

1. Summary Statistics & Missing-Value Report

- Load dataset with `pandas.read_csv()`.
- Convert 'Timestamp' to datetime.
- Use `df.describe()` to summarize numeric data.
- Identify missing data using `df.isna().sum()` and calculate percentage.
- Filter columns with >5% missing values.

2. Outlier Detection & Basic Cleaning

- Apply Z-score using `scipy.stats.zscore()` on ['GHI', 'DNI', 'DHI', 'ModA', 'ModB', 'WS', 'WSgust'].
- Flag rows where $|Z| > 3$ as outliers.
- Impute missing values in key columns with median using `fillna()`.
- Export cleaned data to `data/<country>_clean.csv`.

3. Time Series Analysis

- Plot GHI, DNI, DHI, Tamb over time using line charts.
- Analyze monthly and daily trends by grouping by `Timestamp.dt.month`.
- Visualize seasonality and anomalies.

4. Cleaning Impact

- Group by cleaning flag and compute mean ModA and ModB.

- Use bar plots to compare performance pre/post-clean.

5. Correlation & Relationship Analysis

- Generate a heatmap of correlations among GHI, DNI, DHI, TModA, TModB using ``seaborn.heatmap()``.
- Create scatter plots: WS vs GHI, RH vs Tamb, RH vs GHI to visualize dependencies.

6. Wind & Distribution Analysis

- Plot histograms for GHI and WS.
- Use ``WindroseAxes`` from ``windrose`` library to plot wind rose for WS/WD.

7. Temperature Analysis

- Explore the effect of RH on Tamb and GHI using scatter plots and trend lines.

8. Bubble Chart

- Plot GHI vs Tamb with bubble size representing RH and color as BP using Plotly.
- Useful for visualizing how humidity and pressure affect solar potential.

Conclusion & Key Insights

- Cleaned data ready for modeling and comparison.
- Outliers removed and missing values imputed.
- Strong correlations found between solar irradiance and temperature.
- Cleaning improves module performance significantly.
- Humidity inversely affects temperature, affecting sensor performance.