

PROJET 3

Concevez une application au service de la santé publique

Projected sales of main products in 2013

Distribution of market share among the major industry players



Distribution of market share among the major industry players: IT & C and BN & T was 74% and 26% percent respectively. A further change in the economic situation in the market will be characterized by a more equal distribution of market share major players

Share of market activity

Changes in the activity of the active and passive market in 2013. Estimated data based on various sources

Projected sales of main products in 2013



Passive market share

Application de conseil nutritionnel

Recherche de produits en fonction de critères choisis

- Femme enceinte
 - Vitamine B12
 - Fer
 - Calcium
- Sportif haut niveau prise de masse
 - Protéines en quantités importantes
 - Bon Nutriscore et score Nova Group
- Régime sans sel et minimum sucre

Proposition d'un produit de remplacement avec de meilleurs indicateurs

Proposition de menus sur une semaine avec un équilibre sur la semaine



Le jeu de donnée OpenFoodFacts

Données disponibles à l'adresse <https://world.openfoodfacts.org/data>

Constitué de 1 630 000 lignes de 184 colonnes

Conserve les données pour la France soit 770 000 lignes

On va travailler principalement avec les colonnes :

- Energy-kj_100g : valeur énergétique pour 100g
- nutriscore_score : score qui permet de donner le grade NUTRIScore (A,B,C,D,E)
- Fat_100g : teneur en graisse pour 100g
- Saturated-fat_100g : teneur en graisse saturée pour 100g
- Sugars_100g : teneur en sucres pour 100g
- Fiber_100g : teneur en fibre pour 100g
- Proteins_100g : teneur en protéines pour 100g
- Sodium_100g : teneur en sodium pour 100g



Nettoyage du jeu de données OpenFoodFacts

- ❖ calcul à partir de la colonne energy-kcal de la colonne energy-kj ($1\text{kcal}=4,18\text{kJ}$)
- ❖ suppression des incohérences
 - ❖ Rapport energy-kj/energy-kcal <3
 - ❖ ligne où colonne xxxxxxxx_100g $> 100\text{g}$
 - ❖ ligne où la somme graisse, sucre, fibres, proteines et sel $> 100\text{g}$
- ❖ suppression des duplicatas
- ❖ valeurs aberrantes pour energy-kj_100g
 - ❖ deux données : bruschetta et yaki nori
 - ❖ Beaucoup de données au delà de 3200kj correspondent à des erreurs de saisie entre kJ et kcal, j'ai choisi de les supprimer



Nettoyage du jeu de données OpenFoodFacts

TRAITEMENT DES VALEURS MANQUANTES pour les colonnes suivantes :

- energy-kj_100g : 0,28 % de valeurs manquantes
- nutriscore_score : 55,74 % de valeurs manquantes
- fat_100g : 0,61 % de valeurs manquantes
- saturated-fat_100g : 0,46 % de valeurs manquantes
- sugars_100g : 0,37 % de valeurs manquantes
- proteins_100g : 0,27 % de valeurs manquantes

On va remplacer les valeurs manquantes grâce à un algorithme KNN (k-plus proches voisins) pour les petits pourcentages. Pour la colonne NutrisScore nous allons repartir de la formule de base.

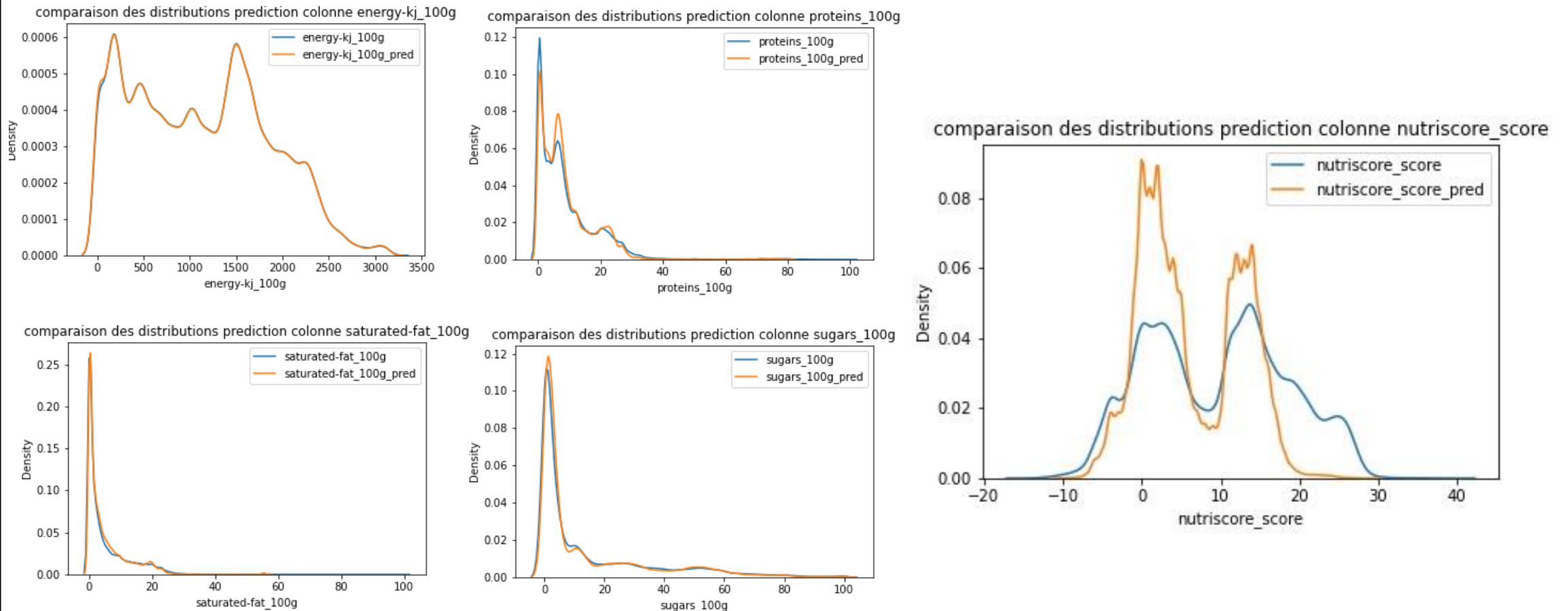
Nous arrivons à la fin de cette phase à un fichier de 577 742 lignes.

