

Projet 3

Une application au service de la santé publique

Eva Bookjans



En Un Coup d'Oeil

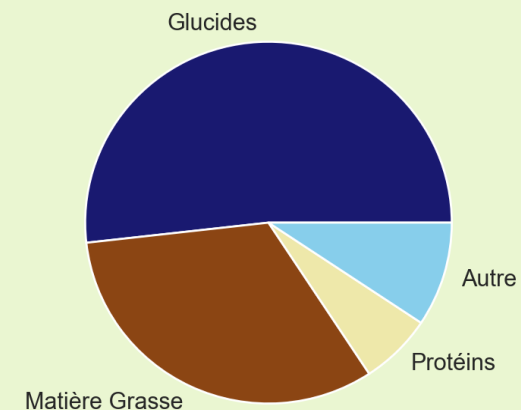
Etape 1

Scanner le code bar



Etape 2

Les informations nutritionnelles



ACIDES GRAS SATURES



SUCRES



SEL



Etape 3

Options

Moins de grasse

Moins de glucides

Plus de protéins

Etape 4

Afficher les meilleurs options

Nutri-Score



Moins de acides gras saturés



57 g



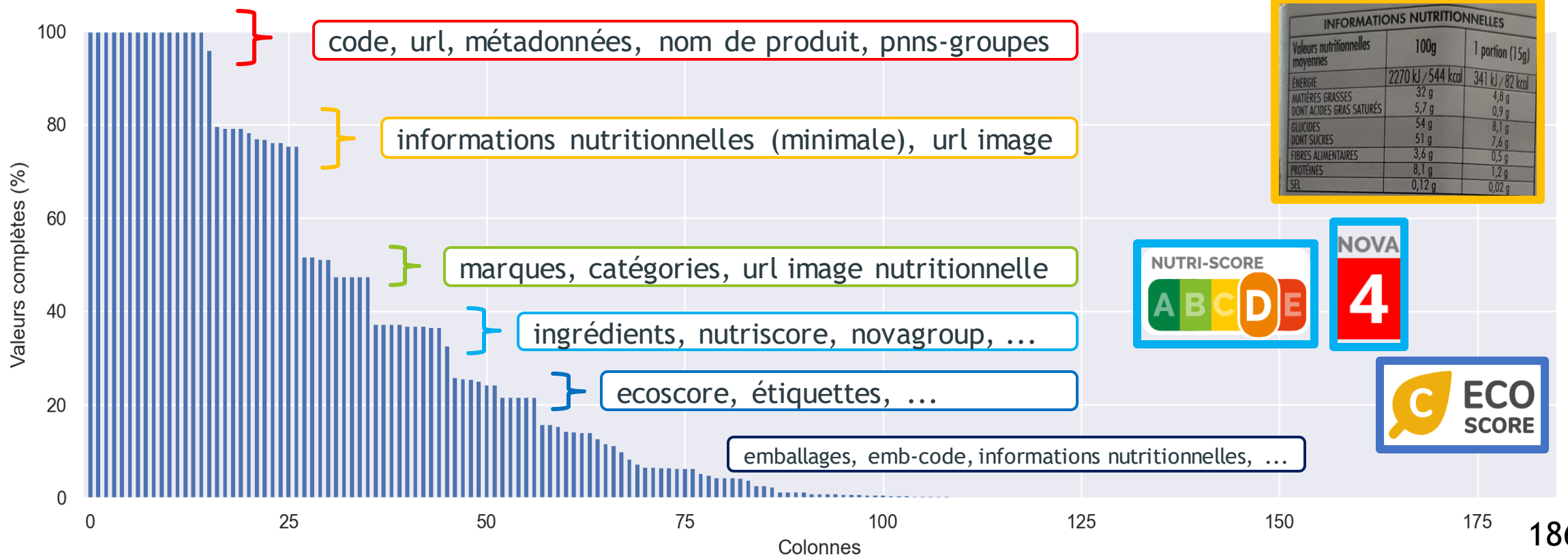
14 g



8 g

Les Données d'OpenFoodFacts

- ▶ > 1.9 Millions produits d'alimentation renseignés
- ▶ **ATTENTION:** valeurs manquantes, entrées erronées et dupliquées !!



Les Catégories Alimentaires

38.4% des catégories PNNS* sont renseignées (\neq 'unknown')

10 (+1) catégories principales

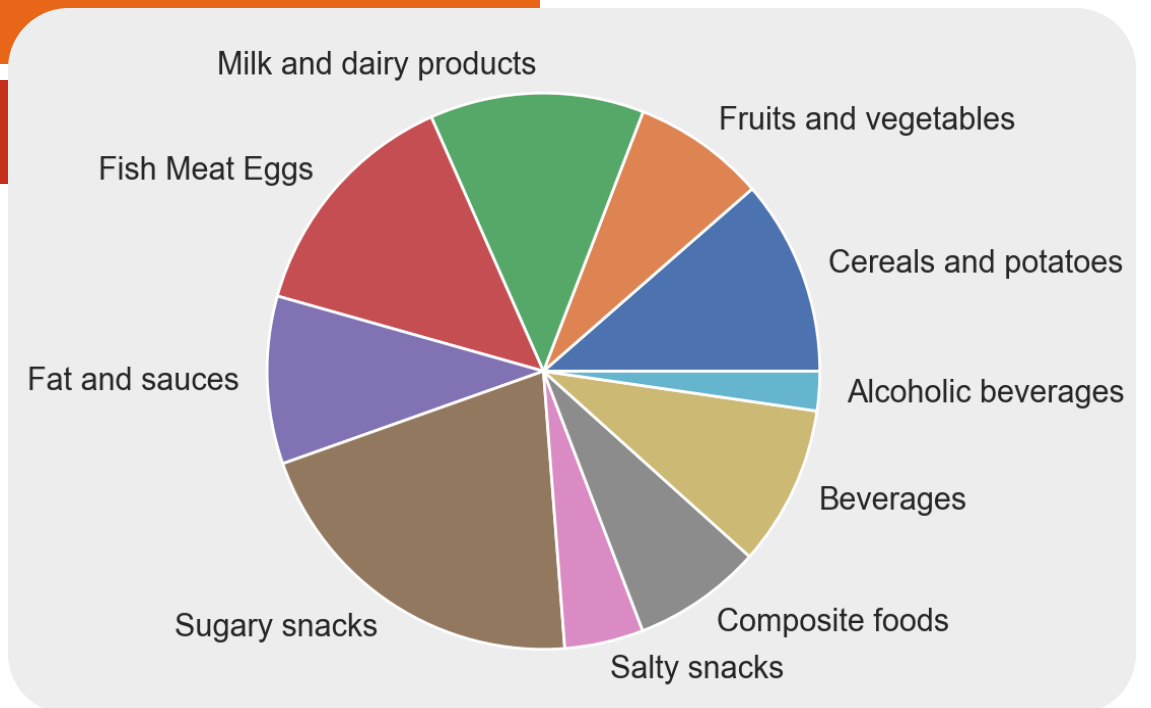
39 (+1) sous-catégories

Un peu de nettoyage à faire :

