

Effectifs_freeze

Manon Santrisse

13/10/2022

Récupération du jeu Effectifs.xlsx

```
data=read_excel('../Donnees/Effectifs_freeze.xlsx')
head(data)

## # A tibble: 6 x 33
##   Num Ctr Coll~1 Lib E~2 Code ~3 Condi~4 Produ~5 Date ~6 Date ~7 Code ~8 Depar~9
##   <chr>          <chr>  <chr>  <chr>  <chr>      <dbl>  <dbl> <chr>  <chr>
## 1 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## 2 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## 3 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## 4 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## 5 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## 6 1-coll          1-Lib ~ NCA    NCA    1-Prod~  2.02e7  2.05e7 2932Z  @
## # ... with 24 more variables: 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>,
## #   `Date Effet Radiation Num Personne` <dbl>, `Type Assure` <chr>, Sexe <chr>,
## #   `Date Naissance` <dtm>, `R/NR` <chr>, `Lien entreprise Anonyme` <chr>,
## #   `Numéro contrat coll Grands Comptes` <chr>, `Indexation 2018` <dbl>, ...
```

Formatage

Simplifications des valeurs et typologie des variables

```
## # A tibble: 6 x 33
##   Num Ctr Coll~1 Lib E~2 Code ~3 Condi~4 Produ~5 Date ~6 Date ~7 Code ~8 Depar~9
##   <int>  <int> <fct>  <fct>      <int>  <dbl>  <dbl> <fct>  <fct>
## 1      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## 2      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## 3      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## 4      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## 5      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## 6      1      1 NCA    NCA          1  2.02e7  2.05e7 2932Z  @
## # ... with 24 more variables: REGROUP_PROD_1 <fct>, REGROUP_PROD_2 <fct>,
## #   REGROUP_PROD_3 <fct>, `REGROUP_PROD_4 Anonyme` <fct>, REGROUP_PROD_5 <fct>,
## #   `Num Personne Anonyme` <int>, `Num Ctr Indiv Anonyme` <int>,
## #   `Date Effet Adhesion Num Personne` <dtm>,
## #   `Date Effet Radiation Num Personne` <dbl>, `Type Assure` <fct>, Sexe <fct>,
## #   `Date Naissance` <dtm>, `R/NR` <fct>, `Lien entreprise Anonyme` <int>,
```

```
## # `Numéro contrat coll Grands Comptes` <fct>, `Indexation 2018` <chr>, ...
```

Il y a une seule catégorie pour regroup_prod 2 et 3 donc on ne va pas les étudier.

Gestion du format date

```
## Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Condition.Vente
## 1 1 1 1 NCA NCA
## 2 1 1 1 NCA NCA
## 3 1 1 1 NCA NCA
## 4 1 1 1 NCA NCA
## 5 1 1 1 NCA NCA
## 6 1 1 1 NCA NCA
## Produit.Anonyme Date.Effet.Adhesion.Contrat.Coll
## 1 1 2022-01-01
## 2 1 2022-01-01
## 3 1 2022-01-01
## 4 1 2022-01-01
## 5 1 2022-01-01
## 6 1 2022-01-01
## Date.Effet.Radiation.Contrat.Coll Code.Ape Departement REGROUP_PROD_1
## 1 2050-12-31 2932Z @ 1_ERCAC
## 2 2050-12-31 2932Z @ 1_ERCAC
## 3 2050-12-31 2932Z @ 1_ERCAC
## 4 2050-12-31 2932Z @ 1_ERCAC
## 5 2050-12-31 2932Z @ 1_ERCAC
## 6 2050-12-31 2932Z @ 1_ERCAC
## REGROUP_PROD_4.Anonyme REGROUP_PROD_5 Num.Personne.Anonyme
## 1 1-Regroup_Prod_4 5_BASE 1
## 2 1-Regroup_Prod_4 5_BASE 2
## 3 1-Regroup_Prod_4 5_BASE 3
## 4 1-Regroup_Prod_4 5_BASE 4
## 5 1-Regroup_Prod_4 5_BASE 5
## 6 1-Regroup_Prod_4 5_BASE 6
## Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 1 1 2022-01-01
## 2 2 2022-01-01
## 3 3 2022-01-01
## 4 4 2022-01-01
## 5 5 2022-01-01
## 6 6 2022-01-01
## Date.Effet.Radiation.Num.Personne Type.Assure Sexe Date.Naissance R.NR
## 1 2050-12-31 ENFANT F 2005-03-02 R
## 2 2050-12-31 ENFANT F 2010-05-21 R
## 3 2050-12-31 ENFANT M 1999-01-08 R
## 4 2050-12-31 CONJOI F 1991-08-27 R
## 5 2050-12-31 ENFANT F 2017-12-02 R
## 6 2050-12-31 ASSPRI M 1980-10-10 R
## Lien.entreprise.Anonyme Numéro.contrat.coll.Grands.Comptes Indexation.2018
## 1 1 Grands comptes <NA>
## 2 1 Grands comptes <NA>
## 3 1 Grands comptes <NA>
## 4 1 Grands comptes <NA>
## 5 1 Grands comptes <NA>
## 6 1 Grands comptes <NA>
```

```
##      Indexation.2019 Indexation.2020 Indexation.2021 Indexation.2022
## 1              <NA>              <NA>              <NA>              <NA>
## 2              <NA>              <NA>              <NA>              <NA>
## 3              <NA>              <NA>              <NA>              <NA>
## 4              <NA>              <NA>              <NA>              <NA>
## 5              <NA>              <NA>              <NA>              <NA>
## 6              <NA>              <NA>              <NA>              <NA>
##      Indexation.2023 Renégo.2020 Renégo.2021 Renégo.2022
## 1              6.0%          FALSE          FALSE          FALSE
## 2              6.0%          FALSE          FALSE          FALSE
## 3              6.0%          FALSE          FALSE          FALSE
## 4              6.0%          FALSE          FALSE          FALSE
## 5              6.0%          FALSE          FALSE          FALSE
## 6              6.0%          FALSE          FALSE          FALSE
```

