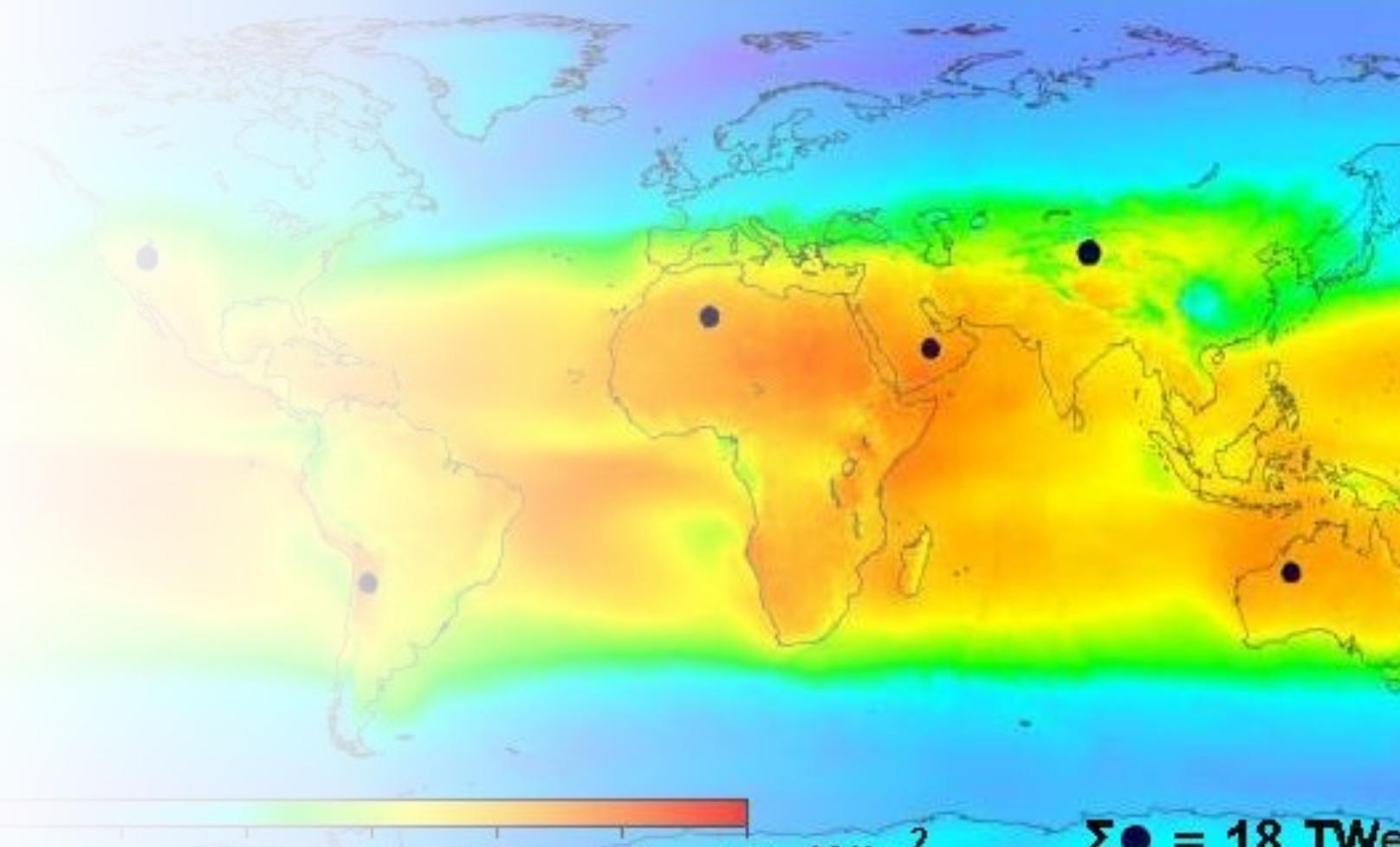




ML SOLAR ENERGY

J. Garcia



INTRODUCTION

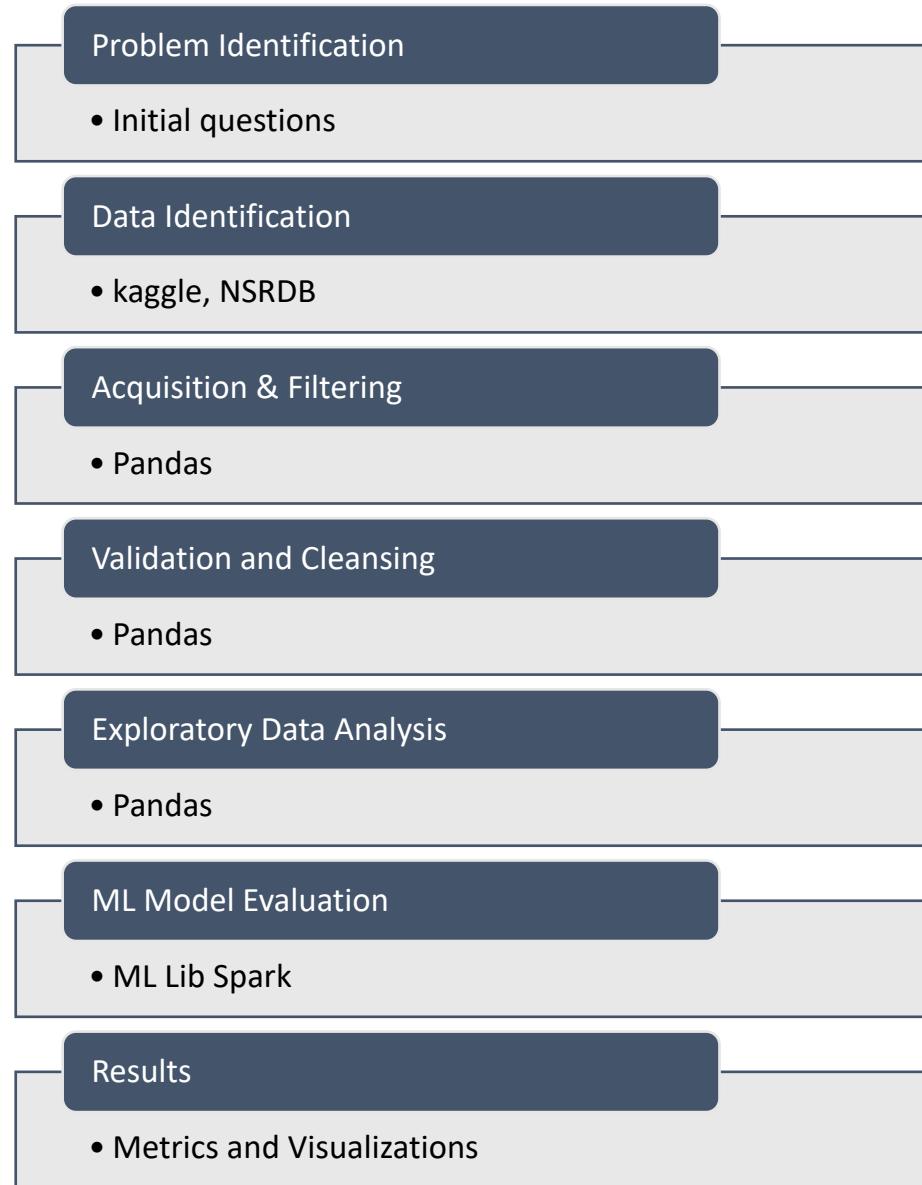
MOTIVATION

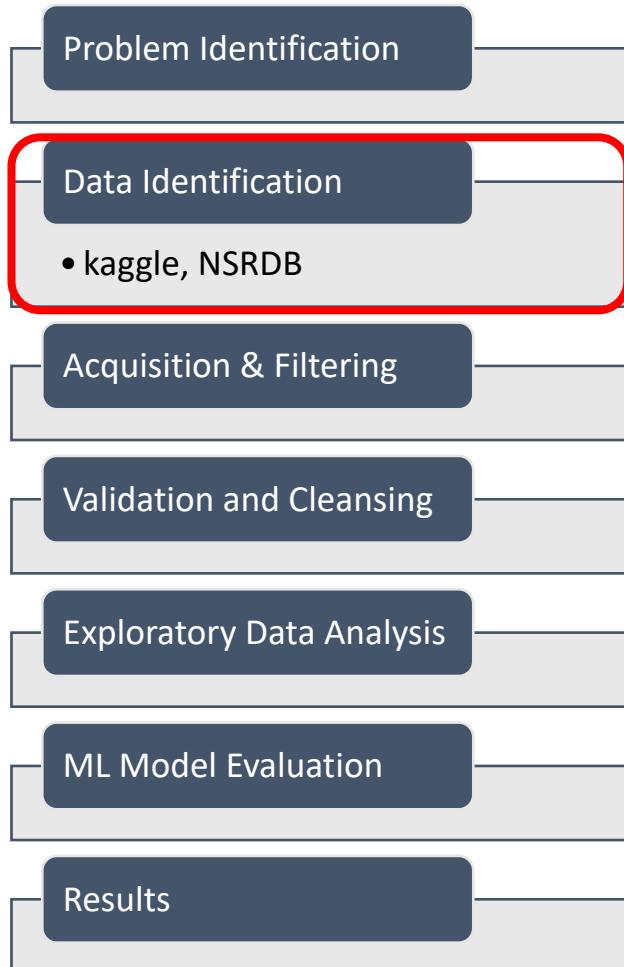
- Role of Artificial Intelligence in the energy industry.
- Environmental challenges.
- Innovation in solar panels.

QUESTION

- **Which weather variables influence in solar energy production?**
- **How accurate can a ML model predict solar energy production?**

METHODOLOGY





Kaggle

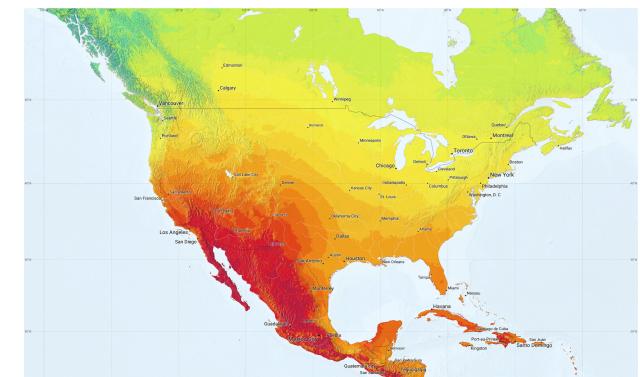
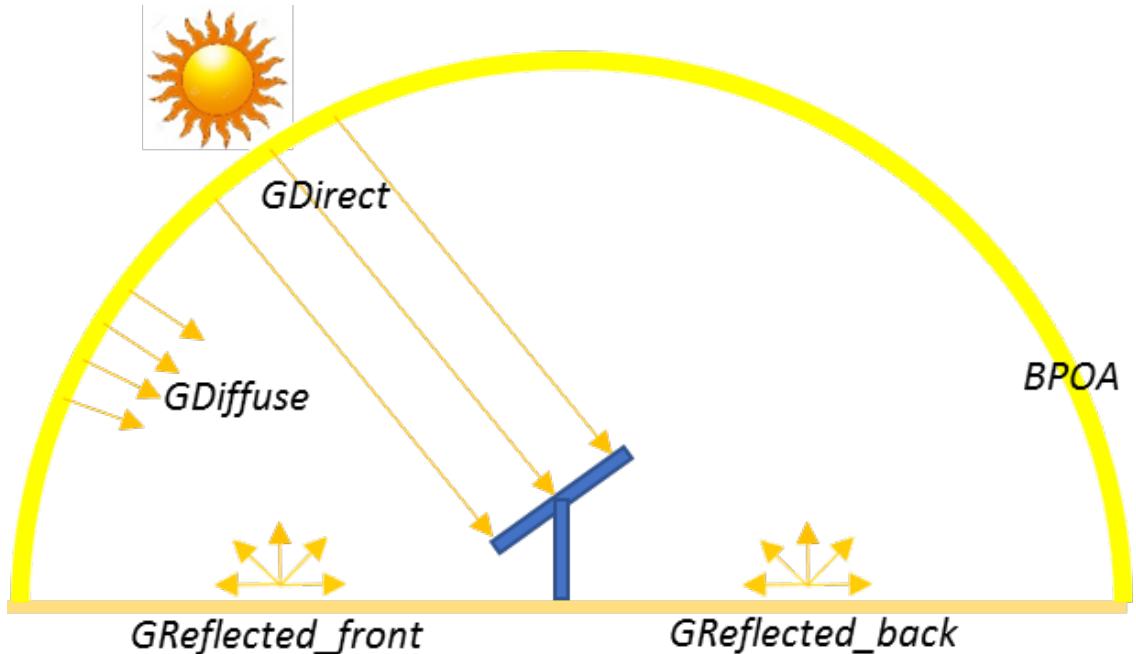
- Data complied from a solar power system installed in Berkeley, CA.
 - Date
 - Is Daylight
 - Distance to Solar Noon
 - Temperature
 - Barometric Pressure
 - Wind Speed
 - Sky Cover
 - Visibility
 - Humidity
 - Generated Power

National Solar Radiation Database ([NSRDB](#))

- Is a collection of meteorological and solar irradiance data sets for the United States.
- Solar Irradiance:
 - Global Horizontal (ghi)
 - Direct Normal (dni)
 - Diffuse (dhi)
- Cloud Type
- Dew Point
- Temperature
- Surface Albedo
- Pressure
- Relative Humidity
- Solar Zenith Angle

Variables explanation

- Solar irradiance
 - “**Power per unit area** received from the Sun in the form of electromagnetic radiation. [W/m²]”
- Direct Normal Irradiance: DNI
 - “Is measured at the surface of the Earth at a given location with a **surface element perpendicular to the Sun.**”
- Diffuse Horizontal Irradiance: DHI
 - “Is the radiation at the Earth’s surface from light **scattered or reflected** by the atmospheric components”
- Global Horizontal Irradiance: GHI
 - “Is the total irradiance from the sun on a horizontal surface on Earth, is the sum of direct irradiance (after accounting for the solar zenith angle of the sun z), and diffuse horizontal irradiance” **GHI= DHI + DNI x cos(z)**



Long-term average of GHI (1999–2018)
Yearly total: 400 1000 1200 1400 1600 1800 2000 kWh/m²
Daily total: 2.19 2.74 3.26 3.83 4.38 4.93 5.48 6.02

