Titanic Dataset Report

YData and the open source community.

2025-07-30 08:21:23.968964

[Titanic Dataset Report](#top)

##### Titanic Dataset Report

* [Overview](#overview)
* [Variables](#variables-dropdown)
* [Interactions](#interactions)
* [Correlations](#correlations_tab)
* [Missing values](#missing)
* [Sample](#sample)
* [Duplicate rows](#duplicate)

# Overview

Brought to you by [YData](https://ydata.ai/?utm_source=opensource&utm_medium=ydataprofiling&utm_campaign=report)

* Overview
* Alerts 20
* Reproduction

Dataset statistics

|  |  |
| --- | --- |
| Number of variables | 15 |
| Number of observations | 891 |
| Missing cells | 869 |
| Missing cells (%) | 6.5% |
| Duplicate rows | 53 |
| Duplicate rows (%) | 5.9% |
| Total size in memory | 313.7 KiB |
| Average record size in memory | 360.5 B |

Variable types

|  |  |
| --- | --- |
| Categorical | 8 |
| Numeric | 4 |
| Boolean | 3 |

Alerts

Dataset has 53 (5.9%) [duplicate rows](#duplicate)

Duplicates

[adult\_male](#pp_var_646073766436733809) is highly overall correlated with alive and 3 other fields

High correlation

[age](#pp_var_2684574419960095649) is highly overall correlated with who

High correlation

[alive](#pp_var_5599174402278627879) is highly overall correlated with adult\_male and 3 other fields

High correlation

[alone](#pp_var_3893282888255956381) is highly overall correlated with parch and 1 other fields

High correlation

[class](#pp_var_6771237817665225922) is highly overall correlated with deck and 1 other fields

High correlation

[deck](#pp_var_3015878985090200450) is highly overall correlated with class and 1 other fields

High correlation

[embark\_town](#pp_var_-234494340988115811) is highly overall correlated with embarked

High correlation

[embarked](#pp_var_5003508584533474652) is highly overall correlated with embark\_town

High correlation

[parch](#pp_var_5885521624025493630) is highly overall correlated with alone

High correlation

[pclass](#pp_var_4031927281142044962) is highly overall correlated with class and 1 other fields

High correlation

[sex](#pp_var_-7373172888818346473) is highly overall correlated with adult\_male and 3 other fields

High correlation

[sibsp](#pp_var_7919890951915530931) is highly overall correlated with alone

High correlation

[survived](#pp_var_-844320206311774204) is highly overall correlated with adult\_male and 3 other fields

High correlation

[who](#pp_var_3267960624249646528) is highly overall correlated with adult\_male and 4 other fields

