

AMÉLIORATION DE BASE DE DONNÉES DE SANTÉ

Pour le compte de « Santé publique France »



Nettoyage des données
brutes



Analyse des valeurs
manquantes



Les différentes analyses
(univariée, bivariée,
multivariée)



ACP (cercles de
corrélations, projection
des individus)



Conclusion

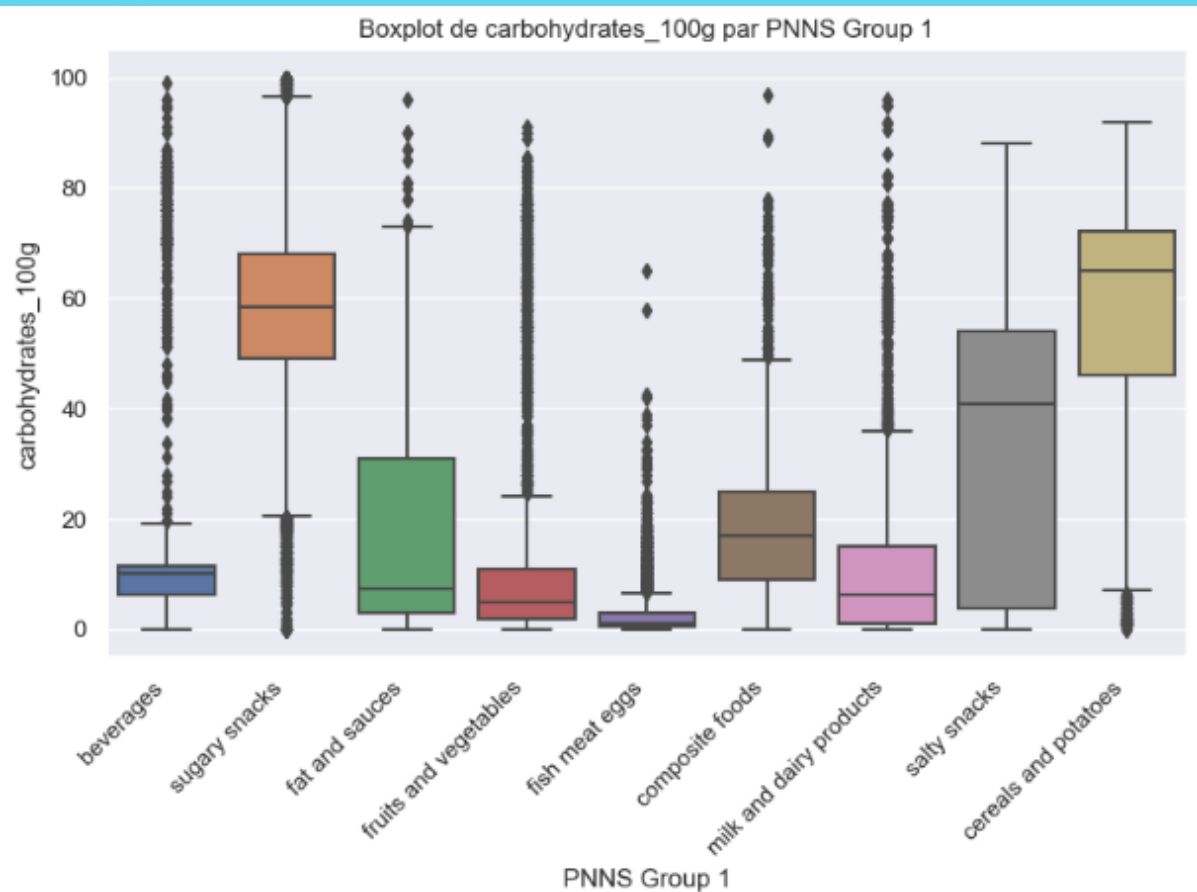
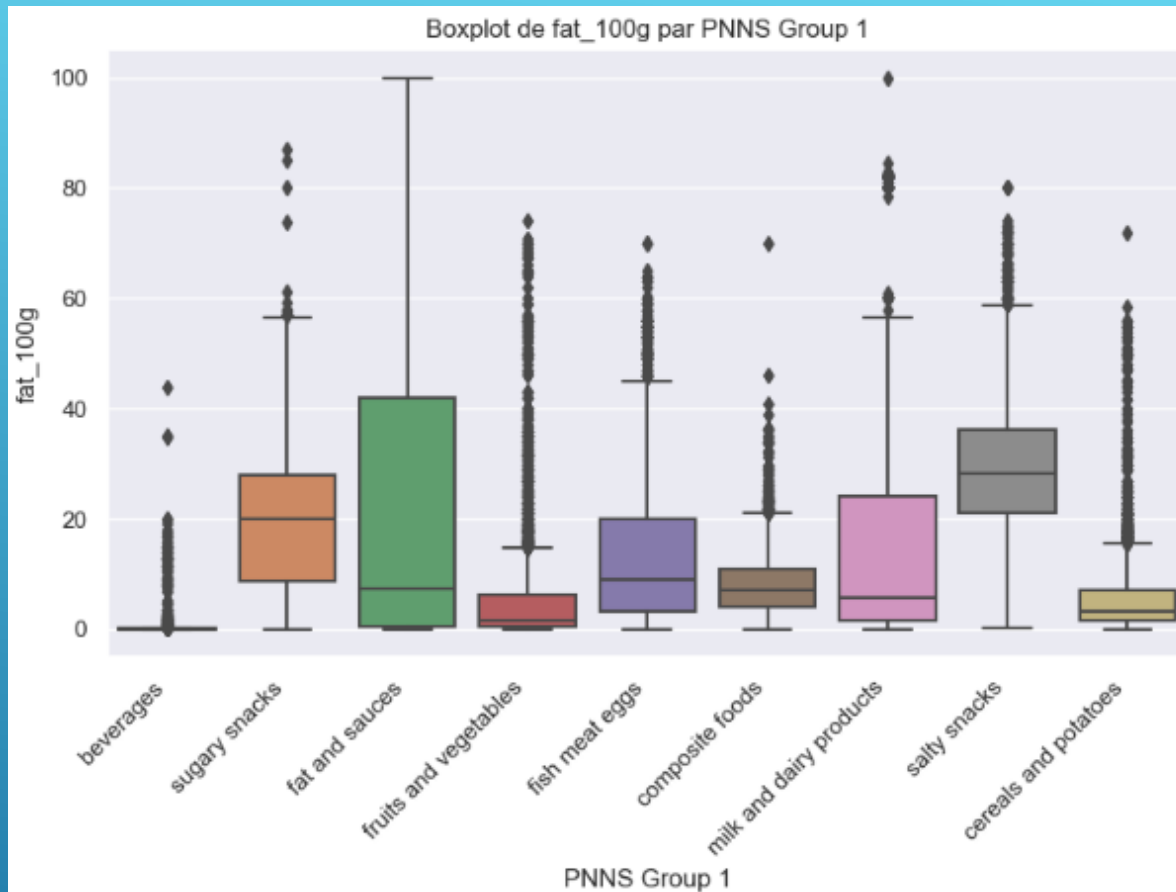
SOMMAIRE

