

Concevez une application au service de la santé publique

Soutenance du 28 avril 2023

Yann Pham-Van

Concevez une application au service de la santé publique

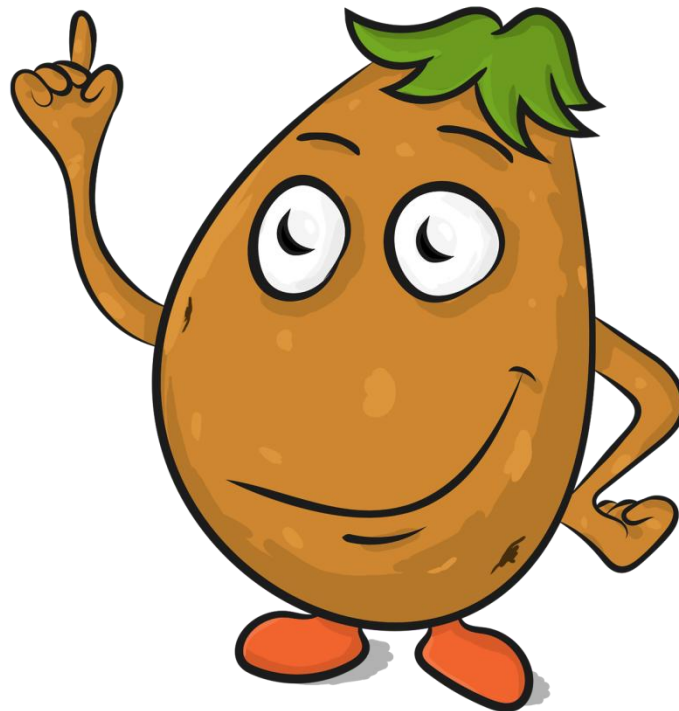
- Appel à projets
- Idée d'application
- Nettoyage des données
- Analyse exploratoire
- Synthèse des conclusions

Concevez une application au service de la santé publique

- **Appel à projets**
- Idée d'application
- Nettoyage des données
- Analyse exploratoire
- Synthèse des conclusions

Appel à projets > Santé publique France

« Trouver des idées innovantes d'applications en
lien avec l'alimentation »



Appel à projets > Open Food Facts

Produits alimentaires



Base de données



NUTRI-SCORE



Appel à projets > Nutri-Score

Composantes négatives

- Energie
- Graisses saturées
- Sucres
- Sodium



Composantes positives

- Fruits & légumes
- Fibres
- Protéines



NUTRI-SCORE



Concevez une application au service de la santé publique

- Appel à projets
- **Idée d'application**
- Nettoyage des données
- Analyse exploratoire
- Synthèse des conclusions

Idée d'application

Composantes négatives

- Energie
- Graisses saturées
- Sucres
- Sodium



Composantes positives

- Fruits & légumes
- Fibres
- Protéines



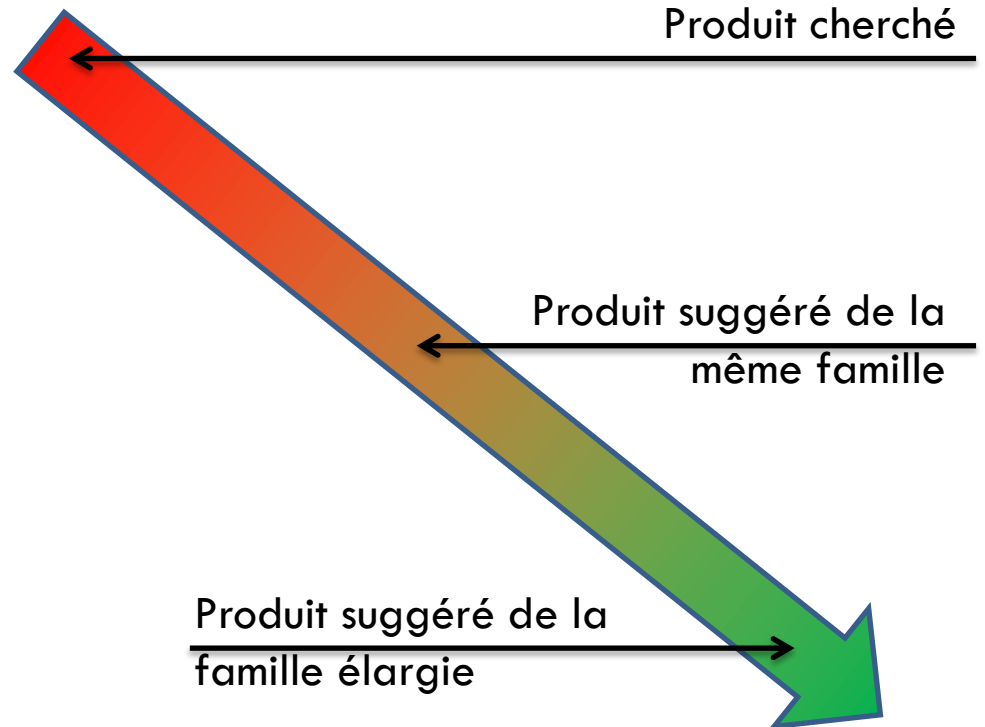
NUTRI-SCORE



Idée d'application

Composantes négatives

- Energie
- Graisses saturées
- Sucres
- Sodium



TOPIZ'LESS

Concevez une application au service de la santé publique

- Appel à projets
- Idée d'application
- **Nettoyage des données**
- Analyse exploratoire
- Synthèse des conclusions

Nettoyage des données > inspection

- 320 772 observations
- 162 variables

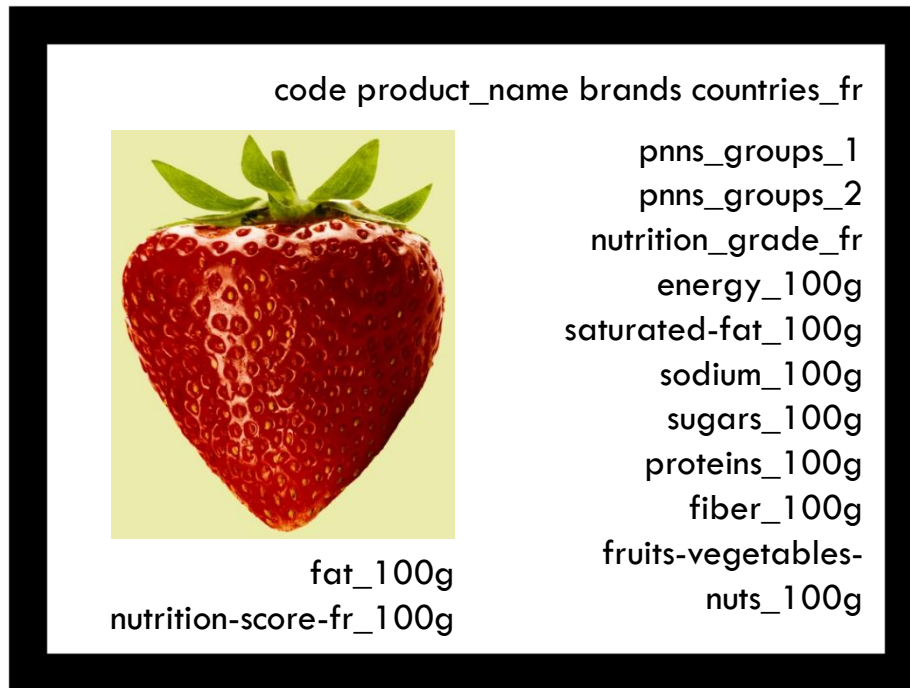
Totalement vides



Remplies à 50 % minimum



Nettoyage des données > sélection des variables



Besoins généraux pour les produits

- Identifier
- Catégoriser
- Caractériser

Besoins impératifs pour l'application

- Nom
- Famille
- Sous-famille

(Subsistent **27.6** % des observations initiales)

