COESIONE ITALIA 21-27 EMILIA-ROMAGNA

Cofinanziato dall'Unione europea

GUARDA AVANTI

Big Data, nuove competenze per nuove professioni

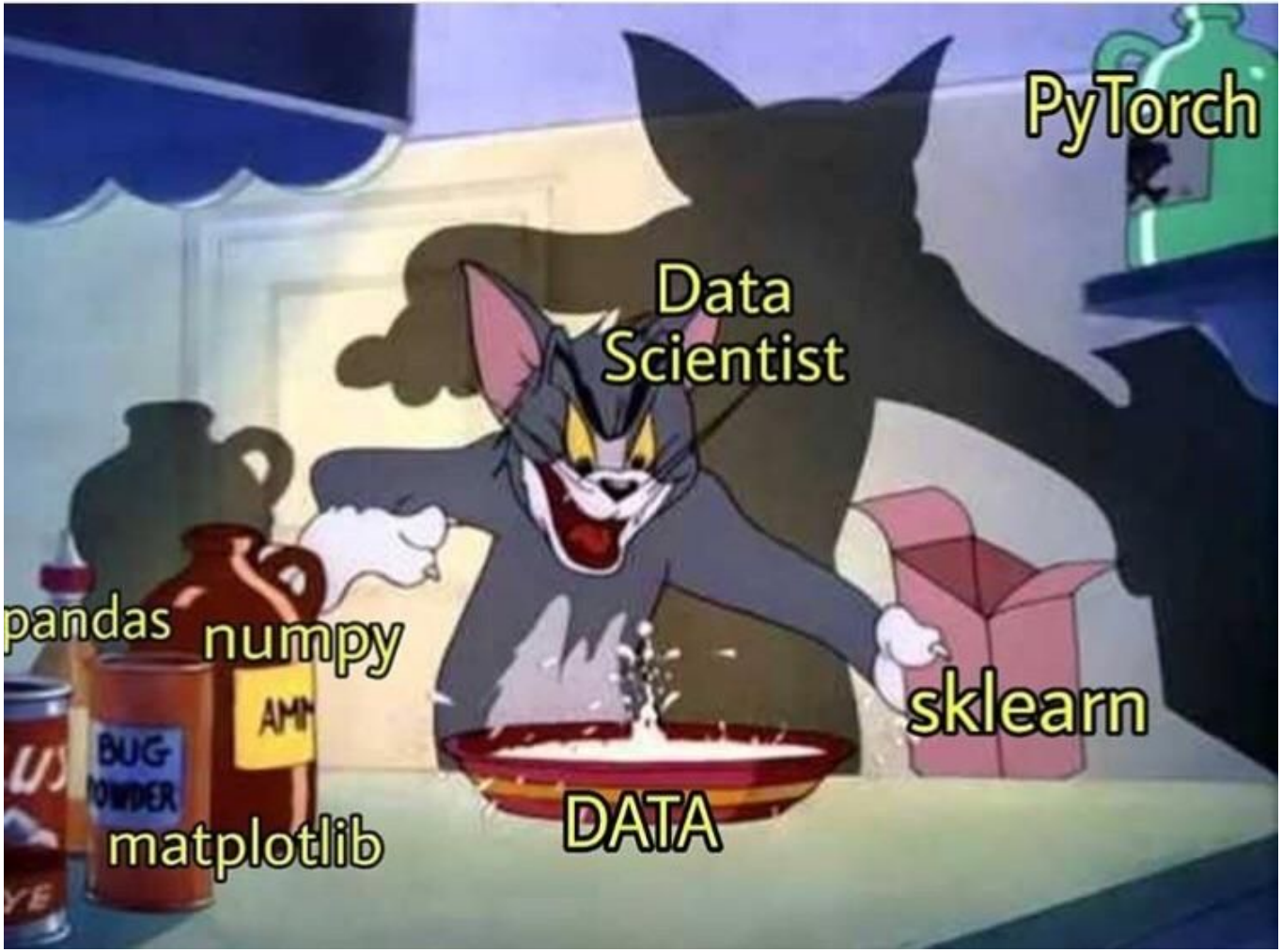
(Progetto rivolto a laureati in tutte le aree disciplinari, co-finanziato dal Fondo Sociale Europeo Plus 2021-2027 Regione Emilia-Romagna)

