

AIR QUALITY ASSESSMENT

Bett Kipkemoi

Data Quality Assessment

- Original dataset size: 9,357 rows × 15 columns
- Cleaned dataset size: 8,991 rows × 16 columns
- Data loss: 3.9%
- Missing value handling: -200 values replaced with NaN
- Date/time conversion: Successfully converted to datetime format

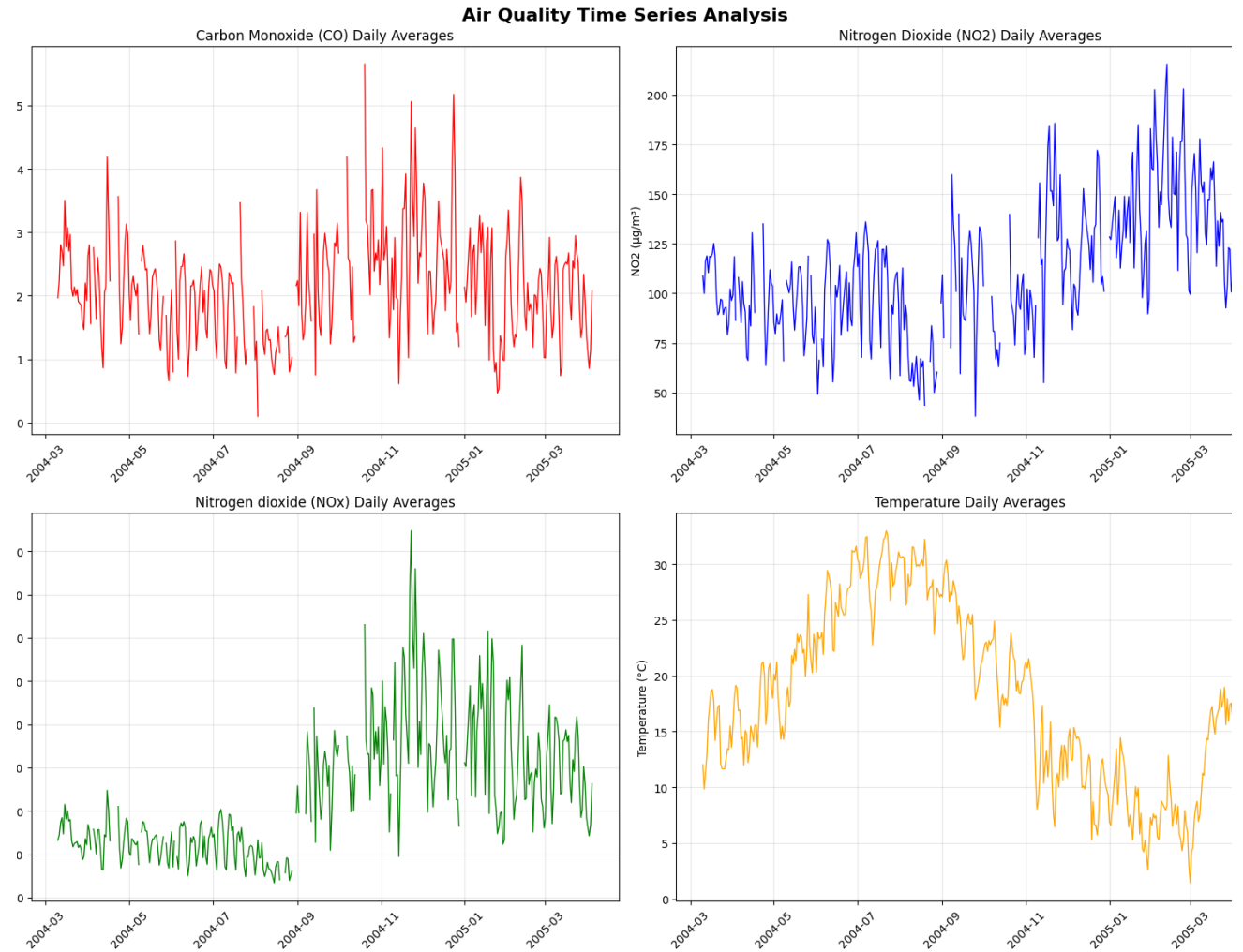
Pollutants Levels Summary

CO: Mean = 2.130, Standard deviation = 1.436

NO₂: Mean = 112.145, Standard deviation = 47.629