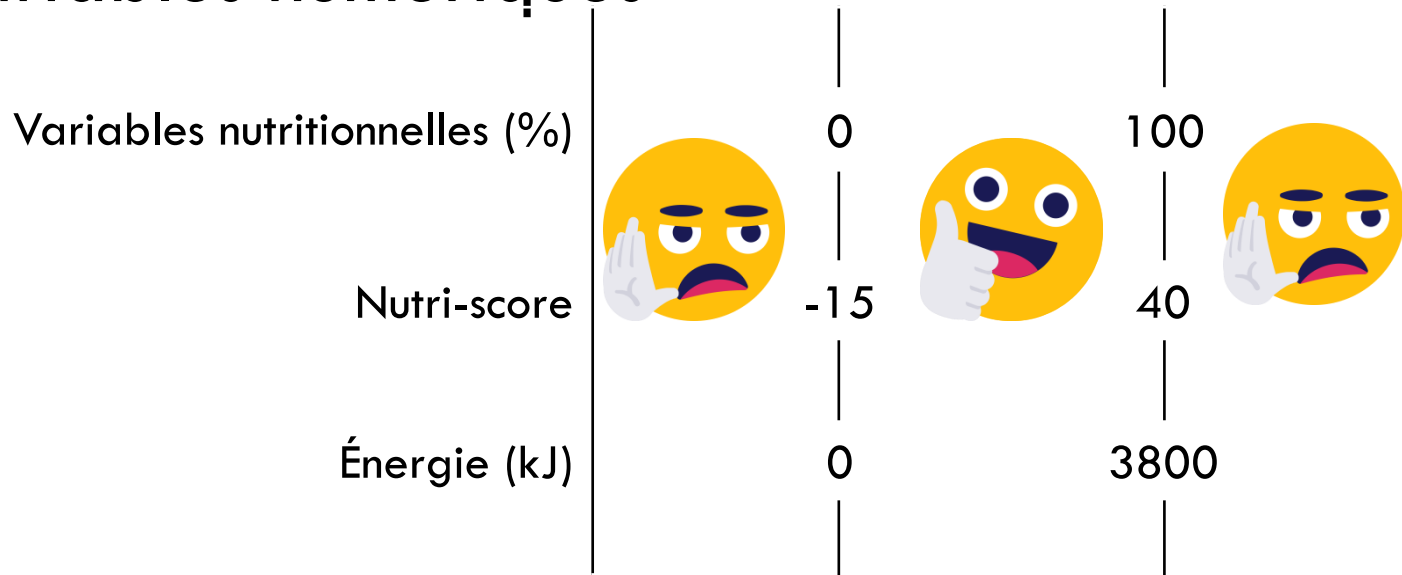
Nettoyage des données > filtrage des observations

- Filtrage sur variables qualitatives impératives
 - product_name
 - pnns_groups_1
 - pnns_groups_2
- Nettoyage des modalités
 - pnns_groups_1
 - pnns_groups_2

(Subsistent **21.3** % des observations initiales)

Nettoyage des données > valeurs aberrantes

- Variables numériques



- Nutrition_grade_fr

A | B | C | D | E | NaN

(Subsistent **21.3** % des observations initiales)

Nettoyage des données > valeurs manquantes

- Variables quantitatives
 - Remplacement par la médiane de sous-famille
 - Sauf *Potatoes & Salty and fatty products* pour *fruits-vegetables-nuts_100g* → 0
- Variables qualitatives
 - *Countries_fr* → *Inconnu*
 - *Brands* → *product_name*
 - *Nutrition_grade_fr* → calcul exact

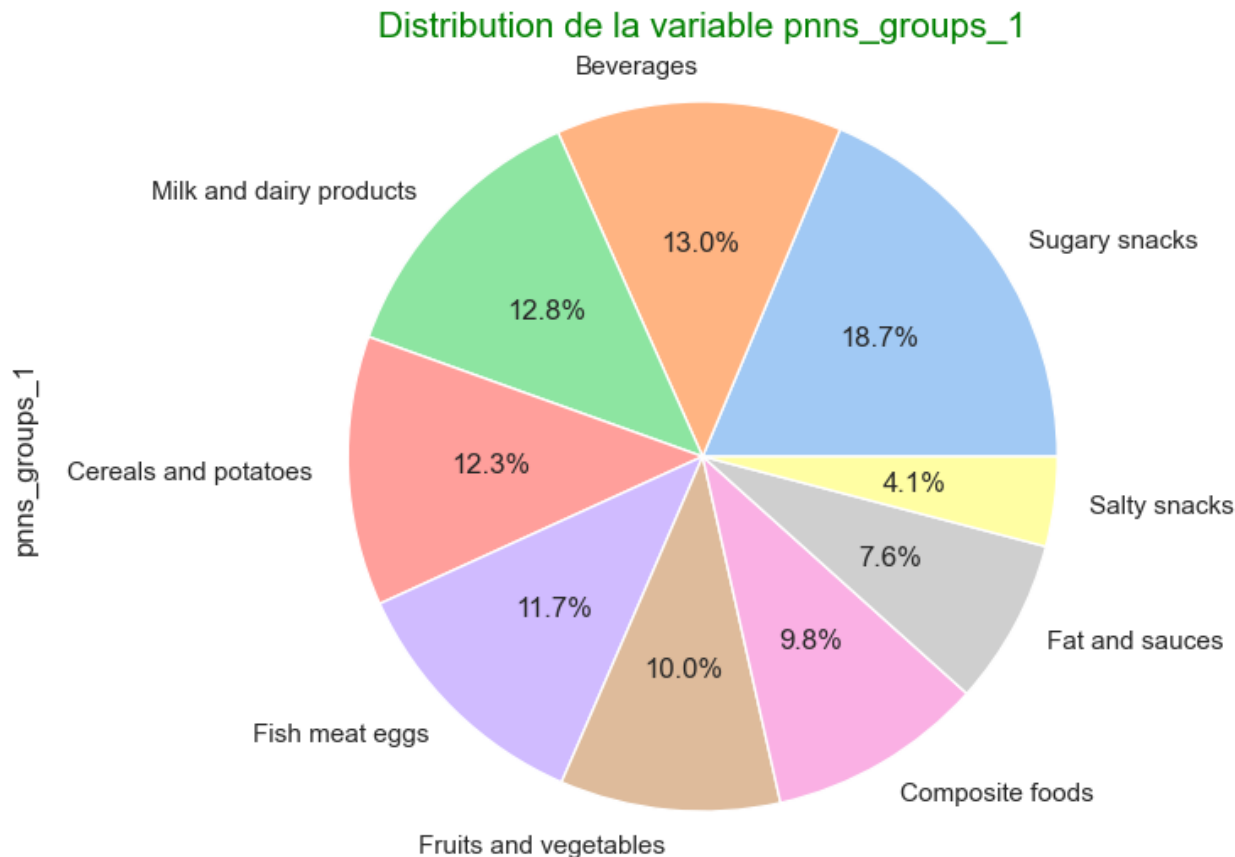


Concevez une application au service de la santé publique

- Appel à projets
- Idée d'application
- Nettoyage des données
- **Analyse exploratoire**
- Synthèse des conclusions

