Dataset: https://archive.ics.uci.edu/dataset/360/air+quality
GROUP 24-003

- ## Prediction Goal:

Our goal is to predict the concentration of carbon monoxide (CO) one hour in the future using sensor measurements, weather variables (wind speed, temperature, and humidity), and hour/ day of the week. This model will identify how CO levels fluctuate at different times with a focus on identifying peaks in morning/ evening traffic.

- ## Project Motivation and Significance:

Air pollution, especially carbon monoxide (CO), is a continuing problem for the environment in urban areas. Specifically, CO levels tend to fluctuate due to many factors such as temperature, humidity, and wind speed. Additionally, human activity such as rush-hour traffic can also have a great effect. By predicting CO levels in advance, we can gain valuable insights for both the citizens of the city and the policy makers.

From a bigger picture, looking at a societal level, accurate short-term forecasts could allow city residents (especially with respiratory conditions) to make safe decisions about when they go outside. Business and industries could leverage information about optimal traffic controls, make efforts to reduce emissions, and prepare to comply with stricter environmental regulations.

Our project aligns with newer developments in data-driven environmental information. As governments and industries shift toward improvements in sustainability, we can reduce the negative environmental problems caused by emissions and pollution, and reduce the impacts associated using the information

- ## Data Overview:

Our data for this project incorporates hourly measurements of multiple air pollutants from outside an urban environment. These various pollutants include Carbon Monoxide (CO), Non-Methane Hydrocarbons (NMHC), Nitrogen Oxides (NOx), and Benzene (C6H6), among others. Along with air pollutants, the data includes meteorological measurements of Temperature (T), Relative Humidity (RH), and Absolute Humidity (AH). The time series data was captured hourly across one year from 2004 to 2005. With the time series nature and amount of meaningful variables, this sets a perfect stage for prediction. We can take this any way, whether that is the future prediction of singular pollutants or relationships between variables such as temperature and CO, and even using predictive models to detect anomalies.

Dataset: https://archive.ics.uci.edu/dataset/360/air+quality
GROUP 24-003

- ## Preliminary Methods:

For this project, we will try a variety of different analytical methods to try to first learn about the data, and then afterwards develop more sophisticated models to see if we can better fit the data. We will start with doing multiple linear regression, with some alterations such as PCR, or feature selection, to see how accurately we can predict the amount of carbon monoxide in the atmosphere. After doing this, and learning about the data, we will likely broaden our scope to other analytical methods such as random forest, neural networks, and especially time series methods such as utilizing lag variables. Before we can do any of this however, we will have to find out what to do with the large amount of missing variables, we may omit the data, or instead if we see a pattern with carbon monoxide concentration, we can instead use methods such as weighted average filling or other fill methods.

- ## Expected Contributions/Outcomes:

By the end of this project, we hope to accomplish many outcomes due to our contributions. To begin, we hope to develop machine learning models that are capable of forecasting hourly carbon monoxide concentrations based on environmental and pollutant data. We also hope to identify which meteorological factors and pollutant levels are the strongest predictors/ contributors of air quality. By being able to identify these insights, we hope our data will be used to help inform environmental monitoring agencies, policymakers, and urban developers/businesses. Beyond the technical deliverables, we hope to have our project demonstrate the importance of the societal value of predictive analytics. Our accurate forecasting will be able to apply to many different sectors such as public health, policy and urban planning, and business decision making with a focus on sustainability impact. In regards to the public health sector, we hope that hospitals and healthcare organizations will be able to prepare for potential spikes in respiratory or cardiovascular admissions by anticipating poor air quality days. With policy and urban planning, local governments will be able to implement dynamic measures to mitigate pollution such as public transportation incentives, traffic control measures, and emissions regulations. Business decision making with a focus on sustainability impact can benefit greatly from our predictive air quality insights. Companies in manufacturing can use forecasts to optimize operations, reduce emissions during high-risk periods, and ensure compliance with environmental regulations- and perhaps be incentivized by it if working with local officials. This not only minimizes financial and reputational risks but also strengthens companies goals to corporate social responsibility. In all, our project is hoped to showcase that predictive analytics can be used and turned into actionable strategies that push towards healthier communities, smarter policies, and more sustainable business practices.

Dataset: https://archive.ics.uci.edu/dataset/360/air+quality

GROUP 24-003