

# Overview

## Dataset statistics

<b>Number of variables</b>	12
<b>Number of observations</b>	891
<b>Missing cells</b>	866
<b>Missing cells (%)</b>	8.1%
<b>Duplicate rows</b>	0
<b>Duplicate rows (%)</b>	0.0%
<b>Total size in memory</b>	83.7 KiB
<b>Average record size in memory</b>	96.1 B

## Variable types

<b>Numeric</b>	5
<b>Categorical</b>	4
<b>Text</b>	3

### Alerts

Survived is highly overall correlated with Sex	High correlation
Sex is highly overall correlated with Survived	High correlation
Age has 177 (19.9%) missing values	Missing
Cabin has 687 (77.1%) missing values	Missing
PassengerId is uniformly distributed	Uniform
PassengerId has unique values	Unique
Name has unique values	Unique
SibSp has 608 (68.2%) zeros	Zeros
Parch has 678 (76.1%) zeros	Zeros
Fare has 15 (1.7%) zeros	Zeros

## Reproduction

<b>Analysis started</b>	2023-09-28 21:45:21.583445
<b>Analysis finished</b>	2023-09-28 21:45:27.262698
<b>Duration</b>	5.68 seconds
<b>Software version</b>	ydata-profiling vv4.5.1 ( <a href="https://github.com/ydataai/ydata-profiling">https://github.com/ydataai/ydata-profiling</a> )
<b>Download configuration</b>	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Profile%20Train%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22crea

# Variables

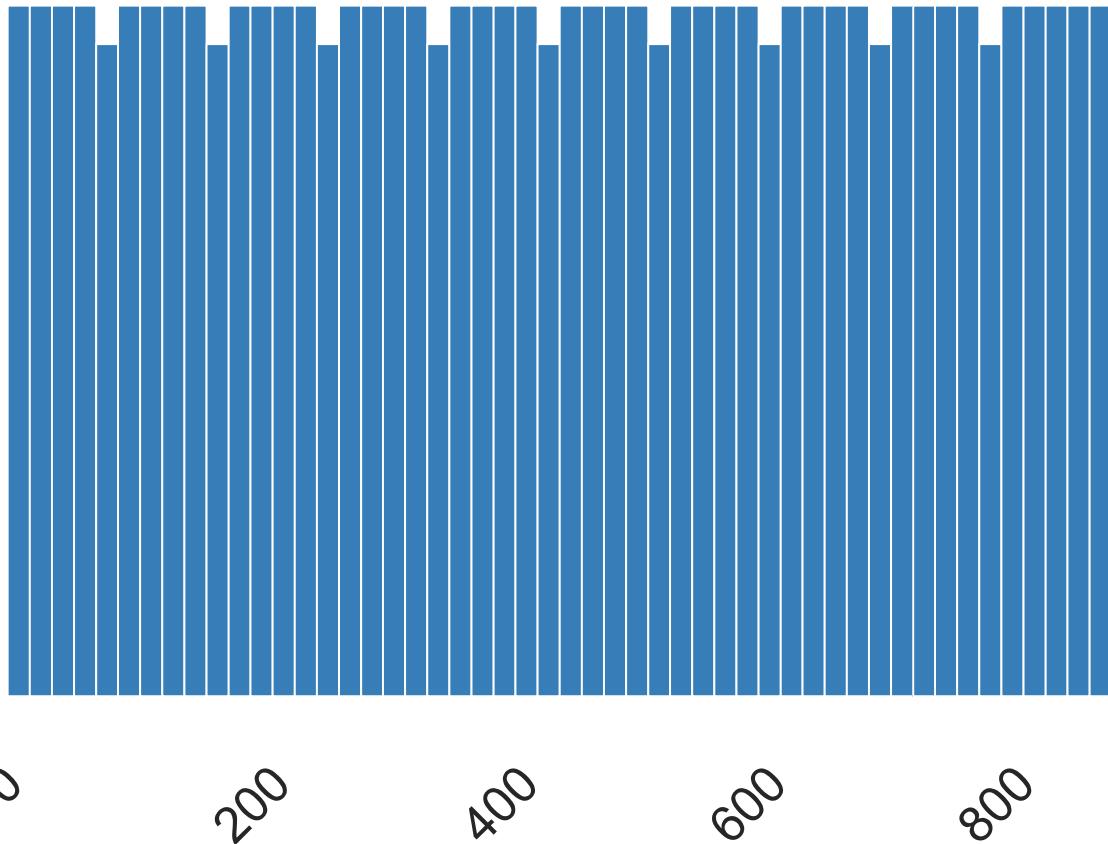
Select Columns ▾

## PassengerId

Real number (ℝ)

UNIFORM UNIQUE

<b>Distinct</b>	891
<b>Distinct (%)</b>	100.0%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	446
<b>Minimum</b>	1
<b>Maximum</b>	891
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB

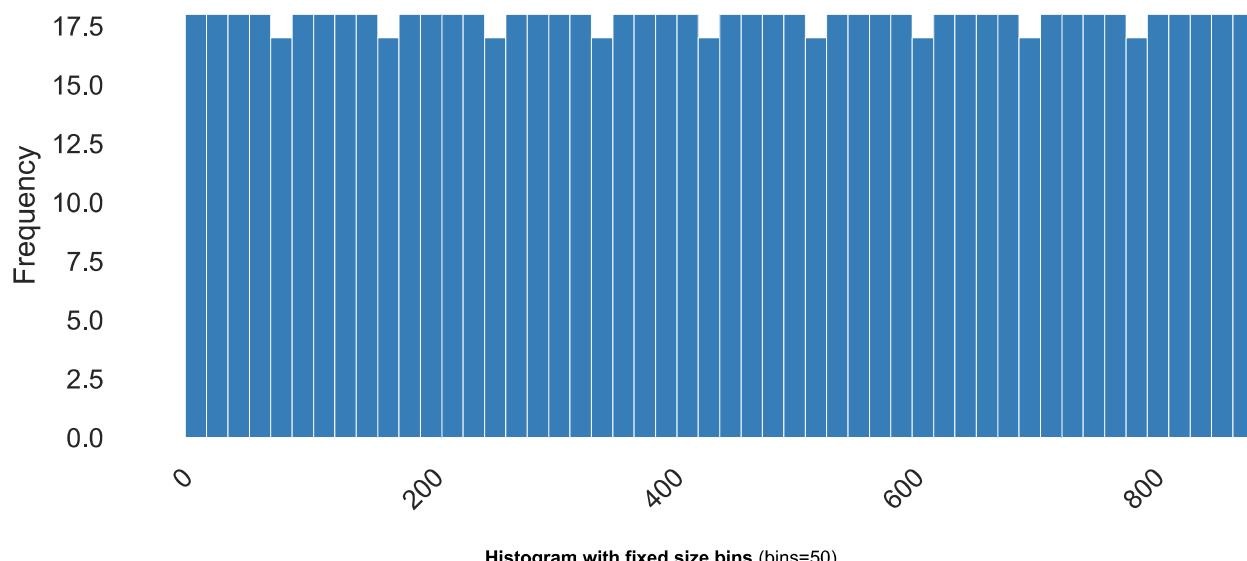


## Quantile statistics

<b>Minimum</b>	1
<b>5-th percentile</b>	45.5
<b>Q1</b>	223.5
<b>median</b>	446
<b>Q3</b>	668.5
<b>95-th percentile</b>	846.5
<b>Maximum</b>	891
<b>Range</b>	890
<b>Interquartile range (IQR)</b>	445

## Descriptive statistics

<b>Standard deviation</b>	257.35384
<b>Coefficient of variation (CV)</b>	0.57702655
<b>Kurtosis</b>	-1.2
<b>Mean</b>	446
<b>Median Absolute Deviation (MAD)</b>	223
<b>Skewness</b>	0
<b>Sum</b>	397386
<b>Variance</b>	66231
<b>Monotonicity</b>	Strictly increasing



Value	Count	Frequency (%)
1	1	0.1%
599	1	0.1%
588	1	0.1%
589	1	0.1%
590	1	0.1%
591	1	0.1%
592	1	0.1%
593	1	0.1%
594	1	0.1%
595	1	0.1%
Other values (881)	881	98.9%

Value	Count	Frequency (%)
1	1	0.1%
2	1	0.1%
3	1	0.1%
4	1	0.1%
5	1	0.1%
6	1	0.1%
7	1	0.1%
8	1	0.1%
9	1	0.1%
10	1	0.1%

Value	Count	Frequency (%)
891	1	0.1%
890	1	0.1%
889	1	0.1%
888	1	0.1%
887	1	0.1%
886	1	0.1%
885	1	0.1%
884	1	0.1%
883	1	0.1%
882	1	0.1%





**Survived**

Categorical

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

<b>Distinct</b>	2
<b>Distinct (%)</b>	0.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB

**Length**

<b>Max length</b>	1
<b>Median length</b>	1
<b>Mean length</b>	1
<b>Min length</b>	1

**Characters and Unicode**

<b>Total characters</b>	891	
<b>Distinct characters</b>	2	
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

**Unique**

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

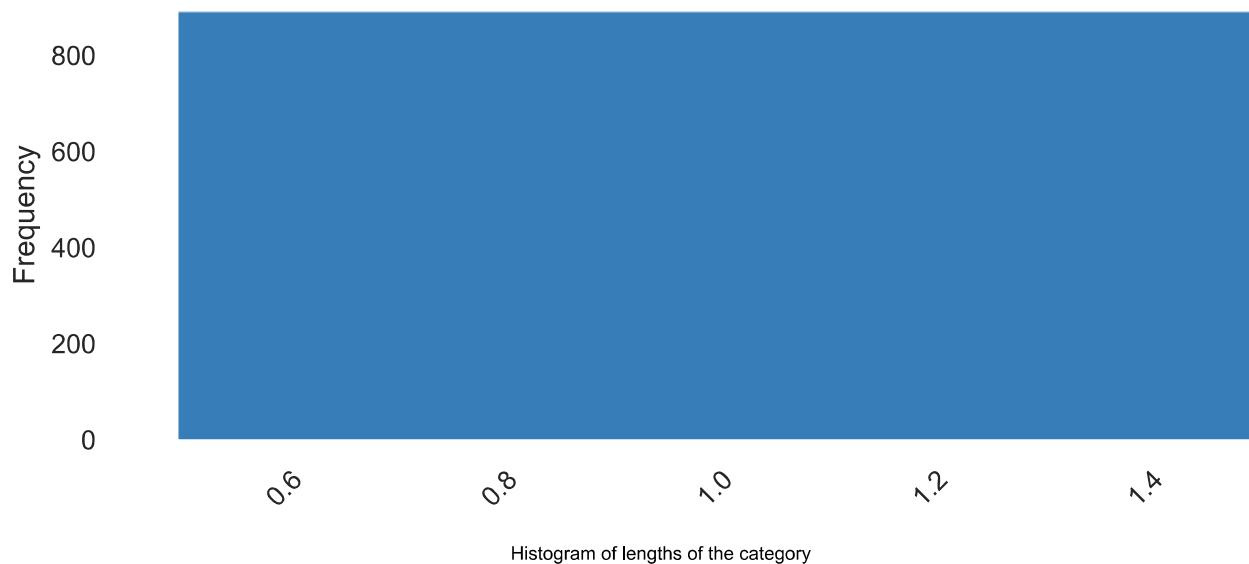
**Sample**

<b>1st row</b>	0
<b>2nd row</b>	1
<b>3rd row</b>	1
<b>4th row</b>	1
<b>5th row</b>	0

## Common Values

Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%

## Length



## Common Values (Plot)



Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%

## Most occurring characters

Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	891	100.0%

## Most frequent character per category

*Decimal Number*

Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%

## Most occurring scripts

Value	Count	Frequency (%)
Common	891	100.0%

## Most frequent character per script

*Common*

Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	891	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	549	61.6%
1	342	38.4%





**Pclass**

Categorical

<b>Distinct</b>	3
<b>Distinct (%)</b>	0.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB

**Length**

<b>Max length</b>	1
<b>Median length</b>	1
<b>Mean length</b>	1
<b>Min length</b>	1

**Characters and Unicode**

<b>Total characters</b>	891
<b>Distinct characters</b>	3
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )

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

**Unique**

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

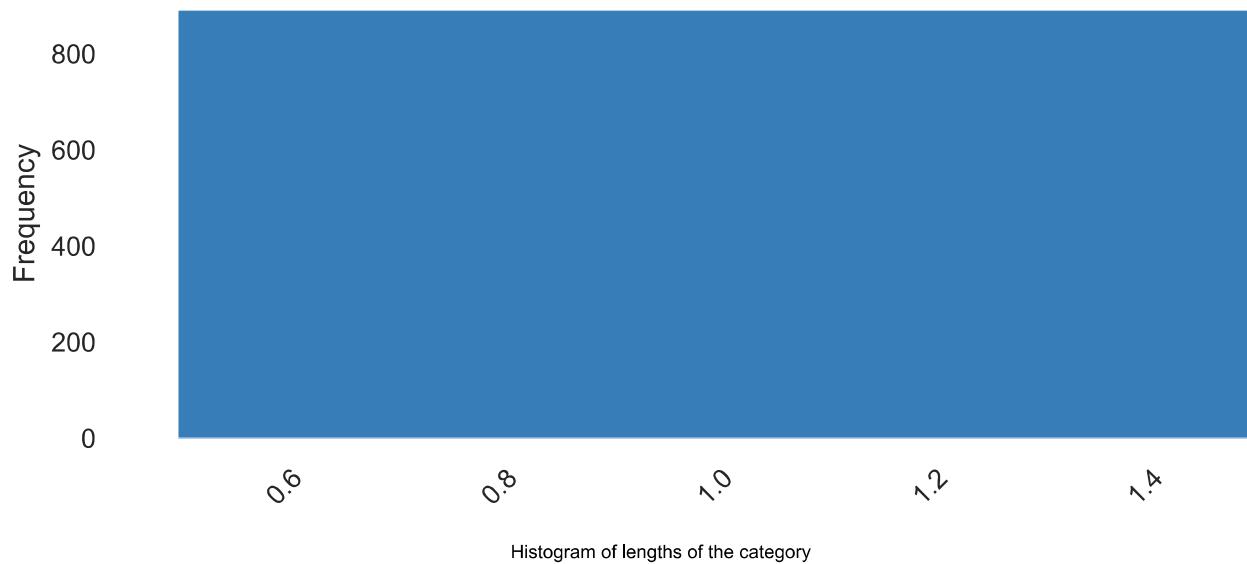
**Sample**

<b>1st row</b>	3
<b>2nd row</b>	1
<b>3rd row</b>	3
<b>4th row</b>	1
<b>5th row</b>	3

## Common Values

Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%

## Length



## Common Values (Plot)



Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%

## Most occurring characters

Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	891	100.0%

## Most frequent character per category

*Decimal Number*

Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%

## Most occurring scripts

Value	Count	Frequency (%)
Common	891	100.0%

## Most frequent character per script

*Common*

Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	891	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
3	491	55.1%
1	216	24.2%
2	184	20.7%





Name \_\_\_\_\_

Text

UNIQUE

<b>Distinct</b>	891
<b>Distinct (%)</b>	100.0%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB



## Length

<b>Max length</b>	82
<b>Median length</b>	52
<b>Mean length</b>	26.965208
<b>Min length</b>	12

## Characters and Unicode

Total characters	24026
Distinct characters	60
Distinct categories	7 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
Distinct scripts	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )

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

Unique

Unique	891	?
Unique (%)	100.0%	

**Sample**

<b>1st row</b>	Braund, Mr. Owen Harris
<b>2nd row</b>	Cumings, Mrs. John Bradley (Florence Briggs Thayer)
<b>3rd row</b>	Heikkinen, Miss. Laina
<b>4th row</b>	Futrelle, Mrs. Jacques Heath (Lily May Peel)
<b>5th row</b>	Allen, Mr. William Henry

Value	Count	Frequency (%)
mr	521	14.4%
miss	182	5.0%
mrs	129	3.6%
william	64	1.8%
john	44	1.2%
master	40	1.1%
henry	35	1.0%
george	24	0.7%
james	24	0.7%
charles	23	0.6%
Other values (1515)	2538	70.0%



## Most occurring characters

Value	Count	Frequency (%)
r	2735	11.4%
e	1958	8.1%
a	1703	7.1%
i	1657	6.9%
n	1325	5.5%
s	1304	5.4%
M	1297	5.4%
	1128	4.7%

Value	Count	Frequency (%)
I	1067	4.4%
O	1008	4.2%
Other values (50)	8844	36.8%

## Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	15446	64.3%
Uppercase Letter	3645	15.2%
Space Separator	2735	11.4%
Other Punctuation	1899	7.9%
Close Punctuation	144	0.6%
Open Punctuation	144	0.6%
Dash Punctuation	13	0.1%

## Most frequent character per category

*Lowercase Letter*

Value	Count	Frequency (%)
r	1958	12.7%
e	1703	11.0%
a	1657	10.7%
i	1325	8.6%
n	1304	8.4%
s	1297	8.4%
l	1067	6.9%
o	1008	6.5%
t	667	4.3%
h	517	3.3%
Other values (16)	2943	19.1%