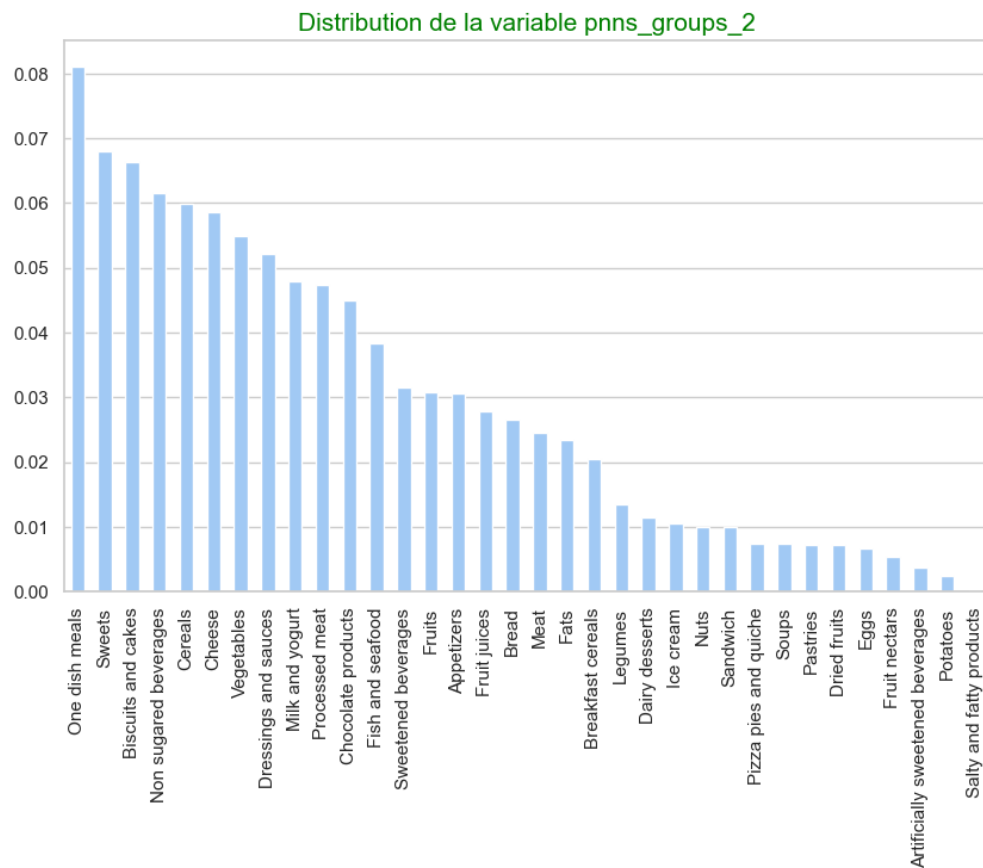
Analyse exploratoire > univariées

Comment se répartissent les familles ?



Analyse exploratoire > univariées

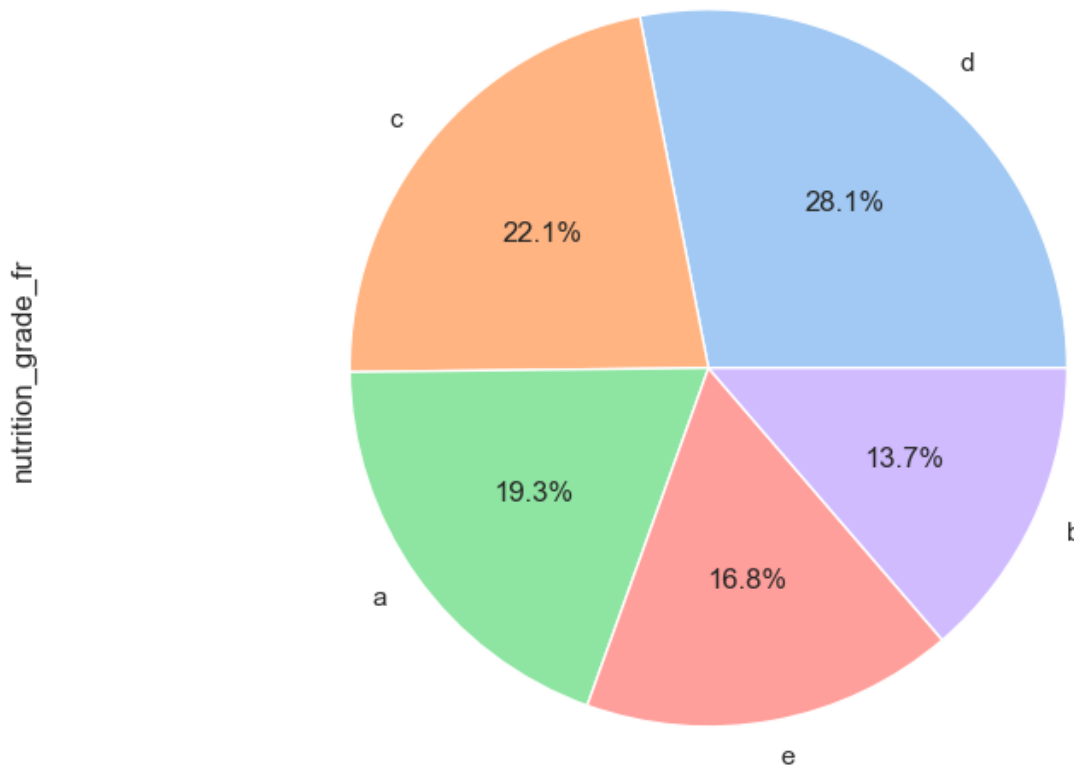
Comment se répartissent les sous-familles ?



Analyse exploratoire > univariées

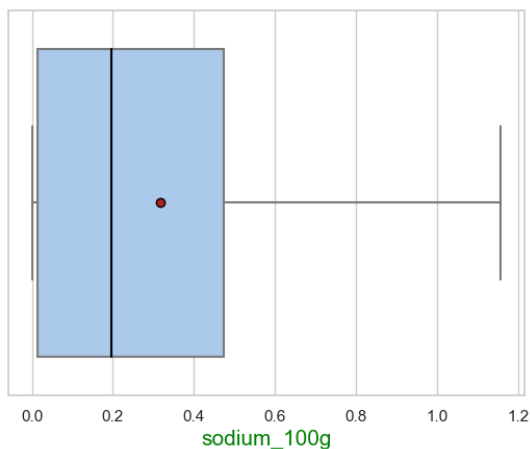
Quel est le Nutri-Score des produits ?

Distribution de la variable nutrition_grade_fr



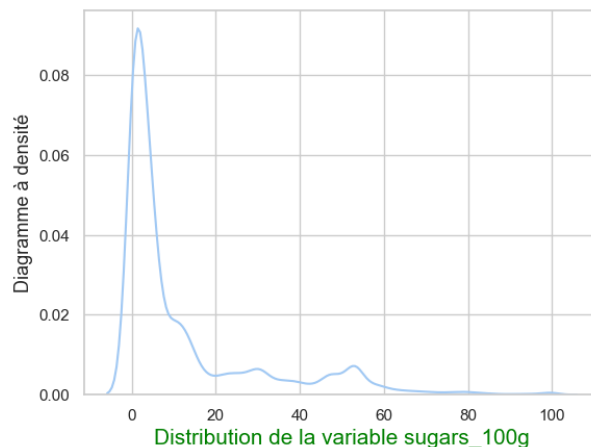
Analyse exploratoire > univariées

Quid des données nutritionnelles ?

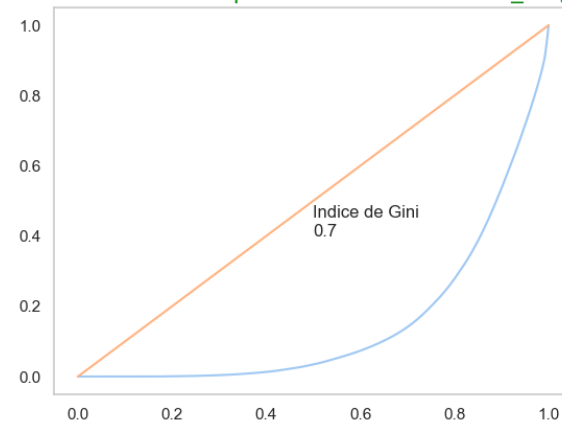


Concentrées en 0+ ...

