Objectifs :

Traiter les données alimentaires de manière à pouvoir :

- Prédire le nutriscore d'un plat ou d'un aliment en fonction de la proportion de ce qu'ils contiennent.

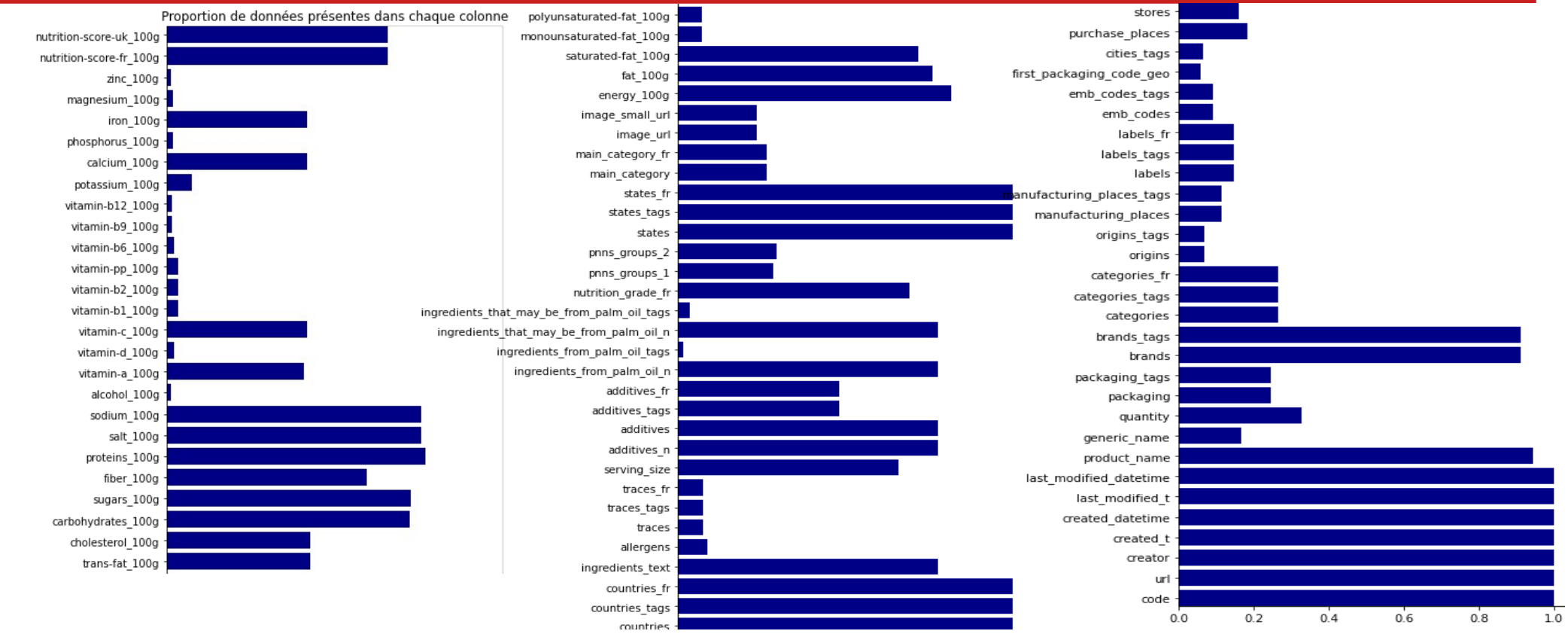
Description des données :

Base de donnée : - information sur des produits alimentaires
(en.openfoodfacts.org.products)

- taille (320 772, 162)

Taille base de donnée après suppression des doublons : - (320 750 x 162)

Visualisation des données



Nettoyage des données

Conservation uniquement des variables dont le remplissage est supérieur à 1 %

Application de règles métiers.

