

Back to list of profiles

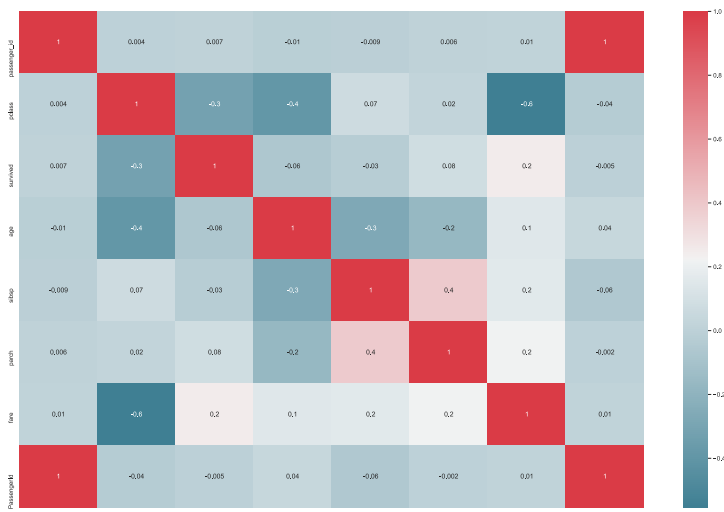
Overview

Dataset statistics

Number of variables	17
Number of observations	2200
Missing cells	6623
Missing cells (%)	17.71%
Duplicate rows	0
Duplicate rows (%)	0%
Total size in memory	292.31 KB
Average record size in memory	136.06 Bytes

Variable types

Numeric	8
Categorical	8
Unsupported	1



Variables

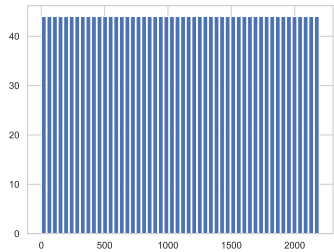
passenger_id

- Real number (ℝ)
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- UNIFORM
- UNIQUE

Distinct	2200
Distinct (%)	100%
Missing	0
Missing (%)	0%
Infinite	0
Infinite (%)	0%
Mean	1100.5

Minimum	1
Maximum	2200
Zeros	0
Zeros (%)	0%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

1	1
2	1
3	1

Quantile statistics

Minimum	1
---------	---

Descriptive statistics

Standard deviation	635.2296172776162
--------------------	-------------------

Minimum	1
5-th percentile	110.95
Q1	550.75
Median	1100.5
Q3	1650.25
95-th percentile	2090.05
Maximum	2200
Range	2199
Interquartile range (IQR)	1099.5

Standard deviation	665.2270172770102
Coefficient of variation (CV)	0.577219097935135
Kurtosis	-1.2000000000000002
Mean	1100.5
Median Absolute Deviation (MAD)	550
Skewness	0
Sum	2421100
Variance	403516.6666666667
Monotonicity	0

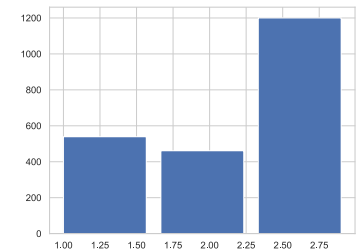
pclass

- Real number (ℝ)
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION

Distinct	3
Distinct (%)	0.14%
Missing	0
Missing (%)	0%
Infinite	0
Infinite (%)	0%
Mean	2.3

Minimum	1
Maximum	3
Zeros	0
Zeros (%)	0%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

3	1200
1	539
2	461

Quantile statistics

Minimum	1
5-th percentile	1
Q1	2
Median	3
Q3	3
95-th percentile	3
Maximum	3
Range	2
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	0.8369587926956457
Coefficient of variation (CV)	0.3638232254357677
Kurtosis	-1.300998263500697
Mean	2.3004545454545453
Median Absolute Deviation (MAD)	0
Skewness	-0.611088766811437
Sum	5061
Variance	0.7005000206705528
Monotonicity	0

survived

- Real number (ℝ)
- HIGH CORRELATION

Distinct	2
Distinct (%)	0.09%

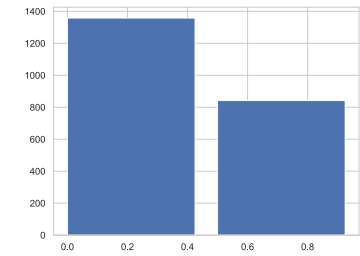
Minimum	0
Maximum	1

Histogram

ZEROS

Missing	0
Missing (%)	0%
Infinite	0
Infinite (%)	0%
Mean	0.38

Zeros	1358
Zeros (%)	61.73%
Negative	0
Negative (%)	0%
Memory size	17.31 KB



Frequent Values

0	1358
1	842

Quantile statistics

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

Descriptive statistics

Standard deviation	0.48616308099282374
Coefficient of variation (CV)	1.2702598315726985
Kurtosis	-1.7684345205139531
Mean	0.38272727272727275
Median Absolute Deviation (MAD)	0
Skewness	0.48288091152896895
Sum	842
Variance	0.2363545413204349
Monotonicity	0

age

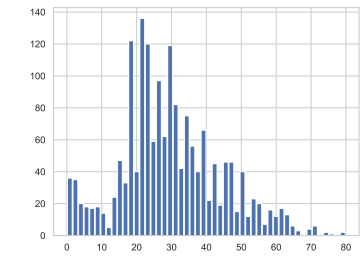
Real number (ℝ)

MISSING

Distinct	98
Distinct (%)	5.57%
Missing	440
Missing (%)	20%
Infinite	0
Infinite (%)	0%
Mean	29.81

Minimum	0.17
Maximum	80
Zeros	0
Zeros (%)	0%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

24.0	77
22.0	70
18.0	65

Quantile statistics

Minimum	0.17
5-th percentile	4
Q1	21
Median	28
Q3	39

Descriptive statistics

Standard deviation	14.45559700527994
Coefficient of variation (CV)	0.4849684208124134
Kurtosis	0.15626392898128927
Mean	29.807295454545457
Median Absolute Deviation (MAD)	8

95-th percentile	57
Maximum	80
Range	79.83
Interquartile range (IQR)	18

Median Absolute Deviation (MAD)	0
Skewness	0.39954706285913594
Sum	52460.84
Variance	208.96428477905835
Monotonicity	0

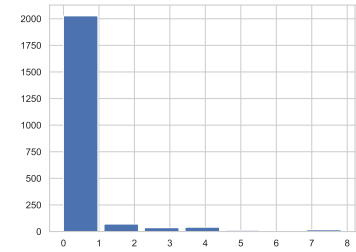
sibsp

Real number (ℝ)
ZEROS

Distinct	7
Distinct (%)	0.32%
Missing	0
Missing (%)	0%
Infinite	0
Infinite (%)	0%
Mean	0.51

Minimum	0
Maximum	8
Zeros	1499
Zeros (%)	68.14%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

0	1499
1	528
2	70

Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
Median	0
Q3	1
95-th percentile	2
Maximum	8
Range	8
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	1.0666373722154538
Coefficient of variation (CV)	2.0970529212457536
Kurtosis	19.075694848822668
Mean	0.5086363636363637
Median Absolute Deviation (MAD)	0
Skewness	3.780740045071688
Sum	1119
Variance	1.1377152838066888
Monotonicity	0

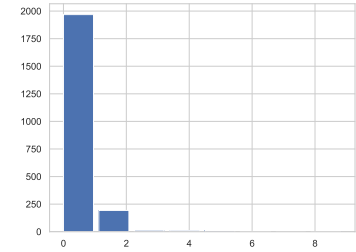
parch

Real number (ℝ)
ZEROS

Distinct	8
Distinct (%)	0.36%
Missing	0
Missing (%)	0%
Infinite	0
Infinite (%)	0%

Minimum	0
Maximum	9
Zeros	1680
Zeros (%)	76.36%
Negative	0
Negative (%)	0%

Histogram



Frequent Values

	Mean	0.38	Memory size	17.31 KB	0	1680
					1	288
					2	193

Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
Median	0
Q3	0
95-th percentile	2
Maximum	9
Range	9
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.841783687628853
Coefficient of variation (CV)	2.1942228824448775
Kurtosis	17.62833624357166
Mean	0.3836363636363636
Median Absolute Deviation (MAD)	0
Skewness	3.347178082271727
Sum	844
Variance	0.7085997767580304
Monotonicity	0

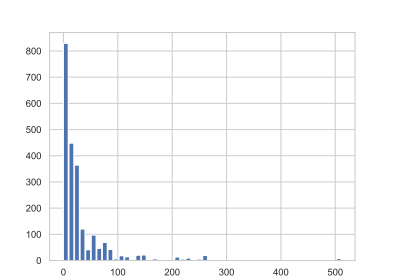
fare

- Real number (ℝ)
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- ZEROS

Distinct	281
Distinct (%)	12.78%
Missing	1
Missing (%)	0.05%
Infinite	0
Infinite (%)	0%
Mean	32.85

