DATASCI 200: Introduction to Data Science Programming

Project 2

Real world solar irradiance and solar panel Photovoltaic(PV) module study

Zejia Jing (jzejia1@berkeley.edu)

Jianyi Teng (jianyiubc@berkeley.edu)

Minghung Cho(ming.cho@berkeley.edu)

Team's GitHub repository:

https://github.com/UC-Berkeley-I-School/SEC6_ZEJIA_JIANYI_MING

1) Overview

Solar produced 4.7% of U.S. electricity in 2022. In solar energy production, solar technologies play a key role by converting sunlight into electrical energy through photovoltaic (PV) panels.

2) Questions

a. What is the efficiency of different types of solar panels? Nowadays We have many types of solar panel materials and technology. In this project, the efficiency of 10 types of solar panel materials will be studied based on a real world measurement dataset.
b. What is the relationship of location data (longitude and latitude), time data(y-m-d h-m-s), weather data (including cloud type) and real-time solar irradiance evaluated in Direct Normal Irradiance(DNI), Diffuse Horizontal Irradiance(DNI) and Global Horizontal Irradiance(GHI)?

3) Data Collection and Cleaning

In this study, we are using open-source databases.

The solar radiation data is downloaded from the National Solar Radiation Database (NSRDB) developed by the National Renewable Energy Laboratory (NREL) [1]. The solar panel power output data is measured and validated in a public NREL report: “User’s Manual for Data for Validating Models for PV Module Performance”[2].

We use real world dataset from 2012 to 2013 from the NSRDB and NREL report including weather, location, timing, solar irradiance and solar panel power output data. The picture of experiment PV Modules is shown in Fig. 1 where real world data is collected by NREL from three locations: the west coast, mid-US, and east coast and solar power output is measured using 10 types of solar panel technologies by 9 manufacturers.

Fig 1. PV modules and solar radiation measurement devices deployed on the Golden, Colorado(left); Cocoa, Florida(middle) and Eugene, Oregon(right) [2]

Table I: PV Modules, technology and their manufacture

Module Identifier	Technology	Manufacturer
xSi12922	Single-crystalline silicon	Manufacturer 1
mSi460A8	Multi-crystalline silicon	Manufacturer 1
mSi0166	Multi-crystalline silicon	Manufacturer 2

CdTe75638	Cadmium telluride	Manufacturer 3
CIGS39017	Copper indium gallium selenide	Manufacturer 4
CIGS8-001	Copper indium gallium selenide	Manufacturer 5
HIT05667	Amorphous silicon/crystalline silicon (HIT)	Manufacturer 6
aSiMicro03036	Amorphous silicon/ microcrystalline silicon	Manufacturer 7
aSiTandem72-46	Amorphous silicon tandem junction	Manufacturer 8
aSiTriple28324	Amorphous silicon triple junction	Manufacturer 9

We also use data cleaning techniques to clean the empty value column and exclude the unrelated parameters and columns in the dataset..

4) Variables

The solar radiation data includes these variables we intend to explore:

- Weather data: temperature, cloud type
- Time data: year, month, day, and hour
- Location data: latitude and longitude
- Solar radiation data: GHI, DHI and DNI

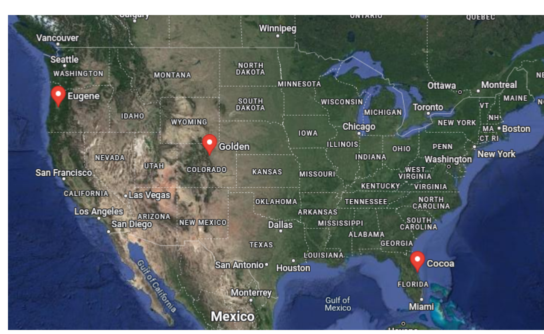
We expect to find a strong relationship between weather data, time data, location data to solar radiation data. Besides, we will summarize the data with the average estimated solar radiation based on different weather, time, and location. This analysis will give users a good understanding and estimate of solar irradiance based on their location, time, and weather conditions.