- ▶ Centralisation de l'analyse sur la France (métropole et DOM-TOM)
- ▶ Nettoyage des différences d'écritures selon les valeurs des variables
- ▶ Imputation de valeurs par médiane pour différentes variables

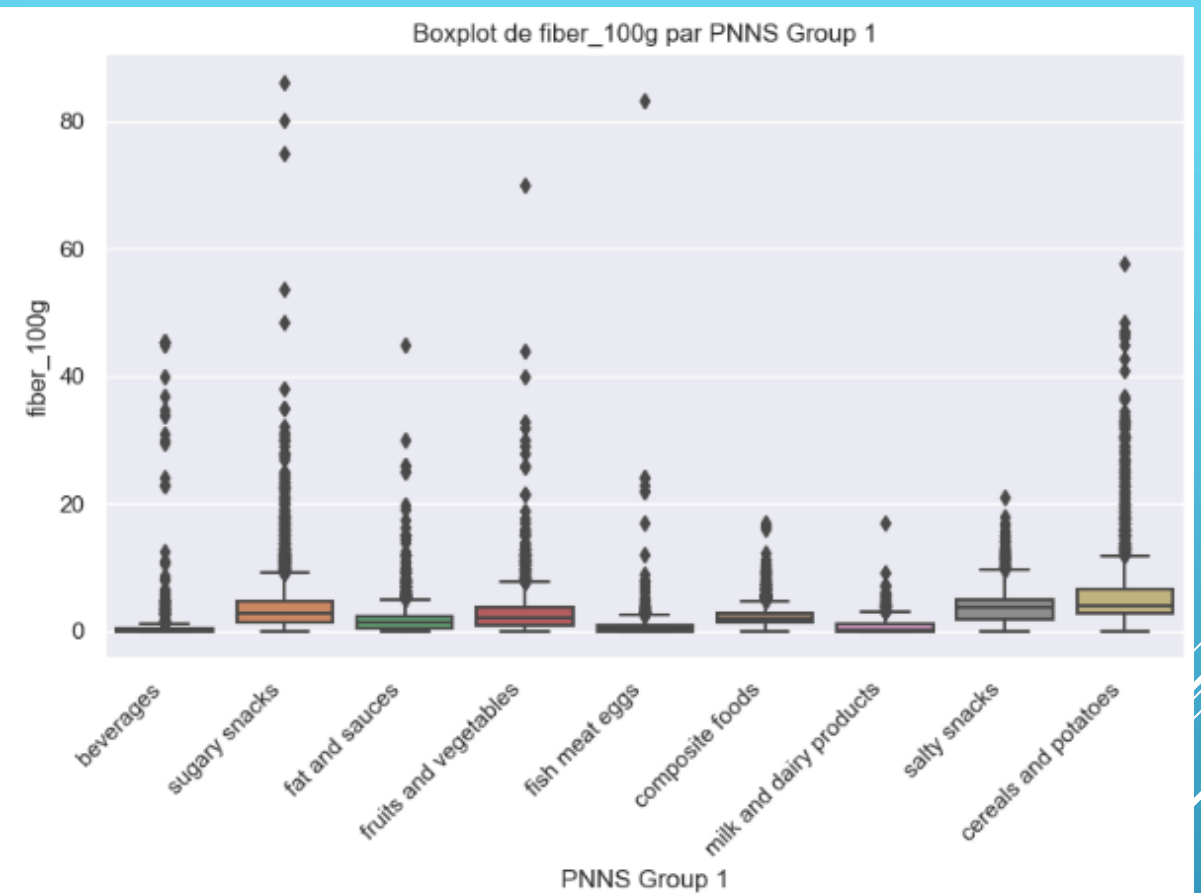
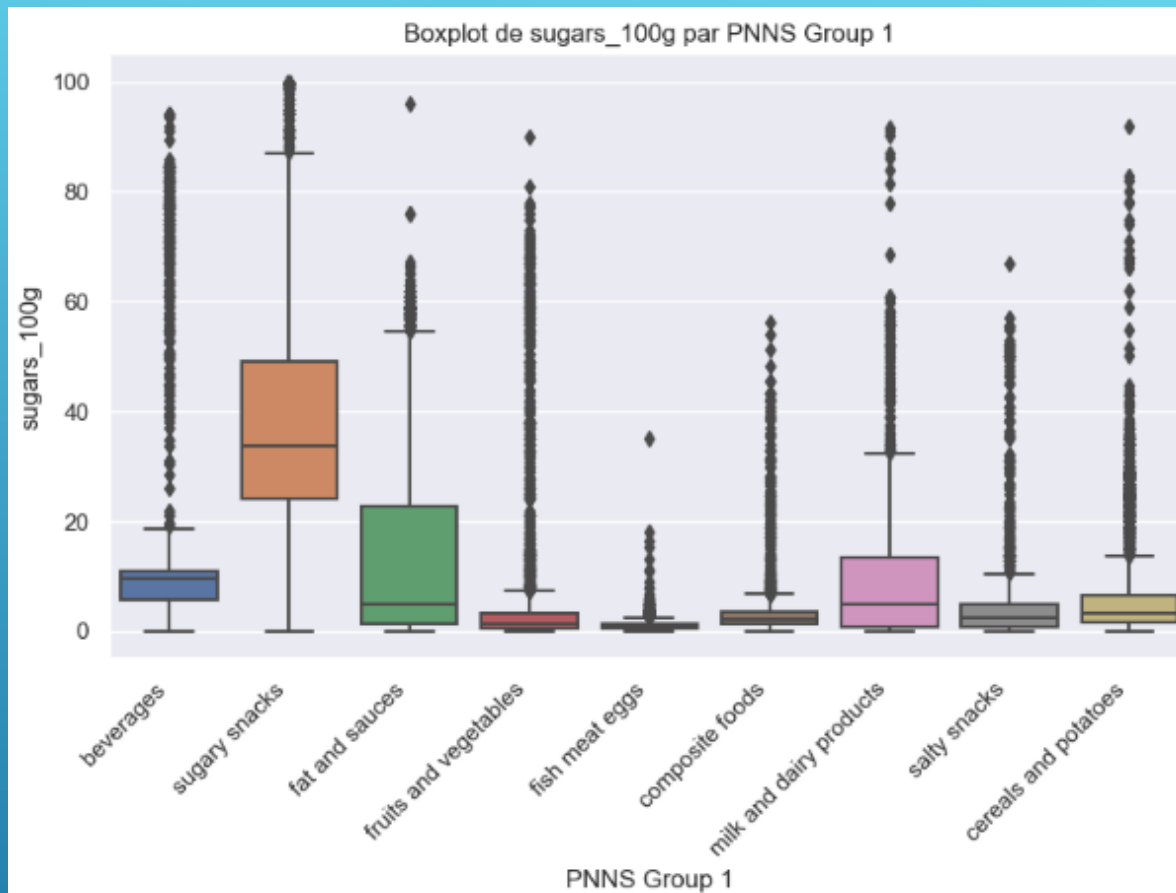
NETTOYAGE DES DONNÉES BRUTES

