

**Machine Learning: Data Foundation**

**Course Overview:**

The **Machine Learning: Data Foundation** course explores how data can be gathered and used to solve problems, define strategy and uncover hidden needs. The session provides an introduction to statistics and data analytics in an effort to drive all developers to be more data-minded. Instructors will refer to a Salesforce Infrastructure data example throughout the session to illustrate concepts and facilitate discussion. This course is a pre-requisite for the **Machine Learning: Algorithms & Applications** course.

**Productivity Objectives:**

**After this course, you will be able to:**

* Describe the role of Machine Learning in your organization and where it fits into Information Technology strategies
* Explain the technical and business drivers that result from using Machine Learning
* Classify data into Supervised and Unsupervised categories
* Recognize the characteristics an organization needs to be data-driven
* Identify visualization strategies for understanding data sets
* Discuss how Machine Learning could lead to inaccurate assumptions

**Course Duration:** This course will be delivered in 2 days.

**Course Outline:**

* Course Introduction
* Data and Data Processing
  + Data
    - Structured
    - Unstructured
  + Metadata
* Data sources
  + RDBMS
  + NoSQL
  + Document store
  + Spreadsheets
  + Streams
  + Events
* Data Driven
  + Basic statistics, distribution, skew
  + Correlation/causation
  + Hypothesis Testing
* Visualization
  + Analytic Graphics
  + Presentation Graphics
* Data Science
* Data Directed
  + Machine Learning
    - Supervised
    - Unsupervised
    - Algorithm tour
      * Classification
      * Regression
      * Clustering
      * Dimensionality Reduction