

Projet 2

Concevez une application au service
de la santé publique



Loridan Adrien



student machine learning
engineer



medium.com/@adrien.loridan



linkedin.com/in/adrien-loridan



github.com/loridan

Everything begins with an idea

**Une idée d'application au service de
la santé publique**

Comment pouvons nous améliorer la santé de la population ?

Everything begins with an idea

Une idée d'application au service de la santé publique

Comment pouvons nous améliorer la santé de la population ?

Problème

L'obésité en France, tout comme le surpoids, augmente régulièrement depuis 1997.

Everything begins with an idea

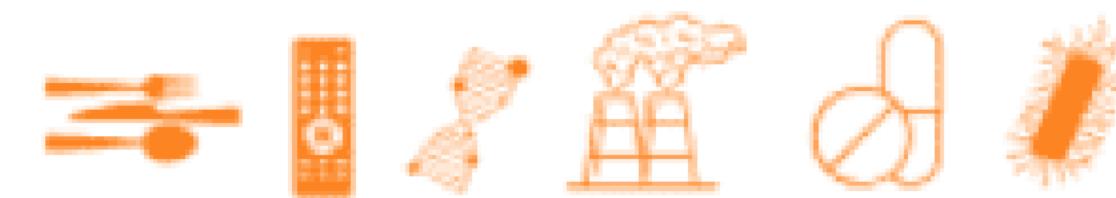
Une idée d'application au service de la santé publique

Source Inserm 

L'obésité concerne
17% des adultes
en France
(13% dans le monde)

De nombreuses complications

✓ Diabète de type 2, ✓ maladies cardiovasculaires, ✓ hépatiques, ✓ rénales, ✓ cancers, ✓ respiratoires, ✓ articulaires, ✓ dermatologiques...



- des **causes multiples**
- des **mécanismes biologiques variés**
pas encore tous entièrement élucidés

Everything begins with an idea

Une idée d'application au service de la santé publique

Démarche de résolution

S'attaquer à la nutrition, c'est à dire en favorisant le choix de produits plus sains par les consommateurs

Everything begins with an idea

Une idée d'application au service de la santé publique

Objectif à atteindre

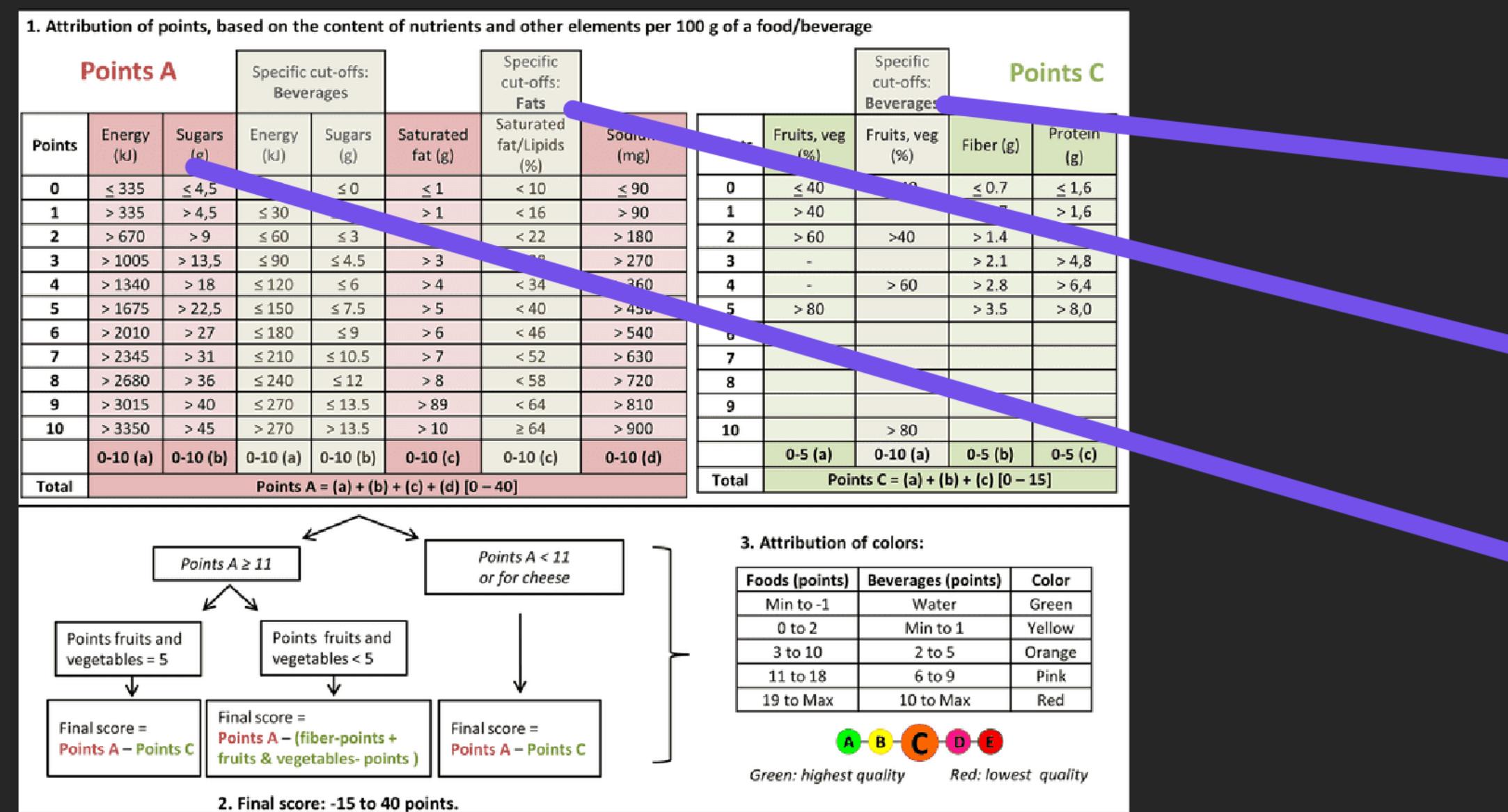
Le consommateur doit obtenir le nutriscore de n'importe quel aliment lors de son choix en magasin