- nan = 'unknown'
- 'sugary-snacks' = 'Sugary snacks'
- 'Pizza pies and quiche' = 'Pizza pies and quiches'
- 'pastries' = 'Pastries'

MANGER BOUGER
PROGRAMME NATIONAL NUTRITION SANTÉ

* programme national nutrition santé



La NOVA classification

- Une classification pour mettre en évidence le degré de transformation des aliments

NOVA

1

Aliments non transformés ou transformés minimalement



NOVA

3

Aliments transformés



NOVA

2

Ingédients culinaires transformés



NOVA

4

Produits alimentaires et boissons ultra-transformés

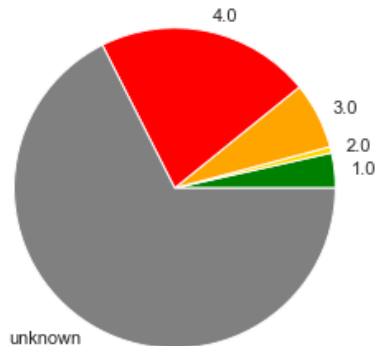


La NOVA classification

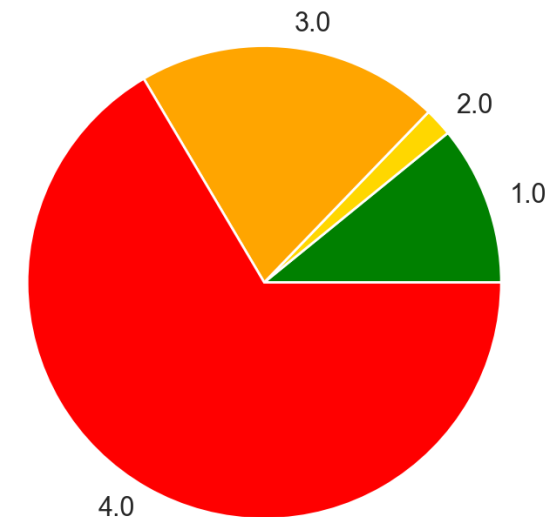
32.4% des éléments ont été attribués à une NOVA catégorie

Nettoyage :

- Définir le variable comme 'string'
- Classer les éléments vides comme 'unknown'



Une grande partie des produits alimentaires renseignés sont ultra-transformés



Les Valeurs Nutritionnelles

68.6% des éléments sont non-vides

- énergie, matière grasse, acides gras saturés, glucides, sucres, protéines, sel (et sodium) pour 100g

De ces éléments, 5.7% ont des valeurs erronées

67.9% des éléments sont 'bien' renseignés

- 1 295 445 produits alimentaires

Nettoyage :

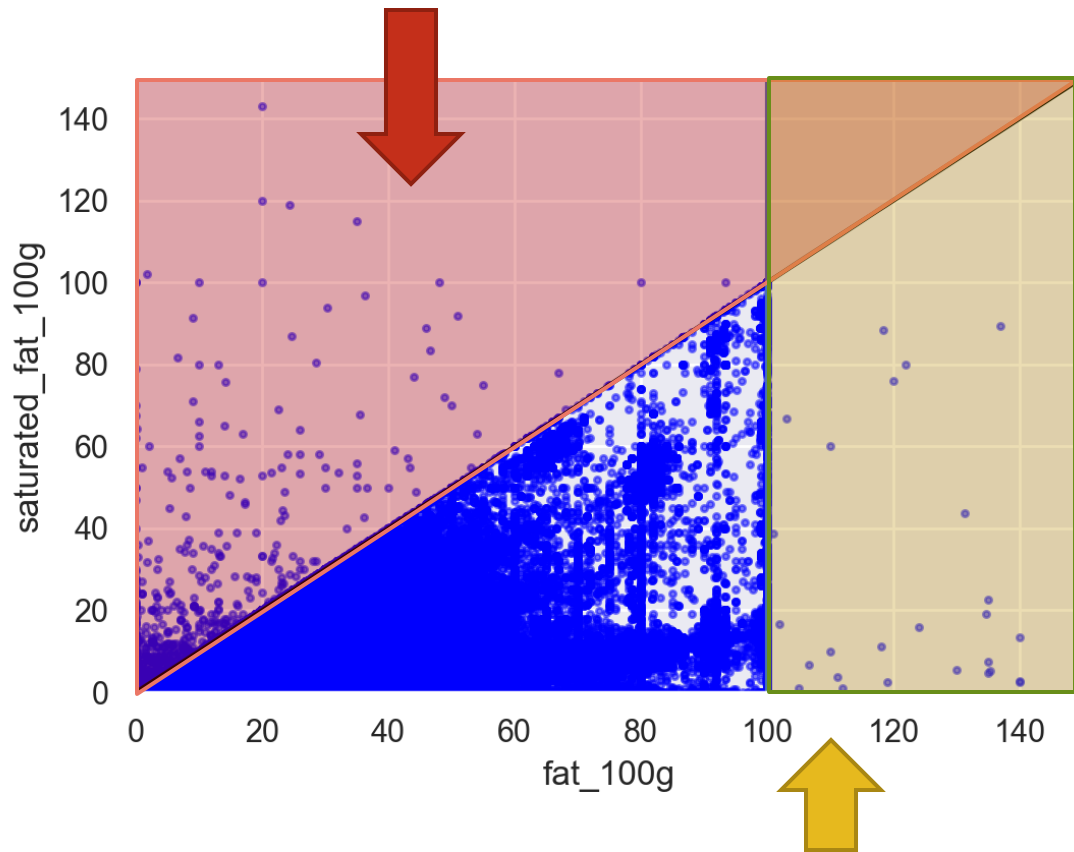
- Écarter les éléments avec des valeurs vides
- Éliminer les éléments avec des valeurs erronées
 - Les variables ne sont pas cohérentes avec
 - leurs définitions
 - leurs énergies nutritionnelles
 - la somme totale
 - l'énergie nutritionnelle totale.

Valeurs nutritionnelles			
Valeurs Nutritionnelles	pour 100g	pour 1/2 tarte soit 175 g	% des AQR* par portion
Énergie	928 kJ 222 kcal	1624 kJ 389 kcal	19%
Matières grasses	13 g	23 g	33%
dont acides gras saturés	6,6 g	12 g	58%
Glucides	20 g	35 g	13%
dont sucres	2,1 g	3,7 g	4%
Fibres alimentaires	1,5 g	2,6 g	
Protéines	5,6 g	9,8 g	20%
Sel	0,82 g	1,4 g	24%

*Apports Quotidiens de Référence pour un adulte type (8400 kJ / 2000 kcal). Cette tarte contient 389 kcal pour une portion de 175 g. Déjeuner et dîner devraient respectivement couvrir 40 et 30% de vos besoins journaliers.

