



Concevez une application au service de la santé publique

Healthy Planet

Une application pour vous aider à
remplacer les viandes et poissons dans
votre alimentation au quotidien.





Scannez votre produit et découvrez des alternatives !



Cette application vous propose des produits avec des apports nutritionnels semblables sans viandes ni poissons.





Opérations de nettoyage du jeu de données

world.openfoodfacts.org

1 - Choix des variables

Codes, noms, marques, pays
Compositions (catégories et ingrédients)
Informations nutritionnelles

2519850
observations

2 - Suppression des doublons

Doublons de code barre : 29
Doublons de marque et produit : 540005

1979815
observations



Opérations de nettoyage du jeu de données

3 - Suppression des variables très peu remplies

Données manquantes : 87,6 %

Suppression d'une variable si taux de remplissage < 15 %

- 22 variables conservées

1979815
observations

4 - Regroupement des colonnes

Tentative de remplissage en croisant les variables

Traduction en français de certaines valeurs

Nettoyage de la variable 'countries'

Choix des variables sur le Nutri-Score

1932601
observations



Opérations de nettoyage du jeu de données

5 - Suppression des valeurs aberrantes

Remplacement par np.nan :

- des valeurs < 0
- des valeurs d'énergie > 12000 J
- des valeurs > 100 pour les autres données quantitatives
- des valeurs pour lesquelles :
 - `saturated_fat` $>$ `fat`
 - `sugars` $>$ `carbohydrates`

Suppression des observations avec 6 données manquantes ou plus



Opérations de nettoyage du jeu de données

6 - Imputation des valeurs manquantes

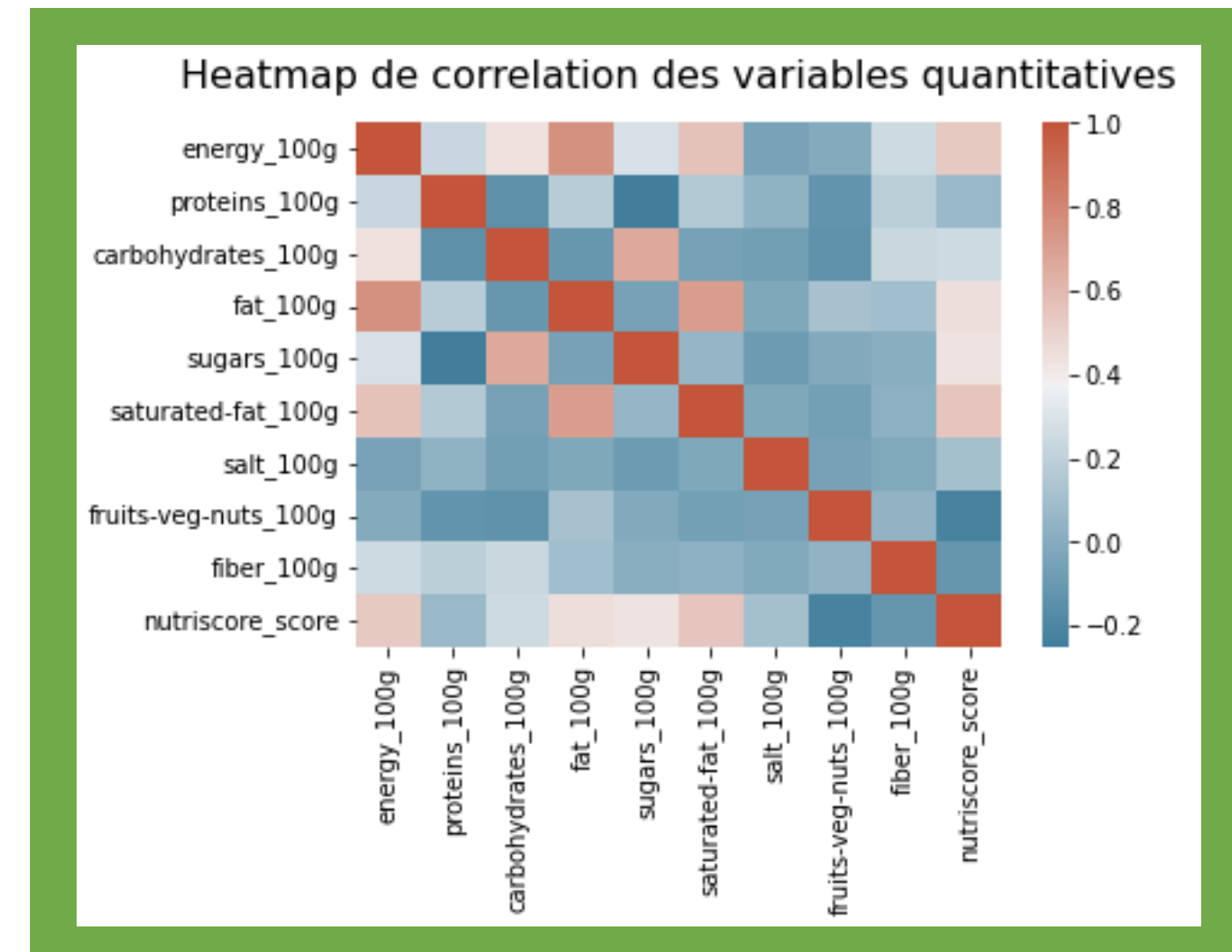
1529573
observations

Remplacement des manquants de chaque catégorie (top 500) par la médiane de chaque variable pour cette même catégorie

Remplacement par des zéros pour les variables 'salt', 'fiber' et 'fruits-veg-nuts'

Utilisation de IterativeImputer pour les variables hautement corrélées et comparaison avec KNNImputer

Calcul des nouveaux Nutri-Scores





Opérations de nettoyage du jeu de données

7 - Préparation du jeu de donnée pour l'analyse

Sélection des produits vendus en France

556360
observations

Suppression des produits sans code barre

551337
observations

Suppression des observations non catégorisées

280035
observations

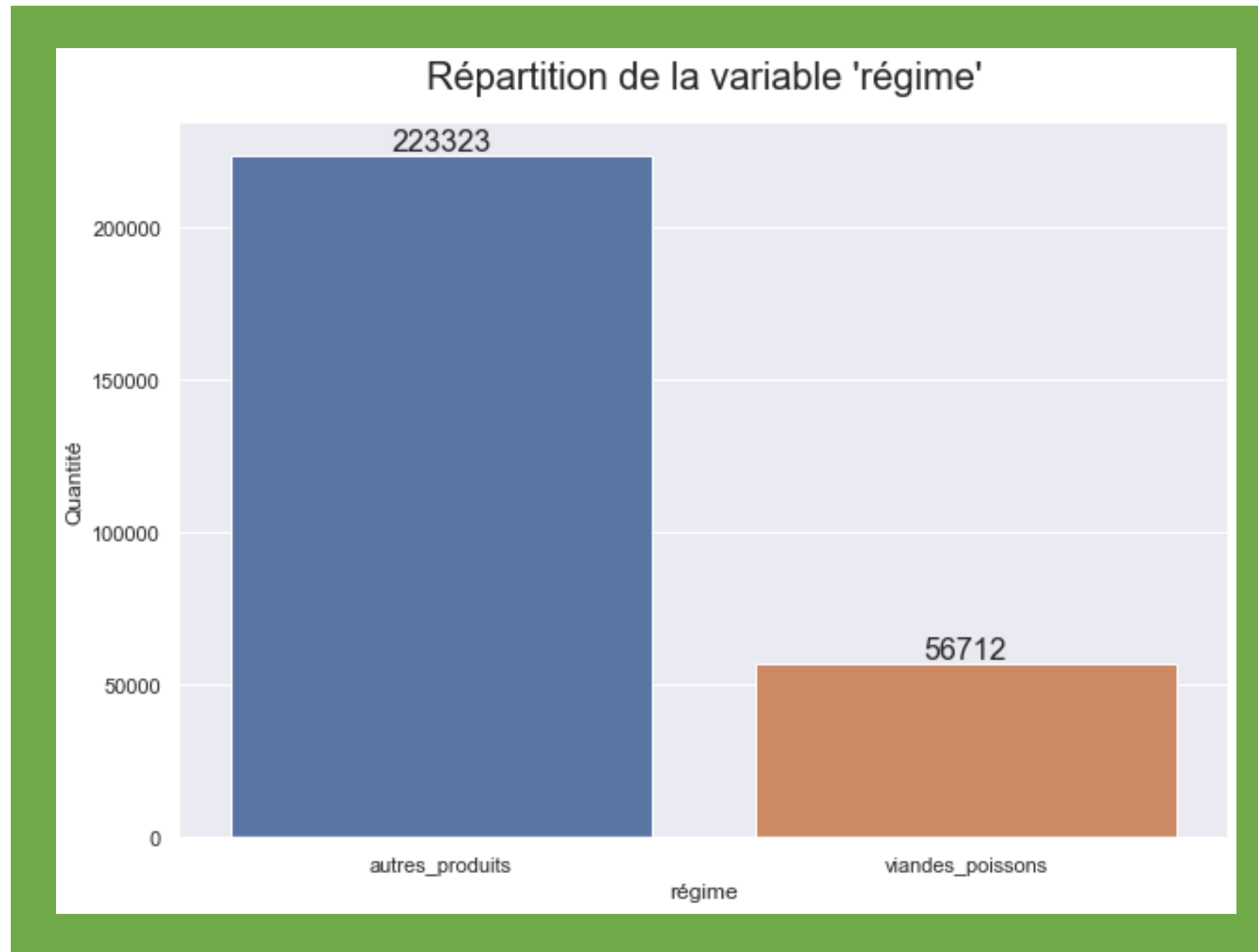
Création de la variable "régime" et remplissage par :

- 'viandes_poissons' à partir de certains mots-clés
- 'autres_produits'

