

Projet 3 : Concevoir une application au service de la santé publique



Courses Ecolo Healthy

Abdessalem BOUZAYANI

8/11/2022

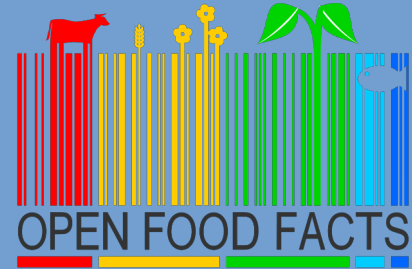
Sommaire

- **Problématique**
- **Application**
- **Jeu de données**
- **Nettoyage**
- **Exploration**
- **Démo d'application**
- **Conclusion**

Problématique



- Appel à projets de l'agence "Santé publique France"
- A partir d'un jeu de données Open Food Facts, on doit répondre aux questions :
 - Réfléchir à une idée d'application
 - Repérer les variables pertinentes
 - Nettoyer les données
 - Analyse univariée et multivariée
 - Justifier l'idée d'application



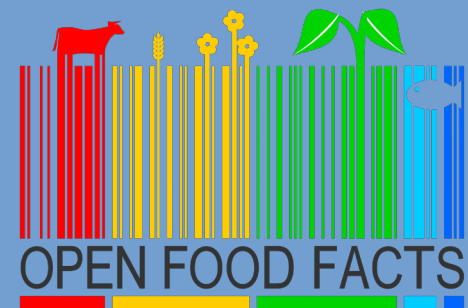
Application : idée

Courses écolo-healthy

- Pour votre santé, évitez de manger trop gras, trop sucré, trop salé
- Pour votre santé, privilégier les fibres et les vitamines
- Pour protéger votre environnement, éviter les déchets polluants
- Pour votre pays, manger des produits français

Application : cas d'utilisation





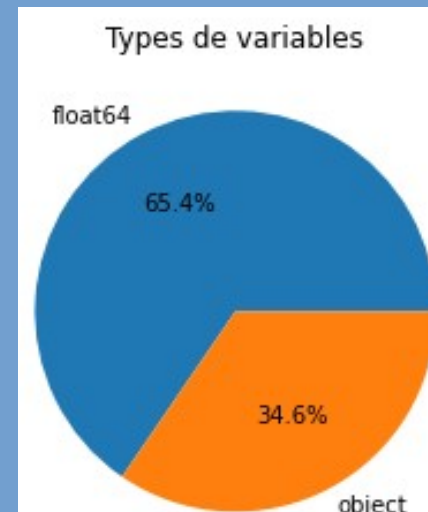
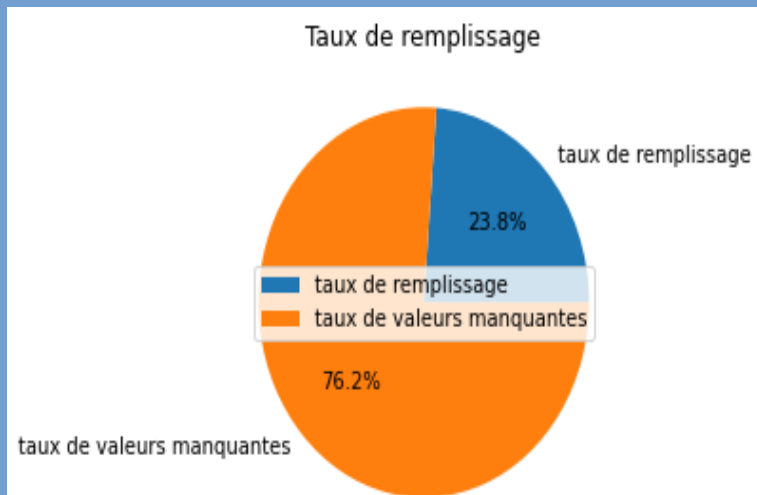
Jeu de données

162 colonnes

320772 lignes

(produits)

(informations générales, ensemble de tags, ingrédients, informations nutritionnelles)



