



VegiFit

NUTRITION TRACKER

Octave POUILLOT
Formation Data Scientist – Projet 3

UNE APPLICATION AU SERVICE DE LA SANTÉ



VegiFit, une application mobile permettant aux sportifs végétariens de surveiller leur apport nutritionnel et de s'assurer qu'ils consomment suffisamment de nutriments essentiels pour maintenir leur forme physique et leur santé.

SOMMAIRE

1

INTRODUCTION

VegiFit dans les grandes lignes

2

OPEN FOOD FACTS

Présentation de la base de données

3

NETTOYAGE DU DATASET

Focus sur les données pertinentes

4

ANALYSE DE FAISABILITÉ

Exploration des données

5

CONCLUSION

VegiFit, qu'en est-il ?



SECTION 1

INTRODUCTION



LE PROJET D'APPLICATION MOBILE

UN ACCOMPAGNEMENT PERSONNALISÉ

- Proposition de complément de repas après le scan d'un produit
- Affichage des informations nutritionnelles détaillée des produits proposés
- Recherche manuelle pour trouver des informations nutritionnelles détaillées
- Pédagogie personnalisée sur l'alimentation
- Suivi des apports nutritionnels

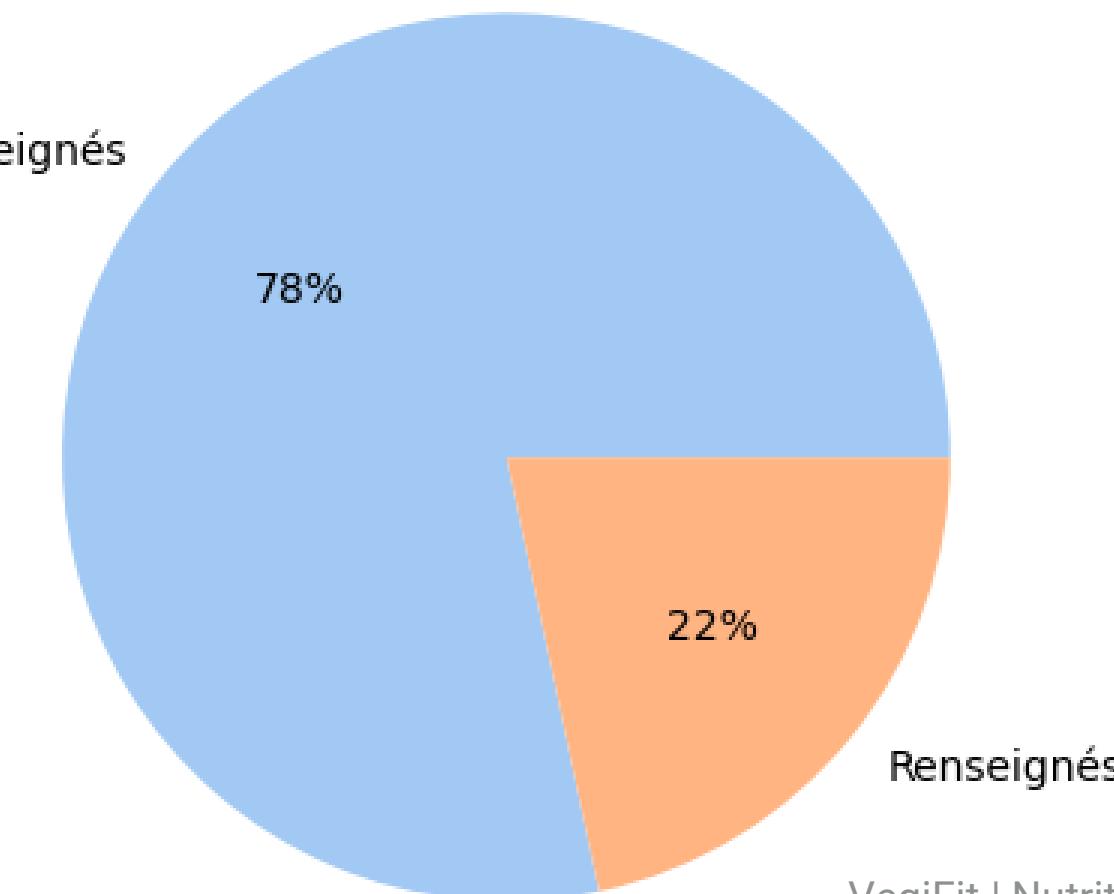


SECTION 2
OPEN FOOD FACTS

Base de données de produits alimentaires

- Plus de 2,5 millions de produits
- 201 indicateurs
- Différents type d'informations :
 - Code barre
 - Nom
 - Date (création, modification)
 - Marque
 - Tags (pays, catégories, ...)
 - Labels (bio, eu, ...)
 - Nutriscore
 - Ingrédients
 - Informations nutritionnelles

