Week 0 – Solar Challenge Report

Name: Chalie Lijalem

Submission Date: 5/21/2025

Solar Potential Assessment: Cross-Country Comparative Study

Project Overview

This project is part of **Week 0 of the 10 Academy Solar Challenge**, with the overarching aim of assessing solar irradiance potential across three West African countries—**Benin**, **Togo**, and **Sierra Leone**. By combining data cleaning, exploratory data analysis (EDA), cross-country comparison, and dashboard development, the project lays the foundation for data-driven decision-making in solar energy planning. It reflects not only on technical execution but also on strategic thinking and reproducibility of insights.

Data Cleaning and Preprocessing

The project began with meticulous cleaning and transformation of raw solar irradiance data. I created reusable functions like `impute_median`, `remove_duplicates`, and `parse_datetime_column`, stored in `src/utils.py`, to standardize the process. These functions ensure that missing values are properly filled (using median imputation where necessary), datetime fields are correctly parsed, and duplicates are removed.

Each cleaned dataset was exported as a `.csv` file and saved in the `data/` directory under standardized filenames such as `benin_clean.csv`, `togo_clean.csv`, and `sierra_leone_clean.csv`. This preprocessing pipeline guarantees data consistency and supports reproducibility across notebooks and applications.

Exploratory Data Analysis (EDA)

To gain a better understanding of each country's solar characteristics, I performed **EDA** in separate country-level notebooks. This phase involved:

- **Time-series visualization** of GHI, DNI, and DHI values across months to assess seasonal patterns.
- **Histograms** and **boxplots** for distribution analysis of irradiance values.
- **Correlation heatmaps** to explore relationships between solar variables.
- **Missing value maps** to inspect data quality.

Key insights from EDA revealed that **Togo** has the broadest seasonal spread in GHI, suggesting both opportunity and volatility. **Benin** demonstrated relatively consistent solar behavior with fewer outliers, whereas **Sierra Leone** showed the lowest GHI but also less variability—potentially ideal for stable but modest solar generation.

Cross-Country Comparison

To synthesize findings across the three datasets, a dedicated notebook (`compare_countries.ipynb`) was created. This notebook focuses on:

- **Boxplots**: GHI, DNI, and DHI values were visualized using side-by-side boxplots, highlighting inter-country variation.
- **Summary Table**: A comparative table with **mean**, **median**, and **standard deviation** for each metric across countries was created.
- **Statistical Testing**: The **Kruskal–Wallis H-test** was applied to GHI values, providing a non-parametric alternative to ANOVA. The low p-values indicated significant differences in solar potential between at least two of the countries.

Key Observations

- **Togo** has the **highest median GHI** but also the **widest spread**, suggesting opportunity with risk.
- **Benin** shows **moderate irradiance values** with less extreme fluctuations.
- **Sierra Leone** ranks lowest in GHI but exhibits **consistent** values across time.

A **bar chart** summarizing average GHI per country was also included as a visual wrap-up.

Interactive Dashboard

To make the analysis more interactive and accessible, I developed a **Streamlit dashboard** under the `app/` directory. Key features of the dashboard include:

- **Dropdown widget** to select one or multiple countries.
- **Dynamic boxplots** displaying the selected irradiance metric by country.
- **Summary stats table** that updates based on selected filters.
- **Bar charts** showcasing average GHI rankings.

All data processing and visualization functions were modularized in `app/utils.py`, promoting code reusability. The dashboard can be launched using:

```
bash
CopyEdit
streamlit run app/main.py
```

The dashboard's interface is clean and intuitive, designed to support both technical and non-technical users in understanding solar patterns across the region.

Folder Structure and Git Hygiene

A consistent and modular folder structure was maintained:

```
bash
CopyEdit
├── data/                       # Cleaned data
├── notebooks/                 # Jupyter notebooks (EDA and comparison)
│   └── compare_countries.ipynb
├── src/
│   └── utils.py               # Reusable preprocessing functions
├── app/                       # Streamlit dashboard
│   ├── main.py
│   └── utils.py
└── scripts/
    └── README.md
```

Version control was followed diligently using feature branches (`compare-countries`, `dashboard-dev`) and atomic commits like:

```
bash
CopyEdit
feat: add boxplot comparison for solar metrics across countries
feat: implement Kruskal-Wallis test for GHI
feat: basic Streamlit UI for solar dashboard
```

Reflections and Challenges

While the project objectives were met, there are a few areas for improvement:

- **Task Prioritization**: The dashboard implementation could have started earlier, allowing more time for enhancements like map visualizations or region-level breakdowns.
- **Feasibility Planning**: Pre-defining expected CSV structure or outlier thresholds might have helped catch anomalies earlier in the pipeline.
- **Risk Assessment**: Greater focus on proactively identifying dependencies—such as data availability for specific columns—would have reduced friction during later stages.

Key Takeaways and Next Steps

This project demonstrates a solid application of data science for real-world impact in renewable energy. It combines preprocessing, EDA, statistical reasoning, and interactive visualization.

To enhance future reports and deliverables, I plan to:

- Implement **risk identification frameworks** at the project start.
- Add **unit testing** for utility functions.
- Deploy the dashboard to **Streamlit Community Cloud** for broader accessibility.
- Refactor for greater **modularity and scalability**, supporting potential multi-year or multi-variable extensions.