DataSpace

Detailed Instructions

Overview

Your project is composed of the following main tasks:

1. Get all our data loaded and in shape.
2. Create a neatly formatted version of the data.
3. Show summaries of the demand over time.
4. Show the effect of weather on demand.
5. Show the impact of holidays on demand.

The Data

AusEnergy has provided a file called **energy_demand.csv** showing daily energy demand and price information from 2015-2020. It also includes details of the weather on each day.

Here are the columns and their definitions:

- Date: datetime, the date of the recording.
- Demand: float (i.e. a decimal value), total daily electricity demand in MWh.
- RRP: float, a recommended retail price in AUD$ / MWh.
- Min_temperature: float, minimum temperature during the day in Celsius.
- Max_temperature: float, maximum temperature during the day in Celsius.
- Solar_exposure: float, total daily sunlight energy in MJ/m^2.
- Rainfall: float, daily rainfall in mm.

I want you to analyse demand on holidays vs. non-holidays, so I've pulled together a spreadsheet showing which of these days are school days and which are public holidays. You can find this in **holidays.xlsx**.

DataSpace

Task 1: Get Data Into Shape

It's a good idea to bring all our data into the spreadsheet and perform any calculations before trying to summarise and report on the data. Remember that we don't always know what additional data and calculations we need right at the start of the project, so be prepared to come back to this step again if necessary!

1.1 Create the Spreadsheet and Load the Data

1. Create a new spreadsheet called **AusEnergy Demand Analysis.**
2. Load the **energy_demand.csv** file into a tab called **Energy Demand.**
3. Open the **holidays.xlsx** file and copy the contents into your main spreadsheet in a tab called **_holidays**.

A quick tip: I like to use the convention of prefixing worksheets that just contain working data and calculations with an underscore. That way, I know which ones to hide later.

1.2 Break the Date Column Into Components

Breaking out the date column will make it easier to analyse the data by things like day of week or month.

1. Create a new column called **Day of Week** and show the day of the week from the date.
2. Create a new column called **Month** and show just the month from the date.
3. Create a new column called **Year** and show just the year from the date.
4. Create a new column called **Week** and compute the week number in the year from the date (i.e. the first week in Jan is week 1, the second is week 2, etc.).

Tip: Use the Excel TEXT function to get the Day of Week and Month as string values rather than using the WEEKDAY and MONTH functions, which return integer values. If you are working in Google Sheets, look here. Note that you can copy/paste the whole date column or reference the original dates in the new column, If you are using Numbers, you should use formulae to do this as explained here.

DataSpace

1.3 Bring in the Holiday Information Using VLOOKUP

You should have the holiday information in a tab called _holidays. You will need to look up dates in this sheet to determine if they are school days or holidays.

1. Use a VLOOKUP to add a **School Day** column to the main sheet, pulling in the school_day column from the Holiday sheet. You should show a Y if it is a school day and an N if it is not.
2. Use a VLOOKUP to add a **Holiday** column to the main sheet, pulling in the holiday column from the Holiday sheet. You should show a Y if it is a holiday and an N if it is not.

Tip: Review the course material here if you are unsure how to do this. Note that the VLOOKUP will only pull in values that are present in the holiday sheet. Other values will show as #N/A. You can use the Excel IFERROR function to trap the #N/A values and turn them into Ns.

1.4 Calculate the Revenue

1. Create a new **Revenue** column as a calculation of demand x Recommended Retail Price (RRP).

Tip: Review the course material here if you are unsure how to do this.

By the end of task 1, you should have:

- **Created the spreadsheet and loaded the data.**
- **Broken out the date column into its components (day of week, month, year, week).**
- **Used VLOOKUP twice to pull holiday information from one sheet to another.**
- **Used a formula to calculate the revenue.**
- **Created two worksheets: Energy Demand and _holidays.**
- **Your Energy Demand worksheet should look something like this:**

