

# Concevez une application au service de la santé publique

Description du travail recopiée du site : L'agence "Santé publique France" a lancé un appel à projets pour trouver des idées innovantes d'applications en lien avec l'alimentation. Vous souhaitez y participer et proposer une idée d'application.

## Les données

Extrait de l'appel à projets :

Le jeu de données Open Food Facts est disponible sur le site officiel (ou disponible à ce lien en téléchargement). Les variables sont définies à cette adresse: (<https://world.openfoodfacts.org/data/data-fields.txt>)

### Données téléchargées dans le dossier "datasets"

Les champs sont séparés en quatre sections : Les informations générales sur la fiche du produit : nom, date de modification, etc. Un ensemble de tags : catégorie du produit, localisation, origine, etc. Les ingrédients composant les produits et leurs additifs éventuels. Des informations nutritionnelles : quantité en grammes d'un nutriment pour 100 grammes du produit.

## Votre mission

Après avoir lu l'appel à projets, voici les différentes étapes que vous avez identifiées :

1) Traiter le jeu de données, en :

- Réfléchissant à une idée d'application.
- Repérant des variables pertinentes pour les traitements à venir, et nécessaires pour votre idée d'application.
- Nettoyant les données en :
  - mettant en évidence les éventuelles valeurs manquantes, avec au moins 3 méthodes de traitement adaptées aux variables concernées,
  - identifiant et en quantifiant les éventuelles valeurs aberrantes de chaque variable.
  - Automatisant ces traitements pour éviter de répéter ces opérations Le programme doit fonctionner si la base de données est légèrement modifiée (ajout d'entrées, par exemple).

2) Tout au long de l'analyse, produire des visualisations afin de mieux comprendre les données. Effectuer une analyse univariée pour chaque variable intéressante, afin de synthétiser son comportement.

L'appel à projets spécifie que l'analyse doit être simple à comprendre pour un public néophyte. Soyez donc attentif à la lisibilité : taille des textes, choix des couleurs, netteté suffisante, et variez les graphiques (boxplots, histogrammes, diagrammes circulaires, nuages de points...) pour illustrer au mieux votre propos.

3) Confirmer ou infirmer les hypothèses à l'aide d'une analyse multivariée. Effectuer les tests statistiques appropriés pour vérifier la significativité des résultats.

4) Justifier votre idée d'application. Identifier des arguments justifiant la faisabilité (ou non) de l'application à partir des données Open Food Facts.

5) Rédiger un rapport d'exploration et pitcher votre idée durant la soutenance du projet.

## Livrables

- Un notebook du nettoyage des données (non cleané, pour comprendre votre démarche).
- Un notebook d'exploration comportant une analyse univariée, multivariée, une réduction dimensionnelle, ainsi que les différentes questions de recherches associées (non cleané, pour comprendre votre démarche).
- Une présentation, à utiliser en support pour la soutenance, comprenant :
  - La présentation de votre idée d'application.
  - Les opérations de nettoyage effectuées.
  - La description et l'analyse univariée des différentes variables importantes avec les visualisations associées.
  - L'analyse multivariée et les résultats statistiques associés, en lien avec votre idée d'application.
  - 3 observations solidement étayées (graphes et/ou tests statistiques à l'appui au besoin) évaluant la pertinence et la faisabilité de votre application.
  - La synthèse des différentes conclusions sur la faisabilité de votre projet.

Conseil additionnel : Il faut rester au stade de l'idée. On doit extraire une conclusion de chaque graphique. Doit-on mettre en oeuvre un algorithme de machine learning ? Non, pas forcément. Si on le temps utiliser KNN

In [184]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

#erreur au chargement -
#ParserError: Error tokenizing data. C error: Expected 34 fields in Line 3, saw 37
#on dirait que le CSV est séparé par des tabs ?
#il faut donc ajouter un séparateur tab au read_csv

myDF = pd.read_csv('.\Datasets\fr.openfoodfacts.org.products.csv', sep ='\t')
print("Données chargées")
```

D:\Users\yag\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (0,3,5,19,20,24,25,26,27,28,35,36,37,38,39,48) have mixed types.Specify dtype option on import or set low\_memory=False.
exec(code\_obj, self.user\_global\_ns, self.user\_ns)

Données chargées

Il m'arrive de voir le message d'erreur suivant pour la cellule précédente : D:\Users\yag\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (0,3,5,19,20,24,25,26,27,28,35,36,37,38,39,48) have mixed types.Specify dtype option on import or set low\_memory=False. exec(code\_obj, self.user\_global\_ns, self.user\_ns)'

MENTOR : Pourquoi? Explication ici : <https://stackoverflow.com/questions/24251219/pandas-read-csv-low-memory-and-dtype-options>

Faut-il faire quelque chose? L'explication ci-dessus ne le suggère pas.

=> regarder la fonction unique qui affiche les modalités unique => Ajouter dans le nettoyage le réglage pour les colonnes problématiques (ssi besoin)

In [185...]

```
myDF.head()
```

Out[185...]

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	quantity	packagin
0	3087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir		NaN	1kg
1	4530	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Banana Chips Sweetened (Whole)		NaN	NaN
2	4559	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Peanuts		NaN	NaN
3	16087	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055731	2017-03-09T10:35:31Z	1489055731	2017-03-09T10:35:31Z	Organic Salted Nut Mix		NaN	NaN
4	16094	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055653	2017-03-09T10:34:13Z	1489055653	2017-03-09T10:34:13Z	Organic Polenta		NaN	NaN

In [186...]

```
myDF.shape
```

Out[186...]

```
(320772, 162)
```

In [187...]

```
myDF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 320772 entries, 0 to 320771
Columns: 162 entries, code to water-hardness_100g
dtypes: float64(106), object(56)
memory usage: 396.5+ MB
```

In [188...]

```
myColTypes = myDF.dtypes
```

```
myColTypes
```

Out[188...]

code	object
url	object
creator	object
created_t	object
created_datetime	object
last_modified_t	object
last_modified_datetime	object
product_name	object
generic_name	object
quantity	object
packaging	object
packaging_tags	object
brands	object
brands_tags	object
categories	object
categories_tags	object
categories_fr	object
origins	object
origins_tags	object
manufacturing_places	object
manufacturing_places_tags	object
labels	object
labels_tags	object
labels_fr	object
emb_codes	object
emb_codes_tags	object
first_packaging_code_geo	object
cities	object
cities_tags	object
purchase_places	object
stores	object
countries	object
countries_tags	object
countries_fr	object
ingredients_text	object
allergens	object
allergens_fr	object
traces	object
traces_tags	object
traces_fr	object
serving_size	object
no_nutriments	float64
additives_n	float64

additives	object
additives_tags	object
additives_fr	object
ingredients_from_palm_oil_n	float64
ingredients_from_palm_oil	float64
ingredients_from_palm_oil_tags	object
ingredients_that_may_be_from_palm_oil_n	float64
ingredients_that_may_be_from_palm_oil	float64
ingredients_that_may_be_from_palm_oil_tags	object
nutrition_grade_uk	float64
nutrition_grade_fr	object
pnns_groups_1	object
pnns_groups_2	object
states	object
states_tags	object
states_fr	object
main_category	object
main_category_fr	object
image_url	object
image_small_url	object
energy_100g	float64
energy-from-fat_100g	float64
fat_100g	float64
saturated-fat_100g	float64
butyric-acid_100g	float64
caproic-acid_100g	float64
caprylic-acid_100g	float64
capric-acid_100g	float64
lauric-acid_100g	float64
myristic-acid_100g	float64
palmitic-acid_100g	float64
stearic-acid_100g	float64
arachidic-acid_100g	float64
behenic-acid_100g	float64
lignoceric-acid_100g	float64
cerotic-acid_100g	float64
montanic-acid_100g	float64
melissic-acid_100g	float64
monounsaturated-fat_100g	float64
polyunsaturated-fat_100g	float64
omega-3-fat_100g	float64
alpha-linolenic-acid_100g	float64
eicosapentaenoic-acid_100g	float64
docosahexaenoic-acid_100g	float64
omega-6-fat_100g	float64
linoleic-acid_100g	float64
arachidonic-acid_100g	float64
gamma-linolenic-acid_100g	float64
dihomo-gamma-linolenic-acid_100g	float64
omega-9-fat_100g	float64
oleic-acid_100g	float64
elaidic-acid_100g	float64
gondoic-acid_100g	float64
mead-acid_100g	float64
erucic-acid_100g	float64
nervonic-acid_100g	float64
trans-fat_100g	float64
cholesterol_100g	float64
carbohydrates_100g	float64
sugars_100g	float64
sucrose_100g	float64
glucose_100g	float64
fructose_100g	float64
lactose_100g	float64
maltose_100g	float64
maltdextrins_100g	float64
starch_100g	float64
polyols_100g	float64
fiber_100g	float64
proteins_100g	float64
casein_100g	float64
serum-proteins_100g	float64
nucleotides_100g	float64
salt_100g	float64
sodium_100g	float64
alcohol_100g	float64
vitamin-a_100g	float64
beta-carotene_100g	float64
vitamin-d_100g	float64
vitamin-e_100g	float64
vitamin-k_100g	float64
vitamin-c_100g	float64
vitamin-b1_100g	float64
vitamin-b2_100g	float64
vitamin-pp_100g	float64
vitamin-b6_100g	float64
vitamin-b9_100g	float64
folates_100g	float64
vitamin-b12_100g	float64
biotin_100g	float64
pantothenic-acid_100g	float64
silica_100g	float64
bicarbonate_100g	float64
potassium_100g	float64
chloride_100g	float64
calcium_100g	float64
phosphorus_100g	float64
iron_100g	float64
magnesium_100g	float64
zinc_100g	float64
copper_100g	float64
manganese_100g	float64
fluoride_100g	float64
selenium_100g	float64
chromium_100g	float64
molybdenum_100g	float64
iodine_100g	float64

```

caffeine_100g           float64
taurine_100g            float64
ph_100g                 float64
fruits-vegetables-nuts_100g float64
collagen-meat-protein-ratio_100g float64
cocoa_100g               float64
chlorophyl_100g          float64
carbon-footprint_100g    float64
nutrition-score-fr_100g   float64
nutrition-score-uk_100g   float64
glycemic-index_100g      float64
water-hardness_100g       float64
dtype: object

```

In [ ]:

In [ ]:

In [ ]:

Je ne comprends pas pourquoi le "shape" me renvoie 162 colonnes, alors que je n'en compte que 20. C'est peut-être une limitation du head() ? De même, le dtypes ne m'indique par toutes les lignes.

=> Oui, c'est bien une histoire de taille d'affichage. Solution à retenir = il faut changer les options d'affichage pour voir toutes les colonnes.

In [189...]: myDF.describe()

Out[189...]:

	no_nutriments	additives_n	ingredients_from_palm_oil_n	ingredients_from_palm_oil	ingredients_that_may_be_from_palm_oil_n	ingredients_that_may_be_from_palm_oil	nutrit
<b>count</b>	0.0	248939.000000	248939.000000	0.0	248939.000000	0.0	
<b>mean</b>	NaN	1.936024	0.019659	NaN	0.055246	NaN	
<b>std</b>	NaN	2.502019	0.140524	NaN	0.269207	NaN	
<b>min</b>	NaN	0.000000	0.000000	NaN	0.000000	NaN	
<b>25%</b>	NaN	0.000000	0.000000	NaN	0.000000	NaN	
<b>50%</b>	NaN	1.000000	0.000000	NaN	0.000000	NaN	
<b>75%</b>	NaN	3.000000	0.000000	NaN	0.000000	NaN	
<b>max</b>	NaN	31.000000	2.000000	NaN	6.000000	NaN	

In [190...]: # Modification de l'affichage de colonnes, lignes et largeurs de colonnes pour avoir un maximum d'information  
pd.set\_option('display.max\_columns', 200)  
pd.set\_option('display.max\_rows', 200)  
pd.set\_option('display.max\_colwidth', 50)

# à retenir : "none" value pour colwidth = illimité

In [191...]: # Modification de l'affichage de colonnes, lignes et largeurs de colonnes pour avoir un maximum d'information  
pd.set\_option('display.max\_columns', 200)  
pd.set\_option('display.max\_rows', 200)  
pd.set\_option('display.max\_colwidth', 50)

# à retenir : "none" value pour colwidth = illimité

myDF.head()

Out[191...]:

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	quantity	packagin
0	3087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir	NaN	1kg	Na
1	4530	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Banana Chips Sweetened (Whole)	NaN	NaN	Na
2	4559	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Peanuts	NaN	NaN	Na
3	16087	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055731	2017-03-09T10:35:31Z	1489055731	2017-03-09T10:35:31Z	Organic Salted Nut Mix	NaN	NaN	Na
4	16094	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055653	2017-03-09T10:34:13Z	1489055653	2017-03-09T10:34:13Z	Organic Polenta	NaN	NaN	Na

```
In [192]: myDF.head()
```

```
Out[192]:
```

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	quantity	packagin
0	3087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir		NaN	1kg
1	4530	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Banana Chips Sweetened (Whole)		NaN	NaN
2	4559	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Peanuts		NaN	NaN
3	16087	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055731	2017-03-09T10:35:31Z	1489055731	2017-03-09T10:35:31Z	Organic Salted Nut Mix		NaN	NaN
4	16094	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055653	2017-03-09T10:34:13Z	1489055653	2017-03-09T10:34:13Z	Organic Polenta		NaN	NaN

```
In [193]: myDF.dtypes
```

```
Out[193]:
```

code	object
url	object
creator	object
created_t	object
created_datetime	object
last_modified_t	object
last_modified_datetime	object
product_name	object
generic_name	object
quantity	object
packaging	object
packaging_tags	object
brands	object
brands_tags	object
categories	object
categories_tags	object
categories_fr	object
origins	object
origins_tags	object
manufacturing_places	object
manufacturing_places_tags	object
labels	object
labels_tags	object
labels_fr	object
emb_codes	object
emb_codes_tags	object
first_packaging_code_geo	object
cities	object
cities_tags	object
purchase_places	object
stores	object
countries	object
countries_tags	object
countries_fr	object
ingredients_text	object
allergens	object
allergens_fr	object
traces	object
traces_tags	object
traces_fr	object
serving_size	object
no_nutrients	float64
additives_n	float64
additives	object
additives_tags	object
additives_fr	object
ingredients_from_palm_oil_n	float64
ingredients_from_palm_oil	float64
ingredients_from_palm_oil_tags	object
ingredients_that_may_be_from_palm_oil_n	float64
ingredients_that_may_be_from_palm_oil	float64
ingredients_that_may_be_from_palm_oil_tags	object
nutrition_grade_uk	float64
nutrition_grade_fr	object
pnns_groups_1	object
pnns_groups_2	object
states	object
states_tags	object
states_fr	object
main_category	object
main_category_fr	object
image_url	object

```
image_small_url          object
energy_100g              float64
energy-from-fat_100g      float64
fat_100g                 float64
saturated-fat_100g       float64
butyric-acid_100g        float64
caproic-acid_100g        float64
caprylic-acid_100g       float64
capric-acid_100g         float64
lauric-acid_100g         float64
myristic-acid_100g       float64
palmitic-acid_100g       float64
stearic-acid_100g        float64
arachidic-acid_100g      float64
 behenic-acid_100g        float64
 lignoceric-acid_100g     float64
 cerotic-acid_100g        float64
 montanic-acid_100g       float64
 melissic-acid_100g       float64
 monounsaturated-fat_100g float64
 polyunsaturated-fat_100g float64
 omega-3-fat_100g         float64
 alpha-linolenic-acid_100g float64
 eicosapentaenoic-acid_100g float64
 docosahexaenoic-acid_100g float64
 omega-6-fat_100g          float64
 linoleic-acid_100g        float64
 arachidonic-acid_100g     float64
 gamma-linolenic-acid_100g float64
 dihomo-gamma-linolenic-acid_100g float64
 omega-9-fat_100g          float64
 oleic-acid_100g           float64
 elaidic-acid_100g         float64
 gondoic-acid_100g         float64
 mead-acid_100g            float64
 erucic-acid_100g          float64
 nervonic-acid_100g        float64
 trans-fat_100g             float64
 cholesterol_100g           float64
 carbohydrates_100g        float64
 sugars_100g                float64
 sucrose_100g               float64
 glucose_100g                float64
 fructose_100g               float64
 lactose_100g                float64
 maltose_100g                float64
 maltodextrins_100g          float64
 starch_100g                 float64
 polyols_100g                float64
 fiber_100g                  float64
 proteins_100g                float64
 casein_100g                  float64
 serum-proteins_100g          float64
 nucleotides_100g             float64
 salt_100g                   float64
 sodium_100g                  float64
 alcohol_100g                  float64
 vitamin-a_100g                float64
 beta-carotene_100g           float64
 vitamin-d_100g                 float64
 vitamin-e_100g                 float64
 vitamin-k_100g                 float64
 vitamin-c_100g                 float64
 vitamin-b1_100g                float64
 vitamin-b2_100g                float64
 vitamin-pp_100g                float64
 vitamin-b6_100g                float64
 vitamin-b9_100g                float64
 folates_100g                 float64
 vitamin-b12_100g               float64
 biotin_100g                  float64
 pantothenic-acid_100g          float64
 silica_100g                  float64
 bicarbonate_100g              float64
 potassium_100g                float64
 chloride_100g                 float64
 calcium_100g                  float64
 phosphorus_100g               float64
 iron_100g                    float64
 magnesium_100g                float64
 zinc_100g                     float64
 copper_100g                   float64
 manganese_100g                 float64
 fluoride_100g                  float64
 selenium_100g                 float64
 chromium_100g                  float64
 molybdenum_100g                float64
 iodine_100g                   float64
 caffeine_100g                  float64
 taurine_100g                  float64
 ph_100g                      float64
 fruits-vegetables-nuts_100g      float64
 collagen-meat-protein-ratio_100g float64
 cocoa_100g                     float64
 chlorophyl_100g                 float64
 carbon-footprint_100g             float64
 nutrition-score-fr_100g            float64
 nutrition-score-uk_100g             float64
 glycemic-index_100g              float64
 water-hardness_100g               float64
dtype: object
```

Les types de données ont l'air correct. \_100g et \_serving désignent une quantité, donc float64. (pas de colonne en \_serving d'ailleurs)

Je vais regarder la liste des colonnes sur lesquelles le chargement signalait une erreur de type.

```
myDF.iloc[ :, [0,3,5,19,20,24,25,26,27,28,35,36,37,38,39,48]]
```

Out[194...]	code	created_t	last_modified_t	manufacturing_places	manufacturing_places_tags	emb_codes	emb_codes_tags	first_packaging_code_geo	cities	cities_tags	al
0	3087	1474103866	1474103893	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	4530	1489069957	1489069957	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4559	1489069957	1489069957	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	16087	1489055731	1489055731	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	16094	1489055653	1489055653	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
320767	9948282780603	1490631299	1491244498	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
320768	99567453	1489059076	1491244499	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
320769	9970229501521	1422099377	1491244499	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
320770	9980282863788	1492340089	1492340089	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
320771	999990026839	1489072709	1491244499	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

320772 rows × 16 columns

```
In [195...]: type(myColTypes)
```

```
Out[195...]: pandas.core.series.Series
```

```
In [196...]: myColTypes[[0,3,5,19,20,24,25,26,27,28,35,36,37,38,39,48]]
```

```
Out[196...]: code          object  
created_t        object  
last_modified_t   object  
manufacturing_places    object  
manufacturing_places_tags    object  
emb_codes        object  
emb_codes_tags   object  
first_packaging_code_geo  object  
cities           object  
cities_tags      object  
allergens        object  
allergens_fr     object  
traces           object  
traces_tags      object  
traces_fr        object  
ingredients_from_palm_oil_tags    object  
dtype: object
```

```
In [197...]: myDictCol = { c : str for c in (myColTypes[[0,3,5,19,20,24,25,26,27,28,35,36,37,38,39,48]].index) }  
#myDictCol['code'] = 'Int64'  
#myDictCol['created_t'] = 'Int64'  
myDictCol['last_modified_t'] = 'Int64'  
myDictCol
```

```
Out[197...]: {'code': str,  
'created_t': str,  
'last_modified_t': 'Int64',  
'manufacturing_places': str,  
'manufacturing_places_tags': str,  
'emb_codes': str,  
'emb_codes_tags': str,  
'first_packaging_code_geo': str,  
'cities': str,  
'cities_tags': str,  
'allergens': str,  
'allergens_fr': str,  
'traces': str,  
'traces_tags': str,  
'traces_fr': str,  
'ingredients_from_palm_oil_tags': str}
```

```
In [198...]: # myDF2 = pd.read_csv('./Datasets/fr.openfoodfacts.org.products.csv', sep ='\t', dtype = myDictCol)  
# print("Données chargées")  
  
# Je laisse tomber, je n'arrive pas a forcer un type lors du chargement.  
# apres discussion avec Anthony, cela ne semble pas très important.
```

```
In [199...]: # myDF2.dtypes
```

```
In [ ]:
```

```
In [200...]: (myDF["code"].unique() )
```

```
#List ici permet de montrer tout le contenu du array.

Out[200... array([3087, 4530, 4559, ..., '9970229501521', '9980282863788',
   '999990026839'], dtype=object)

In [201... list( myDF["manufacturing_places"].unique() )

