PROJECT CHECK-IN

SEASONAL ANALYSIS AND FORECASTING OF AIR QUALITY IN MARICOPA COUNTY

BACKGROUND

- The US Environmental Protection Agency (EPA) has extensive and public data files of air quality data along with an API.
- The data is collected through the EPA ambient air quality monitoring program.
- The measure for air quality used will be the Air Quality Index (AQI) data collected by the EPA’s monitors.
- AQI data is available at the county level and some date back over 40 years.

WHY DO WE CARE ABOUT AIR QUALITY?

- A 2001 study followed 66,820 participants for 10 years to assess the long-term effects of particulate matter on cancer mortality.

- They found that for every 10 microgram per cubic meter (µg/m3) of increased exposure to PM2.5, the risk of dying from any cancer increased by 22 percent.

Chit Ming Wong, Hilda Tsang, Hak Kan Lai, G. Neil Thomas, Kin Bong Lam, King Pan Chan, Qishi Zheng, Jon G. Ayres, Siu Yin Lee, Tai Hing Lam, Thuan Quoc Thach; Cancer Mortality Risks from Long-term Exposure to Ambient Fine Particle. Cancer Epidemiol Biomarkers Prev 1 May 2016; 25 (5): 839–845. https://doi.org/10.1158/1055-9965.EPI-15-0626

PROJECT OVERVIEW

- First the project will aggregate the data across all counties and determine if certain seasons usually have higher AQI values.

- Then seasonal data will be plotted over time to understand the general trends in seasonal AQI values over time.

- Finally, Maricopa County AQI data will be analyzed with a seasonal decomposition and its future AQI values will be predicted with a Seasonal Autoregressive Integrated Moving Average (SARIMA) model.

WHERE IS THE DATA?

- The US Environmental Protection Agency (EPA) has extensive and public data files of air quality data along with an API.
- The data is collected through the EPA ambient air quality. monitoring program
- The measure for air quality used will be the Air Quality Index (AQI) data collected by the EPA’s monitors.
- AQI data is available at the county level and some date back over 40 years.

RELATED WORK

- The EPA publishes their own patterns found from the AQI data they publish, you can review those insights at: https://www.epa.gov/air-trends/air-quality-national-summary

- Additionally, Lung.org publishes their own findings, however, these reports focus on identifying patterns in unhealthy air quality levels using recent data: https://www.lung.org/research/sota/key-findings