The PV panel data includes these variables we intend to explore:

- Solar radiation data: Plane of Array solar irradiance calculated from GHI, DHI and DNI
- Solar panel material type: 10 types
- PV solar power output

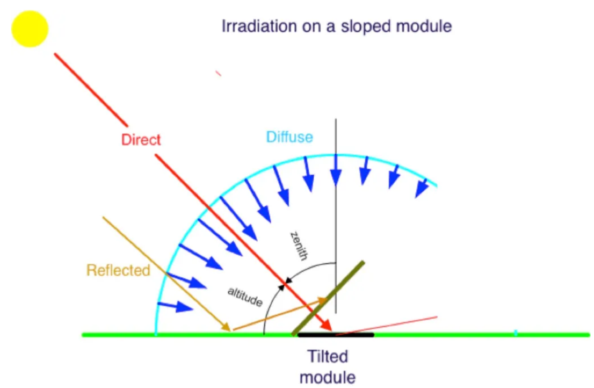
We would like to estimate solar panel efficiency of different types of solar panels with different manufacturers. This analysis will give users a good understanding of the real-world efficiency of different types of solar panels.

5) Ming's Data Stories and Figures: Analysis: Solar panel installation site analysis

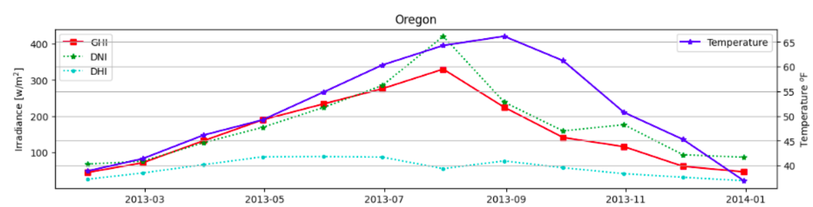


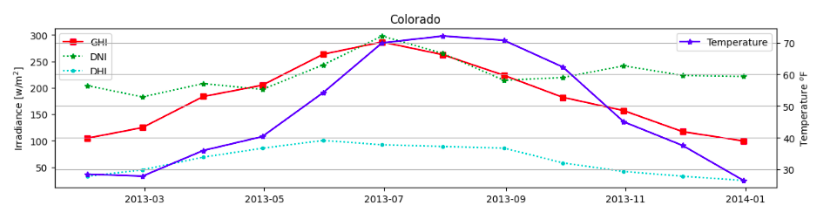
There are many factors considered when choosing a solar panel installation site. The key factors such as sunlight exposure, local climate conditions, roof orientation, or surrounding shading impact the performance of a solar panel. In this analysis, we are looking at data from the National Renewable Energy Laboratory for 3 locations, Eugene, Oregon, Golden, Colorado, and Cocoa, Florida.

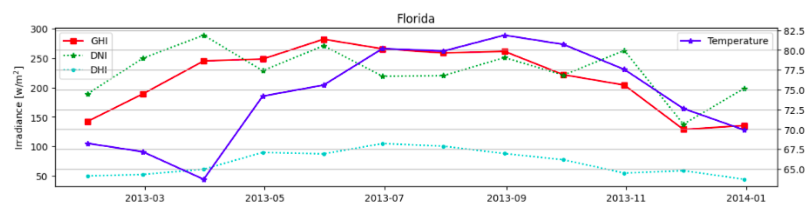
In this analysis, we will focus on analyzing the GHI, DNI, DHI, and temperature from the aforementioned locations in 2013. Here is a summary of what these variables represent.



- **GHI**, Global Horizontal Irradiance
  Higher value means more sunlight.
- **DNI**, Direct Normal Irradiance
  Higher values indicate more direct sunlight.
- **DHI**, Direct Diffuse Horizontal Irradiance
  Higher value can be beneficial for flat solar panels that are not directly oriented towards the sun.
- **Temperature**
  Solar panels perform better around 25 degrees Celsius / 77 degrees Fahrenheit.



