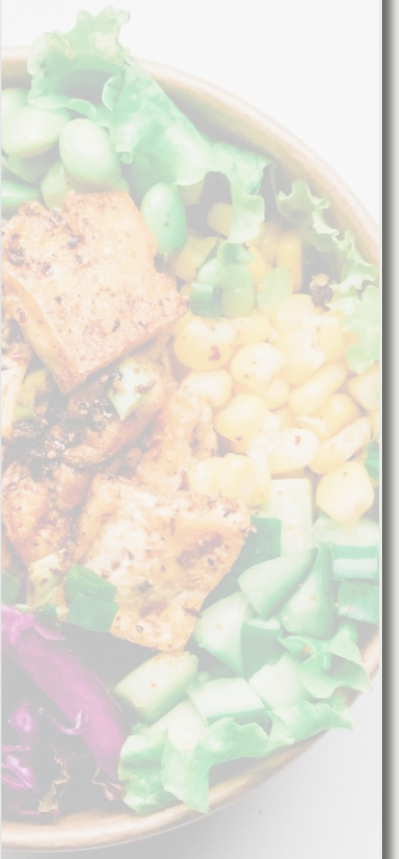
Nudge Score

Projet 3 – Raphaël Koudache--Louvet

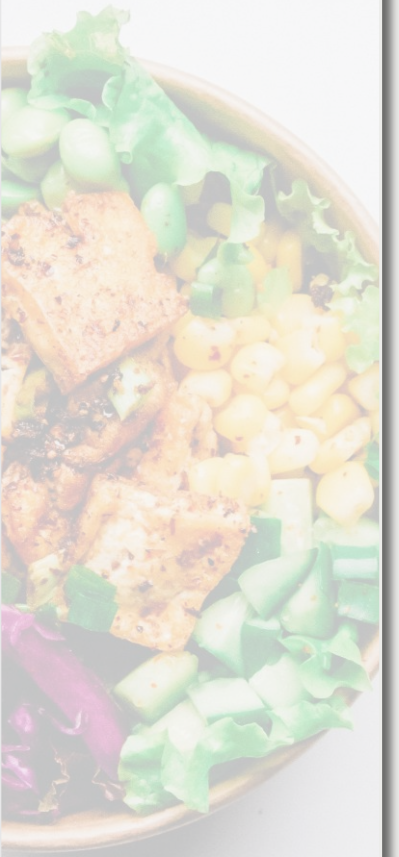


Contexte

Santé Publique
(nutriscore)

Application en lien avec
l'alimentation

Open Food Facts



Pain Points

Nutriscore

Données manquantes

Modèle qui « tire vers le bas »

« Vulgarisation qualitative »

Impossibilité de faire un score total « mathématique » (qualitatif, poids)

Manque de granularité

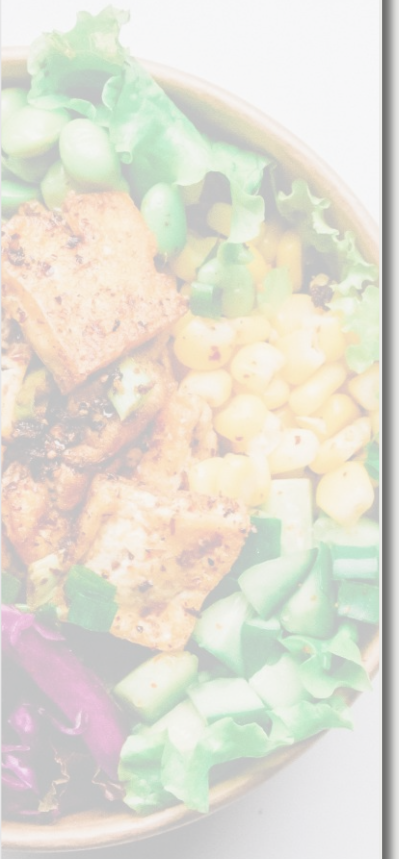
Psychologie du changement

Hype Curve & Binarité

~~Sacrifice~~

Effet pygmalion

Référentiel (autres vs soi-même)



Concept

Compensation/balance

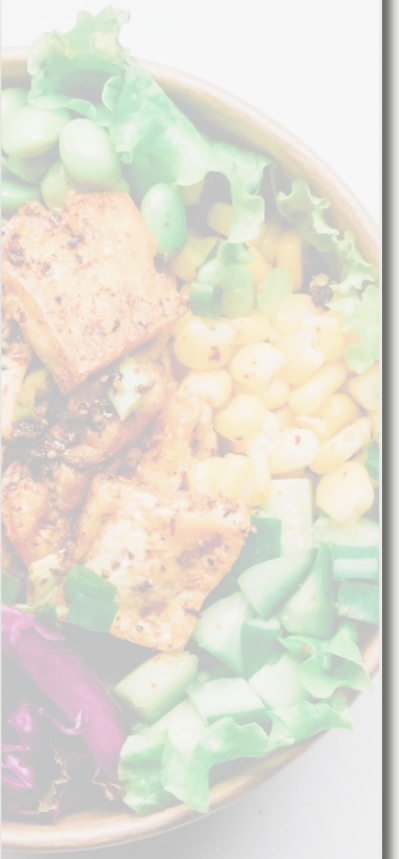
Encouragement

Long terme

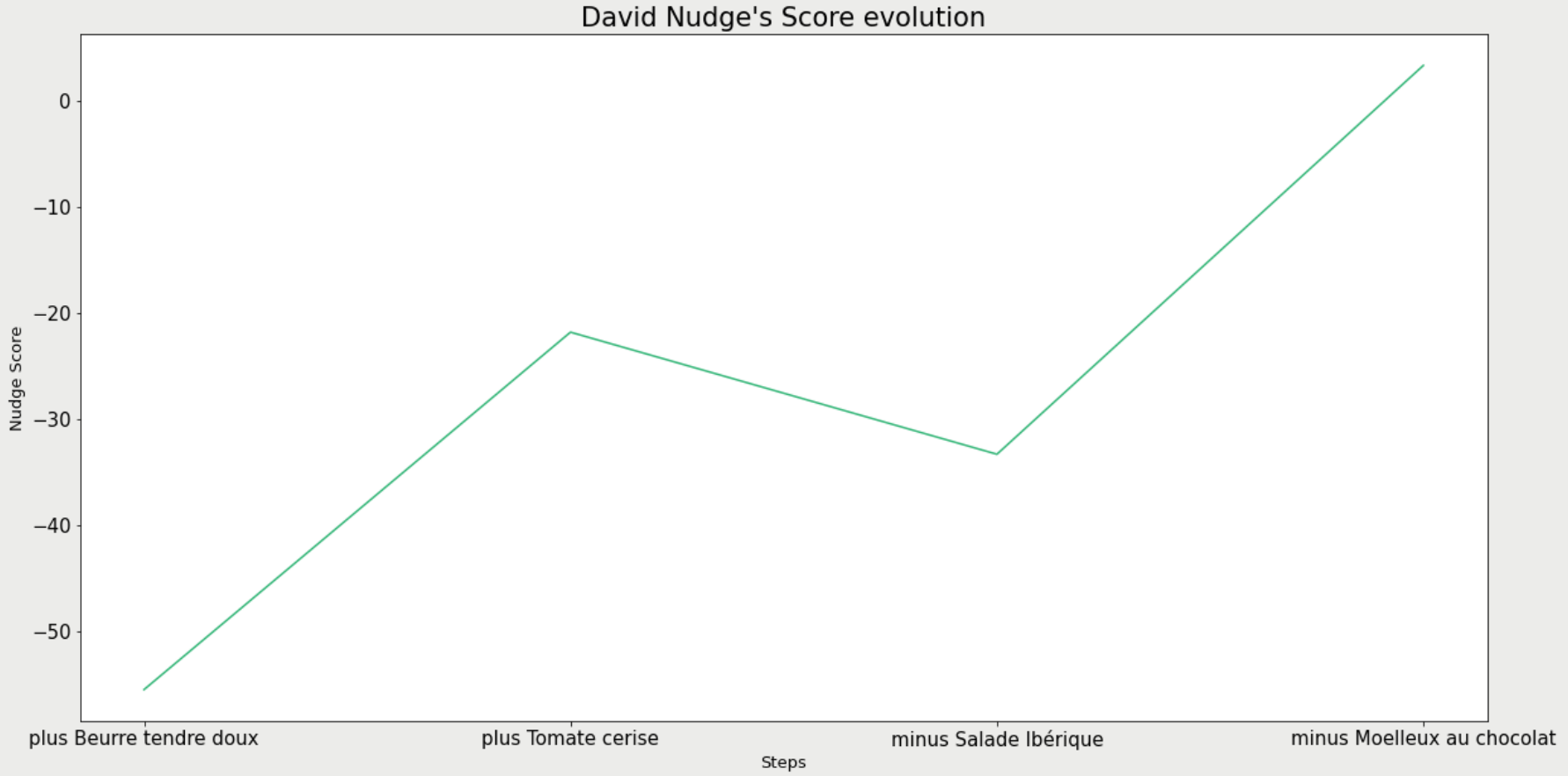
Si 1%

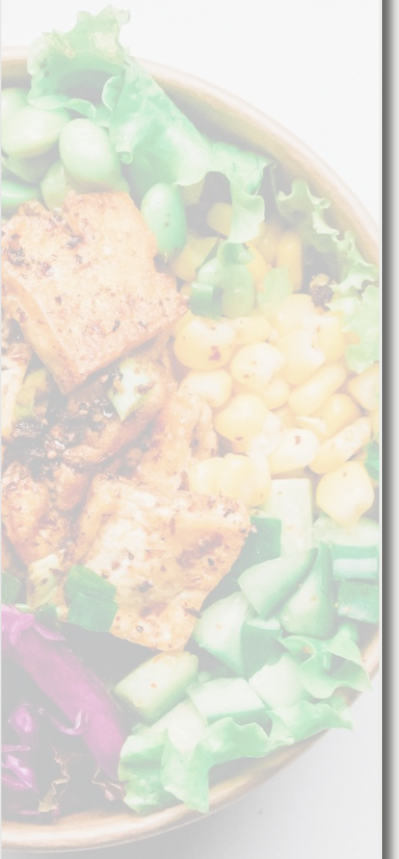
nutriscore.nudged()

Taux d’évolution panier moment m vs m-1



Exemple

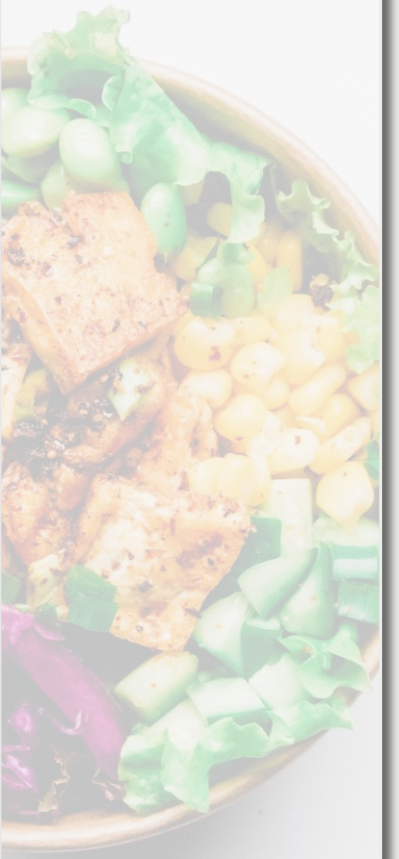

David Nudge's Score evolution
0
−10
−20
−30
−40
−50
Nudge Score
plus Beurre tendre doux
plus Tomate cerise
minus Salade Ibérique
minus Moelleux au chocolat
Steps