*Uppercase Letter*

Value	Count	Frequency (%)
M	1128	30.9%
A	250	6.9%
J	215	5.9%
H	203	5.6%
S	180	4.9%
C	172	4.7%
E	166	4.6%
W	143	3.9%
B	140	3.8%
L	129	3.5%
Other values (15)	919	25.2%

*Other Punctuation*

Value	Count	Frequency (%)
.	892	47.0%

Value	Count	Frequency (%)
,	891	46.9%
"	106	5.6%
'	9	0.5%
/	1	0.1%

*Space Separator*

Value	Count	Frequency (%)
	2735	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	144	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	144	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	13	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	19091	79.5%
Common	4935	20.5%

## Most frequent character per script

*Latin*

Value	Count	Frequency (%)
r	1958	10.3%
e	1703	8.9%
a	1657	8.7%
i	1325	6.9%
n	1304	6.8%
s	1297	6.8%
M	1128	5.9%
I	1067	5.6%
o	1008	5.3%
t	667	3.5%
Other values (41)	5977	31.3%

*Common*

Value	Count	Frequency (%)
.	2735	55.4%
,	892	18.1%
)	891	18.1%
(	144	2.9%
"	144	2.9%
-	106	2.1%
'	13	0.3%
/	9	0.2%
	1	< 0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	24026	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
r	2735	11.4%
e	1958	8.1%
a	1703	7.1%
i	1657	6.9%
n	1325	5.5%
s	1304	5.4%
M	1297	5.4%
l	1128	4.7%
o	1067	4.4%
Other values (50)	1008	4.2%
	8844	36.8%





**Sex**

Categorical

**HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Survived)**

<b>Distinct</b>	2
<b>Distinct (%)</b>	0.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB

**Length**

<b>Max length</b>	6
<b>Median length</b>	4
<b>Mean length</b>	4.704826
<b>Min length</b>	4

**Characters and Unicode**

<b>Total characters</b>	4192	
<b>Distinct characters</b>	5	
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

**Unique**

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

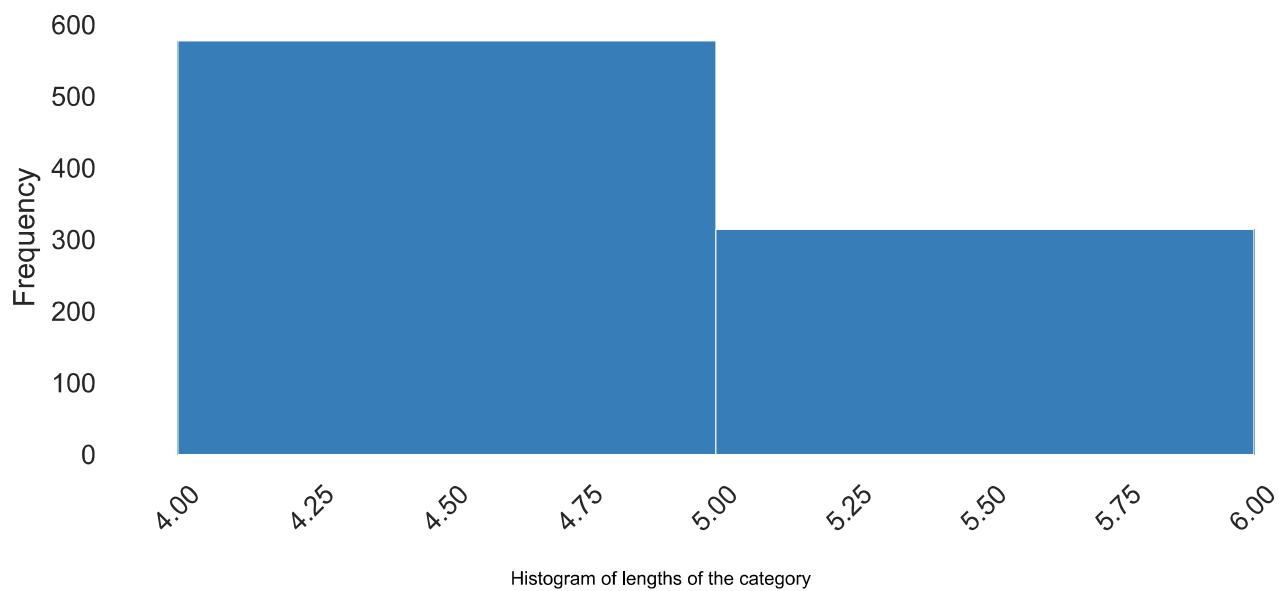
**Sample**

<b>1st row</b>	male
<b>2nd row</b>	female
<b>3rd row</b>	female
<b>4th row</b>	female
<b>5th row</b>	male

## Common Values

Value	Count	Frequency (%)
male	577	64.8%
female	314	35.2%

## Length



## Common Values (Plot)



Value	Count	Frequency (%)
male	577	64.8%
female	314	35.2%

## Most occurring characters

Value	Count	Frequency (%)
e	1205	28.7%
m	891	21.3%
a	891	21.3%
l	891	21.3%
f	314	7.5%

## Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	4192	100.0%

## Most frequent character per category

*Lowercase Letter*

Value	Count	Frequency (%)
e	1205	28.7%
m	891	21.3%
a	891	21.3%
l	891	21.3%
f	314	7.5%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	4192	100.0%

## Most frequent character per script

*Latin*

Value	Count	Frequency (%)
e	1205	28.7%
m	891	21.3%
a	891	21.3%
l	891	21.3%
f	314	7.5%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	4192	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	1205	28.7%
m	891	21.3%
a	891	21.3%
l	891	21.3%
f	314	7.5%





**Age**  
Real number ( $\mathbb{R}$ )

MISSING

Distinct

88

Distinct (%)

12.3%

Missing

177

Missing (%)

19.9%

Infinite

0

Infinite (%)

0.0%

Mean

29.699118

Minimum

0.42

Maximum

80

Zeros

0

Zeros (%)

0.0%

Negative

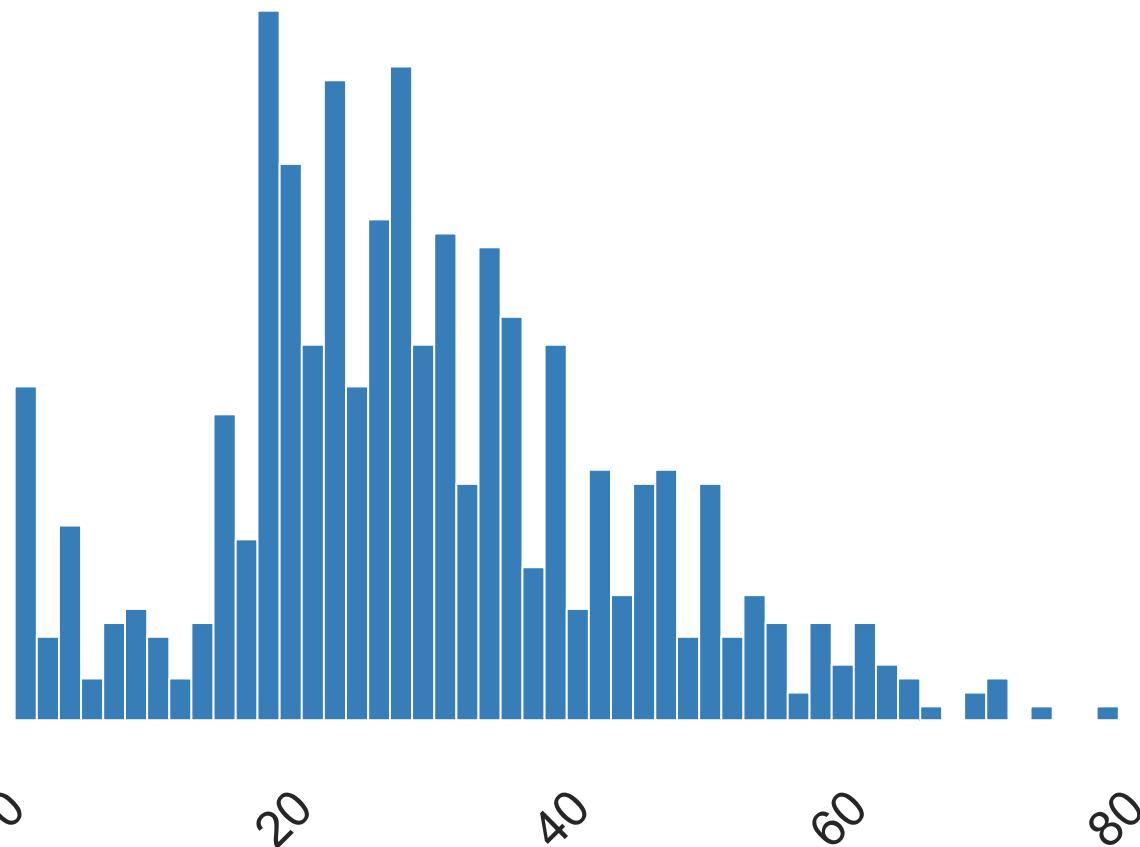
0

Negative (%)

0.0%

Memory size

7.1 KiB



## Quantile statistics

Minimum

0.42

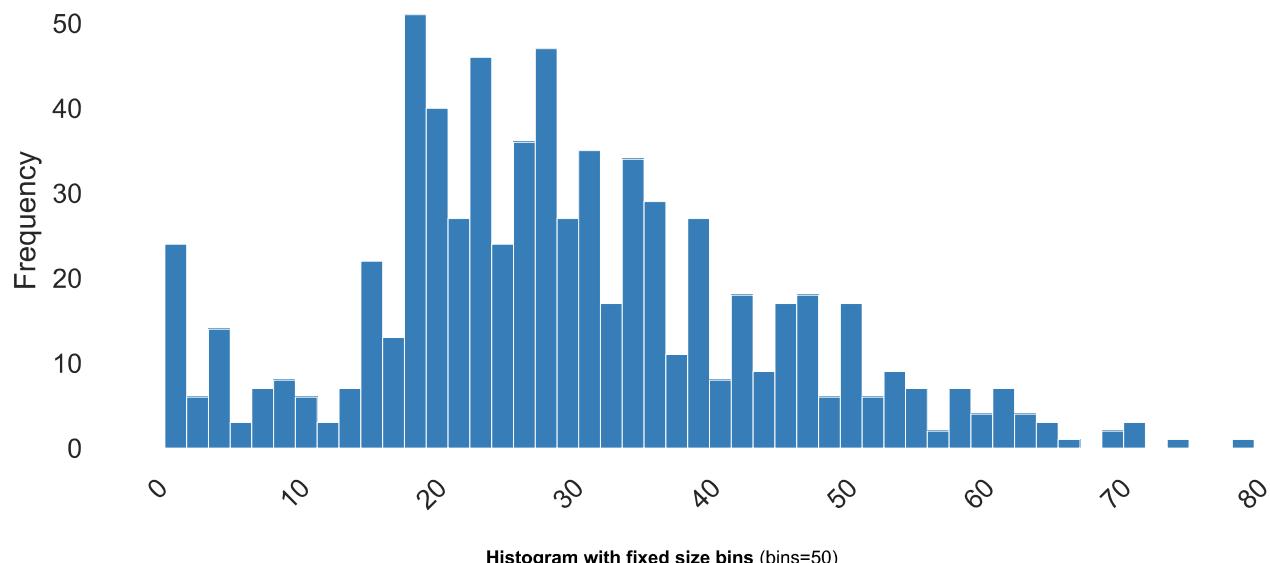
5-th percentile

4

Q1

20.125

<b>median</b>	28
<b>Q3</b>	38
<b>95-th percentile</b>	56
<b>Maximum</b>	80
<b>Range</b>	79.58
<b>Interquartile range (IQR)</b>	17.875
<b>Descriptive statistics</b>	
<b>Standard deviation</b>	14.526497
<b>Coefficient of variation (CV)</b>	0.48912219
<b>Kurtosis</b>	0.17827415
<b>Mean</b>	29.699118
<b>Median Absolute Deviation (MAD)</b>	9
<b>Skewness</b>	0.38910778
<b>Sum</b>	21205.17
<b>Variance</b>	211.01912
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
24	30	3.4%
22	27	3.0%
18	26	2.9%
28	25	2.8%
30	25	2.8%
19	25	2.8%
21	24	2.7%
25	23	2.6%
36	22	2.5%
29	20	2.2%
Other values (78)	467	52.4%
(Missing)	177	19.9%

Value	Count	Frequency (%)
0.42	1	0.1%
0.67	1	0.1%
0.75	2	0.2%
0.83	2	0.2%
0.92	1	0.1%
1	7	0.8%
2	10	1.1%
3	6	0.7%
4	10	1.1%
5	4	0.4%

Value	Count	Frequency (%)
80	1	0.1%
74	1	0.1%
71	2	0.2%
70.5	1	0.1%
70	2	0.2%
66	1	0.1%
65	3	0.3%
64	2	0.2%
63	2	0.2%
62	4	0.4%

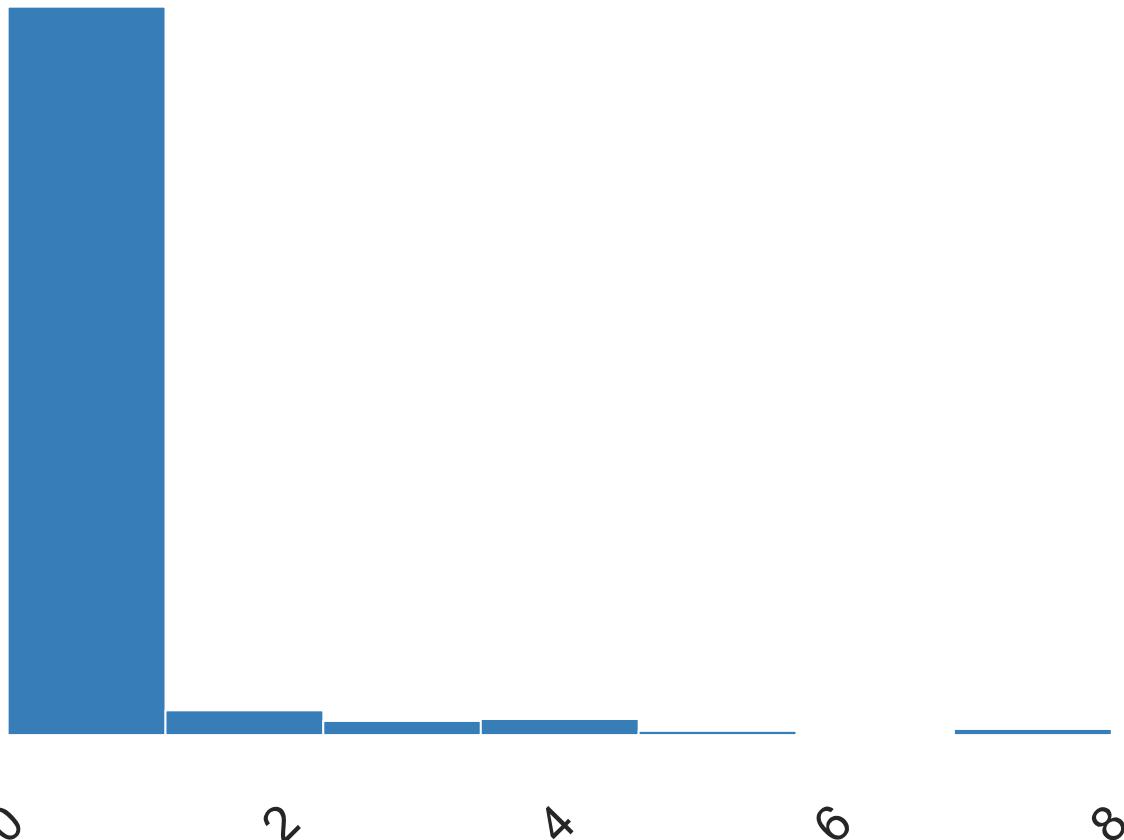




**SibSp**Real number ( $\mathbb{R}$ )

ZEROS

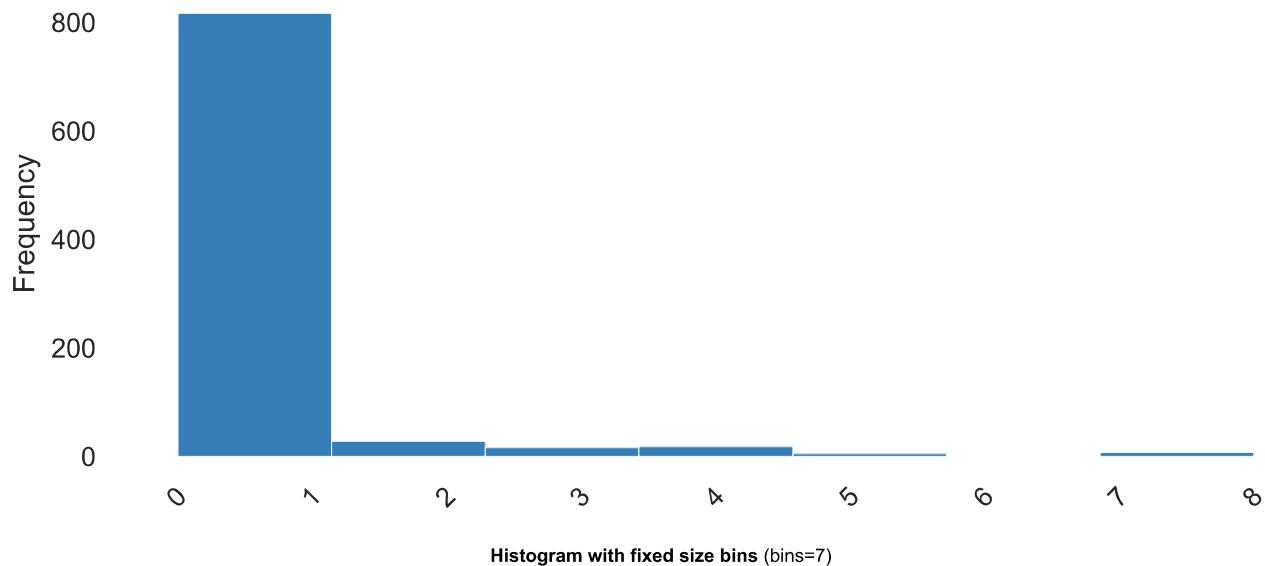
Distinct	7
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.52300786
Minimum	0
Maximum	8
Zeros	608
Zeros (%)	68.2%
Negative	0
Negative (%)	0.0%
Memory size	7.1 KiB



