

Open Food Facts

Analyse exploratoire des données

Cyril Jeanneau, Dylan Chevallier, Jonathan Caillaux



L'information nutritionnelle : un enjeu de santé publique

Augmentation des maladies liées à l'alimentation

- 39% de la population mondiale en surpoids (OMS, 2022)
- Diabète en hausse constante
- Maladies cardiovasculaires : première cause de mortalité

Solution : Adopter une alimentation saine.

Compréhension Métier

Qu'est-ce qu'un élément sain ?

Définition : « Un aliment est considéré comme sain d'un point de vue nutritionnel lorsqu'il fournit des nutriments essentiels (vitamines, minéraux, fibres, protéines) tout en ayant une teneur limitée en éléments à modérer (sucres ajoutés, sel, graisses saturées) »

Compréhension Métier

Qu'est-ce qu'un élément sain ?

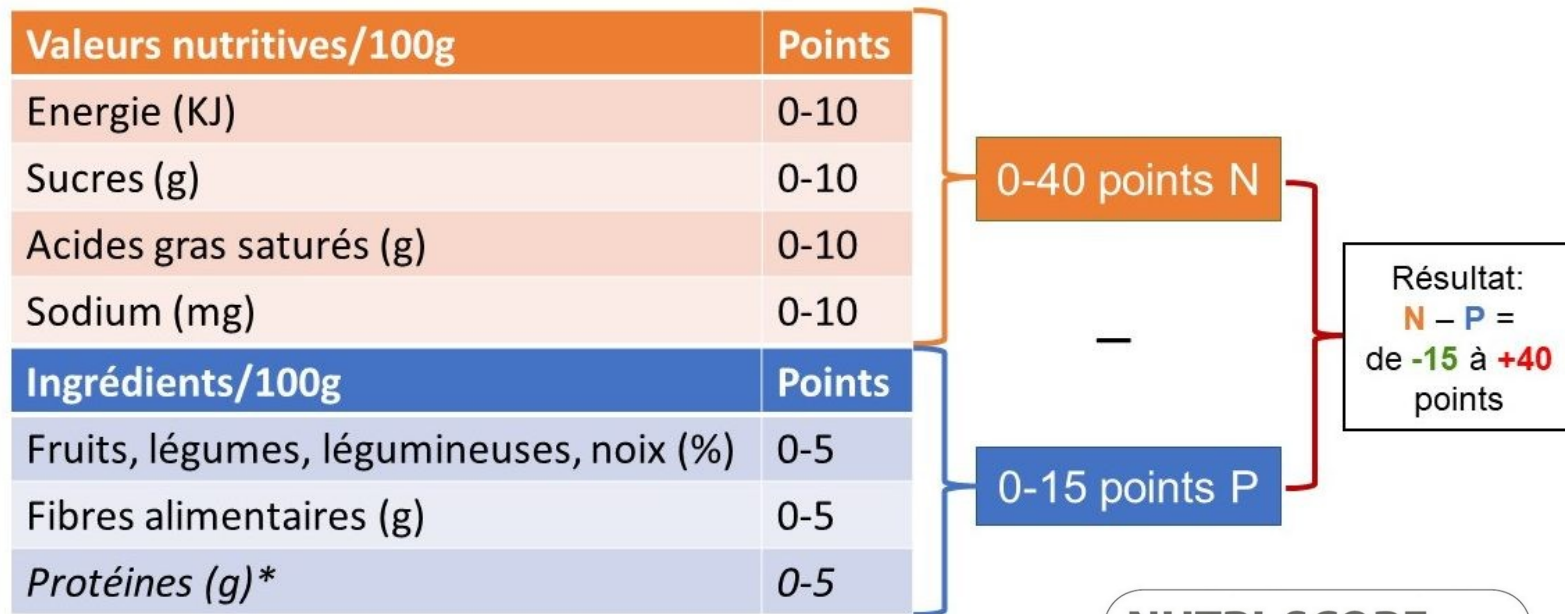
Définition : « Un aliment est considéré comme sain d'un point de vue nutritionnel lorsqu'il fournit des nutriments essentiels (vitamines, minéraux, fibres, protéines) tout en ayant une teneur limitée en éléments à modérer (sucres ajoutés, sel, graisses saturées) »

Comment les repérer ?

Infos nutritionnelles		
Champignons à la Grecque		
Quantité par portion		
Calories	256 1	Lipides 9
% des apports journaliers*		
Lipides	1g 2	2%
Lipides saturés	1g	6%
Sodium	115mg 3	5%
Potassium	1198mg	34%
Glucides	39g 4	13%
Fibres	5g	21%
Sucre	28g	31%
Protéines	10g 5	20%
Vitamine A	262IU 6	5%
Vitamine C	45mg	55%
Calcium	67mg 7	7%
Fer	3mg	17%
* Pourcentage journalier basé sur une consommation de 2000 calories.		

Compréhension Métier

Nutriscore : une seule note pour résumer les informations nutritionnelles



*sans les protéines si la somme des points A >11



Open Food Facts

- Projets libre créer en 2012 par Stéphane Gigandet et Pierre Slamich
- Base de donnée mondiale concernant les informations nutritionnelles
- Soutenu principalement par *Santé publique France* et la Commission Européenne
- Indépendant de l'industrie agro-alimentaire
- Financé par le mécennat et des donations publiques



