Package ‘vvauditor’

February 1, 2023

Title Creates Assertion Tests

Version 0.5.8

Description
Offers a comprehensive set of assertion tests to help users validate the integrity of their data. These tests can be used to check for specific conditions or properties within a dataset and help ensure that data is accurate and reliable. The package is designed to make it easy to add quality control checks to data analysis workflows and to aid in identifying and correcting any errors or inconsistencies in data.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports dplyr, findR, magrittr, stats, stringr

NeedsCompilation no

Author Tomer Iwan [aut, cre, cph]

Maintainer Tomer Iwan <t.iwan@vu.nl>

Repository CRAN

Date/Publication 2023-02-01 10:50:05 UTC

R topics documented:

check_double_columns .................................................. 2
check_rows .............................................................. 2
count_more_than_1 .................................................... 3
drop_na_column_names .................................................. 3
duplicates_in_column ................................................... 4
find_pattern_r ............................................................ 5
get_values ............................................................... 5
md_complete_cases ..................................................... 6
retrieve_functions_and_packages ...................................... 6
str_detect_in_file ..................................................... 7
test_all_equal .......................................................... 7
unique_id ............................................................... 8
check_double_columns

Description
Check whether two dataframes have intersecting column names.

Usage
check_double_columns(x, y, connector = NULL)

Arguments
x Data frame x.
y Data frame y.
connector The connector columns as strings. Also possible as vector.

Value
Message informing about overlap in columns between the dataframes.

See Also
Other tests: duplicates_in_column(), test_all_equal()

Examples
check_double_columns(mtcars, iris)

check_rows

Description
This function prints the number of rows of a data frame. This function is used to check that rows are not deleted or doubled unless expected.

Usage
check_rows(df, name = NULL)

Arguments
df The data frame whose rows are to be counted
name The name of the data file (this will be printed)
**count_more_than_1**

**Value**

A message is printed to the console with the number of rows of the data

**Examples**

```r
count_more_than_1(mtcars)
```

---

**Description**

Function to count the number of values greater than 1 in a vector. This function is used in the function `check_rows` to count duplicate values.

**Usage**

```r
count_more_than_1(x)
```

**Arguments**

- `x` The vector to test

**Value**

Number of values greater than 1.

**Examples**

```r
count_more_than_1(c(1, 1, 4))
```

---

**drop_na_column_names**

**Description**

Deletes columns whose name is NA or whose name is empty.

**Usage**

```r
drop_na_column_names(x)
```
**Arguments**

- `df` dataframe

**Value**

dataframe without columns that are NA

---

**Description**

Searches for duplicates in a data frame column.

**Usage**

duplicates_in_column(df, col)

**Arguments**

- `df` Data frame.
- `col` Column name.

**Value**

Rows containing duplicated values.

**See Also**

Other tests: `check_double_columns()`, `test_all_equal()`

**Examples**

duplicates_in_column(mtcars, "mpg")
**find_pattern_r**

*Find pattern in R scripts*

**Description**

Function to search for a pattern in R scripts.

**Usage**

```r
find_pattern_r(pattern, path = ".", case.sensitive = TRUE, comments = FALSE)
```

**Arguments**

- `pattern`: Pattern to search
- `path`: Directory to search in
- `case.sensitive`: Whether pattern is case sensitive or not
- `comments`: whether to search in commented lines

**Value**

Dataframe containing R script paths

---

**get_values**

*Get values of column*

**Description**

A function to determine what kind of values are present in columns.

**Usage**

```r
get_values(df, column)
```

**Arguments**

- `df`: The dataframe
- `column`: Column to get values from.

**Value**

The class of the column values

**Examples**

```r
get_values(mtcars, "mpg")
```
**md_complete_cases**  
*MD complete cases*

**Description**
Print the complete cases of the data.

**Usage**
```r
md_complete_cases(data, digits = 1)
```

**Arguments**
- **data**: The data frame.
- **digits**: Default: 1. number of digits for rounding.

**Value**
Message with the number of rows, number of rows with missing values and the percentage of complete rows.

**Examples**
```r
# example code
md_complete_cases(iris)
md_complete_cases(iris)
```

---

**retrieve_functions_and_packages**  
*Retrieve functions and packages*

**Description**
Retrieves functions and their corresponding packages used in a given script.

**Usage**
```r
retrieve_functions_and_packages(path)
```

**Arguments**
- **path**: The complete path of the script.
str_detect_in_file

Value

Used_functions

**str_detect_in_file**  
*Detect string in file*

**Description**

Detect string in file

**Usage**

```
str_detect_in_file(file, pattern, only_comments = FALSE, collapse = FALSE)
```

**Arguments**

- `file` Path to file.
- `pattern` Pattern to match.
- `only_comments` default FALSE. Whether to only search in commented lines.
- `collapse` default: FALSE: search file line by line. If true, then pattern is search in the entire file at once after collapsing. (only_comments does not work when collapse is set to TRUE)

**Value**

Boolean whether pattern exists in file.

test_all_equal

**Description**

Test whether all values in a vector are equal.

**Usage**

```
test_all_equal(x, na.rm = FALSE)
```

**Arguments**

- `x` Vector to test.
- `na.rm` default: FALSE. exclude NAs from the test.
Value

Boolean result of the test

See Also

Other tests: `check_double_columns()`, `duplicates_in_column()`

Examples

```r
test_all_equal(c(5, 5, 5))

test_all_equal(c(5, 6, 3))
```

---

**unique_id**

<table>
<thead>
<tr>
<th>unique_id</th>
<th>unique id</th>
</tr>
</thead>
</table>

Description

Check if parsed variable is a unique identifier. This function was adapted from: Source: https://edwinth.github.io/blog/unique_id/

Usage

```r
unique_id(x, ...)
```

Arguments

- `x` vector or dataframe.
- `...` optional variables, e.g. name of column or a vector of names.

Value

Boolean whether variable is a unique identifier.

Examples

```r
unique_id(iris, Species)

mtcars$name <- rownames(mtcars)
unique_id(mtcars, name)
```
Index

* tests
  check_double_columns, 2
  duplicates_in_column, 4
  test_all_equal, 7
* vector calculations
  count_more_than_1, 3
check_double_columns, 2, 4, 8
check_rows, 2
count_more_than_1, 3
drop_na_column_names, 3
duplicates_in_column, 2, 4, 8
find_pattern_r, 5
get_values, 5
md_complete_cases, 6
retrieve_functions_and_packages, 6
str_detect_in_file, 7
test_all_equal, 2, 4, 7
unique_id, 8