## Quantile statistics

Minimum	0
5-th percentile	0
Q1	0

<b>median</b>	0
<b>Q3</b>	1
<b>95-th percentile</b>	3
<b>Maximum</b>	8
<b>Range</b>	8
<b>Interquartile range (IQR)</b>	1
<b>Descriptive statistics</b>	
<b>Standard deviation</b>	1.1027434
<b>Coefficient of variation (CV)</b>	2.1084644
<b>Kurtosis</b>	17.88042
<b>Mean</b>	0.52300786
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	3.6953517
<b>Sum</b>	466
<b>Variance</b>	1.2160431
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
0	608	68.2%
1	209	23.5%
2	28	3.1%
4	18	2.0%
3	16	1.8%
8	7	0.8%
5	5	0.6%

Value	Count	Frequency (%)
0	608	68.2%
1	209	23.5%
2	28	3.1%
3	16	1.8%
4	18	2.0%
5	5	0.6%
8	7	0.8%

Value	Count	Frequency (%)
8	7	0.8%
5	5	0.6%
4	18	2.0%
3	16	1.8%
2	28	3.1%
1	209	23.5%
0	608	68.2%

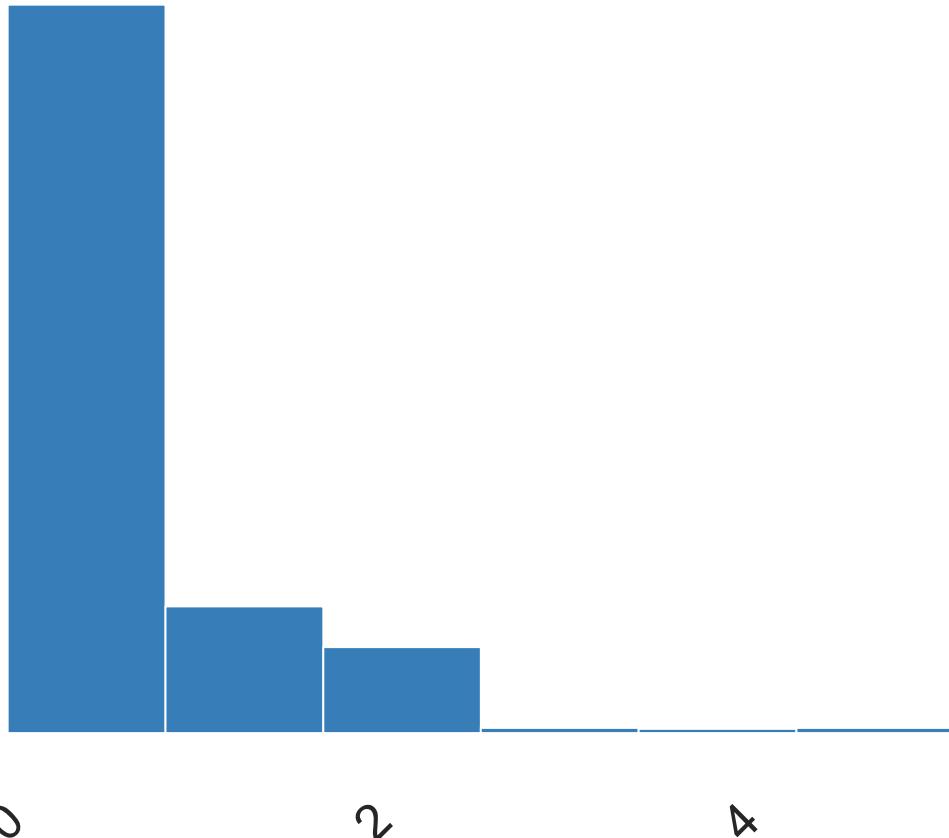




**Parch**Real number ( $\mathbb{R}$ )

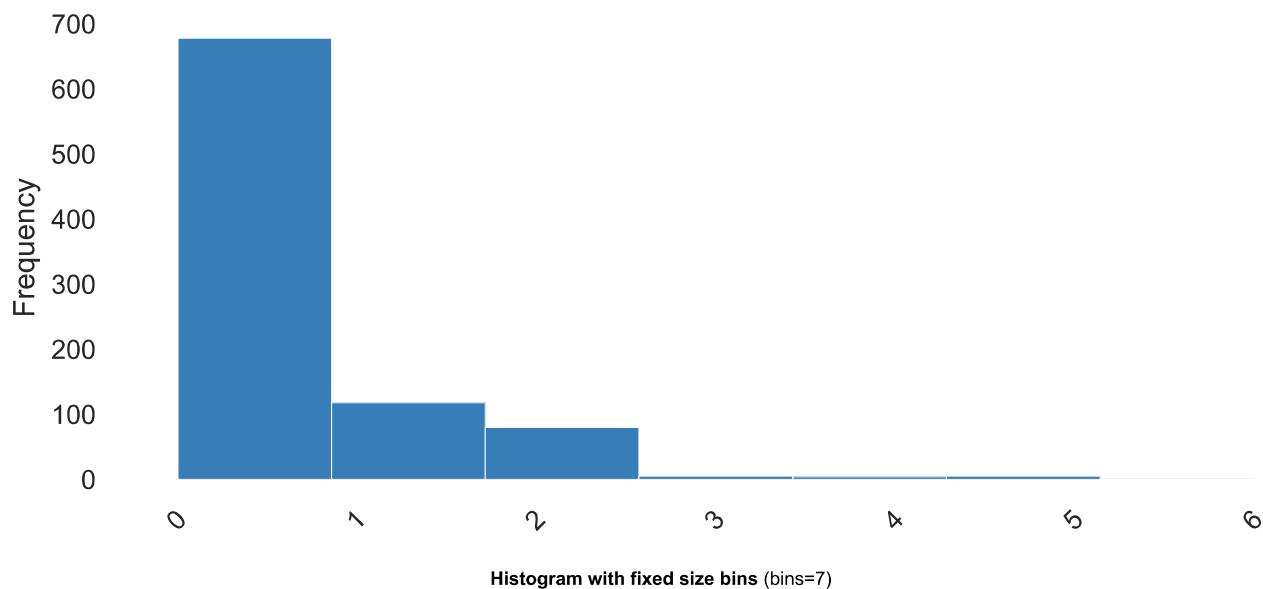
ZEROS

<b>Distinct</b>	7
<b>Distinct (%)</b>	0.8%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	0.38159371
<b>Minimum</b>	0
<b>Maximum</b>	6
<b>Zeros</b>	678
<b>Zeros (%)</b>	76.1%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB

**Quantile statistics**

<b>Minimum</b>	0
<b>5-th percentile</b>	0
<b>Q1</b>	0

<b>median</b>	0
<b>Q3</b>	0
<b>95-th percentile</b>	2
<b>Maximum</b>	6
<b>Range</b>	6
<b>Interquartile range (IQR)</b>	0
<b>Descriptive statistics</b>	
<b>Standard deviation</b>	0.80605722
<b>Coefficient of variation (CV)</b>	2.1123441
<b>Kurtosis</b>	9.7781252
<b>Mean</b>	0.38159371
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	2.749117
<b>Sum</b>	340
<b>Variance</b>	0.64972824
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
0	678	76.1%
1	118	13.2%
2	80	9.0%
5	5	0.6%
3	5	0.6%
4	4	0.4%
6	1	0.1%

Value	Count	Frequency (%)
0	678	76.1%
1	118	13.2%
2	80	9.0%
3	5	0.6%
4	4	0.4%
5	5	0.6%
6	1	0.1%

Value	Count	Frequency (%)
6	1	0.1%
5	5	0.6%
4	4	0.4%
3	5	0.6%
2	80	9.0%
1	118	13.2%
0	678	76.1%





## Ticket

Text

<b>Distinct</b>	681
<b>Distinct (%)</b>	76.4%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.1 KiB



## Length

<b>Max length</b>	18
<b>Median length</b>	17
<b>Mean length</b>	6.7508418
<b>Min length</b>	3

## Characters and Unicode

Total characters	6015
Distinct characters	35
Distinct categories	5 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
Distinct scripts	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )

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

Unique

Unique	547	?
Unique (%)	61.4%	

**Sample**

<b>1st row</b>	A/5 21171
<b>2nd row</b>	PC 17599
<b>3rd row</b>	STON/O2. 3101282
<b>4th row</b>	113803
<b>5th row</b>	373450

Value	Count	Frequency (%)
pc	60	5.3%
c.a	27	2.4%
a/5	17	1.5%
ca	14	1.2%
ston/o	12	1.1%
2	12	1.1%
sc/paris	9	0.8%
w./c	9	0.8%
soton/o.q	8	0.7%
347082	7	0.6%
Other values (709)	955	84.5%



## Most occurring characters

Value	Count	Frequency (%)
3	746	12.4%
1	689	11.5%
2	594	9.9%
7	490	8.1%
4	464	7.7%
6	422	7.0%
0	406	6.7%
5	387	6.4%

Value	Count	Frequency (%)
9	328	5.5%
8	282	4.7%
Other values (25)	1207	20.1%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	4808	79.9%
Uppercase Letter	652	10.8%
Other Punctuation	295	4.9%
Space Separator	239	4.0%
Lowercase Letter	21	0.3%

## Most frequent character per category

*Uppercase Letter*

Value	Count	Frequency (%)
C	151	23.2%
O	100	15.3%
P	98	15.0%
A	82	12.6%
S	74	11.3%
N	40	6.1%
T	36	5.5%
W	16	2.5%
Q	15	2.3%
I	11	1.7%
Other values (6)	29	4.4%

*Decimal Number*

Value	Count	Frequency (%)
3	746	15.5%
1	689	14.3%
2	594	12.4%
7	490	10.2%
4	464	9.7%
6	422	8.8%
0	406	8.4%
5	387	8.0%
9	328	6.8%
8	282	5.9%

*Lowercase Letter*

Value	Count	Frequency (%)
a	6	28.6%
s	5	23.8%
r	4	19.0%
i	4	19.0%

Value	Count	Frequency (%)
I	1	4.8%
e	1	4.8%

*Other Punctuation*

Value	Count	Frequency (%)
.	197	66.8%
/	98	33.2%

*Space Separator*

Value	Count	Frequency (%)
	239	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Common	5342	88.8%
Latin	673	11.2%

## Most frequent character per script

*Latin*

Value	Count	Frequency (%)
C	151	22.4%
O	100	14.9%
P	98	14.6%
A	82	12.2%
S	74	11.0%
N	40	5.9%
T	36	5.3%
W	16	2.4%
Q	15	2.2%
I	11	1.6%
Other values (12)	50	7.4%

*Common*

Value	Count	Frequency (%)
3	746	14.0%
1	689	12.9%
2	594	11.1%
7	490	9.2%
4	464	8.7%
6	422	7.9%
0	406	7.6%
5	387	7.2%
9	328	6.1%
8	282	5.3%
Other values (3)	534	10.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	6015	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
3	746	12.4%
1	689	11.5%
2	594	9.9%
7	490	8.1%
4	464	7.7%
6	422	7.0%
0	406	6.7%
5	387	6.4%
9	328	5.5%
8	282	4.7%
Other values (25)	1207	20.1%





**Fare**

Real number (R)

ZEROS

Distinct

248

Distinct (%)

27.8%

Missing

0

Missing (%)

0.0%

Infinite

0

Infinite (%)

0.0%

Mean

32.204208

Minimum

0

Maximum

512.3292

Zeros

15

Zeros (%)

1.7%

Negative

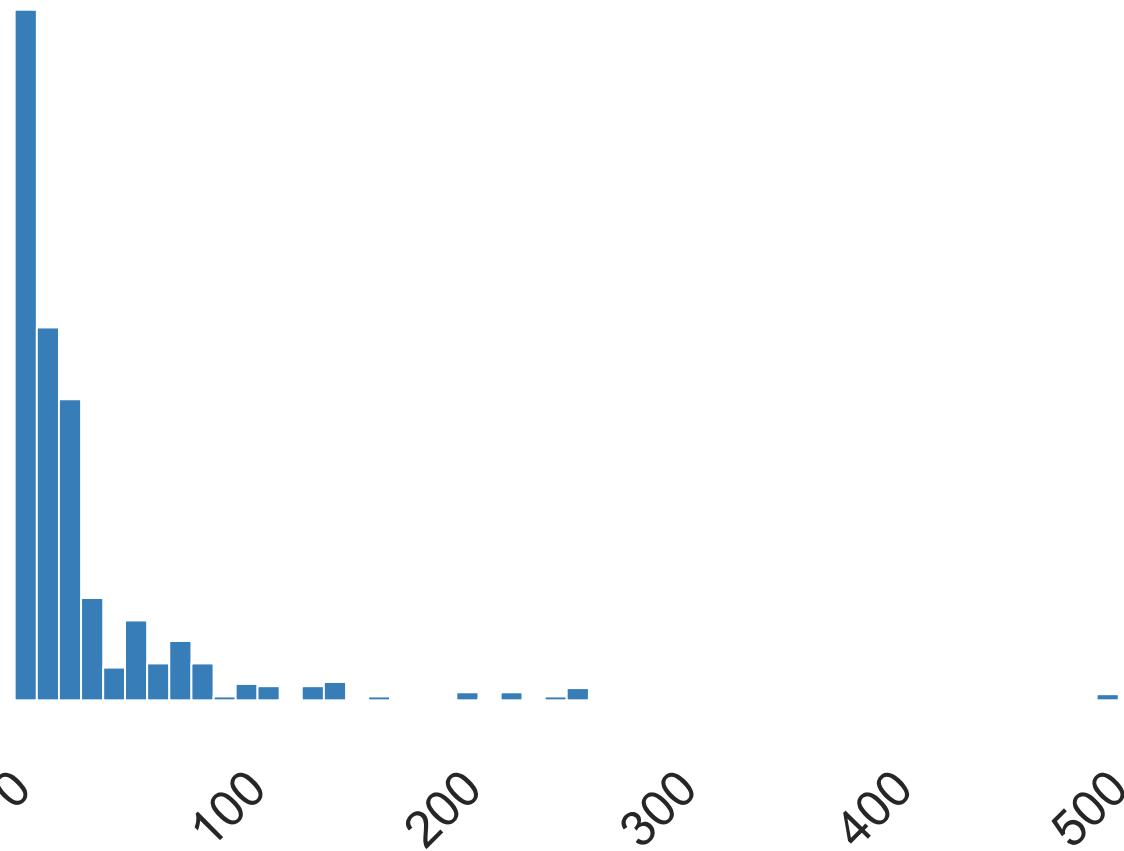
0

Negative (%)