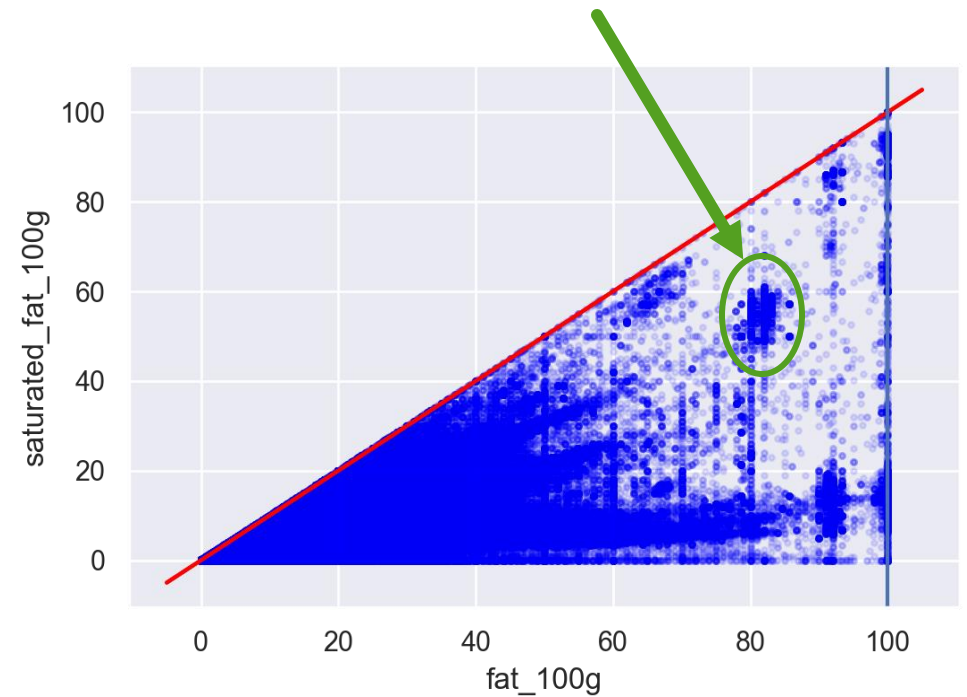
Exemple de Nettoyage

Les acides gras saturés font partie de la matière grasse



Le poids maximale est 100 g

On voit déjà des clusters



Le Nutri-Score

- Un logo qui informe sur la qualité nutritionnelle des produits sous une forme simplifiée et complémentaire à la déclaration nutritionnelle obligatoire (fixée par la réglementation européenne)

À favoriser :

- Fibres
- Protéines
- Fruits, Légumes, Légumineuses, Fruits à coques, Huile de colza, de noix et d'olive

À limiter :

- Énergie
- Acides gras saturés
- Sucres
- Sel

Score / Points		Classement
Aliments solides	Boissons	
-15 à -1	Eaux	A
0 à 2	-15 à 1	B
3 à 10	2 à 5	C
11 à 18	6 à 9	D
19 à 40	10 à 40	E

Le Nutri-Score

36.6% des éléments ont un Nutri-Score renseignées

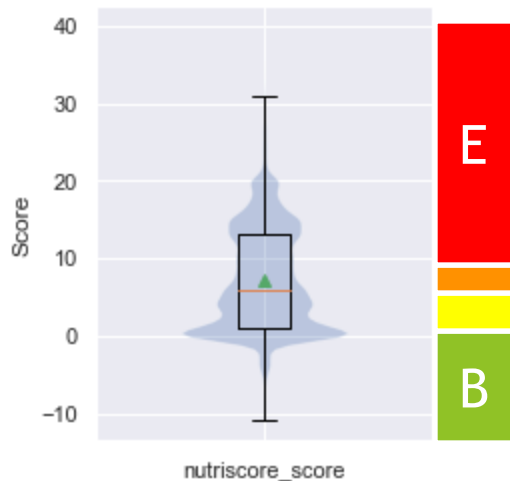
Pour un score renseigné ont a un classement et vice versa

Des éléments renseignées, 98.9% ont un score et classement cohérente

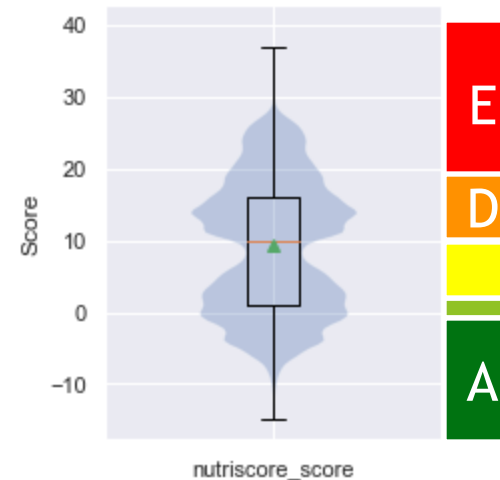
Nettoyage :

- Écarter les éléments vides
- Éliminer les valeurs pour lesquels le score et le classement ne sont pas cohérente
 - Utiliser les catégories alimentaires PNNS (boisson, non-boisson, pas catégorisé)

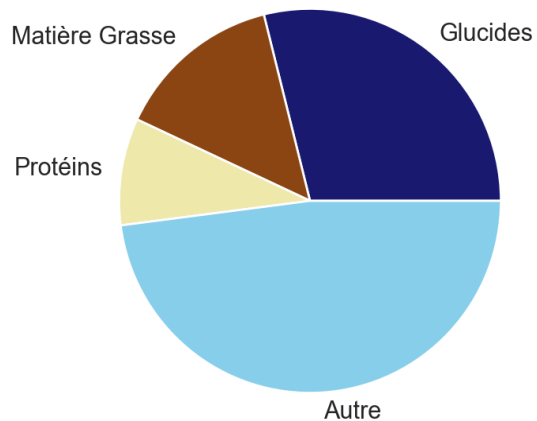
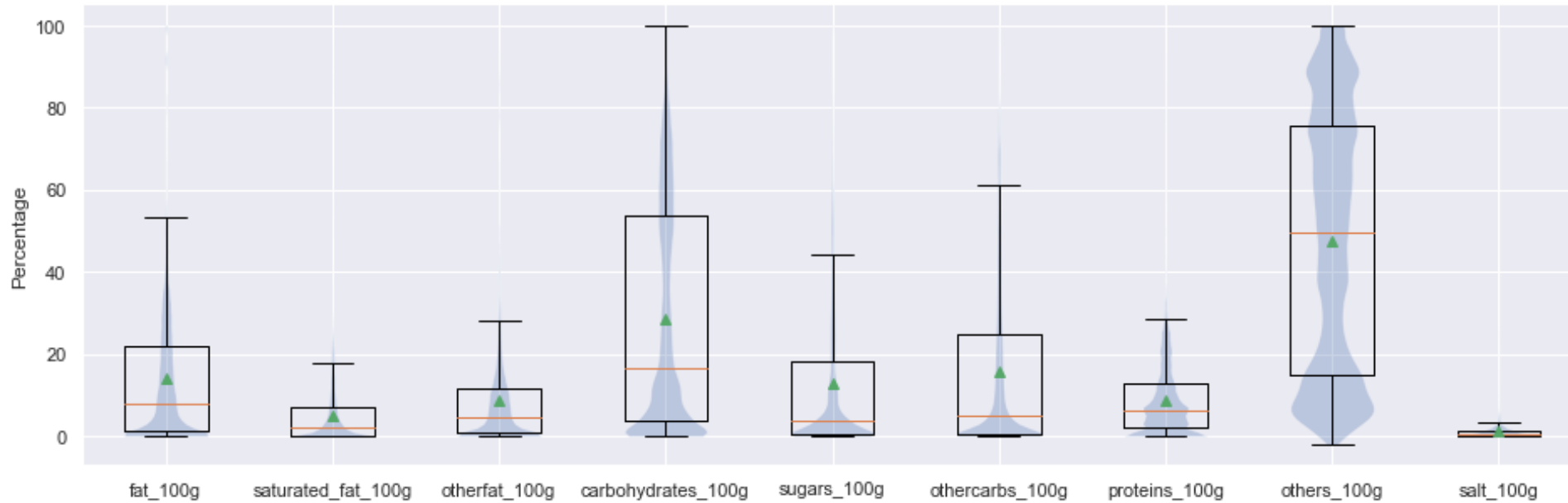
BOISSONS



