

Analyse des données

Application au service de la
santé publique



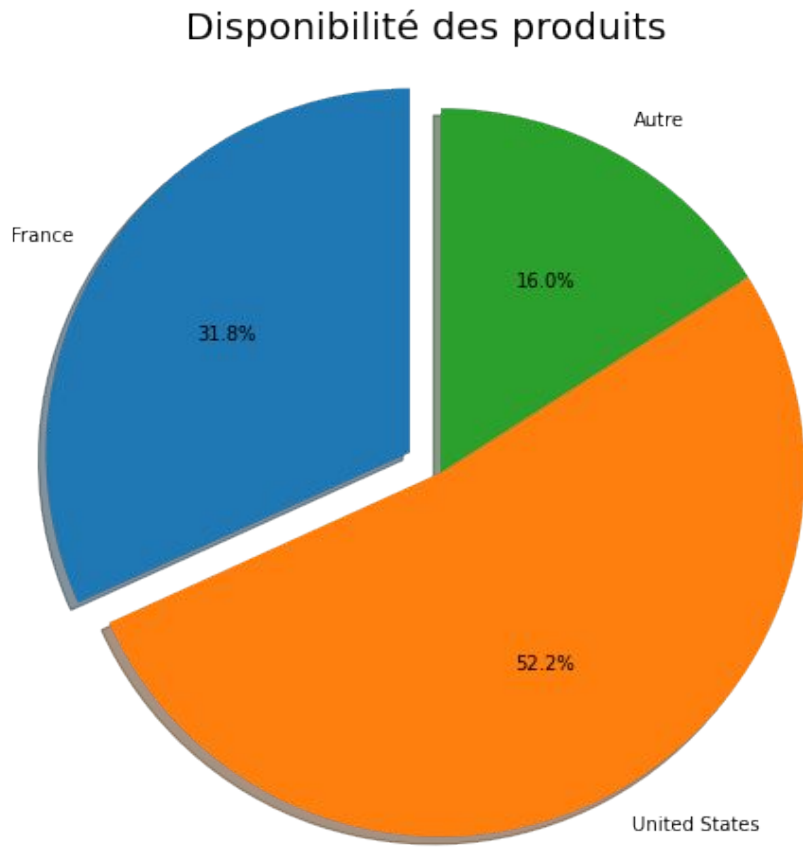
Sommaire

1. Analyses univariées
2. Analyses bivariées
3. Analyses multivariées
4. Idée d'application

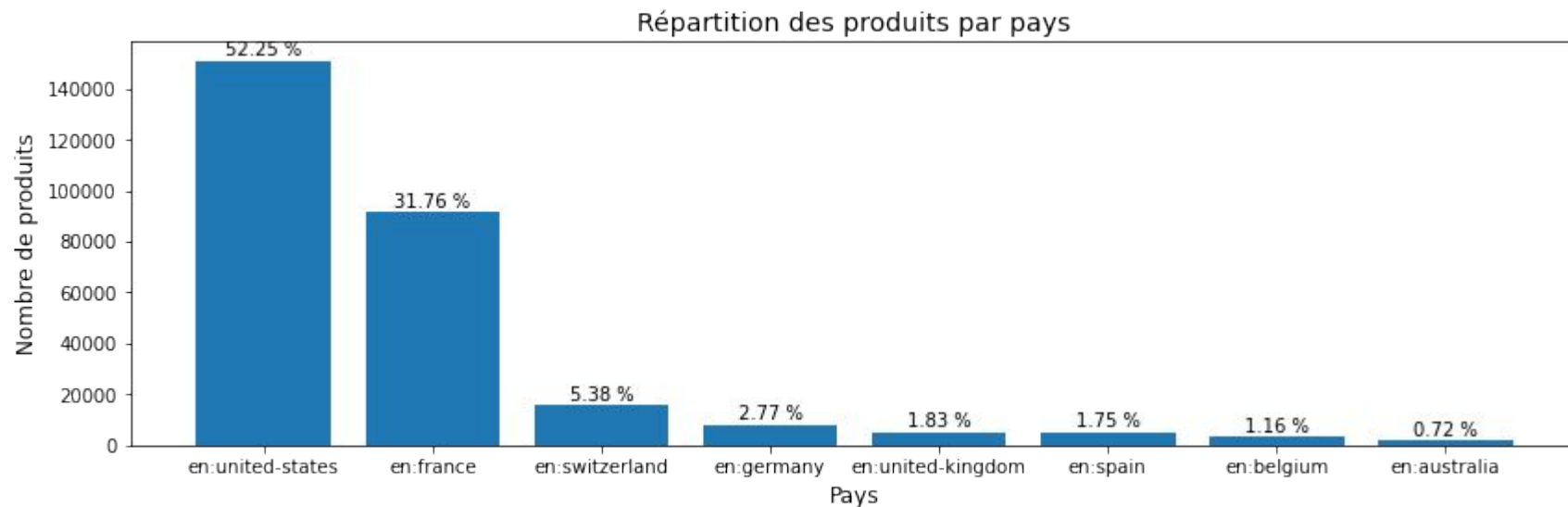
1. Analyses univariées

Disponibilité des produits par pays

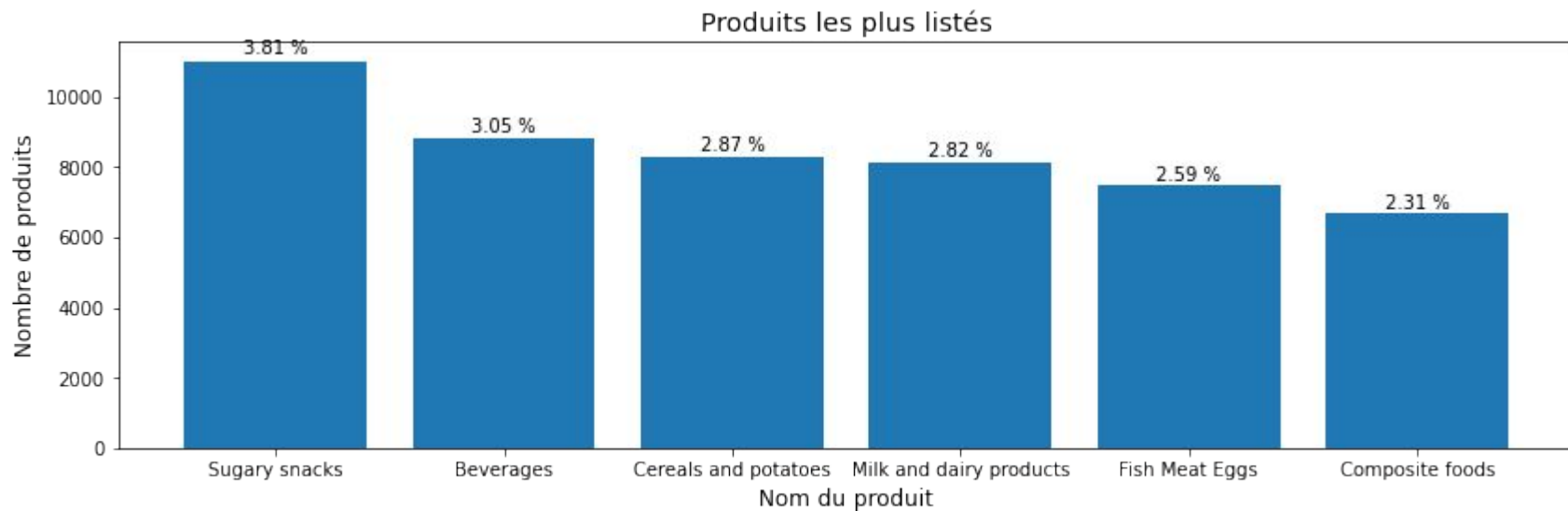
- 84% des produits sont disponibles en France ou aux États-Unis.