DataSpace

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	date	demand	RRP	min_temperature	max_temperature	solar_exposure	rainfall	Day of Week	Month	Year	Week	School Day	Holiday	Revenue
2	01/01/2015	99635.03	25.63369643	13.3	26.9	23.6	0	Thursday	January	2015	1	N	Y	2554014.1
3	02/01/2015	129606.01	33.13898756	15.4	38.8	26.8	0	Friday	January	2015	1	N	N	4295012
4	03/01/2015	142300.54	34.56485483	20	38.2	26.5	0	Saturday	January	2015	1	N	N	4918597.5
5	04/01/2015	104330.715	25.00556024	16.3	21.4	25.2	4.2	Sunday	January	2015	2	N	N	2608848
6	05/01/2015	118132.2	26.72417628	15	22	30.7	0	Monday	January	2015	2	N	N	3156985.7
7	06/01/2015	130672.485	31.28231073	17.7	26	31.6	0	Tuesday	January	2015	2	N	N	4087737.3
8	07/01/2015	153514.82	48.31230938	18.9	37.4	20.7	0	Wednesday	January	2015	2	N	N	7416655.5
9	08/01/2015	142015.655	49.1172803	23.1	28.2	13.5	19.4	Thursday	January	2015	2	N	N	6975422.7
10	09/01/2015	121801.155	34.49067545	16.5	18	3.1	1.2	Friday	January	2015	2	N	N	4201004.1
11	10/01/2015	103043.66	20.2298249	13.6	21.7	5.6	5.2	Saturday	January	2015	2	N	N	2084555.2
12	11/01/2015	99865.755	18.23476806	15.6	27.5	29.9	0	Sunday	January	2015	3	N	N	1821028.9
13	12/01/2015	131261.125	33.69481017	16.1	31.3	31.6	0	Monday	January	2015	3	N	N	4422818.7

Task 2 Create a Nicely Formatted Version of the Data

Spreadsheet programs have some really nice features that can turn pages of boring data into exciting colourful tables! Not only will this impress Simon, but it will also help you to get to know the data.

2.1 Format the Header Row and Column

Working with the Energy Demand worksheet tab:

1. Format the top header row and first column (the date column) using neutral colouring (i.e. grey or black) to help it stand out.
2. Format the column headings to make them easier to read (capitalise, use spaces rather than underscores).
3. Freeze the top row and first column, so they don't move as you scroll through the data.
4. Order the columns, so all the date ones are first, weather ones next and the demand, RRP, and revenue.

Tip: Review the course material here if you are unsure how to do this.

2.2 Format the Values

1. Format the date into a more friendly format, with the month name rather than a number.
2. Format the demand column to hide the decimal places and use comma separators for the thousands.

3. Format the RRP as a currency
   (these numbers are Australian dollars)
4. Format the revenue as a currency (again, these are Australian dollars)
5. Make sure all columns are an appropriate width,
   so all data values are visible

Tip: Review the course material here if you are unsure how to do this.

2.3 Use Conditional Formatting to Add Colour

1. Use conditional formatting to show colour scales for the demand column (low demand in green, high demand in red).
2. Use conditional formatting to show RRP values below 20 in red and RRP values above 200 in green.
3. Use conditional formatting to show Revenue values below 1,000,000 in red and above 10,000,000 in green.
4. Use conditional formatting for the four weather columns. Use a suitable conditional formatting approach, e.g. data bars in Excel, colour scales, or min/max values. Choose appropriate colours for each (think of meaningful colours for each weather condition!).

Tip: Review the course material here if you are unsure how to do this.

By the end of task 2, you should:

- **Have the Energy Demand worksheet attractively formatted with clear headings.**
- **Have the Energy Demand worksheet attractively formatted with clear headings conditional formatting.**
- **Your spreadsheet might look something like this:**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date	Day of Week	Month	Year	Week	School Day	Holiday	Min Temperatu	Max Temperatu	Solar Exposure	Rainfall	Demand	RRP	Revenue
2	01 January 2015	Thursday	January	2015	1	N	Y	13.3	26.9	23.6	0	99,635	$25.63	$2,554,014.11
3	02 January 2015	Friday	January	2015	1	N	N	15.4	38.8	26.8	0	129,606	$33.14	$4,295,011.95
4	03 January 2015	Saturday	January	2015	1	N	N	20	38.2	26.5	0	142,301	$34.56	$4,918,597.51
5	04 January 2015	Sunday	January	2015	2	N	N	16.3	21.4	25.2	4.2	104,331	$25.01	$2,608,847.98
6	05 January 2015	Monday	January	2015	2	N	N	15	22	30.7	0	118,132	$26.72	$3,156,985.74
7	06 January 2015	Tuesday	January	2015	2	N	N	17.7	26	31.6	0	130,672	$31.28	$4,087,737.28
8	07 January 2015	Wednesday	January	2015	2	N	N	18.9	37.4	20.7	0	153,515	$48.31	$7,416,655.48
9	08 January 2015	Thursday	January	2015	2	N	N	23.1	28.2	13.5	19.4	142,016	$49.12	$6,975,422.73
10	09 January 2015	Friday	January	2015	2	N	N	16.5	18	3.1	1.2	121,801	$34.49	$4,201,004.11
11	10 January 2015	Saturday	January	2015	2	N	N	13.6	21.7	5.6	5.2	103,044	$20.23	$2,084,555.20
12	11 January 2015	Sunday	January	2015	3	N	N	15.6	27.5	29.9	0	99,866	$18.23	$1,821,028.88
13	12 January 2015	Monday	January	2015	3	N	N	16.1	31.3	31.6	0	131,261	$33.69	$4,422,818.69

