



Rapport de Stage d'ingénieur

Réalisé Par : Mohamed Haithem Iemsi

Encadrement Professionnel : Mohamed Ali Zaouali

Remerciements	3
Résumé.....	4
Introduction	5
Contexte.....	5
Objectifs	6
3.1 Objectif général	6
3.2 Objectifs spécifiques.....	6
3.3 Contraintes, périmètre et critères d'acceptation	7
3.4 Technologies pressenties.....	8
3.5 Livrables attendus.....	8
Société / Département aperçu.....	9
Brief présentation de Sofrecom	9
4.1 Présentation de Sofrecom Tunis	9
4.2 Département où a eu lieu le stage	9
Projet description.....	10
Énoncé du problème :	10
5.1 Objectifs du projet :.....	10
5.2 Portée du projet.....	11
5.3 Méthodologie Agile et organisation des sprints.....	11
5.4 Outils de gestion de projet (JIRA, GitLab)	12
Analyse et conception de systèmes	13
6 Analyse des besoins	13
6.1 Besoins fonctionnels	13
6.2 Besoins non fonctionnels.....	14
6.3 Conception du système	14
6.4 Architecture générale	14
6.5 Schéma architectural	15
6.6 Modules principaux	16
6.7 Choix technologiques	16
Implémentation et Réalisation.....	17
7.1 Développement du Backend (Flask)	17
7.2 Génération des Rapports (PDF et Excel).....	18
7.3 Développement du Frontend (Tailwind + Chart.js)	18
7.4. Intégration de l'IA (GPT-4)	19
7.5. Tests et Validation.....	19
Conclusion et Perspectives.....	20

8.1 Conclusion	20
8.2 Perspectives d'amélioration.....	21
Annexes.....	22
Annexe A – Captures d'écran de l'application	22
Annexe B – Diagramme de cas d'utilisation.....	27
Annexe C – Diagramme de classes simplifié	28

Remerciements

Je tiens à exprimer ma profonde gratitude à mon encadrant de stage, M. Chekir Ibrahim, pour son accompagnement, ses précieux conseils et son soutien tout au long de cette expérience.

Je souhaite également remercier M. Mohamed Ali Zaouali, mon manager, pour sa confiance, sa disponibilité et ses orientations constructives durant mon stage.

Mes remerciements s'adressent aussi à l'ensemble de l'équipe de Sofrecom Tunisie pour leur accueil chaleureux, leur collaboration et les connaissances qu'ils m'ont transmises.

Enfin, j'adresse mes sincères remerciements à ma famille et à mes amis pour leur encouragement et leur soutien constant tout au long de cette période.

Résumé

Ce rapport présente le travail réalisé au cours de mon stage au sein de Sofrecom Tunis. L'objectif principal de ce projet était de concevoir et développer une solution automatisée pour l'analyse des écarts de données issues de différentes applications du domaine Relation Client (DRC).

Actuellement, les données sont extraites mensuellement depuis plusieurs applications et stockées sous forme de fichiers (CSV, Excel, JSON, etc.). Le processus manuel de comparaison des données à l'aide d'Excel s'avère long et sujet aux erreurs humaines. Pour répondre à cette problématique, j'ai développé une solution intelligente et automatisée permettant de :

- Lire et analyser automatiquement les fichiers provenant d'un emplacement prédéfini ;
- Comparer les données entre plusieurs sources et détecter les écarts ;
- Générer des rapports PDF et Excel enrichis de graphiques et de statistiques ;
- Fournir une interface web moderne et intuitive pour visualiser les résultats ;
- Exploiter l'intelligence artificielle de GPT-4 pour améliorer l'analyse des données, interpréter les résultats et gérer les fichiers corrompus ou incohérents.

La solution repose sur un backend développé avec Flask (Python) et des bibliothèques telles que Pandas, NumPy et Matplotlib pour le traitement et la visualisation des données.

La partie frontend a été développée avec Chart.js pour la création de graphiques interactifs et Tailwind CSS pour un design ergonomique, responsive et agréable à utiliser.

L'intégration de GPT-4 a apporté une dimension d'intelligence supplémentaire, permettant une analyse automatique, contextuelle et explicative des écarts détectés.

L'ensemble du projet a été conçu de manière scalable, performante et maintenable, avec une architecture claire et modulaire.

Ce projet a permis d'améliorer considérablement l'efficacité, la fiabilité et la traçabilité du processus de comparaison des données, tout en réduisant les interventions manuelles et les risques d'erreurs.

Chapitre 1

Introduction

Contexte

Dans le cadre de l'amélioration continue du pilotage de la Relation Client (DRC), l'entreprise s'appuie sur plusieurs applications métier (par exemple: CRM, centre de contacts, ticketing, campagnes marketing, facturation) générant des données hétérogènes. Ces données sont essentielles pour produire des indicateurs de performance (ex. NPS, taux de résolution au premier contact, délais de réponse, volumétrie d'interactions) et pour garantir la qualité des processus opérationnels.

Actuellement, les données sont extraites mensuellement depuis ces différentes applications, puis déposées dans un emplacement prédéfini (dossier réseau ou serveur de fichiers) sous forme de fichiers (CSV, Excel, JSON, etc.). Le rapprochement et la comparaison entre ces sources sont effectués manuellement via Excel. Ce mode opératoire présente plusieurs limites:

- Temps de traitement important et non scalable lorsque le volume ou le nombre de sources augmente.
- Risque d'erreurs humaines lors des manipulations et des formules.
- Manque de traçabilité et de reproductibilité (difficulté à rejouer, vérifier ou auditer les comparaisons).
- Hétérogénéité des formats, schémas et conventions (encodages, types de champs, clés, nomenclatures) complexifiant l'alignement des données.
- Difficulté à synthétiser et visualiser les écarts de manière exploitable pour les équipes métiers.

Dans ce contexte, l'entreprise Sofrecom souhaite disposer d'une solution automatisée, fiable et maintenable pour:

- Ingestion et normalisation: lire de façon régulière et contrôlée les fichiers déposés, comprendre leurs structures, harmoniser les schémas et préparer les données.
- Comparaison multi-sources: détecter de manière robuste les écarts (données manquantes, divergences de valeur, doublons, incohérences) selon des règles métier explicites.
- Restitution et décision: produire des rapports et tableaux de bord permettant une interprétation rapide, une priorisation des corrections et un suivi dans le temps.

- Gouvernance et qualité: assurer la traçabilité des traitements, l'historisation des écarts, et la mise en place d'indicateurs de qualité des données.

