

Description du Projet 1 :**Création d'une Application Streamlit pour
l'étude de la Qualité de l'Air et de la Pollution****1. Introduction**

Dans ce projet, vous jouerez le rôle d'un Data Scientist environnemental travaillant pour un organisme public chargé de surveiller la qualité de l'air au Canada.

La pollution atmosphérique est un enjeu de santé publique majeur. Elle est influencée par plusieurs facteurs tels que la concentration de polluants chimiques (PM2.5, PM10, NO₂, SO₂, CO), les conditions météorologiques (température, humidité), la densité de population et la proximité des zones industrielles.

L'objectif principal est de comprendre comment ces variables interagissent et influencent le niveau de qualité de l'air d'une région donnée.

À travers ce projet, vous effectuerez une analyse statistique descriptive, une étude de corrélations et une visualisation dynamique des données afin de déterminer les principaux contributeurs à la pollution atmosphérique. Vous utiliserez Streamlit, un outil de création d'applications web interactives en Python, pour concevoir une interface permettant aux utilisateurs d'explorer les données, visualiser les tendances et interpréter les résultats de manière intuitive.

2. Objectifs du projet

- Explorer et comprendre la structure du jeu de données.
- Identifier les relations entre les différents facteurs environnementaux et la qualité de l'air.
- Déterminer les variables les plus corrélées avec la pollution atmosphérique.
- Créer une **application Streamlit interactive** permettant d'afficher les résultats, graphiques et analyses.
- Présenter un **rapport complet et professionnel** résumant les observations principales et les recommandations.

3. Directives

1. Charger et explorer le jeu de données.
2. Identifier les types de variables (quantitatives, qualitatives).
3. Effectuer un **nettoyage des données** : détection et traitement des valeurs manquantes et aberrantes.
4. Réaliser une **analyse descriptive** (moyenne, médiane, écart-type, quartiles).
5. Étudier les **corrélations** entre les variables et la qualité de l'air.
6. Créer des **visualisations dynamiques** (histogrammes, heatmaps, boxplots, etc.).
7. Présenter vos **conclusions et analysés** dans un rapport clair et concis.

4. Description des données

Ce jeu de données se concentre sur l'évaluation de la qualité de l'air dans différentes régions. Il contient des facteurs environnementaux et démographiques essentiels qui influencent les niveaux de pollution.

- Température (°C) : Température moyenne de la région.
- Humidité (%) : Humidité relative enregistrée dans la région.
- Concentration de PM2.5 ($\mu\text{g}/\text{m}^3$) : Niveau de particules fines en suspension.
- Concentration de PM10 ($\mu\text{g}/\text{m}^3$) : Niveau de particules grossières en suspension.
- Concentration de NO₂ (ppb) : Niveau de dioxyde d'azote.
- Concentration de SO₂ (ppb) : Niveau de dioxyde de soufre.
- Concentration de CO (ppm) : Niveau de monoxyde de carbone.
- Proximité des zones industrielles (km) : Distance jusqu'à la zone industrielle la plus proche.
- Densité de population (habitants/km²) : Nombre d'habitants par kilomètre carré dans la région.

Variable cible : Niveau de qualité de l'air :

Bonne (0) : Air propre avec de faibles niveaux de pollution.

Modérée (1) : Qualité de l'air acceptable avec la présence de certains polluants.

Mauvaise (2) : Pollution notable pouvant causer des problèmes de santé pour les groupes sensibles.

Dangereuse (3) : Air fortement pollué posant de graves risques pour la santé de la population.

5. Questions d'analyse à traiter

Lors de votre analyse et de la conception de votre application, répondez aux questions suivantes à l'aide des statistiques et visualisations produites :

1. Quels sont les types de données présents dans le jeu de données ?
2. Combien d'échantillons et de colonnes contient le jeu de données ?
3. Y a-t-il des valeurs manquantes dans le jeu de données ? Si oui, dans quelles colonnes ?
4. Appliquer sur le jeu de données une technique pour remplacer les valeurs manquantes.
5. Quels sont les facteurs environnementaux les plus corrélés avec la qualité de l'air ?
6. Quels sont les trois principaux contributeurs aux niveaux de pollution dans le dataset ?
7. Existe-t-il des valeurs aberrantes dans les concentrations de PM2.5 ou PM10 ?
8. Quelle est la moyenne, la médiane et l'écart-type des concentrations de PM2.5 et PM10 ?
9. Quelle est la corrélation entre l'humidité et la qualité de l'air ?
10. Quel est le lien entre la densité de population et les niveaux de PM2.5 ?
11. Quels sont les quartiles des concentrations de monoxyde de carbone (CO) ?

6. Livrables attendus

1. **Application Streamlit interactive** présentant :
 - o L'analyse descriptive complète des données.
 - o Les visualisations statistiques et environnementales pertinentes.
 - o L'interprétation des corrélations et des tendances.
2. **Rapport final** détaillant :
 - o Les résultats obtenus.
 - o L'analyse des principaux facteurs influençant la qualité de l'air.
 - o Des recommandations pour la prévention et la gestion de la pollution.
3. **Lien vers le dépôt GitHub** contenant le code source et la documentation.