Phase 2: Exploratory Data Analysis (EDA) on Time Series Data

The analysis provides trends and relationships in the air quality dataset. We focus on three pollutants- Carbon Monoxide (CO), Nitrogen Oxides (NOx) and Benzene (C6H6).

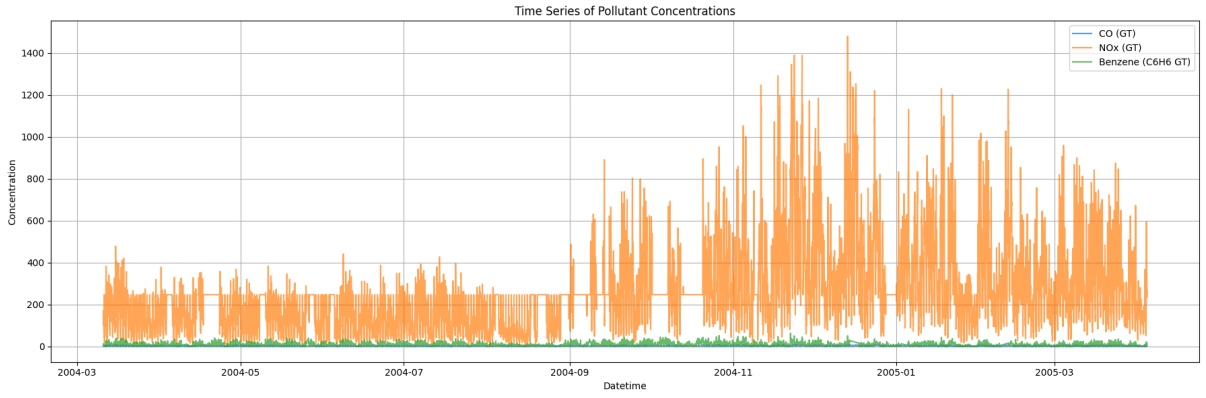
Key Pollutants Tracked:

- CO(GT): Carbon monoxide
- C6H6(GT): Benzene
- NOx(GT): Nitrogen oxides

Basic Visualizations

1. Time-Series Analysis of Pollutant Concentrations

    - The time series plot of the pollutants show clear fluctuations in their concentrations overtime, with noticeable peaks during the day, especially during early mornings and evenings. This suggests that CO levels might be because of human intervention, such as traffic patterns or human activities such as working industries.
    - NOx concentrations also show a similar pattern, mixing with CO analysis, which might be because of vehicular emissions during peak hours.
    - Benzene does show some peaks, but is typically related to chemical processes and industrial processes.

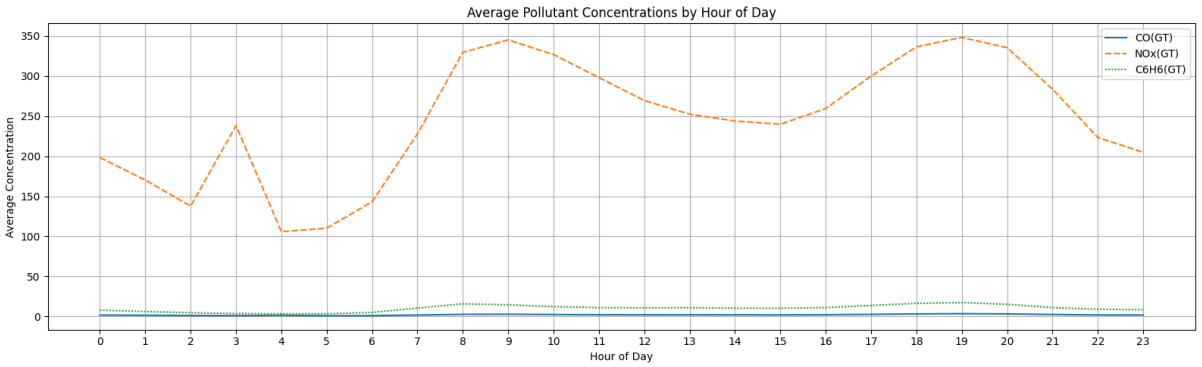


2. Hourly and Daily Patterns

Hourly trends:

- NOx peaks during morning and evening rush hours, aligning with traffic patterns.
- CO shows the same pattern but with less intensity.
- Benzene remains the same/stable with minor variations.