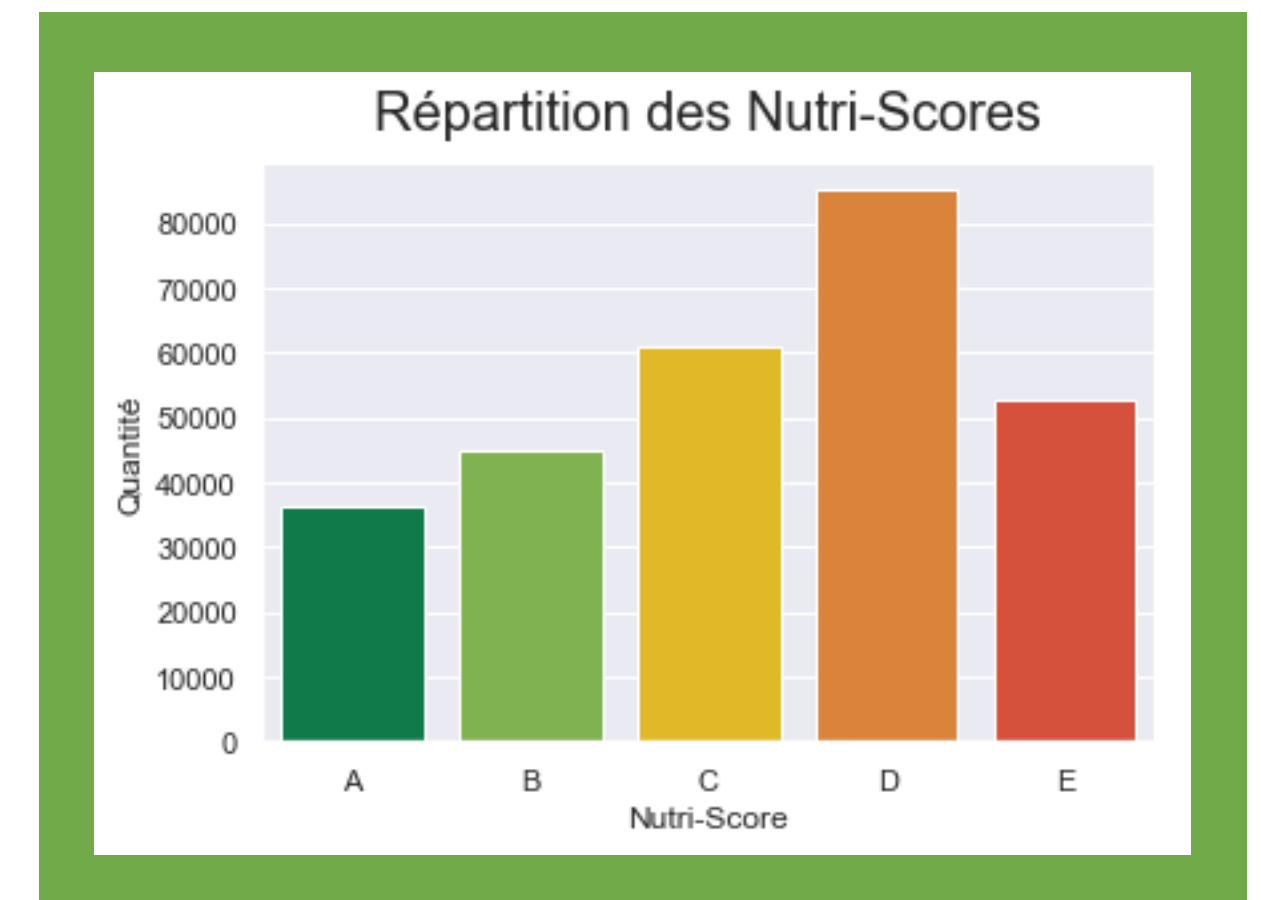
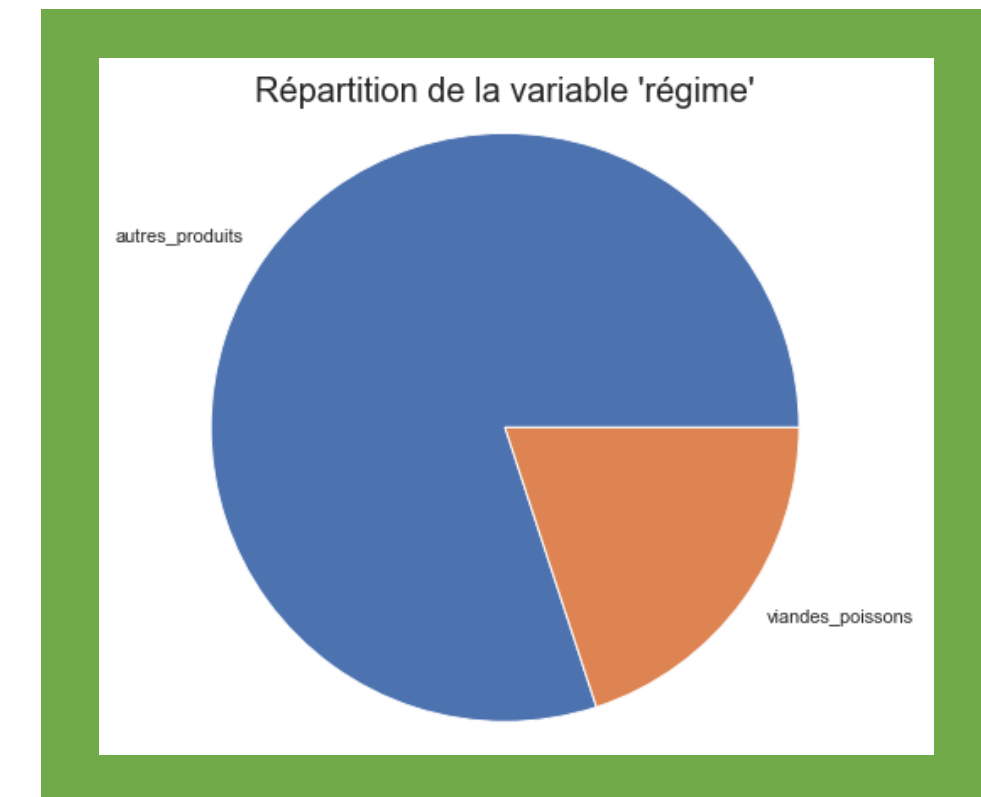
280035
observations



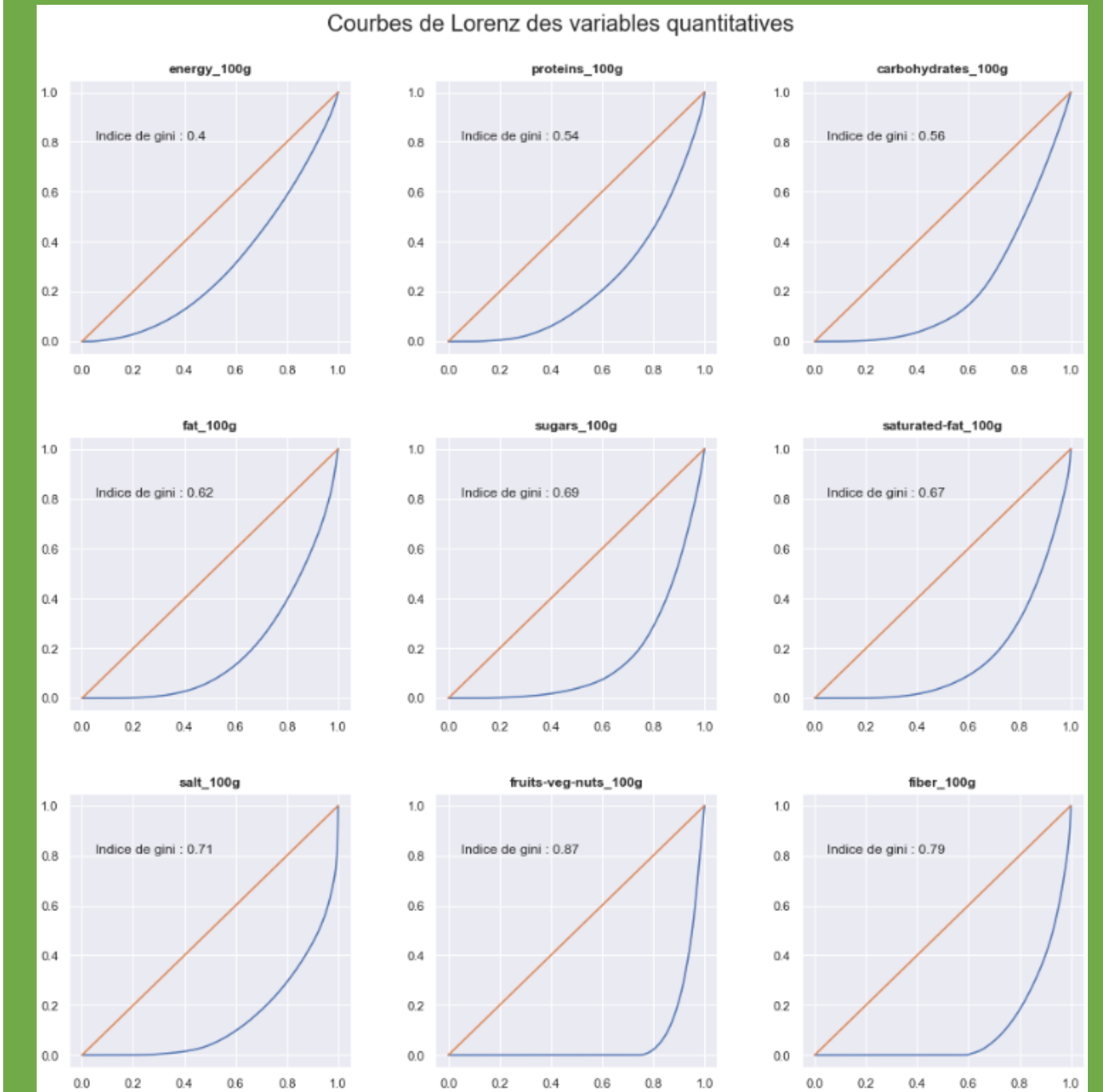
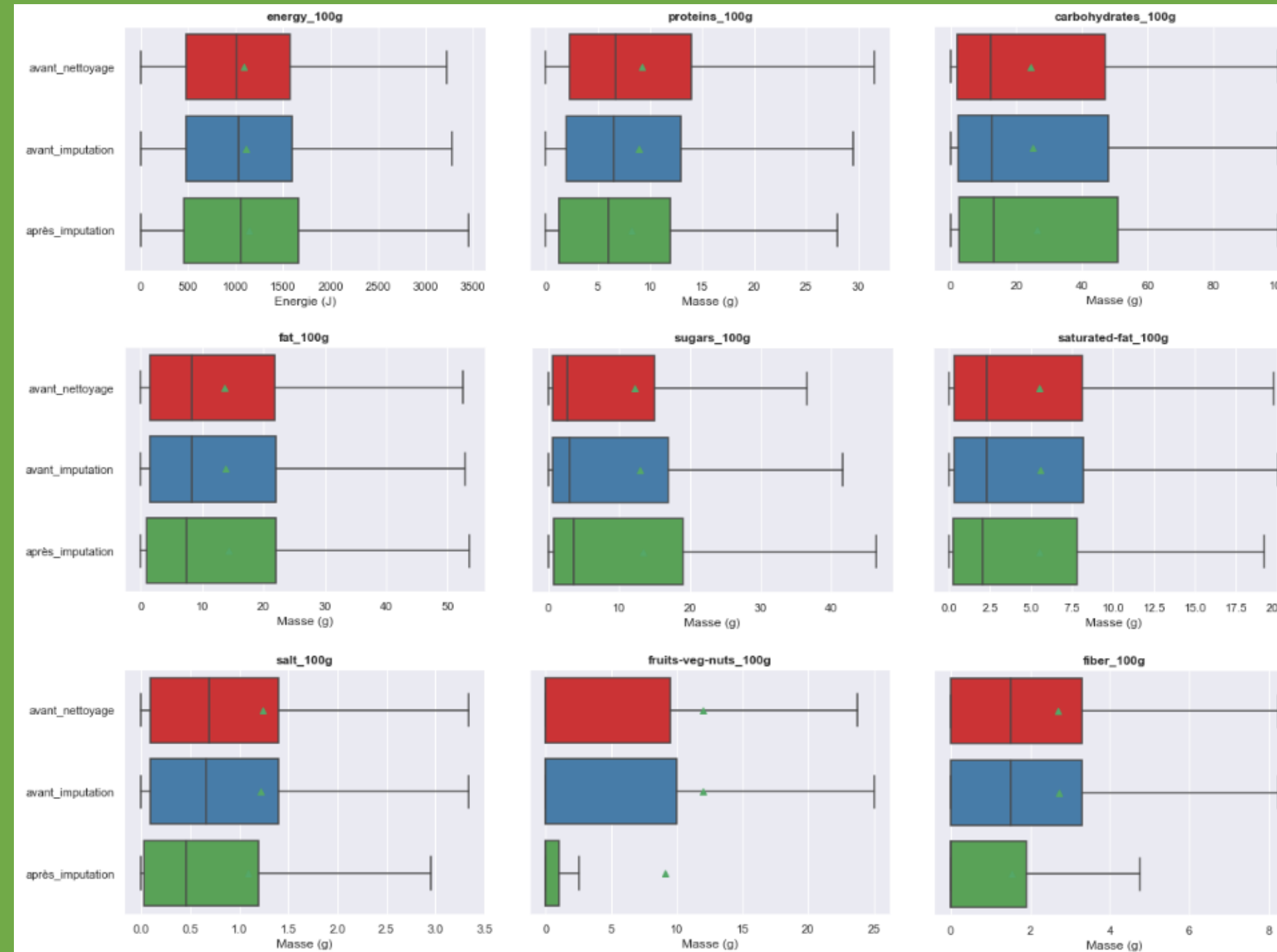
Analyse univariée



