Industrial Energy Efficiency Data Explorer: A Data-Driven Support Tool for Industrial Energy Modelers

Technical Documentation submitted in partial satisfaction of the requirements for the degree of Master of Environmental Data Science for the
Bren School of Environmental Science & Management

By:
Oksana Protsukha
Naomi Moraes
Eva Newby
Yos Ramirez

Committee in Charge:
Dr. Carmen Galaz Garcia
Dr. Eric Masanet

Client:
Dr. Eric Masanet for The Industrial Sustainability Analysis Laboratory at the University of California, Santa Barbara

June 2025

Industrial Energy Efficiency Data Explorer: A Data-Driven Support Tool for Industrial Energy Modelers.

As developers of this Capstone Project documentation, we archive this documentation on the Bren School's website such that the results of our research are available for all to read. Our signatures on the document signify our joint responsibility to fulfill the archiving standards set by the Bren School of Environmental Science & Management.

____

Oksana Protsukha

____

Naomi Moraes

____

Eva Newby

____

Yos Ramirez

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principle of the Bren School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions.

The Capstone Project is required of all students in the Master of Environmental Data Science (MEDS) Program. The project is a six-month-long activity in which small groups of students contribute to data science practices, products or analyses that address a challenge or need related to a specific environmental issue. This MEDS Capstone Project Technical Documentation is authored by MEDS students and has been reviewed and approved by:

____

Dr. Eric Masanet

____

Dr. Carmen Galaz García

____

DATE

Acknowledgments

Faculty Advisor & Client
Dr. Eric Masanet
The Bren School of Environmental Science & Management, U.S. Department of Energy

Capstone Advisor
Dr. Carmen Galaz Garcia
The Bren School of Environmental Science & Management

Testing, Research & Development
Aiden O'Neil
The Bren School of Environmental Science & Management

Antoine Merlo
The Bren School of Environmental Science & Management

Jason Ye
The Bren School of Environmental Science & Management

Dylan Radovic
The Bren School of Environmental Science & Management

Abstract

The industrial sector is a major contributor to greenhouse gas (GHG) emissions in the United States, accounting for 30% of energy-related CO_2 emissions with 1,360 million metric tons released in 2020[1]. Carbon dioxide (CO_2) is a greenhouse gas that absorbs and retains heat in Earth's atmosphere, contributing to global warming, disrupting ecosystems, accelerating ice melt, and amplifying extreme weather events. Dr. Eric Masanet and his research team at the Industrial Sustainability Analysis Laboratory, at the University of California, Santa Barbara, address this challenge by identifying industrial decarbonization pathways that illuminate needed technology transitions. One of their key data sources is the Industrial Assessment Centers (IAC) database, which contains over 40 years of industrial energy efficiency audits and recommendations, a program funded by the U.S. Department of Energy (DOE) and managed by over 50 universities across the country. Even though it is a valuable resource, the IAC database has several limitations. It lacks information on potential savings in GHG emissions and local air pollutants, and does not adjust for changes in technology costs over time. Additionally, it provides limited data visualization capabilities and its frequent updates complicate real-time analysis. This project aims to address these gaps by developing a publicly accessible interface that integrates GHG emissions, air pollutants, and other key data with the IAC database, and adjusts economic metrics to present values for historical data. By streamlining access to critical data and expanding analytical and data visualization capabilities, this tool will accelerate research into industrial energy efficiency and support broader decarbonization efforts in the manufacturing sector.

[1] Energy.gov. “DOE Industrial Decarbonization Roadmap.” Accessed September 23, 2024. https://www.energy.gov/industrial-technologies/doe-industrial-decarbonization-roadmap.

Table of Contents

Executive Summary

Industrial processes and outputs are vital in supporting global commerce and local economies, which enable our modern standard of living. However, the industrial sector is one of the largest contributors to energy use and greenhouse gas (GHG) emissions in the United States, accounting for 30% of energy-related Carbon dioxide (CO_2) emissions in 2020. CO_2 is a greenhouse gas that traps heat in Earth's atmosphere, driving issues such as global warming, ecosystem disruption, melting ice, and intensified extreme weather events. To address this challenge, the Industrial Sustainability Analysis Laboratory (ISAL) at the University of California, Santa Barbara, led by Dr. Eric Masanet, identifies industrial decarbonization pathways that promote energy efficiency and sustainable technology transitions. A key resource in this work is the U.S. Department of Energy's Industrial Assessment Centers (IAC) database, which contains over 40 years of energy audit data and efficiency recommendations for small to midsize manufacturing plants across the United States. The IAC data contains information covering facility types, energy usage, and recommendations for improvement, such as potential energy and cost savings. While the IAC database offers a rich foundation for research, it lacks critical data on GHG emissions, air pollutants, evolving technology costs, and offers limited visualization capabilities. Enhancing the IAC dataset through integration with additional public datasets and developing tools for data analysis and visualization will provide deeper insights, facilitate data-driven decision-making, and promote research on energy efficiency and low-carbon industrial pathways for industrial energy data modelers.

Problem Statement

The Industrial Symbiosis Analysis Lab (ISAL) currently employs manual processes to integrate supplementary datasets that address information gaps within the Industrial Assessment Center (IAC) data repository. These critical datasets encompass the Environmental Protection Agency's AP-42 Compilation of Air Emissions Factors from Stationary Sources, electricity emissions and generation data from the Energy Information Administration, and Producer Price Index data from the U.S. Bureau of Labor Statistics. The existing manual integration workflow, conducted through Excel-based operations, presents significant operational inefficiencies, such as extending processing timelines and introducing elevated error-risk. To mitigate these operational challenges, the student capstone team developed a comprehensive solution comprised of three core components: a semi-automated integration pipeline designed to reduce processing time and minimize error introduction, an interactive platform enabling data filtering and visualization capabilities for efficient preliminary data analysis, and an automated pipeline infrastructure to ensure dataset relevancy through systematic updates as newer data becomes available. This approach addresses the client's immediate challenges while establishing a scalable framework for ongoing data management requirements.

Objectives

1. Create a system (primarily through python code, and detailed in the user guides) that streamlines the addition of new data and enables complex queries using multiple inputs, with the flexibility to incorporate future data as it becomes available.

2. Develop an interactive dashboard that helps researchers visualize trends, uncover insights, and propose solutions for decarbonizing the industrial sector through improved energy efficiency.

Deliverables

1. **Integrated Dataset**: Expand the IAC data by incorporating power grid emissions, fuel-specific air pollutant factors, and economic data to adjust costs over time. The dataset will feature a scalable structure for adding new data and support estimates of GHGs, sulfur dioxide (SO_2), nitrogen oxides (NO_x), and other emissions tied to electricity and onsite fuel use.
2. **Interactive Web Dashboard**: A user-friendly tool built on the integrated dataset, enabling data exploration, custom queries, filtering, and visualizations to support industrial decarbonization analysis.
3. **User Documentation**: Includes a dictionary of the data terms and definitions, methodology for data integration, and a full list of sources to guide users in navigating and applying the tools effectively.

Approach

The following chart (Figure 1), illustrates the key tasks in the project, along with the corresponding methods of completion for each task, outlining the sequential steps involved for the completion of the deliverables.

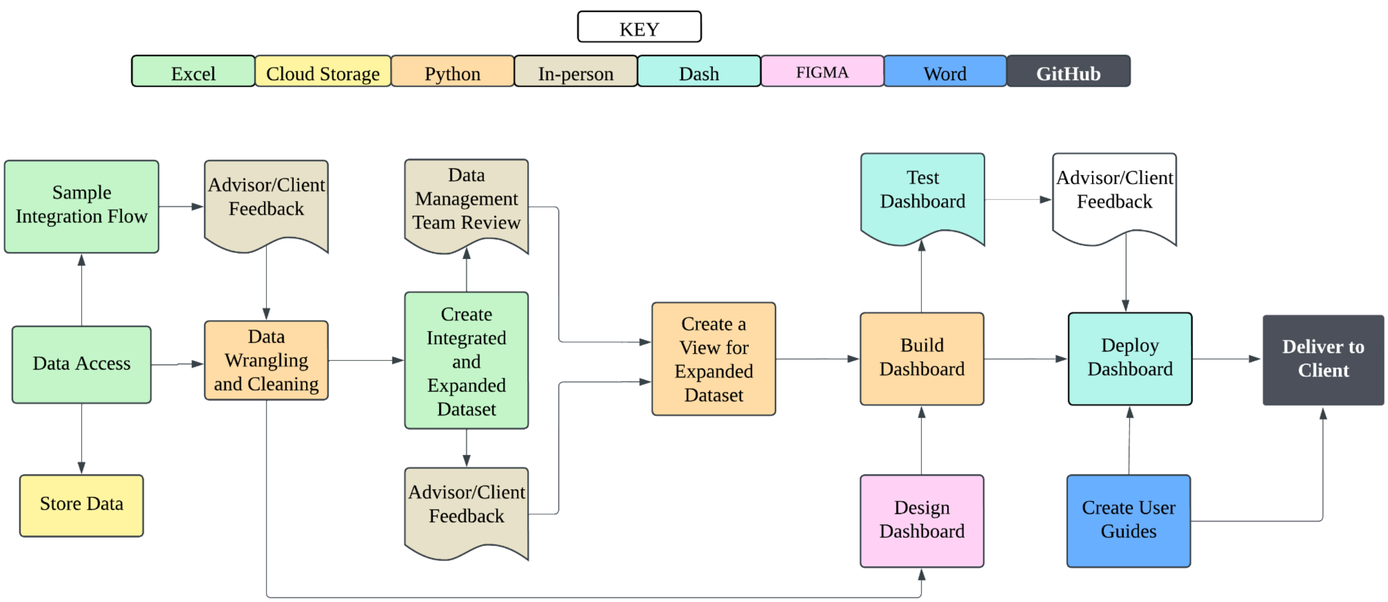


Figure 1. Capstone Project Workflow. A flowchart outlining the sequential steps involved for the completion of the deliverables.