Chaque ligne est visible à l'adresse https://fr.openfoodfacts.org/produit/code_barre

Ex : <https://fr.openfoodfacts.org/produit/3061990141354/>

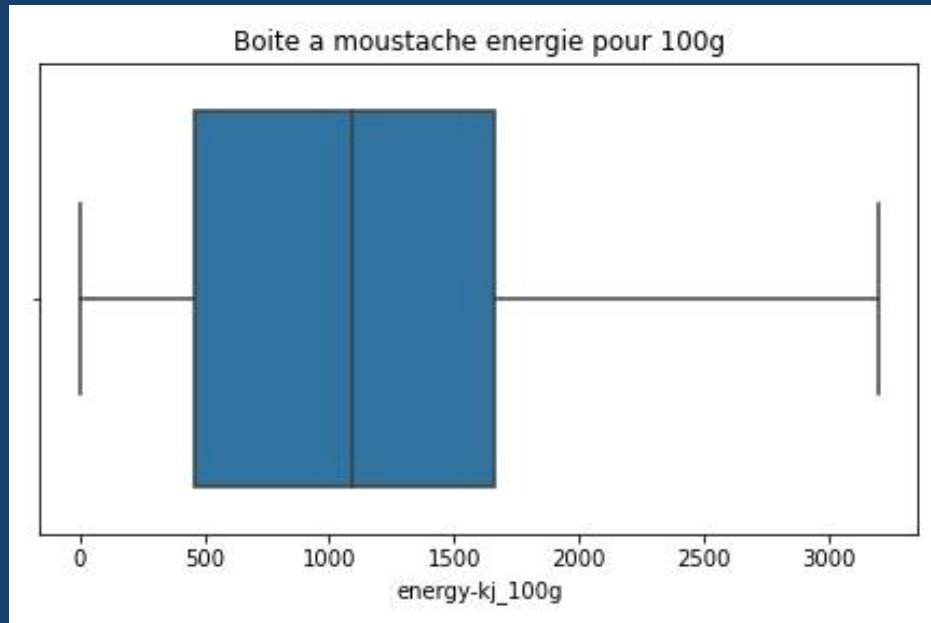


Nettoyage du jeu de données OpenFoodFacts



analyse exploratoire

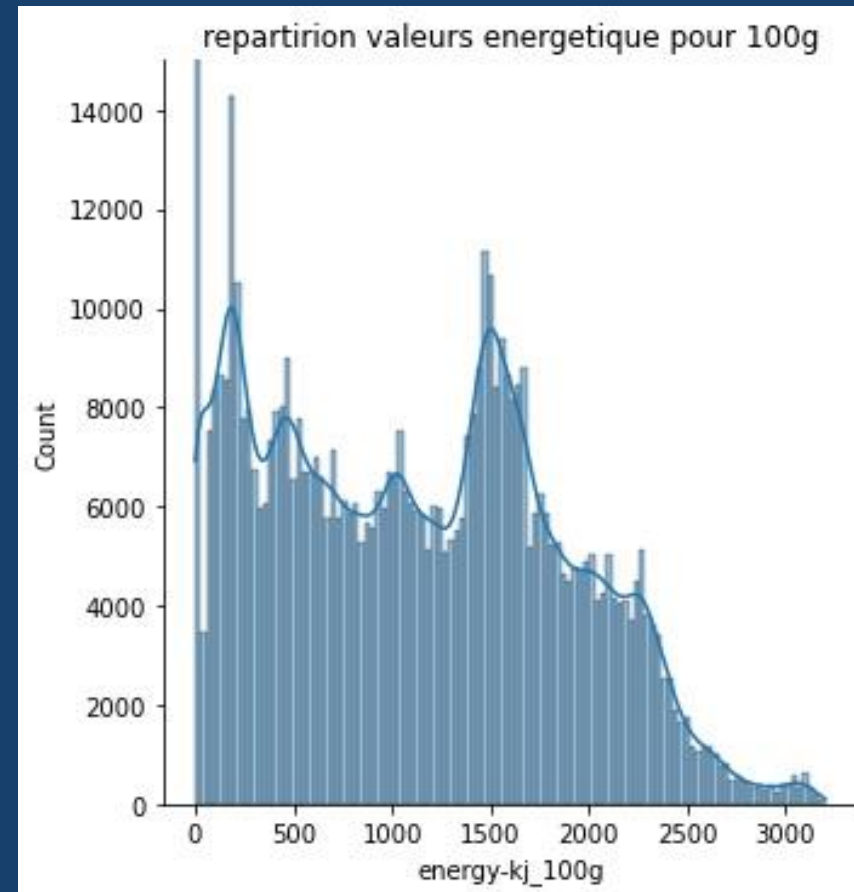
analyse univariée energy-kj_100g



