Package 'edm1.df'

July 26, 2024

Title Set of functions for dataframe manipulation

Version 2.0.0.0	
Description Provides multiple functions to manipulate data in dataframe according to different algorithms for different goals.	У
License GPL (==3)	
Encoding UTF-8	
Roxygen list(markdown = TRUE)	
RoxygenNote 7.3.1	
Imports stringr, stringi, dplyr, openxlsx	
Contents	
nestr_datf1 nestr_datf2 paste_datf see_datf see_datf swipr unique_datf val_replacer vec_in_datf vlookup_datf	10 11 12 13 13 13 15 19
Index	2

2 colins_datf

Description

Allow to insert vectors into a dataframe.

Usage

```
colins_datf(inpt_datf, target_col = list(), target_pos = list())
```

Arguments

```
inpt_datf is the dataframe where vectors will be inserted
target_col is a list containing all the vectors to be inserted
target_pos is a list containing the vectors made of the columns names or numbers where
the associated vectors from target_col will be inserted after
```

```
datf1 <- data.frame("frst_col"=c(1:5), "scd_col"=c(5:1))</pre>
print(colins_datf(inpt_datf=datf1, target_col=list(c("oui", "oui", "oui", "non", "non"),
            c("u", "z", "z", "z", "u")),
              target_pos=list(c("frst_col", "scd_col"), c("scd_col"))))
# frst_col cur_col scd_col cur_col.1 cur_col
       1 oui 5 oui
#1
#2
         2
              oui
                       4
                              oui
                                        7.
        3
                       3
#3
              oui
                              oui
                                       Z
                       2
#4
        4
                              non
                                        Z
              non
#5
              non
                               non
print(colins_datf(inpt_datf=datf1, target_col=list(c("oui", "oui", "oui", "non", "non"),
            c("u", "z", "z", "z", "u")),
              \verb|target_pos=list(c(1, 2), c("frst_col")))|
# frst_col cur_col scd_col cur_col cur_col
#1
        1 oui 5 u oui
                       4
#2
         2
              oui
                              Z
                                    oui
                       3
                              Z
#3
        3
              oui
                                    oui
#4
        4
                       2
              non
                               Z
                                    non
#5
        5
                       1
              non
                                     non
```

cumulated_rows 3

```
cumulated_rows cumulated_rows
```

Description

Output a vector of size that equals to the rows number of the input dataframe, with TRUE value at the indices corresponding to the row where at least a cell of any column is equal to one of the values inputed in values_v

Usage

```
cumulated_rows(inpt_datf, values_v = c())
```

Arguments

Examples

```
datf_teste <- data.frame(c(1:10), c(10:1))</pre>
print(datf_teste)
   c.1.10. c.10.1.
1
        1
                10
2
         2
                 9
3
         3
                 8
4
         4
                 7
5
         5
                 6
6
         6
                 5
7
         7
8
         8
                 3
9
         9
                 2
10
        10
                 1
print(cumulated_rows(inpt_datf = datf_teste, values_v = c(2, 3)))
      FALSE TRUE TRUE
                                        FALSE FALSE TRUE TRUE
[1]
                         FALSE
                                FALSE
                                                                     FALSE
```

Description

Output a vector of size that equals to the rows number of the input dataframe, with TRUE value at the indices corresponding to the row where at least a cell of any column is equal to NA.

4 cut_v

Usage

```
cumulated_rows_na(inpt_datf)
```

Arguments

```
inpt_datf is the input data.frame
```

Examples

```
datf_teste <- data.frame(c(1, 2, 3, 4, 5, NA, 7), c(10, 9, 8, NA, 7, 6, NA))
print(datf_teste)
 c.1..2..3..4..5..NA..7. c.10..9..8..NA..7..6..NA.
1
                       1
2
                        2
3
                        3
4
                                                 NA
5
                       5
                                                  7
6
                       NA
                                                  6
                                                 NA
print(cumulated_rows_na(inpt_datf = datf_teste))
[1] FALSE FALSE FALSE TRUE FALSE TRUE TRUE
```

```
cut_v cut_v
```

Description

Allow to convert a vector to a dataframe according to a separator.