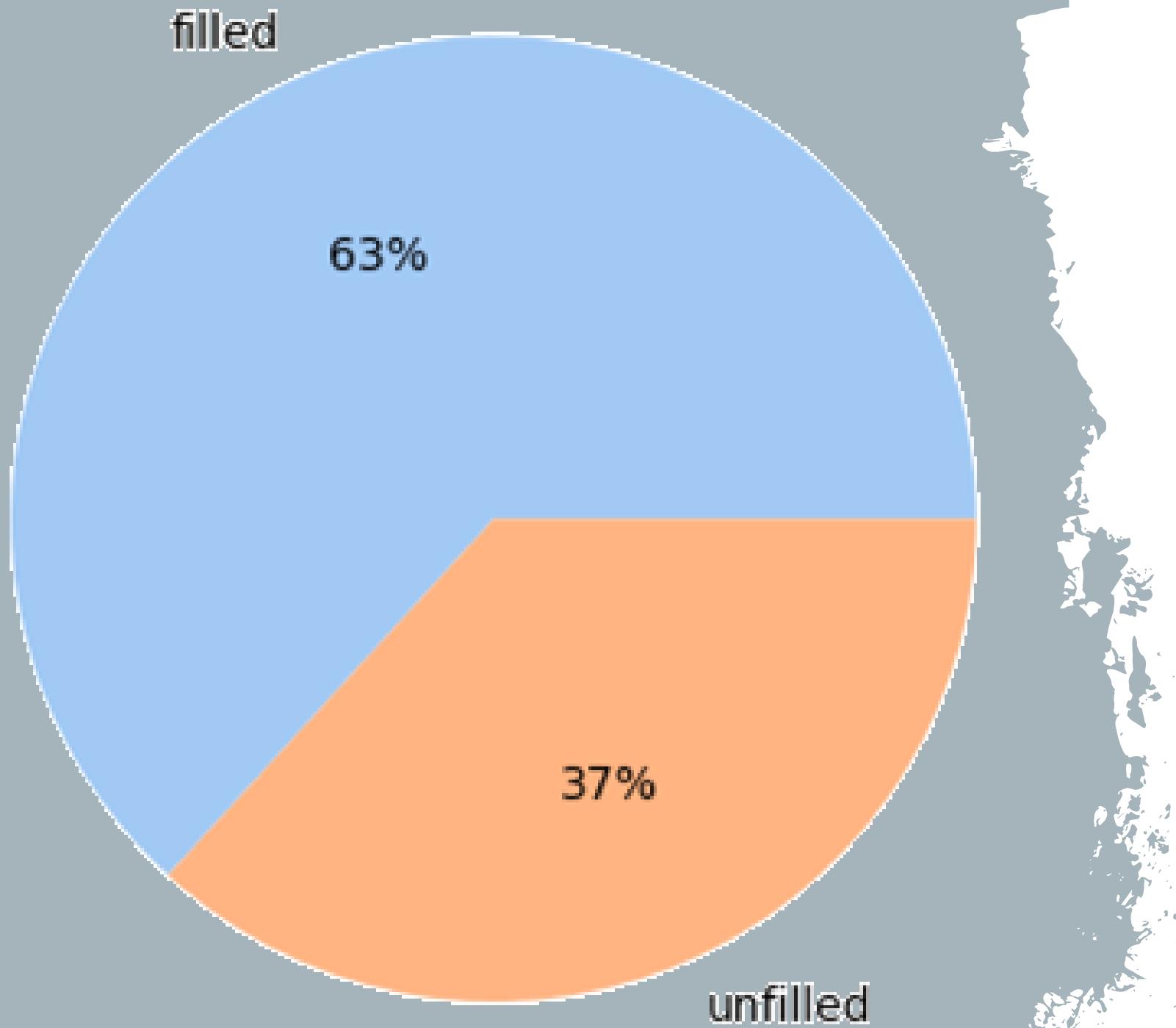
Taux de remplissage des données





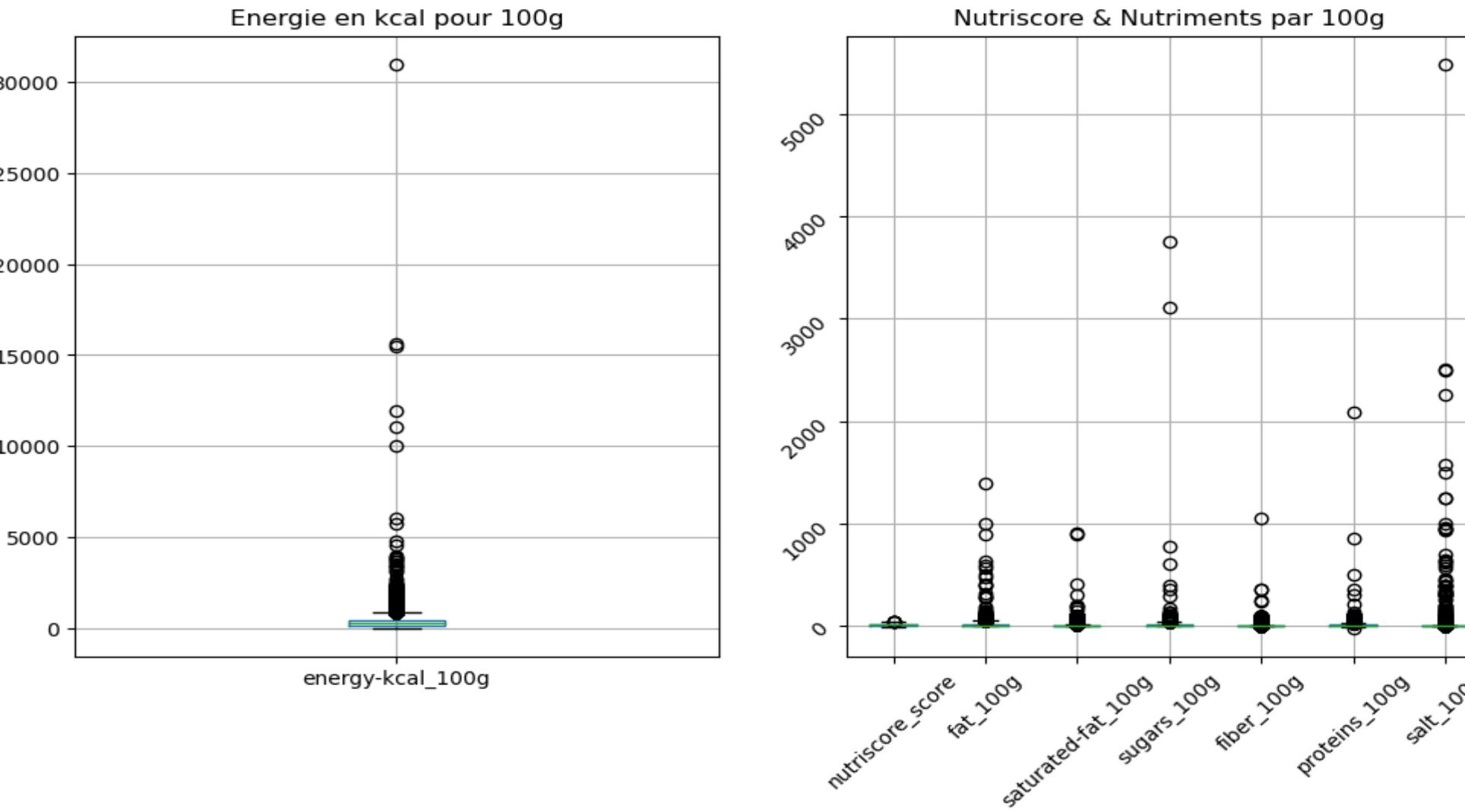
SECTION 3
NETTOYAGE DU DATASET

Taux de remplissage des données



RÉDUCTION DES DONNÉES

-
- Pays : France → 966 573 produits
 - Code barre EAN-13 uniquement → 894 028 produits
 - Features pertinents uniquement → Informations, labels, nutriscore, nutriments
 - Pas de doublons → 894 016 produits



VALEURS MANQUANTES & ABERRANTES

MANQUANTES

Si moins de 4 données renseignées (= 30%)

(code + nom + 2 variables)

→ 181 724 produits concernés

ABERRANTES

- Nutriments < 0g → 5 produits
- Nutriments > 100g → 113 produits

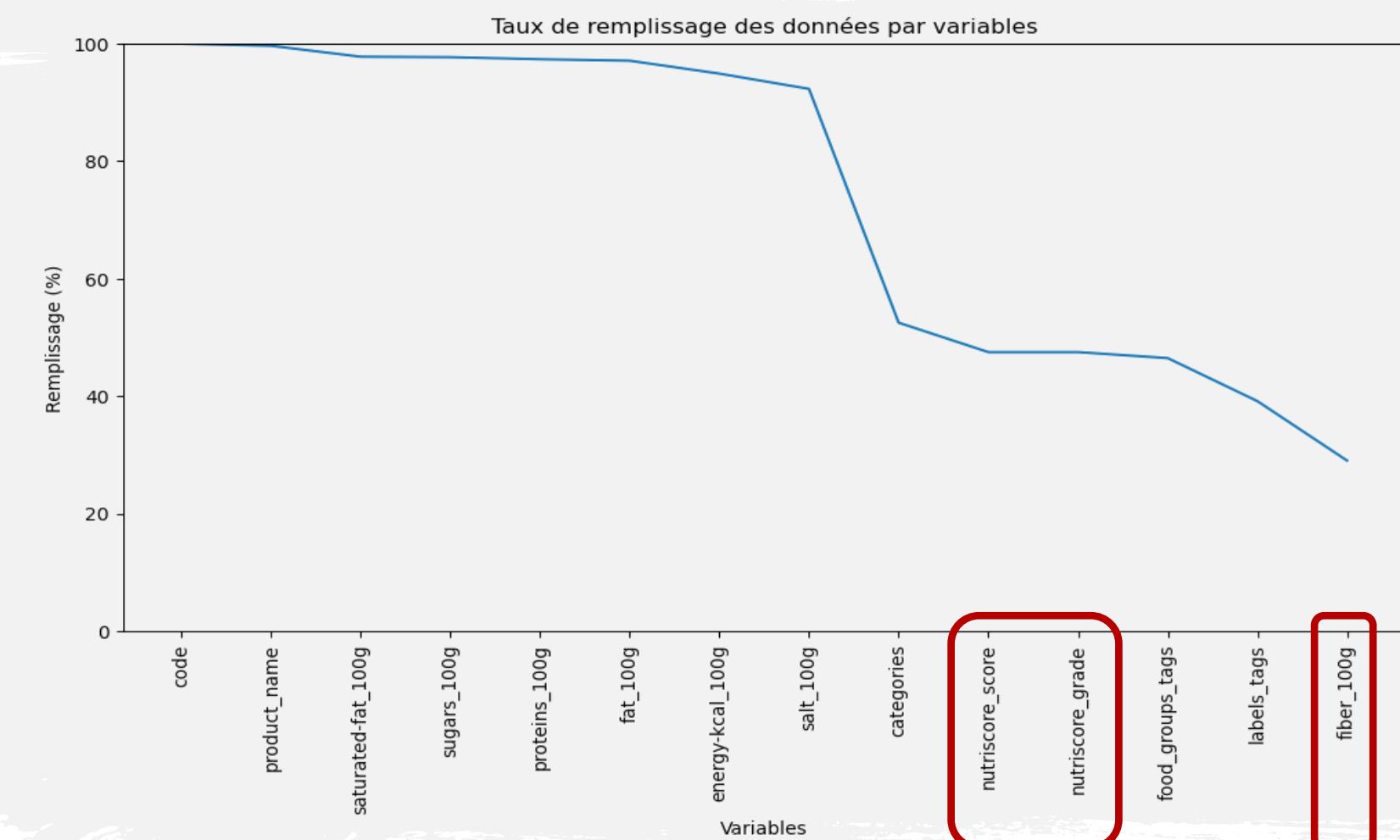
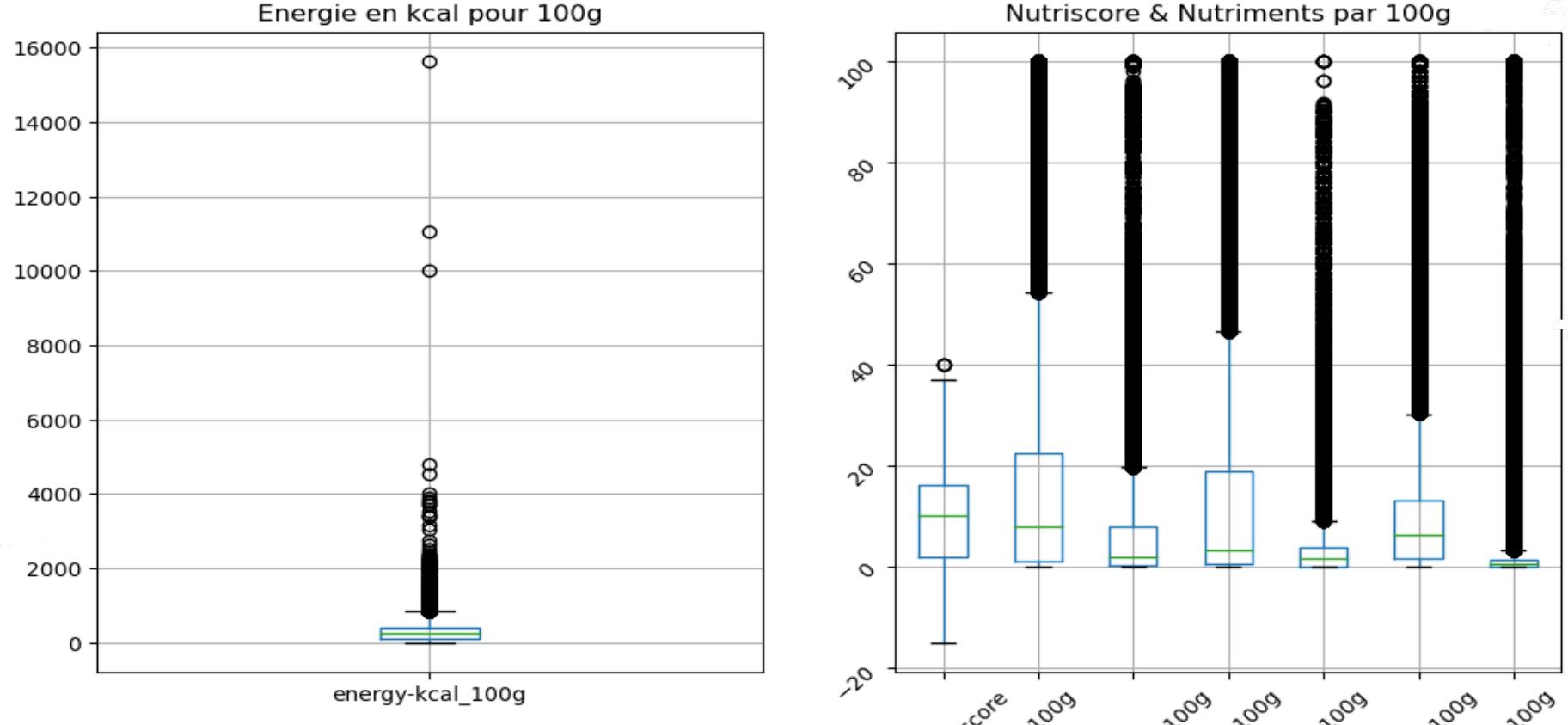
DISTRIBUTION FINALE