Moyenne 1116 kj

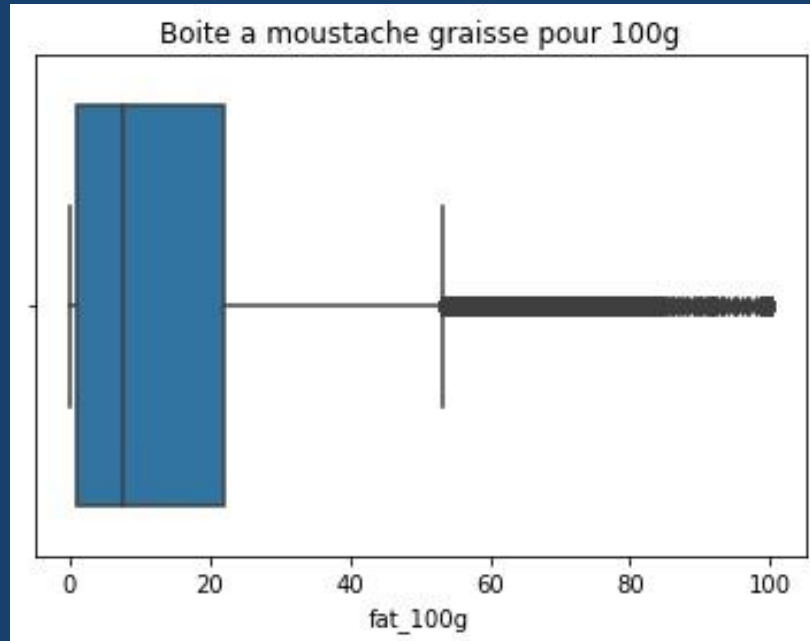
Médiane 1095 kj

Ecart-type 737



analyse exploratoire

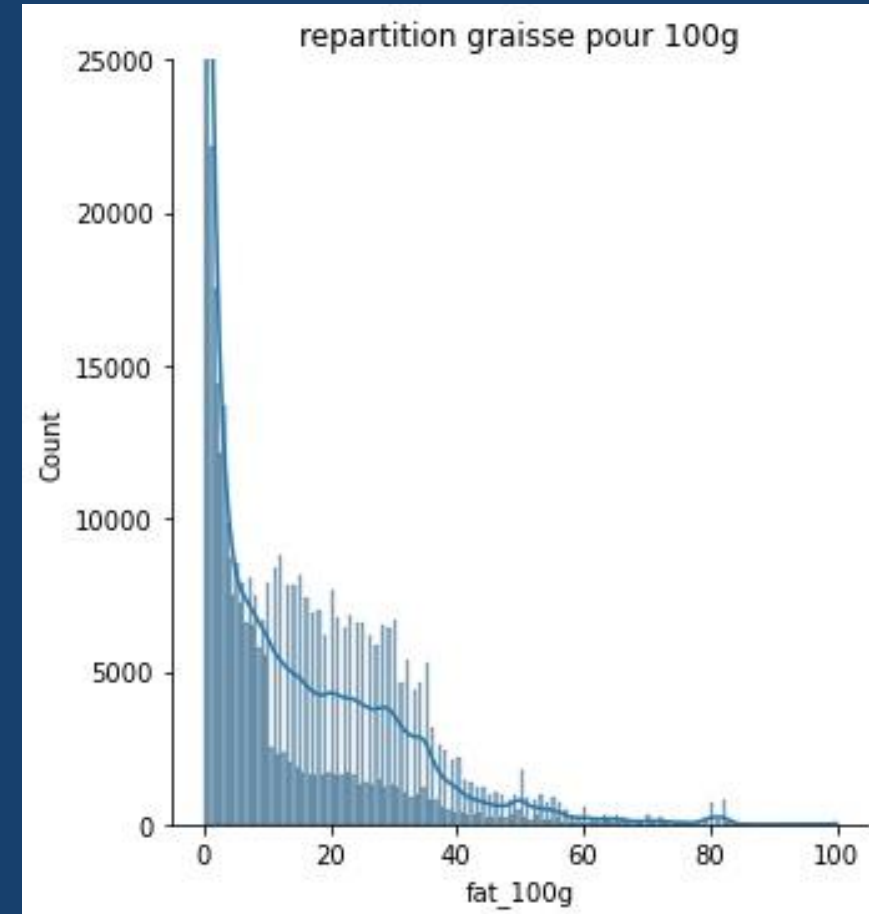
analyse univariée graisse fat_100g



Moyenne 13g

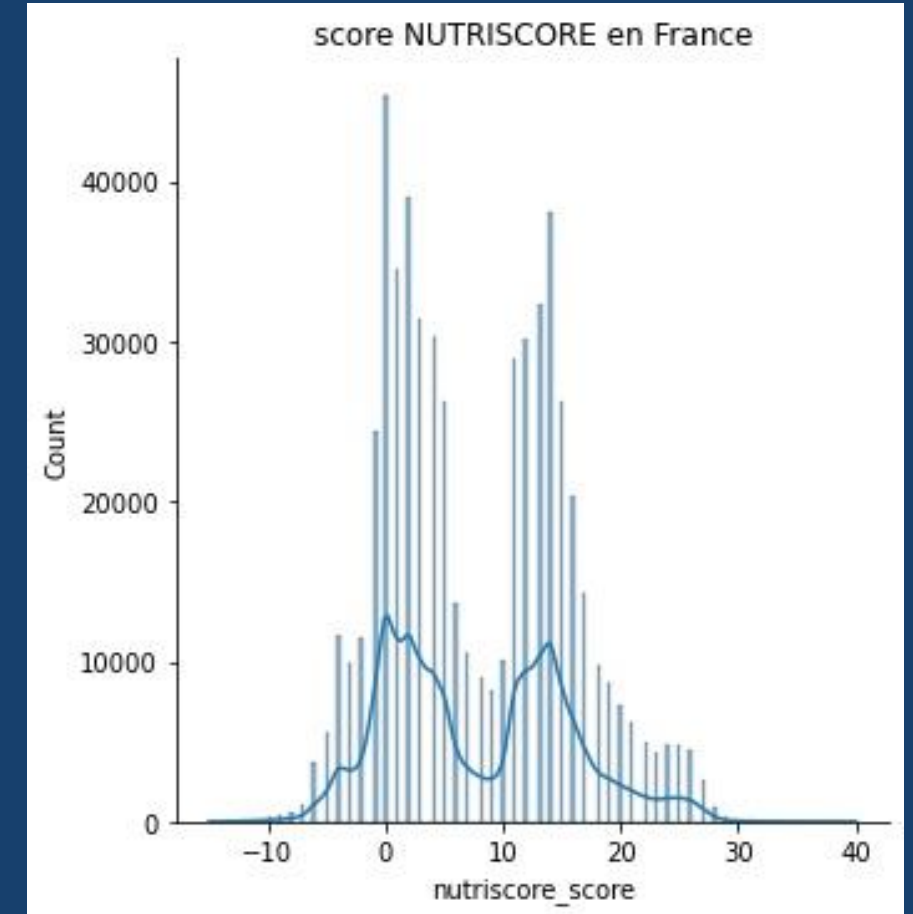
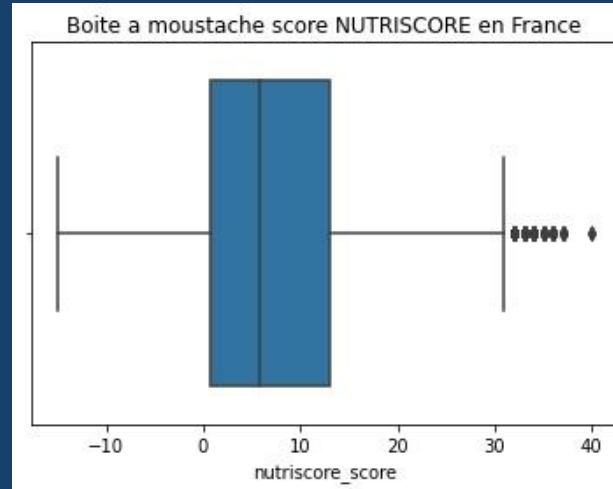
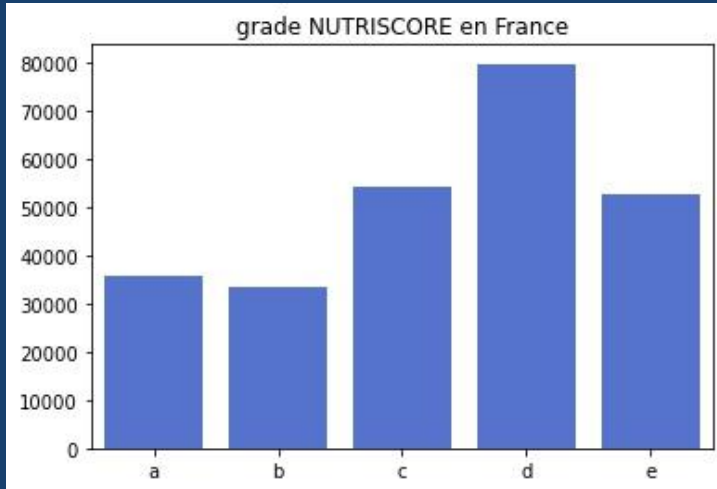
Médiane 7,5g

