

# Concevez une application au service de la santé publique

...

3 mars 2023  
Yoann Poupart

# Food-Tracker

## Story:

- ➡ Suivre ses habitudes alimentaires
- ➡ Mesurer la qualité des apports

## Potential features:

- ➡ Ajout d'aliment
- ➡ Ajout de recette
- ➡ Planification des repas
- ➡ Liste de course

## Features:

- ➡ Suivi des AJR (statistiques sur la durée)
- ➡ Recherche & recommandation d'aliments



# Éléments clés des données

## Composition:

- ➡ Indication précise sur les habitudes alimentaires
- ➡ Suivi détaillé des AJR

## Nom et catégorie:

- ➡ Recherche d'aliments
- ➡ Recommandation et regroupement d'aliments

## Énergie:

- ➡ Résumé synthétique d'un aliment
- ➡ Suivi globale des AJR

## Mesure de qualité:

- ➡ Qualité des aliments (nutri-score|nutri-grade)
- ➡ Qualité de la production (eco-score)



# Démarche expérimentale

## Nettoyage

- Sélection des caractéristiques et individus
- Filtrage des valeurs aberrantes

- Remplissage des valeurs omises
- Prédiction des valeurs manquantes

## Exploration

- Description des caractéristiques
- Analyse des corrélations

- Tests de dépendance | d'indépendance
- Analyse en composantes principales

# Nettoyage des données

# Caractéristiques

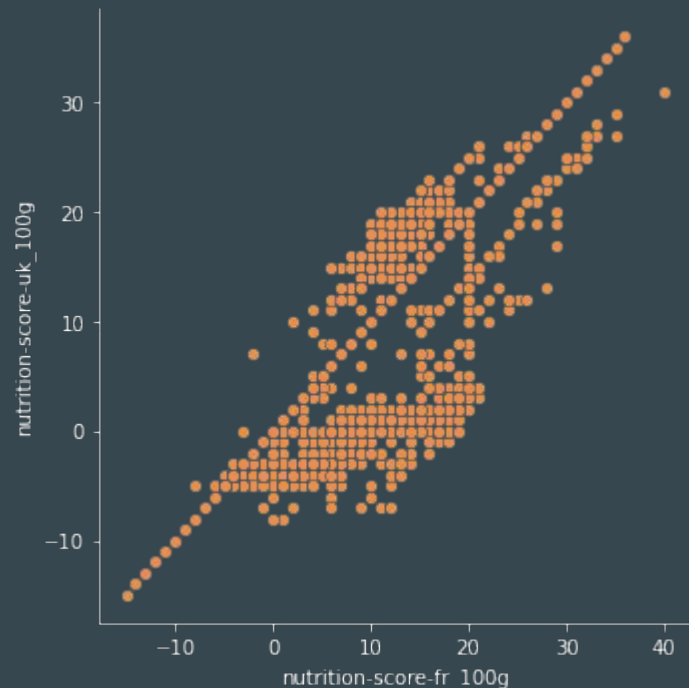
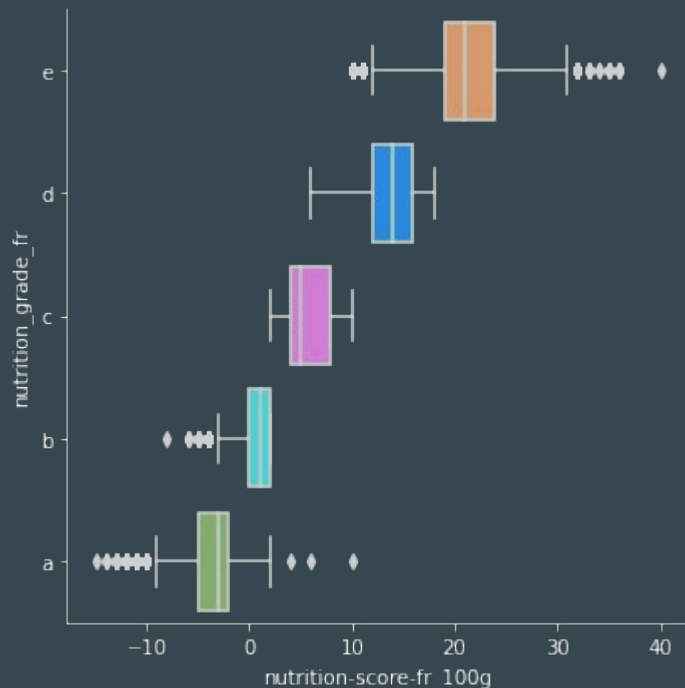
## Sélection

- Composition nutritionnelle
  - ➔ Suivi détaillé des AJR
- Mesures globale (énergie, nutri-grade, nutri-score)
  - ➔ Suivi de la qualité et des AJR
- Nom et différentes catégories
  - ➔ Regroupement d'aliments

## Filtrage

- Caractéristiques relativement remplies (seuil à 1%)
  - ➔ Réduire la sparsité (Perte de variables intéressantes eco-score)
- Fusion de caractéristiques
  - ➔ Remplir des valeurs manquantes

