

PROJET 4 DATA ANALYST

Réalisez une étude de santé publique avec R ou Python

PARTIE NETTOYAGE DES DONNEES

```
In [2]: import pandas as pd
```

1 - Analyse exploratoire du fichier aide alimentaire

```
In [3]: aide_alimentaire=pd.read_csv('aide_alimentaire.csv')
aide_alimentaire.head()
```

```
Out[3]:
```

	Pays bénéficiaire	Année	Produit	Valeur
0	Afghanistan	2013	Autres non-céréales	682
1	Afghanistan	2014	Autres non-céréales	335
2	Afghanistan	2013	Blé et Farin	39224
3	Afghanistan	2014	Blé et Farin	15160
4	Afghanistan	2013	Céréales	40504

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1475 entries, 0 to 1474
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Pays bénéficiaire     1475 non-null  object
1   Année                 1475 non-null  int64
2   Produit               1475 non-null  object
3   Valeur                1475 non-null  int64
dtypes: int64(2), object(2)
memory usage: 46.2+ KB
```

```
In [5]: #changement du nom de la colonne Pays bénéficiaire par Zone
aide_alimentaire.rename(columns={'Pays bénéficiaire':'Zone','Valeur': 'Aide_alimentaire'})
```

```
In [6]: #Multiplication de la colonne Aide_alimentaire qui contient des tonnes par 1000 pour a
aide_alimentaire['Aide_alimentaire']=aide_alimentaire['Aide_alimentaire']*1000
aide_alimentaire.head()
```

Out[6]:	Zone	Année	Produit	Aide_alimentaire
0	Afghanistan	2013	Autres non-céréales	682000
1	Afghanistan	2014	Autres non-céréales	335000
2	Afghanistan	2013	Blé et Farin	39224000
3	Afghanistan	2014	Blé et Farin	15160000
4	Afghanistan	2013	Céréales	40504000

```
In [7]: sorted(aide_alimentaire['Zone'].unique())
```

```
Out[7]: ['Afghanistan',
        'Algérie',
        'Angola',
        'Bangladesh',
        'Bhoutan',
        'Bolivie (État plurinational de)',
        'Burkina Faso',
        'Burundi',
        'Bénin',
        'Cambodge',
        'Cameroun',
        'Chine, continentale',
        'Colombie',
        'Comores',
        'Congo',
        'Cuba',
        "Côte d'Ivoire",
        'Djibouti',
        'El Salvador',
        'Gambie',
        'Ghana',
        'Guatemala',
        'Guinée',
        'Guinée-Bissau',
        'Géorgie',
        'Haïti',
        'Honduras',
        "Iran (République islamique d')",
        'Iraq',
        'Jordanie',
        'Kenya',
        'Kirghizistan',
        'Lesotho',
        'Liban',
        'Libye',
        'Libéria',
        'Madagascar',
        'Malawi',
        'Mali',
        'Mauritanie',
        'Mozambique',
        'Myanmar',
        'Nicaragua',
        'Niger',
        'Népal',
        'Ouganda',
        'Pakistan',
        'Palestine',
        'Philippines',
        'Rwanda',
        'République arabe syrienne',
```

```

'République centrafricaine',
'République dominicaine',
'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Sao Tomé-et-Principe',
'Sierra Leone',
'Somalie',
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Swaziland',
'Sénégal',
'Tadjikistan',
'Tchad',
'Timor-Leste',
'Togo',
'Vanuatu',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Équateur',
'Éthiopie']

```

```

In [8]: liste=['Barbade','Bermudes','Belize','Timor-Leste']
aide_alimentaire.drop(aide_alimentaire[aide_alimentaire['Zone'].isin(liste)].index, inplace=True)
sorted(aide_alimentaire['Zone'].unique())

```

```

Out[8]: ['Afghanistan',
'Algérie',
'Angola',
'Bangladesh',
'Bhoutan',
'Bolivie (État plurinational de)',
'Burkina Faso',
'Burundi',
'Bénin',
'Cambodge',
'Cameroun',
'Chine, continentale',
'Colombie',
'Comores',
'Congo',
'Cuba',
"Côte d'Ivoire",
'Djibouti',
'El Salvador',
'Gambie',
'Ghana',
'Guatemala',
'Guinée',
'Guinée-Bissau',
'Géorgie',
'Haïti',
'Honduras',
'Iran (République islamique d')',
'Iraq',
'Jordanie',
'Kenya',
'Kirghizistan',
'Lesotho',
'Liban',
'Libye',
'Libéria',

```

