



# 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**

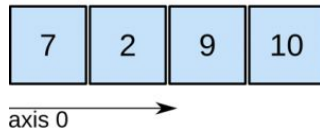
- Dataframe
- Caricare dati
- Leggere dati
- Selezionare dati
- Iterare dati
- Ordinare dati
- Salvare dati
- Filtrare dati
- Manipolare i dati
- Pulire dati
- Visualizzare i dati



# Cos'è NumPy?

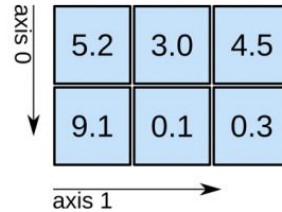
NumPy è una libreria di Python che permette di lavorare con gli array.

1D array



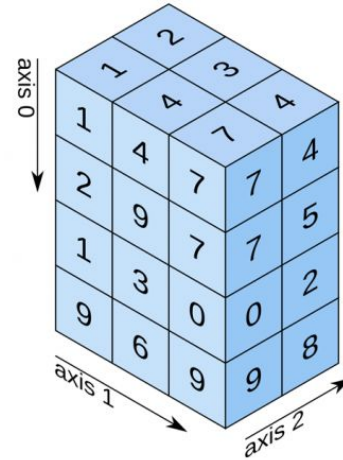
shape: (4,)

2D array



shape: (2, 3)

3D array



shape: (4, 3, 2)

## Section Navigation

### Getting started

[What is NumPy?](#)[Installation !\[\]\(799877f5c2f906134441300079881630\_img.jpg\)](#)[NumPy quickstart](#)[NumPy: the absolute basics for beginners](#)

### Fundamentals and usage

[NumPy fundamentals](#)[NumPy for MATLAB users](#)[NumPy Tutorials !\[\]\(4a7b4ce770af8456e11a71f9565c8c2b\_img.jpg\)](#)[NumPy how-tos](#)

### Advanced usage and interoperability

[Building from source](#)[Using NumPy C-API](#)[F2PY user guide and reference manual](#)[Under-the-hood documentation for developers](#)[Interoperability with NumPy](#)

### Extras

[Glossary](#)[Release notes](#)[NumPy license](#)[🏠](#) > [NumPy user guide](#)

# NumPy user guide

This guide is an overview and explains the important features; details are found in [NumPy reference](#).

## Getting started

[What is NumPy?](#)[Installation](#)[NumPy quickstart](#)[NumPy: the absolute basics for beginners](#)

## Fundamentals and usage

[NumPy fundamentals](#)[Array creation](#)[Indexing on `ndarrays`](#)[I/O with NumPy](#)[Data types](#)[Broadcasting](#)[Copies and views](#)[Structured arrays](#)[Universal functions \(`ufunc`\) basics](#)[NumPy for MATLAB users](#)[NumPy Tutorials](#)[NumPy how-tos](#)

[NumPy Documentation](#)





## NumPy Tutorial

### NumPy HOME

NumPy Intro  
NumPy Getting Started  
NumPy Creating Arrays  
NumPy Array Indexing  
NumPy Array Slicing  
NumPy Data Types  
NumPy Copy vs View  
NumPy Array Shape  
NumPy Array Reshape  
NumPy Array Iterating  
NumPy Array Join  
NumPy Array Split  
NumPy Array Search  
NumPy Array Sort  
NumPy Array Filter

### NumPy Random

Random Intro

# NumPy Tutorial

[← Home](#)

NumPy is a Python library.

NumPy is used for working with arrays.

NumPy is short for "Numerical Python".

## Learning by Reading

We have created 43 tutorial pages for you to learn more about NumPy.

Starting with a basic introduction and ends up with creating and plotting random data sets, and working with NumPy functions:

[NumPy tutorial di W3 School](#)

# Cos'è Pandas?

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

Column Label/ Header		0	1	2	3	4
Index Label		Name	Age	Marks	Grade	Hobby
0	S1	Joe	20	85.10	A	Swimming
1	S2	Nat	21	77.80	B	Reading
2	S3	Harry	19	91.54	A	Music
3	S4	Sam	20	88.78	A	Painting
4	S5	Monica	22	60.55	B	Dancing

Annotations:

- Column Index: Points to the header row (0-4).
- Row Index: Points to the first column (S1-S5).
- Column: Points to the 'Marks' column.
- Row: Points to the 'Painting' row.
- Element/ Value/ Entry: Points to the value '88.78'.

# 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

## Series

	apples
0	3
1	2
2	0
3	1

+

## Series

	oranges
0	0
1	3
2	7
3	2

=

## DataFrame

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

[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:

[Pandas Documentation](#)

☰ On this page

How to read these guides

Guides

 [Show Source](#)

## 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](#)





