

Rapport d'exploration

— Projet pour Lamarmite —

Par Xavier Montamat

Problématique

L'objectif de ce projet est de présenter une liste de produits permettant de réaliser des recettes de cuisine saines.

Ce à partir d'une liste importantes de produits et de leur composition nutritionnelle, provenant du site openfoodfacts.org

Axe d'approche choisi

Catégoriser les produits en sous listes

Identifier les produits contenant de bons ingrédients

Identifier les composants de mauvais ingrédients

Croiser les informations

Ordonner les listes

Classement

Proposer les meilleurs aliments possibles

Autres axes possibles

Un unique score global

Inconvénients :

- Peu de variété, produits 'à tout faire' en haut
- Exclusion des produits spécialisés
- Nutriscore?

Analyse des ingrédients

Inconvénients :

- Plus complexe
- Ingrédients ne sont pas toujours indiqués

Plan de l'analyse nutritionnelle

Nettoyage

- Données initiales
- Actions menées
- Solutions choisies

Analyse des données

- Le Nutriscore
- Analyse des graisses
- Analyse des sucres
- Sels et additifs
- Croisement et affinage des listes
- Choix d'un ordonnancement

Propositions

- Exploration des résultats
- Conclusion sur l'analyse
- Axes d'amélioration

Nettoyage des données

Les données

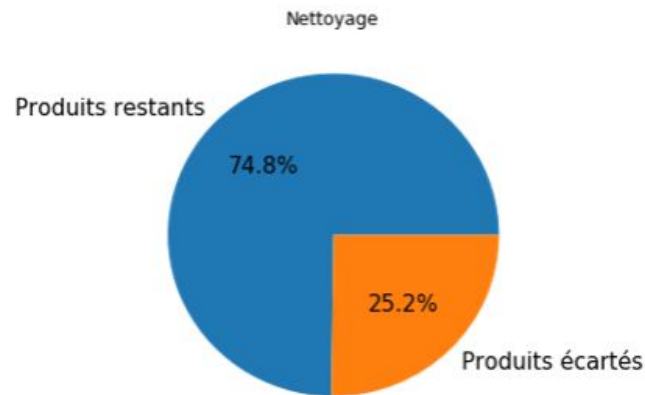
- 320 772 produits et 162 attributs

Filtrer les produits

- 17 762 produits sans nom
- 43 809 produits sans composants
- 798 produits ne sont pas de la nourriture

product_name	categories	categories_tags
shampooing	Open Beauty Facts	en:non-food-products,en:open-beauty-facts

- 18 234 produits dupliqués



Filtrer les attributs

- 16 attributs jamais remplis pour aucun produit
- 41 attributs non utiles pour l'analyse

Remplissage des données

Les données aberrantes

- 243 produits dont attributs en % non compris entre 0 et 100.
 - Supprimés (faible nombre, non fiables)

