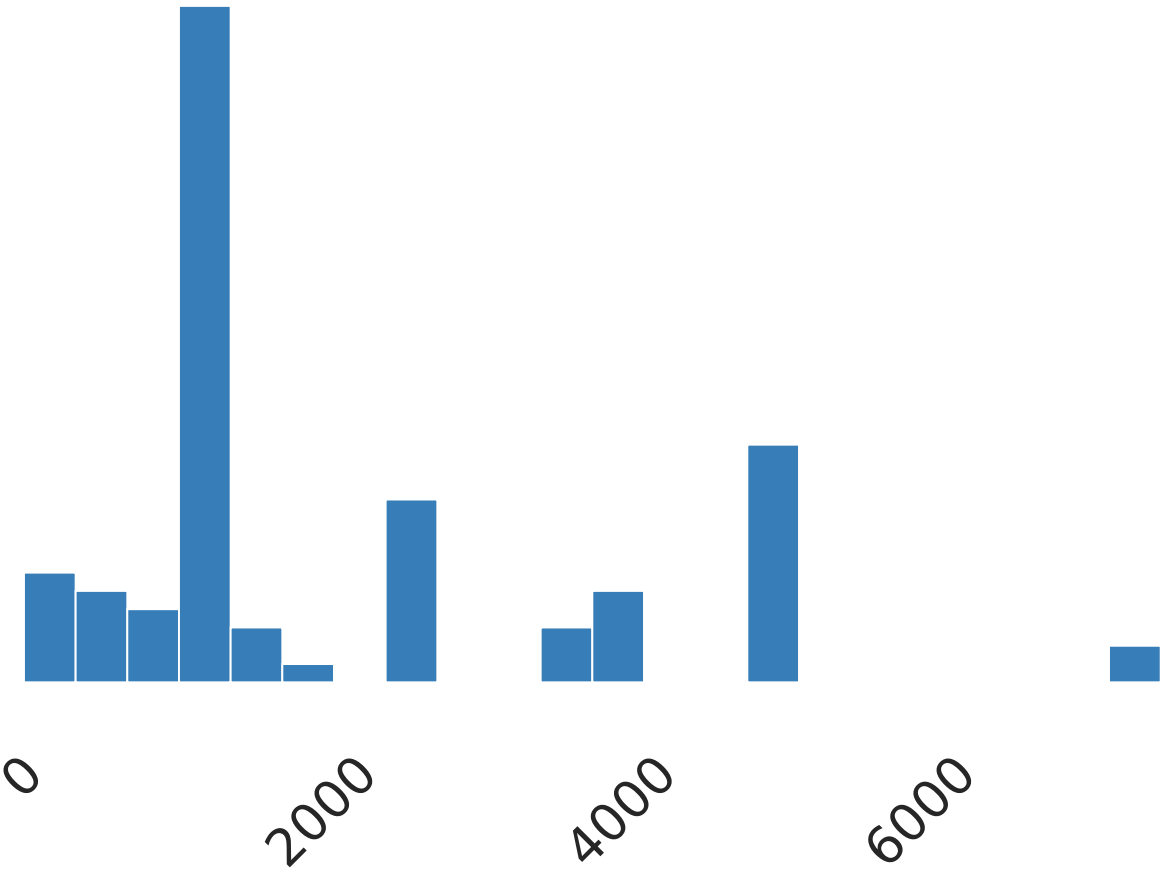
Quantile statistics

Minimum	0
5-th percentile	308
Q1	1252
median	1266
Q3	3756
95-th percentile	5064
Maximum	7592
Range	7592
Interquartile range (IQR)	2504

Descriptive statistics

Standard deviation	1730.3309
Coefficient of variation (CV)	0.78500259
Kurtosis	0.6260944
Mean	2204.236
Median Absolute Deviation (MAD)	646
Skewness	1.1635479
Sum	196177
Variance	2994045.1
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=22)

DATA ANALYSIS REPORT
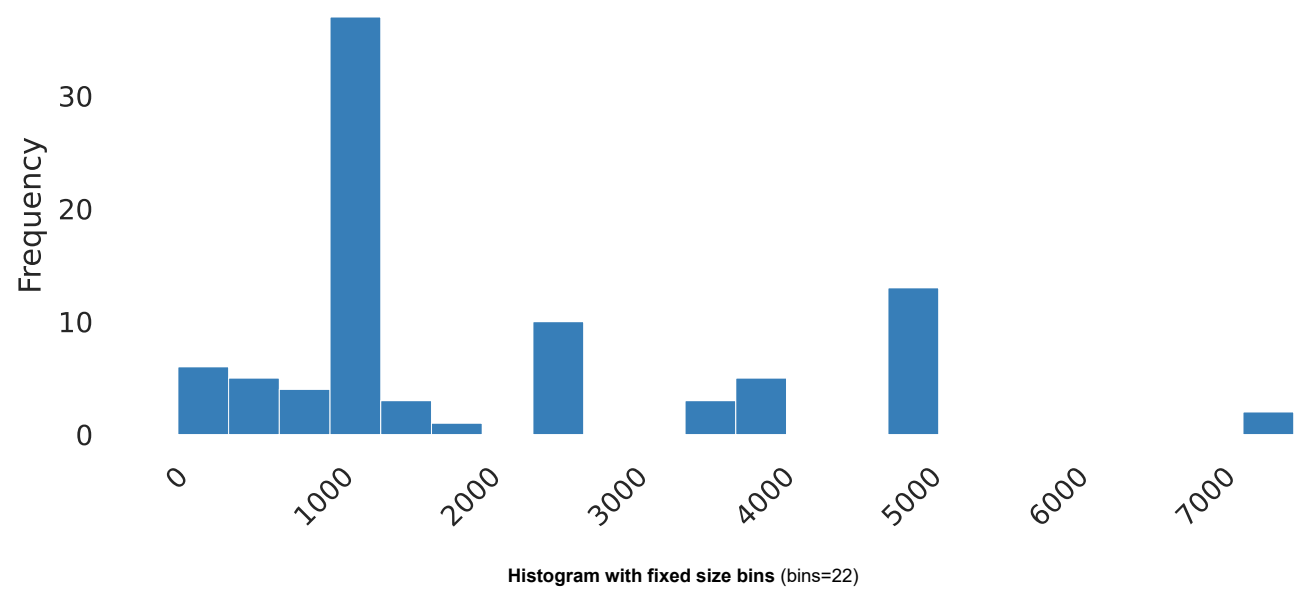
Value	Count	Frequency (%)
1252	16	18.0%
1266	16	18.0%
5008	9	10.1%
2504	7	7.9%
5064	4	4.5%
3798	4	4.5%
420	4	4.5%
3756	3	3.4%
1680	3	3.4%
2532	3	3.4%
Other values (12)	20	22.5%

Value	Count	Frequency (%)
0	2	2.2%
295	2	2.2%
308	2	2.2%
420	4	4.5%
590	1	1.1%
840	3	3.4%
885	1	1.1%
1180	1	1.1%
1232	2	2.2%
1252	16	18.0%

Value	Count	Frequency (%)
7592	2	2.2%
5064	4	4.5%
5008	9	10.1%
3798	4	4.5%
3796	1	1.1%
3756	3	3.4%
2532	3	3.4%
2504	7	7.9%
1912	1	1.1%
1680	3	3.4%

Revenue
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 5 fields: Unit Cost , Unit Price , Profit , Cost , Order Quantity)   ZEROS

Distinct	22
Distinct (%)	24.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4043.9438

Minimum	0
Maximum	13500
Zeros	2
Zeros (%)	2.2%
Negative	0
Negative (%)	0.0%
Memory size	840.0 B
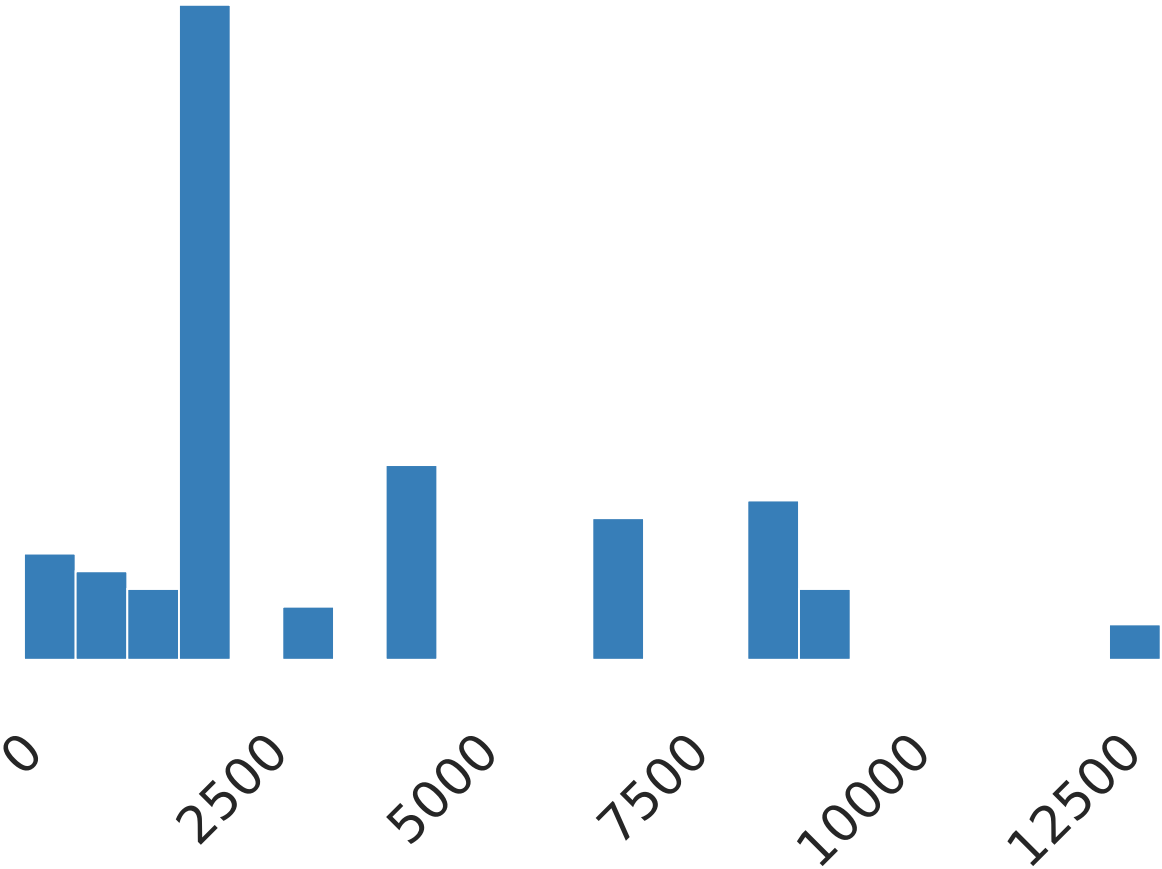
Quantile statistics

Minimum	0
5-th percentile	565
Q1	2295
median	2320
Q3	6750
95-th percentile	9280
Maximum	13500
Range	13500
Interquartile range (IQR)	4455

Descriptive statistics

Standard deviation	3139.7586
Coefficient of variation (CV)	0.77641004
Kurtosis	0.39544936
Mean	4043.9438
Median Absolute Deviation (MAD)	1240
Skewness	1.1063349
Sum	359911
Variance	9858083.9
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=22)
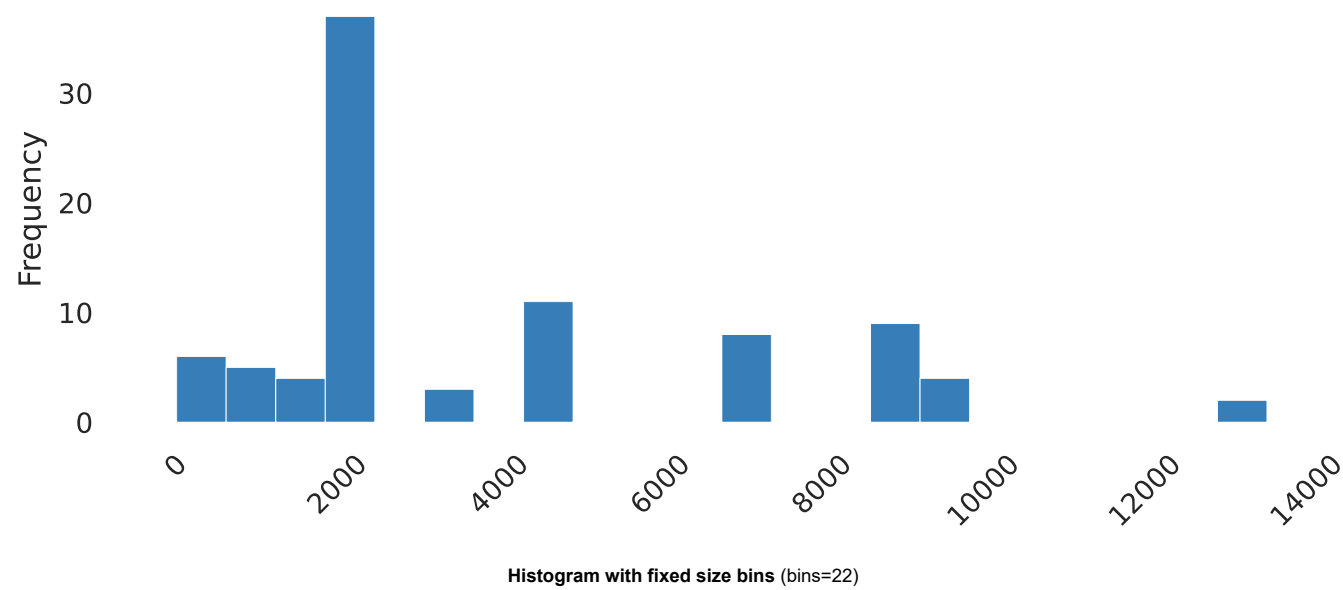
Value	Count	Frequency (%)
2295	16	18.0%
2320	16	18.0%
9180	9	10.1%
4590	8	9.0%
6960	4	4.5%
9280	4	4.5%
769	4	4.5%
4640	3	3.4%
6885	3	3.4%
1538	3	3.4%
Other values (12)	19	21.3%

Value	Count	Frequency (%)
0	2	2.2%
540	2	2.2%
565	2	2.2%
769	4	4.5%
1080	1	1.1%
1538	3	3.4%
1620	1	1.1%
2160	1	1.1%
2260	2	2.2%
2295	16	18.0%

Value	Count	Frequency (%)
13500	2	2.2%
9280	4	4.5%
9180	9	10.1%
6960	4	4.5%
6885	3	3.4%
6750	1	1.1%
4640	3	3.4%
4590	8	9.0%
3400	1	1.1%
3076	2	2.2%