Le reste des outliers sont des valeurs atypiques
(protéines en poudre, sirops de sucres,)

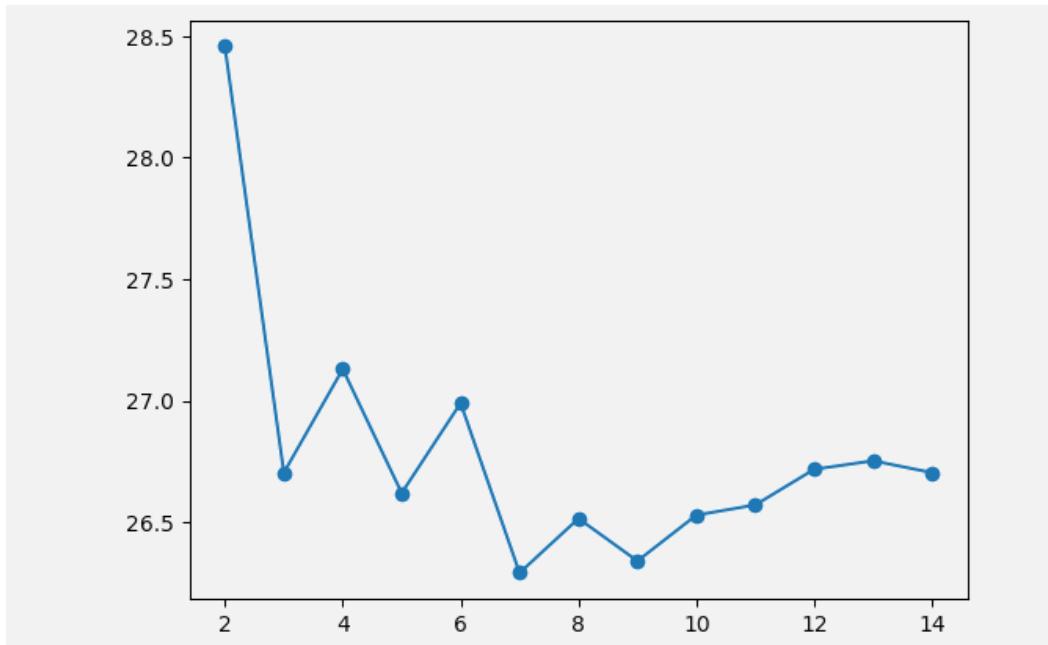
74% de valeurs renseignées

Variables peu renseignées :
Nutriscore
Fibres

TAUX DE REMPLISSAGE



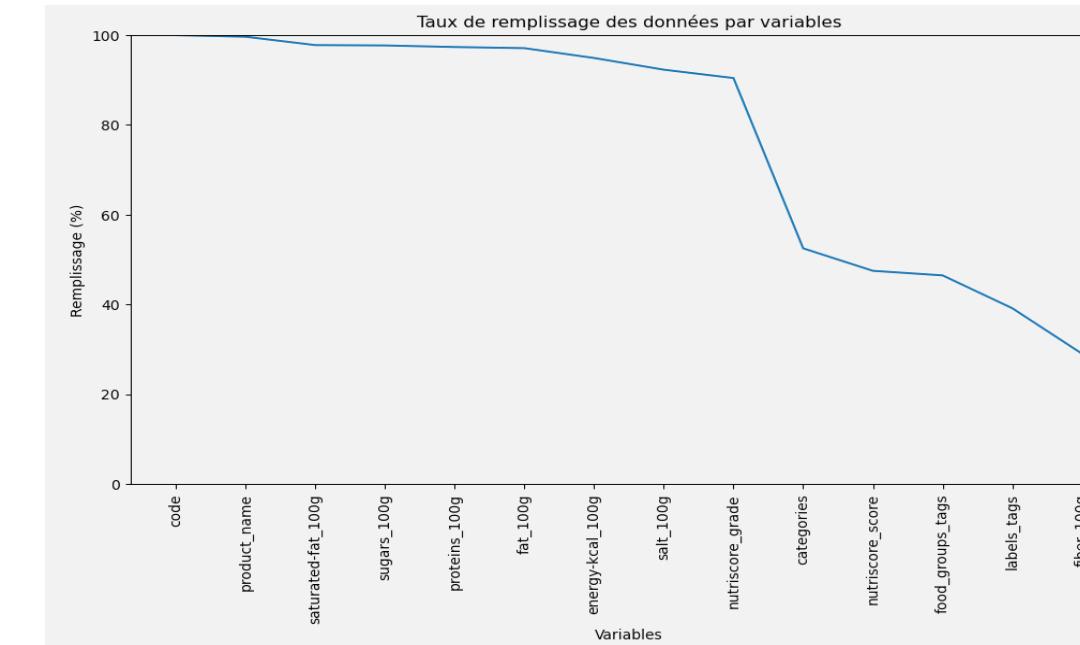
UTILISATION D'UN ALGORITHME KNN POUR NUTRISCORE



TAUX D'ERREUR

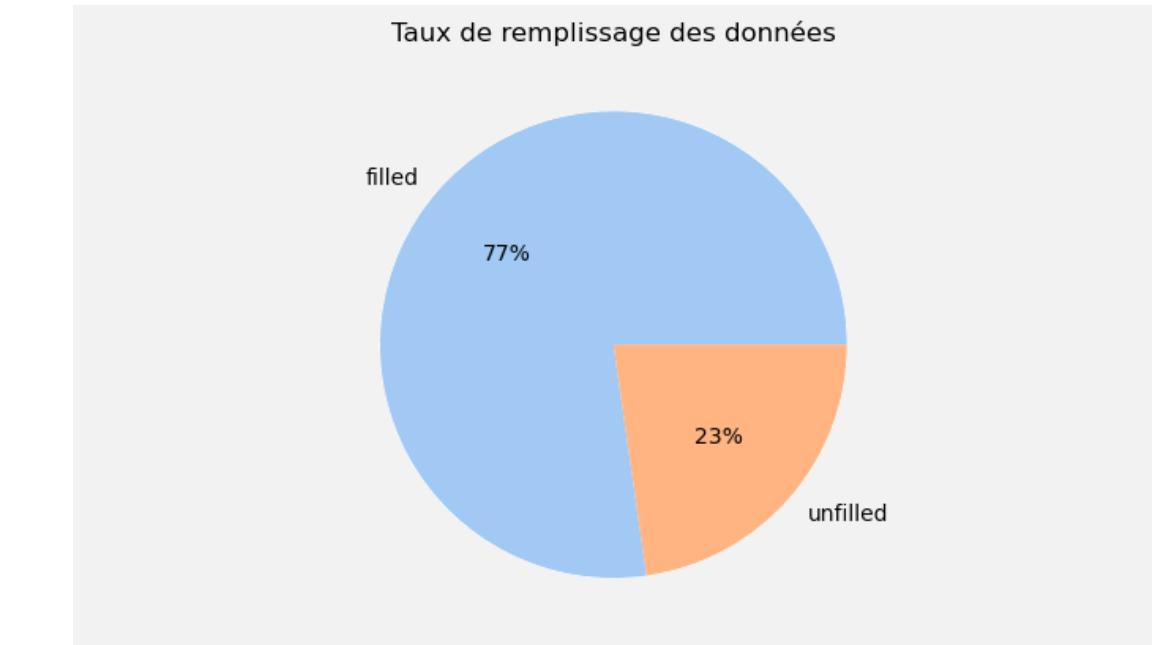
Test pour $k = 2$ à $k = 14$

Meilleurs performances pour 7-NN avec une erreur de ~26%



REmplissage APRÈS 7-NN

Nutriscore (grade) remonté au niveau de salt_100g (nutriment minimum)



