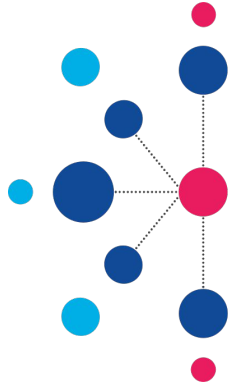


OPENCLASSROOMS



Santé
publique
France



Projet 3 : Concevez une application au service de la santé publique

Anissa TALEB Mai 2023



AGRAS

B O N O U M A U V A I S G R A S ?

Objectif :

⇒ distinguer les aliments contenant de bon gras du mauvais gras :

- Taux Acide Gras (gras saturé)
- Ratio Oméga 6/ Oméga 3
- Nutriscore

Sommair :

Partie 1 : Importation des librairies et de jeu de données

Partie 2 : Inspection de qualité de jeu de données

Partie 3 : Nettoyage de jeu de données

Partie 4 : Analyse exploratoire

Partie 5 : Présentation de l'idée d'application

Partie 1 : Importation de librairies et de jeu de données

Importation des librairies : Pandas
Numpy
Matplotlib
Seaborn
SciPy
sklearn

source de dataSet: données Open Food Facts

Partie 2 : Inspection de qualité de jeu de données

	code	url	creator	created_t	created_datetime	last_modified_t	last_modified_datetime	product_name	generic_
0	0000000003087	http://world-fr.openfoodfacts.org/produit/0000...	openfoodfacts-contributors	1474103866	2016-09-17T09:17:46Z	1474103893	2016-09-17T09:18:13Z	Farine de blé noir	

Nombre de lignes 320772

Nombre de colonnes : 162

Valeurs nulles ++++++

Pas conversion de type d'objets nécessaire

Nombre des valeurs manquants :

```
df.isnull().sum()
```

```
code                23
url                 23
creator             2
created_t           3
created_datetime     9
...
carbon-footprint_100g  320504
nutrition-score-fr_100g  99562
nutrition-score-uk_100g  99562
glycemic-index_100g    320772
water-hardness_100g    320772
Length: 162, dtype: int64
```

df.dtypes

```
code                object
url                 object
creator             object
created_t           object
created_datetime     object
...
carbon-footprint_100g  float64
nutrition-score-fr_100g  float64
nutrition-score-uk_100g  float64
glycemic-index_100g    float64
water-hardness_100g    float64
Length: 162, dtype: object
```

Partie 3 : Nettoyage des données

Sélection sur les colonnes :

- uniquement celles en rapport avec l'acide gras + Nutriscore FR
- Éliminer celles qui n'ont aucune valeur renseignée

Partie 3 : Nettoyage des données

Sélection sur les lignes:

- Suppression des produits sans nom: aucun intérêt métier
- uniquement individus ayant au moins 5 observations statistiques renseignées

Partie 3 : Nettoyage des données

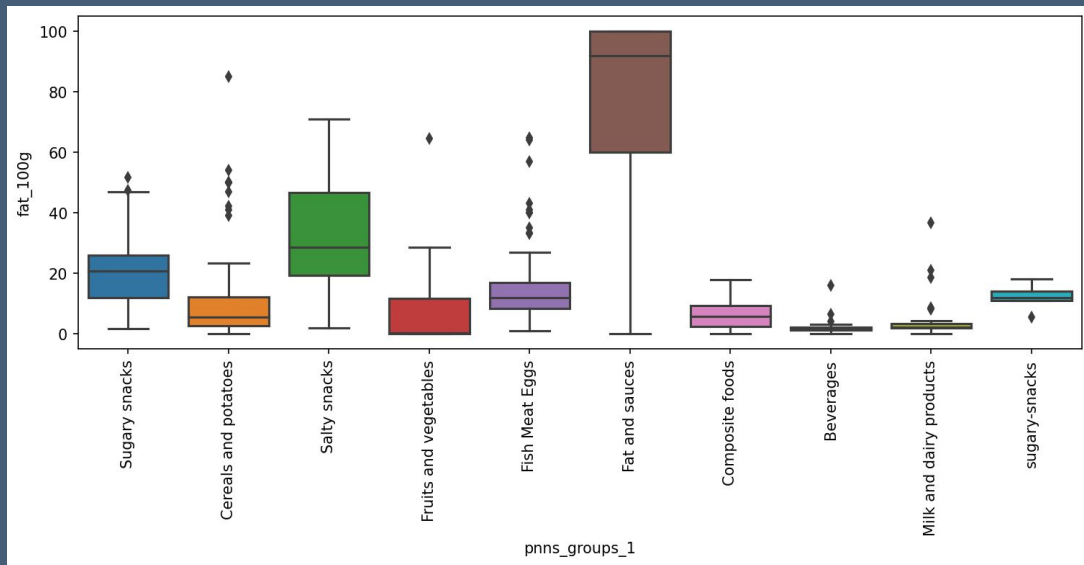
Gestion de valeurs aberrantes :

- Taux supérieur à 100 dans 100 grammes de nutriment
⇒ Transformer les valeurs supérieures à 100 en NaN.

Partie 3 : Nettoyage des données

Gestion de valeurs aberrantes :

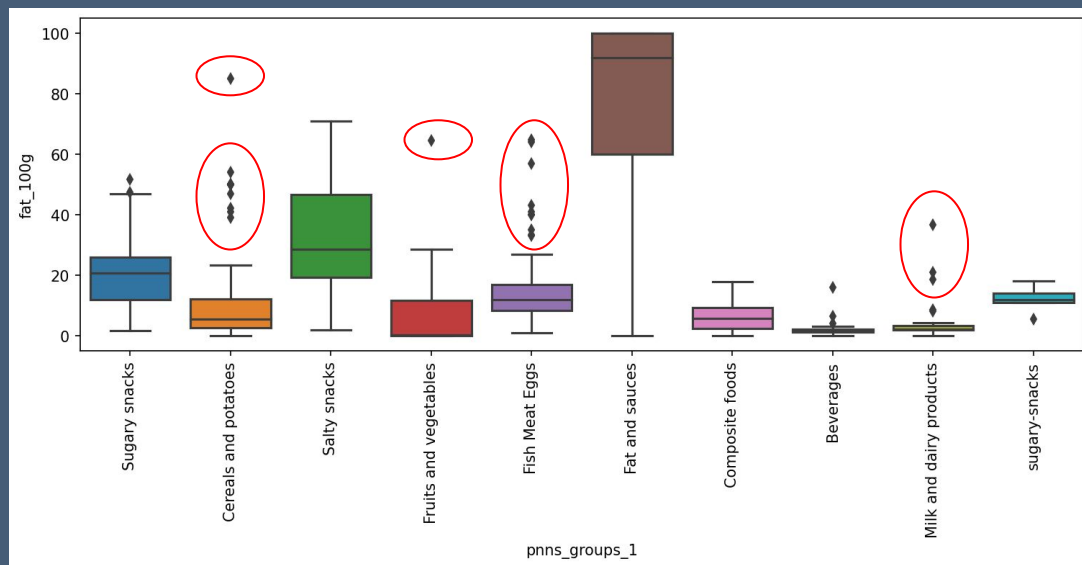
- Visualisation de répartition des valeurs par catégorie



Partie 3 : Nettoyage des données

Gestion de valeurs aberrantes :

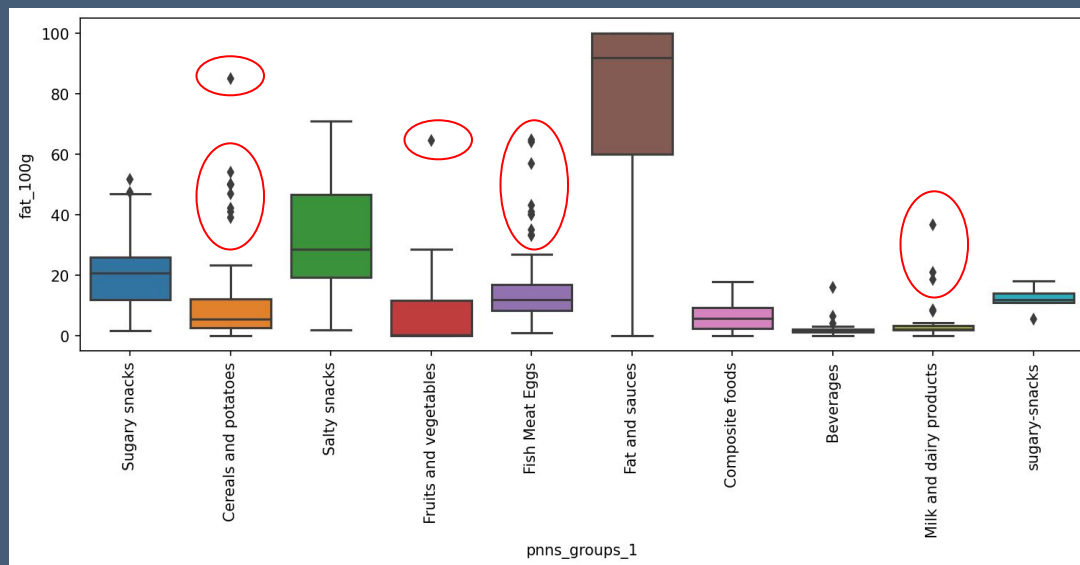
- Visualisation de répartition des valeurs par catégorie



Partie 3 : Nettoyage des données

Gestion de valeurs aberrantes :

- Visualisation de répartition des valeurs par catégorie



⇒ Correctes !
donc atypiques !

