My Report

To better understand the applied machine learning techniques to investigate air quality trends using atmospheric data from Tiantan, Beijing, China.

Introduction

Cities are major producers of particulate matter, contributing to influencing the urban atmospheric environment. Residents who live in polluted areas are often exposed to chemicals and at risk for more health problems. For example, researchers found that there is a positive correlation between PM2.5, tropospheric ozone, and preterm birth (Mekonnen et al. 2021).

Additionally, atmospheric chemical composition can directly affect cloud formation and feedback loops, which can impact radiative effects within a city (Vo, T. T., Hu, L., Xue, L., & Chen, S., 2023). Aerosols can act as cloud condensation nuclei (CCN), where water vapor condenses over the suspended particles. Coupled with preexisting excess heat from the Urban Heat Island (UHI) effect, cities with high density are more prone to having areas with dangerously elevated extreme heat (Chapman, S., Watson, J. E. M., Salazar, A., Thatcher, M., & McAlpine, C. A., 2017).

In China, air quality criteria follow a different index than in the United States. Here are two tables with the corresponding Air Quality Index (AQI) values for each nation:

China AQI

AQI Value	*Description*	*PM2.5 Concentration (ug/m3)*
0 – 50	Excellent	0 – 35
51 – 100	Good	35 – 75
101 – 150	Lightly Polluted	75 – 115
151 – 200	Moderately Polluted	115 – 150
201 – 300	Heavily Polluted	150 – 250
301 – 500	Severely Polluted	250 – 500

US AQI

AQI Value	*Description*	*PM2.5 Concentration (ug/m3)*
0 – 50	Good	0 – 12
51 – 100	Moderate	12.1 – 35.4
101 – 150	Unhealthy for Sensitive Groups	35.5 – 55.5
151 – 200	Unhealthy	55.6 – 150.4

201 – 300	Very Unhealthy	150.5 – 250.4
301 – 500	Hazardous	250.5 – 500

As climate change alters environmental trends, air quality measurements, including particulate matter and aerosols, will be altered in many ways. Within this data set, machine learning can be used to assist in predicting air quality days based on environmental factors. With predictions from climate models, machine learning can help with simulating air quality and be useful for public safety and warnings. Vulnerable communities with fewer resources need to be aware of risks and be assisted in policies.

Data

The data used was sourced from the University of California Machine Learning repository that intentionally serves as a database for many machine learning examples. Specifically, the data used comes from Beijing at the Tiantan air quality monitoring site between the years 2013 - 2017. This data included a variety of variables such as PM2.5, PM10, SO2, NO2, O3, temperature, and precipitation. To understand the data set, the data was separated into each of its variables and related over time.

To understand the climate of the area, the variables were separated into heatmaps which showcase the average for each variable (temperature, precipitation, air pollutants). This method illustrates areas where there is a higher or magnitudes magnitude in a brief summarizing visualization.

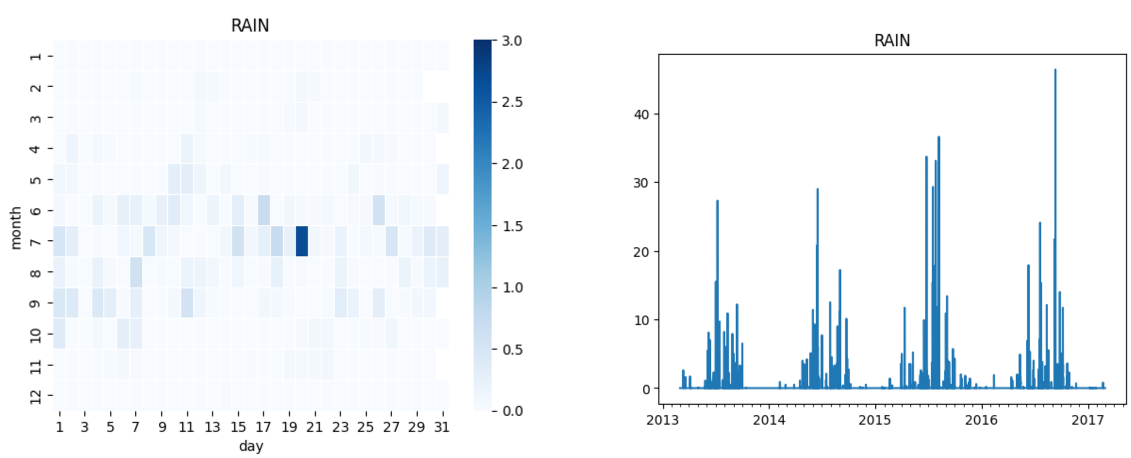

Fig. 1: Precipitation averages heat map (left), precipitation time series (right)

