

Concevez une application au service de la santé publique

Noura Romari – Parcours Ingénieur Machine Learning depuis le 04/03/22
Soutenance Projet n°2 du 05/05/22

Mentor : Hamza Tajmouati
Evaluateur : Nyemo Koumadi



Sommaire

- **Introduction**
 - Le contexte
 - Les données Open Food Facts
 - Ma proposition d'application
- **Nettoyage des données**
 - Stratégie
 - Résultats
- **Analyse**
 - Stratégie
 - Univariées
 - Bivariées
 - Multifactorielle
- **Ma proposition d'application**
- **Conclusion**
 - Bilan analytique
 - Faisabilité de l'application
 - Axes d'amélioration

Introduction

- **Objectif**

- Répondre à l'appel à projet de **Santé publique France**
- Proposer une application innovante en lien avec la Santé
- Utiliser les données d'**Open Food Facts**



- **Open Food Facts**

- Projet collaboratif en libre accès
- Réunir le maximum d'informations sur les produits alimentaires par scan de leur gencode
- Le consommateur est acteur : utilisation d'une application



Introduction

- **Les données**

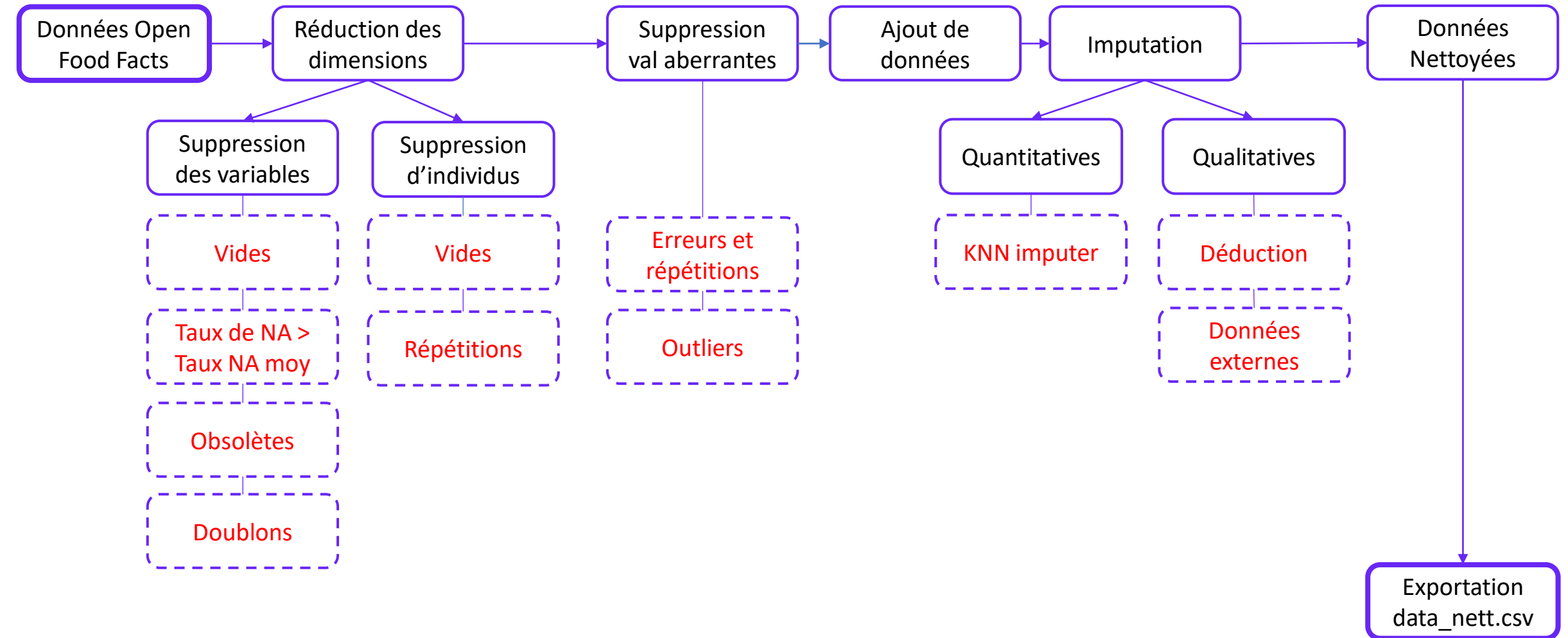
- Informations très variées :
 - Ingrédients et allergènes
 - Valeurs nutritionnels
 - Parcours de vie du produit : du lieu de fabrication au lieu de consommation
 - Marque et packaging
 - Nutiscore
- Dimension du jeux de données : (320772, 162)

- **Ma proposition d'application**

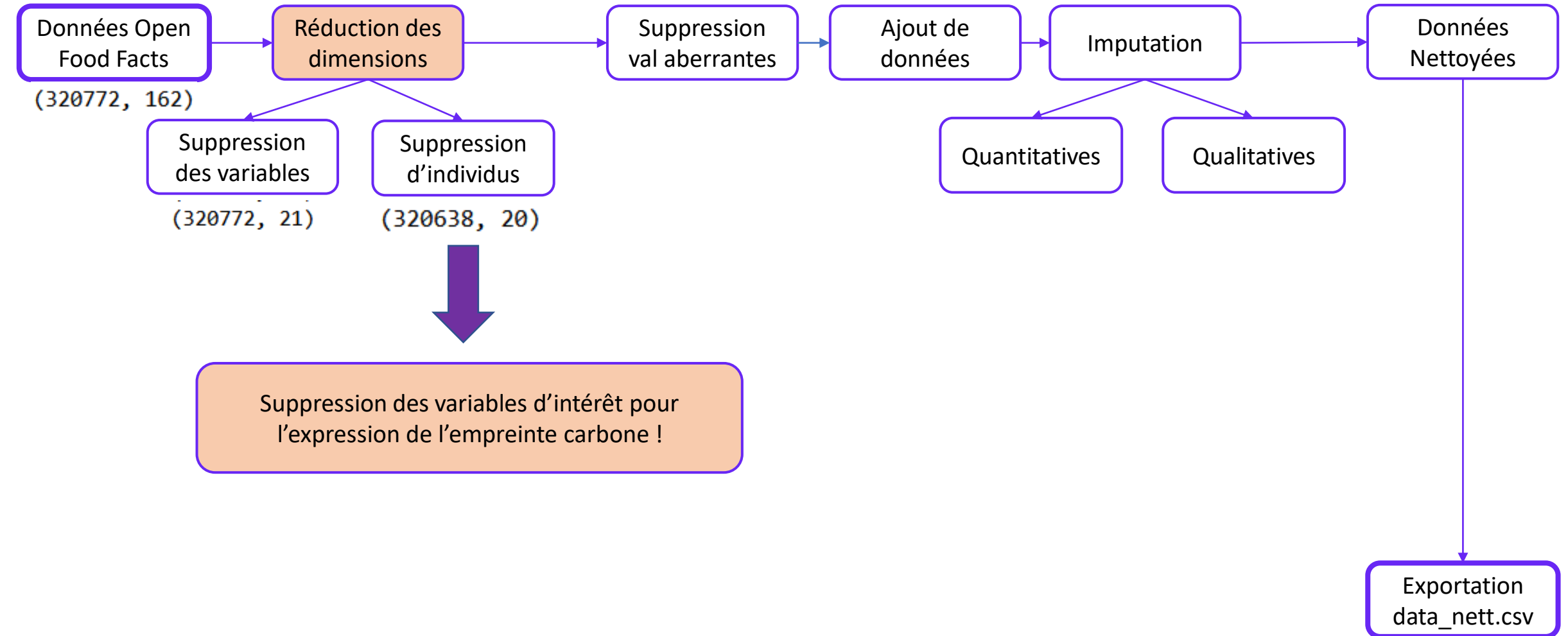
- Quelles sont les préoccupations du « consom'acteur » d'aujourd'hui ?
 - La sécurité alimentaire
 - La santé par l'alimentation
 - La santé de notre planète
- Proposer une application qui aide le consommateur à se maintenir en bonne santé, tout en agissant pour la planète → **Allier Nutri-score et empreinte carbone.**