Out[201... [nan,
 'Brossard Québec',
 'United Kingdom',
 'Brossard,Québec',
 'France,Avranches',
 'Etats-unis',
 'France',
 'Thailand',
 'Belgien',
 'Las Ventas de Retamosa,Toledo (provincia),Castilla-La Mancha,España',
 'Germany',
 'France,Limousin,87500,Saint-Yrieux',
 'France,87500',
 '87500,France',
 'Pays-Bas,Netherlands',
 'Royaume-Uni',
 'Royaume Uni',
 'Ireland',
 'UK',
 'Royaume-Uni,Irelande',
 'California,USA',
 'Italie',
 'Switzerland',
 'Canada',
 'United States',
 'Taiwan',
 'Avery Island, Luisiana, Estados Unidos',
 'Topco',
 'Eden Prairie,Minnesota,USA',
 'Denmark',
 'Indonesia',
 'Purchase,NY,USA',
 'NY,USA',
 'Mexico',
 'Italy',
 'USA',
 'Berkeley,CA,USA',
 'États-Unis',
 'Norwalk,CT',
 'Norwalk,CT,USA',
 'U.S.A.',
 'Luxembourg',
 'allemande',
 'Allemagne',
 'china',
 'Minneapolis,MN,USA',
 'Thailande',
 'Thailande',
 'Thai Agri Foods Public Company Limited,155/1 Moo 1,Theparak Rd,Samutprakarn,10540,Thailand',
 'Israel',
 'Californie',
 'Taiwan',
 'Autriche',
 'Ohio,USA',
 'Toledo,Ohio,USA',
 '07000 Privas,France',
 'Ohio,United States',
 'Grèce',
 '[IMPORTADOR],Las Ventas de Retamosa,Toledo (provincia),Castilla-La Mancha,España',
 'Importé de Grèce',
 'Griechenland',
 'Italien',
 'Crèmerie Soignon,Poitou-Charente,France',
 'The Netherlands',
 'Serbia',
 'Etats-Unis',
 'Plano,Texas,USA',
 'Texas,USA',
 'Belgique',
 'California,United States',
 'Danemark',
 'United States if America',
 "États-Unis d'Amérique",
 'Minnesota USA',
 'Poland',
 'Suisse',
 'Battle Creek,Michigan,USA',
 'Chicago,Illinois,USA',
 'Japan',
 'Philippines',
 'Westlake Village CA,California,USA',
 'Uruguay',
 'Walkers Shortbread Ltd Aberiour-on-Spey A838 9PD Scotland',
 'Ecosse',
 'Scotland',
 'the Netherlands',
 'Morocco',
 'Dominican Republic',
 '[ENVASADOR],República Dominicana,[DISTRIBUIDOR],Casarrubios del Monte,Toledo (provincia),Castilla-La Mancha,España',
 'Marlborough,Massachusetts',
 'U.S.A',
 'Springfield,USA',
 'US,California',
 'Agrón (parroquia),Ames,La Coruña (provincia),Galicia,España',
 'Agrón (parroquia), Ames, La Coruña (provincia), Galicia, España',
 'ETATS UNIS',
 'Etats Unis',
 'Tiptree,Essex,England,United Kingdom',
```

'Quetigny',  
'Royaume Uni,Ecosse',  
'[FABRICANTE], Ecuador, [DISTRIBUIDOR], Massanassa, Valencia (provincia), Comunidad Valenciana, España',  
'[FABRICANTE],Ecuador,[DISTRIBUIDOR],Massanassa, Valencia (provincia), Comunidad Valenciana, España',  
'Ravens Oak Dairy,Nantwich,Cheshire,England,United Kingdom',  
'Tyskland',  
'Sauce Crafters Inc.,3695 Interstate Parkway #2,Riviera Beach,Florida,33404,USA',  
'China',  
'The JM Smucker Company,Orrville,OH 44667,USA',  
'Angleterre',  
'Spain',  
'Oregon,USA',  
'Blainville,Québec',  
'Dieppe (N.-B.)',  
'more than one country',  
'Moray,United Kingdom',  
'Blaine,Washington,USA',  
'St-Donat (Québec)',  
'Boisbriand (Québec)',  
'Tägerwilen (Suisse)',  
'Ancaster,Ontario',  
'Québec,Canada',  
'Montréal,Québec',  
'Lévis,Québec',  
'Intario,Canada',  
'Dy-Hyacinthe,Québec',  
'Hong Kong',  
'Australia',  
'Saint-Antoine-de-Tilly (Québec)',  
'Turquie,Montréal',  
'Canada,Nustef Foods Ltd',  
'Brampton,Ontario',  
'St Hyacinthe (Quebec)',  
'Saint-Augustin-de-Desmaures (Québec)',  
'Scotland,UK',  
'Longeuil,QC,Markham,ON,Victoria,BC',  
'Québec Canada',  
'St-Hubert,Québec',  
'BC Canada',  
'Québec',  
'Mississauga ON Canada',  
'Ontario',  
'Mississauga ON,Quebec QC',  
'Saint-Léonard (Québec)',  
'Toronto,Ontario',  
'Don Mills,Ontario',  
'Toronto ON,CANADA',  
'Adstock,Qc,Canada',  
'Adstock,Québec,Canada',  
'Adstock Québec',  
'UK,Polen',  
'Англия',  
'united kingdom',  
'united kingdom,Andover,Hampshire,SP10 5AQ',  
'Southway,Andover,Hampshire,SP10 5AQ,United Kingdom',  
'Pologne',  
'EU',  
'Stockton,CA',  
'Maine,USA',  
'Foster Dairy Farms,Modesto,California,USA',  
'Vacaville,CA',  
'Bridgeville,Delaware,USA',  
'Jelly Belly Candy Compagny,One Jelly Belly Lane,Fairfield,CA 94533,USA',  
'MIN / É.-U.A.',  
'Peru',  
'Tamma Industrie Alimentari Di Capitanata Srl,Italy',  
'Niigata,Japan',  
'Luzern,Suisse',  
'Napa Valley,California,USA',  
'Cincinnati OH',  
'Corona CA',  
'Malaisie',  
'Écosse',  
'Hereford,Texas,United States',  
'Portugal',  
'Corée',  
'Korea',  
'Japan,Japon',  
'Finlande',  
'Winston-Salem,North Carolina',  
'isvire',  
'Porto Rico',  
'Heinfels,Tyrol,Austria',  
'Hong-Kong,Chine',  
'Denmark ',  
'Malaysia',  
'Charlotte,NC',  
'Chine',  
'中國製造',  
'chine',  
'Salem,Oregon,USA',  
'Chivas Brothers Ltd,Keith,Ecosse',  
'Modena,Italy',  
'Tuscany,Italy',  
'Brazil',  
'Grande-Bretagne',  
'Lynchburg,TN,USA',  
'Kenya,India',  
'Pays-Bas',  
'Sonoma,California,USA',  
'thailande',  
'Thailand,Thai World import & export co.',  
'Bangkok,Thailande',  
'Great Britain',  
'Loretto,KY,USA',  
'Fort Collins CO USA',  
'Wales,United Kingdom',

'Corée du Sud',  
'South Korea',  
'Tahiti',  
'Londonderry,New Hampshire',  
'New Jersey,Usa',  
'Biars,France',  
'Royaume-Uni,Angleterre',  
'Espagne',  
'Kenya',  
'Royaume-Uni,Écosse',  
'England,United Kingdom',  
"Fleury Michon Traiteur - FMT (Filiale Fleury Michon Solution Repas) - Parc d'Activité Polaris - 85110 Chantonnay,Vendée,Pays de la Loire,France",  
'Ottawa (Ontario)',  
'É. - U.A.',  
'Ottawa,ON',  
'Ontario,Canada',  
'É. - U.A.',  
'United States,Yuma az',  
'Costa Rica',  
'India,Kenya',  
'Utebo,Zaragoza (provincia),Aragón,España',  
'Belgium',  
'italy',  
'Colombia',  
'Pittsburg,United States',  
'Moulin de Beau-Rivage,Anjou,France',  
'England',  
'austria',  
"Les Délices de la Mer (Filiale Groupe Le Graët) - ZA des 4 Voies - 22170 Plélo,Côtes-d'Armor,Bretagne,France",  
'france',  
'germany',  
'France,Anjou',  
'South Africa',  
'Yorkshire,United Kingdom',  
'Tunisia',  
'Bretagne,France',  
'uk,wington vale creamery',  
'switzerland',  
'Forli,Italy',  
'portugal',  
'Lesserevil Brand Snack Co,CT,United States',  
'Hawaii,USA',  
'Pertshire,United kingdom',  
'Fife,United Kingdom',  
'STOFFLER ZONE INDUSTRIELLE BD D EUROPE BP 35 67210 OBERNAI CEDEX France,France',  
'St Louis,Missouri',  
'Sains,UK',  
'Rochester,New York,USA',  
'Livermore ça or USA ?',  
'USA,UK',  
'Saint-André-de-la-Marche,France',  
'Société des Affineurs Franc-Comtois Réunis - ZA Sur le Jura - 25690 Avoudrey,Doubs,Franche-Comté,France',  
'Dänemark',  
'Netherlands',  
'Pyrénéefrom (Filiale Groupe Lactalis),64120 Larceveau-Arros-Cibits,Pyrénées-Atlantiques,Aquitaine,France',  
'Auvergne',  
'Saint-Jacut-du-Mené,France',  
'Poligny,Jura,Bourgogne-Franche-Comté',  
'Colmar,Alsace,France',  
'France,Abattoir : Domvallier,Découpé : Alençon',  
"Abattoi,Groupe Bigard - ZI Grévaux-Les-Guides - Rue Daniel Gaillard - 59750 Feignies,Nord,Nord-Pas-de-Calais,Découpé,Groupe Bigard - Rue du 60ème Régiment d'Infanterie - 80470 Ailly-sur-Somme,Somme,Picardie,France",  
'Lescar,France',  
'Courtenay,Loiret,France',  
'Barjouville chartres,France',  
"Salaisons Celtilques (Filiale du Groupement Les Mousquetaires) - Parc d'Activités Tréhonin - 56300 Le Sourn,Morbihan,Bretagne,France",  
'Cantal,France',  
'Isère,France',  
'Villefort,Lozère,France',  
'France,Le Kremlin Bicêtre',  
'France,Vendée',  
'Saint-Flour',  
'France,Loué',  
'France,Sarthe',  
'LDC Sablé (Filiale Groupe Lambert Dodard Chancereul - LDC) - ZI Saint Laurent - 72300 Sablé-sur-Sarthe,Sarthe,Pays de la Loire,France',  
'LDC Bourgogne (Filiale Groupe Lambert Dodard Chancereul - LDC) - ZI des Marosses - 71500 Branges,Saône-et-Loire,Bourgogne,France',  
'Savoie,France',  
'Valencia,Espagne',  
"Fromapac (Filiale Bongrain) - Parc d'Activités des Rouyardières - 35220 Châteaubourg,Ille-et-Vilaine,Bretagne,France",  
'SOCIETE MORLAISIENNE SALAISONS LA LANDE DU CHENOT 56380 BEIGNON France,France',  
'DUC - Grande Rue - 89770 Chailley,Yonne,Bourgogne,France',  
'29150,France',  
'European Union',  
"Fromagerie Le Centurion - FLC (Filiale Sodagraal) - 9 Rue Blaise Pascal - ZI Les Portes du Nord - 62820 Libercourt,Pas-de-Calais,Nord-Pas-de-Calais,France",  
'Fermiers du Gers 32100 CONDOM',  
"GPVL (Groupement des Producteurs de Volailles de Licques) - 7 Rue de l'Abbé Pruvost - 62850 Licques,Pas-de-Calais,Nord-Pas-de-Calais,France",  
'Nord Pas-de-Calais,France',  
'Nord Pas-de-Calais Picardie,France',  
'France,Nord Pas-de-Calais Picardie',  
'Simply Orange Juice Company,Florida,USA',  
'Indiana,USA,Indianapolis',  
'Carvin,France',  
'Nord,France',  
'Croyle,United Kingdom',  
'Courrières,France',  
'France,Bretagne',  
'Royaume-Uni,France',  
'Nouvelle-Zélande,France',  
"Fromagerie Mulin (Filiale Sodagraal) - Lieu dit Champs Breland - 25170 Noironte,Doubs,Franche-Comté,France",  
"Les Délices d'Auzan [Filiale ITM Entreprises (Groupement Les Mousquetaires)] Lieu dit Domaine d'Archam - 32440 Castelnau-d'Auzan,Gers,Midi-Pyrénées,France",  
'France,Bretagne,Côtes-d'Armor",  
'Great-Britain',  
'Norway',  
'Conditionné par :Monts & Terroirs (Filiale Sodiaal International),Vevy,Jura,Franche-Comté,France',  
'France,Alsace',

"Ets Schoepfer - 7 Avenue Fontcouverte - Zone Industrielle - 84000 Avignon,Vaucluse,Provence-Alpes-Côte d'Azur,France",  
'Gard',  
'Biscuits Panier (Filiale Groupe Poult) - ZI des Pays-Bas - 29510 Briec,Finistère,Bretagne,France',  
'Jus de Fruits d'Alsace - JFA (Filiale Laiterie de Saint-Denis-de-l'Hôtel - LSDH) - Zone Industrielle Rimsdorf - 67260 Sarre-Union,Bas-Rhin,Alsace-Champagne-Ardenne-Lorraine,France',  
'Brasserie Castelain,62410,Bénifontaine,France',  
'Hershey,PA,USA',  
'Hershey,Pennsylvania,USA',  
'Quéven,Lorient (distrito),Morbihan,Bretaña,Francia',  
'Product of USA',  
'Missisauga,Ontario',  
'Austria',  
'Somerset,United Kingdom',  
'Somerset, United Kingdom',  
'Hampshire, England, United Kingdom',  
'CWA',  
'Saint Louis 974',  
'Markham ON',  
'Hiifoots Village,Alva,Ecosse',  
'Michigan,USA',  
'Londres',  
'Santa Monica,CA,USA',  
'90 Peffermill riad,Edinburgh,Écosse,Royaume-Uni',  
'France,Chambery',  
'Union Européenne',  
'Pays Bas',  
'Amsterdam,Pays Bas',  
'Pays-bas',  
'Pakistan',  
'La Baie',  
'Libanon',  
'Toronto (Ontario)',  
'Sainte-Cécile-de Milton (Québec)',  
'Edmonton,Alberta,Canada',  
'Longueuil (Québec)',  
'Région Parisienne,France',  
'Scottsdale Arizona',  
'Scotland, United Kingdom',  
'Scotland,United Kingdom',  
'Écosse,Scotland',  
'Orgaz (Toledo),Spain',  
'France,Normandie',  
'România',  
'Ecosse,Royaume-Uni',  
'Produit et emballé en Grèce,Importé par:The Tassos Group,Oakbrook Terrace,IL,É.-U.A.',  
'Santa Monica,CA,California,United Stated of America,USA',  
'Bethesda,MD,United States',  
'Plovdiv',  
'Ontario CDA',  
'Notre-Dame-de-Lourdes,Québec',  
'Ferme-Neuve,Québec,Canada',  
'Toronto,Ontario,Canada',  
'日本',  
'Norvège',  
'Zhaqing,Guangdong,China',  
'Montreal,Québec',  
'Manitoba CDN',  
'Watou,Belgium',  
'45730 Saint-Benoit-sur-Loire,Loiret,Centre-Val de Loire,France',  
'6 st Colme Street,Edinburgh,EH3 6AD,United Kingdom',  
'Emirats arabes unis',  
'Coombs Family Farms,Brattleboro,VT,05302,United States',  
'Watsonville,CA',  
'Boulder,Colorado,United States',  
'israel',  
'Product of France',  
'Ciudad Guadalupe,Nuevo León.',  
'Fort Bragg,California',  
'Santa Catarina,Nuevo León,Mexico',  
'6 St Colme Street,EH3 6AD,Edinburgh,United Kingdom',  
'Lebanon',  
'Freiburg,Deutschland',  
'Broomfield,Colorado,USA',  
'Oloron-Sainte-Marie,France',  
'Fremont,Ohio,USA',  
'India',  
'California, USA',  
'Aurora,USA',  
'Mondelez France SAS BP100 92146 CLAMART CEDEX',  
'Mississauga,Ontario',  
'Lou Gascoun (Filiale Financière Martin) - La Prade - Parc d'Activité de La Prade - 33650 Saint-Médard-d'Eyrans,Gironde,Aquitaine,France',  
'Cuisiné par Fleury Michon',  
'quebec,Saint-jean baptiste',  
'Saint-Jean-Baptiste,Québec,Canada',  
'Boisbriand,Québec',  
'Maroc',  
'Maryland,USA',  
'CA 91402,Los Angeles,USA',  
'Los Angeles,CA 91402,USA',  
'City of Industry,CA',  
'Carib Brewery Limited,Champs fleurs,Trinidad',  
'Estado de México,México',  
'Baie de San Francisco,San Francisco,Californie,Etats-Unis',  
'Au Canada à partir d'ingrédients canadiens et importés",  
'Yucatán,Mexico',  
'Chicago IL',  
'Las Vegas,Nevada,USA',  
'Carlifornia,USA',  
'Chestnut,Illinois,USA',  
'Vietnam',  
'Listowel,Ontario',  
'Équateur',  
'canada',  
'Argentina',  
"St-Théodore d'Acton (Québec)",  
'USA,Vancouver (BC),Acton (ON)',  
'Montréal (Québec)',

'Brooklyn,USA',  
'Long Beach,California,USA',  
'MEXIQUE',  
'Minnesota,United States',  
'California USA',  
'New York,USA,Astoria',  
'US',  
'Philippines',  
'Mountain View,CA',  
'Pennsylvania,USA',  
'Berkeley,CA',  
'Crowny Food Industries BV,Zaandam,Nederland',  
'Vancouver,BC',  
'Vancouver',  
'belgium',  
'Brasserie De Brabandere,8531,Bavikhove,Belgique',  
'Россия',  
'La Jolla,California,USA',  
'singapour',  
'China,USA',  
'Schweden',  
'Suède',  
'United kingdom',  
'Lima',  
'Seychelles',  
'Munnar,India',  
'Importé pour IKEA CANADA,BURLINGTON,ONTARIO',  
'Sweden',  
'Inde',  
'Northumberland,Pennsylvania',  
'Sonoma,California',  
'Fiorette Mâcon (Filiale Agrial) - 335 Rue des Pouilly Vinzelles - Espace Entreprises Loche - 71000 Mâcon,Saône-et-Loire,Bourgogne,France',  
'Finland',  
'A.Pitenis Bros S.A.,Grèce',  
'Alpina Savoie (Filiale Groupe Galapagos) - 209 Rue Aristide Bergès - 73000 Chambéry,Savoie,Rhône-Alpes,France',  
'MAS VEZIAN,66350,TOULOGUES',  
'Laiterie Grand'Ouche (Filiale Savencia Fromage & Dairy) - 20 Rue de la Grand'Ouche - 16200 Réparsac,Charente,Poitou-Charentes,France",  
'Saint-Paul en Jarez',  
'Lithuanie',  
'fr:Suède',  
'Alemania',  
'californie',  
'Turquie',  
'Agen',  
'France,Châteaudun, Eure-et-Loir',  
'пос. Сидоровка,г. Набережные Челны,Республика Татарстан,Россия',  
'De Katjang BV,Koekoeksedijk 14,4761 PJ,Sevenbergen,Nederland',  
'Société Fromagère de Sainte-Cécile (Filiale Groupe Lactalis) - L'Acherie - 50800 Sainte-Cécile,Manche,Basse-Normandie,France",  
'Вьетнам',  
'посёлок Новый Свет,Гатчинский район,Ленинградская область,Россия',  
'COCA COLA ENTREPRISE,27 RUE CAMILLE DESMOULINS,92130,ISSY LES MOULINEAUX,FRANCE',  
'Manantial Amer Palatin,Amer,Selva (comarca),Gerona (provincia),Cataluña,España',  
'Morón de la Frontera, Sevilla (provincia), Andalucía, España',  
'Coopérative Isigny Sainte Mère,14230 Isigny-sur-Mer,Calvados,Basse-Normandie,France,Normandie',  
'Manantial Sigüenza,Moratilla de Henares,Sigüenza,Guadalajara (provincia),Castilla-La Mancha,España',  
'Valdastillas,Cáceres (provincia),Extremadura,España',  
'Estepa,Sevilla (provincia),Andalucía,España',  
'La Guardia,Toledo (provincia),Castilla-La Mancha,España',  
'Logroño, La Rioja, España',  
'Villamalea,Albacete (provincia),Castilla-La Mancha,España',  
'Tercia (pedanía),Lorca,Murcia (comunidad autónoma),España',  
'La Fromagerie de la Lémance - ZA du Haut Agenais - 47500 Montayral,Lot-et-Garonne,Aquitaine,France',  
'Játiva, Valencia (provincia), Comunidad Valenciana, España',  
'Charente,France',  
'Pâtisseries Gourmandes (Filiale Groupe Roullier) - Le Bourg - 56540 Saint-Tugdual,Morbihan,Bretagne,France',  
'Arróniz, Navarra, España',  
'Villaconejos,Madrid (comunidad autónoma),España',  
'Motril,Granada (provincia),Andalucía,España',  
'Société Laitière de Vitré (Filiale Groupe Lactalis) - Les Guichardières - 35500 Vitré,Ille-et-Vilaine,Bretagne,France',  
'Yoplait France (Filiale General Mills/Sodiaal International) - 49 Route d'Auxerre - 89470 Monéteau,Yonne,Bourgogne,France",  
'Benifairó de les Valls,Valencia (provincia),Comunidad Valenciana,España',  
'Berlin,Alemania',  
'Antequera,Málaga (provincia),Andalucía,España',  
'La Calahorra,Granada (provincia),Andalucía,España',  
'Wiesbaum,Vulkaneifel (distrito),Renania-Palatinado,Alemania',  
'Pfarrkirchen,Rottal-Inn (distrito),Baviera,Alemania',  
'Roding-Neubäu,Roding,Cham (distrito),Baviera,Alemania',  
'[DISTRIBUIDOR],Oberkohlstätten,Unterkohlstätten,Oberwart (distrito),Burgenland,Austria',  
'Halle (Sajonia-Anhalt), Sajonia-Anhalt (estado), Alemania',  
'Lepe, Huelva (provincia), Andalucía, España',  
'Société Coopérative Agricole des Producteurs de Reblochons de la Vallée de Thônes (Sté Coop Agric Product Reblochon Vall) - Route d'Annecy - 74230 Th ônes,Haute-Savoie,Rhône-Alpes,France",  
'Etablissements Paul Paulet - EPP (Filiale MWBrands) - 4 Rue du Pont Dinou - ZI de Pouldavid - 29100 Douarnenez,Finistère,Bretagne,France',  
'Soulisé Restauration (Filiale William Saurin) - 392 Rue des Artisans - ZA Bel Air - 12000 Rodez,Aveyron,Midi-Pyrénées,France',  
'Etablissements Berni et Cie - Rue du Port Sec - ZI de Tavannes - 55100 Verdun,Meuse,Lorraine,France',  
'Yffiniac,Côtes d'Armor,Bretagne,France",  
'[FABRICANTE],Italia,[DISTRIBUIDOR],Mülheim an der Ruhr,Düsseldorf (región),Renania del Norte-Westfalia,Alemania',  
'Salerno,Italy,campania',  
'Normandie',  
'grèce',  
'Arpajon',  
'Fabriqué en Anjou,par,Marcillat,à,Saint-Clément,Maine-et-Loire,Pays de la Loire,France',  
'Rumilly,France',  
'Suisse,Laufen',  
'Rumilly,Haute-Savoie,France',  
'Maltagliati SpA - Via Puccini 49 - 51016 Montecatini Terme,(PT : Pistoia),Toscane,Italie',  
'Deville les Rouen,Val de Saône,Saint-Martin de Londres',  
'Queyssac,Dordogne',  
'Evian,France',  
'Athis-Mons,France',  
'Vittel,France',  
'Hamal Signature - Steenweg op Tielen 53 - 2300 Turnhout,Province d'Anvers,Région flamande,Belgique",  
'Marie Surgelés - MS (Filiale Groupe Lambert Dodard Chancereul - LDC) - Zone Economique de Coquine - Le Pont de Barroux - 79600 Airvault,Deux-Sèvres,Poitou-Charentes,France',  
'Senoble Desserts Premium (Filiale Senoble) - 9 Rue Galilée - ZAC de Belle Aire - 17440 Aytré,Charente-Maritime,Poitou-Charentes,France',  
'Allemagne,Unterhaching',  
'[ENVASADOR],Halle (Sajonia-Anhalt),Sajonia-Anhalt (estado),Alemania',

'Autol,La Rioja,España',  
'Alsace',  
'Montbrun,Lozère,France',  
'Le Pompidou,Lozère,France',  
'La Hoya (pedanía), Lorca, Murcia (comunidad autónoma), España',  
'[DISTRIBUIDOR],Alemania',  
'Waldbaur Feine Schokoladen GmbH,Stollwerck GmbH,Köln,Köln (Regierungsbezirk),Nordrhein-Westfalen,Deutschland',  
'Senagral (Filial Senoble) - Avenue des Saulniers - Zone Industrielle - 57170 Château-Salins,Moselle,Lorraine,France',  
'Estepa, Sevilla (provincia), Andalucía, España',  
'Lidl',  
'Orgaz, Toledo (provincia), Castilla-La Mancha, España',  
'Italia parma',  
'Lebor (pedanía), Totana, Murcia (comunidad autónoma), España',  
'Antequera, Málaga (provincia), Andalucía, España',  
'france 45 la ferte st aubin',  
'[DISTRIBUIDOR],Candreita,Navarra,España',  
'San Adrián,Navarra,España',  
'London,England,United Kingdom',  
'Dakar,Senegal',  
'italia',  
'[DISTRIBUIDOR], Candreita, Navarra, España',  
'Italia,[DISTRIBUIDOR],Alemania',  
'Bamako',  
'Sri Lanka',  
'ITALY',  
'Sarthe,France',  
'italie',  
'Murcia',  
'28310 Janville,France,SIRET:50292362600019',  
'La Ferme de Sigy,chemin départemental 62,77520 Sigy,France,SIRET:79448186100019',  
'622310 Créquy,France',  
'[FABRICANTE],Cañada, Alicante (provincia), Comunidad Valenciana, España',  
'Casa Taradellas Barcelona',  
'19500 Curemonte,Corrèze,France',  
"Société Fromagère de Jort,14170 Bernières d'Ailly,Normandie,France",  
'Fromagerie du Val-de-Sienne,50450 Gavray,Normandie,France',  
'03530 La Nucia',  
'61120 Camembert,Normandie,France',  
'Chine,CN',  
'Fiorette Food Service,66440,Torreilles',  
'LOZERE',  
'Bockum-Hövel',  
'TOGO',  
'Lome TOGO',  
'22120,Quesoy,France',  
'Grande Bretagne',  
'Ardèche',  
'E.A. SAS,Annecy,74000',  
'Tarn,Midi-Pyrnes,France',  
'Italie,France',  
"Les Salaisons de l'Arrée - Parc d'Entreprises - 29690 Brennilis,Finistère,Bretagne,France",  
'Bialystok',  
'Sant Gregori, Gerona (provincia), Cataluña, España',  
'Benisanó, Valencia (provincia), Comunidad Valenciana, España',  
'Point Lobos,California,USA,Monterey',  
'Pulpí, Almería (provincia), Andalucía, España',  
'Valdastillas, Cáceres (provincia), Extremadura, España',  
'Narros de Cuéllar,Samboal,Segovia (provincia),Castilla y León,España',  
'Abarán, Murcia (comunidad autónoma), España',  
'Salem,Tamil Nadu,India',  
'España',  
'Villafranca,Navarra,España',  
'Portland,USA',  
'Paris,France',  
'San Francisco,USA',  
'Slovaquie',  
'Reykjavik,Islande',  
'Deutschland,Stuttgart',  
'Strasbourg,France',  
'BK - 67210 - Obernai,France',  
'Scamark,94859,Ivry sur Seine,France',  
'Berlin,Deutschland',  
'Frima NV - Napoleonlaan 16 - 8400 Ostende,Province de Flandre-Occidentale,Région flamande,Belgique',  
'BK,67200,Strasbourg,France',  
'InBev Belgium s.a.,1070,Bruxelles,Belgique',  
'Le Mans',  
'Cacolac,33850,Léognan,France',  
'BK,67200,Strasbourg,France',  
'Refresco France 2885 Route des Pangons 26260 Margès,France',  
'Mendavia,Navarra,España',  
'Molkerei Gropper GmbH & Co. KG Am Mühlberg 2 D-86657 BISSINGEN Allemagne',  
'SUISSE LINDT&SPRÜNGLI CH-8802 KILCHBERG SUISSE',  
'Groupe Fructa Partner 119 rue Michel Aulas F-69400 Limas,France,belgique',  
'France,Provence',  
'Vilassar de Mar,Barcelona (provincia),Cataluña,España',  
'PAYS BAS,COROOIS Conserven B.V. 4420 AC Kapelle',  
'Ruoms',  
'ALLEMAGNE,Molkerei Gropper GMBH & CO KG Am Mühlberg 2 -D-86657 BISSINGEN',  
'Saint Pierre sur Dives,Calvados,France',  
'Moosehead Breweries Ltd,Saint John,New Brunswick',  
'Novelda,Alicante (provincia),Comunidad Valenciana,España',  
'[ENVASADOR],Greven,Distrito de Steinfurt,Renania del Norte-Westfalia,Alemania',  
'Deutschland,Nürnberg',  
'Villé-Morgon,France',  
'Deutschland',  
'[FABRICANTE],Alemania,[DISTRIBUIDOR],Hamburgo,Alemania',  
'Bedarrides,France',  
'Autol, La Rioja, España',  
'Turkey',  
'Istanbul,Turkey',  
'Eskişehir,Turkey',  
'Belgique,Retinne',  
'spain',  
'France,Bias',  
'Sant Pere Pescador, Gerona (provincia), Cataluña, España',  
'Bremen, Bremen (estado), Alemania',  
'Flandes,Bélgica',

'Lima Perú',  
'Dakar',  
'Cameroun',  
'Ruoms,Ardèche,France',  
'Rully,France',  
'Meloisey,France',  
'Chagny,Bourgogne,France',  
'Cave de Saint Cyr en Bourg',  
'Vilassar de Mar, Barcelona (provincia), Cataluña, España',  
'Andalousie',  
'Domagné,Bretagne,france',  
'tunisie',  
'Brooklyn,NY 11211',  
'Brooklyn,NY 11211,USA',  
'Plouvier',  
'Corrèze,France',  
'Türkîye',  
'Ankara,Türkîye',  
'İstanbul Turkey',  
'REMOND Didier,05150,Montmorin,France',  
'Algérie',  
'Saint Grégoire',  
'Deutschland,Langenburg',  
'Irland,Deutschland',  
'68350 BRUNSTATT',  
'35230 Bourgbarré',  
'Oliveira de Azeméis,Portugal',  
'provence,France',  
'город Ижевск,Удмуртская Республика,Россия',  
'Soup' idéale - 223 Allée d'Allemagne - Zone Artoipôle - 62223 Feuchy,Pas-de-Calais,Nord-Pas-de-Calais,France",  
'Mesnay Jura France',  
'Mesnay',  
'3 chemin de la sandlach 67501 haguenau cedex',  
'Grasse,France',  
'Voiron,France',  
'France,Grasse-France',  
'Greece',  
'Augsburg,D-86179,Allemagne',  
'France,84700',  
'Roumanie,Județul Mureș',  
'França',  
'France,Pas-de-Calais',  
'Olhão,Portugal',  
'Iran',  
'FRANCE',  
'Roquefort',  
'Roquefort,France',  
'Россия,Чехов,Московская область',  
'Marie (Filiale Groupe Lambert Dodard Chancereul - LDC) - ZI de l'Aubrée - 72300 Sablé-sur-Sarthe,Sarthe,Pays de la Loire,France",  
'Lou Gascoun-Mercadier La Prade 33650 Saint-Ménard d'Eyrans,France",  
'Freiburg',  
'Les Courgons,04140,Provence-Alpes-Côte d'Azur,SELONNET',  
'France,83140,Var',  
'France,Sagy,95450',  
[http://www.episaveurs.fr/produit/epicerie/produit\\_0004596.html](http://www.episaveurs.fr/produit/epicerie/produit_0004596.html)',  
'42008 St Etienne',  
'Camembert',  
'GERMANY (Allemagne) - BERLIN',  
'Angleterre,Royaume-Uni',  
'Berlin',  
'PC@CP,CCCP,Россия',  
'Produit élaboré en France.',  
'France,Rennes,Bretagne',  
'France,Thank you my deer',  
'Indre-et-Loire,37160',  
'France,Clisson',  
'Asturias',  
'Le Sapapet SARL Produits laitiers La Combe 1658 Rossinière VD Suisse',  
'Société de Fromagerie du Brassus Le Rocher 4 1348 Le Brassus VD Suisse',  
'Hongrie',  
'Москва,СССР,РСФСР',  
'inconnue',  
'Tailandia',  
'France,Nimes',  
'Denmark,Tunisia,Malta',  
'Oosterhout,The Netherlands',  
'BATOOK CHEWING GUM - SAUDI ARABIA',  
'BATOOK CHEWING GUM',  
'Saint-Jean-sur-Richelieu',  
'france,italie',  
'Belgica',  
'Grecia ?',  
'Eskişehir Türkiye',  
'Karaman,Türkiye',  
'Dudullu,İstanbul,Türkiye',  
'Ümraniye',  
'Idaho Falls,U.S.A.',  
'Neu-Isenburg,Allemagne',  
'Cuba',  
'France,Saint-Nazaire',  
'Lituanie',  
'75009 PARIS',  
'Saint Martin La Plaine,Loire,France',  
'W.S,f-77407 lagny-sur-marne cedex',  
'UE',  
'Japon',  
'Jandu,Strasbourg',  
'Lyon',  
'Corbas',  
'Lyon,France',  
'Saint Jean de Monts',  
'Aveyron,France',  
'France,30460,Lasalle-en-Cévennes',  
'Amsterdam,Pays-Bas,Becky's BV",  
'Rausch Schokoladen GmbH,Peine,Peine (Landkreis),Niedersachsen,Deutschland',  
'Union européenne',  
'Zanetti SpA,Via Madonna 1,24040,Lallio,Italy',

'Copack Tiefkühlkost Produktionges.mbH - Am Lunedeich 116 - 27572 Bremerhaven,Brême,Greenland Seafood Wilhelmshaven GmbH - Flutstraße 84 - 26386 Wilhelmshaven,Basse-Saxe,Allemagne',  
'Greenland Seafood Wilhelmshaven GmbH - Flutstraße 84 - 26386 Wilhelmshaven,Basse-Saxe,Allemagne',  
'Allemagne,T.M.A. GmbH An den Breiten D-01454 - Leppersdorf',  
'Privatmolkerei Bechtel, Molkereistrasse 5, D-92521 Schwarzenfeld',  
'Aurich,Niedersachsen,Deutschland',  
'Deutschland,Schwäbisch Hall',  
'Bourges,France',  
'AF Industrie Alimentari SpA - Loc Incoronata Zona ASI,71100 Foggia,(FG : Foggia),Pouilles,Italie',  
'Nürnberger Wein Kellerei GmbH,90451,Nürnberg,Deutschland',  
'Kentucky,Etats-Unis,USA',  
'Übach-Palenberg,Heinsberg (Kreis),Köln (Regierungsbezirk),Nordrhein-Westfalen,Deutschland',  
'Alemania',  
'Hochland Deutschland GmbH - Kemptener Straße 17 - 88178 Heimenkirch, Bavière, Allemagne',  
'Valle Lomellina,Italie',  
'41012 CARPI Italie',  
'Géo - Zone d'Activités Nord - ZA Ablis Nord - 78660 Ablis,Yvelines,Île-de-France,France",  
'Landshut,Allemagne',  
'Puente Genil,Córdoba (provincia),Andalucía,España',  
'Porto,Portugal',  
'cote du rhone',  
'44140,Le Bignon,France',  
'France,Saint-Hippolyte',  
'Normandie,France',  
'Lactalis Nestlé Ultra-Frais Laval - LNUF Laval (Filiale Groupe Lactalis) - Boulevard Arago - ZI des Touches - 53810 Changé,Mayenne,Pays de la Loire,France',  
'Dolcissimo Srl - Via Fratelli Cervi 41 - 20010 Ossona,(MI : Milan),Lombardie,Italie',  
'Senagral 89150 Jouy,France',  
'LAMPAULaise DES SALAISONS,LDS SAS,29403 LANDIVISIAU,Lampaul-Guimiliau,BRETAGNE,FINISTERE,France',  
'La Lampaulaise de Salaisons - 1 Rue Galilée - ZA de Kerourvois Sud - 29500 Ergué-Gabéric,Finistère,Bretagne,France',  
'Société des Rillettes Bahier (SDRB) - Route de Paris - D 323 - 72160 Sceaux-sur-Huisne,Sarthe,Pays de la Loire,France',  
'Sceaux sur Huisne',  
'Aoste SNC - A SNC (Filière Campofrio Food Group) - ZI Les Jacquards - 42520 Maclas,Loire,Rhône-Alpes,France',  
'Canelia Petit Fayt Beurre (Filière Groupe Lactalis) - 49 Rue du Village - 59244 Petit-Fayt,Nord,Nord-Pas-de-Calais,Société Laitière de Rodez (Filière Groupe Lactalis) - Rue de la Prade - ZI de Cantaranne - 12850 Onet-le-Château,Aveyron,Midi-Pyrénées,Société Laitière Montauban (Filière Groupe Lactalis) - 25 Impasse de Maastricht - ZI Albasud - 82000 Montauban,Tarn-et-Garonne,Midi-Pyrénées,France',  
'Martinet (Filière FPM) - 24 Rue du Limousin - Zone Industrielle - 38070 Saint-Quentin-Fallavier,Isère,Rhône-Alpes,France',  
'William Saurin (WS) - 26 Rue de Crécy - D35 - 02270 Pouilly-sur-Serre,Aisne,Picardie,France',  
'William Saurin (WS) - 81 à 89 Avenue du Général Leclerc - 77400 Saint-Thibault-des-Vignes,Seine-et-Marne,Île-de-France,France',  
'Fromageries Saint-Saviol (Filière Groupe Savencia Saveurs & Spécialités) - 86400 Saint-Saviol,Vienne,Poitou-Charentes,France',  
'Frankreich',  
'Compagnie des Fromages & Richesmonts - CF&R (Filière Sodiaal International) - 2 Rue de la Laiterie - 57670 Bénestroff,Moselle,Lorraine,France',  
'Fromagerie du Dauphiné (Filière Groupe Lactalis) - Le Village - Route Départementale 1092,38470 Têche,Isère,Rhône-Alpes,France',  
'FLS,12 bis boulevard Pont Achard,86000,Poitiers,France',  
'France,Poitou-Charentes',  
'Société Fromagère de Lons-le-Saunier (Filière Groupe Lactalis) - 39 Avenue Camille Prost - 39000 Lons-le-Saunier,Jura,Franche-Comté,France',  
'Société Fromagère de Lons le Saunier,39 Rue Camille Prost,39008,Lons,France',  
'Grécia',  
'Löningen,Germany',  
'[1],Osuna,Sevilla (provincia),Andalucía,España,[2],Alicante,Alicante (provincia),Comunidad Valenciana,España',  
'Morón de la Frontera,Sevilla (provincia),Andalucía,España',  
'Milch-Union Hocheifel eG (MUH) - Im Scheid 1 - 54597 Pronsfeld,Rhénanie-Palatinat,Allemagne',  
'Charcuterie Pierre Schmidt (CAPS) - 3 Rue du Ried - 67720 Weyersheim,Bas-Rhin,Alsace,France',  
'Orgaz,Toledo (provincia),Castilla-La Mancha,España',  
'Gádor, Almería (provincia), Andalucía, España',  
'Arroyo de la Encomienda, Valladolid (provincia), Castilla y León, España',  
'Espanha',  
'Aguilar de Campoo,Palencia (provincia),Castilla y León,España',  
'70230 Loulans-Verchamp,Haute-Saône,Bourgogne-Franche-Comté,France',  
'Delpierre-Mer et Tradition - DMT (Filière Continentale Nutrition) - 16 Boulevard de l'Europe - Parc d'Activités de Landacres - 62360 Hesdin-l'Abbé,Pas-de-Calais,Nord-Pas-de-Calais,France',  
'Villalonga,Valencia (provincia),Comunidad Valenciana,España',  
'Duc (Filière Groupe Duc) Château de Flamboin 77114 Gouaix,Seine et Marne,Île de France,France,89771,Chailly',  
'Socopa Viandes (Filière Groupe Bigard) - Lieu dit Les Bordes - D323 - 72400 Cherré,Sarthe,Pays de la Loire,France',  
'Allemagne,Westfalenland Fleischwaren GmbH Hessenweg 2 - DE-48157 Münster',  
'Strasbourg',  
'Lactalis Nestlé Ultra-Frais MDD - (Filière Groupe Lactalis) - ZA Les Placis - 35230 Bourgarré,FR 47.185.001 CE',  
'Solaried (Filière Laiterie des Ardennes) - Recogne - Rue de Saint-Hubert 75 - 6800 Libramont-Chevigny,Province de Luxembourg,Région wallonne,Belgique,Allemagne,FRANCE',  
'Lidl Stiftung & Co,Neckarsulm',  
'Arguedas, Navarra, España',  
'Allemagne,Molkerei Gropper GMBH & CO. KG Am Mühlberg 2 D-86657 Bissingen',  
'Cap Océan,rue des alizés,14790 VERNON France',  
'Delabli (Filière Labeyrie Fine Foods) - 369 Rue des Artisans - Zone d'Activités - 14670 Troarn,Calvados,Basse-Normandie,France",  
'Allemagne,Espagne',  
'Société Fromagère de Craon,53400,Craon,France',  
'Royal Greenland Seafood GmbH - Flutstraße 84 - 26386 Wilhelmshaven,Basse-Saxe,Allemagne',  
'Saint Pierre du Mont,Landes',  
'Wiefelstede-Dringenburg',  
'Coopérative Isigny Sainte Mère,14230 Isigny-sur-Mer,Calvados,Basse-Normandie,France',  
'Spanien',  
'с. Староселье,Брянский район,Брянская область,Россия',  
'д. Староселье,Брянский р-н,Брянская обл.,РФ,241525',  
'п. Молкомбината,г. Рудня,Российская Федерация,Смоленская область',  
'поселок Краснолит,г. Краснодар,Россия',  
'с. Кумовское,Бабынинский район,Калужская область,Российская Федерация',  
'France,Loire-Atlantique',  
'Hollande,Pays-Bas',  
'France,Tarn-et-Garonne',  
'Limoges',  
'Allemagne,Molkerei Gropper GmbH & Co KG Am Mühlberg 2 D-86657 Bissingen',  
'Nahe,Deutschland',  
'Borgo Vercelli,Vercelli (provincia),Piamonte,Italia',  
'Tops Foods NV - Lammerdries-Oost 26 - 2250 Olen,Province d'Anvers,Région flamande,Belgique",  
'Brie de l'Ordet,France",  
'La Compagnie du Biscuit (Filière Groupe Biscuits Bouvard) - 69 Avenue Jean Jaurès - 33600 Pessac,Gironde,Aquitaine-Limousin-Poitou-Charentes,France',  
'Graz,Österreich',  
'Aisne,Picardie,France',  
'Deutschland,Simbach am Inn,Antersdorfer Mühle',  
'Polonia,Poland,towicz',  
'Privatmolkerei Bechtel Molkereistrasse 5,D-92521 Schwarzenfeld,Allemagne',  
'Niederlande',  
'Société Fromagère d'Orbec (Filière Groupe Lactalis) - 8 Rue de Vimoutier - 14290 Orbec,Calvados,Basse-Normandie,France",  
'Crailsheim,Allemagne',  
'Loyer Woessen SA - Rue Saint-Patrick - Parc d'Activités - 59133 Phalempin,Nord,Nord-Pas-de-Calais,France",  
'Beuralia (Filière Sodiaal International) - 18 Rue de la Charme - 63100 Clermont-Ferrand,Puy-de-Dôme,Auvergne,France',  
'France,Villefranche d'Allier",

'France,Roland Monterrat SAS - ZA Sud - F-01570 Fellens',  
'Buss Fertiggerichte GmbH - Am Damm 26 - 28870 Ottersberg,Basse-Saxe,Allemagne',  
'DE EV 3 EG',  
'Murcia,lote AA13',  
'Grecia',  
'67723 HOERDT,Aisace lait',  
'Mexico City,Mexico',  
'Princes Industrie Alimentari Srl - Corso Vittorio Emanuele II.23 - 84012 Angri,(SA : Salerne),Campanie,Italie',  
'Arguedas,Navarra,España',  
'France,Rodez',  
'PAYS-BAS,COROOS Conserven 4420 AV Kapelle',  
'Novandie (Filiale Groupe Andros) - Rue de la Gravelle - 35560 Marcillé-Raoul,Illle-et-Vilaine,Bretagne,Novandie [Filiale Andros et Compagnie (Groupe A  
ndros)] - 19 Rue de la République - 76150 Maromme,Seine-Maritime,Haute-Normandie,Novandie (Filiale Groupe Andros) - Route de Roinville - 28700 Auneau,E  
ure-et-Loir,Centre-Val de Loire,France',  
'город Набережные Челны,Республика Татарстан,Россия',  
'Malgara Chiari & Forti SpA - Localita' Campagnole 1 Borghetto - 38063 Avio,(TN : Trente),Trentin-Haut-Adige,Italie',  
'Deutschland,Breuberg',  
'56110 Gourin,Morbihan,Bretagne,France',  
'FAO 67,FAO 61',  
'Lactalis Nestlé Ultra-Frais Bayeux - LNUF Bayeux (Filiale Groupe Lactalys) - Rue Jean Mermoz - ZAC des Longchamps - 14400 Saint-Martin-des-Entrées,Ca  
lvados,Basse-Normandie,France',  
'Compagnie Générale de Produits Alimentaires Peny - CGPA [Filiale Union Fermière Morbihannaise - UFM (Filiale Groupe CECAB)] - Pont Hellec - 29380 Sai  
nt-Thurien,Finistère,Bretagne,France',  
'[DISTRIBUIDOR],Milagro, Navarra, España',  
'Brasserie Duyck,59144,Jenlain,France',  
'Österreich',  
'Duitsland',  
'Delpeyrat (DSAS) [Filiale Groupe Maisadour (Holding MVVH)] - 150 Rue de la Téoulère - ZA Téoulère - 40280 Saint-Pierre-du-Mont,Landes,Aquitaine,Franc  
e',  
'Princes Industrie Alimentari S.R.L. C.so Vittorio Emanuele II-23 -84012 ANGRI (SA) - Italie',  
'Société Aubret (Filiale Groupe d'aucy) - ZI Route de Maumusson - 44540 Saint-Mars-la-Jaille,Loire-Atlantique,Pays de la Loire,France',  
'Allemagne,Basse-Saxe',  
'Fromageries Terres d'Or - 17 Rue des Varennes,71140 Bourbon-Lancy,Saône-et-Loire,Bourgogne,France",  
'Tierras del Condado SL,Ctra Almonte El Rocio km 22.8,Finca La cañada,21730,Huelva,Spain',  
'Barcelona,Barcelona (provincia),Cataluña,España',  
'Italia',  
'Montélimar',  
'Totana, Murcia (comunidad autónoma), España',  
'Garrel, Allemagne',  
'Covi (Groupe Covi) - 341 Route de Clisson - 44230 Saint-Sébastien-sur-Loire,Loire-Atlantique,Pays de la Loire,France',  
'Société Nouvelle Laiterie de la Montagne - Saillant - 63710 Saint-Nectaire,Puy-de-Dôme,Auvergne,France',  
'Les 3 Moulins,Vitaprim,la Brosse Tenaud,44310,Saint Philibert de grand Lieu,France',  
'Villafranca del Bierzo,León (provincia),Castilla y León,España',  
'Biosabor SAT,Ctra San Isidro,Campohermoso,Níjar,Almería,Espagne',  
'[ENVASADOR],La Murada, Alicante (provincia), Comunidad Valenciana, España',  
'88148,Bulgneville,France',  
'Estubeny,Valencia (provincia),Comunidad Valenciana,España',  
'El Palmar (pedanía), Murcia, Murcia (comunidad autónoma), España',  
'Alcarrás, Lérida (provincia), Cataluña, España',  
'Borox, Toledo (provincia), Castilla-La Mancha, España',  
'Águilas, Murcia (comunidad autónoma), España',  
'Vícár,Almería (provincia),Andalucía,España',  
'Terrea,13210,Saint Remy de Provence,France',  
'inconnu',  
'Millagro,Navarra,España',  
'Oliva,Valencia (provincia),Comunidad Valenciana,España',  
'Nouvelle Fromagerie de Vaudes - 5 Grande Rue,10260 Vaudes,Aube,Champagne-Ardenne,France',  
'Laiterie de Bresse (Filiale Laiteries H. Triballat - LHT) - 60 Route des Charmettes - 71480 Varennes-Saint-Sauveur,Saône-et-Loire,Bourgogne,France',  
'Jouffruit,84300,Cavaillon,France',  
'Les 3 Moulins Vitaprim,La Brosse Tenaud,44130,St Philibert de Grandlieu,France',  
'Devos Group BVBA,3460,Bekkevoort,Belgium',  
'El Ejido,Almería (provincia),Andalucía,España',  
'Sanchonuño,Segovia (provincia),Castilla y León,España',  
'Schwarzenfeld,Alemania',  
'Fraga,Huesca (provincia),Aragón,España',  
'Blue whale,1205 avenue de falguières,82000,Montauban,France',  
'Granada (provincia),Andalucía,España',  
'UK FR040EC',  
'Cuéllar,Segovia (provincia),Castilla y León,España',  
'Covinor S.A.S.,Raismes,Nord,Nord-Pas-de-Calais,France',  
'Alaquas,Valencia (provincia),Comunidad Valenciana,España',  
'Reino Unido',  
'Grèce,47042,Filothei Artas',  
'25690 Avoudrey',  
'China,[IMPORTADOR],Calahorra,La Rioja,España',  
'strasbourg',  
'Luxemburg',  
'72400 La Ferté Bernard',  
'Don Benito, Badajoz (provincia), Extremadura, España',  
'Les Brasseurs de Gayant,59502,Douai,France',  
'Société Fromagère de Charchigné (Filiale Groupe Lactalys) - Route de Lassay - 53250 Charchigné,Mayenne,Société Fromagère de Bouvron (Filiale Groupe L  
actalys) - Route de Fay-de-Bretagne - 44130 Bouvron,Loire-Atlantique,Pays de la Loire,France',  
'Alliance Océane SAS,Verson,Calvados,Basse-Normandie,France',  
'Conserverie STÉPHAN (Filiale Groupe Le Graët) - ZI de Kerprat - 22970 Ploumagoar,Côtes-d'Armor,Bretagne,France',  
'Citricos del Andarax S.A.,Pasaje Llano El Ron S/N,ES-04560 Gador (almeira) Spain',  
'Noel alimentaria S.A.U.,Girona',  
'Molkerei Gropper GmbH & Co. Kg,Am Mühlberg 2 D - 86657 Bissingen Allemagne,Allemagne',  
'Olivet,Loiret,France',  
'Olivet,France',  
'Noyelles-Godault,France',  
'Narros de Cuéllar, Samboal, Segovia (provincia), Castilla y León, España',  
'Granada (provincia), Andalucía, España',  
'netherlands',  
'Benifaió, Valencia (provincia), Comunidad Valenciana, España',  
'São Miguel,Ponta Delgada,Açores,Portugal',  
'Roumanie',  
'Francia',  
'Haute-Marne,France',  
'Tunisie',  
'Compagnie des Fromages & Richesmonts - CF&R (Filiale Sodiaal International) - 91 Rue d'Aunay - 14500 Vire,Calvados,Basse-Normandie,France",  
'Pilar de la Horadada, Alicante (provincia), Comunidad Valenciana, España',  
'France,L.N.U.F. MOD. ZA Les Placis 35230 BOURGBARRE',  
'Bissingen,Molkerei Gropper GMBH& Co KG AM Muhberg 2,D-86657 Bissingen',  
'Allemagne,Molkerei Gropper GmbH Co. KG AM Mühlberg 2 D-86657 Bissingen',  
'Liechtenstein',  
'Hilcona AG - Liechtenstein',  
'Ribeira Grande,São Miguel,Açores,Portugal',

'Essen,Allemagne',  
 'Pays bas',  
 "Union Laitière Vitelloise (ULV)-Fromagerie de l'Ermitage - 718 Rue de la Division Leclerc - 88140 Bulgnéville,Vosges,Lorraine,France",  
 'ZA les placis 35230 bourgbarré',  
 'Arroyo de la Encienda,Valladolid (provincia),Castilla y León,España',  
 'Alginet,Valencia (provincia),Comunidad Valenciana,España',  
 'Novandie (Filiale Groupe Andros) - Rue de la Gravelle - 35560 Marcillé-Raoul,Ille-et-Vilaine,Bretagne,Novandie (Filiale Groupe Andros) - 11 Rue du Centre - 62240 Vieil-Moutier,Pas-de-Calais,Nord-Pas-de-Calais,France',  
 'Novandie,19 rue de la République,76153,Maromme,France',  
 "Mix'Buffet - ZA du Val Coric - 56380 Guer,Morbihan,Bretagne,France",  
 'Duitsland,Allemagne',  
 'Nutrisun GmbH,21218,Seevetal,Allemagne',  
 'Long Clawson Dairy LTD; Long Clawson; Milton Mowbray; LE14 4P1,UK',  
 'Italie,Ariano Polesine',  
 'Fontanilles,Gerona (provincia),Cataluña,España',  
 'Urkabe Benetan SA,Pol. Ind. Masti Loidi 9-11,Errenerteria,Gipuzkoa,Espagne',  
 'Fontanilles, Gerona (provincia), Cataluña, España',  
 '[DISTRIBUIDOR],Mendavia, Navarra, España',  
 'Tarragona,Espagne',  
 'San Cibrao das Viñas, Orense (provincia), Galicia, España',  
 "Arrivé Sainte-Hermine (Filiale Groupe Lambert Dodard Chancereul - LDC) - Avenue des Frênes - Parc d'Activités Vendée Atlantique - 85210 Saint-Jean-de-Beugné,Vendée,Pays de la Loire,France",  
 'La Puebla de Vícar (entidad de población), Vícar, Almería (provincia), Andalucía, España',  
 'Santa Eulália,Almargem do Bispo,Portugal',  
 'Huber GmbH & Co. - Karwendelstrasse 6-16 - D-86802 Buchloe',  
 'Murcia,Murcia (comunidad autónoma),España',  
 'Douarnenez,France',  
 'Terrea,Mas de Sarraset,Ancienne Voie Aurelia,13210,St. Remy de Provence,France',  
 'Ecuador',  
 'Weiland GmbH,Industriestrasse 25,49201,Dissen,Allemagne',  
 'Murcia,España',  
 'Ivory Coast',  
 'Marie - MSAS (Filiale Groupe Lambert Dodard Chancereul - LDC) - Route de Landrevarezec - 29510 Brie,Finistère,Bretagne,France',  
 'Polen',  
 'Novandie (Filiale Groupe Andros) - Rue de la Gravelle - 35560 Marcillé-Raoul,Ille-et-Vilaine,Bretagne,Pomly - 92 Rue Principale - 02210 Rozet-Saint-Albin,Aisne,Picardie,Novandie (Filiale Groupe Andros) - Route de Roinville - 28700 Auneau,Eure-et-Loir,Centre-Val de Loire,France',  
 'France,Kervignac',  
 'Cité Marine - Carrefour Industriel du Porzo - 56700 Kervignac,Morbihan,Bretagne,France',  
 '[DISTRIBUIDOR],Tudela,Navarra,España',  
 'Ludwig Weinrich GmbH,32020,Herford,Allemagne',  
 'Gironde,Aquitaine,France',  
 'Reus,Tarragona (provincia),Cataluña,España',  
 '03120,Lapalisse,France',  
 'Losar de la Vera, Cáceres (provincia), Extremadura, España',  
 'Pelussin,Loire,Rhône-Alpes,France',  
 'Paterna,Valencia (provincia),Comunidad Valenciana,España',  
 'D-74167 Neckarsulm Allemagne,Allemagne',  
 'Frankreich,Normandie',  
 '[FABRICANTE],Perú,[DISTRIBUIDOR],Mendavia,Navarra,España',  
 'Ramales de la Victoria, Cantabria, España',  
 'F-38490 LES ABRETS',  
 'Kervignac',  
 'Marcillé-Raoul,Fougères-Vitré (distrito),Ille-et-Vilaine,Bretaña,Francia',  
 'C Lorca 22,30120,El Palmar,Murcia,Spain',  
 'Candreita, Navarra, España',  
 'Chypre',  
 '39800 Poligny',  
 'Allemagne,TMA. Handelsgesellschaft mbh An den Breiten D-01454 Leppersdorf',  
 'France,Arcens,07310',  
 'Consukaa,Papendrecht,Pays Bas',  
 'Équateur',  
 'france,Normandie',  
 'Rovaniemi,Finland',  
 'Saupiquet (S. SAS) - Allée de Kergolvez - ZI de Kergolvez - 29000 Quimper,Finistère,Bretagne,France',  
 'Sardaigne,Italie',  
 'Murcia, Murcia (comunidad autónoma), España',  
 'Beneixama, Alicante (provincia), Comunidad Valenciana, España',  
 ...

In [ ]:

In [ ]:

In [202...]

myDF.describe()

Out[202...]

	no_nutriments	additives_n	ingredients_from_palm_oil_n	ingredients_from_palm_oil	ingredients_that_may_be_from_palm_oil_n	ingredients_that_may_be_from_palm_oil	nutrit
count	0.0	248939.000000	248939.000000	0.0	248939.000000	0.0	
mean	NaN	1.936024	0.019659	NaN	0.055246	NaN	
std	NaN	2.502019	0.140524	NaN	0.269207	NaN	
min	NaN	0.000000	0.000000	NaN	0.000000	NaN	
25%	NaN	0.000000	0.000000	NaN	0.000000	NaN	
50%	NaN	1.000000	0.000000	NaN	0.000000	NaN	
75%	NaN	3.000000	0.000000	NaN	0.000000	NaN	
max	NaN	31.000000	2.000000	NaN	6.000000	NaN	

In [203...]

