

Analyse exploratoire des données

Objectif : **générateur de recettes saines**

Pour :  Lamarmite

Données :  Openfood facts



Pour des recettes saines

**Face à l'abondance du choix
Comment bien manger ?
Comment bien choisir ?**



DEMARCHE PROPOSEE

- 1/ Il faut définir les critères d'une alimentation de qualité**
- 2/ Identifier les sources de données pour ces critères**
- 3/ Faire un scoring des aliments sur la base de ces critères**
- 4/ Conserver les meilleurs aliments**
- 5/ Faire des compromis ? Budget / Choix / ...**

Nettoyage des données

320k Lignes, 163 Colonnes

Dataset initial

Réduction à 98 lignes

On a conservé uniquement les produits vendus en France

75 C Colonnes

On a enlevé tous les attributs < 0.01% de complétude

34 C

On a conservé uniquement les critères d'une alimentation de qualité et supprimé les attributs redondants
=> **Voir notebook pour la justification détaillée de chaque choix**

10 C

Feature engineering : on a créé 10 variables de scoring



Les critères d'une alimentation saine

Feature engineering

Critère	Données retenues	Features créées
Une bonne proportion d'énergie, sucre, sel, graisses, fruits et légumes, fibres, protéines	nutrition-score-fr_100g energy_100g salt_100g sodium_100g sugars_100g saturated-fat_100g fiber_100g proteins_100g	Un scoring entre 1 et 5 pour chaque des données retenues Source pour les intervalles de scoring : <u>Nutri-score_reglement_usage_041019.pdf</u>
Consommer un maximum de produits bruts non transformés	Liste des ingrédients, y compris les additifs : ingredients_text	no_ingredients_scoring additives_nocive_scoring ➔ Alimentation des additifs nocifs à partir d'une source de données externe : la collecte exhaustive de ces données reste à réaliser par le client
Privilégier le bio français	labels_tags	Un scoring de 1 à 5 : 5 : bio français 4 : bio européen 3 : bio 2: fabriqué en France 1: ni bio ni français

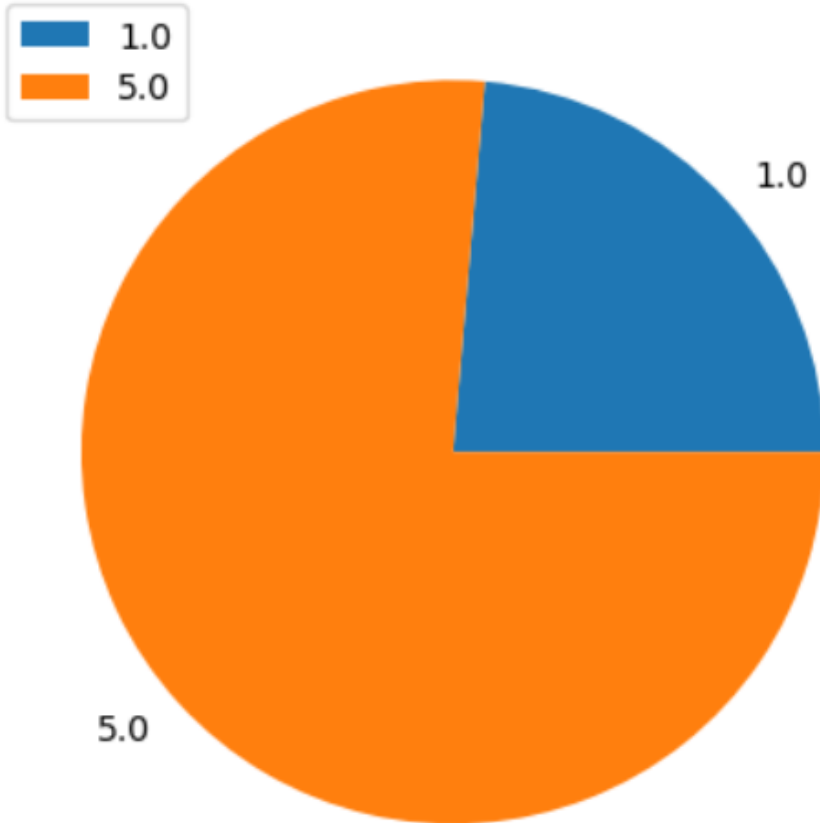
Critère	Données retenues	Features créées
Manger des produits de saison	<p>On pourrait construire un tableau contenant les produits par saison de manière à proposer des recettes ciblées selon la date.</p> <p>En récupérant par exemple les informations ici : https://alimentsain.fr/aliment/calendrier-fruits-legumes/</p> <p>Mais il s'agit d'une source de données externe hors périmètre de ce projet.</p>	NA
Dimension écologique	<p>Le nombre d'ingrédients issus de l'huile de palme est disponible</p> <p>Mais la dimension écologique n'influe pas à elle seule sur la qualité nutritionnelle des aliments. Elle n'a donc pas été retenue ici</p>	NA