Plan

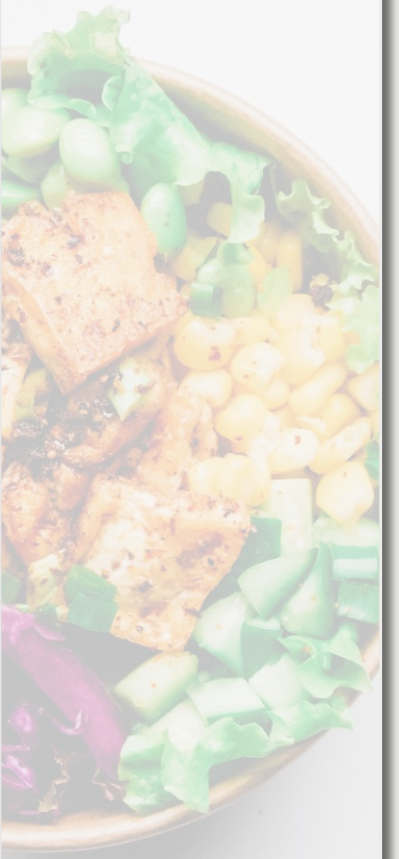
I. Préparation de la Base de Donnée

II. Analyse Statistique



I. Préparation de la Base de Donnée

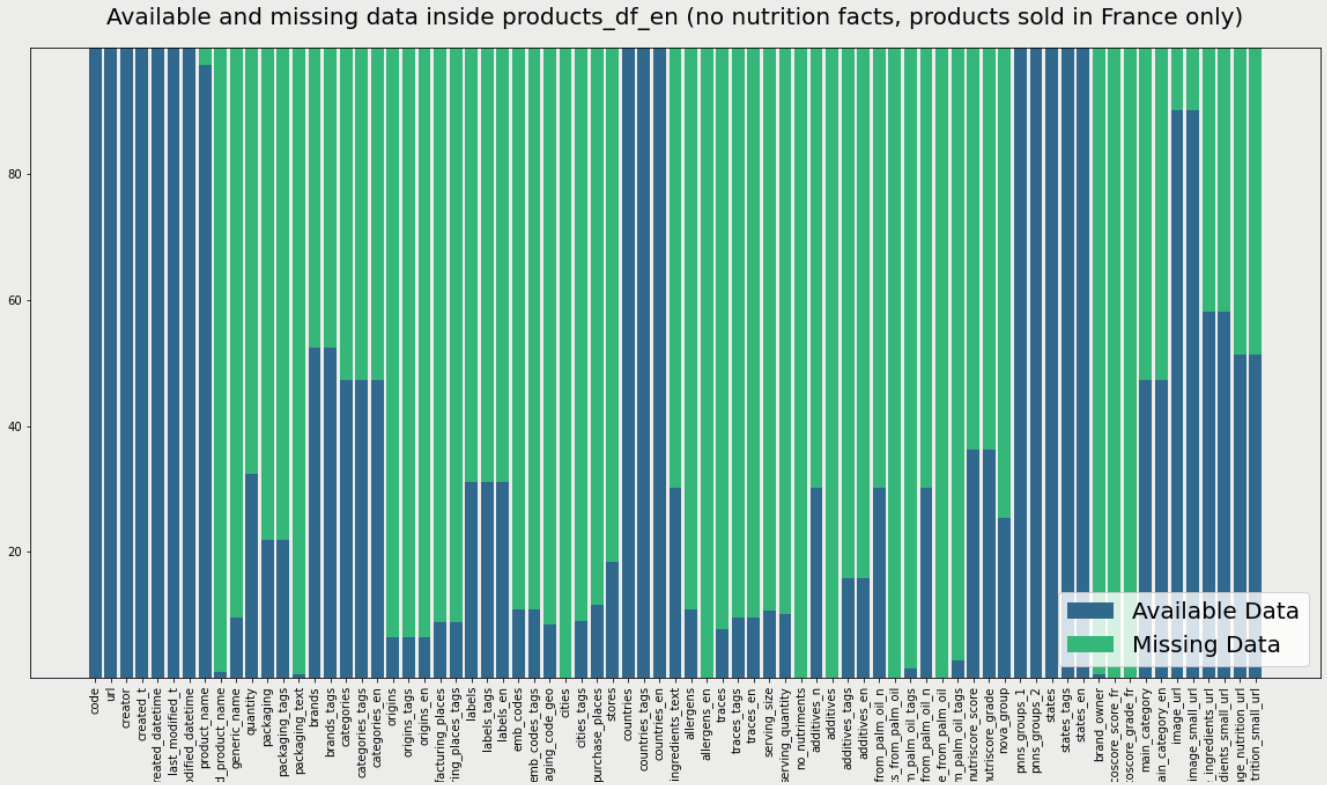
1. Découverte des Données

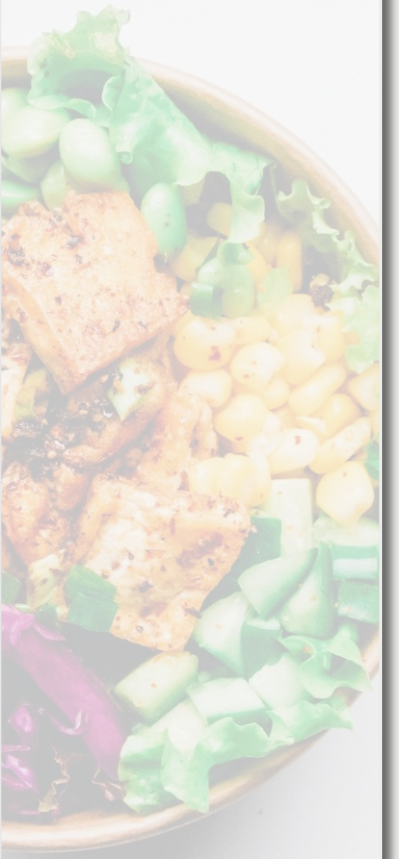


2,9 GB

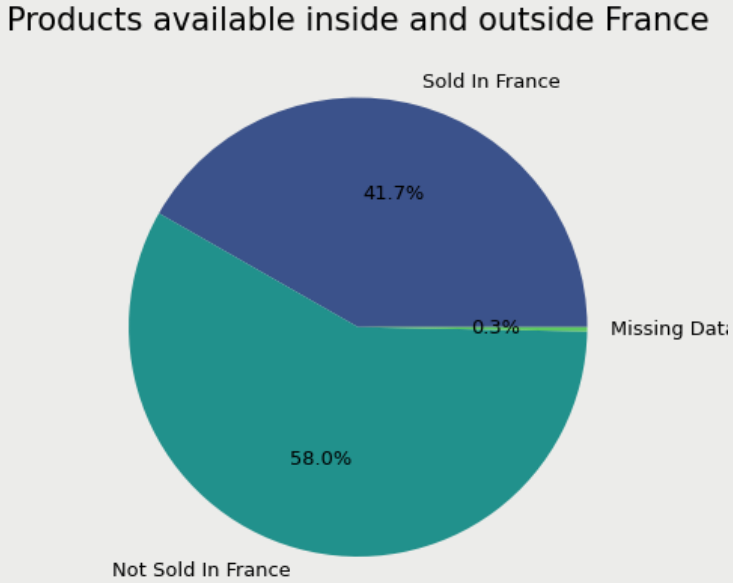
187 Features

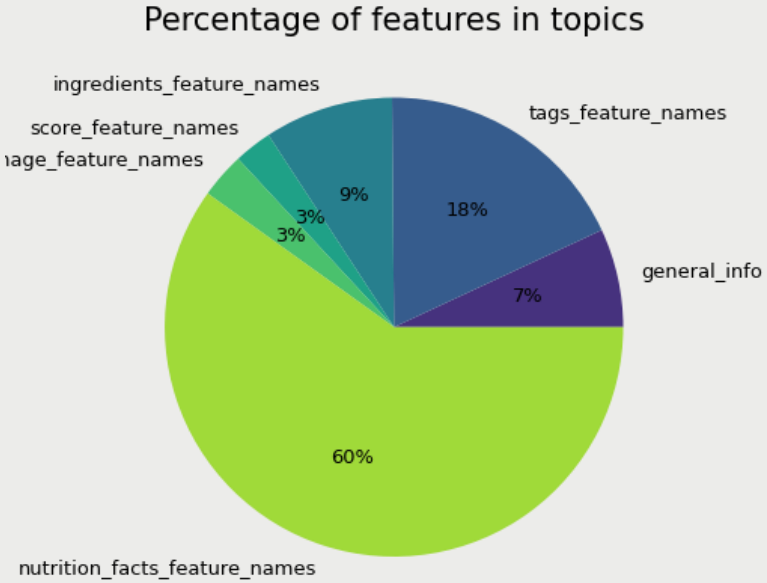
2 059 358 Samples





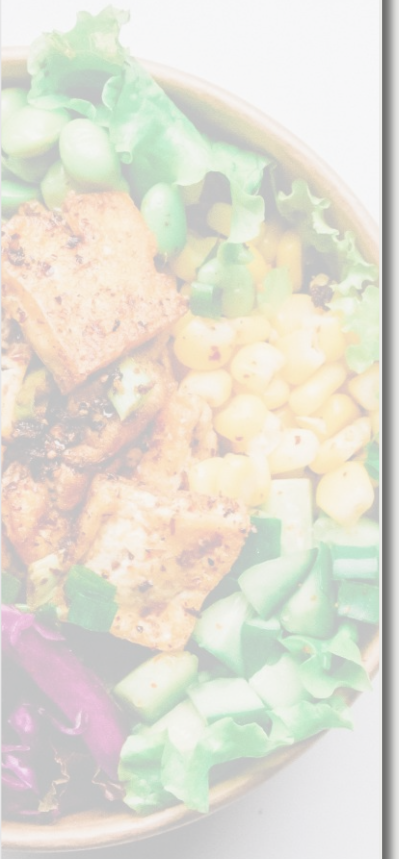
2. Sélection des features (Pas de nutriments)





Concept, Documentation, Topics, Géographie

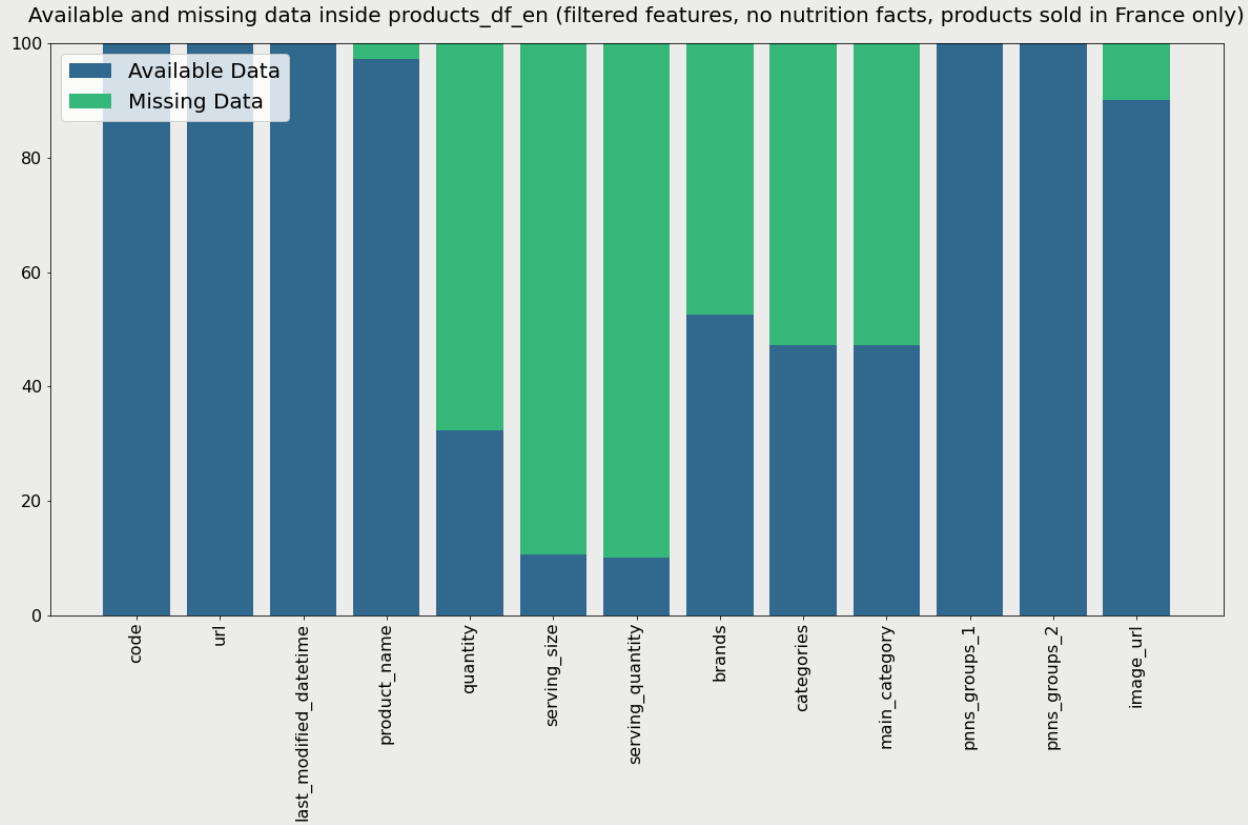
2. Sélection des features (Pas de nutriments)



