codeop

The Top 9 Skills Needed to Break into Data Science in 2022

Table of contents

The growing demand for data professionals

It's easy to see the appeal of a career in and around data science and analytics.

With emerging technologies increasingly driven by data and a seemingly endless number of tools available, anyone with a creative mind and a willingness to work with maths and statistics has the opportunity to build/ advance in an exciting career and work for companies that are using data in ever more interesting ways.

There's **more demand than ever for data scientists and employees with data analytics skills**. Still, with the huge influence of data science across such a wide range of sectors, it's clear that actually becoming a data scientist is not easy. **It's an interdisciplinary field,** with many different roles that require a range of skill sets.

With tons of theories, techniques and tools tied into the field of data science, it can be tricky to unpick what **you** need to get ahead in your chosen career – whether you want to upskill and develop your abilities or start your data science/data analytics journey from scratch. We want to support you on your journey with as much information

as possible to help you get started on the right track.

So let's delve into the details and **break down the essential data science skills needed** in 2022, while finding out exactly how these skills can be applied to your chosen career.

1.
Programming languages

Using code or 'programming languages' is a fundamental part of data science. It's how those working in the field of data science communicate and collaborate and, like verbal languages, there are many different programming languages.

Depending on your chosen data science career, you'll need to be fluent in some important programming languages, while having a good knowledge of others. Here are some of the most widely used programming languages and what they can do:

Python

The world's number one programming language, as well as the most widely used data science programming language. Extremely adapt-able, flexible and user-friendly, Python is an all-purpose language, great for anything from data analysis and data mining to artificial intelligence and machine learning.

Not surprisingly, it has the largest community of any programming

language and a very comprehensive library base, making Python the best code for collaborating with other developers.

WHY YOU NEED THIS SKILL:

- It's beginner friendly – this should be the first code you learn
- It's comprehensive – you can do almost anything with Python
- The python community is the largest programming language community in the world, making it great for collaboration

SQL

SQL, otherwise known as Structured Query Language, is a programming language used to create, maintain, retrieve and access data within databases. It does this by using simple, easily understood queries.

Developed by IBM in the early 1970s, SQL is as relevant today as it's ever been. It works by understanding and analyzing large stacks of data. In fact, it can handle data pools of virtually any size and enables you to perform many functions at high efficiency and speed, including:

- Counting rows
- Aggregating functions
- Extreme value identification
- Slicing
- Sorting
- Filtering
- Groupings

For anyone using databases within the field of data science, knowing SQL is essential. It's a universal language that is transferable to other disciplines and languages and, as it's open-source, it has a large community of developers, making it an excellent tool for collaboration. SQL is also a highly in-demand skill, appearing in [42.7% of all job postings on Indeed.](#)

WHY YOU NEED THIS SKILL:

- Almost every organisation in tech uses SQL
- It's easy to use and open-source
- It's in demand
- It's been used for almost 50 years and continues to be a linchpin in data science
- It helps you to understand the workings of other languages such as Python and Java

PARTICULARLY GOOD IF YOU WANT TO UPSKILL IN:

Marketing, Product Management

R

A very powerful programming language, R is capable of handling huge amounts of data. A key tool for anyone working with statistics and big data, R has extensive libraries for plots and statistical ML (machine learning models).

One of the fastest growing languages in data science, R comes with impressive functionality, enabling data scientists to create custom- ised tools with specialised techniques. And while it's not as easy to get to grips with as Python, it's relatively user-friendly and certainly one of the first languages you need to know if you're planning on a career in data science.

WHY YOU NEED THIS SKILL:

- It's a very powerful tool capable of handling large amounts of data. Essential for statisticians
- It gives you the ability to customise tools for specialised use
- It's excellent for collaborating with others on open- source projects

Julia

Developed in 2009 at MIT, Julia is one of the fastest growing programming languages. It was designed to have syntax similar to that of Python and R, while bringing power and performance to rival C.

Julia is particularly useful for anyone looking to get into deep learning or data mining and is focused on performance and productivity, making it ideal for low-level programming as well as intensive projects. And while its community is relatively small, it's growing by the day.

