

# Mini-projet technique

## Recrutement Sustaain 2024-02

Votre client est un industriel impliqué dans la chaîne de production de chocolat et de café. Il a des fournisseurs et partenaires dans différents pays du monde, en particulier dans quelques pays d'Afrique.

Ce test technique pose des questions volontairement ouvertes. N'hésitez pas à formuler des hypothèses, à justifier votre démarche, à expliquer ce que vous feriez même si vous ne l'implémentez pas ici-même. Ne passez pas plus d'une journée sur ce travail ! 😊

Si vous êtes invité à un entretien technique, vous aurez l'occasion de parler de ce travail.

### **Les livrables minimaux attendus pour ce test :**

(merci de fournir des livrables dans lesquels les codes développés puissent être relancés et répétés)

- un notebook jupyter/colab (ne pas hésiter à implémenter un pipeline, une petite app, par exemple pour présenter l'accès structuré à la donnée et/ou présenter des visualisations - tant que c'est testable par le relecteur)
- un document attaché selon le besoin (ex. pour dessiner un schéma)
- une présentation slides (pdf, maximum 8 slides) à montrer au client dans une réunion de rendu synthétique.

## Question 1

Le client voudrait avoir une vue d'ensemble de la filière cacao et café dans les pays de cette partie d'Afrique. Il vous demande une analyse de ces données, et un rendu synthétique de cette analyse lui présentant des indicateurs et leur évolution temporelle, selon les pays, etc. Le client est spécialement attentif à la transparence de l'analyse, il souhaite connaître la fiabilité des chiffres / visualisations présentés.

## Question 2

Le client aimerait intégrer des textes réglementaires issus de documents PDF dans sa base de données.

Il souhaite donc disposer d'un outil permettant de

1. extraire le texte d'un document PDF
2. structurer le contenu extrait (par exemple, distinguer les titres des paragraphes, les métadonnées du document...)
3. le stocker dans un format adapté sur un système de stockage adapté
4. résumer un paragraphe voire le texte entier d'un document

Proposez un schéma de principe de la structure qu'aurait un tel outil ; n'hésitez pas à dessiner un schéma, et à décrire les pipelines de données (acquisition, traitement...), comment les déclencher, quels stockages, quelles librairies etc. Écrivez un code (/ des morceaux de code) réalisant une ou plusieurs parties des fonctionnalités demandées.

Utilisez les deux documents suivants comme exemples de documents PDF concernés.

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2464>

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023R1115>

## Question 2-bis

Pouvez-vous préciser ces solutions (texte/schéma) dans le cas où le client a

- 100 ensembles de documents différents ?
- 1000 ensembles de documents différents + des données tabulaires statiques + des données temporelles avec fréquence d'update annuel pour certaines, hebdomadaire pour d'autres ?

## Question 3

Pouvez-vous donner une estimation de l'évolution des rendements ("yield") dans chaque pays concerné, pour le cacao et le café, sur les 5 prochaines années ? Merci d'expliquer votre démarche et de l'illustrer / quantifier autant que possible.

## Question 4

En supposant que, sur les exploitations du client dans cette zone du monde :

- Les réglementations internationales dans la filière cacao imposent une taxe de 30% de la valeur exportée pour la fève de cacao brute et 2% pour les produits transformés issus du cacao

... et en formulant toute hypothèse nécessaire,

Le client a-t-il intérêt à investir plutôt pour

- Augmenter la surface des cultures de cacao ?
- Améliorer la production de produits transformés à base de cacao ?

## Question 5

Quelles autres sources de données seraient pertinentes pour améliorer le service apporté au client ?