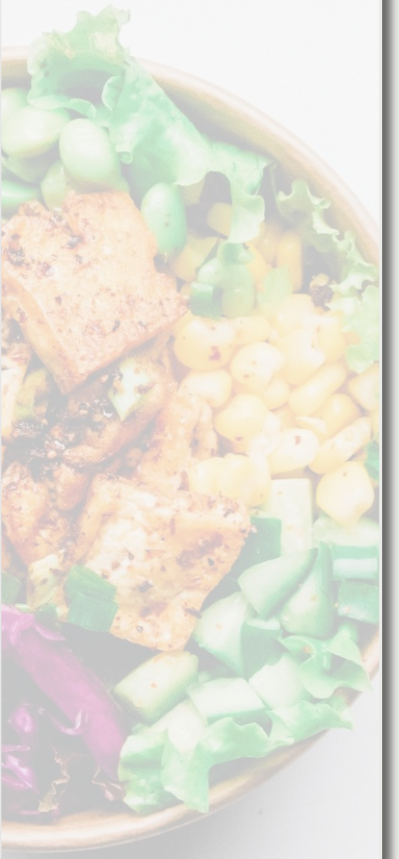
['code',
 'product_name',
 'quantity',
 'pnns_groups_2']

Format quantity, quantity outliers

Drop quantity, product_name

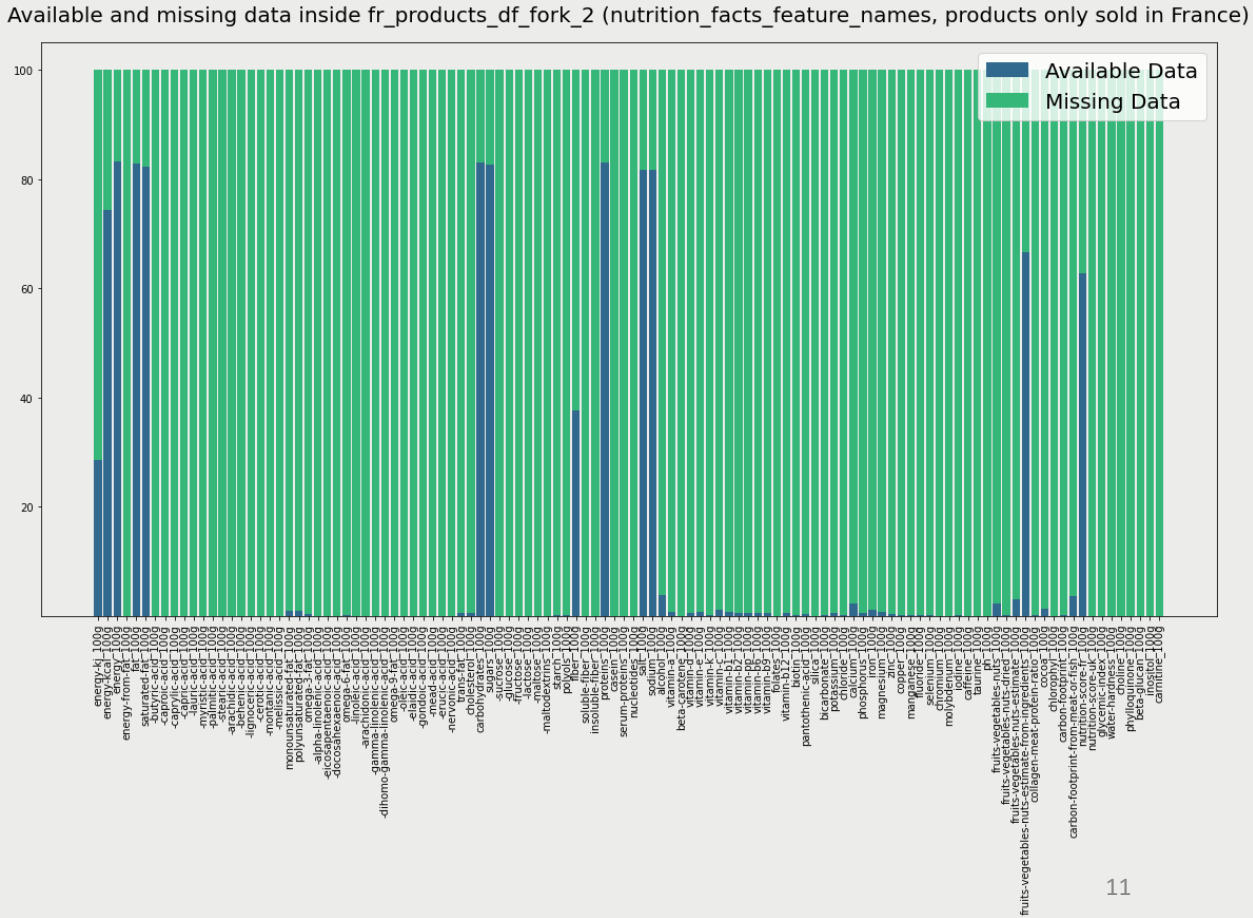


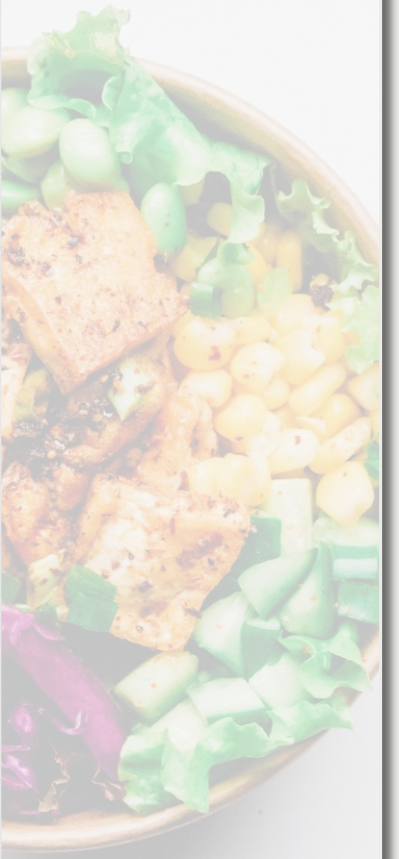
2. Sélection des features (Nutriments)



Seuil > 4%

Inspiration Nutriscore



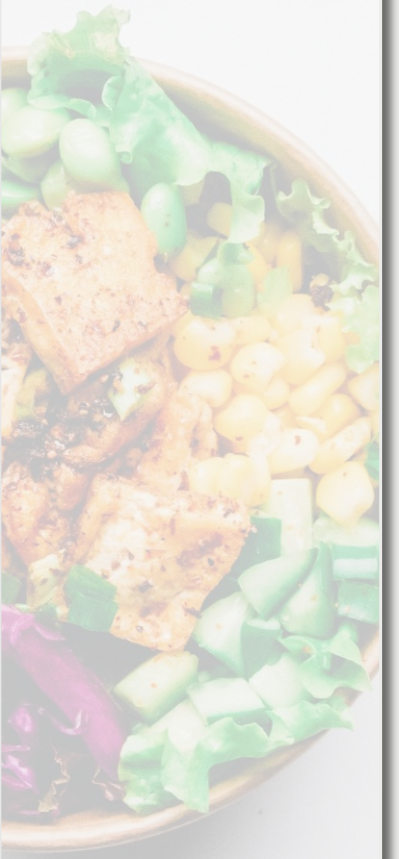


2. Sélection des features (Nutriments)

energy_100g, fat_100g, saturated-fat_100g,

carbohydrates_100g, sugars_100g, fiber_100g, proteins_100g,

salt_100g, sodium_100g, fruits-vegetables-nuts-estimate-from-ingredients_100g

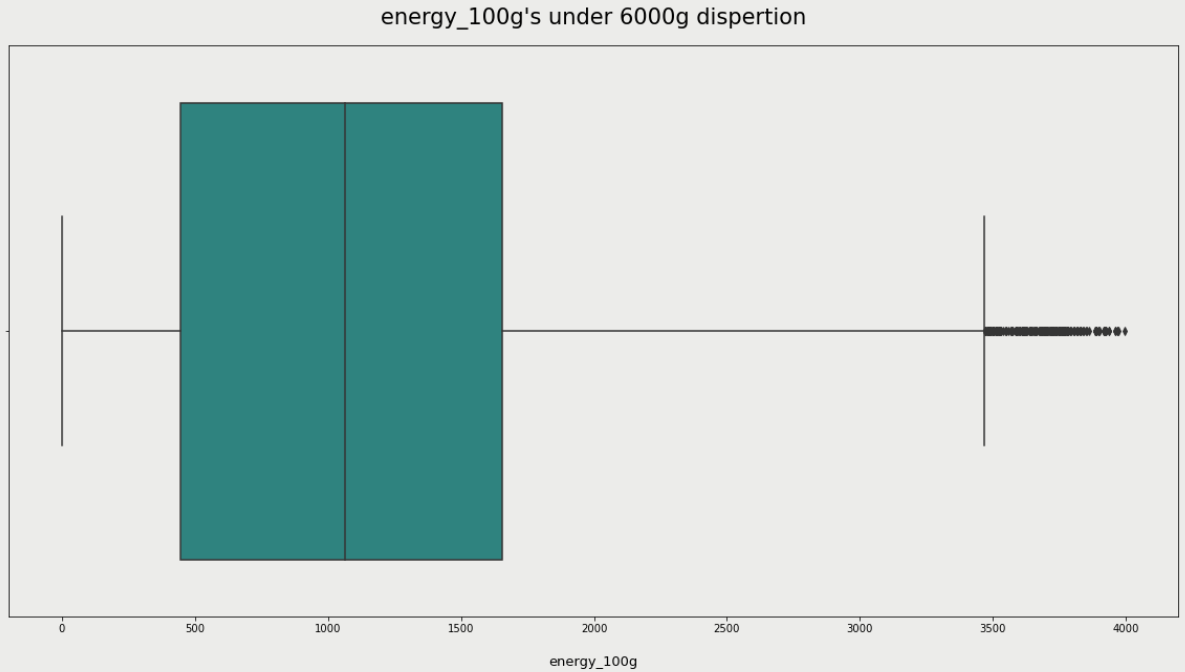


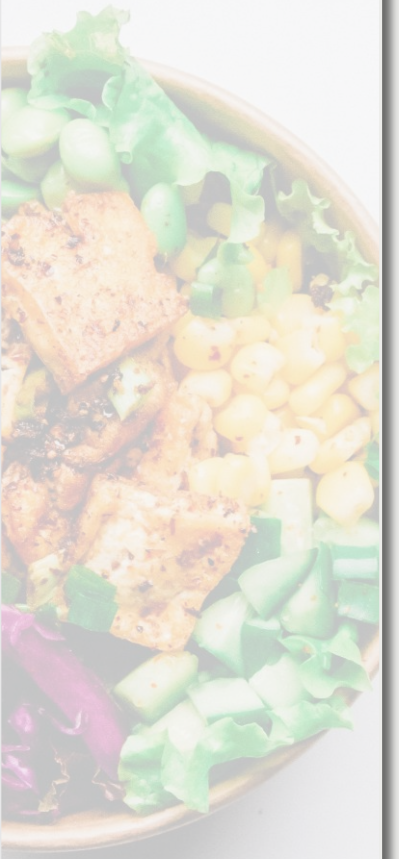
3. Outliers

Kj <= 4000

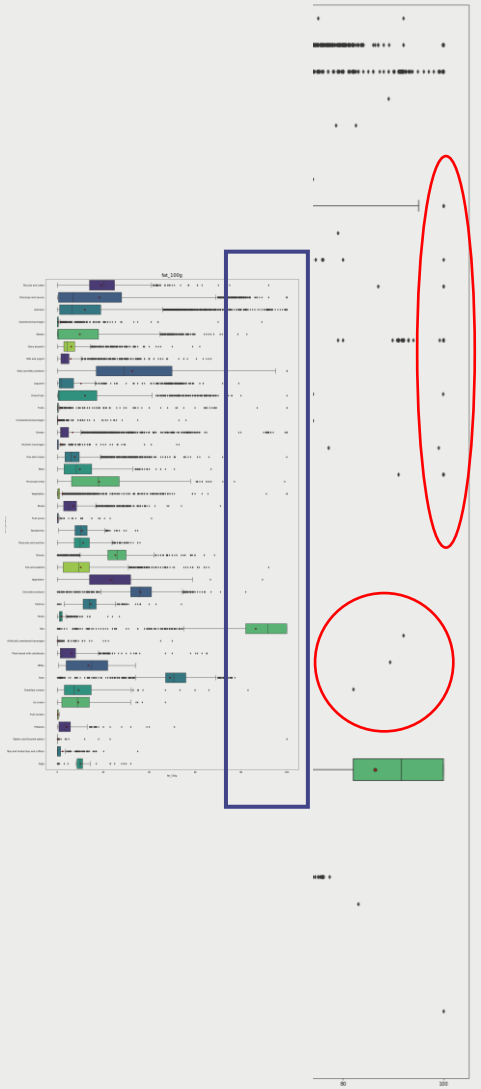
100g => [0,100]