Usage

```
cut_v(inpt_v, sep_ = "")
```

Arguments

```
inpt_v is the input vector
sep_ is the separator of the elements in inpt_v, defaults to ""
```

diff_datf 5

diff_datf

 $diff_datf$

Description

Returns a vector with the coordinates of the cell that are not equal between 2 dataframes (row, column).

Usage

```
diff_datf(datf1, datf2)
```

Arguments

```
datf1 is an an input dataframe datf2 is an an input dataframe
```

Examples

```
datf1 <- data.frame(c(1:6), c("oui", "oui", "oui", "oui", "oui", "oui", "oui"), c(6:1))
datf2 <- data.frame(c(1:7), c("oui", "oui", "oui", "oui", "non", "oui", "zz"))
print(diff_datf(datf1=datf1, datf2=datf2))
#[1] 5 1 5 2</pre>
```

groupr_datf

groupr_datf

Description

Allow to create groups from a dataframe. Indeed, you can create conditions that lead to a flag value for each cell of the input dataframeaccording to the cell value. This function is based on see_datf and nestr_datf2 functions.

6 historic_sequence

Usage

```
groupr_datf(
  inpt_datf,
  condition_lst,
  val_lst,
  conjunction_lst,
  rtn_val_pos = c()
)
```

Arguments

Examples

```
interactive()
datf1 <- data.frame(c(1, 2, 1), c(45, 22, 88), c(44, 88, 33))
val_lst <- list(list(c(1), c(1)), list(c(2)), list(c(44, 88)))
condition_lst <- list(c(">", "<"), c("%%"), c("==", "=="))
conjunction_lst <- list(c("|"), c(), c("|"))
rtn_val_pos <- c("+", "++", "+++")
print(groupr_datf(inpt_datf=datf1, val_lst=val_lst, condition_lst=condition_lst, conjunction_lst=conjunction_lst, rtn_val_pos=rtn_val_pos))
# X1 X2 X3
#1 <NA> + +++
#2 ++ +++++
#3 <NA> ++++
```

historic_sequence historic_sequence

Description

Allow to perform a pivot wider on a sequencial dataset (here the type is dataframe), each variable will be dupplicated in a column to show the value to this variable at n - 1 for each individual, see examples.

historic_sequence 7

Usage

```
historic_sequence(inpt_datf, bf_ = 1)
```

Arguments

inpt_datf is the input dataframe
bf_ is the number of previous value of the individual it will search for, see examples

Examples

3

20

peut1

101

110

115

14

17

18

```
set.seed(123)
var1 < - round(runif(n = 14, min = 100, max = 122))
set.seed(123)
var2 \leftarrow round(runif(n = 14, min = 14, max = 20))
datf <- data.frame("ids" = c(20, 20, 20, 20, 19, 19, 19, 18, 18, 18, 18,
                         17, 17, 17),
                 "individual" = c("oui", "non", "peut1", "peut2",
                                "oui", "peut1", "peut2"),
                "var1" = var1,
                 "var2" = var2)
print(datf)
  ids individual var1 var2
1
   20
          oui 106 16
            non 117
2
   20
                       19
          peut1 109
3
  20
                      16
          peut2 119
4
  20
                      19
           oui 121
5
  19
                       2.0
6
  19
         peut1 101
                      14
          peut2 112
7
  19
                      17
8
  18
           oui 120
9
  18
           non 112
                      17
10 18
         peut1 110
                      17
11 18
         peut2 121
                       20
12 17
           oui 110
                      17
13 17
          peut1 115
                       18
14 17
          peut2 113
                      17
historic\_sequence(inpt\_datf = datf, bf\_ = 2)
 id_seq individual var1-1 var1-2 var2-1 var2-2
     20
         oui 121 120
                                20
1
                                       19
2
     20
              non
                     NA
                           112
                                  NA
                                         17
3
     20
            peut1
                     101
                           110
                                   14
                                         17
4
     20
            peut2
                     112
                           121
                                   17
                                         20
                                  19
5
     19
             oui
                     120
                           110
                                         17
                                  17
6
     19
                     110
                           115
                                         18
            peut1
     19
                     121
                           113
                                  20
                                         17
            peut2
historic_sequence(inpt_datf = datf, bf_ = 3)
 id_seq individual var1-1 var1-2 var1-3 var2-1 var2-2 var2-3
1
     20 oui 121 120 110 20 19 17
     20
             non
                     NA
                           112
                                  NA
                                         NA
                                                17
                                                      NA
```

id_keepr

```
4 20 peut2 112 121 113 17 20 17
```

id_keepr_datf

Description

Allow to get the original indexes after multiple equality comparaison according to the original number of row

