AUTOMATING DATA COLLECTION AND ANALYSIS FOR SOLAR ENERGY INITIATIVES

David Dai ‘27, Kelli Eng ‘26, Jenny Le ‘26, Ruth Tilahun ‘26, Gioanna Zhao ‘26
Field Site: Philadelphia Solar Energy Association
Field Advisors: Liz Robinson, Ron Celentano, Faculty Advisor: Jennifer Spohrer

Background

Founded in 1980, the Philadelphia Solar Energy Association (PSEA) is a non-profit with the goal of promoting solar energy usage in Pennsylvania through education, community outreach, and policy advocacy. From organizing community events to supporting clean energy legislation, PSEA empowers future innovators and plays an important role in shaping the state’s future with sustainable energy.

Although interest in solar power is growing, maintaining up-to-date, well-organized data on solar installations has been a persistent challenge. PSEA relied heavily on one person to manually gather and compile data from scattered, inconsistently formatted sources. This time-consuming process made it difficult to update visual materials regularly or respond quickly to new developments. This project aimed to streamline this process and make it significantly easier to organize and update solar data, allowing PSEA to refresh their outreach graphics more frequently and make their resources more impactful and accessible.

Objectives

In order to create solutions for PSEA’s needs, this project aimed to:

- Develop a sustainable process for gathering relevant solar installation data
- Ensure data is structured, consistent, and usable for long-term reference
- Create clear, interpretable visualizations to support analysis and communication
- Build a system that allows PSEA to efficiently update and rebuild visualizations as public source data changes

Responsibilities

- Find a suitable code environment that runs the final deliverable for PSEA
- Write scripts in Python that gather data from public sources (AEPS, SEIA, PJM) and generate visuals
- Write documentation on set-up and implementation of code to maintain data longevity and access
- Touch base with PSEA biweekly to present development of project and reassess goals
- Organize project deliverables into a consolidated drive to transfer ownership access

Methodology

- Download data from SEIA, AEPS, and PJM
- Filter and clean the data
- Generate plots that convey relevant information on solar energy (histograms, bubble plots, bar graphs, maps)
- Present final deliverable in the form of a Jupyter Notebook file
- Deploy solar visualizations to PSEA’s website: www.phillysolar.org
- Organize project components in Google Drive

Visualization generated by automated Python script

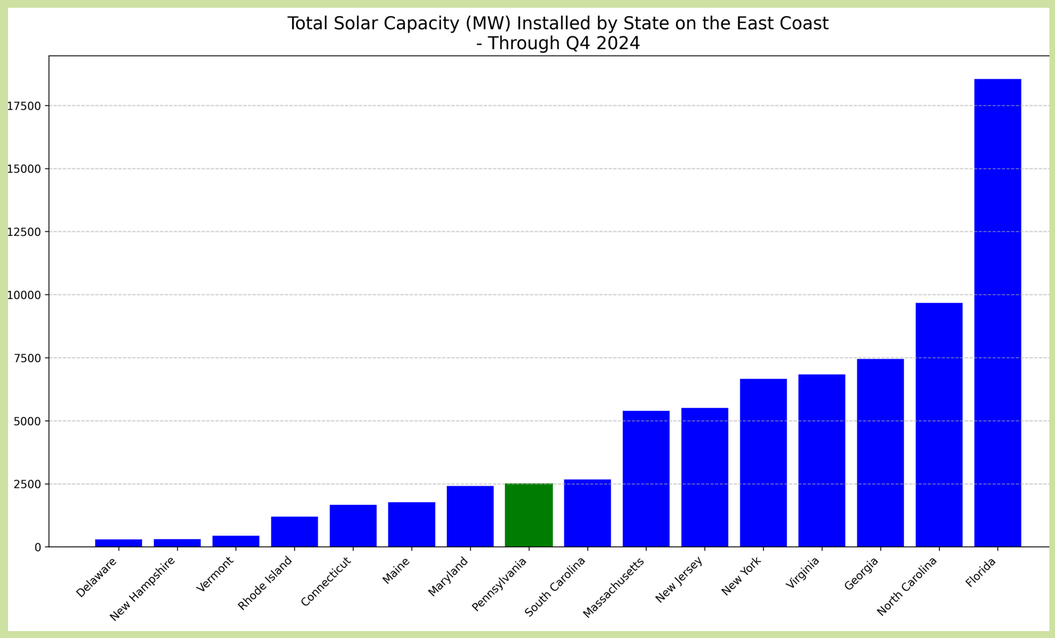

Figure 1: Total Solar Capacity (MW) Installed by State on the East Coast - Through Q4 2024

Visualization generated by automated Python script

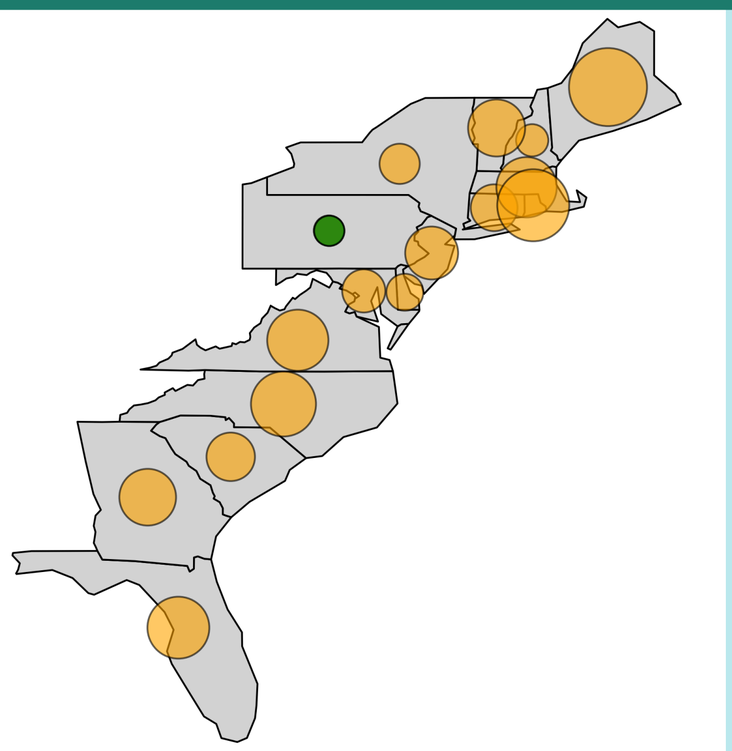

Figure 2: Total Solar PV Installed per Capita (MW/Million) - Through Q4 2024

Key Takeaways

- Adapted the project approach by experimenting with multiple tools and methods to find solutions that balanced functionality with usability
- Learned to navigate challenges related to inconsistent data formats, website structure changes, and privacy considerations in web scraping
- Saw the impact of technical work in a real-world context, contributing to broader sustainability and advocacy efforts in PA

References: seia.org, https://pennaeps.com/reports/, pjm.com
Tools/Libraries: Jupyter Notebook, Python, Request, Re, Pandas, Matplotlib

PSEA

BRYN MAWR COLLEGE