

# Santé publique

France

Rendre les  
données de  
santé plus  
accessibles

---

# Projet

L'agence Santé publique France souhaite rendre les données de santé publique plus accessibles, pour qu'elles soient utilisables par ses agents.



# Sommaire

Démarche de préparation des données

Prototype d'application

# Démarche de préparation des données

Les données en bref

Nettoyage des données

Exploration des données

# Les données en bref

320 772 produits référencés  
162 caractéristiques

Information générale sur le produit  
Tags sur le produit  
Ingrédients composant le produit et ses additifs éventuels  
Informations nutritionnelles

76% de données manquantes

# Nettoyage des données

Données dupliquées

Données manquantes

Données aberrantes

# Données dupliquées

22 lignes dupliquées

La variable code doit être unique car elle représente l'identifiant du produit

On garde la dernière occurrence, en supposant que les données sont plus à jour

# Données manquantes

218 463 produits

Suppression des variables qui ont plus de 50% de valeurs manquantes

Suppression les lignes dont la variable "code" est vide

Suppression des lignes dont la variable "product\_name" (nom du produit) est vide

Suppression des lignes dont "nutrition\_grade\_fr" (nutri-score) est vide

Remplacement par la valeur la plus fréquente pour la variable "brands » lorsqu'elle celle-ci est manquante

Remplacement par le chiffre 0 des nutriments pour lesquels la valeur est manquante

# Les nutriments qui entrent dans le calcul des portions pour 100g

- Matière grasse/Lipide (fat\_100g)
- Glucide (carbohydrates\_100g)
- Fibre alimentaire (fiber\_100g)
- Protéine (protein\_100g)
- Sel (salt\_100g)

VALEURS NUTRITIONNELLES MOYENNES	Pour 100 g
Énergie	645 kJ / 154 kcal
Matières grasses - dont acides gras saturés	9 g 1,5 g
Glucides - dont sucres	2,1 g 0,4 g
Fibres alimentaires	0,5 g
Protéines	16 g
Sel	0,5 g

Oméga 3 : 0,7g / 100g.

# Données aberrantes

Les observations dont le total pour les nutriments de base est supérieur à 100g sont supprimées

Les observations dont le total pour les nutriments de base est égal à 0 sont supprimées

