

# Applied Python for Data Science - TTPS4876

Gain advanced skills to handle complex data sets, understand machine learning algorithms, and translate data into actionable insights

**Duration:** 2 Days

**Skill Level:** Intermediate

**Available Format:** Instructor-Led Online; Instructor-Led, Onsite In Person ; Blended; On Public Schedule

Geared for experienced Python users with basic data science skills, Applied Python for Data Science is a comprehensive hands-on course that deep dives the advanced skills and tools used to perform exploratory data analysis, create complex visualizations, and perform large-scale distributed processing on Big Data.

## What You'll Learn

### Overview

This course provides students with Python data science skills that can immediately be applied in real life. The course focuses on Pandas as the primary tool, using related packages such as NumPy and Seaborn to enhance processing and visualization.

### Objectives

Working in a hands-on, applied learning environment, participants will learn to:

- **Advanced Data Ingestion & Preparation:** Efficiently import, clean, and export complex datasets using Pandas, preparing data for deeper analysis and reuse.
- **Sophisticated Data Selection & Indexing:** Confidently navigate and subset data using advanced indexing techniques, Boolean logic, and multi-indexing for hierarchical datasets.
- **Data Aggregation & Summarization:** Apply groupby() operations and aggregation functions to analyze trends, patterns, and summaries across large datasets.

- **Data Transformation & Reshaping:** Transform, merge, and reshape datasets to support more effective analysis and streamlined analytical workflows.
- **Functional Data Processing:** Apply user-defined and third-party functions to Pandas objects to extend analytical capabilities and customize data processing.
- **Advanced Data Visualization:** Create clear, informative, and visually compelling data visualizations using advanced Matplotlib features and Seaborn enhancements.
- **NumPy for Analytical Efficiency:** Utilize NumPy arrays and operations to manipulate large numerical datasets and improve analytical performance.
- **Applied Scientific Computing with SciPy:** Gain practical exposure to key SciPy subpackages to support statistical analysis, optimization, and scientific workflows.

## Audience

This course is designed for data professionals who already have foundational Python and Pandas skills and want to apply Python more effectively to real-world data analysis problems. Typical roles include data analysts, data scientists, engineers, and researchers.

## Pre-Requisites

Participants should have a solid foundation in Python and introductory Pandas concepts, equivalent to those covered in Python Fundamentals for Data Science (TTPS4874). This course assumes prior experience with Python syntax, basic data structures, and simple data manipulation in Pandas.

TTPS4873	Fast Track to Python for Data Science and/or Machine Learning
TTPS4874	Python Fundamentals for Data Science

## Agenda

### 1) Pandas input and output

Reading data into Pandas dataframes and exporting to various formats.

- General input considerations
- Reading CSV Files
- Data cleaning
- Reading other data formats
- Exporting data

### 2) Pandas filtering and sorting

Selecting subsets of dataframes for focused analysis.

- Indexing rows and columns
- Multi-indexing
- Selection by conditions
- Sorting data

### **3) Pandas grouping and aggregation**

Consolidating data and providing sums and other aggregate values

- Using groupby()
- Aggregate functions
- Using data summaries
- Alternate approaches

### **4) Pandas Data Transformation**

Manipulating datasets for simpler analysis

- Applying functions to data
- Renaming columns and indexes
- Inserting and removing data
- Combining and merging dataframes
- Reshaping datasets

### **5) Advanced Matplotlib**

Going beyond the basics with Matplotlib

- Components of a figure
- Multiple plots
- Complex plots
- Matplotlib options and settings
- Customizing styles (and everything else)

### **6) Seaborn**

Learning how Seaborn supplements and improves on Matplotlib

- What does Seaborn provide?
- Using themes
- Advanced plot types

- Fine-tuning the details

## 7) Using NumPy

Loading large datasets into NumPy arrays for further analysis

- NumPy basics
- Creating arrays
- Indexing and slicing
- Builtin functions()
- Reading and writing data

## 8) Useful SciPy subpackages

A look at some of the 20-odd SciPy subpackages

- What is SciPy?
- `scipy.stats`
- `scipy.interpolate`
- `scipy.optimize`

The following are required:

- The Anaconda Python distribution (or equivalent), including
  - iPython
  - Jupyter
  - Pandas
  - Matplotlib
  - Seaborn
  - Numpy
  - SciPy

## For More Information

Please [contact us](#) or call 844-475-4559 toll free for more information about our training services (instructor-led, self-paced or blended), coaching and mentoring services, public course enrollment or questions, partner programs, courseware licensing options and more.