# Corrélations des caractéristiques



# Corrélations des caractéristiques

## Interprétation:

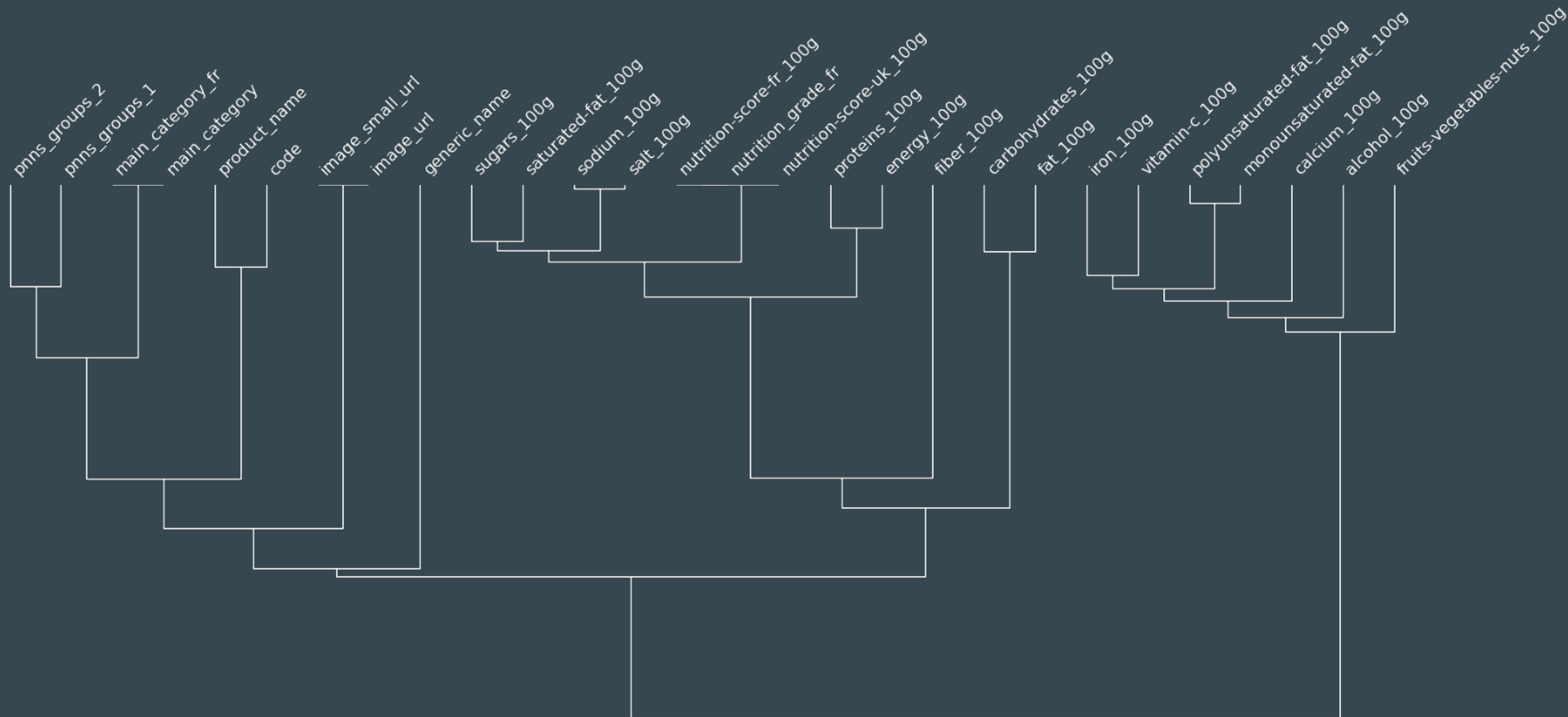
- ➡ Caractéristiques corrélées mais différentes
- ➡ Donnent des informations légèrement différentes

## Procédure adoptée:

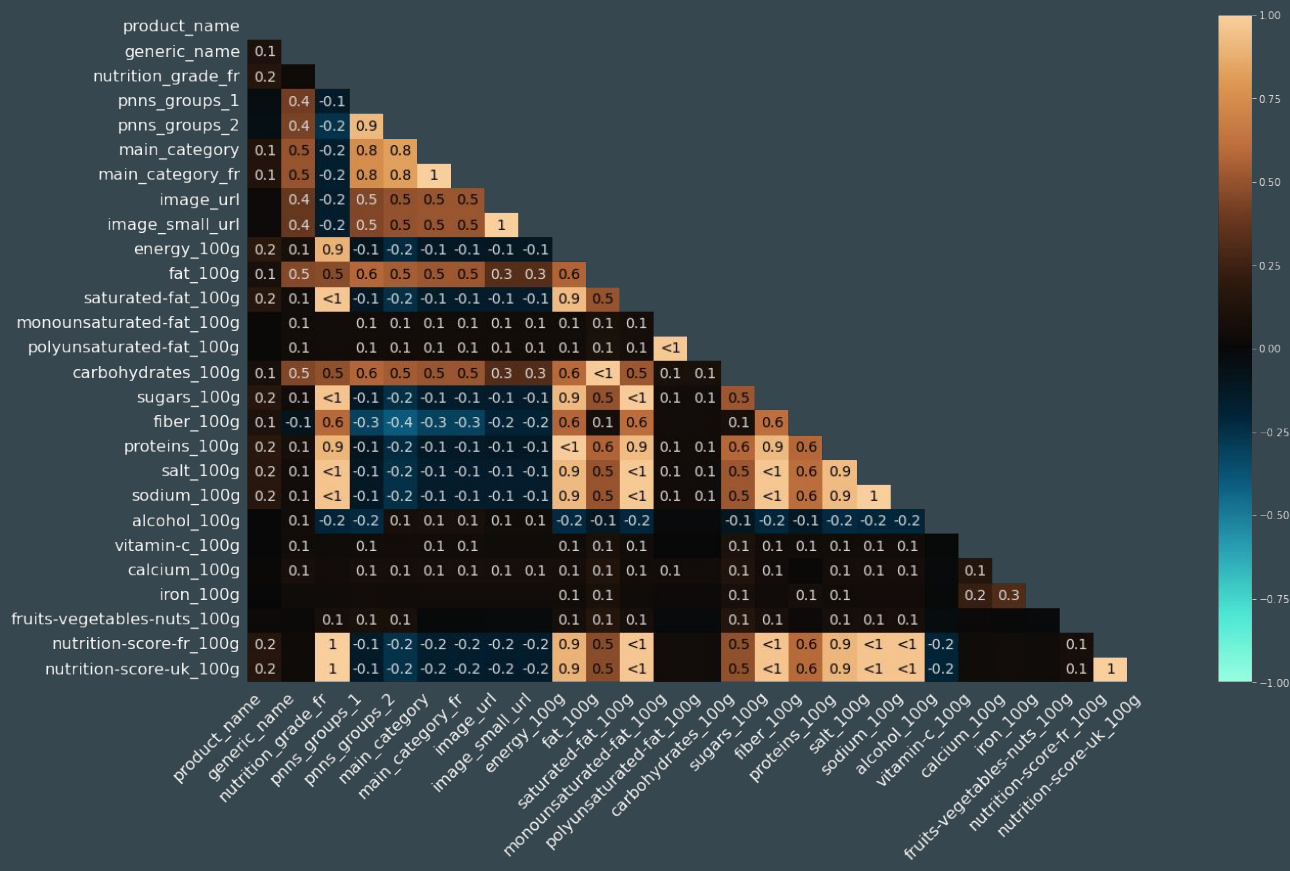
- ➡ Garder les 3 scores
- ➡ Créer un nutri-score moyen



