

ITS INFORMATION AND COMMUNICATIONS TECHNOLOGY Academy

NOME MODULO: IA
UNITÀ DIDATTICA: IA.1
Lezione 1
Stefano Puglia

Biennio 2023-2025



INDICE DEGLI ARGOMENTI

- Introduzione a DataFrame e Series di pandas
- Leggere dati in un DataFrame
- Selezione colonne
- Selezione righe
- “Affettare” un DataFrame
- “Filtrare” un DataFrame in base a criteri
- Scrivere un DataFrame in un file





```
import pandas as pd
```



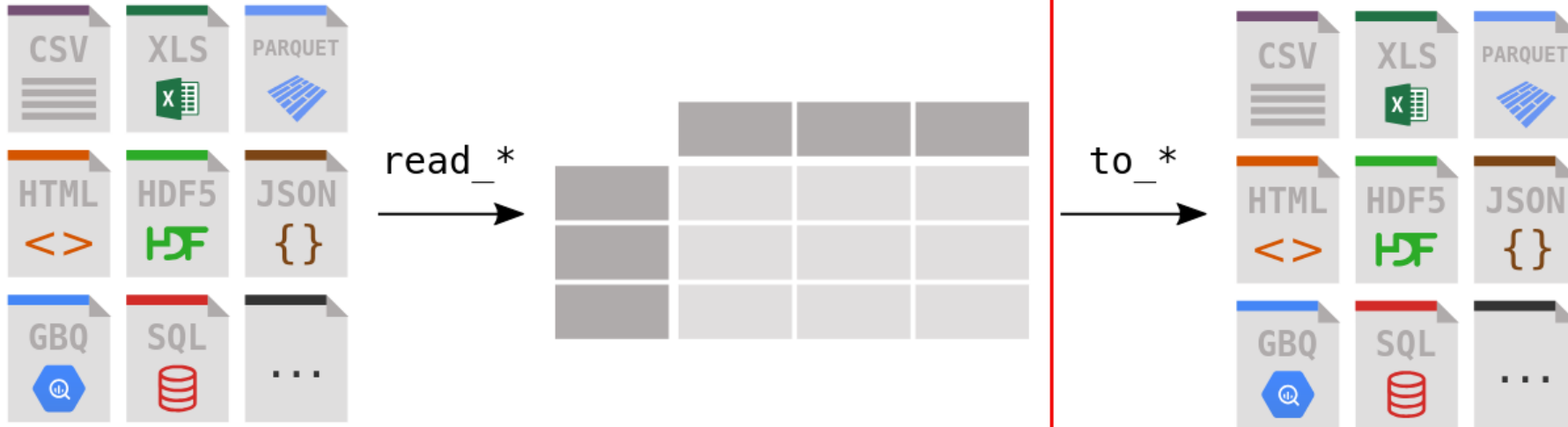
DataFrame

column

row



Creazione di un DataFrame



- Da un dizionario Python

```
df = pd.DataFrame(dict)
```

- Da un file (es. csv, Excel)

```
df = pd.read_csv("../dati/clean_data.csv")
```

- Da tabelle di un database

Creazione di un DataFrame

```
import pandas as pd

dict = {
    "Name": [
        "Braund, Mr. Owen Harris",
        "Allen, Mr. William Henry",
        "Bonnell, Miss. Elizabeth",
        "Taylor, Miss. Jane"
    ],
    "Age": [22, 35, 58, 55],
    "Sex": ["male", "male", "female", "female"],
    "Location": ["Rome", "London", "Berlin", "New York"],
}

df = pd.DataFrame(dict)
print("Un dataframe creato da un dizionario Python")
print(df)
print("\n")
```

```
In [3]: runfile('/home/stefano/Documents/Personal/Courses/
ITS_Academy/Lezioni_IA.1/Lezione1/codice/primeprovepandas.py',
wdir='/home/stefano/Documents/Personal/Courses/ITS_Academy/
Lezioni_IA.1/Lezione1/codice')
```

Un dataframe creato da un dizionario Python

	Name	Age	Sex	Location
0	Braund, Mr. Owen Harris	22	male	Rome
1	Allen, Mr. William Henry	35	male	London
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

```
In [4]:
```



Creazione di un DataFrame

```
import pandas as pd
```

```
students = pd.read_csv("../dati/clean_data.csv")  
print("Un dataframe creato da un file csv")  
print(students)  
print("\n")
```