biotin_100g	pantothenic_acid_100g
-1.000000	-1.000000

Les données manquantes

- Informations nutritionnelles non remplies

Considérées comme absentes (0) . Non listées sur emballages

- Attributs de texte non remplis

Considérés comme vides

Analyse exploratoire

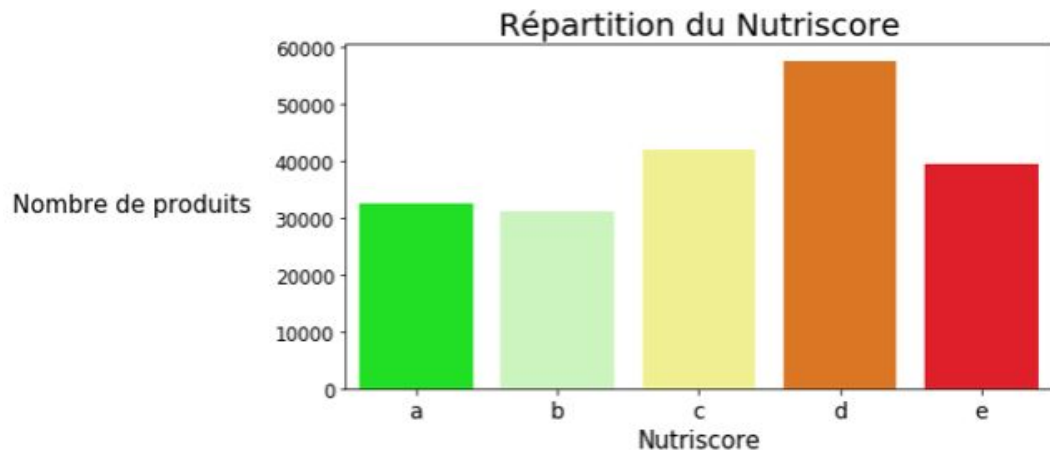
La démarche que j'ai choisie est de se baser sur le Nutriscore pour démontrer la valeur nutritionnelle de différents composants.

Puis de réaliser mes propres listes de bons et mauvais produits.

Le Nutriscore est un indice global de valeur nutritionnelle, basé sur divers métriques.

Le Nutriscore

- Présent sur 85% des produits
- Numérique ou alphabétique (de A à E)
- Les produits noté **A en vert** sont les meilleurs pour la santé
- Les produits en **E en rouge** sont les plus mauvais



D légèrement surreprésentés

Groupes globalement équilibrés

Analyse des graisses

- Indispensables pour notre santé
- Trois sous groupes
 - Acides gras saturés (Lait, fromages, viandes rouges)
 - Acides gras non saturés (Poissons, noix)
 - Dont les omegas-x
 - Les trans gras (Aliments frits, gateaux)

Impact des mauvaises et bonnes graisses

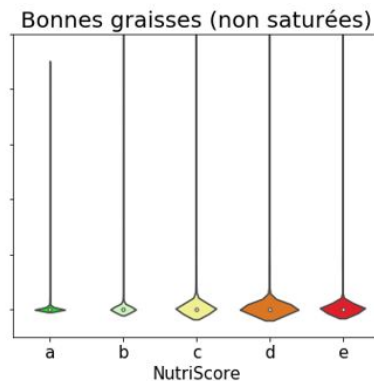
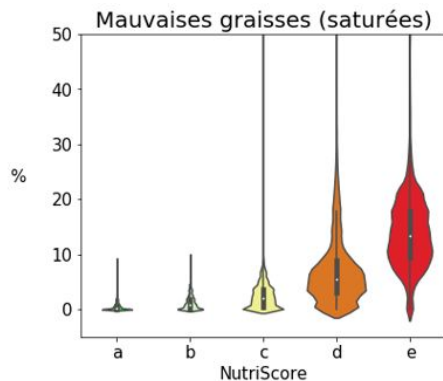
Mauvaises graisses =

Saturées + Trans

Impact très négatif

En grande quantité

90% notes **E** ont plus de 6.5% d'acide gras saturés

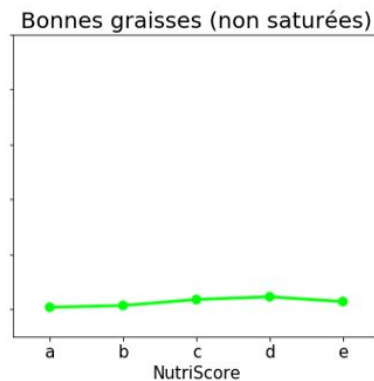
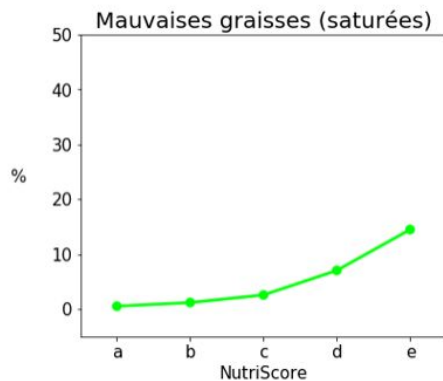


Bonnes graisses =

Non saturés (Omega etc...)