Everything begins with an idea

Une idée d'application au service de la santé publique

Calcul du Nutriscore suivant les catégories



Boissons

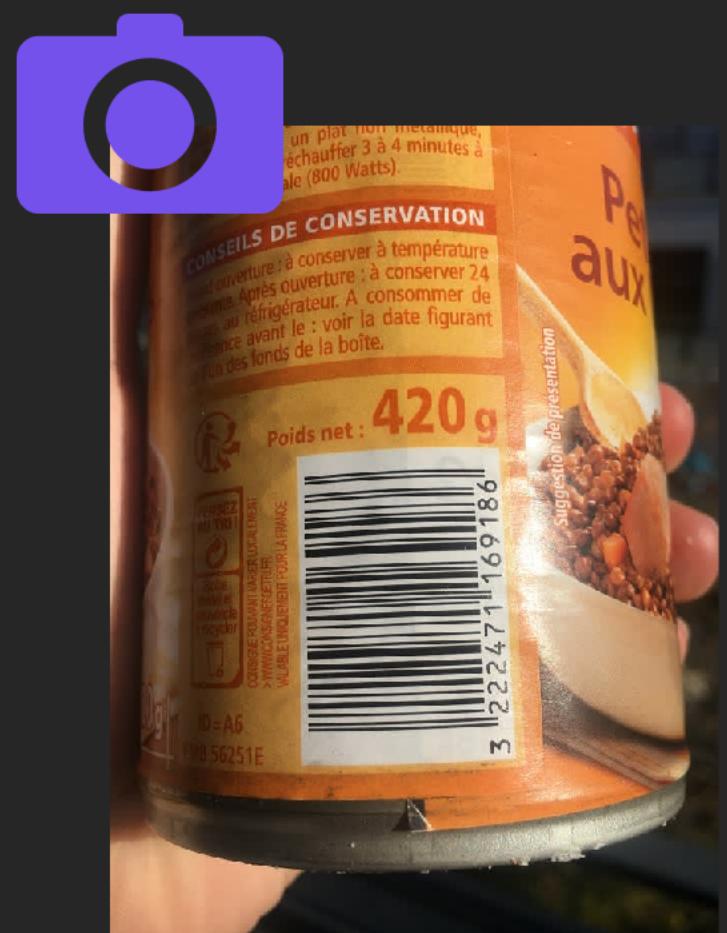
Matières grasses ajoutées

Divers produits transformés

Everything begins with an idea

Une idée d'application au service de la santé publique

Cas 1 : Utilisation du code-barre



consommateur

prend
une photo



serveur

service intelligent

bdd

api

Everything begins with an idea

Une idée d'application au service de la santé publique

Cas 2 : Utilisation des nutriments



consommateur

serveur

prend
une photo



api

service intelligent

bdd



Everything begins with an idea

Une idée d'application au service de la santé publique

Le service intelligent sera capable d'extraire les données de nutrition à partir d'une photo mais ...

Hypothèse principale

Est-il possible de prédire le Nutriscore à partir des nutriments d'un produit ?

Everything begins with an idea

Une idée d'application au service de la santé publique

Le service intelligent sera capable d'extraire les données de nutrition à partir d'une photo mais ...

Hypothèse secondaire

Pouvons nous classer nos produits en 3 catégories:
{divers, matières grasses ajoutées, boissons} ?

La science des données

Expérimentation

1 Collecter des données

2 Préparer les données

3 Explorer les données



La science des données

Expérimentation



Téléchargement au format CSV

<https://static.openfoodfacts.org/data/en.openfoodfacts.org.products.csv>

Quelle est la réglementation sur la protection de ces données ?

Aucune, la source de données est publique

Où sont stockées les données ?

Elles sont sur les serveurs distants publics et sur notre ordinateur en sdd local

Quel est le type des données

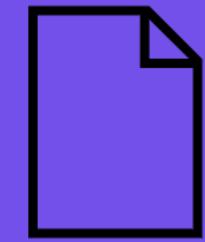
Structurée

La science des données

Expérimentation



- Importer et observer
- Identifier et traiter les doublons
- Identifier et traiter les valeurs manquantes
- Identifier et traiter les valeurs aberrantes
- Renommer les colonnes
- Convertir
- Exporter pour analyse