```
# Recopié de mon précédent notebook pour contrôler les données

print("Taille du Dataset = ", myDF.shape[0], "lignes + ", myDF.shape[1], "colonnes ")

print("Nombres de NaN = ", myDF.isna().sum().sum())

print("Pourcent de NaN = ", (myDF.isna().sum().sum() / (myDF.shape[0]*myDF.shape[1])) * 100)
```

```

print("Nb lignes vides = ", myDF.shape[0] - myDF.dropna(axis="rows", how="all").shape[0])

# a enlever par la suite
myCols_Empty = [col for col in myDF.columns if myDF[col].isna().all()]
if (len(myCols_Empty) == 0) :
    print("Pas de colonnes vides")
else :
    print("Nb colonnes vides = ", len(myCols_Empty))
    print("Les colonnes vides = ", myCols_Empty)

# lignes en doublon
myDFrowsDuplicated = myDF[myDF.duplicated()]
print("Nbre lignes en doublon ", myDFrowsDuplicated.shape[0])

# regarder les noms en doublon (si il y en a, choisir un regle - comme dropduplicate qui ne garde que le premier)

```

Taille du Dataset = 320772 lignes + 162 colonnes  
 Nombres de Nan = 39608589  
 Pourcent de NaN = 76.22157263195135  
 Nb lignes vides = 0  
 Nb colonnes vides = 16  
 Les colonnes vides = ['no\_nutriments', 'ingredients\_from\_palm\_oil', 'ingredients\_that\_may\_be\_from\_palm\_oil', 'nutrition\_grade\_uk', 'butyric-acid\_100g', 'caproic-acid\_100g', 'lignoceric-acid\_100g', 'cerotic-acid\_100g', 'melissic-acid\_100g', 'elaidic-acid\_100g', 'mead-acid\_100g', 'erucic-acid\_100g', 'neronic-acid\_100g', 'chlorophyl\_100g', 'glycemic-index\_100g', 'water-hardness\_100g']  
 Nbre lignes en doublon 0

## Commençons le nettoyage

On peut enlever les colonnes vides.

On regarde ensuite quelles collonnes / lignes ont beaucoup de Nan, pour décider ce qu'on fait avec.

Ensuite on regardera les valeurs pour 100g - pour vérifier qu'on a pas de chiffres négatifs ou supérieurs à 100.

```
In [204...]: myDF2 = myDF.dropna(axis="columns", how="all").dropna(axis="rows", how="all")
print("Taille du Dataset = ", myDF2.shape[0], "lignes + ", myDF2.shape[1], "colonnes ")
```

Taille du Dataset = 320772 lignes + 146 colonnes

```
In [205...]: # on a bien enlevé les 16 colonnes vides.
# regardons maintenant les doublons sur les noms des produits :

myDF_counts_name = myDF2.product_name.value_counts()
myDF_counts_name
```

```
Out[205...]: Ice Cream           410
Extra Virgin Olive Oil      303
Potato Chips                281
Premium Ice Cream            226
Tomato Ketchup                182
...
Seltzer Water, Orange Cream      1
Holiday Sanded Trees & Snowmen, Sweet & Chewy Candy      1
Holiday Cinnamon Santas Candy      1
Ciliegine Fresh Mozzarella      1
Tomates aux Vermicelles          1
Name: product_name, Length: 221347, dtype: int64
```

```
In [206...]: myDF_counts_name[ myDF_counts_name > 1]
```

```
Out[206...]: Ice Cream           410
Extra Virgin Olive Oil      303
Potato Chips                281
Premium Ice Cream            226
Tomato Ketchup                182
...
Tic tac 100                  2
Chopped Macadamia Nuts      2
Shredded Sharp White Cheddar Cheese  2
Roasted Garlic Vinaigrette      2
Gluten Free Muffins          2
Name: product_name, Length: 24432, dtype: int64
```

```
In [207...]: myDF_counts_name[ myDF_counts_name > 1].sum()
```

Out[207...]: 106095

```
In [208...]: myDF_counts_name.unique()
```

```
Out[208...]: array([410, 303, 281, 226, 182, 167, 162, 157, 155, 150, 142, 140, 133,
       132, 124, 123, 120, 119, 117, 111, 108, 107, 106, 104, 103, 102,
       101, 99, 97, 94, 92, 91, 89, 88, 86, 83, 82, 80, 79,
       78, 77, 76, 74, 73, 72, 71, 70, 69, 68, 67, 66, 64,
       63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51,
       50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38,
       37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25,
       24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12,
       11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1], dtype=int64)
```

In [ ]:

```
In [209...]: # on va sortir les doublons.
# Mais les doublons dupliqués plus d'une centaines de fois ne sont-ils pas des produits différents?
# je vais identifier les doublons comme même produit et même marque.

myDF_mask_name_brand = myDF2.duplicated(subset=['product_name', 'brands_tags'], keep=False)
```

```
myDF_mask_name_brand.sum()
#myDoubLons = myDF_counts_name[ myDF_counts_name > 1].index
#print(myDoubLons)
```

Out[209...]

In [210...]: print(myDF2[myDF\_mask\_name\_brand].sort\_values(by = 'product\_name'))

```
      code                               url \
294518  7613269752103 http://world-fr.openfoodfacts.org/produit/7613...
315118  871478999784 http://world-fr.openfoodfacts.org/produit/8714...
33952   36800499423 http://world-fr.openfoodfacts.org/produit/0036...
33145   36800314207 http://world-fr.openfoodfacts.org/produit/0036...
33951   36800499416 http://world-fr.openfoodfacts.org/produit/0036...
...
320747   9800895250 http://world-fr.openfoodfacts.org/produit/9800...
320748    9802735 http://world-fr.openfoodfacts.org/produit/9802735
320752  98709879087 http://world-fr.openfoodfacts.org/produit/9870...
320755   988639 http://world-fr.openfoodfacts.org/produit/988639
320766  99410148 http://world-fr.openfoodfacts.org/produit/9941...

           creator  created_t  created_datetime \
294518  date-limite-app 1479823833 2016-11-22T14:10:33Z
315118  openfoodfacts-contributors 1463765792 2016-05-20T17:36:32Z
33952   usda-ndb-import 1489052504 2017-03-09T09:41:44Z
33145   usda-ndb-import 1489052504 2017-03-09T09:41:44Z
33951   usda-ndb-import 1489052503 2017-03-09T09:41:43Z
...
320747  tacite-mass-editor 1490554697 2017-03-26T18:58:17Z
320748  date-limite-app 1472728165 2016-09-01T11:09:25Z
320752  openfoodfacts-contributors 1472254210 2016-08-26T23:30:10Z
320755  tacite-mass-editor 1490554698 2017-03-26T18:58:18Z
320766  date-limite-app 1463936572 2016-05-22T17:02:52Z

  last_modified_t last_modified_datetime product_name generic_name \
294518  1479825877 2016-11-22T14:44:37Z          NaN
315118  1463765792 2016-05-20T17:36:32Z          NaN
33952  1489052504 2017-03-09T09:41:44Z #Error!          NaN
33145  1489052505 2017-03-09T09:41:45Z #Error!          NaN
33951  1489052504 2017-03-09T09:41:44Z #Error!          NaN
...
320747  1490554697 2017-03-26T18:58:17Z          NaN          NaN
320748  1472728167 2016-09-01T11:09:27Z          NaN          NaN
320752  1474729393 2016-09-24T15:03:13Z          NaN          NaN
320755  1490554698 2017-03-26T18:58:18Z          NaN          NaN
320766  1491244498 2017-04-03T18:34:58Z          NaN          NaN

      quantity packaging packaging_tags     brands brands_tags \
294518      NaN      NaN          NaN      NaN      NaN
315118      NaN      NaN          NaN      NaN      NaN
33952      NaN      NaN          NaN  Food Club  food-club
33145      NaN      NaN          NaN  Food Club  food-club
33951      NaN      NaN          NaN  Food Club  food-club
...
320747      ...      ...          ...      ...      ...
320748      NaN      NaN          NaN      NaN      NaN
320752      NaN      NaN          NaN      NaN      NaN
320755      33cl      NaN          NaN      NaN      NaN
320766      NaN      Plastique      plastique      NaN      NaN

           categories \
294518          NaN
315118          NaN
33952          NaN
33145          NaN
33951          NaN
...
320747  fr:Pâtes à tartiner aux noisettes et au cacao          NaN
320748          NaN
320752          Colas
320755  fr:Pâtes à tartiner aux noisettes et au cacao          NaN
320766          Peanuts

           categories_tags \
294518          NaN
315118          NaN
33952          NaN
33145          NaN
33951          NaN
...
320747  en:spreads,en:breakfasts,en:sweet-spreads,fr:p...          NaN
320748          NaN
320752  en:beverages,en:carbonated-drinks,en:sodas,en:...
320755  en:spreads,en:breakfasts,en:sweet-spreads,fr:p...
320766  en:plant-based-foods-and-beverages,en:plant-ba...