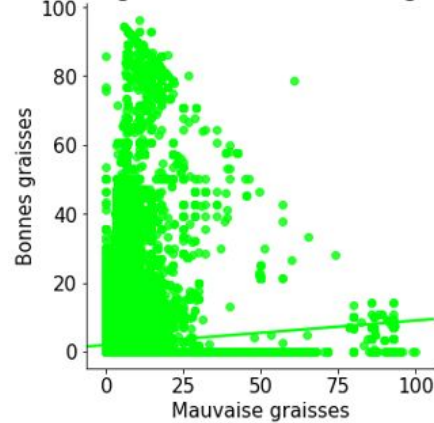
Impact moins négatif

En petite quantité



Privilégier les bonnes graisses

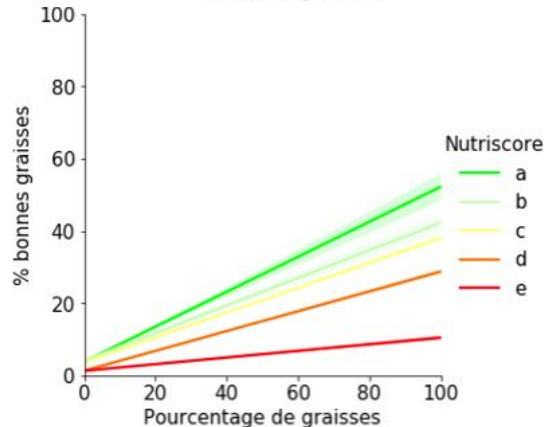
Bonnes graisses & Mauvaises graisses



- Nous avons besoin de graisses
- Il faut privilégier les graisses non saturés

Deux listes :

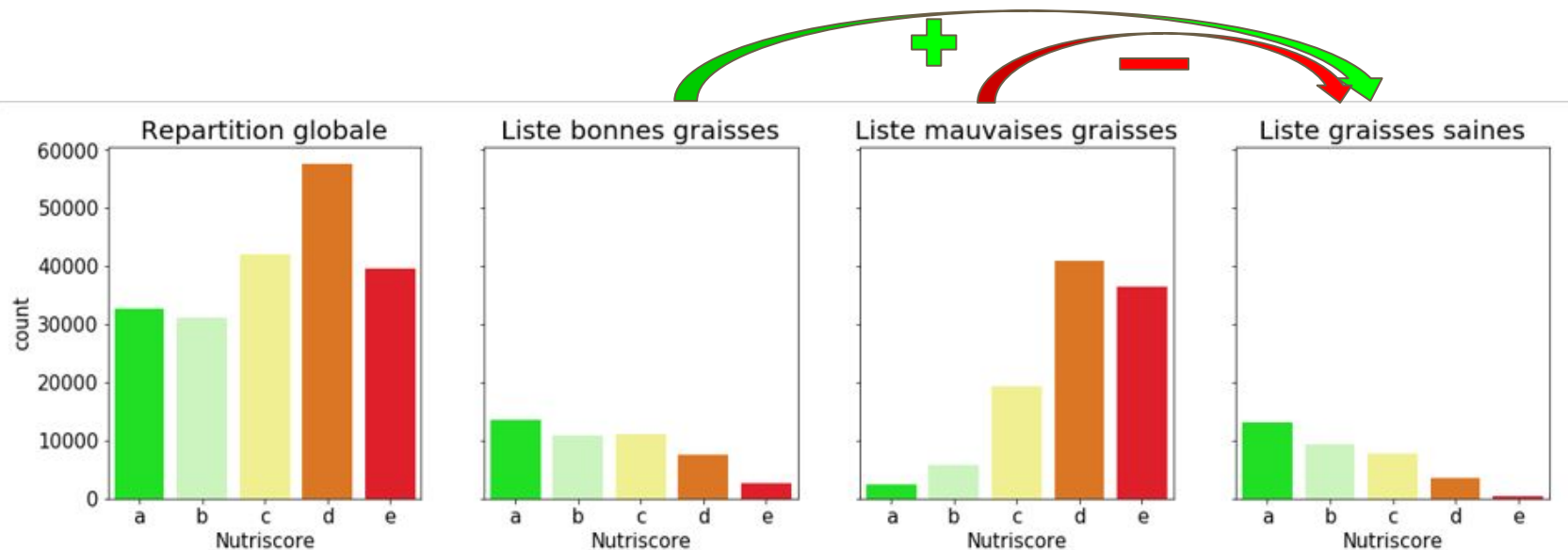
- Bonnes graisses (à garder)
 - De 1 et 10% de graisses on saturés
 - Ou entre 1 et 5% de graisses
- Mauvaises graisses (à retirer)
 - Plus de 2% d'acides gras saturés
 - Ou plus de 10% de graisses
 - Ou plus de 0.01% de trans gras



