Aidan Wiktorowicz
DS 4002
May 5, 2024


Predicting Air Quality Trends and Identifying Risk Factors Using Data Science

Prompt:

Cities across the globe have increasingly faced deteriorating air quality. Elevated levels of PM 2.5s, O3, and other pollutants pose significant health risks to individuals, including elevated rates of various cancers, asthma, and heart disease. The World Health Organization estimates that over 7 million individuals die prematurely every single year due to air pollution. While emerging markets have garnered significant media attention for bad air days, Western cities are not immune from poor air quality and high levels of pollution. It is, thus, imperative for policymakers to understand the drivers of air pollution and adapt policy to improve health outcomes and quality of life for citizens. Additionally, it is important for citizens and organizations to have accurate pollution forecasts in order to better adapt their schedules, such as workout times and outdoor events, to minimize the health impacts of poor air. By identifying the sources of pollution and providing accurate information on future levels of pollution, cities may be able to reduce pollution and its health/quality of life implications.

Deliverable:

This case study requires you to create a model that can accurately predict daily average levels of pollution. This study also requires you to identify and rank the main drivers of poor air quality, assessing their relative impact. This study requires you to use models, such as the ARIMA model, Multi-linear regression, etc., and encourages the use of data visualizations. The case study will provide you with data from St. Louis, MO, as the United States government collects and publishes significant amounts of data and St. Louis represents an interesting case study due to its central location. The data provided is compiled from the US Federal Reserve, NOAA, and EPA. This dataset includes economic, transit, weather, and air quality data in order to provide comprehensive insight into air quality trends/drivers. This case study challenges you to be creative and enables you to create a unique model. You may use additional data/resources as you see fit. The goal of this project is to create a model that can predict St. Louis's air quality. Optionally, groups may see how their model works on other cities and can train their models on additional city data (so as to not overfit the model for St. Louis). A good project will provide compelling analysis on the drivers of pollution and be able to forecast daily average AQI levels accurately.