

Projet 3

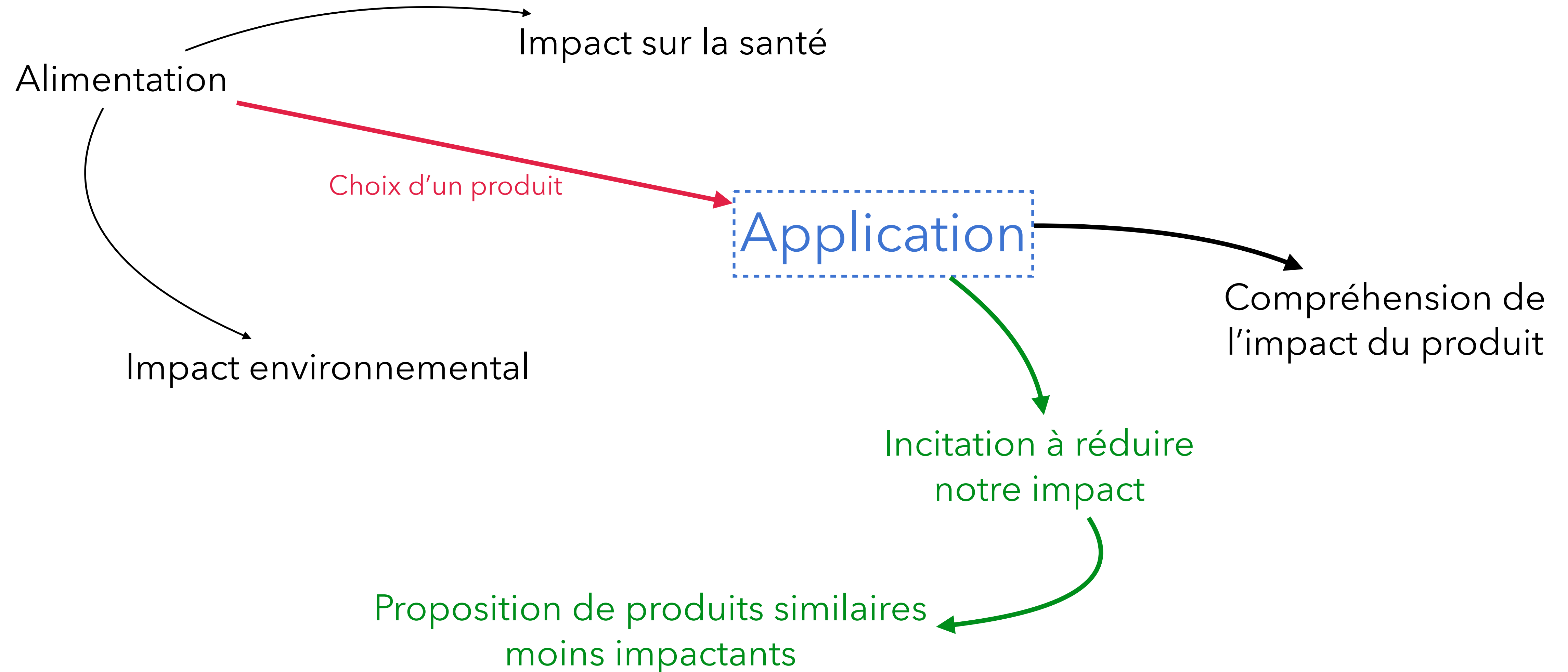
Concevez une application au service de la santé publique

Appel à projet de santé publique France

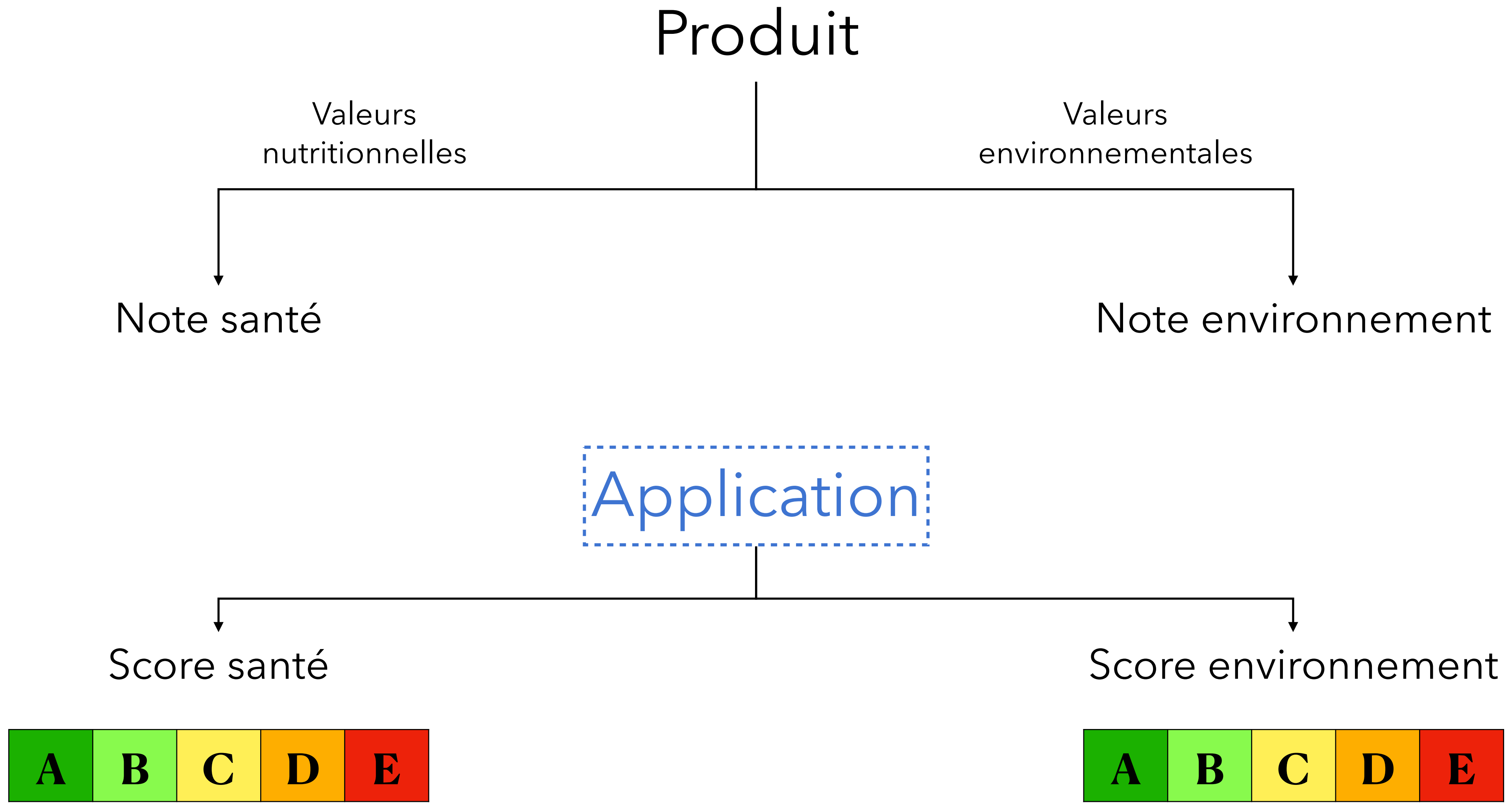
Trouver une idée innovante d'application en lien avec l'alimentation

