

**OpenClassrooms**

**Projet P3:**

**Concevez une application au service de la santé publique**

N. Alves

# Concevez une application au service de la santé publique

## Sommaire

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Conclusion et perspectives

# Concevez une application au service de la santé publique

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Conclusion et perspectives

# Concevez une application au service de la santé publique

## Contexte

Le NUTRI-SCORE:

Le nutriscore est un indicateur permettant de comparer la qualité nutritionnelle d'un produit avec un autre. Les comparaisons peuvent se faire entre aliments d'une même catégorie. Il s'inscrit dans un cadre de repères alimentaires plus large comme le programme national nutrition santé <sup>(1,2)</sup>.

Le score de A à E est calculé en fonction des éléments à favoriser (fibres, protéines, fruits et légumes, légumes secs...) et ceux à limiter (calories, acides gras saturés, sucres, sel)<sup>(3)</sup>. Il ne tient pas compte des additifs, des pesticides et du niveau de transformations des produits.

(1) Manger Bouger. Programme National Nutrition Santé. Objectif et Principes Généraux

(2) Manger Bouger. Programme National Nutrition Santé. Les objectifs de santé publique

(3) Santé publique France – Règlement d'usage du Logo « Nutri-Score » . Calcul du nutriscore p18.



# Concevez une application au service de la santé publique

## Contexte

Le NUTRI-SCORE:

Le nutriscore est un indicateur permettant de comparer la qualité nutritionnelle d'un produit avec un autre. Les comparaisons peuvent se faire entre aliments d'une même catégorie. Il s'inscrit dans un cadre de repères alimentaires plus large comme le programme national nutrition santé <sup>(1,2)</sup>.

Le score de A à E est calculé en fonction des éléments à favoriser (fibres, protéines, fruits et légumes, légumes secs...) et ceux à limiter (calories, acides gras saturés, sucres, sel)<sup>(3)</sup>. Il ne tient pas compte des additifs, des pesticides et du niveau de transformations des produits.

(1) Manger Bouger. Programme National Nutrition Santé. Objectif et Principes Généraux

(2) Manger Bouger. Programme National Nutrition Santé. Les objectifs de santé publique

(3) Santé publique France – Règlement d'usage du Logo « Nutri-Score » . Calcul du nutriscore p18.



# Concevez une application au service de la santé publique

## Contexte

Certains produits, malgré une notation nutritionnelle favorable, contiennent des additifs en grands nombres. Les additifs sont utilisés pour leurs propriétés d'émulsifiants, de colorants, de conservateurs etc.

Les effets sur la santé sont peu connus, la nature délétère de certains est suspectée.<sup>(4)</sup>

### ⇒ Idée d'application :

Fournir une nouvelle information qualitative pour un produit donné : un indicateur agrégeant la notion d'additif et de nutriscore.

L'objectif sera de limiter le nombre d'additifs tout en améliorant la qualité nutritionnelle des produits transformés.



[Détail de l'analyse des ingrédients »](#)

#### Additifs :

- E331 - Citrates de sodium
- E300 - Acide ascorbique
- E316 - Érythorbate de sodium
- E392 - Extrait de romarin
- E14XX - Amidons modifiés
- E336 - Tartrates de potassium
- E500 - Carbonates de sodium
- E1100 - Amylases

<sup>(4)</sup> INSERM. Nutrition et santé. La santé passe par l'assiette et l'activité physique.

# Concevez une application au service de la santé publique

- Contexte
- **Présentation des informations disponibles**
- Étude du jeu de données
- Analyse
- Conclusion et perspectives

# Concevez une application au service de la santé publique

## Présentation des informations disponibles

Open Food Facts met à disposition une base de données de produits alimentaires.

- Elle répertorie une quantité importante de produits dans le monde : France , États-unis, etc.
- Ces produits possèdent des caractéristiques diverses (186 au total),
- Ces caractéristiques sont classées en cinq thèmes (informations générales, tags, ingrédients, divers, informations nutritionnelles),

Une recherche par mot-clé est disponible.



