

S2.04 : Exploitation d'une base de données

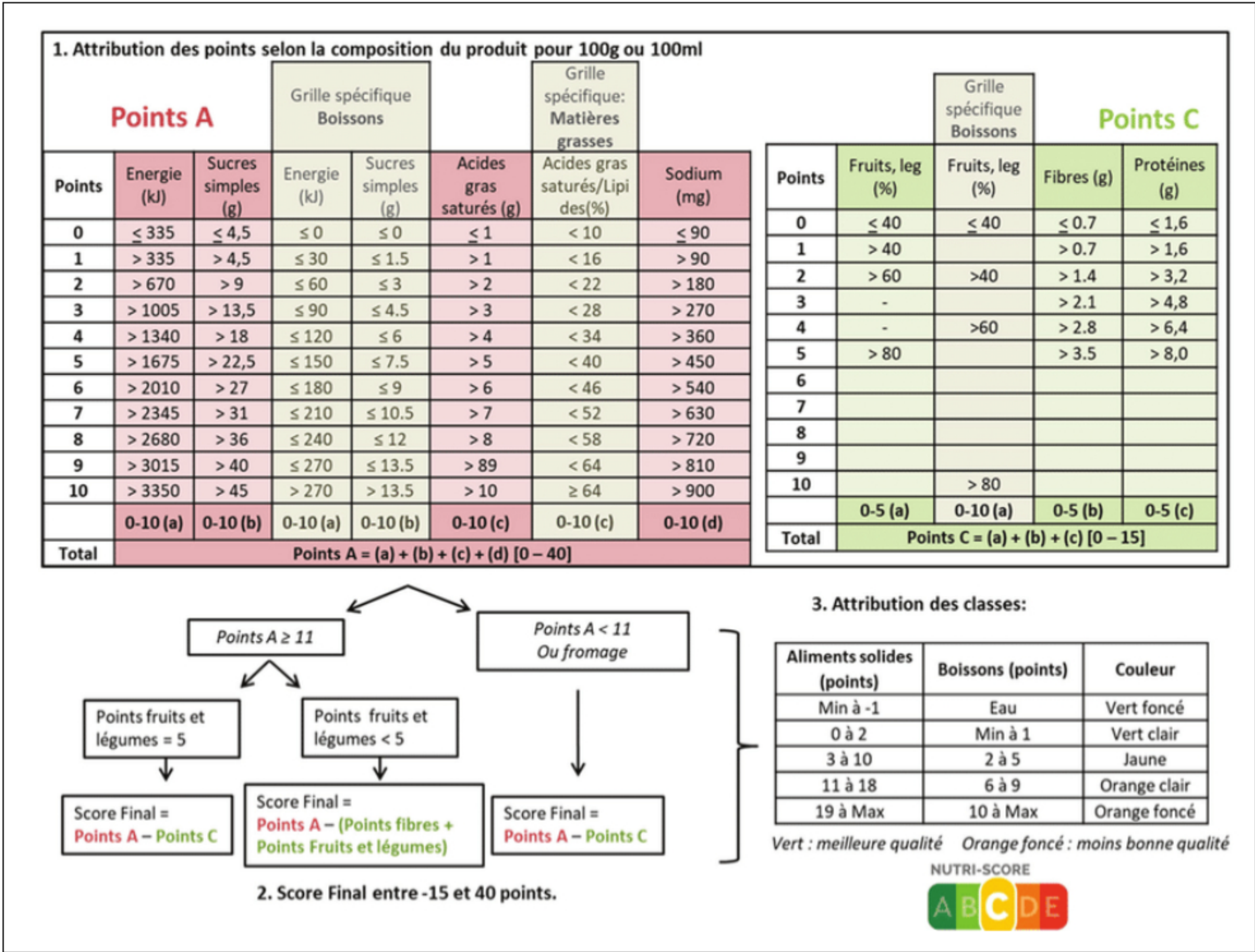
Cette SAÉ est un mini-projet constituant une première sensibilisation à ce que peut être une mission en *Data Science*.

1. Mise en situation

Vous êtes deux membres d’une petite association sensible à l'accès libre aux données et souhaitant informer le consommateur sur le **nutriscore**.

La vidéo suivante donne quelques éléments sur ce qu’est le nutriscore et ce qu’il permet ou non de mesurer : [\[VIDEO\]](#).

Il est calculé selon l’algorithme suivant :



Source : [H. Aguenau, C. Julia, A. El Hajjab, P. Galan, H. Berri, M. Brahim, J. Belkhadir, J. Heikel, S. Hercberg \(2018\). Le logo nutritionnel Nutri-Score: un outil au service du consommateur marocain. 271-276.](#)

2. Objectif et organisation du projet

A partir d'une base de données disponible en *opendata* et renseignée par des contributions libres (donc mal remplie, et contenant plus d'informations que nécessaire etc.), réaliser une étude simple contenant des graphiques appropriés pour répondre à une question précise, et rédigée en anglais.

La base de données utilisée est celle proposée par <https://fr.openfoodfacts.org/> où vous pouvez trouver la courte présentation suivante :

Open Food Facts est une base de données sur les produits alimentaires faite par tout le monde, pour tout le monde. Elle vous permet de faire des choix plus informés, et comme les données sont ouvertes (open data), tout le monde peut les utiliser pour tout usage.

Dans le cadre de cette SAÉ, nous vous proposons d'organiser le travail en trois phases :

1. Exploration et nettoyage des données : encadrée par l'équipe pédagogique de bases de données
2. Production des graphiques : encadrée par l'équipe pédagogique de statistiques
3. Rédaction en anglais : encadrée par l'équipe pédagogique d'anglais

Ce découpage peut être un peu artificiel : en réalité les trois phases se mélangent en des aller-retours successifs, et les outils pourraient être partagés (on pourrait utiliser PostgreSQL pour calculer des grandeurs statistiques ou R pour explorer et nettoyer les données). Il a l'avantage de clairement identifier les trois phases et de faciliter le partage de l'encadrement entre les équipes pédagogiques.

2.1. Calendrier

Ces phases s'organisent selon le calendrier suivant, sur trois semaines :

- Phase préliminaire :
 - semaine 13 : pas de séance; constitution des binômes, découverte du sujet, début du travail de cadrage du projet.
- Phase 1 (BD) :
 - semaine 14 : 3 séances de 2h;
- Phase 2 (Stats) :
 - semaine 14 : 1 séance de 3h;
- Phase 3 (Anglais) :
 - semaine 15 : 1 séance de 2h.

Chaque équipe pédagogique vous informera des attendus pour chaque étape et du mode d'évaluation choisi.

2.2. Equipes

Dans le cadre de ces trois semaines, le travail est à réaliser en binômes. Votre première tâche est de constituer votre équipe de deux en répondant à l'enquête accessible depuis le *Parcours* de la page chamilo de la SAÉ <https://chamilo.iut2.univ-grenoble-alpes.fr/courses/BUTSAE23INFO204/> :

- **Chaque étudiant doit répondre à l'enquête pour désigner son binôme au sein du même groupe TD.**
- Seules les équipes dont les deux membres se seront mutuellement désignés seront acceptées.
- Dans le cas contraire, M. Gérot constituera des équipes arbitrairement.
- Si un groupe TD contient un nombre impair d'étudiants, M. Gérot ne constituera qu'une seule équipe constituée d'un étudiant.

L'enquête doit être complétée par chaque étudiant avant Mercredi 27 Mars 2024 à 12:00 (midi).