# Corrélations des valeurs manquantes



# Corrélations des valeurs manquantes



# Corrélations des valeurs manquantes

## Interprétation:

- ➡ Valeurs manquantes corrélées par nutriments similaires
- ➡ Relève de l'omission plus que de l'erreur

## Procédure adoptée:

- ➡ Nutriments omis remplacés par 0

# Individus

## Sélection

- Aliments vendus en France
  - ➔ Application française
- Aliments nommés
  - ➔ Affichage pour l'application
- Aliments uniques

## Filtrage

- Valeurs aberrantes
  - ➔ Intervall des caractéristiques (positif,  $\text{energie} < 3765.6$ ,  $\text{nutriment} < 100$ )
- Cohérence des caractéristiques
  - ➔ Lien caractéristiques & sous-caractéristiques
    - ➔ Total inférieur à 100

# Outliers

## Interprétation:

- ➡ Nutriments: distributions à queues lourdes
- ➡ Certains aliments sont des outliers par nature (sel, huile, ...)

## Procédure adoptée:

- ➡ Pas de filtrage des outliers

# Prédiction des valeurs manquantes

## Variable catégorielle (nutri-grade)

- Classifieur *randomForest*
  - ➔ Robuste aux outliers
  - ➔ Interprétable
- Encodage numérique
  - ➔ Par indigage
  - ➔ Encodage one-hot

## Variables numériques

- Imputeur KNN
  - ➔ Prédictions simultanées
  - ✗ Lent et sensible aux outliers
- Régresseur *randomForest*
  - ➔ Robuste aux outliers
  - ➔ Interprétable

# Prédiction des valeurs manquantes

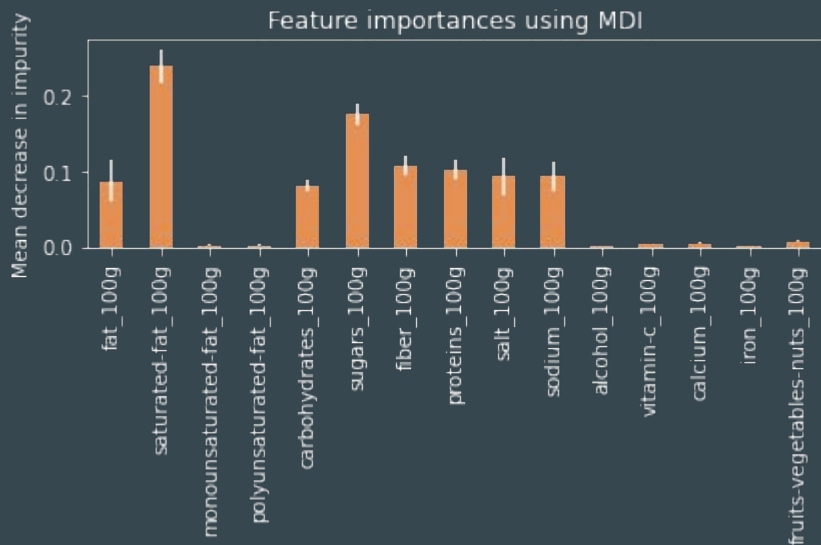
## Résultats:

- ➡ RandomForest (RF) meilleure que KNN
- ➡ Overfitting (avec 20 arbres et profondeur max à 20)

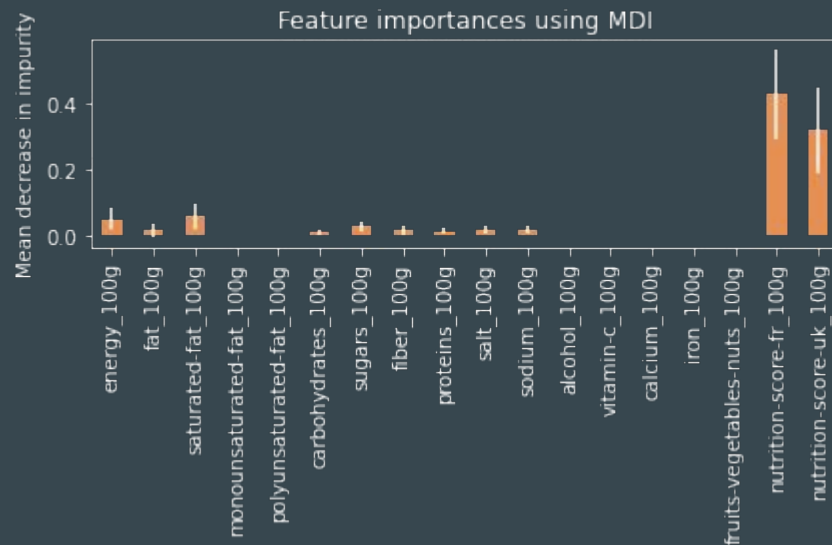
	Train	Test
KNN	-	$R^2$ : 0.919
RF	$R^2$ : 0.975	$R^2$ : 0.921
RF (nut)	A: 0.980 - F1: 0.978	A: 0.877 - F1: 0.873
RF	A: 1.000 - F1: 1.000	A: 0.999 - F1: 0.999

