

Project Paper

Watts Up Team Members:

Frank Vasquez – frankvasquez7

Jacob Trout – jacobtrout

Praveen Devarajan – pravchand

Xiaoyue Wei – Wxy-23

Project Abstract:

This project analyzes trends in the energy generation, price and carbon emission across states within the United States. More specifically, we aim to gather data on the composition of sources of energy generation and energy prices in each state, and then explore the trends in pricing, production and carbon emission over time. We also construct maps of the US to support detailed visualization energy generation of each plant by geography. We also create a diverse visualization including line plot, bar chart, tree map, to present our analysis of trends, and we create a prediction model to predict the year for each state to reach the goal of generating 60% renewable energy among total energy generation .

Overall Structure of the Project:

The project is structured into 4 main parts:

1. Data collection and this stage are presented in the “data” folder, which is divided into “raw data”, “intermediate data” and “final data”.

- The “raw data” folder contains the 4 raw data sources :
 - a. Energy protection and carbon emission data: in the form of Excel files downloaded from the United States Environmental Protection Agency(EPA)
 - b. Energy price data: API dashboard from the US. Energy Information Administration(E.I.A)
 - c. Population and GDP data: in the form of CSV files downloaded from the U.S Census Bureau and U.S. Bureau of Economic Analysis (BEA).
- The “intermediate data” folder contains a file necessary to run the data cleaning program to create the final database.
- The “final data” folder contains the SQL database, plant.db, which integrates energy generation, carbon emission, and energy price data.

2. Data preprocessing and this stage is presented in the “data processing” folder, which is divided into 3 folders: “clean_data”, “extract_data” and “load_data”.

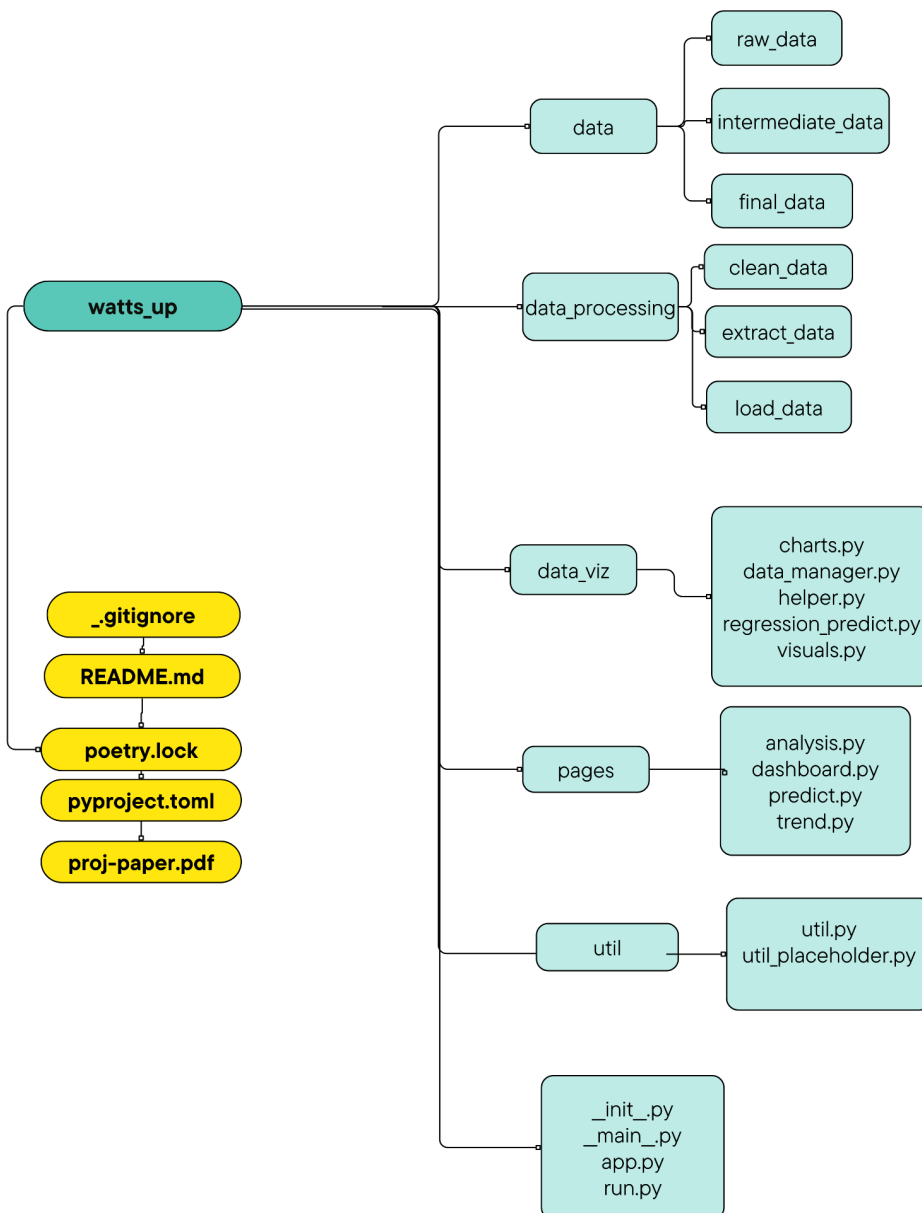
- The “clean_data” folder has files for cleaning all 4 data sources.
- The “load_data” folder has files presenting the functions to create the plants.db SQL database.