Project Data

The project relies on publicly available data sources that do not involve intellectual property restrictions or require non-disclosure agreements. Table 1 provides a comprehensive list of all data sources needed to complete the project.

Table 1: Research Data Summary by Source and Variable		
Variable	Type, Approximate Size	Source
Industrial energy audits data: - ASSESS table (industrial audit assessments): 30 variables - RECC table (industrial audit recommendations): 55 variables	Excel 2003 (.xls), 13.2 MB	U.S. Department of Energy, Industrial Assessment Centers (IAC) Program data portal. [3]
Electric power emissions and Air Pollutants: CO_2, SO_2, NO_x by state and year (1990-present)	Excel (xlsx), 2.1 MB	U.S. Energy Information Administration (EIA), Electric Power Annual , Detailed EIA-923 emissions survey data, Electric power industry estimated emissions by state (back to 1990). [4]
Total electricity generation by state and year (1990–present)	Excel (xlsx), 6.8 MB	U.S. Energy Information Administration (EIA), Electric Power Annual, Generation and thermal output, Detailed preliminary EIA-923 monthly and annual survey data, State-level generation and fuel consumption data annual (back to 1990) [5]
Producer price index data (PPI) for technologies corresponding to each corresponding Assessment recommendation code (ARC), 1990-present	Excel (xlsx)	Shared via Google Drive by the client, Dr. Eric Masanet. Original source: Bureau of labor statistics present price index data (BLS PPI) [6]

Table 1: Research Data Summary by Source and Variable		
Variable	Type, Approximate Size	Source
Combustion emission Factors for CO_2, SO_x, NO_x by fuel type	PDF	Compilation of Air Pollutant Emissions Factors from Stationary Sources (AP-42) [7]
North American Industry Classification System Codes	Excel (xlsx) 340 KB	2022 NAICS Code List from NAICS Association [8]
Standard Industrial Classification System Codes	Excel (xlsx) 149 KB	2022 NAICS to SIC Crosswalk from NAICS Association [9]
Assessment Recommendation Codes (ARCs)	PDF	Assessment Recommendation Code from Revised ARC Manual [10]

Methods

User Research and Requirements Collection

In addition to the initial project proposal, weekly meetings with the client served as a primary source for gathering and refining requirements. These discussions were essential for clarifying deliverables, aligning expectations, and developing a deep understanding of both the technical constraints and user needs. A central element of our methodology was the development of a user story to guide design decisions and ensure that the dashboard is relevant and usable for the target audience.

This iterative, user centered approach allowed us to tailor the dashboard to the real-world context in which it will be used, ensuring its value as a practical tool for modeling and analysis. User feedback was continuously gathered during development and testing phases through informal demonstrations and weekly meetings, directly informing improvements to the layout, functionality, and visualizations.

Data Management

The data flow and integration pipeline used to construct the unified dataset is shown in Figure 2, which outlines how raw data from various sources listed in Table 1 were processed to produce the enhanced integrated dataset deliverable. These sources include industrial energy audit records, electricity generation and emissions data, producer price indices, and emissions factors for different fuel types. All datasets used are publicly available and free from intellectual property restrictions.

Data Integration

Utilising Python, in a series of Jupyter Notebooks, key datasets were established or explored, cleaned and integrated. Datasets included; the ‘Assess’ and ‘Recc’ tables from the U.S Department of Energy’s Industrial Assessment Centers detailing assessment and recommendation codes and information from audits, emissions and electricity generation data categorized by state and year from the U.S. Energy Information Administration, emission factors by fuel type sourced front the U.S. Environmental Protection Agency’s AP-42 database, and the U.S Bureau of Labor Statistics Producer Price Index.

Industrial Assessment Centers Database

The Industrial Assessment Centers Database, maintained by the U.S. Department of Energy, contains detailed records from energy audits conducted at small and medium-sized manufacturers across the United States. The database includes two key tables: ‘ASSESS’, which documents general audit and facility information, and ‘RECC’, which details recommended energy efficiency measures, associated implementation costs, projected savings, and implementation status.

Fuel Combustion Emission Factors

Fuel combustion emission factors were integrated for CO_2, NO_x, and SO_2 to ensure comprehensive air pollutant accounting. CO_2 combustion factors were provided by the client, while NO_x and SO_2 emission factors were derived from EPA's AP-42 compilation. Emission factors for various fuel types in large wall-fired boilers were converted from pounds to kilograms and normalized to kg/MMBTU.

Electricity Generation and Emission Factors

State-specific electricity emissions and generation data by year provided electricity emission factors for calculating emissions generated and avoided for each plant and recommendation in units of kg of pollutant per kWh of electricity (Figure 2).

Producer Price Index

Assessment recommendation costs (ARC) were normalized across time using a Producer Price Index ratio (Figure 2) multiplied by the ARC implementation cost.

NAICS, SIC, ARC Codes and Description

NAICS and SIC codes and descriptions were sourced from the 2022 NAICS Keyword list and 2022 NAICS to SIC Crosswalk. Assessment recommendation code descriptions were derived from the Industrial Assessment Center Assessment Recommendation Codes Manual.

Integrated Dataset

The Figure 2 below illustrates the different components of industrial energy audits data, electric power emissions and generation data, Producer Price Index data, and combustion emission factors integrated into one.

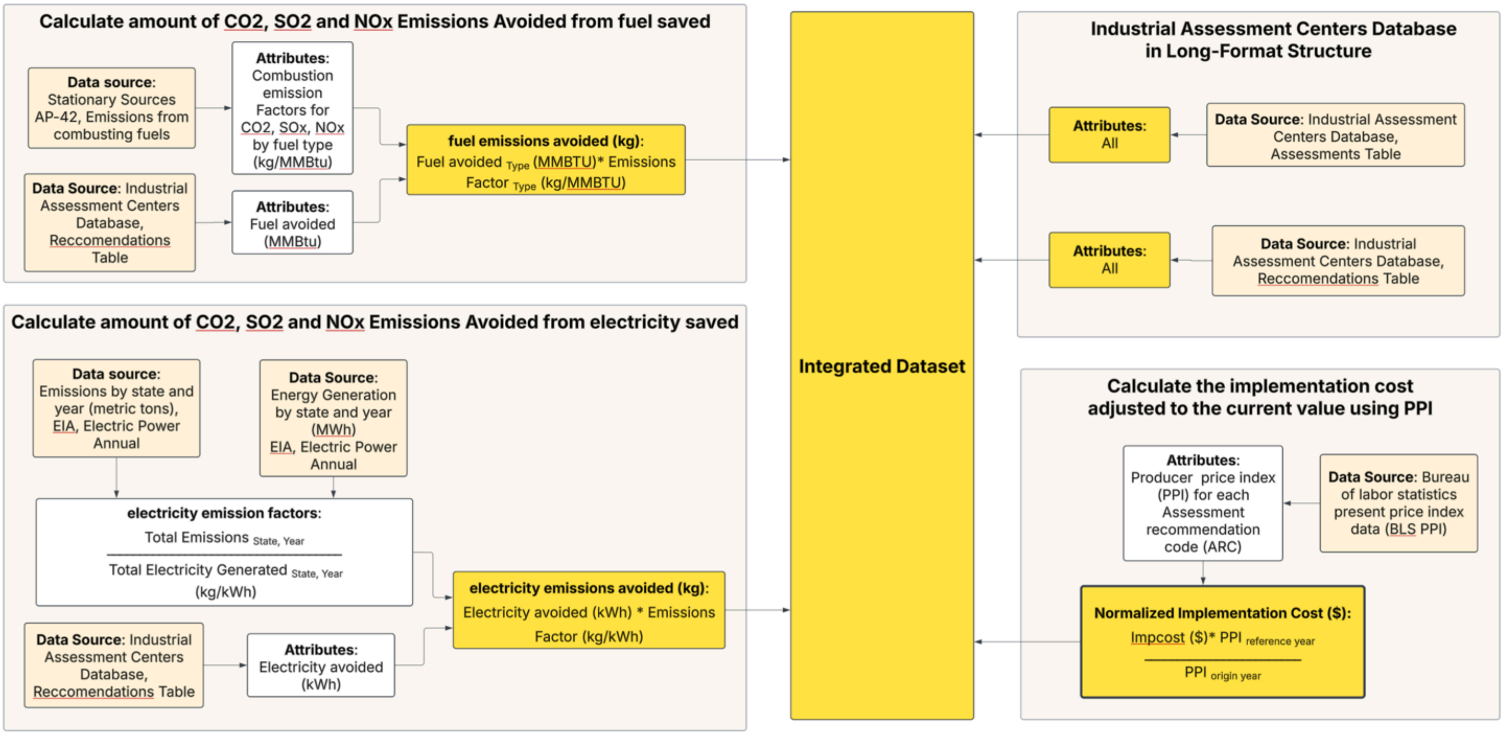


Figure 2. Industrial Energy Integrated Dataset: Data flow and integration process of raw data sources for the completion of the enhanced integrated dataset deliverable.