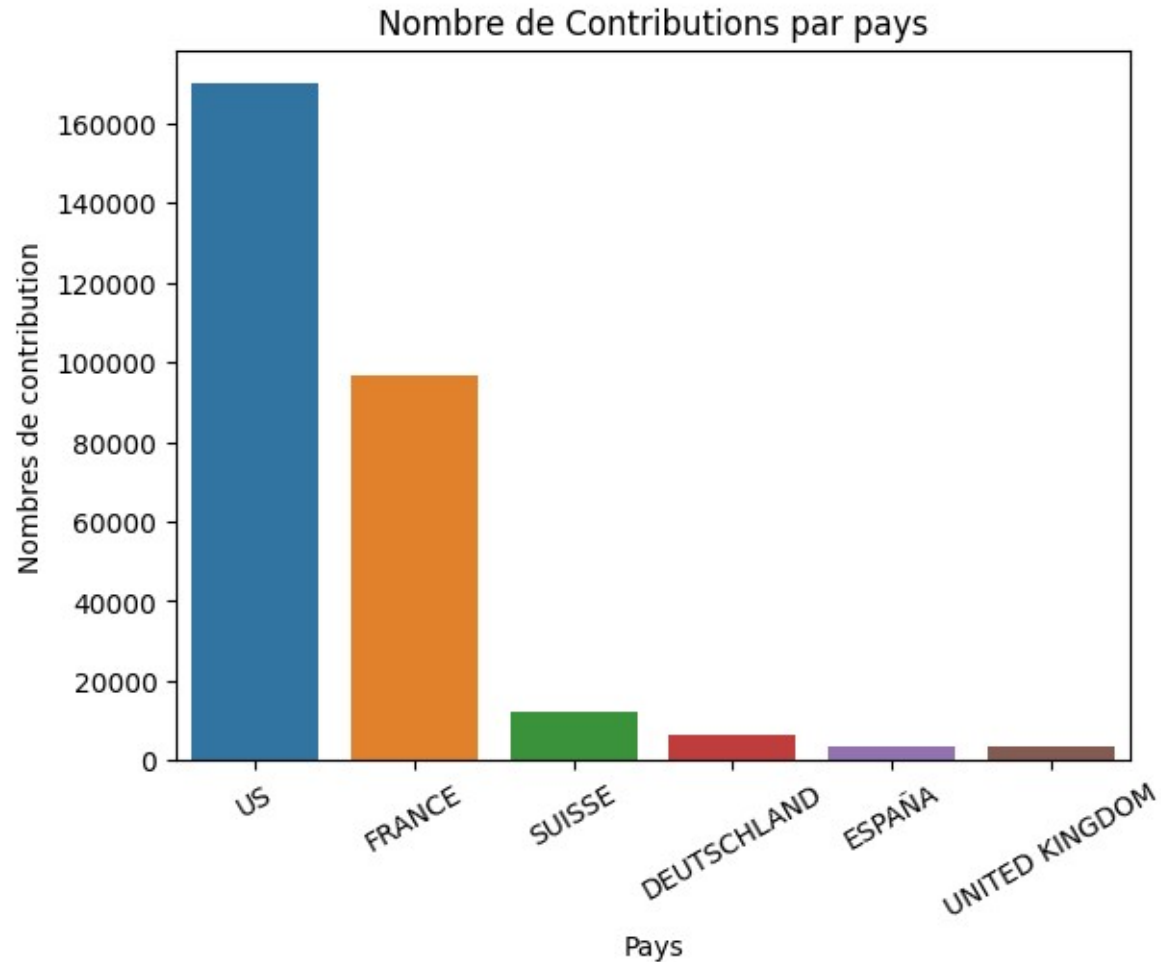
Stéphane Gigandet



Pierre Slamich



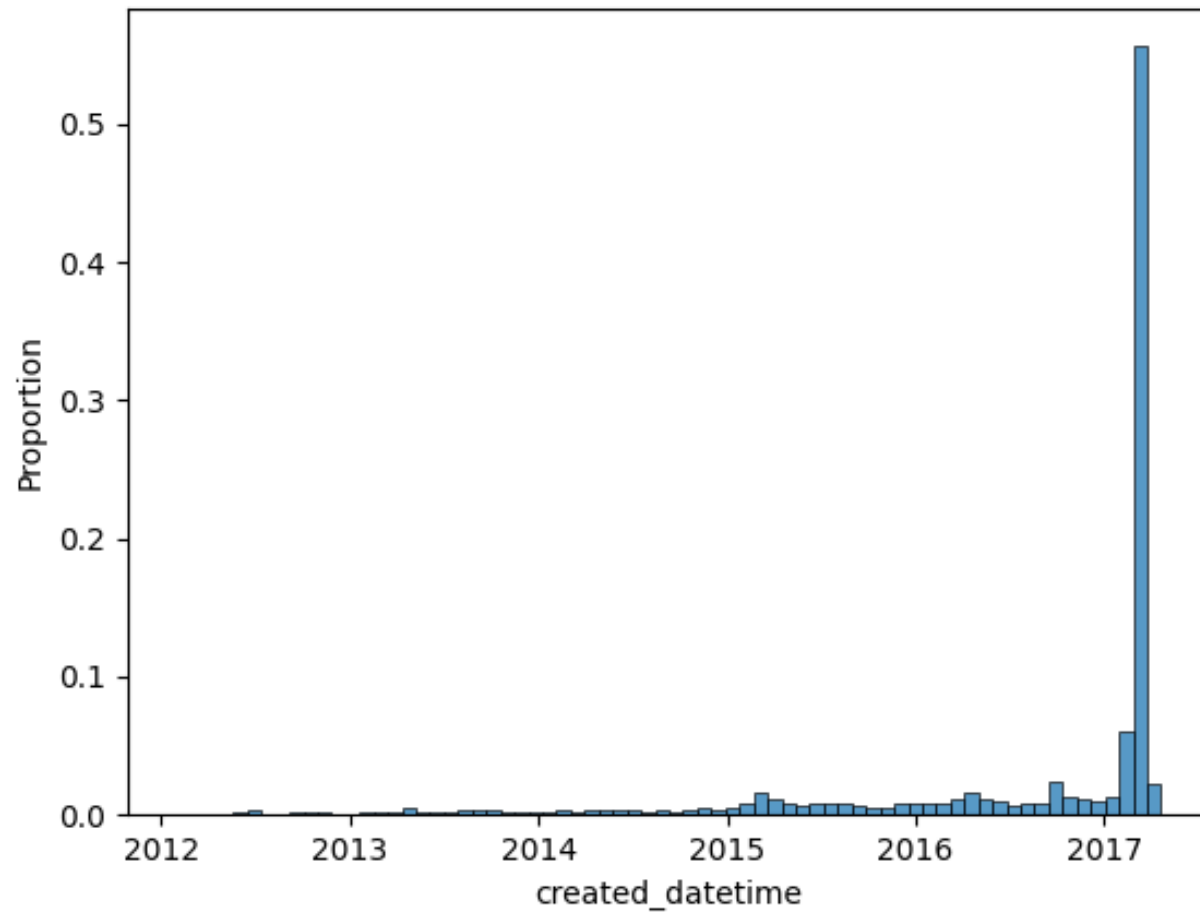
Qualification des données



Biais géographiques importants : la plupart des produits viennent des US et de la France