DATA LAB

Programma della lezione

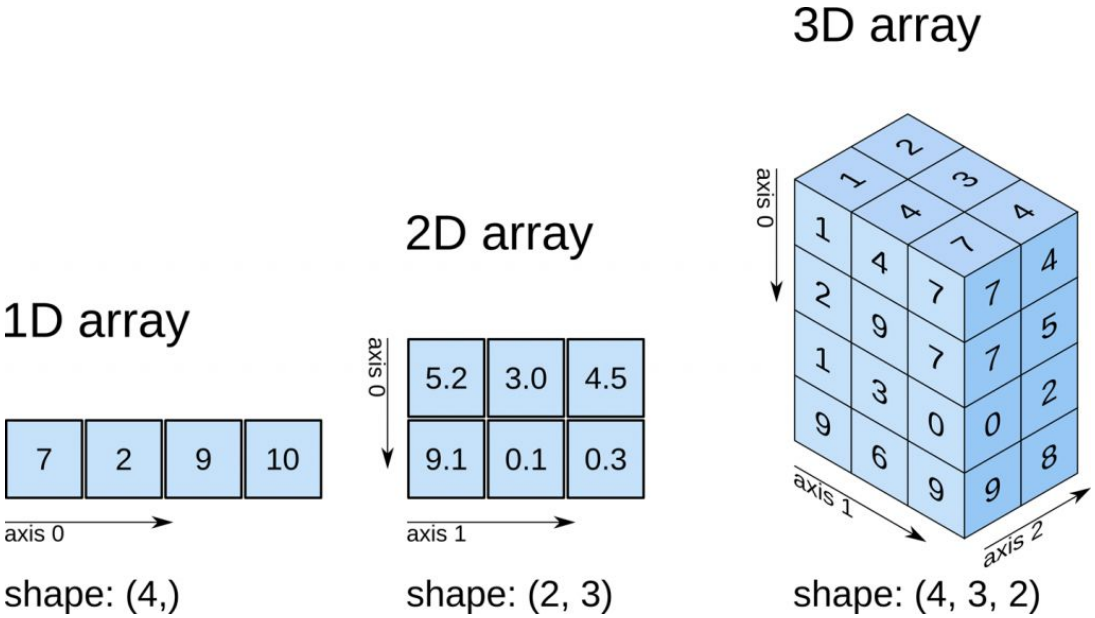
- **NumPy**
  - NumPy array Vs lista
  - Creare array
  - Operazioni tra array
  - Cambiare shape di un array

