Final Report

1. Project Overview

The Solar Data Analysis Challenge focused on analyzing large-scale solar radiation datasets from **Benin, Sierra Leone, and Togo**. The goal was to develop data engineering, exploratory data analysis (EDA), statistical testing, and interactive visualization skills while generating actionable insights to guide MoonLight Energy Solutions' solar installation strategy across West Africa.

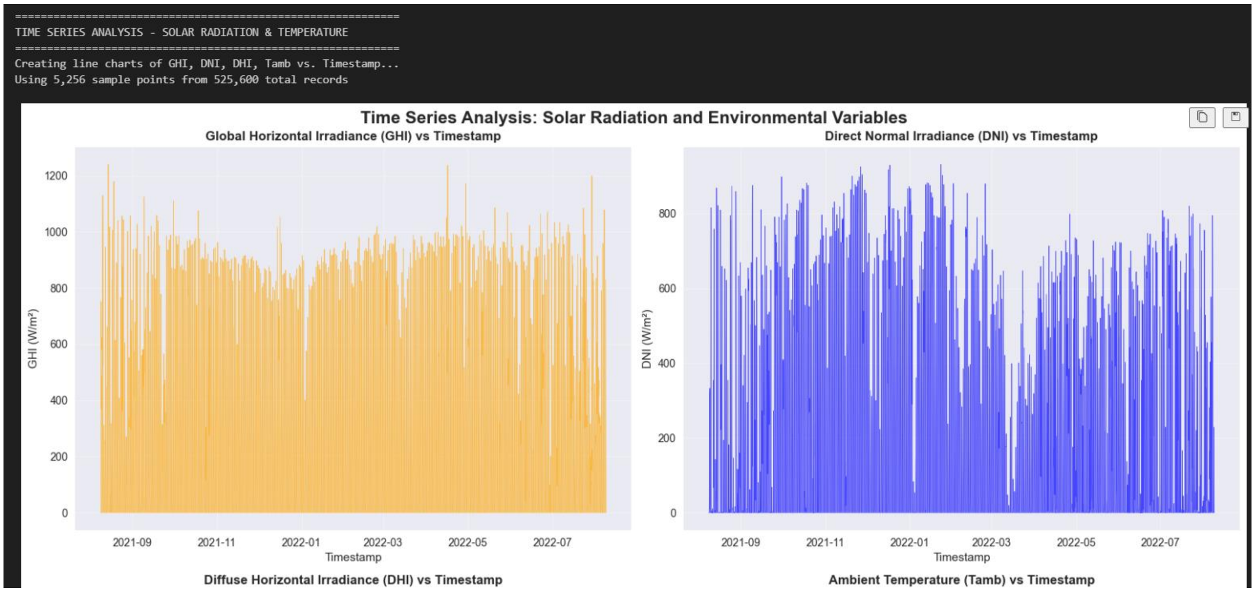
This report summarizes the full project lifecycle, including methodologies, data cleaning, cross-country comparison, and recommendations aligned with the company's long-term sustainability and growth objectives.

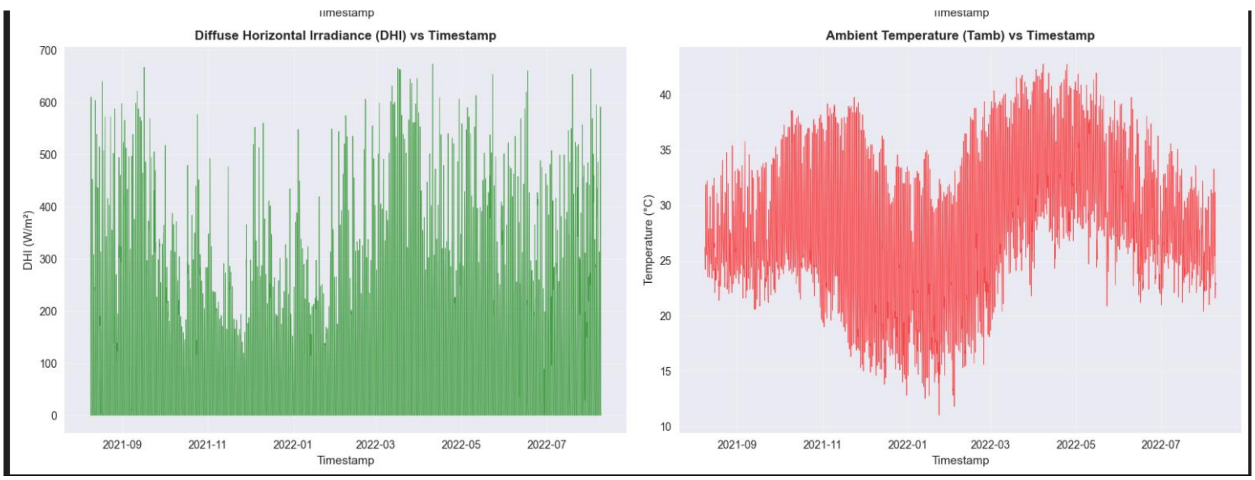
2. Environment Setup and Version Control

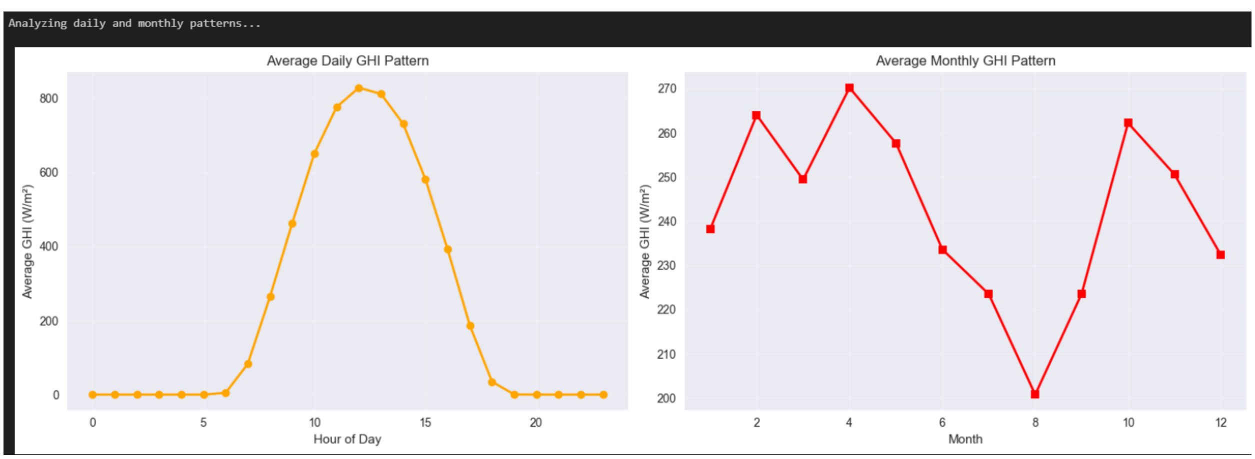
- Initialized a GitHub repository to ensure reproducibility, collaboration, and code versioning.
- Configured a Python 3.13 virtual environment with key libraries: pandas, numpy, matplotlib, scipy, Streamlit, etc.
- Established .gitignore and requirements.txt for clean repo management and environment replication.
- Integrated GitHub Actions CI/CD workflows for automated testing and code quality checks.
- Documented environment setup comprehensively in README for onboarding and reproducibility.