$$\sum features \leq 100$$





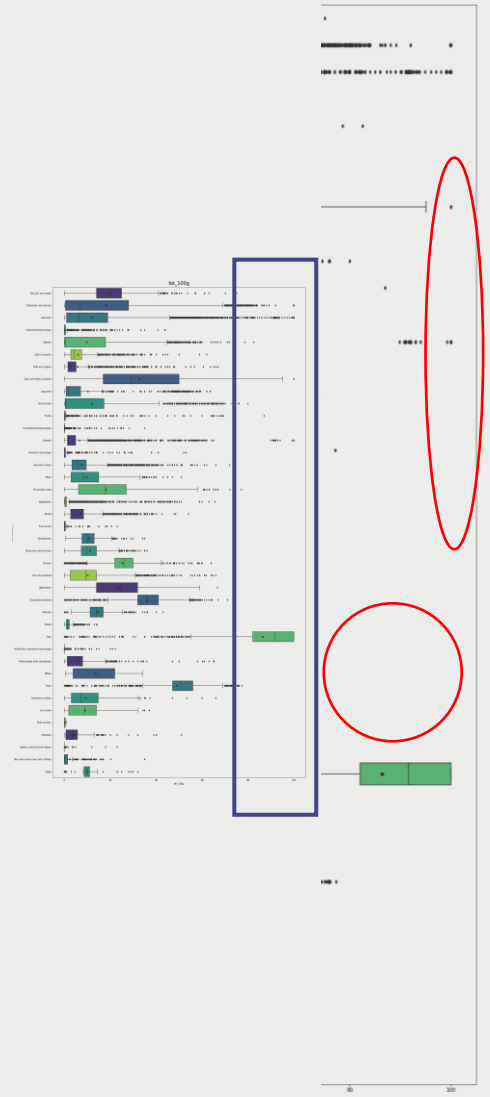
3. Outliers

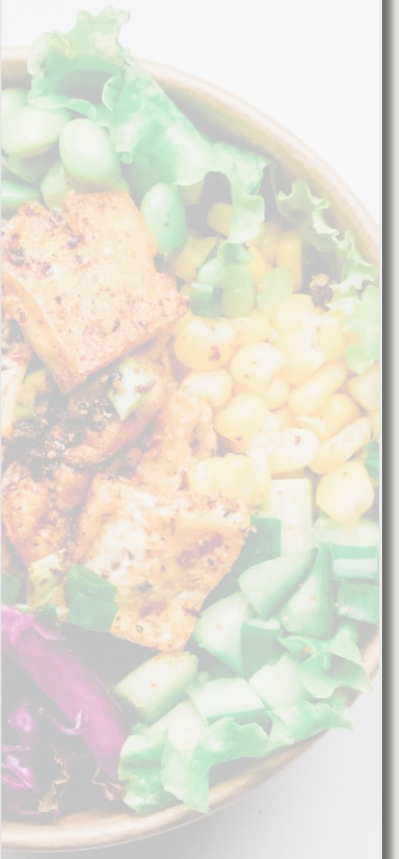


Pnns2

Local Outlier Factor

98 percentile

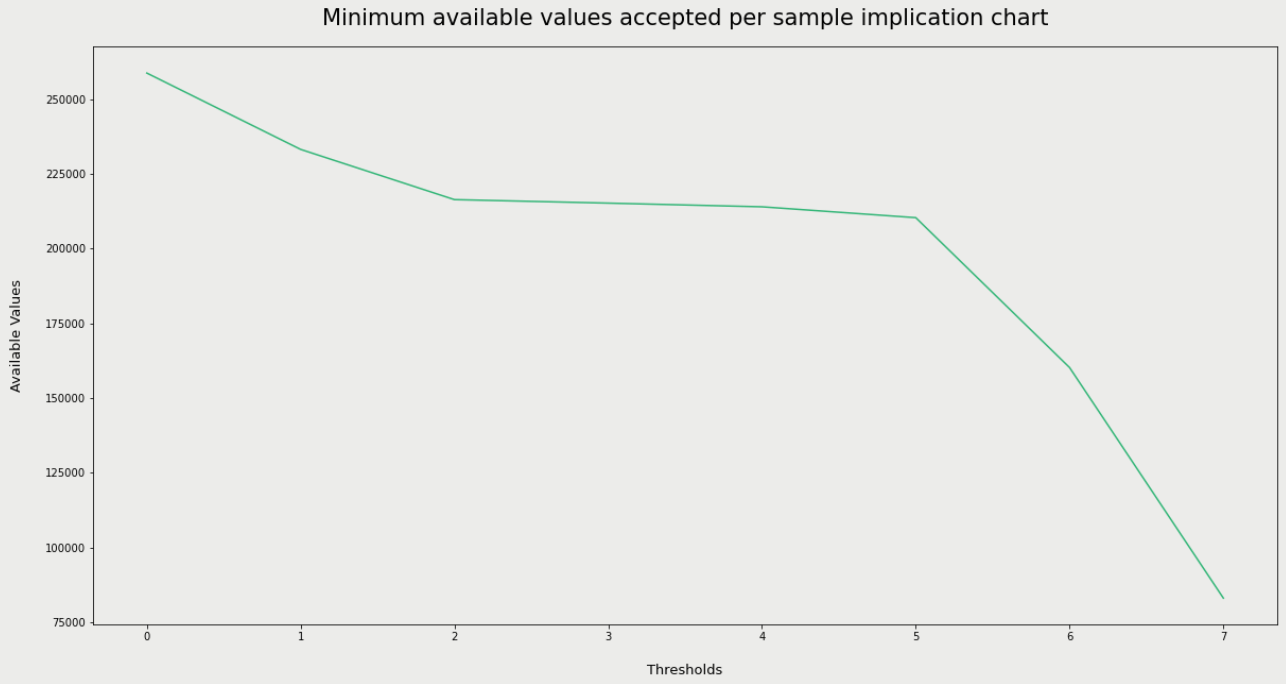




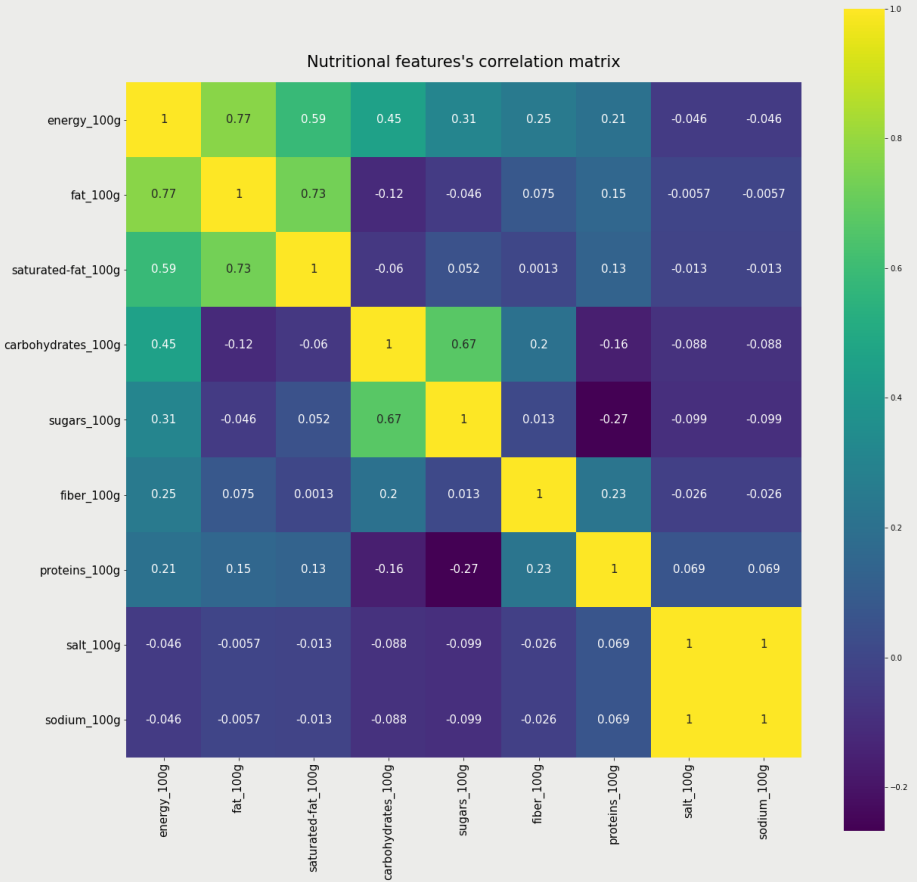
4. Valeurs Manquantes

Proprieties : Fruits Case

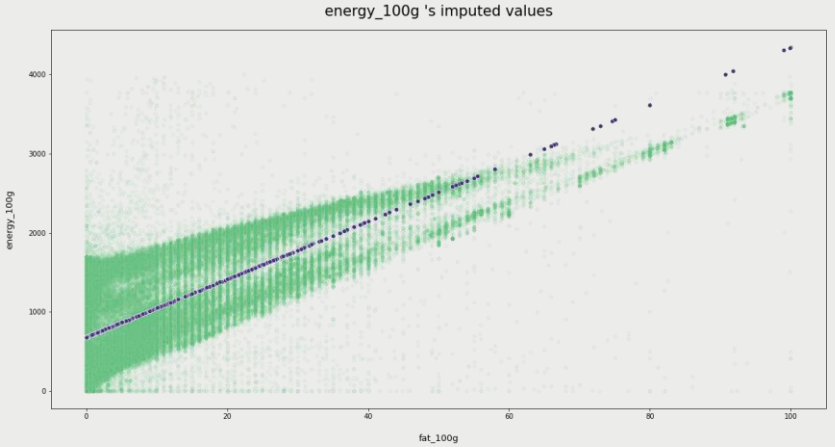
Dropping Threshold = 4

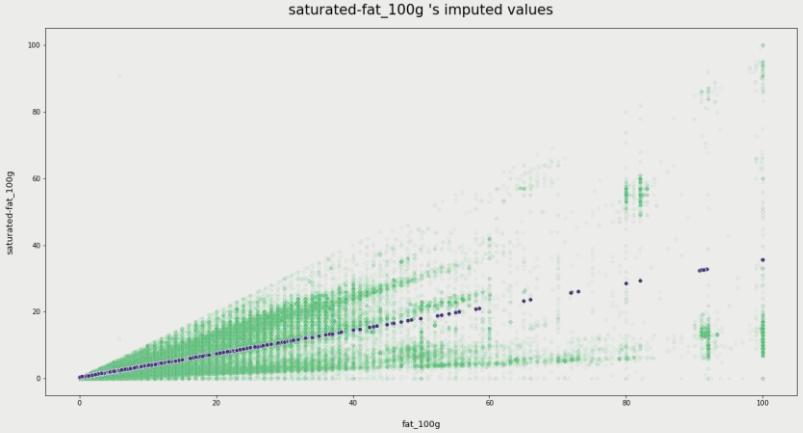


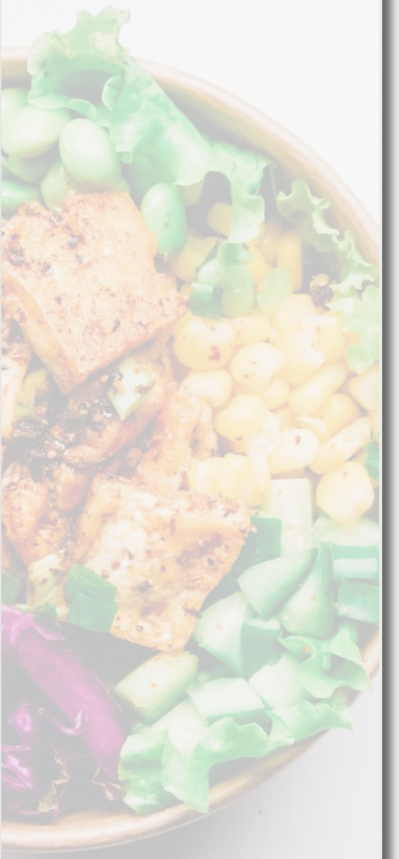
4. Valeurs Manquantes



Drop fat_100g

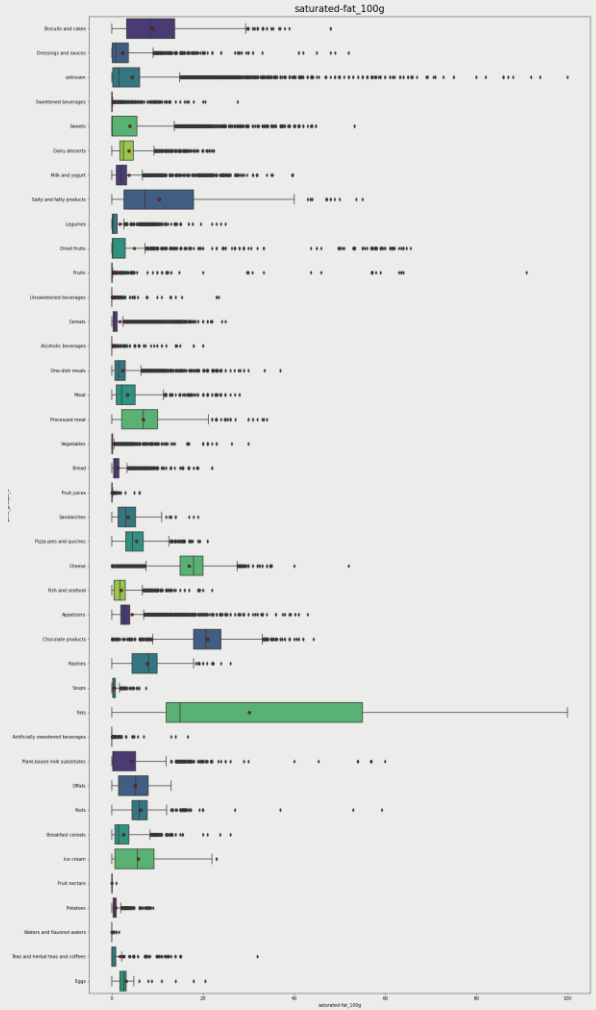