REmplissage TOTAL

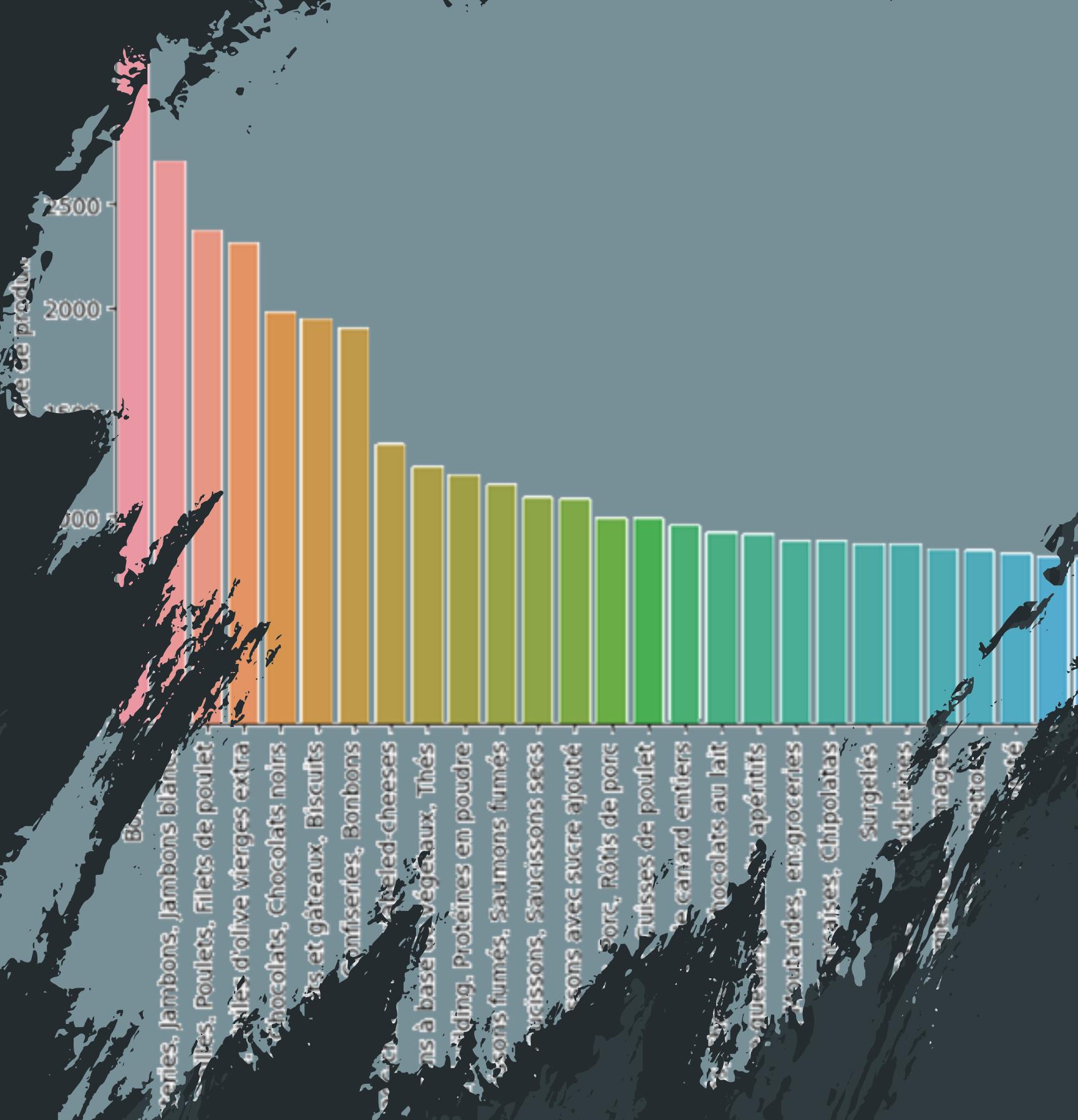
Taux de remplissage à plus de 75%



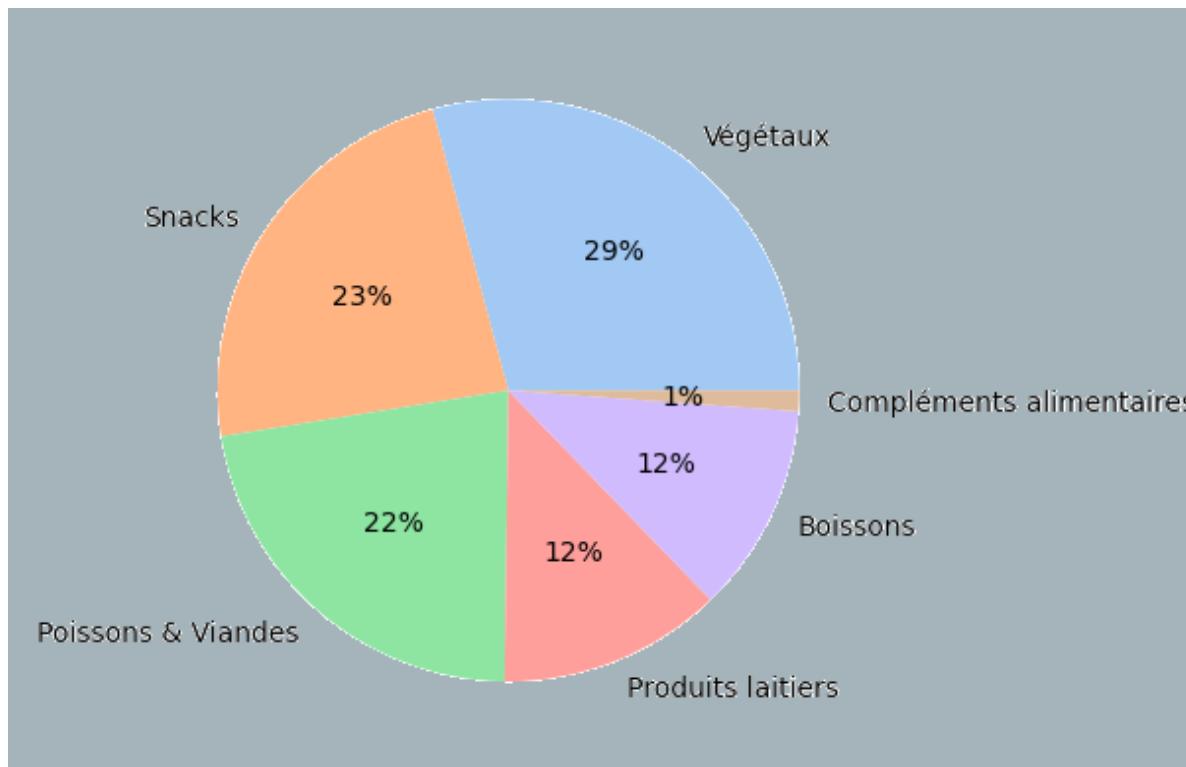
SECTION 4

ANALYSE DE FAISABILITÉ

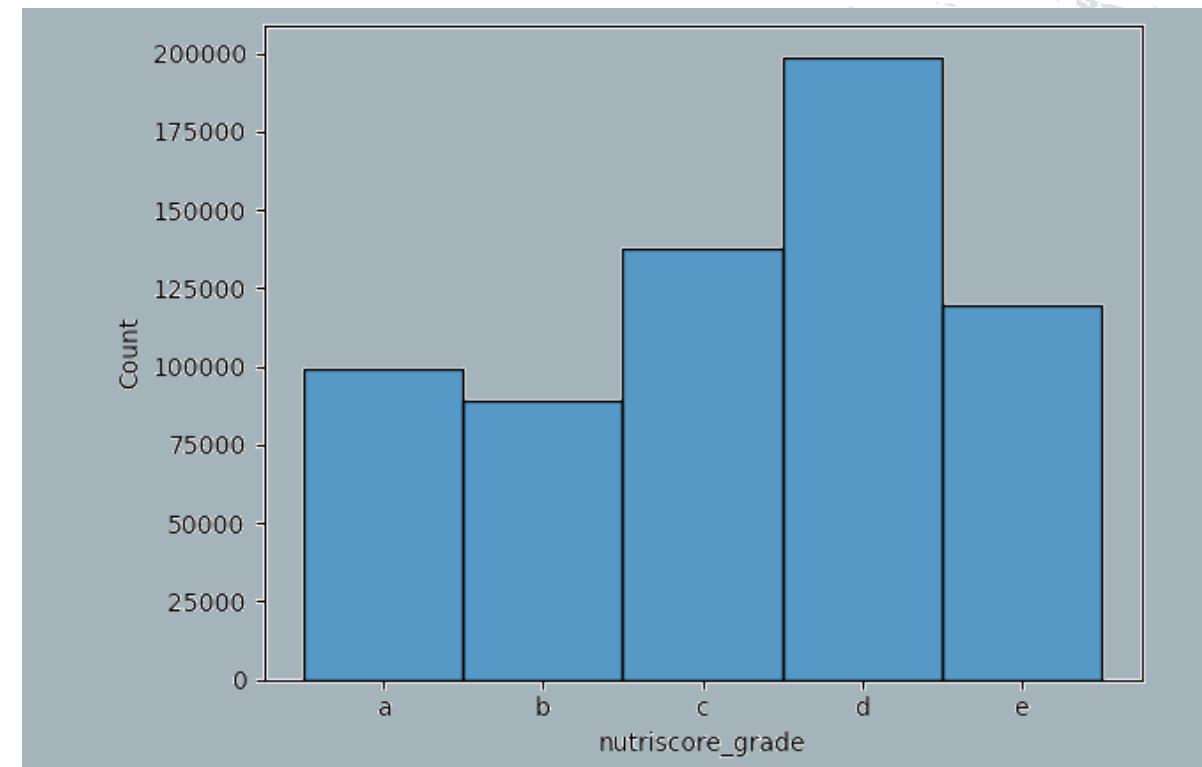
ANALYSE UNIVARIÉE