Disponibilité des produits par pays



Produits les plus listés

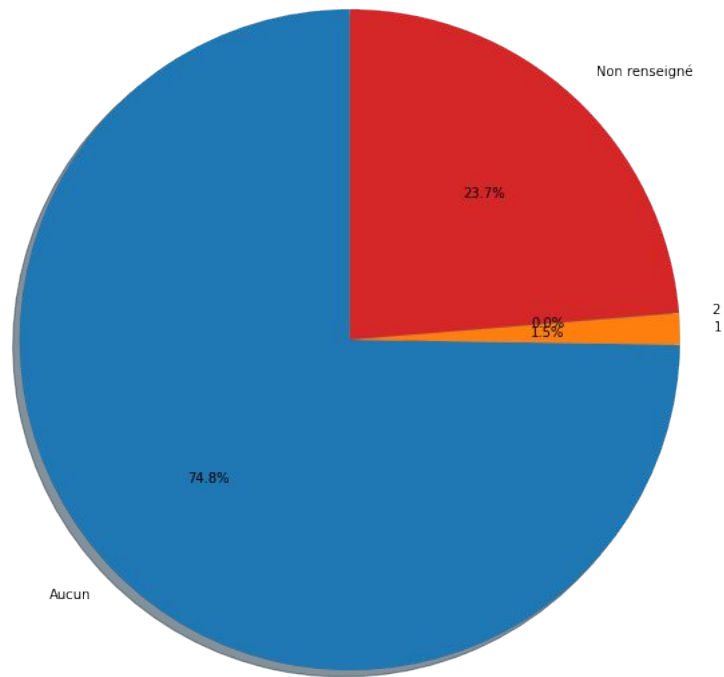


- Le type de produit le plus listé est *sucreries*, représentant 3,81 % des produits.

Nombre d'ingrédients provenant d'huile de palme

- 75% des produits ne contiennent aucun ingrédient provenant d'huile de palme
- 24% des produits n'ont pas cette information renseignée

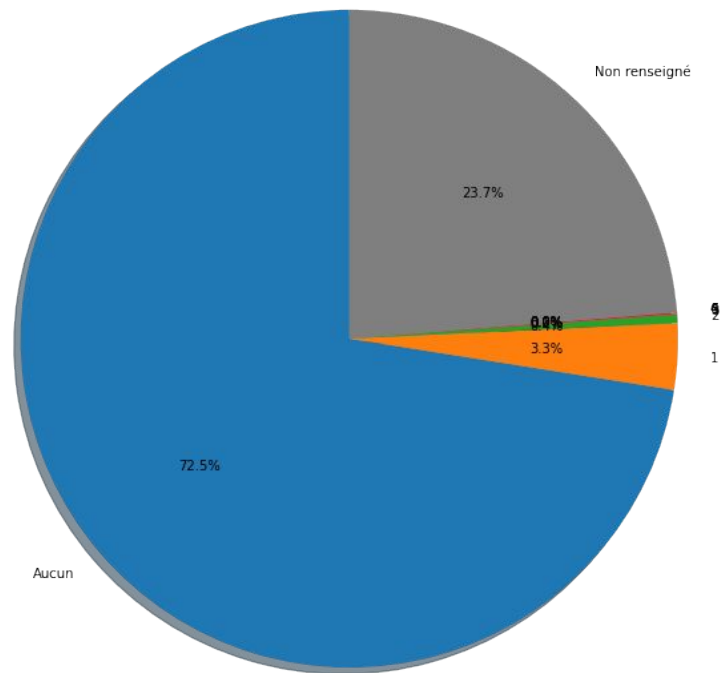
Nombre d'ingrédients provenant d'huile de palme



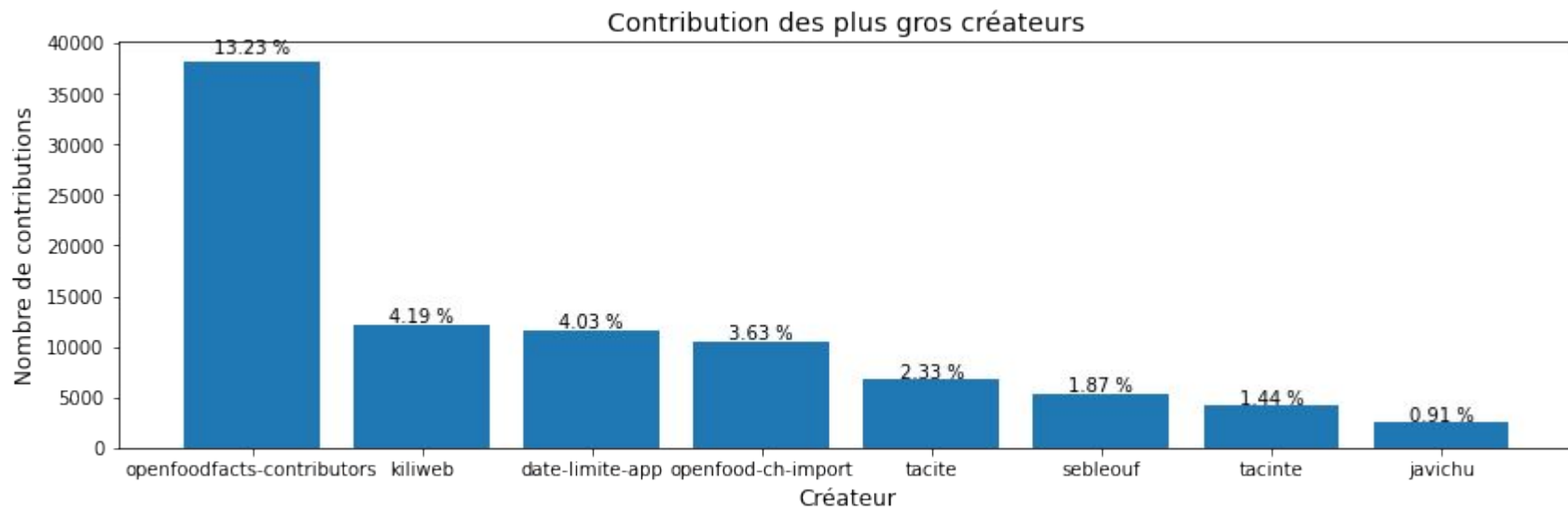
Nombre d'ingrédients pouvant provenir d'huile de palme

- 72,5% des produits ne contiennent aucun ingrédient provenant d'huile de palme
- 3,3% des produits pourraient en provenir, soit le double de ceux dont on est sûr

