Worksheet designed by Jessica Kulp and Chris Orban

Solar Panel Data Analysis:

Solar panels convert the energy from sunlight, a renewable resource, into electricity. You probably realize that their energy production fluctuates throughout the day, peaking around noon and dropping at night. But how does their energy production change over longer periods of time, like months or even years? In this activity, you will analyze multiple years of daily energy production measurements from some solar cells on a teacher's home in Washington, DC.

PART 1: Exploring the raw data

1. Our dataset will have two columns: Energy produced in units of Watt hours (Wh), and time/date. Daily measurements are reported from 2018 to 2022. **Before you look at the data, make some predictions:**
   - What will you see in the data? What story might it tell us?
   - What might impact the energy produced by solar panels over the months of one year? Over several years?
2. Energy production is expected to change not just over the course of each day, but over the course of a year. For example, higher temperatures and longer days in summer mean more energy received from the sun. You might've also considered how solar panels get worn down (become less efficient) over multiple years. **Draw a rough sketch or explain in words what you think the energy production graph might look like.**

3. **Open the data file: http://go.osu.edu/solar1 and save solar1.csv to your computer**. We will use this data to investigate annual and year-to-year variation in energy generation.

4. **Plot this data**. In a spreadsheet program (Excel or Sheets) highlight all the data and then click insert chart. Scatterplots and line graphs are usually good for showing numerical data over time. Energy generation should be on the y axis and time is on the x-axis. **Consider these questions:** What do you see in this plot? What can you *not* see in this plot? Can you tell whether energy production is increasing, decreasing or staying the same when comparing the first year of data to later years? Why or why not?

5. Although there is a lot of day-to-day variability, you should see clear seasonal changes in the plot with increased energy duration in the summer and the least energy generation during the winter. Long-term, year-to-year changes are harder to discern. Just by looking at the plot, you may be able to see a slight decrease in peak energy generation comparing the first year to later years, but there is some variation and some winters are more harsh than others. A statistician might say that there is “noise” in the energy generation data on both day-to-day and year-to-year timescales. As an example of how to deal with this, below is a screenshot of how some professional software provides a summary of energy generation data. **What do you notice about the year-to-year trend below?**

   Figure 1.

   This report provides up to five years energy production ending with the year 2021.

Month	2017	2018	2019	2020	2021
January	338 kWh	392 kWh	365 kWh	318 kWh	337 kWh
February	463 kWh	328 kWh	380 kWh	424 kWh	392 kWh
March	698 kWh	755 kWh	606 kWh	546 kWh	839 kWh
April	759 kWh	730 kWh	699 kWh	710 kWh	775 kWh
May	778 kWh	882 kWh	772 kWh	753 kWh	793 kWh
June	876 kWh	765 kWh	799 kWh	836 kWh	738 kWh
July	817 kWh	863 kWh	905 kWh	869 kWh	784 kWh
August	844 kWh	801 kWh	825 kWh	784 kWh	740 kWh
September	744 kWh	625 kWh	885 kWh	782 kWh	767 kWh
October	616 kWh	623 kWh	701 kWh	522 kWh	548 kWh
November	495 kWh	342 kWh	469 kWh	569 kWh	539 kWh
December	391 kWh	347 kWh	395 kWh	287 kWh	332 kWh
Total	**7,818 kWh**	**7,452 kWh**	**7,801 kWh**	**7,401 kWh**	**7,584 kWh**

PART 2: Reorganizing and aggregating the data

1. **Breaking down and aggregating the data can help us see broader trends through the noise**. We can break the data into monthly and yearly sums (totals), like they did in Figure 1 to better understand the year-to-year trends in solar1.csv

First, separate the raw data by year. We will do this in a new sheet/tab. When transforming or reorganizing data, it's always a good idea to make changes in a new version in case you make mistakes. Create a new sheet/tab by clicking the "+" near the bottom of the screen. Copy all the data in the original sheet (CTRL+A to select all) and paste it in the new sheet. Here, we will keep Column A as our date. Column B will be energy production for 2018. Column C will be energy production for 2019, and so on. Use Cut and Paste to reorganize the data this way. Your first row of values should look like this:

1/1/2018	12,481	5,768	14,443	1,482	1,846

Note: It doesn't necessarily make sense to include the year in the dates because we are using Column A to describe the dates for all the years. Ideally we would have things labeled clearly and accurately like this:

Date/Time	Energy Produced (Wh)				
	2018	2019	2020	2021	2022
1/1	12,481	5,768	14,443	1,482	1,846

Optional tasks for more advanced spreadsheet skills: Fix the year labeling issue by getting your spreadsheet in the format shown above. Change the format of all the date cells to MM/DD (no year). Insert a new row below the headers (Date/Time and Energy Produced) and put the year labels there. Merge the "Energy Produced" header cell with the 4 cells to its right using Merge and Center. Feel free to ask questions in a google search for advice, it can be surprisingly helpful.

2. **Set up a new table to hold the monthly total energy generation, somewhere to the right of the raw data.** It should be organized like Figure 1 above, with month labels in the first column and year labels in its first row. There should also be a "Total" label after December. Note that our data goes from 2018-2022.

3. **Learn about the SUM function**: SUM is one of many built-in spreadsheet functions. Google Sheets and Excel use the same syntax for it: `=SUM(individual cells names or a range of cells)`. We always type `=` before a spreadsheet function name. A computer scientist would say the parentheses

hold the “argument(s)” of the spreadsheet function. In math we might call it the “input(s)”. Range names are written like A1:A5, which corresponds to cells A1, A2, A3, A4, and A5, or A1:B2, which corresponds to cells A1, A2, B1, and B2. You can type this out manually, OR: After you’ve typed =SUM(, highlight the range of cells by clicking and dragging, then press enter. Excel/Sheets will fill in the range name for you.

4. **Use this function to calculate the total energy production in January 2018**. Type your formula with the correct range into the January 2018 cell of your new table. You should find that the total energy generation in January 2018 was 409,419 Wh.
5. **Repeat the calculation for each month of 2018**. If you do this using autofill, be very careful because it is easy to make a mistake! Be sure to check the cell values to make sure it is summing up the energy generation during the period of time appropriate to that month.
6. It can be tedius to correctly configure the monthly energy generation for 2018. You might wonder if there is an easier way to do 2019-2022, and there is! Autofilling down the 2018 column is not a great idea, but we can save time using an **autofill horizontaly across each month’s row to complete the table**.
7. **Calculate the total energy production for each year using the sum function to add up the monthly energy production.** After calculating the total energy generation in 2018, try using a horizontal autofill for the other years. **Looking at your yearly totals, what can you notice about the change over the years?**

   Optional: The total energy generation for one year in Wh is a pretty big number! In your spreadsheet create a new row in order to convert your units from Wh to kWh for the total energy generation in each year

PART 3: Modeling and prediction

1. **Visualize the change in yearly energy production.** Highlight the data you want to plot (yearly totals and years, i.e., the first and last rows of your new table) and insert a scatterplot. **Do you see a clear year-to-year trend?**

2. **Add a linear trendline and display the best-fit line equation**. In Excel, you can find trendline options in the Chart Elements menu ("+" icon that appears by the top right corner of the plot). In Google Sheets, go to the Customize tab of the Chart Editor and you'll find trendline options under Series. **Does a linear model adequately represent any trends in your data? At what rate is the energy production changing over time?**
3. An important fact is that the efficiency of solar panels can change over time, which can lead to year-to-year decreases in energy generation. **According to your linear model, in how many years will the efficiency decrease by 20%** (i.e. compared to the efficiency in 2018)?
4. Using what you know now about how the efficiency decreases, **after how many years should you replace the solar panels?** Assume that it costs $15,000 to replace the solar panels and that the cost of electricity is $0.15 per kWh.