Interactions

+2.617e5

+2.617e5

+2.617e5

+2.617e5

+2.617e5

+2.617e5
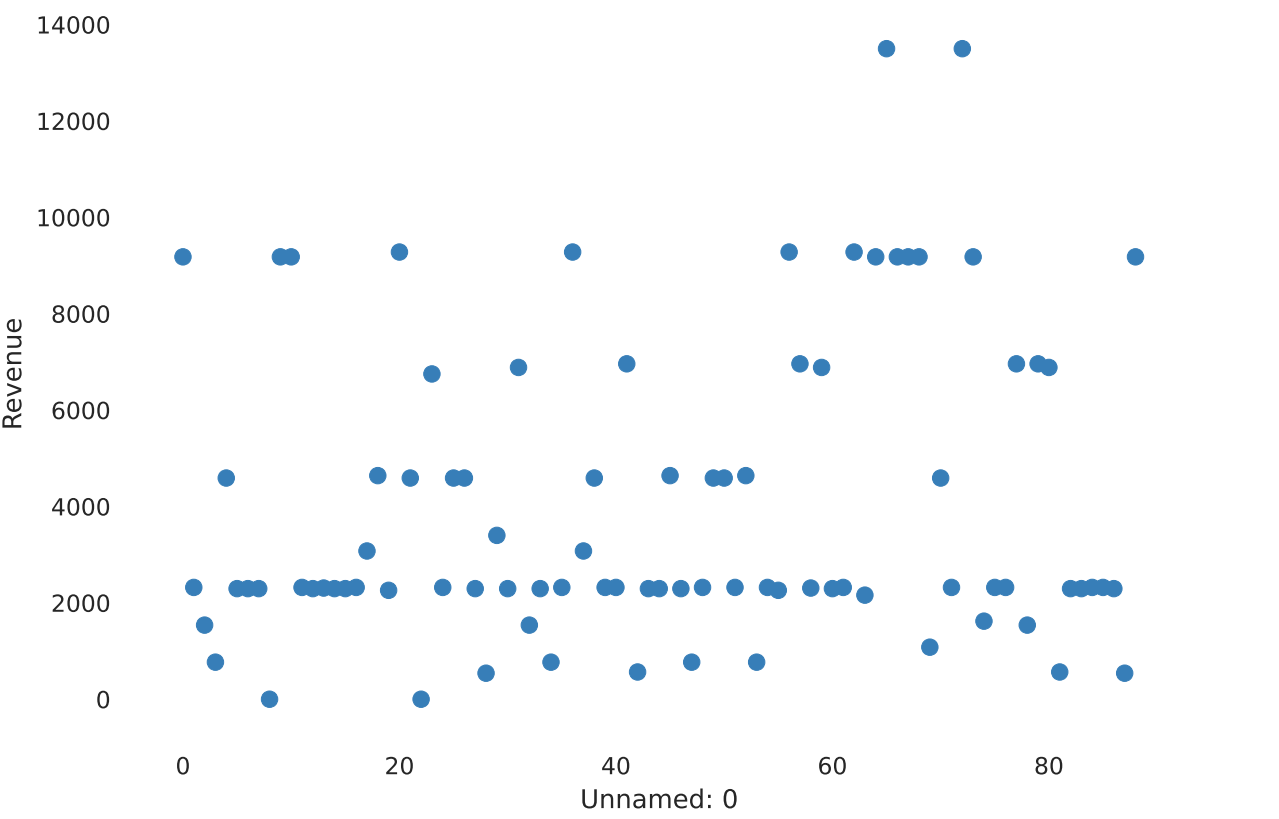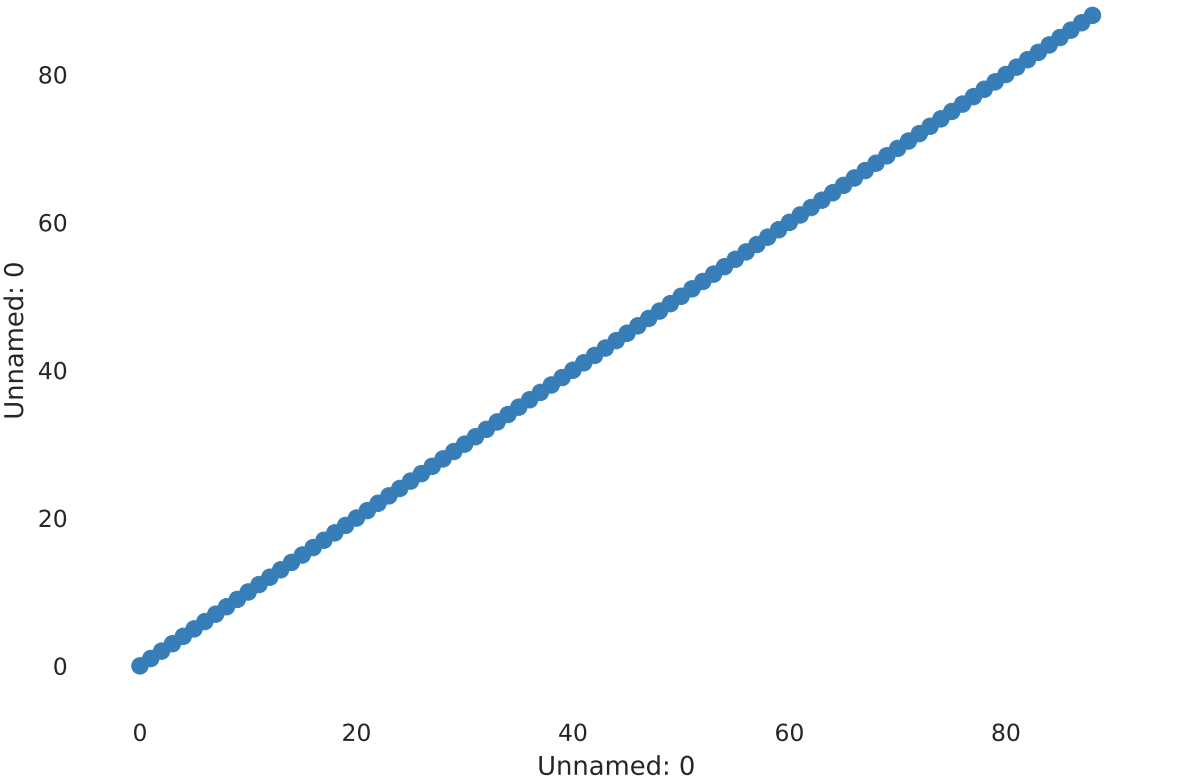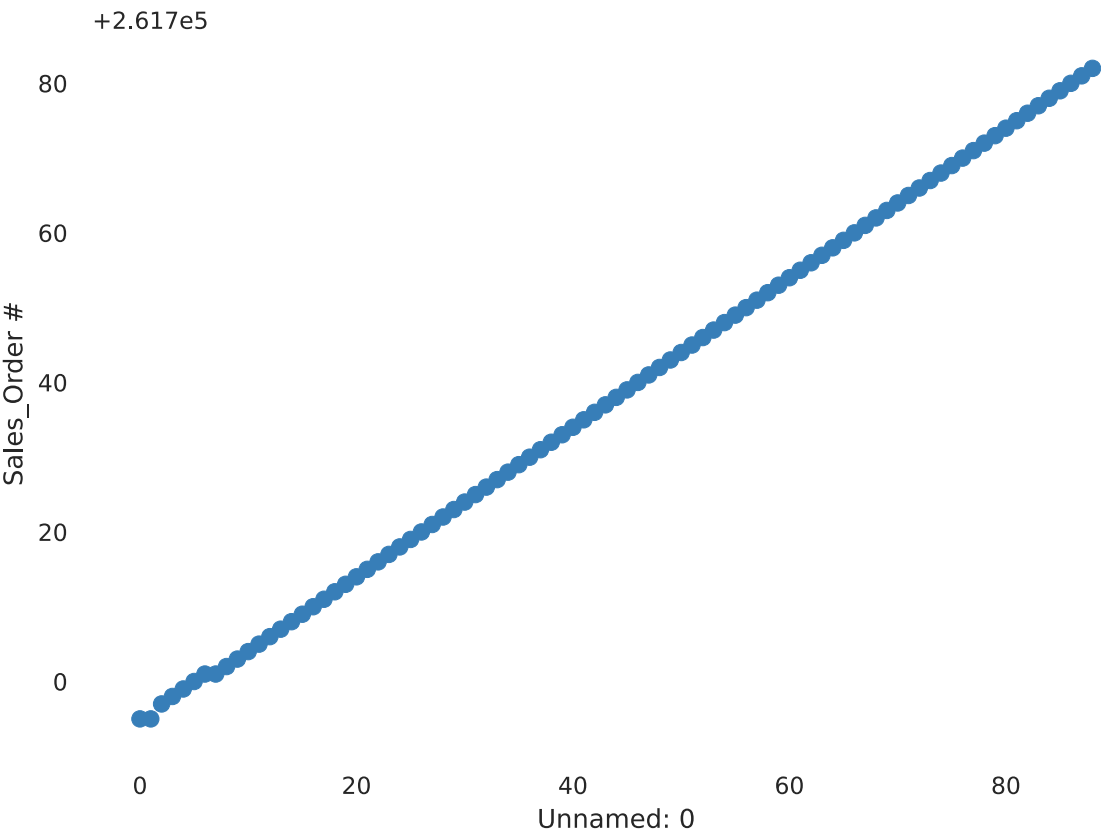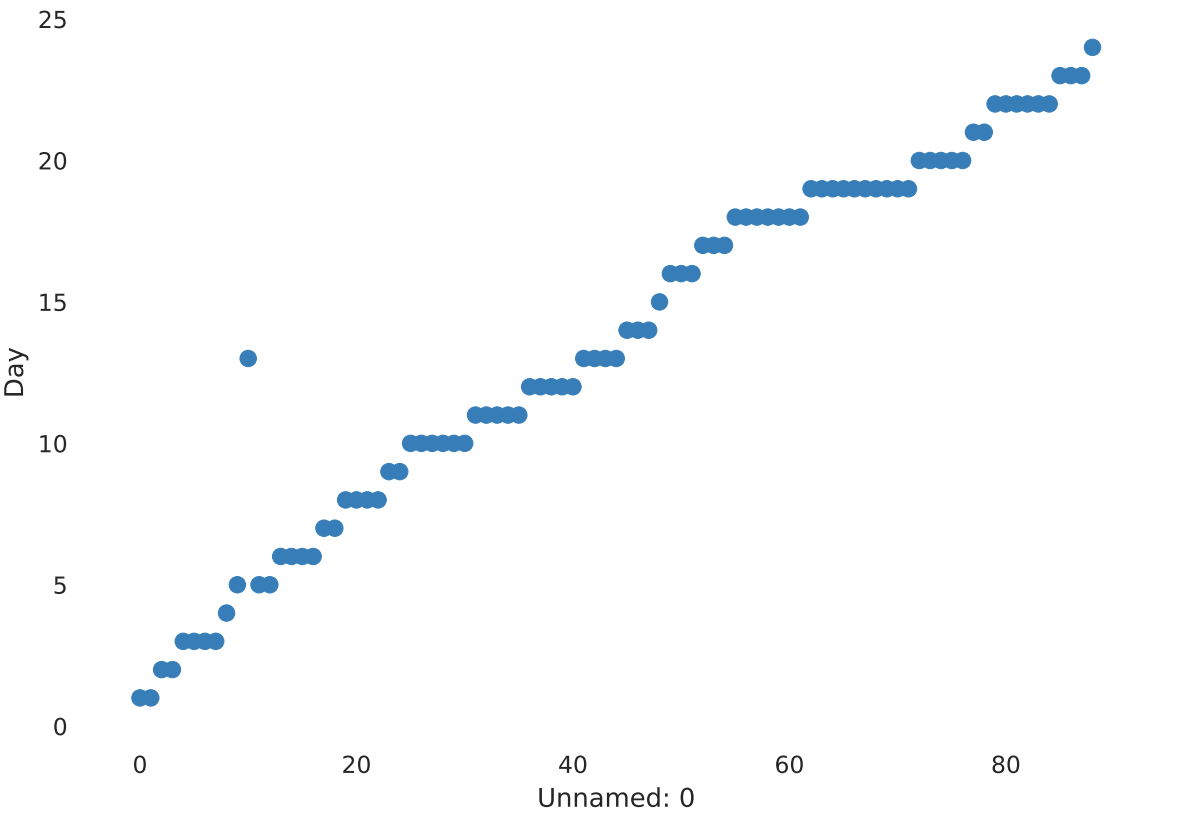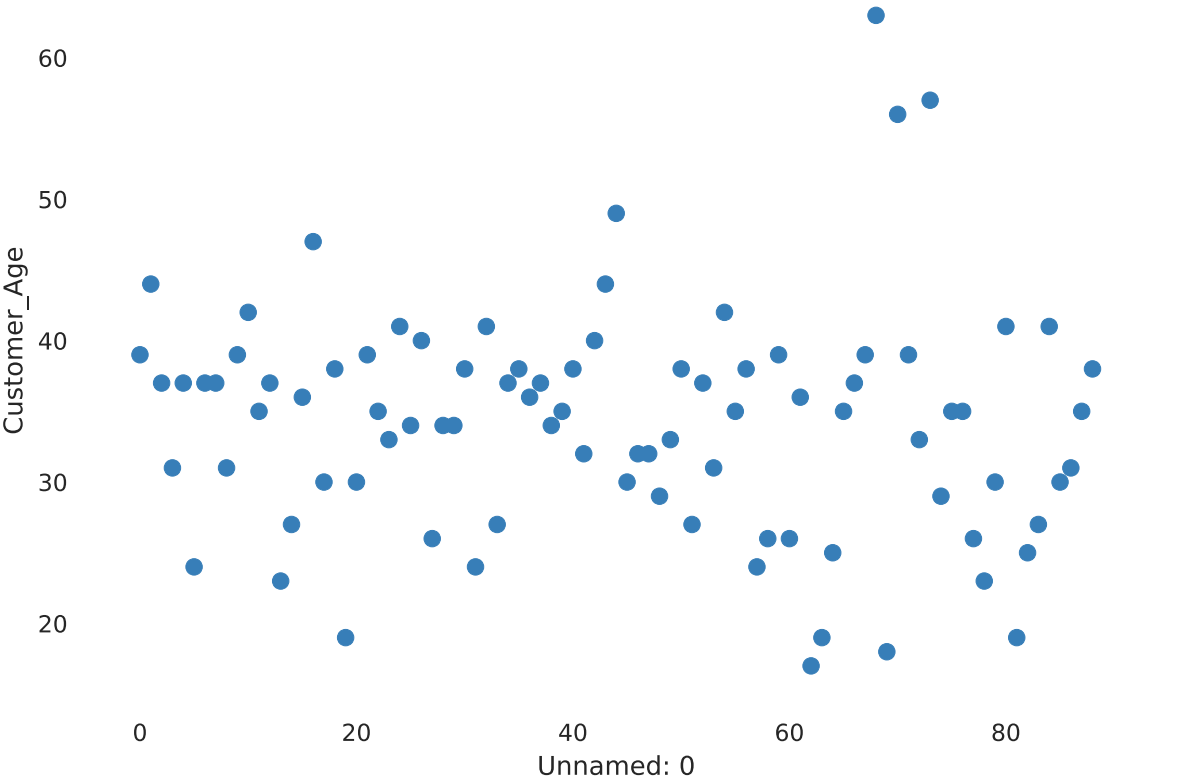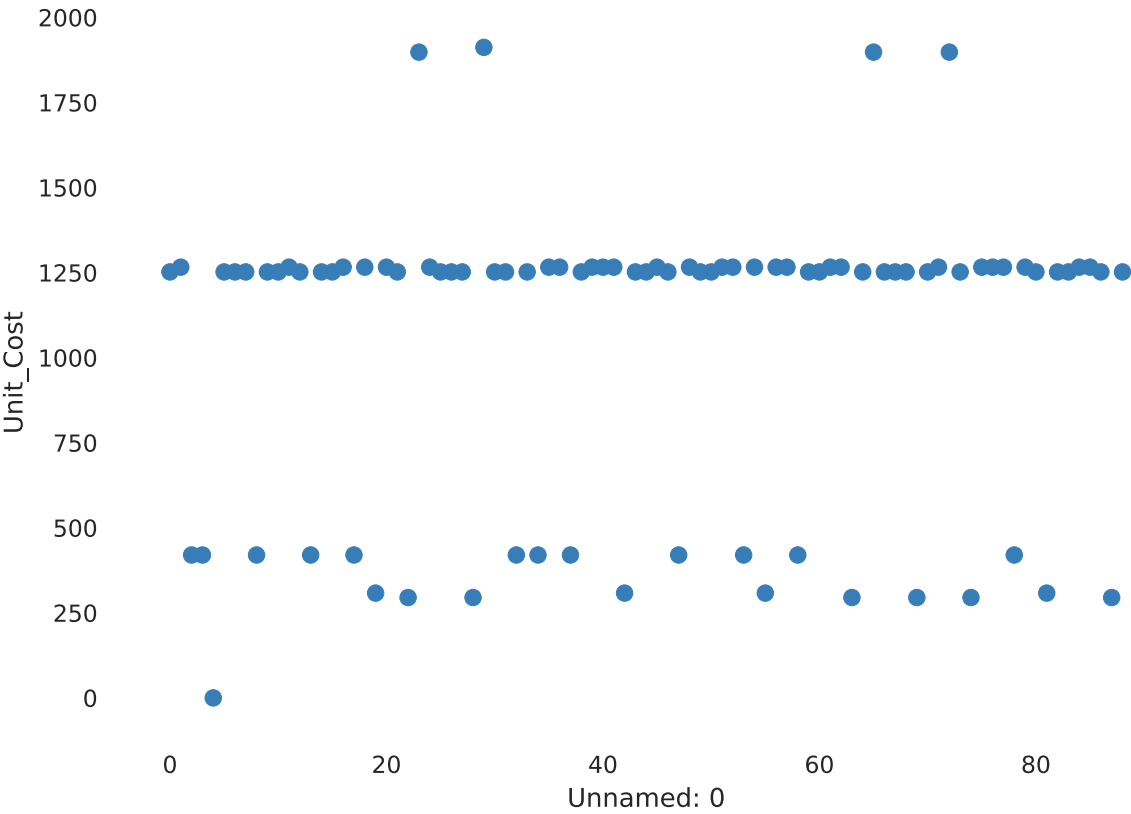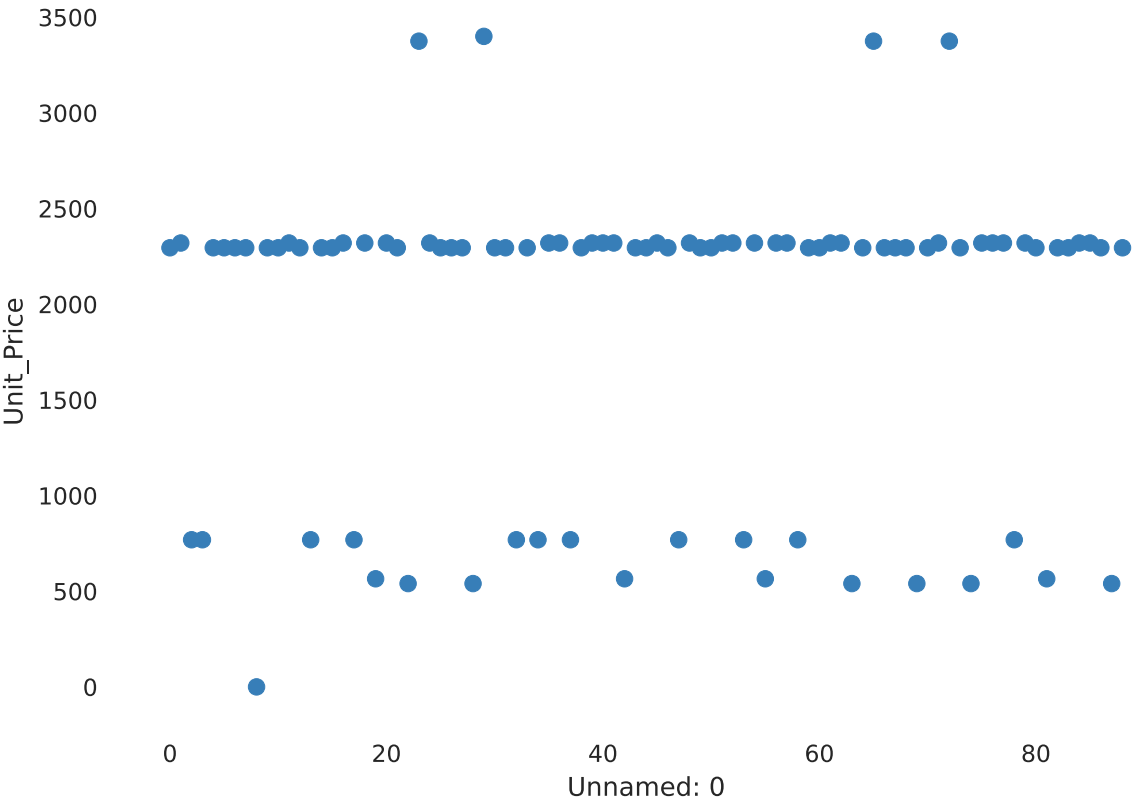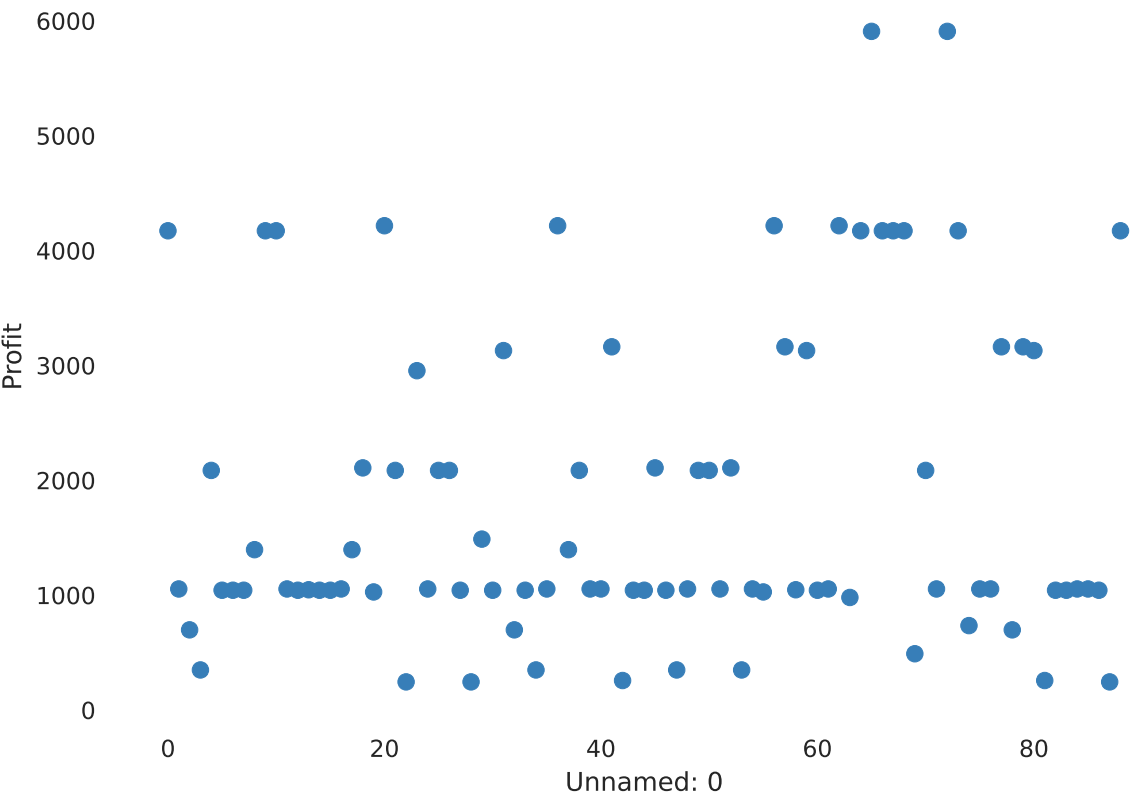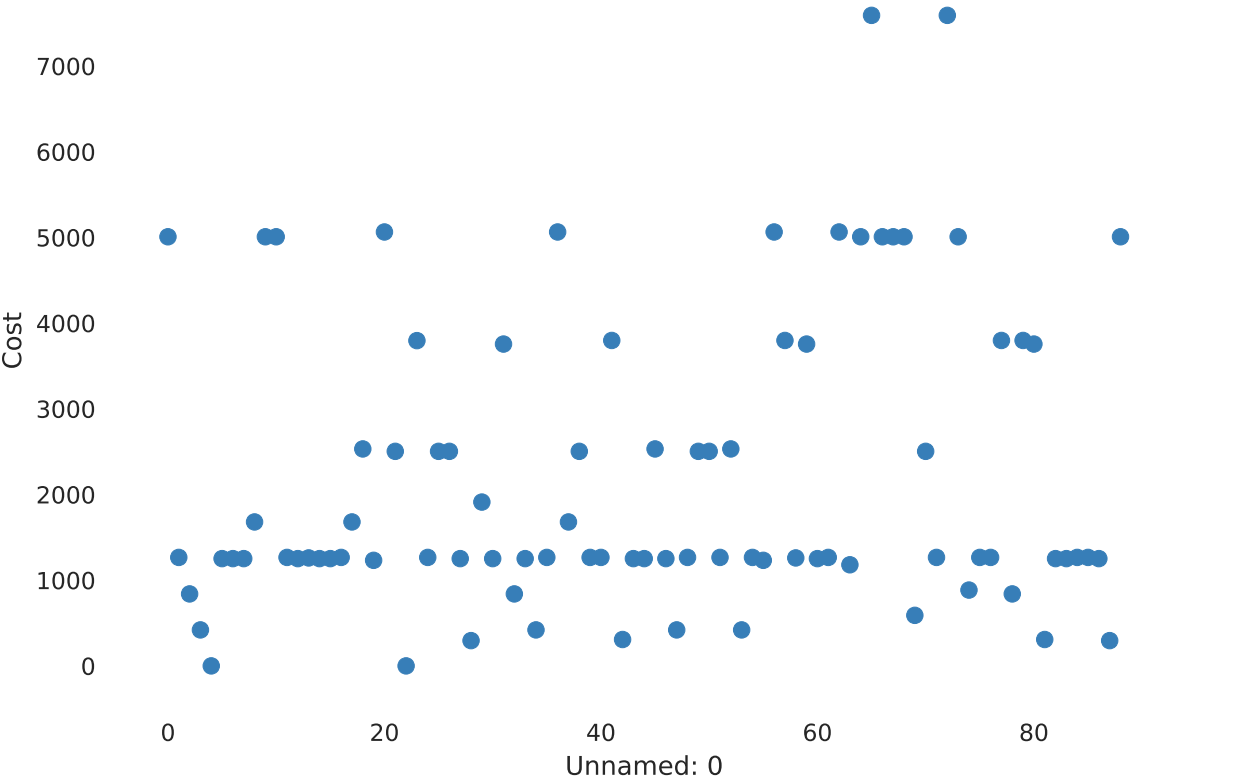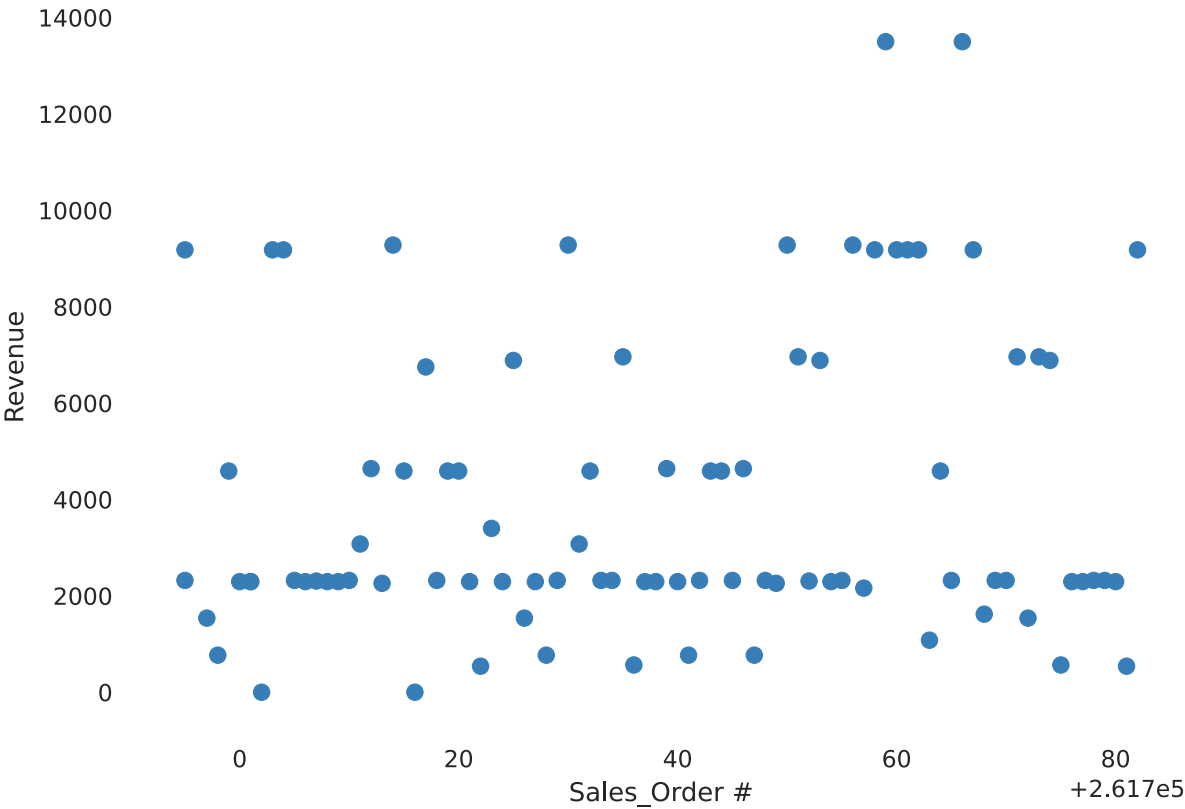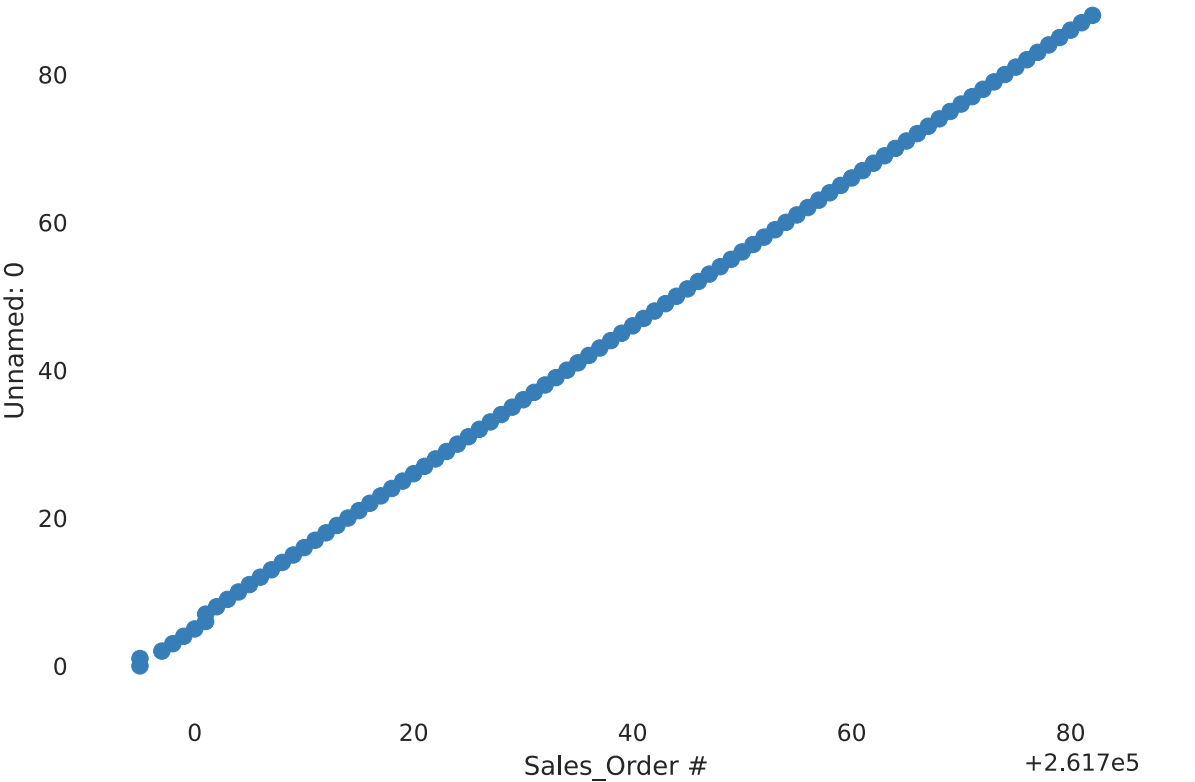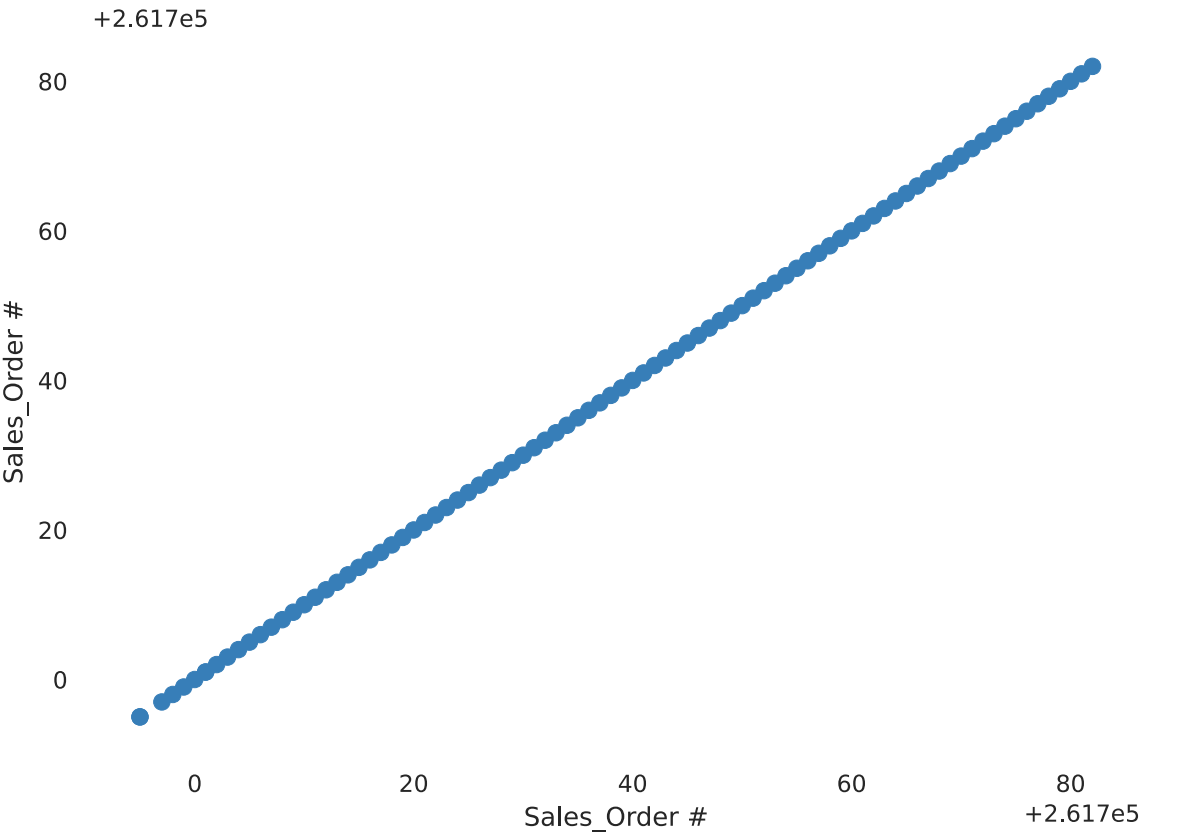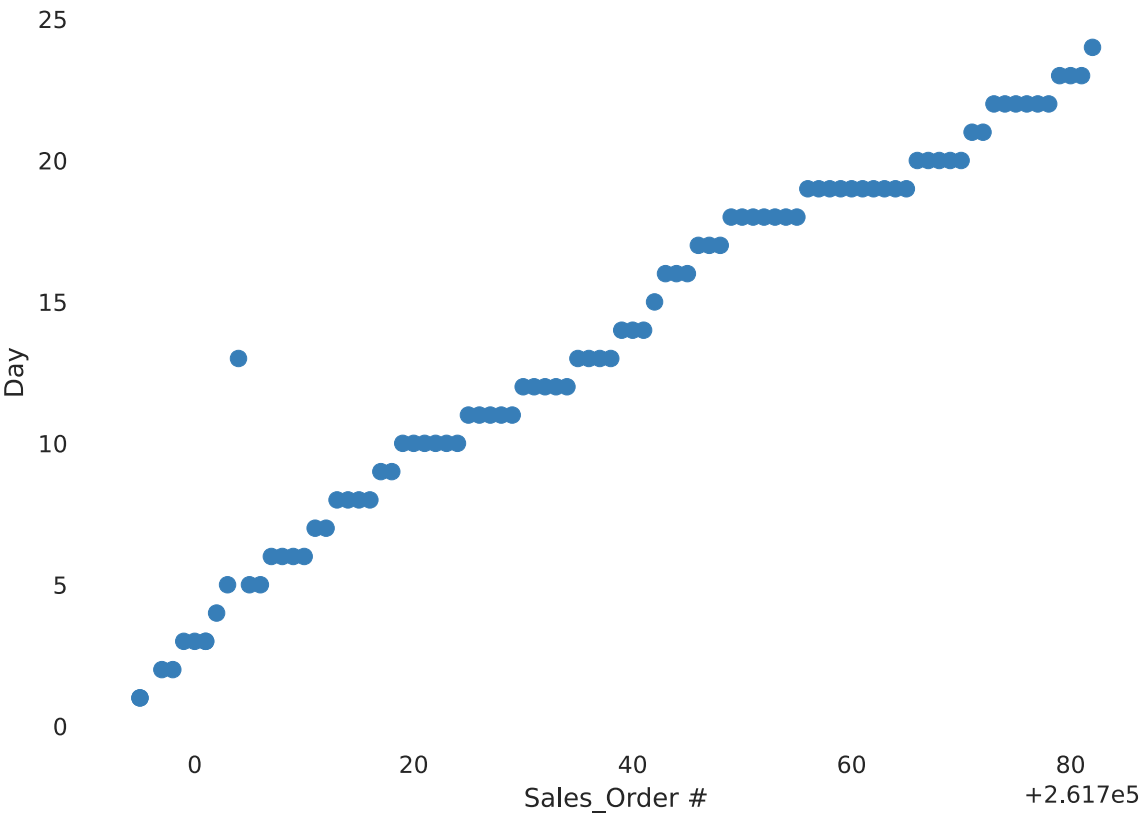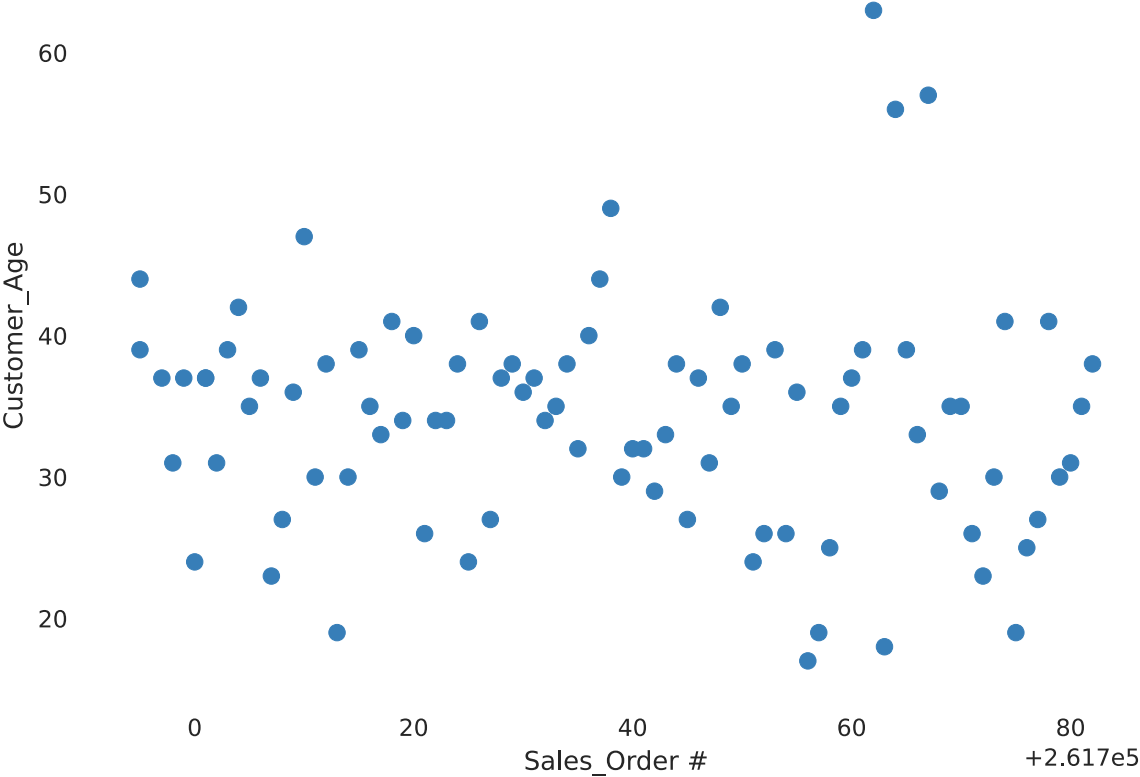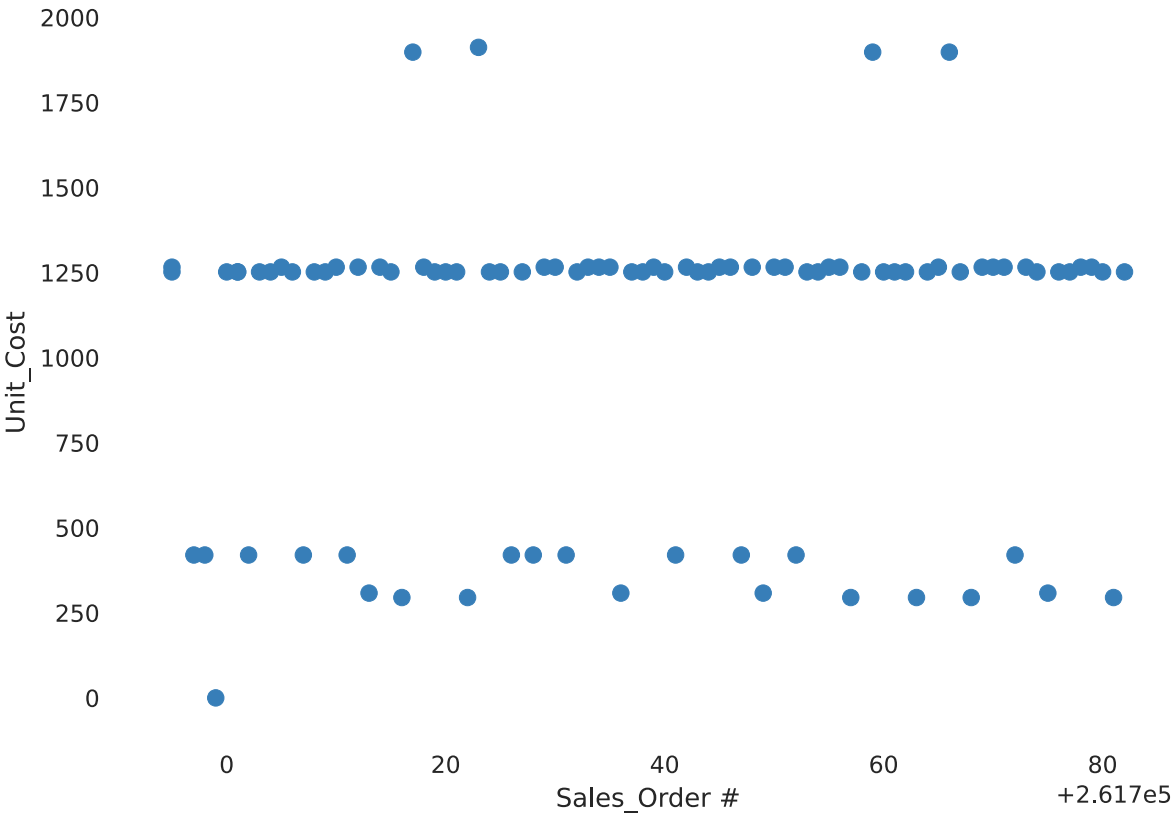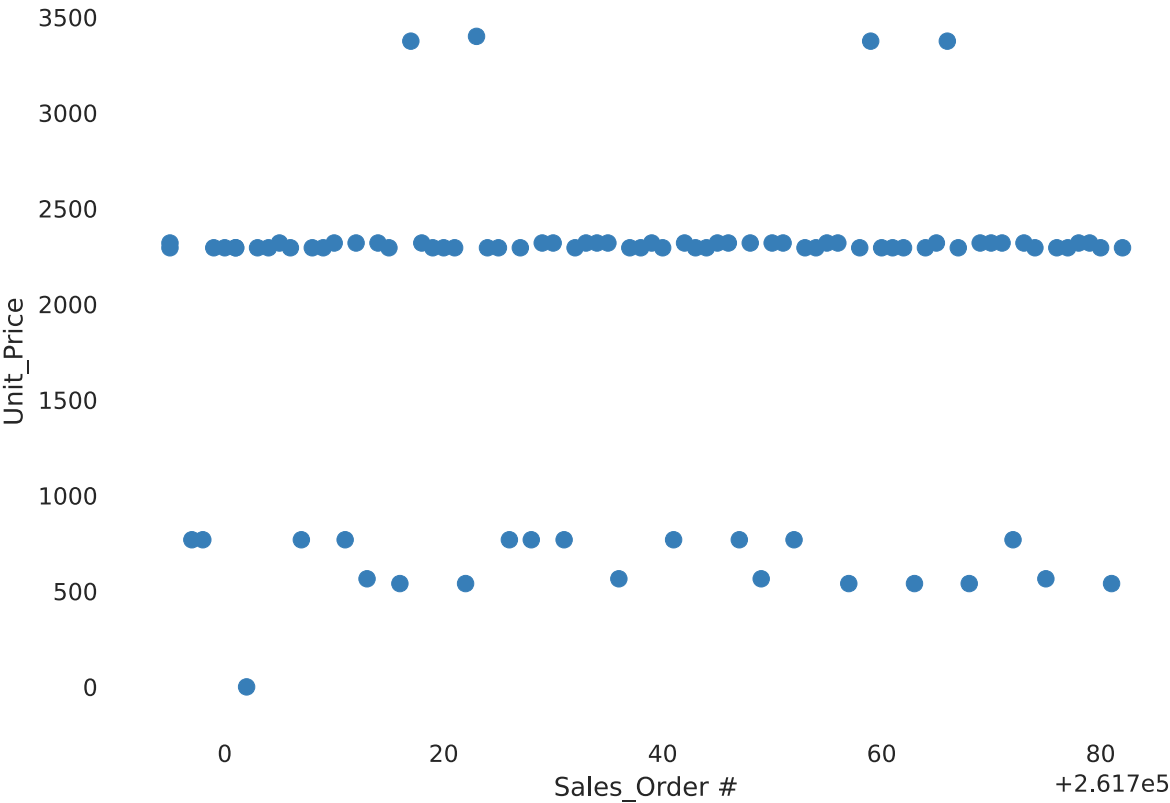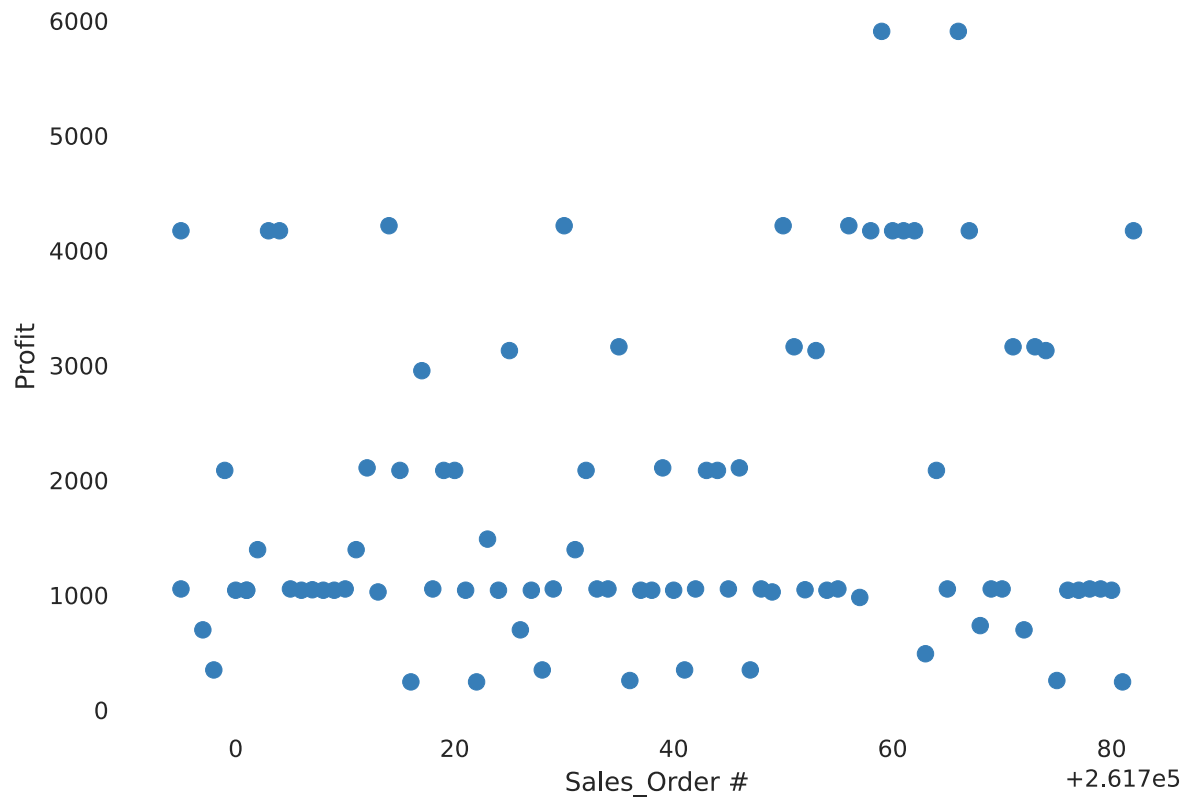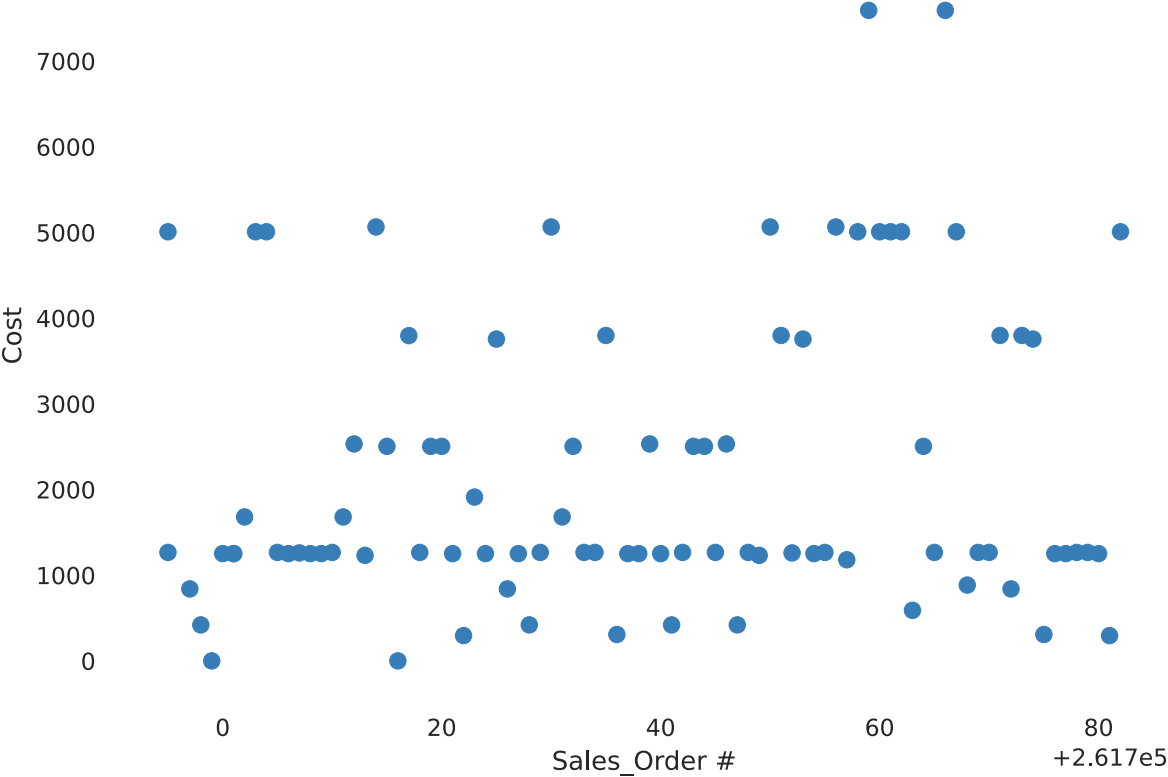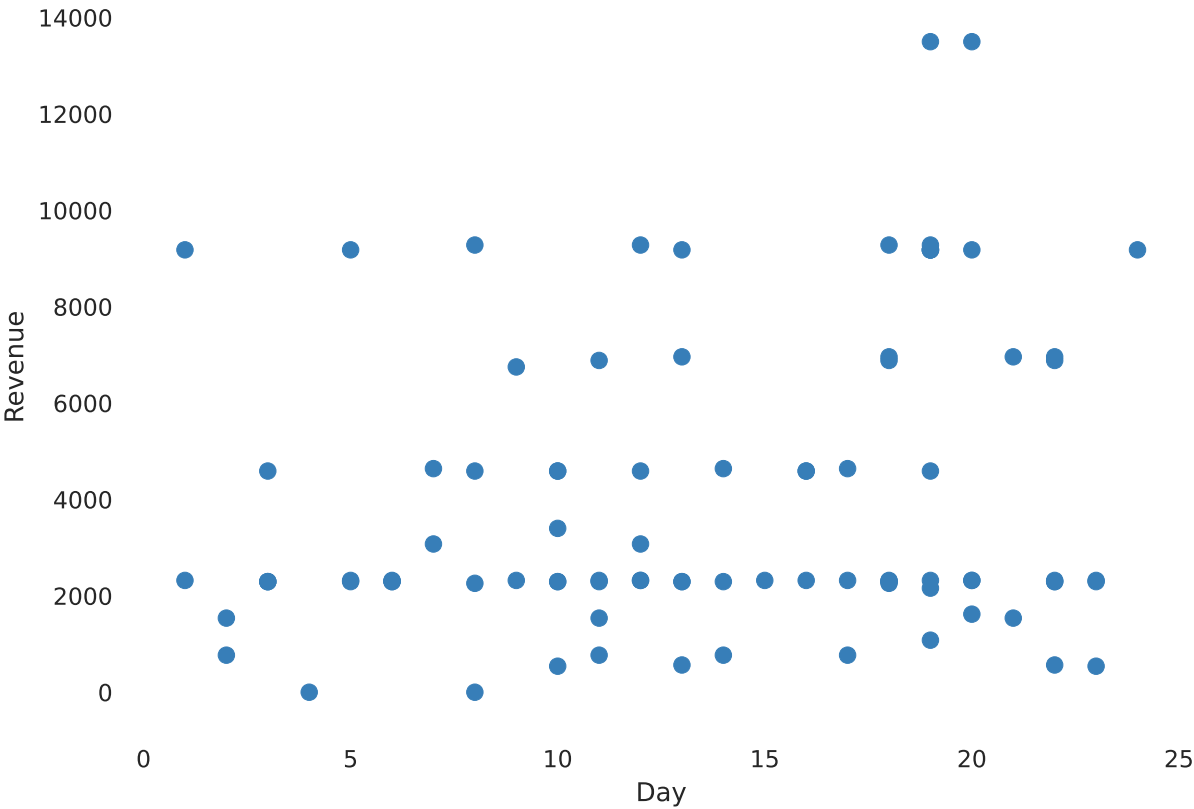