

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

```
project
├── Dockerfile
├── data
│   └── inflation_data.csv
├── docker-compose.yaml
├── project.ipynb
├── requirements.txt
└── utils
    ├── __init__.py
    └── insert_data.py
```

- *data/*: ici on a les données source (qu'on va insérer dans la base de données)
- *utils/insert_data.py*: le script python qui fera l'insertion des données dans postgres (**à compléter**)
- *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 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**)
- *Dockerfile*: contient l'image docker de l'app python qui fait l'insertion (déjà fait rien à faire)

Indications:

- Cloner le repos et se mettre dans le dossier *project*
- Préparer votre environnement (virtuelle), Compléter votre requirements.txt
- Compléter le script d'insertion (*utils/insert_data.py*)
- Compléter votre docker-compose.yaml
- Lancer vos conteneurs (docker-compose up)
- Vérifier que tout se passe bien
- Lancer votre notebook et connecter vous à votre source et commencer à travailler dessus
- Toutes les écritures doivent se faire sur le stockage objet

Deadline:

Lundi 19/02/2025 à 23h59

Rendu:

Un dossier comme ci haut avec tous les composants à envoyer par mail (idrissamagassa@gmail.com)

Le notebook doit expliquer votre démarche