Usage

```
id_keepr(inpt_datf, col_v = c(), el_v = c(), rstr_l = NA)
```

Arguments

inpt_datf	is the input dataframe
col_v	is the vector containing the column numbers or names to be compared to their respective elements in "el_v" $$
el_v	is a vector containing the elements that may be contained in their respective column described in " col_v "
rstr_l	is a list containing the vector composed of the indexes of the elements chosen for each comparison. If the length of the list is inferior to the length of comparisons, so the last vector of rstr_l will be the same as the last one to fill make rstr_l equal in term of length to col_v and el_v

insert_datf 9

insert_datf	edm1 insert_datf
-------------	------------------

Description

Allow to insert dataframe into another dataframe according to coordinates (row, column) from the dataframe that will be inserted

Usage

```
insert_datf(datf_in, datf_ins, ins_loc)
```

Arguments

```
datf_in is the dataframe that will be inserted

datf_ins is the dataset to be inserted

ins_loc is a vector containg two parameters (row, column) of the begining for the insertion
```

```
datf1 \leftarrow data.frame(c(1, 4), c(5, 3))
datf2 \leftarrow data.frame(c(1, 3, 5, 6), c(1:4), c(5, 4, 5, "ereer"))
print(insert_datf(datf_in=datf2, datf_ins=datf1, ins_loc=c(4, 2)))
    c.1..3..5..6. c.1.4. c.5..4..5...ereer..
# 1
                1
                     1
# 2
                3
                       2
                                             4
# 3
                5
                       3
                                             5
# 4
                6
                       1
print(insert_datf(datf_in=datf2, datf_ins=datf1, ins_loc=c(3, 2)))
    c.1..3..5..6. c.1.4. c.5..4..5...ereer..
# 1
                       1
# 2
                3
                       2
                                             4
# 3
                5
                       1
                                             5
# 4
                6
print(insert_datf(datf_in=datf2, datf_ins=datf1, ins_loc=c(2, 2)))
    c.1..3..5..6. c.1.4. c.5..4..5...ereer..
# 1
                1
                      1
# 2
                3
                       1
                                             5
# 3
                5
                       4
                                             3
# 4
                       4
```

10 intersect_mod

Description

Returns the mods that have elements in common

Usage

```
intersect_mod(datf, inter_col, mod_col, n_min, descendly_ordered = NA)
```

Arguments

inter_col is the column name or the column number of the values that may be commun
betwee the different mods

mod_col is the column name or the column number of the mods in the dataframe

n_min is the minimum elements in common a mod should have to be taken in count
ordered_descendly
in case that the elements in commun are numeric, this option can be enabled by

giving a value of TRUE or FALSE see examples

```
datf <- data.frame("col1"=c("oui", "oui", "oui", "oui", "oui", "oui",</pre>
                     "non", "non", "non", "ee", "ee", "ee"), "col2"=c(1:6, 2:5, 1:
print(intersect_mod(datf=datf, inter_col=2, mod_col=1, n_min=2))
   col1 col2
2
   oui
          2
3
    oui
           3
7
           2
    non
8
           3
    non
12
           2
    ee
13
           3
    ee
print(intersect_mod(datf=datf, inter_col=2, mod_col=1, n_min=3))
   col1 col2
2
    oui
3
    oui
           3
4
    oui
           4
5
           5
    oui
7
           2
    non
8
           3
    non
9
           4
    non
10 non
print(intersect_mod(datf=datf, inter_col=2, mod_col=1, n_min=5))
  coll col2
```

