Data Science

What is data science?

Data science is a multidisciplinary approach to extracting actionable insights from the large and ever-increasing volumes of data collected and created by today’s organizations. Data science encompasses preparing data for analysis and processing, performing advanced data analysis, and presenting the results to reveal patterns and enable stakeholders to draw informed conclusions.

Data preparation can involve cleansing, aggregating, and manipulating it to be ready for specific types of processing. Analysis requires the development and use of algorithms, analytics and AI models. It’s driven by software that combs through data to find patterns within to transform these patterns into predictions that support business decision-making. The accuracy of these predictions must be validated through scientifically designed tests and experiments. And the results should be shared through the skillful use of data visualization tools that make it possible for anyone to see the patterns and understand trends.

As a result, data scientists (as data science practitioners are called) require computer science and pure science skills beyond those of a typical data analyst. A data scientist must be able to do the following:

* Apply mathematics, statistics, and the scientific method
* Use a wide range of tools and techniques for evaluating and preparing data—everything from SQL to data mining to data integration methods
* Extract insights from data using predictive analytics and [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence), including [machine learning](https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks) and [deep learning](https://www.ibm.com/cloud/learn/deep-learning) models
* Write applications that automate data processing and calculations
* Tell—and illustrate—stories that clearly convey the meaning of results to decision-makers and stakeholders at every level of technical knowledge and understanding
* Explain how these results can be used to solve business problems

This combination of skills is rare, and it’s no surprise that data scientists are currently in high demand. According to an [IBM survey (PDF, 3.9 MB)](https://www.ibm.com/downloads/cas/3RL3VXGA), the number of job openings in the field continues to grow at over 5% per year, with over 60,000 forecast for 2020.