Critère	Données retenues	Features créées
Respecter les AJR (Apports Journaliers Recommandés)	<p>Les données concernées ne sont pas assez représentées dans les données fournies. On ne les a donc pas retenues :</p> <p>vitamin-b1_100g 0.008736 magnesium_100g 0.008279 vitamin-e_100g 0.008035 vitamin-b6_100g 0.007578 vitamin-pp_100g 0.007213 vitamin-b9_100g 0.007009 vitamin-b2_100g 0.006562 vitamin-d_100g 0.006004 vitamin-a_100g 0.005983 phosphorus_100g 0.005719 vitamin-b12_100g 0.005638</p>	NA

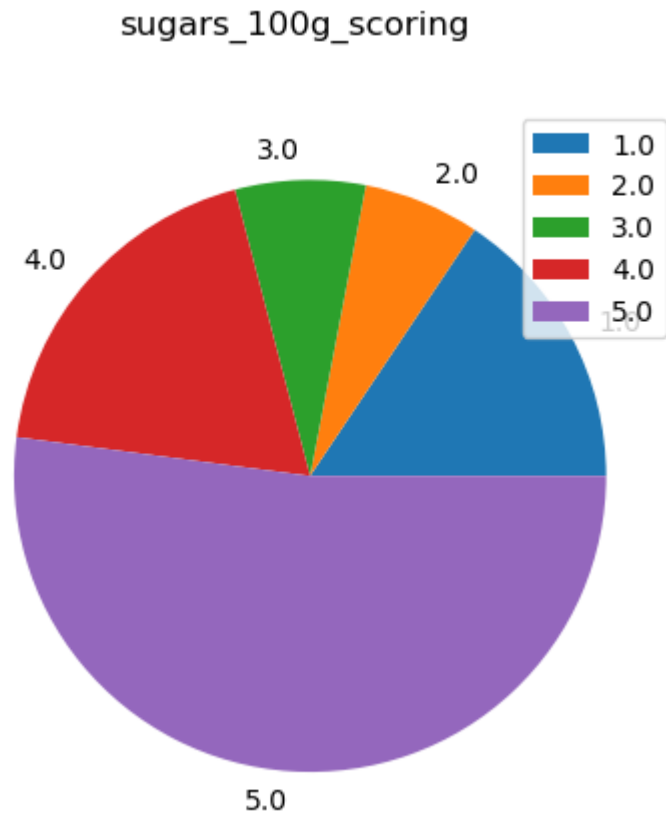
**Quelques constats sur les données analysées,
qui montrent l'intérêt de bien choisir**



additives_nocive_scoring

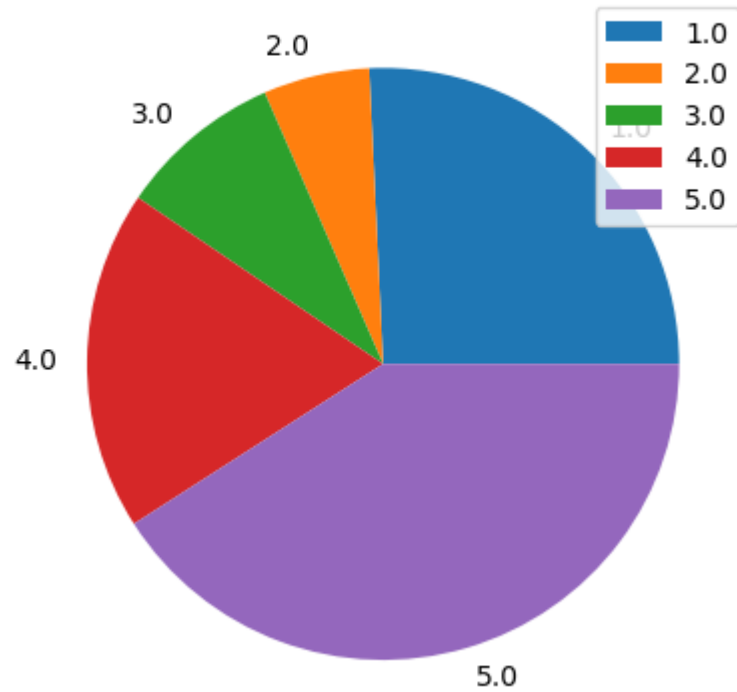


Près d'un quart des produits
comportent au moins 1 additif
considéré comme nocif
(scoring 1)