Minimum	0
Maximum	512.3292
Zeros	32
Zeros (%)	1.45%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

8.05	103
13.0	101
7.75	89

Quantile statistics

Minimum	0
5-th percentile	7.23
Q1	7.9
Median	14.45
Q3	31.28
95-th percentile	120
Maximum	512.3292
Range	512.33

Descriptive statistics

Standard deviation	50.923367956762846
Coefficient of variation (CV)	1.5500222858258184
Kurtosis	29.339907776025157
Mean	32.85331341518872
Median Absolute Deviation (MAD)	6.9042
Skewness	4.525607283282722
Sum	72244.44
Variance	2593.189404059861

Interquartile range (IQR)	23.38	Monotonicity	0
---------------------------	-------	--------------	---

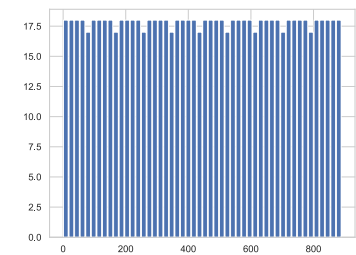
PassengerId

- Real number (\mathbb{R})
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- HIGH CORRELATION
- MISSING
- UNIFORM

Distinct	891
Distinct (%)	100%
Missing	1309
Missing (%)	59.5%
Infinite	0
Infinite (%)	0%
Mean	446

Minimum	1
Maximum	891
Zeros	0
Zeros (%)	0%
Negative	0
Negative (%)	0%
Memory size	17.31 KB

Histogram



Frequent Values

119.0	1
544.0	1
711.0	1

Quantile statistics

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

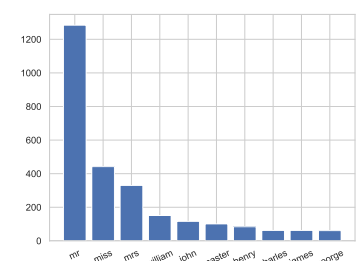
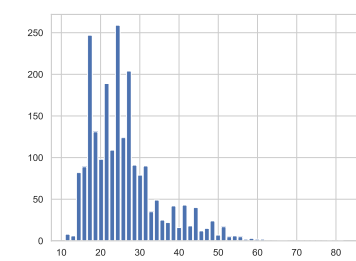
Descriptive statistics

Standard deviation	257.3538420152301
Coefficient of variation (CV)	0.5770265516036549
Kurtosis	-1.1999999999999997
Mean	446
Median Absolute Deviation (MAD)	223
Skewness	0
Sum	397386
Variance	66231
Monotonicity	2

name

- Categorical
- HIGH CARDINALITY
- UNIFORM

Distinct	2198
Distinct (%)	99.91%
Missing	0
Missing (%)	0%
Memory size	17.31 KB



Length

Max length	82
Median length	25
Mean length	26.2

Unicode

Total characters	0
Distinct characters	0
Distinct categories	0

Unique

Unique	2196
Unique (%)	99.82%

Sample

1st Row	Braund, Mr. Owen Harris
2nd Row	Cumings, Mrs. John Bradley (Florence Briggs Thayer)

Min length	11	Distinct scripts	0
		Distinct blocks	0
3rd Row	Heikkinen, Miss. Laina		
4th Row	Futrelle, Mrs. Jacques Heath (Lily May Peel)		
5th Row	Allen, Mr. William Henry		

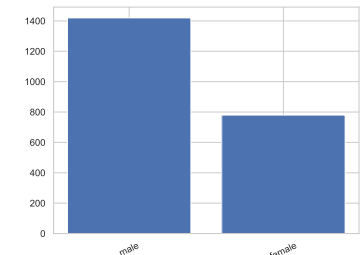
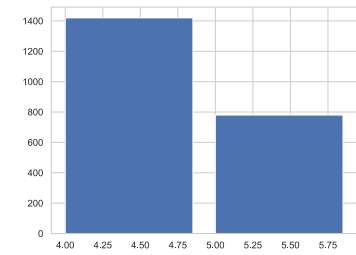
sex

Categorical

HIGH CORRELATION

HIGH CORRELATION

Distinct	2
Distinct (%)	0.09%
Missing	0
Missing (%)	0%
Memory size	17.31 KB



Length

Max length	6
Median length	4
Mean length	4.71
Min length	4

Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

Unique

Unique	0
Unique (%)	0%

Sample

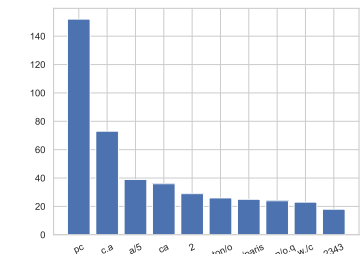
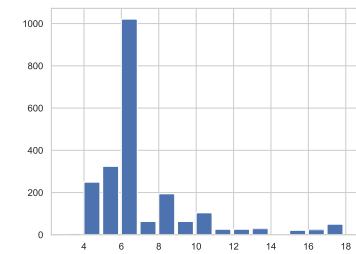
1st Row	male
2nd Row	female
3rd Row	female
4th Row	female
5th Row	male

ticket

Categorical

HIGH CARDINALITY

Distinct	929
Distinct (%)	42.23%
Missing	0
Missing (%)	0%
Memory size	17.31 KB



Length

Max length	18
Median length	6
Mean length	6.77
Min length	3

Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

Unique

Unique	232
Unique (%)	10.55%

Sample

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

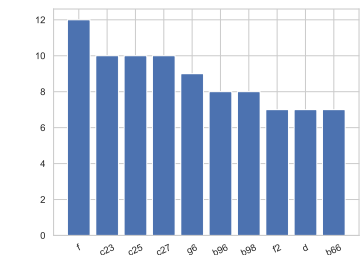
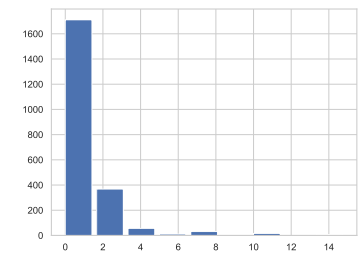
Sum Row		Sum Row	

cabin

Categorical

HIGH CARDINALITY

Distinct	187
Distinct (%)	8.5%
Missing	0
Missing (%)	0%
Memory size	17.31 KB



Length

Max length	15
Median length	0
Mean length	0.83
Min length	0

Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

Unique

Unique	32
Unique (%)	1.45%

Sample

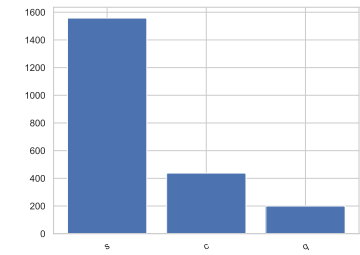
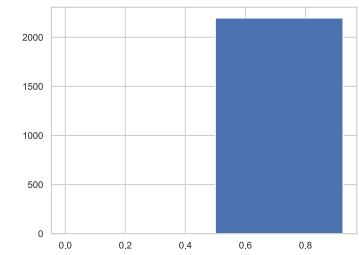
1st Row	
2nd Row	C85
3rd Row	
4th Row	C123
5th Row	

embarked

Categorical

HIGH CORRELATION

Distinct	4
Distinct (%)	0.18%
Missing	0
Missing (%)	0%
Memory size	17.31 KB



Length

Max length	1
Median length	1
Mean length	1
Min length	0

Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

Unique

Unique	0
Unique (%)	0%

Sample

1st Row	S
2nd Row	C
3rd Row	S
4th Row	S
5th Row	S

boat

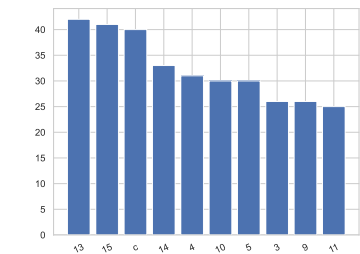
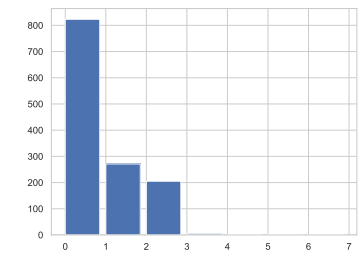
Categorical

