

De l'équilibre dans votre caddie!

Une application innovante en lien avec l'alimentation

Appel à projets de :



À partir de la base de données :



Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

Nettoyage

Suppression de lignes / code

Récupération des doses servies

Traitement des valeurs aberrantes

Suppression de lignes / nutriments

Détermination de groupes

Nettoyage des groupes

Imputation des groupes

Imputation des nutriments

Imputation des serving_size

Exploration

Conclusion

Sommaire

- Présentation ÉQUI CADDIE
- Sélection des données
- Nettoyage
 - Suppression de lignes / code
 - Récupération des doses servies
 - Traitement des valeurs aberrantes
 - Suppression de lignes / nutriments
 - Détermination de groupes
 - Nettoyage des groupes
 - Imputation des groupes
 - Imputation des nutriments
 - Imputation des serving_size
- Exploration
- Conclusion

Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size




Exploration

Conclusion

Vous faites vos courses et scannez vos produits avec votre smartphone ou la douchette du magasin



Vous savez en temps réel si l'équilibre alimentaire et les besoins nutritionnels recommandés par l'ANSES sont respectés

Énergie	Macro-nutriments
54 000 kcal Caddie pour 10 jours pour 2 adultes	 
Vitamines	Minéraux et fibres
vitamine D : 600 / 300 µg ↘ vitamine A : 600 / 15000 µg ↗	Fibres : 200 / 600 g ↗ 

Et vous bénéficiez d'éventuelles recommandations pour + d'équilibre!

Recommandations

Ajout de 400g de fruits à coques : Vitamine A
Fibres



à remplacer par

pour réduire
l'apport en
vitamine D



Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

Nom du produit

Code à barres

Catégorie ou
groupe d'aliment

Quantité de
l'aliment servie

Pays de vente

Nombre d'additifs

Indication de non
présence gluten

Pays de production

Adapté au vegan ou
végétarien

Présence d'huile de
palme

Empreinte carbone

Produit biologique

Liste des
nutriments repris
par l'ANSES pour
les apports
nutritionnels
journaliers
conseillés

Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration

Conclusion

nutriment observé	unité ANC	unité openfoodfacts	Remarques	Fourchette acceptée
energie	kcal	kJ/100g	la valeur maximale est apportée par les lipides et correspond à 37kJ/g	0-3700 kJ
fibre	g	g/100g	teneur max pour la cannelle de 52g/100g	0-75 g
proteine	%(proteine+lipide+glucide)	g/100g	teneur max pour la gélatine alimentaire de 87,6g/100g	0-100 g
lipide	%(proteine+lipide+glucide)	g/100g	teneur max pour l'huile de 100g/100g	0-100 g
glucide	%(proteine+lipide+glucide)	g/100g	teneur max pour le sucre blanc de 99,6g/100g	0-100 g
vitamine A	µgER	g/100g	teneur max pour les abats de dinde de 10737µg/100g	0-15 mg
vitamine B1	mg	g/100g	teneur max pour la levure alimentaire de 40mg/100g	0-60 mg
vitamine B2	mg	g/100g	teneur max pour le foie d'agneau de 4,3mg/100g	0-10 mg
vitamine B3	mg	g/100g	teneur max pour les viandes blanches de 20mg/100g	0-50 mg
vitamine B5	mg	g/100g	teneur max pour les champignons shiitakes séchés de 22mg/100g	0-50 mg
vitamine B6	mg	g/100g	teneur max pour la levure de bière de 2,6mg/100g	0-10 mg
vitamine B9	µgEFA	g/100g	teneur max pour les abats de volaille de 770µg/100g	0-1,5 mg
vitamine B12	µg	g/100g	teneur max pour le foie de boeuf de 65µg/100g	0-100 µg
vitamine C	mg	g/100g	teneur max pour le cassis de 200mg/100g	0-500 mg
vitamine D	µg	g/100g	teneur max pour l'huile de foie de morue de 250µg/100g	0-500 µg
vitamine E	mg	g/100g	teneur max pour l'huile de germe de blé de 21mg/100g	0-50 mg
calcium	mg	g/100g	teneur max pour les fromages à pâte dure de 1000mg/100g	0-2 g
cuivre	mg	g/100g	teneur max pour le foie de veau cuit de 15mg/100g	0-30 mg
fer	mg	g/100g	teneur max pour le meloukhia en poudre de 87mg/100g	0-200 mg
iode	µg	g/100g	teneur max pour le sel enrichi en iode de 1860µg/100g	0-3 mg
magnesium	mg	g/100g	teneur max pour le meloukhia en poudre de 609mg/100g	0-1 g
manganese	mg	g/100g	teneur max pour les germes de blé de 11mg/100g	0-20 mg
phosphore	mg	g/100g	teneur max pour la levure alimentaire de 1300mg/100g	0-2 g
selenium	µg	g/100g	teneur max pour les huîtres de 154µg/100g	0-300 µg
zinc	mg	g/100g	teneur max pour les huîtres de 33mg/100g	0-50 mg
sel	g	g/100g		0-100 g

