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.

<table>
<thead>
<tr>
<th>Modules</th>
<th>Case Studies</th>
<th>Takeaways</th>
<th>Key Exercises</th>
</tr>
</thead>
</table>
| Module 1              | Data 101             | • Explain why data collection is important  
• Identify factors that may affect data quality  
• Recognize that not all data is numerical  
• Explain how the organization of data can affect the information you are able to extract from it | • List sources of data  
• Discuss what can be done with data  
• Categorize data by various factors  
• Determine whether data is high-quality or not |
| Module 2              | Predictions and      | • Understand the basic structure of a predictive algorithm  
• Identify where human decisions shape predictive systems  
• Evaluate the success of a predictive system | • Examine how weather forecasts work  
• Use data to create a prediction  
• Sort types of training data  
• Simulate a predictive system |
| Module 3              | Cause and Effect     | • Explain why it is important to establish causal relationships  
• Identify barriers to establishing causal relationships in a variety of settings  
• Identify why randomization can help establish a causal relationship but also create other problems | • Classify relationships based on correlation or causation  
• Examine the relationship between variables  
• Identify potential common causes for correlated events |
| Module 4              | Data Governance and  | • Explain why data privacy is important  
• Describe what can constitute a violation of privacy  
• Critique existing privacy policies  
• Create a set of ethical tenets to guide data work at their own organizations | • Formulate data privacy guidelines  
• Discuss the risks of data re-identification  
• Evaluate existing data privacy policies for ethics |
|                       | Privacy              |                                                                                                                                           |                                                                              |

© Copyright 2021. President and Fellows of Harvard College. All Rights
<table>
<thead>
<tr>
<th>Modules</th>
<th>Case Studies</th>
<th>Takeaways</th>
<th>Key Exercises</th>
</tr>
</thead>
</table>
| Module 5 | Burning Glass and Text Data | • Identify sources of non-numerical data  
• Explain why it would be useful to use non-numerical data  
• Describe the differences in approach for supervised and unsupervised learning  
• Identify use cases for neural networks | • Perform a sentiment analysis  
• Determine what types of data an algorithm cannot read  
• Examine how computers intake visual and audio data  
• Experiment with facial recognition |
| Module 6 | Reducing food waste with Shelf Engine | • Describe some algorithms commonly used in data science  
• Understand basic workhorse algorithms in data science such as regression  
• Explain why and how such tools are made substantially more complex  
• Explain the crucial role humans have in overseeing and maintaining algorithms  
• Explain some of the trade-offs between more sophisticated algorithms, including the costs of running and evaluating their success | • Examine how to evaluate the performance of an algorithm  
• Identify variables that can be used to predict consumer demand  
• Select appropriate algorithms for different purposes |
| Module 7 | Harvard Link | • Explain the importance of data transformation and wrangling  
• List the common technologies used within data science ecosystems  
• Describe the connection between data science tasks, software tools, and hardware tools  
• Identify potential sources of bottlenecks in the data science process | • Identify and order the lifecycle of data  
• Define what “the cloud” is  
• Estimate the size of various data streams |
| Module 8 | Healthcare Prioritization | • Recognize a problem that an algorithm might be able to solve  
• Recognize the challenges created by using data science tools in ways outside their intended use  
• Identify steps within the data science process that need auditing | • Choose types of data to ingest into an algorithm  
• Evaluate the risks of solely using an algorithm to make decisions  
• Discuss how algorithms can reinforce biases  
• Create a set of guidelines to evaluate projects |

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.