

Panda or panel data : (کتابخانه پایتون که برای دستکاری و تجزیه و تحلیل داده ها نوشته شده است . به ویژه، سافتمان داده و عملیات را برای دستکاری جداول عددی و سری های زمانی ارائه می دهد.)

}

Pandas یک کتابخانه قدرتمند در پایتون برای تجزیه و تحلیل داده های پویایی است که مجموعه ای از توابع و ابزارهای مفید را برای کار با داده های جدولی ارائه می دهد.

<https://pandas.pydata.org/>

سافتمان داده های پاندا :

سری : یک آرایه برپسب دار یک بعدی که داده ها را از هر نوع نگهداری می کند .

`s = pd.Series([1, 3, 5, np.nan, 6, 8])`

دیتا فریم : یک سافتار داده دو بعدی که داده ها را مانند یک آرایه دوبعدی یا یک جدول با سطرها و ستون ها نگهداری می کند.

`df = pd.DataFrame(data=a , (index for row) index= [0,1,2,3,4])`

1. read_csv:1

- برای خواندن داده ها از یک فایل CSV به یک DataFrame Pandas استفاده می شود.

- آرگومان های کلیدی:

○ filepath:مسیر فایل CSV

○ sep:م جداکننده ستون (پیش فرض: ",")

○ header:شماره ردیف حاوی نام ستون ها (پیش فرض: 0)

○ encoding:رمزگذاری کاراکتر فایل

2. head:

- برای نمایش ردیف های ابتدایی یک DataFrame استفاده می شود.

- آرگومان های کلیدی:

○ n:تعداد ردیف های نمایش داده شده (پیش فرض: 5)

3. tail:

- برای نمایش ردیف های انتهایی یک DataFrame استفاده می شود.

- آرگومان های کلیدی:

○ n:تعداد ردیف های نمایش داده شده (پیش فرض: 5)

4. info:

- اطلاعات کلی در مورد یک DataFrame مانند ابعاد، نام ستون ها، و انواع داده را نمایش می دهد.

5. describe:

- آمار توصیفی مانند میانگین، انحراف معیار، و مقادیر مداخل و مداکثر را برای هر ستون عددی مناسبه و نمایش می دهد.

6. loc:

- برای انتخاب زیرمجموعه ای از ردیف ها و ستون ها با استفاده از برپسب ها یا موقعیت های آنها استفاده می شود.

- آرگومان های کلیدی:

○ axis:معمور انتساب (0 برای ردیف، 1 برای ستون)

○ labels:برچسب های ردیف یا ستون (لیست یا آرایه)

○ loc:انتخاب بر اساس موقعیت (اعداد صمیع)

7. iloc:

● برای انتخاب زیرمجموعه ای از ردیف ها و ستون ها با استفاده از موقعیت های آنها استفاده می شود.

● آرگومان های کلیدی:

○ axis:معمور انتساب (0 برای ردیف، 1 برای ستون)

○ positions:موقعیت های ردیف یا ستون (اعداد صمیع یا لیست)

8. sort_values:

● برای مرتب سازی DataFrame بر اساس یک یا چند ستون استفاده می شود.

● آرگومان های کلیدی:

○ by:نام ستون یا ستون هایی که بر اساس آنها مرتب سازی انجام می شود

○ ascending:ترتیب صعودی (True) یا نزولی(False)

9. groupby:

● برای گروه بندی DataFrame بر اساس یک یا چند ستون و انجام عملیات آماری یا دستکاری داده ها بر روی هر گروه استفاده می شود.

● آرگومان های کلیدی:

○ by:نام ستون یا ستون هایی که بر اساس آنها گروه بندی انجام می شود

○ agg:تابع یا دیکشنری از توابع برای اعمال بر روی هر گروه

10. merge:

● برای ترکیب دو یا چند DataFrame بر اساس ستون های مشترک استفاده می شود.

● آرگومان های کلیدی:

○ left_on:نام ستون در DataFrame سمت چپ برای ترکیب

○ right_on:نام ستون در DataFrame سمت راست برای ترکیب

○ how:نوع پیوستن (مانند "inner", "outer", "left", "right")

11. fillna:

● مقادیر گمشده (NaN) را در DataFrame با مقادیر دلخواه جایگزین می کند.

● آرگومان های کلیدی:

○ value:مقدار جایگزینی برایNaN

○ method:روش جایگزینی (مانند "ffill", "bfill", "mean", "median")

12. dropna:

● ردیف هایی را که حاوی مقادیر گمشده (NaN) هستند از DataFrame حذف می کند.

● آرگومان های کلیدی:

○ axis:معمور حذف (0 برای ردیف، 1 برای ستون)

○ how: شرط حذف (مانند "all", "any")

13. reset_index:

- ستون شاخص را از DataFrame حذف می کند یا آن را به عنوان یک ستون داده جدید تبدیل می کند.

14. set_index:

- ستون دلفواه را به عنوان ستون شاخص DataFrame تنظیم می کند.
- آرگومان های کلیدی:
 - col: ستونی که به عنوان شاخص استفاده می شود
 - drop: آیا ستون شاخص اصلی را بعد از تنظیم حذف کند (پیش فرض True):

15. apply:

- یک تابع را به هر ردیف یا ستون DataFrame اعمال می کند و نتیجه را به عنوان یک DataFrame جدید برمی گرداند.
- آرگومان های کلیدی:
 - func: تابعی که باید اعمال شود
 - axis: محور اعمال (0 برای ردیف، 1 برای ستون)
 - result_type: نوع داده DataFrame خروجی

16. plot:

- برای ایجاد نمودارهای مختلف مانند نمودار خطی، هیستوگرام، و نمودار پراکندگی از DataFrame استفاده می شود.
- آرگومان های کلیدی:
 - kind: نوع نمودار (مانند "line", "bar", "scatter")
 - x: نام ستون محور X
 - y: نام ستون محور Y

17. to_csv:

- DataFrame را به یک فایل CSV ذخیره می کند.
- آرگومان های کلیدی:
 - filepath: مسیر فایل CSV
 - sep: جداکننده ستون
 - header: آیا نام ستون ها را در ردیف اول شامل کند

18. to_sql:

- DataFrame را به یک جدول پایگاه داده ذخیره می کند.
- آرگومان های کلیدی:
 - name: نام جدول در پایگاه داده
 - con: اتصال پایگاه داده
 - if_exists: if_exists: نحوه مدیریت وجود جدول قبلی (مانند "fail", "replace", "append")

19. merge_asof:

- دو DataFrame را بر اساس یک ستون زمانبندی شده با نزدیکترین مطابقت زمانی ادغام می کند.

- آرگومان های کلیدی:

- left_on: ستون زمانبندی شده در DataFrame سمت چپ

- right_on: ستون زمانبندی شده در DataFrame سمت راست

- direction: جهت جستجوی مطابقت (مانند "left", "right", "nearest")

20. resample:

- DataFrame را بر اساس یک قاعده فرکانس (مانند "D", "H", "M") به گروه های زمانی تبدیل می کند و میانگین، مجموع یا سایر آمار را برای هر گروه مناسبه می کند.

- آرگومان های کلیدی:

- rule: قاعده فرکانس نمونه گیری

- agg: تابع یا دیکشنری از توابع برای اعمال بر روی هر گروه

اینها فقط تعدادی از توابع ضروری Pandas هستند که باید بدانید Pandas. مجموعه ای گسترده از ابزارهای قدرتمند برای تجزیه و تحلیل داده ها را ارائه می دهد که می تواند به شما در مل طیف وسیعی از مشکلات مربوط به داده کمک کند.

{