

Concevez une application au service de la santé publique

OPENCLASSROOMS

Bayram DONAT

Sommaire

- idée d'application
- opérations de nettoyage effectuées.
- description et analyse univariée des différentes variables importantes
- analyse multivariée et résultats statistiques associés
- 3 observations évaluant la pertinence du projet
- faisabilité du projet

1. Mon idée d'application

- Trouver des aliments équivalents
 - En fonction des ingrédients
 - Dans la même catégorie
 - Vendu dans le même pays
 - Avec un nutriscore meilleur

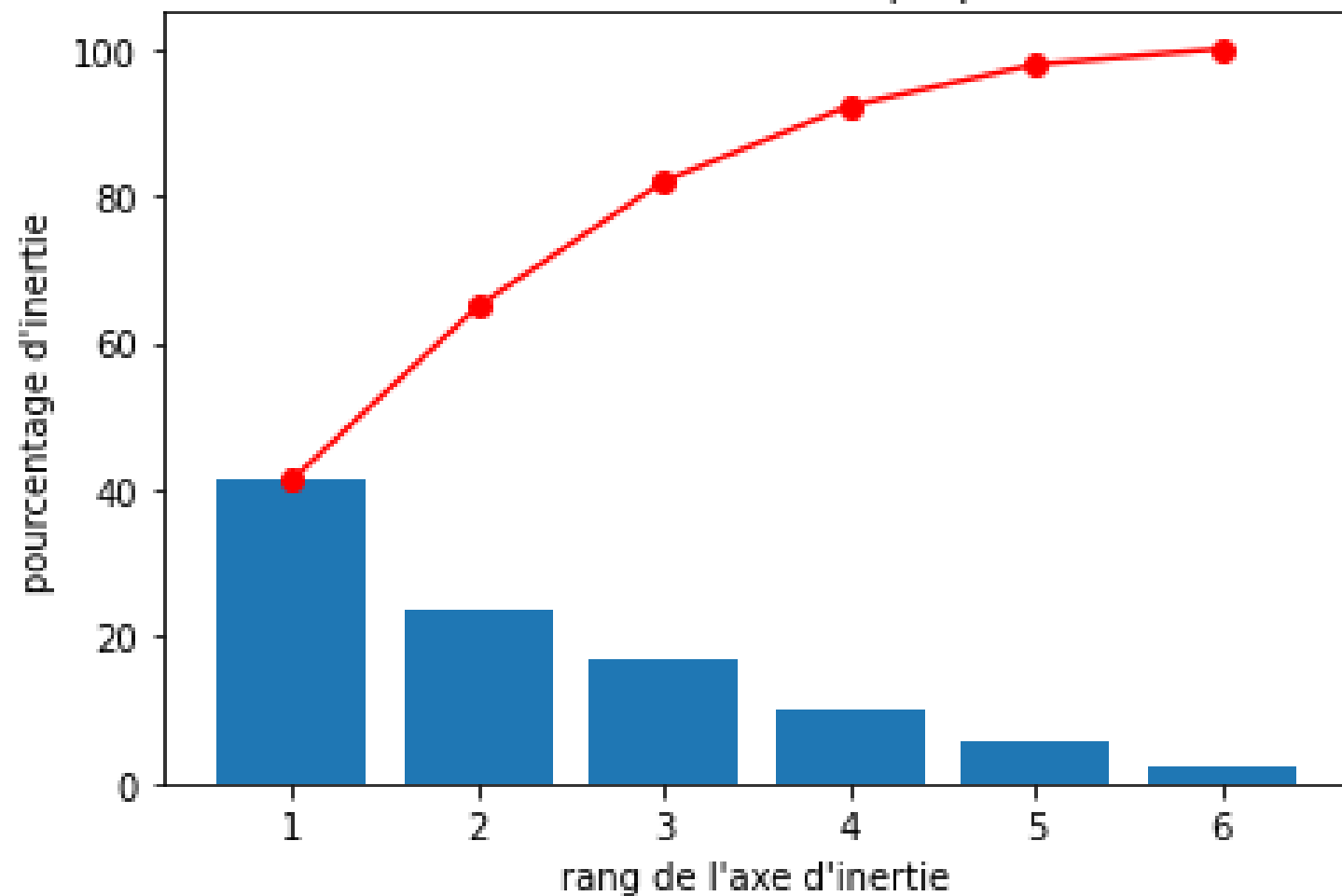
2. Nettoyage de données

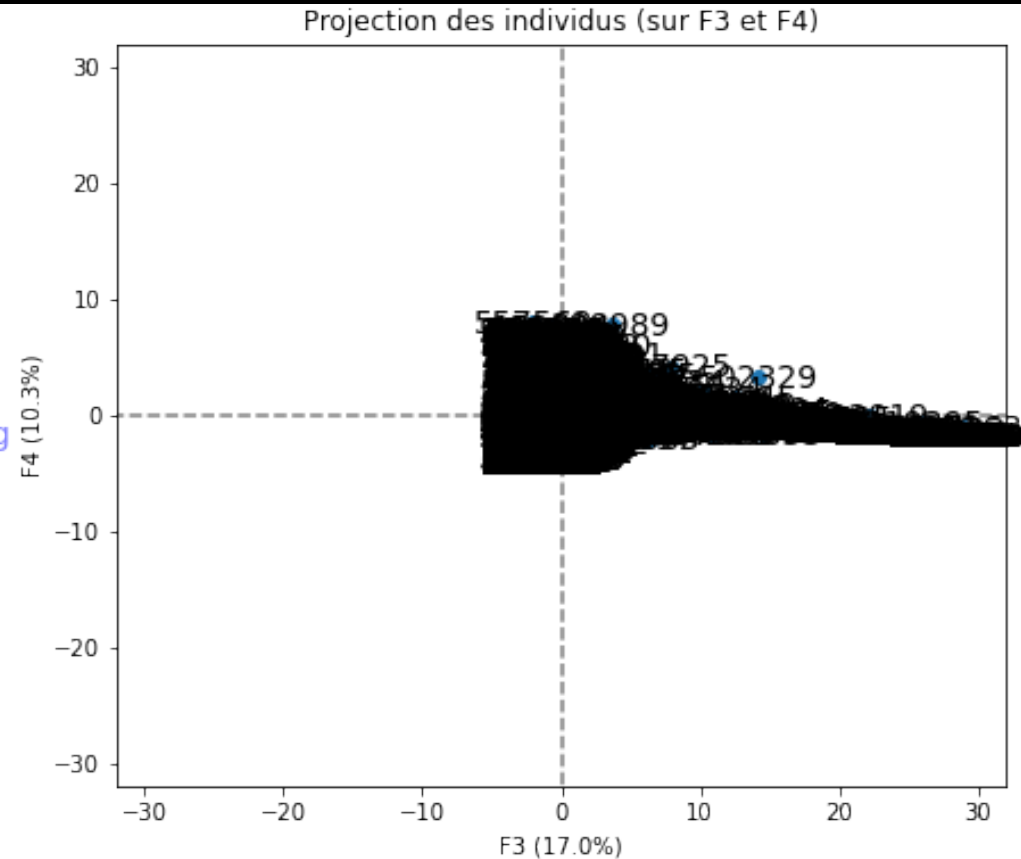
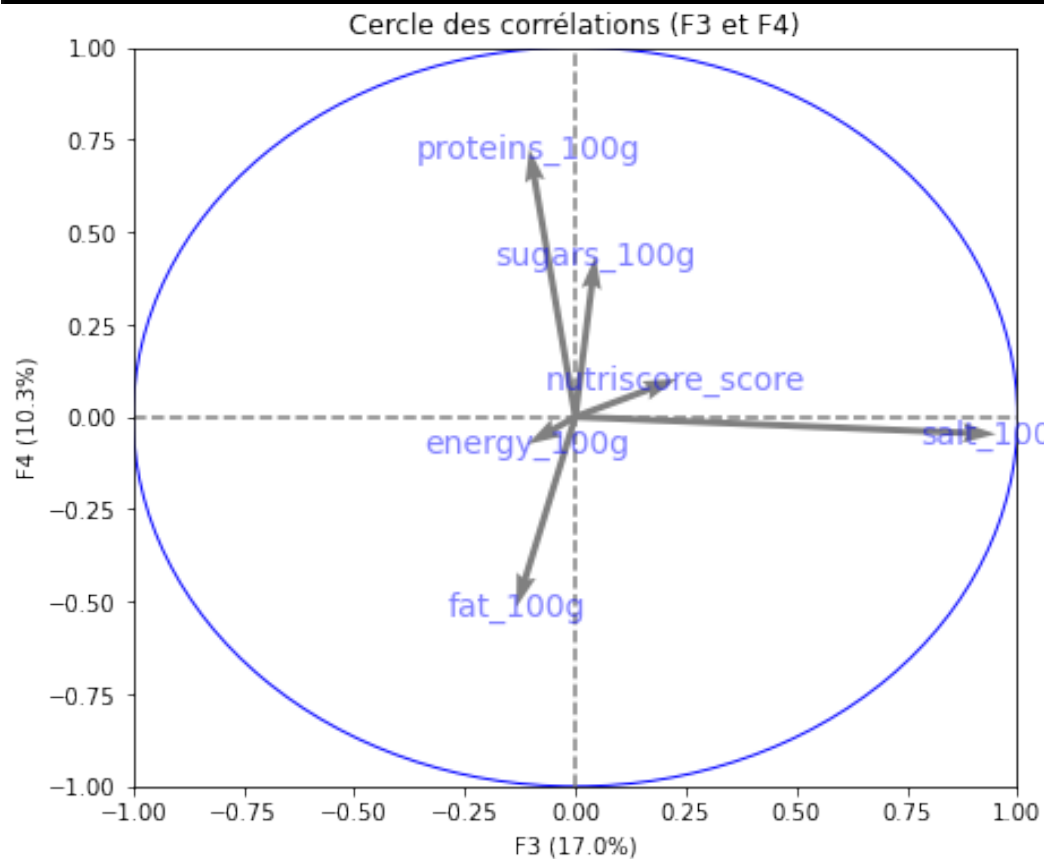
- Critères choisies
 - 6 Variables quantitatives
 - nutriscore_score
 - energy_100g, fat_100g, sugars_100g, proteins_100g, salt_100g
 - 4 variables qualitatives :
 - nutriscore_score
 - pnns_groups_1, pnns_groups_2, main_category_en