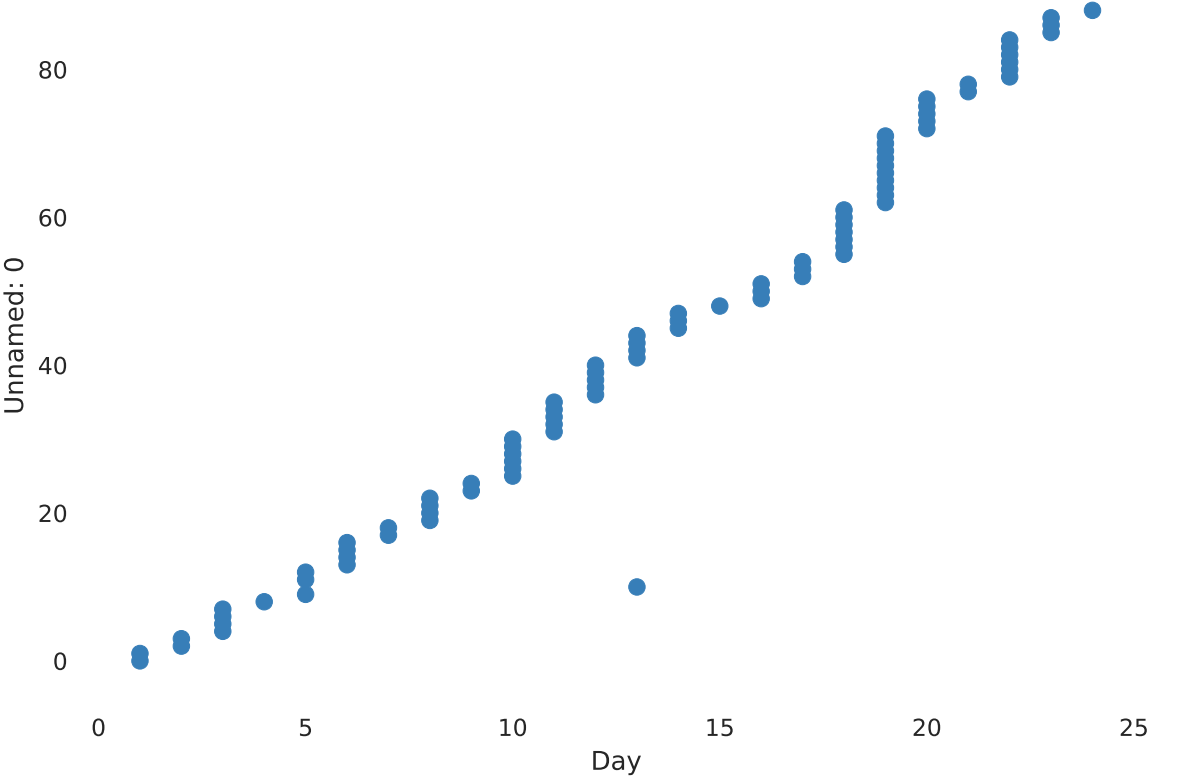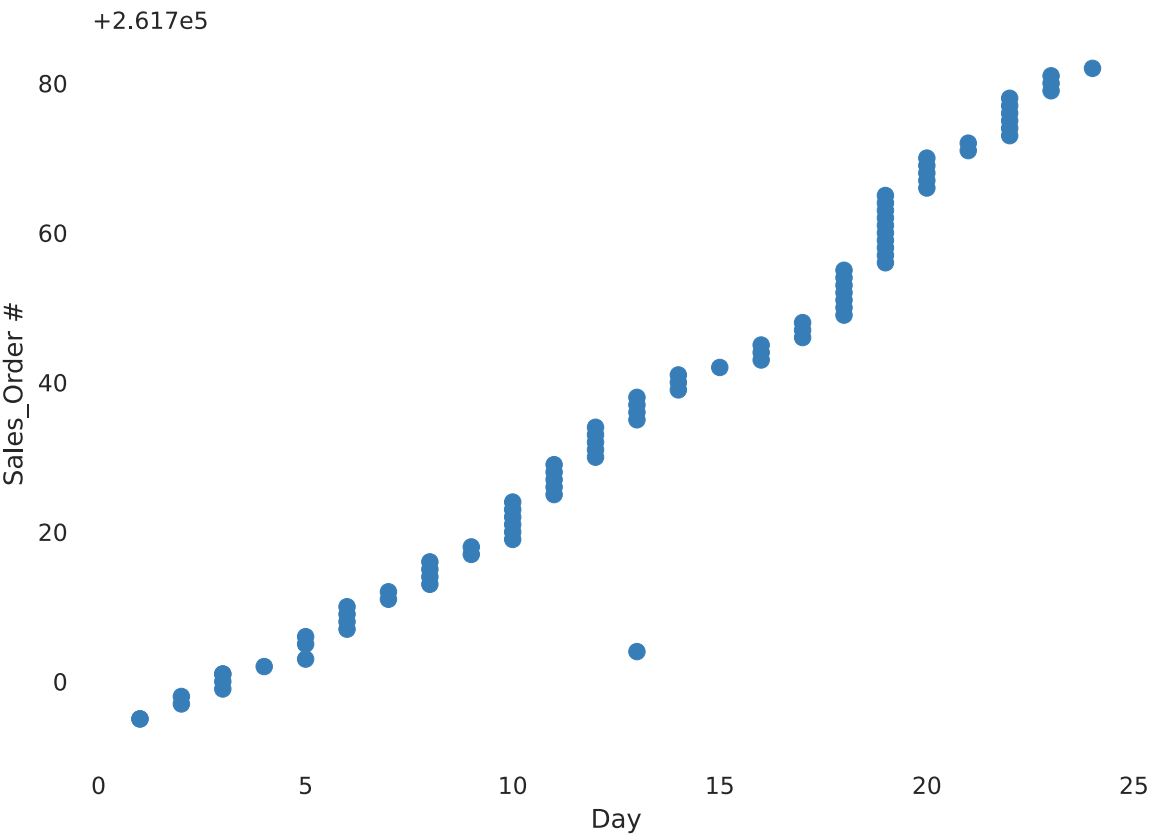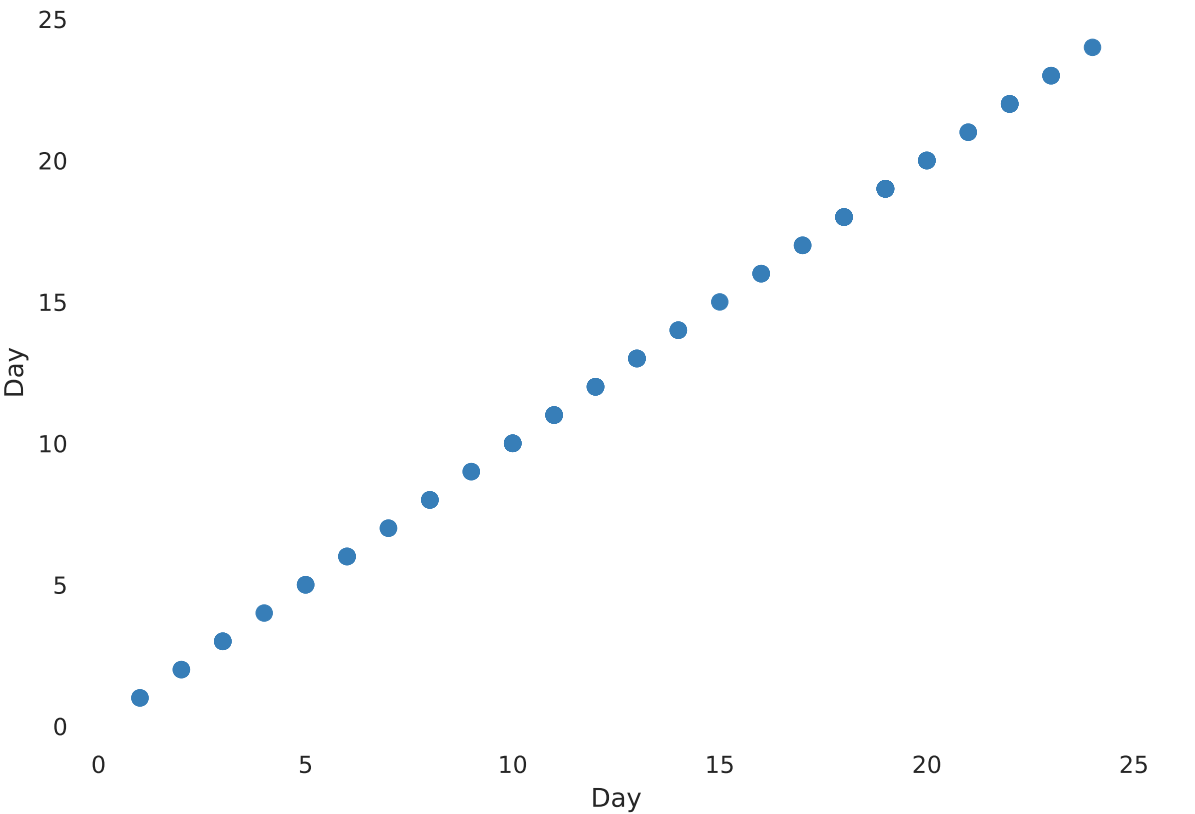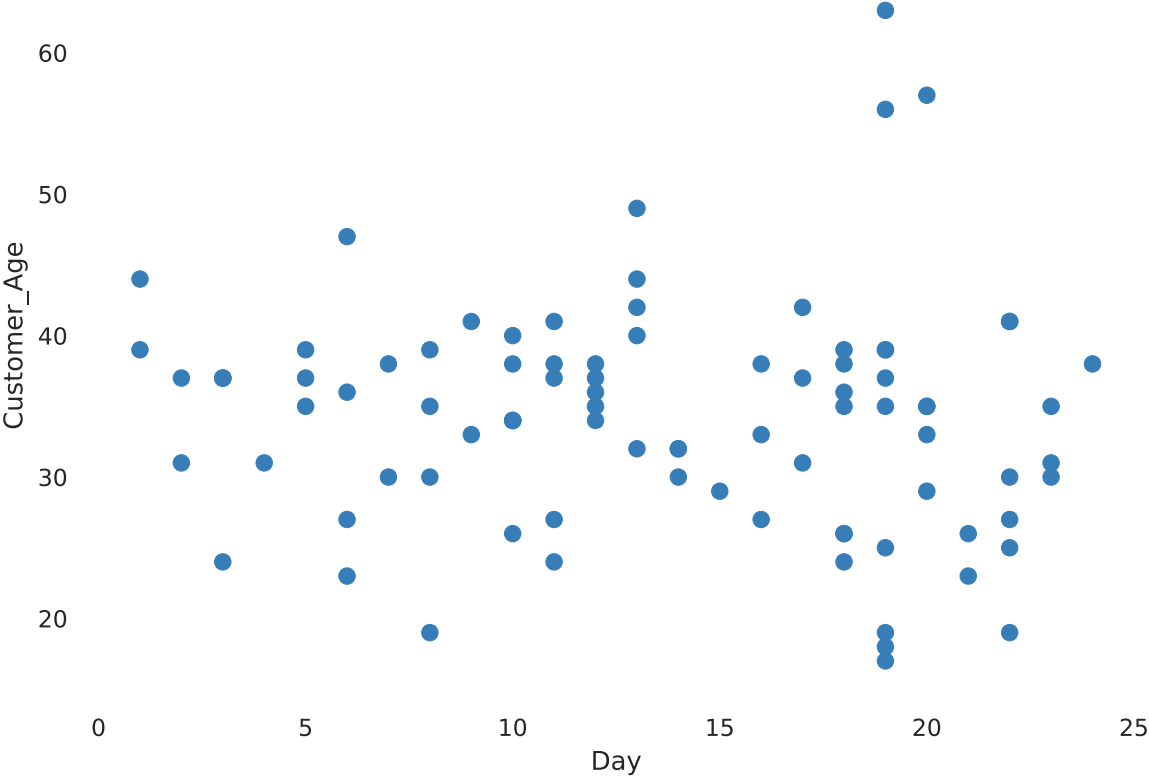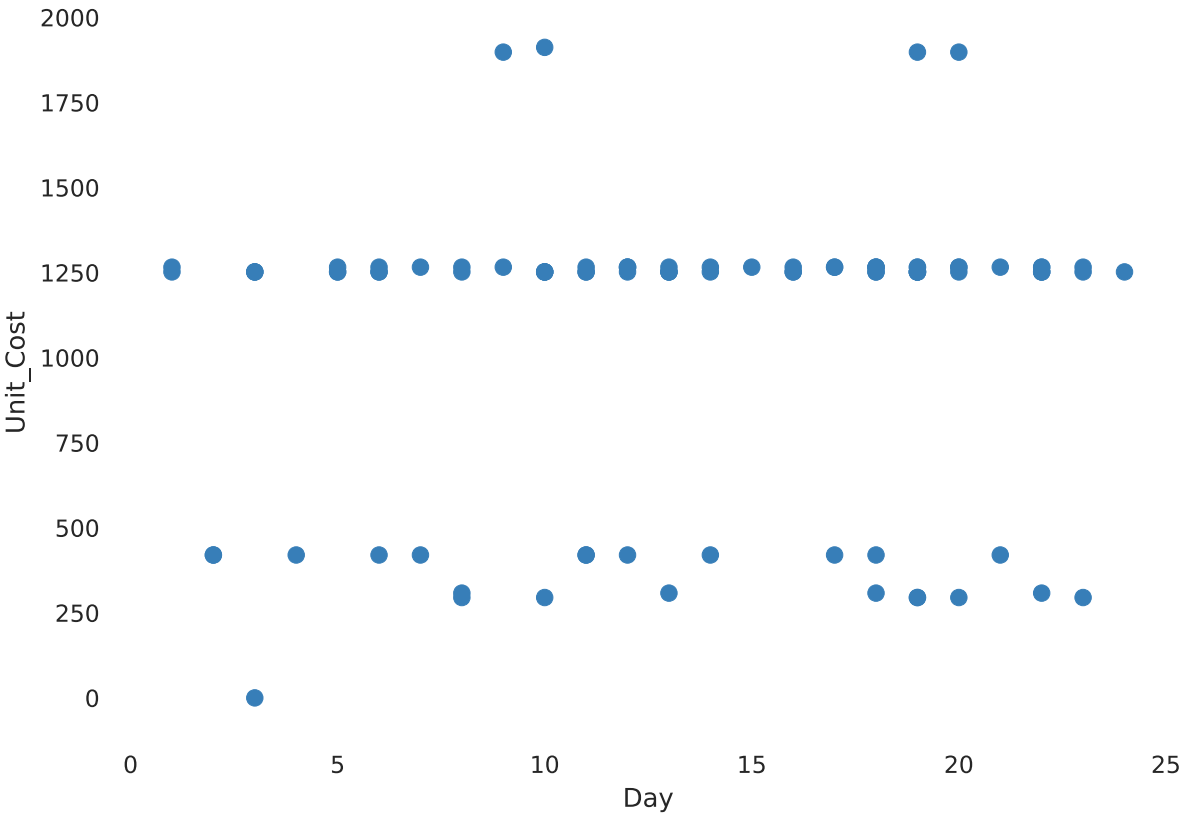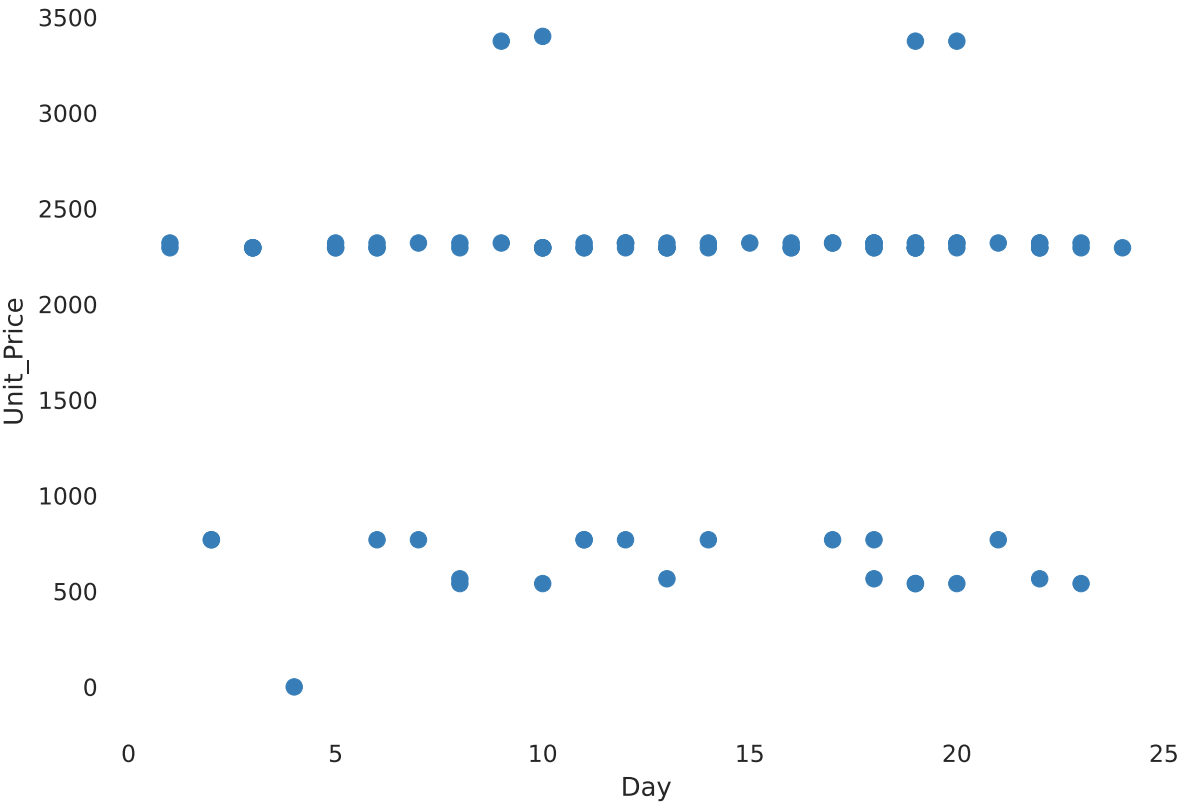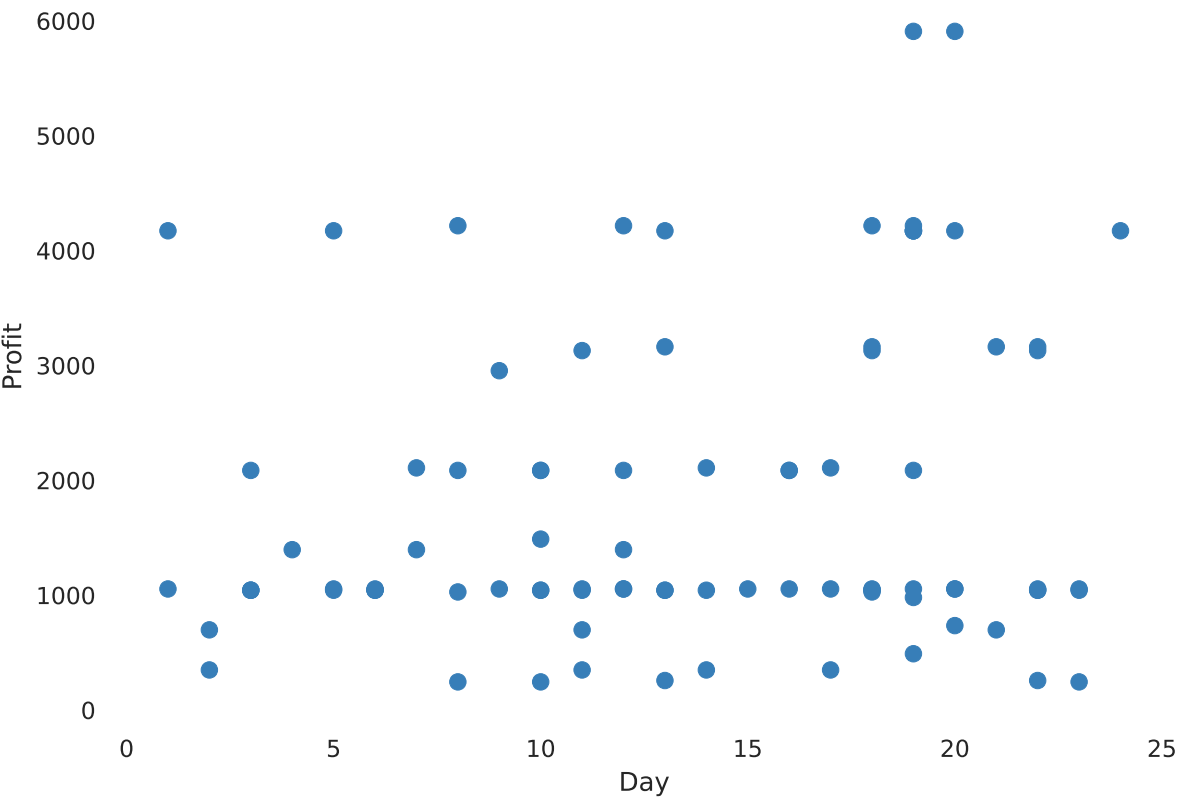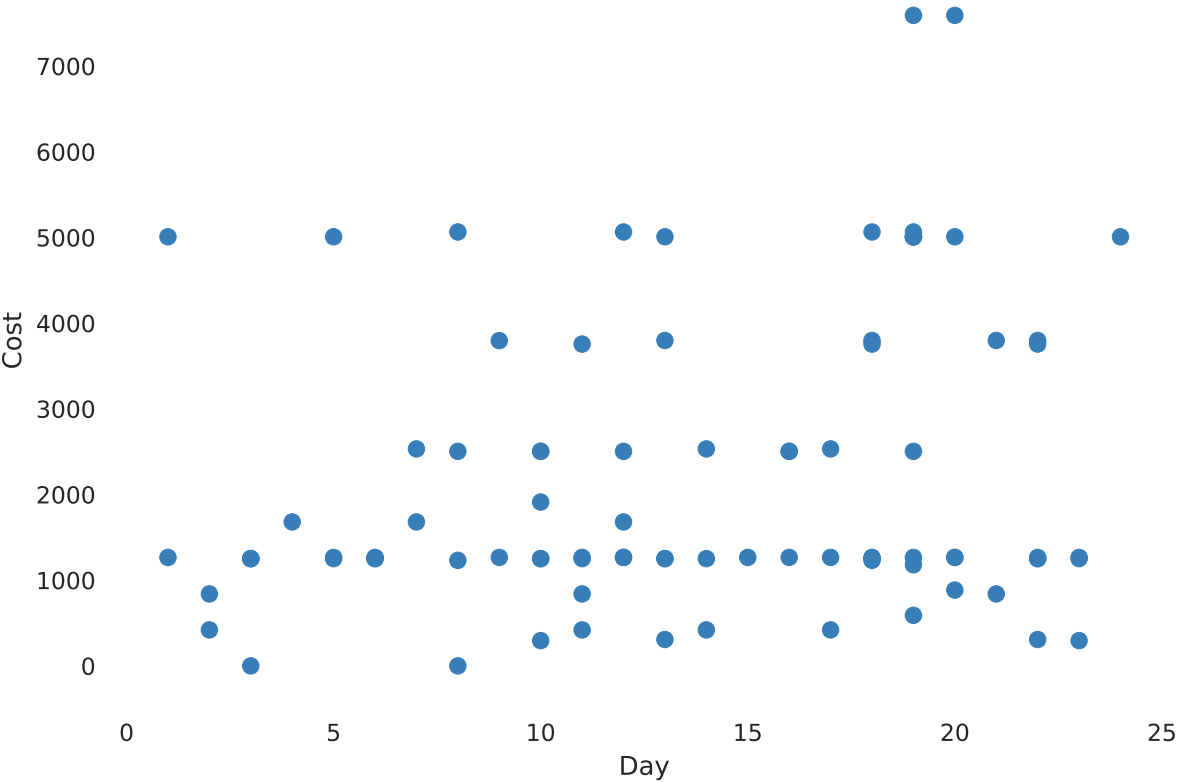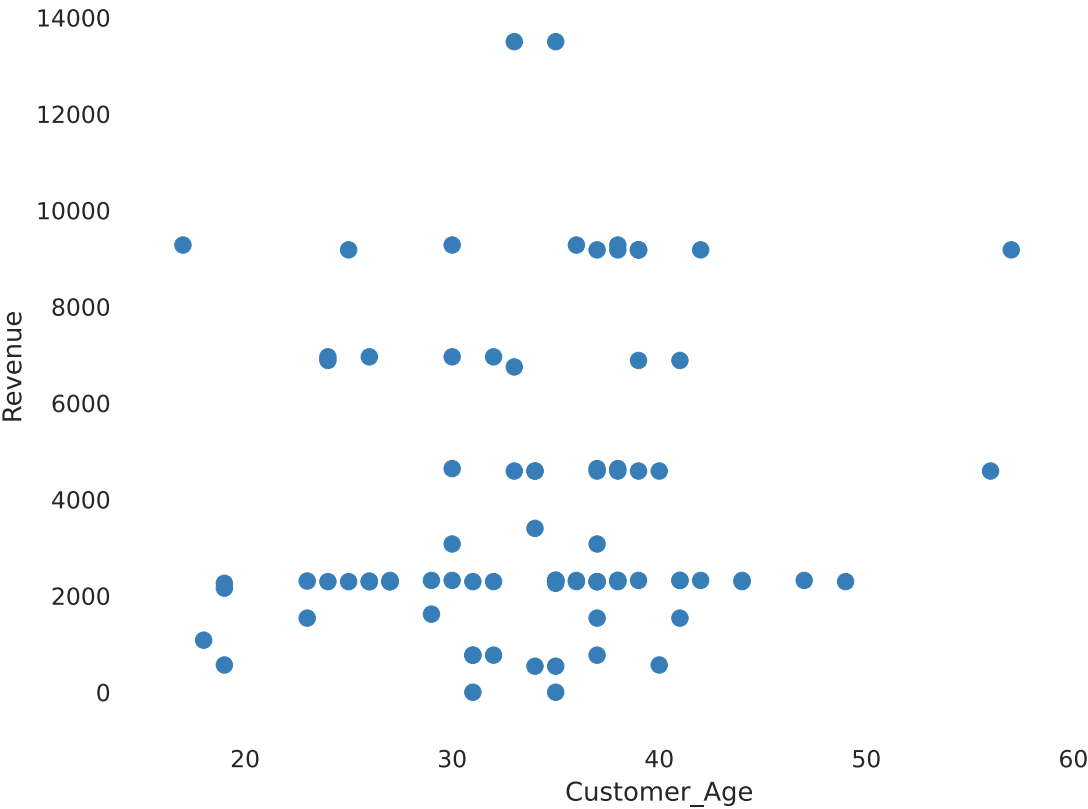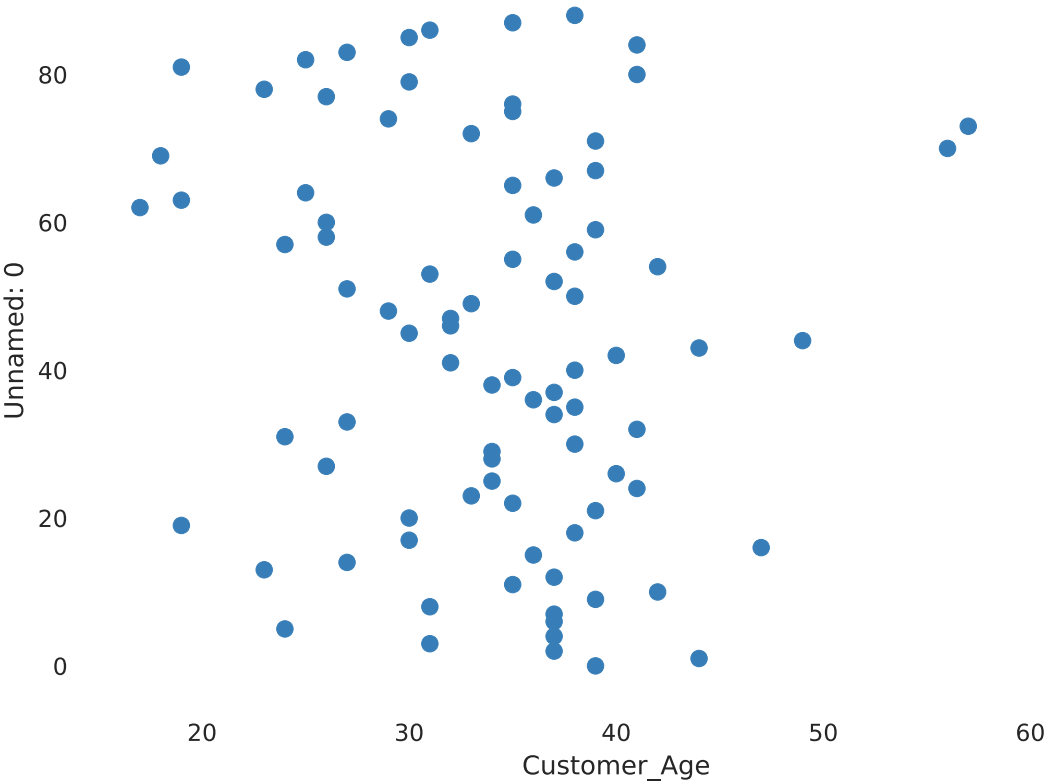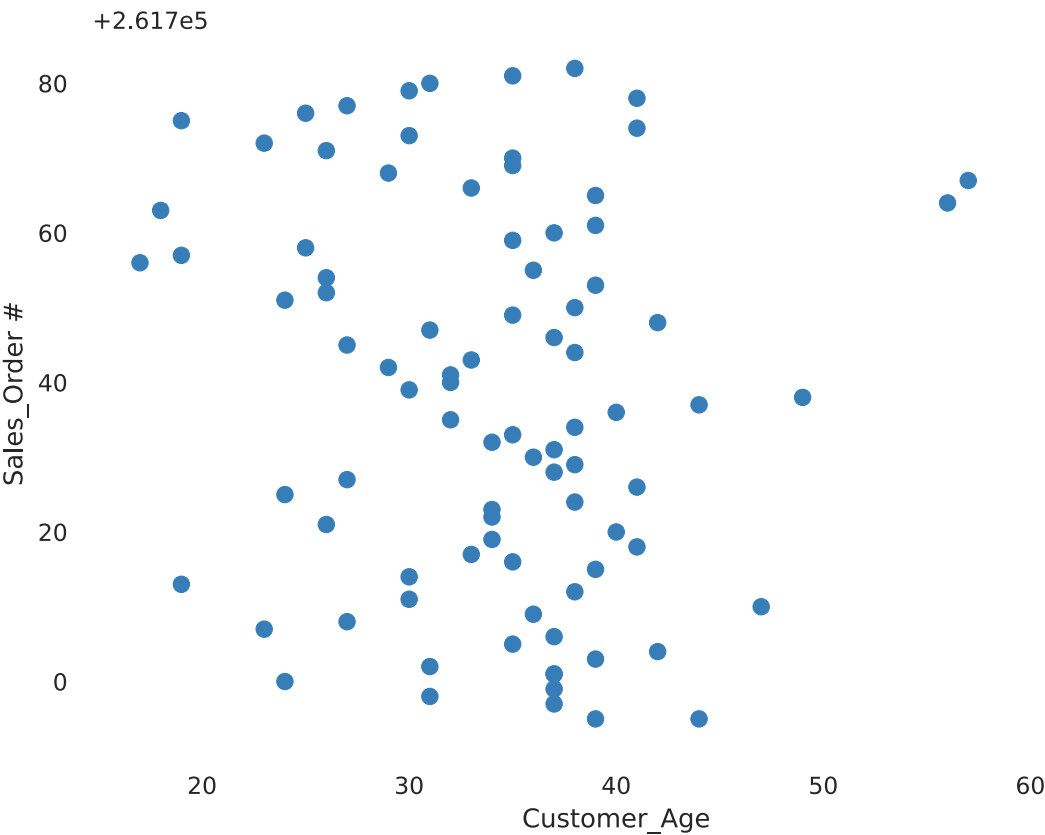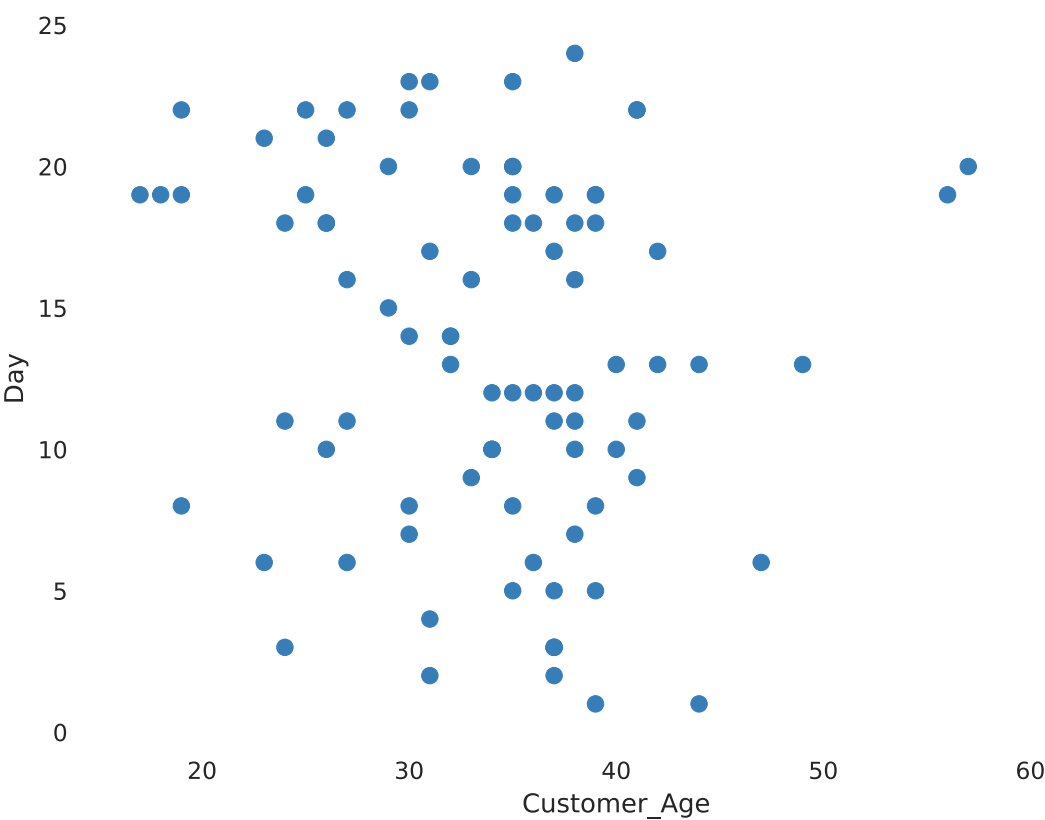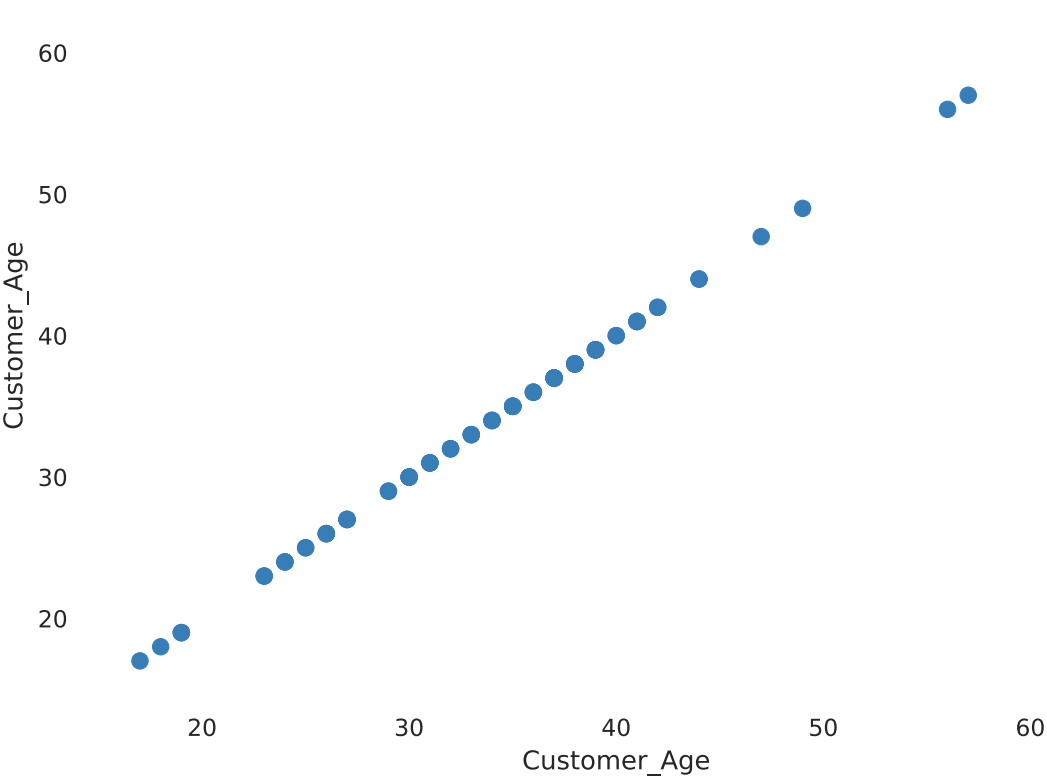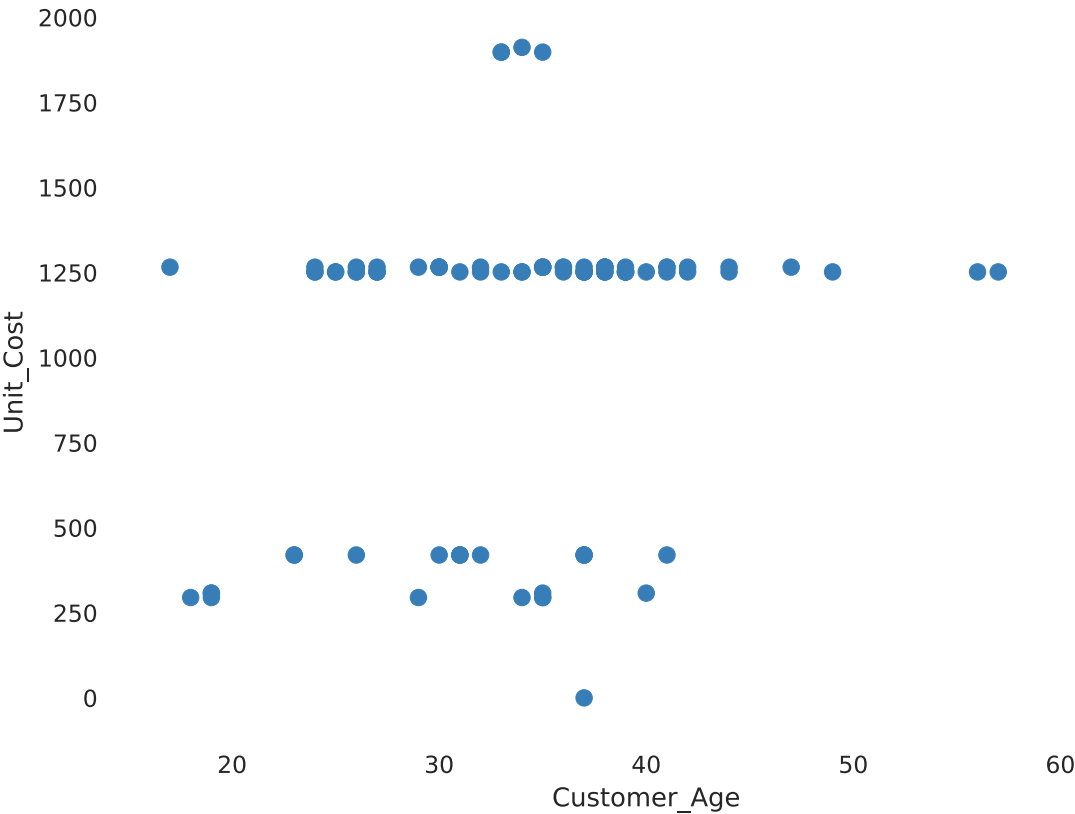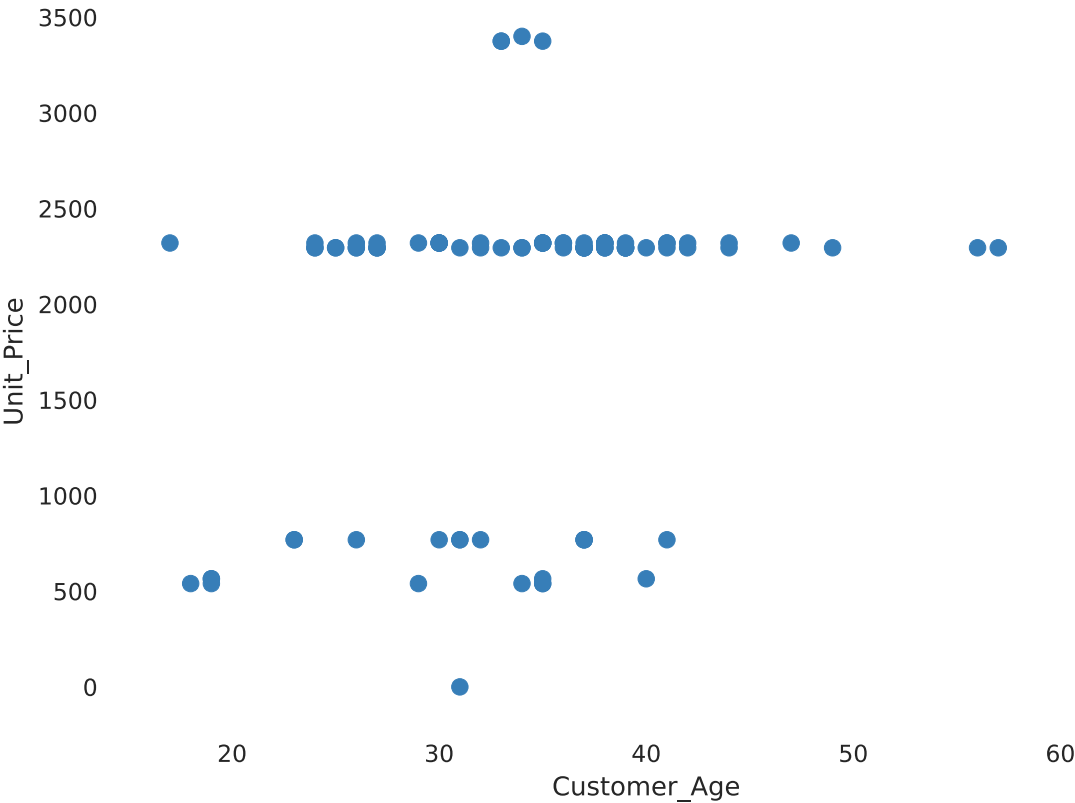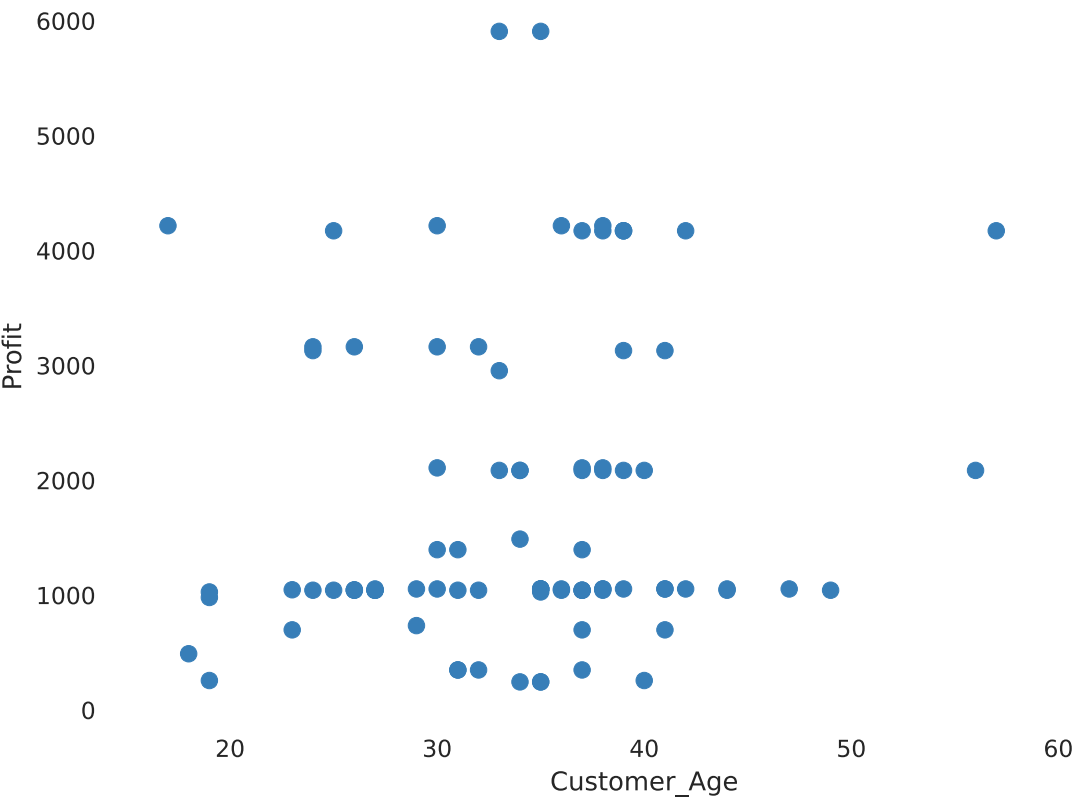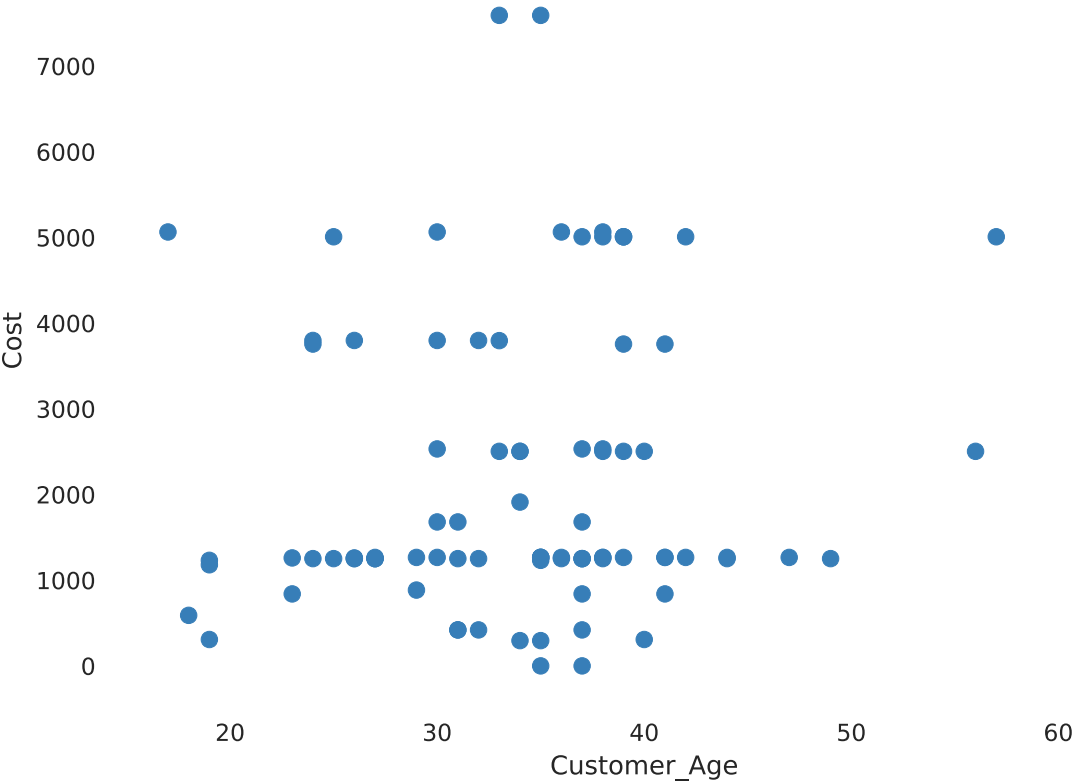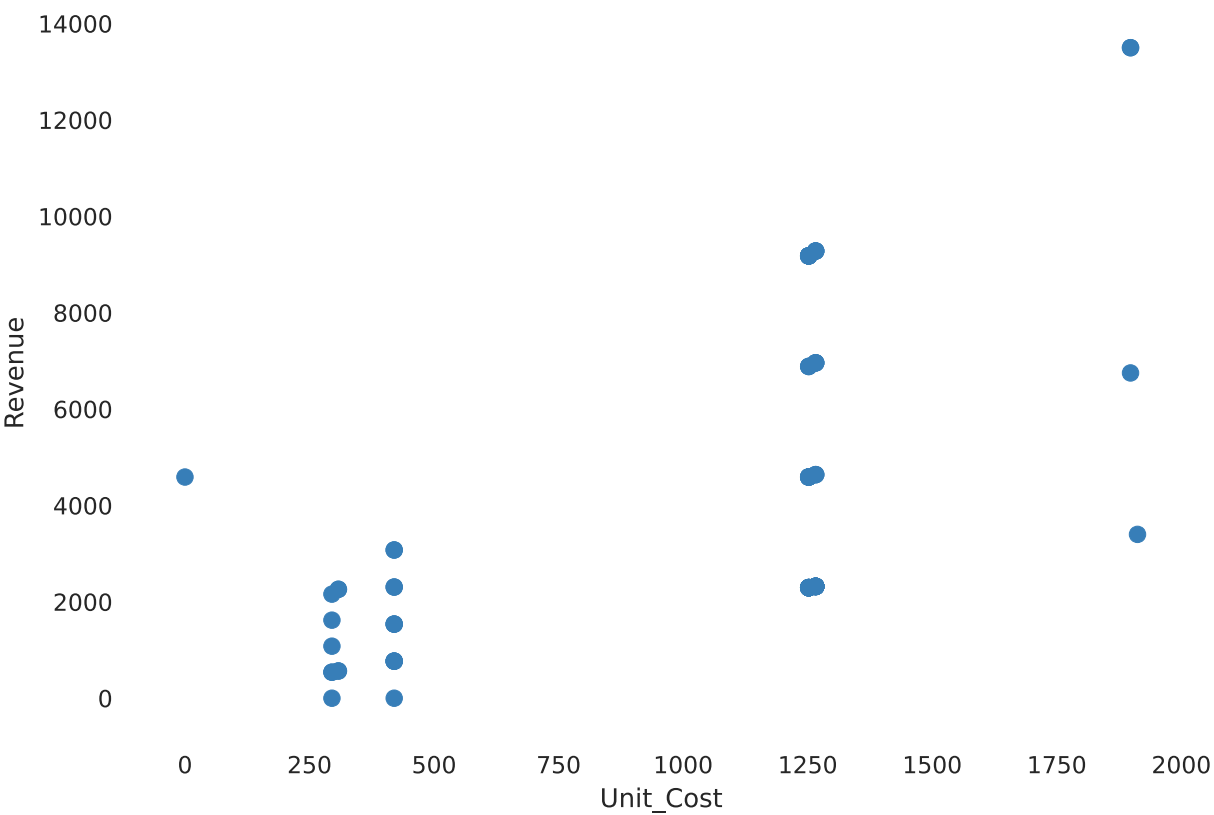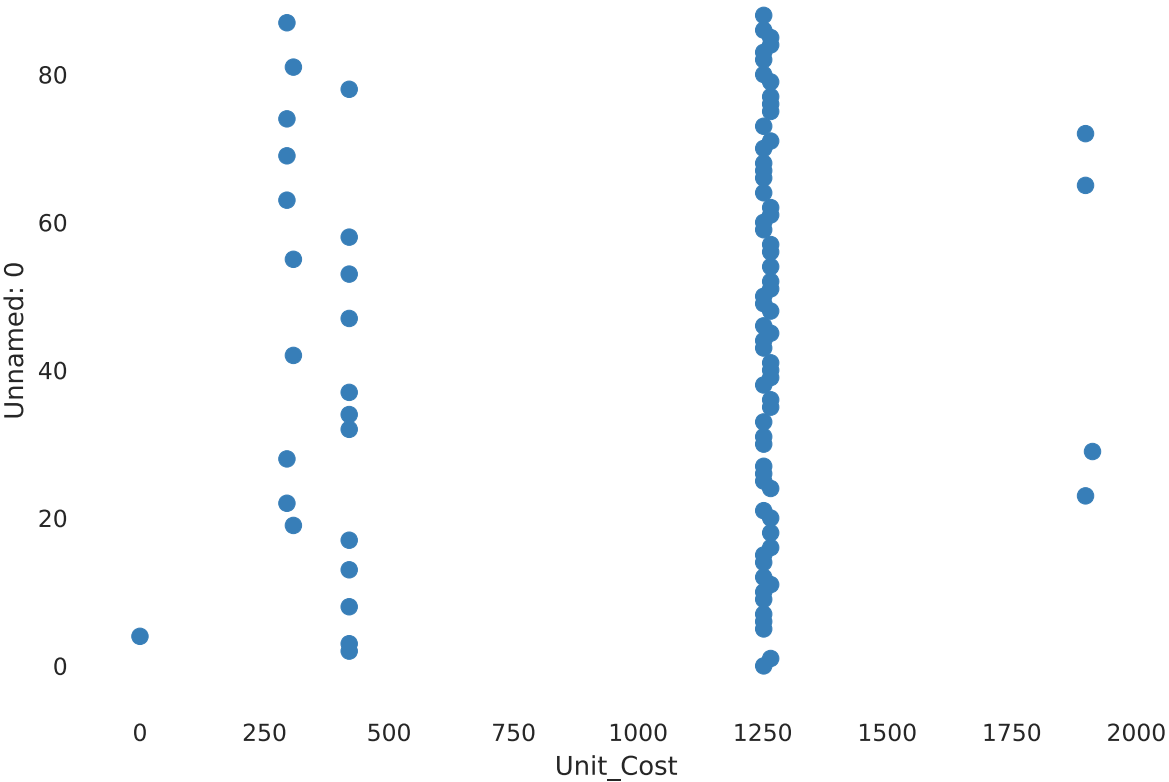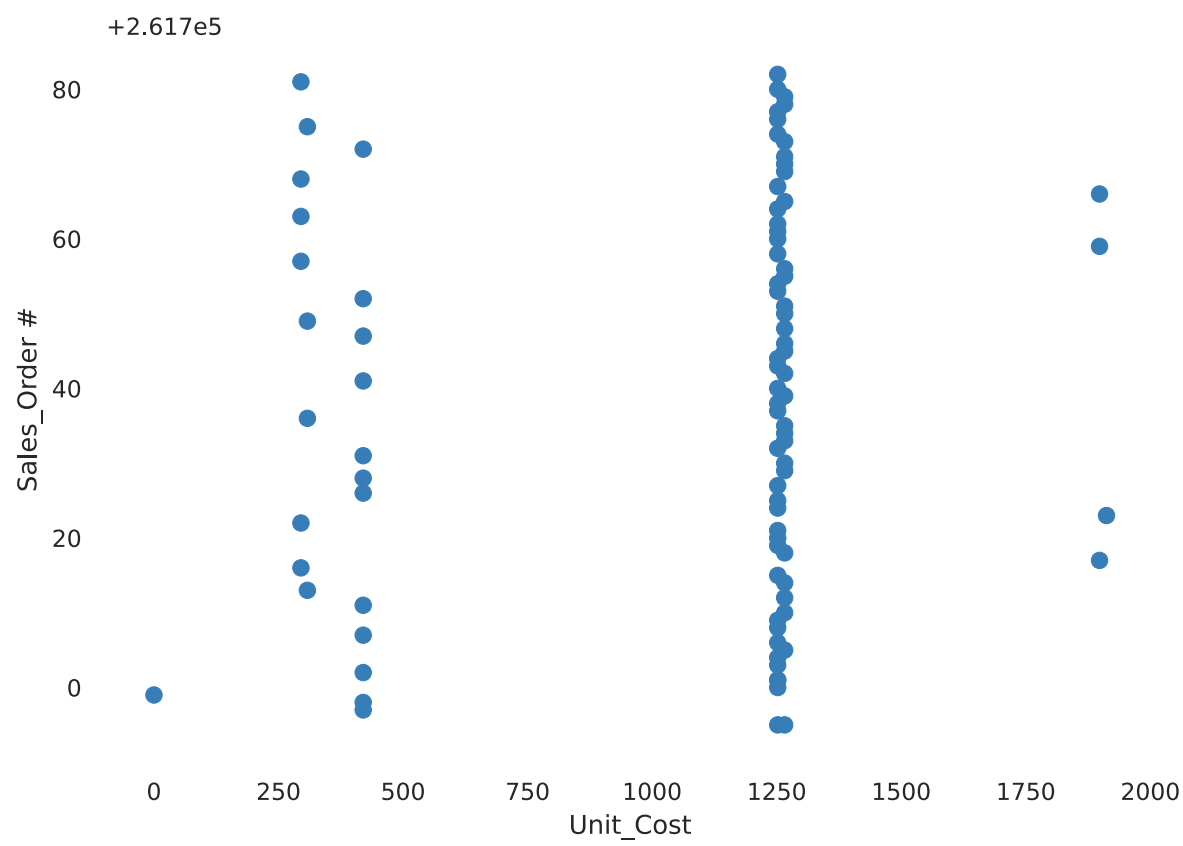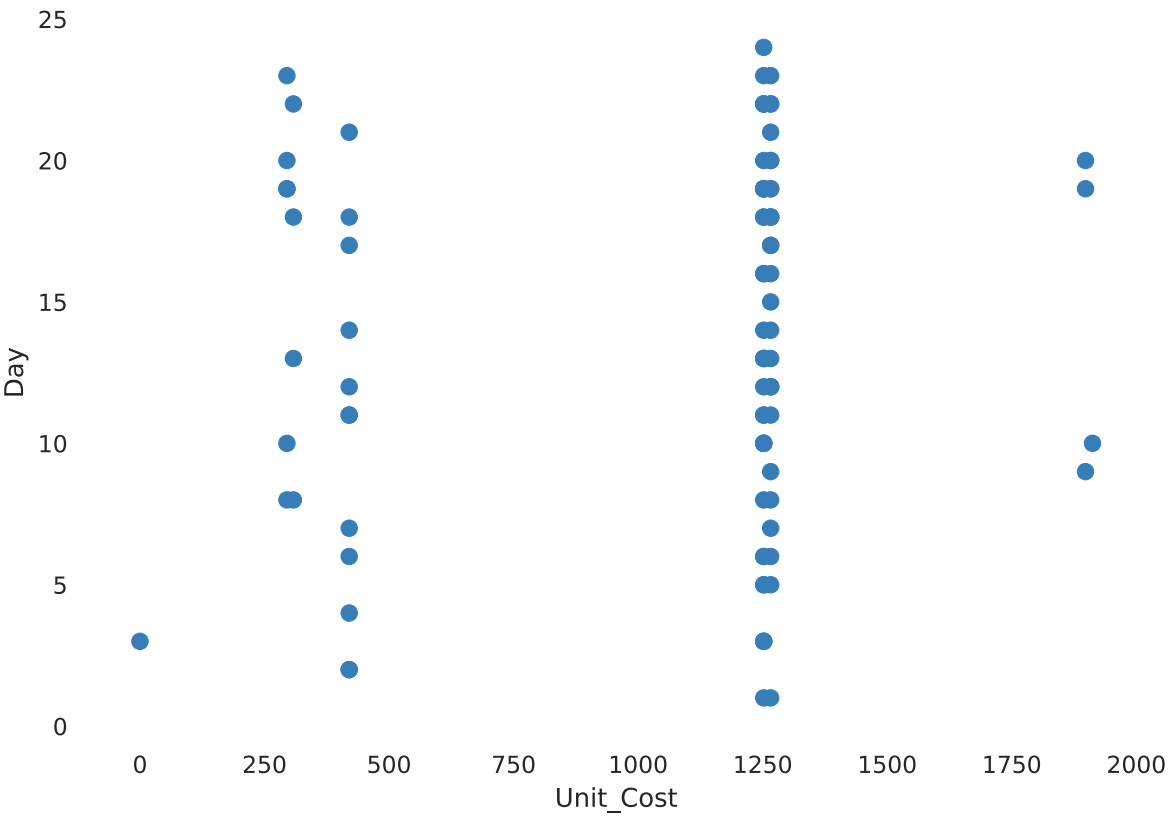