







♠ > API reference > DataFrame > pandas.DataFrame

pandas.DataFrame

class pandas.DataFrame(data=None, index=None, columns=None, dtype=None, [source] copy=None)

Two-dimensional, size-mutable, potentially heterogeneous tabular data.

Data structure also contains labeled axes (rows and columns). Arithmetic operations align on both row and column labels. Can be thought of as a dict-like container for Series objects. The primary pandas data structure.

Parameters:

data: ndarray (structured or homogeneous), Iterable, dict, or DataFrame

Dict can contain Series, arrays, constants, dataclass or list-like objects. If data is a dict, column order follows insertion-order. If a dict contains Series which have an index defined, it is aligned by its index. This alignment also occurs if data is a Series or a DataFrame itself. Alignment is done on Series/DataFrame inputs.

If data is a list of dicts, column order follows insertion-order.

index: Index or array-like

Index to use for resulting frame. Will default to RangeIndex if no indexing information part of input data and no index provided.

columns: Index or array-like

Column labels to use for resulting frame when data does not have them, defaulting to RangeIndex(0, 1, 2, ..., n). If data contains column labels, will perform column selection instead.

dtype: dtype, default None

Data type to force. Only a single dtype is allowed. If None, infer.

copy: bool or None, default None

Copy data from inputs. For dict data, the default of None behaves like copy=True. For

is a dict containing one or more Series (possibly of different dtypes), copy=False will ensure that these inputs are not copied.

① Changed in version 1.3.0.

See also

```
DataFrame.from_records
```

Constructor from tuples, also record arrays.

```
DataFrame.from_dict
```

From dicts of Series, arrays, or dicts.

```
read_csv
```

Read a comma-separated values (csv) file into DataFrame.

```
read_table
```

Read general delimited file into DataFrame.

```
read_clipboard
```

Read text from clipboard into DataFrame.

Notes

Please reference the User Guide for more information.

Examples

Constructing DataFrame from a dictionary.

```
>>> d = {'col1': [1, 2], 'col2': [3, 4]}
>>> df = pd.DataFrame(data=d)
>>> df
    col1 col2
0    1    3
1    2    4
```

Notice that the inferred dtype is int64.

```
>>> df.dtypes
col1 int64
```

```
col2 int64 dtype: object
```

To enforce a single dtype:

```
>>> df = pd.DataFrame(data=d, dtype=np.int8)
>>> df.dtypes
col1   int8
col2   int8
dtype: object
```

Constructing DataFrame from a dictionary including Series:

```
>>> d = {'col1': [0, 1, 2, 3], 'col2': pd.Series([2, 3], index=[2, 3])}
>>> pd.DataFrame(data=d, index=[0, 1, 2, 3])
      col1    col2
0      0      NaN
1      1      NaN
2      2      2.0
3      3      3.0
```

Constructing DataFrame from numpy ndarray:

Constructing DataFrame from a numpy ndarray that has labeled columns:

Constructing DataFrame from Series/DataFrame:

Attributes

T	The transpose of the DataFrame.
	The transpose of the Data raine

at Access a single value for a row/column label pair.

attrs Dictionary of global attributes of this dataset.

axes Return a list representing the axes of the DataFrame.

columns The column labels of the DataFrame.

dtypes Return the dtypes in the DataFrame.

empty Indicator whether Series/DataFrame is empty.

flags Get the properties associated with this pandas object.

iloc	(DEPRECATED) Purely integer-location based indexing for selection by position.
index	The index (row labels) of the DataFrame.
<u>loc</u>	Access a group of rows and columns by label(s) or a boolean array.
ndim	Return an int representing the number of axes / array dimensions.
shape	Return a tuple representing the dimensionality of the DataFrame.
size	Return an int representing the number of elements in this object.
style	Returns a Styler object.
values	Return a Numpy representation of the DataFrame.

Methods

<u>abs</u> ()	Return a Series/DataFrame with absolute numeric value of each element.
add (other[, axis, level, fill_value])	Get Addition of dataframe and other, element-wise (binary operator <i>add</i>).
<pre>add_prefix (prefix[, axis])</pre>	Prefix labels with string <i>prefix</i> .
<pre>add_suffix (suffix[, axis])</pre>	Suffix labels with string <i>suffix</i> .
agg ([func, axis])	Aggregate using one or more operations over the specified axis.
aggregate ([func, axis])	Aggregate using one or more operations over the specified axis.
<pre>align (other[, join, axis, level, copy,])</pre>	Align two objects on their axes with the specified join method.
all ([axis, bool_only, skipna])	Return whether all elements are True, potentially over an axis.

