

COM6018 Data Science with Python

Week 4 - Introducing Pandas

Varvara Papazoglou

Overview

- Review of Lab Class 3
- Common traps when using NumPy
- Introducing Pandas
- Preview of Lab Class 4
- Some words about the 1st assignment

Review of Lab Class 3

- We continued our analysis of the atmospheric gas concentration data.
- We want to better understand the oscillations in the data.
- We are interested in extracting information including:
 - The period of the oscillations.
 - The amplitude of the oscillations.
 - Dates at which the peaks and dips occur.

The Problem

Finding the peaks and dips in the oscillations is not so straightforward.

Problem:

- The data is very noisy.
- The daily measurements can bounce up and down in a seemingly random way.
- There are many local peaks and dips that are not part of the oscillations we are interested in.

We need to smooth the data to reduce the noise.

Tasks

- Reading datasets from files.
- Smoothing the data
 - Using a simple moving average
 - By applying a weighted window function
- Finding the peaks and dips in the smoothed data.
- Visualising the peaks and dips in the data.
- Answering questions about the period and size of the oscillations, etc.

We will do all of this using NumPy.

The Solution Notebook

[Open the Solution Notebook](#)

Common Traps When Using NumPy

[Open Notebook](#)

Introducing Pandas

What is Pandas?

- **Data Analysis Library**: Pandas is a fast, open-source Python library for working with structured and labelled data.
- **DataFrame Object**: Provides a powerful, table-like structure built on top of NumPy, ideal for exploring, transforming, and summarising datasets.
- **Rich Data Operations**: Includes high-level tools for data cleaning, reshaping, joining, grouping, and statistical analysis.
- **Interoperability**: Works seamlessly with NumPy, Matplotlib, scikit-learn, and modern data tools like Polars or PyArrow.
- **Time-Series and Indexing**: Excellent support for sequential data analysis.

Pandas Tutorial

[Link to the Pandas tutorial](#)

Lab Class 4 Preview

- We will use Pandas to extend our analysis of greenhouse gas data.
- We will see how using Pandas makes it very easy to read and manipulate datasets.

Lab Class 4

[Link to lab class](#)

Next Steps

- Review the Week 3 materials:
 - Review the Lab 3 solution notebook if you have not already done so.
 - Read the Week 3 tutorial on NumPy
- Prepare for the Week 4 lab class:
 - Read through the Pandas tutorial.
- Check out the reading list for Week 4 on Blackboard.