

Cosa contiene?

- un tipo di dato di nome `Series` che rappresenta un array *etichettato* (gli indici sono delle etichette) monodimensionale in grado di contenere qualsiasi tipo di dato,
- un tipo di dato di nome `DataFrame` che rappresenta una tabella con colonne di tipi potenzialmente differenti (`Series`), e
- un insieme *routine* che implementano operazioni su `Series` e `DataFrame`

- La documentazione della libreria si trova in
<https://pandas.pydata.org/>
- Per utilizzare la libreria: `import pandas as pd`



Costruire un oggetto Series

- `s = pd.Series(data=dati, index=etichette)`, dove dati è una collezione di valori e etichette (facoltativo) è una elenco di etichette.
- `s = pd.read_csv(...).squeeze()`, per utilizzare i dati in un file.

Costruire un oggetto DataFrame

- `s = pd.DataFrame(data=dati)`, dove dati, ad esempio, può essere:
 - un dict di Series,
 - un dict di list
- `s = pd.read_csv(...)`, per utilizzare i dati in un file di testo.
- `s = pd.read_excel(...)`, per utilizzare i dati in un file Excel.

Indicizzazione: Series

- `s[a]` oppure `s.loc[a]` dove l'argomento `a` può essere:
 - un'etichetta per selezionare l'elemento di etichetta `a`,
 - una lista di etichette per selezionare gli elementi le cui etichette corrispondono ai valori della lista `a`.
In sequenza booleana della stessa dimensione delle etichette per selezionare gli elementi corrispondenti ai valori True di `a`.
- `s.at[a]` è l'elemento con etichetta `a` di `s`,
- `s.iat[a]` è l'elemento di posizione `a` di `s`,
- `s.iloc[a]` dove l'argomento `a` può essere:
 - un intero per selezionare l'elemento di posizione `a`,
 - una lista di interi per selezionare gli elementi le cui posizioni corrispondono ai valori della lista `a`.
 - un oggetto slice (ad esempio `2:9`) per selezionare gli elementi le cui posizioni allo slice `a`
In sequenza booleana della stessa dimensione degli etichette per selezionare le righe corrispondenti ai valori True di `a`.

Indicizzazione: DataFrame

- `df.at[r, c]` e `df.iat[r, c]` per selezionare l'elemento di coordinate (r, c) (etichette/numeri),
- `df[a]` dove `a` può essere:
 - un'etichetta per selezionare la colonna di etichetta `a`,
 - una lista di etichette per selezionare le colonne le cui etichette corrispondono ai valori della lista `a`.
- `df.loc[a]` per indicizzare le righe,
- `df.loc[a, b]` per indicizzare separatamente le righe e le colonne,
- `df.iloc` per indicizzare con numeri interi,
- in `df.loc` e `df.iloc` può essere utilizzata un sequenza booleana per indicizzare le righe e colonne.

Alcuni attributi/metodi su oggetti Series

- name: nome della serie
- dtypes: tipo degli oggetti
- array: array dei valori
- size: numero di valori
- index: metodo per ottenere le etichette
- rename(): metodo per cambiare il nome
- astype(): metodo per cambiare il tipo
- add, mul, div, sub, abs, cumsum, cumprod
- sum, prod, mean, var

Alcuni attributi/metodi su oggetti DataFrame

- `columns`: etichette delle colonne
- `index`: etichette delle righe
- `dtypes`: tipi di dati delle colonne
- `size`: numero di valori
- `shape`: numero di (righe, colonne)
- `index`: metodo per ottenere le etichette
- `add, mul, div, sub, abs, cumsum, cumprod`
- `diff, sum, prod, mean, var`
- `merge, join, groupby`