The exact format will depend on whether you use Excel, Sheets or Numbers.

Task 3 Summarise the Extremes in Demand and Revenue

Now that you have all your data in place, you can start to do some analysis. First summarise the min, max and mean values for demand and revenue.

3.1 Create a Summary Table

1. Create a new worksheet table called **Summary**.
2. Create a small summary table that looks like this:

Demand and Revenue Summary				
	Demand	Date	Revenue	Date
min				
max				
mean				

3. Use formulae to enter the min, max and mean demand.
4. Use formulae to enter the min, max and mean revenue.
5. Use formulae to enter the date on which the min and max demand occurred.

   Tip: You can use the Excel INDEX function to find the value in a particular row and the Excel MATCH function to find the row that matches a particular value.

6. Use formulae to enter the date on which the min and max revenue occurred.
7. Format the demand to 0 decimal places and thousand separators.
8. Format the revenue to 0 decimal places and currency format (Australian dollars).

3.2 Create a Summary Table for Weekends

1. Create another small summary table in the Summary worksheet that looks like this:

	Demand	Date
max Saturday		
max Sunday		

2. Use formulae to enter the minimum and maximum demand for a Saturday and Sunday.

   Tip: You can use the Excel MAXIFS function to find the maximum value given a particular condition. In this case, the condition is that the day of the week is a Saturday or a Sunday.

3. Apply the same formatting as you did for task 3.1.

By the end of task 2, you should:

- **Have a new worksheet called Summary containing two small summary tables for demand and revenue.**

Task 4 Show Summaries of the Demand Over Time

Now that you have all your data in place, you can start to do some analysis. First, try to understand a bit about the demand over time.

4.1 Create a Table With the Top and Bottom 20 Demand Days

1. Copy the Energy Demand data into another worksheet tab and call it **Top and Bottom 20 Demand**.
2. Sort the data on the Demand column (highest to lowest).
3. Add a Rank column as the first column. This will show the demand rank (1 for the day with the highest demand, 2 for the next highest, etc.).
4. Remove rows below the top 20 and above the bottom 20. You should have two blocks of rows - the top 20 and bottom 20. They should look something like this, but of course, with the values in the tables filled in.

Top 20 Days for Demand

Rank	Date	Day of Week	Month	Year	Week	School Day	Holiday	Min Temperature	Max Temperature	Solar Exposure	Rainfall	Demand	RRP	Revenue
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														

Bottom 20 Days for Demand

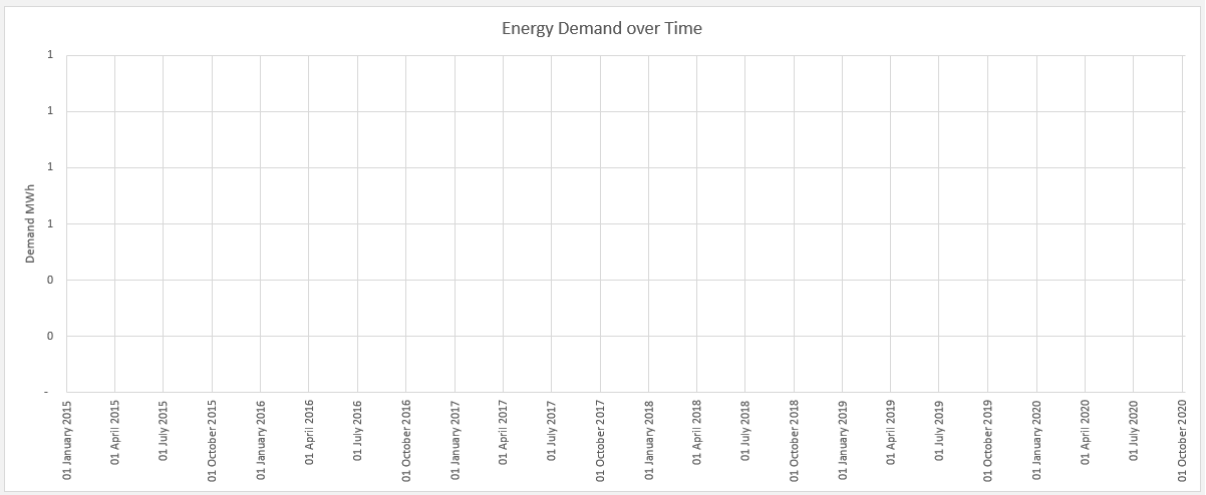
Rank	Date	Day of Week	Month	Year	Week	School Day	Holiday	Min Temperature	Max Temperature	Solar Exposure	Rainfall	Demand	RRP	Revenue
2087														
2088														
2089														
2090														
2091														
2092														
2093														
2094														
2095														
2096														
2097														
2098														
2099														
2100														
2101														
2102														
2103														
2104														
2105														
2106														