High correlation

[age](#pp_var_2684574419960095649) has 177 (19.9%) missing values

Missing

[deck](#pp_var_3015878985090200450) has 688 (77.2%) missing values

Missing

[sibsp](#pp_var_7919890951915530931) has 608 (68.2%) zeros

Zeros

[parch](#pp_var_5885521624025493630) has 678 (76.1%) zeros

Zeros

[fare](#pp_var_1765197558829918918) has 15 (1.7%) zeros

Zeros

Reproduction

|  |  |
| --- | --- |
| Analysis started | 2025-07-30 08:21:23.968964 |
| Analysis finished | 2025-07-30 08:21:26.909350 |
| Duration | 2.94 seconds |
| Software version | [ydata-profiling vv4.16.1](https://github.com/ydataai/ydata-profiling) |
| Download configuration | [config.json](data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Titanic%20Dataset%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22creator%22%3A%20%22%22%2C%20%22author%22%3A%20%22%22%2C%20%22copyright_holder%22%3A%20%22%22%2C%20%22copyright_year%22%3A%20%22%22%2C%20%22url%22%3A%20%22%22%7D%2C%20%22variables%22%3A%20%7B%22descriptions%22%3A%20%7B%7D%7D%2C%20%22infer_dtypes%22%3A%20true%2C%20%22show_variable_description%22%3A%20true%2C%20%22pool_size%22%3A%200%2C%20%22progress_bar%22%3A%20true%2C%20%22vars%22%3A%20%7B%22num%22%3A%20%7B%22quantiles%22%3A%20%5B0.05%2C%200.25%2C%200.5%2C%200.75%2C%200.95%5D%2C%20%22skewness_threshold%22%3A%2020%2C%20%22low_categorical_threshold%22%3A%205%2C%20%22chi_squared_threshold%22%3A%200.999%7D%2C%20%22text%22%3A%20%7B%22length%22%3A%20true%2C%20%22words%22%3A%20true%2C%20%22characters%22%3A%20true%2C%20%22redact%22%3A%20false%7D%2C%20%22cat%22%3A%20%7B%22length%22%3A%20true%2C%20%22characters%22%3A%20true%2C%20%22words%22%3A%20true%2C%20%22cardinality_threshold%22%3A%2050%2C%20%22percentage_cat_threshold%22%3A%200.5%2C%20%22imbalance_threshold%22%3A%200.5%2C%20%22n_obs%22%3A%205%2C%20%22chi_squared_threshold%22%3A%200.999%2C%20%22coerce_str_to_date%22%3A%20false%2C%20%22redact%22%3A%20false%2C%20%22histogram_largest%22%3A%2050%2C%20%22stop_words%22%3A%20%5B%5D%2C%20%22dirty_categories%22%3A%20false%2C%20%22dirty_categories_threshold%22%3A%200.85%7D%2C%20%22image%22%3A%20%7B%22active%22%3A%20true%2C%20%22exif%22%3A%20true%2C%20%22hash%22%3A%20true%7D%2C%20%22bool%22%3A%20%7B%22n_obs%22%3A%203%2C%20%22imbalance_threshold%22%3A%200.5%2C%20%22mappings%22%3A%20%7B%22t%22%3A%20true%2C%20%22f%22%3A%20false%2C%20%22yes%22%3A%20true%2C%20%22no%22%3A%20false%2C%20%22y%22%3A%20true%2C%20%22n%22%3A%20false%2C%20%22true%22%3A%20true%2C%20%22false%22%3A%20false%7D%7D%2C%20%22path%22%3A%20%7B%22active%22%3A%20true%7D%2C%20%22file%22%3A%20%7B%22active%22%3A%20true%7D%2C%20%22url%22%3A%20%7B%22active%22%3A%20true%7D%2C%20%22timeseries%22%3A%20%7B%22active%22%3A%20false%2C%20%22sortby%22%3A%20null%2C%20%22autocorrelation%22%3A%200.7%2C%20%22lags%22%3A%20%5B1%2C%207%2C%2012%2C%2024%2C%2030%5D%2C%20%22significance%22%3A%200.05%2C%20%22pacf_acf_lag%22%3A%20100%2C%20%22autolag%22%3A%20%22AIC%22%2C%20%22maxlag%22%3A%20null%7D%7D%2C%20%22sort%22%3A%20null%2C%20%22missing_diagrams%22%3A%20%7B%22bar%22%3A%20true%2C%20%22matrix%22%3A%20true%2C%20%22heatmap%22%3A%20true%7D%2C%20%22correlation_table%22%3A%20true%2C%20%22correlations%22%3A%20%7B%22auto%22%3A%20%7B%22key%22%3A%20%22auto%22%2C%20%22calculate%22%3A%20true%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%2C%20%22spearman%22%3A%20%7B%22key%22%3A%20%22spearman%22%2C%20%22calculate%22%3A%20false%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%2C%20%22pearson%22%3A%20%7B%22key%22%3A%20%22pearson%22%2C%20%22calculate%22%3A%20false%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%2C%20%22phi_k%22%3A%20%7B%22key%22%3A%20%22phi_k%22%2C%20%22calculate%22%3A%20false%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%2C%20%22cramers%22%3A%20%7B%22key%22%3A%20%22cramers%22%2C%20%22calculate%22%3A%20false%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%2C%20%22kendall%22%3A%20%7B%22key%22%3A%20%22kendall%22%2C%20%22calculate%22%3A%20false%2C%20%22warn_high_correlations%22%3A%2010%2C%20%22threshold%22%3A%200.5%2C%20%22n_bins%22%3A%2010%7D%7D%2C%20%22interactions%22%3A%20%7B%22continuous%22%3A%20true%2C%20%22targets%22%3A%20%5B%5D%7D%2C%20%22categorical_maximum_correlation_distinct%22%3A%20100%2C%20%22memory_deep%22%3A%20true%2C%20%22plot%22%3A%20%7B%22missing%22%3A%20%7B%22force_labels%22%3A%20true%2C%20%22cmap%22%3A%20%22RdBu%22%7D%2C%20%22image_format%22%3A%20%22svg%22%2C%20%22correlation%22%3A%20%7B%22cmap%22%3A%20%22RdBu%22%2C%20%22bad%22%3A%20%22%23000000%22%7D%2C%20%22dpi%22%3A%20800%2C%20%22histogram%22%3A%20%7B%22bins%22%3A%2050%2C%20%22max_bins%22%3A%20250%2C%20%22x_axis_labels%22%3A%20true%2C%20%22density%22%3A%20false%7D%2C%20%22scatter_threshold%22%3A%201000%2C%20%22cat_freq%22%3A%20%7B%22show%22%3A%20true%2C%20%22type%22%3A%20%22bar%22%2C%20%22max_unique%22%3A%2010%2C%20%22colors%22%3A%20null%7D%2C%20%22font_path%22%3A%20null%7D%2C%20%22duplicates%22%3A%20%7B%22head%22%3A%2010%2C%20%22key%22%3A%20%22%23%20duplicates%22%7D%2C%20%22samples%22%3A%20%7B%22head%22%3A%2010%2C%20%22tail%22%3A%2010%2C%20%22random%22%3A%200%7D%2C%20%22reject_variables%22%3A%20true%2C%20%22n_obs_unique%22%3A%2010%2C%20%22n_freq_table_max%22%3A%2010%2C%20%22n_extreme_obs%22%3A%2010%2C%20%22report%22%3A%20%7B%22precision%22%3A%208%7D%2C%20%22html%22%3A%20%7B%22style%22%3A%20%7B%22primary_colors%22%3A%20%5B%22%230d6efd%22%2C%20%22%23dc3545%22%2C%20%22%23198754%22%5D%2C%20%22logo%22%3A%20%22%22%2C%20%22theme%22%3A%20null%7D%2C%20%22navbar_show%22%3A%20true%2C%20%22minify_html%22%3A%20true%2C%20%22use_local_assets%22%3A%20true%2C%20%22inline%22%3A%20true%2C%20%22assets_prefix%22%3A%20null%2C%20%22assets_path%22%3A%20null%2C%20%22full_width%22%3A%20false%7D%2C%20%22notebook%22%3A%20%7B%22iframe%22%3A%20%7B%22height%22%3A%20%22800px%22%2C%20%22width%22%3A%20%22100%25%22%2C%20%22attribute%22%3A%20%22srcdoc%22%7D%7D%7D) |

