# Mise en situation

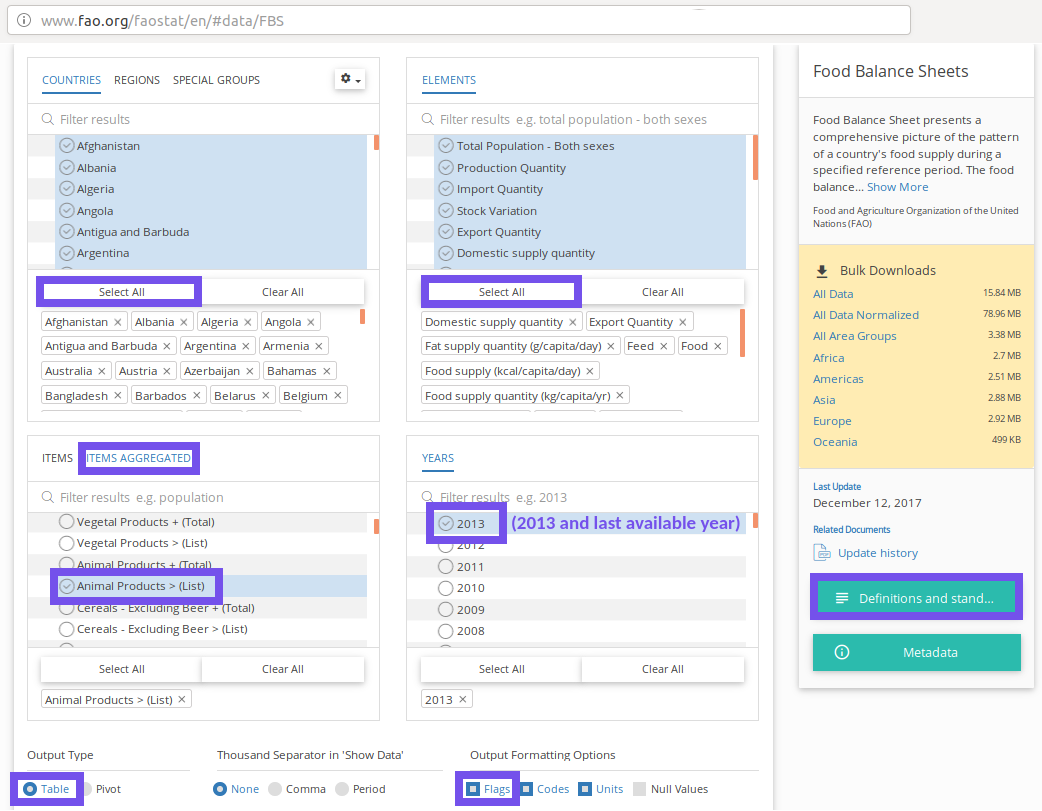
Vous êtes intégré à une nouvelle équipe de chercheurs de la [Food and Agriculture Organization of the United Nations (FAO)](https://fr.wikipedia.org/wiki/Organisation_des_Nations_unies_pour_l%27alimentation_et_l%27agriculture), l'un des organes qui compose l'ONU et dont l'objectif est d' « *aider à construire 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.

# 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***.



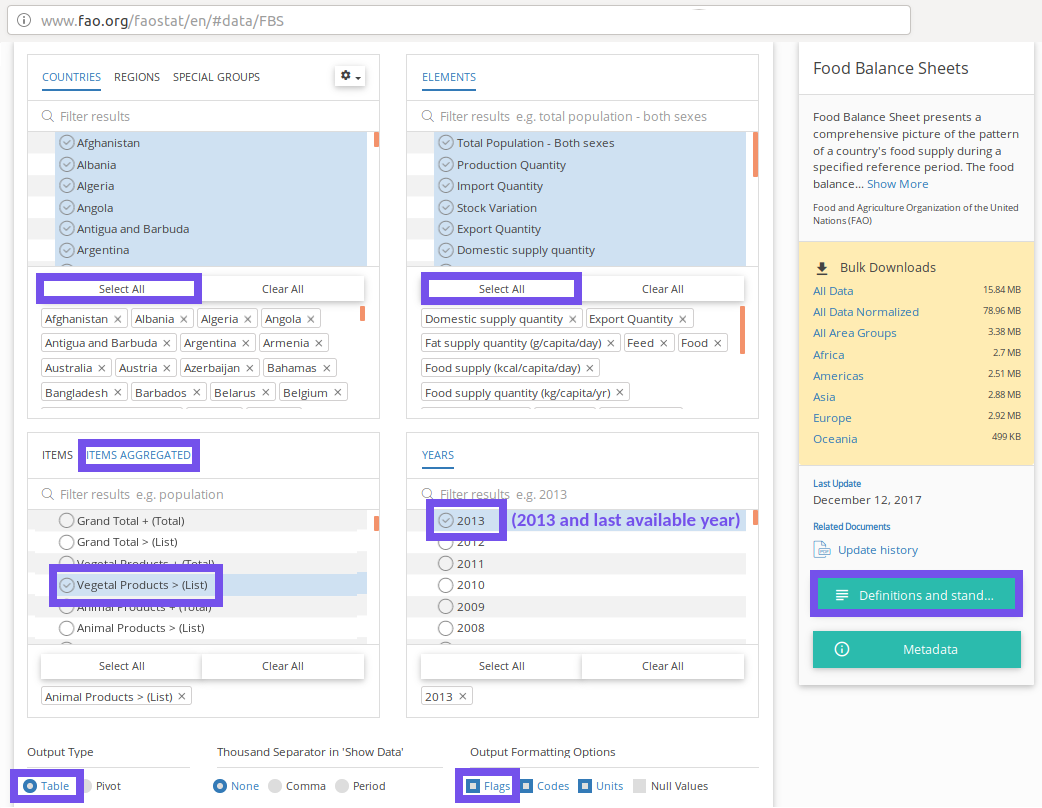
Il faudra prendre un peu de temps pour se familiariser avec cette interface. Avant de télécharger les fichiers CSV, il est possible de tester le filtrage grâce à l’option de prévisualisation (bouton “Montrer les données”).

