



Générateur de recettes saines  
pour le site Lamarmite



# Contexte et Travaux réalisés

- Contexte
  - Le site Lamarmite souhaite mettre en place un générateur de recettes saines à partir de la base de données OpenFoodFacts à destination d'une clientèle française.
  - Définition d'une alimentation saine.
- Travaux Réalisés
  - Analyse descriptive de la base de donnée
  - Sélection des variables
  - Nettoyage des données
  - Calcul d'indicateurs (Univariés et Multivariés)
  - Mise en perspective des résultats donnés par les indicateurs avec l'objectif poursuivi
  - Conclusion et Recommandations

# Contexte et Travaux Réalisés

## Définition du score nutritionnel

Nutri-Score - Score nutritionnel et notes de couleurs

### Un système de notes de A à E pour simplifier l'étiquetage nutritionnel

Ces notes de couleur NutriScore sont établies en calculant un score nutritionnel qui tient compte d'une part de l'énergie, des graisses saturées, des sucres, du sodium (les teneurs élevées sont considérées comme mauvaises pour la santé), et d'autre part de la proportion de fruits, légumes et noix, des fibres et des protéines (les teneurs élevées sont considérées comme bonnes pour la santé).

Seuils des notes Nutri-Score

Les seuils utilisés pour attribuer les notes de A à E sont :

Aliments solides :

A : jusqu'à -1

B : de 0 à 2

C : de 3 à 10

D : de 11 à 18

E : 19 et plus

Boissons :

Seule l'eau (minérale et de source, avec exclusion des eaux aromatisées) est classée A.

A : Eaux minérales et eaux de source

B : jusqu'à 1

C : de 2 à 5

D : de 6 à 9

E : 10 et plus

NUTRI-SCORE



NUTRI-SCORE



NUTRI-SCORE



NUTRI-SCORE



NUTRI-SCORE



# Contexte et Travaux Réalisés

## Définition du score nutritionnel

- **Avertissement : ce score peut encore ne pas correspondre au score préconisé :**
- La teneur en fibre n'est pas obligatoirement présente dans le tableau nutritionnel. Lorsque celle-ci n'est pas indiquée, l'éventuelle contribution positive des fibres n'est pas prise en compte. Nous encourageons tous les producteurs à indiquer la teneur en fibre de leurs produits sur l'emballage de ceux-ci.
- Les points correspondants aux teneurs en fruits et légumes ne sont pas pris en compte pour tous les produits. La teneur en fruits, légumes et n'est en effet pas indiquée sur l'emballage dans le tableau nutritionnel, et nous allons devoir la déterminer grâce à d'autres informations (comme les mentions "teneur minimum en fruits", la catégorie des produits (par exemple les confitures contiennent 50% de fruits), et/ou la liste des ingrédients.)
  - Un nouveau champ "Fruits, légumes (minimum)" a été créé pour noter les valeurs des mentions "teneur minimum en fruits/légumes" etc.
  - A noter : les tubercules comme les pommes de terres et patates douces ne sont pas considérés comme des légumes pour le calcul du score nutritionnel.





# Contexte et Travaux réalisés

- Qu'est ce qu'une alimentation saine ?
  - Une alimentation saine est une alimentation variée et équilibrée. Elle contient tous les éléments nécessaires, sans exagération. Le régime méditerranéen est l'exemple type d'une alimentation saine:
    - il se compose essentiellement d'aliments d'origine végétale (fruits, légumes, légumineuses, fruits secs, pain, pâtes);
    - l'huile d'olive y est la principale matière grasse;
    - viande, produits laitiers, œufs et vin y sont consommés avec modération.
  - La graisse consommée doit se limiter à 30-35% de l'apport total d'énergie. Les graisses saturées doivent être inférieures à 10% et les acides gras trans (= graisses végétales durcies artificiellement, utilisées dans l'industrie alimentaire) inférieurs à 1%.



# Contexte et Travaux réalisés

- Qu'est ce qu'une alimentation saine ?
- **Acides gras saturés et insaturés**
  - Les acides gras sont les éléments qui composent les graisses. Les graisses saturées se composent d'acides gras saturés; elles sont dures. Les graisses saturées d'origine animale contiennent de grandes quantités de cholestérol. Ce sont donc de mauvaises graisses.
  - Les "bonnes" graisses contiennent proportionnellement plus d'acides gras insaturés et moins d'acides gras saturés; elles sont plus molles et plus liquides.



# Contexte et Travaux réalisés

- Qu'est ce qu'une alimentation saine ?
- **Acides gras insaturés, oméga-3 et 6**
  - Les acides gras insaturés sont considérés comme faisant partie des bonnes graisses, pauvres en cholestérol.
  - On distingue les acides gras monoinsaturés (par exemple l'huile d'olive)
  - et les acides gras polyinsaturés dont font notamment partie les acides gras oméga-3 et oméga-6.





# Base de données OpenFoodFacts

- Analyse descriptive a priori
  - 162 colonnes (variables)
  - 320790 lignes (données)
  - Données nutritionnelles (macro- et micro-nutrinements), techniques (pays de vente), classification (catégories)
  - Beaucoup de données manquantes ou aberrantes
  - Version de travail fournie assez ancienne
  - Versions plus récentes disponible corrigent beaucoup d'incohérences et de données manquantes



## Nettoyage des données et sélection des variables

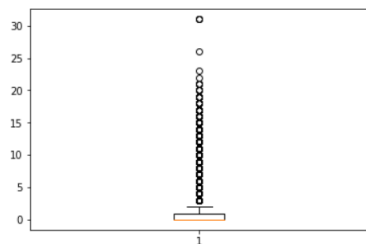
- Conservation des produits vendus en France
- Les produits n'ayant pas de noms (donc pas identifiables) ont été supprimés
- Calcul du pourcentage de données manquantes
- Choix du critère de conservation des données : si la variable possède moins de 60% de données manquantes, celle-ci est conservée
- Suppression des valeurs supérieures à 100 car il ne peut avoir plus de 100g d'un certain nutriment dans une valeur rapportée pour 100g (variables finissant par \_100g)
- Pour les données numériques (float) , les données manquantes ont été remplacées par des 0
- Pour les données alphanumériques (float) , les données manquantes ont été remplacées par des '.' (point)
- Les valeurs aberrantes ont été traitées manuellement au cas par cas (avec l'aide des visualisations)
- Les données une fois nettoyées et sélectionnées , il reste 24 colonnes (variables) et 91235 lignes (données)
- Parmi les données sélectionnées avec le critère précédent la liste des variables ci-dessous a été conservée :  
`"code","product_name","labels_fr","ingredients_text","serving_size","categories_fr","main_category_fr","additives_n","additives_fr","energy_100g","proteins_100g","carbohydrates_100g","sugars_100g","fat_100g","saturated-fat_100g","trans-fat_100g","cholesterol_100g","fiber_100g","sodium_100g","vitamin-a_100g","calcium_100g","iron_100g","nutrition_grade_fr","nutrition-score-fr_100g"`



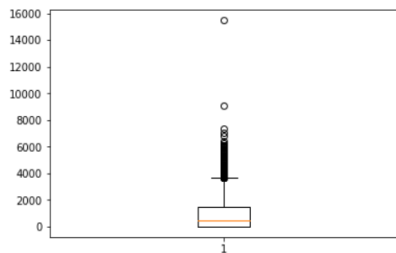
# Analyse Univariée

# Base de données OpenFoodFacts

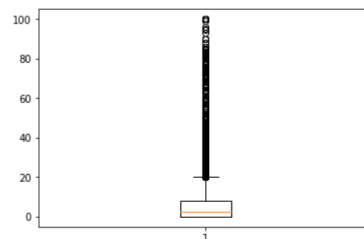
- Analyse Univariée



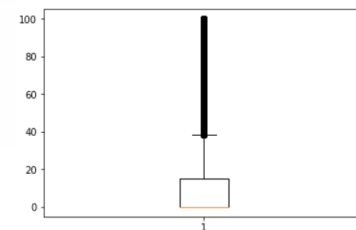
Nombre (en g) d'additifs  
dans un produit  
Moyenne 1.09



