The Last Air Benders

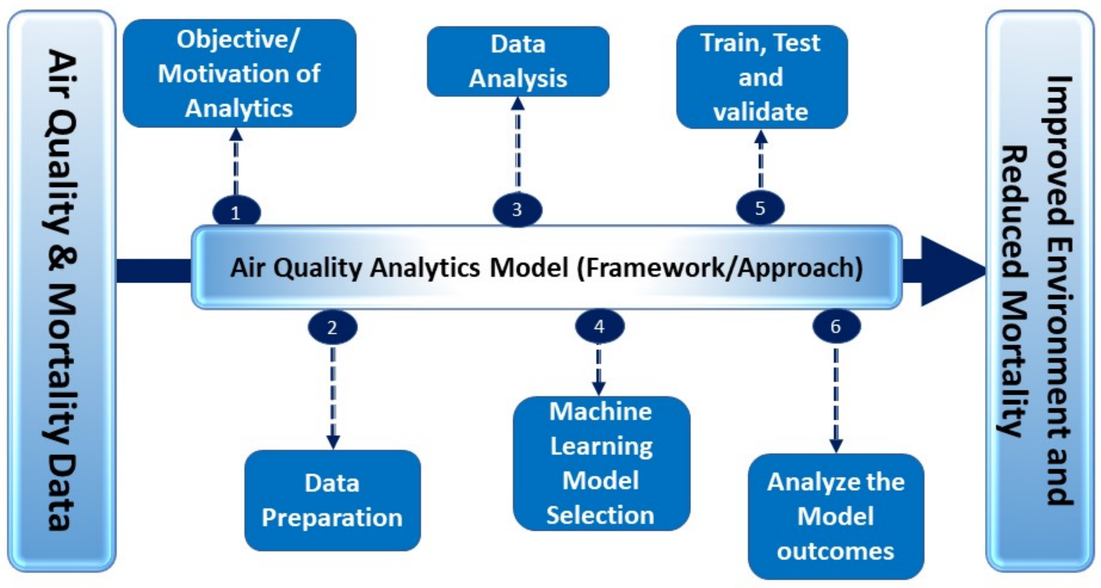
Air Quality and Mortality in the United States

Joseph McAndrews, Colleen Morse, Sunil Prasad, Valerie Schnapp, Jonathan Yang

Summary and Motivation:

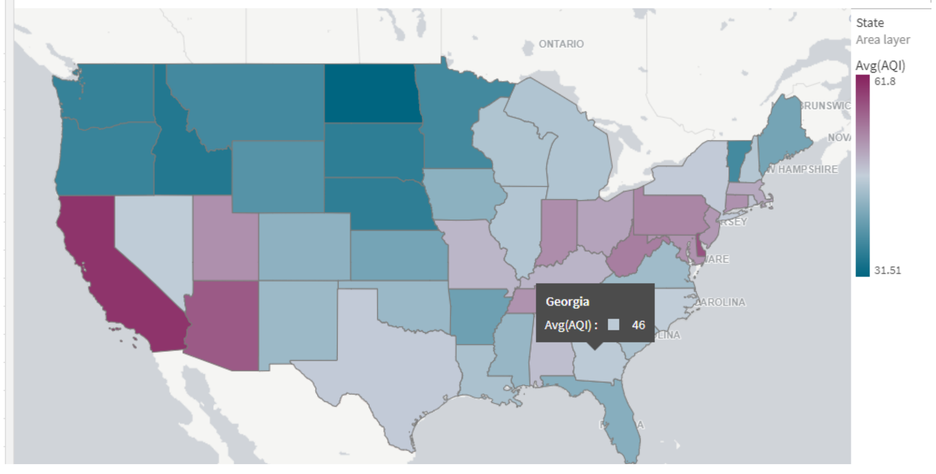
Our project seeks to create a tool to help decision makers in public health and environmental studies quickly and easily see how their data collection efforts coincide and explore the relationship between air quality and mortality. Currently, no online resource exists that shows correlation between air quality and health results.

Concept of Operations:



Our team will follow the above scheme to ingest data, analyze it, and create predictive models to inform decision makers

Visualization:



We used QLIK Dashboards to host visuals. This is an example of a tool tip for AQI showing data for the state of Georgia

Data Sources:

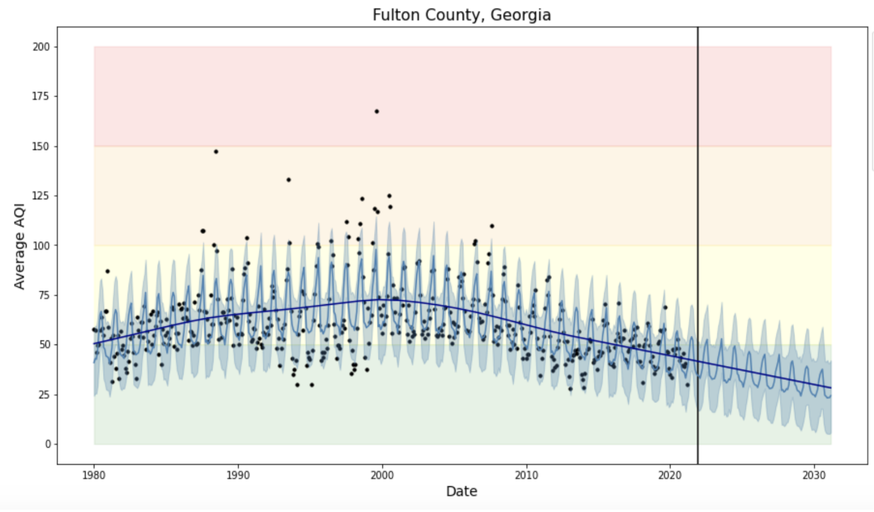
Our data consists of Air Quality Data from the Environmental Protection Agency's Air Quality Index and mortality data from the Centers for Disease Control and Prevention National Center for Health Statistics Mortality Data. Both datasets are freely available to the public on each respective website. Since the data is near 700 MB, we hosted it in the Google Cloud Platform File Browser.