280035 observations

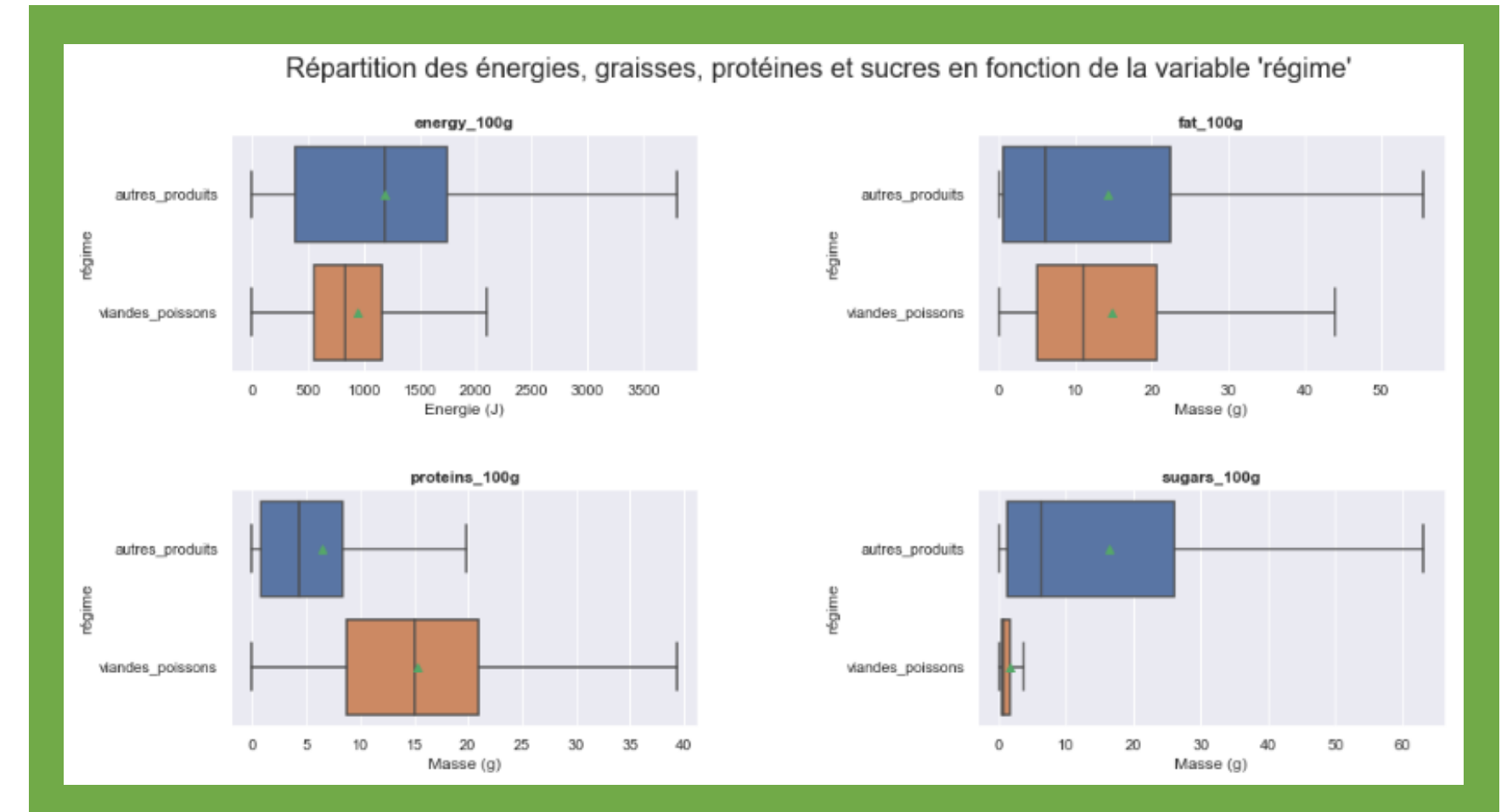
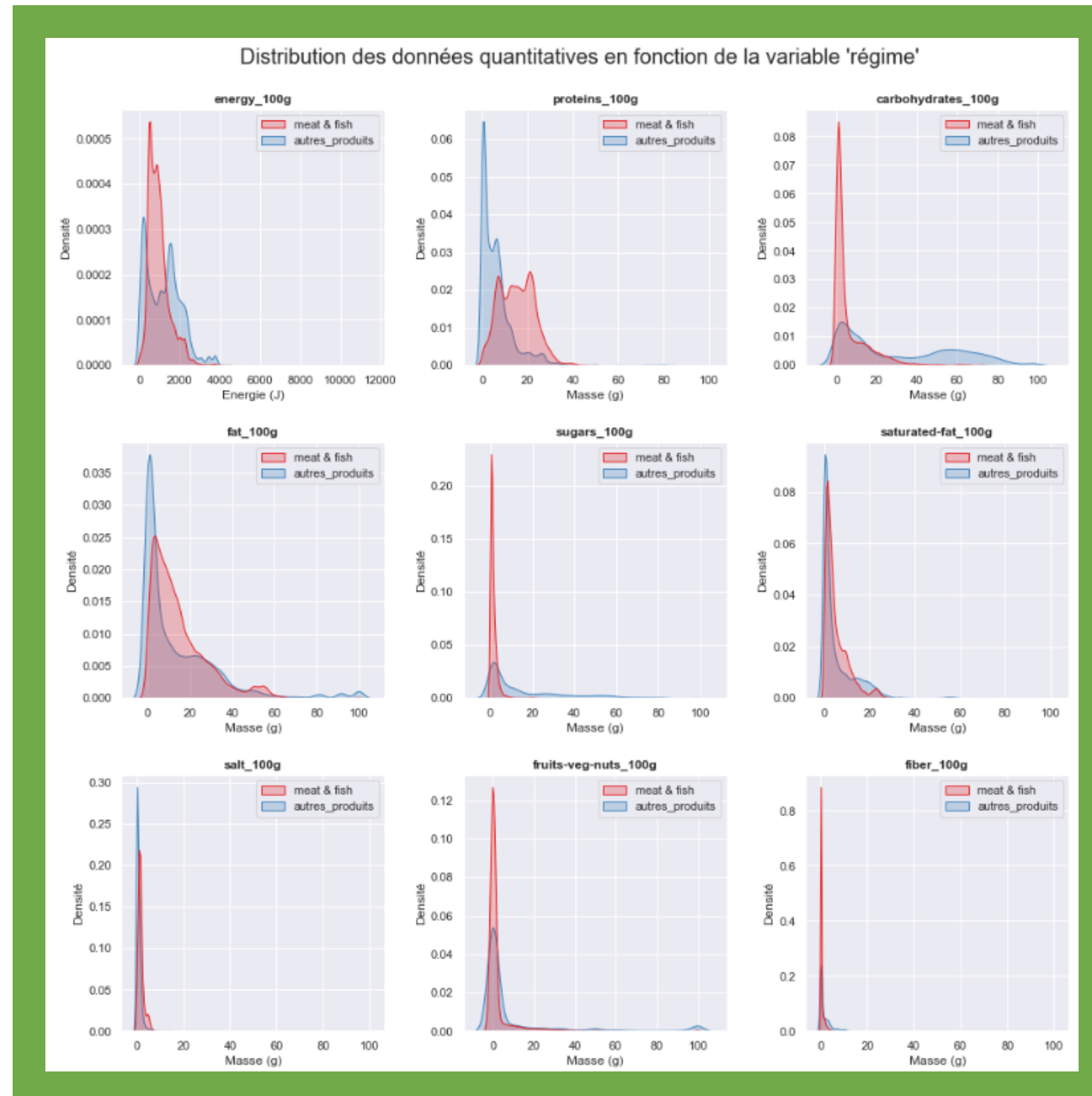