           categories_fr origins \
294518          NaN      NaN
315118          NaN      NaN
33952          NaN      NaN
33145          NaN      NaN
33951          NaN      NaN
...
320747  Produits à tartiner,Petit-déjeuners,Produits à...      NaN
320748          NaN      NaN
320752  Boissons,Boissons gazeuses,Sodas,Sodas au cola...      NaN
320755  Produits à tartiner,Petit-déjeuners,Produits à...      NaN
320766  Aliments et boissons à base de végétaux,Alimen...      NaN

      origins_tags manufacturing_places manufacturing_places_tags labels \
294518      NaN          NaN          NaN      NaN      NaN
315118      NaN          NaN          NaN      NaN      NaN
33952      NaN          NaN          NaN      NaN      NaN
33145      NaN          NaN          NaN      NaN      NaN
33951      NaN          NaN          NaN      NaN      NaN
```

... ... ... ... ...

320747 NaN NaN NaN NaN NaN  
 320748 NaN NaN NaN NaN NaN  
 320752 NaN NaN NaN NaN NaN  
 320755 NaN NaN NaN NaN NaN  
 320766 NaN NaN NaN NaN NaN

labels\_tags labels\_fr emb\_codes emb\_codes\_tags \

294518 NaN NaN NaN NaN  
 315118 NaN NaN NaN NaN  
 33952 NaN NaN NaN NaN  
 33145 NaN NaN NaN NaN  
 33951 NaN NaN NaN NaN  
 ... ... ... ... ...  
 320747 NaN NaN NaN NaN  
 320748 NaN NaN NaN NaN  
 320752 NaN NaN NaN NaN  
 320755 NaN NaN NaN NaN  
 320766 NaN NaN NaN NaN

first\_packaging\_code\_geo cities cities\_tags purchase\_places stores \

294518 NaN NaN NaN NaN NaN  
 315118 NaN NaN NaN NaN NaN  
 33952 NaN NaN NaN NaN NaN  
 33145 NaN NaN NaN NaN NaN  
 33951 NaN NaN NaN NaN NaN  
 ... ... ... ... ...  
 320747 NaN NaN NaN NaN  
 320748 NaN NaN NaN NaN  
 320752 NaN NaN NaN NaN  
 320755 NaN NaN NaN NaN  
 320766 NaN NaN NaN NaN

countries countries\_tags countries\_fr \

294518 Switzerland en:switzerland Suisse  
 315118 en:FR en:france France  
 33952 US en:united-states États-Unis  
 33145 US en:united-states États-Unis  
 33951 US en:united-states États-Unis  
 ... ... ... ...  
 320747 en:FR en:france France  
 320748 en:DZ en:algeria Algérie  
 320752 Switzerland en:switzerland Suisse  
 320755 en:FR en:france France  
 320766 United Kingdom en:united-kingdom Royaume-Uni

ingredients\_text allergens \

294518 NaN NaN  
 315118 NaN NaN  
 33952 Citric acid, maltodextrin\*, potassium citrate,... NaN  
 33145 Citric acid, instant tea, maltodextrin\*, corn ... NaN  
 33951 Citric acid, maltodextrin\*, aspartame\*\*, corns... NaN  
 ... ...  
 320747 NaN NaN  
 320748 NaN NaN  
 320752 NaN NaN  
 320755 NaN NaN  
 320766 NaN NaN

allergens\_fr traces traces\_tags traces\_fr serving\_size \

294518 NaN NaN NaN NaN NaN  
 315118 NaN NaN NaN NaN NaN  
 33952 NaN NaN NaN 2 g (8 fl oz)  
 33145 NaN NaN NaN 1.1 g (8 fl oz)  
 33951 NaN NaN NaN 1.3 g (8 fl oz)  
 ... ... ... ...  
 320747 NaN NaN NaN NaN  
 320748 NaN NaN NaN NaN  
 320752 NaN NaN NaN NaN  
 320755 NaN NaN NaN NaN  
 320766 NaN NaN NaN NaN

additives\_n additives \

294518 NaN NaN  
 315118 NaN NaN  
 33952 6.0 [ citric-acid -> en:e330 -> exists -- ok ]...  
 33145 7.0 [ citric-acid -> en:e330 -> exists -- ok ]...  
 33951 10.0 [ citric-acid -> en:e330 -> exists -- ok ]...  
 ... ...  
 320747 NaN NaN  
 320748 NaN NaN  
 320752 NaN NaN  
 320755 NaN NaN  
 320766 NaN NaN

additives\_tags \

294518 NaN  
 315118 NaN  
 33952 en:e330,en:e331,en:e951,en:e950,en:e341iiii,en:...  
 33145 en:e330,en:e951,en:e110,en:e341iiii,en:e129,en:...  
 33951 en:e330,en:e951,en:e331,en:e341iiii,en:e129,en:...  
 ... ...  
 320747 NaN  
 320748 NaN  
 320752 NaN  
 320755 NaN  
 320766 NaN

additives\_fr \

294518 NaN  
 315118 NaN  
 33952 E330 - Acide citrique,E331 - Citrates de sodiu...  
 33145 E330 - Acide citrique,E951 - Aspartame,E110 - ...  
 33951 E330 - Acide citrique,E951 - Aspartame,E331 - ...  
 ... ...  
 320747 NaN  
 320748 NaN

320752			NaN		
320755			NaN		
320766			NaN		
294518	ingredients_from_palm_oil_n	ingredients_from_palm_oil_tags	\		
315118	NaN	NaN			
33952	NaN	NaN			
33145	0.0	NaN			
33951	0.0	NaN			
...	...	...			
320747	NaN	NaN			
320748	NaN	NaN			
320752	NaN	NaN			
320755	NaN	NaN			
320766	NaN	NaN			
294518	ingredients_that_may_be_from_palm_oil_n	\			
315118	NaN	NaN			
33952	NaN	NaN			
33145	0.0	NaN			
33951	0.0	NaN			
...	...	...			
320747	NaN	NaN			
320748	NaN	NaN			
320752	NaN	NaN			
320755	NaN	NaN			
320766	NaN	NaN			
294518	ingredients_that_may_be_from_palm_oil_tags	nutrition_grade_fr	\		
315118	NaN	NaN			
33952	NaN	NaN			
33145	NaN	NaN			
33951	NaN	NaN			
...	...	...			
320747	NaN	NaN			
320748	NaN	NaN			
320752	NaN	NaN			
320755	NaN	NaN			
320766	NaN	NaN			
294518	pnns_groups_1	pnns_groups_2	\		
315118	unknown	unknown			
33952	NaN	NaN			
33145	NaN	NaN			
33951	NaN	NaN			
...	...	...			
320747	Sugary snacks	Sweets			
320748	NaN	NaN			
320752	Beverages	Sweetened beverages			
320755	Sugary snacks	Sweets			
320766	Salty snacks	Nuts			
294518		states	\		
315118	en:to-be-completed, en:nutrition-facts-to-be-c...				
33952	en:to-be-completed, en:nutrition-facts-to-be-c...				
33145	en:to-be-completed, en:nutrition-facts-complet...				
33951	en:to-be-completed, en:nutrition-facts-complet...				
...	...				
320747	en:to-be-completed, en:nutrition-facts-to-be-c...				
320748	en:to-be-completed, en:nutrition-facts-to-be-c...				
320752	en:to-be-completed, en:nutrition-facts-to-be-c...				
320755	en:to-be-completed, en:nutrition-facts-to-be-c...				
320766	en:to-be-completed, en:nutrition-facts-to-be-c...				
294518		states_tags	\		
315118	en:to-be-completed, en:nutrition-facts-to-be-co...				
33952	en:to-be-completed, en:nutrition-facts-complete...				
33145	en:to-be-completed, en:nutrition-facts-complete...				
33951	en:to-be-completed, en:nutrition-facts-complete...				
...	...				
320747	en:to-be-completed, en:nutrition-facts-to-be-co...				
320748	en:to-be-completed, en:nutrition-facts-to-be-co...				
320752	en:to-be-completed, en:nutrition-facts-to-be-co...				
320755	en:to-be-completed, en:nutrition-facts-to-be-co...				
320766	en:to-be-completed, en:nutrition-facts-to-be-co...				
294518		states_fr	\		
315118	A compléter, Informations nutritionnelles à com...				
33952	A compléter, Informations nutritionnelles à com...				
33145	A compléter, Informations nutritionnelles compl...				
33951	A compléter, Informations nutritionnelles compl...				
...	...				
320747	A compléter, Informations nutritionnelles à com...				
320748	A compléter, Informations nutritionnelles à com...				
320752	A compléter, Informations nutritionnelles à com...				
320755	A compléter, Informations nutritionnelles à com...				
320766	A compléter, Informations nutritionnelles à com...				
294518	main_category	main_category_fr	image_url	image_small_url	\
315118	NaN	NaN	NaN	NaN	
33952	NaN	NaN	NaN	NaN	
33145	NaN	NaN	NaN	NaN	
33951	NaN	NaN	NaN	NaN	
...	...	...	...	...	
320747	fr:pates-a-tartiner	Pâtes à tartiner	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	en:sodas	Sodas	NaN	NaN	
320755	fr:pates-a-tartiner	Pâtes à tartiner	NaN	NaN	
320766	en:nuts	Fruits à coques	NaN	NaN	

	energy_100g	energy-from-fat_100g	fat_100g	saturated-fat_100g	\
294518	NaN	NaN	NaN	NaN	
315118	NaN	NaN	NaN	NaN	
33952	1046.0	NaN	0.0	NaN	
33145	1904.0	NaN	0.0	NaN	
33951	1611.0	NaN	0.0	NaN	
...	...	...	...	...	
320747	NaN	NaN	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	NaN	NaN	NaN	NaN	
320755	NaN	NaN	NaN	NaN	
320766	NaN	NaN	NaN	NaN	
	caprylic-acid_100g	capric-acid_100g	lauric-acid_100g	\	
294518	NaN	NaN	NaN	NaN	
315118	NaN	NaN	NaN	NaN	
33952	NaN	NaN	NaN	NaN	
33145	NaN	NaN	NaN	NaN	
33951	NaN	NaN	NaN	NaN	
...	...	...	...	...	
320747	NaN	NaN	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	NaN	NaN	NaN	NaN	
320755	NaN	NaN	NaN	NaN	
320766	NaN	NaN	NaN	NaN	
	myristic-acid_100g	palmitic-acid_100g	stearic-acid_100g	\	
294518	NaN	NaN	NaN	NaN	
315118	NaN	NaN	NaN	NaN	
33952	NaN	NaN	NaN	NaN	
33145	NaN	NaN	NaN	NaN	
33951	NaN	NaN	NaN	NaN	
...	...	...	...	...	
320747	NaN	NaN	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	NaN	NaN	NaN	NaN	
320755	NaN	NaN	NaN	NaN	
320766	NaN	NaN	NaN	NaN	
	arachidic-acid_100g	behenic-acid_100g	montanic-acid_100g	\	
294518	NaN	NaN	NaN	NaN	
315118	NaN	NaN	NaN	NaN	
33952	NaN	NaN	NaN	NaN	
33145	NaN	NaN	NaN	NaN	
33951	NaN	NaN	NaN	NaN	
...	...	...	...	...	
320747	NaN	NaN	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	NaN	NaN	NaN	NaN	
320755	NaN	NaN	NaN	NaN	
320766	NaN	NaN	NaN	NaN	
	monounsaturated-fat_100g	polyunsaturated-fat_100g	omega-3-fat_100g	\	
294518	NaN	NaN	NaN	NaN	
315118	NaN	NaN	NaN	NaN	
33952	NaN	NaN	NaN	NaN	
33145	NaN	NaN	NaN	NaN	
33951	NaN	NaN	NaN	NaN	
...	...	...	...	...	
320747	NaN	NaN	NaN	NaN	
320748	NaN	NaN	NaN	NaN	
320752	NaN	NaN	NaN	NaN	
320755	NaN	NaN	NaN	NaN	
320766	NaN	NaN	NaN	NaN	
	alpha-linolenic-acid_100g	eicosapentaenoic-acid_100g	\		
294518	NaN	NaN	NaN		
315118	NaN	NaN	NaN		
33952	NaN	NaN	NaN		
33145	NaN	NaN	NaN		
33951	NaN	NaN	NaN		
...	...	...	...		
320747	NaN	NaN	NaN		
320748	NaN	NaN	NaN		
320752	NaN	NaN	NaN		
320755	NaN	NaN	NaN		
320766	NaN	NaN	NaN		
	docosahexaenoic-acid_100g	omega-6-fat_100g	linoleic-acid_100g	\	
294518	NaN	NaN	NaN		
315118	NaN	NaN	NaN		
33952	NaN	NaN	NaN		
33145	NaN	NaN	NaN		
33951	NaN	NaN	NaN		
...	...	...	...		
320747	NaN	NaN	NaN		
320748	NaN	NaN	NaN		
320752	NaN	NaN	NaN		
320755	NaN	NaN	NaN		
320766	NaN	NaN	NaN		
	arachidonic-acid_100g	gamma-linolenic-acid_100g	\		
294518	NaN	NaN			
315118	NaN	NaN			
33952	NaN	NaN			
33145	NaN	NaN			
33951	NaN	NaN			
...	...	...			
320747	NaN	NaN			
320748	NaN	NaN			
320752	NaN	NaN			
320755	NaN	NaN			
320766	NaN	NaN			
	dihomo-gamma-linolenic-acid_100g	omega-9-fat_100g	oleic-acid_100g	\	
294518	NaN	NaN	NaN		

315118		NaN	NaN	NaN		
33952		NaN	NaN	NaN		
33145		NaN	NaN	NaN		
33951		NaN	NaN	NaN		
...		...	...	...		
320747		NaN	NaN	NaN		
320748		NaN	NaN	NaN		
320752		NaN	NaN	NaN		
320755		NaN	NaN	NaN		
320766		NaN	NaN	NaN		
	gondoic-acid_100g	trans-fat_100g	cholesterol_100g	\		
294518		NaN	NaN	NaN		
315118		NaN	NaN	NaN		
33952		NaN	NaN	NaN		
33145		NaN	NaN	NaN		
33951		NaN	NaN	NaN		
...	...	...	...	...		
320747		NaN	NaN	NaN		
320748		NaN	NaN	NaN		
320752		NaN	NaN	NaN		
320755		NaN	NaN	NaN		
320766		NaN	NaN	NaN		
	carbohydrates_100g	sugars_100g	sucrose_100g	glucose_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		0.0	NaN	NaN	NaN	
33145		0.0	NaN	NaN	NaN	
33951		0.0	NaN	NaN	NaN	
...	...	...	...	...	...	
320747		NaN	NaN	NaN	NaN	
320748		NaN	NaN	NaN	NaN	
320752		NaN	NaN	NaN	NaN	
320755		NaN	NaN	NaN	NaN	
320766		NaN	NaN	NaN	NaN	
	fructose_100g	lactose_100g	maltose_100g	maltodextrins_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		NaN	NaN	NaN	NaN	
33145		NaN	NaN	NaN	NaN	
33951		NaN	NaN	NaN	NaN	
...	...	...	...	...	...	
320747		NaN	NaN	NaN	NaN	
320748		NaN	NaN	NaN	NaN	
320752		NaN	NaN	NaN	NaN	
320755		NaN	NaN	NaN	NaN	
320766		NaN	NaN	NaN	NaN	
	starch_100g	polyols_100g	fiber_100g	proteins_100g	casein_100g	\
294518		NaN	NaN	NaN	NaN	NaN
315118		NaN	NaN	NaN	NaN	NaN
33952		NaN	NaN	NaN	0.0	NaN
33145		NaN	NaN	NaN	0.0	NaN
33951		NaN	NaN	NaN	0.0	NaN
...	...	...	...	...	...	...
320747		NaN	NaN	NaN	NaN	NaN
320748		NaN	NaN	NaN	NaN	NaN
320752		NaN	NaN	NaN	NaN	NaN
320755		NaN	NaN	NaN	NaN	NaN
320766		NaN	NaN	NaN	NaN	NaN
	serum-proteins_100g	nucleotides_100g	salt_100g	sodium_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		NaN	NaN	2.54	1.0	
33145		NaN	NaN	0.00	0.0	
33951		NaN	NaN	0.00	0.0	
...	...	...	...	...	...	
320747		NaN	NaN	NaN	NaN	
320748		NaN	NaN	NaN	NaN	
320752		NaN	NaN	NaN	NaN	
320755		NaN	NaN	NaN	NaN	
320766		NaN	NaN	NaN	NaN	
	alcohol_100g	vitamin-a_100g	beta-carotene_100g	vitamin-d_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		NaN	NaN	NaN	NaN	
33145		NaN	NaN	NaN	NaN	
33951		NaN	NaN	NaN	NaN	
...	...	...	...	...	...	
320747		NaN	NaN	NaN	NaN	
320748		NaN	NaN	NaN	NaN	
320752		NaN	NaN	NaN	NaN	
320755		NaN	NaN	NaN	NaN	
320766		NaN	NaN	NaN	NaN	
	vitamin-e_100g	vitamin-k_100g	vitamin-c_100g	vitamin-b1_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		NaN	NaN	NaN	NaN	
33145		NaN	NaN	NaN	NaN	
33951		NaN	NaN	NaN	NaN	
...	...	...	...	...	...	
320747		NaN	NaN	NaN	NaN	
320748		NaN	NaN	NaN	NaN	
320752		NaN	NaN	NaN	NaN	
320755		NaN	NaN	NaN	NaN	
320766		NaN	NaN	NaN	NaN	
	vitamin-b2_100g	vitamin-pp_100g	vitamin-b6_100g	vitamin-b9_100g	\	
294518		NaN	NaN	NaN	NaN	
315118		NaN	NaN	NaN	NaN	
33952		NaN	NaN	NaN	NaN	
33145		NaN	NaN	NaN	NaN	

33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	folates_100g	vitamin-b12_100g	biotin_100g	pantothenic-acid_100g
294518	NaN	NaN	NaN	NaN
315118	NaN	NaN	NaN	NaN
33952	NaN	NaN	NaN	NaN
33145	NaN	NaN	NaN	NaN
33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	silica_100g	bicarbonate_100g	potassium_100g	chloride_100g
294518	NaN	NaN	NaN	NaN
315118	NaN	NaN	NaN	NaN
33952	NaN	NaN	NaN	NaN
33145	NaN	NaN	NaN	NaN
33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	calcium_100g	phosphorus_100g	iron_100g	magnesium_100g
294518	NaN	NaN	NaN	NaN
315118	NaN	NaN	NaN	NaN
33952	NaN	NaN	NaN	NaN
33145	NaN	NaN	NaN	NaN
33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	copper_100g	manganese_100g	fluoride_100g	selenium_100g
294518	NaN	NaN	NaN	NaN
315118	NaN	NaN	NaN	NaN
33952	NaN	NaN	NaN	NaN
33145	NaN	NaN	NaN	NaN
33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	chromium_100g	molybdenum_100g	iodine_100g	caffeine_100g
294518	NaN	NaN	NaN	NaN
315118	NaN	NaN	NaN	NaN
33952	NaN	NaN	NaN	NaN
33145	NaN	NaN	NaN	NaN
33951	NaN	NaN	NaN	NaN
...	...	...	...	...
320747	NaN	NaN	NaN	NaN
320748	NaN	NaN	NaN	NaN
320752	NaN	NaN	NaN	NaN
320755	NaN	NaN	NaN	NaN
320766	NaN	NaN	NaN	NaN
	taurine_100g	ph_100g	fruits-vegetables-nuts_100g	
294518	NaN	NaN	NaN	
315118	NaN	NaN	NaN	
33952	NaN	NaN	NaN	
33145	NaN	NaN	NaN	
33951	NaN	NaN	NaN	
...	...	...	...	
320747	NaN	NaN	NaN	
320748	NaN	NaN	NaN	
320752	NaN	NaN	NaN	
320755	NaN	NaN	NaN	
320766	NaN	NaN	NaN	
	collagen-meat-protein-ratio_100g	cocoa_100g	carbon-footprint_100g	
294518	NaN	NaN	NaN	
315118	NaN	NaN	NaN	
33952	NaN	NaN	NaN	
33145	NaN	NaN	NaN	
33951	NaN	NaN	NaN	
...	...	...	...	
320747	NaN	NaN	NaN	
320748	NaN	NaN	NaN	
320752	NaN	NaN	NaN	
320755	NaN	NaN	NaN	
320766	NaN	NaN	NaN	
	nutrition-score-fr_100g	nutrition-score-uk_100g		
294518	NaN	NaN		
315118	NaN	NaN		
33952	NaN	NaN		
33145	NaN	NaN		
33951	NaN	NaN		
...	...	...		
320747	NaN	NaN		

```

320748           NaN
320752           NaN
320755           NaN
320766           NaN

```

[52831 rows x 146 columns]

In [211...  
`## a voir avec Anthony : je ne suis pas sur que je devrais tous les retirer ?  
# on ne peut pas un peu mieux voir ce qu'on fait, surtout avec tout les "ice-cream" ?  
# je refais la même chose avec le nom, la marque, et le pays :`

```
myDF_mask_name_brand = myDF2.duplicated(subset=['product_name', 'countries_fr', 'brands_tags'], keep=False)
myDF_mask_name_brand.sum()
```

Out[211...  
50838

In [212...  
`# c'est pas beaucoup mieux. on rattrappe 2000 produits.  
# on va utiliser ça.`

```
myDoublonSeries = myDF2.duplicated(subset=['product_name', 'countries_fr', 'brands_tags'], keep='first')
myDF2_test = myDF2[~myDoublonSeries]
```

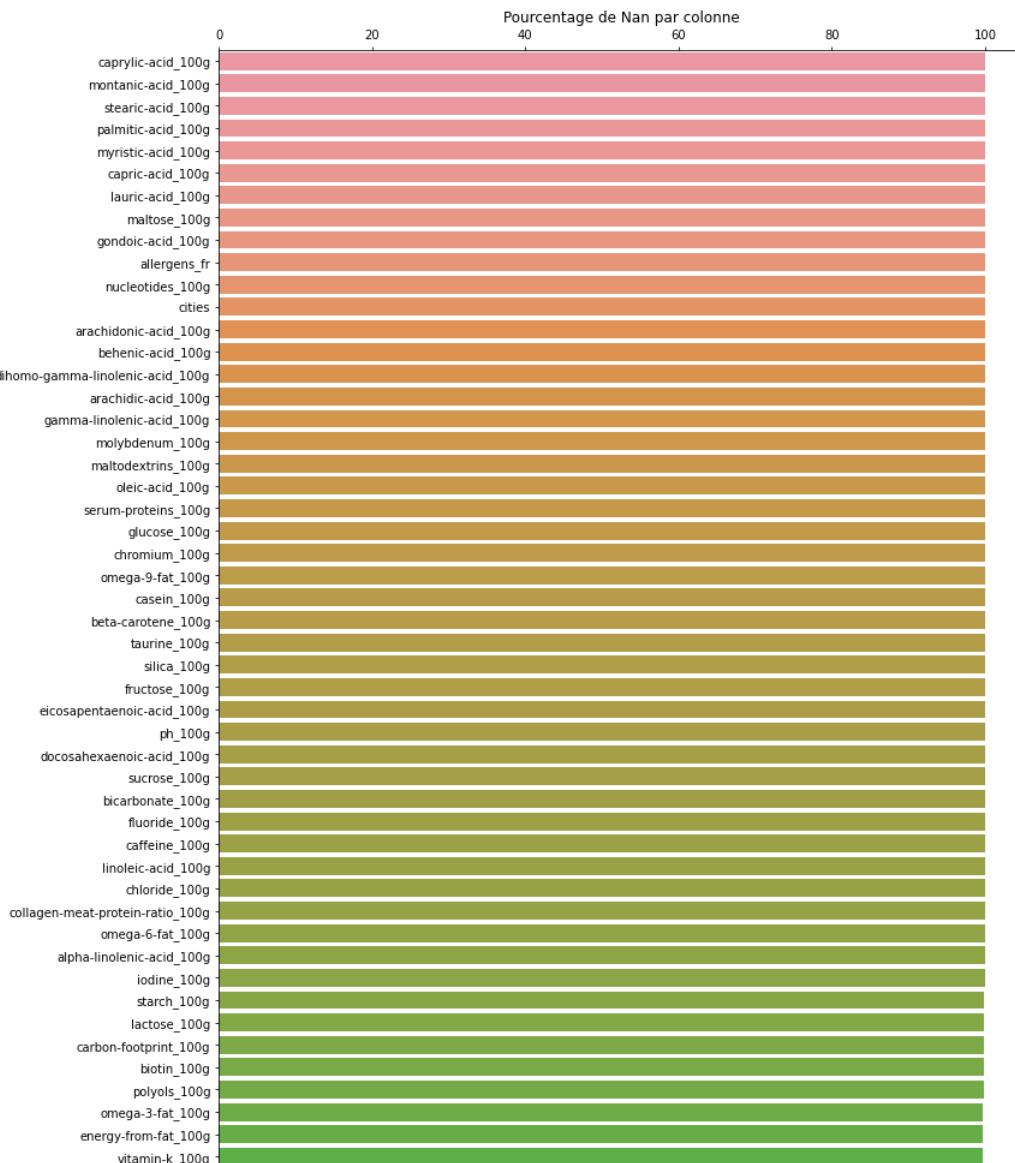
In [213...  
`print("Taille du Dataset myDF2 = ", myDF2.shape[0], "lignes + ", myDF2.shape[1], "colonnes ")
print("Taille du Dataset myDF2_test = ", myDF2_test.shape[0], "lignes + ", myDF2_test.shape[1], "colonnes ")`

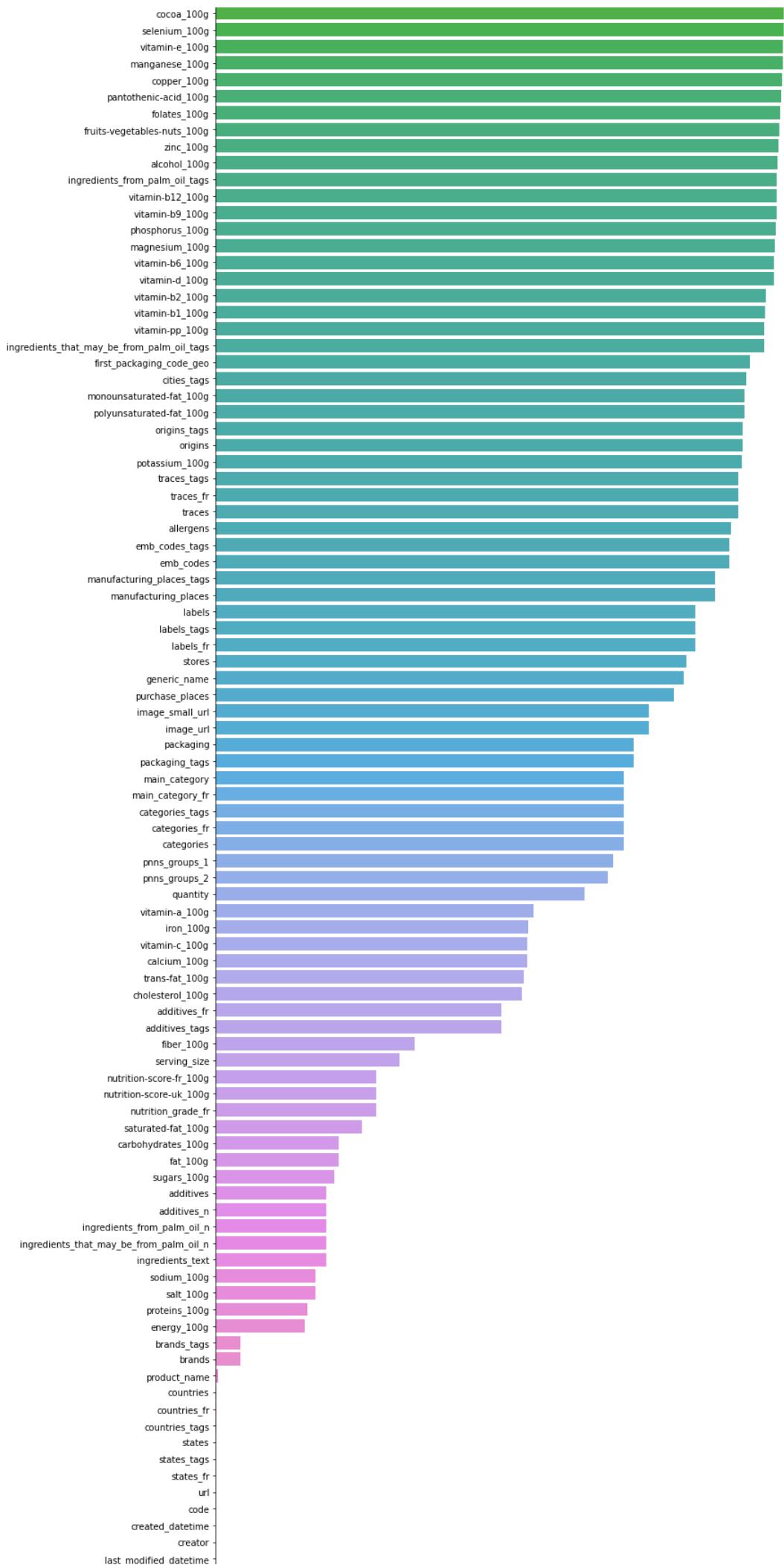
Taille du Dataset myDF2 = 320772 lignes + 146 colonnes  
Taille du Dataset myDF2\_test = 284230 lignes + 146 colonnes

In [214...  
`# ca a marché. Je mets mes 248k Lignes dans DF2 :`  
`myDF2 = myDF2_test`

In [215...  
`pourcent_nan = myDF2.isna().sum().divide(myDF2.shape[0]/100).sort_values(ascending=False)`