Sommaire

Présentation ÉQUI CADDIE Sélection des variables

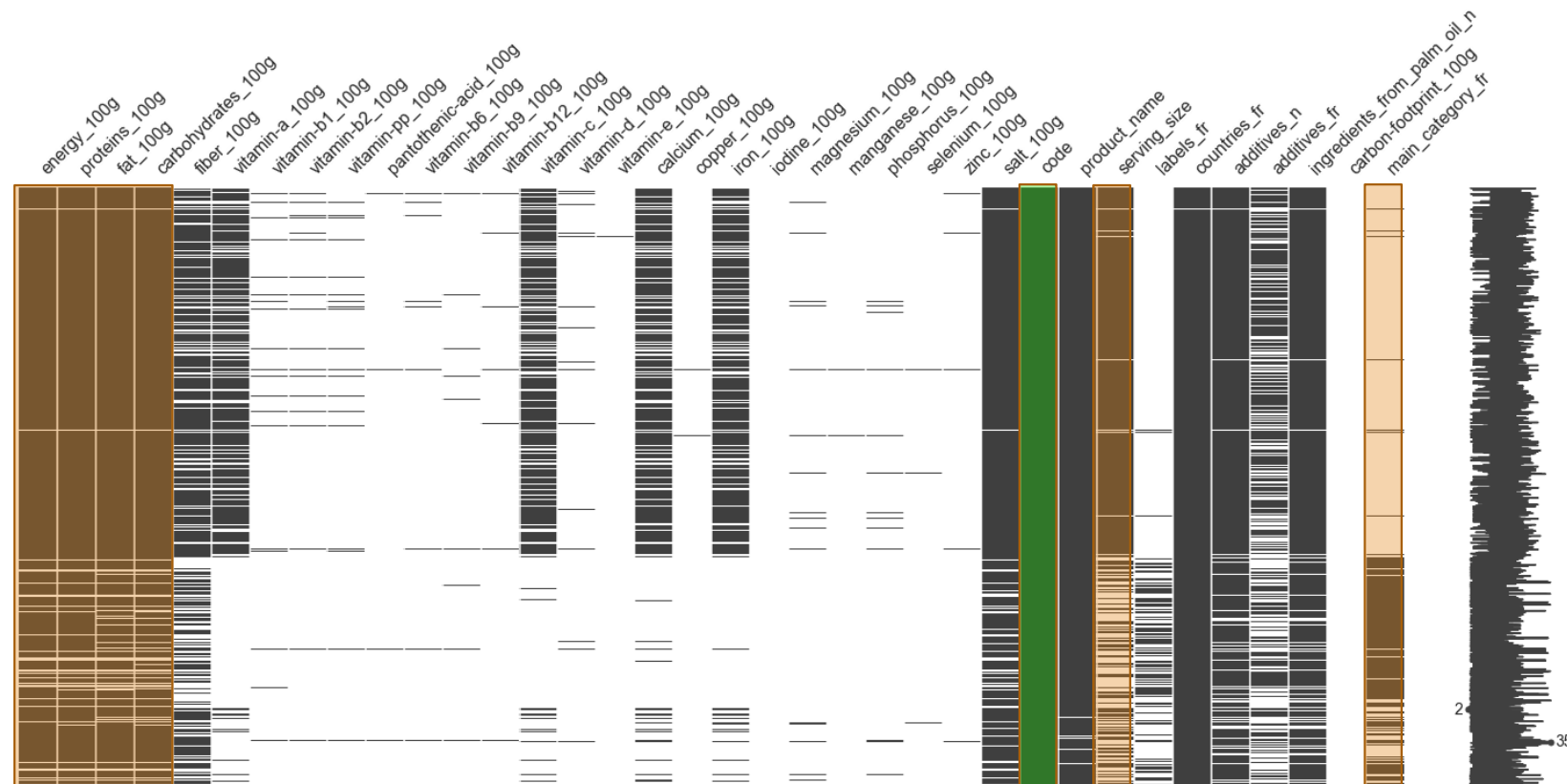
Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Sélection des variables	Nombre de lignes	Avant : 320 772 Après : 320 772	Nombre de valeurs manquantes	Avant : 39 608 milliers Après : 7 500 milliers
-------	-------------------------	------------------	------------------------------------	------------------------------	---

- Passage de 162 colonnes à 36 colonnes



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Suppression des produits sans code	Nombre de lignes	Avant : 320 772 Après : 320 749	Nombre de valeurs manquantes	Avant : 7 500 milliers Après : 7 499 milliers
-------	------------------------------------	------------------	------------------------------------	------------------------------	--

- Besoin de code pour le scan des codes-barres : Suppression de produits sans valeur sur la variable code, ainsi que des doublons (aucun doublon recensé)



Étape	Suppression des produits sans les nutriments de base	Nombre de lignes	Avant : 320 749 Après : 266 525	Nombre de valeurs manquantes	Avant : 7 499 milliers Après : 5 778 milliers
-------	--	------------------	------------------------------------	------------------------------	--

Glucides

Lipides

Protides



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

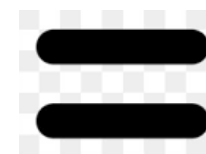
Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

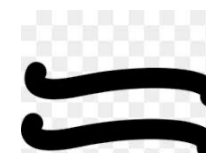
Étape	Serving_size : format texte vers numérique	Nombre de lignes	Avant : 266 525 Après : 266 525	Nombre de valeurs manquantes	Avant : 7 499 milliers Après : 7 785 milliers
-------	---	---------------------	------------------------------------	---------------------------------	--

85 g (85 GRM)
20 g (0.7 oz)
1 g (0.5 PACKET)
...



85
20
1
...

3.2 ml (0.75 TSP | ABOUT)
100 ml (100 ml)
330 ml
...



3.2
100
330
...

Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

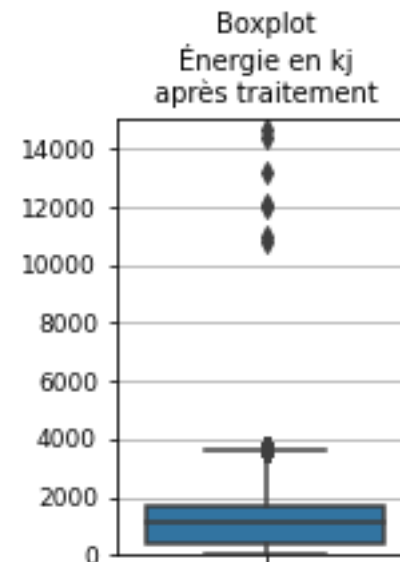
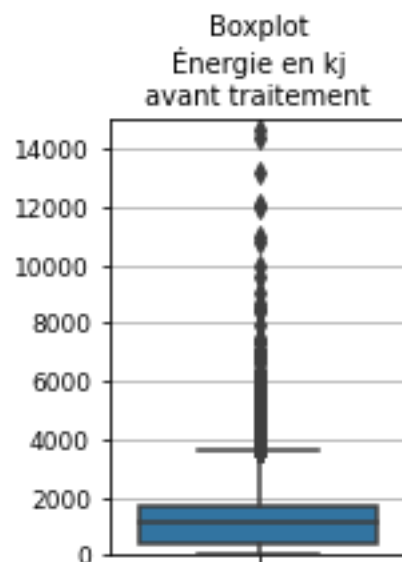
Étape	Traitement des valeurs négatives	Nombre de lignes	Avant : 266 525 Après : 266 525	Nombre de valeurs manquantes	Avant : 7 785 milliers Après : 7 785 milliers
-------	----------------------------------	------------------	------------------------------------	------------------------------	--

-800



800

Étape	Traitement n°1 de la variable energy_100g	Nombre de lignes	Avant : 266 525 Après : 266 525	Nombre de valeurs manquantes	Avant : 7 785 milliers Après : 7 785 milliers
-------	---	------------------	------------------------------------	------------------------------	--



Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

Nettoyage

Suppression de lignes / code

Récupération des doses servies

Traitement des valeurs aberrantes

Suppression de lignes / nutriments

Détermination de groupes

Nettoyage des groupes

Imputation des groupes

Imputation des nutriments

Imputation des serving_size

Exploration

Conclusion

Étape	Valeurs aberrantes des nutriments		Nombre de lignes	Avant : 266 525 Après : 266 525	Nombre de valeurs manquantes	Avant : 7 785 milliers Après : 7 785 milliers
Nutriment	1ère division par 1000		2ème division par 1000	suppression		
energy_100g	19		0	0		
fiber_100g	44		0	0		
proteins_100g	3		0	0		
fat_100g	4		0	0		
carbohydrates_100g	19		0	0	0	
vitamin-a_100g	43		1	0		
vitamin-b1_100g	4815		6	0		
vitamin-b2_100g	5890		13	0		
vitamin-pp_100g	232		0	0		
pantothenic-acid_100g	49		1	0		
vitamin-b6_100g	86		2	0		
vitamin-b9_100g	31		3	0		
vitamin-b12_100g	32		1	0		
vitamin-c_100g	287		1	0		
vitamin-d_100g	3		0	0		
vitamin-e_100g	42		0	0		
calcium_100g	133		0	0		
copper_100g	7		1	0		
iron_100g	31		0	0		
iodine_100g	7		0	0		
magnesium_100g	36		0	0		
manganese_100g	10		0	0		
phosphorus_100g	19		0	0	0	
selenium_100g	7		1	0		
zinc_100g	23		0	0		
salt_100g	158		0	0		

g
mg
µg

Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration Conclusion

Étape	Suppression des lignes aux nutriments incohérents	Nombre de lignes	Avant : 266 525 Après : 254 998	Nombre de valeurs manquantes	Avant : 7 785 milliers Après : 5 564 milliers
-------	---	---------------------	------------------------------------	---------------------------------	--

Σ des poids de nutriments pour 100g d'aliment ≤ 100 g

Sommaire

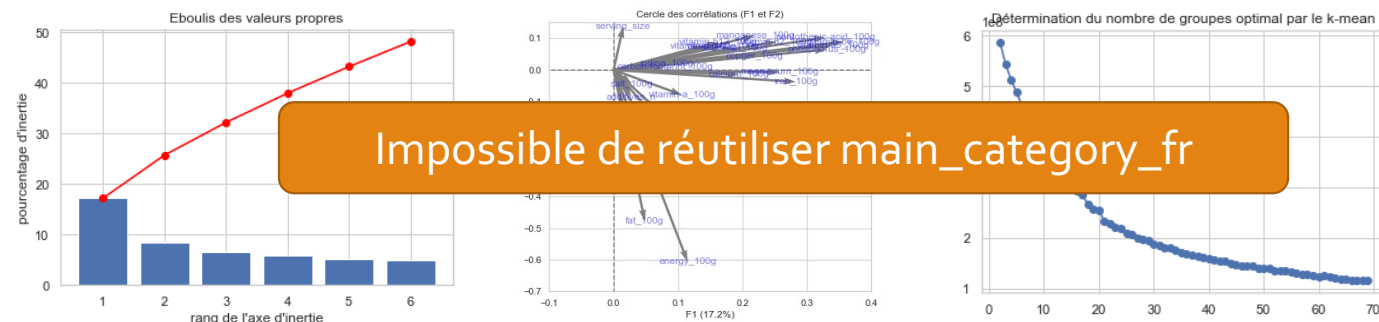
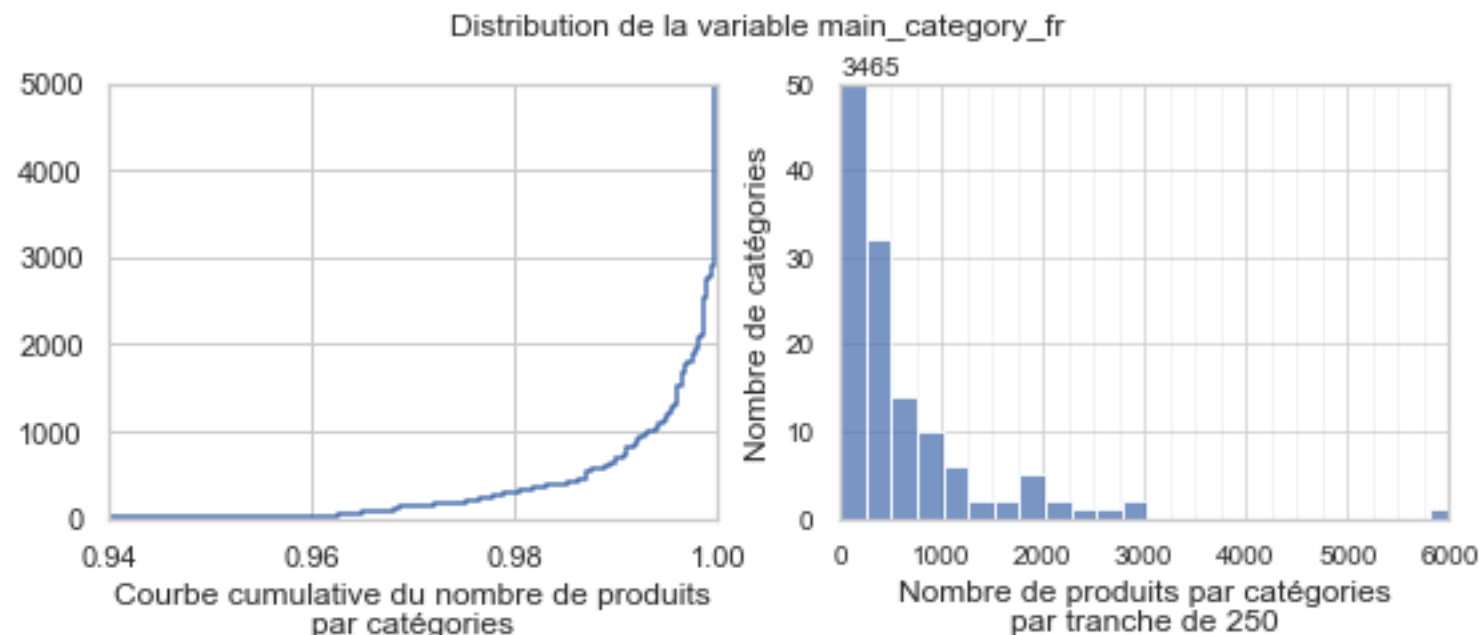
Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration Conclusion

Étape	Détermination des groupes d'aliments	Nombre de lignes	Avant : 254 998 Après : 254 998	Nombre de valeurs manquantes	Avant : 5 564 milliers Après : 5 564 milliers
-------	--------------------------------------	------------------	------------------------------------	------------------------------	--



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Détermination des groupes d'aliments	Nombre de lignes	Avant : 254 998 Après : 254 998	Nombre de valeurs manquantes	Avant : 5 564 milliers Après : 5 564 milliers
	Produits sucrés				
	Céréales et féculents		Plats préparés	Condiments	
	Fruits et légumes		Fruits secs	Chocolats	
	Boissons		Fruits à coques	Fromages	
	Produits laitiers		Foies	Olives	
	Matières grasses				
	Viandes poissons œufs				

Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

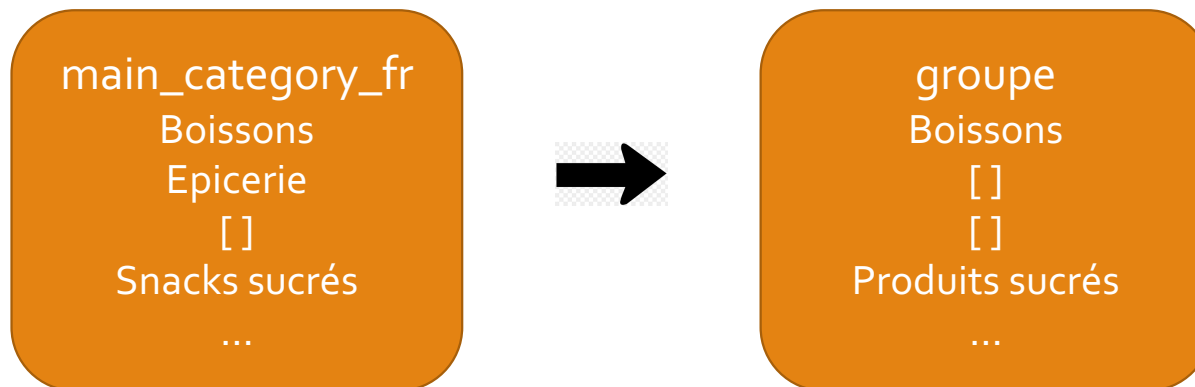
Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

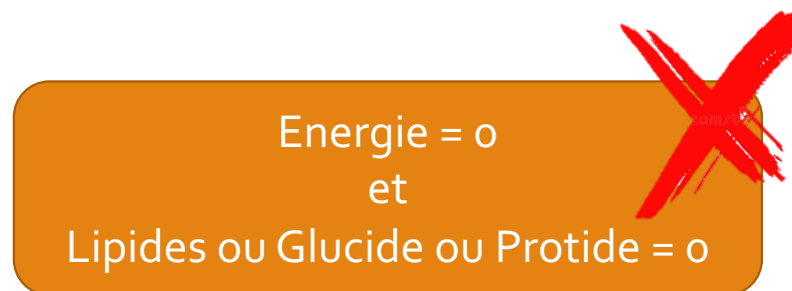
Exploration

Conclusion

Étape	Imputation n°1 des groupes	Nombre de lignes	Avant : 254 998 Après : 254 998	Nombre de valeurs manquantes	Avant : 5 564 milliers Après : 5 763 milliers
-------	----------------------------	------------------	------------------------------------	------------------------------	--



Étape	Suppression de lignes / énergie, LPG	Nombre de lignes	Avant : 254 998 Après : 223 458	Nombre de valeurs manquantes	Avant : 5 763 milliers Après : 4 867 milliers
-------	--------------------------------------	------------------	------------------------------------	------------------------------	--



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration Conclusion

Étape	Nettoyage des groupes	Nombre de lignes	Avant : 223 458 Après : 223 458	Nombre de valeurs manquantes	Avant : 4 867 milliers Après : 4 868 milliers
-------	-----------------------	------------------	------------------------------------	------------------------------	--