<pre>any (*[, axis, bool_only, skipna])</pre>	Return whether any element is True, potentially over an axis.
apply (func[, axis, raw, result_type, args,])	Apply a function along an axis of the DataFrame.
<pre>applymap (func[, na_action])</pre>	(DEPRECATED) Apply a function to a Dataframe elementwise.
asfreq (freq[, method, how, normalize,])	Convert time series to specified frequency.
(where[, subset])	Return the last row(s) without any NaNs before <i>where</i> .
assign (**kwargs)	Assign new columns to a DataFrame.
astype (dtype[, copy, errors])	Cast a pandas object to a specified dtype dtype.
<pre>at_time (time[, asof, axis])</pre>	Select values at particular time of day (e.g., 9:30AM).
backfill (*[, axis, inplace, limit, downcast])	(DEPRECATED) Fill NA/NaN values by using the next valid observation to fill the gap.
between_time (start_time, end_time[,])	Select values between particular times of the day (e.g., 9:00-9:30 AM).
<pre>bfill (*[, axis, inplace, limit, limit_area,])</pre>	Fill NA/NaN values by using the next valid observation to fill the gap.
bool ()	(DEPRECATED) Return the bool of a single element Series or DataFrame.
boxplot ([column, by, ax, fontsize, rot,])	Make a box plot from DataFrame columns.
clip ([lower, upper, axis, inplace])	Trim values at input threshold(s).

<pre>combine (other, func[, fill_value, overwrite])</pre>	Perform column-wise combine with another DataFrame.
<pre>combine_first (other)</pre>	Update null elements with value in the same location in <i>other</i> .
<pre>compare (other[, align_axis, keep_shape,])</pre>	Compare to another DataFrame and show the differences.
<pre>convert_dtypes ([infer_objects,])</pre>	Convert columns to the best possible dtypes using dtypes supporting pd.NA.
copy ([deep])	Make a copy of this object's indices and data.
<pre>corr ([method, min_periods, numeric_only])</pre>	Compute pairwise correlation of columns, excluding NA/null values.
<pre>corrwith (other[, axis, drop, method,])</pre>	Compute pairwise correlation.
count ([axis, numeric_only])	Count non-NA cells for each column or row.
cov ([min_periods, ddof, numeric_only])	Compute pairwise covariance of columns, excluding NA/null values.
cummax ([axis, skipna])	Return cumulative maximum over a DataFrame or Series axis.
cummin ([axis, skipna])	Return cumulative minimum over a DataFrame or Series axis.
cumprod ([axis, skipna])	Return cumulative product over a DataFrame or Series axis.
cumsum ([axis, skipna])	Return cumulative sum over a DataFrame or Series axis.

div (other[, axis, level, fill_value])	Get Floating division of dataframe and other, element-wise (binary operator <i>truediv</i>).
divide (other[, axis, level, fill_value])	Get Floating division of dataframe and other, element-wise (binary operator <i>truediv</i>).
dot (other)	Compute the matrix multiplication between the DataFrame and other.
drop ([labels, axis, index, columns, level,])	Drop specified labels from rows or columns.
<pre>drop_duplicates ([subset, keep, inplace,])</pre>	Return DataFrame with duplicate rows removed.
droplevel (level[, axis])	Return Series/DataFrame with requested index / column level(s) removed.
dropna (*[, axis, how, thresh, subset,])	Remove missing values.
<pre>duplicated ([subset, keep])</pre>	Return boolean Series denoting
	duplicate rows.
eq (other[, axis, level])	
eq (other[, axis, level]) equals (other)	duplicate rows. Get Equal to of dataframe and other,
	duplicate rows. Get Equal to of dataframe and other, element-wise (binary operator eq). Test whether two objects contain the
equals (other)	duplicate rows. Get Equal to of dataframe and other, element-wise (binary operator eq). Test whether two objects contain the same elements. Evaluate a string describing operations
equals (other) eval (expr, *[, inplace])	duplicate rows. Get Equal to of dataframe and other, element-wise (binary operator eq). Test whether two objects contain the same elements. Evaluate a string describing operations on DataFrame columns. Provide exponentially weighted (EW)

