

effectifs_Cleaning

Manon Santrisse

11/10/2022

```
library(dplyr)

##
## Attachement du package : 'dplyr'
## Les objets suivants sont masqués depuis 'package:stats':
##
##   filter, lag
## Les objets suivants sont masqués depuis 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(corrplot)

## corrplot 0.92 loaded

library(readxl)
library(FactoMineR)
library(factoextra)

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(cowplot)
library("NbClust")
library(missForest)

## Le chargement a nécessité le package : randomForest
## randomForest 4.7-1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attachement du package : 'randomForest'
## L'objet suivant est masqué depuis 'package:ggplot2':
##
##   margin
## L'objet suivant est masqué depuis 'package:dplyr':
##
##   combine

## Le chargement a nécessité le package : foreach
## Le chargement a nécessité le package : itertools
## Le chargement a nécessité le package : iterators
```

```
library(writexl)
```

```
## Warning: le package 'writexl' a été compilé avec la version R 4.1.3
```

Récupération des données Effectifs entreprises

```
data=read_excel("Donnees/Effectifs_cleaning.xlsx")
data
```

```
## # A tibble: 287,976 x 33
##   `Num Ctr Coll Anonyme` `Lib Entreprise An` `Code Grp Assu` `Condition Ven`
##   <chr>                  <chr>              <chr>          <chr>
## 1 1-coll                1-Lib Entreprise    NCA            NCA
## 2 1-coll                1-Lib Entreprise    NCA            NCA
## 3 1-coll                1-Lib Entreprise    NCA            NCA
## 4 1-coll                1-Lib Entreprise    NCA            NCA
## 5 1-coll                1-Lib Entreprise    NCA            NCA
## 6 1-coll                1-Lib Entreprise    NCA            NCA
## 7 1-coll                1-Lib Entreprise    NCA            NCA
## 8 1-coll                1-Lib Entreprise    NCA            NCA
## 9 1-coll                1-Lib Entreprise    NCA            NCA
## 10 1-coll              1-Lib Entreprise    NCA            NCA
## # ... with 287,966 more rows, and 29 more variables: `Produit Anonyme` <chr>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <chr>,
## #   Departement <chr>, REGROUP_PROD_1 <chr>, REGROUP_PROD_2 <chr>,
## #   REGROUP_PROD_3 <chr>, `REGROUP_PROD_4 Anonyme` <chr>, REGROUP_PROD_5 <chr>,
## #   `Num Personne Anonyme` <chr>, `Num Ctr Indiv Anonyme` <chr>,
## #   `Date Effet Adhesion Num Personne` <dtm>, ...
```

Renégo oui (1) /non (0)

Transformé les valeurs Na en non (0)

```
for (i in 1:nrow(data)){
  if (is.na(data$`Renégo 2020`[i])){
    data$`Renégo 2020`[i]=0
  }
  if (is.na(data$`Renégo 2021`[i])){
    data$`Renégo 2021`[i]=0
  }
  if (is.na(data$`Renégo 2022`[i])){
    data$`Renégo 2022`[i]=0
  }
}
data
```

```
## # A tibble: 287,976 x 33
##   `Num Ctr Coll Anonyme` `Lib Entreprise An` `Code Grp Assu` `Condition Ven`
##   <chr>                  <chr>              <chr>          <chr>
## 1 1-coll                1-Lib Entreprise    NCA            NCA
## 2 1-coll                1-Lib Entreprise    NCA            NCA
## 3 1-coll                1-Lib Entreprise    NCA            NCA
```

```

## 4 1-coll          1-Lib Entreprise      NCA          NCA
## 5 1-coll          1-Lib Entreprise      NCA          NCA
## 6 1-coll          1-Lib Entreprise      NCA          NCA
## 7 1-coll          1-Lib Entreprise      NCA          NCA
## 8 1-coll          1-Lib Entreprise      NCA          NCA
## 9 1-coll          1-Lib Entreprise      NCA          NCA
## 10 1-coll         1-Lib Entreprise      NCA          NCA
## # ... with 287,966 more rows, and 29 more variables: `Produit Anonyme` <chr>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <chr>,
## #   Departement <chr>, REGROUP_PROD_1 <chr>, REGROUP_PROD_2 <chr>,
## #   REGROUP_PROD_3 <chr>, `REGROUP_PROD_4 Anonyme` <chr>, REGROUP_PROD_5 <chr>,
## #   `Num Personne Anonyme` <chr>, `Num Ctr Indiv Anonyme` <chr>,
## #   `Date Effet Adhesion Num Personne` <dtm>, ...
write_xlsx(data, './Donnees/Effectifs_freeze.xlsx')

```