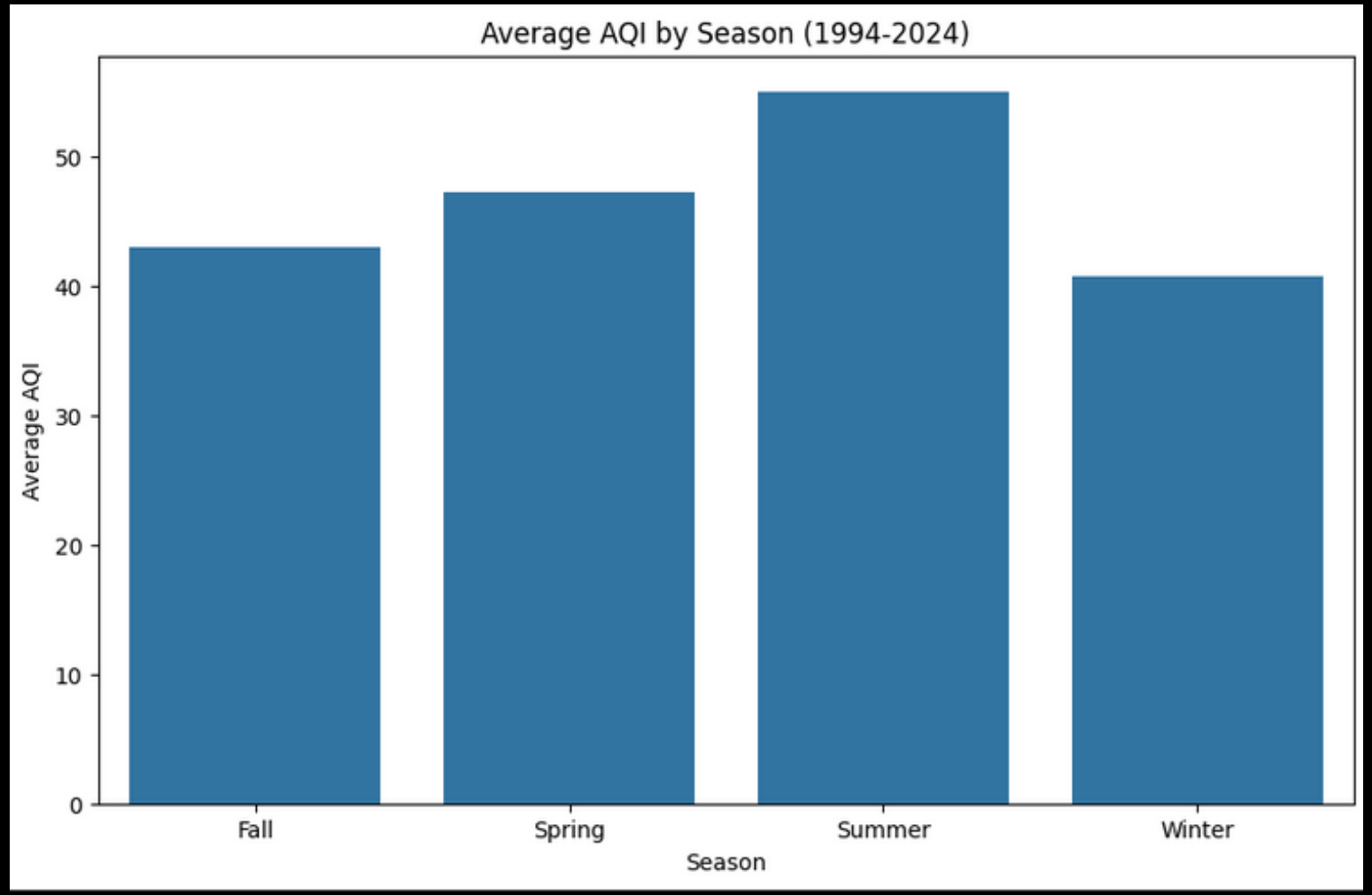
- This project will be focused on using long term data to identify general trends in county level AQI data.

PROPOSED WORK

Using the data from the EPA's data sets this project will aim to analyze the data and provide insights on air quality trends.

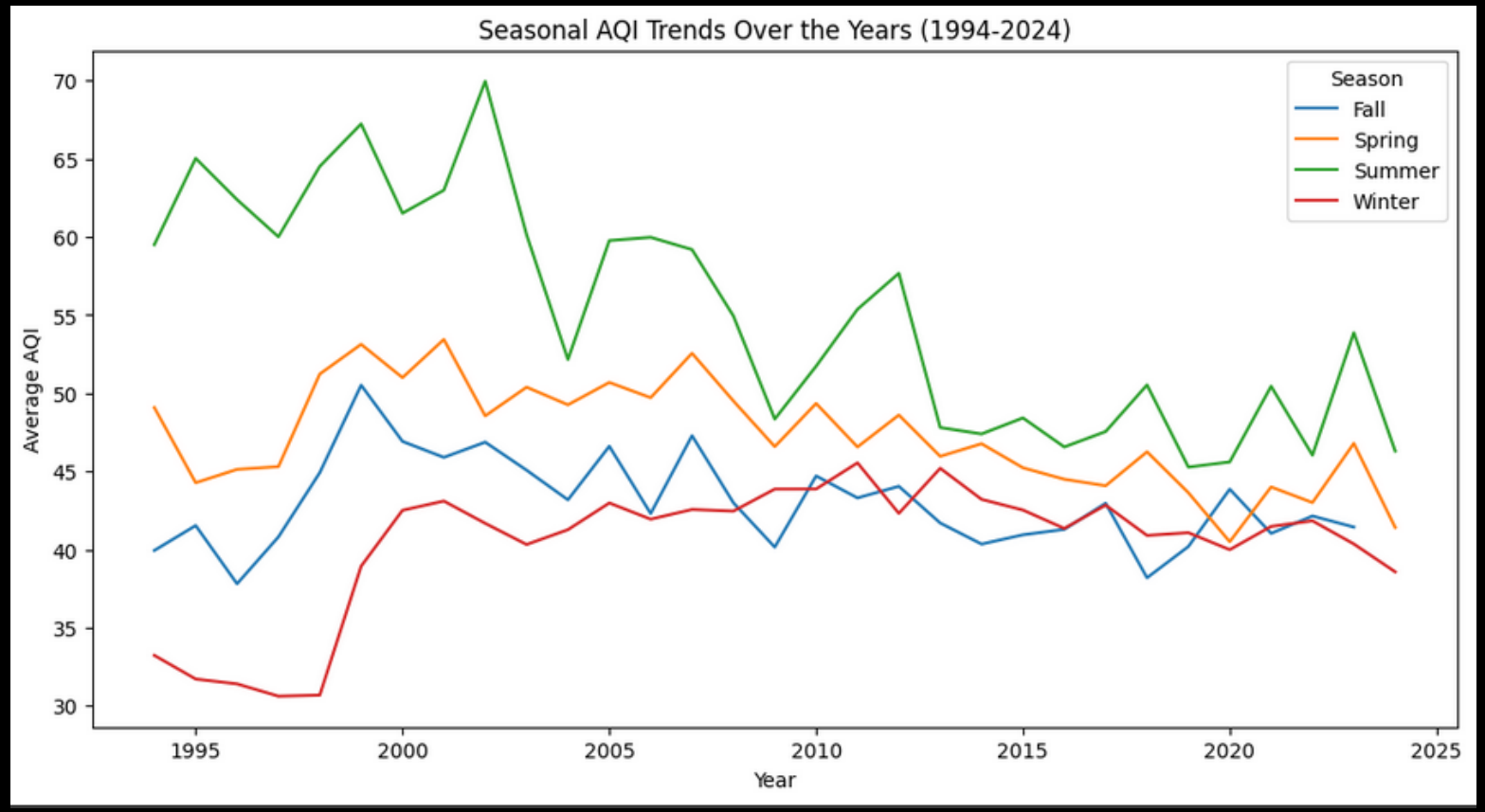
- The data will need to be aggregated and cleaned prior to performing any analysis on it.
- The project will determine if some seasons generally have higher AQI values.
- The AQI averages for each season will be plotted by year to determine seasonal trends in AQI values.
- A SARIMA model will be implemented and evaluated for the prediction of AQI values in Maricopa County.