Le stage s'inscrit donc dans un enjeu à la fois technique et métier: fiabiliser le patrimoine de données de la Relation Client, accélérer les cycles d'analyse, et améliorer la prise de décision en réduisant l'effort manuel et les risques d'erreurs.

Objectifs

3.1 Objectif général

Concevoir et développer une solution automatisée et scalable pour la détection et l'analyse des écarts entre les données issues de plusieurs applications du domaine DRC, intégrant des capacités de visualisation et une interface utilisateur facilitant l'exploitation des résultats.

3.2 Objectifs spécifiques

- Cadrage et analyse des sources
 - Recenser les applications et fichiers (CSV, Excel, JSON, etc.), leurs structures et dictionnaires de données.
 - Identifier les clés de rapprochement, les règles métier et les types d'écarts à détecter.
 - Définir le périmètre initial (applications, tables, champs critiques) et les KPI de qualité/écarts.
- Ingestion et préparation des données
 - Développer des scripts Python pour lire automatiquement les fichiers depuis l'emplacement prédéfini (planification, gestion des chemins/versions, logs).
 - Normaliser les schémas: harmonisation des types, encodages, formats de dates, nomenclatures; gestion des valeurs manquantes et doublons.
 - Structurer un mécanisme d'historisation des chargements et de traçabilité (journalisation).
- Moteur de comparaison et détection d'écarts
 - Implémenter une logique robuste de Matching (clés simples/composées, jointures tolérantes) et de comparaison paramétrable.

- Classer les écarts (absence, divergence de valeur, incohérence de référentiels) et calculer des indicateurs de sévérité/priorité.
 - Prévoir des règles configurables (seuils, tolérances, listes de champs à exclure) et une gestion d'erreurs explicite.
- Restitution et visualisation
 - Générer des rapports lisibles (tableaux de synthèse, statistiques, exports CSV/Excel).
 - Produire des visualisations via Matplotlib/Plotly; optionnellement, concevoir un tableau de bord Power BI/Tableau pour les besoins métiers.
- Interface utilisateur
 - Proposer une interface web légère (Flask ou équivalent) permettant:
 - de déclencher ou planifier des traitements;
 - de consulter, filtrer et exporter les écarts;
 - d'annoter/qualifier certains écarts si nécessaire.
- Industrialisation et qualité logicielle
 - Structurer le code (architecture modulaire), écrire des tests unitaires et d'intégration, et documenter l'ensemble.
 - Préparer la conteneurisation Docker (optionnel) et un guide de déploiement.
 - Mettre en place des métriques de performance (temps de traitement, volumétrie), de robustesse (taux d'échec, reprise sur incident) et de qualité (couverture des règles, faux positifs).

3.3 Contraintes, périmètre et critères d'acceptation

- Contraintes et périmètre
 - Sources: fichiers déposés mensuellement dans un répertoire/serveur partagé.
 - Formats cibles: CSV, Excel, JSON (priorité aux formats les plus utilisés).
 - Périmètre initial: [à préciser: liste des applications/tables/champs].
 - Sécurité et conformité: respect des politiques internes, gestion des accès, RGPD (pseudonymisation si nécessaire).
- Critères d'acceptation (exemples à ajuster)
 - Performance: traitement de [X] fichiers totalisant [Y] lignes en moins de [Z] minutes.

- Robustesse: reprise sur incident et traçabilité complète (logs horodatés, erreurs explicites).
- Qualité: couverture des principales règles de comparaison ($> [N]\%$), taux de faux positifs $< [p]\%$.
- Utilisabilité: interface claire, filtres, export; documentation utilisateur disponible.
- Maintenabilité: code testé, configuré par fichiers/variables d'environnement, documentation technique à jour.

3.4 Technologies pressenties

- Python (orchestration et automatisation)
- Pandas / NumPy (ingestion, transformation, comparaison)
- Matplotlib / Plotly (visualisation)
- Flask (interface web légère) ou alternative équivalente
- Power BI / Tableau (optionnel, pour les besoins de reporting avancé)
- Docker (optionnel, pour le packaging et le déploiement)

3.5 Livrables attendus

- Code source de la solution et modules de comparaison
- Documentation technique (architecture, schémas, règles de rapprochement, procédures)
- Manuel utilisateur (prise en main, cas d'usage, FAQ)
- Jeux d'essai et rapports/Dashboard de démonstration
- Guide de déploiement (et éventuellement image Docker)

Chapitre 2

Société / Département aperçu

Brief présentation de Sofrecom

4.1 Présentation de Sofrecom Tunis

Sofrecom, filiale du groupe **Orange**, est une société de conseil et d'ingénierie spécialisée dans les télécommunications et la transformation numérique. Forte de plus de 50 ans d'expérience, elle accompagne ses clients – opérateurs télécoms, gouvernements et grandes entreprises – dans la conception, le déploiement et l'optimisation de leurs solutions technologiques.

Basée à **Tunis**, **Sofrecom Tunisie** constitue l'un des pôles stratégiques du groupe, contribuant activement à des projets internationaux à forte valeur ajoutée. Le centre tunisien se distingue par ses **compétences en développement logiciel, data engineering, cloud, cybersécurité, et intelligence artificielle**, ainsi que par son **expertise dans le domaine de la Relation Client (DRC)**.

L'entreprise met un point d'honneur à promouvoir **l'innovation, la qualité et la collaboration**, tout en offrant un environnement stimulant pour les ingénieurs et les stagiaires. Sofrecom Tunis joue ainsi un rôle essentiel dans la digitalisation des services du groupe Orange et de ses partenaires à travers le monde.

4.2 Département où a eu lieu le stage

Mon stage s'est déroulé au **deuxième étage des locaux de Sofrecom Tunis**, au sein du **département DRC (Domaine Relation Client)**. Cette équipe, dirigée par **M. Mohamed Ali Zaouali**, est en **contact direct avec le client** et joue un rôle essentiel dans la conception, le développement et le maintien des solutions liées à la relation client pour les filiales du groupe Orange.

J'ai été encadré par **M. Chekir Ibrahim**, membre de cette même équipe, qui m'a accompagné tout au long du stage en me guidant sur les aspects techniques et méthodologiques du projet.

Le département se distingue par une forte **collaboration entre ses membres**, combinant expertise technique et compréhension des besoins métiers. Cette expérience m'a permis de découvrir un environnement professionnel dynamique, où l'innovation et la qualité du service client sont au cœur des priorités.

