**README - Gestion des Données de Patients avec MongoDB et Python**

Description

Ce projet permet l'importation, le nettoyage, l'insertion et la validation de données de patients dans une base de données MongoDB à partir d'un fichier CSV. Il comprend un service MongoDB et un service Python pour exécuter les scripts d'analyse et de test.

Structure du Projet

├── docker-compose.yml

├── Dockerfile

├── python/

│ ├── healthVisual.py

│ ├── testVisual.py

├── csv/

│ ├── healthcare\_dataset\_nettoye.csv

├── requirements.txt

└── README.md

Prérequis

* Docker et Docker Compose
* Python 3.10+
* MongoDB

Installation

1. Cloner le dépôt :
2. git clone <URL\_REPO>

cd <NOM\_DU\_PROJET>

1. Installer les dépendances Python :

pip install -r requirements.txt

Exécution avec Docker

1. Démarrer les services :

docker-compose up --build

1. Les scripts Python s'exécutent automatiquement et procèdent à :
   * Chargement et nettoyage des données CSV
   * Insertion dans MongoDB
   * Validation de l'importation avec des tests unitaires
2. Arrêter les services :

docker-compose down

Scripts Principaux

healthVisual.py

* Charge les données depuis le fichier CSV
* Effectue le nettoyage des données
* Convertit certains champs en formats compatibles avec MongoDB
* Insère les données dans la collection patients
* Affiche des statistiques sur l'insertion

testVisual.py

* Teste la correspondance entre les données du fichier CSV et MongoDB
* Vérifie :
  + L'égalité du nombre de lignes
  + La cohérence du nombre de colonnes
* Effectue des tests unitaires avec unittest

MongoDB - Opérations CRUD

* Create : Insérer un document
* Read : Récupérer des données
* Update : Modifier un document
* Delete : Supprimer un document

Variables d'Environnement

Les identifiants de connexion MongoDB sont définis dans docker-compose.yml :

environment:

- MONGO\_INITDB\_ROOT\_USERNAME=admin

- MONGO\_INITDB\_ROOT\_PASSWORD=admin#75\*Db

Améliorations Possibles

* Ajouter un fichier de configuration .env
* Implémenter une interface web pour visualiser les données
* Améliorer la gestion des erreurs et des logs