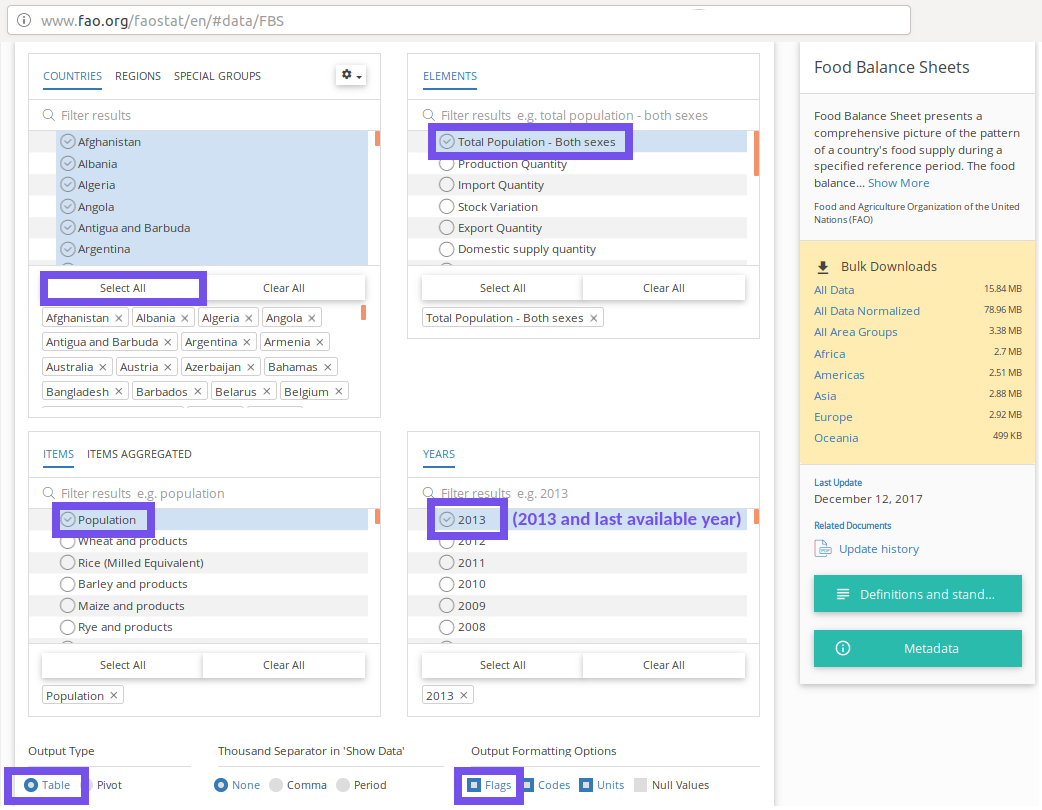
Je vous conseille de se limiter aux cinq tables suivantes (toutes les années depuis 2013) :

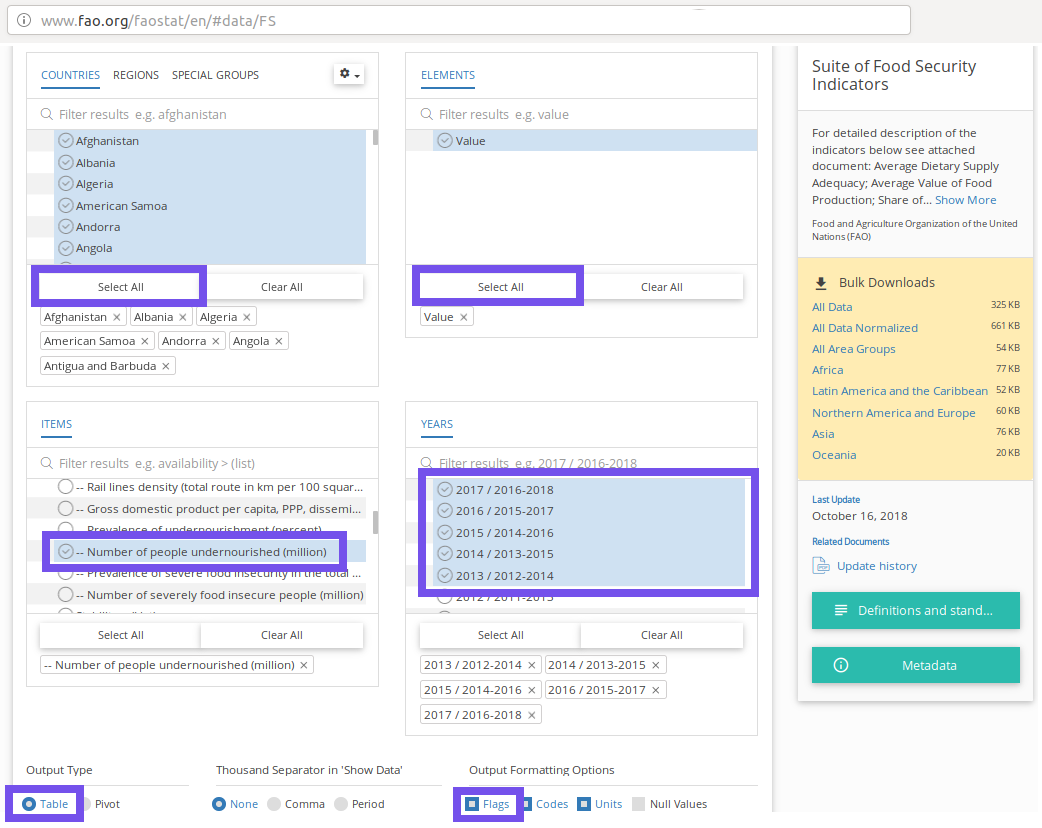
1. Une table des bilans alimentaires pour les produits animaux
2. Une table des bilans alimentaires pour les produits végétaux
3. Une table donnant la population de chaque pays
4. Une table donnant les informations sur le nombre de personnes en sous-nutrition par pays
5. Une liste des produits considérés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

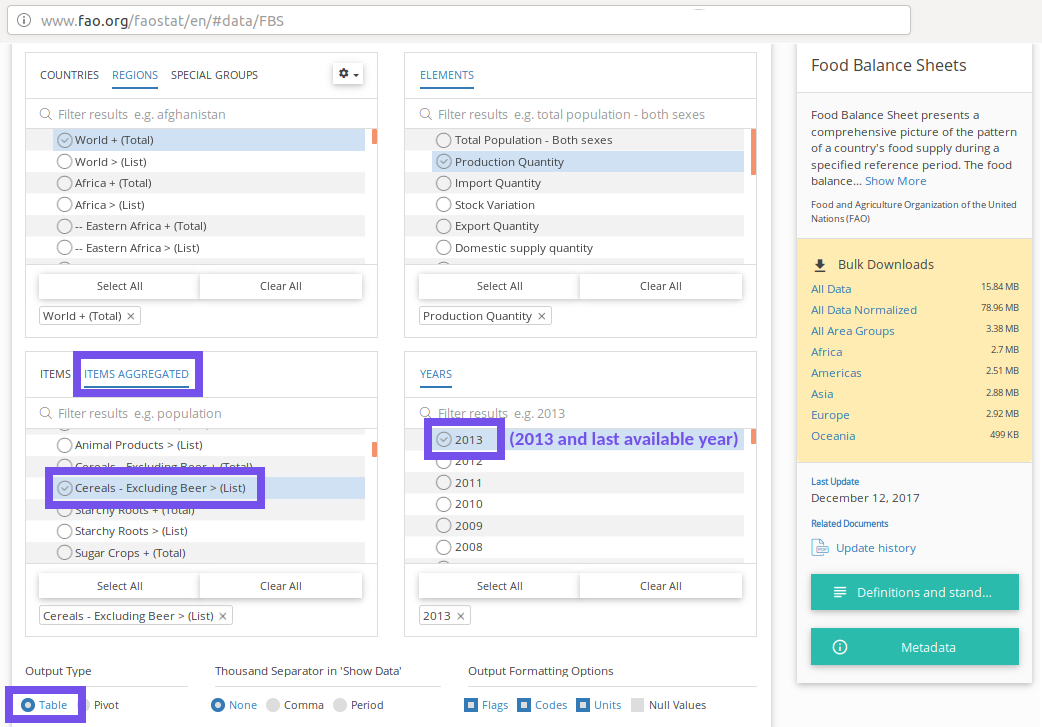
Pour les tables 1 et 2, il faut bien faire attention à ne sélectionner QUE “Produits Végétaux > (Liste)” et “Produits animaux > (Liste)” (voir les captures d’écran ci-dessous). Sinon, il risque d’y avoir de la redondance dans les données (exemple : une ligne contenant “viande bovine” et une ligne contenant “viande”. La ligne “viande” étant une agrégation de “viande bovine” et d’autres types de viandes).