- 1 - Présentation de l'application
- 2 - Sélection et nettoyage des variables
- 3 - Réduction dimensionnelle
- 4 - Construction des scores
- 5 - Influence des variables sur le score
- 6 - Conclusion

1- Présentation de l'application



1- Présentation de l'application



2- Sélection et nettoyage des variables

Jeu de données du site Openfoodfacts

Identification des produits

Code

Nom du produit

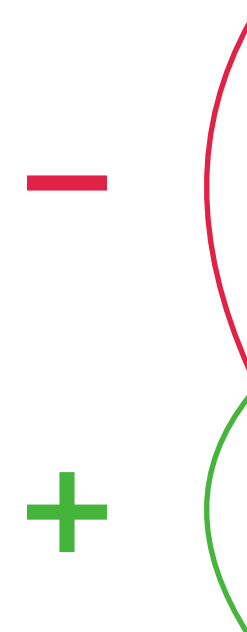
- product_name
- generic_name

Catégorie du produit

- categories
- categories_tags
- pnns_groups_1
- pnns_groups_2

Score santé

Valeurs nutritionnelles

- 
- energie (kJ)
 - gras / gras saturés
 - sel
 - sucre
 - alcool
 - vitamines
 - fibres
 - protéines

Valeurs pour 100g
de produit

Additifs

Score environnement

Présence d'huile de palme

- ingredients_from_palm_oil_n
- ingredients_that_may_be_from_palm_oil_n

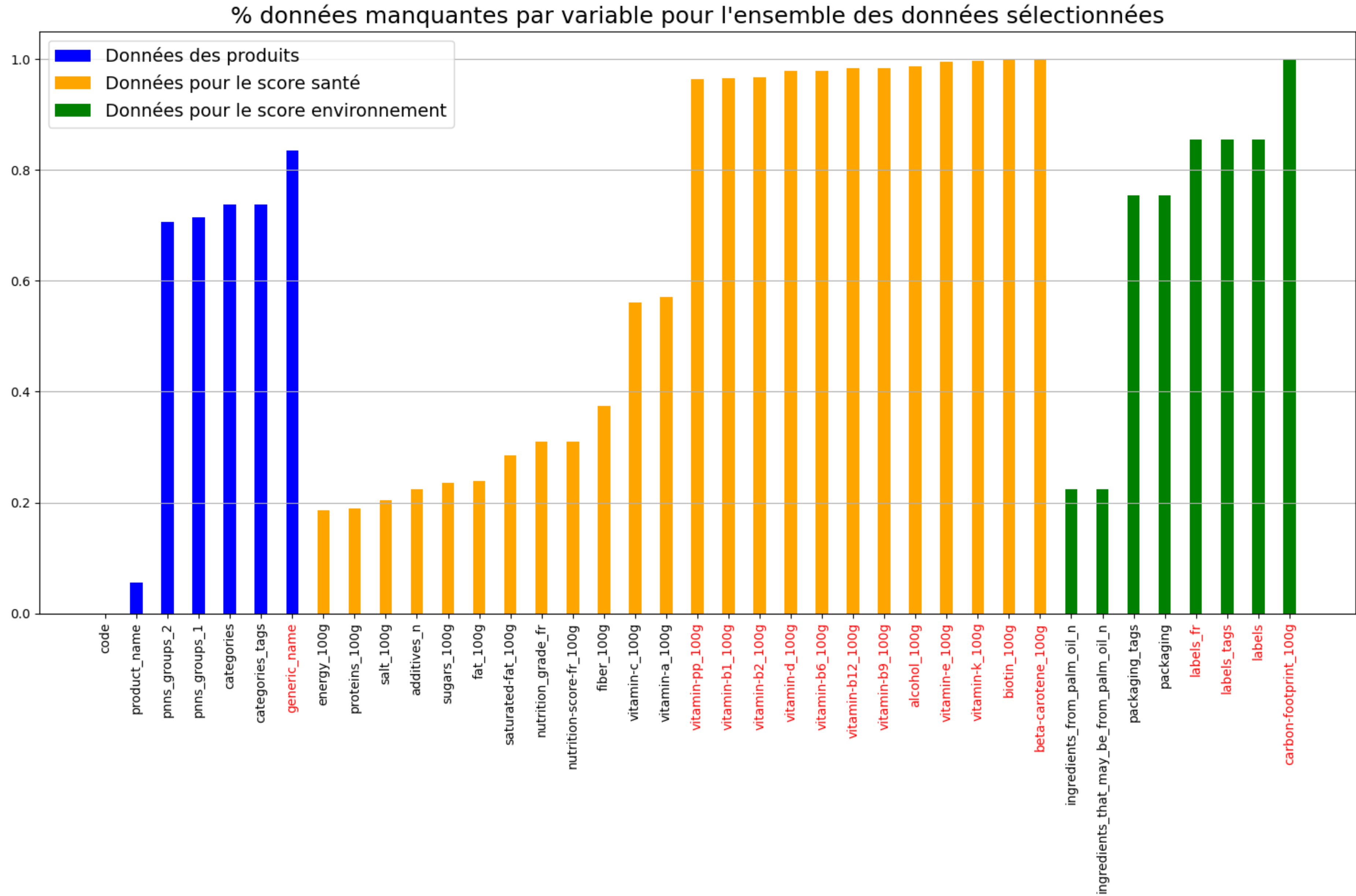
Produit issu de l'AB
/ végétarien

- Labels
- labels_tags
- labels_fr

Packagings utilisés

- packaging
- packaging_tags

2- Sélection et nettoyage des variables



2- Sélection et nettoyage des variables

Traitement des valeurs redondantes pour les données produits :