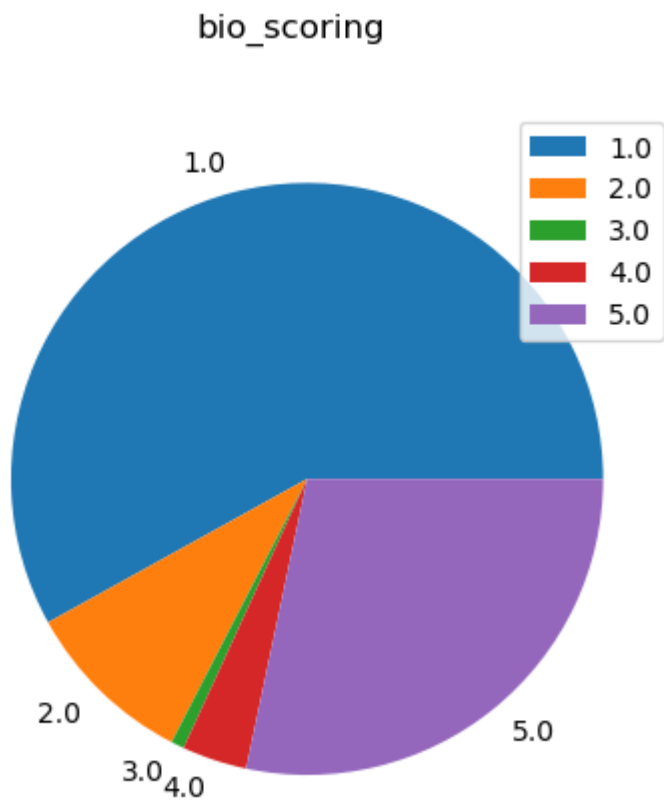


La moitié des produits comportent plus de 4,5g de sucre /100g (ils sont en dessous du **scoring 5** pour le sucre)

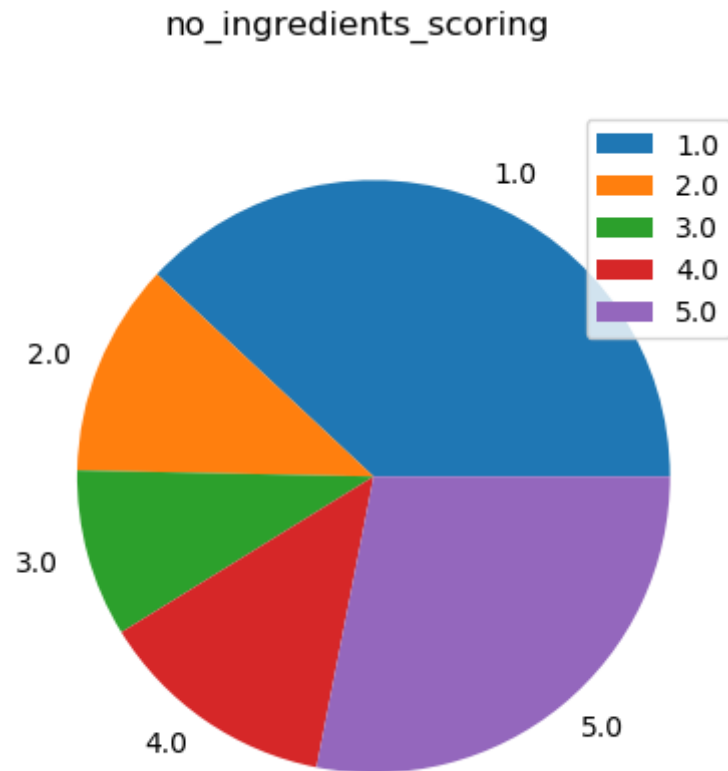
saturated-fat_100g_scoring



Un quart des produits
comportent plus de 7g de
graisses saturées
(**scoring 1**)



Plus d'un quart des produits sont
du bio français (**scoring 5**)