Nettoyage des données

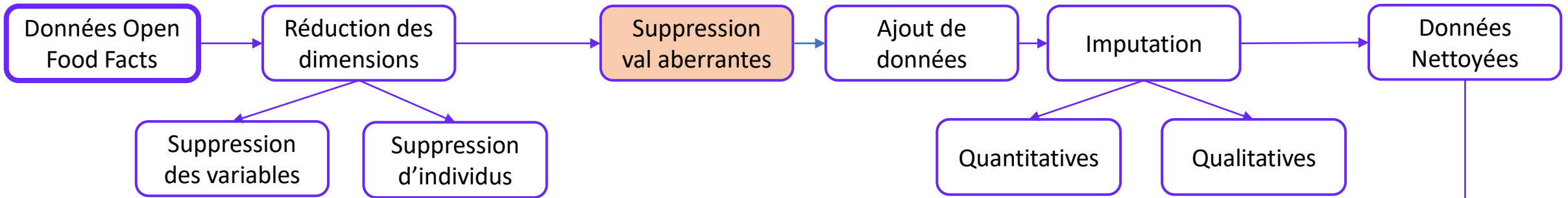
Stratégie de nettoyage



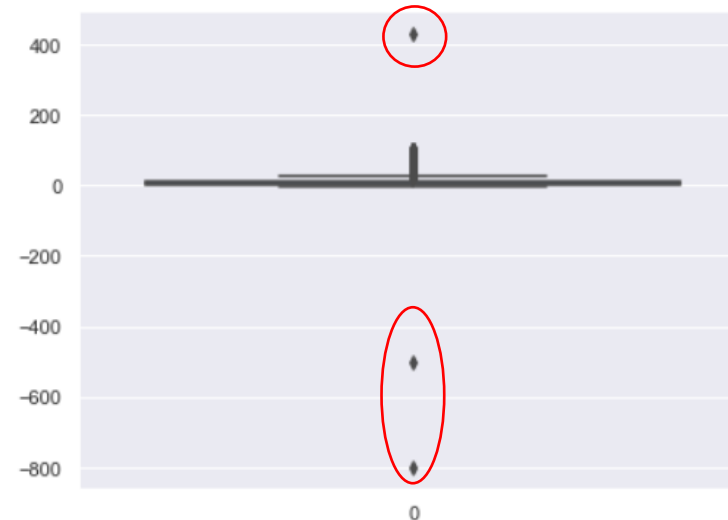
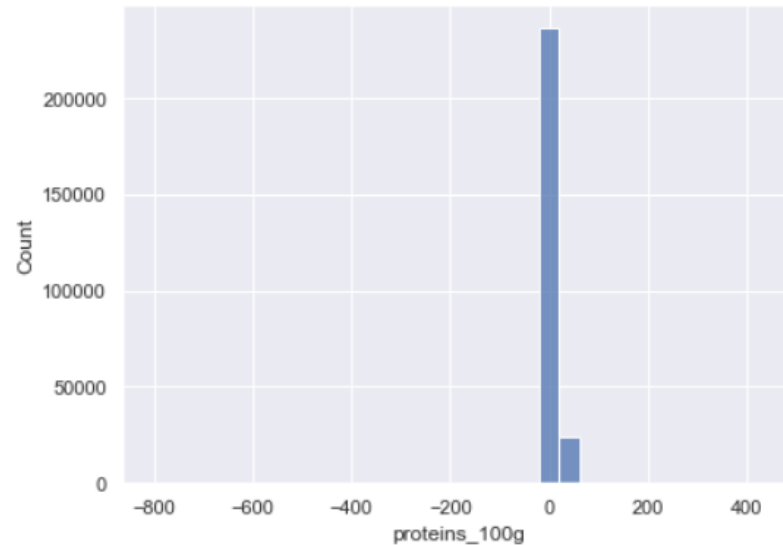
Résultats du nettoyage



Résultats du nettoyage

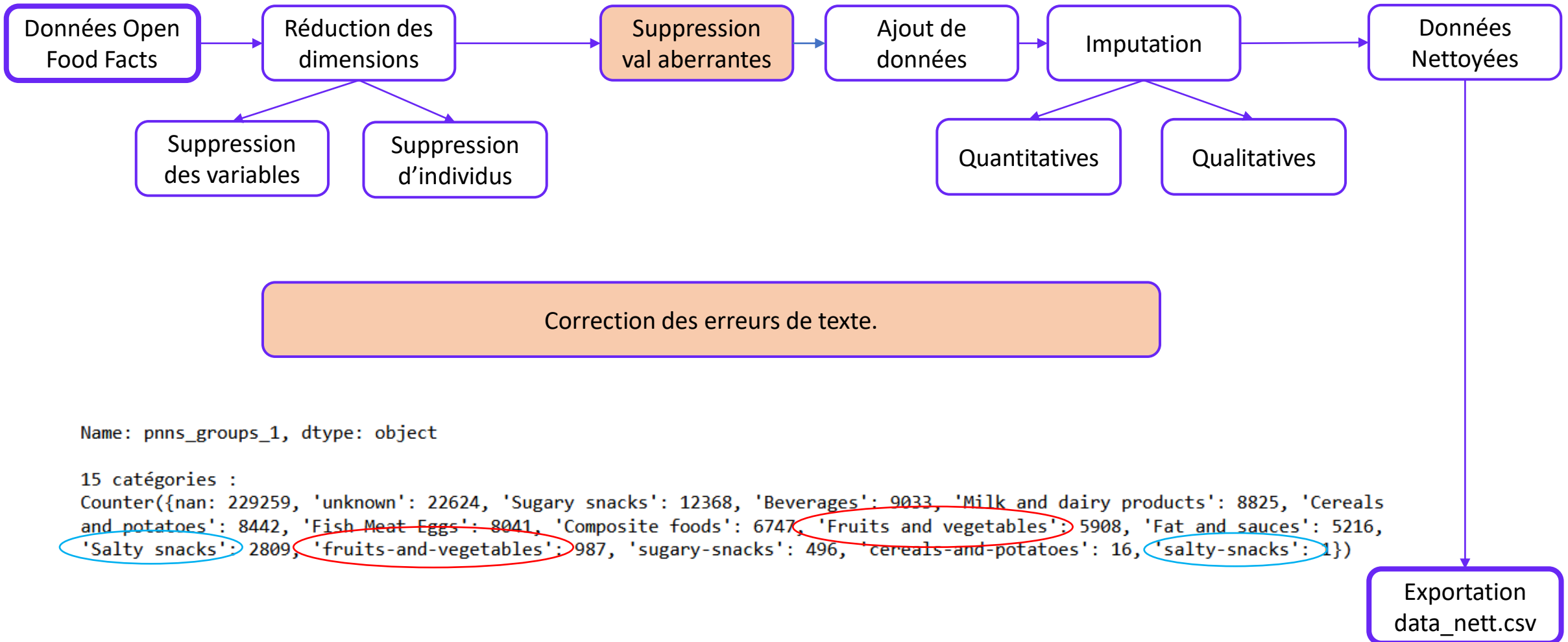


Suppression des valeurs négatives et des valeurs supérieures à 100 pour toutes les données nutritionnelles exprimées en g/100g

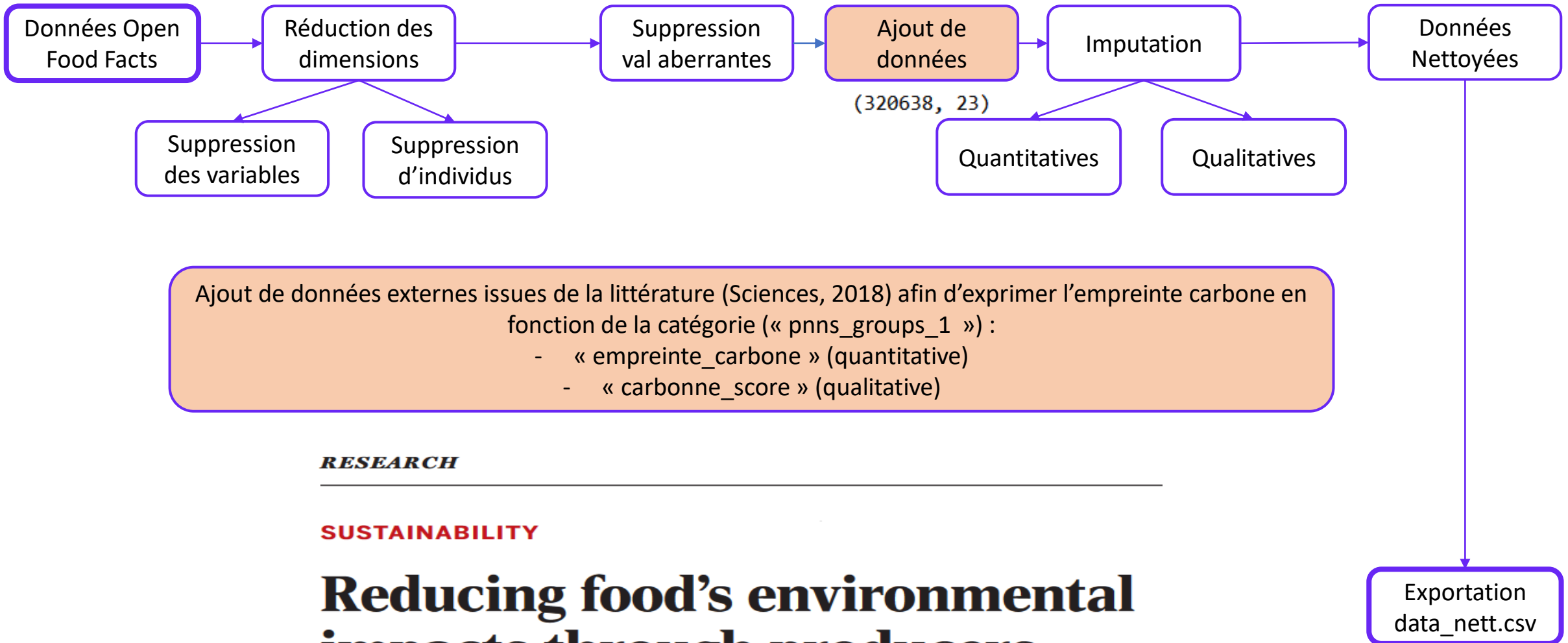


Exportation
data_nettt.csv

Résultats du nettoyage



Résultats du nettoyage



RESEARCH

SUSTAINABILITY

Reducing food's environmental impacts through producers and consumers

J. Poore^{1,2*} and T. Nemecek³

RESEARCH

SUSTAINABILITY

Reducing food's environmental impacts through producers and consumers

J. Poore^{1,2*} and T. Nemecek³

Food's environmental impacts are created by millions of diverse producers. To identify solutions that are effective under this heterogeneity, we consolidated data covering five environmental indicators; 38,700 farms; and 1600 processors, packaging types, and retailers. Impact can vary 50-fold among producers of the same product, creating substantial mitigation opportunities. However, mitigation is complicated by trade-offs, multiple ways for producers to achieve low impacts, and interactions throughout the supply chain. Producers have limits on how far they can reduce impacts. Most strikingly, impacts of the lowest-impact animal products typically exceed those of vegetable substitutes, providing new evidence for the importance of dietary change. Cumulatively, our findings support an approach where producers monitor their own impacts, flexibly meet environmental targets by choosing from multiple practices, and communicate their impacts to consumers.

