

## Projet 3 : Concevez une application au service de la santé publique

Lancelot LECLERCQ

novembre 2021

# Problématique

- Application au service de la santé publique
- Idées d'application en lien avec l'alimentation



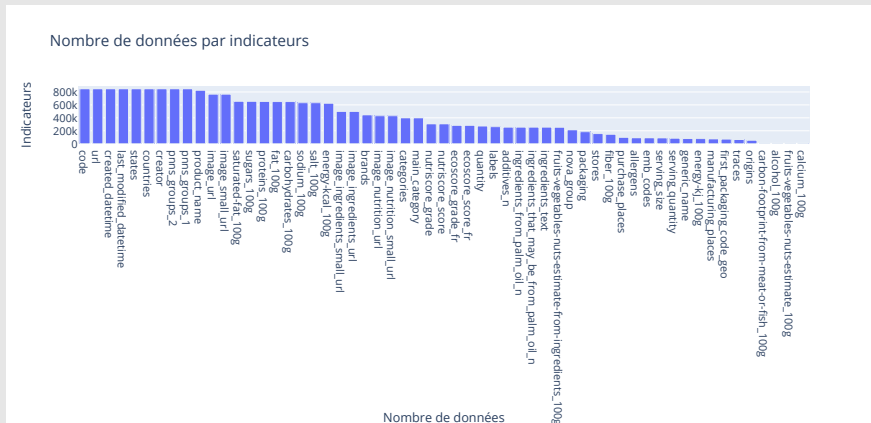
# Jeu de données

- Jeu de données issu de Open Food Facts
- Base de donnée ouverte et participative
- Repertoire les ingrédients, les allergènes, composition nutritionnelle, les labels, l'origine, etc
- Grosse base de données (environ 4,5 Go)



# Produits vendus en France

- Sélection des produits vendus en France
- 850 000 entrées sur les 2 milliards d'origine
- Suppression des colonnes dans lesquelles on a moins de 1% de données et des colonnes contenant des tags et des versions anglaises qui sont redondantes avec les colonnes du même nom
- 57 colonnes restantes sur les 186 d'origine



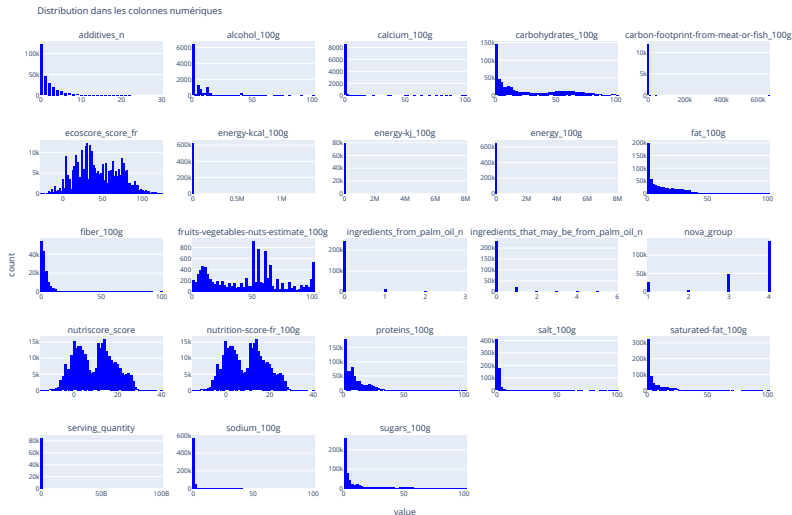
# Nettoyage des outliers

## - Valeurs extrêmes pour :

- Quantité servie
- Empreinte carbone
- Énergie

## - Des valeurs très importantes là où le nombre de valeurs est très faible

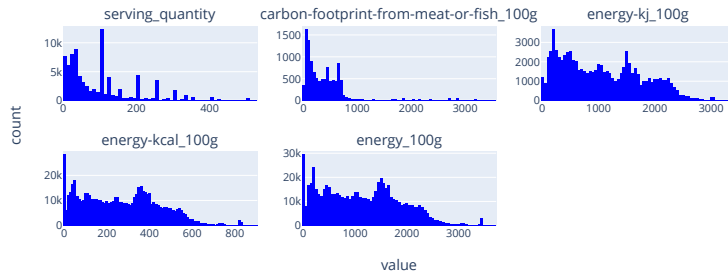
## - Nous écartons ces valeurs lorsqu'elles sont en dehors du quartile qui contient 99% des valeurs.



# Nettoyage des outliers

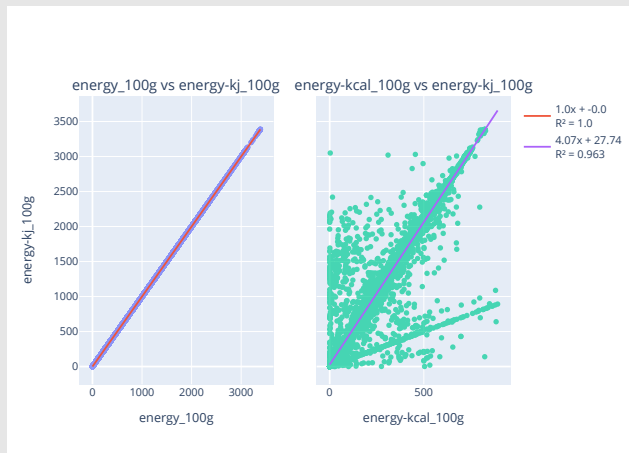
- On obtient des distributions plus intéressantes
- Fonctionnement assez similaire des différents types d'énergies

Distribution dans les colonnes quantité, empreinte carbone et energies



# Données redondantes

- Mêmes valeurs energy\_100g et energy-kj\_100g
- Variation pour ce qui est de l'énergie en kcal pour 100g
- Nous n'allons garder que les colonnes avec des unités d'énergie : kcal et kj
- De même nous allons supprimer les colonnes finissant par \_tags et \_en qui contiennent des données redondantes des colonnes du même nom sans suffixes



- Pouvoir scanner un produit et trouver des alternatives similaires de meilleure qualité
  - Meilleures pour la santé (nutriscore plus élevé)
    - Avec moins d'additifs (nombre d'additifs inférieurs)
  - Meilleures pour l'environnement (ecoscore plus élevé)
    - Origine plus proche (France, Europe)
  - Voir les deux
- Pouvoir choisir selon un régime alimentaire particulier
  - Végétarien
  - Vegan
  - Halal
  - Kasher
  - Sans gluten
  - etc

Implémenter un système de reconnaissance d'image permettant d'extraire les données manquantes des photos des étiquettes