

# Python для Data Science

## **Заняття 3.**

Знайомство з бібліотекою Pandas

# План заняття

---



- Короткий опис бібліотеки Pandas
- Основні структури даних в Pandas
- Робота з Pandas Series
- Робота з Pandas DataFrame

# Pandas

---

**Pandas** — це швидкий, потужний, гнучкий і простий у використанні інструмент аналізу та обробки даних з відкритим кодом, створений на основі мови програмування Python.

На високому рівні **Pandas** працює дуже схоже на електронну таблицю (наприклад, Microsoft Excel або Google Sheets), оскільки ви працюєте з рядками та стовпцями.

Але краще =)

Pandas є однією з основоположних бібліотек будь-якого робочого процесу обробки даних, оскільки дозволяє виконувати обробку, аналіз та обмін даними. Це особливо важливо, оскільки багато хто вважає, що етап попередньої обробки даних займає **до 80% часу спеціаліста з обробки даних**.



# Pandas

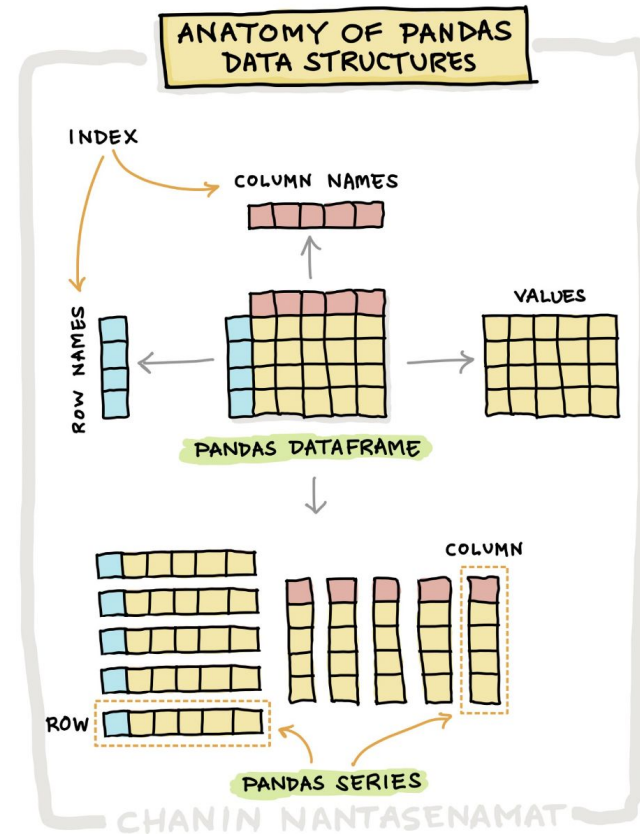
---

**Pandas** побудовано на основі пакета **NumPy**, тобто велика частина структури NumPy використовується або копіюється в Pandas. Дані в Pandas часто використовуються для статистичного аналізу в **SciPy**, побудови графіків функцій із **Matplotlib** і алгоритмів машинного навчання в **Scikit-learn**.

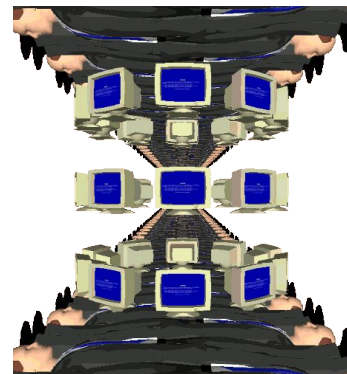


# Pandas data structures

- **Series** — це одновимірний масив з мітками (labeled), здатний містити елементи будь-яких типів (цілі числа, рядки, числа з плаваючою комою, об'єкти Python тощо). Всі мітки разом називаються індексом. Можете сприймати як колонку excel sheet.
- **DataFrame** — це двовимірна, потенційно неоднорідна таблична структура даних із позначеними осями (рядки та стовпці) і стовпцями потенційно різних типів. Ви можете думати про DataFrame як про електронну таблицю чи SQL-таблицю, або як набір об'єктів Series. Як правило, це найбільш часто використовуваний об'єкт Pandas. Pandas DataFrame складається з трьох основних компонентів: даних, індексів рядків і стовпців.



# Live coding / Практика



**Питання**

???

???