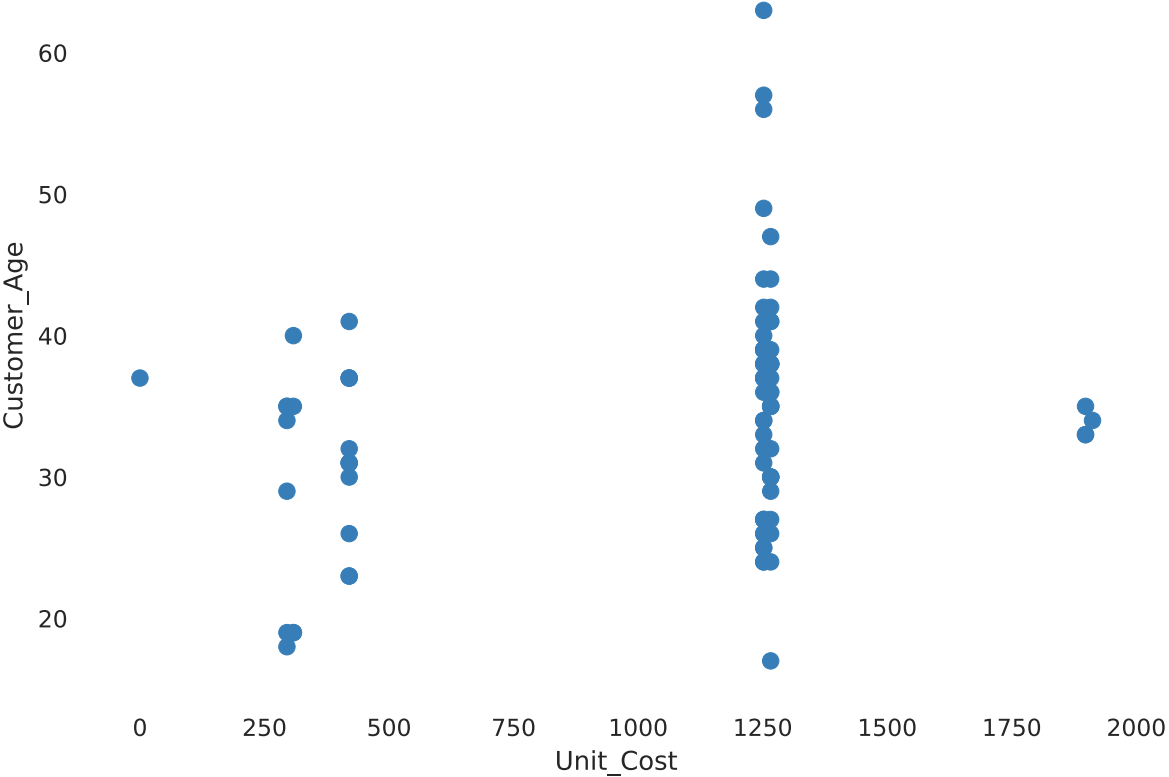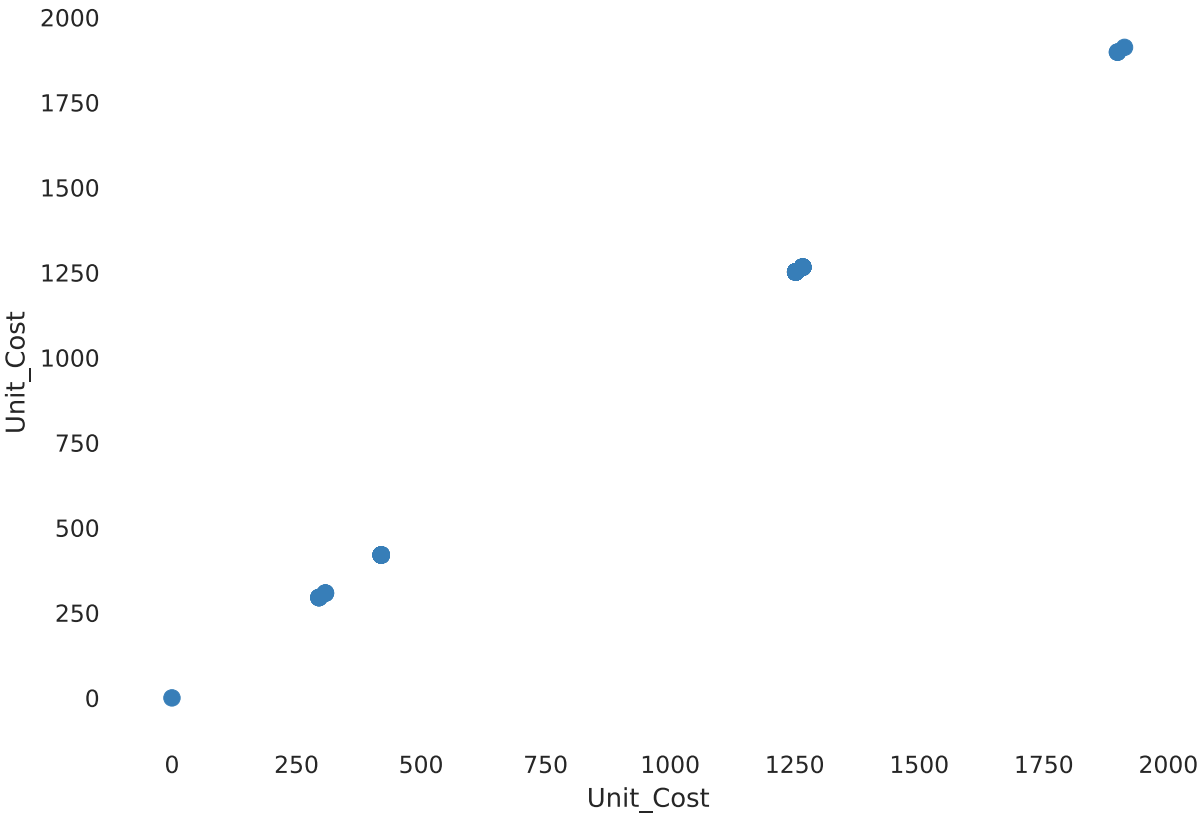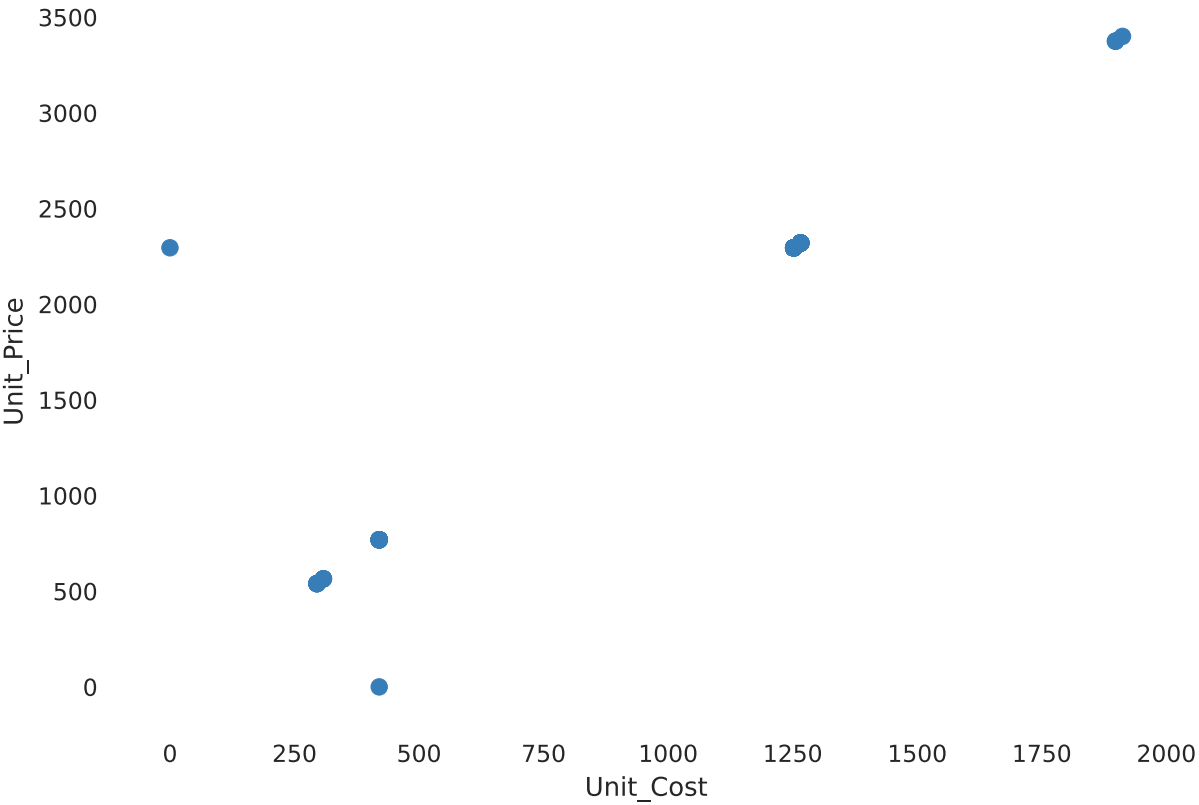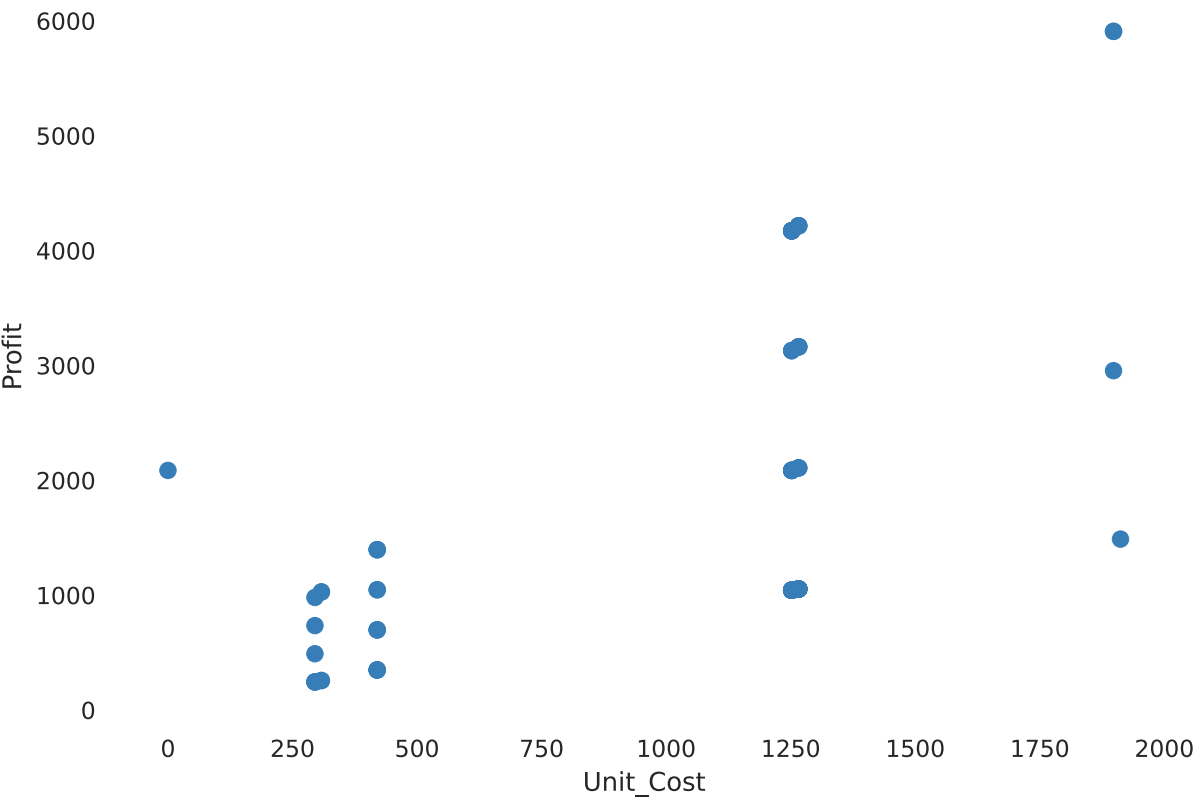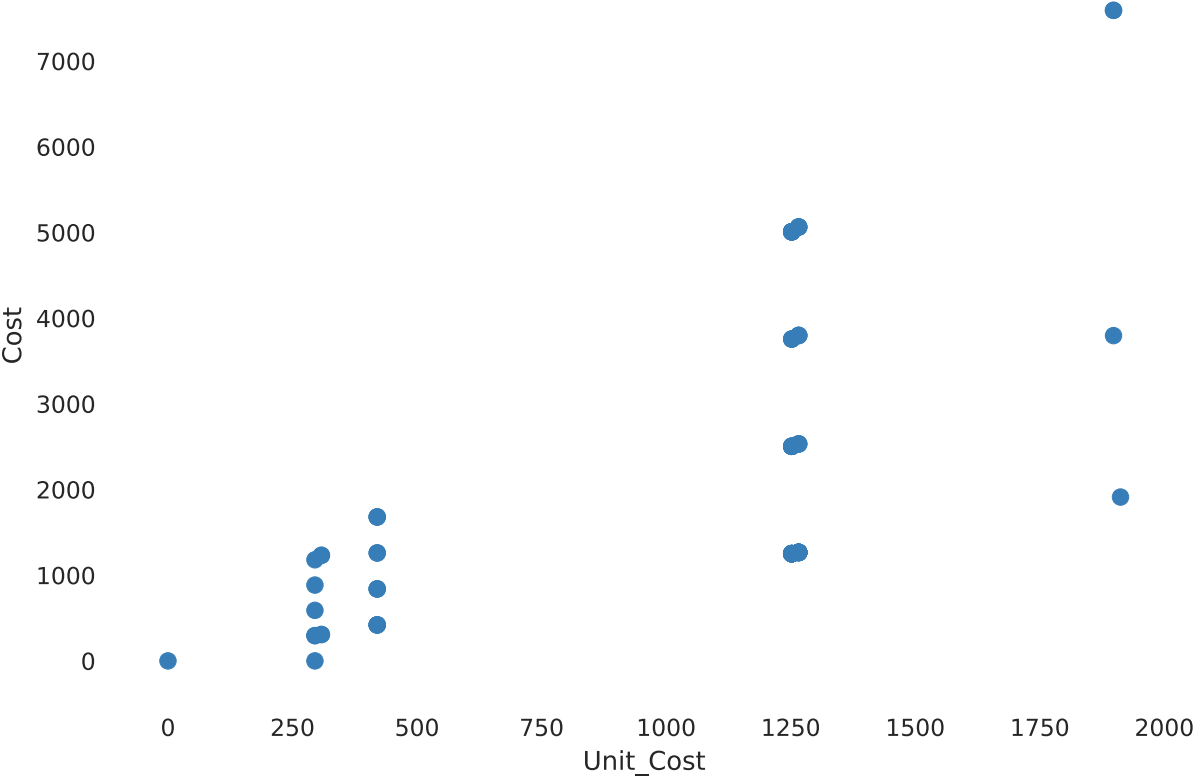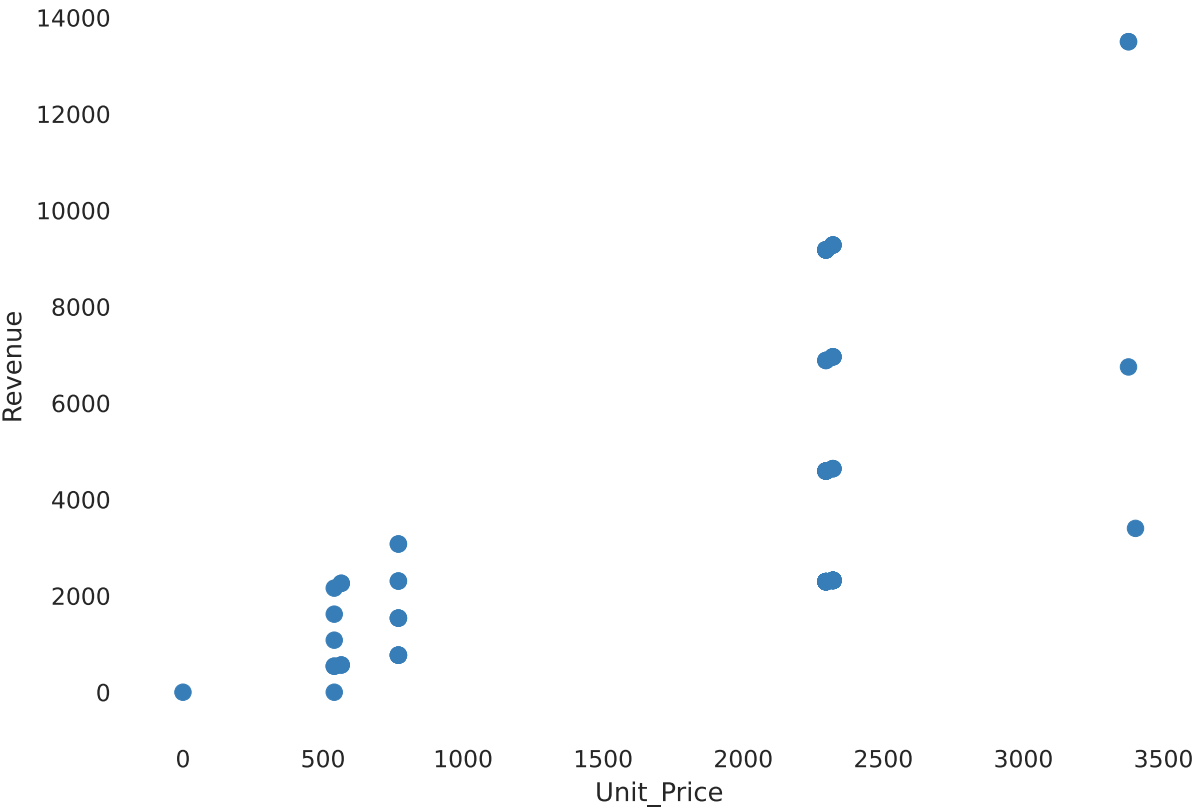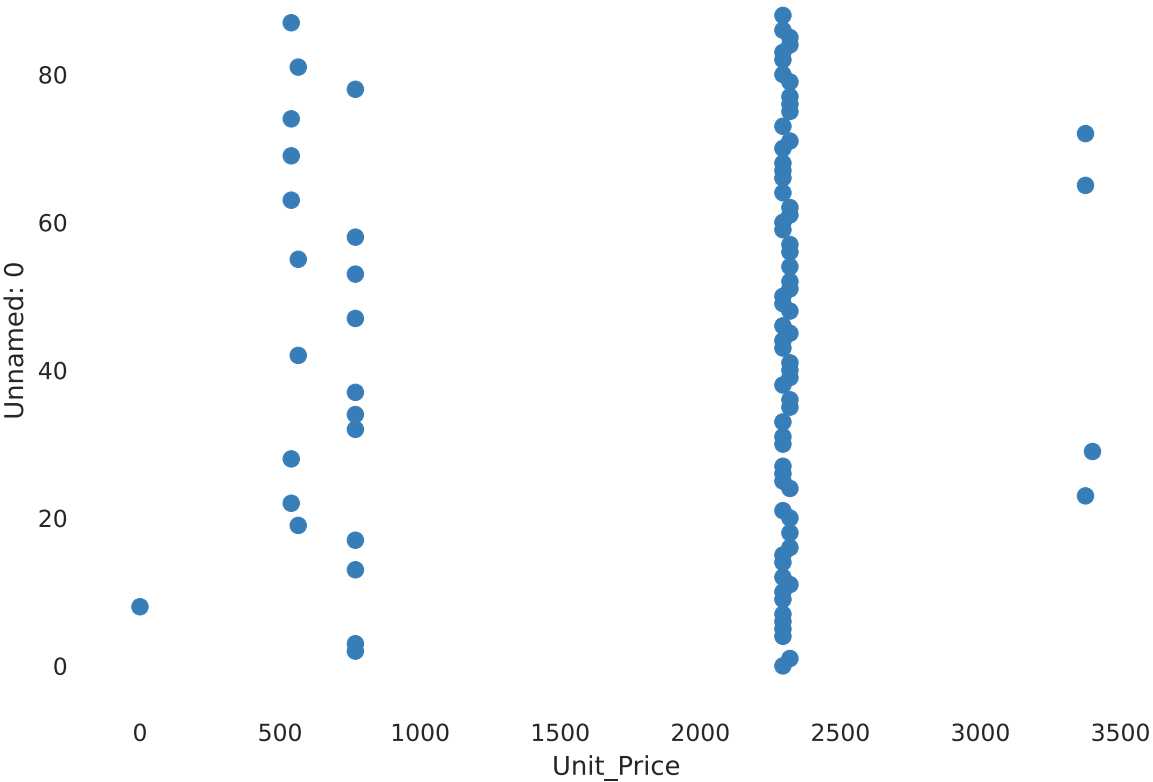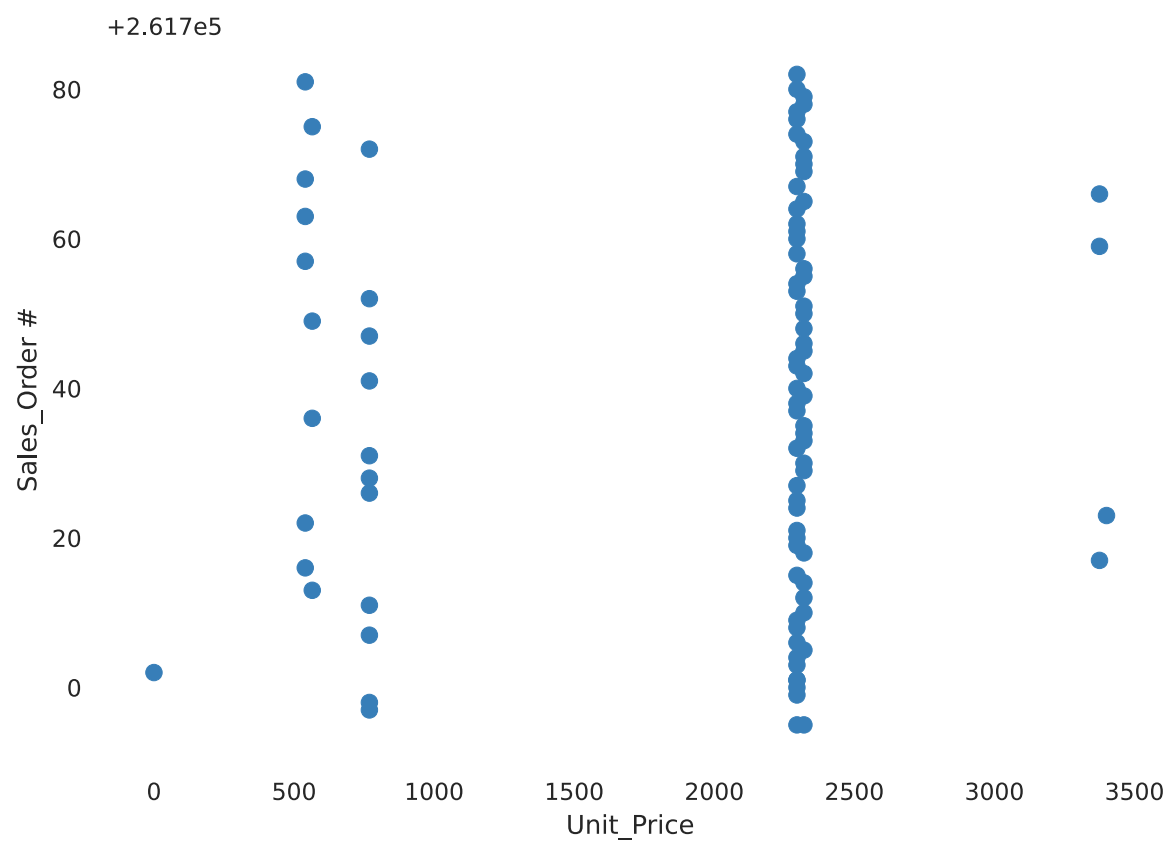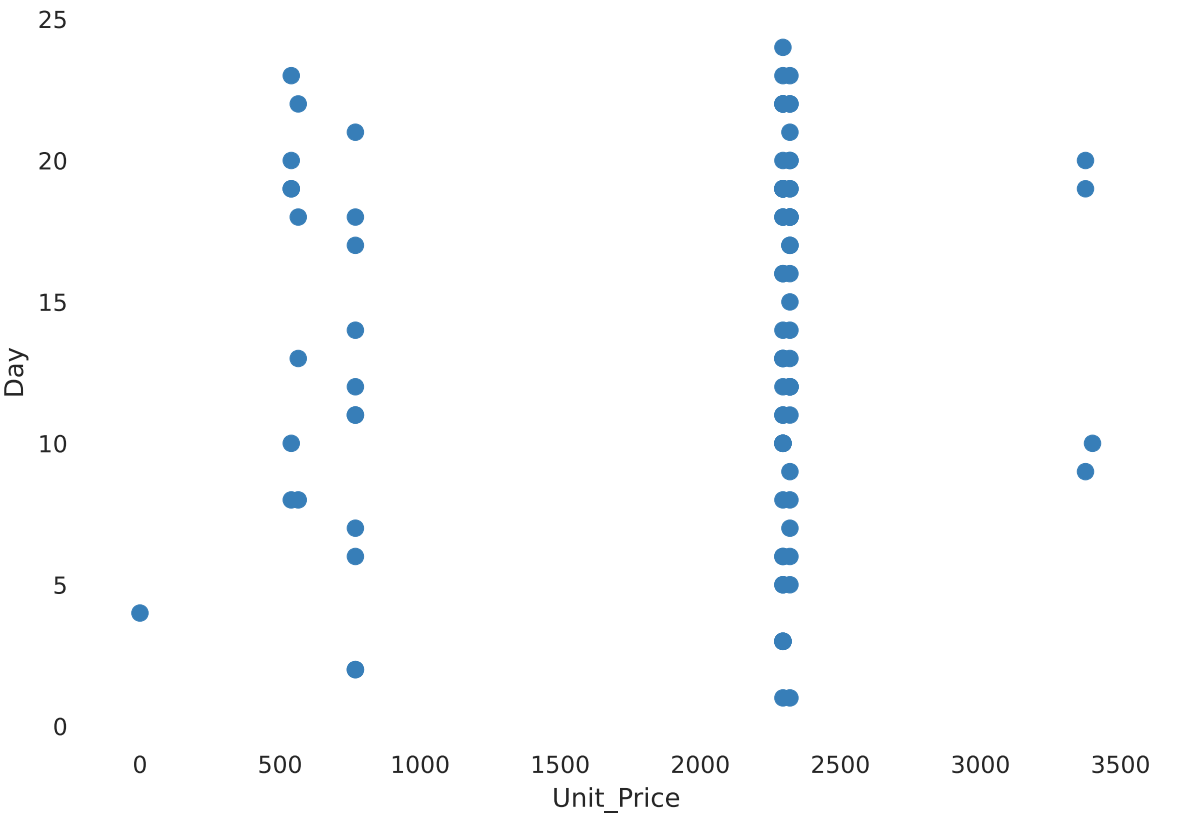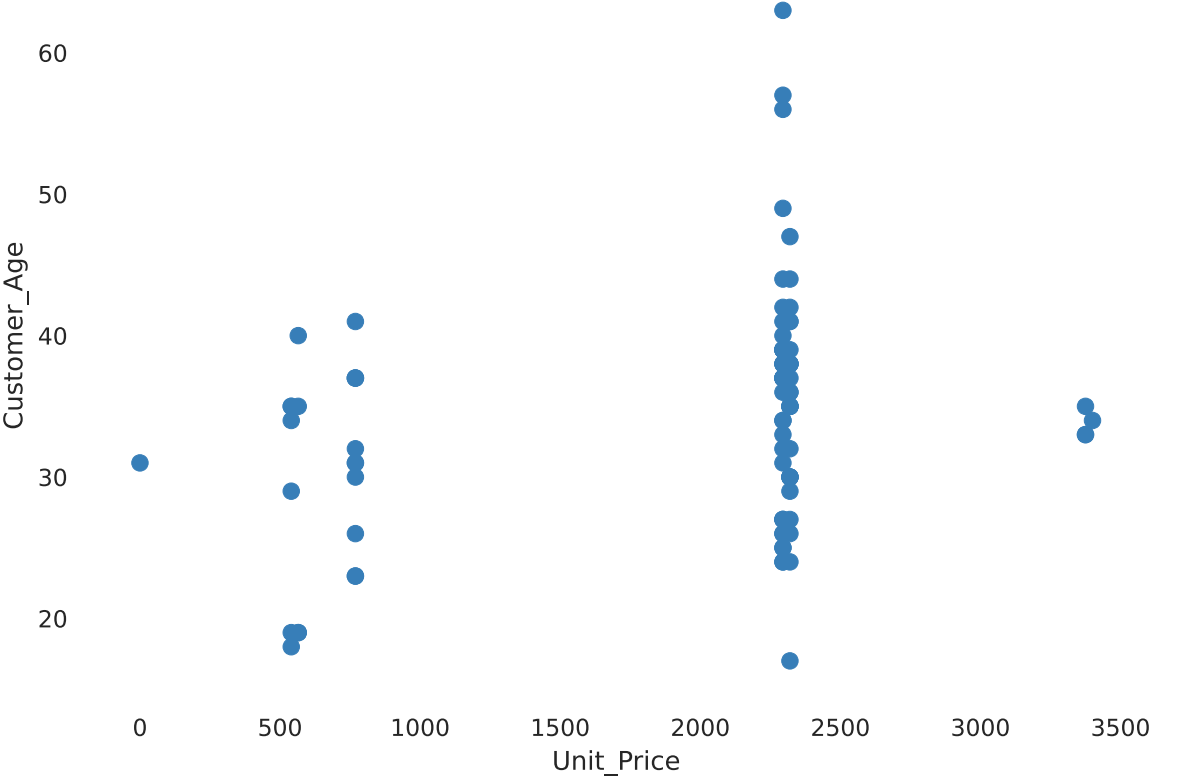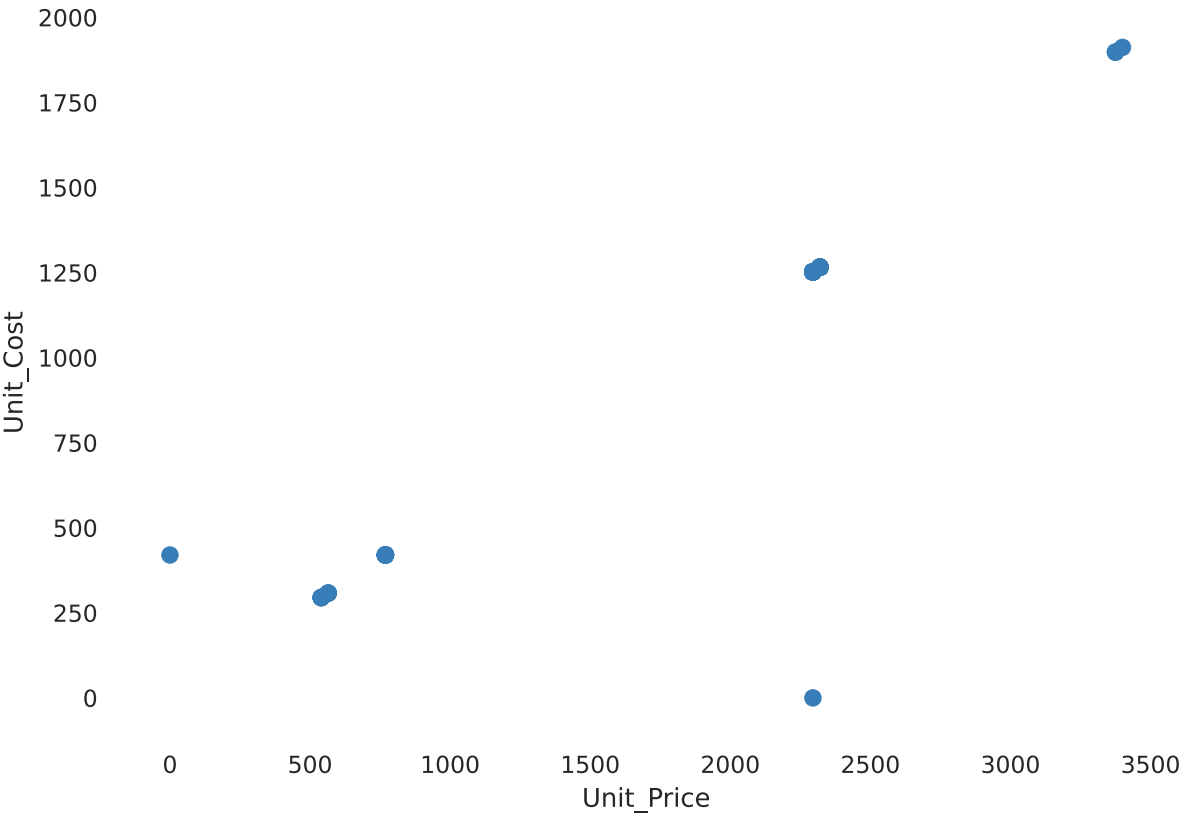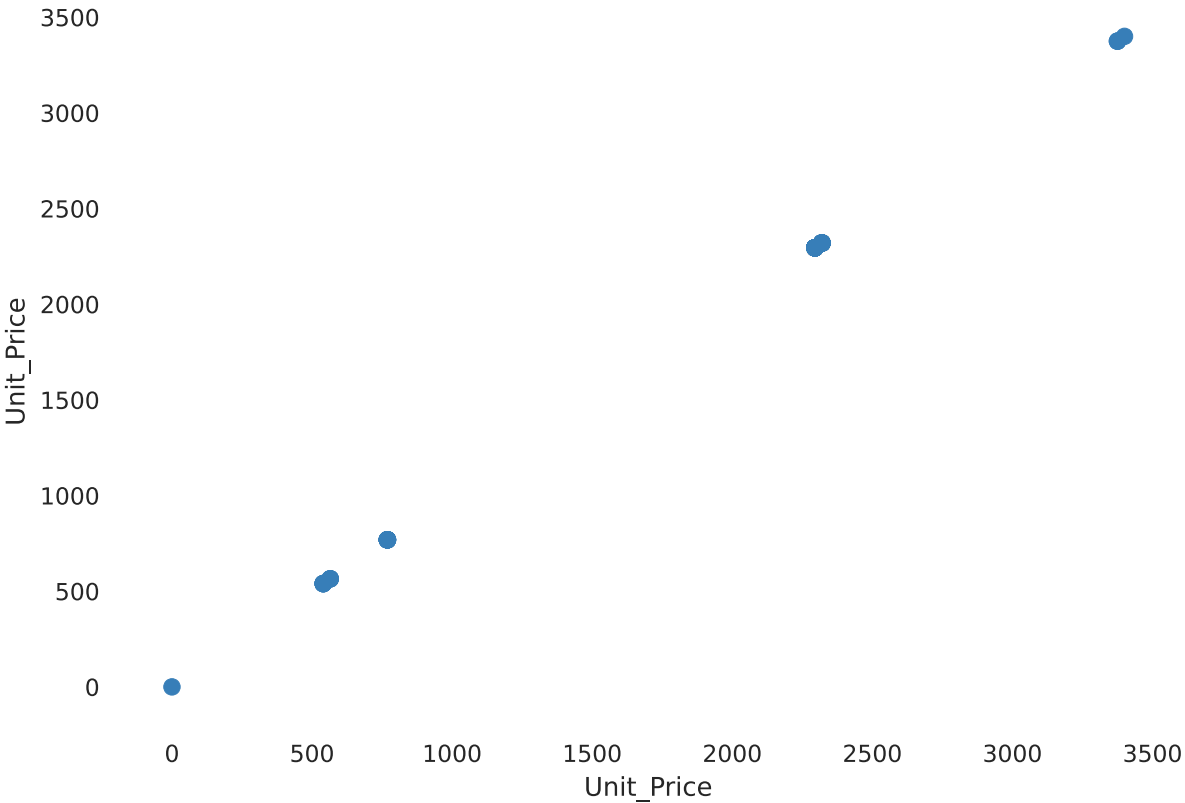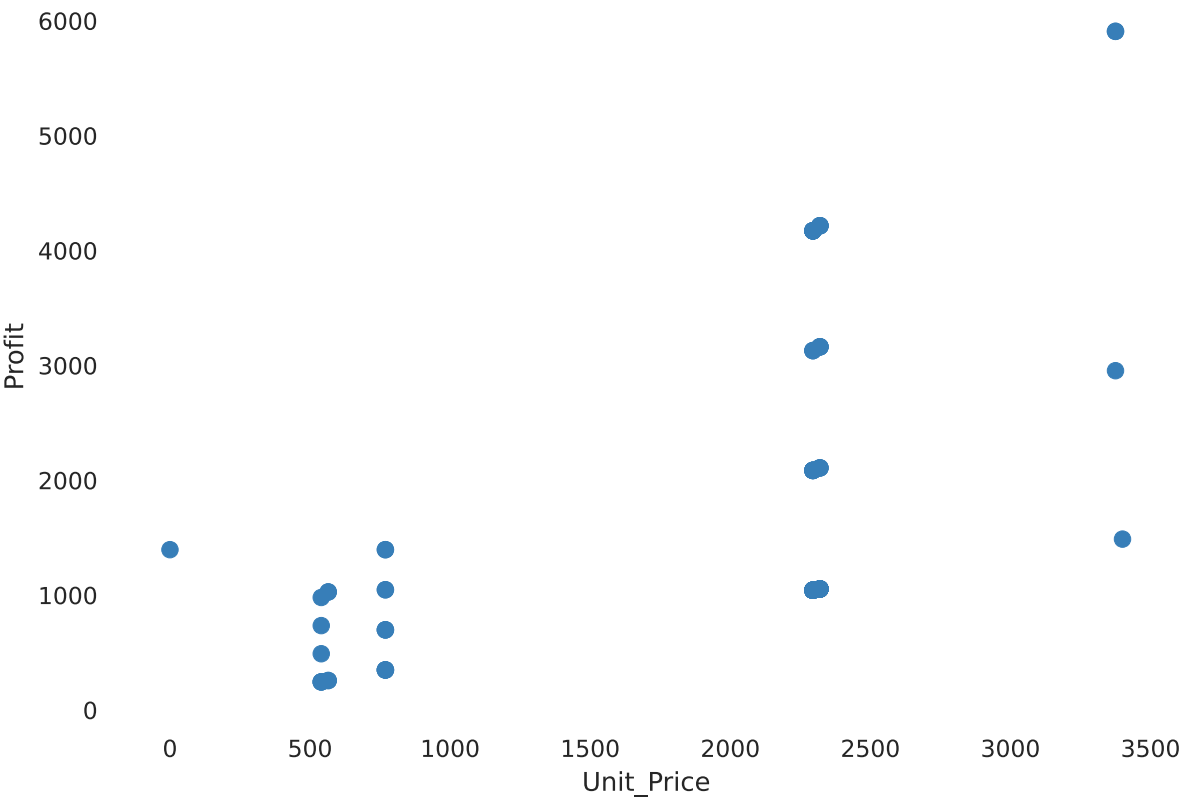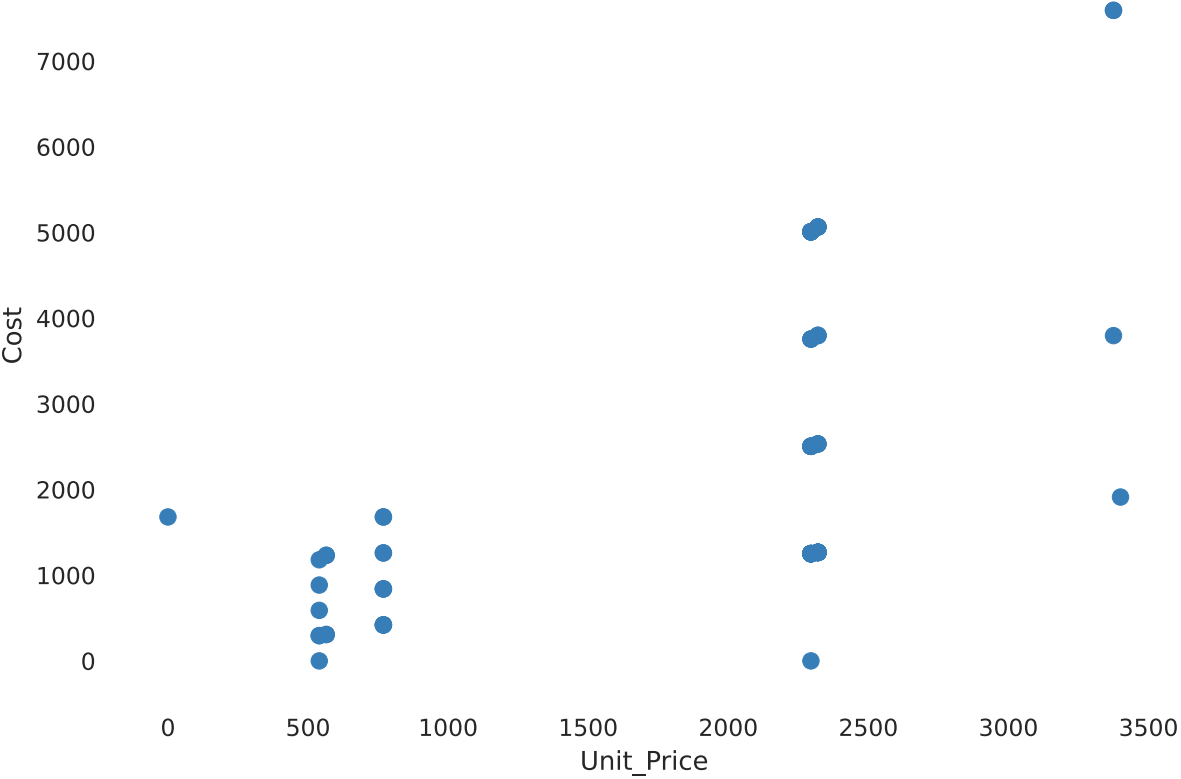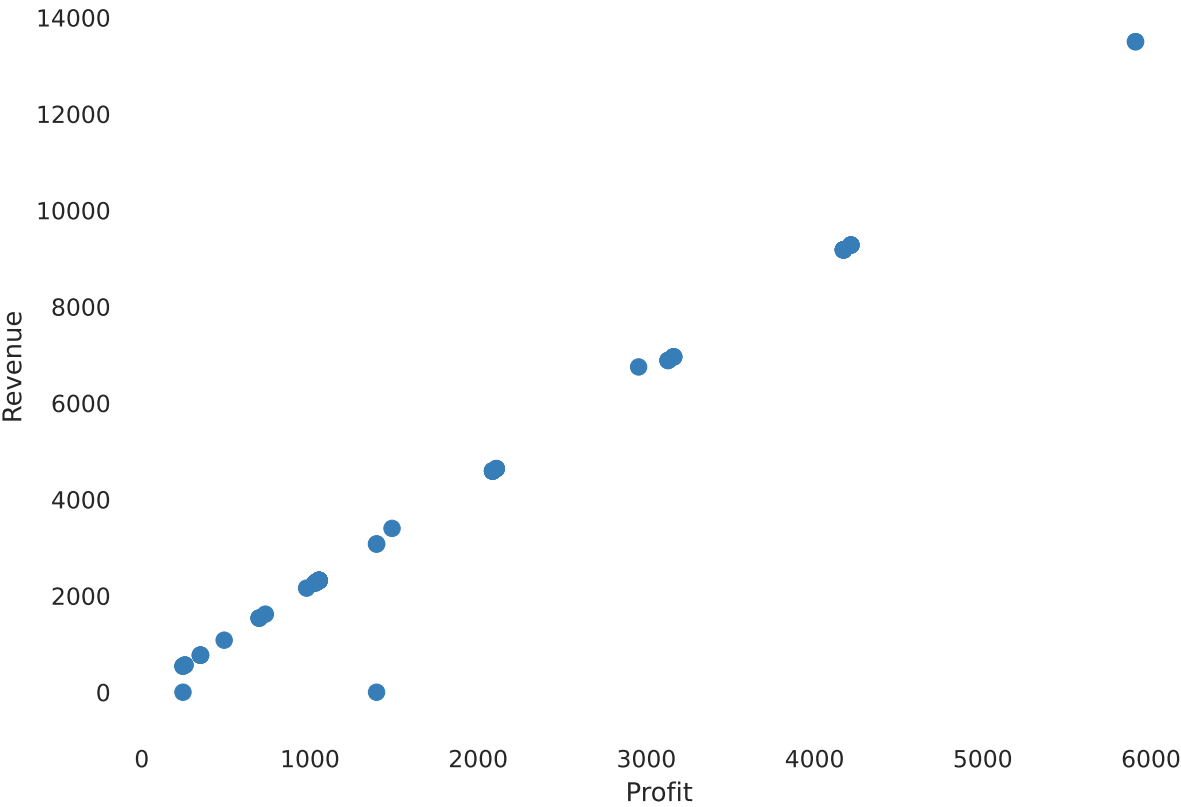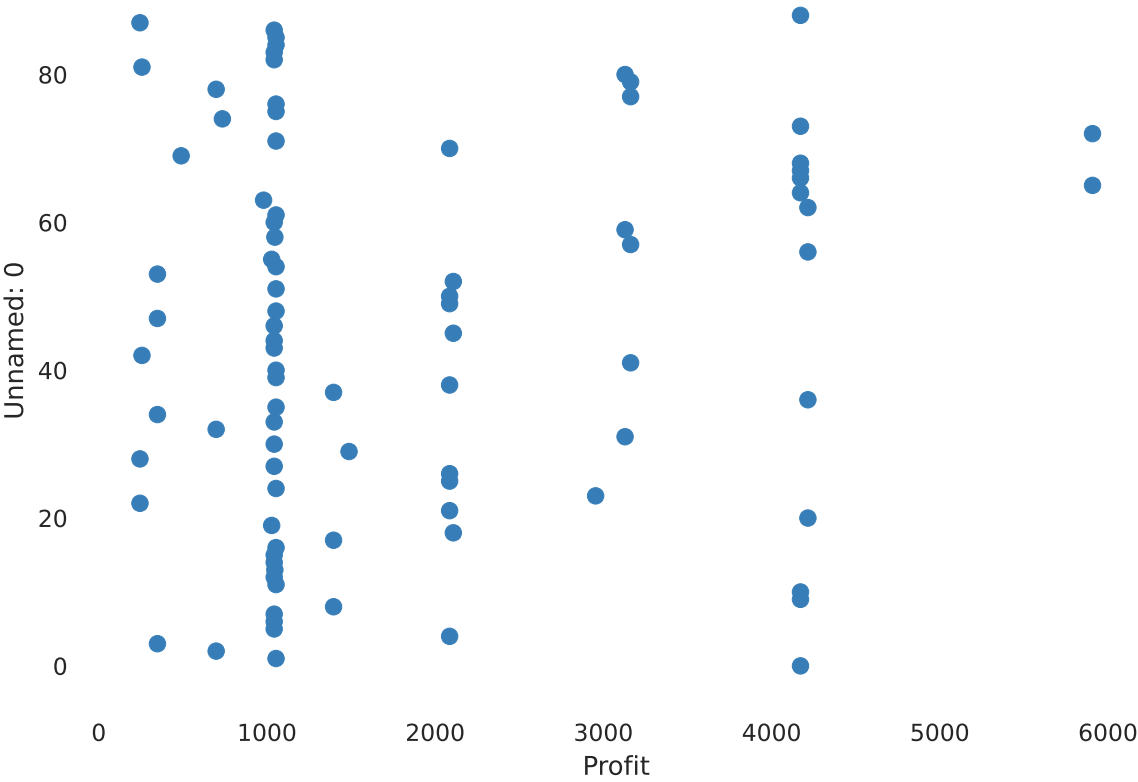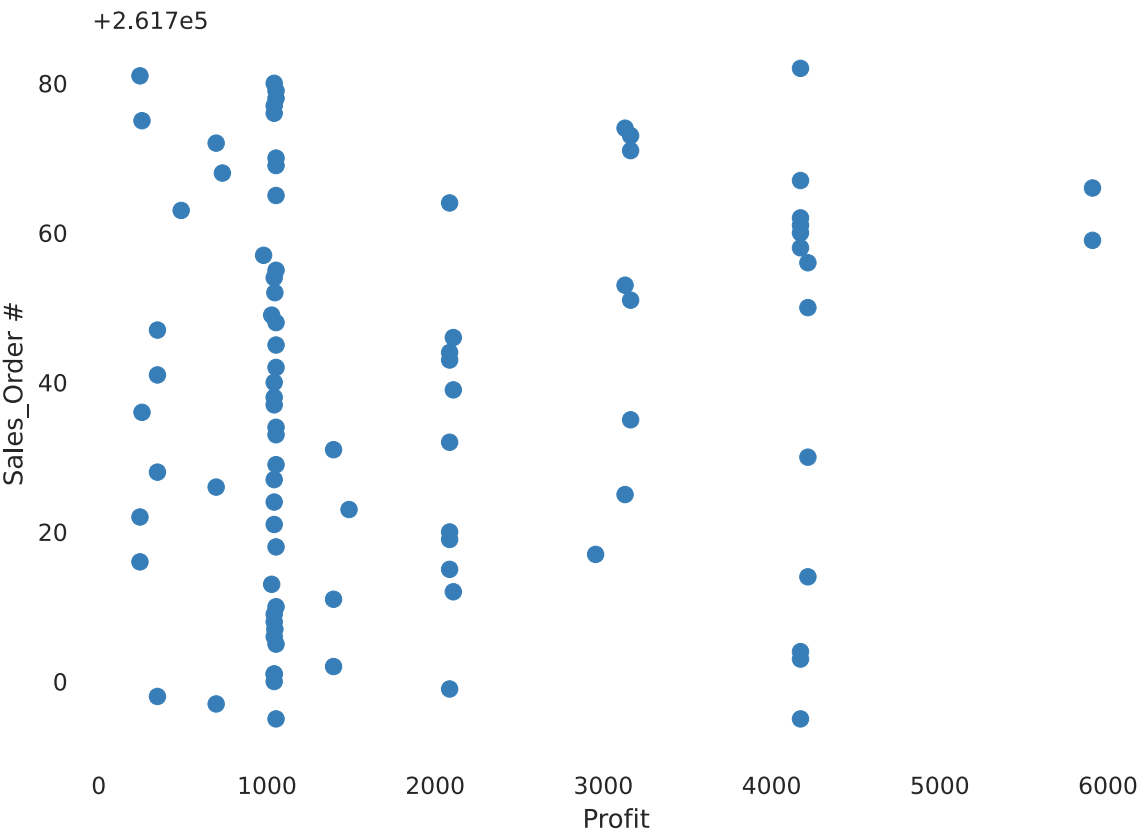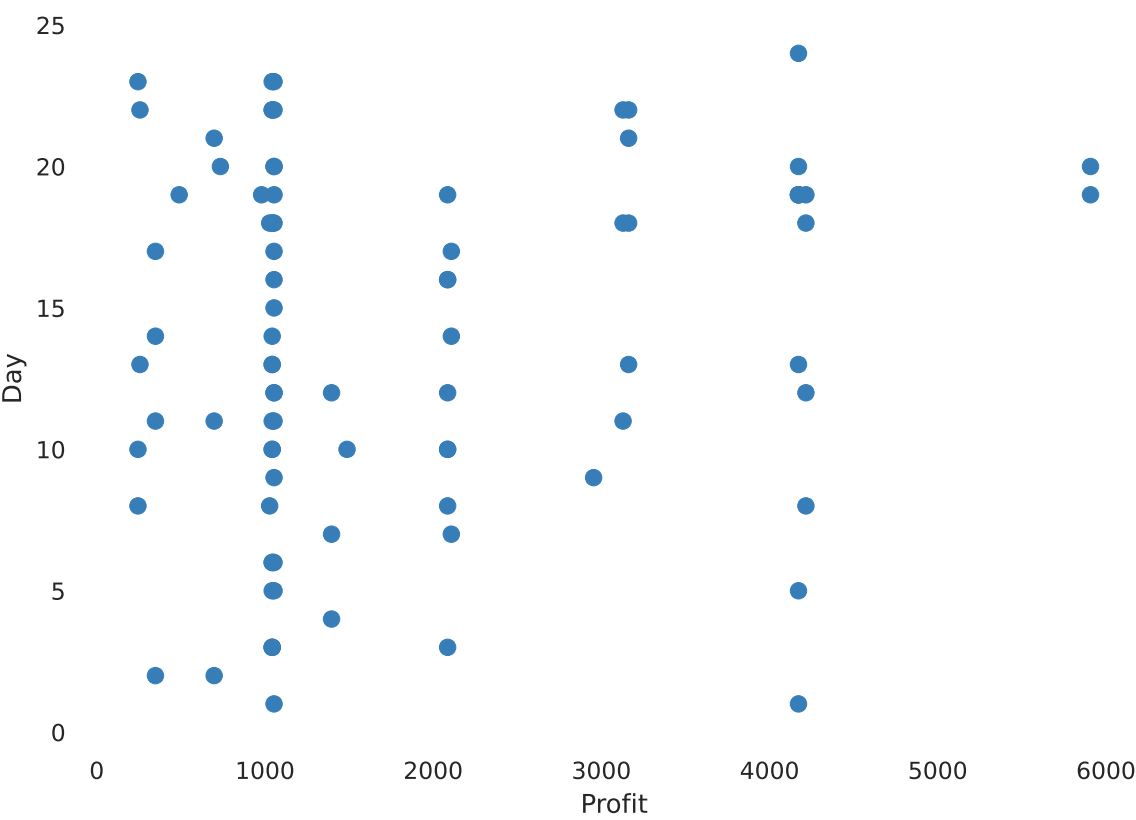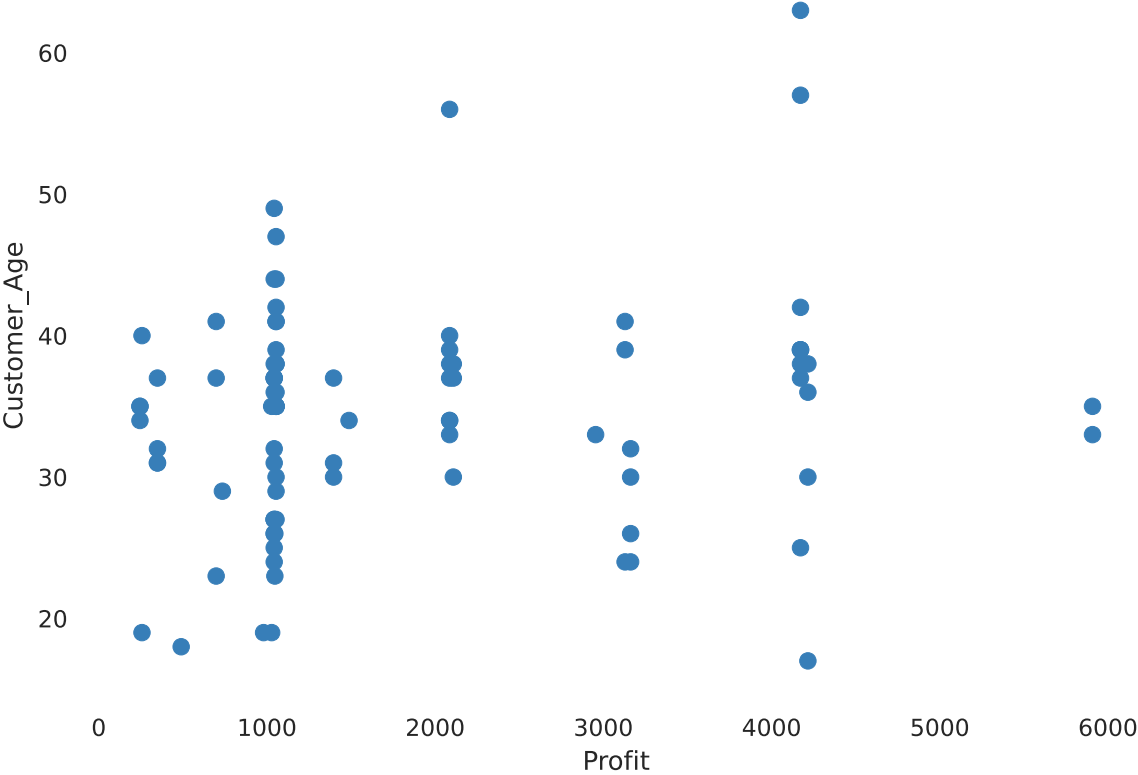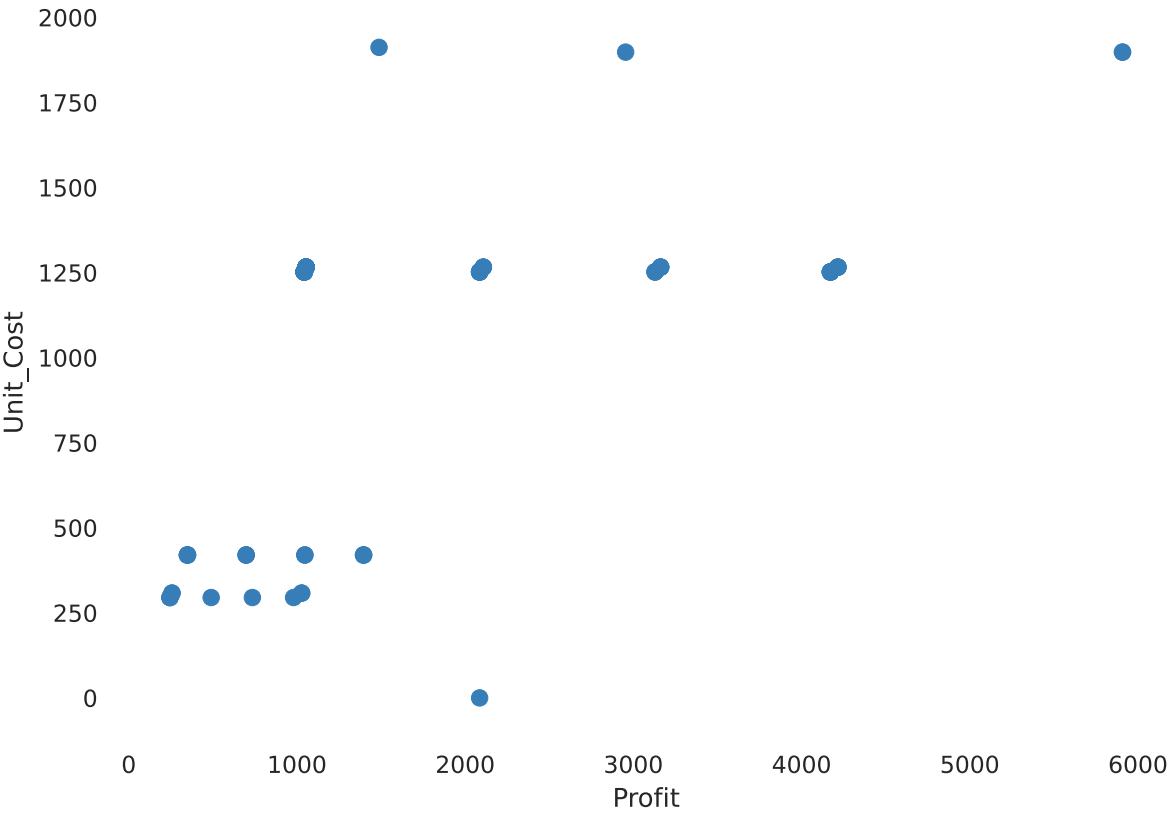