Attention, les données de la FAO changent parfois suite à des actualisations. Il est possible alors que vos résultats d’analyse diffèrent suivant les dates de téléchargement. Mettez une stratégie pour avoir toujours des résultats à jour.









# Entrer vos données dans une BD relationnelle

Lors de la première réunion, vous avez été désigné pour mettre en place la base de données que votre équipe pourra requêter (en SQL) à souhait pour réaliser cette étude.

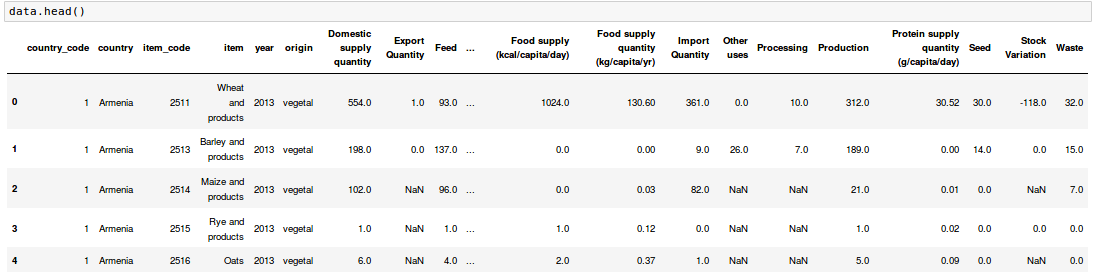
Indice : Pour exporter un dataframe de R ou Python vers une base de données, l'une des solutions est de passer via un fichier CSV.

Sous Python, utilisez to\_csv : df.to\_csv("export.csv", index =False)

Sous R, utilisez write\_csv : readr::write\_csv(df, file ="export.csv")

Avant de passer à la création de la BD, je vous conseille vivement de travailler sur un dataframe principal (qui fusionne *animal* et *vegetal*) de manière à obtenir une colonne pour chaque élément (Production, Importation, Disponibilité intérieure, Semences, etc.) de cette manière :





Vous pouvez pour ceci vous aider de la méthode [**pivot\_table**](https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.pivot_table.html?highlight=pivot_table#pandas.DataFrame.pivot_table), aussi via [cet exemple](https://pbpython.com/pandas-pivot-table-explained.html).

**Votre base devra contenir ces différentes tables :**

# Table population

* Une table appelée **population**, contenant l*a population de chaque pays pour chaque année comprise entre 2013 et l’année courante. Elle devra contenir 4 colonnes* : **country**, **country\_code**, **year**, **population**.

## Q : Taille de la population mondiale ?

En se basant sur le dataframe construit avec Python ou R (à partir des CSV brutes), calculer la taille de la population mondiale (pour chaque année) ?Indice : pour 2013 ; 6 997 326 000 humains. Refaire le même calcul avec SQL ?Indice :même chiffreque la 1è Q. Constater que vosrésultats selon la FAOsontalors bien au-dessus de la taille réelle de la population mondiale (par ex La Banque Mondiale). Trouver la source d’anomalie et effectuer les corrections nécessaires.

**Indice** **:**Il faut remarquer que la Chine est présente 2 fois dans les tables. C’est un problème que vous devriez corriger. Il faut exclure les lignes dont le code\_pays est 351 car il correspond à l’agrégation des pays 41, 96, 241 et 128.

# Table dispo\_alim

* Une table appelée **dispo\_alim** contenant pour chaque pays, pour chaque produit, et pour chaque année, les informations suivantes :
  + La nature du produit (deux valeurs possibles : “*animal*” ou “*végétal*”)
  + Disponibilité alimentaire en tonnes
  + Disponibilité alimentaire en Kcal/personne/jour
  + Disponibilité alimentaire de protéines en g/personne/jour
  + Disponibilité alimentaire de matières grasses en g/personne/jour

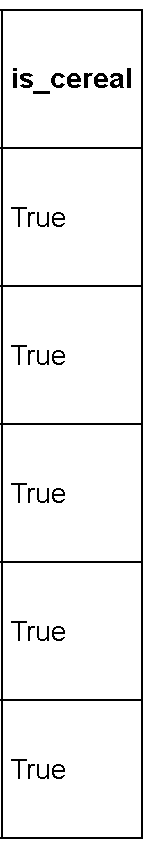
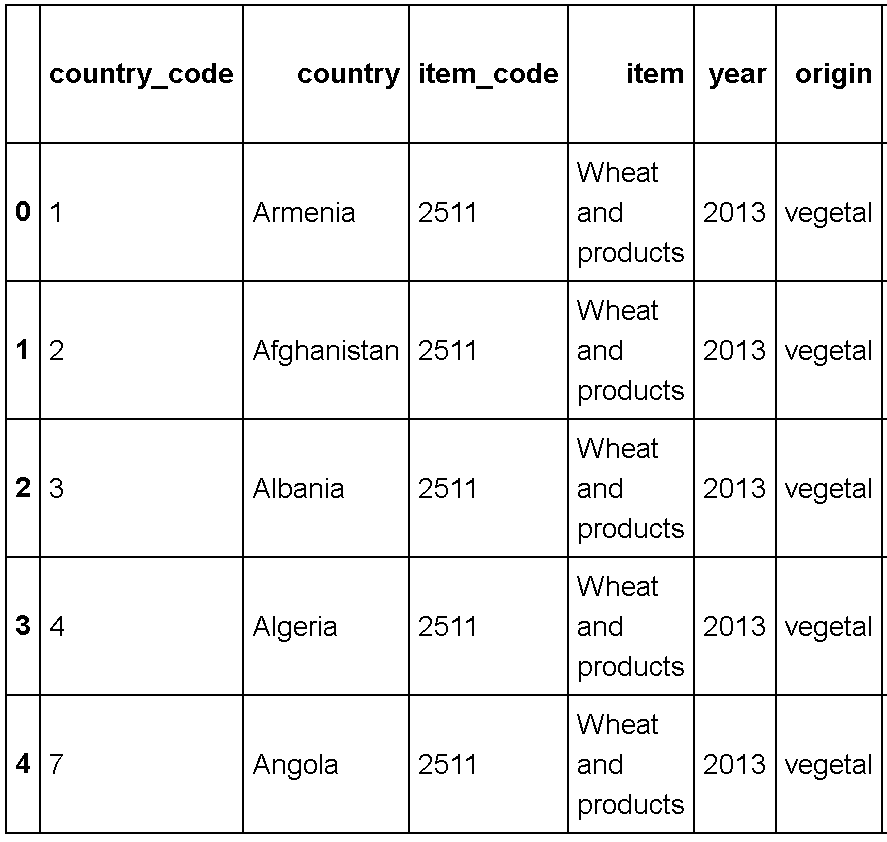
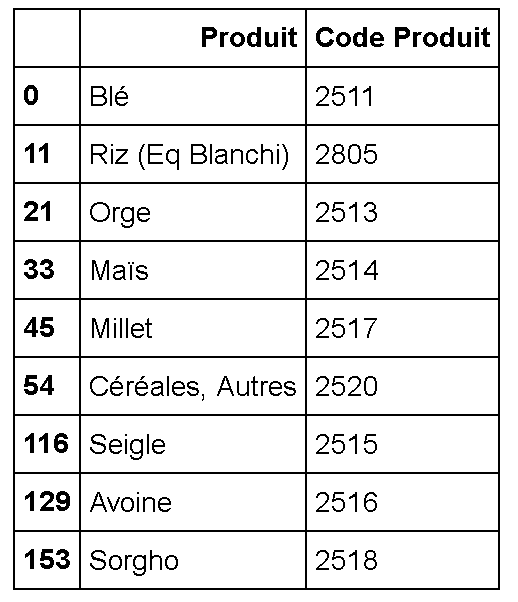
Elle devra contenir ces colonnes : pays, code\_pays, année, produit, code\_produit, origin, dispo\_alim\_tonnes, dispo\_alim\_kcal\_p\_j,dispo\_prot, dispo\_mat\_gr .