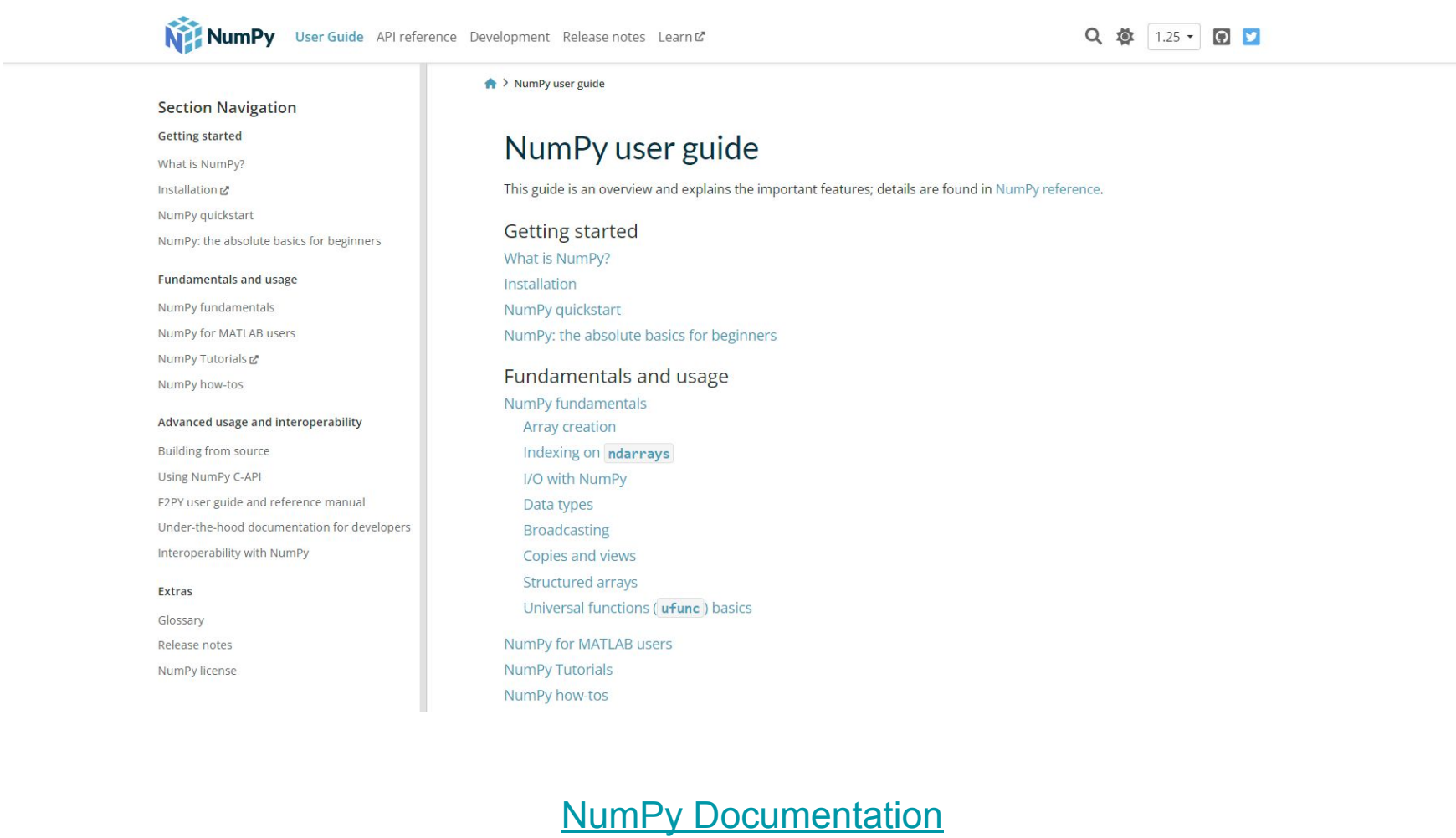
- Cos'è **Pandas**
- Dataframe
- Caricare dati
- Leggere dati
- Selezionare dati
- Iterare dati
- Ordinare dati
- Salvare dati
- Filtrare dati
- Manipolare i dati
- Pulire dati
- Visualizzare i dati


PyTorch
Data
Scientist
pandas
numpy
sklearn
BUG
matplotlib
DATA

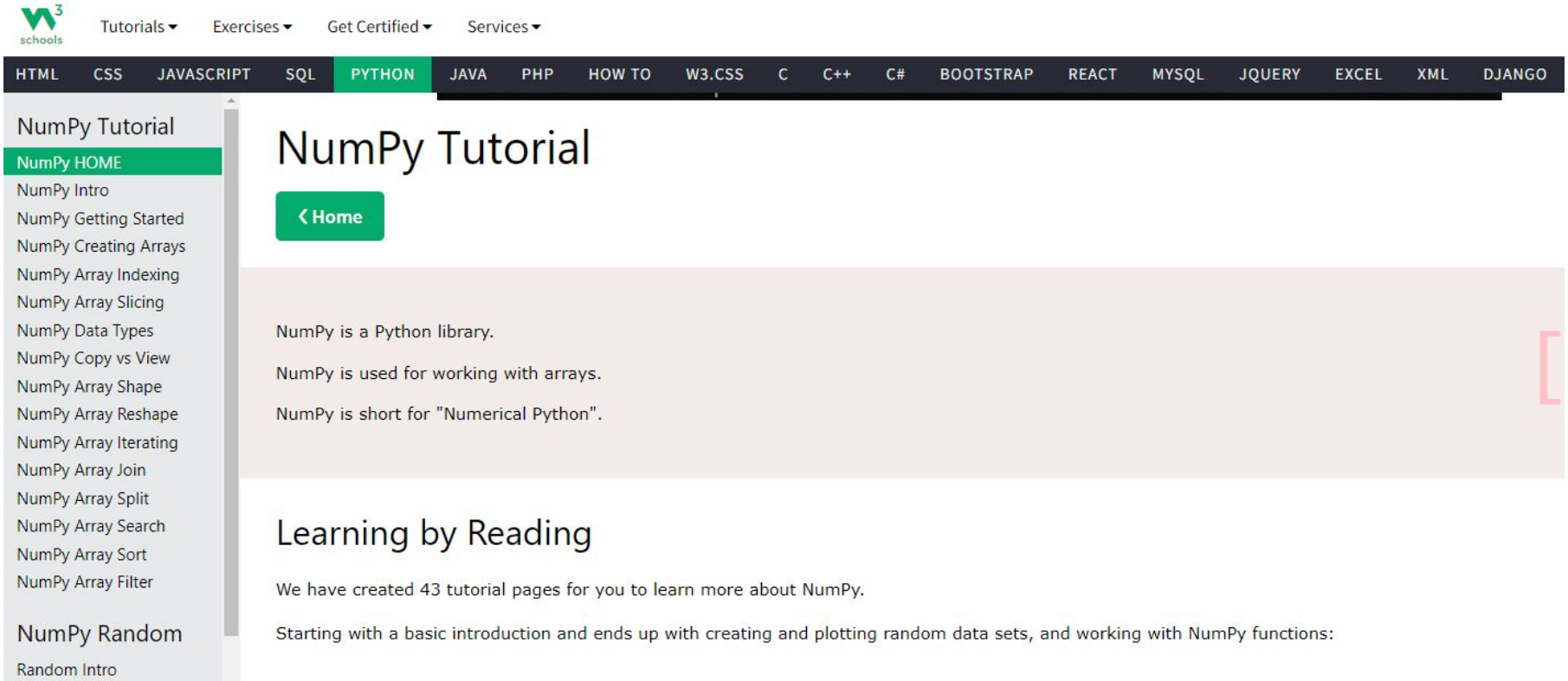
Cos’è NumPy?