Variables explanation

- **Albedo**

- Is the measure of the reflection of solar radiation out of the total solar radiation. Scale (0-1)

- **Cloud Types**

- {**Clear**: 0, **Probably Clear**: 1, **Fog**: 2, **Water**: 3, **Super Cooled Water**: 4, **Mixed**: 5, **Opaque Ice**: 6, **Cirrus**: 7, **Overlapping**: 8, **Overshooting**: 9, **Unknown**: 10, **Dust**: 11, **Somke**: 12}

- **Sky cover**

- {**Clear**: 0, **Partially Clear**: 1, **Partially Cloudy**: 2, **Cloudy** : 3, **Over casted**: 4}

- **Dew Point**

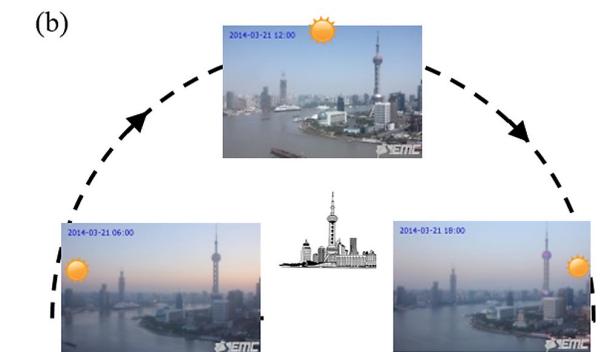
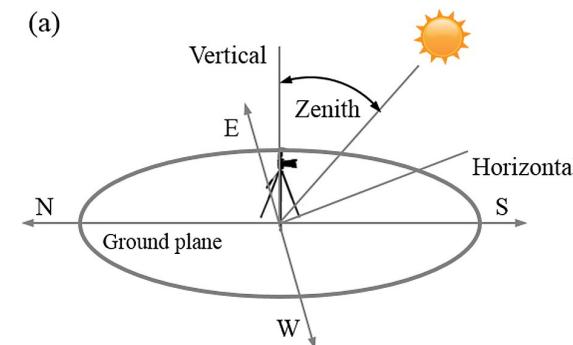
- Is the temperature to which air must be cooled to become saturated with water vapor. Related with humidity and expressed in Celsius.

- **Solar noon distance**

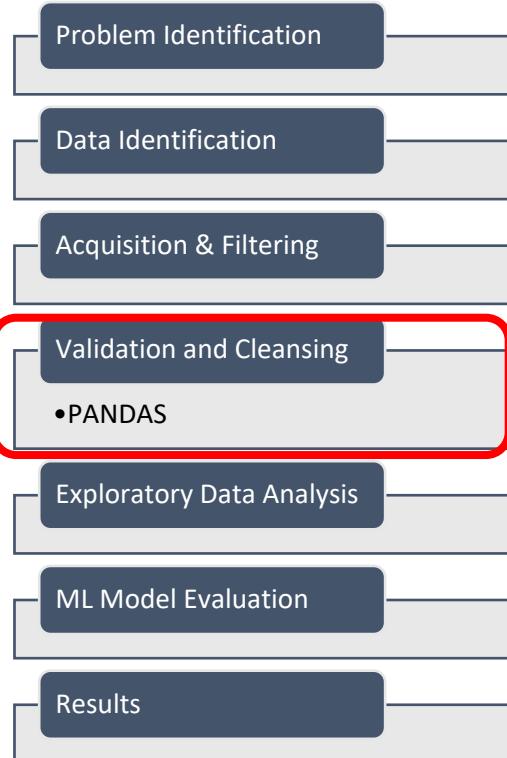
- Is the moment when the sun crosses the local meridian and reaches its highest position in the sky.

- **Solar Zenith Angle**

- Is the angle between the sun's rays and the vertical.