Analyse univariée





Analyse bivariable

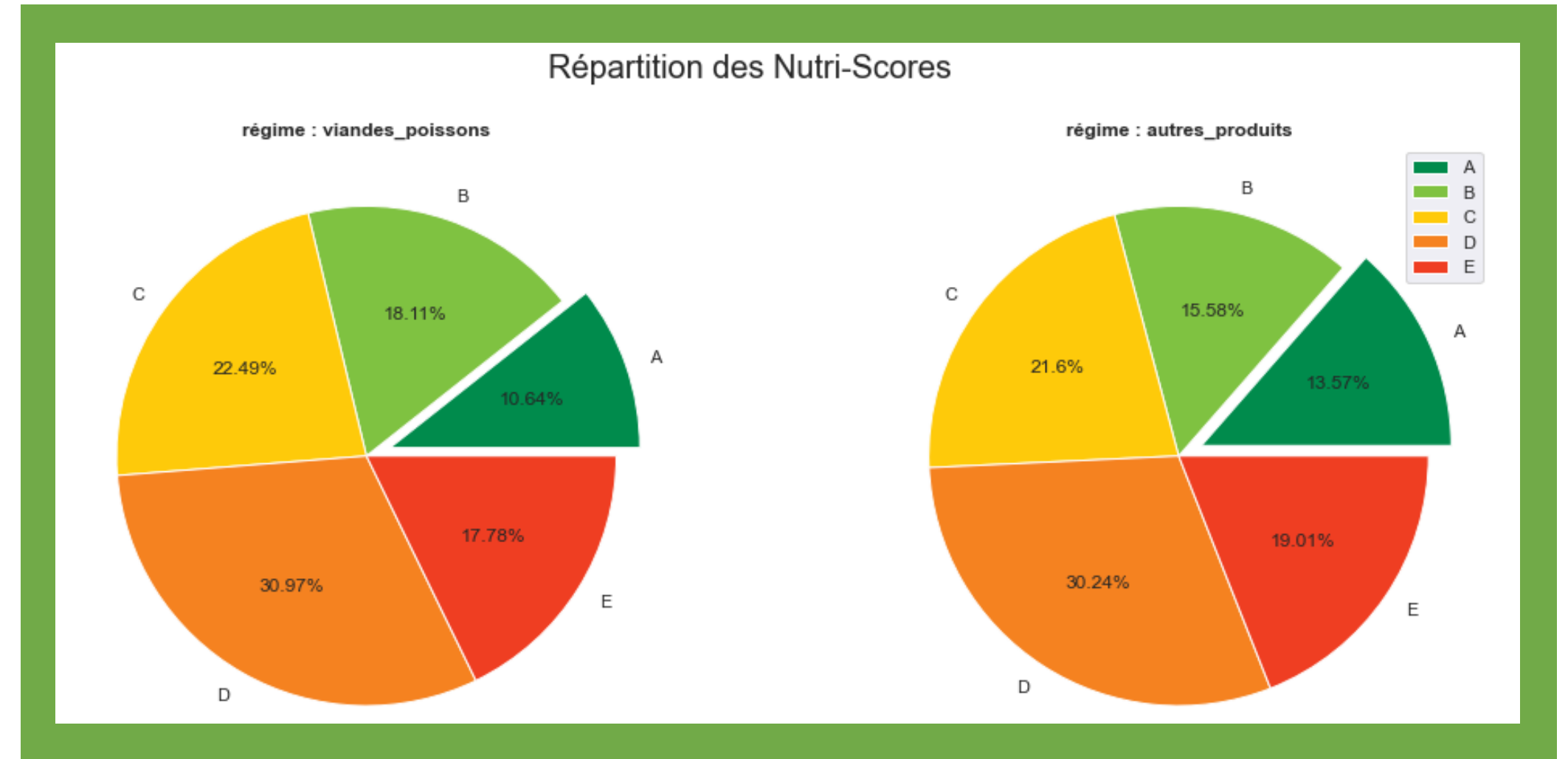
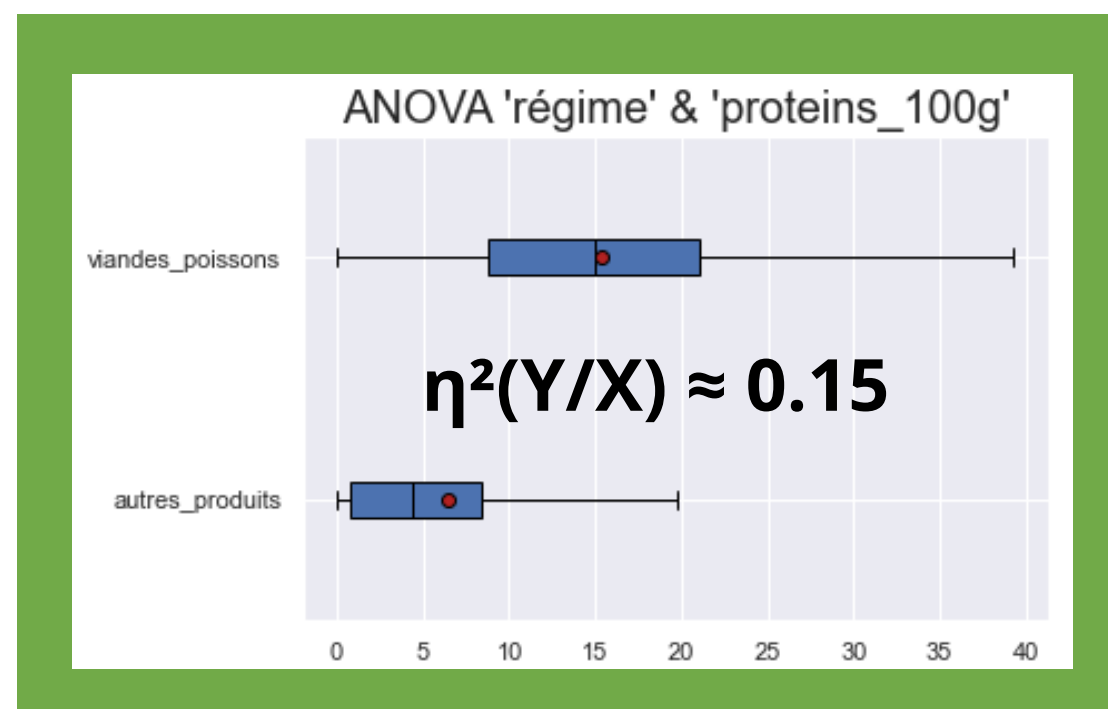
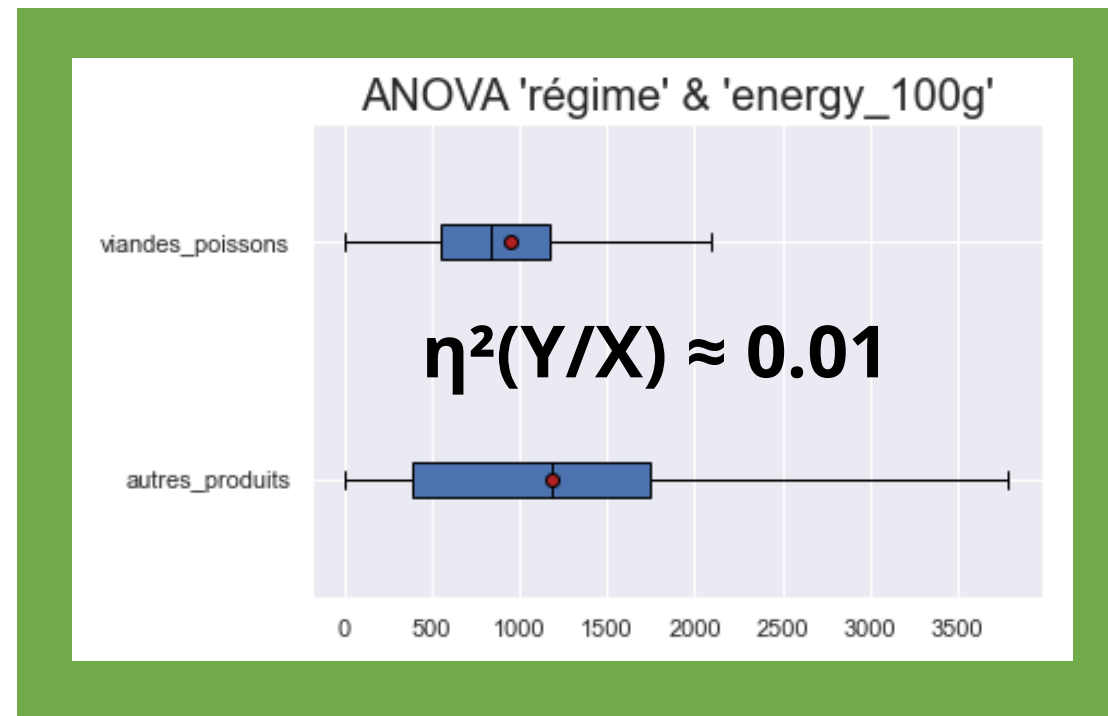


Produits "viandes_poissons" :

- plus gras
- moins sucrés
- plus riches en protéines
- autant énergétiques ?