# Prédiction des valeurs manquantes

Classifieur (RF) à partir des nutriments



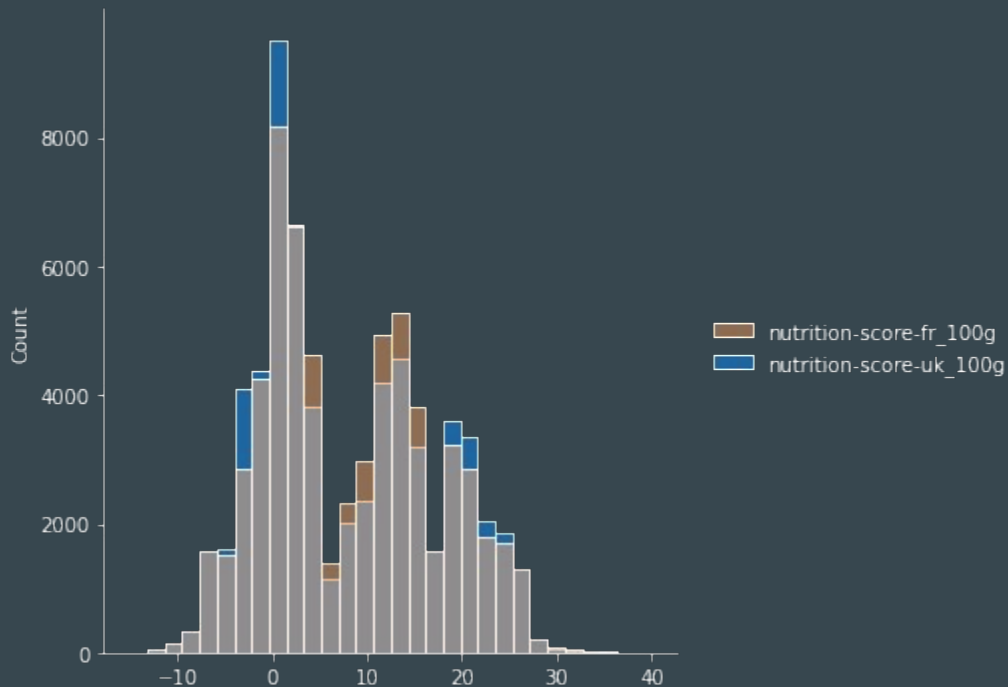
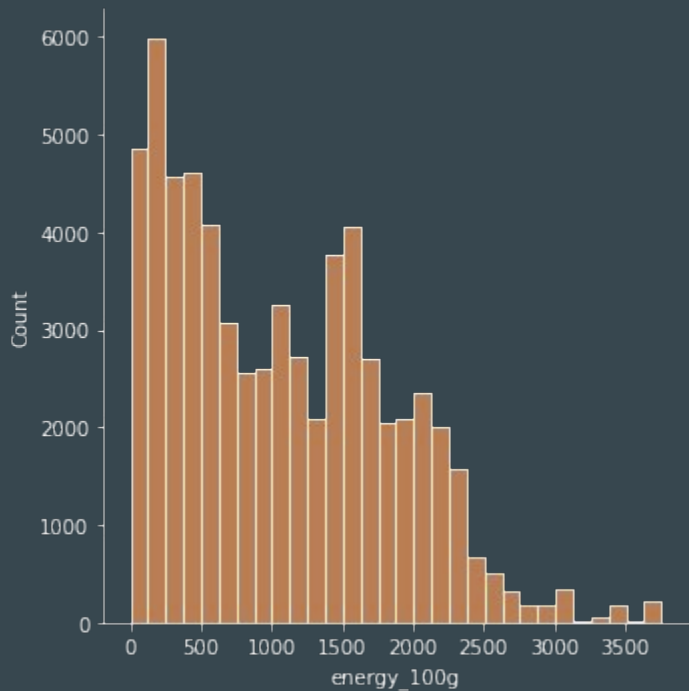
Classifieur (RF) à partir des variables numériques