3. Data Profiling, Cleaning & Exploratory Data Analysis (EDA)

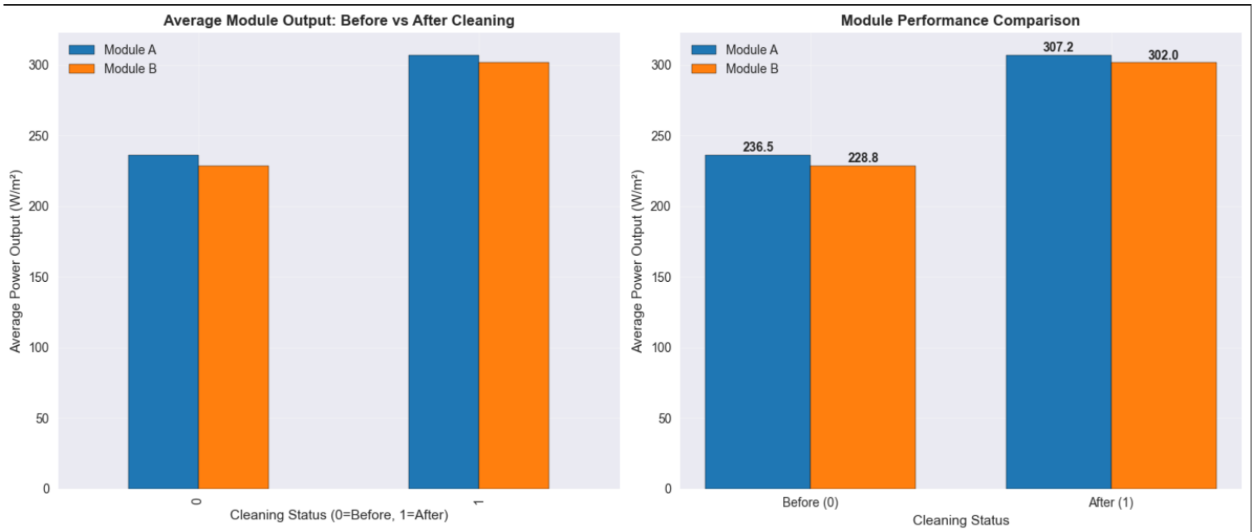
- Processed Benin's dataset (~525,600 records, 19 variables), followed by Sierra Leone and Togo datasets with similar rigor.
- Corrected negative solar radiation values (up to 49% in some datasets) by replacing with zeros or median values to maintain physical validity.



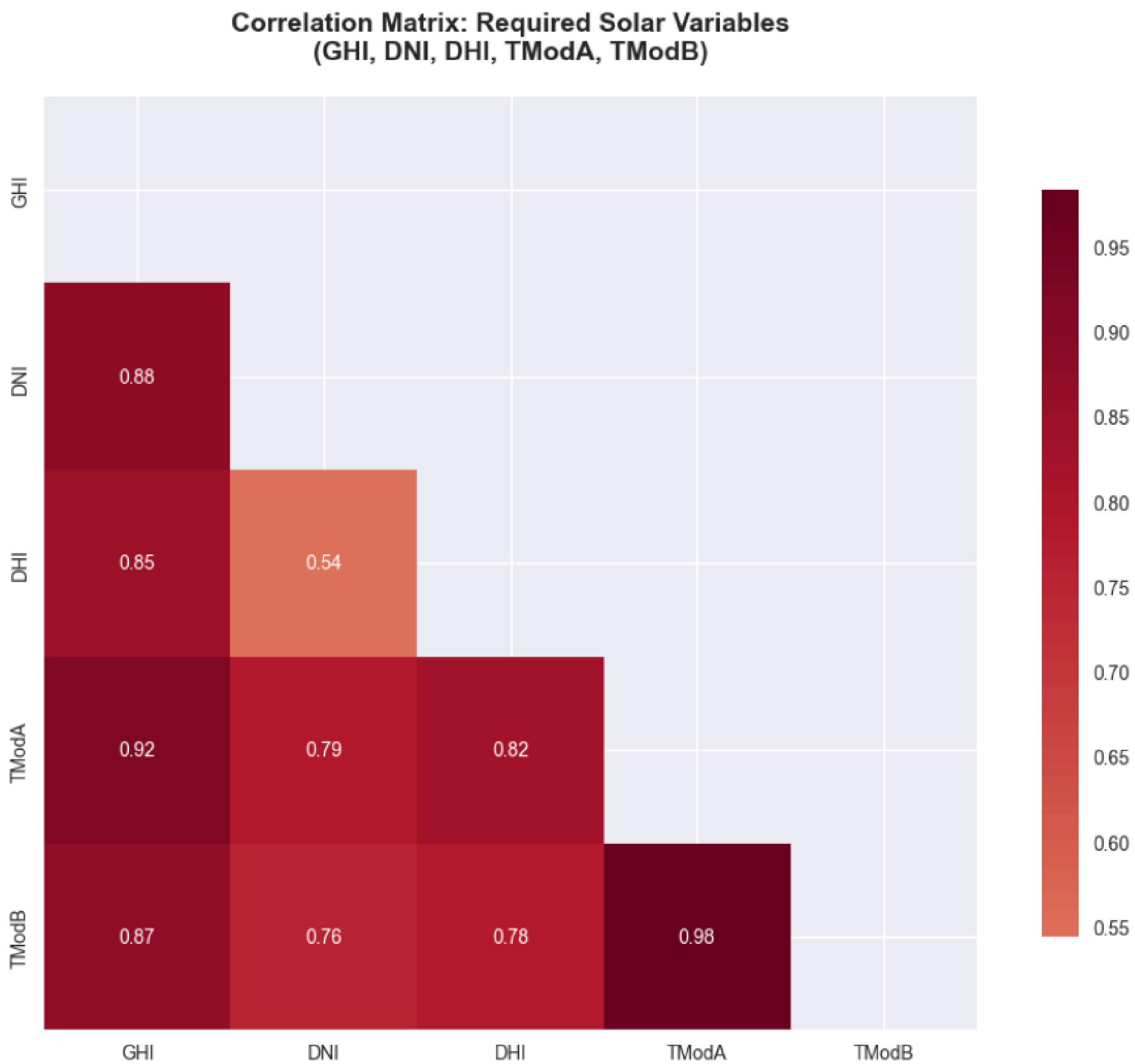




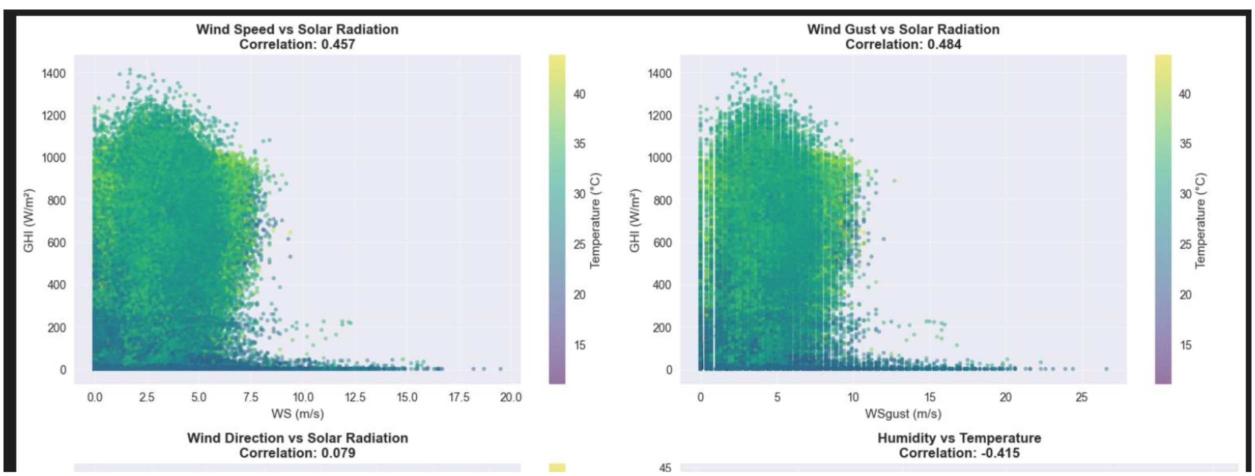
- Removed statistical outliers (Z-score threshold |Z| > 3) on solar and environmental variables to reduce noise.