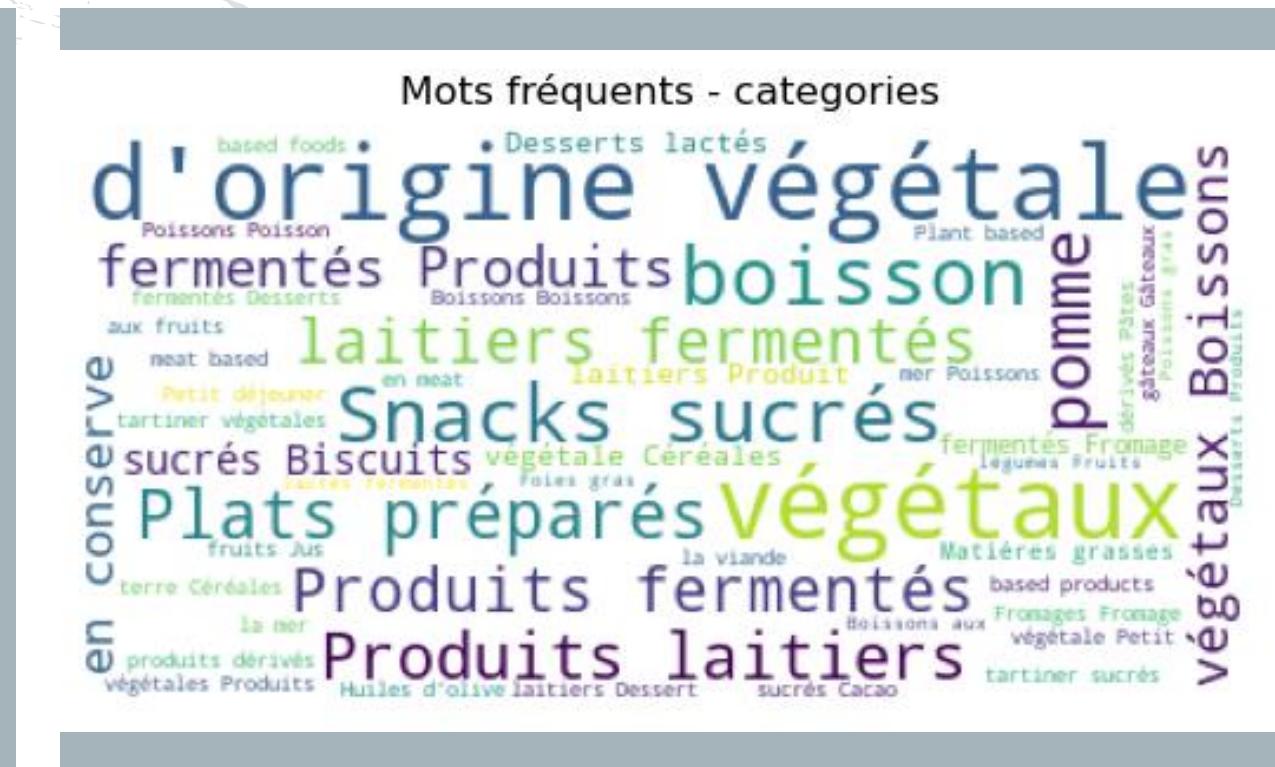
ANALYSE UNIVARIÉE



RÉPARTITION DES CATÉGORIES PRINCIPALES



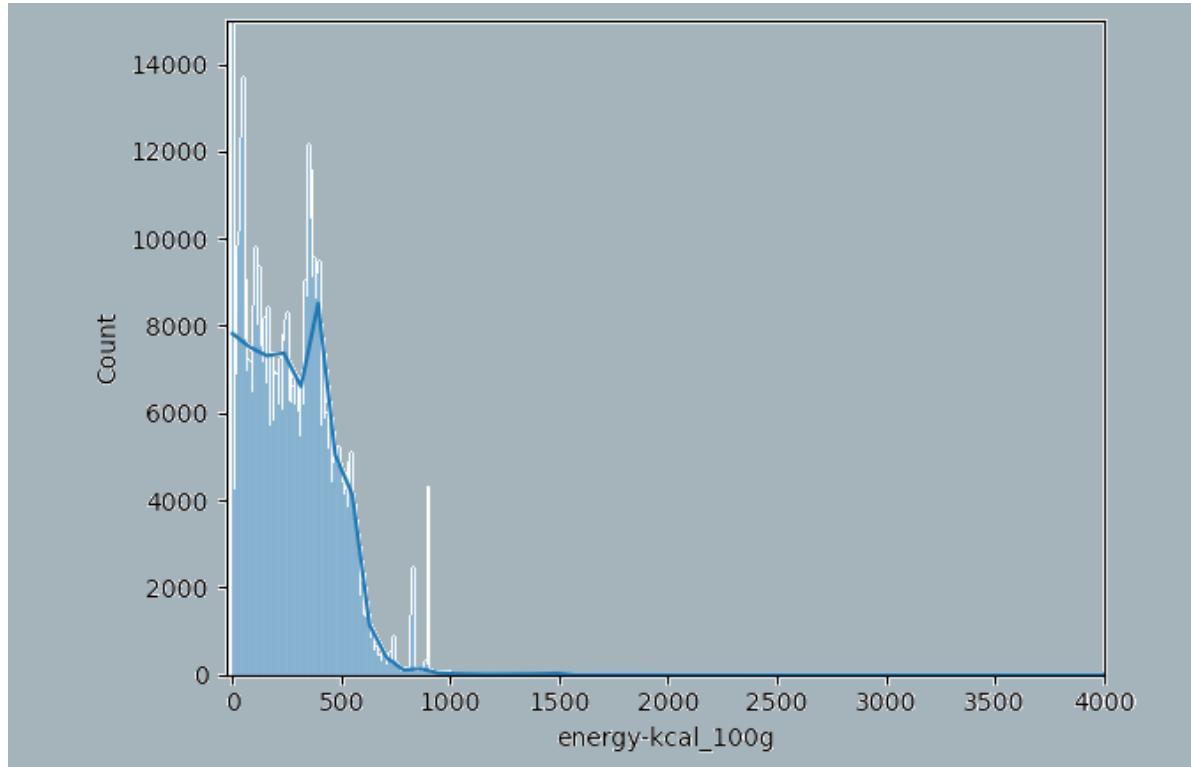
NOMBRE DE PRODUITS PAR NUTRISCORE



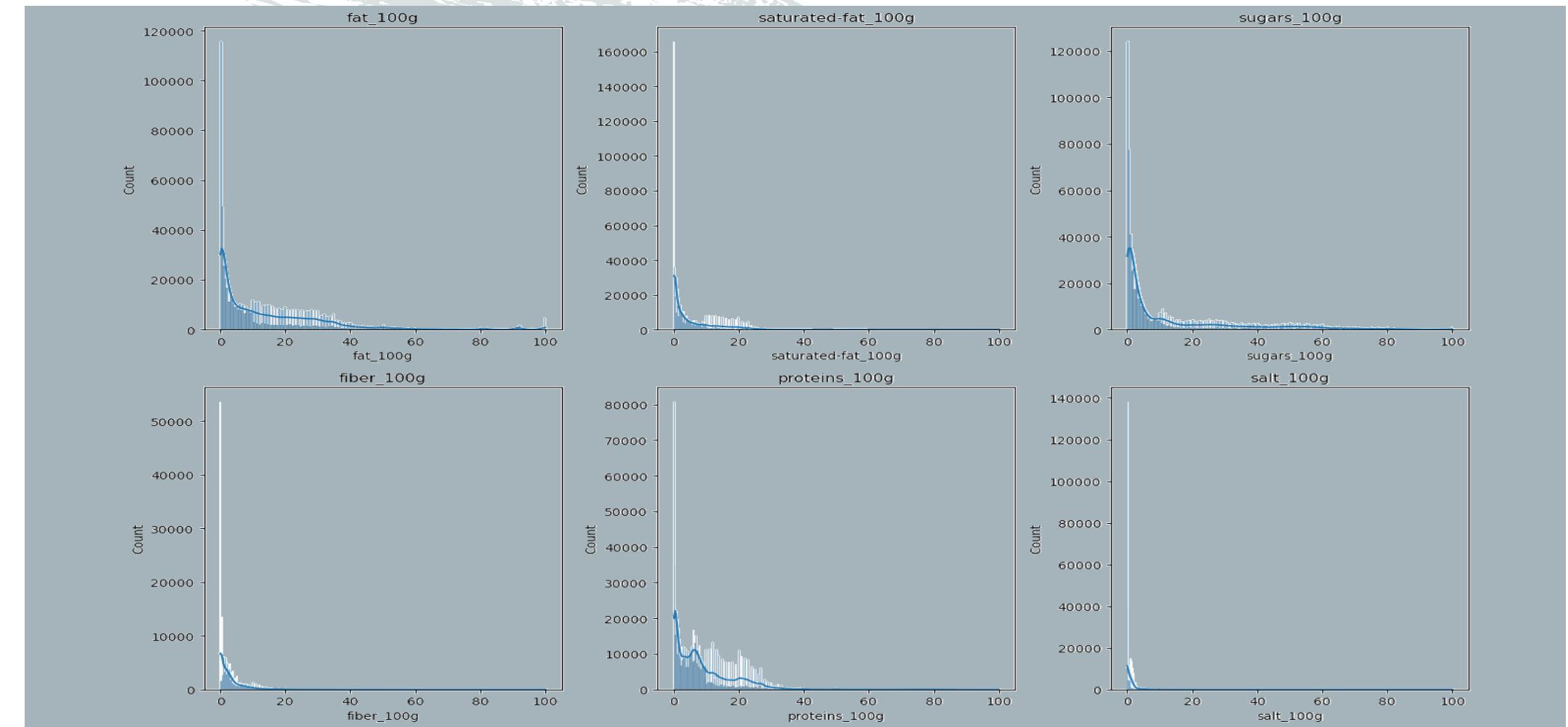
NUAGE DE MOT - FEATURE « CATEGORIES »

