AUSTRALIAN DATA SCIENCE EDUCATION INSTITUTE

Weather prediction accuracy project

We often hear people complain that the weather forecast didn't match the weather that actually showed up. The Bureau of Meteorology (BoM) uses past weather to predict future weather, but the climate is changing, which means the way it behaved in the past might not be all that useful in predicting what it does tomorrow. In this project we're going to look at how accurate weather forecasts are, and compare the BoM forecasts with ones from other sources like weather.com, weatherzone, and accuweather. We will also look at how accurate the long range forecasts are, and how different the long range accuracy is compared to the predictions for the next day.

In this project you're going to learn a lot of Data Science skills, including collecting, analysing, and visualising your data.

Year levels

Years 5 and 6

Curriculum Areas

CURRICULUM AREA	STRAND	SUB-STRAND
Science	Science Understanding	Science as a human endeavour
	Science Inquiry Skills	Questioning and predicting Recording and processing Analysing and evaluating Communicating
Mathematics	Number and Algebra	Number and Place value Fractions and Decimals
	Statistics and Probability	Data representation & interpretation
Digital Technologies	Data and Information	
English	Reading and Viewing Writing Speaking and Listening	Interpreting, analysing, evaluating Creating texts Interacting with others
Capabilities	Critical and Creative Thinking Questions and possibilities Reasoning	

Equipment needed

- Access to the Internet and a computer
- Excel or spreadsheet program

Success Criteria

In this unit the student will improve their data literacy and data analysis skills by :

Knowing:
- What data is and how it is collected
- That data can help tell a story about the world around us and help us solve problems

Doing:
- Collecting data
- Applying basic mathematical skills to make sense of data
- Sharing the data story with others in different ways.

Understanding:
- How weather forecasts are created using data
- How to check your data for possible errors
- That data can be interpreted in different ways by different people for different reasons

Step 1

- How accurate do you think the temperature forecasts are for your area?

- Estimate the accuracy for 6 day long range forecasts, 1 day forecasts, and on the day forecasts, in numbers of degrees - ie you might expect the 6 day forecast to be off by 4 degrees, the 1 day forecast by only 2 degrees, and the on the day forecast by 0 degrees.

- How different do you expect them to be? Put your estimates in a table. How did you decide what your accuracy prediction should be?

WEATHER FORECAST MELBOURNE	**6 DAY**	**Weather forecast accuracy - my prediction**	**1 DAY**	**Weather forecast accuracy - my prediction**	**Actual daily temp**
Monday, 24th March	*25*	*+ 4 deg C*	*23*	*+2 deg C*	
Tueday, 25th March	*26*				

Sample table: My weather forecast accuracy prediction

Step 2

- For two weeks, use a spreadsheet to record the 6 day and 1 day forecast maximum and minimum temperatures for your area, as well as the forecast for the current day, and the recorded maximum and minimum for the day before.
  As a class, make sure you're **not** all using the same website for the forecasts. Some might use the Bureau of Meteorology, http://www.bom.gov.au/, others might use Weather.com, accuweather, google search results, or a weather app. Use as many different sources within the class as possible.
- Make sure you all do this task **on the same days**, to make sure your data is comparable.
- Make sure your sheet records the source of your data as well, and use the same source every time. You might use a layout like this:

date	6 day max	6 day min	1 day max	1 day min	morning ma	morning min	recorded ma	recorded min
1/1/20	31	11	39	38	11	12	39	10
2/1/20	32	10	34	32	7	8	33	8
3/1/20	26	14	23	25	9	9	20	7
4/1/20	35	20	31	32	12	14	31	13
5/1/20	20	10	15	13	12	12	13	12
6/1/20	22	17	26	25	19	20	26	20

Note: it will actually take nearly 3 weeks, rather than 2, to fill this table out completely, because you want to fill out the above table for 14 days, which means you need to start 6 days in advance (to capture your 6 day forecast period)!

Step 3

- Once you have recorded two weeks' worth of data, make a copy of your spreadsheet so that you can work with the numbers without risking deleting them. This is good practice with any dataset, but it's especially important when working with a dataset like this one that is impossible to recover if you lose it.

- Highlight the columns for 6 day forecasts and compare them with the recorded temperatures for those days. How big a difference is there between the predicted values and the actual temperatures? Now look at the 1 day, and on the day forecasts. Are they closer? Is it easy to tell the size of the difference?