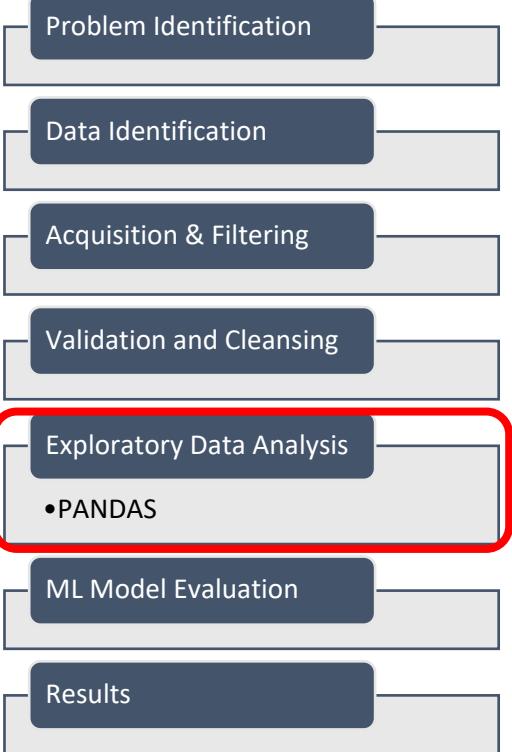


DATA CLEANSING

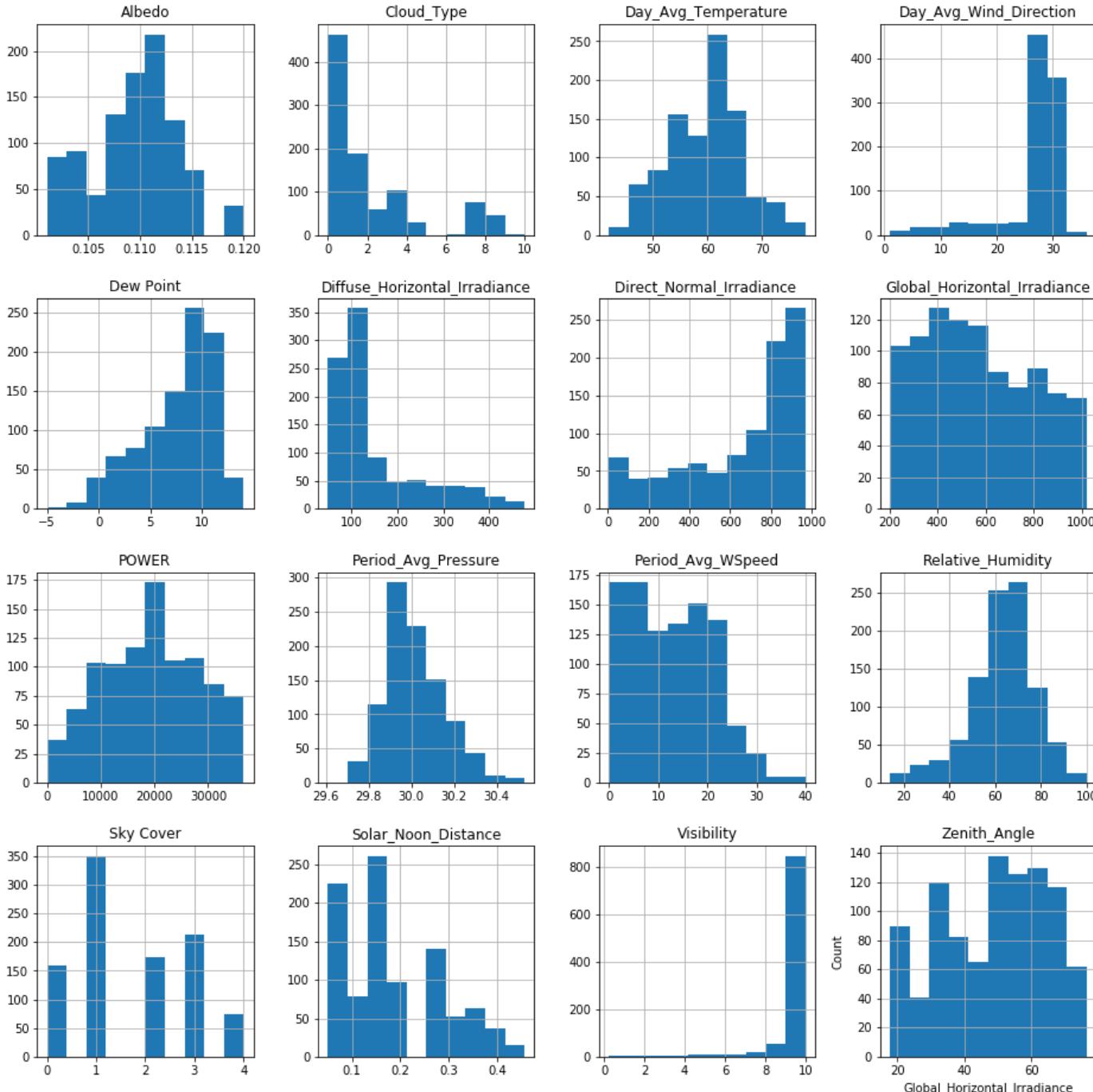


- One null value.
- **Different date time formats.**
- Most of the variables came in string format.
- Different headers.
- Drop and eliminate common columns like month, day, hour.
- Merge two datasets overlapping their resolution
 - 4 hours vs 15 min resolution.

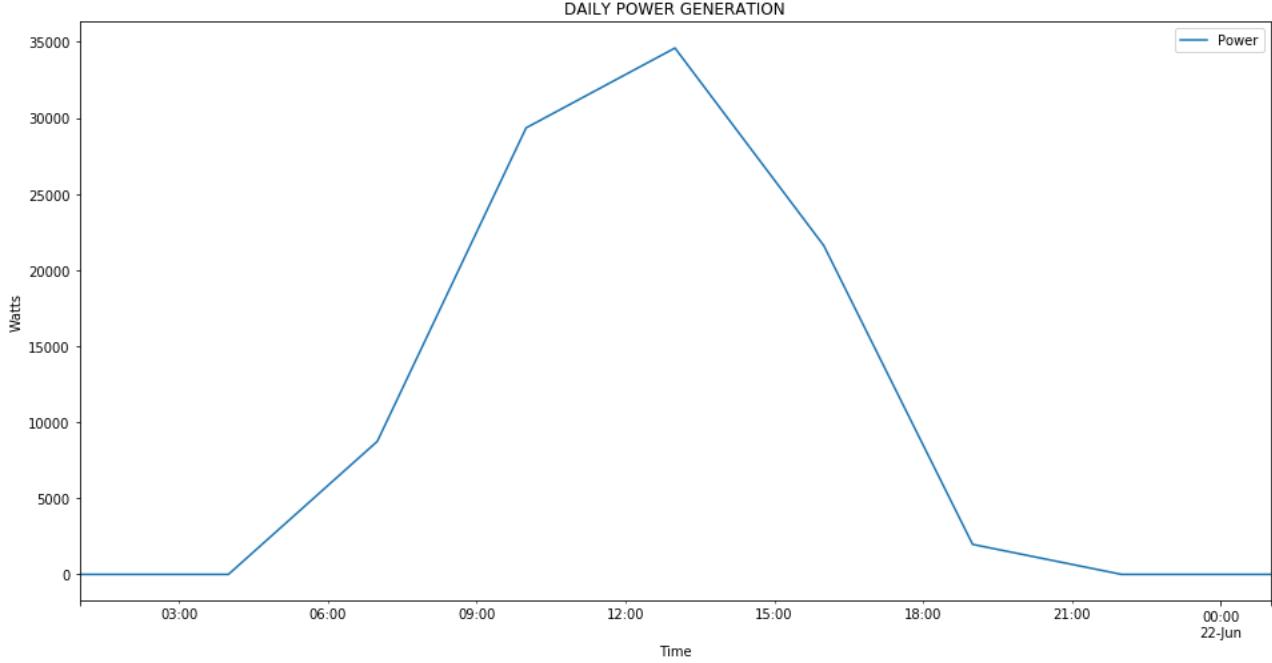
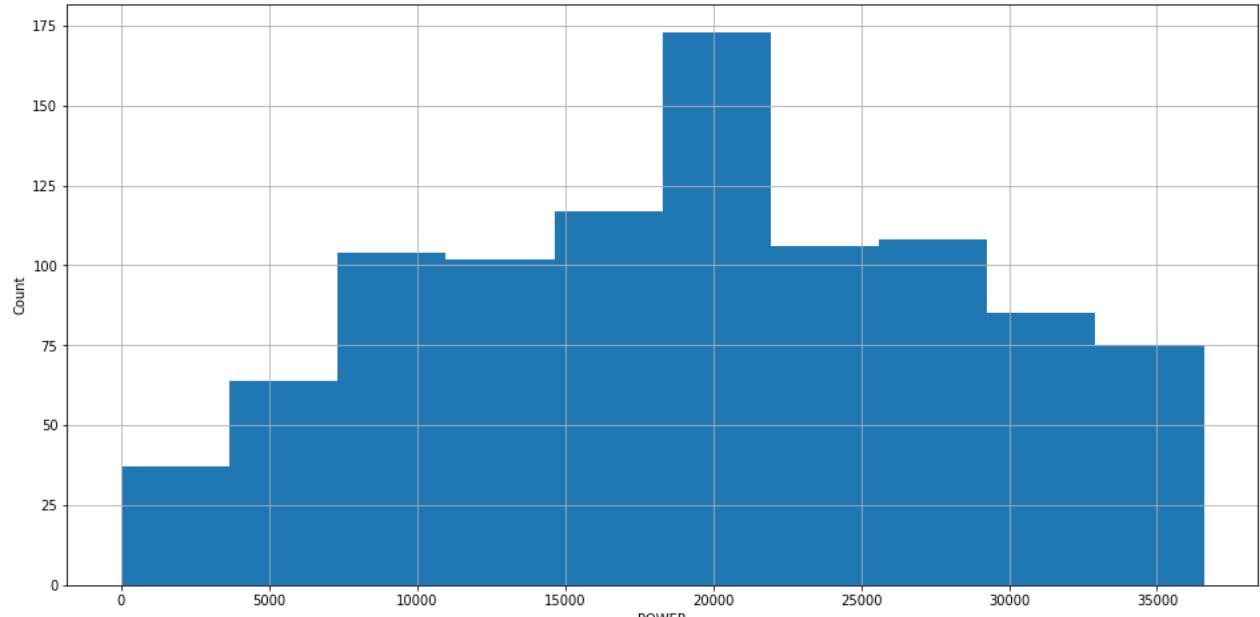
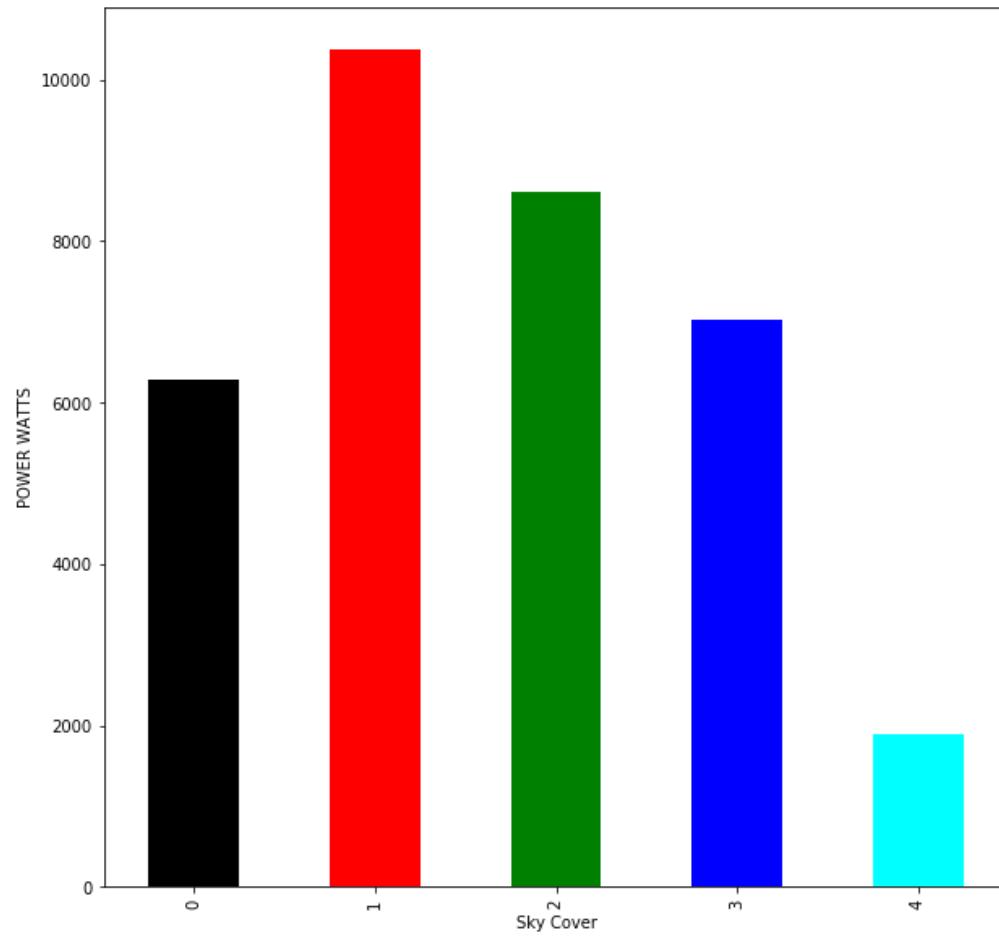
EDA



