

Projet 3

Concevez une application au service de la santé publique

Recommandation alimentaire pour maladies d' Homocystinurie

ETUDIANTE	: BOUCHRA MEKHALDI
MENTOR	: SOUHAIL TOUMDI
EVALUATEUR	: DAOUDA THIOYE
DATE	: 31/03/22

Problématique

L'agence "Santé publique France" a lancé un appel à projets pour rendre les données de santé plus accessibles a fin d'élaborer une idée innovante d'application en lien avec l'alimentation

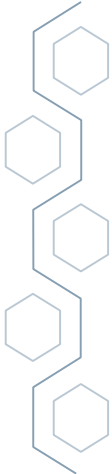


Ce qui est attendu :



- Réaliser une première exploration et visualisation des données, afin que les agents puissent ensuite s'appuyer sur les résultats.
- L'analyse sera basée sur le jeu de données **d'Open Food Facts**, portant sur des produits alimentaires disponibles dans la grande distribution.

Plan de Présentation

- 
- 1 _ Idée d'application
 - 2 _ Nettoyage des données
 - 3 _ Analyse exploratoire
 - 4 _ Exemple d'application
 - 5 _ Conclusion

Idée d'application

Moteur de Recommandation Alimentaire pour les personnes atteintes de **Homocystinurie**
(Maladies Métaboliques Héréditaires)

➤ **Homocystinurie ?**

C'est une maladie métabolique héréditaire.

L'HCU est provoquée par une carence en l'enzyme nommée cystathionine bêta-synthase (CBS). Cela rend l'organisme incapable de métaboliser un acide aminé nommé méthionine (MET). Cela provoque une accumulation nocive de méthionine et d'une autre substance nommée homocystéine (HCY) dans le sang



Régime hypo protidique → Peu de protéines

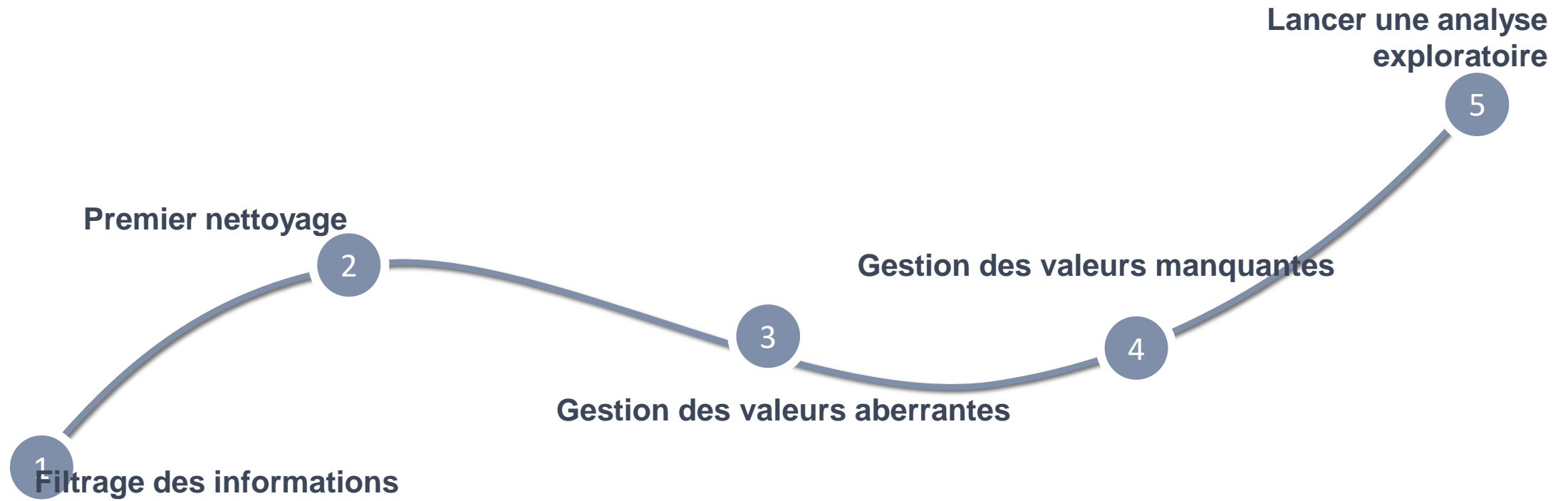
Alimentation saine → Nutri-score faible
Peu de sel
Peu d'additifs
Pas huile de palme
Sucre limité
Peu calorique

Produits disponibles

- Nom du produit
- Photo
- Catégorie
- Marque
- Vendus en France

Nettoyage des données

Approche méthodologique

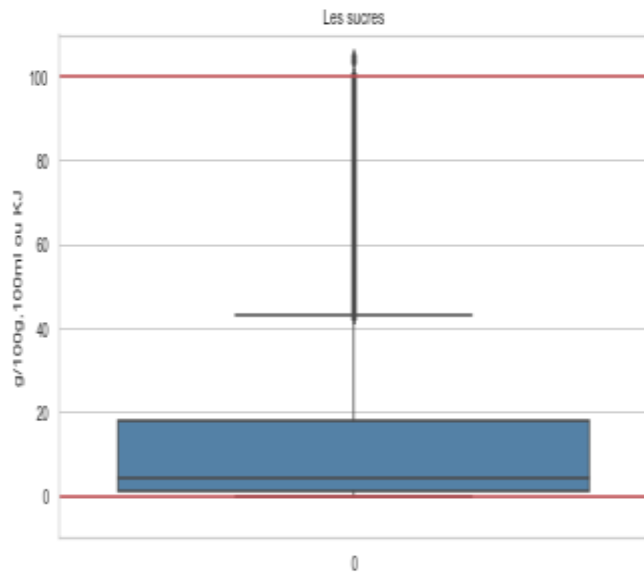


Filtrage des informations → Conservation des produits vendus en France

Premier nettoyage →

- ✓ Suppression des métadonnées
- ✓ Suppression des 2 doublons sur le code produit
- ✓ Sélection Informations généraux
- ✓ Suppression des variables avec %NaN > 60%
- ✓ Suppression des lignes sans nom produit

Gestion des valeurs aberrantes



Outliers pour 100g ou ml de produit

- énergie nutritionnelle > 900 kcal ou > 3765.6 kJ
- nutriments _100g > 100g (ou 100ml)
- nutriments _100g < 0g (ou 0ml)

Macro-nutriment et sous-groupe de nutriment

- la masse totale des glucides inférieure à la masse de sucre
- la masse totale des lipides est inférieure à la masse des acides gras saturés.
- la masse de sodium est supérieure à la masse de sel (en g).