Les lignes contenant des valeurs négatives pour les nutriments sont supprimées

# Données aberrantes

suite

196 863 produits

Énergie pour 100g



Kcal = Protéines \* 4 + Glucides \* 4 +  
Lipides \* 9 + Alcool \* 7

1 Kcal = 4,184 Kj

Les lignes ayant plus de 3769 Kj pour 100g sont supprimées

# Analyse des données

Quelques variables

Relations entre les variables

Hypothèses

# Quelques variables

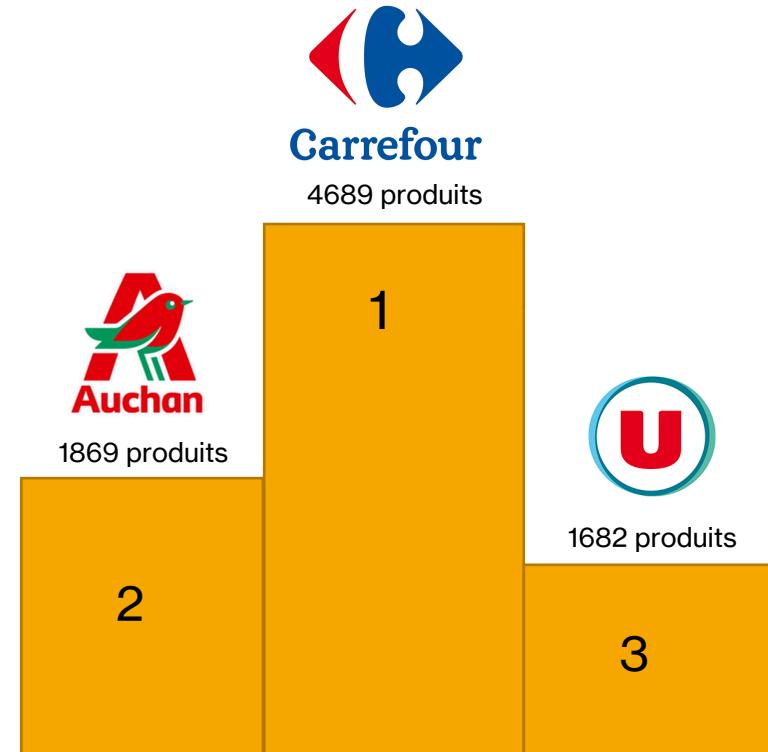
Marques

Énergie

Nutri-score

37 553 marques représentées

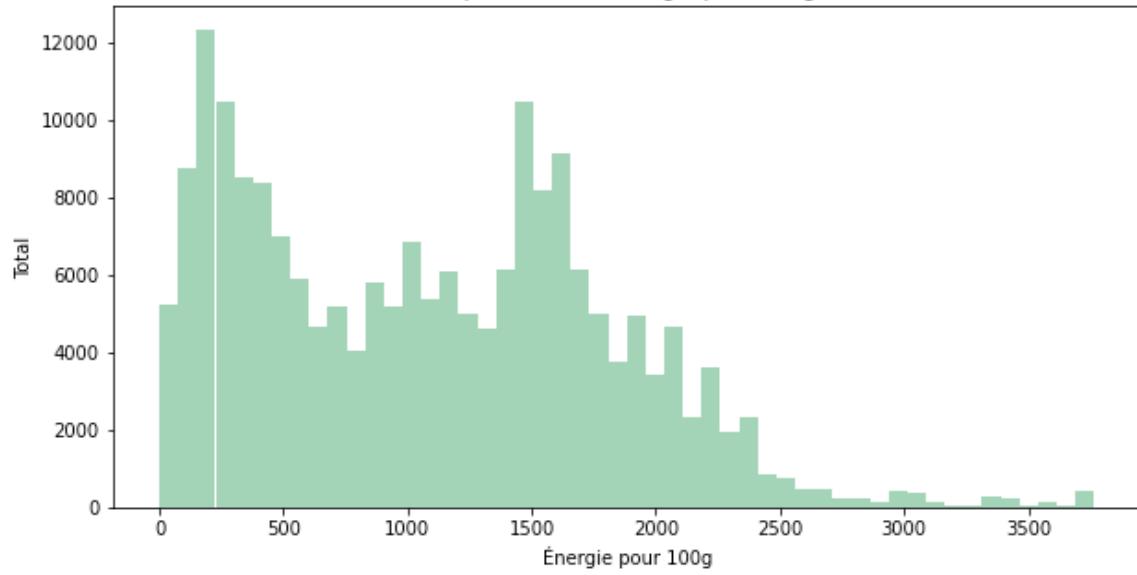
# Marques



## Quelques variables

# Énergie

Répartition de l'énergie pour 100g



**Cola Carrefour pour 100g**

- **Energie: 0kJ/0kcal**
- Matière grasse/Lipides: 0g
- Glucides: 0g
- Fibre alimentaire: 1,70g
- Protéine: 5,70g
- Sel: 0,13g



