



Data Leader's Guide to Upskilling

Skill up at scale. Data training designed for your business.

Join 2000+ companies and 80% of the Fortune 1000 who use DataCamp to upskill their teams.

Visit DataCamp

Google



ebay



PayPal

Uber

Deloitte.

CREDIT SUISSE



T-Mobile



1 The march to data literacy

Data transformation is at the heart of digital transformation

Over the past two decades, digital-first startups such as Uber, Amazon, Airbnb, and Stripe have disrupted vital industries such as transportation, commerce, travel, and banking. Organizations across all industries have recognized the need for digital transformation to compete in the new information economy. This is especially true of the COVID-19 economy, which has been accelerating the digitization of their processes and services ([PwC](#)). Despite massive digitalization investments, the painful truth is that approximately 70% of digital transformation initiatives fail to reach their stated goal ([McKinsey](#)).

While there may be many culprits for why digital transformation programs may fail, a key reason is not recognizing that having sustainable [organization-wide data skills](#) is a prerequisite for successful digital transformation. [Gartner](#) finds that fewer than 50% of documented corporate strategies mention data analytics as a critical lever for delivering enterprise-wide value.

“Leaders need to look at data first to succeed in their digital initiatives, rather than treating them as an afterthought to help with ad hoc projects.”

[Mike Rollings, Research Vice President at Gartner](#)

The underutilization of data science and business analytics dooms digital transformation initiatives from the start—to improve digital services and processes, you must have insights into the data they generate. Forrester estimates that an average of [60 to 73 percent of organization data is untouched](#) for analysis. The key differentiators between the disruptors and the incumbents is not technology-based but in their data-driven culture, the insights they draw from data while examining and iterating upon their services, and the data literacy skills they foster.

Data transformation requires moving beyond the center of excellence

Many organizations have tried to bridge the data gap by creating data science centers of excellence and investing in data tools. But merely laying out a data infrastructure and building a data science team isolates data science as a service center and underutilizes data science as a strategic pillar of the organization's success.

When the data science team must always react to the demands of other teams, they cannot be forward-thinking. Much of their time is spent on routine reporting rather than advancing the strategic gains enabled from becoming a data-driven organization. Other teams control their time since they are a support center to the rest of the organization. And once the data team fulfills a request, they face the challenge of communicating the results to outside stakeholders, as other groups may not speak the same language of data.

The hurdle of translating their findings and creating actionable insights is compounded by the data tools they use. Data scientists typically use tools such as SQL, Python, and R, while other teams may only be familiar with Excel. Conveying data results outside of data scientists' preferred toolkit is frustrating, time-consuming, and often done poorly. Moreover, not all members of the organization have the data literacy skills to interpret the insights drawn from data scientists and critically understand the business implications of their findings.

The business stakeholders requesting data from the centralized data science team aren't happy in this scenario either. There is often a tedious—and sometimes political—process in place for prioritizing data requests. This is because the data science organization drives the roadmap on how they provide value to the organization. Data science leaders may guide priorities to protect their teams from what they might view as egregious requests. And data scientists on the receiving end of a request may not understand the full context of what they're trying to solve, since they aren't as familiar with the same daily business needs as the person making the request.

In a data-driven organization, data science—and more broadly, data literacy—is an inclusive methodology for answering organizational questions where everyone is equipped to answer questions with data. For example, a marketing analyst would be able to use data to optimize their marketing spend, and a business analyst would visualize and describe data to prescribe actionable insights.

This frees up the data science team to spend its time working on projects with high return on investment that reinforces data science as a strategic pillar for the organization. The success of your digital transformation pivots on having the appropriate data skills across the organization.

The data literacy skill gap

To become data-driven, many organizations are recognizing that they need to address data literacy skill gaps within their organizations. A [McKinsey survey](#) of over a thousand businesses from various industries found that the most pressing skill gap to be addressed was data analytics—with 43% of respondents believing it to be the most urgent priority when it comes to upskilling. Similarly, [PwC's 2019 annual CEO survey](#) found that 34% of CEOs believe skill gaps in data analytics are the most crucial threat for their organization. This skill gap exists across all organization levels—an [Accenture study](#) found that executives are almost twice as likely as middle managers to value their "gut feeling" over data-driven insights.

