

P3: CONCEVEZ UNE APPLICATION AU SERVICE DE LA SANTE PUBLIQUE

Concevez une application au service de la sante publique



SOMMAIRE

1. Objectifs du projet
2. Pitch de l'idée d'application
3. Fonctionnement
4. Démonstration
5. Nettoyage du jeu de données
6. Traitement des valeurs manquantes
7. Présentation de l'analyse exploratoire
8. Conclusion



CONCEVEZ UNE APPLICATION AU SERVICE DE LA SANTÉ PUBLIQUE

L'agence d'Etat **Santé publique France** a lancé un **appel à projets** pour trouver des idées innovantes d'**applications en lien avec l'alimentation**.

Une **étude de faisabilité** est menée à partir des données disponibles sur **Open Food Facts**.

OPEN FOOD FACTS

Open Food Facts est un **projet collaboratif** dont le but est de constituer une **base de données** libre et ouverte sur les **produits alimentaires** commercialisés dans le monde entier.

Open Food Facts est disponible via un **site web** et plusieurs **applications mobiles**.





MyRepasPlanner

MyRepasPlanner est une application mobile nutritionnelle d'aide à la composition de repas équilibrés.

Renseignez vos caractéristiques morphologiques, précisez vos préférences culinaires et c'est parti !

join now

log in



A partir de notre base de données MyRepasPlanner générera une combinaison de produits adaptés à vos besoins pour chaque repas de la journée



Fonctionnement



Repas :

Pour **chaque repas** une combinaison :

- Entrée
- Plat
- Dessert

Pour **chaque journée** un plan :

- Petit déjeuner
- Déjeuner
- Collation
- Diner

Contraintes :

- **AJR*** comprenant :
 - Apports **caloriques** (2500 kcal/jour/adulte)
 - Apports **nutritionnels** (protéines, lipides, glucides, calcium,...)
- Caractéristiques **morphologiques** de l'utilisateur :
 - Age
 - Sexe
 - Taille/Poids
 - Activité physique
- Allergies
- Préférences culinaires

Démonstration

Je suis

Homme

Femme

Hauteur

180

cm

Poids

80

kg

Âge

40

années

Graisse corporelle

Meugler

Moyen

Haute

Niveau d'activité

Modérément actif

Fixer un objectif de poids ?

Non merci

Ouais faisons-le !

Calculer

Spécifications

Plan de repas du jour

Régénérer

2516 Calories

Petit déjeuner

717 Calories

Wrap aux œufs brouillés et aux saucisses

2 portion

Fromage cottage et compote de pommes

2 portion

Déjeuner

870 Calories

Sandwich à la dinde et à la suisse

2 portion

Dîner

930 Calories

Sauté au boeuf

2 portion

Épinards sautés aux pignons de pin

2 portion

Plan journalier

POURCENTAGE DE CALORIES PROVENANT DE

Protéines

38%

Matières grasses

28%

Glucides

34%

TOTAUX ACTUELS

CIBLES

Modifier

Calories

2489

2500

Crabes

213,3 g

125g - 313g

Gros

80g

56g - 139g

Protéine

239,7 g

125g - 313g

Fibre

40,2 g

25 grammes

Glucides nets

173,1 g

-

Sodium

2829mg

-

Cholestérol

1113mg

-

Est. Coût

8,48 \$

< 10 \$

Nutrition plus détaillée

Nutrition

Votre liste de courses

Les ingrédients principaux

Sauce piquante

Nécessaire : 4 cuillères à soupe

Sauce piquante pure cristal, 6 oz

1,39 \$ (10,23 \$/oz liq)

Poulet en conserve

Nécessaire : 1 tasse

Blanc de poulet en morceaux de q...

4,99 \$ (0,51 \$/once)

Objets similaires

Tomates

Nécessaire : 2

Fresh Brand – Tomates raisins, 10...

2,32 \$ (3,71 \$/lb)