Codes :

0,007 % codes manquants

Suppression

Nom de produit :

product_name

~~generic_name~~

Suppression des produits
avec des noms manquants

Catégories de produit :

~~categories~~

~~categories_tags~~

pnns_groups

pnns_groups_1

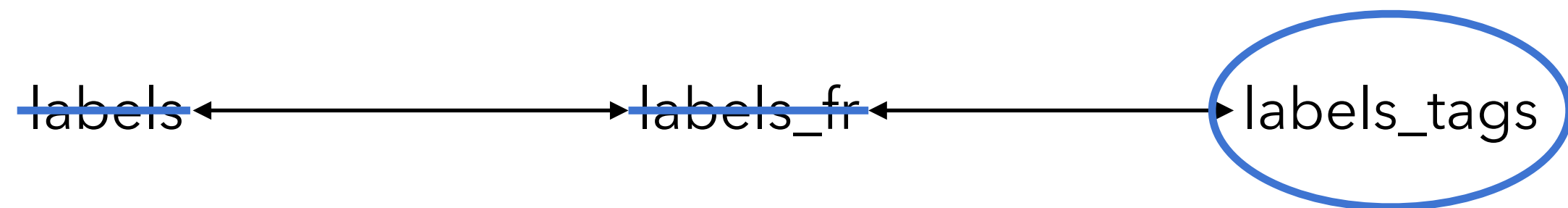
pnns_groups_2

Complémentaires

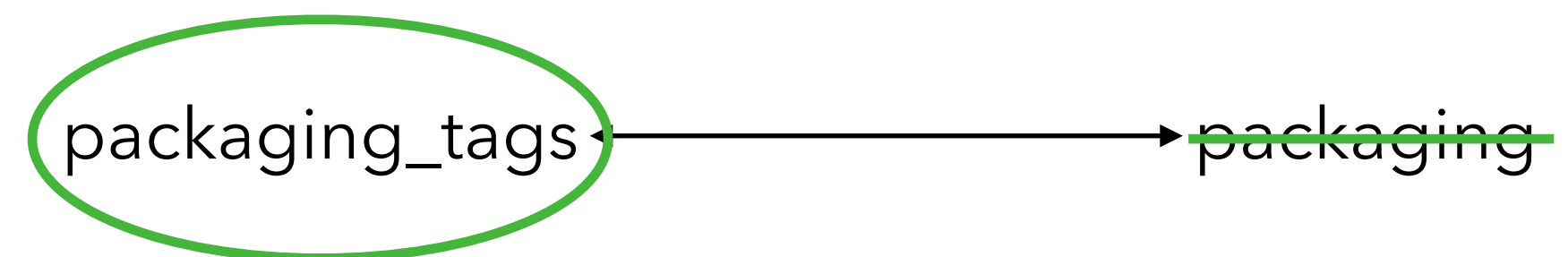
2- Sélection et nettoyage des variables

Traitement des valeurs redondantes pour les données environnement :

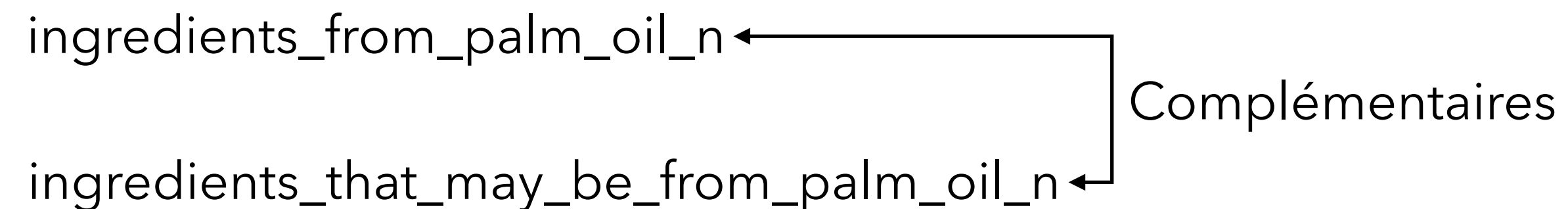
Labels :



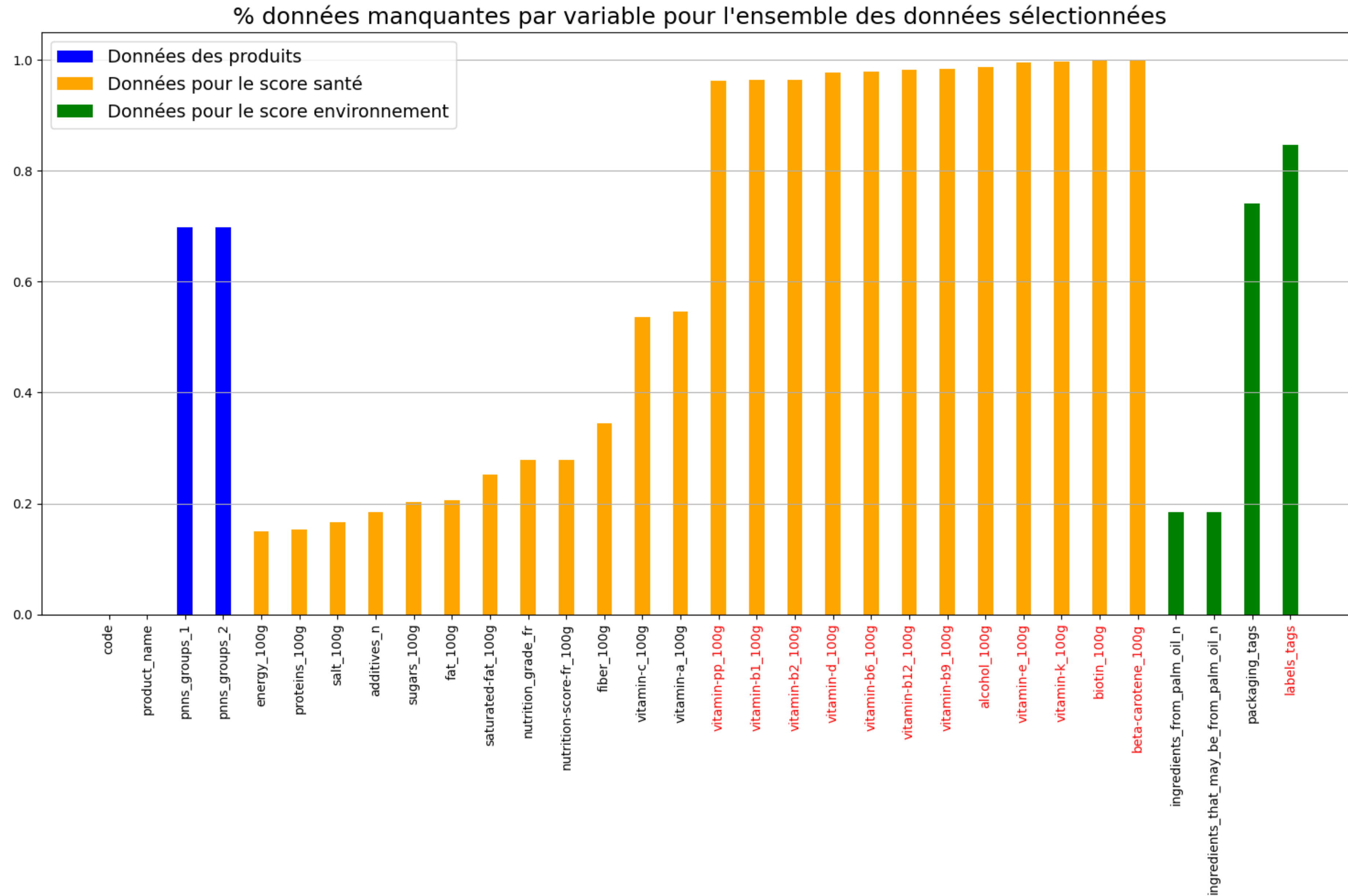
Packagings :