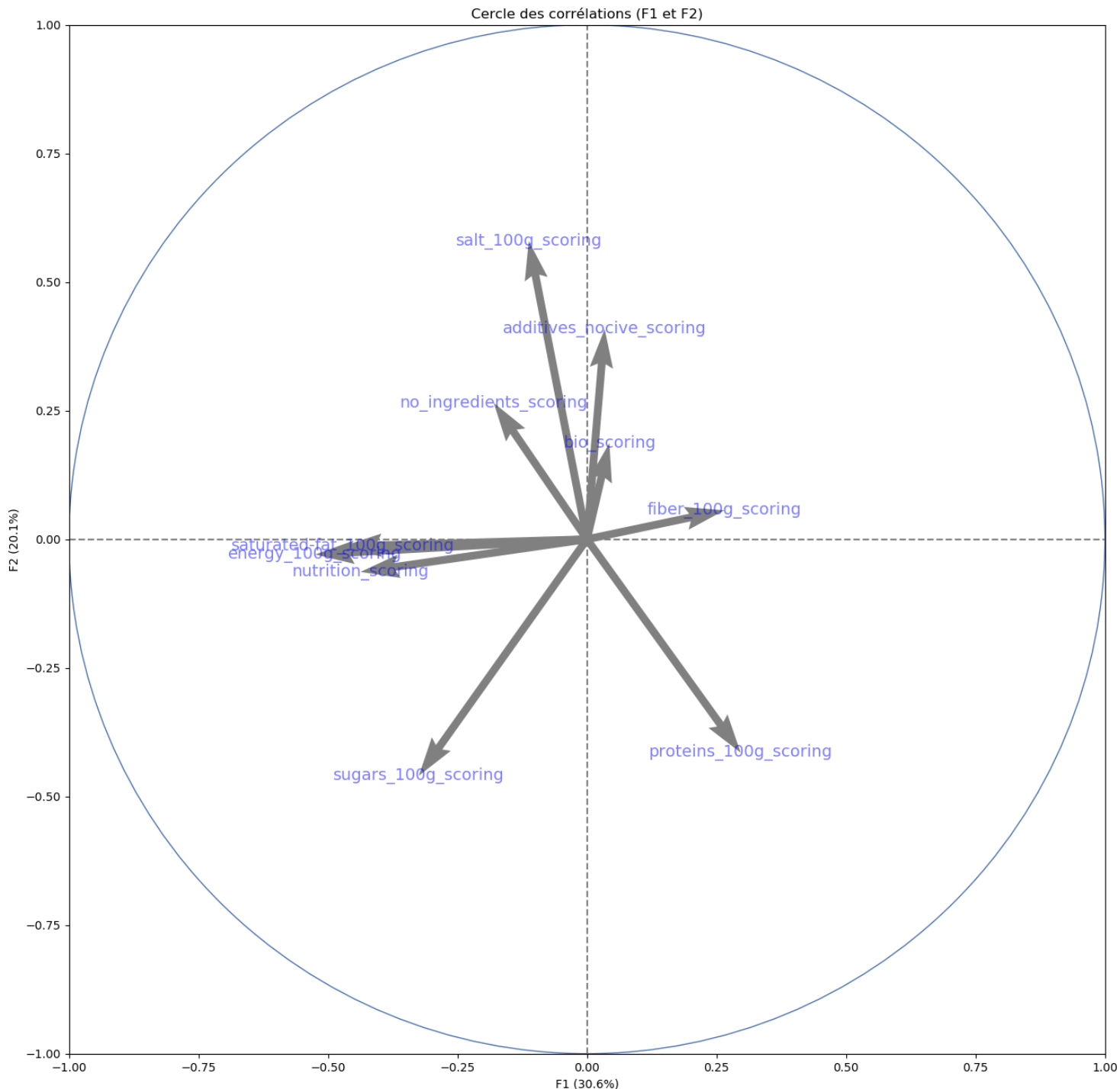
La moitié des produits ont plus de 7 ingrédients
(**scoring 1 : > 10 ingrédients**,
scoring 2 : > 7 ingrédients)

à partir de 5 ingrédients on commence à considérer le produit comme transformé

→ À nuancer par rapport aux produits préparés, pour lesquels on peut tolérer davantage d'ingrédients

Les variables de scoring et leurs relations





On a 3 groupes de variables intéressants :

- Additifs, nombre d'ingrédients, bio/français
- Nutrition scoring, graisses, énergie, sucres
- Fibres, et protéines (on peut se permettre de ne pas les considérer car elles sont déjà prises en compte dans le nutrition scoring)

Le sel joue un rôle important. On peut ne pas le considérer car il est inclus dans le nutrition_scoring

Conserver les meilleurs aliments et faire des compromis



Si on ne conserve que les aliments ayant tous les scorings à 5 on se rend compte qu'il reste très peu de choix (voir slide suivante)

On peut faire un compromis si on ne conserve que les aliments avec :

⇒ le nutrition scoring = 5

⇒ un scoring bio ≥ 2

⇒ (ce qui signifie au minimum des produits français, qu'ils soient bio ou non. Et pas d'additif nocif)

