EFFICIENT DATA PROCESSING USING PYTHON

COURSE DESCRIPTION

The *Efficient Data Processing training* is a hands-on course designed for individuals aiming to accelerate their data processing workflows. In general, data processing can be highly time-consuming due to the size and complexity of datasets, and typically requires significantly more time than the initial data collection. However, with efficient data processing, this time can be shortened, allowing for a quicker analysis and interpretation of the dataset.

This course demonstrates, through multiple case studies related to pharmaceutical (R&D) manufacturing processes, how data can be effectively cleaned and transformed using Python. Python modules such as numpy, pandas, scikit-learn and basic visualisation using matplotlib seaborn or plotly will be used. These case studies make the training particularly relevant for scientists and academics involved in R&D. Each case study is introduced with the necessary theoretical background, covering essential Python code and modules, and then applied in a practical context.

The course offers a variety of approaches to eliminate repetitive actions, optimize workflows, and visually analyze multiple experiments. By focusing on these methods, participants gain insights and tools to enhance efficiency in their data processing tasks.

COURSE PROGRAM (Day 1)

- Extracting the relevant data from log files
- Processing a large number of files in a loop
- Dataset merging: e.g., linking process information to the corresponding quality attributes
- Interactive visualization for repetitive experimental measurements
- Repetitive visualization for efficient outlier removal
- Threshold selections for efficient outlier removal

TARGET AUDIENCE

The target audience is professionals or students in chemical, pharmaceutical & healthcare sciences and engineering.

REQUIREMENTS

Basic understanding of the Python programming language is required.