Spencer Rubin
Springboard Data Science
Capstone 2: Project Proposal

Modeling U.S. Air Quality Pollutants (Ozone) In Major Urban Areas

Problem & Context:
Air quality throughout the U.S. fluctuates based on various environmental, climatic, and anthropogenic features, and varies at both geographic and temporal scales. Air quality measurements include particulate matter, nitrous oxides, sulfurous oxides, and ozone, all which have an impact on human health depending on threshold, quantity, and exposure limits, and that are monitored by either the state or federal EPA. Understanding what features are influencing air quality will allow environmental regulators, city planners, and government agencies opportunities to improve poor air quality by understanding the influence of the changes at a temporal scale in the U.S. Specifically, ground level ozone (O3), a direct byproduct of fossil fuel consumption, plays a major role in human health. How does ground-level air quality fluctuate between various urban areas throughout the U.S.?

Target Market:
The target market of this project includes state and federal environmental agencies such as the EPA and policy-makers, various high-polluting industries (e.g. refineries, agriculture and farms), and urban planners, along with housing developers and the public. Air quality modeling can help manage potential worsening of air quality indices and improve human health through understanding seasonal and temporal fluctuations, and influence management of air quality features influencing these changes such as automobile traffic.

Data:
The datasets for this project will be acquired through the Environmental Protection Agency's Air Quality System (AQS) API, which is freely accessible through their API and available at https://aqs.epa.gov/aqsweb/documents/data_api.html. The data is available in JSON format, and includes, but is not limited to, daily and annual summary data throughout various EPA monitoring stations on significant air quality monitoring features. Data will be acquired for major urban populations by county and specific to ozone measurements.

Constraints & Scope:

- Modeling air quality data based on local weather events (e.g. severe weather), population density, geography.
- Choosing data based on geographic locations for analysis (picking specific areas throughout the U.S)
- Correlating features of air quality with anthropogenic influencers, air pollutant emissions;

Data Analysis Approach:

- Data from the various datasets will be acquired, cleaned, and transformed into a single dataframe using pandas.
- Data will then be fitted to show correlations between air quality and demographic information and weather events
- Modeling of air quality data to further understand features and influencers of air quality throughout major areas of the U.S.

Deliverables:

- Jupyter Notebook to outline: i) the process of acquiring, cleaning, and transforming data; ii) Python code used; iii) data preprocessing and modeling; and iv) data visualization process and code.
- Slide Deck for presentation.
- Summary report of process and findings.