... avec asymétrie à gauche
et queue à droite...



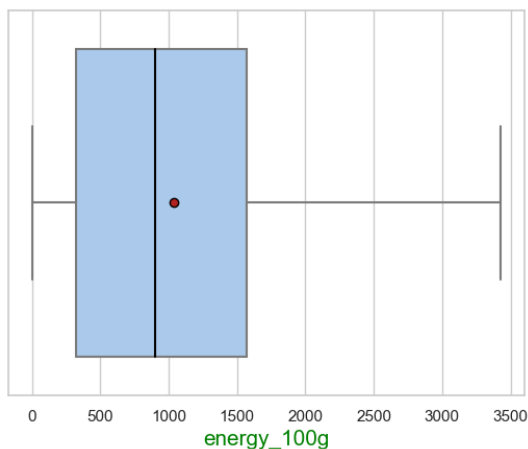
Courbe de Lorenz pour la variable saturated-fat_100g



... et discriminantes.

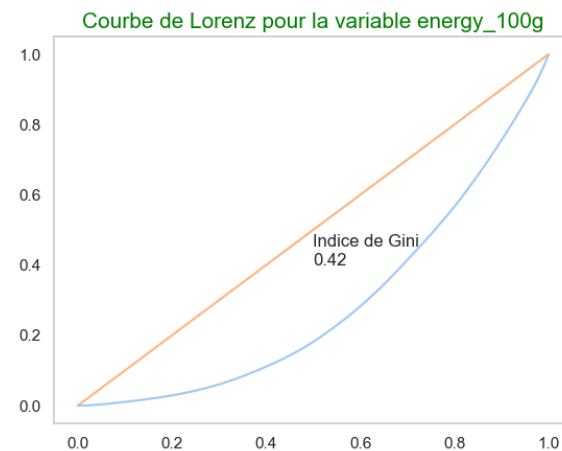
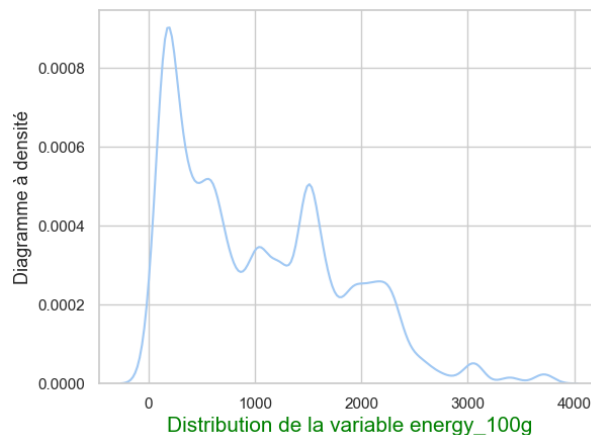
Analyse exploratoire > univariées

L'énergie fait cavalier seul !



Moyenne : 1041.7
Médiane : 899.0

Skewness : 0.69
Kurtosis : -0.17

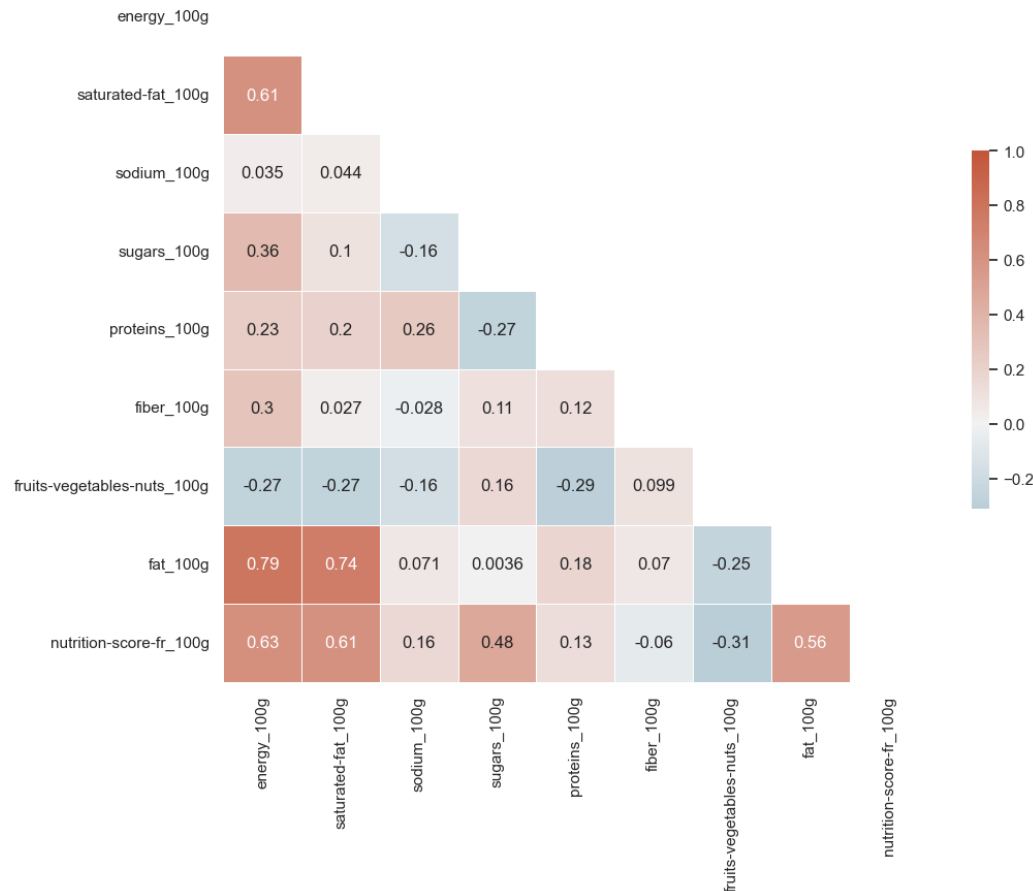


... et moins discriminante.

Analyse exploratoire > bivariées

Des couples assortis ?