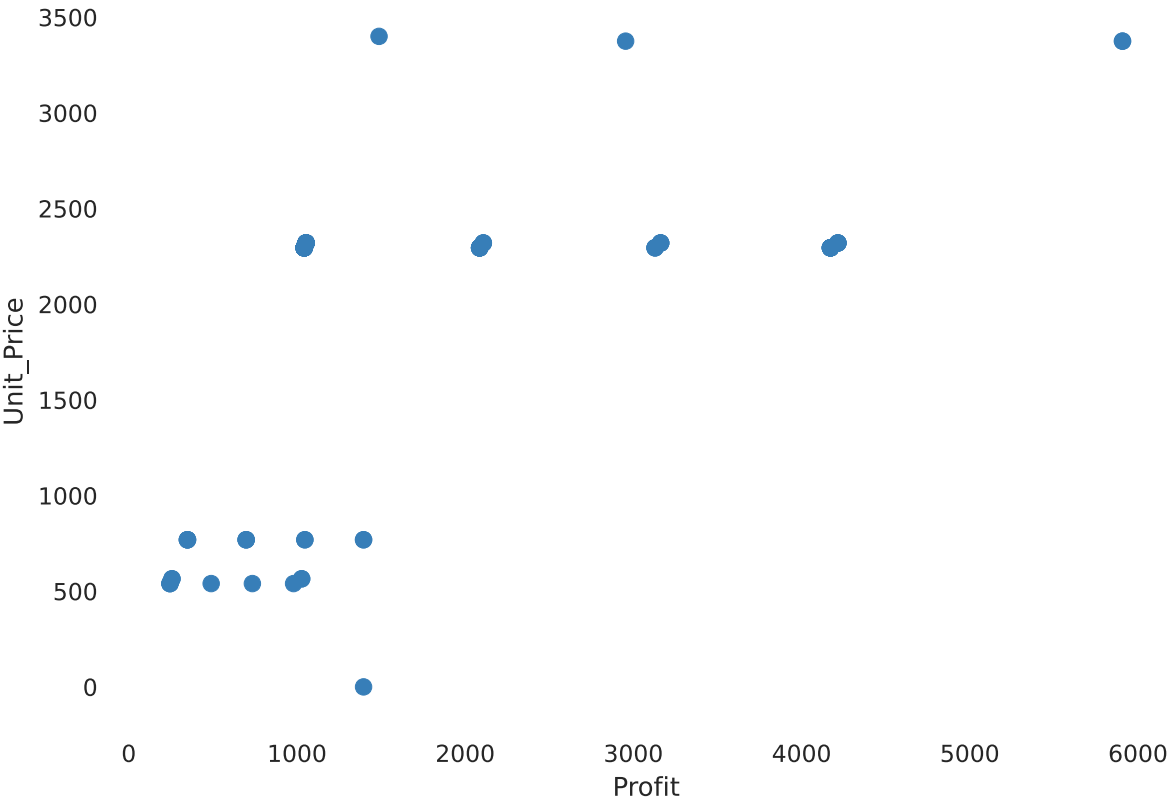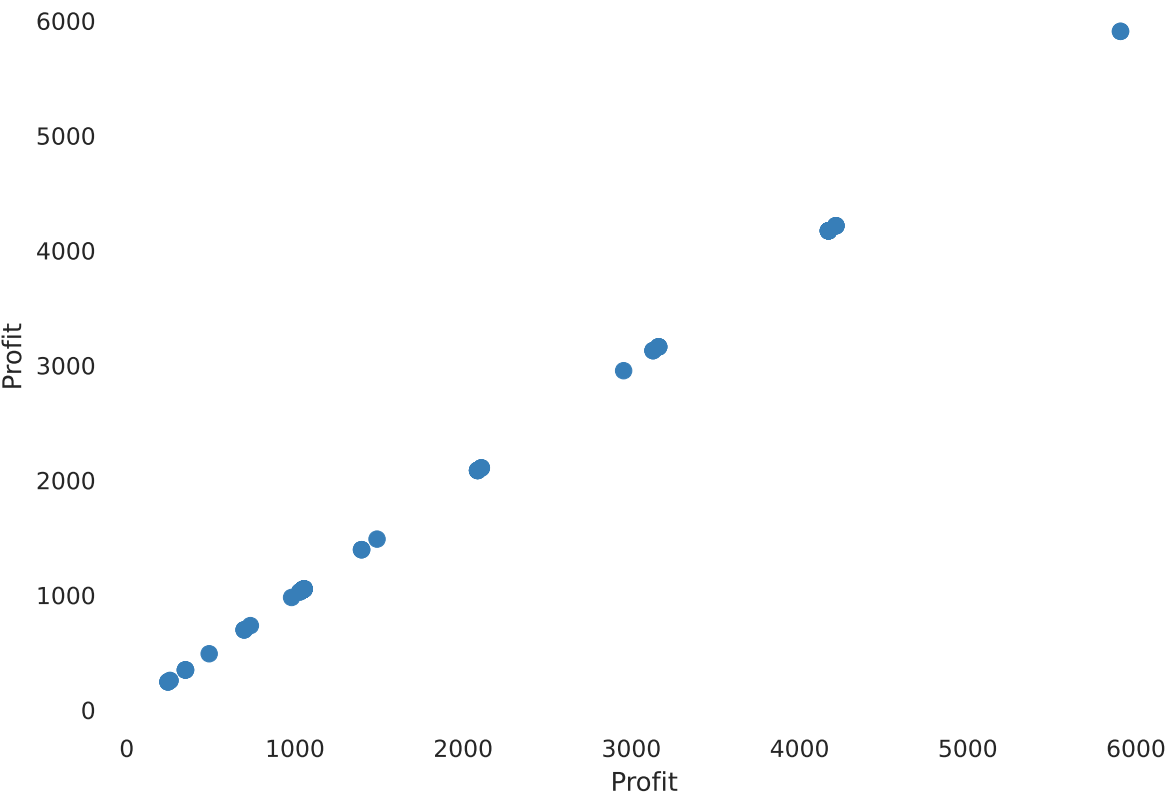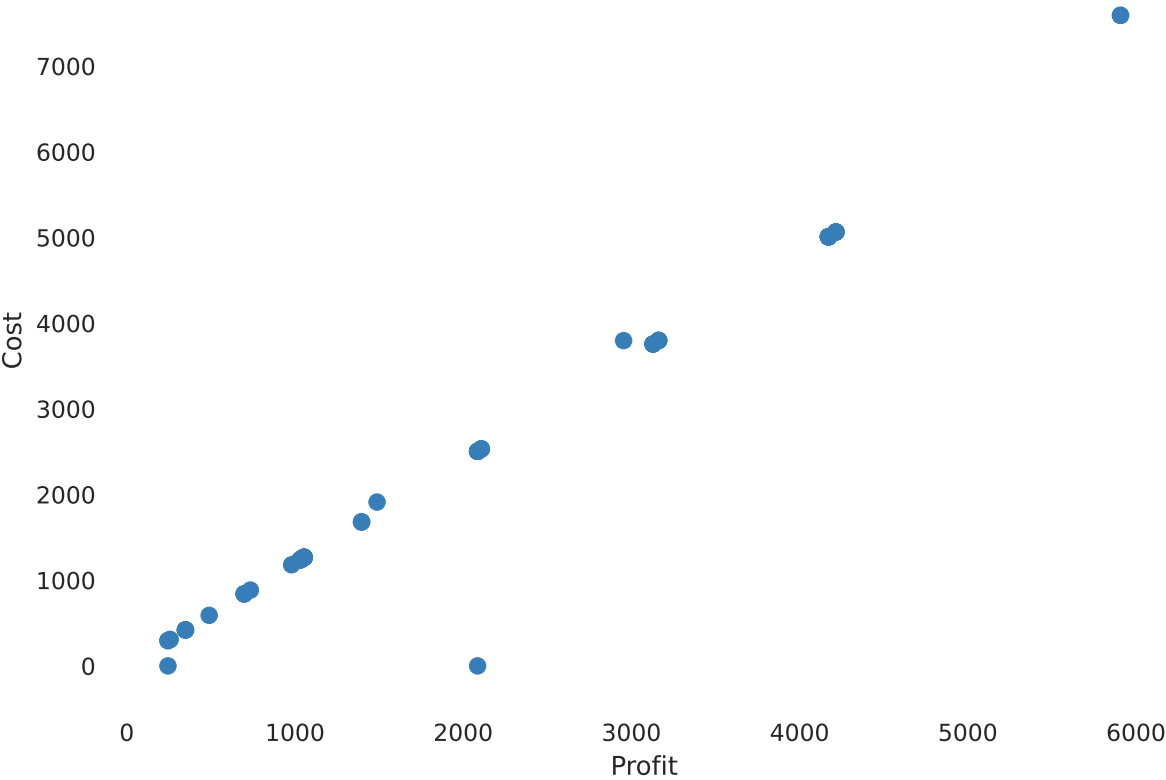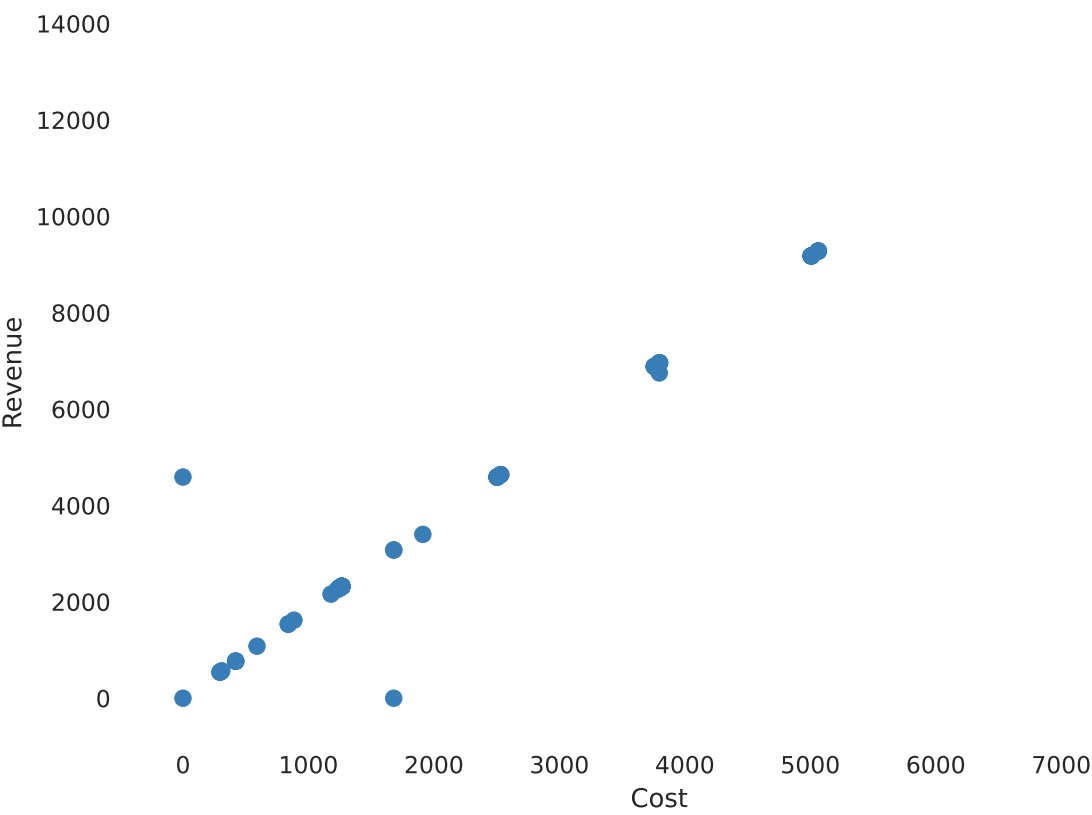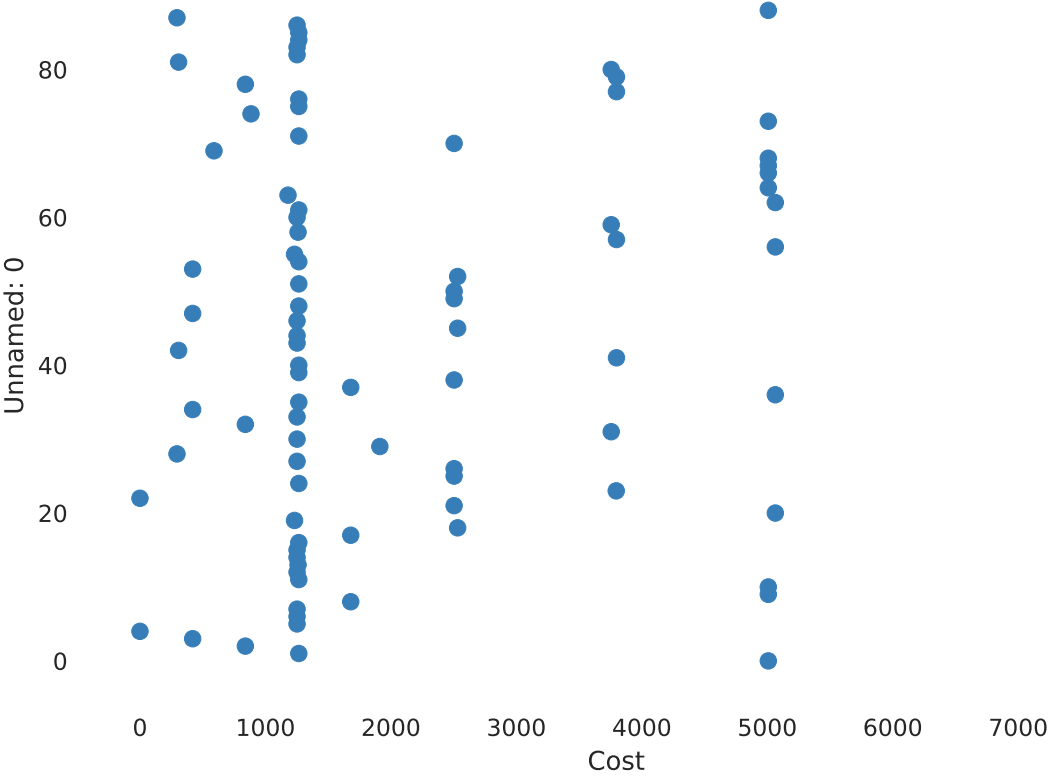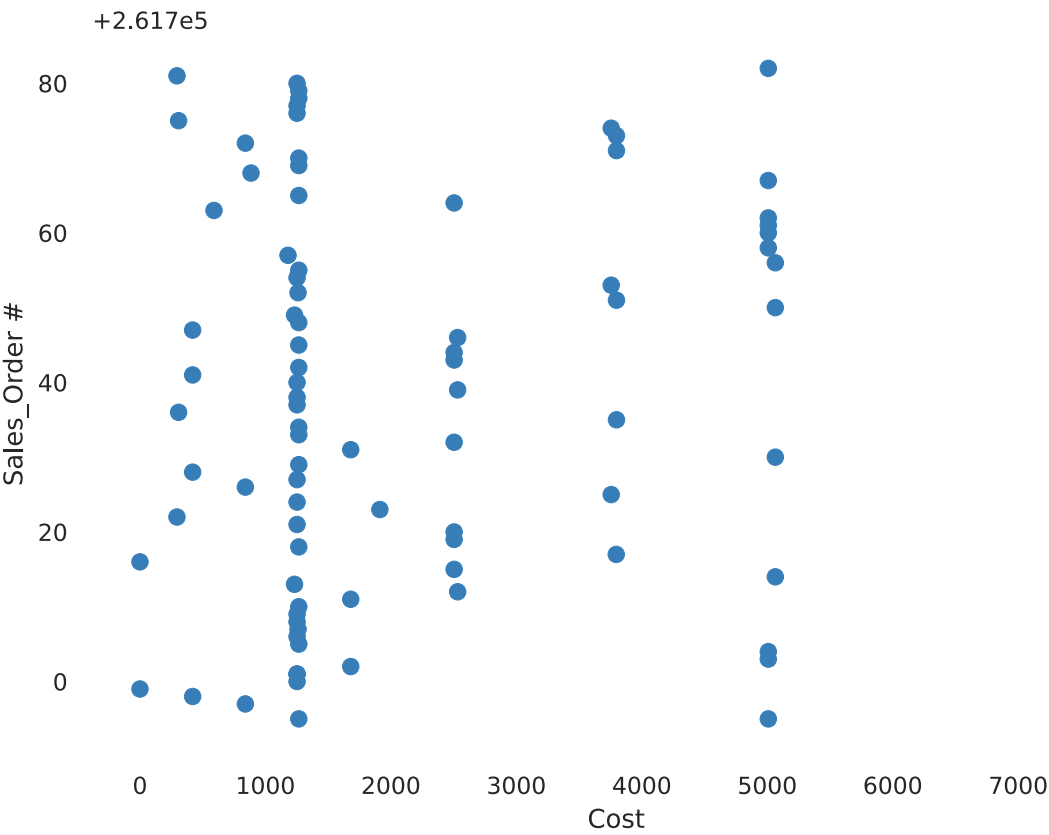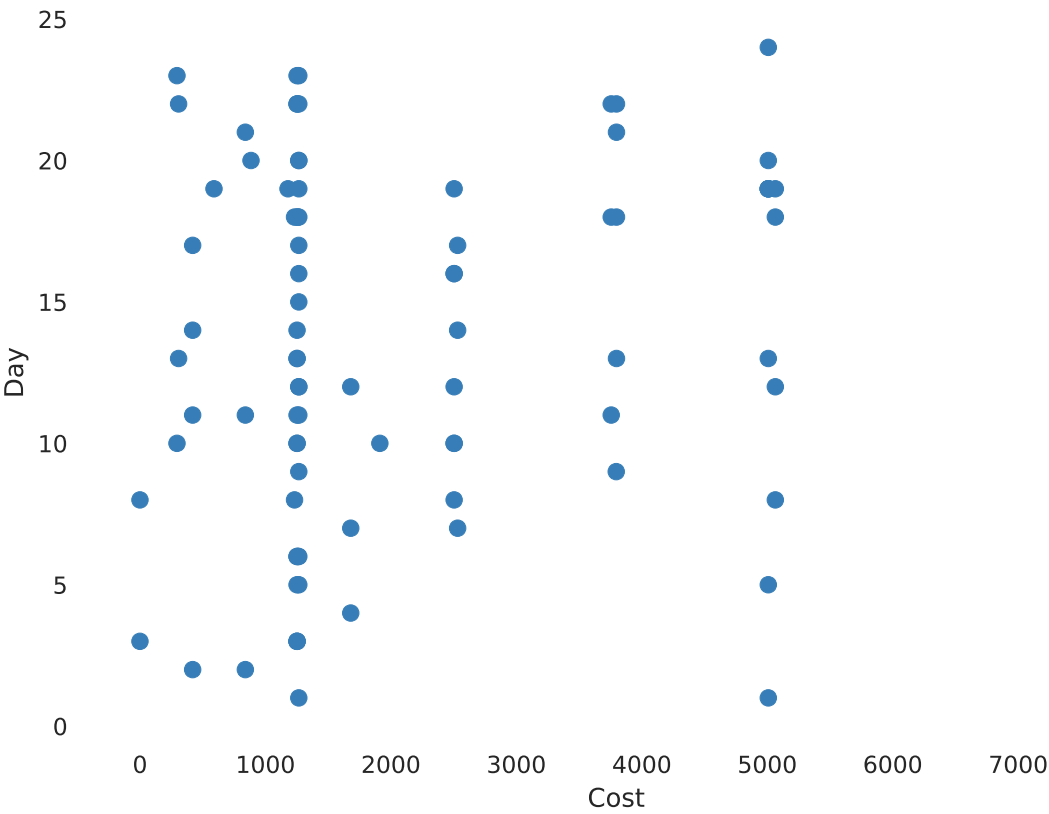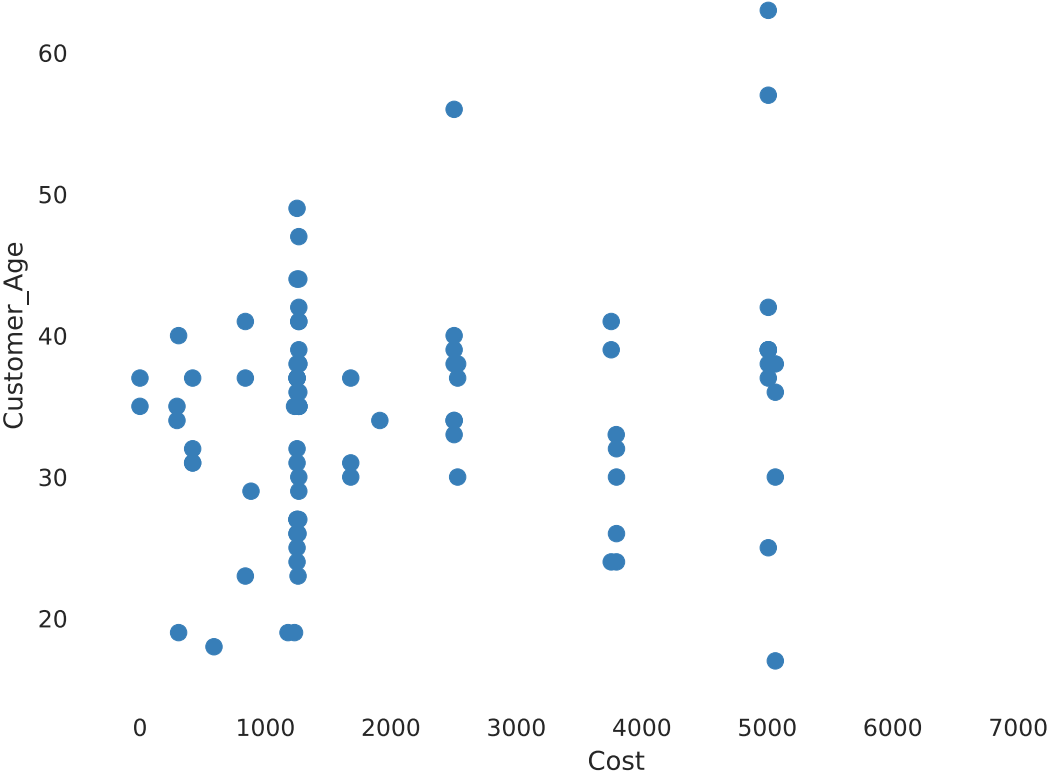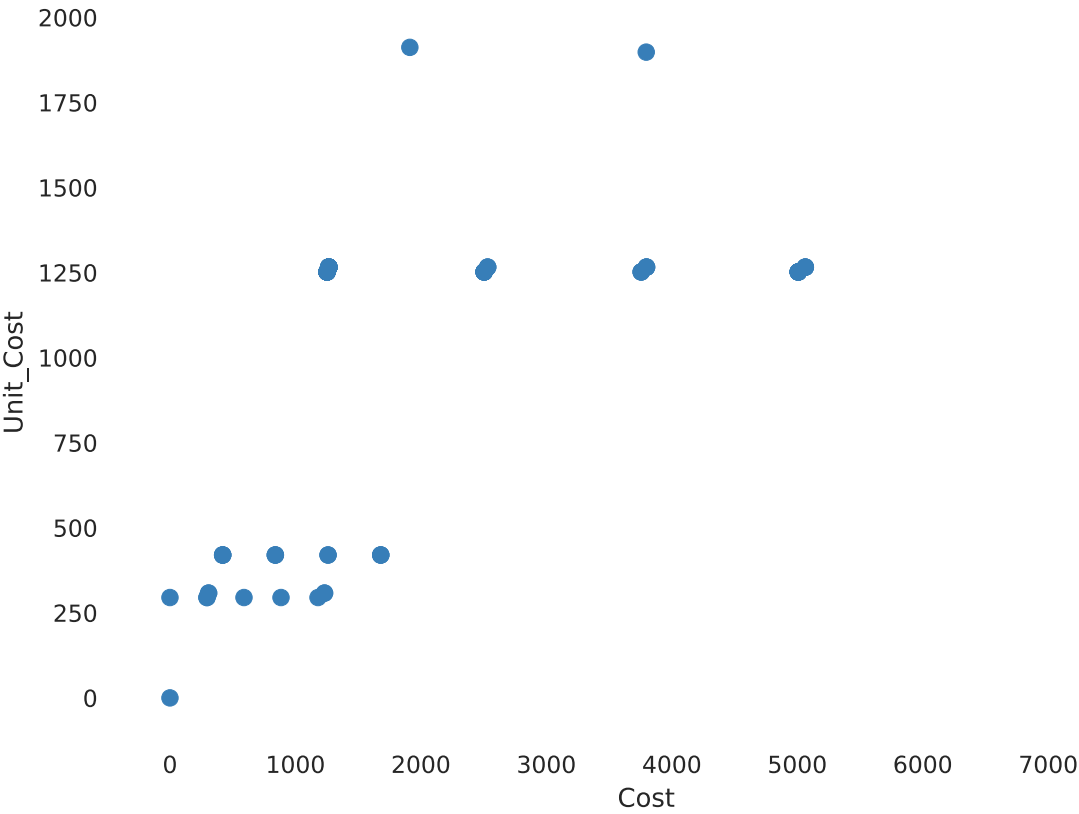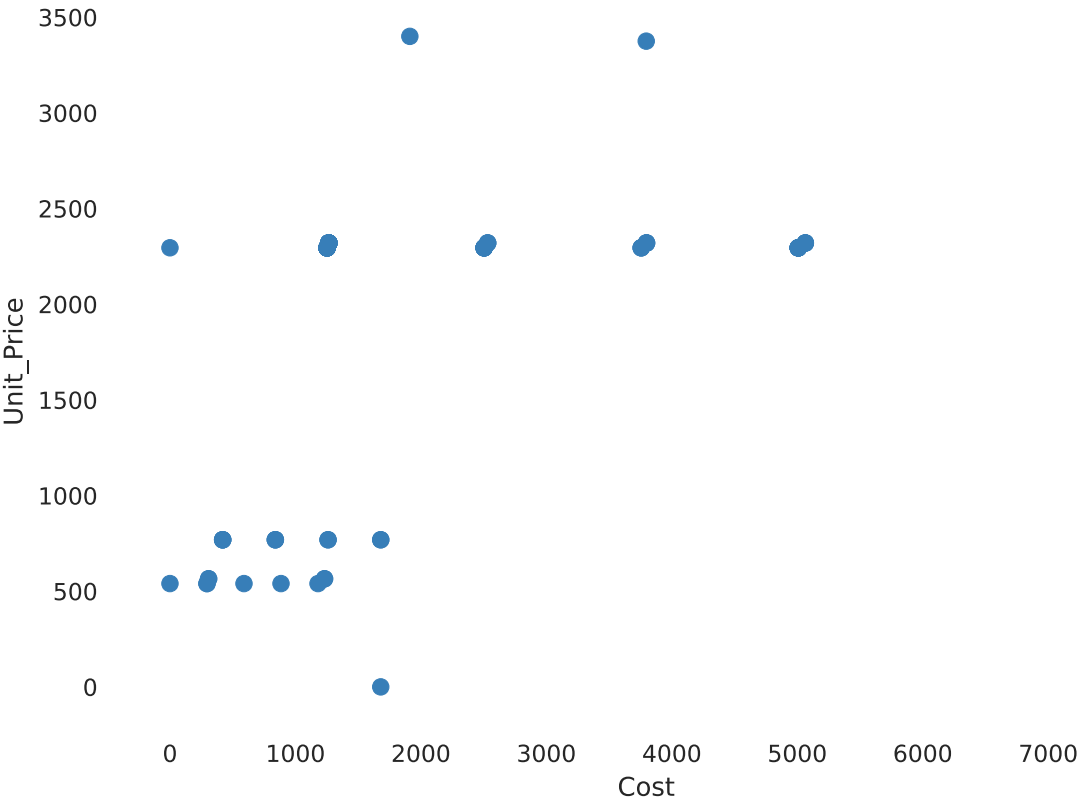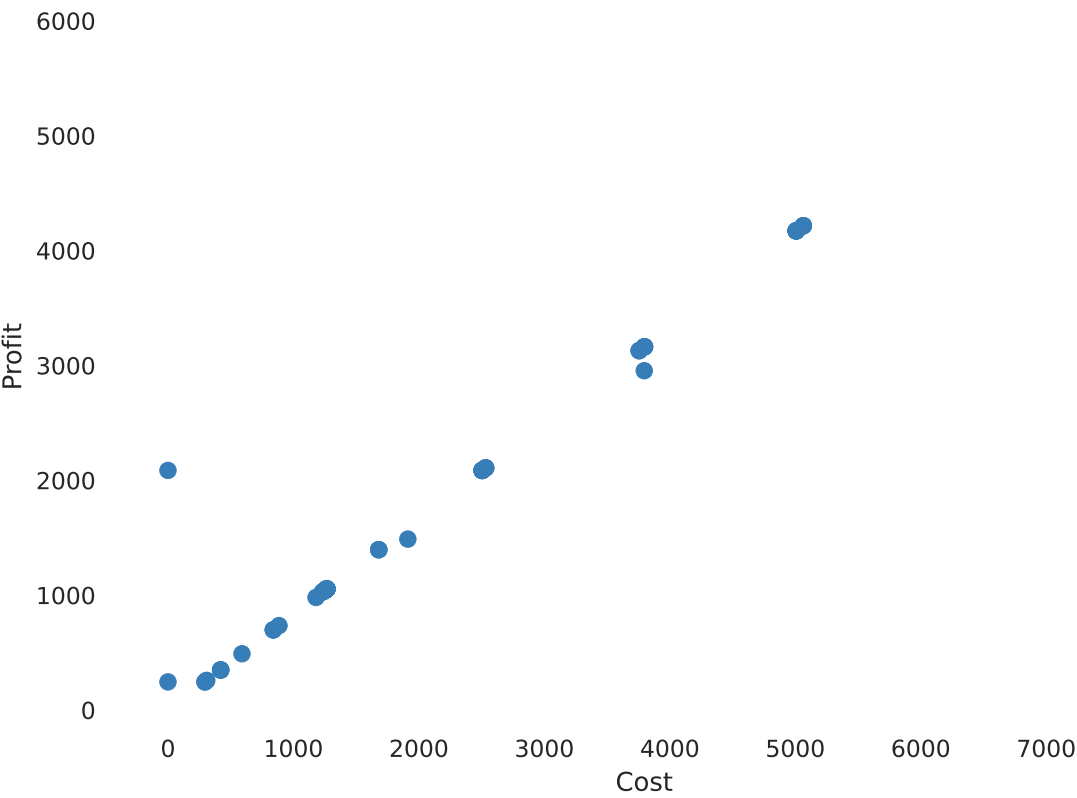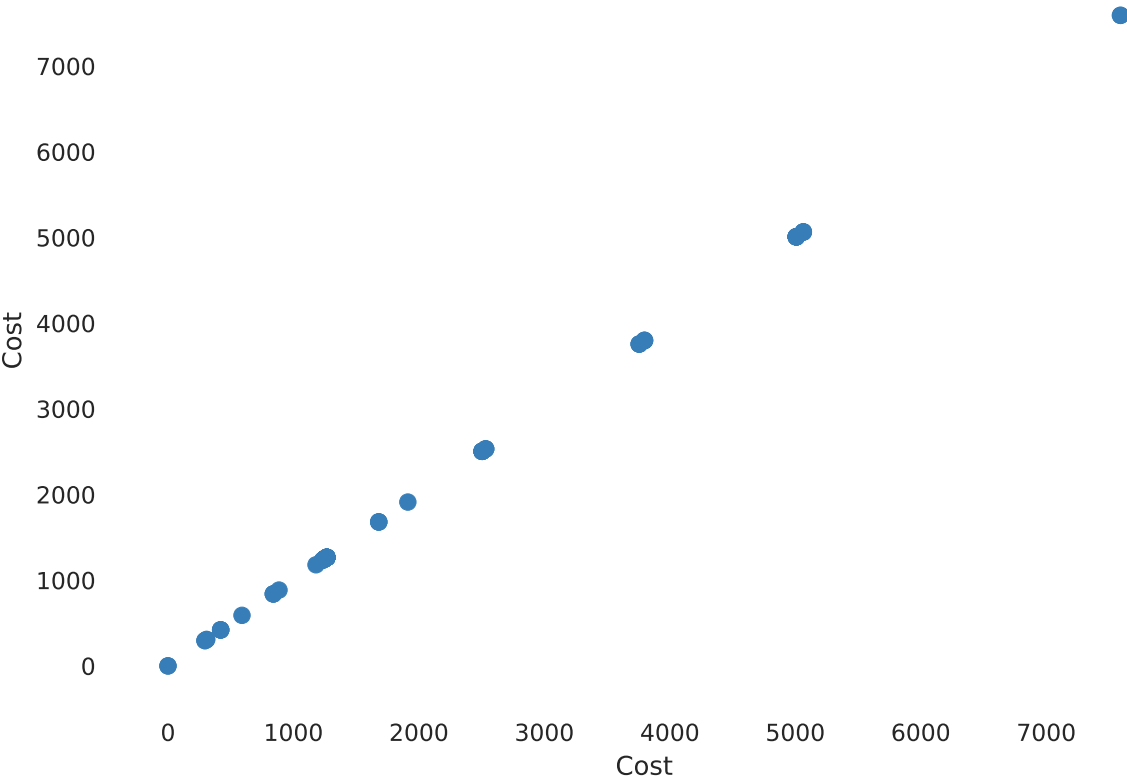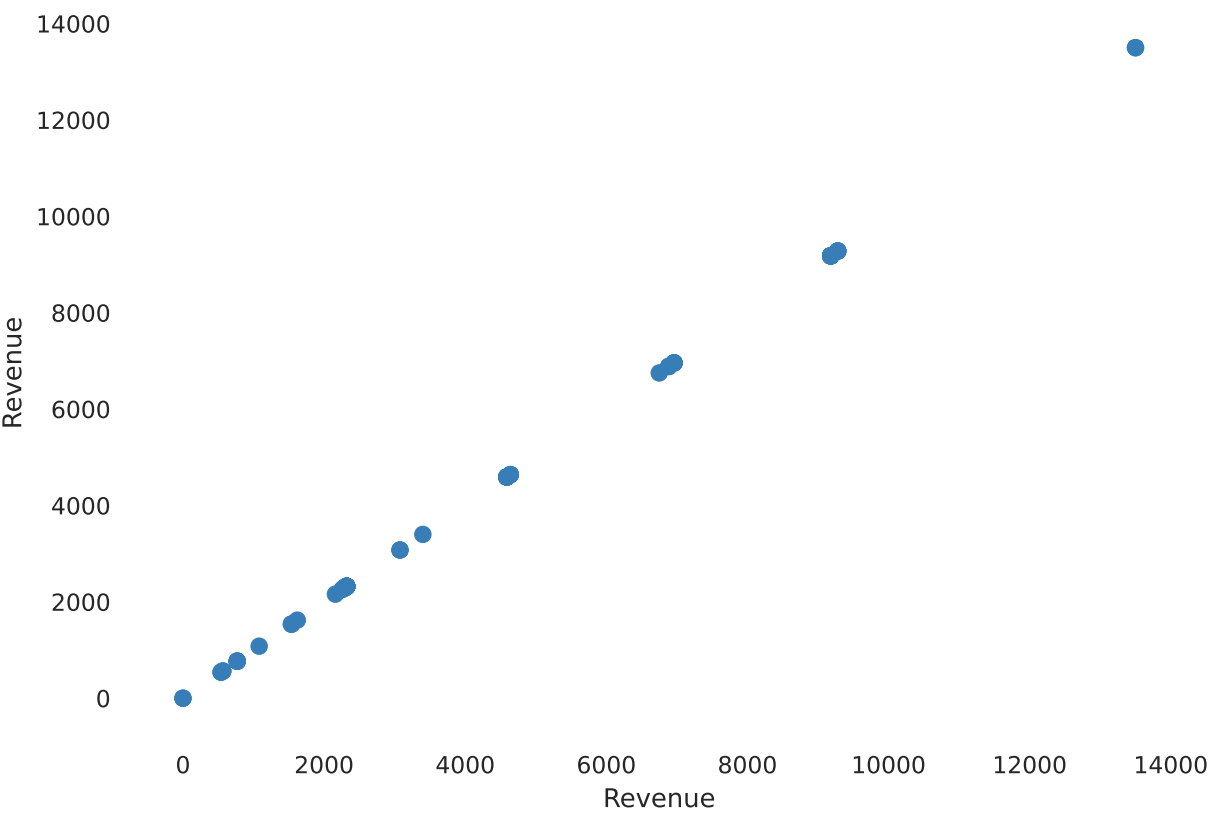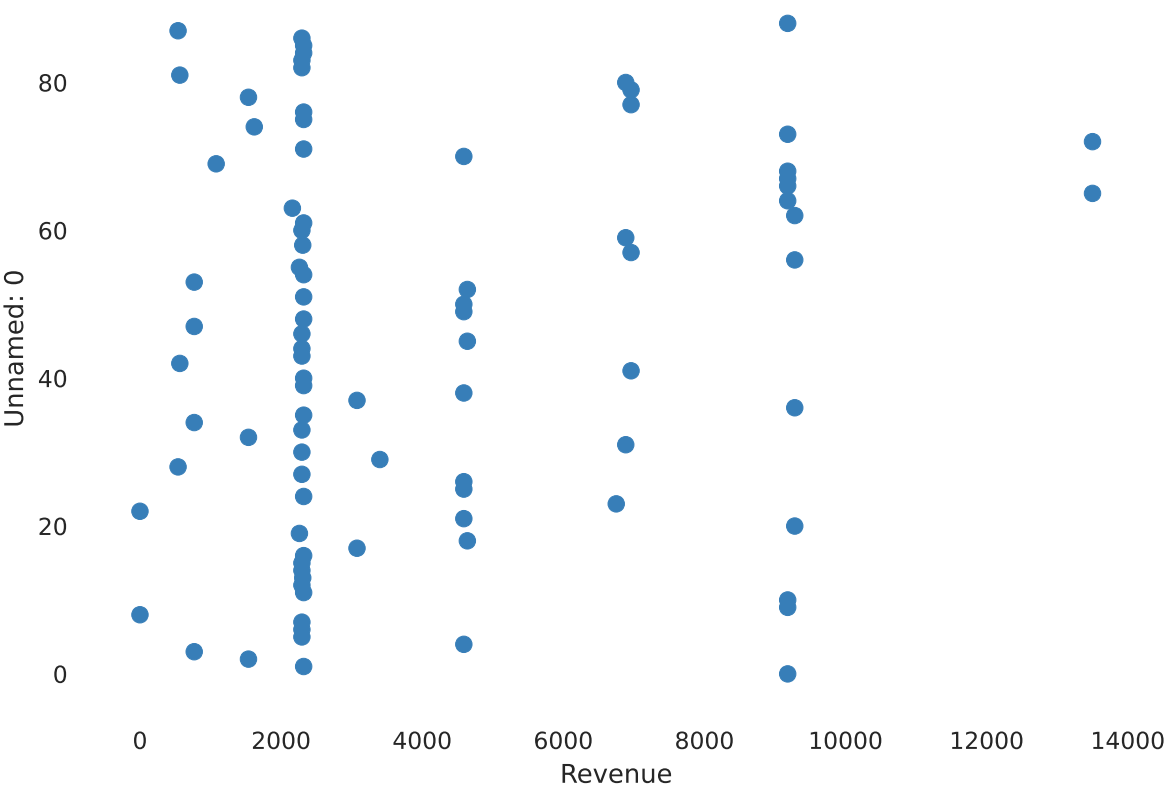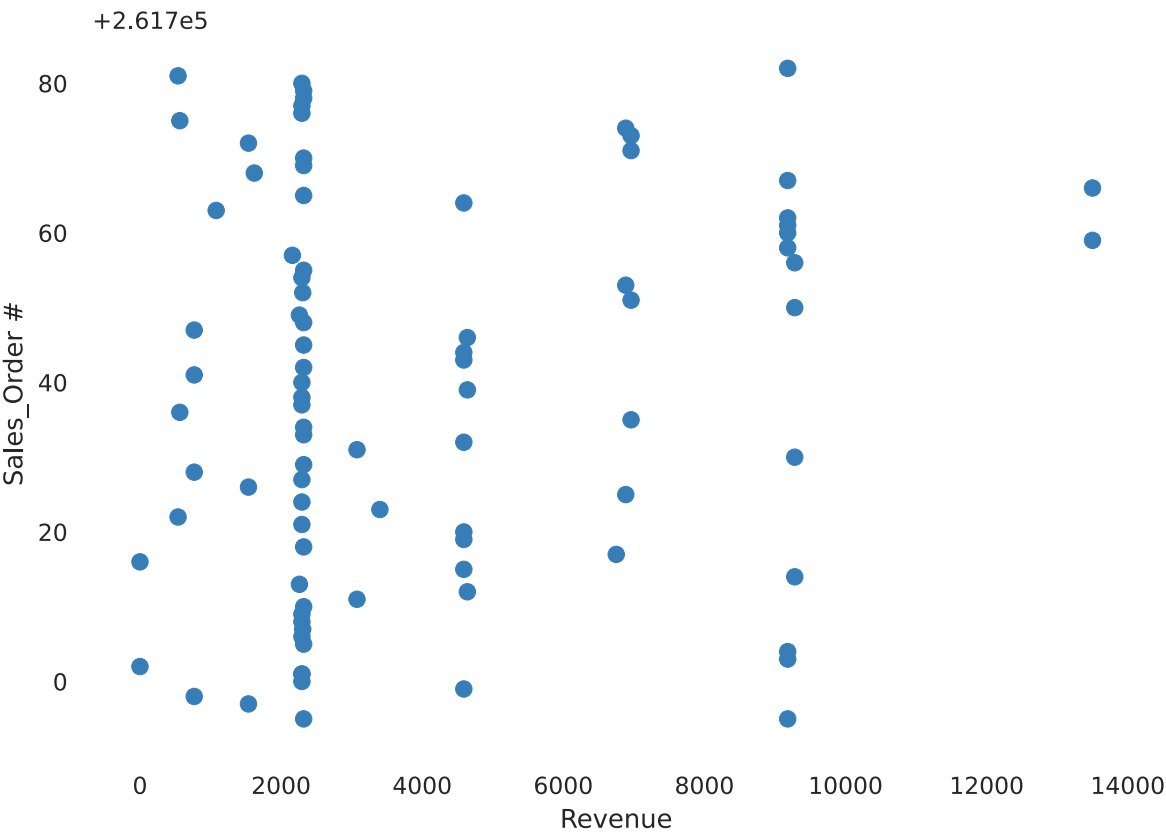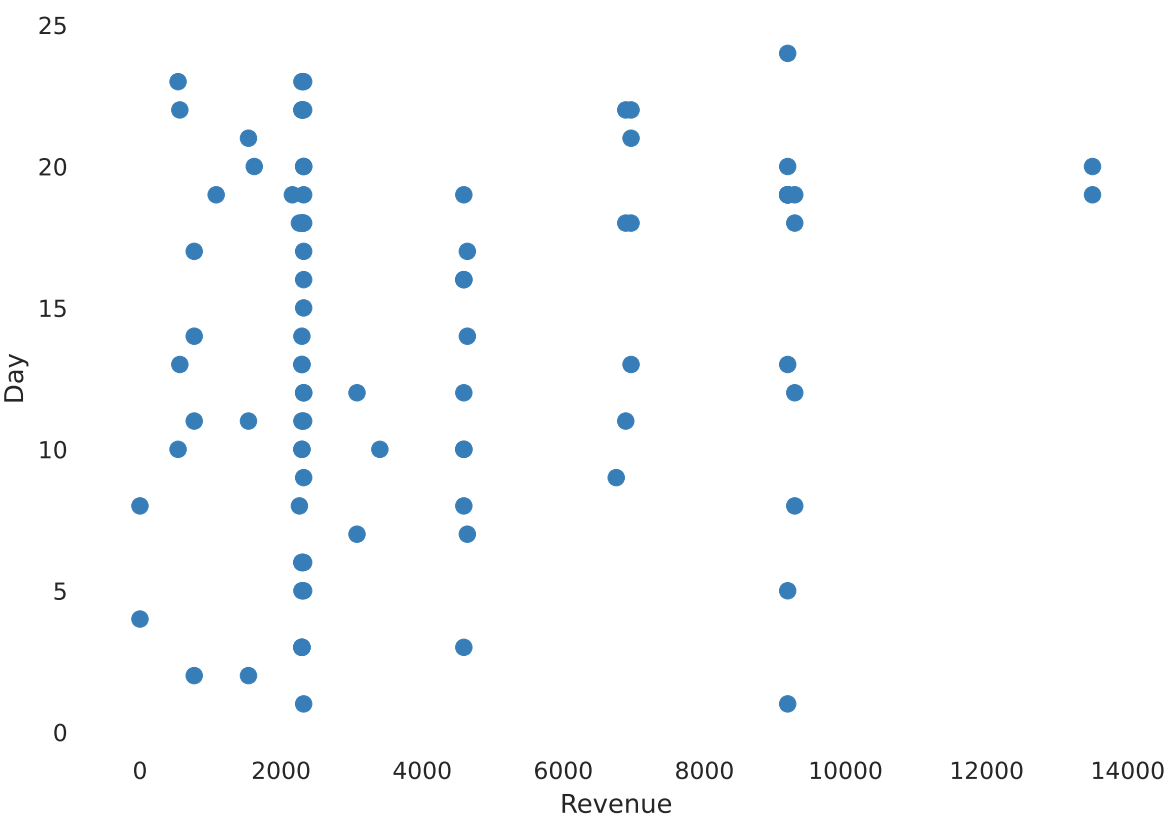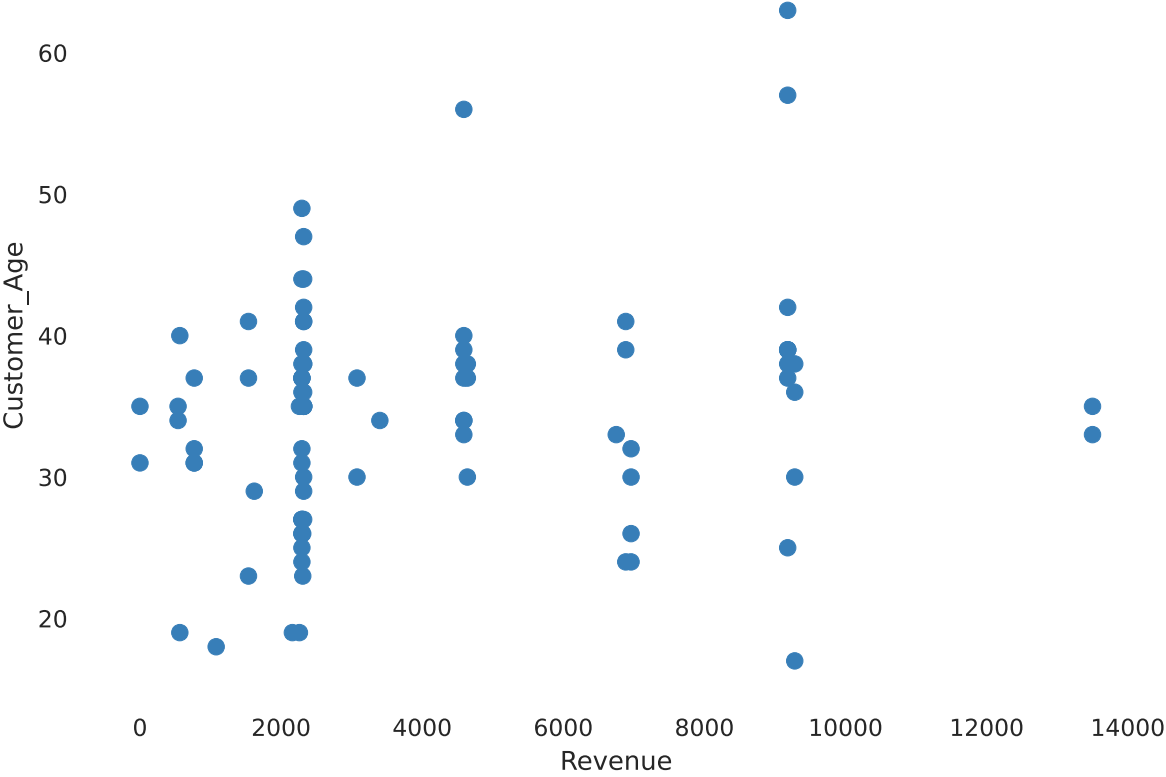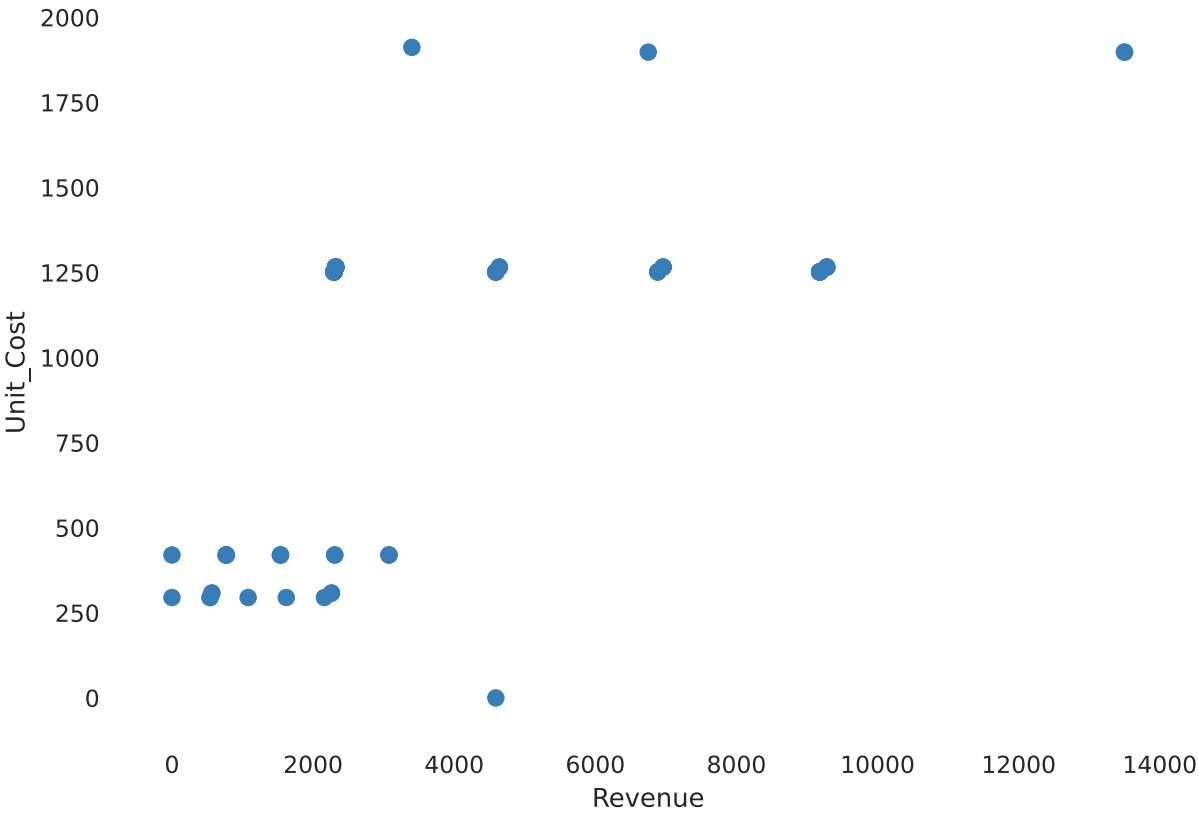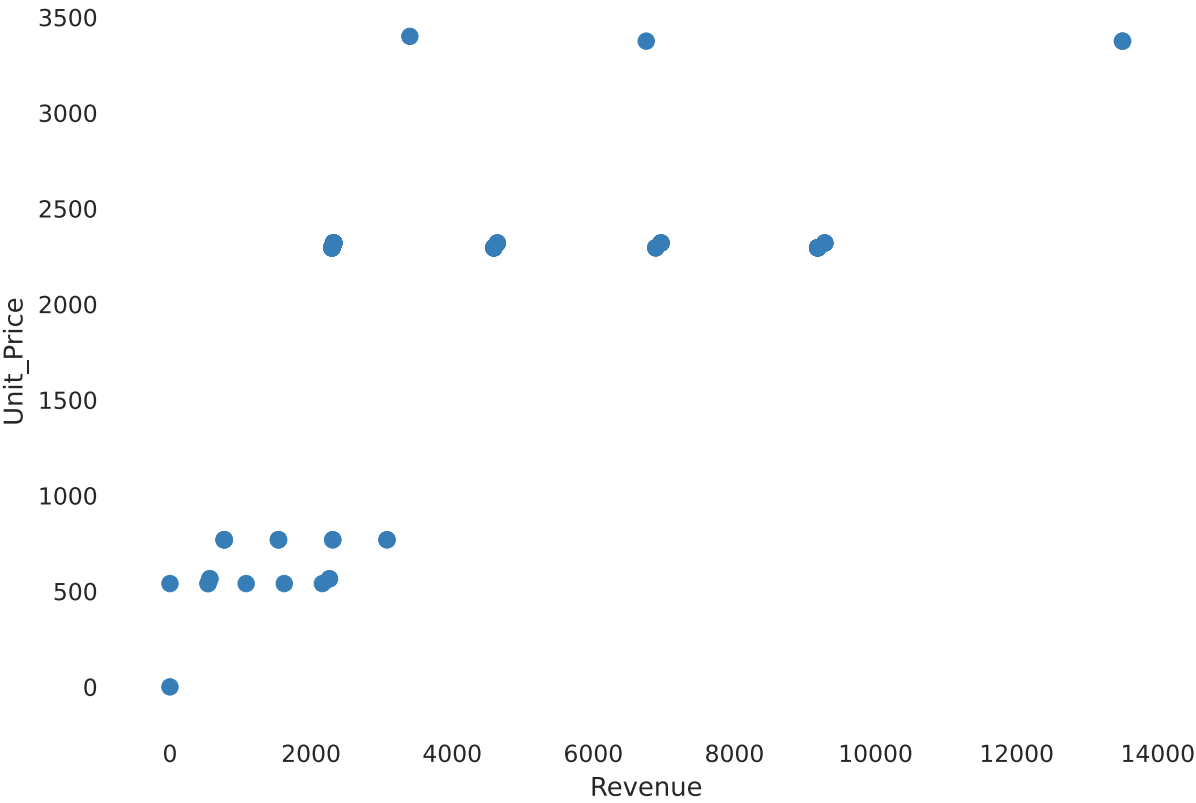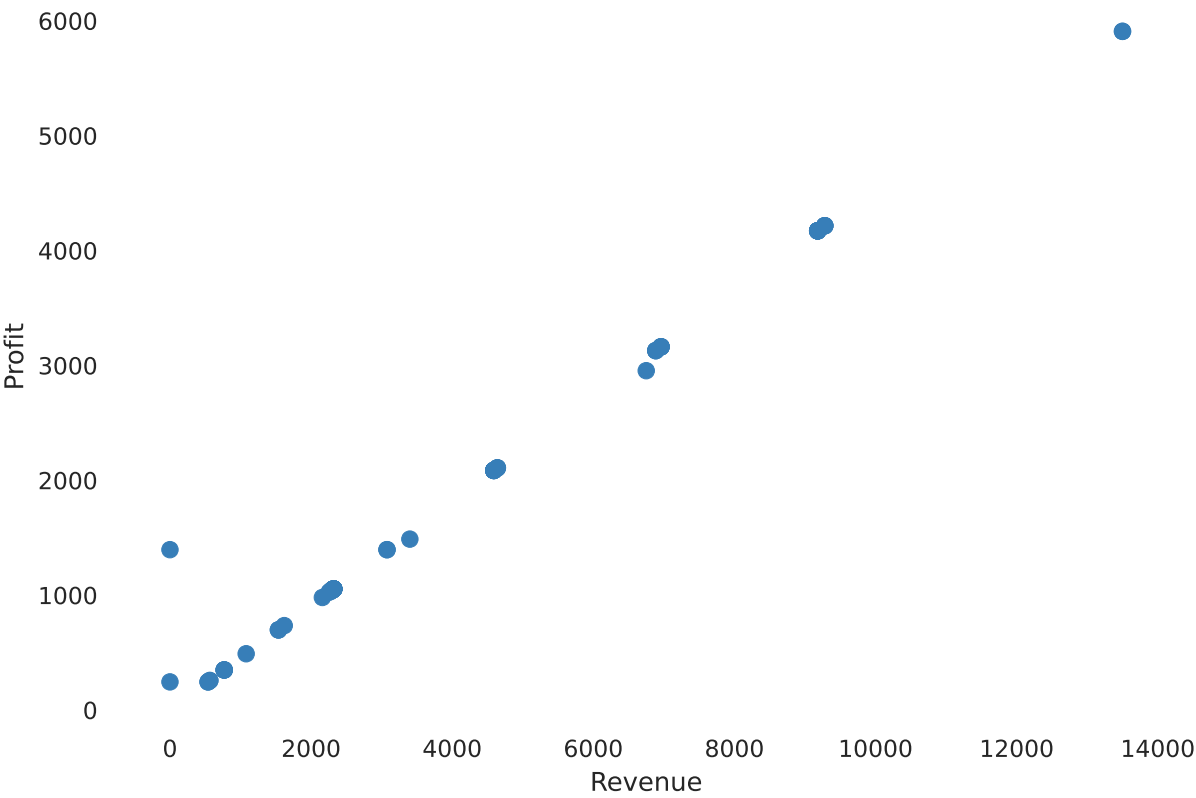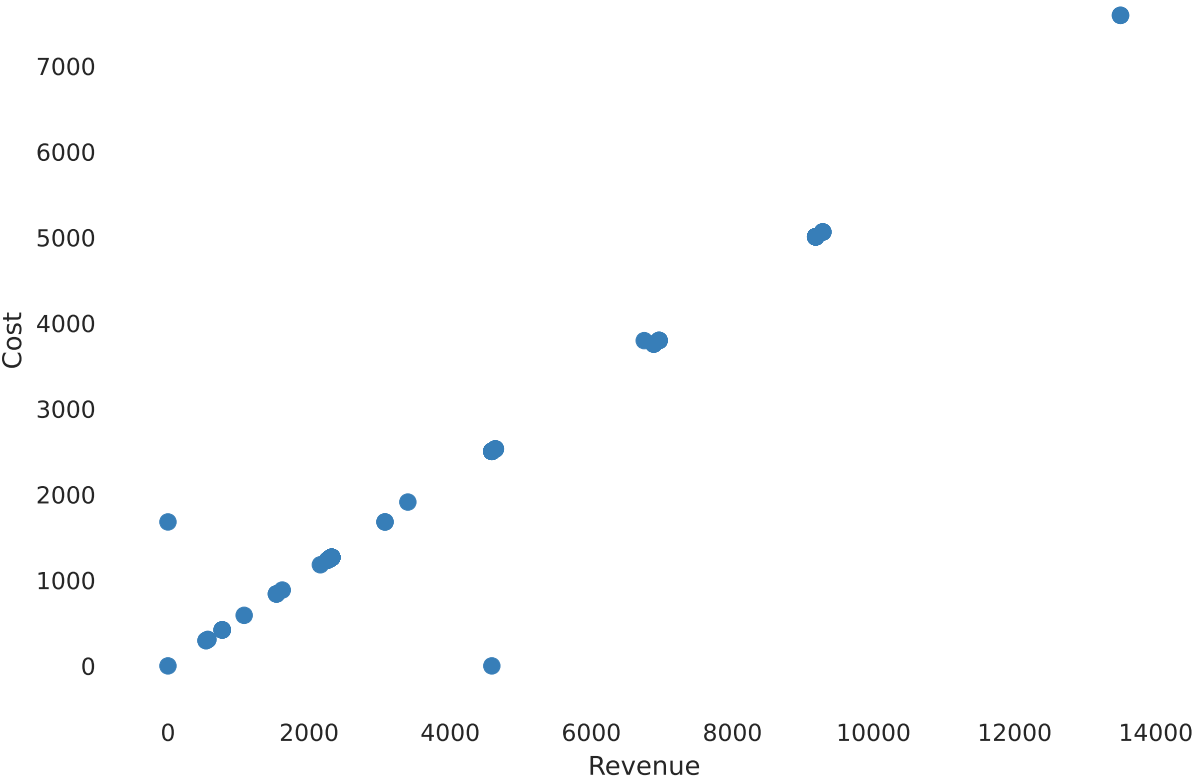