Product Design

Intended User(s)

The primary users of the Industrial Energy Efficiency Dashboard are experienced industrial energy analysts or modelers focused on identifying energy efficiency pathways for the industrial sector. These users are familiar with the Industrial Assessment Centers (IAC) dataset and use it to explore, analyze, and retrieve empirical data on the energy-saving measures they prioritize. Their goal is to obtain data on energy savings, implementation costs, payback periods, and avoided greenhouse gas and air pollutant emissions to support models or analyses that guide strategies for reducing energy consumption and emissions in industrial systems.

Dashboard Development Approach

Prior to the dashboard development, the team created a dashboard interface prototype using Figma, a cloud-based design platform. This early design phase played a critical role in defining the dashboard's necessary elements. Through collaborative design sessions, the team clarified the dashboard's key use cases, data visualizations, and interactive features helping to ensure the final dashboard aligned with user needs and offered a clear, intuitive interface.

To maintain consistency across the project the dashboard was developed using Python, aligning with the tools and processes used during data cleaning and integration. The dashboard application was built using Dash, an open-source Python framework developed by Plotly Inc. Dash is specifically designed for creating interactive data science and machine learning web applications, as well as responsive and visually rich data visualizations. The interactive visual components in the dashboard are powered by Plotly.py, enabling users to explore the integrated dataset with dynamic filters and real-time visual updates. Key data attributes such as audit year, state, recommendation category, and energy savings can be easily selected and compared, providing valuable insights for experienced energy efficiency modelers.

The dashboard code is version-controlled and publicly hosted on GitHub to support transparency, team collaboration, and long-term reproducibility: https://github.com/IndustrialEnergy.

Dashboard Design

The dashboard design was informed by the input from the project client, Dr. Eric Masanet. Its primary objectives are to provide streamlined access to an enhanced IAC dataset and to support automated exploratory data analysis.

To meet these goals, the dashboard is organized into the following sections:

- **Navigation Bar**
  - Includes page tab links and a logo for easy navigation.
- **Dashboard Page**
  - Dynamic visualizations categorized by three key metrics (energy, cost, and emissions)

    - Filters available help users customize their data view

      Must-have Filters

        - Assessment recommendation description
        - Sector: Manufacturing type based on NAICS codes
        - States: A dropdown menu listing all states where audits were conducted
        - Date Range: Selectable start and end years
        - Recommendation implementation status

      Nice-to-have Filters

        - Avoid–Shift–Improve (ASI) measures
        - Recommendation sub-categories based on 2-, 3-, and 4-digit classification codes
        - Outliers: An option to hide/show outliers on the boxplot charts
- **About Page**
    - Contains information about the dashboard, the team, and access.
- **Data Page**
    - Contains the data in excel and csv format for download
- **Documentation Page**
    - Contains the User Guide

Product Roadmap

The development of the dashboard followed the product roadmap illustrated in Figure 3 below. This roadmap outlines key development phases, beginning with initial data exploration and integration, to the implementation of interactive visualizations and user testing. The timeline summarizes the major milestones and key deliverables required to successfully complete the capstone project.

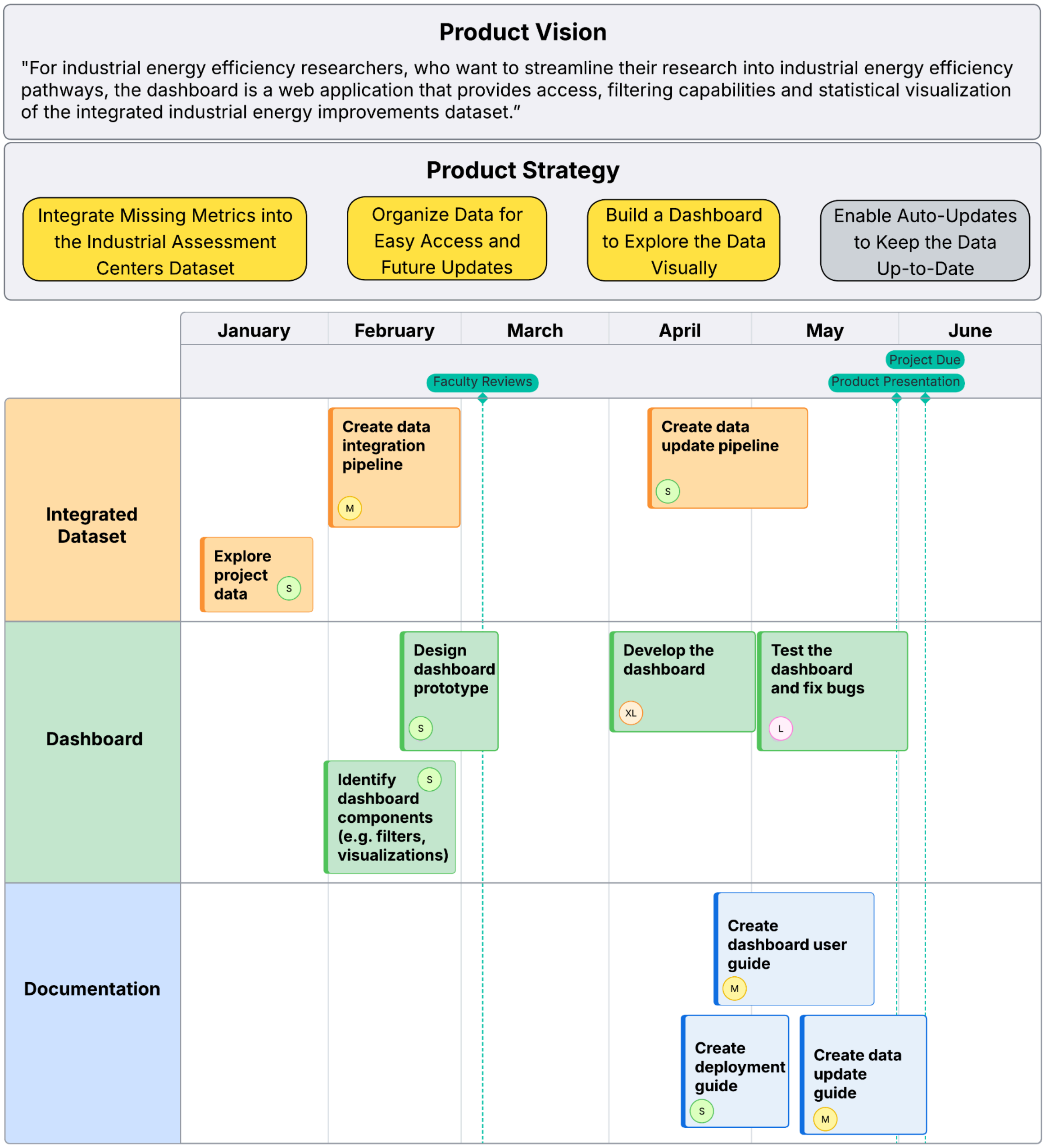


Figure 3. Product Roadmap

Product Testing

A critical phase in the product development lifecycle is the testing of the product prior to its release. Testing ensures that the product, which includes all final deliverables, align with the client's expectations. All final deliverables must ensure a seamless user experience. This includes a responsive dashboard, a fully functional integrated dataset, and clear, accurate documentation. The testing phase focused on the presentation of the integrated data in the dashboard and resolving any functional issues that a user may encounter.

Process

Active participation from target users and team members played a key role to identify inconsistencies, bugs, and other improvement opportunities. Target users included PhD students and post-doctoral researchers from the Industrial Sustainability Lab as well as Dr. Masanet, who provided domain specific insights. Engagement began during the mid-development phase and continued through to final testing. Feedback was collected through Zoom demonstrations of the prototype, and in person meetings during which users interacted with the tool and provided real time input. Their feedback informed refinements to both functionality and usability. Further testing details are sectioned by deliverable below.

Integrated dataset

1. To ensure the accuracy of the integrated dataset, team members reviewed the expected columns after integration, naming conventions, and calculated fields by inspecting the underlying code, later confirmed by the client.

2. Completeness of values expected in rows of the dataset was assessed just before the initial dashboard deployment, also validated by the client.