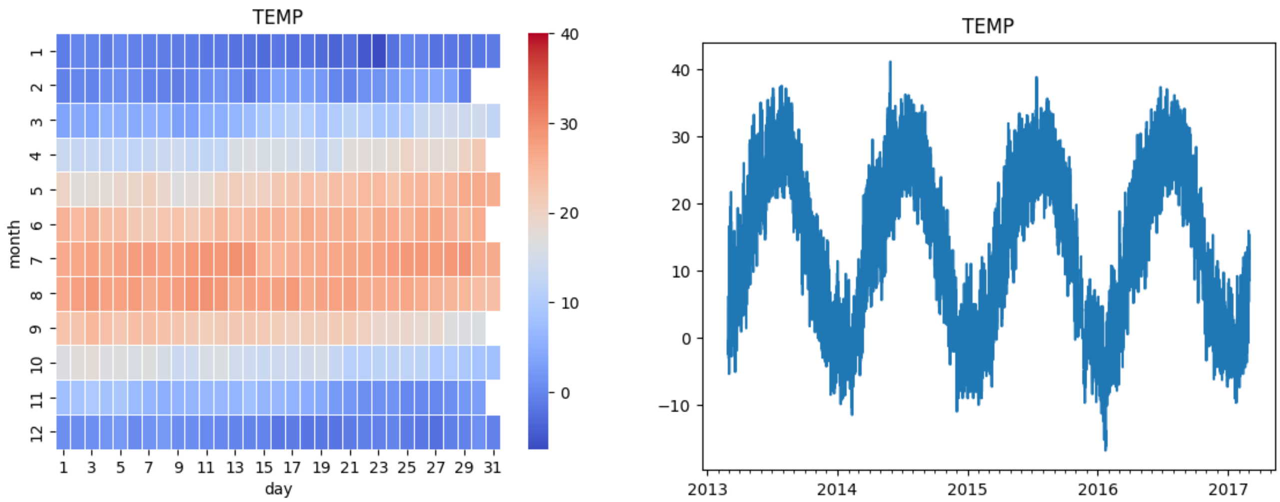

Fig 2: Temperature (deg C) averages heat map (left), temperature time series (right)

Based on this data, there is a clear seasonality that is expected for a midlatitude area in the Northern Hemisphere. Warmer temperatures are within the main summer months of June, July, and August. Precipitation maximums occur over the summer as well, and there is minimal to no precipitation during the winter months.

To allow for easier manipulation, this research will focus solely on particulate matter. Since China's AQI uses PM2.5 as a standard, this will allow for easy adaption into applicable material for policy or implications. Furthermore, research already conducted suggests that there is a negative relationship between particulate matter and precipitation. As rain droplets fall, they take with them particulate matter and other suspended particles in a process known as wet deposition (Tao, W.-K., Chen, J.-P., Li, Z., Wang, C., & Zhang, C., 2012). In the following figure, this relationship is clearly demonstrated. Where there is greater rainfall, there is a local minimum for the particulate matter.