Variables Qualitatives

ANALYSE UNIVARIÉE



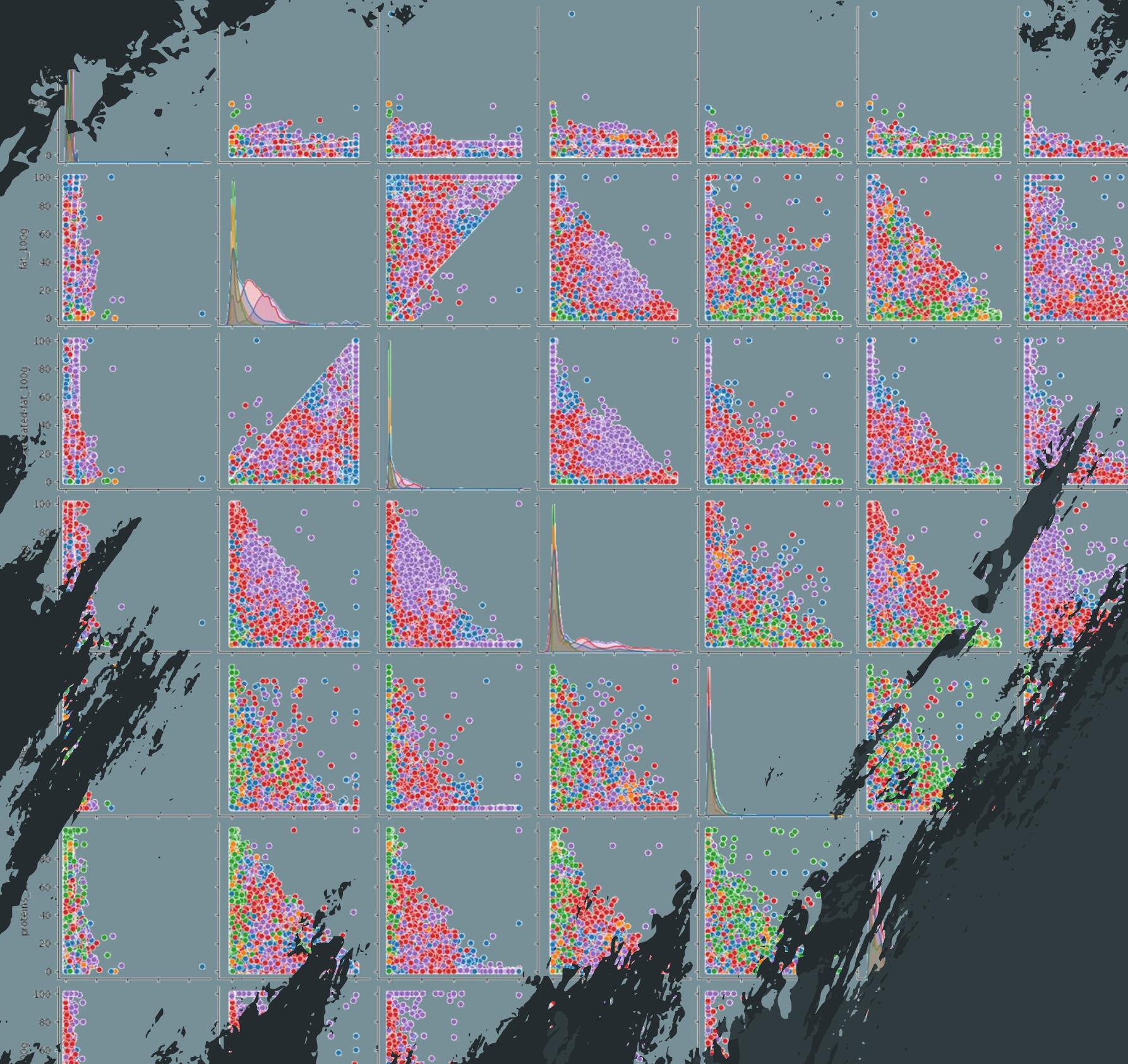
RÉPARTITION DE L'ÉNERGIE (POUR 100G)



RÉPARTITION DES NUTRIMENTS (POUR 100G)