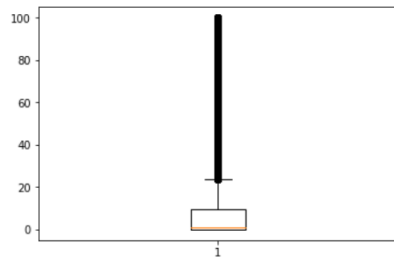
Quantité (en g) d'Énergie  
pour 100g de Produit  
Moyenne 785.82



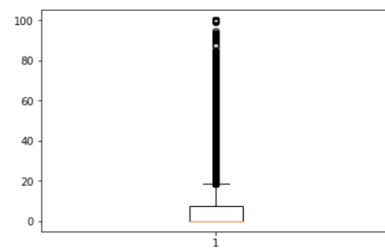
Quantité (en g) de protéines  
pour 100g de Produit  
Moyenne 5.44



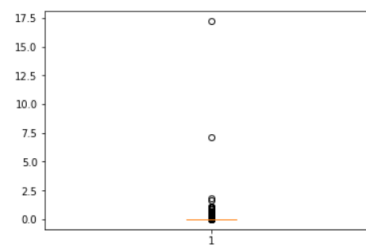
Quantité (en g) de carbohydrates  
(glucides) pour 100g de Produit  
Moyenne 14.23



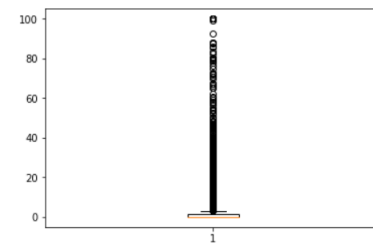
Quantité (en g) de sucres  
pour 100g de Produit  
Moyenne 9.14



Quantité (en g) de graisses  
pour 100g de Produit  
Moyenne 6.90



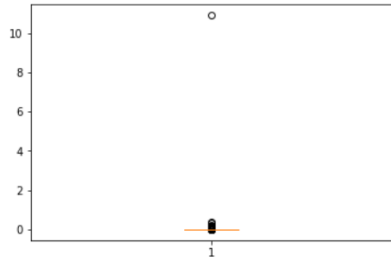
Quantité (en g) d'acides gras  
trans pour 100g de Produit  
Moyenne 8.85e-04



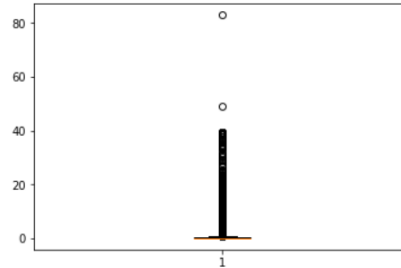
Quantité (en g) de fibres  
pour 100g de Produit  
Moyenne 1.27

# Base de données OpenFoodFacts

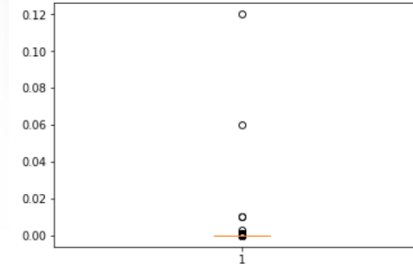
- Analyse Univariée



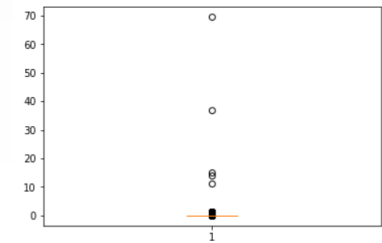
Quantité (en g) de cholestérol  
pour 100g de Produit  
Moyenne 1.81e-04



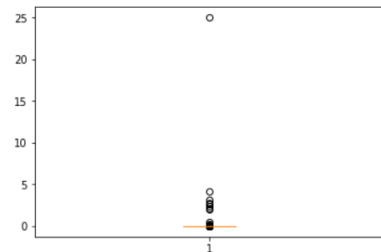
Quantité (en g) de sodium  
pour 100g de Produit  
Moyenne 0.31



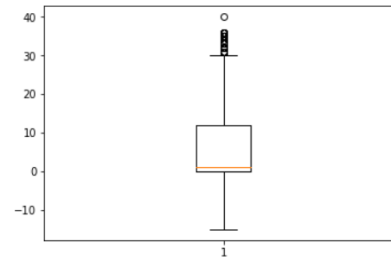
Quantité (en g) de vitamine A  
pour 100g de Produit  
Moyenne 3.88e-06



Quantité (en g) de calcium pour  
100g de Produit  
Moyenne 8.02e-03



Quantité (en g) de fer pour  
100g de Produit  
Moyenne 5.75e-04



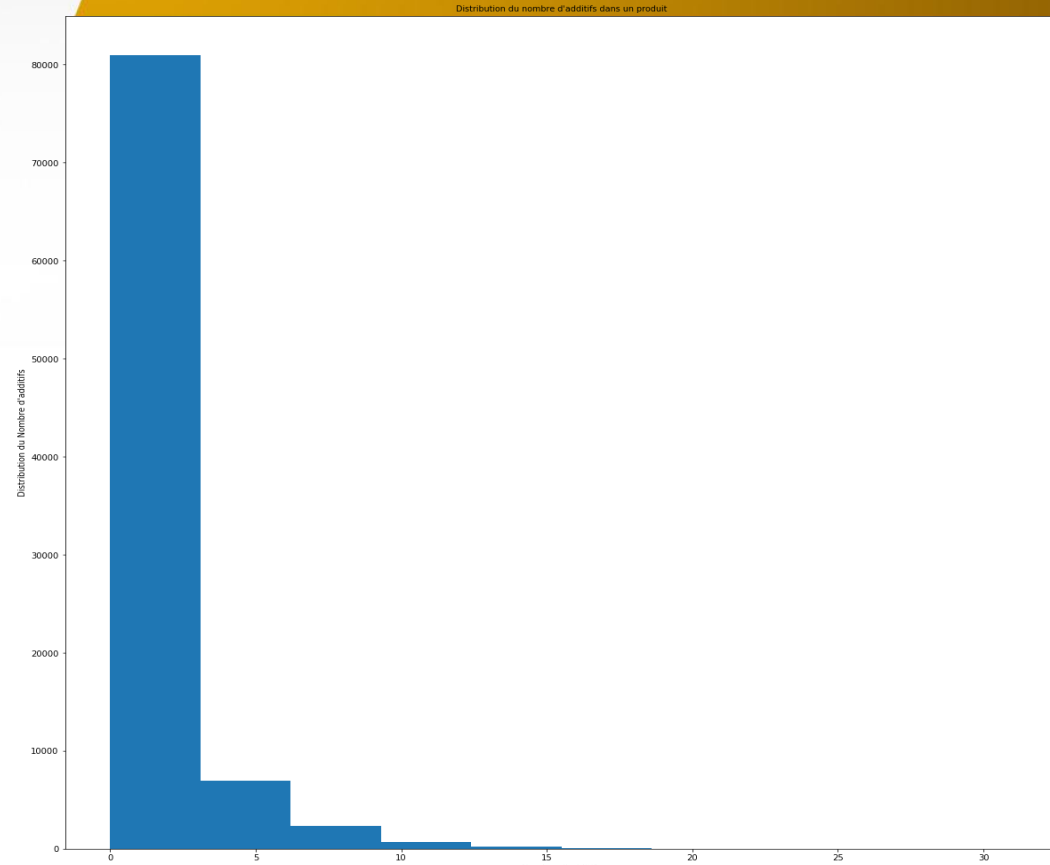
Score Nutritionnel (variant  
entre -15 et 40)  
Moyenne 5.81

A propos du score nutritionnel, celui-ci est calculé selon des modalités complexes, j'ai effectué une tentative de recalcul de ce score mais le nombre trop important de données manquantes à son recalcul ainsi que la différence avec les données entrée, ont fait que je n'ai pas persisté dans cette voie