Chapitre 3

Projet description

Énoncé du problème :

Au sein du domaine Relation Client (DRC) de Sofrecom Tunis, le processus de comparaison et d'analyse des données issues de différentes applications était auparavant réalisé de manière manuelle, principalement à l'aide de fichiers Excel. Cette approche, bien que fonctionnelle, présentait plusieurs limitations : elle était longue, fastidieuse et sujette aux erreurs humaines.

L'absence d'un système centralisé et automatisé rendait difficile la détection rapide des écarts entre les jeux de données, ainsi que la génération de rapports clairs et fiables. De plus, la manipulation manuelle des fichiers augmentait les risques d'incohérences et de pertes d'informations.

Face à ces constats, il est apparu nécessaire de développer une solution automatisée et intelligente capable de comparer, analyser et visualiser les écarts de manière efficace. Cette solution devait permettre de réduire le temps de traitement, améliorer la fiabilité des résultats, et offrir une interface intuitive pour faciliter la prise de décision et le suivi des analyses.

5.1 Objectifs du projet :

L'objectif principal de ce projet était de **concevoir et développer une application web entièrement automatisée** permettant de **déetecter, analyser et visualiser les écarts de données** entre plusieurs applications du domaine **Relation Client (DRC)**.

Les objectifs spécifiques du projet étaient les suivants :

- **Automatiser** le processus de lecture, de comparaison et d'analyse des fichiers issus de différentes sources (CSV, Excel, JSON, etc.) ;
- **DéTECTER et signaler les écarts** entre les jeux de données de manière précise et rapide ;
- **Générer des rapports dynamiques** sous forme de fichiers **PDF et Excel**, accompagnés de **graphiques interactifs** réalisés avec **Chart.js** ;

- **Mettre en place une interface web intuitive et ergonomique** grâce à **Tailwind CSS**, facilitant la visualisation et la compréhension des résultats par les utilisateurs ;
- **Intégrer l'intelligence artificielle de GPT-4** afin d'améliorer l'analyse, interpréter les écarts détectés et traiter les fichiers corrompus ou incohérents ;
- **Réduire le temps de traitement et les erreurs humaines**, tout en augmentant la fiabilité et la traçabilité des résultats ;
- **Garantir une architecture évolutive et maintenable**, permettant d'étendre la solution à d'autres domaines ou types de données à l'avenir.

5.2 Portée du projet

La portée de ce projet englobe l'ensemble des étapes nécessaires à la **conception, au développement et à la mise en place d'une solution automatisée** pour l'analyse des écarts de données entre plusieurs applications du domaine **Relation Client (DRC)**.

Le périmètre du projet inclut :

- **Le développement du frontend et du backend** de l'application web ;
- **La mise en place de processus d'automatisation** pour la lecture, la comparaison et l'analyse des fichiers.
- **L'intégration d'un module d'intelligence artificielle (GPT-4)** pour renforcer l'analyse, la compréhension et la gestion des données corrompues.
- **La visualisation des résultats** à travers des **graphiques interactifs** créés avec **Chart.js** et une interface ergonomique développée avec **Tailwind CSS**.
- **La génération automatique de rapports PDF et Excel** détaillant les écarts détectés ;
- **Le déploiement, les tests et la validation** de la solution dans l'environnement interne de Sofrecom.

Le projet se concentre exclusivement sur **l'automatisation interne du processus d'analyse des écarts de données** au sein du domaine DRC et **n'englobe pas les systèmes externes** en dehors de cet environnement.

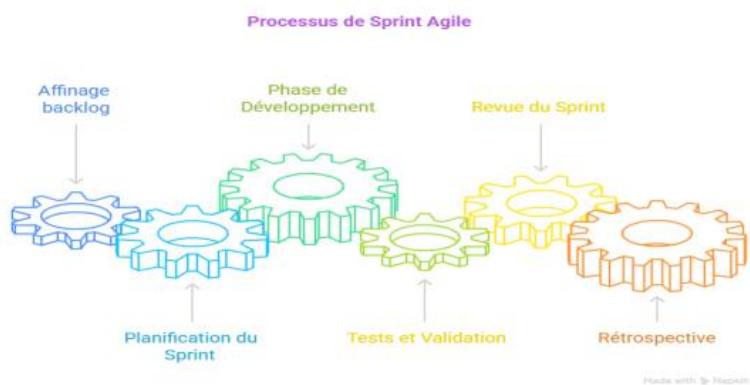
Cependant, la solution a été conçue de manière **modulaire et évolutive**, afin de pouvoir être **étendue et améliorée** pour d'autres domaines ou besoins futurs.

5.3 Méthodologie Agile et organisation des sprints

Sofrecom travaille selon la méthodologie Agile Scrum, qui structure le travail en sprints de trois semaines. Cette approche vise à fournir des livrables réguliers, testés, validés et

directement exploitables. Elle encourage aussi la communication continue, la transparence et l'adaptation rapide aux besoins changeants. Le sprint se déroule en plusieurs étapes clés :

- Affinage
- Sprint Planning
- Phase de développement
- Tests et validation
- Sprint Review & Rétrospective



5.4 Outils de gestion de projet (JIRA, GitLab)

L'environnement de travail agile s'appuie sur une boîte à outils cohérente et intégrée, facilitant la gestion des tâches, la collaboration technique, ainsi que la validation des développements. Les trois outils principaux sont :

- JIRA : utilisé pour la gestion des tickets, le suivi de l'avancement des sprints, et l'organisation du backlog. Chaque tâche du sprint est détaillée sous forme de ticket, auquel sont associées des checklists, des commentaires, et des statuts évolutifs. JIRA permet également la planification des travaux à l'aide de la roadmap, offrant ainsi une visibilité sur les jalons et les priorités à moyen terme. Il est essentiel pour assurer la traçabilité du travail et la répartition des responsabilités
- GitLab : plateforme de gestion du code source. Elle permet à chaque développeur de créer des branches spécifiques, de versionner son code, et de soumettre des Merge Requests pour relecture et validation. GitLab facilite le travail collaboratif et renforce la qualité du code produit grâce à des revues croisées systématiques.

Chapitre 4

Analyse et conception de systèmes

6 Analyse des besoins

Le système a été conçu dans le but d'automatiser et d'optimiser le processus de détection et d'analyse des écarts de données entre plusieurs applications du domaine Relation Client (DRC) au sein de Sofrecom Tunis.

Cette automatisation vise à réduire les interventions manuelles, les erreurs humaines et le temps de traitement, tout en offrant une interface intuitive pour la visualisation et l'interprétation des résultats.