0.0%

Memory size

7.1 KiB

**Quantile statistics**

Minimum

0

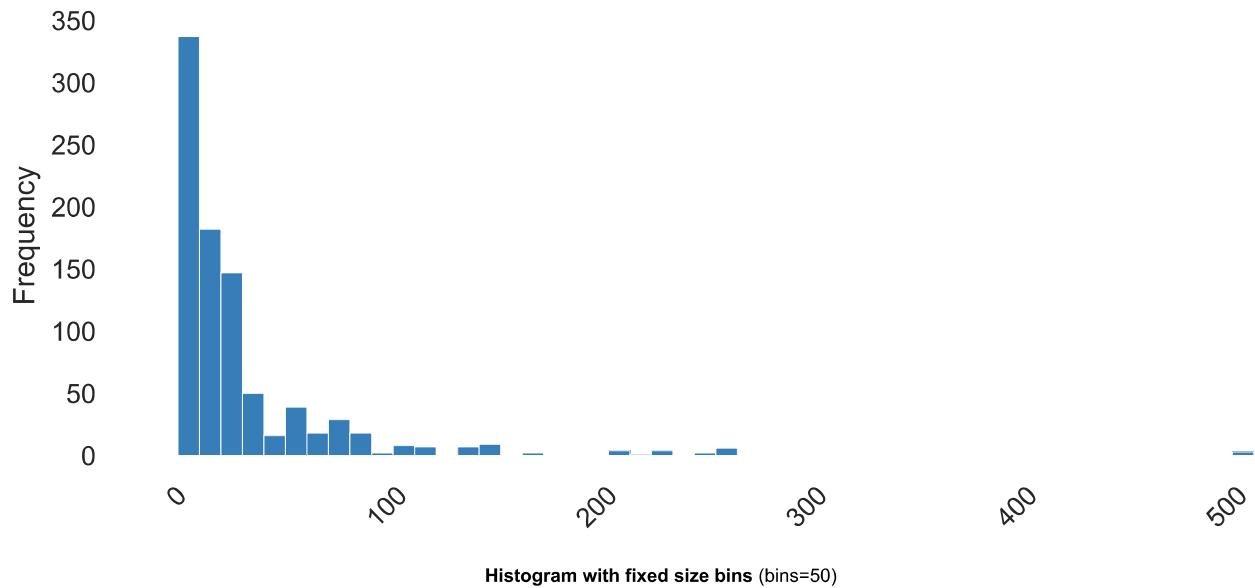
5-th percentile

7.225

Q1

7.9104

<b>median</b>	14.4542
<b>Q3</b>	31
<b>95-th percentile</b>	112.07915
<b>Maximum</b>	512.3292
<b>Range</b>	512.3292
<b>Interquartile range (IQR)</b>	23.0896
<b>Descriptive statistics</b>	
<b>Standard deviation</b>	49.693429
<b>Coefficient of variation (CV)</b>	1.5430725
<b>Kurtosis</b>	33.398141
<b>Mean</b>	32.204208
<b>Median Absolute Deviation (MAD)</b>	6.9042
<b>Skewness</b>	4.7873165
<b>Sum</b>	28693.949
<b>Variance</b>	2469.4368
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
8.05	43	4.8%
13	42	4.7%
7.8958	38	4.3%
7.75	34	3.8%
26	31	3.5%
10.5	24	2.7%
7.925	18	2.0%
7.775	16	1.8%
7.2292	15	1.7%
0	15	1.7%
Other values (238)	615	69.0%

Value	Count	Frequency (%)
0	15	1.7%
4.0125	1	0.1%
5	1	0.1%
6.2375	1	0.1%
6.4375	1	0.1%
6.45	1	0.1%
6.4958	2	0.2%
6.75	2	0.2%
6.8583	1	0.1%
6.95	1	0.1%

Value	Count	Frequency (%)
512.3292	3	0.3%
263	4	0.4%
262.375	2	0.2%
247.5208	2	0.2%
227.525	4	0.4%
221.7792	1	0.1%
211.5	1	0.1%
211.3375	3	0.3%
164.8667	2	0.2%
153.4625	3	0.3%





## Cabin

Text

MISSING

<b>Distinct</b>	147
<b>Distinct (%)</b>	72.1%
<b>Missing</b>	687
<b>Missing (%)</b>	77.1%
<b>Memory size</b>	7.1 KiB



### Length

<b>Max length</b>	15
<b>Median length</b>	3
<b>Mean length</b>	3.5882353
<b>Min length</b>	1

## Characters and Unicode

Total characters	732
Distinct characters	19
Distinct categories	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
Distinct scripts	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )

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

Unique

Unique	101	?
Unique (%)	49.5%	

**Sample**

<b>1st row</b>	C85
<b>2nd row</b>	C123
<b>3rd row</b>	E46
<b>4th row</b>	G6
<b>5th row</b>	C103

Value	Count	Frequency (%)
c23	4	1.7%
c27	4	1.7%
g6	4	1.7%
b96	4	1.7%
b98	4	1.7%
f	4	1.7%
c25	4	1.7%
f33	3	1.3%
e101	3	1.3%
f2	3	1.3%
Other values (151)	201	84.5%



## Most occurring characters

Value	Count	Frequency (%)
2	72	9.8%
C	71	9.7%
B	64	8.7%
1	61	8.3%
3	59	8.1%
6	51	7.0%
5	45	6.1%
4	37	5.1%

Value	Count	Frequency (%)
8	37	5.1%
	34	4.6%
Other values (9)	201	27.5%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	460	62.8%
Uppercase Letter	238	32.5%
Space Separator	34	4.6%

## Most frequent character per category

*Decimal Number*

Value	Count	Frequency (%)
2	72	15.7%
1	61	13.3%
3	59	12.8%
6	51	11.1%
5	45	9.8%
4	37	8.0%
8	37	8.0%
7	34	7.4%
9	33	7.2%
0	31	6.7%

*Uppercase Letter*

Value	Count	Frequency (%)
C	71	29.8%
B	64	26.9%
D	34	14.3%
E	33	13.9%
A	15	6.3%
F	13	5.5%
G	7	2.9%
T	1	0.4%

*Space Separator*

Value	Count	Frequency (%)
	34	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Common	494	67.5%
Latin	238	32.5%

## Most frequent character per script

*Common*

Value	Count	Frequency (%)
2	72	14.6%
1	61	12.3%
3	59	11.9%
6	51	10.3%
5	45	9.1%
4	37	7.5%
8	37	7.5%
	34	6.9%
7	34	6.9%
9	33	6.7%

*Latin*

Value	Count	Frequency (%)
C	71	29.8%
B	64	26.9%
D	34	14.3%
E	33	13.9%
A	15	6.3%
F	13	5.5%
G	7	2.9%
T	1	0.4%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	732	100.0%

## Most frequent character per block

ASCII

Value	Count	Frequency (%)
2	72	9.8%
C	71	9.7%
B	64	8.7%
1	61	8.3%
3	59	8.1%
6	51	7.0%
5	45	6.1%
4	37	5.1%
8	37	5.1%
	34	4.6%
Other values (9)	201	27.5%





**Embarked**

Categorical

<b>Distinct</b>	3
<b>Distinct (%)</b>	0.3%
<b>Missing</b>	2
<b>Missing (%)</b>	0.2%
<b>Memory size</b>	7.1 KiB

**Length**

<b>Max length</b>	1
<b>Median length</b>	1
<b>Mean length</b>	1
<b>Min length</b>	1

**Characters and Unicode**

<b>Total characters</b>	889
<b>Distinct characters</b>	3
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )

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

**Unique**

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

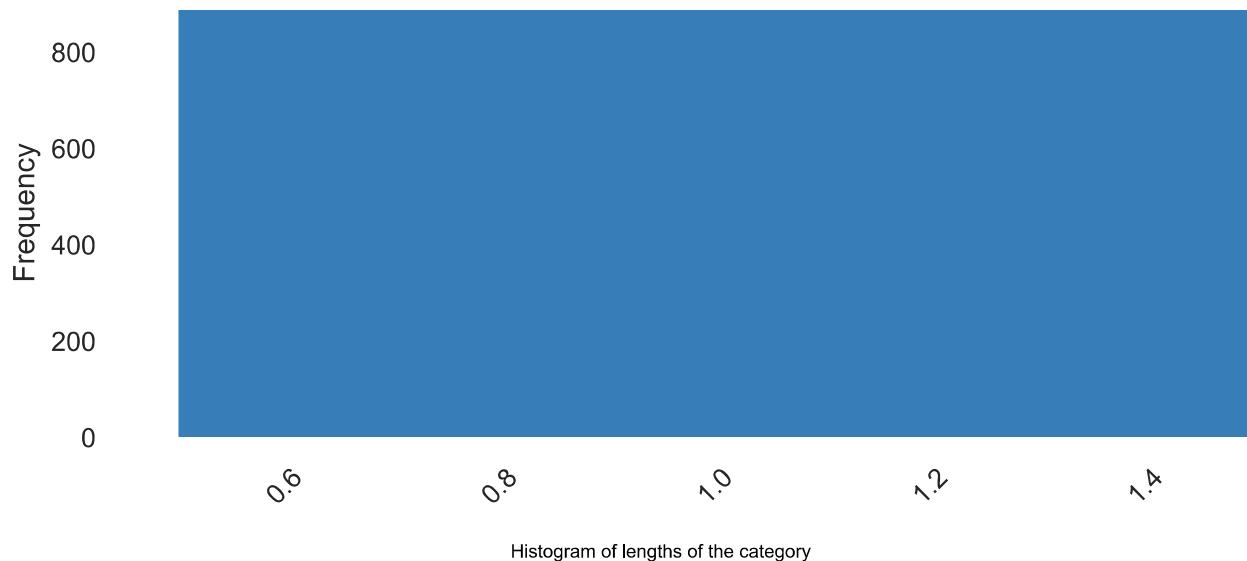
**Sample**

<b>1st row</b>	S
<b>2nd row</b>	C
<b>3rd row</b>	S
<b>4th row</b>	S
<b>5th row</b>	S

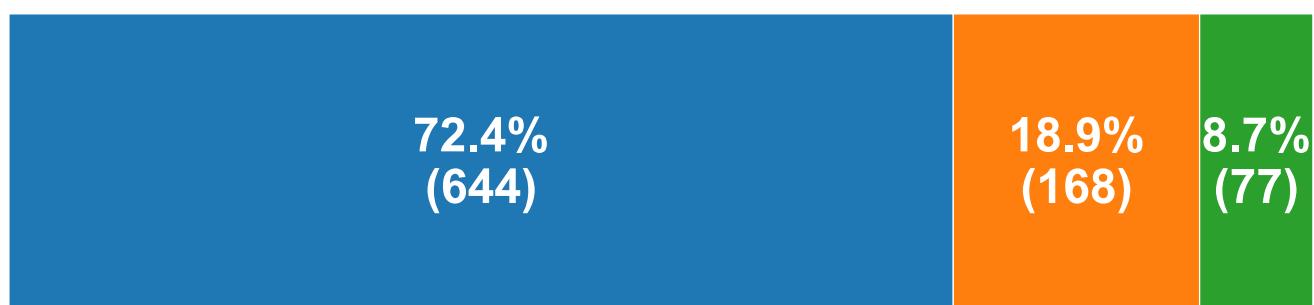
## Common Values

Value	Count	Frequency (%)
S	644	72.3%
C	168	18.9%
Q	77	8.6%
(Missing)	2	0.2%

## Length



## Common Values (Plot)



Value	Count	Frequency (%)
s	644	72.4%
c	168	18.9%
q	77	8.7%

## Most occurring characters

Value	Count	Frequency (%)
S	644	72.4%
C	168	18.9%
Q	77	8.7%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	889	100.0%

## Most frequent character per category

*Uppercase Letter*

Value	Count	Frequency (%)
S	644	72.4%
C	168	18.9%
Q	77	8.7%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	889	100.0%

## Most frequent character per script

*Latin*

Value	Count	Frequency (%)
S	644	72.4%
C	168	18.9%
Q	77	8.7%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	889	100.0%

## Most frequent character per block

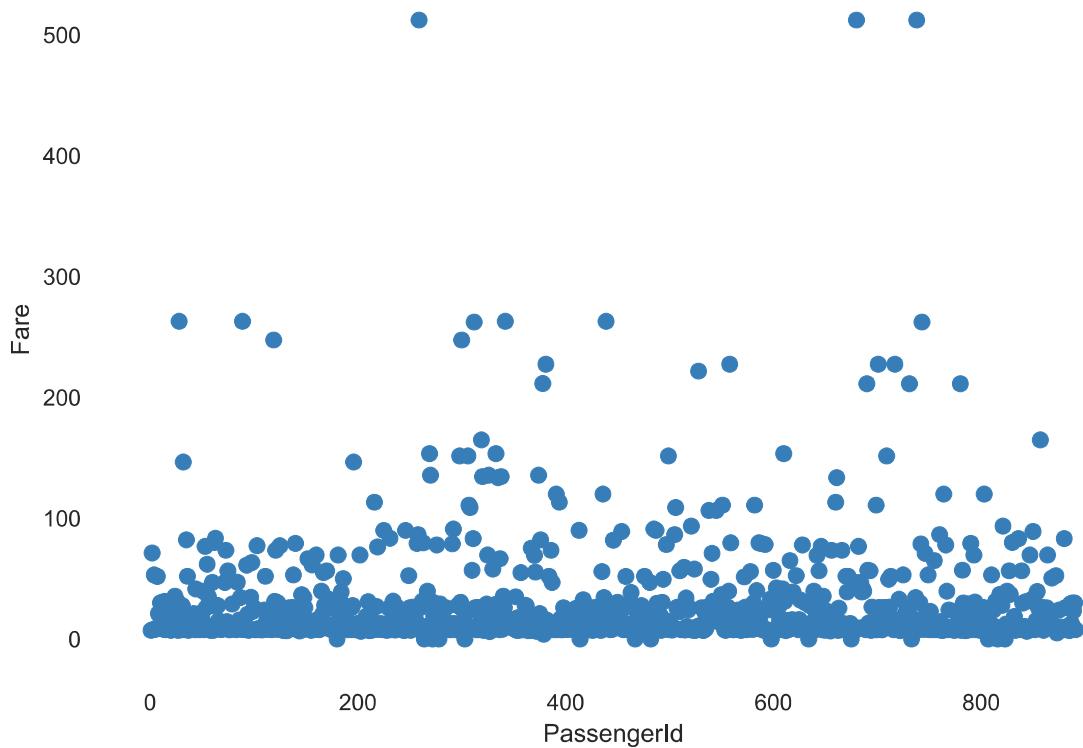
ASCII

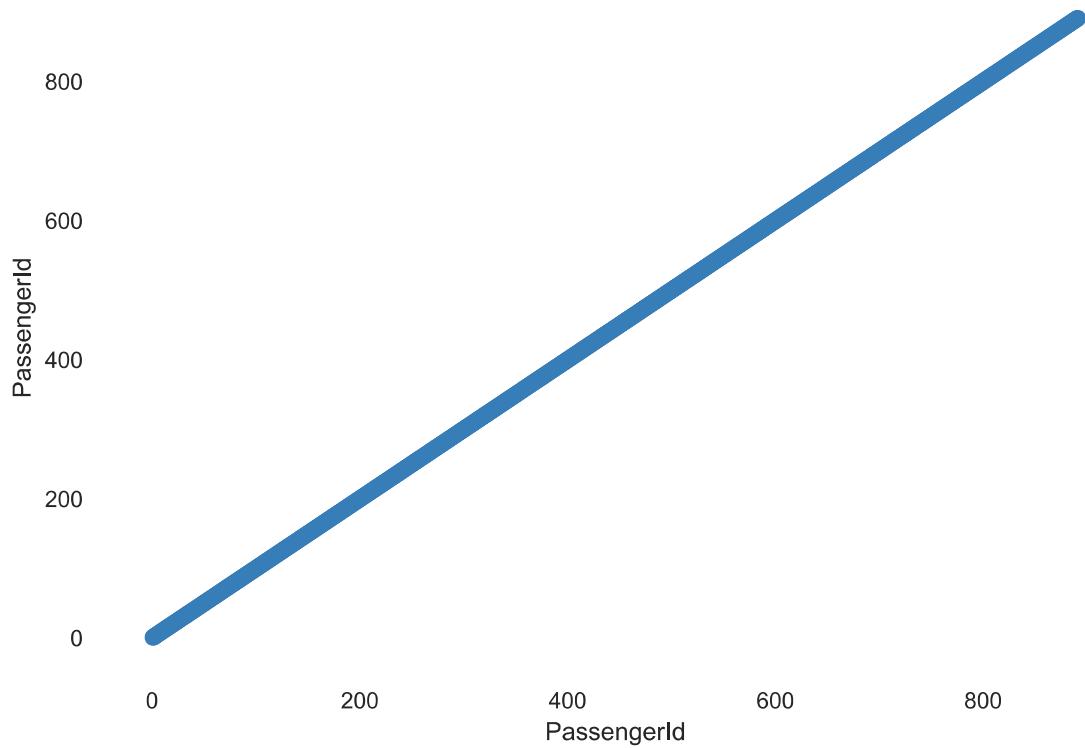
Value	Count	Frequency (%)
S	644	72.4%
C	168	18.9%
Q	77	8.7%