3. Analyse univariée

- Variables quantitatives
 - 0 à 100g pour les données nutritionnelles en g
 - 0 à 900kCal et 0 à 3766J pour l'énergie
- Suppression des doublons et valeurs manquantes
- Analyse des composantes principales
 - Variables quantitatives
 - 'additives_n', 'nova_group',
 - 'ingredients_from_palm_oil_n', 'ingredients_that_may_be_from_palm_oil_n',
 - 'nutriscore_score', 'nutrition-score-fr_100g'
 - 'energy_100g', 'sugars_100g', 'fiber_100g', 'proteins_100g',
 - 'fat_100g', 'saturated-fat_100g',
 - 'salt_100g', 'sodium_100g',

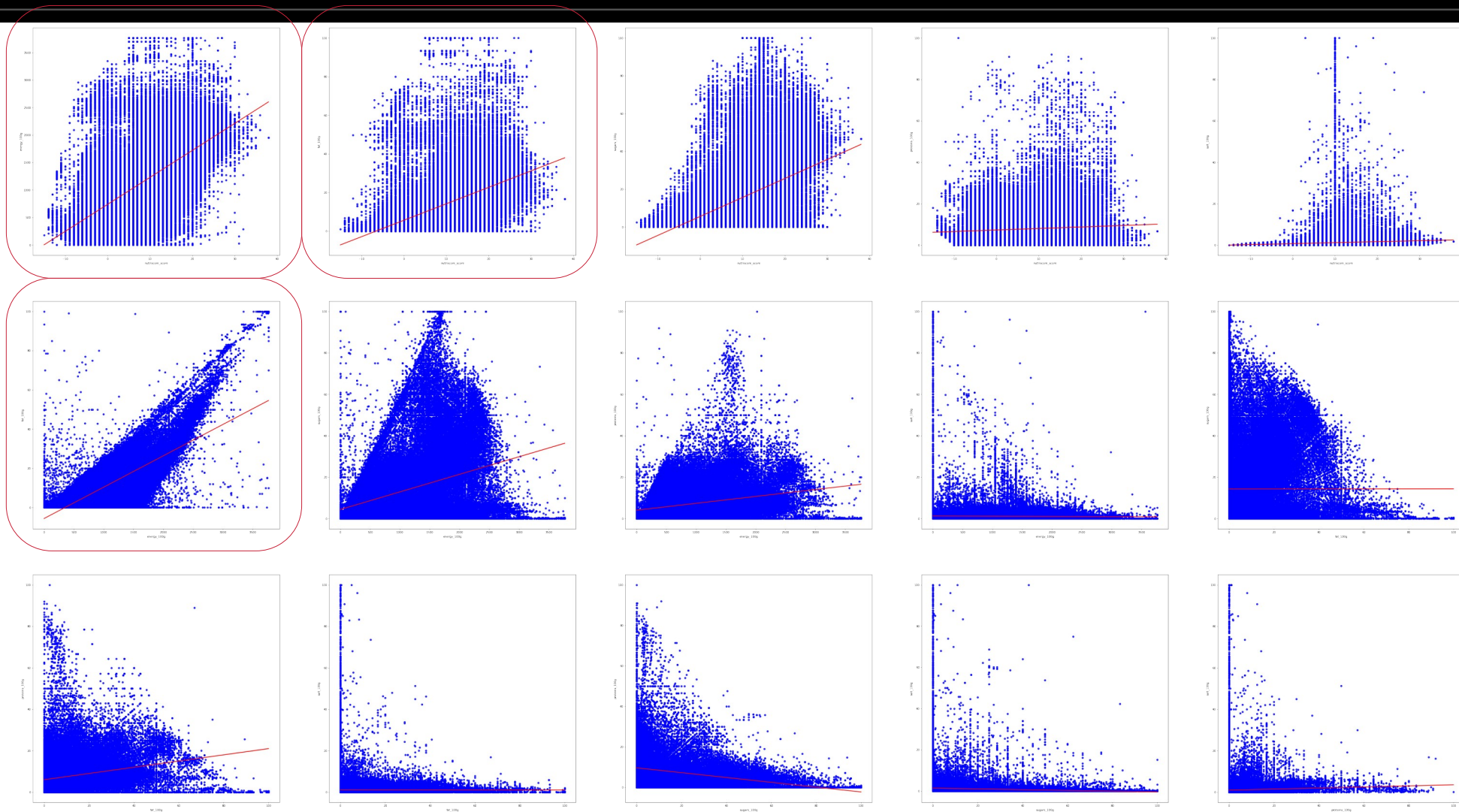
Eboulis des valeurs propres

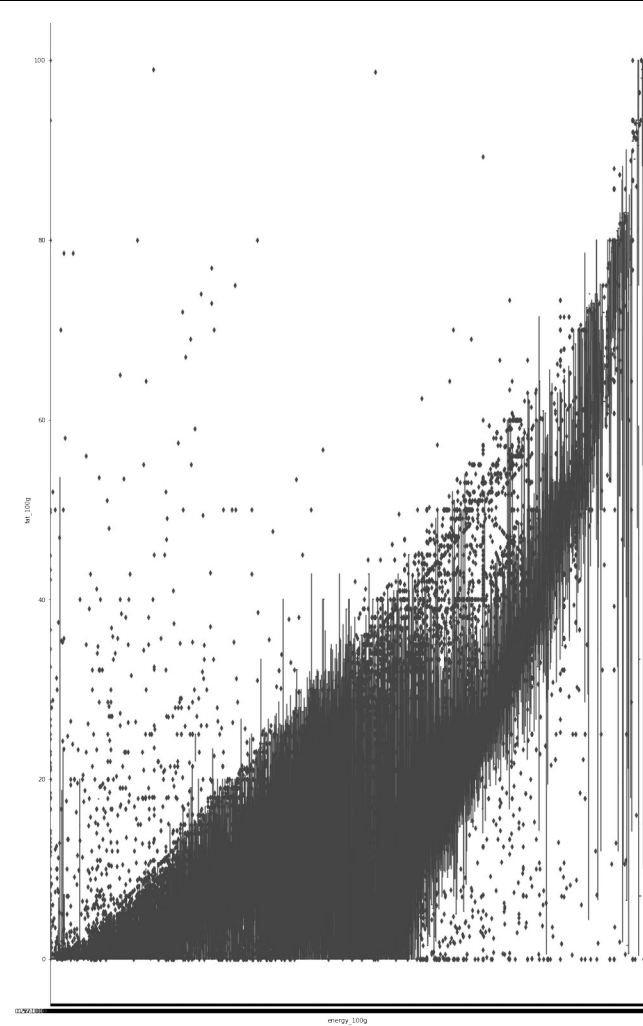
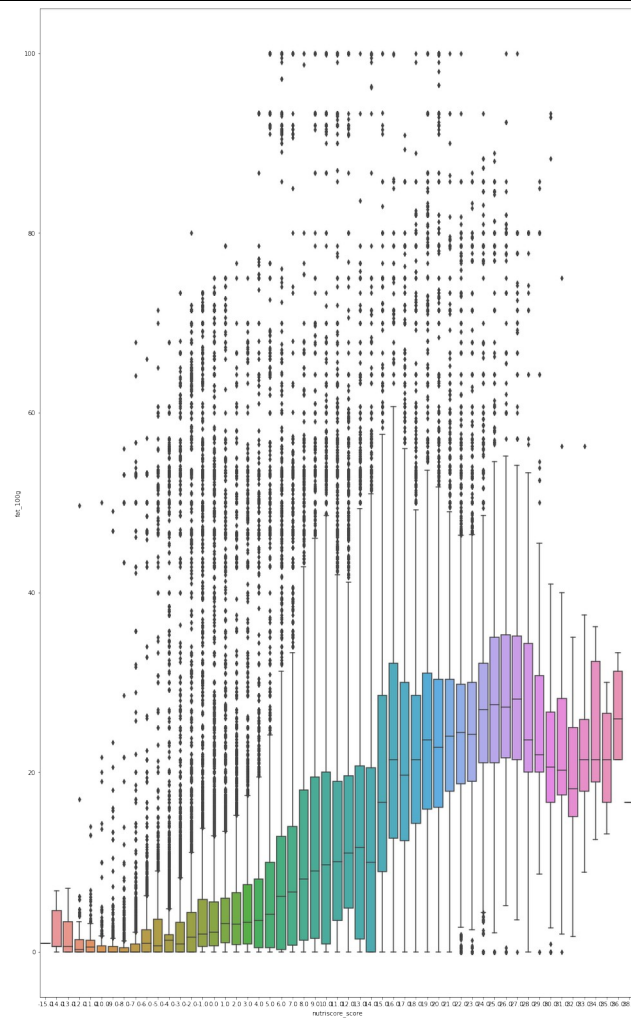
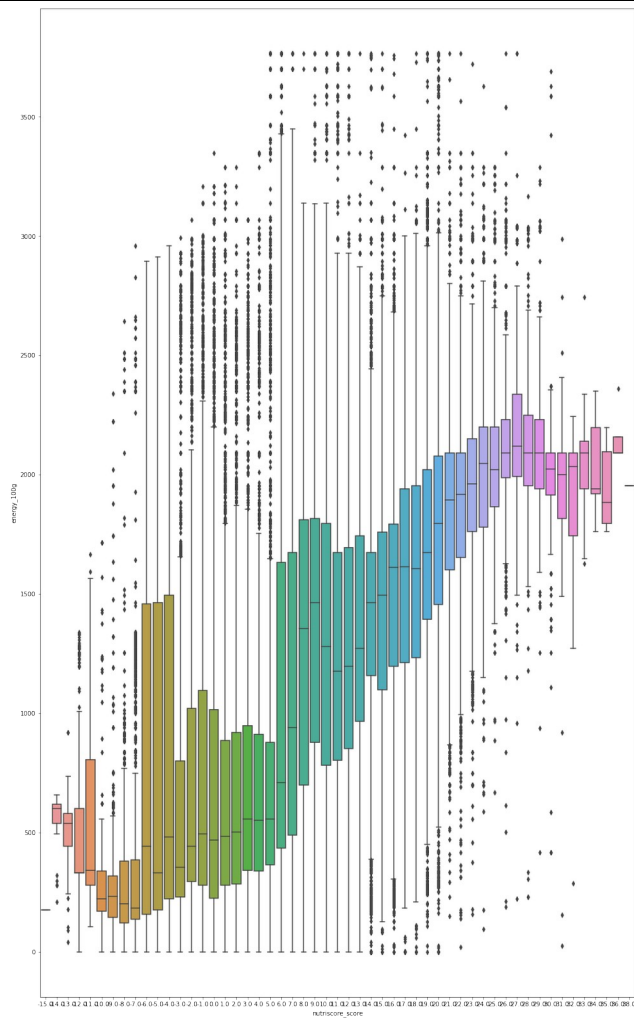