6.1. Besoins fonctionnels

Les besoins fonctionnels identifiés pour ce projet sont les suivants :

- **Lecture automatique des fichiers** provenant d'un emplacement défini (CSV, Excel, JSON, etc.) .
- **Analyse et comparaison des données** entre plusieurs sources afin de détecter les écarts.
- **Détection intelligente des anomalies** grâce à l'intégration du modèle **GPT-4**, capable d'interpréter les différences et de proposer des explications ou recommandations.
- **Génération de rapports automatiques** au format **PDF** et **Excel**, contenant des tableaux, graphiques et statistiques de comparaison .
- **Visualisation des écarts détectés** à travers des **graphiques interactifs** réalisés avec **Chart.js**.
- **Interface web ergonomique** développée avec **Tailwind CSS**, permettant à l'utilisateur de naviguer, d'analyser et d'exporter facilement les résultats .
- **Historisation et suivi des analyses précédentes**, afin de conserver une traçabilité complète des traitements effectués.
- **Possibilité d'évolution** du système pour intégrer de nouvelles sources de données ou d'autres domaines applicatifs à l'avenir.

6.2 Besoins non fonctionnels

En plus des aspects fonctionnels, certains critères de qualité et de performance ont été définis :

- Le système doit être **sécurisé**, en protégeant l'accès aux fichiers et aux résultats d'analyse.
- Il doit être **rapide et performant**, même avec des volumes importants de données .
- L'application doit être **accessible via une interface web** simple et responsive.
- Les traitements doivent être **entièrement automatisés**, sans intervention manuelle.
- Le système doit être **fiable et maintenable**, avec une architecture claire et modulaire.
- L'interface doit être **conviviale et intuitive**, afin de faciliter la compréhension et l'interprétation des résultats.
- Le code doit respecter les **bonnes pratiques de développement** pour garantir la qualité et la pérennité du projet.

6.3 Conception du système

La conception du système vise à proposer une architecture claire, modulaire et évolutive permettant d'automatiser efficacement le processus de **comparaison, d'analyse et de génération de rapports** sur les écarts de données.

Le système repose sur une approche **Full Stack** combinant un **backend Flask**, une **interface web interactive** et des **composants d'intelligence artificielle**.

6.4 Architecture générale

L'architecture du système est basée sur un modèle **3-tiers** :

1. Couche de présentation (Frontend)

- Développée en **HTML, CSS (Tailwind) et JavaScript**.
- Offre une interface utilisateur simple, responsive et interactive.
- Permet de téléverser les fichiers, lancer les comparaisons et visualiser les résultats sous forme de tableaux et de graphiques dynamiques (Chart.js).

2. Couche logique (Backend Flask)

- Assure la **gestion des requêtes, la comparaison des fichiers, la génération de rapports et la communication avec le modèle d'IA**.

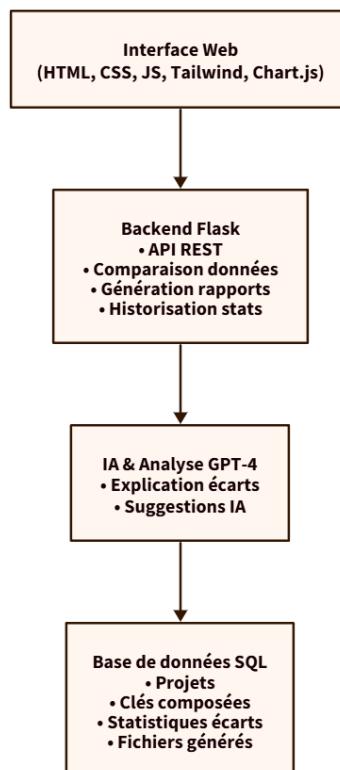
- Les principales fonctionnalités incluent :
 - Lecture et traitement des fichiers (pandas)
 - Comparaison automatique des données selon des clés composées
 - Calcul des statistiques (écart, correspondances, taux d'erreurs)
 - Génération de rapports **PDF** et **Excel** stylisés (avec logo et graphiques)
 - Historisation des exécutions et stockage des statistiques dans une base de données

3. Couche d'intelligence artificielle (GPT Intégration)

- Intègre un modèle **GPT-4** pour analyser les différences détectées.
- Fournit une **analyse contextuelle des écarts**, des **suggestions d'amélioration** et des **explications automatisées** dans les rapports.

6.5 Schéma architectural

Le schéma global du système peut être représenté ainsi :



6.6 Modules principaux

Le système est organisé en plusieurs **modules indépendants** :

- **Module Importation & Lecture** : permet à l'utilisateur de téléverser des fichiers CSV, Excel ou JSON.
- **Module de comparaison** : applique des clés de correspondance composées et détecte les écarts ligne par ligne.
- **Module d'analyse IA** : exploite GPT-4 pour interpréter les écarts détectés et générer des commentaires pertinents.
- **Module de génération de rapports** : crée des fichiers PDF et Excel enrichis de graphiques, statistiques et logos institutionnels.
- **Module de gestion des projets** : stocke les configurations et les résultats d'analyse dans une base de données relationnelle.
- **Module de visualisation** : affiche les statistiques sous forme de tableaux et de graphiques interactifs sur le tableau de bord.

6.7 Choix technologiques

Composant	Technologie	Rôle principal
Frontend	HTML5, Tailwind CSS, Chart.js	Interface utilisateur et visualisation des données
Backend	Flask (Python)	API, logique métier, génération de rapports
IA	GPT-4 (OpenAI API)	Analyse sémantique et suggestions automatisées
Base de données	SQLite / MySQL	Stockage des configurations et statistiques
Formats de rapports	PDF (ReportLab), Excel (openpyxl)	Génération et export de rapports
Librairies principales	pandas, pypandoc, matplotlib	Traitement et visualisation des données

Chapitre 5

Implémentation et Réalisation

L'implémentation du projet a consisté à transformer la conception théorique en une application web fonctionnelle, performante et ergonomique.

Cette phase s'est déroulée sur plusieurs étapes : le développement du **backend Flask**, la création d'une **interface utilisateur moderne avec Tailwind CSS**, la mise en place de **graphiques interactifs** via Chart.js, et l'intégration d'un module **d'analyse intelligente GPT-4**.

7.1 Développement du Backend (Flask)

Le cœur du système repose sur le Framework **Flask (Python)**, choisi pour sa légèreté et sa flexibilité.

Il a permis de construire une **API RESTful** capable de gérer les différentes opérations nécessaires à la comparaison et à l'analyse des fichiers.

Les principales fonctionnalités implémentées :

- **Importation et lecture des fichiers :**

Utilisation de la bibliothèque **pandas** pour lire et traiter différents formats de fichiers (CSV, Excel, JSON).