Racourcis noms variables

```
Num_E=data$Num.Ctr.Coll.Anonyme
Nom_E=data$Lib.Entreprise.Anonyme
Gp_assures=data$Code.Grp.Assures
cond_vente=data$Condition.Vente
Prod=data$Produit.Anonyme
Date_adh_coll=data$Date.Effet.Adhesion.Contrat.Coll
Date_rad_coll=data$Date.Effet.Radiation.Contrat.Coll
Secteur=data$Code.Ape
REGROUP_PROD_4=data$REGROUP_PROD_4.Anonyme
Num_P=data$Num.Personne.Anonyme
Num_Fam=data$Num.Ctr.Indiv.Anonyme
Date_adh_pers=data$Date.Effet.Adhesion.Num.Personne
Date_rad_pers=data$Date.Effet.Radiation.Num.Personne
Type=data$Type.Assure
Date_Naissance=data$Date.Naissance
Lien=data$Lien.entreprise.Anonyme
Index2018=data$Indexation.2018
Index2019=data$Indexation.2019
Index2020=data$Indexation.2020
Index2021=data$Indexation.2021
Index2022=data$Indexation.2022
Index2023=data$Indexation.2023
Renego2020=data$Renégo.2020
Renego2021=data$Renégo.2021
Renego2022=data$Renégo.2022
```

Summary

Summary pour voir la tendance de répartition des variables.

```
summary(data)
```

```
##      Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Condition.Vente
## Min.      : 1           Min.      : 1           ACT      :125100 COL      :120616
## 1st Qu.: 909           1st Qu.: 884           RET      : 22189 ACT      : 84885
## Median :1512           Median :1458           ACT1     : 21997 RET      : 10426
## Mean    :1709           Mean    :1645           PORT     : 18730 ACT1     :  8843
```

```

## 3rd Qu.:2437          3rd Qu.:2364          NCA      : 12554  ACC      : 8189
## Max.      :3843          Max.      :3716          ACT4     : 9363  PORT     : 7754
##                                     (Other): 78043  (Other): 47263
## Produit.Anonyme Date.Effet.Adhesion.Contrat.Coll
## Min.      : 1      Min.      :1900-01-01
## 1st Qu.   : 641    1st Qu.   :2013-01-01
## Median    : 972    Median    :2016-01-01
## Mean      :1216    Mean      :2014-07-22
## 3rd Qu.   :1915    3rd Qu.   :2019-01-01
## Max.      :2927    Max.      :2022-10-01
##
## Date.Effet.Radiation.Contrat.Coll      Code.Ape      Departement
## Min.      :2023-01-01      @      :102399      @      :185809
## 1st Qu.   :2050-12-31      7010Z   : 16481      12      : 33457
## Median    :2050-12-31      3030Z   : 13857      66      : 14080
## Mean      :2050-12-19      6512Z   : 8877      46      : 11989
## 3rd Qu.   :2050-12-31      8810A   : 8128      11      : 9318
## Max.      :2050-12-31      8899B   : 7852      15      : 8712
##                                     (Other):130382  (Other): 24611
## REGROUP_PROD_1      REGROUP_PROD_4.Anonyme      REGROUP_PROD_5
## 1_ERCAC:180099      27-Regroup_Prod_4: 58043      5_BASE      :254674
## 1_ER24 : 25388      5-Regroup_Prod_4 : 36581      5_SURCOMP   : 19214
## 1_ER11 : 23335      19-Regroup_Prod_4: 19895      5_BASE_SURCOMP: 6800
## 1_CPM  : 22433      11-Regroup_Prod_4: 10666      5_OPTION_PH15 : 4689
## 1_ER66 : 12470      51-Regroup_Prod_4: 8656      5_OPTION_Sante: 1360
## 1_ER76 : 10295      39-Regroup_Prod_4: 7629      5_OPTION_IJ  : 1063
## (Other): 13956      (Other)      :146506      (Other)      : 176
## Num.Personne.Anonyme Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## Min.      : 1      Min.      : 1      Min.      :1958-10-01
## 1st Qu.   : 35706    1st Qu.   :24920    1st Qu.   :2015-12-01
## Median    : 67550    Median    :45058    Median    :2018-05-01
## Mean      : 68847    Mean      :45289    Mean      :2017-03-16
## 3rd Qu.   :101925    3rd Qu.   :66757    3rd Qu.   :2021-01-01
## Max.      :135458    Max.      :87197    Max.      :2023-04-01
##
## Date.Effet.Radiation.Num.Personne Type.Assure      Sexe
## Min.      :2018-01-01      ASSPRI:172432      F:150628
## 1st Qu.   :2021-01-01      AUTRE : 24      I: 96
## Median    :2050-12-31      CONJOI: 46898      M:137252
## Mean      :2037-08-04      ENFANT: 68622
## 3rd Qu.   :2050-12-31
## Max.      :2050-12-31
##
## Date.Naissance      R.NR      Lien.entreprise.Anonyme
## Min.      :1913-02-22      NR: 15188      Min.      : 1.0
## 1st Qu.   :1959-05-02      R :272788      1st Qu.   : 72.0
## Median    :1977-08-04      Median    : 299.0
## Mean      :1977-12-23      Mean      : 382.7
## 3rd Qu.   :1998-03-19      3rd Qu.   : 645.0
## Max.      :2022-09-13      Max.      :1008.0
##
## Numéro.contrat.coll.Grands.Comptes Indexation.2018      Indexation.2019
## Grands comptes: 89253      Length:287976      Length:287976
## NA's      :198723      Class :character      Class :character