Addressing data literacy skill gaps has become even more important, especially when considering the cost of not doing so. Organizations digitizing their processes and services expect their employees to work with data, despite not having the necessary skills to do their best work. This results in employees who aren't empowered to act on their data and a culture where, according to Accenture, about [50% of employees](#) avoid data-related tasks or find alternative methods to solve tasks without relying on data. All of this hurts the organization's decision making process and ability to iterate, and diminishes the opportunities for career growth across teams.

Organizations suffer when there is a clash between well-meaning digital transformation initiatives and a lack of necessary data skills to accommodate them. The same Accenture study found that 61% of employees believe that not having the required skills to extract insights from data has contributed to their workplace stress. When accounting for data-related procrastination and stress-related sick leave resulting from data and technology issues, organizations are losing around five working days per employee. This costs the US economy around \$100 billion yearly.

Upskilling is the only way to become data literate

Upskilling is the only way forward. Forward-thinking organizations are already pouring in investments to upskill their people to compete in the digital age. For example, [Marks & Spencer](#) created a retail data academy to upskill over 1,000 employees. Amazon launched a [Machine Learning University](#) to equip their engineers with the skills needed to deploy machine learning at scale in their products and services. Airbnb developed its own [Data University](#) to provide every level of the organization with the skills to make data-driven decisions. [AT&T](#) embarked on a \$1 billion, 10-year long project to upskill more than half of its 250,000 people workforce.

“This is our biggest digital investment in our people to date and the creation of the M&S Data Academy will upskill colleagues and provide them with an in-depth level of digital literacy as well as a Data Analytics qualification. Transformation of our business is key to survival and a huge part of this lies with our colleagues. We need to change their digital behaviors, mindsets, and our culture to make the business fit for the digital age.”

Steve Row, CEO of Marks and Spencer

These companies are ushering in a new era for upskilling in the digital age. A [McKinsey study](#) found that more than two-thirds of organizations have or plan on having an upskilling initiative to address skill gaps. More importantly, the same study found that 70% of organizations that invested in upskilling efforts are reporting positive business impacts that exceed the initial investment in upskilling. For example, 48% of organizations have reported moderate to significant positive effects on bottom-line growth due to upskilling—and 73% of organizations have reported moderate to substantial improvements in employee satisfaction. A [Deloitte survey](#) asking executives about the data maturity of their organizations found that 88% of organizations who have undergone organization-wide analytics upskilling have exceeded business goals.

The challenges in upskilling for data literacy

Data literacy is a methodology for answering business questions rather than a singular skill to be taught and learned, like traditional learning and development initiatives. Learning journeys will vary depending on the level of interaction different individuals may have with data. For example, a marketing analyst who regularly works with Excel may need to learn R or Python to succeed at their job, while a manager or leader may only need to know how to make educated decisions using data.

This is why a role-based, persona-driven learning journey is more effective at scaling data literacy training programs. Every persona has a different relationship with data and would need to acquire competencies in different tools, and grow different skills to thrive in the digital age.

Creating scalable and personalized learning paths for data literacy across the organization requires familiarity with the broad landscape of data tools, and a nuanced view of the different types of data personas found in data-driven organizations. In short, there's no one-size-fits all when it comes to data literacy.

A breakdown of data tools and data personas

A breakdown of crucial data tools

While programming languages are at the forefront of data literacy upskilling initiatives, it's also essential to consider the entire landscape of data tools. Just like data science—or more broadly, data literacy—can be considered a means to an end to solve business questions, data tools can be considered a means to an end to perform data-related tasks.

1 Programming Languages

Open-source programming languages have skyrocketed in popularity over the last two decades for data workflows. Apart from being free to use, open-source programming languages provide a plethora of tools and packages that allow practitioners to hone skills and perform data tasks across all data literacy competency areas. The most relevant open-source programming languages for data science are Python, R, and Scala.

