In this, the final chapter, we will review what we have covered in this book and look at how the skills we’ve learned can be used in different contexts. With the journey all but over, what does the destination look like and where might we go from here? What else can we use our data science toolkit for?

13.1 Am I a Data Scientist?

Congratulations on making it this far. If you’ve followed the chapters, worked through the examples and everything has made sense, then you have a solid foundation in data science and the ability to apply it in real-world situations.

Of course, just like studying the basics of biology does not make one a doctor, so having a solid foundation in data science does not make one a data scientist. The information and techniques outlined in this book are a stepping-stone: useful in and of themselves and potentially the first step toward becoming a professional data scientist. But what is a “professional data scientist”?

Doctors, nurses, architects, plumbers: disciplines like these require accreditation before the title can be rightfully used. That involves studying a prescribed body of knowledge and passing through some or other
qualification process. There is no universally accepted data scientist accreditation, no governing body, and no final exam: anyone can call themselves a data scientist.

As to the question of whether you are a data scientist: in your job or day-to-day life do you work with data, process and analyze that data, interpret it and produce actionable output? If so, then maybe you are.

13.2 Becoming a Data Scientist

While the loose/informal definition of what makes someone a data scientist is interesting, when companies are hiring data scientists they generally have a clear idea of what they are looking for. This tends to be a set of skills/experiences that relate to how the company already undertakes data science, and it is very common to see jobs advertised requiring specific skills such as coding (e.g., in Python), SQL, statistics and machine learning.

If you are interested in exploring data science further, potentially making a living from it, there are many disciplines to study and a huge range of directions to go in. One of the best ways to get an understanding of what skills matter in the field of data science is to look at job boards.

Most of the skills required can be learned, and if you’ve found that you are comfortable with the material in the book then you probably have the aptitude required to do so. One theme has been consistent throughout this journey, and that is the role that curiosity plays in the process. The skills mentioned are valuable, but what makes a great data scientist is their curiosity, their drive to ask questions. Intuition is also highly prized: being curious about the right things and being able to ask the right questions are traits that computers are nowhere near able to replicate. And remember, you don’t need to be able to think like a computer: we have computers to do that for us.
13.3 Debunking Some Myths

From a data science point of view, myths are hypotheses/facts that have been accepted by people to be true/proven, but which are not based on any reliable or verifiable data. It is somewhat ironic that a field like data science has its own set of myths. The problem with myths are that they often look like barriers to people on the other side, so we’ll take a look at a few and try to debunk them:

Only data scientists do real data science.

- Data science is what we have been doing throughout this book – using scientific methods to gather, process, and perform all manner of operations on data.
- Until a professional standard is established, a data scientist is someone who is employed in a job with “Data Scientist” in the title. The term is used professionally to refer to the analytical aspects of the process. The team member who builds and codes the sensor arrays would not be called the data scientist. But they are participating in data science: it is a team operation, and the data scientist is a member of that team.

Real data scientists do not use Excel/spreadsheets.

- A novice uses Excel excessively.
- An expert never uses Excel.
- A guru uses Excel when it is appropriate.

1Myths that are found to be true cease to be myths and become facts, of course.
You have to be a strong coder to be a data scientist.

- Data scientists will find themselves in situations where coding is required. It is a great skill to have.
- But when some code needs to be written, a data scientist can use a developer to write it for them.
- The key is knowing what to ask the developer for: what are the outcomes needed from the code? Any good developer should have no problem converting well thought-out natural language code into real code.

Only big corporations and universities can afford it.

- It is possible for a single individual on a budget to undertake real, affordable and novel data science experimentation.
- Given the sheer volume of data that is being generated daily, and with so much of it being free and easy to access online, how many fascinating, groundbreaking and useful correlations are there waiting to be found, perhaps only a few clicks away?

You have to be exceptionally intelligent to be a data scientist.

- Data science is an intellectual field, but anyone with the right curiosity and drive can learn the techniques needed.
- Of course highly intelligent people tend to do well in a range of intellectual fields, including data science.
Market research is not real data science.

- Especially in academia, market research is looked down on as being substandard; some might say it is guilty of “selling out” – putting profit before quality.
- There is good and bad market research just like there is good and bad academic research.
- The real difference is that we so often see bad market research splashed on billboards and thrust in our faces on advertising. This does not represent the full range and potential of the market research industry.
- An example of nontrivial market research in post-apartheid South Africa will help illustrate a more positive application.

Back in the 1990s, advisers to the ANC used a small market research company based in Durban (the now defunct Research Surveys) who had an innovative product called “The Conversion Model”. This was a survey technique which was surprisingly effective at measuring predicted behavioral responses from different groups of people to specific messages. In a nutshell, would this message incite or unite?