Correlations
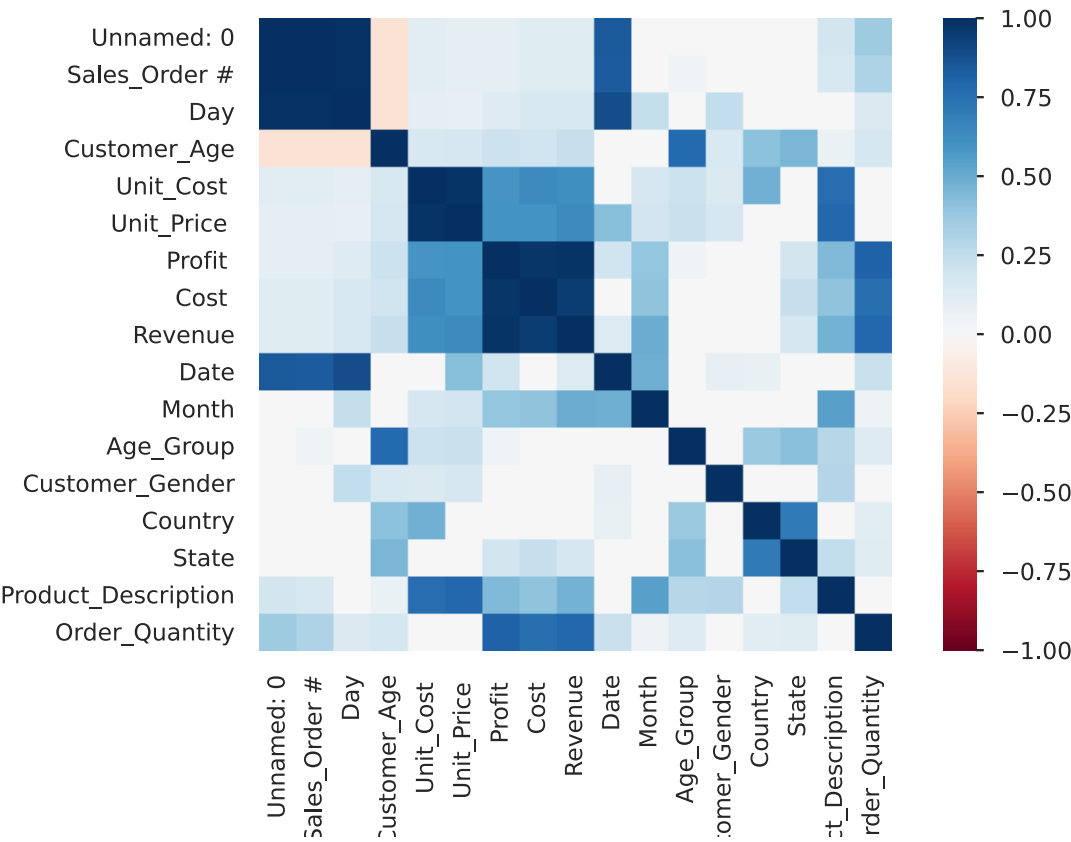
	Unnamed: 0	Sales_Order #	Day	Customer_Age	Unit_Cost	Unit_Price	Profit	Cost	Revenue	Date	Month
Unnamed: 0	1.000	1.000	0.989	-0.145	0.106	0.091	0.086	0.123	0.118	0.840	0.000
Sales_Order #	1.000	1.000	0.989	-0.145	0.106	0.090	0.086	0.123	0.118	0.832	0.000
Day	0.989	0.989	1.000	-0.146	0.099	0.084	0.127	0.161	0.158	0.888	0.236
Customer_Age	-0.145	-0.145	-0.146	1.000	0.158	0.170	0.218	0.203	0.229	0.000	0.000
Unit_Cost	0.106	0.106	0.099	0.158	1.000	0.983	0.590	0.633	0.611	0.000	0.167
Unit_Price	0.091	0.090	0.084	0.170	0.983	1.000	0.600	0.601	0.636	0.424	0.188
Profit	0.086	0.086	0.127	0.218	0.590	0.600	1.000	0.970	0.977	0.197	0.390
Cost	0.123	0.123	0.161	0.203	0.633	0.601	0.970	1.000	0.947	0.000	0.405
Revenue	0.118	0.118	0.158	0.229	0.611	0.636	0.977	0.947	1.000	0.137	0.499