Nombre d'ingrédients pouvant provenir d'huile de palme



Contribution des plus gros créateurs

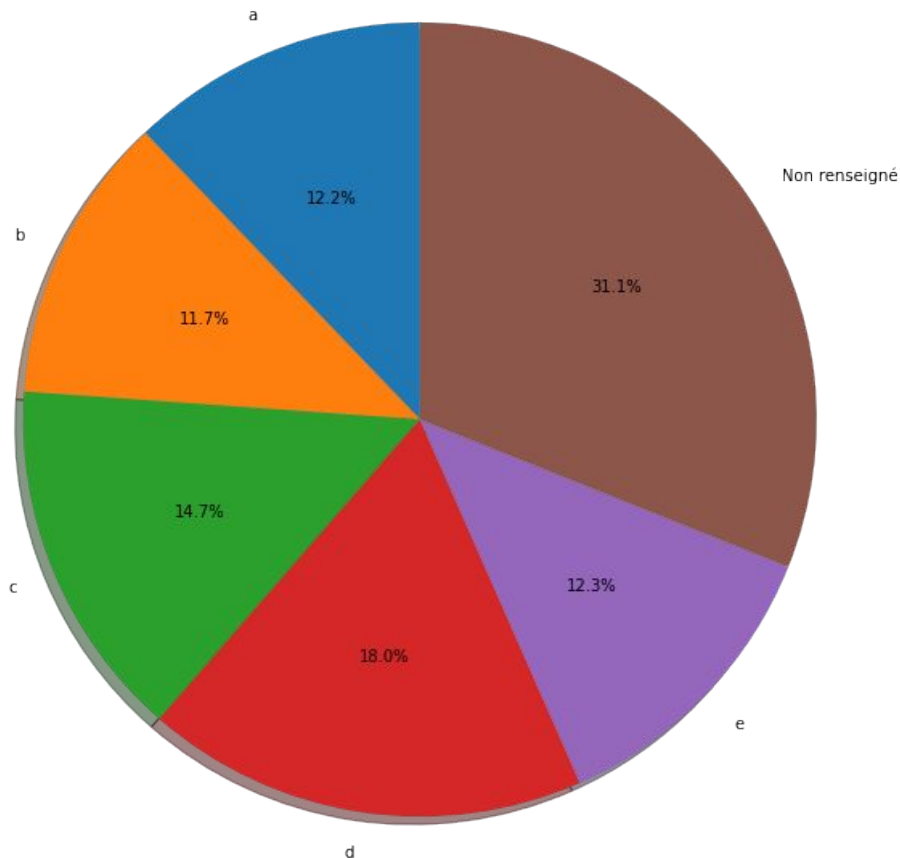


- Le créateur ayant le plus contribué à la database est *openfoodfacts-contributors* avec 13,23 % des produits, suivi de *kiliweb* avec 4,19 % des produits créés.

Répartition des produits par nutrition grade

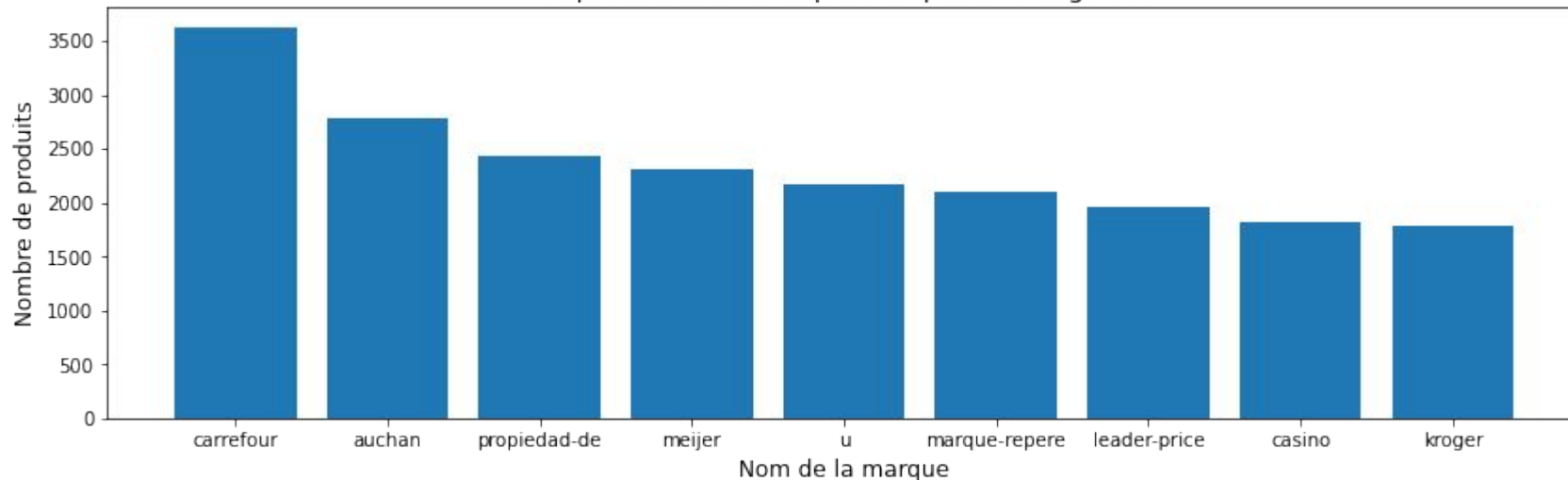
- Répartition relativement égale de la variable *nutrition grade* entre les produits

