

**PROJET 5- UTILISER LES DONNÉES PUBLIQUES DE  
L'OPENFOODFACTS**

Le programme développé a pour but de proposer aux utilisateurs un moyen simple de changer leurs habitudes alimentaires. Il permet de sélectionner des aliments issus de la base de données publiques OpenFoodFacts et de les comparer en fonction de leurs score nutritionnel. Le programme a été développé en deux parties: la première partie comprend l'analyse des données et l'insertion dans la base de données, et la deuxième partie concerne l'interaction des données avec l'utilisateur.

Les données sont stockées dans une base de données relationnelle MySQL. La base de données est constituée de 4 tables(catégories, sous-catégories, produits, et produits de remplacement), liées entre elles par un couple de clés primaires et étrangères.

L'analyse des données a été faite en utilisant la librairie Pandas. Pandas fait partie des librairies de base pour la data science. L'utilisation d'un "DataFrame" pour lire et trier les données à été utilisé. Le "DataFrame" est généré à partir d'un fichier csv qui est téléchargé directement depuis le site d'OpenFoodFacts. Divers masques sont utilisés afin de trier les données qui ne font pas partie de la France, et suppriment des données si elles sont nulles. La classe DataToMySQL permettant d'insérer les données du "DataFrame" dans la base de données est une classe qui hérite de la classe Singleton. Cette classe permet de sécuriser l'utilisation de la classe DataToMySQL en ne permettant l'instanciation de cette dernière qu'une seule fois au sein du programme. L'insertion des données se fait par l'utilisation d'une librairie nommée "Records". Cette librairie permet une utilisation simplifiée de requêtes SQL par la création d'objet faisant le lien avec la base de données.

La gestion de l'interface entre l'utilisateur et la base de données a été créée en 3 classes. Chaque classe a une fonction définie. La classe Data permet de sélectionner ou de mettre à jour les données dans la base. Elle permet aussi de mettre à jour la base de données en téléchargeant à nouveau le fichier csv auprès du site internet d'OpenFoodFacts et de réinsérer des données dans la base. Cependant cette méthode supprime toutes les données de la base, l'utilisateur n'a pas moyen de garder les produits qu'il a substitué auparavant. La classe UserChoice permet de stocker temporairement les différents choix de l'utilisateur au travers des attributs de classe. Ces attributs sont accessibles et modifiables si besoin au travers d'accesseurs et de mutateurs.

La classe Interface est la classe principale de notre projet permettant l'interaction avec l'utilisateur. Cette classe affiche dans le terminal les données de la base, et reçoit les choix de l'utilisateur. Il a été donné à l'utilisateur la possibilité de sélectionner ses choix par les touches numériques du clavier. L'utilisateur a aussi la possibilité d'avoir accès à d'autres fonctions telles que revenir à la page d'accueil, de changer de pages, d'afficher de l'aide et d'afficher plus d'informations sur un produit choisi. Ces choix sont autorisés par un "mapping" de certains caractères unique du clavier, ou par combinaison de touche pour le choix d'afficher des informations supplémentaires d'un produit.

La première difficulté sur ce projet aura été la manipulation des données fournies par le site OpenFoodFacts. Le fichier téléchargé au format csv est un fichier très lourd qui est difficilement exploitable en l'état. Il m'aura fallu apprendre le fonctionnement de la librairie Pandas pour pouvoir l'exploiter au mieux. Cependant l'utilisation de "Pandas" ne permet pas de régler tous les types de problèmes associés au fichier, car au final on trouve beaucoup d'incohérence dans ce fichier. Par exemple, j'ai choisi de n'utiliser que les produits situés en France, et pourtant des produits anglophones sont insérés dans la base de données car ces produits sont mal renseignés. J'ai fait le choix de les laisser car je voulais traiter les données de manière dynamique et non pas supprimer une par une les valeurs qui ne convenaient pas.

La seconde difficulté a été de gérer l'interaction avec l'utilisateur. La gestion de plusieurs types d'interactions clavier, qu'ils soient avec un chiffre entier, un caractère ou une combinaison d'un caractère et d'un chiffre n'ont pas été aisés, et notamment dans la vérification du type d'interaction. Il m'a fallu pour cela comprendre quand et comment récupérer ses informations mais aussi sous quelle forme. En implémentant un type d'interaction à la fois j'ai pu surmonter cette difficulté et créer tous les types d'interactions que je souhaitais avec le message d'erreur associé en cas de non respect des interactions.

Ce projet m'aura permis d'utiliser une base de données relationnelles et d'interagir avec cette base grâce au langage SQL. Il m'aura aussi permis de m'améliorer en Python avec les interactions utilisateurs, les formatages de chaîne et la gestion des erreurs au sein d'un programme.