Open Food Facts répertorie les produits alimentaires du monde entier.



# Concevez une application au service de la santé publique

## Présentation des informations disponibles

Cette base est publique et peut fournir les informations par différents moyens dont le téléchargement d'un fichier au format csv.

Ce fichier comporte un volume important d'informations :

- 1,8 millions de lignes, 186 colonnes,
- 4,1 Go



Open Food Facts répertorie les produits alimentaires du monde entier.

# Concevez une application au service de la santé publique

- Contexte
- Présentation des informations disponibles
- **Étude du jeu de données**
- Analyse
- Conclusion et perspectives

# Concevez une application au service de la santé publique

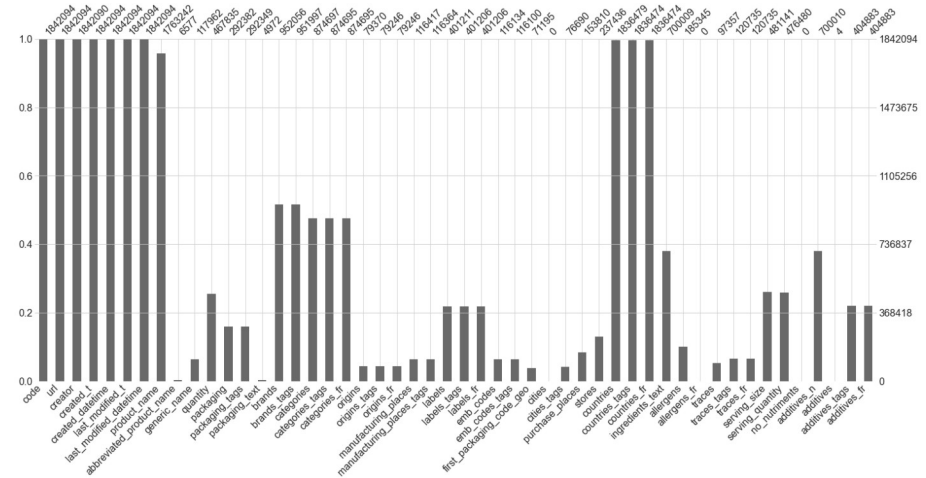
## Étude du jeu de données

Sélection d'informations :

Le jeu de données comporte beaucoup d'informations manquantes :

- Plus de 50 % des colonnes sont remplies à moins de 1 %.

On procède à une sélection de colonnes et de lignes pour limiter le volume d'information.



Taux de remplissage par colonne du fichier

# Concevez une application au service de la santé publique

## Étude du jeu de données

### Sélection de colonnes :

- Sélection de colonnes remplies à au moins 1 %.
- Sélection de colonnes dans chacun des cinq thèmes<sup>(\*)</sup>.

Ex : nom du produit, label éventuel associé, additif, nutri-score, eco-score, informations nutritionnelles etc.

### Sélection de lignes :

- Après un renommage des catégories du PNNS<sup>(\*\*)</sup>, on récupère uniquement les produits des catégories connues.
- Puis on sélectionne les produits pour lesquels une information nutritionnelle (l'énergie) est renseignée.

<sup>(\*)</sup> Informations générales, Tags, Ingrédients, Divers, Informations nutritionnelles