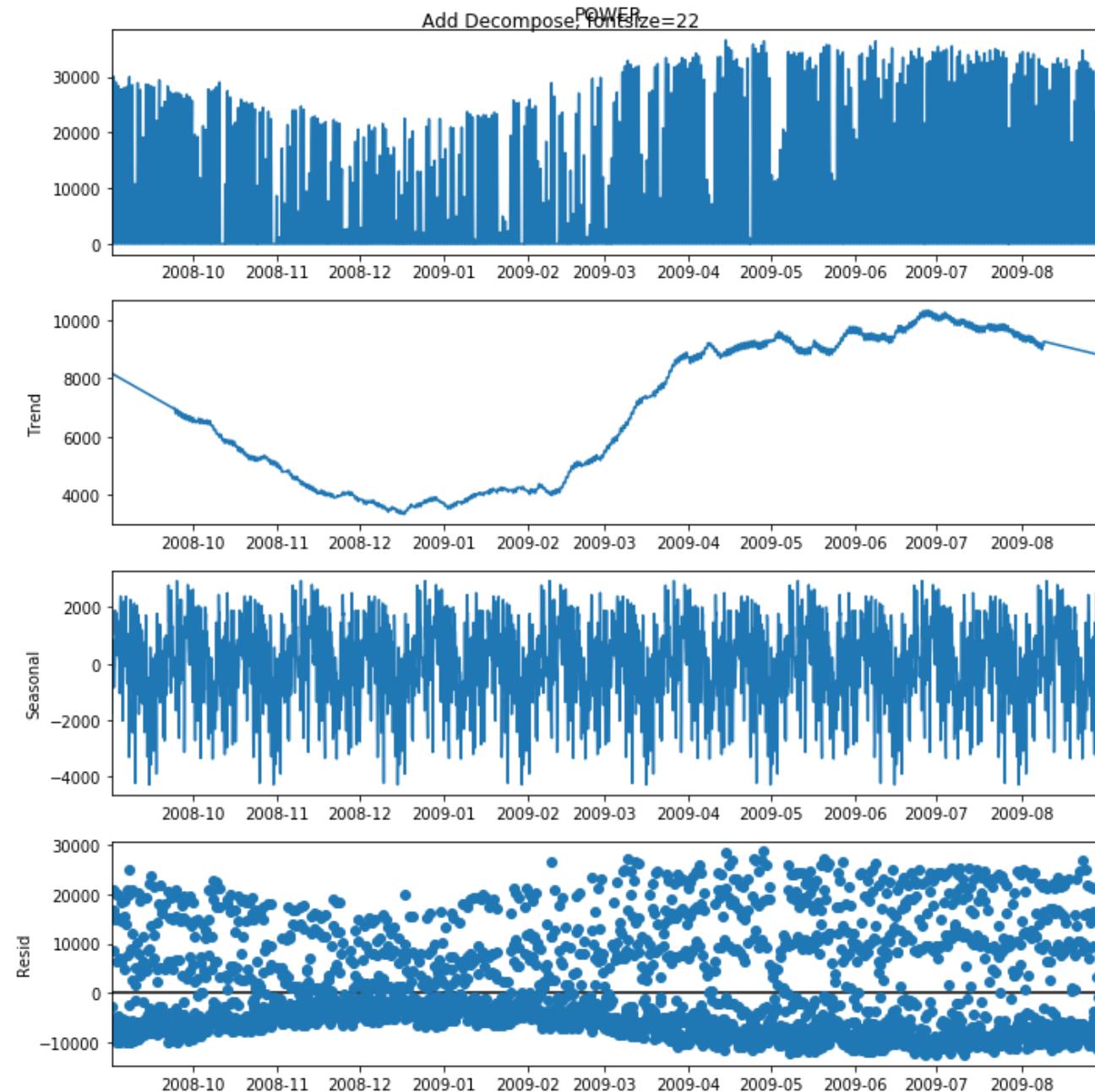
- Time series dataset
- Shape: 2920 x 24
 - 2008-09-01 - 2009-08-31
- Resolution: 3 hour
- Target Variable: Power Generation
- Seasonality
 - Daylight