Ecart-type 14,8g



analyse exploratoire

analyse univariée nutriscore



Moyenne 7,3

Médiane 5,8

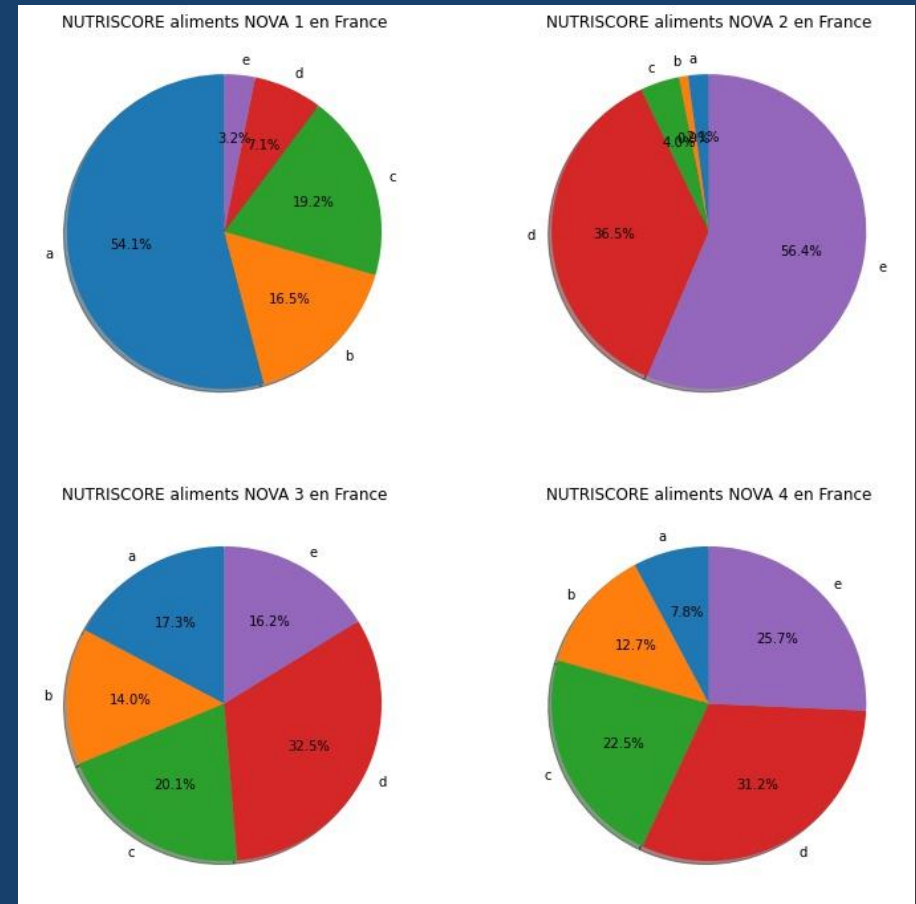
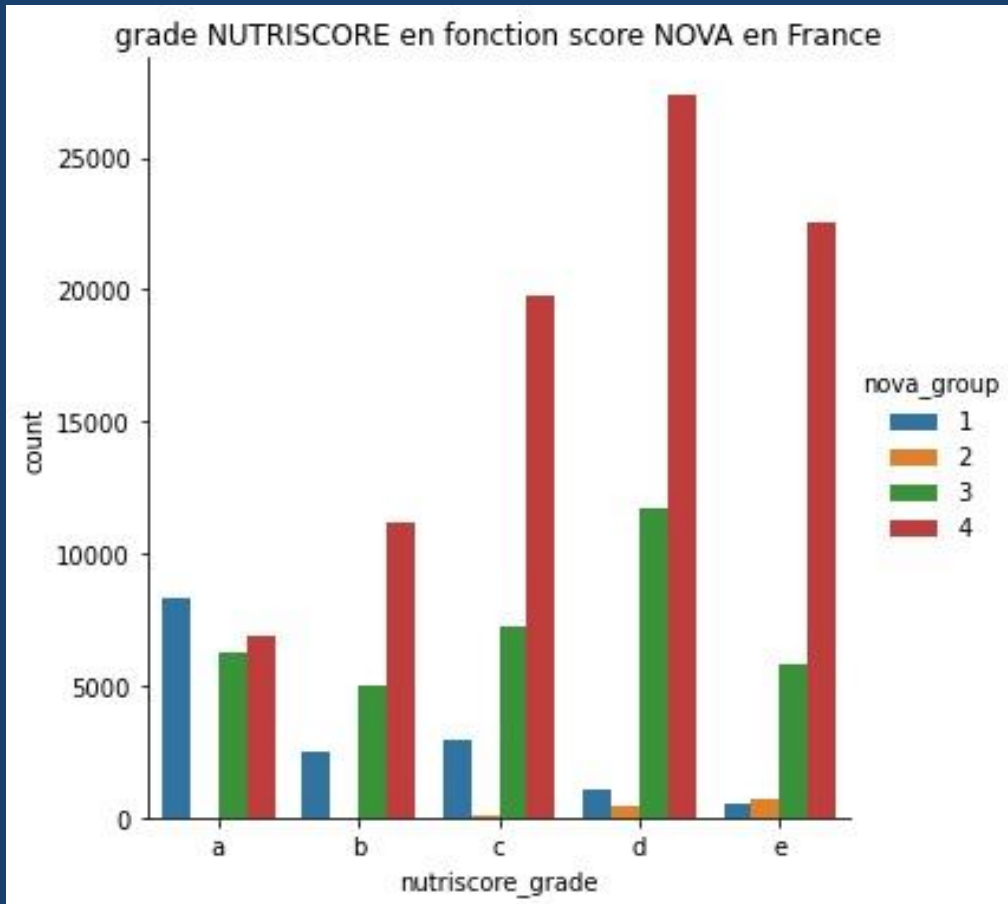
Ecart-type 7,8

Rappel A (<0), B(0-2), C(3-10), D(11-18), E(>19)