<sup>(\*\*)</sup> Programme National Nutrition Santé

# Concevez une application au service de la santé publique

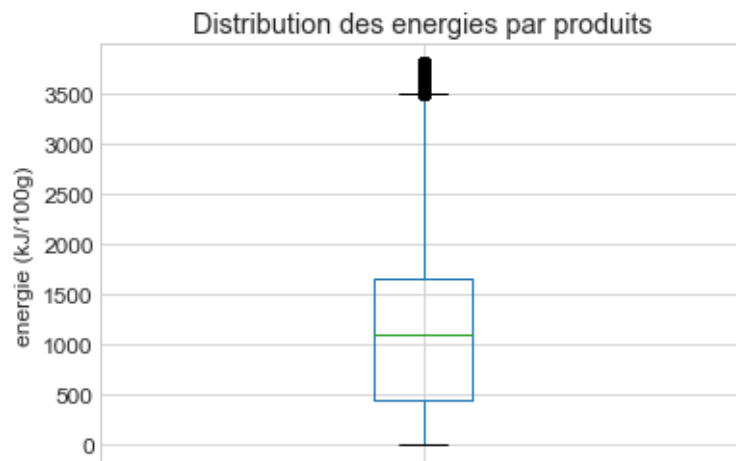
## Étude du jeu de données

### Valeurs aberrantes :

Le traitement des valeurs aberrantes est réalisé sur l'énergie et la masse.

- L'énergie pour 100g est limitée à 3700 kJ.
- Le masse de certains nutriments pour 100 g (lipide, glucide, protéine, sel) est limitée à 100g
- La masse totale pour 100g est limitée à 100g

Les produits en dehors de ce cadre ne sont pas pris en compte.



⇒ Après traitement, le jeu de donnée comporte 600 000 lignes et 44 colonnes.

La taille du fichier est de 400 Mo.

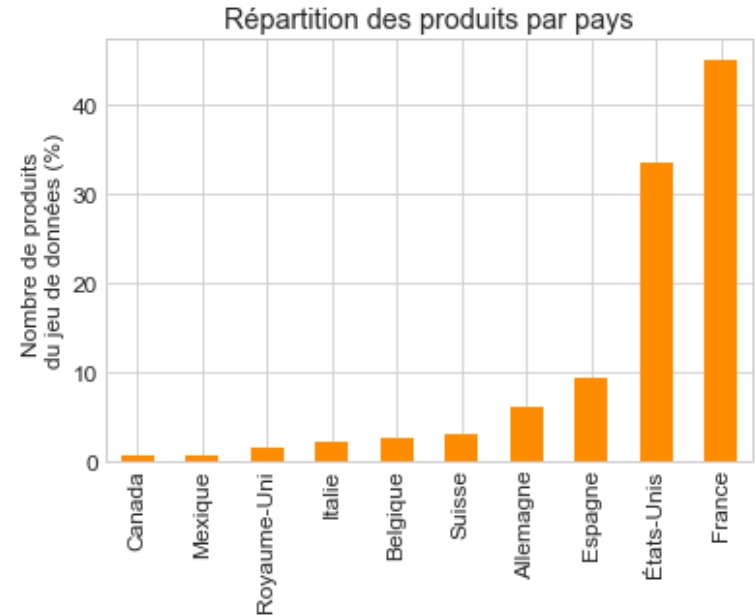
# Concevez une application au service de la santé publique

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- **Analyse**
- Conclusion et perspectives

# Concevez une application au service de la santé publique

## Analyse

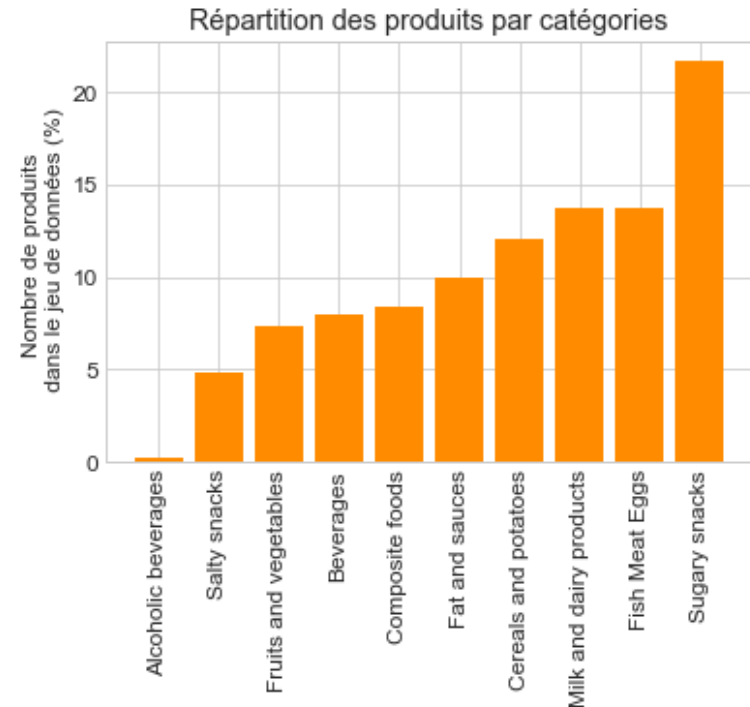
Les produits sont essentiellement disponibles en France et au États-Unis, ils représentent 80 % du jeu de données.