HIGH CORRELATION

HIGH CORRELATION

MISSING

Distinct	28
Distinct (%)	2.14%
Missing	891



Missing (%)	40.5%
-------------	-------

Memory size	17.31 KB
-------------	----------

Length

Max length	7
Median length	0
Mean length	0.55

Min length	0
------------	---

Unicode

Total characters	0
Distinct characters	0

Distinct categories	0
---------------------	---

Distinct scripts	0
------------------	---

Distinct blocks	0
-----------------	---

Unique

Unique	4
--------	---

Unique (%)	0.31%
------------	-------

Sample

1st Row

2nd Row

--

3rd Row

4th Row

5th Row

body

Categorical

HIGH CARDINALITY

MISSING

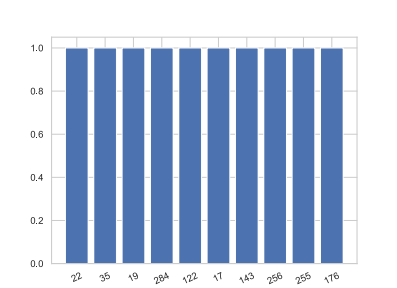
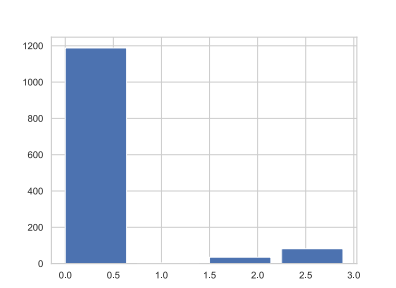
Distinct	122
----------	-----

Distinct (%)	9.32%
--------------	-------

Missing	891
---------	-----

Missing (%)	40.5%
-------------	-------

Memory size	17.31 KB
-------------	----------



Length

Max length	3
Median length	0
Mean length	0.24

Min length	0
------------	---

Unicode

Total characters	0
Distinct characters	0

Distinct categories	0
---------------------	---

Distinct scripts	0
------------------	---

Distinct blocks	0
-----------------	---

Unique

Unique	121
--------	-----

Unique (%)	9.24%
------------	-------

Sample

1st Row

2nd Row

3rd Row

4th Row

5th Row

home.dest

Categorical

HIGH CARDINALITY

MISSING

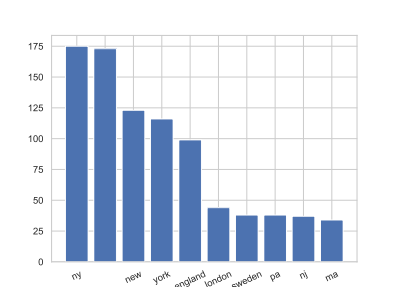
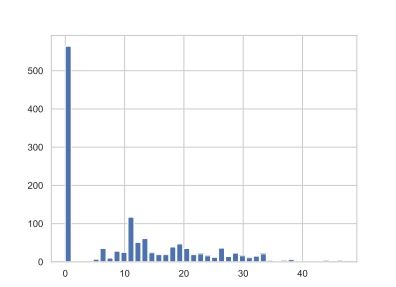
Distinct	370
----------	-----

Distinct (%)	28.27%
--------------	--------

Missing	891
---------	-----

Missing (%)	40.5%
-------------	-------

Memory size	17.31 KB
-------------	----------



Length

Max length	47
------------	----

Median length	10
---------------	----

Unicode

Total characters	0
------------------	---

Distinct characters	0
---------------------	---

Unique

Unique	217
--------	-----

Unique (%)	16.58%
------------	--------

Sample

1st Row

2nd Row

Mean length	10.25
-------------	-------

Min length	0
------------	---

Distinct categories	0
---------------------	---

Distinct scripts	0
------------------	---

Distinct blocks	0
-----------------	---

3rd Row

4th Row

5th Row

home_dest

- Unsupported
- MISSING
- REJECTED
- UNSUPPORTED

Distinct	0
----------	---

Distinct (%)	0%
--------------	----

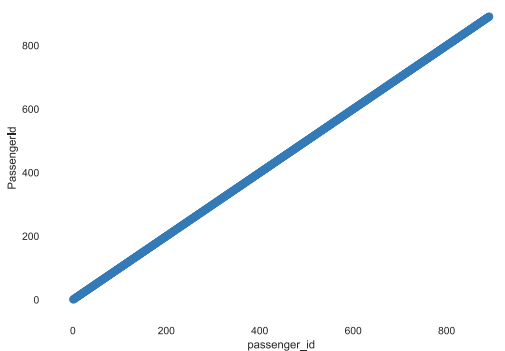
Missing	2200
---------	------

Missing (%)	100%
-------------	------

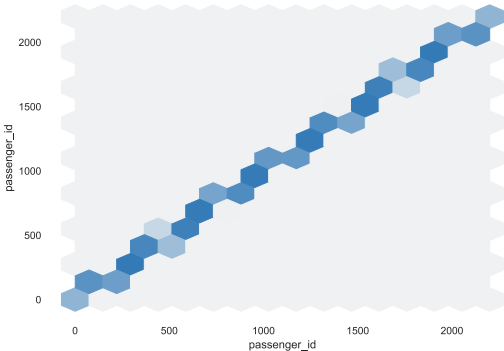
Memory size	17.31 KB
-------------	----------

Common Values

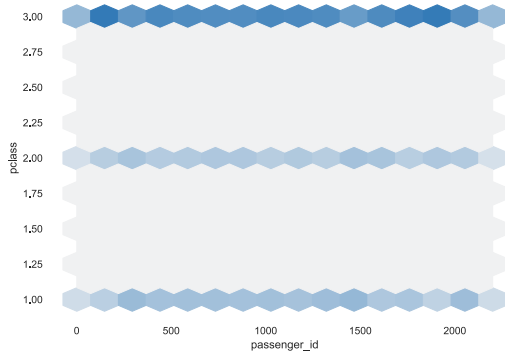
Interactions



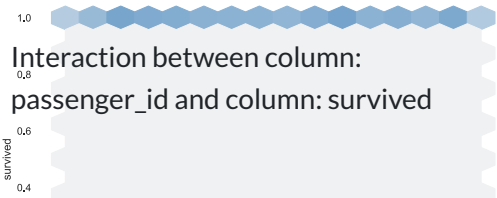
Interaction between column: passenger_id and column: PassengerId



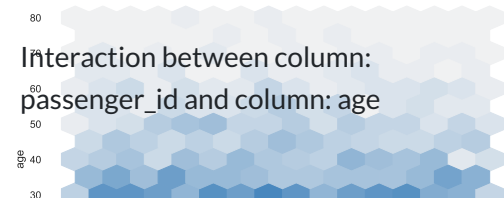
Interaction between column: passenger_id and column: passenger_id



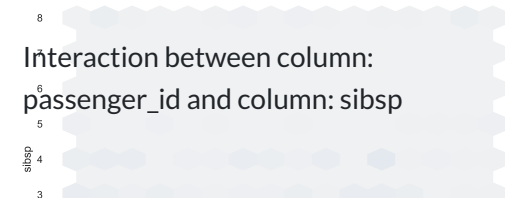
Interaction between column: passenger_id and column: pclass



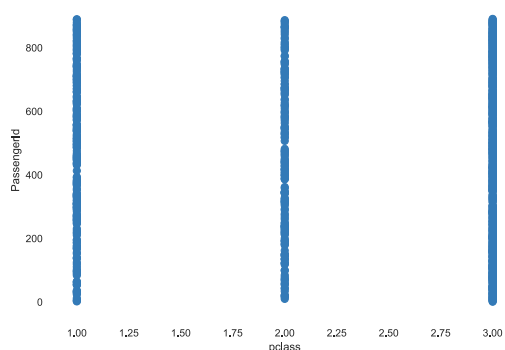
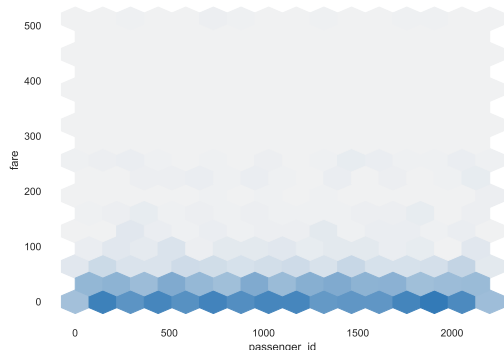
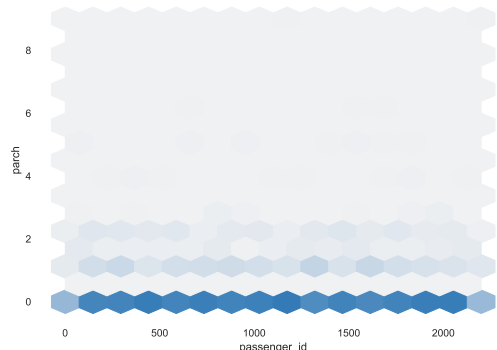
Interaction between column: passenger_id and column: survived



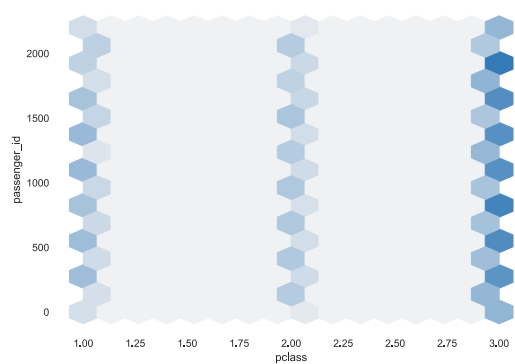
Interaction between column: passenger_id and column: age