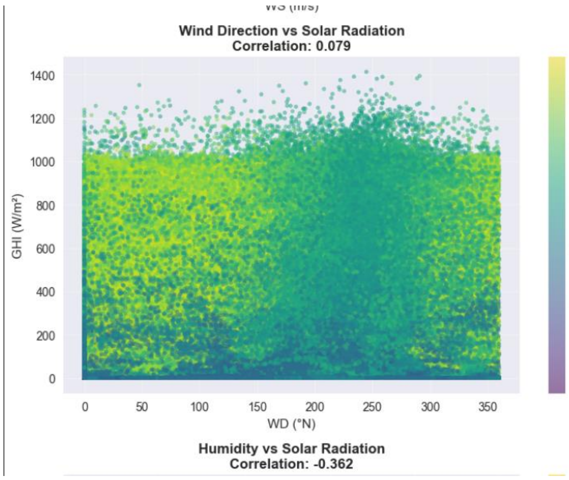
- Created temporal line charts highlighting diurnal and seasonal solar radiation patterns.
- Generated correlation heatmaps revealing relationships among GHI, temperature, humidity, wind speed, and module performance.

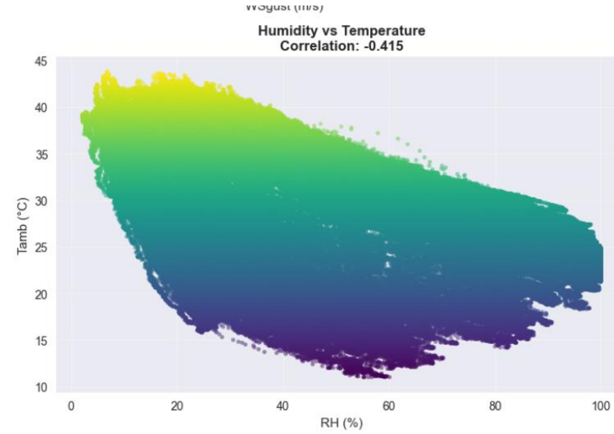


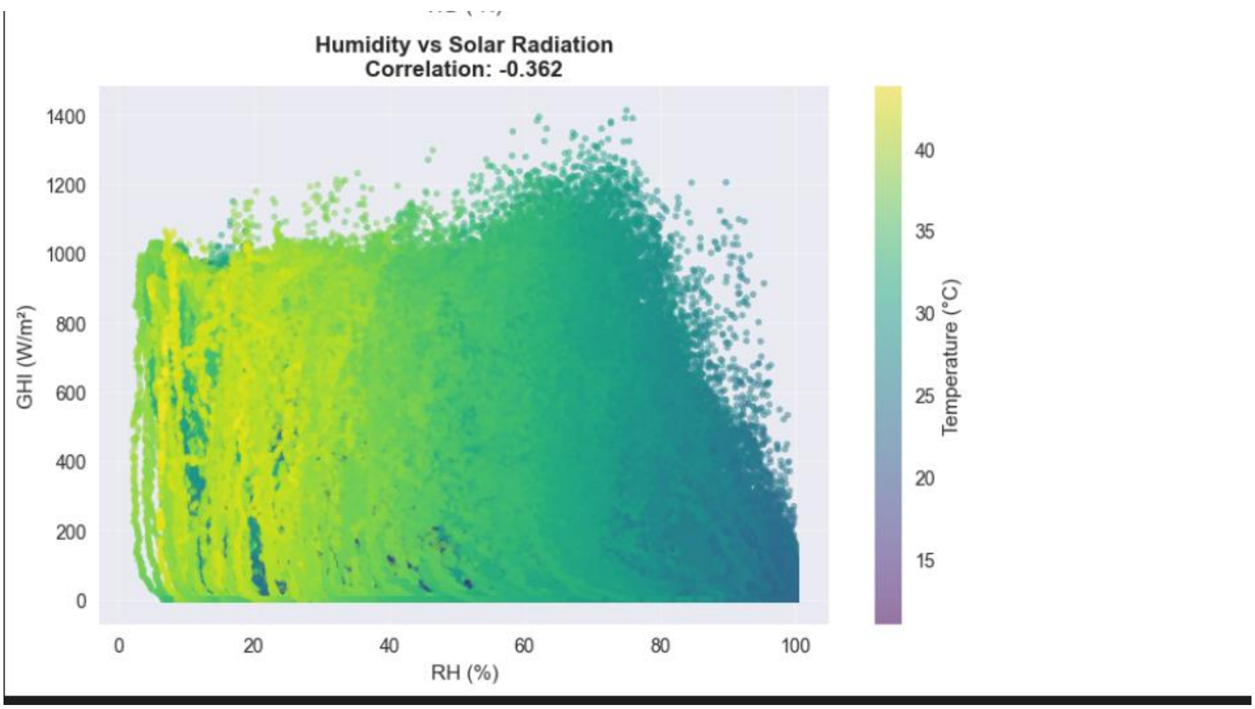
Correlation Matrix: Required Solar Variables
(GHI, DNI, DHI, TModA, TModB)
GHI
DNI
DHI
TModA
TModB
0.88
0.85
0.54
0.92
0.79
0.82
0.87
0.76
0.78
0.98
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55