Objectifs : retirer les valeurs aberrantes

energy_100g <= 3700

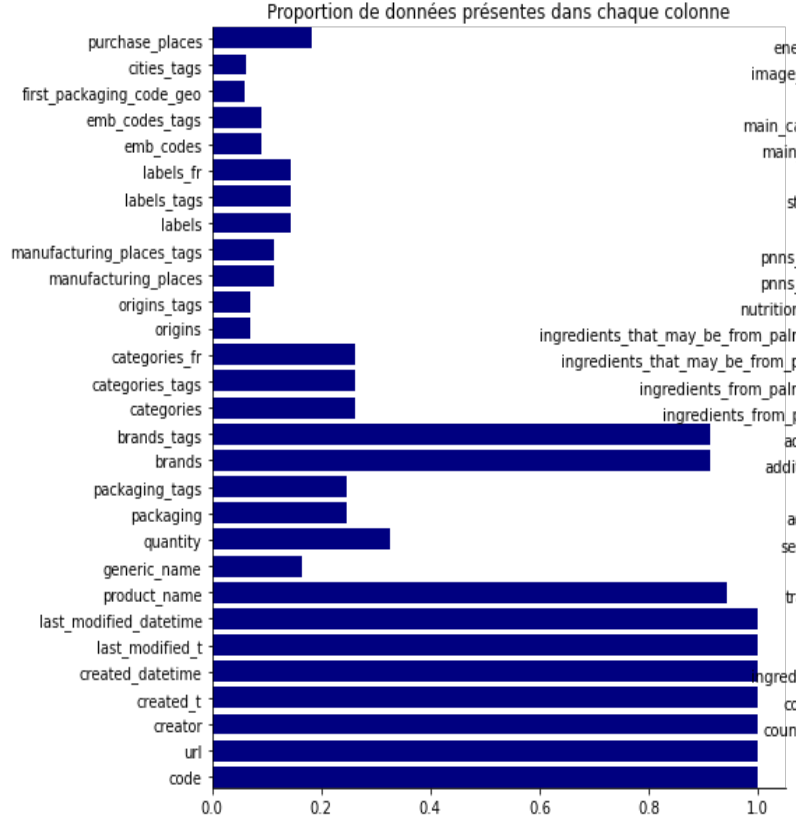
saturated_fat_100g <= fat_100g

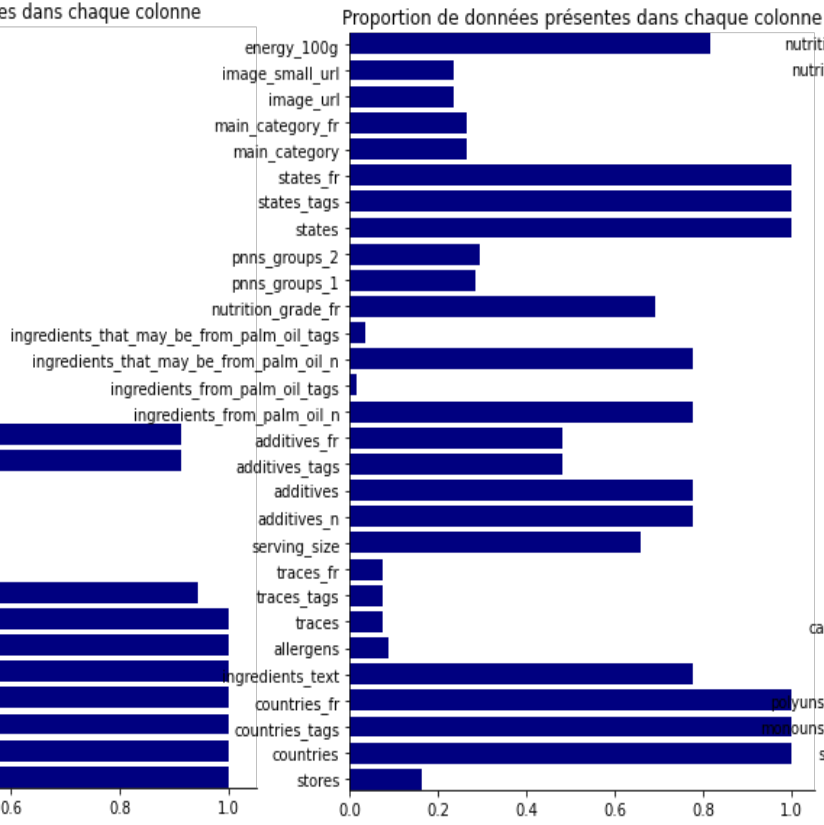
sugars_100g <= carbohydrates_100g

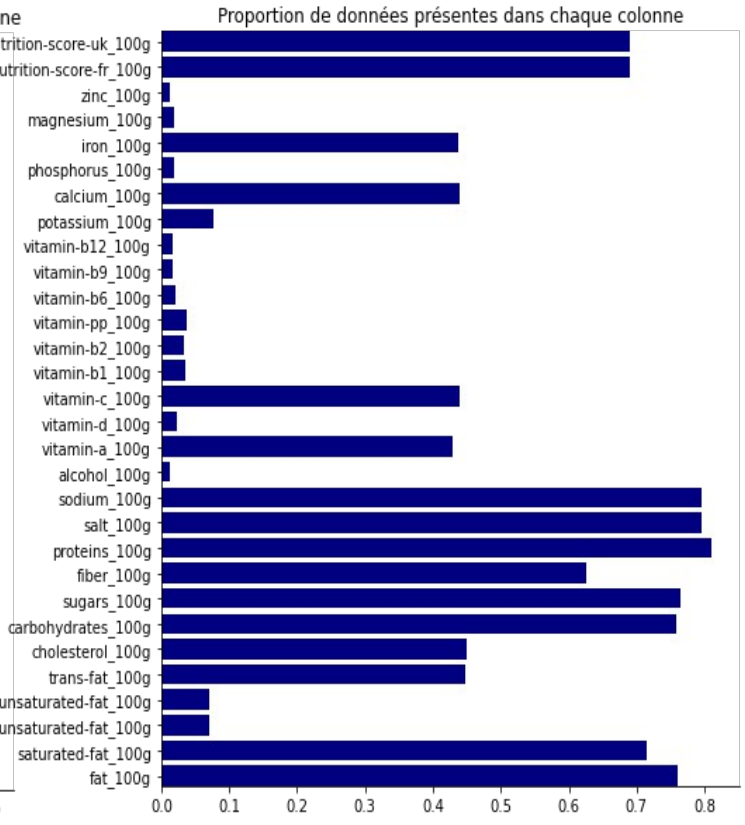
sodium_100g <= salt_100g

(fat_100g + carbohydrates_100g + proteins_100g + salt_100g <= 100

2 Données après netoyage







Algorithme du K-NN

Classe que l'algorithme doit déterminer :

nutriscore_grade_fr

Variables sur lesquelles il peut se baser :

energy_100g, fat_100g, saturated-fat_100g, salt_100g, sodium_100g, proteins_100g, carbohydrates_100g, sugars_100g, iron_100g

Échantillonnage : on prend 15 000 aliments aléatoires

Sur ces 15 000 aliments, 80 % servent à entraîner l'algorithme, 20 % à tester la précision de l'algorithme