Produits issus de l'huile de palme :

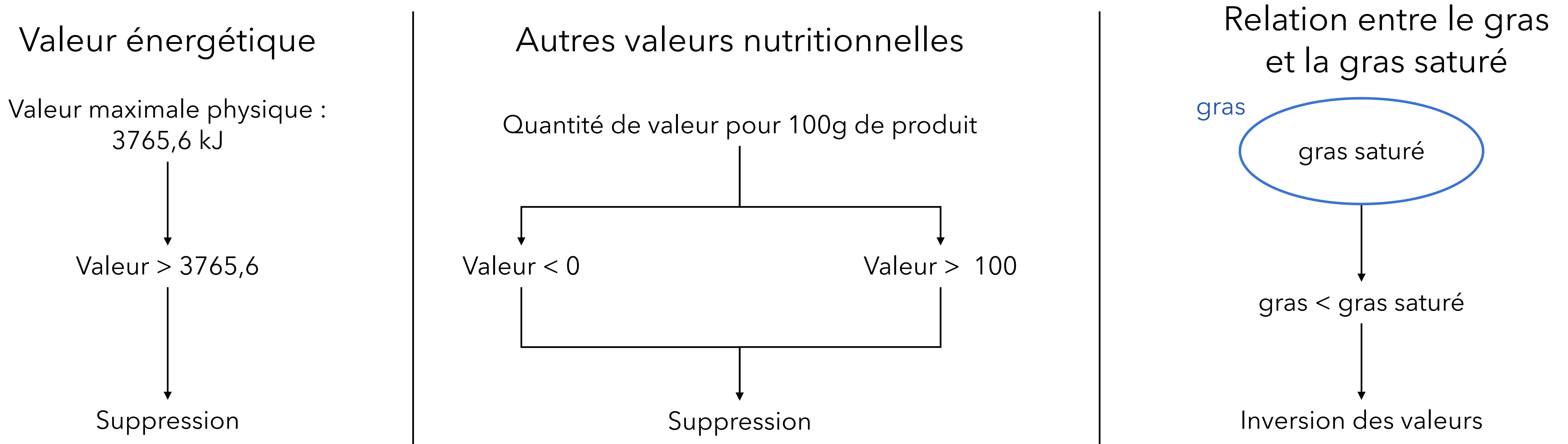


2- Sélection et nettoyage des variables



2- Sélection et nettoyage des variables

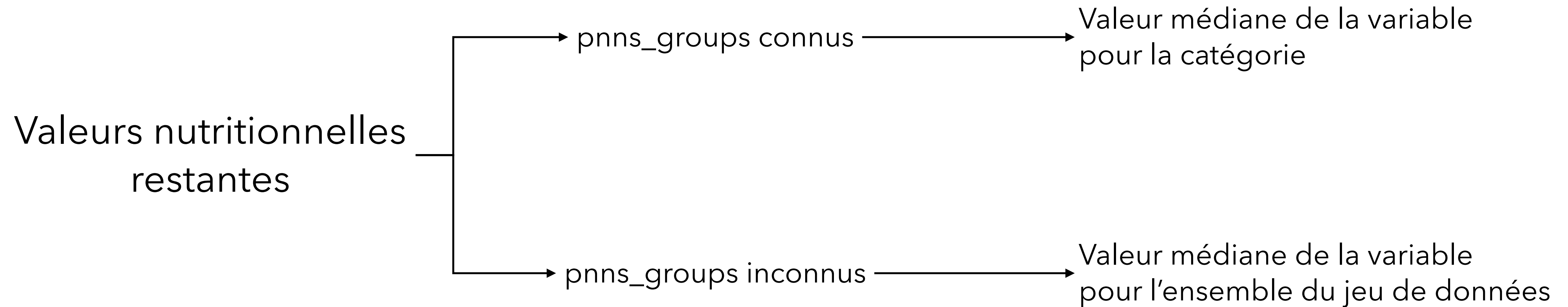
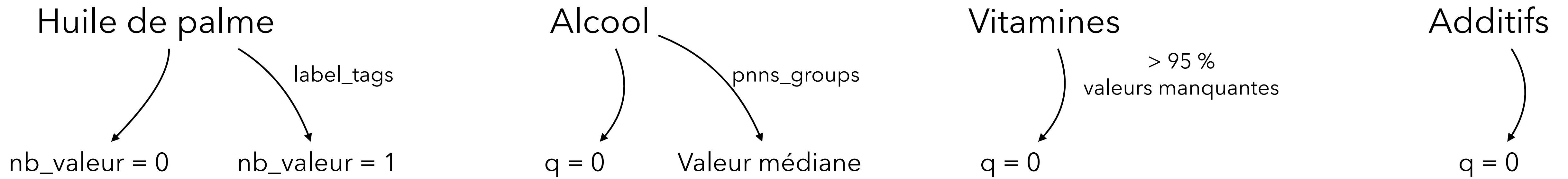
Traitement des valeurs nutritionnelles aberrantes :