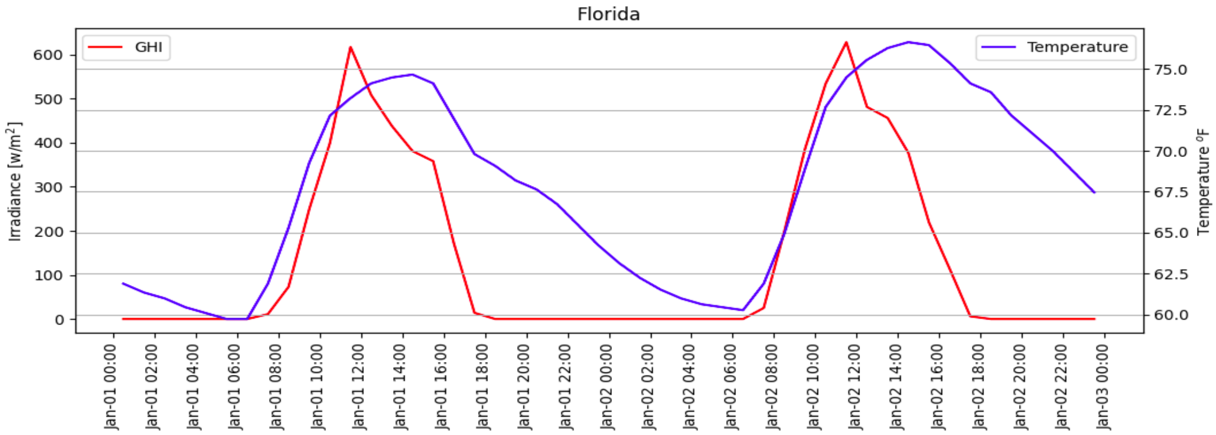




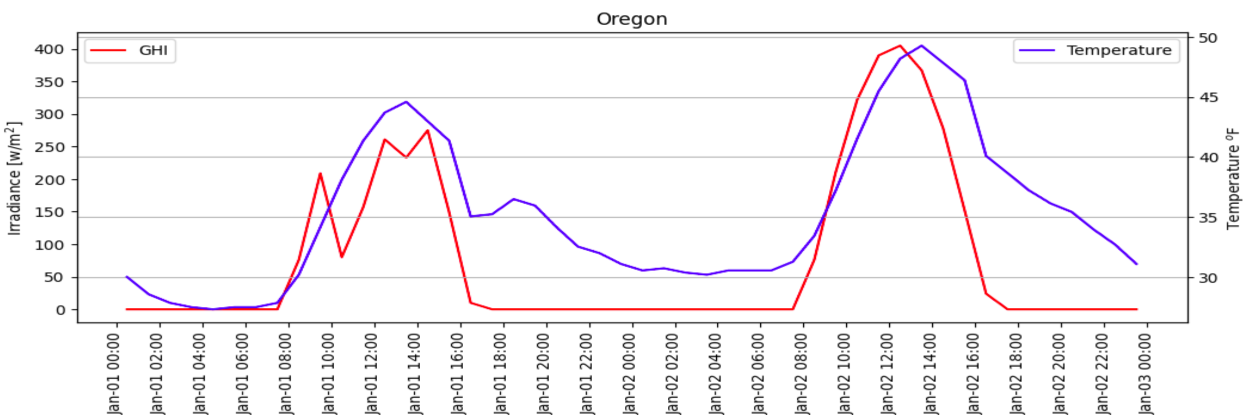
To begin, we created a chart to illustrate 4 variables using their monthly average so that we can get an idea how these variables change throughout the year. At a quick glance, you can see that all 3 locations share similar patterns. For instance, GHI stays above 150 from May to November. Except for Florida whereas its GHI values stays above 150 all year around. Clearly, all three locations present good environmental conditions to develop solar energy during summer time when sunlight exposure is abundant.

However, what about during the winter time when sunshine time is reduced?