# Concevez une application au service de la santé publique

## Analyse

Les snacks sucrés (biscuit, brioche, barre chocolat etc.) représentent plus de 20 % du jeu de données. C'est la catégorie la plus présente.

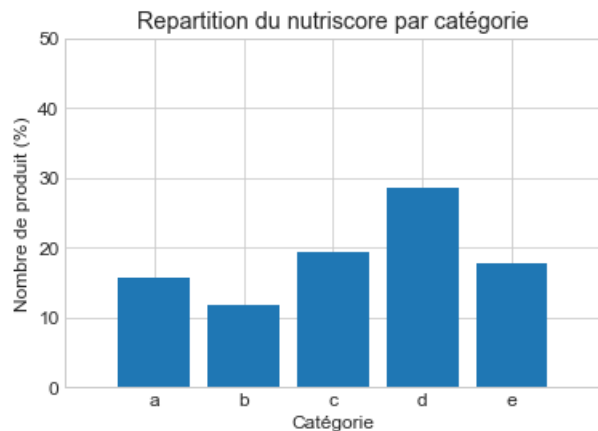




# Concevez une application au service de la santé publique

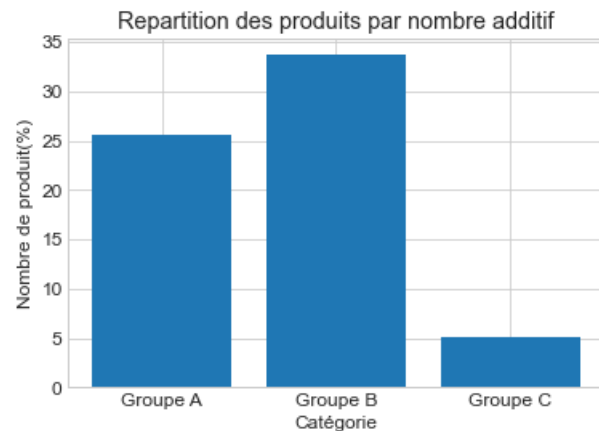
## Analyse

### Nutri-score :



⇒ Le nutri-score se répartit sur des catégories de A à E. La catégorie D est la plus présente. Le nutri-score est renseigné à plus de 90% dans la base.

### Additifs : répartition catégorielle



On regroupe les produits suivant trois catégories :  
Groupe A (0 additif)  
Groupe B (1 à 5 additifs)  
Groupe C (6+ additifs)  
⇒ Le groupe B est le plus présent