Le système détecte automatiquement le type de fichier et extrait les colonnes nécessaires à la comparaison.

- **Comparaison automatique :**

Implémentation d'un algorithme qui combine plusieurs colonnes pour former une **clé composée**.

Cette clé permet d'effectuer une comparaison précise entre les deux jeux de données et de repérer :

- Les correspondances exactes
- Les lignes manquantes ou ajoutées
- Les différences de valeurs

- **Calcul des statistiques :**

Après chaque comparaison, le système génère automatiquement :

- Le **nombre total d'écart**

- Le **taux de correspondance**
- Le **nombre de lignes analysées**
- Le **temps d'exécution**
- **Sauvegarde des résultats :**
Les résultats et statistiques sont stockés dans une **base de données relationnelle** pour un suivi et un historique des analyses.

7.2 Génération des Rapports (PDF et Excel)

Le système offre la possibilité de générer automatiquement des rapports professionnels en **PDF** et **Excel** :

- Les rapports **PDF** sont produits à l'aide de **pdfkit** et **matplotlib**, intégrant :
 - Le **logo de Sofrecom**
 - Les **statistiques clés**
 - Les **graphiques de performance**
 - Les **analyses générées par GPT-4**
- Les rapports **Excel** sont générés via **openpyxl**, permettant :
 - Un style clair avec couleurs alternées
 - L'ajout de graphiques intégrés
 - La mise en page dynamique et lisible

Ces rapports sont automatiquement horodatés et classés par projet dans le tableau de bord.

7.3 Développement du Frontend (Tailwind + Chart.js)

L'interface utilisateur a été développée en **HTML, JavaScript et Tailwind CSS** pour garantir un design moderne, épuré et responsive.

Elle permet à l'utilisateur de :

- Téléverser les fichiers sources à comparer
- Sélectionner les clés de comparaison

- Lancer le processus de comparaison
- Visualiser les résultats sous forme de **graphiques interactifs (Chart.js)**
- Télécharger les rapports PDF et Excel

L'utilisation de **Tailwind CSS** a permis de gagner en productivité et d'obtenir un **affichage ergonomique** adapté aux besoins internes de Sofrecom.

7.4. Intégration de l'IA (GPT-4)

Une partie innovante du projet réside dans l'intégration du modèle **GPT-4**, utilisé pour enrichir les rapports avec des analyses contextuelles.

Après la comparaison, le système envoie un résumé des écarts au modèle GPT-4, qui fournit :

- Une **interprétation sémantique** des différences détectées
- Des **suggestions d'actions correctives**
- Une **analyse prédictive** sur les causes possibles des incohérences
- Une **lecture automatisée des fichiers corrompus** lorsque c'est possible

Cette intégration permet d'apporter une **valeur ajoutée intelligente** au système, rendant les rapports plus pertinents et exploitables.

7.5. Tests et Validation

Des tests unitaires et fonctionnels ont été réalisés pour assurer la fiabilité du système.

Les principaux cas testés incluent :

- Compatibilité multiformat (CSV, Excel, JSON)
- Exactitude des comparaisons et des clés composées
- Génération correcte des rapports PDF et Excel
- Intégration stable avec GPT-4
- Performance et temps de réponse du backend Flask

Tous les tests ont donné des résultats satisfaisants, démontrant la robustesse et la fiabilité du système.

Chapitre 6

Conclusion et Perspectives

8.1 Conclusion

Ce stage au sein de **Sofrecom Tunis**, intégré à l'équipe **DRC (Domaine Relation Client)**, m'a permis de mettre en pratique mes compétences en développement web, traitement de données et intégration de l'intelligence artificielle.

L'objectif du projet est de **concevoir et développer une solution automatisée d'analyse des écarts de données** a été pleinement atteint.

La solution **DataAlign** que j'ai développée permet désormais :

- D'automatiser la lecture et la comparaison de fichiers (CSV, Excel, JSON) .
- De générer des rapports PDF et Excel professionnels, incluant des graphiques et statistiques.
- D'afficher les résultats dans une interface web moderne et ergonomique (Tailwind CSS + Chart.js).
- D'intégrer **GPT-4** pour enrichir l'analyse, interpréter les écarts et corriger les incohérences.

Au-delà de l'aspect technique, ce projet m'a permis de découvrir le **fonctionnement d'une équipe Agile**, la rigueur des outils de gestion (JIRA, GitLab), et l'importance de la communication dans un environnement professionnel collaboratif.

J'ai pu renforcer mes compétences en **Python, Flask, visualisation de données, UX/UI**, ainsi qu'en **intégration d'API d'intelligence artificielle**.

Ce stage a également constitué une **expérience humaine enrichissante**, m'offrant la possibilité de collaborer avec des ingénieurs expérimentés et de comprendre les enjeux réels du domaine Relation Client au sein d'une entreprise internationale comme Sofrecom.

8.2 Perspectives d'amélioration

Plusieurs axes d'évolution peuvent être envisagés pour la suite du projet **DataAlign** :

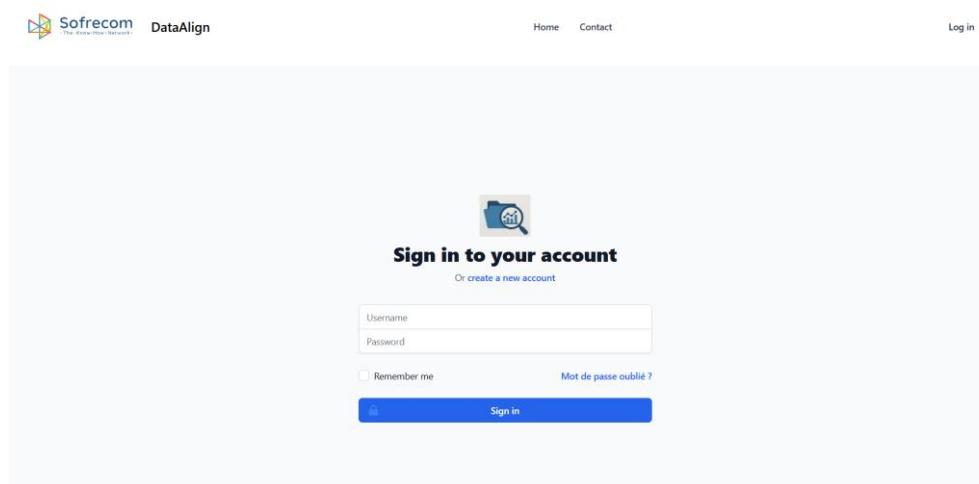
- **Authentification et rôles utilisateurs** : ajouter un module de connexion et de gestion des accès pour sécuriser les traitements.
- **Intégration temps réel** : connecter l'application directement aux bases de données ou APIs des applications DRC pour une analyse continue.
- **Optimisation des performances** : paralléliser certains traitements et améliorer la gestion mémoire pour supporter des volumes de données plus importants.
- **Tableau de bord avancé** : intégrer des visualisations plus riches (ex. Power BI, Grafana) et un suivi historique complet.
- **Apprentissage automatique** : entraîner un modèle ML pour prédire les écarts récurrents et recommander des actions correctives automatiques.
- **Déploiement Docker et CI/CD** : conteneuriser entièrement la solution pour un déploiement rapide et reproductible sur différents environnements.