ffill (*[, axis, inplace, limit, limit_area,])	Fill NA/NaN values by propagating the last valid observation to next valid.
fillna ([value, method, axis, inplace,])	Fill NA/NaN values using the specified method.
filter ([items, like, regex, axis])	Subset the dataframe rows or columns according to the specified index labels.
first (offset)	(DEPRECATED) Select initial periods of time series data based on a date offset.
<pre>first_valid_index ()</pre>	Return index for first non-NA value or None, if no non-NA value is found.
floordiv (other[, axis, level, fill_value])	Get Integer division of dataframe and other, element-wise (binary operator <i>floordiv</i>).
<pre>from_dict (data[, orient, dtype, columns])</pre>	Construct DataFrame from dict of array-like or dicts.
from_records (data[, index, exclude,])	Convert structured or record ndarray to DataFrame.
ge (other[, axis, level])	Get Greater than or equal to of dataframe and other, element-wise (binary operator <i>ge</i>).
get (key[, default])	Get item from object for given key (ex: DataFrame column).
groupby ([by, axis, level, as_index, sort,])	Group DataFrame using a mapper or by a Series of columns.
gt (other[, axis, level])	Get Greater than of dataframe and other, element-wise (binary operator gt).
head ([n])	Return the first <i>n</i> rows.

hist ([column, by, grid, xlabelsize, xrot,])	Make a histogram of the DataFrame's columns.
idxmax ([axis, skipna, numeric_only])	Return index of first occurrence of maximum over requested axis.
idxmin ([axis, skipna, numeric_only])	Return index of first occurrence of minimum over requested axis.
<pre>infer_objects ([copy])</pre>	Attempt to infer better dtypes for object columns.
info ([verbose, buf, max_cols, memory_usage,])	Print a concise summary of a DataFrame.
<pre>insert (loc, column, value[, allow_duplicates])</pre>	Insert column into DataFrame at specified location.
interpolate ([method, axis, limit, inplace,])	Fill NaN values using an interpolation method.
<pre>isetitem (loc, value)</pre>	Set the given value in the column with position <i>loc</i> .
<u>isin</u> (values)	Whether each element in the DataFrame is contained in values.
isna ()	Detect missing values.
isnull ()	DataFrame.isnull is an alias for DataFrame.isna.
<u>items</u> ()	Iterate over (column name, Series) pairs.
iterrows ()	Iterate over DataFrame rows as (index, Series) pairs.
<pre>itertuples ([index, name])</pre>	Iterate over DataFrame rows as namedtuples.
join (other[, on, how, Isuffix, rsuffix,])	Join columns of another DataFrame.

keys ()	Get the 'info axis' (see Indexing for more).
kurt ([axis, skipna, numeric_only])	Return unbiased kurtosis over requested axis.
kurtosis ([axis, skipna, numeric_only])	Return unbiased kurtosis over requested axis.
last (offset)	(DEPRECATED) Select final periods of time series data based on a date offset.
<pre>last_valid_index ()</pre>	Return index for last non-NA value or None, if no non-NA value is found.
le (other[, axis, level])	Get Less than or equal to of dataframe and other, element-wise (binary operator <i>le</i>).
1t (other[, axis, level])	Get Less than of dataframe and other, element-wise (binary operator <i>lt</i>).
map (func[, na_action])	Apply a function to a Dataframe elementwise.
mask (cond[, other, inplace, axis, level])	Replace values where the condition is True.
([axis, skipna, numeric_only])	Return the maximum of the values over the requested axis.
mean ([axis, skipna, numeric_only])	Return the mean of the values over the requested axis.
median ([axis, skipna, numeric_only])	Return the median of the values over the requested axis.
<pre>melt ([id_vars, value_vars, var_name,])</pre>	Unpivot a DataFrame from wide to long format, optionally leaving identifiers set.

<pre>memory_usage ([index, deep])</pre>	Return the memory usage of each column in bytes.
<pre>merge (right[, how, on, left_on, right_on,])</pre>	Merge DataFrame or named Series objects with a database-style join.
min ([axis, skipna, numeric_only])	Return the minimum of the values over the requested axis.
<pre>mod (other[, axis, level, fill_value])</pre>	Get Modulo of dataframe and other, element-wise (binary operator <i>mod</i>).
mode ([axis, numeric_only, dropna])	Get the mode(s) of each element along the selected axis.
<pre>mul (other[, axis, level, fill_value])</pre>	Get Multiplication of dataframe and other, element-wise (binary operator <i>mul</i>).
<pre>multiply (other[, axis, level, fill_value])</pre>	Get Multiplication of dataframe and other, element-wise (binary operator <i>mul</i>).
ne (other[, axis, level])	Get Not equal to of dataframe and other, element-wise (binary operator <i>ne</i>).
<pre>nlargest (n, columns[, keep])</pre>	Return the first <i>n</i> rows ordered by columns in descending order.
notna ()	Detect existing (non-missing) values.
notnull ()	DataFrame.notnull is an alias for DataFrame.notna.
<u>nsmallest</u> (n, columns[, keep])	Return the first <i>n</i> rows ordered by columns in ascending order.
nunique ([axis, dropna])	Count number of distinct elements in