- ▶ L'analyse des valeurs manquantes a montré beaucoup de valeurs non remplies
- ▶ Selon les différentes variables, certaines ne seront pas choisies, car non pertinentes
- ▶ Une fois les variables choisies, un nettoyage supplémentaire sera mis en place
- ▶ Le nettoyage dans son ensemble a permis de passer de 320 772 à 41 234 individus

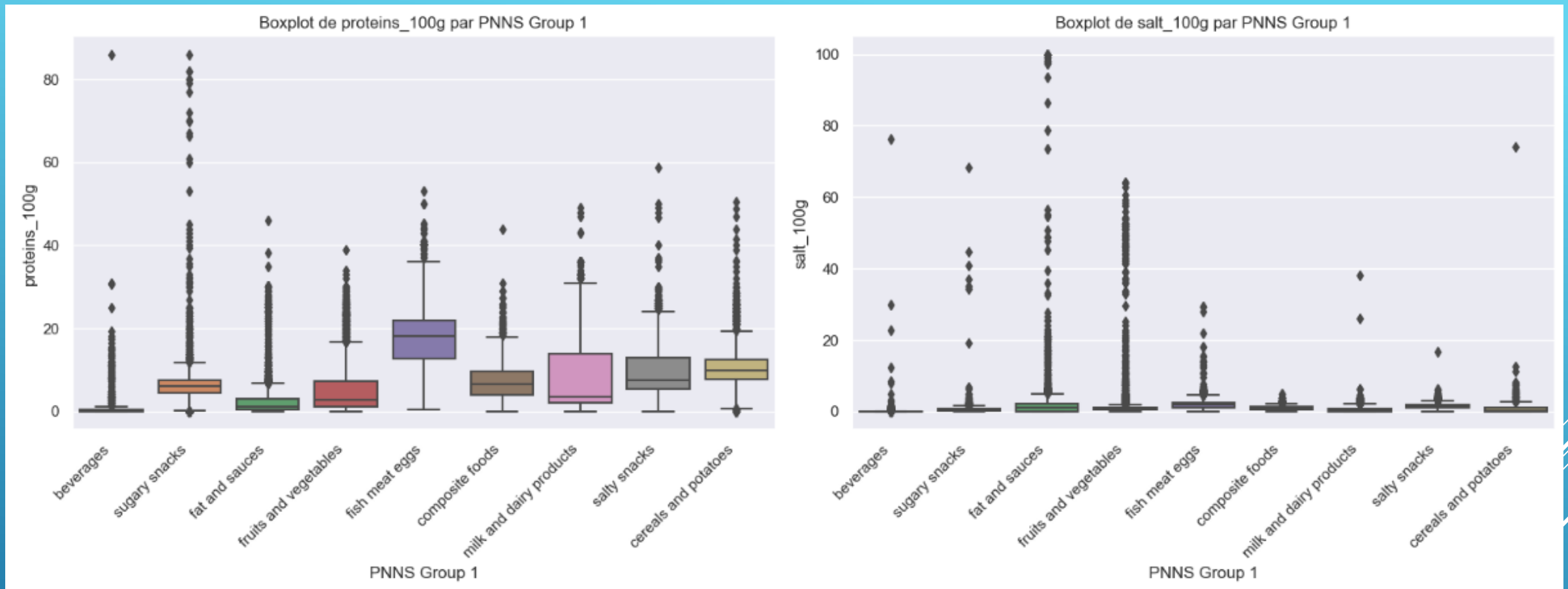