If we take a closer look at Florida's data during winter to understand how its GHI and temperature change from Jan-1 to Jan-2, you can clearly see the GHI stays above 100 from 8 AM to 5 PM, which coincides with the time of sunrise and sunset. The location provides a health GHI with good temperature for 9 hours



By comparison, if we run the same analysis for Oregon during the same time, you will see a very different result. Not only is the average GHI about half compared to Florida, but its effective sunlight exposure is only about 7 hours (9 AM - 4PM).



As a result, Oregon’s primary electricity source is Hydro (39%) with Solar less than 2% whereas Florida’s primary electricity source is Natural gas (74%) with Solar around 3.7%. Floria is also ranked 3rd in solar capacity in the United States in 2023.

6) Jianyi’s Data Stories and Figures: Temperature and Weather ‘s influence toward GHI (Global Horizontal Irradiance)

Data Source: https://nsrdb.nrel.gov/
City chosen for analysis: Cocoa Florida, Golden Colorado, Eugene Oregon
Time Range: 2013-01 to 2014-01

OverView:
The Main purpose of the analysis is to Find out if GHI has correlation with Temperature and Weather. We confirmed that source data contains clean data about temperature for further plotting, while lacking categorized data to indicate weather types. At this point, we choose Cloud type as an explanatory variable, which will be driveable to get weather information in next step analysis.

Cloud type table:

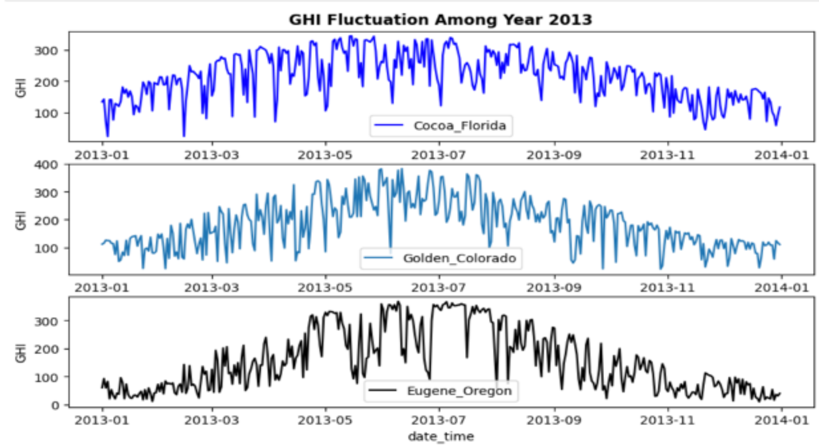
Cloud Type Index	0	1	2	3	4	5	6	7	8
Explanation	Clear	Probably Clear	Fog	Water	Super-cooled Water	Mixed	Opaque Ice	Cirrus	Overlapping

Cloud Type Index	9	10	11	12
Explanation	Overshooting	Unknown	Dust	Smoke

Question One:

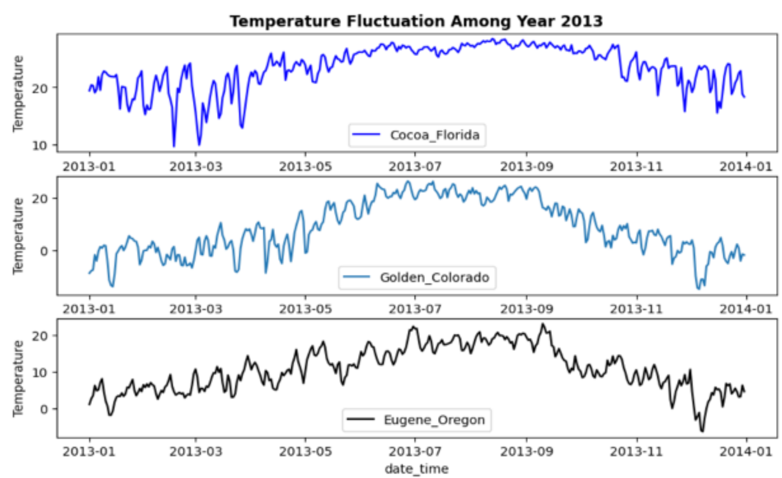
How does GHI and Temperature change during the entire year 2013?