Analyse bivariable



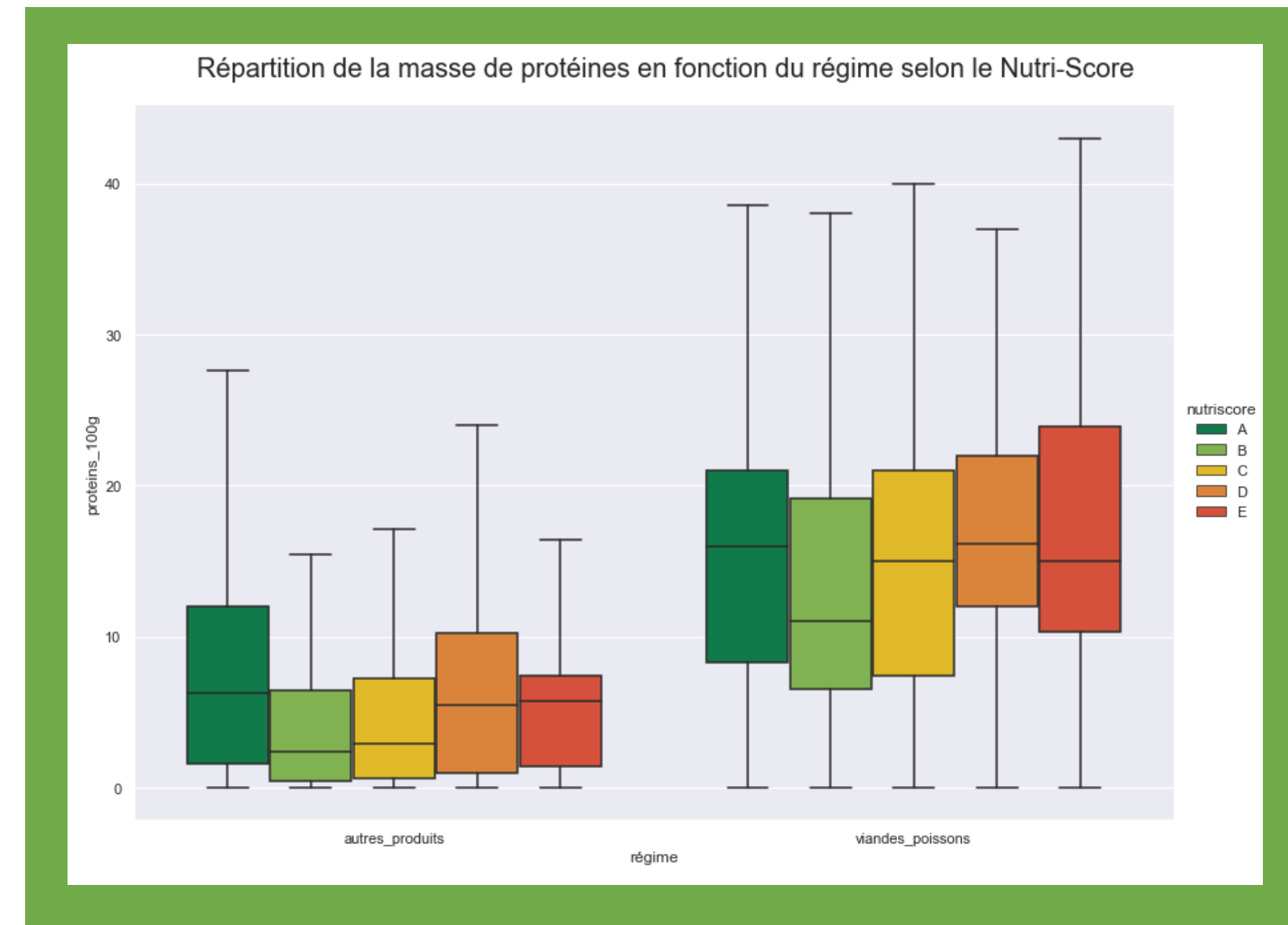
Distribution des Nutri-Scores analogue



Analyse multivariée



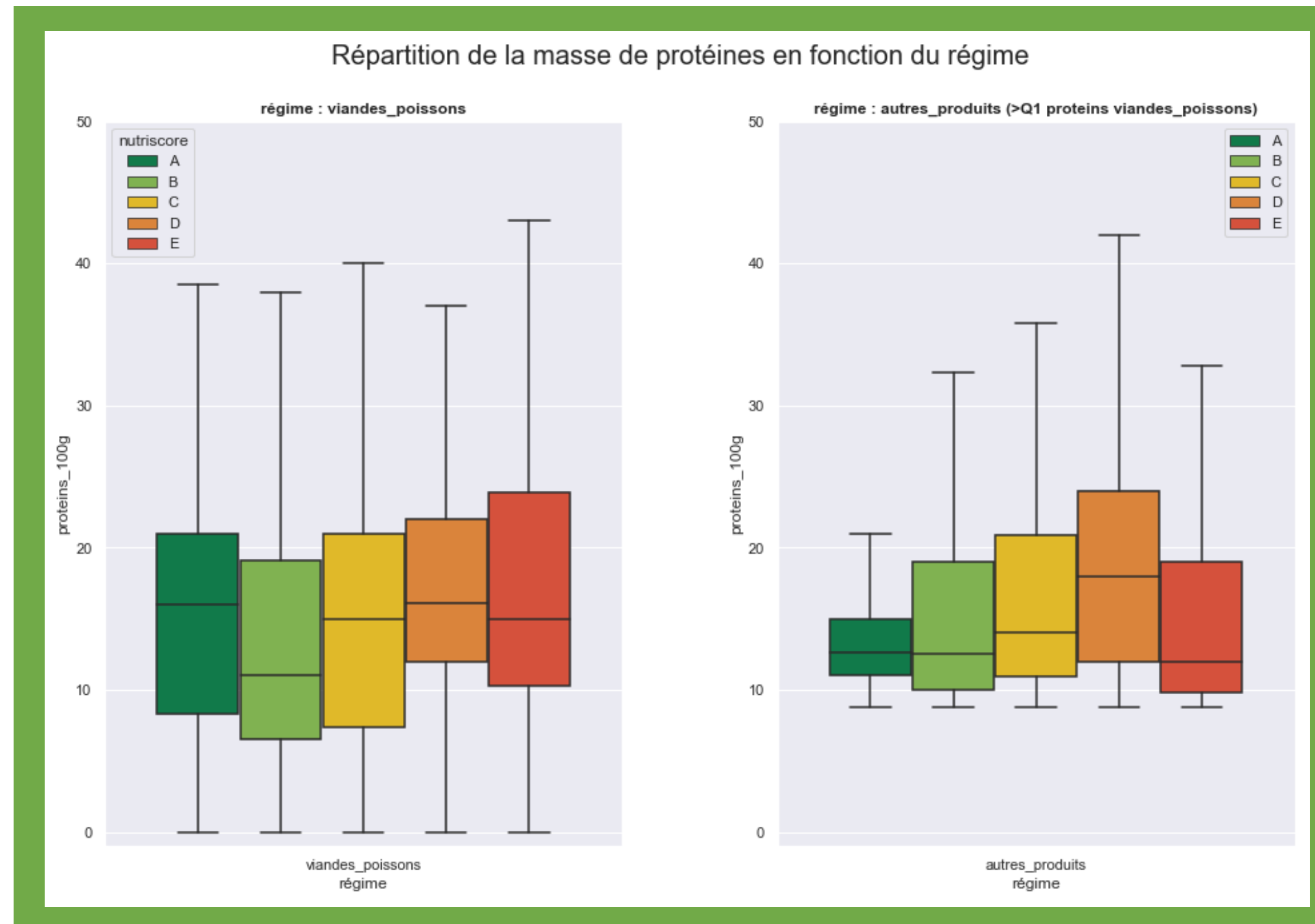
Les produits les plus énergétiques
ont un Nutri-Score élevé.



La quantité de protéines
n'influence pas le Nutri-Score.



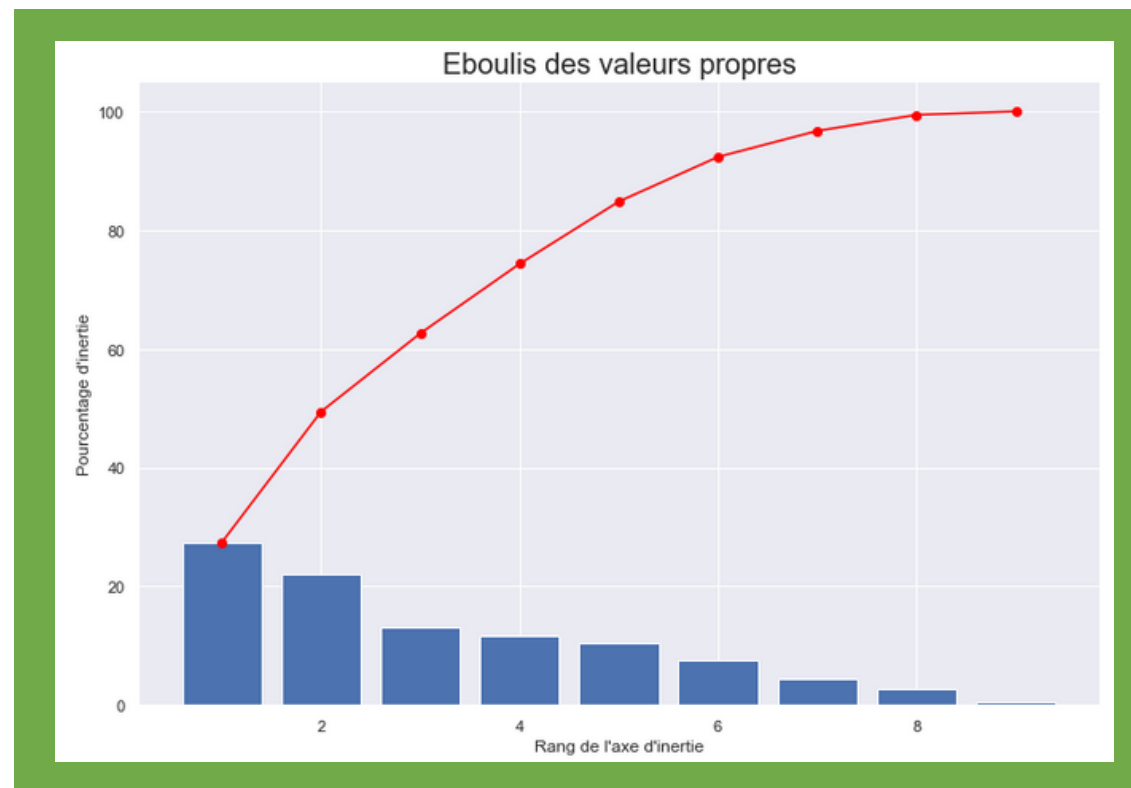
Analyse multivariée



Possibilité de trouver des produits autant protéinés dans ceux catégorisés "autres_produits"