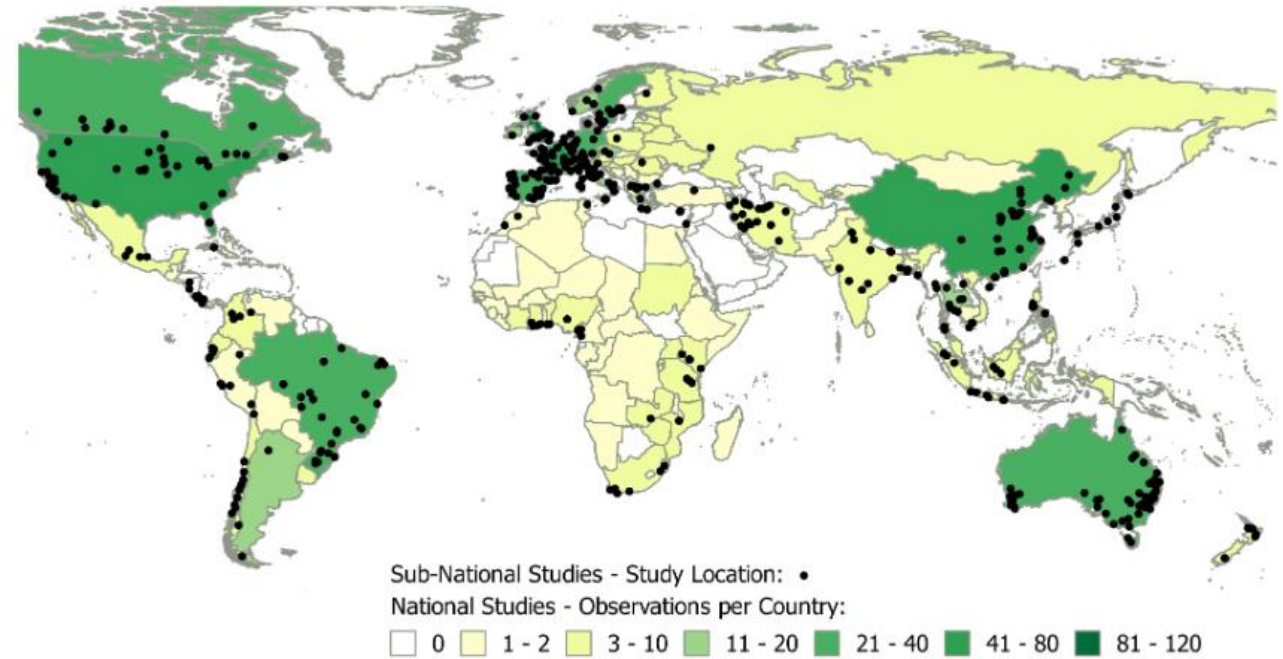


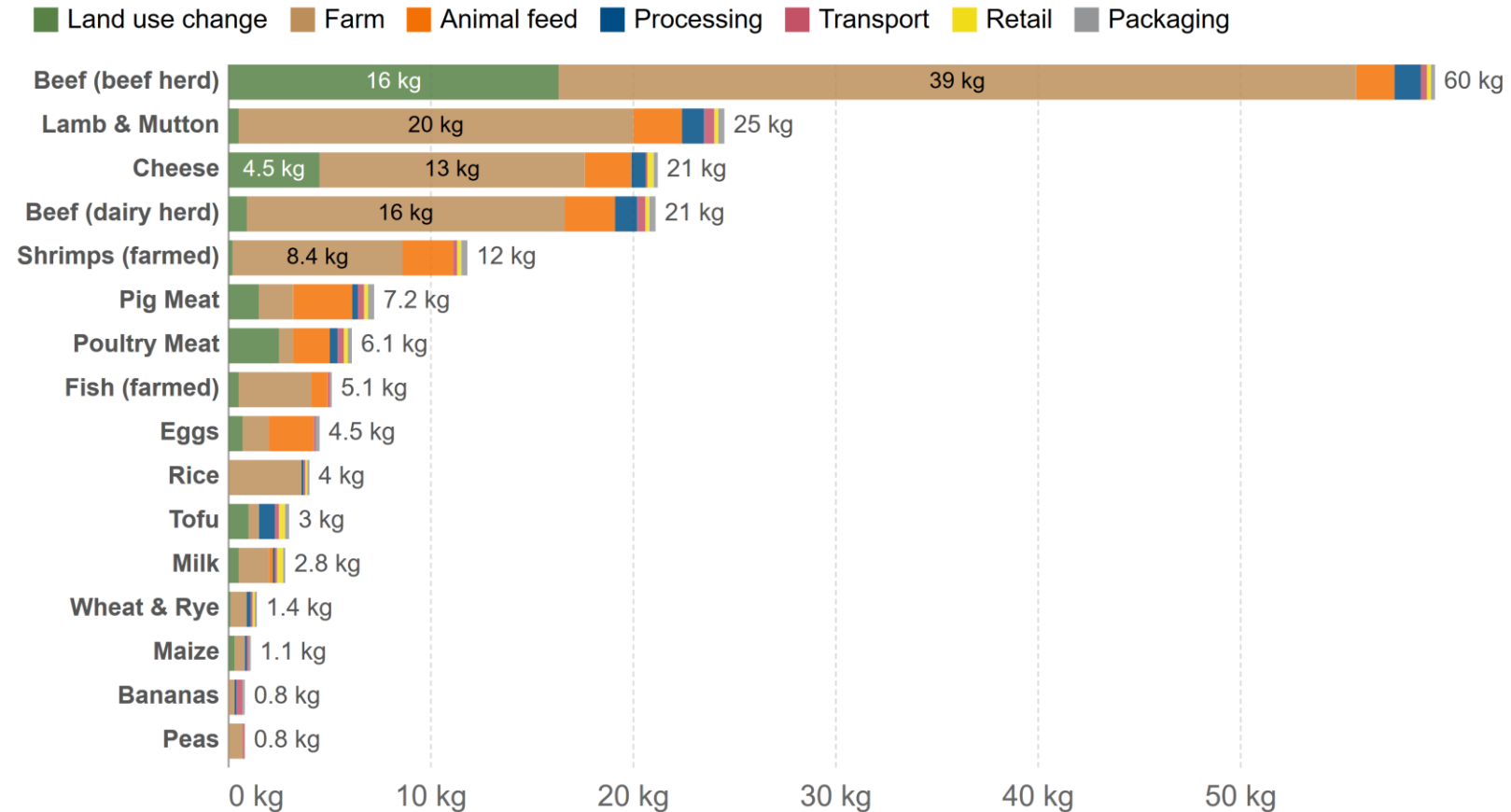
Fig. S2. Map of study locations for all products. Black circles represent locations of subnational studies (n observations = 1160); country shading represents the number of national-level studies per country (n observations = 1118).

Données externes

Food: greenhouse gas emissions across the supply chain

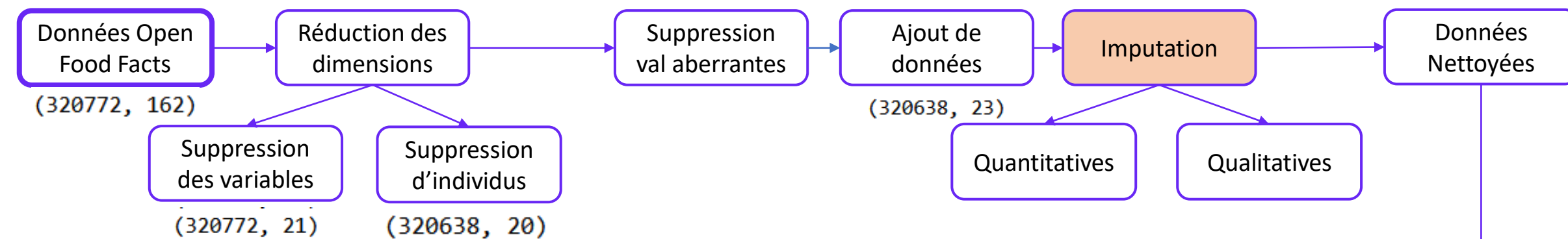
Our World
in Data

Greenhouse gas emissions are measured in kilograms of carbon dioxide equivalents (kgCO₂eq) per kilogram of food. This means non-CO₂ greenhouse gases are included and weighted by their relative warming impact.



Source: Poore, J., & Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*.
OurWorldInData.org/environmental-impacts-of-food • CC BY

Résultats du nettoyage

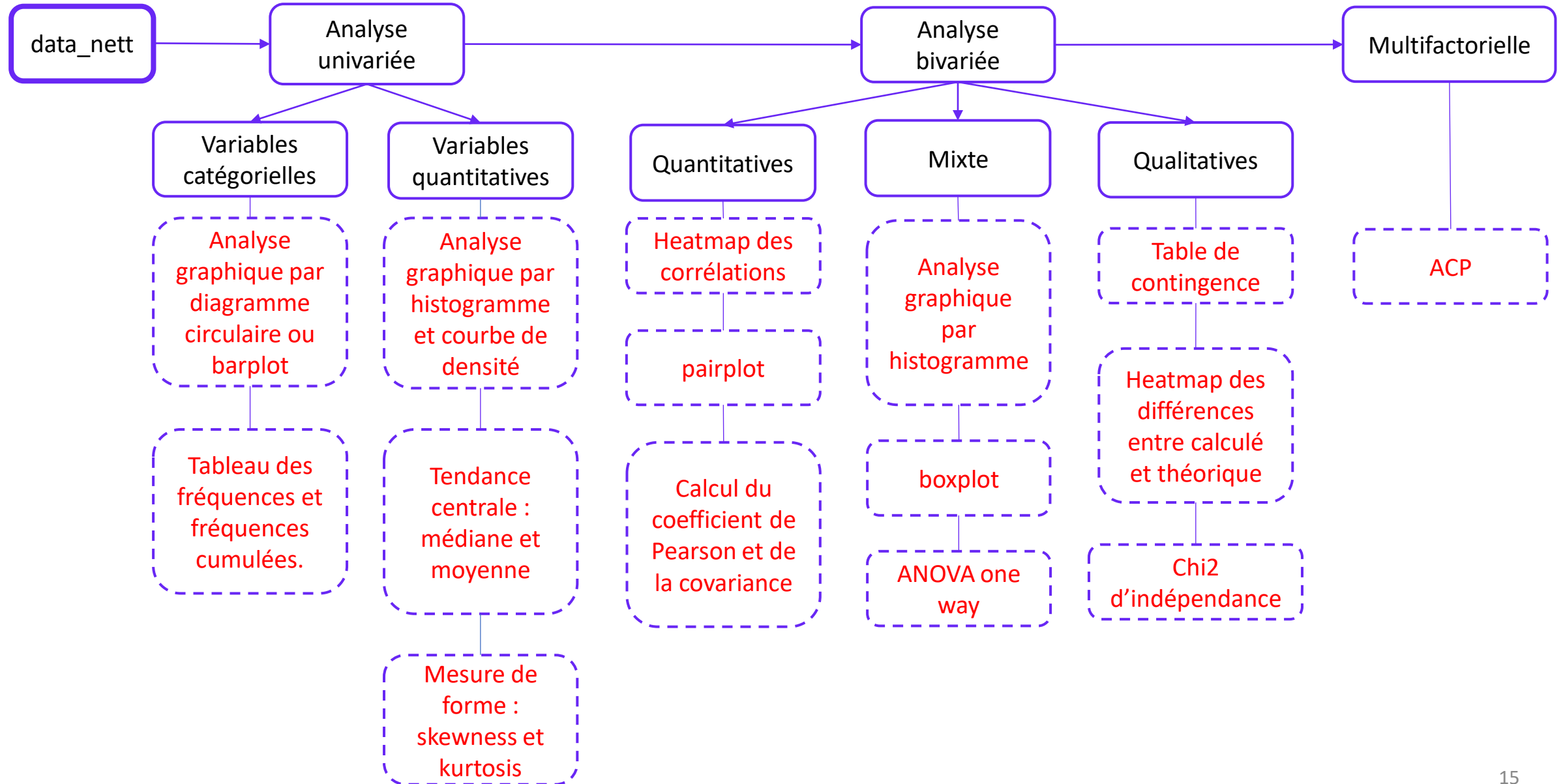


```
data_net.isna().mean()
```

```
: additives_n 0.000000
  ingredients_from_palm_oil_n 0.000000
  energy_100g 0.000000
  fat_100g 0.000000
  saturated-fat_100g 0.000000
  trans-fat_100g 0.000000
  cholesterol_100g 0.000000
  carbohydrates_100g 0.000000
  sugars_100g 0.000000
  fiber_100g 0.000000
  proteins_100g 0.000000
  salt_100g 0.000000
  sodium_100g 0.000000
  vitamin-a_100g 0.000000
  vitamin-c_100g 0.000000
  calcium_100g 0.000000
  iron_100g 0.000000
  nutrition_score_fr_100g 0.000000
  empreinte_carbone 0.000000
  pnns_groups_1 0.714778
  new_grade 0.000000
  carbone_grade 0.000000
  categories_additives_n 0.000000
dtype: float64
```