Somme valeurs nutritionnelles > 100 → Normalisation entre 0 et 100

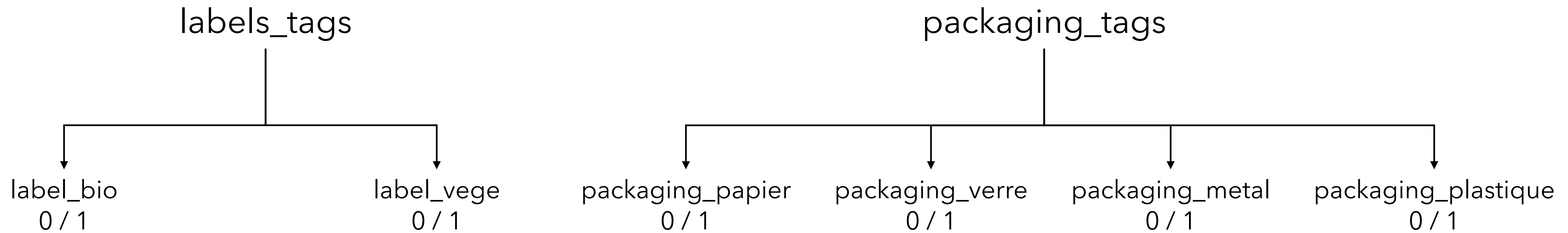
2- Sélection et nettoyage des variables

Complétion des valeurs :

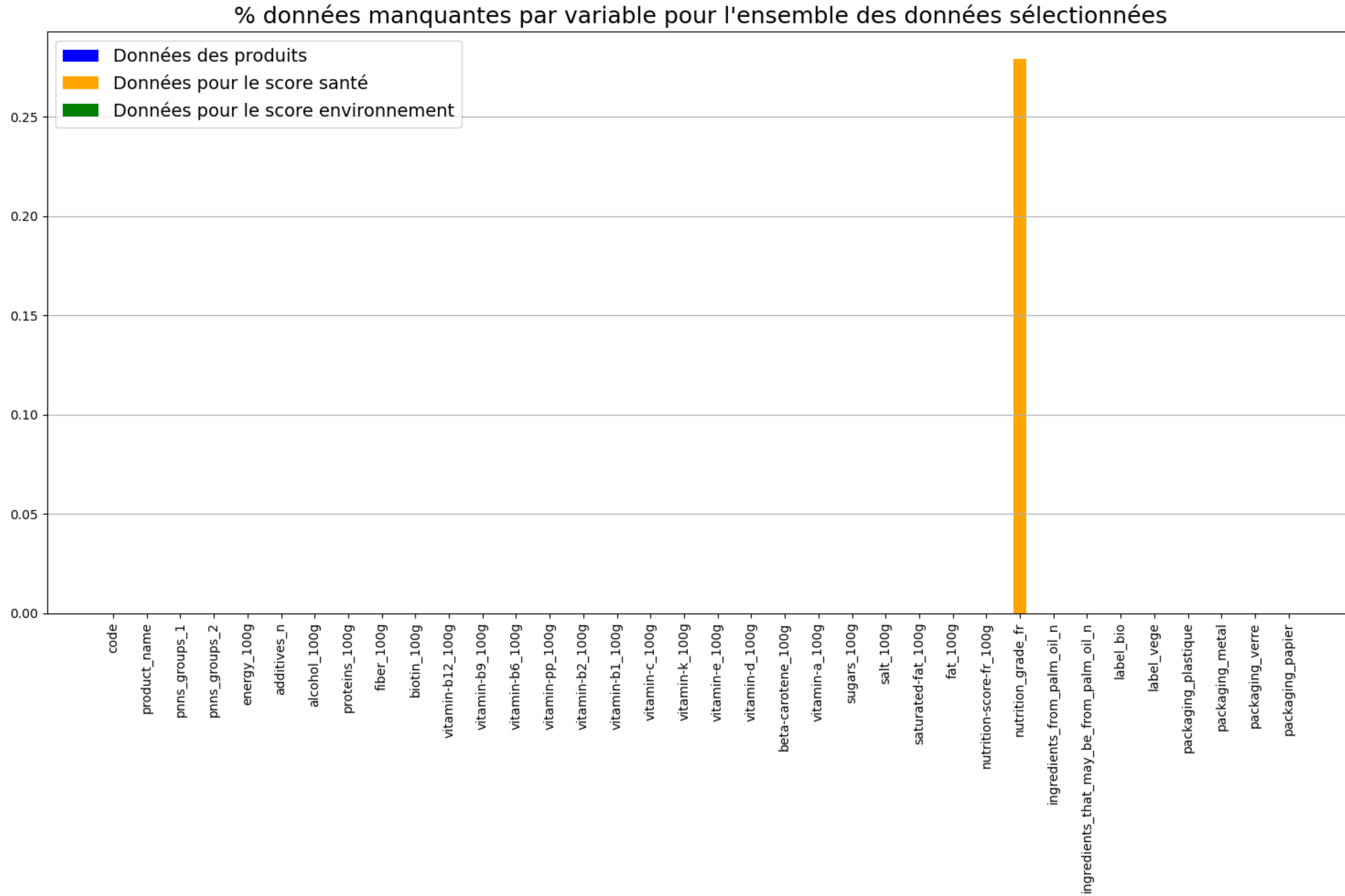


2- Sélection et nettoyage des variables

Encodage des variables qualitatives pour le score environnement



2- Sélection et nettoyage des variables



3- Réduction dimensionnelle

Réalisation d'une ACP sur les variables pour le calcul des scores
(cherche a expliquer la variance totale)

Score Sante :

