Course Length

3 days

Course Description

Data is the new oil, but it comes crude. To do anything meaningful - modelling, visualization, machine learning, for predictive analysis – you first need to wrestle and wrangle with data. Data Wrangling with Python teaches you the essentials that will get you up and running with data wrangling in no time.

Overview

For data to be useful and meaningful, it must be curated and refined. Data Wrangling with Python teaches you the core ideas behind these processes and equips you with knowledge of the most popular tools and techniques in the domain.

The course starts with the absolute basics of Python, focusing mainly on data structures. It then delves into the fundamental tools of data wrangling like NumPy and Pandas libraries. You'll explore useful insights into why you should stay away from traditional ways of data cleaning, as done in other languages, and take advantage of the specialized pre-built routines in Python. This combination of Python tips and tricks will also demonstrate how to use the same Python backend and extract/transform data from an array of sources including the Internet, large database vaults, and Excel financial tables. To help you prepare for more challenging scenarios, you'll cover how to handle missing or wrong data, and reformat it based on the requirements from the downstream analytics tool. The course will further help you grasp concepts through real-world examples and datasets.

By the end of this course, you will be confident in using a diverse array of sources to extract, clean, transform, and format your data efficiently.

Scope

This course teaches concepts by deep-dive on-hand exercises. Throughout the course, you will learn data wrangling with hands-on exercises and activities. You' ll find checklists, best practices, and critical points mentioned throughout the lessons, making things more interesting.

Target Audience

Data Wrangling with Python takes a practical approach to equip beginners with the most essential data analysis tools in the shortest possible time. It contains multiple activities that use real-life business scenarios for you to practice and apply your new skills in a highly relevant context.

Technical Requirements

Hardware:

For an optimal student experience, we recommend the following hardware configuration:

- Processor: Intel Core i5 or equivalent
- Memory: 8GB RAM or higher
- Internet Connection

Software:

You' ll also need the following software installed in advance:

- OS: Windows 7 SP1 64-bit, Windows 8.1 64-bit or Windows 10 64-bit, Ubuntu Linux, or the latest version of OS X
- Browser: Google Chrome/Mozilla Firefox Latest Version
- Notepad++/Sublime Text as IDE (Optional, as you can practice everything using Jupyter note course on your browser)
- Python 3.4+ (latest is Python 3.7) installed (from https://python.org)
- Python libraries as needed (Jupyter, NumPy, Pandas, Matplotlib, BeautifulSoup4, and so)

Course Outline

Lesson 1: Introduction to Data Structure using Python

- Python for Data Wrangling
- Lists, Sets, Strings, Tuples, and Dictionaries

Lesson 2: Advanced Operations on Built-In Data Structure

- Advanced Data Structures
- Basic File Operations in Python

Lesson 3: Introduction to NumPy, Pandas, and Matplotlib

- NumPy Arrays
- Pandas DataFrames
- Statistics and Visualization with NumPy and Pandas
- Using NumPy and Pandas to Calculate Basic Descriptive Statistics on the DataFrame

Lesson 4: Deep Dive into Data Wrangling with Python

- Subsetting, Filtering, and Grouping
- Detecting Outliers and Handling Missing Values
- Concatenating, Merging, and Joining
- Useful Methods of Pandas

Lesson 5: Get Comfortable with a Different Kind of Data Sources

- Reading Data from Different Text-Based (and Non-Text-Based) Sources
- Introduction to BeautifulSoup4 and Web Page Parsing

Lesson 6: Learning the Hidden Secrets of Data Wrangling

- Advanced List Comprehension and the zip Function
- Data Formatting

Lesson 7: Advanced Web Scraping and Data Gathering

- Basics of Web Scraping and BeautifulSoup libraries
- Reading Data from XML

Lesson 8: RDBMS and SQL

- Refresher of RDBMS and SQL
- Using an RDBMS (MySQL/PostgreSQL/SQLite)

Lesson 9: Application in real life and Conclusion of course

- Applying Your Knowledge to a Real-life Data Wrangling Task
- An Extension to Data Wrangling