Data science 101

More employees need to think like data scientists

Companies know they need to embrace the power of data. Research has shown insights-driven businesses are growing more quickly, are more likely to acquire customers and retain them, and be more profitable. But at the same time, only a small percentage of companies provide data access to the majority of their employees.

Becoming a data-driven business is overdue. And that means getting every employee involved. Data scientists will continue to drive analytics, but employees need to act like citizen data scientists. Data is the key.

Taming the data explosion

Businesses know they need to turn trusted data into real business insights and value but are unsure how to move beyond proof-of-concept to deliver actual results.

Analytics-driven organizations build information infrastructures and tools that enable employees to answer the questions that create the most business value. Successful infrastructure includes considerations around people, process, technology, data, and culture. These different elements need to be open,

to quickly absorb new technologies and thought leadership to meet competitive challenges and opportunities.

There is no one-size-fits-all solution. Technology, data, and people will differ for every company. It can depend on a company's existing technology stack, appetite for intelligent decision-making, innovative people culture, data governance and control of data, structured processes and workflows, data-driven approaches to business challenges, and team collaboration with educational community building.

Employees need to think like a data scientist

Not every employee will be a data scientist but thinking around data should not be limited to one group. Industry disruptors are using this cross-functional access very successfully by demonstrating that improved access to data can result in citizen data scientists.

A successful analytics and data-science roadmap needs clear goals that embed insights into every-day decision-making and initiatives. This is a culture shift to a more data-driven organization. This is the only way data science can be a game-changer.

Four key pillars of data science

Business decision-making:
Deep sector knowledge to determine aspirations, knowledge share, success metrics, and long-term success

Technical skills:
Understanding technologies such as hosting (AWS, Azure), machine learning (Python, TensorFlow), model development (DS platforms, Jupyter and Zeppelin notebooks), and others. Managing unstructured data, visualization, engineering, and architectures

Soft and social skills:
Combines teamwork, business acumen, creativity, intuition, training, mentors, and the ability to go beyond

Education (academic and citizen data science):
Team members with advanced degrees, such as an MS, MBA, or PhD with many years of experience, are very helpful, as are SMEs that acquired an in-depth business knowledge and carry out data-science functions.
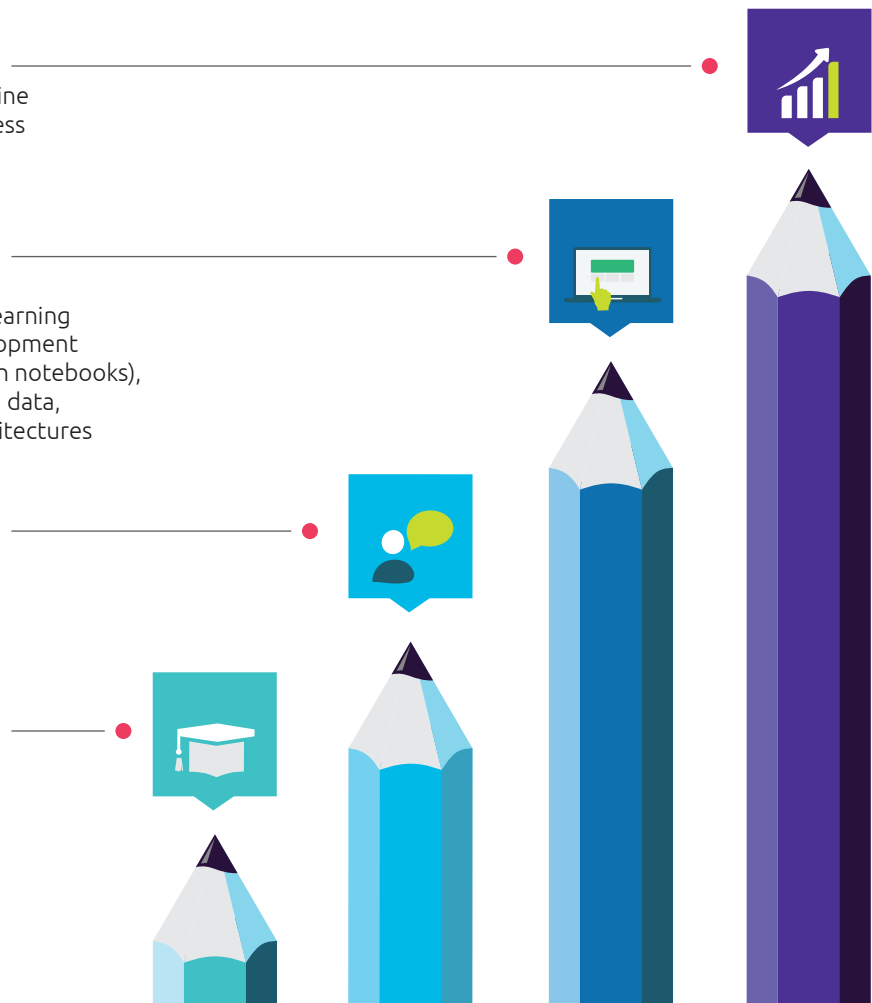
Data science will deliver on the promise of new digital technologies. Artificial intelligence and machine learning rely on data for success. By planning a system that delivers trusted data pipelines, companies will tap into the power of AI and ML to automate other applications and processes.