- The “extract_data” folder contains one script that converts the EPA data from excel to JSON and combines the files. The script also contains a function to extract the energy price data from the API.

3. Data visualization, and this stage is presented in the “data_viz” and “pages” folder, which contains files to create the final visualization of a dashboard with 4 pages and the visualization of each page.

4. Program execution: This stage is presented in app.py and run.py.

Project folder structure:



Responsibilities of each member:

Name	Model	Task	Files
Frank	Data visualization, Data analysis, Project folder design Poetry virtual environment GITHUB	<ul style="list-style-type: none"> - Centralized data visualization through app.py, enabling pagination. - Created analysis, and visualized the prediction page for energy production. - create helper functions to enable data visualizations. - Github management - 	Pages Folder: analysis.py, predict.py Data Viz Folder: Helper files for data visualization. app.py
Jacob	Data processing, Data cleaning, Project folder design Poetry virtual environment Document writing	<ul style="list-style-type: none"> - Download energy generation data, and identify the changes in data structure in the excel files over the years and columns needed for the project. - Clean energy generation and emission data, join price population and GDP data to create the plants.db sql database - Compiled all functions, scripts, and modules in order to get 'watts_up' to run from the command line as an application. 	Data folder Import_data.py Clean_data.py Make_db.py Schema.py Util.py README.md Run.py
Praveen	Data scraping, Data cleaning, Data analysis and Viz	<ul style="list-style-type: none"> -Pull energy price data from EIA API, cleaned and stored it. -Pulled GDP and POP data, cleaned and stored it. - Wrote schema for sqlite3 for the above data tables - Wrote a data analysis file to predict when states hit x% renewable energy in their mix - Wrote the chart file to create line and treemaps - Wrote the trends file to display the charts from the chart file to the DASH app. The charts include a line chart and a tree map with dropdowns. - Wrote a labelling function to label rows based on the type of energy they generate 	trend.py regression_predict.py charts.py Joint author: Import_data.py Clean_data.py Make_db.py helper.py schema.py
Xiaoyue	Data processing, Data visualization Communication Coordination Document writing	<ul style="list-style-type: none"> -Download energy generation data, and identify the changes in data structure in the excel files over the years and columns needed for the project. -Pull data from the sql database and 	dashboard.py proj-paper pdf.

		<p>generate pandas data frames for making a trend plot dashboard.</p> <p>-Wrote functions to create a dashboard that generates trend plots for each state showing the trends of the energy generation, price and carbon emission over the years.</p> <p>-Coordinated with TA and facilitated group meetings.</p> <p>-Wrote the proj-paper.</p>	
Team Effort	Research, Project Design	Research available data sources and topics, draft outline for the project	

Application Guide

There are 4 pages in the final product dashboard: Analysis, State Trends, National Trends and Projections. Users can navigate to each page through the header section of the dashboard.