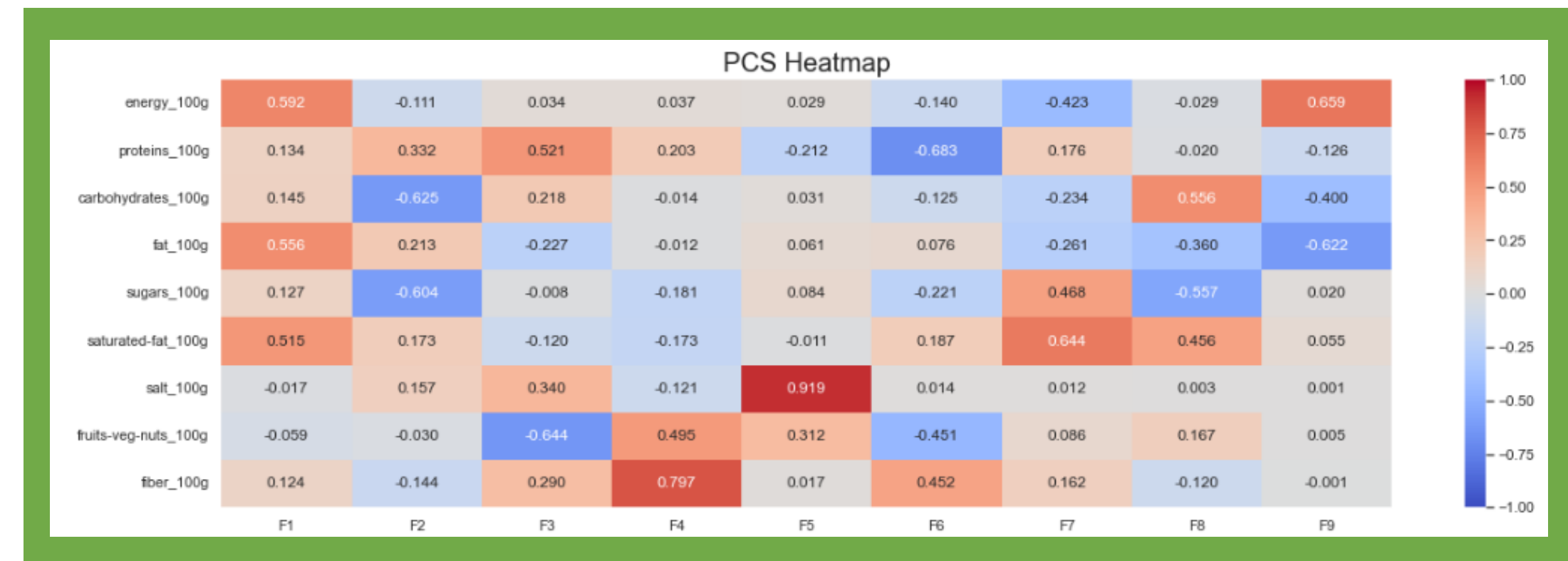
Réductions dimensionnelles (ACP)



Variable illustrative : "régime"

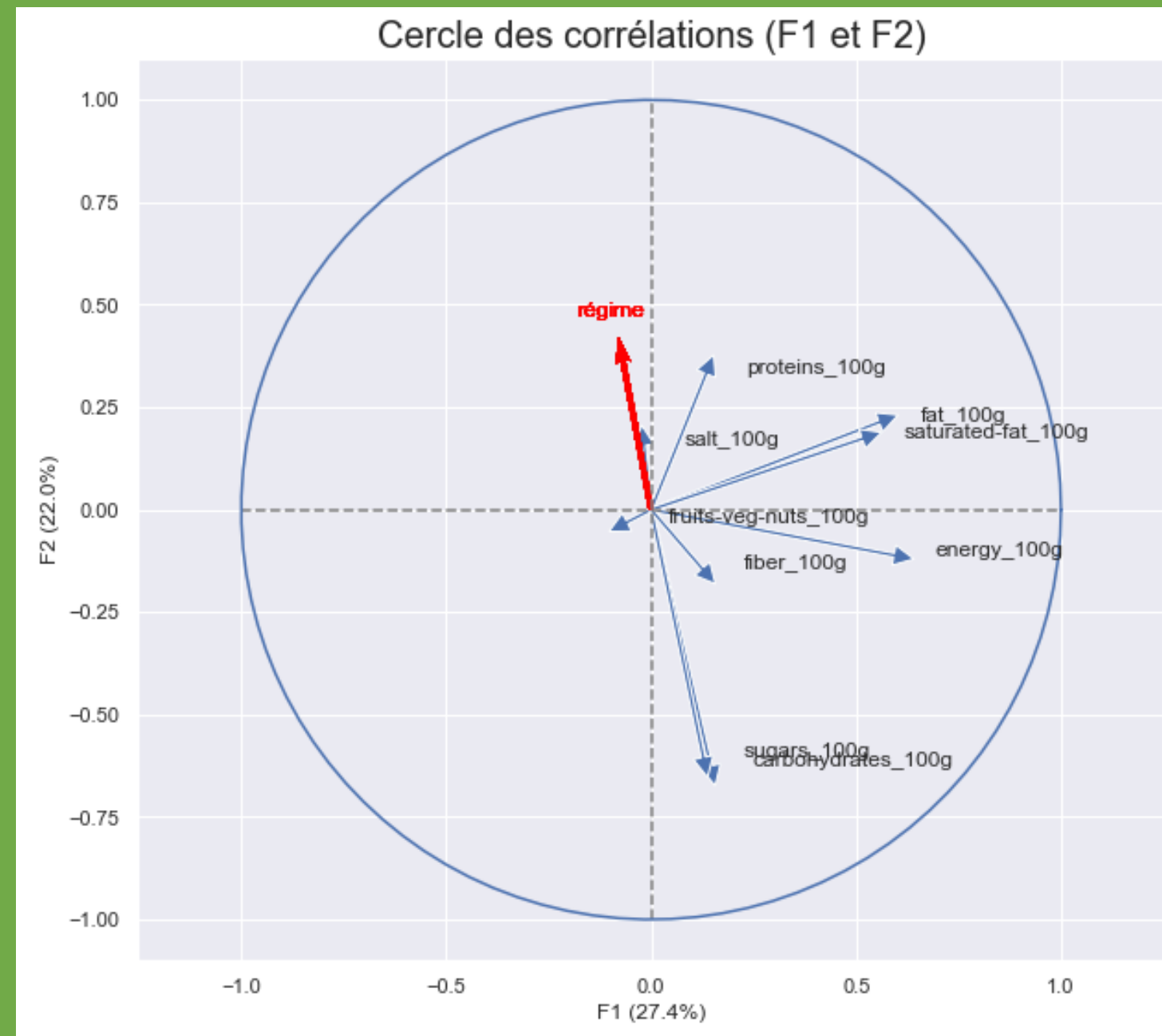
Inertie couverte par les 2 premières composantes : 49,4 %

Inertie couverte par les 4 premières composantes : 74,3 %

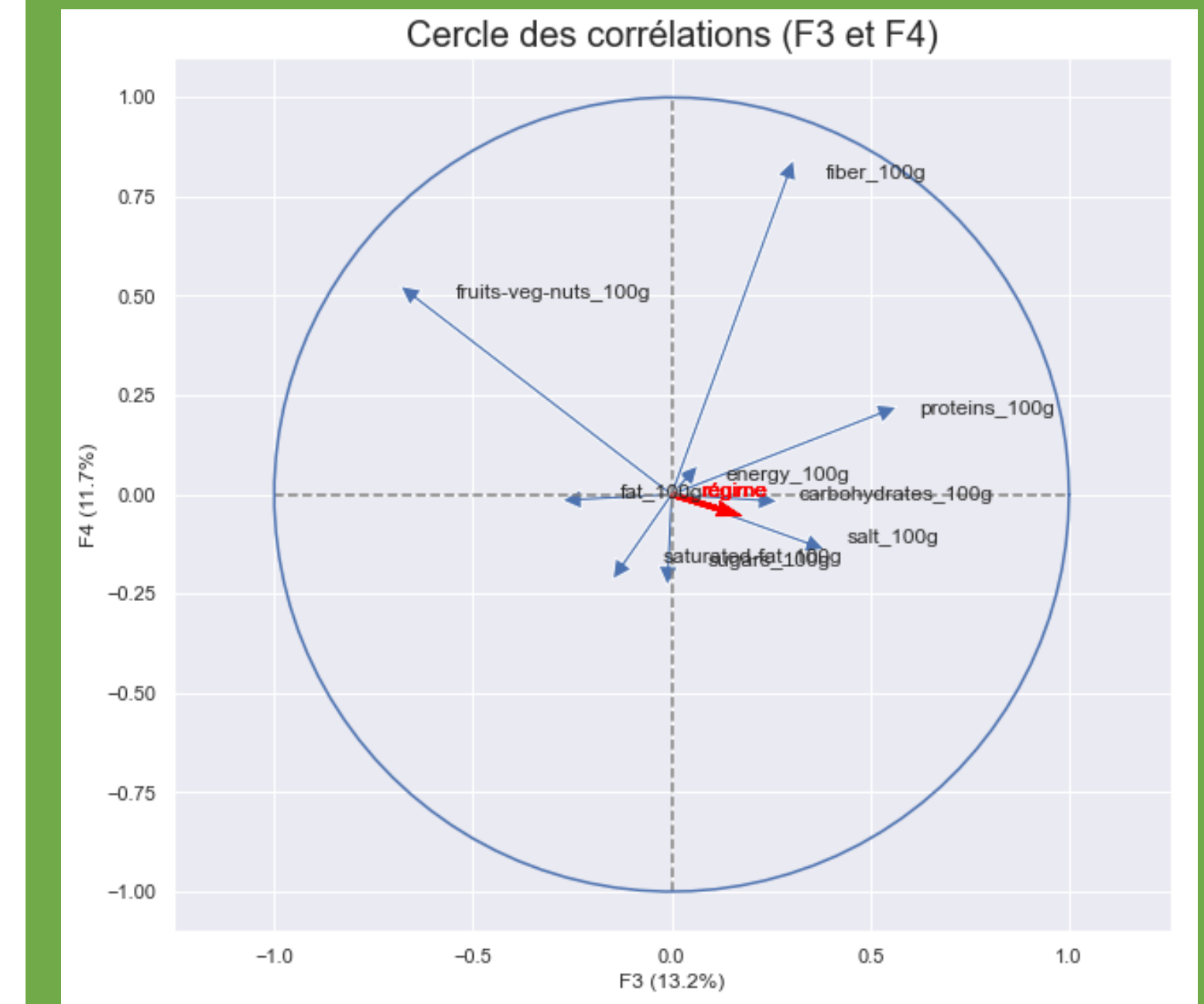




Réductions dimensionnelles (ACP)



