

Python 3: Web Scraping and Machine Learning

For Business and Finance Professionals

Summary

During this hands-on course, participants will work with Python and several popular packages to create programs that acquire, consolidate, analyze and present large data sets. Whether originated from websites or internal databases, this course will demonstrate core techniques to efficiently manage and explore business data.



Prerequisites

A general understanding of programming principles and Python is recommended for this course. The course material will build on the content of “Python 1: Core Data Analysis” and “Python 2: Visualization and Analysis”. Participants should be familiar with Python packages and their installation. Participants are expected to download & install Anaconda or an equivalent Python distribution in advance of the course.



Timing

This course requires 1 day.

Learning Topics

1. Acquiring Data from Websites (“Web Scraping”)

- ✓ Automate corporate due diligence and data gathering by designing programs to download publicly available information from websites
- ✓ Aggregate alternative data from industry websites
- ✓ Create programs for competitor analysis and price comparisons
- ✓ Review API's and Python packages used for web scraping, such as *Requests*, *Urllib* and *Beautiful Soup* to parse downloaded data into a format that can be analyzed and visualized
- ✓ Automate user interactions with websites using the *Selenium* package
- ✓ Extract financial data from Yahoo Finance, EDGAR and other sources
- ✓ Learn to import data from various types of websites (HTML, JSON, XML, PDFs)

2. Machine Learning (ML) & AI Applications

- ✓ Overview popular Machine Learning algorithms and how companies are leveraging Python's ML packages
- ✓ Use advanced language processing packages for natural language processing (NLP) to extract key information from news articles and press releases
- ✓ Review the *NLTK* and *SpaCy* packages used for NLP
- ✓ Use image processing packages to extract text and key information from images

3. Automation, Visualization and Best Practices

- ✓ Tips for moving and creating folders on the fly and importing data from multiple source files
- ✓ Automate extracting and cleaning tables from PDF files
- ✓ Build powerful visualizations using more advanced visualization packages such as *Bokeh*, *Seaborn*, and *Plotly*
- ✓ Create interactive dashboards and charts using *Dash* and *Streamlit* packages
- ✓ Explore the integration of Python with Power BI, Microsoft's powerful dashboarding and visualization tool