NO_x: Mean = 242.189, Standard deviation = 206.312

Benzene: Mean = 10.083, Standard deviation = 7.450



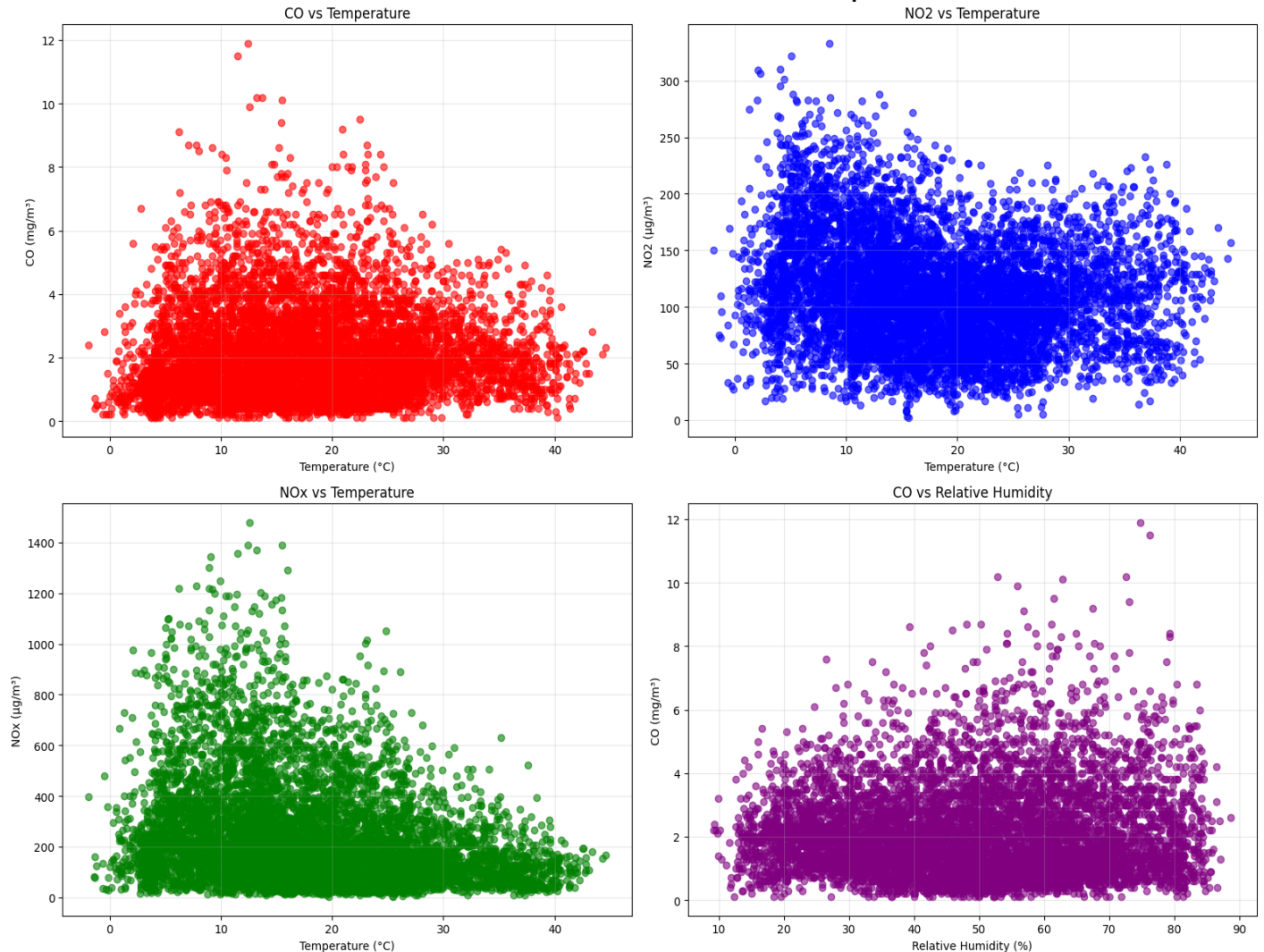
Environmental Factors Summary

Temperature: Mean = 18.318, Standard deviation = 8.832

Relative Humidity: Mean = 49.234, Standard deviation = 17.317

Absolute Humidity: Mean = 1.026, Standard deviation = 0.404

Pollutant vs Environmental Factor Relationships

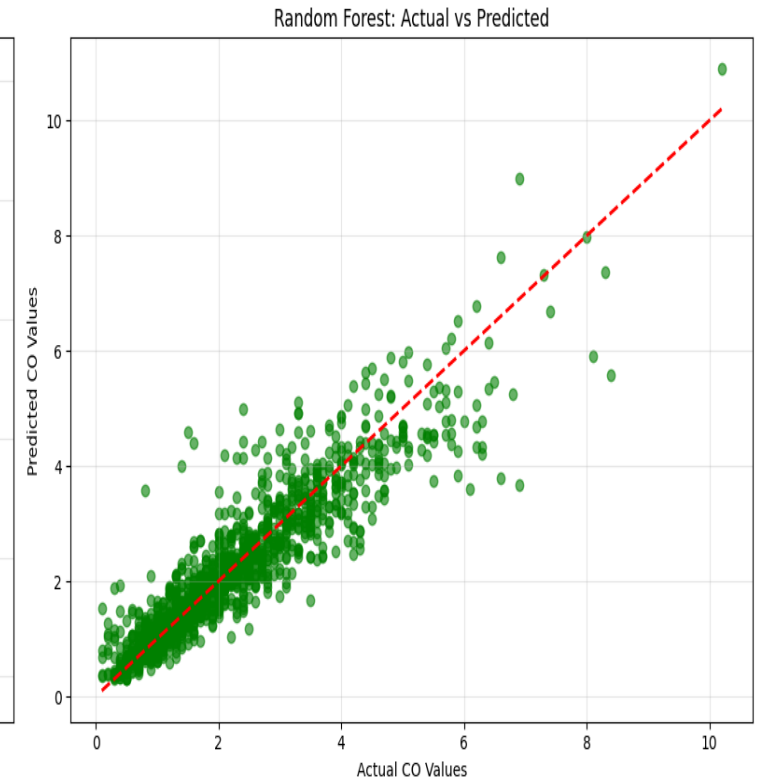