Qualification des données

Contributions au cours du temps



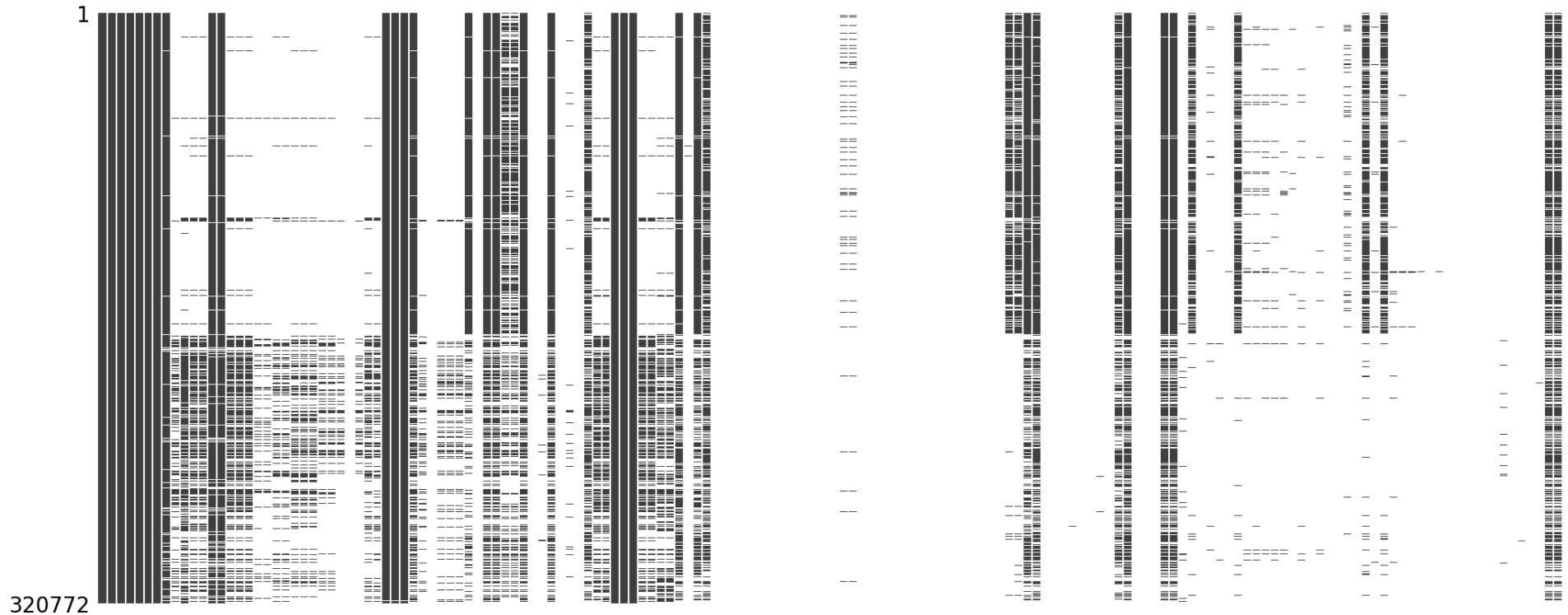
La plupart du dataset date de 2017

Qualification des données

	creator	count	filling_rate
0	usda-ndb-import	169868	0.515054
1	openfoodfacts-contributors	40117	0.410590
2	kiliweb	13891	0.373552
3	date-limite-app	11918	0.284100
4	openfood-ch-import	11478	0.365856
5	tacite	7214	0.662406
6	sebleouf	5864	0.649794
7	tacinte	4398	0.471811
8	javichu	2768	0.695856
9	stephane	2691	0.532723

Les contributeurs les plus importants sont des projets.
Leur taux de remplissage n'est pas très élevé.

Nettoyage des données



Dataset original très éparse

Shape : 320772 x 162

Nettoyage des données

1. Conversion des types des colonnes
2. Suppression des doublons et colonnes avec un taux de valeurs non nulles trop faible (< 3.5 %)
3. Suppression des aliments ayant un code invalide

	Column	Dtype	Non-null Count (%)
6	last_modified_datetime	object	100.000000
5	last_modified_t	object	100.000000
2	creator	object	99.999377
3	created_t	object	99.999065
4	created_datetime	object	99.997194
0	code	object	99.992830
1	url	object	99.992830

Nettoyage des données

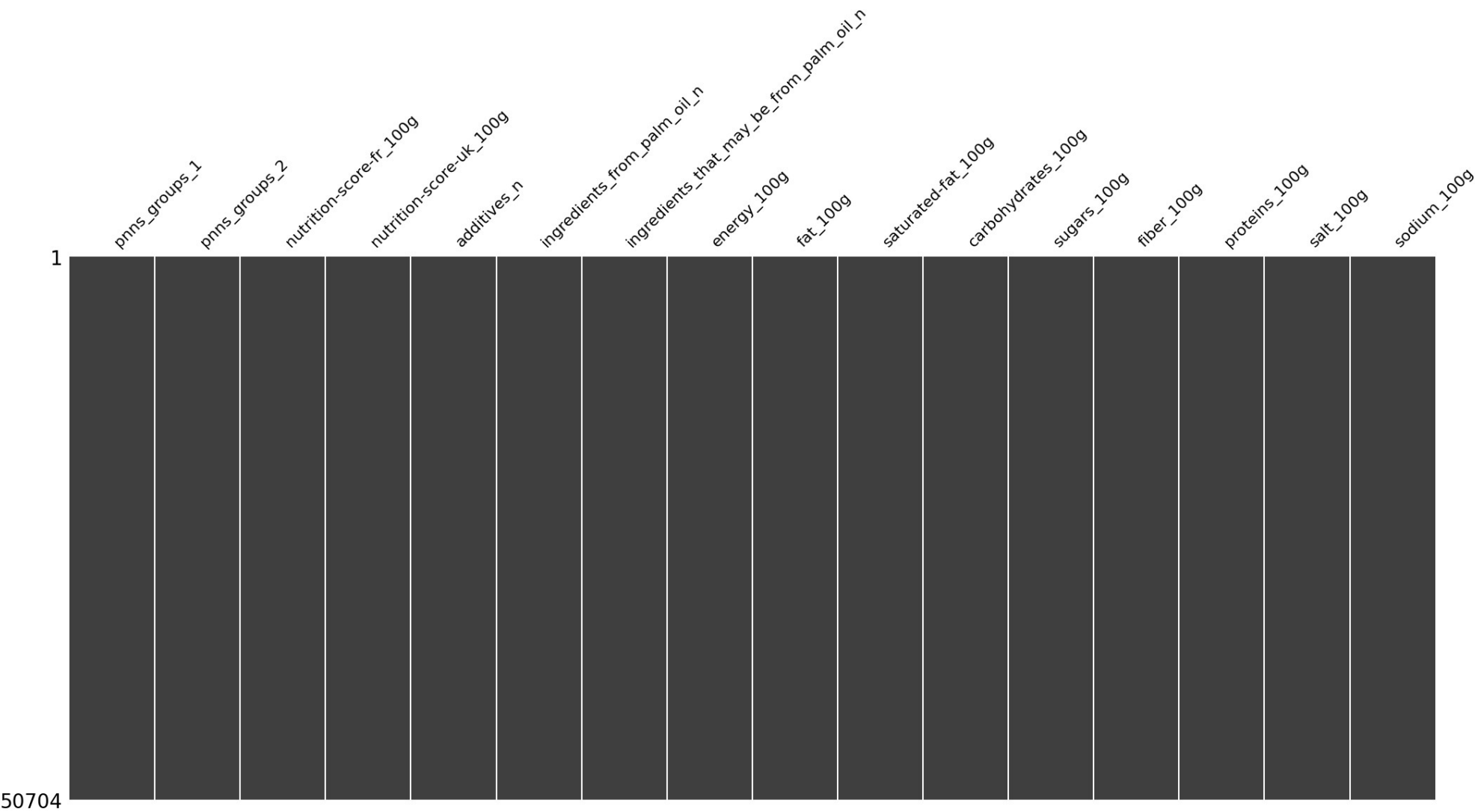
1. Conversion des types des colonnes
2. Suppression des colonnes avec un taux de valeurs non nulles trop faible ($< 3.5\%$)
3. Suppression des aliments ayant un code invalide