The start point of our analysis is to conduct an observation about the pattern of GHI and Temperature in 2013, thereby evaluating possible similarity and correlationship. To answer the first question, we choose time series plotting.



Observation:

Among entire year, highest GHI happened around 2013-05 to 2013-08
Among entire year, highest Temperature happened around 2013-07 to 2013-08

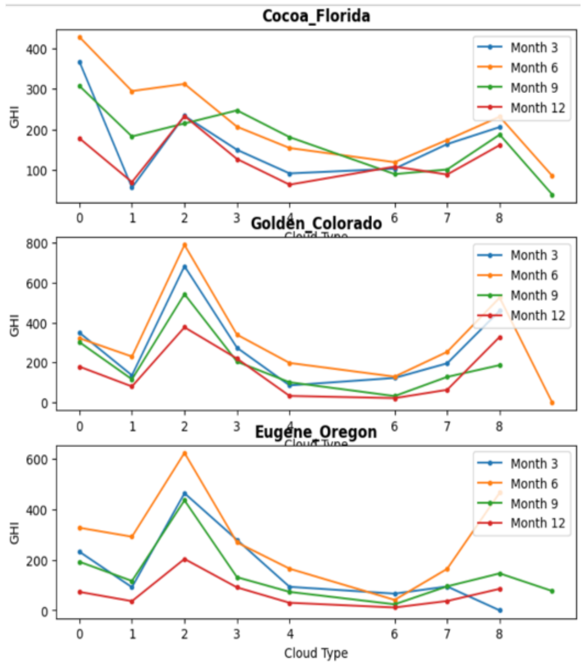


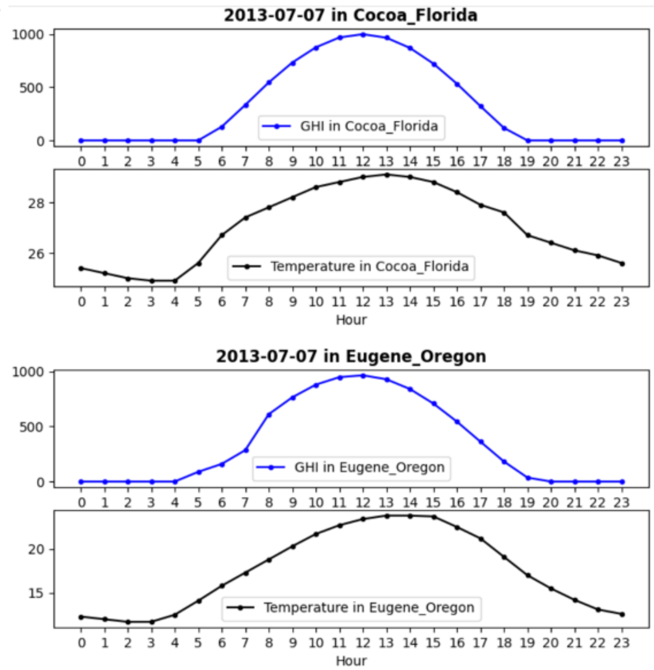
From the above figure, trends of GHI Fluctuation over 2013 shows similarity as temperature Fluctuation. To get further confirmation of our discovery, analysis about GHI and temperature in 24 hours (one specific day) is established.

Question Two -- How does GHI and Temperature change on one specific day?

The day (2013-07-07) with highest GHI among the year was chosen for question two analysis to show greater numerical fluctuations.Figure shown in Page 6 right side

In addition, Cocoa Florida and Eugene Oregon are chosen as locations for the sake of controlling cloud type variables -- cloud type of selected city in 2013-07-07 kept static at 0.





Conclusion:
There exist positive correlation between temperature and GHI

Question Three: Consider about Cloud Type and GHI, which Cloud Types shows highest GHI and which shows lowest?