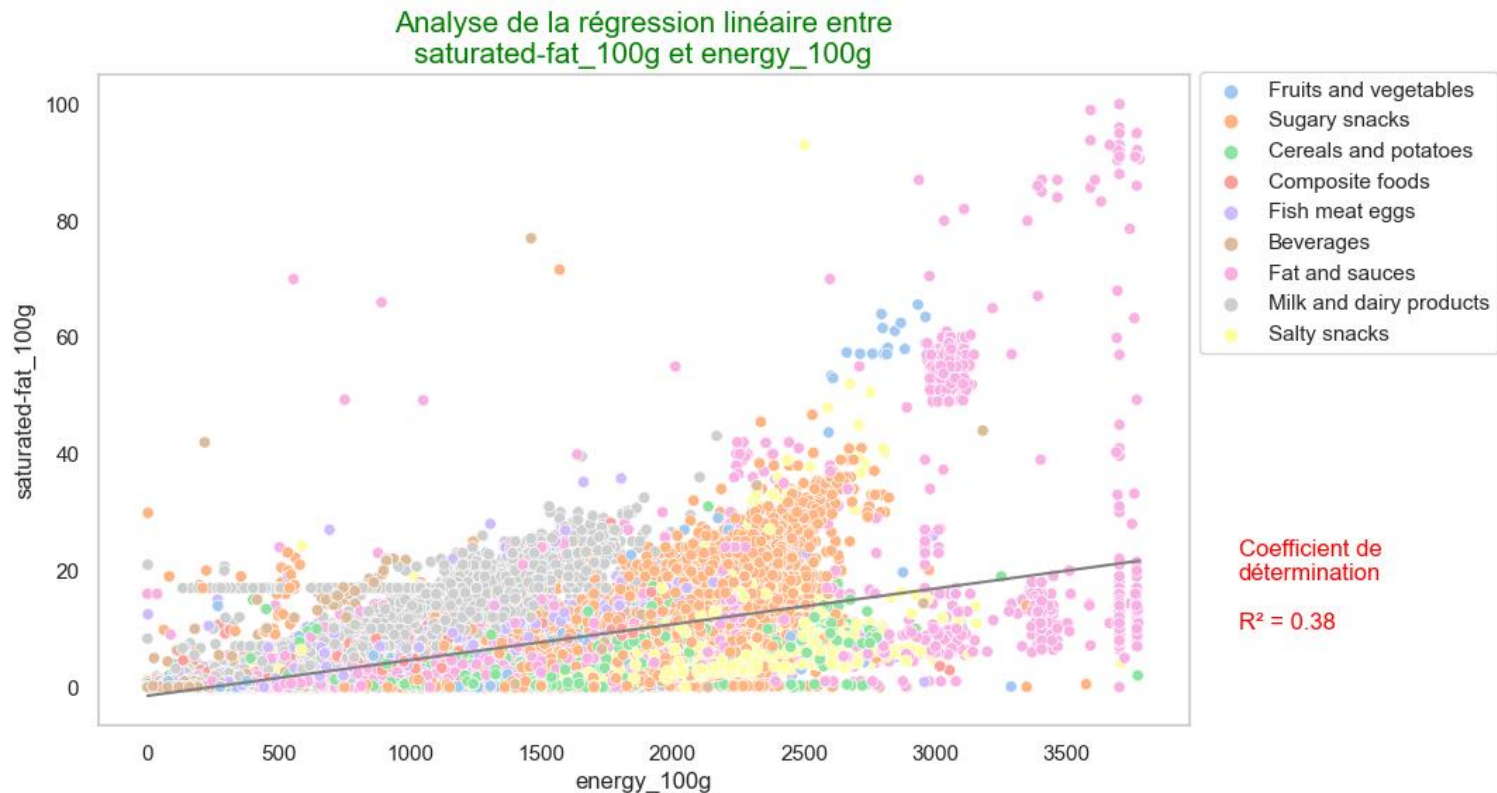
Recherche des corrélations linéaires avec le coefficient de Pearson



Une petite
accointance
entre l'énergie et
les graisses
saturées...?

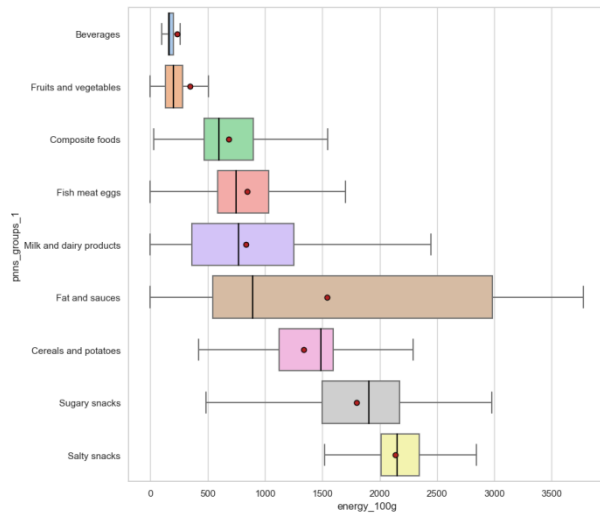
Analyse exploratoire > bivariées

... Ouf, pas vraiment, on évite la redondance !

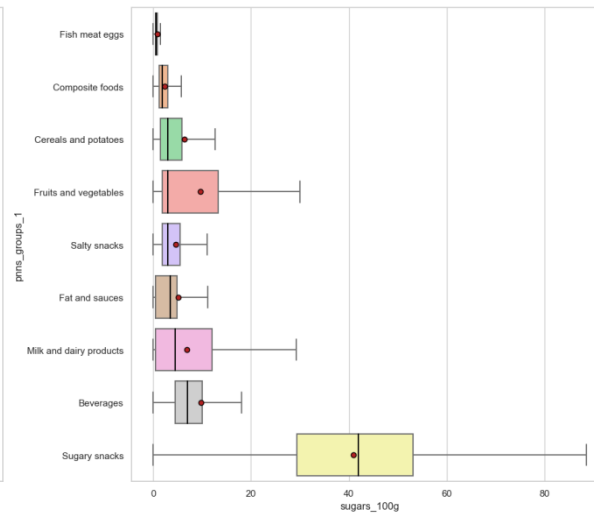


Analyse exploratoire > bivariées

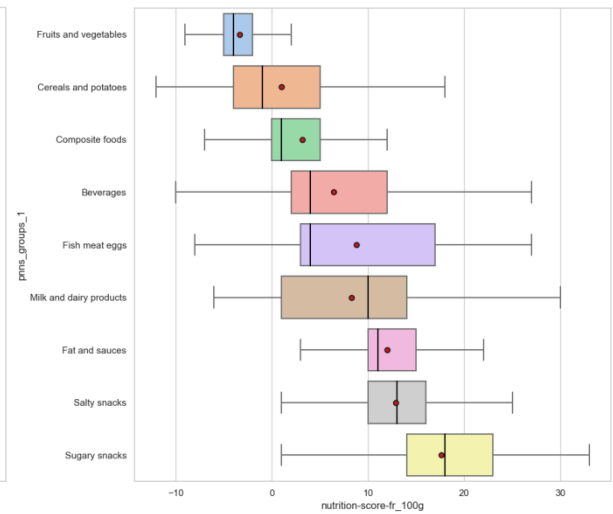
Familles malsaines ?