- Python is an open-source programming language used for statistical and data analysis, big data processing, data engineering, and machine learning. It's considered one of the most popular programming languages for data work and is replacing legacy tools like SPSS and SAS.
- R is an open-source programming language most commonly used in research and development, statistical analysis, data analysis, and dashboard creation.
- Scala is an open-source programming language—it's especially used for maintaining and processing big data and big data applications.

2 SQL

SQL is a structured query language that allows data professionals to query, access, and manipulate data inside of database management systems. Just like one human language can have many different dialects, there are many different dialects of SQL (i.e., PostgreSQL, Oracle SQL, SQL Server) that are used by different organizations. The differences are generally minor as they [share many commonalities](#) in syntax and features.

3 Business Intelligence Tools (BI Tools)

BI tools have gained momentum over the past decade. Business intelligence tools are essentially supercharged spreadsheet tools made for the digital age. They allow for the organization, aggregation, and visualization of data in easy to use point-and-click dashboards, with no coding required. While there are many business intelligence tools, commonly used tools include the following:

- Tableau offers a robust, flexible, and intuitive interface to connect to raw data and create beautiful and interactive visualizations that allow teams to get an overview of their data.
- Power BI offers seamless connectivity with various raw data sources and an easy point-and-click interface to visualize and process data. A key feature of Power BI is that one version of it comes with Office 365 for the Enterprise and has an interface that is slightly reminiscent of Excel.

4 Spreadsheets

Spreadsheets have always been the go-to tool for data practitioners. They allow for easy, intuitive, and drag-and-drop interfaces for manipulating, aggregating, and visualizing data. However, they fall short when processing large amounts of data and often produce bottlenecks when creating reproducible shareable analysis. The most popular spreadsheet tools are Microsoft Excel and Google Sheets.

5 Big data tools

As organizations started collecting more and more data, it became imperative to efficiently structure, organize, and store big data. Many solutions have emerged that accommodate manipulating big data and orchestrating big data workflows. Notable examples include:

- Spark is a framework that allows for large-scale data processing and manipulation. It can be used using Python, Scala, or R.
- Airflow is an open-source workflow management tool that allows you to schedule data pipelines to ensure consistency and reliability across data workflows.

6 Command line tools

Command line tools are used to systematize file handling, enable version control, work with cloud tools, execute data pipelines developed using other data tools—especially programming languages—easily and scalably. The two most notable command line tools are the following:

- Shell is a command line interface that allows running programs, automating tasks, and accessing file directories.
- Git is a version control tool that allows for easily tracking and experimenting with changes done to code repositories.

7 Cloud platforms

Cloud platforms provide tools for organizations and teams to store and process data, host applications, and deploy data pipelines all using remote computing resources hosted by cloud providers. They have become the de-facto solution for data infrastructure for many organizations as they are easier to maintain, scale, and more resilient. The most widely used cloud platform providers are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud.

The different relationships with data

While each organization and the data they produce is different, there are commonalities in the different relationships individuals have with data. A useful way of thinking about and scaling data-focused upskilling efforts is with data personas. Each data persona has a different relationship with data and requires different data literacy competencies to become empowered to do their best work. Organizations can then map different roles within the organization to that persona and create a curated, personalized learning experience depending on what they need to learn.

1 Data Consumers and Leaders

Data Consumers and Leaders often work in nontechnical roles, but they consume data insights and analytics to make data-driven decisions. They often need to have conversations with data professionals and should be able to distinguish when data can and cannot be used to answer business questions.

DATA SKILLS:

Beginner

- Understand what data scientists, machine learning scientists, and data engineers do.
- Know which questions can (and can't) be answered with data.
- Interpret the results of data projects, including calculations and visualizations.

Intermediate

- Draw common visualizations and extract simple descriptive statistics from data.

Advanced

- Have a strong grasp of the fundamentals of business intelligence and BI tools.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Spreadsheets: Google Sheets, Microsoft Excel
- Business Intelligence: Power BI, Tableau

EXAMPLE JOB TITLES:

Chief Marketing Officer, Human Resources Manager, Head of Sa

2 Business Analysts

Business Analysts are responsible for tying data insights to actionable results that increase profitability or efficiency. They have deep knowledge of the business domain and often use SQL alongside non-coding tools to communicate insights derived from data.