ALIMENTS SOLIDES



Analyse des Valeurs Nutritionnelles

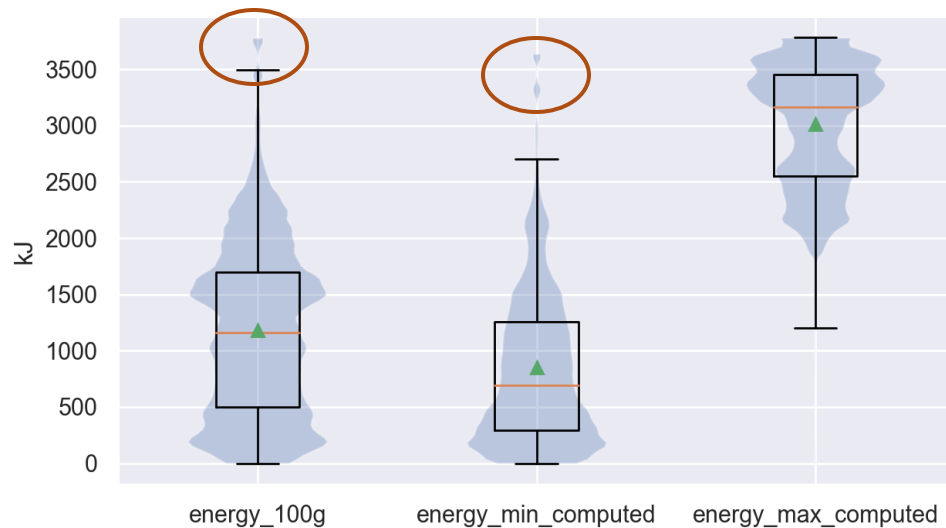


En moyenne, la moitié d'un aliment est composé d'autre chose que glucides, matière grasse ou protéines

- de l'eau, des fibres alimentaires...



Analyse des Valeurs Nutritionnelles

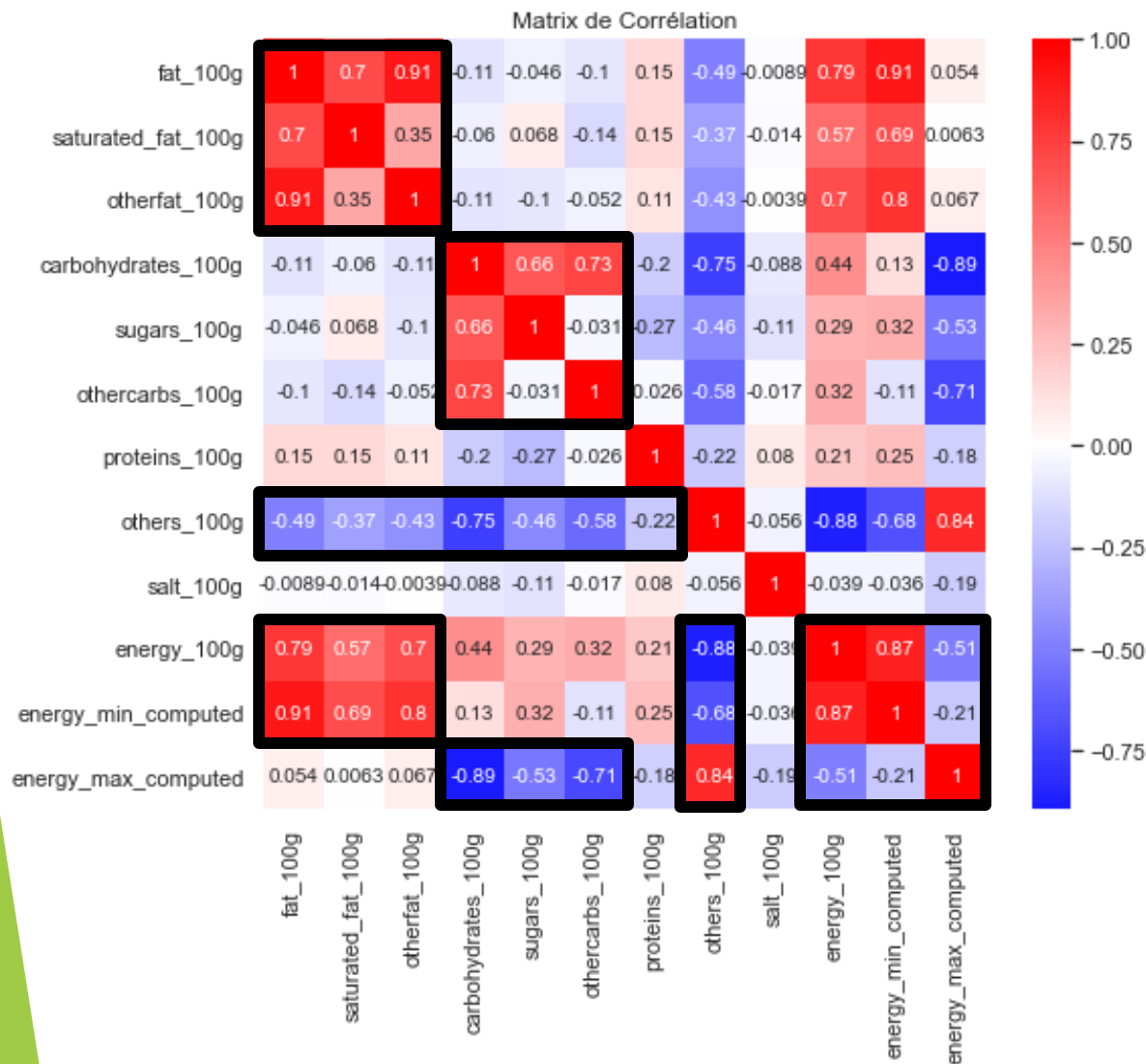


Maximum énergie
= 100g matière grasse
= 3700 kJ



- ▶ Énergie minimum = $36.2 * \text{matière grasse} + 16.5 * \text{protéines} + 15.3 * \text{sucres}$ (kJ / g)
- ▶ Énergie maximum = $37.8 * \text{matière grasse} + 17.5 * \text{protéines} + 18.7 * \text{glucides} + 37.8 * \text{"autres"}$ (kJ / g)