```

'Madagascar',
'Malawi',
'Mali',
'Mauritanie',
'Mozambique',
'Myanmar',
'Nicaragua',
'Niger',
'Népal',
'Ouganda',
'Pakistan',
'Palestine',
'Philippines',
'Rwanda',
'République arabe syrienne',
'République centrafricaine',
'République dominicaine',
'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Sao Tomé-et-Principe',
'Sierra Leone',
'Somalie',
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Swaziland',
'Sénégal',
'Tadjikistan',
'Tchad',
'Togo',
'Vanuatu',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Équateur',
'Éthiopie']

```

```
In [9]: aide_alimentaire.to_csv('aide_alimentaire_clean.csv', index=False)
```

2 - Analyse exploratoire du fichier disponibilité alimentaire

</div>

```
In [10]: dispo_alimentaire = pd.read_csv('dispo_alimentaire.csv')
dispo_alimentaire.head()
```

Out[10]:

	Zone	Produit	Origine	Aliments pour animaux	Autres Utilisations	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité alimentaire en quantité (kg/personne/an)	Dispon matière c
0	Afghanistan	Abats Comestible	animale	NaN	NaN	5.0	1.72	
1	Afghanistan	Agrumes, Autres	vegetale	NaN	NaN	1.0	1.29	
2	Afghanistan	Aliments	vegetale	NaN	NaN	1.0	0.06	

			pour enfants				
3	Afghanistan	Ananas	vegetale	NaN	NaN	0.0	0.00
4	Afghanistan	Bananes	vegetale	NaN	NaN	4.0	2.70

```
In [11]: #Afficher les dimensions du dataset
print("Le tableau comporte {} observation(s) ou article(s)".format(dispo_alimentaire
print("Le tableau comporte {} colonne(s)".format(dispo_alimentaire.shape[1]))
```

Le tableau comporte 15605 observation(s) ou article(s)
Le tableau comporte 18 colonne(s)

```
In [12]: dispo_alimentaire.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15605 entries, 0 to 15604
Data columns (total 18 columns):
 #   Column                                Non-Null Count
Dtype
---  ---
0    Zone                                15605 non-null
object
1    Produit                             15605 non-null
object
2    Origine                             15605 non-null
object
3    Aliments pour animaux               2720 non-null
float64
4    Autres Utilisations                 5496 non-null
float64
5    Disponibilité alimentaire (Kcal/personne/jour) 14241 non-null
float64
6    Disponibilité alimentaire en quantité (kg/personne/an) 14015 non-null
float64
7    Disponibilité de matière grasse en quantité (g/personne/jour) 11794 non-null
float64
8    Disponibilité de protéines en quantité (g/personne/jour) 11561 non-null
float64
9    Disponibilité intérieure             15382 non-null
float64
10   Exportations - Quantité              12226 non-null
float64
11   Importations - Quantité              14852 non-null
float64
12   Nourriture                           14015 non-null
float64
13   Pertes                               4278 non-null
float64
14   Production                           9180 non-null
float64
15   Semences                             2091 non-null
float64
16   Traitement                           2292 non-null
float64
17   Variation de stock                   6776 non-null
float64
dtypes: float64(15), object(3)
memory usage: 2.1+ MB
```

```
In [13]: dispo_alimentaire.duplicated(subset=['Zone', 'Produit']).sum()
```

Out[13]: 0

```
In [14]: dispo_alimentaire.isna().sum()
```

```
Out[14]: Zone 0
Produit 0
Origine 0
Aliments pour animaux 12885
Autres Utilisations 10109
Disponibilité alimentaire (Kcal/personne/jour) 1364
Disponibilité alimentaire en quantité (kg/personne/an) 1590
Disponibilité de matière grasse en quantité (g/personne/jour) 3811
Disponibilité de protéines en quantité (g/personne/jour) 4044
Disponibilité intérieure 223
Exportations - Quantité 3379
Importations - Quantité 753
Nourriture 1590
Pertes 11327
Production 6425
Semences 13514
Traitement 13313
Variation de stock 8829
dtype: int64
```

```
In [15]: dispo_alimentaire.fillna(0,inplace=True)
dispo_alimentaire.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15605 entries, 0 to 15604
Data columns (total 18 columns):
#   Column                                     Non-Null Count
Dtype
---  ---
0    Zone                                     15605 non-null
object
1    Produit                                 15605 non-null
object
2    Origine                                15605 non-null
object
3    Aliments pour animaux                  15605 non-null
float64
4    Autres Utilisations                   15605 non-null
float64
5    Disponibilité alimentaire (Kcal/personne/jour) 15605 non-null
float64
6    Disponibilité alimentaire en quantité (kg/personne/an) 15605 non-null
float64
7    Disponibilité de matière grasse en quantité (g/personne/jour) 15605 non-null
float64
8    Disponibilité de protéines en quantité (g/personne/jour) 15605 non-null
float64
9    Disponibilité intérieure                15605 non-null
float64
10   Exportations - Quantité                15605 non-null
float64
11   Importations - Quantité                15605 non-null
float64
12   Nourriture                            15605 non-null
float64
13   Pertes                                15605 non-null
float64
14   Production                            15605 non-null
float64
15   Semences                              15605 non-null
float64
16   Traitement                            15605 non-null
```

