Capstone project 1: Fossil fuel economies in the United States

**What is the problem that you want to solve?**

Typically, the debate over fossil fuel usage revolves around climate change issues. However, fossil fuels continue to be an important part of the US economy. Energy independence is one of the top goals for the current US administration, and was one for the last administration as well.

Different parts of the country depend on fossil fuel production in different ways - for some states and counties, it is their primary source of livelihood. When the oil price drops, people living in these areas are very adversely affected. In other states, for example in California, fossil fuels are not that important to the economy, except as a commodity to be used.

The hypothesis I am working with is that many counties and states around the US continue to be deeply dependent on fossil fuel production and use. A starting point for this was the Enigma Labs project: [Enigma Labs: Boom to Bust](https://labs.enigma.com/north-dakota-boom-to-bust/). This project centered around the effects of the oil price crash on North Dakota, and correlated the price crash with the unemployment rate in the state.

Another dimension to this problem is about public spending on infrastructure projects. If industries in a given country/state are doing well, it should follow that they have more money for big projects. This spending should go down in bad times, and reflect in the state/county’s spending on bridges, roads and other public works.

The most important part of this problem is trying to answer the question on how the US, and subsequently the world, will manage a transition to a carbon-free, or carbon-neutral economy. With the rapid advances being made in battery technology and electric cars, these questions affect the futures of more than a few entrepreneurs. With this project, I am trying to guess what could happen to a subset of the US economy.

**Who is your client and why do they care about this problem?**

My client could be from a number of different industries. As is evident from the study on North Dakota, oil price crashes deeply impact unemployment rates in states which are the most dependent on fossil fuels. The same is true for other kinds of fossil fuels like coal and natural gas. The first obvious application of such a project would be for planning a state or county’s budgets around these outcomes. So individual state governments could be clients.

There are many other industries which indirectly are affected by the fortunes of the fossil fuel industry - since everyone relies on availability of cheap energy. These include, and are not limited to transportation, infrastructure, and manufacturing. Investors in these industries, or analysts looking for exposure of companies’ to these sectors could also be potential clients. Regulators for each of those industries could also potentially be clients.

**What data are you going to use for this?**

The Enigma Labs project is a start-off point, which provides unemployment & oil production data for North Dakota. Similar, and related datasets are available through public sources:

1. Oil and gas production by state: <https://public.enigma.com/browse/oil-gas-production/805e7d28-8477-4be1-a8a5-5967e7edf454>
2. Top industries in each county in the US: <https://www.ers.usda.gov/data-products/county-typology-codes.aspx>
3. Studies on infrastructure in the US:

[Washinton Post article on US infrastructure](https://www.washingtonpost.com/graphics/national/maps-of-american-infrastrucure)

<https://hifld-dhs-gii.opendata.arcgis.com>

<https://blogs.esri.com/esri/esri-insider/2016/02/24/open-data-for-economic-resiliency/>

d. Rig counts are usually taken as a macro-indicator of the health of the oil industry:

<http://phx.corporate-ir.net/phoenix.zhtml?c=79687&p=irol-rigcountsoverview>

e. Other potentially relevant datasets:

i) Industrial sector data: <http://pages.stern.nyu.edu/~adamodar/New_Home_Page/>

ii) Building permits: Building permits:

<https://www.census.gov/construction/bps/definitions/>

<https://www.census.gov/construction/bps/how_the_data_are_collected/>

<https://tinyletter.com/data-is-plural/letters/data-is-plural-2016-06-22-edition>

iii) Interstate commodity flows : <http://faf.ornl.gov/fafweb/>

https://www.axios.com/the-flow-of-goods-2463665414.html

**Outline your approach to solving this problem.**

My initial approach will be to identify all the different variables which could be relevant to the problem. That is why I am trying to work with very disparate datasets, which could lead to new insights for the problem.

Having found these variables, I would like to do sensitivity analyses for the influence of various factors, like oil price, regulations on carbon emissions etc, on the livelihood of people affected.

**What are your deliverables?**

1. A new, clean dataset, which should incorporate all the different variables important for this problem. Currently, there doesn’t exist a dataset which combines fossil fuel production/use by county/state, and correlates it with other macroeconomic parameters like unemployment, infrastructure spending, state GDP.
2. Testing my hypotheses, namely that wide swathes of the US economy continue to be deeply intertwined with the fate of fossil fuels.
3. Using statistical models, testing what effect oil price crashes, or a general commodity price crash has on the state’s spending, and wider macroeconomic parameters.
4. If possible, to try and answer the question about how to manage the transition to a carbon-free or carbon-neutral economy, without causing too much pain to the people affected adversely. I think this is a much harder question, and may not be possible to answer given the time constraints.