

Thématique : Contrôle qualité et valorisation des données géospatiales des coopératives agricoles

Livrables :

1. **Dossier technique complet** (rapport PDF ou PowerPoint)
2. **Fichiers produits** (scripts, fichiers QGIS, exports Excel/Power BI, cartes, etc.)
3. **Présentation orale (15 min)** expliquant la démarche, les difficultés rencontrées et les choix techniques effectués

1. Contexte général

Notre entreprise accompagne plusieurs **coopératives de cacao** dans la mise en conformité de leurs données de traçabilité et de durabilité. Chaque coopérative nous transmet régulièrement :

- Un **registre Excel** de producteurs et de fermes déclarées (identifiants, superficies, coordonnées, etc.)
- Un **fichier géospatial** (GeoJSON) représentant les polygones de parcelles cartographiées

Ces données présentent souvent des problèmes : doublons, chevauchements, erreurs topologiques, différences entre surfaces déclarées et calculées, incohérences de coordonnées, etc.

Le rôle du Geo Data Analyst est de **concevoir un pipeline de contrôle qualité et d'analyse spatiale**, puis de **fournir un rapport clair et exploitable** pour la direction et les partenaires.

2. Objectifs du projet

Le candidat doit :

1. **Analyser et nettoyer les données** reçues (Excel + GeoJSON)
2. **Mettre en place un système de contrôle qualité géospatial et tabulaire**
3. **Identifier et documenter** les anomalies : doublons, chevauchements, distances avec forêts protégées, incohérences de surface
4. **Produire un tableau de bord ou rapport analytique** permettant de suivre la qualité des données
5. **Présenter la démarche complète** (outils, méthodes, résultats, limites, recommandations)

3. Données à fournir aux candidats

Fichier	Type	Description
<code>cooperatives_data.xlsx</code>	Excel	Liste des fermes (colonnes : Id, Cooperative, Producteur, Surface_déclarée_ha, Longitude, Latitude)
<code>parcelles.geojson</code>	GeoJSON	Polygones des fermes avec attributs (Id, Surface_calculee, geometry)

4. Travaux demandés

Étape 1 — Préparation et exploration

- Charger les données (Excel + GeoJSON) dans QGIS, Python ou tout programme équivalent
- Vérifier les projections, les types de géométries et la cohérence des identifiants
- Calculer les superficies réelles et comparer avec les surfaces déclarées
- Calculer la Longitude, la Latitude de chaque parcelle

Étape 2 — Contrôle qualité et nettoyage

Mettre en place un protocole de contrôle qualité :

1. Détection et suppression des **doublons de polygones**
2. Détection et suppression des **doublons d'attributs**
3. Détection des **chevauchements > 15 %** entre polygones
4. Identification des **polygones invalides** et apporté une correction si possible
5. Comparaison des surfaces déclarées vs calculées (+/- 10 %)
6. Calcul de la **distance minimale entre chaque ferme et les forêts classée**
7. Croisé le fichier **coopératives_data.xlsx** et **parcelles.geojson** pour détectés les producteurs qui ont effectivement des parcelles
8. Synthèse des anomalies dans un fichier de suivi

Étape 3 — Automatisation

- Proposer un script **Python** (ex. via geopandas, shapely, pandas) ou un **modèle QGIS** (graphical modeler) automatisant le processus
- Documenter le fonctionnement (étapes, fonctions, structure du code)

Étape 4 — Restitution et visualisation

- Produire une **carte finale** avec les parcelles valides et celles supprimées (en couleurs distinctes)
- Générer un **rapport Excel ou tableau de bord Power BI** présentant :
 - Nombre de parcelles valides / invalides
 - Surface totale avant / après nettoyage
 - Anomalies par coopérative
 - Distribution des distances forêt–fermes

5. Livrables attendus

Élément	Format	Détails
<i>Rapport technique</i>	PDF / PPT	Présentation structurée : objectifs, méthodes, résultats, limites, recommandations
<i>Données nettoyées</i>	GeoJSON / Excel	Fichiers corrigés et cohérents
<i>Script / modèle</i>	.py ou .model3	Automatisation du pipeline
<i>Tableau de bord</i>	.pbix ou .xlsx	Synthèse visuelle des résultats
<i>Carte finale</i>	.pdf / .png	Carte thématique claire et légendée
<i>Présentation orale</i>	10–15 min	Présentation du travail réalisé

6. Éléments d'évaluation

Critère	Pondération	Détails
<i>Qualité du nettoyage et rigueur analytique</i>	30 %	Respect des consignes, pertinence du protocole
<i>Qualité technique (scripts, outils, automatisation)</i>	30 %	Code clair, réutilisable, bien documenté
<i>Visualisation et rapport final</i>	20 %	Clarté, pertinence, lisibilité
<i>Démarche méthodologique et présentation orale</i>	20 %	Cohérence, compréhension, capacité d'explication

7. Consignes de rendu

- Tous les livrables doivent être déposés dans un dossier compressé (Nom_Prenom_GeoDataAnalyst.zip)
- Dossier contenant :
 - Rapport PDF
 - Script(s) / modèles QGIS
 - Données nettoyées
 - Carte(s) et tableau(x) de bord
- La présentation orale sera organisée après analyse du dossier.