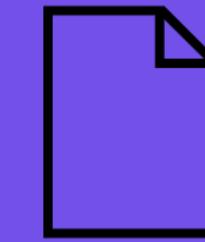
P02_01_notebook_nettoyage.ipynb

La science des données

Expérimentation

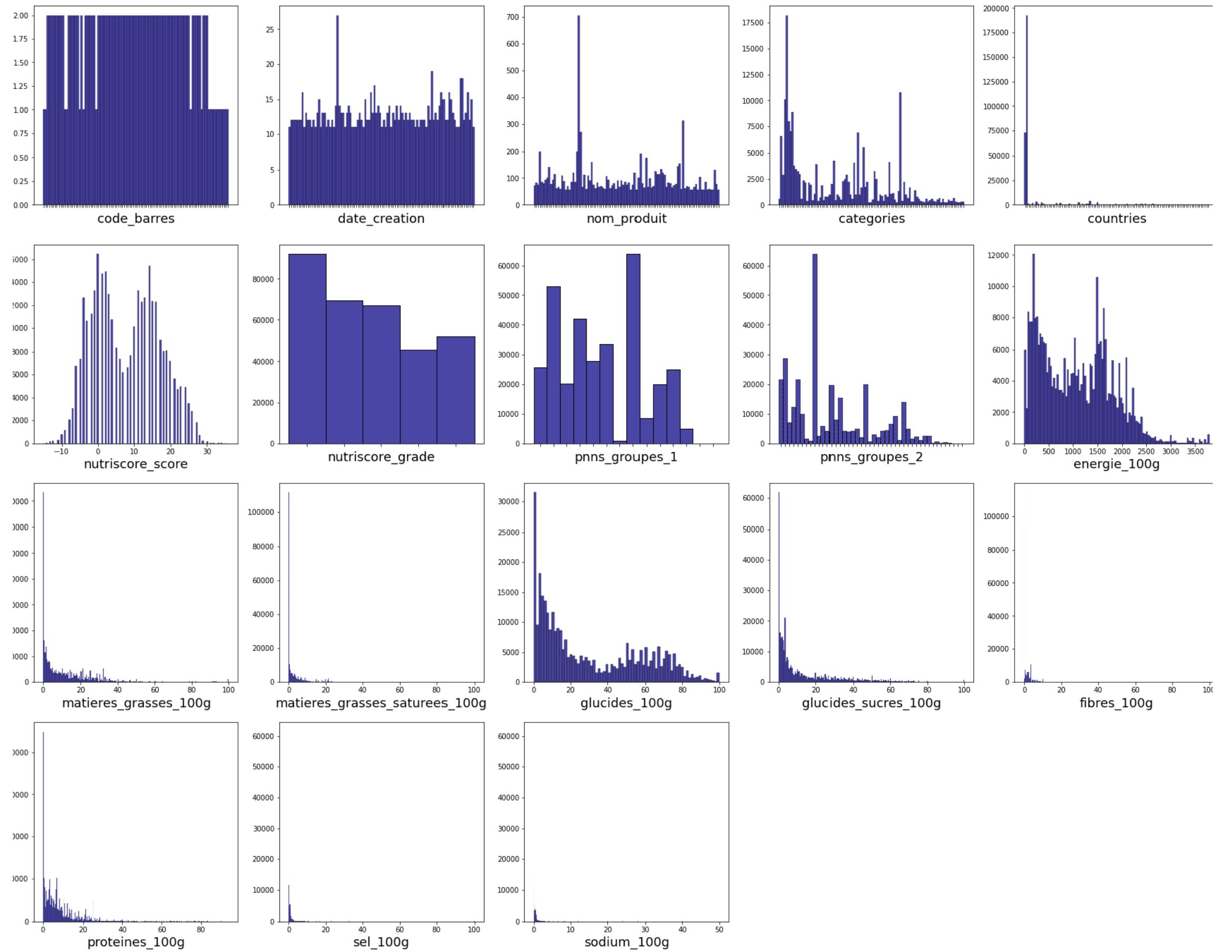


- Import des dépendances , fonctions, constantes et des données pré-traitées
- Visualiser les données
- Analyser les données
- Réaliser des tests statistiques
- Conclusion