DATA SKILLS:

Beginner

- Draw common visualizations and extract simple descriptive statistics from data.
- Understand the business applications of data.

Intermediate

- A deep knowledge of the business domain and the ability to report and communicate insights using data.

Advanced

- Democratize access to data insights by creating dashboards and organizing data to answer organizational questions.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Spreadsheets: Google Sheets, Microsoft Excel
- Business Intelligence: Power BI, Tableau
- SQL: PostgreSQL, SQL Server, Oracle SQL

EXAMPLE JOB TITLES:

Business Analyst, Supply Chain Analyst, Operations Analyst, Financial Analyst

3 Data Analysts

Similar to Business Analysts, Data Analysts are responsible for analyzing data and reporting insights from their analysis. They have a deep understanding of the data analysis workflow and report their insights through a combination of coding and non-coding tools.

DATA SKILLS:

Beginner

- Can draw common visualizations and extract simple descriptive statistics from data.
- Understands the business applications of data.

Intermediate

- A deep understanding of the data analysis workflow, which includes importing, manipulating, cleaning, calculating, and reporting on organization data.
- A strong grasp of business intelligence and BI tools.

Advanced

- Democratize access to data insights by creating dashboards and organizing data to answer organizational questions.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Programming languages: Python, R
- Spreadsheets: Google Sheets, Microsoft Excel
- Business Intelligence: Power BI, Tableau
- SQL: PostgreSQL, SQL Server, Oracle SQL

EXAMPLE JOB TITLES:

Data Analyst, Business Analyst, Supply Chain Analyst, Operations Analyst, Financial Ana

4 Data Scientists

Data Scientists investigate, extract, and report meaningful insights in the organization's data. They communicate these insights to nontechnical stakeholders and have a good understanding of machine learning workflows and how to tie them back to business applications. They work almost exclusively with coding tools, conduct analysis, and often work with big data tools.

DATA SKILLS:

Beginner

- Deeply understand the data analysis workflow, which includes importing, manipulating, cleaning, calculating, and reporting on organization data.

Intermediate

- Understand fundamental statistics, including distributions, modeling, and inference.
- Understand supervised and unsupervised machine learning workflows.
- Can create dashboards using coding tools such as Python and R.

Advanced

- Can apply analyses and machine learning workflows to business applications such as finance, marketing, and healthcare.
- Work with non-standard data types, such as time series, text, geospatial, and images.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Programming languages: Python, R, Scala
- SQL: PostgreSQL, SQL Server, Oracle SQL
- Big data tools: Airflow, Spark

EXAMPLE JOB TITLES:

Data Scientist, Data Analyst, can include a “citizen data scientist” (i.e., someone who performs the tasks of a data scientist, but does not have the title “Data Scientist”).

5 Machine Learning Scientists

Machine Learning Scientists are responsible for developing machine learning systems at scale. They derive predictions from data using machine learning models to solve problems like predicting churn and customer lifetime value, and are responsible for deploying these models for the organization to use.

DATA SKILLS:

Beginner

- Deeply understand the data analysis workflow, which includes importing, manipulating, cleaning, calculating, and reporting on organization data.

Intermediate

- Perform supervised and unsupervised machine learning workflows including feature engineering, training models, testing accuracy and making predictions.
 - Can apply analyses and machine learning workflows to business applications such as finance, marketing, and healthcare.

Advanced

- Perform deep learning workflows.
- Work with APIs and coding best practices.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Programming languages: Python, R, Scala
- SQL: PostgreSQL, SQL Server, Oracle SQL
- Big data tools: Airflow, Spark
- Command line tools: Git, Shell

EXAMPLE JOB TITLES:

Data Scientist, Research Scientist, Machine Learning Scientist, Machine Learning Engineer

6 Statisticians

Similar to Data Scientists, Statisticians work on highly rigorous analysis, which involves designing and maintaining experiments such as A/B tests and hypothesis testing. They focus on quantifying uncertainty and presenting findings that require exceptional degrees of rigor, like in finance or healthcare.