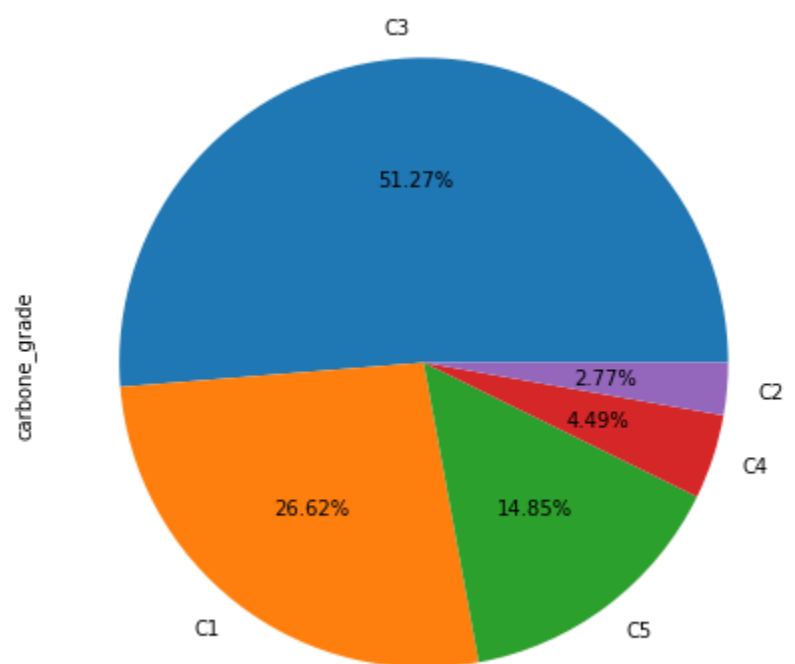
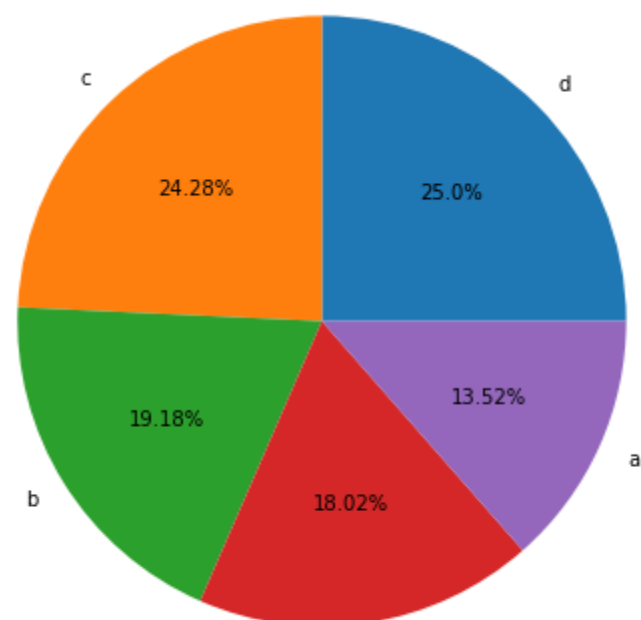
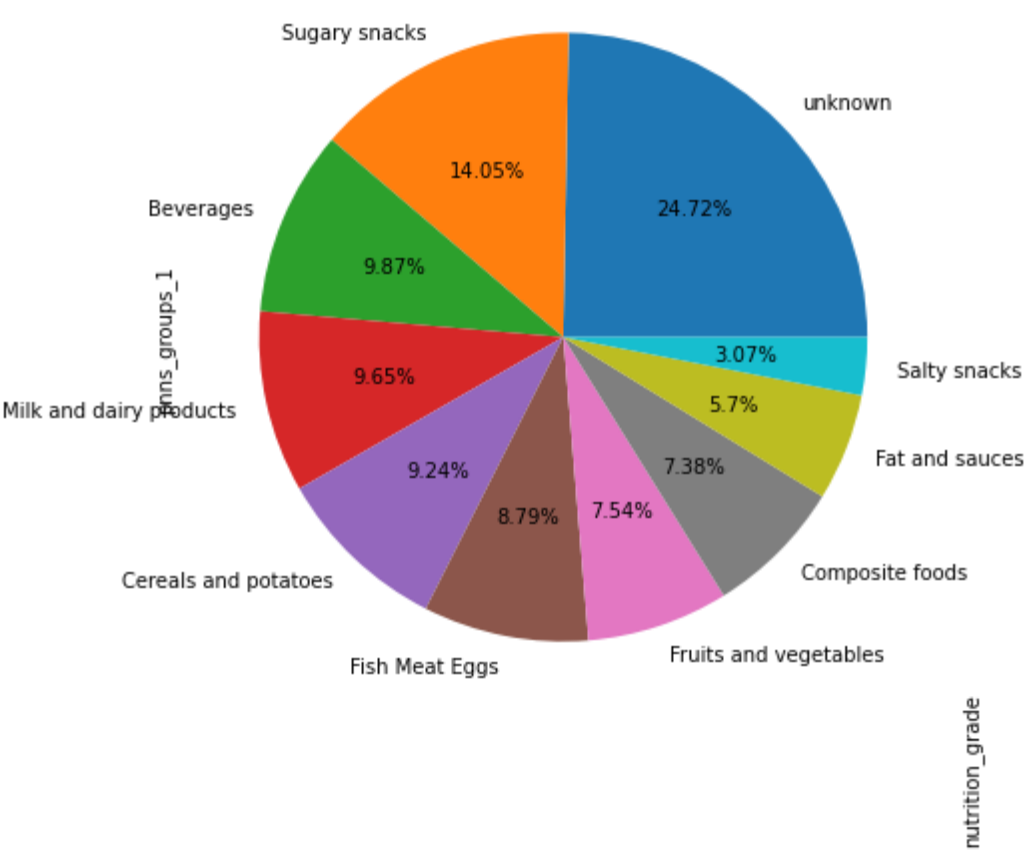
Analyse des données