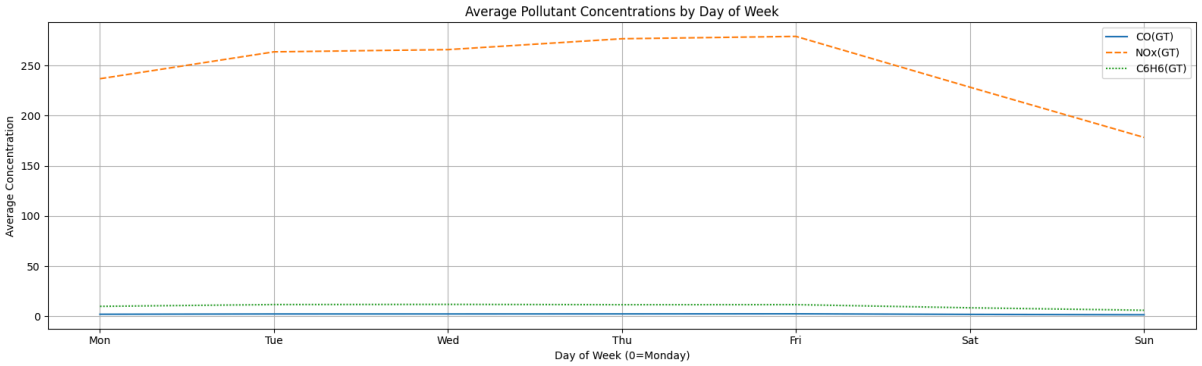
The highest pollutant levels occur during early morning and late evening, likely due to rush hours when traffic congestion is at its peak.
Notably, pollutant levels drop during late night and early mornings when traffic decreases and industrial activities slow down.



Weekly trends:

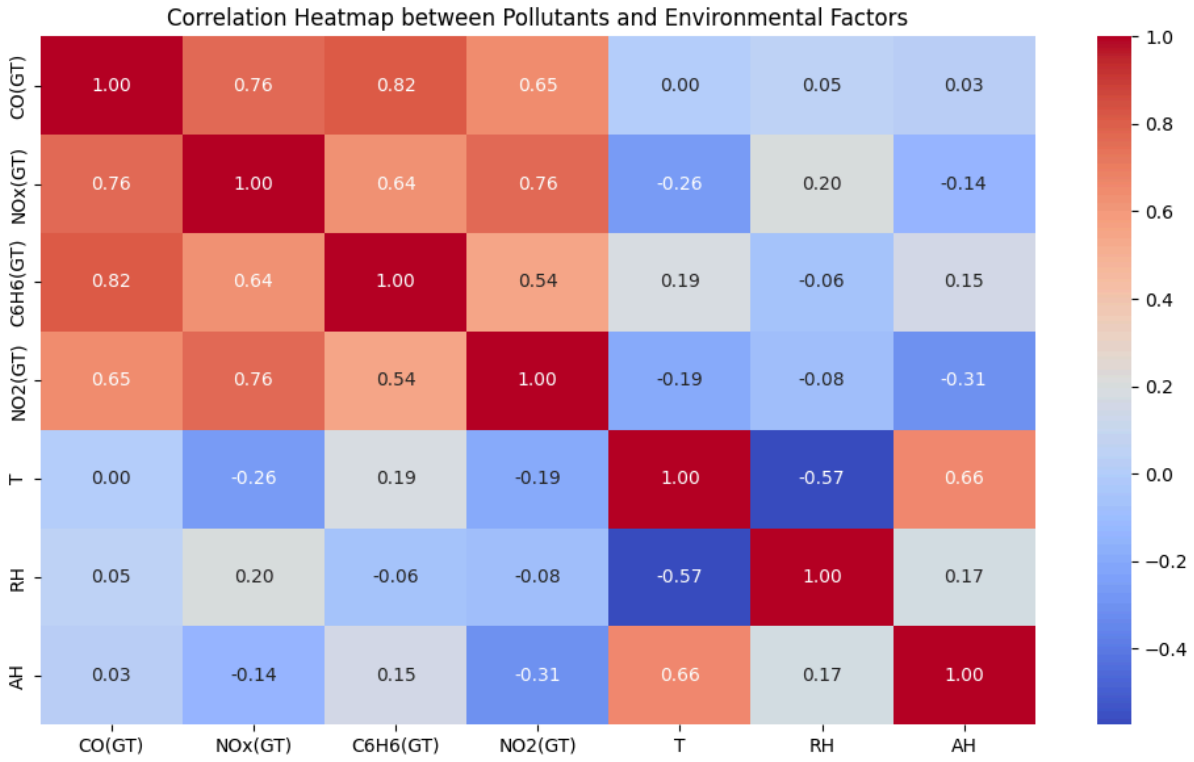
- Higher NOx trends on weekdays, than on weekends. This is consistent with the theory of vehicular emissions and their relation to traffic.
- CO and Benzene show minor variations but remain stable most of the time.