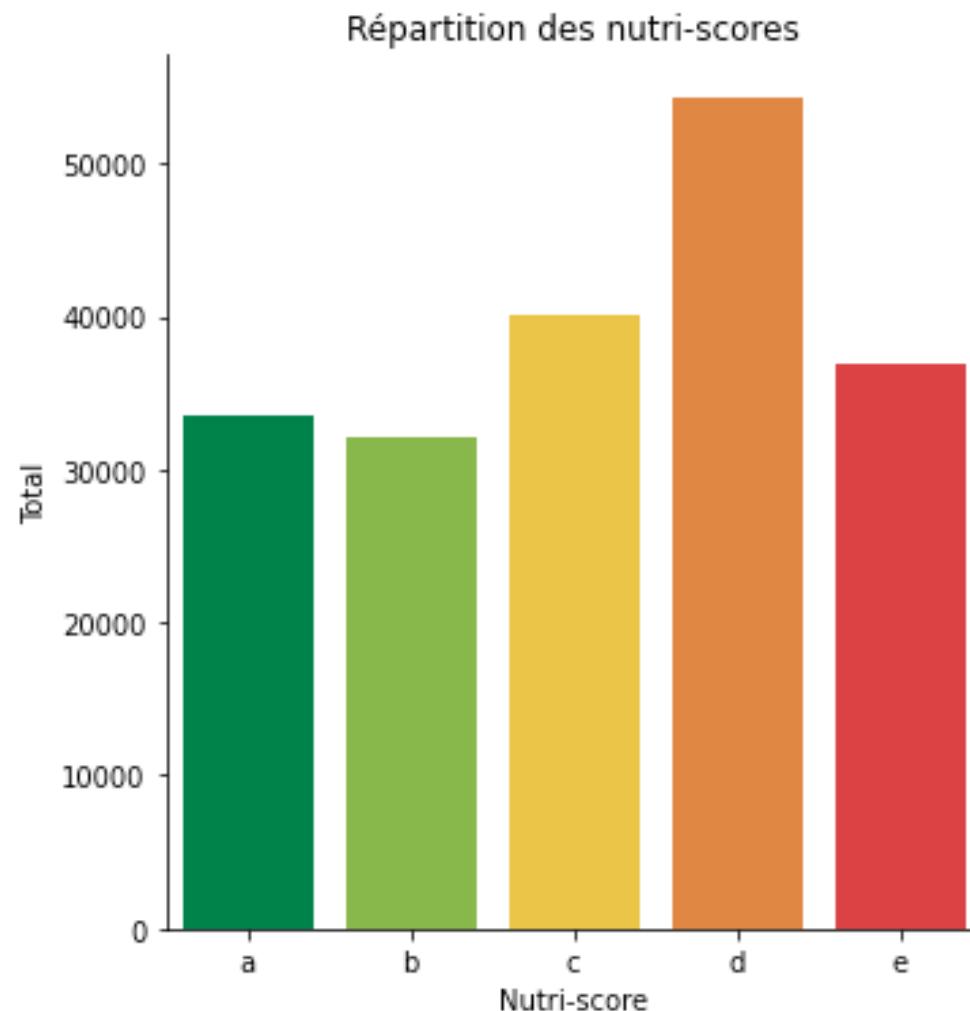
**Huile d'olive vierge douce pour 100g**

- **Energie: 3766kJ/828kcal**
- Matière grasse/Lipides: 92g
- Glucides: 11g
- Fibre alimentaire: 0g
- Protéine: 0g
- Sel: 0g

## Quelques variables

# Nutri-score

Le nutri-score est un système d'étiquetage nutritionnel à cinq niveaux, allant de A à E et du vert au rouge, établi en fonction de la valeur nutritionnelle d'un produit alimentaire.(source: Wikipédia)