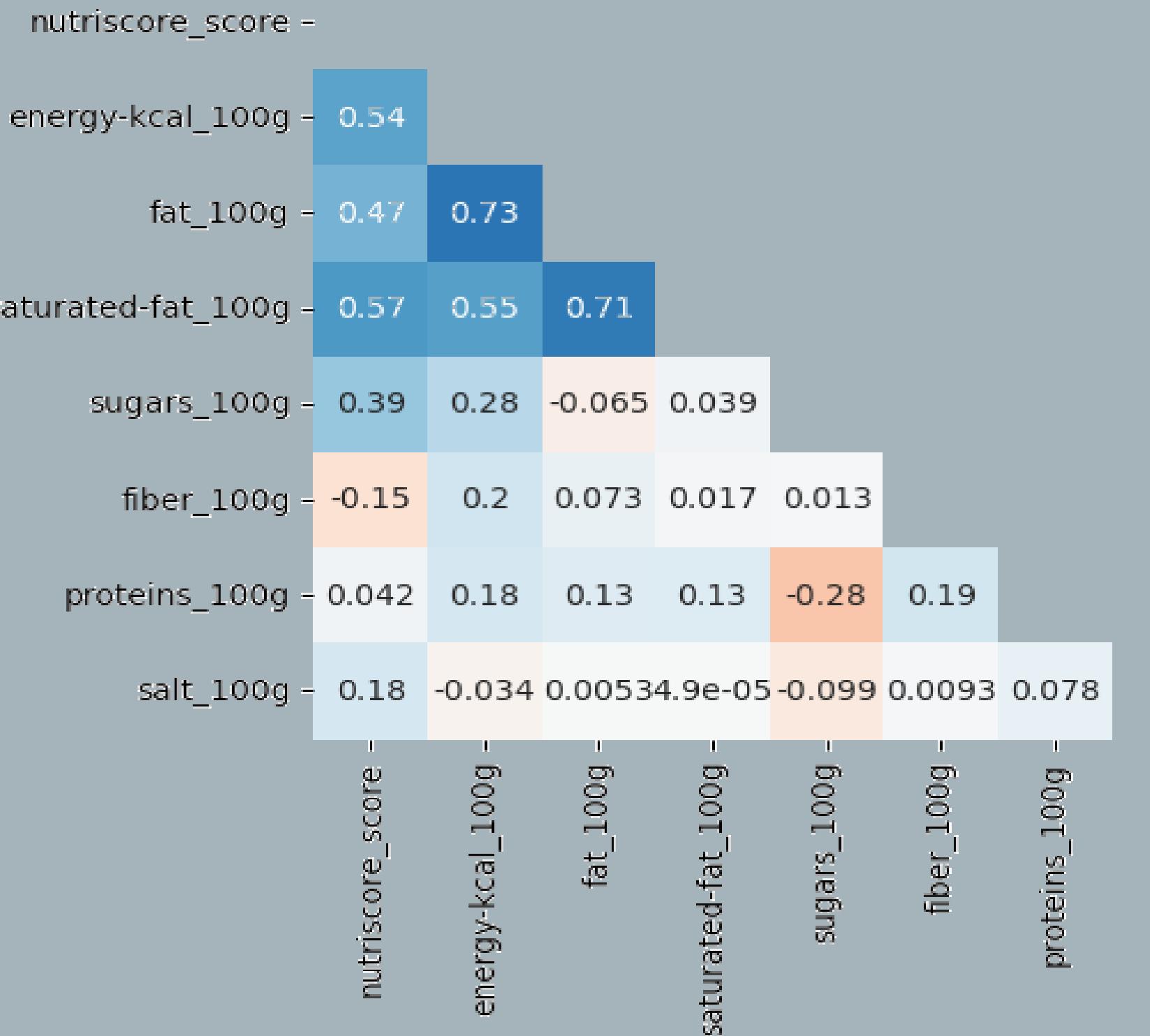
Variables Quantitatives

ANALYSE MULTIVARIÉE

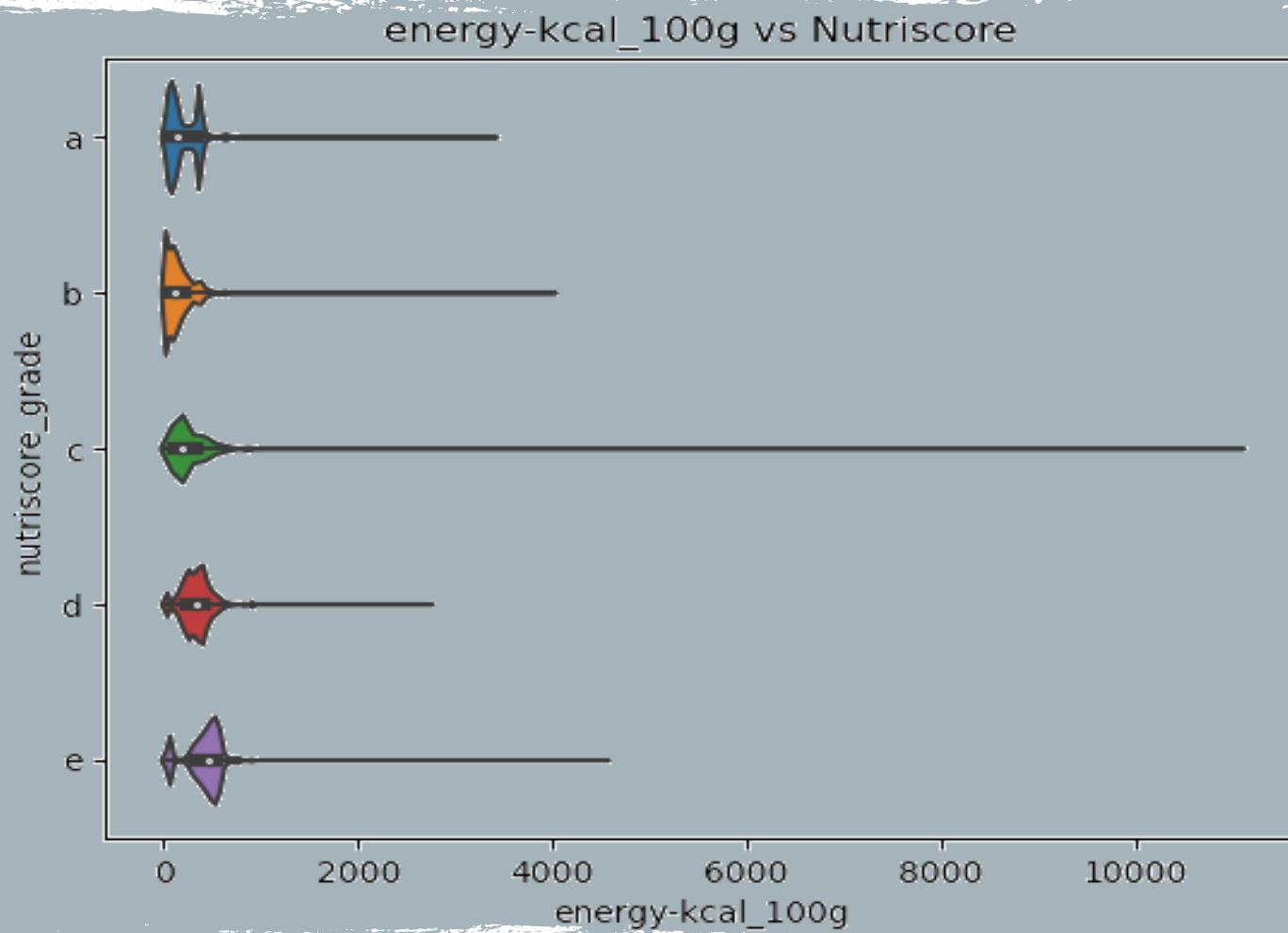


MATRICE DE CORRÉLATION

Matrice de corrélation des variables numériques



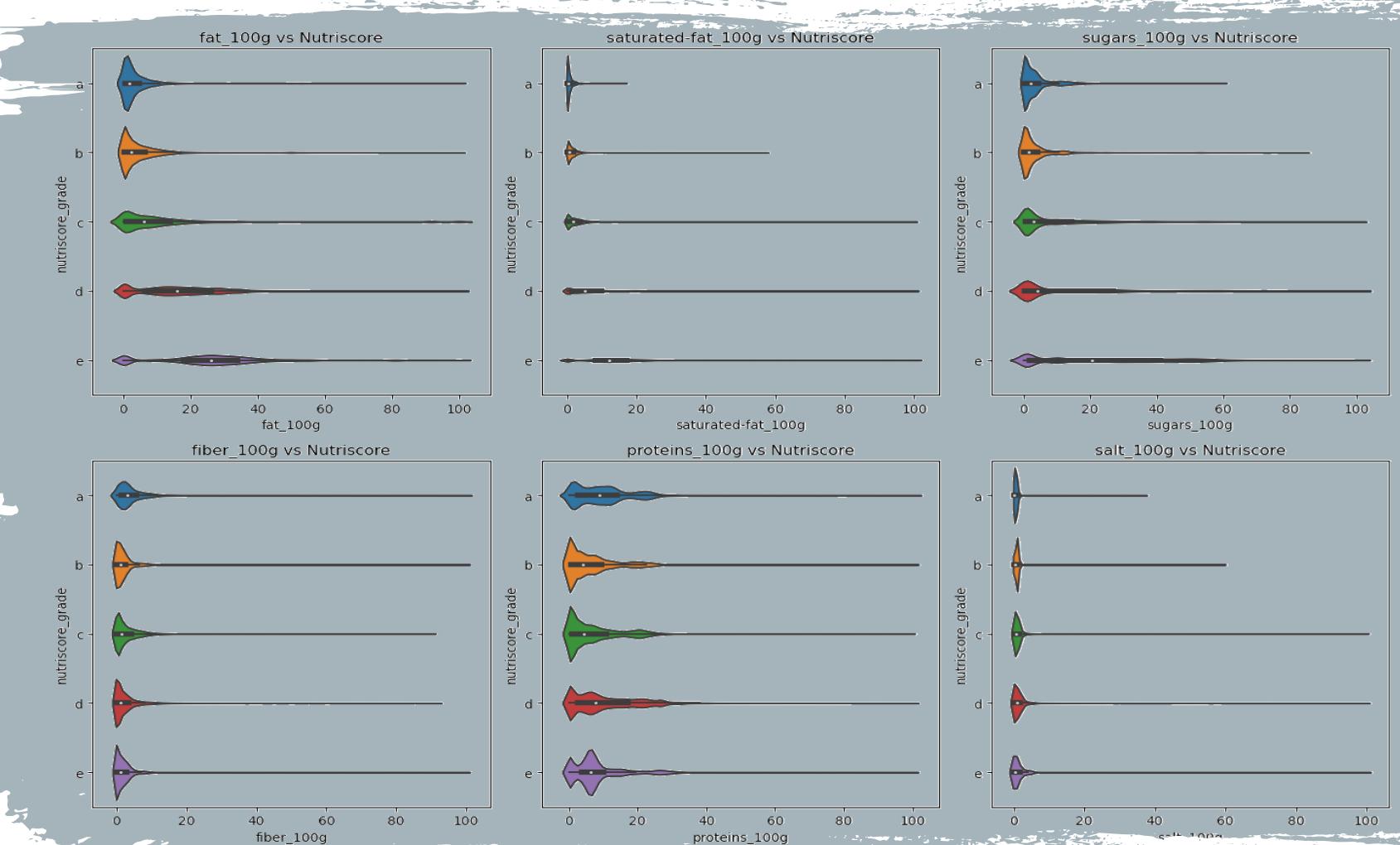
NUTRIENTS VS NUTRISCORE



Rapport de corrélation (η^2) Nutriscore vs :

- Fat : 0.19
- Saturated fat : 0.31
- Sugars : 0.14
- Fiber : 0.03
- Proteins : 0.03
- Salt : 0.02
- Energy : 0.21