5. Format the spreadsheet so the two blocks are clearly labelled, as shown above.
6. Write some comments in a cell about what you observe about the top and bottom 20 rows.

4.2 Create a Chart of Demand Over Time

1. Create a new worksheet called **Demand Over Time**.
2. Create a line chart of electricity demand over time (with time on the x-axis).
3. Set the x-axis data labels to show only every quarter.

4. Label the y-axis (show the name of the axis and the units of measure - refer to the list of columns and definitions at the top of this document for this information).
5. Add a title to the chart.
6. Format to use a thin orange line.
7. Use vertical and horizontal gridlines, making sure they help and don't hinder the reading of the chart. The chart structure will look something like this, but of course, with an orange line showing how the data changes over time.



8. Explain the shape of the chart in comment in a cell below

Tip: Review the course material here if you are unsure how to do this.

By the end of task 4, you should:

- **Have a new worksheet/tab called Top and Bottom 20 Demand, showing the top and bottom days for energy demand.**
- **Have a new worksheet/tab called Demand over Time, showing a line chart.**

Task 5: Show the Effect of Weather on Demand With Scatterplots

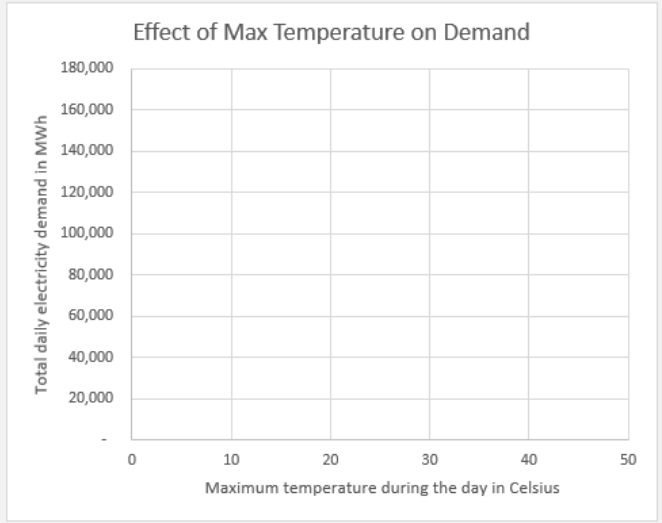
---

Now that you know a bit about how demand changes over time, you can do some more insightful analysis. Some clients love scatterplots! It's often a really great data visualization and helps insights jump off the page! Check out examples like the ones on this Pinterest page, just for inspiration so you can see what you are aiming for! Let's see which weather conditions affect the demand and try to explain the effect.

Tip: Review the course material here to remind yourself how to create charts in Excel.

5.1 Effect of Temperature on Demand

1. Create a new worksheet tab called **Effect of Weather**.
2. Create a scatter plot of max temperature against demand.
3. Add a chart title.
4. Add axes labels, including the units of measure (refer to the list of columns and definitions at the top of this document).
   The chart structure will look something like this, with circles showing the temperature and demand for each day.



5. Explain the shape of the chart.

5.2 Effect of Solar Exposure on Demand

1. Add an additional scatter plot to the Effect of Weather sheet showing the effect of solar exposure on demand.
2. Explain the shape of the chart.