3. Ensuring the reliability and alignment of rows and columns of the integrated dataset was critical, as any flaws could compromise the functionality and trustworthiness of the entire dashboard.

Dashboard

1. Interface responsiveness was evaluated through local deployments between implementing changes that were necessary to address misalignments and bugs, with internal testing conducted by team members.

2. Layout clarity was initially guided by a prototype feedback from the client, and confirmed after the deployment of the dashboard by the client and PhD students.

3. Intuitive navigation was assessed through targeted questions aimed at uncovering usability concerns, with input gathered from both team members and the client, who provided final approval.

4. Ensuring responsiveness, clarity, and ease of navigation was essential and constantly checked by team members after any changes made to the underlying code, as a dashboard should not be unresponsive, confusing, or frustrating to use.

Documentation

1. The clarity of the documentation was refined through multiple reruns of the integration process and dashboard deployment, allowing team members to finalize each step outlined for manual updates.

2. Accuracy of the documentation was ensured through both independent reviews and collaborative sessions, leading to necessary edits based on feedback from the client and team members.

3. Depth was ensured by considering the expected user experience and the technical knowledge needed to understand the content, with assessment by the client, PhD students, and team members to align with anticipated interaction.

The team members and client prioritized a usable interface that met the final deliverables.

For any issues not addressed due to time constraints, it was determined that logging them and making them available in the project's GitHub as issues would benefit any future improvements.

Dashboard architecture

The Industrial Energy Dashboard architecture separates data storage and processing from the application code, enabling independent data updates without modifying code.

Data Component

The integrated dataset and all supporting datasets are stored on the Bren server at apps.bren.ucsb.edu/capstone/industrialenergy/dashboard/data/. Users can access this data directly on the Bren VPN at http://128.111.89.73:3010 or connect via SSH to apps.bren.ucsb.edu/ and then navigate to the folder home/capstone/industrialenergy/data/. For SSH setup instructions, consult the guide Using VS Code to SSH into Bren Servers . The Dataset can be updated with the latest data using the data pipeline workflow. This process, except for the data upload step, is fully automated. Refer to the Data Update and Deployment guide section on how to update data and redeploy the dashboard.

Code Component

The complete application codebase is maintained in the GitHub repository at https://github.com/IndustrialEnergy/dashboard. The application architecture utilizes Docker containerization, which means anyone can deploy the dashboard without installing additional software dependencies or managing complex library configurations. The Docker-based environment setup is illustrated in Figure 4.

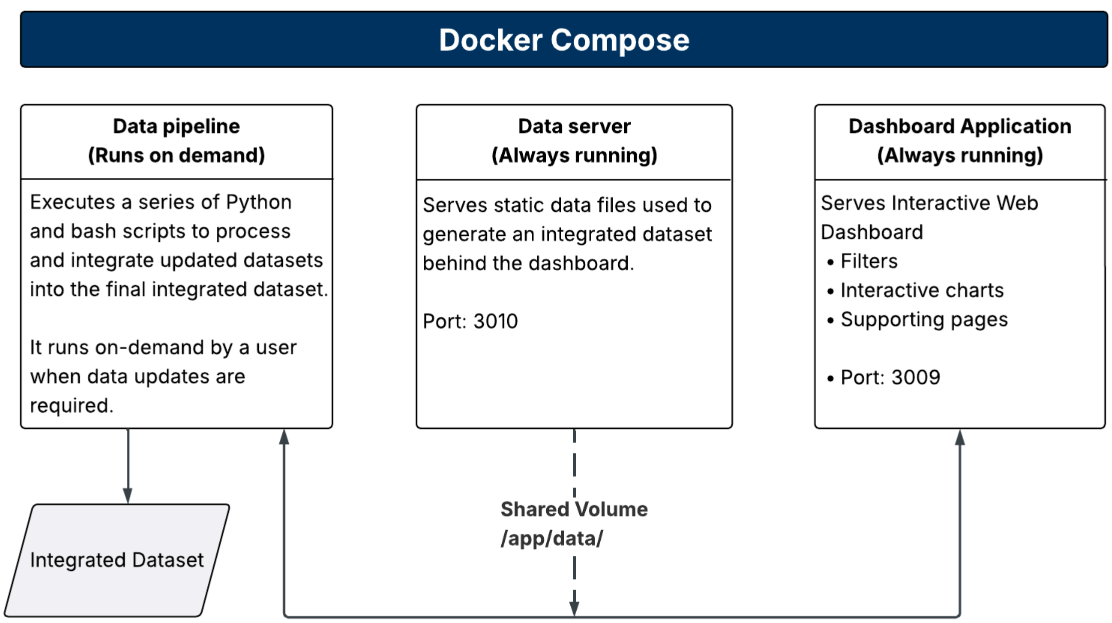


Figure 4. Dashboard Environment Setup

Modular Structure of the Dashboard

The dashboard follows a modular design, as shown in Figure 5 and described below.

The application is built around three core Python modules that handle server setup, application logic, and data access:

- **server.py:** Serves as the entry point of the application, managing web server configuration and initialization. This module provides the technical infrastructure that enables users to access the dashboard via their web browsers.
- **app.py:** Acts as the central application controller, coordinating all dashboard components. It handles page routing, imports necessary modules, and defines the interactive relationships between different parts of the system.
- **data_loader.py**: Loads the integrated CSV dataset and transforms it into pandas DataFrames. This module ensures the data is properly formatted for visualization.

In addition to these core modules, the dashboard is organized into four specialized layers that simplify development, promote reuse, and support easy maintenance:

- **Pages Layer (pages/ directory)**
  Defines the structure and layout of each dashboard page. It controls how information is arranged and presented to users.
- **Components Layer (components/** directory**)**
  Includes reusable interface elements like filters, menus, and buttons. These elements keep the interface consistent and reduce code duplication.
- **Charts Layer (charts/** directory**)**
  Contains functions that display data as interactive visualizations. Each function creates a chart based on user defined filters.
- **Callbacks Layer (callbacks/** directory**)**
  Handles the interactive behavior of the dashboard. It connects user actions—like clicks or filter changes—to real-time updates in charts and other elements.

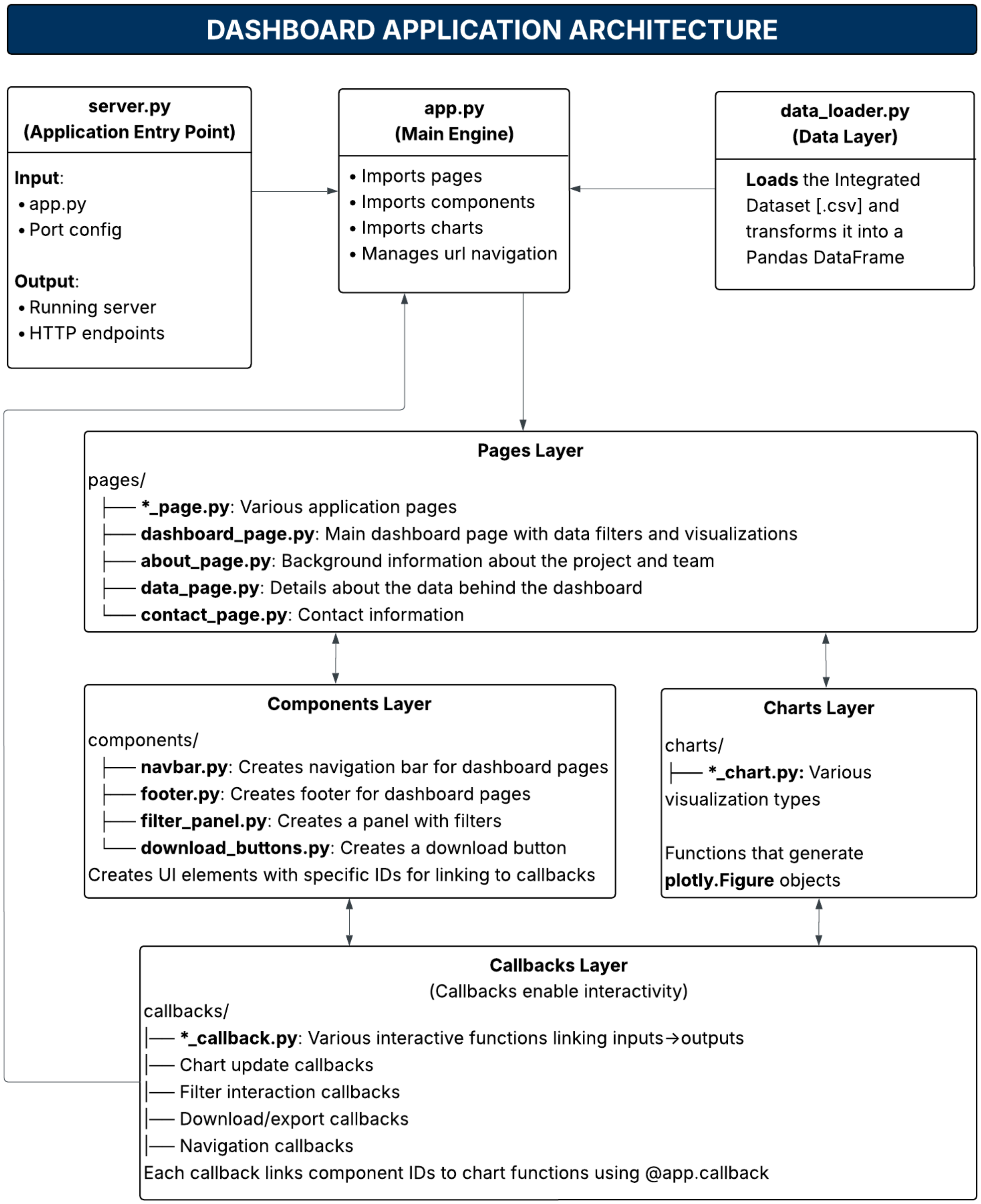


Figure 5. Dashboard Architecture

User Guide

Introduction

This website was built in order to host an industrial energy efficiency data explorer. Key components of the website include a dashboard, access to the underlying integrated industrial dataset and its creation, as well as documentation necessary for users to utilize and update as new data becomes available.

