Solar Farm Data Analysis Documentation

Done by: Mihret Akalu

Submission Date: December 8, 2024

Solar Farm Data Analysis Documentation

Objective

This document presents an analysis of three solar farm datasets (**Benin**, **Sierra Leone**, and **Togo**) focusing on their descriptive statistics, missing data analysis, anomalies, correlation issues, and memory usage. The goal is to compare these datasets and determine their quality, consistency, and readiness for further analysis.

Data Datasets

The datasets analyzed are related to Global Horizontal Irradiance (GHI)**,** Direct Normal Irradiance (DNI)**,** Diffuse Horizontal Irradiance (DHI)**,** and other related solar radiation metrics.

Datasets Analyzed:

1. Benin Dataset
2. Sierra Leone Dataset
3. Togo Dataset

Analytical Approach

The analysis focused on the following evaluation criteria:

1. **Descriptive Statistics Analysis:** Compare key statistics like mean values for GHI, DNI, and DHI**.**
2. **Missing Data Analysis:** Analyze columns with null entries or anomalies.
3. **Outliers/Anomalies:** Identify any extreme anomalies like negative GHI values.
4. **Memory Usage:** Compare how much memory each dataset consumes.
5. **Correlation Issues:** Analyze anomalies/errors related to string timestamps interfering with correlation computations.

Dataset Analysis

The key metrics evaluated are **Mean GHI, Mean DNI, Mean DHI,** and **standard deviation.**

Table 1: Comparison

Metric	Benin	Sierra Leone	Togo
Total Entries	525,600	525,600	525,600
Mean GHI (Global Irradiance)	240.56	201.96	230.55
Mean DNI (Direct Irradiance)	167.18	116.38	151.26
Mean DHI (Diffuse Irradiance)	115.36	113.72	116.44
Standard Deviation GHI	331.13	298.49	322.53
Max GHI Value	1413.0	1499.0	1424.0

Observations

- Benin has the **highest average GHI values** (240.56) compared to the other two regions.
- Sierra Leone shows a much **lower mean DNI value**, indicating differences in solar patterns across regions.
- **Benin & Togo exhibit the most variability** in their GHI data as shown by their high standard deviation values.

Missing Data Analysis

The **Comments column** was found to have 525,600 missing entries across all datasets.

Table 2:

Dataset	Missing Entries in Comments
Benin	525,600
Sierra Leone	525,600
Togo	525,600

Observation:

- The **Comments column contains only null entries**.
- Recommendation: This column should likely be dropped to streamline analysis computations and avoid errors.

Anomalies

Anomalies were identified by looking at negative GHI values across datasets.

Findings:

1. **Togo's GHI Dataset:**
    - Contains **257,385 negative GHI values.**
    - These anomalies suggest data collection errors or preprocessing issues.
2. **Benin and Sierra Leone:**
    - No significant anomalies (negative values) were observed, indicating cleaner data.

Memory Usage

Table 3:

Dataset	Memory Usage (approx.)
Benin	76.2 MB
Sierra Leone	76.2 MB
Togo	76.2 MB

Findings:

- The memory usage for all three datasets is **similar (~76.2 MB)**, indicating no memory-related performance differences.

Recommendations

After comparing the datasets across all the factors:

Table 4:

Criteria	Best Dataset
Clean Data Quality	Benin & Sierra Leone
Fewest anomalies	Sierra Leone & Benin
Data Variability	Benin (most variability observed)

Final Recommendation

- If **minimizing anomalies and errors** is a priority: **Sierra Leone or Benin** are preferred for analysis.
- If **solar patterns over time** are the main goal and anomalies can be preprocessed: **Togo's dataset** can be used by addressing these anomalies.