Objets similaires

Panier de courses

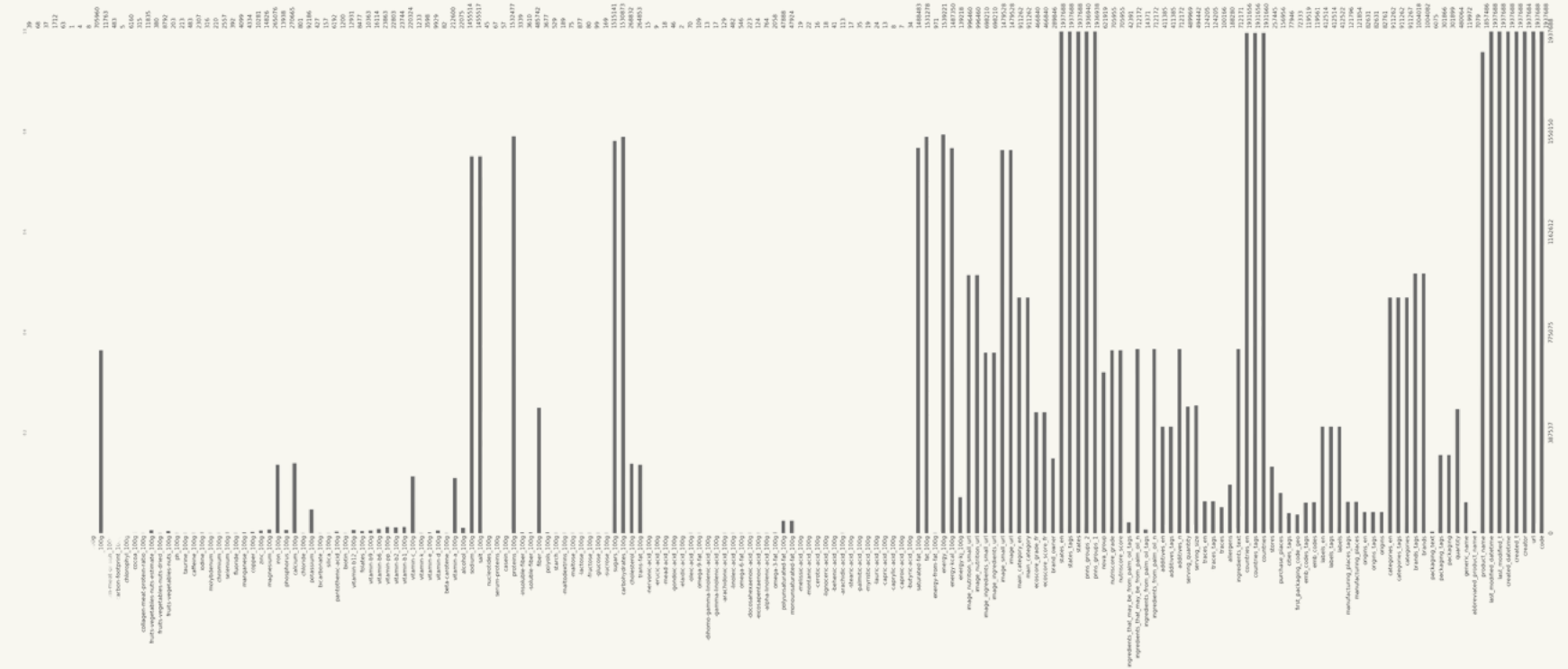
5

Romain Vaillant : Octobre 2021

Présentation du jeu de données

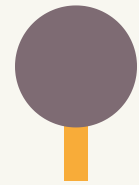
Les champs se regroupent en quatre catégories :

- Des **informations générales** : nom du produit, marque, pays de commercialisation,...
- Des **informations nutritionnelles** : énergie, macronutriments, vitamines,...
- Des **ingrédients** et des **additifs** éventuels
- Un **ensemble de tags** : catégorie du produit, labels qualité, lieu de fabrication,...



- **180** colonnes
- **1 937 688** lignes
- **79%** de valeurs manquantes

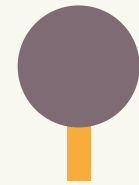
Nettoyage du jeu de données



FONCTIONNEL

Analyse des indicateurs

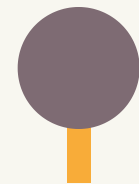
- Sélection
- Filtrage



STATISTIQUE

Outliers

- NaN



OUTLIERS

Visualisation

- Méthode interquartile
- Ajustement des résultats
- Visualisation de contrôle



NAN

Algorithmes de classification et régression

- Imputations par Zéro

Nettoyage fonctionnel



RECHERCHE DES VARIABLES PAR MOT CLE

Après **étude** des variables nutritionnelles **usuelles**

Utilisation d'une **fonction de filtrage**

Analyse et sélection de la variable **la plus appropriée** de la liste

180



21 colonnes

FILTRAGE

Sélection des **produits** vendus en **France**

Sélection des produits disposant des informations :

- **code**
- **product-name**
- **brands**
- **countries**

1 937 688



498 459 lignes

TAUX DE VALEURS MANQUANTES

79



50%

Variables sélectionnées

INDEX

code
product_name
brands
countries
nutriscore_grade

MACRONUTRIMENTS

proteins_100g
carbohydrates_100g'
fat_100g
saturated-fat_100g
sugars_100g
salt_100g

MINEREAUX

calcium_100g
magnesium_100g
potassium_100g
iron_100g

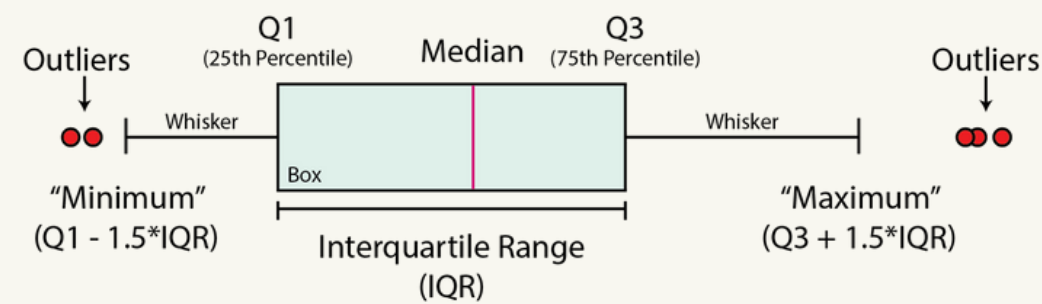
VITAMINES

vitamin-a_100g
vitamin-d_100g
vitamin-b9_100g
vitamin-c_100g

AUTRES

energy-kcal_100g
fiber_100g

Nettoyage statistique



0

OUTLIERS

- **Méthode interquartile** ($IQR = Q3 - Q1$)
- Ajustement aux **valeurs cohérentes** pour les variables exprimées pour **100g** [0;100]

VALEURS MANQUANTES

Algorithmes supervisés :