Partie 3 : Nettoyage des données

Gestion de valeurs nulles :

- Déterminer la valeur monounsaturated-fat_100g et polyunsaturated-fat_100g à partir de la relation :
$$\text{fat_100g} = \text{saturated-fat_100g} + \text{monounsaturated-fat_100g} + \text{polyunsaturated-fat_100g}$$
- Déterminer la valeur de Omega 9
$$\text{omega-9-fat_100g} = \text{monounsaturated-fat_100g}$$
- Remplacer les NaN par la moyenne de chaque catégorie

Partie 3 : Nettoyage des données

Gestion de valeurs nulles :

```
dfP3.isnull().sum()
```

nutrition_grade_fr	93
pnns_groups_1	0
product_name	4
fat_100g	0
saturated-fat_100g	0
monounsaturated-fat_100g	20
polyunsaturated-fat_100g	18
omega-3-fat_100g	456
omega-6-fat_100g	654
omega-9-fat_100g	756
trans-fat_100g	472
cholesterol_100g	381
dtype: int64	

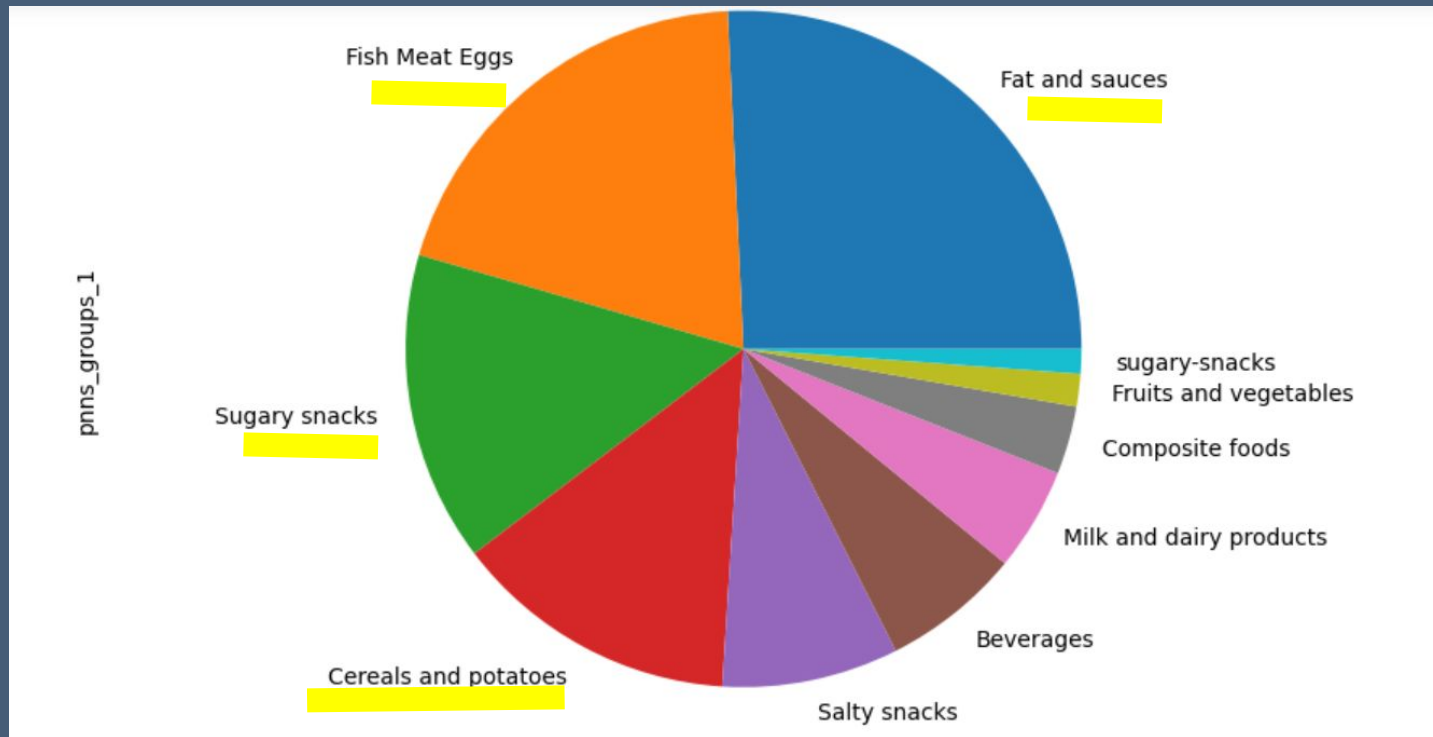


```
dfP3.isnull().sum()
```

nutrition_grade_fr	91
pnns_groups_1	0
product_name	0
fat_100g	0
saturated-fat_100g	0
monounsaturated-fat_100g	0
polyunsaturated-fat_100g	0
omega-3-fat_100g	12
omega-6-fat_100g	21
omega-9-fat_100g	0
trans-fat_100g	0
cholesterol_100g	0
dtype: int64	

Partie 4 : Analyse exploratoire

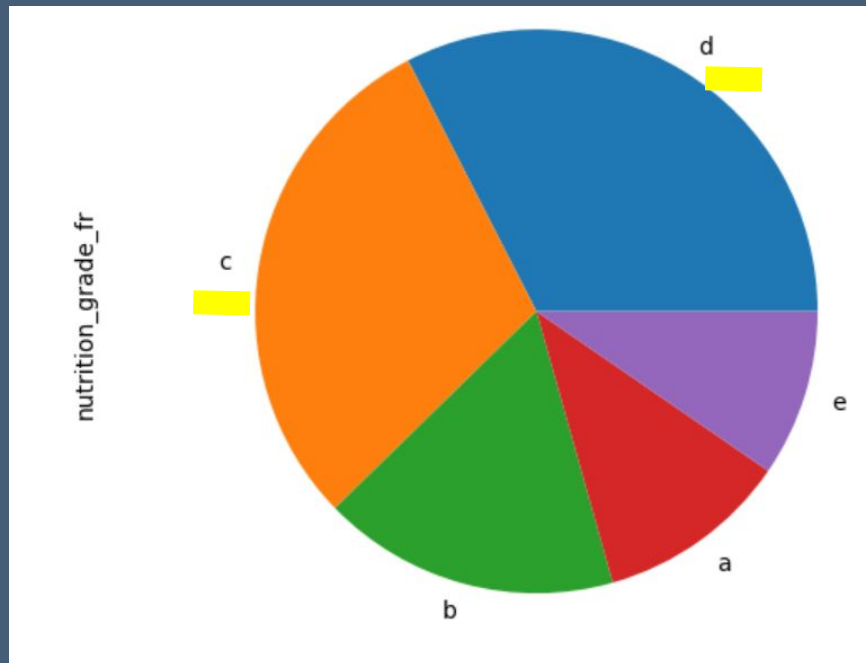
Analyse univariée :