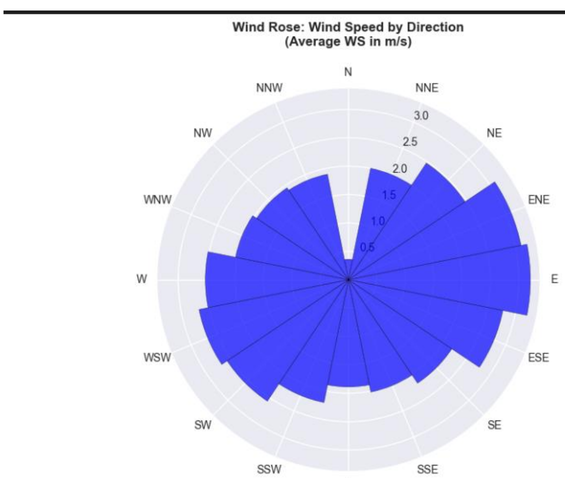


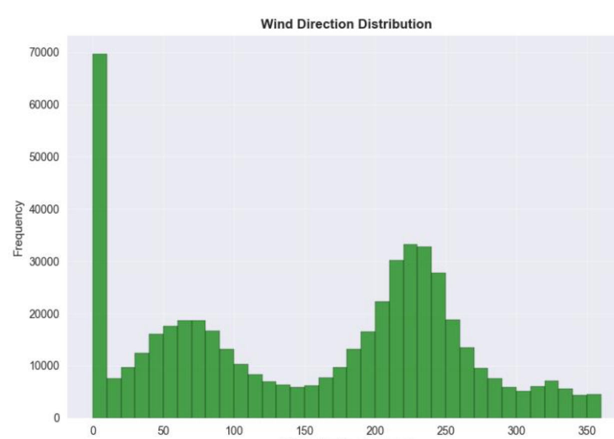
Wind Speed vs Solar Radiation
Correlation: 0.457
GHI (W/m²)
WS (m/s)
Temperature (°C)
Wind Gust vs Solar Radiation
Correlation: 0.484
WSgust (m/s)
Wind Direction vs Solar Radiation
Correlation: 0.079
Humidity vs Temperature
Correlation: -0.415

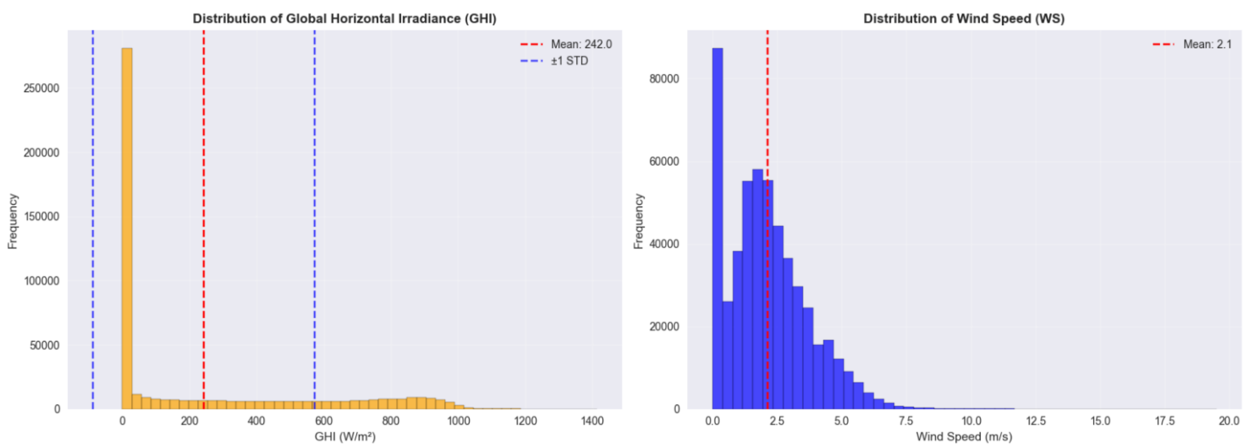

WS (m/s)
Wind Direction vs Solar Radiation
Correlation: 0.079
GHI (W/m²)
WD (°N)
Temperature (°C)
Humidity vs Solar Radiation
Correlation: -0.362

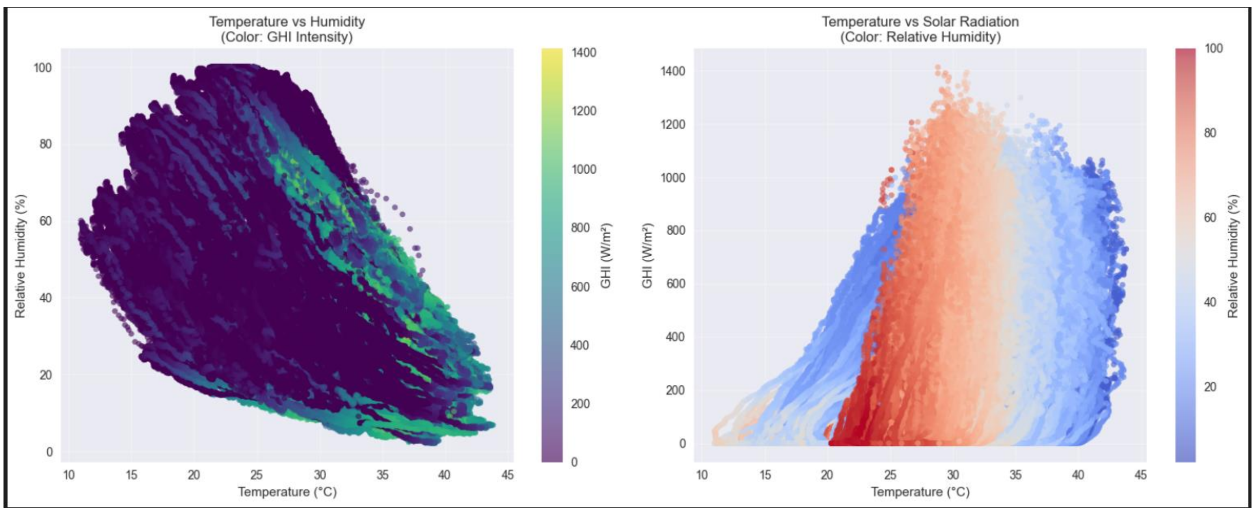

WSgust (m/s)
Humidity vs Temperature
Correlation: -0.415
Tamb (°C)
RH (%)

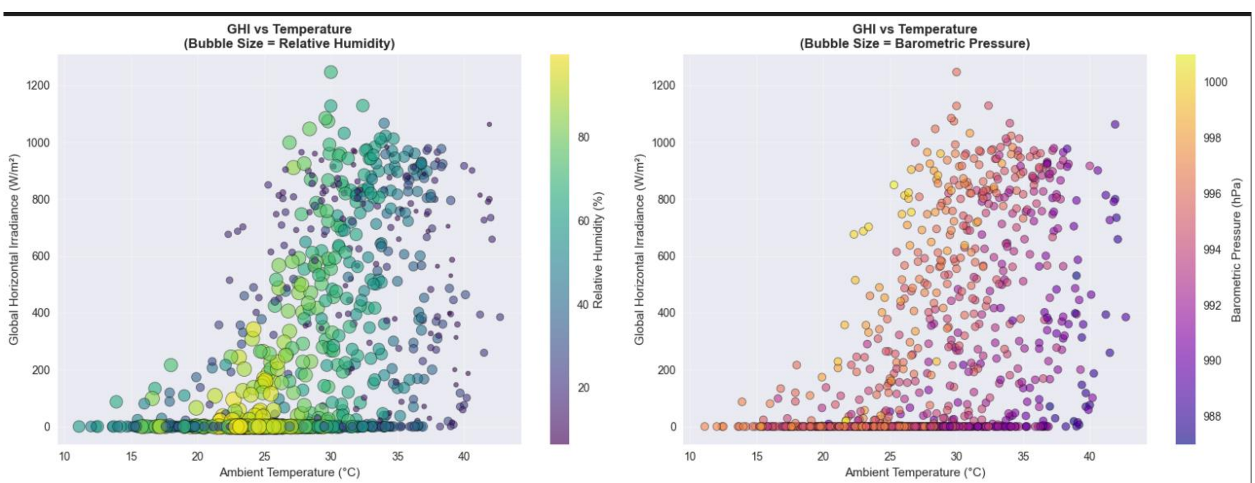

Humidity vs Solar Radiation
Correlation: -0.362
GHI (W/m²)
RH (%)
Temperature (°C)