pad (*[, axis, inplace, limit, downcast])	(DEPRECATED) Fill NA/NaN values by propagating the last valid observation to next valid.
<pre>pct_change ([periods, fill_method, limit, freq])</pre>	Fractional change between the current and a prior element.
pipe (func, *args, **kwargs)	Apply chainable functions that expect Series or DataFrames.
<pre>pivot (*, columns[, index, values])</pre>	Return reshaped DataFrame organized by given index / column values.
<pre>pivot_table ([values, index, columns,])</pre>	Create a spreadsheet-style pivot table as a DataFrame.
pop (item)	Return item and drop from frame.
(other[, axis, level, fill_value])	Get Exponential power of dataframe and other, element-wise (binary operator <i>pow</i>).
prod ([axis, skipna, numeric_only, min_count])	Return the product of the values over the requested axis.
<pre>product ([axis, skipna, numeric_only, min_count])</pre>	Return the product of the values over the requested axis.
quantile ([q, axis, numeric_only,])	Return values at the given quantile over requested axis.
<pre>query (expr, *[, inplace])</pre>	Query the columns of a DataFrame with a boolean expression.
radd (other[, axis, level, fill_value])	Get Addition of dataframe and other, element-wise (binary operator <i>radd</i>).
rank ([axis, method, numeric_only,])	Compute numerical data ranks (1 through n) along axis.

rdiv (other[, axis, level, fill_value])	Get Floating division of dataframe and other, element-wise (binary operator <i>rtruediv</i>).
reindex ([labels, index, columns, axis,])	Conform DataFrame to new index with optional filling logic.
<pre>reindex_like (other[, method, copy, limit,])</pre>	Return an object with matching indices as other object.
rename ([mapper, index, columns, axis, copy,])	Rename columns or index labels.
rename_axis ([mapper, index, columns, axis,])	Set the name of the axis for the index or columns.
<pre>reorder_levels (order[, axis])</pre>	Rearrange index levels using input order.
replace ([to_replace, value, inplace, limit,])	Replace values given in <i>to_replace</i> with <i>value</i> .
resample (rule[, axis, closed, label,])	Resample time-series data.
	nesample time series data.
reset_index ([level, drop, inplace,])	Reset the index, or a level of it.
<pre>reset_index ([level, drop, inplace,])</pre> <pre>rfloordiv (other[, axis, level, fill_value])</pre>	·
	Reset the index, or a level of it. Get Integer division of dataframe and other, element-wise (binary operator
rfloordiv (other[, axis, level, fill_value])	Reset the index, or a level of it. Get Integer division of dataframe and other, element-wise (binary operator rfloordiv). Get Modulo of dataframe and other,
rmod (other[, axis, level, fill_value]) rmod (other[, axis, level, fill_value])	Reset the index, or a level of it. Get Integer division of dataframe and other, element-wise (binary operator rfloordiv). Get Modulo of dataframe and other, element-wise (binary operator rmod). Get Multiplication of dataframe and other, element-wise (binary operator

rpow (other[, axis, level, fill_value])	Get Exponential power of dataframe and other, element-wise (binary operator <i>rpow</i>).
(other[, axis, level, fill_value])	Get Subtraction of dataframe and other, element-wise (binary operator <i>rsub</i>).
rtruediv (other[, axis, level, fill_value])	Get Floating division of dataframe and other, element-wise (binary operator <i>rtruediv</i>).
sample ([n, frac, replace, weights,])	Return a random sample of items from an axis of object.
<pre>select_dtypes ([include, exclude])</pre>	Return a subset of the DataFrame's columns based on the column dtypes.
<pre>sem ([axis, skipna, ddof, numeric_only])</pre>	Return unbiased standard error of the mean over requested axis.
<pre>set_axis (labels, *[, axis, copy])</pre>	Assign desired index to given axis.
<pre>set_flags (*[, copy, allows_duplicate_labels])</pre>	Return a new object with updated flags.
<pre>set_index (keys, *[, drop, append, inplace,])</pre>	Set the DataFrame index using existing columns.
<pre>shift ([periods, freq, axis, fill_value, suffix])</pre>	Shift index by desired number of periods with an optional time <i>freq</i> .
skew ([axis, skipna, numeric_only])	Return unbiased skew over requested axis.
<pre>sort_index (*[, axis, level, ascending,])</pre>	Sort object by labels (along an axis).
<pre>sort_values (by, *[, axis, ascending,])</pre>	Sort by the values along either axis.
squeeze ([axis])	Squeeze 1 dimensional axis objects into scalars.