ANALYSE DES VALEURS MANQUANTES



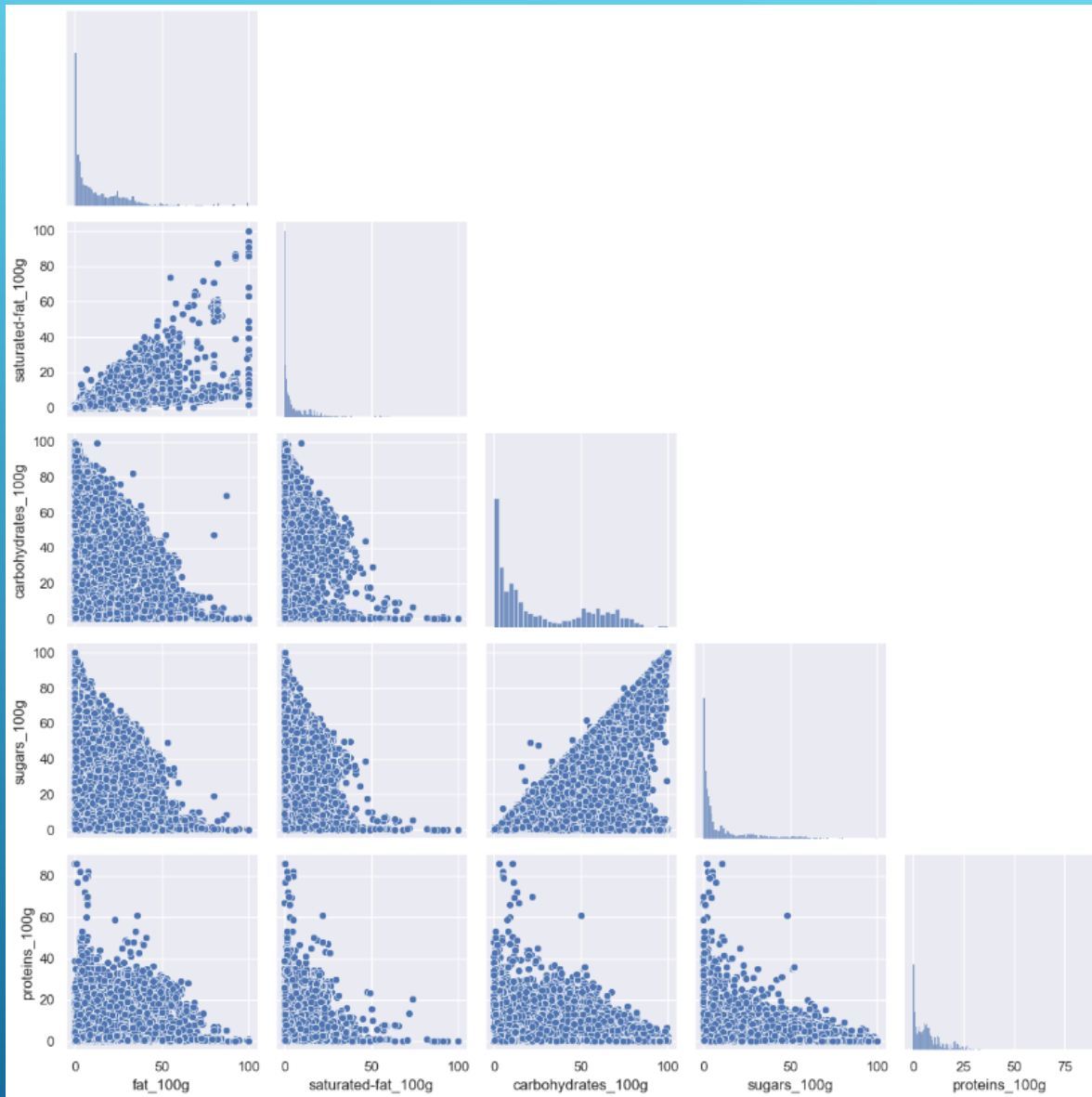
ANALYSE UNIVARIÉE - BOXPLOT



ANALYSE UNIVARIÉE - BOXPLOT

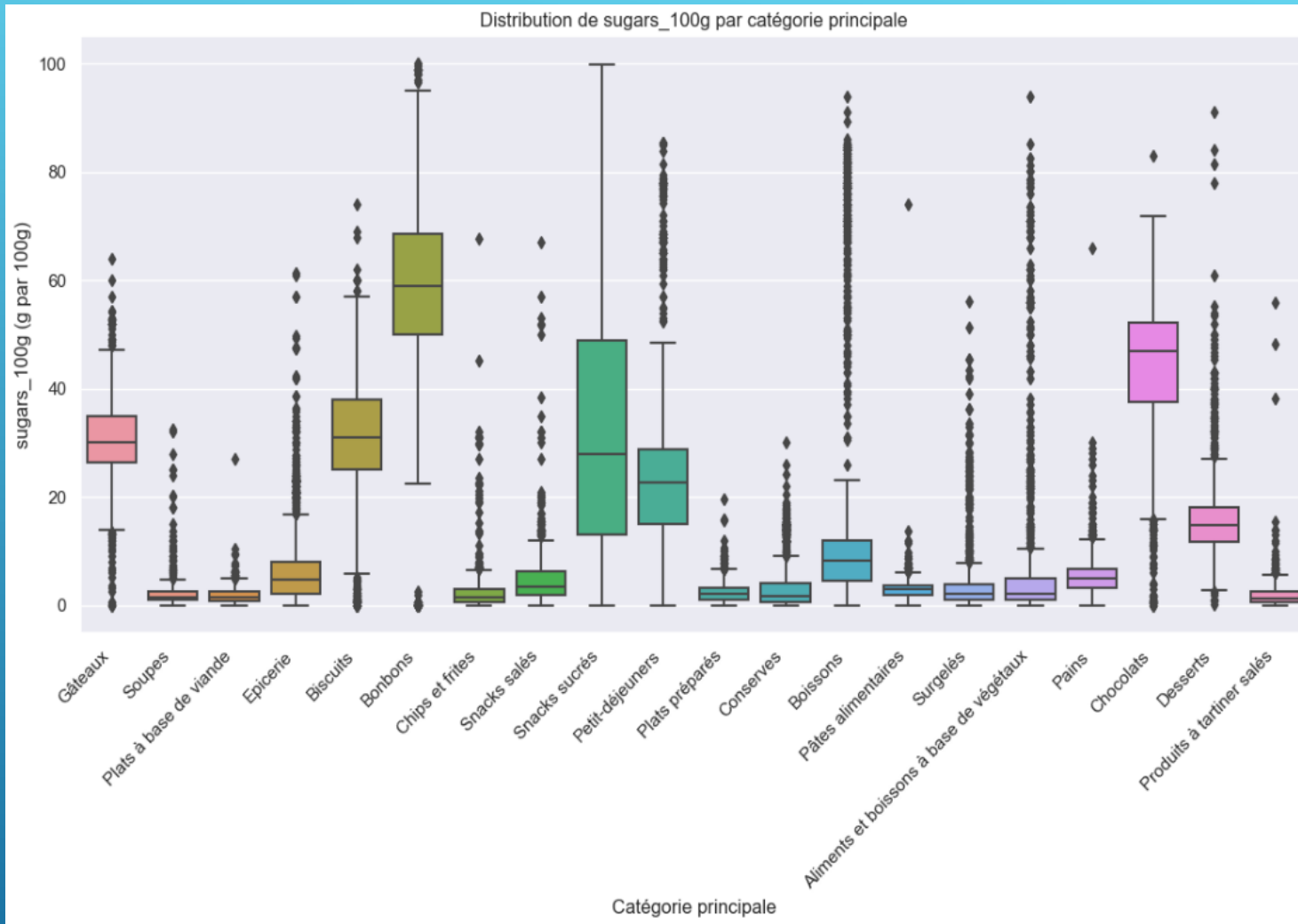


ANALYSE UNIVARIÉE - BOXPLOT



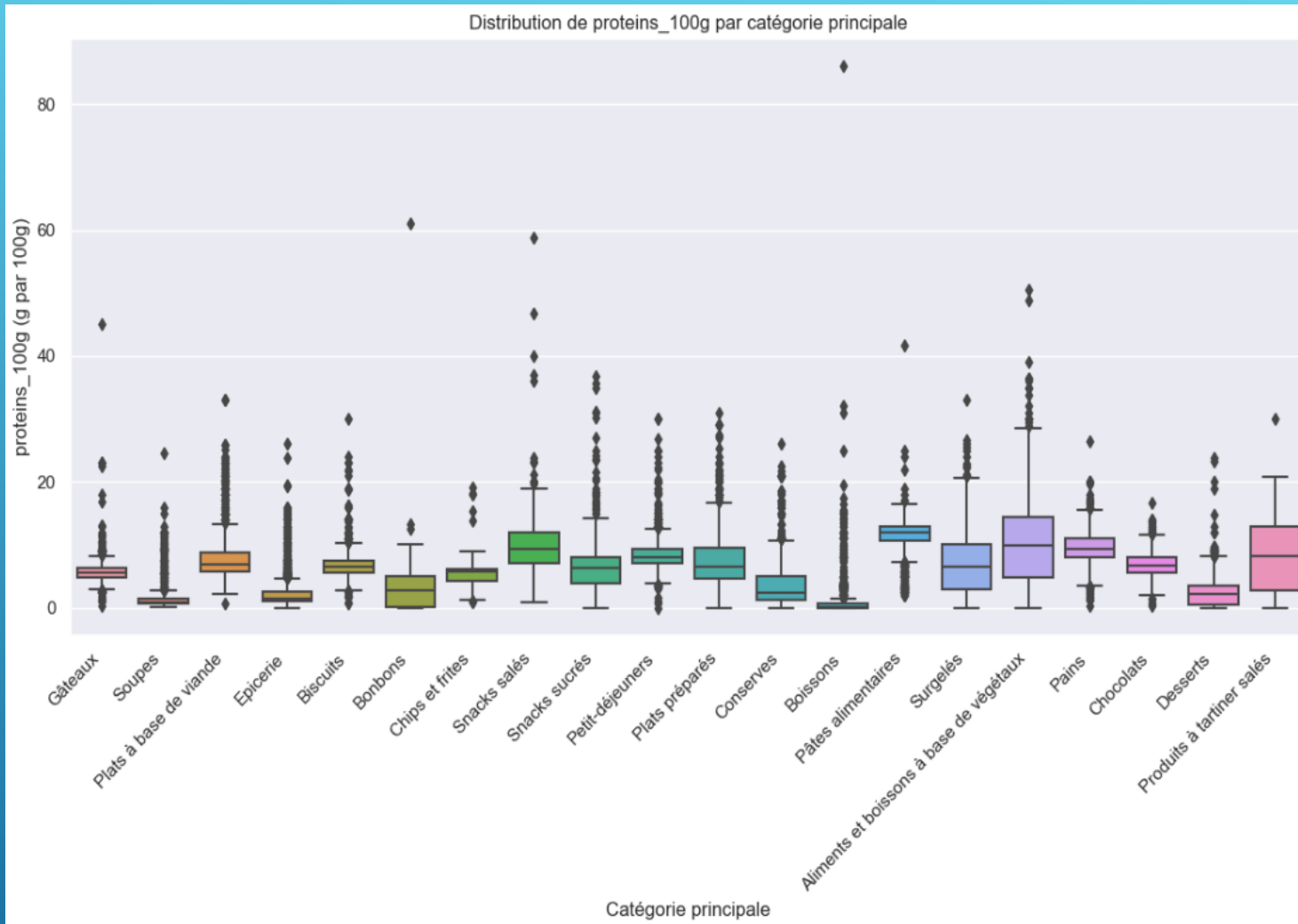
- ▶ Chaque distribution est plus ou moins asymétrique
- ▶ Cela indique une majorité de valeurs basse et quelques valeurs extrêmes