To reduce the bias coming from temperature's inference in the analysis. Specific time ranges are selected to ensure all temperature levels are included. Figure shown on page 6 left side.
Month chosen:
3 (Spring),6(Summer),9(Fall),12(Winter)

Summarization from the left figure:
4 lines representing different season own similar trend and prove that there exist correlation between Cloud Type and GHI

Highest GHI: Cloud Type 2 -- Fog
Cloud Type 8 -- Overlapping
Lowest GHI: Cloud Type 6 -- Opaque Ice

7) Zejia's Data Stories and Figures: PV Module Efficiency Analysis

Question 7a Print the shape of the data:

```
(38929, 422)
```

The data including 38928 measurements from 2011-01-21 to 2022-03-04

Question 7b Print out the first five rows of the dataset. How do the columns Time Stamps, POA irradiance CMP22 pyranometer (W/m2) and Pmp (W) look?

	Time Stamp (local standard time) yyyy-mm-ddThh:mm:ss	POA irradiance CMP22 pyranometer (W/m2)	POA irradiance uncertainty (%)	PV module back surface temperature (degC)	PV module back surface temperature uncertainty (degC)	Isc (A)	Isc uncertainty (%)	Pmp (W)	Pmp uncertainty (%)	Imp (A)	...	Global horizontal irradiance standard deviation of 1-second samples of 5-second average (W/m2)	Diffuse horizontal irradiance (W/m2)	Diffuse horizontal irradiance uncertainty (%)
0	2011-01-21 08:15:05	35.1	10.7	19.0	1.9	0.1806	4.2	2.1674	4.2	0.1536	...	0.1	37.6	9.9
1	2011-01-21 08:25:05	21.0	17.1	19.0	1.9	0.1092	6.7	1.1815	6.7	0.0899	...	0.1	22.8	15.7
2	2011-01-21 08:35:04	29.9	12.3	19.1	1.9	0.1616	4.6	1.8936	4.7	0.1362	...	0.0	29.8	12.2
3	2011-01-21 09:10:05	29.5	12.1	18.9	1.9	0.1593	4.6	1.8719	4.6	0.1348	...	0.1	34.6	10.7
4	2011-01-21 09:15:04	52.1	7.0	19.5	1.9	0.2868	2.7	3.6956	2.7	0.2486	...	0.2	59.9	6.9

5 rows × 42 columns

Question 7c Print the average time interval of the datasets

Answer: -333.5586021788327

The average time interval for the datasets is 332 seconds which is around 5.5 minutes.

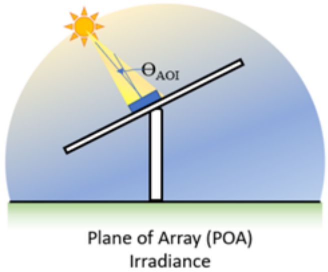
Question 7d Would you recommend dropping any columns based on the number of nulls?
Answer: Null column is with column name including 'Unnamed'. I would recommend dropping those columns with the title “Unnamed”.

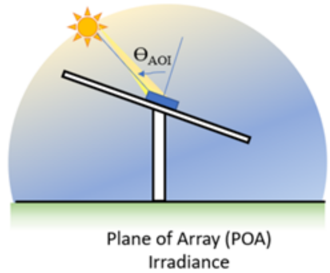
```
Time Stamp (local standard time) yyyy-mm-ddThh:mm:ss      38929
POA irradiance CMP22 pyranometer (W/m2)                   38929
POA irradiance uncertainty (%)                            38929
PV module back surface temperature (degC)                 38929
PV module back surface temperature uncertainty (degC)     38929
                                                            ...
Unnamed: 417                                               4327
Unnamed: 418                                               1736
Unnamed: 419                                               1736
Unnamed: 420                                                165
Unnamed: 421                                                165
Length: 422, dtype: int64
```

Question 7e: Delete columns that are NaN or unnamed. Print the shape of the data
(38929, 42)
Answer: The current datasets are in good shape. I reduce the column number from 422 to 42 and drop those NaN and Unnamed columns.
Question 7f: Print out the values for the column POA irradiance CMP22 pyranometer (W/m2). In your own words, based on the documentation, what information does the POA irradiance CMP22 pyranometer (W/m2) variable contain?

```
0         35.1
1         21.0
2         29.9
3         29.5
4         52.1
         ...
38924    152.0
38925    118.1
38926     89.2
38927     38.0
38928     27.2
Name: POA irradiance CMP22 pyranometer (W/m2), Length: 38929, dtype: float64
```