nestr_datf1 11

```
1
   oui
          1
2
   oui
3
   oui
          3
4
   oui
          4
5
          5
   oui
6
          6
   oui
datf <- data.frame("col1"=c("non", "non", "oui", "oui", "oui", "oui",</pre>
                       "non", "non", "non", "ee", "ee", "ee"), "col2"=c(1:6, 2:5, 1
print(intersect_mod(datf=datf, inter_col=2, mod_col=1, n_min=3))
   col1 col2
8
    non
           3
9
    non
           4
10 non
           5
           3
3
    oui
4
           4
    oui
5
           5
    011 i
```

nestr_datf1

nestr_datf1

Description

Allow to write a value (1a) to a dataframe (1b) to its cells that have the same coordinates (row and column) than the cells whose value is equal to a another special value (2a), from another another dataframe (2b). The value (1a) depends of the cell value coordinates of the third dataframe (3b). If a cell coordinates (1c) of the first dataframe (1b) does not correspond to the coordinates of a good returning cell value (2a) from the dataframe (2b), so this cell (1c) can have its value changed to the same cell coordinates value (3a) of a third dataframe (4b), if (4b) is not set to NA.

Usage

```
nestr_datf1(
   inptf_datf,
   inptt_pos_datf,
   nestr_datf,
   yes_val = TRUE,
   inptt_neg_datf = NA
)
```

Arguments

12 nestr_datf2

Examples

```
print(nestr_datf1(inptf_datf=data.frame(c(1, 2, 1), c(1, 5, 7)),
inptt_pos_datf=data.frame(c(4, 4, 3), c(2, 1, 2)),
inptt_neg_datf=data.frame(c(44, 44, 33), c(12, 12, 12)),
nestr_datf=data.frame(c(TRUE, FALSE, TRUE), c(FALSE, FALSE, TRUE)), yes_val=TRUE))
  c.1..2..1. c.1..5..7.
#1
           4
#2
           44
                      12
#3
            3
                       2
print(nestr_datf1(inptf_datf=data.frame(c(1, 2, 1), c(1, 5, 7)))
inptt_pos_datf=data.frame(c(4, 4, 3), c(2, 1, 2)),
inptt_neg_datf=NA,
nestr_datf=data.frame(c(TRUE, FALSE, TRUE), c(FALSE, FALSE, TRUE)), yes_val=TRUE))
    c.1..2..1. c.1..5..7.
#1
        4
#2
           2
                       5
#3
           3
                       2
```

Description

Allow to write a special value (1a) in the cells of a dataframe (1b) that correspond (row and column) to whose of another dataframe (2b) that return another special value (2a). The cells whose coordinates do not match the coordinates of the dataframe (2b), another special value can be written (3a) if not set to NA.

Usage

```
nestr_datf2(inptf_datf, rtn_pos, rtn_neg = NA, nestr_datf, yes_val = T)
```

Arguments

```
inptf_datf is the input dataframe (1b)
rtn_pos is the special value (1a)
rtn_neg is the special value (3a)
nestr_datf is the dataframe (2b)
yes_val is the special value (2a)
```

paste_datf 13

```
#3 yes yes
```

Description

Return a vector composed of pasted elements from the input dataframe at the same index.

Usage

```
paste_datf(inpt_datf, sep = "")
```

Arguments

```
inpt_datf is the input dataframe
sep is the separator between pasted elements, defaults to ""
```

Examples

```
print(paste_datf(inpt_datf=data.frame(c(1, 2, 1), c(33, 22, 55))))
#[1] "133" "222" "155"
```

```
see_datf
```

see_datf

Description

Allow to return a dataframe with special value cells (ex: TRUE) where the condition entered are respected and another special value cell (ex: FALSE) where these are not

Usage

```
see_datf(
  datf,
  condition_l,
  val_l,
  conjunction_l = c(),
  rt_val = TRUE,
  f_val = FALSE
)
```