- ▶ « fat » et « saturated-fat » une corrélation positive assez forte
- ▶ Résultat similaire pour « sugars » et « carbohydrates »

ANALYSE MULTIVARIÉE - PAIRPLOT



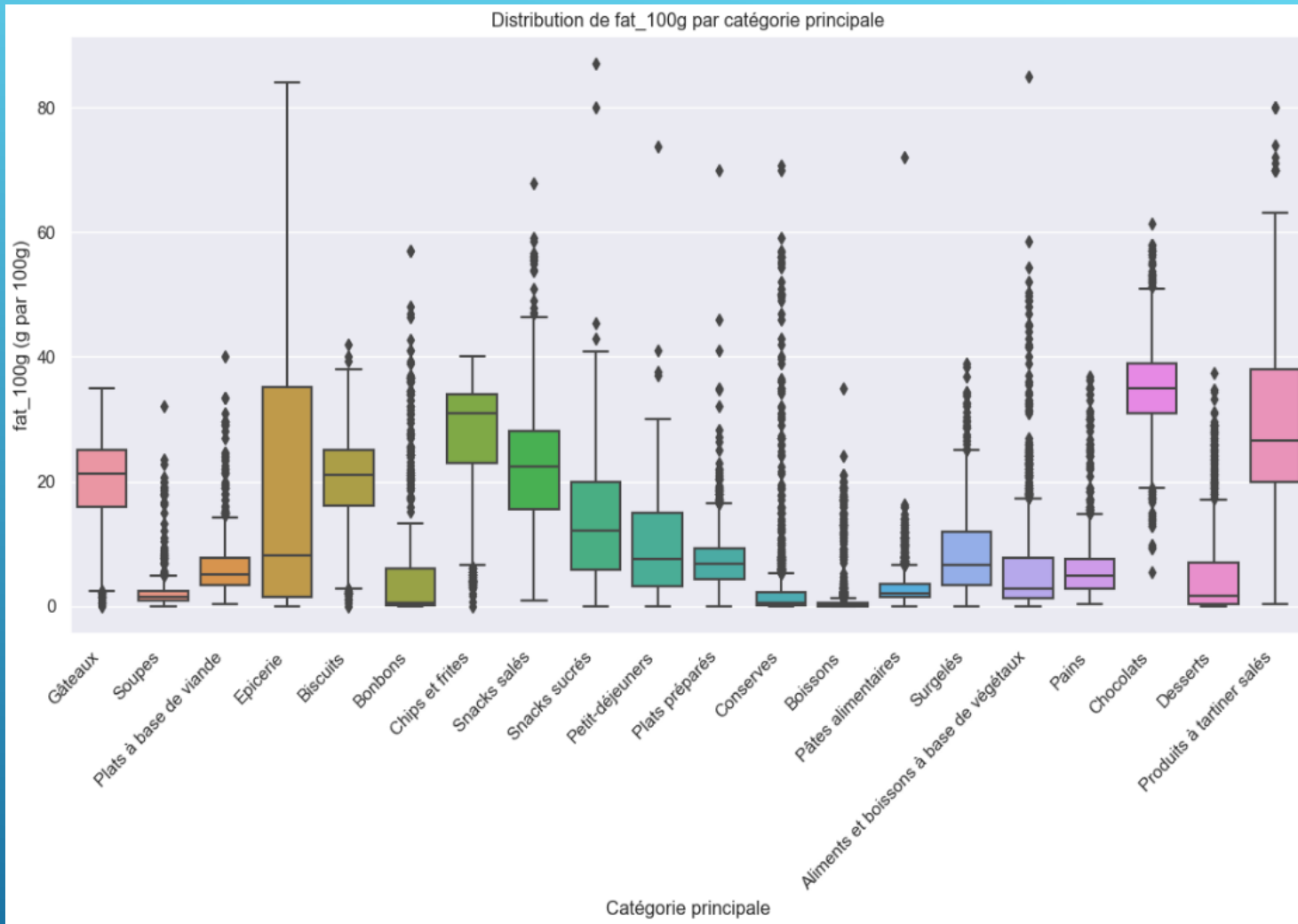
- ▶ On retrouve une grande diversité de teneurs en sucre selon les catégories
- ▶ De nombreuses valeurs extrêmes
- ▶ Les plats préparés et à base de viande ont une faible teneur

