**Hands-On Data Analysis with Pandas (3 days)**

**Quick Start to Using the Pandas Library to Reshape, Clean, Aggregate, Analyze & Visualize Your Data**

Overview

Data analysis has become a necessary skill in a variety of domains where knowing how to work with data and extract insights can generate significant value. Geared for data team members with incoming Python scripting experience, **Hands-On Data Analysis with Pandas** will show you how to analyze your data, get started with machine learning, and work effectively with Python libraries often used for data science, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn.

Using real-world datasets, you will learn how to use the powerful pandas library to perform data wrangling to reshape, clean, and aggregate your data. Then, you will be able to conduct exploratory data analysis by calculating summary statistics and visualizing the data to find patterns. In the concluding lessons, you will explore some applications of anomaly detection, regression, clustering, and classification to make predictions based on past data. Students will leave the course armed with the skills required to use pandas to ensure the veracity of their data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets.

Learning Objectives

This course is approximately **50% hands-on**, combining expert lecture, real-world demonstrations and group discussions with machine-based practical labs and exercises.

Working in a hands-on learning environment, guided by our expert team, attendees will learn to:

* Understand how data analysts and scientists gather and analyze data
* Perform data analysis and data wrangling using Python
* Combine, group, and aggregate data from multiple sources
* Create data visualizations with pandas, matplotlib, and seaborn
* Apply machine learning (ML) algorithms to identify patterns and make predictions
* Use Python data science libraries to analyze real-world datasets
* Use pandas to solve common data representation and analysis problems
* Build Python scripts, modules, and packages for reusable analysis code
* Perform efficient data analysis and manipulation tasks using pandas
* Apply pandas to different real-world domains with the help of step-by-step demonstrations
* Get accustomed to using pandas as an effective data exploration tool.

Audience & Pre-Requisites

This course is geared for Python-experienced attendees who wish to be equipped with the skills you need to use pandas to ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets.

**Take Before:** Students should have skills at least equivalent to the following course(s) or should have attended as a pre-requisite:

* TTDS6600 Understanding Data Science | A Technical Overview – 1 day (helpful but not required)
* TTPS4800 Introduction to Python Programming (3 days)

**Follow On Courses**: Our core Python, data science and machine learning training coursesprovide students with a solid foundation for continued learning based on role, goals, or their areas of specialty. Our learning paths offer a wide variety of related follow-on courses such as:

* TTPS4876 Next Level Python in Data Science (Intermediate) (5 days)
* TTML5506-P Machine Learning Essentials with Python (3 days)
* Please see the attached **Python Training Suite** or **Machine Learning Series** list for additional topics and titles.

Course Topics / Agenda

Please note topics, agenda and labs are subject to change, and may adjust during live delivery based on audience interests, skill-level and participation.

1. Introduction to Data Analysis

* Fundamentals of data analysis
* Statistical foundations
* Setting up a virtual environment

1. Working with Pandas DataFrames

* Pandas data structures
* Bringing data into a pandas DataFrame
* Inspecting a DataFrame object
* Grabbing subsets of the data
* Adding and removing data

1. Data Wrangling with Pandas

* What is data wrangling?
* Collecting temperature data
* Cleaning up the data
* Restructuring the data
* Handling duplicate, missing, or invalid data

1. Aggregating Pandas DataFrames

* Database-style operations on DataFrames
* DataFrame operations
* Aggregations with pandas and numpy
* Time series

1. Visualizing Data with Pandas and Matplotlib

* An introduction to matplotlib
* Plotting with pandas
* The pandas.plotting subpackage

1. Plotting with Seaborn and Customization Techniques

* Utilizing seaborn for advanced plotting
* Formatting
* Customizing visualizations

1. Financial Analysis - Bitcoin and the Stock Market

* Building a Python package
* Data extraction with pandas
* Exploratory data analysis
* Technical analysis of financial instruments
* Modeling performance

1. Rule-Based Anomaly Detection

* Simulating login attempts
* Exploratory data analysis
* Rule-based anomaly detection

1. Getting Started with Machine Learning in Python

* Learning the lingo
* Exploratory data analysis
* Preprocessing data
* Clustering
* Regression
* Classification

1. Making Better Predictions - Optimizing Models

* Hyperparameter tuning with grid search
* Feature engineering
* Ensemble methods
* Inspecting classification prediction confidence
* Addressing class imbalance
* Regularization

1. Machine Learning Anomaly Detection

* Exploring the data
* Unsupervised methods
* Supervised methods
* Online learning

1. The Road Ahead

* Data resources
* Practicing working with data
* Python practice