NumPy è una libreria di Python che permette di lavorare con gli array.





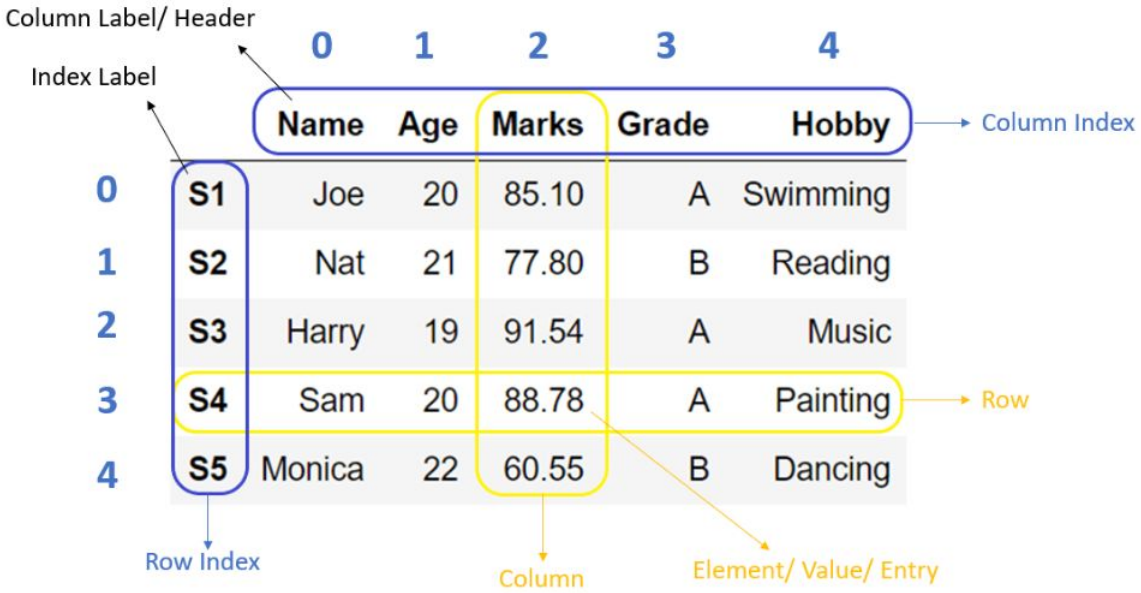
[NumPy Documentation]



[NumPy tutorial di W3 School]

Cos’è Pandas?

Pandas è una libreria **costruita sopra NumPy** che si usa per lavorare con i dati.