Wind Rose: Wind Speed by Direction
(Average WS in m/s)
N
NNE
NE
ENE
E
ESE
SE
SSE
S
SSW
SW
WSW
W
WNW
NW
NNW


Wind Direction Distribution
Frequency





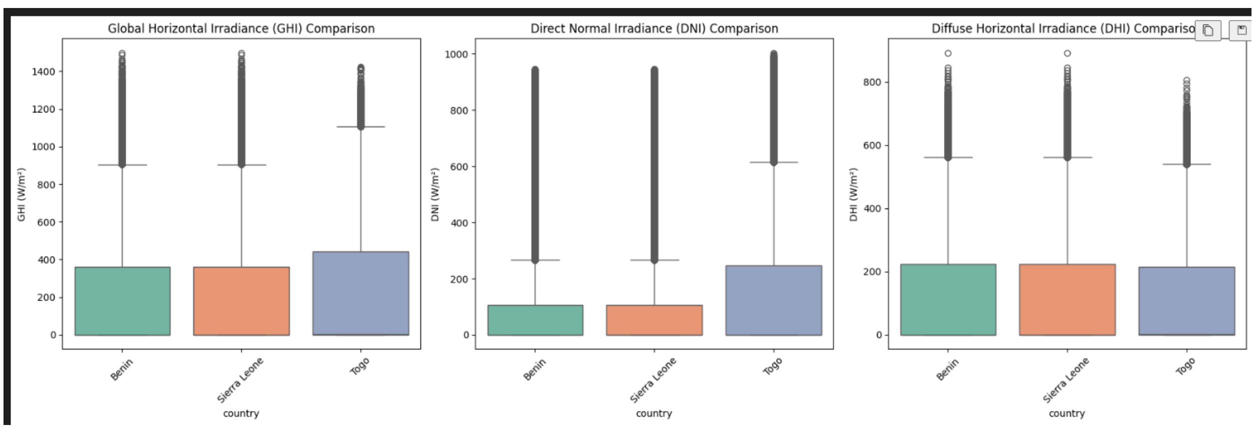


- Exported cleaned datasets for downstream analysis, ensuring consistency and quality.

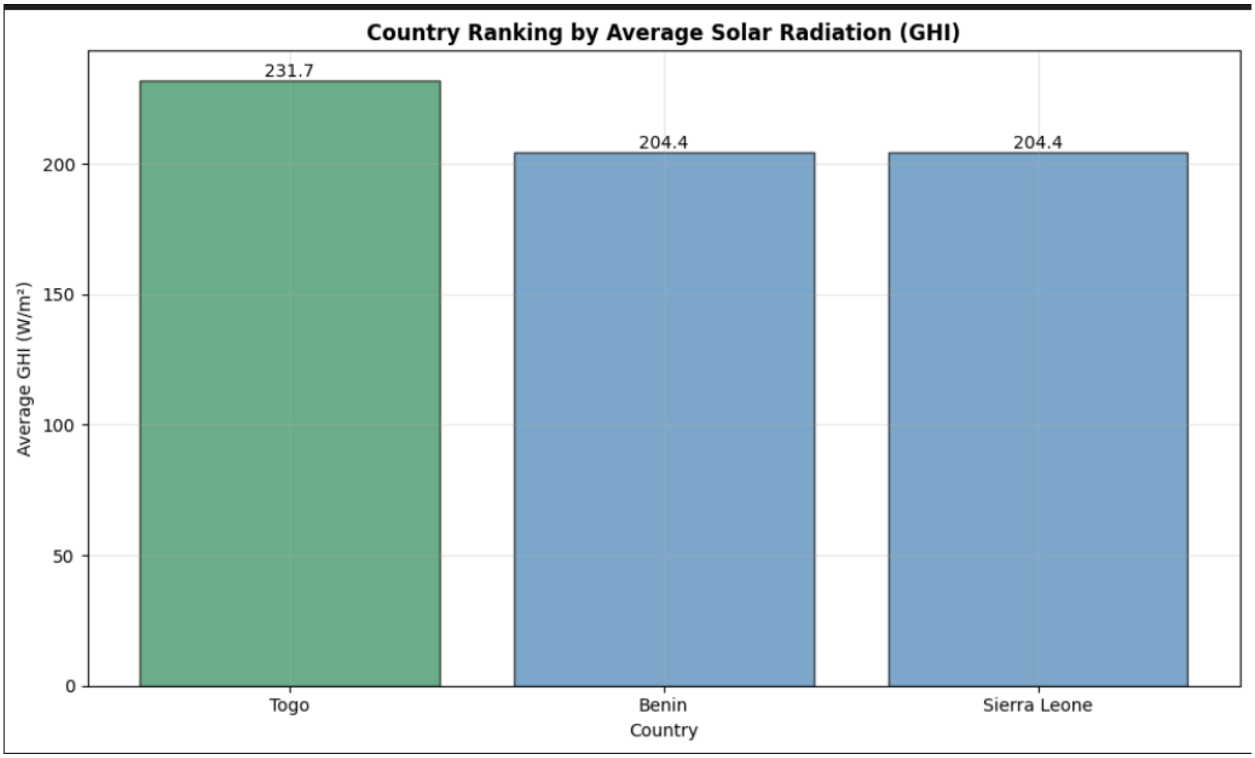
4. Cross-Country Comparison and Statistical Testing

Metric	Benin	Sierra Leone	Togo	Highest Median	Most Stable (Lowest Std)
GHI (W/m²)	204.4	204.4	**231.7**	Togo (2.1)	Benin (296.8)
DNI (W/m²)	116.5	116.5	**151.3**	Togo (0.0)	Benin (218.6)
DHI (W/m²)	116.3	116.3	116.4	Togo (2.5)	Benin (157.0)

- **Statistical tests (ANOVA and Kruskal-Wallis)** confirmed **significant differences** ($p < 0.05$) in solar radiation profiles between countries.
- **Togo shows the highest solar potential** based on mean and median Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI), beneficial for both photovoltaic (PV) and concentrating solar power (CSP) systems.
- **Benin demonstrates the most consistent solar radiation** (lowest standard deviation), which is advantageous for stable energy yield and grid integration.



- The **solar radiation is positively correlated with temperature** ($r \approx 0.56$), and negatively correlated with humidity ($r \approx -0.26$), highlighting climate dependencies critical for solar system performance optimization.