Interaction between column: passenger_id and column: sibsp

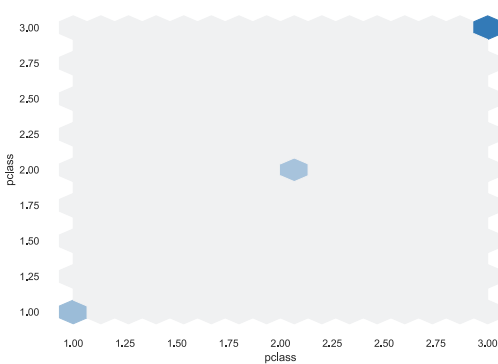


Interaction between column:
passenger_id and column: parch



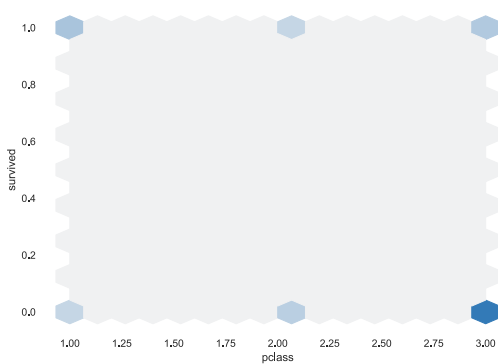
Interaction between column: pclass and
column: passenger_id

Interaction between column:
passenger_id and column: fare

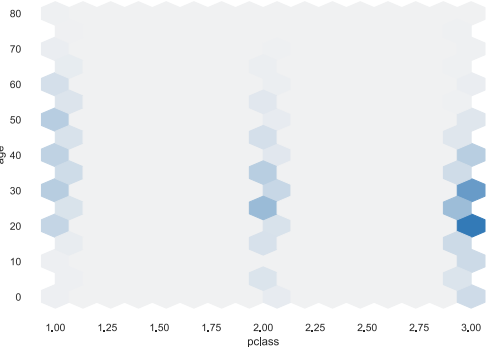


Interaction between column: pclass and
column: pclass

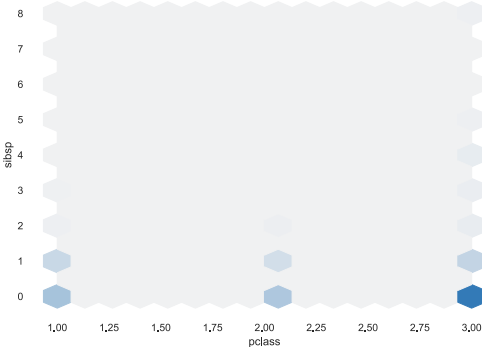
Interaction between column: pclass and
column: PassengerId



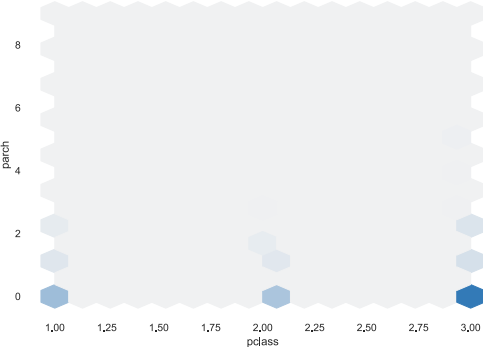
Interaction between column: pclass and
column: survived



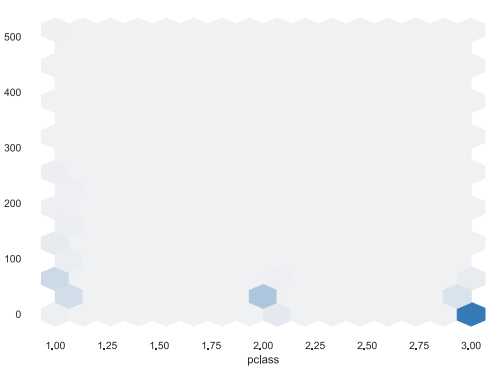
Interaction between column: pclass and
column: age



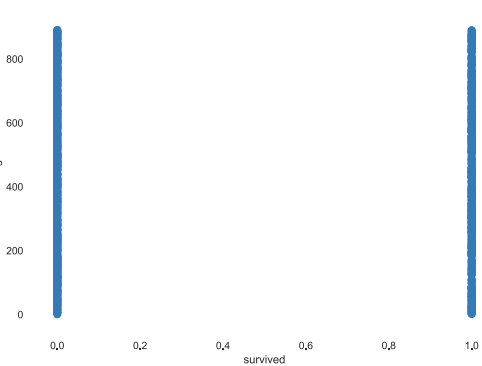
Interaction between column: pclass and
column: sibsp



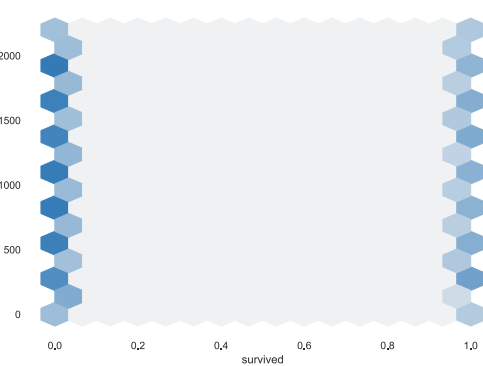
Interaction between column: pclass and
column: parch



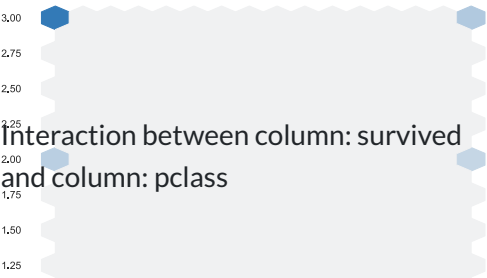
Interaction between column: pclass and
column: fare



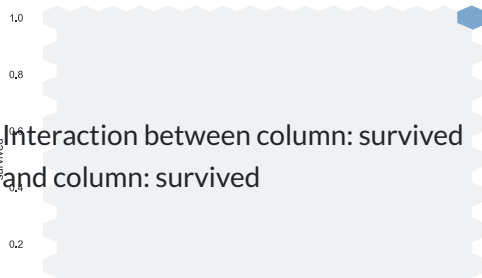
Interaction between column: survived
and column: PassengerId



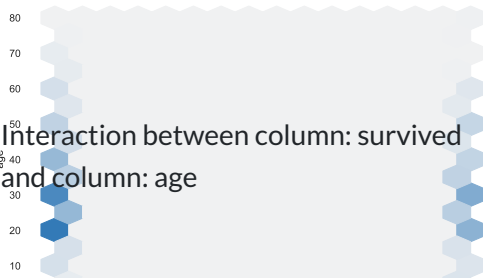
Interaction between column: survived
and column: passenger_id



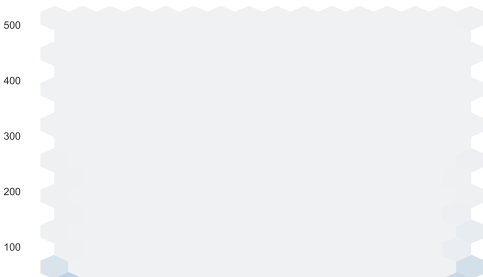
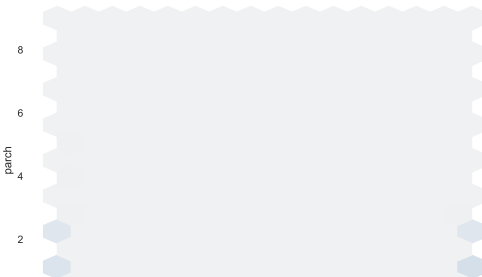
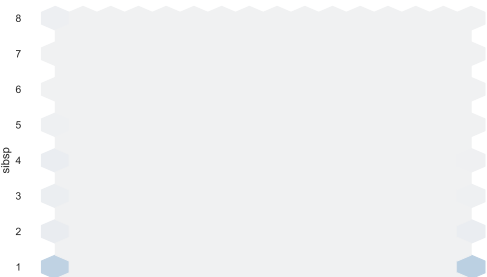
Interaction between column: survived
and column: pclass



Interaction between column: survived
and column: survived



Interaction between column: survived
and column: age





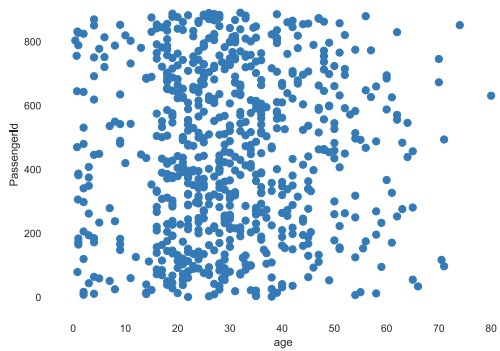
Interaction between column: survived
and column: sibsp