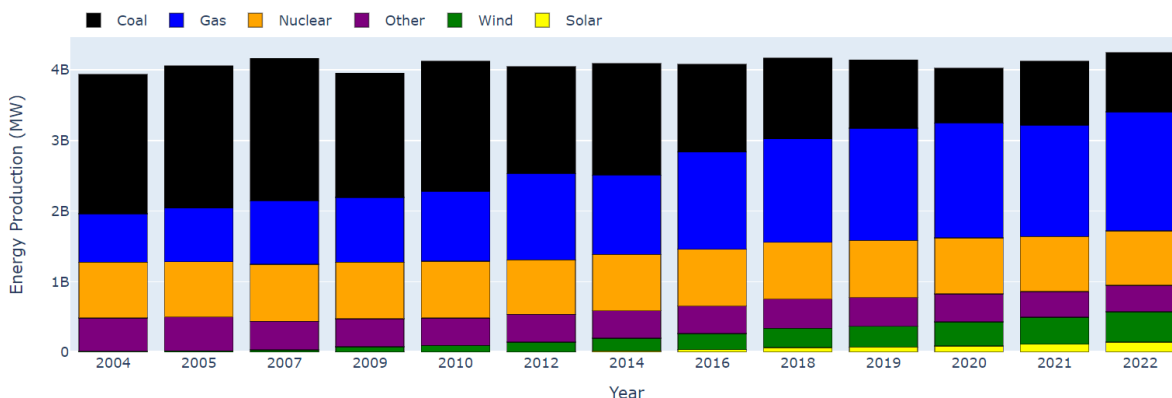
Watts Up

CAPP-30122 Project - E-Grid Data Visualized

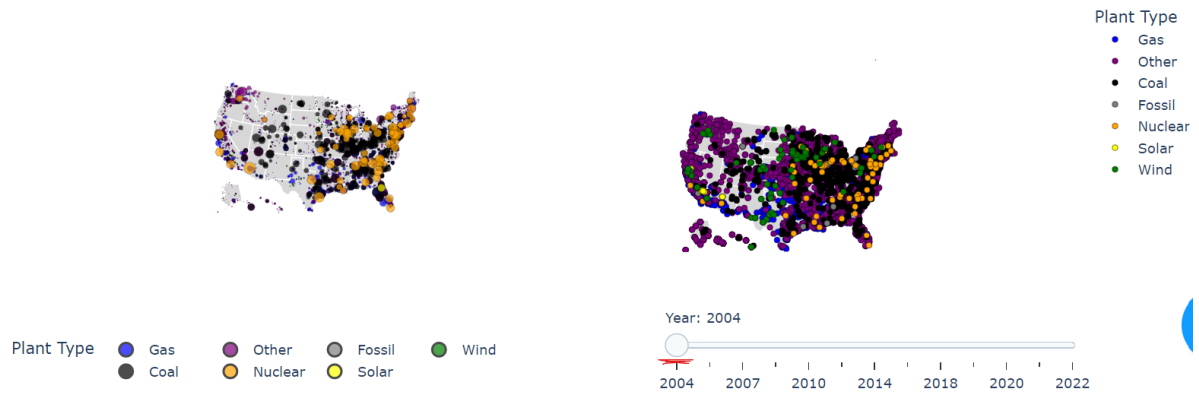
Analysis State Trends National Trends Projections

1. **Analysis:** This page also presents a summary bar chart of energy production by energy generation type of each year from 2004 to 2022.