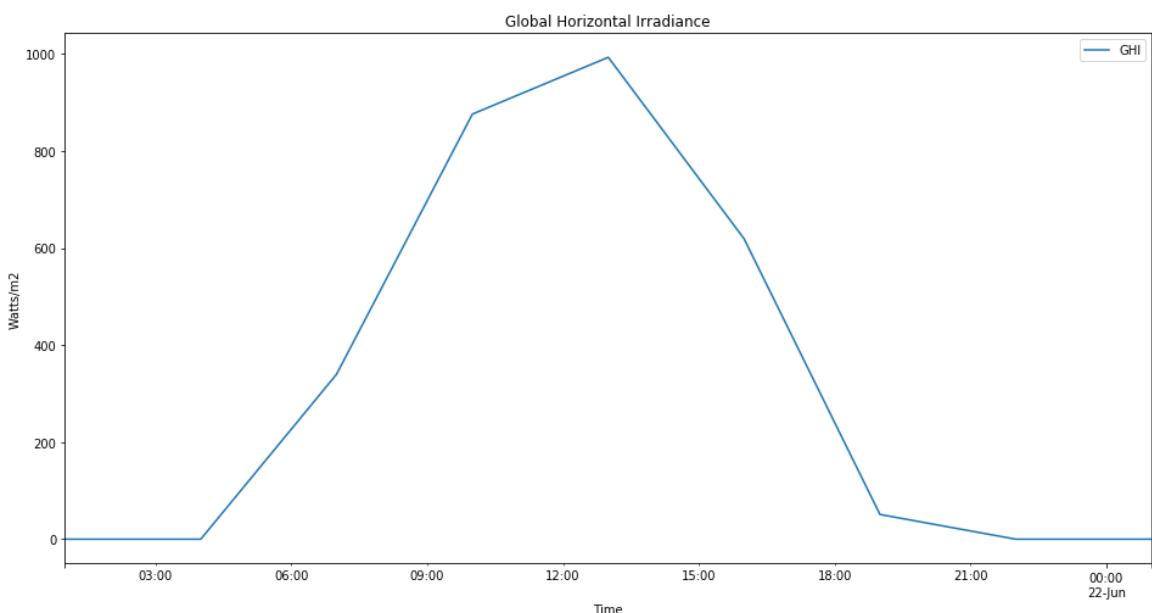
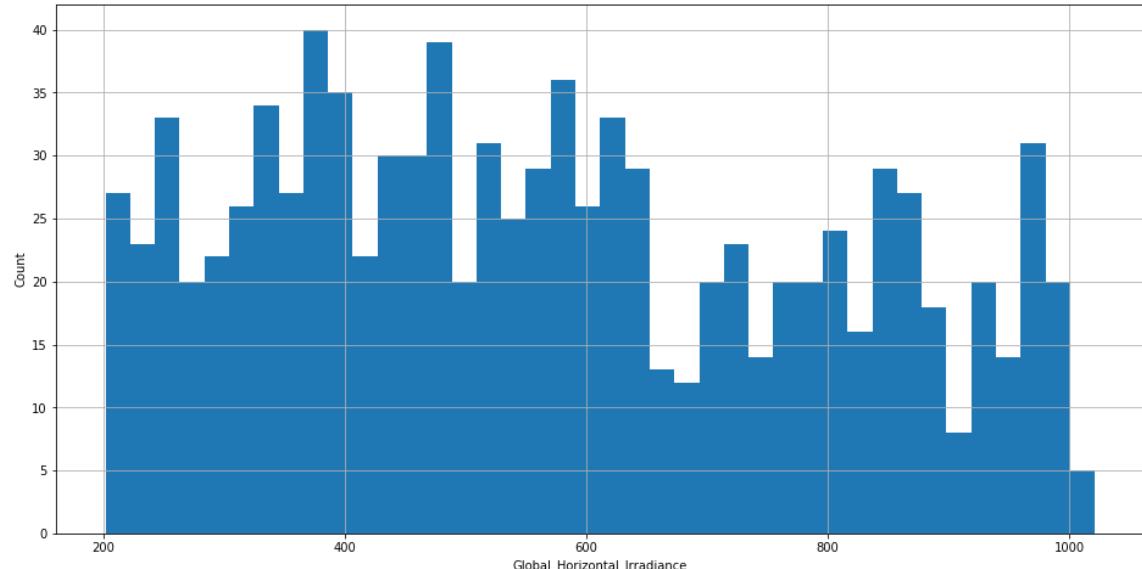
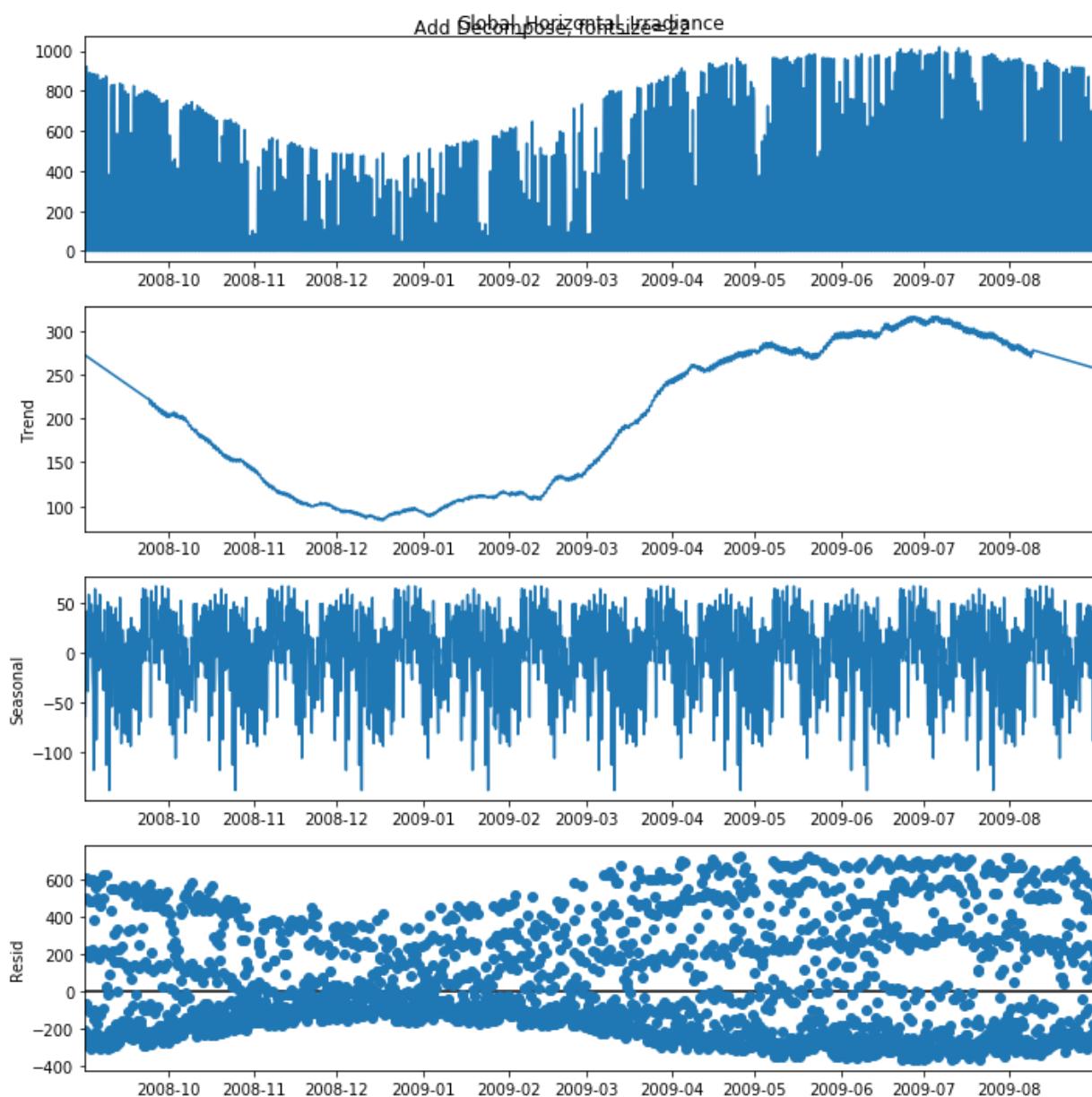
Power Generation



