

# Data Science for Investment Professionals Certificate Syllabus

## Course 1: Data and Statistics Foundation for Investment Professionals

### Module 1: Measures of Central Tendency

- Lesson 1: Data Types
- Lesson 2: Frequency and Contingency Tables
- Lesson 3: Measures of Central Tendency - Arithmetic Mean
- Lesson 4: Measures of Central Tendency - Median and Mode
- Lesson 5: Measures of Central Tendency - Weighted, Geometric, and Harmonic Mean

### Module 2: Measures of Dispersion

- Lesson 1: Quantiles
- Lesson 2: Measures of Dispersion
- Lesson 3: Skew and Kurtosis

### Module 3: Introduction to Distributions

- Lesson 1: Uniform Distributions and Basic Probability
- Lesson 2: The Binomial Distribution
- Lesson 3: Decision Trees and Option Pricing
- Lesson 4: Normal and Lognormal Distributions
- Lesson 5: Monte Carlo Simulation

### Module 4: Data Visualization Techniques

- Lesson 1: Histograms and Bar Charts
- Lesson 2: Pie Charts, Donuts, and Tree Maps
- Lesson 3: Clouds, Plots, Maps, and Ways to Mislead
- Lesson 4: Storytelling and Narrative Techniques

### Module 5: Sampling Theory

- Lesson 1: Sampling and the Distribution of the Sample Means
- Lesson 2: Point Estimators, Confidence Intervals, and the T-Distribution
- Lesson 3: Good Estimators and Bias

### Module 6: Hypothesis Testing

- Lesson 1: Steps in Hypothesis Testing
- Lesson 2: Hypothesis Testing of Mean and Variance

- Lesson 3: Multiple Testing Problem & False Discovery
- Lesson 4: Curiosity and Creativity

## **Course 2: Statistics for Machine Learning for Investment Professionals**

### **Module 1: Data and Patterns**

- Lesson 1: Data and the Importance of Patterns
- Lesson 2: Modeling Relationships between Variables

### **Module 2: Randomness and Probability**

- Lesson 1: Random Processes and Probability Distributions
- Lesson 2: Probability Concepts

### **Module 3: Linear Regression**

- Lesson 1: From Patterns to Regressions
- Lesson 2: Linear Regression Assumptions
- Lesson 3: Linear Regression Using Dummy Variables
- Lesson 4: Linear Regression Assumption Violations

### **Module 4: Introduction to Advanced Regression Concepts**

- Lesson 1: Gradient Descent
- Lesson 2: Bias-Variance and Linear Regression Limitations
- Lesson 3: Bias-Variance Python Case Study
- Lesson 4: Bias and Variance In-Practice Examples
- Lesson 5: Logistic Regression
- Lesson 6: Logistic Regression Using Python
- Lesson 7: Logistic Regression Real-World Examples

### **Module 5: Introduction to Time Series Analysis**

- Lesson 1: Use and Value of Time-Series Models
- Lesson 2: Trend and Log-Trend Relationships
- Lesson 3: Autoregressive Models and Covariance Stationarity
- Lesson 4: Random Walks and Markets

## **Course 3: Machine Learning for Investment Professionals**

### **Module 1: Machine Learning**

- Lesson 1: What Is Machine Learning?
- Lesson 2: Overfitting and How to Address It

### **Module 2: Supervised Learning**

- Lesson 1: Penalized Regression
- Lesson 2: Support Vector Machine
- Lesson 3: K-Nearest Neighbor
- Lesson 4: Classification and Regression Tree
- Lesson 5: Ensemble Learning and Random Forest

### **Module 3: Unsupervised Learning**

- Lesson 1: Principal Components Analysis
- Lesson 2: Clustering

### **Module 4: Deep Learning**

- Lesson 1: Neural Networks
- Lesson 2: Deep Neural Networks and Reinforcement Learning
- Lesson 3: Choosing the Appropriate ML Algorithm

