

Projet : Analyse de la Pollution de l'Air

Contexte :

La pollution de l'air est un problème majeur affectant la santé publique et l'environnement. Ce projet vise à analyser les niveaux de pollution de l'air dans plusieurs régions, en examinant les corrélations possibles avec des facteurs géographiques et démographiques.

Objectif :

Votre mission est de récupérer des données de pollution de l'air à partir d'une API gratuite, de les analyser en les combinant avec des données démographiques et géographiques fournies, et de présenter les résultats sous la forme d'un tableau de bord interactif.

Tâches à Réaliser :

1. Récupération des Données de Pollution de l'Air :

- Utilisez l'API [OpenWeatherMap](#) ou toute autre API gratuite pour collecter les données de pollution de l'air, y compris les indices de qualité de l'air (AQI) et les concentrations de divers polluants (PM2.5, PM10, O3, NO2, SO2, CO).
- Vous êtes libre de récupérer d'autres types de données (météorologiques, etc.) si vous estimez qu'elles sont nécessaires pour enrichir votre analyse.

2. Données Fournies :

- Données Démographiques et Géographiques : Ces données vous sont fournies et incluent des informations telles que la population, la densité de population, le taux d'urbanisation, les revenus moyens, le niveau d'éducation, l'altitude, et la proximité des sources de pollution.

3. Développement du Programme :

- **Collecte des Données** : Développez un programme pour automatiser la récupération des données de l'API de pollution de l'air.
- **Stockage des Données** : Organisez et stockez les données récupérées pour faciliter leur analyse ultérieure.

4. Nettoyage et Transformation des Données :

- **Nettoyage** : Assurez-vous que les données sont cohérentes, complètes et prêtes pour l'analyse.
- **Transformation** : Préparez les données (agrégation, normalisation, etc.) pour l'analyse et les visualisations.

5. Analyse des Données :

- **Exploration des Corrélations** : Analysez les corrélations entre les niveaux de pollution et les facteurs démographiques et géographiques.
- **Identification des Tendances** : Identifiez les tendances significatives dans les données.

6. Automatisation avec Apache Airflow :

- **Pipeline ETL** : Utilisez Apache Airflow pour automatiser l'extraction, la transformation et le chargement des données (ETL).
- **Planification** : Mettez en place des tâches planifiées pour mettre à jour les données et analyses régulièrement.

7. Création du Tableau de Bord :

- **Visualisations** : Créez des visualisations interactives pour présenter les résultats de votre analyse.
- **Outil** : Utilisez un outil de visualisation de données tel que Tableau, Power BI, ou Dash.

Livrables :

- **Programme de Collecte et d'Analyse des Données** : Code source documenté et instructions pour l'exécution.
- **Rapport d'Analyse** : Document décrivant les données utilisées, la méthodologie, les analyses réalisées, et les conclusions.
- **Tableau de Bord Interactif** : Présentation visuelle des résultats, incluant des cartes, des graphiques et d'autres éléments interactifs.

Évaluation :

Votre travail sera évalué sur la base de :

- La qualité et la précision des données récupérées.
- La pertinence des analyses effectuées et des corrélations identifiées.
- L'efficacité et la fiabilité du pipeline ETL automatisé avec Airflow.
- L'esthétique et la fonctionnalité du tableau de bord interactif.

Délai :

Le projet doit être complété et soumis avant le 11 août.

Bonne chance !

Bonne chance et n'oubliez pas de documenter chaque étape de votre travail pour faciliter la compréhension et la reproduction de vos résultats.