



# Data Science Principles

Data Science Principles makes the fundamental topics in data science approachable and relevant by using real-world examples and prompts learners to think critically about applying these new understandings to their own workplace. Get an overview of data science with a nearly code- and math-free introduction to prediction, causality, visualization, data wrangling, privacy, and ethics.

Modules	Case Studies	Takeaways	Key Exercises
<b>Module 1</b> Data 101	Flu Detection	<ul style="list-style-type: none"> <li>• Explain why data collection is important</li> <li>• Identify factors that may affect data quality</li> <li>• Recognize that not all data is numerical</li> <li>• Explain how the organization of data can affect the information you are able to extract from it</li> </ul>	<ul style="list-style-type: none"> <li>• List sources of data</li> <li>• Discuss what can be done with data</li> <li>• Categorize data by various factors</li> <li>• Determine whether data is high-quality or not</li> </ul>
<b>Module 2</b> Predictions and	Predicting Sepsis	<ul style="list-style-type: none"> <li>• Understand the basic structure of a predictive algorithm</li> <li>• Identify where human decisions shape predictive systems</li> <li>• Evaluate the success of a predictive system</li> </ul>	<ul style="list-style-type: none"> <li>• Examine how weather forecasts work</li> <li>• Use data to create a prediction</li> <li>• Sort types of training data</li> <li>• Simulate a predictive system</li> </ul>
<b>Module 3</b> Cause and Effect	The Google Tax	<ul style="list-style-type: none"> <li>• Explain why it is important to establish causal relationships</li> <li>• Identify barriers to establishing causal relationships in a variety of settings</li> <li>• Identify why randomization can help establish a causal relationship but also create other problems</li> </ul>	<ul style="list-style-type: none"> <li>• Classify relationships based on correlation or causation</li> <li>• Examine the relationship between variables</li> <li>• Identify potential common causes for correlated events</li> </ul>
<b>Module 4</b> Data Governance and Privacy	Privacy and Facial Recognition	<ul style="list-style-type: none"> <li>• Explain why data privacy is important</li> <li>• Describe what can constitute a violation of privacy</li> <li>• Critique existing privacy policies</li> <li>• Create a set of ethical tenets to guide data work at their own organizations</li> </ul>	<ul style="list-style-type: none"> <li>• Formulate data privacy guidelines</li> <li>• Discuss the risks of data re-identification</li> <li>• Evaluate existing data privacy policies for ethics</li> </ul>



Modules	Case Studies	Takeaways	Key Exercises
<b>Module 5</b> Beyond the Spreadsheet	Burning Glass and Text Data	<ul style="list-style-type: none"> <li>Identify sources of non-numerical data</li> <li>Explain why it would be useful to use non-numerical data</li> <li>Describe the differences in approach for supervised and unsupervised learning</li> <li>Identify use cases for neural networks</li> </ul>	<ul style="list-style-type: none"> <li>Perform a sentiment analysis</li> <li>Determine what types of data an algorithm cannot read</li> <li>Examine how computers intake visual and audio data</li> <li>Experiment with facial recognition</li> </ul>
<b>Module 6</b> Introduction to Algorithms	Reducing food waste with Shelf Engine	<ul style="list-style-type: none"> <li>Describe some algorithms commonly used in data science</li> <li>Understand basic workhorse algorithms in data science such as regression</li> <li>Explain why and how such tools are made substantially more complex</li> <li>Explain the crucial role humans have in overseeing and maintaining algorithms</li> <li>Explain some of the trade-offs between more sophisticated algorithms, including the costs of running and evaluating their success</li> </ul>	<ul style="list-style-type: none"> <li>Examine how to evaluate the performance of an algorithm</li> <li>Identify variables that can be used to predict consumer demand</li> <li>Select appropriate algorithms for different purposes</li> </ul>
<b>Module 7</b> Data Science Ecosystems	Harvard Link	<ul style="list-style-type: none"> <li>Explain the importance of data transformation and wrangling</li> <li>List the common technologies used within data science ecosystems</li> <li>Describe the connection between data science tasks, software tools, and hardware tools</li> <li>Identify potential sources of bottlenecks in the data science process</li> </ul>	<ul style="list-style-type: none"> <li>Identify and order the lifecycle of data</li> <li>Define what “the cloud” is</li> <li>Estimate the size of various data streams</li> </ul>
<b>Module 8</b> The Road Ahead	Healthcare Prioritization	<ul style="list-style-type: none"> <li>Recognize a problem that an algorithm might be able to solve</li> <li>Recognize the challenges created by using data science tools in ways outside their intended use</li> <li>Identify steps within the data science process that need auditing</li> </ul>	<ul style="list-style-type: none"> <li>Choose types of data to ingest into an algorithm</li> <li>Evaluate the risks of solely using an algorithm to make decisions</li> <li>Discuss how algorithms can reinforce biases</li> <li>Create a set of guidelines to evaluate projects</li> </ul>

Learning requirements: In order to earn a Certificate of Completion from Harvard Online and Harvard Business School Online, participants must thoughtfully complete all 8 modules, including associated quizzes, by stated deadlines.