

Data_exploration_and_dataset_preparation

September 4, 2018

1 CDV study - data exploration & dataset preparation

```
In [1]: from pathlib import Path
import pandas as pd
import numpy as np
from datetime import datetime
import time
import matplotlib.pyplot as plt
%matplotlib inline
import pickle
#%%pylab inline

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score, GridSearchCV
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, f1_score, precision_score, recall_score
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import LabelBinarizer
from sklearn.preprocessing import OneHotEncoder
from sklearn.svm import SVC
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV, RFE
from sklearn.utils import resample

In [2]: path_project = Path.home() / Path('Google Drive/Felix')
path_data = path_project / Path("data")
path_dump = path_project / Path("dump")
```

1.1 I - CDV study - Data exploration

```
In [3]: # loading cdv data
file = path_data / Path("felix.csv")
with Path.open(file, 'rb') as fp:
    cdv = pd.read_csv(fp, encoding='cp1252', low_memory=False)
```

```
In [4]: cdv.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11131 entries, 0 to 11130
Columns: 354 entries, INTER6 to an_nais
dtypes: float64(47), int64(15), object(292)
memory usage: 30.1+ MB
```

```
In [5]: # loading cdv data without format
file = path_data / Path("felix_ssfmt.csv")
with Path.open(file, 'rb') as fp:
    cdv_ssfmt = pd.read_csv(fp, encoding='cp1252', low_memory=False)
```

```
In [6]: cdv_ssfmt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11131 entries, 0 to 11130
Columns: 354 entries, INTER6 to an_nais
dtypes: float64(153), int64(190), object(11)
memory usage: 30.1+ MB
```

1.2 1) Dataset Size and missing values analysis

1.2.1 a) Differences between "cdv.csv" & "cdv_ssfmt.csv"

```
In [7]: print(f"Number of different columns between 'cdv.csv' & 'cdv_ssfmt.csv' : \
{np.sum(cdv_ssfmt.columns != cdv.columns)}")
```

```
Number of different columns between 'cdv.csv' & 'cdv_ssfmt.csv' : 0
```

```
In [8]: # insight on exemple
A = cdv.loc[:, ["INTER6", "INTER", "ANNEEFUZ", "ANNEFUZ2", "AGGLO9",
               "DIPLOME", "COLLECTE", "CHAMP", "POND"]].dropna()
B = cdv_ssfmt.loc[:, ["INTER6", "INTER", "ANNEEFUZ", "ANNEFUZ2", "AGGLO9",
                     "DIPLOME", "COLLECTE", "CHAMP", "POND"]].dropna()
```

```
In [9]: A.head()
```

```
Out[9]:
```

	INTER6	INTER	ANNEEFUZ	ANNEFUZ2	AGGLO9	\
0	373001	3001	2015	2015 online	5 000 à 10 000 habitants	
1	373002	3002	2015	2015 online	2 000 à 5 000 habitants	
2	373003	3003	2015	2015 online	Agglomération parisienne	
3	373004	3004	2015	2015 online	[Nsp]	
4	373005	3005	2015	2015 online	2 000 à 5 000 habitants	

	DIPLOME	COLLECTE	CHAMP	POND
0	BEPC, brevet des collèges	Online	18 ans et + métropole	1.313554

1	BEPC, brevet des collèges	Online	18 ans et + métropole	2.009015
2	Bac + 3 et supérieur	Online	18 ans et + métropole	0.217607
3	Bac + 3 et supérieur	Online	18 ans et + métropole	0.539351
4	Bac + 3 et supérieur	Online	18 ans et + métropole	0.270204

```
In [10]: B.head()
```

```
Out[10]:
```

	INTER6	INTER	ANNEEFUZ	ANNEFUZ2	AGGLO9	DIPLOME	COLLECTE	CHAMP	\
0	373001	3001	37	371	3	3	2	1	
1	373002	3002	37	371	2	3	2	1	
2	373003	3003	37	371	9	7	2	1	
3	373004	3004	37	371	10	7	2	1	
4	373005	3005	37	371	2	7	2	1	


```

POND
0  1.313554
1  2.009015
2  0.217607
3  0.539351
4  0.270204

```

```
In [11]: diff = np.sum(cdv != cdv_ssfmt)
```

Variable without any différences

```
In [12]: len(diff[diff == 0])
```

```
Out[12]: 17
```

```
In [13]: diff[diff == 0]
```

```
Out[13]:
```

INTER6	0
INTER	0
POND	0
DPT	0
COMINSEE	0
AGE	0
NB0003	0
NB0306	0
NB0610	0
NB1016	0
NB1620	0
NB2099	0
REVENQ	0
AUTREREV	0
REVTOT6	0
an_enq	0
an_nais	0

```
dtype: int64
```

```
In [14]: cdv[diff[diff == 0].index].dtypes
```

```
Out[14]: INTER6      int64
INTER      int64
POND       float64
DPT        object
COMINSEE   object
AGE        int64
NB0003     int64
NB0306     int64
NB0610     int64
NB1016     int64
NB1620     int64
NB2099     int64
REVENQ     int64
AUTREREV   int64
REVTOT6    int64
an_enq     int64
an_nais    int64
dtype: object
```

```
In [15]: cdv[diff[diff == 0].index].head()
```

```
Out[15]:
```

	INTER6	INTER	POND	DPT	COMINSEE	AGE	NB0003	NB0306	NB0610	NB1016	\
0	373001	3001	1.313554	29	29027	56	0	0	0	0	
1	373002	3002	2.009015	46	46137	64	0	0	0	0	
2	373003	3003	0.217607	34	34172	71	0	0	0	0	
3	373004	3004	0.539351	91	91021	21	0	0	0	0	
4	373005	3005	0.270204	23	23021	70	0	0	0	0	

	NB1620	NB2099	REVENQ	AUTREREV	REVTOT6	an_enq	an_nais
0	0	0	1100	0	2	2015	1959
1	0	0	2000	0	4	2015	1951
2	0	0	5000	0	5	2015	1944
3	0	0	200	0	1	2015	1994
4	0	0	998	0	3	2015	1945

```
In [16]: scope_identical = diff[diff == 0].index
```

Variable with différences

```
In [17]: diff[diff != 0]
```

```
Out[17]: ANNEEFUZ      11131
ANNEFUZ2      11131
COLLECTE      11131
CHAMP         11131
identifiant    8115
SEXE          11131
```

AGE5	11131
PCSENQ8	11131
TYPOSQT	11131
DIPL4	11131
AGGLO5	11131
UDA10	11131
SITUEMP3	11131
AGEDIP2	11131
AGGLO9	11131
DIPLOME	11131
EXERCPR0	11131
SITUEMP	11131
SITUEMP5	11131
SITUEMP6	11131
SALCOMP	11131
TYPEEMPL	11131
INTERIM	11131
TYPCONT	11131
TEMPSTRA	11131
nbheures	5242
NBHEUR39	11131
NBHEUR35	11131
PREFPALI	11131
SALCOMPI	11131
...	
poppeud	5134
poptrpeu	5134
agedip	8115
AUTREAL	11131
age_OW	8115
UDA5	11131
CSP6	8115
CP	8115
TYPLOG	8115
inseel	2259
inseenum	8115
couple2	11131
cpt	8115
AGE6	11131
PCSRED10	11131
prescaf	11130
refus2	8115
info	8115
med	8232
i	8115
com	8115
type99	5095
AGGLOINS	2045

DEPCOM	5134
pmun	5134
QUOTAAGE	11131
PRIVPUB	11131
interim2	11131
EMP7	11131
typcont2	11131

Length: 337, dtype: int64

In [18]: `cdv[diff[diff != 0].index].dtypes`

Out[18]:

ANNEEFUZ	int64
ANNEFUZ2	object
COLLECTE	object
CHAMP	object
identifiant	object
SEXE	object
AGE5	object
PCSENQ8	object
TYPOSQT	object
DIPL4	object
AGGL05	object
UDA10	object
SITUEMP3	object
AGEDIP2	object
AGGL09	object
DIPL0ME	object
EXERCPRO	object
SITUEMP	object
SITUEMP5	object
SITUEMP6	object
SALCOMP	object
TYPEEMPL	object
INTERIM	object
TYPCONT	object
TEMPSTRA	object
nbheures	float64
NBHEUR39	object
NBHEUR35	object
PREFPALI	object
SALCOMPI	object
...	
poppeud	float64
poptrpeu	float64
agedip	float64
AUTREAL	object
age_OW	float64
UDA5	object

CSP6	float64
CP	float64
TYPL0G	float64
inseel	object
inseenum	float64
couple2	object
cpt	float64
AGE6	object
PCSRED10	object
prescaf	float64
refus2	float64
info	object
med	float64
i	float64
com	object
type99	object
AGGLOINS	float64
DEPCOM	object
pmun	float64
QUOTAAGE	object
PRIVPUB	object
interim2	object
EMP7	object
typcont2	object
Length: 337, dtype: object	

```
In [19]: scope_diff = diff[diff != 0].index
```

Understanding of the differences

```
In [20]: typ = cdv[scope_diff].dtypes
scope_obj = typ[typ=='object'].index
scope_notobj = typ[typ!='object'].index
print(f"Number variable with difference having 'object' as type :\
{len(scope_obj)}")
print(f"Number variable with difference having other type :\
{len(scope_notobj)}")
```

Number variable with difference having 'object' as type :290

Number variable with difference having other type :47

1) Zoom in variable with 'object' type

```
In [21]: scope_obj_long = []
scope_obj_diff = []
scope_no_diff = []
for c in scope_obj:
    cdv_c_dropna = cdv[c].dropna()
```

```

cdv_ssfmt_c_dropna = cdv_ssfmt[c].dropna()
if np.sum(cdv_c_dropna != cdv_ssfmt_c_dropna) > 0:
    cdv_valct = cdv[c].value_counts()
    cdv_ssfmt_valct = cdv_ssfmt[c].value_counts()
    if len(cdv_valct) > 15:
        print("XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX")
        print(f"!!!!!!! {len(cdv_ssfmt_valct)} modalités !!!!!!!!")
        print("XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX")
        scope_obj_long.append(c)
    scope_obj_diff.append(c)
    print(f"***{c}***")
    print()
    print(cdv_ssfmt[c].value_counts())
    print()
    print(cdv[c].value_counts())
    print("-----\n")
else:
    scope_no_diff.append(c)
    print(f"***{c}***")
    print("NO DIFFERENCE")
    print("-----\n")

```

ANNEFUZ2

```

380    3050
390    3020
400    3016
371    2045
Name: ANNEFUZ2, dtype: int64

```

```

2016    3050
2017    3020
2018    3016
2015 online  2045
Name: ANNEFUZ2, dtype: int64
-----

```

COLLECTE

```

2    11131
Name: COLLECTE, dtype: int64

```

```

Online    11131
Name: COLLECTE, dtype: int64
-----

```

CHAMP


```

1      10574
2        557
Name: CHAMP, dtype: int64

18 ans et + métropole      10574
15-17 ans + DOM + Corse    557
Name: CHAMP, dtype: int64

```

```

***identifiant***
NO DIFFERENCE

```

```

***SEXE***

```

```

2      5848
1      5283
Name: SEXE, dtype: int64

```

```

Femme      5848
Homme       5283
Name: SEXE, dtype: int64

```

```

***AGE5***

```

```

3      3945
2      2203
5      1732
1      1686
4      1565
Name: AGE5, dtype: int64

```

```

40 à 59 ans      3945
25 à 39 ans      2203
70 ans et plus   1732
Moins de 25 ans  1686
60 à 69 ans      1565
Name: AGE5, dtype: int64

```

```

***PCSENQ8***

```

```

7      2968
4      1909
3      1570
5      1379
8      1237

```

2 1011
6 595
1 462
Name: PCSENQ8, dtype: int64

Retraité	2968
Employé	1909
Profession Intermédiaire	1570
Ouvrier	1379
Autre inactif	1237
Cadre et profession intellectuelle supérieure	1011
Personne au foyer	595
Agri. exploitant, artisan, commerçant et chef d'entreprise	462

Name: PCSENQ8, dtype: int64

TYPOSQT

3 2873
1 2815
2 2610
4 2486
5 347
Name: TYPOSQT, dtype: int64

Classes moyennes supérieures	2873
Bas revenus	2815
Classes moyennes inférieures	2610
Hauts revenus	2486
Non réponse	347

Name: TYPOSQT, dtype: int64

DIPL4

4 4942
2 3142
3 2461
1 586
Name: DIPL4, dtype: int64

Diplômé du supérieur	4942
BEPC	3142
BAC	2461
Non diplômé	586

Name: DIPL4, dtype: int64

AGGL05

4	3443
1	2467
2	1914
5	1776
3	1531

Name: AGGL05, dtype: int64

Plus de 100 000 habitants	3443
Communes rurales	2467
2 000 à 20 000 habitants	1914
Agglomération parisienne	1776
20 000 à 100 000 habitants	1531

Name: AGGL05, dtype: int64

UDA10

1	1988
4	1544
9	1351
8	1332
7	1174
3	1035
6	929
2	823
5	716
10	239

Name: UDA10, dtype: int64

Ile de France	1988
Ouest	1544
Méditerranée	1351
Sud-Est	1332
Sud-Ouest	1174
Bassin parisien Ouest	1035
Est	929
Bassin parisien Est	823
Nord	716
DOM	239

Name: UDA10, dtype: int64

SITUEMP3

1	5889
2	4498

3 744
Name: SITUEMP3, dtype: int64

Actif occupé 5889
Inactif 4498
Chômeur 744
Name: SITUEMP3, dtype: int64

AGEDIP2

5 2538
3 1404
7 1128
4 1103
9 1014
11 974
10 758
6 714
1 703
8 551
2 244
Name: AGEDIP2, dtype: int64

30-49 ans diplôme bac et +	2538
20-29 ans diplôme bac et +	1404
50-59 ans diplôme bac et +	1128
30-49 ans diplôme inférieur au bac	1103
60-69 ans diplôme bac et +	1014
70 ans et + diplôme bac et +	974
70 ans et + diplôme inférieur au bac	758
50-59 ans diplôme inférieur au bac	714
Moins de 20 ans tous diplômes confondus	703
60-69 ans diplôme inférieur au bac	551
20-29 ans diplôme inférieur au bac	244

Name: AGEDIP2, dtype: int64

AGGLO9

8 2516
9 1968
1 1783
2 961
6 867
7 733
3 703
5 671

4 591
 10 338
 Name: AGGL09, dtype: int64

Plus de 200 000 habitants	2516
Agglomération parisienne	1968
Communes rurales	1783
2 000 à 5 000 habitants	961
50 000 à 100 000 habitants	867
100 000 à 200 000 habitants	733
5 000 à 10 000 habitants	703
20 000 à 50 000 habitants	671
10 000 à 20 000 habitants	591
[Nsp]	338

Name: AGGL09, dtype: int64

DIPLOME

7 2832
 5 2461
 4 2294
 6 2110
 3 848
 1 306
 2 280
 Name: DIPLOME, dtype: int64

Bac + 3 et supérieur	2832
Bac généraliste, technologique ou professionnel	2461
Diplôme professionnel court de type CAP ou BEP	2294
Bac + 2 : Deug, IUT, DUT, BTS	2110
BEPC, brevet des collèges	848
Aucun diplôme	306
Certificat d études primaires	280

Name: DIPLOME, dtype: int64

EXERCPRO

1 10141
 2 990
 Name: EXERCPRO, dtype: int64

Exerce ou a déjà exercé une profession	10141
N'a jamais exercé de profession	990

Name: EXERCPRO, dtype: int64

SITUEMP

1	5710
6	2972
8	955
3	622
4	405
2	179
7	154
9	122
5	12

Name: SITUEMP, dtype: int64

Exerce un emploi	5710
Retraité, retiré des affaires	2972
Étudiant ou élève, sans activité professionnelle	955
Chômeur (ayant déjà travaillé)	622
Reste au foyer sans chercher d'emploi	405
Étudiant qui travaille	179
Invalide ou malade de longue durée	154
Recherche un premier emploi	122
Aide un membre de sa famille, sans rémunération	12

Name: SITUEMP, dtype: int64

SITUEMP5

5	4498
4	3633
3	1706
2	744
1	550

Name: SITUEMP5, dtype: int64

Autre	4498
Privé	3633
Public	1706
Chômeur	744
À son compte	550

Name: SITUEMP5, dtype: int64

SITUEMP6

6	4499
4	2987
3	1359

5 992
2 744
1 550
Name: SITUEMP6, dtype: int64

Autre 4499
Privé CDI 2987
Public CDI 1359
CDD, intérim 992
Chômeur 744
À son compte 550
Name: SITUEMP6, dtype: int64

SALCOMP

1.0 5339
2.0 550
Name: SALCOMP, dtype: int64

Salarié 5339
À votre compte 550
Name: SALCOMP, dtype: int64

TYPEEMPL

2.0 3462
1.0 1389
3.0 317
4.0 171
Name: TYPEEMPL, dtype: int64

D'une entreprise privée 3462
De la fonction publique 1389
D'une entreprise publique 317
D'une association 171
Name: TYPEEMPL, dtype: int64

INTERIM

2.0 5076
1.0 263
Name: INTERIM, dtype: int64

Non 5076
Oui 263

Name: INTERIM, dtype: int64

TYPCONT

2.0 4346

1.0 729

3.0 1

Name: TYPCONT, dtype: int64

À durée indéterminée 4346

À durée déterminée 729

[Nsp] 1

Name: TYPCONT, dtype: int64

TEMPSTRA

1.0 4869

2.0 1019

3.0 1

Name: TEMPSTRA, dtype: int64

Temps plein 4869

Temps partiel 1019

[Nsp] 1

Name: TEMPSTRA, dtype: int64

NBHEUR39

1.0 3698

2.0 631

3.0 602

4.0 390

5.0 248

6.0 180

7.0 140

Name: NBHEUR39, dtype: int64

Moins de 39 heures 3698

39 heures 631

40 heures 602

41 à 45 heures 390

46 à 50 heures 248

Plus de 50 heures 180

[Nsp] 140

Name: NBHEUR39, dtype: int64

NBHEUR35

2.0 1805
3.0 1407
1.0 1117
4.0 992
5.0 428
6.0 140

Name: NBHEUR35, dtype: int64

35 heures 1805
36 à 39 heures 1407
Moins de 35 heures 1117
40 à 45 heures 992
46 heures et plus 428
[Nsp] 140

Name: NBHEUR35, dtype: int64

PREFPALI

1.0 4259
2.0 1616
3.0 14

Name: PREFPALI, dtype: int64

Une amélioration de votre pouvoir d'achat 4259
Un temps libre plus long 1616
[Nsp] 14

Name: PREFPALI, dtype: int64

SALCOMPI

1.0 3864
2.0 308
3.0 155

Name: SALCOMPI, dtype: int64

Salarié 3864
À votre compte 308
N'a jamais travaillé 155

Name: SALCOMPI, dtype: int64

CHERCHEM

```

2      8816
1      2279
3        36
Name: CHERCHEM, dtype: int64

```

```

Non      8816
Oui      2279
[Nsp]     36
Name: CHERCHEM, dtype: int64
-----

```

IMAGTRAV

```

2.0      2076
3.0       592
1.0       333
4.0        15
Name: IMAGTRAV, dtype: int64

```

```

Un moyen de gagner sa vie      2076
Une source d'épanouissement, de fierté  592
Une source de stress, de mal-être      333
[Nsp]                               15
Name: IMAGTRAV, dtype: int64
-----

```

COUPLE

```

1.0      5437
3.0      2900
2.0       749
Name: COUPLE, dtype: int64

```

```

Oui, avec une personne qui vit dans le logement      5437
Non                                                    2900
Oui, avec une personne qui ne vit pas dans le logement  749
Name: COUPLE, dtype: int64
-----

```

STATMAT

```

2      4736
1      3157
3      1863
4       979
5       396
Name: STATMAT, dtype: int64

```

Marié(e)	4736
Célibataire	3157
Vivant maritalement (ou PACS)	1863
Séparé(e), divorcé(e)	979
Veuf(ve)	396

Name: STATMAT, dtype: int64

STATMAT4

2 6599

1 3157

3 979

4 396

Name: STATMAT4, dtype: int64

En ménage, marié 6599

Célibataire 3157

Séparé(e), divorcé(e) 979

Veuf(ve) 396

Name: STATMAT4, dtype: int64

ACTCONJ

1.0 3311

4.0 2367

2.0 525

3.0 268

5.0 145

6.0 16

Name: ACTCONJ, dtype: int64

Oui, à temps plein 3311

Non, a travaillé mais ne travaille plus 2367

Oui, à temps partiel 525

Non, est au chômage 268

N'a jamais travaillé 145

[Nsp] 16

Name: ACTCONJ, dtype: int64

SALCOMPC

1.0 3424

2.0 394

3.0 18

Name: SALCOMPC, dtype: int64

Salarié 3424

A son compte 394

[Nsp] 18

Name: SALCOMPC, dtype: int64

ENFANTS

1 7107

2 4024

Name: ENFANTS, dtype: int64

Oui 7107

Non 4024

Name: ENFANTS, dtype: int64

NBENF6

1 4024

3 3129

2 1883

4 1469

5 417

6 209

Name: NBENF6, dtype: int64

Aucun enfant 4024

Deux enfants 3129

Un enfant 1883

Trois enfants 1469

Quatre enfants 417

Cinq enfants et plus 209

Name: NBENF6, dtype: int64

FAMILLE

1 6089

2 4994

3 48

Name: FAMILLE, dtype: int64

Oui 6089

Non 4994

[Nsp] 48

Name: FAMILLE, dtype: int64

UNIONGAY

1 4853

2 3125

3 1627

4 1493

5 33

Name: UNIONGAY, dtype: int64

Tout à fait d'accord 4853

Assez d'accord 3125

Peu d'accord 1627

Pas du tout d'accord 1493

[Nsp] 33

Name: UNIONGAY, dtype: int64

ADOPTGAY

1 3387

4 2826

2 2750

3 2098

5 70

Name: ADOPTGAY, dtype: int64

Tout à fait d'accord 3387

Pas du tout d'accord 2826

Assez d'accord 2750

Peu d'accord 2098

[Nsp] 70

Name: ADOPTGAY, dtype: int64

TRAVFEM

3 8623

2 878

1 853

4 652

5 69

6 56

Name: TRAVFEM, dtype: int64

Elles devraient travailler quand elles le désirent

8623

Elles ne devraient travailler que si nécessaire	878
Elles ne devraient jamais travailler avec des enfants en bas âge	853
Elles devraient toujours travailler	652
Elles ne devraient jamais travailler	69
[Nsp]	56

Name: TRAVFEM, dtype: int64

NB03_4

1	10527
2	552
3	49
4	3

Name: NB03_4, dtype: int64

Aucun enfant de moins de 3 ans	10527
Un enfant de moins de 3 ans	552
Deux enfants de moins de 3 ans	49
Trois et plus enfants de moins de 3 ans	3

Name: NB03_4, dtype: int64

NB06_4

1	10046
2	811
3	251
4	23

Name: NB06_4, dtype: int64

Aucun enfant de moins de 6 ans	10046
Un enfant de moins de 6 ans	811
Deux enfants de moins de 6 ans	251
Trois et plus enfants de moins de 6 ans	23

Name: NB06_4, dtype: int64

NB10_4

1	9453
2	1015
3	549
4	114

Name: NB10_4, dtype: int64

Aucun enfant de moins de 10 ans	9453
Un enfant de moins de 10 ans	1015

Deux enfants de moins de 10 ans	549
Trois et plus enfants de moins de 10 ans	114

Name: NB10_4, dtype: int64

NB16_4

1	8658
2	1209
3	949
4	315

Name: NB16_4, dtype: int64

Aucun enfant de moins de 16 ans	8658
Un enfant de moins de 16 ans	1209
Deux enfants de moins de 16 ans	949
Trois et plus enfants de moins de 16 ans	315

Name: NB16_4, dtype: int64

NB20_4

1	8147
2	1336
3	1190
4	458

Name: NB20_4, dtype: int64

Aucun enfant de moins de 20 ans	8147
Un enfant de moins de 20 ans	1336
Deux enfants de moins de 20 ans	1190
Trois et plus enfants de moins de 20 ans	458

Name: NB20_4, dtype: int64

NB99_4

1	7543
2	1579
3	1409
4	600

Name: NB99_4, dtype: int64

Aucun enfant de 20 ans et plus	7543
Un enfant de 20 ans et plus	1579
Deux enfants de 20 ans et plus	1409
Trois et plus enfants de 20 ans et plus	600

Name: NB99_4, dtype: int64

NB03_2

2 10527

1 604

Name: NB03_2, dtype: int64

Non 10527

Oui, enfant de moins de 3 ans 604

Name: NB03_2, dtype: int64

NB06_2

2 10046

1 1085

Name: NB06_2, dtype: int64

Non 10046

Oui, enfant de moins de 6 ans 1085

Name: NB06_2, dtype: int64

NB10_2

2 9453

1 1678

Name: NB10_2, dtype: int64

Non 9453

Oui, enfant de moins de 10 ans 1678

Name: NB10_2, dtype: int64

NB16_2

2 8658

1 2473

Name: NB16_2, dtype: int64

Non 8658

Oui, enfant de moins de 16 ans 2473

Name: NB16_2, dtype: int64

NB20_2

2 8147
1 2984
Name: NB20_2, dtype: int64

Non 8147
Oui, enfant de moins de 20 ans 2984
Name: NB20_2, dtype: int64

NB99_2

2 7543
1 3588
Name: NB99_2, dtype: int64

Non 7543
Oui, enfant 3588
Name: NB99_2, dtype: int64

SITUFAM

4.0 3598
6.0 2499
1.0 2219
2.0 1778
5.0 543
3.0 485
Name: SITUFAM, dtype: int64

Couple sans enfants	3598
Couple avec enfants	2499
Personne sans enfants vivant seul	2219
Personne sans enfants vivant avec d'autres personnes (colocataires, parents...)	1778
Couple sans enfants, vivant avec d'autres personnes (colocataires, parents...)	543
Foyer monoparental	485

Name: SITUFAM, dtype: int64

RESIDALT

1.0 2319
2.0 590
4.0 60
3.0 47
Name: RESIDALT, dtype: int64

La résidence alternée chez le père et la mère 2319

La résidence principale chez la mère	590
[Nsp]	60
La résidence principale chez le père	47

Name: RESIDALT, dtype: int64

NBPIECE6

4	2775
5	2367
6	2281
3	2207
2	1151
1	350

Name: NBPIECE6, dtype: int64

Quatre pièces	2775
Cinq pièces	2367
Six et plus	2281
Trois pièces	2207
Deux pièces	1151
Une pièce	350

Name: NBPIECE6, dtype: int64

LOGSUFFI

1	9788
2	1279
3	64

Name: LOGSUFFI, dtype: int64

Oui	9788
Non	1279
[Nsp]	64

Name: LOGSUFFI, dtype: int64

STATLOGB

2.0	3877
3.0	1861
1.0	1509
4.0	1049
5.0	762
6.0	28

Name: STATLOGB, dtype: int64

Propriétaire	3877
Locataire ou sous-locataire parc privé	1861
Accédant à la propriété en cours de remboursement d'emprunt	1509
Locataire ou sous-locataire parc social	1049
Logé gratuitement	762
Autre	28

Name: STATLOGB, dtype: int64

STATLOG4

2	4857
3	4381
1	1855
4	28
5	10

Name: STATLOG4, dtype: int64

Propriétaire	4857
Locataire,sous-locataire ou logé gratuitement	4381
Accédant à la propriété	1855
Autre	28
[Nsp]	10

Name: STATLOG4, dtype: int64

DEPLOG

2	4622
3	4445
4	1025
1	719
5	268
6	52

Name: DEPLOG, dtype: int64

Une charge que vous pouvez supporter sans difficulté	4622
Une lourde charge	4445
Une très lourde charge	1025
Une charge négligeable	719
Une charge à laquelle vous ne pouvez faire face	268
[Nsp]	52

Name: DEPLOG, dtype: int64

DEPLOG3

2	5738
---	------

1 5341
 3 52
 Name: DEPLOY3, dtype: int64

Une charge lourde, très lourde 5738
 Une charge négligeable, supportable 5341
 [Nsp] 52
 Name: DEPLOY3, dtype: int64

CADVIE

2 6844
 1 2781
 3 1184
 4 275
 5 47
 Name: CADVIE, dtype: int64

Satisfait 6844
 Très satisfait 2781
 Peu satisfait 1184
 Pas satisfait du tout 275
 [Nsp] 47
 Name: CADVIE, dtype: int64

CADVIE3

1 9625
 2 1459
 3 47
 Name: CADVIE3, dtype: int64

Satisfait de son cadre de vie 9625
 Non satisfait de son cadre de vie 1459
 [Nsp] 47
 Name: CADVIE3, dtype: int64

SECURITE

2 6773
 1 2740
 3 1271
 4 286
 5 61
 Name: SECURITE, dtype: int64

Assez en sécurité	6773
Tout à fait en sécurité	2740
Peu en sécurité	1271
Pas du tout en sécurité	286
[Nsp]	61

Name: SECURITE, dtype: int64

SECUR3

1	9513
2	1557
3	61

Name: SECUR3, dtype: int64

En sécurité	9513
Pas en sécurité	1557
Nsp, NR	61

Name: SECUR3, dtype: int64

MODCHAUF

1	3866
3	3643
4	1030
6	1028
8	560
7	331
2	243
10	241
9	126
11	35
5	28

Name: MODCHAUF, dtype: int64

Électricité	3866
Gaz naturel (gaz de réseau)	3643
Fioul, mazout	1030
Bois	1028
NSP (pour certains chauffages collectifs)	560
Chauffage urbain	331
Gaz butane, propane	243
Géothermie, pompe à chaleur	241
Autre	126
Aucun	35
Charbon	28

Name: MODCHAUF, dtype: int64

TYPCHAUF

1 8961

2 1967

3 203

Name: TYPCHAUF, dtype: int64

Une installation individuelle propre à votre logement 8961

Un système de chauffage collectif 1967

[Nsp] 203

Name: TYPCHAUF, dtype: int64

VOITURE

1 9656

2 1420

3 55

Name: VOITURE, dtype: int64

Oui 9656

Non 1420

[Nsp] 55

Name: VOITURE, dtype: int64

CONDUIT

1 9319

3 1170

2 574

4 68

Name: CONDUIT, dtype: int64

Oui 9319

N'a pas le permis de conduire 1170

N'a pas conduit, mais a le permis de conduire 574

[Nsp] 68

Name: CONDUIT, dtype: int64

CHOIXNUC

1 4749

2 3314

3 3068
Name: CHOIXNUC, dtype: int64

Plutôt des avantages 4749
Plutôt des inconvénients 3314
Sans avis 3068
Name: CHOIXNUC, dtype: int64

TAXENV

2 8243
1 2810
3 78
Name: TAXENV, dtype: int64

Non 8243
Oui 2810
[Nsp] 78
Name: TAXENV, dtype: int64

HANDICAP

2 8273
1 2766
3 92
Name: HANDICAP, dtype: int64

Non 8273
Oui 2766
[Nsp] 92
Name: HANDICAP, dtype: int64

SOUFFTET

2 6630
1 4293
3 208
Name: SOUFFTET, dtype: int64

Non 6630
Oui 4293
[Nsp] 208
Name: SOUFFTET, dtype: int64

SOUFFDOS

1 5859
2 5114
3 158
Name: SOUFFDOS, dtype: int64

Oui 5859
Non 5114
[Nsp] 158
Name: SOUFFDOS, dtype: int64

SOUFFNER

2 7306
1 3589
3 236
Name: SOUFFNER, dtype: int64

Non 7306
Oui 3589
[Nsp] 236
Name: SOUFFNER, dtype: int64

SOUFFDEP

2 9124
1 1743
3 264
Name: SOUFFDEP, dtype: int64

Non 9124
Oui 1743
[Nsp] 264
Name: SOUFFDEP, dtype: int64

SOUFFINS

2 7102
1 3801
3 228
Name: SOUFFINS, dtype: int64

Non 7102
Oui 3801

[Nsp] 228
Name: SOUFFINS, dtype: int64

ETATSAN

2 7386
1 1681
3 1650
4 333
5 81
Name: ETATSAN, dtype: int64

Satisfaisant 7386
Très satisfaisant 1681
Peu satisfaisant 1650
Pas satisfaisant du tout 333
[Nsp] 81
Name: ETATSAN, dtype: int64

LIMVIAND

3.0 1320
1.0 1069
2.0 588
4.0 39
Name: LIMVIAND, dtype: int64

Vous ne l'envisagez pas du tout 1320
Vous limitez déjà votre consommation 1069
Vous envisagez de limiter votre consommation 588
[Nsp] 39
Name: LIMVIAND, dtype: int64

WHYLIM

1.0 623
2.0 463
3.0 379
4.0 178
5.0 14
Name: WHYLIM, dtype: int64

C'est bon pour la santé 623
C'est bon pour l'environnement, la planète 463
Ca limite les souffrances animales 379

Ce sont des produits qui coûtent cher	178
[Nsp]	14

Name: WHYLIM, dtype: int64

ADNSTIC

2.0	1187
1.0	978
3.0	420
4.0	395
5.0	36

Name: ADNSTIC, dtype: int64

Assez d'accord	1187
Tout à fait d'accord	978
Pas très d'accord	420
Pas du tout d'accord	395
[Nsp]	36

Name: ADNSTIC, dtype: int64

ADNCB

4.0	1102
3.0	929
2.0	683
1.0	265
5.0	37

Name: ADNCB, dtype: int64

Pas du tout d'accord	1102
Pas très d'accord	929
Assez d'accord	683
Tout à fait d'accord	265
[Nsp]	37

Name: ADNCB, dtype: int64

ADNORDI

2.0	1050
3.0	756
4.0	728
1.0	441
5.0	41

Name: ADNORDI, dtype: int64

```

Assez d'accord          1050
Pas très d'accord       756
Pas du tout d'accord    728
Tout à fait d'accord    441
[Nsp]                   41
Name: ADNORDI, dtype: int64
-----

***ROBOT1***

2.0    1616
1.0    1321
3.0      79
Name: ROBOT1, dtype: int64

Cest plutôt une mauvaise chose    1616
Cest plutôt une bonne chose       1321
[Nsp]                             79
Name: ROBOT1, dtype: int64
-----

***ROBOT2***

2.0    1551
1.0    1388
3.0      77
Name: ROBOT2, dtype: int64

Cest plutôt une mauvaise chose    1551
Cest plutôt une bonne chose       1388
[Nsp]                             77
Name: ROBOT2, dtype: int64
-----

***ROBOT3***

2.0    1527
1.0    1413
3.0      76
Name: ROBOT3, dtype: int64

Cest plutôt une mauvaise chose    1527
Cest plutôt une bonne chose       1413
[Nsp]                             76
Name: ROBOT3, dtype: int64
-----

***NBPERS5***

```

```
2    4390
1    2219
3    1881
4    1717
5     924
```

Name: NBPERS5, dtype: int64

```
Deux personnes      4390
Une personne        2219
Trois personnes     1881
Quatre personnes    1717
Cinq personnes et plus  924
```

Name: NBPERS5, dtype: int64

SEXE_2

```
1.0    3916
2.0    3860
```

Name: SEXE_2, dtype: int64

```
Homme    3916
Femme    3860
```

Name: SEXE_2, dtype: int64

SEXE_3

```
2.0    2054
1.0    1925
```

Name: SEXE_3, dtype: int64

```
Femme    2054
Homme     1925
```

Name: SEXE_3, dtype: int64

SEXE_4

```
2.0    1197
1.0    1134
```

Name: SEXE_4, dtype: int64

```
Femme    1197
Homme     1134
```

Name: SEXE_4, dtype: int64

SEXE_5

1.0 442

2.0 389

Name: SEXE_5, dtype: int64

Homme 442

Femme 389

Name: SEXE_5, dtype: int64

SEXE_6

2.0 118

1.0 93

Name: SEXE_6, dtype: int64

Femme 118

Homme 93

Name: SEXE_6, dtype: int64

SEXE_7

1.0 28

2.0 25

Name: SEXE_7, dtype: int64

Homme 28

Femme 25

Name: SEXE_7, dtype: int64

SEXE_8

2.0 18

1.0 6

Name: SEXE_8, dtype: int64

Femme 18

Homme 6

Name: SEXE_8, dtype: int64

SEXE_9

1.0 5

2.0 2
Name: SEXE_9, dtype: int64

Homme 5
Femme 2
Name: SEXE_9, dtype: int64

LIEN_2

1.0 6725
3.0 960
2.0 655
4.0 354
5.0 101
6.0 32
7.0 27
Name: LIEN_2, dtype: int64

Conjoint ou compagnon	6725
Parent, beau-parent	960
Enfant, beau-fils, belle-fille	655
Frère, soeur	354
Autre	101
6	32
7	27

Name: LIEN_2, dtype: int64

LIEN_3

2.0 3147
3.0 820
4.0 325
5.0 99
1.0 57
6.0 17
7.0 11
Name: LIEN_3, dtype: int64

Enfant, beau-fils, belle-fille	3147
Parent, beau-parent	820
Frère, soeur	325
Autre	99
Conjoint ou compagnon	57
6	17
7	11

Name: LIEN_3, dtype: int64

LIEN_4

2.0 1807
4.0 497
3.0 163
6.0 55
5.0 50
1.0 29
7.0 12

Name: LIEN_4, dtype: int64

Enfant, beau-fils, belle-fille	1807
Frère, soeur	497
Parent, beau-parent	163
6	55
Autre	50
Conjoint ou compagnon	29
7	12

Name: LIEN_4, dtype: int64

LIEN_5

2.0 562
4.0 222
3.0 56
6.0 28
5.0 22
1.0 9
7.0 4

Name: LIEN_5, dtype: int64

Enfant, beau-fils, belle-fille	562
Frère, soeur	222
Parent, beau-parent	56
6	28
Autre	22
Conjoint ou compagnon	9
7	4

Name: LIEN_5, dtype: int64

LIEN_6

2.0 113
4.0 67

3.0 17
 5.0 9
 6.0 9
 1.0 8
 7.0 2

Name: LIEN_6, dtype: int64

Enfant, beau-fils, belle-fille	113
Frère, soeur	67
Parent, beau-parent	17
Autre	9
6	9
Conjoint ou compagnon	8
7	2