```
plt.figure(figsize=(12, 50)) #necessaire pour que les labels soient lisibles
ax = sns.barplot(y = pourcent_nan.index, x=pourcent_nan.values)
ax.xaxis.set_ticks_position('top')
plt.title('Pourcentage de Nan par colonne')
plt.show()
```





In [216]:

pourcent\_nan

	last_modified_t	created_t
caprylic-acid_100g	99.999648	
montanic-acid_100g	99.999648	
stearic-acid_100g	99.999648	
palmitic-acid_100g	99.999648	
myristic-acid_100g	99.999648	
capric-acid_100g	99.999296	
lauric-acid_100g	99.998945	
maltose_100g	99.998593	
gondoic-acid_100g	99.998593	
allergens_fr	99.997185	
nucleotides_100g	99.997185	
cities	99.997185	
arachidonic-acid_100g	99.997185	
behenic-acid_100g	99.997185	
dihomo-gamma-linolenic-acid_100g	99.997185	
arachidic-acid_100g	99.996834	
gamma-linolenic-acid_100g	99.996834	
molybdenum_100g	99.996482	
maltooligosaccharide_100g	99.996130	
oleic-acid_100g	99.995426	
serum-proteins_100g	99.994723	
glucose_100g	99.993315	
chromium_100g	99.993315	
omega-9-fat_100g	99.992612	
casein_100g	99.990852	
beta-carotene_100g	99.990852	
taurine_100g	99.990149	
silica_100g	99.989093	
fructose_100g	99.988390	
eicosapentaenoic-acid_100g	99.986982	
ph_100g	99.985927	
docosahexaenoic-acid_100g	99.977835	
sucrose_100g	99.976428	
bicarbonate_100g	99.975724	
fluoride_100g	99.974668	
caffeine_100g	99.972557	
linoleic-acid_100g	99.950040	
chloride_100g	99.949337	
collagen-meat-protein-ratio_100g	99.943004	
omega-6-fat_100g	99.938782	
alpha-linolenic-acid_100g	99.937023	
iodine_100g	99.915913	
starch_100g	99.912043	
lactose_100g	99.912043	
carbon-footprint_100g	99.910988	
biotin_100g	99.897970	
polyols_100g	99.859269	
omega-3-fat_100g	99.721001	
energy-from-fat_100g	99.718890	
vitamin-k_100g	99.714668	
cocoa_100g	99.675967	
selenium_100g	99.664005	
vitamin-e_100g	99.573585	
manganese_100g	99.489498	
copper_100g	99.348415	
pantothenic-acid_100g	99.215424	
folates_100g	99.038455	
fruits-vegetables-nuts_100g	98.965275	
zinc_100g	98.776343	
alcohol_100g	98.616261	
ingredients_from_palm_oil_tags	98.428385	
vitamin-b12_100g	98.393906	
vitamin-b9_100g	98.347465	
phosphorus_100g	98.202512	
magnesium_100g	98.074447	
vitamin-b6_100g	97.918235	
vitamin-d_100g	97.889737	
vitamin-b2_100g	96.547866	
vitamin-b1_100g	96.444781	
vitamin-pp_100g	96.280125	
ingredients_that_may_be_from_palm_oil_tags	96.171762	
first_packaging_code_geo	93.649157	
cities_tags	93.145340	
monounsaturated-fat_100g	92.845583	
polyunsaturated-fat_100g	92.830454	
origins_tags	92.491292	
origins	92.478626	
potassium_100g	92.372374	
traces_tags	91.689477	
traces_fr	91.686662	
traces	91.686310	
allergens	90.393343	
emb_codes_tags	90.115048	
emb_codes	90.113640	
manufacturing_places_tags	87.574500	
manufacturing_places	87.572389	
labels	84.201527	
labels_tags	84.172677	
labels_fr	84.169862	
stores	82.509587	
generic_name	82.183226	
purchase_places	80.406713	
image_small_url	76.041938	
image_url	76.041938	
packaging	73.367343	
packaging_tags	73.366992	
main_category	71.636351	
main_category_fr	71.636351	
categories_tags	71.628259	
categories_fr	71.625444	

```

categories 71.625444
pnns_groups_1 69.700243
pnns_groups_2 68.723217
quantity 64.655385
vitamin-a_100g 55.748162
iron_100g 54.820744
vitamin-c_100g 54.696549
calcium_100g 54.637442
trans-fat_100g 54.035464
cholesterol_100g 53.739225
additives_fr 50.079161
additives_tags 50.079161
fiber_100g 34.934384
serving_size 32.207367
nutrition-score-fr_100g 28.216585
nutrition-score-uk_100g 28.216585
nutrition_grade_fr 28.216585
saturated-fat_100g 25.621856
carbohydrates_100g 21.634944
fat_100g 21.526932
sugars_100g 20.784928
additives 19.444112
additives_n 19.433206
ingredients_from_palm_oil_n 19.433206
ingredients_text 19.430391
sodium_100g 17.484783
salt_100g 17.471062
proteins_100g 16.048974
energy_100g 15.661612
brands_tags 4.380255
brands 4.377441
product_name 0.476375
countries 0.082680
countries_fr 0.082680
countries_tags 0.082680
states 0.010907
states_tags 0.010907
states_fr 0.010907
url 0.002815
code 0.002815
created_datetime 0.001407
creator 0.000352
last_modified_datetime 0.000000
last_modified_t 0.000000
created_t 0.000000
dtype: float64

```

In [217]...

```

# La Liste des ingrédients avec beaucoup de Nan n'est pas choquante (acides rares)
# le champs (non numérique) cities est aussi dans la liste - on ne va sans doute pas l'utiliser de toutes façons

# c'est un peu plus gênant pour des valeurs comme fructose, ph, energy from fat (qu'on pourrait si besoin reconstruire)

# pour la plupart de ces champs (ceux terminant en _100g)
# on peut considérer qu'un Nan est équivalent à une valeur nulle.

#list_100g = filter(Lambda x: '_100g' in x , list(myDF2))
# MENTOR => est-ce la méthode la plus simple ?
# cela donne une erreur (de type peut-être, il fallait des crochets?)

list_100g = [x for x in list(myDF2) if '_100g' in x]
print(list_100g)

```

```

['energy_100g', 'energy-from-fat_100g', 'fat_100g', 'saturated-fat_100g', 'caprylic-acid_100g', 'capric-acid_100g', 'lauric-acid_100g', 'myristic-acid_100g', 'palmitic-acid_100g', 'stearic-acid_100g', 'arachidic-acid_100g', 'behenic-acid_100g', 'montanic-acid_100g', 'monounsaturated-fat_100g', 'polyunsaturated-fat_100g', 'omega-3-fat_100g', 'alpha-linolenic-acid_100g', 'eicosapentaenoic-acid_100g', 'docosahexaenoic-acid_100g', 'omega-6-fat_100g', '1-inoleic-acid_100g', 'arachidonate-acid_100g', 'gamma-linolenic-acid_100g', 'dihomo-gamma-linolenic-acid_100g', 'omega-9-fat_100g', 'oleic-acid_100g', 'gondoic-acid_100g', 'trans-fat_100g', 'cholesterol_100g', 'carbohydrates_100g', 'sugars_100g', 'sucrose_100g', 'glucose_100g', 'fructose_100g', 'lactose_100g', 'maltose_100g', 'maltodextrins_100g', 'starch_100g', 'polyols_100g', 'fiber_100g', 'proteins_100g', 'casein_100g', 'serum-proteins_100g', 'nucleotides_100g', 'salt_100g', 'sodium_100g', 'alcohol_100g', 'beta-carotene_100g', 'vitamin-d_100g', 'vitamin-e_100g', 'vitamin-k_100g', 'vitamin-c_100g', 'vitamin-b1_100g', 'vitamin-b2_100g', 'vitamin-pp_100g', 'vitamin-b6_100g', 'vitamin-b9_100g', 'folates_100g', 'vitamin-b12_100g', 'biotin_100g', 'pantothenic-acid_100g', 'silica_100g', 'bicarbonate_100g', 'potassium_100g', 'chloride_100g', 'calcium_100g', 'phosphorus_100g', 'iron_100g', 'magnesium_100g', 'zinc_100g', 'copper_100g', 'manganese_100g', 'fluoride_100g', 'selenium_100g', 'chromium_100g', 'molybdenum_100g', 'iodine_100g', 'caffeine_100g', 'taurine_100g', 'ph_100g', 'fruits-vegetables-nuts_100g', 'collagen-meat-protein-ratio_100g', 'cocoa_100g', 'carbon-footprint_100g', 'nutrition-score-fr_100g', 'nutrition-score-uk_100g']

```

In [218]...

```

# sélection des valeurs qu'on va mettre à 0 (car elle doivent être entre 0 et 100)
# on enlève energie nutriscore

```

```

list_ingredient_100g = ['fat_100g', 'saturated-fat_100g',
    'caprylic-acid_100g', 'capric-acid_100g', 'lauric-acid_100g', 'myristic-acid_100g',
    'palmitic-acid_100g', 'stearic-acid_100g', 'arachidic-acid_100g', 'behenic-acid_100g',
    'montanic-acid_100g', 'monounsaturated-fat_100g', 'polyunsaturated-fat_100g',
    'omega-3-fat_100g', 'alpha-linolenic-acid_100g', 'eicosapentaenoic-acid_100g',
    'docosahexaenoic-acid_100g', 'omega-6-fat_100g', 'linoleic-acid_100g', 'arachidonic-acid_100g',
    'gamma-linolenic-acid_100g', 'dihomo-gamma-linolenic-acid_100g', 'omega-9-fat_100g', 'oleic-acid_100g',
    'gondoic-acid_100g', 'trans-fat_100g', 'cholesterol_100g', 'carbohydrates_100g', 'sugars_100g', 'sucrose_100g',
    'glucose_100g', 'fructose_100g', 'lactose_100g', 'maltose_100g', 'maltodextrins_100g', 'starch_100g',
    'polyols_100g', 'fiber_100g', 'proteins_100g', 'casein_100g', 'serum-proteins_100g', 'nucleotides_100g',
    'salt_100g', 'sodium_100g', 'alcohol_100g', 'beta-carotene_100g', 'vitamin-d_100g',
    'vitamin-e_100g', 'vitamin-k_100g', 'vitamin-c_100g', 'vitamin-b1_100g', 'vitamin-b2_100g', 'vitamin-pp_100g',
    'vitamin-b6_100g', 'vitamin-b9_100g', 'folates_100g', 'vitamin-b12_100g', 'biotin_100g', 'pantothenic-acid_100g',
    'silica_100g', 'bicarbonate_100g', 'potassium_100g', 'chloride_100g', 'calcium_100g', 'phosphorus_100g',
    'iron_100g', 'magnesium_100g', 'zinc_100g', 'copper_100g', 'manganese_100g', 'fluoride_100g', 'selenium_100g',
    'chromium_100g', 'molybdenum_100g', 'iodine_100g', 'caffeine_100g', 'taurine_100g', 'ph_100g',
    'fruits-vegetables-nuts_100g', 'collagen-meat-protein-ratio_100g', 'cocoa_100g'
]

```

In [219]...

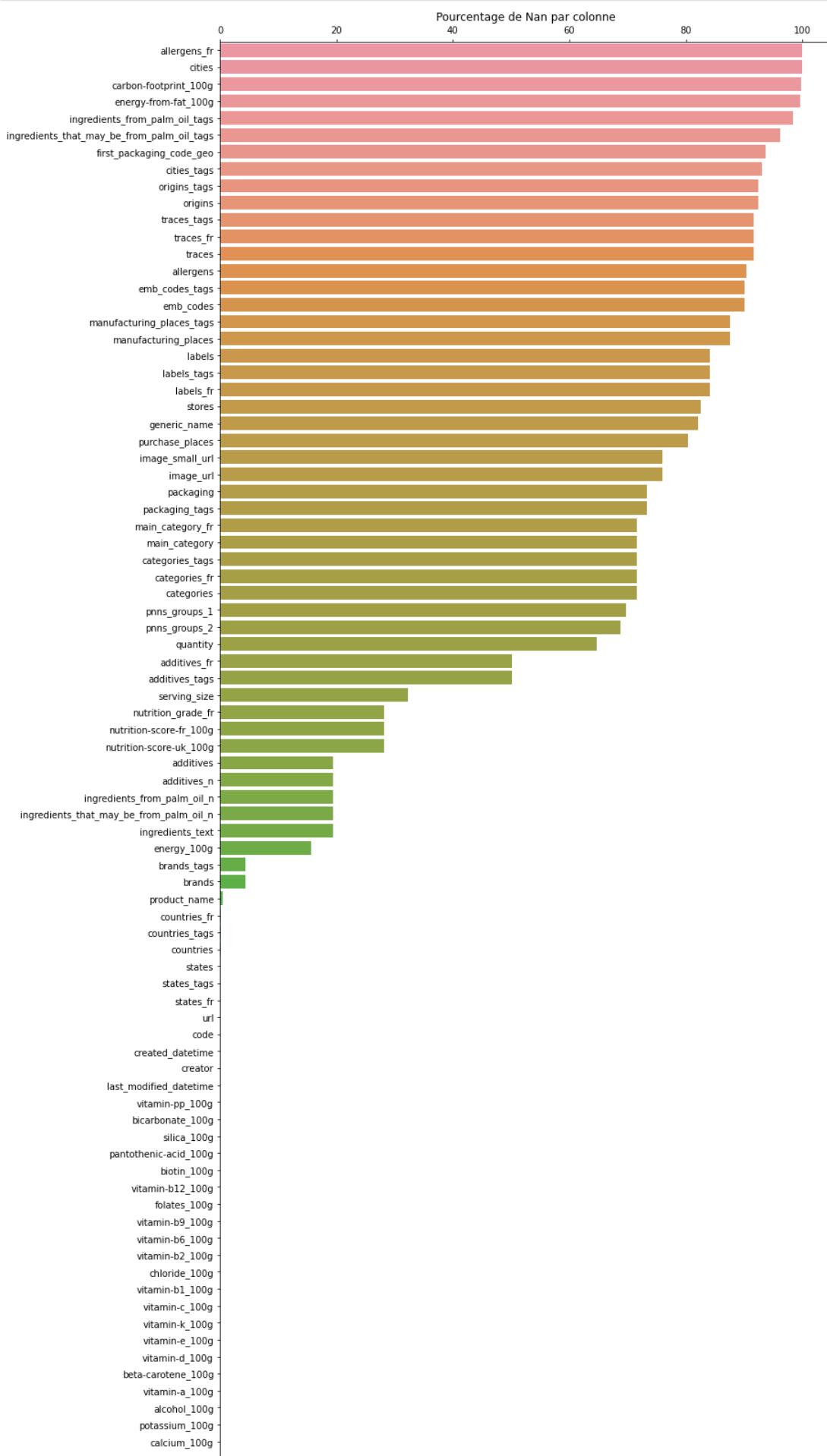
```

myDF3 = myDF2
myDF3[list_ingredient_100g] = myDF2[list_ingredient_100g].fillna(value=0)

```

```
In [220]: pourcent_nan = myDF3.isna().sum().divide(myDF3.shape[0]/100).sort_values(ascending=False)
```

```
plt.figure(figsize=(12, 50)) #necessaire pour que les labels soient lisibles
ax = sns.barplot(y = pourcent_nan.index, x=pourcent_nan.values)
ax.xaxis.set_ticks_position('top')
plt.title('Pourcentage de Nan par colonne')
plt.show()
```



```

        salt_100g
phosphorus_100g
        created_t
cocoa_100g
collagen-meat-protein-ratio_100g
fruits-vegetables-nuts_100g
        ph_100g
taurine_100g
caffeine_100g
iodine_100g
molybdenum_100g
chromium_100g
selenium_100g
fluoride_100g
manganese_100g
copper_100g
zinc_100g
magnesium_100g
iron_100g
sodium_100g
nucleotides_100g
last_modified_t
behenic-acid_100g
linoleic-acid_100g
omega-6-fat_100g
docosahexaenoic-acid_100g
eicosapentaenoic-acid_100g
alpha-linolenic-acid_100g
omega-3-fat_100g
monounsaturated-fat_100g
montanic-acid_100g
arachidic-acid_100g
serum-proteins_100g
stearic-acid_100g
palmitic-acid_100g
myristic-acid_100g
lauric-acid_100g
capric-acid_100g
caprylic-acid_100g
saturated-fat_100g
fat_100g
arachidonic-acid_100g
gamma-linolenic-acid_100g
dihomo-gamma-linolenic-acid_100g
omega-9-fat_100g
casein_100g
proteins_100g
fiber_100g
polyols_100g
starch_100g
maltodextrins_100g
maltose_100g
lactose_100g
fructose_100g
glucose_100g
sucrose_100g
sugars_100g
carbohydrates_100g
cholesterol_100g
trans-fat_100g
gondoic-acid_100g
oleic-acid_100g
polyunsaturated-fat_100g

```

In [221]:

```

# on se débarrasse des valeurs au dessus de 100 et en dessous de 0
# en les remplaçant pas 0 et 100

# MENTOR = est-ce la bonne chose à faire?

myDF4 = myDF2.copy() # MENTOR est-ce que je dois faire un deep copy?

nb_exces = 0
nb_intra = 0
nb_OK = 0
for x in list_ingredient_100g :
    if ~(myDF4[x].max() > 100.0) and ~(myDF4[x].min() < 0.0) :
        print("colonne {} a toutes ses valeurs entre 0 et 100g".format(x))
        nb_OK += 1
    else :
        print()
        if myDF4[x].max() > 100.0 :
            valeurs_aberrantes = myDF4[x] > 100.0
            print("colonne {} a {} valeurs supérieures a 100g".format(x, sum(valeurs_aberrantes)))
            print(pd.concat([myDF4["product_name"],myDF4[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
            myDF4.loc[valeurs_aberrantes , x] = 100.0
            nb_exces +=1
        if myDF4[x].min() < 0.0 :
            valeurs_aberrantes = myDF4[x] < 0.0
            print("colonne {} a {} valeurs inférieures a 0g".format(x, sum(valeurs_aberrantes)))
            print(pd.concat([myDF4["product_name"],myDF4[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
            myDF4.loc[valeurs_aberrantes , x] = 0.0
            nb_intra +=1
print("{} colonnes avec valeurs >100g, {} avec valeurs <0g".
      format(nb_exces, nb_intra))

myDF4[list_ingredient_100g] = myDF4[list_ingredient_100g].fillna(value=0)

```

```
# utiliser methode KNN (cf notebook) pour calculer valeur manquantes plutot que redressement maxi
```

```
colonne fat_100g a 4 valeurs supérieures a 100g
```

		product_name	fat_100g
112681	Premium Extra Virgin Olive Oil	714.29	
245311	mini choux goût fromage de chèvre - poivre	380.00	
303101	Ekstra Jomfru Olivenolie	101.00	
305054	Graine de couscous moyen	105.00	

```
colonne saturated-fat_100g a 3 valeurs supérieures a 100g
```

		product_name	saturated-fat_100g
119389	Raw 100% Cacao, With Bits Of Delicate Dates	175.38	
245311	mini choux goût fromage de chèvre - poivre	210.00	
313506	Nusco, Chocolate Spread, Chocolate	550.00	

```
colonne monounsaturated-fat_100g a 1 valeurs supérieures a 100g
```

		product_name	monounsaturated-fat_100g
112681	Premium Extra Virgin Olive Oil	557.14	

```
colonne trans-fat_100g a 5 valeurs supérieures a 100g
```

		product_name	trans-fat_100g
11941	12 Round Roasted Pepper And Smoked Mozzarella ...	369.00	
81625	Pinto Beans	130.00	
81627	Roman Beans	130.00	
120930	Gourmet Ice Cream, Maple Nut	110.00	
138451	Country Fried Beef Steak	173.26	

```
colonne trans-fat_100g a 3 valeurs inférieures a 0g
```

		product_name	trans-fat_100g
23576	Xtra Butter Microwave Pop Corn, Butter	-3.03	
107990	Gelato Truffle On A Stick	-0.70	
120692	Rugen Fisch, Salmon Fillets Skinless And Boneless	-1.00	

```
colonne carbohydrates_100g a 16 valeurs supérieures a 100g
```

		product_name	carbohydrates_100g
71000	Toaster Pastries, Strawberry	209.38	
119783	Naturally Caffeinated Pure Empower Mint Dietar...	120.00	
133567	Mango Jalapeno	125.00	
155609	Beurre d'érable	164.00	
155803	Tamarind Juice With Nata De Coco, Tamarind	2916.67	
202571	Sirop d'Agave brun bio	104.00	
234726	Palets pur beurre	106.00	
239347	Sirop d'agave	105.00	
245311	mini choux goût fromage de chèvre - poivre	190.00	
260598	Agave Syrup dark	103.50	
260599	Sirop d'agave pur	103.50	
260611	Agavendicksaft	103.50	
265108	Agavendicksaft	103.50	
265110	Agavendicksaft Dunkel	103.00	
286482	Sauce Caramel	101.00	
317635	Banane sèche	139.00	

```
colonne sugars_100g a 10 valeurs supérieures a 100g
```

		product_name	sugars_100g
133567	Mango Jalapeno	110.71	
155803	Tamarind Juice With Nata De Coco, Tamarind	166.67	
202571	Sirop d'Agave brun bio	104.00	
239347	Sirop d'agave	105.00	
260598	Agave Syrup dark	103.50	
260599	Sirop d'agave pur	103.50	
260611	Agavendicksaft	103.50	
265108	Agavendicksaft	103.50	
265110	Agavendicksaft Dunkel	103.00	
313506	Nusco, Chocolate Spread, Chocolate	3520.00	

```
colonne sugars_100g a 7 valeurs inférieures a 0g
```

		product_name	sugars_100g
8582	Grade A Fancy Chopped Spinach	-1.20	
18209	Select, Spicy Red Bell Pepper Pasta Sauce	-0.80	
117739	Hummous, Black Truffle	-3.57	
146284	Crackers	-6.67	
150858	Italianavera, Tomato Sauce With Gaeta Olives &...	-6.25	
164030	Organic Pumpkin Seeds	-17.86	
195761	Caprice des dieux	-0.10	

```
colonne fructose_100g a 1 valeurs supérieures a 100g
```

		product_name	fructose_100g
202571	Sirop d'Agave brun bio	101.0	

colonne fiber\_100g a 4 valeurs supérieures a 100g  
product\_name fiber\_100g  
64352 Chef Merito, Creole Style Adobo 166.7  
155803 Tamarind Juice With Nata De Coco, Tamarind 250.0  
298897 Milka Schmunzelhase Alpenmilch 175G 178.0  
313506 Nusco, Chocolate Spread, Chocolate 5380.0  
colonne fiber\_100g a 1 valeurs inférieures a 0g  
product\_name fiber\_100g  
23784 Traditional Tender Cracklins Chicharrones -6.7

colonne proteins\_100g a 1 valeurs supérieures a 100g  
product\_name proteins\_100g  
313506 Nusco, Chocolate Spread, Chocolate 430.0  
colonne proteins\_100g a 2 valeurs inférieures a 0g  
product\_name proteins\_100g  
115310 Gourmet Blends, Seasoning, Garlic Pepper -500.0  
169119 Lightly Dried Cilantro -800.0

colonne salt\_100g a 142 valeurs supérieures a 100g  
product\_name salt\_100g  
501 Mediterranean Sea Salt 105.83418  
1435 Dill Pickle Cashews 870.85678  
3472 Coarse Kosher Salt 101.60000  
7037 Fine Sea Salt 101.60000  
7983 Butter Beans 781.53768  
11129 Autumn Cookie Selection 327.74128  
18588 Mediterranean Sea Salt Grinder 105.83418  
18848 Himalayan Pink Salt 106.68000  
18851 Olde Thompson, Mediterranean Sea Salt 101.60000  
18858 Himalayan Pink Salt 105.66400  
18869 Mediterranean Sea Salt 101.60000  
19014 Adolph's, Prime Cut, Tenderizer Seasoned With ... 101.60000  
20821 Special Blend, Peas & Carrots 714.37500  
21619 Fine Mediterranean Sea Salt 101.60000  
21626 Sea Salt 103.41356  
21627 Mediterranean Sea Salt, Fine 101.60000  
21632 Coarse Kosher Salt 102.00000  
21633 Kosher Salt 101.60000  
21892 Marie Sharp's, Orange Pulp Habanero Pepper Sau... 594.86800  
23590 Original Barbecue Sauce 2032.00000  
25671 Fine Sea Salt 101.60000  
28787 Fratelli Comino, Tochetti With Sea Salt 177.80000  
32371 Light & Fit Greek Crunch - S'mores 153.00000  
33630 Hawaiian Black Sea Salt 103.55326  
39679 Garlic Salt 234.95000  
40408 Fleur De Sel, Hand-Harvested French Sea Salt 101.60000  
40409 Grey Sea Salt 101.60000  
40410 Fine Crystals Sea Salt 101.60000  
40411 Coarse Crystals Sea Salt 101.60000  
40412 Sea Salt 101.60000  
43049 Premium Spice Blends, Seasoning Salt 124.46000  
43772 Coarse Kosher Salt 101.60000  
43773 Fine Sea Salt 101.60000  
49970 Mediterranean Sea Salt 105.83418  
49972 Himalayan Pink Salt 106.68000  
51348 Crispy Fried Onions, White Cheddar 2177.14322  
55514 Pure Carrot With Organic Lemon Juice 158.75000  
58007 String Cheese Snacks 1799.16582  
61552 Sea Salt Grinder 101.60000  
63824 Steak Seasoning Salt 104.14000  
65202 Salt 104.14000  
68379 Rhodes, Bake N Serv, Cinnamon Rolls With Cream... 1098.37728  
68388 Rhodes, Bake N Serv, Cinnamon Rolls With Cream... 858.59112  
68567 Kosher Salt 101.60000  
74261 Fine Sea Salt 101.60000  
74685 Coarse Sea Salt 101.60000  
74686 All Natural Sea Salt 101.60000  
74687 Kosher Sea Salt 101.60000  
75834 Crawfish, Shrimp & Crab Boil 101.60000  
75838 Crawfish, Shrimp And Crab Boil 101.60000  
79161 Pickling & Canning Salt 100.01250  
81378 Adobo Seasoning With Pimento 190.50000  
81379 All Purpose Adobo Seasoning With Sazon 190.50000  
90930 Sea Salt Fine 101.60000  
90982 Coarse Kosher Salt 101.60000  
90985 Nature's Supreme, Pure Fine Sea Salt 101.60000  
93914 La Flor, Grinder Sea Salt 101.60000  
93915 Alderwood Smoked Sea Salt Grinder 101.60000  
94631 Swiss Recipe Smoked Almonds & Sea Salt In Dark... 1318.38192  
105006 Bastoncini Whole Wheat 2452.41318  
106330 Small Potato Dumplings, Potato Gnocchetti 1139.15190  
106360 Fine Sea Salt 113.32210  
106361 Coarse Sea Salt 113.32210  
106544 Sea Salt 107.04322  
106545 Sea Salt, Fine Grind 106.68000  
106880 Popcorn Salt 114.30000  
108367 Natural Yogurt Pretzels 666.75000  
109090 Mediterranean Sea Salt Grinder 105.83418  
109154 Himalayan Pink Salt 107.00000  
109195 Lisy, Seasoned Meat Tenderizer 101.60000  
113216 Himalayan Pink Salt Grinder 106.68000  
115253 All Natural Sea Salt 105.83418  
115865 Mediterranean Sea Saalt 105.83418  
115866 Himalayan Pink Salt 106.68000  
115872 Himalayan Pink Sea Salt 106.68000  
117667 Sea Salt 149.86000  
119185 Premium Iodized Sea Salt 101.60000  
119186 Coarse Crystals 101.60000  
119860 Flower Of Salt, Smoked 108.71200  
119863 Natural Sea Salt, Spicy 100.33000  
120800 Coarse Kosher Salt 101.60000  
120946 Evolution Salt Co, Gourmet Pink Himalaya Salt 254.00000  
120952 Granlund's, Gummi Peaches 123.90120

120960	Granlund's, Poco Brittle	136.07034
120994	Seasoning Salt	677.33418
120995	Anisa's Secrets, Steak Rub	677.33418
120996	Anisa's Secrets, Rosemary Lemon Pepper	677.33418
120999	Anisa's Secrets, Moroccan Spice	677.33418
121005	Anisa's Secrets, Poultry Seasoning	677.33418
121006	Anisa's Secrets, Lemon Dill	677.33418
121766	Primordial Himalayan Sea Salt	112.18418
121767	Primordial Himalayan Sea Salt	107.69600
121892	Inc., California Thompson Raisins	317.50000
122061	Smokey Tomato Speard	604.76130
125882	Mediterranean Sea Salt	101.60000
131415	Gold, Mediterranean Sea Salt	105.83418
132084	Roasted Garlic & Himalayan Pink Salt Seasoning	107.10418
132086	Habanero & Himalayan Pink Salt Seasoning	101.23678
132087	Himalayan Pink Salt Seasoning	112.18418
132090	Sriracha & Himalayan Pink Salt Seasoning	104.86644
135064	Chiles Habaneros Rojos	3556.00000
137618	Piri-Piri Hot Sauce	187.96000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	421.83812
141332	Seasoned Salt	111.76000
141333	Seasoned Salt	111.76000
141334	Pegs, Pink Salt	121.92000
142408	Fine Iodized Sea Salt	103.71582
145844	Prickly Syrup	64312.80000
146800	Pur Beverages, Energy Beverage, Acai Berry	100.85324
147845	Truffle & Salt	109.22000
148914	Sweet Goodies White Chocolate Balls	181.42966
149081	Jim Kelly, Sea Salt Adjustable Grinder	100.01250
150717	Quinoa Pasta Fusilli	133.68528
150718	Quinoa Curvo Rigate Pasta	133.68528
152783	Himalayan Chef, Himalayan Pink Salt Grinder	106.68000
152784	Himalayan Pink Salt	106.68000
155803	Tamarind Juice With Nata De Coco, Tamarind	147.32000
160684	Sausage Patties, Veggie Sausage Patties	1669.14322
165311	Dressing & Marinade	1016.00000
165331	Bacon Salt	118.11000
165332	Molten Hot Salt	118.36400
165333	Garlic And Rosemary Salt	118.11000
165334	Mountain Mushroom Salt	118.11000
166102	Pink Salt	106.68000
170832	Gold, Adjustable Grinder, Pink Himalayan Salt	106.68000
171990	Seasoning	147.32000
172421	Bruschetta Blend	381.00000
172425	Sweet Bread & Butter Pickling Seasoning	296.33418
176694	flake	276.86000
176696	herdhey's	124.46000
177374	Sea Salt	105.83418
177375	Applewood Smoked Sea Salt	101.60000
177376	Himalayan Pink Salt	106.68000
177379	Mediterranean Coarse Sea Salt	105.83418
184025	Himalayan Pink Salt	106.68000
232239	Keeny Bio	211.00000
233372	Sel de Guérande Label Rouge	125.00000
274767	Cheese salad	297.60000
279614	Organic Dill Salt	100.51034
284225	Bamboo Shoots	448.23634
298059	Kids Cristal	930.00000
305114	Tartufi Di Fassia, Grey Salt With White Truffles	106.68000

colonne sodium_100g a 33 valeurs supérieures a 100g		
1435	product_name	sodium_100g
7983	Dill Pickle Cashews	342.857000
11129	Butter Beans	307.692000
20821	Autumn Cookie Selection	129.032000
21892	Special Blend, Peas & Carrots	281.250000
23590	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	234.200000
51348	Original Barbecue Sauce	800.000000
58007	Crispy Fried Onions, White Cheddar	857.143000
68379	String Cheese Snacks	708.333000
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	432.432000
94631	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	338.028000
105006	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	519.048000
106330	Bastoncini Whole Wheat	965.517000
108367	Small Potato Dumplings, Potato Gnocchetti	448.485000
120994	Natural Yogurt Pretzels	262.500000
120995	Seasoning Salt	266.667000
120996	Anisa's Secrets, Steak Rub	266.667000
120999	Anisa's Secrets, Rosemary Lemon Pepper	266.667000
121005	Anisa's Secrets, Moroccan Spice	266.667000
121006	Anisa's Secrets, Poultry Seasoning	266.667000
121892	Anisa's Secrets, Lemon Dill	266.667000
122061	Inc., California Thompson Raisins	125.000000
1272421	Smokey Tomato Speard	238.095000
135064	Chiles Habaneros Rojos	1400.000000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	166.078000
145844	Prickly Syrup	25320.000000
160684	Sausage Patties, Veggie Sausage Patties	657.143000
165311	Dressing & Marinade	400.000000
172421	Bruschetta Blend	150.000000
172425	Sweet Bread & Butter Pickling Seasoning	116.667000
176694	flake	109.000000
274767	Cheese salad	117.165354
284225	Bamboo Shoots	176.471000
298059	Kids Cristal	366.141732

colonne vitamin-a_100g a 1 valeurs inférieures a 0g		
80440	product_name	vitamin-a_100g
Flavor Aid, Soft Drink Mix, Lemon	-0.00034	

colonne vitamin-c_100g a 1 valeurs supérieures a 100g		
product_name	vitamin-c_100g	