- Valeurs nutritionnelles
- Additifs

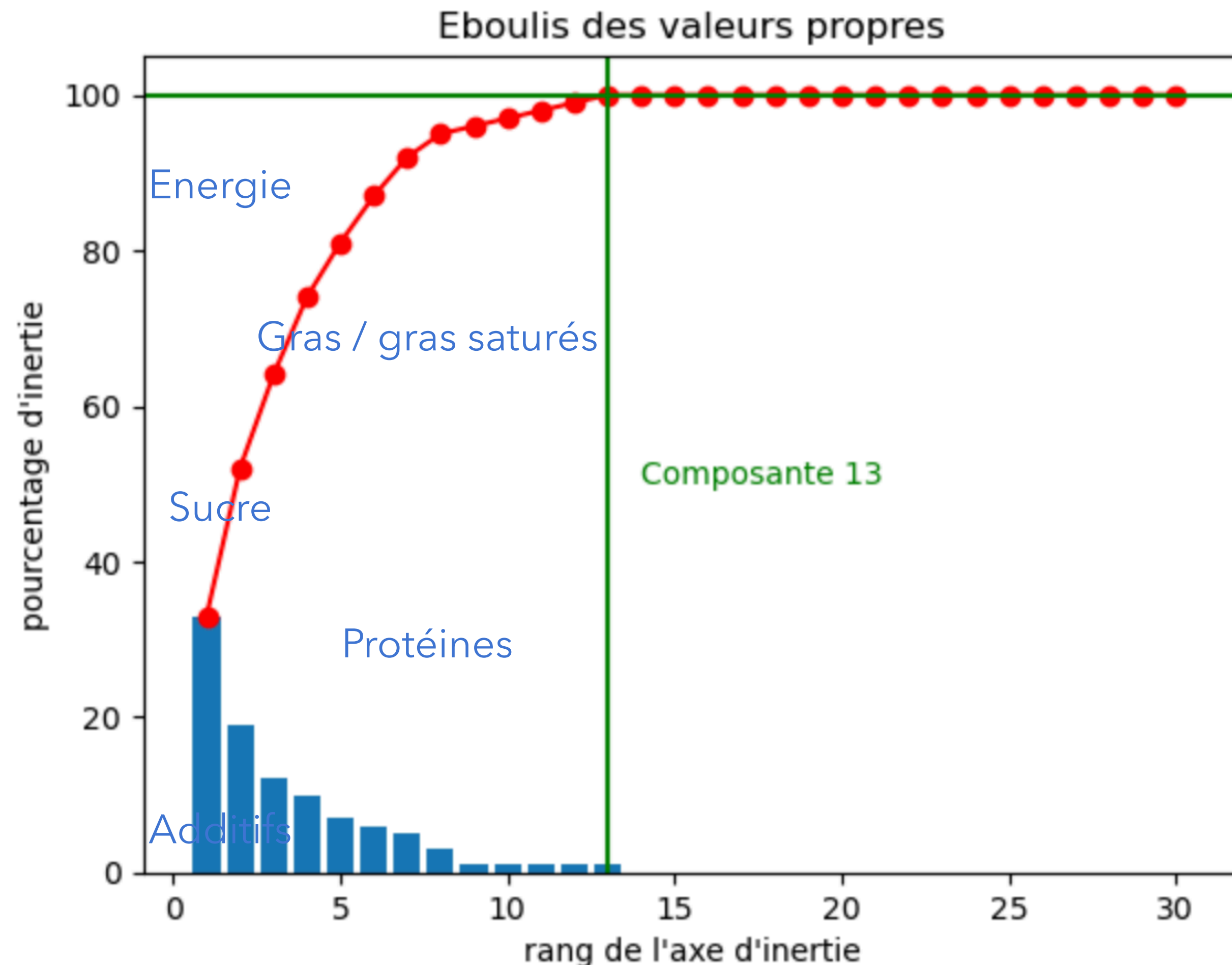
Score environnement :

- Labels bio / vege
- Labels packaging
- Ingrédients issus de l'huile de palme.

Labels bio / vege

Labels packaging

Ingrédients issus de l'huile de palme.



4-Construction des scores

Note santé

0 → 50

Variables nutritionnelles

variable_1

variable_2

...

note_variable_1

note_variable_2

....