### **Module 5: The Translator**

- Lesson 1: T-Shaped Teams
- Lesson 2: Collaboration and Communication
- Lesson 3: Practitioners Discuss Skills Needed for Machine Learning Projects
- Lesson 4: Practitioners Offer Lessons Learned from Past Projects
- Lesson 5: Ethical Decision Making for Machine Learning

## **Course 4: Natural Language Processing for Investment Professionals**

### **Module 1: Cleaning and Wrangling Text Data**

- Lesson 1: Cleaning Text Data
- Lesson 2: Wrangling Text Data

### **Module 2: Exploratory Data Analysis, Feature Selection and Feature Engineering with Text Data**

- Lesson 1: Exploratory Data Analysis with Text Data
- Lesson 2: Feature Selection with Text Data
- Lesson 3: Feature Engineering with Text Data

### **Module 3: Selecting, Training, Evaluating, and Tuning a Natural Language Processing Model**

- Lesson 1: Selecting NLP Models and Methods
- Lesson 2: Training and Evaluating an NLP Model
- Lesson 3: Tuning an NLP Model

#### **Module 4: Developing an NLP Model for Predicting Sentiment of Financial Text**

- Lesson 1: Curating, Cleaning and Wrangling of Financial Text Data
- Lesson 2: EDA, Feature Selection, and Engineering with Financial Text Data
- Lesson 3: Training, Selecting, Evaluating, Tuning, and Testing a Sentiment Classifier Model

#### **Module 5: Applications of NLP in Investments**

- Lesson 1 Part 1: Applications of NLP-Based Sentiment Analysis
- Lesson 1 Part 2: Collaboration around NLP Projects
- Lesson 2 Part 1: Applications of NLP-Based Topic Classification
- Lesson 2 Part 2: Innovative Applications of NLP

#### **Course 5: Mitigating Biases in the Data Science Pipeline**

##### **Module 1: Investment Context, Some Ethical Dilemmas, Biases and Practical Issues**

- Lesson 1: What is Bias?
- Lesson 2: How should We think About Bias?
- Lesson 3: What should We do About Bias?
- Lesson 4: Trade-off between Interpretability and Accuracy

##### **Module 2: Biases and Mitigations in Study Design and Hypothesis Generation**

- Lesson 1: Define Success
- Lesson 2: Resources Inventory
- Lesson 3: How Should Good Training Samples Look?
- Lesson 4: Making Your Analysis Convincing

##### **Module 3: Biases and Mitigations in Data Collection and Exploration**

- Lesson 1: Data Collection Biases
- Lesson 2: Data Exploration Biases
- Lesson 3: Reliance on Published Research
- Lesson 4: Documenting Data Collection and Exploration
- Lesson 5: Mitigating Data Collection Biases
- Lesson 6: Mitigating Data Exploration Biases

##### **Module 4: Biases and Mitigations for Model Development, Testing, and Monitoring**

- Lesson 1: Identifying Biases in Variable Selection

- Lesson 2: Identifying Biases in Model Building
- Lesson 3: Mitigating Variable Biases
- Lesson 4: Mitigating Model Biases
- Lesson 5: Monitoring Models and Data

#### **Module 5: Biases and Mitigations in Model Interpretation, Communication, and Governance**

- Lesson 1: Biases in Context of Fairness, Transparency, and Accountability Principles
- Lesson 2: Fairness Considerations and Failures
- Lesson 3: Governance Framework Overview in Principle
- Lesson 4: Human Intervention, Model Interpretability and Explainability

#### **Module 6: Case Studies and Code Labs: Mitigating Biases in the Data Science Pipeline**

- Lesson 1: Identifying and Mitigating Biases in Consumer Lending Use Case (Classifying Good vs. Bad Credits) - Case Study, Instructional Videos and Code Lab
- Lesson 2: Identifying and Mitigating Biases in Investment Management Use Case (Equity Selection Models Using Machine Learning) - Case Study, Instructional Videos and Code Lab