



Air Quality Analysis Model

Table of Contents

1. [Introduction](#)
2. [Features](#)
3. [Installation](#)
4. [Usage](#)
5. [Data](#)
6. [Dependencies](#)
7. [Contributing](#)
8. [License](#)

1. Introduction

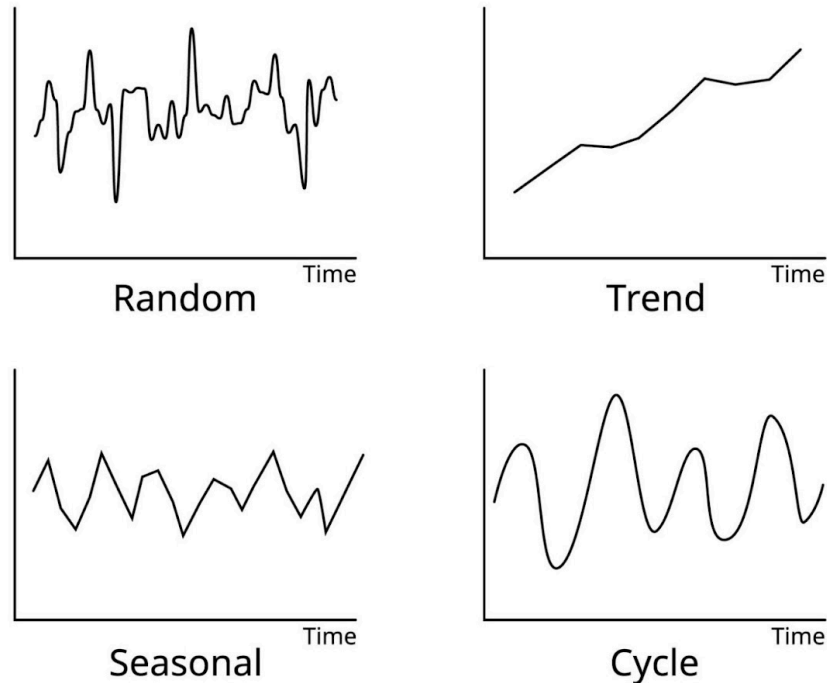
This repository hosts a Python-based model designed for analyzing and visualizing air quality data. The model is built using common data science libraries, providing tools for data cleaning, exploratory data analysis (EDA), time-series forecasting of pollutants (e.g., PM2.5, NO2, O3), and source apportionment insights.

The primary goal is to provide a reliable, easily reproducible framework for understanding air pollution trends and predicting future air quality index (AQI) values.

2. Features

- **Data Ingestion:** Supports loading time-series air quality data from CSV or JSON files.
- **Data Preprocessing:** Handles missing values, performs data normalization, and resamples time-series data to standard intervals (e.g., hourly, daily).
- **Exploratory Data Analysis (EDA):** Generates plots and statistical summaries of key pollutants.
 - Time-series plots of pollutant concentrations.
 - Correlation matrices between pollutants and meteorological factors (temperature, humidity).
 - Seasonal and diurnal variation analysis.
- **Modeling & Forecasting:** Implements time-series forecasting techniques (e.g., ARIMA, Prophet, or LSTM models) to predict future pollutant concentrations and AQI.

Time Series Components



Getty Images

- **Visualization:** Produces interactive and static visualizations using Matplotlib and Seaborn/Plotly.
- **Report Generation:** Capability to generate a summary report of the analysis findings.

3. Installation

Prerequisites

You need Python 3.8+ installed on your system.

Steps

1. **Clone the repository:**

```
git clone
```

```
[https://github.com/yourusername/air-quality-analysis.git](https://github.com/yourusername/air-quality-analysis.git)
```

```
cd air-quality-analysis
```

2. **Create and activate a virtual environment (recommended):**

```
python -m venv venv  
source venv/bin/activate # On macOS/Linux  
# venv\Scripts\activate # On Windows
```

3. **Install the required packages:**

```
pip install -r requirements.txt
```

4. Usage

The main entry point for the analysis is typically a Jupyter Notebook or a main Python script.

Running the Analysis

To run a full analysis pipeline, execute the main script:

```
python main_analysis.py --data_file data/raw_air_quality.csv --pollutant PM2.5 --model_type  
ARIMA
```

Key Scripts

File Name	Description
main_analysis.py	Command-line interface for running the full analysis pipeline.
data_preprocessing.py	Contains functions for cleaning, normalizing, and preparing raw data.
eda.ipynb	Jupyter Notebook for interactive exploratory data analysis and visualization.
forecasting_model.py	Implementation of the selected time-series forecasting model.
requirements.txt	List of all Python dependencies.

5. Data

The model is designed to work with time-series air quality datasets.

Format

The input data file (e.g., data/raw_air_quality.csv) must contain:

- A **DateTime** column (must be parseable as a date/time object).
- Columns for various **Pollutants** (e.g., PM2.5, PM10, NO2, CO, O3).
- Optional **Meteorological** columns (e.g., Temperature, Humidity, WindSpeed).

Example Data

A sample data file is provided in the data/ directory (sample_data.csv) for testing purposes.

6. Dependencies

All required packages are listed in requirements.txt. Key dependencies include:

- pandas: Data manipulation and analysis.
- numpy: Numerical operations.
- scikit-learn: Machine learning utilities (scaling, metrics).
- matplotlib & seaborn: Static data visualization.
- statsmodels / Prophet / tensorflow (optional): Time-series modeling and forecasting.

7. Contributing

We welcome contributions! If you have suggestions or find a bug, please open an issue or submit a pull request.

8. License

This project is licensed under the MIT License. See the LICENSE file for details.