# Base de données OpenFoodFacts

- Analyse Univariée

Distribution du nombre  
d'additifs dans un produit

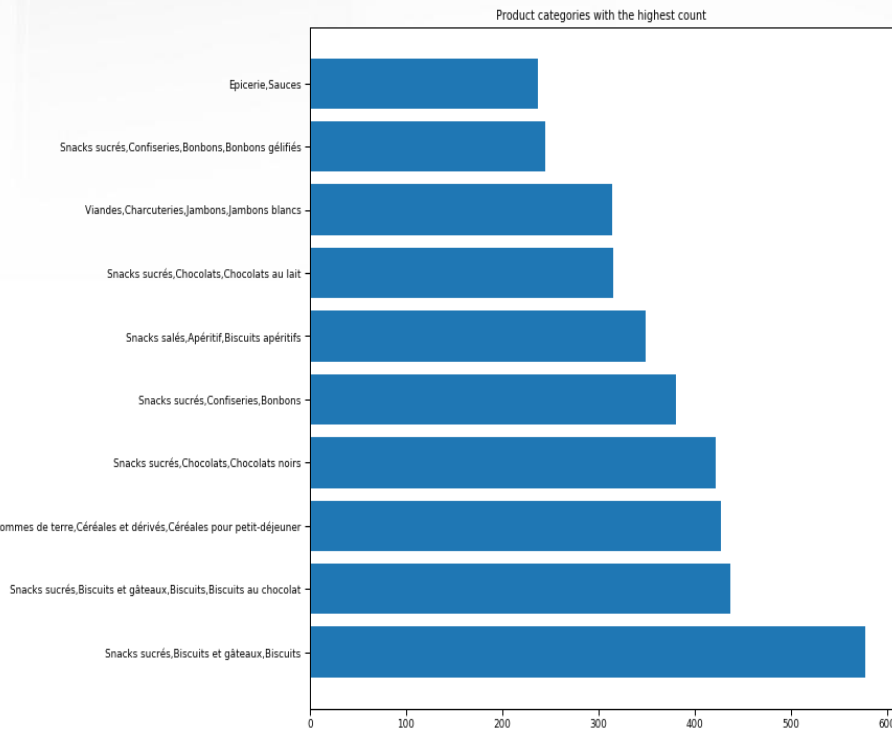


# Base de données OpenFoodFacts

- Analyse Univariée

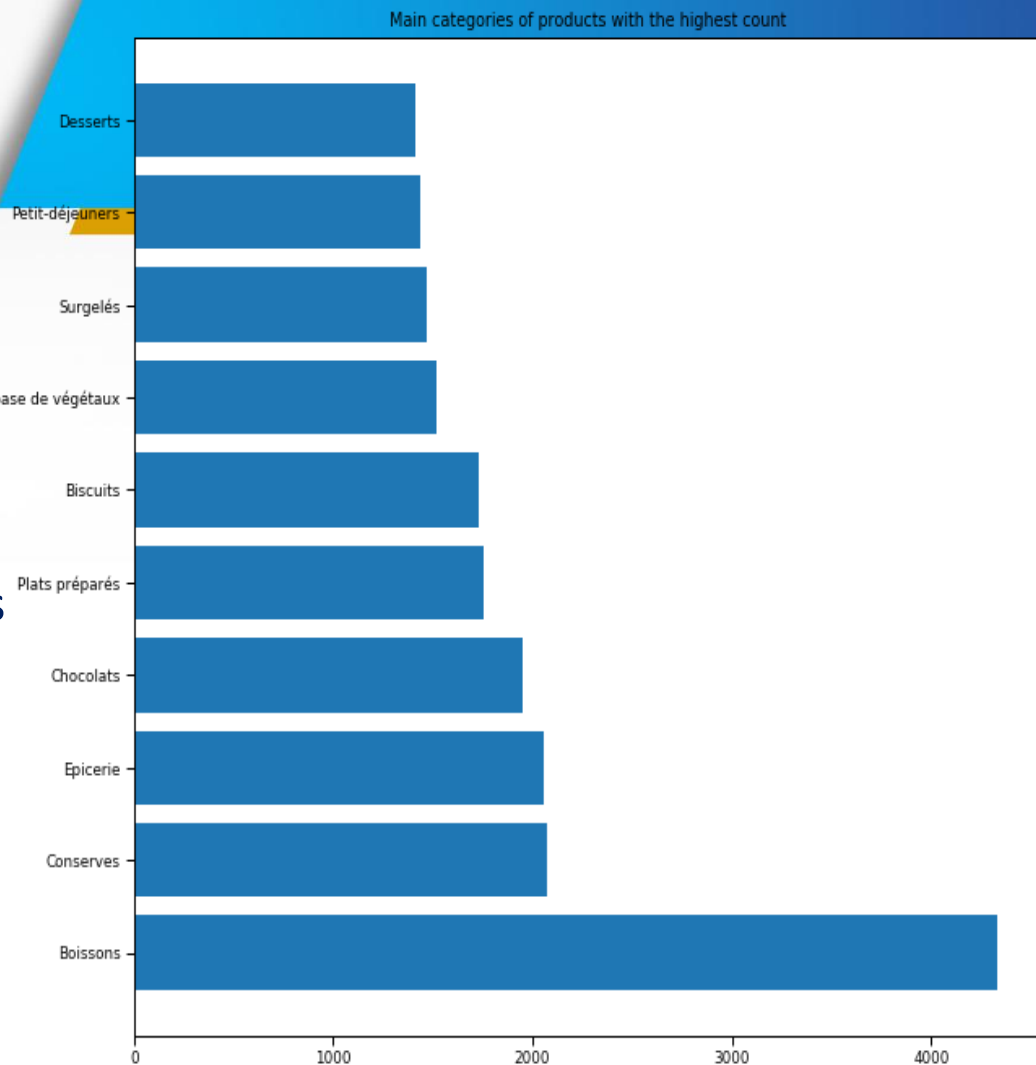
Catégories de produits ayant le plus grand nombre de produits (TOP 10) en partant du bas

Aliments et boissons à base de végétaux, Aliments d'origine végétale, Petit-déjeuners, Céréales et pommes de terre, Céréales et dérivés, Céréales pour petit-déjeuner



- Analyse Univariée

Catégories principales de produits  
Ayant le plus grand nombre de produits  
(TOP 10) en partant du bas