Stratégie d'analyse



Principaux résultats de l'analyse univariée

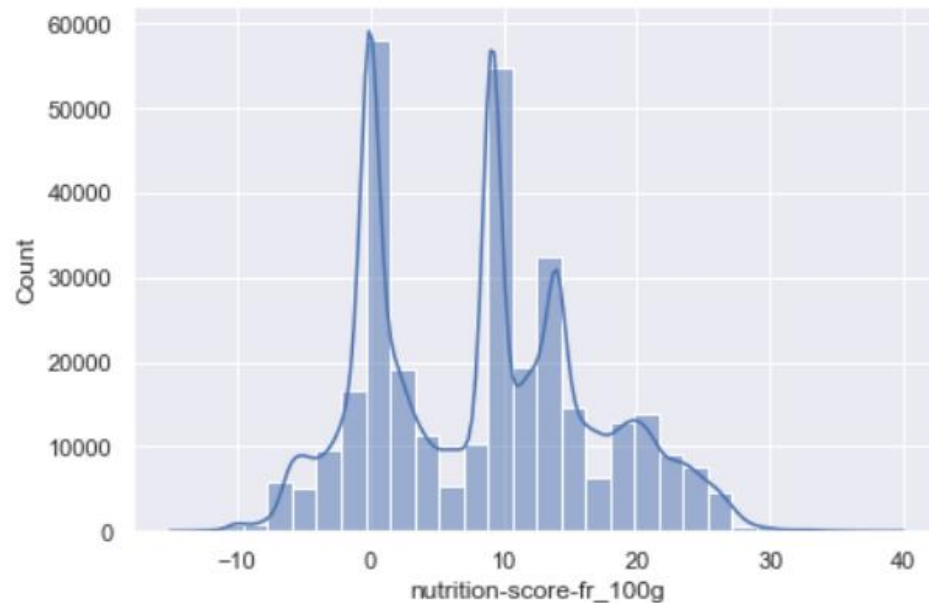
- Variables catégorielles



Principaux résultats de l'analyse univariée

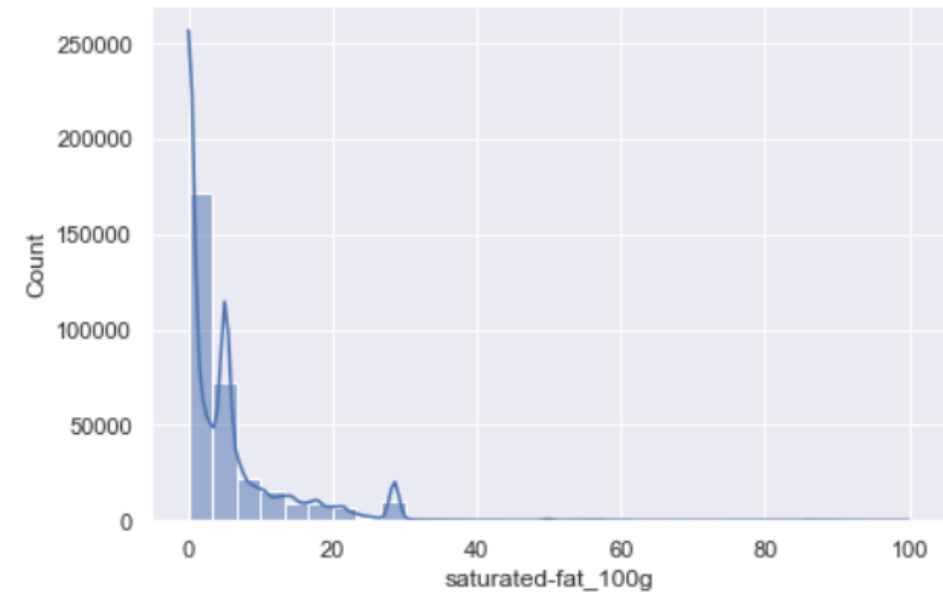
- Variables quantitatives

Distribution homogène



moyenne nutrition-score-fr_100g : 8.28
mediane nutrition-score-fr_100g : 9.17
skewness nutrition-score-fr_100g : 0.27
kurtosis nutrition-score-fr_100g : -0.74

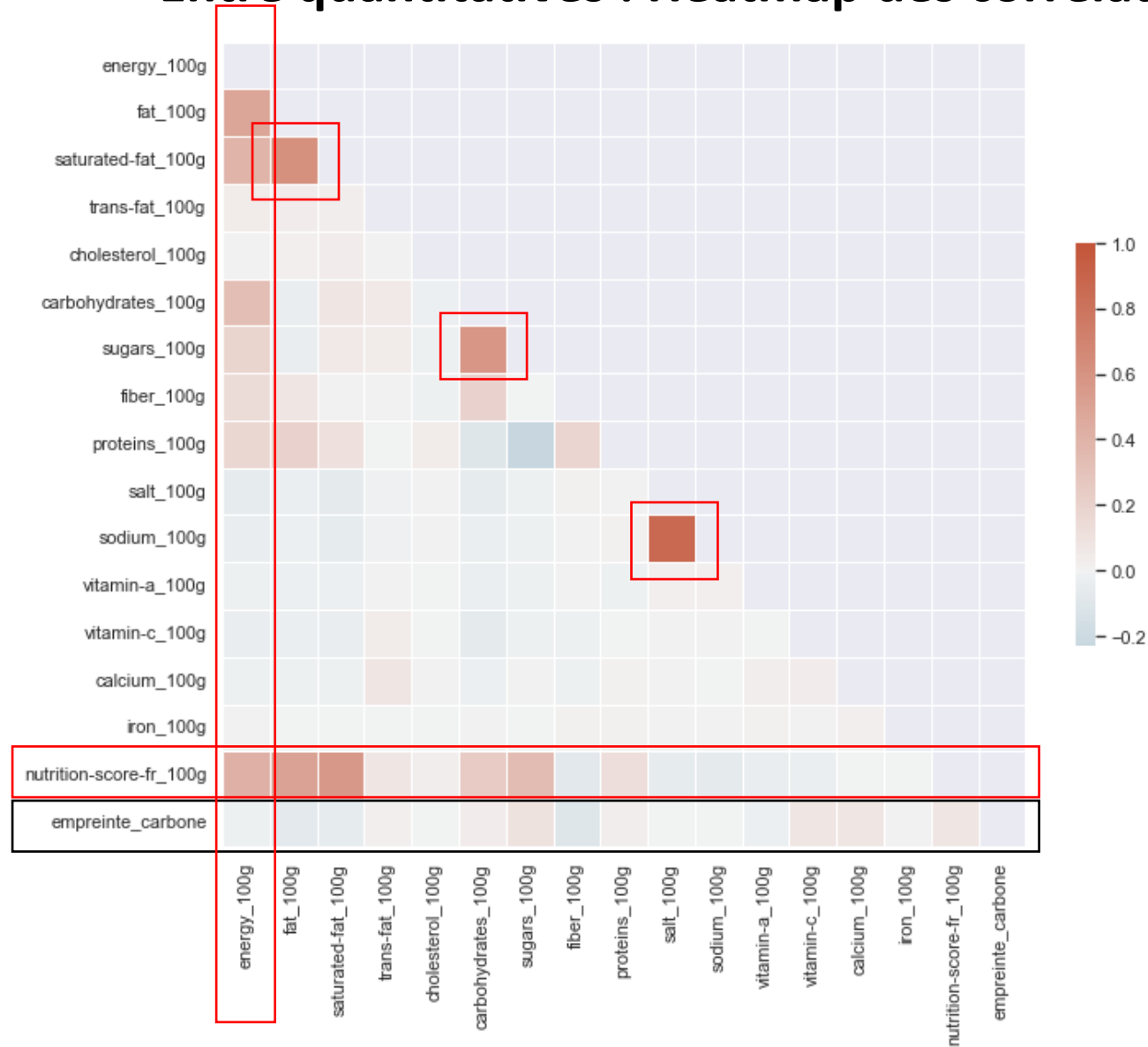
Distribution hétérogène



moyenne saturated-fat_100g : 5.33
mediane saturated-fat_100g : 2.5
skewness saturated-fat_100g : 3.06
kurtosis saturated-fat_100g : 16.17

Principaux résultats de l'analyse bivariée

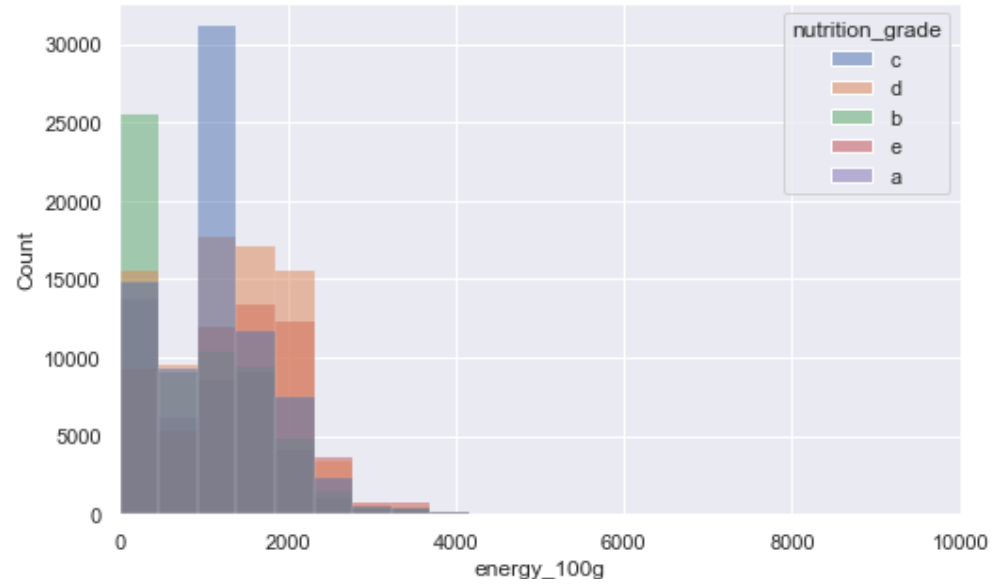
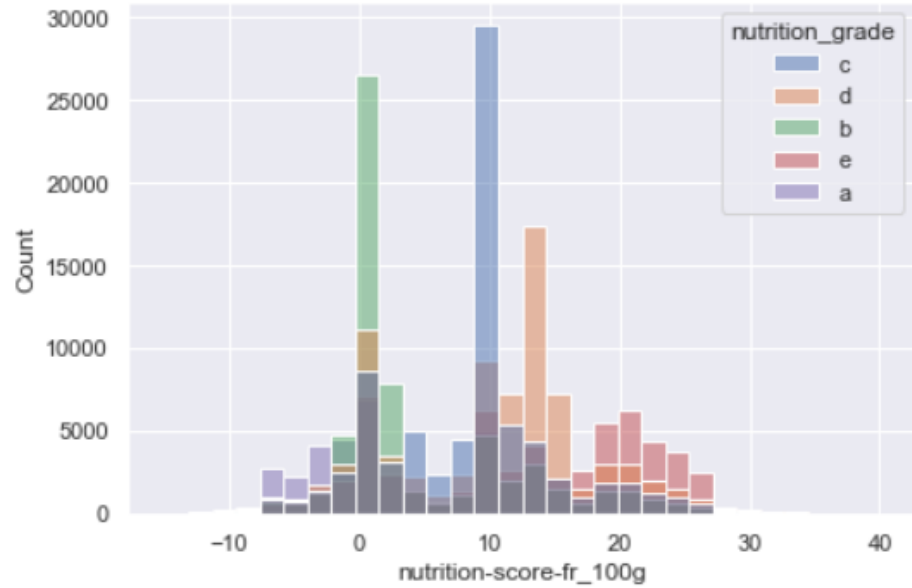
- Entre quantitatives : Heatmap des corrélation et coefficients de Pearson