5.3 Effect of Rainfall on Demand

1. Add an additional scatter plot to the Effect of Weather sheet showing the effect of rainfall on demand.
2. Explain the shape of the chart.

Another quick tip: Adjust the style of the plot points to enhance the charts. Use suitable colours for each weather condition! Remember, you want consistency!

5.4 Effect of Temperature and School/Holidays on Demand

Now repeat the scatter plot of the effect of temperature on demand, but split the data so you can see how school days and holidays affect things.

1. Create a new worksheet tab called energy_holidays.
2. Add four new columns: School, No School, Holiday, Working Day.
3. Use formulae in the four columns to select the demand value only in the appropriate columns.

   Another quick tip: You can use the IF function for this. Note that in Excel or Sheets where there is no match, you use NA() to specify this, but in Numbers, you can use an empty string (“”).

   For clarity, here is an example output:

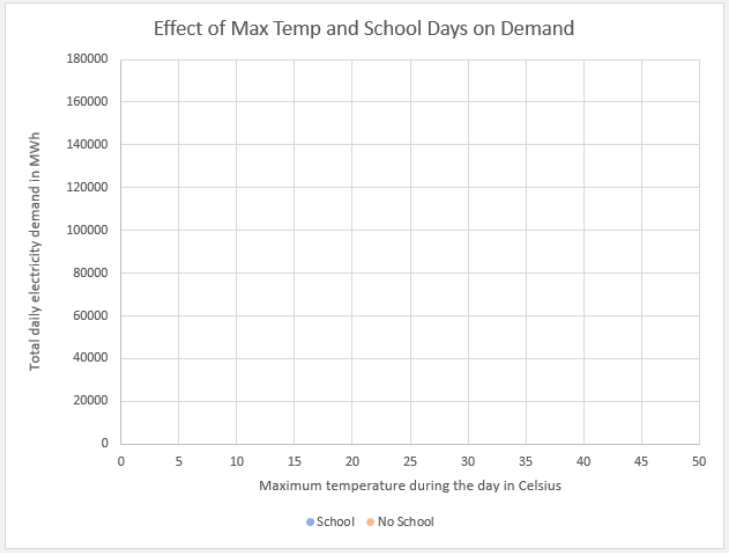
Date	School Day	Holiday	Demand	School	No School	Holiday	Working Day
01 January 2015	N	Y	99635.03	#N/A	99635.03	99635.03	#N/A
02 January 2015	N	N	129606.01	#N/A	129606.01	#N/A	129606.01
03 January 2015	N	N	142300.54	#N/A	142300.54	#N/A	142300.54
04 January 2015	N	N	104330.715	#N/A	104330.715	#N/A	104330.715
...							
22 January 2015	N	N	153232.1	#N/A	153232.1	#N/A	153232.1
23 January 2015	N	N	138095.2	#N/A	138095.2	#N/A	138095.2
24 January 2015	N	N	116310.59	#N/A	116310.59	#N/A	116310.59
25 January 2015	N	N	97959.46	#N/A	97959.46	#N/A	97959.46
26 January 2015	N	Y	103769.48	#N/A	103769.48	103769.48	#N/A
27 January 2015	N	N	118393.31	#N/A	118393.31	#N/A	118393.31

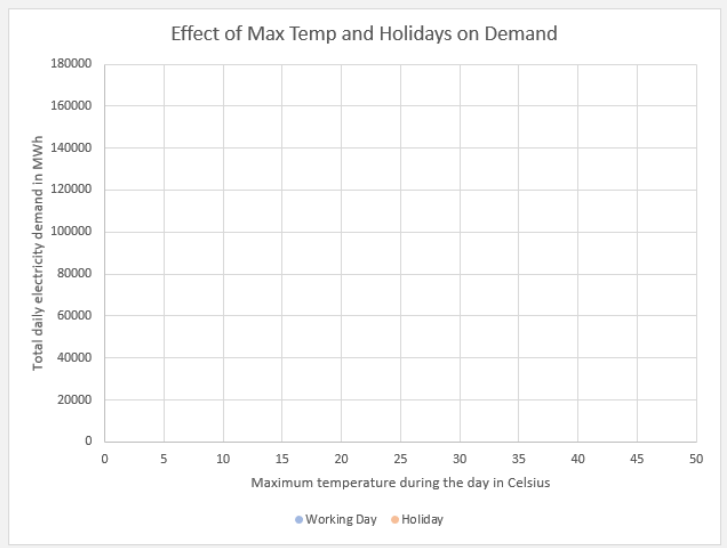
   As you can see in the above screenshot, 1 January 2015 is not a school day, so the demand value of 99635.03 is placed in the ‘No School’ column. It is also a holiday, so the demand value is also placed in the ‘Holiday’ column.