## Céréales

### Q : Liste des céréales selon la FAO ?

Établir la liste des produits (ainsi que leur code) considéré comme des céréales selon la FAO et créer une colonne de type booléen nommée "is\_cereal" ?

**Indice** **:**



### Q : Proportion des céréales pour l’alimentation animale ?

En ne prenant en compte que les céréales destinées à l'alimentation (humaine et animale), quelle proportion (en termes de poids) est destinée à l'alimentation animale ?**Indice : 46 %**

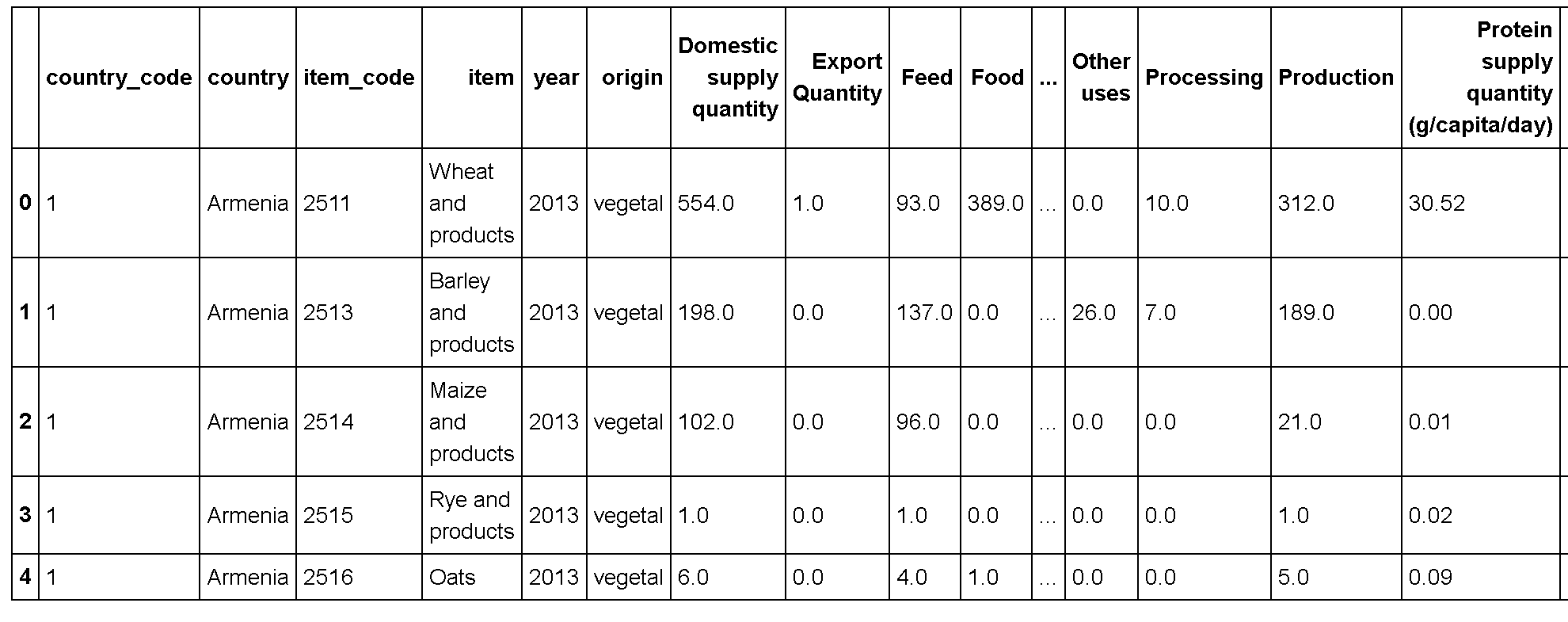
## Comment mesure-t-on la disponibilité alimentaire ?