ANALYSE MULTIVARIÉE - BOXPLOT

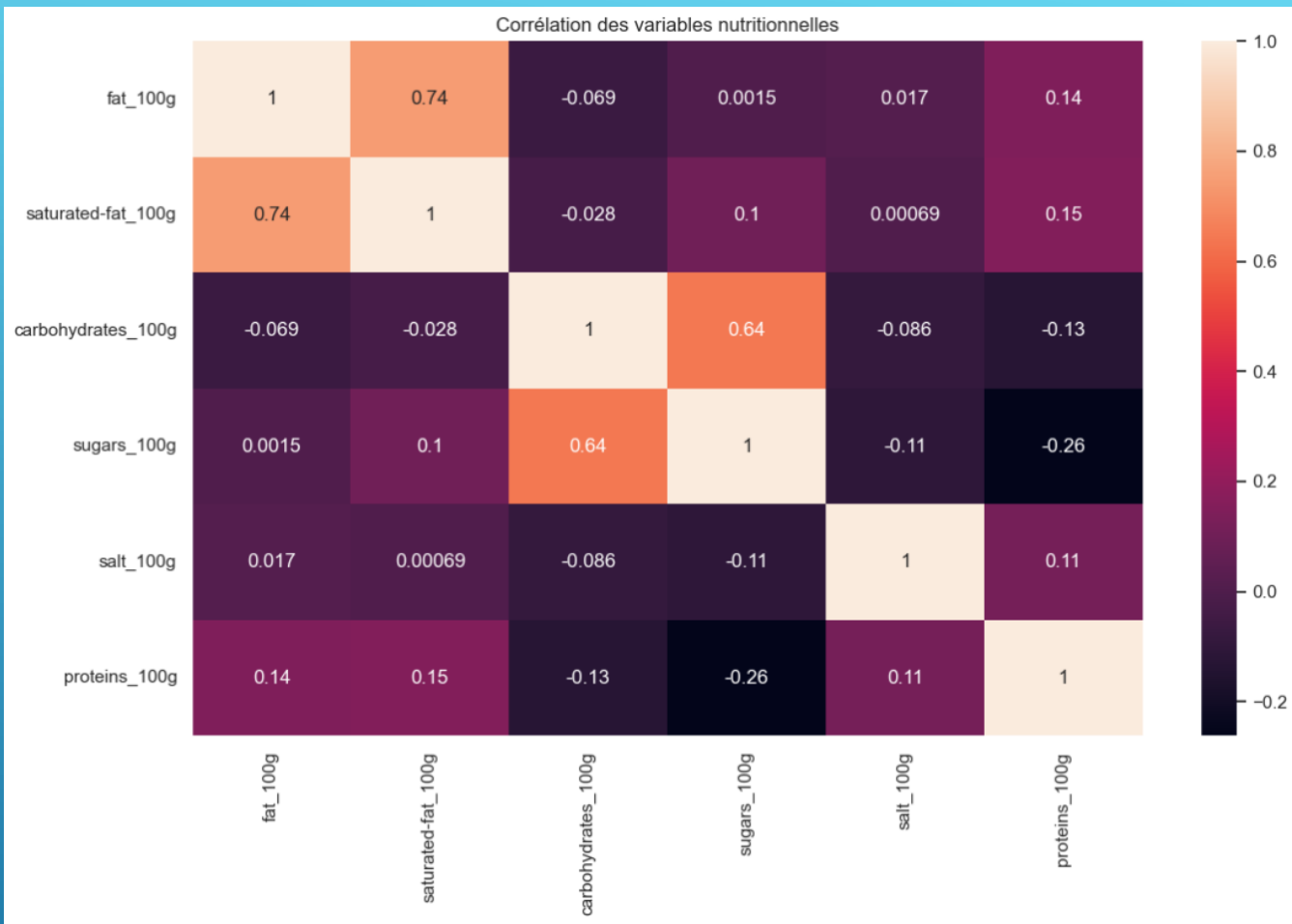


- ▶ On retrouve des différences en teneurs selon les catégories
- ▶ Présence notable dans les plats à base de viande
- ▶ C'est un nutriment important pour suivre une alimentation équilibrée