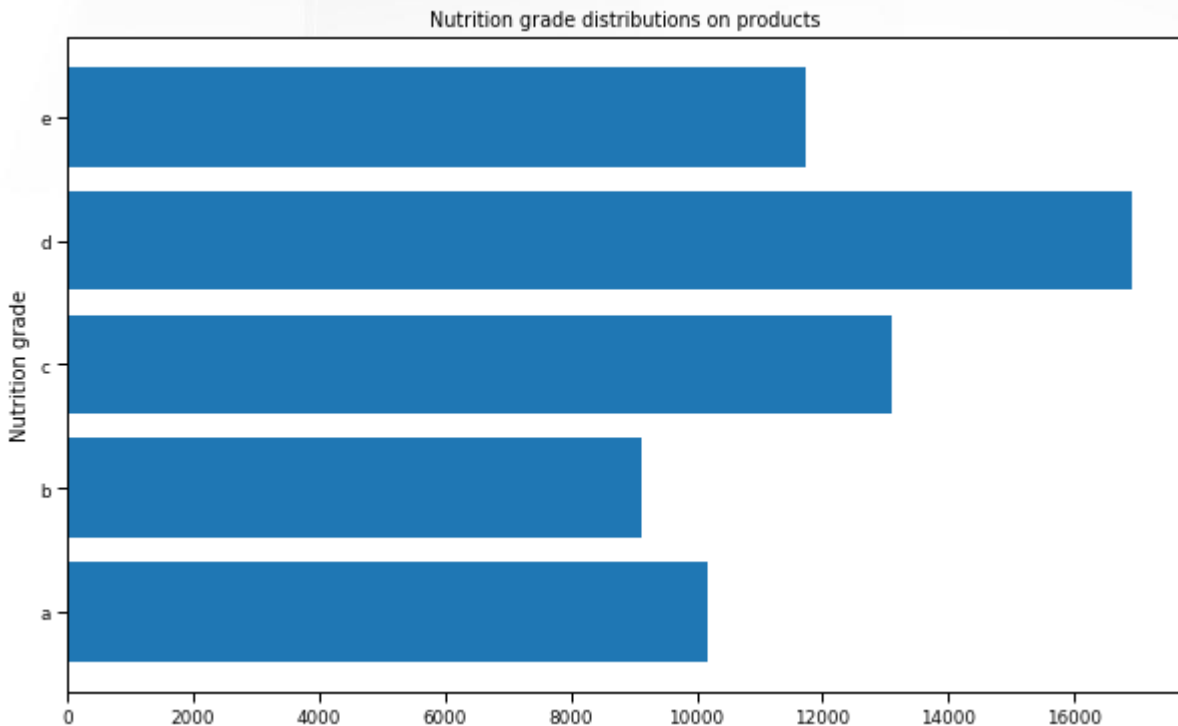




# Base de données OpenFoodFacts

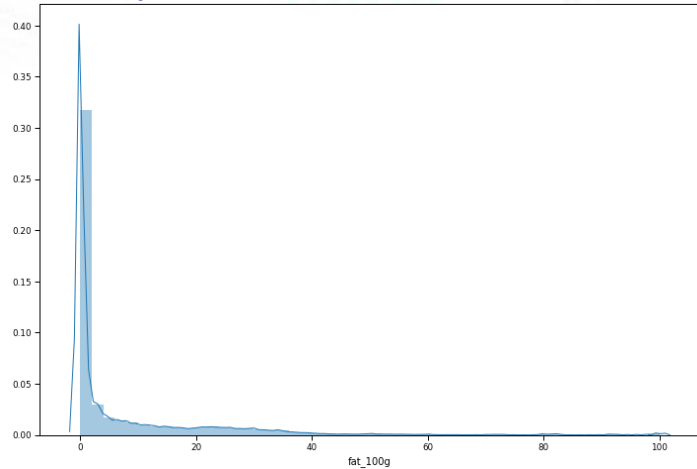
- Analyse Univariée

Le nombre de produits par note nutritionnelle

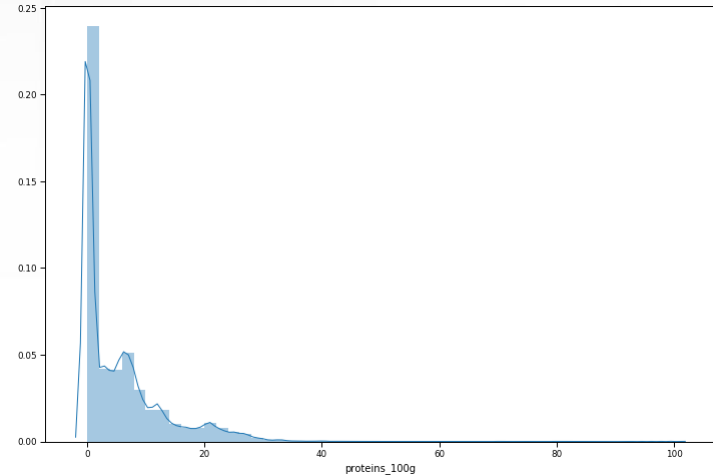


# Base de données OpenFoodFacts

## Analyse Univariée



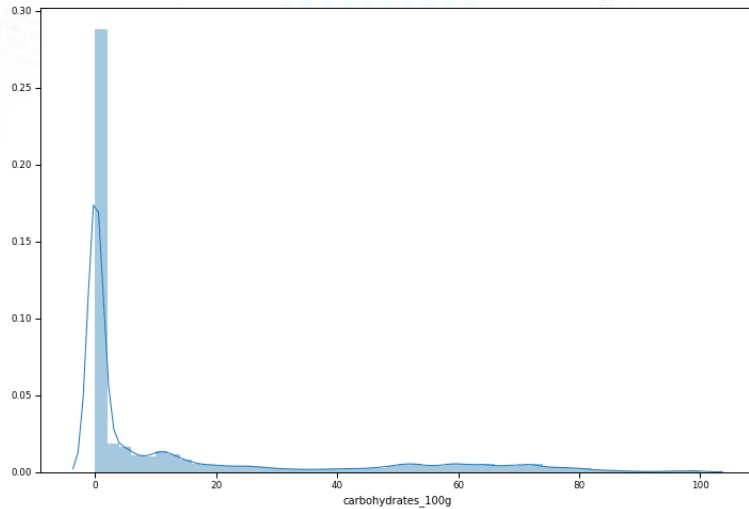
Histogramme de la quantité de graisses par produits pour 100g de produits



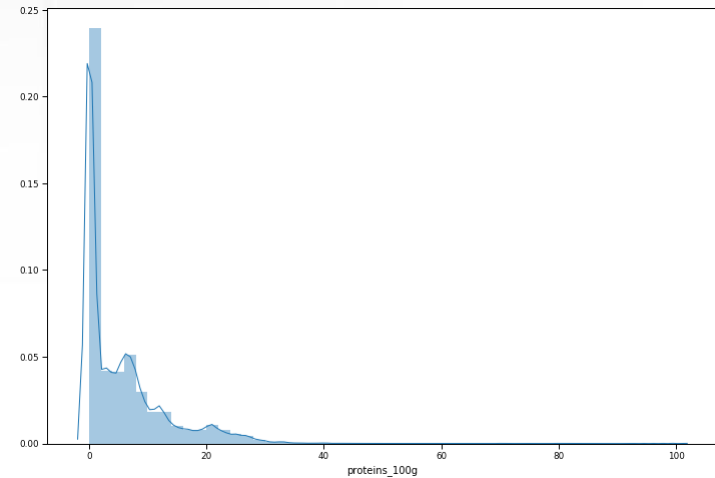
Histogramme de la quantité de protéines par produits pour 100g de produits

# Base de données OpenFoodFacts

## Analyse Univariée



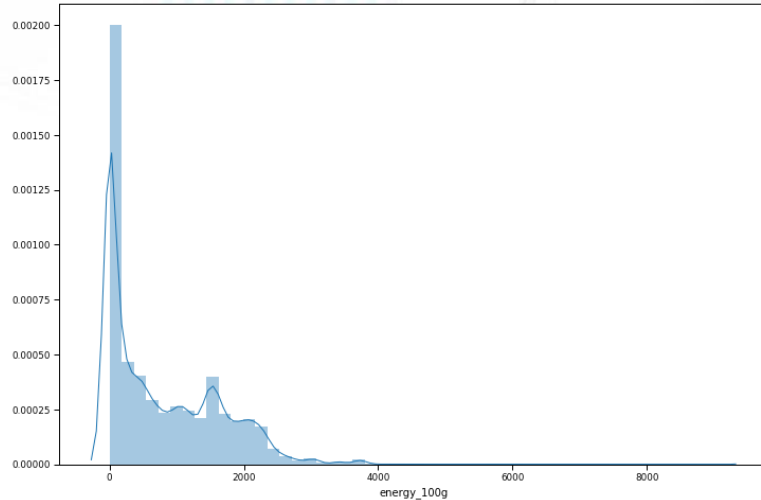
Histogramme de la quantité de glucides par produits pour 100g de produits



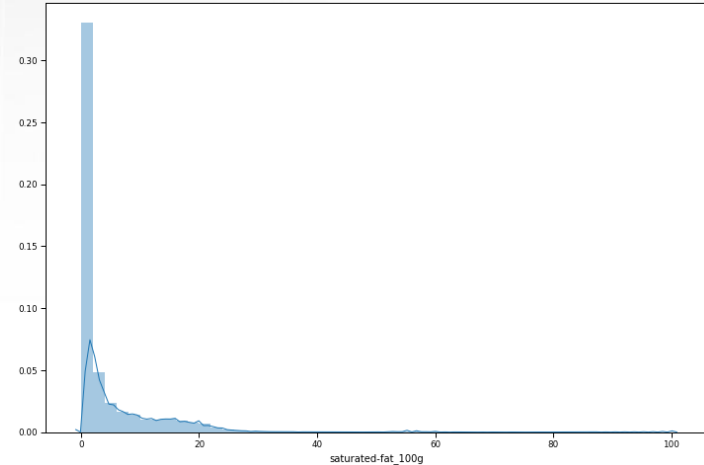
Histogramme de la quantité de protéines par produits pour 100g de produits