$$\eta^2 = 0.549$$



$$\eta^2 = 0.593$$

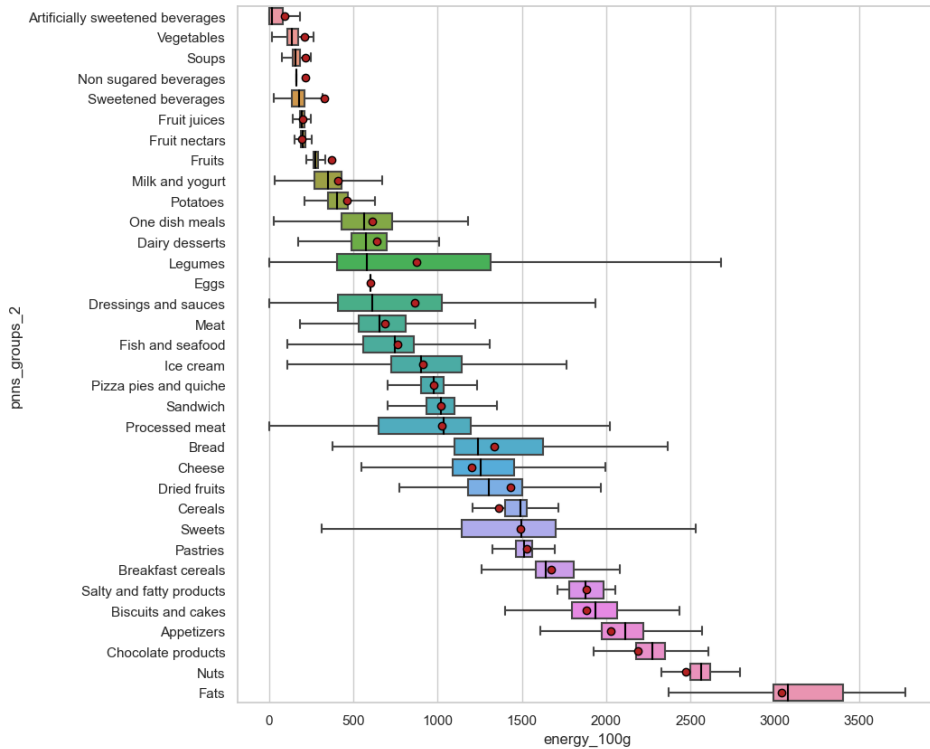


$$\eta^2 = 0.516$$

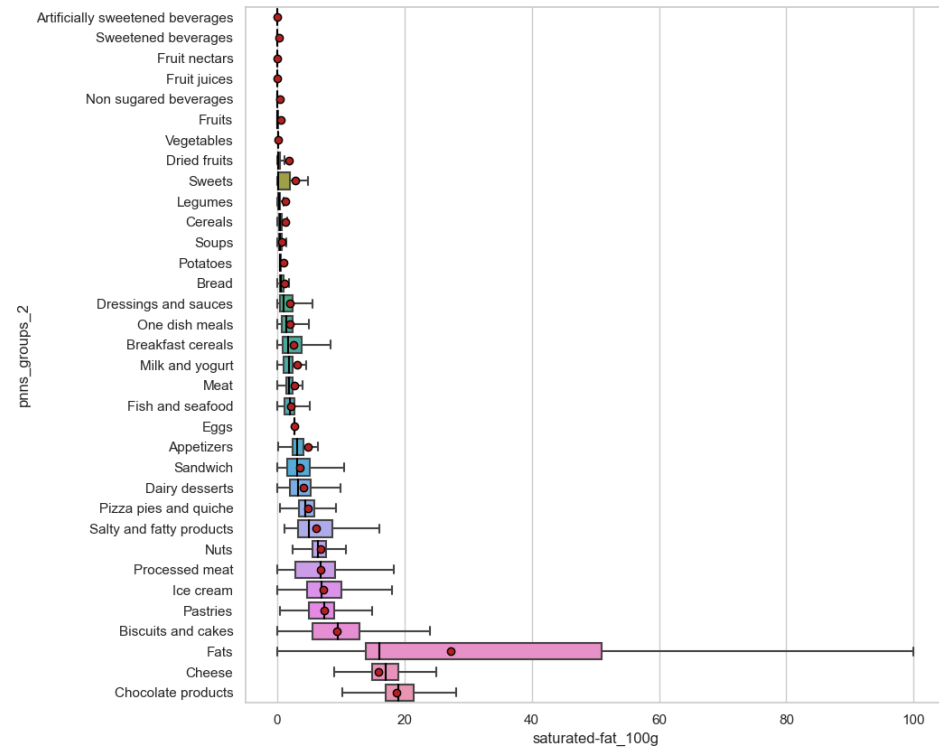
Rapports de corrélation

Analyse exploratoire > bivariées

Encore mieux en comités restreints !



$$\eta^2 = 0.761$$

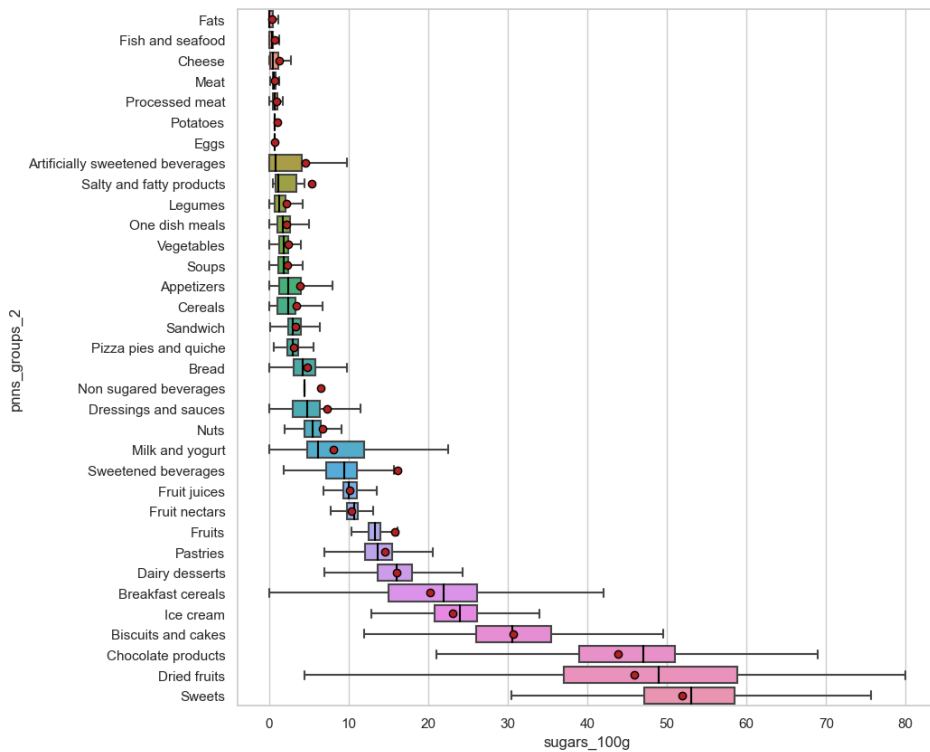


$$\eta^2 = 0.599$$

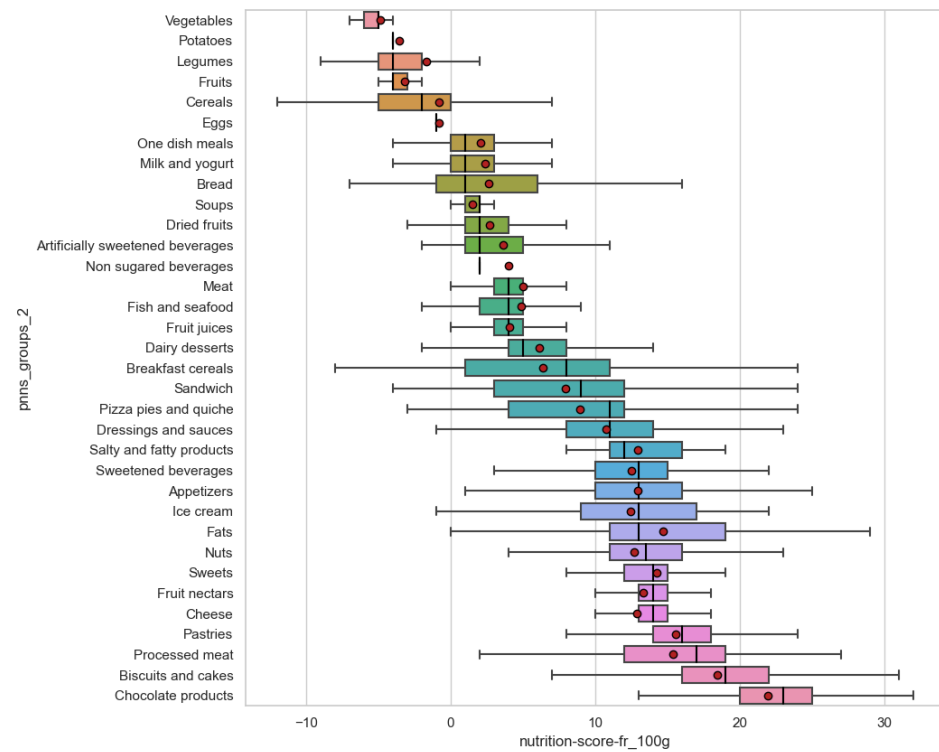
Rapports de corrélation

Analyse exploratoire > bivariées

Encore mieux en comités restreints !



