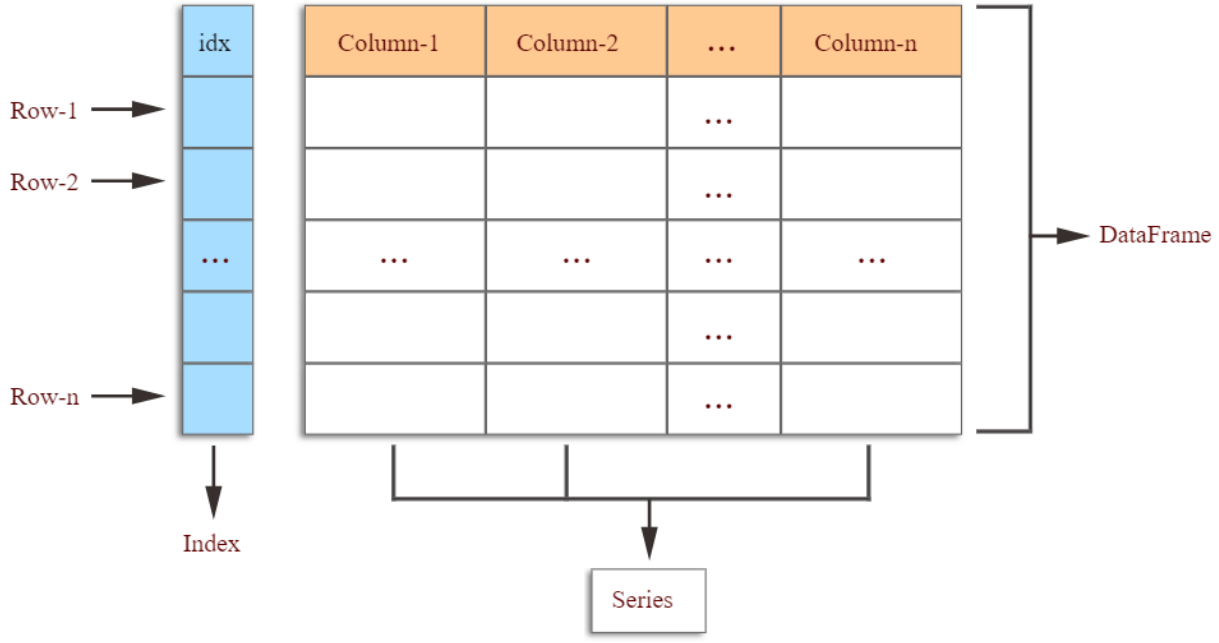
Pandas
Python for data analysis

Pandas

- libreria Python che fornisce ***strutture*** dati di ***alto livello*** progettate per facilitare e ottimizzare le operazioni sui dati
- le strutture dati sono in formato sequenziale (***Series***) o tabellare (***DataFrame***)
- caratteristiche principali:
  - caricamento e salvataggio di formati standard per dati tabellari
    - ***CSV*** (Comma-separated Values), TSV (Tab-separated Values) …
  - operazioni di indicizzazione e aggregazione di dati semplici e potenti
  - funzioni numeriche e statistiche