⇒ un scoring sur le nombre d'ingrédients ≥ 2

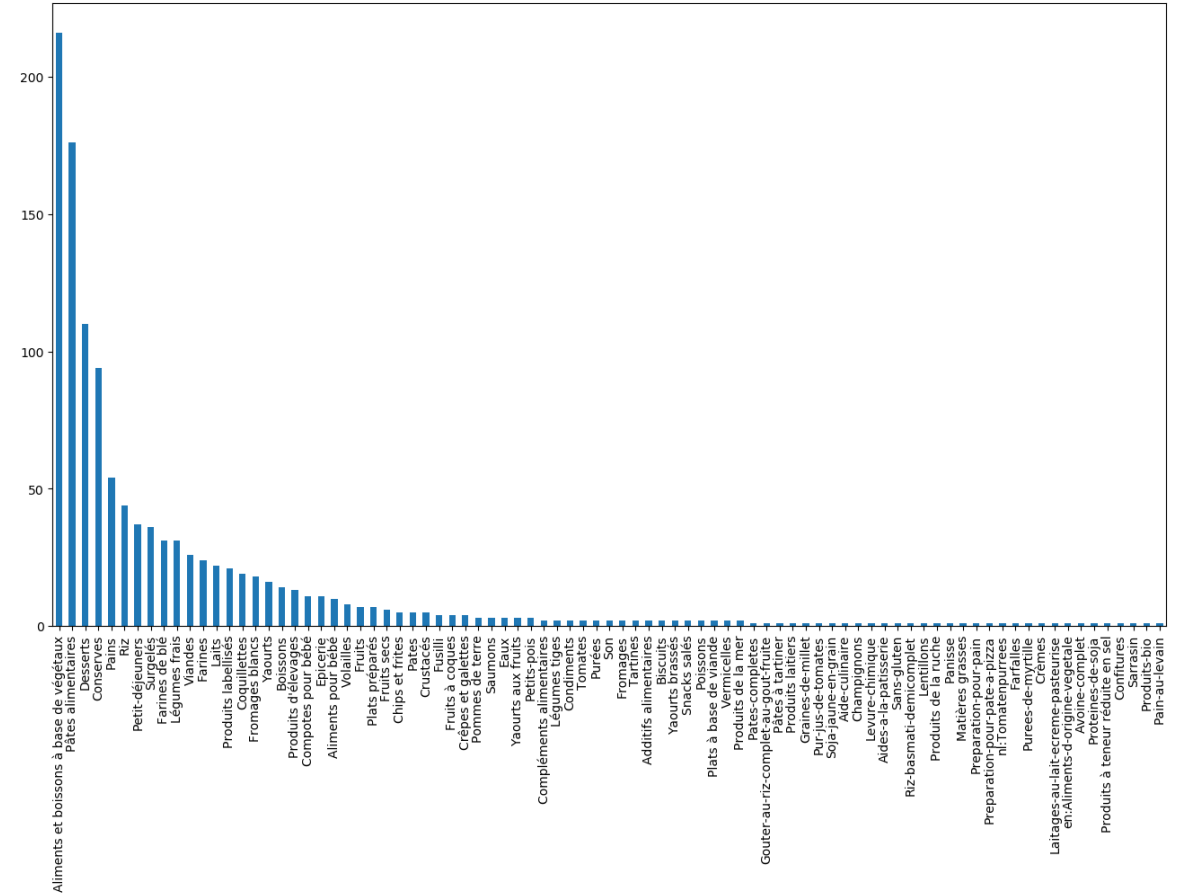
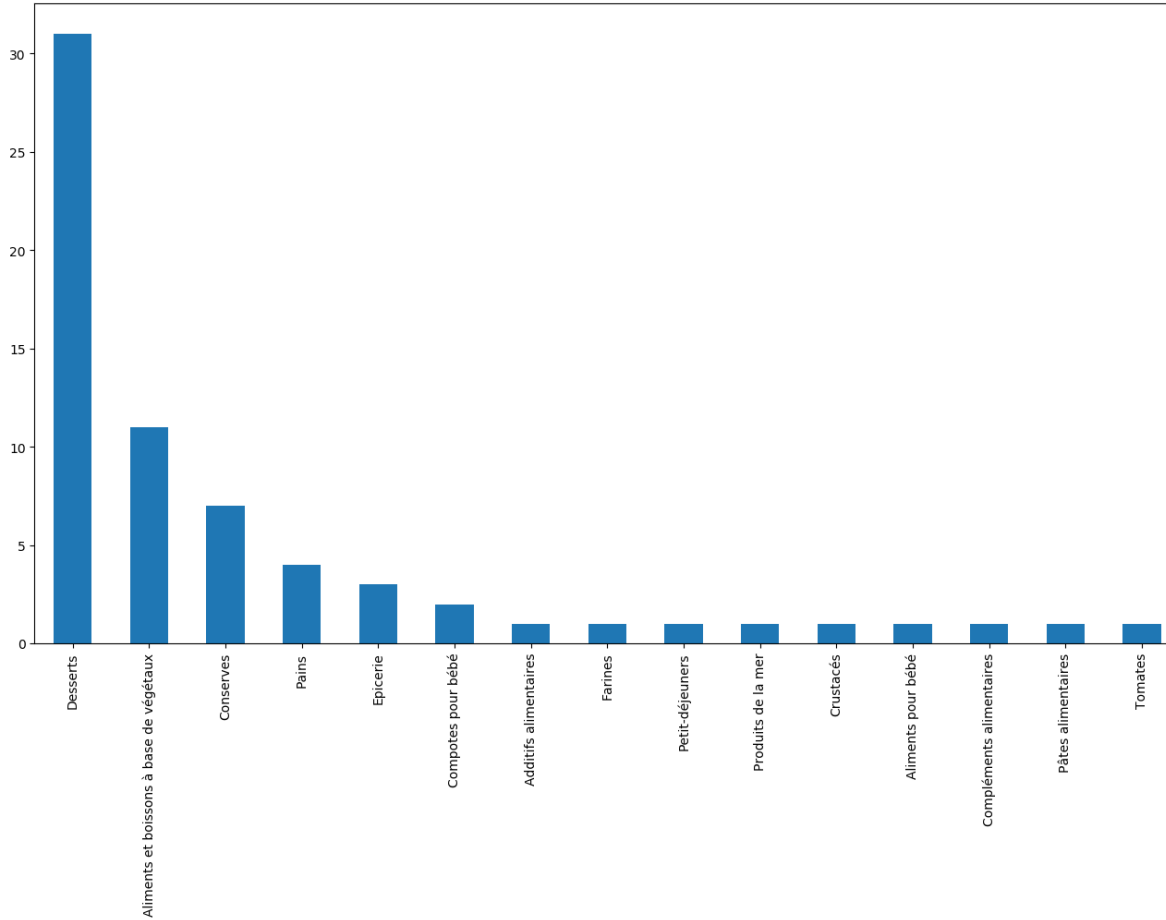
⇒ (c'est-à-dire pas plus de 10 ingrédients pour éviter les produits transformés)