Sortir tous les produits qui ne correspondent pas au groupe

Médianes,
moyennes

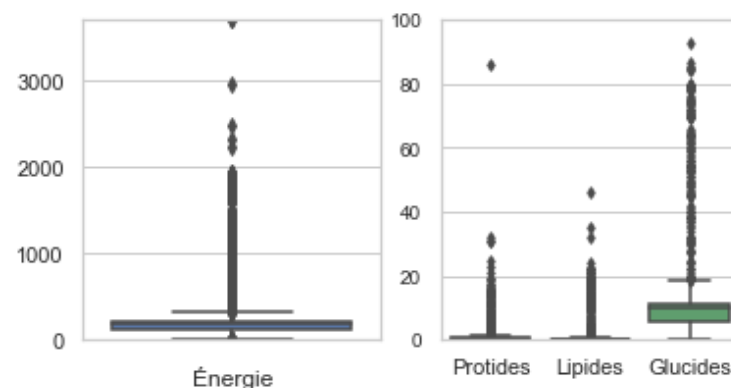
Ne pas s'en préoccuper

Observations des outliers ✓

Chasses aux outliers ✗

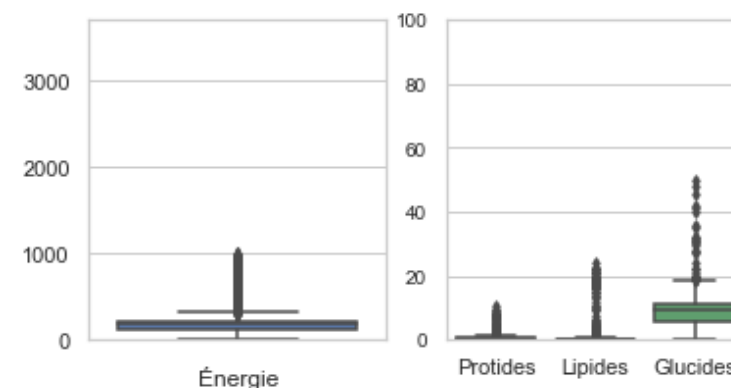
Avant nettoyage

Distribution des principaux nutriments des boissons



Après nettoyage

Distribution des principaux nutriments des boissons



Sommaire

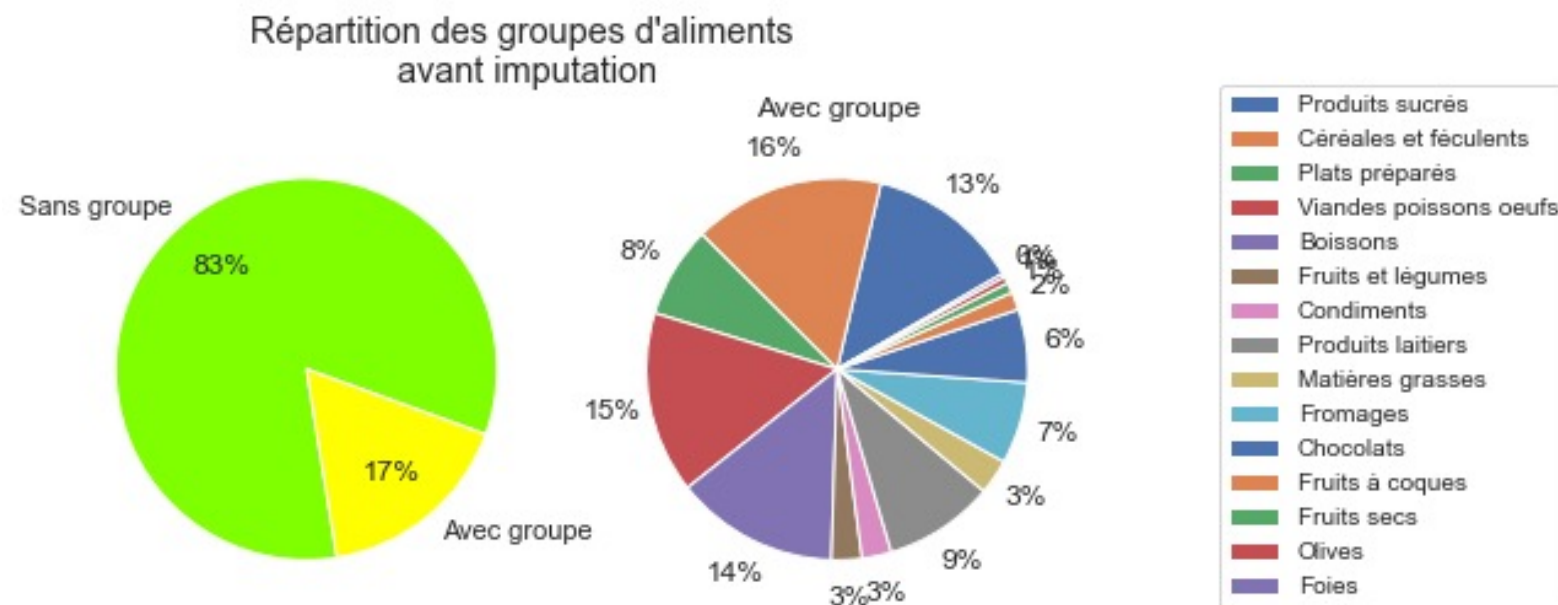
Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Imputation des groupes	Nombre de lignes	Avant : 223 458 Après : 223 458	Nombre de valeurs manquantes	Avant : 4 868 milliers Après : 4 683 milliers
-------	------------------------	------------------	------------------------------------	------------------------------	--



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

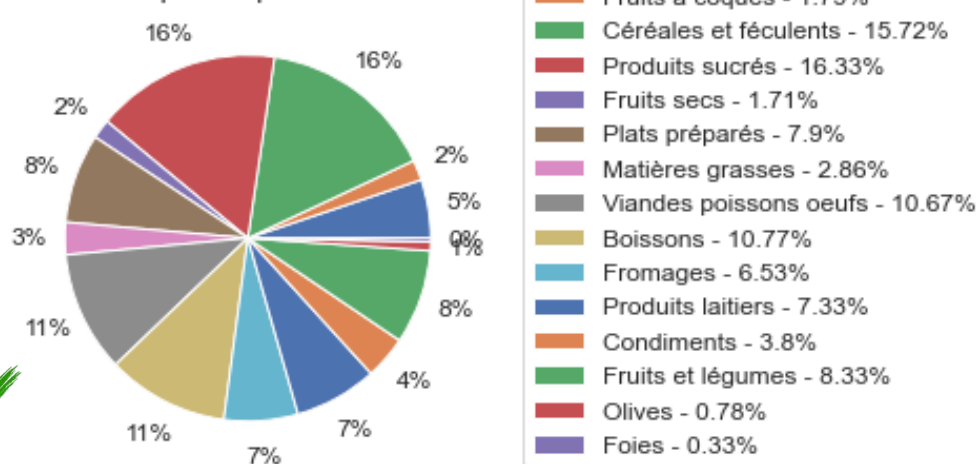
Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Imputation des groupes	Nombre de lignes	Avant : 223 458 Après : 223 458	Nombre de valeurs manquantes	Avant : 4 868 milliers Après : 4 683 milliers
-------	------------------------	------------------	------------------------------------	------------------------------	--

Méthode	Score sur le jeu de données d'entraînement	Score sur les données non connues
reg_avant_nettoyage - o:26	60%	74%
knn_avant_nettoyage - o:26	80%	75%
reg_après_nettoyage - o:26	67%	76%
knn_après_nettoyage - o:26	86%	85%
knn_après_nettoyage - o:5	81%	80%

Répartition des groupes d'aliments après imputation



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

Suppression de lignes / code
Récupération des doses servies
Traitement des valeurs aberrantes
Suppression de lignes / nutriments
Détermination de groupes
Nettoyage des groupes
Imputation des groupes
Imputation des nutriments
Imputation des serving_size

Exploration Conclusion

Étape	Imputation des nutriments		Nombre de lignes		Avant : 223 458 Après : 223 458		Nombre de valeurs manquantes		Avant : 4 683 milliers Après : 2 066 milliers	
groupe	energy_100g	carbohydrates_100g	fat_100g	proteins_100g	calcium_100g	phosphorus_100g	vitamin-a_100g	vitamin-b2_100g		
Boissons	180.0	9.60	0.00	0.30	NaN	NaN	NaN	NaN		
Produits laitiers	335.0	6.10	2.90	3.39	0.1200	0.094	0.000063	0.000205		
Produits sucrés	1510.0	61.00	3.95	3.40	NaN	NaN	NaN	NaN		
Viandes poissons oeufs	791.0	1.00	10.20	19.00	0.0357	NaN	0.000000	NaN		
Fruits et légumes	127.0	3.60	0.30	1.40	NaN	NaN	0.000340	NaN		
Céréales et féculents	1493.0	57.70	4.00	8.00	0.0357	0.209	NaN	0.000455		
Condiments	652.0	9.35	2.35	4.20	NaN	NaN	NaN	NaN		
Matières grasses	3358.0	0.00	91.00	0.00	NaN	NaN	NaN	NaN		
Chocolats	2290.0	50.55	35.00	6.80	0.1655	NaN	NaN	NaN		
Fruits à coques	2570.0	11.40	51.00	20.00	NaN	0.466	NaN	NaN		
Plats préparés	587.0	13.30	5.80	6.40	NaN	NaN	NaN	NaN		
Fromages	1292.0	1.00	25.00	19.00	NaN	NaN	NaN	NaN		
Olives	682.5	1.00	16.20	1.20	NaN	NaN	NaN	NaN		
Fruits secs	1268.5	63.25	0.50	2.50	NaN	NaN	NaN	NaN		
Foies	2104.0	2.20	53.00	6.80	NaN	NaN	NaN	NaN		

