

## Scénario

Vous travaillez chez La poule qui chante, une entreprise française d'agroalimentaire. Elle souhaite se développer à l'international.



# La poule qui chante

L'international, oui, mais pour l'instant, le champ des possibles est bien large : aucun pays particulier ni aucun continent n'est pour le moment choisi. Tous les pays sont envisageables !

Votre manager, Patrick, vous briefe par un e-mail :

**Objet** : Lancement mission data international

**De** : Patrick

**À** : Moi

Salut !

Comme on en a déjà brièvement parlé, je te fais ce mail pour te briefe sur la mission d'analyse pour le lancement à l'international.

Ton objectif sera de proposer une première analyse des groupements de pays que l'on peut cibler pour exporter nos poulets. Nous approfondirons ensuite l'étude de marché.

Tu seras en totale autonomie sur ce projet, notamment sur le choix des données à analyser et même du langage à utiliser (R ou Python).

Pars des données de la FAO (*Food and Agriculture Organization*) que je te mets en pièce jointe pour faire ton analyse. Si tu veux aller plus loin,

notamment avec les critères de l'analyse PESTEL, tu peux récupérer et utiliser toutes les données en *open data* que tu souhaites sur [le site de la FAO](#).

Pour la partie analyse, dans un premier temps j'aimerais que tu testes la classification ascendante hiérarchique, avec un dendrogramme comme visualisation. Ensuite tu pourras utiliser la méthode des k-means, afin d'affiner l'analyse et comparer les résultats des deux méthodes de clustering. N'hésite pas à prendre le temps d'analyser les centroïdes de tes classes. Tu peux également réaliser une ACP afin de visualiser les résultats de ton analyse, comprendre les groupes, les liens entre les variables, les liens entre les individus...

Voici ce que j'attends de toi :

- un notebook Jupyter ou un fichier R avec tes analyses ;
- une heatmap avec les croisements entre les clusters de pays et les différentes variables (en format image), afin que je puisse tout de suite me faire une idée des différents groupes que tu auras définis.

Bon courage ! 😊

**Patrick**

**Pièce jointe :** [Données New Food Balances \(FAO\)](#)

Vous avez toutes les cartes en main. À vous de jouer !

## Livrables

1. Un notebook R Markdown ou Jupyter contenant vos analyses.
2. Une heatmap avec vos clusters et les différentes variables (format image).

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

- *Nom\_Prénom\_1\_notebook\_mmaaaa*
- *Nom\_Prénom\_2\_heatmap\_mmaaaa*

Par exemple, le premier livrable peut être nommé comme suit : *Dupont\_Jean\_1\_notebook\_012022*.

## **Soutenance**

Pendant la soutenance, l'évaluateur jouera le rôle de votre manager à qui vous présentez vos résultats :

- **Présentation (15 minutes) avec ou sans support de présentation**
  - Vous rappellerez le contexte du projet de data analyse ;
  - Vous expliquerez votre démarche (sans entrer dans les détails mathématiques) ;
  - Vous présenterez vos résultats et recommandations.
- **Discussion (10 minutes)**
  - Toujours dans le rôle de manager, l'évaluateur vous posera des questions sur votre travail et sur vos choix.
- **Débriefing (5 minutes)**
  - À la fin de la soutenance, l'évaluateur arrêtera de jouer le rôle du responsable du pôle Data 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**

- Effectuer un clustering simple
- Explorer des données pour synthétiser des variables