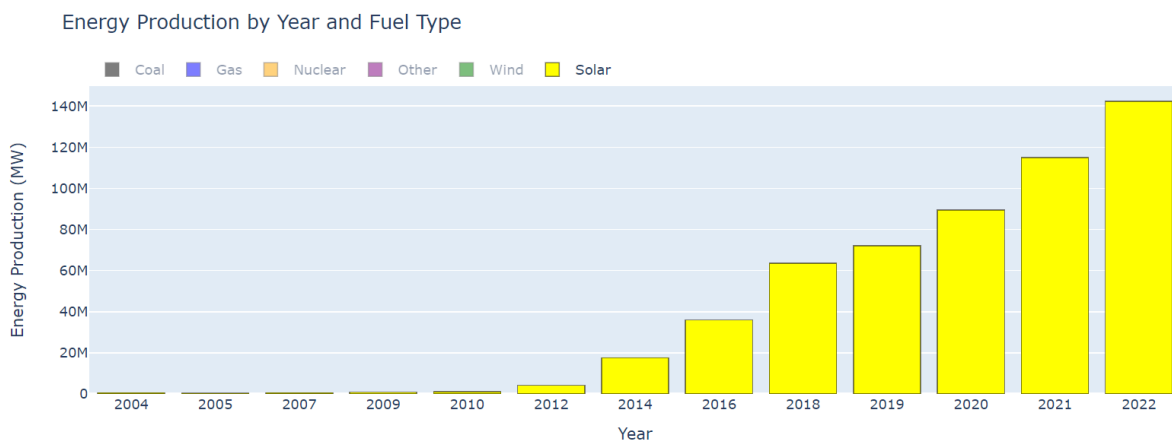
Energy Production by Year and Fuel Type



Users can choose a target year to create two maps presenting the plant based energy generation data across the US with each circle presenting a single plant and the size of the circle presenting the corresponding value. The one on the left side presents the change in total generating capacity of each plant compared to the previous year. The map on the right is an interactive map which allows users to scroll through the timeline to changes of plant locations by their energy generation types over time.

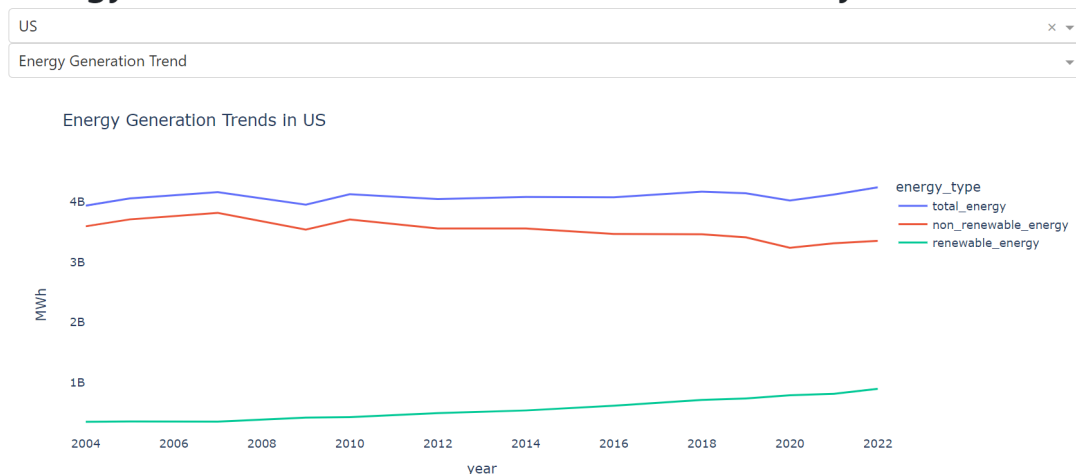


When double clicking the energy type, it will filter out other energy types and present the total national energy generation from 2004 to 2022 of the selected energy type.



