

Data Analysis with Python

NumPy, pandas, Matplotlib and remote data sources

FA-0783 · 3 days

Data & Analytics · Software Development

Introduction

Use NumPy, pandas and Matplotlib to acquire, organise, extract and present data for reports. Build on existing basic Python knowledge through array operations, data-frame workflows and time-series visualisation.

Learning outcomes

- Create and index NumPy arrays and apply array operations.
- Work with pandas Series/DataFrames, missing data, grouping and joins.
- Import and export selected data and produce Matplotlib/pandas charts.
- Retrieve prepared remote datasets using a supported DataReader provider or Nasdaq Data Link interface.

Prerequisites

Basic Python programming knowledge. A compatible Python environment with the course libraries and access to prepared data. Remote providers may require an API key, suitable account or dataset entitlement; exercises use data made available for the course.

Training outline

1. Module 1 – NumPy foundation

- Introduction to NumPy

- NumPy Arrays
- Numpy Operations
- Numpy Indexing

2. Module 2 — pandas data workflows

- Introduction to Pandas
- Series
- DataFrames
- Missing Data
- Group By with Pandas
- Merging, Joining, and Concatenating DataFrames
- Pandas Common Operations
- Data Input and Output

3. Module 3 — Reporting and visualisation

- Introduction to Visualization in Python
- Matplotlib Basics
- Pandas Visualization Overview
- Pandas Time Series Visualization

4. Module 4 — Remote data acquisition

- Introduction to data sources.
- pandas-datareader as a separate package; select a currently maintained provider.
- Nasdaq Data Link, formerly Quandl, using its supported Python client/API.
- Review retrieved schemas, authentication and provider access requirements before using data in reports.