```
float64
17 Variation de stock
float64
dtypes: float64(15), object(3)
memory usage: 2.1+ MB
```

15605 non-null

```
In [16]: #multiplication de toutes les lignes contenant des milliers de tonnes en Kg
columns_to_multiply = [
    'Aliments pour animaux',
    'Disponibilité intérieure',
    'Exportations - Quantité',
    'Importations - Quantité',
    'Nourriture',
    'Pertes',
    'Production',
    'Semences',
    'Traitement',
    'Variation de stock']

for column in columns_to_multiply:
    dispo_alimentaire[column] = dispo_alimentaire[column] * 1000
```

```
In [17]: sorted(list(dispo_alimentaire['Zone'].unique()))
```

```
Out[17]: ['Afghanistan',
'Afrique du Sud',
'Albanie',
'Algérie',
'Allemagne',
'Angola',
'Antigua-et-Barbuda',
'Arabie saoudite',
'Argentine',
'Arménie',
'Australie',
'Autriche',
'Azerbaïdjan',
'Bahamas',
'Bangladesh',
'Barbade',
'Belgique',
'Belize',
'Bermudes',
'Bolivie (État plurinational de)',
'Bosnie-Herzégovine',
'Botswana',
'Brunéi Darussalam',
'Brésil',
'Bulgarie',
'Burkina Faso',
'Bélarus',
'Bénin',
'Cabo Verde',
'Cambodge',
'Cameroun',
'Canada',
'Chili',
'Chine - RAS de Hong-Kong',
'Chine - RAS de Macao',
'Chine, Taiwan Province de',
'Chine, continentale',
'Chypre',
'Colombie',
'Congo',
'Costa Rica',
```

'Croatie',
'Cuba',
"Côte d'Ivoire",
'Danemark',
'Djibouti',
'Dominique',
'El Salvador',
'Espagne',
'Estonie',
'Eswatini',
'Fidji',
'Finlande',
'France',
'Fédération de Russie',
'Gabon',
'Gambie',
'Ghana',
'Grenade',
'Grèce',
'Guatemala',
'Guinée',
'Guinée-Bissau',
'Guyana',
'Géorgie',
'Haïti',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
"Iran (République islamique d')",
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',
'Kirghizistan',
'Kiribati',
'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libéria',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',
'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Maurice',
'Mauritanie',
'Mexique',
'Mongolie',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nicaragua',

'Niger',
'Nigéria',
'Norvège',
'Nouvelle-Calédonie',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Panama',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Polynésie française',
'Portugal',
'Pérou',
'Roumanie',
'Royaume-Uni',
'Rwanda',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Saint-Kitts-et-Nevis',
'Saint-Vincent-et-les Grenadines',
'Sainte-Lucie',
'Samoa',
'Sao Tomé-et-Principe',
'Serbie',
'Sierra Leone',
'Slovaquie',
'Slovénie',
'Soudan',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie (la)',
'Thaïlande',
'Timor-Leste',
'Togo',
'Trinité-et-Tobago',
'Tunisie',
'Turkménistan',
'Turquie',
'Ukraine',
'Uruguay',
'Vanuatu',
'Venezuela (République bolivarienne du)',
'Viet Nam',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Émirats arabes unis',
'Équateur',
'États-Unis d'Amérique',

```
'Éthiopie',  
'Îles Salomon']
```

```
In [18]: liste=['Chine - RAS de Hong-Kong','Chine - RAS de Macao','Chine, Taiwan Province de'  
             'Saint-Kitts-et-Nevis','Saint-Vincent-et-les Grenadines','Sainte-Lucie','Samoa'  
             'Îles Salomon']  
dispo_alimentaire.drop(dispo_alimentaire[dispo_alimentaire['Zone'].isin(liste)].index)
```

```
In [19]: dispo_alimentaire.replace("Tchéquie (la)","Tchéquie",inplace=True)  
sorted(dispo_alimentaire['Zone'].unique())
```