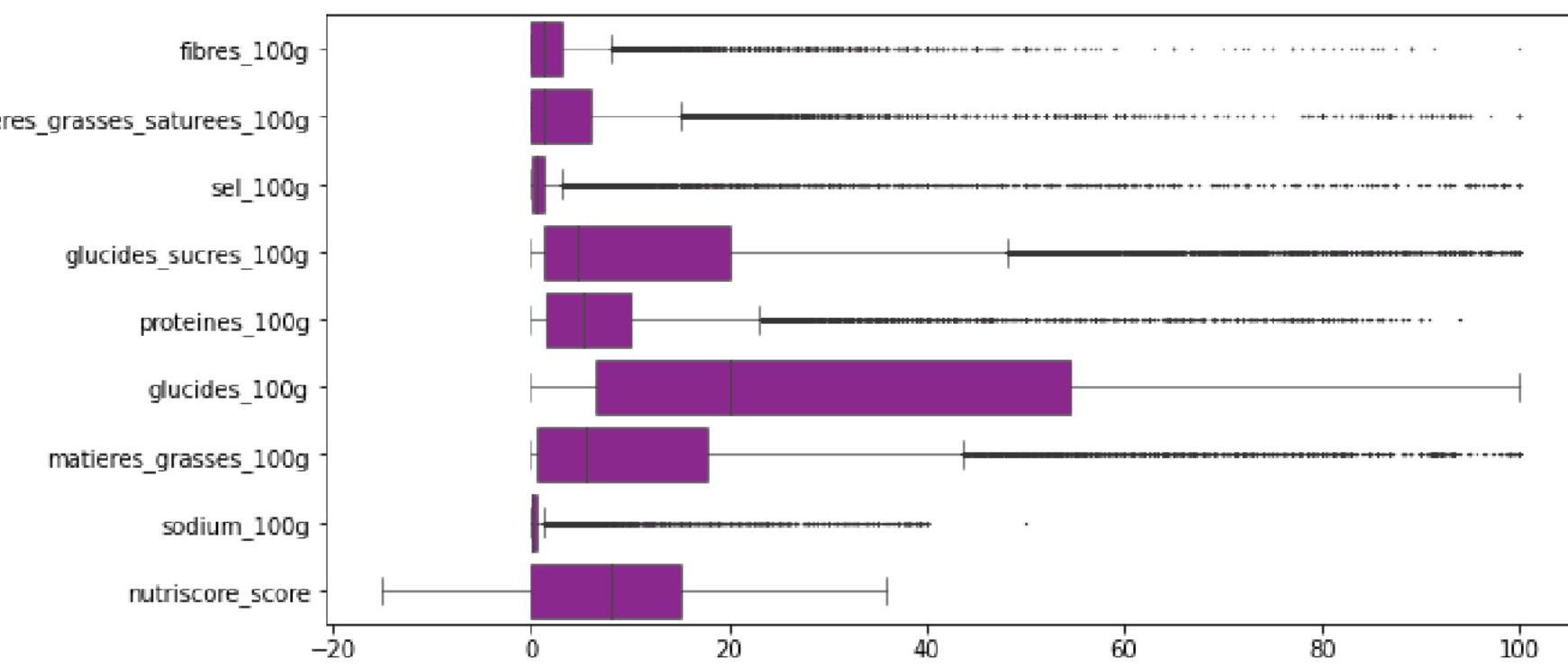


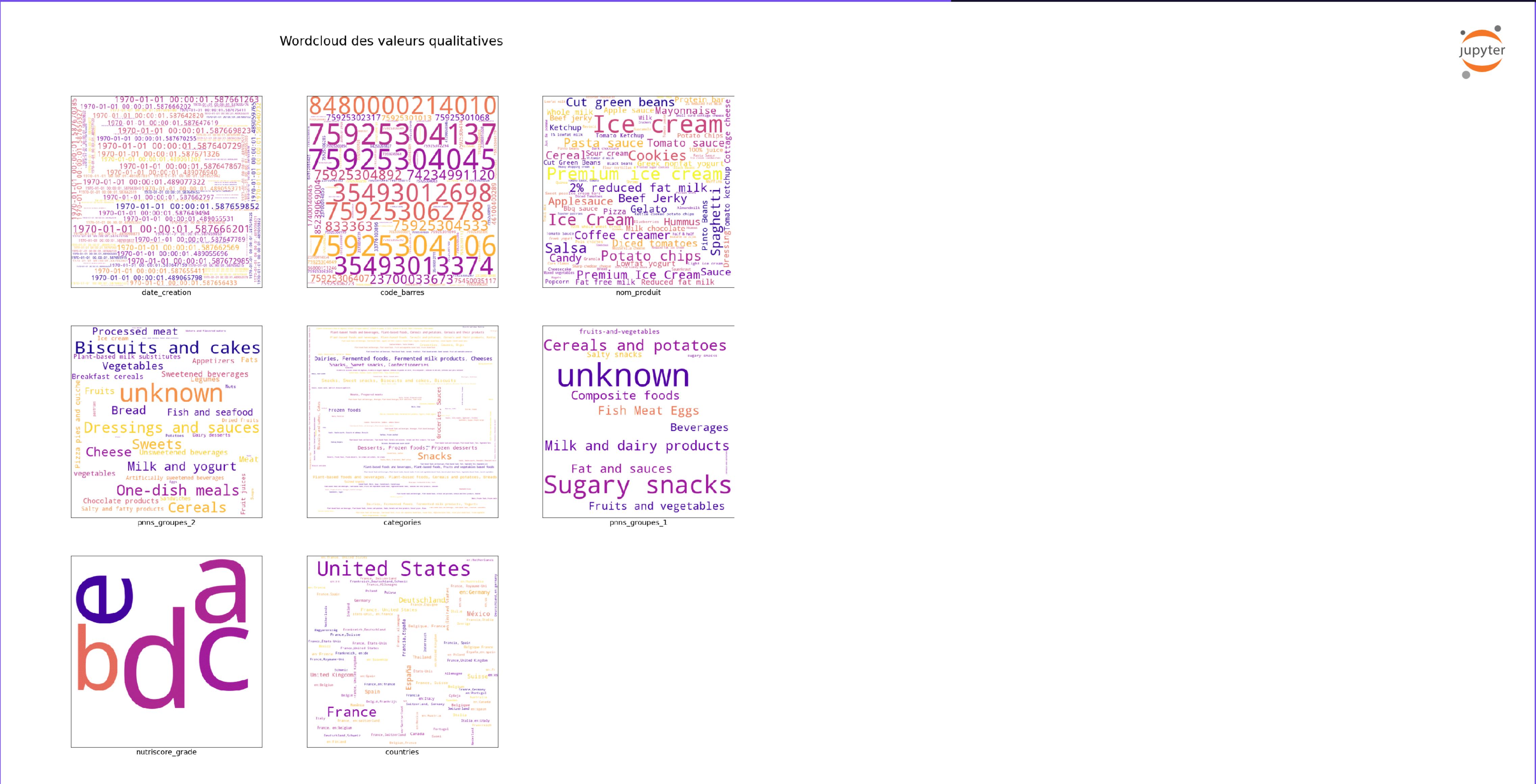
P02_02_notebook_exploration.ipynb

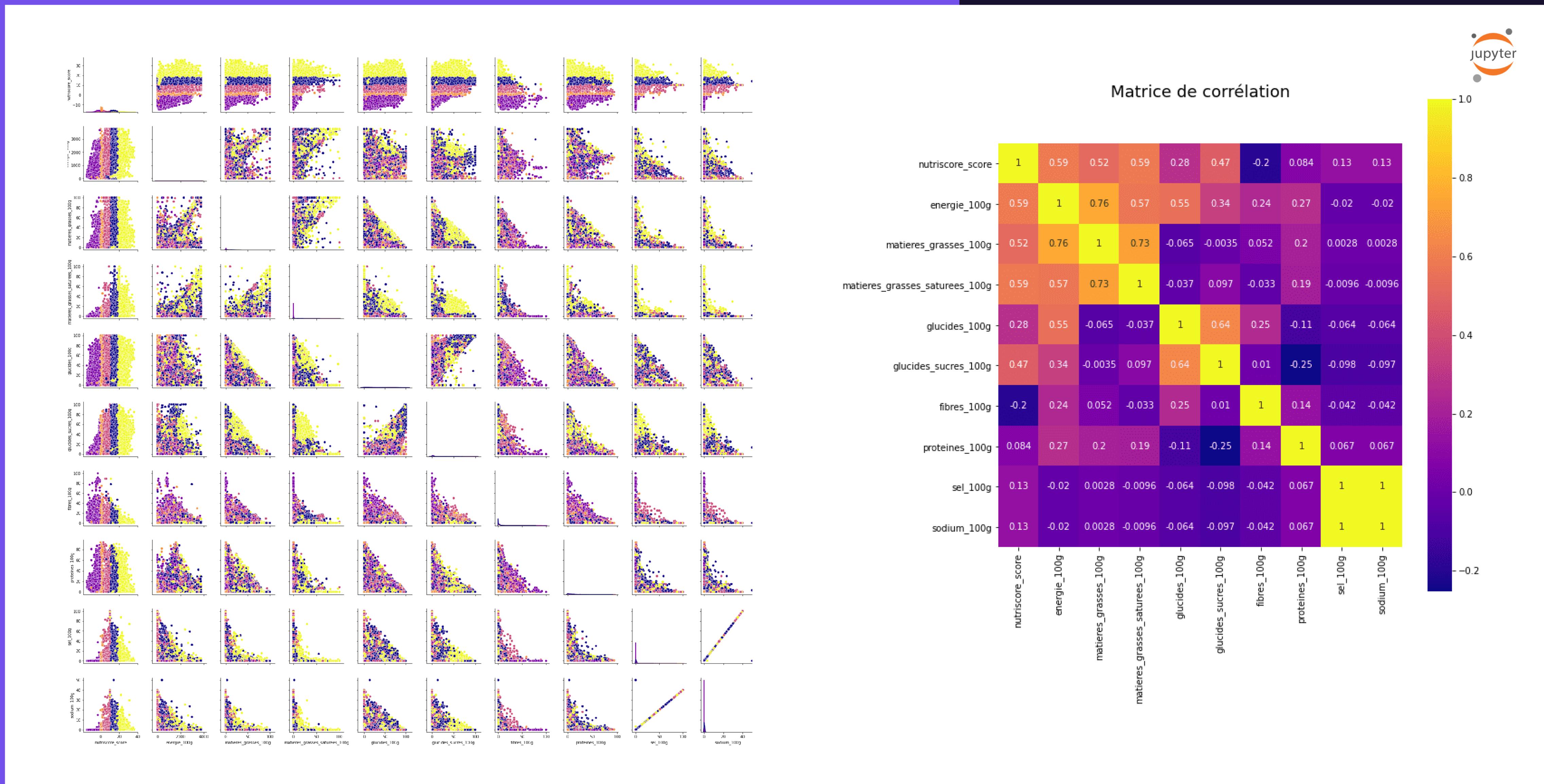