AVERAGE AQI BY SEASON (1994-2024)



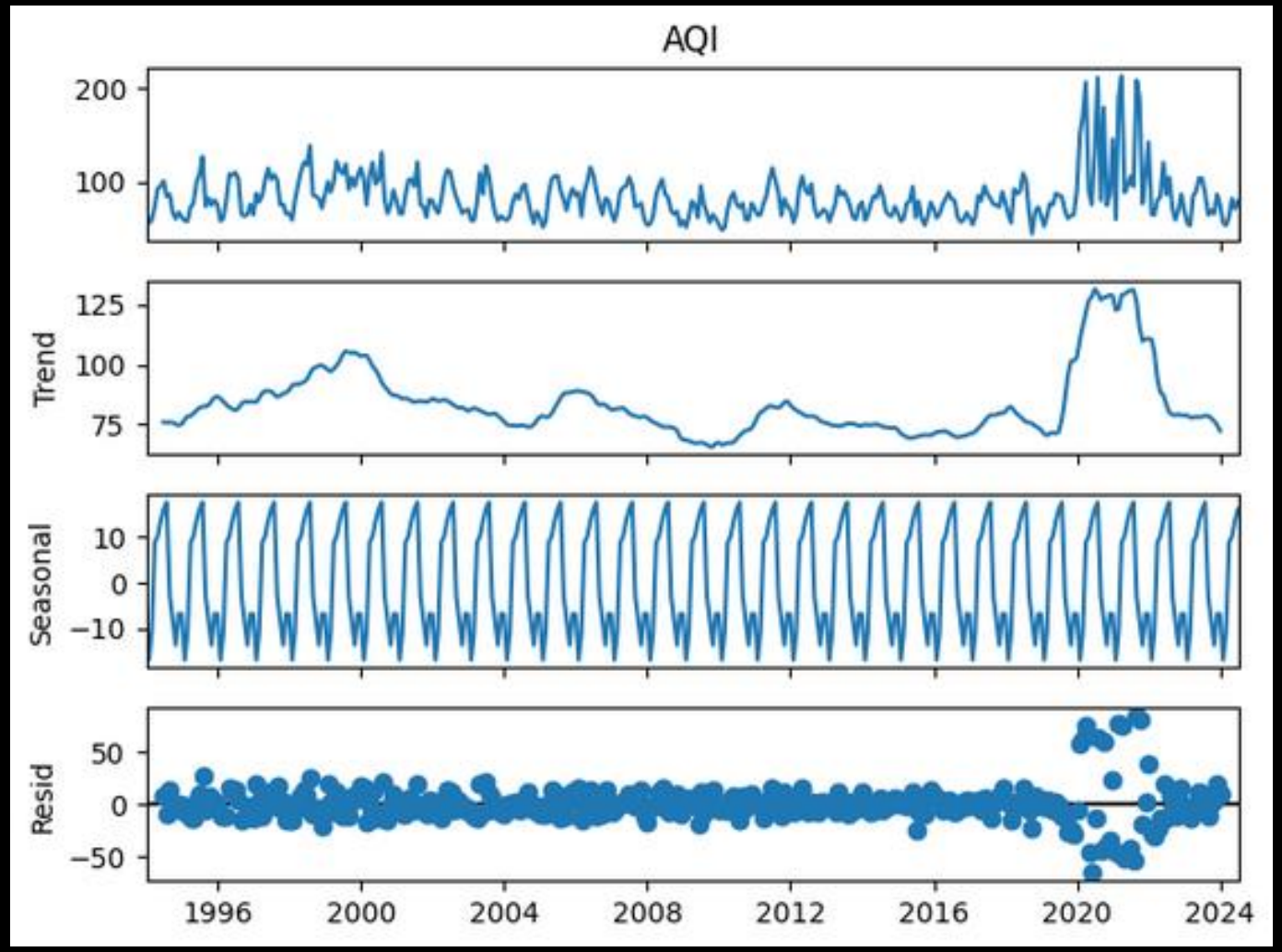
Summer averages AQI values about 22% higher than any other season