Gestion des valeurs manquantes

Remplissage par « 0 »

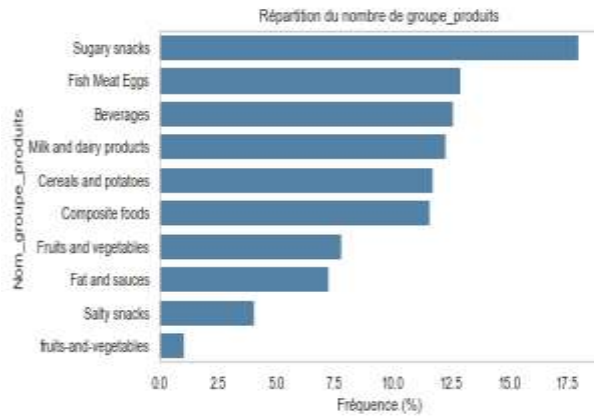
Remplissage par la médiane

imputation par KNNImputer

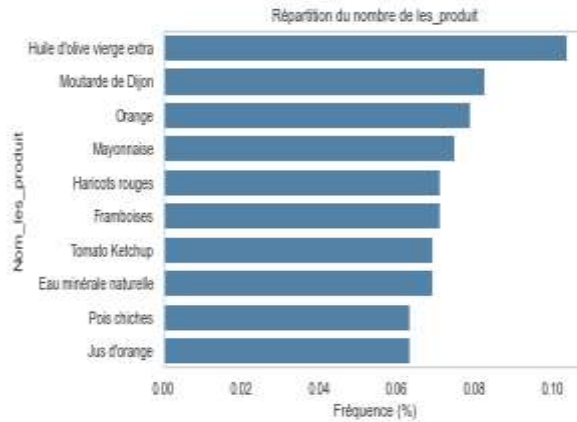
Analyse exploratoire

Analyses univariées

Variables Qualitatives

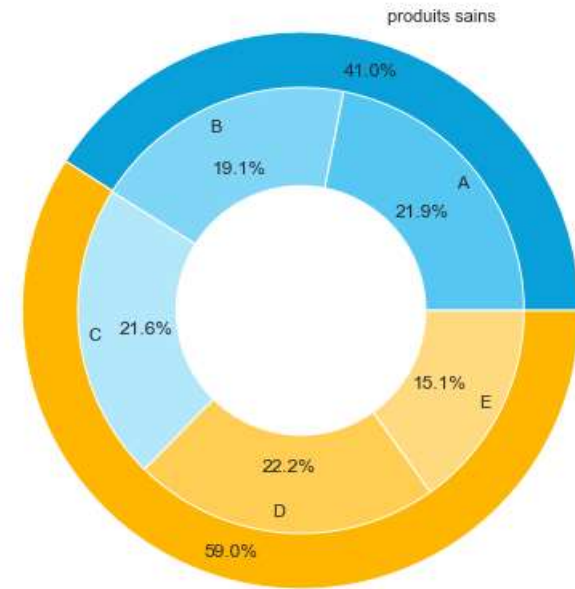
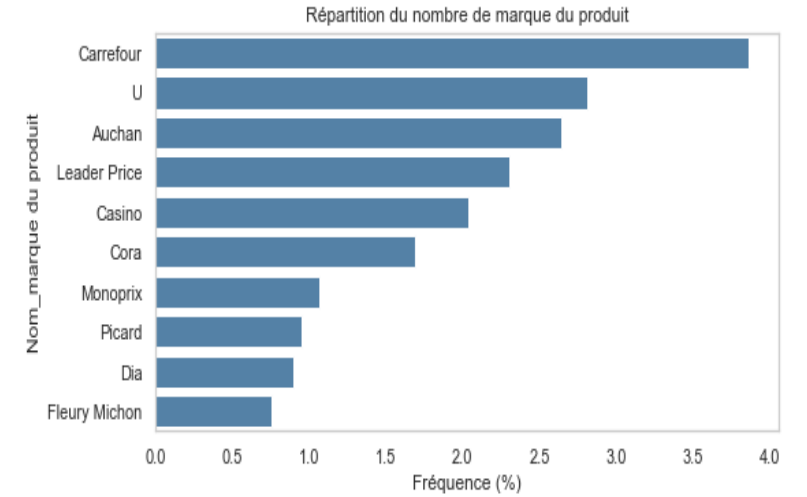