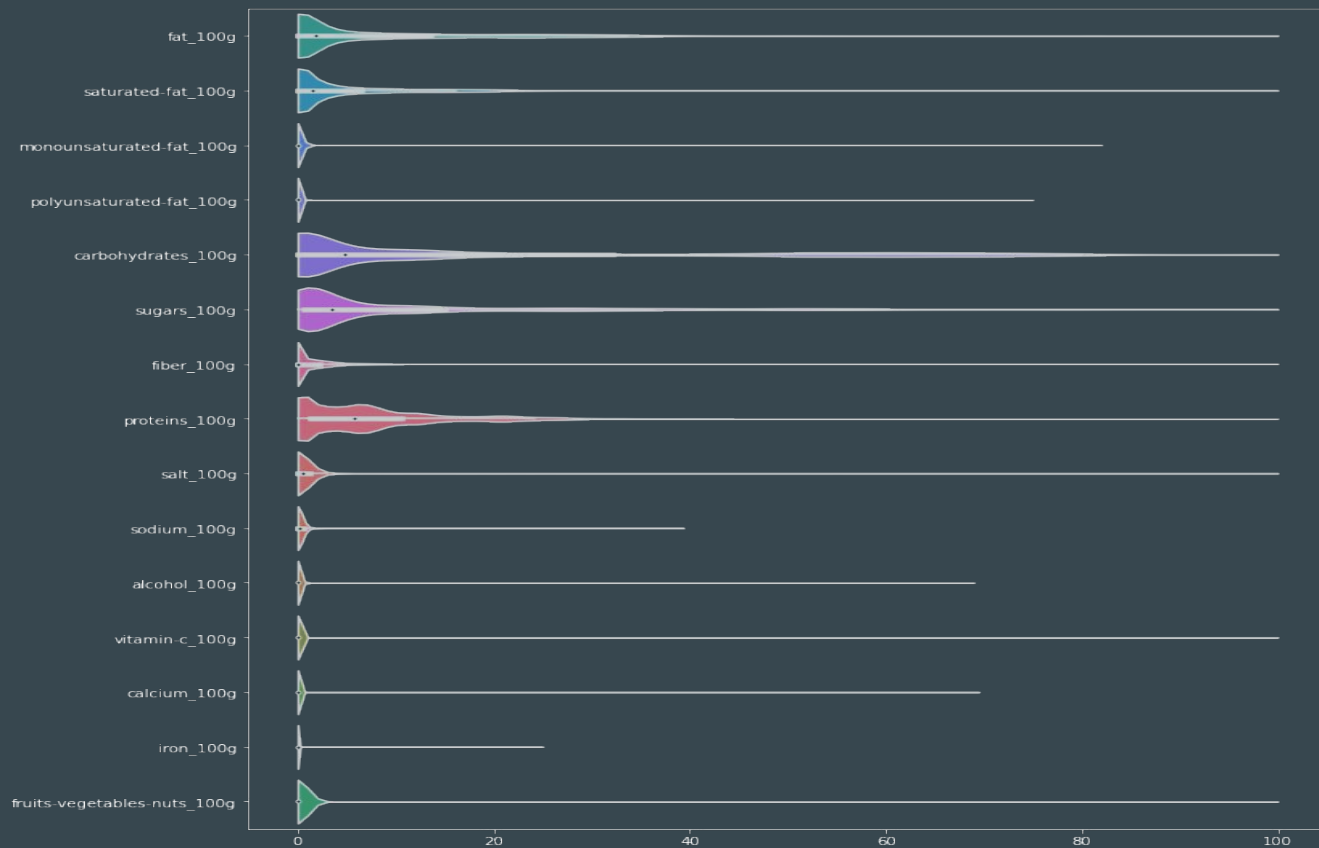


# Exploration des données

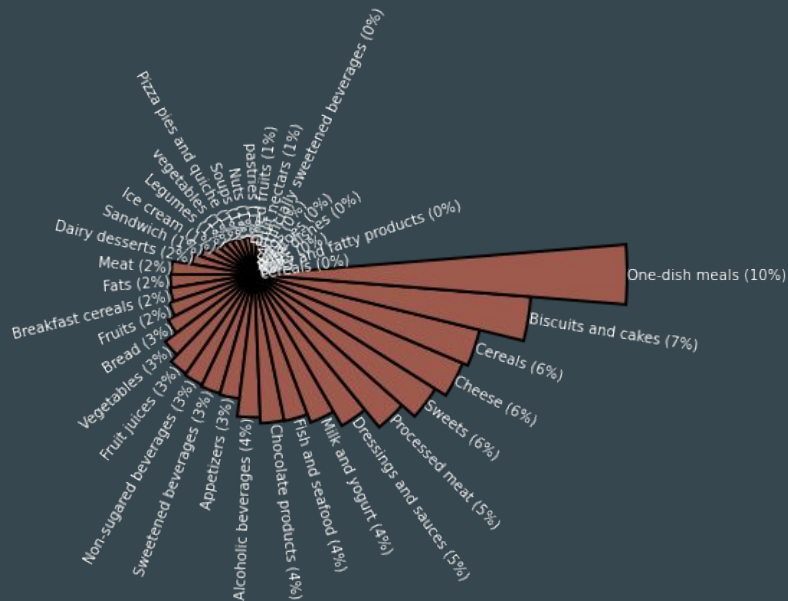
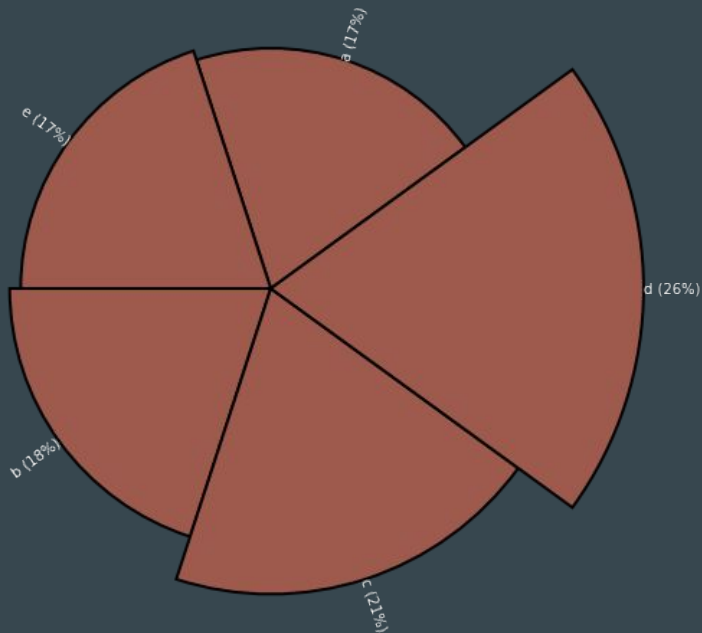
# Numerical features



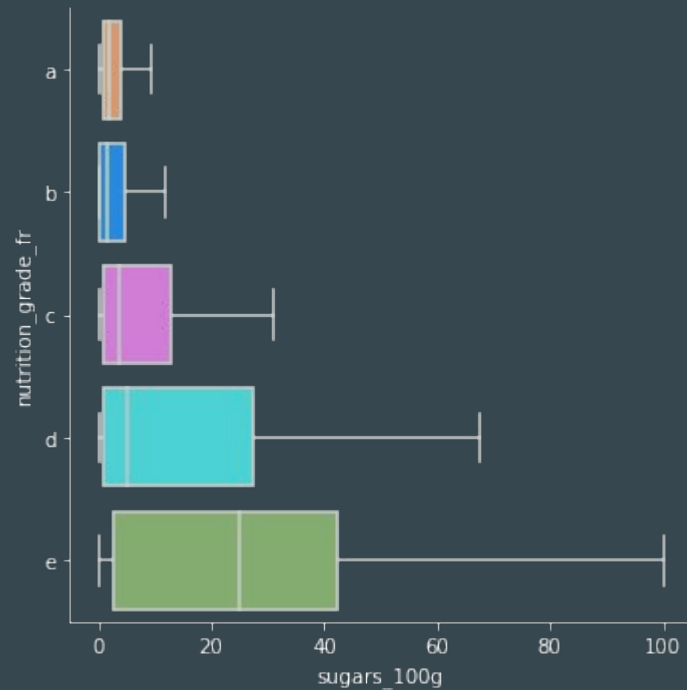
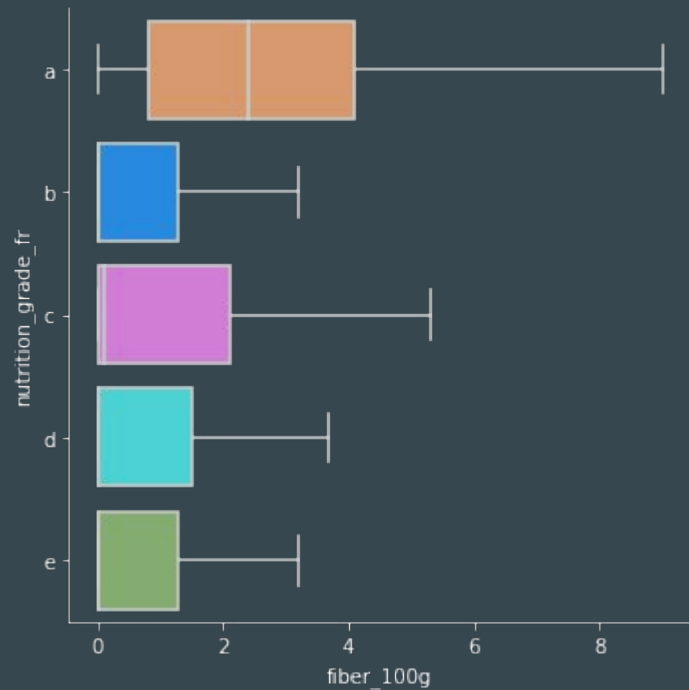
# Numerical features



