Report: Analyzing the Impact of Air Quality Measures on Climate Dynamics in Urban Settings

Introduction

Investigating the interplay between urban air quality and climate dynamics is essential for developing effective public health strategies and environmental policies. Urban centers, with their dense populations, high traffic volumes, and industrial activities, are major sources of air pollution. This pollution not only poses significant health risks but also contributes to climate change. This analysis seeks to answer the question: "How do variations in air pollutant levels in different urban environments correlate with climate change indicators, and what predictive models can be constructed to forecast these impacts on urban air quality?"

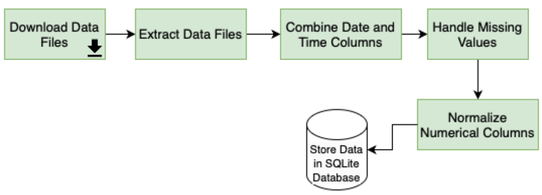
Used Data

The Beijing Multi-Site Air Quality Data, sourced from the UCI Machine Learning Repository, provides hourly measurements of various pollutants, including PM2.5, PM10, SO2, NO2, and O3, along with meteorological data such as temperature and humidity, collected from multiple monitoring stations in Beijing between 2013 and 2017. This dataset is structured in CSV format and is available under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

The Inorganic Gases-2017 dataset, obtained from the European Commission's Joint Research Centre (JRC), includes measurements of PM2.5, PM10, NO2, and O3 from various European locations during the year 2017. This dataset is also provided in CSV format and is licensed under the European Union Open Data Portal license. Both datasets were selected for their high relevance and quality, facilitating a comprehensive analysis of urban air quality and climate dynamics across different regions.

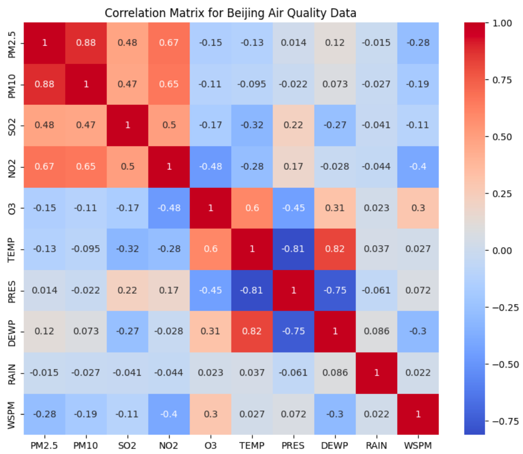
Analysis

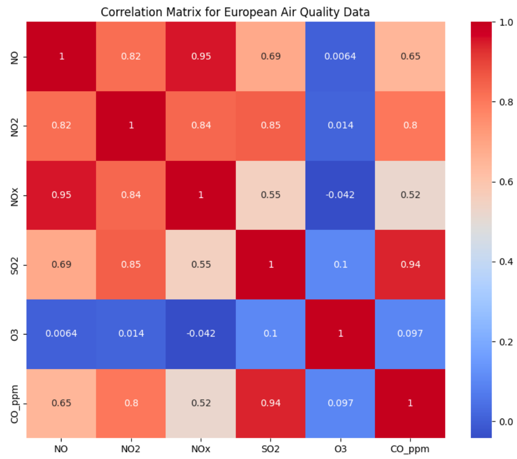
The data processing for this analysis was done via a data pipeline written in Python, most of which automated and made seamless many of the stages in handling the data. This pipeline started by automatically downloading data files from their locations and extracting them. This provided an effective way of consistently retrieving the data. The data processing phase consisted of combining the separate date and time columns into a single datetime column. Handling missing values was one of the important processes, where incomplete data entries were dealt with aptly to avoid any type of data loss. The numerical columns were then normalized so that comparisons between the various data may be done with accuracy. The cleaned and normalized data was then stored back into an SQLite database. This method of storage was chosen for efficiency in both query speed and analysis of large datasets that would facilitate the later steps in analysis.



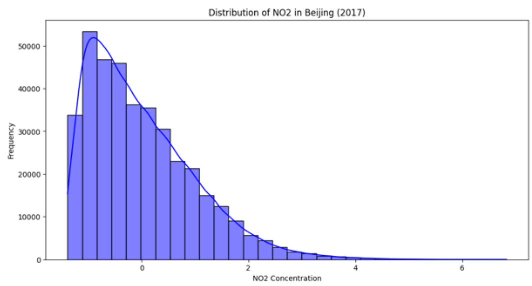
Proceeding to the data analysis, we explored various distributions and correlations to uncover meaningful insights.

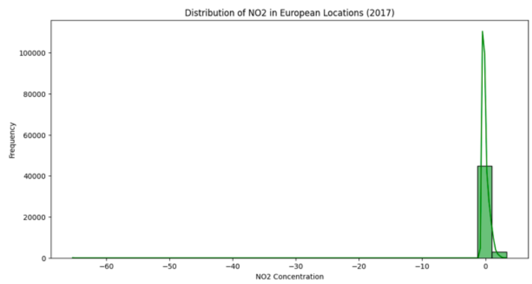
Correlation matrices provided further insights into the pollution sources in each region. In Beijing, there were strong correlations between PM2.5, PM10, and NO2, indicating common pollution sources likely related to industrial activities and vehicular emissions. In European locations, high correlations between NO, NO2, and NOx suggested that vehicle emissions were a major source of pollution.