Mr. Mandela was without question one of the greatest statesmen of his, or any, era, and while the ANC were assured of victory in the national elections it was a priority of his to ensure that all the different race groups in the country were engaged with, and excited by, the process of transforming South Africa into the Rainbow Nation.

Multiple conversion model surveys were commissioned to assess which parts of the ANC’s message resonated the best with different race groups and which ones were most polarizing. Mr. Mandela’s government of national unity did a great job of prioritizing items that helped keep a broad spectrum of the population enthused and optimistic.
13.4 Extrapolating Learnings

With access to the right hardware there is no limit to the range and sophistication of digital instruments that you can build using the techniques outlined in this book. Most of the skills we have covered are transferrable:

**If you can connect one sensor to a microprocessor:**

- Whatever peripheral system you are using (XinaBox, breadboards, SparkFun, etc.), if you have successfully connected one sensor to a microprocessor, then you can usually apply the same method to attach other sensors too. If you choose a peripheral system with a broad range of sensors, you should be able to attach any of these.

- As long as the microprocessor supports it (and most do), you can connect several sensors at the same time and build instruments capable of measuring a broad range of sensor readings.

- And remember, in 2020+ you do not need an engineering degree or a soldering iron to build high-quality and robust digital instruments.

- **Figure 13-1** shows two digital instruments built with xChips. We built the Wi-Fi gateway shown on the left, while the device shown on the right could be used in a high-altitude balloon to record and transmit environmental data.
If you can write block code for micro:bit:

- Then you can write code: you can program.
- There are a LOT of microprocessors that can take block code. And most of them are orders of magnitude more powerful than the micro:bit. You can program these too, using the same techniques.
- **Figure 13-2** shows the Maker.Makecode block coding environment. This is a Microsoft product – the micro:bit MakeCode we used is a “white label” implementation of this block coding environment.
- The interface, syntax, flashing process are very similar – they will be familiar to users of MakeCode.
- **Figure 13-2** shows a selection of the boards that can be coded using block code.
If you can program a microprocessor to read data from a sensor:

- Then you can program it to read data from any sensor. You know how to load libraries/extensions, and the blocks to read sensor data tend to work in a very similar way. Finding the right extension for a specific sensor can be the trickiest part: look for manufacturers who provide easy-to-use (and find) MakeCode extensions for the microprocessor you are using.

- Reading data from more than one sensor is trivial: you should be able to read from an array of many sensors simultaneously.

- Figure 13-3 shows four MakeCode extensions for different types of sensor. The names of the variables change, but in each case the value from the sensor is returned as a simple variable, a number. The paradigm is the same.
If you can analyze temperature and humidity data:

- Then you can analyze any two data sets using the same techniques.
- You should be able to extend this: add other data sets, experiment with different types of visualizations, hunt for correlations, and add trend lines.

If you can connect to an IoT platform:

- Then you can send and receive data from an online platform and access services on the cloud.
- The power of the online platform and the services available can be used to add sophistication to your digital instrument.
If you can process received data from an IoT platform:

- Then you can use/respond to commands or data from that IoT platform. These commands/data have had the processing power of the IoT platform behind them: processing power orders of magnitude greater than is available on the SBC you are using.

- If you can get an IoT platform to send data to a remote digital instrument, then an AI/ML service that the IoT platform is hooked up to can also sent data to it.

If you can hook up a weather station to a ML predictive model:

- Then you can hook up any edge device to any of the many ML predictive models available.

As recently as 20, maybe even 10 years ago, a comparable set of skills would have taken a bright and motivated individual many years to acquire. But with modern technology it is possible to gain those skills more quickly and focus on how to apply them, on the things that the technology can enable, rather than fiddling with the technology itself. Data science is a layer on top of digital technology, and being able to delve into it like we have done here is a luxury that would not have been possible until quite recently: within the lifetime of most readers.

13.5 Applying Our Knowledge to Different Builds

Using the skills outlined in the previous section you could build all sorts of things, such as

- An instrument with an array of sensors that connects to an IoT platform
- And that sends data which is then passed into an ML model
• The ML generates predictions or key metrics which are sent back (perhaps to a completely different device/devices)

• And where a series of real-world electronic switches or motors are turned on/off

• All this time information is output on one or more screens, perhaps on devices in different locations

• And is written to a file and backed up on the IoT platform

While this might sound complex, it is extremely generic: the core architecture of any number of modern inventions/gadgets. With this basic model, you can build a huge range of things, including:

• A tool to monitor the air quality of a classroom, or bedroom, or airport waiting room. Data is collected by sensors in the room and shared by Wi-Fi with an ML service, which analyzes the data and remotely controls a window, blinds, fan, dehumidifier, and air conditioner as appropriate to maintain a healthy environment at minimal energy cost.