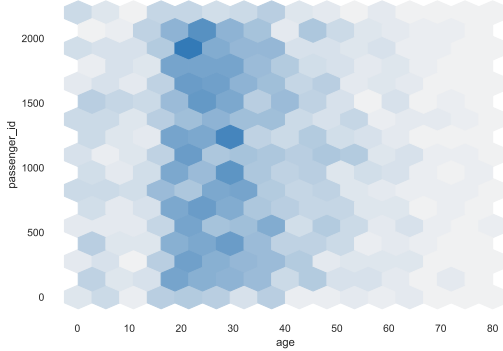
Interaction between column: survived
and column: parch



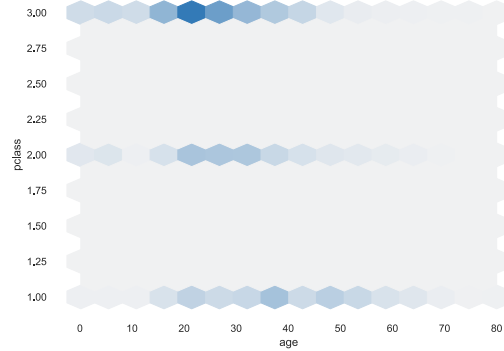
Interaction between column: survived
and column: fare



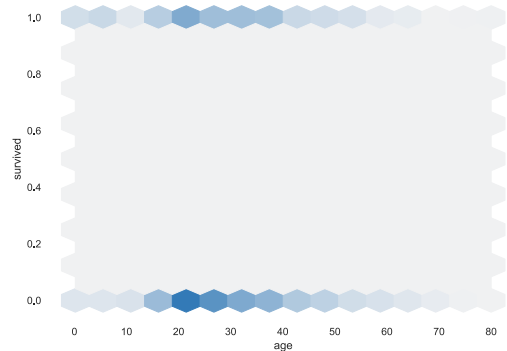
Interaction between column: age and
column: PassengerId



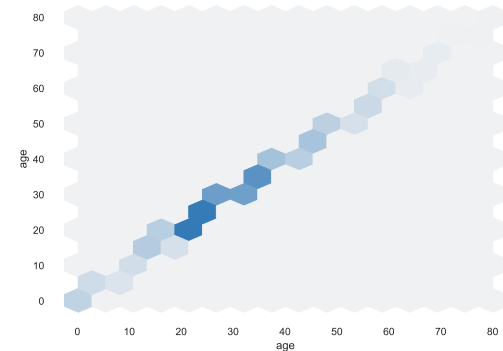
Interaction between column: age and
column: passenger_id



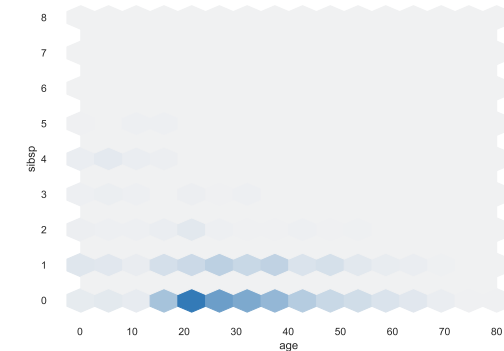
Interaction between column: age and
column: pclass



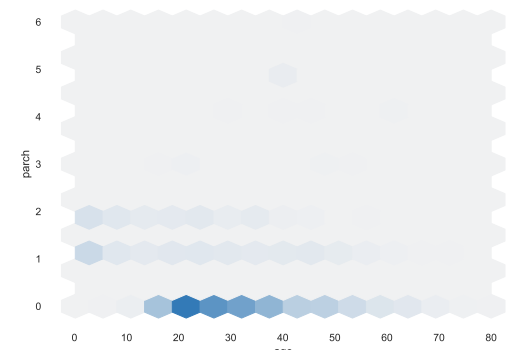
Interaction between column: age and
column: survived



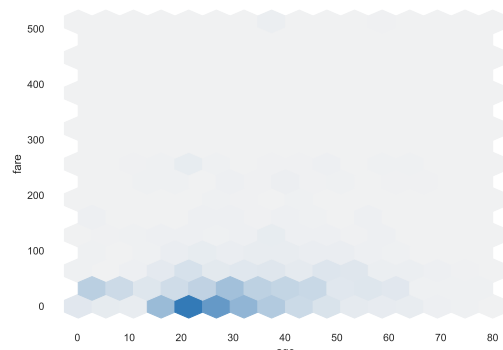
Interaction between column: age and
column: age



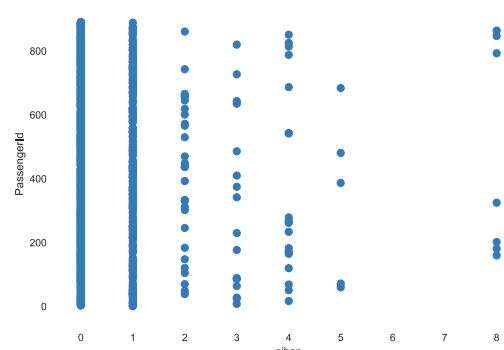
Interaction between column: age and
column: sibsp



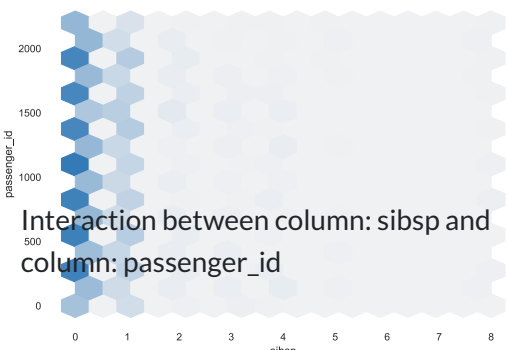
Interaction between column: age and
column: parch



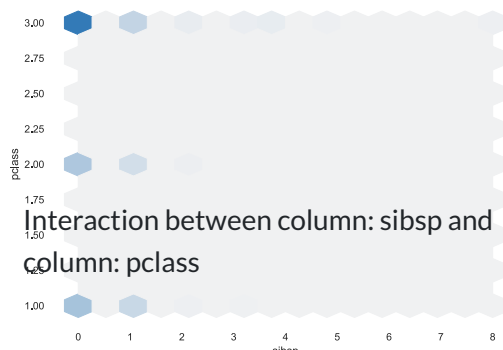
Interaction between column: age and
column: fare



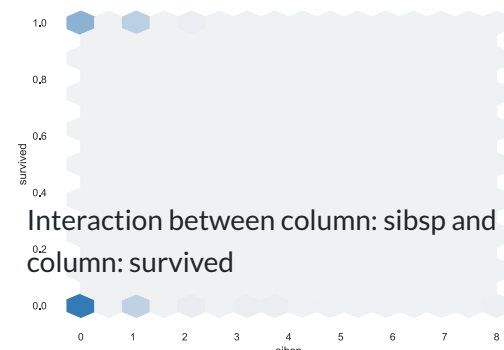
Interaction between column: sibsp and
column: PassengerId



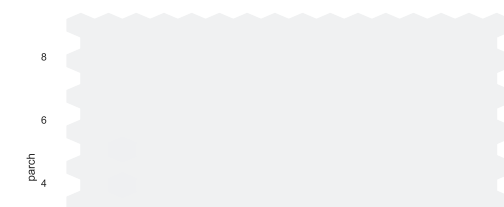
Interaction between column: sibsp and
column: passenger_id

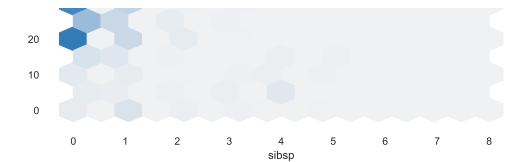


Interaction between column: sibsp and
column: pclass

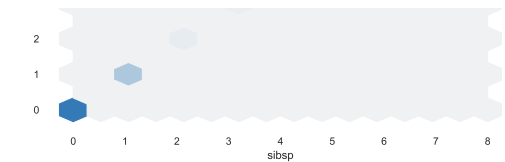


Interaction between column: sibsp and
column: survived

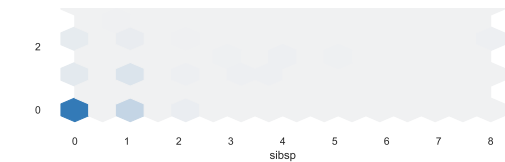




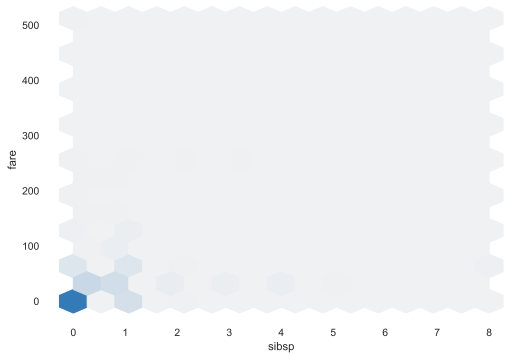
Interaction between column: sibsp and column: age