Répartition des produits par nutrition grade



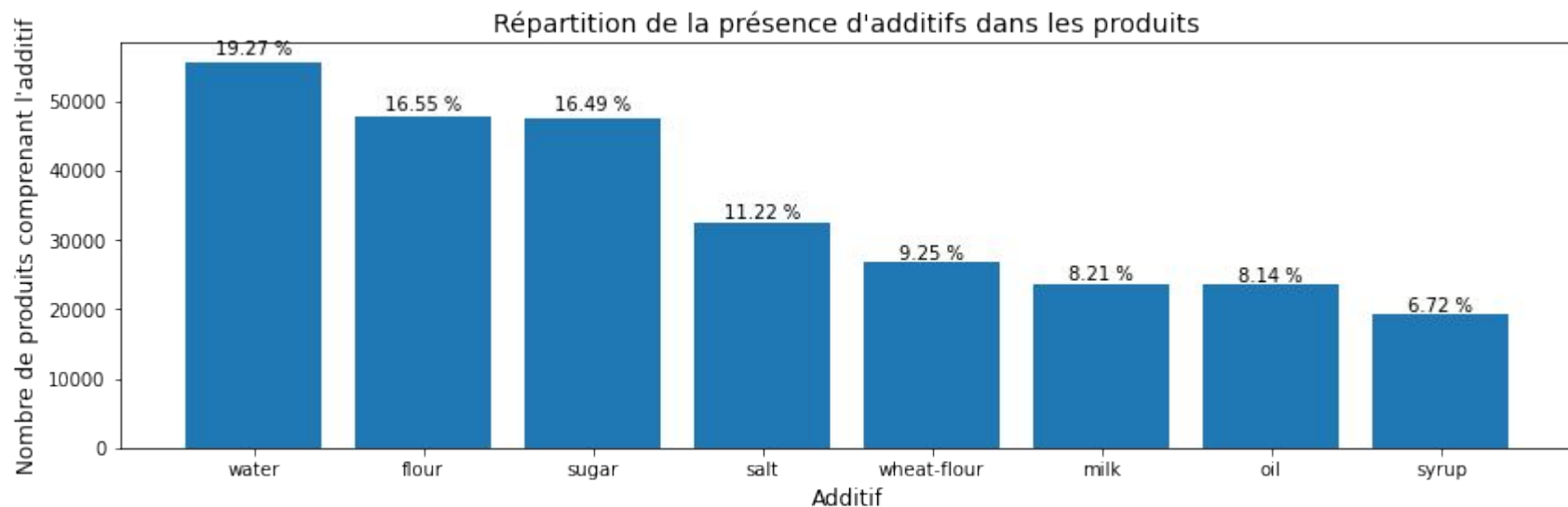
Marques les plus présentes

Répartition des marques les plus renseignées



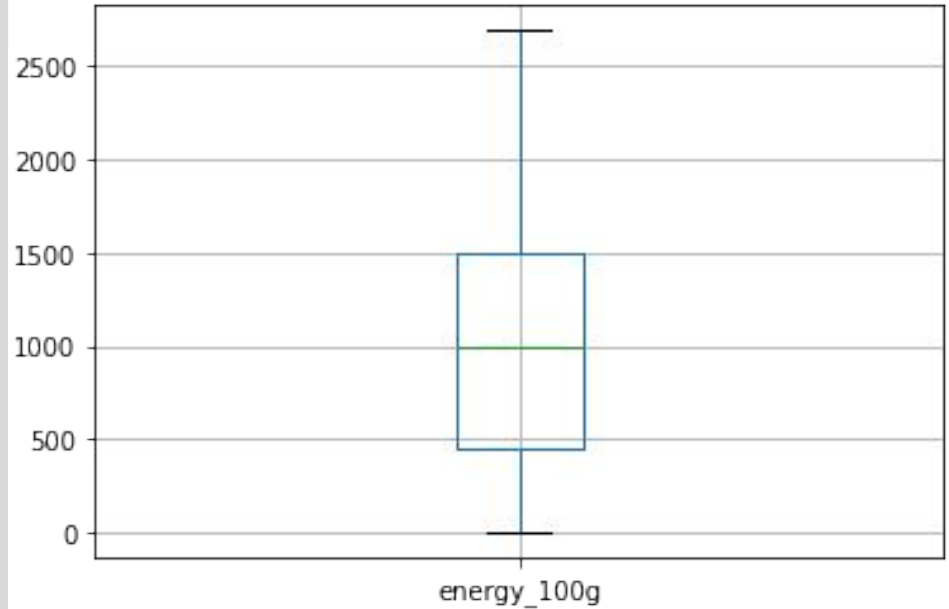
- Les deux marques les plus présentes sont *Carrefour* et *Auchan*.

Répartition des **additifs** dans les produits

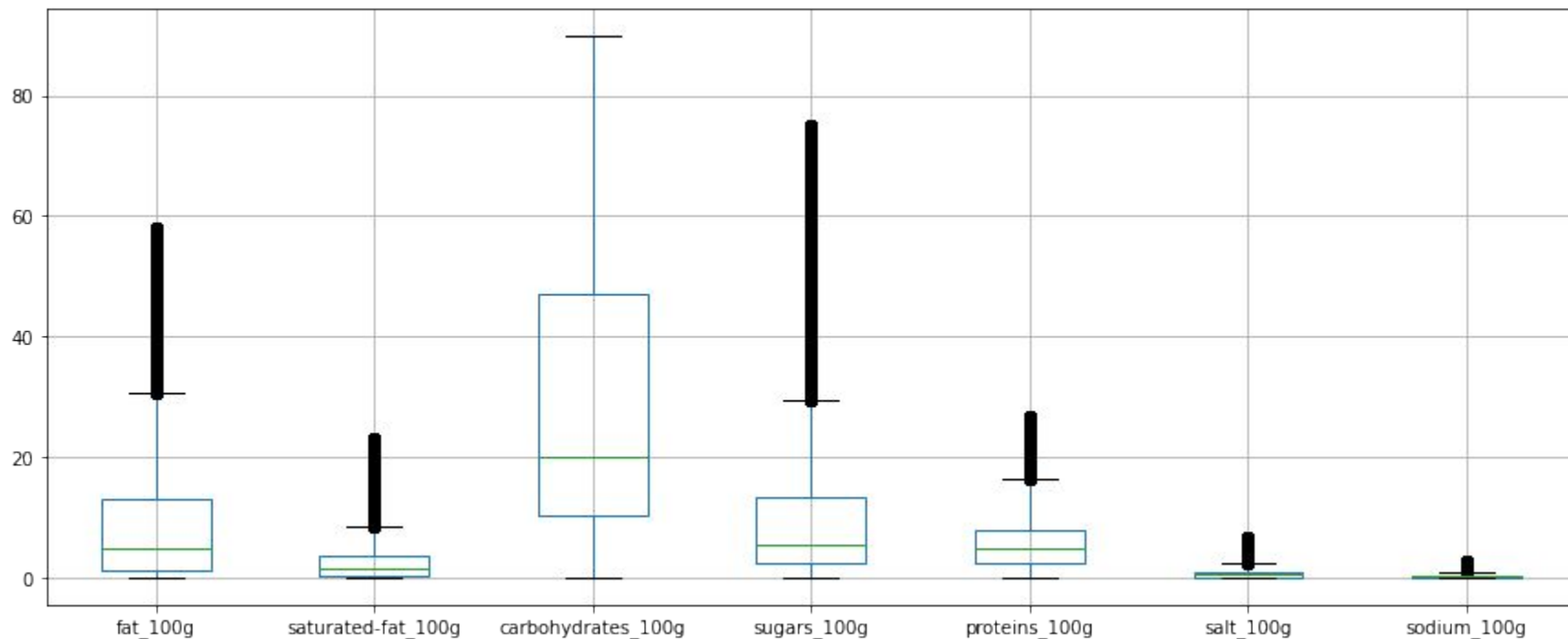


- 3 additifs sont présents dans au moins 15 % des produits : l'eau, la farine et le sucre.

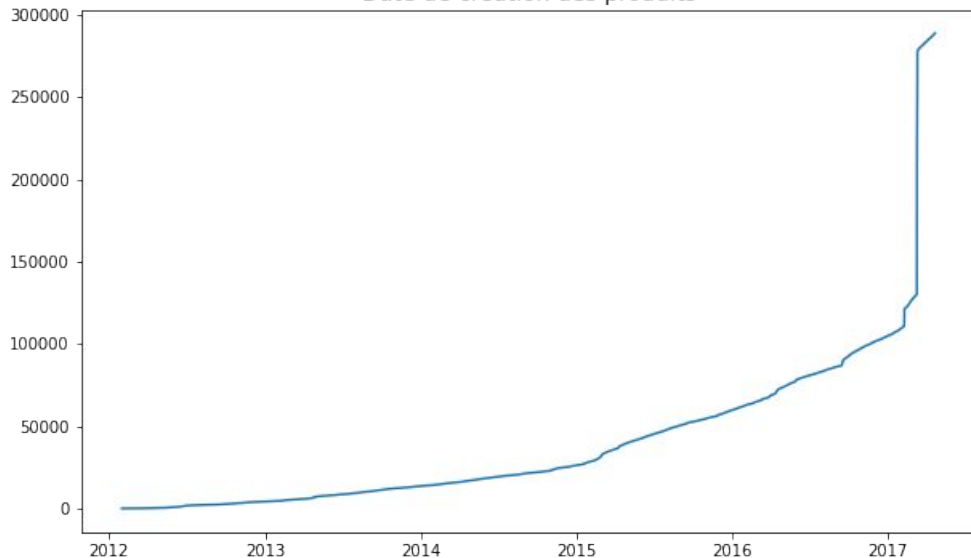
Boxplot : **Valeur
énergétique** par
100g (en kJ)



Boxplot : composants par 100g (en g)