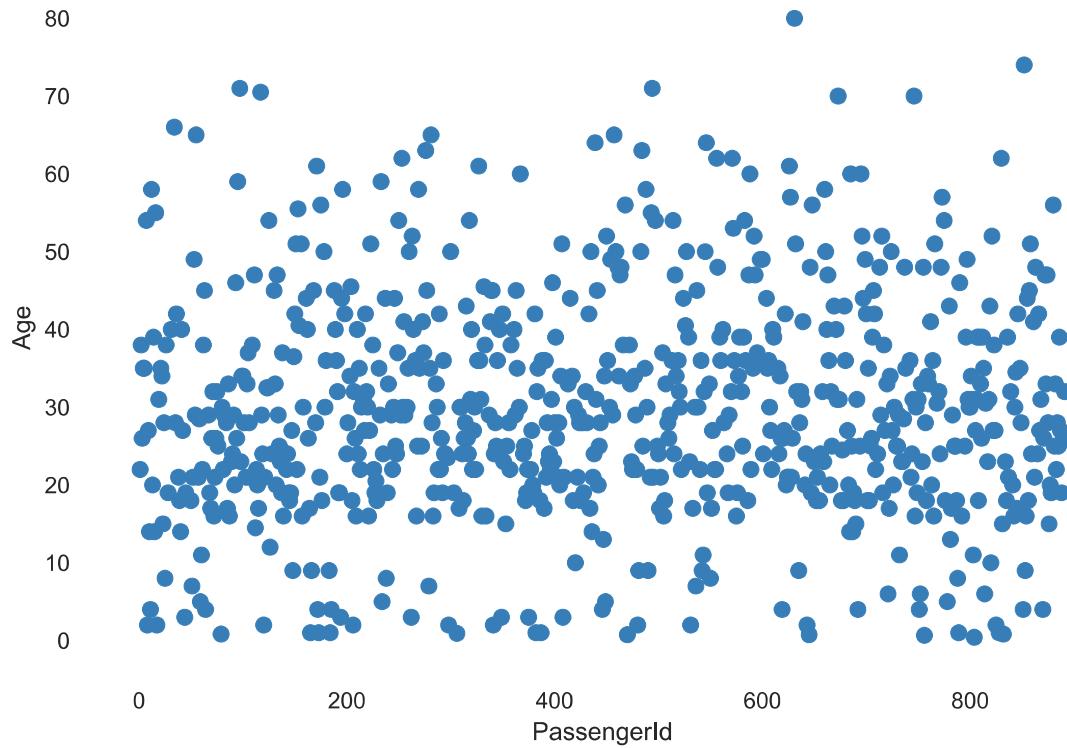


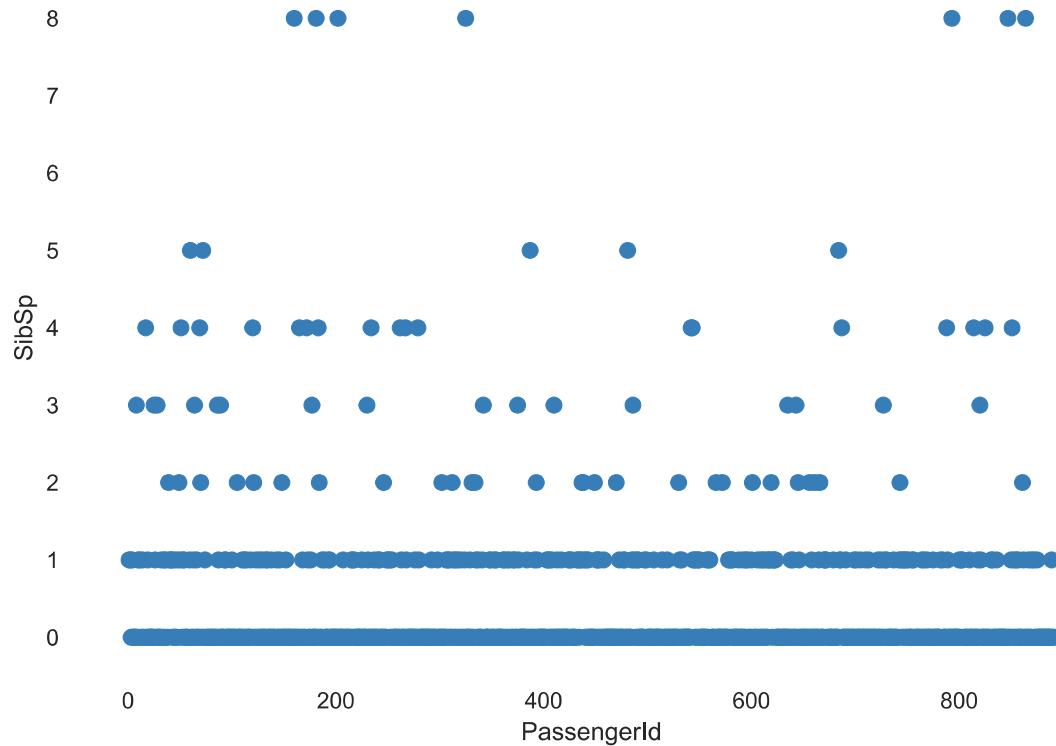


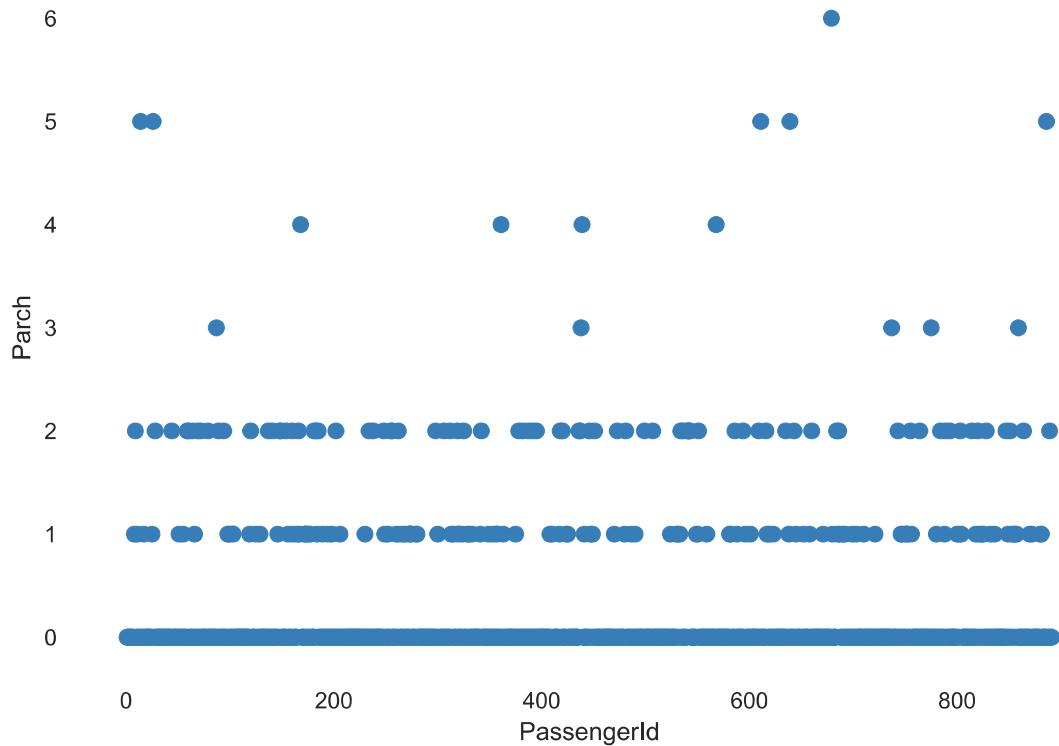
# Interactions



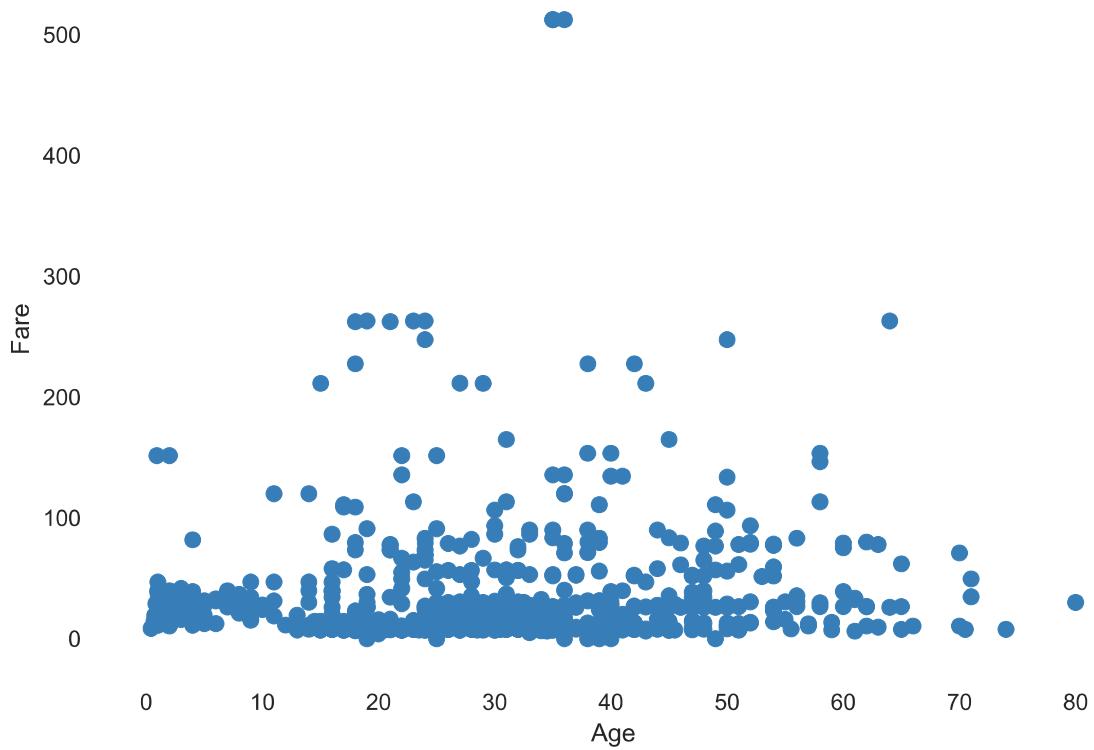


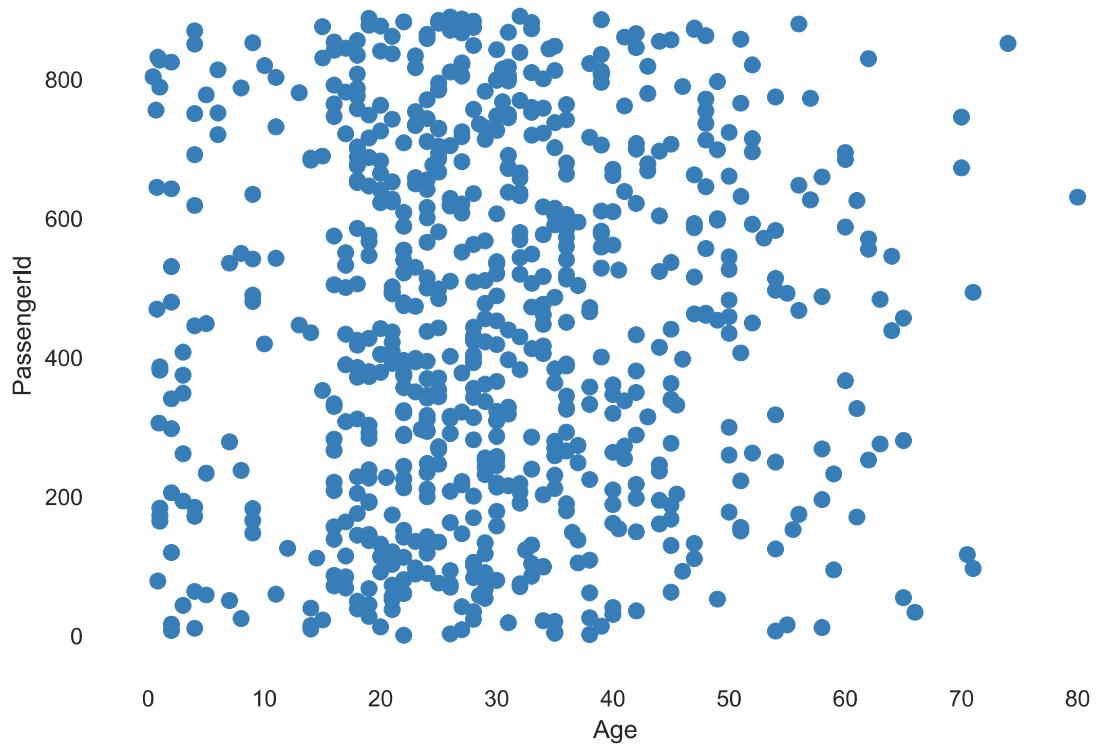


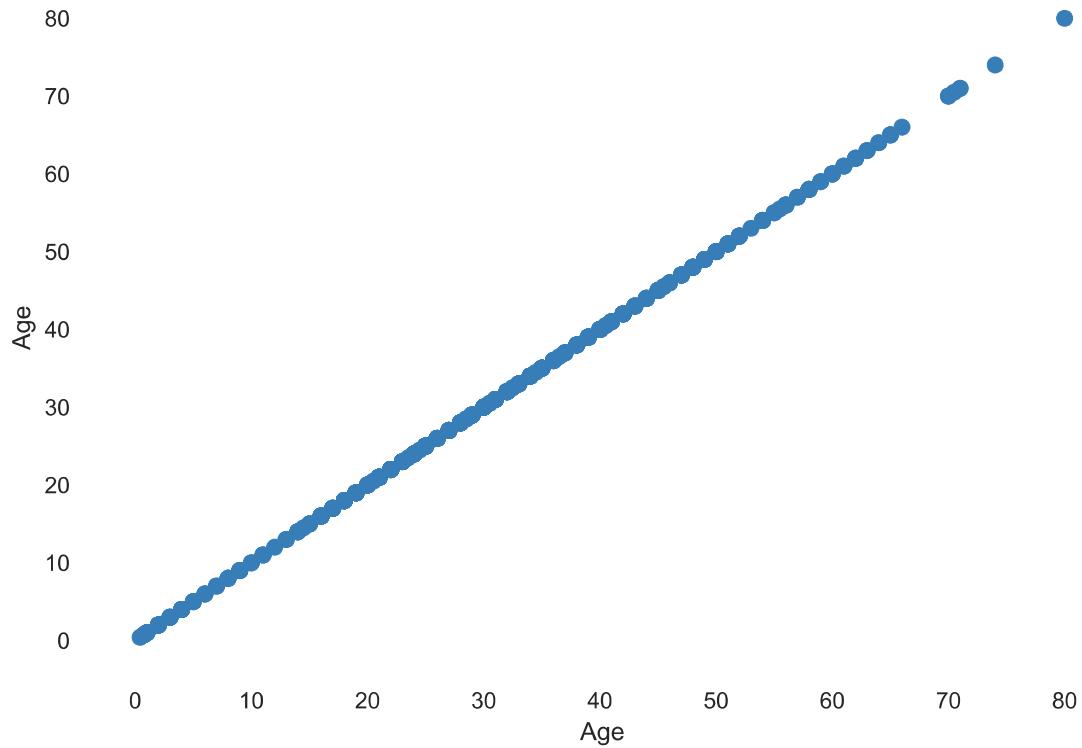


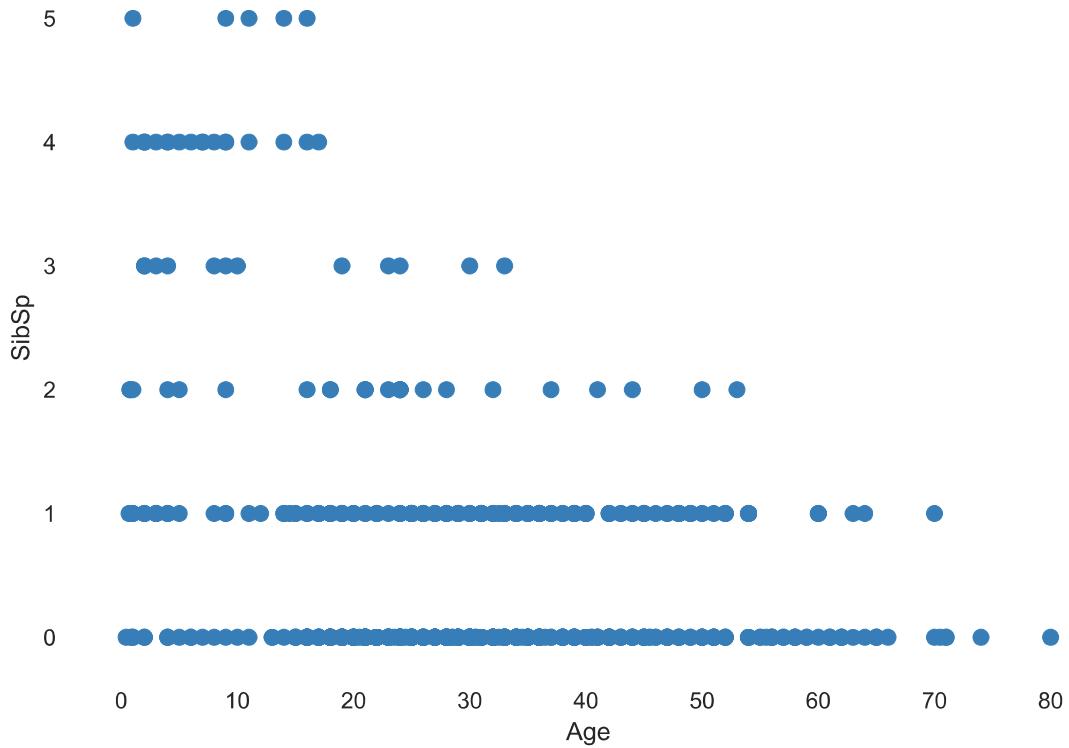


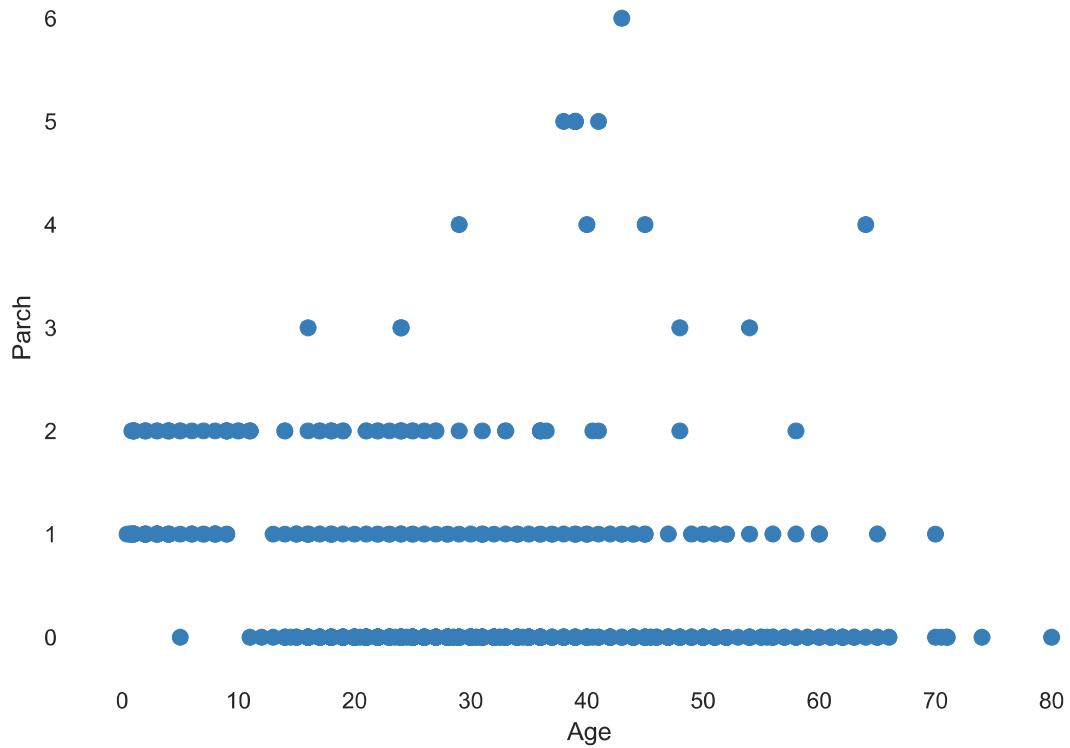




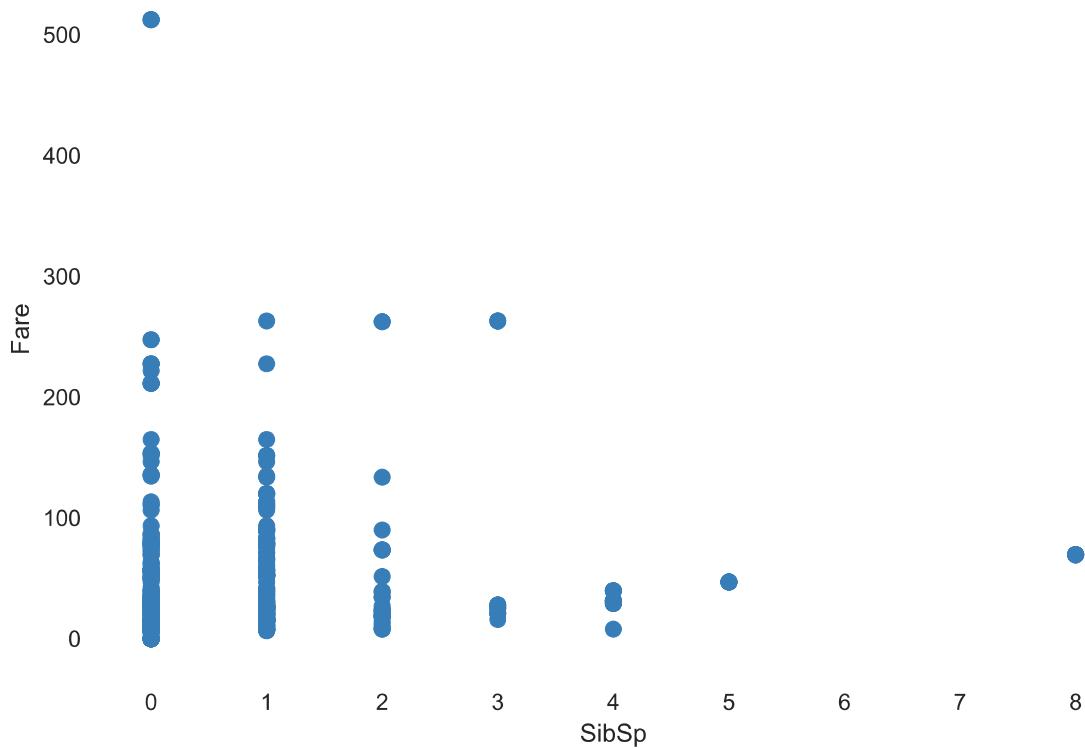


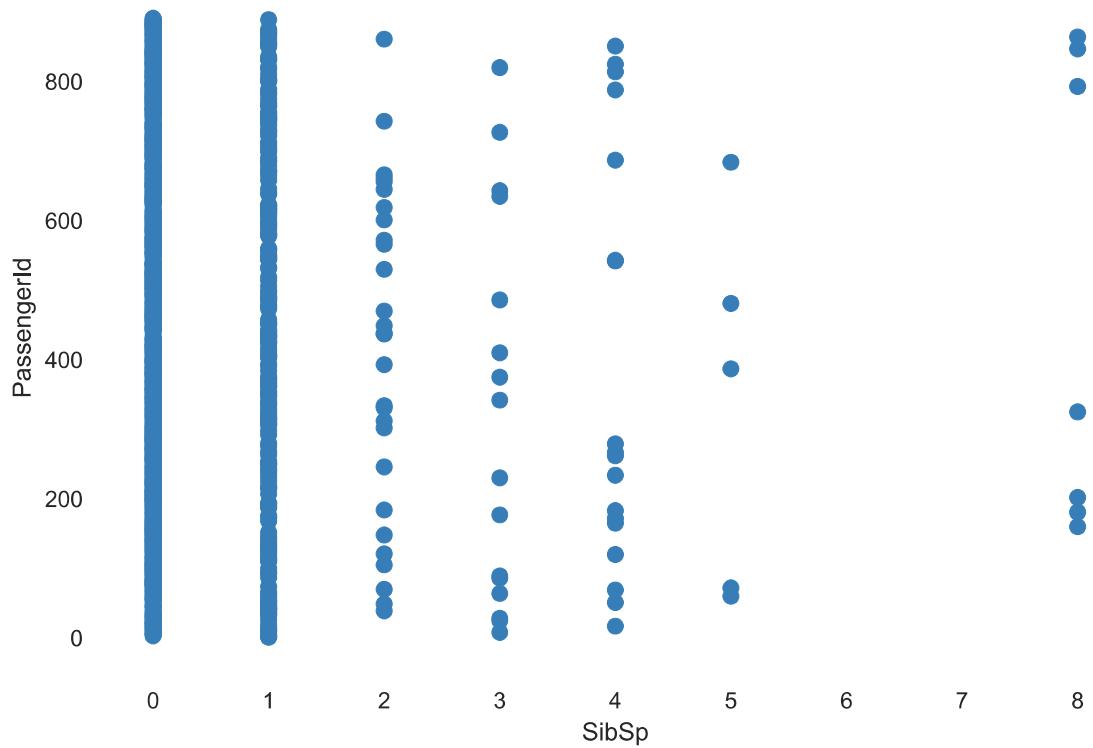


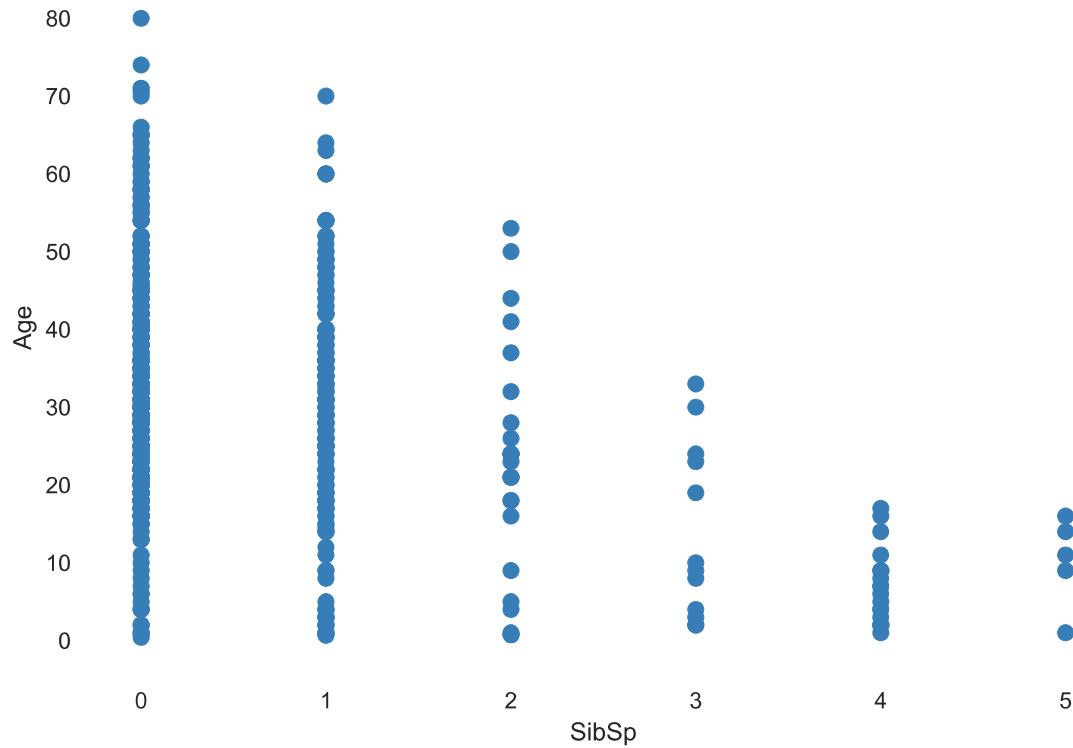


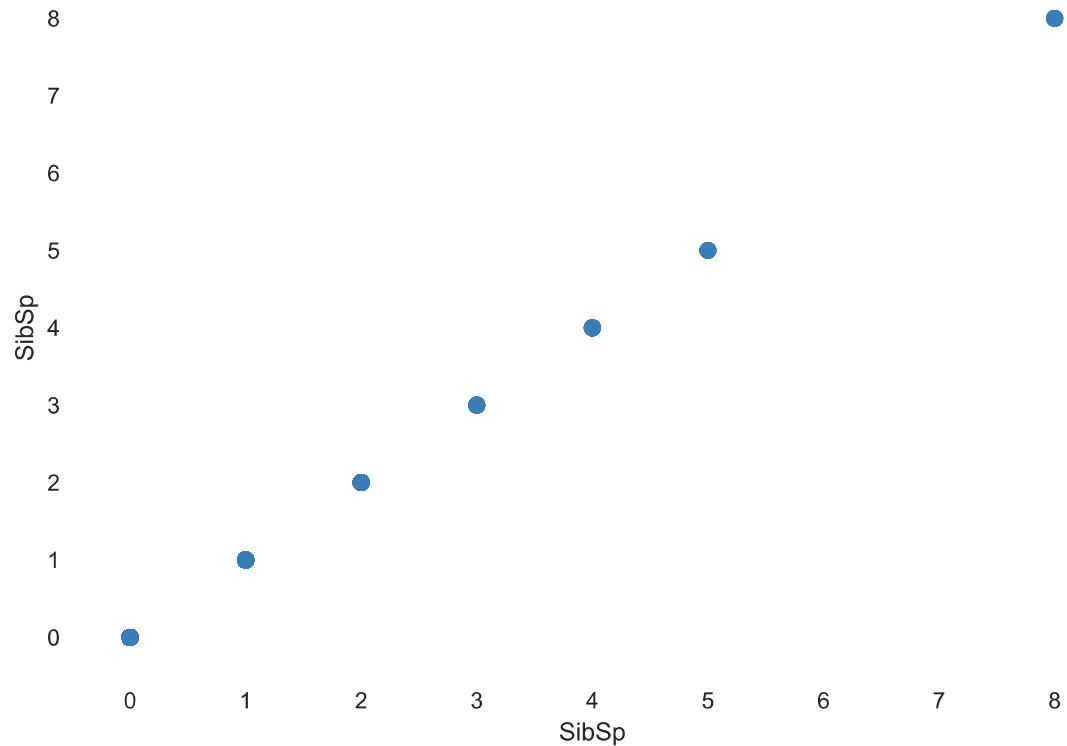


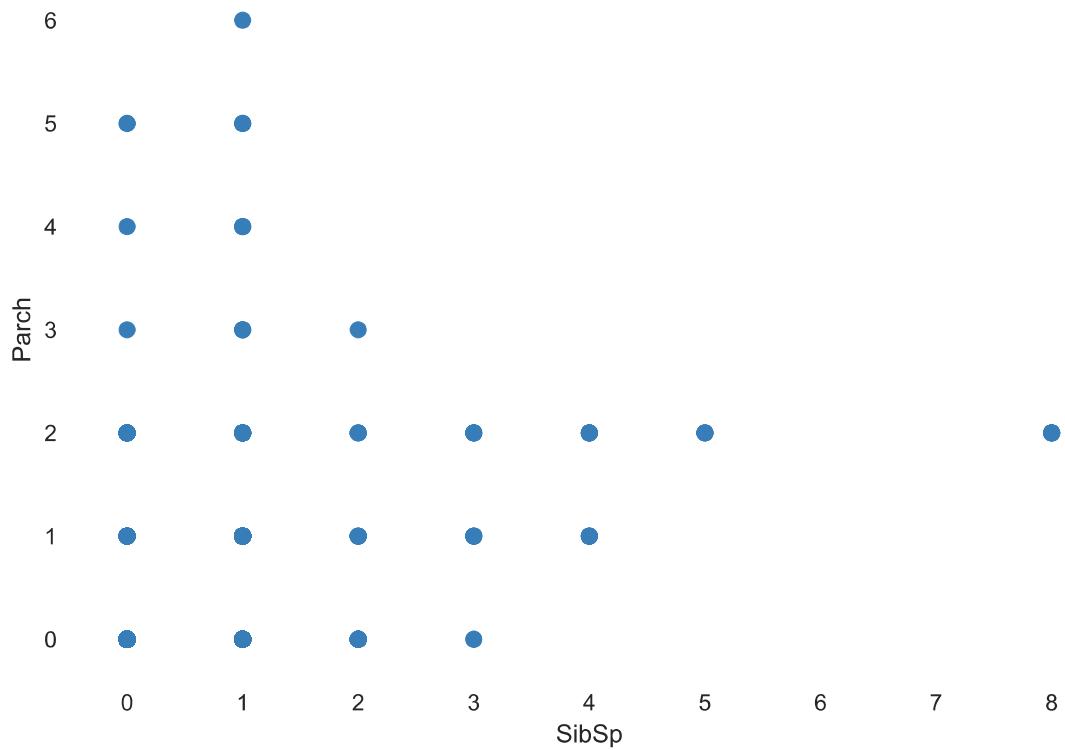




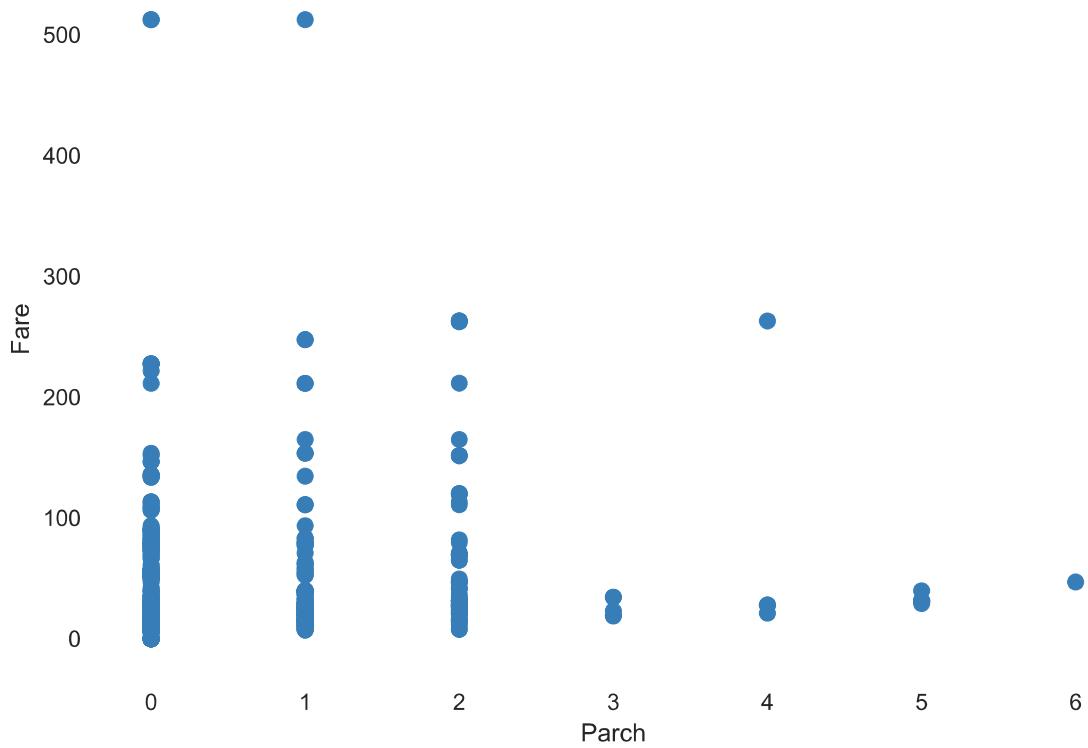


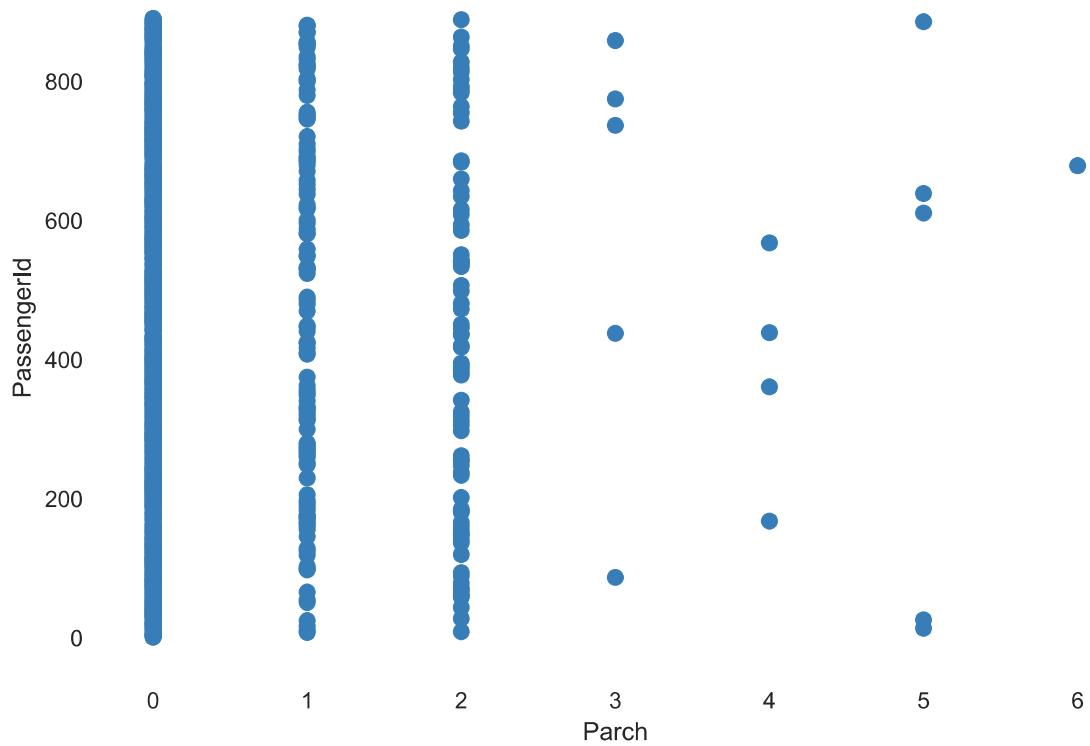


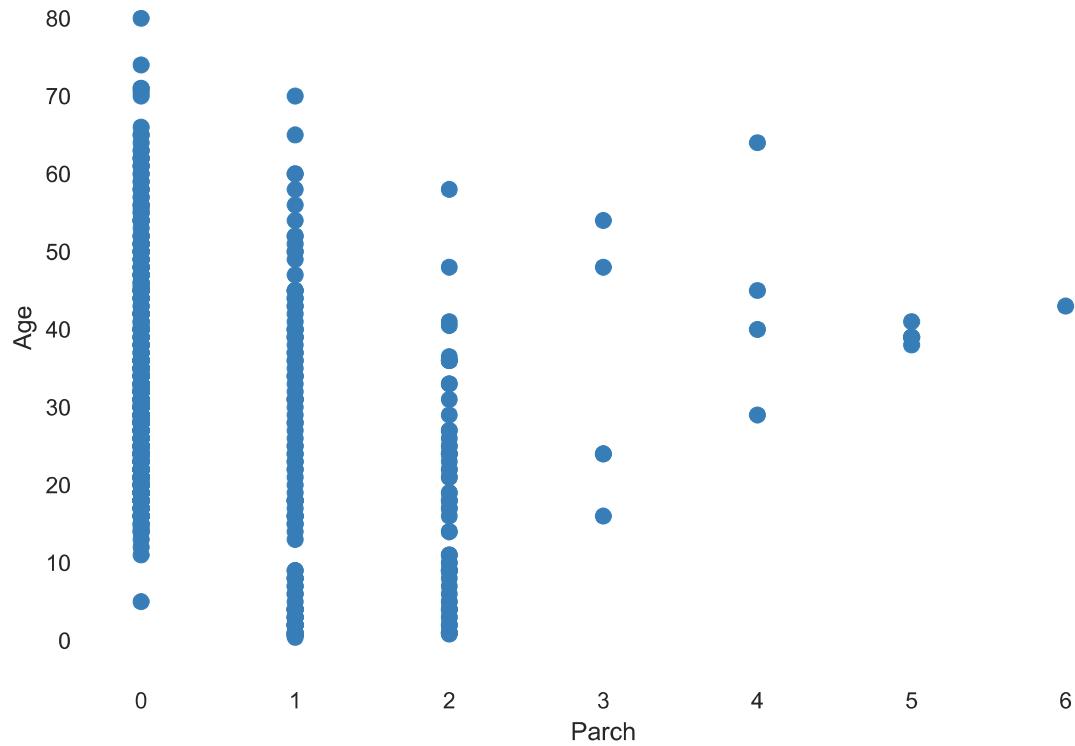


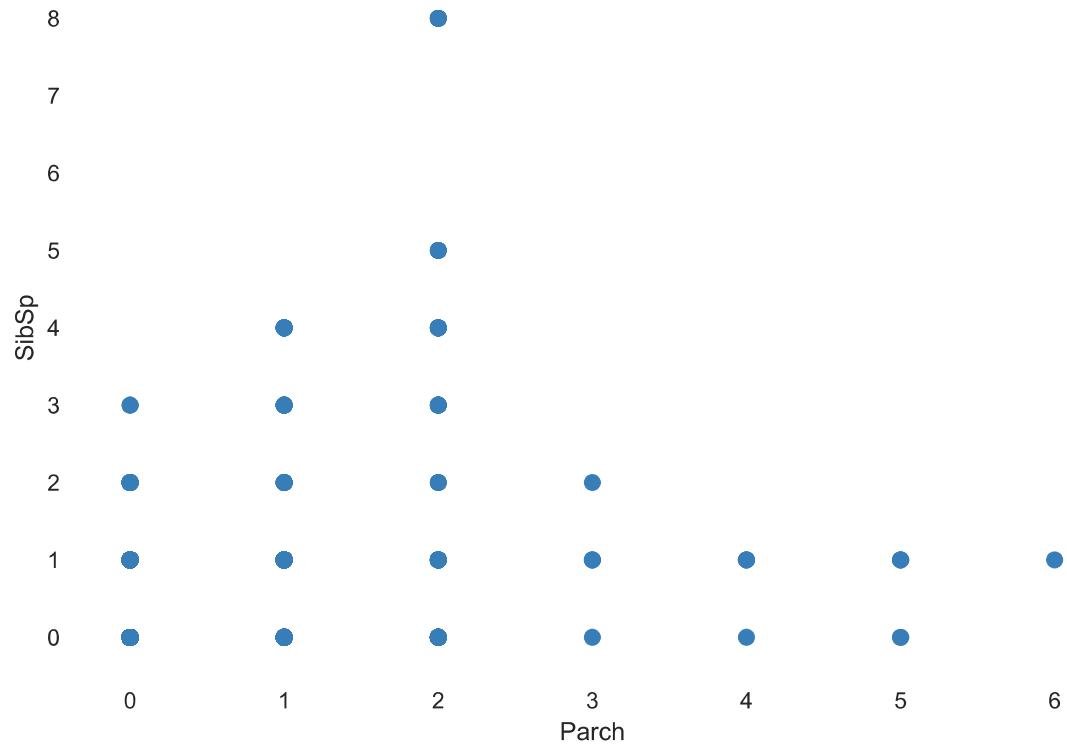


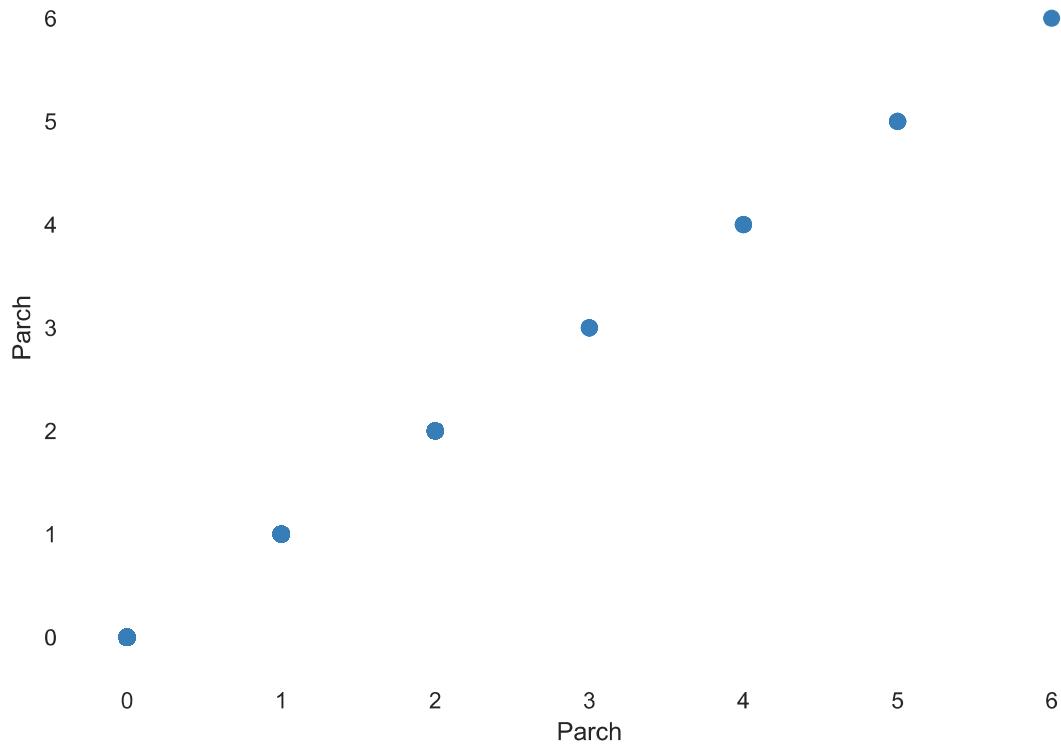




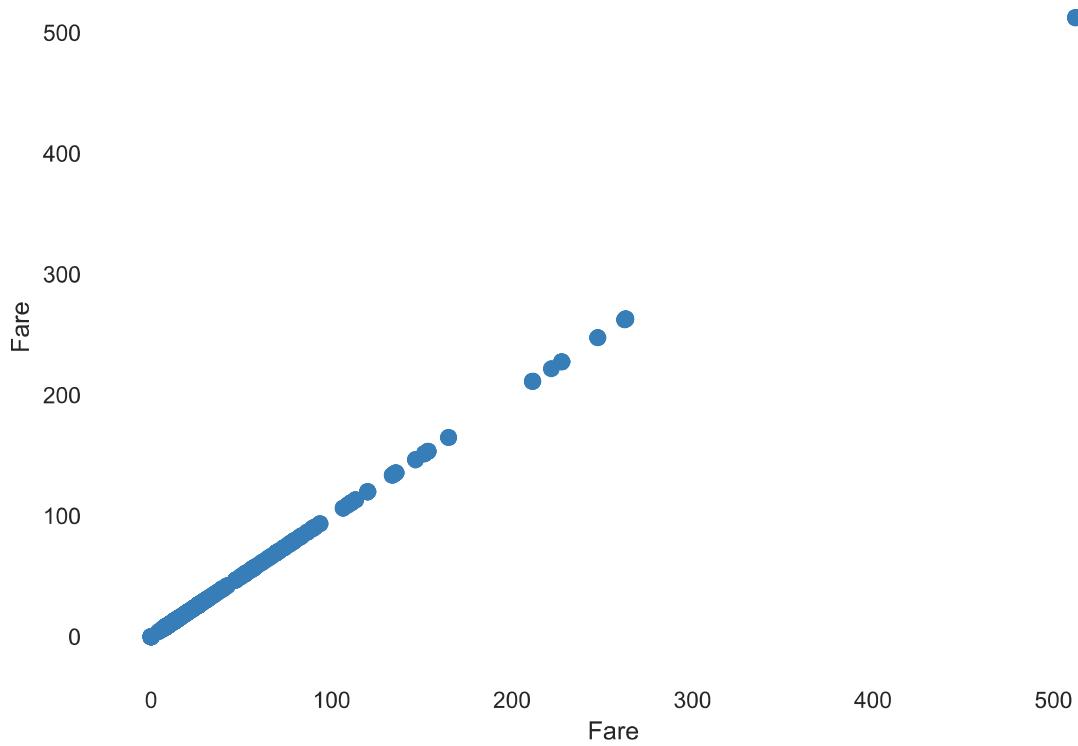