The website can be navigated through the buttons in the header, located along the top-left corner.

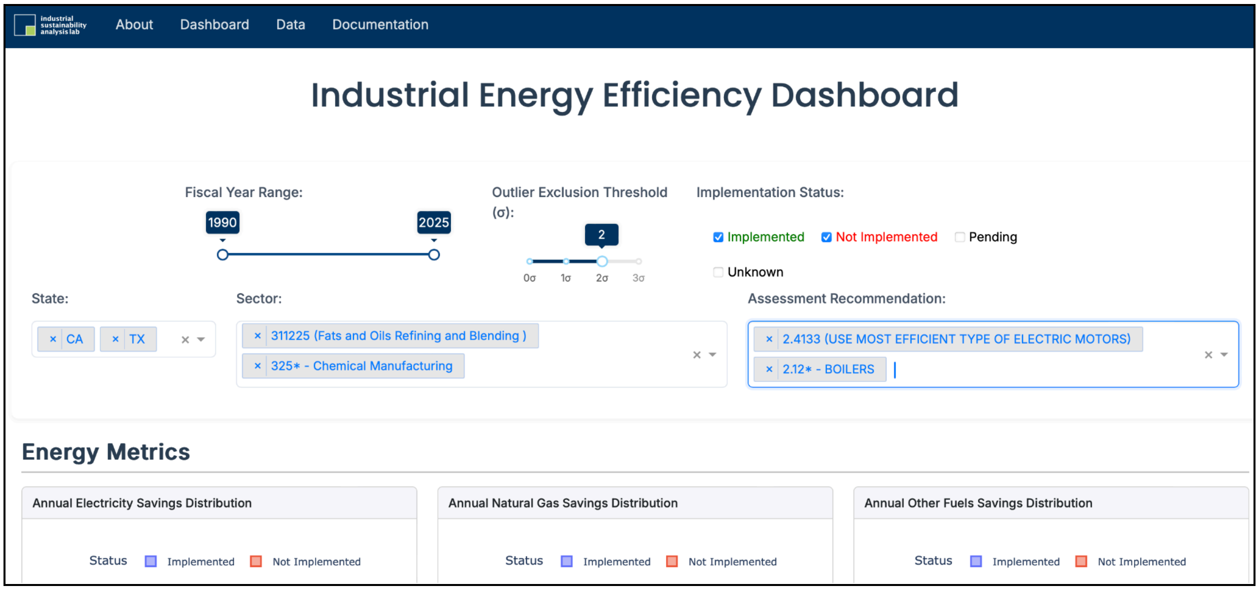
Clicking on the following buttons redirects to the following pages:

About Page

The About page features information about the dashboard as well as the capstone team, data, and code behind the project. To view the summaries, click on the downward arrows (bordered in red below).

Dashboard Page

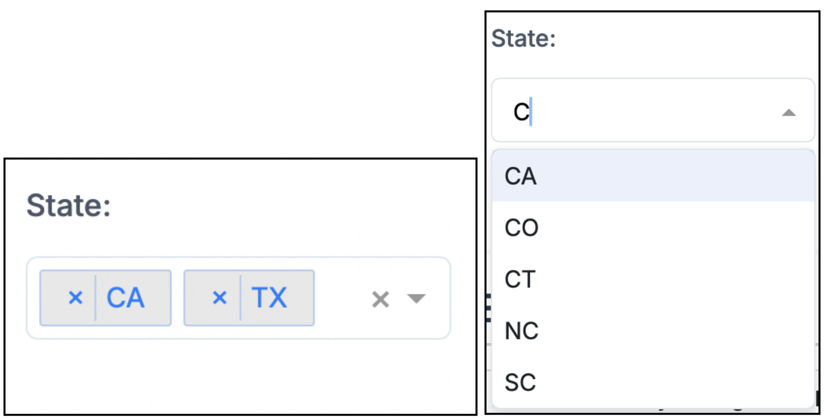
The dashboard presents a centralized interface for exploring industrial energy audit data through potential and realized energy, fiscal and emissions metrics, across facilities and recommendations. It is designed to support data exploration, policy analysis, and decarbonization strategies in industrial energy use.



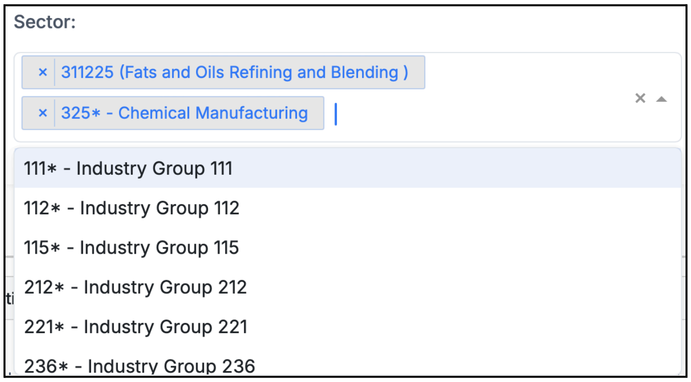
Dashboard Filters

The dashboard includes multiple filters to allow users to tailor data visualizations around their research and modeling needs. A list of filters can be found below.

- State: Select any of the 50 US states, as well as Puerto Rico. Select one, multiple, or leave the selection empty to include all states. Scroll or type to find desired codes.

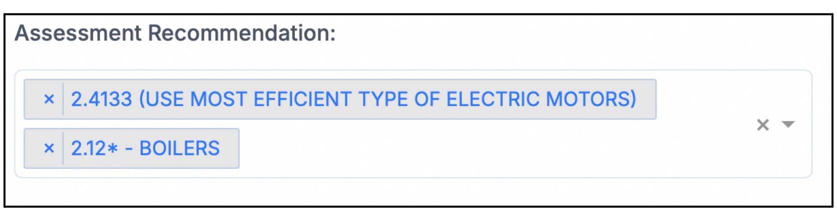


- Sector (NAICS Code): Select relevant NAICS codes or industries. Select one, multiple, or leave the selection empty to include all sectors. Scroll or type to find desired codes.

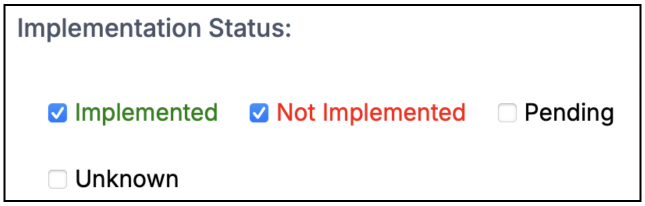


- Fiscal Year Range: Narrow down results to a specific time frame (within the 1990 - 2025 time range) via the sliding scale. Click and drag the white buttons to the desired year(s).

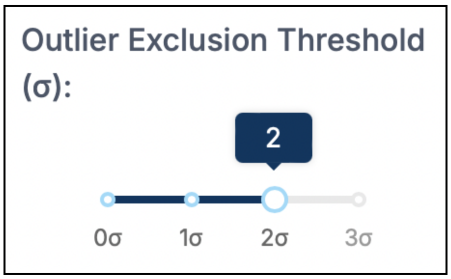
- Assessment Recommendation Code (ARC): Select types of energy-saving recommendations or overarching recommendation categories - produced from Industrial Assessment Centers' (IAC) audits. Select one, multiple, or leave the selection empty to include all recommendations. Scroll down or type to select desired codes.



- Implementation Status: Filter by whether an IAC recommendation was implemented, not implemented, pending, or unknown. Single, multiple, and none can be checked.



- Outlier Exclusion Threshold (σ): Filter the data to focus the visualized data on the desired distribution. Click and drag the white button to choose the desired visualized standard deviations from the mean.



Visualizations

The Importance of Box Plots

A box plot (or box-and-whisker plot) summarizes the distribution of the selected inputs using five key values: the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The box spans from Q1 to Q3, representing the middle 50% of the data (interquartile range), with a line inside showing the median. Whiskers extend from the box to the smallest and largest values within 1.5 times the interquartile range, while any points beyond are considered outliers. The shape of the box and position of the median indicate the data's spread and skewness, helping identify symmetry, variability, and unusual values.