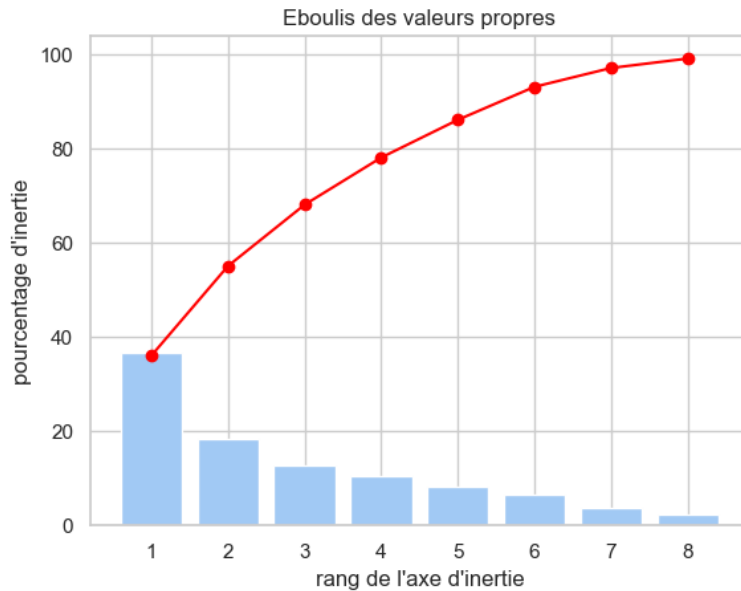
$$\eta^2 = 0.739$$



$$\eta^2 = 0.676$$

Rapports de corrélation

Analyse exploratoire > multivariées

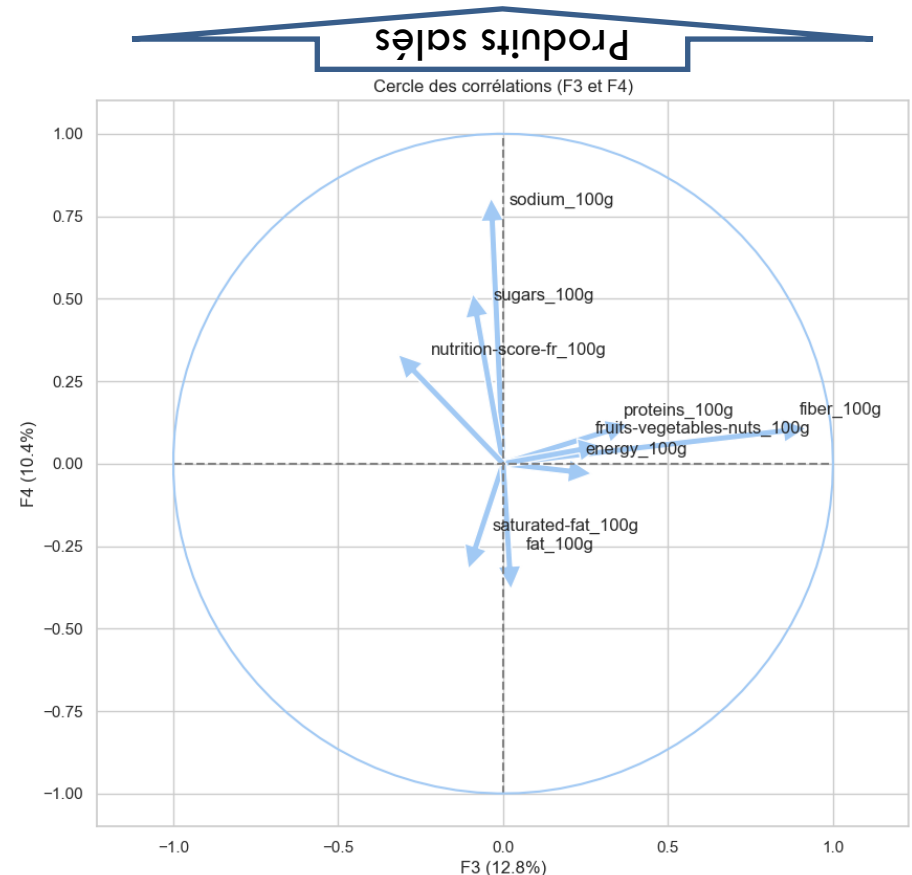
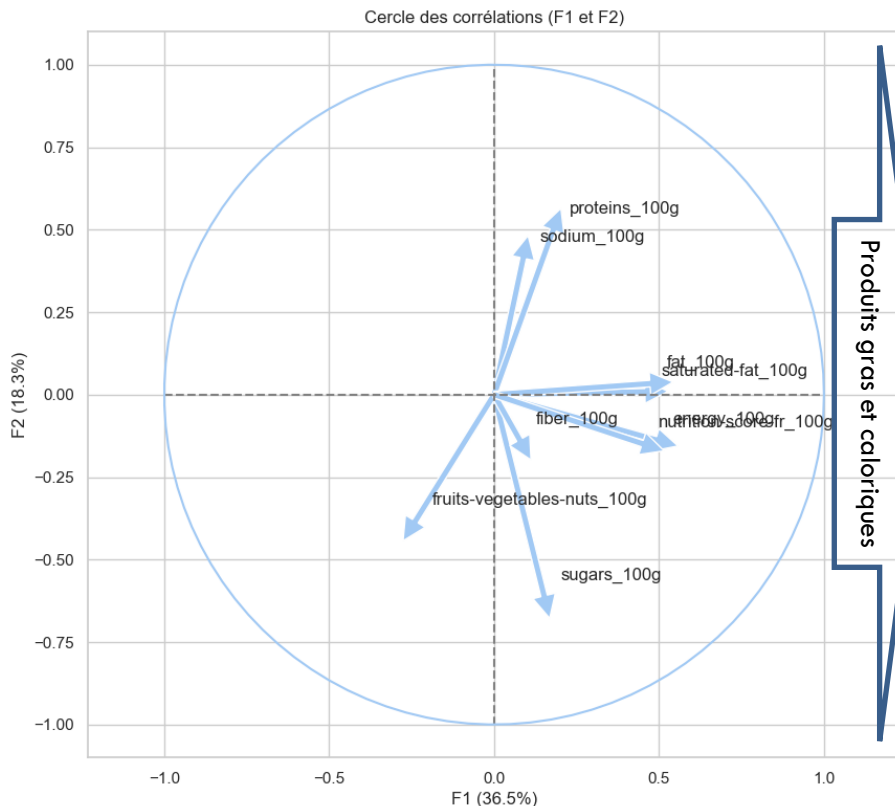


On change de dimensions
avec l'ACP !