# Concevez une application au service de la santé publique

## Analyse

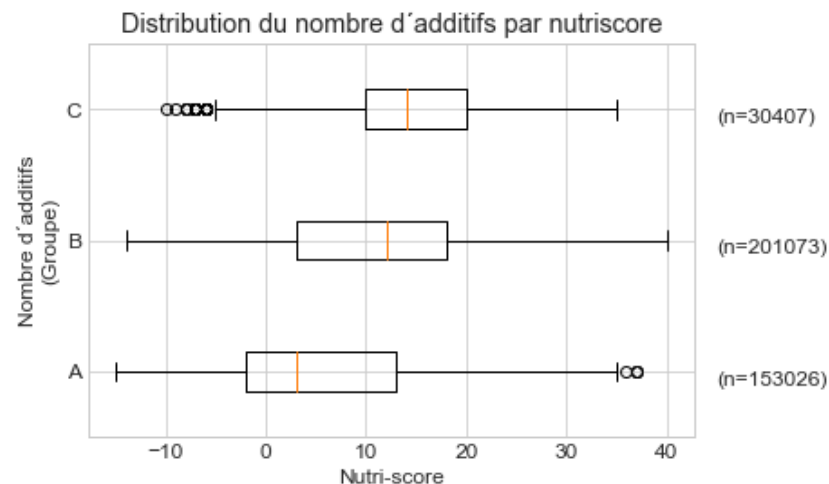
### Nutri-score et additifs:

On cherche à observer l'impact du nombre d'additifs sur le nutri-score.

Pour cela, on regroupe les produits par catégorie de nombre d'additifs : A , B , C

Additif	Moyenne Nutri-score	Écart-type Nutri-score	Nombre de produits
Groupe A (0 additif)	5.31	8.83	153026
Groupe B (1 à 5 additifs)	10,82	8.82	201073
Groupe C (6+ additifs)	13.75	7.57	30407

Le nombre d'individus se répartit entre 30000 et 200000 suivant la catégorie.



# Concevez une application au service de la santé publique

## Analyse

### Nutri-score et additifs:

Analyse de la variance:

- Premier calcul :  
La valeur p est nulle,  $H_0$  est rejeté<sup>(\*)</sup>.  
⇒ le groupe d'additif a un effet significatif sur le nutri-score.
- Second calcul :  
On a utilisé un échantillon où la catégorie 'Sugary snacks' est moins présente, le rapport de corrélation  $\eta^2$  se trouve diminué (i.e. la taille de l'effet).  
⇒ -8 % de produits 'sugary snacks' diminue le rapport de corrélation de -10% .

### Test statistique

Jeu de donnée	$\eta^2$	F	Valeur p
Premier calcul : Jeu Initial	0.1047	22491	0
Second calcul : Jeu échantillonné	0.0939	18271	0

<sup>(\*)</sup> Hypothèses:

$H_0$ : Il est crédible de penser que les groupes suivent la même distribution.

$H_1$ : Au moins une des distributions a une moyenne qui s'écarte des autres

# Concevez une application au service de la santé publique

## Analyse

Nutri-score et additifs :  
Catégorie spécifique de produit 'One dish meal'.

Les produits sont les plus présents:

- pour les additifs : entre 1 à 5 additifs
- pour le nutriscore : entre 0 et 10

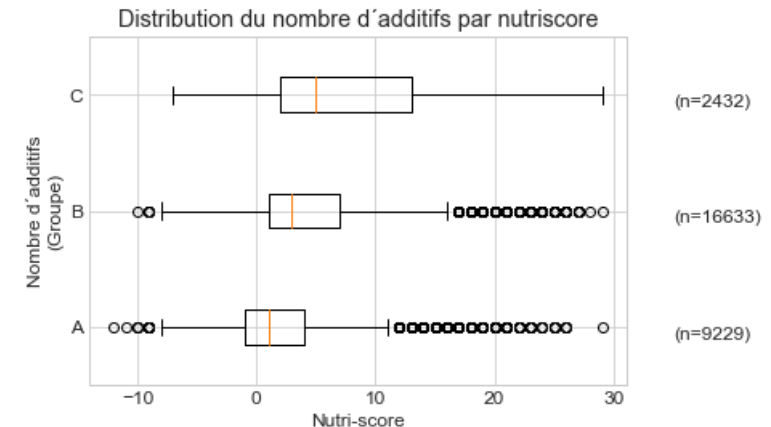
Analyse de la variance : la valeur p est proche de zéro.

Le groupe d'additif a un effet significatif sur le nutri-score.