pnnns_groups_1



product_name

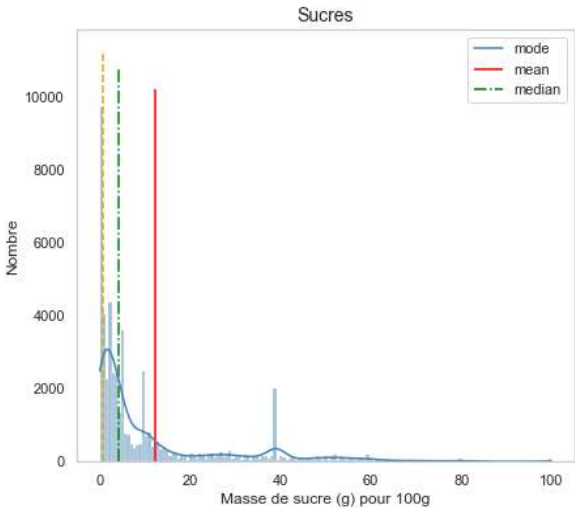
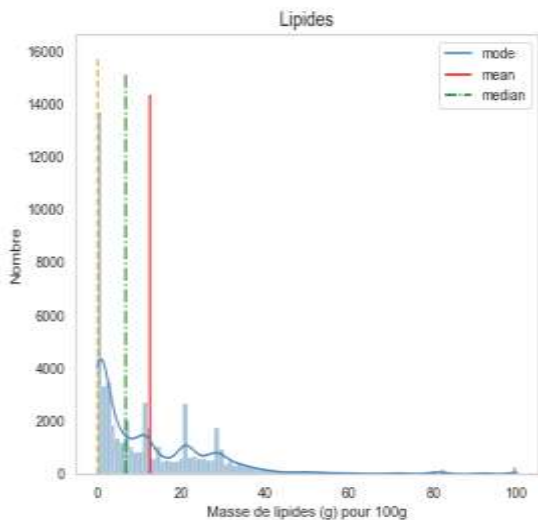
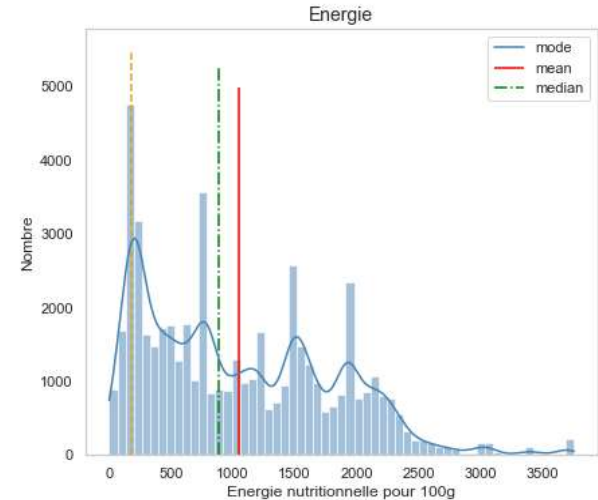
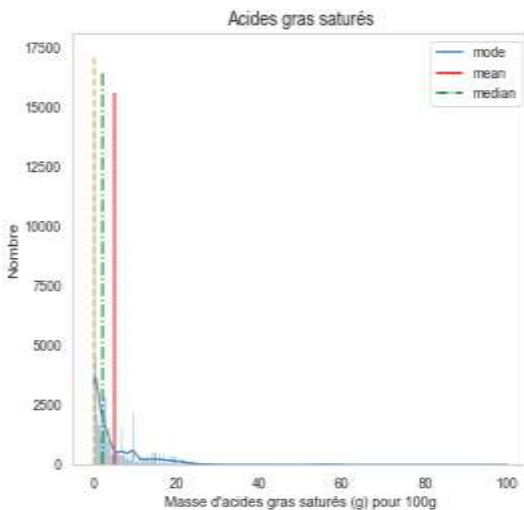
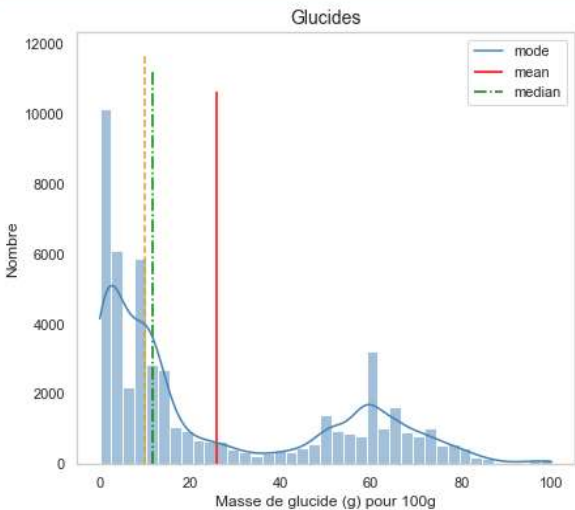
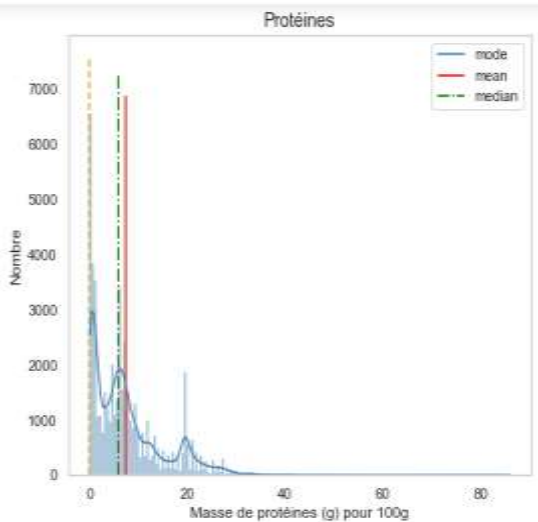
brands