```
Out[19]: ['Afghanistan',  
          'Afrique du Sud',  
          'Albanie',  
          'Algérie',  
          'Allemagne',  
          'Angola',  
          'Antigua-et-Barbuda',  
          'Arabie saoudite',  
          'Argentine',  
          'Arménie',  
          'Australie',  
          'Autriche',  
          'Azerbaïdjan',  
          'Bahamas',  
          'Bangladesh',  
          'Belgique',  
          'Belize',  
          'Bolivie (État plurinational de)',  
          'Bosnie-Herzégovine',  
          'Botswana',  
          'Brunéi Darussalam',  
          'Brésil',  
          'Bulgarie',  
          'Burkina Faso',  
          'Biélarus',  
          'Bénin',  
          'Cabo Verde',  
          'Cambodge',  
          'Cameroun',  
          'Canada',  
          'Chili',  
          'Chine, continentale',  
          'Chypre',  
          'Colombie',  
          'Congo',  
          'Costa Rica',  
          'Croatie',  
          'Cuba',  
          "Côte d'Ivoire",  
          'Danemark',  
          'Djibouti',  
          'El Salvador',  
          'Espagne',  
          'Estonie',  
          'Eswatini',  
          'Fidji',  
          'Finlande',  
          'France',  
          'Fédération de Russie',  
          'Gabon',  
          'Gambie',  
          'Ghana',  
          'Grenade',  
          'Grèce',
```

'Guatemala',
'Guinée',
'Guinée-Bissau',
'Guyana',
'Géorgie',
'Haïti',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
'Iran (République islamique d')',
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',
'Kirghizistan',
'Kiribati',
'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libéria',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',
'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Maurice',
'Mauritanie',
'Mexique',
'Mongolie',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nicaragua',
'Niger',
'Nigéria',
'Norvège',
'Nouvelle-Calédonie',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Panama',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Portugal',
'Pérou',
'Roumanie',
'Royaume-Uni',

```

'Rwanda',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Serbie',
'Sierra Leone',
'Slovaquie',
'Slovénie',
'Soudan',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie',
'Thaïlande',
'Togo',
'Trinité-et-Tobago',
'Tunisie',
'Turkménistan',
'Turquie',
'Ukraine',
'Uruguay',
'Vanuatu',
'Venezuela (République bolivarienne du)',
'Viet Nam',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Émirats arabes unis',
'Équateur',
"États-Unis d'Amérique",
'Éthiopie']

```

```
In [20]: dispo_alimentaire.to_csv('dispo_alimentaire_clean.csv', index=False)
```

3 - Analyse exploratoire du fichier sous nutrition

</div>

```
In [21]: sous_nutrition = pd.read_csv('sous_nutrition.csv')
sous_nutrition.head()
```

Out[21]:

	Zone	Année	Valeur
--	------	-------	--------

0	Afghanistan	2012-2014	8.6
---	-------------	-----------	-----

1	Afghanistan	2013-2015	8.8
---	-------------	-----------	-----

2	Afghanistan	2014-2016	8.9
---	-------------	-----------	-----

3	Afghanistan	2015-2017	9.7
---	-------------	-----------	-----

4	Afghanistan	2016-2018	10.5
---	-------------	-----------	------

```
In [22]: #Afficher les dimensions du dataset
print("Le tableau comporte {} observation(s) ou article(s)".format(sous_nutrition.s
print("Le tableau comporte {} colonne(s)".format(sous_nutrition.shape[1]))
```

Le tableau comporte 1218 observation(s) ou article(s)
Le tableau comporte 3 colonne(s)

```
In [23]: #Consulter le nombre de colonnes
sous_nutrition.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1218 entries, 0 to 1217
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Zone        1218 non-null   object
1   Année       1218 non-null   object
2   Valeur      624 non-null    object
dtypes: object(3)
memory usage: 28.7+ KB
```

```
In [24]: sous_nutrition['Valeur'] = pd.to_numeric(sous_nutrition['Valeur'], errors='coerce')
```

```
In [25]: sous_nutrition['année_convert']=sous_nutrition.Année.apply(lambda x:int(x[0:4])+1)
sous_nutrition.drop(columns='Année',axis=1,inplace=True)
sous_nutrition.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1218 entries, 0 to 1217
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Zone            1218 non-null   object
1   Valeur          504 non-null    float64
2   année_convert   1218 non-null   int64
dtypes: float64(1), int64(1), object(1)
memory usage: 28.7+ KB
```

```
In [26]: sous_nutrition.rename(columns={'année_convert':'Année','Valeur':'Sous_nutrition'},i
sous_nutrition.head()
```

Out[26]:

	Zone	Sous_nutrition	Année
--	------	----------------	-------

0	Afghanistan	8.6	2013
1	Afghanistan	8.8	2014
2	Afghanistan	8.9	2015
3	Afghanistan	9.7	2016
4	Afghanistan	10.5	2017