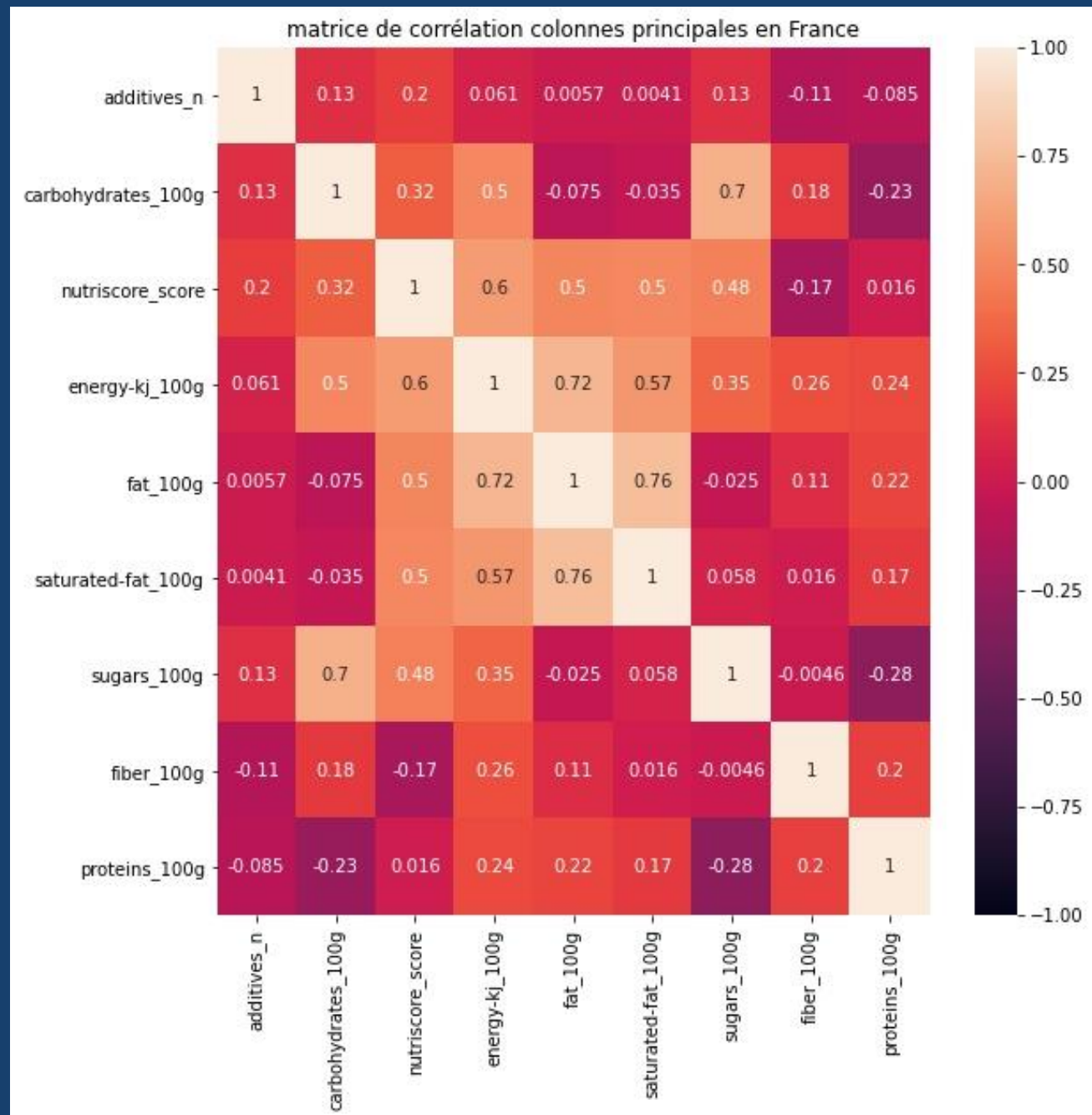


analyse exploratoire

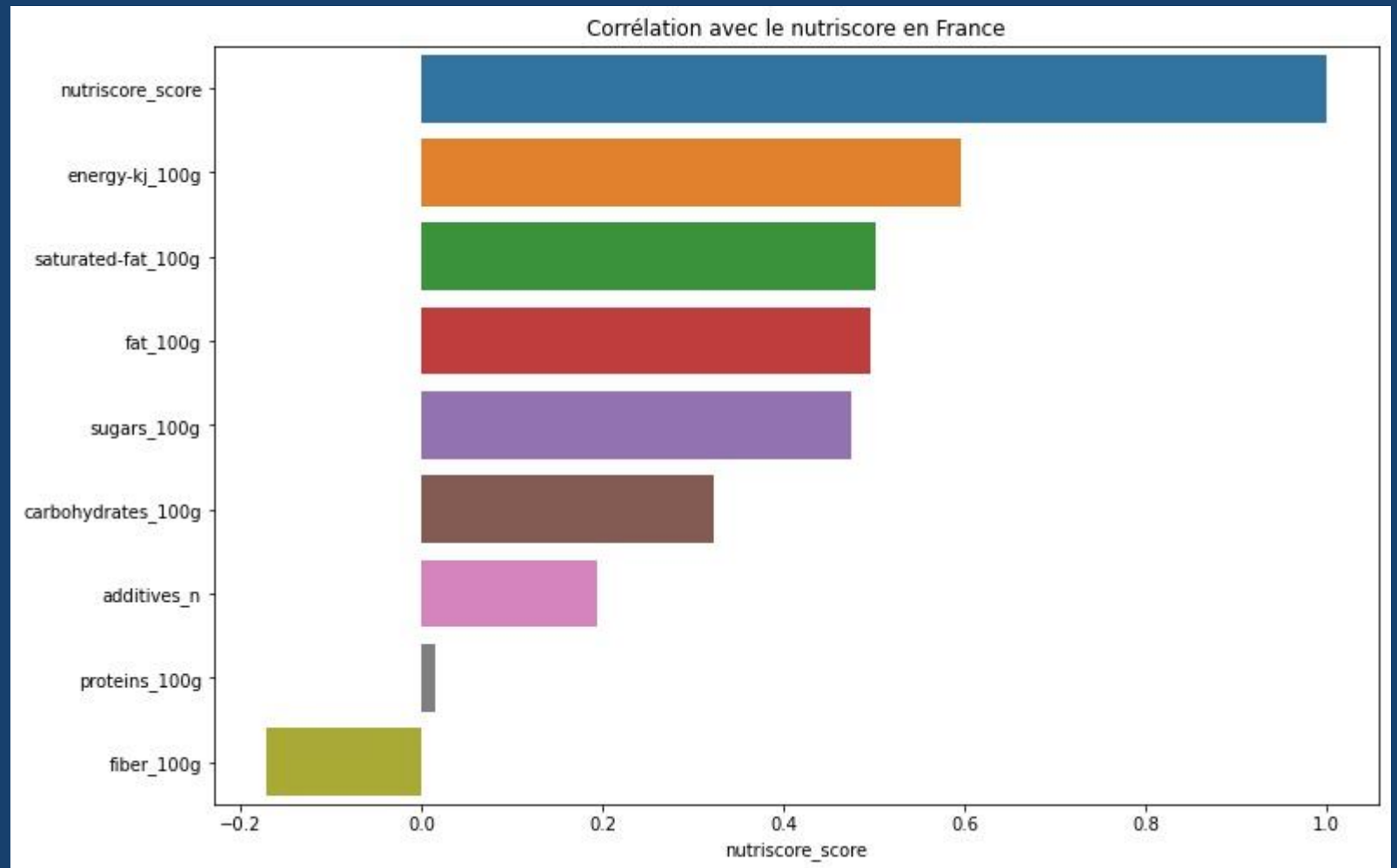
analyse bivariée nutriscore - NOVA



analyse exploratoire analyse bivariée corrélation