produits malsains Nutri_score

Analyses univariées

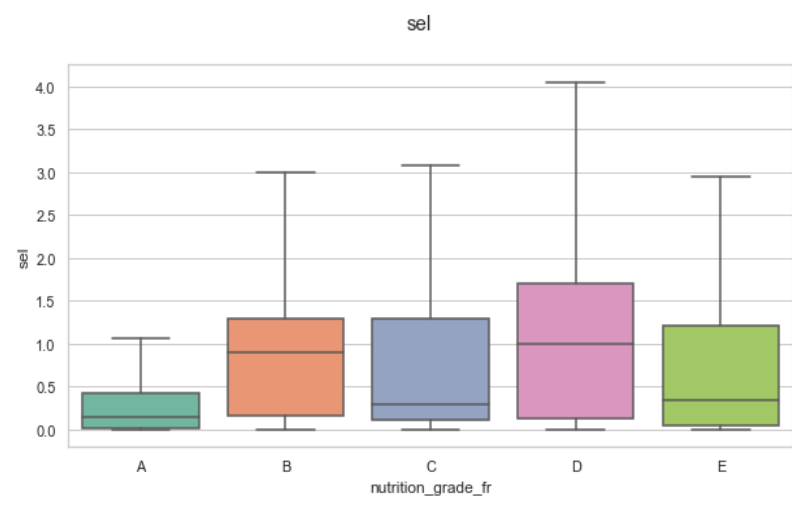
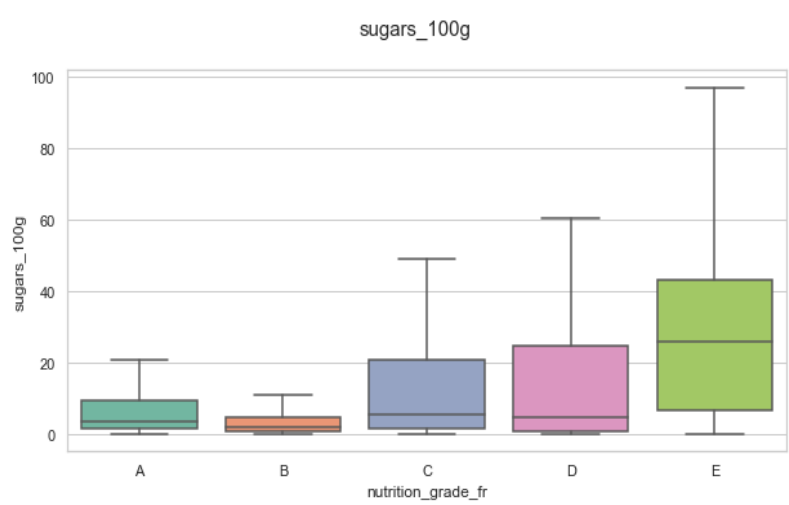
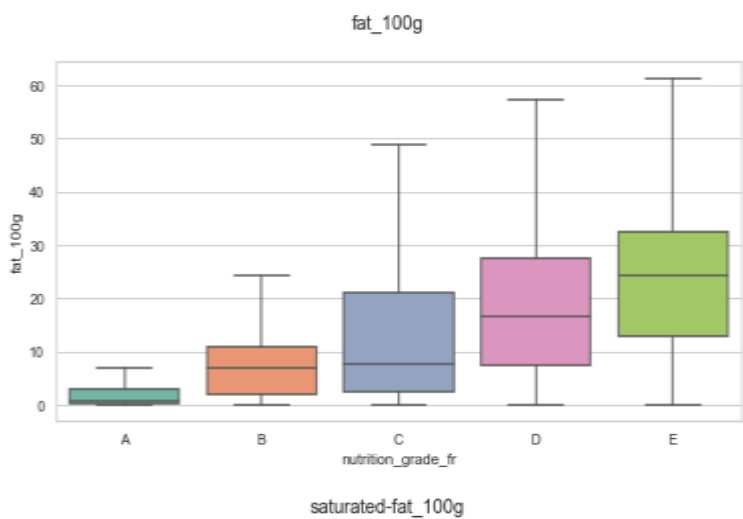
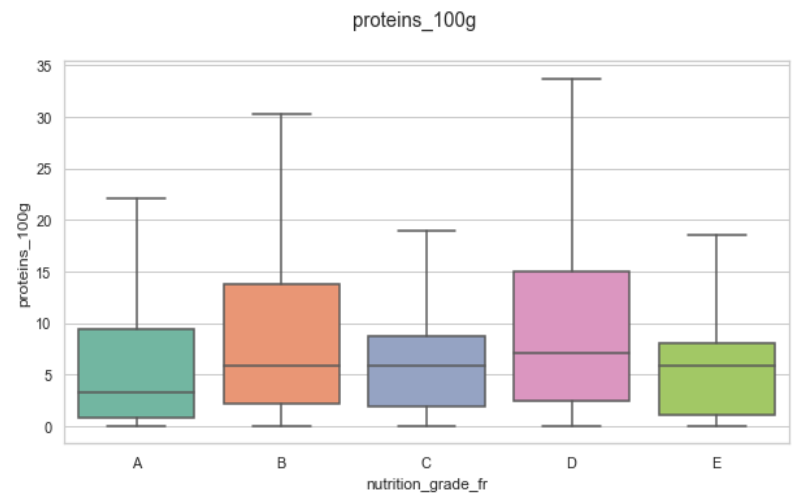
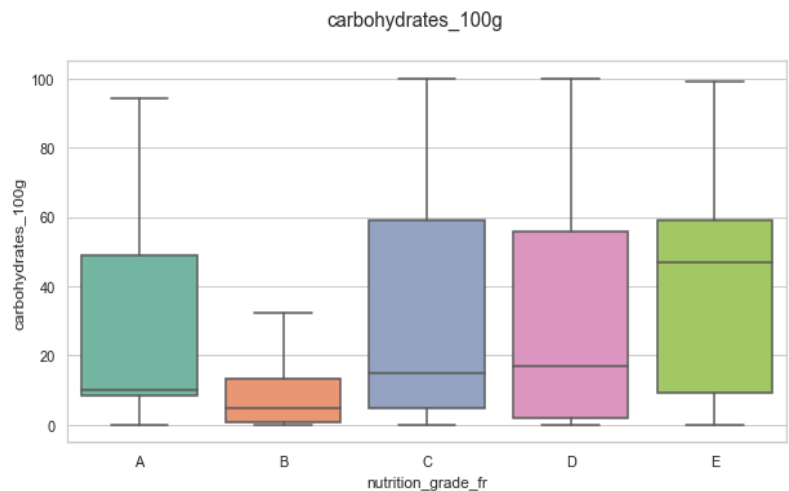
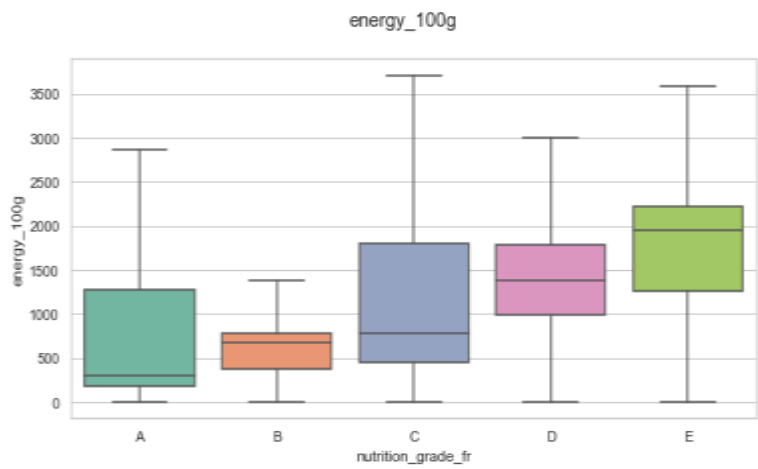
Variables Quantitatives



Desc	energy_100g	fat_100g	saturated-fat_100g	carbohydrates_100g	sugars_100g	proteins_100g	sel
mean	1051.287714	12.680111	4.902931	25.891689	12.165605	7.436662	0.821558
median	885.000000	6.900000	2.000000	11.800000	4.300000	5.800000	0.430000
var	573359.190529	253.621540	57.144687	714.327075	305.768127	52.367700	1.935080
std	757.204854	15.925500	7.559410	26.726898	17.486227	7.236553	1.391071
skew	0.612814	2.319976	3.589049	0.782891	2.020371	1.302318	14.793307
kurtosis	-0.288412	7.643637	21.918410	-0.892303	3.956578	1.782721	541.969359
mode	0 184.0	0 0.0	0 0.0	0 10.0	0 0.6	0 0.0	0 0.0
Min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Max	3760.000000	100.000000	100.000000	100.000000	100.000000	86.000000	76.200000