4 catégories de 10 $\Rightarrow \frac{3}{4}$ de total

Partie 4 : Analyse exploratoire

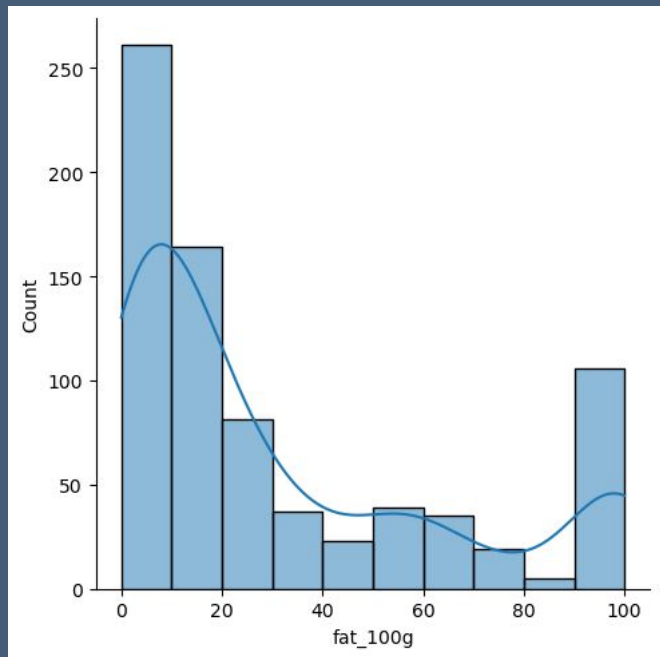
Analyse univariée :



les catégories c et d sont celles les mieux présentés

Partie 4 : Analyse exploratoire

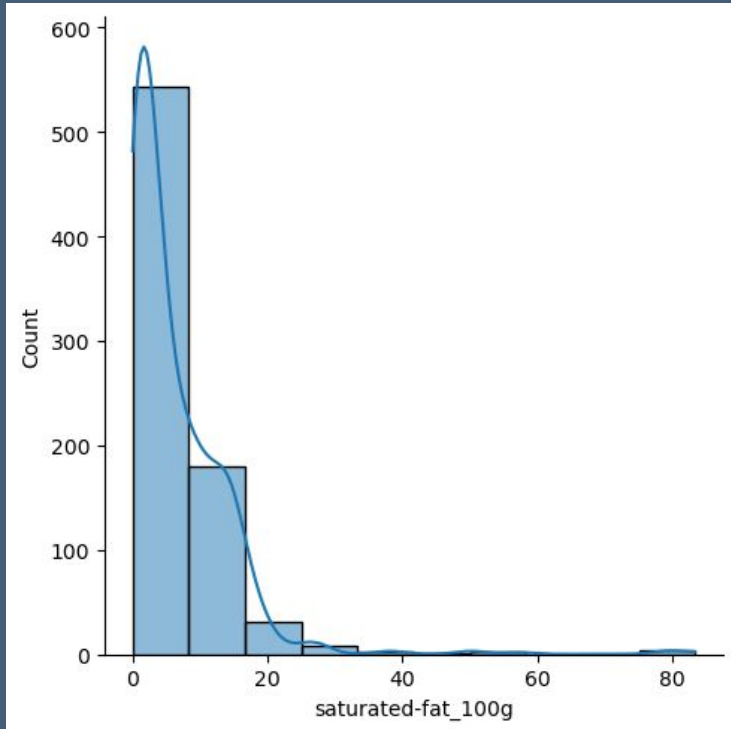
Analyse univariée :



=> +++produits avec taux élevé d'acide gras (100%)

Partie 4 : Analyse exploratoire

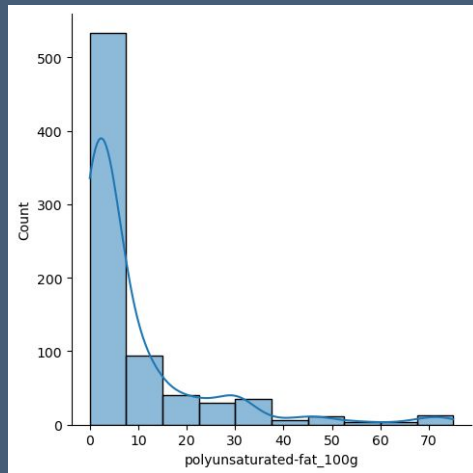
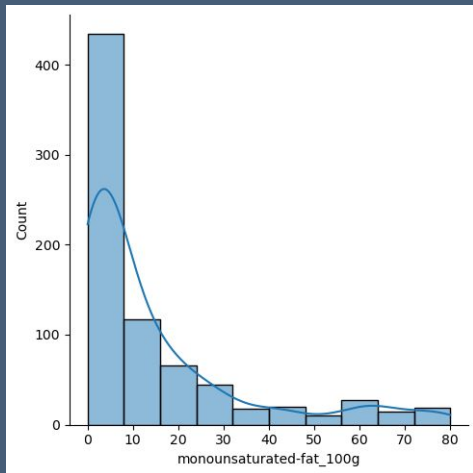
Analyse univariée :



=> ++++produits avec taux très bas d'acide gras saturé

Partie 4 : Analyse exploratoire

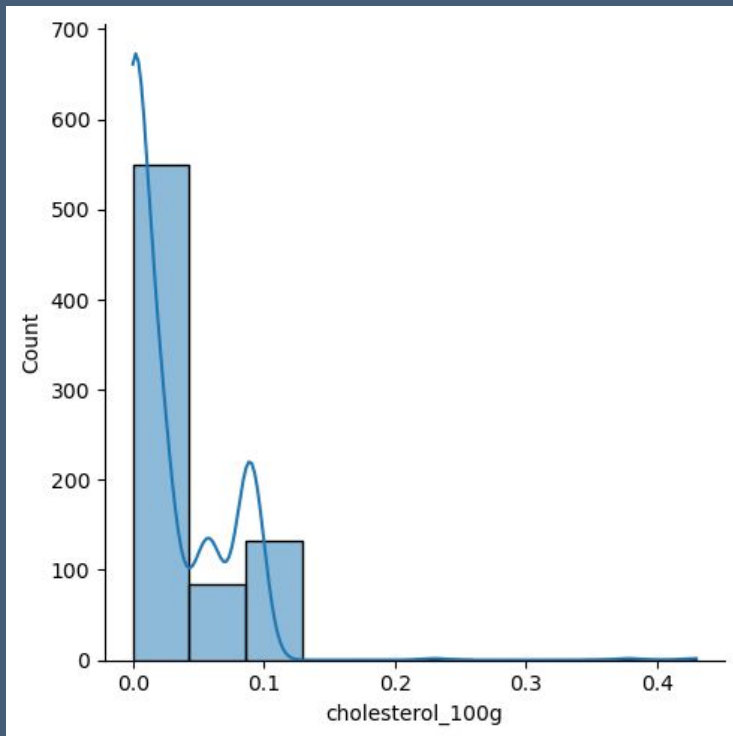
Analyse univariée :



=> ++++produits avec taux bas
d'acide gras insaturé

Partie 4 : Analyse exploratoire

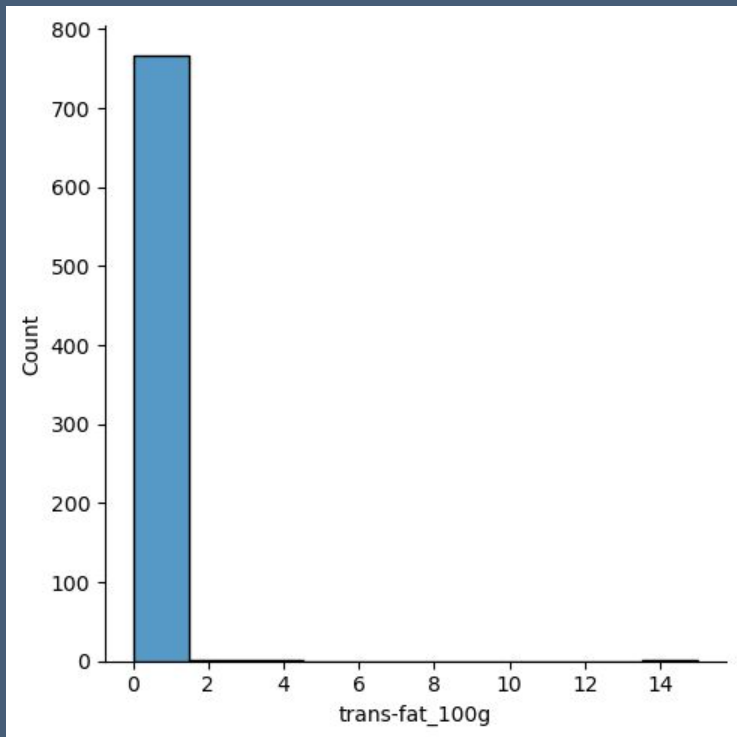
Analyse univariée :



=> ++++produits avec taux bas de cholestérol

Partie 4 : Analyse exploratoire

Analyse univariée :