Distribution des valeurs quantitatives & qualitatives



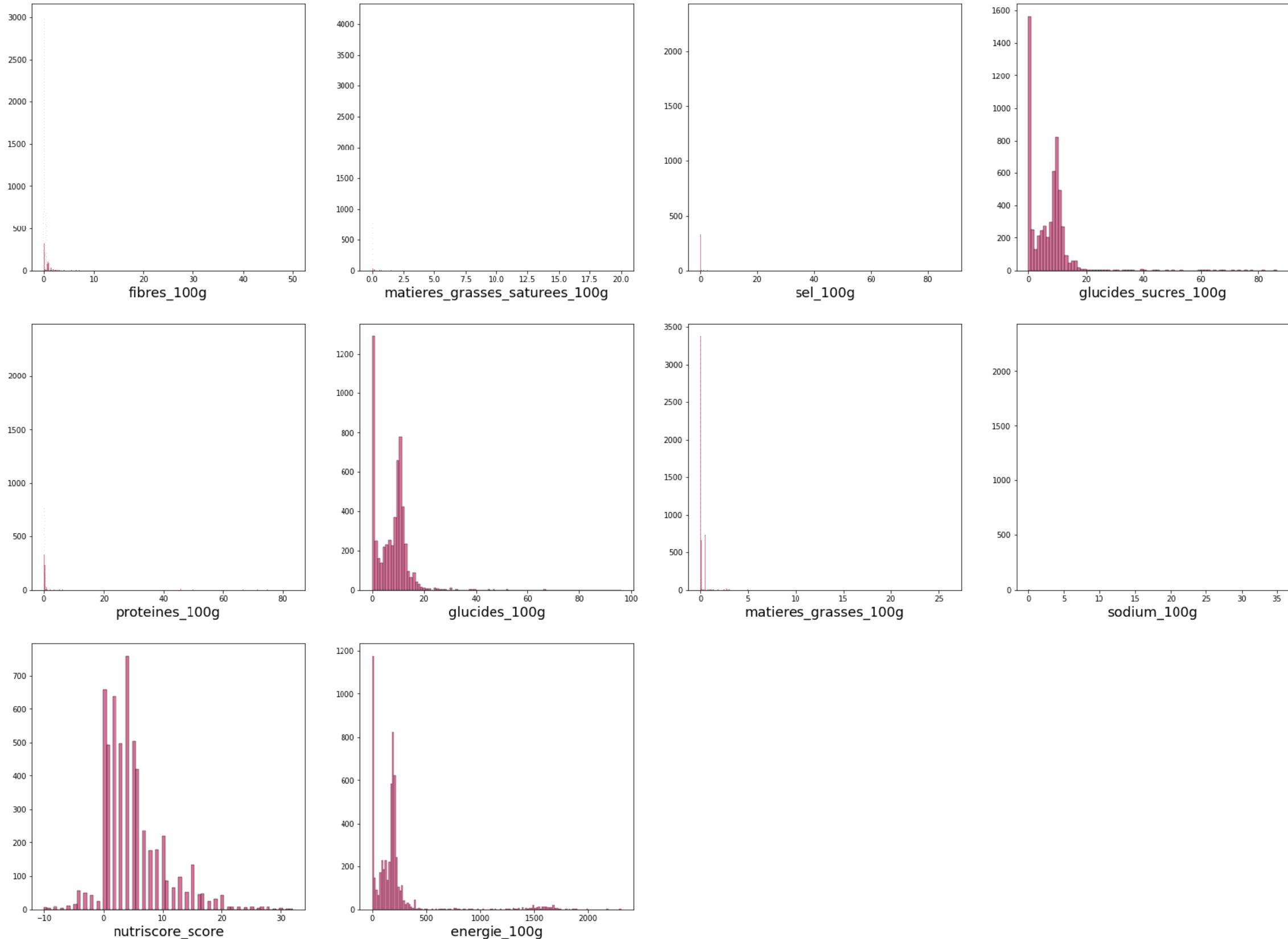
Comparaison des distributions des valeurs quantitatives