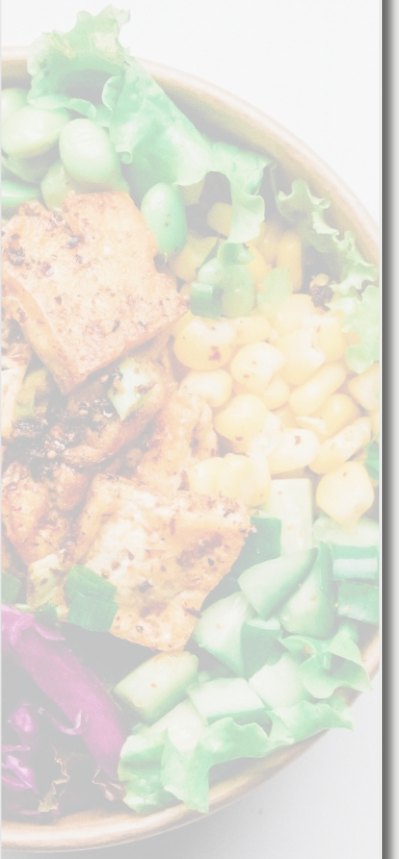




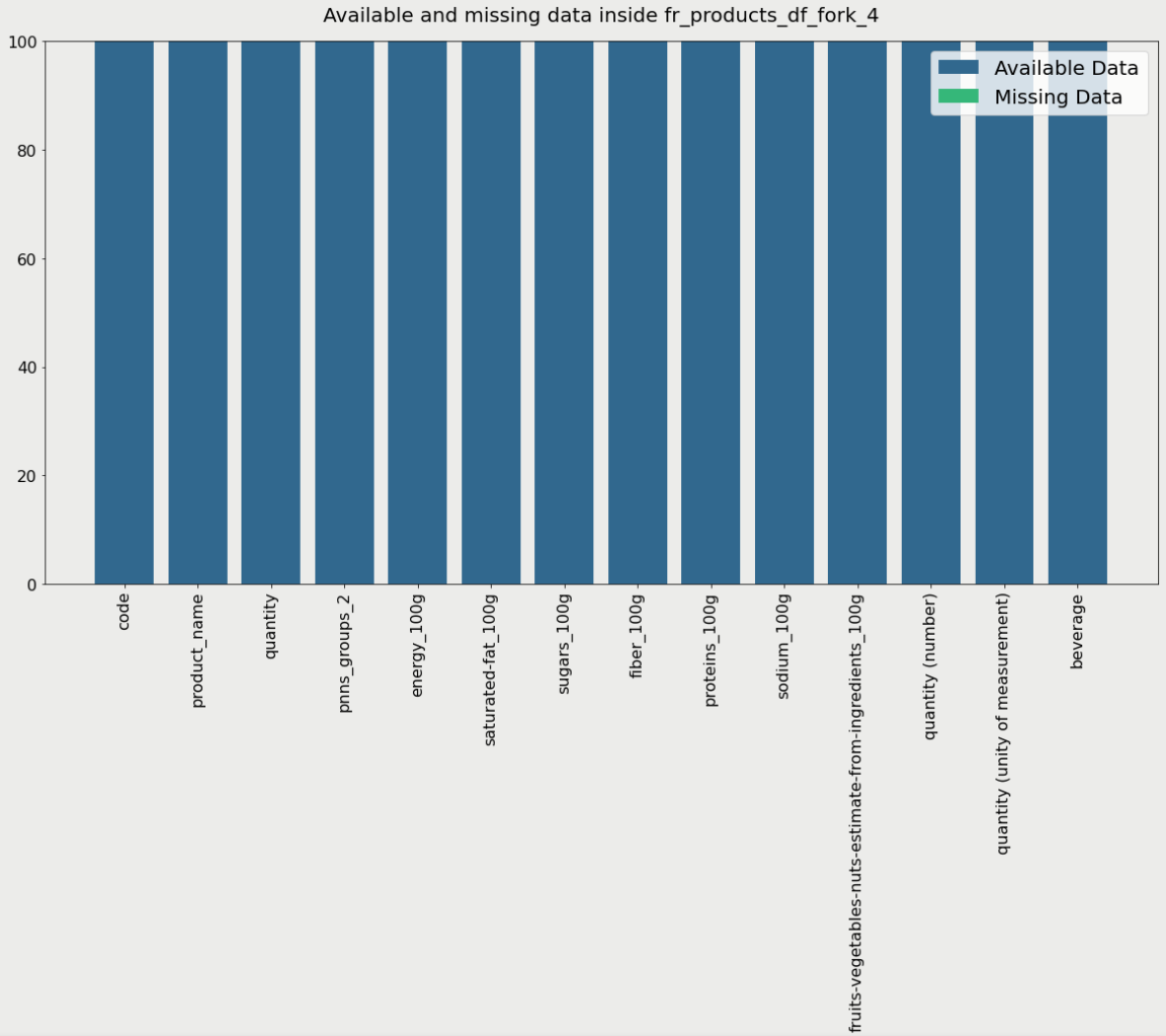
4. Valeurs Manquantes

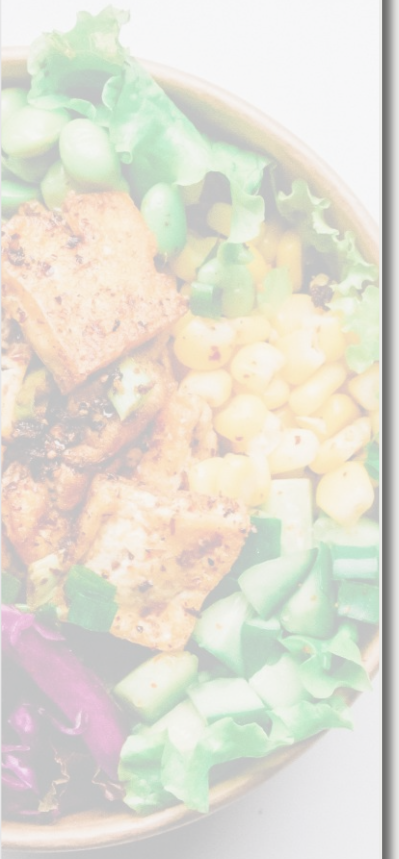
Mean and Medians





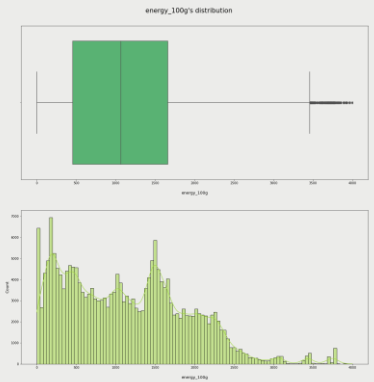
4. Valeurs Manquantes

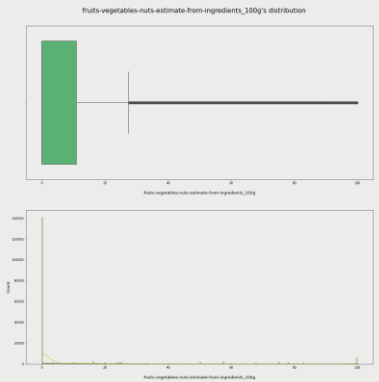


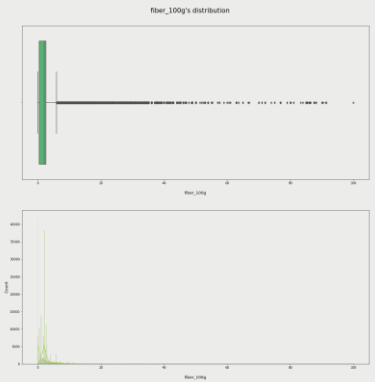


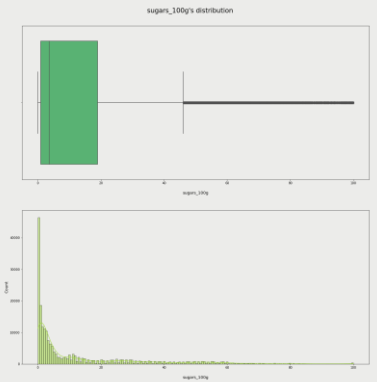
II. Analyse

1. Analyses Univariées





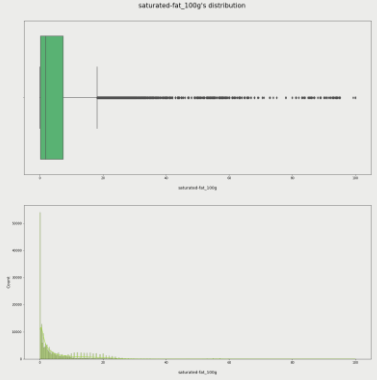


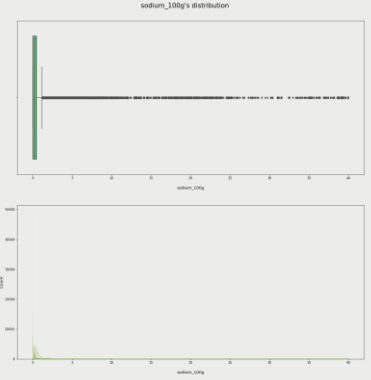


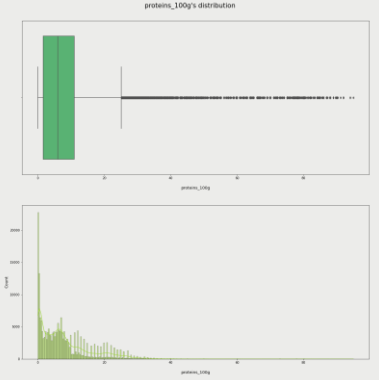
Right Skewed

Atypical Values