# Relations entre les variables

Énergie et lipides

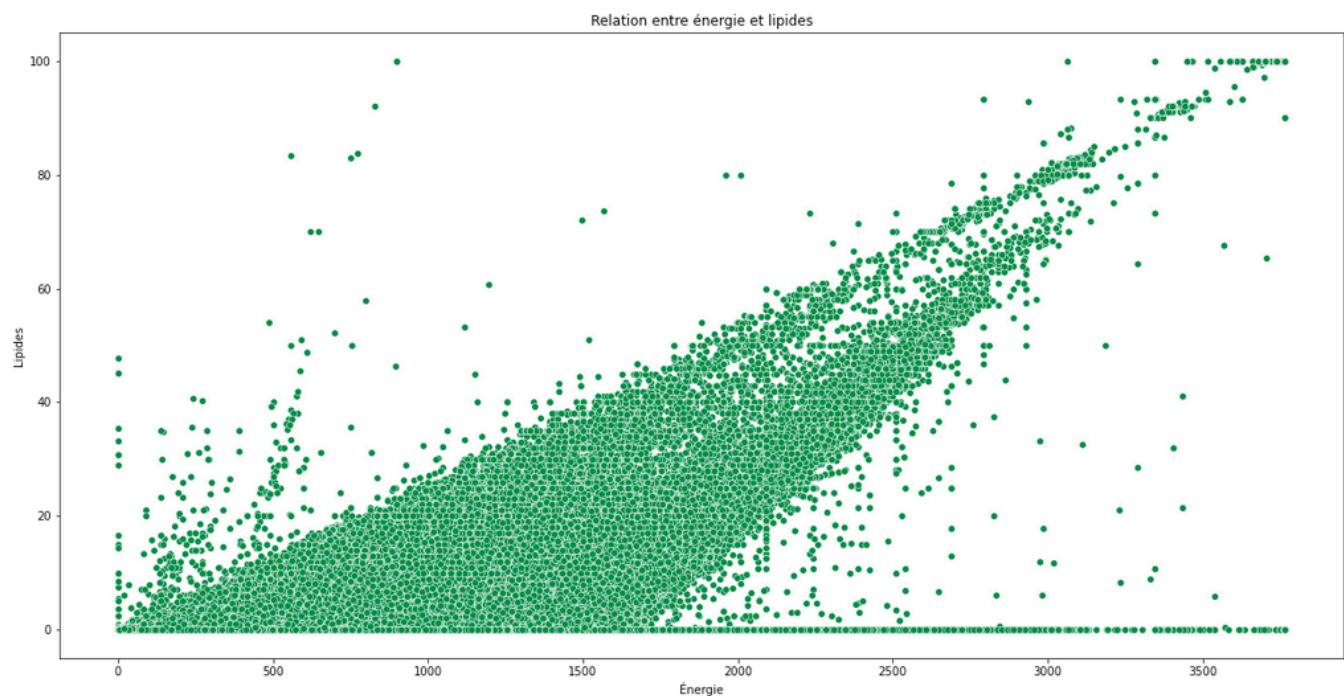
Sel et sodium

Nutri-score et nutriments

Nutri-score et marques

# Énergie et lipides

Les produits qui contiennent beaucoup de lipides sont-ils très énergétiques ?