5) Interactive Dashboard Development

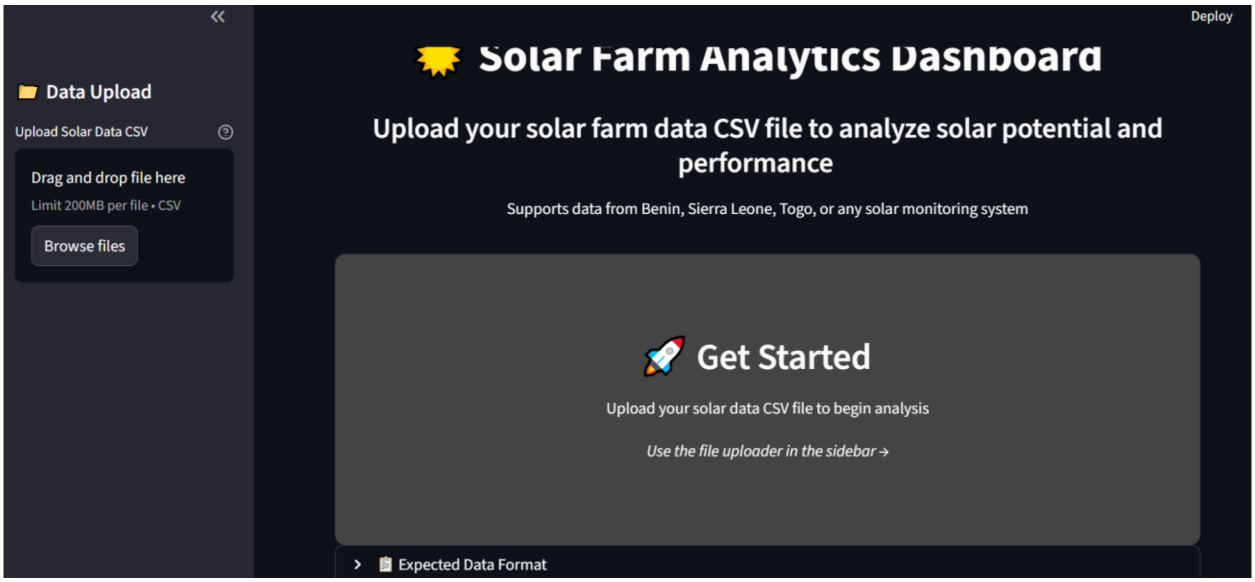
- Built a user-friendly Streamlit dashboard to enable dynamic exploration of solar metrics across countries.

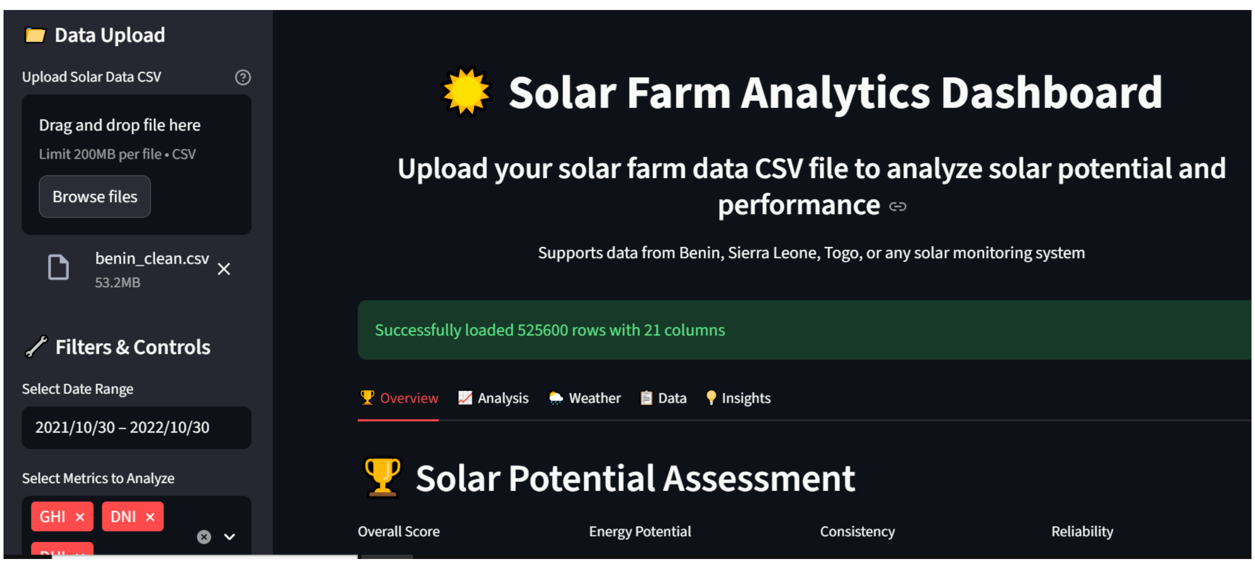
dynamic, user-friendly dashboard that brings solar data to life:

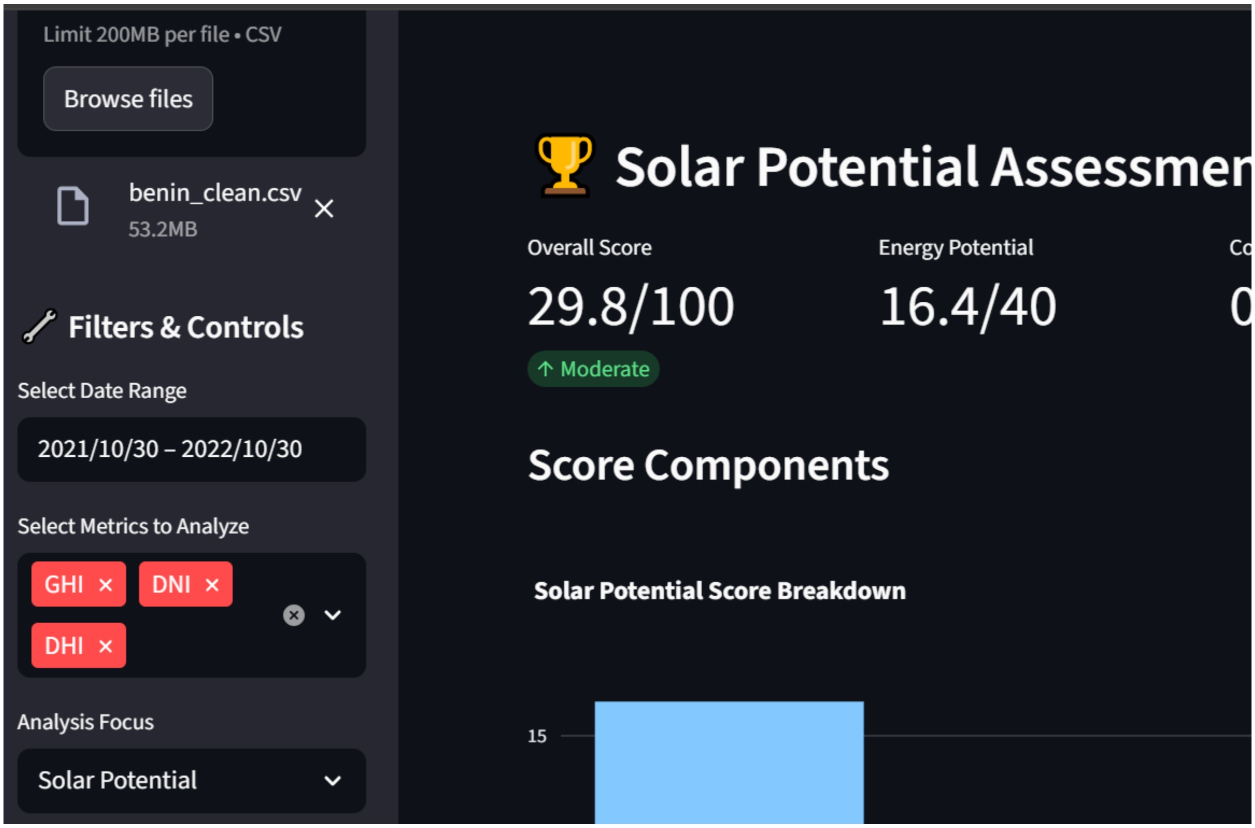
- **Country-Level Analysis**: Easily select Benin, Sierra Leone, or Togo to explore detailed solar radiation metrics.
- **Interactive Visualizations**: Dynamic boxplots, summary statistics, and temporal trends enable intuitive comparison of solar metrics such as GHI, DNI, and Diffuse Horizontal Irradiance (DHI).
- **Real-Time Insights**: The dashboard supports drill-down analysis, helping stakeholders visualize not only average solar potential but also variability and consistency — key factors for risk management and project design.
- **Future-Proofing**: Designed with scalability in mind, this tool can be expanded to include additional countries and predictive modeling, supporting strategic planning and portfolio optimization.
- Key features included:
    - Country selection dropdown widget.