POA irradiance CMP22 pyranometer (W/m2) is Plane-of-Array (POA) Irradiance measured by the CMP 22 device. The CMP 22 pyranometer was used to measure the amount of solar irradiance, expressed in watts per square meter, that the PV module surface received at the specified time.





Question 7g: A column we know that we want to use is the 'POA irradiance CMP22 pyranometer (W/m2)' column. From the documentation each POA irradiance value should be less than 1,500 W/m and greater than -5 W/m. Any issues with 'POA irradiance CMP22 pyranometer (W/m2)?

Out[8]: 0 No issues with data quality

Question 7h: What is the maximum POA irradiance number of the year? At what time do we have the maximum POA irradiance? What is the minimum POA irradiance number of the year? At what time do we have the minimum POA irradiance?

Answer: On November 3rd 11:45, we will have maximum POA irradiance of 1439 W/m^2 and on April 12th we will have minimum POA irradiance of 20 W/m^2.

2011-11-03 11:45:10
1439.6
2011-04-12 18:10:12
20.0

Question 7i: What is the maximum solar power output of the year? At what time do we have the maximum solar power output? What is the minimum solar power output of the year? At what time we have the minimum solar power output? Does this result match the maximum and minimum POA solar irradiance time?

Answer: 2011-11-03 11:45:10
104.842
2011-01-28 17:50:12
0.7277

On November 3rd 11:45, we will have maximum power output of 104.842 W and on April 12th we will have minimum power output of 0.7277 W.

The max time moment of the solar power output is the same as the max time moment of the POA solar irradiance. While the min time moment of the solar power output is not the same as the min time moment of the POA solar irradiance. The reason is because of the 'POA irradiance uncertainty (%).

Question 7j: What are the datatypes for the `Time Stamp (local standard time) yyyy-mm-ddThh:mm:ss`, `Pmp (W)` and `POA irradiance CMP22 pyranometer (W/m2)`?

Answer: Time Stamp (local standard time) yyyy-mm-ddThh:mm:ss:datetime64[ns]
Pmp (W) and POA irradiance CMP22

pyranometer (W/m2): float64

Question 7k: The next column to check for the analysis is the `Time Stamp (local standard time) yyyy-mm-ddThh:mm:ss` column. This column could be the primary key so look for duplicates. How many duplicate entries are there?