# Categorical features



# Corrélations qualitatives



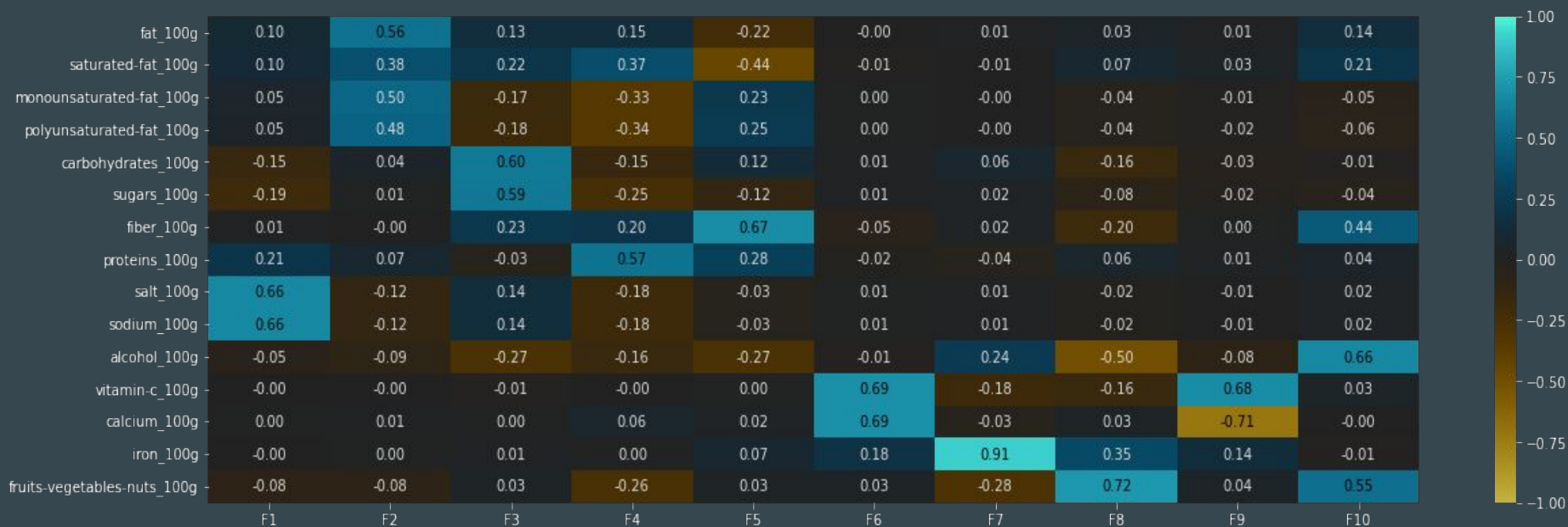
# Corrélations quantitatives

# Corrélations quantitatives

LR	Énergie - Nutri-score (moyen)	$R^2$ : 0.445	p-value: 0.000
	Nutri-score (fr) - Nutri-score (uk)	$R^2$ : 0.928	p-value: 0.000
$\chi^2$	Nutri-grade - Catégorie	-	p-value: 0.000
ANOVA	Nutri-grade - Nutri-score (fr)	$\eta^2$ : 0.914	p-value: 0.000
	Nutri-grade - Énergie	$\eta^2$ : 0.338	p-value: 0.000
	Catégorie - Nutri-score (fr)	$\eta^2$ : 0.391	p-value: 0.000
	Catégorie - Énergie	$\eta^2$ : 0.496	p-value: 0.000

Test de normalité: Énergie (p-value: 0.000) - Nutri-score (p-value: 0.000)