=> ++++produits avec taux bas
d'acide gras trans

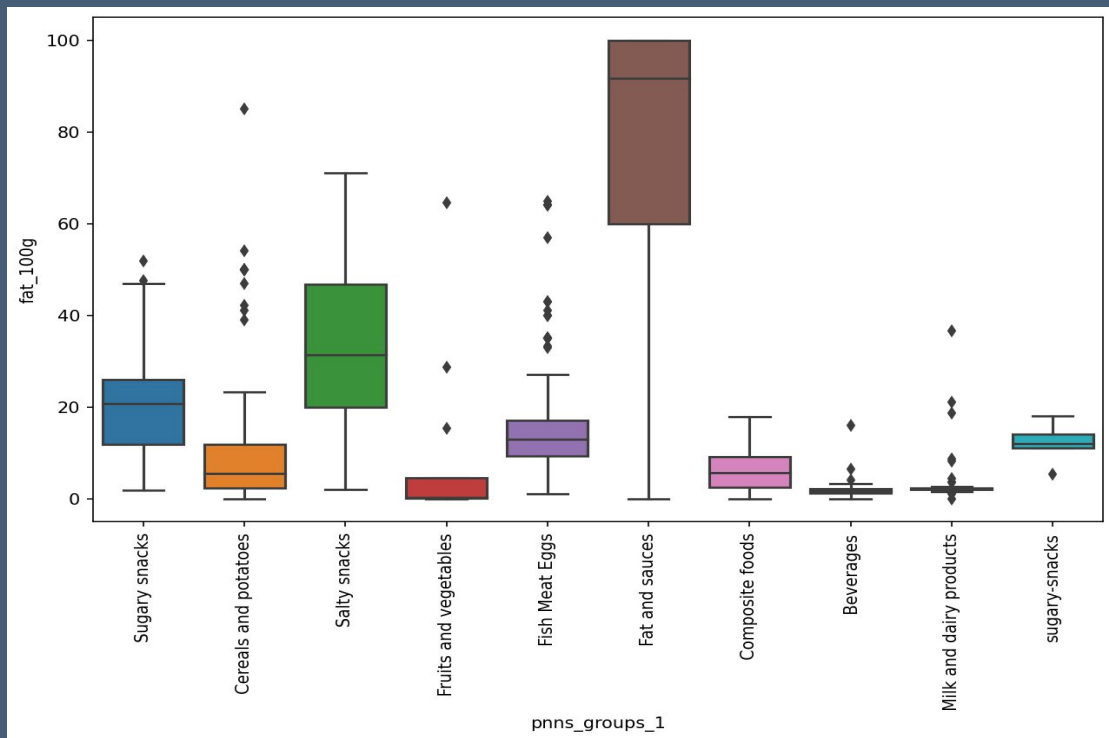
Partie 4 : Analyse exploratoire

Analyse univariée :

⇒ La distribution des variables
quantitatives ne suit pas une loi
normal ⇒ Pas de test ANOVA

Partie 4 : Analyse exploratoire

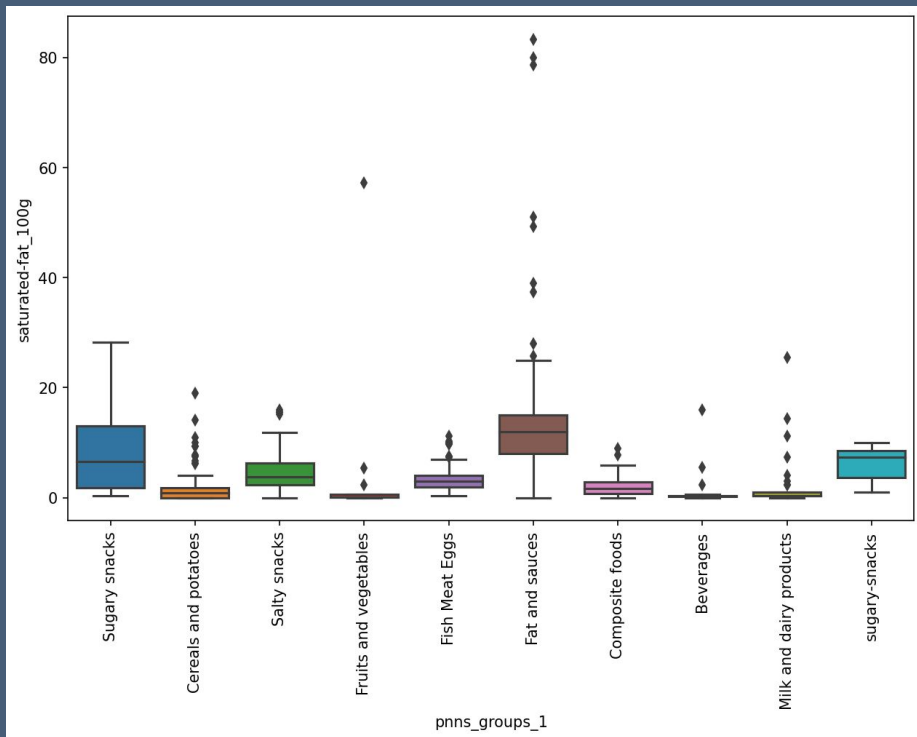
Analyse bivariable:



=> Les mesures de tendance centrale de taux d'acide gras par catégorie ne sont pas stables d'une catégorie à une autre.

Partie 4 : Analyse exploratoire

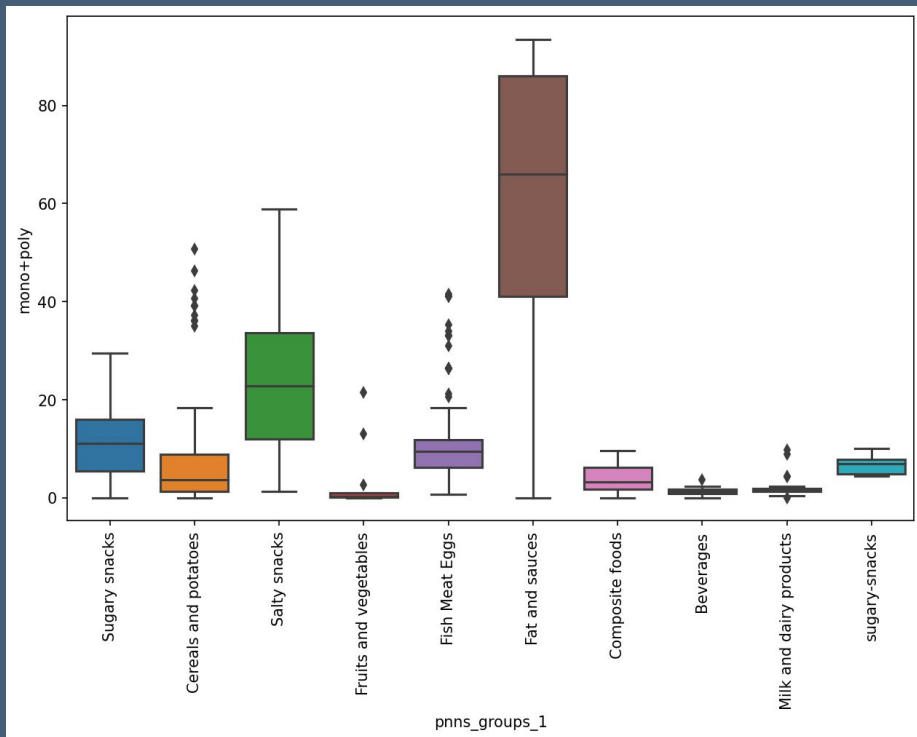
Analyse bivariable:



=> homogènes malgré les points atypiques présents dans certaines catégories spécifiques.

