## Exploratory Data Analysis (EDA) Findings

### Dataset Summary

* Duration: June 2023 to May 2025
* Scope: Rajasthan-level solar generation (CEA) + NASA POWER climate data
* Records: 731 daily entries

### Column Renaming & Structure

Cleaned and normalized columns:

* Solar generation: solar\_mwh
* Temperature: temp\_avg\_c, temp\_max\_c, temp\_min\_c
* Humidity: humidity\_pct
* Wind: wind\_speed\_ms
* Precipitation: precip\_mm
* Solar irradiance: solar\_rad\_allsky\_mj\_m2, solar\_rad\_clrsky\_mj\_m2
* Cloudiness proxy: cloudiness\_index = clrsky - allsky

### 1. 📊 Correlation Heatmap Insights

\*\*Target: \*\*``

| Predictor | Correlation | Direction | Strength |
| --- | --- | --- | --- |
| All-sky Irradiance | +0.86 | ↗ | Very Strong |
| Clear-sky Irradiance | +0.81 | ↗ | Strong |
| Cloudiness Index (derived) | −0.72 | ↘ | Strong Inverse |
| Max Temperature | +0.17 | ↗ | Weak |
| Humidity | −0.13 | ↘ | Weak |
| Rainfall | −0.08 | ↘ | Negligible |
| Wind Speed | −0.03 | ↔ | None |

**Conclusion:**

Solar radiation (especially All-sky) is the dominant driver. Cloudiness has strong negative effect. Other variables provide minor seasonal context.

### 2. 📈 Temporal Trend Analysis

* solar\_mwh and solar\_rad\_allsky\_mj\_m2 show strong seasonal alignment.
  + Peak: April to June
  + Dips: July (monsoon) & December–January (winter haze)
* precip\_mm spikes during monsoon (June–September) correlate with solar troughs.
* cloudiness\_index surges during same low solar periods, confirming its suppressive role.
* Temperatures peak May–June but are not strongly aligned with solar output drops.

### 3. ☁️ Cloudiness as Key Inhibitor

* Difference between clear-sky and all-sky irradiance reveals true cloud burden.
* Cloudiness index is highly anti-correlated with solar output (−0.72).

### 4. 📊 Correlation Matrix Summary

solar\_mwh 1.00  
solar\_rad\_allsky 0.86  
solar\_rad\_clrsky 0.81  
cloudiness\_index -0.72  
temp\_max\_c 0.17  
humidity\_pct -0.13  
precip\_mm -0.08  
wind\_speed\_ms -0.03

### 5. 🧩 Pairplot Insights

* **Strong Linear Cluster:**
  + solar\_mwh vs. solar\_rad\_allsky\_mj\_m2
  + solar\_mwh vs. solar\_rad\_clrsky\_mj\_m2
* **Clear Negative Slope:**
  + solar\_mwh vs. cloudiness\_index
* **Wide Scatter / Weak Correlation:**
  + solar\_mwh vs. humidity\_pct
  + solar\_mwh vs. precip\_mm
  + solar\_mwh vs. wind\_speed\_ms
* **Outliers & Distribution Checks:**
  + All variables generally show Gaussian distributions except for precipitation, which is heavily right-skewed.

### Recommendations for Modeling

* **Primary predictors:** solar\_rad\_allsky\_mj\_m2, cloudiness\_index
* **Optional add-ons:** temp\_max\_c, humidity\_pct
* **Exclude:** wind\_speed\_ms, precip\_mm (low signal)

Would you like me to proceed with feature engineering (lags, rolling means) or start building a baseline predictive model next?