Name: LIEN_6, dtype: int64

LIEN_7

4.0 19
 2.0 19
 3.0 7
 5.0 6
 7.0 2
 6.0 2

Name: LIEN_7, dtype: int64

Enfant, beau-fils, belle-fille	19
Frère, soeur	19
Parent, beau-parent	7
Autre	6
7	2
6	2

Name: LIEN_7, dtype: int64

LIEN_8

4.0 9
 5.0 5
 2.0 5
 3.0 2
 6.0 2
 1.0 1
 7.0 1

Name: LIEN_8, dtype: int64

Frère, soeur	9
--------------	---

Enfant, beau-fils, belle-fille	5
Autre	5
Parent, beau-parent	2
6	2
Conjoint ou compagnon	1
7	1

Name: LIEN_8, dtype: int64

LIEN_9

5.0	3
2.0	3
1.0	1
4.0	1

Name: LIEN_9, dtype: int64

Enfant, beau-fils, belle-fille	3
Autre	3
Conjoint ou compagnon	1
Frère, soeur	1

Name: LIEN_9, dtype: int64

PRESTCAF

2	8248
1	2869
3	14

Name: PRESTCAF, dtype: int64

Non	8248
Oui	2869
[Nsp]	14

Name: PRESTCAF, dtype: int64

RESTRICT

1	6454
2	4571
3	106

Name: RESTRICT, dtype: int64

Oui	6454
Non	4571
[Nsp]	106

Name: RESTRICT, dtype: int64

```

-----

***RE_MEDI***

2.0      3906
1.0      2155
3.0       393
Name: RE_MEDI, dtype: int64

Non       3906
Oui       2155
[Nsp]     393
Name: RE_MEDI, dtype: int64
-----

***RE_VOIT***

1.0      3784
2.0      2310
3.0       360
Name: RE_VOIT, dtype: int64

Oui       3784
Non       2310
[Nsp]     360
Name: RE_VOIT, dtype: int64
-----

***RE_EQUI***

1.0      4670
2.0      1465
3.0       319
Name: RE_EQUI, dtype: int64

Oui       4670
Non       1465
[Nsp]     319
Name: RE_EQUI, dtype: int64
-----

***RE_ALIM***

2.0      3095
1.0      2992
3.0       367
Name: RE_ALIM, dtype: int64

```

Non 3095
Oui 2992
[Nsp] 367
Name: RE_ALIM, dtype: int64

RE_VAC

1.0 5730
2.0 566
3.0 158
Name: RE_VAC, dtype: int64

Oui 5730
Non 566
[Nsp] 158
Name: RE_VAC, dtype: int64

RE_HABI

1.0 4949
2.0 1248
3.0 257
Name: RE_HABI, dtype: int64

Oui 4949
Non 1248
[Nsp] 257
Name: RE_HABI, dtype: int64

RE_LOG

2.0 3836
1.0 2172
3.0 446
Name: RE_LOG, dtype: int64

Non 3836
Oui 2172
[Nsp] 446
Name: RE_LOG, dtype: int64

RE_ENF

2.0 4043

```

1.0      1937
3.0      474
Name: RE_ENF, dtype: int64

Non      4043
Oui      1937
[Nsp]    474
Name: RE_ENF, dtype: int64
-----

***RE_TABAL***

2.0      3513
1.0      2466
3.0      475
Name: RE_TABAL, dtype: int64

Non      3513
Oui      2466
[Nsp]    475
Name: RE_TABAL, dtype: int64
-----

***RE_WEB***

1.0      3083
2.0      2942
3.0      429
Name: RE_WEB, dtype: int64

Oui      3083
Non      2942
[Nsp]    429
Name: RE_WEB, dtype: int64
-----

***NIVPERSO***

3      3445
4      3248
5      1980
2      1727
1      613
6      118
Name: NIVPERSO, dtype: int64

C'est pareil      3445
Un peu moins bien  3248

```

```
Beaucoup moins bien    1980
Un peu mieux           1727
Beaucoup mieux         613
[Nsp]                  118
Name: NIVPERS0, dtype: int64
-----
```

NIVPERS4

```
3    5228
2    3445
1    2340
4     118
Name: NIVPERS4, dtype: int64
```

```
Moins bien    5228
Pareil        3445
Mieux         2340
[Nsp]         118
Name: NIVPERS4, dtype: int64
-----
```

NIVFRAN

```
4    4766
5    2986
3    2331
2     749
6     173
1     126
Name: NIVFRAN, dtype: int64
```

```
Un peu moins bien    4766
Beaucoup moins bien  2986
C'est pareil         2331
Un peu mieux         749
[Nsp]                173
Beaucoup mieux       126
Name: NIVFRAN, dtype: int64
-----
```

NIVFRAN4

```
3    7752
2    2331
1     875
4     173
Name: NIVFRAN4, dtype: int64
```

Moins bien	7752
Pareil	2331
Mieux	875
[Nsp]	173

Name: NIVFRAN4, dtype: int64

CDV5

3	3661
4	3591
2	1821
5	1604
1	357
6	97

Name: CDV5, dtype: int64

Vont rester semblables	3661
Vont se détériorer un petit peu	3591
Vont s'améliorer un petit peu	1821
Vont se détériorer beaucoup	1604
Vont s'améliorer beaucoup	357
[Nsp]	97

Name: CDV5, dtype: int64

CDV5_4

3	5195
2	3661
1	2178
4	97

Name: CDV5_4, dtype: int64

Moins bien	5195
Pareil	3661
Mieux	2178
[Nsp]	97

Name: CDV5_4, dtype: int64

NBCHOM

1	5915
2	4200
3	913
4	103

Name: NBCHOM, dtype: int64

Augmenter pendant plusieurs années	5915
Se stabiliser dans un avenir proche	4200
Diminuer dans un avenir proche	913
[Nsp]	103

Name: NBCHOM, dtype: int64

BANQMOB

2	8311
1	2639
3	181

Name: BANQMOB, dtype: int64

Non	8311
Oui	2639
[Nsp]	181

Name: BANQMOB, dtype: int64

BANQEPA

1	8546
2	2451
3	134

Name: BANQEPA, dtype: int64

Oui	8546
Non	2451
[Nsp]	134

Name: BANQEPA, dtype: int64

BANQVIE

1	5552
2	5410
3	169

Name: BANQVIE, dtype: int64

Oui	5552
Non	5410
[Nsp]	169

Name: BANQVIE, dtype: int64

INNOVTEC

2 5421
3 3548
1 1328
4 734
5 100

Name: INNOVTEC, dtype: int64

Assez attiré 5421
Peu attiré 3548
Très attiré 1328
Pas du tout attiré 734
[Nsp] 100

Name: INNOVTEC, dtype: int64

TELMOB

1 10630
2 407
3 94

Name: TELMOB, dtype: int64

Oui 10630
Non 407
[Nsp] 94

Name: TELMOB, dtype: int64

PRATCOLL

4.0 5518
2.0 1600
3.0 1497
1.0 390
5.0 54

Name: PRATCOLL, dtype: int64

Jamais 5518
De temps en temps 1600
Rarement 1497
Souvent 390
[Nsp] 54

Name: PRATCOLL, dtype: int64

CONFWEB


```
2.0    4504
3.0    3226
4.0     906
1.0     336
5.0     114
Name: CONFWEB, dtype: int64
```

```
Assez confiance    4504
Peu confiance      3226
Pas du tout confiance    906
Très confiance     336
[Nsp]              114
Name: CONFWEB, dtype: int64
```

VACANCES

```
1    7457
2    3588
3      86
Name: VACANCES, dtype: int64
```

```
Oui    7457
Non    3588
[Nsp]    86
Name: VACANCES, dtype: int64
```

VISITFAM

```
1.0    6550
2.0    1459
3.0     72
Name: VISITFAM, dtype: int64
```

```
Oui    6550
Non    1459
[Nsp]    72
Name: VISITFAM, dtype: int64
```

RECEP

```
4.0    3203
3.0    2907
2.0    1256
5.0     482
```

1.0 177
 6.0 56
 Name: RECEP, dtype: int64

Plus rarement 3203
 En moyenne, une fois par mois 2907
 En moyenne, une fois par semaine 1256
 Jamais 482
 Tous les jours ou presque 177
 [Nsp] 56
 Name: RECEP, dtype: int64

YOGA

5.0 2128
 4.0 408
 2.0 228
 3.0 139
 1.0 73
 6.0 40
 Name: YOGA, dtype: int64

Jamais 2128
 Plus rarement 408
 En moyenne, une fois par semaine 228
 En moyenne, une fois par mois 139
 Tous les jours ou presque 73
 [Nsp] 40
 Name: YOGA, dtype: int64

FREQSPOR

3 5458
 2 2907
 1 2674
 4 92
 Name: FREQSPOR, dtype: int64

Jamais 5458
 Exceptionnellement 2907
 Régulièrement 2674
 [Nsp] 92
 Name: FREQSPOR, dtype: int64

FREQBIBL

```

3    5381
2    3437
1    2210
4     103
Name: FREQBIBL, dtype: int64

```

```

Jamais          5381
Exceptionnellement 3437
Régulièrement   2210
[Nsp]           103
Name: FREQBIBL, dtype: int64

```

```

-----

```

```

***FREQCINE***

```

```

2    5575
1    3256
3    2196
4     104
Name: FREQCINE, dtype: int64

```

```

Exceptionnellement 5575
Régulièrement      3256
Jamais             2196
[Nsp]              104
Name: FREQCINE, dtype: int64

```

```

-----

```

```

***FREQTELE***

```

```

1    7354
2    2172
3    1038
4     479
5      88
Name: FREQTELE, dtype: int64

```

```

Tous les jours    7354
Assez souvent     2172
Pas très souvent  1038
Jamais            479
[Nsp]             88
Name: FREQTELE, dtype: int64

```

```

-----

```

```

***ASSOSPOR***

```

```
2      8132
1      2594
3       405
Name: ASSOSPOR, dtype: int64
```

```
Non      8132
Oui      2594
[Nsp]     405
Name: ASSOSPOR, dtype: int64
-----
```

ASSOCULT

```
2      8421
1      2294
3       416
Name: ASSOCULT, dtype: int64
```

```
Non      8421
Oui      2294
[Nsp]     416
Name: ASSOCULT, dtype: int64
-----
```

ASSOCONF

```
2      10039
1        573
3        519
Name: ASSOCONF, dtype: int64
```

```
Non      10039
Oui        573
[Nsp]     519
Name: ASSOCONF, dtype: int64
-----
```

ASSOJEUN

```
2      10128
3        542
1        461
Name: ASSOJEUN, dtype: int64
```

```
Non      10128
[Nsp]     542
Oui        461
Name: ASSOJEUN, dtype: int64
```

ASSOSYND

2 9744
1 850
3 537
Name: ASSOSYND, dtype: int64

Non 9744
Oui 850
[Nsp] 537
Name: ASSOSYND, dtype: int64

ASSOENVI

2 10069
1 533
3 529
Name: ASSOENVI, dtype: int64

Non 10069
Oui 533
[Nsp] 529
Name: ASSOENVI, dtype: int64

ASSOPARE

2 10043
1 546
3 542
Name: ASSOPARE, dtype: int64

Non 10043
Oui 546
[Nsp] 542
Name: ASSOPARE, dtype: int64

ASSOCONS

2 9890
1 693
3 548
Name: ASSOCONS, dtype: int64

```
Non      9890
Oui      693
[Nsp]    548
Name: ASSOCONS, dtype: int64
-----
```

ASSOPOLI

```
2      9969
1      608
3      554
Name: ASSOPOLI, dtype: int64
```

```
Non      9969
Oui      608
[Nsp]    554
Name: ASSOPOLI, dtype: int64
-----
```

ASSOHUMA

```
2      9375
1      1256
3      500
Name: ASSOHUMA, dtype: int64
```

```
Non      9375
Oui      1256
[Nsp]    500
Name: ASSOHUMA, dtype: int64
-----
```

ASSOAUTR

```
2.0     10762
1.0       310
Name: ASSOAUTR, dtype: int64
```

```
Non      10762
Oui       310
Name: ASSOAUTR, dtype: int64
-----
```

ASS011_2

```
1      5779
2      5352
Name: ASS011_2, dtype: int64
```

Adhérent 5779
Non adhérent 5352
Name: ASS011_2, dtype: int64

ASS011_3

1 5352
2 3127
3 2652
Name: ASS011_3, dtype: int64

Aucune 5352
Une 3127
Deux ou plus 2652
Name: ASS011_3, dtype: int64

ASS06_2

2 6181
1 4950
Name: ASS06_2, dtype: int64

Non adhérent 6181
Adhérent 4950
Name: ASS06_2, dtype: int64

ASS010_2

2 5642
1 5489
Name: ASS010_2, dtype: int64

Non adhérent 5642
Adhérent 5489
Name: ASS010_2, dtype: int64

ASS010_3

1 5642
2 3194
3 2295
Name: ASS010_3, dtype: int64

```
Aucune          5642
Une             3194
Deux ou plus    2295
Name: ASS010_3, dtype: int64
-----
```

HARVEY

```
1.0    1438
2.0    1258
3.0     284
4.0      36
Name: HARVEY, dtype: int64
```

```
On n'en parle pas assez    1438
On en parle suffisamment    1258
On en parle trop            284
[Nsp]                       36
Name: HARVEY, dtype: int64
-----
```

AIDESUFF

```
1    7335
2    3636
3     160
Name: AIDESUFF, dtype: int64
```

```
Suffisante    7335
Insuffisante   3636
[Nsp]          160
Name: AIDESUFF, dtype: int64
-----
```

EFFORTPP

```
3    4656
2    4348
1    1965
4     162
Name: EFFORTPP, dtype: int64
```

```
Ne font pas assez    4656
Font ce qu'ils doivent 4348
Font trop            1965
[Nsp]                162
Name: EFFORTPP, dtype: int64
-----
```


RAISPAUV

1 7157

2 3755

3 219

Name: RAISPAUV, dtype: int64

Plutôt parce qu'elles n'ont pas eu de chance 7157

Plutôt parce qu'elles n'ont pas fait d'effort 3755

[Nsp] 219

Name: RAISPAUV, dtype: int64

CHOAVANT

1 8656

2 2328

3 147

Name: CHOAVANT, dtype: int64

Plutôt d'accord 8656

Plutôt pas d'accord 2328

[Nsp] 147

Name: CHOAVANT, dtype: int64

CHOVOLON

1 6351

2 4614

3 166

Name: CHOVOLON, dtype: int64

Plutôt d'accord 6351

Plutôt pas d'accord 4614

[Nsp] 166

Name: CHOVOLON, dtype: int64

OPIRSA

1 6278

2 4609

3 244

Name: OPIRSA, dtype: int64

Cela risque d'inciter les gens à s'en contenter 6278

Cela leur donnera le coup de pouce nécessaire 4609
[Nsp] 244
Name: OPIRSA, dtype: int64

JUSTICE

3 5365
2 3080
4 2379
1 206
5 101
Name: JUSTICE, dtype: int64

Assez mal 5365
Assez bien 3080
Très mal 2379
Très bien 206
[Nsp] 101
Name: JUSTICE, dtype: int64

RELEG

3.0 1302
2.0 1069
1.0 311
4.0 290
5.0 44
Name: RELEG, dtype: int64

Non, plutôt pas 1302
Oui, plutôt 1069
Oui, tout à fait 311
Non, pas du tout 290
[Nsp] 44
Name: RELEG, dtype: int64

CONFPUB

2 6026
3 3418
4 1030
1 503
5 154
Name: CONFPUB, dtype: int64

Plutôt confiance	6026
Plutôt pas confiance	3418
Pas du tout confiance	1030
Très confiance	503
[Nsp]	154

Name: CONFPUB, dtype: int64

CONFENTR

2	6402
3	3131
4	829
1	569
5	200

Name: CONFENTR, dtype: int64

Plutôt confiance	6402
Plutôt pas confiance	3131
Pas du tout confiance	829
Très confiance	569
[Nsp]	200

Name: CONFENTR, dtype: int64

CONFASSO

2	7061
3	2330
1	852
4	691
5	197

Name: CONFASSO, dtype: int64

Plutôt confiance	7061
Plutôt pas confiance	2330
Très confiance	852
Pas du tout confiance	691
[Nsp]	197

Name: CONFASSO, dtype: int64

CONFPOLI

4	4700
3	4543
2	1524
5	201

1 163
Name: CONFPOLI, dtype: int64

Pas du tout confiance 4700
Plutôt pas confiance 4543
Plutôt confiance 1524
[Nsp] 201
Très confiance 163
Name: CONFPOLI, dtype: int64

CONFBANK

3.0 3624
2.0 2987
4.0 2079
1.0 239
5.0 157
Name: CONFBANK, dtype: int64

Plutôt pas confiance 3624
Plutôt confiance 2987
Pas du tout confiance 2079
Très confiance 239
[Nsp] 157
Name: CONFBANK, dtype: int64

CONFPRES

3.0 4186
2.0 2556
4.0 2008
1.0 180
5.0 156
Name: CONFPRES, dtype: int64

Plutôt pas confiance 4186
Plutôt confiance 2556
Pas du tout confiance 2008
Très confiance 180
[Nsp] 156
Name: CONFPRES, dtype: int64

CONFECOL

2.0 5475

```
3.0    2018
1.0    888
4.0    556
5.0    149
Name: CONFECOL, dtype: int64
```

```
Plutôt confiance      5475
Plutôt pas confiance   2018
Très confiance        888
Pas du tout confiance  556
[Nsp]                 149
Name: CONFECOL, dtype: int64
-----
```

CONFKEUF

```
2.0    5538
1.0    1457
3.0    1415
4.0     530
5.0     146
Name: CONFKEUF, dtype: int64
```

```
Plutôt confiance      5538
Très confiance        1457
Plutôt pas confiance   1415
Pas du tout confiance  530
[Nsp]                 146
Name: CONFKEUF, dtype: int64
-----
```

TRANSFST

```
1     9288
2     1710
3      133
Name: TRANSFST, dtype: int64
```

```
Oui      9288
Non       1710
[Nsp]     133
Name: TRANSFST, dtype: int64
-----
```

TRANSFO5

```
2     4988
1     4231
```

4 1710
 5 133
 3 69
 Name: TRANSF05, dtype: int64

Radical 4988
 Progressiste 4231
 Pas de réforme 1710
 Nsp si réforme 133
 Nsp quelle réforme 69
 Name: TRANSF05, dtype: int64

PROGRAD

2.0 4988
 1.0 4231
 3.0 69
 Name: PROGRAD, dtype: int64

Changements radicaux 4988
 Réformes progressives 4231
 [Nsp] 69
 Name: PROGRAD, dtype: int64

RADIQUOI

NO DIFFERENCE

RADWHY1

0.0 1091
 1.0 81
 Name: RADWHY1, dtype: int64

0 1091
 Position anti-Elite, anti-Etat 81
 Name: RADWHY1, dtype: int64

RADWHY2

0.0 1081
 2.0 91
 Name: RADWHY2, dtype: int64

0 1081

plus de social: tout le monde doit pouvoir en profiter 91
Name: RADWHY2, dtype: int64

RADWHY3

0.0 1116
3.0 56
Name: RADWHY3, dtype: int64

0 1116
Plus de justice sociale, réduction des inégalités 56
Name: RADWHY3, dtype: int64

RADWHY4

0.0 1098
4.0 74
Name: RADWHY4, dtype: int64

0 1098
Plus d'éducation, formation professionnelle des jeunes, s'occuper des étudiants 74
Name: RADWHY4, dtype: int64

RADWHY5

0.0 1107
5.0 65
Name: RADWHY5, dtype: int64

0 1107
Moins de social : fraudes, abus, profiteurs, 65
Name: RADWHY5, dtype: int64

RADWHY6

0.0 1129
6.0 43
Name: RADWHY6, dtype: int64

0 1129
Réduire les impôts, taxes, cotisations patronales 43
Name: RADWHY6, dtype: int64

RADWHY7

0.0 1119

7.0 53

Name: RADWHY7, dtype: int64

0

1119

Création d'emploi, chômage, embauche, salaires 53

Name: RADWHY7, dtype: int64

RADWHY8

0.0 1055

8.0 117

Name: RADWHY8, dtype: int64

0

1055

Moins d'immigration, il faut renvoyer tous les immigrés chez eux 117

Name: RADWHY8, dtype: int64

RADWHY9

0.0 1167

9.0 5

Name: RADWHY9, dtype: int64

0

Protectionnisme économique, contre l'Europe, renforcer les filières industrielles françaises

Name: RADWHY9, dtype: int64

RADWHY10

0.0 1024

10.0 148

Name: RADWHY10, dtype: int64

0

1024

Agir sur la sécurité, plus d'ordre, une justice plus sévère 148

Name: RADWHY10, dtype: int64

RADWHY11

0.0 1158

11.0 14

Name: RADWHY11, dtype: int64

0	1158
Nouvelles formes du vivre ensemble, écologie, décroissance, fin du capitalisme	14

Name: RADWHY11, dtype: int64

RADWHY12

0.0 732

12.0 440

Name: RADWHY12, dtype: int64

0	732
---	-----

Nsp, NR, réformulation de la question ou contournement de la réponse	440
--	-----

Name: RADWHY12, dtype: int64

RADWHY13

0.0 1030

13.0 142

Name: RADWHY13, dtype: int64

0 1030

Autre 142

Name: RADWHY13, dtype: int64

RADWHY14

0.0 1124

14.0 48

Name: RADWHY14, dtype: int64

0

Aligner le public sur le privé, réduire les dépenses publiques (simplifier l'administration, sup

Name: RADWHY14, dtype: int64

ORDLIB

2 4711

1 3699

3 1817

4 739

5 165

Name: ORDLIB, dtype: int64

Un peu plus d'ordre	4711
Beaucoup plus d'ordre	3699
Un peu plus de liberté	1817
Beaucoup plus de liberté	739
[Nsp]	165

Name: ORDLIB, dtype: int64

CONFMEFI

2.0	4931
1.0	4024
3.0	131

Name: CONFMEFI, dtype: int64

On n'est jamais assez méfiant	4931
Il est possible de faire confiance aux autres	4024
[Nsp]	131

Name: CONFMEFI, dtype: int64

PREOCCU1

1	2716
7	1789
10	1465
2	1218
4	1071
3	1063
8	696
5	543
9	178
6	158
11	140
12	94

Name: PREOCCU1, dtype: int64

Le chômage	2716
La violence et l'insécurité	1789
L'immigration	1465
Les maladies graves	1218
La pauvreté en France	1071
La dégradation de l'environnement	1063
Les tensions internationales	696
La pauvreté dans le monde	543
Les conflits sociaux	178
La drogue	158

L'Europe	140
[Nsp, NR]	94

Name: PREOCCU1, dtype: int64

PREOCCU2

7	2126
10	1581
4	1367
1	1141
3	1119
8	932
2	849
5	801
9	348
11	343
6	278
12	246

Name: PREOCCU2, dtype: int64

La violence et l'insécurité	2126
L'immigration	1581
La pauvreté en France	1367
Le chômage	1141
La dégradation de l'environnement	1119
Les tensions internationales	932
Les maladies graves	849
La pauvreté dans le monde	801
Les conflits sociaux	348
L'Europe	343
La drogue	278
[Nsp, NR]	246

Name: PREOCCU2, dtype: int64

PREOCHOM

2	7274
1	3857

Name: PREOCHOM, dtype: int64

Non	7274
Oui	3857

Name: PREOCHOM, dtype: int64

PREOMALA

2 9064
1 2067
Name: PREOMALA, dtype: int64

Non 9064
Oui 2067
Name: PREOMALA, dtype: int64

PREOENV

2 8949
1 2182
Name: PREOENV, dtype: int64

Non 8949
Oui 2182
Name: PREOENV, dtype: int64

PREOPAUF

2 8693
1 2438
Name: PREOPAUF, dtype: int64

Non 8693
Oui 2438
Name: PREOPAUF, dtype: int64

PREOPAUM

2 9787
1 1344
Name: PREOPAUM, dtype: int64

Non 9787
Oui 1344
Name: PREOPAUM, dtype: int64

PREODROG

2 10695
1 436
Name: PREODROG, dtype: int64

Non 10695
Oui 436
Name: PREODROG, dtype: int64

PREOVIO

2 7216
1 3915
Name: PREOVIO, dtype: int64

Non 7216
Oui 3915
Name: PREOVIO, dtype: int64

PREOTENS

2 9503
1 1628
Name: PREOTENS, dtype: int64

Non 9503
Oui 1628
Name: PREOTENS, dtype: int64

PREOCONF

2 10605
1 526
Name: PREOCONF, dtype: int64

Non 10605
Oui 526
Name: PREOCONF, dtype: int64

PREOIMMI

2 8085
1 3046
Name: PREOIMMI, dtype: int64

Non 8085
Oui 3046
Name: PREOIMMI, dtype: int64

PREOEURO

2 10648
1 483
Name: PREOEURO, dtype: int64

Non 10648
Oui 483
Name: PREOEURO, dtype: int64

RADI1

1.0 1691
2.0 1282
3.0 43
Name: RADI1, dtype: int64

Oui, je comprends 1691
Non, je ne comprends pas 1282
[Nsp] 43
Name: RADI1, dtype: int64

RADI2

2.0 2683
1.0 284
3.0 49
Name: RADI2, dtype: int64

Non, je ne comprends pas 2683
Oui, je comprends 284
[Nsp] 49
Name: RADI2, dtype: int64

RADI3

2.0 2609
1.0 359
3.0 48
Name: RADI3, dtype: int64

Non, je ne comprends pas 2609
Oui, je comprends 359

```

[Nsp]                                     48
Name: RAD13, dtype: int64
-----

***OPICULT***

1.0      4603
2.0      4250
3.0       233
Name: OPICULT, dtype: int64

La diversité des cultures est une richesse      4603
La diversité des cultures rend difficile la vie en commun  4250
[Nsp]                                             233
Name: OPICULT, dtype: int64
-----

***COMMU1***

2.0      2284
1.0       594
3.0       138
Name: COMMU1, dtype: int64

Non      2284
Oui       594
[Nsp]     138
Name: COMMU1, dtype: int64
-----

***COMMU2***

2.0      2063
1.0       828
3.0       125
Name: COMMU2, dtype: int64

Non      2063
Oui       828
[Nsp]     125
Name: COMMU2, dtype: int64
-----

***COMMU3***

2.0      2434
1.0       463
3.0       119

```

Name: COMMU3, dtype: int64

Non 2434

Oui 463

[Nsp] 119

Name: COMMU3, dtype: int64

COMMU4

2.0 2030

1.0 874

3.0 112

Name: COMMU4, dtype: int64

Non 2030

Oui 874

[Nsp] 112

Name: COMMU4, dtype: int64

COMMU5

2.0 2513

1.0 386

3.0 117

Name: COMMU5, dtype: int64

Non 2513

Oui 386

[Nsp] 117

Name: COMMU5, dtype: int64

COMMU6

2.0 2252

1.0 637

3.0 127

Name: COMMU6, dtype: int64

Non 2252

Oui 637

[Nsp] 127

Name: COMMU6, dtype: int64

COMMU7

2.0 2709
3.0 160
1.0 147
Name: COMMU7, dtype: int64

Non 2709
[Nsp] 160
Oui 147
Name: COMMU7, dtype: int64

COMMU8

2.0 1951
1.0 1065
Name: COMMU8, dtype: int64

Non 1951
Oui 1065
Name: COMMU8, dtype: int64

MONDIAL

2.0 4065
1.0 3874
3.0 142
Name: MONDIAL, dtype: int64

Plutôt des inconvénients 4065
Plutôt des avantages 3874
[Nsp] 142
Name: MONDIAL, dtype: int64

INQMALAD

1 4516
2 4100
3 1861
4 508
5 146
Name: INQMALAD, dtype: int64

Beaucoup 4516
Assez 4100
Un peu 1861

Pas du tout 508
 [Nsp] 146
 Name: INQMALAD, dtype: int64

INQMALA3

1 8616
 2 2369
 3 146
 Name: INQMALA3, dtype: int64

Inquiet des risques de maladie grave 8616
 Non inquiet 2369
 [Nsp] 146

Name: INQMALA3, dtype: int64

INQAGRES

2 3600
 3 3409
 1 2475
 4 1446
 5 201
 Name: INQAGRES, dtype: int64

Assez 3600
 Un peu 3409
 Beaucoup 2475
 Pas du tout 1446
 [Nsp] 201

Name: INQAGRES, dtype: int64

INQAGRE3

1 6075
 2 4855
 3 201
 Name: INQAGRE3, dtype: int64

Inquiet des risques d'agression dans la rue 6075
 Non inquiet 4855
 [Nsp] 201

Name: INQAGRE3, dtype: int64

INQROUTE

2 4261
3 3312
1 2635
4 718
5 205

Name: INQROUTE, dtype: int64

Assez 4261
Un peu 3312
Beaucoup 2635
Pas du tout 718
[Nsp] 205

Name: INQROUTE, dtype: int64

INQROUT3

1 6896
2 4030
3 205

Name: INQROUT3, dtype: int64

Inquiet des risques d'accident de la route 6896
Non inquiet 4030
[Nsp] 205

Name: INQROUT3, dtype: int64

INQCHOMA

2 3648
3 2840
1 2543
4 1900
5 200

Name: INQCHOMA, dtype: int64

Assez 3648
Un peu 2840
Beaucoup 2543
Pas du tout 1900
[Nsp] 200

Name: INQCHOMA, dtype: int64

INQCHOM3

1 6191
 2 4740
 3 200
 Name: INQCHOM3, dtype: int64

Inquiet des risques de chômage 6191
 Non inquiet 4740
 [Nsp] 200
 Name: INQCHOM3, dtype: int64

INQGUERR

3 3348
 2 3175
 1 2633
 4 1756
 5 219
 Name: INQGUERR, dtype: int64

Un peu 3348
 Assez 3175
 Beaucoup 2633
 Pas du tout 1756
 [Nsp] 219
 Name: INQGUERR, dtype: int64

INQGUER3

1 5808
 2 5104
 3 219
 Name: INQGUER3, dtype: int64

Inquiet des risques d'une guerre 5808
 Non inquiet 5104
 [Nsp] 219
 Name: INQGUER3, dtype: int64

INQNUCLE

3 3781
 2 2688
 4 2561
 1 1880

5 221
Name: INQNUCLE, dtype: int64

Un peu 3781
Assez 2688
Pas du tout 2561
Beaucoup 1880
[Nsp] 221
Name: INQNUCLE, dtype: int64

INQNUCL3

2 6342
1 4568
3 221
Name: INQNUCL3, dtype: int64

Non inquiet 6342
Inquiet des risques d'un accident de centrale nucléaire 4568
[Nsp] 221
Name: INQNUCL3, dtype: int64

INQALIM

3 4245
2 3325
4 1940
1 1412
5 209
Name: INQALIM, dtype: int64

Un peu 4245
Assez 3325
Pas du tout 1940
Beaucoup 1412
[Nsp] 209
Name: INQALIM, dtype: int64

INQALIM3

2 6185
1 4737
3 209
Name: INQALIM3, dtype: int64

Non inquiet	6185
Inquiet des risques alimentaires	4737
[Nsp]	209

Name: INQALIM3, dtype: int64

INQ4SUR6

3	6937
---	------

1	2985
---	------

2	1209
---	------

Name: INQ4SUR6, dtype: int64

Autre	6937
-------	------

Inquiet	2985
---------	------

Non inquiet	1209
-------------	------

Name: INQ4SUR6, dtype: int64

CLASSES0

4	4693
---	------

3	2880
---	------

5	2036
---	------

6	681
---	-----

2	560
---	-----

1	171
---	-----

7	110
---	-----

Name: CLASSES0, dtype: int64

La classe moyenne inférieure	4693
------------------------------	------

La classe moyenne supérieure	2880
------------------------------	------

La classe populaire	2036
---------------------	------

Les défavorisés	681
-----------------	-----

Les gens aisés	560
----------------	-----

Les privilégiés	171
-----------------	-----

[Nsp]	110
-------	-----

Name: CLASSES0, dtype: int64

ISEG0

2.0	2590
-----	------

3.0	2129
-----	------

1.0	871
-----	-----

4.0	356
-----	-----

5.0	90
-----	----

Name: ISEG0, dtype: int64

Assez souvent	2590
Rarement	2129
Très souvent	871
Jamais	356
[Nsp]	90

Name: ISEGO, dtype: int64

HEUREUX

3	5423
2	3665
4	1758
1	203
5	82

Name: HEUREUX, dtype: int64

Assez souvent	5423
Occasionnellement	3665
Très souvent	1758
Jamais	203
[Nsp]	82

Name: HEUREUX, dtype: int64

CONFGOUV

4	4069
3	3625
2	2882
1	430
5	125

Name: CONFGOUV, dtype: int64

Pas du tout confiance	4069
Plutôt pas confiance	3625
Plutôt confiance	2882
Tout à fait confiance	430
[Nsp]	125

Name: CONFGOUV, dtype: int64

EHPOL

4	2967
3	1803
5	1663

2 1418
6 1202
7 1040
8 542
1 496
Name: ECHPOL, dtype: int64

Au centre 2967
Plutôt à gauche 1803
Plutôt à droite 1663
A gauche 1418
À droite 1202
Très à droite 1040
[Nsp, NR] 542
Très à gauche 496
Name: ECHPOL, dtype: int64

OPIIMMIG

2 6013
1 4823
3 295
Name: OPIIMMIG, dtype: int64

Le départ d'un grand nombre d'immigrés 6013
L'intégration des immigrants 4823
[Nsp] 295
Name: OPIIMMIG, dtype: int64

RELIGION

4 4703
3 3306
2 2167
1 830
5 125
Name: RELIGION, dtype: int64

Ni pratique, ni sentiment d'appartenance 4703
Pas de pratique, mais sentiment d'appartenir à une religion 3306
Une pratique religieuse occasionnelle 2167
Une pratique religieuse régulière 830
[Nsp] 125
Name: RELIGION, dtype: int64

REVAUON

2.0 6522
1.0 1144
999999.0 15
Name: REVAUON, dtype: int64

Non 6522
Oui 1144
1E6 15
Name: REVAUON, dtype: int64

REV_TR7

3 2332
4 2281
6 2087
5 1675
2 1544
1 862
7 350
Name: REV_TR7, dtype: int64

De 1500 à 2300 2332
De 2300 à 3100 2281
4000 et plus 2087
3100 à 4000 1675
De 900 à 1500 1544
Moins de 900 862
Non déclaré (ne sait pas, refus) 350
Name: REV_TR7, dtype: int64

revtot7

3 2333
4 2272
6 2096
5 1681
2 1547
1 861
7 341
Name: revtot7, dtype: int64

De 1500 à 2300 2333
De 2300 à 3100 2272
4000 et plus 2096

3100 à 4000	1681
De 900 à 1500	1547
Moins de 900	861
Non déclaré (ne sait pas, refus)	341

Name: revtot7, dtype: int64

decuc

6	1267
3	1165
1	1147
9	1133
5	1095
10	1070
4	1058
7	969
2	965
8	915
11	347

Name: decuc, dtype: int64

Décile 6	1267
Décile 3	1165
Décile 1	1147
Décile 9	1133
Décile 5	1095
Décile 10	1070
Décile 4	1058
Décile 7	969
Décile 2	965
Décile 8	915
Non renseigné	347

Name: decuc, dtype: int64

decsqt

7	1289
3	1136
5	1117
10	1101
9	1089
1	1062
8	1047
2	1025
4	984
6	934

11 347
Name: decsqt, dtype: int64

Décile 7	1289
Décile 3	1136
Décile 5	1117
Décile 10	1101
Décile 9	1089
Décile 1	1062
Décile 8	1047
Décile 2	1025
Décile 4	984
Décile 6	934

Non renseigné 347
Name: decsqt, dtype: int64

TYPOCLAS

3	3420
2	2130
4	1884
1	1147
5	1133
6	1070
7	347

Name: TYPOCLAS, dtype: int64

Classes moyennes inférieures	3420
Catégories modestes	2130
Classes moyennes supérieures	1884
Catégories pauvres	1147
Catégories aisées	1133
Hauts revenus	1070
Non réponse	347

Name: TYPOCLAS, dtype: int64

TYOLOG2

1	6986
2	4145

Name: TYOLOG2, dtype: int64

Maison individuelle	6986
Logement collectif	4145

Name: TYOLOG2, dtype: int64

AGESEX12

9	1835
3	1806
8	1137
4	950
5	930
10	892
12	890
6	842
11	635
2	511
7	459
1	244

Name: AGESEX12, dtype: int64

30-49	ans, femme	1835
30-49	ans, homme	1806
20-29	ans, femme	1137
50-59	ans, homme	950
60-69	ans, homme	930
50-59	ans, femme	892
70-99	ans, femme	890
70-99	ans, homme	842
60-69	ans, femme	635
20-29	ans, homme	511
18-19	ans, femme	459
18-19	ans, homme	244

Name: AGESEX12, dtype: int64

XX

!!!!!!!!!!!! 35 modalités !!!!!!!!!!!!!!!