SEASONAL AQI TRENDS (1994-2024)



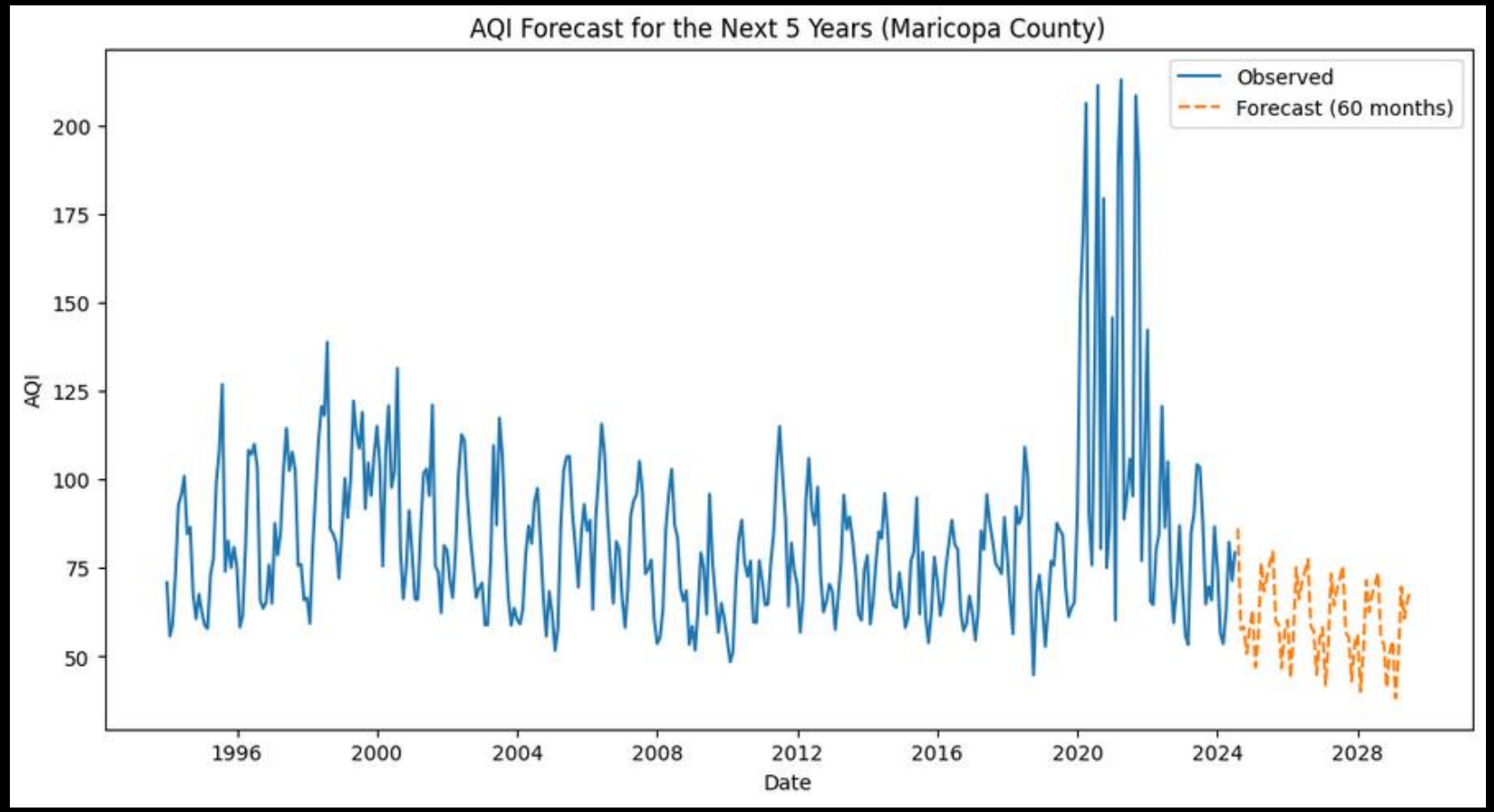
AQI values are vary less across seasons and average AQI values are generally trending downward