En somme, le projet DataAlign constitue une **base solide et extensible** pour une plateforme complète de gouvernance et de qualité de données au sein du domaine DRC.

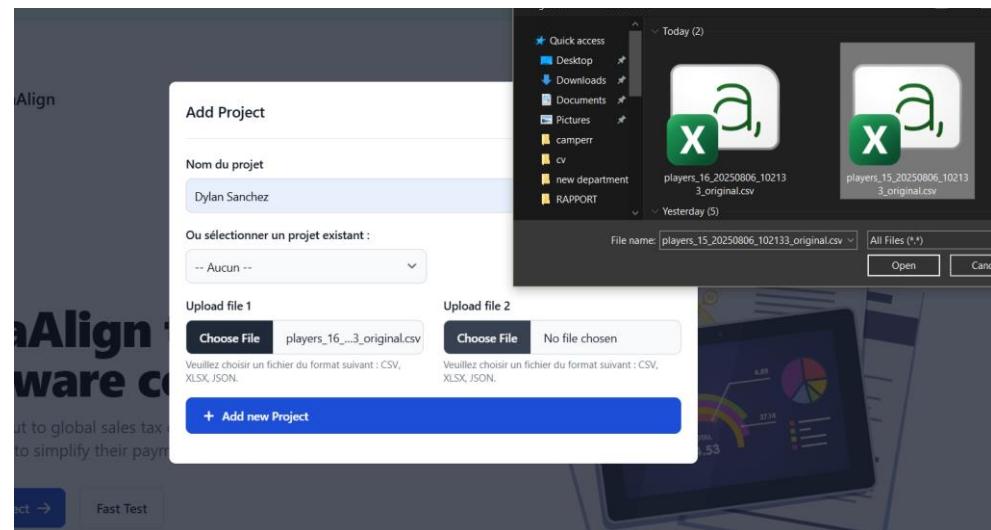
Annexes

Annexe A – Captures d'écran de l'application

1. Page d'accueil de DataAlign



2. Formulaire d'importation des fichiers



3. Tableau comparatif des données

Choisissez les clés pour comparer

Clé dans le fichier 1:

sofifa_id
player_url
short_name
long_name

Vous pouvez sélectionner plusieurs colonnes avec Ctrl / Cmd + clic.

Clé dans le fichier 2:

sofifa_id
player_url
short_name
long_name

Vous pouvez sélectionner plusieurs colonnes avec Ctrl / Cmd + clic.

Comparer

Large Files Detected! One or both of your files are large (File 1: 14,881 rows, 104 columns) (File 2: 15,465 rows, 104 columns). The system will use optimized processing to handle them efficiently. Only a preview is shown below, but the full comparison will process all data.

Données du premier fichier

Voici un aperçu des données extraites automatiquement.

Clé dans le fichier 1:

sofifa_id
player_url
short_name
long_name

Vous pouvez sélectionner plusieurs colonnes avec Ctrl / Cmd + clic.

Clé dans le fichier 2:

sofifa_id
player_url
short_name
long_name

Comparing files and generating results...

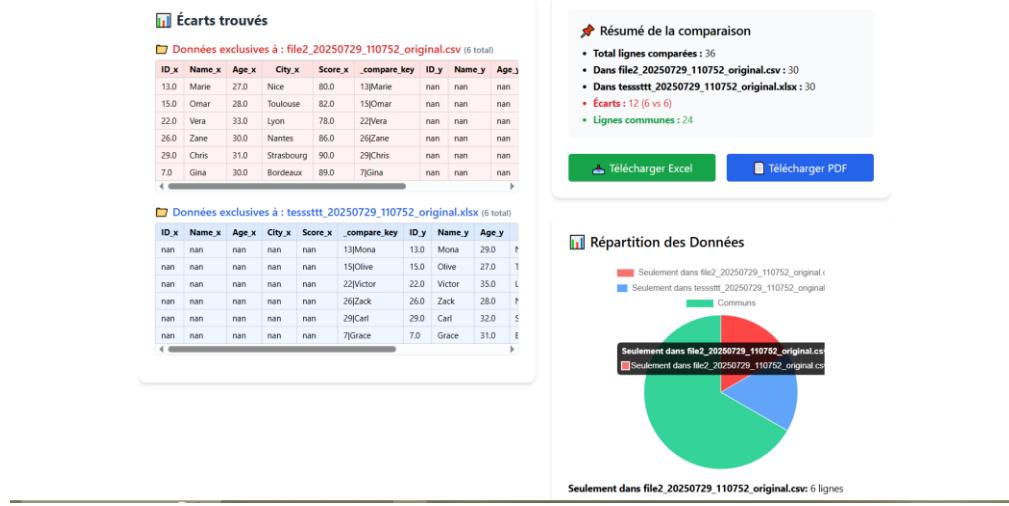
Comparer

Données du deuxième fichier

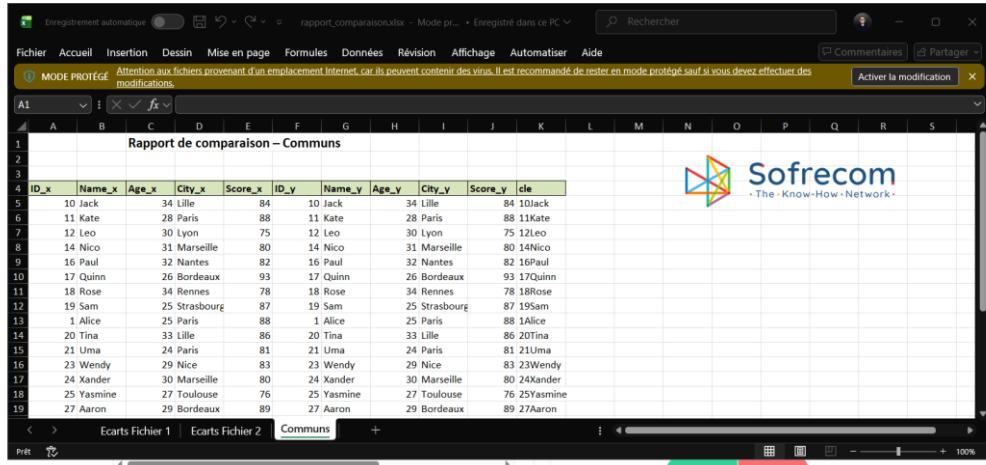
Voici un aperçu des données extraites automatiquement.