	Column	Dtype	Non-null Count (%)
0	code	object	100.000000
1	url	object	100.000000
3	created_t	float64	100.000000
6	last_modified_datetime	datetime64[ns]	100.000000
5	last_modified_t	float64	100.000000
4	created_datetime	datetime64[ns]	99.999688
2	creator	object	99.999376

Nettoyage des données

1. Conversion des types des colonnes
2. Suppression des colonnes avec un taux de valeurs non nulles trop faible ($< 3.5\%$)
3. Suppression des aliments ayant un code invalide
4. Selection des colonnes les plus importantes
5. Remplissage des valeurs manquante (KNN imputer)

Nettoyage des données



Shape 50704 x 16

Nettoyage des données

Nombre de produits par catégories

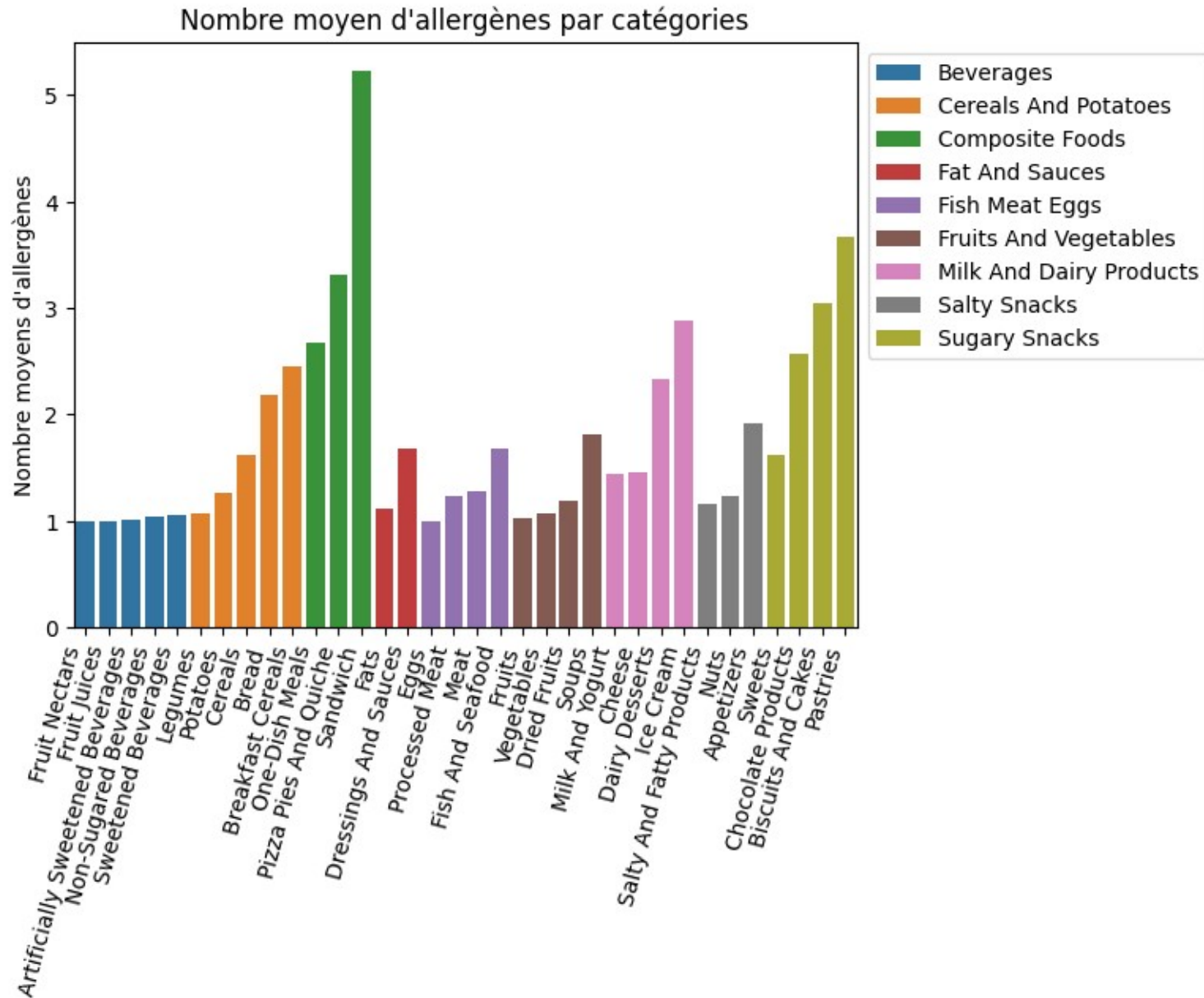
pnns_groups_1	pnns_groups_2	
Beverages	Non-Sugared Beverages	1876
	Sweetened Beverages	1712
	Fruit Juices	1592
	Fruit Nectars	290
	Artificially Sweetened Beverages	227
	Alcoholic Beverages	59
Cereals And Potatoes	Cereals	3315
	Bread	1483
	Breakfast Cereals	1266
	Legumes	661
	Potatoes	77
Composite Foods	One-Dish Meals	4520
	Sandwich	625
	Pizza Pies And Quiche	449
Fat And Sauces	Dressings And Sauces	2560
	Fats	1168
Fish Meat Eggs	Processed Meat	2405
	Fish And Seafood	1805
	Meat	1000
	Eggs	152
	Tripe Dishes	48
Fruits And Vegetables	Vegetables	2224
	Fruits	1186
	Soups	419
	Dried Fruits	366
	Cheese	2951
Milk And Dairy Products	Milk And Yogurt	2606
	Dairy Desserts	692
	Ice Cream	612
	Appetizers	1817
Salty Snacks	Nuts	522
	Salty And Fatty Products	18
Sugary Snacks	Biscuits And Cakes	3877
	Sweets	3212
	Chocolate Products	2517
	Pastries	395