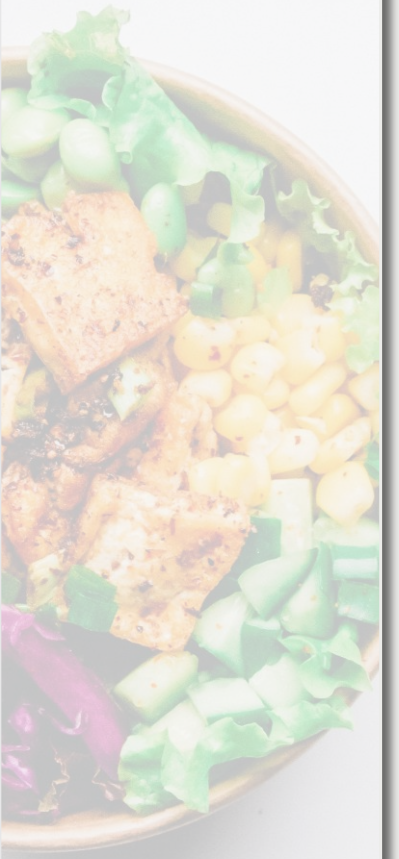
Intuition

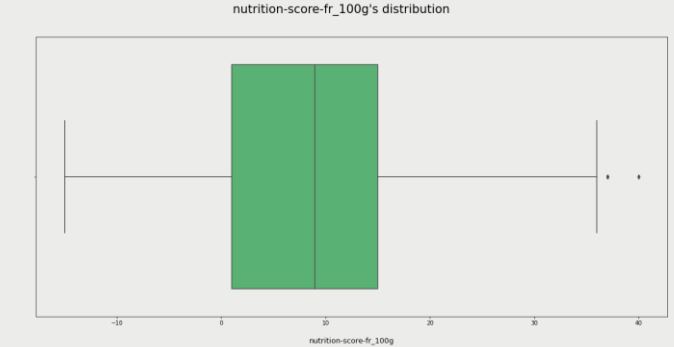


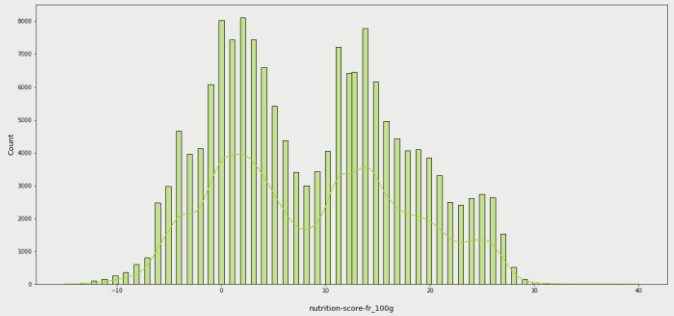


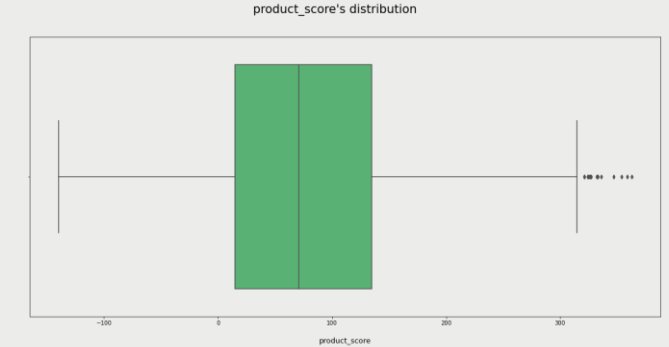


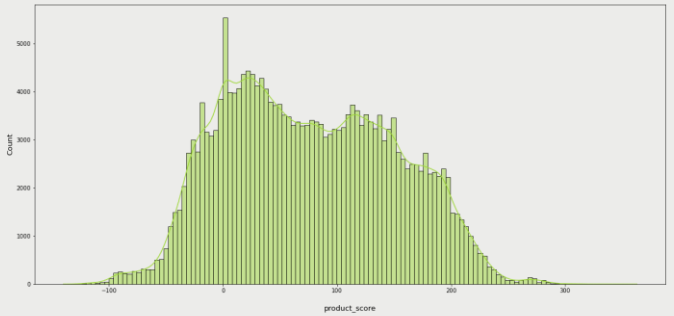
1. Analyses Univariées

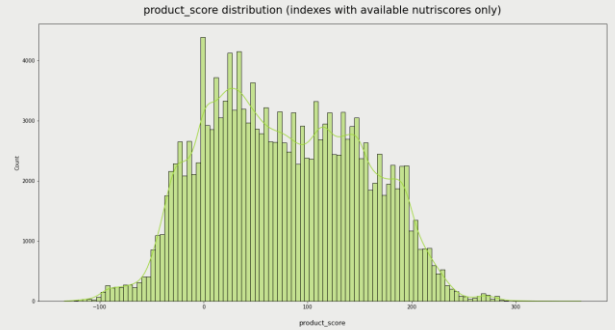










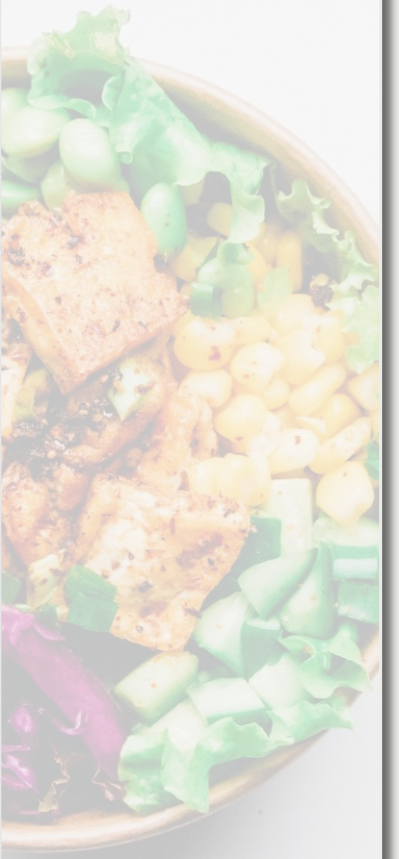


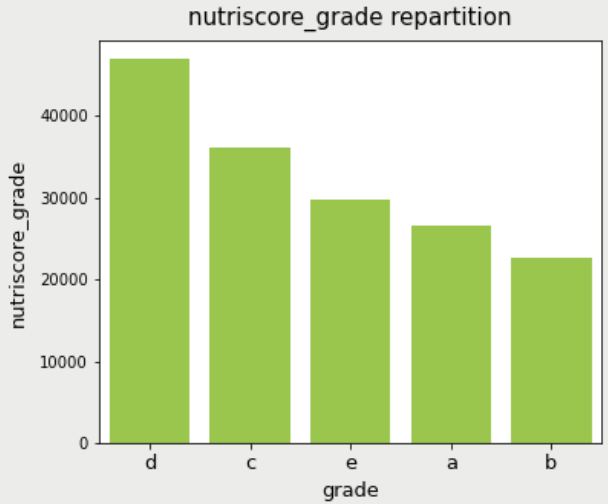
Multimodal, Gap, ⟦-15;40⟧

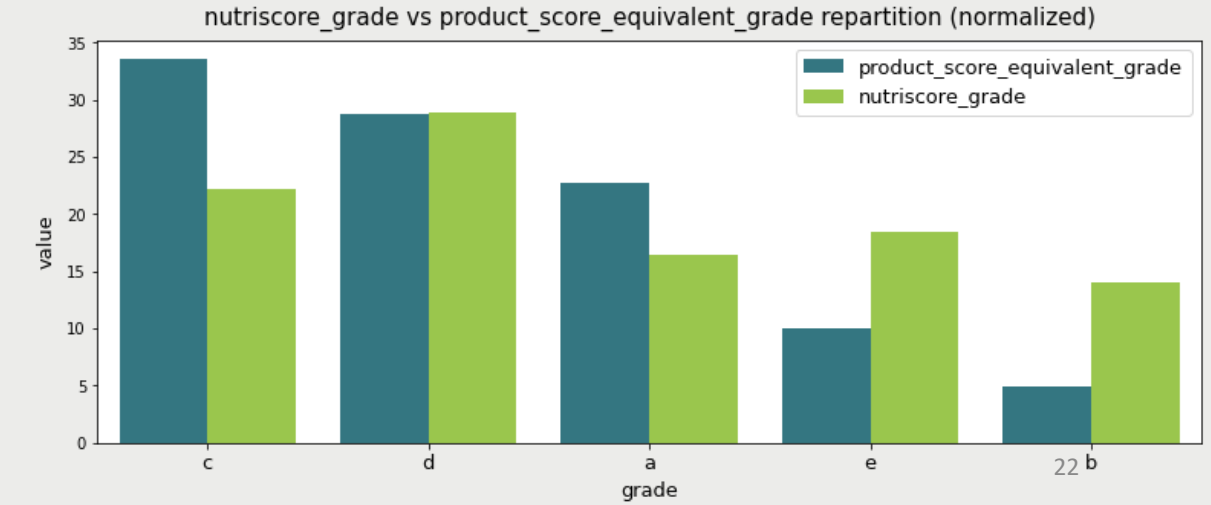
Product score, interpolation ⟦-150;400⟧

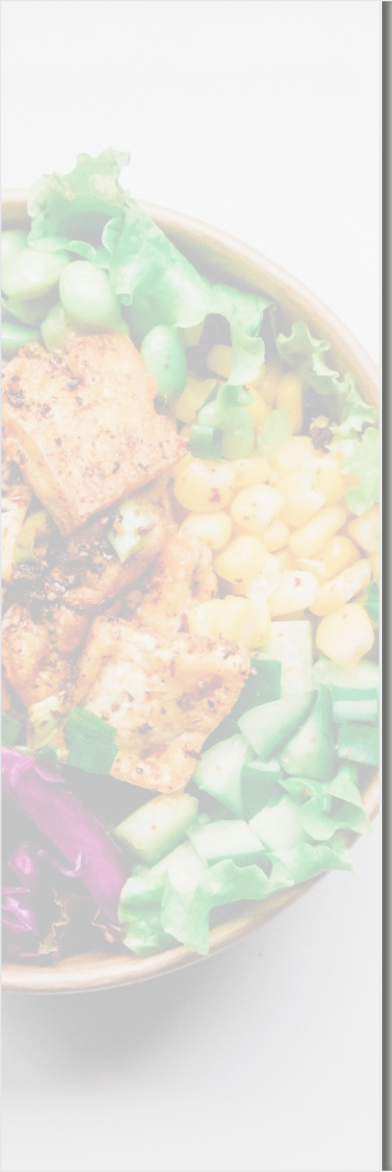
1. Analyses Univariées

« Moyennisation »