Step 4

Make separate sheets within your spreadsheet for the 6 day, 1 day, and current day forecasts. In each sheet you will need the date, the max and min forecasts for your forecast time (eg 6 days), and the actual recorded data for those days. It will look something like this:

date	6 day max	6 day min	recorded max	recorded min
1/1/20	31	11	39	10
2/1/20	32	10	33	8
3/1/20	26	14	20	7
4/1/20	35	20	31	13
5/1/20	20	10	13	12
6/1/20	22	17	26	20
7/1/20	20	8	21	8

Remember to rename each sheet so that you can easily find the data you are looking for!

Step 5

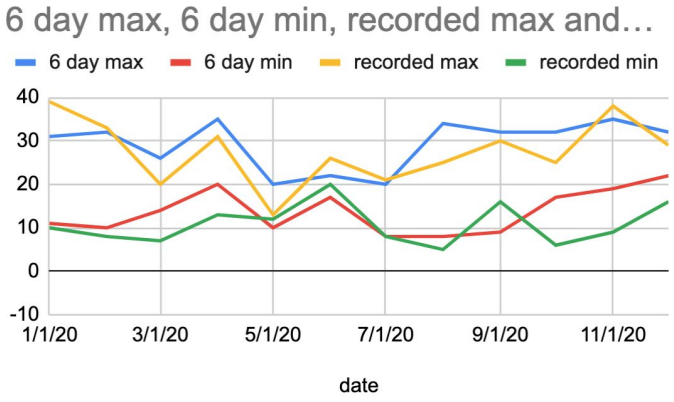
From the sheets, estimate how different from the actual temperature the forecasts usually are. Make a table to record these estimates. It might look something like this:

forecast range	estimated difference from recorded (list)
6 day max	3
6 day min	4
1 day max	3
1 day min	2
on the day max	1
on the day min	1

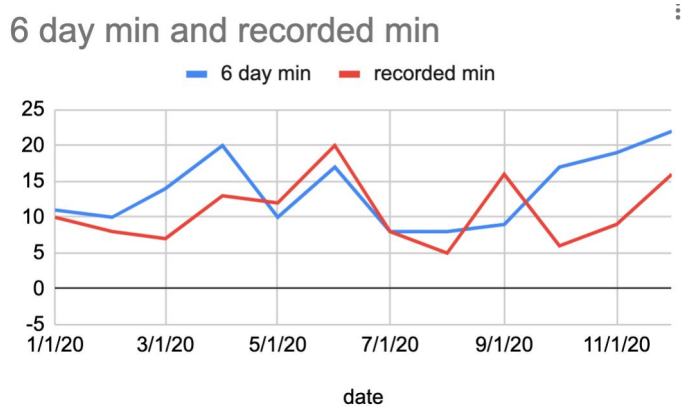
Step 6

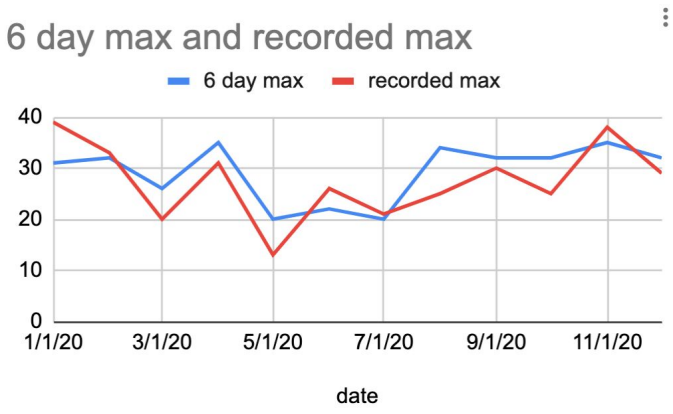
It's hard to understand numbers just from lists of them - sometimes you need a visualisation, or graph, to make it easier to see what's going on.

- Try selecting all of the data in the main sheet and make a line chart with it. There is a LOT of information on this chart.
- Select all of the data in the 6 day sheet, and create a line chart of that data. It may look something like this, although the data in this graph has been faked (ie made up!), so the lines on your graph will probably be quite different!



How close do the 6 day predictions look to the recorded values? Is it easier to see the difference than it was from just looking at the columns? Try graphing the minimum values on a different graph to the maximum values, like the two below. Is it easier to see the differences on the graph with all of the lines, just four lines, or on the graphs with only two? Why do you think that might be?