Note nutritionnelle

Note Additifs

Note santé

Note environnement

0 → 50

Labels bio / vege → note_label

Labels packaging → note_packaging

Ingrédients huile de palme → note_palm_oil

Note environnement

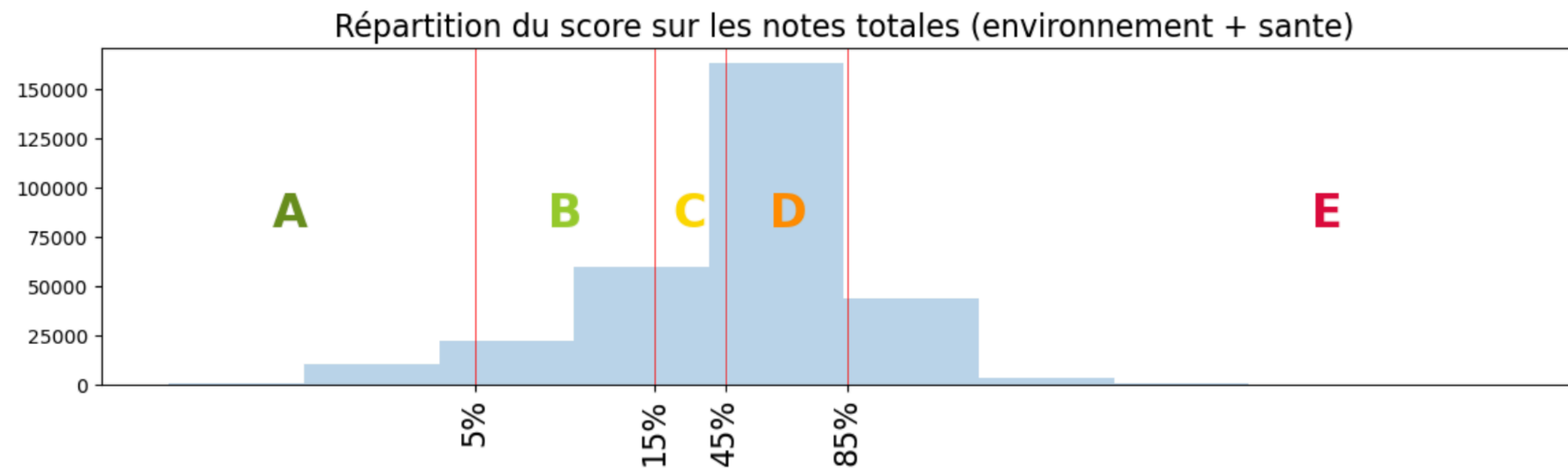
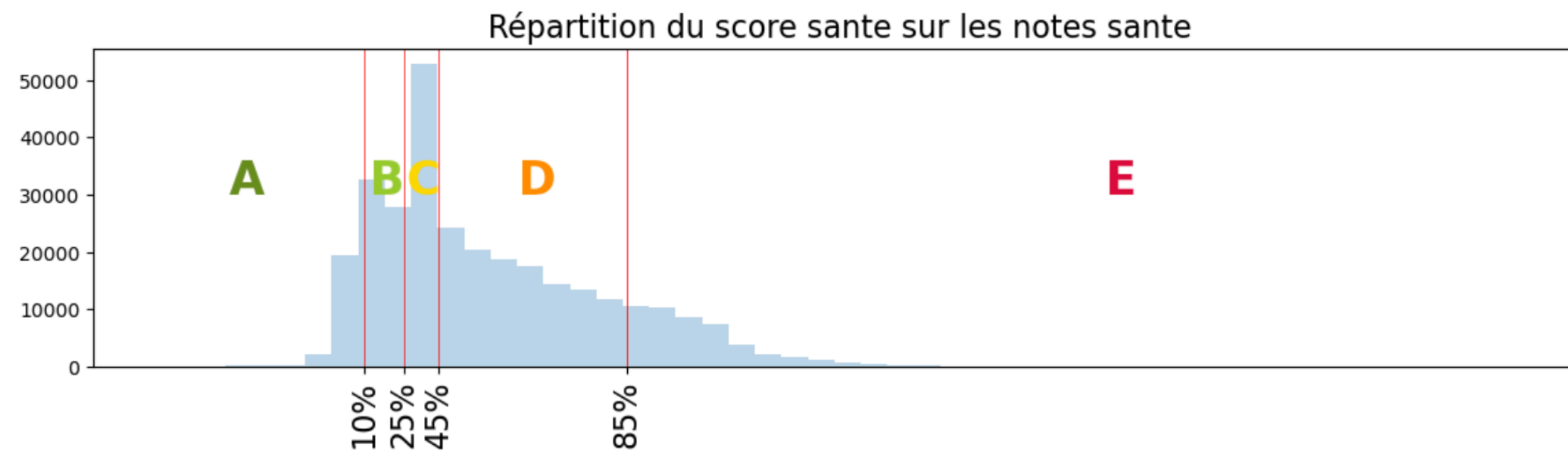
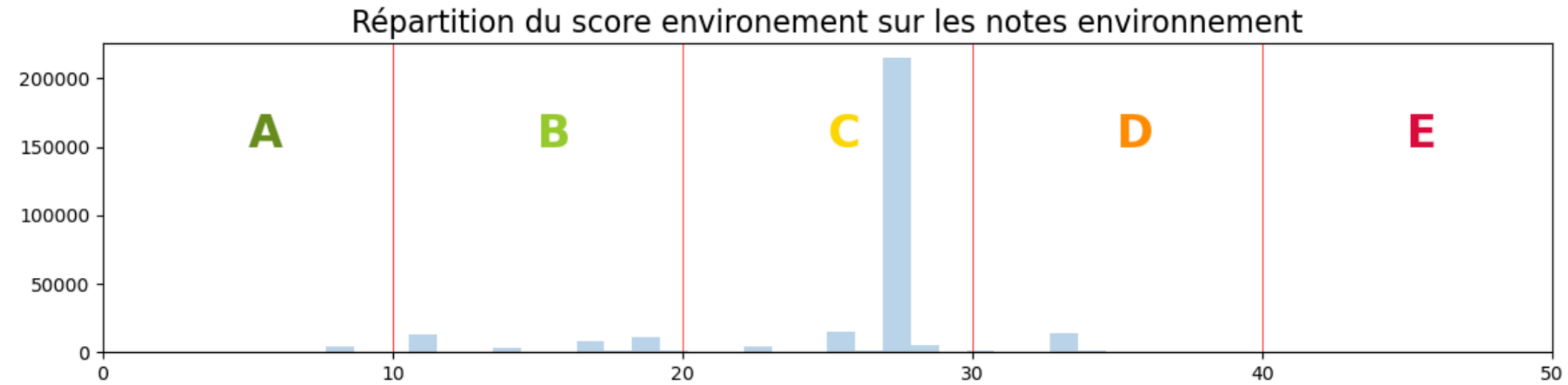
Note total