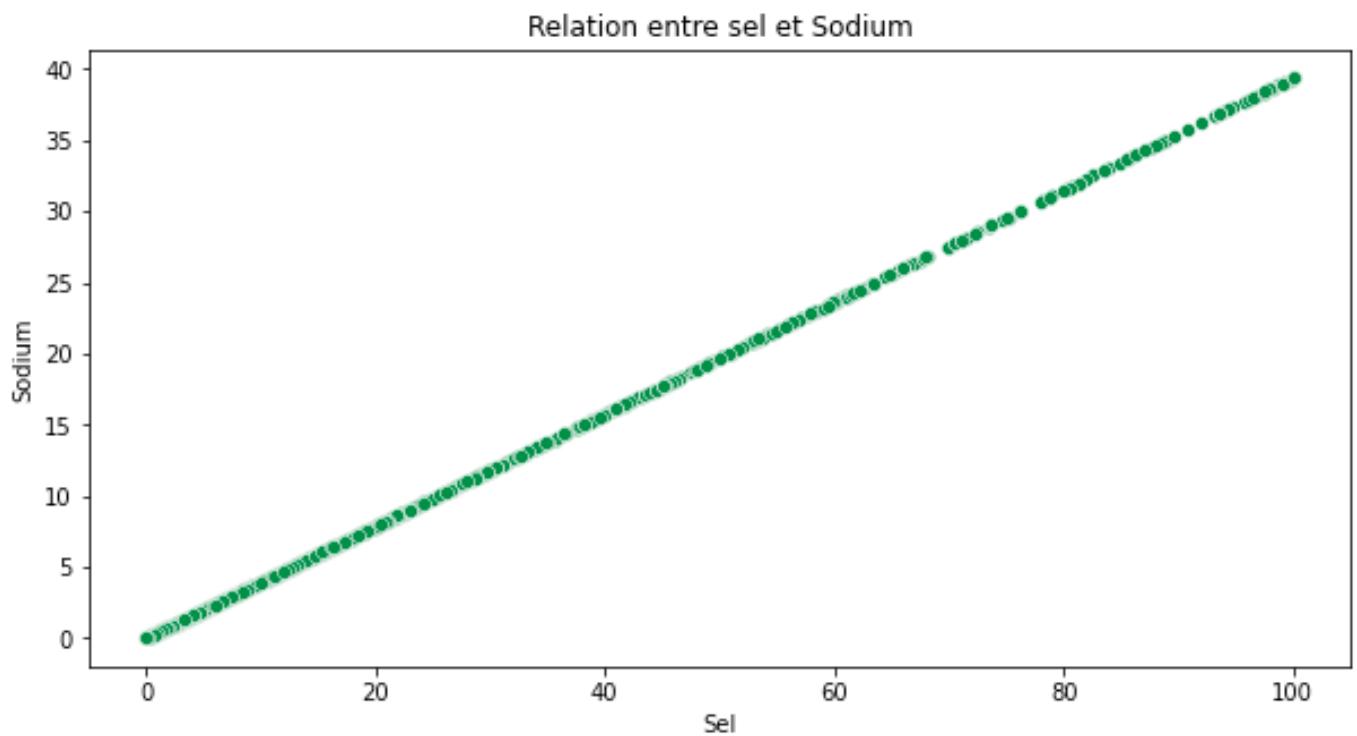


## Relation entre les variables

# Sel et sodium

1g de sel correspond à :

- 400mg de sodium
- 600mg de chlore



Relation entre les variables

# Nutri-score et marques



## Tillamook

- 97,83% de produits avec un nutri-score de E
- 45/46 produits

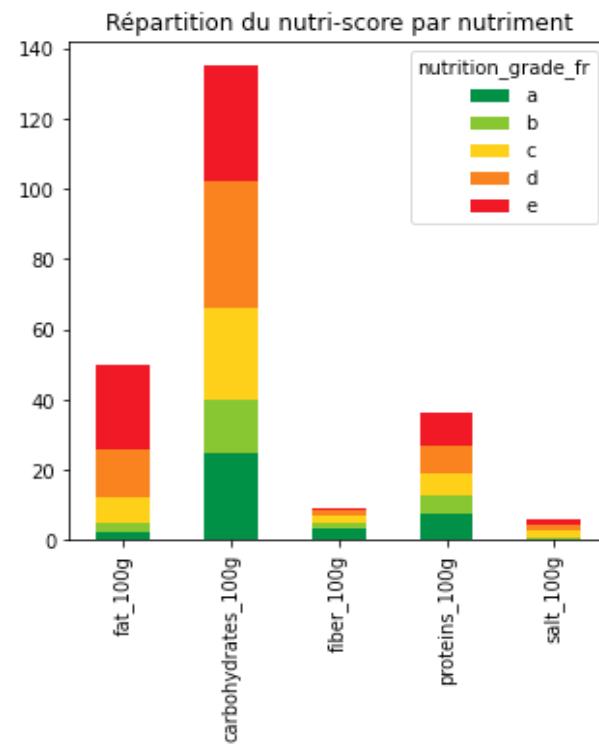


## Mann's

- 96,00% de produits avec un nutri-score de A
- 24/25 produits

Relation entre les variables

# Nutri-score et nutriments



- 2% de lipides,
- 25% de glucides,
- 4% de fibre,
- 8% de protéines,
- 0,3% de sel