SEASONAL DECOMPOSITION PLOT (MARICOPA COUNTY)



The Seasonal plot shows Maricopa county has clear, repeating seasonal trends

SARIMA FORECAST PLOT (MARICOPA COUNTY)



Model predictions for the next 5 years shown in orange

EVALUATION

- The model will initially be evaluated with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

- Our SARIMA prediction model was evaluated against a test dataset for the last 12 months of Maricopa County AQI data.

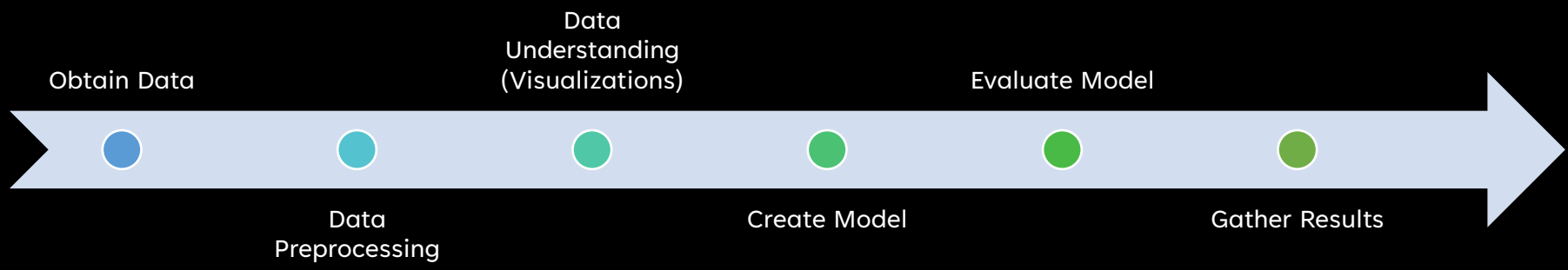
- The test set gave the following values:
    - MAE: 10.56
    - RMSE: 13.38

- The results show that the model is performing reasonably well, since the typical AQI values for Maricopa County range between 50 and 150.

DISCUSSION

- **Potential Challenges:** No challenges are currently foreseen however, as a backup plan, an alternative to the current SARIMA model may be used if model does not pass further evaluation.

- **Project Timeline:**



CONCLUSION

- This project aims to identify seasonal trends in Air Quality and predict future AQI values within Maricopa County.
- A SARIMA time series forecasting model was used to predict future AQI values and seasonal trends in Maricopa County.
- The model was evaluated using MAE and RMSE values and shown to be effective.
- The model will be evaluated using a variety of statistical techniques for fit and significance.

POTENTIAL IMPROVEMENTS

- Incorporate additional variables like weather conditions and industrial activity data to improve the model's prediction accuracy and provide policymakers with more actionable insights.

- Implement the model on all counties in the dataset and create a geospatial map to analyze seasonal AQI variations across different regions.

- Evaluate other machine learning models to determine if they offer better predictive accuracy while still capturing seasonal variations.