- **KNN**
- **RandomForest Classifier**

VALEURS MANQUANTES

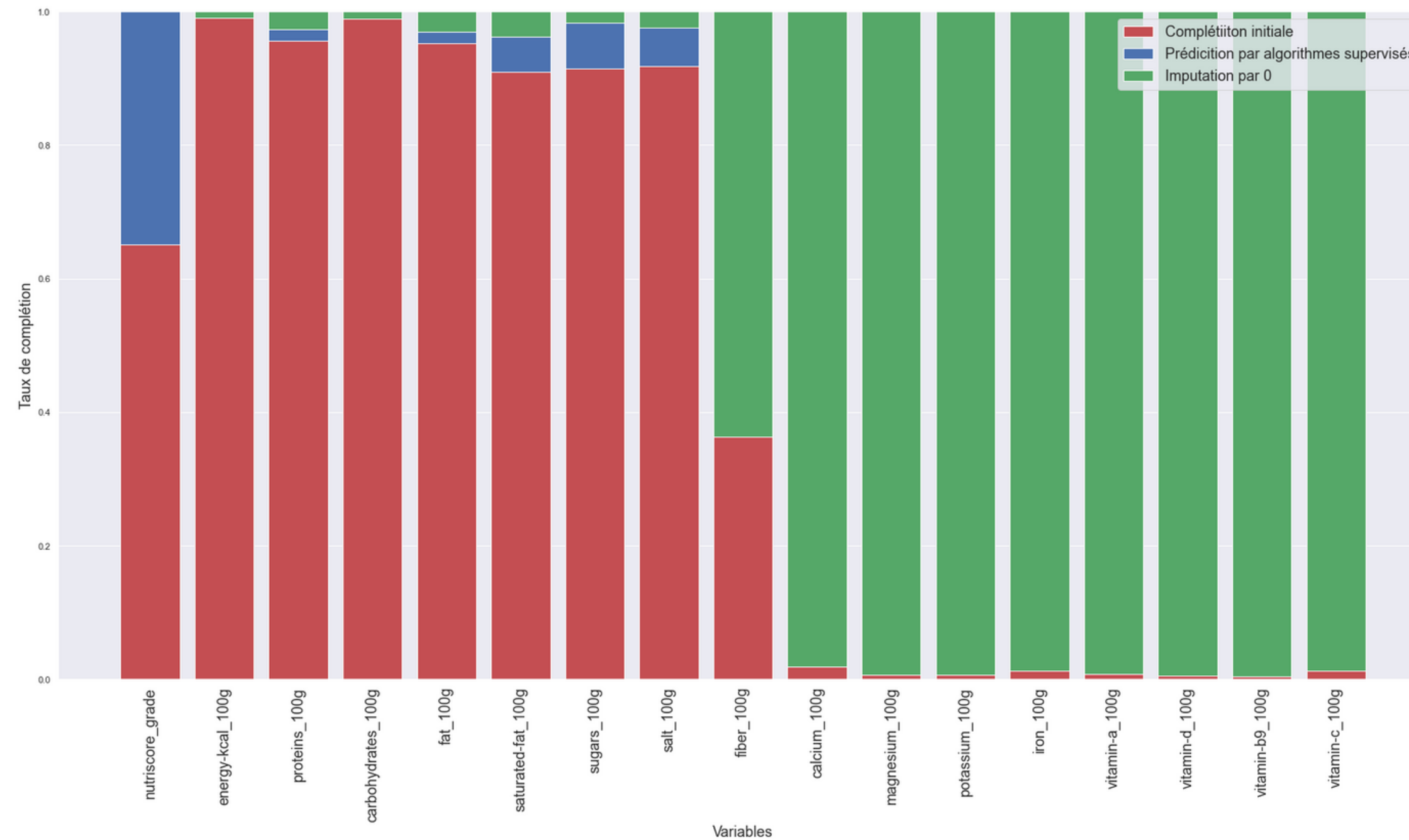
Imputations conventionnelles :

- **Imputation par 0** :

Stratégies d'imputation

	TYPE	KNEIGHBORSREGRESSOR	RANDOMFORESTCLASSIFIER	IMPUTATION PAR 0
energy-kcal_100g	Macronutriment	✓		✓
proteins_100g	Macronutriment	✓		✓
carbohydrates_100g	Macronutriment	✓		✓
fat_100g	Macronutriment	✓		✓
saturated-fat_100g	Macronutriment	✓		✓
sugars_100g	Macronutriment	✓		✓
salt_100g	Macronutriment			✓
fiber_100g	Autre			✓
calcium_100g	Minéraux			✓
magnesium_100g	Minéraux			✓
potassium_100g	Minéraux			✓
iron_100g	Minéraux			✓
vitamin-a_100g	Vitamines			✓
vitamin-d_100g	Vitamines			✓
vitamin-b9_100g	Vitamines			✓
vitamin-c_100g	Vitamines			✓
nutriscore_grade	Autre		✓	✓

Jeu de données final

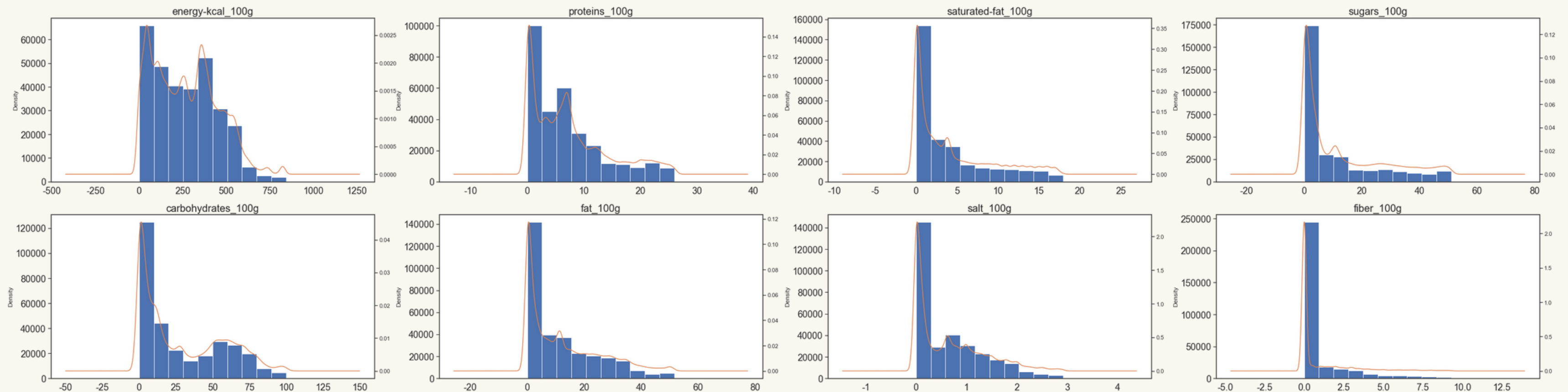


311 603 produits
17 variables
100%
(49%)