WHY YOU NEED THIS SKILL:

- To stay ahead of the curve – it's one of the fastest growing data science languages, with downloads up 78% in the last four years
- It's extremely fast and flexible
- Data science increasingly revolves around deep learning machine learning, both of which Julia excels at

2.
Data visualisation

A big part of data science is creating a bridge between complex data and the people who make decisions based on this data. That's where data visualisation comes in. It takes large amounts of information and paints a picture in a way that's compelling and easy to understand, while looking for trends, patterns and anomalies, and removing any noise that might obscure the picture.

Data visualisation can take many forms, including charts, graphs and maps. Key data visualisation tools include software such as Tableau, D3, QlikView and Datawrapper, as well as plotting libraries for Python including Matplotlib and Seaborn.

When visualised in a clear, concise way, data has the power to persuade, change opinions, shape strategies and drive innovation. That's why the influence of data visualisation will [continue to grow](#) over the next few years.

WHY YOU NEED THIS SKILL:

- Helps make better decisions, faster
- Helps make sense of large amounts of complicated data
- Being able to convert complex data into easily understood visuals is a highly valued skill in an increasingly data-driven world

3.
Statistics

Along with programming, statistics is the cornerstone of data science. It's essentially the use of mathematical theories to form a technical analysis of data and unlock the secrets within the data. Statistics also help make predictions and find structures and forms within the data. Fundamental concepts within the field of statistics include:

- Distribution: mean, median, mode, variance and standard deviations
- Sample vs. population
- Central limit theorem
- Hypothesis testing
- Data generating process

A good knowledge of statistics is crucial for every step of the data analysis life cycle. From data discovery and formation, to preparation, model creation and measuring effectiveness, statistics plays a crucial role in the process.

WHY YOU NEED THIS SKILL:

- It's an underlying element of data science and fundamental to understanding the field
- It helps you understand how to collect data, analyse it and present the results
- It gives you base knowledge to go on and learn complex subjects such as machine learning

4.
Machine learning

While the concept of machine learning has been around for a while, within the last few years it has become hugely influential. The underlying driver of AI systems, machine learning essentially programs computers to program themselves, enabling them to learn, develop and adapt.

Machine learning enables organisations to understand and explore trends, behaviors and operational patterns, as well as develop future products and features. For anyone working in the field of data science, the following ML concepts are key:

- Generalization and (cross) validation
- Feature engineering
- Embeddings
- Regularization

All of the world's leading companies have adopted machine learning as a central part of their operational strategy. As AI systems and the technology around them continue to develop, machine learning will only become more integral to data science, as it becomes the most

accurate and efficient way to analyse large amounts of data.

WHY YOU NEED THIS SKILL:

- It's soon to become the most accurate way to analyse large amounts of complex data
- It's essential in order to understand how to make quality predictions and estimations
- With the emergence of AI and automation, it will only become more and more important for every company and organisation

5.
Deep learning

A subfield of machine learning but hugely important in its own right, deep learning is based on [artificial neural networks](#). It's inspired by the neural networks of the human brain and is tipped to be the tool that takes AI and machine learning to its full potential.

Deep learning algorithms are designed to mimic the way a human would analyse data, albeit on a far larger scale. These deep learning algorithms are created to continually analyse data within logical, multi-layered structures known as neural networks, which help create transferable solutions.

Unlike machine learning algorithms, deep learning algorithms can carry out feature engineering on their own, enabling them to learn and develop at a faster rate. And the list of [applications for deep learning models](#) is growing at a staggering rate. At the moment, virtual assistants, visual and speech recognition, fraud detection and personalization are key areas, with self-driving cars, 3D mapping, precision agriculture, robotics and automated healthcare tipped for mainstream adoption.

Deep learning models range from basic neural network models to advanced models like RNN or LSTM. Experience with Python libraries such as NumPy, Pandas, SyPy,TensorFlow, Keras and PyTorch are needed.

WHY YOU NEED THIS SKILL:

- It's a revolutionary tool that will shape the future of data science
- The scope for innovation is almost endless

6.
Big data

With the rapid advance of Internet of Things (IoT) technologies, we're generating more data than ever before and it's never been so valuable. Big Data is a term used to describe the way in which companies access and analyse massive amounts of information, in order to uncover patterns and correlations, gain valuable insights and make better decisions.

With rapid advances in technology, it's possible for companies to analyse massive amounts of data almost instantly. It also enables them to use predictive modelling through deep learning and machine learning. This saves time and money and provides a clear business strategy.

The key Big Data tools that you'll need to know are Apache Hadoop, Apache Spark, Cassandra, Xplenty and Adverity.