# Variables

Select Columns survived pclass sex age sibsp parch fare embarked class who adult\_male deck embark\_town alive alone

[survived](#pp_var_-844320206311774204)  
Categorical

High correlation

|  |  |
| --- | --- |
| Distinct | 2 |
| Distinct (%) | 0.2% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 50.6 KiB |

|  |  |
| --- | --- |
| 0 | 549 |
| 1 | 342 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 1 |
| Median length | 1 |
| Mean length | 1 |
| Min length | 1 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 891 |
| 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 | 0 |
| 2nd row | 1 |
| 3rd row | 1 |
| 4th row | 1 |
| 5th row | 0 |

#### Common Values

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

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

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

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

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

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per category

##### (unknown)

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

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per script

##### (unknown)

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

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per block

##### (unknown)

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

[pclass](#pp_var_4031927281142044962)  
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

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 1 |
| Median length | 1 |
| Mean length | 1 |
| Min length | 1 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 891 |
| 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 | 3 |
| 2nd row | 1 |
| 3rd row | 3 |
| 4th row | 1 |
| 5th row | 3 |

#### Common Values

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

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

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

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

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

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per category

##### (unknown)

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

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per script

##### (unknown)

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

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 891 | 100.0% |

#### Most frequent character per block

##### (unknown)

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

[sex](#pp_var_-7373172888818346473)  
Categorical

High correlation

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

|  |  |
| --- | --- |
| male | 577 |
| female | 314 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 6 |
| Median length | 4 |
| Mean length | 4.704826 |
| Min length | 4 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 4192 |
| Distinct characters | 5 |
| 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 | male |
| 2nd row | female |
| 3rd row | female |
| 4th row | female |
| 5th row | male |

#### Common Values

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

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

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

* Characters
* Categories
* Scripts
* Blocks

#### 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 (%) |
| --- | --- | --- |
| (unknown) | 4192 | 100.0% |

#### Most frequent character per category

##### (unknown)

| 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 (%) |
| --- | --- | --- |
| (unknown) | 4192 | 100.0% |

#### Most frequent character per script

##### (unknown)

| 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 (%) |
| --- | --- | --- |
| (unknown) | 4192 | 100.0% |

#### Most frequent character per block

##### (unknown)

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

[age](#pp_var_2684574419960095649)  
Real number (ℝ)

High correlation  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 |

More details

* Statistics
* Histogram
* Common values
* Extreme values

Quantile statistics

|  |  |
| --- | --- |
| Minimum | 0.42 |
| 5-th percentile | 4 |
| Q1 | 20.125 |
| median | 28 |
| Q3 | 38 |
| 95-th percentile | 56 |
| Maximum | 80 |
| Range | 79.58 |
| Interquartile range (IQR) | 17.875 |

Descriptive statistics

|  |  |
| --- | --- |
| Standard deviation | 14.526497 |
| Coefficient of variation (CV) | 0.48912219 |
| Kurtosis | 0.17827415 |
| Mean | 29.699118 |
| Median Absolute Deviation (MAD) | 9 |
| Skewness | 0.38910778 |
| Sum | 21205.17 |
| Variance | 211.01912 |
| Monotonicity | Not monotonic |

**Histogram with fixed size bins** (bins=50)

| 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% |

* Minimum 10 values
* Maximum 10 values

| 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](#pp_var_7919890951915530931)  
Real number (ℝ)

High correlation  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 |

More details

* Statistics
* Histogram
* Common values
* Extreme values

Quantile statistics

|  |  |
| --- | --- |
| Minimum | 0 |
| 5-th percentile | 0 |
| Q1 | 0 |
| median | 0 |
| Q3 | 1 |
| 95-th percentile | 3 |
| Maximum | 8 |
| Range | 8 |
| Interquartile range (IQR) | 1 |