Partie 4 : Analyse exploratoire

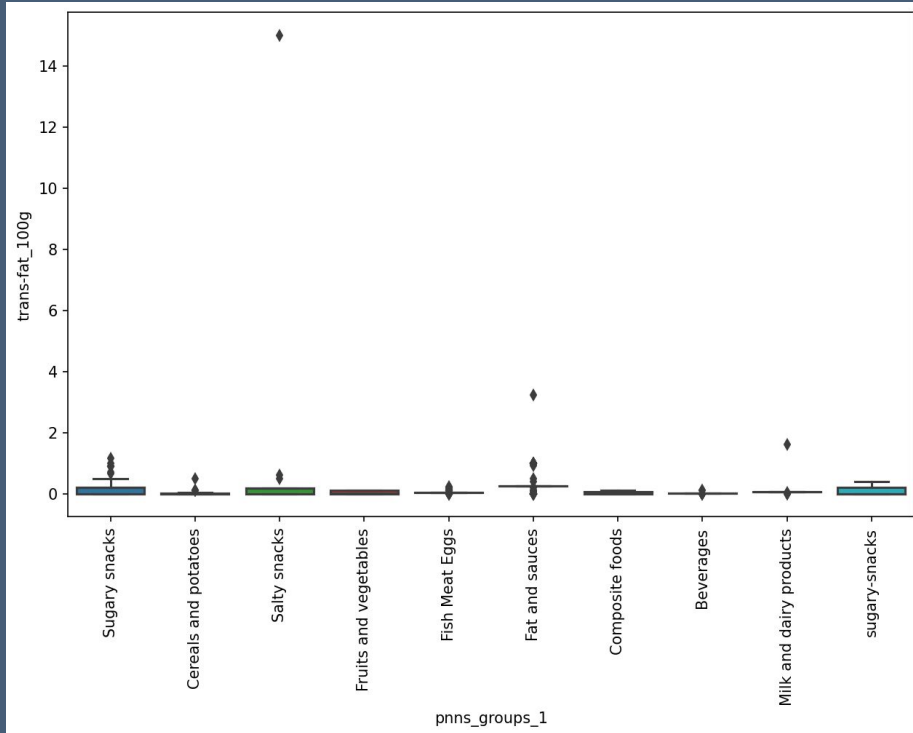
Analyse bivariable:



=> +++ 'Fat and sauce' et 'Salty snacks'

Partie 4 : Analyse exploratoire

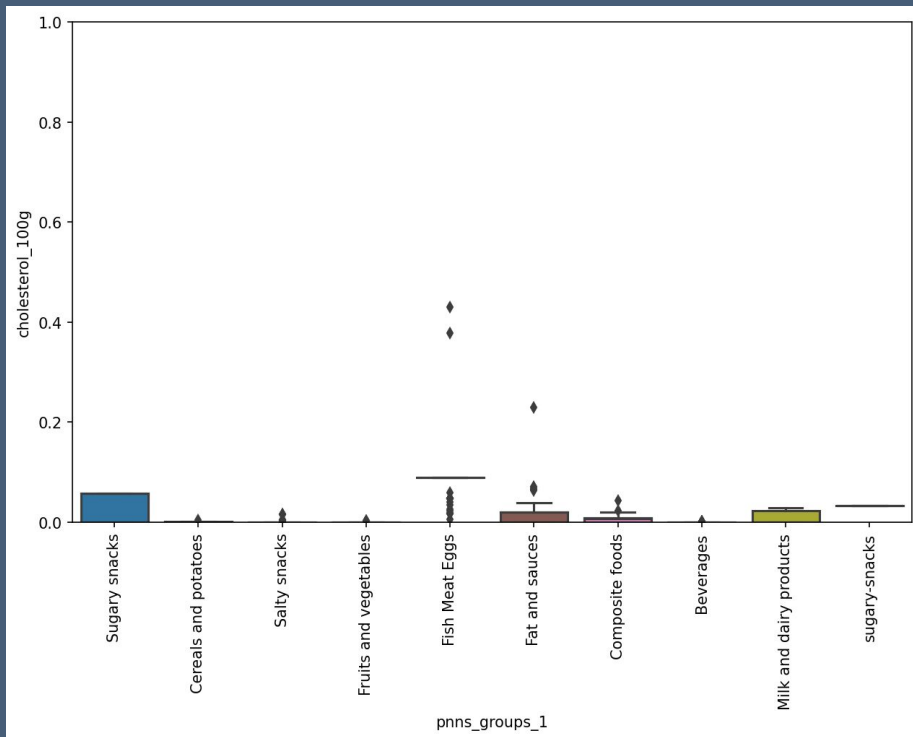
Analyse bivariée:



=> +++ 'Fat and sauce' et 'Salty snacks'

Partie 4 : Analyse exploratoire

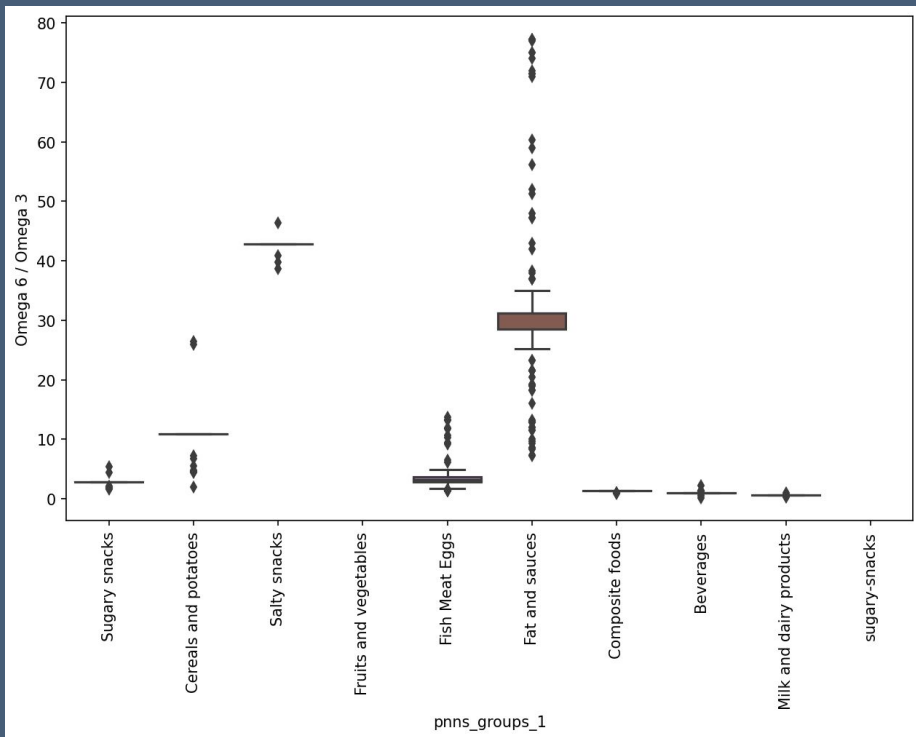
Analyse bivariable:



=> le cholestérol == aliments gras

Partie 4 : Analyse exploratoire

Analyse bivariable:



=> 'fat and sauce' et 'salty snacks' beaucoup de variation

+ ratio très élevé !

Partie 4 : Analyse exploratoire

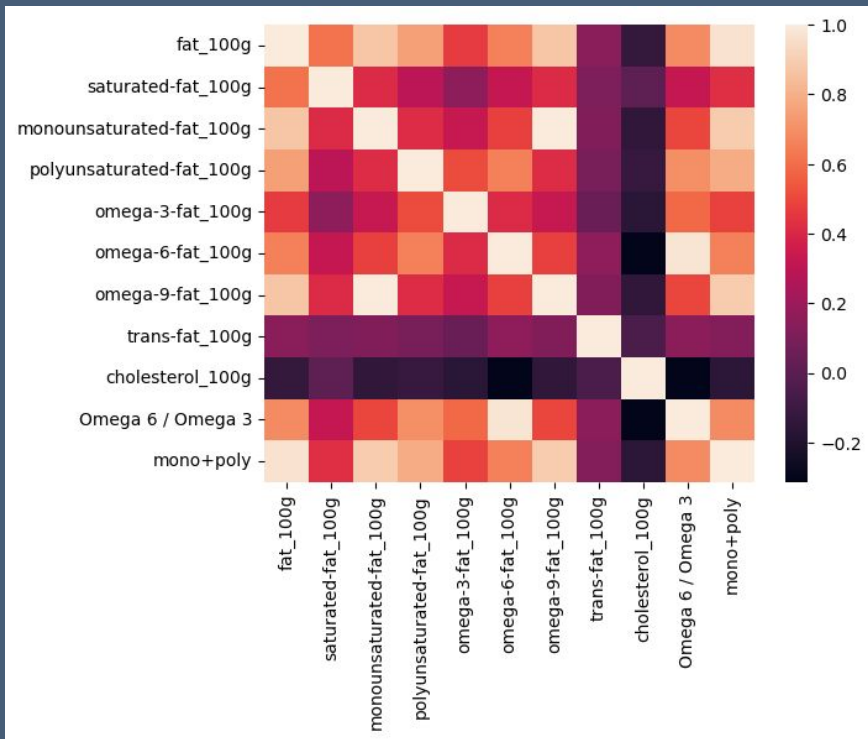
Analyse bivariée:

⇒ 'fat and sauce' et 'salty snacks' sont celles qui contiennent le plus d'AG saturés et d'AG insaturés

⇒ + ratio oméga 6/oméga 3 est très élevé !

