

Effectifs_freeze

Manon Santrisse

13/10/2022

Récupération du jeu Effectifs_freeze.xlsx

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

## # 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

```
head(effectif)

## # 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
##   <fct>          <fct>  <chr>  <chr>  <fct>    <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 <chr>,
## #   REGROUP_PROD_3 <chr>, `REGROUP_PROD_4 Anonyme` <fct>, REGROUP_PROD_5 <fct>,
## #   `Num Personne Anonyme` <fct>, `Num Ctr Indiv Anonyme` <fct>,
## #   `Date Effet Adhesion Num Personne` <dtm>,
```

```
## # `Date Effet Radiation Num Personne` <dbl>, `Type Assure` <chr>, Sexe <chr>,
## # `Date Naissance` <dtm>, `R/NR` <fct>, `Lien entreprise Anonyme` <fct>,
## # `Numéro contrat coll Grands comptes` <chr>, `Indexation 2018` <dbl>, ...
```

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

On ne s'intéressera pas à la variable 'Condition de Vente'.

Gestion du format date

```
## Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Produit.Anonyme
## 1 1 1 NCA 1
## 2 1 1 NCA 1
## 3 1 1 NCA 1
## 4 1 1 NCA 1
## 5 1 1 NCA 1
## 6 1 1 NCA 1
## Date.Effet.Adhesion.Contrat.Coll Date.Effet.Radiation.Contrat.Coll Code.Ape
## 1 2022-01-01 2050-12-31 2932Z
## 2 2022-01-01 2050-12-31 2932Z
## 3 2022-01-01 2050-12-31 2932Z
## 4 2022-01-01 2050-12-31 2932Z
## 5 2022-01-01 2050-12-31 2932Z
## 6 2022-01-01 2050-12-31 2932Z
## Departement REGROUP_PROD_1 REGROUP_PROD_4.Anonyme REGROUP_PROD_5
## 1 @ ERCAC 1 BASE
## 2 @ ERCAC 1 BASE
## 3 @ ERCAC 1 BASE
## 4 @ ERCAC 1 BASE
## 5 @ ERCAC 1 BASE
## 6 @ ERCAC 1 BASE
## Num.Personne.Anonyme Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 1 1 1 2022-01-01
## 2 2 2 2022-01-01
## 3 3 3 2022-01-01
## 4 4 4 2022-01-01
## 5 5 5 2022-01-01
## 6 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	NA	FALSE	FALSE	FALSE
## 2	0.06	FALSE	FALSE	FALSE
## 3	0.06	FALSE	FALSE	FALSE
## 4	0.06	FALSE	FALSE	FALSE
## 5	0.06	FALSE	FALSE	FALSE
## 6	0.06	FALSE	FALSE	FALSE

Racourcis noms variables

```

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

```

Summary

Summary pour voir la tendance de répartition des variables.

```
summary(effectif)
```

##	Num.Ctr.Coll.Anonyme	Lib.Entreprise.Anonyme	Code.Grp.Assures
##	909 : 15298	884 : 15298	Length:279307
##	1018 : 13852	989 : 13852	Class :character
##	1156 : 5234	1122 : 5234	Mode :character
##	1966 : 4892	1911 : 4892	
##	2565 : 4404	2505 : 4404	
##	76 : 4220	72 : 4220	
##	(Other):231407	(Other):231407	

```

## Produit.Anonyme Date.Effet.Adhesion.Contrat.Coll
## 641 : 10546 Min. :1900-01-01
## 856 : 9844 1st Qu.:2013-07-01
## 972 : 8485 Median :2016-01-01
## 855 : 5722 Mean :2014-08-16
## 1954 : 4844 3rd Qu.:2019-01-01
## 86 : 4310 Max. :2022-07-01
## (Other):235556
## Date.Effet.Radiation.Contrat.Coll Code.Ape Departement
## Min. :2023-01-01 @ : 94852 @ :178909
## 1st Qu.:2050-12-31 7010Z : 16472 12 : 33457
## Median :2050-12-31 3030Z : 13857 66 : 12327
## Mean :2050-12-18 6512Z : 8877 46 : 11989
## 3rd Qu.:2050-12-31 8810A : 8049 11 : 9318
## Max. :2050-12-31 8899B : 7851 15 : 8712
## (Other):129349 (Other): 24595
## REGROUP_PROD_1 REGROUP_PROD_4.Anonyme REGROUP_PROD_5
## ERCAC :177949 27 : 58040 BASE :248590
## ER11 : 22673 5 : 36489 SURCOMP : 18937
## CPM : 22415 19 : 19609 BASE_SURCOMP : 6742
## ER24 : 21504 11 : 10627 OPTION_PH15 : 3631
## ER66 : 12224 51 : 8656 OPTION_Sante : 1184
## ER80 : 8846 39 : 7551 OPTION_Assistance: 172
## (Other): 13696 (Other):138335 (Other) : 51
## Num.Personne.Anonyme Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 33163 : 38 36739 : 44 Min. :1958-10-01
## 35276 : 36 72573 : 42 1st Qu.:2016-01-01
## 33198 : 33 36733 : 36 Median :2018-07-01
## 32603 : 32 72574 : 35 Mean :2017-04-12
## 35343 : 32 36977 : 32 3rd Qu.:2021-01-01
## 32584 : 28 72709 : 31 Max. :2023-04-01
## (Other):279108 (Other):279087
## Date.Effet.Radiation.Num.Personne Type.Assure Sexe
## Min. :2018-01-01 Length:279307 Length:279307
## 1st Qu.:2021-01-01 Class :character Class :character
## Median :2050-12-31 Mode :character Mode :character
## Mean :2037-08-26
## 3rd Qu.:2050-12-31
## Max. :2050-12-31
##
## Date.Naissance R.NR Lien.entreprise.Anonyme
## Min. :1913-02-22 NR: 15083 109 : 19609
## 1st Qu.:1959-04-08 R :264224 340 : 16133
## Median :1977-10-28 404 : 13730
## Mean :1977-12-29 64 : 10627
## 3rd Qu.:1998-03-23 39 : 9213
## Max. :2022-09-13 282 : 7551
## (Other):202444
## Numéro.contrat.coll.Grands.comptes Indexation.2018 Indexation.2019
## Length:279307 Min. :0.00 Min. : -0.05
## Class :character 1st Qu.:0.00 1st Qu.: 0.00
## Mode :character Median :0.03 Median : 0.00
## Mean :0.03 Mean : 0.01
## 3rd Qu.:0.04 3rd Qu.: 0.00

```