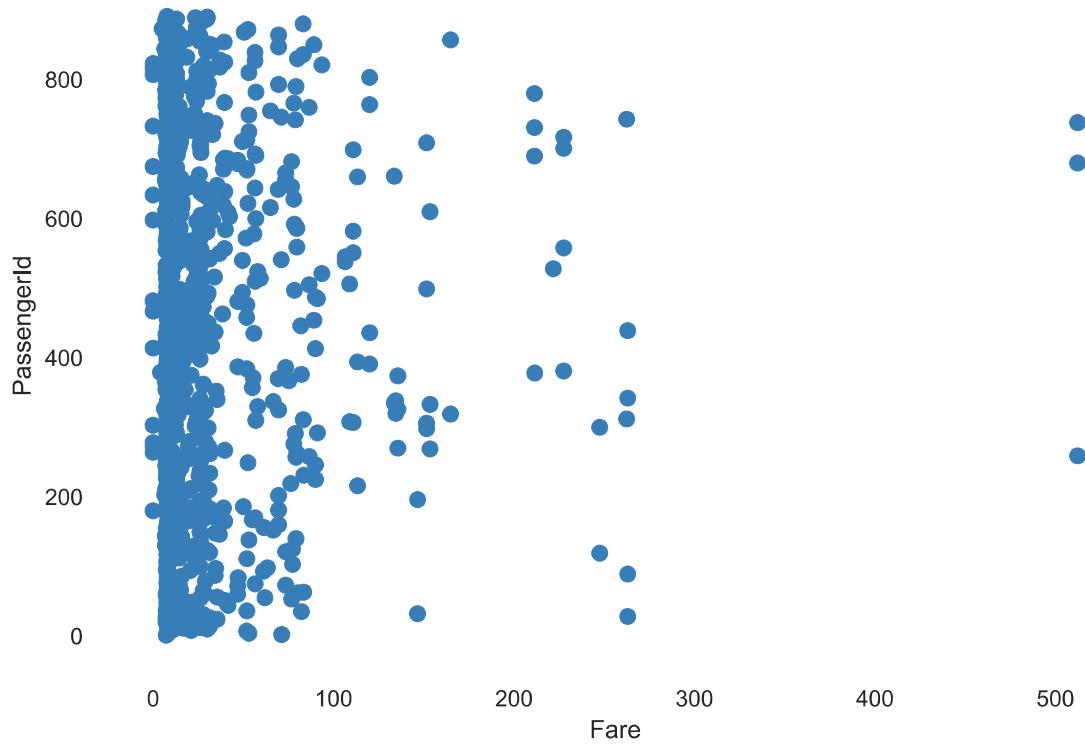


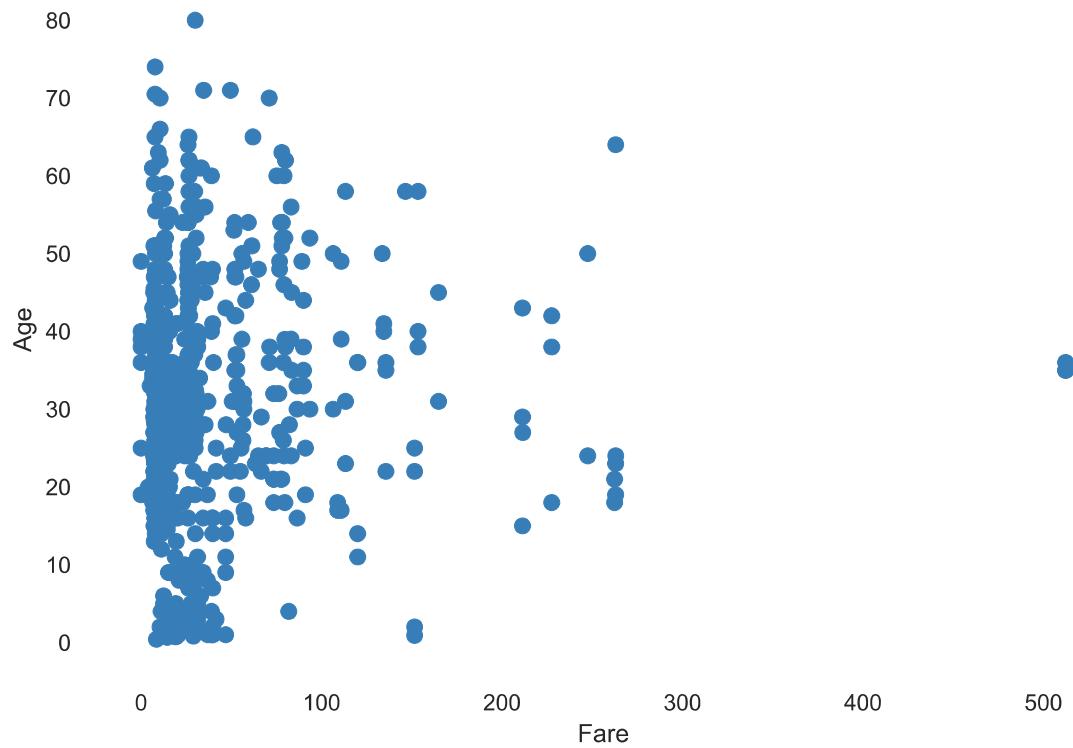


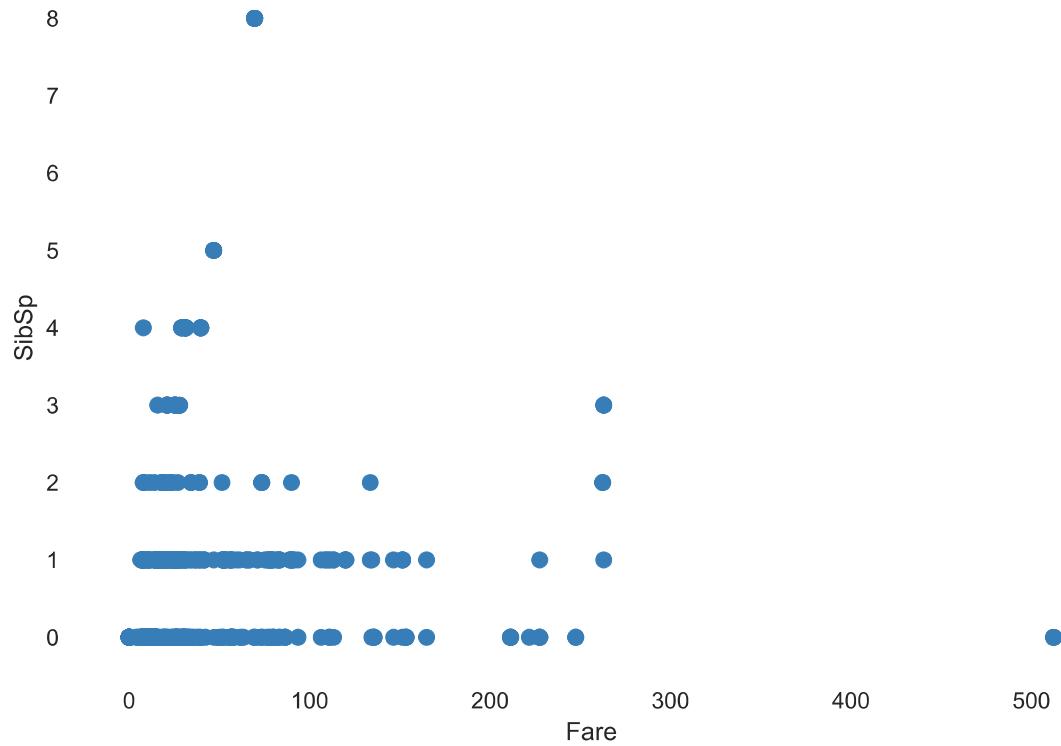


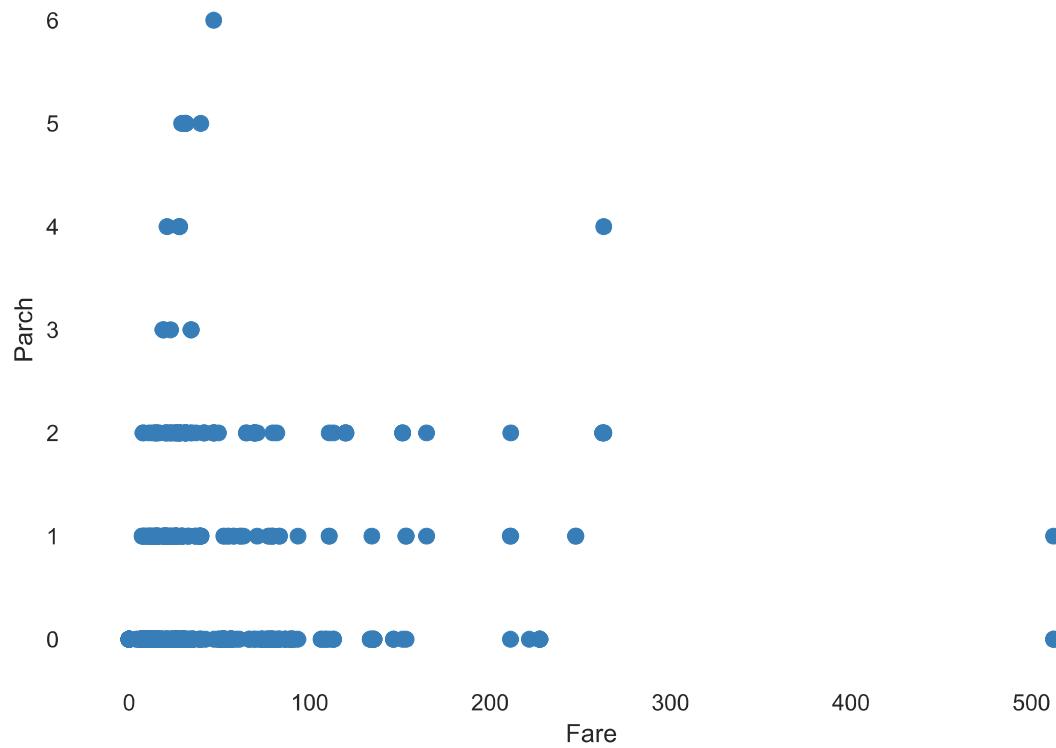






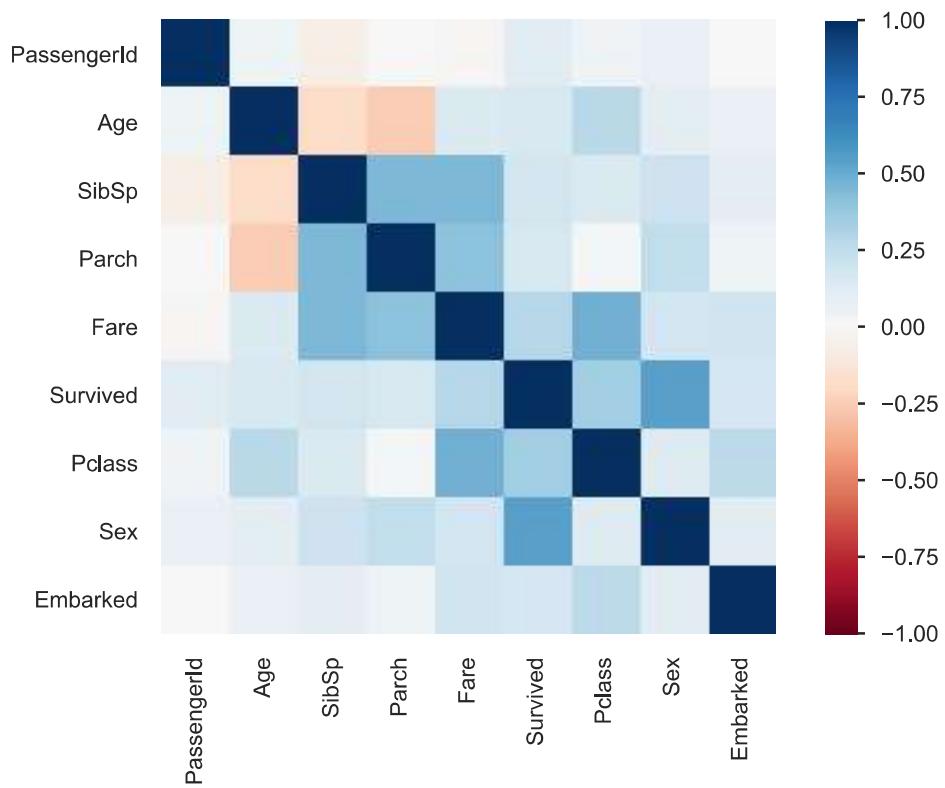








# Correlations

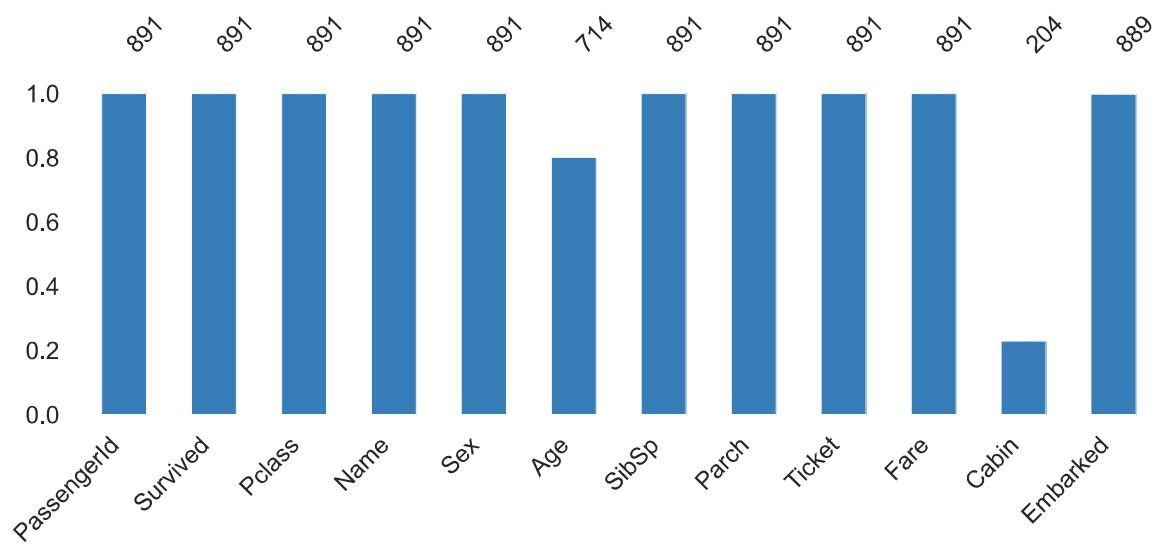


## Profile Train

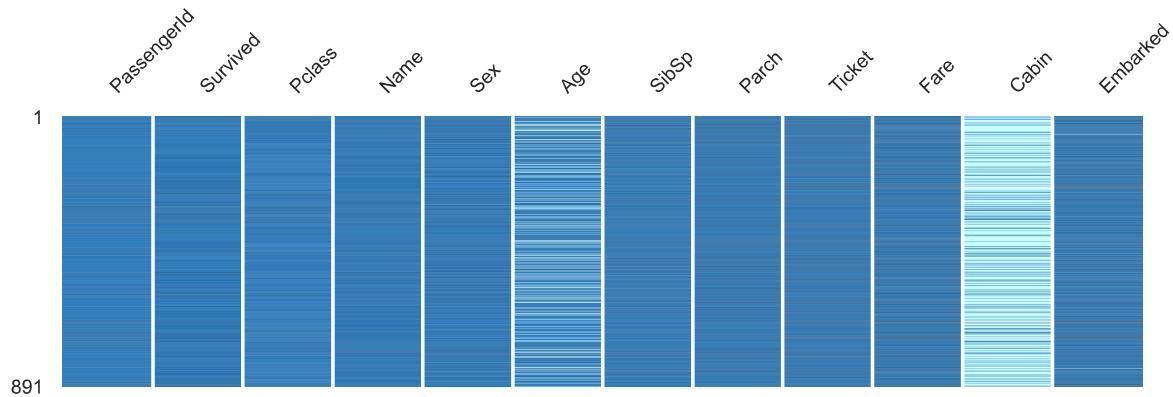
	<b>PassengerId</b>	<b>Age</b>	<b>SibSp</b>	<b>Parch</b>	<b>Fare</b>	<b>Survived</b>	<b>Pclass</b>	<b>Sex</b>	<b>Embarked</b>
<b>PassengerId</b>	1.000	0.041	-0.061	0.001	-0.014	0.104	0.032	0.066	0.000
<b>Age</b>	0.041	1.000	-0.182	-0.254	0.135	0.155	0.269	0.099	0.065
<b>SibSp</b>	-0.061	-0.182	1.000	0.450	0.447	0.187	0.148	0.206	0.092
<b>Parch</b>	0.001	-0.254	0.450	1.000	0.410	0.157	0.022	0.247	0.052
<b>Fare</b>	-0.014	0.135	0.447	0.410	1.000	0.283	0.479	0.189	0.196
<b>Survived</b>	0.104	0.155	0.187	0.157	0.283	1.000	0.337	0.540	0.166