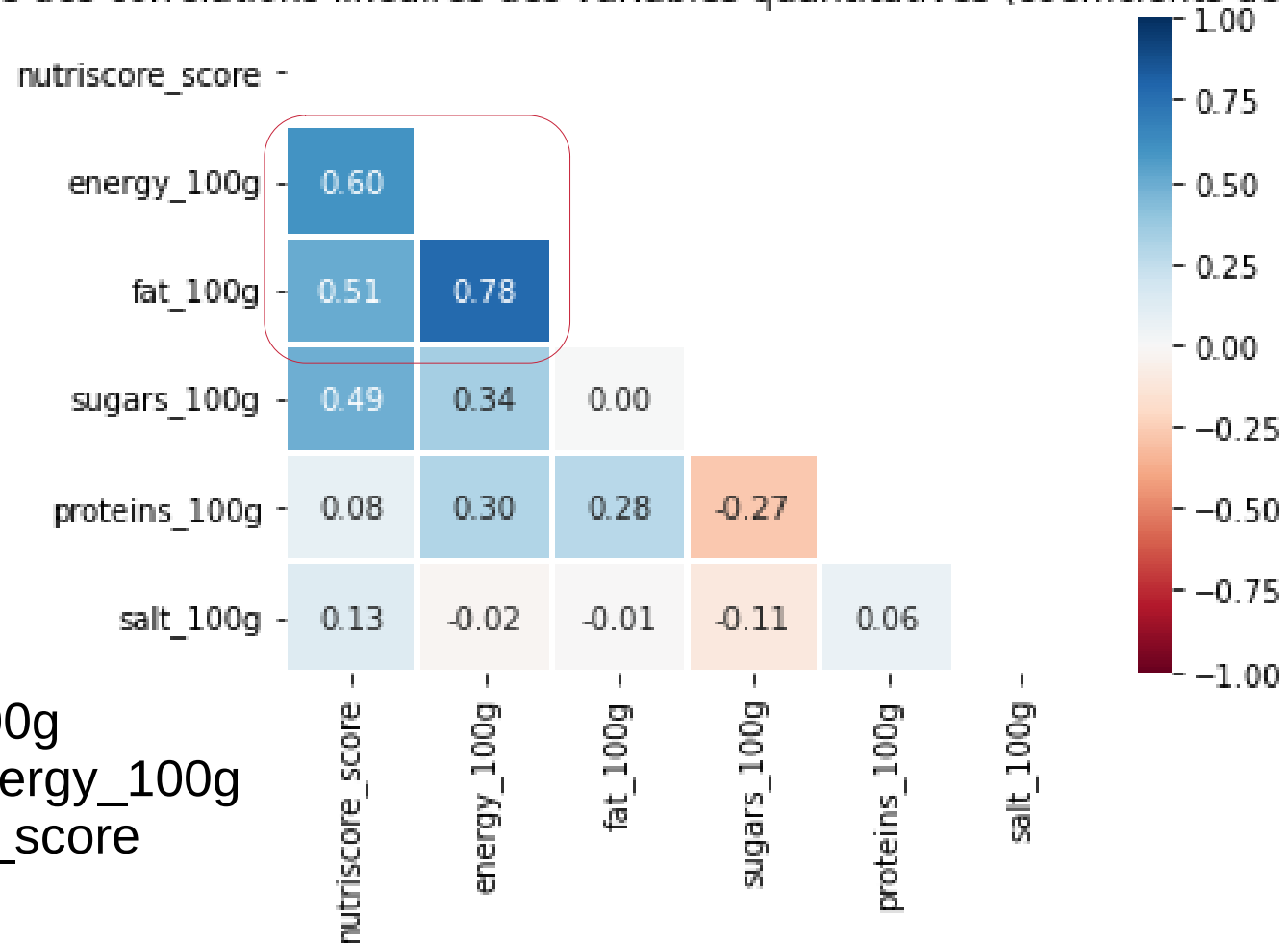
4. Analyse multivariée

- 2 variables quantitatives
 - Nuages de points avec régression linéaire
 - Boîtes à moustaches
 - Coefficient de corrélation de Pearson
 - Covariance



