Mastering Data Analysis: Pandas and NumPy Essentials

WORKSHOP DETAILS

- Instructor: Tanya Khanna
- Email: tk759@scarletmail.rutgers.edu
- Course Materials: Github Link - https://github.com/Tanya-Khanna/DataScienceWorkshop_Fall-2024_NBL
- Workshop Recordings: https://libguides.rutgers.edu/datascience/python
- Spring 2024 Workshops: Link
- Workshop Feedback Form: https://rutgers.libwizard.com/f/graduate_specialist_feedback

WORKSHOPS SCHEDULE

Fall 2024 workshops calendar

Introduction to Python Programming	September 9, 2024; 4 - 5:30 PM
Advanced Python Programming	September 16, 2024; 4 - 5:30 PM
Web Scraping with Python	September 23, 2024; 4 - 5:30 PM
Mastering Data Analysis: Pandas and Numpy	September 30, 2024; 4 - 5:30 PM
Data Management with Python: SQL and NoSQL	October 7, 2024; 4 - 5:30 PM
Python for Visualization and Exploration	October 14, 2024; 4 - 5:30 PM
Introduction to Machine Learning: Supervised Learning	October 21, 2024; 4 - 5:30 PM
Introduction to Machine Learning: Unsupervised Learning	October 28, 2024; 4 - 5:30 PM
Deploying Machine Learning Models	November 4, 2024; 4 - 5:30 PM
Ethical AI and Responsible Data Science	November 11, 2024; 4 - 5:30 PM
Large Language Models (LLMs) and ChatGPT	November 18, 2024; 4 - 5:30 PM

TABLE OF CONTENTS

What is NumPy?

- NumPy, or numerical Python is a library that is used for advanced mathematical computations.
- Provides support for large, multi-dimensional arrays and matrices.
- Can perform high-level mathematical functions to operate on these arrays.

- NumPy is like a super-powered calculator for computers. It helps them handle big lists and tables of numbers and do complex math on them quickly.
- Think of it as a toolbox that lets your computer work with lots of numbers at once, making it much easier and faster to solve math problems.

Uses of NumPy:

NumPy is like a high-speed storage room for numbers compared to Python's usual way of keeping numbers in a list.

It's more organized because you can tell it exactly what type of numbers you're storing (like whole numbers, decimal numbers, etc.), which helps save space and makes things run smoother.

It does math faster. Can create random data with almost any specification that can be thought of.

It lets you play with numbers in many dimensions, like stacking sheets of paper into a book. This is something you can't easily do with Python's lists, which are more like a single string of beads.

Thus, ideal for tasks such as data analysis, machine learning, and mathematical operations.

What is Pandas?

- Pandas is a fast, powerful, flexible and easy-to-use open source data analysis and manipulation tool built on top of the Python programming language.
- It is also one of the most popular libraries used by data experts from all around the world.
- Think of it as a super-smart spreadsheet that can do a lot of tricks very quickly.

Why you should learn Pandas:

- We're all surrounded by tons of data every day (like what we buy, watch, or search online), and being able to understand what all that data means is super important nowadays.
- Companies love to use data to make better decisions, like figuring out what their customers like or how to make their services faster and cheaper.
- Learning pandas is like learning how to fish in the sea of data: once you know how, you can apply it to any data you come across, making you a bit of a data wizard.
- Pandas is friendly for beginners, especially if you've played around with Python, and it's a skill that's in high demand, especially if you're into data science or machine learning (which is like teaching computers to recognize patterns and make decisions based on data).

Uses of Pandas:

Organize and adjust your data, like putting it in neat tables or getting rid of parts you don't need. Also, you can make simple charts to see your data visually, like pie charts or line graphs.

Find and fix any missing or wrong info.

Group your data in ways that make sense for what you're looking at, like how many sales happened in each region.

Thus, ideal for tasks such as data wrangling, data analysis and data visualisation.