Analyse univariée

DISTRIBUTION DES VARIABLES SELECTIONNEES

- Des distributions qui **ne suivent pas** une loi **gaussienne**
- Une large majorité de **valeurs faibles**



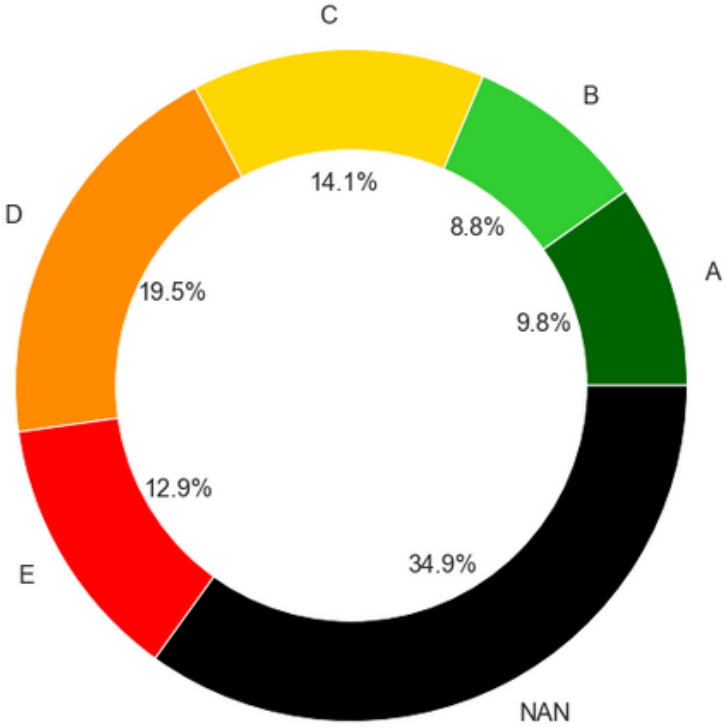
Analyse univariée

RÉPARTITION DE LA PART DU NUTRISCORE PAR CATÉGORIE DES PRODUITS VENDUS EN FRANCE

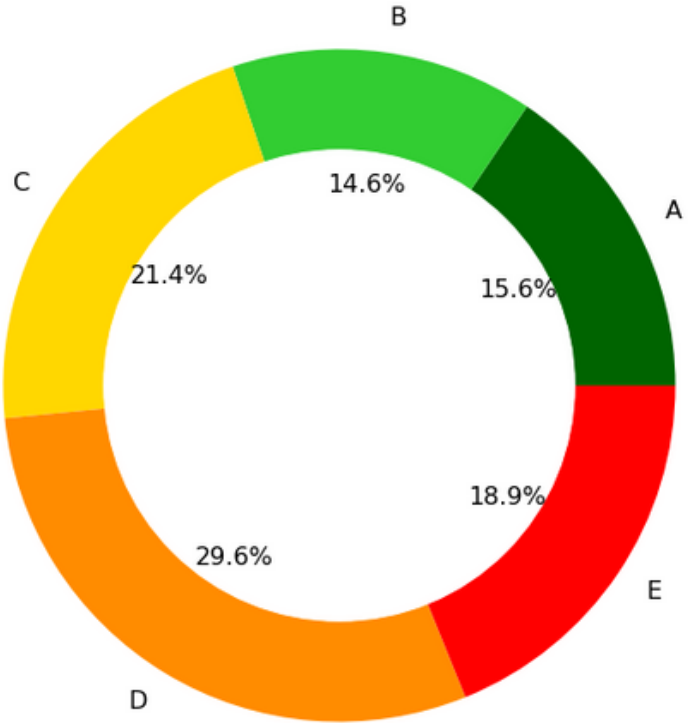
- De nombreuses valeurs manquantes à priori
- Des produits de **qualité** qui sont **minoritaires**



Avant prédiction :