XX

PCSENQ36

28	1258
34	1134
23	1009
29	750
20	632
30	622
18	546
35	477
16	366
21	359
12	338

24	332
9	313
15	288
22	259
17	230
10	218
14	211
3	190
27	179
6	156
5	152
31	148
2	142
7	137
11	132
36	118
19	113
32	103
4	87
1	43
25	38
8	35
26	11
13	5

Name: PCSNQ36, dtype: int64

Ancien cadre	1258
Élève, étudiant	1134
Ouvrier qualifié	1009
Ancienne profession intermédiaire	750
Employé administratif d'entreprise	632
Ancien employé	622
Employé civil, agent de service fonction publique	546
Personne au foyer	477
Technicien	366
Employé de commerce	359
Profession intermédiaire de la santé et du travail social	338
Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et comm. des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Ancien artisan, commerçant, chef d'entreprise	179
Cadre de la fonction publique	156
Profession libérale	152

Ancien ouvrier	148
Artisan	142
Professeur, profession scientifique	137
Professeur des écoles, instituteur et assimilé	132
Autre inactif	118
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Ancien agriculteur exploitant	11
Clergé, religieux	5

Name: PCSENQ36, dtype: int64

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 !!!!!!!!!!! 26 modalités !!!!!!!!!!!!!!!
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 PCSCONJ

23.0	599
18.0	329
9.0	274
20.0	255
21.0	229
16.0	219
10.0	178
22.0	169
24.0	156
12.0	150
6.0	141
14.0	140
5.0	139
15.0	132
17.0	128
7.0	115
2.0	103
3.0	99
11.0	74
19.0	63
1.0	39
4.0	38
25.0	34
8.0	22
26.0	10
13.0	1

Name: PCSCONJ, dtype: int64

Ouvrier qualifié	599
Employé civil, agent de service fonction publique	329
Cadre administratif et commercial d'entreprise	274
Employé administratif d'entreprise	255
Employé de commerce	229
Technicien	219
Ingénieur et cadre technique d'entreprise	178
Personnel des services directs aux particuliers	169
Ouvrier non qualifié	156
Profession intermédiaire santé et travail social	150
Cadre de la fonction publique	141
Profession intermédiaire fonction publique	140
Profession libérale	139
Profession inter. adm. et com.des entreprises	132
Contremaître, agent de maîtrise	128
Professeur, profession scientifique	115
Artisan	103
Commerçant	99
Professeur des écoles, instituteur et assimilé	74
Policier et militaire	63
Agriculteur	39
Chef d'entreprise 10 salariés ou plus	38
Ouvrier agricole	34
Profession de l'information, des arts et des spectacles	22
Nsp, NR	10
Clergé, religieux	1
Name: PCSCONJ, dtype: int64	

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
!!!!!!!!!!!! 34 modalités !!!!!!!!!!!!!
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
***PCSENQ35***

```

28	1258
34	1134
23	1009
29	750
20	632
30	622
35	595
18	546
16	366
21	359
12	338
24	332
9	313

15	288
22	259
17	230
10	218
14	211
3	190
27	179
6	156
5	152
31	148
2	142
7	137
11	132
19	113
32	103
4	87
1	43
25	38
8	35
26	11
13	5

Name: PCSNQ35, dtype: int64

Ancien cadre	1258
Élève, étudiant	1134
Ouvrier qualifié	1009
Ancienne profession intermédiaire	750
Employé administratif d'entreprise	632
Ancien employé	622
Personne au foyer et autre inactif	595
Employé civil, agent de service fonction publique	546
Technicien	366
Employé de commerce	359
Profession intermédiaire santé et travail social	338
Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et comm. des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Ancien artisan, commerçant, chef d'entreprise	179
Cadre de la fonction publique	156
Profession libérale	152
Ancien ouvrier	148
Artisan	142
Professeur, profession scientifique	137

Professeur des écoles, instituteur et assimilé	132
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Ancien agriculteur exploitant	11
Clergé, religieux	5

Name: PCSENQ35, dtype: int64

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 !!!!!!!!!!! 29 modalités !!!!!!!!!!!!!!!
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 PCSENQ32

30	2968
28	1134
23	1009
20	632
27	595
18	546
16	366
21	359
12	338
24	332
9	313
15	288
22	259
17	230
10	218
14	211
3	190
6	156
5	152
2	142
7	137
11	132
19	113
32	103
4	87
1	43
25	38
8	35
13	5

Name: PCSENQ32, dtype: int64

Retraité	2968
Élève, étudiant	1134
Ouvrier qualifié	1009
Employé administratif d'entreprise	632
Personnes au foyer, a.inac.	595
Employé civil, agent de service fonction publique	546
Technicien	366
Employé de commerce	359
Profession intermédiaire santé et travail social	338
Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et com. des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Cadre de la fonction publique	156
Profession libérale	152
Artisan	142
Professeur, profession scientifique	137
Professeur des écoles, instituteur et assimilé	132
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Clergé, religieux	5
Name: PCSENQ32, dtype: int64	

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
!!!!!!!!!! 24 modalités !!!!!!!!!!!!!
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
***PCSENQ24***

```

21	2008
24	1729
16	1009
22	770
8	686
12	659
13	632
7	531
10	366
14	359
17	332

6	328
9	288
15	259
11	230
3	190
20	179
5	152
2	142
23	103
4	87
1	43
18	38
19	11

Name: PCSNQ24, dtype: int64

Ancien cadre et profession intermédiaire	2008
Inactif divers (autre que retraité)	1729
Ouvrier qualifié	1009
Ancien employé et ouvrier	770
Prof. inter. enseignement, santé, fonct. publ. et assimilé	686
Employé de la fonction publique	659
Employé administratifs d'entreprise	632
Cadre d'entreprise	531
Technicien	366
Employé de commerce	359
Ouvrier non qualifié	332
Cadre fonction publique, profession intellect. & artistique	328
Prof. inter. administrative et commerciale des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Commerçant et assimilé	190
Ancien artisan, commerçant, chef d'entreprise	179
Profession libérale et assimilé	152
Artisan	142
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise de 10 salariés ou plus	87
Agriculteur exploitant	43
Ouvrier agricole	38
Ancien agriculteur exploitant	11

Name: PCSNQ24, dtype: int64

PCSENQ15

14	2968
9	1650
7	1570
10	1341

12	1134
5	859
13	595
3	277
11	259
6	152
4	142
15	103
1	43
2	38

Name: PCSENQ15, dtype: int64

Retraité	2968
Employé	1650
Cadre moyen	1570
Ouvrier	1341
Étudiant	1134
Cadre supérieur	859
Personne au foyer	595
Commerçant, chef d'entreprise	277
Personnel de service	259
Profession libérale	152
Artisan	142
Autre inactif	103
Agriculteur exploitant	43
Ouvrier agricole	38

Name: PCSENQ15, dtype: int64

PCSENQ13

13	3071
8	1650
6	1570
9	1341
11	1134
5	1011
12	595
3	277
10	259
4	142
1	43
2	38

Name: PCSENQ13, dtype: int64

Retraité, autre inactif	3071
Employé	1650
Cadre moyen	1570

Ouvrier	1341
Étudiant	1134
Profession libérale, cadre supérieur	1011
Personne au foyer	595
Commerçant, chef d'entreprise	277
Personnel de service	259
Artisan	142
Agriculteur exploitant	43
Ouvrier agricole	38

Name: PCSENQ13, dtype: int64

PCSENQ10

9	2968
10	1832
5	1650
4	1570
7	1341
3	1011
2	419
6	259
1	43
8	38

Name: PCSENQ10, dtype: int64

Retraité	2968
Autre inactif	1832
Employé	1650
Profession Intermédiaire	1570
Ouvrier	1341
Cadre et profession intellectuelle supérieure	1011
Artisan, commerçant et chef d'entreprise	419
Personnel de service	259
Agriculteur exploitant	43
Ouvrier agricole	38

Name: PCSENQ10, dtype: int64

PCSENQ9

8	2968
5	1909
4	1570
6	1379
7	1134
3	1011
9	698

2 419
 1 43
 Name: PCSENQ9, dtype: int64

Retraité	2968
Employé	1909
Profession Intermédiaire	1570
Ouvrier	1379
Étudiant	1134
Cadre et profession intellectuelle supérieure	1011
Autre inactif	698
Artisan, commerçant et chef d'entreprise	419
Agriculteur exploitant	43

Name: PCSENQ9, dtype: int64

PCSENQ7

6 2968
 2 2581
 3 1909
 4 1379
 7 1237
 5 595
 1 462
 Name: PCSENQ7, dtype: int64

Retraité	2968
Cadre, prof. intellectuelle sup., profession intermédiaire	2581
Employé	1909
Ouvrier	1379
Autre inactif	1237
Personne au foyer	595
Agri. exploitant, artisan, commerçant et chef d'entreprise	462

Name: PCSENQ7, dtype: int64

PCSCON7

7.0 4532
 6.0 2770
 2.0 1687
 3.0 1039
 4.0 776
 1.0 272
 5.0 8
 Name: PCSCON7, dtype: int64

Sans conjoint	4532
Conjoint inactif	2770
Cadre, prof. intellectuelle sup., profession intermédiaire	1687
Employé	1039
Ouvrier	776
Agri. exploitant, artisan, commerçant et chef d'entreprise	272
Non déclaré	8

Name: PCSCON7, dtype: int64

ACM1

2	4988
1	4231
4	1710
5	133
3	69

Name: ACM1, dtype: int64

Radical	4988
Progressiste	4231
Pas de réforme	1710
Nsp si réforme	133
Nsp quelle réforme	69

Name: ACM1, dtype: int64

ACM2

3	9275
1	922
2	878
4	56

Name: ACM2, dtype: int64

Devraient travailler selon désir ou toujours	9275
Ne devraient jamais travailler, si enfants	922
Devraient travailler si besoin de 2 salaires	878
[Nsp]	56

Name: ACM2, dtype: int64

ACM3

3	5738
2	4622
1	719
4	52

Name: ACM3, dtype: int64

Charge de logement lourde ou +	5738
Charge de logement supportable	4622
Charge de logement négligeable	719
[Nsp]	52

Name: ACM3, dtype: int64

ACM4

2	6844
1	2781
3	1459
4	47

Name: ACM4, dtype: int64

Satisfait de son cadre de vie	6844
Très satisfait de son cadre de vie	2781
Pas satisfait de son cadre de vie	1459
[Nsp]	47

Name: ACM4, dtype: int64

ACM5

2	7386
3	1983
1	1681
4	81

Name: ACM5, dtype: int64

État de santé satisfaisant	7386
État de santé pas satisfaisant	1983
État de santé très satisfaisant	1681
[Nsp]	81

Name: ACM5, dtype: int64

ACM6

3	4766
4	2986
2	2331
1	875
5	173

Name: ACM6, dtype: int64

Niveau de vie Français : un peu moins bien	4766
Niveau de vie Français : beaucoup moins bien	2986
Niveau de vie Français : pareil	2331
Niveau de vie Français : mieux	875
[Nsp]	173

Name: ACM6, dtype: int64

ACM7

2	3661
3	3591
1	2178
4	1604
5	97

Name: ACM7, dtype: int64

CDV dans 5 ans : semblables	3661
CDV dans 5 ans : détériorer un peu	3591
CDV dans 5 ans : améliorer	2178
CDV dans 5 ans : détériorer beaucoup	1604
[Nsp]	97

Name: ACM7, dtype: int64

ACM8

2	5365
1	3286
3	2379
4	101

Name: ACM8, dtype: int64

Fonctionnement justice : assez mal	5365
Fonctionnement justice : bien	3286
Fonctionnement justice : très mal	2379
[Nsp]	101

Name: ACM8, dtype: int64

ACM9

4	3293
1	2986
2	2612
3	2240

Name: ACM9, dtype: int64

Nombre de maux : trois ou +	3293
Nombre de maux : aucun	2986
Nombre de maux : un	2612
Nombre de maux : deux	2240

Name: ACM9, dtype: int64

ACM10

3	3445
4	3248
5	1980
2	1727
1	613
6	118

Name: ACM10, dtype: int64

Évolution niveau de vie perso : pareil	3445
Évolution niveau de vie perso : un peu moins bien	3248
Évolution niveau de vie perso : beaucoup moins bien	1980
Évolution niveau de vie perso : un peu mieux	1727
Évolution niveau de vie perso : beaucoup mieux	613
[Nsp]	118

Name: ACM10, dtype: int64

ACM11

1	6454
2	4571
3	106

Name: ACM11, dtype: int64

Restrictions : oui	6454
Restrictions : non	4571
[Nsp]	106

Name: ACM11, dtype: int64

ACM12

1	6089
2	4994
3	48

Name: ACM12, dtype: int64

Famille seul endroit où l'on se sente bien : oui	6089
Famille seul endroit où l'on se sente bien : non	4994

[Nsp]
Name: ACM12, dtype: int64

48

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
!!!!!!!!!!!! 23 modalités !!!!!!!!!!!!!
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
UDA23

12	1988
22	1120
21	841
17	716
18	646
6	619
2	584
7	486
16	478
13	472
15	463
19	303
5	295
11	288
20	279
1	268
4	261
23	239
8	225
3	212
10	198
14	112
9	38

Name: UDA23, dtype: int64

Ile de France	1988
Rhône-Alpes	1120
Provence Alpes Côte d'Azur	841
Nord Pas de Calais	716
Pays de la Loire	646
Bretagne	619
Aquitaine	584
Centre	486
Midi-Pyrénées	478
Languedoc-Roussillon	472
Lorraine	463
Picardie	303
Bourgogne	295
Haute-Normandie	288

Poitou-Charente	279
Alsace	268
Basse-Normandie	261
DOM	239
Champagne-Ardenne	225
Auvergne	212
Franche-Comté	198
Limousin	112
Corse	38

Name: UDA23, dtype: int64

UDA14

2	1988
12	1332
6	1019
10	975
7	956
11	950
13	841
8	646
9	619
5	549
4	493
3	486
1	239
14	38

Name: UDA14, dtype: int64

Ile de France	1988
Auvergne - Rhône Alpes	1332
Nord Pas de Calais - Picardie	1019
Aquitaine - Limousin - Poitou Charente	975
Alsace - Champagne - Ardenne Lorraine	956
Languedoc Roussillon - Midi Pyrénées	950
Provence Alpes Côte d'Azur	841
Pays de la Loire	646
Bretagne	619
Normandie	549
Bourgogne - Franche-Comté	493
Centre Val de Loire	486
DOM	239
Corse	38

Name: UDA14, dtype: int64

statut99

NO DIFFERENCE

RURAUURBA

NO DIFFERENCE

zau1999

1.0 6930

2.0 1705

6.0 1195

4.0 615

3.0 545

5.0 36

Name: zau1999, dtype: int64

Commune pôle urbain 6930

Commune monopolarisée 1705

Commune espace dominante rurale 1195

Commune pole emploi espace rural 615

Commune multipolarisée 545

Commune couronne pole emploi espace rural 36

Name: zau1999, dtype: int64

zau2010

111.0 6770

112.0 1897

300.0 541

120.0 526

221.0 456

400.0 414

211.0 340

212.0 44

222.0 22

Name: zau2010, dtype: int64

Commune appartenant à grand pôle (10 000 emplois ou +) 6770

Commune appartenant à couronne d'un grand pôle 1897

Autre commune multipolarisée 541

Commune multipolarisée des grandes aires urbaines 526

Commune appartenant à petit pôle (de 1.500 à 5 000 emplois) 456

Commune isolée hors influence des pôles 414

Commune appartenant à moyen pôle (5 à 10 000 emplois) 340

Commune appartenant à couronne d'un moyen pôle 44

Commune appartenant à couronne d'un petit pôle 22

Name: zau2010, dtype: int64

AUTREAL

1.0 53

4.0 48

2.0 22

3.0 13

Name: AUTREAL, dtype: int64

Voyage, vacances 53

Autre (mariage, études, développement personnel, artistique) 48

Travaux, amélioration de l'habitat 22

Achat d'un véhicule automobile 13

Name: AUTREAL, dtype: int64

UDA5

2.0 713

5.0 695

3.0 666

1.0 531

4.0 326

Name: UDA5, dtype: int64

Nord Ouest 713

Sud Est 695

Nord Est 666

Ile de France 531

Sud Ouest 326

Name: UDA5, dtype: int64

inseel

NO DIFFERENCE

couple2

1.0 1837

3.0 919

2.0 260

Name: couple2, dtype: int64

Oui, avec une personne qui vit dans le logement 1837

Non 919

Oui, avec une personne qui ne vit pas dans le logement 260

Name: couple2, dtype: int64

AGE6

4.0 3216

3.0 1796

6.0 1397

5.0 1250

2.0 1147

1.0 280

Name: AGE6, dtype: int64

40 à 59 ans 3216

25 à 39 ans 1796

70 ans et plus 1397

60 à 69 ans 1250

18 à 24 ans 1147

Moins de 18 ans 280

Name: AGE6, dtype: int64

PCSRED10

9.0 1572

5.0 858

4.0 845

7.0 779

8.0 650

3.0 546

10.0 381

2.0 223

6.0 158

1.0 24

Name: PCSRED10, dtype: int64

Retraité 1572

Employé 858

Profession Intermédiaire 845

Ouvrier 779

Étudiant 650

Cadre et profession intellectuelle supérieure 546

Autre inactif 381

Artisan, commerçant et chef d'entreprise 223

Personnel de service 158

Agriculteur exploitant 24

Name: PCSRED10, dtype: int64

info
NO DIFFERENCE

com
NO DIFFERENCE

type99
NO DIFFERENCE

DEPCOM
NO DIFFERENCE

QUOTAAGE

3.0	981
4.0	491
6.0	471
2.0	434
5.0	416
1.0	227

Name: QUOTAAGE, dtype: int64

30 à 49 ans	981
50 à 59 ans	491
70 ans et plus	471
20 à 29 ans	434
60 à 69 ans	416
Moins de 20 ans	227

Name: QUOTAAGE, dtype: int64

PRIVPUB

1.0	3633
2.0	1706

Name: PRIVPUB, dtype: int64

Privé	3633
Public	1706

Name: PRIVPUB, dtype: int64

interim2

2.0 3416
1.0 217
Name: interim2, dtype: int64

Non 3416
Oui 217
Name: interim2, dtype: int64

EMP7

1 5889
5 2972
2 955
7 744
3 417
4 154
Name: EMP7, dtype: int64

Actif 5889
Retraité 2972
Etudiant 955
Chomeur 744
Reste au foyer 417
Malade, invalide 154
Name: EMP7, dtype: int64

typcont2

2.0 4346
1.0 992
3.0 1
Name: typcont2, dtype: int64

À durée indéterminée 4346
À durée déterminée 992
[Nsp] 1
Name: typcont2, dtype: int64

```
In [22]: print(f"Includes {len(scope_no_diff)} other identical variables :\n\
{' '.join(scope_no_diff)}")
scope_identical = scope_identical.union(pd.Index(scope_no_diff))
```

```
print(f"total number of identical variables {len(scope_identical)}")
```

Includes 9 other identical variables :

identifiant RADIQUEOI statut99 RURAURBA inseel info com type99 DEPCOM

total number of identical variables 26

```
In [23]: print(f"variable with more than 15 modalities :\n{scope_obj_long}")
```

variable with more than 15 modalities :

['PCSENQ36', 'PCSCONJ', 'PCSENQ35', 'PCSENQ32', 'PCSENQ24', 'UDA23']

```
In [24]: df_tmp = cdv.loc[:,scope_obj_long].dropna()
df_tmp.head()
```

```
Out[24]:
```

	PCSENQ36 \
9	Chef d'entreprise 10 salariés ou plus
12	Ancien employé
18	Profession inter. adm. et comm. des entreprises
20	Technicien
32	Contremaître, agent de maîtrise

	PCSCONJ \
9	Profession libérale
12	Employé civil, agent de service fonction publique
18	Cadre administratif et commercial d'entreprise
20	Contremaître, agent de maîtrise
32	Technicien

	PCSENQ35 \
9	Chef d'entreprise 10 salariés ou plus
12	Ancien employé
18	Profession inter. adm. et comm. des entreprises
20	Technicien
32	Contremaître, agent de maîtrise

	PCSENQ32 \
9	Chef d'entreprise 10 salariés ou plus
12	Retraité
18	Profession inter. adm. et com. des entreprises
20	Technicien
32	Contremaître, agent de maîtrise

	PCSENQ24 \
9	Chef d'entreprise de 10 salariés ou plus
12	Ancien employé et ouvrier
18	Prof. inter. administrative et commerciale des...
20	Technicien

```
32                               Contremaître, agent de maîtrise
```

```
                                UDA23
```

```
9   Provence Alpes Côte d'Azur
12                                Franche-Comté
18                                Bretagne
20                                Lorraine
32                                Rhône-Alpes
```

```
In [25]: cdv['PCSENQ36'].value_counts()
```

```
Out[25]: Ancien cadre                                1258
         Élève, étudiant                             1134
         Ouvrier qualifié                             1009
         Ancienne profession intermédiaire             750
         Employé administratif d'entreprise            632
         Ancien employé                               622
         Employé civil, agent de service fonction publique 546
         Personne au foyer                             477
         Technicien                                    366
         Employé de commerce                          359
         Profession intermédiaire de la santé et du travail social 338
         Ouvrier non qualifié                          332
         Cadre administratif et commercial d'entreprise 313
         Profession inter. adm. et comm. des entreprises 288
         Personnel des services directs aux particuliers 259
         Contremaître, agent de maîtrise               230
         Ingénieur et cadre technique d'entreprise     218
         Profession intermédiaire de la fonction publique 211
         Commerçant                                    190
         Ancien artisan, commerçant, chef d'entreprise 179
         Cadre de la fonction publique                 156
         Profession libérale                           152
         Ancien ouvrier                               148
         Artisan                                       142
         Professeur, profession scientifique           137
         Professeur des écoles, instituteur et assimilé 132
         Autre inactif                                118
         Policier et militaire                        113
         Chômeur n'ayant jamais travaillé            103
         Chef d'entreprise 10 salariés ou plus         87
         Agriculteur                                   43
         Ouvrier agricole                             38
         Profession de l'information, des arts et des spectacles 35
         Ancien agriculteur exploitant                 11
         Clergé, religieux                             5
         Name: PCSENQ36, dtype: int64
```

```
In [26]: cdv['UDA23'].value_counts()
```

```

Out[26]: Ile de France          1988
        Rhône-Alpes           1120
        Provence Alpes Côte d'Azur  841
        Nord Pas de Calais        716
        Pays de la Loire          646
        Bretagne                 619
        Aquitaine                 584
        Centre                   486
        Midi-Pyrénées            478
        Languedoc-Roussillon      472
        Lorraine                 463
        Picardie                 303
        Bourgogne                295
        Haute-Normandie          288
        Poitou-Charente          279
        Alsace                   268
        Basse-Normandie          261
        DOM                      239
        Champagne-Ardennes       225
        Auvergne                 212
        Franche-Comté           198
        Limousin                 112
        Corse                    38
        Name: UDA23, dtype: int64

```

2) variable with difference between cdv & cdv_ssfmt with other type

```

In [27]: print(" ".join(scope_notobj))

```

```

ANNEEFUZ nbheures NBENF NBPIECES SENSIENV NBPERS AGE_2 AGE_3 AGE_4 AGE_5 AGE_6 AGE_7 AGE_8 AGE_9

```

```

In [28]: cdv.loc[:,scope_notobj].head()

```

```

Out[28]:   ANNEEFUZ  nbheures  NBENF  NBPIECES  SENSIENV  NBPERS  AGE_2  AGE_3  AGE_4  \
0      2015      35.0    NaN      7.0      7.0      1.0    NaN    NaN    NaN
1      2015      NaN     3.0      5.0      6.0      2.0   61.0    NaN    NaN
2      2015      NaN     3.0      8.0      4.0      2.0   82.0    NaN    NaN
3      2015      NaN    NaN      1.0      6.0      1.0    NaN    NaN    NaN
4      2015      NaN     5.0      7.0      5.0      2.0   60.0    NaN    NaN

```

```

      AGE_5  ...  CP  TYPLOG  inseenum  cpt  prescaf  refus2  med  i  AGGLOINS  \
0    NaN  ...  NaN    NaN      NaN  NaN      NaN    NaN  NaN  NaN    NaN
1    NaN  ...  NaN    NaN      NaN  NaN      NaN    NaN  NaN  NaN    NaN
2    NaN  ...  NaN    NaN      NaN  NaN      NaN    NaN  NaN  NaN    NaN
3    NaN  ...  NaN    NaN      NaN  NaN      NaN    NaN  NaN  NaN    NaN
4    NaN  ...  NaN    NaN      NaN  NaN      NaN    NaN  NaN  NaN    NaN

```

```

pmun

```

```

0    NaN
1    NaN
2    NaN
3    NaN
4    NaN

```

[5 rows x 47 columns]

```
In [29]: cdv_ssfmt.loc[:,scope_notobj].head()
```

```

Out [29]:   ANNEEFUZ  nbheures  NBENF  NBPIECES  SENSIENV  NBPERS  AGE_2  AGE_3  AGE_4  \
0         37        35.0    NaN         7.0         7.0     1.0    NaN    NaN    NaN
1         37         NaN     3.0         5.0         6.0     2.0   61.0    NaN    NaN
2         37         NaN     3.0         8.0         4.0     2.0   82.0    NaN    NaN
3         37         NaN    NaN         1.0         6.0     1.0    NaN    NaN    NaN
4         37         NaN     5.0         7.0         5.0     2.0   60.0    NaN    NaN

```

```

      AGE_5  ...  CP  TYOLOG  inseenum  cpt  prescaf  refus2  med  i  AGGLOINS  \
0    NaN  ...  NaN    NaN        NaN  NaN    NaN    NaN    NaN  NaN    NaN
1    NaN  ...  NaN    NaN        NaN  NaN    NaN    NaN    NaN  NaN    NaN
2    NaN  ...  NaN    NaN        NaN  NaN    NaN    NaN    NaN  NaN    NaN
3    NaN  ...  NaN    NaN        NaN  NaN    NaN    NaN    NaN  NaN    NaN
4    NaN  ...  NaN    NaN        NaN  NaN    NaN    NaN    NaN  NaN    NaN

```

```

      pmun
0    NaN
1    NaN
2    NaN
3    NaN
4    NaN

```

[5 rows x 47 columns]

```

In [30]: scope_notobj_diff = []
scope_notobj_identical = []
for c in scope_notobj:
    cdv_c_dropna = cdv[c].dropna()
    cdv_ssfmt_c_dropna = cdv_ssfmt[c].dropna()
    if np.sum(cdv_c_dropna != cdv_ssfmt_c_dropna) > 0:
        scope_notobj_diff.append(c)
        print(f"{c} : difference")
    else:
        print(f"{c} : identical")
        scope_notobj_identical.append(c)

```

```

ANNEEFUZ : difference
nbheures : identical
NBENF : identical
NBPIECES : identical

```

SENSIENV : identical
 NBPERS : identical
 AGE_2 : identical
 AGE_3 : identical
 AGE_4 : identical
 AGE_5 : identical
 AGE_6 : identical
 AGE_7 : identical
 AGE_8 : identical
 AGE_9 : identical
 REVPF : identical
 NOT_FAMI : identical
 NOT_PROF : identical
 NOT_AMIS : identical
 NOT_COHE : identical
 NOT_POLI : identical
 NOT_LIBR : identical
 NOT_LOG : identical
 NOT_CAD : identical
 REVCONJ : identical
 REVAUTR : identical
 REVTOT : identical
 NBUC : identical
 REVUC : identical
 REVsq : identical
 typodeg : identical
 popdense : identical
 popinter : identical
 poppeud : identical
 poptrpeu : identical
 agedip : identical
 age_OW : identical
 CSP6 : identical
 CP : identical
 TYPLOG : identical
 inseenum : identical
 cpt : identical
 prescaf : identical
 refus2 : identical
 med : identical
 i : identical
 AGGLOINS : identical
 pmun : identical

```

In [31]: print(f"Includes {len(scope_notobj_identical)} other identical variables :\n\n\
{' '.join(scope_notobj_identical)}")
scope_identical = scope_identical.union(pd.Index(scope_notobj_identical))
  
```

Includes 46 other identical variables :

nbheures NBENF NBPIECES SENSIENV NBPERS AGE_2 AGE_3 AGE_4 AGE_5 AGE_6 AGE_7 AGE_8 AGE_9 REVPF NO

```
In [32]: scope_difference = pd.Index(scope_notobj_diff).union(pd.Index(scope_obj_diff))
```

```
In [33]: print(f"Total number of identical variables : {len(scope_identical)}")
         print(f"Total number of variables with difference : {len(scope_difference)}")
```

Total number of identical variables : 72

Total number of variables with difference : 282

Conclusion : The 2 dataset are almost identical. They differ only by encoding on 282 variables. The other 72 variables are totally identical.

1.2.2 b) First analysis regardless of the year of the study

```
In [34]: print(f"Number of records: {cdv.shape[0]}")
         print(f"Number of variables: {cdv.shape[1]}")
```

Number of records: 11131

Number of variables: 354

```
In [35]: print(f"List of {cdv.shape[1]} variables names:\n")
         print(" ".join(cdv.columns))
```

List of 354 variables names:

INTER6 INTER ANNEEFUZ ANNEFUZ2 COLLECTE CHAMP POND identifiant SEXE AGE5 PCSENQ8 TYPOSQT DIPL4 A

```
In [36]: print(f"Number of lines without missing values : \
              {cdv.dropna().shape[0]} out of {cdv.shape[0]}")
```

Number of lines without missing values : 0 out of 11131

```
In [37]: nb_missing_per_var = np.sum(cdv.isnull())
```

```
In [38]: print("Number of missing values per variables :")
         nb_missing_per_var.sort_values(ascending=False).head(50)
```

Number of missing values per variables :

```

Out [38]: prescaf      11130
          SEXE_9       11124
          AGE_9        11123
          LIEN_9       11123
          SEXE_8       11107
          LIEN_8       11106
          AGE_8        11106
          SEXE_7       11078
          AGE_7        11077
          LIEN_7       11076
          AUTREAL      10995
          SEXE_6       10920
          AGE_6        10909
          LIEN_6       10906
          SEXE_5       10300
          LIEN_5       10228
          AGE_5        10226
          RADQUOI      10146
          RADWHY3       9959
          RADWHY9       9959
          RADWHY2       9959
          RADWHY4       9959
          RADWHY1       9959
          RADWHY7       9959
          RADWHY8       9959
          RADWHY5       9959
          RADWHY10      9959
          RADWHY11      9959
          RADWHY12      9959
          RADWHY13      9959
          RADWHY14      9959
          RADWHY6       9959
          WHY LIM       9474
          REVAUTR       9101
          SEXE_4       8800
          AGE_4        8525
          LIEN_4       8518
          med          8232
          UDA5         8200
          COMMU4       8115
          RAD I3       8115
          COMMU1       8115
          COMMU3       8115
          RAD I1       8115
          COMMU5       8115
          COMMU6       8115
          COMMU7       8115
          COMMU8       8115

```



```
RADI2      8115
LIMVIAND    8115
dtype: int64
```

```
In [39]: n_complete = len(nb_missing_per_var[nb_missing_per_var == 0])
n_uncomplete = len(nb_missing_per_var[nb_missing_per_var != 0])
print(f"Number of variables without missing values :\n{n_complete} out of {cdv.shape[1]} variable")
print(f"Number of variables with at least one missing values :\n{n_uncomplete} out of {cdv.shape[1]} variable")
```

```
Number of variables without missing values :193 out of 354 variable
Number of variables with at least one missing values :161 out of 354 variable
```

```
In [40]: complete_variables = nb_missing_per_var[nb_missing_per_var == 0].index
uncomplete_variables = nb_missing_per_var[nb_missing_per_var != 0].index
print(f"List of {n_complete} variables without missing values names:\n")
print(" ".join(complete_variables))
print(f"\nList of {n_uncomplete} variables with at least 1 missing value:\n")
print(" ".join(uncomplete_variables))
```

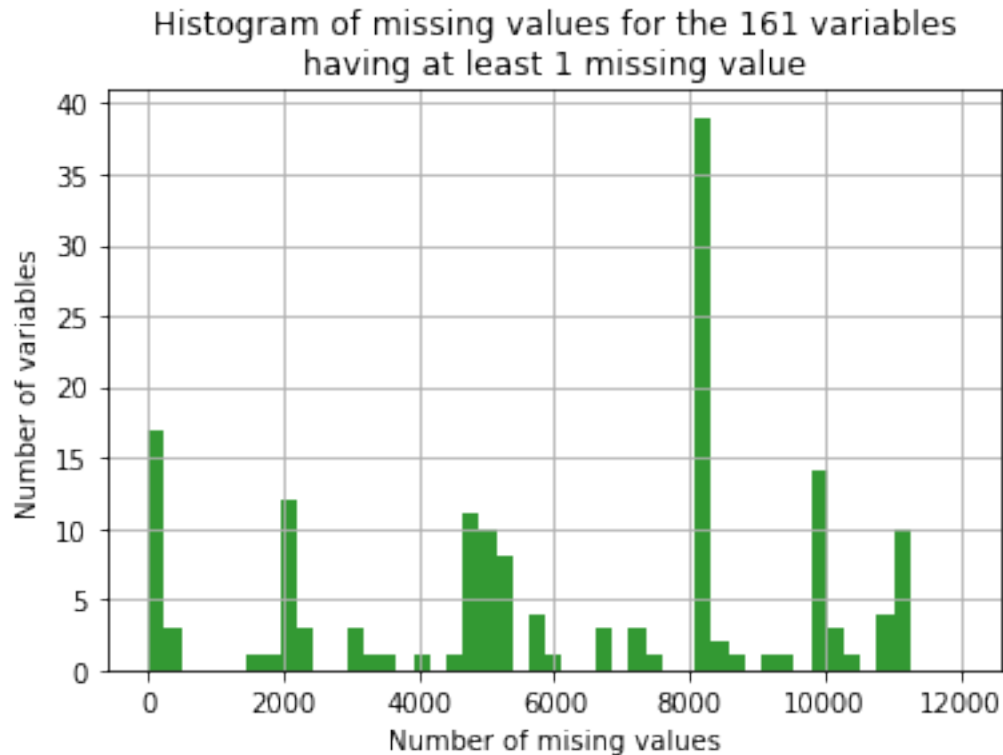
```
List of 193 variables without missing values names:
```

```
INTER6 INTER ANNEFUZ ANNEFUZ2 COLLECTE CHAMP POND SEXE AGE5 PCSENQ8 TYPOSQT DIPL4 AGGLO5 UDA10
```

```
List of 161 variables with at least 1 missing value:
```

```
identifiant SALCOMP TYPEMPL INTERIM TYPCONT TEMPSTRA nbheures NBHEUR39 NBHEUR35 PREFPALI SALCOMP
```

```
In [41]: fig=plt.figure()
plt.title("Histogram of missing values for the 161 variables\n\
having at least 1 missing value")
plt.ylabel(u'Number of variables')
plt.xlabel("Number of missing values")
bins = np.linspace(0, 12000,50)
plt.hist(nb_missing_per_var[uncomplete_variables],
        bins, facecolor='g', alpha=0.8)
plt.grid()
```



1.2.3 b) Year of realisation of the study and missing values

Variables **ANNEEFUZ** & **ANNEFUZ2** seems equivalent

According to the authors of the study :

En **2015**, l'enquête a été menée à la fois en face-à-face (**2 000 personnes** interrogées) et aussi online (**2 000 personnes** également) tous **âgés de 18 ans et plus**, résidant en **France métropolitaine (hors Corse)**. Seuls les 2000 enregistrements correspondants à l'enquête online sont présents dans le dataset.