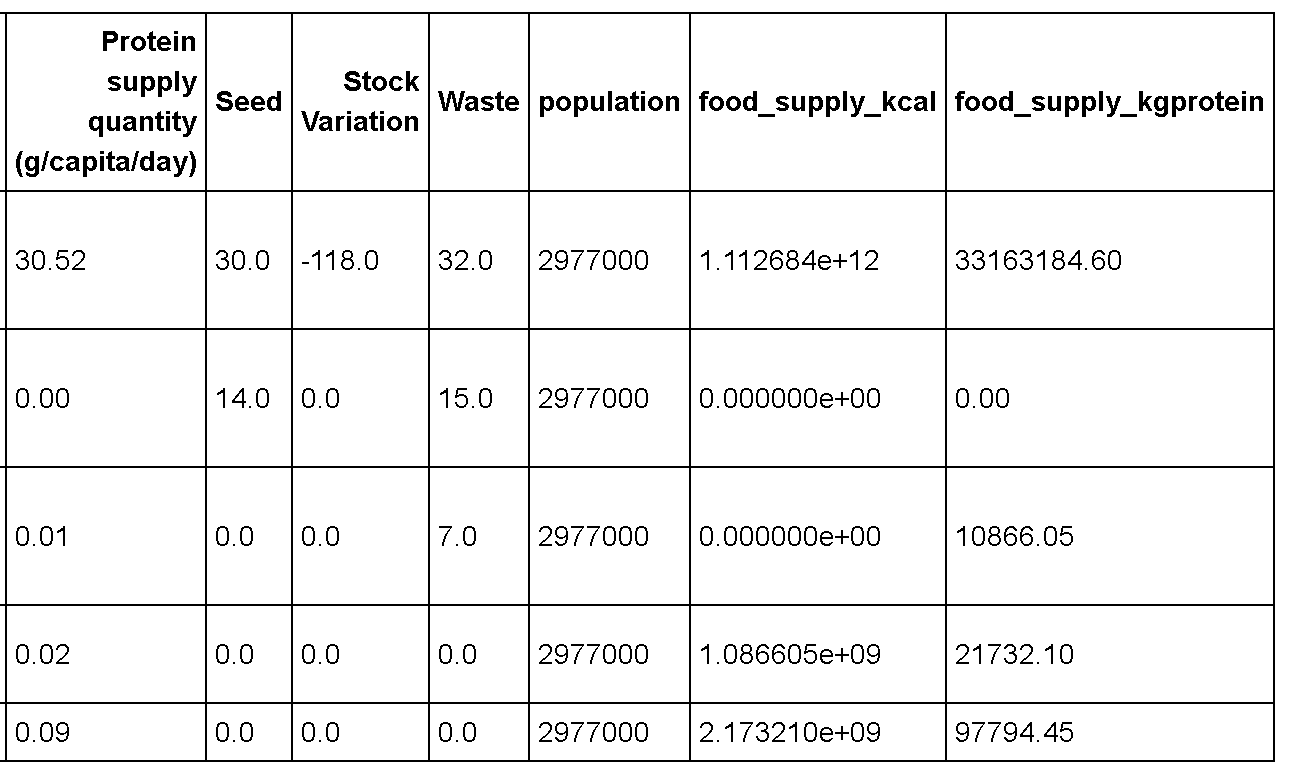
### Q : Dispo alim par pays et par produit ?

Calculer (pour chaque pays et chaque produit) la disponibilité alimentaire en kcal puis en kg de protéines ?

**Indice :**Vous ferez cela à partir de ces informations :

* Population de chaque pays ;
* Disponibilité alimentaire ("Food Supply" en anglais) donnée pour chaque produit et pour chaque pays en kcal/personne/jour.
* Disponibilité alimentaire en protéines ("Protein supply quantity" en anglais) donnée pour chaque produit et pour chaque pays en g/personne/jour.





### Q : Ratio "énergie/poids" ?

A partir de ces dernières informations, et à partir du poids de la disponibilité alimentaire (pour chaque pays et chaque produit), calculez pour chaque produit le ratio "énergie/poids", que vous donnerez en kcal ?

**Indice 1 :**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**](https://fr.wikipedia.org/wiki/%C5%92uf_(aliment)).

**Indice 2 :**La disponibilité alimentaire en kcal/personne/jour est calculée par la FAO en multipliant la quantité Nouriture (Food) par le ratio énergie/poids (en kcal/kg), puis en le divisant par la population du pays puis par 365. Ici, on vous demande juste de retrouver le ratio énergie/poids que la FAO a utilisé dans son calcul.

### Q : % de protéines de chaque produit ?

En suivant la même méthodologie, calculez le pourcentage de protéines de chaque produit (pour chaque pays).

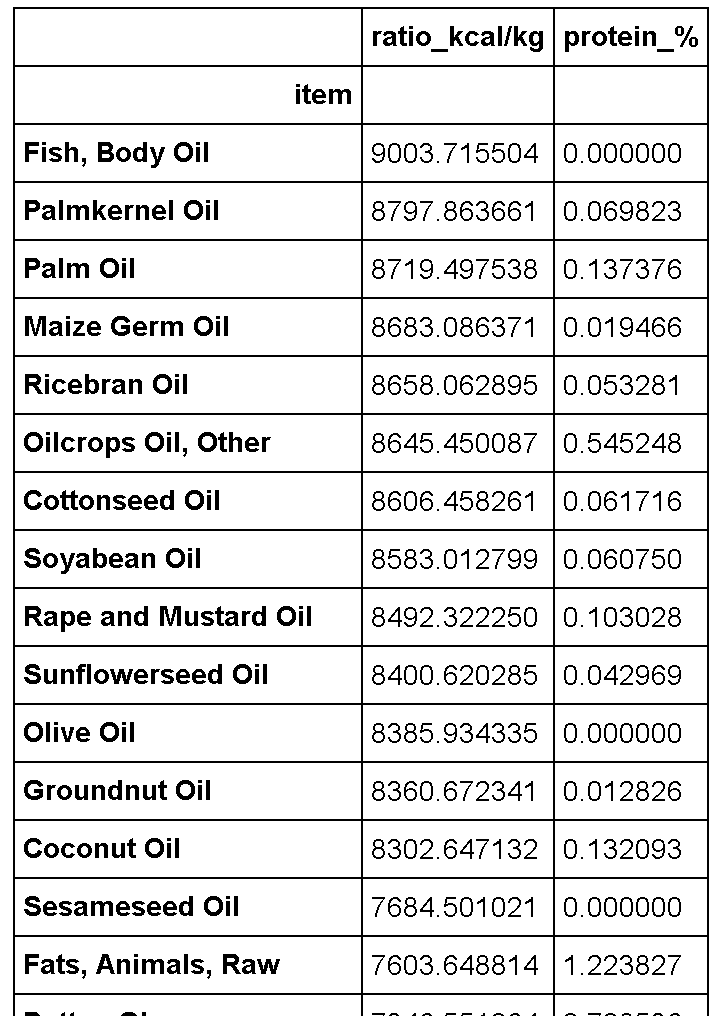
**Indice :** Ce pourcentage est obtenu en calculant le ratio "poids de protéines/poids total" (attention aux unités utilisées). Vous pouvez vérifier la cohérence de votre calcul en comparant ce ratio aux données disponibles sur internet, par exemple en cherchant la teneur en protéines de l'[avoine](https://fr.wikipedia.org/wiki/Avoine_cultiv%C3%A9e).