Nettoyage des données

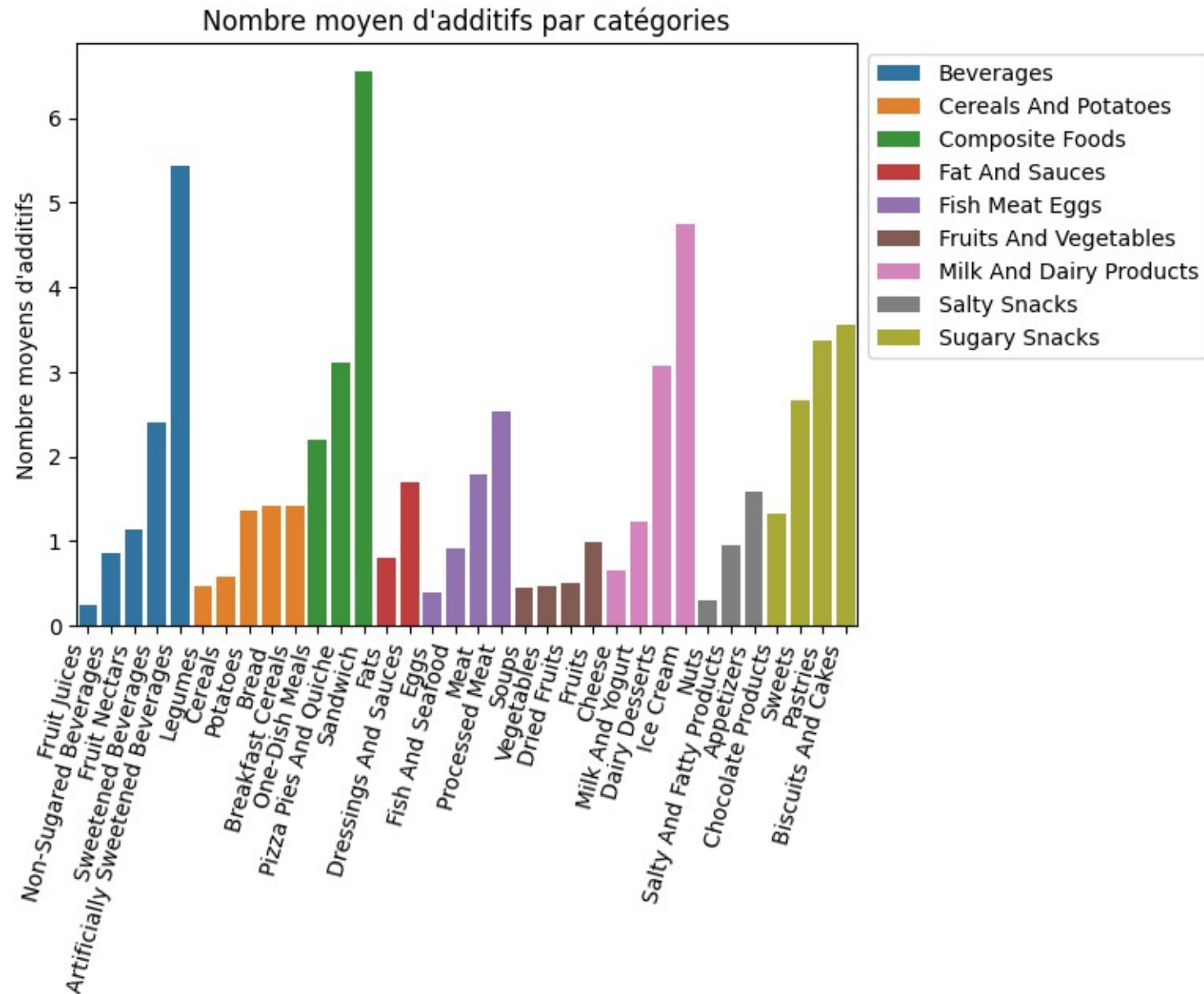
Toute les catégories contiennent des produits sains

pnnns_groups_1	pnnns_groups_2	
Beverages	Non-Sugared Beverages	a
	Sweetened Beverages	a
	Artificially Sweetened Beverages	a
Cereals And Potatoes	Cereals	a
	Bread	a
	Breakfast Cereals	a
	Legumes	a
	Potatoes	a
Composite Foods	One-Dish Meals	a
	Sandwich	a
	Pizza Pies And Quiche	a
Fat And Sauces	Dressings And Sauces	a
	Fats	a
Fish Meat Eggs	Processed Meat	a
	Fish And Seafood	a
	Meat	a
	Eggs	a
	Tripe Dishes	a
Fruits And Vegetables	Vegetables	a
	Fruits	a
	Soups	a
	Dried Fruits	a
Milk And Dairy Products	Cheese	a
	Milk And Yogurt	a
	Dairy Desserts	a
	Ice Cream	a
Salty Snacks	Appetizers	a
	Nuts	a
Sugary Snacks	Salty And Fatty Products	a
	Biscuits And Cakes	a
	Sweets	a
	Chocolate Products	a
	Pastries	a