Corrélations entre les Valeurs Nutritionnelles



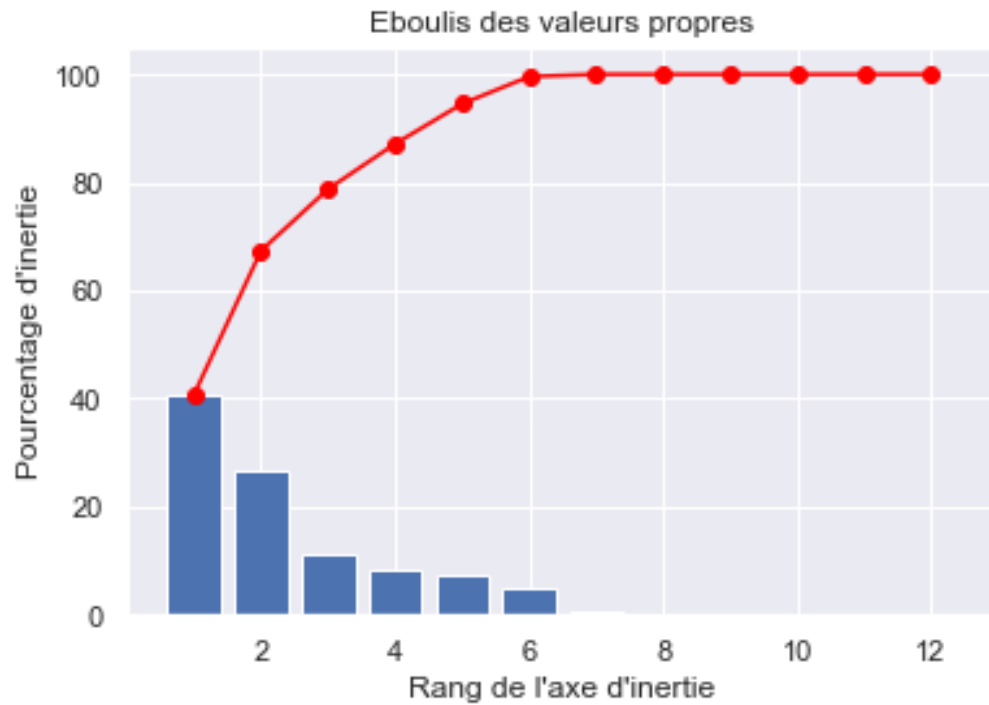
Réduction de dimension

→ "PCA"