- Interactive boxplots for GHI, DNI, and DHI metrics.
- Summary tables presenting descriptive statistics.
- Planned enhancements for time-series charts, bubble plots, and downloadable reports.

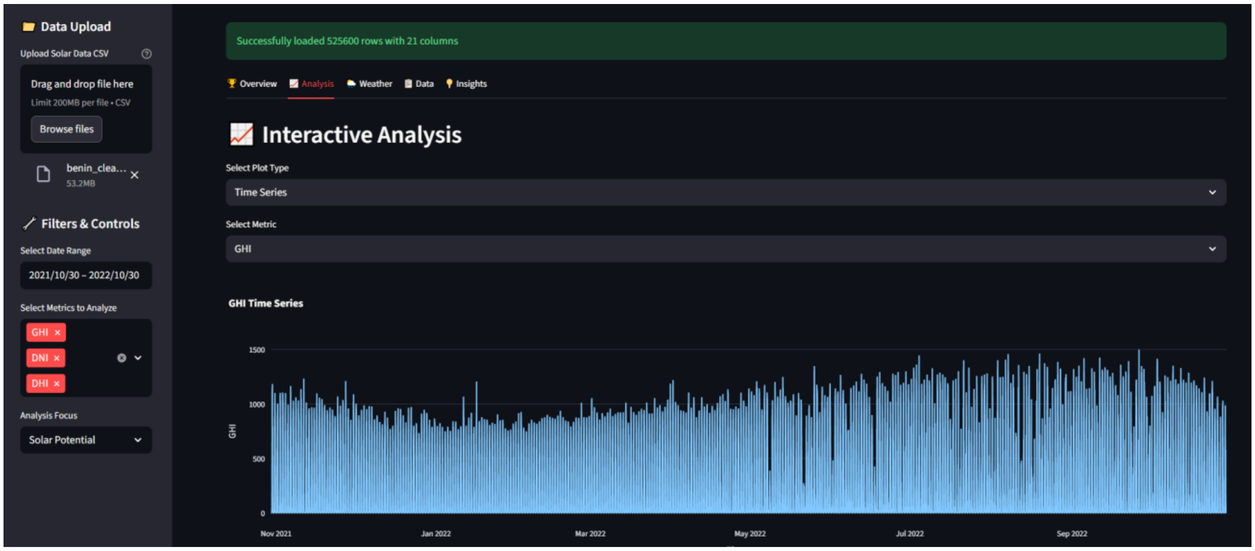
- Addressed challenges with large file handling and optimized data loading to maintain dashboard responsiveness.
- Prepared dashboard for deployment on Streamlit Community Cloud to facilitate broader accessibility.