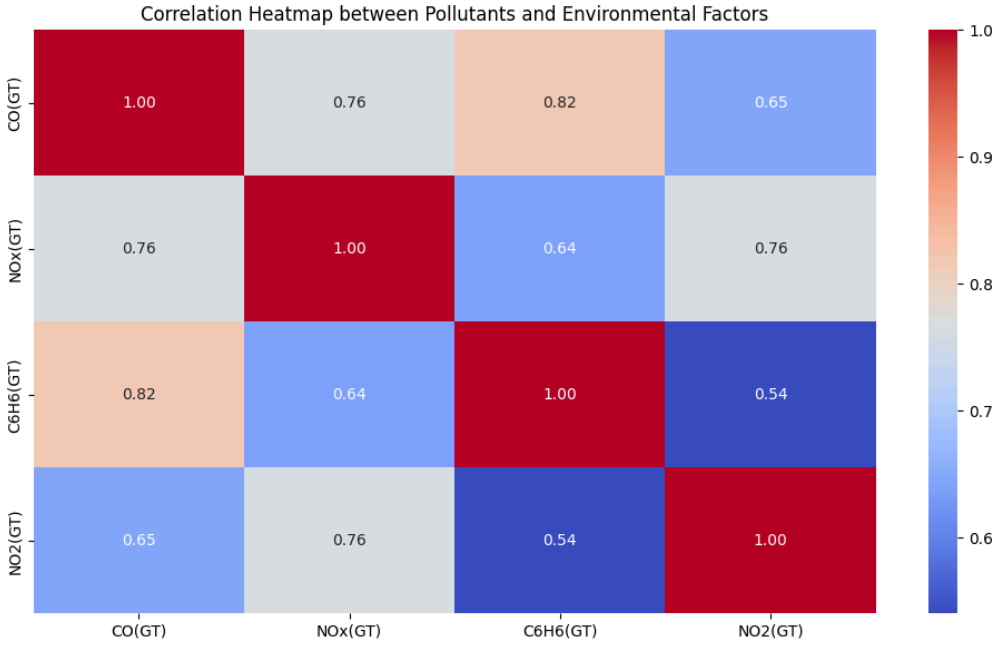
Weekends tend to have lower pollutant concentrations, reflecting a reduction in both traffic and industrial emissions.



3. Relationships between Pollutants

A correlation matrix shows:





The correlation heatmap shows positive correlation between CO, NOx, and Benzene. This means that they are not so independent of each other.
CO and NOx are associated with traffic emissions and thus are correlated.
The correlation between CO and Benzene is moderate, which may suggest that both are not so related and there might be other factors that influence the increase of benzene independently.
The correlation between NOx and Benzene is weaker, indicating that they might share some sources but are mostly influenced by different factors.

4. Potential Influencing Factors

Factor	Influence Level	Description
Traffic Volume	High	Strong rush hours signals in CO and NOx patterns
Day of the week	High	Lower pollutants on weekends indicate human related emissions, be it by traffic or Industrial activities.
Temperature/Humidity (Seasonality)	Medium	Can influence reactions or pollutant dispersion.

5. Implications for Modelling

- Time-Aware Modelling: Models like LSTM or ARIMA can be used to capture seasonality and trend components.
- Multivariate approach: Considering correlated pollutants together may enhance accuracy.
- Anomalies and Outliers: The pollutant levels might change because of some special events or some unusual weather patterns.
- Seasonality: Seasonality can play a huge role in this if we consider temperature a major feature.