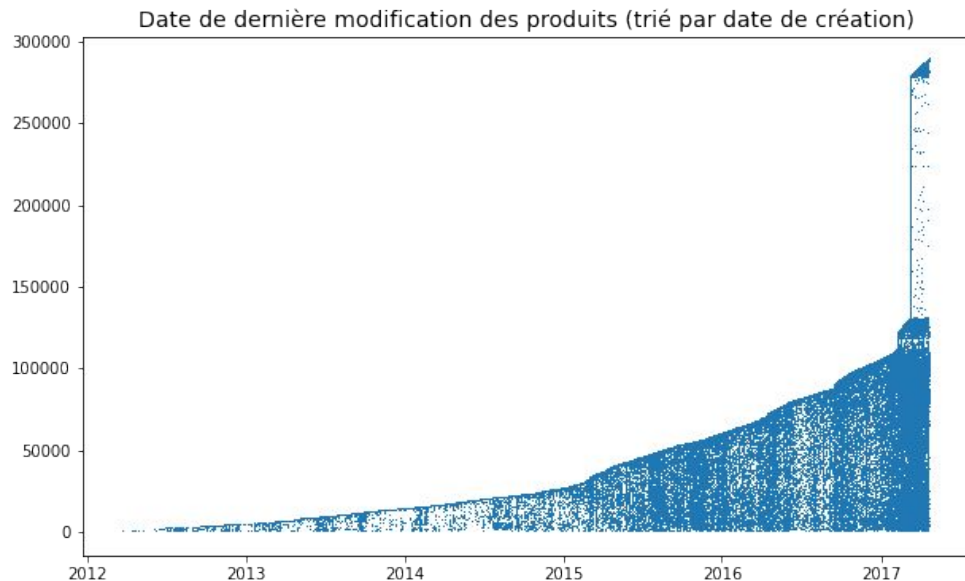


Date de création des produits



Date de création des produits sur la plateforme

- Augmentation soudaine du nombre de produits en 2017
- Avant 2017, les produits sont créés de manière stable dans le temps

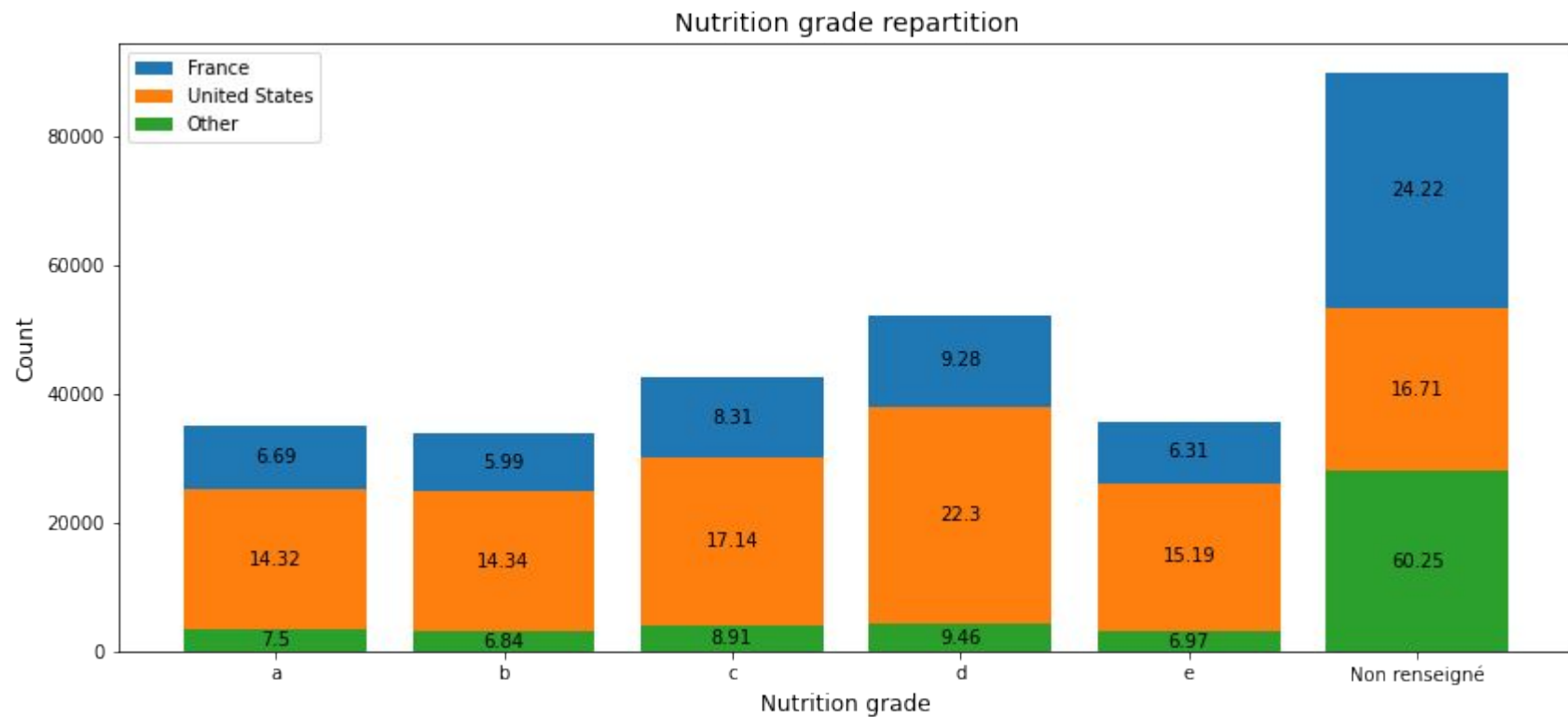


Dates de **dernière** **modification** des produits

- On remarque que même des produits créés il y a longtemps continuent d'être modifiés.

