

TITANIC EXPLORATIVE DATA ANALYSIS



Overview

Brought to you by [YData](#)

Overview

Alerts 20

Reproduction

Dataset statistics

Number of variables	15
Number of observations	891
Missing cells	869
Missing cells (%)	6.5%
Duplicate rows	1
Duplicate rows (%)	0.1%
Total size in memory	313.7 KiB
Average record size in memory	360.5 B

Variable types

Categorical	8
Numeric	4
Boolean	3

Variables

Select Columns

survived

Categorical

High correlation

Distinct	2	0	549
		1	342

Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	50.6 KiB

More details

pclass

Categorical

High correlation

Distinct	3
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	50.6 KiB

3	491
1	216
2	184

More details

sex

Categorical

High correlation

Distinct	2
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	53.8 KiB

male	577
female	314

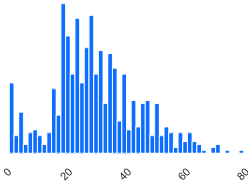
More details

age

Real number (\mathbb{R})

High correlation Missing

Distinct	88	Minimum	0.42
Distinct (%)	12.3%	Maximum	80
Missing	177	Zeros	0
Missing (%)	19.9%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	29.699118	Memory size	7.1 KiB



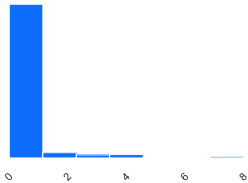
More details

sibsp

Real number (\mathbb{R})

High correlation Zeros

Distinct	7	Minimum	0
Distinct (%)	0.8%	Maximum	8
Missing	0	Zeros	608
Missing (%)	0.0%	Zeros (%)	68.2%
Infinite	0	Negative	0



Infinite (%)0.0%

Mean0.52300786

Negative (%)0.0%

Memory size7.1 KiB

More details

parch

Real number (\mathbb{R})

High correlationZeros

Distinct7

Distinct (%)0.8%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean0.38159371

Minimum0

Maximum6

Zeros678

Zeros (%)76.1%

Negative0

Negative (%)0.0%

Memory size7.1 KiB

A histogram showing the distribution of the 'parch' variable. The x-axis represents the number of parents/children aboard, with labels at 0, 1, 2, 3, and 4. The y-axis represents frequency. The bar for 0 is the tallest, reaching a frequency of approximately 678. The bar for 1 is significantly shorter, and bars for 2, 3, and 4 are very short, indicating a highly right-skewed distribution.

More details

fare

Real number (\mathbb{R})

Zeros

file:///C:/Users/Francois/Documents/Simplon/P1 - POEC Data Analyst/P1 - POEC Data Analyst/3. Ressources_cours/Sem 1/Titanic - GitHub.html

4/12

Distinct248

Distinct (%)27.8%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean32.204208

Minimum0

Maximum512.32

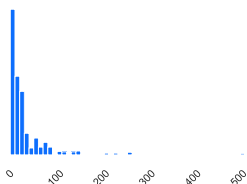
Zeros15

Zeros (%)1.7%

Negative0

Negative (%)0.0%

Memory size7.1 KiB



More details

embarked

Categorical

High correlation

Distinct3

Distinct (%)0.3%

Missing2

Missing (%)0.2%

Memory size50.6 KiB

S644

C168

Q77

More details

class

Categorical

High correlation

Distinct	3	Third	491
Distinct (%)	0.3%	First	216
Missing	0	Second	184
Missing (%)	0.0%		
Memory size	1.3 KiB		

More details

who

Categorical

High correlation

Distinct	3	man	537
Distinct (%)	0.3%	woman	271
Missing	0	child	83
Missing (%)	0.0%		
Memory size	53.0 KiB		

More details

adult_male

Boolean

High correlation

Distinct	2	True	537
Distinct (%)	0.2%	False	354
Missing	0		
Missing (%)	0.0%		
Memory size	1019.0 B		

More details

deck

Categorical

High correlation Missing

Distinct	7	C	59
Distinct (%)	3.4%	B	47
Missing	688	D	33
Missing (%)	77.2%	E	32
Memory size	1.7 KiB	A	15
		Other va...	17

More details

embark_town

Categorical

High correlation

Distinct	3	Southam...	644
Distinct (%)	0.3%	Cherbou...	168
Missing	2	Queenst...	77
Missing (%)	0.2%		
Memory size	58.8 KiB		

