
Concevez une application au service de la santé publique

Thibault Grandjean¹

¹ Étudiant auprès d'Openclassrooms

24 janvier 2020

L'obésité est aujourd'hui une maladie au centre de l'attention. En effet, ce sont 17% des adultes en France (13 % dans le monde) qui sont concernés. Le nombre de cas d'obésité a presque triplé depuis 1975.[1] [2] Cette maladie résulte d'un déséquilibre entre les apports et les dépenses énergétiques. Pour enrayer la tendance, l'État français a mis en place un plan de prévention national au travers du ministère de la santé. Le présent document étudie la faisabilité d'une application permettant de mieux manger dans le cadre du plan de prévention national.

1 Introduction

Le surpoids et l'obésité se définissent comme une accumulation anormale ou excessive de graisse corporelle qui représente un risque pour la santé.

Le surpoids et l'obésité sont des facteurs de risque majeurs pour un certain nombre de maladies chroniques, parmi lesquelles le diabète, les maladies cardiovasculaires et le cancer. [2]

1.1 Contexte

L'agence santé publique France a lancé un appel à projets pour trouver des idées innovantes d'applications en lien avec l'alimentation. Nous répondons donc à cet appel à projets avec une idée d'application pour smartphone, permettant sur base des informations nutritionnelles de proposer des produits équivalents à ceux désirés par le consommateur mais avec de meilleures propriétés nutritionnelles.

2 Objectifs

L'objectif principal est d'étudier la faisabilité d'une telle application. Pour ce faire, une analyse minutieuse des données contenues dans la base de données d'openfoodfacts est nécessaire. Il s'agit dans un premier temps de cerner les variables nécessaires au fonctionnement de l'application et ensuite vérifier que les données permettent bien de répondre à la problématique.

3 Problématique

La problématique est double :

La première chose est de s'assurer que les données dont on dispose contiennent bien les informations nécessaires pour créer une telle application.

La deuxième chose :

Peut-on trouver dans la base de données, un produit équivalent avec de meilleures propriétés nutritionnelles à partir d'un produit scanné dans un magasin ?

4 Démarche

Pour répondre à la problématique, on se base sur la base de données d'openfoodfacts (à l'heure actuelle la plus grosse base de produits alimentaire). Une fois les données téléchargées, ces dernières subissent un nettoyage. On réalise alors une analyse statistique descriptive sur une sélection de variables d'intérêt. L'analyse statistique est divisée en deux parties, les analyses univariées et les analyses bivariées.

4.1 Outils

4.1.1 Matériel

L'analyse a été réalisée sur un ordinateur personnel (processeur intel i7 4 coeurs, 16 Go de RAM.) et ne nécessite pas de matériel particulier.

4.1.2 Logiciels

L'analyse a été réalisée avec le langage Python et des notebooks Jupyter sont disponibles dans le répertoire Github (voir section 10)

5 Les données utilisées

5.1 Généralités

Les données utilisées sont disponibles gratuitement auprès d'Openfoodfacts et sont publiées sous licence "Open Database License".

Les données sont entrées par les utilisateurs (Applications mobiles : Openfoodfacts et Yuka). Par conséquent les données sont régulièrement mal complétées ou erronées. La base de données est assez conséquente (2.2 Go), elle contient (à l'heure actuelle ¹) 1 054 801 ² d'entrées et 177 colonnes.

5.2 Contenu de la base de données

Les champs sont séparés en quatre sections :

- Les informations générales sur la fiche du produit : nom, marque, date de création...
- Un ensemble de tags : catégorie du produit, localisation, origine, etc.
- La liste des ingrédients et les additifs éventuels.
- Des informations nutritionnelles : quantités au 100g (graisse, sucre, etc.).

5.3 Nettoyage de la base de données

5.3.1 Données concernant la France

On ne récupère que les données pour la France et les pays limitrophes francophones.