```

```

##                               Mode :character   Mode :character
##
##
##
##
##  Indexation.2020      Indexation.2021      Indexation.2022      Indexation.2023
##  Length:287976      Length:287976      Length:287976      Length:287976
##  Class :character    Class :character    Class :character    Class :character
##  Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
##
##  Renégo.2020      Renégo.2021      Renégo.2022
##  Mode :logical    Mode :logical    Mode :logical
##  FALSE:277164    FALSE:269176    FALSE:275305
##  TRUE :10812     TRUE :18800     TRUE :12671
##
##
##
##

```

On relève au total :

- * 3843 entreprises
- * 2927 produits
- * 135458 personnes
- * 87197 familles
- * 1008 liens entreprises

Il y a de nombreuses catégories pour les variables :

- * Code.Grp.Assures * Condition.Vente * REGROUP_PROD_5. Nous allons rassembler certaines catégories ensembles.

Dans type assurés on retrouve une catégorie Autre, celle-ci doit correspondre aux ascendants. Nous allons les ranger avec les conjoints.

Dans Sexe, il y a une catégorie I avec 24 personnes, au vu du nombre peu élevé nous allons mettre 16 personnes de sexe féminin et 12 de sexe masculin.

Dans indexation 2019 et 2020 on relève un pourcentage négatif. En regardant à quelle entreprise et lien cela est lié on voit qu'il semble s'agir de CNN. Les CNN ont des contrats avec des particularités imposés par l'état ce que peut expliquer cette réduction d'indexation.

Il y a tout de même de nombreuses données manquantes dans les indexations entre 2018 et 2023 (voir pourcentage ci-dessous).

```

## [1] "Pourcentage de données manquantes pour indexation 2018 : 32.71"
## [1] "Pourcentage de données manquantes pour indexation 2019 : 23.43"
## [1] "Pourcentage de données manquantes pour indexation 2020 : 37.53"
## [1] "Pourcentage de données manquantes pour indexation 2021 : 32.92"
## [1] "Pourcentage de données manquantes pour indexation 2022 : 21.14"
## [1] "Pourcentage de données manquantes pour indexation 2023 : 4.95"

```

De plus quand nous n'avions pas de données pour les Renégociation nous avons défini la valeur à 'FALSE'.

Nettoyage

Code.Grp.Assures

Condition.Vente

Date.Effet.Adhesion.Contrat.Coll et Date.Effet.Radiation.Contrat.Coll

Code.Ape

Departement

REGROUP_PROD_1

REGROUP_PROD_4.Anonyme

REGROUP_PROD_5

Date.Effet.Adhesion.Num.Personne et Date.Effet.Radiation.Num.Personne

Type.Assure

Sexe

Date.Naissance