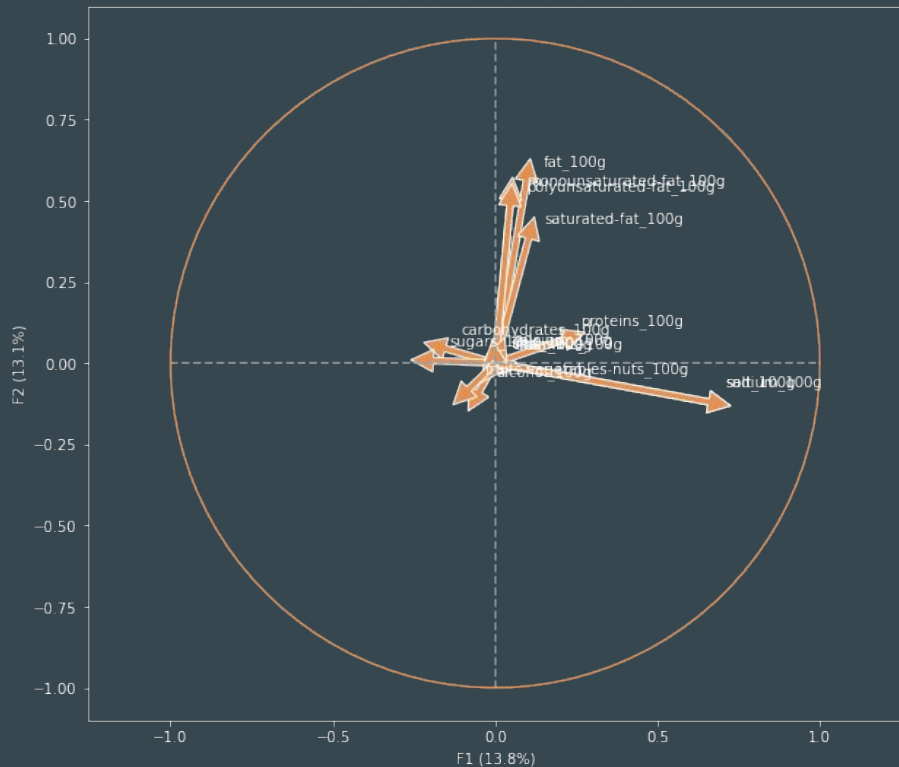
# ACP partielle



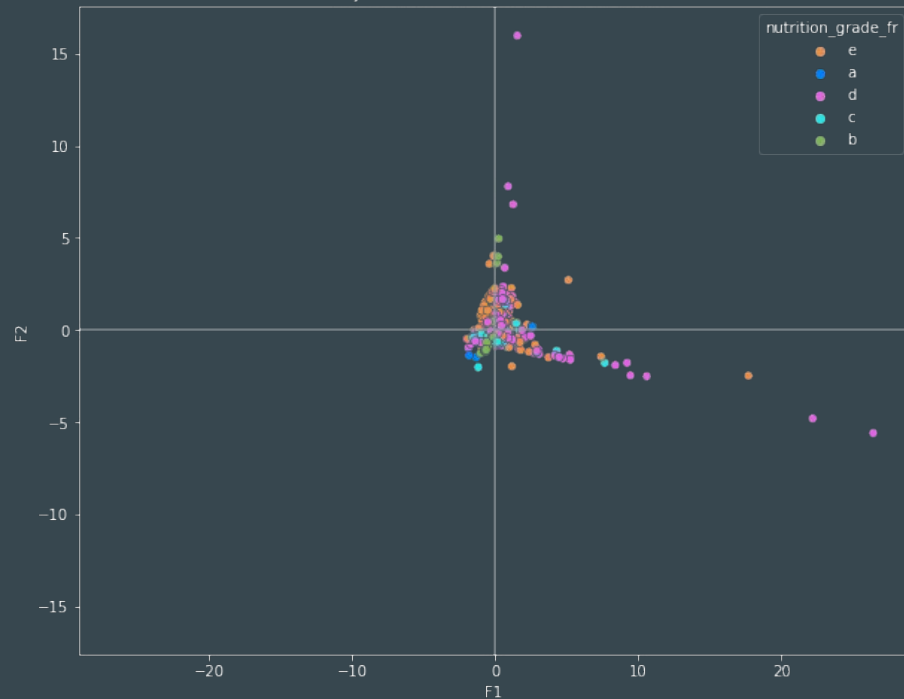


# ACP partielle

Cercle des corrélations (F1 et F2)

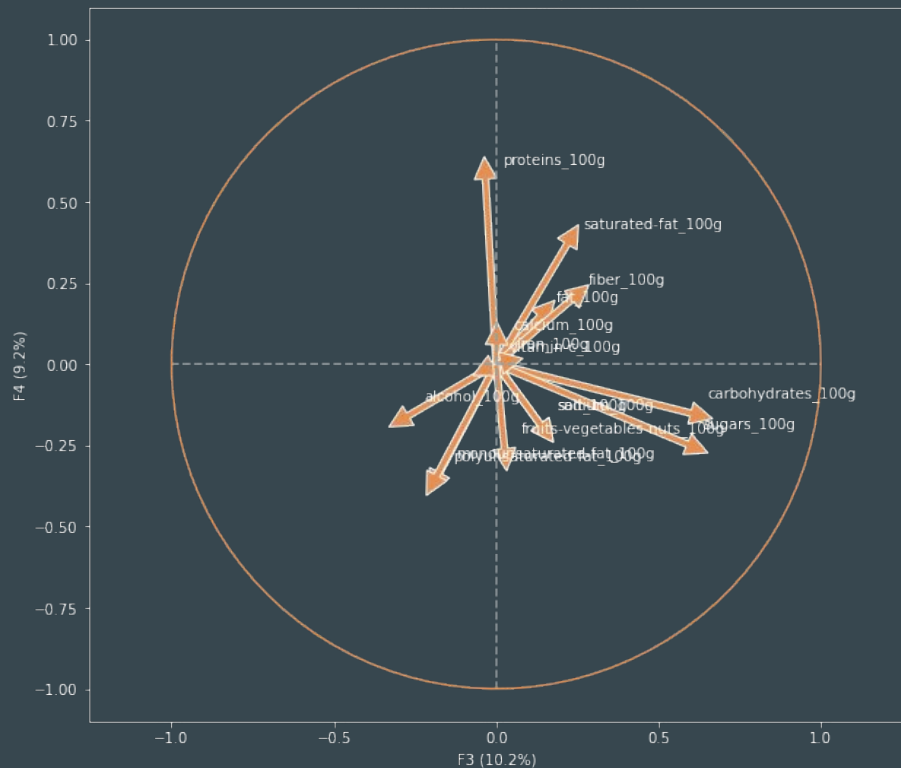


Projection des individus (sur F1 et F2)