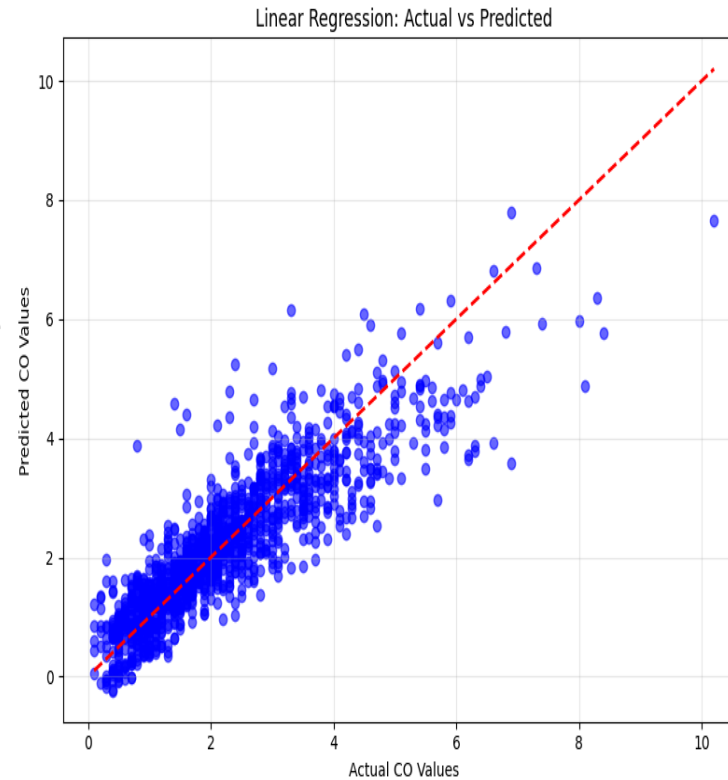


Correlations

- The top correlations with temperature are:
 - Absolute Humidity: 0.656
 - Relative Humidity: 0.579
 - PT08_S4_NO2: 0.561
 - NMHC: 0.392

Predictive Modelling Results

- Linear Regression R^2 : 0.7679
- Random Forest R^2 : 0.8395
- From the R^2 , the best performing model is Random Forest with a score of 0.8395, or 83.95%.
- The models are plotted against each other as shown in the images.