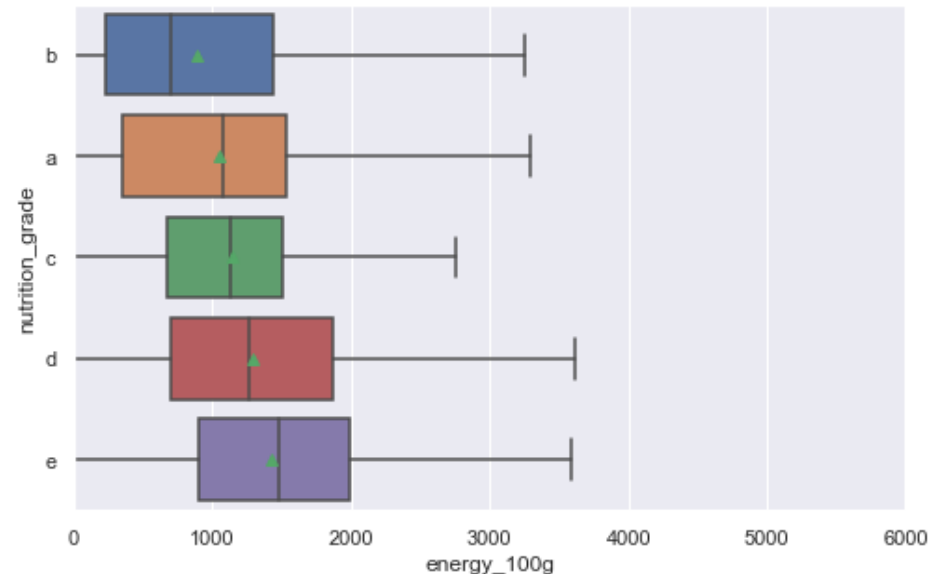
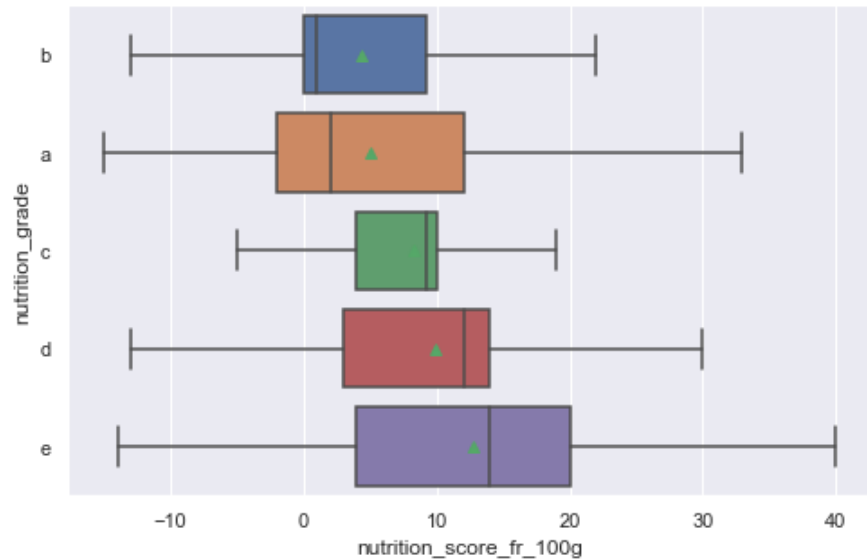
- Le nutri-score est corrélé avec 5 variables nutritionnelles : coef de Pearson de 0.12 à 0.57
- L'énergie est corrélée avec les variables nutritionnelles les plus riches : coef de Pearson de 0.14 à 0.49
- Autres corrélations significatives
 - Sel et Sodium : 0.87
 - Sucre et carbohydrates : 0.59
 - Graisses et graisses saturées : 0.62
- L'empreinte carbone n'est corrélée avec aucune variable, ce qui est cohérent avec mon projet d'application.
- L'absence de lien entre l'empreinte carbone et les caractéristiques nutritionnelles garantit au consommateur d'avoir une information supplémentaire utile et indépendante du nutri-score.

Principaux résultats de l'analyse bvariée

- Entre quantitatives et qualitatives : exemple du Nutri-grade

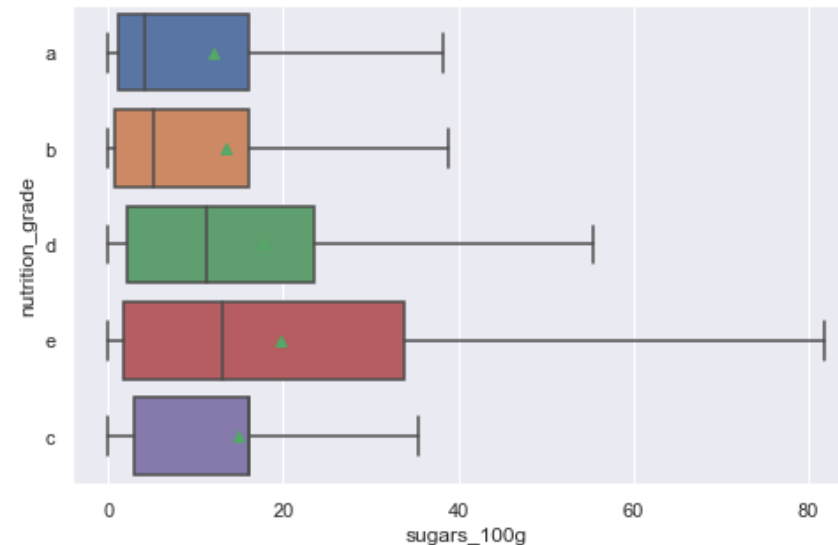
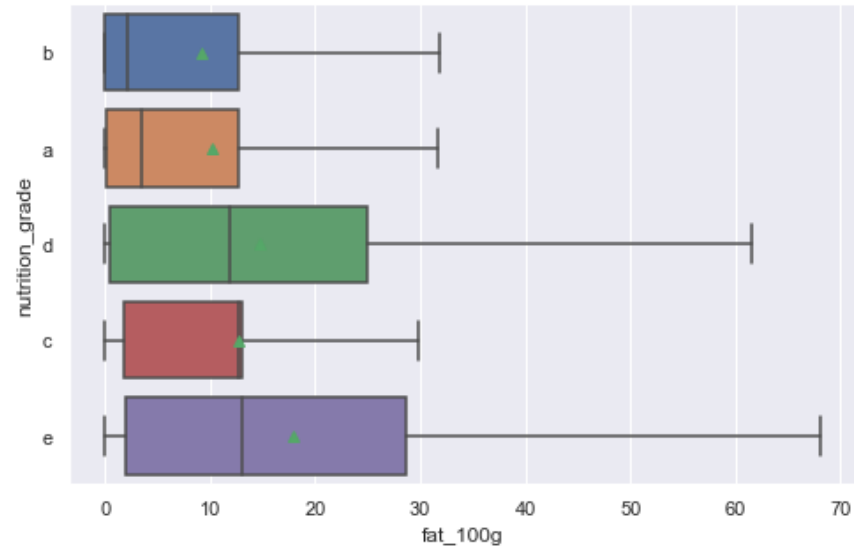
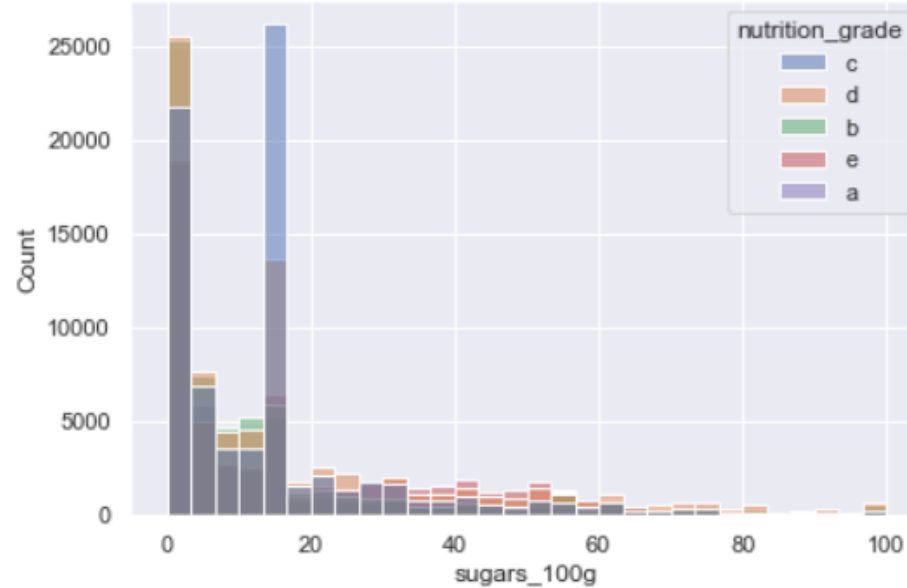
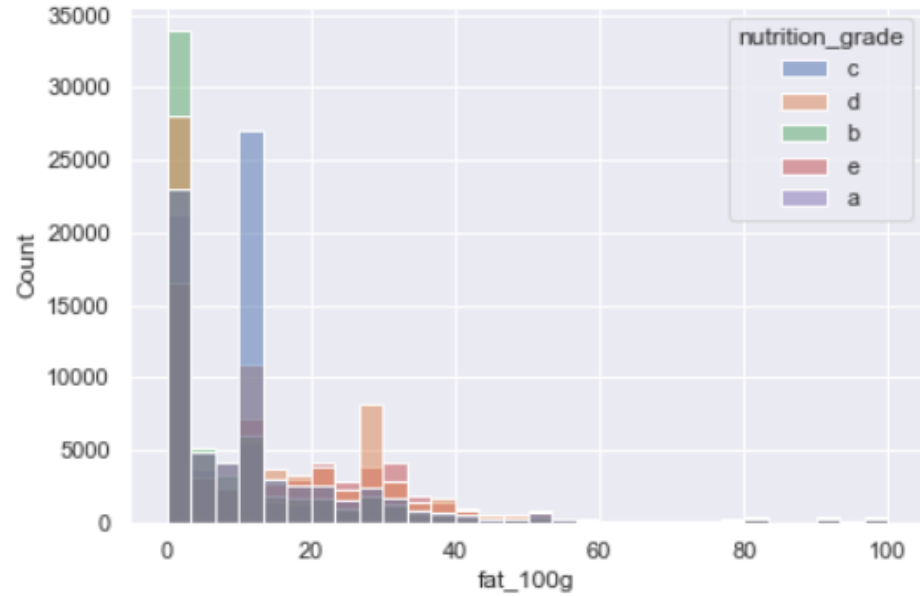


➤ Le Nutri-grade est lié au Nutri-score, et à l'énergie (confirmé par ANOVA).



