# Brief Projet Data Analyst au sein de la FAO

En complément des informations données dans le brief projet sur Simplonline, vous trouverez ci-joint les attendus de ce projet.

## Analyses descriptives et exploratoire des données

Il s’agit ici de réaliser des analyses descriptives statistiques et graphiques sur les différentes variables des jeux de données :

* Analyses univariées
* Analyses bi-variées
* Analyses multi-variées

Réaliser également des analyses exploratoires sur les différentes variables

## Etude guidée n°1

Il s’agit ici de répondre à toutes les questions ci-dessous.

### Question 1 : Nombre d'humains sur la planète

Calculez le nombre total d’humains sur la planète.  
Critiquez votre résultat. En cas d’anomalie, analysez et effectuer les corrections nécessaires.  
Donnez le résultat de votre calcul pour l'année 2013.

### Question 2 : Redondances

Identifiez ces redondances, en donnant votre réponse sous forme de formule mathématique (pas besoin de coder ici).  
C'est une équation à 3 termes de type (a\_1 + a2 + [...] = b\_1 + b\_2 + [...] = c\_1 + c\_2 + [...]) ) faisant intervenir chacune des 11 quantités données ci dessus.  
Illustrez cette équation avec l'exemple du blé en France.

### Question 3 : Disponibilité alimentaire (calories, protéines)

Calculez (pour chaque pays et chaque produit) la disponibilité alimentaire en kcal puis en kg de protéines. Vous ferez cela à partir de ces informations :   
- Population de chaque pays  
- Disponibilité alimentaire donnée pour chaque produit et pour chaque pays en kcal/personne/jour, - - Disponibilité alimentaire en protéines donnée pour chaque produit et pour chaque pays en g/personne/jour.

Pour cette étape vous avez besoin de constituer une seule et même table à partir des tables animaux et végétaux.

### Question 4 : 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/kg. 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.

### Question 5 : Aliments les plus caloriques et protéiques

En considérant les aliments végétaux et animaux, citez 5 aliments parmi les 20 aliments les plus caloriques, en utilisant le ratio énergie/poids. Citez 5 aliments parmi les 20 aliments les plus riches en protéines.

### Question 6 : Dispo. intérieure mondiale des végétaux

Calculez, pour les produits végétaux uniquement, la disponibilité intérieure mondiale exprimée en kcal.

### Question 7 : Potentiel alimentaire des végétaux

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.

### Question 8 : Potentiel alimentaire des végétaux (destinés aux animaux pertes)

Combien d'humains pourraient être nourris si toute la disponibilité alimentaire en produits végétaux la nourriture végétale destinée aux animaux et les pertes de produits végétaux 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.

### Question 9 : Potentiel alimentaire de la dispo. 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.

### Question 10 : Proportion de la sous-nutrition mondiale

A partir des données téléchargées qui concernent la sous-nutrition, répondez à cette question : Quelle proportion de la population mondiale est considérée comme étant en sous-nutrition ?

### Question 11 : Céréales

Établissez la liste des produits (ainsi que leur code) considéré comme des céréales selon la FAO. 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 ?

### Question 12 : Sous-nutrition

Sélectionnez parmi les données des bilans alimentaires les informations relatives aux pays dans lesquels la FAO recense des personnes en sous-nutrition.  
Repérez les 15 produits les plus exportés par ce groupe de pays.  
Parmi les données des bilans alimentaires au niveau mondial, sélectionnez les 200 plus grandes importations de ces produits (1 importation = une quantité d'un produit donné importée par un pays donné).  
Groupez ces importations par produit, afin d'avoir une table contenant 1 ligne pour chacun des 15 produits. Ensuite, calculez 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)

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

### Question 13 : USA

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

### Question 14 : Thaïlande

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

## Etude guidée n°2

Il s’agit ici d’intégrer toutes les données dans une base de données relationnelle puis de répondre à toutes les questions ci-dessous.  
La base de données contiendra 4 tables : **population, dispo\_alim, equilibre\_prod, sous\_nutrition.**

### Question 15 : Table population

Définir une clé primaire pertinente pour la table population. Intégrer la table dans la base de données relationnelle.

### Question 16 : Table dispo\_alim

La table appelée dispo\_alim contient pour chaque pays, pour chaque produit, sur l'année 2013, 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

La table contient ces colonnes : pays, code\_pays, année, produit, code\_produit, origin, dispo\_alim\_tonnes, dispo\_alim\_kcal\_p\_j, dispo\_prot, dispo\_mat\_gr .

Définir une clé primaire pertinente pour la table dispo\_alim. Intégrer la table dans la base de données relationnelle.

### Question 17 : Table equilibre\_prod

La table appelée equilibre\_prod contient pour chaque pays, pour chaque produit, et pour l'année disponible 2013, les quantités suivantes :  
- disponibilité intérieure  
- aliments pour animaux  
- semences  
- pertes  
- transformés  
- nourriture  
- autres utilisations

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

Définir une clé primaire pertinente pour la table equilibre\_prod. Intégrer la table dans la base de données relationnelle.

### Question 18 : Table sous\_nutrition

La table appelée sous\_nutrition, contient le nombre de personnes en sous-alimentation pour chaque pays et pour chaque année. Elle contient 4 colonnes : pays, code\_pays, année, nb\_personnes.

Définir une clé primaire pertinente pour la table sous\_nutrition. Intégrer la table dans la base de données relationnelle.

### Question 19 : Requêtes SQL

Formuler les requêtes SQL pour répondre aux questions suivantes :

* Quels sont 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 ?
* Quels sont les 10 pays ayant le plus faible ratio disponibilité alimentaire/habitant en termes de protéines (en kg) par habitant ?
* Quelle est la quantité totale (en kg) de produits perdus par pays?
* Quels sont les 10 pays pour lesquels la proportion de personnes sous-alimentées est la plus forte ?
* Quels sont les 10 produits pour lesquels le ratio Autres utilisations/Disponibilité intérieure est le plus élevé ?

### Question 20 : Autres utilisations des produits

Pour quelques-uns des produits identifiés dans cette dernière requête SQL, supposez quelles sont ces "autres utilisations" possibles (recherchez sur internet !).

## Analyses complémentaire

Fort de votre expérience de Data Analyst, il s’agit maintenant de réaliser vos propres analyses et exploration.