ANALYSE MULTIVARIÉE - BOXPLOT

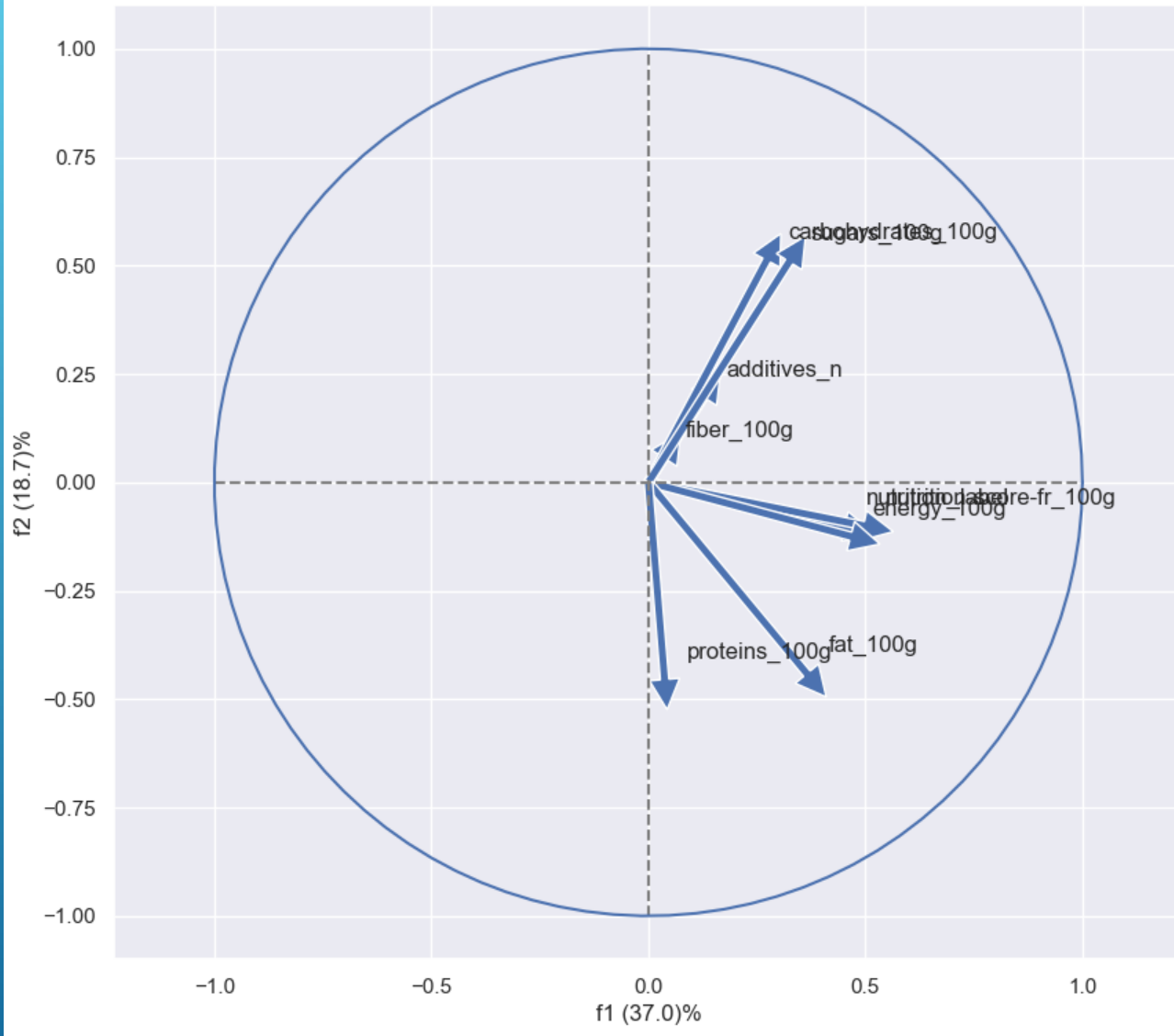


- On retrouve une haute teneur avec une grande variabilité dans les catégories « épicerie » et « produits à tartiner salés »
- Des catégories comme « soupes » et « plats préparés » ont des teneurs plus basses



