

Pandas - високорівнева бібліотека
описання, оброблення, аналіз даних
у табличному форматі (.csv, .tsv, .xlsx...)

Основний інструмент для аналізу даних,
• маніпуляцій над ними,
• побудови моделей на них.

Pandas - таблиці
Numpy - матриці та вектори
Matplotlib, Seaborn - візуалізація
scikit-learn - реалізація методів ML

Основні структури даних Pandas

Series - одновимірний індексований масив
даних деякого дискретного типу
(str, int...)

Data Frame - двовимірна структура даних,
що являє собою таблицю, кожен
стовпець якої містить дані одного типу.

```
dict = {"brand": "Ford",  
        "model": "Must.",  
        "year": "1964"}  
df = {"series1 name": [...],  
      "series2 name": [...],  
      : }
```

DF можна розуміти як словник
об'єктів типу Series.

Структура добре підходить для
зберігання реальних табличних даних:
рядки відповідають описам окремих об'єктів,
а стовпці відповідають ознакам.

Example Series

| | Name | Team | Number |
|-----|---------------|----------------|--------|
| 0 | Avery Bradley | Boston Celtics | 0.0 |
| 1 | ... | ... | 30.0 |
| 2 | ... | ... | NaN |
| 3 | ... | ... | ... |
| ... | ... | ... | ... |
| ... | ... | ... | ... |

`ser = pd.Series(df['Name'])`
`['Team']`
`['Number']`

Data Frame — таблица

Series — колонка

у списка (list) бывается только элементы
разных типов;
у Series же можно, все не принято

Data Frame Example

Columns
axis = 1

Column names

| | Name | Team | Number | Position | Age |
|---|------|------|--------|----------|------|
| 0 | ... | ... | | | |
| 1 | ... | | | | |
| 2 | ... | | | | |
| 3 | | NaN | | C | 21.0 |
| 4 | | | 8.0 | | |
| 5 | | | | | |
| 6 | | | | | |

Index label

Index
axis = 0

Missing value

Data

Axis 0 - perokke

Axis 1 - konatku

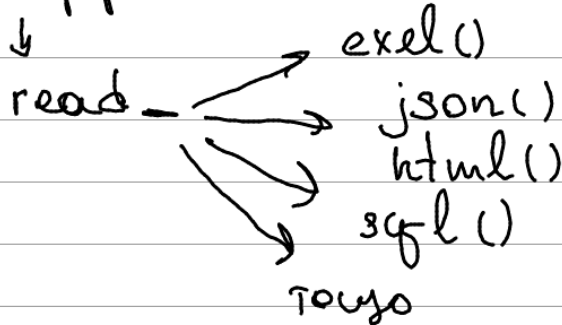
Зчитування даних з файлу

CSV- найпопулярніший

`pd.read_csv()` - автоматично розпізнає роздільник та створює `DataFrame` з даних.

налаштування: вибір рядка заголовка
вказання конкретних стовпців
обробка пропущених значень
визначення типів даних, тощо

інші формати:



os - стандартна бібліотека, не потребує встановлення
`os.listdir()` - вивести весь вміст поточної директорії