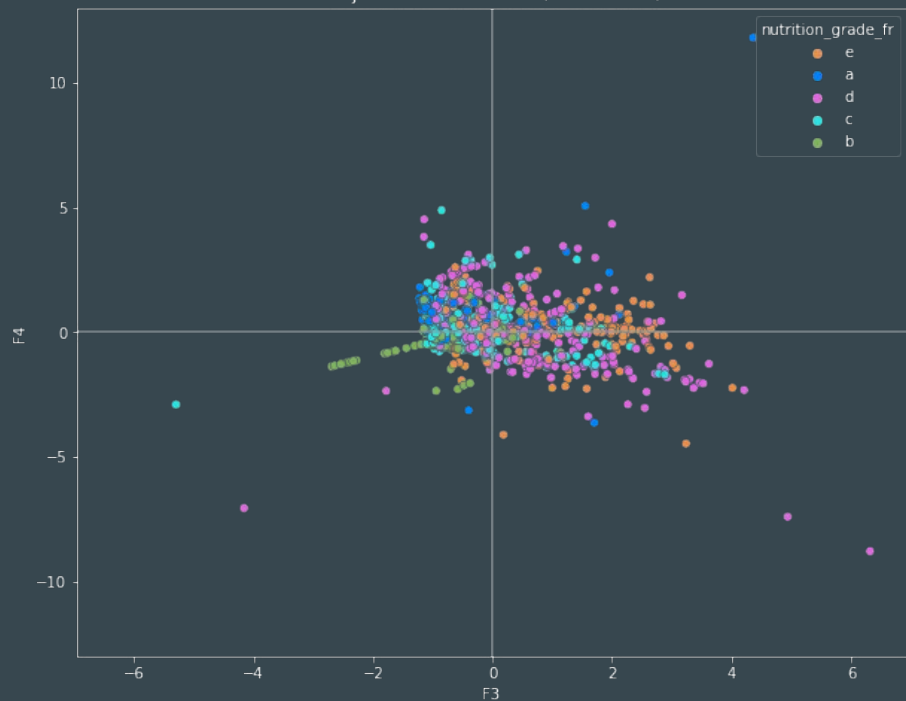


# ACP partielle

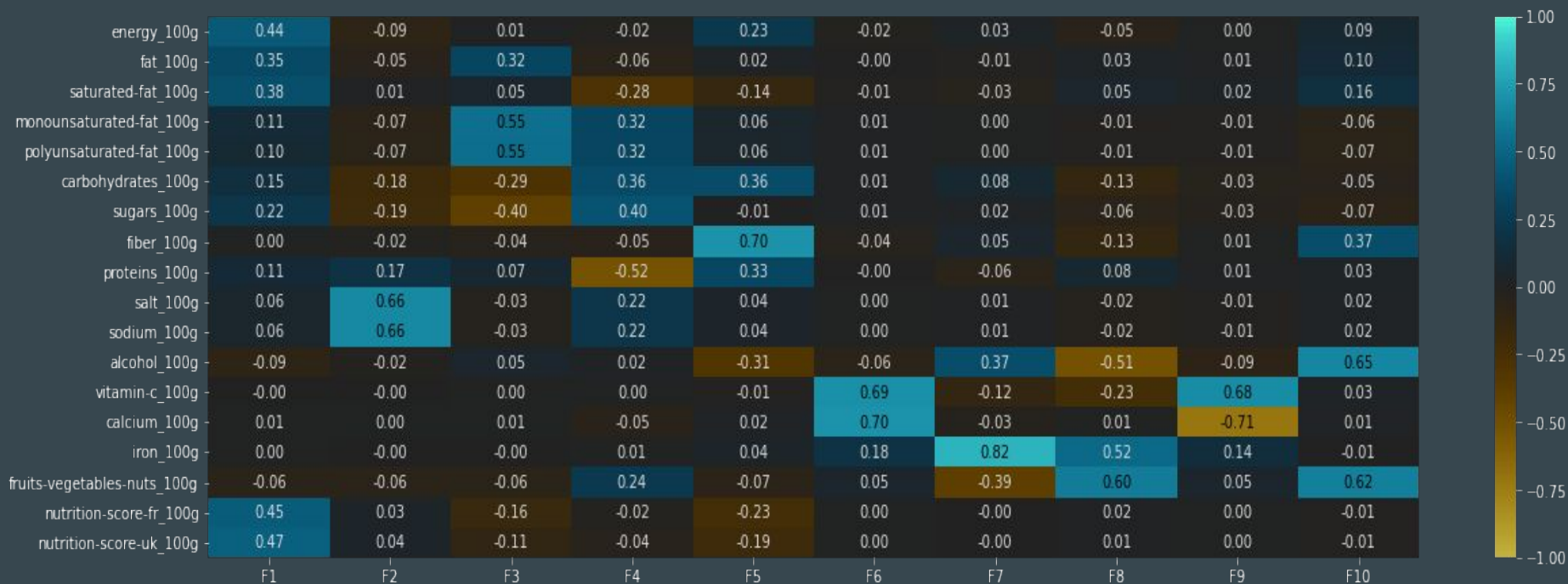
Cercle des corrélations (F3 et F4)



Projection des individus (sur F3 et F4)

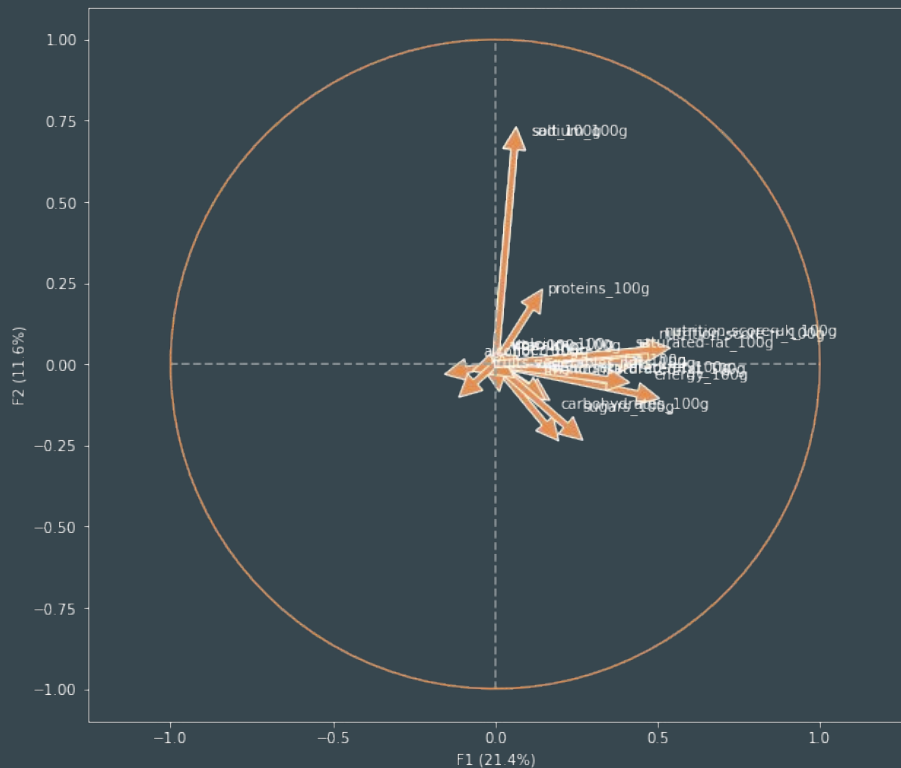


# ACP totale



# ACP totale

Cercle des corrélations (F1 et F2)



Projection des individus (sur F1 et F2)



# Merci de votre attention.

...

Des questions ?