Répartition listes graisses

Croisement des listes

Graisses 'saines' = bonnes graisses moins mauvaises graisses



Analyse des glucides

- Nécessaires pour notre santé
- Deux groupes :
 - Les glucides simples (**sucre**s tels que le glucose, fructose, lactose..)

Sont absorbés rapidement par l'organisme

- Les glucides complexes (**amidon**, **fibres**)

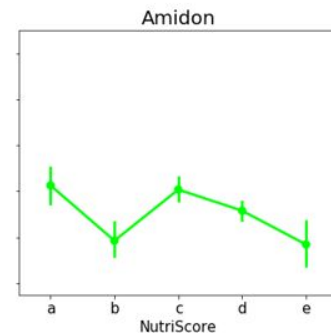
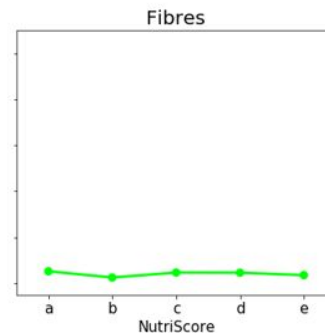
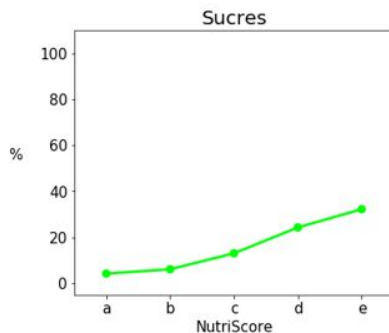
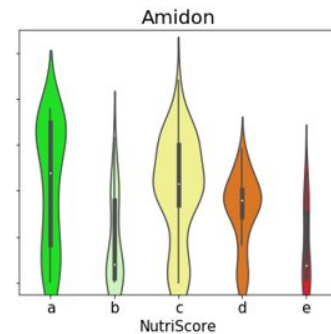
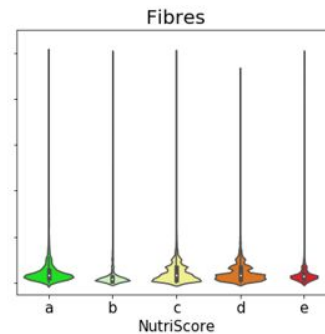
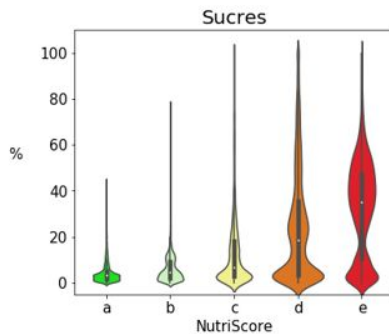
Pas ou peu absorbés