strutture dati

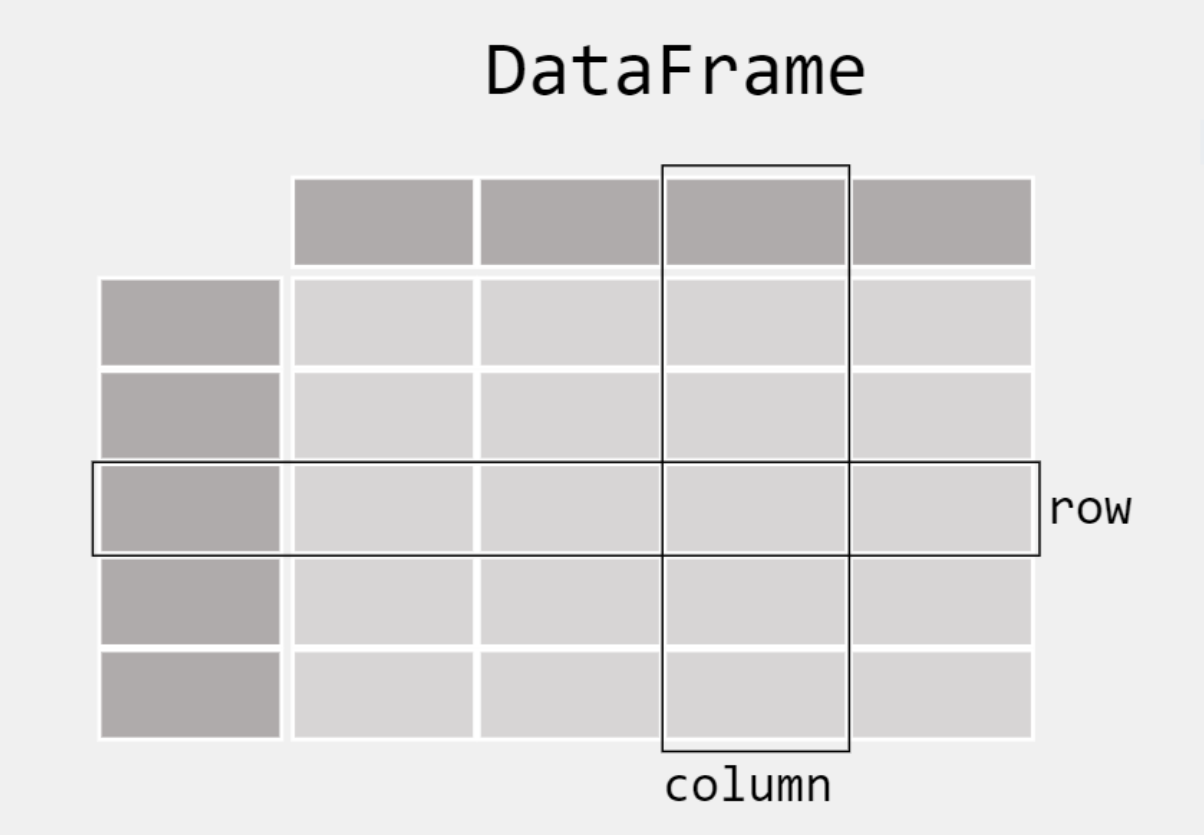
- ***strutture dati*** fondamentali
  - ***Series (1d)***
  - ***DataFrame (2d)***
    - ogni colonna di un DataFrame è una Series



Series

- una ***Series*** è un ***vettore mono-dimensionale*** i cui elementi sono etichettati con un ***index***
- similitudine con liste Python e array di Numpy
  - possibile accedere in sequenza agli elementi
- similitudine con i dizionari Python
  - accesso agli elementi tramite indice

DataFrame



DataFrame

- ***tabella*** di oggetti eterogenei
  - equivalente bidimensionale di una Series
- ***indici*** sia per le righe che per le colonne
  - ***index*** rappresenta le etichette delle righe
  - ***columns*** rappresenta le etichette delle colonne
- l'attributo ***shape*** descrive le dimensioni della tabella
- gni colonna di un DataFrame è una Series
- tutte le operazioni sulle Series possono essere applicate a colonne estratte da un DataFrame
- molte delle operazioni definite per le Series possono essere applicate direttamente su un DataFrame

iportazione / esportazione dati

importazione dati

- l'importazione dei dati è il primo passo in qualsiasi progetto di data science
- i file CSV (Comma Separated Value) sono uno degli standard di interscambio dati fra procedure diverse.
- la funzione read_csv () fornisce un modo estremamente semplice per importare i dati da un file CSV a un DataFrame

```
iris = pd.read_csv("iris.csv")
```