The AQI data included the state name, county name, exact date of data collection, AQI number, AQI category, the pollutant measured, and the latitude and longitude of the measurement. Dates of collection range from January 1, 1980 through May 6, 2021.

The CDC's mortality data includes yearly death counts down to a county level. The CDC Wonder querying app was used to pull both cardiopulmonary-associated death counts and general death counts across each state in the continental US for each year available back to 1980.

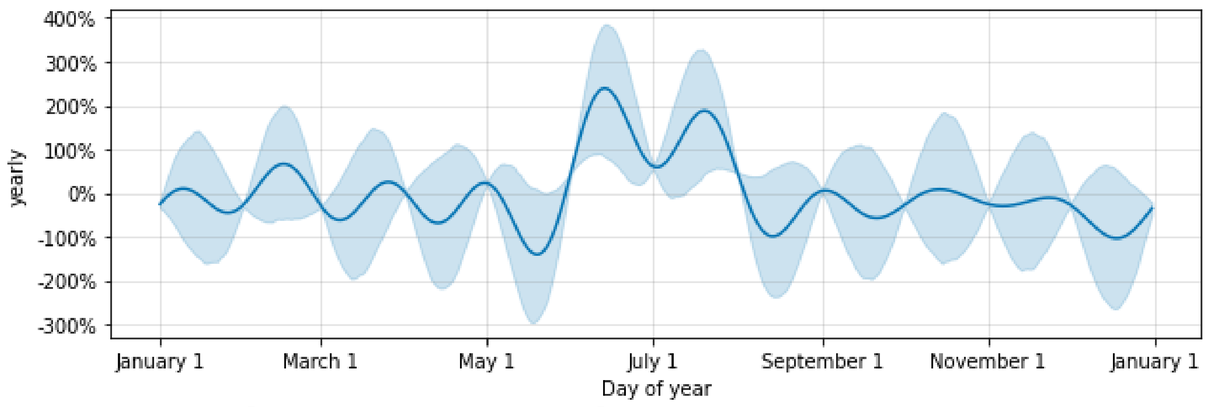
Visual Analysis:

One of our project's main goals was to create many tools to visualize air pollution for all states.

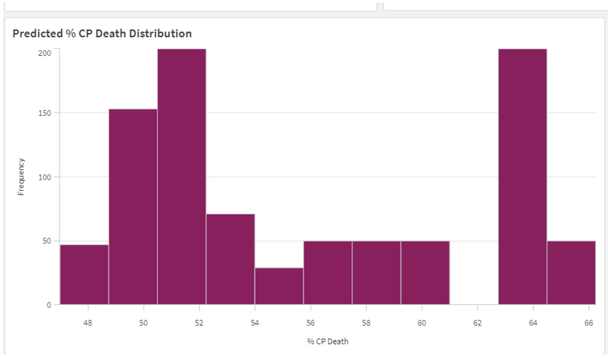


To the left is an example showing the seasonality of air polution predicted future values for the Fulton County, Georgia.

Below is a detailed view of air pollution seasonality, which is very typical across the United States.

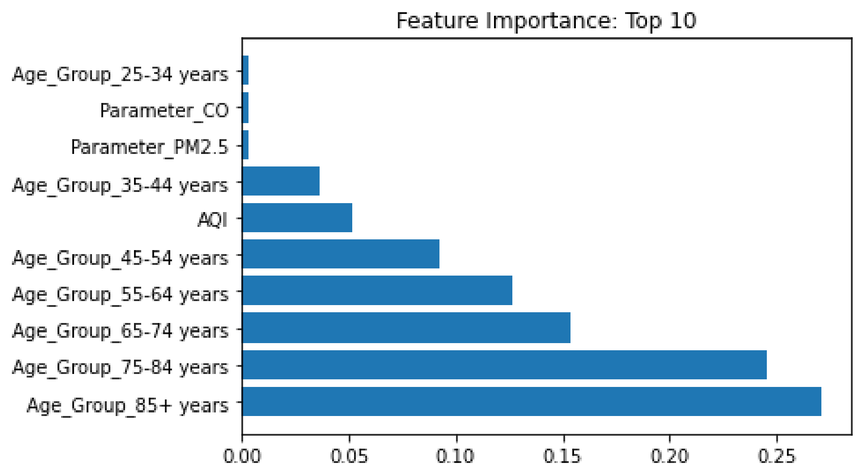


Predictive Analytics:



Above is a predictive model for mortality in the population age 65 and older for all air pollutants in the EPA category of "Unhealthy for Sensitive Groups". Although it seems like this is a strong predictor, we also learned that age was the most important factor. Changing the factors in the model did not lead to a significant change in mortality rate for a given age group.

Results and Evaluation:



Random Forest Regression showed age group was the most significant factor in predicting mortality

The most important factor in predicting mortality was observed to be age - indicating that the correlation between air quality and mortality may not be as strong as we hypothesized.

GitHub link: https://github.com/schnappv/CSE_6242_Project