Partie 4 : Analyse exploratoire

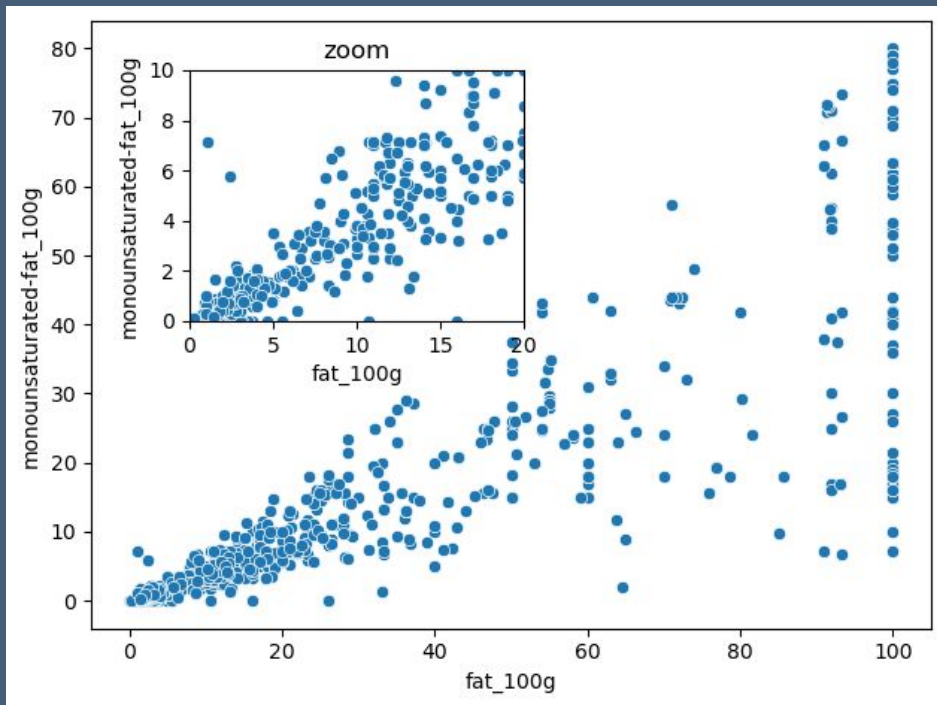
Carte thermique de corrélation :



=> pas de corrélation détectée entre le cholestérol/AG Trans et les autres indicateurs.

Partie 4 : Analyse exploratoire

Analyse bivariable:

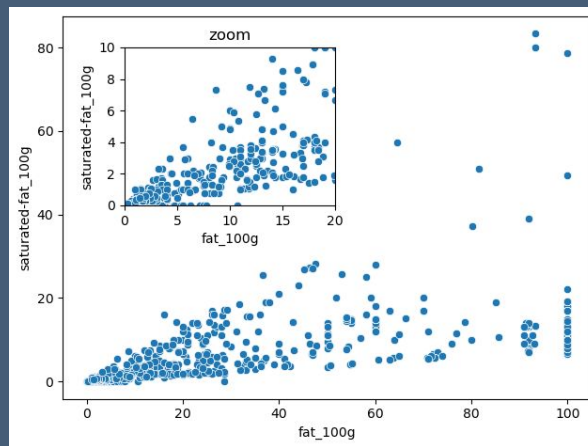
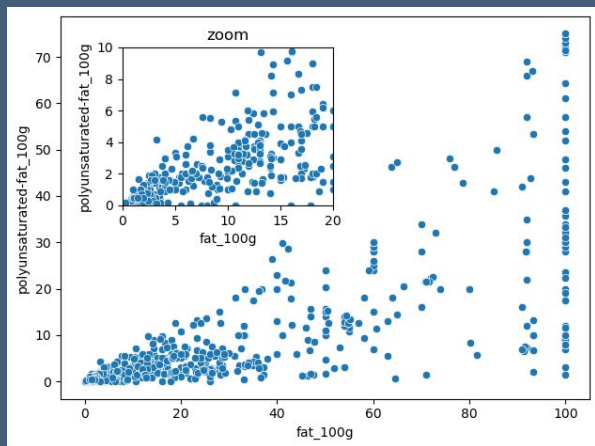


=> Corrélation mise en évidence notamment pour les petites valeurs de fat_100g

=> Cette relation est perturbée au niveau des valeurs atypiques!

Partie 4 : Analyse exploratoire

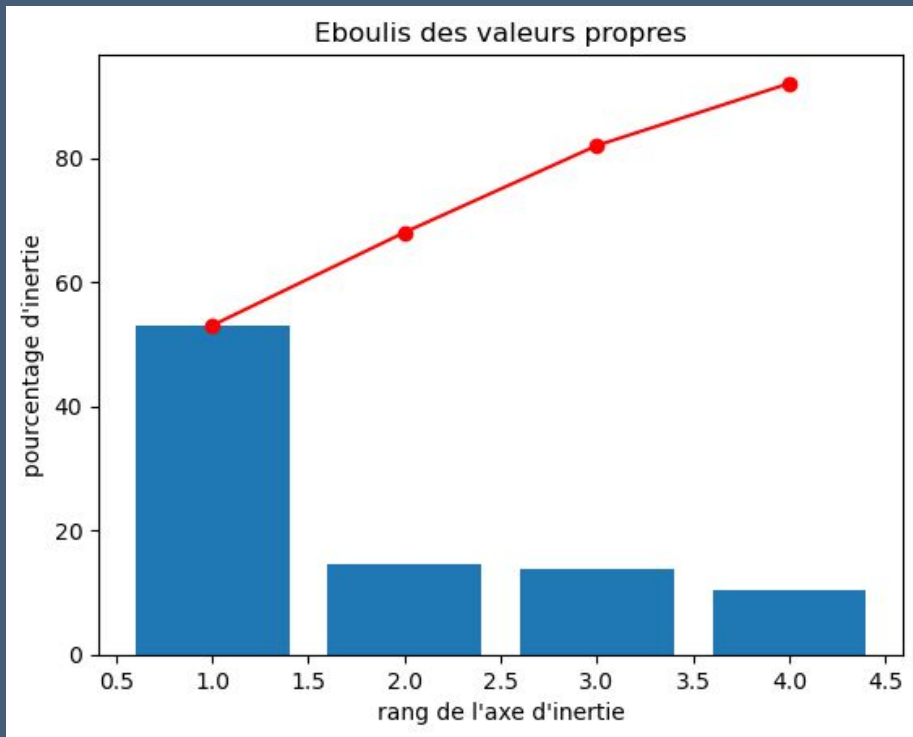
Analyse bivariable:



=> Corrélation perturbée au niveau des valeurs atypiques!

Partie 4 : Analyse exploratoire

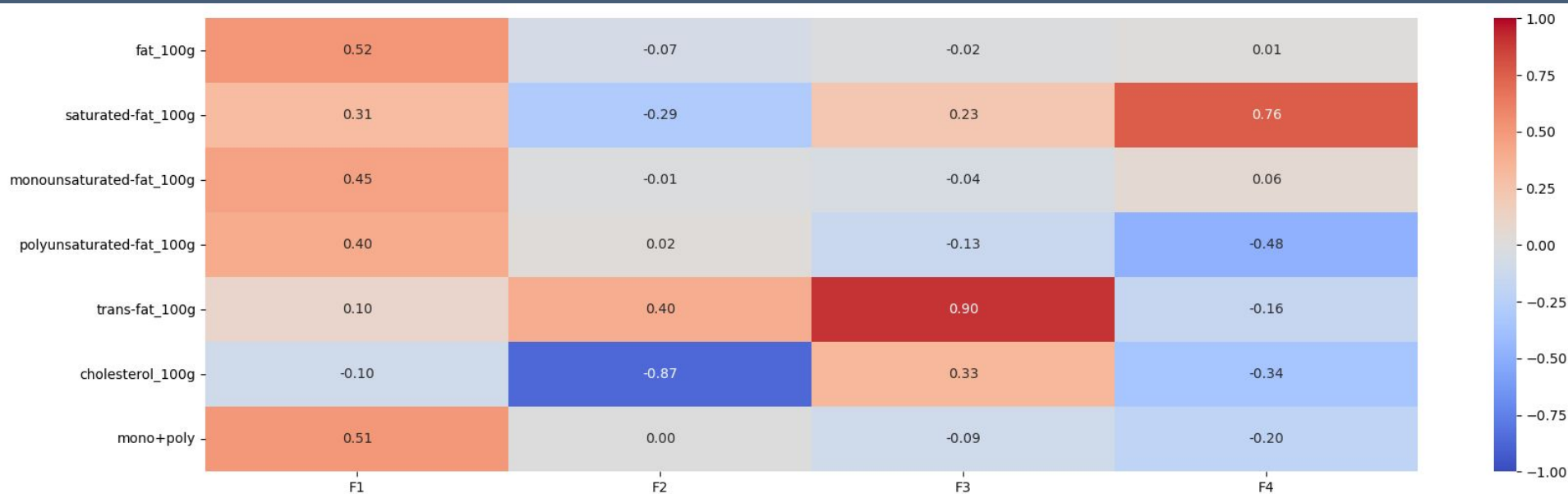
Analyse en composantes principales (ACP):



