Data Wrangling in Python

10 hours of instruction (Python)

Course Description

Python programming is among the most powerful and widely used tools by data scientists today.

This course will cover how to leverage Python's capabilities to manipulate and explore data. By the end of this program, attendees will be able to outline use cases for Python and automate tedious data processes, clean data for analysis, summarize and reshape data.

Objectives
1. Identify what data science is and why it's important
2. Program proficiently in Python
3. Automate data cleaning and processing

Software Requirements:
Python & Anaconda

- Anaconda and Jupyter Notebook

Prerequisites
Students must have learned the basic features of Python and have an understanding of these areas before attending class:

- Variables: defining, swapping, printing and deleting variables
- Main data types: numbers, logicals, strings
- Basic data structures: lists, sets, dictionaries, tuples.
- Control flow structures and modular code
- Conditional statements
- "For" loops and list comprehensions
- "While" loops and break/continue statements

Topics Covered
Day One:

- Programming across industries and core functions of data scientists
- Data science use cases for Python

- Functions in Python

Day Two:

- Introduction to NumPy
- Arrays: filtering and reshaping
- Working with Pandas

Day Three:

- Basic operations on series
- Dataframes and basic operations
- Loading data into Python using Pandas

Day Four:

- Summarizing data using Pandas
- Reshaping data using Pandas