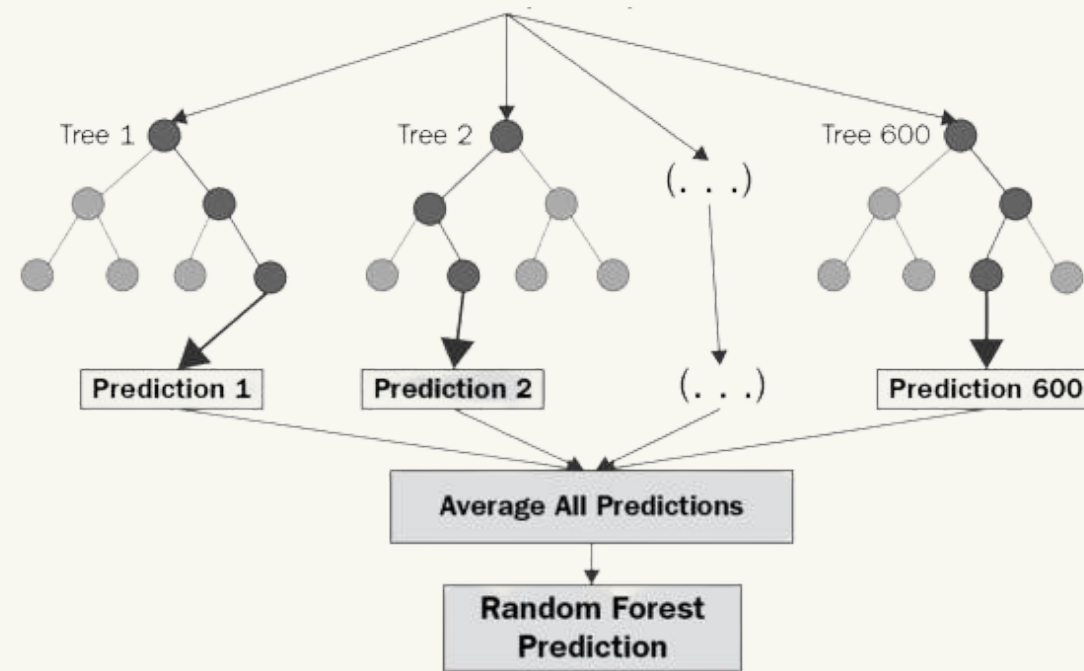


Après prédiction :



Nutriscore : Prédiction des valeurs manquantes

RANDOM FOREST CLASSIFIER



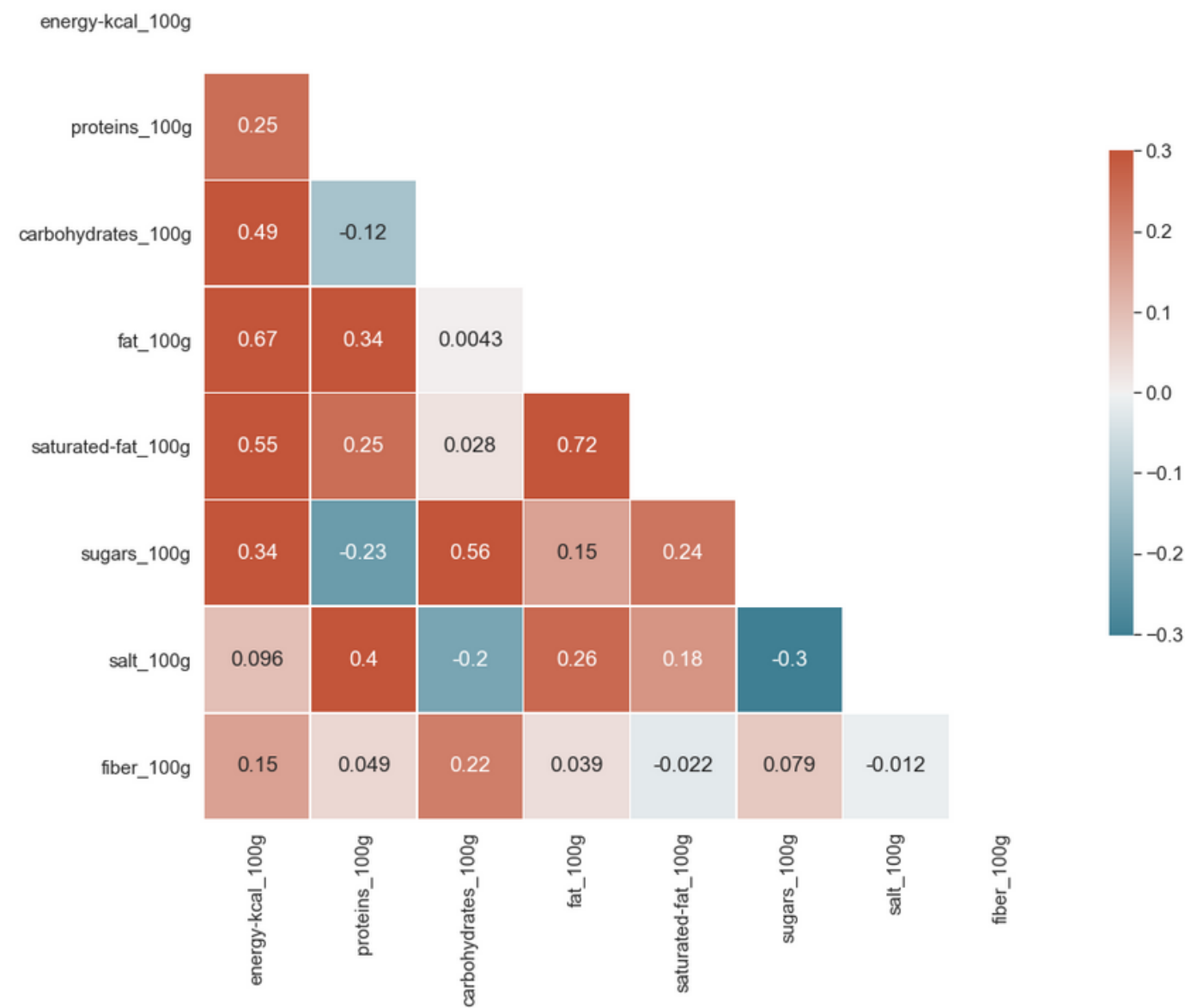
- Un **algorithme** de classification avec une précision **satisfaisante**
- **81%** en moyenne