A **partir de 2016**, le mode de collecte est passé en ligne et on interroge désormais **3 000 individus âgés de 15 ans et plus en France entière** (France métropolitaine, Corse et DOM-TOM).

```
In [42]: cdv["ANNEEFUZ"].unique()
```

```
Out[42]: array([2015, 2016, 2017, 2018])
```

```
In [43]: cdv["ANNEFUZ2"].unique()
```

```
Out[43]: array(['2015 online', '2016', '2017', '2018'], dtype=object)
```

```
In [44]: nb_enregistrements_anneefuz = cdv["ANNEEFUZ"].value_counts().sort_values(ascending = F
print("Number of records per year 'ANNEEFUZ':")
nb_enregistrements_anneefuz
```

Number of records per year 'ANNEEFUZ':

```
Out[44]: 2016    3050
         2017    3020
         2018    3016
         2015    2045
         Name: ANNEEFUZ, dtype: int64
```

```
In [45]: nb_enregistrements_annefuz2 = cdv["ANNEFUZ2"].value_counts()
         print("Number of records per year 'ANNEFUZ2':")
         nb_enregistrements_annefuz2
```

Number of records per year 'ANNEFUZ2':

```
Out[45]: 2016            3050
         2017            3020
         2018            3016
         2015 online    2045
         Name: ANNEFUZ2, dtype: int64
```

```
In [46]: B = cdv.ANNEEFUZ.astype(str)
         R = cdv.loc[B != cdv["ANNEFUZ2"], ["ANNEEFUZ", "ANNEFUZ2"]]
         print(R["ANNEFUZ2"].unique())
         print(R["ANNEEFUZ"].unique())
```

```
['2015 online']
[2015]
```

```
In [47]: # number of missing value per variable for a given year
         na_2015 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2015].isnull())
         na_2016 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2016].isnull())
         na_2017 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2017].isnull())
         na_2018 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2018].isnull())
```

```
In [48]: complete_2015 = set(na_2015[na_2015==0].index)
         complete_2016 = set(na_2016[na_2016==0].index)
         complete_2017 = set(na_2017[na_2017==0].index)
         complete_2018 = set(na_2018[na_2018==0].index)
```

```
In [49]: print(f"Number of variable without any missing values in 2015: {len(complete_2015)}")
         print(f"Number of variable without any missing values in 2016: {len(complete_2016)}")
         print(f"Number of variable without any missing values in 2017: {len(complete_2017)}")
         print(f"Number of variable without any missing values in 2018: {len(complete_2018)}")
```

Number of variable without any missing values in 2015: 199
Number of variable without any missing values in 2016: 224
Number of variable without any missing values in 2017: 224
Number of variable without any missing values in 2018: 257

```
In [50]: missing_2015 = set(na_2015[na_2015==2045].index)
missing_2016 = set(na_2016[na_2016==3050].index)
missing_2017 = set(na_2017[na_2017==3020].index)
missing_2018 = set(na_2018[na_2018==3016].index)
```

```
In [51]: print(f"Number of variable totally missing in 2015: {len(missing_2015)}")
print(f"Number of variable totally missing in 2016: {len(missing_2016)}")
print(f"Number of variable totally missing in 2017: {len(missing_2017)}")
print(f"Number of variable totally missing in 2018: {len(missing_2018)}")
```

```
Number of variable totally missing in 2015: 82
Number of variable totally missing in 2016: 73
Number of variable totally missing in 2017: 56
Number of variable totally missing in 2018: 1
```

```
In [52]: full_scope = set(cdv.columns)
scope_2015 = full_scope - missing_2015
scope_2016 = full_scope - missing_2016
scope_2017 = full_scope - missing_2017
scope_2018 = full_scope - missing_2018
```

```
In [53]: print(f"Number of variable used 2015: {len(scope_2015)}")
print(f"Number of variable used 2016: {len(scope_2016)}")
print(f"Number of variable used 2017: {len(scope_2017)}")
print(f"Number of variable used 2018: {len(scope_2018)}")
```

```
Number of variable used 2015: 272
Number of variable used 2016: 281
Number of variable used 2017: 298
Number of variable used 2018: 353
```

Synthesis of variable evolution over the period

```
In [54]: print(f"2016 vs 2015\n\tNew variable ({len(scope_2016 - scope_2015)}):")
print(" ".join(scope_2016 - scope_2015))
print(f"\tVariable dropped ({len(scope_2015 - scope_2016)}):")
print(" ".join(scope_2015 - scope_2016))
print(f"\n2017 vs 2016\n\tNew variable ({len(scope_2017 - scope_2016)}):")
print(" ".join(scope_2017 - scope_2016))
print(f"\tVariable dropped ({len(scope_2016 - scope_2017)}):")
print(" ".join(scope_2016 - scope_2017))
print(f"\n2018 vs 2017\n\tNew variable ({len(scope_2018 - scope_2017)}):")
print(" ".join(scope_2018 - scope_2017))
print(f"\tVariable dropped ({len(scope_2017 - scope_2018)}):")
print(" ".join(scope_2017 - scope_2018))
```

```
2016 vs 2015
New variable (13):
```

```
CONFMEFI CONFECOL CONFKEUF insee1 CONFPRES OPICULT PRATCOLL CONFBANK AGE6 STATLOGB CONFWEB AGGLO
Variable dropped (4):
MONDIAL VISITFAM REVAUON RECEP
```

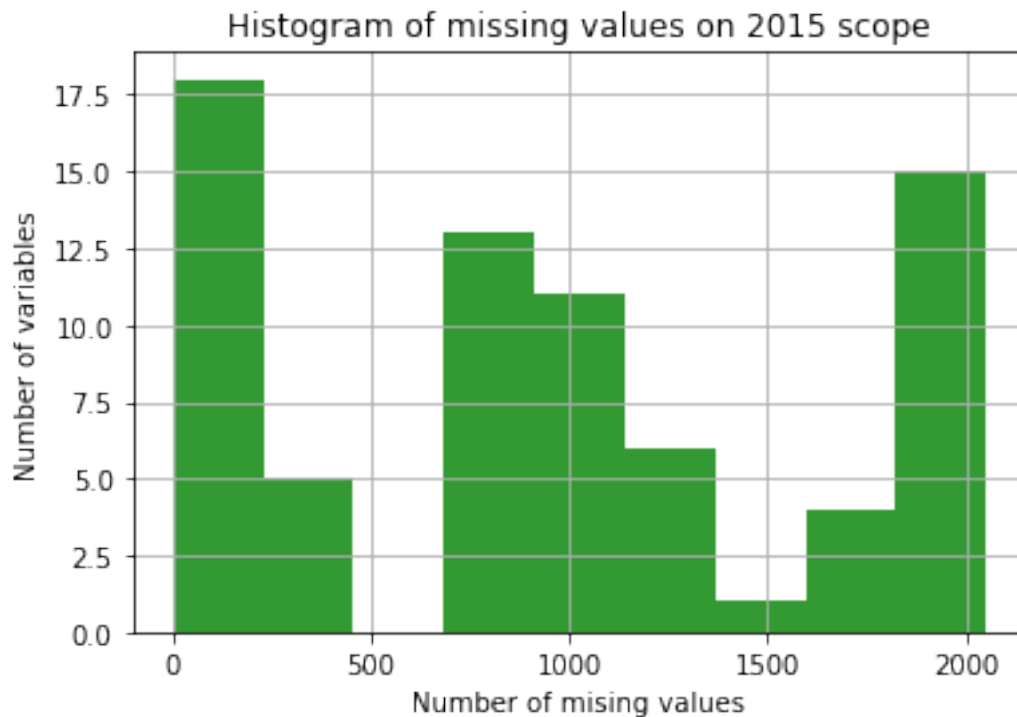
2017 vs 2016

```
New variable (17):
pmun type99 poppeud typodeg MONDIAL ISEGO PCSRED10 DEPCOM popdense RECEP VISITFAM poptrpeu REVAU
Variable dropped (0):
```

2018 vs 2017

```
New variable (56):
RADWHY4 RADWHY5 COMMU4 med ADNORDI COMMU7 ROBOT1 inseeenum ADNCB AUTREAL couple2 TYPLOG info RELE
Variable dropped (1):
QUOTAAGE
```

```
In [55]: nb_missing_per_var_2015 = np.sum(cdv.loc[cdv["ANNEEFUZ"]==2015].isnull())
fig=plt.figure()
plt.title("Histogram of missing values on 2015 scope")
plt.ylabel(u'Number of variables')
plt.xlabel("Number of missing values")
bins = np.linspace(0, 2050,10)
plt.hist(nb_missing_per_var_2015[scope_2015 - complete_2015],
        bins, facecolor='g', alpha=0.8)
plt.grid()
```



```
In [56]: nb_missing_per_var_2015[scope_2015 - complete_2015].sort_values(ascending=False)
```

```
Out[56]: SEXE_9      2043
        LIEN_9      2043
        AGE_9       2043
        AGE_8       2040
        LIEN_8      2040
        SEXE_8      2040
        SEXE_7      2038
        AGE_7       2038
        LIEN_7      2038
        SEXE_6      2014
        LIEN_6      2014
        AGE_6       2014
        LIEN_5      1886
        AGE_5       1886
        SEXE_5      1886
        REVAUTR     1715
        SEXE_4      1601
        AGE_4       1601
        LIEN_4      1601
        interim2    1405
        SALCOMPC    1321
        PCSCONJ     1321
        LIEN_3      1289
        SEXE_3      1289
        AGE_3       1289
        SALCOMPI    1236
        TYPCONT     1089
        typcont2    1063
        TYPEEMPL    1063
        INTERIM     1063
        ...
        RE_ENF      794
        RE_VAC      794
        RE_HABI     794
        RE_MEDI     794
        ACTCONJ     737
        REVCONJ     737
        NBENF       687
        REVAUON     400
        SEXE_2      400
        AGE_2       400
        LIEN_2      400
        PROGRAD     303
```

zau1999	105
statut99	105
zau2010	103
RURAUURBA	93
REVsq	63
REVUC	63
REVTOT	57
NOT_PROF	25
SENSIENV	24
NOT_COHE	23
NOT_AMIS	23
NOT_POLI	20
NOT_LIBR	19
NOT_FAMI	11
NBPERS	9
SITUFAM	9
NBUC	9
NBPIECES	9

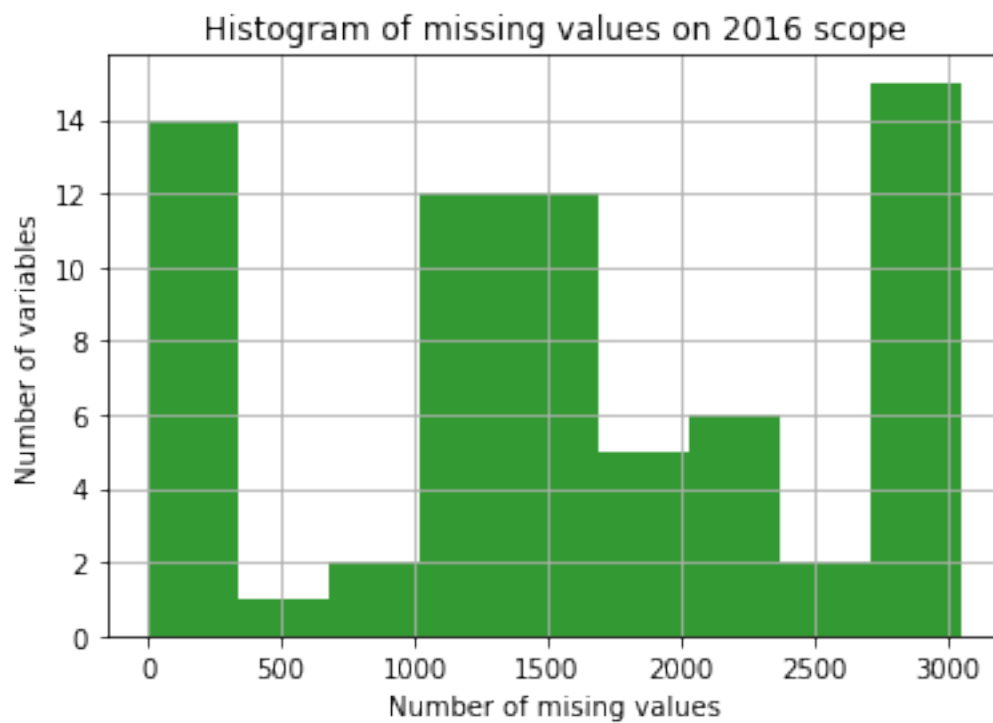
Length: 73, dtype: int64

```
In [57]: print("List of variable with more than 75% missing values in 2015:\n")
l = nb_missing_per_var_2015[scope_2015 -
                             complete_2015][nb_missing_per_var_2015 >
                                                0.75*2045]
print(" ".join(l.index))
```

List of variable with more than 75% missing values in 2015:

SEXE_9 LIEN_9 AGE_6 LIEN_6 LIEN_7 SEXE_8 SEXE_6 LIEN_8 AGE_8 AGE_4 AGE_7 SEXE_7 AGE_9 SEXE_4 LIE

```
In [58]: nb_missing_per_var_2016 = np.sum(cdv.loc[cdv["ANNEEFUZ"]==2016].isnull())
fig=plt.figure()
plt.title("Histogram of missing values on 2016 scope")
plt.ylabel(u'Number of variables')
plt.xlabel("Number of missing values")
bins = np.linspace(0, 3050,10)
plt.hist(nb_missing_per_var_2016[scope_2016 - complete_2016],
         bins, facecolor='g', alpha=0.8)
plt.grid()
```



```
In [59]: nb_missing_per_var_2016[scope_2016
        - complete_2016].sort_values(ascending=False)
```

```
Out [59]: SEXE_9      3048
          LIEN_9      3047
          AGE_9       3047
          SEXE_8      3045
          AGE_8       3044
          LIEN_8      3044
          SEXE_7      3040
          AGE_7       3039
          LIEN_7      3038
          SEXE_6      3010
          AGE_6       2999
          LIEN_6      2996
          SEXE_5      2881
          LIEN_5      2809
          AGE_5       2807
          SEXE_4      2630
          SEXE_3      2388
          AGE_4       2355
          LIEN_4      2348
          REVAUTR     2164
          interim2    2120
```


PCSCONJ	2079
SALCOMPC	2079
AGE_3	1901
LIEN_3	1891
SALCOMPI	1884
SEXE_2	1804
TYPCONT	1737
PRIVPUB	1677
typcont2	1677
...	
REVCONJ	1376
ACTCONJ	1340
RE_ALIM	1241
RE_WEB	1241
RE_VOIT	1241
RE_HABI	1241
RE_VAC	1241
RE_MEDI	1241
RE_EQUI	1241
RE_ENF	1241
RE_LOG	1241
RE_TABAL	1241
NBENF	1178
AGE_2	809
LIEN_2	726
PROGRAD	497
inseel	95
REVUC	74
REVsq	74
REVTOT	74
NOT_PROF	35
NOT_COHE	34
NOT_FAMI	33
NOT_AMIS	32
NOT_POLI	32
NOT_LIBR	29
PRATCOLL	27
zau2010	18
SENSIENV	15
RURAUABA	4

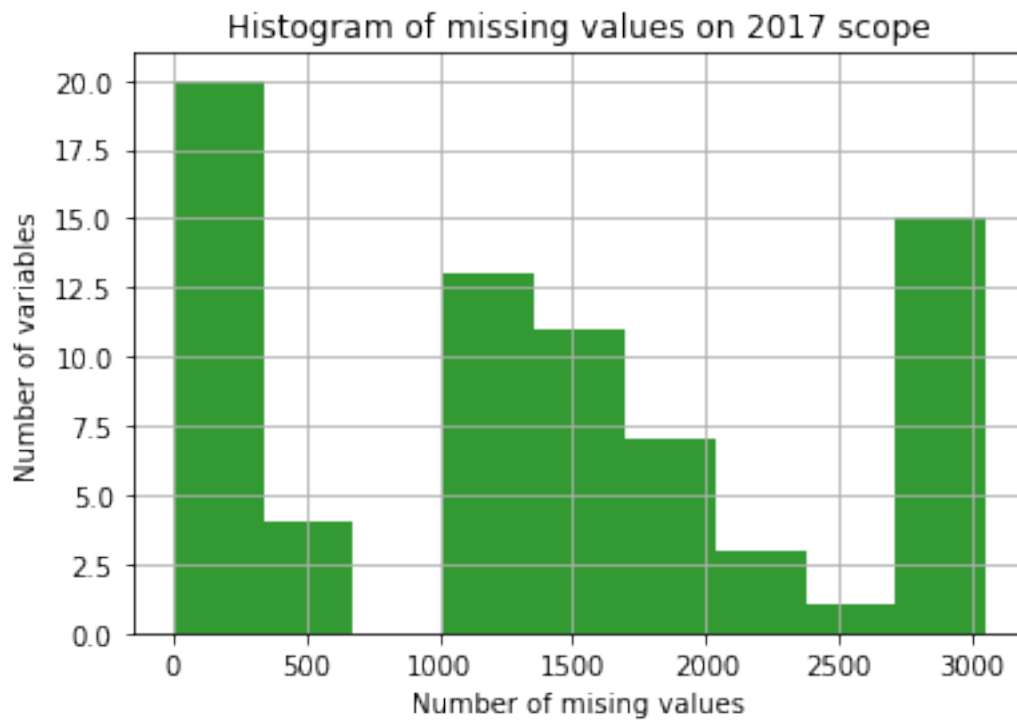
Length: 69, dtype: int64

```
In [60]: print("List of variable with more than 75% missing values in 2016:\n")
l = nb_missing_per_var_2016[scope_2016 -
                             complete_2016][nb_missing_per_var_2016 >
                                                0.75*3050]
print(" ".join(l.index))
```

List of variable with more than 75% missing values in 2016:

SEXE_3 SEXE_9 LIEN_9 AGE_6 LIEN_6 LIEN_7 SEXE_8 SEXE_6 LIEN_8 AGE_8 AGE_4 AGE_7 SEXE_7 AGE_9 SEX

```
In [61]: nb_missing_per_var_2017 = np.sum(cdv.loc[cdv["ANNEEFUZ"]==2017].isnull())
fig=plt.figure()
plt.title("Histogram of missing values on 2017 scope")
plt.ylabel(u'Number of variables')
plt.xlabel("Number of mising values")
bins = np.linspace(0, 3050,10)
plt.hist(nb_missing_per_var_2017[scope_2017 - complete_2017],
        bins, facecolor='g', alpha=0.8)
plt.grid()
```



```
In [62]: nb_missing_per_var_2017[scope_2017
        - complete_2017].sort_values(ascending=False)
```

```
Out [62]: SEXE_9      3018
          LIEN_9      3018
          AGE_9       3018
          AGE_8       3013
          LIEN_8       3013
          SEXE_8       3013
          AGE_7       2999
```

SEXE_7	2999
LIEN_7	2999
SEXE_6	2946
AGE_6	2946
LIEN_6	2946
SEXE_5	2775
AGE_5	2775
LIEN_5	2775
REVAUTR	2518
LIEN_4	2278
SEXE_4	2278
AGE_4	2278
interim2	1967
SALCOMPC	1920
PCSCONJ	1920
SALCOMPI	1869
AGE_3	1715
LIEN_3	1715
SEXE_3	1715
TYPCONT	1590
typcont2	1523
INTERIM	1523
TYPEEMPL	1523
...	
RE_LOG	1327
RE_MEDI	1327
RE_TABAL	1327
REVCONJ	1232
ACTCONJ	1203
NBENF	1063
SEXE_2	578
LIEN_2	578
AGE_2	578
PROGRAD	435
inseel	119
REVUC	105
REVsq	105
REVTOT	105
NOT_PROF	55
NOT_LIBR	52
NOT_POLI	47
NOT_CAD	44
NOT_LOG	43
NOT_COHE	39
NOT_FAMI	39
NOT_AMIS	37
DEPCOM	24
poptrpeu	24

```

typodeg      24
poppeud      24
popdense     24
popinter     24
pmun         24
SENSIENV     15
Length: 74, dtype: int64

```

```

In [63]: print("List of variable with more than 75% missing values in 2017:\n")
         l = nb_missing_per_var_2017[scope_2017 -
                                     complete_2017][nb_missing_per_var_2017 >
                                                     0.75*3050]

         print(" ".join(l.index))

```

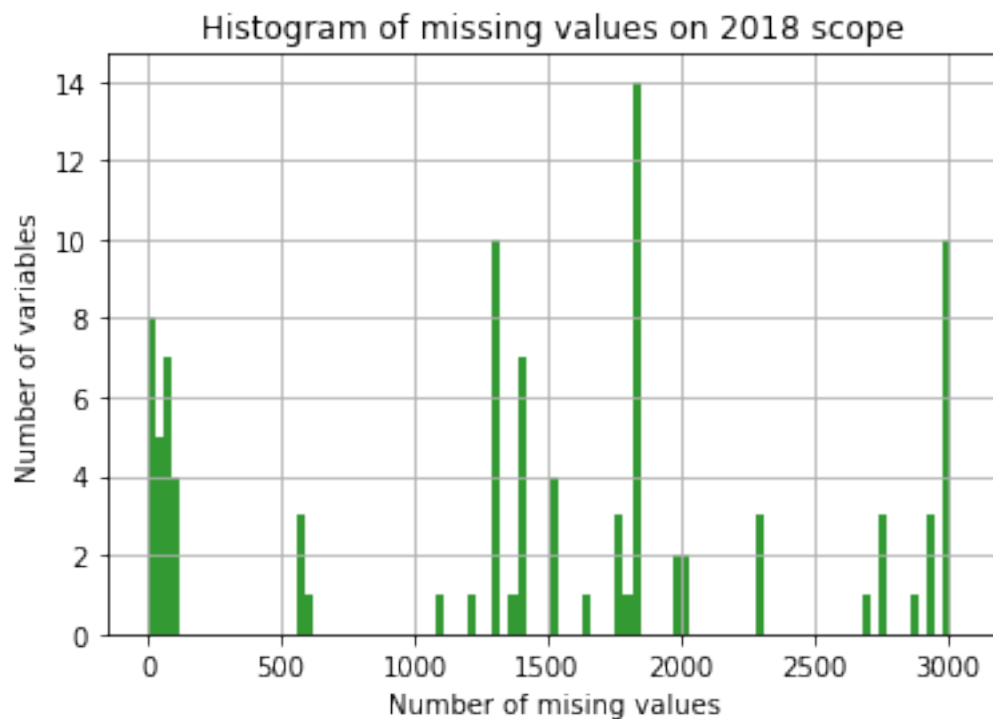
List of variable with more than 75% missing values in 2017:

SEXE_9 LIEN_9 AGE_6 LIEN_6 LIEN_7 SEXE_8 SEXE_6 LIEN_8 AGE_8 AGE_7 SEXE_7 AGE_9 SEXE_5 AGE_5 LIE

```

In [64]: nb_missing_per_var_2018 = np.sum(cdv.loc[cdv["ANNEEFUZ"]==2018].isnull())
         fig=plt.figure()
         plt.title("Histogram of missing values on 2018 scope")
         plt.ylabel(u'Number of variables')
         plt.xlabel("Number of missing values")
         bins = np.linspace(0, 3050,100)
         plt.hist(nb_missing_per_var_2018[scope_2018 - complete_2018],
                  bins, facecolor='g', alpha=0.8)
         plt.grid()

```



```
In [65]: nb_missing_per_var_2018[scope_2018
        - complete_2018].sort_values(ascending=False)
```

```
Out [65]: SEXE_9      3015
          LIEN_9      3015
          prescaf     3015
          AGE_9       3015
          SEXE_8      3009
          LIEN_8      3009
          AGE_8       3009
          SEXE_7      3001
          LIEN_7      3001
          AGE_7       3001
          AGE_6       2950
          SEXE_6      2950
          LIEN_6      2950
          AUTREAL     2880
          AGE_5       2758
          SEXE_5      2758
          LIEN_5      2758
          REVAUTR     2704
          AGE_4       2291
          SEXE_4      2291
          LIEN_4      2291
          RADQUOI     2031
          interim2    2006
          SALCOMPC    1975
          PCSCONJ     1975
          RADWHY8     1844
          RADWHY14    1844
          RADWHY7     1844
          RADWHY1     1844
          RADWHY11    1844
          ...
          ACTCONJ     1219
          NBENF       1096
          PROGRAD      608
          SEXE_2       573
          AGE_2        573
          LIEN_2       573
          med          117
          REVsqst     105
          REVUC        105
          REVTOT       105
          NOT_POLI      85
```

UDA5	85
NOT_COHE	81
NOT_LIBR	71
NOT_PROF	70
NOT_CAD	67
NOT_AMIS	65
NOT_LOG	61
ASSOAUTR	59
SENSIENV	54
NOT_FAMI	50
PCSCON7	47
pmun	15
poptrpeu	15
typodeg	15
popinter	15
DEPCOM	15
poppeud	15
popdense	15
RURAUROA	6

Length: 96, dtype: int64

```
In [66]: print("List of variable with more than 75% missing values in 2017:\n")
l = nb_missing_per_var_2018[scope_2018 -
                             complete_2018][nb_missing_per_var_2018 >
                                                0.75*3050]
print(" ".join(l.index))
```

List of variable with more than 75% missing values in 2017:

SEXE_9 LIEN_9 SEXE_4 AGE_5 AGE_6 LIEN_6 LIEN_7 SEXE_6 AGE_4 AGE_8 SEXE_5 REVAUTR SEXE_8 LIEN_8 S

1.3 2) Selection and classification of features

1.3.1 a) Variable to be predicted - "HEUREUX"

```
In [67]: cdv['HEUREUX'].value_counts().sort_values(ascending = False)
```

```
Out[67]: Assez souvent      5423
Occasionnellement      3665
Très souvent      1758
Jamais      203
[Nsp]      82
Name: HEUREUX, dtype: int64
```

1.3.2 b) Variable common to all years

```
In [68]: common_variables = (scope_2015 & scope_2016 & scope_2017 & scope_2018)
len(common_variables)
```

```
Out[68]: 268
```

```
In [69]: cdv_restricted_common = cdv.loc[:,common_variables]
```

```
In [70]: l = list(common_variables)
l.sort()
print("List of variable common to all years")
print(l)
```

List of variable common to all years

```
['ACM1', 'ACM10', 'ACM11', 'ACM12', 'ACM2', 'ACM3', 'ACM4', 'ACM5', 'ACM6', 'ACM7', 'ACM8', 'ACM9']
```

1.3.3 c) variable analysis - link with CDV study

```
In [71]: print(list(cdv.columns))
```

```
['INTER6', 'INTER', 'ANNEEFUZ', 'ANNEFUZ2', 'COLLECTE', 'CHAMP', 'POND', 'identifiant', 'SEXE', 'AGE']
```

```
In [72]: cdv.loc[:,["POND","INTER6",
                    "INTER","COLLECTE",
                    "CHAMP","identifiant"]].head()
```

```
Out[72]:
```

	POND	INTER6	INTER	COLLECTE	CHAMP	identifiant
0	1.313554	373001	3001	Online	18 ans et + métropole	NaN
1	2.009015	373002	3002	Online	18 ans et + métropole	NaN
2	0.217607	373003	3003	Online	18 ans et + métropole	NaN
3	0.539351	373004	3004	Online	18 ans et + métropole	NaN
4	0.270204	373005	3005	Online	18 ans et + métropole	NaN

```
In [73]: cdv["CHAMP"].unique()
```

```
Out[73]: array(['18 ans et + métropole', '15-17 ans + DOM + Corse'], dtype=object)
```

```
In [74]: cdv["COLLECTE"].unique()
```

```
Out[74]: array(['Online'], dtype=object)
```

```
In [75]: # Variables not present in the list ???
cdv["RURAURBA"].unique()
```

```
Out[75]: array(['PR', 'PU', nan, 'IN'], dtype=object)
```

```
In [76]: cdv["AGGLOINS"].unique()
```

```
Out[76]: array([ nan,  0.,  2.,  1.,  7.,  4.,  8.,  5.,  3.,  6.])
```

```
In [77]: # List of variable explained in the exceel file provided
liste_explained = {"INTER6", "ANNEEFUZ", "ANNEEFUZ2", "COLLECTE", "SEXE",
                  "AGE5", "PCSENQ8", "TYPOSQT",
                  "DIPL4", "AGGLO5", "UDA10", "SITUEMP3", "AGGLO9", "AGE",
                  "EXERCPR0", "SITUEMP", "SITUEMP5",
                  "SITUEMP6", "SALCOMP", "INTERIM", "TYPCONT", "TEMPSTRA",
                  "nbheures", "NBHEUR39",
                  "NBHEUR35", "PREFPALI", "SALCOMPI", "CHERCHEM", "NBCHOM",
                  "STATMAT", "ACTCONJ", "SALCOMPC",
                  "ENFANTS", "NBENF", "NBENF6", "DIPLOME", "FAMILLE",
                  "UNIONGAY", "ADOPTGAY", "TRAVFEM", "NB0003", "NB0306",
                  "NB0610", "NB1016",
                  "NB1620", "NB2099", "NBPIECE6", "LOGSUFFI", "DEPLOG",
                  "DEPLOG3", "CADVIE", "CADVIE3", "SECUR3",
                  "MODCHAUF", "TYPCHAUF", "TELFIXE", "TELMOB", "SENSIENV",
                  "TAXENV", "HANDICAP", "SOUFFTET", "SOUFFDOS",
                  "SOUFFNER", "SOUFFDEP", "SOUFFINS", "ETATSAN", "NBPERS",
                  "NBPERS5", "SEXE_2", "SEXE_3",
                  "SEXE_4", "SEXE_5", "SEXE_6", "SEXE_7", "SEXE_8", "SEXE_9",
                  "AGE_2", "AGE_3",
                  "AGE_4", "AGE_5", "AGE_6", "AGE_7", "AGE_8", "AGE_9",
                  "LIEN_2", "LIEN_3", "LIEN_4", "LIEN_5", "LIEN_6",
                  "LIEN_7", "LIEN_8", "LIEN_9", "RESTRICT", "NIVPERSO",
                  "NIVFRAN", "NIVFRAN4", "CDV5", "BANQEPA",
                  "BANQVIE", "ASSOSPOR", "ASSOCULT", "ASSOCONF",
                  "ASSOJEUN", "ASSOSYND", "ASSOENVI",
                  "ASSOPARE", "ASSOCONS", "ASSOPOLI", "ASSOHUMA",
                  "ASSOAUTR", "FREQSPOR",
                  "FREQTELE", "RAISPAUV", "CHOAVANT", "CHOVOLON",
                  "OPIRSA", "JUSTICE", "TRANSFST", "PREOCCU1",
                  "PREOCCU2", "INQAGRE3", "INQALIM", "CLASSES0",
                  "HEUREUX", "CONFGOUV", "revtot7",
                  "NBUC", "TYPLOG2", "TYPLOG3", "AGESEX12",
                  "PCSENQ36", "UDA14", "zau1999", "POND", "dpt"}
```

```
In [78]: print(f"Number of variable explained in the exceel file \
... :{len(liste_explained)}")
```

Number of variable explained in the exceel file :135

```
In [79]: columns = set(cdv.columns)
```

```
In [80]: print(f"Variables explained but not present in the dataset :\
{len(liste_explained - columns)}\n")
print(" ".join(liste_explained - columns))
```

Variables explained but not present in the dataset :4

TYPLOG3 TELFIXE dpt ANNEEFUZ2

```
In [81]: print(f"Variables present in the dataset but not explained :\n\n"
          {len(columns - liste_explained)}\n")
          print(" ".join(columns - liste_explained))
```

Variables present in the dataset but not explained :223

REVTOT6 ASS06_2 CONFPOLI poppeud SITUFAM RE_WEB RE_TABAL NBPIECES INQROUTE INQMALA3 PCSCON7 ADNC

```
In [82]: print(f"Variables present in the dataset for all years but not explained :\n\n"
          {len(common_variables - liste_explained)}\n")
          print(" ".join(common_variables - liste_explained))
```

Variables present in the dataset for all years but not explained :137

ASS06_2 REVTOT6 CONFPOLI RE_WEB SITUFAM RE_TABAL NBPIECES INQROUTE INQMALA3 PCSCON7 NOT_AMIS INQ

1.3.4 d) bottom up...

Variable **inseel**, **com** & **DEPCOM**

```
In [83]: # analysis of differences between inseel com & DEPCOM
          df_tmp = cdv.loc[:,['inseel','com','DEPCOM']].dropna()
```

```
In [84]: df_tmp.loc[df_tmp['inseel'] != df_tmp['com']].head(10)
```

```
Out [84]:
```

	inseel	com	DEPCOM
8128	75056SR19	75119	75056
8132	69123SR03	69383	69123
8144	69123SR01	69381	69123
8157	69123SR03	69383	69123
8171	75056SR12	75112	75056
8175	69123SR07	69387	69123
8180	75056SR12	75112	75056
8181	75056SR18	75118	75056
8182	75056SR15	75115	75056
8217	75056SR16	75116	75056

```
In [85]: df_tmp.loc[df_tmp['inseel']!= df_tmp['DEPCOM']].head(10)
```

```
Out [85]:
```

	inseel	com	DEPCOM
8128	75056SR19	75119	75056
8132	69123SR03	69383	69123
8144	69123SR01	69381	69123
8157	69123SR03	69383	69123

```

8171  75056SR12  75112  75056
8175  69123SR07  69387  69123
8180  75056SR12  75112  75056
8181  75056SR18  75118  75056
8182  75056SR15  75115  75056
8217  75056SR16  75116  75056

```

```
In [86]: cdv["REVENQ"].describe()
```

```

Out[86]: count      11131.000000
         mean       71370.251101
         std        253880.504617
         min         0.000000
         25%        1100.000000
         50%        1800.000000
         75%        2800.000000
         max        999999.000000
         Name: REVENQ, dtype: float64

```

```
In [87]: cdv["AUTREREV"].describe()
```

```

Out[87]: count      11131.000000
         mean      192360.568682
         std       393925.875729
         min         0.000000
         25%         0.000000
         50%         0.000000
         75%        1000.000000
         max       999999.000000
         Name: AUTREREV, dtype: float64

```

```
In [88]: cdv['REV_TR7'].value_counts()
```

```

Out[88]: De 1500 à 2300      2332
         De 2300 à 3100      2281
         4000 et plus        2087
         3100 à 4000        1675
         De 900 à 1500       1544
         Moins de 900        862
         Non déclaré (ne sait pas, refus)  350
         Name: REV_TR7, dtype: int64

```

N – THEMES GENERAUX*

```
In [89]: # N1
         cdv['JUSTICE'].value_counts()
```

```

Out[89]: Assez mal      5365
         Assez bien     3080

```

```

Très mal      2379
Très bien     206
[Nsp]         101
Name: JUSTICE, dtype: int64

```

```

In [90]: # N2
cdv['RELEG'].value_counts()

```

```

Out[90]: Non, plutôt pas      1302
        Oui, plutôt          1069
        Oui, tout à fait      311
        Non, pas du tout      290
        [Nsp]                 44
        Name: RELEG, dtype: int64

```

```

In [91]: # N3 done
cdv['CONFPUB'].value_counts()

```

```

Out[91]: Plutôt confiance      6026
        Plutôt pas confiance    3418
        Pas du tout confiance   1030
        Très confiance          503
        [Nsp]                  154
        Name: CONFPUB, dtype: int64

```

```

In [92]: # N4
cdv['TRANSFST'].value_counts()

```

```

Out[92]: Oui      9288
        Non      1710
        [Nsp]     133
        Name: TRANSFST, dtype: int64

```

```

In [93]: # N5
cdv['PROGRAD'].value_counts()

```

```

Out[93]: Changements radicaux    4988
        Réformes progressives    4231
        [Nsp]                   69
        Name: PROGRAD, dtype: int64

```

```

In [94]: # N6 only if N5 "Changements radicaux"
cdv.loc[cdv['RADIQUOI'].notnull(), 'RADIQUOI'].head(20)

```

```

Out[94]: 8116    Simplifier les papiers administratifs pour les...
        8117    diminuer le nombre de fonctionnaires et mettre...
        8120    Redevenons ce que nous avons !t! pendant des s...
        8121                      revenir ! des valeurs saines
        8124    ne pas attendre tout de l' !tat.,Arr!ter l' !t...