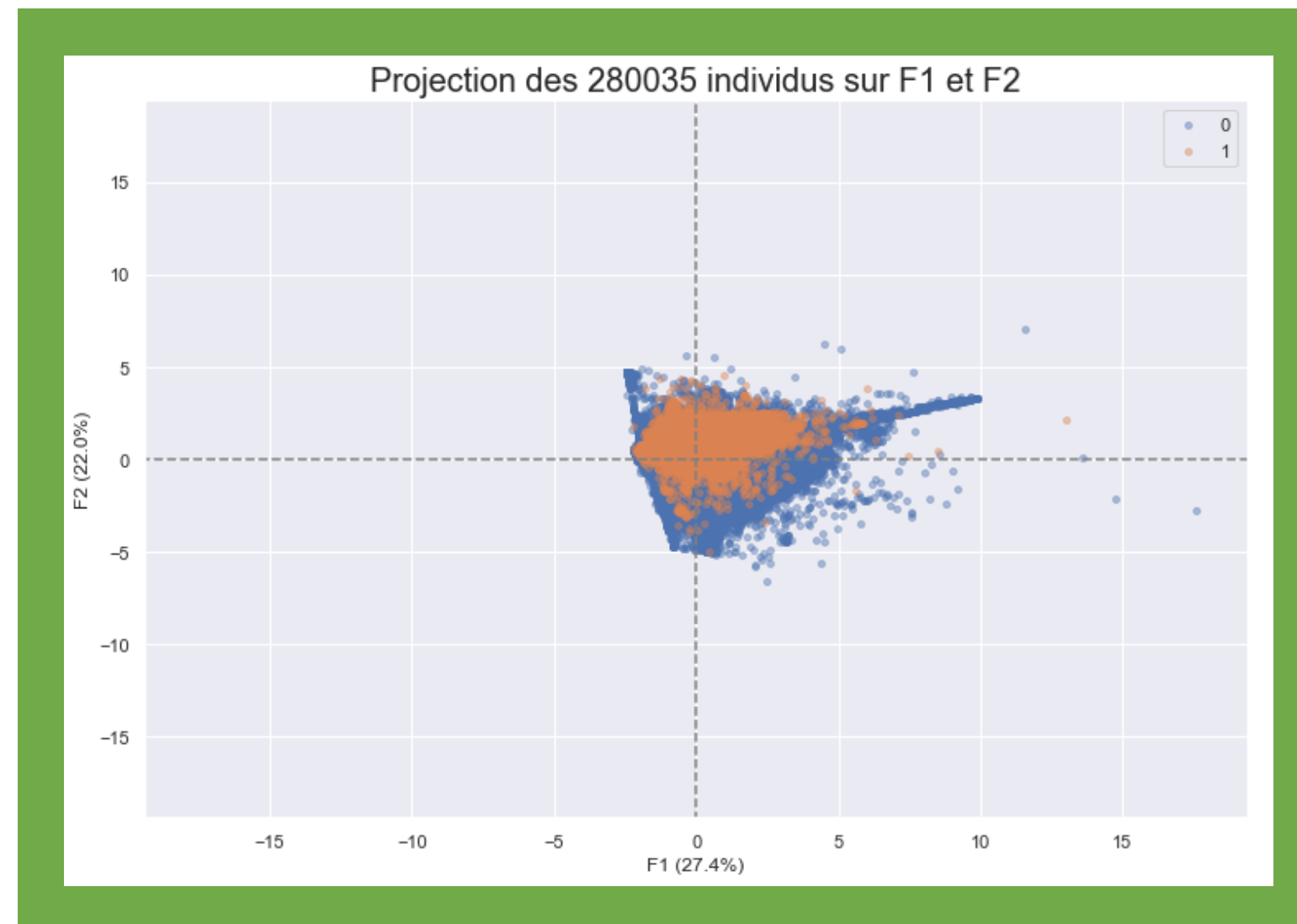
F1 : facteur énergétique
F2 : caractère sucré / salé



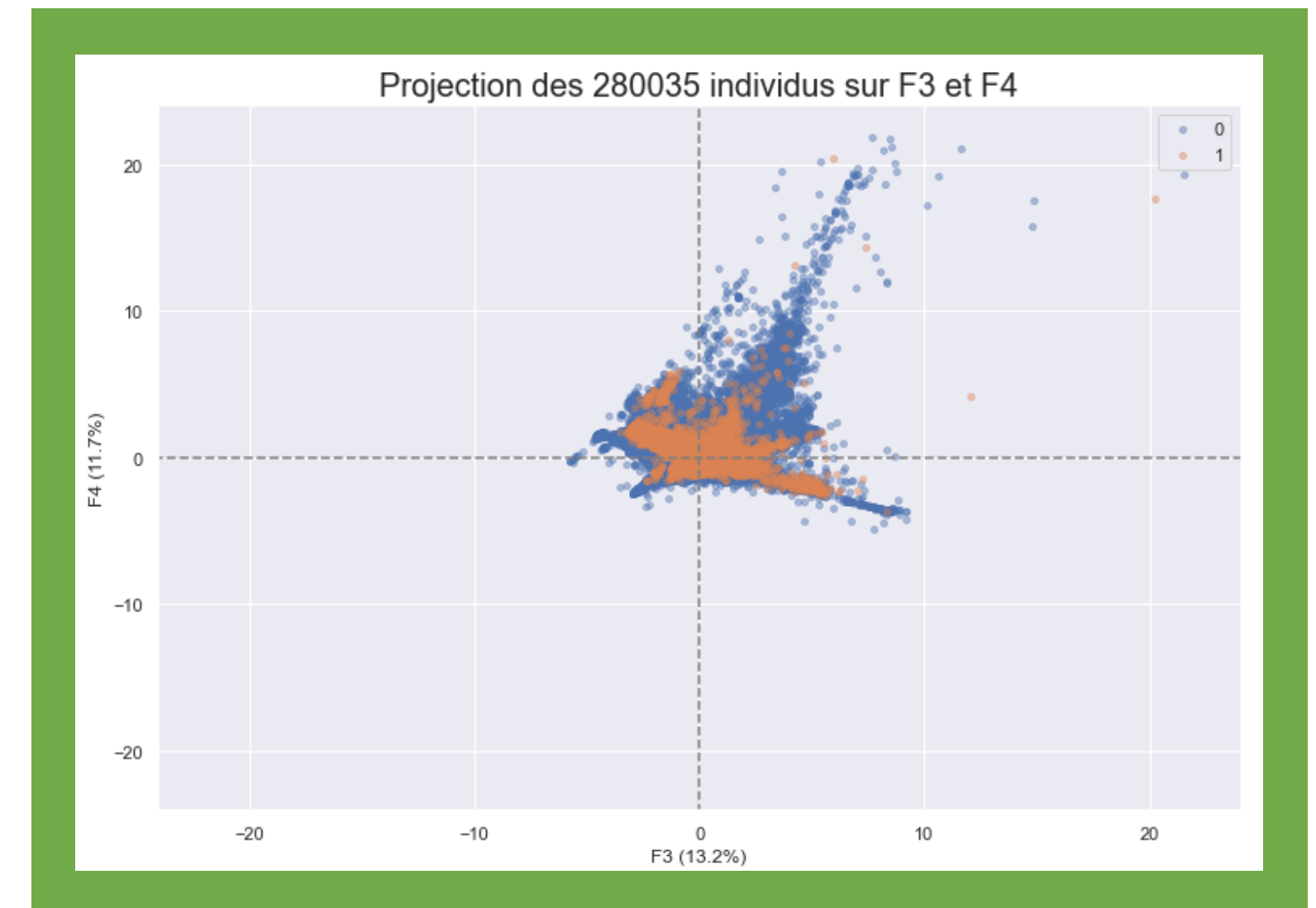
F3 : caractère gras / non-gras
F4 : caractère légumineux



Réductions dimensionnelles (ACP)



