

变形

透视表

Pivot

一般情况下，数据在DataFrame会以stacked状态存放，pivot函数可以将某一列作为新的cols，
`DataFrame.pivot(self, index=None, columns=None, values=None)`

Reshape data (produce a "pivot" table) based on column values. Uses unique values from specified index / columns to form axes of the resulting DataFrame. This function does not support data aggregation, multiple values will result in a MultiIndex in the columns.

不允许values中出现重复的行列索引对

pivot_table

功能更多，速度不如pivot

`pandas.pivot_table(data, values=None, index=None, columns=None, aggfunc='mean', fill_value=None, margins=False, dropna=True, margins_name='All', observed=False)`

- aggfunc聚合函数，默认为mean，可以传入list
- margins 汇总边际状态，默认False
- index columnns values都可以为多级

crosstab

一种特殊的透视表，典型的用途：分组统计

计算两个或更多因子的简单交叉表，默认情况下为frequency

`pandas.crosstab(index, columns, values=None, rownames=None, colnames=None, aggfunc=None, margins=False, margins_name: str = 'All', dropna: bool = True, normalize=False)`

- values & aggfunc：分组对某些数据进行聚合操作，这两个参数必须成对出现
- normalize: bool, {'all', 'index', 'columns'}, or {0,1}, default False
Normalize by dividing all values by the sum of values.
If passed 'all' or True, will normalize over all values.
If passed 'index' will normalize over each row.
If passed 'columns' will normalize over each column.
If margins is True, will also normalize margin values.

其他变形方法

melt

可以看作pivot函数的逆操作

`pandas.melt(frame: pandas.core.frame.DataFrame, id_vars=None, value_vars=None, var_name=None, value_name='value', col_level=None)`

- id_vars:作为identifier variables的变量
- value_vars: 要unpivot的列

stack

Stack the prescribed level(s) from columns to index.

两个参数: lavel, dropna

unstack

Pivot a level of the (necessarily hierarchical) index labels.

哑变量与因子化

get_dummies

`pandas.get_dummies(data, prefix=None, prefix_sep='_', dummy_na=False, columns=None, sparse=False, drop_first=False, dtype=None)`

factorize

Encode the object as an enumerated type or categorical variable.
This method is useful for obtaining a numeric representation of an array when all that matters is identifying distinct values.

`pandas.factorize(values, sort: bool = False, na_sentinel: int = -1, size_hint: Union[int, NoneType] = None) → Tuple[numpy.ndarray, Union[numpy.ndarray, pandas.core.dtypes.generic.ABCIndex]]`