Visualized Metrics

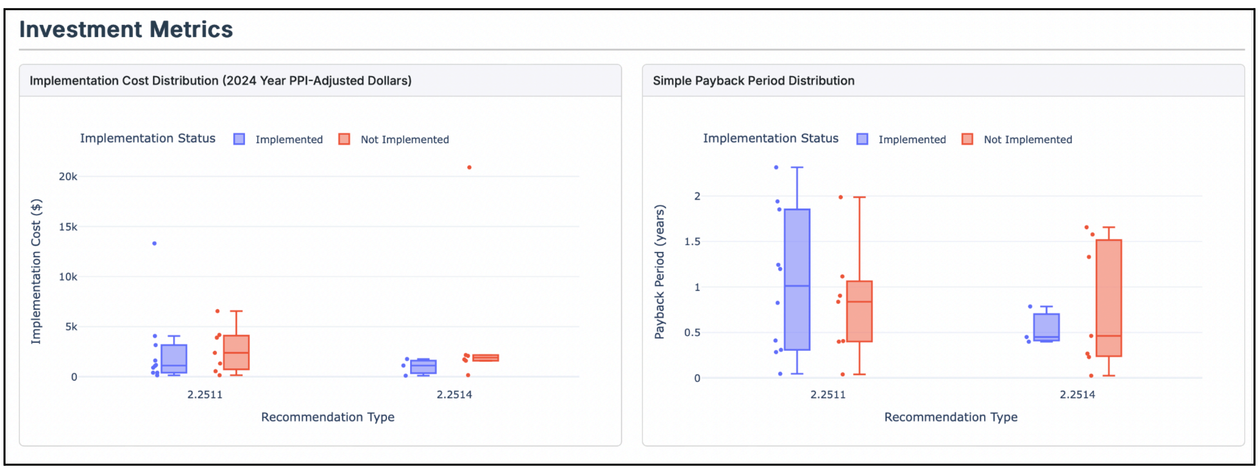
The following below is a list of boxplot titles, brief descriptions, and units currently available on the dashboard:

Energy Metrics

- "Annual Electricity Savings Distribution" - Electricity saved per recommendation type in kilowatt-hours (kWh) per year.
- "Annual Natural Gas Savings Distribution" - Natural gas saved per recommendation type in million British thermal units (MMBtu) per year.
- "Annual Other Fuels Savings Distribution" - Fuels other than natural gas saved per recommendation type in million British thermal units (MMBtu) per year.

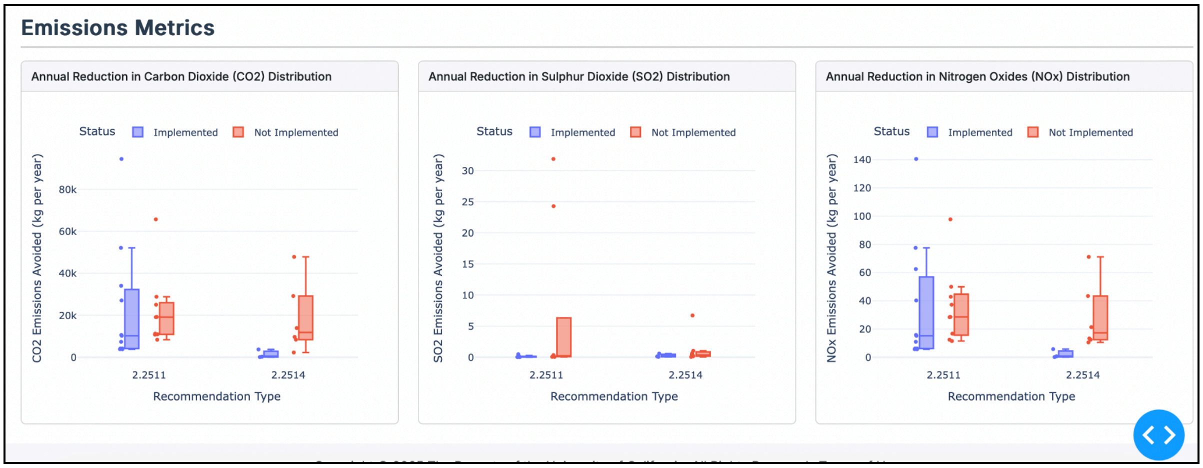
Investment Metrics

- “Implementation Cost Distribution (2024 Year PPI- Adjusted Dollars)” - Implementation cost per recommendation type in United States Dollars (USD 2024$).
- “Simple Payback Period Distribution” - Payback period per recommendation type in years. Simple payback (years) is calculated as the implementation cost divided by the annual energy savings times energy cost.



Emissions Metrics

- “Annual Reduction in Carbon Dioxide (CO_2) Distribution” - CO_2 emissions avoided per recommendation type in kilograms (kg) per year.
- “Annual Reduction in Sulphur Dioxide (SO_2) Distribution” - SO_2 emissions avoided per recommendation type in kilograms (kg) per year.
- “Annual Reduction in Nitrogen Oxides (NOx) Distribution” - NOx emissions avoided per recommendation type in kilograms (kg) per year.

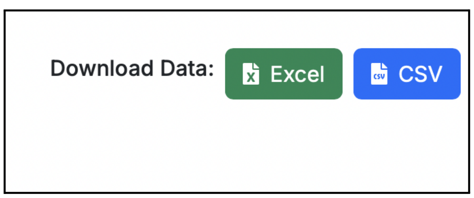


Data Page

The Data page hosts the integrated dataset that the dashboard is utilizing at the time of deployment.

Download Data

Users can download data by selecting the “Data” tab. The downloaded files are provided in either CSV or Excel format.



Latest Data Update, Time of the Most Recent Update to the Dashboard

The dashboard is updated periodically. The latest update date is displayed at the top or bottom of the homepage or data tab (e.g., Last updated: May 2025). Please check this date to ensure you are working with the most current data available. As multiple datasets across several sources are used to generate the integrated dataset and the information on the dashboard, the dataset version information will be available under the “about” tab on the dashboard.

Documentation Page

The Documentation page contains this User Guide, including a Frequently Asked Questions section listed below.

FAQ (Frequently Asked Questions)

Can I use this data for academic research or publication?

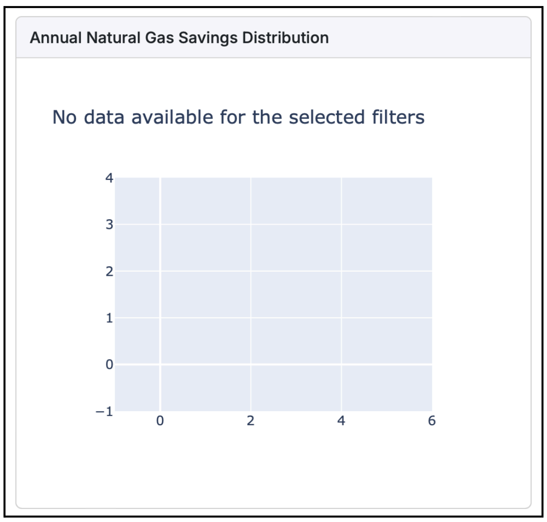
Yes, provided proper citation. Please refer to the dataset's license.

Are cost savings in present-day dollars?

Yes, all cost savings are represented in present-day USD and are adjusted for inflation and fluctuating costs of materials. This was done using the Producer Price Index (PPI), which measures the average change over time in the cost of materials. PPI data is published by the U.S. Bureau of Labor Statistics (BLS).

Why are some data points missing or incomplete?

Not all facilities report full implementation status, and some audits may lack complete metadata. It may also be expected that certain recommendations do not provide data on certain metrics. In such cases, visualizations will return the following message ["No data available for the selected filters"] and an empty chart.



I think I found an error on the dashboard, who do I contact?

For maintenance and upkeep of the dashboard, please contact Dr. Eric Masanet at emasanet@ucsb.edu. Please include screenshots of the error and a description.

Data Update and Deployment Guide

Dashboard Installation on a Local Machine

Development Environment Setup

1. Install VS Code if you don't have it yet. https://code.visualstudio.com/

2. Install Anaconda if you don't have it yet.

3. Clone the repo from Git. In the terminal, run:

   `git clone git@github.com:IndustrialEnergy/dashboard.git`

4. Create a new conda environment. In the terminal, run:
   `conda env create -f environment.yml`

5. Activate the conda environment. In the terminal, run:
   `conda activate industrialenergy`

6. Update the file structure on your local machine by creating missing folders (e.g. some folders are in .gitignore and are not in the remote repo): refer to the section Repository Organization in this README.md.

7. Copy the latest available integrated dataset file from the Bren server directly apps.bren.ucsb.edu/capstone/industrialenergy/dashboard/data/final folder using SSH or from https://apps.bren.ucsb.edu/IE-Data/ to the `data/final`

   a. For SSH connection instructions, consult the guide Using VS Code to SSH into Bren Servers

   b. Alternatively, generate an updated dataset by running a data pipeline. Refer to section Data Update for step-by-step instructions.