MATRICE DE CONFUSION

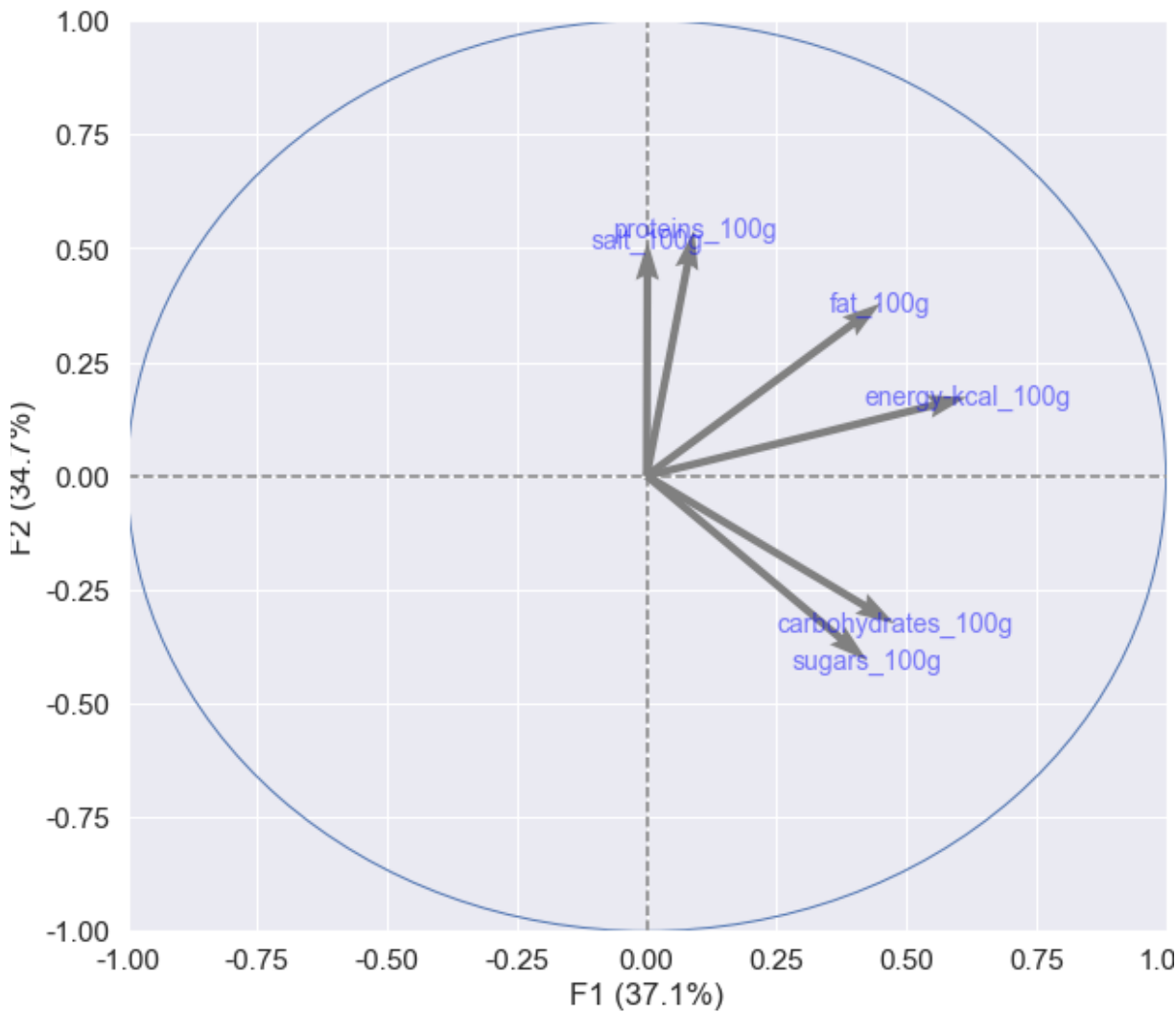
True label\	a	b	c	d	e
	0.83	0.1	0.048	0.0074	0.011
	0.11	0.73	0.13	0.024	0.011
	0.03	0.055	0.76	0.13	0.026
	0.012	0.011	0.078	0.85	0.044
	0.011	0.012	0.035	0.089	0.85
	a	b	c	d	e