```
139242 Organic Fruit Snacks, Berry Medley      370.3704
colonne vitamin-c_100g a 1 valeurs inférieures a 0g
    product_name vitamin-c_100g
41538 Cheez Waffles      -0.0021
```

```
colonne vitamin-b1_100g a 1 valeurs supérieures a 100g
    product_name vitamin-b1_100g
67437 Céréales rice krispies      161.0
```

```
colonne folates_100g a 2 valeurs supérieures a 100g
    product_name folates_100g
3594 Cereal Toasted Oats      166.66667
4494 Toasted Oats      178.571429
```

```
colonne biotin_100g a 3 valeurs inférieures a 0g
    product_name biotin_100g
189152 France      -1.0
189160 France      -2.0
189272 France      -1.0
```

```
colonne pantothenic-acid_100g a 3 valeurs inférieures a 0g
    product_name pantothenic-acid_100g
189152 France      -1.0
189160 France      -2.0
189272 France      -1.0
```

```
colonne potassium_100g a 2 valeurs supérieures a 100g
    product_name potassium_100g
149716 Active Natural Electrolyte Enhanced Drink Tabs      1870.37
172984 Deglet Noor Pitted Dates      600.00
```

```
colonne calcium_100g a 9 valeurs supérieures a 100g
    product_name calcium_100g
10494 Sweeteners, Organic Molasses      522.727
112298 Dulce De Leche, Milk Caramel Spread      285.714
113049 5 Year, Gouda Cheese      125.000
117077 Promenade Red Apple Market, B.B.Q Seasoning      136.000
123679 Sweet Rice      173.333
124143 Tastyshake, Thermodynamic Protein Complex Prot...
124222 Turkish Sun Dried Figs      161.560
165426 Spreadable Gouda Cheese With Red Pepper      432.143
285611 Jaleas Ujarrá's, Jelly, Strawberry      120.000
```

```
colonne phosphorus_100g a 5 valeurs supérieures a 100g
    product_name phosphorus_100g
117363 Soy Protein & Multigrain Cereal      400.000
117364 Simply Fiber, All Natural Crunchy O's Cereal      559.459
123836 Kamut, Premium Khorasan Wheat      410.870
123837 Premium Buckwheat      315.556
123838 Premium Teff      320.000
```

```
colonne iron_100g a 1 valeurs inférieures a 0g
    product_name iron_100g
316892 Lamthong, Fruit Cocktail In Syrup      -0.00026
```

```
colonne magnesium_100g a 6 valeurs supérieures a 100g
    product_name magnesium_100g
19681 Essentials, Organic Ground Chia Seed Dietary S...      333.333
19683 Organic Ground Premium Flaxseed Dietary Supple...
149841 Goji Basil Superfood + Cashews      196.429
149845 Hemp Seeds      573.333
166358 Cacao Nibs      300.000
166363 Raw Cacao Powder      657.143
```

```
colonne selenium_100g a 1 valeurs inférieures a 0g
    product_name selenium_100g
28324 Venus, Cuttlefish Balls      -0.000002
```

```
18 colonnes avec valeurs >100g, 10 avec valeurs <0g
```

```
In [222]: myDF5 = myDF2.copy()

# même fonction que précédemment, mais cette fois on remplace les valeurs aberrantes par NaN

nb_exces = 0
nb_infra = 0
nb_OK= 0
for x in list_ingredient_100g :
    print()
    if myDF5[x].max() > 100.0 :
        valeurs_aberrantes = myDF5[x] > 100.0
        print("colonne {} a {} valeurs supérieures a 100g".format(x, sum(valeurs_aberrantes)))
        print(pd.concat([myDF5["product_name"],myDF5[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        myDF5.loc[valeurs_aberrantes , x] = np.nan
        nb_exces +=1
    if myDF5[x].min() < 0.0 :
```

```

        valeurs_aberrantes = myDF5[x] < 0.0
        print("colonne {} a {} valeurs inférieures a 0g".format(x, sum(valeurs_aberrantes)))
        print(pd.concat([myDF5["product_name"],myDF5[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        myDF5.loc[valeurs_aberrantes , x] = np.nan
        nb_intra +=1
print("{} colonnes avec valeurs >100g, {} avec valeurs <0g".
      format( nb_exces, nb_intra))

colonne fat_100g a 4 valeurs supérieures a 100g
                                         product_name  fat_100g
112681          Premium Extra Virgin Olive Oil    714.29
245311  mini choux goût fromage de chèvre - poivre   380.00
303101           Ekstra Jomfru Olivenolie   101.00
305054            Graine de couscous moyen   105.00

colonne saturated-fat_100g a 3 valeurs supérieures a 100g
                                         product_name saturated-fat_100g
119389  Raw 100% Cacao, With Bits Of Delicate Dates     175.38
245311  mini choux goût fromage de chèvre - poivre     210.00
313506          Nusco, Chocolate Spread, Chocolate    550.00

colonne monounsaturated-fat_100g a 1 valeurs supérieures a 100g
                                         product_name monounsaturated-fat_100g
112681  Premium Extra Virgin Olive Oil                557.14

colonne trans-fat_100g a 5 valeurs supérieures a 100g
                                         product_name trans-fat_100g
11941  12 Round Roasted Pepper And Smoked Mozzarella ...    369.00
81625                  Pinto Beans                   130.00
81627                  Roman Beans                  130.00
120930          Gourmet Ice Cream, Maple Nut       110.00
138451  Country Fried Beef Steak                 173.26
colonne trans-fat_100g a 3 valeurs inférieures a 0g
                                         product_name trans-fat_100g
23576          Xtra Butter Microwave Pop Corn, Butter      -3.03
107990          Gelato Truffle On A Stick      -0.70
120692  Rugen Fisch, Salmon Fillets Skinless And Boneless   -1.00

colonne carbohydrates_100g a 16 valeurs supérieures a 100g
                                         product_name carbohydrates_100g
71000          Toaster Pastries, Strawberry             209.38
119783  Naturally Caffeinated Pure Empower Mint Dietar...   120.00
133567                  Mango Jalapeno                125.00
155609          Beurre d'érable                     164.00
155803  Tamarind Juice With Nata De Coco, Tamarind      2916.67
202571          Sirop d'Agave brun bio                 104.00
234726          Palets pur beurre                  106.00
239347          Sirop d'agave                      105.00
245311  mini choux goût fromage de chèvre - poivre     190.00
260598          Agave Syrup dark                  103.50
260599          Sirop d agave pur                  103.50
260611          Agavendicksaft                  103.50
265108          Agavendicksaft                  103.50
265110          Agavendicksaft Dunkel                103.00
286482          Sauce Caramel                    101.00
317635          Banane sèche                      139.00

colonne sugars_100g a 10 valeurs supérieures a 100g
                                         product_name sugars_100g
133567          Mango Jalapeno                  110.71
155803  Tamarind Juice With Nata De Coco, Tamarind      166.67
202571          Sirop d'Agave brun bio                 104.00
239347          Sirop d'agave                      105.00
260598          Agave Syrup dark                  103.50
260599          Sirop d agave pur                  103.50
260611          Agavendicksaft                  103.50
265108          Agavendicksaft                  103.50
265110          Agavendicksaft Dunkel                103.00
313506          Nusco, Chocolate Spread, Chocolate    3520.00
colonne sugars_100g a 7 valeurs inférieures a 0g
                                         product_name sugars_100g
8582          Grade A Fancy Chopped Spinach            -1.20
18209          Select, Spicy Red Bell Pepper Pasta Sauce   -0.80
117739          Hummous, Black Truffle                  -3.57
146284          Crackers                           -6.67
150858  Italianavera, Tomato Sauce With Gaeta Olives &...   -6.25
164030          Organic Pumpkin Seeds                -17.86
195761          Caprice des dieux                  -0.10

colonne fructose_100g a 1 valeurs supérieures a 100g
                                         product_name fructose_100g
202571  Sirop d'Agave brun bio                  101.0

```

colonne fiber_100g a 4 valeurs supérieures a 100g		
	product_name	fiber_100g
64352	Chef Merito, Creole Style Adobo	166.7
155803	Tamarind Juice With Nata De Coco, Tamarind	250.0
298897	Milka Schmunzelhase Alpenmilch 175G	178.0
313506	Nusco, Chocolate Spread, Chocolate	5380.0
colonne fiber_100g a 1 valeurs inférieures a 0g		
	product_name	fiber_100g
23784	Traditional Tender Cracklins Chicharrones	-6.7
colonne proteins_100g a 1 valeurs supérieures a 100g		
	product_name	proteins_100g
313506	Nusco, Chocolate Spread, Chocolate	430.0
colonne proteins_100g a 2 valeurs inférieures a 0g		
	product_name	proteins_100g
115310	Gourmet Blends, Seasoning, Garlic Pepper	-500.0
169119	Lightly Dried Cilantro	-800.0

colonne salt_100g a 142 valeurs supérieures a 100g		
	product_name	salt_100g
501	Mediterranean Sea Salt	105.83418
1435	Dill Pickle Cashews	870.85678
3472	Coarse Kosher Salt	101.60000
7037	Fine Sea Salt	101.60000
7983	Butter Beans	781.53768
11129	Autumn Cookie Selection	327.74128
18588	Mediterranean Sea Salt Grinder	105.83418
18848	Himalayan Pink Salt	106.68000
18851	Olde Thompson, Mediterranean Sea Salt	101.60000
18858	Himalayan Pink Salt	105.66400
18869	Mediterranean Sea Salt	101.60000
19014	Adolph's, Prime Cut, Tenderizer Seasoned With ...	101.60000
20821	Special Blend, Peas & Carrots	714.37500
21619	Fine Mediterranean Sea Salt	101.60000
21626	Sea Salt	103.41356
21627	Mediterranean Sea Salt, Fine	101.60000
21632	Coarse Kosher Salt	102.00000
21633	Kosher Salt	101.60000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	594.86800
23590	Original Barbecue Sauce	2032.00000
25671	Fine Sea Salt	101.60000
28787	Fratelli Comino, Tochetti With Sea Salt	177.80000
32371	Light & Fit Greek Crunch - S'mores	153.00000
33630	Hawaiian Black Sea Salt	103.55326
39679	Garlic Salt	234.95000
40408	Fleur De Sel, Hand-Harvested French Sea Salt	101.60000
40409	Grey Sea Salt	101.60000
40410	Fine Crystals Sea Salt	101.60000
40411	Coarse Crystals Sea Salt	101.60000
40412	Sea Salt	101.60000
43049	Premium Spice Blends, Seasoning Salt	124.46000
43772	Coarse Kosher Salt	101.60000
43773	Fine Sea Salt	101.60000
49970	Mediterranean Sea Salt	105.83418
49972	Himalayan Pink Salt	106.68000
51348	Crispy Fried Onions, White Cheddar	2177.14322
55514	Pure Carrot With Organic Lemon Juice	158.75000
58007	String Cheese Snacks	1799.16582
61552	Sea Salt Grinder	101.60000
63824	Steak Seasoning Salt	104.14000
65202	Salt	104.14000
68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	1098.37728
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	858.59112
68507	Kosher Salt	101.60000
74261	Fine Sea Salt	101.60000
74685	Coarse Sea Salt	101.60000
74686	All Natural Sea Salt	101.60000
74687	Kosher Sea Salt	101.60000
75834	Crawfish, Shrimp & Crab Boil	101.60000
75838	Crawfish, Shrimp And Crab Boil	101.60000
79161	Pickling & Canning Salt	100.01250
81378	Adobe Seasoning With Pimento	190.50000
81379	All Purpose Adobo Seasoning With Sazon	190.50000
90930	Sea Salt Fine	101.60000
90982	Coarse Kosher Salt	101.60000
90985	Nature's Supreme, Pure Fine Sea Salt	101.60000
93914	La Flor, Grinder Sea Salt	101.60000
93915	Alderwood Smoked Sea Salt Grinder	101.60000
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	1318.38192
105006	Bastoncini Whole Wheat	2452.41318
106330	Small Potato Dumplings, Potato Gnocchetti	1139.15190
106360	Fine Sea Salt	113.32210
106361	Coarse Sea Salt	113.32210
106544	Sea Salt	107.04322
106545	Sea Salt, Fine Grind	106.68000
106880	Popcorn Salt	114.30000
108367	Natural Yogurt Pretzels	666.75000
109090	Mediterranean Sea Salt Grinder	105.83418
109154	Himalayan Pink Salt	107.00000
109195	Lisy, Seasoned Meat Tenderizer	101.60000
113216	Himalayan Pink Salt Grinder	106.68000
115253	All Natural Sea Salt	105.83418
115865	Mediterranean Sea Salt	105.83418
115866	Himalayan Pink Salt	106.68000
115872	Himalayan Pink Sea Salt	106.68000
117667	Sea Salt	149.86000
119185	Premium Iodized Sea Salt	101.60000
119186	Coarse Crystals	101.60000

119860	Flower Of Salt, Smoked	108.71200
119863	Natural Sea Salt, Spicy	100.33000
120800	Coarse Kosher Salt	101.60000
120946	Evolution Salt Co, Gourmet Pink Himalaya Salt	254.00000
120952	Granlund's, Gummi Peaches	123.90120
120960	Granlund's, Peco Brittle	136.07034
120994	Seasoning Salt	677.33418
120995	Anisa's Secrets, Steak Rub	677.33418
120996	Anisa's Secrets, Rosemary Lemon Pepper	677.33418
120999	Anisa's Secrets, Moroccan Spice	677.33418
121005	Anisa's Secrets, Poultry Seasoning	677.33418
121006	Anisa's Secrets, Lemon Dill	677.33418
121786	Primordial Himalayan Sea Salt	112.18418
121787	Primordial Himalayan Sea Salt	107.69600
121892	Inc., California Thompson Raisins	317.50000
122061	Smokey Tomato Speard	604.76130
125882	Mediterranean Sea Salt	101.60000
131415	Gold, Mediterranean Sea Salt	105.83418
132084	Roasted Garlic & Himalayan Pink Salt Seasoning	107.10418
132086	Habanero & Himalayan Pink Salt Seasoning	101.23678
132087	Himalayan Pink Salt Seasoning	112.18418
132090	Sriracha & Himalayan Pink Salt Seasoning	104.86644
135064	Chiles Habaneros Rojos	3556.00000
137618	Piri-Piri Hot Sauce	187.96000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	421.83812
141332	Seasoned Salt	111.76000
141333	Seasoned Salt	111.76000
141334	Pegs, Pink Salt	121.92000
142408	Fine Iodized Sea Salt	103.71582
145844	Prickly Syrup	64312.80000
146800	Pur Beverages, Energy Beverage, Acai Berry	100.85324
147845	Truffle & Salt	109.22000
148914	Sweet Goodies White Chocolate Balls	181.42966
149081	Jim Kelly, Sea Salt Adjustable Grinder	100.01250
150717	Quinoa Pasta Fusilli	133.68528
150718	Quinoa Curvo Rigate Pasta	133.68528
152783	Himalayan Chef, Himalayan Pink Salt Grinder	106.68000
152784	Himalayan Pink Salt	106.68000
155803	Tamarind Juice With Nata De Coco, Tamarind	147.32000
160684	Sausage Patties, Veggie Sausage Patties	1669.14322
165311	Dressing & Marinade	1016.00000
165331	Bacon Salt	118.11000
165332	Molten Hot Salt	118.36400
165333	Garlic And Rosemary Salt	118.11000
165334	Mountain Mushroom Salt	118.11000
166102	Pink Salt	106.68000
170832	Gold, Adjustable Grinder, Pink Himalayan Salt	106.68000
171990	Seasoning	147.32000
172421	Bruschetta Blend	381.00000
172425	Sweet Bread & Butter Pickling Seasoning	296.33418
176694	flake	276.86000
176696	herdhey's	124.46000
177374	Sea Salt	105.83418
177375	Applewood Smoked Sea Salt	101.60000
177376	Himalayan Pink Salt	106.68000
177379	Mediterranean Coarse Sea Salt	105.83418
184025	Himalayan Pink Salt	106.68000
232239	Keeny Bio	211.00000
233372	Sel de Guérande Label Rouge	125.00000
274767	Cheese salad	297.60000
279614	Organic Dill Salt	100.51034
284225	Bamboo Shoots	448.23634
298059	Kids Cristal	930.00000
305114	Tartufi Di Fassia, Grey Salt With White Truffles	106.68000

colonne sodium\_100g a 33 valeurs supérieures a 100g

	product_name	sodium_100g
1435	Dill Pickle Cashews	342.857000
7983	Butter Beans	307.692000
11129	Autumn Cookie Selection	129.032000
20821	Special Blend, Peas & Carrots	281.250000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	234.200000
23590	Original Barbecue Sauce	800.000000
51348	Crispy Fried Onions, White Cheddar	857.143000
58007	String Cheese Snacks	708.333000
68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	432.432000
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	338.028000
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	519.048000
105006	Bastoncini Whole Wheat	965.517000
106330	Small Potato Dumplings, Potato Gnocchetti	448.485000
108367	Natural Yogurt Pretzels	262.500000
120994	Seasoning Salt	266.667000
120995	Anisa's Secrets, Steak Rub	266.667000
120996	Anisa's Secrets, Rosemary Lemon Pepper	266.667000
120999	Anisa's Secrets, Moroccan Spice	266.667000
121005	Anisa's Secrets, Poultry Seasoning	266.667000
121006	Anisa's Secrets, Lemon Dill	266.667000
121892	Inc., California Thompson Raisins	125.000000
122061	Smokey Tomato Speard	238.095000
135064	Chiles Habaneros Rojos	1400.000000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	166.078000
145844	Prickly Syrup	25320.000000
160684	Sausage Patties, Veggie Sausage Patties	657.143000
165311	Dressing & Marinade	400.000000
172421	Bruschetta Blend	150.000000
172425	Sweet Bread & Butter Pickling Seasoning	116.667000
176694	flake	109.000000
274767	Cheese salad	117.165354
284225	Bamboo Shoots	176.471000
298059	Kids Cristal	366.141732

colonne vitamin-a\_100g a 1 valeurs inférieures a 0g

product_name	vitamin-a_100g
80440 Flavor Aid, Soft Drink Mix, Lemon	-0.00034

```

colonne vitamin-c_100g a 1 valeurs supérieures a 100g
    product_name vitamin-c_100g
139242 Organic Fruit Snacks, Berry Medley      370.3704
colonne vitamin-c_100g a 1 valeurs inférieures a 0g
    product_name vitamin-c_100g
41538 Cheez Waffies      -0.0021

colonne vitamin-b1_100g a 1 valeurs supérieures a 100g
    product_name vitamin-b1_100g
67437 Céréales rice krispies      161.0


colonne folates_100g a 2 valeurs supérieures a 100g
    product_name folates_100g
3594 Cereal Toasted Oats      166.666667
4494 Toasted Oats      178.571429

colonne biotin_100g a 3 valeurs inférieures a 0g
    product_name biotin_100g
189152 France      -1.0
189160 France      -2.0
189272 France      -1.0

colonne pantothenic-acid_100g a 3 valeurs inférieures a 0g
    product_name pantothenic-acid_100g
189152 France      -1.0
189160 France      -2.0
189272 France      -1.0

colonne potassium_100g a 2 valeurs supérieures a 100g
    product_name potassium_100g
149716 Active Natural Electrolyte Enhanced Drink Tabs      1870.37
172984 Deglet Noor Pitted Dates      600.00

colonne calcium_100g a 9 valeurs supérieures a 100g
    product_name calcium_100g
10494 Sweeteners, Organic Molasses      522.727
112298 Dulce De Leche, Milk Caramel Spread      285.714
113049 5 Year, Gouda Cheese      125.000
117077 Promenade Red Apple Market, B.B.Q Seasoning      136.000
123679 Sweet Rice      173.333
124143 Tastyshake, Thermodynamic Protein Complex Prot...      613.636
124222 Turkish Sun Dried Figs      161.560
165426 Spreadable Gouda Cheese With Red Pepper      432.143
285611 Jaleas Ujarrá's, Jelly, Strawberry      120.000

colonne phosphorus_100g a 5 valeurs supérieures a 100g
    product_name phosphorus_100g
117363 Soy Protein & Multigrain Cereal      400.000
117364 Simply Fiber, All Natural Crunchy O's Cereal      559.459
123836 Kamut, Premium Khorasan Wheat      410.870
123837 Premium Buckwheat      315.556
123838 Premium Teff      320.000

colonne iron_100g a 1 valeurs inférieures a 0g
    product_name iron_100g
316892 Lamthong, Fruit Cocktail In Syrup      -0.00026

colonne magnesium_100g a 6 valeurs supérieures a 100g
    product_name magnesium_100g
19681 Essentials, Organic Ground Chia Seed Dietary S...      333.333
19683 Organic Ground Premium Flaxseed Dietary Supple...      364.286
149841 Goji Basil Superfood + Cashews      196.429
149845 Hemp Seeds      573.333
166358 Cacao Nibs      300.000
166363 Raw Cacao Powder      657.143


colonne selenium_100g a 1 valeurs inférieures a 0g
    product_name selenium_100g
28324 Venus, Cuttlefish Balls      -0.000002

```

18 colonnes avec valeurs >100g, 10 avec valeurs <0g

```

In [223]: myDF5.select_dtypes(['int32','float64']).columns
Out[223]: Index(['additives_n', 'ingredients_from_palm_oil_n',
   'ingredients_that_may_be_from_palm_oil_n', 'energy_100g',
   'energy-from-fat_100g', 'fat_100g', 'saturated-fat_100g',
   'caprylic-acid_100g', 'capric-acid_100g', 'lauric-acid_100g',
   'myristic-acid_100g', 'palmitic-acid_100g', 'stearic-acid_100g',
   'arachidic-acid_100g', 'behenic-acid_100g', 'montanic-acid_100g',
   'monounsaturated-fat_100g', 'polyunsaturated-fat_100g',
   'omega-3-fat_100g', 'alpha-linolenic-acid_100g',
   'eicosapentaenoic-acid_100g', 'docosahexaenoic-acid_100g',
   'vitamin-a_100g', 'vitamin-d_100g', 'vitamin-e_100g',
   'vitamin-k_100g', 'vitamin-b1_100g', 'vitamin-b2_100g',
   'vitamin-b3_100g', 'vitamin-b5_100g', 'vitamin-b6_100g',
   'vitamin-b9_100g', 'vitamin-b12_100g', 'vitamin-c_100g',
   'folates_100g', 'biotin_100g', 'pantothenic-acid_100g',
   'potassium_100g', 'calcium_100g', 'phosphorus_100g',
   'iron_100g', 'magnesium_100g', 'selenium_100g'],
  dtype='object')

```

```
'omega-6-fat_100g', 'linoleic-acid_100g', 'arachidonic-acid_100g',
'gamma-linolenic-acid_100g', 'dihomo-gamma-linolenic-acid_100g',
'omega-9-fat_100g', 'oleic-acid_100g', 'gondoic-acid_100g',
'trans-fat_100g', 'cholesterol_100g', 'carbohydrates_100g',
'sugars_100g', 'sucrose_100g', 'glucose_100g', 'fructose_100g',
'lactose_100g', 'maltose_100g', 'maltodextrins_100g', 'starch_100g',
'polyols_100g', 'fiber_100g', 'proteins_100g', 'casein_100g',
'serum-proteins_100g', 'nucleotides_100g', 'salt_100g', 'sodium_100g',
'alcohol_100g', 'vitamin-a_100g', 'beta-carotene_100g',
'vitamin-d_100g', 'vitamin-e_100g', 'vitamin-k_100g', 'vitamin-c_100g',
'vitamin-b1_100g', 'vitamin-b2_100g', 'vitamin-pp_100g',
'vitamin-b6_100g', 'vitamin-b9_100g', 'folates_100g',
'vitamin-b12_100g', 'biotin_100g', 'pantothenic-acid_100g',
'silica_100g', 'bicarbonate_100g', 'potassium_100g', 'chloride_100g',
'calcium_100g', 'phosphorus_100g', 'iron_100g', 'magnesium_100g',
'zinc_100g', 'copper_100g', 'manganese_100g', 'fluoride_100g',
'selenium_100g', 'chromium_100g', 'molybdenum_100g', 'iodine_100g',
'caffeine_100g', 'taurine_100g', 'ph_100g',
'fruits-vegetables-nuts_100g', 'collagen-meat-protein-ratio_100g',
'cocoa_100g', 'carbon-footprint_100g', 'nutrition-score-fr_100g',
'nutrition-score-uk_100g'],
dtype='object')
```

In [224]: `len(myDF5.select_dtypes(['int32', 'float64']).columns)`

Out[224]: 90

In [225]: `len(list_ingredient_100g)`

Out[225]: 82

In [226]: `myDF6 = myDF5[list_ingredient_100g]`

In [ ]:

In [227]: # je travaille sur la liste des ingrédients plutot que toutes les colonnes numeriques

```
from sklearn.impute import KNNImputer

imputer = KNNImputer(n_neighbors=5, weights="distance")
myDF6 = pd.DataFrame(imputer.fit_transform(myDF6), columns = myDF6.columns)
myDF6
```

Out[227]:

	fat_100g	saturated-fat_100g	caprylic-acid_100g	capric-acid_100g	lauric-acid_100g	myristic-acid_100g	palmitic-acid_100g	stearic-acid_100g	arachidic-acid_100g	behenic-acid_100g	montanic-acid_100g	monounsaturated-fat_100g	polyunsaturated-fat_100g	omega-3-fat_100g
0	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	28.57	28.57	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	17.86	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	57.14	5.36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.43	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
284225	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
284226	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
284227	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
284228	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
284229	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

284230 rows × 82 columns

In [228]: `myDF7 = myDF5.copy()
myDF7[list_ingredient_100g] = myDF6
myDF7`

Out[228]:

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	qua
0	3087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir	NaN	
1	4530	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Banana Chips Sweetened (Whole)	NaN	
2	4559	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Peanuts	NaN	

		code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	qua
3	16087	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055731	2017-03-09T10:35:31Z	1489055731	2017-03-09T10:35:31Z	Organic Salted Nut Mix		NaN	
4	16094	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055653	2017-03-09T10:34:13Z	1489055653	2017-03-09T10:34:13Z	Organic Polenta		NaN	
...	...	...	...	...	...	...	...	...	...	...	
320767	9948282780603	http://world-fr.openfoodfacts.org/produit/9948...	openfoodfacts-contributors	1490631299	2017-03-27T16:14:59Z	1491244498	2017-04-03T18:34:58Z	Tomato & ricotta		NaN	
320768	99567453	http://world-fr.openfoodfacts.org/produit/9956...	usda-ndb-import	1489059076	2017-03-09T11:31:16Z	1491244499	2017-04-03T18:34:59Z	Mint Melange Tea A Blend Of Peppermint, Lemon ...		NaN	
320769	9970229501521	http://world-fr.openfoodfacts.org/produit/9970...	tomato	1422099377	2015-01-24T11:36:17Z	1491244499	2017-04-03T18:34:59Z	乐吧泡菜味薯片	Leba pickle flavor potato chips		
320770	9980282863788	http://world-fr.openfoodfacts.org/produit/9980...	openfoodfacts-contributors	1492340089	2017-04-16T10:54:49Z	1492340089	2017-04-16T10:54:49Z	Tomates aux Vermicelles		NaN	
320771	999990026839	http://world-fr.openfoodfacts.org/produit/9999...	usda-ndb-import	1489072709	2017-03-09T15:18:29Z	1491244499	2017-04-03T18:34:59Z	Sugar Free Drink Mix, Peach Tea		NaN	

284230 rows × 146 columns

