

## Course Expectations

### Module 1: Introduction

- What Is Cluster Analysis?
- Cluster Analysis Use Cases
- The Challenge of Clustering Data

### Module 2 – Data Sets Used in the Course

- The Iris Data Set
- The Hand-Written Digits Data Set
- The Heart Data Set

### Module 3 – Types of Clusterings and Clusters

- Hierarchical, Partitional, and Overlapping Clustering
- Prototype Clusters
- Density-Based Clusters

### Module 4 – K-Means Clustering

- Introducing K-Means
- The K-Means Algorithm
- Euclidian Distance
- The Problem with Outliers
- Data Standardization
- K-Means Caveats
- Hands-On Lab #1

### Module 5 – Optimizing K-Means

- Evaluating Clusters
- Cluster Cohesion
- Evaluating Cohesion with the Elbow Method
- The Silhouette Coefficient
- Evaluating a Clustering Using the Silhouette Score
- Hands-On Lab #2

**Module 6 – DBSCAN Clustering**

- Introducing DBSCAN
- The DBSCAN Algorithm
- DBSCAN Caveats

**Module 7 – Optimizing DBSCAN**

- Considerations for Optimizing DBSCAN
- Calculating min\_samples
- Choosing the eps Value
- Introducing Nearest Neighbors
- Evaluating eps Using the Elbow Method
- K-Means Versus DBSCAN
- Hands-On Lab #3

**Module 8 – Dimensionality Reduction**

- Introducing Dimensionality Reduction
- Introducing Principal Component Analysis (PCA)
- PCA Concepts
- Hands-On Lab #4

**Module 9 – Categorical Data**

- The Problem with Categories
- One-Hot Encoding
- Factor Analysis of Mixed Data (FAMD)

**Module 10 – Additional Resources****Hands-On Lab #5**