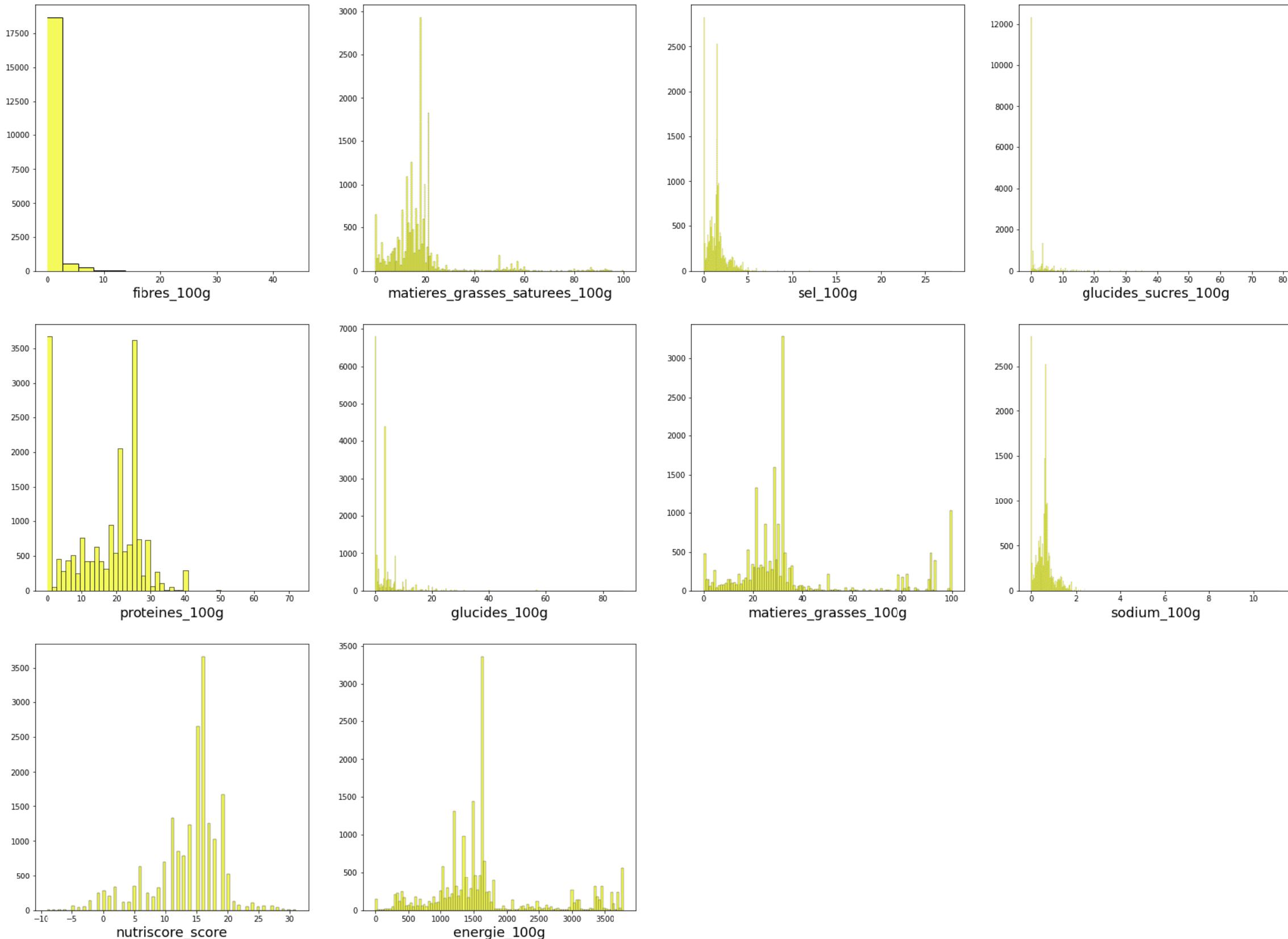




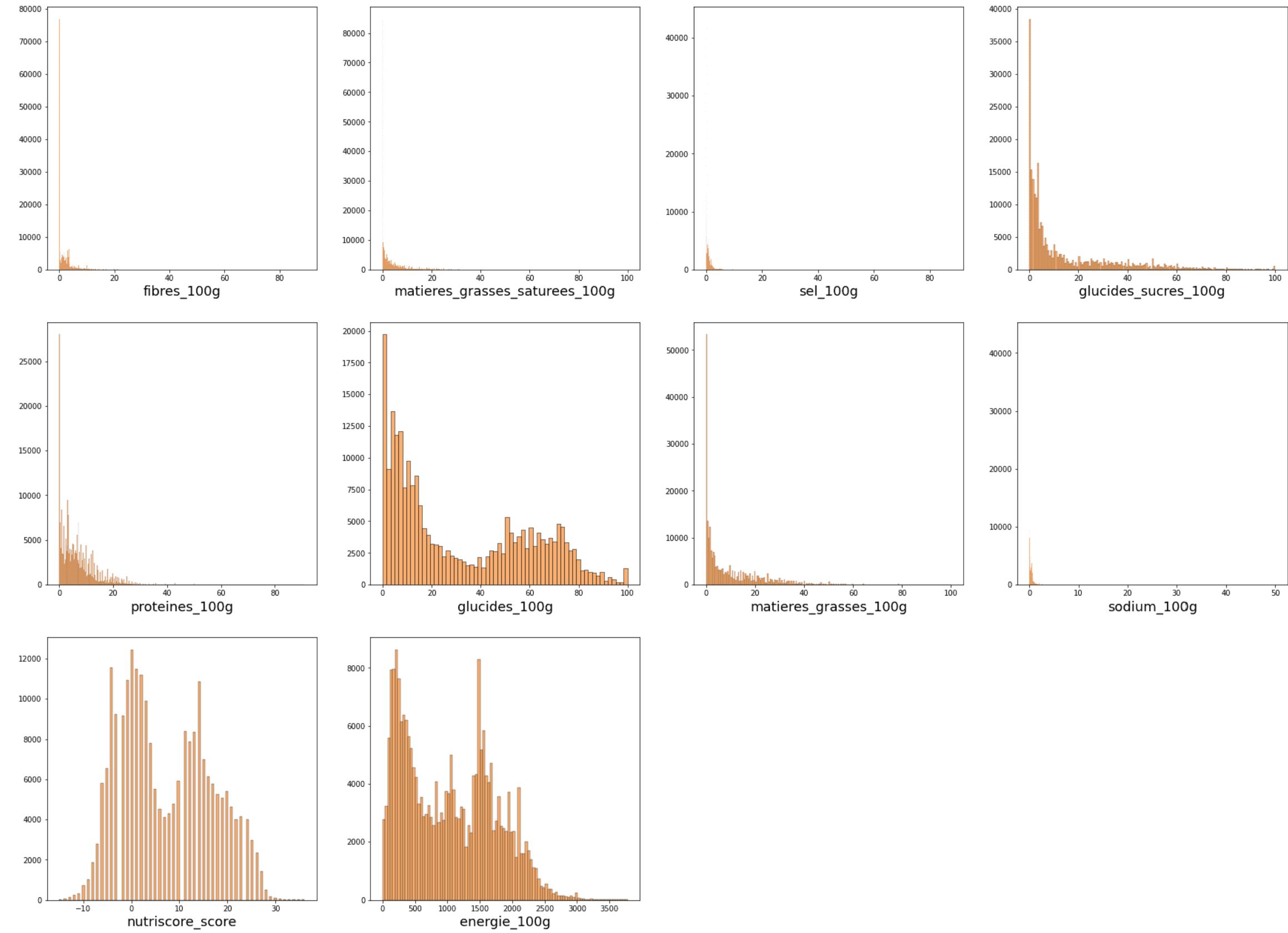
Distribution des valeurs quantitatives pour les boissons