```
In [27]: sous_nutrition['Sous_nutrition']=sous_nutrition['Sous_nutrition']*1000000
sous_nutrition
```

Out[27]:

	Zone	Sous_nutrition	Année
--	------	----------------	-------

0	Afghanistan	8600000.0	2013
1	Afghanistan	8800000.0	2014
2	Afghanistan	8900000.0	2015
3	Afghanistan	9700000.0	2016

4	Afghanistan	10500000.0	2017
...
1213	Zimbabwe	NaN	2014
1214	Zimbabwe	NaN	2015
1215	Zimbabwe	NaN	2016
1216	Zimbabwe	NaN	2017
1217	Zimbabwe	NaN	2018

1218 rows × 3 columns

```
In [29]: sous_nutrition['Sous_nutrition'].fillna(0,inplace=True)
sous_nutrition.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1218 entries, 0 to 1217
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Zone            1218 non-null   object
1   Sous_nutrition  1218 non-null   float64
2   Année           1218 non-null   int64
dtypes: float64(1), int64(1), object(1)
memory usage: 28.7+ KB
```

```
In [30]: sorted(sous_nutrition['Zone'].unique())
```

```
Out[30]: ['Afghanistan',
'Afrique du Sud',
'Albanie',
'Algérie',
'Allemagne',
'Andorre',
'Angola',
'Antigua-et-Barbuda',
'Arabie saoudite',
'Argentine',
'Arménie',
'Australie',
'Autriche',
'Azerbaïdjan',
'Bahamas',
'Bahreïn',
'Bangladesh',
'Barbade',
'Belgique',
'Belize',
'Bermudes',
'Bhoutan',
'Bolivie (État plurinational de)',
'Bosnie-Herzégovine',
'Botswana',
'Brunéi Darussalam',
'Brésil',
'Bulgarie',
'Burkina Faso',
'Burundi',
'Bélarus',
'Bénin',
'Cabo Verde',
'Cambodge',
```

'Cameroun',
'Canada',
'Chili',
'Chine - RAS de Hong-Kong',
'Chine - RAS de Macao',
'Chine, Taiwan Province de',
'Chine, continentale',
'Chypre',
'Colombie',
'Comores',
'Congo',
'Costa Rica',
'Croatie',
'Cuba',
'Côte d'Ivoire',
'Danemark',
'Djibouti',
'Dominique',
'El Salvador',
'Espagne',
'Estonie',
'Eswatini',
'Fidji',
'Finlande',
'France',
'Fédération de Russie',
'Gabon',
'Gambie',
'Ghana',
'Grenade',
'Groenland',
'Grèce',
'Guatemala',
'Guinée',
'Guinée équatoriale',
'Guinée-Bissau',
'Guyana',
'Géorgie',
'Haïti',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
'Iran (République islamique d')',
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',
'Kirghizistan',
'Kiribati',
'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libye',
'Libéria',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',

'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Maurice',
'Mauritanie',
'Mexique',
'Micronésie (États fédérés de)',
'Mongolie',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nauru',
'Nicaragua',
'Niger',
'Nigéria',
'Nioué',
'Norvège',
'Nouvelle-Calédonie',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Palaos',
'Palestine',
'Panama',
'Papouasie-Nouvelle-Guinée',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Polynésie française',
'Porto Rico',
'Portugal',
'Pérou',
'Qatar',
'Roumanie',
'Royaume-Uni de Grande-Bretagne et d'Irlande du Nord',
'Rwanda',
'République arabe syrienne',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',
'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Saint-Kitts-et-Nevis',
'Saint-Vincent-et-les Grenadines',
'Sainte-Lucie',
'Samoa',
'Samoa américaines',
'Sao Tomé-et-Principe',
'Serbie',
'Seychelles',
'Sierra Leone',
'Singapour',
'Slovaquie',
'Slovénie',
'Somalie',


```
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie',
'Thaïlande',
'Timor-Leste',
'Togo',
'Tokélaou',
'Tonga',
'Trinité-et-Tobago',
'Tunisie',
'Turkménistan',
'Turquie',
'Tuvalu',
'Ukraine',
'Uruguay',
'Vanuatu',
'Venezuela (République bolivarienne du)',
'Viet Nam',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Émirats arabes unis',
'Équateur',
'Érythrée',
"États-Unis d'Amérique",
'Éthiopie',
'îles Cook',
'îles Marshall',
'îles Salomon']
```

```
In [31]: liste=['Polynésie française','Barbade','Bermudes','Chine - RAS de Hong-Kong','Chine
             'Haïti','Saint-Kitts-et-Nevis','Saint-Vincent-et-les Grenadines','Sainte-Luc
             'Trinité-et-Tobago','Îles Cook','Îles Marshall','Îles Salomon','Timor-Leste'
sous_nutrition.drop(sous_nutrition[sous_nutrition['Zone'].isin(liste)].index, inplace=True)
```