stack ([level, dropna, sort, future_stack])	Stack the prescribed level(s) from columns to index.
std ([axis, skipna, ddof, numeric_only])	Return sample standard deviation over requested axis.
(other[, axis, level, fill_value])	Get Subtraction of dataframe and other, element-wise (binary operator <i>sub</i>).
<pre>subtract (other[, axis, level, fill_value])</pre>	Get Subtraction of dataframe and other, element-wise (binary operator <i>sub</i>).
sum ([axis, skipna, numeric_only, min_count])	Return the sum of the values over the requested axis.
swapaxes (axis1, axis2[, copy])	(DEPRECATED) Interchange axes and swap values axes appropriately.
<pre>swaplevel ([i, j, axis])</pre>	Swap levels i and j in a MultiIndex.
tail ([n])	Return the last <i>n</i> rows.
take (indices[, axis])	Return the elements in the given positional indices along an axis.
<pre>to_clipboard (*[, excel, sep])</pre>	Copy object to the system clipboard.
to_csv ([path_or_buf, sep, na_rep,])	Write object to a comma-separated values (csv) file.
to_dict ([orient, into, index])	Convert the DataFrame to a dictionary.
to_excel (excel_writer, *[, sheet_name,])	Write object to an Excel sheet.
to_excel_(excel_writer, *[, sheet_name,]) to_feather (path, **kwargs)	Write object to an Excel sheet. Write a DataFrame to the binary Feather format.

to_hdf (path_or_buf, *, key[, mode,])	Write the contained data to an HDF5 file using HDFStore.
to_html ([buf, columns, col_space, header,])	Render a DataFrame as an HTML table.
to_json ([path_or_buf, orient, date_format,])	Convert the object to a JSON string.
to_latex ([buf, columns, header, index,])	Render object to a LaTeX tabular, longtable, or nested table.
to_markdown ([buf, mode, index, storage_options])	Print DataFrame in Markdown-friendly format.
to_numpy ([dtype, copy, na_value])	Convert the DataFrame to a NumPy array.
to_orc ([path, engine, index, engine_kwargs])	Write a DataFrame to the ORC format.
to_parquet ([path, engine, compression,])	Write a DataFrame to the binary parquet format.
to_period ([freq, axis, copy])	Convert DataFrame from DatetimeIndex to PeriodIndex.
<pre>to_period ([freq, axis, copy]) to_pickle (path, *[, compression, protocol,])</pre>	
	to PeriodIndex. Pickle (serialize) object to file.
to_pickle (path, *[, compression, protocol,])	to PeriodIndex. Pickle (serialize) object to file. Convert DataFrame to a NumPy record
to_pickle (path, *[, compression, protocol,]) to_records ([index, column_dtypes, index_dtypes])	to PeriodIndex. Pickle (serialize) object to file. Convert DataFrame to a NumPy record array. Write records stored in a DataFrame to a
to_pickle (path, *[, compression, protocol,]) to_records ([index, column_dtypes, index_dtypes]) to_sql (name, con, *[, schema, if_exists,])	to PeriodIndex. Pickle (serialize) object to file. Convert DataFrame to a NumPy record array. Write records stored in a DataFrame to a SQL database. Export DataFrame object to Stata dta

<

to_xarray ()	Return an xarray object from the pandas object.
to_xml ([path_or_buffer, index, root_name,])	Render a DataFrame to an XML document.
transform (func[, axis])	Call func on self producing a DataFrame with the same axis shape as self.
<pre>transpose (*args[, copy])</pre>	Transpose index and columns.
truediv (other[, axis, level, fill_value])	Get Floating division of dataframe and other, element-wise (binary operator <i>truediv</i>).
truncate ([before, after, axis, copy])	Truncate a Series or DataFrame before and after some index value.
Previausvert (tz[, axis, level, copy]) DataFrame	Convert tz-aware axis to target til Next zone. pandas.DataFrame.index

© 2024, pandas via <u>NumFOCUS, Inc.</u> Hosted by <u>OVHcloud</u>.

Created using **Sphinx** 8.0.2.

Built with the <u>PyData Sphinx Theme</u> 0.14.4.