WHY YOU NEED THIS SKILL:

- Makes data analysis much more efficient
- Helps to harness huge amounts of data and use it to make better decisions
- Will become increasingly influential over the next few decades

7.
Experimentation & causal inference

An important part of data science is understanding the reason behind trends and patterns. This "why" is known as causal inference and helps to shape data-driven decisions and solve problems for leading organisations around the world.

Experimentation, in the form of randomised controlled trials (RCTs), is the standard method for causal inference. The most common and basic RCT is A/B testing, where two versions of a variable are compared to see which performs best.

There are times when RCTs or A/B testing is not possible or desirable for ethical or financial reasons. In this case, there are alternatives – advancements in econometrics have led to methods such as instrumental variables and propensity score matching, which can be used on observed data to make an empirical assessment.

Other types of causal inference include causal machine learning, inverse probability weighting and interrupted time series designs.

With an experimentation and causal inference skill set, data scientists can formulate and test causal hypotheses, design quasi-experiments, identify successful natural experiments and help shape business strategy, while driving future innovation.

WHY YOU NEED THIS SKILL:

- Helps to go deep and understand the reasons behind trends and patterns in data
- Equips you to gather data in a way that is creative and forward-thinking
- Data scientists with an in-depth knowledge of causal inference are now highly sought-after by top tech companies

8.
Communication: visualisation, storytelling, & business intelligence

An often-overlooked aspect of data science, good communication skills are an essential part of the field. Having a varied and in-depth technical skill set is great, but if you don't know how to communicate your findings, those skills will go to waste.

Data science poses questions and finds answers to these questions, extracting insights from data. These insights are only valuable if they can be properly communicated to stakeholders: key decision-makers who, more often than not, aren't from a data science background. This involves a subset of skills, including visualisation, storytelling and structured thinking.

We've touched on the importance of data visualisation (see skill #2), which helps make sense of large amounts of complicated data using illustrated forms. Storytelling is the ability to shape data insights into a compelling, concise and easy to understand narrative, one that creates an emotional connection with the audience.

It's also very important for people working in the field of data science to have a good grasp of business. Good decisions come from good collaborations, so being able to switch into a business mindset and discuss data in a business context is vital.

WHY YOU NEED THIS SKILL:

- Being able to convey results in a clear and accessible way can be the difference between good and bad decisions
- Having a well-rounded knowledge of the business world helps a data scientist be the bridge between the data and the strategy of the organisation
- Discussion, collaboration and the exchange of ideas cultivates innovation

9.
Curiosity

Finally, aside from all the knowledge and technical skill, a data scientist must have one core value that never wavers—the curiosity to find out more! To always ask questions and look beyond surface results, to uncover hidden patterns and find ingenious solutions. This inquisitive spirit is what drives organisations to break new ground.

So... are you ready to break new ground? Want to know how to gain all these skills and find your ideal job in data science?

We're here to help...

Your data science journey

We can't brush off the fact that gender disparity exists in the tech space, and we know it will take support from all sides to fix it.

This is why we created CodeOp, a tech school for women+ (inclusive of trans and nonbinary) who want to transition to tech or upskill their careers.

WHY CODEOP IS DIFFERENT:

- We provide the safe space, training, and resources needed to encourage, support, and equip women+ with the skills they need to thrive in the tech industry
- We keep class sizes small, guaranteeing a ratio of 1 instructor for every 5 students to provide 1-1 attention and faster learning
- Our students enjoy a transformative learning experience in an environment that fosters collaborative learning over competitive learning
- We provide education and preparation in translating past careers into technical careers, salary negotiation, and leadership

- Our students get access to role models and mentors who understand the journey and experience unique to underrepresented groups in tech.

WHY OUR DATA SCIENCE COURSE IS DIFFERENT:

- Our rigorous Data Science course is based on the strongest technical curricula in tech education
- Students learn from highly experienced data science experts from both world leading companies and academia
- We expose you to a hands-on, collaborative, experiential method of learning, with real-life application cases
- You'll have the choice to study online, or at our fantastic campuses in Barcelona and Malaysia
- Our support extends far beyond the classroom, with tailored career advice, workshops, and focused 1-1 coaching sessions to help you get where you want to go in your tech career

You've got the motivation, so why not get the skills? Download our Data Science Course Guide and get started on your data science journey!