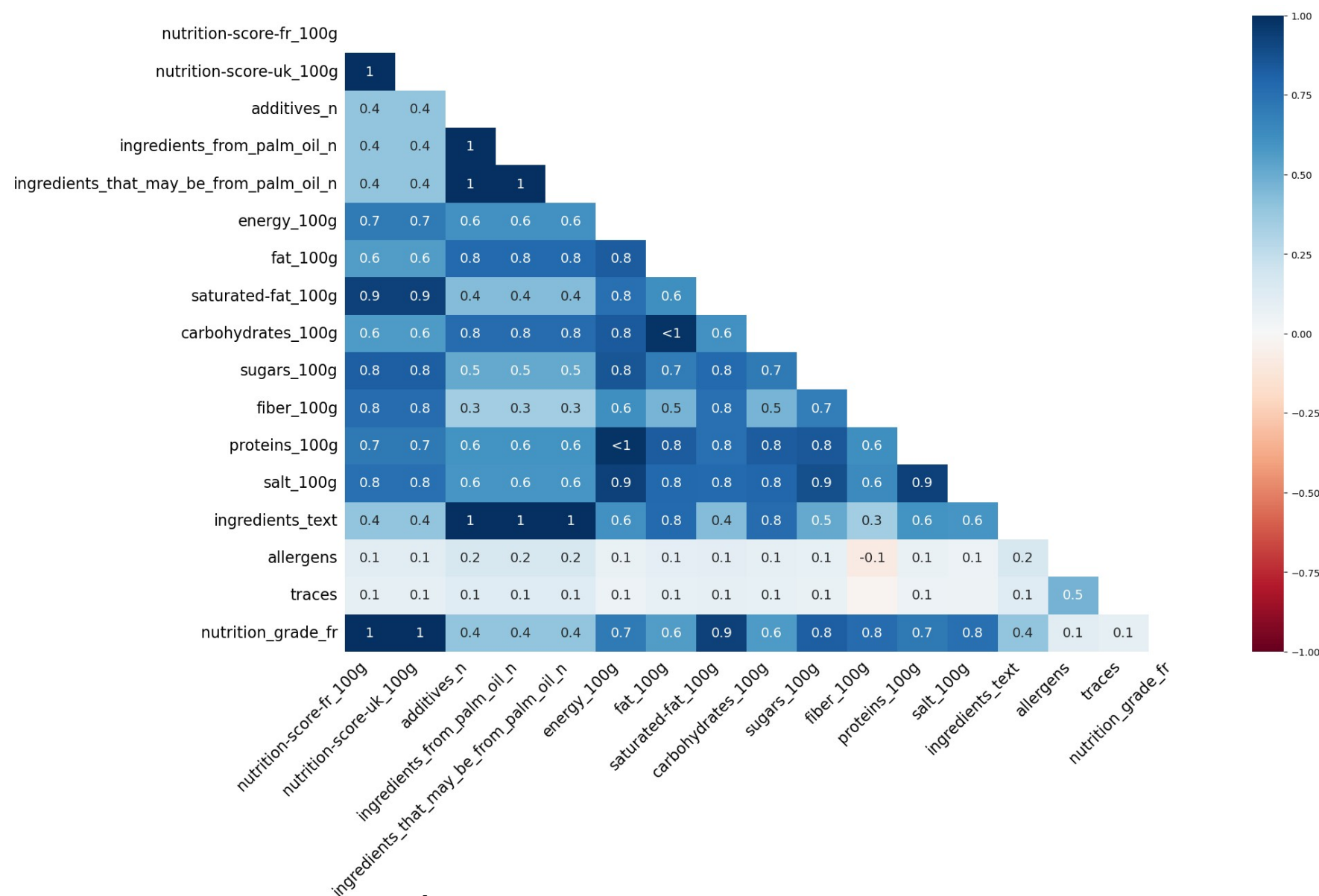
Exploration de données



Exploration de données

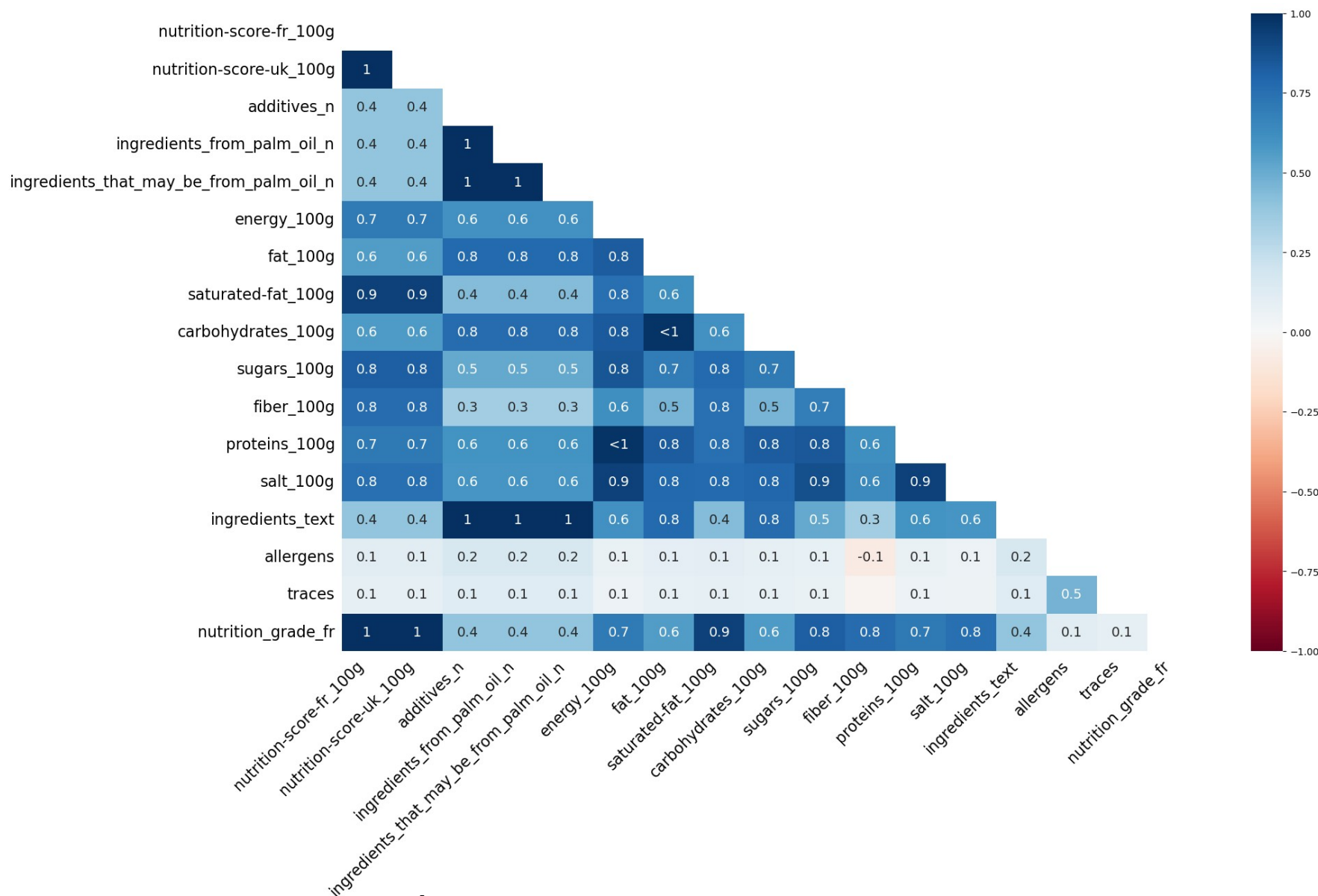


Exploration de données



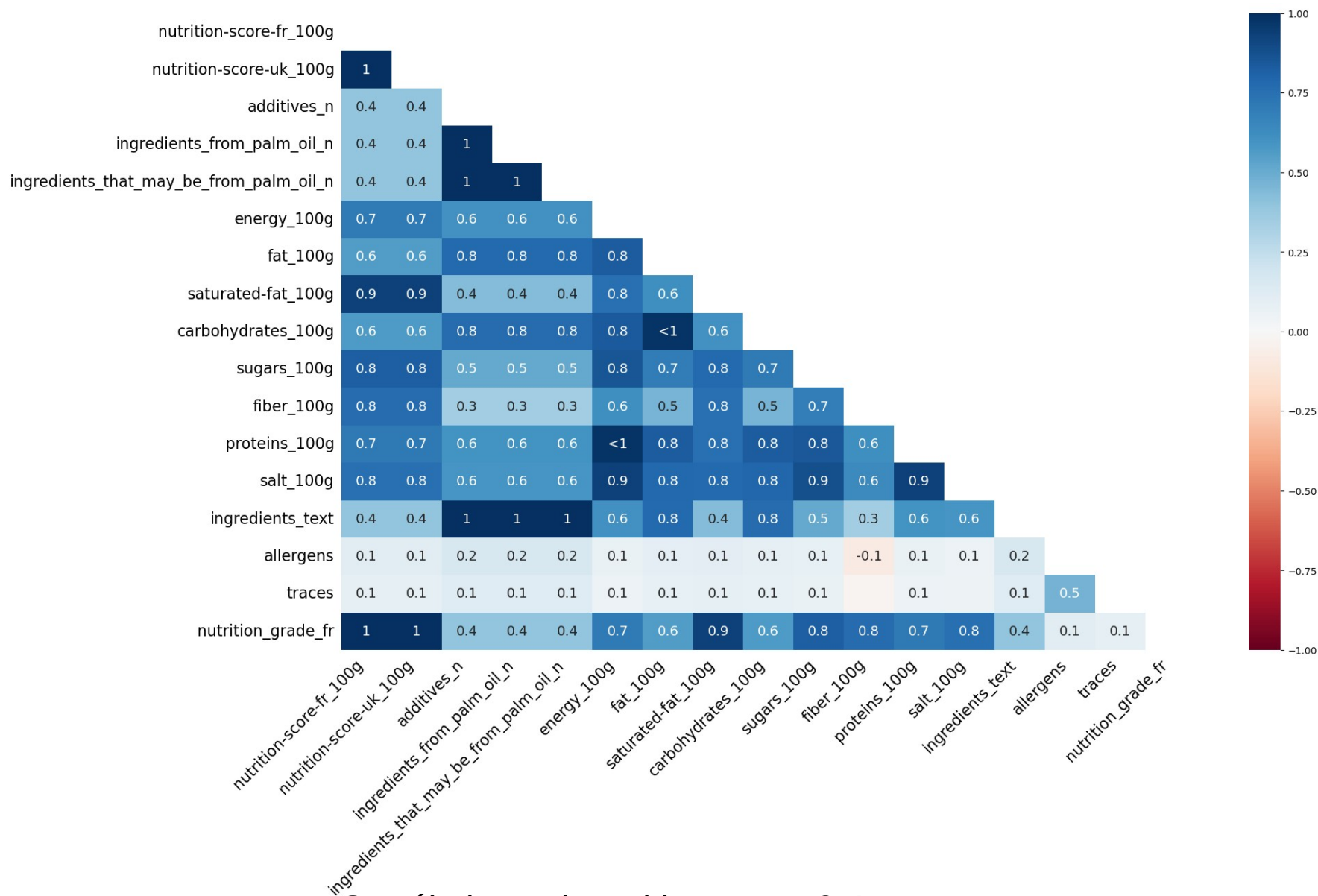