Running the Dashboard Locally

Locally without docker

1. Ensure that you are in the root folder.

2. In the terminal run:
   `python dashboard_app/server.py`

3. Run the dashboard in a browser on your local machine:
   http://localhost:3009/dashboard

Locally in docker

1. Install docker desktop
   https://docs.docker.com/desktop/

2. Launch the docker desktop

3. Build the docker image

i. Navigate to the dashboard application root folder:

   `cd ../dashboard/`

ii. Remove existing containers and volumes:

   `docker-compose down --volumes`

iii. Rebuild images with --no-cache:

   `docker-compose build --no-cache`

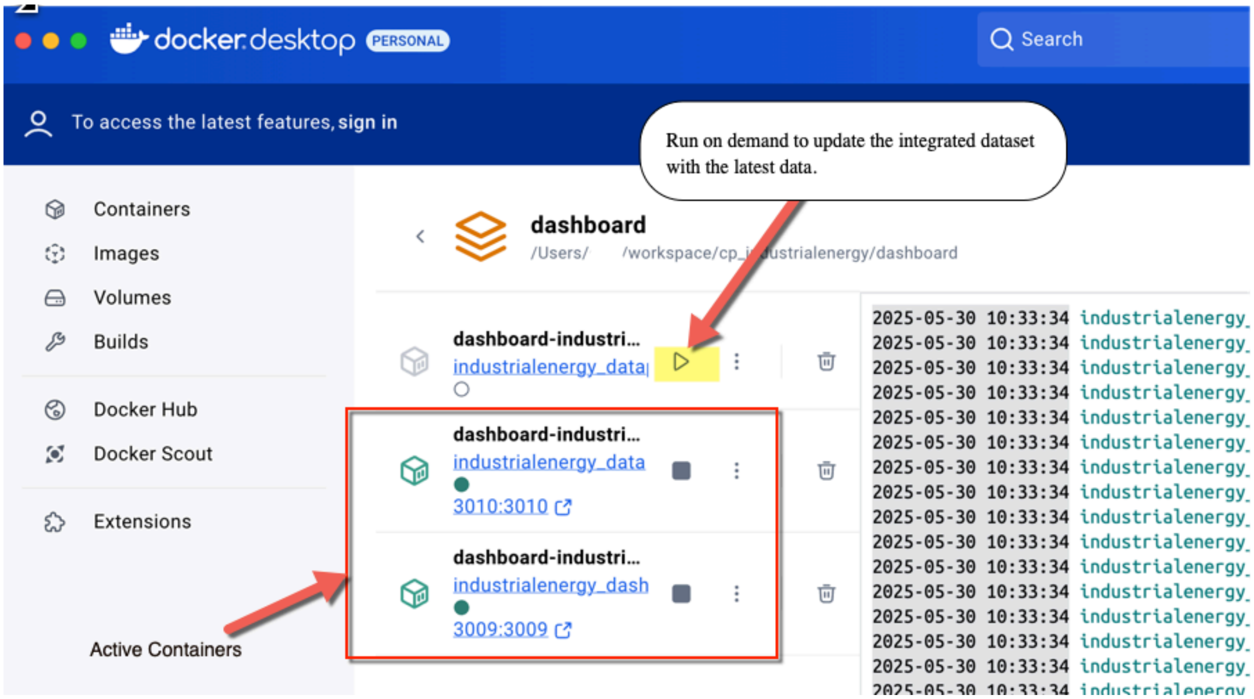
iv. Start docker containers with data and dashboard

`docker-compose up -d industrialenergy_data industrialenergy_dashboard`

4. Open the dashboard in a browser on your local machine:
   http://localhost:3009/dashboard

   Important: You can start the dashboard either using Docker or by running `python dashboard_app/server.py` directly—but not both at the same time. Doing so will result in a “port already in use” error.

5. You can view all docker containers and their status in the Docker desktop:



Data Update

Before running the data pipeline, ensure that all required files are present, correctly named, and organized according to the folder structure described below.

- **raw/ folder**
  Include only the updated files in this folder. Files that haven't changed since the last update are already stored in the processed/ folder and do not need to be reuploaded.
  Note: When the data pipeline runs, files in raw/ will be moved to the archive/ folder, and the processed versions will be saved to the processed/ folder.

- **processed/ folder**
  This folder contains two types of files:

  1. Files automatically generated by the data pipeline

  2. Static files that do not change between updates (e.g., arc_descriptions.csv)
     Important: If any static file needs to be updated, manually edit it directly in the processed/ folder.

- **final/ folder**
  The fully integrated dataset is automatically saved here after the data pipeline is executed.

Example of the correct folder structure:

```
├── data
│   ├── Dockerfile
│   ├── archive
│   ├── final
│   │   └── iac_integrated.csv
│   ├── processed
│   │   ├── NAICS_SIC_Xwalk.csv
│   │   ├── arc_descriptions.csv
│   │   ├── ec_emission_factors.csv
│   │   ├── fuel_emission_factors.csv
│   │   ├── iac.csv
│   │   ├── naics_hierarchy.csv
│   │   ├── ppi.csv
│   │   └── sic_to_naics.csv
│   └── raw
│   │   ├── ARC_PPI_YYYY.xlsx   (e.g. ARC_PPI_2024.xlsx)
│   │   ├── IAC_Database_YYYYMMDD.xls  (e.g. IAC_Database_20250518.xlsx)
│   │   ├── annual_generation_state_YYYY.xls  (e.g.  annual_generation_state_2024.xls)
│   │   └── emission_annual_YYYY.xlsx  (emission_annual_2024.xlsx)
```

---

1. Ensure you have dashboard installed:
    a. If working on a local machine, complete steps 1-7 following instructions in the [Configure the environment] section to install the dashboard locally.
    b. If working on the Bren server, no additional installation is required. The dashboard is available in the following location on the Bren server: apps.bren.ucsb.edu/capstone/industrialenergy/dashboard/
2. Upload updated files to the **raw/ folder**
    a. Only upload the files that changed since the last data update.
    b. Ensure that the files have correct titles:

Expected File Name	Example	Download URL
ARC_PPI_YYYY.xlsx	ARC_PPI_2024.xlsx	Google drive
IAC_Database_YYYYMMDD.xls	IAC_Database_20250518.xlsx	IAC website
annual_generation_state_YYYY.xls	annual_generation_state_2024.xls	EIA State Generation Excel
emission_annual_YYYY.xlsx	emission_annual_2024.xlsx	EIA Annual Emission Excel

3. Navigate to the root folder
4. The data update script can be executed in Docker (**Recommended**) or locally
    a. To run in Docker (**Recommended**) - In the terminal run the following command:

```
docker-compose --profile pipeline run --rm industrialenergy_datapipeline
```

    b. To directly trigger a script - in the terminal run the following command:

```
Python tools/data_pipeline/run_pipeline.sh
```

5. The update process might take a few minutes to run. When completed, the updated file “iac_integrated.csv” will be generated and saved into the /data/final folder

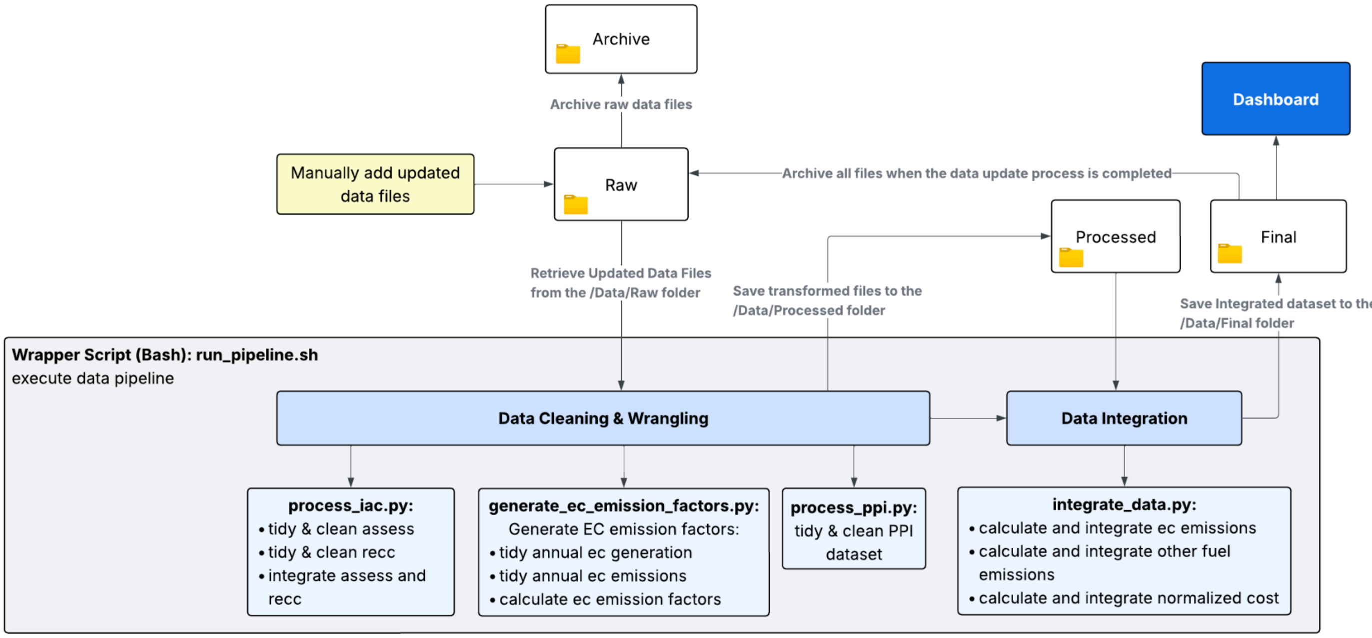


Figure 6. Data processing pipeline

Deploying the Dashboard to Production

Note: Deployment to the production environment is done using Docker containers.