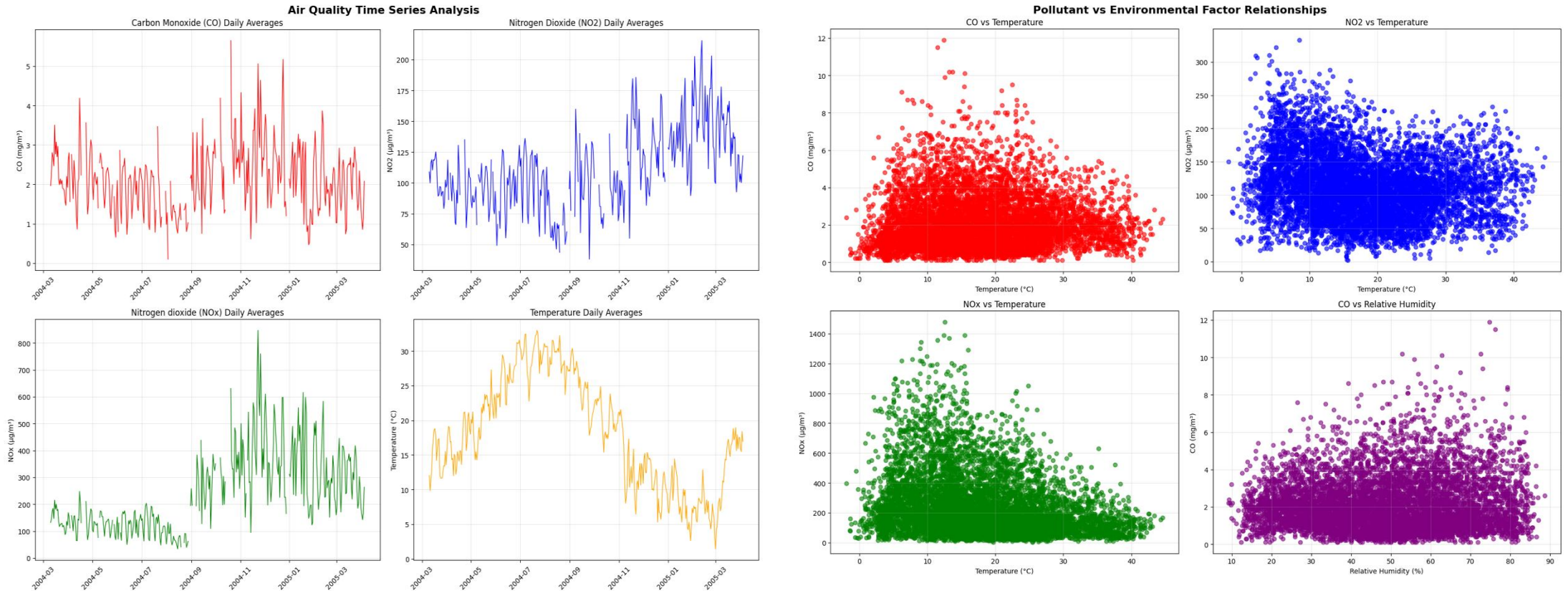
Environmental Insights

Temperature shows significant correlation with CO, NO₂, and Nox pollutants

Relative humidity has varying effects on different pollutants

Seasonal patterns are evident in the time series data

CO levels show strong correlation with temperature and other pollutants



Public Health Implications

- Monitoring CO levels is crucial for public health
- Temperature and humidity significantly influence air quality
- Seasonal variations suggest the need for adaptive air quality management
- Predictive models can help in early warning systems

Recommendations for Future Research

- Extend analysis to include more recent data
- Investigate spatial variations in air quality
- Develop more sophisticated time series forecasting models
- Analyze the impact of specific weather events on air quality
- Study the relationship between air quality and health outcomes