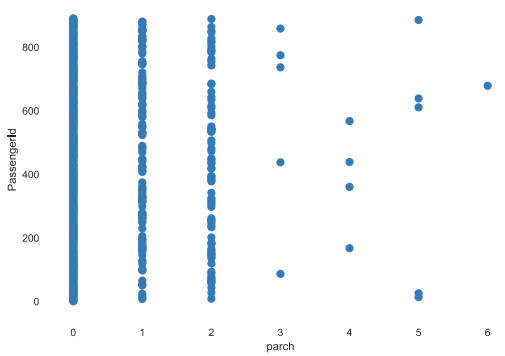
Interaction between column: sibsp and column: sibsp



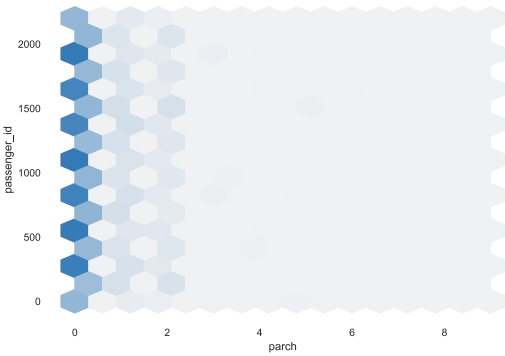
Interaction between column: sibsp and column: parch



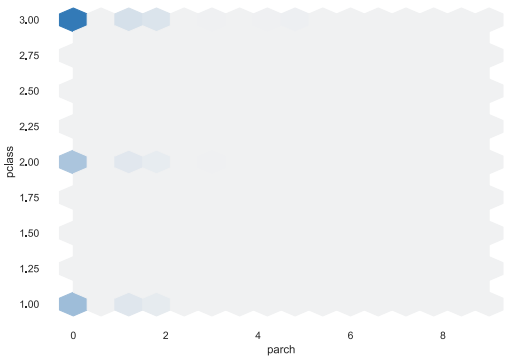
Interaction between column: sibsp and column: fare



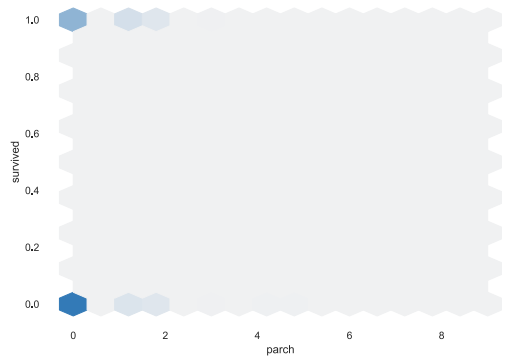
Interaction between column: parch and column: PassengerId



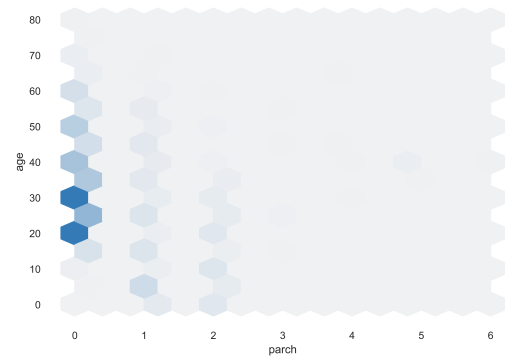
Interaction between column: parch and column: passenger_id



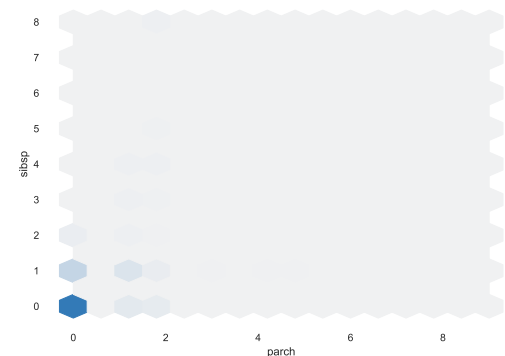
Interaction between column: parch and column: pclass



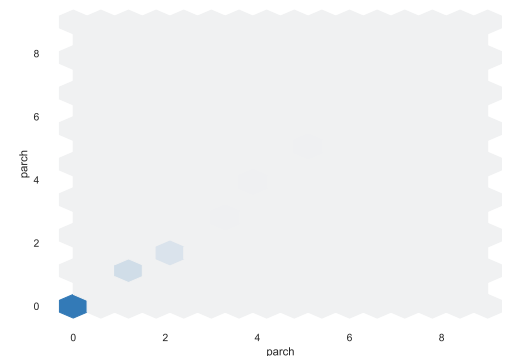
Interaction between column: parch and column: survived



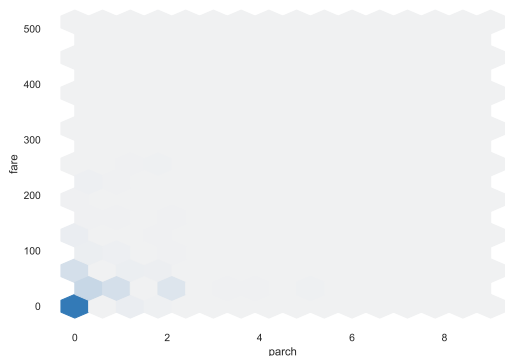
Interaction between column: parch and column: age



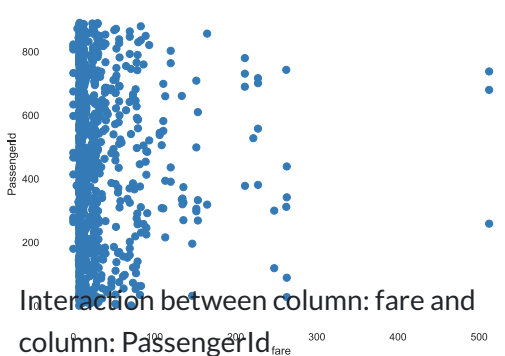
Interaction between column: parch and column: sibsp



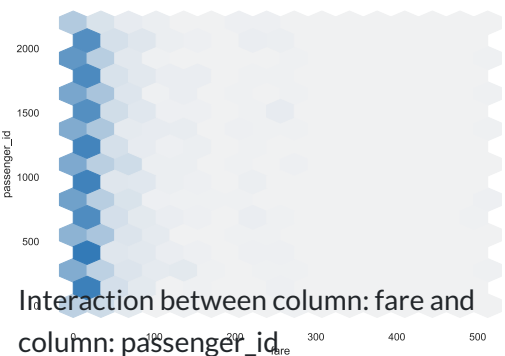
Interaction between column: parch and column: parch



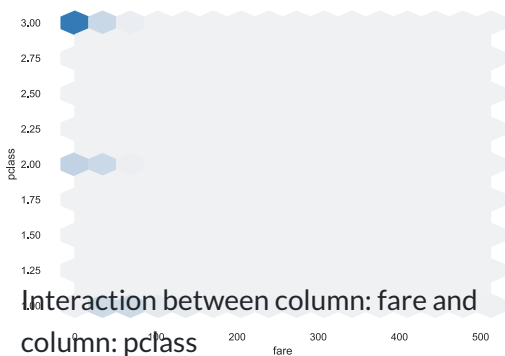
Interaction between column: parch and column: fare



Interaction between column: fare and column: PassengerId



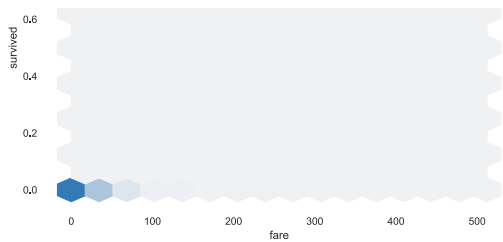
Interaction between column: fare and column: passenger_id



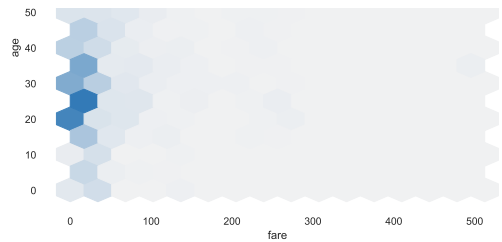
Interaction between column: fare and column: pclass



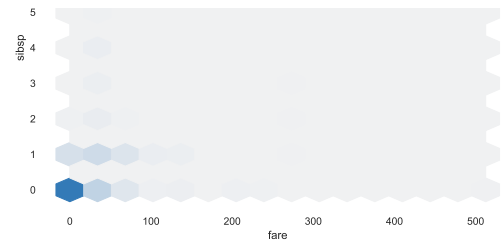
Interaction between column: fare and column: parch