ÉQUIPAGE

Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

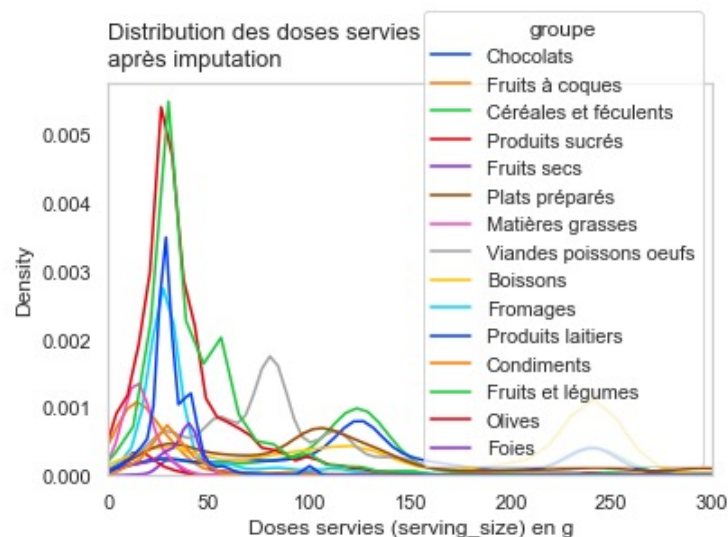
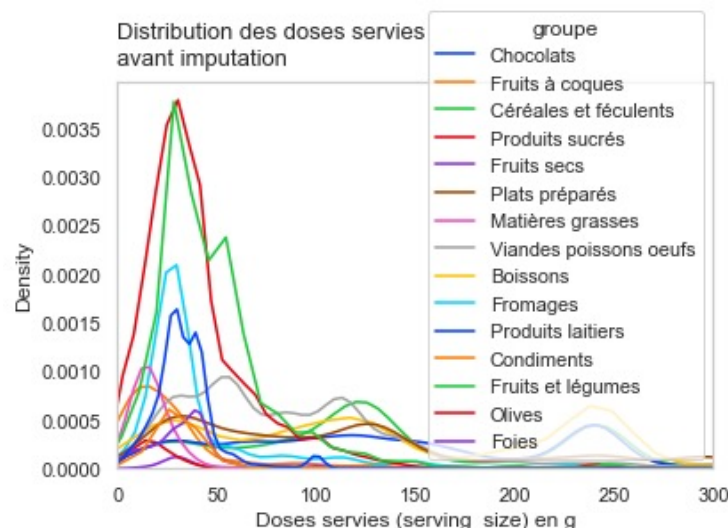
Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

Étape	Imputation des serving_size	Nombre de lignes	Avant : 223 458 Après : 223 458	Nombre de valeurs manquantes	Avant : 2 066 milliers Après : 2 023 milliers
-------	-----------------------------	------------------	------------------------------------	------------------------------	--



groupe	mode	mediane	skewness	kurtosis
Chocolats	28.0	31.0	6.468112	99.425079
Fruits à coques	30.0	30.0	48.898182	2642.961385
Céréales et féculents	28.0	40.0	6.872373	143.068347
Produits sucrés	28.0	32.0	5.717744	72.490101
Fruits secs	40.0	33.0	7.472563	106.906498
Plats préparés	130.0	103.9	1.185824	1.222161
Matières grasses	14.0	15.0	8.624180	101.423069
Viandes poissons oeufs	56.0	80.0	2.026781	6.190716
Boissons	240.0	125.0	3.082442	39.661366
Fromages	28.0	28.0	3.843552	26.203441
Produits laitiers	240.0	125.0	0.732320	1.832550
Condiments	15.0	17.0	4.373520	28.567034
Fruits et légumes	240.0	125.0	0.982423	1.713910
Olives	15.0	15.0	21.109255	587.176930
Foies	30.0	30.0	7.134972	67.488184

Sommaire

Présentation ÉQUI CADDIE

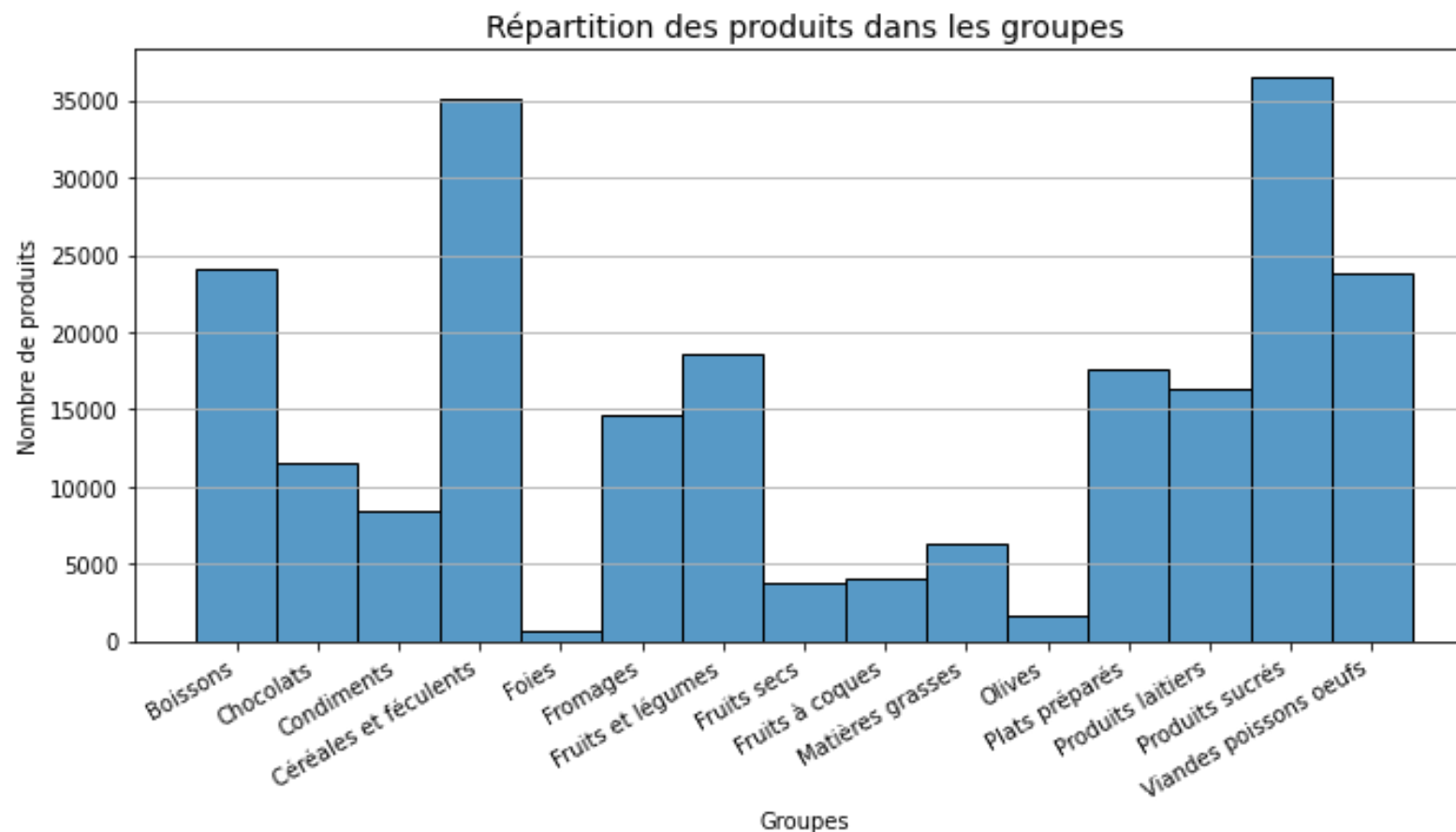
Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion



Énergie, glucides, lipides, protides par groupe d'aliment

Sommaire

Présentation ÉQUI CADDIE

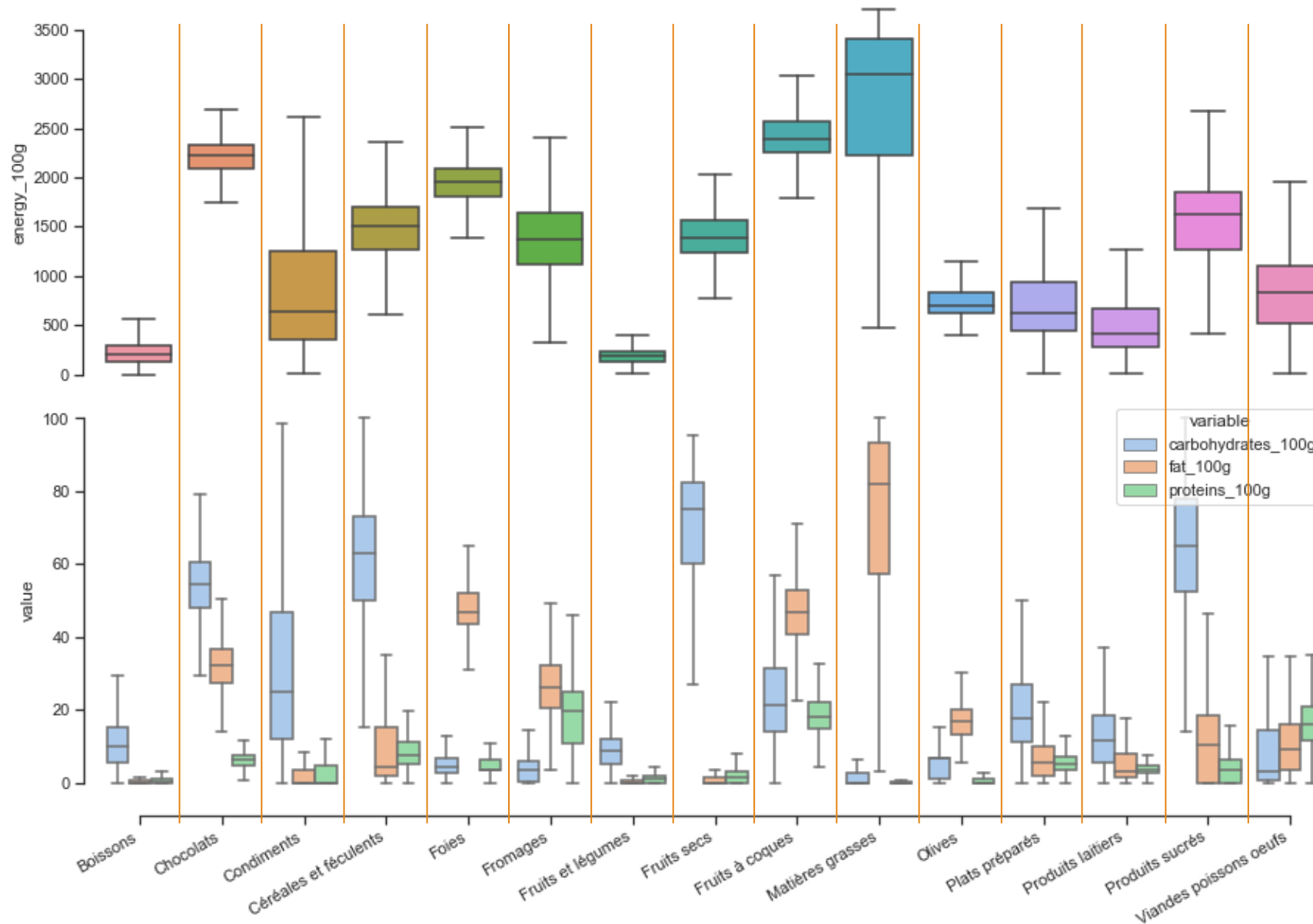
Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion



Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

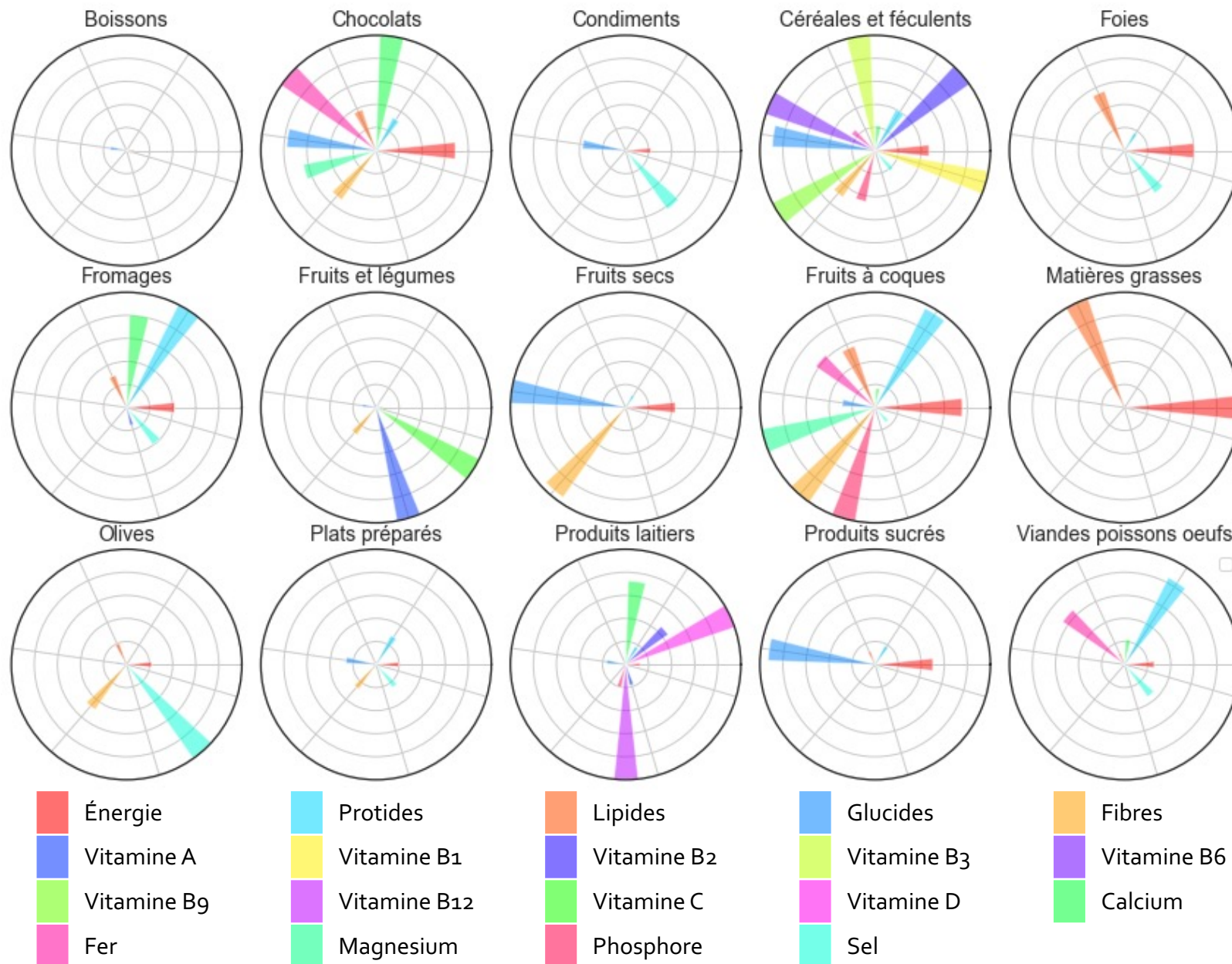
Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

