AQI Prediction Using Machine Learning

A machine learning project to forecast the **Air Quality Index (AQI)** across U.S. cities using pollution, demographic, and seasonal data.

📌 Overview

Air pollution poses a serious threat to public health. Forecasting AQI can help governments, citizens, and urban planners take proactive steps. This project builds a predictive ML model using AQI trends, pollutant levels, population estimates, and seasonal factors to estimate future AQI levels.

🎯 Objectives

- Predict AQI for a given U.S. city and date.
- Provide interpretable insights into pollution patterns and their drivers.
- Assist policymakers and citizens with proactive planning.

🛠️ Tech Stack

- **Languages & Tools:** Python, Pandas, Scikit-learn, Seaborn, Streamlit
- **ML Model:** Random Forest Regressor
- **Visualization:** Matplotlib, Seaborn, Streamlit UI
- **Deployment:** Streamlit App (local)

🧠 Features Engineered

- Percentage of Unhealthy Days
- PM2.5 and Ozone Days
- Seasonal encoding (Spring, Summer, Winter)
- Total pollutant days
- Population estimates

📊 Dataset Highlights

- **Source**: U.S. EPA, State-Level Pollution Stats, Population Census
- ~3,450 rows after cleaning and merging
- Monthly aggregated AQI across multiple cities
- Handled missing data, capped extreme outliers (AQI > 500)

🔍 Exploratory Analysis

- Positive correlation between AQI and population
- Cities with higher % of PM2.5 and Unhealthy Days showed higher AQI
- Seasonal trends visible in monthly breakdown

🤖 Model Performance

Model	R^2 Score	RMSE	MAE
Random Forest	0.41	17.02	12.35

Random Forest provided the best generalization, interpretability, and stability.

🚀 How to Run

Clone the repo:
git clone https://github.com/YourUsername/AQI-Prediction-Model.git
cd AQI-Prediction-Model

1. Install dependencies:
   pip install -r requirements.txt

2. Run the Streamlit app:
   streamlit run AQI_Prediction_Model.ipynb

📁 Files

- AQI_Prediction_Model.ipynb: Full ML pipeline with code & visuals
- merged_aqi_dataset.csv: Final dataset used for training
- rf_aqi_model.pkl: Saved Random Forest model
- aqi_model_ready.csv: Model-ready cleaned dataset
- monthly_avg_aqi.csv: Initial aggregated AQI dataset

📌 Future Improvements

- Add deep learning models like LSTM for time-based AQI prediction
- Integrate real-time AQI feeds via APIs
- Enhance UI and deploy on cloud (e.g., AWS or Streamlit Cloud)