ICAEW
HELPSHEET

Data science skills for Finance

The continuing growth in data opens up a wide range of opportunities for finance professionals. To take advantage of these opportunities new skills will be required. The introduction and framework below are designed to help organisations and individuals map out the necessary skill development.

23 October 2018

INTRODUCTION

Our interactions with the world are increasingly digital meaning they can be tracked, stored and analysed. Digital data is collected on the products we buy, the people we interact with, what we read, our interests and values, our movements around the world and so on.

With the growth in the 'Internet of Things' there is also the increasing ability to record the status and location of physical objects in real-time. Useable data is no longer just structured data, in the form of defined database fields and numbers, but also unstructured data such as free text, speech, sound and video. To make use of this data requires the application of analytical tools and algorithms, often using machine learning.

In this digital and data centric business environment the composition of finance teams will evolve as finance professionals will increasingly find themselves working in multi-disciplinary teams including data scientists. There will be an increasing demand for 'purple people' – individuals with the skills to act as intermediaries between data scientist ('red people') and decision makers in the wider business ('blue people').

To continue to be a valued business partner at the heart of shaping and delivering the analytics agenda, finance teams will need to develop new skills to manage the increased complexity of data as well as using financial and non-financial data to deliver greater insights to the business.

In the table below we define the capabilities that will be increasingly expected from finance enabling them to integrate wider functional and business skills with data science skills.

Capability	Descriptor
Data strategy	Organisations are increasingly thinking more holistically about data and its value to the business. Indeed a data strategy often needs to be an integral part of an organisation's overall business strategy. Consideration needs to be given to data governance and how much to invest in data collection and data science programmes including both expertise and the relevant technologies. Choosing the right technology in a rapidly changing market is a significant challenge. Making the right decisions in these areas will require wider strategic thinking, overall knowledge of the digital landscape and adapting investment appraisal techniques to new business models.
Ethics	The accounting professions code of ethics provide a good foundation for data analysts. However, these will need to be further developed given the ways in which sensitive data can now be used and abused on a large scale. Privacy, data security, transparency, decision-making bias in algorithms and so on all require careful thought before starting new data analysis projects.
Data intuition	Developing a sense of when data can be applied to solving business problems requires a deep understanding of the business. This is developed through building strong relationships with business stakeholders, getting out into the business and gathering industry information from a wide range of sources. In the data science jargon this is 'domain knowledge'. This knowledge will help you determine how to frame a problem in a way it can be answered with data, what relevant data is available, what additional data will be valuable and the types of analysis to undertake. In the main we want to focus on the most important business problems, the most relevant data and the most promising analytical techniques. However, there is also value in allowing some time for experimentation and playing with data to see if unexpected insights emerge. This will further help develop your data intuition.
Data analysis and interpretation	There are a wide range of tools, algorithms, statistical and mathematical techniques that can be used to analyse data, some of which are discussed below. At a higher level there is a need to understand some basic principles of data analysis and interpretation. These include the implications of the quality of data, sampling method and assumptions going in to any particular approach and understanding the implications, limitations and accuracy of any outputs. Overall it is about being in a position to ask the right, challenging questions of those undertaking the analysis to ensure that the correct conclusions are being drawn.

Communication and data visualisation	As a data analyst your job is to not only interpret the data but to also effectively communicate your findings to other stakeholders, so they can make data-informed decisions. Many stakeholders will not be interested in the technical details behind your analysis. That's why it's very important for you to be able to communicate and present your findings in a way that is easy to understand for your audience, both technical and non-technical. A fundamental element is deciding what the audience needs to know so they can make an informed decision, such as key assumptions and limitations of the analysis, and what you can take responsibility for in terms of recommendations. It is also through effective communication skills that accountants can provide the bridge between data scientists and the business. It can be immensely helpful to be familiar with data visualization tools like Tableau, QlickView and Power BI. It is important to not just be familiar with the tools necessary to visualise data, but also the principles behind visually encoding data, visual design, storytelling and communicating information.
Programming and tools	Data analysts and scientists use a variety of tools and programming languages in their everyday work. You'll use these tools to query and retrieve data from databases, transform and summarise data, develop models and build machine learning algorithms. You should be able to analyse data in one or more tools/programming languages, and have a good grasp of the landscape of the most commonly used data analytics libraries and packages. Excel and SQL are a good place to start. For more sophisticated analysis and building algorithms, Python, the 'r' statistical programming language and SAS are commonly used by data scientists. It is worth noting Python and 'r ' are open source and many other tools can be experimented with for free if not used commercially.
Statistics, maths. modelling and algorithms	Sufficient understanding of the interrelated fields of statistics, maths, modelling, algorithms and machine learning is vital as a data analyst. In particular you should know which techniques to use in a given situation. The range of techniques, some of which can be extremely complex, means that no one can keep on top of all developments in these fields. So it is important to develop information sources and relationships with other analysts to discuss possible approaches. However, at minimum analysts should have an understanding of: • **Statistics –** descriptive statistics to summarise data, sampling approaches, basic probability, distribution types, hypothesis testing, time series analysis, linear regression, inferential statistics to draw conclusions on a population based on a sample, statistical significance and effect sizes

Statistics, maths. modelling and algorithms (continued)	• **Maths** – compounding, functions, basic linear algebra for simple optimisation problems • **Modelling** – developing hypotheses, rules and scenarios about the connections between different business variables and applying data to those models e.g. financial models and Monte Carlo simulation • **Algorithms and machine learning –** using data can be used to create rules and relationships, for example regression. From these foundations analysts will be able to delve more deeply into techniques that are most relevant for the business problems they are trying to solve.
Data wrangling	A less celebrated part of data analysis is collecting, cleaning and linking data so it can be easily explored and analysed. This process is known as ‘data wrangling’ or ‘data munging’ in the data science community. While not as exciting as building advanced models, data wrangling is a task that data scientists can spend up to 50-80% of their time doing. So why do you need to wrangle data? Often, the data you’re analysing is sourced from legacy business systems, disparate external sources and is going to be messy and/or difficult to work with. Because of this, it’s really important to know how to deal with imperfections in data.