DATA SKILLS:

Beginner

- Deeply understand the data analysis workflow, which includes importing, manipulating, cleaning, calculating, and reporting on organization data.

Intermediate

- Perform statistical modeling workflows, including feature engineering, training models, testing goodness of fit, and inferring significance.
- Test hypotheses and design simple experiments such as A/B tests.

Advanced

- Design more complex experiments and understand Bayesian statistics.
- Understand specialist models, such as survival models, generalized additive models, mixture models, and structural equation models.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Programming languages: Python, R
- SQL: PostgreSQL, SQL Server, Oracle SQL

EXAMPLE JOB TITLES:

Quantitative Analyst, Inference Data Scientist, Data Scientist

7 Programmers

Programmers are highly technical individuals that work on data teams and work on automating repetitive tasks when accessing and working with an organization's data. They bridge the gap between traditional software engineering and data science and have a thorough understanding of deploying and sharing code at scale.

DATA SKILLS:

Beginner

- Write functions to avoid repetitive code.
- Benchmark and optimize code to improve performance.

Intermediate

- Deeply understand coding best practices.
- Work with web APIs and develop packages for sharing code.

Advanced

- Develop data pipelines and work with parallel programming.
- Understand programming paradigms, such as functional programming and object-oriented programming.

COMMONLY USED TECHNOLOGY AND TOOLS:

- Programming languages: Python, R, Scala
- Command line tools: Git, Shell

EXAMPLE JOB TITLES:

Software Engineer, Data Scientist, Dev-Ops Engineer

8 Data Engineers

Data Engineers are responsible for getting the right data in the hands of the right people. They create and maintain the infrastructure and data pipelines that take terabytes of raw data coming from different sources into one centralized location with clean, relevant data for the organization.

DATA SKILLS:

Beginner

- Efficiently extract, transform, and load data from raw data sources into organization databases.

Intermediate

- Process data and automate data flows using the command line.
- Process data using cloud platforms.

Advanced

- Manage and optimize databases and process big datasets.

COMMONLY USED TECHNOLOGY AND TOOLS:

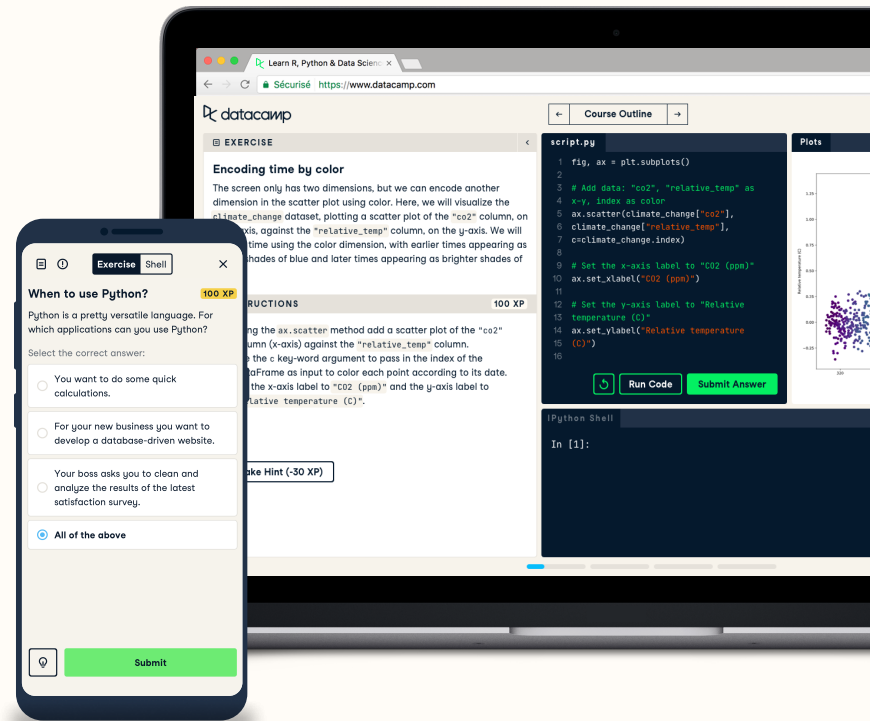
- Programming languages: Python, R, Scala
- SQL: PostgreSQL, SQL Server, Oracle SQL
- Command line tools: Git, Shell
- Big data tools: Airflow, Spark
- Cloud Platforms (e.g., Amazon Web Services)