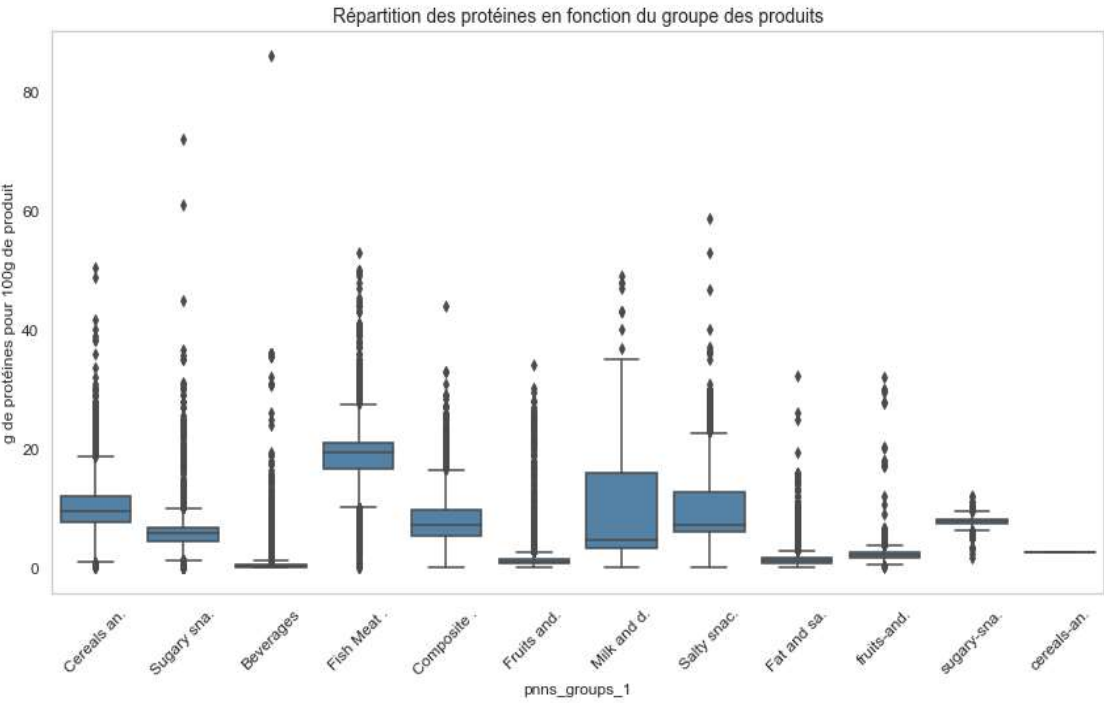
Analyses bivariées

Nutri-grade/Variables quantitatives

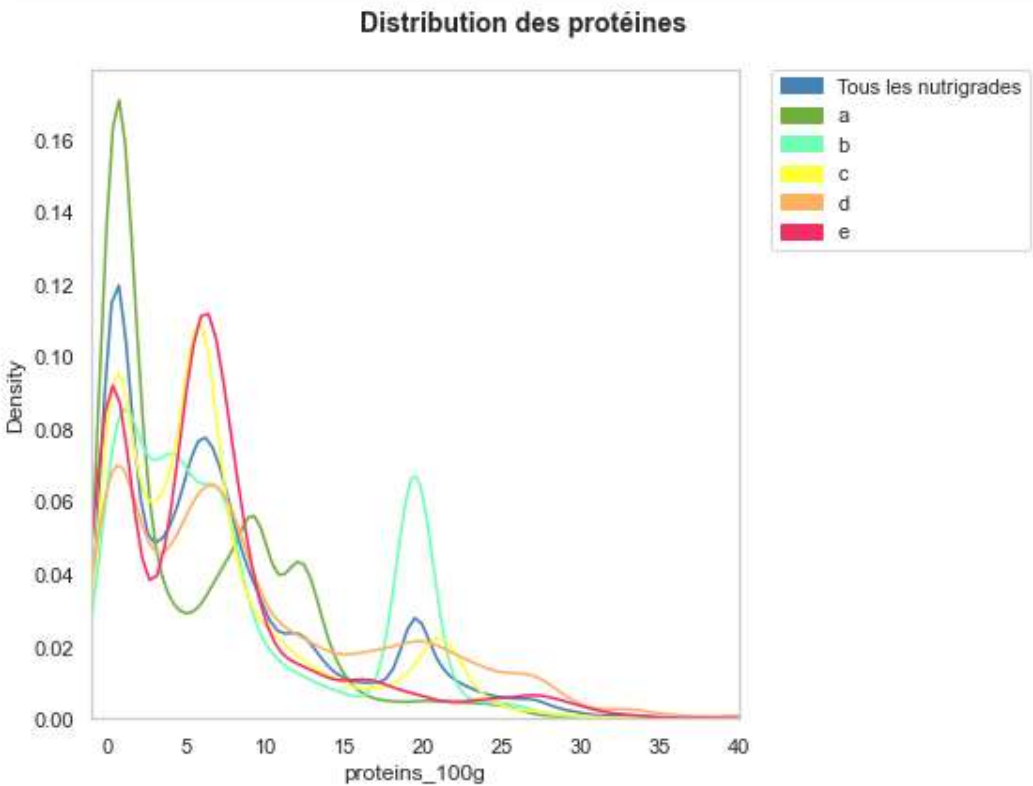


Analyses bivariées