Descriptive statistics

|  |  |
| --- | --- |
| Standard deviation | 1.1027434 |
| Coefficient of variation (CV) | 2.1084644 |
| Kurtosis | 17.88042 |
| Mean | 0.52300786 |
| Median Absolute Deviation (MAD) | 0 |
| Skewness | 3.6953517 |
| Sum | 466 |
| Variance | 1.2160431 |
| Monotonicity | Not monotonic |

**Histogram with fixed size bins** (bins=7)

| 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% |

* Minimum 10 values
* Maximum 10 values

| 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](#pp_var_5885521624025493630)  
Real number (ℝ)

High correlation  Zeros

|  |  |
| --- | --- |
| Distinct | 7 |
| Distinct (%) | 0.8% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Infinite | 0 |
| Infinite (%) | 0.0% |
| Mean | 0.38159371 |

|  |  |
| --- | --- |
| Minimum | 0 |
| Maximum | 6 |
| Zeros | 678 |
| Zeros (%) | 76.1% |
| Negative | 0 |
| Negative (%) | 0.0% |
| Memory size | 7.1 KiB |

More details

* Statistics
* Histogram
* Common values
* Extreme values

Quantile statistics

|  |  |
| --- | --- |
| Minimum | 0 |
| 5-th percentile | 0 |
| Q1 | 0 |
| median | 0 |
| Q3 | 0 |
| 95-th percentile | 2 |
| Maximum | 6 |
| Range | 6 |
| Interquartile range (IQR) | 0 |

Descriptive statistics

|  |  |
| --- | --- |
| Standard deviation | 0.80605722 |
| Coefficient of variation (CV) | 2.1123441 |
| Kurtosis | 9.7781252 |
| Mean | 0.38159371 |
| Median Absolute Deviation (MAD) | 0 |
| Skewness | 2.749117 |
| Sum | 340 |
| Variance | 0.64972824 |
| Monotonicity | Not monotonic |

**Histogram with fixed size bins** (bins=7)

| 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% |

* Minimum 10 values
* Maximum 10 values

| 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% |

[fare](#pp_var_1765197558829918918)  
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 |

More details

* Statistics
* Histogram
* Common values
* Extreme values

Quantile statistics

|  |  |
| --- | --- |
| Minimum | 0 |
| 5-th percentile | 7.225 |
| Q1 | 7.9104 |
| median | 14.4542 |
| Q3 | 31 |
| 95-th percentile | 112.07915 |
| Maximum | 512.3292 |
| Range | 512.3292 |
| Interquartile range (IQR) | 23.0896 |

Descriptive statistics

|  |  |
| --- | --- |
| Standard deviation | 49.693429 |
| Coefficient of variation (CV) | 1.5430725 |
| Kurtosis | 33.398141 |
| Mean | 32.204208 |
| Median Absolute Deviation (MAD) | 6.9042 |
| Skewness | 4.7873165 |
| Sum | 28693.949 |
| Variance | 2469.4368 |
| Monotonicity | 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% |
| 26.55 | 15 | 1.7% |
| Other values (238) | 615 | 69.0% |

* Minimum 10 values
* Maximum 10 values

| 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% |

[embarked](#pp_var_5003508584533474652)  
Categorical

High correlation

|  |  |
| --- | --- |
| Distinct | 3 |
| Distinct (%) | 0.3% |
| Missing | 2 |
| Missing (%) | 0.2% |
| Memory size | 50.6 KiB |

|  |  |
| --- | --- |
| S | 644 |
| C | 168 |
| Q | 77 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 1 |
| Median length | 1 |
| Mean length | 1 |
| Min length | 1 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 889 |
| 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 | S |
| 2nd row | C |
| 3rd row | S |
| 4th row | S |
| 5th row | S |

#### Common Values

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

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

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

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

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

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 889 | 100.0% |

#### Most frequent character per category

##### (unknown)

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

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 889 | 100.0% |

#### Most frequent character per script

##### (unknown)

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

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 889 | 100.0% |

#### Most frequent character per block

##### (unknown)

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

[class](#pp_var_6771237817665225922)  
Categorical

High correlation

|  |  |
| --- | --- |
| Distinct | 3 |
| Distinct (%) | 0.3% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 1.3 KiB |

|  |  |
| --- | --- |
| Third | 491 |
| First | 216 |
| Second | 184 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 6 |
| Median length | 5 |
| Mean length | 5.2065095 |
| Min length | 5 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 4639 |
| Distinct characters | 13 |
| 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 | Third |
| 2nd row | First |
| 3rd row | Third |
| 4th row | First |
| 5th row | Third |

#### Common Values

| Value | Count | Frequency (%) |
| --- | --- | --- |
| Third | 491 | 55.1% |
| First | 216 | 24.2% |
| Second | 184 | 20.7% |

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| third | 491 | 55.1% |
| first | 216 | 24.2% |
| second | 184 | 20.7% |