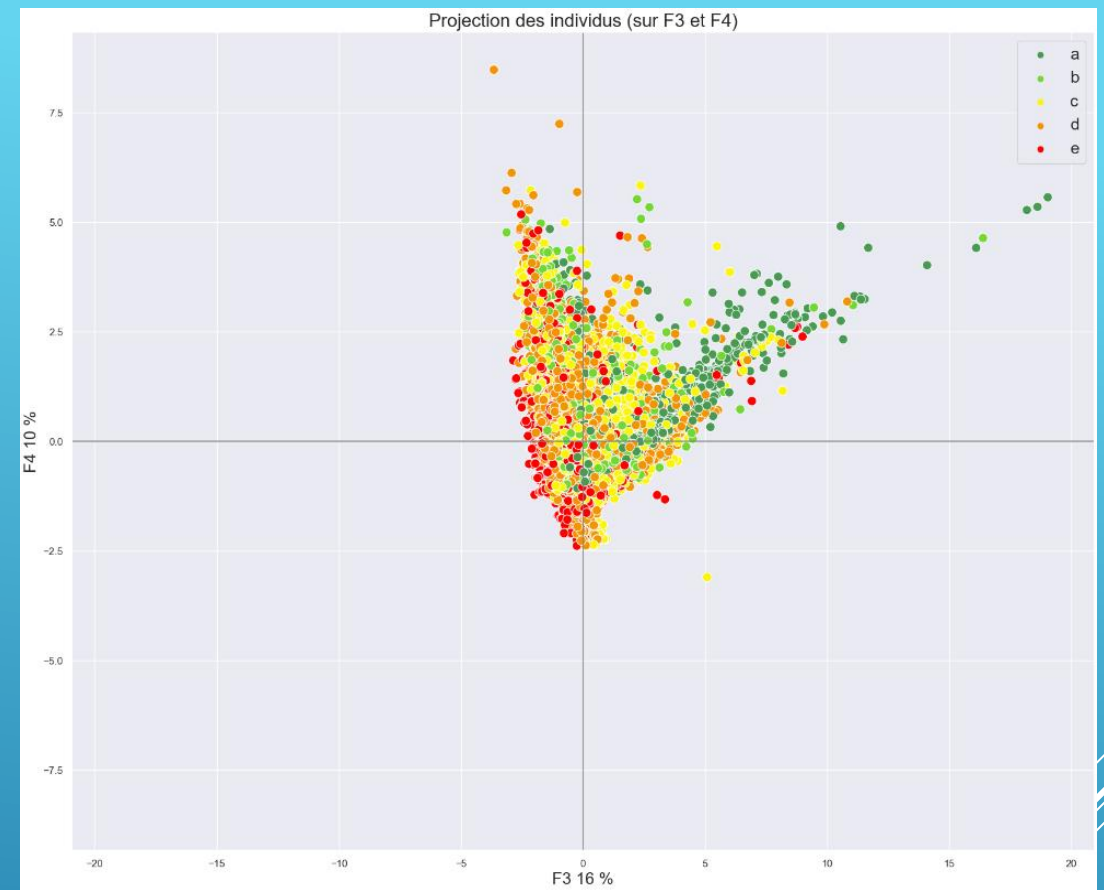
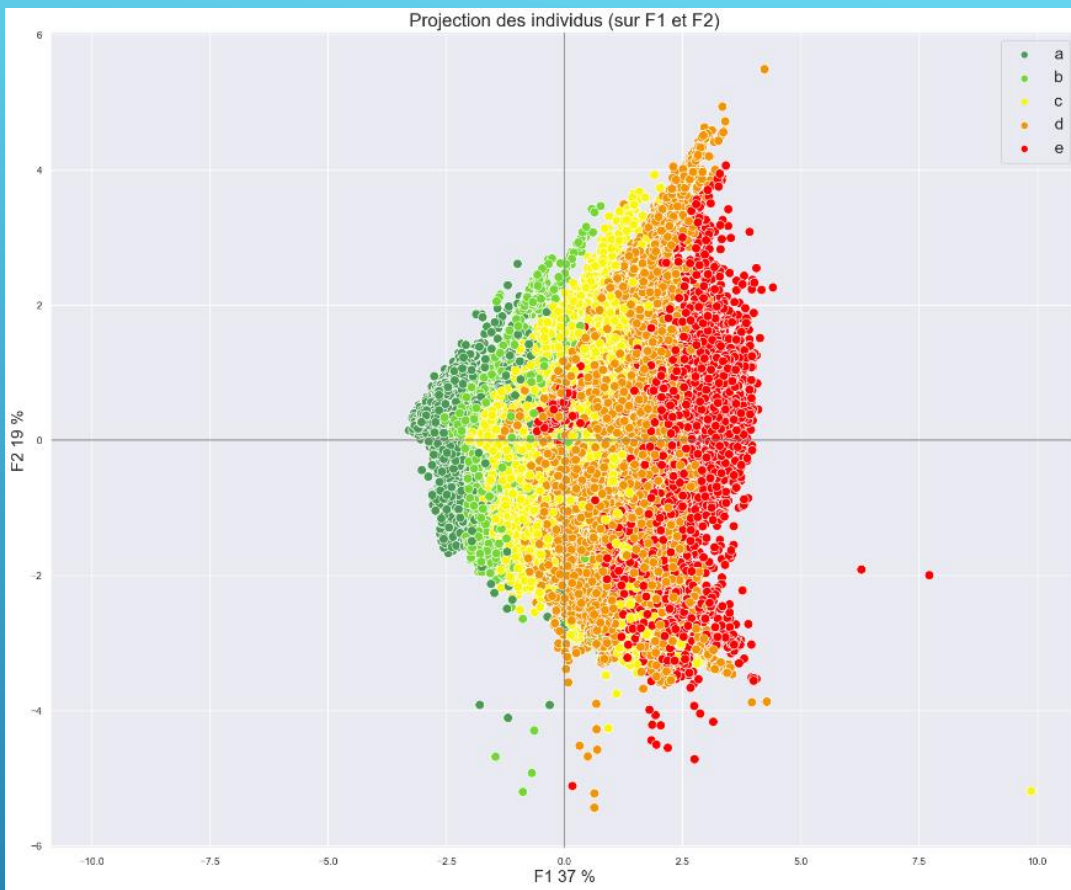
- La heatmap montre une corrélation importante entre « saturated fat » et « fat »
- Il y a aussi une corrélation entre « carbohydrates » et « sugars »
- On peut voir aussi une corrélation négative importante entre « proteins » et « sugars »

Cercle des corrélations (F1 et F2)



- ▶ « energy », « nutriscore_score », et « fat » sont fortement corrélées entre elles
- ▶ « proteins » et « carbohydrates » ont une corrélation négative

ACP – CERCLE DE CORRÉLATION



- Les projections montrent une séparation nette selon le Nutri-Score (F1 et F2)
- Les produits notés A/B ont des profils nutritionnels plus sains, avec moins de sucres/additifs
- Les additifs jouent un rôle crucial sur le Nutri-Score (F4)

ACP – PROJECTION DES INDIVIDUS

OLS Regression Results							
Dep. Variable:	additives_n		R-squared:		0.053		
Model:	OLS		Adj. R-squared:		0.052		
Method:	Least Squares		F-statistic:		566.1		
Date:	Thu, 16 May 2024		Prob (F-statistic):		0.00		
Time:	18:27:37		Log-Likelihood:		-95143.		
No. Observations:	40811		AIC:		1.903e+05		
Df Residuals:	40806		BIC:		1.903e+05		
Df Model:	4						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
Intercept	0.8539	0.029	29.157	0.000	0.797	0.911	
nutrition_grade_fr[T.b]	1.1006	0.044	24.971	0.000	1.014	1.187	
nutrition_grade_fr[T.c]	1.2447	0.040	31.510	0.000	1.167	1.322	
nutrition_grade_fr[T.d]	1.6156	0.038	43.075	0.000	1.542	1.689	
nutrition_grade_fr[T.e]	1.6570	0.041	40.679	0.000	1.577	1.737	
Omnibus:	17282.189	Durbin-Watson:		1.297			
Prob(Omnibus):	0.000	Jarque-Bera (JB):		92832.548			
Skew:	2.000	Prob(JB):		0.00			
Kurtosis:	9.212	Cond. No.		6.22			

- ▶ Le nombre d'additifs est significatif sur le résultat du Nutri-Score
- ▶ Plus un produit contient un nombre élevé d'additifs, plus ce sera proche de E

- ▶ Les résultats offrent une base pour des décisions stratégiques et opérationnelles dans le secteur agroalimentaire et médicale
- ▶ Diverses solutions peuvent être mises en place pour améliorer ce secteur dans le futur
- ▶ Il est tout de même conseillé de compléter cette base en consultant des experts dans le domaine pour plus de pertinence

CONCLUSION