ANOVA





CHI - 2

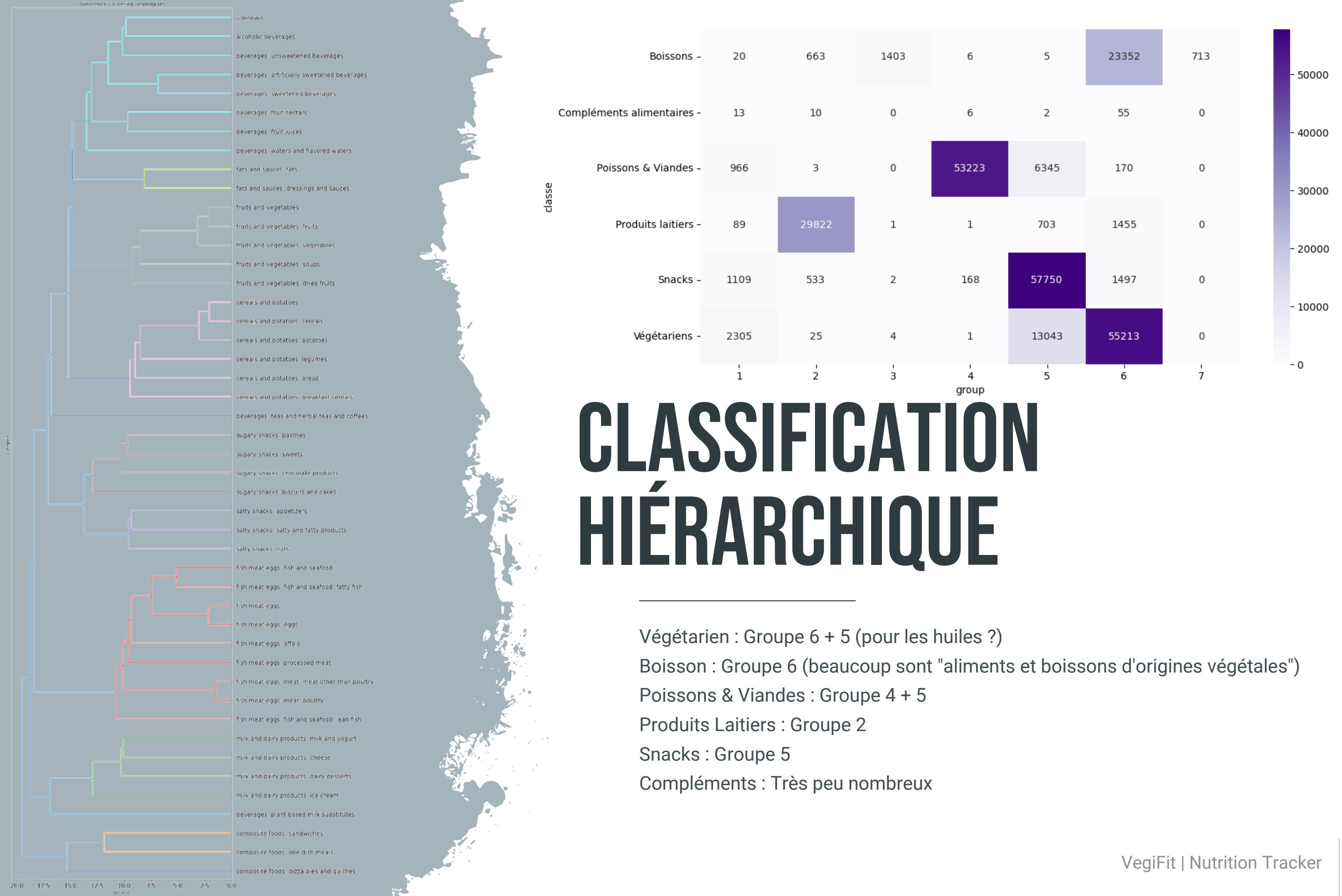
NUTRISCORE VS CLASSE

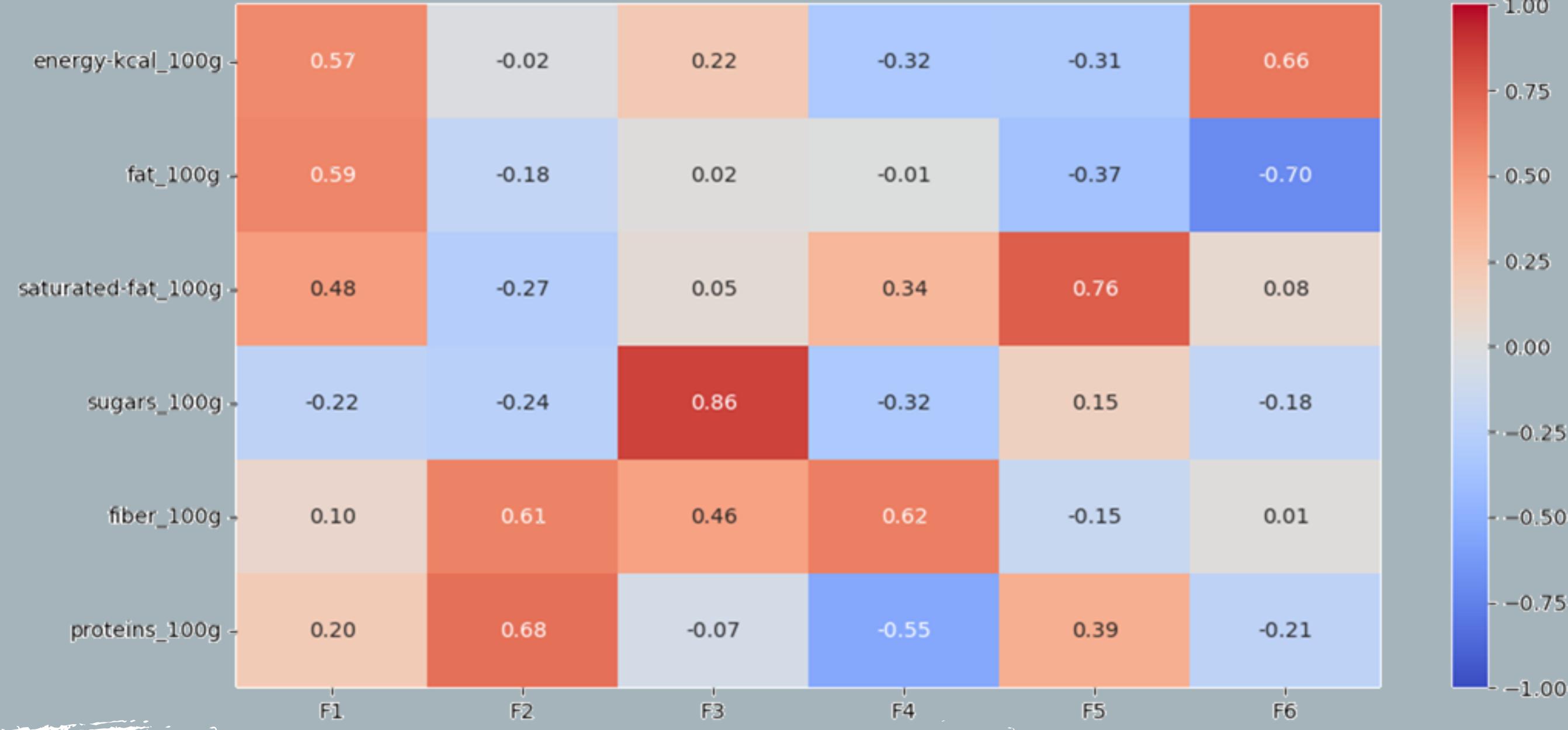
Chi-2: 83004.72766530629

p-value: 0.0

Degrés de liberté: 20

Corrélation polarisée entre Végétarien et nutriscore :
A pour produits alimentaires, E pour huile ?





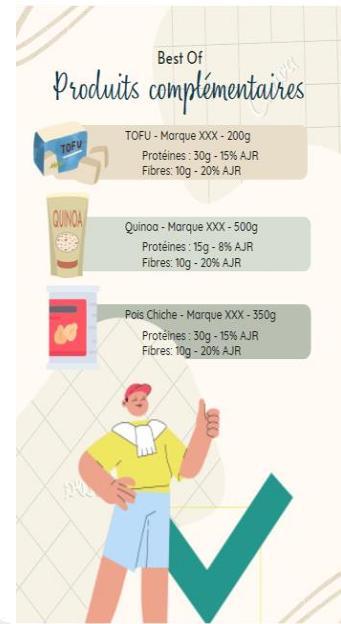
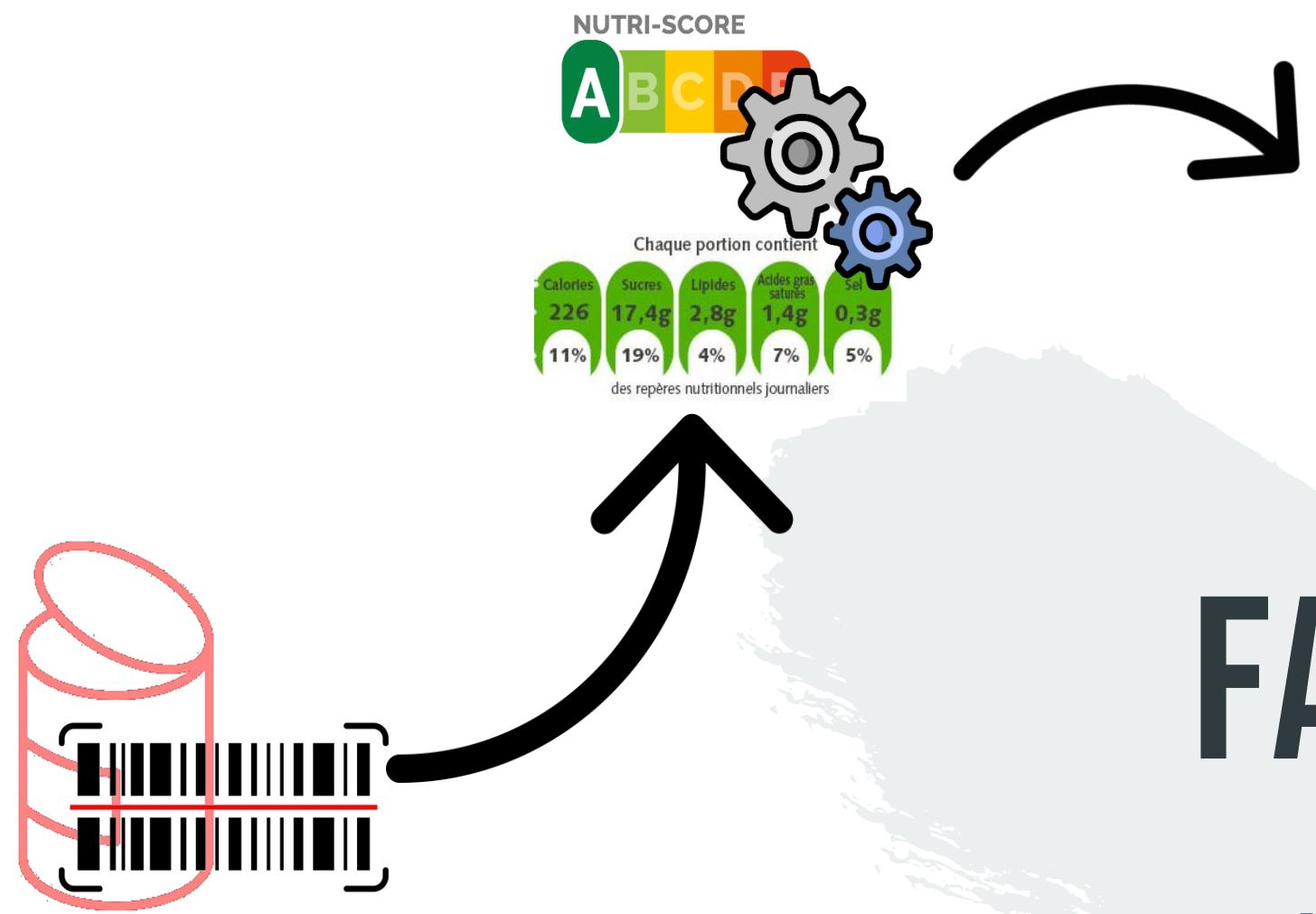
ACP SUR NUTRIMENTS

Pas de simplification possible par composantes principales

(il en faudrait 5 ou 6, pour 6 features pertinents)



SECTION 5
CONCLUSION



FAISABILITÉ

- Scan d'un produit alimentaire
- Recherche de compléments végétariens via « categories » et « food_group_tags » basés sur le **nutriscore** et les AJR en **nutriments principaux**
- Liste de plusieurs produits avec informations nutritionnelles

Nutriments	Limite à ne pas dépasser ou à atteindre
Calories	2000-2500 kcal/jour pour les femmes et 2500-3000 kcal/jour pour les hommes
Graisses saturées	Ne pas dépasser 10 % de l'apport calorique quotidien total
Graisses trans	Aussi bas que possible
Glucides	45 à 65 % de l'apport calorique quotidien total
Sucres ajoutés	Aussi bas que possible
Protéines	0,8 à 1,2 gramme par kilogramme de poids corporel
Sel	Moins de 5 grammes par jour
Fibres	25-30 grammes par jour pour les adultes





LIMITES OPEN FOOD FACTS

- Produits alimentaires uniquement (et non aliments)
- Manque de données sur les micro-nutriments (vitamines et minéraux)

- Nécessité de travailler avec des professionnels de la nutrition pour développer une telle application
 - Accompagnement personnalisé à prévoir pour prévenir les risques de dérives

MERCI DE VOTRE ATTENTION!

DES QUESTIONS?

Octave POUILLOT

Formation Data Scientist – OpenClassRooms – Projet 3