

PY-DW



Data Wrangling with Python

DURATION	LEVEL	TECHNOLOGY	DELIVERY METHOD	TRAINING CREDITS
3 Days	Intermediate	Python	Instructor Led	NA

INTRODUCTION

Data is the new oil, but it comes crude. To do anything meaningful - modeling, visualization, machine learning, for predictive analysis – you first need to wrestle and wrangle with data. Data Wrangling with Python teaches you the essentials that will get you up and running with data wrangling in no time.

For data to be useful and meaningful, it must be curated and refined. Data Wrangling with Python teaches you the core ideas behind these processes and equips you with knowledge of the most popular tools and techniques in the domain.

AUDIENCE PROFILE

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.

PREREQUISITES

Basic computer literacy: Familiarity with using a computer and navigating operating systems (Windows, macOS, Ubuntu) is assumed. The course starts with the basics of Python, focusing on data structures. This suggests the course is suitable for beginners with no prior Python experience.

COURSE OBJECTIVES

- Understand the fundamentals of data wrangling and its importance in data analysis.
- Grasp core Python concepts, particularly data structures (lists, sets, strings, tuples, dictionaries).
- Learn and utilize popular data-wrangling libraries like NumPy and Pandas.
- Gain practical knowledge in data extraction, cleaning, transformation, and formatting from various sources (internet, databases, spreadsheets).
- Develop proficiency in handling missing or incorrect data and reformatting based on analytical needs.
- Apply data wrangling techniques through real-world examples and datasets.
- Gain confidence in using Python for efficient data manipulation across diverse data sources.

COURSE CONTENT

Module 1: Introduction to Data Structure using Python

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

Module 2: Advanced Operations on Built-In Data Structure

- Advanced Data Structures
- Basic File Operations in Python

Module 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

Module 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

Module 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

Module 6: Learning the Hidden Secrets of Data Wrangling

- Advanced List Comprehension and the zip Function
- Data Formatting

Module 7: Advanced Web Scraping and Data Gathering

- Basics of Web Scraping and BeautifulSoup libraries
- Reading Data from XML

Module 8: RDBMS and SQL

- Refresher of RDBMS and SQL

- Using an RDBMS (MySQL/PostgreSQL/SQLite)

ASSOCIATED CERTIFICATIONS & EXAM

None