The Different Data Science Roles

Before we dive into what skills you need to become a data scientist, you should be aware that there are different roles in data science. Oftentimes, a data science team will rely on different team members for different skill sets, or the skill set needed may depend on the type of company and part of the organization you work in.

Let's look at some broad categories of roles that are lumped under the umbrella term "Data Science."

Data Scientists

One definition of a data scientist is someone who knows more about programming that a statistician, and more statistics than a software engineer. A data scientist will be able to run with data science projects from end-to-end: they will store and clean large amounts of data, explore data sets to identify potential insights, build predictive models, and weave a story around the findings.

Data scientists fine-tune the statistical and mathematical models that are applied onto that data. This could involve applying theoretical knowledge of statistics and algorithms to find the best way to solve a data problem.

Data scientists are the bridge between programming and algorithmic thinking. A data scientist might use historical data to build a model that predicts the number of credit card defaults in the following month, and use their data engineering skills to implement a simulation of their model on some sample data.

Additionally, within the broad category of data scientists, you may encounter statisticians who focus on statistical approaches to data, and data managers who focus on running data science teams.

Data Analysts and Business Analysts

Data analysts sift through data and provide reports and visualizations to explain what the data can offer. When somebody helps people from across the company understand specific queries with charts, they are filling the data analyst (or business analyst) role. In some ways, you can think of them as junior data scientists, or the first step on the way to a data science job.

Business analysts are adjacent to data analysts, but are more concerned with the business implications of the data. Should the company invest more in project X or project Y? Business analysts will leverage the work of data science teams to visualize and communicate what insight can be gained from the data to answer those questions.

Data Engineers

Data engineers are software engineers who handle large amounts of data, and often lay the groundwork for data scientists to do their jobs effectively. They are responsible for managing database systems, scaling the data architecture to multiple servers, and writing complex queries to sift through the data. They might also clean up data sets, and implement complex requests from data scientists (e.g. they take the predictive model from the data scientist and implement it into production-ready code).

Data engineers, in addition to knowing a breadth of programming languages (e.g. Ruby or Python), will usually know some Hadoop-based technologies (e.g. MapReduce, Hive, and Pig) and database technologies (e.g. MySQL, Cassandra, and MongoDB).

Within the broad category of data engineers, you'll find data architects who focus on structuring the technology that manages data models, and database administrators who focus on managing data storage solutions.