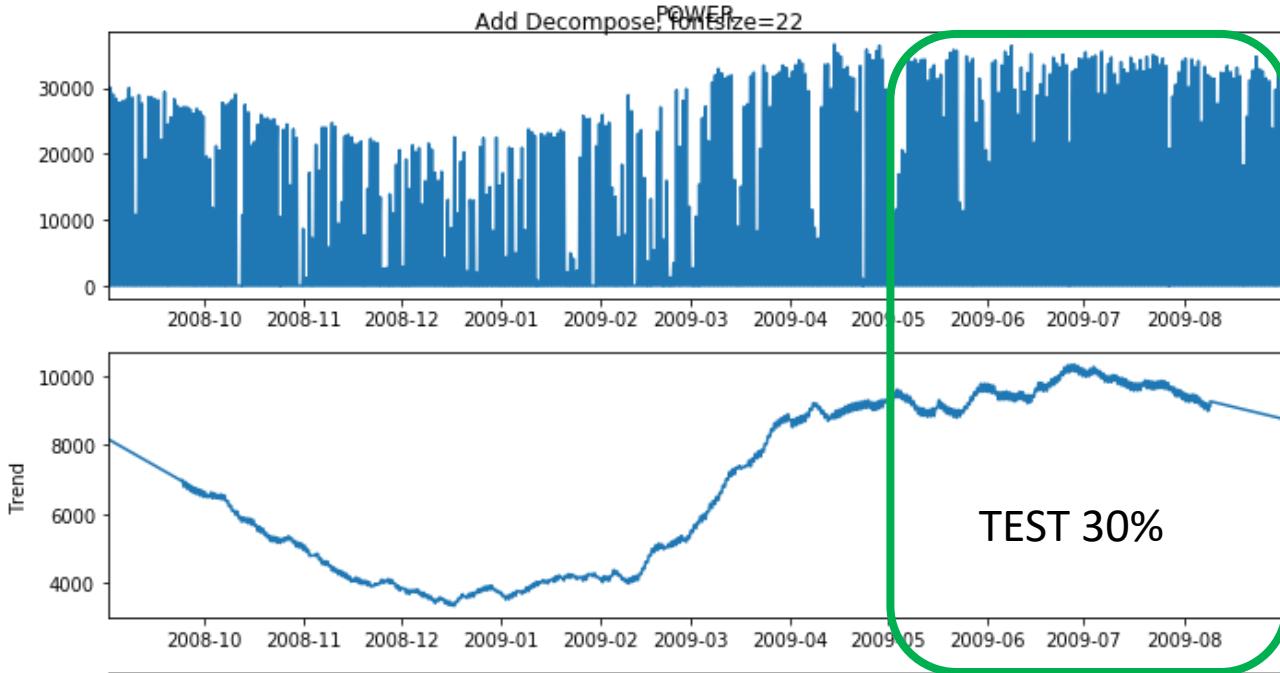
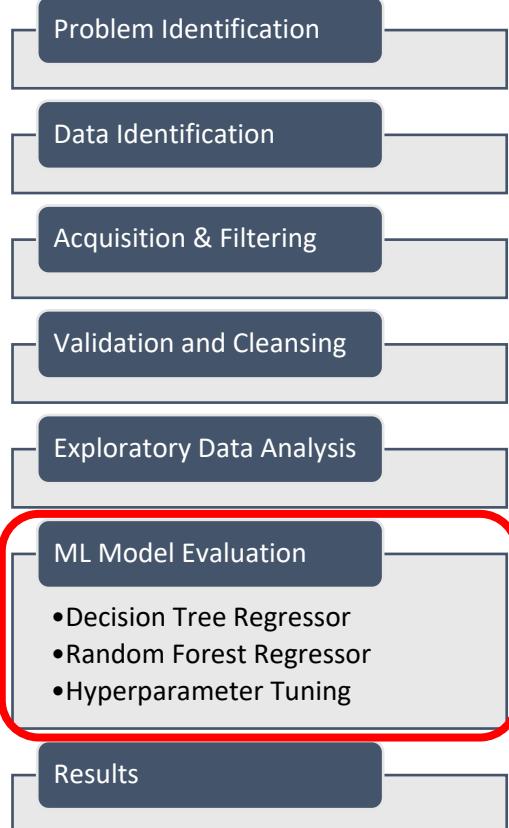
TIME SERIES POWER GENERATION