# Base de données OpenFoodFacts

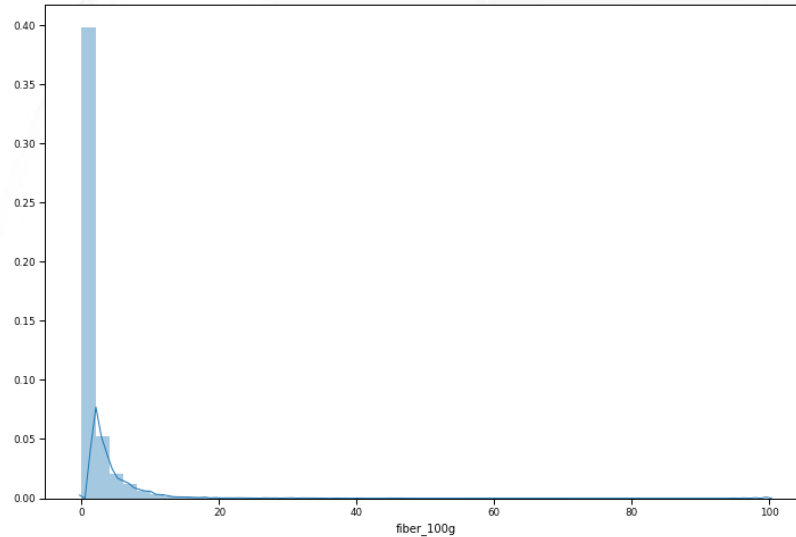
## Analyse Univariée



Histogramme de la quantité d'énergie par produits pour 100g de produits



Histogramme de la quantité de graisses saturées par produits pour 100g de produits




Histogramme de la quantité de fibre par produits pour 100g de produits



# Analyse Multivariée

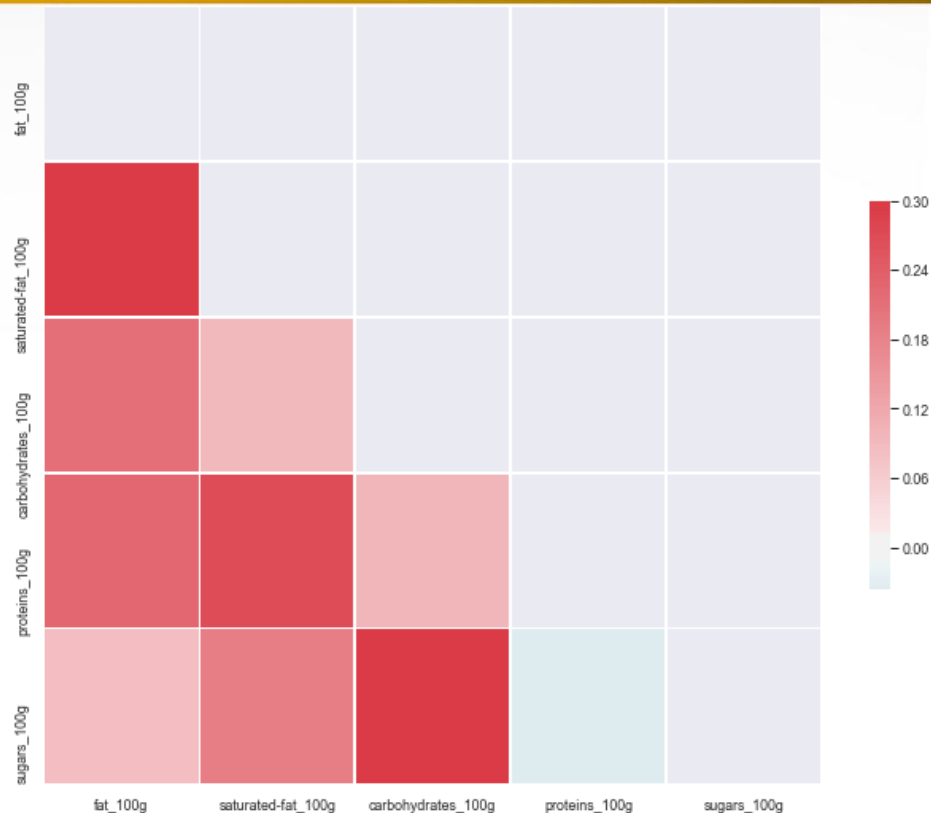
# Base de données OpenFoodFacts



	fat_100g	saturated-fat_100g	carbohydrates_100g	proteins_100g	sugars_100g
fat_100g	1	0.57	0.21	0.23	0.086
saturated-fat_100g	0.57	1	0.092	0.27	0.19
carbohydrates_100g	0.21	0.092	1	0.098	0.5
proteins_100g	0.23	0.27	0.098	1	-0.037
sugars_100g	0.086	0.19	0.5	-0.037	1

- **Analyse Multivariée**

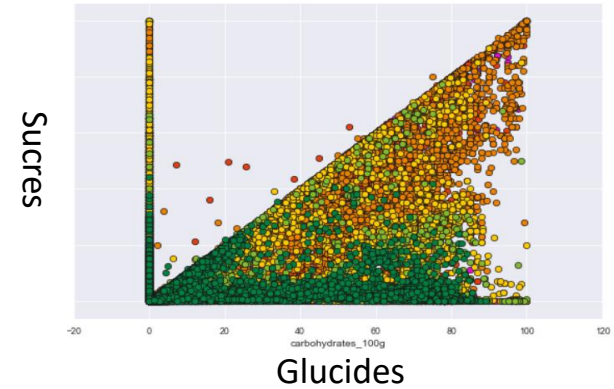
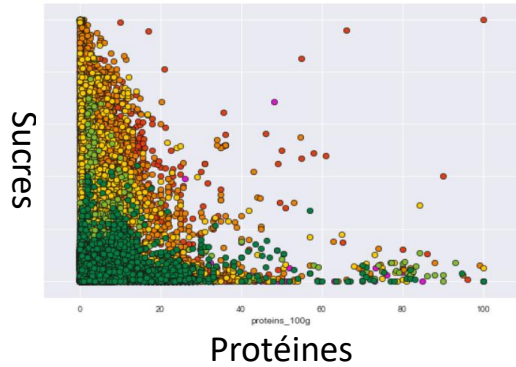
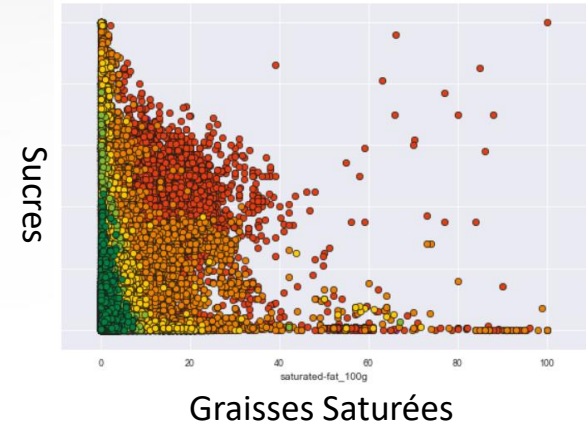
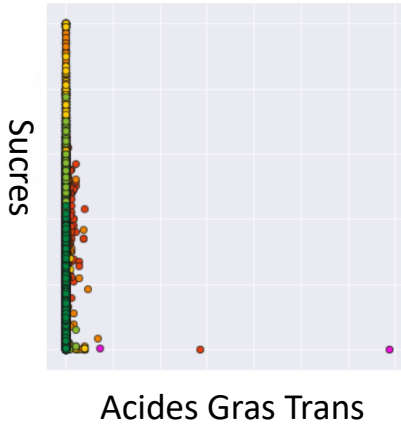
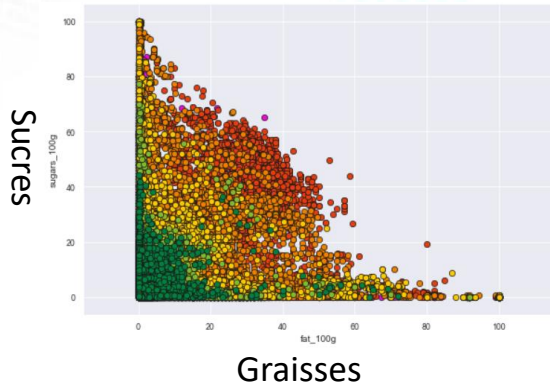
- Matrice de corrélation entre les graisses, les glucides, les protéines et les sucres (macronutriments)
- Les graisses et les graisses saturées ainsi que les glucides (carbohydrates) et les sucres sont relativement corrélés
- Les protéines sont plus corrélées avec les graisses qu'avec les glucides et les sucres
- Les glucides et les graisses sont plus corrélés que le sucre et les graisses
- Le sucre et les graisses saturées sont plus corrélés que les glucides et les graisses saturées