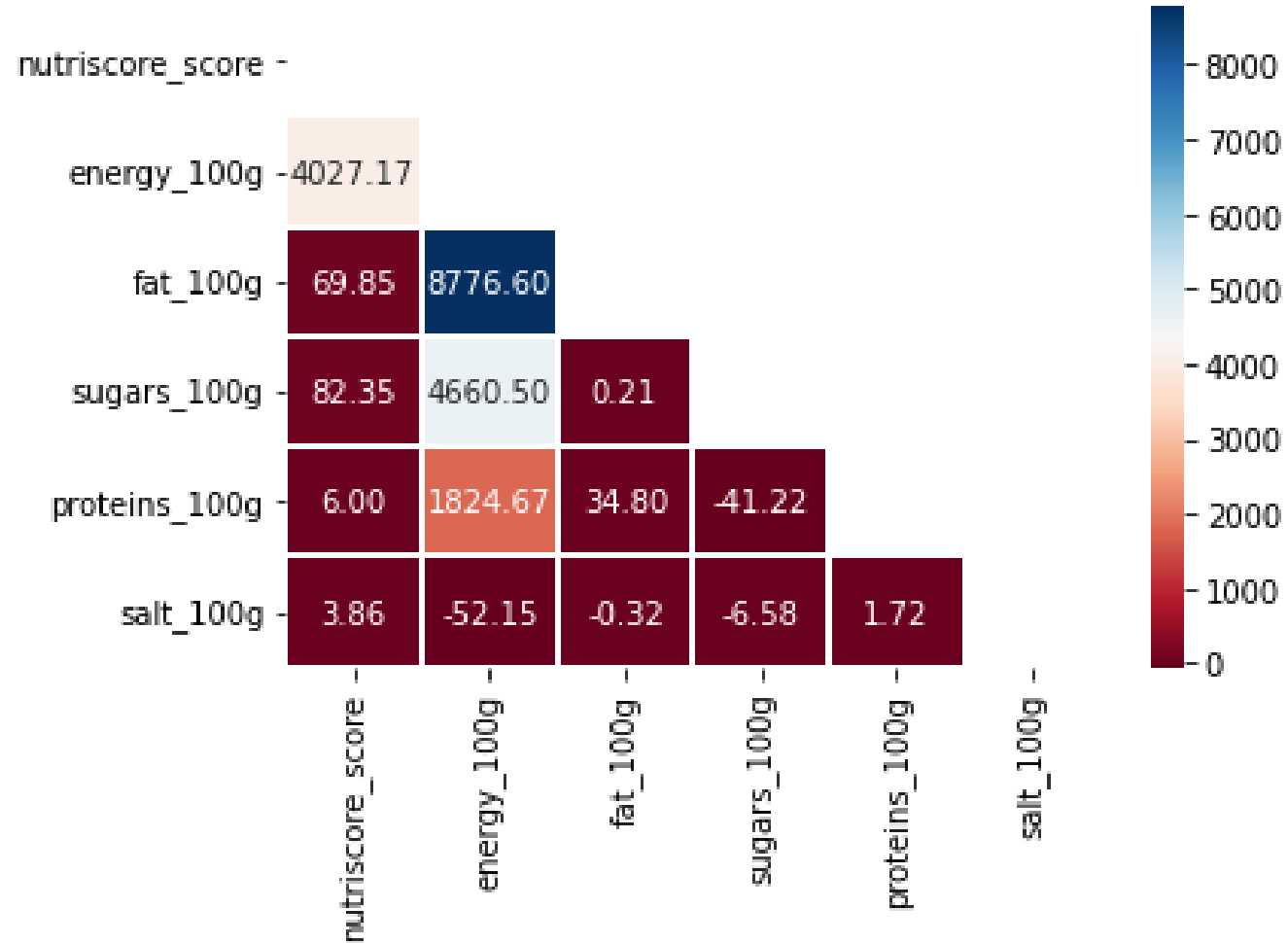


Carte des corrélations linéaires des variables quantitatives (coefficients de Pearson)



- fat_100g & energy_100g
- nutriscore_score & energy_100g
- fat_100g & nutriscore_score

Carte des covariances des variables quantitatives



4. Analyse multivariée

- 1 variable quantitative et 1 variable qualitative
 - Analyse de la variance (ANOVA)
 - $\eta^2 > 0,5 \Rightarrow$ très forte corrélation
 - $0,5 > \eta^2 > 0,14 \Rightarrow$ forte corrélation
 - $0,14 > \eta^2 > 0,06 \Rightarrow$ corrélation moyenne
 - $0,06 > \eta^2 > 0,01 \Rightarrow$ corrélation faible

qualitative	quantitative	Eta 2	corrélation
'pnns_groups_1	energy_100g	0,3604048360352514	forte
'pnns_groups_1	fat_100g	0,19042740448321085	forte
'pnns_groups_1	nutriscore_score	0,3731868235434252	forte
'pnns_groups_1	proteins_100g	0,2503387185664257	forte
'pnns_groups_1	salt_100g	0,028496824825580174	faible
'pnns_groups_1	sugars_100g	0,4179171340991356	forte
'pnns_groups2	energy_100g	0,507895216773527	très forte
'pnns_groups2	fat_100g	0,3764968273193561	forte
'pnns_groups2	nutriscore_score	0,46725870355162297	forte
'pnns_groups2	proteins_100g	0,3744929873905267	forte
'pnns_groups2	salt_100g	0,041699850934901285	faible
'pnns_groups2	sugars_100g	0,5312767698822433	très forte
nutriscore_grade	energy_100g	0,2955716551557354	forte
nutriscore_grade	fat_100g	0,21215339412082945	forte
nutriscore_grade	nutriscore_score	0,9215696422291599	très forte
nutriscore_grade	proteins_100g	0,015728436541087538	faible
nutriscore_grade	salt_100g	0,021793998728242482	faible
nutriscore_grade	sugars_100g	0,21338886358752499	forte

4. Analyse multivariée

- 2 variables qualitatives
 - Test du Chi2

pnns_groups_1 + pnns_groups_2	2807620
pnns_groups_1 + nutriscore_grade	151384,491368516
pnns_groups_1 + main_category_en	2528699,03235631
pnns_groups_2 + nutriscore_grade	215676,392360408
pnns_groups_2 + main_category_en	9383238,60824625