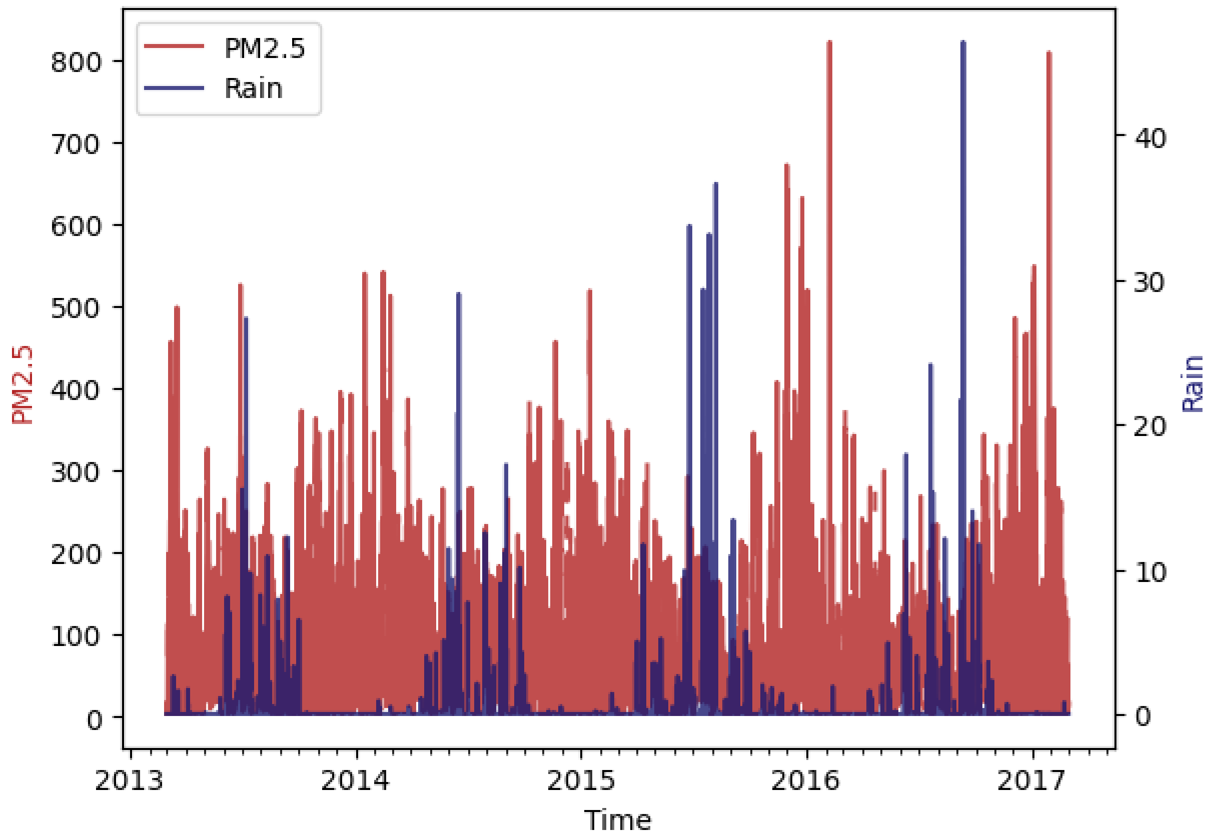

Fig 3: PM2.5 in red and precipitation in blue.

Since there is an excess amount of data, there are many points that make it difficult to create a clear trend. Therefore, data was grouped by day and averaged for particulate matter.

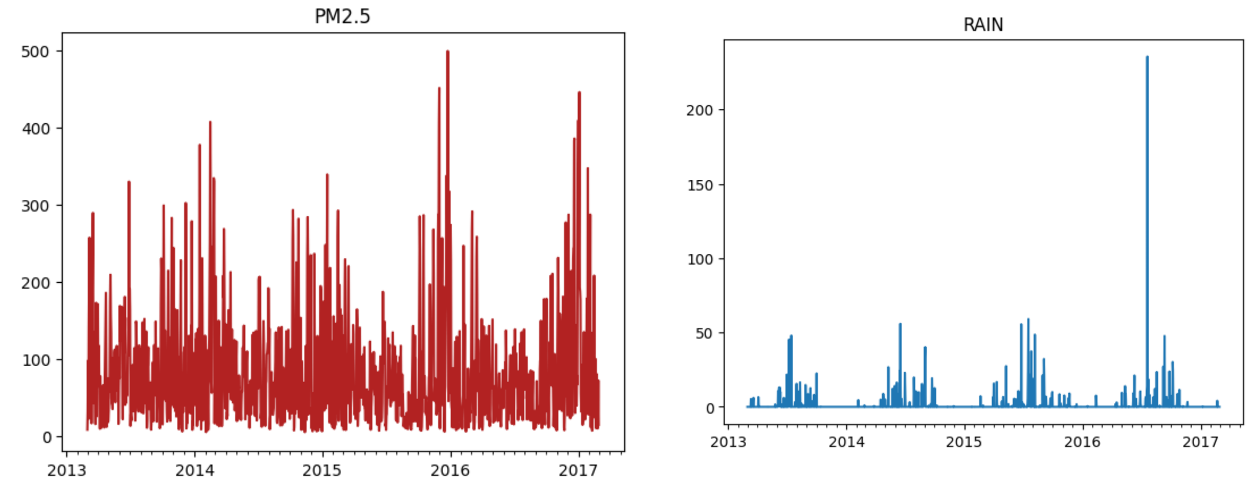

Fig 4: PM2.5 averages per day (left), daily total precipitation (right)

Modelling

For this data set, I began with some trial and error. At first, polynomial linear regression using Ridge regression was performed, and it was found that order 3 had the best fit. However, clearly, in Figure 5, this does not accurately represent the data at all and needs more fitting to have a conclusive result.