```

```

8127 Am!liorer le pouvoir d'achat, revoir la fiscal...
8128 lutter contre les in!galit!s sociales,les !car...
8129 la gestion de l'!tat
8130 tout est a changer
8132 les migrants doivent retourner dans leur payi
8134 lib!ralisation que l'!tat soit moins pr!sent, ...
8135 requilibre de la fortune
8144 Limiter le pouvoir de la sph!re financi!re
8148 code du travail, fiscalit!
8152 Tout
8153 Punir les voyous beaucoup plus s!v!rement et r...
8157 plus d'!galit! entre jeune et vieux
8159 D!velopper l?initiative et l?entrepreneariat. R...
8165 La justice
8174 M!me m!thode pour TOUS dans la fa!on de calcul...
Name: RADIQUEOI, dtype: object

```

```
In [95]: cdv_ssfmt.loc[cdv_ssfmt['RADIQUEOI'].notnull(),'RADIQUEOI'].head(20)
```

```

Out[95]: 8116 Simplifier les papiers administratifs pour les...
8117 diminuer le nombre de fonctionnaires et mettre...
8120 Redevenons ce que nous avons !t! pendant des s...
8121 revenir ! des valeurs saines
8124 ne pas attendre tout de l' !tat.,Arr!ter l' !t...
8127 Am!liorer le pouvoir d'achat, revoir la fiscal...
8128 lutter contre les in!galit!s sociales,les !car...
8129 la gestion de l'!tat
8130 tout est a changer
8132 les migrants doivent retourner dans leur payi
8134 lib!ralisation que l'!tat soit moins pr!sent, ...
8135 requilibre de la fortune
8144 Limiter le pouvoir de la sph!re financi!re
8148 code du travail, fiscalit!
8152 Tout
8153 Punir les voyous beaucoup plus s!v!rement et r...
8157 plus d'!galit! entre jeune et vieux
8159 D!velopper l?initiative et l?entrepreneariat. R...
8165 La justice
8174 M!me m!thode pour TOUS dans la fa!on de calcul...
Name: RADIQUEOI, dtype: object

```

```
In [96]: # N7
cdv['ORDLIB'].value_counts()
```

```

Out[96]: Un peu plus d'ordre          4711
Beaucoup plus d'ordre          3699
Un peu plus de liberté          1817
Beaucoup plus de liberté          739

```

```

[Nsp] 165
Name: ORDLIB, dtype: int64

In [97]: # N8
cdv['CONFMEFI'].value_counts()

Out[97]: On n'est jamais assez méfiant 4931
Il est possible de faire confiance aux autres 4024
[Nsp] 131
Name: CONFMEFI, dtype: int64

In [98]: # N9
cdv['PREOCCU1'].value_counts()

Out[98]: Le chômage 2716
La violence et l'insécurité 1789
L'immigration 1465
Les maladies graves 1218
La pauvreté en France 1071
La dégradation de l'environnement 1063
Les tensions internationales 696
La pauvreté dans le monde 543
Les conflits sociaux 178
La drogue 158
L'Europe 140
[Nsp, NR] 94
Name: PREOCCU1, dtype: int64

In [99]: cdv['PREOCCU2'].value_counts()

Out[99]: La violence et l'insécurité 2126
L'immigration 1581
La pauvreté en France 1367
Le chômage 1141
La dégradation de l'environnement 1119
Les tensions internationales 932
Les maladies graves 849
La pauvreté dans le monde 801
Les conflits sociaux 348
L'Europe 343
La drogue 278
[Nsp, NR] 246
Name: PREOCCU2, dtype: int64

In [100]: # N10 RADI1 RADI2 RADI3
cdv['RADI1'].value_counts()

Out[100]: Oui, je comprends 1691
Non, je ne comprends pas 1282
[Nsp] 43
Name: RADI1, dtype: int64

```

```

In [101]: cdv['RADI2'].value_counts()

Out[101]: Non, je ne comprends pas      2683
          Oui, je comprends             284
          [Nsp]                         49
          Name: RADI2, dtype: int64

In [102]: cdv['RADI3'].value_counts()

Out[102]: Non, je ne comprends pas      2609
          Oui, je comprends             359
          [Nsp]                         48
          Name: RADI3, dtype: int64

In [103]: # N11 OPICULT
          cdv['OPICULT'].value_counts()

Out[103]: La diversité des cultures est une richesse      4603
          La diversité des cultures rend difficile la vie en commun  4250
          [Nsp]                                           233
          Name: OPICULT, dtype: int64

In [104]: # N12 COMMU1COMMU8
          cdv['COMMU1'].value_counts()

Out[104]: Non      2284
          Oui       594
          [Nsp]     138
          Name: COMMU1, dtype: int64

In [105]: cdv['COMMU2'].value_counts()

Out[105]: Non      2063
          Oui       828
          [Nsp]     125
          Name: COMMU2, dtype: int64

In [106]: # N13
          cdv['MONDIAL'].value_counts()

Out[106]: Plutôt des inconvénients      4065
          Plutôt des avantages          3874
          [Nsp]                        142
          Name: MONDIAL, dtype: int64

In [107]: # N14
          # INQMALAD - INQAGRES - INQROUTE -
          # INQCHOMA - INQGUERR - INQNUCLE- INQALIM
          cdv['INQMALAD'].value_counts()

```

```

Out[107]: Beaucoup      4516
          Assez         4100
          Un peu        1861
          Pas du tout    508
          [Nsp]         146
          Name: INQMALAD, dtype: int64

In [108]: cdv['INQAGRES'].value_counts()

Out[108]: Assez         3600
          Un peu        3409
          Beaucoup      2475
          Pas du tout    1446
          [Nsp]         201
          Name: INQAGRES, dtype: int64

In [109]: # N15
          cdv['CLASSES0'].value_counts()

Out[109]: La classe moyenne inférieure    4693
          La classe moyenne supérieure    2880
          La classe populaire             2036
          Les défavorisés                 681
          Les gens aisés                  560
          Les privilégiés                 171
          [Nsp]                          110
          Name: CLASSES0, dtype: int64

In [110]: # N16
          cdv['ISEGO'].value_counts()

Out[110]: Assez souvent    2590
          Rarement         2129
          Très souvent     871
          Jamais           356
          [Nsp]            90
          Name: ISEGO, dtype: int64

In [111]: # N19 CONFGOUV
          cdv['CONFGOUV'].value_counts()

Out[111]: Pas du tout confiance    4069
          Plutôt pas confiance     3625
          Plutôt confiance         2882
          Tout à fait confiance    430
          [Nsp]                   125
          Name: CONFGOUV, dtype: int64

In [112]: # N20 ECHPOL
          cdv['ECHPOL'].value_counts()

```

```
Out[112]: Au centre          2967
          Plutôt à gauche   1803
          Plutôt à droite   1663
          A gauche          1418
          À droite          1202
          Très à droite     1040
          [Nsp, NR]         542
          Très à gauche     496
          Name: ECHPOL, dtype: int64
```

```
In [113]: cdv_ssfmt['ECHPOL'].value_counts()
```

```
Out[113]: 4      2967
          3      1803
          5      1663
          2      1418
          6      1202
          7      1040
          8       542
          1       496
          Name: ECHPOL, dtype: int64
```

```
In [114]: cdv_ssfmt['NB0003'].value_counts()
```

```
Out[114]: 0      10527
          1       552
          2        49
          3         2
          4         1
          Name: NB0003, dtype: int64
```

```
In [115]: cdv_ssfmt['NB0610'].value_counts()
```

```
Out[115]: 0      10185
          1       814
          2       126
          3         6
          Name: NB0610, dtype: int64
```

```
In [116]: obj_df = cdv.select_dtypes(include=['object']).copy()
```

```
In [117]: # INTER6 index candidate
```

```
In [118]: np.sum(cdv["INTER6"].isnull())
```

```
Out[118]: 0
```

```
In [119]: len(cdv["INTER6"].unique()) == cdv.shape[0]
```

```
Out[119]: True
```



```
In [120]: tech_var = {"ANNEEFUZ", "ANNEFUZ2", "COLLECTE", "CHAMP",
                    "identifiant", "an_enq", "INTER"}
cdv.loc[:,tech_var].dropna().head()
```

```
Out[120]:
```

	ANNEEFUZ	CHAMP	INTER	identifiant	COLLECTE	an_enq	\
8115	2018	18 ans et + métropole	1	0a22333	Online	2018	
8116	2018	18 ans et + métropole	2	0a26329	Online	2018	
8117	2018	18 ans et + métropole	3	0a26791	Online	2018	
8118	2018	18 ans et + métropole	4	0a27001	Online	2018	
8119	2018	18 ans et + métropole	5	0a27217	Online	2018	

	ANNEFUZ2
8115	2018
8116	2018
8117	2018
8118	2018
8119	2018

```
In [121]: com_var = {'COMINSEE', 'DEPCOM', 'com', 'inseel', 'inseenum', 'CP'}
cdv.loc[:,tech_var].dropna().head()
```

```
Out[121]:
```

	ANNEEFUZ	CHAMP	INTER	identifiant	COLLECTE	an_enq	\
8115	2018	18 ans et + métropole	1	0a22333	Online	2018	
8116	2018	18 ans et + métropole	2	0a26329	Online	2018	
8117	2018	18 ans et + métropole	3	0a26791	Online	2018	
8118	2018	18 ans et + métropole	4	0a27001	Online	2018	
8119	2018	18 ans et + métropole	5	0a27217	Online	2018	

	ANNEFUZ2
8115	2018
8116	2018
8117	2018
8118	2018
8119	2018

```
In [122]: bizz_var = {'NB0003', 'NB0306', 'NB0610', 'NB1016', 'NB1620', 'NB2099',
                    'an_nais', 'decuc', 'decsqt', 'info', 'typodeg', 'refus2',
                    'cpt', 'prescaf', 'poptrpeu', 'REVUC', 'i', 'REVTOT',
                    'poppeud', 'popdense', 'popinter', 'pmun', 'agedip', 'age_OW',
                    'REVsqt', 'NBUC', 'AGGLOINS', 'med', 'CSP6', 'REVTOT6'}
cdv.loc[:,bizz_var].dropna().head()
```

```
Out[122]:
```

	pmun	REVTOT6	REVsqt	poppeud	med	NB1016	decsqt	\
10457	2017.0	3	1356.937913	0.04195	1900.0	0	Décile 4	

	NB1620	NB0306	info ...	AGGLOINS	NBUC	cpt	poptrpeu	\
10457	0	0	revtranc seul ...	2.0	1.5	0.0	0.0	

	NB0610	agedip	CSP6	age_OW	refus2	i
--	--------	--------	------	--------	--------	---

```
10457      0    21.0   5.0    4.0    1.0   6.0
```

```
[1 rows x 30 columns]
```

```
In [123]: cdv.loc[:, 'RADIQUOI'].dropna().head()
```

```
Out[123]: 8116    Simplifier les papiers administratifs pour les...
          8117    diminuer le nombre de fonctionnaires et mettre...
          8120    Redevenons ce que nous avons !t! pendant des s...
          8121                      revenir ! des valeurs saines
          8124    ne pas attendre tout de l' !tat.,Arr!ter l' !t...
          Name: RADIQUOI, dtype: object
```

Type 'object' in cdv file

```
In [124]: obj_cdv = cdv.select_dtypes(include=['object'])
          obj_var = set(obj_cdv.columns)
          print(f"{len(obj_var)} variables having 'object' type in 'cdv.csv'")
          cat_max9_var = set()
          cat_min10_var = set()
          for c in obj_var:
              obj_cdv_valcpt = obj_cdv[c].value_counts()
              print(f"*****{c}*****")
              print(obj_cdv_valcpt)
              print("-----\n")
              if len(obj_cdv_valcpt) > 10:
                  cat_min10_var.add(c)
              else:
                  cat_max9_var.add(c)
```

```
292 variables having 'object' type in 'cdv.csv'
```

```
*****SEXE_9*****
```

```
Homme      5
```

```
Femme     2
```

```
Name: SEXE_9, dtype: int64
```

```
-----
```

```
*****ASS06_2*****
```

```
Non adhérent    6181
```

```
Adhérent        4950
```

```
Name: ASS06_2, dtype: int64
```

```
-----
```

```
*****TEMPSTRA*****
```

```
Temps plein    4869
```

```
Temps partiel  1019
```

```
[Nsp]          1
```

```
Name: TEMPSTRA, dtype: int64
```

*****SITUFAM*****

Couple sans enfants	3598
Couple avec enfants	2499
Personne sans enfants vivant seul	2219
Personne sans enfants vivant avec d'autres personnes (colocataires, parents...)	1778
Couple sans enfants, vivant avec d'autres personnes (colocataires, parents...)	543
Foyer monoparental	485

Name: SITUFAM, dtype: int64

*****RE_WEB*****

Oui	3083
Non	2942
[Nsp]	429

Name: RE_WEB, dtype: int64

*****CONFPOLI*****

Pas du tout confiance	4700
Plutôt pas confiance	4543
Plutôt confiance	1524
[Nsp]	201
Très confiance	163

Name: CONFPOLI, dtype: int64

*****LIEN_9*****

Enfant, beau-fils, belle-fille	3
Autre	3
Conjoint ou compagnon	1
Frère, soeur	1

Name: LIEN_9, dtype: int64

*****RE_TABAL*****

Non	3513
Oui	2466
[Nsp]	475

Name: RE_TABAL, dtype: int64

*****MODCHAUF*****

Électricité	3866
Gaz naturel (gaz de réseau)	3643
Fioul, mazout	1030
Bois	1028

NSP (pour certains chauffages collectifs)	560
Chauffage urbain	331
Gaz butane, propane	243
Géothermie, pompe à chaleur	241
Autre	126
Aucun	35
Charbon	28

Name: MODCHAUF, dtype: int64

*****ETATSAN*****

Satisfaisant	7386
Très satisfaisant	1681
Peu satisfaisant	1650
Pas satisfaisant du tout	333
[Nsp]	81

Name: ETATSAN, dtype: int64

*****INQROUTE*****

Assez	4261
Un peu	3312
Beaucoup	2635
Pas du tout	718
[Nsp]	205

Name: INQROUTE, dtype: int64

*****INQMALA3*****

Inquiet des risques de maladie grave	8616
Non inquiet	2369
[Nsp]	146

Name: INQMALA3, dtype: int64

*****PCSCON7*****

Sans conjoint	4532
Conjoint inactif	2770
Cadre, prof. intellectuelle sup., profession intermédiaire	1687
Employé	1039
Ouvrier	776
Agri. exploitant, artisan, commerçant et chef d'entreprise	272
Non déclaré	8

Name: PCSCON7, dtype: int64

*****ADNCB*****

Pas du tout d'accord	1102
----------------------	------

Pas très d'accord 929
Assez d'accord 683
Tout à fait d'accord 265
[Nsp] 37
Name: ADN CB, dtype: int64

*****HEUREUX*****

Assez souvent 5423
Occasionnellement 3665
Très souvent 1758
Jamais 203
[Nsp] 82
Name: HEUREUX, dtype: int64

*****INQAGRES*****

Assez 3600
Un peu 3409
Beaucoup 2475
Pas du tout 1446
[Nsp] 201
Name: INQAGRES, dtype: int64

*****TELMOB*****

Oui 10630
Non 407
[Nsp] 94
Name: TELMOB, dtype: int64

*****FREQBIBL*****

Jamais 5381
Exceptionnellement 3437
Régulièrement 2210
[Nsp] 103
Name: FREQBIBL, dtype: int64

*****SECURITE*****

Assez en sécurité 6773
Tout à fait en sécurité 2740
Peu en sécurité 1271
Pas du tout en sécurité 286
[Nsp] 61
Name: SECURITE, dtype: int64

*****SOUFFDOS*****

Oui 5859
Non 5114
[Nsp] 158
Name: SOUFFDOS, dtype: int64

*****NIVPERSO*****

C'est pareil 3445
Un peu moins bien 3248
Beaucoup moins bien 1980
Un peu mieux 1727
Beaucoup mieux 613
[Nsp] 118
Name: NIVPERSO, dtype: int64

*****OPIRSA*****

Cela risque d'inciter les gens à s'en contenter 6278
Cela leur donnera le coup de pouce nécessaire 4609
[Nsp] 244
Name: OPIRSA, dtype: int64

*****CONFECOL*****

Plutôt confiance 5475
Plutôt pas confiance 2018
Très confiance 888
Pas du tout confiance 556
[Nsp] 149
Name: CONFECOL, dtype: int64

*****REV_TR7*****

De 1500 à 2300 2332
De 2300 à 3100 2281
4000 et plus 2087
3100 à 4000 1675
De 900 à 1500 1544
Moins de 900 862
Non déclaré (ne sait pas, refus) 350
Name: REV_TR7, dtype: int64

*****HARVEY*****

On n'en parle pas assez 1438
On en parle suffisamment 1258

```

On en parle trop                284
[Nsp]                            36
Name: HARVEY, dtype: int64
-----

*****RELIGION*****
Ni pratique, ni sentiment d'appartenance        4703
Pas de pratique, mais sentiment d'appartenir à une religion  3306
Une pratique religieuse occasionnelle            2167
Une pratique religieuse régulière                830
[Nsp]                                             125
Name: RELIGION, dtype: int64
-----

*****ACTCONJ*****
Oui, à temps plein                            3311
Non, a travaillé mais ne travaille plus        2367
Oui, à temps partiel                          525
Non, est au chômage                           268
N'a jamais travaillé                          145
[Nsp]                                           16
Name: ACTCONJ, dtype: int64
-----

*****CONFBANK*****
Plutôt pas confiance        3624
Plutôt confiance            2987
Pas du tout confiance       2079
Très confiance              239
[Nsp]                       157
Name: CONFBANK, dtype: int64
-----

*****ANNEFUZ2*****
2016                3050
2017                3020
2018                3016
2015 online        2045
Name: ANNEFUZ2, dtype: int64
-----

*****SALCOMPC*****
Salarié            3424
A son compte       394
[Nsp]              18
Name: SALCOMPC, dtype: int64
-----

```

*****ADOPTGAY*****

Tout à fait d'accord	3387
Pas du tout d'accord	2826
Assez d'accord	2750
Peu d'accord	2098
[Nsp]	70

Name: ADOPTGAY, dtype: int64

*****decuc*****

Décile 6	1267
Décile 3	1165
Décile 1	1147
Décile 9	1133
Décile 5	1095
Décile 10	1070
Décile 4	1058
Décile 7	969
Décile 2	965
Décile 8	915
Non renseigné	347

Name: decuc, dtype: int64

*****RADI2*****

Non, je ne comprends pas	2683
Oui, je comprends	284
[Nsp]	49

Name: RADI2, dtype: int64

*****PRESTCAF*****

Non	8248
Oui	2869
[Nsp]	14

Name: PRESTCAF, dtype: int64

*****PCSENQ32*****

Retraité	2968
Élève, étudiant	1134
Ouvrier qualifié	1009
Employé administratif d'entreprise	632
Personnes au foyer, a.inac.	595
Employé civil, agent de service fonction publique	546
Technicien	366
Employé de commerce	359
Profession intermédiaire santé et travail social	338

Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et com. des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Cadre de la fonction publique	156
Profession libérale	152
Artisan	142
Professeur, profession scientifique	137
Professeur des écoles, instituteur et assimilé	132
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Clergé, religieux	5

Name: PCSNQ32, dtype: int64

*****SOUFFTET*****

Non	6630
Oui	4293
[Nsp]	208

Name: SOUFFTET, dtype: int64

*****JUSTICE*****

Assez mal	5365
Assez bien	3080
Très mal	2379
Très bien	206
[Nsp]	101

Name: JUSTICE, dtype: int64

*****ACM4*****

Satisfait de son cadre de vie	6844
Très satisfait de son cadre de vie	2781
Pas satisfait de son cadre de vie	1459
[Nsp]	47

Name: ACM4, dtype: int64

*****AGE6*****

40 à 59 ans	3216
25 à 39 ans	1796
70 ans et plus	1397
60 à 69 ans	1250
18 à 24 ans	1147
Moins de 18 ans	280

Name: AGE6, dtype: int64

*****ROBOT3*****

Cest plutôt une mauvaise chose	1527
Cest plutôt une bonne chose	1413
[Nsp]	76

Name: ROBOT3, dtype: int64

*****YOGA*****

Jamais	2128
Plus rarement	408
En moyenne, une fois par semaine	228
En moyenne, une fois par mois	139
Tous les jours ou presque	73
[Nsp]	40

Name: YOGA, dtype: int64

*****ADNSTIC*****

Assez d'accord	1187
Tout à fait d'accord	978
Pas très d'accord	420
Pas du tout d'accord	395
[Nsp]	36

Name: ADNSTIC, dtype: int64

*****OPICULT*****

La diversité des cultures est une richesse	4603
La diversité des cultures rend difficile la vie en commun	4250
[Nsp]	233

Name: OPICULT, dtype: int64

*****SEXE_4*****

Femme	1197
Homme	1134

Name: SEXE_4, dtype: int64

*****NIVFRAN*****

Un peu moins bien	4766
Beaucoup moins bien	2986
C'est pareil	2331
Un peu mieux	749
[Nsp]	173
Beaucoup mieux	126

Name: NIVFRAN, dtype: int64

*****ASSOPARE*****

Non	10043
Oui	546
[Nsp]	542

Name: ASSOPARE, dtype: int64

*****EFFORTPP*****

Ne font pas assez	4656
Font ce qu'ils doivent	4348
Font trop	1965
[Nsp]	162

Name: EFFORTPP, dtype: int64

*****INQCHOM3*****

Inquiet des risques de chômage	6191
Non inquiet	4740
[Nsp]	200

Name: INQCHOM3, dtype: int64

*****PCSENQ35*****

Ancien cadre	1258
Élève, étudiant	1134
Ouvrier qualifié	1009
Ancienne profession intermédiaire	750
Employé administratif d'entreprise	632
Ancien employé	622
Personne au foyer et autre inactif	595
Employé civil, agent de service fonction publique	546
Technicien	366
Employé de commerce	359
Profession intermédiaire santé et travail social	338
Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et comm. des entreprises	288
Personnel des services directs aux particuliers	259

Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Ancien artisan, commerçant, chef d'entreprise	179
Cadre de la fonction publique	156
Profession libérale	152
Ancien ouvrier	148
Artisan	142
Professeur, profession scientifique	137
Professeur des écoles, instituteur et assimilé	132
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Ancien agriculteur exploitant	11
Clergé, religieux	5

Name: PCSNQ35, dtype: int64

*****INQUERR*****

Un peu	3348
Assez	3175
Beaucoup	2633
Pas du tout	1756
[Nsp]	219

Name: INQUERR, dtype: int64

*****RADI3*****

Non, je ne comprends pas	2609
Oui, je comprends	359
[Nsp]	48

Name: RADI3, dtype: int64

*****EHPOL*****

Au centre	2967
Plutôt à gauche	1803
Plutôt à droite	1663
A gauche	1418
À droite	1202
Très à droite	1040
[Nsp, NR]	542
Très à gauche	496

Name: EHPOL, dtype: int64

*****PREOCCU1*****

Le chômage	2716
La violence et l'insécurité	1789
L'immigration	1465
Les maladies graves	1218
La pauvreté en France	1071
La dégradation de l'environnement	1063
Les tensions internationales	696
La pauvreté dans le monde	543
Les conflits sociaux	178
La drogue	158
L'Europe	140
[Nsp, NR]	94

Name: PREOCCU1, dtype: int64

*****PREOVIO*****

Non 7216

Oui 3915

Name: PREOVIO, dtype: int64

*****RADWHY14*****

0

Aligner le public sur le privé, réduire les dépenses publiques (simplifier l'administration, sup

Name: RADWHY14, dtype: int64

*****RE_ALIM*****

Non 3095

Oui 2992

[Nsp] 367

Name: RE_ALIM, dtype: int64

*****SEXE_2*****

Homme 3916

Femme 3860

Name: SEXE_2, dtype: int64

*****ASSOCONF*****

Non 10039

Oui 573

[Nsp] 519

Name: ASSOCONF, dtype: int64

*****TRANSF05*****

Radical	4988
Progressiste	4231
Pas de réforme	1710
Nsp si réforme	133
Nsp quelle réforme	69

Name: TRANSF05, dtype: int64

*****ACM5*****

État de santé satisfaisant	7386
État de santé pas satisfaisant	1983
État de santé très satisfaisant	1681
[Nsp]	81

Name: ACM5, dtype: int64

*****ACM6*****

Niveau de vie Français : un peu moins bien	4766
Niveau de vie Français : beaucoup moins bien	2986
Niveau de vie Français : pareil	2331
Niveau de vie Français : mieux	875
[Nsp]	173

Name: ACM6, dtype: int64

*****ENFANTS*****

Oui	7107
Non	4024

Name: ENFANTS, dtype: int64

*****COLLECTE*****

Online	11131
--------	-------

Name: COLLECTE, dtype: int64

*****RADWHY2*****

0	1081
plus de social: tout le monde doit pouvoir en profiter	91

Name: RADWHY2, dtype: int64

*****ASSOAUTR*****

Non	10762
Oui	310

Name: ASSOAUTR, dtype: int64

*****DEPLOG3*****

Une charge lourde, très lourde 5738

Une charge négligeable, supportable 5341

[Nsp] 52

Name: DEPLOG3, dtype: int64

*****CONFMEFI*****

On n'est jamais assez méfiant 4931

Il est possible de faire confiance aux autres 4024

[Nsp] 131

Name: CONFMEFI, dtype: int64

*****CONFENTR*****

Plutôt confiance 6402

Plutôt pas confiance 3131

Pas du tout confiance 829

Très confiance 569

[Nsp] 200

Name: CONFENTR, dtype: int64

*****NB99_4*****

Aucun enfant de 20 ans et plus 7543

Un enfant de 20 ans et plus 1579

Deux enfants de 20 ans et plus 1409

Trois et plus enfants de 20 ans et plus 600

Name: NB99_4, dtype: int64

*****SEXE_3*****

Femme 2054

Homme 1925

Name: SEXE_3, dtype: int64

*****RESTRICT*****

Oui 6454

Non 4571

[Nsp] 106

Name: RESTRICT, dtype: int64

*****COMMU7*****

Non 2709
[Nsp] 160
Oui 147
Name: COMMU7, dtype: int64

*****INTERIM*****

Non 5076
Oui 263
Name: INTERIM, dtype: int64

*****SALCOMPI*****

Salarié 3864
À votre compte 308
N'a jamais travaillé 155
Name: SALCOMPI, dtype: int64

*****TAXENV*****

Non 8243
Oui 2810
[Nsp] 78
Name: TAXENV, dtype: int64

*****FAMILLE*****

Oui 6089
Non 4994
[Nsp] 48
Name: FAMILLE, dtype: int64

*****LIEN_6*****

Enfant, beau-fils, belle-fille	113
Frère, soeur	67
Parent, beau-parent	17
Autre	9
6	9
Conjoint ou compagnon	8
7	2

Name: LIEN_6, dtype: int64

*****NB10_2*****

Non 9453
Oui, enfant de moins de 10 ans 1678
Name: NB10_2, dtype: int64

*****SEXE_6*****

Femme 118

Homme 93

Name: SEXE_6, dtype: int64

*****LIEN_7*****

Enfant, beau-fils, belle-fille 19

Frère, soeur 19

Parent, beau-parent 7

Autre 6

7 2

6 2

Name: LIEN_7, dtype: int64

*****RE_ENF*****

Non 4043

Oui 1937

[Nsp] 474

Name: RE_ENF, dtype: int64

*****RE_VAC*****

Oui 5730

Non 566

[Nsp] 158

Name: RE_VAC, dtype: int64

*****RELEG*****

Non, plutôt pas 1302

Oui, plutôt 1069

Oui, tout à fait 311

Non, pas du tout 290

[Nsp] 44

Name: RELEG, dtype: int64

*****PRIVPUB*****

Privé 3633

Public 1706

Name: PRIVPUB, dtype: int64

*****zau2010*****

Commune appartenant à grand pôle (10 000 emplois ou +)	6770
Commune appartenant à couronne d'un grand pôle	1897
Autre commune multipolarisée	541
Commune multipolarisée des grandes aires urbaines	526
Commune appartenant à petit pôle (de 1.500 à 5 000 emplois)	456
Commune isolée hors influence des pôles	414
Commune appartenant à moyen pôle (5 à 10 000 emplois)	340
Commune appartenant à couronne d'un moyen pôle	44
Commune appartenant à couronne d'un petit pôle	22

Name: zau2010, dtype: int64

*****INQALIM*****

Un peu	4245
Assez	3325
Pas du tout	1940
Beaucoup	1412
[Nsp]	209

Name: INQALIM, dtype: int64

*****AGE5*****

40 à 59 ans	3945
25 à 39 ans	2203
70 ans et plus	1732
Moins de 25 ans	1686
60 à 69 ans	1565

Name: AGE5, dtype: int64

*****RADWHY7*****

0	1119
Création d'emploi, chômage, embauche, salaires	53

Name: RADWHY7, dtype: int64

*****RADWHY10*****

0	1024
Agir sur la sécurité, plus d'ordre, une justice plus sévère	148

Name: RADWHY10, dtype: int64

*****CLASSES0*****

La classe moyenne inférieure	4693
La classe moyenne supérieure	2880
La classe populaire	2036
Les défavorisés	681
Les gens aisés	560

Les privilégiés	171
[Nsp]	110

Name: CLASSES0, dtype: int64

*****COMMU6*****

Non	2252
Oui	637
[Nsp]	127

Name: COMMU6, dtype: int64

*****NBCHOM*****

Augmenter pendant plusieurs années	5915
Se stabiliser dans un avenir proche	4200
Diminuer dans un avenir proche	913
[Nsp]	103

Name: NBCHOM, dtype: int64

*****PREOEURO*****

Non	10648
Oui	483

Name: PREOEURO, dtype: int64

*****UDA5*****

Nord Ouest	713
Sud Est	695
Nord Est	666
Ile de France	531
Sud Ouest	326

Name: UDA5, dtype: int64

*****com*****

31555	33
33063	24
44109	22
67482	18
83137	16
34172	15
75120	15
66136	14
06088	14
37261	13
75111	13
76540	12

75115	12
35238	12
75114	12
75112	12
29019	11
62193	11
75119	11
49007	11
80021	10
59350	10
75113	9
69383	9
69387	9
59599	8
54395	8
21231	8
84007	8
67437	8
	..
71092	1
63210	1
42240	1
13102	1
76410	1
38257	1
62502	1
14754	1
79311	1
71194	1
85295	1
86128	1
44035	1
59670	1
85219	1
62321	1
2A006	1
68084	1
66231	1
44187	1
89163	1
35322	1
33483	1
60120	1
14352	1
28191	1
39453	1
50516	1
29212	1

```

93015      1
Name: com, Length: 1906, dtype: int64
-----

*****CONFASS0*****
Plutôt confiance          7061
Plutôt pas confiance      2330
Très confiance            852
Pas du tout confiance     691
[Nsp]                     197
Name: CONFASS0, dtype: int64
-----

*****RADI1*****
Oui, je comprends         1691
Non, je ne comprends pas  1282
[Nsp]                     43
Name: RADI1, dtype: int64
-----

*****DEPLOG*****
Une charge que vous pouvez supporter sans difficulté  4622
Une lourde charge                                     4445
Une très lourde charge                                1025
Une charge négligeable                                719
Une charge à laquelle vous ne pouvez faire face       268
[Nsp]                                                  52
Name: DEPLOG, dtype: int64
-----

*****COMMU3*****
Non          2434
Oui          463
[Nsp]        119
Name: COMMU3, dtype: int64
-----

*****ISEGO*****
Assez souvent  2590
Rarement      2129
Très souvent   871
Jamais         356
[Nsp]          90
Name: ISEGO, dtype: int64
-----

*****ACM1*****
Radical          4988

```

Progressiste	4231
Pas de réforme	1710
Nsp si réforme	133
Nsp quelle réforme	69

Name: ACM1, dtype: int64

*****RADWHY1*****

0	1091
Position anti-Elite, anti-Etat	81

Name: RADWHY1, dtype: int64

*****RESIDALT*****

La résidence alternée chez le père et la mère	2319
La résidence principale chez la mère	590
[Nsp]	60
La résidence principale chez le père	47

Name: RESIDALT, dtype: int64

*****CDV5*****

Vont rester semblables	3661
Vont se détériorer un petit peu	3591
Vont s'améliorer un petit peu	1821
Vont se détériorer beaucoup	1604
Vont s'améliorer beaucoup	357
[Nsp]	97

Name: CDV5, dtype: int64

*****INQAGRE3*****

Inquiet des risques d'agression dans la rue	6075
Non inquiet	4855
[Nsp]	201

Name: INQAGRE3, dtype: int64

*****ACM3*****

Charge de logement lourde ou +	5738
Charge de logement supportable	4622
Charge de logement négligeable	719
[Nsp]	52

Name: ACM3, dtype: int64

*****EXERCPRO*****

Exerce ou a déjà exercé une profession	10141
--	-------

N'a jamais exercé de profession 990

Name: EXERCPRO, dtype: int64

*****PREOPAUF*****

Non 8693

Oui 2438

Name: PREOPAUF, dtype: int64

*****SEXE_5*****

Homme 442

Femme 389

Name: SEXE_5, dtype: int64

*****SOUFFINS*****

Non 7102

Oui 3801

[Nsp] 228

Name: SOUFFINS, dtype: int64

*****FREQSPOR*****

Jamais 5458

Exceptionnellement 2907

Régulièrement 2674

[Nsp] 92

Name: FREQSPOR, dtype: int64

*****type99*****

URBAIN 4672

RURAL 1364

Name: type99, dtype: int64

*****SEXE*****

Femme 5848

Homme 5283

Name: SEXE, dtype: int64

*****UNIONGAY*****

Tout à fait d'accord 4853

Assez d'accord 3125

Peu d'accord 1627

Pas du tout d'accord 1493

[Nsp] 33
Name: UNIONGAY, dtype: int64

*****SOUFFNER*****

Non 7306
Oui 3589
[Nsp] 236
Name: SOUFFNER, dtype: int64

*****MONDIAL*****

Plutôt des inconvénients 4065
Plutôt des avantages 3874
[Nsp] 142
Name: MONDIAL, dtype: int64

*****PREOIMMI*****

Non 8085
Oui 3046
Name: PREOIMMI, dtype: int64

*****PCSRED10*****

Retraité	1572
Employé	858
Profession Intermédiaire	845
Ouvrier	779
Étudiant	650
Cadre et profession intellectuelle supérieure	546
Autre inactif	381
Artisan, commerçant et chef d'entreprise	223
Personnel de service	158
Agriculteur exploitant	24

Name: PCSRED10, dtype: int64

*****INQNUCL3*****

Non inquiet	6342
Inquiet des risques d'un accident de centrale nucléaire	4568
[Nsp]	221

Name: INQNUCL3, dtype: int64

*****FREQCINE*****

Exceptionnellement	5575
Régulièrement	3256

Jamais	2196
[Nsp]	104

Name: FREQCINE, dtype: int64

*****typcont2*****

À durée indéterminée	4346
À durée déterminée	992
[Nsp]	1

Name: typcont2, dtype: int64

*****CONFIGOUV*****

Pas du tout confiance	4069
Plutôt pas confiance	3625
Plutôt confiance	2882
Tout à fait confiance	430
[Nsp]	125

Name: CONFIGOUV, dtype: int64

*****PRATCOLL*****

Jamais	5518
De temps en temps	1600
Rarement	1497
Souvent	390
[Nsp]	54

Name: PRATCOLL, dtype: int64

*****NB06_4*****

Aucun enfant de moins de 6 ans	10046
Un enfant de moins de 6 ans	811
Deux enfants de moins de 6 ans	251
Trois et plus enfants de moins de 6 ans	23

Name: NB06_4, dtype: int64

*****PCSENQ36*****

Ancien cadre	1258
Élève, étudiant	1134
Ouvrier qualifié	1009
Ancienne profession intermédiaire	750
Employé administratif d'entreprise	632
Ancien employé	622
Employé civil, agent de service fonction publique	546
Personne au foyer	477
Technicien	366

Employé de commerce	359
Profession intermédiaire de la santé et du travail social	338
Ouvrier non qualifié	332
Cadre administratif et commercial d'entreprise	313
Profession inter. adm. et comm. des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Ingénieur et cadre technique d'entreprise	218
Profession intermédiaire de la fonction publique	211
Commerçant	190
Ancien artisan, commerçant, chef d'entreprise	179
Cadre de la fonction publique	156
Profession libérale	152
Ancien ouvrier	148
Artisan	142
Professeur, profession scientifique	137
Professeur des écoles, instituteur et assimilé	132
Autre inactif	118
Policier et militaire	113
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise 10 salariés ou plus	87
Agriculteur	43
Ouvrier agricole	38
Profession de l'information, des arts et des spectacles	35
Ancien agriculteur exploitant	11
Clergé, religieux	5
Name: PCSENQ36, dtype: int64	

*****SITUEMP3*****

Actif occupé	5889
Inactif	4498
Chômeur	744
Name: SITUEMP3, dtype: int64	

*****QUOTAAGE*****

30 à 49 ans	981
50 à 59 ans	491
70 ans et plus	471
20 à 29 ans	434
60 à 69 ans	416
Moins de 20 ans	227
Name: QUOTAAGE, dtype: int64	

*****PREOENV*****

Non	8949
-----	------

Oui 2182
Name: PREOENV, dtype: int64

*****TYPCONT*****
À durée indéterminée 4346
À durée déterminée 729
[Nsp] 1
Name: TYPCONT, dtype: int64

*****ASSOCONS*****
Non 9890
Oui 693
[Nsp] 548
Name: ASSOCONS, dtype: int64

*****WHYLIM*****
C'est bon pour la santé 623
C'est bon pour l'environnement, la planète 463
Ca limite les souffrances animales 379
Ce sont des produits qui coûtent cher 178
[Nsp] 14
Name: WHYLIM, dtype: int64

*****RADWHY4*****
0 1098
Plus d'éducation, formation professionnelle des jeunes, s'occuper des étudiants 74
Name: RADWHY4, dtype: int64

*****PREOTENS*****
Non 9503
Oui 1628
Name: PREOTENS, dtype: int64

*****INQ4SUR6*****
Autre 6937
Inquiet 2985
Non inquiet 1209
Name: INQ4SUR6, dtype: int64

*****STATLOG4*****
Propriétaire 4857

Locataire,sous-locataire ou logé gratuitement	4381
Accédant à la propriété	1855
Autre	28
[Nsp]	10

Name: STATLOG4, dtype: int64

*****COMMU4*****

Non	2030
Oui	874
[Nsp]	112

Name: COMMU4, dtype: int64

*****TYPOCLAS*****

Classes moyennes inférieures	3420
Catégories modestes	2130
Classes moyennes supérieures	1884
Catégories pauvres	1147
Catégories aisées	1133
Hauts revenus	1070
Non réponse	347

Name: TYPOCLAS, dtype: int64

*****EMP7*****

Actif	5889
Retraité	2972
Etudiant	955
Chomeur	744
Reste au foyer	417
Malade, invalide	154

Name: EMP7, dtype: int64

*****ADNORDI*****

Assez d'accord	1050
Pas très d'accord	756
Pas du tout d'accord	728
Tout à fait d'accord	441
[Nsp]	41

Name: ADNORDI, dtype: int64

*****RE_VOIT*****

Oui	3784
Non	2310
[Nsp]	360

Name: RE_VOIT, dtype: int64

*****LOGSUFFI*****

Oui 9788

Non 1279

[Nsp] 64

Name: LOGSUFFI, dtype: int64

*****RECEP*****

Plus rarement 3203

En moyenne, une fois par mois 2907

En moyenne, une fois par semaine 1256

Jamais 482

Tous les jours ou presque 177

[Nsp] 56

Name: RECEP, dtype: int64

*****ASSOSPOR*****

Non 8132

Oui 2594

[Nsp] 405

Name: ASSOSPOR, dtype: int64

*****UDA23*****

Ile de France 1988

Rhône-Alpes 1120

Provence Alpes Côte d'Azur 841

Nord Pas de Calais 716

Pays de la Loire 646

Bretagne 619

Aquitaine 584

Centre 486

Midi-Pyrénées 478

Languedoc-Roussillon 472

Lorraine 463

Picardie 303

Bourgogne 295

Haute-Normandie 288

Poitou-Charente 279

Alsace 268

Basse-Normandie 261

DOM 239

Champagne-Ardenne 225

Auvergne 212

Franche-Comté	198
Limousin	112
Corse	38

Name: UDA23, dtype: int64

*****CHERCHEM*****

Non 8816

Oui 2279

[Nsp] 36

Name: CHERCHEM, dtype: int64

*****SEXE_8*****

Femme 18

Homme 6

Name: SEXE_8, dtype: int64

*****BANQVIE*****

Oui 5552

Non 5410

[Nsp] 169

Name: BANQVIE, dtype: int64

*****PCSENQ15*****

Retraité 2968

Employé 1650

Cadre moyen 1570

Ouvrier 1341

Étudiant 1134

Cadre supérieur 859

Personne au foyer 595

Commerçant, chef d'entreprise 277

Personnel de service 259

Profession libérale 152

Artisan 142

Autre inactif 103

Agriculteur exploitant 43

Ouvrier agricole 38

Name: PCSENQ15, dtype: int64

*****BANQEPA*****

Oui 8546

Non 2451

[Nsp] 134

Name: BANQEPA, dtype: int64

*****CONFWEB*****

Assez confiance 4504

Peu confiance 3226

Pas du tout confiance 906

Très confiance 336

[Nsp] 114

Name: CONFWEB, dtype: int64

*****ACM11*****

Restrictions : oui 6454

Restrictions : non 4571

[Nsp] 106

Name: ACM11, dtype: int64

*****ACM9*****

Nombre de maux : trois ou + 3293

Nombre de maux : aucun 2986

Nombre de maux : un 2612

Nombre de maux : deux 2240

Name: ACM9, dtype: int64

*****couple2*****

Oui, avec une personne qui vit dans le logement 1837

Non 919

Oui, avec une personne qui ne vit pas dans le logement 260

Name: couple2, dtype: int64

*****UDA10*****

Ile de France 1988

Ouest 1544

Méditerranée 1351

Sud-Est 1332

Sud-Ouest 1174

Bassin parisien Ouest 1035

Est 929

Bassin parisien Est 823

Nord 716

DOM 239

Name: UDA10, dtype: int64

*****SOUFFDEP*****

Non 9124
 Oui 1743
 [Nsp] 264
 Name: SOUFFDEP, dtype: int64

*****LIEN_3*****

Enfant, beau-fils, belle-fille 3147
 Parent, beau-parent 820
 Frère, soeur 325
 Autre 99
 Conjoint ou compagnon 57
 6 17
 7 11
 Name: LIEN_3, dtype: int64

*****NB03_2*****

Non 10527
 Oui, enfant de moins de 3 ans 604
 Name: NB03_2, dtype: int64

*****INQNUCLE*****

Un peu 3781
 Assez 2688
 Pas du tout 2561
 Beaucoup 1880
 [Nsp] 221
 Name: INQNUCLE, dtype: int64

*****RAISPAUV*****

Plutôt parce qu'elles n'ont pas eu de chance 7157
 Plutôt parce qu'elles n'ont pas fait d'effort 3755
 [Nsp] 219
 Name: RAISPAUV, dtype: int64

*****IMAGTRAV*****

Un moyen de gagner sa vie 2076
 Une source d'épanouissement, de fierté 592
 Une source de stress, de mal-être 333
 [Nsp] 15
 Name: IMAGTRAV, dtype: int64

