Analysis Report

Basic Visualizations:

1. Time series plots

CO

A graph showing a number of blue lines

AI-generated content may be incorrect.

Nox

A graph showing a number of blue lines

AI-generated content may be incorrect.

Benzene

A graph showing a number of times

AI-generated content may be incorrect.

These graphs shows the CO, NOx and Benzene concentration levels in the Air over the year. The peaks show the higher emission levels and traffic levels, and it is seen that the emission levels are higher in the winter months compared to the summer and fall. This indicates a seasonal influence on pollutant levels except for Benzene.

1. Daily/Weekly Pattern

A graph of pollutants

AI-generated content may be incorrect.A graph with orange line

AI-generated content may be incorrect.

The peaks seen at 8AM and 8PM suggests the effect of peak traffic on the pollutants emission on the road, with NOx being the most emitted pollutant in terms of quantity. CO and C6H6 have visible morning and evening peaks as well. The higher average over the first few days of the week suggests increased traffic and industrial activities during the weekdays compared to the weekends.

1. Correlation matrix

The graph shows strong correlations between CO, NOx, and Benzene which indicates shared emission sources. There is a mild correlation between time-based features like Hour and pollutant levels.

A colorful chart with numbers and a graph

AI-generated content may be incorrect.

Advanced Visualizations

1. Autocorrelation and Partial Correlation for CO

A graph with blue dots

AI-generated content may be incorrect. A graph with blue dots

AI-generated content may be incorrect.

1. Partial correlation CO

High autocorrelation at lag 1 and every 24 hour cycle suggest the influence of peak traffic and industrial times on the quantity of CO in the air. After controlling for intermediate lags, the partial correlation graphs show the directly predictive lags. The lags around 3 and 24 suggest the influence of traffic cycles. This is useful for a time based modelling.

1. Seasonal Trend

A graph of different types of data

AI-generated content may be incorrect.

The decomposition shows a clear long-term trend and some repeating seasonality. The residuals show variability not explained by the trend or seasonality.

potential factors influencing air quality variations

1. Traffic Patterns: There is a visible correlation between the peak traffic hours and amounts of pollutants emitted in the air, indicating vehicular emissions to be a main source of pollution, especially for NOx
2. Weather and Seasons: Its seen that there is a pattern exhibited in the amount of pollutants emitted that varies with season. There are significantly more pollutants emitted during the winter when compared the summer and fall.
3. Industrial Activity: More pollutants emitted during the weekdays than the weekends show that industrial emissions during the work week also contributes to the NOx , CO and Benzene emissions.
4. Time of the day: Human activities (traffic and work hours) play a role in the emissions as well.

Modelling approach

Based on these observed patterns and factors, time-aware models like ARIMA, SARIMA, or ML models with lagged/time features such as Random Forests, XGBoost with time features looks to be more suitable for this dataset.

Some of the reasons for choosing such models are that variables like 'Hour' and 'DayOfWeek' should be included as categorical or cyclical features and that long-term trend and seasonality justify decomposition-based models or seasonal regressors