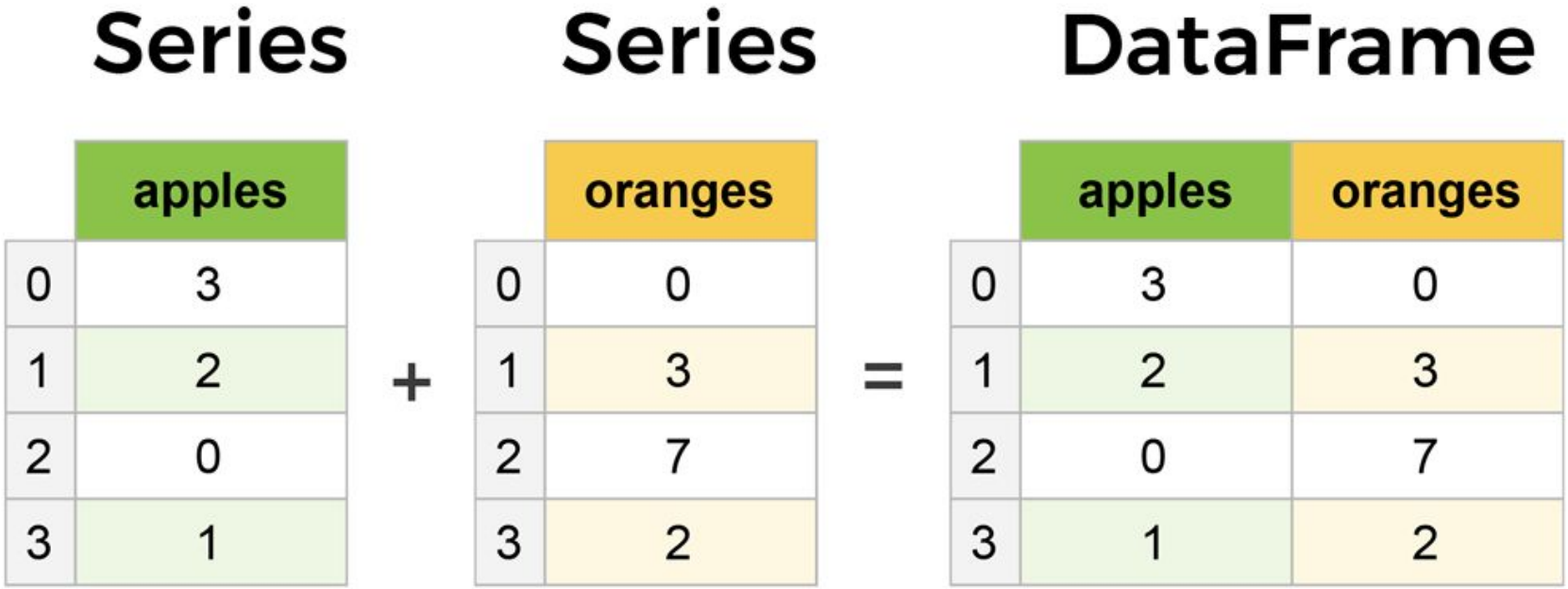
Perché Pandas?

1. Permette di **analizzare**, **manipolare** e **visualizzare** grandi quantità di dati

2. Offre un’ampia gamma di metodi per la pulizia dei dati

3. Ha due strutture dati molto potenti: Dataframe e Series

Dataframe Vs Series



10 minutes to pandas
Intro to data structures
Essential basic functionality
IO tools (text, CSV, HDF5, ...)
PyArrow Functionality
Indexing and selecting data
MultiIndex / advanced indexing
Copy-on-Write (CoW)
Merge, join, concatenate and compare
Reshaping and pivot tables
Working with text data
Working with missing data
Duplicate Labels
Categorical data
Nullable integer data type
Nullable Boolean data type
Chart visualization
Table Visualization

User Guide

The User Guide covers all of pandas by topic area. Each of the subsections introduces a topic (such as "working with missing data"), and discusses how pandas approaches the problem, with many examples throughout.

Users brand-new to pandas should start with 10 minutes to pandas.

For a high level summary of the pandas fundamentals, see Intro to data structures and Essential basic functionality.

Further information on any specific method can be obtained in the API reference.

How to read these guides

In these guides you will see input code inside code blocks such as:

```
import pandas as pd
pd.DataFrame({'A': [1, 2, 3]})
```

or:

On this page
How to read these guides
Guides

Show Source

[Pandas Documentation]

Pandas Tutorial
Pandas HOME
Pandas Intro
Pandas Getting Started
Pandas Series
Pandas DataFrames
Pandas Read CSV
Pandas Read JSON
Pandas Analyzing Data

Cleaning Data
Cleaning Data
Cleaning Empty Cells
Cleaning Wrong Format
Cleaning Wrong Data
Removing Duplicates

Correlations
Pandas Correlations

Pandas Introduction

‹ Previous Next ›

What is Pandas?

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Why Use Pandas?

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

Relevant data is very important in data science.

[Pandas Tutorial di W3 Schools]