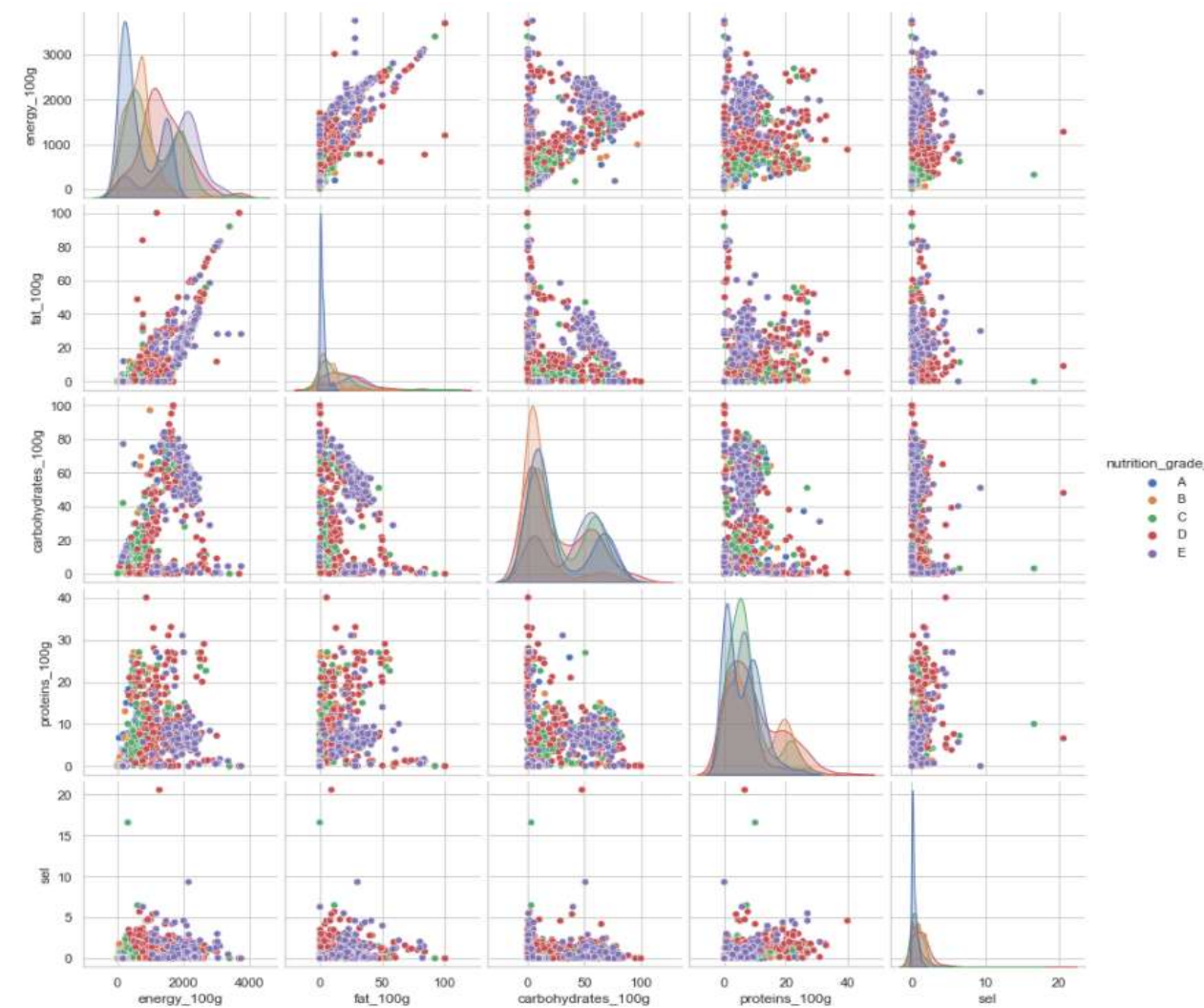
protéines / Groupes des produits



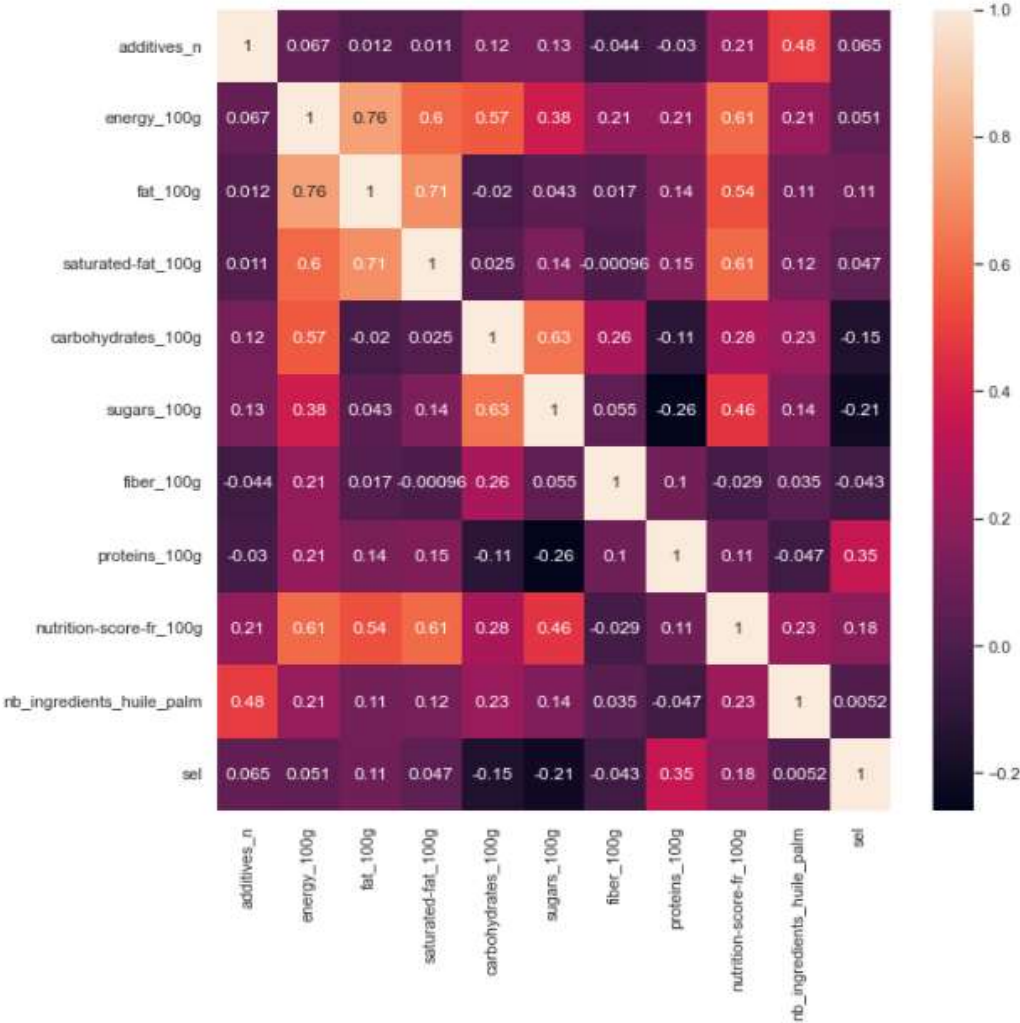
protéines / nutri_garde



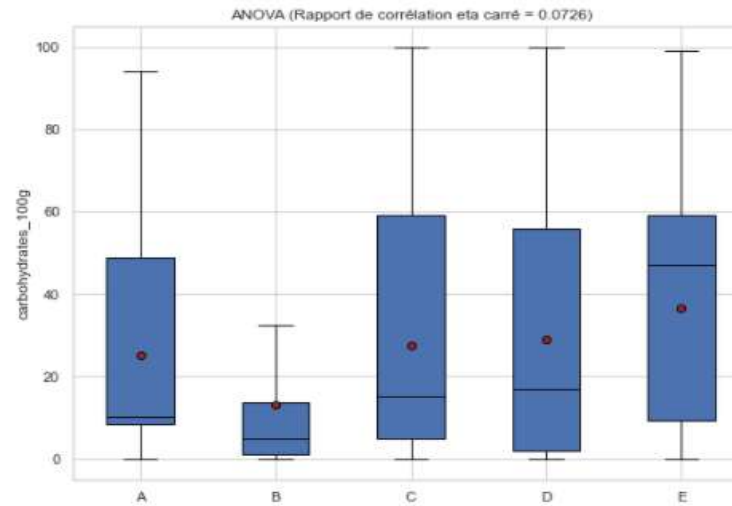
