

Projet Fil Rouge

1- Mise en situation

Vous êtes intégré à une nouvelle équipe de chercheurs de la [Food and Agriculture Organization of the United Nations \(FAO\) FAOSTAT](#) l'un des organes qui compose l'ONU et dont l'objectif est d'« aider à construire un monde libéré de la faim ».

Le problème de la faim est en effet complexe et peut avoir de multiples causes, différentes selon les pays. Votre équipe est chargée de **réaliser une étude** de grande ampleur sur le thème de **la sous-nutrition dans le monde**.

2- Télécharger les données

L'interface du site de la FAO permet de sélectionner les types de données que vous souhaitez télécharger. Comme le volume de celles-ci est important, l'interface permet un premier filtrage selon les **pays** (ou *groupements de pays et régions du monde*), les **années**, les **éléments**, les **produits** (et groupements de produits : céréales, produits animaux, etc.).

N.B : La FAO sépare ses données en 2 catégories de produits alimentaires distincts : les produits issus des **animaux** et ceux issus de **végétaux**.

www.fao.org/faostat/en/#data/FBS

Il faudra prendre un peu de temps pour se familiariser avec cette interface. Avant de télécharger les fichiers CSV, il est possible de tester le filtrage grâce à l'option de prévisualisation (bouton "Montrer les données").

On cible pour ce projet une Analyse pour ttes les années depuis 2014 qui nécessite les données suivantes :

1. les bilans alimentaires pour les produits animaux
2. les bilans alimentaires pour les produits végétaux
3. la population de chaque pays
4. les informations sur le nombre de personnes en sous-nutrition par pays
5. Une liste des produits considérés comme "céréales"

* Pour les données citées en 1 et 2, il faut bien faire attention à ne sélectionner QUE "Produits Végétaux > (Liste)" et "Produits animaux > (Liste)" (voir les captures d'écran ci-dessus).

Sinon, il risque d'y avoir de la redondance dans les données (exemple : une ligne contenant "viande bovine" et une ligne contenant "viande". La ligne "viande" étant une agrégation de "viande bovine" et d'autres types de viandes et donc ça peut fausser vos résultats par la suite).



*Pour le point 3 : les données de populations :

The screenshot shows the FAO FAOSTAT Suite of Food Security Indicators interface. The page is titled "Suite of Food Security Indicators" and includes a search bar and navigation tabs: "Data", "Selected Indicators", "Compare Data", "Definitions and Standards", and "FAQ". The main content area is divided into several sections: "COUNTRIES", "ELEMENTS", "ITEMS", and "YEARS".

- COUNTRIES:** Filter results e.g. afghanistan. Selected: Afghanistan, Albania, Algeria, American Samoa, Andorra, Angola.
- ELEMENTS:** Filter results e.g. value. Selected: Value.
- ITEMS:** Filter results e.g. availability > (list). Selected: Number of people undernourished (million).
- YEARS:** Filter results e.g. 2019 / 2018-2020. Selected: 2019 / 2018-2020, 2018 / 2017-2019.

At the bottom, there are options for "Output Type" (Table, Pivot), "File Type" (CSV, XLS), "Thousand Separator in 'Show Data'" (None, Comma, Period), and "Output Formatting Options" (Flags, Codes, Units, Null Values). A "file_download" button is visible in the bottom right corner.

* Pour le point 4 : les données liées à la sous nutrition [FAOSTAT](#)

* Pour le point 5 : les céréales :

The screenshot shows the FAO New Food Balances interface. The page is titled "New Food Balances" and includes a search bar and navigation tabs: "DOWNLOAD DATA", "VISUALIZE DATA", "METADATA", and "REPORT". The main content area is divided into several sections: "COUNTRIES", "ELEMENTS", "ITEMS", and "YEARS".

- COUNTRIES:** Filter results e.g. afghanistan. Selected: World + (Total), Africa + (Total), Africa > (List), Eastern Africa + (Total), Eastern Africa > (List).
- ELEMENTS:** Filter results e.g. total population - both sexes. Selected: Production Quantity.
- ITEMS:** Filter results e.g. population. Selected: Cereals - Excluding Beer > (List).
- YEARS:** Filter results e.g. 2018. Selected: 2018, 2017, 2016, 2015, 2014.

At the bottom, there are options for "Output Type" (Table, Pivot), "File Type" (CSV, XLS), "Thousand Separator in 'Show Data'" (None, Comma, Period), and "Output Formatting Options" (Flags, Codes, Units, Null Values). A "file_download" button is visible in the bottom right corner.

3- Travail demandé :

- Modélisation + Mise en place d'une BDD : MCD, MPD
- définition d'au moins 5 requêtes (« pertinentes ») sur cette BDD
- Nettoyage + Preprocessing du jeu de données.
- Répondre aux axes d'analyse suivants :

- * calculer la population mondiale

- *calculer la proportion des céréales pour l'alimentation animale

- *calculer pour chaque pays et pour chaque produit, la disponibilité alimentaire en Kcal et en Kg de protéines.

- *calculez pour chaque produit le ratio "énergie/poids", que vous donnerez en kcal ?

Indication : Vous pouvez vérifier la cohérence de votre calcul en comparant ce ratio aux données disponibles sur internet, par exemple en cherchant la **valeur calorique d'un oeuf**.