2. Analyses bivariées

Répartition de nutrition grade

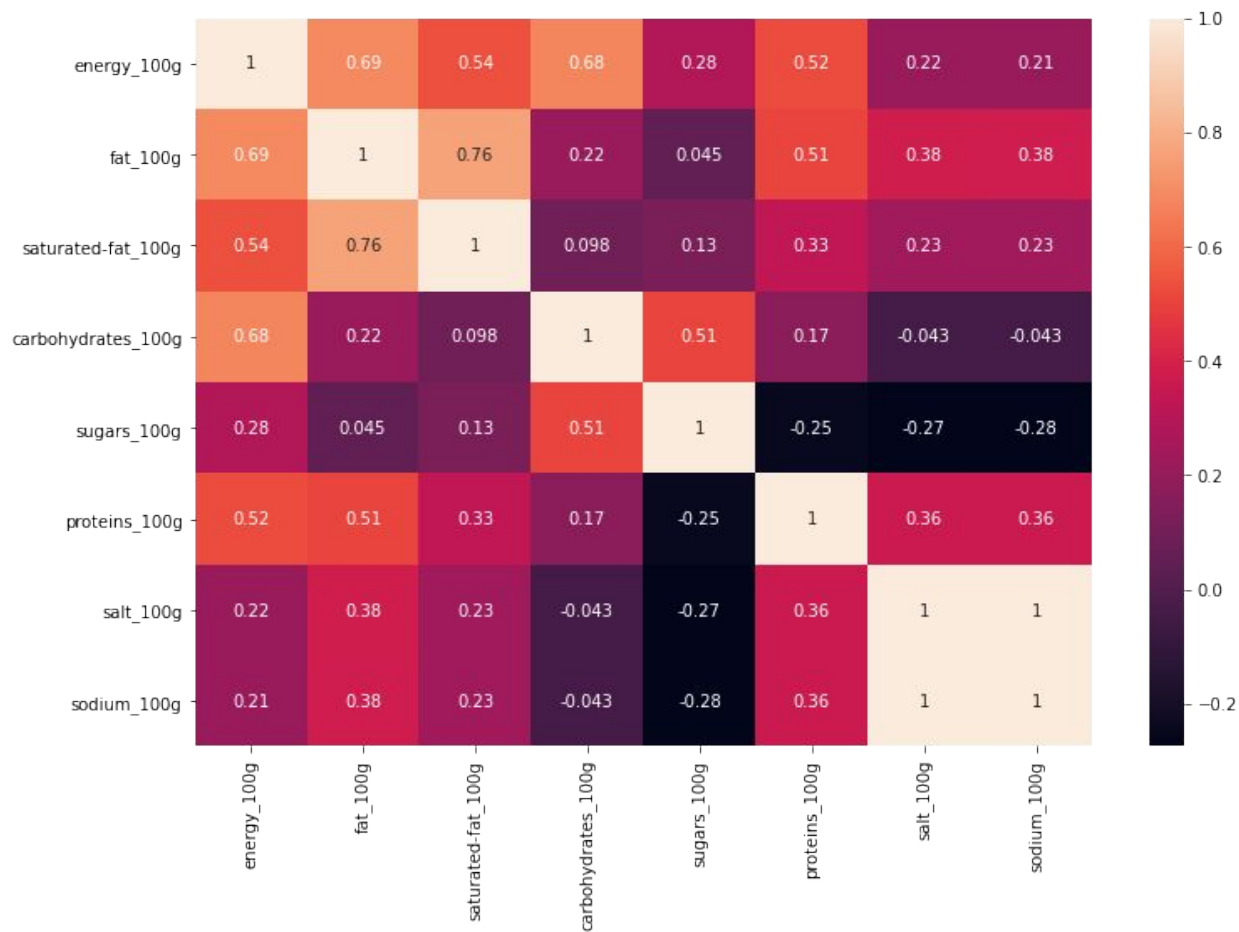


Matrice de corrélations

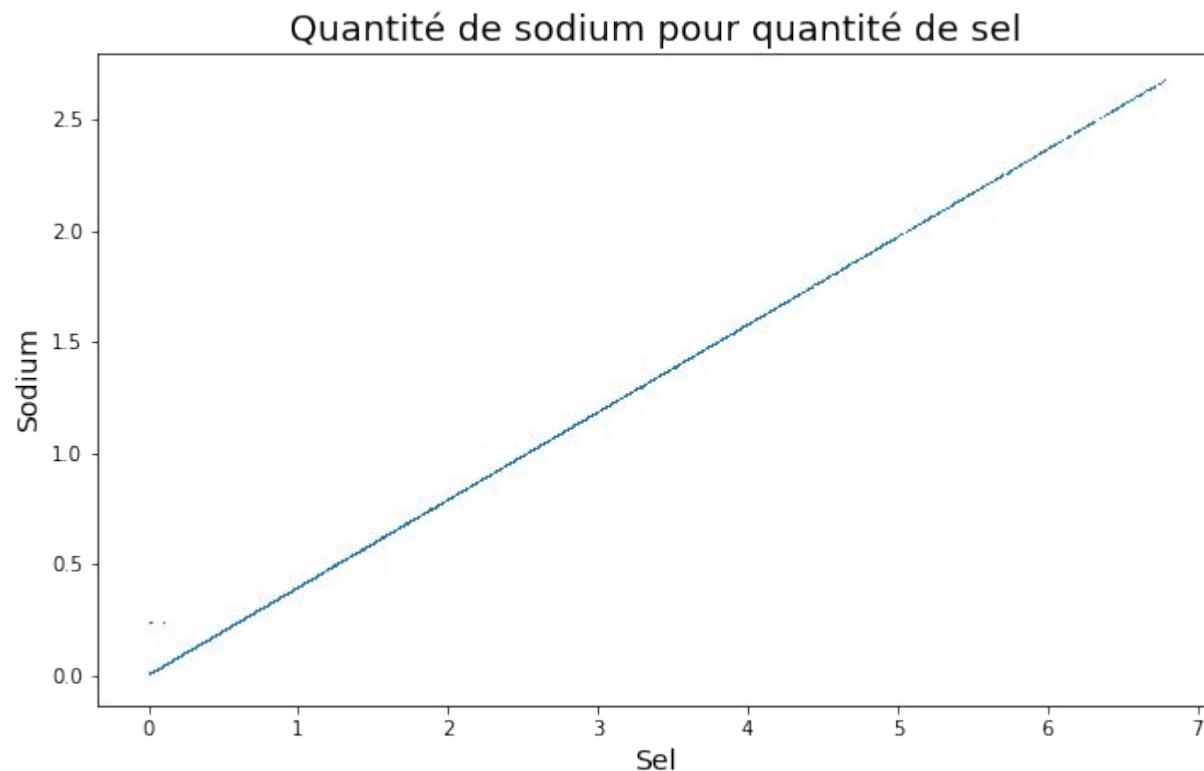
	energy_100g	fat_100g	saturated-fat_100g	carbohydrates_100g	sugars_100g	proteins_100g	salt_100g	sodium_100g
energy_100g	1.000000	0.689586	0.536699	0.678074	0.284490	0.524271	0.215135	0.214941
fat_100g	0.689586	1.000000	0.764492	0.216467	0.044966	0.506197	0.377444	0.377258
saturated-fat_100g	0.536699	0.764492	1.000000	0.098153	0.128336	0.331357	0.233353	0.233203
carbohydrates_100g	0.678074	0.216467	0.098153	1.000000	0.505356	0.169942	-0.042687	-0.043012
sugars_100g	0.284490	0.044966	0.128336	0.505356	1.000000	-0.248996	-0.274733	-0.275101
proteins_100g	0.524271	0.506197	0.331357	0.169942	-0.248996	1.000000	0.358487	0.358353
salt_100g	0.215135	0.377444	0.233353	-0.042687	-0.274733	0.358487	1.000000	0.999730
sodium_100g	0.214941	0.377258	0.233203	-0.043012	-0.275101	0.358353	0.999730	1.000000

- Sel et sodium très corrélés (1)
- Autres variables relativement corrélées :
 - Énergie et gras (0.69)
 - Énergie et glucides (0.68)
 - Gras et gras saturé (0.76)

