Faysal Aziz

Python for Data

A 2-day refresher course

FA-0381 · 2-day

Software Development · Data & Analytics

Introduction

Welcome to the Python Refresher Course on Data Munging and Wrangling! In today's data-driven world, the ability to efficiently process and manipulate data is an invaluable skill. Data munging, also known as data wrangling, involves transforming and mapping data from its raw form into a more usable format. This process is essential for data analysis, machine learning, and decision-making in various industries.

Python, with its rich ecosystem of libraries and tools, has become a go-to language for data professionals. This course aims to refresh your Python skills while diving deep into techniques and best practices for data munging and wrangling. Whether you are preparing data for analysis, cleaning datasets, or performing complex transformations, this course will equip you with the knowledge and tools to handle data efficiently and effectively.

Learning outcomes

By the end of this course, participants will be able to:

- Understand the importance and applications of data munging and wrangling in data science.
- Utilize Python libraries such as pandas, numpy, and others to manipulate and clean data.
- Perform data preprocessing tasks including handling missing values, data normalization, and feature extraction.
- Implement data transformation techniques such as pivoting, melting, and reshaping data.
- Apply advanced data manipulation techniques including merging, joining, and concatenating datasets.
- Use Python for exploratory data analysis (EDA) to uncover insights and patterns in data.
- Automate repetitive data wrangling tasks to improve efficiency.
- Ensure data integrity and quality through validation and verification methods.

Prerequisites

To get the most out of this course, participants should have:

- Basic understanding of Python programming.
- Familiarity with fundamental data structures in Python (lists, dictionaries, tuples, etc.).
- Basic knowledge of using Python libraries such as pandas and numpy.
- Understanding of basic statistical concepts and data analysis techniques.

Training outline

1. Introduction to Data Munging and Wrangling

1. Definition and significance
2. Applications in various industries
3. Common challenges and solutions

2. Review of Python Fundamentals

1. Data types and structures
2. Control flow and loops
3. Functions and modules

3. Introduction to Pandas and Numpy

1. Overview of pandas and numpy
2. Installing and importing libraries
3. Basic operations and data structures (Series, DataFrame, ndarray)

4. Loading and Inspecting Data

1. Reading data from various sources (CSV, Excel, SQL, JSON)
2. Inspecting data (head, tail, info, describe)
3. Data types and conversions

5. Handling Missing Data

1. Identifying missing values

2. Methods to handle missing data (dropna, fillna)
3. Interpolation and imputation techniques

6. Data Transformation

1. Renaming columns
2. Changing data types
3. Handling duplicates
4. Applying functions to columns and rows (apply, map, applymap)

7. Data Cleaning

1. Removing unwanted characters
2. Standardizing text data (lowercase, uppercase, trimming)
3. Handling inconsistent data formats

8. Filtering and Selecting Data

1. Indexing and slicing
2. Conditional filtering
3. Selection based on conditions (loc, iloc)

9. Advanced Data Manipulation

1. Merging and joining datasets
2. Concatenation of dataframes
3. Pivoting and melting data
4. Grouping data and aggregation (groupby, agg)
5. Working with multi-index dataframes

10. Feature Engineering

1. Creating new features from existing data
2. Binning and discretization
3. Encoding categorical variables (one-hot encoding, label encoding)
4. Feature scaling and normalization

11. Exploratory Data Analysis (EDA)

1. Summary statistics and data visualization
2. Detecting outliers and anomalies
3. Correlation analysis and pair plots
4. Time series analysis (if applicable)

12. Data Validation and Quality Assurance

1. Ensuring data consistency and integrity
2. Validating data against defined criteria
3. Handling and reporting errors

13. Automating Data Wrangling Tasks

1. Writing reusable functions for repetitive tasks
2. Scripting and scheduling data wrangling workflows
3. Using Jupyter notebooks for documenting and sharing processes

14. Case Studies and Practical Exercises

1. Real-world data wrangling scenarios
2. Hands-on exercises and projects
3. Applying learned techniques to solve practical problems

15. Resources and Further Learning

1. Recommended books and courses
2. Online resources and communities
3. Staying updated with the latest trends and tools in data wrangling

This course is designed to provide a comprehensive and practical understanding of data munging and wrangling using Python. We hope you find it informative, engaging, and highly beneficial for your data processing needs. Let's dive in and start transforming raw data into actionable insights!

© 2026 Faysal Aziz · Kuala Lumpur