Missing values
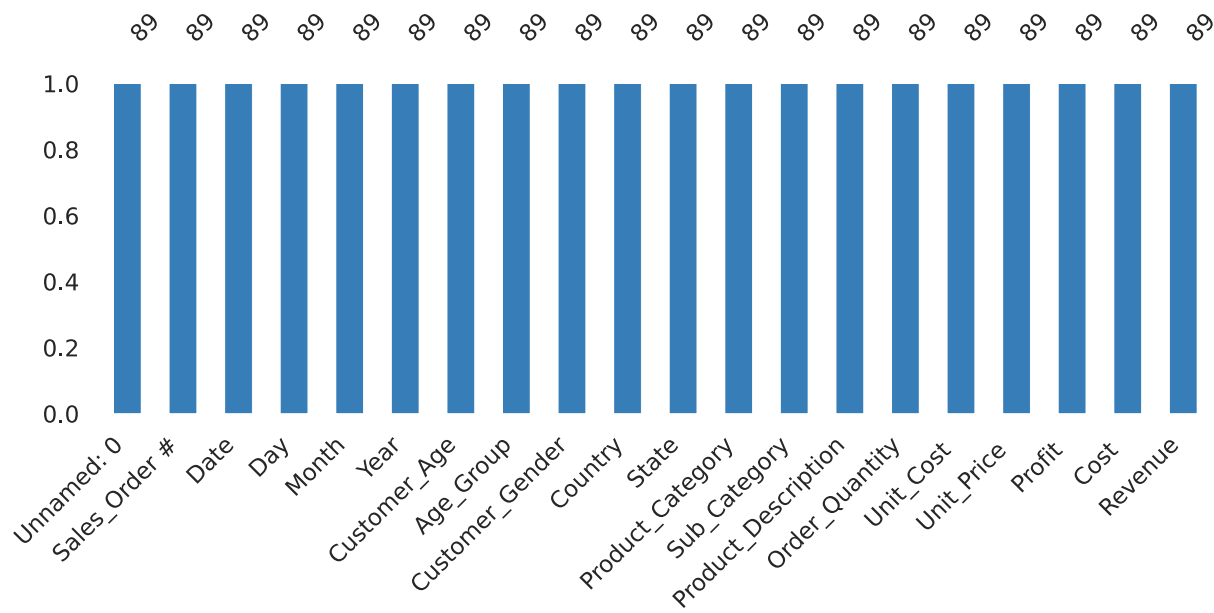
A simple visualization of nullity by column.
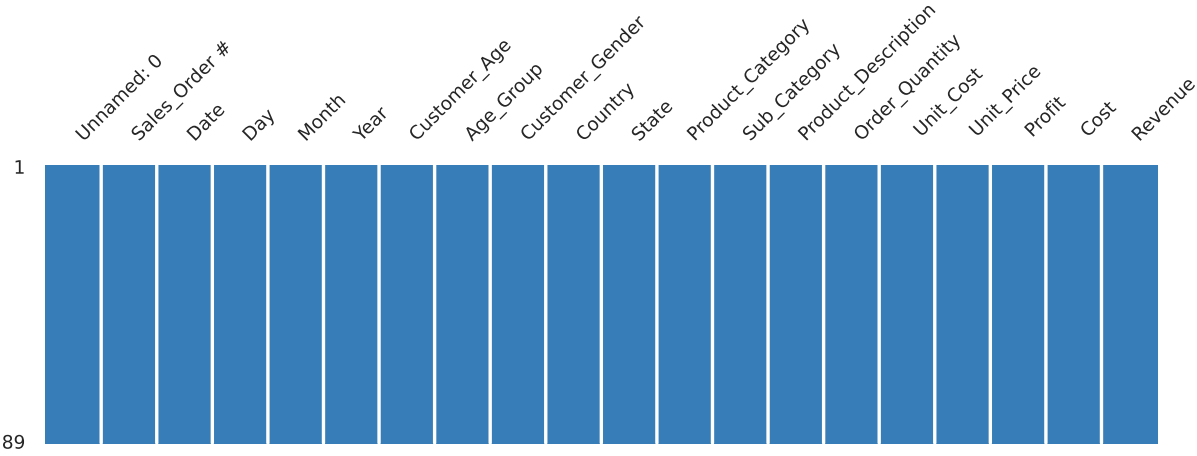
Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