Analyse bivariée

MATRICE DES CORRELATIONS



CERCLES DES CORRELATIONS

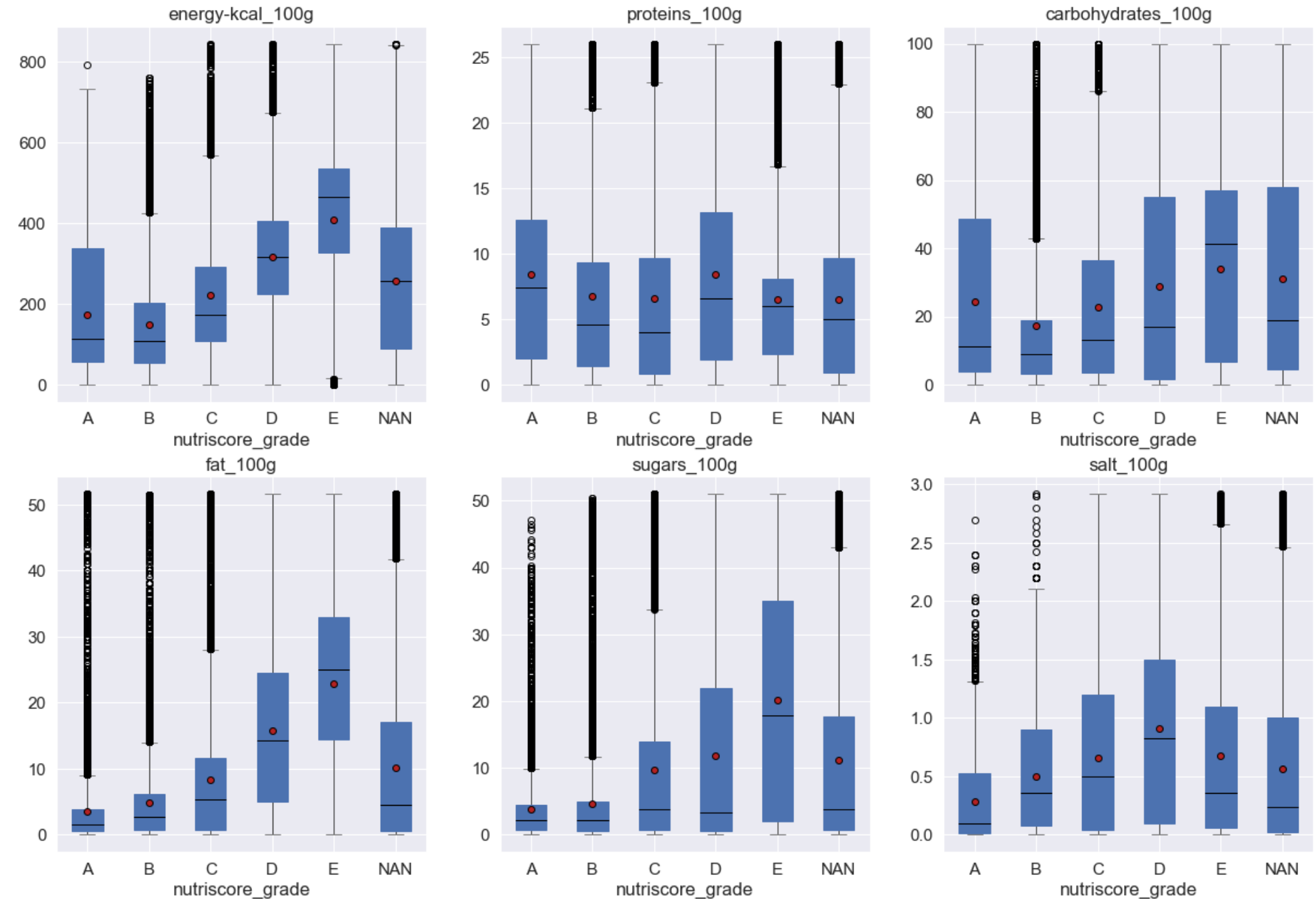


- Des **coefficients** de corrélation qui ont une **décroissance rapide**
- Un **choix** pertinent de **variables**

Analyse bivariée

VARIANCE DES VARIABLES POUR CHAQUE CATÉGORIE DE NUTRISCORE

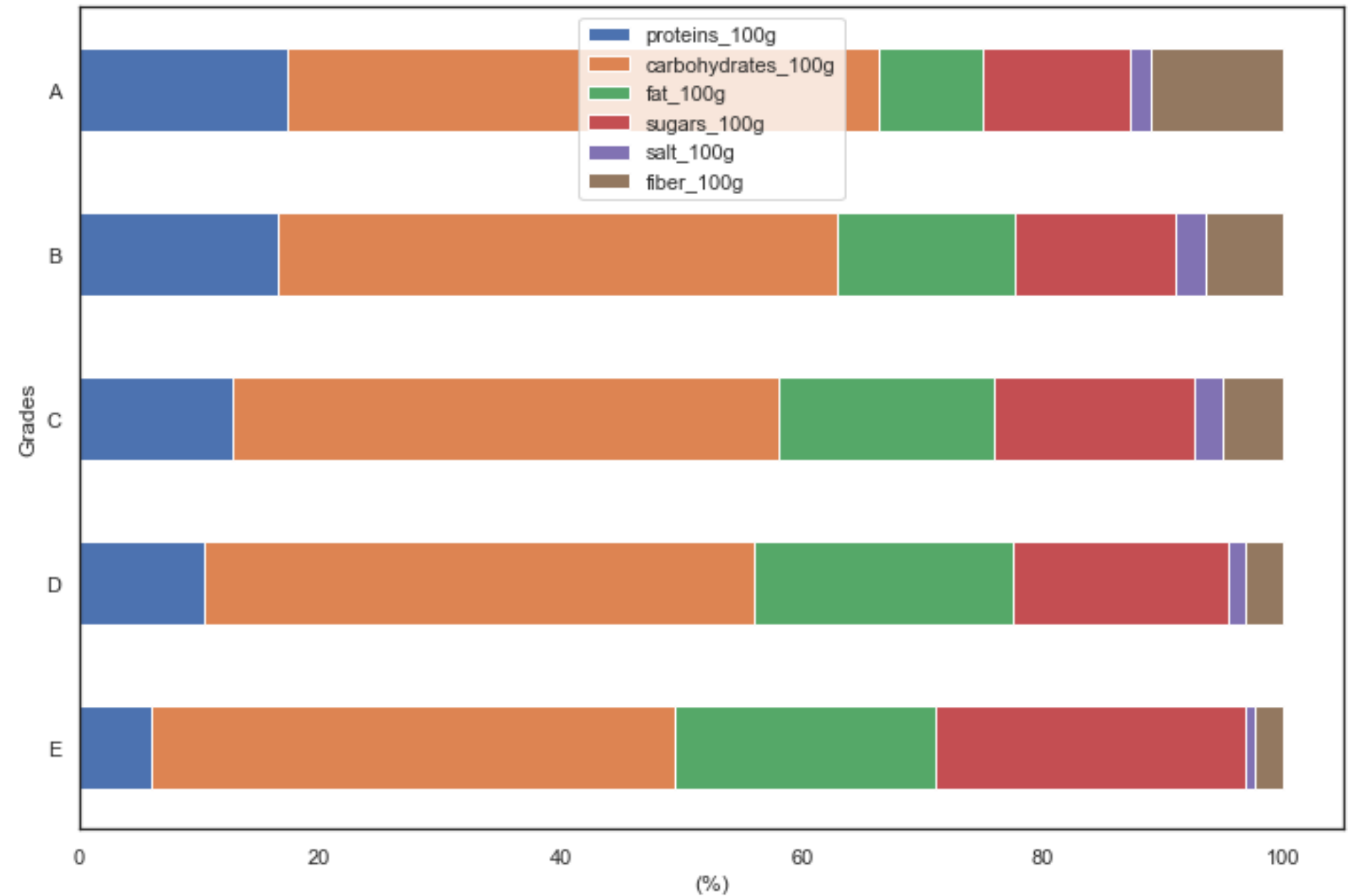
- Les produits à **nutriscore manquant** apparaissent **semblables** aux produits à nutriscore **C,D et E**
- Confirmé par une analyse ANOVA (test de Tuckey)
- La variable **protéines** semble **indépendante** de la catégorie du **nutriscore**
- La **graisse** et le **sucre** impactent immédiatement la **qualité du produit**