```
In [229...]: myDF7.shape
Out[229...]: (284230, 146)

In [230...]: # Regardons ce que cela a changé :
myDF5 = myDF2.copy()

nb_exces = 0
nb_infra = 0
nb_OK= 0
for x in list_ingredient_100g :
    print()
    if myDF5[x].max() > 100.0 :
        valeurs_aberrantes = myDF5[x] > 100.0
        print("colonne {} a {} valeurs supérieures à 100g".format(x, sum(valeurs_aberrantes)))
        print(pd.concat([myDF5["product_name"],myDF5[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        print("Changed to ",pd.concat([myDF7["product_name"],myDF7[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        nb_exces +=1
    if myDF5[x].min() < 0.0 :
        valeurs_aberrantes = myDF5[x] < 0.0
        print("colonne {} a {} valeurs inférieures à 0g".format(x, sum(valeurs_aberrantes)))
        print(pd.concat([myDF5["product_name"],myDF5[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        print("Changed to ", pd.concat([myDF7["product_name"],myDF7[x]],join = "outer", axis = 1).loc[valeurs_aberrantes])
        nb_infra +=1
    print("{} colonnes avec valeurs >100g, {} avec valeurs <0g".
          format( nb_exces, nb_infra))

colonne fat_100g a 4 valeurs supérieures à 100g
product_name fat_100g
112681 Premium Extra Virgin Olive Oil 714.29
245311 mini choux goût fromage de chèvre - poivre 380.00
303101 Ekstra Jomfru Olivenolie 101.00
305054 Graine de couscous moyen 105.00
Changed to product_name fat_100g
112681 Premium Extra Virgin Olive Oil 22.22
245311 mini choux goût fromage de chèvre - poivre 2.68
303101 Ekstra Jomfru Olivenolie NaN
305054 Graine de couscous moyen NaN

colonne saturated-fat_100g a 3 valeurs supérieures à 100g
product_name saturated-fat_100g
119389 Raw 100% Cacao, With Bits Of Delicate Dates 175.38
245311 mini choux goût fromage de chèvre - poivre 210.00
313506 Nusco, Chocolate Spread, Chocolate 550.00
Changed to product_name saturated-fat_100g
119389 Raw 100% Cacao, With Bits Of Delicate Dates 0.0
245311 mini choux goût fromage de chèvre - poivre 0.0
313506 Nusco, Chocolate Spread, Chocolate NaN
```

colonne monounsaturated-fat\_100g a 1 valeurs supérieures a 100g  
product\_name monounsaturated-fat\_100g  
112681 Premium Extra Virgin Olive Oil 557.14  
Changed to product\_name monounsaturated-fat\_100g  
112681 Premium Extra Virgin Olive Oil 0.0

colonne trans-fat\_100g a 5 valeurs supérieures a 100g  
product\_name trans-fat\_100g  
11941 12 Round Roasted Pepper And Smoked Mozzarella ... 369.00  
81625 Pinto Beans 130.00  
81627 Roman Beans 130.00  
120930 Gourmet Ice Cream, Maple Nut 110.00  
138451 Country Fried Beef Steak 173.26  
Changed to product\_name trans-fat\_100g  
11941 12 Round Roasted Pepper And Smoked Mozzarella ... 0.0  
81625 Pinto Beans 0.0  
81627 Roman Beans 0.0  
120930 Gourmet Ice Cream, Maple Nut 0.0  
138451 Country Fried Beef Steak 0.0  
colonne trans-fat\_100g a 3 valeurs inférieures a 0g  
product\_name trans-fat\_100g  
23576 Xtra Butter Microwave Pop Corn, Butter -3.03  
107990 Gelato Truffle On A Stick -0.70  
120692 Rugen Fisch, Salmon Fillets Skinless And Boneless -1.00  
Changed to product\_name trans-fat\_100g  
23576 Xtra Butter Microwave Pop Corn, Butter 0.0  
107990 Gelato Truffle On A Stick 0.0  
120692 Rugen Fisch, Salmon Fillets Skinless And Boneless 0.0

colonne carbohydrates\_100g a 16 valeurs supérieures a 100g  
product\_name carbohydrates\_100g  
71000 Toaster Pastries, Strawberry 209.38  
119783 Naturally Caffeinated Pure Empower Mint Dietar... 120.00  
133567 Mango Jalapeno 125.00  
155609 Beurre d'éable 164.00  
155803 Tamarind Juice With Nata De Coco, Tamarind 2916.67  
202571 Sirop d'Agave brun bio 104.00  
234726 Palets pur beurre 106.00  
239347 Sirop d'agave 105.00  
245311 mini choux goût fromage de chèvre - poivre 190.00  
260598 Agave Syrup dark 103.50  
260599 Sirop d'agave pur 103.50  
260611 Agavendicksaft 103.50  
265108 Agavendicksaft 103.50  
265110 Agavendicksaft Dunkel 103.00  
286482 Sauce Caramel 101.00  
317635 Banane sèche 139.00  
Changed to product\_name carbohydrates\_100g  
71000 Toaster Pastries, Strawberry 6.67  
119783 Naturally Caffeinated Pure Empower Mint Dietar... 4.58  
133567 Mango Jalapeno 28.24  
155609 Beurre d'éable 8.33  
155803 Tamarind Juice With Nata De Coco, Tamarind 12.77  
202571 Sirop d'Agave brun bio 0.00  
234726 Palets pur beurre 0.50  
239347 Sirop d'agave 1.00  
245311 mini choux goût fromage de chèvre - poivre 73.21  
260598 Agave Syrup dark 73.30  
260599 Sirop d'agave pur 0.00  
260611 Agavendicksaft 49.20  
265108 Agavendicksaft 15.00  
265110 Agavendicksaft Dunkel 13.00  
286482 Sauce Caramel NaN  
317635 Banane sèche NaN

colonne sugars\_100g a 10 valeurs supérieures a 100g  
product\_name sugars\_100g  
133567 Mango Jalapeno 110.71  
155803 Tamarind Juice With Nata De Coco, Tamarind 166.67  
202571 Sirop d'Agave brun bio 104.00  
239347 Sirop d'agave 105.00  
260598 Agave Syrup dark 103.50  
260599 Sirop d'agave pur 103.50  
260611 Agavendicksaft 103.50  
265108 Agavendicksaft 103.50  
265110 Agavendicksaft Dunkel 103.00  
313506 Nusco, Chocolate Spread, Chocolate 3520.00  
Changed to product\_name sugars\_100g  
133567 Mango Jalapeno 14.12  
155803 Tamarind Juice With Nata De Coco, Tamarind 12.34  
202571 Sirop d'Agave brun bio 0.00  
239347 Sirop d'agave 1.00  
260598 Agave Syrup dark 23.70  
260599 Sirop d'agave pur 0.00  
260611 Agavendicksaft 44.60  
265108 Agavendicksaft 4.50  
265110 Agavendicksaft Dunkel 4.00  
313506 Nusco, Chocolate Spread, Chocolate NaN

colonne sugars\_100g a 7 valeurs inférieures a 0g

	product_name	sugars_100g
8582	Grade A Fancy Chopped Spinach	-1.20
18209	Select, Spicy Red Bell Pepper Pasta Sauce	-0.80
117739	Hummous, Black Truffle	-3.57
146284	Crackers	-6.67
150858	Italianavera, Tomato Sauce With Gaeta Olives &...	-6.25
164030	Organic Pumpkin Seeds	-17.86
195761	Caprice des dieux	-0.10

Changed to

	product_name	sugars_100g
8582	Grade A Fancy Chopped Spinach	3.64
18209	Select, Spicy Red Bell Pepper Pasta Sauce	37.50
117739	Hummous, Black Truffle	41.00
146284	Crackers	36.67
150858	Italianavera, Tomato Sauce With Gaeta Olives &...	3.64
164030	Organic Pumpkin Seeds	0.00
195761	Caprice des dieux	11.00

colonne fructose\_100g a 1 valeurs supérieures a 100g

	product_name	fructose_100g
202571	Sirop d'Agave brun bio	101.0

Changed to

	product_name	fructose_100g
202571	Sirop d'Agave brun bio	0.0

colonne fiber\_100g a 4 valeurs supérieures a 100g

	product_name	fiber_100g
64352	Chef Merito, Creole Style Adobo	166.7
155803	Tamarind Juice With Nata De Coco, Tamarind	250.0
298897	Milka Schmunzelhase Alpenmilch 175G	178.0
313506	Nusco, Chocolate Spread, Chocolate	5380.0

Changed to

	product_name	fiber_100g
64352	Chef Merito, Creole Style Adobo	2.5
155803	Tamarind Juice With Nata De Coco, Tamarind	0.0
298897	Milka Schmunzelhase Alpenmilch 175G	NaN
313506	Nusco, Chocolate Spread, Chocolate	NaN

colonne fiber\_100g a 1 valeurs inférieures a 0g

	product_name	fiber_100g
23784	Traditional Tender Cracklins Chicharrones	-6.7

Changed to

	product_name	fiber_100g
23784	Traditional Tender Cracklins Chicharrones	3.3

colonne proteins\_100g a 1 valeurs supérieures a 100g

	product_name	proteins_100g
313506	Nusco, Chocolate Spread, Chocolate	430.0

Changed to

	product_name	proteins_100g
313506	Nusco, Chocolate Spread, Chocolate	NaN

colonne proteins\_100g a 2 valeurs inférieures a 0g

	product_name	proteins_100g
115310	Gourmet Blends, Seasoning, Garlic Pepper	-500.0
169119	Lightly Dried Cilantro	-800.0

Changed to

	product_name	proteins_100g
115310	Gourmet Blends, Seasoning, Garlic Pepper	0.0
169119	Lightly Dried Cilantro	5.2

colonne salt\_100g a 142 valeurs supérieures a 100g

	product_name	salt_100g
501	Mediterranean Sea Salt	105.83418
1435	Dill Pickle Cashews	870.85678
3472	Coarse Kosher Salt	101.60000
7037	Fine Sea Salt	101.60000
7983	Butter Beans	781.53768
11129	Autumn Cookie Selection	327.74128
18588	Mediterranean Sea Salt Grinder	105.83418
18848	Himalayan Pink Salt	106.68000
18851	Olde Thompson, Mediterranean Sea Salt	101.60000
18858	Himalayan Pink Salt	105.66400
18869	Mediterranean Sea Salt	101.60000
19014	Adolph's, Prime Cut, Tenderizer Seasoned With ...	101.60000
20821	Special Blend, Peas & Carrots	714.37500
21619	Fine Mediterranean Sea Salt	101.60000
21626	Sea Salt	103.41356
21627	Mediterranean Sea Salt, Fine	101.60000
21632	Coarse Kosher Salt	102.00000
21633	Kosher Salt	101.60000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	594.86800
23590	Original Barbecue Sauce	2032.00000
25671	Fine Sea Salt	101.60000
28787	Fratelli Comino, Tochetti With Sea Salt	177.80000
32371	Light & Fit Greek Crunch - S'mores	153.00000
33630	Hawaiian Black Sea Salt	103.55326
39679	Garlic Salt	234.95000
40408	Fleur De Sel, Hand-Harvested French Sea Salt	101.60000
40409	Grey Sea Salt	101.60000
40410	Fine Crystals Sea Salt	101.60000
40411	Coarse Crystals Sea Salt	101.60000
40412	Sea Salt	101.60000
43049	Premium Spice Blends, Seasoning Salt	124.46000
43772	Coarse Kosher Salt	101.60000
43773	Fine Sea Salt	101.60000
49979	Mediterranean Sea Salt	105.83418
49972	Himalayan Pink Salt	106.68000
51348	Crispy Fried Onions, White Cheddar	2177.14322
55514	Pure Carrot With Organic Lemon Juice	158.75000
58007	String Cheese Snacks	1799.16582
61552	Sea Salt Grinder	101.60000
63824	Steak Seasoning Salt	104.14000
65202	Salt	104.14000

68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	1098.37728
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	858.59112
68507	Kosher Salt	101.60000
74261	Fine Sea Salt	101.60000
74685	Coarse Sea Salt	101.60000
74686	All Natural Sea Salt	101.60000
74687	Kosher Sea Salt	101.60000
75834	Crawfish, Shrimp & Crab Boil	101.60000
75838	Crawfish, Shrimp And Crab Boil	101.60000
79161	Pickling & Canning Salt	100.01250
81378	Adobo Seasoning With Pimento	190.50000
81379	All Purpose Adobo Seasoning With Sazon	190.50000
90930	Sea Salt Fine	101.60000
90982	Coarse Kosher Salt	101.60000
90985	Nature's Supreme, Pure Fine Sea Salt	101.60000
93914	La Flor, Grinder Sea Salt	101.60000
93915	Alderwood Smoked Sea Salt Grinder	101.60000
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	1318.38192
105006	Bastoncini Whole Wheat	2452.41318
106330	Small Potato Dumplings, Potato Gnocchetti	1139.15190
106360	Fine Sea Salt	113.32210
106361	Coarse Sea Salt	113.32210
106544	Sea Salt	107.04322
106545	Sea Salt, Fine Grind	106.68000
106880	Popcorn Salt	114.30000
108367	Natural Yogurt Pretzels	666.75000
109090	Mediterranean Sea Salt Grinder	105.83418
109154	Himalayan Pink Salt	107.00000
109195	Lisy, Seasoned Meat Tenderizer	101.60000
113216	Himalayan Pink Salt Grinder	106.68000
115253	All Natural Sea Salt	105.83418
115865	Mediterranean Sea Saalt	105.83418
115866	Himalayan Pink Salt	106.68000
115872	Himalayan Pink Sea Salt	106.68000
117667	Sea Salt	149.86000
119185	Premium Iodized Sea Salt	101.60000
119186	Coarse Crystals	101.60000
119860	Flower Of Salt, Smoked	108.71200
119863	Natural Sea Salt, Spicy	100.33000
120800	Coarse Kosher Salt	101.60000
120946	Evolution Salt Co, Gourmet Pink Himalaya Salt	254.00000
120952	Granlund's, Gummi Peaches	123.90120
120960	Granlund's, Peco Brittle	136.07034
120994	Seasoning Salt	677.33418
120995	Anisa's Secrets, Steak Rub	677.33418
120996	Anisa's Secrets, Rosemary Lemon Pepper	677.33418
120999	Anisa's Secrets, Moroccan Spice	677.33418
121005	Anisa's Secrets, Poultry Seasoning	677.33418
121006	Anisa's Secrets, Lemon Dill	677.33418
121706	Primordial Himalayan Sea Salt	112.18418
121707	Primordial Himalayan Sea Salt	107.69600
121892	Inc., California Thompson Raisins	317.50000
122061	Smokey Tomato Speard	604.76130
125882	Mediterranean Sea Salt	101.60000
131415	Gold, Mediterranean Sea Salt	105.83418
132084	Roasted Garlic & Himalayan Pink Salt Seasoning	107.10418
132086	Habanero & Himalayan Pink Salt Seasoning	101.23678
132087	Himalayan Pink Salt Seasoning	112.18418
132090	Sriracha & Himalayan Pink Salt Seasoning	104.86644
135064	Chiles Habaneros Rojos	3556.00000
137618	Piri-Piri Hot Sauce	187.96000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	421.83812
141332	Seasoned Salt	111.76000
141333	Seasoned Salt	111.76000
141334	Pegs, Pink Salt	121.92000
142408	Fine Iodized Sea Salt	103.71582
145844	Prickly Syrup	64312.80000
146800	Pur Beverages, Energy Beverage, Acai Berry	100.85324
147845	Truffle & Salt	109.22000
148914	Sweet Goodies White Chocolate Balls	181.42966
149081	Jim Kelly, Sea Salt Adjustable Grinder	100.01250
150717	Quinoa Pasta Fusilli	133.68528
150718	Quinoa Curvo Rigate Pasta	133.68528
152783	Himalayan Chef, Himalayan Pink Salt Grinder	106.68000
152784	Himalayan Pink Salt	106.68000
155803	Tamarind Juice With Nata De Coco, Tamarind	147.32000
160684	Sausage Patties, Veggie Sausage Patties	1669.14322
165311	Dressing & Marinade	1016.00000
165331	Bacon Salt	118.11000
165332	Molten Hot Salt	118.36400
165333	Garlic And Rosemary Salt	118.11000
165334	Mountain Mushroom Salt	118.11000
166102	Pink Salt	106.68000
170832	Gold, Adjustable Grinder, Pink Himalayan Salt	106.68000
171990	Seasoning	147.32000
172421	Bruschetta Blend	381.00000
172425	Sweet Bread & Butter Pickling Seasoning	296.33418
176694	flake	276.86000
176696	herdhey's	124.46000
177374	Sea Salt	105.83418
177375	Applewood Smoked Sea Salt	101.60000
177376	Himalayan Pink Salt	106.68000
177379	Mediterranean Coarse Sea Salt	105.83418
184025	Himalayan Pink Salt	106.68000
232239	Keeny Bio	211.00000
233372	Sel de Guérande Label Rouge	125.00000
274767	Cheese salad	297.60000
279614	Organic Dill Salt	100.51034
284225	Bamboo Shoots	448.23634
298059	Kids Cristal	930.00000
305114	Tartufi Di Fassia, Grey Salt With White Truffles	106.68000
Changed to	product_name	salt_100g
501	Mediterranean Sea Salt	0.15748
1435	Dill Pickle Cashews	0.65024
3472	Coarse Kosher Salt	0.00000
7037	Fine Sea Salt	0.05588
7983	Butter Beans	0.00000

11129		Autumn Cookie Selection	1.21920
18588		Mediterranean Sea Salt Grinder	0.10000
18848		Himalayan Pink Salt	0.84582
18851	Olde Thompson,	Mediterranean Sea Salt	0.84582
18858		Himalayan Pink Salt	0.00000
18869		Mediterranean Sea Salt	0.21336
19014	Adolph's, Prime Cut, Tenderizer Seasoned With ...		1.27000
20821		Special Blend, Peas & Carrots	1.73482
21619		Fine Mediterranean Sea Salt	2.54000
21626		Sea Salt	0.00000
21627		Mediterranean Sea Salt, Fine	0.00000
21632		Coarse Kosher Salt	0.00000
21633		Kosher Salt	0.00000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...		1.78308
23590		Original Barbecue Sauce	0.00000
25671		Fine Sea Salt	15.35430
28787	Fratelli Comino, Tochetti With Sea Salt		0.06858
32371		Light & Fit Greek Crunch - S'mores	1.05918
33630		Hawaiian Black Sea Salt	0.10000
39679		Garlic Salt	0.41910
40408	Fleur De Sel, Hand-Harvested French Sea Salt		1.86182
40409		Grey Sea Salt	11.43000
40410		Fine Crystals Sea Salt	1.94818
40411		Coarse Crystals Sea Salt	1.94818
40412		Sea Salt	1.27000
43049	Premium Spice Blends, Seasoning Salt		0.00000
43772		Coarse Kosher Salt	0.06350
43773		Fine Sea Salt	0.01016
49970		Mediterranean Sea Salt	0.34544
49972		Himalayan Pink Salt	0.22352
51348	Crispy Fried Onions, White Cheddar		1.01600
55514	Pure Carrot With Organic Lemon Juice		2.35966
58007		String Cheese Snacks	1.60782
61552		Sea Salt Grinder	0.49276
63824		Steak Seasoning Salt	3.17500
65202		Salt	0.22352
68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...		0.00000
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...		0.00000
68507		Kosher Salt	0.00000
74261		Fine Sea Salt	4.06400
74685		Coarse Sea Salt	0.71882
74686		All Natural Sea Salt	0.55372
74687		Kosher Sea Salt	0.00000
75834	Crawfish, Shrimp & Crab Boil		0.01016
75838	Crawfish, Shrimp And Crab Boil		0.91948
79161		Pickling & Canning Salt	0.19050
81378	Adobo Seasoning With Pimento		0.20574
81379	All Purpose Adobo Seasoning With Sazon		2.04724
90930		Sea Salt Fine	1.81356
90982		Coarse Kosher Salt	1.02108
90985	Nature's Supreme, Pure Fine Sea Salt		0.79248
93914		La Flor, Grinder Sea Salt	0.31750
93915	Alderwood Smoked Sea Salt Grinder		0.39624
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...		2.76352
105006	Bastoncini Whole Wheat		2.08534
106330	Small Potato Dumplings, Potato Gnocchetti		0.97282
106360		Fine Sea Salt	1.90500
106361		Coarse Sea Salt	0.00000
106544		Sea Salt	10.66800
106545	Sea Salt, Fine Grind		81.28000
106880	Popcorn Salt		0.00000
108367	Natural Yogurt Pretzels		0.86106
109090	Mediterranean Sea Salt Grinder		2.41046
109154	Himalayan Pink Salt		0.00000
109195	Lisy, Seasoned Meat Tenderizer		0.64770
113216	Himalayan Pink Salt Grinder		0.10668
115253	All Natural Sea Salt		1.68402
115865	Mediterranean Sea Salt		0.68072
115866	Himalayan Pink Salt		1.10490
115872	Himalayan Pink Sea Salt		0.63500
117667	Sea Salt		1.17856
119185	Premium Iodized Sea Salt		0.99568
119186	Coarse Crystals		0.60452
119860	Flower Of Salt, Smoked		0.03810
119863	Natural Sea Salt, Spicy		0.00000
120800	Coarse Kosher Salt		0.63500
120946	Evolution Salt Co, Gourmet Pink Himalaya Salt		0.63500
120952	Granlund's, Gummi Peaches		0.00000
120960	Granlund's, Peco Brittle		0.84582
120994		Seasoning Salt	0.84582
120995	Anisa's Secrets, Steak Rub		0.40132
120996	Anisa's Secrets, Rosemary Lemon Pepper		1.82626
120999	Anisa's Secrets, Moroccan Spice		1.86182
121005	Anisa's Secrets, Poultry Seasoning		0.31750
121006	Anisa's Secrets, Lemon Dill		0.28956
121706	Primordial Himalayan Sea Salt		0.04572
121707	Primordial Himalayan Sea Salt		0.04572
121892	Inc., California Thompson Raisins		1.45034
122061	Smokey Tomato Speard		0.00000
125882	Mediterranean Sea Salt		1.31318
131415	Gold, Mediterranean Sea Salt		0.14478
132084	Roasted Garlic & Himalayan Pink Salt Seasoning		0.00000
132086	Habanero & Himalayan Pink Salt Seasoning		1.75260
132087	Himalayan Pink Salt Seasoning		0.46600
132090	Sriracha & Himalayan Pink Salt Seasoning		0.08890
135064	Chiles Habaneros Rojos		0.12700
137618	Piri-Piri Hot Sauce		0.00000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend		7.82320
141332	Seasoned Salt		0.02032
141333	Seasoned Salt		0.01016
141334	Pegs, Pink Salt		0.02032
142408	Fine Iodized Sea Salt		0.04826
145844	Prickly Syrup		1.99644
146800	Pur Beverages, Energy Beverage, Acai Berry		2.24790
147845	Truffle & Salt		3.62966
148914	Sweet Goodies White Chocolate Balls		0.34290
149081	Jim Kelly, Sea Salt Adjustable Grinder		0.00000

150717	Quinoa Pasta Fusilli	0.00000
150718	Quinoa Curvo Rigate Pasta	0.00000
152783	Himalayan Chef, Himalayan Pink Salt Grinder	0.45466
152784	Himalayan Pink Salt	0.15240
155803	Tamarind Juice With Nata De Coco, Tamarind	0.00000
160684	Sausage Patties, Veggie Sausage Patties	1.70000
165311	Dressing & Marinade	0.00000
165331	Bacon Salt	0.00000
165332	Molten Hot Salt	0.00000
165333	Garlic And Rosemary Salt	0.00000
165334	Mountain Mushroom Salt	0.00000
166102	Pink Salt	0.00000
170832	Gold, Adjustable Grinder, Pink Himalayan Salt	0.77000
171990	Seasoning	0.35000
172421	Bruschetta Blend	0.00000
172425	Sweet Bread & Butter Pickling Seasoning	0.00000
176694	flake	0.00000
176696	herdhey's	0.00000
177374	Sea Salt	0.00000
177375	Applewood Smoked Sea Salt	0.00000
177376	Himalayan Pink Salt	0.00000
177379	Mediterranean Coarse Sea Salt	0.00000
184025	Himalayan Pink Salt	0.97000
232239	Keeny Bio	0.00000
233372	Sel de Guérande Label Rouge	0.61000
274767	Cheese salad	0.00000
279614	Organic Dill Salt	0.04000
284225	Bamboo Shoots	0.00000
298059	Kids Cristal	Nan
305114	Tartufi Di Fassia, Grey Salt With White Truffles	Nan

colonne sodium\_100g a 33 valeurs supérieures a 100g

	product_name	sodium_100g
1435	Dill Pickle Cashews	342.857000
7983	Butter Beans	307.692000
11129	Autumn Cookie Selection	129.032000
20821	Special Blend, Peas & Carrots	281.250000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	234.200000
23590	Original Barbecue Sauce	800.000000
51348	Crispy Fried Onions, White Cheddar	857.143000
58007	String Cheese Snacks	708.333000
68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	432.432000
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	338.028000
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	519.048000
105006	Bastoncini Whole Wheat	965.517000
106330	Small Potato Dumplings, Potato Gnocchetti	448.485000
108367	Natural Yogurt Pretzels	262.500000
120994	Seasoning Salt	266.667000
120995	Anisa's Secrets, Steak Rub	266.667000
120996	Anisa's Secrets, Rosemary Lemon Pepper	266.667000
120999	Anisa's Secrets, Moroccan Spice	266.667000
121005	Anisa's Secrets, Poultry Seasoning	266.667000
121006	Anisa's Secrets, Lemon Dill	266.667000
121892	Inc., California Thompson Raisins	125.000000
122061	Smokey Tomato Speard	238.095000
135064	Chiles Habaneros Rojos	1400.000000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	166.078000
145844	Prickly Syrup	25320.000000
160684	Sausage Patties, Veggie Sausage Patties	657.143000
165311	Dressing & Marinade	400.000000
172421	Bruschetta Blend	150.000000
172425	Sweet Bread & Butter Pickling Seasoning	116.667000
176694	flake	109.000000
274767	Cheese salad	117.165354
284225	Bamboo Shoots	176.471000
298059	Kids Cristal	366.141732

Changed to

	product_name	sodium_100g
1435	Dill Pickle Cashews	0.256000
7983	Butter Beans	0.000000
11129	Autumn Cookie Selection	0.480000
20821	Special Blend, Peas & Carrots	0.683000
21892	Marie Sharp's, Orange Pulp Habanero Pepper Sau...	0.702000
23590	Original Barbecue Sauce	0.000000
51348	Crispy Fried Onions, White Cheddar	0.400000
58007	String Cheese Snacks	0.633000
68379	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	0.000000
68388	Rhodes, Bake N Serv, Cinnamon Rolls With Cream...	0.000000
94631	Swiss Recipe Smoked Almonds & Sea Salt In Dark...	1.088000
105006	Bastoncini Whole Wheat	0.821000
106330	Small Potato Dumplings, Potato Gnocchetti	0.383000
108367	Natural Yogurt Pretzels	0.339000
120994	Seasoning Salt	0.333000
120995	Anisa's Secrets, Steak Rub	0.158000
120996	Anisa's Secrets, Rosemary Lemon Pepper	0.719000
120999	Anisa's Secrets, Moroccan Spice	0.733000
121005	Anisa's Secrets, Poultry Seasoning	0.125000
121006	Anisa's Secrets, Lemon Dill	0.114000
121892	Inc., California Thompson Raisins	0.571000
122061	Smokey Tomato Speard	0.000000
135064	Chiles Habaneros Rojos	0.050000
140616	Gold Raw & Dry Roasted Combo Trail Mix Blend	3.080000
145844	Prickly Syrup	0.786000
160684	Sausage Patties, Veggie Sausage Patties	0.669291
165311	Dressing & Marinade	0.000000
172421	Bruschetta Blend	0.000000
172425	Sweet Bread & Butter Pickling Seasoning	0.000000
176694	flake	0.000000
274767	Cheese salad	0.000000
284225	Bamboo Shoots	0.000000
298059	Kids Cristal	Nan

colonne vitamin-a\_100g a 1 valeurs inférieures a 0g

	product_name	vitamin-a_100g
80440	Flavor Aid, Soft Drink Mix, Lemon	-0.00034
Changed to	product_name	vitamin-a_100g
80440	Flavor Aid, Soft Drink Mix, Lemon	0.0

colonne vitamin-c\_100g a 1 valeurs supérieures a 100g  
product\_name vitamin-c\_100g  
139242 Organic Fruit Snacks, Berry Medley 370.3704  
Changed to product\_name vitamin-c\_100g  
139242 Organic Fruit Snacks, Berry Medley 0.0  
colonne vitamin-c\_100g a 1 valeurs inférieures a 0g  
product\_name vitamin-c\_100g  
41538 Cheez Waffies -0.0021  
Changed to product\_name vitamin-c\_100g  
41538 Cheez Waffies 0.0

colonne vitamin-b1\_100g a 1 valeurs supérieures a 100g  
product\_name vitamin-b1\_100g  
67437 Céréales rice krispies 161.0  
Changed to product\_name vitamin-b1\_100g  
67437 Céréales rice krispies 0.0

colonne folates\_100g a 2 valeurs supérieures a 100g  
product\_name folates\_100g  
3594 Cereal Toasted Oats 166.666667  
4494 Toasted Oats 178.571429  
Changed to product\_name folates\_100g  
3594 Cereal Toasted Oats 0.0  
4494 Toasted Oats 0.0

colonne biotin\_100g a 3 valeurs inférieures a 0g  
product\_name biotin\_100g  
189152 France -1.0  
189160 France -2.0  
189272 France -1.0  
Changed to product\_name biotin\_100g  
189152 France 0.0  
189160 France 0.0  
189272 France 0.0

colonne pantothenic-acid\_100g a 3 valeurs inférieures a 0g  
product\_name pantothenic-acid\_100g  
189152 France -1.0  
189160 France -2.0  
189272 France -1.0  
Changed to product\_name pantothenic-acid\_100g  
189152 France 0.0  
189160 France 0.0  
189272 France 0.0

colonne potassium\_100g a 2 valeurs supérieures a 100g  
product\_name potassium\_100g  
149716 Active Natural Electrolyte Enhanced Drink Tabs 1870.37  
172984 Deglet Noor Pitted Dates 600.00  
Changed to product\_name potassium\_100g  
149716 Active Natural Electrolyte Enhanced Drink Tabs 0.0  
172984 Deglet Noor Pitted Dates 0.0

colonne calcium\_100g a 9 valeurs supérieures a 100g  
product\_name calcium\_100g  
10494 Sweeteners, Organic Molasses 522.727  
112298 Dulce De Leche, Milk Caramel Spread 285.714  
113049 5 Year, Gouda Cheese 125.000  
117077 Promenade Red Apple Market, B.B.Q Seasoning 136.000  
123679 Sweet Rice 173.333  
124143 Tastyshake, Thermodynamic Protein Complex Prot... 613.636  
124222 Turkish Sun Dried Figs 161.560  
165426 Spreadable Gouda Cheese With Red Pepper 432.143  
285611 Jaleas Ujarra's, Jelly, Strawberry 120.000  
Changed to product\_name calcium\_100g  
10494 Sweeteners, Organic Molasses 0.000  
112298 Dulce De Leche, Milk Caramel Spread 0.061  
113049 5 Year, Gouda Cheese 0.040  
117077 Promenade Red Apple Market, B.B.Q Seasoning 0.000  
123679 Sweet Rice 0.000  
124143 Tastyshake, Thermodynamic Protein Complex Prot... 0.067  
124222 Turkish Sun Dried Figs 0.014  
165426 Spreadable Gouda Cheese With Red Pepper 0.000  
285611 Jaleas Ujarra's, Jelly, Strawberry NaN

colonne phosphorus\_100g a 5 valeurs supérieures a 100g  
product\_name phosphorus\_100g  
117363 Soy Protein & Multigrain Cereal 400.000  
117364 Simply Fiber, All Natural Crunchy O's Cereal 559.459  
123836 Kamut, Premium Khorasan Wheat 410.870  
123837 Premium Buckwheat 315.556  
123838 Premium Teff 320.000  
Changed to product\_name phosphorus\_100g  
117363 Soy Protein & Multigrain Cereal 0.0  
117364 Simply Fiber, All Natural Crunchy O's Cereal 0.0  
123836 Kamut, Premium Khorasan Wheat 0.0  
123837 Premium Buckwheat 0.0  
123838 Premium Teff 0.0

colonne iron\_100g a 1 valeurs inférieures a 0g  
product\_name iron\_100g  
316892 Lamthong, Fruit Cocktail In Syrup -0.00026  
Changed to product\_name iron\_100g  
316892 Lamthong, Fruit Cocktail In Syrup NaN

