**Title**: Data Science Lecture

**Duration:** Half Day (3 hours)

**Pre-requisite:** Basic programming knowledge will be helpful but not required

**Description:**

This lecture will give a broad overview of Data Science. We’ll clarify the relationship between Data Science and Machine Learning and explore the Data Science process. We’ll talk about identifying an effective data analysis question that is actionable, and how to get the right data to answer the question. We’ll discuss data exploration and the importance of clean data, complete data, and the quantity and variety of data. We’ll also cover how to effectively apply and evaluate Machine Learning models. The lecture will briefly demonstrate how to work through a Data Science project using Pandas and Scikit-learn, highlighting the variety of choices that need to be made throughout the process that determine its success.

**Learning Objectives:**

* Understand the relationship between Data Science and Machine Learning
* Become familiar with the Data Science process
* Identify effective data analysis questions that are actionable
* Identify effective data sources
* Understand the importance of clean, complete, and quantity of data
* Understand how Machine Learning is applied and evaluated within the Data Science process
* Become familiar with some of the tools used throughout the process

**Topic Outline:**

* Introduction to lecture
* Data Science vs. Machine Learning
* The Data Science process
* The importance of data
* Exploring and transforming data
* Creating and evaluating Machine Learning models
* Developing an effective Data Science strategy
* Demonstration of the Data Science process using pandas and scikit-learn
* Next Steps

**Structured Activity/Case Studies:**

Demonstration – the Data Science process using pandas and scikit-learn