Corrélation sucre-nutrition score 0.8

Exploration de données



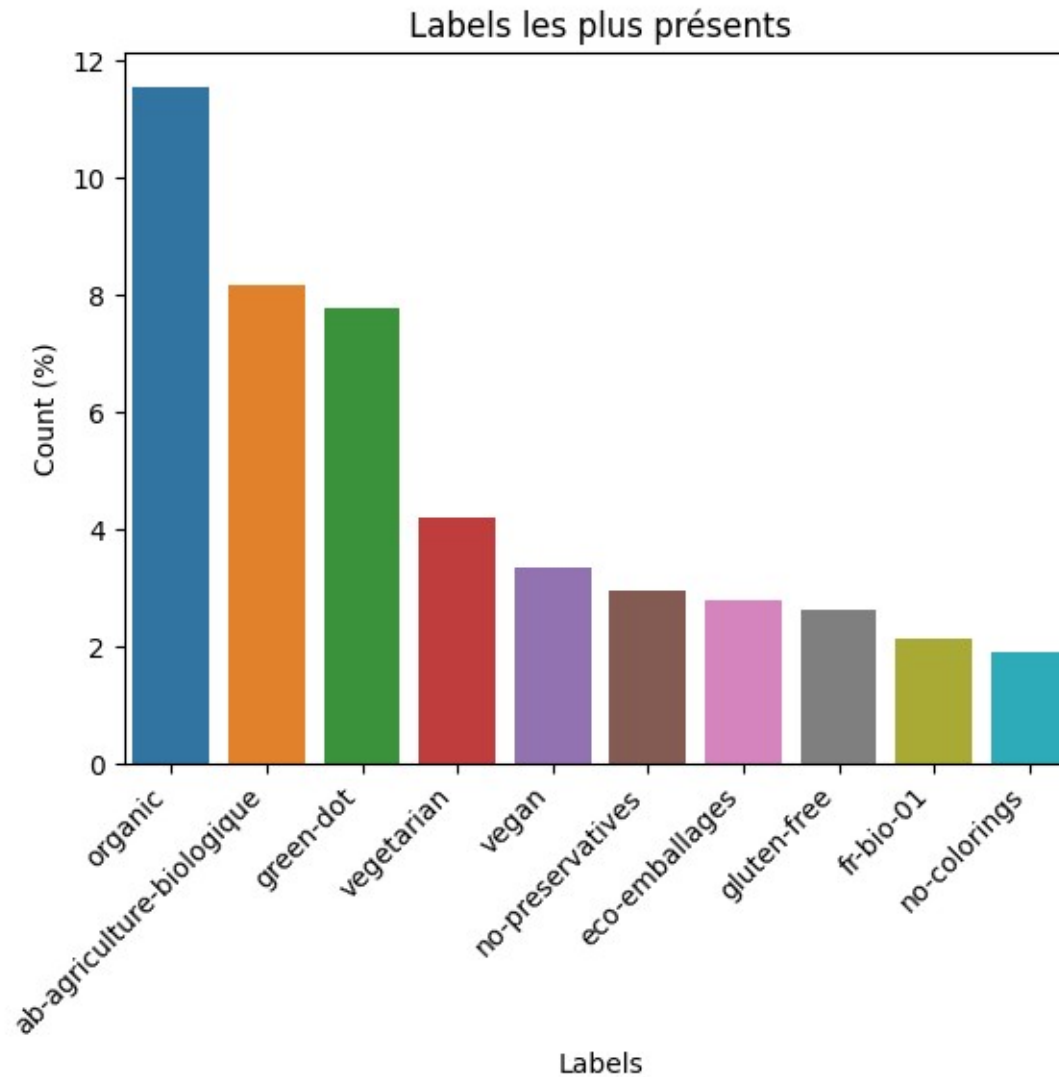
Corrélation positive additifs-nutrition score 0.4

Exploration de données



