

# Projet Module 1 Intro au Big Data

**Contexte:** L'objectif ici est d'extraire des insights (informations importantes agrégées) des données sources tout en appliquant les bonnes pratiques qu'on a vu le long du cours et des TPs. Ces insights doivent être sauvegardés (ici au niveau du stockage objet).

## Données:

La source de données est ici <https://datahub.io/core/inflation>, il s'agit de données de plusieurs années sur différents pays sur l'inflation (en %)

On a 4 colonnes dans le dataset qui sont:

*Country*: le pays en question (string)

*Country Code*: Trigramme du pays (string)

*Year*: l'année (int)

*Inflation*: Le taux de l'inflation (float)

## Consignes Technique:

Voilà à quoi doit ressembler votre rendu



- *\_notebooks/project.ipynb*: Le dossier *\_notebooks* contiendra votre ou vos notebook(s) dans le(s)quel(s) se feront le processing, la lecture et l'écriture de données (**à faire**)
- *data/*: ici on a les données source (qu'on va insérer dans la base de données)
- *docker/app/utils/insert\_data.py*: le script python qui fera l'insertion des données dans postgres (**à compléter**)
- *docker/docker-compose.yaml*: ici vous définissez vos conteneurs (un pour la base postgres, un autre pour l'app python qui fait l'insertion des données, un pour jupyter notebook et un autre pour le stockage objet) - **à faire**
- *project.ipynb*: votre fichier notebook pour faire le processing (**à faire**)
- *requirements.txt*: pour les requirements (**à compléter**)
- *docker/\*/Dockerfile*: pour faire les images des différents container dont vous allez utiliser (**à faire**)

### Indications:

- Cloner le repos et se mettre dans le dossier *Project*
- Préparer votre environnement (virtuelle), Compléter vos fichier de requirements.txt
- Compléter le script d'insertion (*utils/insert\_data.py*)
- Compléter les différents Dockerfile
- Compléter le docker-compose.yaml
- Lancer vos conteneurs (docker-compose up)
- Vérifier que tout se passe bien en vérifiant les différents containers
- Ouvrez jupyter notebook et connecter vous à votre source et commencer à travailler dessus
- Toutes les écritures doivent se faire sur le stockage objet
- Vous pouvez prendre l'exemple du TP2

**Deadline:**

Lundi 14/07/2025 à 23h59

**Rendu:**

Un dossier comme ci haut avec tous les composants à envoyer par mail ([idrissamagassa@gmail.com](mailto:idrissamagassa@gmail.com)) ou partager un lien github

Le notebook doit expliquer votre démarche