*****inseel*****

31555	85
44109	59
33063	56
34172	52
67482	51
06088	48
35238	48
75056SR15	42
59350	41
54395	40
75056SR13	36
75056SR20	35
75056SR12	35
75056SR19	35
69123SR03	32
49007	32
83137	30
75056SR11	30
37261	30
75056SR14	29
69266	28
29019	27
51454	27
63113	26
21231	26
76351	25
66136	25
38185	24
14118	23
76540	22
	..
54094	1
44147	1
53107	1
36045	1
45338	1
69275	1
06122	1
35345	1
57447	1
55132	1
28357	1
83102	1
18070	1
70529	1
37006	1
25442	1

17258	1
33061	1
73243	1
35294	1
39097	1
76212	1
48156	1
33442	1
17420	1
14636	1
55155	1
35273	1
57752	1
29272	1

Name: inseel, Length: 3917, dtype: int64

*****VOITURE*****

Oui	9656
Non	1420
[Nsp]	55

Name: VOITURE, dtype: int64

*****ASSOJEUN*****

Non	10128
[Nsp]	542
Oui	461

Name: ASSOJEUN, dtype: int64

*****RURAUURBA*****

PU	5162
PR	3199
IN	2667

Name: RURAUURBA, dtype: int64

*****LIEN_8*****

Frère, soeur	9
Enfant, beau-fils, belle-fille	5
Autre	5
Parent, beau-parent	2
6	2
Conjoint ou compagnon	1
7	1

Name: LIEN_8, dtype: int64

*****DIPL4*****

Diplômé du supérieur	4942
BEPC	3142
BAC	2461
Non diplômé	586

Name: DIPL4, dtype: int64

*****INQALIM3*****

Non inquiet	6185
Inquiet des risques alimentaires	4737
[Nsp]	209

Name: INQALIM3, dtype: int64

*****NBPIECE6*****

Quatre pièces	2775
Cinq pièces	2367
Six et plus	2281
Trois pièces	2207
Deux pièces	1151
Une pièce	350

Name: NBPIECE6, dtype: int64

*****TYPCHAUF*****

Une installation individuelle propre à votre logement	8961
Un système de chauffage collectif	1967
[Nsp]	203

Name: TYPCHAUF, dtype: int64

*****SITUEMP6*****

Autre	4499
Privé CDI	2987
Public CDI	1359
CDD, intérim	992
Chômeur	744
À son compte	550

Name: SITUEMP6, dtype: int64

*****OPIIMMIG*****

Le départ d'un grand nombre d'immigrés	6013
L'intégration des immigrants	4823
[Nsp]	295

Name: OPIIMMIG, dtype: int64

*****NBHEUR35*****

35 heures	1805
36 à 39 heures	1407
Moins de 35 heures	1117
40 à 45 heures	992
46 heures et plus	428
[Nsp]	140

Name: NBHEUR35, dtype: int64

*****AGGL09*****

Plus de 200 000 habitants	2516
Agglomération parisienne	1968
Communes rurales	1783
2 000 à 5 000 habitants	961
50 000 à 100 000 habitants	867
100 000 à 200 000 habitants	733
5 000 à 10 000 habitants	703
20 000 à 50 000 habitants	671
10 000 à 20 000 habitants	591
[Nsp]	338

Name: AGGL09, dtype: int64

*****interim2*****

Non	3416
Oui	217

Name: interim2, dtype: int64

*****TYPLOG2*****

Maison individuelle	6986
Logement collectif	4145

Name: TYPLOG2, dtype: int64

*****RADWHY8*****

0	1055
Moins d'immigration, il faut renvoyer tous les immigrés chez eux	117

Name: RADWHY8, dtype: int64

*****ACM12*****

Famille seul endroit où l'on se sente bien : oui	6089
Famille seul endroit où l'on se sente bien : non	4994
[Nsp]	48

Name: ACM12, dtype: int64

*****SEXE_7*****

Homme 28

Femme 25

Name: SEXE_7, dtype: int64

*****PREOCONF*****

Non 10605

Oui 526

Name: PREOCONF, dtype: int64

*****COMMU8*****

Non 1951

Oui 1065

Name: COMMU8, dtype: int64

*****identifiant*****

9789g6baev28 1

b43jd8d05c 1

c17bn3n09rdt 1

8896d5di6i9 1

9107g3v62366 1

9374ir77wg7g 1

x54mk1k72pdt 1

c47hw4w59c 1

9200bu6c5425 1

jsz47dbe5efz 1

9fblcdh7qr1dfywq 1

xpd43363994z 1

crn49762d76b 1

cft528e885ev 1

r38qn8n05c 1

8464y91nm21m 1

b35hl4l92c 1

8465q6m78222 1

ay6buxobaezmf8zw 1

11532z73jh2h 1

y38tr8r82c 1

unl7ptfrwmgoltpn 1

q6j34287mep3 1

10154c3oo39z 1

cyt4f288373f 1

110303059828 1

d44xp6p02c	1
zysqglpfsefvcrvm	1
b2j9i9gan2zmcjyp	1
110272707031	1
..	
y64rl5l34udt	1
110279884280	1
110179472943	1
ryq44de2fefx	1
h06ya9a32udt	1
110322109988	1
n571l16l47c	1
gvwde7c32mi3	1
u16ki6i00c	1
11594ixv7eg6	1
ces52254679a	1
gsf49a4a92ex	1
a72oe2e71c	1
w87dn3n78c	1
s91gp2p67c	1
pmn477f88bbj	1
110256880356	1
v60tq9q58c	1
h85py4y28c	1
110039936735	1
u14gl0l65c	1
w7k76293b	1
8u3q654x	1
p65lo6o01c	1
x97nc8c47ibd2	1
10679t34xn7n	1
co0xh77xhfa3	1
110304063664	1
11228o90qp8p	1
8791c9f98468	1

Name: identifiant, Length: 3016, dtype: int64

*****revtot7*****

De 1500 à 2300	2333
De 2300 à 3100	2272
4000 et plus	2096
3100 à 4000	1681
De 900 à 1500	1547
Moins de 900	861
Non déclaré (ne sait pas, refus)	341

Name: revtot7, dtype: int64

*****AGGL05*****

Plus de 100 000 habitants	3443
Communes rurales	2467
2 000 à 20 000 habitants	1914
Agglomération parisienne	1776
20 000 à 100 000 habitants	1531

Name: AGGL05, dtype: int64

*****STATLOGB*****

Propriétaire	3877
Locataire ou sous-locataire parc privé	1861
Accédant à la propriété en cours de remboursement d'emprunt	1509
Locataire ou sous-locataire parc social	1049
Logé gratuitement	762
Autre	28

Name: STATLOGB, dtype: int64

*****RADIQUOI*****

je ne sais pas
 tout
 Tout
 nsp
 .
 ras
 ?
 l'immigration
 la justice
 ...
 NSP
 Nsp
 immigration
 Je sais pas
 La justice
 Je ne sais pas
 sait pas
 !ducation
 ne sait pas
 Ras

la scolarit! des enfants n'est pas du tout adapt!e, les aides sont trop nombreuses, les !tudes n
 Arr!ter d'assister les tricheurs (travail au noir),contr!ler efficacement les pensions de retrain
 Virer tous les politiciens !narques, tous les guignols qui sont d!put!s/s!nateurs ! vie et ne d!
 d!velopper la notion d'interet public. Diminuer les moyens de recours qui entravent toute !volut
 une justice equitable,des aides cibl!es et control!es,des prestations prioritairement aux franca
 egalite
 revisite des remun!rations des hauts postes public, idem pour les politiques, limitation des aid

reconnaitre le travail
la suppression des !lites politiques, des audits sur les syndicats, la fins des aides et subvent
Code du travail,Imp!ts,Immigration

il y en a beaucoup : ,les peines d'emprisonnement pour commencer qui sont trop cl!mentes
y en a tellement
Arr!ter l'immigration, r!tablir la peine de mort, supprimer toutes les aides et allocations aux
le chômage et la discrimination au travail
D!velopper l?initiative et l?entreprenariat. R!server l?assistance ! ceux qui en ont vraiment be
l' immigration, la s!curit!
Sant! Ecole justice administration
Remettre ! plat le r!le de l'Etat, supprimer l'accessoire et centrer les moyens sur ce qui est l
ne pas tenir compe des demandes des arrivants
r!organiser les aides sociales,renforcer la police et la justice
trop de chose
aucune idee
EGALITE !!!
Je sait pas
Stop aux aides,Stop ! la corruption,Stop au gaspillage de nos imp!ts,Stop ! la politique r!mun!r
Donner plus de pouvoirs aux jeunes,, faire une scission compl!te avec le pass! !!
EMPLOI,AIDES SOCIALES,MIGRATIONS
La nationalit! fran!aise
l'interdiction de faire gr!ve pour les fonctionnaires
encourager le travail,assurer la s!curit! des biens et des personnes
sterilisation
lib!ralisme
Soci!t! plus capitaliste
Tout ?,Anarchie
Arr!t de l' immigration, arr!t des avantages aux familles de d!linquants
arr!ter l'assistanat et valoriser les travailleurs, supprimer certains postes publics inutiles (
Passage a l !cologie , agriculture biologique ,
refonte du syst!me politique ,refonte du syst!me scolaire ,changement dans le syst!me d'!volutio
retablir la peine de mort
DISCRIMINATION, injustice
Name: RADIQUEOI, Length: 916, dtype: int64

*****NB03_4*****

Aucun enfant de moins de 3 ans	10527
Un enfant de moins de 3 ans	552
Deux enfants de moins de 3 ans	49
Trois et plus enfants de moins de 3 ans	3

Name: NB03_4, dtype: int64

*****RADWHY13*****

0	1030
Autre	142

Name: RADWHY13, dtype: int64

*****CDV5_4*****

Moins bien 5195

Pareil 3661

Mieux 2178

[Nsp] 97

Name: CDV5_4, dtype: int64

*****INQMALAD*****

Beaucoup 4516

Assez 4100

Un peu 1861

Pas du tout 508

[Nsp] 146

Name: INQMALAD, dtype: int64

*****PREOPAUM*****

Non 9787

Oui 1344

Name: PREOPAUM, dtype: int64

*****NIVFRAN4*****

Moins bien 7752

Pareil 2331

Mieux 875

[Nsp] 173

Name: NIVFRAN4, dtype: int64

*****CHOVOLON*****

Plutôt d'accord 6351

Plutôt pas d'accord 4614

[Nsp] 166

Name: CHOVOLON, dtype: int64

*****zau1999*****

Commune pôle urbain 6930

Commune monopolarisée 1705

Commune espace dominante rurale 1195

Commune pole emploi espace rural 615

Commune multipolarisée 545

Commune couronne pole emploi espace rural 36

Name: zau1999, dtype: int64

*****STATMAT4*****

En ménage, marié	6599
Célibataire	3157
Séparé(e), divorcé(e)	979
Veuf(ve)	396

Name: STATMAT4, dtype: int64

*****NB20_4*****

Aucun enfant de moins de 20 ans	8147
Un enfant de moins de 20 ans	1336
Deux enfants de moins de 20 ans	1190
Trois et plus enfants de moins de 20 ans	458

Name: NB20_4, dtype: int64

*****LIEN_5*****

Enfant, beau-fils, belle-fille	562
Frère, soeur	222
Parent, beau-parent	56
6	28
Autre	22
Conjoint ou compagnon	9
7	4

Name: LIEN_5, dtype: int64

*****PCSENQ7*****

Retraité	2968
Cadre, prof. intellectuelle sup., profession intermédiaire	2581
Employé	1909
Ouvrier	1379
Autre inactif	1237
Personne au foyer	595
Agri. exploitant, artisan, commerçant et chef d'entreprise	462

Name: PCSENQ7, dtype: int64

*****COMMU2*****

Non	2063
Oui	828
[Nsp]	125

Name: COMMU2, dtype: int64

*****CONDUIT*****

Oui	9319
N'a pas le permis de conduire	1170
N'a pas conduit, mais a le permis de conduire	574
[Nsp]	68

Name: CONDUIT, dtype: int64

*****CADVIE*****

Satisfait	6844
Très satisfait	2781
Peu satisfait	1184
Pas satisfait du tout	275
[Nsp]	47

Name: CADVIE, dtype: int64

*****ASS010_2*****

Non adhérent	5642
Adhérent	5489

Name: ASS010_2, dtype: int64

*****RADWHY5*****

0	1107
Moins de social : fraudes, abus, profiteurs,	65

Name: RADWHY5, dtype: int64

*****NBPERS5*****

Deux personnes	4390
Une personne	2219
Trois personnes	1881
Quatre personnes	1717
Cinq personnes et plus	924

Name: NBPERS5, dtype: int64

*****BANQMOB*****

Non	8311
Oui	2639
[Nsp]	181

Name: BANQMOB, dtype: int64

*****INQUER3*****

Inquiet des risques d'une guerre	5808
Non inquiet	5104

[Nsp] 219
Name: INQGUER3, dtype: int64

*****HANDICAP*****
Non 8273
Oui 2766
[Nsp] 92
Name: HANDICAP, dtype: int64

*****ASS011_2*****
Adhérent 5779
Non adhérent 5352
Name: ASS011_2, dtype: int64

*****FREQTELE*****
Tous les jours 7354
Assez souvent 2172
Pas très souvent 1038
Jamais 479
[Nsp] 88
Name: FREQTELE, dtype: int64

*****DIPLOME*****
Bac + 3 et supérieur 2832
Bac généraliste, technologique ou professionnel 2461
Diplôme professionnel court de type CAP ou BEP 2294
Bac + 2 : Deug, IUT, DUT, BTS 2110
BEPC, brevet des collèges 848
Aucun diplôme 306
Certificat d études primaires 280
Name: DIPLOME, dtype: int64

*****CHOAVANT*****
Plutôt d'accord 8656
Plutôt pas d'accord 2328
[Nsp] 147
Name: CHOAVANT, dtype: int64

*****TRAVFEM*****
Elles devraient travailler quand elles le désirent 8623
Elles ne devraient travailler que si nécessaire 878
Elles ne devraient jamais travailler avec des enfants en bas âge 853

Elles devraient toujours travailler	652
Elles ne devraient jamais travailler	69
[Nsp]	56

Name: TRAVFEM, dtype: int64

*****NB10_4*****

Aucun enfant de moins de 10 ans	9453
Un enfant de moins de 10 ans	1015
Deux enfants de moins de 10 ans	549
Trois et plus enfants de moins de 10 ans	114

Name: NB10_4, dtype: int64

*****CONFKEUF*****

Plutôt confiance	5538
Très confiance	1457
Plutôt pas confiance	1415
Pas du tout confiance	530
[Nsp]	146

Name: CONFKEUF, dtype: int64

*****decsqt*****

Décile 7	1289
Décile 3	1136
Décile 5	1117
Décile 10	1101
Décile 9	1089
Décile 1	1062
Décile 8	1047
Décile 2	1025
Décile 4	984
Décile 6	934
Non renseigné	347

Name: decsqt, dtype: int64

*****ACM7*****

CDV dans 5 ans : semblables	3661
CDV dans 5 ans : détériorer un peu	3591
CDV dans 5 ans : améliorer	2178
CDV dans 5 ans : détériorer beaucoup	1604
[Nsp]	97

Name: ACM7, dtype: int64

*****DEPCOM*****

75056	252
69123	76
13055	65
31555	60
44109	41
33063	38
67482	33
06088	31
35238	29
9D411	28
34172	26
59350	25
9D416	25
54395	25
83137	24
49007	22
37261	22
66136	20
29019	19
76540	19
51454	19
14118	17
80021	16
21231	15
62193	15
63113	15
38185	15
76351	15
59599	14
57463	14
...	
68082	1
71194	1
42225	1
70180	1
44203	1
74268	1
37241	1
79156	1
44089	1
33506	1
13029	1
85219	1
67478	1
62321	1
14366	1
9B218	1
21200	1

11441	1
33483	1
89163	1
60106	1
25442	1
49377	1
29160	1
14471	1
14352	1
66231	1
68084	1
40296	1
29272	1

Name: DEPCOM, Length: 3073, dtype: int64

*****INNOVTEC*****

Assez attiré	5421
Peu attiré	3548
Très attiré	1328
Pas du tout attiré	734
[Nsp]	100

Name: INNOVTEC, dtype: int64

*****NB99_2*****

Non	7543
Oui, enfant	3588

Name: NB99_2, dtype: int64

*****ROBOT1*****

C'est plutôt une mauvaise chose	1616
C'est plutôt une bonne chose	1321
[Nsp]	79

Name: ROBOT1, dtype: int64

*****AIDESUFF*****

Suffisante	7335
Insuffisante	3636
[Nsp]	160

Name: AIDESUFF, dtype: int64

*****ASS011_3*****

Aucune	5352
Une	3127

Deux ou plus 2652
Name: ASS011_3, dtype: int64

*****NB20_2*****

Non 8147
Oui, enfant de moins de 20 ans 2984
Name: NB20_2, dtype: int64

*****AUTREAL*****

Voyage, vacances 53
Autre (mariage, études, développement personnel, artistique) 48
Travaux, amélioration de l'habitat 22
Achat d'un véhicule automobile 13
Name: AUTREAL, dtype: int64

*****AGEDIP2*****

30-49 ans diplôme bac et + 2538
20-29 ans diplôme bac et + 1404
50-59 ans diplôme bac et + 1128
30-49 ans diplôme inférieur au bac 1103
60-69 ans diplôme bac et + 1014
70 ans et + diplôme bac et + 974
70 ans et + diplôme inférieur au bac 758
50-59 ans diplôme inférieur au bac 714
Moins de 20 ans tous diplômes confondus 703
60-69 ans diplôme inférieur au bac 551
20-29 ans diplôme inférieur au bac 244
Name: AGEDIP2, dtype: int64

*****VACANCES*****

Oui 7457
Non 3588
[Nsp] 86
Name: VACANCES, dtype: int64

*****SITUEMP*****

Exerce un emploi 5710
Retraité, retiré des affaires 2972
Étudiant ou élève, sans activité professionnelle 955
Chômeur (ayant déjà travaillé) 622
Reste au foyer sans chercher d'emploi 405
Étudiant qui travaille 179
Invalide ou malade de longue durée 154

Recherche un premier emploi	122
Aide un membre de sa famille, sans rémunération	12

Name: SITUEMP, dtype: int64

*****ASSOCULT*****

Non	8421
Oui	2294
[Nsp]	416

Name: ASSOCULT, dtype: int64

*****NBENF6*****

Aucun enfant	4024
Deux enfants	3129
Un enfant	1883
Trois enfants	1469
Quatre enfants	417
Cinq enfants et plus	209

Name: NBENF6, dtype: int64

*****SALCOMP*****

Salarié	5339
À votre compte	550

Name: SALCOMP, dtype: int64

*****PCSCONJ*****

Ouvrier qualifié	599
Employé civil, agent de service fonction publique	329
Cadre administratif et commercial d'entreprise	274
Employé administratif d'entreprise	255
Employé de commerce	229
Technicien	219
Ingénieur et cadre technique d'entreprise	178
Personnel des services directs aux particuliers	169
Ouvrier non qualifié	156
Profession intermédiaire santé et travail social	150
Cadre de la fonction publique	141
Profession intermédiaire fonction publique	140
Profession libérale	139
Profession inter. adm. et com.des entreprises	132
Contremaître, agent de maîtrise	128
Professeur, profession scientifique	115
Artisan	103
Commerçant	99
Professeur des écoles, instituteur et assimilé	74

Policier et militaire	63
Agriculteur	39
Chef d'entreprise 10 salariés ou plus	38
Ouvrier agricole	34
Profession de l'information, des arts et des spectacles	22
Nsp, NR	10
Clergé, religieux	1

Name: PCSCONJ, dtype: int64

*****ACM8*****

Fonctionnement justice : assez mal	5365
Fonctionnement justice : bien	3286
Fonctionnement justice : très mal	2379
[Nsp]	101

Name: ACM8, dtype: int64

*****info*****

ok	1826
pb	772
revtranc seul	301
rien	105
revtot seul	12

Name: info, dtype: int64

*****CHAMP*****

18 ans et + métropole	10574
15-17 ans + DOM + Corse	557

Name: CHAMP, dtype: int64

*****VISITFAM*****

Oui	6550
Non	1459
[Nsp]	72

Name: VISITFAM, dtype: int64

*****REVAUON*****

Non	6522
Oui	1144
1E6	15

Name: REVAUON, dtype: int64

*****PCSENQ10*****

Retraité	2968
Autre inactif	1832
Employé	1650
Profession Intermédiaire	1570
Ouvrier	1341
Cadre et profession intellectuelle supérieure	1011
Artisan, commerçant et chef d'entreprise	419
Personnel de service	259
Agriculteur exploitant	43
Ouvrier agricole	38

Name: PCSEMQ10, dtype: int64

*****AGESEX12*****

30-49 ans, femme	1835
30-49 ans, homme	1806
20-29 ans, femme	1137
50-59 ans, homme	950
60-69 ans, homme	930
50-59 ans, femme	892
70-99 ans, femme	890
70-99 ans, homme	842
60-69 ans, femme	635
20-29 ans, homme	511
18-19 ans, femme	459
18-19 ans, homme	244

Name: AGESEX12, dtype: int64

*****CHOIXNUC*****

Plutôt des avantages	4749
Plutôt des inconvénients	3314
Sans avis	3068

Name: CHOIXNUC, dtype: int64

*****SITUEMP5*****

Autre	4498
Privé	3633
Public	1706
Chômeur	744
À son compte	550

Name: SITUEMP5, dtype: int64

*****LIEN_2*****

Conjoint ou compagnon	6725
Parent, beau-parent	960

Enfant, beau-fils, belle-fille	655
Frère, soeur	354
Autre	101
6	32
7	27

Name: LIEN_2, dtype: int64

*****RE_HABI*****

Oui 4949

Non 1248

[Nsp] 257

Name: RE_HABI, dtype: int64

*****RADWHY12*****

0 732

Nsp, NR, réformulation de la question ou contournement de la réponse 440

Name: RADWHY12, dtype: int64

*****SECUR3*****

En sécurité 9513

Pas en sécurité 1557

Nsp, NR 61

Name: SECUR3, dtype: int64

*****PREFPALI*****

Une amélioration de votre pouvoir d'achat 4259

Un temps libre plus long 1616

[Nsp] 14

Name: PREFPALI, dtype: int64

*****ASSOENVI*****

Non 10069

Oui 533

[Nsp] 529

Name: ASSOENVI, dtype: int64

*****PREOMALA*****

Non 9064

Oui 2067

Name: PREOMALA, dtype: int64

*****COMMU1*****

Non 2284
 Oui 594
 [Nsp] 138
 Name: COMMU1, dtype: int64

*****PCSENQ24*****

Ancien cadre et profession intermédiaire	2008
Inactif divers (autre que retraité)	1729
Ouvrier qualifié	1009
Ancien employé et ouvrier	770
Prof. inter. enseignement, santé, fonct. publ. et assimilé	686
Employé de la fonction publique	659
Employé administratifs d'entreprise	632
Cadre d'entreprise	531
Technicien	366
Employé de commerce	359
Ouvrier non qualifié	332
Cadre fonction publique, profession intellect. & artistique	328
Prof. inter. administrative et commerciale des entreprises	288
Personnel des services directs aux particuliers	259
Contremaître, agent de maîtrise	230
Commerçant et assimilé	190
Ancien artisan, commerçant, chef d'entreprise	179
Profession libérale et assimilé	152
Artisan	142
Chômeur n'ayant jamais travaillé	103
Chef d'entreprise de 10 salariés ou plus	87
Agriculteur exploitant	43
Ouvrier agricole	38
Ancien agriculteur exploitant	11

Name: PCSENQ24, dtype: int64

*****TYPOSQT*****

Classes moyennes supérieures	2873
Bas revenus	2815
Classes moyennes inférieures	2610
Hauts revenus	2486
Non réponse	347

Name: TYPOSQT, dtype: int64

*****PCSENQ8*****

Retraité	2968
Employé	1909
Profession Intermédiaire	1570

Ouvrier	1379
Autre inactif	1237
Cadre et profession intellectuelle supérieure	1011
Personne au foyer	595
Agri. exploitant, artisan, commerçant et chef d'entreprise	462

Name: PCSNQ8, dtype: int64

*****RE_MEDI*****

Non	3906
Oui	2155
[Nsp]	393

Name: RE_MEDI, dtype: int64

*****ASSOHUMA*****

Non	9375
Oui	1256
[Nsp]	500

Name: ASSOHUMA, dtype: int64

*****PROGRAD*****

Changements radicaux	4988
Réformes progressives	4231
[Nsp]	69

Name: PROGRAD, dtype: int64

*****CADVIE3*****

Satisfait de son cadre de vie	9625
Non satisfait de son cadre de vie	1459
[Nsp]	47

Name: CADVIE3, dtype: int64

*****RADWHY3*****

0	1116
Plus de justice sociale, réduction des inégalités	56

Name: RADWHY3, dtype: int64

*****PREOCCU2*****

La violence et l'insécurité	2126
L'immigration	1581
La pauvreté en France	1367
Le chômage	1141
La dégradation de l'environnement	1119

Les tensions internationales	932
Les maladies graves	849
La pauvreté dans le monde	801
Les conflits sociaux	348
L'Europe	343
La drogue	278
[Nsp, NR]	246

Name: PREOCCU2, dtype: int64

*****PCSENQ13*****

Retraité, autre inactif	3071
Employé	1650
Cadre moyen	1570
Ouvrier	1341
Étudiant	1134
Profession libérale, cadre supérieur	1011
Personne au foyer	595
Commerçant, chef d'entreprise	277
Personnel de service	259
Artisan	142
Agriculteur exploitant	43
Ouvrier agricole	38

Name: PCSENQ13, dtype: int64

*****statut99*****

C	3877
B	3655
R	2503
I	991

Name: statut99, dtype: int64

*****PREODROG*****

Non	10695
Oui	436

Name: PREODROG, dtype: int64

*****DPT*****

59	491
75	482
69	399
33	304
13	287
78	255
44	251

92	248
97	239
31	238
77	237
62	224
94	222
83	206
29	200
35	200
76	197
34	190
91	188
38	185
57	185
67	182
95	180
93	176
06	169
54	150
49	148
42	146
45	130
14	128
	...
11	57
36	55
79	54
04	52
47	50
70	48
53	44
39	44
16	39
82	39
08	37
58	35
61	35
55	35
19	33
12	32
65	31
52	31
32	30
09	24
46	23
90	23
43	21
05	19

2A	19
2B	19
15	19
23	14
48	5
20	3

Name: DPT, Length: 98, dtype: int64

*****TRANSFST*****

Oui	9288
Non	1710
[Nsp]	133

Name: TRANSFST, dtype: int64

*****NB16_4*****

Aucun enfant de moins de 16 ans	8658
Un enfant de moins de 16 ans	1209
Deux enfants de moins de 16 ans	949
Trois et plus enfants de moins de 16 ans	315

Name: NB16_4, dtype: int64

*****ROBOT2*****

C'est plutôt une mauvaise chose	1551
C'est plutôt une bonne chose	1388
[Nsp]	77

Name: ROBOT2, dtype: int64

*****STATMAT*****

Marié(e)	4736
Célibataire	3157
Vivant maritalement (ou PACS)	1863
Séparé(e), divorcé(e)	979
Veuf(ve)	396

Name: STATMAT, dtype: int64

*****PCSENQ9*****

Retraité	2968
Employé	1909
Profession Intermédiaire	1570
Ouvrier	1379
Étudiant	1134
Cadre et profession intellectuelle supérieure	1011
Autre inactif	698

Artisan, commerçant et chef d'entreprise	419
Agriculteur exploitant	43

Name: PCSNQ9, dtype: int64

*****LIEN_4*****

Enfant, beau-fils, belle-fille	1807
Frère, soeur	497
Parent, beau-parent	163
6	55
Autre	50
Conjoint ou compagnon	29
7	12

Name: LIEN_4, dtype: int64

*****COMINSEE*****

75056	482
69123	151
13055	111
31555	108
44109	72
33063	65
34172	65
06088	60
67482	60
35238	60
59350	52
54395	46
49007	38
97411	37
83137	37
69266	35
37261	34
51454	33
66136	33
21231	32
29019	32
97416	32
38185	30
63113	28
14118	27
57463	26
76540	25
87085	25
30189	25
29232	25

...

02741	1
17024	1
29105	1
62869	1
35273	1
55155	1
91156	1
14636	1
28357	1
21461	1
83102	1
18070	1
13214	1
70529	1
37006	1
76259	1
34028	1
63102	1
25650	1
59332	1
89309	1
33061	1
73243	1
35294	1
39097	1
48156	1
76212	1
33442	1
17420	1
29272	1

Name: COMINSEE, Length: 4467, dtype: int64

*****CONFPUB*****

Plutôt confiance	6026
Plutôt pas confiance	3418
Pas du tout confiance	1030
Très confiance	503
[Nsp]	154

Name: CONFPUB, dtype: int64

*****NB16_2*****

Non	8658
Oui, enfant de moins de 16 ans	2473

Name: NB16_2, dtype: int64

*****COMMU5*****

Non 2513
 Oui 386
 [Nsp] 117
 Name: COMMU5, dtype: int64

*****ACM2*****

Devraient travailler selon désir ou toujours 9275
 Ne devraient jamais travailler, si enfants 922
 Devraient travailler si besoin de 2 salaires 878
 [Nsp] 56
 Name: ACM2, dtype: int64

*****NBHEUR39*****

Moins de 39 heures 3698
 39 heures 631
 40 heures 602
 41 à 45 heures 390
 46 à 50 heures 248
 Plus de 50 heures 180
 [Nsp] 140
 Name: NBHEUR39, dtype: int64

*****NB06_2*****

Non 10046
 Oui, enfant de moins de 6 ans 1085
 Name: NB06_2, dtype: int64

*****INQROUT3*****

Inquiet des risques d'accident de la route 6896
 Non inquiet 4030
 [Nsp] 205
 Name: INQROUT3, dtype: int64

*****RADWHY6*****

0 1129
 Réduire les impôts, taxes, cotisations patronales 43
 Name: RADWHY6, dtype: int64

*****LIMVIAND*****

Vous ne l'envisagez pas du tout 1320
 Vous limitez déjà votre consommation 1069

Vous envisagez de limiter votre consommation	588
[Nsp]	39
Name: LIMVIAND, dtype: int64	

*****PREOCHOM*****	
Non	7274
Oui	3857
Name: PREOCHOM, dtype: int64	

*****INQCHOMA*****	
Assez	3648
Un peu	2840
Beaucoup	2543
Pas du tout	1900
[Nsp]	200
Name: INQCHOMA, dtype: int64	

*****ACM10*****	
Évolution niveau de vie perso : pareil	3445
Évolution niveau de vie perso : un peu moins bien	3248
Évolution niveau de vie perso : beaucoup moins bien	1980
Évolution niveau de vie perso : un peu mieux	1727
Évolution niveau de vie perso : beaucoup mieux	613
[Nsp]	118
Name: ACM10, dtype: int64	

*****UDA14*****	
Ile de France	1988
Auvergne - Rhône Alpes	1332
Nord Pas de Calais - Picardie	1019
Aquitaine - Limousin - Poitou Charente	975
Alsace - Champagne - Ardennes Lorraine	956
Languedoc Roussillon - Midi Pyrénées	950
Provence Alpes Côte d'Azur	841
Pays de la Loire	646
Bretagne	619
Normandie	549
Bourgogne - Franche-Comté	493
Centre Val de Loire	486
DOM	239
Corse	38
Name: UDA14, dtype: int64	

*****RE_EQUI*****

Oui 4670
Non 1465
[Nsp] 319
Name: RE_EQUI, dtype: int64

*****NIVPERS4*****

Moins bien 5228
Pareil 3445
Mieux 2340
[Nsp] 118
Name: NIVPERS4, dtype: int64

*****ASSOSYND*****

Non 9744
Oui 850
[Nsp] 537
Name: ASSOSYND, dtype: int64

*****RADWHY11*****

0 1158
Nouvelles formes du vivre ensemble, écologie, décroissance, fin du capitalisme 14
Name: RADWHY11, dtype: int64

*****ASS010_3*****

Aucune 5642
Une 3194
Deux ou plus 2295
Name: ASS010_3, dtype: int64

*****RE_LOG*****

Non 3836
Oui 2172
[Nsp] 446
Name: RE_LOG, dtype: int64

*****ASSOPOLI*****

Non 9969
Oui 608
[Nsp] 554
Name: ASSOPOLI, dtype: int64

*****RADWHY9*****

0

Protectionnisme économique, contre l'Europe, renforcer les filières industrielles françaises

Name: RADWHY9, dtype: int64

*****ORDLIB*****

Un peu plus d'ordre 4711

Beaucoup plus d'ordre 3699

Un peu plus de liberté 1817

Beaucoup plus de liberté 739

[Nsp] 165

Name: ORDLIB, dtype: int64

*****TYPEEMPL*****

D'une entreprise privée 3462

De la fonction publique 1389

D'une entreprise publique 317

D'une association 171

Name: TYPEEMPL, dtype: int64

*****CONFPRES*****

Plutôt pas confiance 4186

Plutôt confiance 2556

Pas du tout confiance 2008

Très confiance 180

[Nsp] 156

Name: CONFPRES, dtype: int64

*****COUPLE*****

Oui, avec une personne qui vit dans le logement 5437

Non 2900

Oui, avec une personne qui ne vit pas dans le logement 749

Name: COUPLE, dtype: int64

In [125]: ord_var = {

"CONFPOLI", "AGE5", "SECURITE", "ACM7", "INQUERR", "NBPiece6", "INNOVTEC",
"JUSTICE", "EFFORTPP", "ACM10", "NIVFRAN4", "NBPERS5", "INQCHOMA", "CDV5_4",
"CONFGOUV", "ADOPTGAY", "ACM8", "FREQCINE", "CONFPUB", "FREQSPOR", "INQALIM",
"ASSO10_3", "FREQBIBL", "DEPLOG", "NBCHOM", "CONFENTR", "ORDLIB", "ACM5",
"INQMALAD", "FREQTELE", "NBENF6", "ACM9", "revtot7", "INROUTE", "NIVPERS4",

```

        "ETATSAN", "INQNUCLE", "NIVPERSON", "CONFASSO", "ACM6", "CDV5", "UNIONGAY",
        "ACM4", "INQAGRES", "CADVIE", "NIVFRAN", "REV_TR7", "ISEGO", "RECEP", "AGE6",
        "ADNCB", "PRATCOLL", "NBHEUR39", "HARVEY", "QUOTAAGE", "NBHEUR35", "RELEG",
        "CONFKEUF", "CONFECOL", "ADNSTIC", "ADNORDI", "CONFPRES", "CONFWEB", "CONFBANK"
    }

```

```

In [126]: # review
          for c in ord_var:
              print(f'\n-----{c}-----')
              print(cdv[c].value_counts())

```

```

-----CONFENTR-----
Plutôt confiance          6402
Plutôt pas confiance      3131
Pas du tout confiance     829
Très confiance            569
[Nsp]                     200
Name: CONFENTR, dtype: int64

```

```

-----FREQTELE-----
Tous les jours           7354
Assez souvent            2172
Pas très souvent        1038
Jamais                   479
[Nsp]                    88
Name: FREQTELE, dtype: int64

```

```

-----CONFPOLI-----
Pas du tout confiance    4700
Plutôt pas confiance     4543
Plutôt confiance         1524
[Nsp]                    201
Très confiance           163
Name: CONFPOLI, dtype: int64

```

```

-----CONFKEUF-----
Plutôt confiance         5538
Très confiance           1457
Plutôt pas confiance     1415
Pas du tout confiance     530
[Nsp]                    146
Name: CONFKEUF, dtype: int64

```

```

-----ADNORDI-----
Assez d'accord           1050
Pas très d'accord        756
Pas du tout d'accord     728

```


Tout à fait d'accord 441
[Nsp] 41
Name: ADNORDI, dtype: int64

-----ACM7-----
CDV dans 5 ans : semblables 3661
CDV dans 5 ans : détériorer un peu 3591
CDV dans 5 ans : améliorer 2178
CDV dans 5 ans : détériorer beaucoup 1604
[Nsp] 97
Name: ACM7, dtype: int64

-----INNOVTEC-----
Assez attiré 5421
Peu attiré 3548
Très attiré 1328
Pas du tout attiré 734
[Nsp] 100
Name: INNOVTEC, dtype: int64

-----INQROUTE-----
Assez 4261
Un peu 3312
Beaucoup 2635
Pas du tout 718
[Nsp] 205
Name: INQROUTE, dtype: int64

-----ETATSAN-----
Satisfaisant 7386
Très satisfaisant 1681
Peu satisfaisant 1650
Pas satisfaisant du tout 333
[Nsp] 81
Name: ETATSAN, dtype: int64

-----RECEP-----
Plus rarement 3203
En moyenne, une fois par mois 2907
En moyenne, une fois par semaine 1256
Jamais 482
Tous les jours ou presque 177
[Nsp] 56
Name: RECEP, dtype: int64

-----ADNCB-----
Pas du tout d'accord 1102
Pas très d'accord 929

Assez d'accord	683
Tout à fait d'accord	265
[Nsp]	37

Name: ADN CB, dtype: int64

-----CONFWEB-----

Assez confiance	4504
Peu confiance	3226
Pas du tout confiance	906
Très confiance	336
[Nsp]	114

Name: CONFWEB, dtype: int64

-----INQAGRES-----

Assez	3600
Un peu	3409
Beaucoup	2475
Pas du tout	1446
[Nsp]	201

Name: INQAGRES, dtype: int64

-----FREQBIBL-----

Jamais	5381
Exceptionnellement	3437
Régulièrement	2210
[Nsp]	103

Name: FREQBIBL, dtype: int64

-----SECURITE-----

Assez en sécurité	6773
Tout à fait en sécurité	2740
Peu en sécurité	1271
Pas du tout en sécurité	286
[Nsp]	61

Name: SECURITE, dtype: int64

-----ACM8-----

Fonctionnement justice : assez mal	5365
Fonctionnement justice : bien	3286
Fonctionnement justice : très mal	2379
[Nsp]	101

Name: ACM8, dtype: int64

-----NBENF6-----

Aucun enfant	4024
Deux enfants	3129
Un enfant	1883
Trois enfants	1469

Quatre enfants 417
Cinq enfants et plus 209
Name: NBENF6, dtype: int64

-----INQNUCLE-----

Un peu 3781
Assez 2688
Pas du tout 2561
Beaucoup 1880
[Nsp] 221
Name: INQNUCLE, dtype: int64