Analyse multivariée

PART DES NUTRIMENTS PAR CATEGORIE DE NUTRISCORE

- Des **nutriscores élevés** qui sont essentiellement composés de **protéines** et de **sucres lents**
- Des **nutriscores bas** qui sont essentiellement composés de **sucres** et de **graisse**



Conclusion

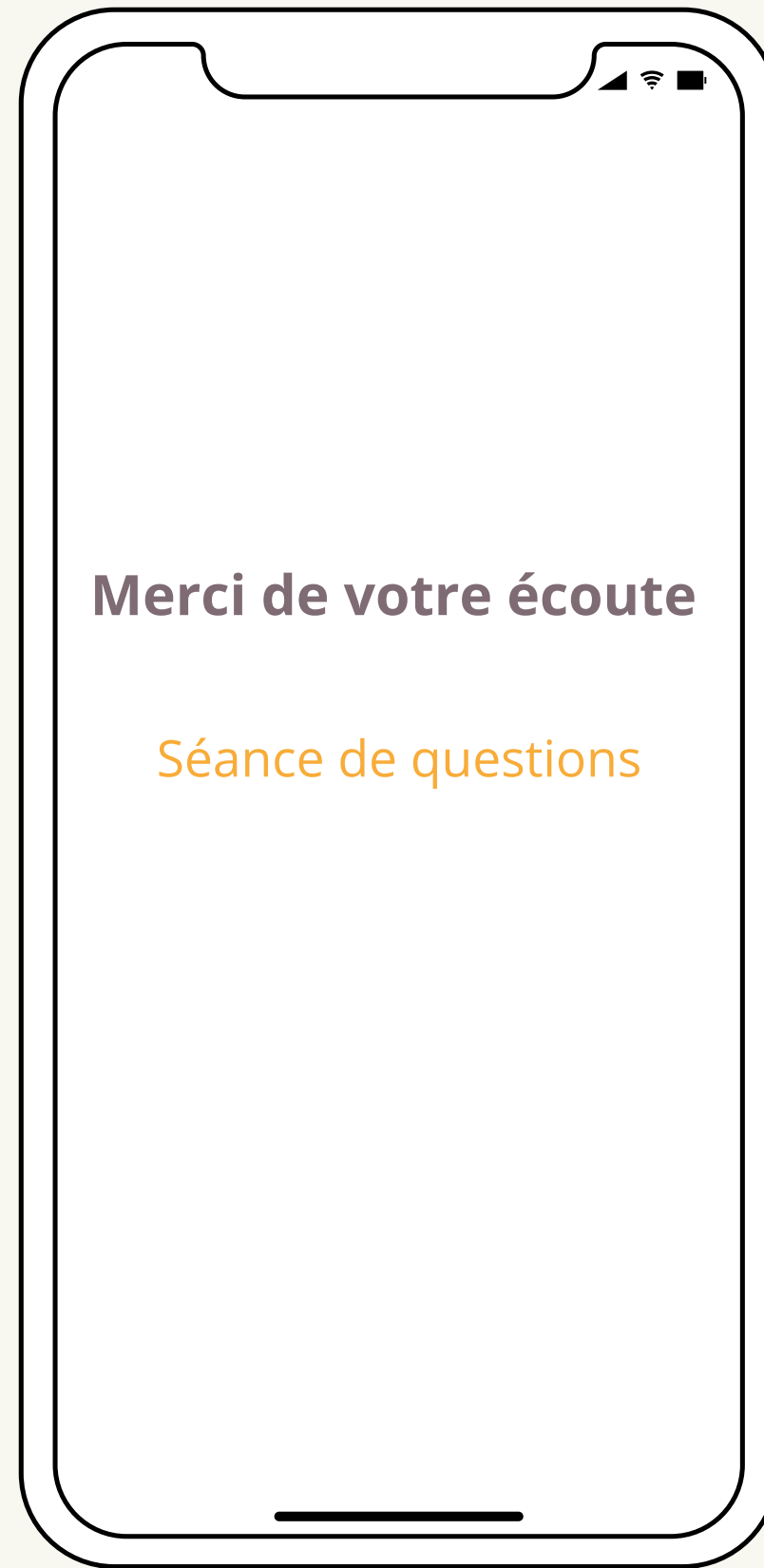
La base de données OpenFoodFacts est **satisfaisante** sur plusieurs points :

- Quantité de produits
- Information générales
- Macronutriments
- Nutriscore : **pierre angulaire**

Socle solide de départ pour **initier le développement** de l'application

Cependant les produits de **qualité** sont **minoritaires**

Cependant **peu de données** renseignées concernant les **minéraux** et les **vitamines** qui sont des **données marketing importantes**



Benchmarking

[HTTPS://WWW.EATTHISMUCH.COM/](https://www.eatthismuch.com/)

(EN)
Site américain

[HTTPS://WWW.STRONGRFASTR.COM/](https://www.strongrfastr.com/)

(EN)
Site américain

[HTTPS://WWW.COOKANDBE.COM/](https://www.cookandbe.com/)

(FR)
Site français

[HTTPS://WWW.LA-FABRIQUE-A-MENUS.FR/](https://www.la-fabrique-a-menus.fr/)

(FR)
Site français

Sponsorisé par :
Ministère de la Santé,
Santé publique France,
MANGER BOUGER (Programme National Nutrition Santé)