Un dataframe creato da un file csv

	age	gender	screen_time_hours	...	math_score	science_score	is_healthy
0	15	M	2.0	...	75	69	False
1	14	M	4.8	...	73	75	False
2	15	F	6.8	...	74	70	False
3	12	F	3.5	...	80	76	False
4	12	F	6.4	...	75	78	False
..
495	15	F	7.7	...	81	78	False
496	13	M	5.0	...	72	75	False
497	14	M	2.8	...	75	61	False
498	14	F	6.2	...	74	57	False
499	15	F	8.0	...	77	65	False

[500 rows x 9 columns]



Informazioni su un DataFrame df

- `df.head()`, prime righe (default 5)
- `df.tail()`, ultime righe (default 5)
- `df.describe()`, statistiche di base
- `df.dtypes`, tipi di dato colonne
- `df.info()`, sommario tecnico
- `df.columns`, lista intestazioni colonne
- `df.index`, lista indici di riga
- `df.shape`, formato del DataFrame



Selezione colonne

Metodo veloce (solo colonne con nome “unico”)

- `df.column_name`, restituisce una Series

Metodo robusto (qualsiasi nome colonna, più colonne)

- `df['column']`, restituisce una Series
- `df[['column 1', 'column 2']]`, restituisce un DataFrame



Selezione colonne

```
print(df.Location)
print("\n")
print(df[['Age', 'Sex']])
```

Un dataframe creato da un dizionario Python

	Name	Age	Sex	Location
0	Braund, Mr. Owen Harris	22	male	Rome
1	Allen, Mr. William Henry	35	male	London
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

```
0    Rome
1    London
2    Berlin
3    New York
Name: Location, dtype: object
```

```
   Age  Sex
0   22  male
1   35  male
2   58 female
3   55 female
```



Selezione righe

Filtro per etichetta dell'indice/colonna

- `df.loc[label]`, restituisce una Series
- `df.loc[[label]]`, restituisce un DataFrame

Filtro per posizione dell'indice/colonna

- `df.iloc[index]`, restituisce una Series
- `df.iloc[[index]]`, restituisce un DataFrame



Selezione righe

```
print(df.iloc[[2]])  
print("\n")  
df.set_index('Name', inplace=True)  
print(df.loc[['Taylor, Miss. Jane']])
```

Un dataframe creato da un dizionario Python

	Name	Age	Sex	Location
0	Braund, Mr. Owen Harris	22	male	Rome
1	Allen, Mr. William Henry	35	male	London
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

	Name	Age	Sex	Location
2	Bonnell, Miss. Elizabeth	58	female	Berlin

	Age	Sex	Location
Name			
Taylor, Miss. Jane	55	female	New York



'Affettare' un DataFrame



'Affettare' un DataFrame

Selezione righe:

- `df.iloc[0:2]` → Righe 0 and 1 (esclude 2)
- `df.loc[0:2]` → Righe 0, 1, and 2 (include 2)

Selezione colonne:

- `df.iloc[0:3]` → Colonne in posizioni 0, 1, 2
- `df.loc['Name':'Sex']` → Colonne da Name a Sex



'Affettare' un DataFrame

```
print(df.iloc[0:2, 0:3])
print("\n")
print(df.loc[2:])
print("\n")
print(df.loc[0:2, 'Name': 'Sex'])
print("\n")
print(df.loc[0:2][['Age', 'Location']])
```

Un dataframe creato da un dizionario Python

	Name	Age	Sex	Location
0	Braund, Mr. Owen Harris	22	male	Rome
1	Allen, Mr. William Henry	35	male	London
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

	Name	Age	Sex
0	Braund, Mr. Owen Harris	22	male
1	Allen, Mr. William Henry	35	male

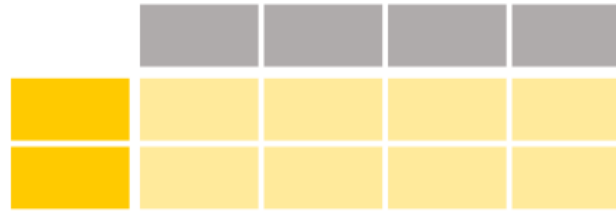
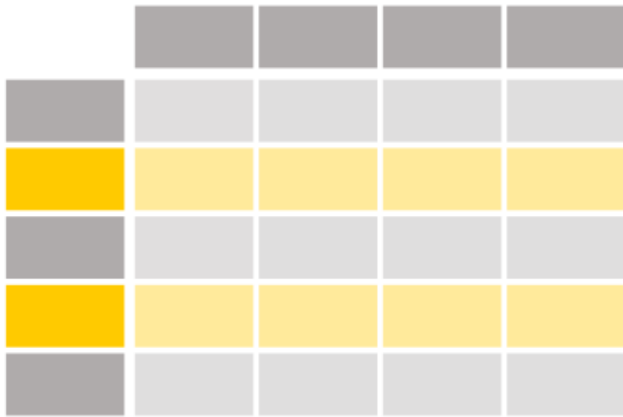
	Name	Age	Sex	Location
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