• Having trouble sleeping? Build a sensor array to monitor factors in your bedroom. Collect a few weeks’ worth of sensor data and add in data about how well you slept. Identify sensor readings that correlate with you sleeping well; then build an instrument that can manage the relevant factors to give you an optimal environment.

• What about a chess playing robot? Or a robot that can sort different colored jelly beans? Or one that can pick ripe strawberries?
• Building and maintaining a controlled environment such as an enclosure for an “exotic” pet.

• Cities like London have large flood gates which are closed when water levels get too high. But it takes time to close them, and it is often costly to do so. A range of sensor arrays upstream and weather report data all being fed into an ML model which predicts flooding could make for a very efficient control system.

• There are rural villages in India that are visited by wild roaming elephants who are known to eat crops and cause damage. A series of arrays and actuators spread around a village might be capable of setting off loud noises or lights to divert elephants away.

13.6 Ethical Considerations

So, you can now undertake all manner of experimentation and build all sorts of innovative and potentially impactful contraptions. But what kind of things shouldn’t you build?

Obviously nothing that risks causing physical injury or criminal damage: we know a line exists. So, if there is a line between what is OK and what isn’t, then where is that line?

To find the answer, we need to look into ethics:

What is/are “ethics”? Many people would agree that at the core of most ethical arguments is the tenet that if you are engaged in any activity that impacts on other people then you have some kind of responsibility for any effect this activity has on those people. Ethics shines a light on that responsibility.
Data science is used to inform decisions and policies that touch our lives in a wide range of different ways, as we have seen throughout this book. Is ethical consideration necessary though?

Data science can appear to be scientific, unbiased, and impartial: the methods and tools are independent of the data that is being analyzed. The same techniques used to analyze temperature and humidity data would work for religious affiliation and IQ data: a correlation coefficient does not have an opinion.

A study into religious affiliation and IQ data is without doubt going to arouse passion, anger, and hurt and could very easily be manipulated (e.g., different IQ tests have cultural biases). Whether it would be ethical or not to actually gather and study this data is not a judgment we will make here. At the point of conception of an idea, the impact of implementing that idea should be considered though; you are your own ethical guardian.

The field of ethics is fascinating and vast, but we have almost run out of space. The important point to make here is that data science is not exempt: good people can produce bad science.

Most important is to remember that, with the growth of AI and ML, the human component of data science may change or even shrink. But it is humans who create the need for ethical considerations and who need to be the guardians of ethics. The best AI/ML cannot come near to a human in terms of empathy and passion for the value of ethics. As we hand over aspects of the data science process to AI/ML, we need to cling to ownership of this key consideration.

So, before we conclude, an ethical dilemma to ponder:

Is it OK to scare off elephants who are just foraging, and what is the impact on the elephants? Will everyone who answers this question have the same answer? Do the opinions of people whose lives are not affected by the elephants matter? What happens if we extend the metaphor and substitute elephants with groups of migrant people?
13.7 Summary

The original plan for this final summary was to list all the bits and bobs that have been featured in the book: make you feel good about attaining a very useful and practical body of skills and techniques.

Instead the authors would like to completely break the fourth wall and end on a somewhat less conventional note:

Hi. Thank you for investing your time and money in this book.

We started writing back in August 2019, in the pre-covid19 world, and we are finishing it in April 2020. So much has changed since then, and we share your loss and sadness.

We still want to end on an upbeat note though.

This crisis will pass and those that remain will emerge stronger from it. When the war on Coronavirus is finally won, society must, and will, remember the heroes who got us through it: the nurses, doctors, orderlies, truck drivers, shelf stackers, teachers, police, farmers, shopkeepers... the list goes on.

And it includes data scientists, front and center.

During the height of this crisis it is hard to look at a screen and not see a chart showing infection/survival rates, or hear experts talk about models projecting predicted outcomes, or read about the expected impact on the global economy. Data science is at the heart of all of these endeavors, informing the life-or-death decisions that governments are being forced to make on a daily, even hourly, basis. The strategy and implementation of self-isolation is based on data science modeling and the development and testing of new medicines will use data science techniques. Next to medicine, data science might be the best weapon we have in our arsenal.

We will leave you with this final thought: what we have looked at in this book is small and frivolous in comparison to the work described earlier, but our efforts share the same DNA as these endeavors. The great scientists who are leading us through this crisis were once at the same stage of their data science journey as you are now.

Thank you and stay safe.