```
In [32]: sous_nutrition.replace("Royaume-Uni de Grande-Bretagne et d'Irlande du Nord","Royaume-Uni")
```

```
In [33]: sorted(sous_nutrition['Zone'].unique())
```

```
Out[33]: ['Afghanistan',
'Afrique du Sud',
'Albanie',
'Algérie',
'Allemagne',
'Andorre',
'Angola',
'Antigua-et-Barbuda',
'Arabie saoudite',
'Argentine',
'Arménie',
'Australie',
'Autriche',
'Azerbaïdjan',
'Bahamas',
'Bahreïn',
'Bangladesh',
'Belgique',
```

'Belize',
'Bhoutan',
'Bolivie (État plurinational de)',
'Bosnie-Herzégovine',
'Botswana',
'Brunéi Darussalam',
'Brésil',
'Bulgarie',
'Burkina Faso',
'Burundi',
'Bélarus',
'Bénin',
'Cabo Verde',
'Cambodge',
'Cameroun',
'Canada',
'Chili',
'Chine, continentale',
'Chypre',
'Colombie',
'Comores',
'Congo',
'Costa Rica',
'Croatie',
'Cuba',
'Côte d'Ivoire',
'Danemark',
'Djibouti',
'Dominique',
'El Salvador',
'Espagne',
'Estonie',
'Eswatini',
'Fidji',
'Finlande',
'France',
'Fédération de Russie',
'Gabon',
'Gambie',
'Ghana',
'Grenade',
'Groenland',
'Grèce',
'Guatemala',
'Guinée',
'Guinée équatoriale',
'Guinée-Bissau',
'Guyana',
'Géorgie',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
'Iran (République islamique d')',
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',
'Kirghizistan',
'Kiribati',

'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libye',
'Libéria',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',
'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Maurice',
'Mauritanie',
'Mexique',
'Micronésie (États fédérés de)',
'Mongolie',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nauru',
'Nicaragua',
'Niger',
'Nigéria',
'Nioué',
'Norvège',
'Nouvelle-Calédonie',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Palaos',
'Palestine',
'Panama',
'Papouasie-Nouvelle-Guinée',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Porto Rico',
'Portugal',
'Pérou',
'Qatar',
'Roumanie',
'Royaume-Uni',
'Rwanda',
'République arabe syrienne',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',
'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Serbie',
'Seychelles',
'Sierra Leone',
'Singapour',

```
'Slovaquie',
'Slovénie',
'Somalie',
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie',
'Thaïlande',
'Togo',
'Tokélaou',
'Tonga',
'Tunisie',
'Turkménistan',
'Turquie',
'Tuvalu',
'Ukraine',
'Uruguay',
'Vanuatu',
'Venezuela (République bolivarienne du)',
'Viet Nam',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Émirats arabes unis',
'Équateur',
'Érythrée',
"États-Unis d'Amérique",
'Éthiopie']
```

```
In [34]: sous_nutrition.to_csv('sous_nutrition_clean.csv', index=False)
```

4 - Analyse exploratoire du fichier population

</div>

```
In [59]: population = pd.read_csv('population.csv')
population.head()
```

```
Out[59]:
```

	Zone	Année	Valeur
0	Afghanistan	2013	32269.589
1	Afghanistan	2014	33370.794
2	Afghanistan	2015	34413.603
3	Afghanistan	2016	35383.032
4	Afghanistan	2017	36296.113

```
In [60]: #Afficher les dimensions du dataset
print("Le tableau comporte {} observation(s) ou article(s)".format(population.shape[0]))
print("Le tableau comporte {} colonne(s)".format(population.shape[1]))
```

Le tableau comporte 1416 observation(s) ou article(s)
Le tableau comporte 3 colonne(s)

In [61]: `population.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1416 entries, 0 to 1415
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Zone        1416 non-null   object
1   Année       1416 non-null   int64
2   Valeur      1416 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 33.3+ KB
```

In [62]: `population['Valeur']=population['Valeur']*1e3`

In [63]: `population.rename(columns={'Valeur':'Population'},inplace=True)`
`population.head()`

Out[63]:

	Zone	Année	Population
--	------	-------	------------

0	Afghanistan	2013	32269589.0
1	Afghanistan	2014	33370794.0
2	Afghanistan	2015	34413603.0
3	Afghanistan	2016	35383032.0
4	Afghanistan	2017	36296113.0