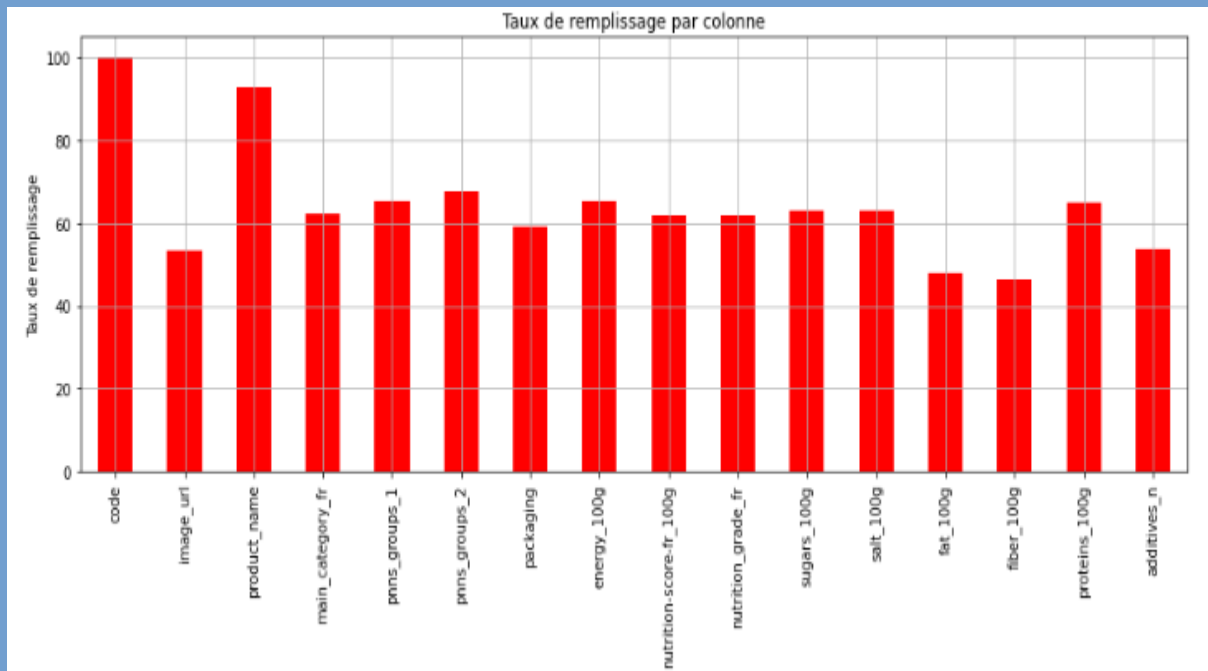
04/11/22

Nettoyage



Variables pertinentes

- code
- image_url
- product_name
- main_category_fr
- pnns_groups_1
- pnns_groups_2
- packaging
- energy_100g
- nutrition-score-fr_100g
- nutrition_grade_fr
- sugars_100g
- salt_100g
- fat_100g
- fiber_100g
- proteins_100g
- additives_n



Valeurs manquantes

- **Suppression**

taux valeurs manquantes > 80 %

- **Constante**

colonnes textuelles

- **Autre variable**

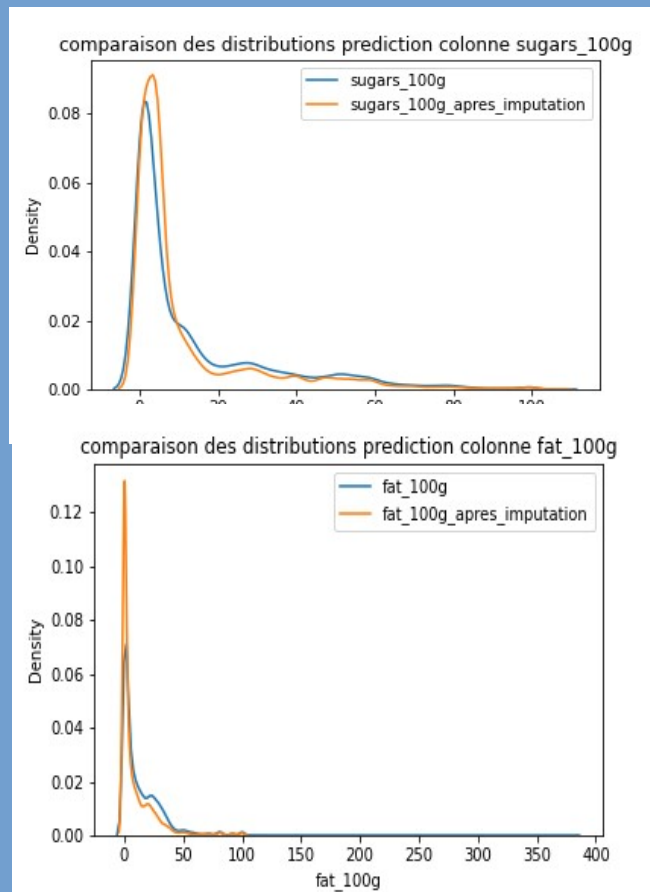
product_name, nutrition_grade_fr et nutrition-score-fr_100g