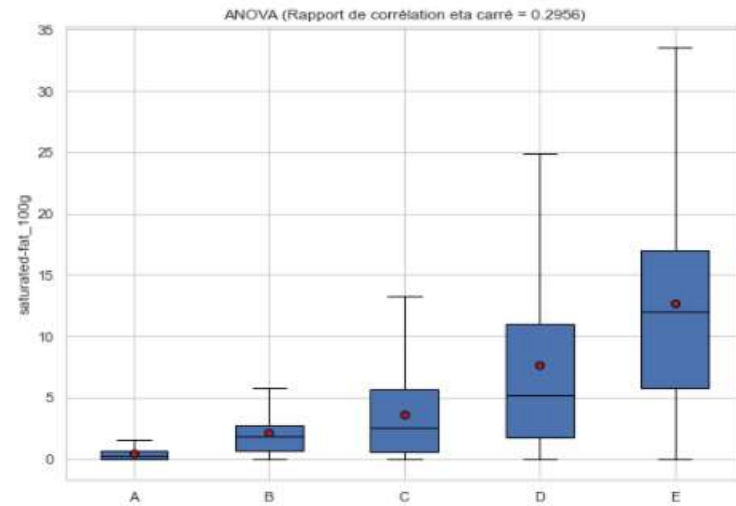
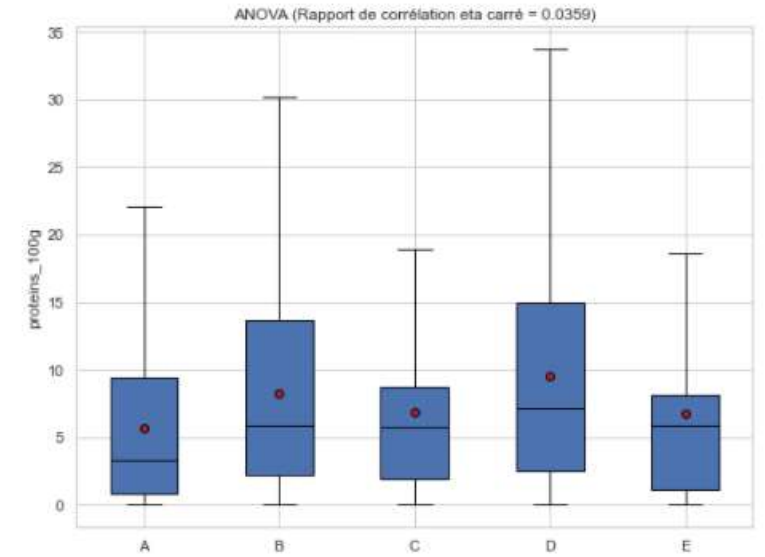
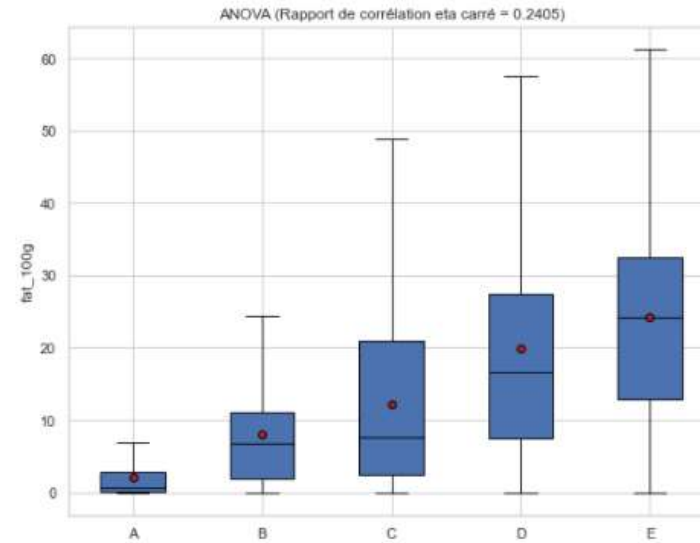
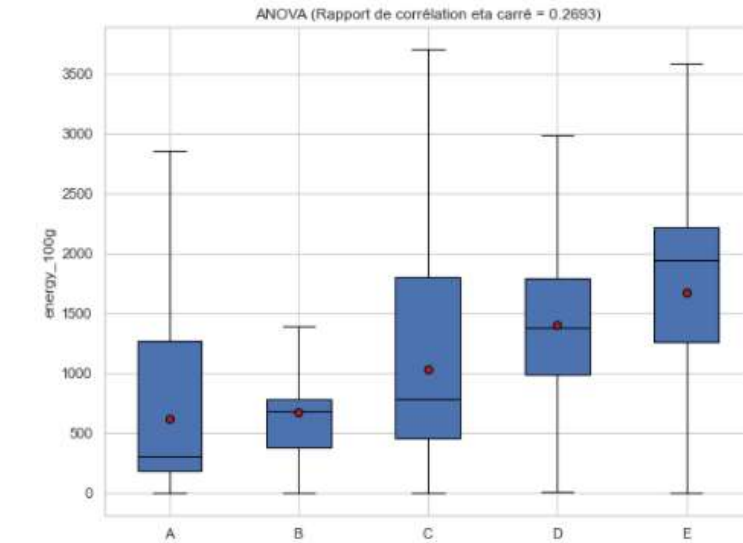
Analyses multivariée