Alors on se rend compte qu'il y a plus de choix (voir slide suivante)

Diversité des choix

Tous les scorings = 5 :

Nutrition score = 5
Bio score ≥ 2
Ingredients score ≥ 2

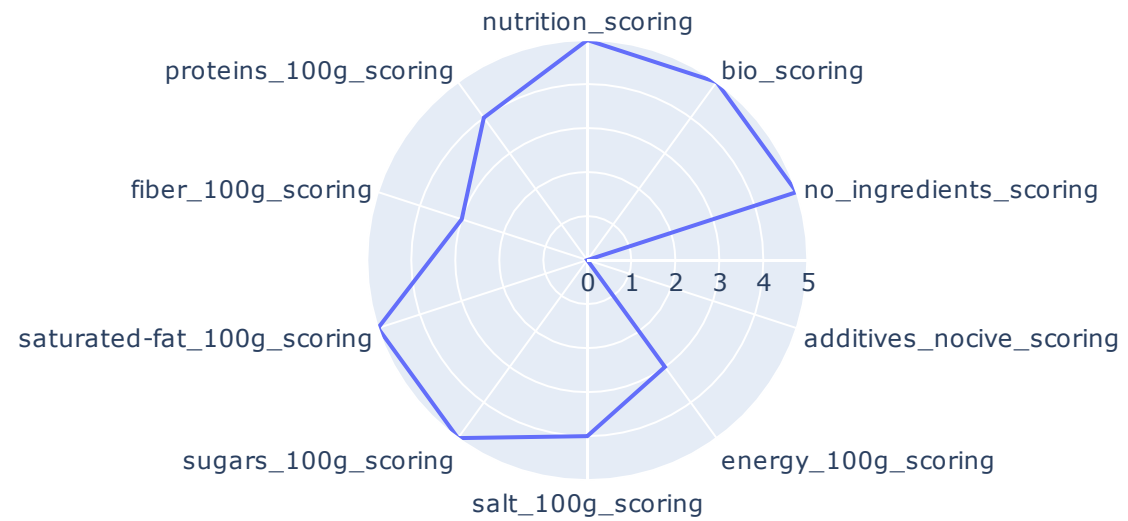