Sample

	Unnamed: 0	Sales_Order #	Date	Day	Month	Year	Customer_Age	Age_Group	Customer_Gender	Country	St
0	0	261695	2021-12-01	1	December	2021	39	Adults (35-64)	F	United States	Ca
1	1	261695	2021-12-01	1	December	2021	44	Adults (35-64)	M	United Kingdom	Er
2	2	261697	2021-12-02	2	December	2021	37	Adults (35-64)	M	United States	Ca
3	3	261698	2021-12-02	2	December	2021	31	Young Adults (25-34)	F	Australia	No
4	4	261699	2021-12-03	3	December	2021	37	Adults (35-64)	F	United States	Ca
5	5	261700	2021-12-03	3	December	2021	24	Youth (<25)	F	United Kingdom	Er
6	6	261701	2021-12-03	3	December	2021	37	Adults (35-64)	M	United States	W
7	7	261701	2021-12-03	3	December	2021	37	Adults (35-64)	M	United States	W
8	8	261702	2021-12-04	4	December	2021	31	Young Adults (25-34)	F	Australia	No
9	9	261703	2021-12-05	5	December	2021	39	Adults (35-64)	F	United States	Ca

	Unnamed: 0	Sales_Order #	Date	Day	Month	Year	Customer_Age	Age_Group	Customer_Gender	Country	Sta
79	79	261773	2021-12-22	22	December	2021	30	Young Adults (25-34)	F	United States	Wa
80	80	261774	2021-12-22	22	December	2021	41	Adults (35-64)	M	United States	Cal
81	81	261775	2021-12-22	22	December	2021	19	Youth (<25)	F	Australia	Nev
82	82	261776	2021-12-22	22	December	2021	25	Young Adults (25-34)	M	France	Sei
83	83	261777	2021-12-22	22	December	2021	27	Young Adults (25-34)	F	Canada	Brit
84	84	261778	2021-12-22	22	December	2021	41	Adults (35-64)	M	Germany	Hes
85	85	261779	2021-12-23	23	December	2021	30	Young Adults (25-34)	F	United States	Ore
86	86	261780	2021-12-23	23	December	2021	31	Young Adults (25-34)	F	Canada	Brit
87	87	261781	2021-12-23	23	December	2021	35	Adults (35-64)	F	United States	Cal
88	88	261782	2021-12-24	24	December	2021	38	Adults (35-64)	M	Australia	Qu

Report generated by YData (https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).