- **Médiane**

sugars_100g, salt_100g, fiber_100g et proteins_100g

- **KNN**

energy_100g, nutrition-score-fr_100g et fat_100g



Valeurs aberrantes

- **Poids positif et inférieur à 100g**

sugars_100g, salt_100g, fat_100g, fiber_100g et proteins_100g

- **Maximum possible**

energy_100g, proteins_100g et fiber_100g

- **Nombre de caractères**

code et product_name

- **Aucun traitement**

main_catgory_fr, pnns_groups_1 et pnns_groups_2

Score Ecolo-healthy

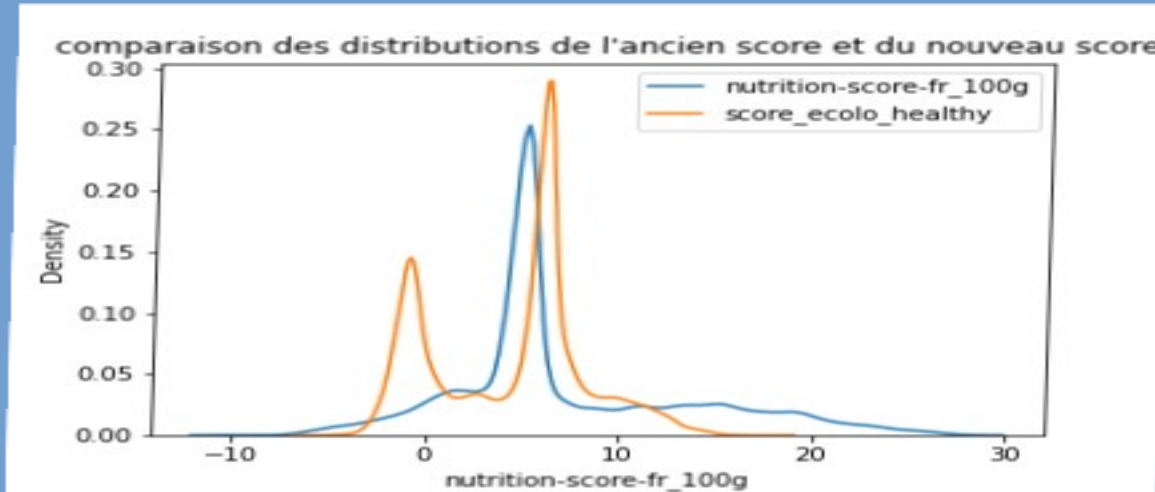
- Normalisation

Min Max (-15,40)

- Score ecole_healthy

$\text{moyenne}(2 \times \text{score}, 2 \times \text{emballage}, \text{sel}, \text{sucres}, \text{energie}, \text{graisses}, \text{fibres}, \text{proteines}, \text{additifs})$