- 24% de lipides,
- 33% de glucides,
- 1% de fibre,
- 9% de protéines,
- 1% de sel

# Hypothèses

Notre intuition

Les produits à forte teneur en lipides ont un mauvais nutri-score

Les produits pauvres en fibres ont tendance à avoir un bon nutri-score

Les glucides n'entrent pas en ligne de compte pour le nutri-score

# Hypothèses

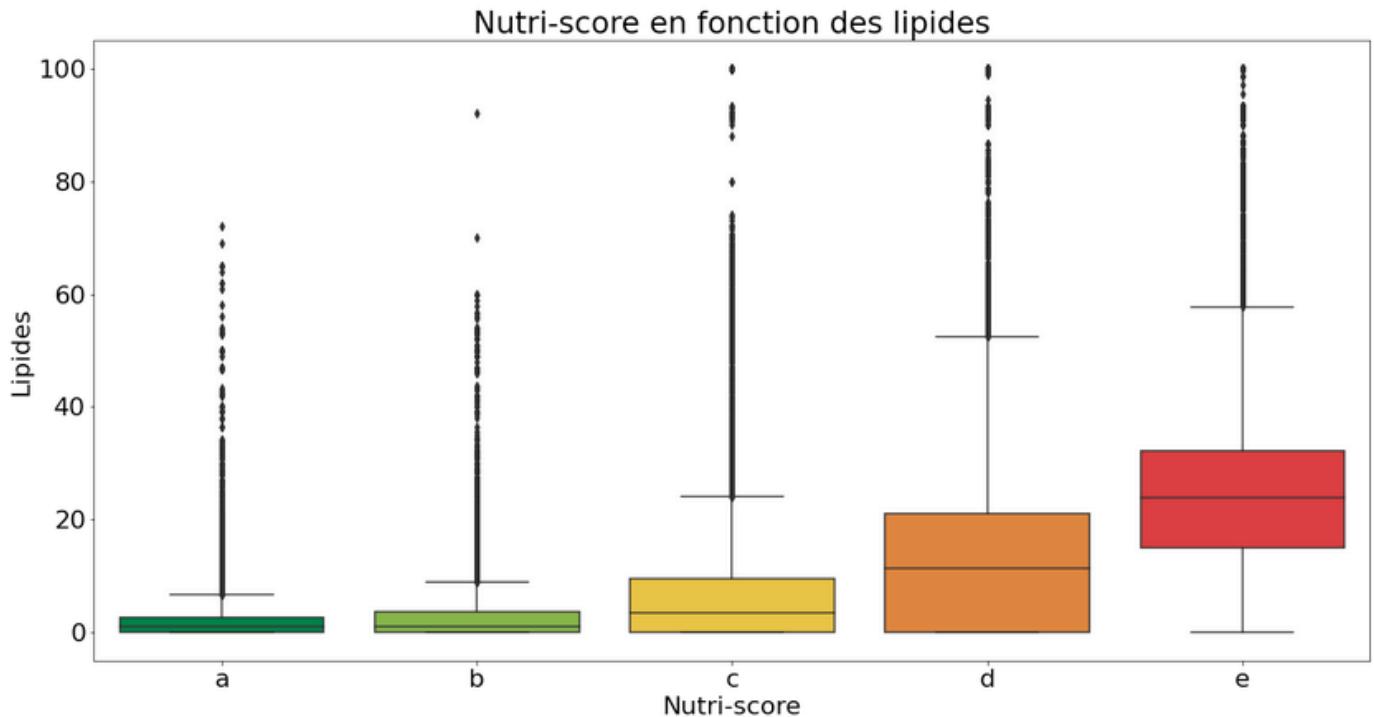
La masse en lipides pour 100g est EGALE pour les produits ayant un bon nutri-score et ceux ayant un mauvais nutri-score

