

Dataframe Conversion API Misused

Context

In Pandas, `df.values()` can turn a DataFrame to a NumPy array.

Problem

`df.values()` has **an inconsistency problem**. With `.values()` it is unclear whether the returned value **would be the actual array**, some transformation of it, or one of the Pandas custom arrays.

Solution

When converting DataFrame to NumPy array, **it is better to use `df.to_numpy()` than `df.values()`**.

Existing Stage

Data Cleaning

Effect

Consistency & Error-prone

Example

```
Python
### NumPy & Pandas
import numpy as np
import pandas as pd
index = [1, 2, 3, 4, 5, 6, 7]
a = [np.nan, np.nan, np.nan, 0.1, 0.1, 0.1, 0.1]
b = [0.2, np.nan, 0.2, 0.2, 0.2, np.nan, np.nan]
c = [np.nan, 0.5, 0.5, np.nan, 0.5, 0.5, np.nan]
df = pd.DataFrame({'A': a, 'B': b, 'C': c}, index=index)
df = df.rename_axis('ID')
- arr = df.values
+ arr = df.to_numpy()
```