Répartition des nutriments dans chaque groupe



Sommaire

Présentation ÉQUI CADDIE

Sélection des variables

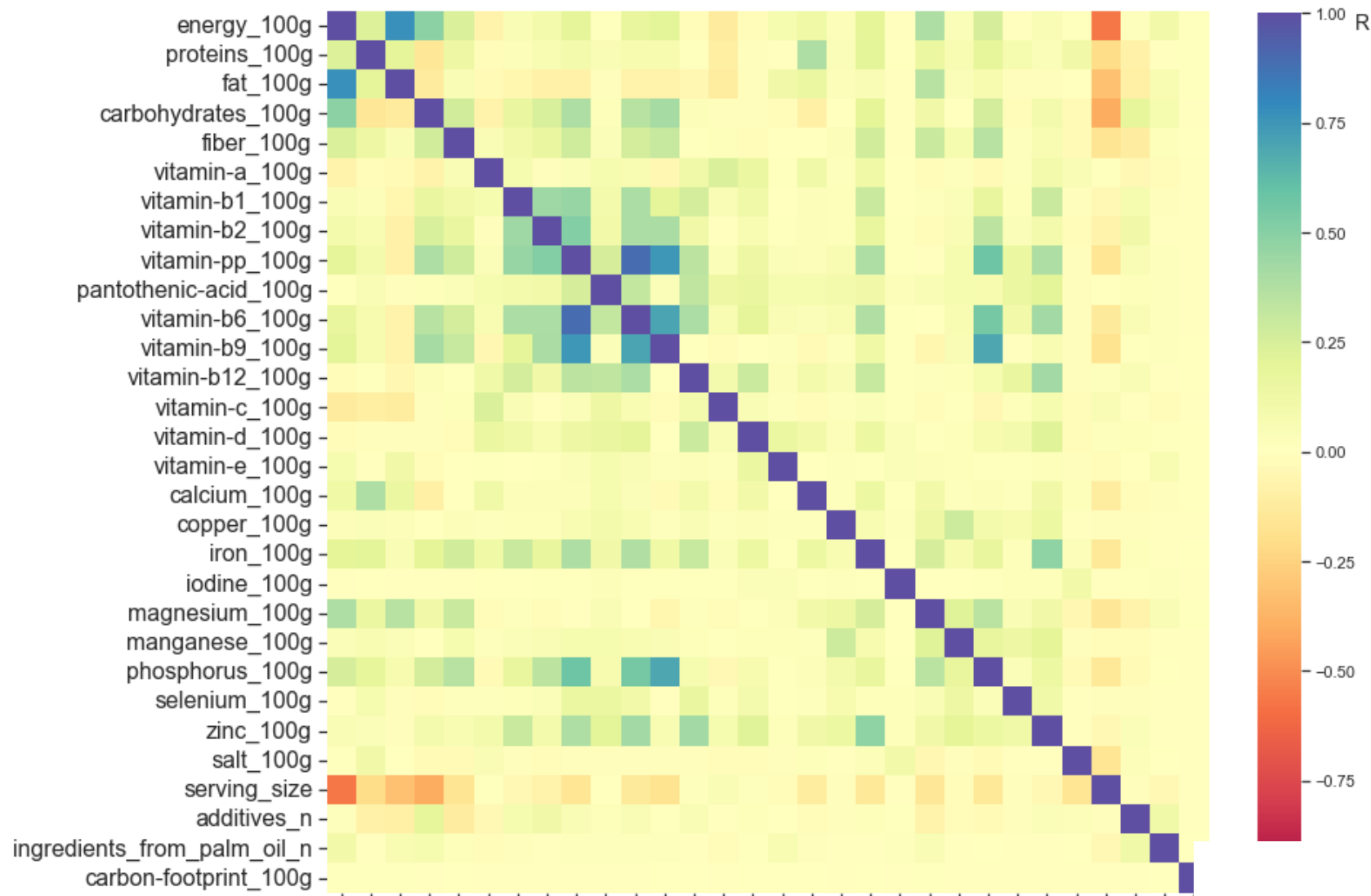
Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

Corrélations entre les différentes variables



Sommaire

Présentation ÉQUI CADDIE Sélection des variables

Nettoyage

- Suppression de lignes / code
- Récupération des doses servies
- Traitement des valeurs aberrantes
- Suppression de lignes / nutriments
- Détermination de groupes
- Nettoyage des groupes
- Imputation des groupes
- Imputation des nutriments
- Imputation des serving_size

Exploration

Conclusion

85% des produits bien catégorisés

Proposition de substitut dans le même groupe ✓

Groupes d'aliments bien distincts
d'un point de vue nutritionnel

Bonne cohérence des groupes –
Proposer des substituts d'un autre groupe ✓

Nutriments indépendants
les uns des autres

Un substitut peut être proposé pour un nutriment
sans dégrader les apports des autres nutriments ✓

Doses servies informées

Pour éviter les biais générés par les achats
long terme (ex : kilo de sel, litre d'huile) ✓

70% des produits conservés,
soit 223 458 produits

Base de donnée très complète,
utilisable en France ou à l'étranger ✓

Merci pour votre attention !