## Top 20 des aliments

### Q : les plus caloriques

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

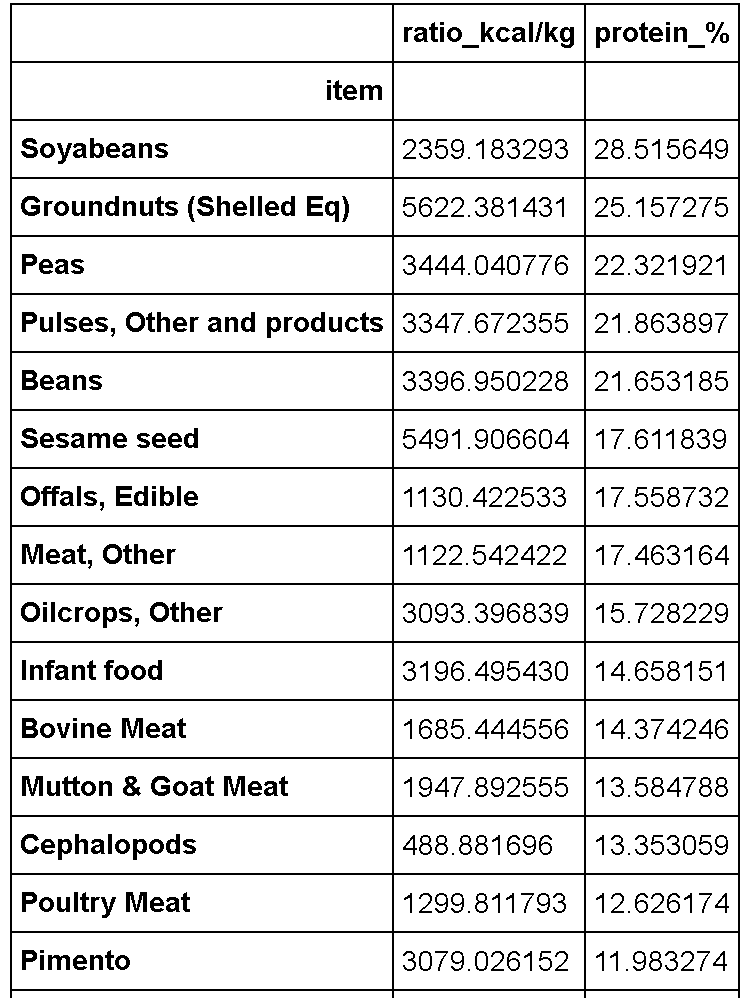
**Indication :** Étonnamment, il arrive que ce ratio soit différent en fonction du pays. Il faudra donc réaliser pour chaque aliment une moyenne sur les différents pays. Vous créerez donc une nouvelle table grâce à une agrégation. Attention à bien retirer les valeurs égales à 0 afin de ne pas fausser le calcul de la moyenne.



### Q : les plus protéinés

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

**Indice :**

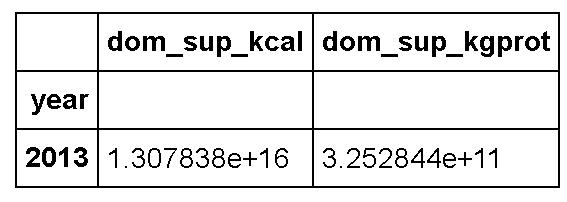


## Disponibilité mondiale

### Q : En végétaux uniquement

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 ?

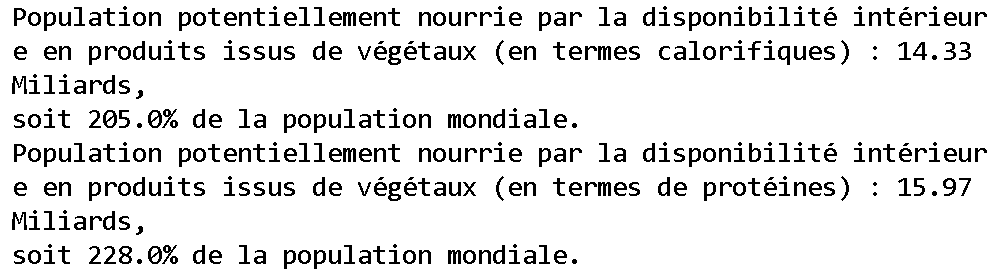
**Indice :**



### Q : Tous végétariens ?

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.

**Indice :**



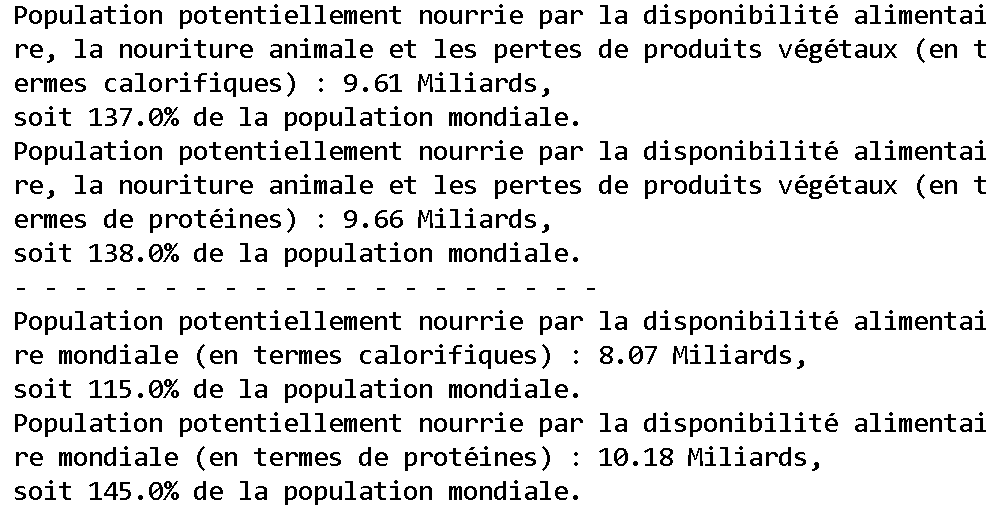
### Q : Rien ne se pert, tout se transforme

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.

### Q : Tous bien nourris ?

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.

**Indice pour les deux dernières questions :**



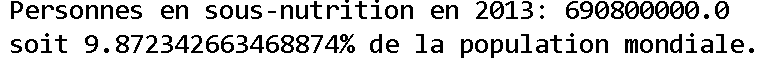
# Table sous\_nutrition

Une table appelée **sous\_nutrition**, contenant le nombre de personnes en sous-alimentation pour chaque pays et pour chaque année. Elle devra contenir 4 colonnes : **pays**, **code\_pays**, **année**, **nb\_personnes**.

## Q : Proportion de la pop en sous-nutrition ?

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

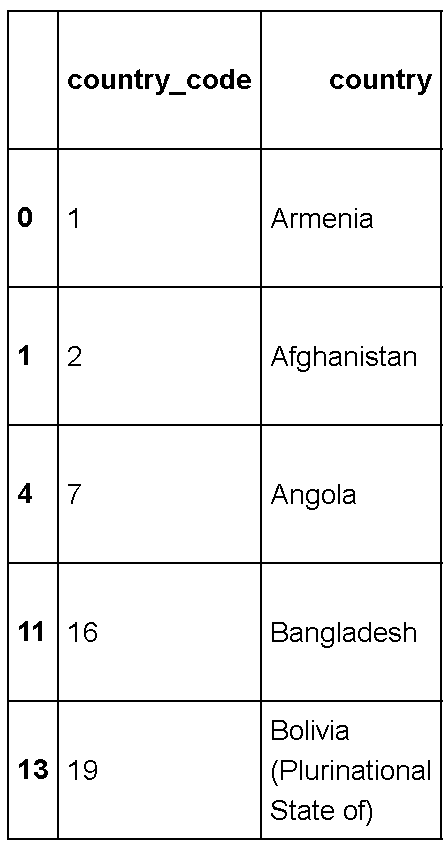
**Indice :**



## Q : Liste des pays en sous-nutrition

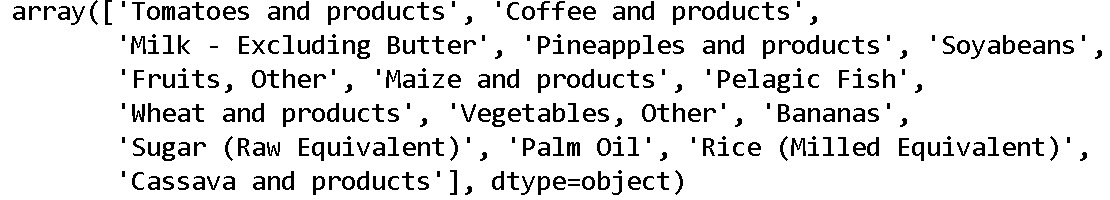
Sélectionnez parmi les données des bilans alimentaires les informations relatives aux pays dans lesquels la FAO recense des personnes en sous-nutrition.