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

| Value | Count | Frequency (%) |
| --- | --- | --- |
| i | 707 | 15.2% |
| r | 707 | 15.2% |
| d | 675 | 14.6% |
| T | 491 | 10.6% |
| h | 491 | 10.6% |
| F | 216 | 4.7% |
| s | 216 | 4.7% |
| t | 216 | 4.7% |
| S | 184 | 4.0% |
| e | 184 | 4.0% |
| Other values (3) | 552 | 11.9% |

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 4639 | 100.0% |

#### Most frequent character per category

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| i | 707 | 15.2% |
| r | 707 | 15.2% |
| d | 675 | 14.6% |
| T | 491 | 10.6% |
| h | 491 | 10.6% |
| F | 216 | 4.7% |
| s | 216 | 4.7% |
| t | 216 | 4.7% |
| S | 184 | 4.0% |
| e | 184 | 4.0% |
| Other values (3) | 552 | 11.9% |

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 4639 | 100.0% |

#### Most frequent character per script

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| i | 707 | 15.2% |
| r | 707 | 15.2% |
| d | 675 | 14.6% |
| T | 491 | 10.6% |
| h | 491 | 10.6% |
| F | 216 | 4.7% |
| s | 216 | 4.7% |
| t | 216 | 4.7% |
| S | 184 | 4.0% |
| e | 184 | 4.0% |
| Other values (3) | 552 | 11.9% |

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 4639 | 100.0% |

#### Most frequent character per block

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| i | 707 | 15.2% |
| r | 707 | 15.2% |
| d | 675 | 14.6% |
| T | 491 | 10.6% |
| h | 491 | 10.6% |
| F | 216 | 4.7% |
| s | 216 | 4.7% |
| t | 216 | 4.7% |
| S | 184 | 4.0% |
| e | 184 | 4.0% |
| Other values (3) | 552 | 11.9% |

[who](#pp_var_3267960624249646528)  
Categorical

High correlation

|  |  |
| --- | --- |
| Distinct | 3 |
| Distinct (%) | 0.3% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 53.0 KiB |

|  |  |
| --- | --- |
| man | 537 |
| woman | 271 |
| child | 83 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 5 |
| Median length | 3 |
| Mean length | 3.7946128 |
| Min length | 3 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 3381 |
| Distinct characters | 10 |
| 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 | man |
| 2nd row | woman |
| 3rd row | woman |
| 4th row | woman |
| 5th row | man |

#### Common Values

| Value | Count | Frequency (%) |
| --- | --- | --- |
| man | 537 | 60.3% |
| woman | 271 | 30.4% |
| child | 83 | 9.3% |

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| man | 537 | 60.3% |
| woman | 271 | 30.4% |
| child | 83 | 9.3% |

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

| Value | Count | Frequency (%) |
| --- | --- | --- |
| m | 808 | 23.9% |
| a | 808 | 23.9% |
| n | 808 | 23.9% |
| w | 271 | 8.0% |
| o | 271 | 8.0% |
| c | 83 | 2.5% |
| h | 83 | 2.5% |
| i | 83 | 2.5% |
| l | 83 | 2.5% |
| d | 83 | 2.5% |

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 3381 | 100.0% |

#### Most frequent character per category

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| m | 808 | 23.9% |
| a | 808 | 23.9% |
| n | 808 | 23.9% |
| w | 271 | 8.0% |
| o | 271 | 8.0% |
| c | 83 | 2.5% |
| h | 83 | 2.5% |
| i | 83 | 2.5% |
| l | 83 | 2.5% |
| d | 83 | 2.5% |

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 3381 | 100.0% |

#### Most frequent character per script

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| m | 808 | 23.9% |
| a | 808 | 23.9% |
| n | 808 | 23.9% |
| w | 271 | 8.0% |
| o | 271 | 8.0% |
| c | 83 | 2.5% |
| h | 83 | 2.5% |
| i | 83 | 2.5% |
| l | 83 | 2.5% |
| d | 83 | 2.5% |

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 3381 | 100.0% |

#### Most frequent character per block

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| m | 808 | 23.9% |
| a | 808 | 23.9% |
| n | 808 | 23.9% |
| w | 271 | 8.0% |
| o | 271 | 8.0% |
| c | 83 | 2.5% |
| h | 83 | 2.5% |
| i | 83 | 2.5% |
| l | 83 | 2.5% |
| d | 83 | 2.5% |

[adult\_male](#pp_var_646073766436733809)  
Boolean

High correlation

|  |  |
| --- | --- |
| Distinct | 2 |
| Distinct (%) | 0.2% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 1019.0 B |

|  |  |
| --- | --- |
| True | 537 |
| False | 354 |

More details

* Common Values (Table)
* Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| True | 537 | 60.3% |
| False | 354 | 39.7% |