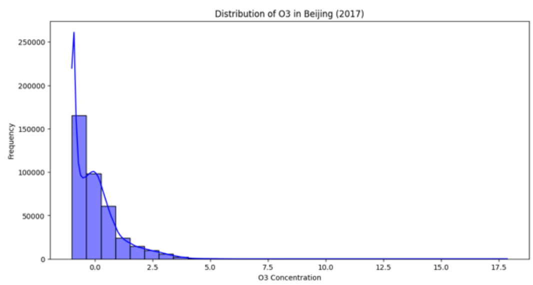


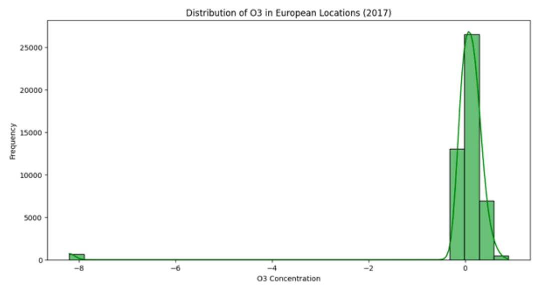


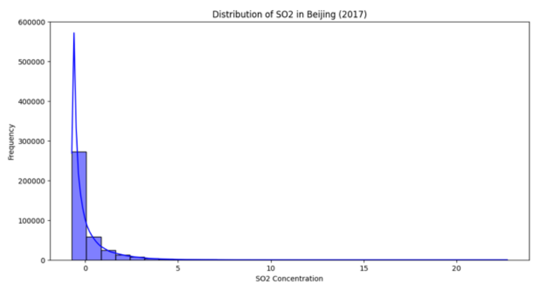
The analysis of NO2, O3, and SO2 concentration distributions revealed distinct patterns between Beijing and European locations for the overlapping period of 2017. In Beijing, the NO2 distribution exhibited a right-skewed pattern with a long tail, indicating a higher frequency of lower NO2 concentrations. In contrast, European locations displayed a sharp peak near zero, with some negative values suggesting potential data quality issues or differences in measurement standards. Similarly, the O3 concentration distribution in Beijing was right-skewed, with most values concentrated at the lower end, while European locations showed a sharp peak around zero, mirroring the NO2 data pattern. The SO2 concentration distribution also differed significantly between the two regions: in Beijing, the distribution was heavily right-skewed with very high frequencies at low concentrations, whereas in European locations, the unusual distribution with negative values pointed to potential data errors. These discrepancies highlight the challenges in comparing air quality data across different regions with varying data collection standards and practices.