Analyse exploratoire > multivariées

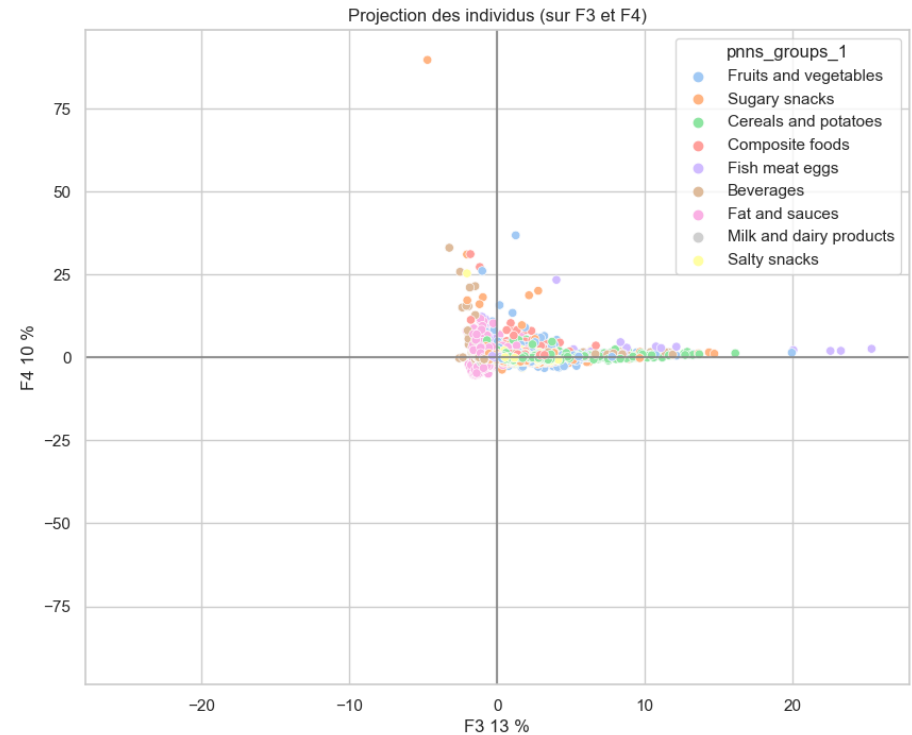
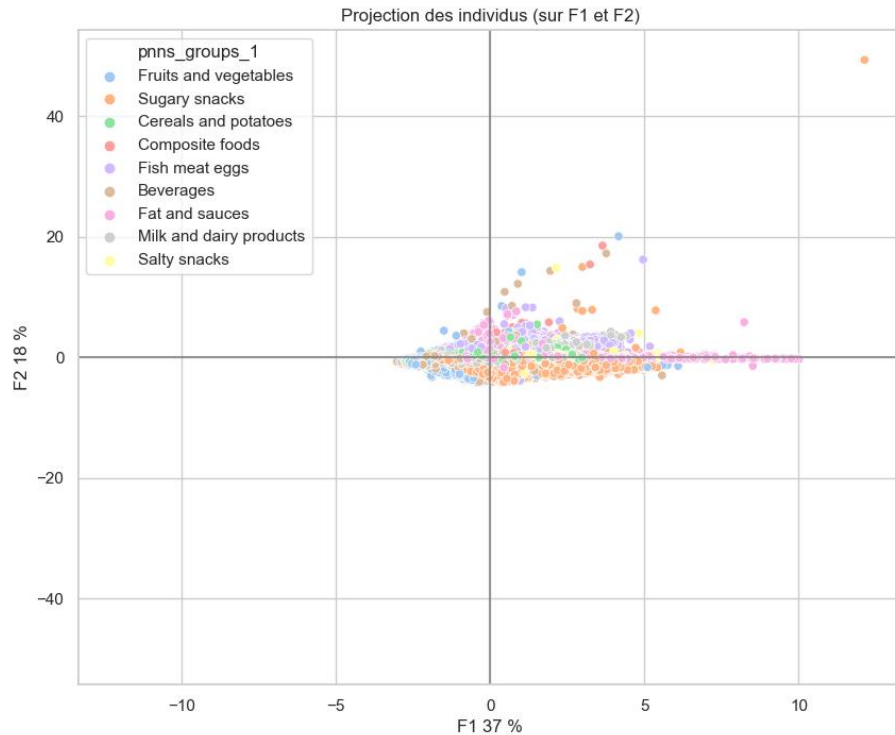
Qui sont les nouveaux chefs ?



Produits sucrés

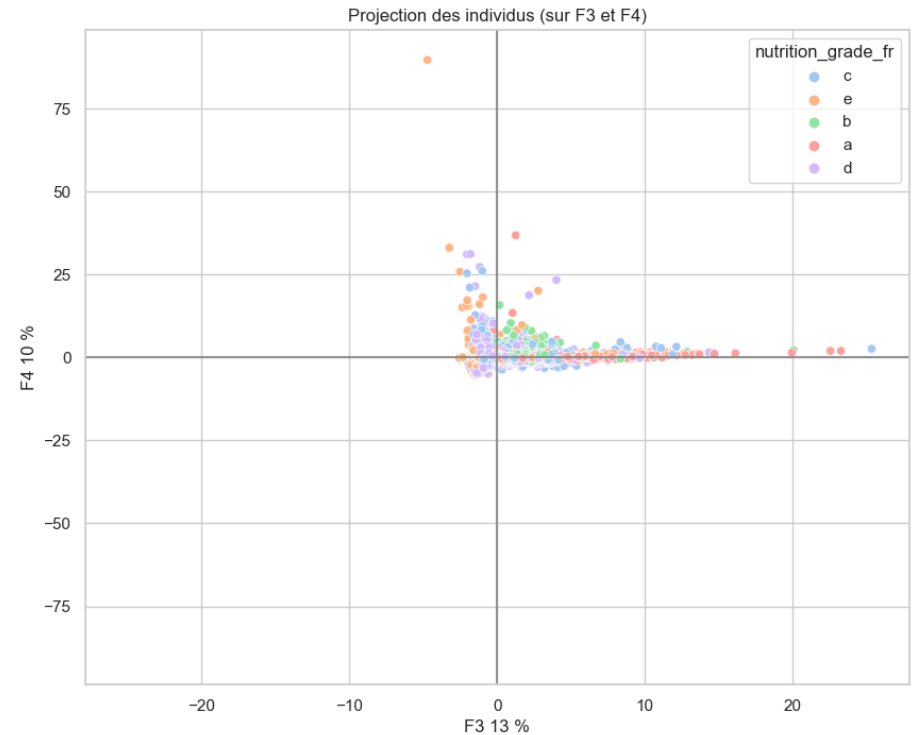
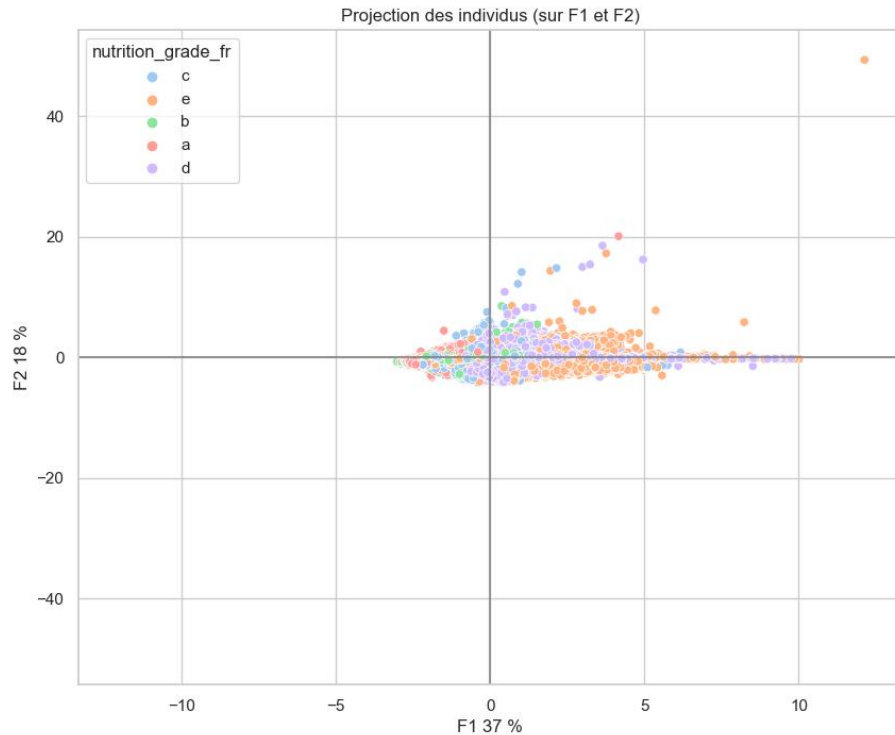
Analyse exploratoire > multivariées

Soldats, en rangs par familles !



Analyse exploratoire > multivariées

Pour les dernières photos, en rangs par grades !



Concevez une application au service de la santé publique

- Appel à projets
- Idée d'application
- Nettoyage des données
- Analyse exploratoire
- **Synthèse des conclusions**

Synthèse des conclusions

- Représentativité des données
- Données nutritionnelles choisies discriminantes
- Indépendance des 4 variables choisies pour le scoring de l'application
- Caractérisation des familles et sous-familles
- Réduction dimensionnelle possible

Avez-vous des questions ?