EXAMPLE JOB TITLES:

Software Engineer, Data Engineer, Dev-Ops Engineer

3 How DataCamp supports organizations on their path to data literacy

DataCamp's proven learning methodology provides a cyclical process for learning and retention. This learning methodology enables learners across the data literacy spectrum to assess their skills and identify gaps, develop a learning plan based on these gaps, practice skills, and apply them in a real-world setting. Experienced data scientists can upskill on new techniques in their target domain, and domain experts can learn the fundamentals of data literacy and data science.



Assess

Effective learning starts with understanding skill gaps and strengths. With [DataCamp Signal™](#), learners can understand specific skill gaps they have across various topics and tools. From data literacy assessments like understanding and interpreting data to programming and machine learning assessments in R, Python, and SQL, our 10-minute adaptive evaluations provide learners with personalized skill gaps and learning paths to address their skill gaps.

Learn

DataCamp's growing course library houses more than 350 expert-led, hands-on courses across various technologies and domains for all data skills and levels. Learners can hit the ground running with our learn-by-doing approach—our bite-sized videos and interactive coding exercises allow them to start working with their preferred tool and topic right in the browser.

Practice









The next step in DataCamp's proven learning methodology is to practice all the information retained in courses. Using practice mode, learners can practice what they've learned with short challenges to test critical concepts. With over 3,400 practice questions, learners can practice their skills across various technologies and topics. Our [mobile app](#) is the perfect way to practice and learn on the go.

Once skills have been assessed, cultivated through courses, and sharpened through practice, learners are ready to apply their skills in a project-based environment. With [DataCamp projects](#), learners can solve a variety of real-world R and Python data science projects.

Learners can opt for guided projects, where they can follow step-by-step tasks and receive helpful feedback as they apply their newfound skills. They can also opt for unguided projects, which are open-ended, offering a variety of possible solutions and a live-code-along video to follow how an expert data scientist would approach a solution.

Track skills with skill matrix

Track the data skills your team has today and map a path to the skills they need tomorrow. Using the Skill Matrix, admin users can easily filter to identify individuals with the skills you need to take on specific projects or teams with low use or data skills gaps. They can then create and assign custom tracks to help bridge these gaps and report on skill development.

NAME	TEAMS	ASSESSMENTS				
		Python Programming	Machine Learning Fundamental with Python 1	Data Manipulation with Python	Statistics Fundamental with Python	Importing & Cleaning with Python
 Alicia Puma alicia@datacamp.com	DATA ANALYSTS	49 ↗	50 ↘	61	80	112
 Dushane Daniel dushane@datacamp.com	DATA ANALYSTS	91	121	49	113	112
 Gauthier Drewitt gauthier@datacamp.com	DATA ANALYSTS	73	78	123	90	49
 Ray Cooper ray@datacamp.com	DATA ANALYSTS	101	115	49 ↗	📄	📄
 Mar Rueda mar@datacamp.com	DATA ANALYSTS	82 ↘	📄	91	129	📄
 Sergio Pliego sergio@datacamp.com	DATA ANALYSTS	88	125	91	101	📄
 Caspar Sawrey caspar@datacamp.com	DATA ANALYSTS	152	112	112	148	167
 Xu Xuefeng xu@datacamp.com	DATA ANALYSTS	158	152	163 ↗	172 ↗	182

A robust training experience

We work with the largest brands in the world, including PayPal, Uber, HSBC, and Google, to help them transform their data skills. Our experienced team provides you with resources and guidance on everything from adoption best practices to SSO and LMS integrations—giving you the tools you need to upskill your team with confidence. Join over 7 million learners around the world. Close your talent gap. Visit datacamp.com.

Learn how you can bridge your team's data literacy gap and become more data-driven.

Visit DataCamp