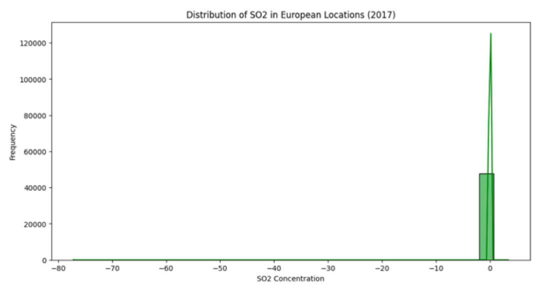




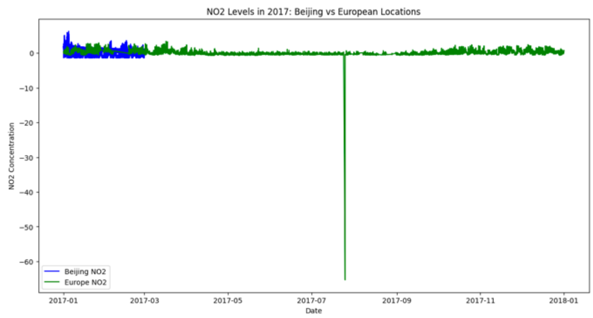


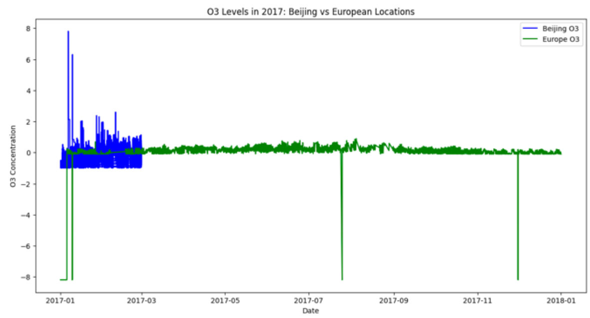


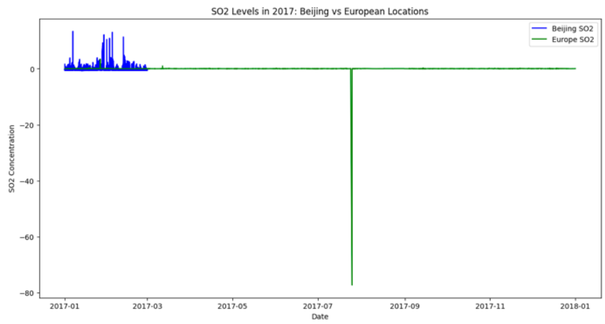




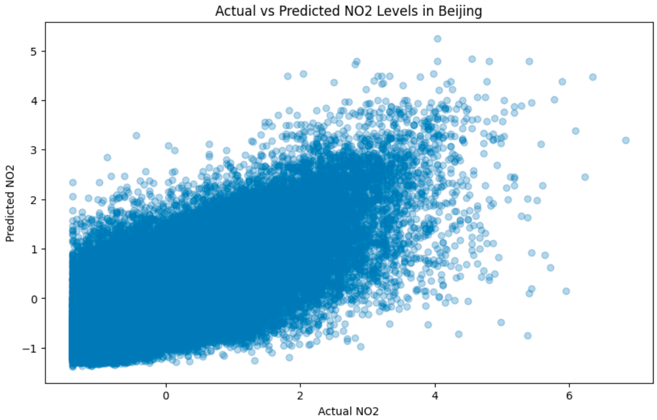
The temporal variations in NO2, O3, and SO2 levels for the overlapping period of 2017 highlight different patterns in pollutant levels between Beijing and European locations. In Beijing, the pollutant levels showed significant fluctuations with numerous peaks, particularly in NO2 and SO2 concentrations, indicating higher pollution events likely driven by industrial activities and vehicular emissions. In contrast, the European locations displayed relatively stable pollutant levels with minimal fluctuations around zero, though occasional negative values suggest potential data inconsistencies or differences in measurement standards. These variations reflect regional differences in pollution sources and control measures, underscoring the complexities in managing urban air quality across diverse geographic regions. The observed discrepancies highlight the necessity for tailored strategies to address specific pollution sources and patterns in different regions, guiding more effective environmental policies and urban planning.

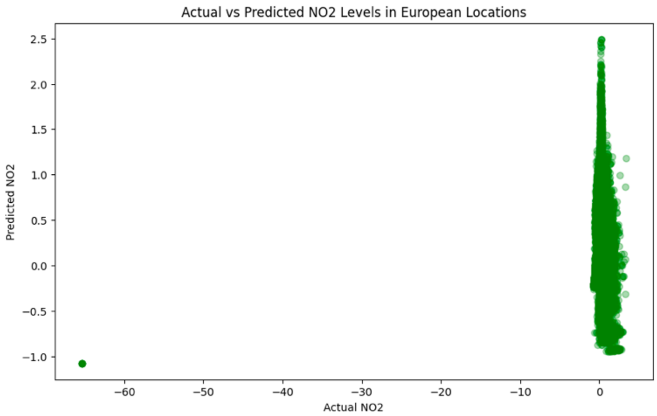






Predictive models for NO2 levels were developed for both regions, with the scatter plots comparing actual versus predicted NO2 levels illustrating the models' accuracies. The model for Beijing showed better performance, likely due to the more consistent and higher quality data available from Beijing compared to the European dataset. The choice to model NO2 was based on its significant presence and variability in urban air pollution, its strong correlations with other pollutants, and its critical impact on public health. NO2 is a key indicator of air quality and is often associated with respiratory issues and other health problems. Moreover, the distinct patterns in NO2 distribution between Beijing and European locations provided a valuable basis for developing and testing predictive models, allowing for a comprehensive understanding of the factors influencing urban air quality in different regions.





Conclusion

This study explored the correlation between air pollutant levels and climate change indicators in urban environments, focusing on Beijing and various European locations for the overlapping period of 2017. We found distinct patterns, with Beijing showing right-skewed distributions for NO2, O3, and SO2, indicating frequent lower concentrations but significant pollution events from industrial activities and traffic. In contrast, European locations had stable pollutant levels with sharp peaks around zero, although occasional negative values suggested data inconsistencies.

Predictive models for NO2 levels were more accurate for Beijing, likely due to the consistent and high-quality data available. NO2 was chosen for modeling because of its variability in urban pollution, strong correlations with other pollutants, and critical public health impact. While the study highlighted important regional differences, future research should standardize data collection and explore additional pollutants and comprehensive climate data to enhance model robustness. These insights can inform more effective environmental policies and urban planning, improving public health and climate resilience in urban areas.

The primary limitation of this study was finding good datasets that contained useful data and had appropriate licensing, which resulted in the study's topic changing twice. Additionally, the datasets only overlapped in 2017, providing just one year of common data. A longer period of overlapping data would have enabled a more comprehensive analysis. Furthermore, the datasets only included common measurements for NO2, O3, and SO2, limiting the ability to check for other important pollutants. Expanding the range of pollutants and the duration of data overlap would enhance the robustness and applicability of future analyses.