

Data Wrangling with Python

Course Overview

In this course you will start with the absolute basics of Python, focusing mainly on data structures. Then you will delve into the fundamental tools of data wrangling like NumPy and Pandas libraries. You'll explore useful insights into why you should stay away from traditional ways of data cleaning, as done in other languages, and take advantage of the specialized pre-built routines in Python.

This combination of Python tips and tricks will also demonstrate how to use the same Python backend and extract/transform data from an array of sources including the Internet, large database vaults, and Excel financial tables. To help you prepare for more challenging scenarios, you'll cover how to handle missing or wrong data, and reformat it based on the requirements from the downstream analytics tool. The course will further help you grasp concepts through real-world examples and datasets.

Who Should Attend

Data Wrangling with Python takes a practical approach to equip beginners with the most essential data analysis tools in the shortest possible time. It contains multiple activities that use real-life business scenarios for you to practice and apply your new skills in a highly relevant context.

Course Objectives

By the end of this course, you will be confident in using a diverse array of sources to extract, clean, transform, and format your data efficiently.

Course Outline

1 INTRODUCTION TO DATA STRUCTURE USING PYTHON

Python for Data Wrangling
Lists, Sets, Strings, Tuples, and Dictionaries

2 ADVANCED OPERATIONS ON BUILT-IN DATA STRUCTURE

Advanced Data Structures
Basic File Operations in Python

This is a 3-day class



Data Wrangling with Python

3 INTRODUCTION TO NUMPY, PANDAS, AND MATPLOTLIB

NumPy Arrays
Pandas DataFrames
Statistics and Visualization with NumPy and Pandas
Using NumPy and Pandas to Calculate Basic Descriptive Statistics on the DataFrame

4 DEEP DIVE INTO DATA WRANGLING WITH PYTHON

Subsetting, Filtering, and Grouping
Detecting Outliers and Handling Missing Values
Concatenating, Merging, and Joining
Useful Methods of Pandas

5 GET COMFORTABLE WITH A DIFFERENT KIND OF DATA SOURCES

Reading Data from Different Text-Based (and Non-Text-Based) Sources
Introduction to BeautifulSoup4 and Web Page Parsing

6 LEARNING THE HIDDEN SECRETS OF DATA WRANGLING

Advanced List Comprehension and the zip Function
Data Formatting

7 ADVANCED WEB SCRAPING AND DATA GATHERING

Basics of Web Scraping and BeautifulSoup libraries
Reading Data from XML

8 RDBMS AND SQL

Refresher of RDBMS and SQL
Using an RDBMS (MySQL/PostgreSQL/SQLite)

9 APPLICATION IN REAL LIFE AND CONCLUSION OF COURSE

Applying Your Knowledge to a Real-life Data Wrangling Task
An Extension to Data Wrangling