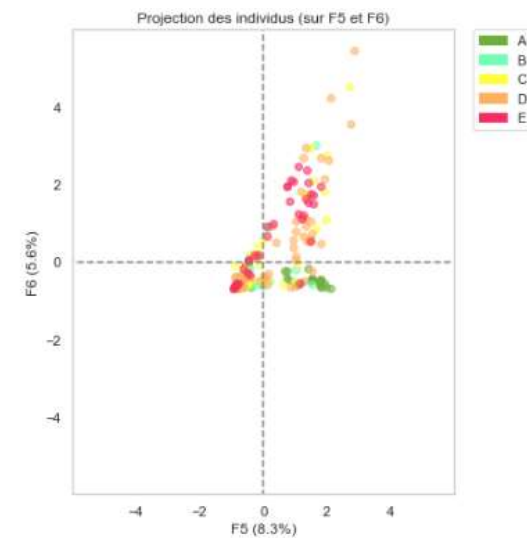
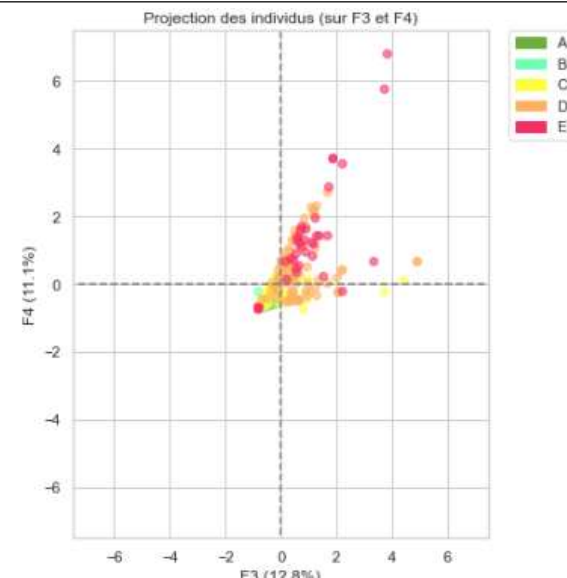
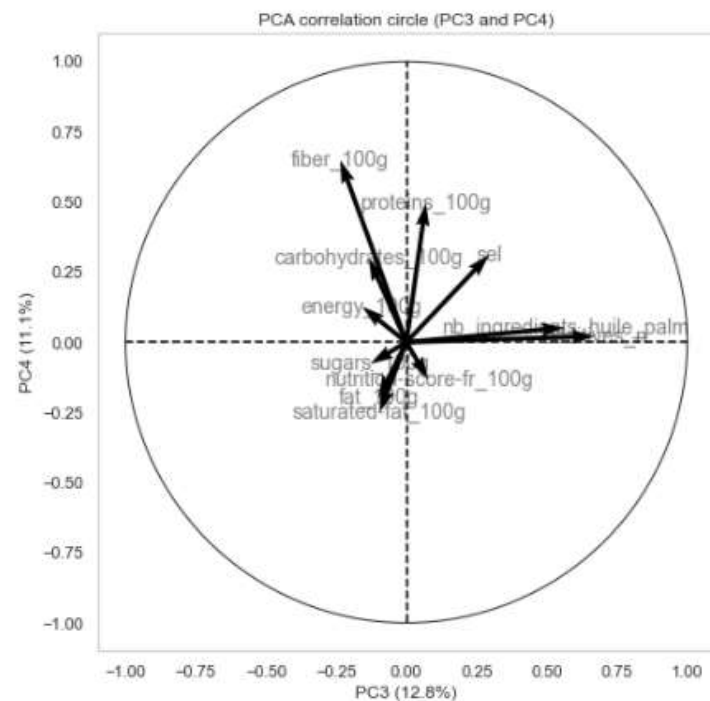
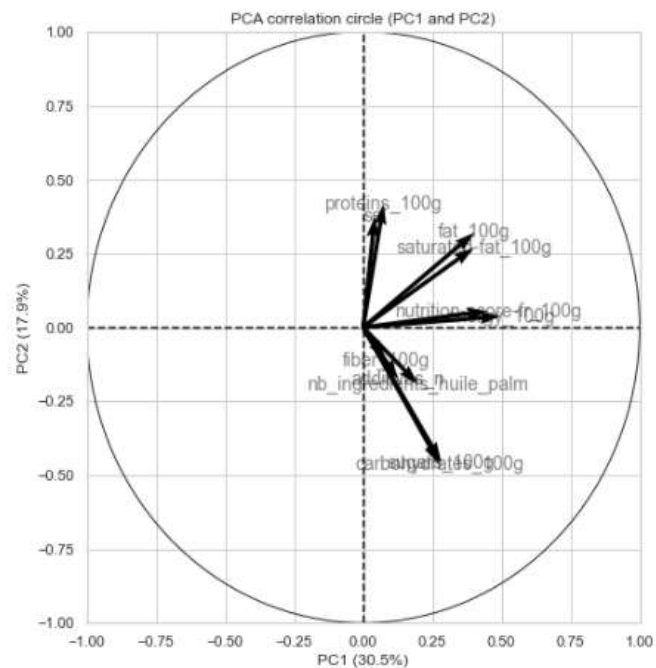
matrice de corrélation



Analyses ANOVA



ACP : Analyse en Composante Principale



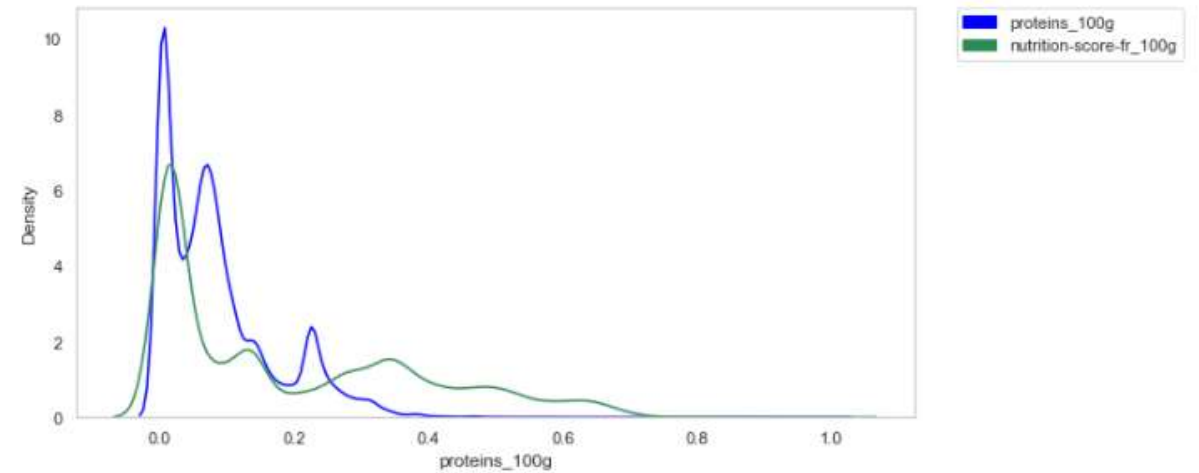
Exemple d'application

Moteur de recommandation -Score

Sélection des variables de
scoring : protéines, nutri score

Mise à l'échelle :
MinMaxScaler

Scoring avec pondération :
Protéines : 10 Nutri score : 2



Moteur de recommandation

Prétraitement des noms de produit:

Vectorisation :
TfidfVectorizer
de scikit learn

DataFrame de comparaison :





Moteur de recommandation :

```
# test moteur -  
produit = input(str('Quel est le produit recherché ? : '))  
rech_produits(produit)
```

Quel est le produit recherché ? : yaourt

Produits recommandés :

Remarque : L_G_Su_Se_Ad_Hp : lipides - Glucides - Sucre - Sel - Additif - huile palme

	image_url	proteins_100g	nutrition_grade_fr	nutrition-score-fr_100g	brands	L_G_Su_Se_Ad_Hp
product_name						
Yaourt nature		3.0	C	3.000000	Delisse	12.0-4.8-11.0-0.103-0.0-0.0
Yaourt nature	NaN	3.8	A	0.508772	Leader Price	1.1-5.0-4.7-0.16-0.0-0.0
Yaourt nature		3.8	A	0.508772	Bonnetterre	3.5-5.9-4.7-0.16-0.0-0.0
Yaourt Nature	NaN	3.9	A	0.508772	Dia	1.0-5.0-4.5-0.1016-0.0-0.0
Yaourt Nature		4.0	A	0.508772	Auchan	1.4-4.2-4.2-0.127-0.0-0.0
Yaourt nature		4.0	A	0.508772	Carrefour	1.0-4.3-4.3-0.127-0.0-0.0

Conclusion

- Utilisation du jeu de donnée Open Food Fact
- Analyse du data set et modélisation du Nutri-score
- Erreur du modèle correcte
- Réalisation de l'Application

Amélioration

- Produits vendus hors de France (vacances, voyages scolaires)
- Conseils des **médecins/diététiciens**
- Magasin où **trouver** le produit **proche de chez soi**
- Consulter les **associations** de personnes atteintes de HUC
- Scanner pour les enfants avec logos interdits, autorisés