Solar Irradiance decomposition



ML LIB SPARK ML FLOW



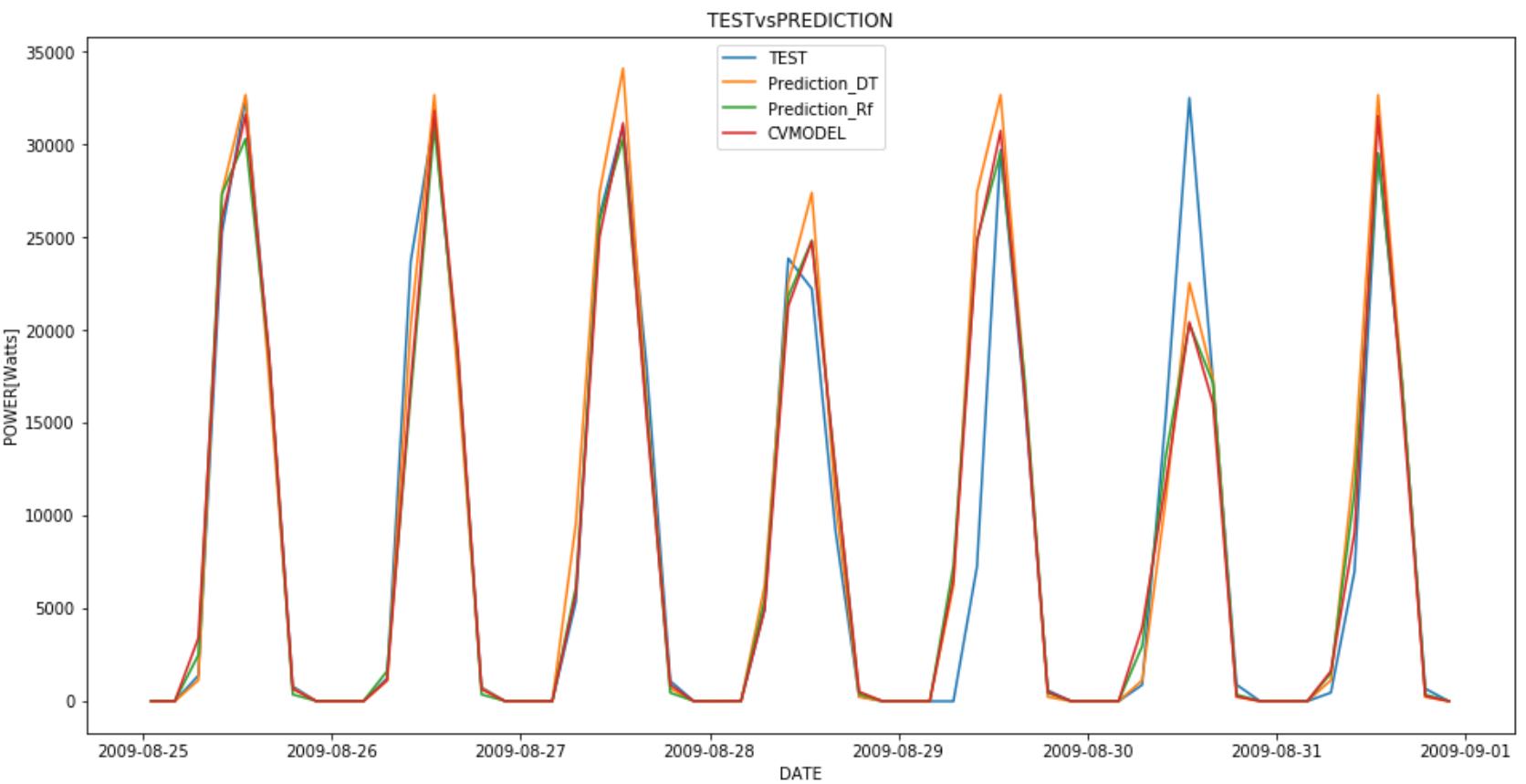
Input Variables: Day light, Distance to solar noon, temperature, sky cover, visibility, humidity, Irradiance, Cloud Type, Zenith Angle, Albedo, Pressure, Dew Point.

Decision Tree Regression

Random Forest Regression

Hyperparameter Tuning and Cross Validation

RESULTS



MODEL	PARAMETERS	R2	RMSE
Decision Tree	default	0.92	3228
Random Forest	Default	0.928	3227.5
Cross Validation Random Forest	nTrees= 100 Max Depth=10 Crossfold= 5	0.922	3351

We can predict accurately solar energy production with weather parameters as input applying different ML models.

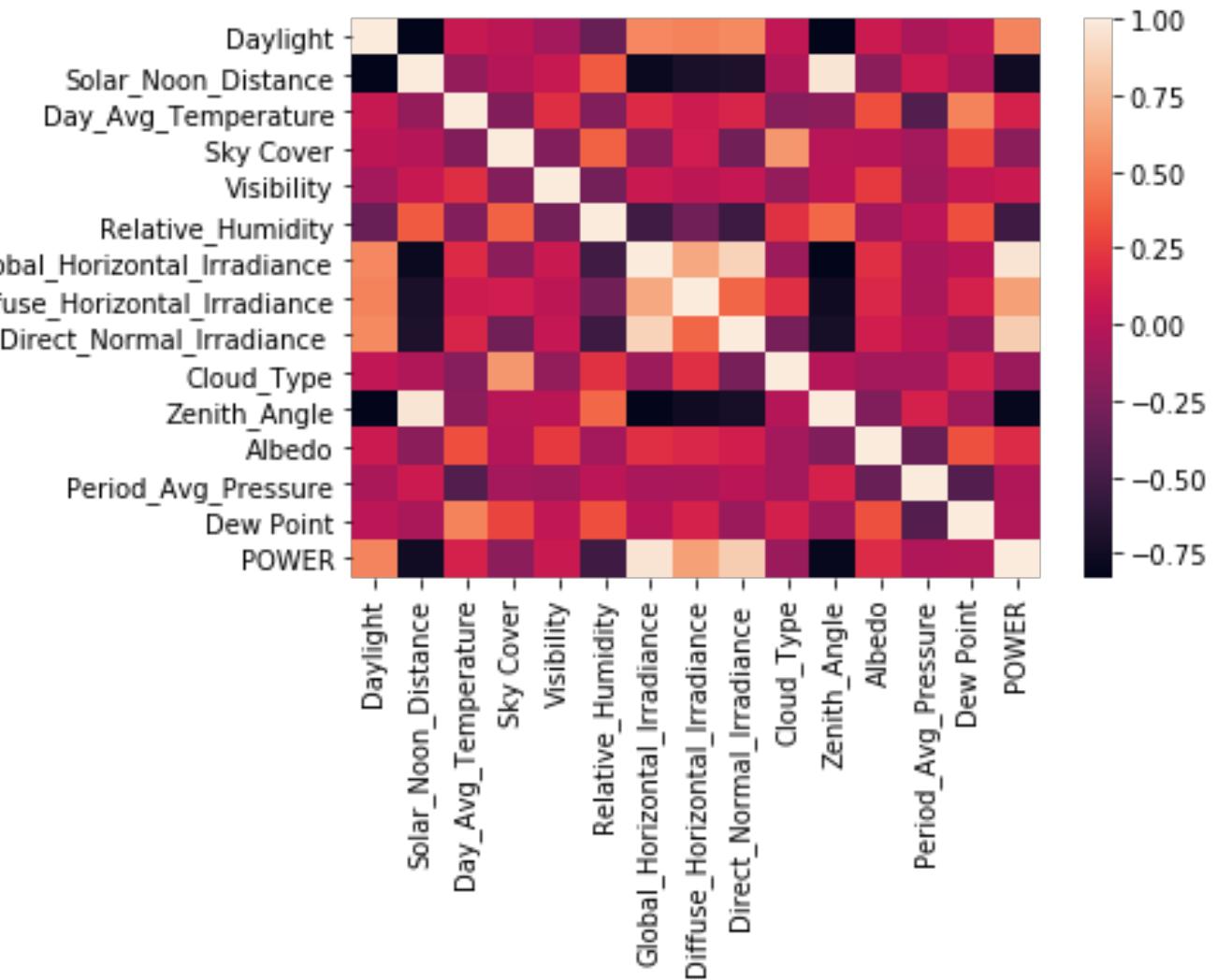
BUT, this model should only be applied within this facility.

RESULTS

	feature	importance
6	Global_Horizontal_Irradiance	0.374424
10	Zenith_Angle	0.207031
1	Solar_Noon_Distance	0.143825
7	Diffuse_Horizontal_Irradiance	0.100521
8	Direct_Normal_Irradiance	0.080387
5	Relative_Humidity	0.020478
9	Cloud_Type	0.016966
3	Sky_Cover	0.015989
12	Period_Avg_Pressure	0.010269
13	Dew_Point	0.009368
11	Albedo	0.008886
2	Day_Avg_Temperature	0.007829
0	Daylight	0.003018
4	Visibility	0.001008

The weather parameters that control solar production are:

- GHI
 - Zenith_Angle
 - Solar_Noon_Distance



THANK YOU