[deck](#pp_var_3015878985090200450)  
Categorical

High correlation  Missing

|  |  |
| --- | --- |
| Distinct | 7 |
| Distinct (%) | 3.4% |
| Missing | 688 |
| Missing (%) | 77.2% |
| Memory size | 1.7 KiB |

|  |  |
| --- | --- |
| C | 59 |
| B | 47 |
| D | 33 |
| E | 32 |
| A | 15 |
| Other values (2) | 17 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 1 |
| Median length | 1 |
| Mean length | 1 |
| Min length | 1 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 203 |
| Distinct characters | 7 |
| 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 | C |
| 2nd row | C |
| 3rd row | E |
| 4th row | G |
| 5th row | C |

#### Common Values

| Value | Count | Frequency (%) |
| --- | --- | --- |
| C | 59 | 6.6% |
| B | 47 | 5.3% |
| D | 33 | 3.7% |
| E | 32 | 3.6% |
| A | 15 | 1.7% |
| F | 13 | 1.5% |
| G | 4 | 0.4% |
| (Missing) | 688 | 77.2% |

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| c | 59 | 29.1% |
| b | 47 | 23.2% |
| d | 33 | 16.3% |
| e | 32 | 15.8% |
| a | 15 | 7.4% |
| f | 13 | 6.4% |
| g | 4 | 2.0% |

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

| Value | Count | Frequency (%) |
| --- | --- | --- |
| C | 59 | 29.1% |
| B | 47 | 23.2% |
| D | 33 | 16.3% |
| E | 32 | 15.8% |
| A | 15 | 7.4% |
| F | 13 | 6.4% |
| G | 4 | 2.0% |

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 203 | 100.0% |

#### Most frequent character per category

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| C | 59 | 29.1% |
| B | 47 | 23.2% |
| D | 33 | 16.3% |
| E | 32 | 15.8% |
| A | 15 | 7.4% |
| F | 13 | 6.4% |
| G | 4 | 2.0% |

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 203 | 100.0% |

#### Most frequent character per script

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| C | 59 | 29.1% |
| B | 47 | 23.2% |
| D | 33 | 16.3% |
| E | 32 | 15.8% |
| A | 15 | 7.4% |
| F | 13 | 6.4% |
| G | 4 | 2.0% |

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 203 | 100.0% |

#### Most frequent character per block

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| C | 59 | 29.1% |
| B | 47 | 23.2% |
| D | 33 | 16.3% |
| E | 32 | 15.8% |
| A | 15 | 7.4% |
| F | 13 | 6.4% |
| G | 4 | 2.0% |

[embark\_town](#pp_var_-234494340988115811)  
Categorical

High correlation

|  |  |
| --- | --- |
| Distinct | 3 |
| Distinct (%) | 0.3% |
| Missing | 2 |
| Missing (%) | 0.2% |
| Memory size | 58.9 KiB |

|  |  |
| --- | --- |
| Southampton | 644 |
| Cherbourg | 168 |
| Queenstown | 77 |

More details

* Overview
* Categories
* Words
* Characters

Length

|  |  |
| --- | --- |
| Max length | 11 |
| Median length | 11 |
| Mean length | 10.535433 |
| Min length | 9 |

Characters and Unicode

|  |  |
| --- | --- |
| Total characters | 9366 |
| Distinct characters | 17 |
| 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 | Southampton |
| 2nd row | Cherbourg |
| 3rd row | Southampton |
| 4th row | Southampton |
| 5th row | Southampton |

#### Common Values

| Value | Count | Frequency (%) |
| --- | --- | --- |
| Southampton | 644 | 72.3% |
| Cherbourg | 168 | 18.9% |
| Queenstown | 77 | 8.6% |
| (Missing) | 2 | 0.2% |

#### Length

Histogram of lengths of the category

#### Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| southampton | 644 | 72.4% |
| cherbourg | 168 | 18.9% |
| queenstown | 77 | 8.7% |

* Characters
* Categories
* Scripts
* Blocks

#### Most occurring characters

| Value | Count | Frequency (%) |
| --- | --- | --- |
| o | 1533 | 16.4% |
| t | 1365 | 14.6% |
| u | 889 | 9.5% |
| h | 812 | 8.7% |
| n | 798 | 8.5% |
| S | 644 | 6.9% |
| a | 644 | 6.9% |
| m | 644 | 6.9% |
| p | 644 | 6.9% |
| r | 336 | 3.6% |
| Other values (7) | 1057 | 11.3% |

#### Most occurring categories

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 9366 | 100.0% |

#### Most frequent character per category

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| o | 1533 | 16.4% |
| t | 1365 | 14.6% |
| u | 889 | 9.5% |
| h | 812 | 8.7% |
| n | 798 | 8.5% |
| S | 644 | 6.9% |
| a | 644 | 6.9% |
| m | 644 | 6.9% |
| p | 644 | 6.9% |
| r | 336 | 3.6% |
| Other values (7) | 1057 | 11.3% |