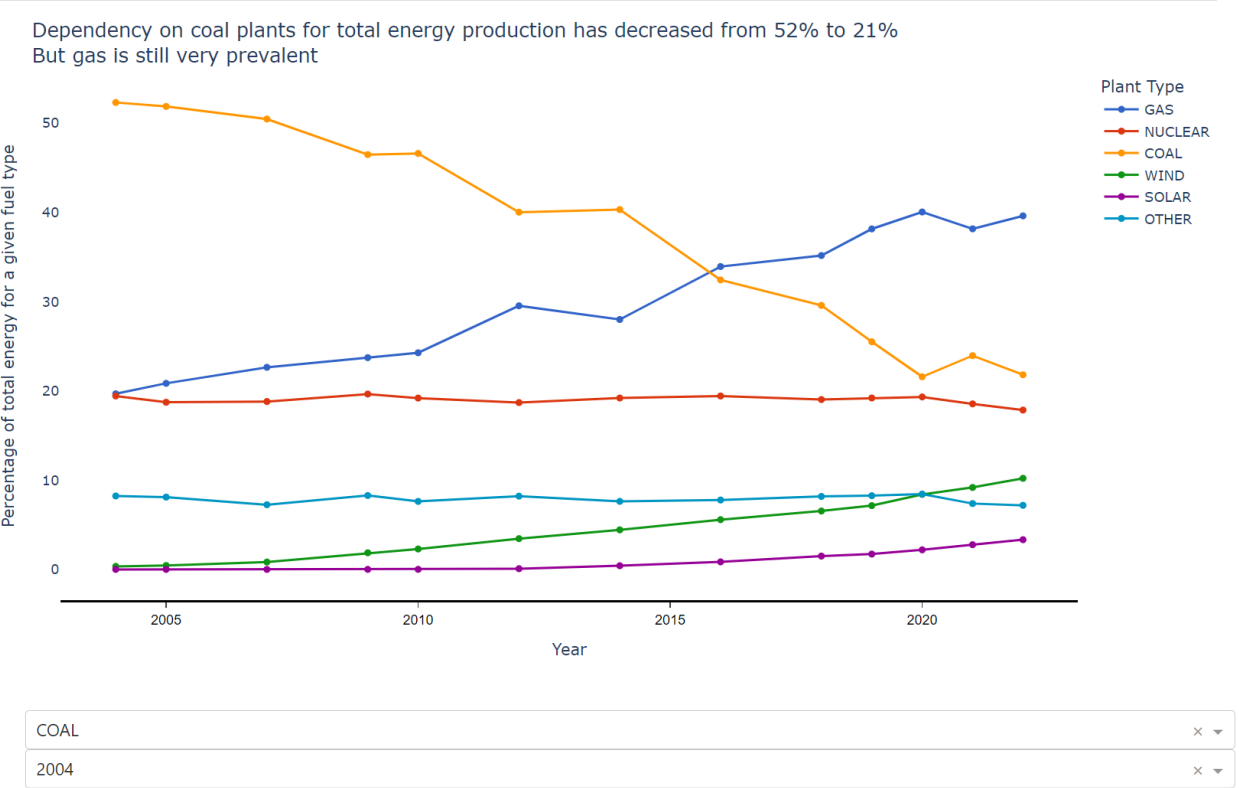
2. **State Trends:** presents the general trend plots of energy generation, price and carbon emission from 2004 to 2022. Users can choose a target state and trend in the dashboard dropdown section, and the default state is set to US, which presents the national trend.

