Forecasting Greenhouse Gas Emissions

A UVA Data Science Case Study by Adam Cook, Spring 2023

Prompt: Since the industrial revolution and the advent of energy produced by the burning of fossil fuels, humankind has been pumping thousands upon thousands of tons of greenhouse gasses into the atmosphere. These gasses are the single largest contributor to the climate change that is having detrimental effects on our world's natural systems and resources. A rise in global temperatures, sea levels, and extreme weather events, among other things, pose a threat to nature and humans alike. Contemporarily, private organizations and governments worldwide are pledging to reduce their greenhouse gas emissions. The U.S.A., specifically, has set a goal to reduce emissions by 50% from 2005 levels by 2030. It will be your job to use emission data to make a prediction on whether or not this goal will be achieved.

Analysis Tips: To address this prompt, you will use the "CO2 and Greenhouse Gas Emissions" dataset from Our World in Data. This is a very large dataset, so you will need to clean it down to only the data you need: Greenhouse gas emissions in the United States. It is up to you if you want to keep additional columns to analyze. I recommend using the dplyr package in R to extract the data you need out of the large dataset. To then use this data to make a projection for 2030, I recommend using an auto-regressive integrated moving average (ARIMA) model. Helpful resources to assist in using dplyr and ARIMA can be found in the materials document.

Deliverable: Using this prompt and the tips and materials provided, you will create an annotated markdown document in which you:

- Import and clean the data
- Run the model
- Draw Conclusions

You will then use the work you did in this document to create a github repository and a presentation. Details regarding these two products can be found in the rubric.