

# P4 - Réalisez une étude de santé publique avec R ou Python

90 heures



Félicitations ! Vous avez récemment rejoint une équipe de chercheurs de la [Food and Agriculture Organization of the United Nations \(FAO\)](#), en tant que data analyst.



**Food and Agriculture  
Organization of the  
United Nations**

Logo de la FAO

Son rôle ? C'est l'un des organes qui composent l'ONU et dont l'objectif est « d'aider à construire un monde libéré de la faim ».

Le responsable de l'équipe, Marc, un chercheur en économie de la santé, est enthousiaste à votre arrivée. En effet, Julien, l'ancien data analyst, a récemment été muté, et votre équipe vient de se voir confier la réalisation d'une **étude de grande ampleur** sur le thème de l'alimentation et plus particulièrement sur la sous-nutrition dans le monde.

Dès votre arrivée, vous trouvez un mail de Marc :

**De :** Marc

**À :** Moi

**Objet :** Bienvenue et briefing du projet

Bienvenue !

Tout d'abord, sache que je suis très content de t'accueillir au sein notre équipe !

Comme tu le sais, nous sommes en train de réaliser une grosse étude sur l'alimentation dans le monde. Je souhaite t'intégrer à ce projet afin que tu puisses nous aider à produire les différentes analyses.

Avant qu'il ne parte, nous avons validé avec Julien les analyses que nous pourrions réaliser.

Tu trouveras en pièces jointes :

- l'**ébauche de présentation** de Julien
- les **données**
- le **lexique**.

Il me semble d'ailleurs que Julien avait rajouté des commentaires dans son ébauche de présentation, j'espère que cela pourra t'aider.

N'hésite pas également à ajouter des analyses complémentaires que tu trouveras pertinentes.

J'ai failli oublier ! Julien avait également préparé un **notebook** avec les différentes étapes à réaliser pour répondre aux questions. Il ne savait pas encore s'il utiliserait Python ou R pour réaliser cette étude.

Est ce que tu as une préférence de ton côté ?

Dans tous les cas, il m'avait recommandé deux cours sur la plateforme OpenClassrooms, n'hésite pas à aller les voir :

- [Cours sur Python](#)
- [Cours sur R](#)

Dernière précision, tu n'es pas obligé d'utiliser les trames du notebook ou du PowerPoint mais elles peuvent être une aide précieuse pour ta première analyse !

N'hésite pas à venir me voir si tu as la moindre question !

Marc

Expert en nutrition et sécurité alimentaire - FAO

PJ1 : [Ébauche de trame de présentation](#)

PJ2 : [Données FAO.zip](#)

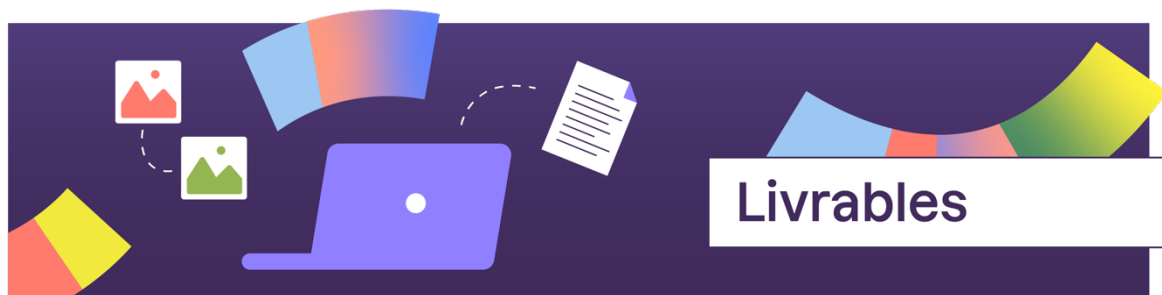
PJ3 : [Lexique](#)

PJ4 : Notebooks [Python](#) et [R](#)

Vous êtes prêt ? Votre mission démarre :

- Coupez dès à présent toutes les sources de distraction.
- Préparez votre environnement de travail (onglet, documents, raccourcis...)
- Démarrez votre "exploration".

Bon courage !



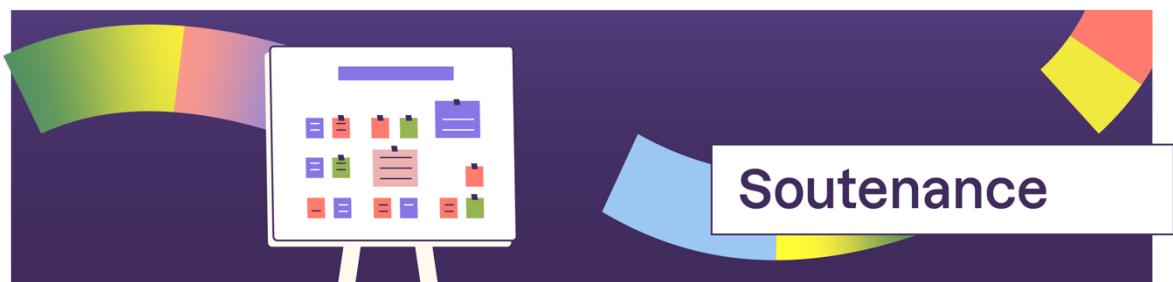
1. Le **notebook R** Markdown ou **Jupyter** (en Python) vous ayant permis de calculer/trouver l'ensemble des résultats demandés.
2. Le **notebook R** Markdown ou **Jupyter** (en Python) en version **PDF**.
3. Le **support de présentation** contenant toutes les informations demandées en version **PPT** ou **Gslides**.

Pour faciliter votre passage devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre\_du\_projet\_nom\_prenom**", tous les livrables du projet comme suit : **Nom\_Prenom\_n° du livrable\_nom du livrable\_date de démarrage du projet**. Cela donnera :

- Nom\_Prenom\_1\_notebook\_mmaaaa ;
- Nom\_Prenom\_2\_notebook-pdf\_mmaaaa ;
- Nom\_Prenom\_3\_presentation\_mmaaaa.

Par exemple, le premier livrable peut être nommé comme suit :

*Dupont\_Jean\_1\_notebook\_012023*



**Soutenance (30 minutes)**

Durant la présentation orale, l'évaluateur interprétera le rôle de Marc, le responsable de l'équipe. La soutenance est structurée de la manière suivante :

- **Présentation des résultats (15 minutes).**
  - Vous présenterez votre notebook et votre support de présentation répondant aux questions.
- **Discussion (10 minutes) :**
  - L'évaluateur jouera le rôle de Marc. Il vous challengera sur vos choix, et vérifiera vos connaissances sur certaines fonctions en R ou Python, notamment :
    - la propreté et l'organisation du code :
    - les opérations d'algèbre relationnelle ( jointure), la restriction (.loc et/ou sélection de colonnes) et l'agrégation
  - Il vous demandera si le RGPD est respecté dans les data set nettoyés
- **Debrief (5 minutes)**

À la fin de la soutenance, l'évaluateur arrêtera de jouer le rôle de Marc pour vous permettre de débriefer ensemble.

Votre présentation devrait durer 15 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, les présentations en dessous de 10 minutes ou au-dessus de 20 minutes peuvent être refusées.

## **Compétences évaluées**

- Manipuler des DataFrames
- Rédiger et présenter une méthodologie d'exploration et d'analyse des données
- Utiliser des librairies spécialisées pour les traitements data