Analyse en Composantes Principales

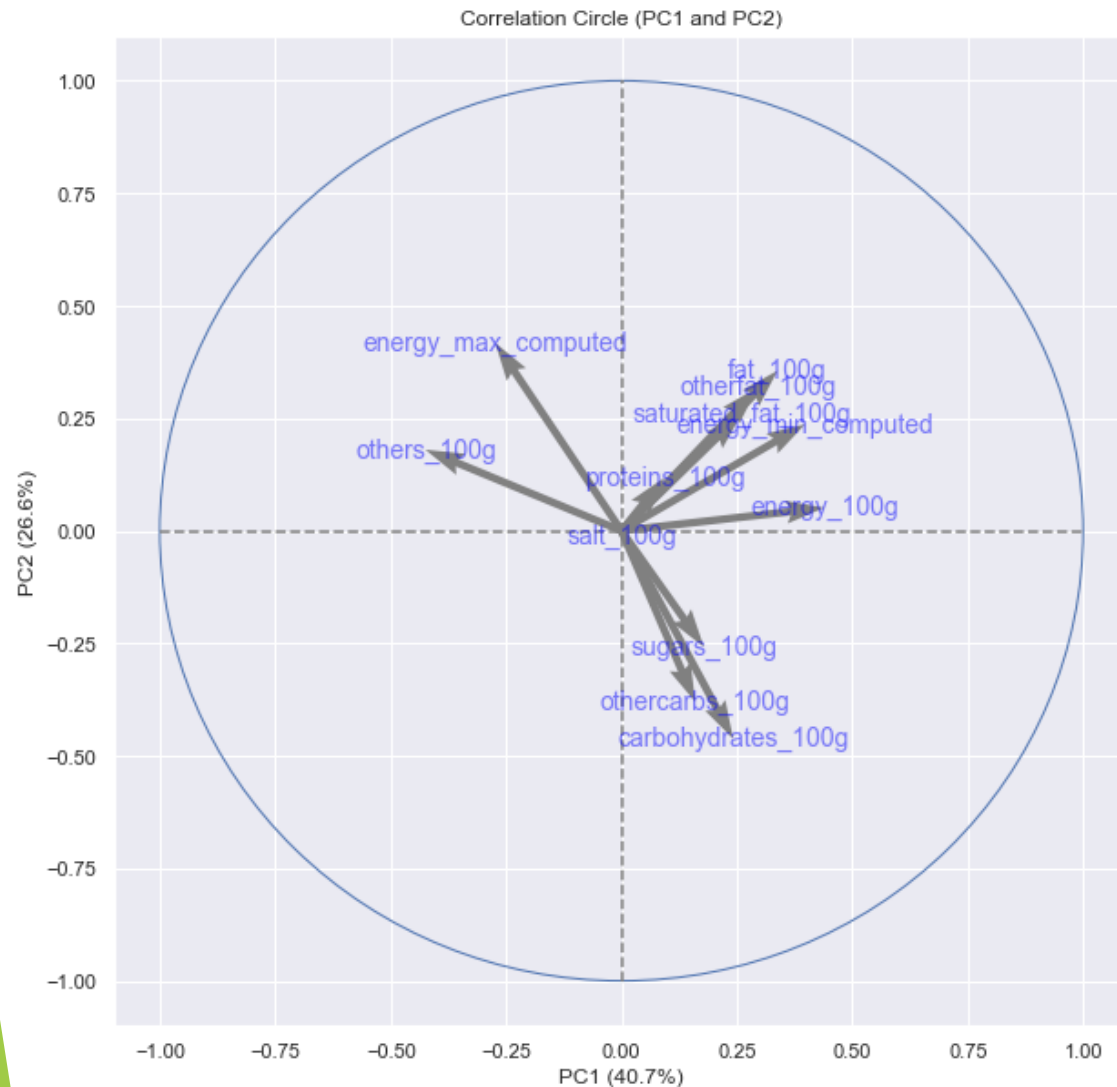
Analyse en Composantes Principales

Réduction de 12 à 6 variables sans perdre de l'information



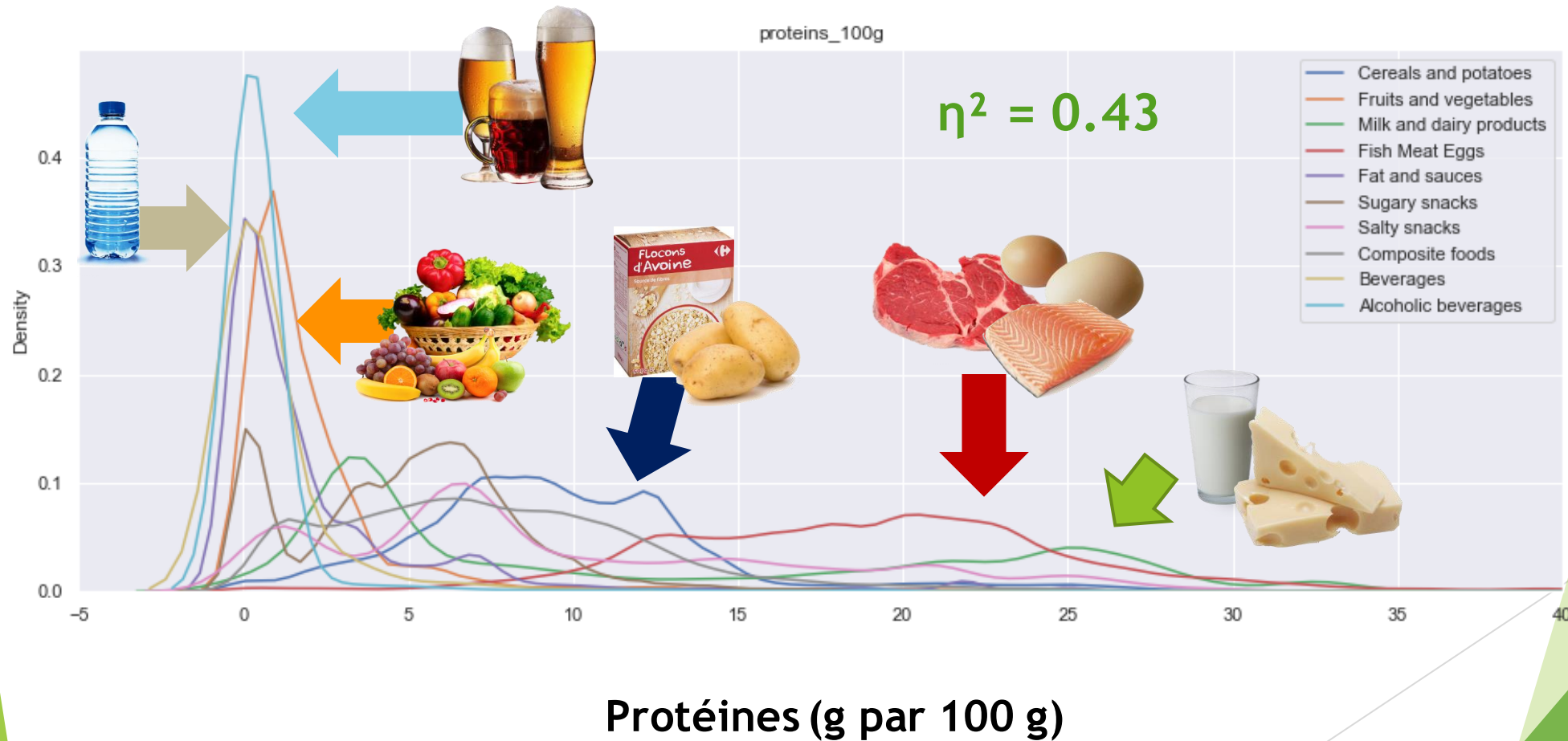
/100	Matière Grasse	Acides Gras Saturés	Acides Gras Non-saturé	Glucides	Sucres	Autres Glucides	Protéins	Autre	Sel	Energie	Min. Energie	Max. Energie
PC1	34	36	29	24	18	16	9.3	-42	0.04	44	40	-27
PC2	36	26	32	-46	-26	-38	12	18	-1.1	5.1	23	42
PC3	-2.4	-9.5	2.3	-13	-54	31	58	-9.3	43	1.8	-12	-19
PC4	-8.5	16	-20	-6.1	41	-45	8.7	-0.8	71	-8.3	11	-15
PC5	17	-17	32	3.9	-19	22	-69	2.4	51	1.9	-8.2	8.2
PC6	2.4	80	-43	2.8	-22	24	-21	2.3	3.2	-1.5	-12	4.1

Réduction des variables



- La matière grasse et les protéines sont la même axe
- Les glucides sont orthogonales à la matière grasse et aux protéines
- L'énergie max. calculée est opposée aux glucides

Différents groupes alimentaires différents valeurs nutritionnelles



Différents groupes alimentaires différents valeurs nutritionnelles

Analysis of
Variance (ANOVA)

	PNNS - groupes 1	PNNS - groupes 2	NOVA groupes
Matière grasse	0.26	0.61	0.089
Acides gras saturés	0.17	0.46	0.067
Ac. gr. non-saturés	0.28	0.55	0.064
Glucides	0.63	0.73	0.031
Sucres	0.51	0.65	0.062
Autre glucides	0.52	0.66	0.012
Protéines	0.42	0.60	0.014
Autres	0.50	0.68	0.033
Sel	0.11	0.19	0.010
Énergie	0.36	0.65	0.054

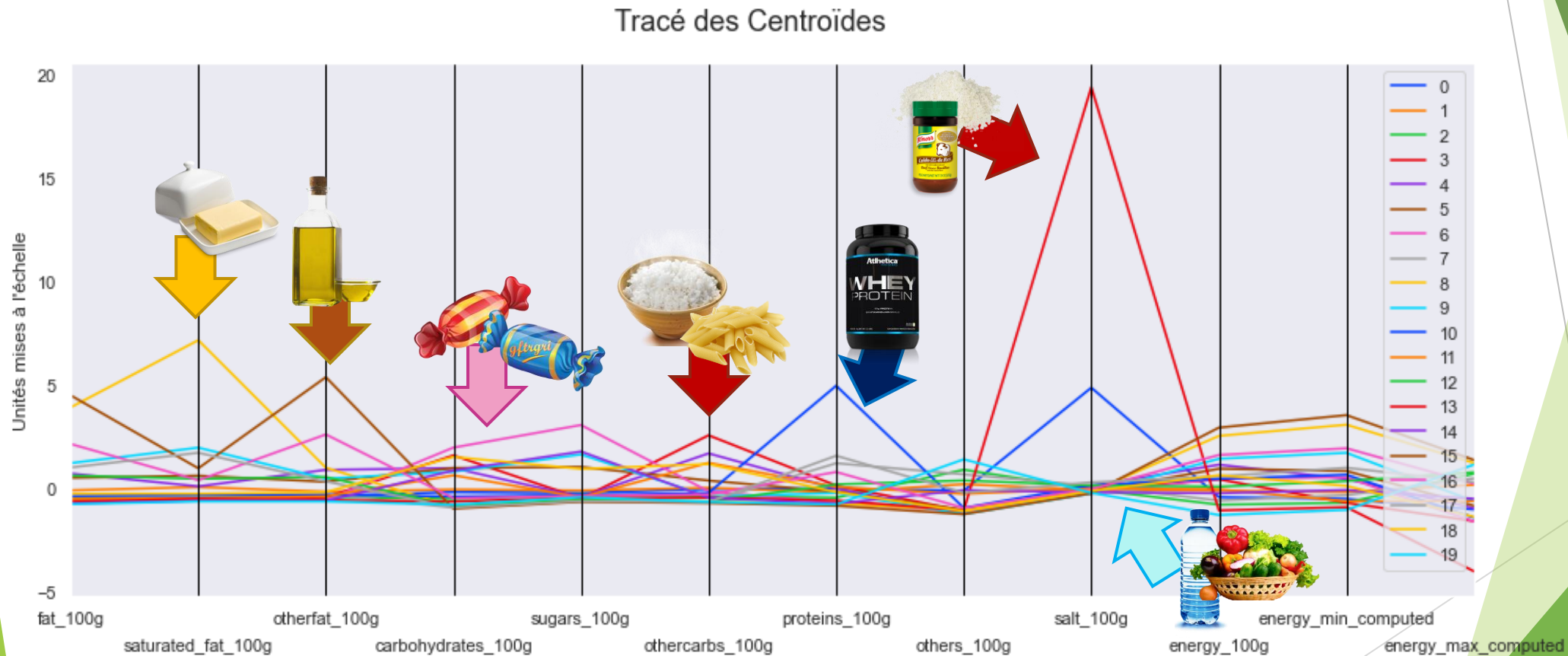
$$\eta^2 = \text{Variance_interclass} / V_{\text{total}}$$