5.4 Taux de remplissage des champs

On peut regarder le taux de remplissage des champs de manière graphique à l'aide d'un graphique type matrice. (chaque valeur est alors représentée par un tiret (les colonnes noires sont alors totalement complètes et les colonnes vides sont blanches.) Voir graphique 1

On peut regarder la distribution du nombre de champs complétés.

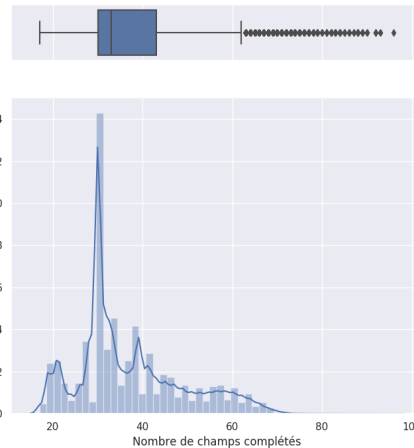


FIGURE 2 – Distribution du nombre de champs complétés.

On a au minimum 18 champs complétés et 97 champs complétés maximum.

6 Analyses univariées

6.1 Variables sélectionnées

Pour assurer la faisabilité du projet, nous avons besoins de :

- Le nom du produit
- Le code bar du produit
- La marque du produit
- La catégorie à la quelle appartient le produit
- Le nutriscore (nutrition grade [A, ..., E])
- Certaines données nutritionnelles :

- Valeur énergétique au 100g
- Teneur en protéines
- Teneur en sucres
- Teneur en graisses
- Teneur en sel

1. Novembre 2019

2. 600 654 pour les pays francophones (France, Suisse, Belgique et Luxembourg)

6.2 Répartition du nutriscore

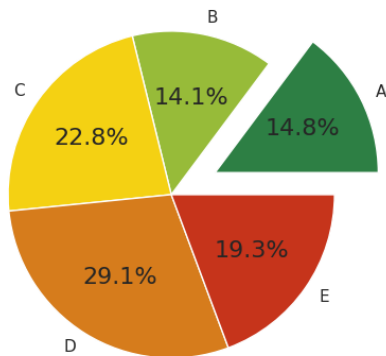


FIGURE 3 – Répartition du nutriscore des produits

6.3 Valeurs énergétiques des produits

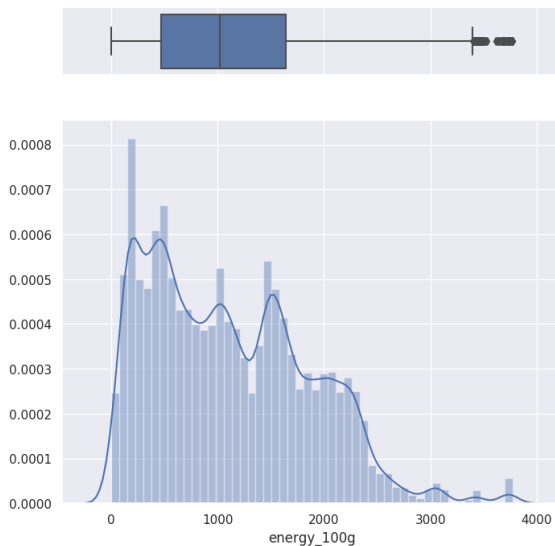


FIGURE 4 – Distribution des valeurs énergétiques des produits dans la base de données.

7 Analyses multivariées

Toujours dans l'objectif de trouver un produit équivalent, il est important de vérifier que pour chaque catégorie (ici groupe Programme nutrition santé PNS) il existe des produits pour chaque nutriscore (de A à E).

8 Conclusion

9 Perspectives

10 Liens internet

- <https://github.com/tgrandjean/OC-sante-publique-france>
- <https://world.openfoodfacts.org/data>

Bibliographie

- [1] *Inserm Dossier information : Obésité.* <https://www.inserm.fr/information-en-sante/dossiers-information/obesite>. Accessed : 2020-01-23.
- [2] *Organisation mondiale de la santé Thème de santé, Obésité.* <https://www.who.int/topics/obesity/fr/>. Accessed : 2020-01-23.