How data science will drive growth

- A data-driven approach will change your business: build models that predict and optimize business outcomes to become more competitive

- Move from descriptive to predictive analytics: stop looking at dashboards. Put your data to better use so you can improve product quality, uncover new opportunities, and speed-up decision-making.

- Determining the right data-science approach for your company: a truly successful approach to data science is unique to every organization.

- Taking control of all your data: it is a hybrid subject and involves disparate disciplines. Blend people, process, technology, data, and culture, and determine the approach that helps people use data.

- Building strategic roadmaps to success: companies need a cohesive roadmap to align strategic goals supported by the right executives and integrated with the infrastructure.

The real value of analytics and data science comes when people realize that the insights generated actually will change a decision, a process, or behavior for the better. From data-derived insights to action and then business value is the key pathway to true competitive advantage.

Tying people, process, technology, data, and culture together with operations and data-science workflows (DSWF) and governance policies, Capgemini has developed a set of ready-made tools and blueprints to successfully execute on data science. Tailored to a company's DNA, Capgemini provides standardized data-science workflows, use-case prioritization matrices, maturity framework maps, structured data-science governance rules, technology roadmaps, technology stack alternatives, and specific data-science job-profiling charts to support the move to being a data-driven organization.

Uncover new growth and opportunities in the numbers. Make every employee think more like a data scientist, tame your data, and become a truly data-driven organization.

Glossary of data-science terms

- **Data scientist:** An individual who performs statistical analyses on large amounts of data to derive actionable insights in the form of patterns, trends, graphs, and other information. Data is usually stored in data warehouses and databases to help solve a variety of business challenges for better decision-making utilizing data science, analytics, machine learning, or AI.

- **Unstructured data:** Data and information that does not have a pre-defined form or is not structured in any pre-defined manner. Usually, unstructured data does not have a recognizable structure. Rather, it is unorganized, raw, or (non) textual. As such, unstructured data is of a text, video, graphics, or audio-visual nature.

- **Data lake:** A massive, easily accessible, centralized repository of large volumes of both structured and unstructured data either on-premises or in a cloud architecture environment. Such a system usually stores data as object blobs or files.

- **Data platform:** A system that allows a standardized method of collecting harmonized data. The goal is to allow users to access the data with standard definitions and values across multiple assets and data sources in a repeatable way.

- **Open source:** A philosophy that allows free access and distribution of a software product for which the original source code is made freely available and can be re-distributed and modified. Source code open to the public fosters a diverse technical community as well as creating new communication paths and innovative models.

- **Infrastructure:** Usually referring to information technology, infrastructure is a composite term for software, hardware, and network resources as well as services needed to maintain, operate, and manage an enterprise IT environment. IT allows the delivery of solutions and services to customers, employees, and partners, and is usually a specific internal organization within a company.

- **Architecture:** Enables the logical and physical applications and data components as well as the architecture vision for a company. Candidate architecture roadmaps are improvement requests or concerns from stakeholders.

- **DSWF:** data-science workflow. A business enabler covering the entire process from posing a business challenge to the final insight sharing. Encompasses the full spectrum of people, process, technology, and data as it involves applying data, modelling and scoring, and generating actionable insights and overcoming the barriers encountered along the way.

About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of over 200,000 team members in more than 40 countries. The Group reported 2018 global revenues of EUR 13.2 billion (about $15.6 billion USD at 2018 average rate).

Learn more about us at
www.capgemini.com

People matter, results count.