Bons pour la santé

Les glucides

Analyse des glucides

Forte teneur en **sucre** =
mauvaises notes



Tendance inverse pour les
Fibres et l'amidon

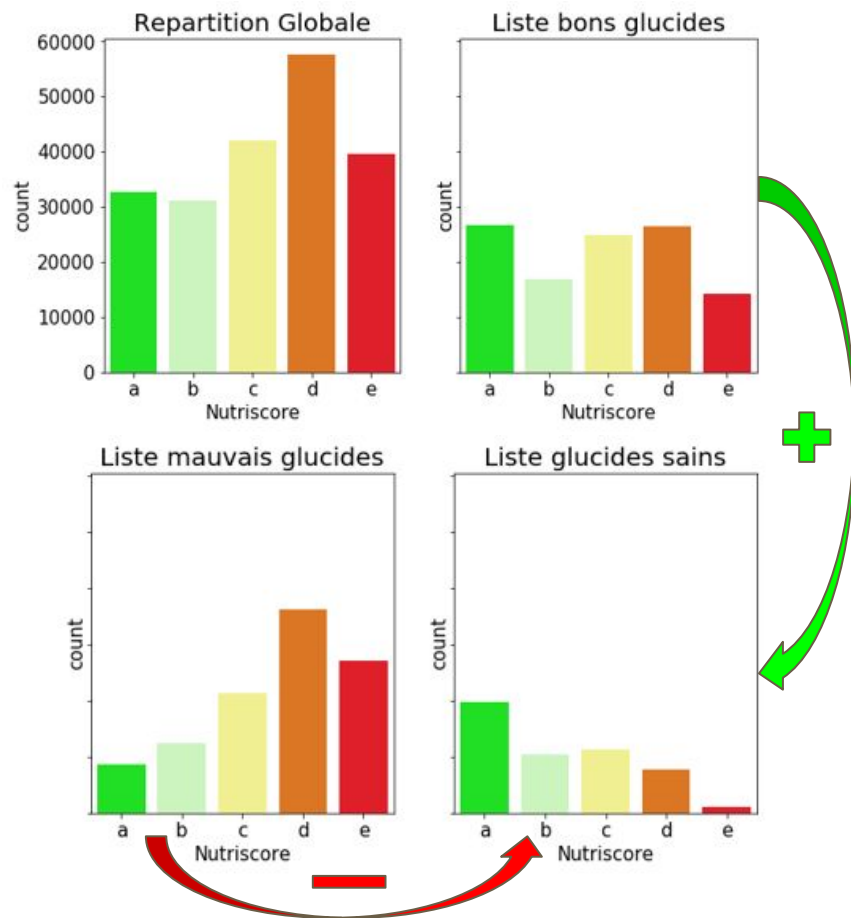
Résultats sur les glucides

- Avoir un minimum de fibres et amidon
- Eviter les sucres

Deux listes

- Bons glucides (à garder)
 - Au moins 0.5% de fibres
 - Ou bien 1% d'amidon
- Mauvais glucides (à retirer)
 - Plus de 5% de sucres au sens large
 - Ou plus de 75% de glucides