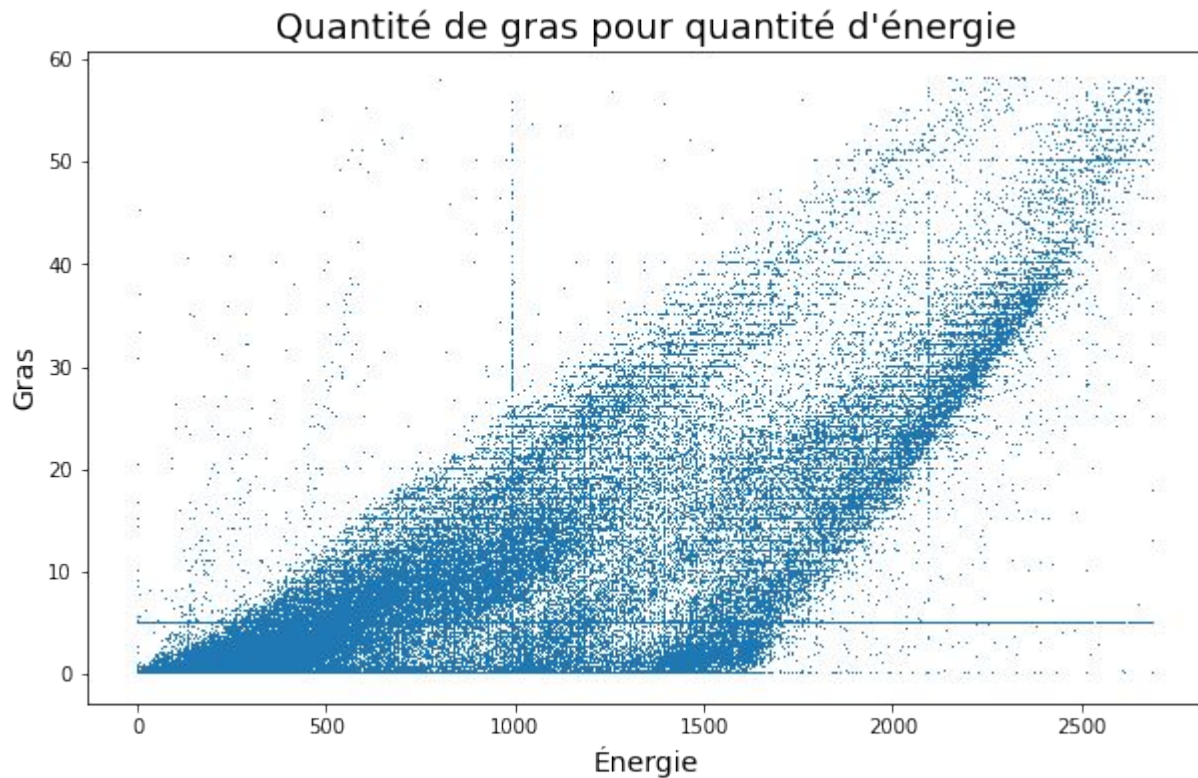
Heatmap des corrélations



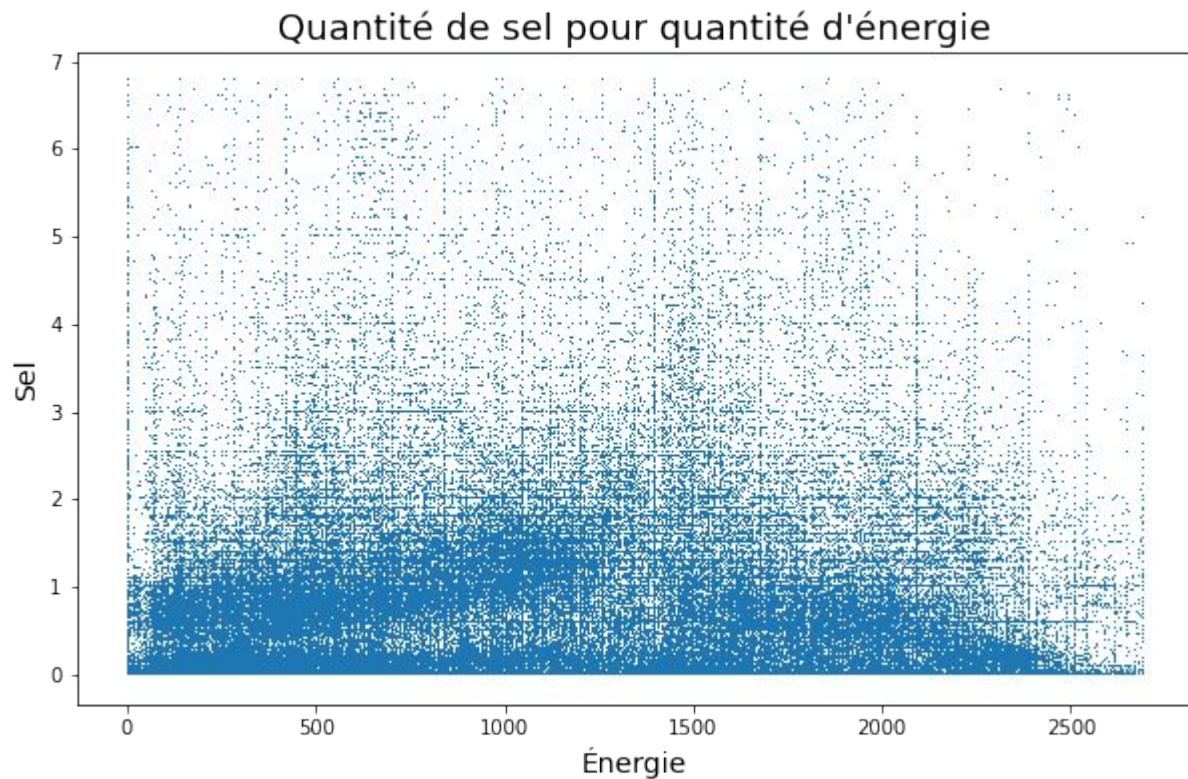
Relation entre **sel** et **sodium** dans les produits : 1