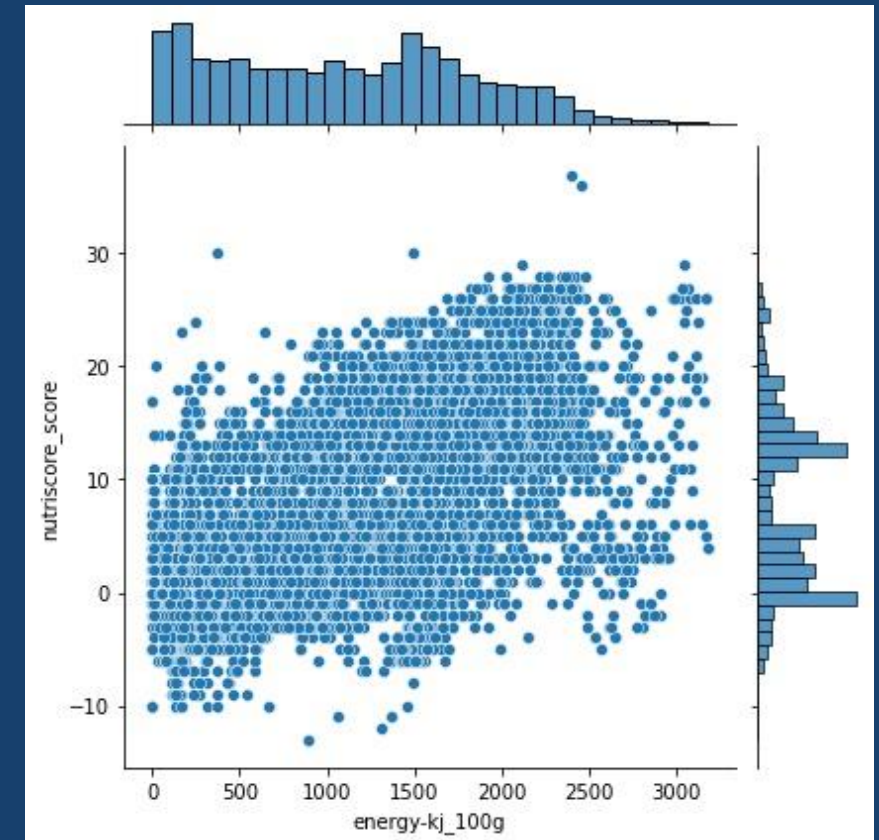
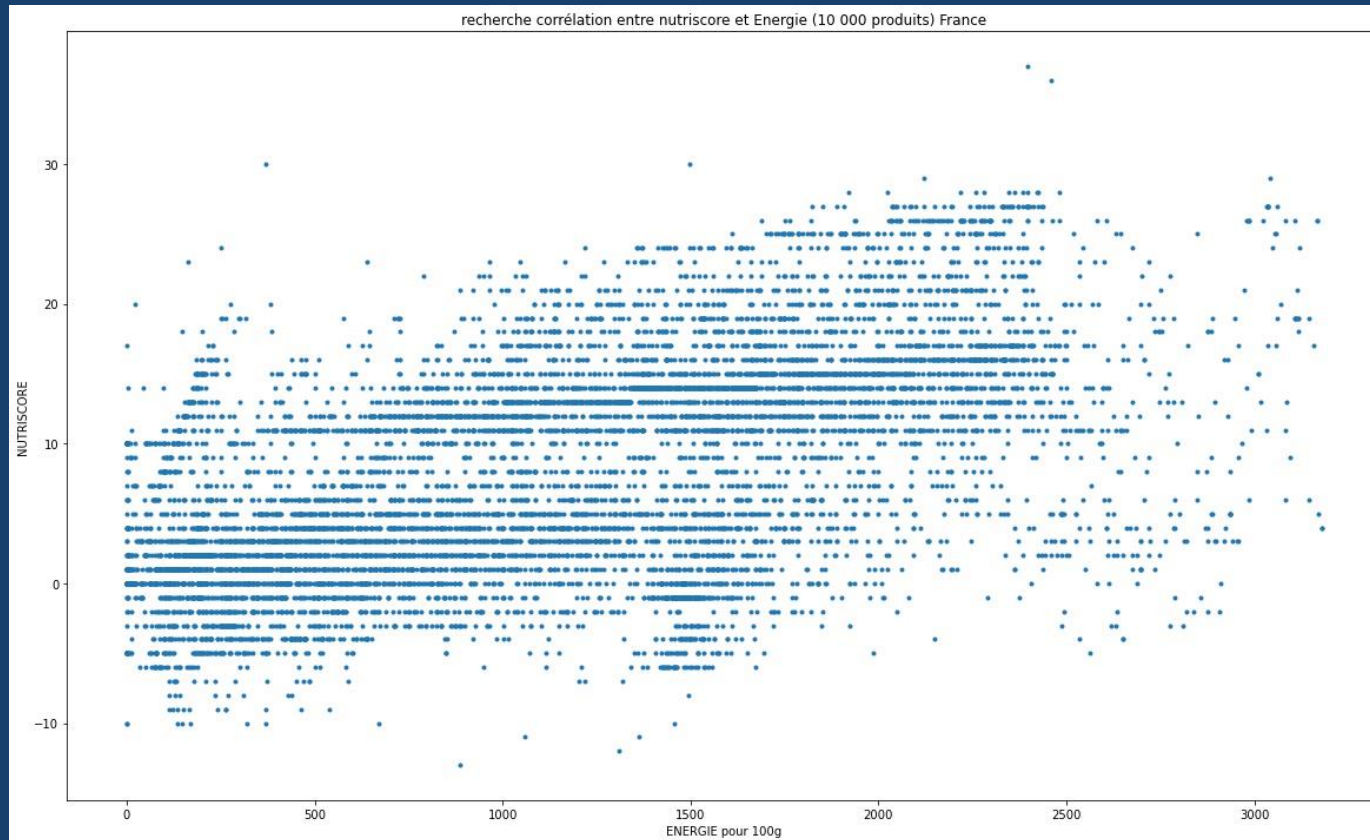
analyse
exploratoire
analyse bivariée
corrélation
nutriscore