=> Le premier plan factoriel couvre environ 70% de variance

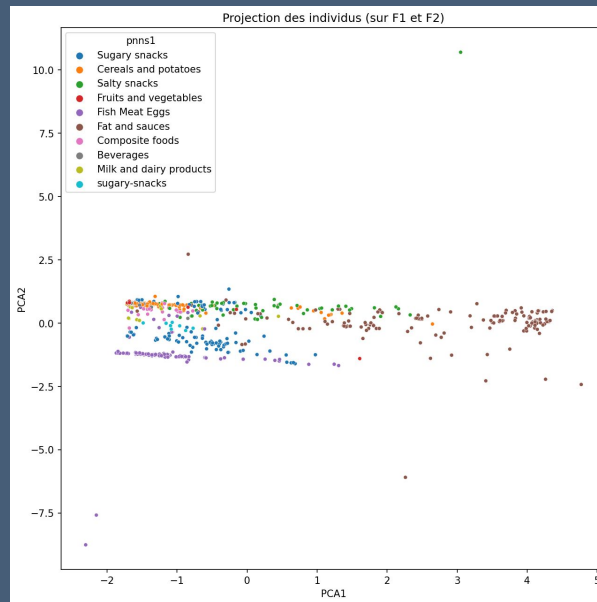
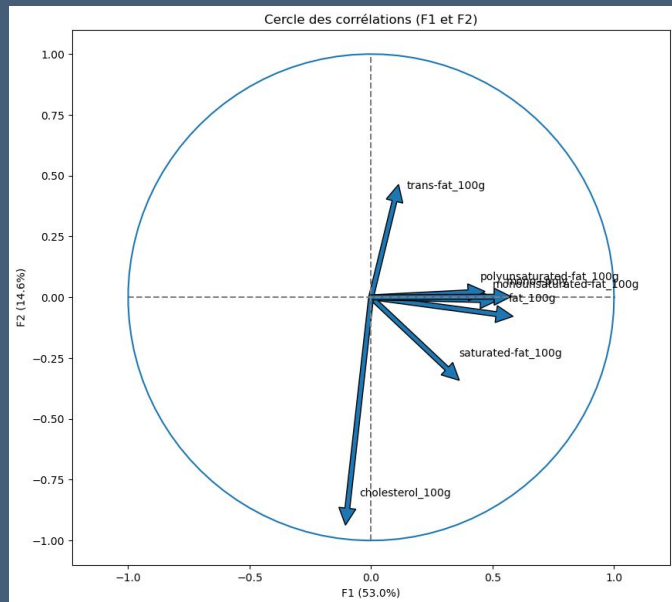
Partie 4 : Analyse exploratoire

Analyse en composantes principales (ACP):



Partie 4 : Analyse exploratoire

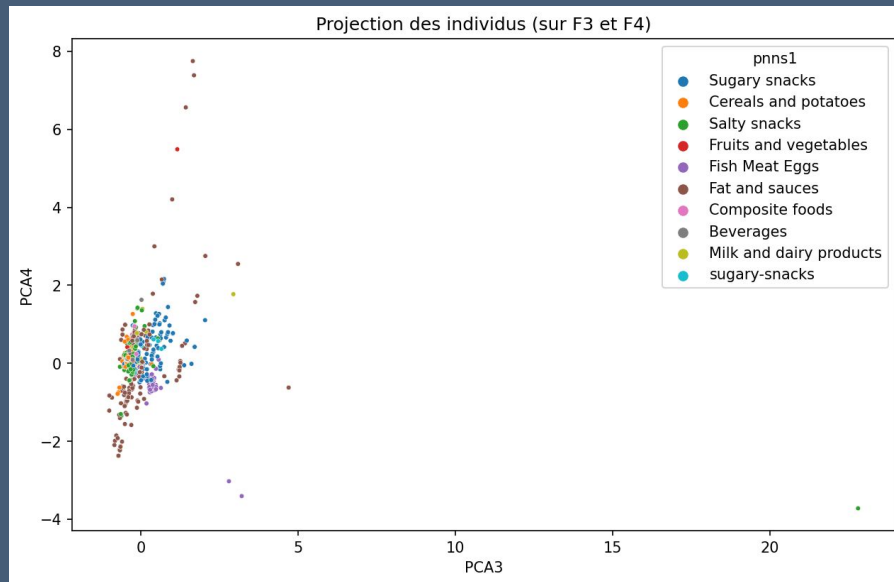
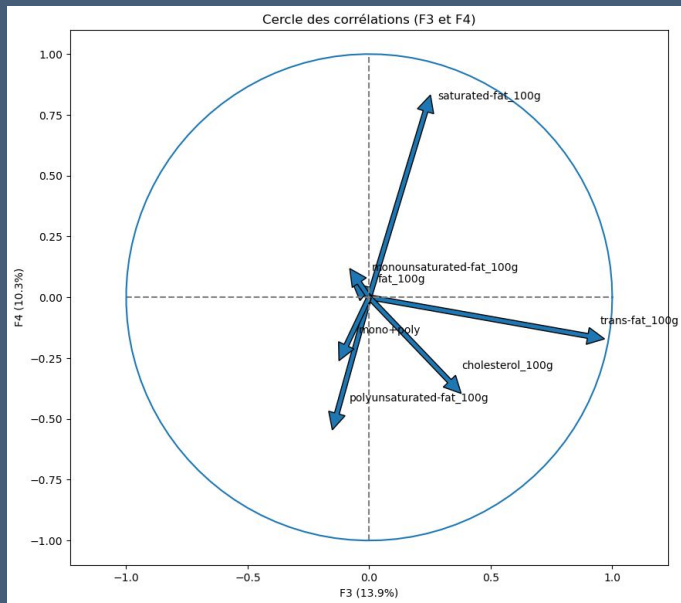
Analyse en composantes principales (ACP):



L'axe F1 mesure le taux d'acide gras en général, et F2 le taux de cholestérol et acide trans

Partie 4 : Analyse exploratoire

Analyse en composantes principales (ACP):



L'axe F3 reflète le taux d'AG trans, et F4 le taux d'AG saturé. Les deux marqueurs du mauvais gras.

Partie 4 : Analyse exploratoire

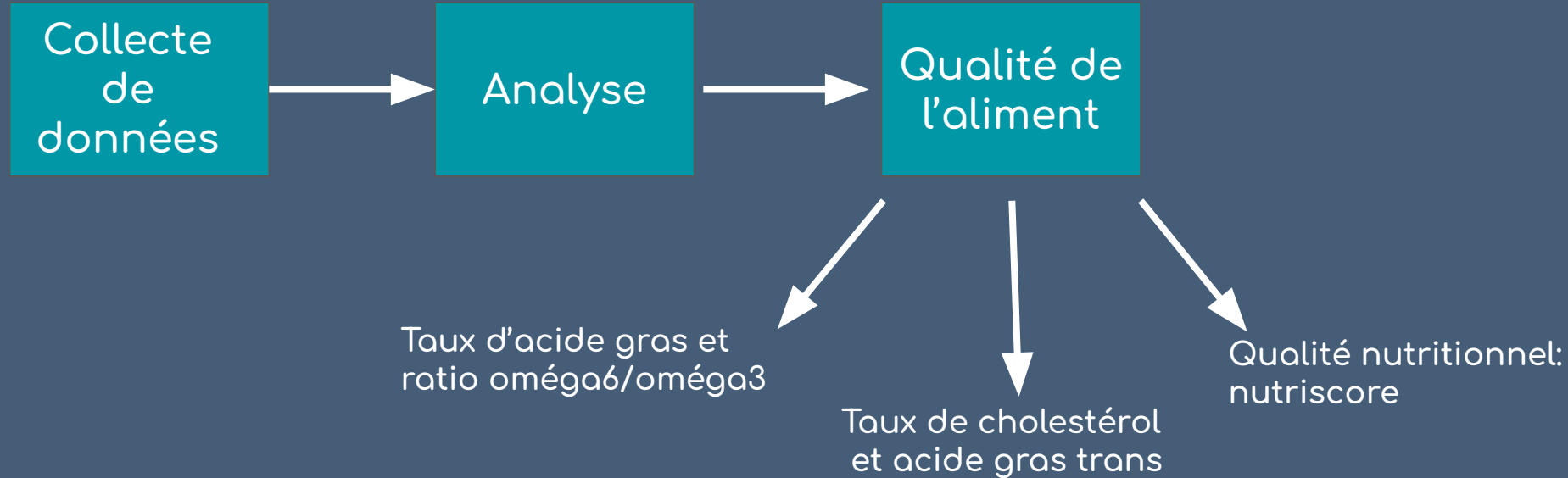
Analyse en composantes principales (ACP):

