1. Determine any correlation between environmental factors with solar generation

2. Solar Generation is an essential part of any Climate and Sustainability Conversation. Without Solar Generation, topics of Climate and Sustainability would be much harder to find solutions for.

3.

Dataset 1

The first dataset I would like to work with is provided by a solar plant owner. The plant is located in India. The person posting this has received an award from the website and has received lots of upvotes for the dataset.

https://www.kaggle.com/datasets/anikannal/solar-power-generation-data

Under this link the file I am particularly interested in is "Plant_1_Generation_Data"

Dataset 2

The second dataset provides data for the ambient temperature, module temperature, and irradiance for the plant. The plant is located in India. The person posting this has received an award from the website and has received lots of upvotes for the dataset.

https://www.kaggle.com/datasets/anikannal/solar-power-generation-data

Under this link the file I am particularly interested in is "Plant_1_Weather_Sensor_Data"

Dataset 3

This last dataset is provided by the National Solar Radiation Database. It is provided by the United States Government. This dataset provides a lot of detailed weather information down to the minute. We can use this data to further understand weather correlations.

The data can be found here: https://nsrdb.nrel.gov/data-viewer

Unfortunately, the link to the dataset I am particularly interested in expires after 24 hours. I can provide the excel datasheet I extracted if needed.

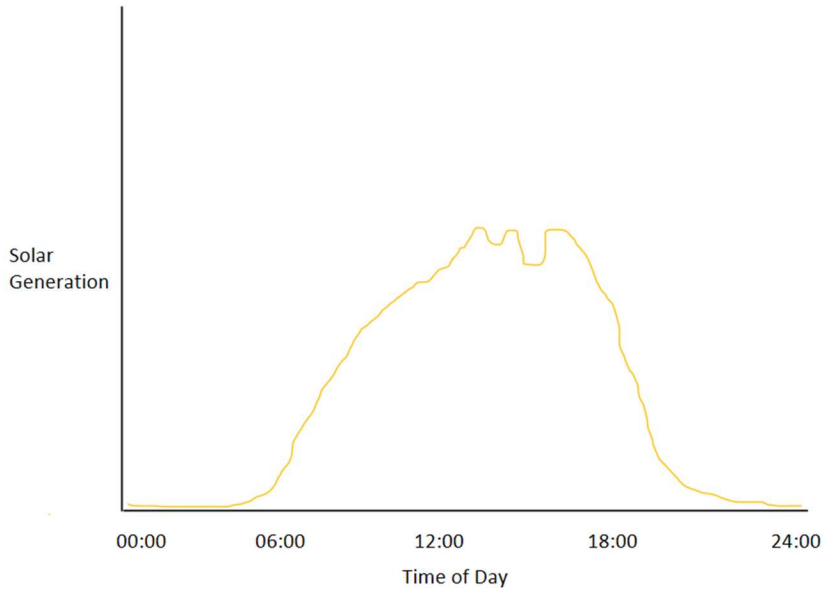
4a. Using these datasets, we can identify the correlations between the energy generated with weather information. We can do this using linear regressions on the models. This is the best analysis method to use for this class of data as we expect direct correlations between weather events and solar energy generation. Linear Regressions are also much simpler and provide easier computational efficiency.