4. Create a new worksheet tab called **Effect of Temp and Hols**.
5. Replot the Effect of Temperature on Demand scatter plot. showing the split between school days and non-school days.
6. Replot the Effect of Temperature on Demand scatter plot showing the split between holidays and working days.
7. Add trend lines to the two charts. For example, add one trendline for School and one for No School for the school holidays chart. Likewise, add one trendline for Working Day and one for Holiday on the holidays chart.

   Tip: You can find out about adding trendlines here. Use the polynomial trendline option (you need to select More Options to see it). This will give you a *curve* that fits better than a straight line. You can find out about polynomial trend lines here.

8. Add titles, axes labels, legends etc.
9. Adjust the style of the plot points to enhance the charts (use suitable colour for each type of day).
   The chart structure will look something like this, with circles showing the temperature and demand for each day. Note the legend at the bottom of each chart. You should have two sets of coloured points on each.





6. Explain the shape of the chart and what the trend lines tell you.

By the end of task 5, you should have:

- A new worksheet called Effect of Weather with three scatter plots.
- A new worksheet called Effect of Temp and Hols with two scatter plots.

DataSpace

Task 6 Show Demand by Different Time Slices With Pivot Tables

---

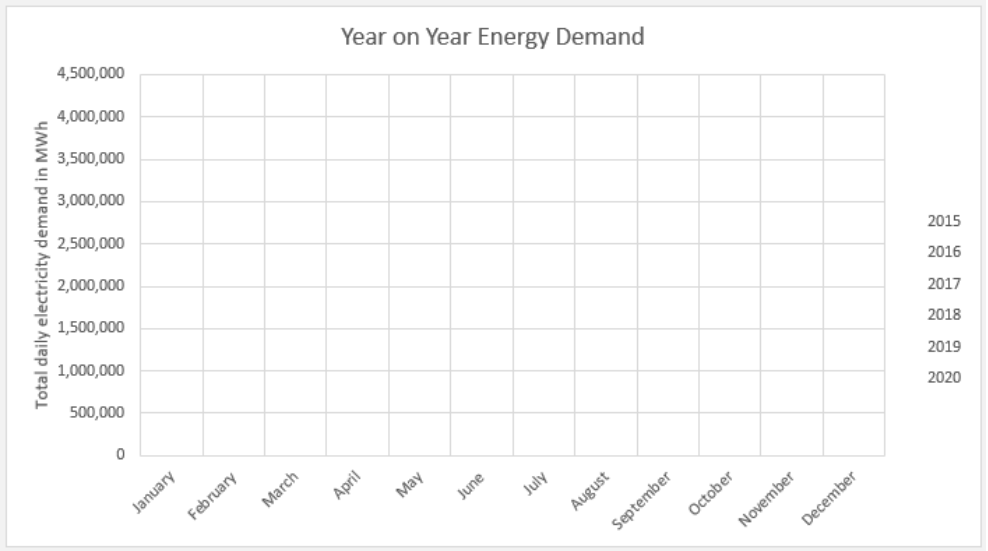
You can now take your analysis even further by examining different slices of the data. Refer back to the course notes on Pivot Tables to refresh your memory on how this functionality works. Note that if you are using Numbers, pivot tables don't exist in the same way as in Excel and Google Sheets. Instead, you can use the categories feature to get a similar effect.

6.1 Year-on-Year Comparisons of Demand

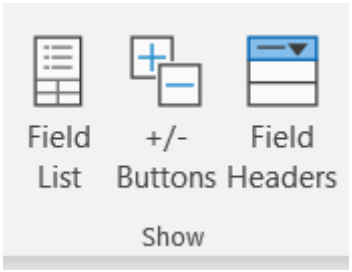
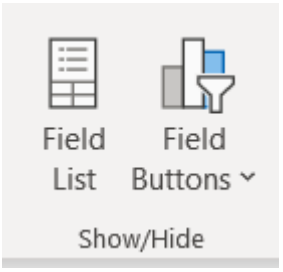
1. Create a pivot chart using all the data in the **Energy Demand** worksheet tab except October 2020. Create this in a new worksheet tab called **YoY Demand**. Exclude Oct 2020 because there is only data for a few days, so it will give misleading results for the sum of Demand if included!
2. Select **Year** on the columns (Series), **Month** on the rows (Categories), and use the **sum of Demand** as the value.
   The pivot table will look something like this, but with the values in the cells:

Year on Year Demand							
Sum of Demand	2015	2016	2017	2018	2019	2020	Grand Total
January							
February							
March							
April							
May							
June							
July							
August							
September							
October							
November							
December							
Grand Total							

3. Plot a **line chart** with one line per year.
   The line chart will look something like this, but with a line for each year:



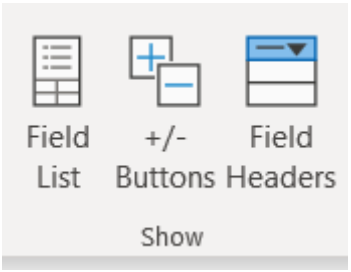
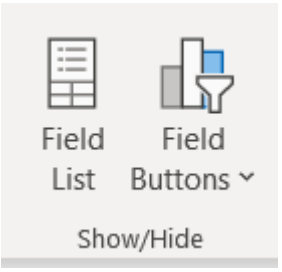
4. If you are using Excel, remove field headers, field buttons and +/- buttons (You can do this by clicking on these buttons in the PivotTable Analyse menu:)

 Pivot Table	 Pivot Chart

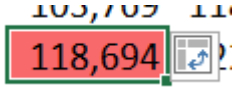
5. Apply suitable formatting to make the chart look nice.
6. Write some comments explaining what you observe below the chart.

6.2 Variation of Demand by Day of Week

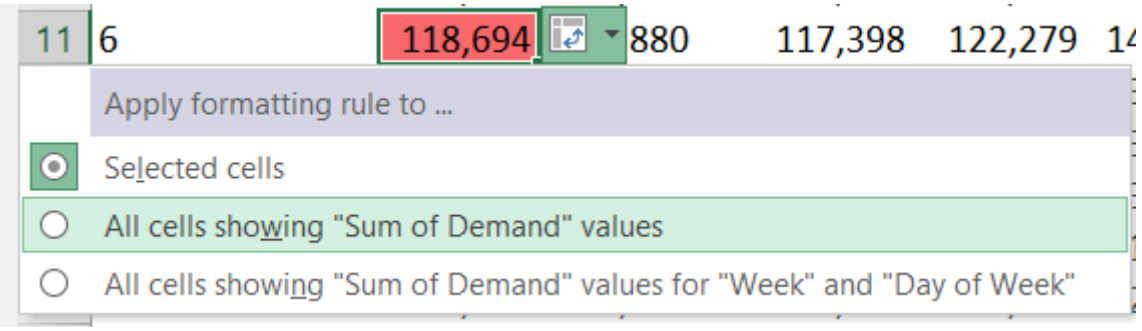
1. Create a pivot table using all the data in the **Energy Demand** worksheet tab. Create and name this new worksheet tab: **Demand by Day of Week**.
2. Select both the **Year** and **Day of Week** on the columns and **Week** on the rows.
3. Remove the totals.
4. If you are using Excel, remove field headers, field buttons and +/- buttons (You can do this by clicking on these buttons in the PivotTable Analyse menu).

 Pivot Table	 Pivot Chart

5. Format numbers with 0 decimal places and comma separators.
6. Apply conditional formatting to show the demand level (green for low, red for high).
   For Excel: Don't select all the cells in the pivot table. Just select one cell and apply the formatting. You will see a little button appear next to the cell:



Click on the button and then you can apply the formatting to the whole pivot table:



7. Click on the button and then apply the formatting to the whole pivot table.
8. Format with bold lines around the years.

DataSpace

By the end of task 6, you should:

- Have a new worksheet called YoY Demand showing a pivot table and chart of demand across the months for each year.
- Have a new worksheet called Demand by Day of Week showing a pivot table with conditional formatting of the high/low demand days split by week number and day of week.

Task 7 Tidy Up

---

1. Add a cover sheet, introducing yourself, your company and summarising what you did.
2. Hide any worksheets that are not part of your report (these should be the data only ones prefixed with _).
3. Make sure all cells are readable.
4. Make sure all charts are clear and labels readable.

By the end of task 7, you should:

- Have a clean and tidy spreadsheet ready to present.