⇒ 'fat and sauce' et 'salty snacks' sont celles qui contiennent le plus de variance en terme de gras
+ transfat/cholestérol

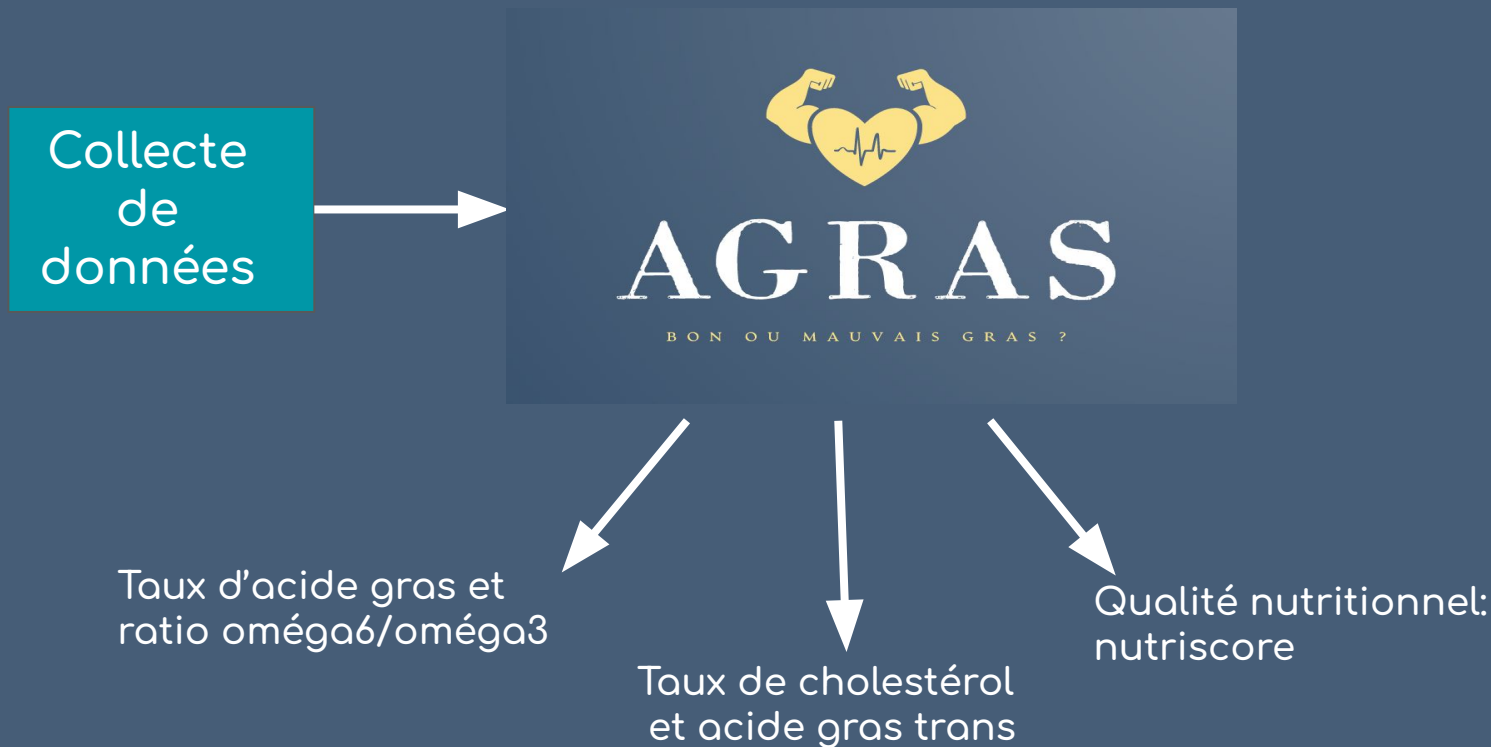
⇒ essentiellement du mauvais gras

Partie 5 : Présentation de l'idée d'application

Faisabilité de projet :



Partie 5 : Présentation de l'idée d'application

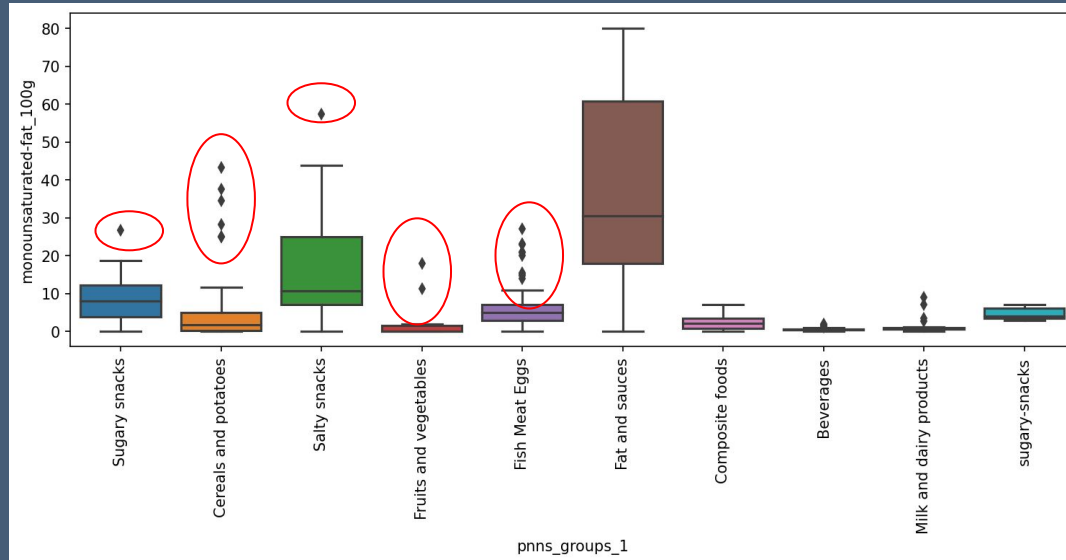


Merci pour votre attention :)

Partie 3 : Nettoyage des données

Gestion de valeurs aberrantes :

- Visualisation de répartition des valeurs par catégorie

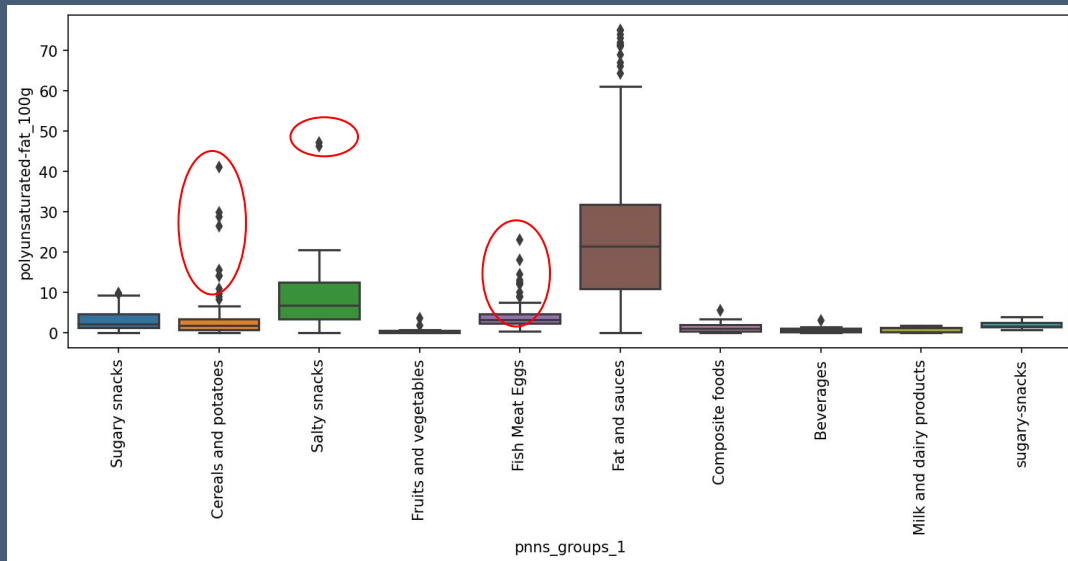


⇒ Correctes !

Partie 3 : Nettoyage des données

Gestion de valeurs aberrantes :

- Visualisation de répartition des valeurs par catégorie



⇒ Correctes !