⇒ Il est utile de construire un nouvel indicateur permettant de discriminer les additifs du nutriscore

Test statistique

Jeu de donnée	$\eta^2$	F	Valeur p
'One-dish meal'	0,050	741	$2.10^{-314}$



# Concevez une application au service de la santé publique

## Analyse

### Catégorie de produit :

Il s'agit ici d'observer les associations entre :

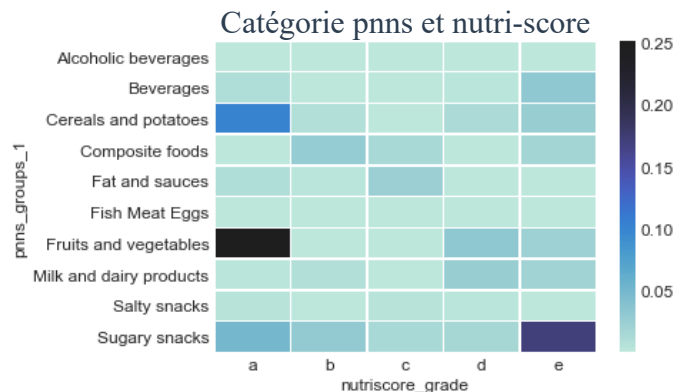
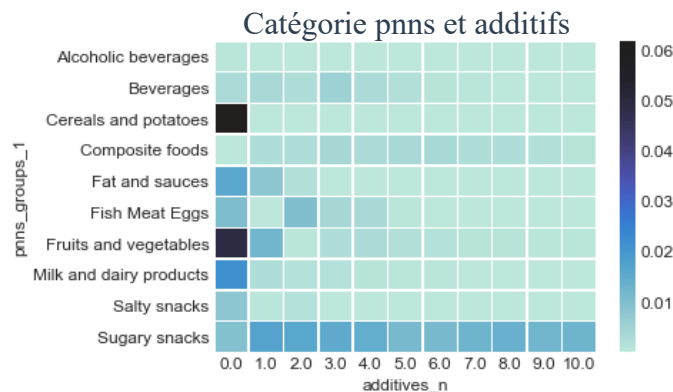
- La catégorie de produit et le nombre d'additifs
- La catégorie de produit et le nutriscore

⇒ Produits céréaliers, fruits et légumes :

Il s'agit des catégories où sont le moins présents les additifs. Ce sont aussi les produits avec des nutriscore le plus souvent favorables.

⇒ Snacks sucrés :

Les additifs sont souvent présents. Les nutriscore E et A/B sont les plus présents pour cette catégorie