analyse exploratoire

analyse bivariée corrélation 0,64

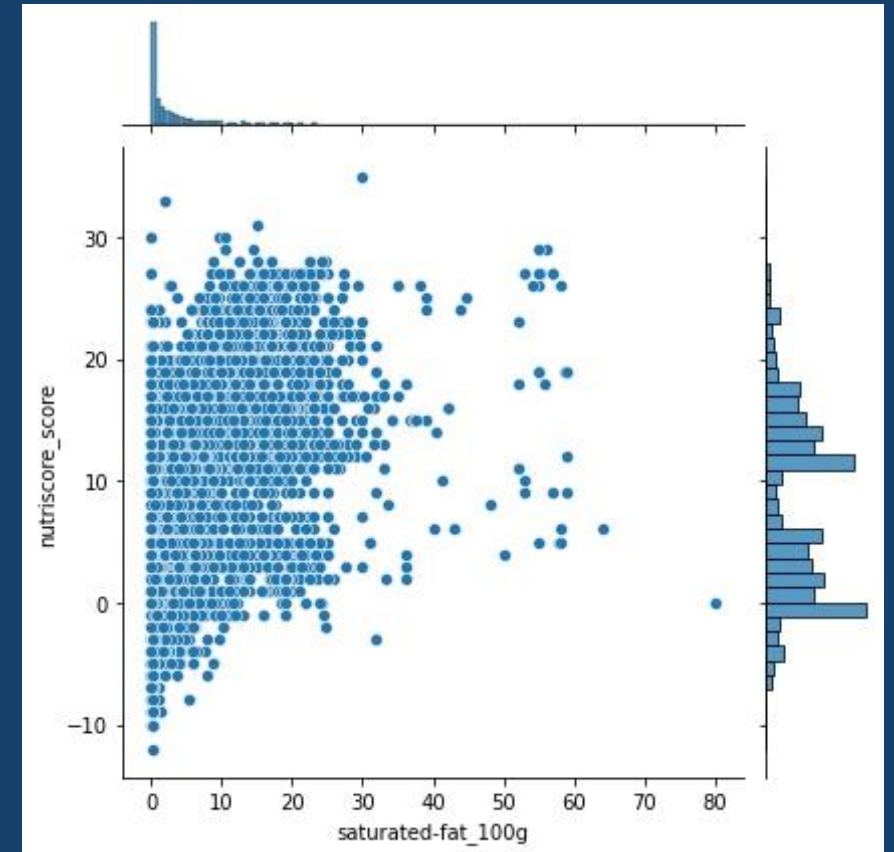
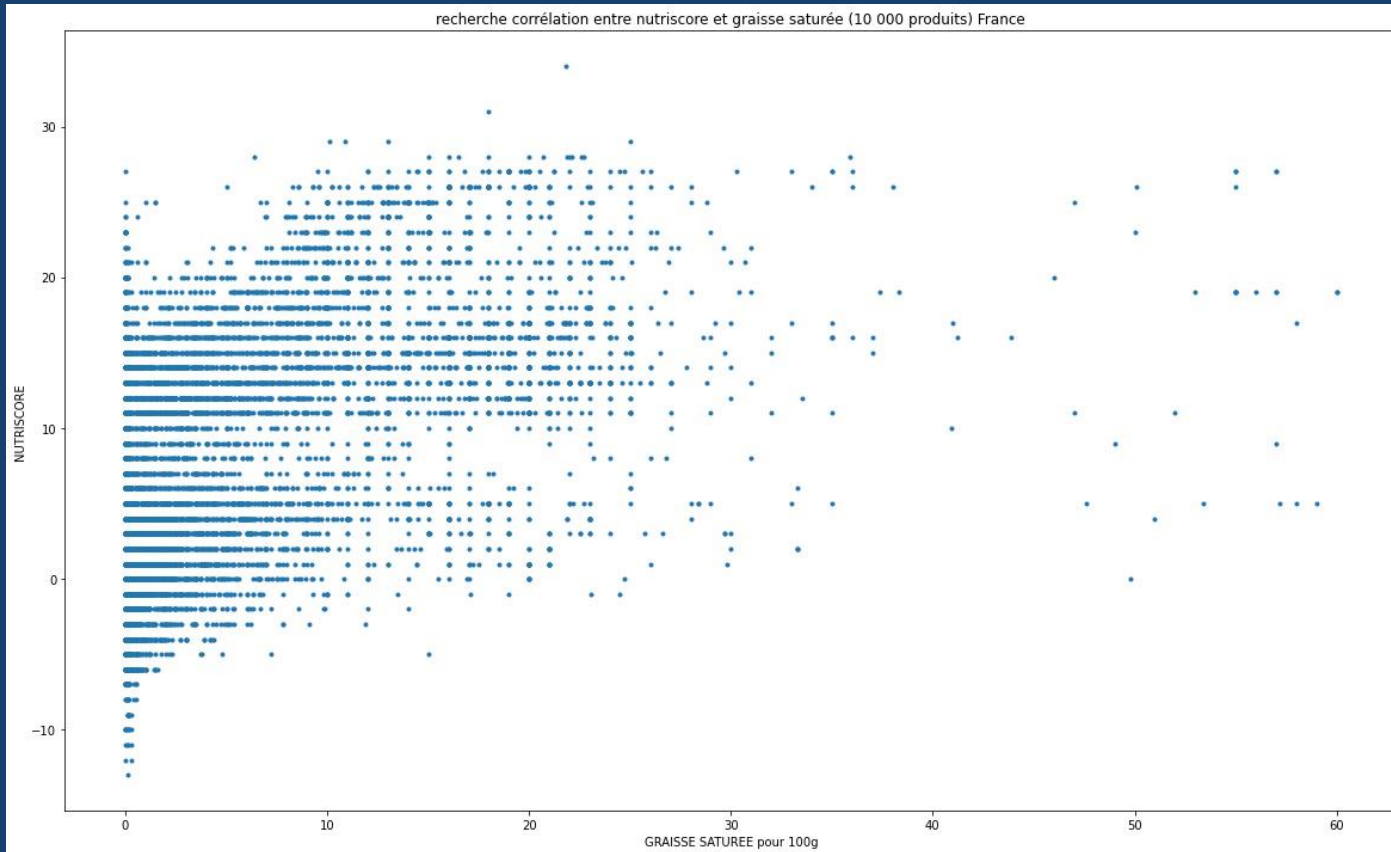
nutriscore – energy-kj



