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

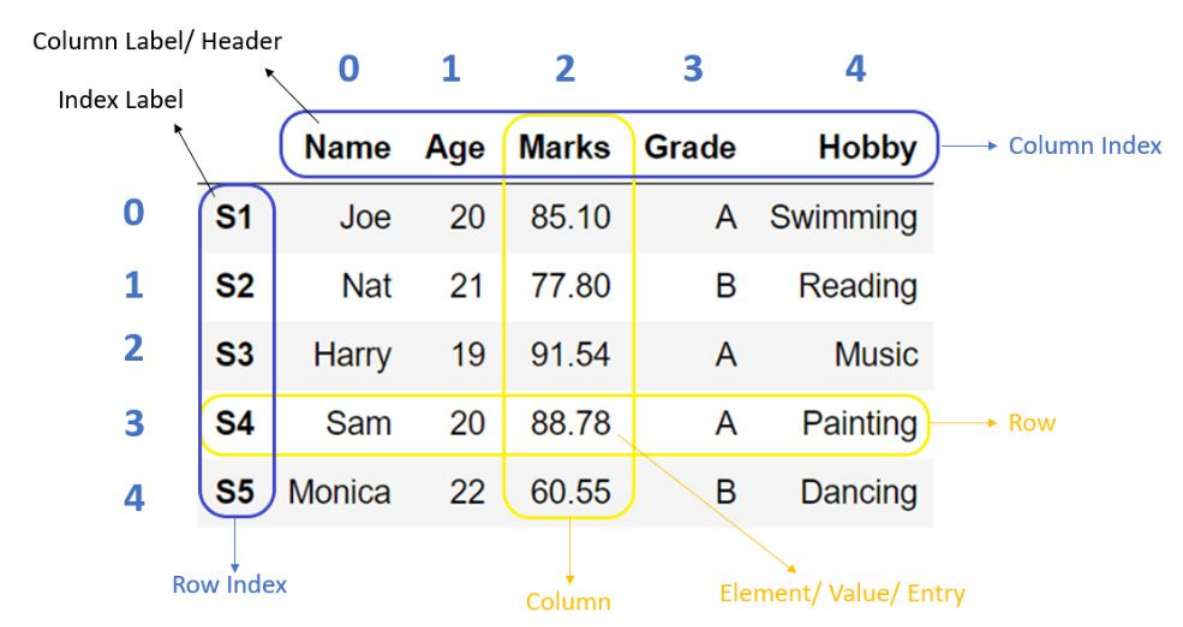
Operazione Rif. PA 2019-11596/RER "Anticipare la crescita con le nuove competenze sui Big Data", approvata dalla Regione Emilia-Romagna con DGR n° 789 del 20 maggio 2019 e co-finanziata dal Fondo Sociale Europeo PO 2014-2020

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.



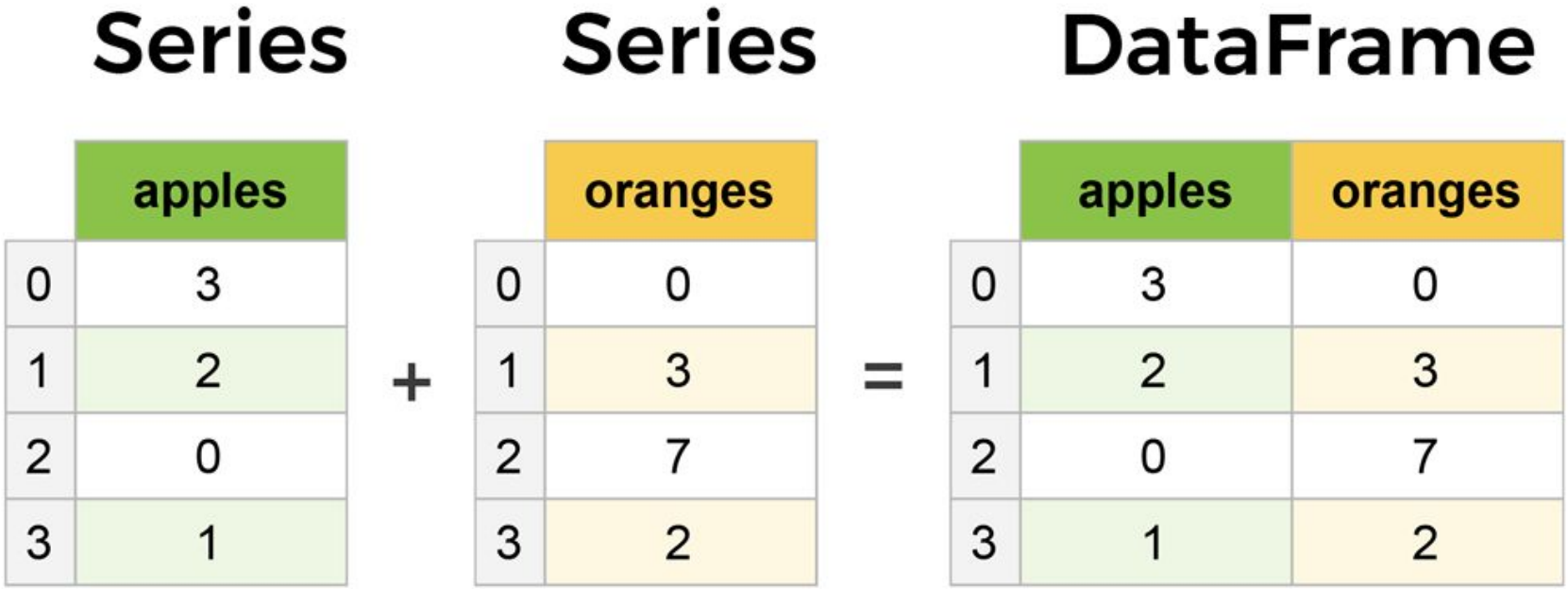
Perché Pandas?

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



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:

Show Source

[Pandas Documentation]

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]