# Concevez une application au service de la santé publique

## Analyse

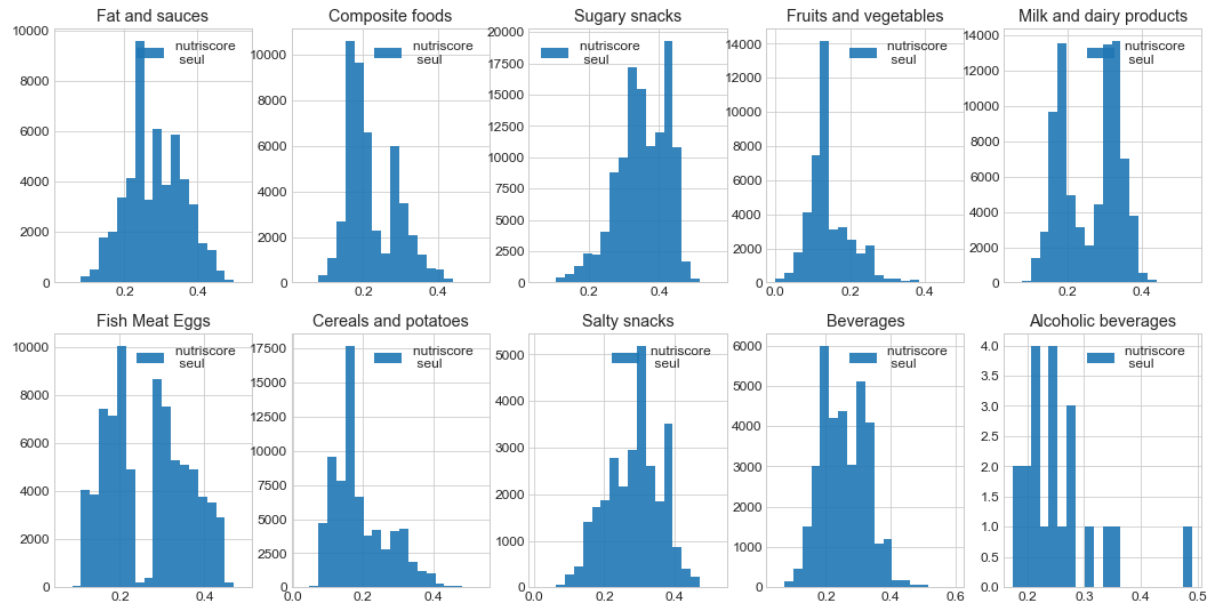
### Nutri-score et catégories PNNS

Un score proche de zéro est le plus favorable à la santé.

Il y a des distributions avec deux pics :

- Catégorie 'Milk and dairy product', 'Fish meat eggs'.  
Les pics sont dus aux produits bruts et transformés (ex: viande/charcuterie, yaourt/crème dessert).
- La catégorie 'composite food' a une distribution avec deux pics aussi, il y a beaucoup de plats avec nutriscore b.

### Distribution du nutri-score par catégorie PNNS



# Concevez une application au service de la santé publique

## Analyse

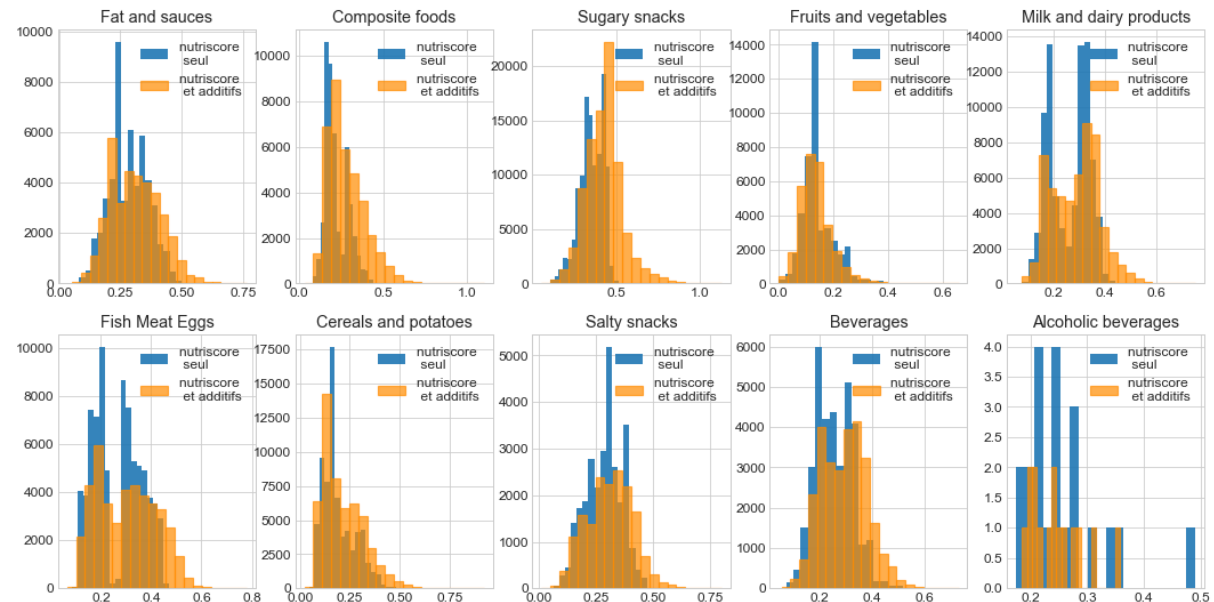
### Intégration des additifs dans le nutri-score :

Construction d'un nouveau score :

Score = 60 % nutri-score + 40 % indicateur additifs.