#### Most occurring scripts

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 9366 | 100.0% |

#### Most frequent character per script

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| o | 1533 | 16.4% |
| t | 1365 | 14.6% |
| u | 889 | 9.5% |
| h | 812 | 8.7% |
| n | 798 | 8.5% |
| S | 644 | 6.9% |
| a | 644 | 6.9% |
| m | 644 | 6.9% |
| p | 644 | 6.9% |
| r | 336 | 3.6% |
| Other values (7) | 1057 | 11.3% |

#### Most occurring blocks

| Value | Count | Frequency (%) |
| --- | --- | --- |
| (unknown) | 9366 | 100.0% |

#### Most frequent character per block

##### (unknown)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| o | 1533 | 16.4% |
| t | 1365 | 14.6% |
| u | 889 | 9.5% |
| h | 812 | 8.7% |
| n | 798 | 8.5% |
| S | 644 | 6.9% |
| a | 644 | 6.9% |
| m | 644 | 6.9% |
| p | 644 | 6.9% |
| r | 336 | 3.6% |
| Other values (7) | 1057 | 11.3% |

[alive](#pp_var_5599174402278627879)  
Boolean

High correlation

|  |  |
| --- | --- |
| Distinct | 2 |
| Distinct (%) | 0.2% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 1019.0 B |

|  |  |
| --- | --- |
| False | 549 |
| True | 342 |

More details

* Common Values (Table)
* Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| False | 549 | 61.6% |
| True | 342 | 38.4% |

[alone](#pp_var_3893282888255956381)  
Boolean

High correlation

|  |  |
| --- | --- |
| Distinct | 2 |
| Distinct (%) | 0.2% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 1019.0 B |

|  |  |
| --- | --- |
| True | 537 |
| False | 354 |

More details

* Common Values (Table)
* Common Values (Plot)

| Value | Count | Frequency (%) |
| --- | --- | --- |
| True | 537 | 60.3% |
| False | 354 | 39.7% |

# Interactions

* age
* sibsp
* parch
* fare

* fare
* age
* sibsp
* parch

* fare
* age
* sibsp
* parch

* fare
* age
* sibsp
* parch

* fare
* age
* sibsp
* parch

# Correlations

* Auto

* Heatmap
* Table

|  | adult\_male | age | alive | alone | class | deck | embark\_town | embarked | fare | parch | pclass | sex | sibsp | survived | who |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| adult\_male | 1.000 | 0.392 | 0.554 | 0.401 | 0.096 | 0.208 | 0.101 | 0.101 | 0.164 | 0.392 | 0.096 | 0.906 | 0.323 | 0.554 | 0.999 |
| age | 0.392 | 1.000 | 0.155 | 0.351 | 0.269 | 0.197 | 0.065 | 0.065 | 0.135 | -0.254 | 0.269 | 0.099 | -0.182 | 0.155 | 0.650 |
| alive | 0.554 | 0.155 | 1.000 | 0.198 | 0.337 | 0.107 | 0.166 | 0.166 | 0.283 | 0.157 | 0.337 | 0.540 | 0.187 | 0.998 | 0.563 |
| alone | 0.401 | 0.351 | 0.198 | 1.000 | 0.127 | 0.196 | 0.110 | 0.110 | 0.304 | 0.686 | 0.127 | 0.300 | 0.837 | 0.198 | 0.451 |
| class | 0.096 | 0.269 | 0.337 | 0.127 | 1.000 | 0.626 | 0.260 | 0.260 | 0.479 | 0.022 | 1.000 | 0.130 | 0.148 | 0.337 | 0.137 |
| deck | 0.208 | 0.197 | 0.107 | 0.196 | 0.626 | 1.000 | 0.147 | 0.147 | 0.182 | 0.136 | 0.626 | 0.238 | 0.114 | 0.107 | 0.273 |
| embark\_town | 0.101 | 0.065 | 0.166 | 0.110 | 0.260 | 0.147 | 1.000 | 1.000 | 0.196 | 0.052 | 0.260 | 0.113 | 0.092 | 0.166 | 0.079 |
| embarked | 0.101 | 0.065 | 0.166 | 0.110 | 0.260 | 0.147 | 1.000 | 1.000 | 0.196 | 0.052 | 0.260 | 0.113 | 0.092 | 0.166 | 0.079 |
| fare | 0.164 | 0.135 | 0.283 | 0.304 | 0.479 | 0.182 | 0.196 | 0.196 | 1.000 | 0.410 | 0.479 | 0.189 | 0.447 | 0.283 | 0.159 |
| parch | 0.392 | -0.254 | 0.157 | 0.686 | 0.022 | 0.136 | 0.052 | 0.052 | 0.410 | 1.000 | 0.022 | 0.247 | 0.450 | 0.157 | 0.386 |
| pclass | 0.096 | 0.269 | 0.337 | 0.127 | 1.000 | 0.626 | 0.260 | 0.260 | 0.479 | 0.022 | 1.000 | 0.130 | 0.148 | 0.337 | 0.137 |
| sex | 0.906 | 0.099 | 0.540 | 0.300 | 0.130 | 0.238 | 0.113 | 0.113 | 0.189 | 0.247 | 0.130 | 1.000 | 0.206 | 0.540 | 0.947 |
| sibsp | 0.323 | -0.182 | 0.187 | 0.837 | 0.148 | 0.114 | 0.092 | 0.092 | 0.447 | 0.450 | 0.148 | 0.206 | 1.000 | 0.187 | 0.373 |
| survived | 0.554 | 0.155 | 0.998 | 0.198 | 0.337 | 0.107 | 0.166 | 0.166 | 0.283 | 0.157 | 0.337 | 0.540 | 0.187 | 1.000 | 0.563 |
| who | 0.999 | 0.650 | 0.563 | 0.451 | 0.137 | 0.273 | 0.079 | 0.079 | 0.159 | 0.386 | 0.137 | 0.947 | 0.373 | 0.563 | 1.000 |