Detected! One or both of your files are large (File 1: 14,881 rows, 104 columns) (File 2: 15,465 rows, 104 columns). The system will use optimized processing to handle them efficiently. Only a preview is shown below, but the full comparison will process all data.

4. Graphiques interactifs (Chart.js)



5. Rapport PDF généré automatiquement



Rapport de comparaison des fichiers

Lignes uniquement dans file2_20250729_110752_original.csv (6):

ID_x	Name_x	Age_x	City_x	Score_x	_compare_key	ID_y	Name_y	Age_y	City_y	Score_y	cle	_merge
13.0	Marie	27.0	Nice	80.0	13 Marie	nan	nan	nan	nan	nan	nan	left_only
15.0	Omar	28.0	Toulouse	82.0	15 Omar	nan	nan	nan	nan	nan	nan	left_only
22.0	Vera	33.0	Lyon	78.0	22 Vera	nan	nan	nan	nan	nan	nan	left_only
26.0	Zane	30.0	Nantes	86.0	26 Zane	nan	nan	nan	nan	nan	nan	left_only
29.0	Chris	31.0	Strasbourg	90.0	29 Chris	nan	nan	nan	nan	nan	nan	left_only
7.0	Gina	30.0	Bordeaux	89.0	7 Gina	nan	nan	nan	nan	nan	nan	left_only

Lignes uniquement dans tesssttt_20250729_110752_original.xlsx (6):

ID_x	Name_x	Age_x	City_x	Score_x	_compare_key	ID_y	Name_y	Age_y	City_y	Score_y	cle	_merge
nan	nan	nan	nan	nan	13 Mona	13.0	Mona	29.0	Nice	89.0	13 Mona	right_only
nan	nan	nan	nan	nan	15 Olive	15.0	Olive	27.0	Toulouse	77.0	15 Olive	right_only
nan	nan	nan	nan	nan	22 Victor	22.0	Victor	35.0	Lyon	74.0	22 Victor	right_only
nan	nan	nan	nan	nan	26 Zack	26.0	Zack	28.0	Nantes	85.0	26 Zack	right_only
nan	nan	nan	nan	nan	29 Carl	29.0	Carl	32.0	Strasbourg	92.0	29 Carl	right_only
nan	nan	nan	nan	nan	7 Grace	7.0	Grace	31.0	Bordeaux	90.0	7 Grace	right_only

Résumé de la comparaison :

Total lignes dans file2_20250729_110752_original.csv : 30

Total lignes dans tesssttt_20250729_110752_original.xlsx : 30

Lignes communes : 24

Lignes uniquement dans file2_20250729_110752_original.csv : 6

Lignes uniquement dans tesssttt_20250729_110752_original.xlsx : 6



6. Interface des traitements

The screenshot displays two windows of the DataAlign software.

Arborescence des Projets (Project Tree) Window:

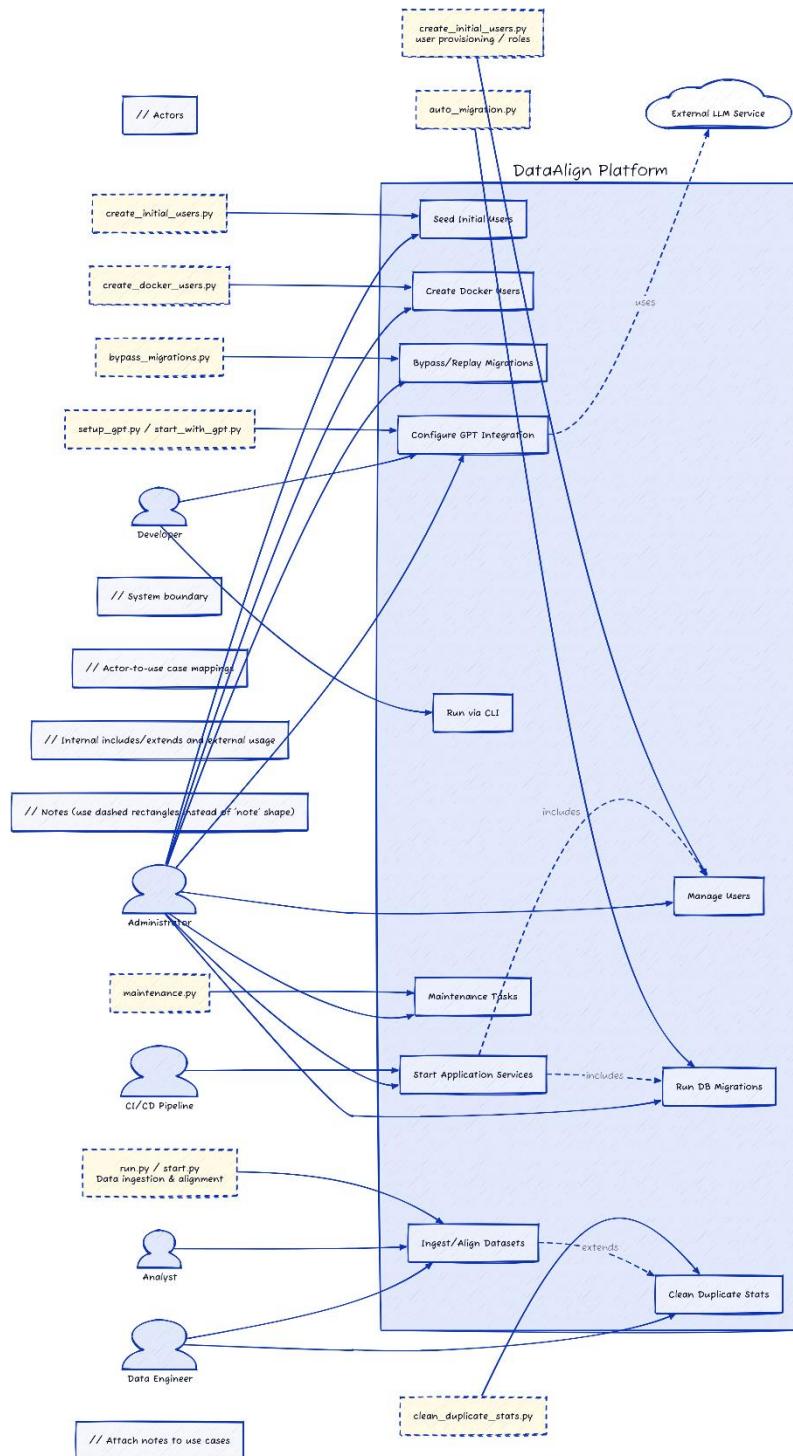
- Header: Sofrecom - The Know-Help-Network, DataAlign, Home, Traitement (highlighted), Notifications, Contact, User icon.
- Section: Dylan Sanchez, 0 traitement(s).
- Section: test, 1 traitement(s).
- Details for the 'test' project:
 - Date: 31/10/2025 à 19:55:40
 - Fichiers traités:
 - Fichier 1: file2_20250729_110752_original_20251031_195532_original.csv
 - Fichier 2: tesssttt_20250729_110752_original_20251031_195532_original.xlsx
 - ID: 1
- Buttons: Details, ZIP, Graphique, Demander.