Même logique de croisement



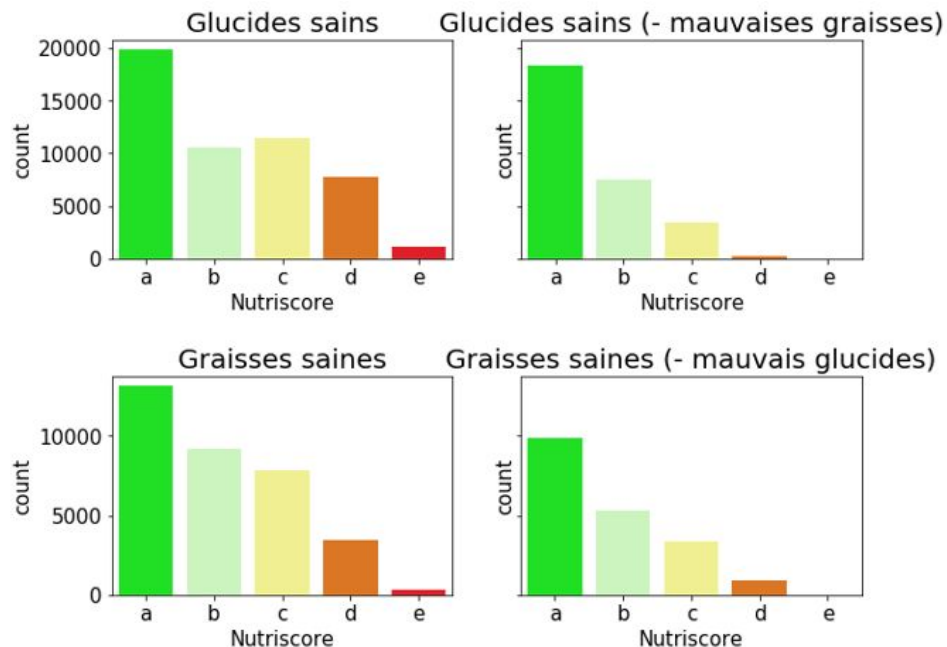
Croisement des listes précédentes

Aller plus loin

Soustraction des mauvais produits

Résultats probants :

- Très peu de notes basses
- Beaucoup de produits gardés



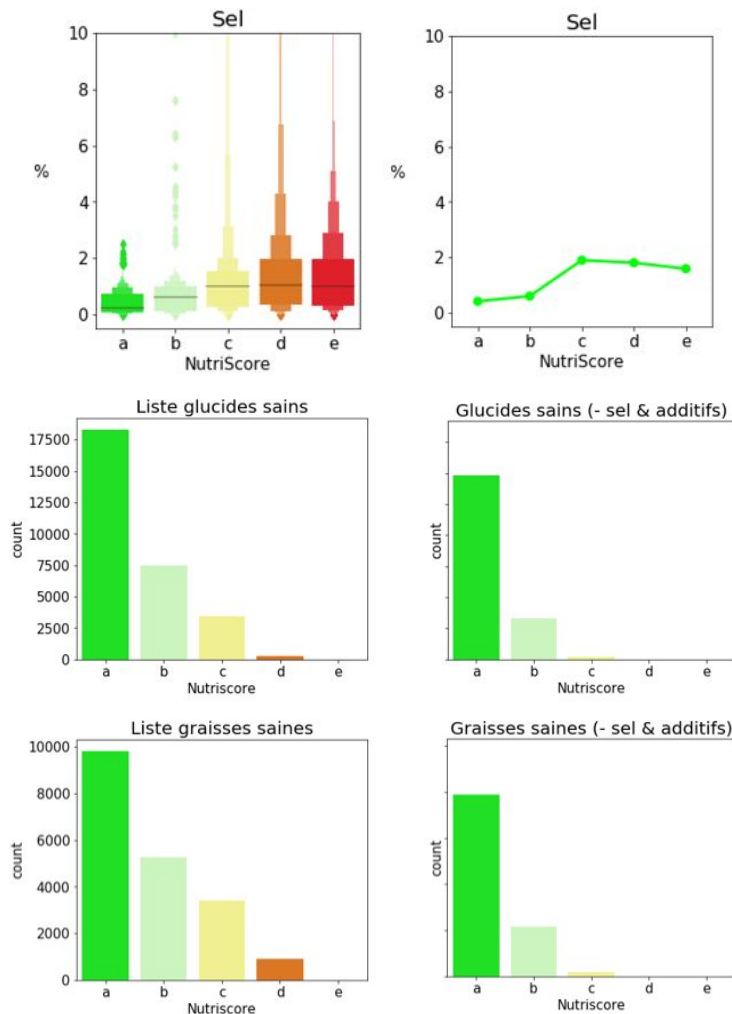
Sel et additifs

- Le sel (à retirer)
 - Consommé en trop grande quantité
 - Sel < 800mg / 100g
- Les additifs (à retirer)
 - Approuvés par l'EU
 - Parfois interdits (Australie)
 - Liste de 9 additifs controversés
 - Ex: e621- Glutamate
 - Seulement 1% des listes impactées