Projections F1/F2 similaires entre les 2 catégories

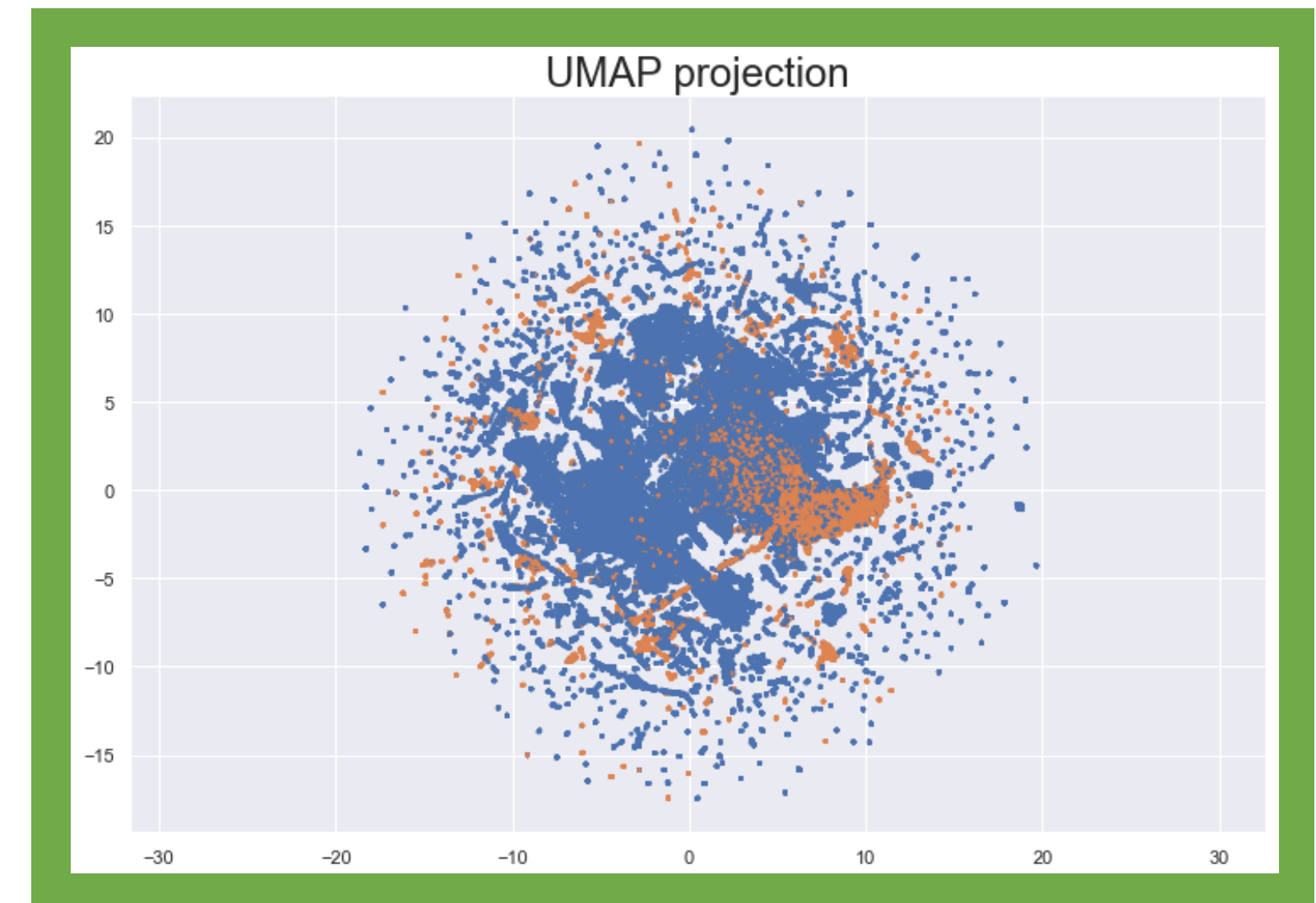
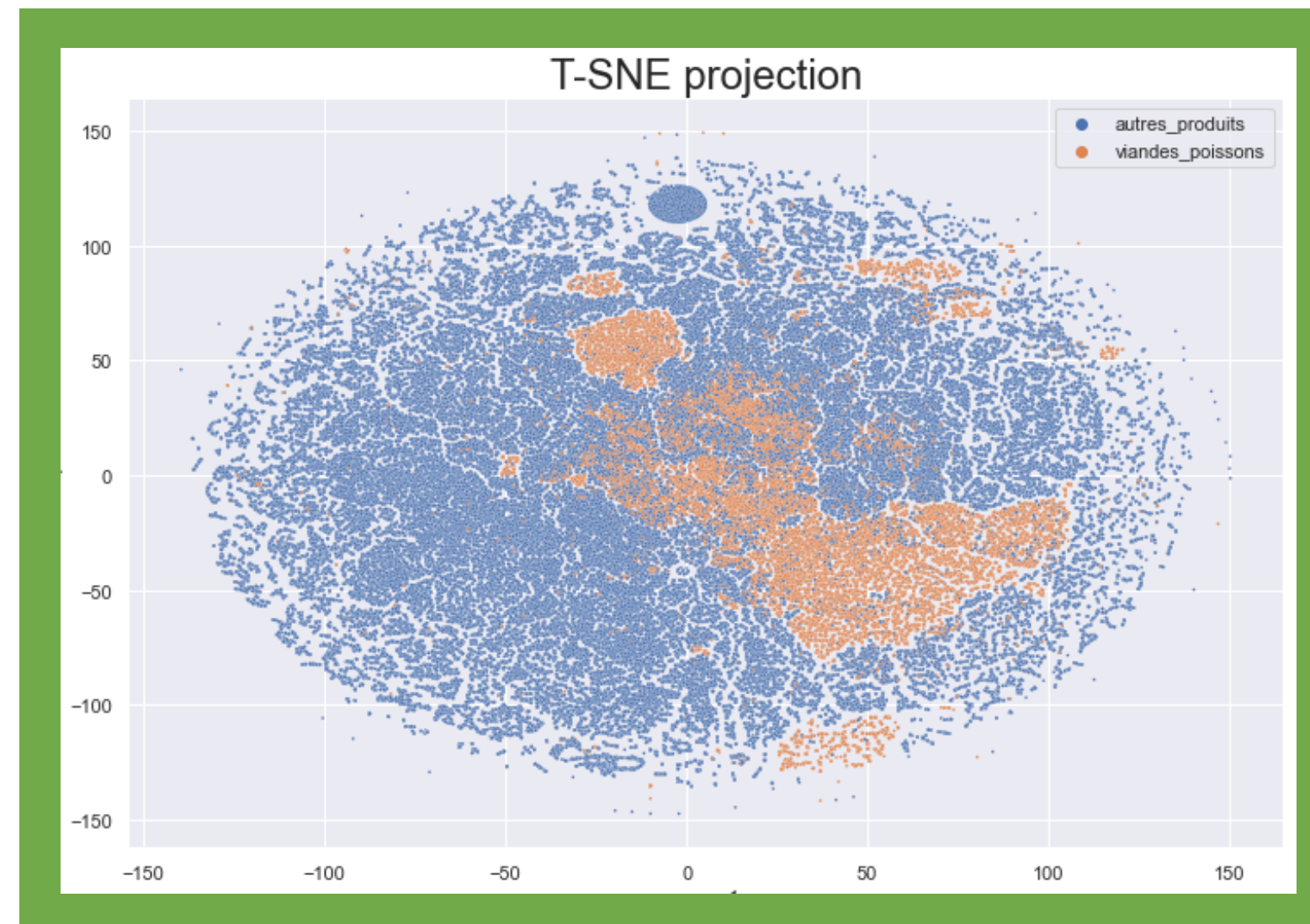


Projections F3/F4 similaires entre les 2 catégories

Les produits avec viandes ou poissons sont bien intégrés avec les autres produits.



Réductions dimensionnelles (autres méthodes)



Fortes similarités entre ces deux projections



Conclusions sur l'Analyse Exploratoire

Les produits avec viandes ou poissons sont :

- plus gras
- moins sucrés
- plus riches en protéines

Les deux catégories de produits sont autant énergétiques.

Il est effectivement possible de trouver des produits sans viandes ou poissons qui ont des apports nutritionnels semblables aux produits qui en ont.



Démonstration de l'application

1 - Des propositions parfois pertinentes ...



NearestNeighbors

code	categories	energy_100g	proteins_100g	carbohydrates_100g	fat_100g	sugars_100g	saturated-fat_100g	salt_100g	fruits-veg-nuts_100g	fiber_100g	nutriscore
3300842240372	Salades piémontaises au jambon	452.0	3.5	11.0	5.5	1.1	0.6	0.7	0.0	0.0	B

code	categories	energy_100g	proteins_100g	carbohydrates_100g	fat_100g	sugars_100g	saturated-fat_100g	salt_100g	fruits-veg-nuts_100g	fiber_100g	nutriscore
3271514167122	Petits épeautres	477.0	2.8	12.00	5.40	2.80	0.8	0.78	0.0	0.0	C
5400119512565	Plats à base de quinoa	460.0	4.7	11.00	4.70	1.00	0.6	0.57	0.0	0.0	B
3083681047929	Poelees rustiques	397.0	3.0	8.50	4.50	1.80	0.9	0.70	0.0	0.0	C
8858135071011	Sauces au curry	385.0	3.0	9.00	5.00	2.00	1.0	1.00	0.0	0.0	C
3770013958342	Onigiri	427.0	3.1	11.20	4.38	0.21	1.4	0.33	0.0	0.0	B
3482970001213	Légumineuses	389.0	3.8	7.75	4.20	0.30	1.3	0.72	0.0	0.0	C
7610800987628	Röstis de pommes de terre surgelés	439.0	2.0	12.00	5.00	0.00	0.5	0.90	0.0	0.0	C
3321950010114	Haricots blancs cuisines	423.0	4.5	11.60	2.90	1.10	1.0	0.71	0.0	0.0	B





Démonstration de l'application

2 - Et d'autres fois inadaptées ...



NearestNeighbors

code	categories	energy_100g	proteins_100g	carbohydrates_100g	fat_100g	sugars_100g	saturated-fat_100g	salt_100g	fruits-veg-nuts_100g	fiber_100g	nutriscore
3104544070936	Foie gras de canard entier du sud ouest	2209.0	7.61	2.7	54.0	0.9	23.0	1.175	0.0	0.0	E

code	categories	energy_100g	proteins_100g	carbohydrates_100g	fat_100g	sugars_100g	saturated-fat_100g	salt_100g	fruits-veg-nuts_100g	fiber_100g	nutriscore
056489278535	Ard	2131.0	5.9	1.9	54.0	0.80	24.0	1.10	0.0	0.0	E
20197070	Produits gastronomiques	2418.0	6.4	2.2	56.1	1.72	23.6	1.25	0.0	0.0	E
760067667192	Igp origine landes	2033.0	9.2	4.4	48.2	0.90	20.0	1.46	0.0	0.0	E
760288360087	Chocolats au lait édulcorés	2042.0	10.0	6.0	47.0	0.00	23.0	0.00	0.0	0.0	D
410721461257	Oignons frits	2531.0	6.0	0.0	45.0	0.00	20.0	1.30	0.0	0.0	D
000915103328	Fromages à tartiner	1916.0	9.9	2.0	46.0	0.50	19.0	1.60	0.0	0.0	D
261081146217	Plats préparés frais	2059.0	6.2	2.4	41.0	1.90	26.0	1.50	0.0	0.0	E
711200457298	Sauces béarnaises	1928.0	7.5	1.9	50.0	1.90	17.0	0.81	0.0	0.5	D





Démonstration de l'application

3 - Faisabilité de l'application

Possibilité d'ajout du traitement du Nutri-Score

Variable indiquant le lieu de disponibilité des produits

Difficulté de catégoriser les produits avec la variable "régime"



Merci pour votre écoute !