Principaux résultats de l'analyse bvariée

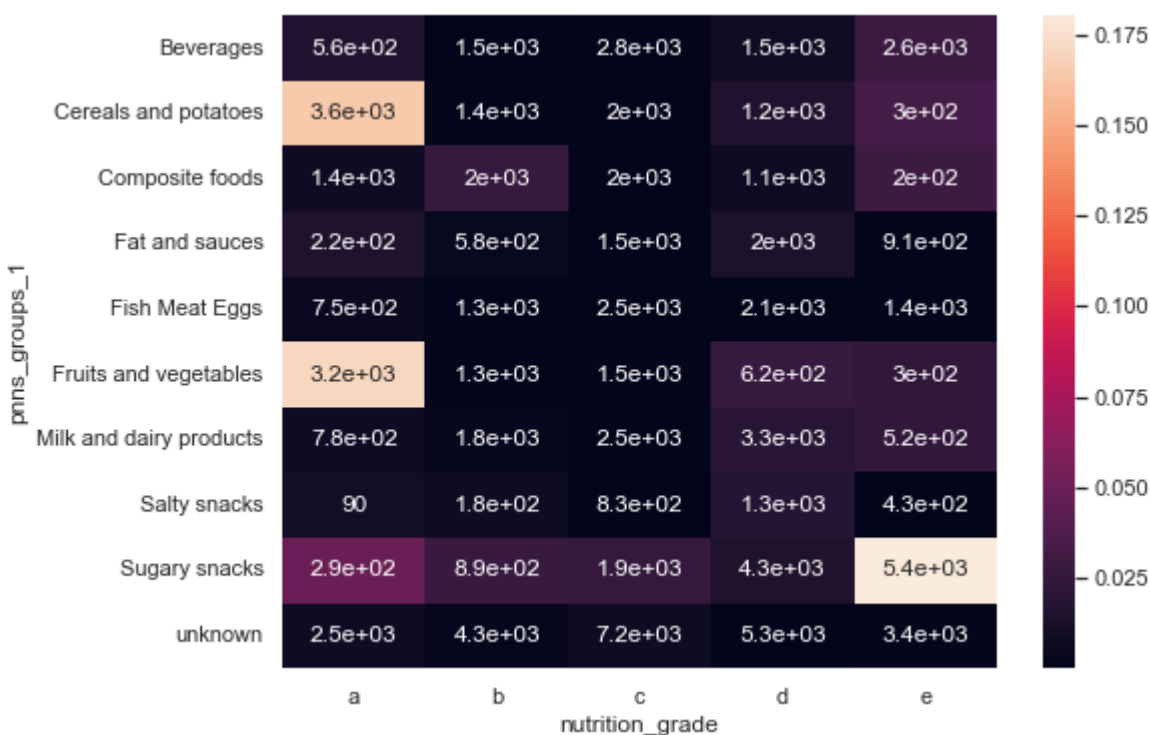
- Entre quantitatives et qualitatives : exemple du Nutri-grade



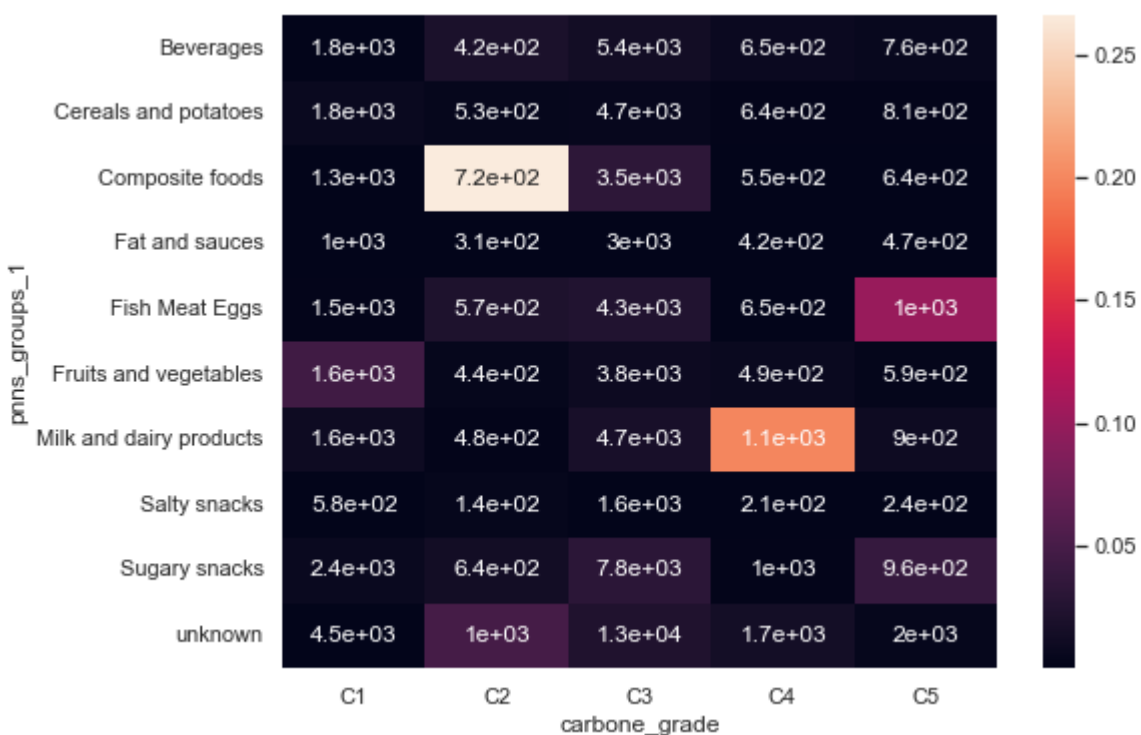
➤ Tendance confirmée avec les caractéristiques nutritionnelles les plus énergétiques.

Principaux résultats de l'analyse bivariée

- Entre qualitatives :

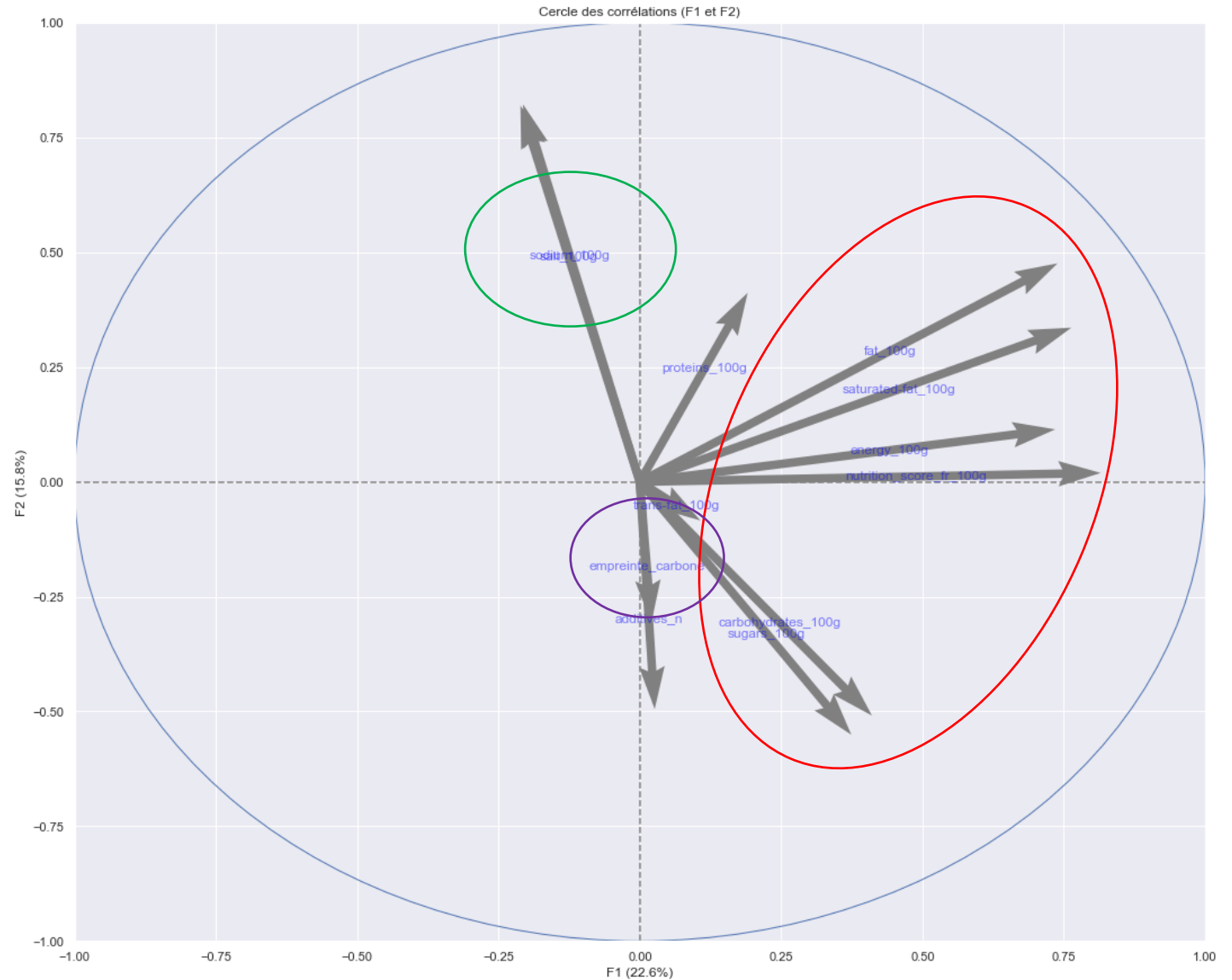
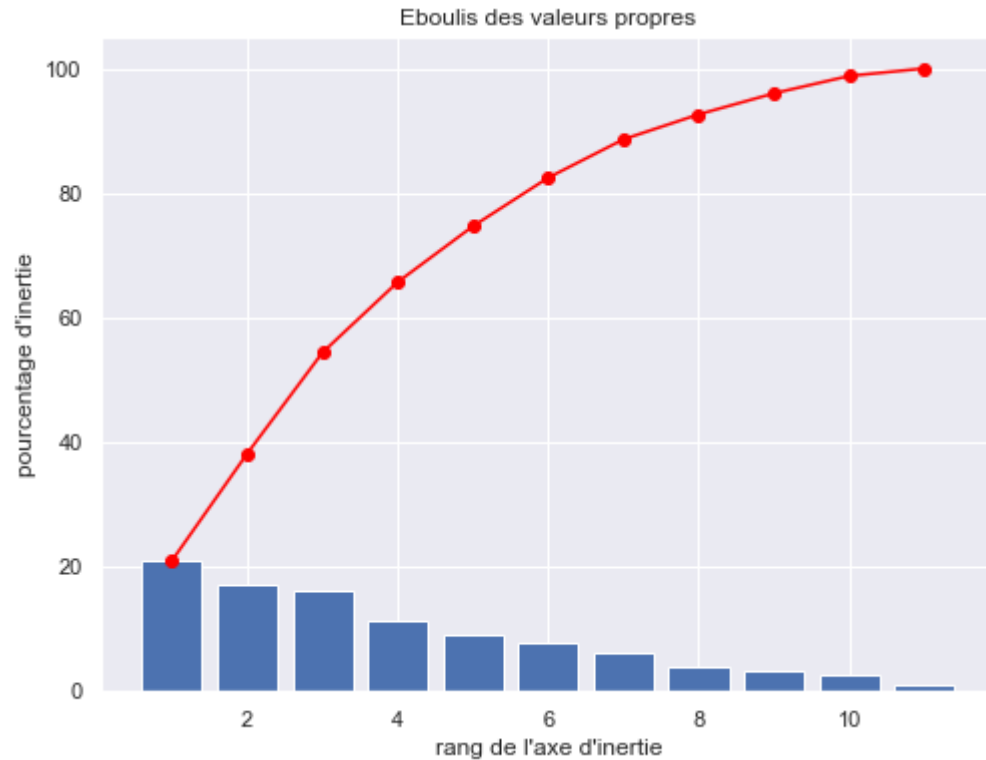