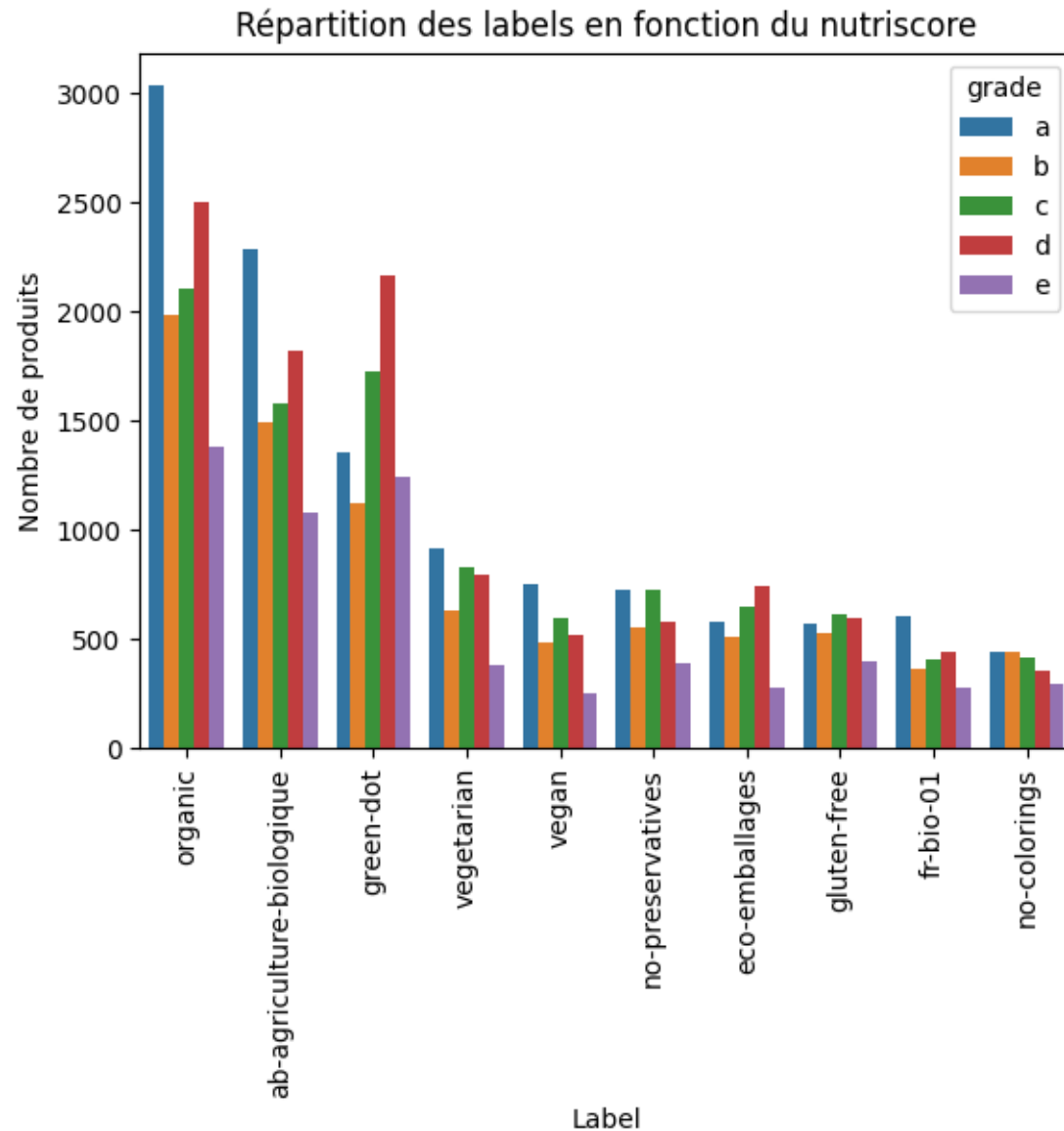
Corrélation sel-nutrition score 0.4

Exploration de données



Labels les plus présents

Exploration de données



Merci pour votre attention