Step 7

- Make the graphs that are easier to read (either all on one, or in separate graphs) from Step 3 for all of the other sheets: 1 day, and on the day forecasts.
- From the graphs, estimate how accurate the forecasts usually are - ie how many degrees different is the 6 day forecast maximum to the recorded maximum temperature?
- Do this for each of the categories: 6 day maximum, 6 day minimum, 1 day max, 1 day min, etc. Make a new column in your table with these estimates.

forecast range	estimated difference from recorded (list)	estimated difference from recorded (graph)
6 day max	3	2.5
6 day min	4	3
1 day max	3	2
1 day min	2	1
on the day max	1	1
on the day min	1	1

Step 8

- Make two new columns in each of your sheets labelled “max difference” and “min difference”. In these columns calculate the difference for each day between the forecast (max or min) and the actual value. You can use a formula to do this, and copy it down the column. The formula for the top cell will look something like this: =D2-B2.
- Take one or two of the values in that column and calculate the difference yourself, to make sure your formula has worked the way you expected.

date	6 day max	6 day min	recorded max	recorded min	max diff	min diff
1/1/20	31	11	39	10	8	-1
2/1/20	32	10	33	8	1	-2
3/1/20	26	14	20	7	-6	-7
4/1/20	35	20	31	13	-4	-7
5/1/20	20	10	13	12	-7	2

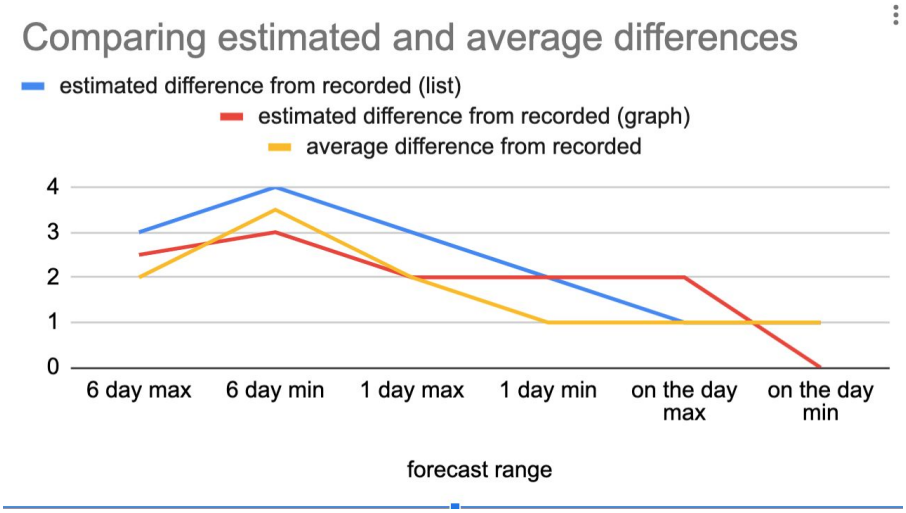
Step 9

- Using those difference columns, calculate the average forecast difference for each forecast category - 6 day max, 6 day min, 1 day max, etc. You can use the average function to do this. Put these averages in your table, next to your estimates. How close were your estimates?

forecast range	estimated difference from recorded (list)	estimated difference from recorded (graph)	average difference from recorded
6 day max	3	2.5	2
6 day min	4	3	3.5
1 day max	3	2	2
1 day min	2	1	1
on the day max	1	1	1
on the day min	1	1	1

Step 10

- Create a graph from your “differences” table.



- Which was more accurate compared with the average: your estimate from the list, or from the graph? Why do you think that might be?
  Note: Any lines on your graph that overlap will mean you can’t see one of them - like in the example above, where the blue line overlaps with the yellow one for the last value, so the blue one is hidden. This might make your graph a bit confusing!

Step 11

- Team up with someone who used a different source for their weather forecasts. First of all, compare the recorded max & min values from the two sources. Are they always the same?
- Now compare the average accuracy of the 6 day, 1 day, and current day forecasts. Which source was more accurate?
- As a class, create an accuracy table for all of the different sources using the average difference values you calculated, like the one below.

forecast range	average difference from recorded BoM	average difference from recorded Weather.com	average difference from recorded accuweather
6 day max	2	4.7	2.2
6 day min	3.5	4.2	3.3
1 day max	2	3.2	2.1
1 day min	1	2	2
on the day max	1	2.5	4
on the day min	1	2.1	1.3

- Which is the most accurate source? Does it differ for the different forecast range? (ie is one source more accurate for 6 days, another for on the day?) Do some of the sources have precisely the same accuracy as each other? Why might this be?