# Missing values

* Count
* Matrix
* Heatmap

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

* First rows
* Last rows

|  | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult\_male | deck | embark\_town | alive | alone |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |
| 5 | 0 | 3 | male | NaN | 0 | 0 | 8.4583 | Q | Third | man | True | NaN | Queenstown | no | True |
| 6 | 0 | 1 | male | 54.0 | 0 | 0 | 51.8625 | S | First | man | True | E | Southampton | no | True |
| 7 | 0 | 3 | male | 2.0 | 3 | 1 | 21.0750 | S | Third | child | False | NaN | Southampton | no | False |
| 8 | 1 | 3 | female | 27.0 | 0 | 2 | 11.1333 | S | Third | woman | False | NaN | Southampton | yes | False |
| 9 | 1 | 2 | female | 14.0 | 1 | 0 | 30.0708 | C | Second | child | False | NaN | Cherbourg | yes | False |

|  | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult\_male | deck | embark\_town | alive | alone |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 881 | 0 | 3 | male | 33.0 | 0 | 0 | 7.8958 | S | Third | man | True | NaN | Southampton | no | True |
| 882 | 0 | 3 | female | 22.0 | 0 | 0 | 10.5167 | S | Third | woman | False | NaN | Southampton | no | True |
| 883 | 0 | 2 | male | 28.0 | 0 | 0 | 10.5000 | S | Second | man | True | NaN | Southampton | no | True |
| 884 | 0 | 3 | male | 25.0 | 0 | 0 | 7.0500 | S | Third | man | True | NaN | Southampton | no | True |
| 885 | 0 | 3 | female | 39.0 | 0 | 5 | 29.1250 | Q | Third | woman | False | NaN | Queenstown | no | False |
| 886 | 0 | 2 | male | 27.0 | 0 | 0 | 13.0000 | S | Second | man | True | NaN | Southampton | no | True |
| 887 | 1 | 1 | female | 19.0 | 0 | 0 | 30.0000 | S | First | woman | False | B | Southampton | yes | True |
| 888 | 0 | 3 | female | NaN | 1 | 2 | 23.4500 | S | Third | woman | False | NaN | Southampton | no | False |
| 889 | 1 | 1 | male | 26.0 | 0 | 0 | 30.0000 | C | First | man | True | C | Cherbourg | yes | True |
| 890 | 0 | 3 | male | 32.0 | 0 | 0 | 7.7500 | Q | Third | man | True | NaN | Queenstown | no | True |

# Duplicate rows

#### Most frequently occurring

|  | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult\_male | deck | embark\_town | alive | alone | # duplicates |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 35 | 0 | 3 | male | NaN | 0 | 0 | 7.8958 | S | Third | man | True | NaN | Southampton | no | True | 13 |
| 36 | 0 | 3 | male | NaN | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True | 12 |
| 33 | 0 | 3 | male | NaN | 0 | 0 | 7.7500 | Q | Third | man | True | NaN | Queenstown | no | True | 8 |
| 46 | 1 | 3 | female | NaN | 0 | 0 | 7.7500 | Q | Third | woman | False | NaN | Queenstown | yes | True | 7 |
| 9 | 0 | 2 | male | NaN | 0 | 0 | 0.0000 | S | Second | man | True | NaN | Southampton | no | True | 6 |
| 30 | 0 | 3 | male | NaN | 0 | 0 | 7.2250 | C | Third | man | True | NaN | Cherbourg | no | True | 5 |
| 31 | 0 | 3 | male | NaN | 0 | 0 | 7.2292 | C | Third | man | True | NaN | Cherbourg | no | True | 5 |
| 40 | 0 | 3 | male | NaN | 8 | 2 | 69.5500 | S | Third | man | True | NaN | Southampton | no | False | 4 |
| 2 | 0 | 2 | male | 23.0 | 0 | 0 | 13.0000 | S | Second | man | True | NaN | Southampton | no | True | 3 |
| 3 | 0 | 2 | male | 25.0 | 0 | 0 | 13.0000 | S | Second | man | True | NaN | Southampton | no | True | 3 |

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