Il y a une corrélation
entre les valeurs
nutritionnelles et les
groupes PNNS.

Une corrélation entre les
valeurs nutritionnelles et
les groupes NOVA n'est
pas apparent.

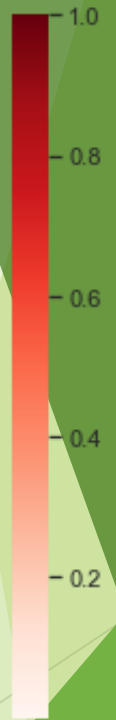
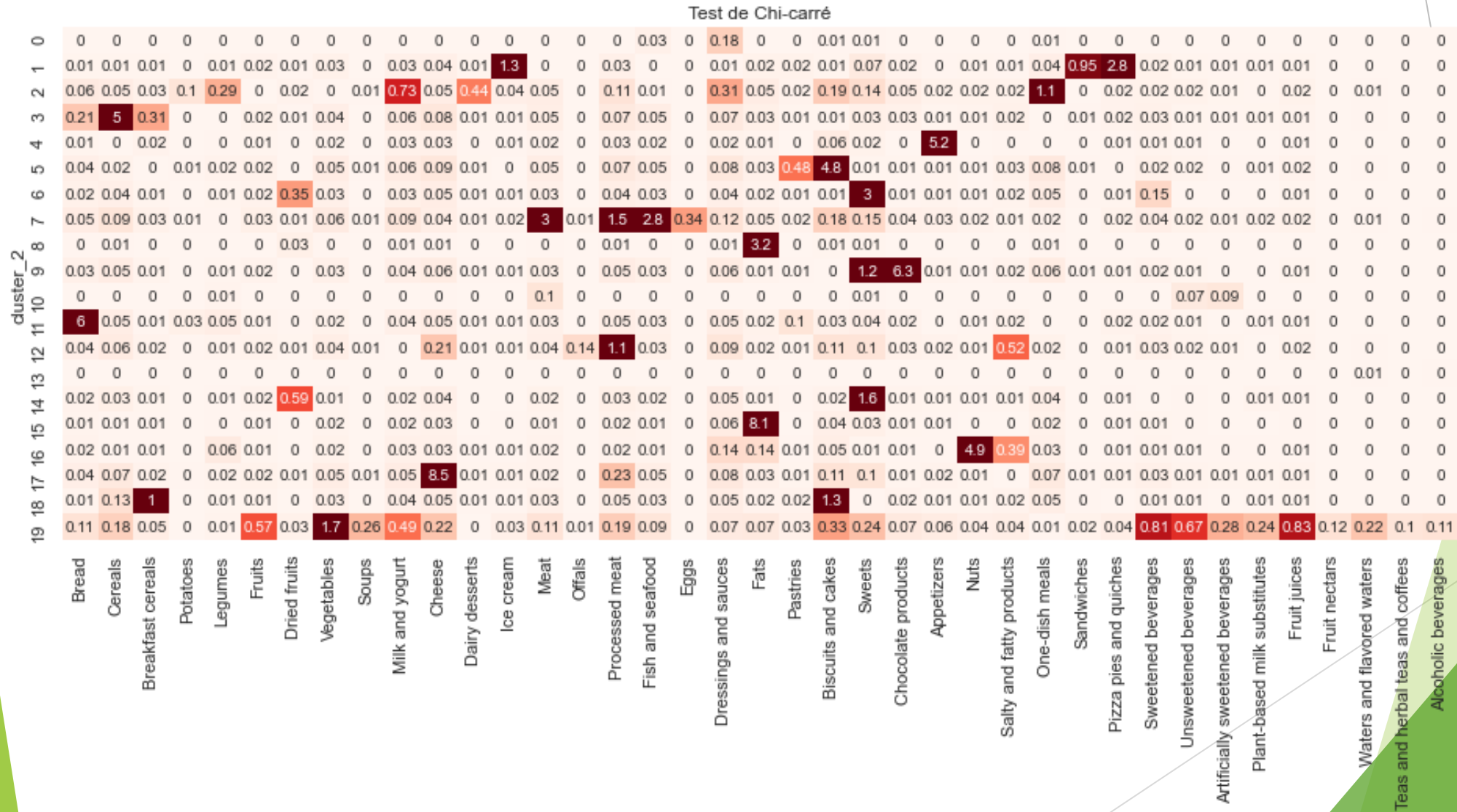
Est-ce qu'on peut trouver les groupes d'alimentation des valeurs nutritionnelles ?

K-Means
Clustering



Regard-on de plus près

ξ - une mesure de dépendance



Regard-on de plus près

Dressings and sauces

Ice cream,
Sanwiches, Pizza
pies and quiches

	PNNS groups
0	(Dressings and sauces)
1	Ice cream, Sandwiches, Pizza pies and quiches
2	Legumes, Potatoes, Milk and yogurt, Dairy desserts, Dressing and sauces, One-dish meals
3	Cereals, (Breakfast cereals, Bread)
4	Appetizers
5	Pastries, Biscuits and cakes
6	Sweets, (Dried Fruits)
7	Meat, Processed meat, Fish and seafood, Eggs
8	(Fats)
9	Chocolate products, (Sweets)

	PNNS groups
10	?? (Meat, Artificially sweetened beverages, Unsweetend beverages)
11	Bread, (Pastries)
12	Offals, Salty and fatty products, (Processed Meat, Cheese)
13	?? (Waters and flavored waters)
14	Dried fruits, (Sweets)
15	Fats
16	Nuts, (Salty and fatty products, Dressing and sauces, Fats)
17	Cheese, (Processed meat)
18	Breakfast cereals, (Biscuits and cakes)
19	Fruits, Vegetables, Soups, All Beverages, (Cheese, Milk and Yogurt, Biscuits and cakes)