Step 12

- As a class, discuss what the weather was like during the two weeks when you recorded your data. Is the weather in your area more consistent and predictable in some seasons than in others? Can you assume that your results work for the whole year? Is the weather exactly the same from year to year? How different do you think your results would have been if you'd chosen the fortnight prior to the one you used, or the fortnight after? Why?
- Design a data collection experiment more likely to give you results that apply across the whole year.

Step 13

- Make a poster, presentation, talk, or video that shows your family and friends which is the most accurate source of weather forecasts.

Advanced Exercise

- Find data for your area from previous years and compare how different your data is to the historical data. If you can't find data from day to day, look for monthly averages. Is your data above or below the average? Are the maximum and minimum different from the average in the same ways? Can you make generalisations about the way the climate has changed based on these two weeks of data? Why/Why not?

Curriculum Links (VCAA 2016)

Curriculum Area: **Strand:** • **Sub-strands**	**Level 5**	**Level 6**
Science: Science Understanding: • Science as a human endeavour	Scientific understandings, discoveries and inventions are used to inform personal and community decisions and to solve problems that directly affect people's lives.	
Science: Science Inquiry Skills: • Questioning and predicting	With guidance,pose questions to clarify practical problems or inform a scientific investigation, and predict what the findings of an investigation might be based on previous experiences or general rules.	
Science: Science Inquiry Skills: • Recording and processing	Construct and use a range of representations, including tables and graphs, to record, represent and describe observations, patterns or relationships in data.	
Science: Science Inquiry Skills: • Analysing and evaluating	Compare data with predictions and use as evidence in developing explanations.	
Science: Science Inquiry Skills: • Communicating	Communicate ideas and processes using evidence to develop explanations of events and phenomena to identify simple cause and effect relationships.	
Mathematics: Number and Algebra • Number and Place value	Using efficient mental and written strategies and apply appropriate digital technologies to solve problems.	Select and apply efficient mental and written strategies and appropriate digital technologies to solve problems involving all four operations with whole numbers and make estimates for these computations. Investigate everyday situations that use integers. Locate and represent on a number line.
Mathematics: Number and Algebra • Fractions and decimal	Compare, order and represent decimals	
Mathematics: Statistics and Probability • Data representation & interpretation	Pose questions and collect categorical or numerical data by observation or survey Construct displays, including column graphs, dot plots and tables, appropriate for data type, with and without the use of digital technologies	Pose and refine questions to collect categorical or numerical data by observation or survey Interpret secondary data presented in digital media and elsewhere

	Describe and interpret different data sets in context	Construct, interpret and compare a range of data displays, including side-by-side column graphs for two categorical variable
Digital Technologies ● Data and Information	Examine how whole numbers are used as the basis for representing all types of data in digital systems. Acquire, store and validate different types of data and use a range of software to interpret and visualise data to create information Plan, create and communicate ideas, information and online collaborative projects, applying agreed ethical, social and technical protocols.	
English <u>Reading and Viewing:</u> Interpreting, analysing, evaluating <u>Writing</u>: Creating texts <u>Speaking and Listening:</u> Interacting with others	Use comprehension strategies to analyse information, integrating and linking ideas from a variety of print and digital sources Clarify understanding of content as it unfolds in formal and informal situations, connecting ideas to students' own experiences, and present and justify a point of view or recount an experience using interaction skills Use a range of software including word processing programs to construct, edit and publish written text, and select, edit and place visual, print and audio elements	Use comprehension strategies to interpret and analyse information and ideas, comparing content from a variety of textual sources including media and digital texts Participate in and contribute to discussions, clarifying and interrogating ideas, developing and supporting arguments, sharing and evaluating information, experiences and opinions, and use interaction skills, varying conventions of spoken interactions according to group size, formality of interaction and needs and expertise of the audience Use a range of software, including word processing programs, learning new functions as required to create texts
Capabilities ● Critical and Creative Thinking ● Questions and Possibilities Reasoning	Examine how different kinds of questions can be used to identify and clarify information, ideas and possibilities. ● Identify and form links and patterns from multiple information sources to generate non-routine ideas and possibilities. ● Investigate common reasoning errors including contradiction and inconsistency, and the influence of context.	