- Histogramme

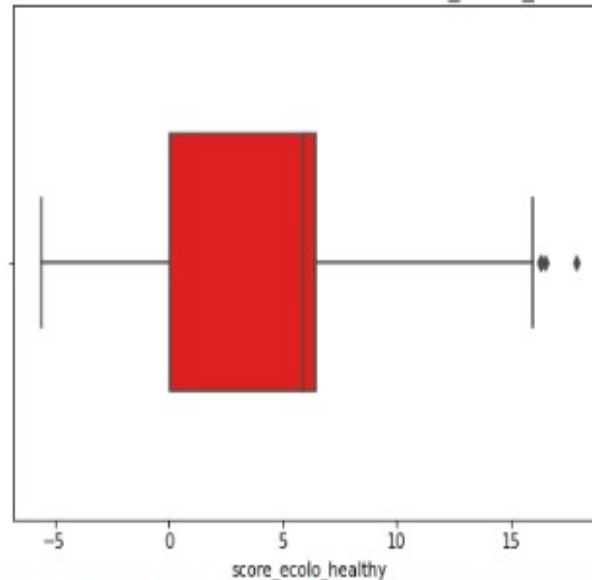


Analyse univariée

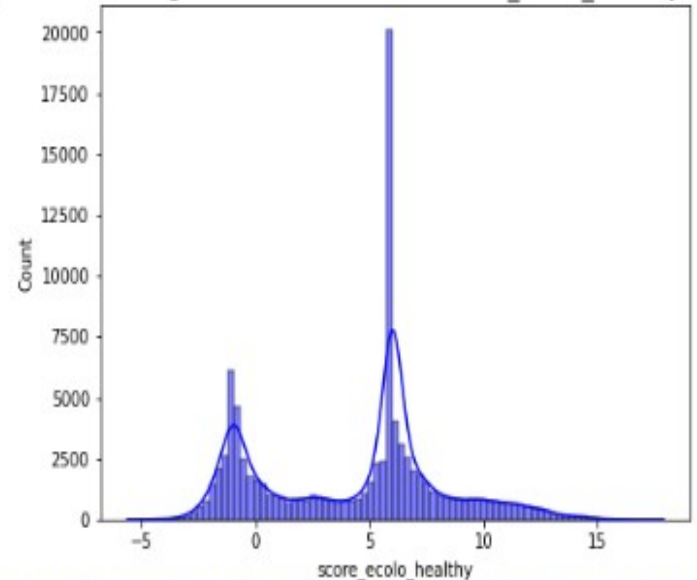
- Score_ecolo_healthy

```
moyenne : 4.36  
mediane : 5.92  
mode : 0 6.0  
dtype: float64  
variance : 15.86  
skewness : -0.02  
kurtosis : -0.79  
ecart type : 3.98  
min : -5.61  
25% : 0.09  
50% : 5.92  
75% : 6.45  
max : 17.9
```