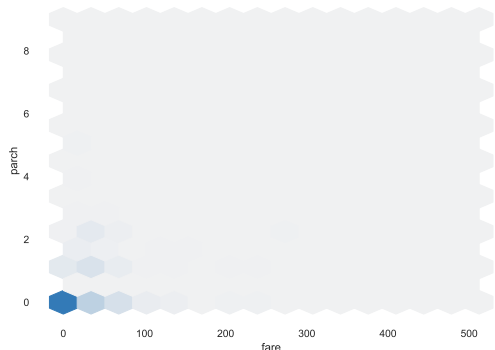
Interaction between column: fare and column: survived



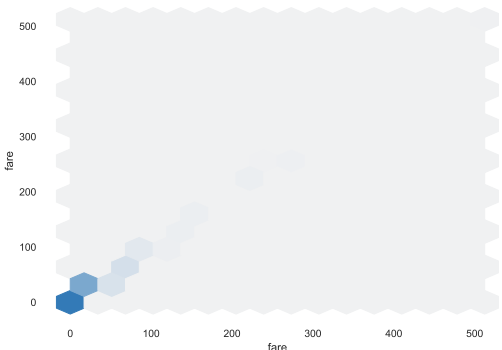
Interaction between column: fare and column: age



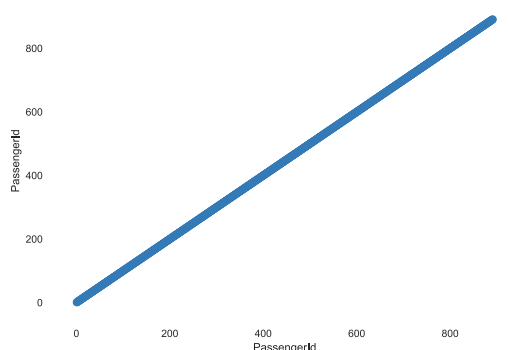
Interaction between column: fare and column: sibsp



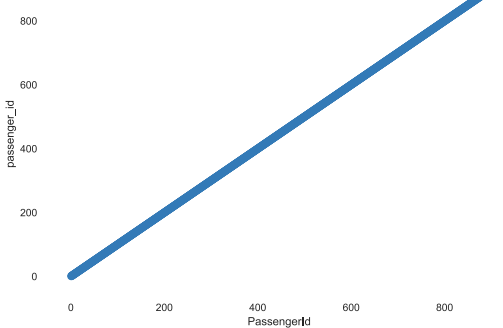
Interaction between column: fare and column: parch



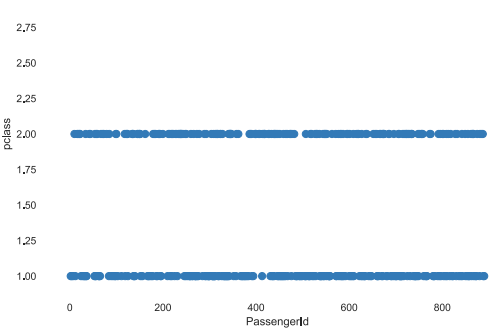
Interaction between column: fare and column: fare



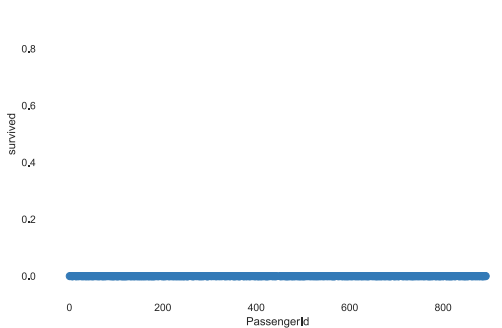
Interaction between column: PassengerId and column: PassengerId



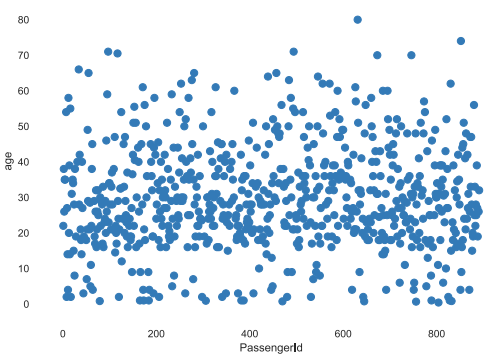
Interaction between column: PassengerId and column: passenger_id



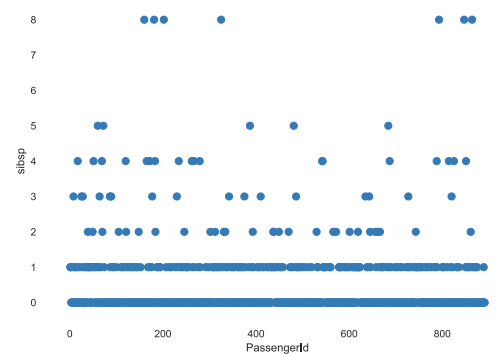
Interaction between column: PassengerId and column: pclass



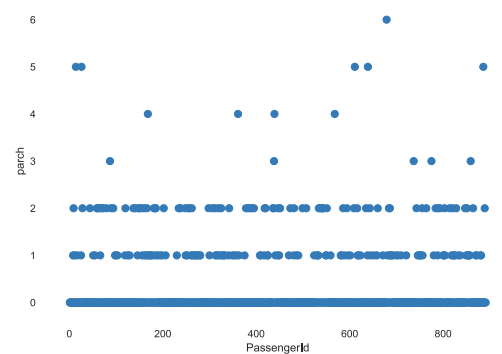
Interaction between column: PassengerId and column: survived



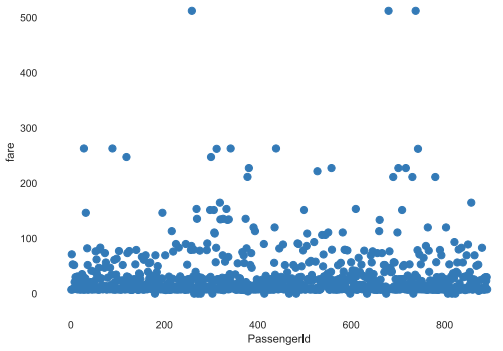
Interaction between column: PassengerId and column: age