**Indice : liste non exhaustive (et spécifique à l’année 2013)**



## Q : Liste des produits les plus exportés

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



## Q : Les plus grandes importations ?

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)

**Indice : liste non exhaustive**

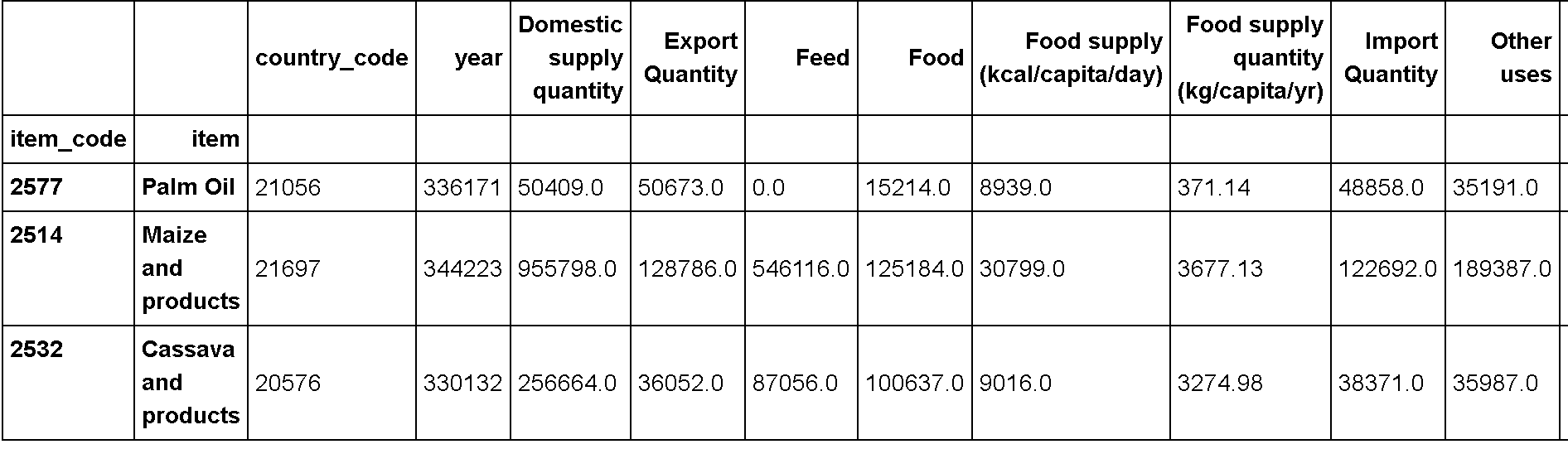


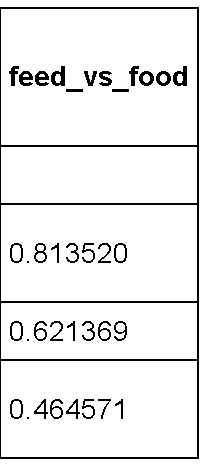
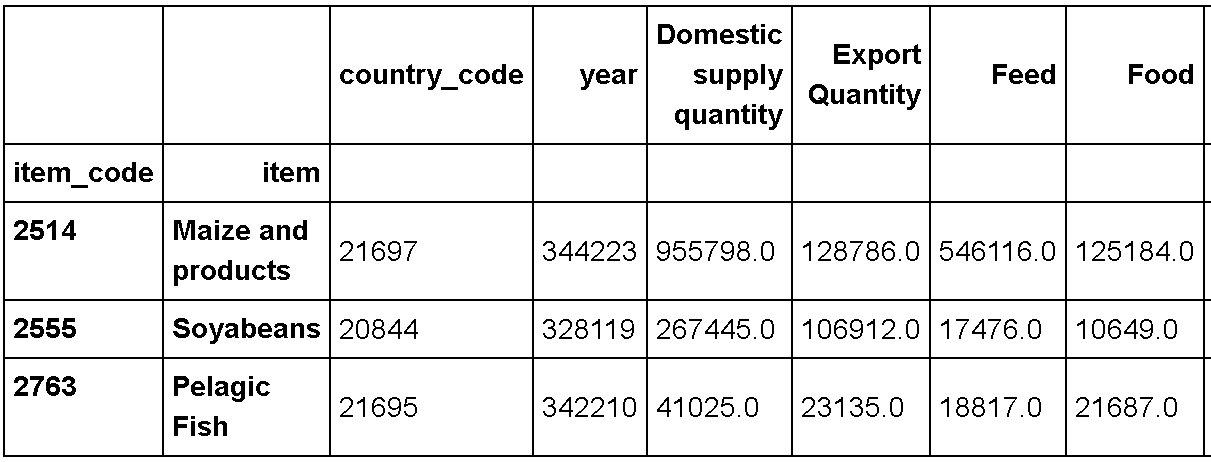
## Q : Regrouper les importations par produit

Grouper ces importations par produit, afin d'avoir une table contenant 1 ligne pour chacun des 15 produits. Ensuite, calculer pour chaque produit les 2 quantités suivantes :

* 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)