La masse en fibres pour 100g est EGALE pour les produits ayant un bon nutri-score et ceux ayant un mauvais nutri-score

La masse en glucide pour 100g est EGALE pour les produits ayant un bon nutri-score et ceux ayant un mauvais nutri-score

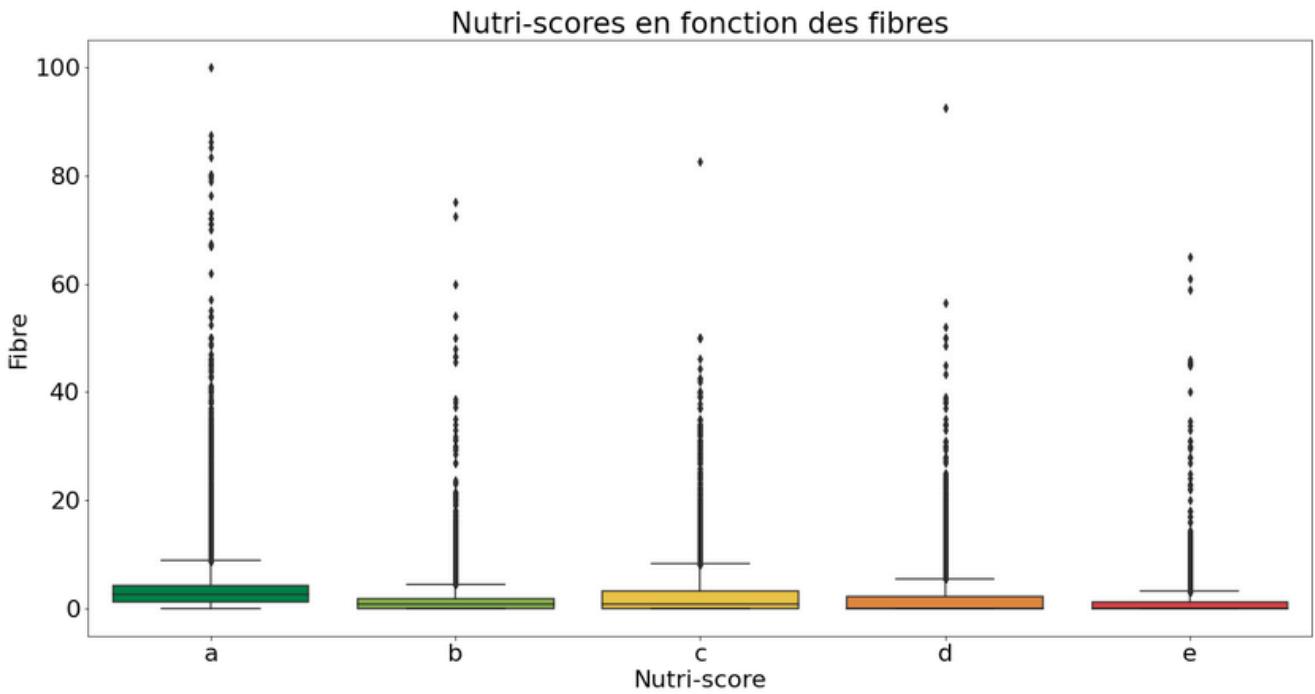
# Hypothèse

La masse en lipides pour 100g est EGALE pour les produits ayant un bon nutri-score et ceux ayant un mauvais nutri-score



