



DATAFRAME ATTRIBUTES

PYTHON for DATA SCIENCE



Jaume Boguñá

Dive into Python

Pandas DataFrame

Pandas is a **Python** library used for **data manipulation** and **analysis**.

Attributes and Underlying Data

I. Attributes

- ✓ Provide metadata and structure about the DataFrame.
 - ✓ `index`
 - ✓ `columns`
 - ✓ ...

II. Underlying Data

- ✓ Can be accessed and manipulated efficiently using NumPy and pandas functions.
 - ✓ `to_numpy()`
 - ✓ `memory_usage()`
 - ✓ ...



DataFrame Attributes

1. df.index

Returns the **index** (row labels) of the **DataFrame**.

```
import pandas as pd

people = {
    "Name": ["Jaume", "Paula", "David", "Berta"],
    "Age": [25, 34, 46, 40],
    "City": ["Barcelona", "Valencia", "Cadiz", "Bilbao"],
    "Profession": ["Engineer", "Doctor", "Artist", "Teacher"],
}

df = pd.DataFrame(people)

```

	Name	Age	City	Profession
0	Jaume	25	Barcelona	Engineer
1	Paula	34	Valencia	Doctor
2	David	46	Cadiz	Artist
3	Berta	40	Bilbao	Teacher

```
df.index
RangeIndex(start=0, stop=4, step=1)
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

2. `df.columns`

Returns the **column labels** of the **DataFrame**.

```
import pandas as pd
```

```
...
```

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

	Name	Age	City	Profession
0	Jaume	25	Barcelona	Engineer
1	Paula	34	Valencia	Doctor
2	David	46	Cadiz	Artist
3	Berta	40	Bilbao	Teacher

```
df.columns
```

```
Index(['Name', 'Age', 'City', 'Profession'], dtype='object')
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

3. `df.dtypes`

Return the **dtypes** in the **DataFrame**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

df.dtypes
name           object
age            int64
height        float64
is_student      bool
dtype: object
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

4. `df.select_dtypes()`

Return a subset of the DataFrame's columns based on the column dtypes.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

df.select_dtypes(include="bool")
is_student
0      True
1     False
2      True
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

5. `df.shape`

Return a **tuple** representing the **dimensionality** (rows, cols) of the **DataFrame**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)
name age height is_student
0 Marc 16 5.9 True
1 Blanca 14 6.0 False
2 Nadia 15 5.7 True

df.shape
(3, 4)
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

6. `df.size`

Returns the **total number of elements** in the **DataFrame** (rows * columns).

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

```

	name	age	height	is_student
0	Marc	16	5.9	True
1	Blanca	14	6.0	False
2	Nadia	15	5.7	True

```
df.size
12
```



DataFrame Attributes

7. `df.ndim`

Returns the **number of dimensions** of the **DataFrame**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)
name  age  height  is_student
0   Marc   16     5.9         True
1  Blanca   14     6.0        False
2   Nadia   15     5.7         True

df.ndim
2
```



DataFrame Attributes

8. `df.empty`

Returns **True** if the **DataFrame** is **empty** (has no elements), otherwise **False**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)
name age height is_student
0 Marc 16 5.9 True
1 Blanca 14 6.0 False
2 Nadia 15 5.7 True

df.empty
False
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

9. `df.axes`

Returns a **list** representing the **axes** (row and column labels) of the **DataFrame**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

df.axes
[RangeIndex(start=0, stop=3, step=1),
Index(['name', 'age', 'height',
'is_student'], dtype='object')]
```



Jaume Boguñá

Dive into Python

DataFrame Attributes

10. `df.info()`

Display **DataFrame** information (like data types, non-null values, ...)

```
import pandas as pd

planets = {
    'planet': ['Mercury', 'Jupiter', 'Neptune'],
    'distancefromsun (AU)': [0.39, 5.20, None],
    'type': ['Terrestrial', 'Gas Giant', 'Ice Giant']
}

df = pd.DataFrame(planets)

df.info()
RangeIndex: 3 entries, 0 to 2
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  -
0   planet                 3 non-null     object
1   distancefromsun (AU)   2 non-null     float64
2   type                   3 non-null     object
dtypes: float64(1), object(2)
memory usage: 200.0+ bytes
```



Jaume Boguñá

Dive into Python

Underlying Data

1. `df.to_numpy()`

Convert the **DataFrame** to a **NumPy** array.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

df.to_numpy()
[['Marc' 16 5.9 True]
 ['Blanca' 14 6.0 False]
 ['Nadia' 15 5.7 True]]
```



Jaume Boguñá

Dive into Python

Underlying Data

2. `df.memory_usage()`

Return the **memory usage** of each **column** in **bytes**.

```
import pandas as pd

teens = {
    "name": ["Marc", "Blanca", "Nadia"],
    "age": [16, 14, 15],
    "height": [5.9, 6.0, 5.7],
    "is_student": [True, False, True],
}

df = pd.DataFrame(teens)

df.memory_usage()
Index      128
name        24
age         24
height      24
is_student   3
dtype: int64
```



Jaume Boguñá

Dive into Python

Summary

DataFrame Attributes

index

columns

dtypes

select_dtypes()

shape

size

ndim

empty

axes

info()

Underlying Data

to_numpy()

memory_usage()



Jaume Boguñá

Dive into Python