



ANTICIPARE LA CRESCITA CON LE NUOVE COMPETENZE SUI BIG DATA – EDIZIONE 2

Operazione Rif. PA 2019-11596/RER “Anticipare la crescita con le nuove competenze sui Big Data”, approvata dalla Regione Emilia-Romagna con DGR n° 789 del 20 maggio 2019 e co-finanziata dal Fondo Sociale Europeo PO 2014-2020



Programma della lezione

- 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 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)
- Row Index (points to the first column)
- Row (points to the S4 row)
- Column (points to the Marks column)
- 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:

☰ 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](#)