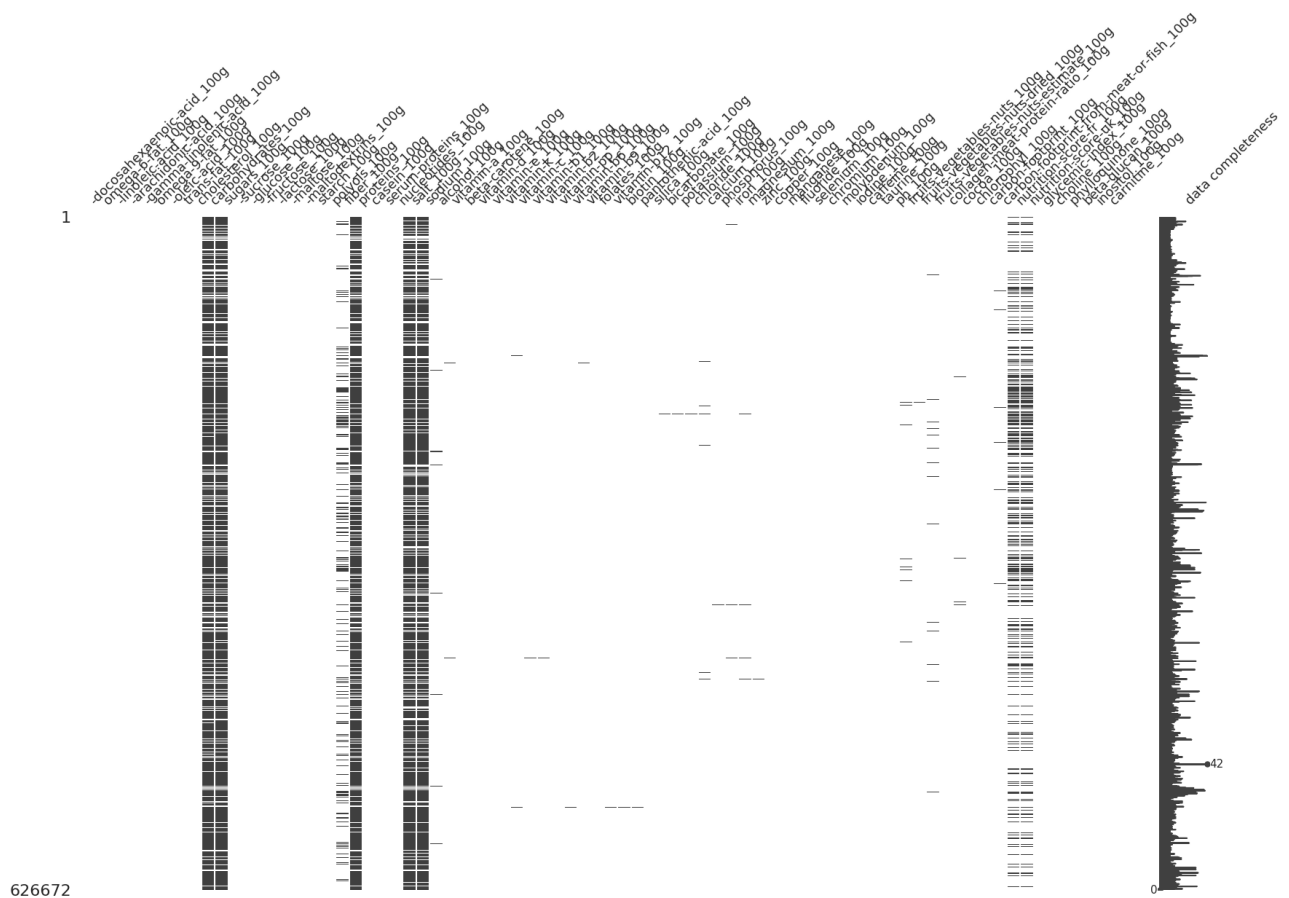
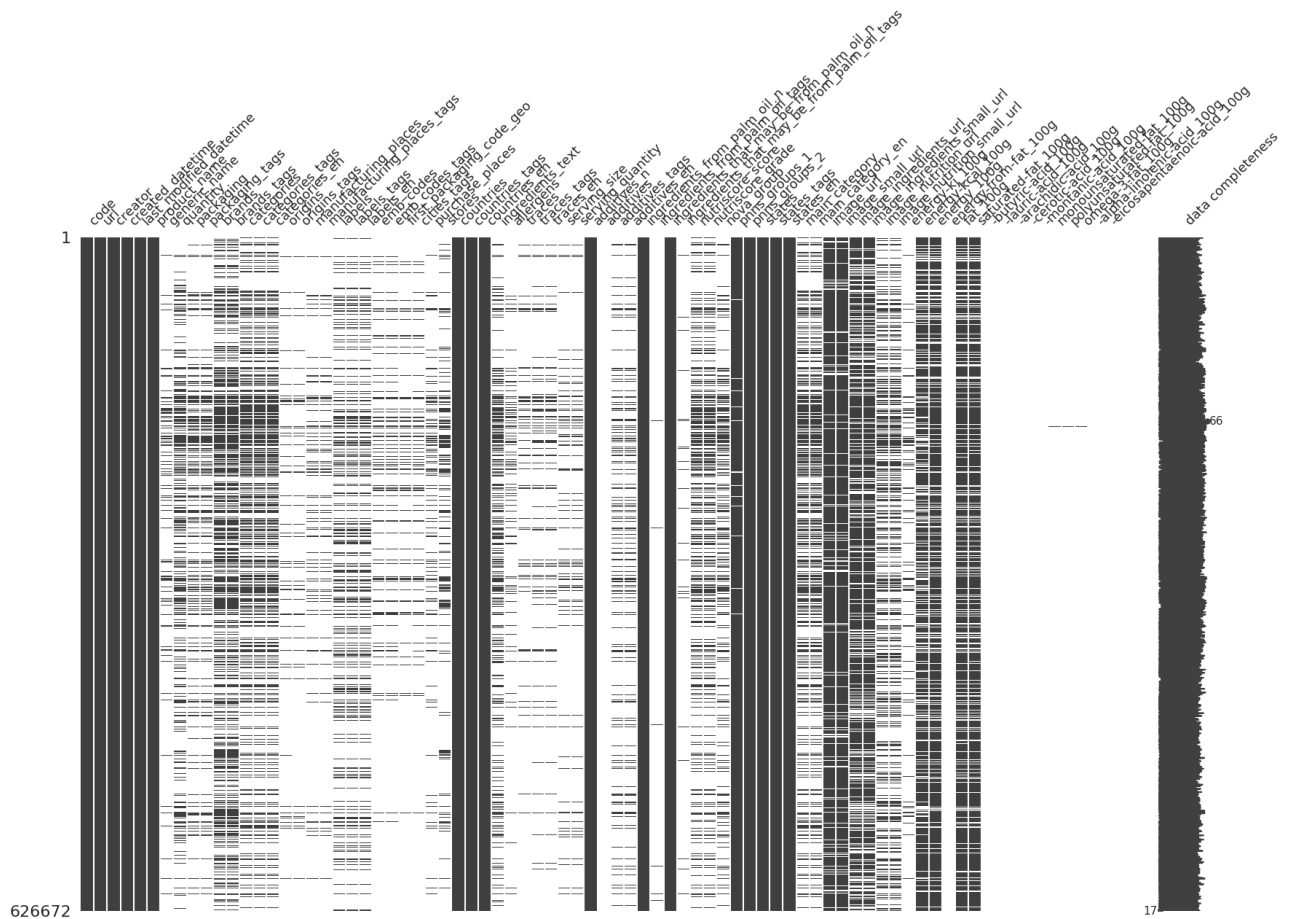


FIGURE 1 – Taux de remplissage des variables. En haut les colonnes de 1 à 78 et en bas le reste des colonnes. Les colonnes vides ont été retirées