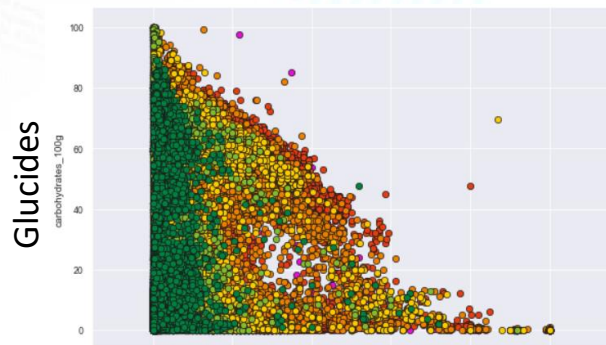
# Base de données OpenFoodFacts

Analyse Multivariée : SUCRES vs Graisses | Acides Gras Trans | Graisses Saturées | Protéines | Glucides vs Note Nutritionnelle

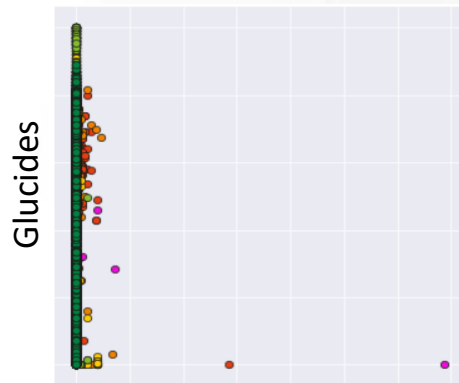


# Base de données OpenFoodFacts

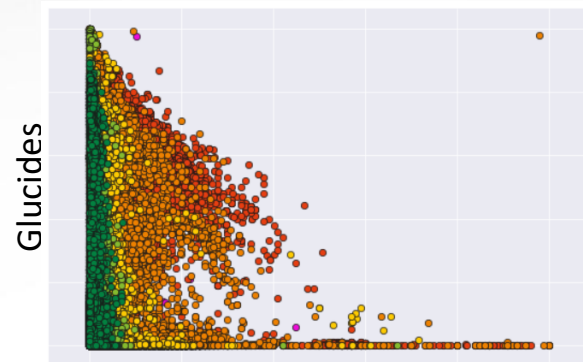
Analyse Multivariée : Glucides vs Graisses | Acides Gras Trans |  
Graisses Saturées | Protéines vs Note Nutritionnelle



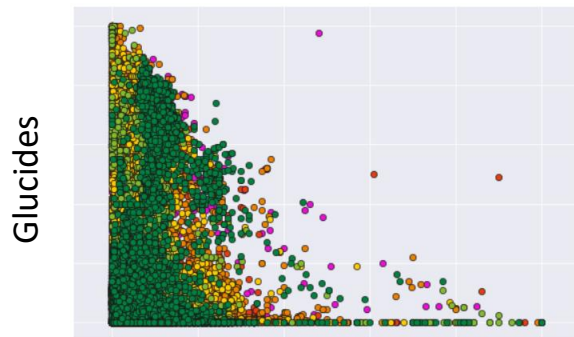
Graisses



Acides Gras Trans



Graisses Saturées



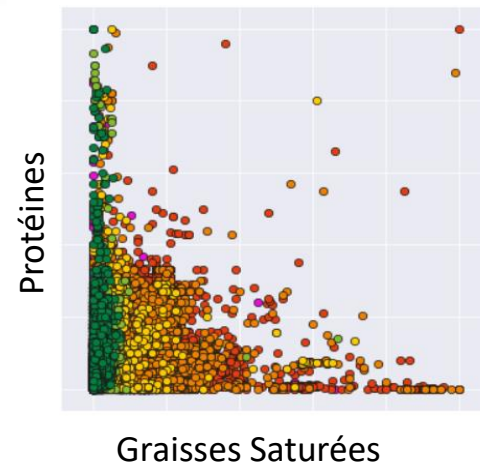
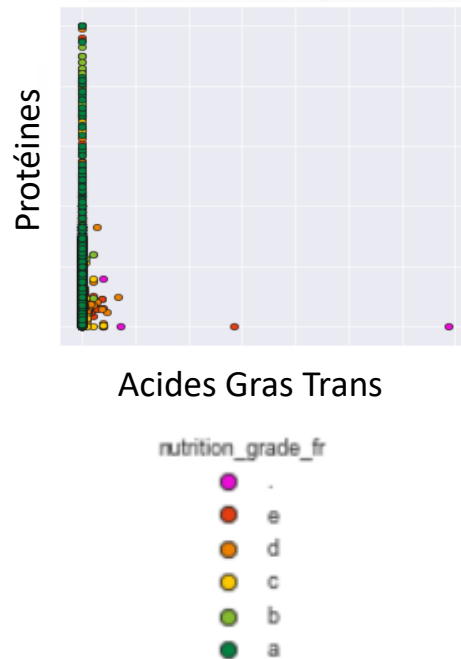
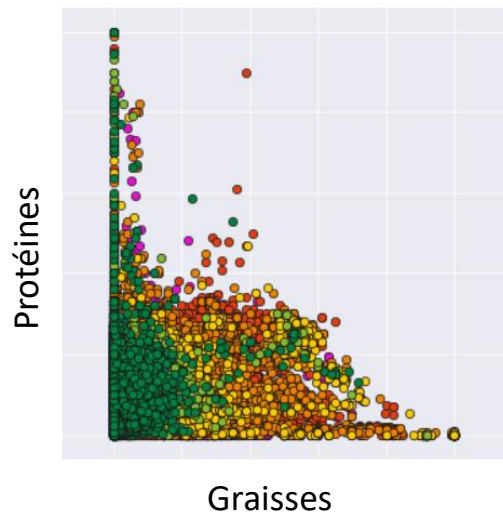
Protéines

nutrition\_grade\_fr



# Base de données OpenFoodFacts

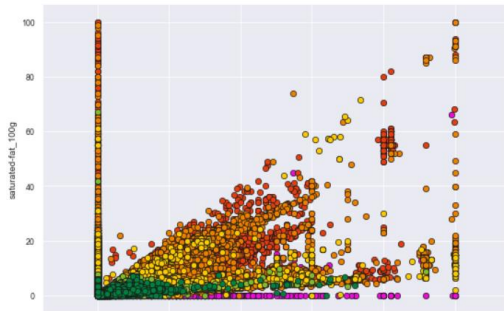
## Analyse Multivariée : Protéines vs Graisses | Acides Gras Trans | Graisses Saturées vs Note Nutritionnelle



# Base de données OpenFoodFacts

## Analyse Multivariée : Graisses Saturées vs Graisses | Acides Gras Trans vs Note Nutritionnelle

Graisses Saturées



Graisses

Graisses Saturées



Acides Gras Trans



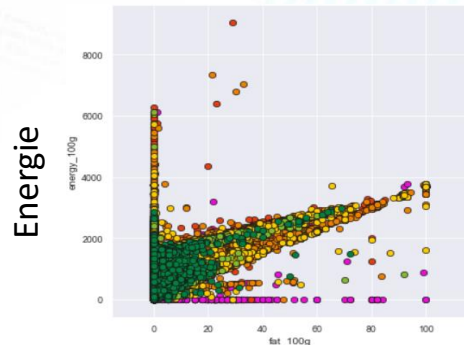
Acides Gras Trans



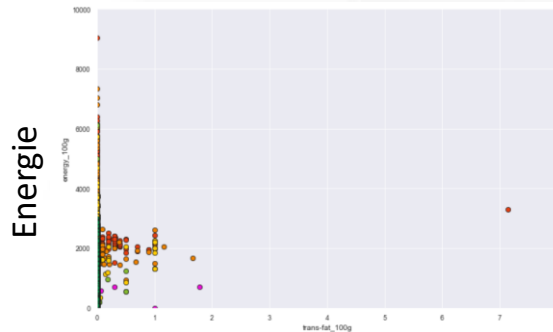
Graisses

# Base de données OpenFoodFacts

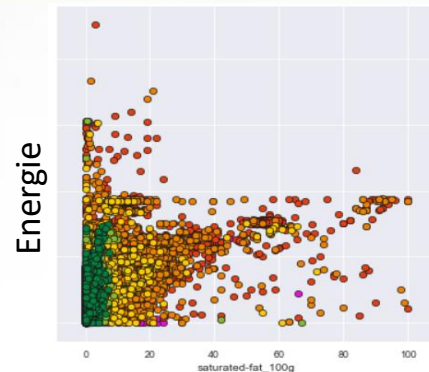
Analyse Multivariée : Energie vs Graisses | Acides Gras Trans | Graisses Saturées | Glucides | Sucres | Protéines vs Note Nutritionnelle