- *calculez le pourcentage de protéines de chaque produit (pour chaque pays).

Indication :Ce pourcentage est obtenu en calculant le ratio "poids de protéines/poids total" (attention aux unités utilisées). Vous pouvez vérifier la cohérence de votre calcul en comparant ce ratio aux données disponibles sur internet, par exemple en cherchant la teneur en protéines de l'**avoine**.

- * Citer les 5 aliments parmi les 20 aliments les plus caloriques, en utilisant le ratio énergie/poids ?

Attention le ratio peut être différent entre les pays !!

- * Citez 5 aliments parmi les 20 aliments les plus riches en protéines ?

- * Calculer, pour les produits végétaux uniquement, la disponibilité intérieure mondiale exprimée en kcal et en Kg protéines pour chaque année et tracer la viz correspondante ?

- * Combien d'humains pourraient être nourris si toute la disponibilité intérieure mondiale de produits végétaux était utilisée pour de la nourriture ? Donnez les résultats en termes de calories, puis de protéines, et exprimez ensuite ces 2 résultats en pourcentage de la population mondiale.

- * Combien d'humains pourraient être nourris si toute la disponibilité alimentaire en produits végétaux (Food), la nourriture végétale destinée aux animaux (Feed) et les pertes de produits végétaux (Waste) étaient utilisés pour de la nourriture ? Donnez les résultats en termes de calories, puis de protéines, et exprimez ensuite ces 2 résultats en pourcentage de la population mondiale.

- * Combien d'humains pourraient être nourris avec la disponibilité alimentaire mondiale ? Donnez les résultats en termes de calories, puis de protéines, et exprimez ensuite ces 2 résultats en pourcentage de la population mondiale.

- * Quelle proportion de la population mondiale est considérée comme étant en sous-nutrition ?

- * Retrouver dans les données des bilans alimentaires les informations relatives aux pays dans lesquels la FAO recense des personnes en sous-nutrition.

- *Repérer les 15 produits les plus exportés par ce groupe de pays.

- * Parmi les données des bilans alimentaires au niveau mondial, sélectionner les 200 plus grandes importations de ces produits (1 importation = une quantité d'un produit donné importée par un pays donné sur l'année choisie)

- * Grouper ces importations par produit, et calculer pour chaque produit :

- * Le ratio entre la quantité destinés aux "Autres utilisations" (Other uses) et la disponibilité intérieure.

- * Le ratio entre la quantité destinée à la nourriture animale et la quantité destinée à la nourriture (animale + humaine)

- * Top 3 des produits qui ont la plus grande valeur pour chacun des 2 ratios

- *Combien de tonnes de céréales pourraient être libérées si les USA diminuaient leur production de produits animaux de 10% ?

- * Parmi les documents sur les Bilans alimentaires, il y a des informations redondantes. En effet, pour un pays donné, certaines de ces informations peuvent se calculer à partir d'autres :

- Production (*Production*)
- Importations (*Import Quantity*)

- Exportations (*Export Quantity*)
- Variation de stock (*Stock Variation*)
- Disponibilité intérieure (*Domestic supply*)
- Semences (*Seed*)
- Pertes (*Waste*)
- Nourriture (*Food*), aussi appelée Disponibilité alimentaire (*Food Supply*)
- Aliments pour animaux (*Feed*)
- Traitement (*Processing*)
- Autres utilisations (*Other uses*)

=> Identifiez ces redondances en donnant votre réponse sous forme de formule mathématique). Illustrez cette relation avec l'exemple du blé en France.

Soyez force de proposition sur la suite d'analyse !

4- la soutenance :

- 7 min Introduction / Mise en contexte . Donner les chiffres clés (nombre de personnes qui décèdent à cause de la faim, chiffres de la sous-nutrition mondiale, évolution au cours des années précédentes, évolution possible au cours des années suivantes). Donner différentes causes de la faim...
- 7 min Approfondir l'une de ces causes (ou plusieurs de ces causes) en présentant les résultats des analyses faites. Vous pouvez conclure cette partie en répondant à ces 2 questions par exemple :
- La faim dans le monde résulte t'elle d'un manque de production, ou de problèmes technologiques ?
 - Quelles sont les prévisions de population en 2050 ? Aura t'on besoin d'augmenter la production alimentaire ?
 - Il n'est pas nécessaire de répondre scolairement à chacune des questions du projet durant la soutenance, mais vous devez présenter les chiffres les plus importants que vous avez trouvé en les incluant dans votre analyse.
- 3 min Détaillez des données téléchargées : source, combien de fichiers CSV, à quoi correspondent-ils, à quoi correspond chaque ligne d'un fichier, etc. (il est possible de montrer dans votre présentation des premières lignes des fichiers en les détaillant). Détaillez de la même manière les principales tables (dataframes ou tables sql) utilisées lors de l'analyse.
- 3 min Détaillez vos requêtes avec au moins :
- Une agrégation
 - Une jointure (en justifiant le type de jointure : interne, externe gauche, etc.)
 - Une restriction
- indiquez à chaque fois, les ontraintes d'intégrités utilisées pour chaque opération. n'hésitez pas à mettre des captures d'écran des tables pour expliquer les opérations utilisées
- 5 min Montrez le résultat de chacune des requêtes (sous forme de capture d'écran). Détaillez le code SQL de 2 d'entre elles.
- 5 min Questions-réponses