Exemples d'aliments de qualité obtenus par les scorings



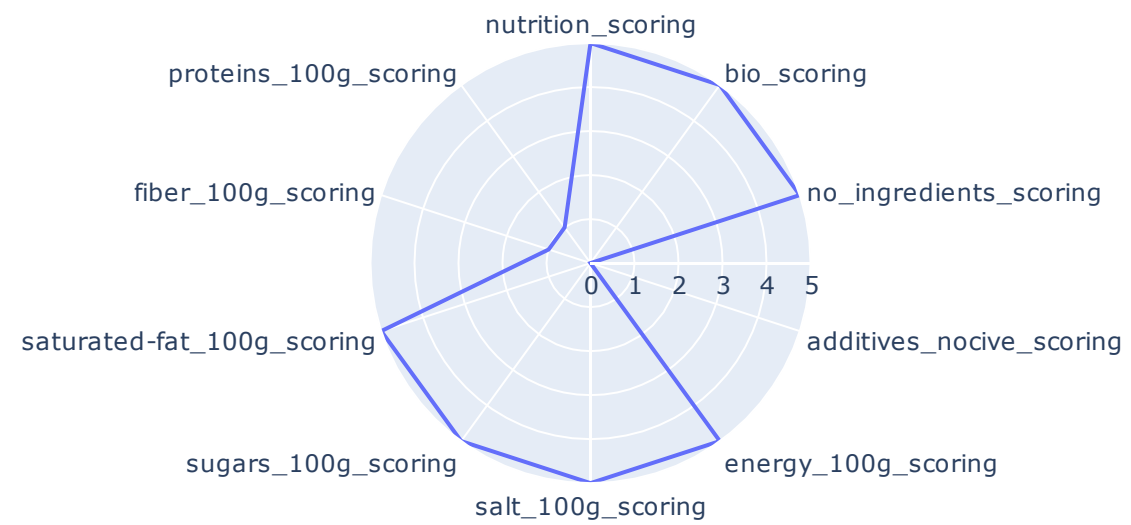


Fines Galettes de Riz



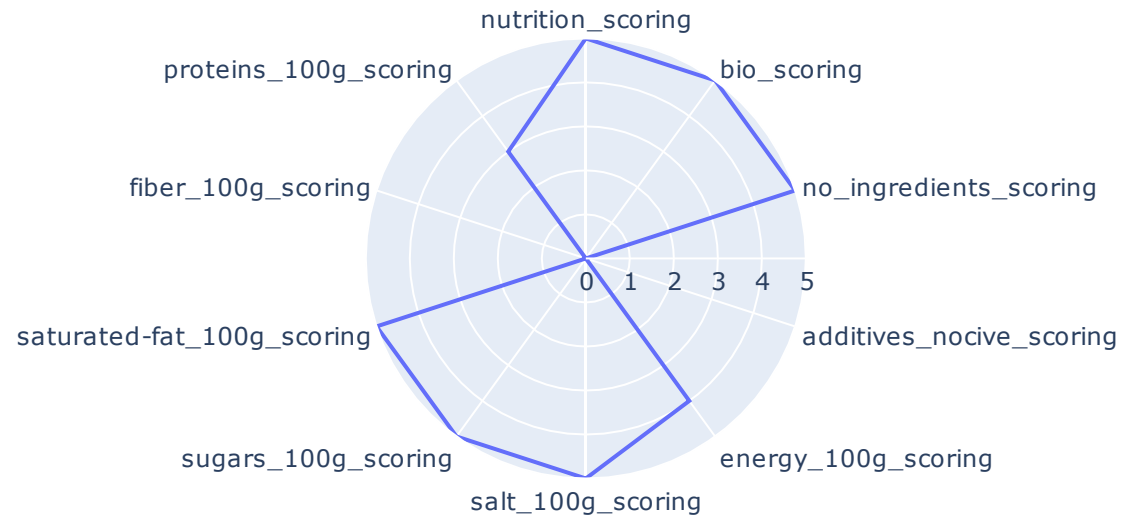


Tomacouli Bio



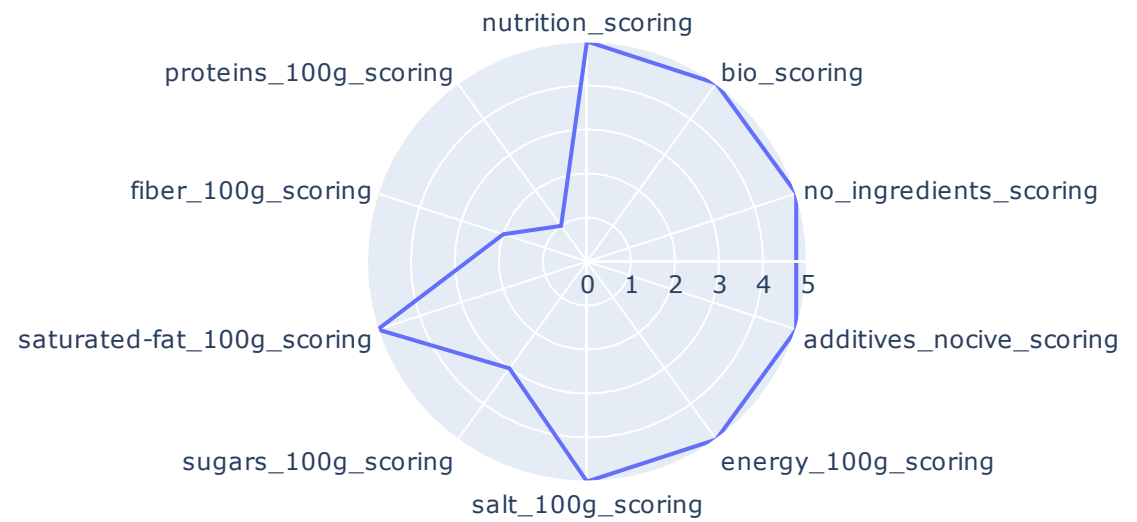


Fusilli



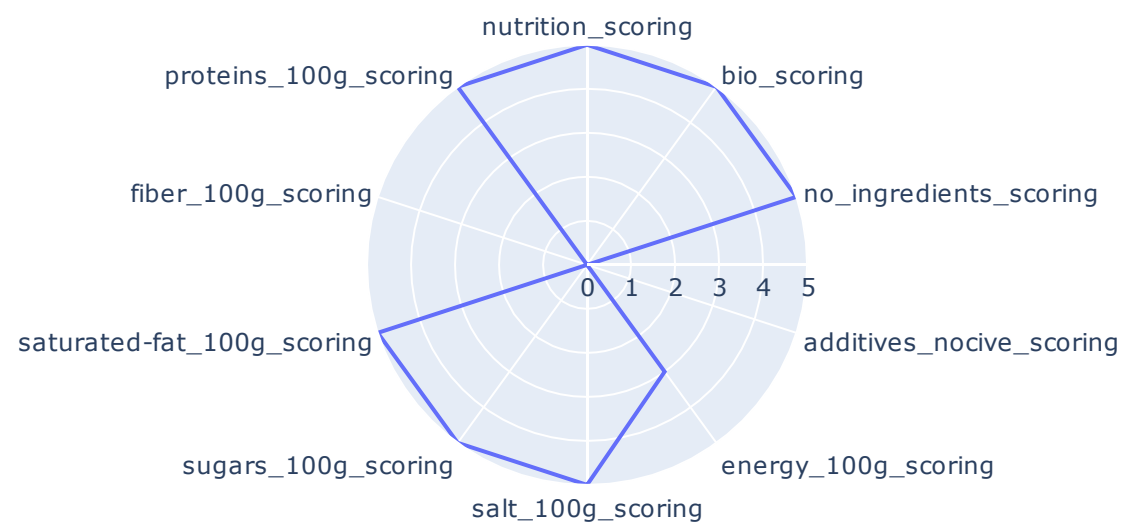