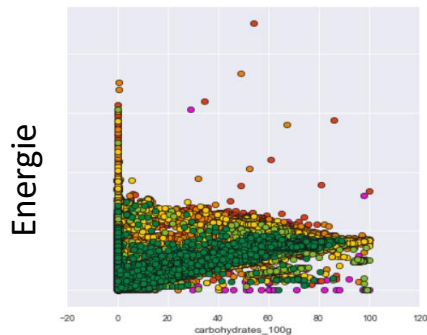
Graisses



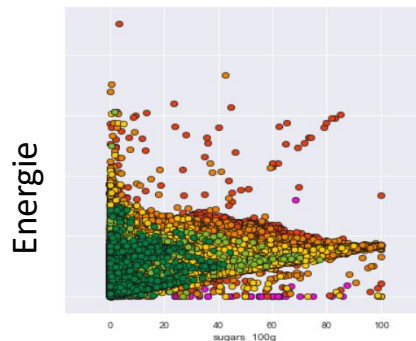
Acides Gras Trans



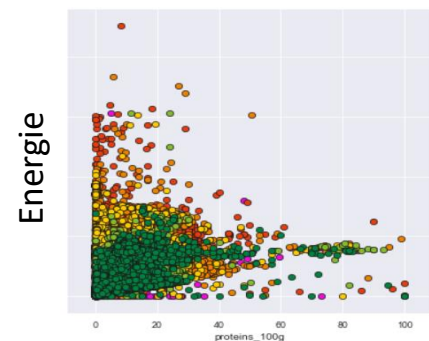
Graisses Saturées



Glucides



Sucres

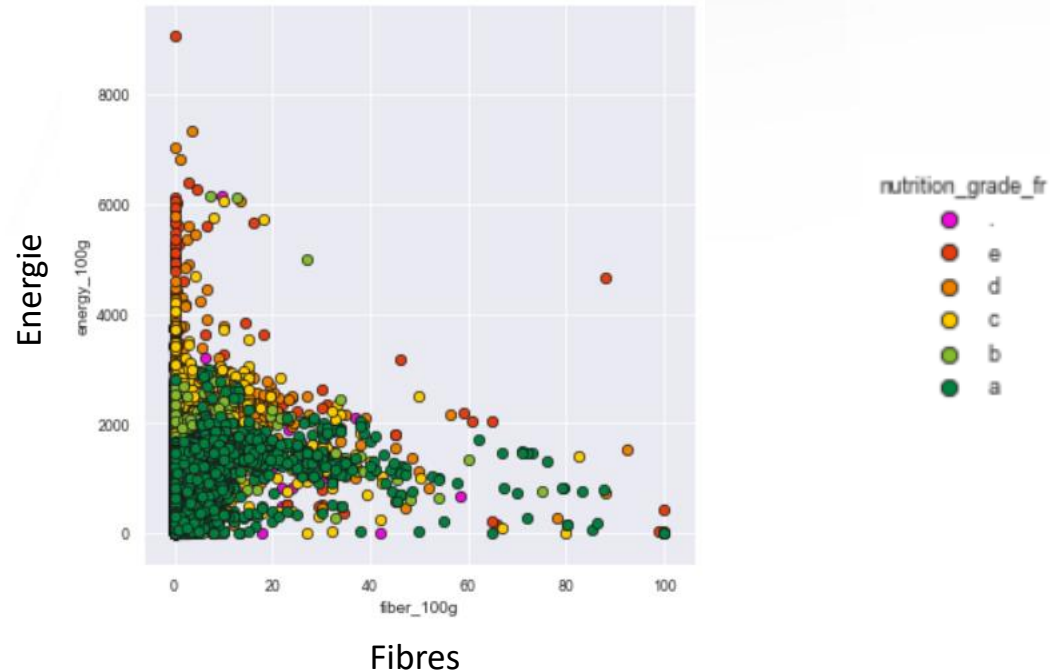


Protéines



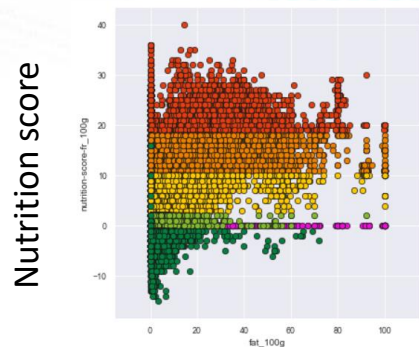
# Base de données OpenFoodFacts

## Analyse Multivariée : Energie vs Fibres vs Note Nutritionnelle

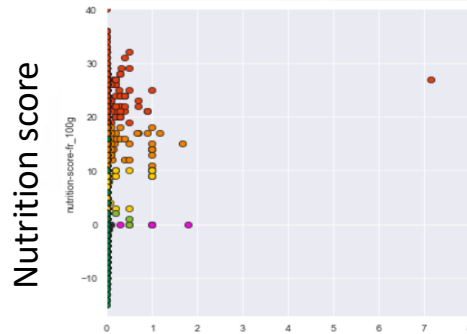


# Base de données OpenFoodFacts

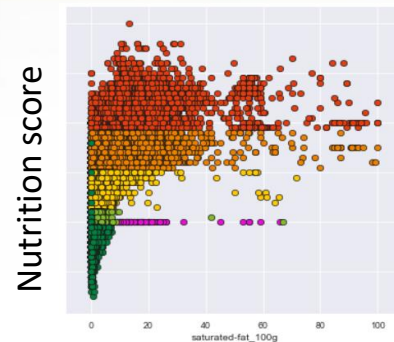
Analyse Multivariée : Nutrition Score vs Graisses | Acides Gras Trans |  
Graisses Saturées | Glucides | Sucres | Protéines vs Note Nutritionnelle