5. Pertinence et faisabilité du projet

- Test du knn sur les ingrédients des produits vendus dans le même pays avec un meilleur nutriscore
- Test du kmeans sur les ingrédients des produits vendus dans le même pays avec un meilleur nutriscore
- Distance euclidienne sur les ingrédients

5. Pertinence et faisabilité du projet

	product_name	countries_en	ingredients_text	nutriscore_score	pnns_groups_1
179486	nutella	France	sucres, huile de palme, NOISETTES 13%, LAIT écrémé en poudre 8,7%, cacao maigre 7,4%, émulsifiants: lécithines [soja]	26.0	Sugary snacks

	product_name	countries_en	ingredients_text	nutriscore_s core	pnns_grou ps_1
31	Paupiette de volaille sauce forestière brocolis purée	France,United Kingdom	Sucre • Beurre (_Lait_) • Œufs_ de poules élevées en plein air pasteurisés - Farine de riz • Cacao maigre en poudre • _Lait_ entier • Humectant : Glycérol • Masse de cacao • Sirop de sucre inversi • Fécule de pomme de terre • Butteroil (_Lait_) • Amidon de tapioca • Sirop de glucose déshydraté • Stabilisant : Gomme xanthane • Beurre de cacao • Sel • Poudre à lever : E341 , Bicarbonate de soude • _Lait_ en poudre écrémé • Huile de palme • Emulsifiant : E481 , E471 , E475, Lécithine de _soja_...	-1.0	Sugary snacks
146	Kashi® Chewy Granola Bars Trail Mix	France,United States	Whole grain oats, brown rice syrup, almonds, dried brown rice syrup, brown rice flour, oat fiber sunflower oil, sunflower seeds, raisins, glycerin, cranberries, cane syrup, sea salt, natural fiavor, TOsemary extract for freshness, peanut flour.	6.0	Sugary snacks
39	Tarte aux framboises	France	Framboises 45%, farine de FROMENT (agent de traitement de la farine : E300), sucre, sirop de glucose-fructose, eau, BEURRE, huile de palme, graisse de palme fractionnée, huile de colza, OEUF entier, poudre de petit LAIT enrichie en protéines, gélifiant : E440, acidifiant: E330, correcteurs d'acidité E331, E332, E327, stabilisant : E333, LACTOSERUM en poudre, poudres à lever (E450, E500), sel, epaissisant E412	7.0	Sugary snacks
80	Annie's Organic Honey Graham Crackers	France,United States	Organic whole wheat flour (graham flour), organic wheat flour, organic cane sugar, organic sunflower oil, organic honey, leavening (baking soda, ammonium bicarbonate, cream of tartar), organic molasses, sea salt, organic brown sugar, organic dried molasses, natural flavor, mixed tocopherols (vitamin e) to protect flavor, organic rosemary extract.	9.0	Sugary snacks
55	Peanut, caramel & cocoa treat bars	France,United Kingdom	Arachides (28%) • Abricots déshydratés • Riz soufflé sucré (Riz • Sucre) • Sirop de glucose • Fibres de chicorée • Miel • Arôme caramel (4%) (Sucre muscovado • Babeurre déshydraté (Lait) • Arôme • Crème déshydratée (Lait) • Farine de riz • Farine de graines de caroube) • Cacao en poudre (4%) Noix de coco déshydratée Huile de colza • Sel de mer. Pour les allergènes, voir les ingrédients indiqués en gras, Ne convient pas aux personnes allergiques aux fruits à coque, au blé, au sésame et au soj...	10.0	Sugary snacks
82	Annie's Organic Friends Bunny Grahams	France,United States	organic wheat flour, organic cane sugar, organic expeller-pressed sunflower oil, organic whole wheat flour (graham flour), organic invert cane syrup, organic corn flour, organic chocolate cookie bits (organic wheat flour, organic cane sugar, organic expeller-pressed sunflower oil, organic malted black barley flour, organic cocoa, baking soda), calcium carbonate, organic honey, organic cocoa powder, organic semi-sweet chocolate chips (organic cane sugar, organic unsweetened chocolate, organic...	11.0	Sugary snacks
83	Annie's Homegrown Organic Friends Bunny Grahams	France,United States	Organic wheat flour, organic cane sugar, organic expeller-pressed sunflower oil, organic whole wheat flour (graham flour), organic invert cane syrup, organic corn flour, organic chocolate cookie bits (organic wheat flour, organic sugar, organic expeller-pressed sunflower oil, organic malted black barley flour, organic cocoa, baking soda), calcium carbonate, organic honey, organic cocoa powder, organic semi-sweet chocolate chips (organic cane sugar, organic unsweetened chocolate, organic coco...	11.0	Sugary snacks
35	Tarte citron meringuée	France	Eau, farine de BLE, sucre, OEUFs entiers, meringue en poudre 6,5% (sucre, poudre de blanc d'OEUF, amidon modifié, acidifiant : acide citrique, arôme naturel), préparation à base de matière grasse végétale (huiles et graisses végétales de palme et colza, eau, émulsifiant : E471, sucre, acidifiant : E330, colorant : E160a, arômes), préparation pour CREME PATISSIERE (sucre, amidon modifié de pommes de terre, dextrose de BLE, LACTOSERUM, matière grasse végétale (matière grasse non hydrogénée de ...	12.0	Sugary snacks
407	Gaufres Vanille Chocolat	France	Gaufres (masseqet cacao) édulàofariti€; maltitol; beurre de cacaOòém0\$1fiifiantléçithinetde soja;ti' : polycicinoîéatde 'de polyglycérol'i Vàngt(e ŒufSde poules élevéeSaus@t, graisseS (canola concentré	13.0	Sugary snacks
81	Annie's Organic Strawberry Toaster Pastries with Frosting 6 Count	France,United States	Whole Wheat Flour*, Invert Sugar*, Water, Sugar*, Palm Oil*, Powdered Sugar*, Honey*, Wheat Flour*, Dried Apples*, Non Fat Dry Milk*, Cornstarch*, Lemon Juice Concentrate*, Salt, Calcium Carbonate, Strawberry Puree Concentrate*, Natural Flavor*, Rice Flour*, Baking Soda, Cream of Tartar, Carrot, Apple and Blackcurrant Juice (Color)*, Tapioca Starch*, Whey Protein Concentrate*, Sodium Alginate (Thickener), Calcium Citrate, Sunflower Oil*, Sodium Citrate, Red Radish Juice (Color)*. *ORGANIC	13.0	Sugary snacks