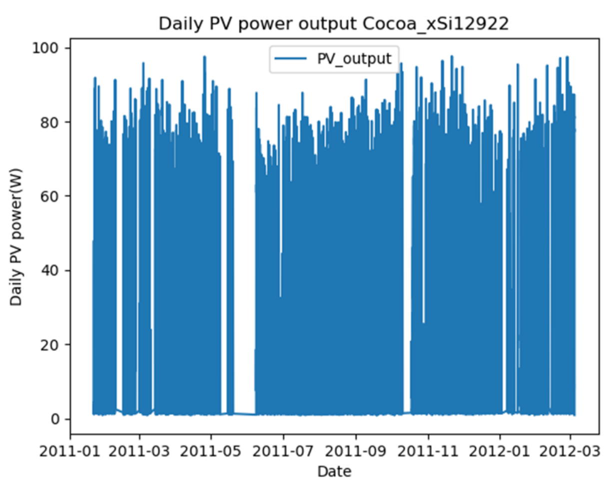
Answer: No duplicate date

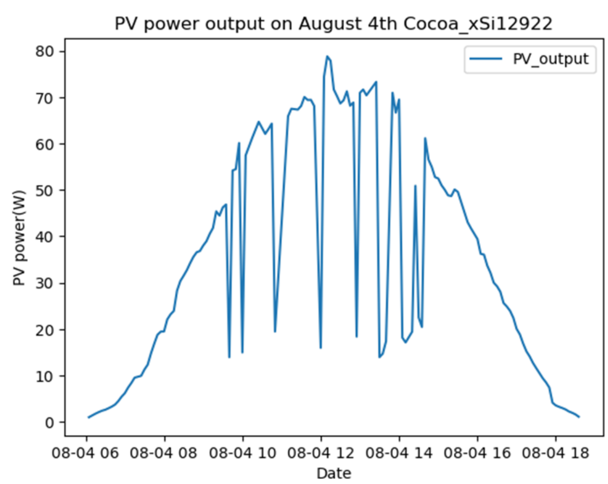
Question 7l: Now that we have a basic understanding of the data, what is the efficiency of Cocoa_xSi12922 PV module?

$$Efficiency = \frac{Solar\ irradiance}{Solar\ power\ output}$$

0.07336392059464329

Question 7m: Plot the PV power output for all samples in the dataset for Cocoa_xSi12922 PV module. Plot the PV power output for one typical summer day August 4th in the dataset for Cocoa_xSi12922 PV module.





Here we can see the solar power output in one year and one typical summer day in Florida. From the measurement, the peak solar output is usually at noon time from 12-2pm. Actually from the yearly solar power output measurement data, we can find that solar panels can also generate lots of solar power in Florida during winter time. That's why Florida is a good place to deploy solar projects that can generate good solar power all year long. Besides, we can also find that solar energy is super fluctuating which is a big challenge to the power system. That's why our power system engineers need to get prepared for challenging renewable energy brought to us.

Question 7n: Now we have a clear idea of the PV output pattern of xSi12922 PV module in Cocoa, Florida. As calculated in question 3l, its efficiency is 7.34%. Do the same thing from question 3a to 3n for other 9 PV modules and plot their efficiency in one histogram plot. Which type of PV module has the best efficiency and which has the worst efficiency?

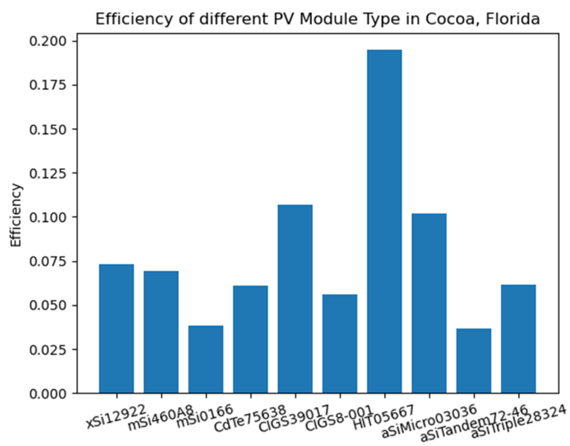
Answer: {'xSi12922': 0.07336392059464338, 'mSi460A8': 0.06905144635822871,

'mSi0166': 0.038229992447086664, 'CdTe75638': 0.061019986832179386,

'CIGS39017': 0.10671866402343981, 'CIGS8-001': 0.05613830801399357,

'HIT05667': 0.1946911521440573, 'aSiMicro03036': 0.10212198348052247,

'aSiTandem72-46': 0.036632746928432276, 'aSiTriple28324': 0.061544757747671916}



Conclusion

PV module mSi0166(Multi-crystalline silicon) has the worst efficiency of 3.82% while HIT05667 (Amorphous silicon/crystalline silicon (HIT)) has the best efficiency of 19.47%. Even using the same technology Multi-crystalline silicon to manufacture PV modules, manufacturer 1's mSi460A8 has a better efficiency 6.9% than manufacturer 2's mSi0166 efficiency 3.8%.

8) Reference

[1] National Solar Radiation Database (NSRDB): https://nsrdb.nrel.gov/

[2] W. Marion et al. , "User's Manual for Data for Validating Models for PV Module Performance," 2014, Accessed: Jul. 31, 2022. [Online]. Available: https://www.nrel.gov/docs/fy14osti/61610.pdf.