- Les produits avec beaucoup d'additifs affichent un score moins favorable à la santé.
- La prise en compte des additifs dégrade le score et atténue les pics.  
Ex : les produits transformés avec additifs se retrouvent moins bien classés ('salty/sugary snacks', 'composite food', etc.) .
- Les fruits, légumes et les produits céréaliers sont toujours très bien placés.

### Distribution du nutri-score et des additifs par catégorie PNNS



# Concevez une application au service de la santé publique

- Contexte
- Présentation des informations disponibles
- Étude du jeu de données
- Analyse
- Conclusion et perspectives



# Concevez une application au service de la santé publique

## Conclusion et perspectives

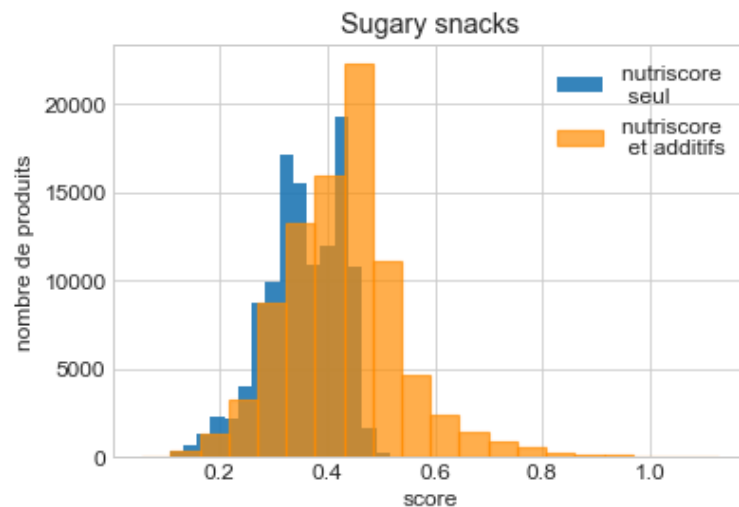
Le jeu de données fourni par Open Food Facts permet de créer une nouvelle information: un indicateur intégrant la présence d'additif au nutri-score. Par mesure de précaution, il met plus en avant les produits avec peu d'additifs.

Une limite de cet exercice est l'absence des prix, il s'agit d'une étude spécifiquement qualitative.

D'autres informations qualitatives pourront être intégrées par la suite:

- Des repères nutritionnelles : quantité et fréquence de consommation à privilégier pour un produit<sup>(5)</sup>, repères du pnns, etc.
- L'empreinte écologique : l'Eco-score

(5) Haut conseil de la Santé publique. Révision des repères alimentaires pour les adultes



# Annexe

## Référence :

1 Manger Bouger. Programme National Nutrition Santé. Objectif et Principes Généraux.

<https://www.mangerbouger.fr/PNNS/Le-PNNS/Qu-est-ce-que-le-PNNS>

2 Manger Bouger. Programme National Nutrition Santé. Les objectifs de santé publique.

<https://www.mangerbouger.fr/PNNS/Le-PNNS/Les-objectifs-de-sante-publique>

3 Santé publique France – Règlement d’usage du Logo « Nutri-Score » . Calcul du nutriscore p18.

<https://www.santepubliquefrance.fr/media/files/02-determinants-de-sante/nutrition-et-activite-physique/nutri-score/reglement-usage>

4 INSERM. Nutrition et santé. La santé passe par l’assiette et l’activité physique.

<https://www.inserm.fr/information-en-sante/dossiers-information/nutrition-et-sante>

5 Haut conseil de la Santé publique. Révision des repères alimentaires pour les adultes.

[https://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=hcspa20170216\\_reperesalimentairesactua2017.pdf](https://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=hcspa20170216_reperesalimentairesactua2017.pdf)



**Avez vous des questions ?**