



## About Data Science Dojo

Data Science Dojo has trained more than 3200 professionals from over 600+ companies around the world through a variety of workshops and bootcamps. Unlike other offerings in the industry, our trainings are short, hands-on, and personal. We enable you to solve real-world business problems in the shortest duration possible. Visit our website, [www.datasciencedojo.com](http://www.datasciencedojo.com), to learn more.

## Prerequisites

You should have interest in data science and data engineering and knowledge of at least one programming language. However, many of our attendees come to us with little to no programming experience. Our pre-bootcamp materials will get you where you need to be to hit the ground running.

## Format

- 50% Lectures + 50% Labs, Exercises, and Demos
- 10 hours of pre-bootcamp coursework
- 50 hours of in-class training
- 10 hours of (optional) post-bootcamp coursework
- Mentored Kaggle project participation

## Duration\*

5 days of in-person training including an IoT project

\* Trainings in California are 4 days only and do not include an IoT project.

## Course Outline

### Preparatory Material (via webinars)

- Introduction to Big Data, Data Science, and Predictive Analytics
- Introduction to Azure ML Studio
- Fundamentals of Data Mining
- Introduction to R Programming
- Introduction to Amazon Machine Learning

### Fundamentals of Data Science

- Data Exploration, Visualization, and Feature Engineering
- Hands-On Labs: Data Exploration, Visualization, and Feature Engineering
- Machine Learning Fundamentals

### Classification Algorithms

- Introduction to Predictive Modeling
- Decision Tree Learning
- Logistic Regression
- Naïve Bayes
- Hands-On Lab: Building a Classifier

### Regression Algorithms

- Linear Regression
- Regularized Regression Models
- Hands-On Lab: Building a Regression Model

### Unsupervised Learning

- K-Means Clustering
- Hands-On Lab: Using K-Means Clustering

### Recommender Systems

- Text Analytics
- Content-Based and Collaborative Filtering
- Evaluation of Recommendation Systems. DCG, nDCG
- Hands-On Lab

### Ensemble Methods

- Bootstrapping, Bagging, and Boosting
- AdaBoost
- Random Forests
- Hands-On Lab: Building a Random Forest Classifier

### Operationalizing Machine Learning Models

- Metrics and Methods for Evaluating Classification and Regression Models
- Tuning Machine Learning Algorithm Parameters
- Hands-On Lab: Building a Classification Model in Azure ML Studio
- Hands-On Lab: Deploying a Predictive Model as a Service

### Data Science Using Amazon Machine Learning

- Hands-On Lab: Loading and Visualizing Data
- Hands-On Lab: Building and Evaluating a Predictive Model
- Hands-On Lab: Deploying a Real-Time Prediction Endpoint

### Fundamentals of Big Data Engineering

- Introduction to Large-Scale Online Systems
- Hive Tutorial
- Hands-On Labs: Creating a Hadoop Cluster and Writing Hive Queries

### Handling Real-Time and Streaming Data\*

- Message Queues and Real-time Analytics
- Hands-On Lab: Creating a Streaming Analytics Pipeline

### Distributed Databases and Data Warehousing\*

- Hands-On Lab: Setting Up Relational Databases in the Cloud
- NoSQL Databases and HBase
- Hands-On Lab: Twitter and HBase

### Data Science Essentials\*

- Introduction to Online Experimentation and A/B Testing

## IoT Project\*

During the bootcamp, spend time building an 'Internet of Things' solution using Azure EventHubs, Azure Stream Analytics, Power BI and Azure ML studio. You will be using a smartphone to obtain real-time accelerometer data and observe various analytics in a cloud BI dashboard in real-time.

\*Not included in the California bootcamp curriculum.