Boite a moustache de la colonne score_ecolo_healthy



histogramme de la colonne score_ecolo_healthy



Analyse bivariable

- qualitatives / qualitatives

Test d'indépendance Chi-square

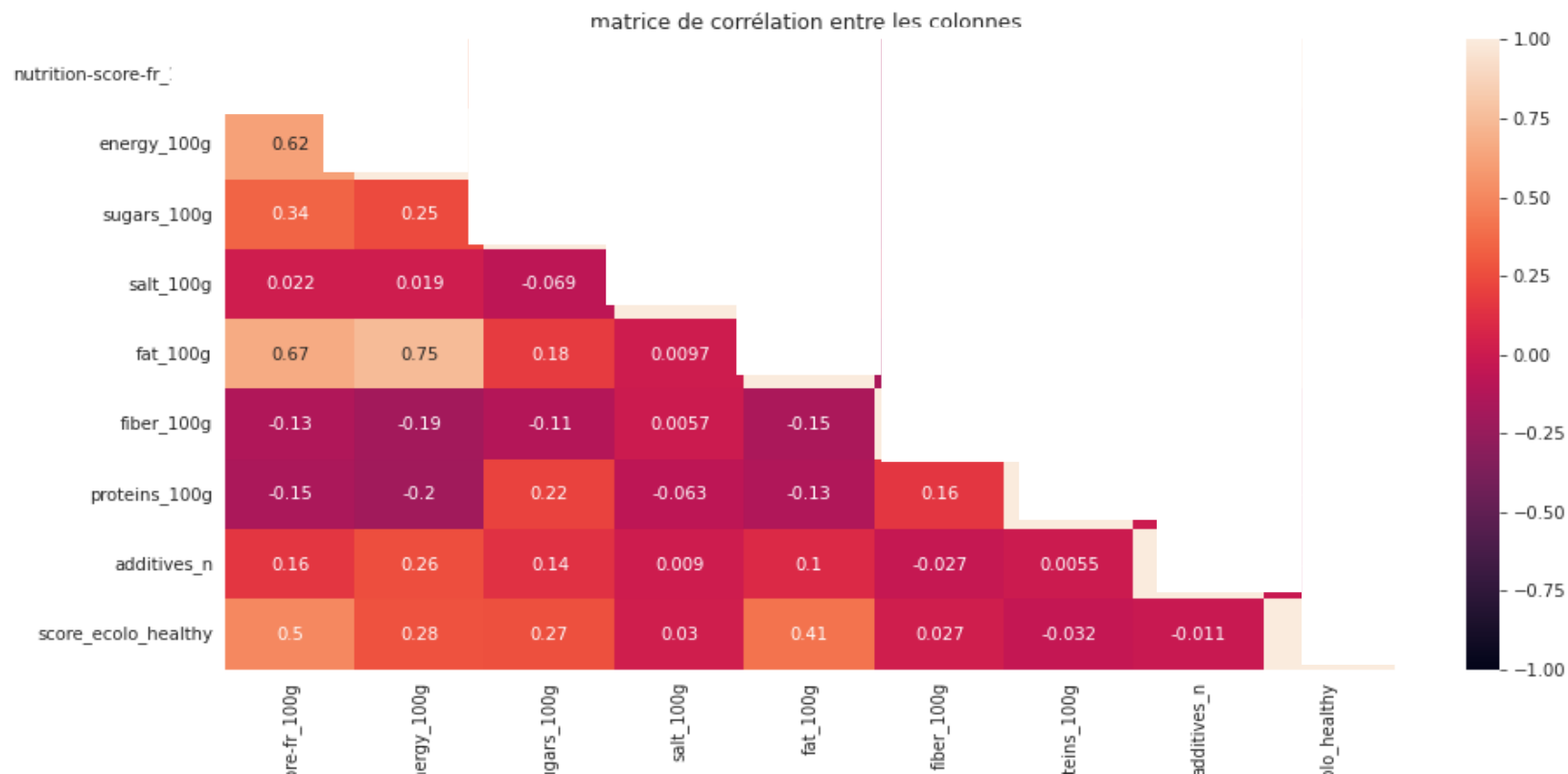
- quantitatives / qualitatives

Test Anova 1 facteur

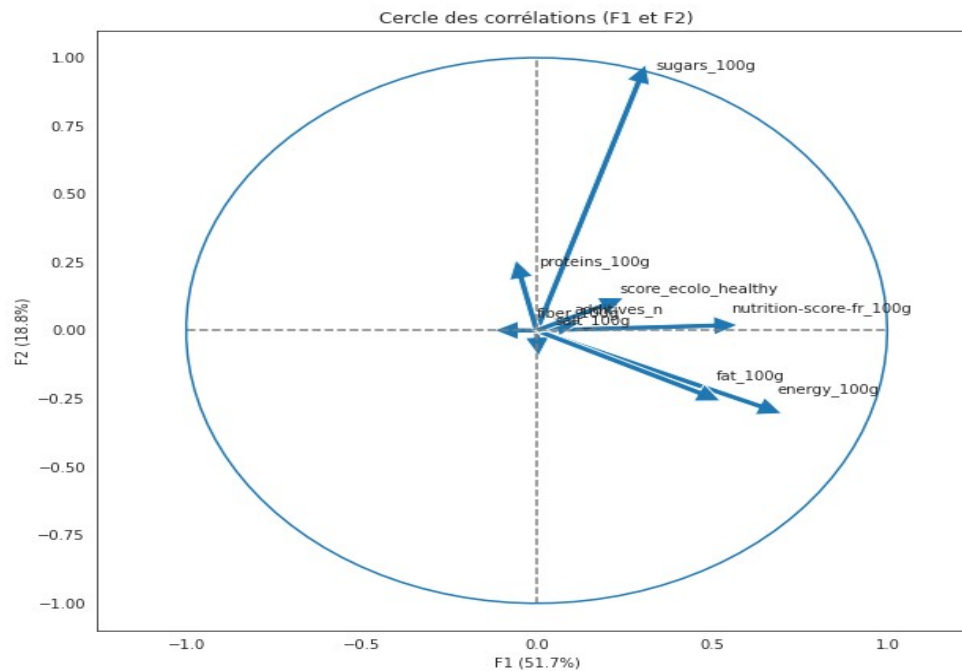
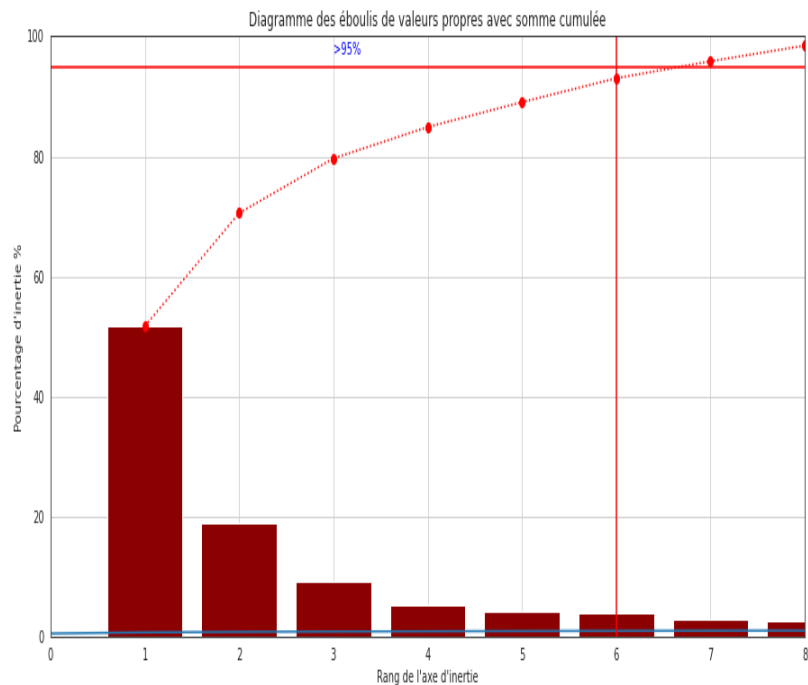
- quantitatives / quantitatives

Matrice de corrélation

Analyse bivariée



Analyse multivariée



Exemple de recommandation

Soupe

'Potage bio Potimarron'



'Pot au Feu de legumes'



'Déli Soup' Brocoli Ciboulette'



'Velouté de légumes bio'



'Harira Marocaine'



Riz

'Riz long incollable'



'Riz Complet naturellement bon'



'Riz pour risotto Arborio Monoprix'



'Riz long grain semi-complet'



'Riz basmati complet'



Chocolat

'Dessert Noir (lot de 2)'



'Dark 72%'



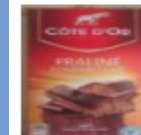
'Extra Fine White'



'Côte d'Or Mignonnette Noir'



'Praliné fondant lait'



Conclusion

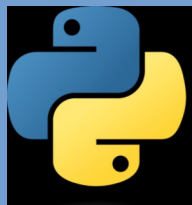
- **Jeu de données large mais avec plusieurs erreurs (données saisies par le public)**
- **Plusieurs opérations de nettoyage (suppression, imputation, valeurs aberrantes)**
- **Plusieurs analyses (univariée, bivariée, multivariée)**
- **Moteur de recommandation des produits saines et écologiques**

Perspectives

- **Produits hors France**
- **Prix**
- **Traitement des variables textuelles (TAL)**

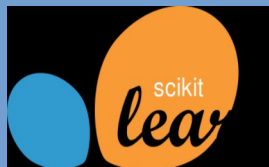
Boîte à outils

Environnement



IP[y]:
IPython

Bibliothèques



Visualisation



Merci de votre attention



Contact : bouzaieni@gmail.com