14 see_datf

Arguments

datf is the input dataframe is the vector of the possible conditions ("==", ">", "<", "!=", "%%", "reg", condition_l "not_reg", "sup_nchar", "inf_nchar", "nchar") (equal to some elements in a vector, greater than, lower than, not equal to, is divisible by, the regex condition returns TRUE, the regex condition returns FALSE, the length of the elements is strictly superior to X, the length of the element is strictly inferior to X, the length of the element is equal to one element in a vector), you can put the same condition n times. val_l is the list of vectors containing the values or vector of values related to condition_l (so the vector of values has to be placed in the same order) conjunction_l contains the and or conjunctions, so if the length of condition_1 is equal to 3, there will be 2 conjunctions. If the length of conjunction_l is inferior to the length of condition_1 minus 1, conjunction_1 will match its goal length value with its last argument as the last arguments. For example, c("&", "I", "&") with a goal length value of $5 \rightarrow c("\&", "|", "\&", "\&", "\&")$ is a special value cell returned when the conditions are respected rt_val f_val is a special value cell returned when the conditions are not respected

Details

This function will return an error if number only comparative conditions are given in addition to having character values in the input dataframe.

```
datf1 <- data.frame(c(1, 2, 4), c("a", "a", "zu"))</pre>
print(see_datf(datf=datf1, condition_l=c("nchar"), val_l=list(c(1))))
     X1
           X2
#1 TRUE TRUE
#2 TRUE TRUE
#3 TRUE FALSE
print(see_datf(datf=datf1, condition_l=c("=="), val_l=list(c("a", 1))))
    X1
           X2
#1
   TRUE TRUE
#2 FALSE
          TRUE
#3 FALSE FALSE
print(see_datf(datf=datf1, condition_l=c("nchar"), val_l=list(c(1, 2))))
          X2
     X1
#1 TRUE TRUE
#2 TRUE TRUE
#3 TRUE TRUE
print(see_datf(datf=datf1, condition_l=c("not_reg"), val_l=list("[a-z]")))
```

swipr 15

```
# X1 X2
#1 TRUE FALSE
#2 TRUE FALSE
#3 TRUE FALSE
```

```
swipr swipr
```

Description

Returns an ordered dataframes according to the elements order given. The input datafram has two columns, one with the ids which can be bonded to multiple elements in the other column.

Usage

```
swipr(inpt_datf, how_to = c(), id_w = 2, id_ids = 1)
```

Arguments

```
    inpt_datf
    is the input dataframe
    how_to
    is a vector containing the elements in the order wanted
    id_w
    is the column number or the column name of the elements
    id_ids
    is the column number or the column name of the ids
```

Examples

9

10

Arm

Al

В

```
datf <- data.frame("col1"=c("Af", "Al", "Al", "Al", "Arg", "Arg", "Arg", "Arm", "Arm",
                                                                    "col2"=c("B", "B", "G", "S", "B", "S", "G", "B", "G", "B"))
print(swipr(inpt_datf=datf, how_to=c("G", "S", "B")))
                         col1 col2
 1
                                       Αf
                                                                                     В
 2
                                        Αl
                                                                                            G
 3
                                       Al
                                                                                           S
 4
                                       Al
                                                                                           В
 5
                                                                                         G
                                Arg
 6
                                Arg
                                                                                         S
 7
                                                                                        В
                                Arg
 8
                                                                                         G
                                Arm
```

16 unique_datf

unique_datf unique_datf

Description

Returns the input dataframe with the unique columns or rows.

Usage

```
unique_datf(inpt_datf, col = FALSE)
```

Arguments

inpt_datf is the input dataframe
col is a parameter that specifies if the dataframe returned should have unique columns
or rows, defaults to F, so the dataframe returned by default has unique rows

```
datf1 <- data.frame(c(1, 2, 1, 3), c("a", "z", "a", "p"))</pre>
print(datf1)
  c.1..2..1..3. c..a...z...a...p.. c.1..2..1..3..1
1
            1
                                 а
             2
3
             1
             3
4
                                                  3
print (unique_datf(inpt_datf=datf1))
#
   c.1..2..1..3. c..a...z...a...p..
#1
          1
#2
              2
datf1 \leftarrow data.frame(c(1, 2, 1, 3), c("a", "z", "a", "p"), c(1, 2, 1, 3))
print(datf1)
  c.1..2..1..3. c..a...z...a...p..
1
       1
2
             2
3
             1
                                   а
print(unique_datf(inpt_datf=datf1, col=TRUE))
# cur_v cur_v
#1
     1 a
      2
#2
           Z
#3
      1
#4
      3
```

val_replacer 17

```
val_replacer val_replacer
```

Description

Allow to replace value from dataframe to another one.

