Data science roles and tools

1. Data science roles and tools

00:00 - 00:05

In this lesson, you'll learn about the different data roles and the tools they use.

2. Roles in data science

00:05 - 00:21

You might be surprised to learn that there isn't a single job within data science. Generally, there's four jobs: Data Engineer, Data Analyst, Data Scientist, and Machine Learning Scientist. Let's explore each one.

3. Data engineer

00:21 - 00:40

Data engineers control the flow of data: they build custom data pipelines and storage systems. They design infrastructure so that data is not only collected, but easy to obtain and process. Within the data science workflow, they focus on the first stage: data collection and storage.

4. Data engineering tools

00:40 - 01:06

Data engineers are proficient in SQL, which they use to store and organize data. They also use one of the following programming languages like Java, Scala, or Python to process data. They use Shell on the command line to automate and run tasks. Finally, data engineers, now more than ever, need to be comfortable with cloud computing to ingest and store large amounts of data.

5. Data analyst

01:06 - 01:31

Data analysts describe the present via data. They do this by exploring the data and creating visualizations and dashboards. To do these tasks, they often have to clean data first. Analysts have less programming and stats experience than the other roles. Within the workflow, they focus on the middle two stages: data preparation and exploration and visualization.

6. Data analyst tools

01:31 - 02:11

Data analysts use SQL, the same language used by data engineers, to query data. While data engineers build and configure SQL storage solutions, analysts use existing databases to retrieve and aggregate data relevant to their analysis. Data analysts use spreadsheets to perform simple analyses on small quantities of data. Analysts also use Business Intelligence, or BI Tools, such as

Tableau, Power BI, or Looker, to create dashboards and share their analyses. More advanced data analysts may be comfortable with Python or R for cleaning and analyzing data.

7. Data scientist

02:11 - 02:35

Data Scientists have a strong background in statistics, enabling them to find new insights from data, rather than solely describing data. They also use traditional machine learning for prediction and forecasting. Within the workflow, they focus on the last three stages: data preparation and exploration and visualization, and experimentation and prediction.

8. Data scientist tools

02:35 - 02:56

Similar to analysts, data scientists have strong skills in SQL. Data scientists must be proficient in at least Python on R. Within these languages, they use popular data science libraries, such as pandas or tidyverse. These libraries contain reusable code for common data science tasks.

9. Machine learning scientist

02:56 - 03:30

Machine learning scientists are similar to data scientists, but with a machine learning specialization. Machine learning is perhaps the buzziest part of Data Science; it's used to extrapolate what's likely to be true from what we already know. These scientists use training data to classify larger, unrulier data, whether its to classify images that contain a car, or create a chatbot. They go beyond traditional machine learning with deep learning. Within the workflow, they do the last three stages with a strong focus on prediction.

10. Machine learning tools

03:30 - 03:44

Machine learning scientists use either Python or R to create their predictive models. Within these languages, they use popular machine learning libraries, such as TensorFlow, to run powerful deep learning algorithms.

11. Review: members of your team

03:44 - 04:10

Here's a slide to recap what we've covered. It may be intimidating to see all these tools and languages, but they aren't as difficult to learn as spoken languages. If you know English, it may take you years to learn French. Programming languages are more similar to power tools. If you know how to use a power drill, you don't necessarily know how to use an electric saw, but you can learn with a little training!

12. Let's practice!

04:10 - 04:15

Alright, time for some exercises!