Traitement (Treatment) Window:

- Title: Traitement:
- Informations du traitement:
 - Nom du projet: 2025-10-31 19:55:39
 - Date d'exécution: 2025-10-31 19:55:39
 - ID du traitement: 1
 - ID du projet: 4
- Fichiers traités:
 - Fichier 1: file2_20250729_110752_original_20251031_195532_original.csv
 - Fichier 2: tesssttt_20250729_110752_original_20251031_195532_original.xlsx
- Emplacement d'archive: uploads/archive/test_20251031_195532
- Statistiques de comparaison:
 - Statistiques spécifiques à ce traitement (31/10/2025 à 19:55:40)
 - Uniquement fichier 1: 6
 - Uniquement fichier 2: 6
 - Lignes communes: 24
 - Total écarts: 12
 - Date d'exécution: 2025-10-31 19:55:39
- Fichiers disponibles:
 - Excel (green checkmark)
 - PDF (green checkmark)
 - Graphique (green checkmark)

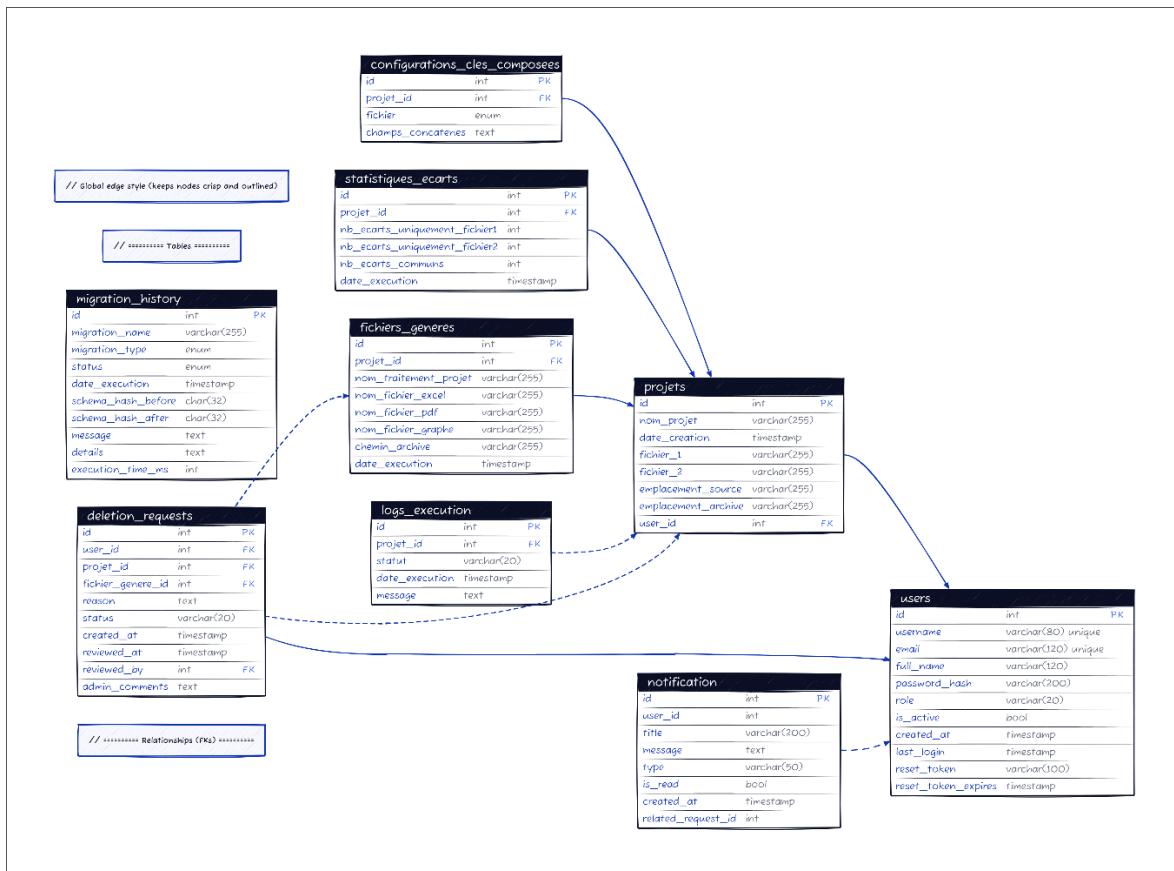
Annexe B – Diagramme de cas d'utilisation

Voici le diagramme UML de cas d'utilisation proposé pour DataAlign



Annexe C – Diagramme de classes simplifié

Voici un diagramme de classes simplifié pour représenter la structure principale de l'application :



Ce diagramme illustre les liens logiques entre les entités :

- Un Projet contient plusieurs Comparaisons.
- Chaque Comparaison peut générer un ou plusieurs Rapports (traitements).
- Une Analyse IA (GPT-4) est associée à une comparaison pour enrichir l'interprétation.



ESPRIT SCHOOL OF ENGINEERING

www.esprit.tn - E-mail : contact@esprit.tn

Siège Social : 18 rue de l'Usine - Charguia II - 2035 - Tél. : +216 71 941 541 - Fax. : +216 71 941 889

Annexe : 1-2 rue André Ampère - 2083 - Pôle Technologique - El Ghazala - Tél +216 70 250 000 - Fax +216 70 685454