-----NIVPERSO-----

C'est pareil 3445
Un peu moins bien 3248
Beaucoup moins bien 1980
Un peu mieux 1727
Beaucoup mieux 613
[Nsp] 118
Name: NIVPERSO, dtype: int64

-----REV_TR7-----

De 1500 à 2300 2332
De 2300 à 3100 2281
4000 et plus 2087
3100 à 4000 1675
De 900 à 1500 1544
Moins de 900 862
Non déclaré (ne sait pas, refus) 350
Name: REV_TR7, dtype: int64

-----RELEG-----

Non, plutôt pas 1302
Oui, plutôt 1069
Oui, tout à fait 311
Non, pas du tout 290
[Nsp] 44
Name: RELEG, dtype: int64

-----HARVEY-----

On n'en parle pas assez 1438
On en parle suffisamment 1258
On en parle trop 284
[Nsp] 36
Name: HARVEY, dtype: int64

-----CONFECOL-----

Plutôt confiance 5475

Plutôt pas confiance	2018
Très confiance	888
Pas du tout confiance	556
[Nsp]	149

Name: CONFECOL, dtype: int64

-----INQALIM-----

Un peu	4245
Assez	3325
Pas du tout	1940
Beaucoup	1412
[Nsp]	209

Name: INQALIM, dtype: int64

-----AGE5-----

40 à 59 ans	3945
25 à 39 ans	2203
70 ans et plus	1732
Moins de 25 ans	1686
60 à 69 ans	1565

Name: AGE5, dtype: int64

-----CONFBANK-----

Plutôt pas confiance	3624
Plutôt confiance	2987
Pas du tout confiance	2079
Très confiance	239
[Nsp]	157

Name: CONFBANK, dtype: int64

-----NBPIECE6-----

Quatre pièces	2775
Cinq pièces	2367
Six et plus	2281
Trois pièces	2207
Deux pièces	1151
Une pièce	350

Name: NBPIECE6, dtype: int64

-----AGE6-----

40 à 59 ans	3216
25 à 39 ans	1796
70 ans et plus	1397
60 à 69 ans	1250
18 à 24 ans	1147
Moins de 18 ans	280

Name: AGE6, dtype: int64

-----NBCHOM-----

Augmenter pendant plusieurs années	5915
Se stabiliser dans un avenir proche	4200
Diminuer dans un avenir proche	913
[Nsp]	103

Name: NBCHOM, dtype: int64

-----NBHEUR35-----

35 heures	1805
36 à 39 heures	1407
Moins de 35 heures	1117
40 à 45 heures	992
46 heures et plus	428
[Nsp]	140

Name: NBHEUR35, dtype: int64

-----ADOPTGAY-----

Tout à fait d'accord	3387
Pas du tout d'accord	2826
Assez d'accord	2750
Peu d'accord	2098
[Nsp]	70

Name: ADOPTGAY, dtype: int64

-----CONFASSO-----

Plutôt confiance	7061
Plutôt pas confiance	2330
Très confiance	852
Pas du tout confiance	691
[Nsp]	197

Name: CONFASSO, dtype: int64

-----DEPLOG-----

Une charge que vous pouvez supporter sans difficulté	4622
Une lourde charge	4445
Une très lourde charge	1025
Une charge négligeable	719
Une charge à laquelle vous ne pouvez faire face	268
[Nsp]	52

Name: DEPLOG, dtype: int64

-----ISEGO-----

Assez souvent	2590
Rarement	2129
Très souvent	871
Jamais	356
[Nsp]	90

Name: ISEGO, dtype: int64

-----JUSTICE-----

Assez mal	5365
Assez bien	3080
Très mal	2379
Très bien	206
[Nsp]	101

Name: JUSTICE, dtype: int64

-----ACM4-----

Satisfait de son cadre de vie	6844
Très satisfait de son cadre de vie	2781
Pas satisfait de son cadre de vie	1459
[Nsp]	47

Name: ACM4, dtype: int64

-----CDV5-----

Vont rester semblables	3661
Vont se détériorer un petit peu	3591
Vont s'améliorer un petit peu	1821
Vont se détériorer beaucoup	1604
Vont s'améliorer beaucoup	357
[Nsp]	97

Name: CDV5, dtype: int64

-----CONFPRES-----

Plutôt pas confiance	4186
Plutôt confiance	2556
Pas du tout confiance	2008
Très confiance	180
[Nsp]	156

Name: CONFPRES, dtype: int64

-----ADNSTIC-----

Assez d'accord	1187
Tout à fait d'accord	978
Pas très d'accord	420
Pas du tout d'accord	395
[Nsp]	36

Name: ADNSTIC, dtype: int64

-----ACM9-----

Nombre de maux : trois ou +	3293
Nombre de maux : aucun	2986
Nombre de maux : un	2612
Nombre de maux : deux	2240

Name: ACM9, dtype: int64

```

-----revtot7-----
De 1500 à 2300                2333
De 2300 à 3100                2272
4000 et plus                  2096
3100 à 4000                   1681
De 900 à 1500                 1547
Moins de 900                  861
Non déclaré (ne sait pas, refus) 341
Name: revtot7, dtype: int64

```

```

-----EFFORTPP-----
Ne font pas assez             4656
Font ce qu'ils doivent       4348
Font trop                     1965
[Nsp]                         162
Name: EFFORTPP, dtype: int64

```

```

-----NIVFRAN-----
Un peu moins bien            4766
Beaucoup moins bien         2986
C'est pareil                 2331
Un peu mieux                 749
[Nsp]                        173
Beaucoup mieux               126
Name: NIVFRAN, dtype: int64

```

```

-----INQGUERR-----
Un peu                       3348
Assez                        3175
Beaucoup                     2633
Pas du tout                  1756
[Nsp]                        219
Name: INQGUERR, dtype: int64

```

```

-----CDV5_4-----
Moins bien                   5195
Pareil                       3661
Mieux                        2178
[Nsp]                         97
Name: CDV5_4, dtype: int64

```

```

-----CONFPUB-----
Plutôt confiance             6026
Plutôt pas confiance         3418
Pas du tout confiance        1030
Très confiance               503
[Nsp]                        154
Name: CONFPUB, dtype: int64

```

```

-----INQMALAD-----
Beaucoup      4516
Assez         4100
Un peu        1861
Pas du tout    508
[Nsp]          146
Name: INQMALAD, dtype: int64

```

```

-----FREQSPOR-----
Jamais         5458
Exceptionnellement 2907
Régulièrement  2674
[Nsp]           92
Name: FREQSPOR, dtype: int64

```

```

-----NBHEUR39-----
Moins de 39 heures  3698
39 heures           631
40 heures           602
41 à 45 heures      390
46 à 50 heures      248
Plus de 50 heures   180
[Nsp]               140
Name: NBHEUR39, dtype: int64

```

```

-----NIVFRAN4-----
Moins bien      7752
Pareil          2331
Mieux           875
[Nsp]           173
Name: NIVFRAN4, dtype: int64

```

```

-----INQCHOMA-----
Assez           3648
Un peu          2840
Beaucoup        2543
Pas du tout     1900
[Nsp]           200
Name: INQCHOMA, dtype: int64

```

```

-----ACM5-----
État de santé satisfaisant 7386
État de santé pas satisfaisant 1983
État de santé très satisfaisant 1681
[Nsp]                        81
Name: ACM5, dtype: int64

```



```

-----ACM10-----
Évolution niveau de vie perso : pareil          3445
Évolution niveau de vie perso : un peu moins bien 3248
Évolution niveau de vie perso : beaucoup moins bien 1980
Évolution niveau de vie perso : un peu mieux      1727
Évolution niveau de vie perso : beaucoup mieux    613
[Nsp]                                              118
Name: ACM10, dtype: int64

```

```

-----UNIONGAY-----
Tout à fait d'accord      4853
Assez d'accord             3125
Peu d'accord               1627
Pas du tout d'accord       1493
[Nsp]                      33
Name: UNIONGAY, dtype: int64

```

```

-----ACM6-----
Niveau de vie Français : un peu moins bien      4766
Niveau de vie Français : beaucoup moins bien    2986
Niveau de vie Français : pareil                 2331
Niveau de vie Français : mieux                  875
[Nsp]                                            173
Name: ACM6, dtype: int64

```

```

-----NIVPERS4-----
Moins bien      5228
Pareil          3445
Mieux           2340
[Nsp]           118
Name: NIVPERS4, dtype: int64

```

```

-----FREQCINE-----
Exceptionnellement  5575
Régulièrement      3256
Jamais             2196
[Nsp]              104
Name: FREQCINE, dtype: int64

```

```

-----CONFGOUV-----
Pas du tout confiance  4069
Plutôt pas confiance  3625
Plutôt confiance      2882
Tout à fait confiance  430
[Nsp]                  125
Name: CONFGOUV, dtype: int64

```

```

-----ASS010_3-----

```

Aucune	5642
Une	3194
Deux ou plus	2295

Name: ASS010_3, dtype: int64

-----PRATCOLL-----

Jamais	5518
De temps en temps	1600
Rarement	1497
Souvent	390
[Nsp]	54

Name: PRATCOLL, dtype: int64

-----ORDLIB-----

Un peu plus d'ordre	4711
Beaucoup plus d'ordre	3699
Un peu plus de liberté	1817
Beaucoup plus de liberté	739
[Nsp]	165

Name: ORDLIB, dtype: int64

-----QUOTAAGE-----

30 à 49 ans	981
50 à 59 ans	491
70 ans et plus	471
20 à 29 ans	434
60 à 69 ans	416
Moins de 20 ans	227

Name: QUOTAAGE, dtype: int64

-----CADVIE-----

Satisfait	6844
Très satisfait	2781
Peu satisfait	1184
Pas satisfait du tout	275
[Nsp]	47

Name: CADVIE, dtype: int64

-----NBPERS5-----

Deux personnes	4390
Une personne	2219
Trois personnes	1881
Quatre personnes	1717
Cinq personnes et plus	924

Name: NBPERS5, dtype: int64

Doubtfull ... classified as categorial

Classes moyennes inférieures 3420 Catégories modestes 2130 Classes moyennes supérieures 1884
Catégories pauvres 1147 Catégories aisées 1133 Hauts revenus 1070 Non réponse 347 Name:
TYPOCLAS, dtype: int64

La classe moyenne inférieure 4693 La classe moyenne supérieure 2880 La classe populaire 2036
Les défavorisés 681 Les gens aisés 560 Les privilégiés 171 [Nsp] 110 Name: CLASSES0, dtype:
int64

Aucune 5352 Une 3127 Deux ou plus 2652 Name: ASSO11_3, dtype: int64

NB10_4 Aucun enfant de moins de 10 ans 9453 Un enfant de moins de 10 ans 1015 Deux enfants de moins de 10 ans 549 Trois et plus enfants de moins de 10 ans 114 Name: NB10_4, dtype: int64

Aucun enfant de moins de 20 ans 8147 Un enfant de moins de 20 ans 1336 Deux enfants de moins de 20 ans 1190 Trois et plus enfants de moins de 20 ans 458 Name: NB20_4, dtype: int64

Aucun enfant de moins de 6 ans 10046 Un enfant de moins de 6 ans 811 Deux enfants de moins de 6 ans 251 Trois et plus enfants de moins de 6 ans 23 Name: NB06_4, dtype: int64

Aucun enfant de 20 ans et plus 7543 Un enfant de 20 ans et plus 1579 Deux enfants de 20 ans et plus 1409 Trois et plus enfants de 20 ans et plus 600 Name: NB99_4, dtype: int64

Jamais 2128 Plus rarement 408 En moyenne, une fois par semaine 228 En moyenne, une fois par mois 139 Tous les jours ou presque 73 [Nsp] 40 Name: YOGA, dtype: int64

```
In [127]: # exclusion of features with order
```

```
cat_var = obj_var - ord_var
cat_max9_var = cat_max9_var - ord_var
cat_min10_var = cat_min10_var - ord_var
```

```
In [128]: scope = cat_min10_var - com_var - tech_var - bizz_var
cdv.loc[:,scope].dropna().head()
```

```
Out[128]: PCSENQ13 \
```

8148	Profession libérale, cadre supérieur
8152	Personne au foyer
8182	Profession libérale, cadre supérieur
8186	Ouvrier
8192	Personne au foyer

	PCSCONJ	UDA14	\
8148	Cadre administratif et commercial d'entreprise	Auvergne - Rhône Alpes	
8152	Artisan	Pays de la Loire	
8182	Professeur, profession scientifique	Ile de France	
8186	Personnel des services directs aux particuliers	Auvergne - Rhône Alpes	
8192	Ingénieur et cadre technique d'entreprise	Ile de France	

RADIQUOI \

8148		code du travail, fiscalit!
8152		Tout
8182	Revoir le syst!me des aides par exemple. R!dui...	
8186		LES IMMIGRANTS
8192		moins de paperasses

		PCSENQ32 DPT \
8148	Cadre administratif et commercial d'entreprise	69
8152	Personnes au foyer, a.inac.	44
8182	Ingénieur et cadre technique d'entreprise	75
8186	Ouvrier non qualifié	69
8192	Personnes au foyer, a.inac.	78

	MODCHAUF	AGESEX12 \
8148	Électricité	20-29 ans, femme
8152	Gaz naturel (gaz de réseau)	30-49 ans, femme
8182	Gaz naturel (gaz de réseau)	30-49 ans, homme
8186	Bois	30-49 ans, homme
8192	Gaz naturel (gaz de réseau)	50-59 ans, femme

	PCSENQ35	UDA23 \
8148	Cadre administratif et commercial d'entreprise	Rhône-Alpes
8152	Personne au foyer et autre inactif	Pays de la Loire
8182	Ingénieur et cadre technique d'entreprise	Ile de France
8186	Ouvrier non qualifié	Rhône-Alpes
8192	Personne au foyer et autre inactif	Ile de France

	PCSENQ36	PCSENQ15 \
8148	Cadre administratif et commercial d'entreprise	Cadre supérieur
8152	Personne au foyer	Personne au foyer
8182	Ingénieur et cadre technique d'entreprise	Cadre supérieur
8186	Ouvrier non qualifié	Ouvrier
8192	Personne au foyer	Personne au foyer

	PCSENQ24	AGEDIP2 \
8148	Cadre d'entreprise	20-29 ans diplôme bac et +
8152	Inactif divers (autre que retraité)	30-49 ans diplôme bac et +
8182	Cadre d'entreprise	30-49 ans diplôme bac et +
8186	Ouvrier non qualifié	30-49 ans diplôme inférieur au bac
8192	Inactif divers (autre que retraité)	50-59 ans diplôme bac et +

	PREOCCU1	PREOCCU2
8148	Les conflits sociaux	La drogue
8152	La pauvreté en France	La violence et l'insécurité
8182	La pauvreté dans le monde	La violence et l'insécurité
8186	L'immigration	La violence et l'insécurité
8192	Les maladies graves	La dégradation de l'environnement

```

In [129]: cdv_dtypes = cdv.dtypes
          cdv_dtypes.value_counts()

Out[129]: object      292
          float64      47
          int64       15
          dtype: int64

In [130]: int_var = set(cdv_dtypes[cdv_dtypes == 'int64'].index)
          scope = int_var - ( tech_var | com_var | bizz_var)
          cdv.loc[:,scope].dropna().head()

Out[130]:   INTER6  AUTREREV  REVENQ  AGE
0  373001          0    1100    56
1  373002          0    2000    64
2  373003          0    5000    71
3  373004          0     200    21
4  373005          0     998    70

In [131]: cdv["REVTOT6"].value_counts()

Out[131]: 5    3777
          3    2333
          4    2272
          2    1547
          1     861
          6     341
          Name: REVTOT6, dtype: int64

In [132]: cdv["NB2099"].value_counts()

Out[132]: 0    10210
          1     733
          2     166
          3       19
          4        3
          Name: NB2099, dtype: int64

In [133]: float_var = set(cdv_dtypes[cdv_dtypes == 'float64'].index)
          float_var

Out[133]: {'AGE_2',
          'AGE_3',
          'AGE_4',
          'AGE_5',
          'AGE_6',
          'AGE_7',
          'AGE_8',
          'AGE_9',

```

```

'AGGLOINS',
'CP',
'CSP6',
'NBENF',
'NBPERS',
'NBPIECES',
'NBUC',
'NOT_AMIS',
'NOT_CAD',
'NOT_COHE',
'NOT_FAMI',
'NOT_LIBR',
'NOT_LOG',
'NOT_POLI',
'NOT_PROF',
'POND',
'REVAUTR',
'REVCONJ',
'REVPF',
'REVTOT',
'REVUC',
'REVsqt',
'SENSIENV',
'TYPLOG',
'age_OW',
'agedip',
'cpt',
'i',
'inseenum',
'med',
'nbheures',
'pmun',
'popdense',
'popinter',
'poppeud',
'poptrpeu',
'prescaf',
'refus2',
'typodeg'}

```

```

In [134]: for c in float_var:
            print(f"\n*****{c}*****\n")
            print(cdv.loc[:,c].dropna().head())
            print()
            print(cdv.loc[:,c].value_counts())
            print('\n-----')

```

```

*****pmun*****

```

5095	3666.0
5096	1759.0
5097	2108.0
5098	89683.0
5099	15193.0

Name: pmun, dtype: float64

2243833.0	252
484344.0	76
850726.0	65
441802.0	60
284970.0	41
239157.0	38
271782.0	33
343304.0	31
207178.0	29
139041.5	28
257351.0	26
227560.0	25
105421.0	25
73154.0	25
164532.0	24
147571.0	22
134817.0	22
117419.0	20
141303.0	19
179992.0	19
110933.0	19
108954.0	17
133448.0	16
73636.0	15
139860.0	15
151212.0	15
155637.0	15
175497.0	15
145150.0	14
48597.0	14

...

7976.0	1
2914.0	1
3084.0	1
2522.0	1
3308.0	1
2648.0	1
1963.0	1
894.0	1
1722.0	1

2472.0	1
4844.0	1
307.0	1
5939.0	1
829.0	1
5724.0	1
1650.0	1
2046.0	1
1205.0	1
361.0	1
334.0	1
3301.0	1
28252.0	1
7899.0	1
1071.0	1
6518.0	1
16040.0	1
3791.0	1
2970.0	1
8142.0	1
4117.0	1

Name: pmun, Length: 2613, dtype: int64

*****REVsqst*****

0	1300.000000
1	1626.345597
2	3535.533906
3	200.000000
4	1378.858223

Name: REVsqst, dtype: float64

1909.188309	212
2545.584412	119
1350.000000	118
1351.563902	113
1991.778381	113
1558.845727	108
2489.722977	106
1357.645020	105
2474.873734	96
1356.937913	91
1408.400000	82
2499.622471	82
1963.989085	79
898.025612	73

1378.858223	73
892.722311	70
1103.547304	68
1626.280238	67
1750.000000	65
1800.000000	63
955.700000	63
656.500000	62
2032.850298	57
737.634251	57
643.000000	57
903.413766	56
1603.590373	54
1107.935167	54
1207.476708	53
1500.000000	52

...

2675.000000	1
355.674711	1
1983.481076	1
1542.055087	1
5356.950000	1
1227.446672	1
698.164000	1
227.334830	1
501.000000	1
2532.000000	1
1773.022325	1
2166.000000	1
687.000000	1
3165.504927	1
2390.807465	1
2653.213162	1
1344.450407	1
1322.657376	1
1326.023457	1
4140.053635	1
934.574000	1
3521.391770	1
3348.631561	1
977.832000	1
739.000000	1
1363.187537	1
1754.059400	1
1934.870427	1
4161.546113	1
1730.320000	1

Name: REVsqt, Length: 2700, dtype: int64

*****poppeud*****

5095 0.916634
5096 0.951524
5097 0.992725
5098 0.029086
5099 0.000000

Name: poppeud, dtype: float64

0.000000 1154
0.002968 65
1.000000 62
0.003304 60
0.003109 41
0.001663 38
0.000597 33
0.005062 31
0.006939 29
0.019994 28
0.005479 26
0.051877 25
0.000463 25
0.000190 25
0.008149 24
0.006897 22
0.004035 22
0.017736 20
0.000685 19
0.010591 19
0.001516 19
0.005091 16
0.006232 15
0.008267 15
0.005114 15
0.002866 15
0.001582 14
0.002416 14
0.083082 13
0.017469 12
...
0.002635 1
0.914222 1
0.867999 1
0.889811 1
0.940653 1

0.653233	1
0.810985	1
0.104468	1
0.038498	1
0.938351	1
0.935377	1
0.013047	1
0.076611	1
0.091583	1
0.934559	1
0.913124	1
0.975590	1
0.162927	1
0.994314	1
0.002494	1
0.206173	1
0.147660	1
0.958484	1
0.133873	1
0.083382	1
0.397907	1
0.979203	1
0.955548	1
0.223838	1
0.982799	1

Name: poppeud, Length: 2631, dtype: int64

*****med*****

8115	1900.0
8116	1250.0
8117	3500.0
8118	2750.0
8119	5400.0

Name: med, dtype: float64

1900.0	625
2750.0	598
5400.0	537
1250.0	438
3500.0	430
650.0	271

Name: med, dtype: int64

*****POND*****

```
0    1.313554
1    2.009015
2    0.217607
3    0.539351
4    0.270204
```

Name: POND, dtype: float64

```
0.296629    43
1.351546    26
0.729224    25
0.220198    20
0.837901    20
1.418262    20
0.976527    20
0.489276    19
0.919534    19
0.304132    19
1.073495    19
3.503147    18
0.535591    17
1.014747    16
1.435315    16
0.754585    16
1.063953    15
1.021787    14
0.901699    14
0.424336    14
0.577182    13
0.266499    13
0.804324    13
0.254228    13
0.751297    13
1.050573    12
0.504567    12
0.430487    12
0.289693    12
0.287580    12
..
1.018006     1
0.546925     1
0.942931     1
1.602928     1
0.683508     1
0.695874     1
0.677009     1
1.009876     1
```

1.099839	1
1.539247	1
0.235651	1
0.619655	1
0.483961	1
0.836043	1
0.617186	1
1.211959	1
1.760275	1
1.695799	1
0.753918	1
0.566479	1
0.708472	1
0.446195	1
2.620703	1
1.063263	1
0.711988	1
0.610579	1
0.441396	1
2.955229	1
1.646769	1
1.745518	1

Name: POND, Length: 7200, dtype: int64

*****nbheures*****

0	35.0
7	40.0
9	50.0
16	35.0
18	38.0

Name: nbheures, dtype: float64

35.0	1805
39.0	631
40.0	602
37.0	357
38.0	282
45.0	202
50.0	202
20.0	191
30.0	166
999999.0	140
36.0	137
42.0	118
25.0	88

60.0	87
32.0	77
28.0	74
24.0	67
15.0	53
10.0	44
70.0	41
41.0	33
48.0	32
27.0	31
21.0	29
26.0	27
18.0	26
55.0	22
8.0	20
31.0	20
43.0	20
	...
16.0	17
22.0	16
23.0	14
17.0	14
6.0	12
33.0	12
29.0	11
4.0	10
19.0	9
5.0	9
14.0	8
13.0	7
7.0	7
65.0	7
3.0	6
56.0	6
2.0	6
49.0	6
52.0	5
46.0	4
47.0	4
11.0	4
54.0	3
64.0	3
1.0	3
9.0	3
57.0	2
51.0	2
68.0	1
59.0	1

Name: nbheures, Length: 63, dtype: int64

*****NBPIECES*****

0	7.0
1	5.0
2	8.0
3	1.0
4	7.0

Name: NBPIECES, dtype: float64

4.0	2775
5.0	2367
3.0	2207
6.0	1224
2.0	1151
7.0	556
1.0	350
8.0	284
9.0	110
99.0	63
10.0	21
11.0	5
13.0	3
12.0	3
20.0	3

Name: NBPIECES, dtype: int64

*****AGE_6*****

98	12.0
125	19.0
194	8.0
306	68.0
358	38.0

Name: AGE_6, dtype: float64

6.0	19
1.0	16
9.0	15
5.0	13
4.0	12
12.0	11
14.0	11

7.0	11
2.0	10
10.0	10
11.0	10
3.0	9
8.0	8
15.0	6
16.0	6
17.0	6
13.0	5
0.0	4
18.0	3
19.0	3
42.0	2
50.0	2
25.0	2
36.0	2
21.0	2
22.0	2
52.0	1
20.0	1
53.0	1
30.0	1
65.0	1
39.0	1
23.0	1
24.0	1
56.0	1
48.0	1
40.0	1
44.0	1
49.0	1
68.0	1
66.0	1
46.0	1
47.0	1
34.0	1
41.0	1
75.0	1
38.0	1
71.0	1

Name: AGE_6, dtype: int64

*****inseenum*****

8115 87203.0

8116	24068.0
8117	71294.0
8118	76351.0
8119	41241.0

Name: inseenum, dtype: float64

31555.0	33
33063.0	24
44109.0	22
67482.0	18
83137.0	16
34172.0	15
75120.0	15
6088.0	14
66136.0	14
37261.0	13
75111.0	13
75115.0	12
35238.0	12
76540.0	12
75112.0	12
75114.0	12
62193.0	11
29019.0	11
75119.0	11
49007.0	11
80021.0	10
59350.0	10
69387.0	9
69383.0	9
75113.0	9
84007.0	8
59599.0	8
67437.0	8
54395.0	8
21231.0	8
	..
74208.0	1
77085.0	1
39245.0	1
81143.0	1
61454.0	1
29226.0	1
31070.0	1
77053.0	1
77032.0	1
33397.0	1
16113.0	1

```

77037.0    1
85231.0    1
1283.0     1
75108.0    1
49060.0    1
85234.0    1
89331.0    1
95476.0    1
33069.0    1
36127.0    1
85239.0    1
13215.0    1
62688.0    1
93057.0    1
91514.0    1
59005.0    1
13091.0    1
63102.0    1
1138.0     1
Name: inseeenum, Length: 1906, dtype: int64

```

*****NOT_AMIS*****

```

0    6.0
1    6.0
2    6.0
3    6.0
4    6.0
Name: NOT_AMIS, dtype: float64

```

```

6.0    3571
5.0    3192
7.0    1613
4.0    1605
3.0     573
2.0     251
1.0     169
Name: NOT_AMIS, dtype: int64

```

*****REVPF*****

```

0    200.0
1     0.0
2     0.0

```

```

3      0.0
4      800.0
Name: REVPF, dtype: float64

```

```

0.0      6669
9999.0    245
129.0     123
200.0     116
100.0     105
300.0      98
500.0      95
120.0      81
250.0      78
400.0      68
150.0      60
130.0      50
600.0      49
700.0      49
350.0      48
800.0      43
450.0      36
180.0      36
194.0      34
1000.0     28
184.0      24
160.0      24
50.0       23
125.0      21
190.0      20
220.0      20
80.0       20
175.0      19
1200.0     19
750.0      18
...
71.0       1
1160.0     1
509.0      1
1907.0     1
636.0      1
193.0      1
303.0      1
341.0      1
724.0      1
285.0      1
410.0      1
274.0      1
139.0      1

```

386.0	1
510.0	1
597.0	1
950.0	1
288.0	1
9.0	1
208.0	1
234.0	1
54.0	1
18.0	1
13.0	1
44.0	1
36.0	1
11.0	1
569.0	1
345.0	1
3000.0	1

Name: REVPF, Length: 495, dtype: int64

*****NOT_LOG*****

5095	7.0
5096	5.0
5097	7.0
5098	6.0
5099	4.0

Name: NOT_LOG, dtype: float64

6.0	2058
5.0	1546
7.0	1277
4.0	709
3.0	207
2.0	81
1.0	54

Name: NOT_LOG, dtype: int64

*****TYPLOG*****

8115	1.0
8116	1.0
8117	1.0
8118	2.0
8119	1.0

Name: TYPL0G, dtype: float64

1.0 1884

2.0 1127

3.0 5

Name: TYPL0G, dtype: int64

*****NBPERS*****

0 1.0

1 2.0

2 2.0

3 1.0

4 2.0

Name: NBPERS, dtype: float64

2.0 4390

1.0 2219

3.0 1881

4.0 1717

5.0 690

6.0 170

7.0 30

8.0 17

9.0 5

10.0 3

Name: NBPERS, dtype: int64

*****typodeg*****

5095 3.0

5096 3.0

5097 3.0

5098 1.0

5099 1.0

Name: typodeg, dtype: float64

1.0 2260

2.0 1855

3.0 1696

4.0 186

Name: typodeg, dtype: int64

*****AGE_4*****

16	62.0
20	17.0
48	15.0
49	21.0
51	23.0

Name: AGE_4, dtype: float64

15.0	133
14.0	133
12.0	126
13.0	122
17.0	114
18.0	110
11.0	109
10.0	107
6.0	107
7.0	103
16.0	103
4.0	98
1.0	96
3.0	93
19.0	90
5.0	89
20.0	87
8.0	85
9.0	79
2.0	76
21.0	69
22.0	47
23.0	45
0.0	44
25.0	37
24.0	30
26.0	17
51.0	15
55.0	14
52.0	14
...	
65.0	4
58.0	4
34.0	4
42.0	3
70.0	3
36.0	3
68.0	3

31.0	3
33.0	3
32.0	2
57.0	2
80.0	2
61.0	2
39.0	2
92.0	2
29.0	2
37.0	2
69.0	2
88.0	1
67.0	1
85.0	1
83.0	1
41.0	1
79.0	1
90.0	1
81.0	1
63.0	1
86.0	1
78.0	1
66.0	1

Name: AGE_4, Length: 79, dtype: int64

*****NOT_PROF*****

0	6.0
1	6.0
2	6.0
3	5.0
4	7.0

Name: NOT_PROF, dtype: float64

5.0	3061
6.0	2528
4.0	1888
7.0	1445
1.0	876
3.0	650
2.0	498

Name: NOT_PROF, dtype: int64

*****NOT_POLI*****

```

0    3.0
1    5.0
2    5.0
3    5.0
4    6.0
Name: NOT_POLI, dtype: float64

```

```

4.0    2578
5.0    2051
3.0    1880
1.0    1613
2.0    1416
6.0     968
7.0     441
Name: NOT_POLI, dtype: int64

```

*****AGE_8*****

```

125    13.0
736     6.0
742     2.0
871     4.0
1581    11.0
Name: AGE_8, dtype: float64

```

```

4.0     4
5.0     3
11.0    2
2.0     2
6.0     2
13.0    2
21.0    1
54.0    1
53.0    1
34.0    1
51.0    1
17.0    1
40.0    1
8.0     1
7.0     1
1.0     1
Name: AGE_8, dtype: int64

```

*****NOT_FAMI*****

0 5.0
1 7.0
2 7.0
3 7.0
4 7.0

Name: NOT_FAMI, dtype: float64

7.0 7110
6.0 1934
5.0 804
4.0 534
1.0 331
3.0 165
2.0 120

Name: NOT_FAMI, dtype: int64

*****NOT_CAD*****

5095 7.0
5096 5.0
5097 7.0
5098 7.0
5099 5.0

Name: NOT_CAD, dtype: float64

6.0 2196
7.0 1603
5.0 1315
4.0 527
3.0 116
2.0 103
1.0 65

Name: NOT_CAD, dtype: int64

*****REVCONJ*****

1 300.0
2 0.0
4 800.0
9 2000.0
12 2000.0

Name: REVCONJ, dtype: float64

0.0	694
1500.0	484
2000.0	468
999999.0	362
1000.0	340
1200.0	337
1800.0	199
2500.0	198
1600.0	189
1400.0	185
3000.0	161
500.0	153
1300.0	148
1700.0	145
800.0	123
1100.0	112
2200.0	111
900.0	98
1900.0	87
600.0	82
700.0	81
2300.0	74
4000.0	72
3500.0	70
2100.0	69
200.0	65
300.0	64
2400.0	58
400.0	56
2800.0	42
	...
620.0	1
1165.0	1
1299.0	1
1290.0	1
2783.0	1
1396.0	1
2850.0	1
1878.0	1
2013.0	1
3030.0	1
2044.0	1
22.0	1
9.0	1
60.0	1
1075.0	1
890.0	1

710.0	1
2523.0	1
2078.0	1
2090.0	1
1222.0	1
1375.0	1
133.0	1
825.0	1
70.0	1
570.0	1
278.0	1
1285.0	1
1470.0	1
1983.0	1

Name: REVCNJ, Length: 337, dtype: int64

*****popinter*****

5095	0.000000
5096	0.000000
5097	0.000000
5098	0.122407
5099	0.000000

Name: popinter, dtype: float64

0.000000	2533
0.002757	76
0.044337	65
0.040686	60
0.016858	41
0.036137	33
0.037514	31
0.014024	29
0.177296	28
0.019312	26
0.946168	25
0.003926	25
0.028595	24
0.025562	22
0.022412	22
0.111928	20
0.007354	19
0.015356	19
0.055694	19
0.036042	17
0.151790	16

0.005711	15
0.056517	15
0.045878	15
0.020174	15
0.018514	15
0.001540	14
0.065212	14
0.914831	13
0.168394	12
...	
0.895050	1
0.918531	1
0.007753	1
0.865622	1
0.838720	1
0.992021	1
0.964405	1
0.111352	1
0.963194	1
0.869003	1
0.724690	1
0.909382	1
0.982764	1
0.013134	1
0.950279	1
0.931794	1
0.900140	1
0.900409	1
0.936877	1
0.622055	1
0.891933	1
0.950136	1
0.002867	1
0.091994	1
0.841527	1
0.718355	1
0.002689	1
0.950624	1
0.000016	1
0.144512	1

Name: popinter, Length: 1426, dtype: int64

*****REVTOT*****

0	1300.0
1	2300.0

```

2    5000.0
3     200.0
4    1950.0
Name: REVTOT, dtype: float64

```

```

2700.000    533
1911.400    314
2816.800    307
1919.000    265
1920.000    259
3521.000    243
3500.000    242
2777.500    228
1262.500    221
3600.000    216
1277.620    208
1950.000    205
1270.000    198
656.500     195
3535.000    194
643.000     153
1262.000    136
686.092     130
3000.000    103
700.000      97
2000.000      92
2500.000      89
5200.000      76
5734.200      75
3030.000      62
2020.000      59
3018.000      57
2200.000      54
5000.000      52
2525.000      52
...
3504.700      1
150.900       1
3960.000      1
2540.000      1
810.000       1
3805.680      1
3472.000      1
2686.020      1
1340.000      1
4090.500      1
1335.000      1
760.530       1

```

5243.272	1
2690.044	1
5915.280	1
5706.500	1
1645.000	1
4130.000	1
6239.212	1
2024.072	1
1764.524	1
3435.490	1
3400.280	1
4971.220	1
2786.620	1
5464.100	1
7102.360	1
1888.700	1
18000.000	1
1933.532	1

Name: REVTOT, Length: 1819, dtype: int64

*****NOT_LIBR*****

0	6.0
1	6.0
2	4.0
3	6.0
4	7.0

Name: NOT_LIBR, dtype: float64

6.0	3549
5.0	3108
7.0	2147
4.0	1476
3.0	388
2.0	180
1.0	112

Name: NOT_LIBR, dtype: int64

*****AGE_7*****

125	16.0
736	8.0
742	5.0
871	5.0

```
1581    13.0
Name: AGE_7, dtype: float64
```

```
4.0     5
9.0     4
13.0    4
10.0    4
3.0     4
2.0     4
1.0     3
63.0    2
12.0    2
8.0     2
5.0     2
7.0     2
0.0     2
16.0    2
6.0     2
19.0    1
21.0    1
46.0    1
70.0    1
89.0    1
27.0    1
11.0    1
15.0    1
37.0    1
41.0    1
```

```
Name: AGE_7, dtype: int64
```

```
-----
```

```
*****AGE_9*****
```

```
125     11.0
871      1.0
4056     1.0
4492     0.0
4807     0.0
```

```
Name: AGE_9, dtype: float64
```

```
0.0     2
1.0     2
19.0    1
37.0    1
60.0    1
11.0    1
```

```
Name: AGE_9, dtype: int64
```

*****NBENF*****

1	3.0
2	3.0
4	5.0
5	1.0
6	1.0

Name: NBENF, dtype: float64

2.0	3129
1.0	1883
3.0	1469
4.0	417
5.0	116
6.0	57
7.0	18
99.0	16
9.0	1
8.0	1

Name: NBENF, dtype: int64

*****popdense*****

5095	0.000000
5096	0.000000
5097	0.000000
5098	0.847785
5099	1.000000

Name: popdense, dtype: float64

0.000000	3577
1.000000	687
0.997243	76
0.951989	65
0.955864	60
0.979959	41
0.998275	38
0.962781	33
0.957408	31
0.978488	29
0.801336	28
0.975110	26
0.999537	25

0.995884	25
0.963025	24
0.970332	22
0.970368	22
0.868544	20
0.991422	19
0.933408	19
0.982881	19
0.963958	17
0.842657	16
0.978475	15
0.936934	15
0.974360	15
0.945107	15
0.994289	15
0.927860	14
0.996878	14
...	
0.859439	1
0.002295	1
0.819878	1
0.808466	1
0.771868	1
0.478912	1
0.937068	1
0.302552	1
0.847565	1
0.971810	1
0.988326	1
0.843231	1
0.897296	1
0.797860	1
0.891764	1
0.627127	1
0.000974	1
0.962132	1
0.912222	1
0.415562	1
0.000082	1
0.859813	1
0.247092	1
0.987733	1
0.000611	1
0.660413	1
0.961878	1
0.823516	1
0.577602	1
0.822051	1

Name: popdense, Length: 400, dtype: int64

*****NOT_COHE*****

0 6.0
1 7.0
3 6.0
4 6.0
5 6.0

Name: NOT_COHE, dtype: float64

5.0 3535
4.0 2982
6.0 1878
3.0 1043
7.0 815
2.0 376
1.0 325

Name: NOT_COHE, dtype: int64

*****REVUC*****

0 1300.000000
1 1533.333333
2 3333.333333
3 200.000000
4 1300.000000

Name: REVUC, dtype: float64

1800.000000 267
2400.000000 134
2000.000000 113
1274.266667 112
1877.866667 112
1300.000000 106
2347.333333 105
1280.000000 105
1500.000000 100
2333.333333 98
1279.333333 86
2356.666667 84
1851.666667 83
1400.000000 81
1350.000000 76