➡ Listes saines terminées

Source :

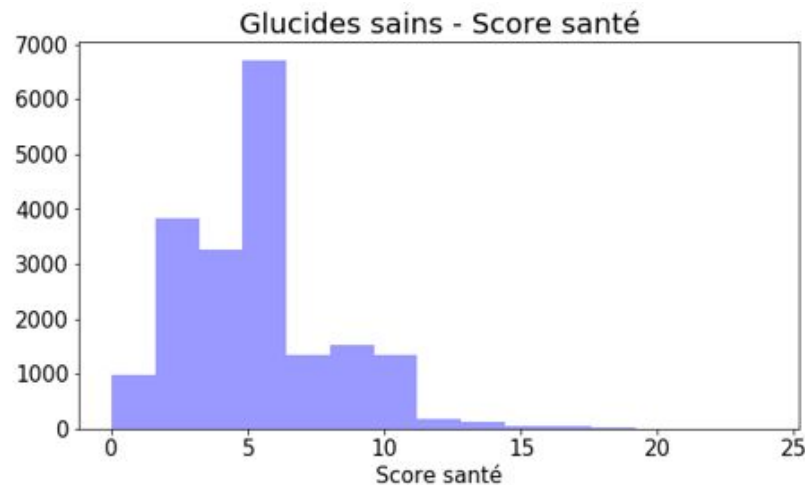
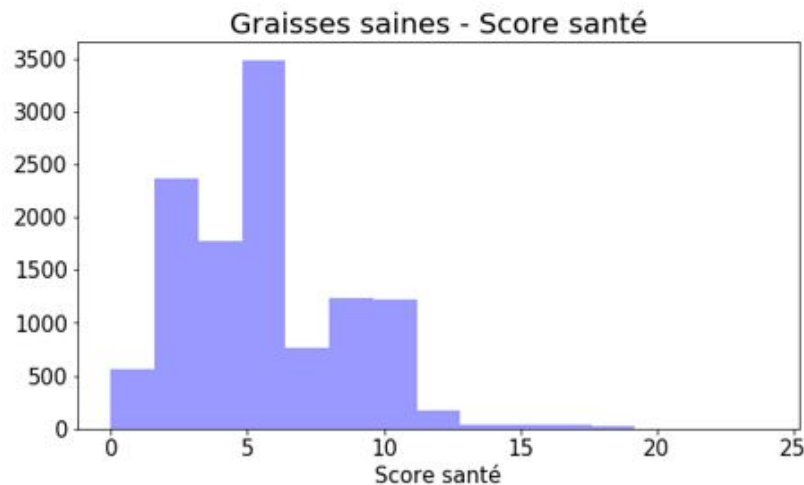
https://www.eatforhealth.gov.au/sites/default/files/files/eatingwell/efh_food_label_example_130621.pdf



Vitamins, Metals & Protein = Score santé

Nouveau "Score Santé" pour ordonner les listes

- + 1 point par type de vitamine
- + 1 point par type de métal
- + 2 à 6 points si le produit contient un bon pourcentage de protéines



Résultats

Nous avons donc désormais deux listes ordonnées des produits “sains”

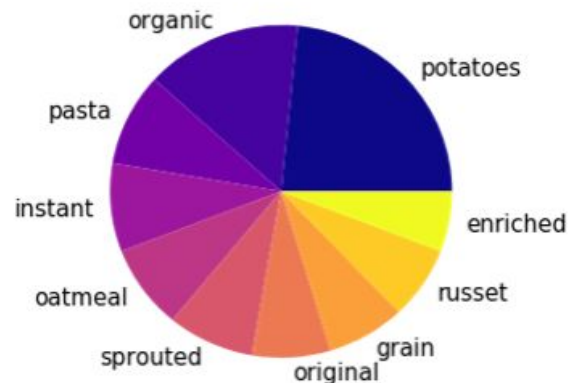
L'une contient des graisses, l'autre des glucides, deux ingrédients essentiels à notre santé