In [64]: `sorted(population['Zone'].unique())`

Out[64]:

```
['Afghanistan',
'Afrique du Sud',
'Albanie',
'Algérie',
'Allemagne',
'Andorre',
'Angola',
'Anguilla',
'Antigua-et-Barbuda',
'Antilles néerlandaises (ex)',
'Arabie saoudite',
'Argentine',
'Arménie',
'Aruba',
'Australie',
'Autriche',
'Azerbaïdjan',
'Bahamas',
'Bahreïn',
'Bangladesh',
'Barbade',
'Belgique',
'Belize',
'Bermudes',
'Bhoutan',
'Bolivie (État plurinational de)',
'Bonaire, Saint-Eustache et Saba',
'Bosnie-Herzégovine',
'Botswana',
'Brunéi Darussalam',
'Brésil',
```

'Bulgarie',
'Burkina Faso',
'Burundi',
'Bélarus',
'Bénin',
'Cabo Verde',
'Cambodge',
'Cameroun',
'Canada',
'Chili',
'Chine - RAS de Hong-Kong',
'Chine - RAS de Macao',
'Chine, Taiwan Province de',
'Chine, continentale',
'Chypre',
'Colombie',
'Comores',
'Congo',
'Costa Rica',
'Croatie',
'Cuba',
'Curaçao',
'Côte d'Ivoire',
'Danemark',
'Djibouti',
'Dominique',
'El Salvador',
'Espagne',
'Estonie',
'Eswatini',
'Fidji',
'Finlande',
'France',
'Fédération de Russie',
'Gabon',
'Gambie',
'Ghana',
'Gibraltar',
'Grenade',
'Groenland',
'Grèce',
'Guadeloupe',
'Guam',
'Guatemala',
'Guinée',
'Guinée équatoriale',
'Guinée-Bissau',
'Guyana',
'Guyane française',
'Géorgie',
'Haïti',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
'Iran (République islamique d')',
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',

'Kirghizistan',
'Kiribati',
'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libye',
'Libéria',
'Liechtenstein',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',
'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Martinique',
'Maurice',
'Mauritanie',
'Mayotte',
'Mexique',
'Micronésie (États fédérés de)',
'Monaco',
'Mongolie',
'Montserrat',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nauru',
'Nicaragua',
'Niger',
'Nigéria',
'Nioué',
'Norvège',
'Nouvelle-Calédonie',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Palaos',
'Palestine',
'Panama',
'Papouasie-Nouvelle-Guinée',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Polynésie française',
'Porto Rico',
'Portugal',
'Pérou',
'Qatar',
'Roumanie',
'Royaume-Uni de Grande-Bretagne et d'Irlande du Nord',
'Rwanda',
'République arabe syrienne',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',

'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Réunion',
'Sahara occidental',
'Saint-Barthélemy',
'Saint-Kitts-et-Nevis',
'Saint-Marin',
'Saint-Martin (partie française)',
'Saint-Pierre-et-Miquelon',
'Saint-Siège',
'Saint-Vincent-et-les Grenadines',
'Sainte-Hélène, Ascension et Tristan da Cunha',
'Sainte-Lucie',
'Samoa',
'Samoa américaines',
'Sao Tomé-et-Principe',
'Serbie',
'Seychelles',
'Sierra Leone',
'Singapour',
'Sint Maarten (partie néerlandaise)',
'Slovaquie',
'Slovénie',
'Somalie',
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie',
'Thaïlande',
'Timor-Leste',
'Togo',
'Tokélaou',
'Tonga',
'Trinité-et-Tobago',
'Tunisie',
'Turkménistan',
'Turquie',
'Tuvalu',
'Ukraine',
'Uruguay',
'Vanuatu',
'Venezuela (République bolivarienne du)',
'Viet Nam',
'Yémen',
'Zambie',
'Zimbabwe',
'Égypte',
'Émirats arabes unis',
'Équateur',
'Érythrée',
'États-Unis d'Amérique',
'Éthiopie',
'Île de Man',
'Îles Anglo-Normandes',
'Îles Caïmanes',
'Îles Cook',
'Îles Falkland (Malvinas)',
'Îles Féroé',


```
'îles Mariannes du Nord',
'îles Marshall',
'îles Salomon',
'îles Turques-et-Caïques',
'îles Vierges américaines',
'îles Vierges britanniques',
'îles Wallis-et-Futuna']
```

```
In [65]: liste=['Polynésie française','Barbade','Bermudes','Chine - RAS de Hong-Kong','Chir
'îles Salomon','Sint Maarten (partie néerlandaise)','Sainte-Hélène, Ascens
'Sao Tomé-et-Principe','Saint-Barthélemy','Saint-Kitts-et-Nevis','Saint-Mar
'Ascension et Tristan da Cunha','Kiribati','Sainte-Lucie', 'Antigua-et-Bark
'Nouvelle-Calédonie','Palaos','Réunion','Île de Man','îles Anglo-Normandes'
'îles Féroé','îles Mariannes du Nord','îles Marshall','îles Turques-et-Caïc
'îles Wallis-et-Futuna','Timor-Leste']
population.drop(population[population['Zone'].isin(liste)].index, inplace=True)
```

```
In [66]: population.replace("Royaume-Uni de Grande-Bretagne et d'Irlande du Nord","Royaume-
sorted(population['Zone'].unique())
```

```
Out[66]: ['Afghanistan',
'Afrique du Sud',
'Albanie',
'Algérie',
'Allemagne',
'Andorre',
'Angola',
'Anguilla',
'Arabie saoudite',
'Argentine',
'Arménie',
'Aruba',
'Australie',
'Autriche',
'Azerbaïdjan',
'Bahamas',
'Bahreïn',
'Bangladesh',
'Belgique',
'Belize',
'Bhoutan',
'Bolivie (État plurinational de)',
'Bonaire, Saint-Eustache et Saba',
'Bosnie-Herzégovine',
'Botswana',
'Brunéi Darussalam',
'Brésil',
'Bulgarie',
'Burkina Faso',
'Burundi',
'Bélarus',
'Bénin',
'Cabo Verde',
'Cambodge',
'Cameroun',
'Canada',
'Chili',
'Chine, continentale',
'Chypre',
'Colombie',
'Comores',
'Congo',
'Costa Rica',
'Croatie',
'Cuba',
```