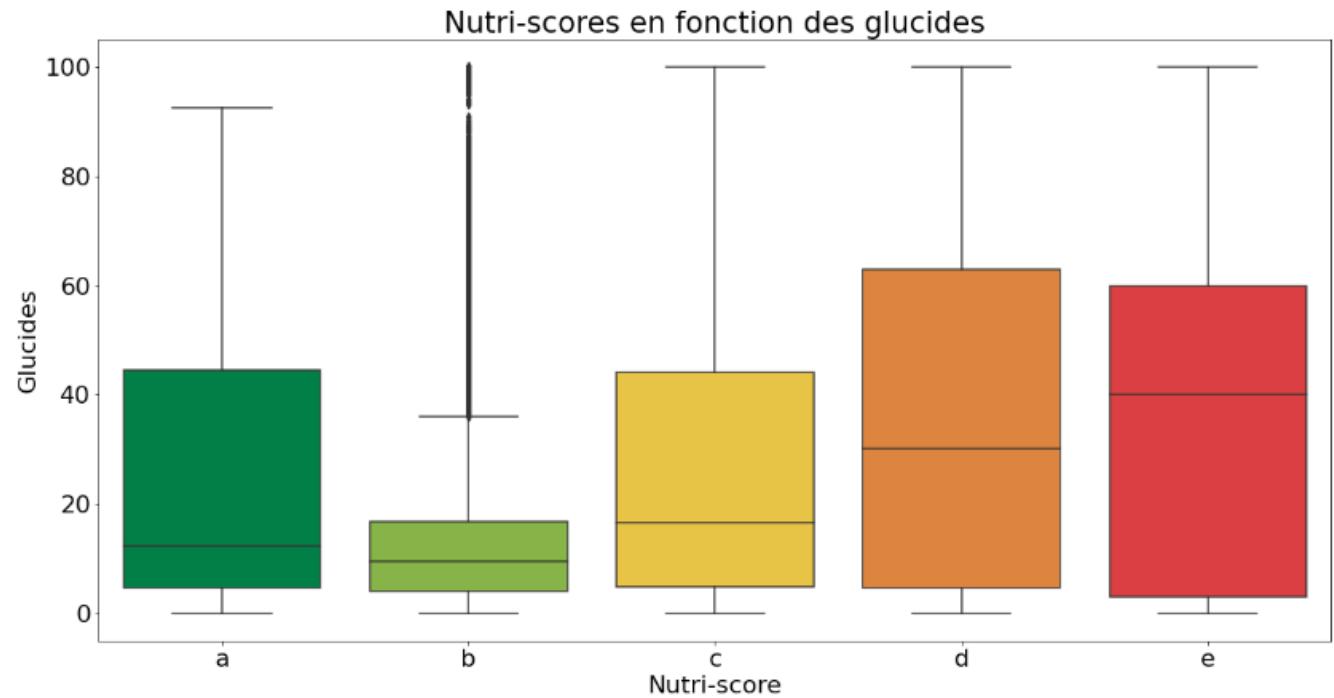
# Hypothèse

La masse en fibres pour 100g est EGALE pour les produits ayant un bon nutriscore et ceux ayant un mauvais nutri-score



# Hypothèse

La masse en glucide pour 100g est EGALE pour les produits ayant un bon nutri-score et ceux ayant un mauvais nutri-score



# Hypothèses

## Conclusions

L'ensemble des trois hypothèses ayant été rejeté, on peut supposer que plus le taux de lipides, de fibres ou de glucides est bas plus le nutri-score sera proche de "A"

# Prototype

Donner une indication de nutri-score en fonction des ingrédients

## Prototype pour la recherche de nutri-score

### Recherche de Nutri-score

Saisissez les Lipides, glucides, fibres, protéines et sel pour 100g afin d'obtenir une évaluation du nutri-score

Lipides	<input type="range"/>	24.50
Glucides	<input type="range"/>	57.50
Fibres	<input type="range"/>	0.00
Protéines	<input type="range"/>	43.50
Sel	<input type="range"/>	54.50

### Resultats



### Quelques produits similaires

Glazed Sour Cream Cake Donut Holes (Kroger, The Kroger Co.)

Double Chocolate & Fudge Love Truffles (Sweet Shop Candies Inc.)

Cinnamon Roll Cookies (Pillsbury)

Chocolate Chip With Candy Cookies (Fareway, Fareway Stores Inc.)

Candy Cookie (Dierberg's Markets Inc.)

# Questions

---