Relation entre **énergie** et **gras** dans les produits : **0.69**

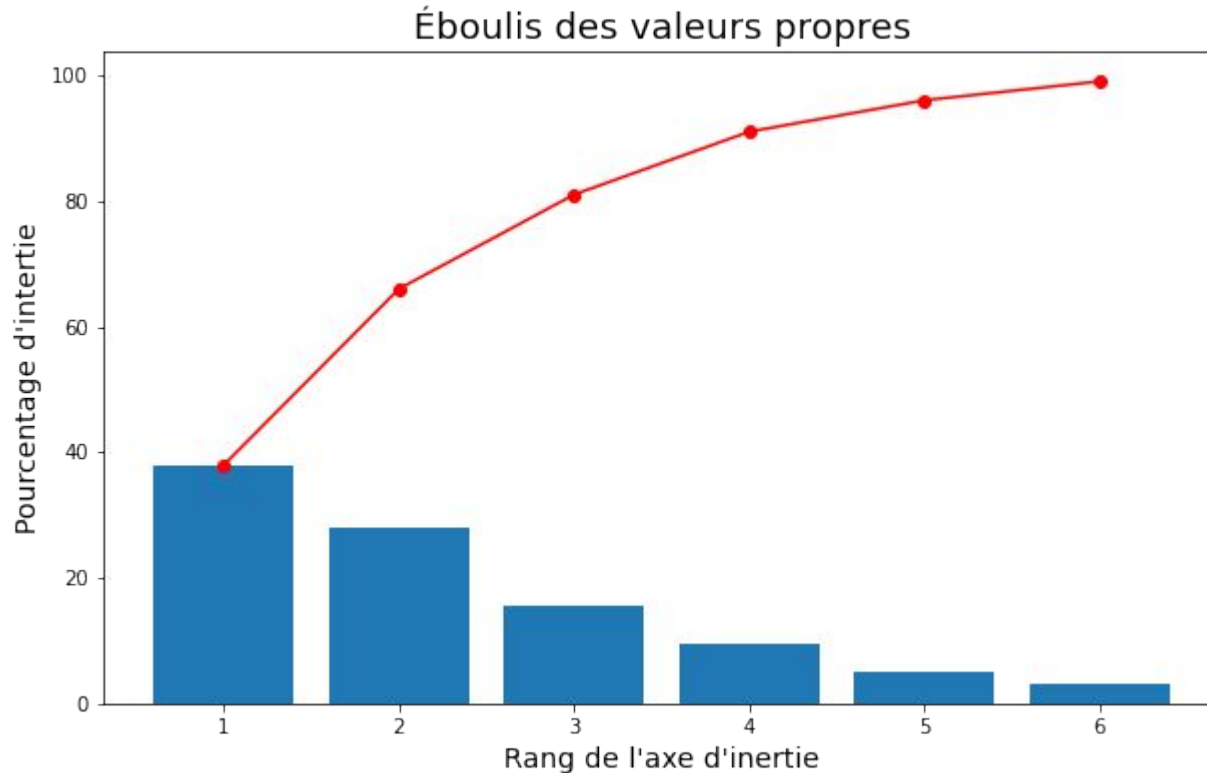


Relation entre **énergie** et **sel** dans les produits : **0.22**



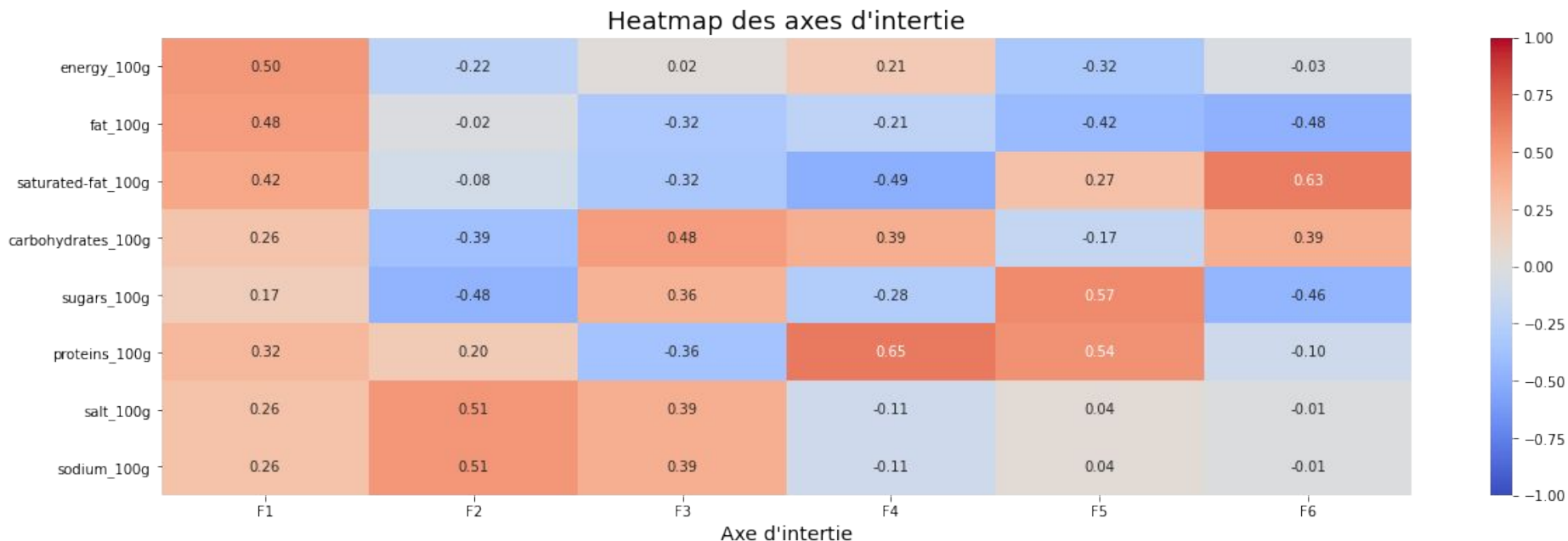
3. Analyses multivariées

ACP : Éboulis des valeurs propres

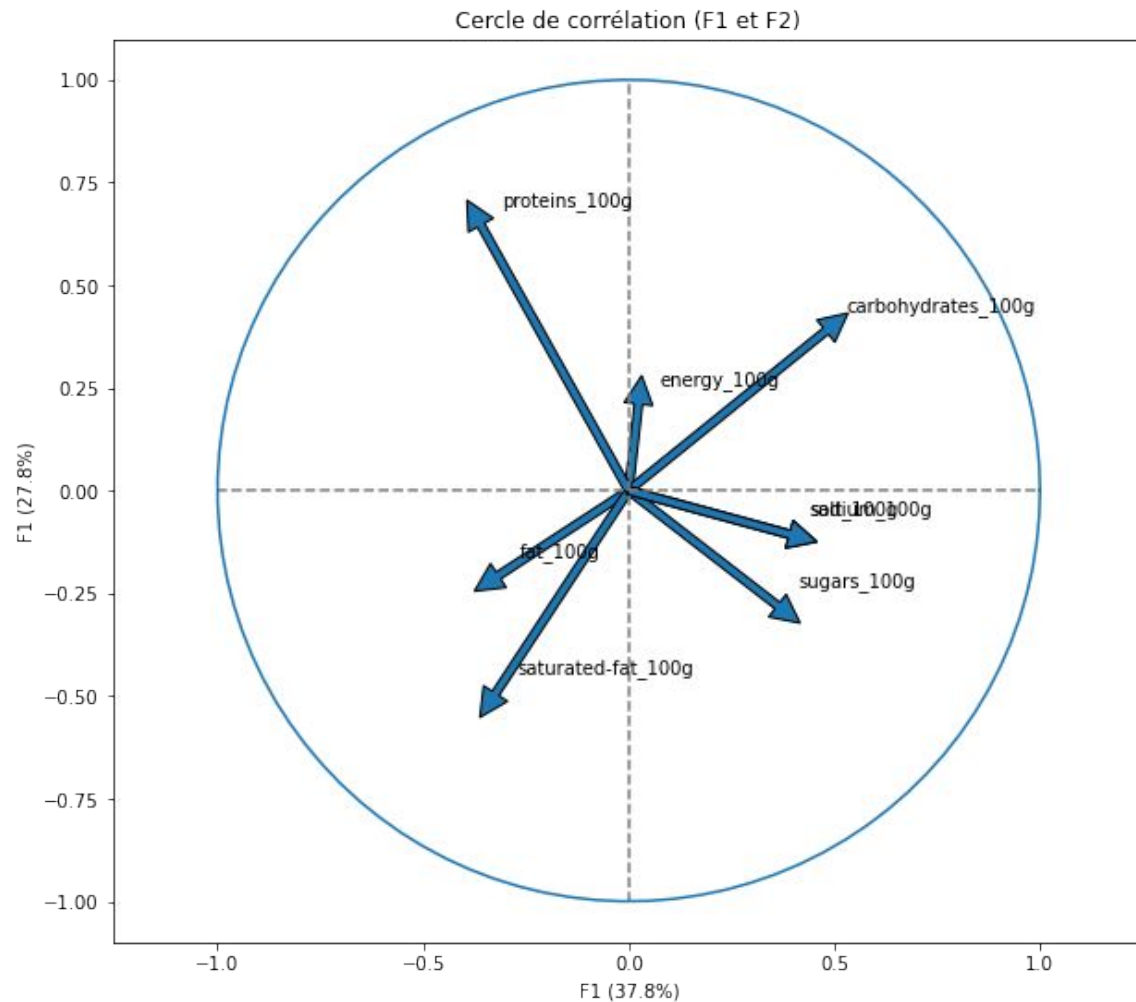


- Les axes F1 et F2 représentent **66%** des données,
- F1 à F4 : **91%**,
- F1 à F6 : **99%** des données.

ACP : Heatmap des axes d'inertie



ACP : Cercle de corrélation - F1 & F2



4. Idée d'application

L'utilisateur input le code ou l'url d'un produit.



L'algorithme recherche des produits du même *product_group*.



L'algorithme affiche ces produits, en priorisant les *nutrition_grade* proches de 'a'.

L'utilisateur peut aussi trier la liste des produits par n'importe quel additif, comme 'sel' ou 'gras', en fonction de son régime nutritionnel préféré.