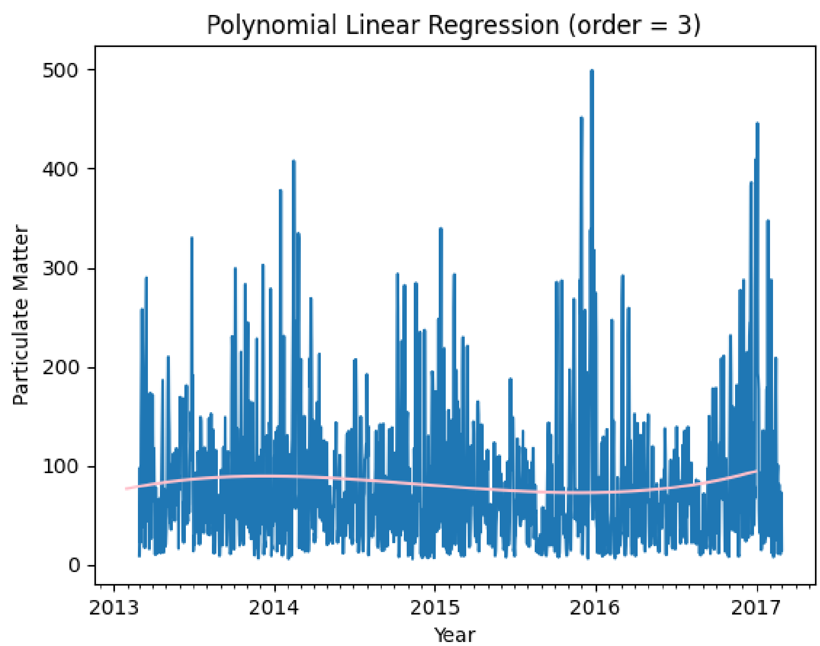

Fig 5: Polynomial Linear Regression (order=3) performed on PM2.5 data

Discussion

As temperature increases, Beijing will only face more dangerous heat waves and extreme heat. Additionally, heat also contributes to the worsening air quality, as seen by PM2.5 concentrations positively correlating with the warmer temperatures. Since Beijing has summer rain, these effects are slightly mitigated. However, rising temperatures will also impact the winter months where rain is much more scarce, contributing to the worse air quality days. As shown by the data, there is a larger portion of days when Beijing is moderately polluted and more (above 150 ug/m3). For the U.S. AQI, this air is unhealthy to hazardous for humans.

Conclusion

Within the data, it is possible to conclude that temperature and precipitation are leading key factors in predicting PM2.5 concentrations. Machine learning is capable of classifying data and air pollutants into different sources as well that was not fully explored in this paper.

References

Chen,Song. (2019). Beijing Multi-Site Air-Quality Data. UCI Machine Learning Repository. https://doi.org/10.24432/C5RK5G.

Chapman, S., Watson, J. E. M., Salazar, A., Thatcher, M., & McAlpine, C. A. (2017). The impact of urbanization and climate change on urban temperatures: A systematic review. Landscape Ecology, 32, 1921-1935. https://doi.org/10.1007/s10980-017-0561-4

Mekonnen, Z.K., Oehlert, J.W., Eskenazi, B. *et al.* (2021). The relationship between air pollutants and maternal socioeconomic factors on preterm birth in California urban counties. *J Expo Sci Environ Epidemiol* 31, 503–51. https://doi.org/10.1038/s41370-021-00323-7

Tao, W.-K., Chen, J.-P., Li, Z., Wang, C., & Zhang, C. (2012). Impact of aerosols on convective clouds and precipitation. Reviews of Geophysics, 50(2). https://doi.org/10.1029/2011RG000369

Vo, T. T., Hu, L., Xue, L., & Chen, S. (2023). Urban effects on local cloud patterns. Proceedings of the National Academy of Sciences, 120(21), e2216765120. https://doi.org/10.1073/pnas.2216765120