```
##                               Max.    :0.30    Max.    : 0.20
##                               NA's     :86495   NA's     :59855
## Indexation.2020 Indexation.2021 Indexation.2022 Indexation.2023
## Min.    :-0.02   Min.    :0.00   Min.    :0.00   Min.    :0.000
## 1st Qu.: 0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.000
## Median : 0.00   Median :0.00   Median :0.02   Median :0.085
## Mean   : 0.01   Mean   :0.01   Mean   :0.02   Mean   :0.079
## 3rd Qu.: 0.01   3rd Qu.:0.01   3rd Qu.:0.03   3rd Qu.:0.125
## Max.    : 0.20   Max.    :0.10   Max.    :0.20   Max.    :0.250
## NA's    :100374  NA's    :87097  NA's    :53222  NA's    :6696
## Renégo.2020   Renégo.2021   Renégo.2022
## Mode :logical Mode :logical Mode :logical
## FALSE:268495  FALSE:260507  FALSE:266636
## 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 : 33.73"
## [1] "Pourcentage de données manquantes pour indexation 2019 : 24.16"
## [1] "Pourcentage de données manquantes pour indexation 2020 : 38.7"
## [1] "Pourcentage de données manquantes pour indexation 2021 : 33.94"
## [1] "Pourcentage de données manquantes pour indexation 2022 : 21.79"
## [1] "Pourcentage de données manquantes pour indexation 2023 : 5.11"
```

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

Correspondance Code Groupe assurés

On va utiliser le doc 'Recodage code groupe assure.xlsx' pour classer les codes groupes assurés

```
code_ass=read_excel('../Donnees/Recodage code groupe assure.xlsx')
head(code_ass)

## # A tibble: 6 x 2
##   `Code Grp Assures` `Code Grp Assures recodée`
##   <chr>              <chr>
## 1 ACC              NON_ACTIFS
## 2 ACC1             NON_ACTIFS
## 3 ACC10            NON_ACTIFS
## 4 ACC11            NON_ACTIFS
## 5 ACC13            NON_ACTIFS
## 6 ACC14            NON_ACTIFS

for (i in 1:nrow(effectif)){
  j=1
  while (effectif$Code.Grp.Assures[i]!=code_ass$`Code Grp Assures`[j]){
    j=j+1
  }
  effectif$Code.Grp.Assures[i]=code_ass$`Code Grp Assures recodée`[j]
}

effectif$Code.Grp.Assures<-as.factor(effectif$Code.Grp.Assures)
levels(effectif$Code.Grp.Assures)

## [1] "ACTIFS"      "NON_ACTIFS"  "PORTABILITE"
```

Type assuré

Autre <- conjoint

```
for (i in 1:nrow(effectif)){
  if (effectif$Type.Assure[i]=='AUTRE'){
    effectif$Type.Assure[i]='CONJOI'
  }
}

effectif$Type.Assure<-as.factor(effectif$Type.Assure)
effectif=droplevels(effectif)
```

Sexe

I<- moitié M et moitié F

```
for (i in 1:nrow(effectif)){
  cpt=1
  if (effectif$Sexe[i]=='I' & (cpt %% 2) == 0){
    effectif$Sexe[i]='M'
  }else if (effectif$Sexe[i]=='I' & (cpt %% 2) != 0){
    effectif$Sexe[i]='F'
  }
}

effectif$Sexe<-as.factor(effectif$Sexe)
effectif=droplevels(effectif)
```

Regroupe levels regroup 5

Faire 2 catégories : base et le reste. Puis faudra les mettre en pourcentage d'effectif après.

```
levels(effectif$REGROUP_PROD_5)[3]="BASE"
levels(effectif$REGROUP_PROD_5)[c(1,3:9)]="OPTION"
summary(effectif$REGROUP_PROD_5)
```

```
## OPTION    BASE
## 23975 255332
```

Grands comptes

```
for (i in 1:nrow(effectif)){
  if (!is.na(effectif$Numéro.contrat.coll.Grands.comptes[i])){
    effectif$Numéro.contrat.coll.Grands.comptes[i]=TRUE
  }else{
    effectif$Numéro.contrat.coll.Grands.comptes[i]=FALSE
  }
}

effectif$Numéro.contrat.coll.Grands.comptes<-as.logical(effectif$Numéro.contrat.coll.Grands.comptes)
```

Code APE & Département

```
for (i in 1:nrow(effectif)){
  if (effectif$Code.Ape[i]=='@'){
    effectif$Code.Ape[i]=NA
  }
  if(effectif$Departement[i]=='@'){
    effectif$Departement[i]=NA
  }
}

effectif=droplevels(effectif)
```

Save en Effectif.Rdata

```
save(effectif, file='../Donnees/Effectifs_freeze.Rdata')
```