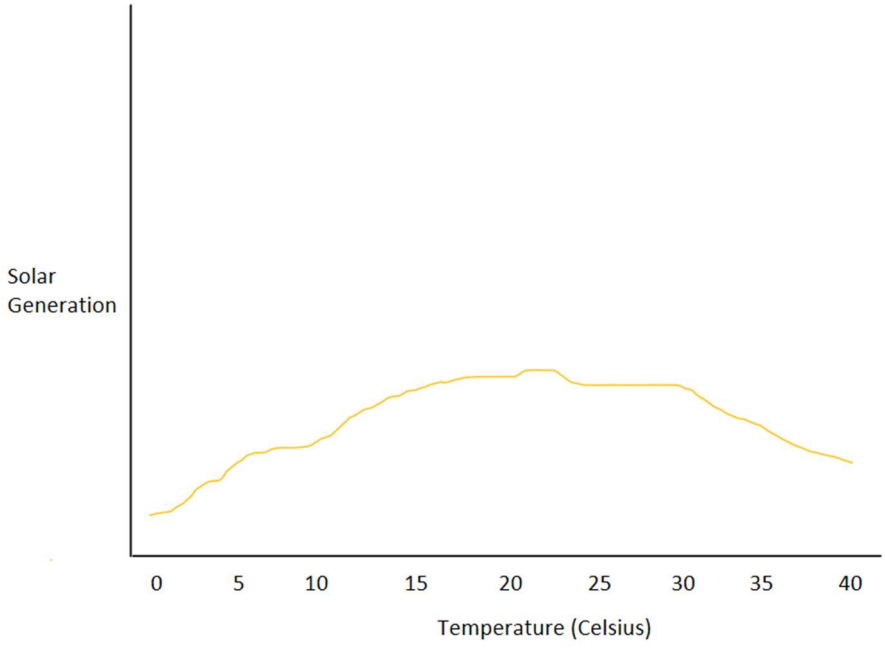
We can also use Neural Networks or Support Vector Machines. These are more complicated and may be unnecessary especially if the dataset can be easily handled through a Linear Regression.

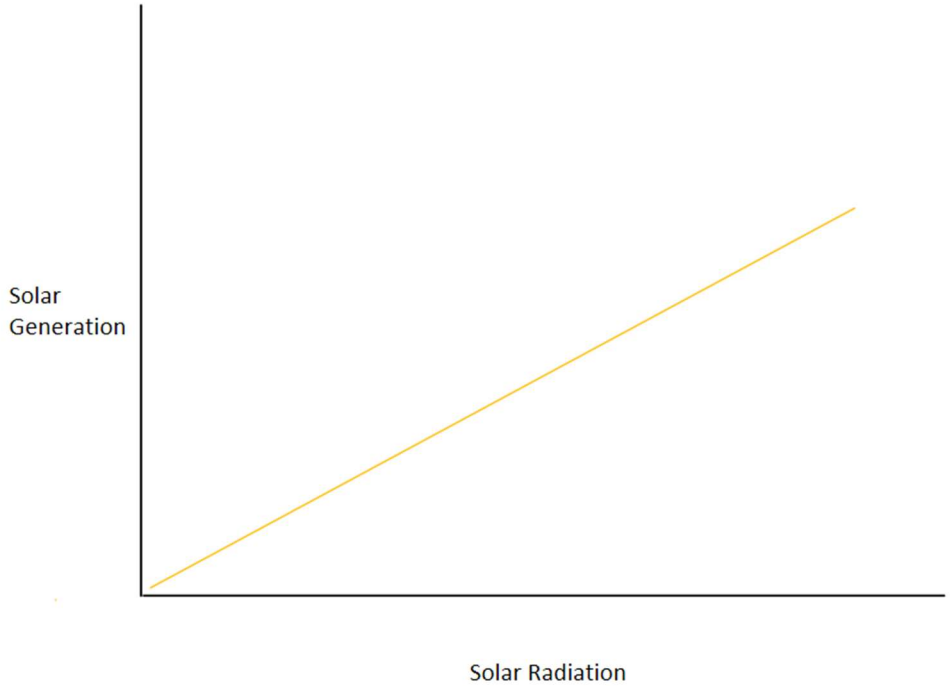
4b. Analyzing solar power generation data is crucial for cities to predict peaks and troughs, optimizing energy distribution, and ensuring grid stability. If cities understand how and when they can produce solar energy, they will be better suited for future solar energy production.

4c. Using a linear regression, we can determine the R^2 number. Using this, we can then determine how meaningful the analysis was. A higher R^2 number indicates a more significant correlation.

5.


Solar Generation
00:00
06:00
12:00
18:00
24:00
Time of Day


Solar Generation
0
5
10
15
20
25
30
35
40
Temperature (Celsius)


Solar
Generation
Solar Radiation