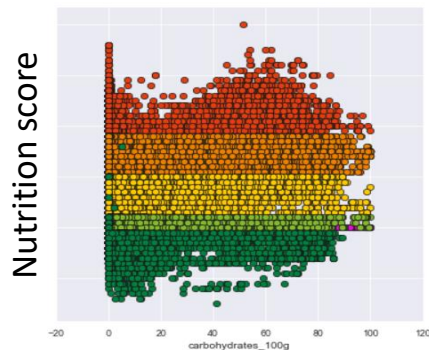
Graisses



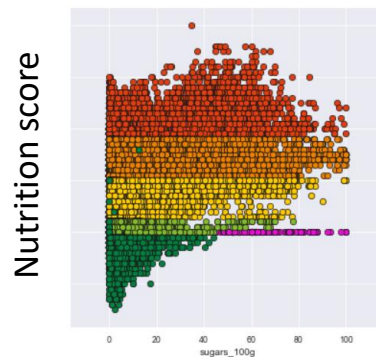
Acides Gras Trans



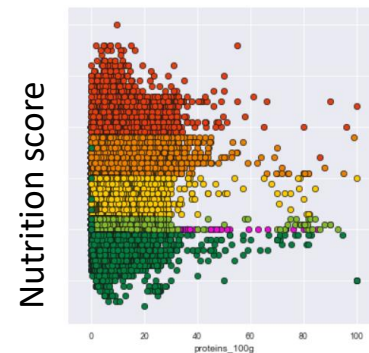
Graisses Saturées



Glucides



Sucres

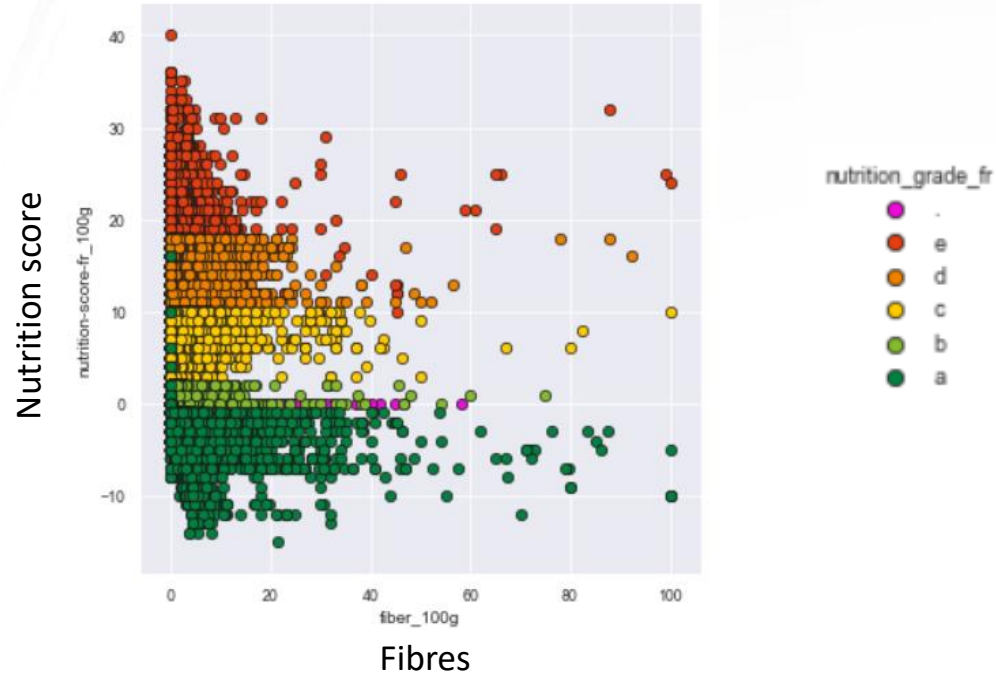


Protéines



# Base de données OpenFoodFacts

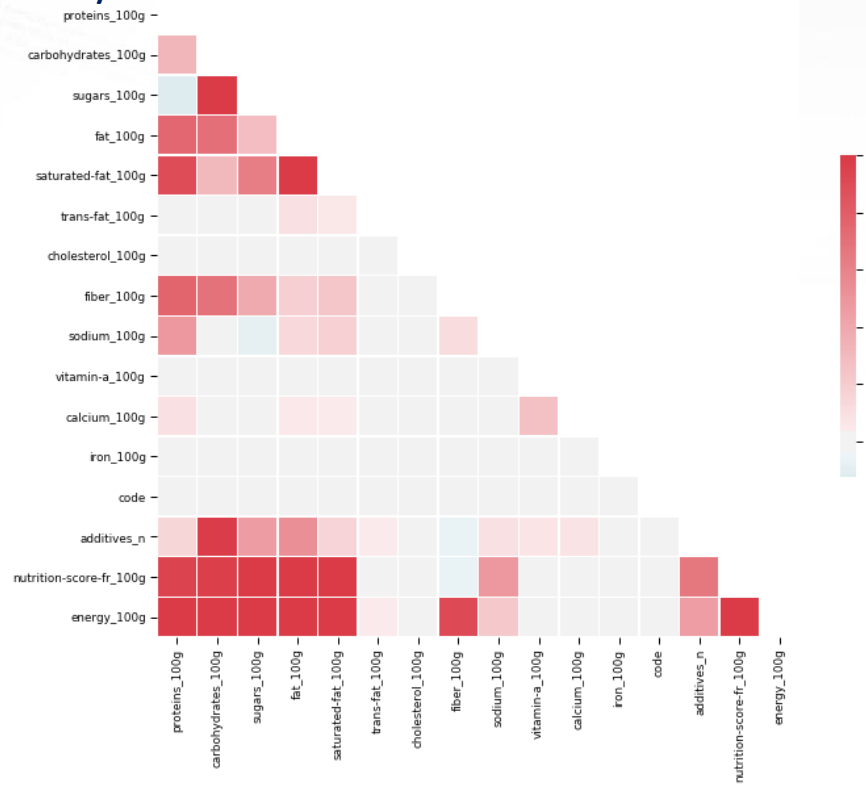
## Analyse Multivariée : Nutrition Score vs Fibres vs Note Nutritionnelle





# Base de données OpenFoodFacts

## Analyse Multivariée - Matrice de Corrélation



	proteins_100g	carbohydrates_100g	sugars_100g	fat_100g	saturated-fat_100g	trans-fat_100g	cholesterol_100g	fiber_100g	sodium_100g	vitamin-a_100g	calcium_100g	iron_100g	code	additives_n	nutrition-score-fr_100g	energy_100g
proteins_100g	1	0.098	-0.037	0.23	0.27	-0.0037	0.0014	0.23	0.14	-0.004	0.029	0.0046	-0.0028	0.046	0.28	0.43
carbohydrates_100g	0.098	1	0.5	0.21	0.092	0.0029	0.0059	0.21	-0.0057	-0.0021	0.0066	0.012	-0.0023	0.29	0.29	0.5
sugars_100g	-0.037	0.5	1	0.086	0.19	0.0011	0.0081	0.12	-0.024	-0.0023	0.0013	0.0018	-0.0021	0.14	0.54	0.44
fat_100g	0.23	0.21	0.086	1	0.57	0.028	0.009	0.059	0.04	0.0071	0.015	0.0037	-0.0019	0.16	0.48	0.63
saturated-fat_100g	0.27	0.092	0.19	0.57	1	0.019	0.0089	0.072	0.055	0.0039	0.015	-0.0022	-0.0019	0.048	0.65	0.61
trans-fat_100g	-0.0037	0.0029	0.0011	0.028	0.019	1	0.0086	-0.0019	0.00045	0.0053	0.0004	-8e-05	-5.2e-05	0.015	0.011	0.013
cholesterol_100g	0.0014	0.0059	0.0081	0.009	0.0089	0.0086	1	0.00074	-0.0004	0.00026	0.00044	0.0055	-1.9e-05	1.5e-05	0.0085	0.0076
fiber_100g	0.23	0.21	0.12	0.059	0.072	-0.0019	0.00074	1	0.037	-0.002	0.0038	0.0068	-0.0014	-0.019	-0.018	0.27
sodium_100g	0.14	-0.0057	-0.024	0.04	0.055	0.00045	-0.0004	0.037	1	-0.00099	0.00012	0.001	-0.00085	0.028	0.15	0.068
vitamin-a_100g	-0.004	-0.0021	-0.0023	0.0071	0.0039	0.0053	0.00026	-0.002	-0.00099	1	0.081	0.0048	-3.3e-05	0.022	0.001	0.00077
calcium_100g	0.029	0.0066	0.0013	0.015	0.015	0.0004	0.00044	0.0038	0.00012	0.081	1	0.0056	-0.00011	0.024	0.012	0.011
iron_100g	0.0046	0.012	0.0018	0.0037	-0.0022	-8e-05	0.0055	0.0068	0.001	0.0048	0.0056	1	-2.6e-05	0.0017	0.0014	0.0075
code	-0.0028	-0.0023	-0.0021	-0.0019	-0.0019	-5.2e-05	-1.9e-05	-0.0014	-0.00085	-3.3e-05	-0.00011	-2.6e-05	1	-0.0019	-0.0027	-0.0036
additives_n	0.046	0.29	0.14	0.16	0.048	0.015	1.5e-05	-0.019	0.028	0.022	0.024	0.0017	-0.0019	1	0.2	0.14
nutrition-score-fr_100g	0.28	0.29	0.54	0.48	0.65	0.011	0.0085	-0.018	0.15	0.001	0.012	0.0014	-0.0027	0.2	1	0.7
energy_100g	0.43	0.5	0.44	0.63	0.61	0.013	0.0076	0.27	0.068	0.00077	0.011	0.0075	-0.0036	0.14	0.7	1

## Conclusions

- Les observations sont cohérentes avec le fait qu'une alimentation équilibrée est un peu de tout (macro et micro nutriments) sans excès
- L'influence croisée des graisses (saturées, trans, et classiques) et des sucres est visible sur le score nutritionnel. Dès que la quantité augmente, cela influe négativement sur le score nutritionnel et donc sur la note nutritionnelle
- L'énergie et le score nutritionnel sont influencés par les protéines, les glucides, les sucres, les graisses, et les fibres
- La base de données surreprésente les produits transformés et industriels

## Recommandations

- Travailler sur une base de données avec des données plus fiables et surtout qui contient plus d'informations sur les variables acides gras mono et poly insaturés (bonne graisses)
- S'assurer d'avoir plus de produits non transformés dans la base de données (matières premières)
- Travailler sur une version plus récente de la base car il semble qu'il y a eu beaucoup de travail de corrections mené dessus
- Tester d'autres méthodes d'imputations des données
- Inclure les données concernant le taux de fruits/légumes/noix dans les analyses multivariées

**MERCI DE VOTRE ATTENTION**  
**QUESTIONS ?**