analyse exploratoire

analyse bivariée corrélation 0,61

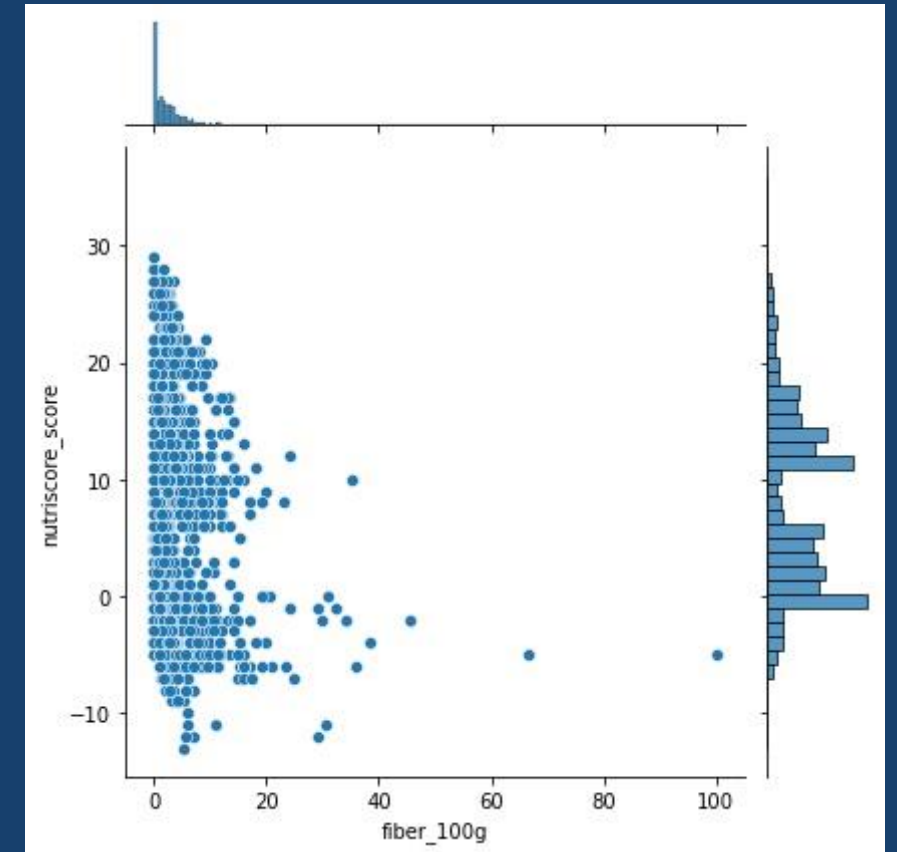
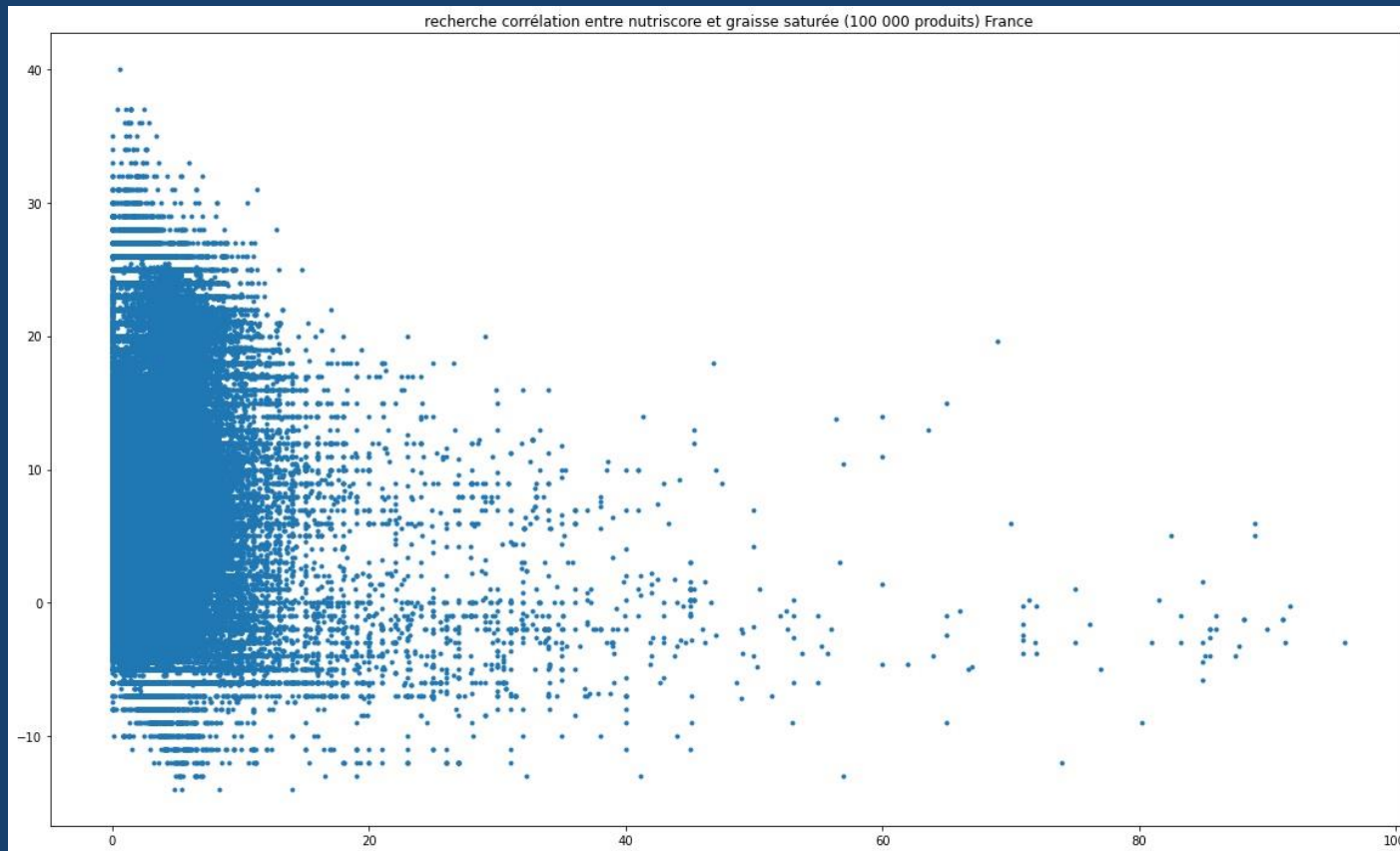
nutriscore – saturated-fat_100g



analyse exploratoire

analyse bivariée corrélation -0,17

nutriscore – fiber_100g

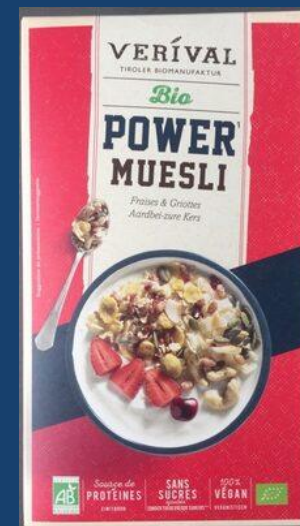


Notre application de conseils exemple 1

Un sportif de haut niveau recherche un produit pour la prise de masse avec de la fraise.

Il souhaite de plus un bon nutriscore (A ou B) et le moins transformé possible

Le résultat fait apparaitre en priorité des muesli



Notre application de conseils exemple 2

Pour une femme enceinte, on va rechercher la vitamine B12, B9, le calcium et le fer.

On atteint là les limites du jeu de données car ces colonnes sont renseignées dans moins de 1% des cas voir 0,25 % pour les vitamines.

J'ai fait une requête dans les produits en cherchant à maximiser le fer et le nutriscore.



Notre application de conseils exemple 3

Un malade cardiaque va chercher un produit sans sel et peu de sucres, avec toujours un bon nutriscore.



Notre application de conseils exemple 4

Un client a scanné un produit : des BNs chocolat et on va chercher un remplaçant de meilleure qualité nutritionnelle.





MERCI

Questions et Réponses