1408.400000	74
846.666667	71
841.666667	70
2020.000000	70
1200.000000	64
656.500000	63
2012.000000	61
1666.666667	58
643.000000	57
851.746667	56
1262.500000	51
1414.000000	50
1285.714286	50
3000.000000	48
1270.000000	47
...	
1568.866667	1
2406.666667	1
263.881538	1
340.123810	1
2390.714286	1
11000.000000	1
1663.333333	1
4090.500000	1
1652.522667	1
1119.284348	1
432.857143	1
1088.376000	1
2031.333333	1
2190.476191	1
1500.549600	1
1403.846154	1
42.083333	1
1236.541667	1
530.944444	1
2645.780000	1
1486.142857	1
3226.577333	1
1830.361111	1
1307.692308	1
506.250000	1
711.000000	1
1461.470000	1
1282.260870	1
1353.333333	1
1677.419355	1

Name: REVUC, Length: 2733, dtype: int64

*****AGE_2*****

1 61.0
2 82.0
4 60.0
5 14.0
6 8.0

Name: AGE_2, dtype: float64

50.0 241
52.0 214
45.0 210
47.0 204
54.0 194
48.0 188
65.0 184
70.0 184
42.0 177
44.0 175
60.0 175
51.0 175
58.0 173
53.0 171
43.0 169
46.0 168
49.0 167
68.0 165
35.0 163
40.0 161
56.0 158
55.0 157
63.0 149
71.0 148
64.0 148
41.0 146
59.0 145
67.0 144
66.0 142
62.0 141
...
12.0 25
80.0 23
1.0 23
11.0 20
81.0 18
5.0 18

83.0	16
6.0	16
7.0	16
8.0	15
82.0	15
3.0	15
84.0	12
2.0	12
4.0	11
85.0	9
86.0	7
87.0	7
92.0	5
90.0	4
91.0	4
88.0	3
93.0	3
98.0	2
89.0	2
0.0	2
97.0	1
100.0	1
94.0	1
95.0	1

Name: AGE_2, Length: 99, dtype: int64

*****AGE_3*****

9	23.0
16	29.0
18	3.0
20	27.0
23	41.0

Name: AGE_3, dtype: float64

16.0	193
17.0	178
18.0	178
15.0	167
8.0	144
13.0	143
14.0	142
12.0	139
6.0	134
11.0	132
1.0	132

19.0	125
9.0	124
10.0	122
21.0	121
22.0	121
20.0	120
7.0	114
5.0	107
4.0	101
3.0	94
23.0	89
2.0	88
25.0	75
24.0	73
50.0	72
46.0	65
47.0	61
45.0	59
51.0	58
...	
38.0	12
64.0	11
61.0	10
70.0	10
34.0	10
68.0	10
66.0	10
67.0	8
72.0	7
80.0	7
69.0	6
71.0	6
84.0	5
76.0	5
77.0	3
74.0	3
83.0	2
75.0	2
78.0	2
96.0	2
73.0	2
82.0	2
81.0	2
91.0	1
89.0	1
93.0	1
92.0	1
90.0	1

```
97.0      1
79.0      1
Name: AGE_3, Length: 92, dtype: int64
```

*****prescaf*****

```
10457     2.0
Name: prescaf, dtype: float64
```

```
2.0      1
Name: prescaf, dtype: int64
```

*****AGGLOINS*****

```
2045     0.0
2046     2.0
2047     1.0
2048     7.0
2049     0.0
Name: AGGLOINS, dtype: float64
```

```
7.0      2393
0.0      1783
8.0      1448
5.0       732
6.0       594
2.0       569
4.0       560
1.0       537
3.0       470
Name: AGGLOINS, dtype: int64
```

*****AGE_5*****

```
16      61.0
49      20.0
51      23.0
97      15.0
98      22.0
Name: AGE_5, dtype: float64
```

```
14.0     50
```

12.0	48
10.0	47
4.0	47
6.0	47
13.0	46
15.0	43
11.0	43
7.0	42
1.0	41
9.0	39
3.0	38
8.0	37
16.0	36
2.0	34
17.0	33
5.0	28
18.0	26
19.0	21
20.0	19
0.0	18
21.0	12
22.0	11
23.0	8
52.0	7
47.0	6
25.0	6
24.0	5
48.0	5
40.0	4
	..
54.0	4
42.0	3
55.0	3
46.0	3
26.0	3
28.0	3
56.0	2
50.0	2
60.0	2
57.0	2
43.0	2
33.0	2
59.0	2
45.0	2
27.0	2
53.0	2
49.0	2
68.0	1

64.0	1
51.0	1
63.0	1
76.0	1
85.0	1
39.0	1
75.0	1
38.0	1
30.0	1
66.0	1
67.0	1
61.0	1

Name: AGE_5, Length: 61, dtype: int64

*****NBUC*****

0	1.0
1	1.5
2	1.5
3	1.0
4	1.5

Name: NBUC, dtype: float64

1.5	4287
1.0	2219
2.0	1167
2.5	719
1.8	656
2.1	589
2.3	396
3.0	229
2.8	177
2.4	148
2.6	136
1.3	103
1.6	58
3.3	37
3.1	37
3.5	36
2.7	33
2.9	29
1.9	13
3.6	9
3.7	6
4.0	6
3.8	6

3.2	5
4.5	4
2.2	4
3.4	3
3.9	2
4.3	2
4.1	2
4.4	1
4.9	1
4.6	1
5.1	1

Name: NBUC, dtype: int64

*****cpt*****

8115	0.0
8116	0.0
8117	0.0
8118	0.0
8119	0.0

Name: cpt, dtype: float64

0.0	2643
1.0	300
2.0	63
4.0	6
3.0	4

Name: cpt, dtype: int64

*****REVAUTR*****

9	1500.0
10	1800.0
16	900.0
17	1400.0
26	3000.0

Name: REVAUTR, dtype: float64

0.0	372
999999.0	212
2000.0	114
1500.0	112
1200.0	90
1000.0	86

3000.0	82
2500.0	53
4000.0	51
1300.0	48
1600.0	46
1400.0	43
3500.0	41
1100.0	40
500.0	38
1800.0	31
5000.0	29
800.0	29
1700.0	27
2200.0	25
600.0	23
400.0	22
700.0	21
6000.0	21
900.0	20
4500.0	19
1900.0	15
2300.0	14
300.0	13
200.0	12
...	
44.0	1
55.0	1
120.0	1
30.0	1
1650.0	1
9500.0	1
930.0	1
2050.0	1
1370.0	1
1850.0	1
550.0	1
55000.0	1
5850.0	1
1025.0	1
2042.0	1
1360.0	1
1930.0	1
460.0	1
728.0	1
154.0	1
468.0	1
865.0	1
1860.0	1

7200.0	1
1660.0	1
567.0	1
4750.0	1
410.0	1
2900.0	1
520000.0	1

Name: REVAUTR, Length: 172, dtype: int64

*****poptrpeu*****

5095	0.083366
5096	0.048476
5097	0.007275
5098	0.000722
5099	0.000000

Name: poptrpeu, dtype: float64

0.000000	1439
0.000705	65
0.000147	60
1.000000	50
0.000073	41
0.000062	38
0.000485	33
0.000017	31
0.000548	29
0.001374	28
0.000098	26
0.001955	25
0.000230	24
0.000323	22
0.000071	22
0.001792	20
0.000306	19
0.000539	19
0.000248	19
0.000462	16
0.000352	15
0.000145	15
0.000747	15
0.000317	15
0.004513	14
0.002088	13
0.000809	12
0.002296	12

0.015703	12
0.001181	12
...	
0.020976	1
0.111860	1
0.085542	1
0.002788	1
0.026020	1
0.001653	1
0.020402	1
0.035002	1
1.000000	1
0.006308	1
0.126853	1
0.002591	1
0.035988	1
0.104737	1
0.116165	1
0.039002	1
0.080403	1
0.057339	1
0.003985	1
0.000295	1
0.011120	1
0.003506	1
0.023807	1
0.000420	1
0.112761	1
0.000368	1
0.398488	1
0.039183	1
0.000392	1
0.006895	1

Name: poptrpeu, Length: 2502, dtype: int64

*****SENSIENV*****

0	7.0
1	6.0
2	4.0
3	6.0
4	5.0

Name: SENSIENV, dtype: float64

5.0	3586
6.0	2925

```
4.0    2035
7.0    1649
3.0     484
2.0     207
1.0     137
Name: SENSIENV, dtype: int64
```

*****CSP6*****

```
8115    8.0
8116    8.0
8117    8.0
8118    8.0
8119    8.0
Name: CSP6, dtype: float64
```

```
8.0    782
4.0    517
7.0    428
3.0    418
5.0    377
2.0    277
1.0    119
6.0     98
Name: CSP6, dtype: int64
```

*****agedip*****

```
8115    22.0
8116    31.0
8117    32.0
8118    31.0
8119    32.0
Name: agedip, dtype: float64
```

```
22.0    968
21.0    506
12.0    490
32.0    467
31.0    410
11.0    175
Name: agedip, dtype: int64
```

*****age_OW*****

8115	4.0
8116	5.0
8117	6.0
8118	5.0
8119	5.0

Name: age_OW, dtype: float64

3.0	978
4.0	496
6.0	461
2.0	449
5.0	416
1.0	216

Name: age_OW, dtype: int64

*****refus2*****

8115	0.0
8116	0.0
8117	0.0
8118	0.0
8119	1.0

Name: refus2, dtype: float64

0.0	2648
1.0	368

Name: refus2, dtype: int64

*****i*****

8115	6.0
8116	6.0
8117	6.0
8118	6.0
8119	6.0

Name: i, dtype: float64

6.0	3016
-----	------

Name: i, dtype: int64

*****CP*****

8115	87260.0
8116	24480.0
8117	71640.0
8118	76620.0
8119	41300.0

Name: CP, dtype: float64

75020.0	15
33000.0	14
67000.0	14
75011.0	13
44000.0	13
83000.0	12
75015.0	12
75014.0	12
75012.0	12
66000.0	12
62100.0	11
75019.0	11
29200.0	11
76000.0	11
6000.0	10
31200.0	9
75013.0	9
31000.0	9
69003.0	9
67700.0	9
69007.0	9
54000.0	8
37000.0	8
49000.0	8
59200.0	8
21000.0	8
75010.0	7
92300.0	7
35000.0	7
75016.0	7
	..
82100.0	1
22310.0	1
94600.0	1
60520.0	1
93700.0	1
2420.0	1
18600.0	1


```

84500.0    1
70700.0    1
56480.0    1
2830.0     1
89520.0    1
94800.0    1
66510.0    1
33260.0    1
57550.0    1
11600.0    1
97233.0    1
50440.0    1
93140.0    1
56230.0    1
20215.0    1
62224.0    1
54740.0    1
27160.0    1
14460.0    1
29280.0    1
77420.0    1
44530.0    1
57905.0    1
Name: CP, Length: 1745, dtype: int64

```

```
In [135]: cdv.loc[:,['age_OW', 'ANNEEFUZ']].dropna().head()
```

```

Out[135]:      age_OW  ANNEEFUZ
8115      4.0      2018
8116      5.0      2018
8117      6.0      2018
8118      5.0      2018
8119      5.0      2018

```

```

In [136]: l = list(cdv.columns)
          ACM_var = [f for f in l if f[0:3] == "ACM"]
          ACM_var

```

```

Out[136]: ['ACM1',
          'ACM2',
          'ACM3',
          'ACM4',
          'ACM5',
          'ACM6',
          'ACM7',
          'ACM8',

```

```
'ACM9',
'ACM10',
'ACM11',
'ACM12']
```

1.4 II - Communal data

```
In [137]: # loading MergeCommunesEnvi data
file = path_data / Path("MergeCommunesEnvi.csv")
with Path.open(file, 'rb') as fp:
    MergeCommunesEnvi = pd.read_csv(fp, encoding='cp1252', low_memory=False, sep=';',
```

```
In [138]: # loading cdv data
file = path_data / Path("felix.csv")
with Path.open(file, 'rb') as fp:
    cdv = pd.read_csv(fp, encoding='cp1252', low_memory=False, index_col = 0)
```

```
In [139]: commune_var = set(MergeCommunesEnvi.columns) - set(cdv.columns)
print(f"{len(commune_var)} additional features")
```

218 additional features

```
In [140]: df = MergeCommunesEnvi.loc[:, commune_var]
np.sum(df.isnull()).sort_values(ascending = False)
```

```
Out[140]: TP6015                2830
PIMP15                2505
Superficie.protection.forte...2017..ha.    840
Part.protection.forte...2017...    840
Part.protection.contractuelle...2017...    840
Superficie.protection.contractuelle...2017..ha.    840
Superficie.forêts.et.milieux.semi.naturels...2012..ha.    358
Superficie.zones.humides.et.surfaces.en.eau...2012..ha.    358
Part.zones.humides.et.surfaces.en.eau...2012...    358
communes                358
Part.forêts.et.milieux.semi.naturels...2012...    358
NBMENFISC15            212
MED15                  212
NB_D106                164
P15_EMPLT              164
NB_F119_NB_COU         164
NB_D303                164
NB_D104                164
NB_D302                164
NB_F102                164
NB_F112                164
NB_E106                164
NB_F101_NB_AIREJEU     164
```

NB_D403	164
NB_F114_NB_AIREJEU	164
P15_EMPLT_SAL	164
NB_D606	164
NB_A406	164
NB_D301	164
NB_F116	164
...	
NB_F120_NB_COU	164
NB_F109	164
NB_D109	164
NB_F103_NB_ECL	164
NB_F121_NB_COU	164
NB_F103_NB_COU	164
NB_F104_NB_ECL	164
NB_F112_NB_ECL	164
NB_F105_NB_ECL	164
NB_F104_NB_COU	164
NB_F104_NB_AIREJEU	164
NB_F118_NB_AIREJEU	164
NB_F203	164
NB_F110_NB_AIREJEU	164
P15_RP_PROP	164
NB_F114	164
NB_D103	164
P10_EMPLT	59
ETTOT15	58
LIBGEO	58
ETBE15	58
ETGU15	58
ETTEFP1015	58
ETAZ15	58
REG	58
ETOQ15	58
DEP	58
ETFZ15	58
ETTEF115	58
ETGZ15	58
Length: 218, dtype: int64	

```
In [141]: df.dtypes.unique()
```

```
Out[141]: array([dtype('float64'), dtype('O')], dtype=object)
```

```
In [142]: df.select_dtypes(include=['O']).columns
```

```
Out[142]: Index(['LIBGEO', 'communes', 'DEP'], dtype='object')
```

1.5 III - Feature sets and engineering - Dataset preparation

```
In [143]: # loading cdv data
file = path_data / Path("felix.csv")
with Path.open(file, 'rb') as fp:
    cdv = pd.read_csv(fp, encoding='cp1252', low_memory=False, index_col = 0)

In [144]: # loading cdv data without format
file = path_data / Path("felix_ssfmt.csv")
with Path.open(file, 'rb') as fp:
    cdv_ssfmt = pd.read_csv(fp, encoding='cp1252', low_memory=False, index_col = 0)
```

1.5.1 Feature scope

```
In [145]: # number of line per year in teh dataset
n_per_year = cdv["ANNEEFUZ"].value_counts()
# number of missing value per variable for a given year
na_2015 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2015].isnull())
na_2016 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2016].isnull())
na_2017 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2017].isnull())
na_2018 = np.sum(cdv.loc[cdv["ANNEEFUZ"] == 2018].isnull())
# column scope per year
cdv_2015_var = set(na_2015[na_2015 < n_per_year[2015]].index)
cdv_2016_var = set(na_2016[na_2016 < n_per_year[2016]].index)
cdv_2017_var = set(na_2017[na_2017 < n_per_year[2017]].index)
cdv_2018_var = set(na_2018[na_2018 < n_per_year[2018]].index)

In [146]: cdv_2015_2018_var = (cdv_2015_var & cdv_2016_var & cdv_2017_var & cdv_2018_var)
cdv_2016_2018_var = (cdv_2016_var & cdv_2017_var & cdv_2018_var)
cdv_2017_2018_var = (cdv_2017_var & cdv_2018_var)

In [147]: print(f"{len(cdv_2015_2018_var)} variables common to all study out of {cdv_ssfmt.shape[0]}")
267 variables common to all study out of 353
```

1.5.2 Special variables

```
In [148]: pred_var = {"HEUREUX"}
tech_var = {"ANNEEFUZ", "ANNEFUZ2", "COLLECTE", "CHAMP",
            "identifiant", "an_enq", "INTER"}
com_var = {"COMINSEE", 'DEPCOM', 'com', 'insee1', 'inseenum', 'CP'}
text_var = {"RADIQUOI"}
bizz_var = {'NB0003', 'NB0306', 'NB0610', 'NB1016', 'NB1620', 'NB2099',
            'an_nais', 'decuc', 'decsqt', 'info', 'typodeg', 'refus2',
            'cpt', 'prescaf', 'poptrepu', 'REVUC', 'i', 'REVTOT',
            'poppeud', 'popdense', 'popinter', 'pmun', 'agedip', 'age_OW',
            'REVsq', 'NBUC', 'AGGLOINS', 'med', 'CSP6', 'REVTOT6',
            'ACM1', 'ACM2', 'ACM3', 'ACM4', 'ACM5', 'ACM6', 'ACM7',
            'ACM8', 'ACM9', 'ACM10', 'ACM11', 'ACM12'}
```

1.5.3 Categorical variable

```
In [149]: obj_cdv = cdv.select_dtypes(include=['object'])
obj_var = set(obj_cdv.columns)
cat_max9_var = set()
cat_min10_var = set()
for c in obj_var:
    obj_cdv_valcpt = obj_cdv[c].value_counts()
    if len(obj_cdv_valcpt) > 10:
        cat_min10_var.add(c)
    else:
        cat_max9_var.add(c)

In [150]: ord_var = {
    "CONFPOLI", "AGE5", "SECURITE", "ACM7", "INQGUERR", "NBPiece6", "INNOVTEC",
    "JUSTICE", "EFFORTPP", "ACM10", "NIVFRAN4", "NBPERS5", "INQCHOMA", "CDV5_4",
    "CONFGOUV", "ADOPTGAY", "ACM8", "FREQCINE", "CONFPUB", "FREQSPOR", "INQALIM",
    "ASSO10_3", "FREQBIBL", "DEPLOG", "NBCHOM", "CONFENTR", "ORDLIB", "ACM5",
    "INQMALAD", "FREQTELE", "NBENF6", "ACM9", "revtot7", "INROUTE", "NIVPERS4",
    "ETATSAN", "INQNUCLE", "NIVPERSO", "CONFASSO", "ACM6", "CDV5", "UNIONGAY",
    "ACM4", "INQAGRES", "CADVIE", "NIVFRAN", "REV_TR7", "ISEGO", "RECEP", "AGE6",
    "ADNCB", "PRATCOLL", "NBHEUR39", "HARVEY", "QUOTAAGE", "NBHEUR35", "RELEG",
    "CONFKEUF", "CONFECOL", "ADNSTIC", "ADNORDI", "CONFPRES", "CONFWEB", "CONFBANK"
}

In [151]: # exclusion of features with order
cat_var = obj_var - ord_var
cat_max9_var = cat_max9_var - ord_var
cat_min10_var = cat_min10_var - ord_var

In [152]: cdv_dtypes = cdv.dtypes

In [153]: int_var = set(cdv_dtypes[cdv_dtypes == 'int64'].index)

In [154]: int_cat_var = {
    'NB0003', 'NB0306', 'NB0610', 'NB1016',
    'NB1620', 'NB2099', 'REVTOT6', 'ANNEEFUZ', 'INTER', 'INTER6'
}
int_cat_max9_var = {
    'NB0003', 'NB0306', 'NB0610', 'NB1016',
    'NB1620', 'NB2099', 'REVTOT6', 'ANNEEFUZ'
}
int_cat_min10_var = {
    'INTER', 'INTER6'
}
int_quant_var = {
    'AGE', 'REVENQ', 'AUTREREV',
    'an_enq', 'an_nais'
}
```

```

In [155]: cat_var = cat_var | int_cat_var
          cat_max9_var = cat_max9_var | int_cat_max9_var
          cat_min10_var = cat_min10_var | int_cat_min10_var
          quant_var = ord_var | int_quant_var

In [156]: float_var = set(cdv_dtypes[cdv_dtypes == 'float64'].index)

In [157]: float_cat_min10_var = {'CP', 'inseenum'}
          float_cat_max9_var = {'refus2', 'cpt', 'prescaf', 'i', 'age_OW', 'TYPLOG', 'AGGLOINS', 'CSP6'}
          float_cat_var = float_cat_min10_var | float_cat_max9_var
          float_quant_var = float_var - float_cat_var

In [158]: cat_var = cat_var | float_cat_var
          cat_max9_var = cat_max9_var | float_cat_max9_var
          cat_min10_var = cat_min10_var | float_cat_min10_var
          quant_var = quant_var | float_quant_var

In [159]: print(f"out of the {cdv.shape[1]} variable :")
          print(f"{len(cat_var)} variables are categorical ")
          print(f"{len(quant_var)} variables are quantitative ")

out of the 353 variable :
248 variables are categorical
106 variables are quantitative

In [160]: print(f"out of the {len(cat_var)} variable categorical:")
          print(f"{len(cat_max9_var)} variables have maximum 9 modalities ")
          print(f"{len(cat_min10_var)} variables have more ")

out of the 248 variable categorical:
221 variables have maximum 9 modalities
27 variables have more

In [161]: dict_var_groups = {
          'cdv_2015_var' : cdv_2015_var,
          'cdv_2016_var' : cdv_2016_var,
          'cdv_2017_var' : cdv_2017_var,
          'cdv_2018_var' : cdv_2018_var,
          'cdv_2015_2018_var' : cdv_2015_2018_var,
          'cdv_2016_2018_var' : cdv_2016_2018_var,
          'cdv_2017_2018_var' : cdv_2017_2018_var,
          'pred_var' : pred_var,
          'tech_var' : tech_var,
          'com_var' : com_var,
          'text_var' : text_var,
          'bizz_var' : bizz_var,
          'cat_var' : cat_var,

```

```

        'cat_max9_var' : cat_max9_var,
        'cat_min10_var' : cat_min10_var,
        'quant_var' : quant_var
    }

```

1.5.4 Adding communal features and levers

```

In [162]: # loading MergeCommunesEnvi data
file = path_data / Path("MergeCommunesEnvi.csv")
with Path.open(file, 'rb') as fp:
    MergeCommunesEnvi = pd.read_csv(fp, encoding='cp1252', low_memory=False, sep=';',

In [163]: # file 'List-of-Actionable-Variables_v0.1_sp' september 01
indiv_act_var = {
    "LIMVIAND", "VACANCES", "VISITFAM", "RECEP", "YOGA", "FREQSPOR", "FREQBIBL", "FREQCINE",
    "FREQTELE", "ASSOSPOR", "ASSOCULT", "ASSOCONF", "ASSOJEUN", "ASSOSYND", "ASSOENVI",
    "ASSOPARE", "ASSOCONS", "ASSOPOLI", "ASSOHUMA", "ASSOAUTR", "NOT_FAMI", "NOT_PROF",
    "NOT_AMIS", "NOT_COHE", "NOT_POLI", "NOT_LIBR", "NOT_LOG", "NOT_CAD", "RELIGION"
}

In [164]: # file 'List-of-Actionable-Variables_v0.1_sp' september 01
indiv_semi_act_var = {
    "SITUEMP5", "SITUEMP6", "TEMPSTRA", "nbheures", "NBHEUR39", "NBHEUR35",
    "IMAGTRAV", "COUPLE", "ENFANTS", "CADVIE", "CADVIE3", "MODCHAUF", "ETATSAN",
    "BANQMOB", "BANQEPA", "BANQVIE", "TELMOB", "CONFPUB", "CONFENTR", "CONFASSO",
    "CONFPOLI", "CONFBANK", "CONFPRES", "CONFECOL", "CONFKEUF", "INQMALAD",
    "INQMALA3", "INQAGRES", "INQAGRE3", "INQROUTE", "INQROUT3", "INQCHOMA",
    "INQCHOM3", "INQGUERR", "INQGUER3", "INQNUCLE", "INQNUCL3", "INQALIM",
    "INQALIM3", "EHPOL"
}

In [165]: admin_act_var = {
    "AIDESUFF", "EFFORTPP", "CHOAVANT", "OPIRSA", "JUSTICE", "RELEG", "RADIQUOI",
    "RADWHY1", "RADWHY2", "RADWHY3", "RADWHY4", "RADWHY5", "RADWHY6", "RADWHY7",
    "RADWHY8", "RADWHY9", "RADWHY10", "RADWHY11", "RADWHY12", "RADWHY13", "RADWHY14",
    "ORDLIB", "PREOCCU1", "PREOCCU2", "CONFGOUV"
}

In [166]: admin_semi_act_var = {
    "SECURITE", "SECUR3", "ADNSTIC", "ADNCB", "ADNORDI", "ROBOT1", "ROBOT2", "ROBOT3",
    "PRESTCAF", "REVPPF", "CONFPUB", "CONFENTR", "CONFASSO", "CONFPOLI", "CONFBANK",
    "CONFPRES", "CONFECOL", "CONFKEUF", "TRANSFST", "TRANSFO5", "PROGRAD", "OPIIMMIG"
}

In [167]: commune_var = set(MergeCommunesEnvi.columns) - set(cdv.columns)

In [168]: dict_var_groups["cat_min10_var"] = dict_var_groups["cat_min10_var"] | {'DEP', 'LIBGEO',
                                          'communes'}
dict_var_groups["cat_var"] = dict_var_groups["cat_var"] | {'DEP', 'LIBGEO', 'communes'}

```

```

In [169]: dict_var_groups["commune_var"] = commune_var

In [170]: df = MergeCommunesEnvi.loc[:,commune_var]
commune_quant_var = set(df.select_dtypes(include=['float64']).columns)
dict_var_groups["quant_var"] = dict_var_groups["quant_var"] | commune_quant_var

In [171]: cdv_2015_var = dict_var_groups['cdv_2015_var']
cdv_2016_var = dict_var_groups['cdv_2016_var']
cdv_2017_var = dict_var_groups['cdv_2017_var']
cdv_2018_var = dict_var_groups['cdv_2018_var']
cdv_2015_2018_var = dict_var_groups['cdv_2015_2018_var']
cdv_2016_2018_var = dict_var_groups['cdv_2016_2018_var']
cdv_2017_2018_var = dict_var_groups['cdv_2017_2018_var']

scope_2015_var = cdv_2015_var | commune_var
scope_2016_var = cdv_2016_var | commune_var
scope_2017_var = cdv_2017_var | commune_var
scope_2018_var = cdv_2018_var | commune_var
scope_2015_2018_var = cdv_2015_2018_var | commune_var
scope_2016_2018_var = cdv_2016_2018_var | commune_var
scope_2017_2018_var = cdv_2017_2018_var | commune_var

dict_var_groups["scope_2015_var"] = scope_2015_var
dict_var_groups["scope_2016_var"] = scope_2016_var
dict_var_groups["scope_2017_var"] = scope_2017_var
dict_var_groups["scope_2018_var"] = scope_2018_var
dict_var_groups["scope_2015_2018_var"] = scope_2015_2018_var
dict_var_groups["scope_2016_2018_var"] = scope_2016_2018_var
dict_var_groups["scope_2017_2018_var"] = scope_2017_2018_var

In [172]: dict_var_groups["indiv_semi_act_var"] = indiv_semi_act_var
dict_var_groups["indiv_act_var"] = indiv_act_var
dict_var_groups["admin_semi_act_var"] = admin_semi_act_var
dict_var_groups["admin_act_var"] = admin_act_var

In [173]: com_var = dict_var_groups['com_var']
tech_var = dict_var_groups['tech_var']
text_var = dict_var_groups['text_var']
bizz_var = dict_var_groups['bizz_var']

cat_max9_var = dict_var_groups['cat_max9_var']
quant_var = dict_var_groups['quant_var']

exclusion = com_var | tech_var | bizz_var | text_var
# suppression of quantitative variables with more than 200 NaN
quant_null = np.sum(MergeCommunesEnvi.loc[:,quant_var].isnull())
quant_var_kept = set(quant_null[quant_null < 200].index)

usual_common_scope = ((cat_max9_var | quant_var_kept) & scope_2015_2018_var) - exclusion

```



```
dict_var_groups["exclusion"] = exclusion
dict_var_groups["usual_common_scope"] = usual_common_scope
```

```
In [174]: filename = path_dump / Path("dict_var_groups.sav")
with open(filename, 'wb') as fp:
    pickle.dump(dict_var_groups,fp,pickle.HIGHEST_PROTOCOL)
```

Dataset

```
In [175]: df = MergeCommunesEnvi.loc[:,:]
df.loc[:,cdv_ssfmt.columns] = cdv_ssfmt.loc[:,:]
df = df.loc[:,usual_common_scope]
df.loc[:,(cat_var & usual_common_scope) - {"HEUREUX"}] = cdv.loc[:,(cat_var & usual_co
```

```
dataset = pd.get_dummies(
    df,
    columns=(cat_var & usual_common_scope) - {"HEUREUX"},
    dummy_na = True,
    drop_first=1
)
```

```
print(f"{dataset.shape[1]} columns after encoding of {len((cat_var & usual_common_scope) - {"HEUREUX"})}
categorical variables in {len((cat_var & usual_common_scope))-1+dataset.shape[1]-df.shape[1]}
binary variables (K-1 one hot encoding)")
```

810 columns after encoding of 150 categorical variables in 553 binary variables (K-1 one hot encoding)

```
In [176]: # saving dataset data
file = path_data / Path("dataset.csv")
with Path.open(file, 'w') as fp:
    dataset.to_csv(fp, encoding='utf-8')
```

Usefull feature sets

```
In [177]: dict_features_sets = dict()
```

```
In [178]: usual_common_features = set(dataset.columns)
dict_features_sets['usual_common_features'] = usual_common_features
```

```
In [179]: df = MergeCommunesEnvi.loc[:,:]
df.loc[:,cdv_ssfmt.columns] = cdv_ssfmt.loc[:,:]
scope = indiv_act_var & usual_common_scope
df = df.loc[:,scope]
df.loc[:,(cat_var & scope) - {"HEUREUX"}] = cdv.loc[:,(cat_var & scope) - {"HEUREUX"}]

df_dummies = pd.get_dummies(
```

```

        df,
        columns=(cat_var & scope),
        dummy_na = True,
        drop_first=1
    )

    print(f"{df_dummies.shape[1]} columns after encoding of {len((cat_var & scope))} categorical variables in {len((cat_var & scope))+df_dummies.shape[1]-df.shape[1]} binary variables (K-1 one hot encoding)")

    indiv_act_features = set(df_dummies.columns)
    dict_features_sets['indiv_act_features'] = indiv_act_features

```

50 columns after encoding of 13 categorical variables in 40 binary variables (K-1 one hot encoding)

```

In [180]: df = MergeCommunesEnvi.loc[:,:]
df.loc[:,cdv_ssfmt.columns] = cdv_ssfmt.loc[:,:]
scope = indiv_semi_act_var & usual_common_scope
df = df.loc[:,scope]
df.loc[:,(cat_var & scope) - {"HEUREUX"}] = cdv.loc[:,(cat_var & scope) - {"HEUREUX"}]

df_dummies = pd.get_dummies(
    df,
    columns=(cat_var & scope),
    dummy_na = True,
    drop_first=1
)

    print(f"{df_dummies.shape[1]} columns after encoding of {len((cat_var & scope))} categorical variables in {len((cat_var & scope))+df_dummies.shape[1]-df.shape[1]} binary variables (K-1 one hot encoding)")

    indiv_semi_act_features = set(df_dummies.columns)
    dict_features_sets['indiv_semi_act_features'] = indiv_semi_act_features

```

73 columns after encoding of 17 categorical variables in 60 binary variables (K-1 one hot encoding)

```

In [181]: df = MergeCommunesEnvi.loc[:,:]
df.loc[:,cdv_ssfmt.columns] = cdv_ssfmt.loc[:,:]
scope = admin_act_var & usual_common_scope
df = df.loc[:,scope]
df.loc[:,(cat_var & scope) - {"HEUREUX"}] = cdv.loc[:,(cat_var & scope) - {"HEUREUX"}]

df_dummies = pd.get_dummies(
    df,
    columns=(cat_var & scope),
    dummy_na = True,

```

```

        drop_first=1
    )

    print(f"{df_dummies.shape[1]} columns after encoding of {len((cat_var & scope))} categor
variables in {len((cat_var & scope))+df_dummies.shape[1]-df.shape[1]} binary variables
(K-1 one hot encoding)")

    admin_act_features = set(df_dummies.columns)
    dict_features_sets['admin_act_features'] = admin_act_features

13 columns after encoding of 3 categorial variables in 9 binary variables (K-1 one hot encoding)

In [182]: df = MergeCommunesEnvi.loc[:,:]
df.loc[:,cdv_ssfmt.columns] = cdv_ssfmt.loc[:,:]
scope = admin_semi_act_var & usual_common_scope
df = df.loc[:,scope]
df.loc[:,(cat_var & scope) - {"HEUREUX"}] = cdv.loc[:,(cat_var & scope) - {"HEUREUX"}]

df_dummies = pd.get_dummies(
    df,
    columns=(cat_var & scope),
    dummy_na = True,
    drop_first=1
)

    print(f"{df_dummies.shape[1]} columns after encoding of {len((cat_var & scope))} categor
variables in {len((cat_var & scope))+df_dummies.shape[1]-df.shape[1]} binary variables
(K-1 one hot encoding)")

    admin_semi_act_features = set(df_dummies.columns)
    dict_features_sets['admin_semi_act_features'] = admin_semi_act_features

25 columns after encoding of 6 categorial variables in 20 binary variables (K-1 one hot encoding)

In [183]: filename = path_dump / Path("dict_features_sets.sav")
    with open(filename, 'wb') as fp:
        pickle.dump(dict_features_sets,fp,pickle.HIGHEST_PROTOCOL)

```

Feature selection and and results recording

```

In [184]: # reducing problem to a 2 class classification problem
df = dataset.loc[:,:]
df["HEUREUX_CLF"] = 0
df.loc[df["HEUREUX"]==4, "HEUREUX_CLF"] = 1
df.loc[df["HEUREUX"]==3, "HEUREUX_CLF"] = 1
df.loc[df["HEUREUX"]==5, "HEUREUX_CLF"] = None

```

```

# treating remaining missing values
features = set(df.columns.drop(['HEUREUX', 'HEUREUX_CLF']))
df = df.loc[:,features | {"HEUREUX_CLF"}].dropna()

```

```

In [185]: X = df.loc[:,features]
          y = df["HEUREUX_CLF"]

```

```

X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.2,
                                                    random_state=42
                                                    )

```

```

scaler = StandardScaler().fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)

```

```

print(f"Number exemple: {y.shape[0]}\n- training set: \
{y_train.shape[0]}\n- test set: {y_test.shape[0]}")
print(f"Number of features: p={X_train.shape[1]}")
print(f"Number of class: {len(np.unique(y))}")
for c in np.unique(y):
    print(f"class {c:0.0f} : {100*np.sum(y==c)/len(y):0.1f}%")

```

```

Number exemple: 10445
- training set: 8356
- test set: 2089
Number of features: p=809
Number of class: 2
class 0 : 35.1%
class 1 : 64.9%

```

```

In [186]: startTime = time.time()
          n_features_to_select = 20
          step = 0.05
          clf = LogisticRegression(C=1,
                                   penalty='l1',
                                   class_weight='balanced',
                                   random_state=42)
          selector = RFE(estimator=clf, n_features_to_select=n_features_to_select, step=step)
          selector.fit(X_train, y_train)
          print(f"Optimal support of size {n_features_to_select} found in {time.time() - startTi

```

```

Optimal support of size 20 found in 745.6 s

```

```

In [187]: lasso_20_features = X.loc[:,selector.support_].columns
          lasso_20_features = set(lasso_20_features)
          dict_features_sets['lasso_20_features'] = lasso_20_features

```

```
In [188]: startTime = time.time()
n_features_to_select = 50
step = 0.05
clf = LogisticRegression(C=1,
                        penalty='l1',
                        class_weight='balanced',
                        random_state=42)
selector = RFE(estimator=clf, n_features_to_select=n_features_to_select, step=step)
selector.fit(X_train, y_train)
print(f"Optimal support of size {n_features_to_select} found in {time.time() - startTi
```

Optimal support of size 50 found in 715.1 s

```
In [189]: lasso_50_features = X.loc[:,selector.support_].columns
lasso_50_features = set(lasso_50_features)
dict_features_sets['lasso_50_features'] = lasso_50_features
```

```
In [191]: startTime = time.time()
n_features_to_select = 100
step = 0.05
clf = LogisticRegression(C=1,
                        penalty='l1',
                        class_weight='balanced',
                        random_state=42)
selector = RFE(estimator=clf, n_features_to_select=n_features_to_select, step=step)
selector.fit(X_train, y_train)
print(f"Optimal support of size {n_features_to_select} found in {time.time() - startTi
```

Optimal support of size 100 found in 757.5 s

```
In [192]: lasso_100_features = X.loc[:,selector.support_].columns
lasso_100_features = set(lasso_100_features)
dict_features_sets['lasso_100_features'] = lasso_100_features
```

```
In [193]: startTime = time.time()
n_features_to_select = 10
step = 0.05
clf = LogisticRegression(C=1,
                        penalty='l1',
                        class_weight='balanced',
                        random_state=42)
selector = RFE(estimator=clf, n_features_to_select=n_features_to_select, step=step)
selector.fit(X_train, y_train)
print(f"Optimal support of size {n_features_to_select} found in {time.time() - startTi
```

Optimal support of size 10 found in 754.7 s

```
In [194]: lasso_10_features = X.loc[:,selector.support_].columns
lasso_10_features = set(lasso_10_features)
dict_features_sets['lasso_10_features'] = lasso_10_features

In [195]: filename = path_dump / Path("dict_features_sets.sav")
with open(filename, 'wb') as fp:
    pickle.dump(dict_features_sets,fp,pickle.HIGHEST_PROTOCOL)
```