

Capstone Project Public Submission Proposal

What is the problem you want to solve: Germany has spent more than \$220 billion on Renewable energy forms such as Wind and Solar, but only modest CO2 reductions have been observed. What might be going in to allow this, and is there a better way?

This Capstone project will seek to rigorously quantify using modern Data Science techniques a longstanding and highly controversial question, namely *“To what extent are the benefits from carbon-free forms of Renewable energy, such as Wind and Solar, offset by the extra carbon emitted by fossil fuel plants that are required to back-up in a highly inefficient manner the inherently intermittent nature of the aggregate power generation from such Renewables?”*

To put the issue in context, Environmental groups claim there is minimal such offset, but some critics claim the offset is almost 100%: the experience of Germany with Renewables has been cited recently as strongly supporting the latter viewpoint. This topic may be too much of a “hot potato” for the academic and NGO communities, who may be censured for appearing to question the conventional wisdom that Renewables are the only way to go. But it is nevertheless a question of fundamental importance and may imply that other alternatives should also be pursued.

The question will be investigated using a voluminous plant-by-plant database for power generation in the US that is published annually by the EIA. A formal “paired t-test” will be done to evaluate whether a statistically significant (and offsetting) increase in the carbon intensity of fossil fuel plants in Texas can be discerned, over a seven year period during which power output from Renewables has grown strongly in Texas.

Incidentally, this course is on Python for Data Science, and about 80% of a data scientist’s time is reportedly spend extracting and cleaning data, which Python is excellent for. Extracting the required information from the EIA database is IMHO non-trivial for a non-expert. In other words, the Machine Learning and statistical inference aspects may be less challenging than extracting the data in the first place. As Amos Tversky, with whom I had studied while at Stanford, had declared, the so-called “*Availability*” bias is the second most common bias in making decisions under uncertainty. As humans we tend to err on the side of using the data that is most readily available, rather than painstakingly seeking out the most relevant data which may be more difficult to acquire. Being agile in concisely extracting the most salient data from a messy and convoluted dataset is therefore of considerable practical value.

Deliverables:

1. A paired t-test on (hundreds of) fossil fuel plants in Texas to see if their carbon intensity has risen in a statistically significant manner over the last seven years. This analysis can very easily be extended to other states, including California.
2. A histogram of the change in carbon intensity of said fossil fuel plants.
3. A calculation of the amount of the carbon offset, if different from zero.
4. A linear regression model that seeks to quantify factors effecting the efficiency change in fossil fuel plants.
5. Commentary on the economic implications.

Parties of Interest: all states and countries which have an interest in most efficiently integrating Renewables with fossil fuel plants, including California, the Midwest, Hawaii, Puerto Rico, China, Germany, other European countries, and India. Also, consulting firms that advise on energy matters, such as BCG, McKinsey, Bloomberg New Energy Finance, and many others.

What might be done differently as a result of this Project : Depending on results, the development of Renewables might be accelerated or slowed. Either way the cost of reducing the overall carbon footprint from power generation should be potentially reduced as a result of the insights gained from this work.

Future Work: An entirely separate question is whether there is a much cheaper and lower-carbon way of dealing with the intermittent nature of Renewables’ power generation than by using fossil-fueled plants in a highly

inefficient manner. One particular such avenue will be explored in a follow-up data science project. Many other detailed projects would also be pertinent.