

Finalized Project Proposal

Basics

Team Lead: Maxwell Tuttle

Recorder: Melissa Best

Spokesperson: Kevin Puduseril

Background & Question

Research Question:

What combination of state characteristics (political leaning, grid capacity, median income, existing energy mix, etc.) and renewable energy policy designs (Renewable Portfolio Standard targets, subsidy types, etc.) best predicts a successful change in fossil fuel shares per state in the U.S.?

Motivation:

As climate concerns grow, U.S. states are progressively adopting policies to lower their reliance on fossil fuels. However, the effectiveness of these policies vary considerably. Some states see

quick adoption of renewables, while others grapple to make meaningful transitions. This inconsistency suggests that success is affected by more than policy alone — foundational state characteristics may play a key role.

Need/Niche:

Present-day literature (Carley, 2009; Fischlein & Smith, 2013; Shrimali & Kniefel, 2011) frequently evaluates the success of energy policies in isolation (e.g., Renewable Portfolio Standard effectiveness), but relatively few studies examine the interaction between policy design and structural state-level features (like income or political climate). Our study fills this niche by recognizing which combinations of conditions and policies are most predictive of fossil fuel displacement success.

Why This Question Matters:

The findings can guide policymakers and other interested parties in creating realistic, context-sensitive energy policies. By comprehending the conditions under which fossil fuel displacement is most successful, states can avoid one-size-fits-all strategies and take on tailored approaches.

Novelty & Originality:

Even though some studies analyze Renewable Portfolio Standard targets or clean energy adoption singly, few blend multi-dimensional state-level predictors with policy features in a

comparative model across all 50 states. This project is original in its comparative approach using predictive modeling and machine learning tools to expose key drivers.

Hypothesis & Prediction

Hypothesis:

States with supportive political climates (e.g., liberal leaning), high grid capacity, higher median incomes, and ambitious, well-designed renewable energy policies (e.g., binding Renewable Portfolio Standard with financial incentives) will be more successful in displacing fossil fuel energy sources.

Prediction:

A machine learning classifier or regression model will show that the above features are significantly and positively associated with a greater percentage-point reduction in fossil fuel energy share over the past decade, compared to other states.

Data & Analysis

1. U.S. EIA “Net Generation by State by Type of Producer by Energy Source”

- We will use the **U.S. Energy Information Administration (EIA)**’s “Net Generation by State by Type of Producer by Energy Source” dataset. It provides detailed annual electricity generation (in MWh) by energy source (e.g., solar, wind, coal, natural gas, etc.) from 1990 to the most recent available year.
- **Link:** <https://www.eia.gov/electricity/data/state/>
- **EIA-923 Power Plant Operations Report (released: 10/4/2024)**
- Net Generation by State by Type of Producer by Energy Source (EIA-906, EIA-920, and EIA-923)¹
- *Date range:* 1990 – 2023
 - *Available formats:* [XLS](#)
- This dataset is ideal because it offers granular, multi-decade data on generation by energy type and geography, allowing for a robust longitudinal state-level analysis.

2. U.S. EIA State Energy Data System

- Use: Annual energy generation mix by source (coal, natural gas, renewables, etc.) for each U.S. state
- Main Site: <https://www.eia.gov/state/seds/>

- Direct Download: <https://www.eia.gov/state/seds/seds-data-complete.php>
- [Use_all_btu.xlsx](#)
 - This file contains total energy consumption by fuel type, sector, state, and year, measured in billion BTUs.
- [Codes_and_Descriptions.xlsx](#)
 - This file defines the MSN codes used in the main data file so we can identify
 - What each row represents
 - Which fuel and sector it corresponds to
- U.S. Energy Information Administration. *State Energy Data System (SEDS)*. Retrieved from <https://www.eia.gov/state/seds/>

3. DSIRE (Database of State Incentives for Renewables & Efficiency)

- Purpose: State-level policy data - Renewable Portfolio Standards (RPS), tax credits, subsidies, etc.
- Main site: <https://www.dsireusa.org/>
- State policy search: <https://www.dsireusa.org/resources/detailed-summary-maps/>
- North Carolina Clean Energy Technology Center. *Database of State Incentives for Renewables & Efficiency (DSIRE)*. Retrieved from <https://www.dsireusa.org/>

4. U.S. Census Bureau - American Community Survey (ACS)

- Purpose: Socioeconomic characteristics like median household income, education levels, etc.
- Main portal: <https://data.census.gov/>
 - Use search terms like:
 - “Median Household Income by State”
 - “Educational Attainment by State”
 - Data found at:
<https://data.census.gov/table/ACSST1Y2010.S1901?q=median+household+income+by+state>
- U.S. Census Bureau. *American Community Survey 5-Year Estimates*. Retrieved from <https://data.census.gov/>

5. MIT Election Data and Science Lab

- Purpose: Contains data for political voting of states for the year range of 1976-2020
- Site:
<https://dataverse.harvard.edu/file.xhtml?fileId=10244938&version=8.0&toolType=PREVIEW>

- MIT Election Data and Science Lab. *State Presidential Election Returns 1976–2020*. Harvard Dataverse,
<https://dataverse.harvard.edu/file.xhtml?fileId=10244938&version=8.0&toolType=PREVIEW>