0 → 100

Note santé

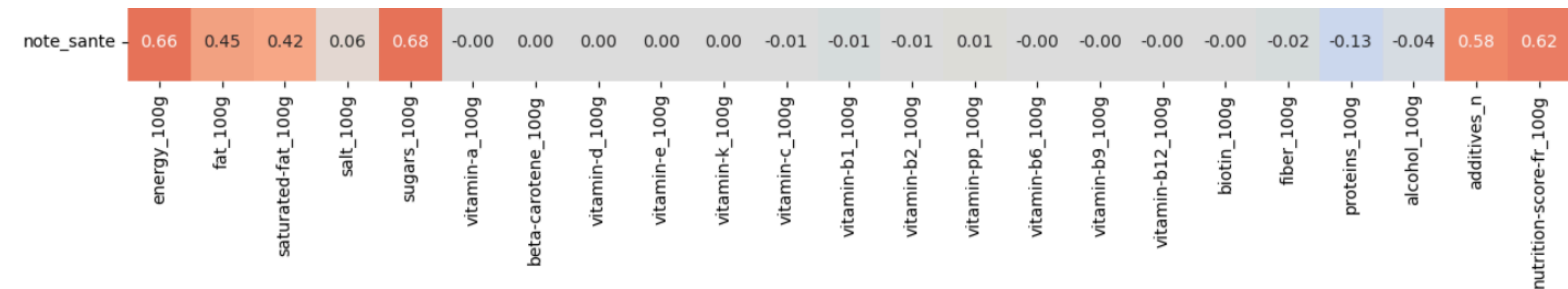
Note environnement

4-Construction des scores

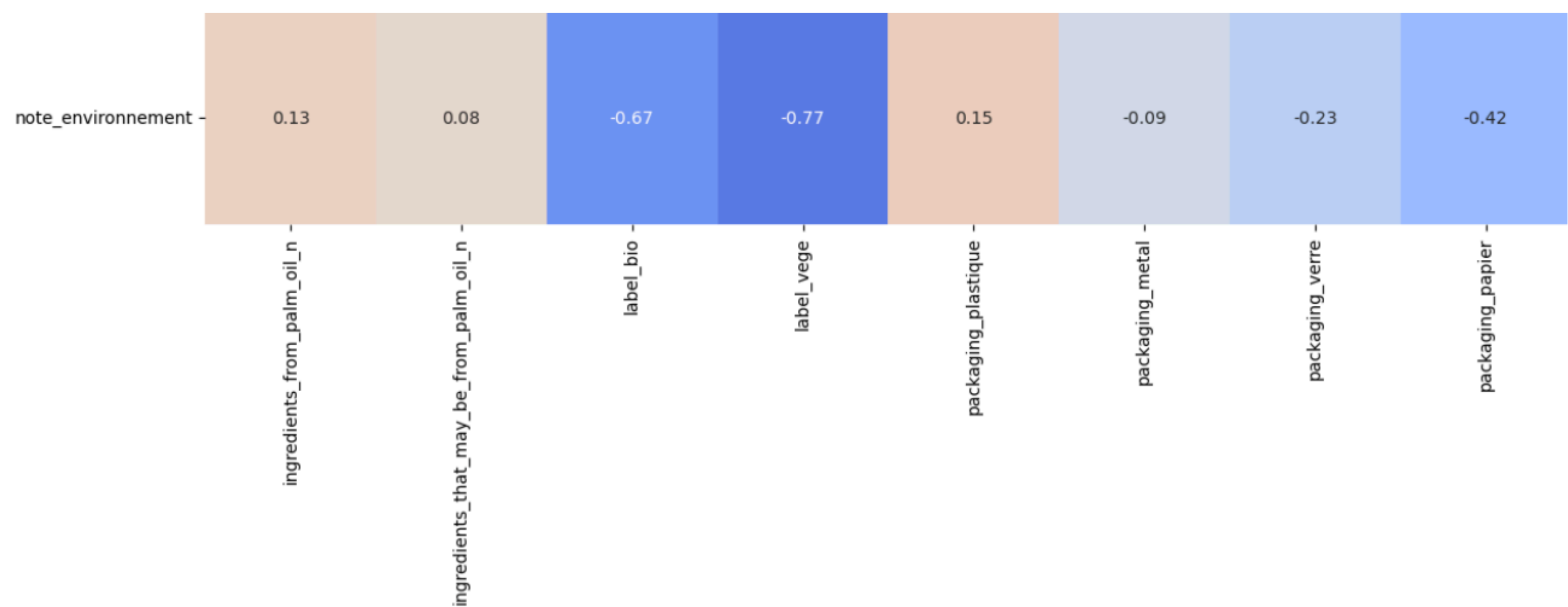


5-Influence des variables sur le score

Variables de construction

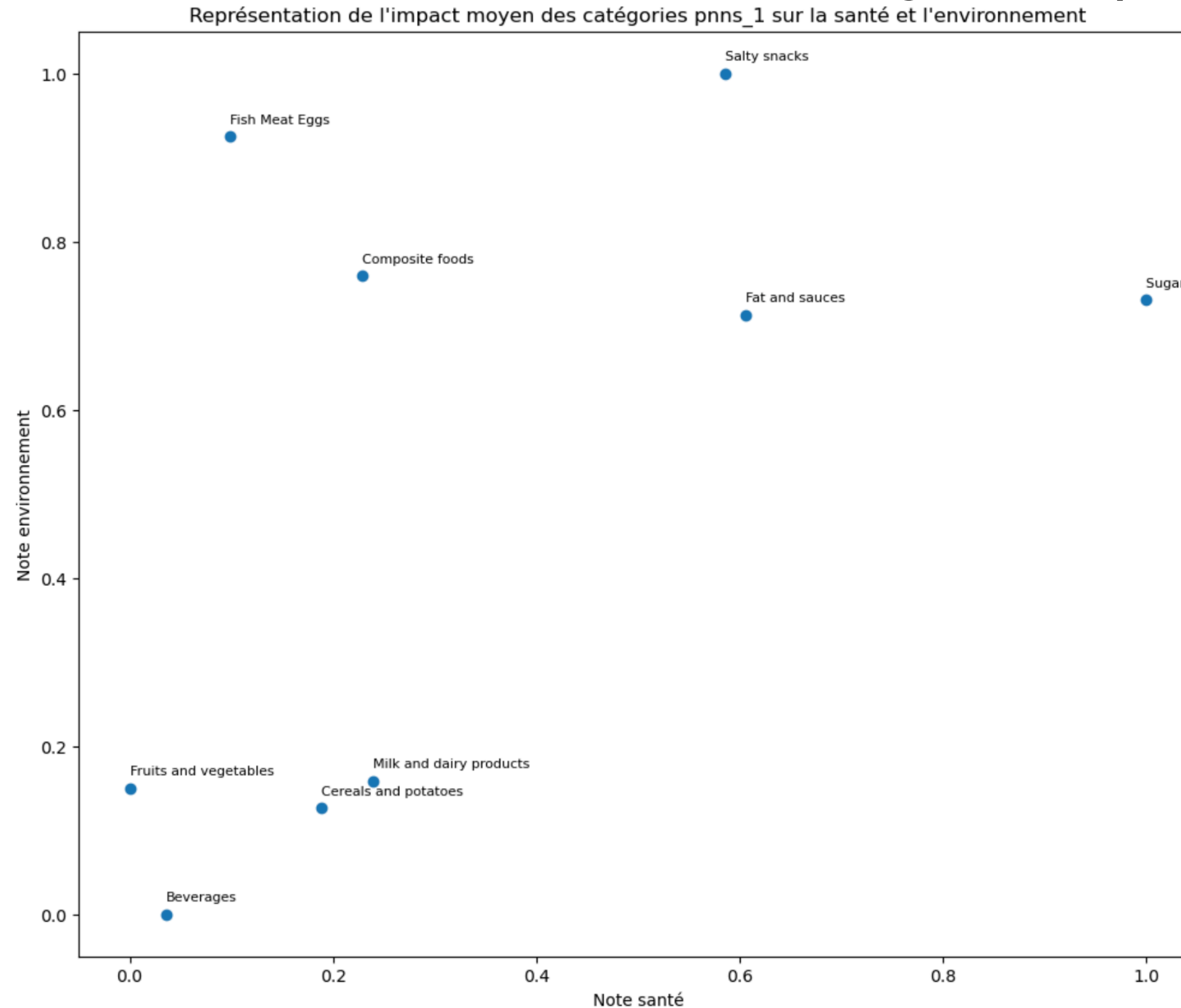


Régression linéaire
sur le score santé :
score de 97,87 %



5-Influence des variables sur le score

Catégories de produit



ANOVA pour le score santé :

$F = 5190,38$

$p = 0$

Eta carré = 13,35 %

ANOVA pour le score environnement :

$F = 11459,34$

$p = 0$

Eta carré = 25,39 %

F → différence entre les moyennes des catégories

p → confirmation de la différence de ces moyennes

Eta carré → explication partielle de la variation des scores

6- Conclusion

Idée d'application viable

- Construction de scores réalisable à partir des variables du jeu de données
- Corrélation importante entre le score santé et le nutriscore
- Validation de l'influence des catégories de produits sur la santé et l'environnement

Perspectives d'amélioration :

- Affiner les constructions des scores.
- Jeu de données mieux référencé, afin d'éviter les approximations

Question ?