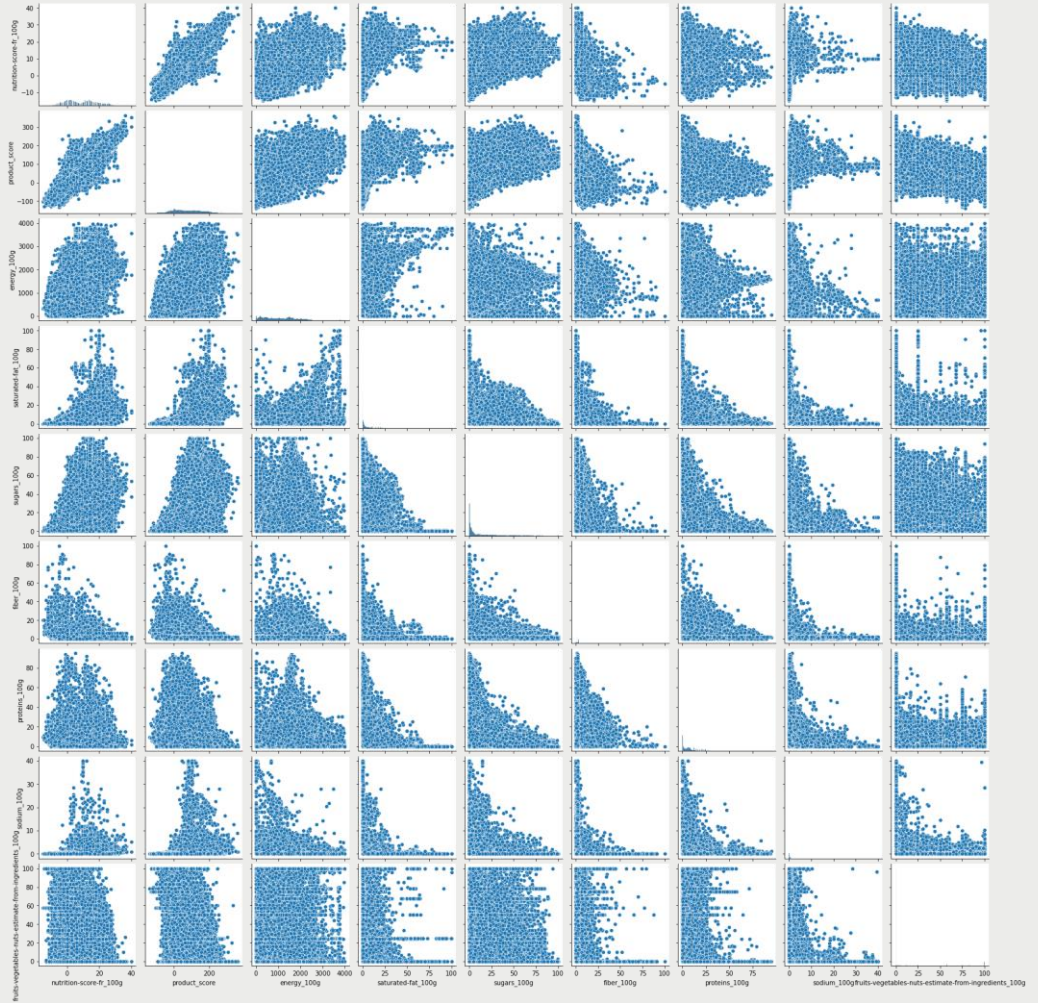




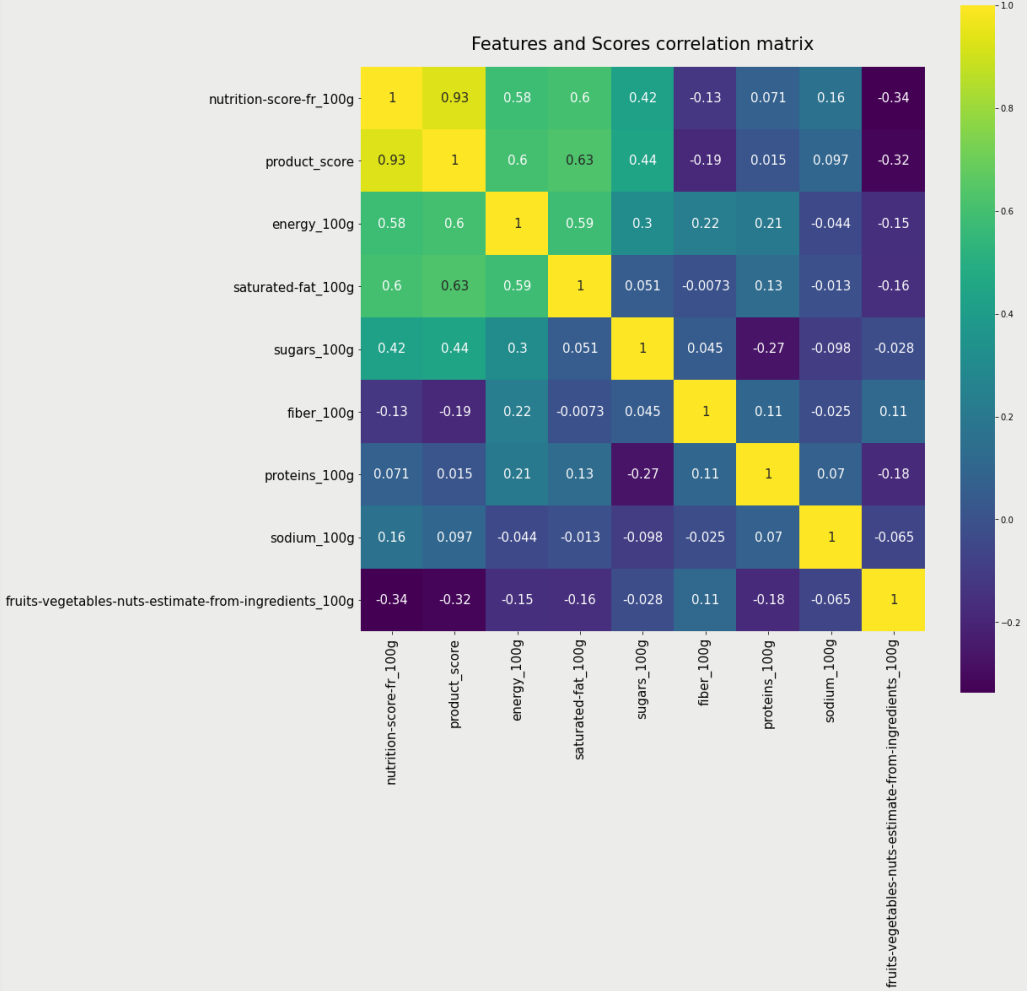
2. Analyses Bivariées

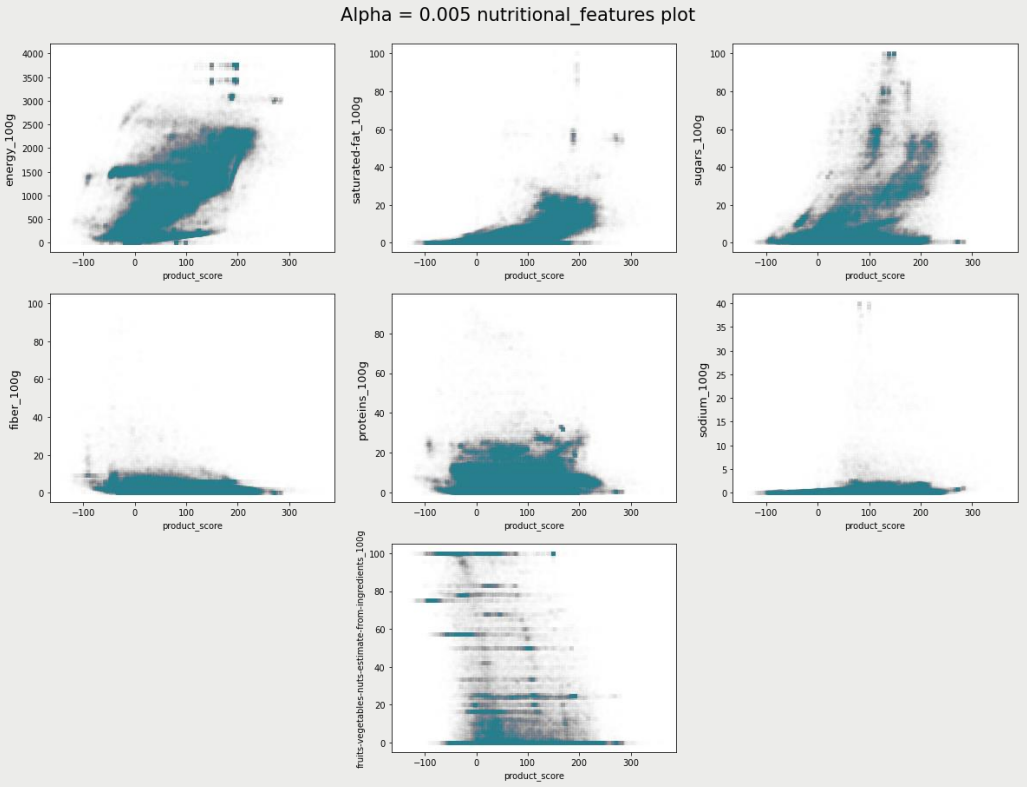
No no linear correlations

Scores



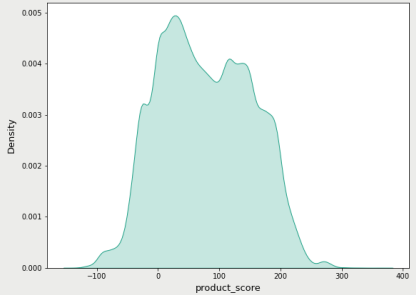
2. Analyses Bivariées

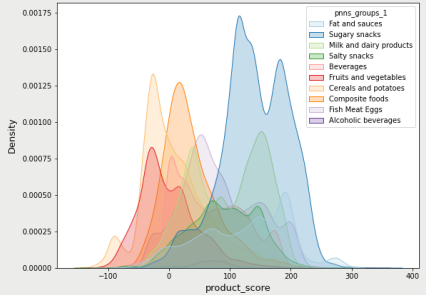




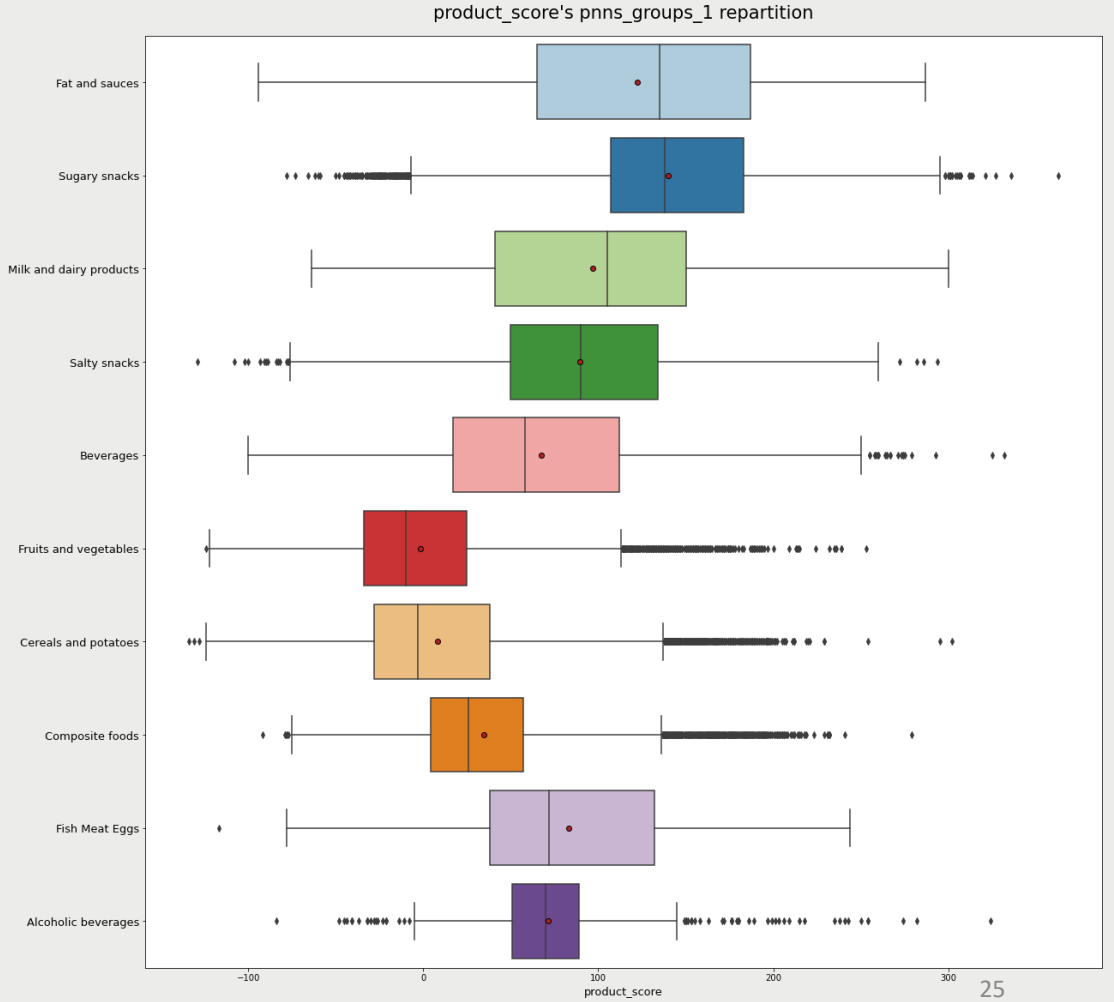
Scores r > 0,7

2. Analyses Bivariées

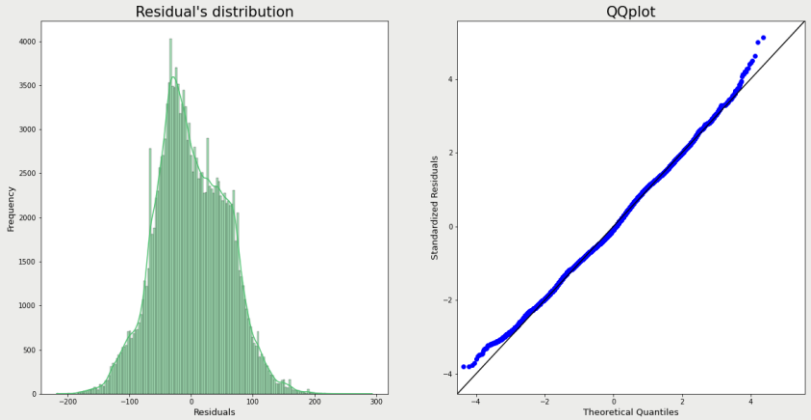




pnns1

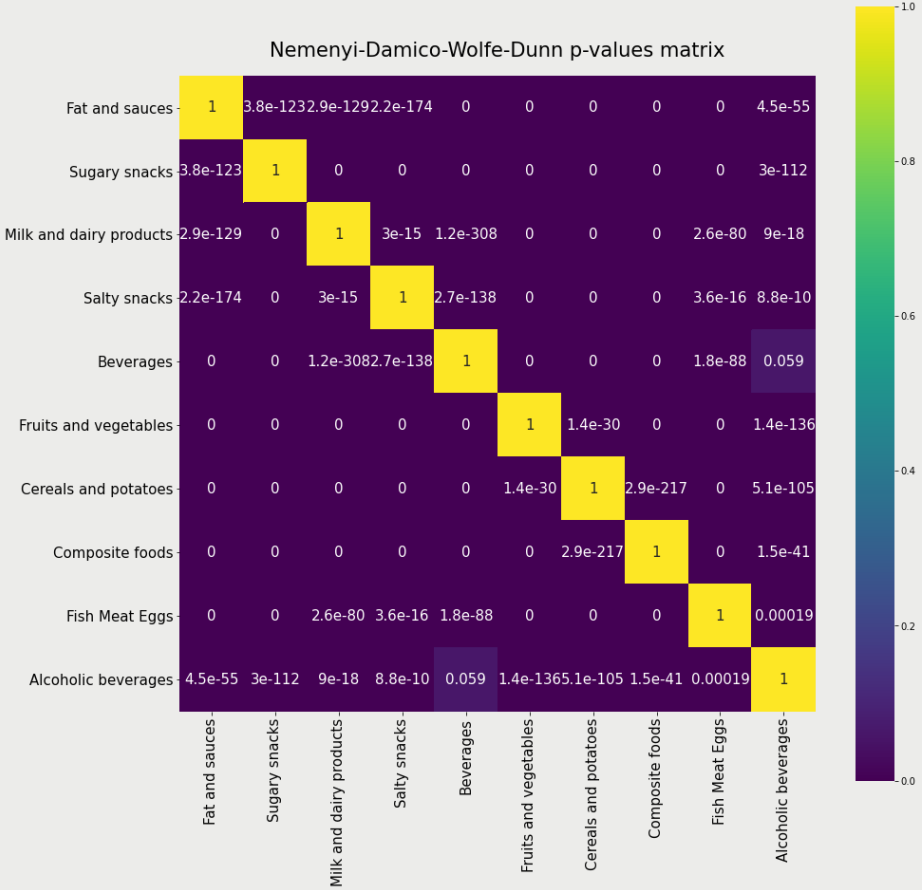