'Curaçao',
"Côte d'Ivoire",
'Danemark',
'Djibouti',
'El Salvador',
'Espagne',
'Estonie',
'Eswatini',
'Fidji',
'Finlande',
'France',
'Fédération de Russie',
'Gabon',
'Gambie',
'Ghana',
'Gibraltar',
'Grenade',
'Groenland',
'Grèce',
'Guadeloupe',
'Guam',
'Guatemala',
'Guinée',
'Guinée équatoriale',
'Guinée-Bissau',
'Guyana',
'Guyane française',
'Géorgie',
'Haïti',
'Honduras',
'Hongrie',
'Inde',
'Indonésie',
"Iran (République islamique d')",
'Iraq',
'Irlande',
'Islande',
'Israël',
'Italie',
'Jamaïque',
'Japon',
'Jordanie',
'Kazakhstan',
'Kenya',
'Kirghizistan',
'Koweït',
'Lesotho',
'Lettonie',
'Liban',
'Libye',
'Libéria',
'Liechtenstein',
'Lituanie',
'Luxembourg',
'Macédoine du Nord',
'Madagascar',
'Malaisie',
'Malawi',
'Maldives',
'Mali',
'Malte',
'Maroc',
'Martinique',
'Maurice',
'Mauritanie',
'Mayotte',

'Mexique',
'Micronésie (États fédérés de)',
'Monaco',
'Mongolie',
'Montserrat',
'Monténégro',
'Mozambique',
'Myanmar',
'Namibie',
'Nauru',
'Nicaragua',
'Niger',
'Nigéria',
'Nioué',
'Norvège',
'Nouvelle-Zélande',
'Népal',
'Oman',
'Ouganda',
'Ouzbékistan',
'Pakistan',
'Palestine',
'Panama',
'Papouasie-Nouvelle-Guinée',
'Paraguay',
'Pays-Bas',
'Philippines',
'Pologne',
'Porto Rico',
'Portugal',
'Pérou',
'Qatar',
'Roumanie',
'Royaume-Uni',
'Rwanda',
'République arabe syrienne',
'République centrafricaine',
'République de Corée',
'République de Moldova',
'République dominicaine',
'République démocratique du Congo',
'République démocratique populaire lao',
'République populaire démocratique de Corée',
'République-Unie de Tanzanie',
'Sahara occidental',
'Serbie',
'Seychelles',
'Sierra Leone',
'Singapour',
'Slovaquie',
'Slovénie',
'Somalie',
'Soudan',
'Soudan du Sud',
'Sri Lanka',
'Suisse',
'Suriname',
'Suède',
'Sénégal',
'Tadjikistan',
'Tchad',
'Tchéquie',
'Thaïlande',
'Togo',
'Tokélaou',
'Tonga',

```
'Trinité-et-Tobago',  
'Tunisie',  
'Turkménistan',  
'Turquie',  
'Tuvalu',  
'Ukraine',  
'Uruguay',  
'Vanuatu',  
'Venezuela (République bolivarienne du)',  
'Viet Nam',  
'Yémen',  
'Zambie',  
'Zimbabwe',  
'Égypte',  
'Émirats arabes unis',  
'Équateur',  
'Érythrée',  
'États-Unis d'Amérique',  
'Éthiopie']
```

```
In [67]: population.to_csv('population_clean.csv', index=False)
```

```
In [ ]:
```

```
In [ ]:
```