	Name	Age	Sex
0	Braund, Mr. Owen Harris	22	male
1	Allen, Mr. William Henry	35	male
2	Bonnell, Miss. Elizabeth	58	female

	Age	Location
0	22	Rome
1	35	London
2	58	Berlin



‘Filtrare’ un DataFrame



‘Filtrare’ un DataFrame

Definire ed applicare una condizione:

- `condition_1 = df['Age'] >= 40`
- `df[condition_1]`

Definire ed applicare più condizioni:

- `condition_2 = df['Location'] == 'New York'`
- `df[condition_1 & condition_2]`



‘Filtrare’ un DataFrame

```
simple_condition_1 = df['Age'] >= 40
print(simple_condition_1)
print("\n")
print(df[simple_condition_1])
print("\n")
simple_condition_2 = df['Location'].str.contains('b', case=False)
print(simple_condition_2)
print("\n")
print(df[simple_condition_1 & simple_condition_2])
```

Un dataframe creato da un dizionario Python

	Name	Age	Sex	Location
0	Braund, Mr. Owen Harris	22	male	Rome
1	Allen, Mr. William Henry	35	male	London
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

```
0    False
1    False
2     True
3     True
Name: Age, dtype: bool
```

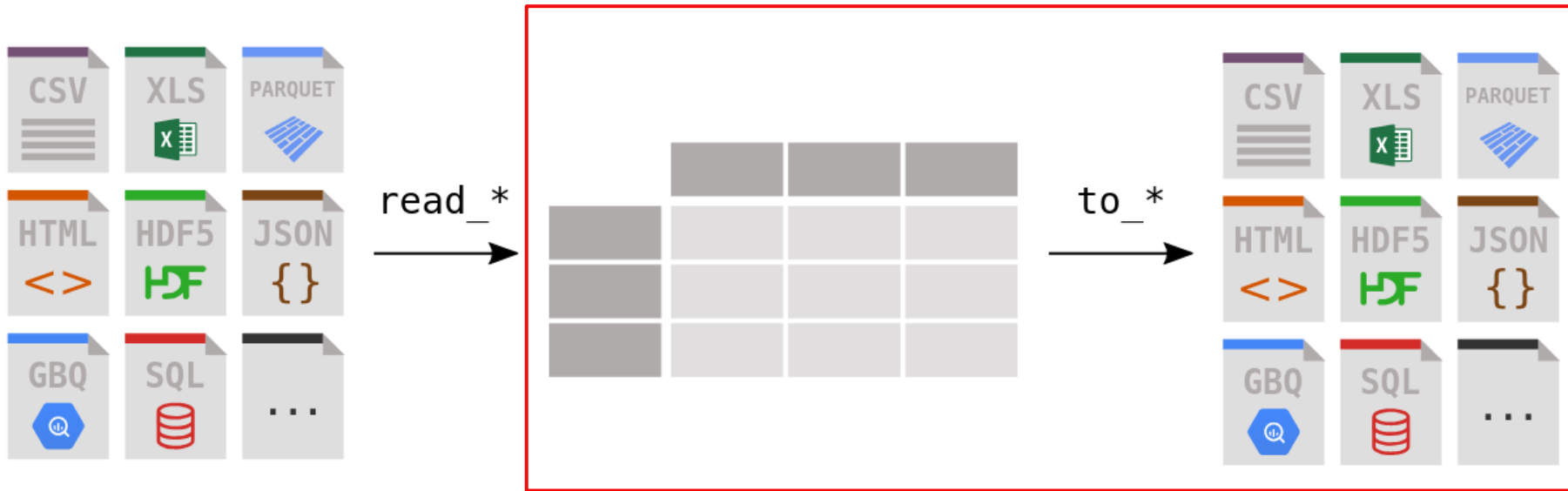
	Name	Age	Sex	Location
2	Bonnell, Miss. Elizabeth	58	female	Berlin
3	Taylor, Miss. Jane	55	female	New York

```
0    False
1    False
2     True
3    False
Name: Location, dtype: bool
```

	Name	Age	Sex	Location
2	Bonnell, Miss. Elizabeth	58	female	Berlin



Scrittura di un DataFrame



- In un dizionario Python

```
df.to_dict()
```

- In un file (es. csv, Excel)

```
df.to_csv("../dati/output.csv", index=False)
```

```
df.to_excel("titanic.xlsx", sheet_name="passengers", index=False)
```

- In una tabella di un database

RIFERIMENTI BIBLIOGRAFICI
- [click to pandas documentation](#)