➤ Cohérence alimentaire confirmée par le Chi2 test.



➤ Cohérence avec le mode de production confirmée par le Chi2 test.

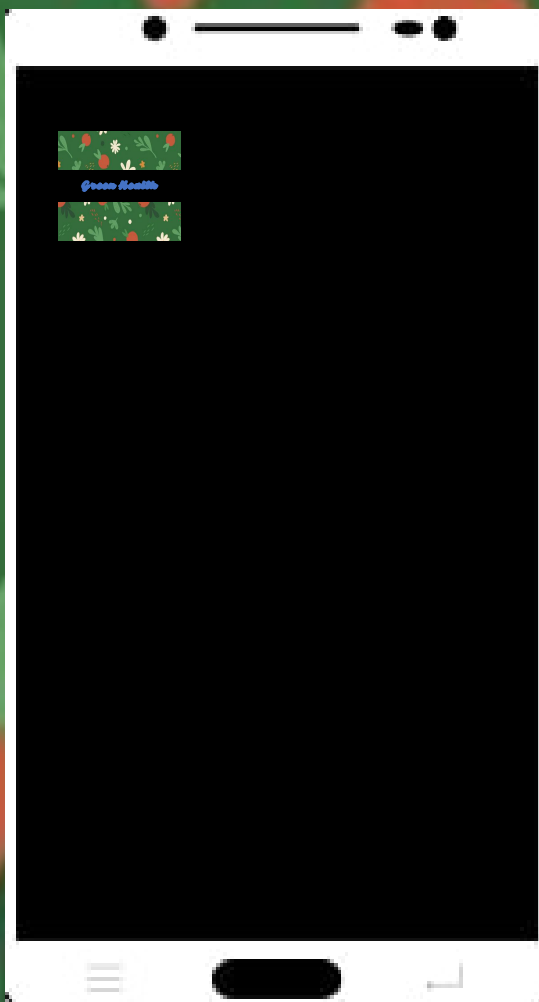
Résultats de l'analyse multifactorielle



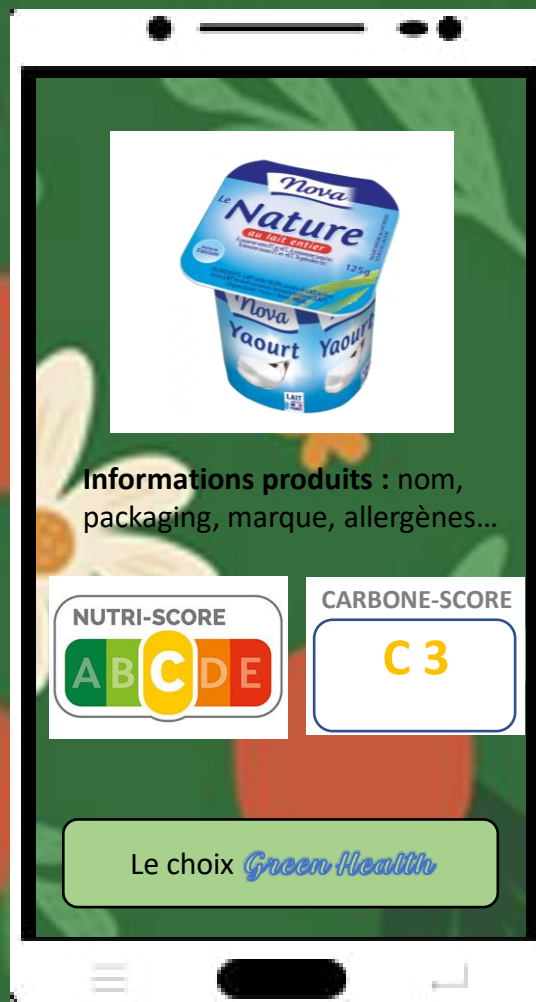
- Les graisses, les sucres, l'énergie et le nutri-score sont les variables les plus corrélées positivement avec le 1^{er} axe.
- Le sel et le sodium sont les variables les plus corrélées positivement avec le 2^{ème} axe.
- L'empreinte carbone n'est pas du tout corrélée au 1^{er} axe.

L'application

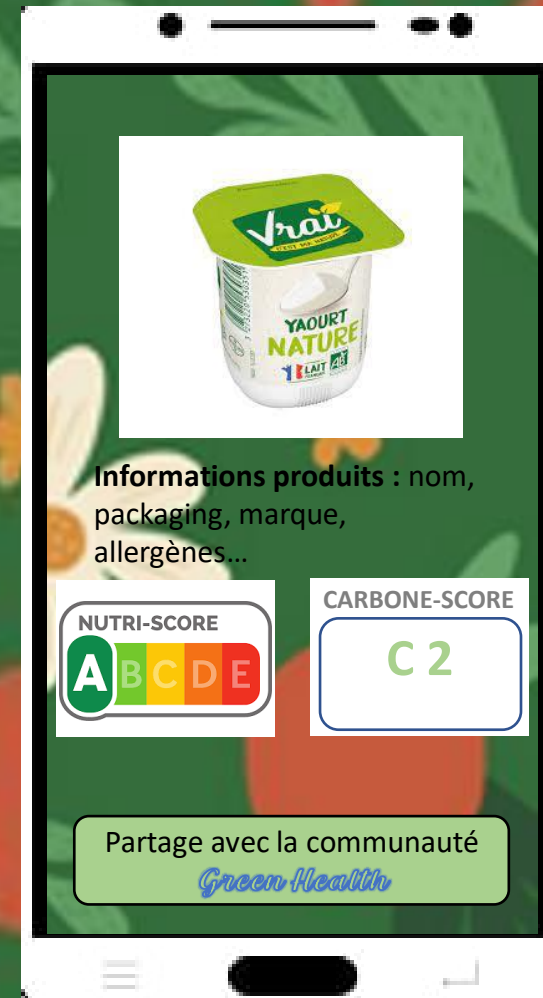




Télécharge l'application
Green Health



Scan ton produit et
obtiens tes infos éco-
santé



Opte pour le choix
Green Health :
préserve ta santé et ta
planète !

Conclusions

➤ Nutri-score

- Les données **Open Food Facts** sont suffisantes pour établir un lien entre **le Nutri-score (ou Nutri-grade) et plusieurs variables nutritionnelles**.
- Les observations réalisées ont pu être validée par des grandeurs et des tests statistiques.
- **Le Nutri-score est une information robuste essentielle à notre application.**

➤ Carbone-score

- Malgré la perte d'informations essentielles au calcul du Carbone-score lors du nettoyage des données, nous avons pu **sourcer des données externes fiables** qui nous ont permis de **taguer notre jeu de données avec un Carbone-score** en fonction de la catégorie du produit.
- Nous avons établi une **corrélation entre le Carbone-score et les produits issus de l'élevage intensif**, connus pour avoir un fort impact écologique du fait des conditions d'élevage.
- Nous avons démontré que le **Carbone-score est décorrélié des caractéristiques nutritionnelles du produit et donc du Nutri-score**, ce qui rend pertinent pour le consommateur un projet d'application associant les 2 niveaux d'information.

Conclusions

➤ Faisabilité

- A la lumière des analyses réalisées, l'application **Green Health** est tout à fait envisageable sur la base des données Open Food Facts.
- Cette application répond aussi à une demande grandissante du consommateur, qui oriente de plus en plus ses choix vers **une consommation éco-responsable et locale**, notamment chez les jeunes.

➤ Axes d'amélioration

- Notre analyse dépend de **l'apport de données externes** pour enrichir notre jeu de données d'informations fiables concernant le carbone-score. Les données externes rajoutées restent néanmoins **pauvres en information concernant les empreintes carbones des produits transformés**. Or ce type de produit est majoritaire dans le jeu de données Open Food Facts. Il faudrait **sourcer plus d'informations sur les empreintes carbone pour ce type de produits**.
- Un autre axe de travail peut être envisager : **conserver les variables du jeu de données initiales** en lien avec l'empreinte carbone. Cela éviterait d'intégrer des données externes, mais il faut accepter de travailler avec un **data-set considérablement tronqué**.

Merci pour votre attention