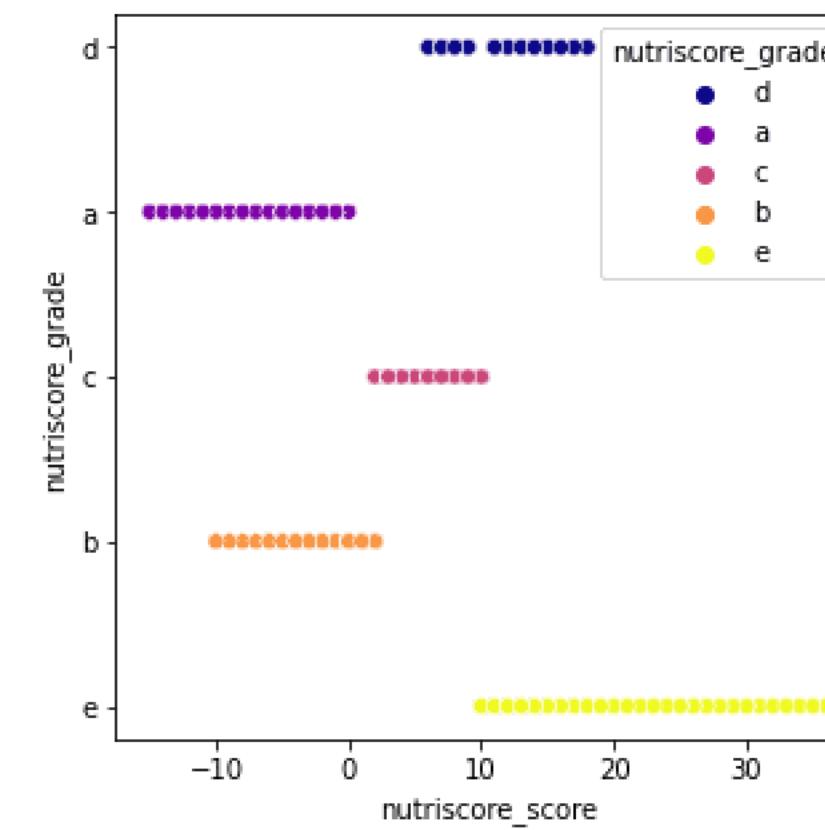
Distribution des valeurs quantitatives pour les fromages & matières grasses ajoutées