1. Connect via SSH to the server apps.bren.ucsb.edu
2. Navigate to the dashboard root folder

apps.bren.ucsb.edu/capstone/industrialenergy/dashboard/

3. When in the server
    a. Open a terminal
    b. In the terminal run the following commands
        i. Navigate to the IndustrialEnergy application root folder:

`cd capstone/industrialenergy/dashboard/`

        ii. Remove existing containers and volumes:

`docker-compose down --volumes`

        iii. Rebuild images with --no-cache:

`docker-compose build --no-cache`

iv. Start docker containers with data and dashboard

```
docker-compose up -d industrialenergy_data industrialenergy_dashboard
```

Archive Access

The integrated dataset generated is not to be archived, as it will be changing and updated on a regular basis as updated input datasets become available. All past data used to create the integrated dataset will remain in the “archive” folder present within the Archive data folder on the Bren server: https://apps.bren.ucsb.edu/IE-Data/.

References

"Industrial Assessment Centers." Retrieved October 3, 2024, from https://iac.university/#database.

Industrial Assessment Center. (2007). "Assessment Recommendation Code." Retrieved February 4, 2025, from https://iac.university/technicalDocs/ARC_list_9.1.pdf.

U.S. Energy Information Administration. (n.d.). *Annual electric power industry emissions data* [Data set]. U.S. Energy Information Administration. Retrieved February 4, 2025, from https://www.eia.gov/electricity/data/state/emission_annual.xlsx.

U.S. Energy Information Administration. (n.d.). *Annual generation data by state*. U.S. Energy Information Administration. Retrieved February 4, 2025, from https://www.eia.gov/electricity/data/state/annual_generation_state.xls.

U.S. Bureau of Labor Statistics. (n.d.). "Producer Price Index (PPI)." Retrieved February 4, 2025, from https://www.bls.gov/ppi/.

U.S. Environmental Protection Agency. (n.d.). "AP-42, Compilation of air pollutant emissions factors from stationary sources." Accessed February 4, 2025. https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors-stationary-sources.

U.S. Census Bureau. (2022). *2022 NAICS codes listed numerically 2-digit through 6-digit* [Data set]. Retrieved February 4, 2025, from https://www.naics.com/search/.

U.S. Census Bureau. (2022). *2022 NAICS to SIC crosswalk* [Data set]. Retrieved February 4, 2025, from https://www.naics.com/search/.

Appendices

Appendix A. Fuel Emissions Factors

IAC Energy Source Type	IAC Fuel Code ID	Description	CO_2, kg/MMBtu	SO_2, kg/MMBtu	NO_x, kg/MMBtu
Natural Gas	E2	Natural gas combustion	53.06	0.00026682	0.078933902
L.P.G	E3	Liquified petroleum gas combustion	61.71	0.00671968	0.066085045
#1 Fuel Oil	E4	Fuel Oil Combustion	75.2	0.07821222	0.055079029
#2 Fuel Oil	E5	Fuel Oil Combustion	73.96	0.02300359	0.055079029
#4 Fuel Oil	E6	Average	74.75	0.28422219	0.073258708
#6 Fuel Oil	E7	Fuel Oil combustion	75.1	0.75145075	0.109618067
Coal	E8	Bituminous and subbituminous coal combustion	94.67	0.78328289	0.109557215
Wood	E9	Wood residue combustion in boilers	93.8	0.20571066	0.140589569
Paper	E10	Same as Wood	93.8	0.20571066	0.140589569
Other Gas	E11	Same as LPG	66.72	0.00671968	0.066085045
Other Energy	E12	Average of all emission factor values	76.28	0.23452991	0.089487518

Appendix B. Integrated Dataset Field Definitions

Integrated Dataset: https://apps.bren.ucsb.edu/IE-Data/final/

The data dictionary lists all fields in the integrated dataset along with their corresponding definitions.

Field	Primary Data Source	Definition
id	IAC. Assess Table	Assessment ID
center	IAC. Assess Table	IAC Center Code
fy	IAC. Assess Table	The fiscal year in which the assessment was conducted
sic	IAC. Assess Table	SIC industrial classification code

naics	IAC. Assess Table	NAICS industrial classification code
state	IAC. Assess Table	US State abbreviation
sales	IAC. Assess Table	Total yearly sales
employees	IAC. Assess Table	Total site employees
plant_area	IAC. Assess Table	Total plant square footage
products	IAC. Assess Table	Types of products
produnits	IAC. Assess Table	Production level units
prodlevel	IAC. Assess Table	Total yearly production
prodhours	IAC. Assess Table	Total yearly hours of operation
numars	IAC. Assess Table	Total number of recommendations
sourccode	IAC. Assess Table	Type of energy source
plant_cost	IAC. Assess Table	Total yearly electricity consumption costs ($)
plant_usage	IAC. Assess Table	Total yearly electricity consumption (kWh)
superid	IAC. Recc Table	Assessment ID + Recommendation Number
ar_number	IAC. Recc Table	Recommendation Number
appcode	IAC. Recc Table	Application Code
arc2	IAC. Recc Table	IAC Assessment Recommendation Code
impstatus	IAC. Recc Table	Implementation Status (I = implemented, N= Not Implement)
impcost	IAC. Recc Table	Total implementation cost
source_rank	IAC. Recc Table	Rank of the energy source (Primary, Secondary, Tertiary, Quaternary)
conserved	IAC. Recc Table	The amount of energy resource conserved
sourconsv	IAC. Recc Table	Energy consumed at the source needed to produce consumed energy at site
saved	IAC. Recc Table	Cost savings
rebate	IAC. Recc Table	Was a rebate involved (yes/no)
incremntal	IAC. Recc Table	Was the recommendation implemented incrementally (yes/no)
ic_capital	IAC. Recc Table	Captial component of implementation cost
ic_other	IAC. Recc Table	Other componenet of implementation cost
payback	IAC. Recc Table	Simple Payback (years)
bptool	IAC. Recc Table	What best practice tools was used (if any)
payback_imputed	New - Added During Capstone Project	Estimated payback value, calculated as saved / impcost_adj.

emissions_avoided	New - Added During Capstone Project	Estimated emissions avoided, calculated as Emission Factor × Conserved.
emission_type	New - Added During Capstone Project	Type of emitted pollutant (e.g., CO_2, SO_2, NO_x).
emission_factor_units	New - Added During Capstone Project	Units associated with the emission factor value.
emission_factor	New - Added During Capstone Project	Emission factor used for calculating avoided emissions.
base_year	New - Added During Capstone Project	Base year for the recommendation, used to adjust implementation costs using the Producer Price Index (PPI).
reference_year	New - Added During Capstone Project	Latest available PPI year used to adjust implementation costs.
reference_ppi	New - Added During Capstone Project	PPI value for the recommendation category in the reference year.
base_ppi	New - Added During Capstone Project	PPI value for the recommendation category in the base year.
impcost_adj	New - Added During Capstone Project	Implementation cost adjusted to the base year using PPI.
main_code	New - Added During Capstone Project	Main ARC code (e.g., 2.3); 2-digit recommendation category code.
main_description	New - Added During Capstone Project	Description of the main ARC code (e.g., *Electrical Power*).
sub_code	New - Added During Capstone Project	Sub ARC code (e.g., 2.32); 3-digit recommendation category code.
sub_description	New - Added During Capstone Project	Description of the sub ARC code (e.g., *Power Factor*).
detail_code	New - Added During Capstone Project	Detail ARC code (e.g., 2.321); 4-digit recommendation category code.
detail_description	New - Added During Capstone Project	Description of the detail ARC code (e.g., *General*).
specific_code	New - Added During Capstone Project	Specific ARC code (e.g., 2.3212); 5-digit detailed recommendation code.
specific_description	New - Added During Capstone Project	Description of the specific ARC code (e.g., *Optimize Plant Power Factor*).
naics_imputed	New - Added During Capstone Project	NAICS code filled by looking up SIC code when NAICS was missing; otherwise, retains the original NAICS code.
naics_description	New - Added During Capstone Project	Industry description corresponding to the NAICS code.