<b>Pclass</b>	0.032	0.269	0.148	0.022	0.479	0.337	1.000	0.130	0.260
<b>Sex</b>	0.066	0.099	0.206	0.247	0.189	0.540	0.130	1.000	0.113
<b>Embarked</b>	0.000	0.065	0.092	0.052	0.196	0.166	0.260	0.113	1.000



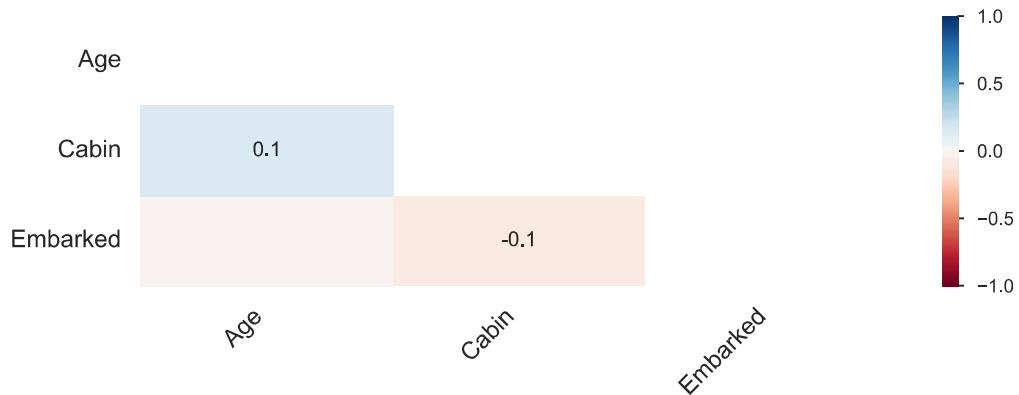
## Missing values



A simple visualization of nullity by column.



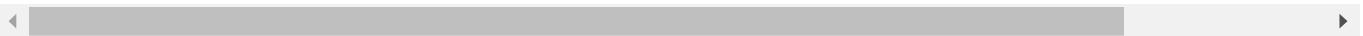
Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.



The correlation heatmap measures nullity correlation: how strongly the presence or absence of one variable affects the presence of another.

# Sample

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.95
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.45
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.3
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.3
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0



## Profile Train

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca	
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	Na
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	Na
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	Na
884	885	0	3	Suttehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	Na
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	Na
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	Na
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B4
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	Na
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C1
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Na



Report generated by YData ([https://ydata.ai/?utm\\_source=opensource&utm\\_medium=pandasprofiling&utm\\_campaign=report](https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report)).