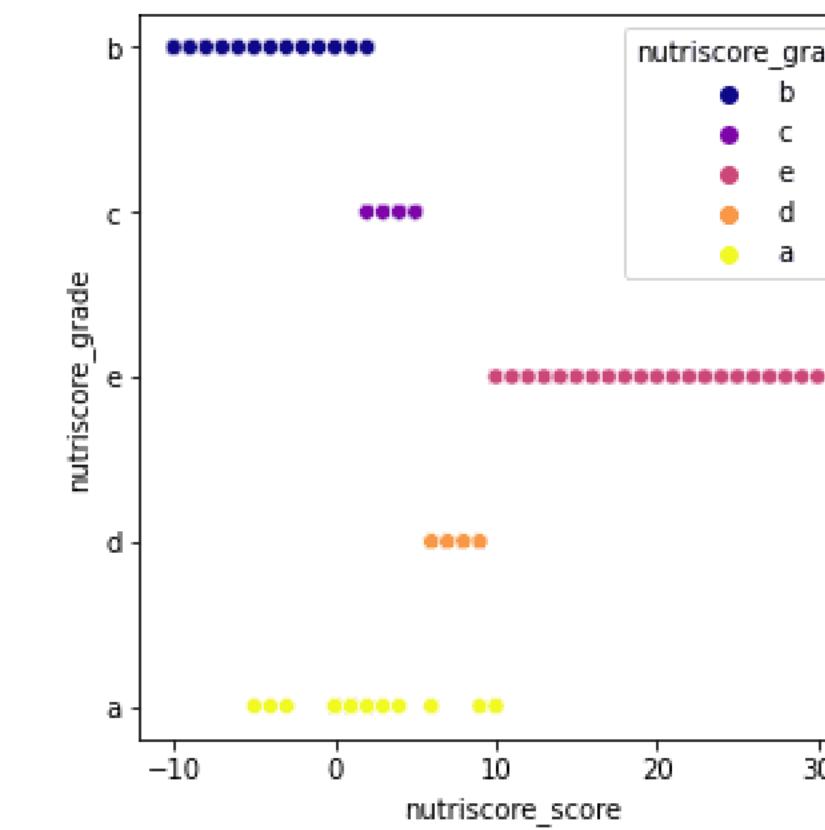
Distribution des valeurs quantitatives pour les produits divers



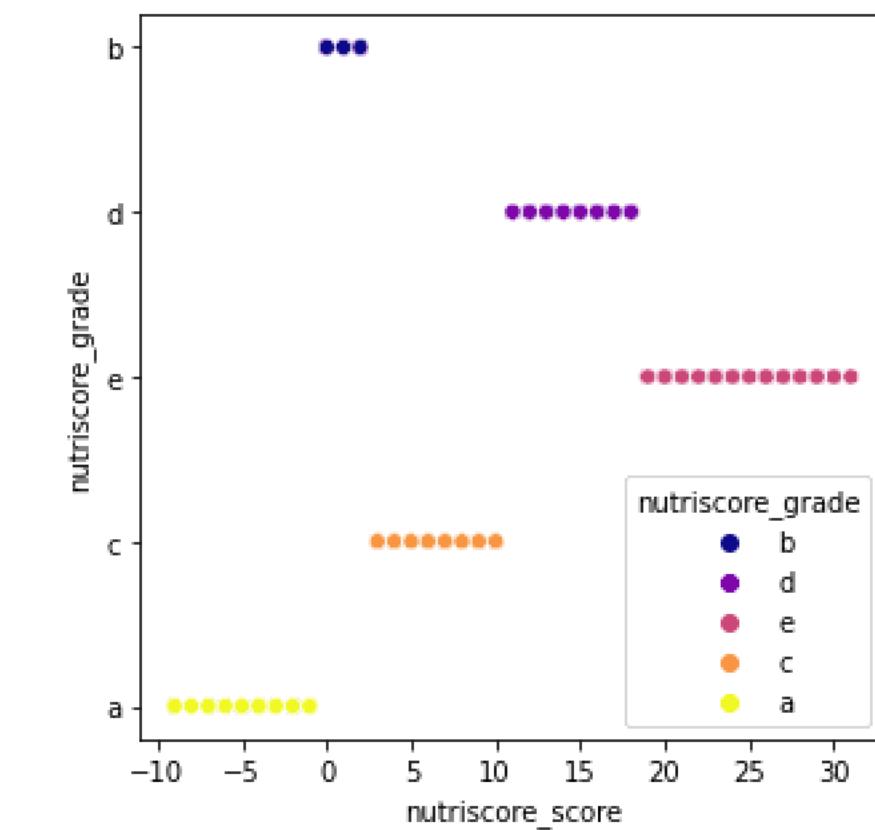
divers



boissons



fromages & matières grasses





Test de Kruskal-Wallis

statistics: 31728.001, p = 0.000

grouplabel: nutriscore_grade, label: fibres_100g, alpha: 0.05

resultat: distributions differentes (rejeter H0)

statistics: 101915.944, p = 0.000

grouplabel: nutriscore_grade, label: matieres_grasses_saturees_100g, alpha: 0.05

resultat: distributions differentes (rejeter H0)

statistics: 41352.990, p = 0.000

grouplabel: nutriscore_grade, label: sel_100g, alpha: 0.05

resultat: distributions differentes (rejeter H0)

statistics: 47770.038, p = 0.000

grouplabel: nutriscore_grade, label: glucides_sucres_100g, alpha: 0.05

resultat: distributions differentes (rejeter H0)

.....

Conclusions

Le nutriscore-grade devrait être identique pour le même nutriscore-score des produits d'une même catégorie.

Malgré des distributions significativement différentes entre les groupes distincts de nutriscore-grade, il sera difficile de classer en fonction des nutriments notre produit afin de réaliser un calcul exact du nutriscore ou de réaliser une prédiction.

Projet 2

Thank you !

