

ETAPE A SUIVRE POUR L'API

I. Installation de WSL (Windows Subsystem for Linux)

Installer WSL :

Ouvrez PowerShell en tant qu'administrateur.

Exécutez la commande suivante pour installer WSL

```
=> wsl --install
```

Configurer une Distribution Linux :

Une fois l'installation terminée, ouvrez votre distribution Linux (Ubuntu) et effectuez les mises à jour :

```
=> sudo apt update && sudo apt upgrade -y
```

II. Configuration de l'Environnement de Développement

Installer les Outils de Développement :

Installez les outils nécessaires :

```
=> sudo apt install build-essential curl git python3 python3-pip -y
```

Configurer un Environnement Virtuel Python :

Créez et activez un environnement virtuel :

```
=> python3 -m venv env
```

```
=> source env/bin/activate
```

Installer les Dépendances du Projet :

Installez les bibliothèques requises :

```
=> pip install requests pandas matplotlib seaborn apache-airflow dash
```

III. Récupération des Données de Pollution de l'Air

Récupérer les Données :

Développez un script pour récupérer les données de pollution de l'air à partir d'une API gratuite comme OpenWeatherMap.

Enregistrer les Données :

Enregistrez les données récupérées dans un fichier JSON dans le dossier source d'Airflow pour une intégration facile avec Power BI.

IV. Collecte et Stockage des Données

Développer le Programme de Collecte :

Créez un script ou une tâche Airflow pour automatiser la collecte des données.

Stocker les Données :

Organisez et stockez les données collectées dans un format structuré, comme un fichier JSON ou une base de données.

V. Nettoyage et Transformation des Données

Nettoyer les Données :

Vérifiez la cohérence et la complétude des données, et effectuez des corrections

si nécessaire.

Transformer les Données :

Préparez les données pour l'analyse en effectuant des opérations telles que l'agrégation et la normalisation.

VI. Analyse des Données

Explorer les Corrélations :

Analysez les relations entre les niveaux de pollution et les facteurs géographiques et démographiques.

Identifier les Tendances :

Recherchez et identifiez des tendances significatives dans les données.

VII. Automatisation avec Apache Airflow

Créer un DAG Airflow :

Développez un DAG pour automatiser le processus d'extraction, de transformation et de chargement des données (ETL).

Configurer les Tâches :

Définissez les tâches dans Airflow pour la collecte des données, le nettoyage et la transformation.

Planifier les Exécutions :

Configurez la planification des tâches pour mettre à jour régulièrement les données et analyses.