HARVARD Online

SHORT COURSE TRAINING

DATA SCIENCE PRINCIPLES



COURSE CODE: **EXE-DSP2025.02**

COURSE DURATION:

Four weeks from August 27 – September 27, 2025
5 – 10 hours per week

LECTURERS:

Dustin Tingley – Harvard University
Huynh Nhat Nam – Fulbright School of Public Policy and Management

Course Description

Data Science Principles is an introduction to data science course for beginners and managers to better understand what data science is, how to work with data scientists, and how to positively contribute to their company's data collection and analysis efforts. Foundational topics in data science are made approachable and relevant by using real-world examples that prompt you to think critically about applying these understandings to your workplace. This course provides a nearly code- and math-free introduction to prediction, causality, visualisation, data wrangling, privacy, and ethics.

The course will be delivered in English by Harvard lecturer and in Vietnamese by Fulbright lecturer.

Learning Outcomes

By the end of the course, participants will be able to:

- Understand the importance of data collection and the factors affecting data quality.
- Recognize different data types and how data organisation impacts information extraction.
- Understand the structure of predictive algorithms and how human decisions influence their design.
- Understand the importance of establishing causal relationships and the challenges involved in different contexts.
- Understand the importance of data privacy, recognise potential violations, and evaluate existing privacy policies.
- Be aware of ethical guidelines to ensure responsible data handling and decision-making.
- Understand the importance of data transformation and wrangling, along with common technologies in the data science ecosystem.
- Explore the relationship between data science tasks, software and hardware tools, and identify potential bottlenecks in the process.
- Identify problems that algorithms can solve and the challenges of using data science tools beyond their intended purpose.
- Recognize key steps in the data science process that require auditing to ensure accuracy and ethical use.

Prospective learners

The course is designed for early and mid-career professionals who want to understand how data is transforming industries and develop a data-driven mindset to advance their career opportunities, as well as for students and recent graduates looking to build a strong foundation in essential data-related concepts, vocabulary, skills, and business intuition to prepare for their careers.

Course requirements

Prior analytics experience would be ideal but not compulsory. The course is not intended to be mathematical intensive. Mathematical details will be kept to a minimum and, where required, will be presented only to help students better understand the analytics concepts and associated techniques that will be introduced in the course.

Participants who complete all 8 modules & instructor-led sessions, including satisfactory completion of the associated quizzes, by stated deadlines, can earn the Dual Certificates from Program, in which a Certificate of Participation from Fulbright School of Public Policy and Management, and a Certificate of Completion from Harvard Online.

Course tuition

25 million Vietnam Dong
(equivalent to US$980)

Discount 5%

will be applied for **early bird** registration before **15 July 2025**

For group registration

Please contact our Executive Education team for more information:
exedu.fsppm@fulbright.edu.vn
091.33.55.911

Application form: Download **Here**

Applicants send the application form to the email: Exedu.fsppm@fulbright.edu.vn no later than **August 15, 2025**.

This course is brought to you by the Fulbright School of Public Policy and Management.

Fulbright
SCHOOL OF PUBLIC POLICY AND MANAGEMENT

Course Schedule

Note: Open/Due date for each module/assignment is at 1.00am GMT+7 on the first and the last day of each module.

Date & time	Mode of delivery	Modules/Assignment	Content
Week 1 ***27 Aug - 03 Sept. 2025***	Online on Harvard Business School Platform	Self-paced study ▪ Module 1: Data 101 Case study: Flu Detection	▪ 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
		Self-paced study ▪ Module 2: Predictions & Recommendations Case Study: Predicting Sepsis	▪ Understand the basic structure of a predictive algorithm ▪ Identify where human decisions shape predictive systems ▪ Evaluate the success of a predictive system
Week 2 ***03 – 11 Sept. 2025***	Online on Harvard Business School Platform	Self-paced study ▪ Module 3: Cause & Effect Case Study: The Google Tax	▪ 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
		Self-paced study ▪ Module 4: Data Governance & Privacy Case Study: Privacy & Facial Recognition	▪ 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
13 Sept. 2025 8.30 – 12.00	On-Campus at Fulbright School of Public Policy and Management (FSPPM) or Online (at learner's selection)	Instructor-led sessions, FSPPM	▪ The data analytics framework CRISP-DM ▪ Exploratory Data Analysis - Data types - Describing and making inferences from a data set - Detecting and handling anomaly values ▪ Supervised learning - Supervised learning versus unsupervised learning - Underfitting and overfitting - Evaluating predictive models ▪ Privacy and ethics - Commercial interest vs individual privacy - Computational approaches to preserving privacy
Week 3 ***11 - 18 Sept. 2025***	Online on Harvard Business School Platform	Self-paced study ▪ Module 5: Beyond the Spreadsheet Case Study: 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
		Self-paced study ▪ Module 6: Introduction to Algorithms Case Study: 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
Week 4 ***18 - 25 Sept. 2025***	Online on Harvard Business School Platform	Self-paced study ▪ Module 7: Data Science Ecosystems Case Study: 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
		Self-paced study ▪ Module 8: The Road Ahead Case Study: 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
27 Sept. 2025 8.30 – 12.00	On-Campus at Fulbright School of Public Policy and Management (FSPPM) or Online (at learner's selection)	Instructor-led sessions, FSPPM	▪ Supervised learning models - Bias variance trade-off - Interpretability versus accuracy - Linear regression ▪ Data science ecosystem - The different skillsets required in a data science team - Common pitfalls