

[🏠](#) > [API reference](#) > [DataFrame](#) > [pandas.DataFrame](#)...

pandas.DataFrame.drop_duplicates

DataFrame.drop_duplicates(*subset=None*, *, *keep='first'*, *inplace=False*, *ignore_index=False*) [\[source\]](#)

Return DataFrame with duplicate rows removed.

Considering certain columns is optional. Indexes, including time indexes are ignored.

Parameters:

subset : *column label or sequence of labels, optional*

Only consider certain columns for identifying duplicates, by default use all of the columns.

keep : {'first', 'last', `False`}, default 'first'

Determines which duplicates (if any) to keep.

- 'first' : Drop duplicates except for the first occurrence.
- 'last' : Drop duplicates except for the last occurrence.
- `False` : Drop all duplicates.

inplace : *bool, default* `False`

Whether to modify the DataFrame rather than creating a new one.

ignore_index : *bool, default* `False`

If `True`, the resulting axis will be labeled 0, 1, ..., n - 1.

Returns:

DataFrame or None

DataFrame with duplicates removed or None if `inplace=True`.

➡ See also

[DataFrame.value_counts](#)

Count unique combinations of columns.

Examples

Consider dataset containing ramen rating.

```
>>> df = pd.DataFrame({
...     'brand': ['Yum Yum', 'Yum Yum', 'Indomie', 'Indomie', 'Indomie'],
...     'style': ['cup', 'cup', 'cup', 'pack', 'pack'],
...     'rating': [4, 4, 3.5, 15, 5]
... })
>>> df
```

	brand	style	rating
0	Yum Yum	cup	4.0
1	Yum Yum	cup	4.0
2	Indomie	cup	3.5
3	Indomie	pack	15.0
4	Indomie	pack	5.0

By default, it removes duplicate rows based on all columns.

```
>>> df.drop_duplicates()
   brand style  rating
0  Yum Yum   cup    4.0
2  Indomie   cup    3.5
3  Indomie  pack   15.0
4  Indomie  pack    5.0
```

To remove duplicates on specific column(s), use `subset`.

```
>>> df.drop_duplicates(subset=['brand'])
   brand style  rating
0  Yum Yum   cup    4.0
2  Indomie   cup    3.5
```

To remove duplicates and keep last occurrences, use `keep`.

```
>>> df.drop_duplicates(subset=['brand', 'style'], keep='last')
   brand style  rating
1  Yum Yum   cup    4.0
2  Indomie   cup    3.5
4  Indomie  pack    5.0
```

[Previous](#)

[pandas.DataFrame.drop](#)

[Next](#)

[pandas.DataFrame.duplicated](#)

[OVHcloud](#).

Built with the [PyData Sphinx Theme](#)

Created using [Sphinx](#) 7.2.6.

0.14.4.