	product_name	countries_en	ingredients_text	nutriscore_score	pnns_groups_1
0	Pâte à Sucre	France	Sucre (74 %), sirop de glucose, huiles et graisses végétales (palme, tournesol, illipé), sirop de sucre inverti, stabilisant : E420ii, eau, humectant : E422, épaississant : E415, émulsifiant : E471, acidifiant : E330, arôme, conservateur.	18.0	Sugary snacks
1	Pastille Vichy	France	Édulcorants (Sorbitol, Aspartame), agent d'enrobage (Sels de magnésium d'acides), Extraits des eaux de Vichy (0,3 %), Arôme. consommation excessive peut avoir des effets laxatifs. Contient une source de phénylalanine.	3.0	Sugary snacks
2	M&M's Peanut Butter	France	Sucre, _Arachide_, pâte de cacao, lait écrémé en poudre, _lactose_ et protéine de _lait_, matières grasses végétales, beurre de cacao, _beurre_ concentré, amidon, sirop de glucose, lécithine de _soja_, gomme arabique, colorant, arôme, sel, huile végétale.	36.0	Sugary snacks
3	Gelée de fleurs de sureau	France	Farine de _blé_, huile de colza, sucre, _oeufs_ frais 18%, sirop de glucose-fructose, stabilisant : glycérol, poudres à lever : carbonates d'ammonium, carbonates de sodium, diphosphates (_blé_), sel, émulsifiants : mono et diglycérides d'acides gras, _lait_ écrémé en poudre, arôme.	14.0	Sugary snacks
4	Madeleines ChocoLait	France	Farine de _blé_, sucre, huile de colza, oeufs frais 14,5%, sirop de glucose-fructose, beurre de cacao, lait entier en poudre, stabilisant : glycérol, pâte de cacao, poudres à lever: carbonates d'ammonium - carbonates de sodium - diphosphates (_blé_), émulsifiants : mono et diglycérides d'acides gras - lécithines (_soja_), sel, lait écrémé en poudre, arômes.	17.0	Sugary snacks

5. Pertinence et faisabilité du projet

les 15 produits les plus proches de nutella

	product_name	nutriscore_score	distance_eucl
0	Pâte à Sucre	18.0	0.0
8175	Eclats d'amandes	19.0	0.0
8176	Eclats de Noisettes	16.0	0.0
5694	Pâte à tartiner aux noisettes	25.0	0.0
5693	Pâtes à tartiner aux noisettes	18.0	0.0
5692	Barre marbrée au cacao	15.0	0.0
2292	Baba au Rhum à la Crème Vanillée	11.0	0.0
5687	Madeleines Longues	15.0	0.0
2299	Gelée Groseilles	11.0	0.0
8174	Sablés aux amandes 100% épeautre	18.0	0.0
2304	Biscoco	21.0	0.0
5679	Sables fourrés noisette	19.0	0.0
9415	Barres cerealieres 4 cereales	10.0	0.0
2309	Galettes au BEURRE	19.0	0.0
9414	barre céréales noisette	9.0	0.0

les 15 produits avec le meilleur nutriscore parmi les 30 plus proches de nutella

	product_name	nutriscore_score	distance_eucl
9414	barre céréales noisette	9.0	0.0
9415	Barres cerealieres 4 cereales	10.0	0.0
2242	Freises Confiture	11.0	0.0
2241	Bonne Maman Konfitüre	11.0	0.0
2236	Framboises - Gelée préparée	11.0	0.0
2234	Confiture orange en tranches	11.0	0.0
2292	Baba au Rhum à la Crème Vanillée	11.0	0.0
2243	Confiture abricot	11.0	0.0
2299	Gelée Groseilles	11.0	0.0
9429	Canistrelli corses	12.0	0.0
5734	Brioche Tressée	12.0	0.0
8180	Sablés pur beurre	14.0	0.0
5687	Madeleines Longues	15.0	0.0
5692	Barre marbrée au cacao	15.0	0.0
8176	Eclats de Noisettes	16.0	0.0

Conclusion

- L'application est réalisable avec même trois versions différentes
 - Knn
 - Kmeans
 - Distance euclidienne
- Axes d'amélioration
 - Prise en compte de tous les ingrédients
 - D'autres critères comme nova_group, données nutritionnelles, traces de, nombre d'ingrédients