More details

alive

Boolean

High correlation

Distinct	2	False	549
Distinct (%)	0.2%	True	342
Missing	0		
Missing (%)	0.0%		
Memory size	1019.0 B		
<div>More details</div>			

alone

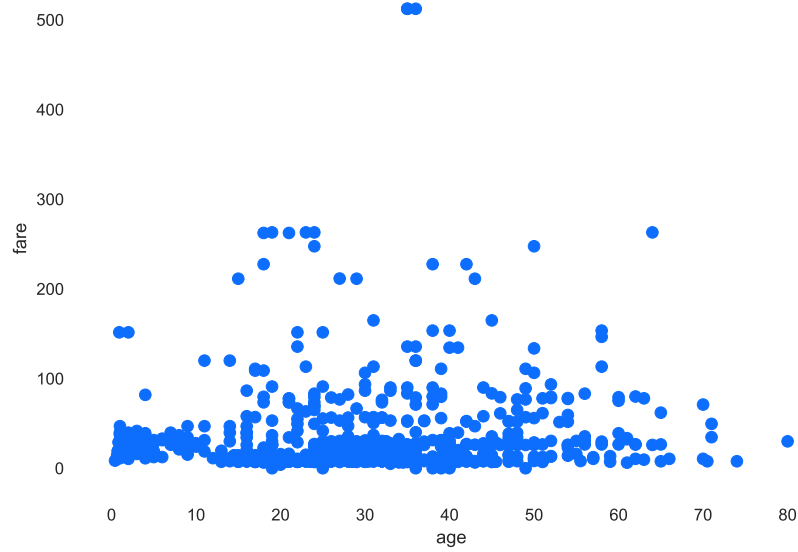
Boolean

High correlation

Distinct	2	True	537
Distinct (%)	0.2%	False	354
Missing	0		
Missing (%)	0.0%		
Memory size	1019.0 B		
<div>More details</div>			

Interactions

age	sibsp	parch	fare
fare	age	sibsp	parch

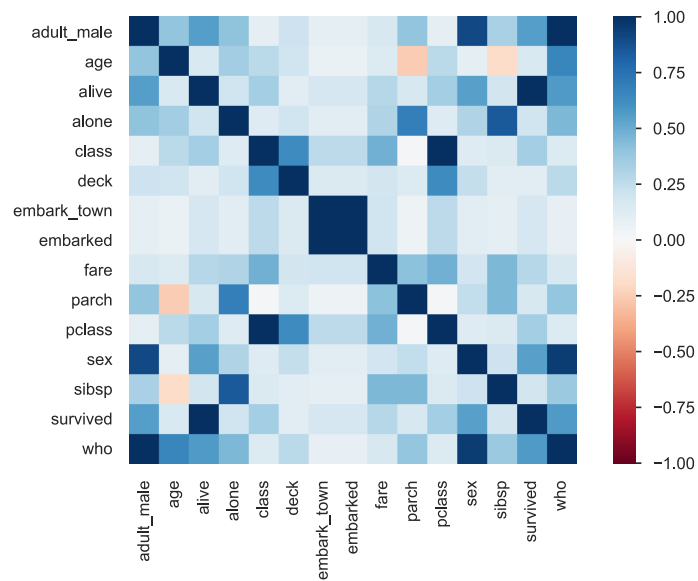


Correlations

Auto

Heatmap

[Table](#)

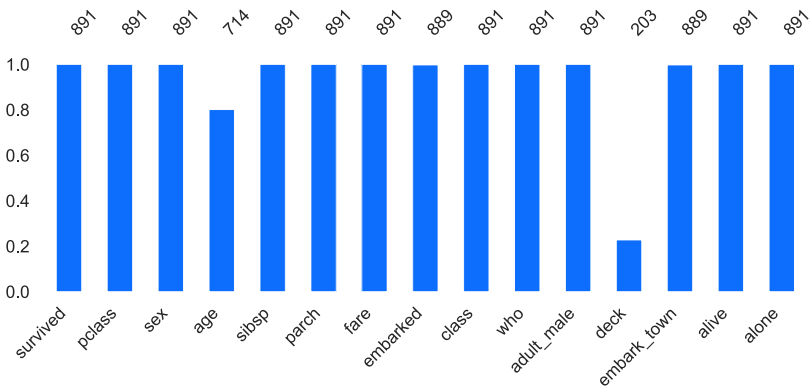


Missing values

Count

Matrix

Heatmap



A simple visualization of nullity by column.

Sample

First rows

Last rows

	survived	pclass	sex	age	sibsp	parch	fare	embarked
0	0	3	male	22.0	1	0	7.2500	S
1	1	1	female	38.0	1	0	71.2833	C
2	1	3	female	26.0	0	0	7.9250	S
3	1	1	female	35.0	1	0	53.1000	S
4	0	3	male	35.0	0	0	8.0500	S
5	0	3	male	NaN	0	0	8.4583	Q
6	0	1	male	54.0	0	0	51.8625	S
7	0	3	male	2.0	3	1	21.0750	S

	survived	pclass	sex	age	sibsp	parch	fare	emba
8	1	3	female	27.0	0	2	11.1333	S
9	1	2	female	14.0	1	0	30.0708	C

Duplicate rows

Most frequently occurring

	survived	pclass	sex	age	sibsp	parch	fare	emba
0	1	1	female	24.0	0	0	69.3	C

Report generated by [YData](#).