Usage

```
val_replacer(datf, val_replaced, val_replacor = TRUE)
```

Arguments

```
datf is the input dataframe
val_replaced is a vector of the value(s) to be replaced
val_replacor is the value that will replace val_replaced
```

Examples

Description

Allow to get if a vector is in a dataframe. Returns the row and column of the vector in the dataframe if the vector is contained in the dataframe.

Usage

```
vec_in_datf(
  inpt_datf,
  inpt_vec = c(),
  coeff = 0,
  stop_untl = 1,
  conventional = FALSE
)
```

vec_in_datf

Arguments

```
inpt_datf is the input dataframe
inpt_vec is the vector that may be in the input dataframe
coeff is the "slope coefficient" of inpt_vec
stop_untl is the maximum number of the input vector the function returns, if in the dataframe
conventional is if a positive slope coefficient means that the vector goes upward or downward
```

```
datf1 <- data.frame(c(1:5), c(5:1), c("a", "z", "z", "z", "a"))</pre>
print(datf1)
# c.1.5. c.5.1. c..a...z...z...z.....
#1
      1
             5
#2
       2
             4
             3
#3
       3
#4
       4
             2
       5
#5
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(5, 4, "z"), coeff=1))
#NULL
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(5, 2, "z"), coeff=1))
#[1] 5 1
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(3, "z"), coeff=1))
#[1] 3 2
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(4, "z"), coeff=-1))
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(2, 3, "z"), coeff=-1))
#[1] 2 1
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(5, 2, "z"), coeff=-1, conventional=TRUE))
#[1] 5 1
datf1[4, 2] <- 1
print(vec_in_datf(inpt_datf=datf1, inpt_vec=c(1, "z"), coeff=-1, conventional=TRUE, stop_
#[1] 4 2 5 2
```

vlookup_datf

vlookup_datf	vlookup_	_datf
--------------	----------	-------

Description

Alow to perform a vlookup on a dataframe

Usage

```
vlookup_datf(datf, v_id, col_id = 1, included_col_id = "yes")
```

Arguments

```
datf is the input dataframe

v_id is a vector containing the ids

col_id is the column that contains the ids (default is equal to 1)

included_col_id

is if the result should return the col_id (default set to yes)
```

Examples

```
datf1 <- data.frame(c("az1", "az3", "az4", "az2"), c(1:4), c(4:1))</pre>
print(vlookup_datf(datf=datf1, v_id=c("az1", "az2", "az3", "az4")))
    c..az1....az3....az4....az2.. c.1.4. c.4.1.
#2
                               az1
#4
                               az2
                                        4
                                               1
#21
                               az3
                                        2
                                               3
                                        3
                                               2
#3
                               az4
```

```
wider_datf wider_datf
```

Description

Takes a dataframe as an input and the column to split according to a seprator.

Usage

```
wider_datf(inpt_datf, col_to_splt = c(), sep_ = "-")
```

Arguments

inpt_datf	is the input dataframe
col_to_splt	is a vector containing the number or the colnames of the columns to split according to a separator
sep	is the separator of the elements to split to new columns in the input dataframe

20 wider_datf

Index

```
colins_datf, 2
\verb|cumulated_rows|, 3
cumulated_rows_na, 3
cut_v,4
diff_datf, 5
groupr_datf,5
historic_sequence, 6
id_keepr, 8
insert\_datf, 9
\verb|intersect_mod, 10|\\
nestr_datf1, 11
\verb"nestr_datf2", 12
paste_datf, 13
see_datf, 13
swipr, 15
{\tt unique\_datf}, 16
val_replacer, 17
vec_in_datf, 17
vlookup_datf, 19
{\tt wider\_datf}, 19
```