2. Analyses Bivariées

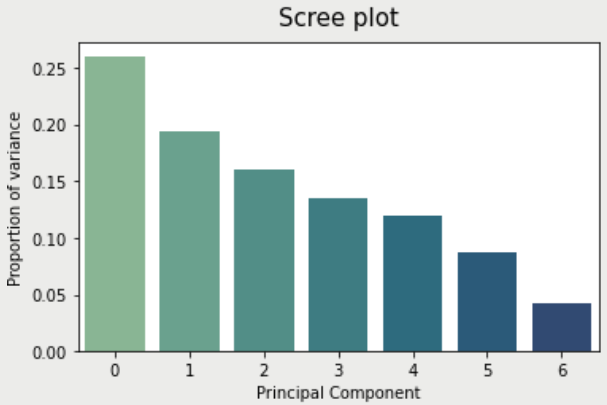


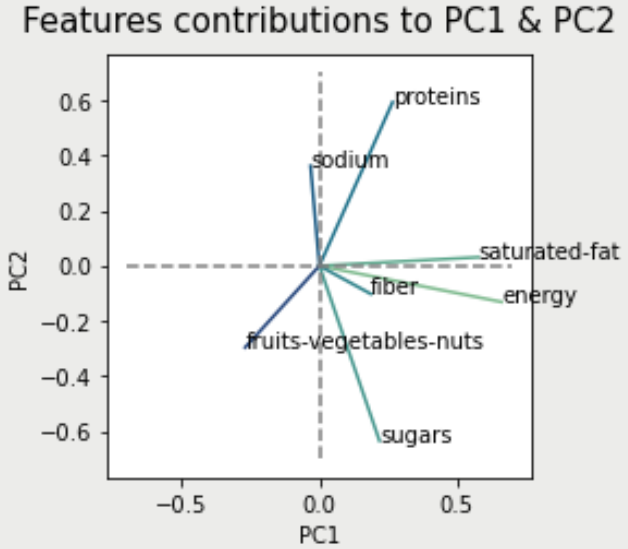
Anova
Komogorov-Smirnov
Levene

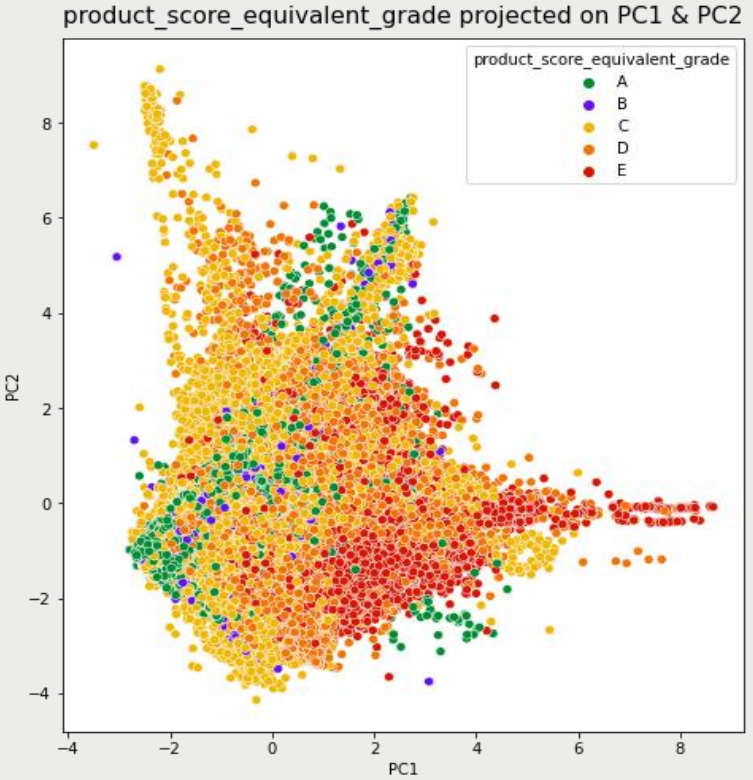
Kruskal-Wallis
Nemenyi-Damico-Wolfe-Dunn



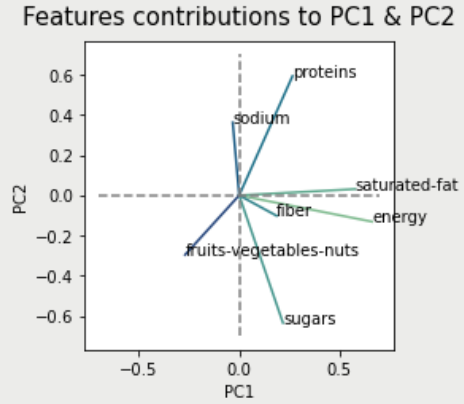
3. Analyses Multivariées

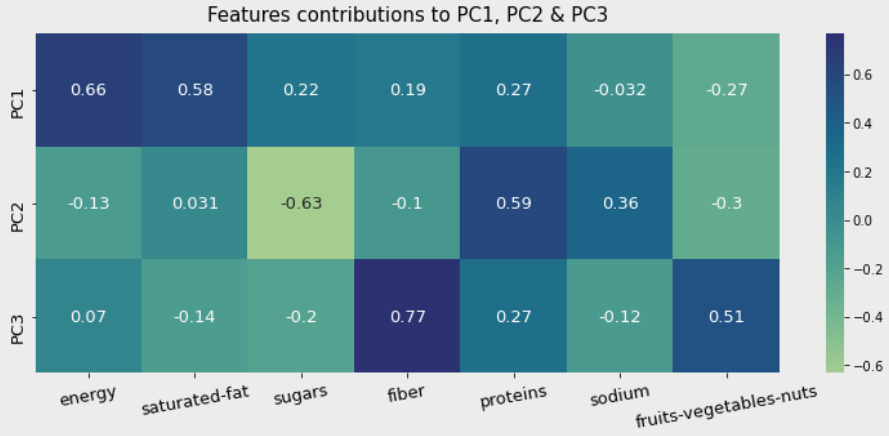


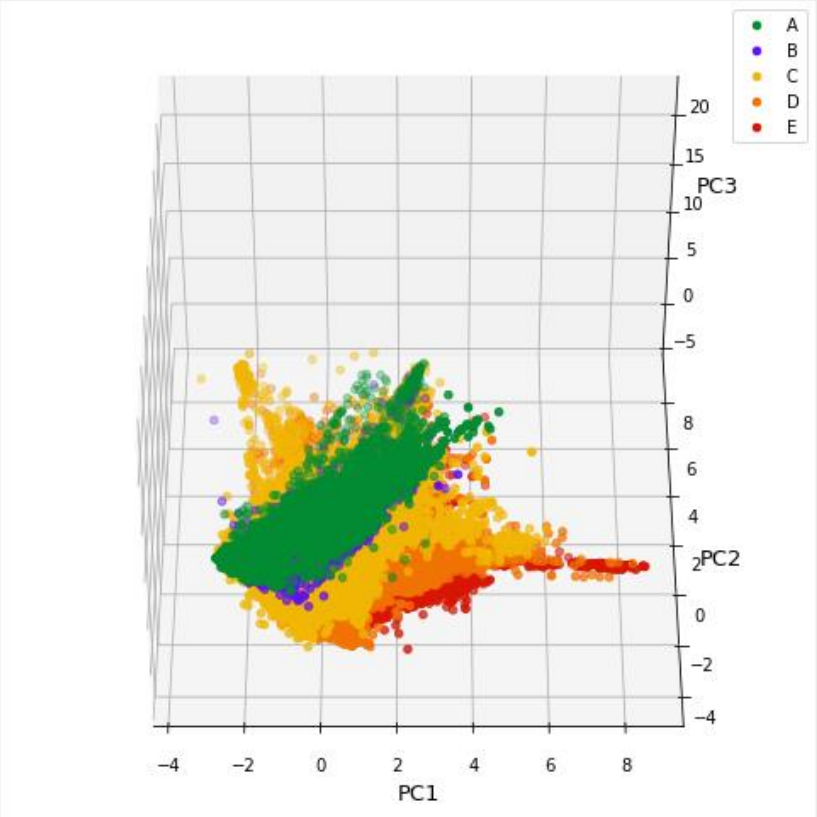




3. Analyses Multivariées







Merci pour votre attention !