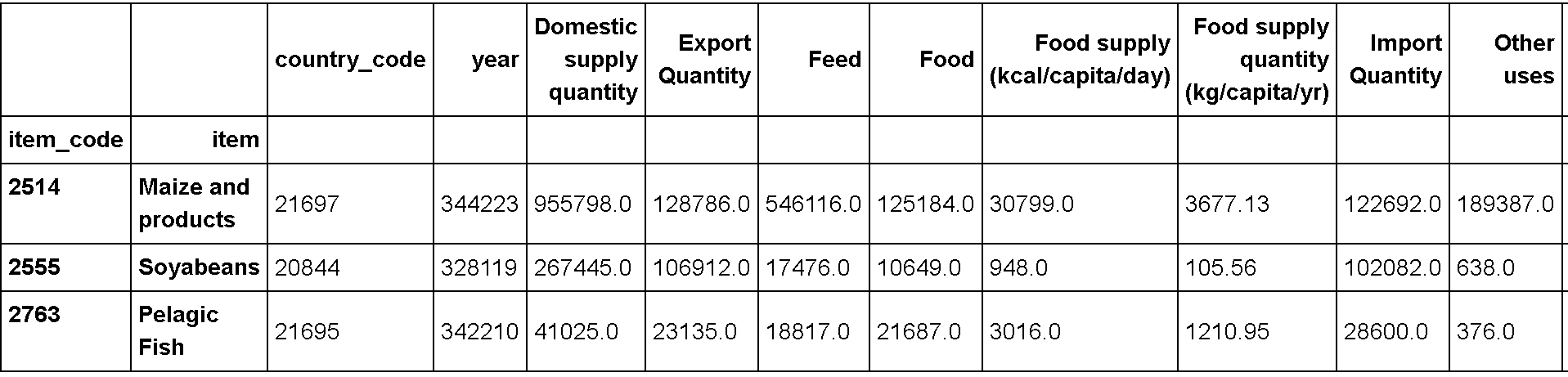
**Indice :**

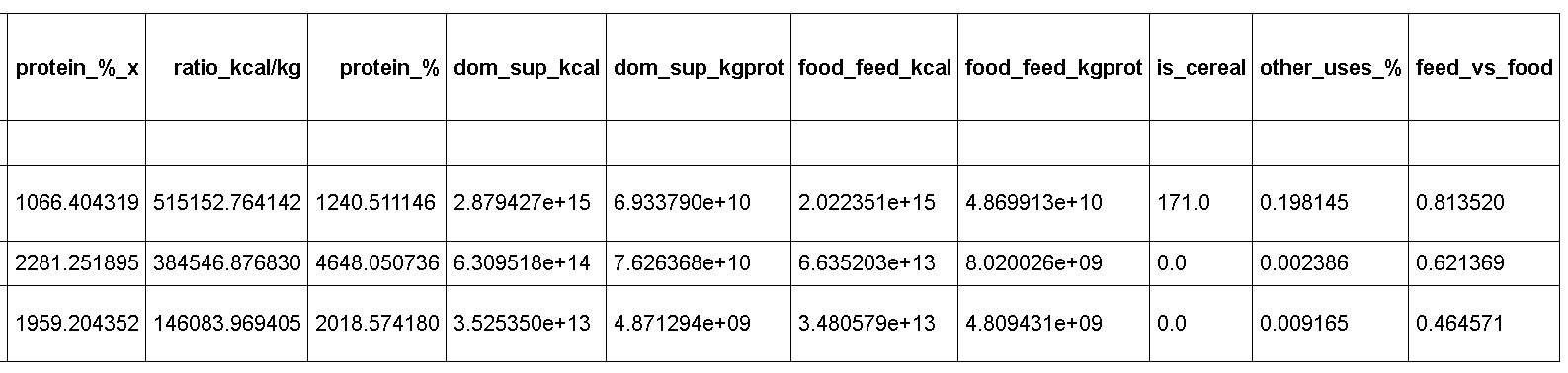




## Q : Top 3 produits

Donnez les 3 produits qui ont la plus grande valeur pour chacun des 2 ratios (vous aurezdonc 6 produits à citer)





## Q :

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

## Q :

En Thaïlande, quelle proportion de manioc est exportée ? Quelle est la proportion de personnes en sous-nutrition ?

# Table equilibre\_prod

* Une table appelée**equilibre\_prod**contenant pour chaque pays, pour chaque produit, et pour chaque année, les quantités suivantes :
  + Disponibilité intérieure
  + Alimentspouranimaux
  + Semences
  + Pertes
  + Transformés
  + Nourriture
  + Autresutilisations

1. Elle devra contenir ces colonnes : pays, code\_pays, année, produit, code\_produit, dispo\_int, alim\_ani, semences, pertes, transfo, nourriture, autres\_utilisations.

## Q :

Parmi les documents sur les Bilans alimentaires que vous avez téléchargés, 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 (*ImportQuantity*)
* Exportations (*ExportQuantity*)
* Variation de stock (*StockVariation*)
* Disponibilité intérieure (*Domesticsupply*)
* Semences (*Seed*)
* Pertes (*Waste*)
* Nourriture (*Food*), aussi appelée Disponibilité alimentaire (*FoodSupply*)
* Aliments pour animaux (*Feed*)
* Traitement (*Processing*)
* Autres utilisations (*Otheruses*)

Identifiez ces redondances en donnant votre réponse sous forme de formule mathématique:soleil:). Illustrez cette relation avec l'exemple du blé en France (indice = 20298).

**Indice** : Cliquez sur "Définitions et Standards" sur [la page de téléchargement](http://www.fao.org/faostat/fr/#data/FBS)des données. Il s’agit de la “balance” des produits alimentaires pour chaque pays.C'est une équation à 3 termes faisant intervenir chacune des 11 quantités données ci-dessus :

a1+a2+[...]=b1+b2+[...]=c1+c2+[...]a1+a2+[...]=b1+b2+[...]=c1+c2+[...]

*Production* + *Importations* - *Exportations* + *Variation de stock*

=

*Disponibilité intérieure*

=

*Semences* + *Pertes* + *Nourriture* + *Alimentspouranimaux* + *Traitement* + *Autresutilisations*

Remarque importante : Il arrive que l'égalité entre les termes 2 et 3 de l’équation ne soit pas toujours vraie (surtout quand domestic supply est négatif), mais les impacts sont négligeables sur le reste des calculs.

# Q : Query SQL

Écrivez les requêtes SQL permettant de connaître… :

1. Les 10 pays ayant **le plus haut ratio disponibilité alimentaire/habitant** en termes de protéines (en kg) par habitant, puis en termes de kcal par habitant.
2. Pour chaque année disponible, les 10 pays ayant **le plus faible ratio disponibilité alimentaire/habitant** en termes de protéines (en kg) par habitant. Le nombre de lignes de la table renvoyée sera donc égal à 10 fois le nombre d'années disponibles.
3. La quantité totale (en kg) de produits perdus par pays et par année. La table renvoyée contiendra donc une ligne par couple (pays, année).
4. Les 10 pays pour lesquels la proportion de personnes sous-alimentées est la plus forte.
5. Les 10 produits pour lesquels le ratio Autres utilisations/Disponibilité intérieure est le plus élevé.

#### Enrichir votre analyse

Chercher sur le site de la FAO les réponses à ces questions :

* Combien de personnes décèdent des causes de la faim ?
* Quelles sont les prévisions de population mondiale en 2050 ?

# Modalités de la soutenance

|  |  |
| --- | --- |
| 7 min | Mise en contexte auprès de votre équipe. Donnez 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). Donnez ensuite différentes causes de la faim. |
| 7 min | Approfondissez l'une de ces causes (ou plusieurs de ces causes) en présentant les résultats de vos calculs. Vous conclurez 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 drastiquement 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 quelques requêtes avec au moins :   * Une agrégation * Une jointure (en justifiant le type de jointure : interne, externe gauche, etc.) * Une restriction   Pour chacune des opérations d'algèbre relationnelle détaillée, indiquez les clés primaires des tables avant et après opération. Il conseillé de donner 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 de l'examinateur |