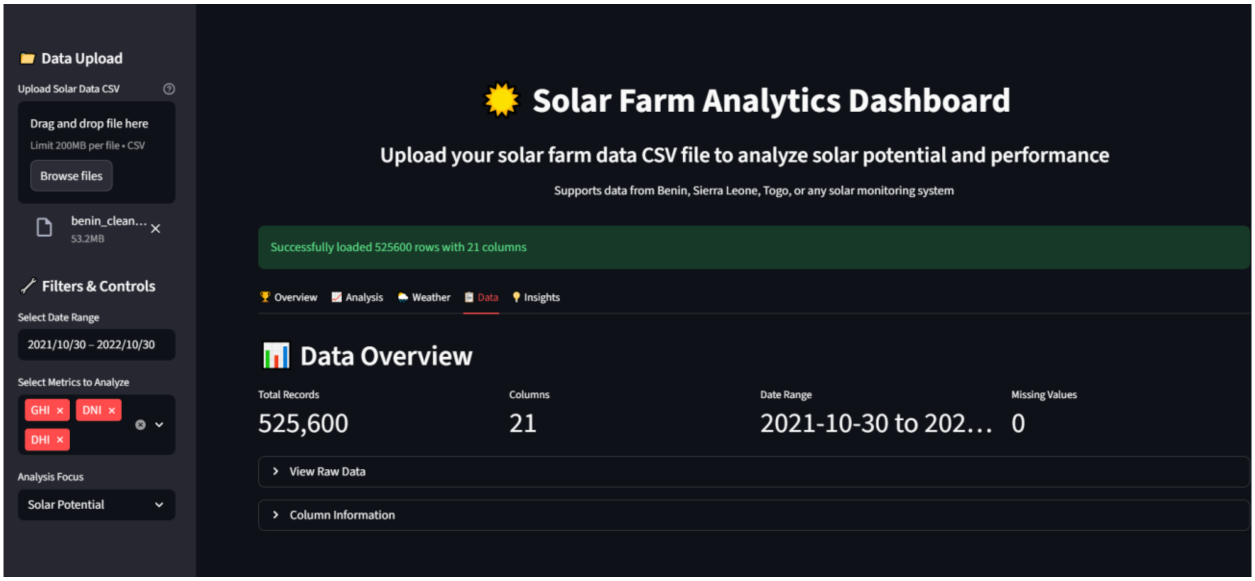


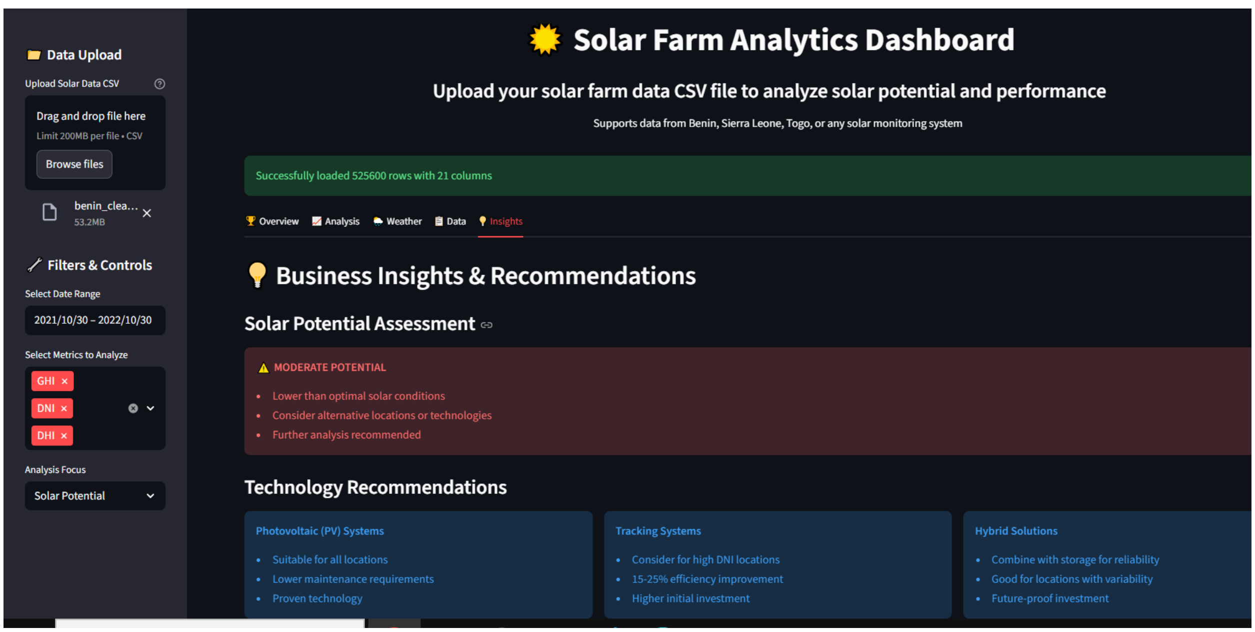

Limit 200MB per file • CSV
Browse files
benin_clean.csv
53.2MB
Filters & Controls
Select Date Range
2021/10/30 – 2022/10/30
Select Metrics to Analyze
GHI
DNI
DHI
Analysis Focus
Solar Potential
Solar Potential Assessmen
Overall Score
29.8/100
Moderate
Energy Potential
16.4/40
Score Components
Solar Potential Score Breakdown
15


15
10
5
0
energy
consistency
reliability
weather
Average GHI
204.4 W/m²
Average Humidity
79.4 %
Average Temperature
26.3 °C
Average Wind Speed
1.1 m/s


Data Upload
Upload Solar Data CSV
Drag and drop file here
Limit 200MB per file • CSV
Browse files
benin_clea...
53.2MB
Filters & Controls
Select Date Range
2021/10/30 – 2022/10/30
Select Metrics to Analyze
GHI
DNI
DHI
Analysis Focus
Solar Potential
Successfully loaded 525600 rows with 21 columns
Overview
Analysis
Weather
Data
Insights
Interactive Analysis
Select Plot Type
Time Series
Select Metric
GHI
GHI Time Series
1500
1000
500
0
GHI
Nov 2021
Jan 2022
Mar 2022
May 2022
Jul 2022
Sep 2022


Overview
Analysis
Weather
Data
Insights
Weather Impact Analysis
Select Weather Variable
Tamb
GHI vs Tamb
1500
1000
500
0
GHI (W/m²)
15
20
25
30
35
40
Tamb (°C)





6). Key Findings from Cleaning and Performance Impact Analysis

- Cleaning solar modules substantially improved output, with Togo seeing over **+136% increase** post-cleaning, emphasizing the importance of maintenance in dusty or soiling-prone regions.
- Data quality improvement through negative value correction and outlier treatment enhanced analysis reliability and module performance monitoring.
- Wind speed positively affects solar radiation by clearing atmospheric particulates but requires structural considerations for panel mounting.

7). Challenges and Solutions

- **Dependency conflicts** managed through environment version pinning and testing alternatives.
- **Git merge conflicts** mitigated by careful branch management and frequent commits.
- **Data quality issues** such as negative readings addressed with domain-informed cleaning.
- **Timestamp inconsistencies** harmonized for accurate cross-country comparisons.
- **Memory management** optimized for dashboard responsiveness.

8. Recommendations

As an Analytics Engineer at MoonLight Energy Solutions, based on this comprehensive analysis and EDA, I recommend the following strategy to identify and prioritize high-potential solar installation regions while aligning with our sustainability goals:

8.1 Strategic Site Selection

- **Prioritize solar installations in Togo**, given its superior solar radiation potential (highest mean and median GHI and DNI), which supports higher energy yield and scalability, especially for CSP technologies.
- **Consider Benin for projects requiring consistent solar output** where lower variability ensures reliable energy supply, easing integration with grid and storage systems.
- For Sierra Leone, adopt a **balanced approach**, leveraging similarities with Benin's radiation profiles but considering local environmental conditions.

Prioritizing Regions for Solar Installation

Based on the comprehensive statistical analysis and exploratory data analysis (EDA), I strongly recommend prioritizing solar farm development in Togo.

- **Why Togo?**
  Our data reveals that Togo exhibits the highest median and mean Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI) values among the three countries analyzed. This means Togo receives the most abundant and consistent sunlight, providing an ideal environment for both photovoltaic (PV) and concentrated solar power (CSP) systems. The high solar energy potential in Togo translates directly into higher power output and better return on investment, ensuring that solar installations there will generate the maximum possible clean energy.
- **Complementary Opportunity in Benin**
  While Togo leads in raw solar potential, Benin offers the most stable solar radiation with the lowest variability. This stability is crucial for energy planning, grid integration, and operational

reliability. Thus, Benin is also a strong candidate for solar farms where consistent energy delivery is critical.

- **Sierra Leone as a Balanced Option**
  Sierra Leone shows similar solar profiles to Benin, suggesting moderate potential and a balance between stability and solar intensity. Projects here can complement regional expansion plans once core sites in Togo and Benin are established.

8.2 Maintenance and Operational Excellence

- Develop a **standardized cleaning and maintenance schedule**, especially in high-dust areas like Togo, to mitigate soiling losses exceeding 100% reduction in module output before cleaning.
- Implement **real-time monitoring** and sensor calibration protocols to minimize data anomalies and enhance operational visibility.

8.3 Environmental Adaptation and System Design

- Factor in **humidity and wind patterns** in system design—protective coatings and robust mounting structures are recommended to maximize system lifespan and performance.
- Optimize **tilt angles and panel spacing** based on predominant wind directions and solar paths to improve efficiency and durability.

8.4 Data Infrastructure and Analytics Advancement

- Continue improving the **interactive dashboard** to enable stakeholders to explore data dynamically and download actionable reports.
- Expand datasets to include **additional countries and socioeconomic/geographic variables** to better inform holistic solar energy planning.
- Incorporate **predictive analytics and forecasting models** to anticipate solar radiation trends and optimize asset management proactively.

9. Lessons Learned

- Robust **version control and CI/CD pipelines** are foundational to reproducible, collaborative analytics projects.
- Thorough data cleaning is indispensable for ensuring accurate, trustworthy insights.
- Interactive visualization tools greatly enhance communication of complex data to diverse teams.
- Modular, well-documented codebases facilitate maintainability and team collaboration.

10 Limitations

- Some datasets lacked detailed metadata, limiting fine-grained spatial or temporal analysis.
- Dashboard deployment requires further optimization for cloud scalability.
- Socioeconomic and geographical factors were not integrated, which could add valuable context to site suitability assessments.

11. Summary of Outcomes

This project successfully applied data engineering, statistical analysis, and software engineering best practices to uncover actionable insights on solar energy potential in Benin, Sierra Leone, and Togo. The established reproducible environment and CI/CD workflows ensure project longevity and ease of collaboration. The developed interactive dashboard provides an accessible platform for dynamic solar data exploration, supporting MoonLight Energy Solutions' mission to expand sustainable solar energy infrastructure in West Africa.

Prepared by: Tsegay Assefa Kidane
Date: November 11, 2025
Program: 10 Academy – Solar Data Discovery Challenge