Vérifier avec des WordClouds

- noms des produits

Dressings and sauces

Cluster 0



Ice cream, Sandwiches, Pizza pies and quiches

Cluster 1



Cereals

Cluster 3



Appetizers

Cluster 4



Pastries, Biscuits and cakes

Cluster 5



Breakfast cereals, (Biscuits and cakes)

Cluster 18



Vérifier avec des WordClouds - noms des produits

Fats



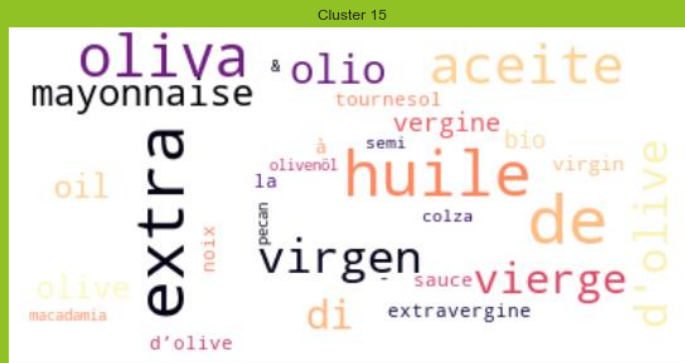
Sweets, (Dried Fruit)



Dried Fruit, (Sweets)



Fats



Chocolate products, (Sweets)



Bread



Vérifier avec des WordClouds - noms des produits

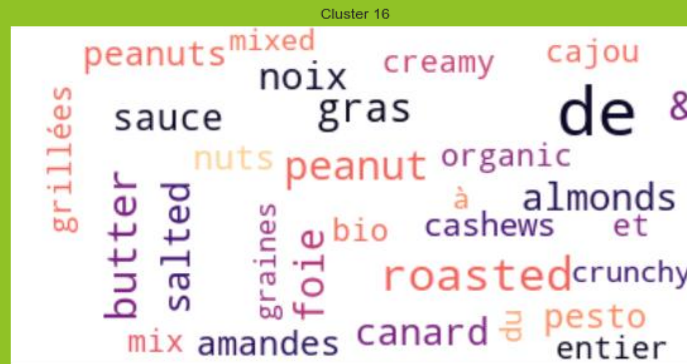
Offals, Salty and fatty products, (Cheese, Processed Meat)



Cheese, (Processed Meat)



Nuts, (Salty and fatty products,
Dressing and sauces, Fats)



Legumes, Potatoes, Milk and yogurt, Dairy desserts, Dressing and sauces, One-dish meals



Meat, Processed meat, Fish and seafood, Eggs



Fruits, Vegetables, Soups, All Beverages

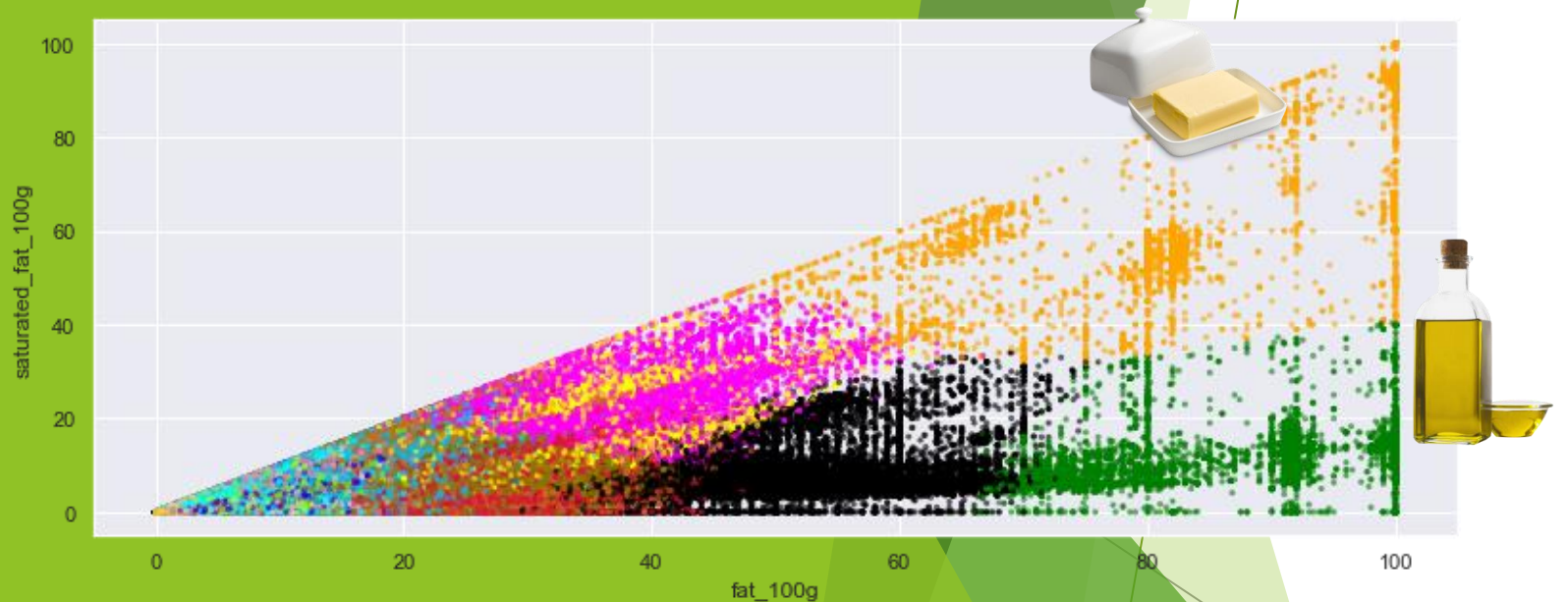


K-Means -une méthode pour détecter des clusters

?? = Proteins



?? = Salt / Seasoning



K-Means -une méthode pour détecter des clusters

Le K-Means clustering est capable de détecter des différents groupes d'alimentation !

- Pour la plupart, c'est cohérente avec les groupes PNNS
- Mais peut être améliorer!
 - Les boissons sont ensemble avec les fruits et légumes
 - Utiliser une autre clustering méthode (e.g. Gaussian Mixture Models)
 - Utiliser plus des information nutritionnelles

Une application nutritionnelle basée sur OpenFoodFacts

- ▶ Nettoyage est important!
 - ▶ Les données ne sont pas toutes standardisées (différents langages et formats)
 - ▶ Valeurs erronées et dupliquées (e.g. les codes à barré)
- ▶ Il y a des beaucoup des informations (pertinentes) !
 - ▶ Information nutritionnelles
 - ▶ Nutri-Score - une scoring nutritionnelle déjà simplifiée
- ▶ La base de données est en évolution.
 - ▶ On peut aussi faire évoluer l'application. Ajouter des fonctions plus tard.

Une application nutritionnelle basée sur OpenFoodFacts

- ▶ Les catégorisations alimentaires peuvent être améliorées
 - ▶ Les préférences des potentiels clients
 - ▶ Les informations nutritionnelles (clustering)
- ▶ L'application doit être faisable
 - ▶ Mettre en place un algorithme pour mieux compléter les valeurs vides
 - ▶ Eventuellement mieux interpréter les informations des ingrédients, catégories, etc. qui sont saisies librement