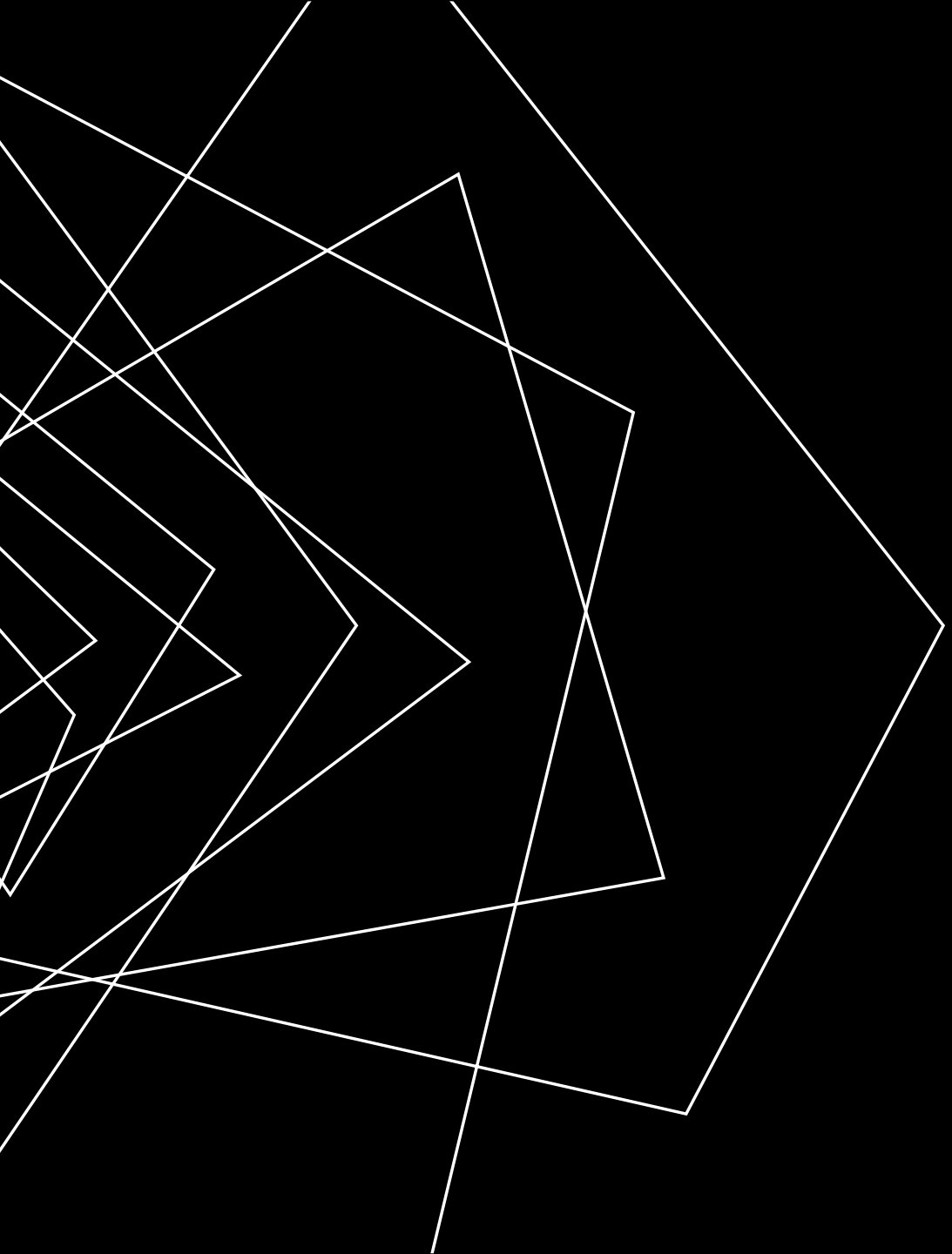
- Improve the seasonal classifier to better capture local anomalies and irregular patterns that might not fit into traditional seasonal definitions.

FUTURE WORK

- Investigate the correlation between AQI values and health data to assess how seasonal pollution affects conditions like asthma or respiratory diseases.

- Explore the relationship between energy consumption patterns and AQI values to develop targeted emission reduction policies.

- Study how the shift to renewable energy sources affects long-term AQI trends and air quality improvements in regions transitioning to cleaner energy.



THANK YOU!