```

colonne magnesium_100g a 6 valeurs supérieures a 100g
      product_name magnesium_100g
19681   Essentials, Organic Ground Chia Seed Dietary S...    333.333
19683   Organic Ground Premium Flaxseed Dietary Supple...    364.286
149841           Goji Basil Superfood + Cashews    196.429
149845           Hemp Seeds    573.333
166358           Cacao Nibs    300.000
166363       Raw Cacao Powder    657.143
Changed to      product_name magnesium_100g
19681   Essentials, Organic Ground Chia Seed Dietary S...    0.0
19683   Organic Ground Premium Flaxseed Dietary Supple...    0.0
149841           Goji Basil Superfood + Cashews    0.0
149845           Hemp Seeds    0.0
166358           Cacao Nibs    0.0
166363       Raw Cacao Powder    0.0

```

```

colonne selenium_100g a 1 valeur inférieure a 0g
      product_name selenium_100g
28324   Venus, Cuttlefish Balls    -0.000002
Changed to      product_name selenium_100g
28324   Venus, Cuttlefish Balls    0.0

```

18 colonnes avec valeurs >100g, 10 avec valeurs <0g

In [ ]:

In [231]: # On fait les graphiques sur toutes les coloones numériques pour avoir des idées de nettoyage supplémentaires

```

list_num_col = myDF4.select_dtypes(include = ['float64']).columns

#for x in list_num_col:
#    sns.displot(myDF4[x], bins=50)
#    plt.title('Valeurs pour la colonne : ' + x)
#    plt.xlabel(x)
#    plt.show()

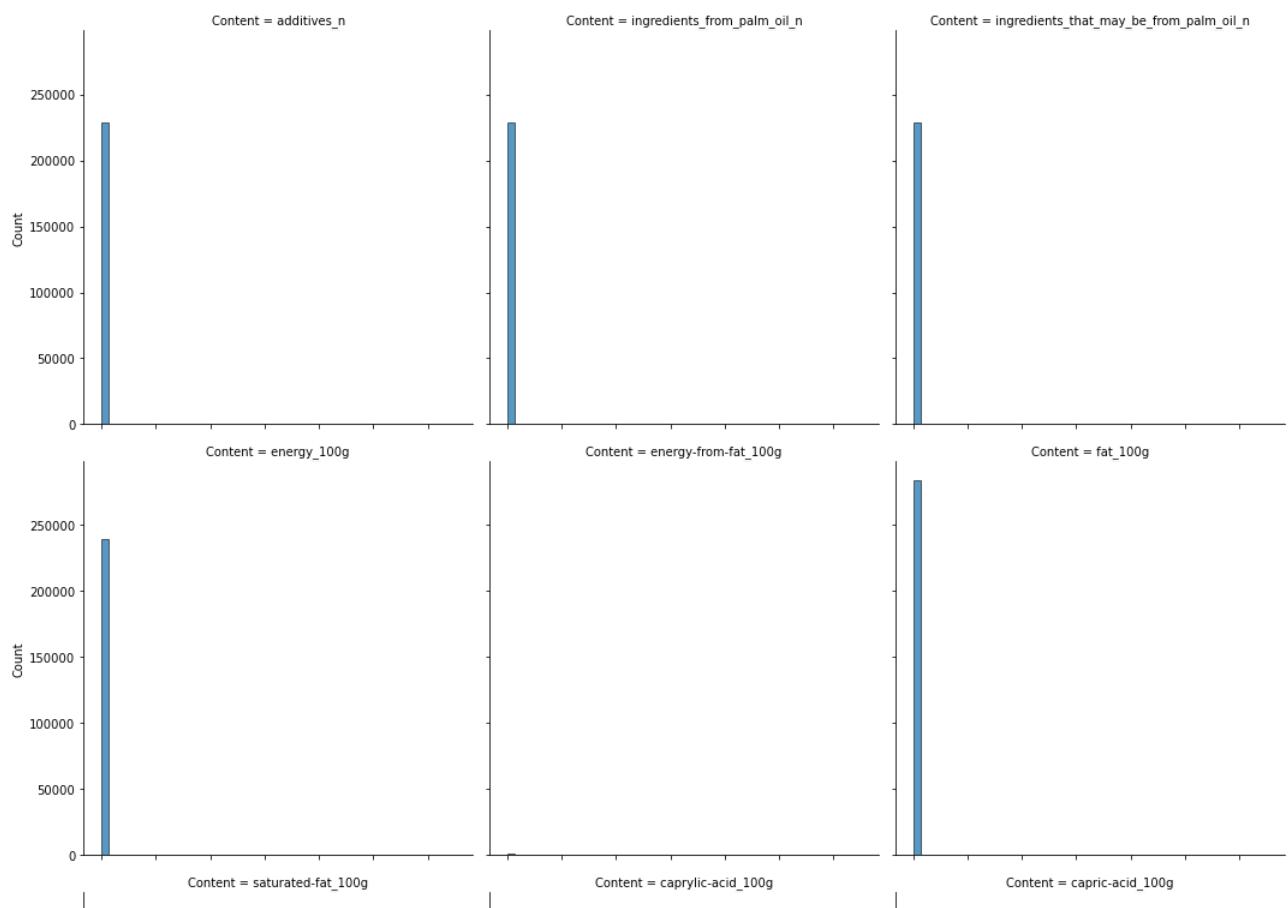
sns.displot(kind='hist', data=myDF4[list_num_col].melt( var_name='Content'), bins = 50,
            col='Content', col_wrap=3, x='value')

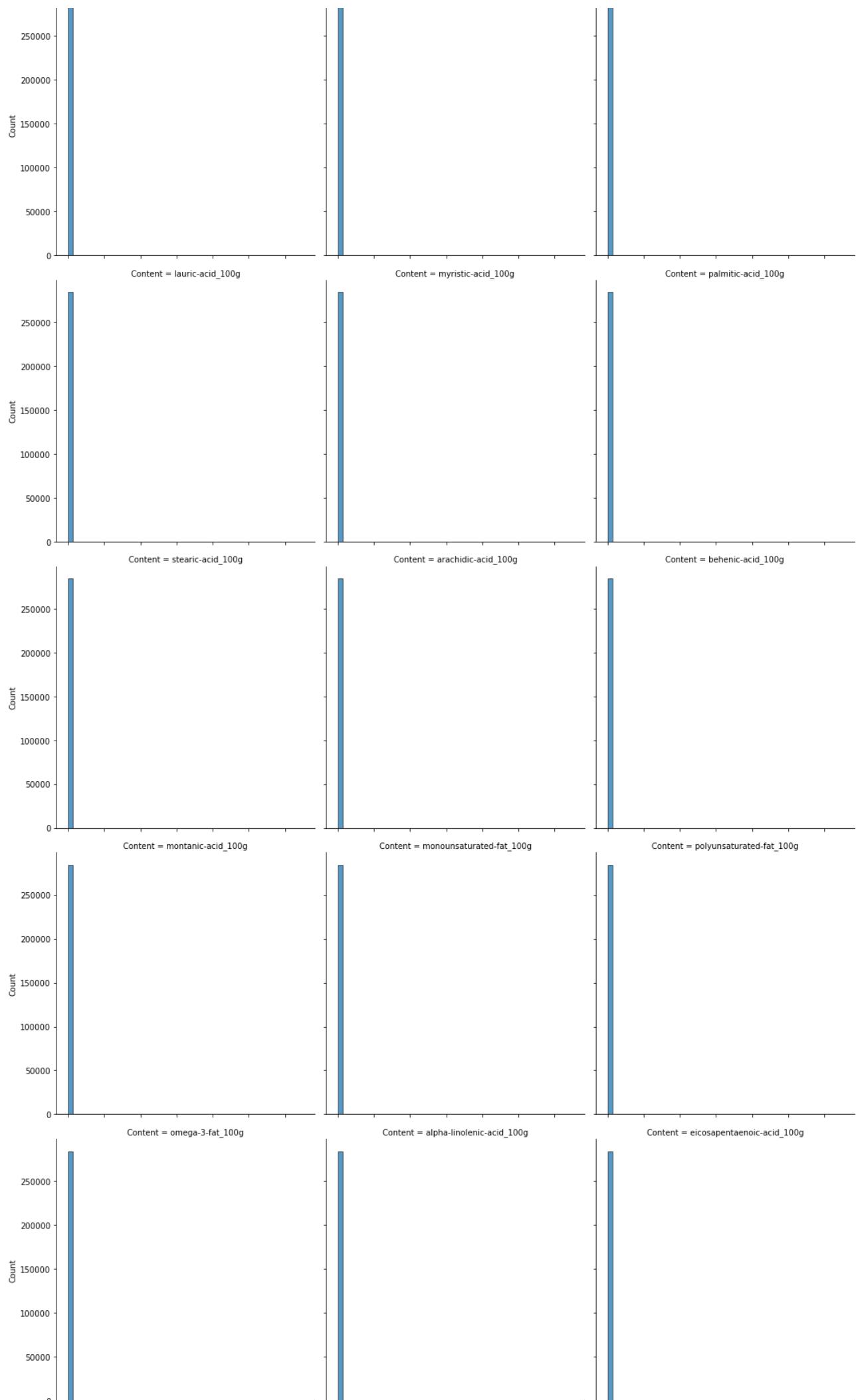
#MENTOR comment voir tout le graphique ? A priori Jupyter ne sait pas gerer hauteur

# Faire des boxplots plutot qu'histo

```

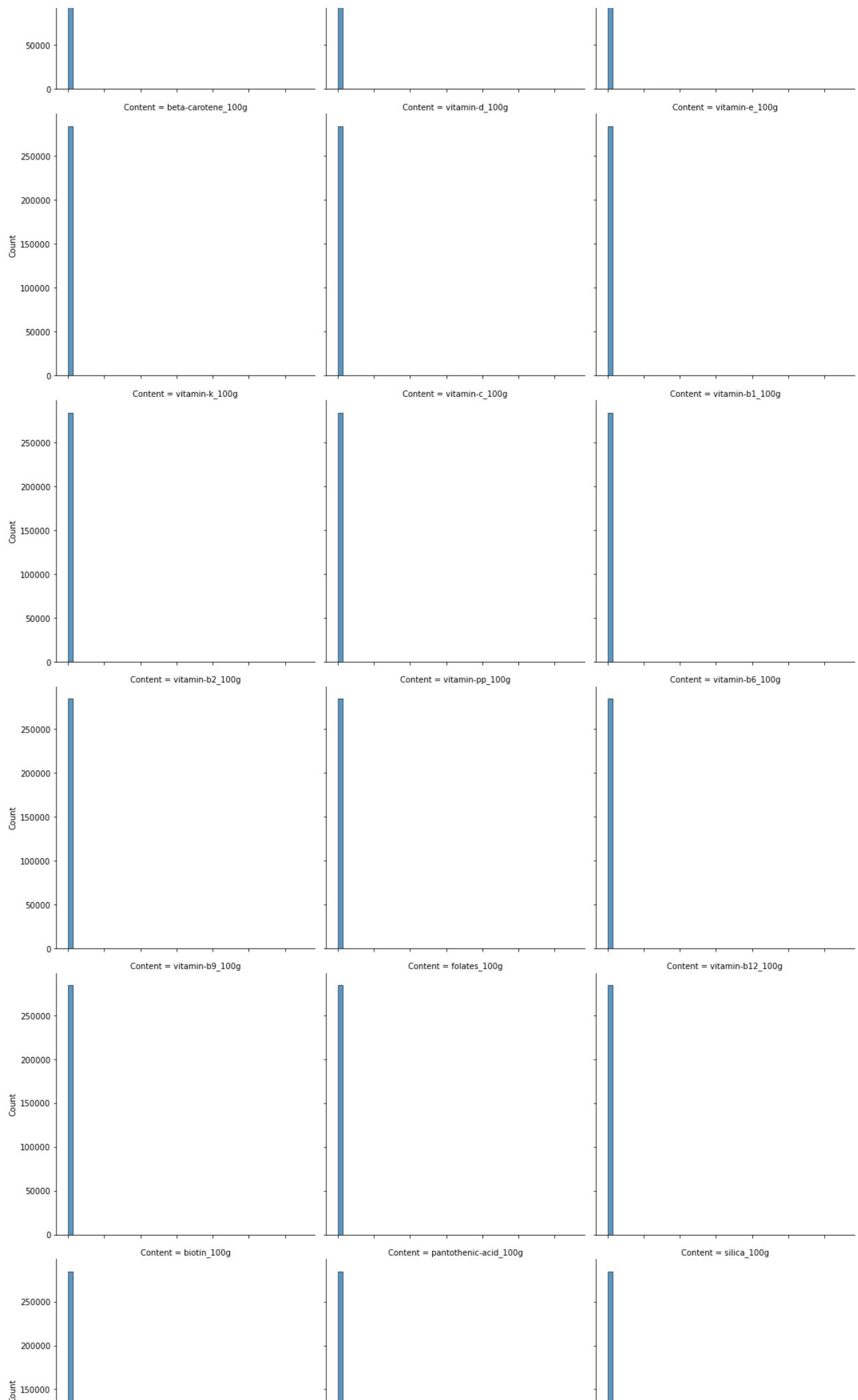
Out[231]: <seaborn.axisgrid.FacetGrid at 0x28d8e58b760>

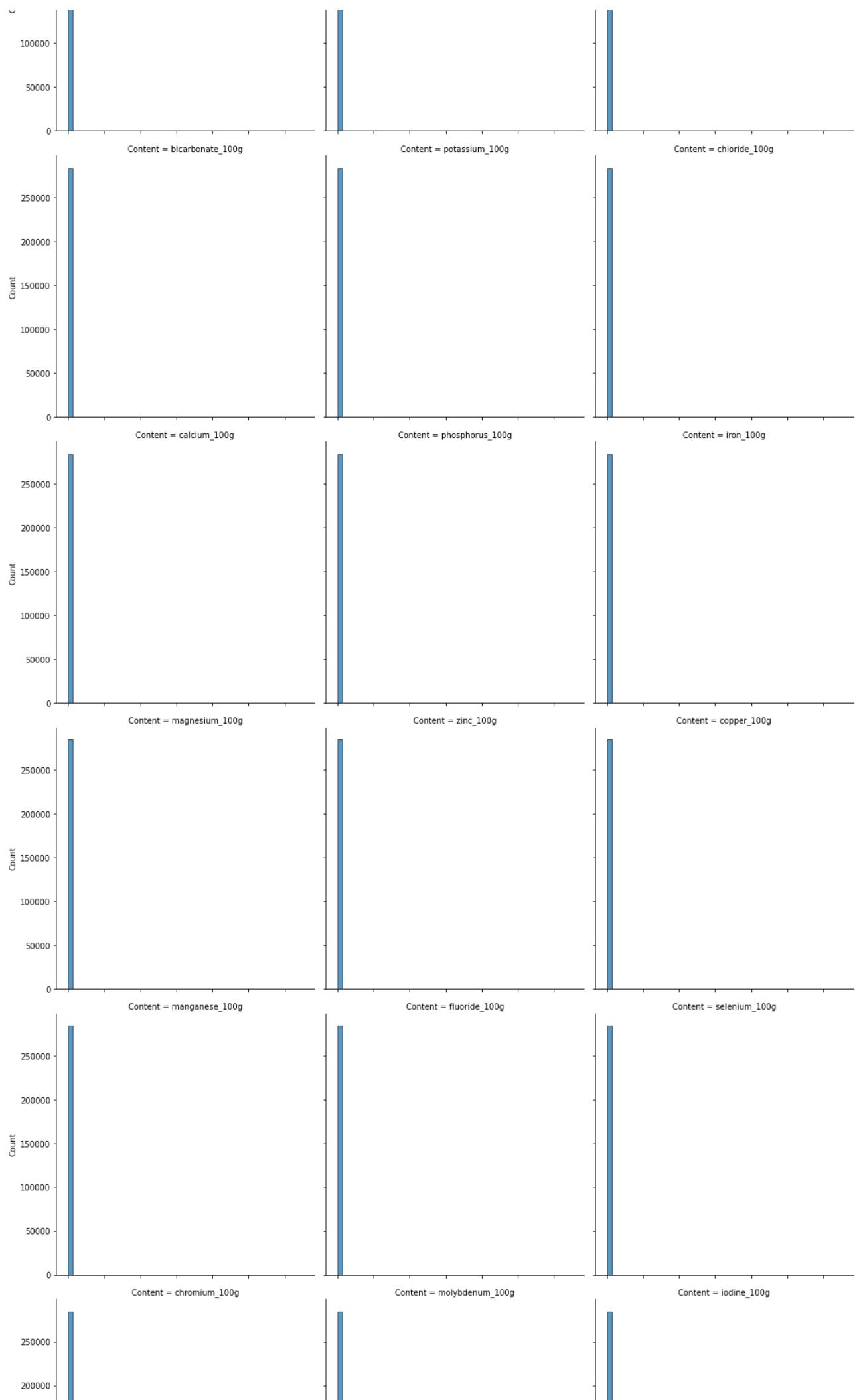


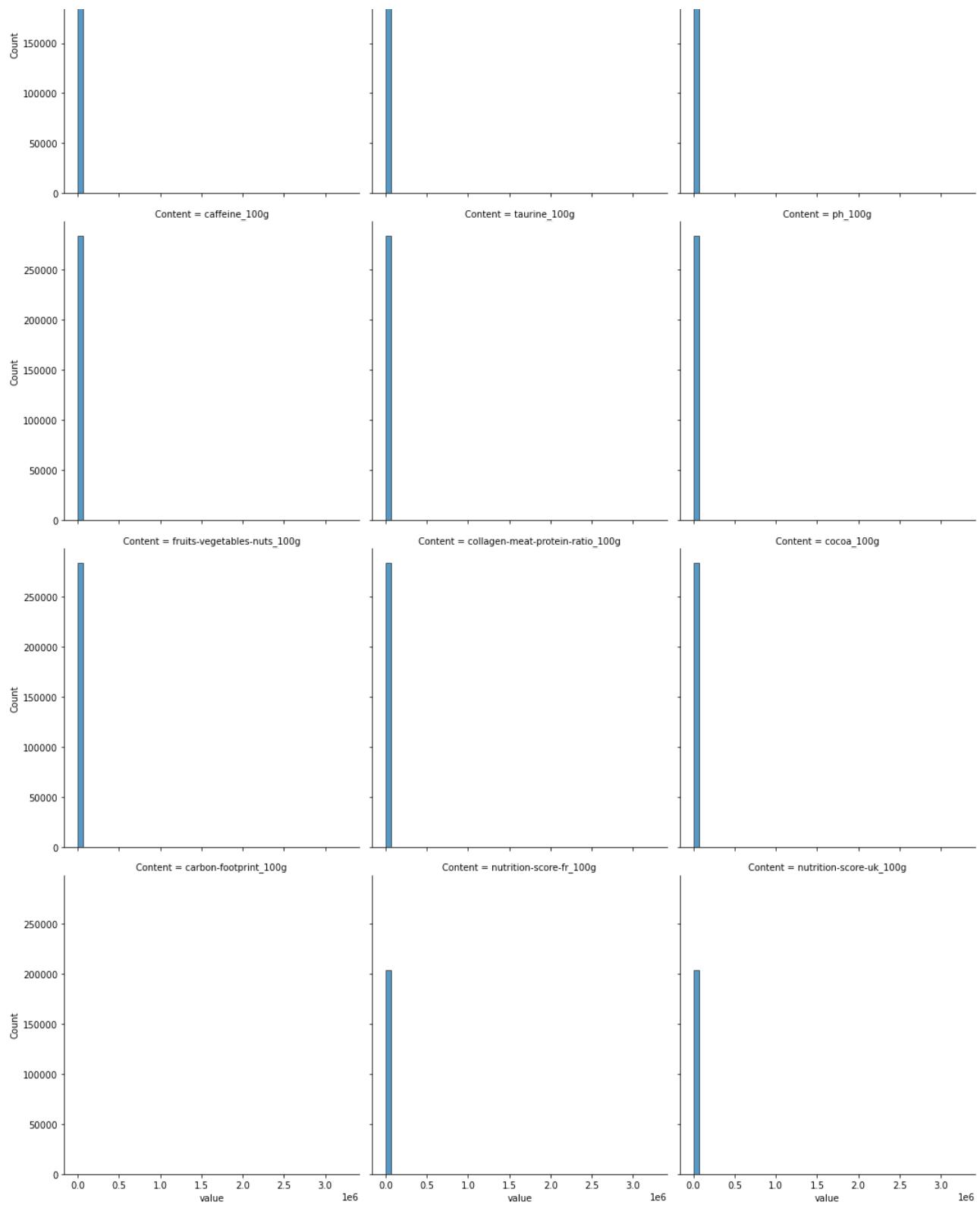












### constats:

ingredients\_from\_palm\_oil\_n et ingredients\_that\_may\_be\_from\_palm\_oil\_n est sans doute un booleen (ne prend que les valeurs 0 ou 1)

In [ ]:

In [ ]:

In [ ]:

In [232]:

```
# myDF7.to_csv("./Datasets/FoodDataNettoyee.csv")
# cf nettoyage supplémentaire ci-dessous
```

### Remarques et conclusion sur le nettoyage

# Quelles idées d'analyses ?

1. Qualifier les produits selon le pays d'origine: Quel pays a les produits les plus sains, faut-il mieux acheter français ou européens ?

2.

Pour rappel, les 3 exemples donnés :

- David = "nutrisport = scanner un produit, suggérer des alternatives"
- Aurélia = choisir de meilleures alternatives (a la place des chips)
- Sylvain = "donner une note à un produit selon l'impact sur l'environnement"

In [233...]

```
# OK valeur manquantes gérée (remplacée par 0), valeur aberrantes (gerer avec KNN)

# OK gestion des doublons ? à faire sur les noms (cf fonction duplicated)

# enlever des variables inutiles ? (selon les besoins du projet)

# le nettoyage pourrait être mieux automatisé
# exemple https://github.com/ydataai/pandas-profiling
# c'est un package qui lance de analyses de bases
```

In [234...]

```
# RESTE A FAIRE = regarder la somme des nutriments supérieurs à 100g.
# - erronés = supprimer
# - possibilité d'hypothèse de "normaliser" = une portion > 100g peut être redressée à 100.

list_nutriments_100g = ['fat_100g', 'carbohydrates_100g', 'proteins_100g']

myDF8 = myDF7
myDF8["total3nutriments_100g"] = myDF7[list_nutriments_100g].sum(axis=1)
myDF8.head()
```

Out[234...]

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	quantity	packagin
0	3087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir		NaN	1kg
1	4530	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Banana Chips Sweetened (Whole)		NaN	NaN
2	4559	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489069957	2017-03-09T14:32:37Z	1489069957	2017-03-09T14:32:37Z	Peanuts		NaN	NaN
3	16087	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055731	2017-03-09T10:35:31Z	1489055731	2017-03-09T10:35:31Z	Organic Salted Nut Mix		NaN	NaN
4	16094	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055653	2017-03-09T10:34:13Z	1489055653	2017-03-09T10:34:13Z	Organic Polenta		NaN	NaN

In [235...]

```
sum(myDF8["total3nutriments_100g"]>100)
```

Out[235...]

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	qua
13	16933	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055651	2017-03-09T10:34:11Z	1489055651	2017-03-09T10:34:11Z	Organic Golden Flax Seeds		NaN
74	33572	http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489138488	2017-03-10T09:34:48Z	1489138488	2017-03-10T09:34:48Z	Organic Whole Brown Rice Flour		NaN

		code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_name	qua
94	34784		http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489138485	2017-03-10T09:34:45Z	1489138485	2017-03-10T09:34:45Z	Unsalted Dry Roasted Almonds		NaN
129	38706		http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489055652	2017-03-09T10:34:12Z	1489055652	2017-03-09T10:34:12Z	Organic Rolled Rye Flakes		NaN
134	39208		http://world-fr.openfoodfacts.org/produit/0000...	usda-ndb-import	1489138488	2017-03-10T09:34:48Z	1489138488	2017-03-10T09:34:48Z	Organic Arboria Rice		NaN
...	...	...	...	...	...	...	...	...	...	...	...
281511	5690527914000		http://world-fr.openfoodfacts.org/produit/5690...	openfood-ch-import	1486510693	2017-02-07T23:38:13Z	1487257466	2017-02-16T15:04:26Z	Icelandic Skyr The Original Lemon Cheesecake		NaN 1
281674	57067094		http://world-fr.openfoodfacts.org/produit/5706...	mathias	1430031527	2015-04-26T06:58:47Z	1430031540	2015-04-26T06:59:00Z	Kartoffel spegepølse		NaN
281878	5760466904919		http://world-fr.openfoodfacts.org/produit/5760...	openfoodfacts-contributors	1463838216	2016-05-21T13:43:36Z	1490762264	2017-03-29T04:37:44Z	Kærgården		NaN 2
283499	6192408100353		http://world-fr.openfoodfacts.org/produit/6192...	openfoodfacts-contributors	1388097715	2013-12-26T22:41:55Z	1388097755	2013-12-26T22:42:35Z	doliprex		NaN
284119	68826707068		http://world-fr.openfoodfacts.org/produit/6882...	usda-ndb-import	1489134802	2017-03-10T08:33:22Z	1489134803	2017-03-10T08:33:23Z	Park Skim Ricotta Cheese		NaN

1572 rows × 147 columns

```

In [ ]:
<ipython-input-237-1333a2f3a2d1>
In [237...]: myDF9 = myDF8.drop(myDF8[myDF8["total3nutrients_100g"] > 100].index)
myDF9.shape
Out[237...]: (282658, 147)

In [238...]: myDF9.to_csv("./Datasets/FoodDataNettoyeed.csv")
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:

```