Energy Generation, Emission and Price Trends by State

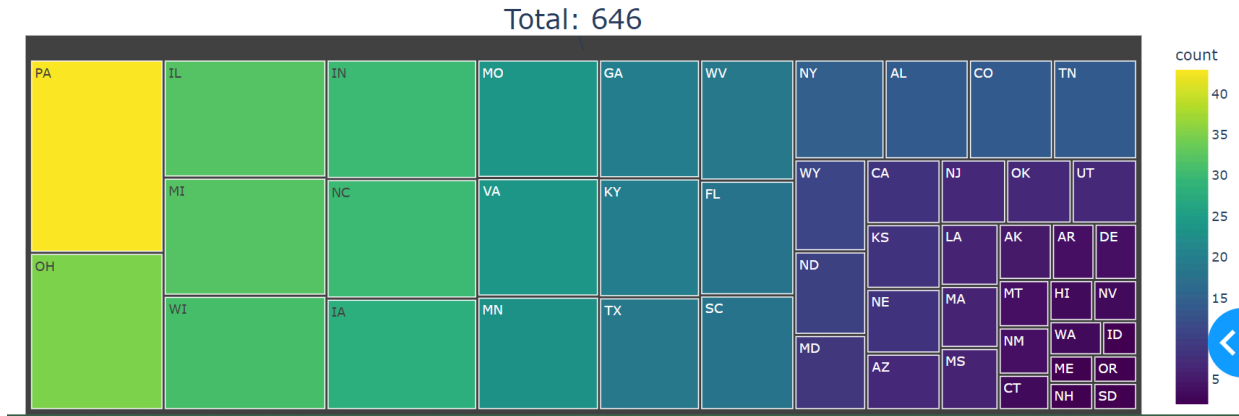


3. National Trends: This page presents a trend plot of the percentage of total plans across the US by their energy generation types from 2004 to 2022.

Users can choose the plant type(gas, nuclear, coal, wind , solar), and select the year(2004, in the dropdown section to generate a tree map showing the total counts of target energy type plants in each state .



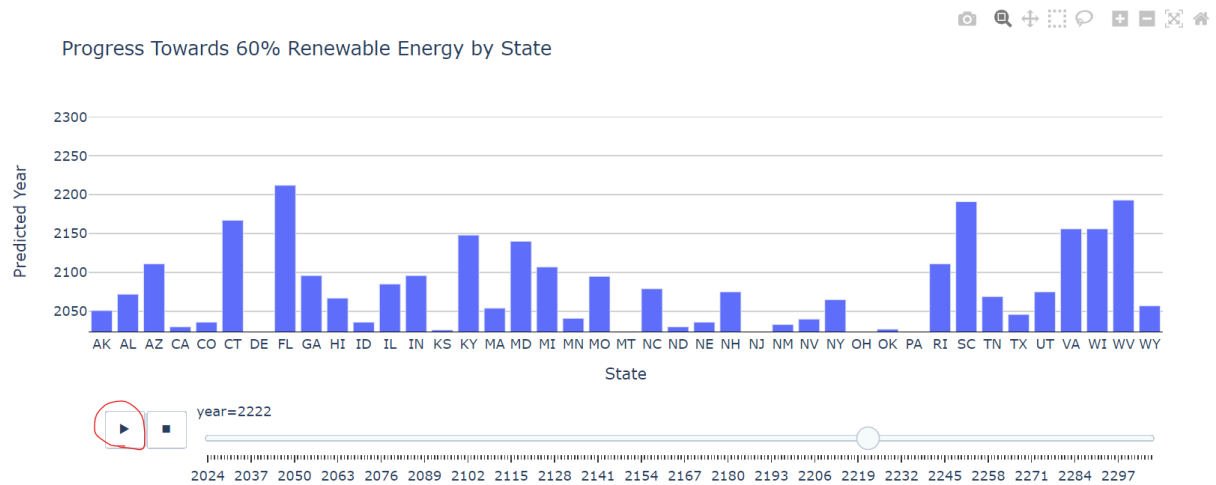
COAL Plants in 2004 by State



4. Projections: This page presents an interactive bar chart of the predicted year for each state to reach the goal of reaching 60% renewable energy generation among total energy generation. Users can click the play button on the left to

play to entire progress, or scroll the timeline to move to the prediction in a certain year.

Renewable Energy Prediction Progress



Note: Predictions are not available for the following states: AR, DC, IA, LA, ME, MS, OR, SD, VT, WA.

Conclusion

The project has accomplished beyond our initial goals. At the beginning, we didn't have a very clear plan on how to utilize our data for visualization and analysis other than explore the general trend of energy production, pricing and carbon emission. After the sql database has been established with all the data sources, it makes it much more convenient to conduct exploratory data analysis, which helps us come up with specific ideas to visualize the data. It is challenging to integrate all the visualizations in one dashboard and we encounter merging issues while working collaboratively. Fortunately we have resolved the issues and successfully created diverse forms of visualizations including line plot, bar chart, tree map and maps to present our analysis in an integrated interactive dashboard. We did not get the chance to build a model on the relationship between changes in fuel sources for energy production and electricity prices. This would be a good area for further research for this project.