Compote de Pomme bio allégée en sucres



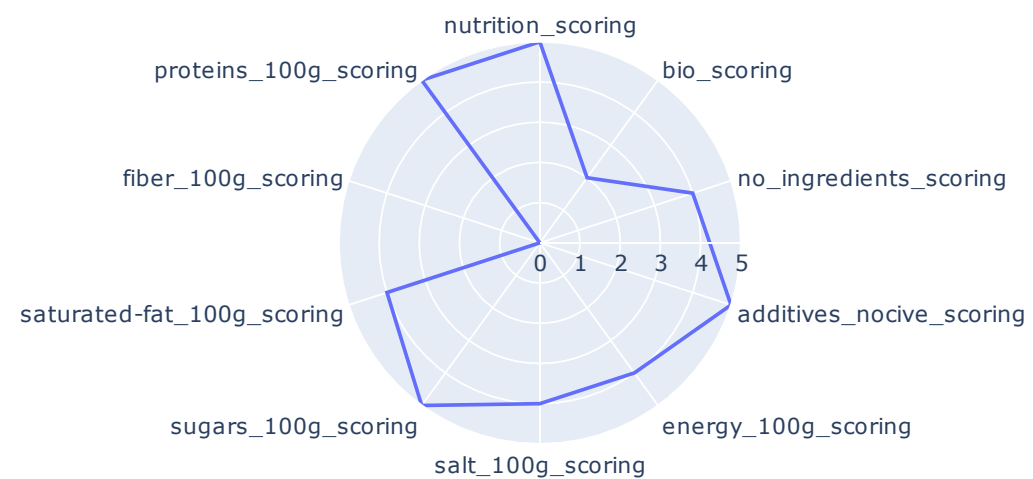


Riz Basmati naturellement parfumé Bio Casino

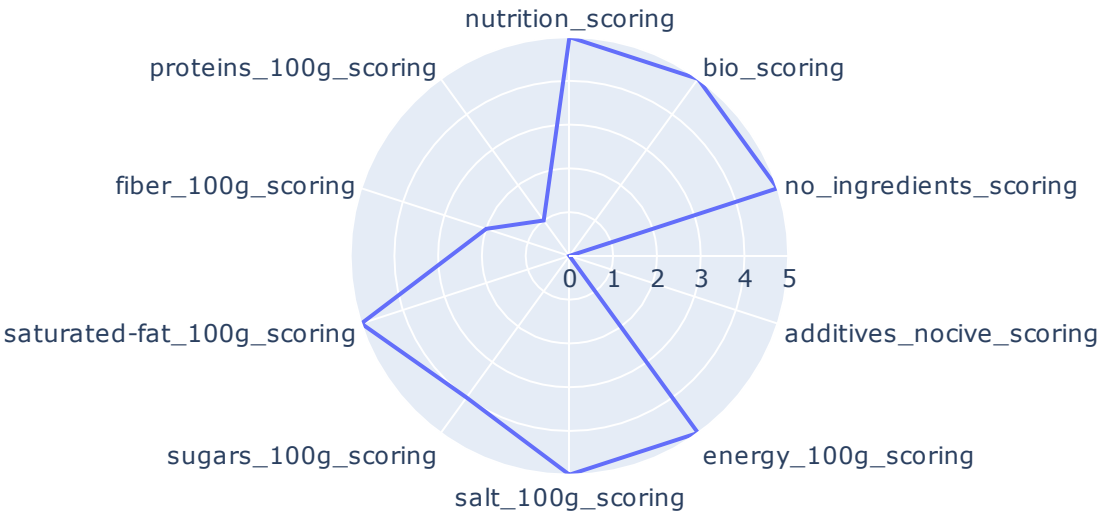




Carpacio charal



Carotte bio



Haricots verts extra fins

