

**Data Science Overview**

**Course Number:** DATA-138  
**Duration:** 1 day

**Overview**

How does data science fit into high-value business analytics? What are the differences between Machine Learning (ML) and Artificial Intelligence (AI)? This live, online Executive Overview of Data Science training teaches participants how data science fits into the overall organizational landscape, demystifies data science buzzwords, compares data science programming languages like R and Python, and more. Participants are given a thorough overview of data science concepts and complete hands-on exercises with their instructor.

**Prerequisites**

No prior experience is presumed.

**Materials**

All Data Science Overview training students receive comprehensive courseware.

**Software Needed on Each Student PC**

Detailed setup will be provided upon request.

**Objectives**

* Understand how data science fits into the existing landscape of traditional organizations
* Place the phrase ‘data science’ in the broader context of implementing high-value analytics
* Describe the changing data environment that has motivated this shift
* Understand the definitions and intuition of key elements of data science such as machine learning and distributed computing
* Differentiate machine learning from deep learning/AI techniques
* Contrast the differences and similarities of open-source analytic solutions like R and Python with commercial software such as SAS and SPSS
* Identify the different roles and related skillsets required to implement high-value data science workflows from a team management perspective

**Outline**

* Introduction
* Using Data to Solve Business Challenges
  + New data sources and new demands on data insight
  + The democratization of data science tools
  + What changed in the past 10 years; why ‘data science’?
  + Coming up with definitions for data science: operational and conceptual
  + How does data science differ from ‘traditional’  reporting?
* Implementing High-value Data Science in the Organization
  + Is big data the right data?
  + Building the right data infrastructure
  + Data versus insights, interesting reports versus high-value products
  + Defining value in data science products
  + The cost of low-value data science
  + The typical data science team
  + Integrating human-centered design principles to increase the value of these products
* Understanding Explanatory Models
  + P-values and hypothesis testing
  + Correlation versus causation, observational versus experimental data
  + Multivariable modeling approaches to explain the relationship between inputs and outputs
  + Assumptions for causal inference and associated interpretation
  + Bayesian modeling: turning the traditional paradigm around
* Developing Predictive Modeling with Machine Learning
  + Clustering versus Supervised models
  + Classification versus Regression
  + Regression example in-depth with example code
  + Validation strategies for avoiding overfitting, understanding model capacity
  + Different families of algorithms: high-level overview
  + Classification example in-depth with example code
  + Understanding accuracy: what do these measures mean?
  + Clustering in-depth: use cases and explaining output
  + Clustering on treatment effects: does the exposure cause a different reaction in different people?
* Deep Learning and AI
  + What is a neural network? How is it different from other ML?
  + Artificial feed-forward neural networks and applications
  + Neural networks for time series data (recurrent neural networks and convolutional neural networks)
  + Neural networks for natural language processing
  + Predictive modeling for image classification
* Building and Maintaining a Highly Effective Data Science Team
  + Traits of high performing (and low performing) organizational analytic cultures
  + What cultural shifts are required for your department?
  + Roles on the data science team:
    - Data architects and engineers (organize, move, and store data)
    - Data managers (extract and transform data for use)
    - Analysts/statisticians (answer questions using data for insight)
    - Topical experts (subject matter experts)
  + Identify roles/skillsets for each of these workflows
  + Combining these skills and roles into a single team
  + Training trajectories for core members of these teams (who needs what)
  + Hiring strategies to build successful data science teams
  + Developing training opportunities for staff doing work in data science
  + Hardware/software infrastructure required
* Conclusion