6. NREL (National Renewable Energy Laboratory) Grid Data & Interconnection

- Purpose: Information on grid capacity, interconnection queues, and renewable integration readiness
- Site: <https://www.nrel.gov/grid/>
- Wind Integration National Dataset (WIND Toolkit)
 - Provides state-level (and more granular) data on wind power potential and capacity
 - Download the “Techno-Economic Summary and Index” CSV from the NREL data catalog—it contains grid-ready data for ~ 120,000 points across the U.S., including hub-height wind generation potential by location and state
 - Further details and aggregated statistics can be accessed via the Wind Toolkit Application, NREL APIs, and AWS-hosted global bulk datasets.
- Solar Integration National Dataset Toolkit
 - Offers simulated output from distributed and utility-scale solar plants

- Navigate to NREL's Solar Power Data for Integration Studies and download individual state-level CSV files (either actual, day-ahead, or 4-hour ahead forecasts) for your state of interest
- These files include simulated power (MWh) by location and capacity, which we can aggregate by state.
- Grid Infrastructure & Capacity Indicators
 - While NREL does not provide a single “grid capacity by state” file, we can derive indicators from:
 - NREL's Grid Modernization Data & Tools, accessible via the Grid Modernization portal, including transmission line maps and data layers — may require geospatial queries or CSV exports
 - The National Climate Database (NCDB) API and AWS datasets may provide regional climate and grid infrastructure overlays
- National Renewable Energy Laboratory. *Interconnection Queue Analysis Tool* and *Grid Modernization Data*. Retrieved from <https://www.nrel.gov/grid/>

Key Variables:

- Outcome Variable:
 - % reduction in fossil fuel energy share (2010 vs. 2022)

- **Predictor Variables:**

- **Structural:**

- Political Leaning
 - Median Household Income
 - Existing energy mix (baseline % fossil, wind, solar)
 - Grid capacity / infrastructure index

- **Policy-Related:**

- Renewable Portfolio Standard (RPS) presence (binary or numeric)
 - RPS strength (target %, year)
 - Subsidy types (grants, tax credits, rebates)
 - Policy start year (or duration)
 - Penalty/enforcement mechanism presence (binary or numeric)
 - Presence of other policies (binary or numeric)

Analysis Plan:

To start our analysis, we will conduct comprehensive data cleaning and feature engineering.

This includes merging all datasets by U.S. state to ensure a coherent unit of analysis across data sources such as energy mix, socioeconomic variables, and policy features. We will

calculate the percentage-point change in fossil fuel energy share between 2010 and 2022 for each state, which will serve as the core metric of fossil fuel displacement. From this, we will

derive a binary outcome variable labeled “success” defined as any state exhibiting a reduction in fossil fuel share greater than the national median change. Additionally, we will encode

categorical policy variables (e.g., types of subsidies, presence of enforcement mechanisms) into suitable formats for analysis.

Following data preparation, our modeling approach begins with exploratory data analysis (EDA). We will use correlation matrices and visualizations like box plots to investigate preliminary relationships between key variables such as political leaning, policy types, and fossil fuel reduction. These EDA steps will help inform variable selection and modeling strategy.

For predictive modeling, we plan to employ both logistic regression and tree-based methods. Logistic regression will be used to classify states as either successful or not in terms of fossil fuel displacement. To gain deeper insights into feature importance and interactions, we will apply ensemble learning techniques such as Random Forests and XGBoost. We will also do further analysis on feature importance to understand what aspects will affect the success of a state.

Our evaluation strategy will depend on the model type. For classification tasks, we will assess performance using accuracy, precision, and ROC-AUC metrics. If regression models are used to predict the magnitude of fossil fuel share reduction, we will evaluate them using R-squared (R^2) and root mean squared error (RMSE). Ultimately, support for our hypothesis will be considered strong if we observe statistically significant performance and if features such as political leaning, income, grid capacity, and renewable policy strength emerge as top predictors in the models.

Technical Details

Data Cleaning & Engineering

- Merge datasets by state
- Calculate change in fossil fuel share:

$$\Delta\text{FossilFuelShare} = \text{FossilShare}_{2010} - \text{FossilShare}_{2022}$$

- Create binary “success” outcome:

$$\text{Success} = 1 \text{ if } \Delta\text{FossilFuelShare} > \text{median}(\Delta\text{FossilFuelShare})$$

- Encode categorical policy features (e.g., subsidy type, enforcement mechanism)

Modeling Approach

- Exploratory Analysis:
 - Correlation matrices, box plots by policy types and political leaning
- Predictive Modeling:
 - Logistic Regression (for binary success)
 - Random Forest / XGBoost for feature importance

Evaluation Criteria

- For classification: Accuracy, Precision, ROC-AUC

- For regression: R^2 , RMSE
- Support for hypothesis:
 - Statistically significant model performance
 - Political leaning, income, grid-capacity, and Renewable Portfolio Standards policy among top predictors

Tools

- Python (Pandas, scikit-learn, matplotlib, seaborn)
- Google Colab for EDA and modeling
- Git for version control
 - GitHub Link: <https://github.com/TaxMuttie/DSE6311>

Sources:

1. Carley, S. (2009). State renewable energy electricity policies: An empirical evaluation of effectiveness. *Energy Policy*, 37(8), 3071–3081.
<https://doi.org/10.1016/j.enpol.2009.03.062>
2. Fischlein, M., & Smith, T. M. (2013). Revisiting renewable portfolio standard effectiveness: Policy design and outcome specification matter. *Policy Sciences*, 46(3), 277–310. <https://doi.org/10.1007/s11077-013-9172-1>

3. Shrimali, G., & Kniefel, J. (2011). Are government policies effective in promoting deployment of renewable electricity resources? *Energy Policy*, 39(9), 4726–4741.
<https://doi.org/10.1016/j.enpol.2011.06.055>