Liste produits glucides sains

Top 250 des produits à base de :

- Des pommes de terre
- Boissons et suppléments protéinés
- Pâtes
- Graines germées
- Flocons d'avoine

Mots récurrents liste glucides



product_name	brands	nutrition_grade_fr	health_score	carbohydrates_100g	fiber_100g	sugars_100g	proteins_100g
Pot-Sized Spaghetti	Anthony's	a	10	73.210000	3.600000	3.570000	12.500000
Black Beans	Hannaford	a	6	15.380000	3.800000	0.770000	5.380000
Organic Sweet Potato Chicken Dinner Baby Food	O Organics	None	6	16.810000	1.800000	3.540000	2.650000
Black Eyed Peas	Luigi Vitelli	a	5	9.850000	4.500000	0.000000	4.620000
Cuisine Millet	Isola Bio	b	0	9.400000	1.600000	0.300000	0.800000

Liste produits graisses saines

Top 250 des produits à base de :

- Macaroni & pâtes
- Boissons et suppléments protéinés
- Blé complet
- Graines



product_name	brands	nutrition_grade_fr	health_score	fat_100g	saturated_fat_100g	carbohydrates_100g	proteins_100g
Fettuccine, Enriched Macaroni Product	Mueller's	a	10	1.790000	0.000000	73.210000	12.500000
Lowfat Cottage Cheese	Lucerne	a	10	1.330000	0.880000	5.310000	11.500000
Protein Plus, Multigrain Pasta, Rotini	Barilla	a	9	3.570000	0.000000	67.860000	17.860000
Stir-Fry Chicken	Dining Smart	a	8	1.010000	0.170000	15.770000	5.370000
Marinara Tomato Sauce	Central Market	b	6	2.030000	0.000000	12.200000	2.440000

Conclusion

Listes terminées
Quelques produits phares
Selection finale manuelle

Axes d'amélioration

Associations automatiques

Trouver produits rares

Web crawling

Retours utilisateurs

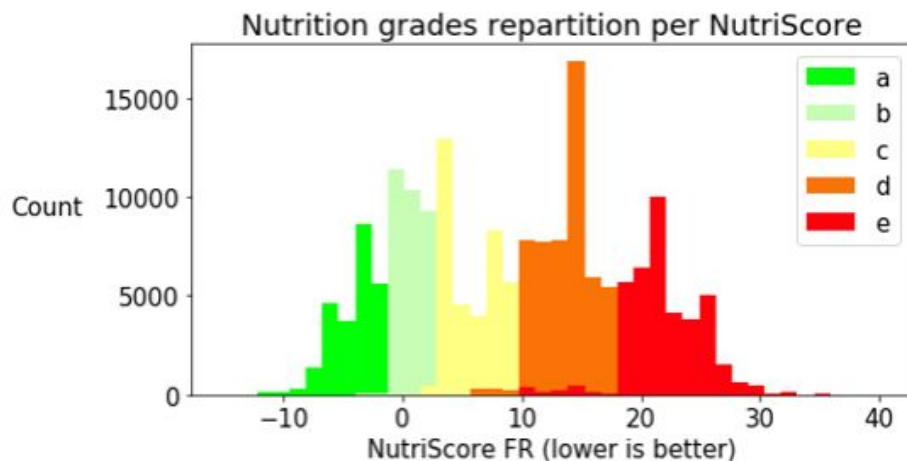
Questions



Xavier Montamat

Le Nutriscore numérique

- Le Nutriscore sous forme numérique (de -15 à 40)
- Plus les produits ont une note élevée, plus ils sont mauvais
- Plus précise mais moins lisible



Notations alphabétique et numérique similaires