Interaction between column: PassengerId and column: sibsp

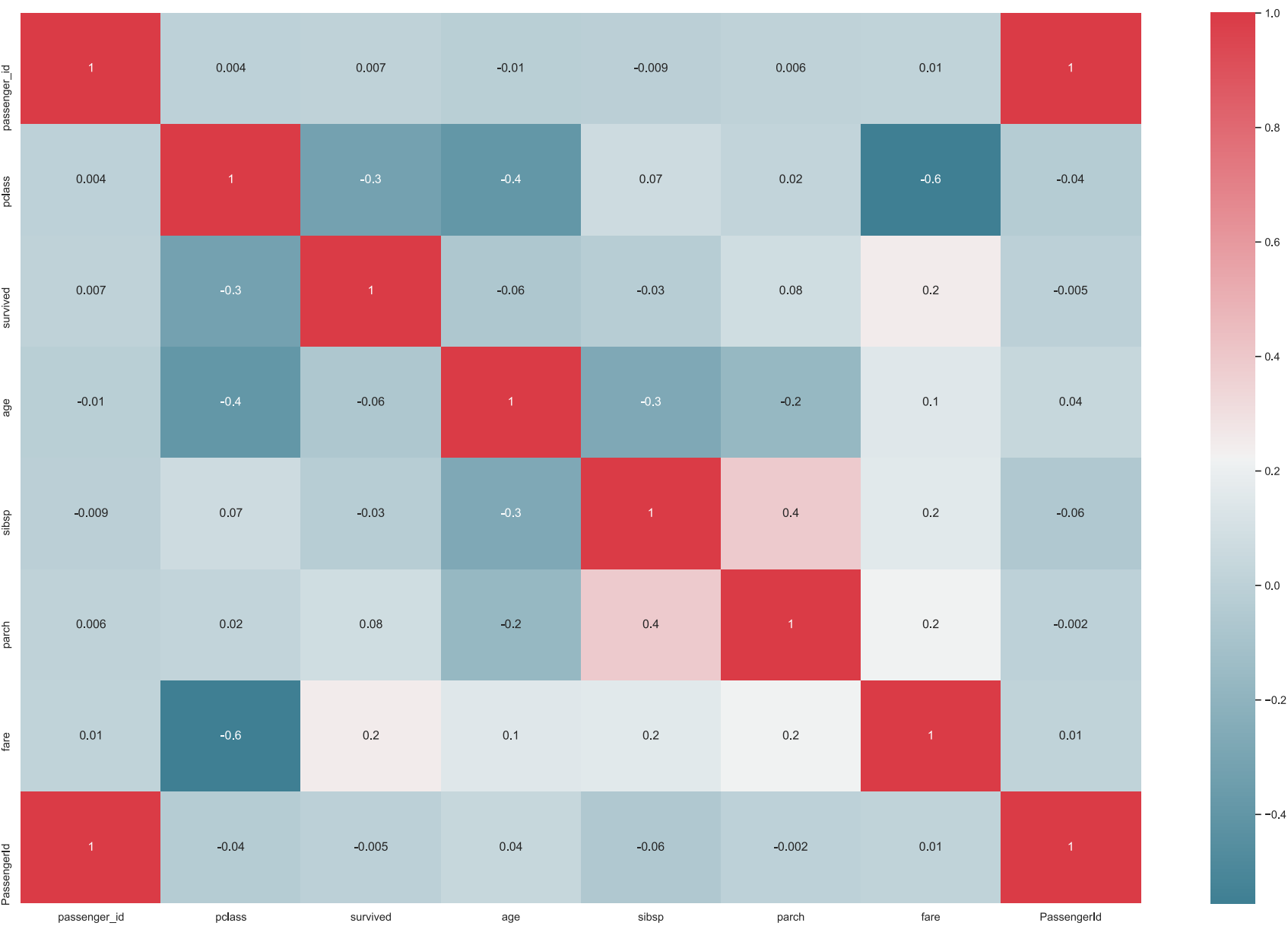


Interaction between column: PassengerId and column: parch



Interaction between column: PassengerId and column: fare

Correlations

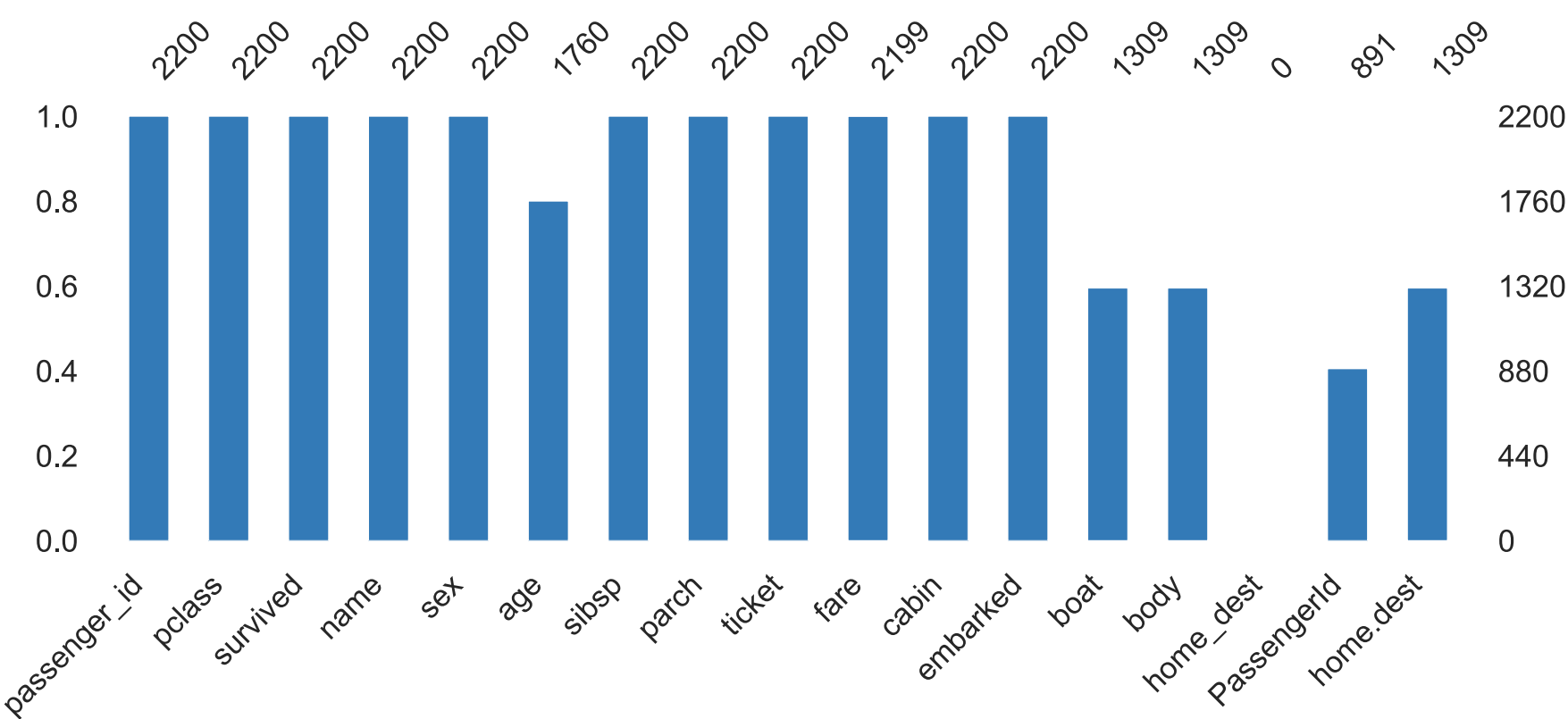


Missing values

Count

Matrix

Dendrogram



A simple visualization of nullity by column.

Sample

First rows

passenger_id	pclass	survived	name	sex	age	sibsp	parch	ticket
1	3	0	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599
3	3	1	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
5	3	0	Allen, Mr. William Henry	male	35	0	0	373450
6	3	0	Moran, Mr. James	male	NaN	0	0	330877
7	1	0	McCarthy, Mr. Timothy J	male	54	0	0	17463
8	3	0	Palsson, Master. Gosta Leonard	male	2	3	1	349909
9	3	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742
10	2	1	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736

Last rows

passenger_id	pclass	survived	name	sex	age	sibsp	parch	ticket	fare
2191	1	1	Chevre Mr. Paul Romaine	male	45	0	0	PC 17594	29.7
2192	1	1	Endres Miss. Caroline Louise	female	38	0	0	PC 17757	227.5
2193	3	0	Attalah Miss. Malake	female	17	0	0	2627	14.45
2194	2	0	Leyson Mr. Robert William Norman	male	24	0	0	C.A. 29566	10.5
2195	3	0	Pettersson Miss. Ellen Natalia	female	18	0	0	347087	7.775
2196	3	0	Rice Master. Eugene	male	2	4	1	382652	29.12
2197	3	0	Dika Mr. Mirko	male	17	0	0	349232	7.895
2198	3	0	Palsson Miss. Torborg Danira	female	8	3	1	349909	21.07
2199	1	0	Widener Mr. George Dunton	male	50	1	1	113503	211.5
2200	3	1	Riordan Miss. Johanna Hannah	female	NaN	0	0	334915	7.720

Warnings(43)

- HIGH CARDINALITY alert on column name
- HIGH CARDINALITY alert on column ticket
- HIGH CARDINALITY alert on column cabin
- HIGH CARDINALITY alert on column body
- HIGH CARDINALITY alert on column home.dest
- passenger_id has HIGH CORRELATION with 1 field: PassengerId
- pclass has HIGH CORRELATION with 1 field: fare
- fare has HIGH CORRELATION with 1 field: pclass
- PassengerId has HIGH CORRELATION with 1 field: passenger_id
- passenger_id has HIGH CORRELATION with 1 field: PassengerId
- pclass has HIGH CORRELATION with 1 field: fare
- fare has HIGH CORRELATION with 1 field: pclass
- PassengerId has HIGH CORRELATION with 1 field: passenger_id
- passenger_id has HIGH CORRELATION with 1 field: PassengerId
- pclass has HIGH CORRELATION with 1 field: fare
- fare has HIGH CORRELATION with 1 field: pclass
- PassengerId has HIGH CORRELATION with 1 field: passenger_id
- sex has HIGH CORRELATION with 1 field: boat
- boat has HIGH CORRELATION with 1 field: sex
- passenger_id has HIGH CORRELATION with 1 field: PassengerId
- pclass has HIGH CORRELATION with 2 fields: fare boat
- survived has HIGH CORRELATION with 2 fields: sex boat
- sex has HIGH CORRELATION with 2 fields: survived boat
- fare has HIGH CORRELATION with 2 fields: pclass boat
- embarked has HIGH CORRELATION with 1 field: boat
- boat has HIGH CORRELATION with 5 fields: pclass survived sex fare embarked
- PassengerId has HIGH CORRELATION with 1 field: passenger_id
- MISSING alert on column age
- MISSING alert on column boat
- MISSING alert on column body
- MISSING alert on column home_dest
- MISSING alert on column PassengerId
- MISSING alert on column home.dest
- REJECTED alert on column home_dest
- UNIFORM alert on column passenger_id
- UNIFORM alert on column name
- UNIFORM alert on column PassengerId
- UNIQUE alert on column passenger_id
- UNSUPPORTED alert on column home_dest
- ZEROS alert on column survived
- ZEROS alert on column sibsp
- ZEROS alert on column parch
- ZEROS alert on column fare