

بنام خدا

موضوع پروژه: کتابخانه pandas

Pandas یک کتابخانه متن باز با کارایی بسیار بالا، ساختاری با قابلیت استفاده آسان و ابزار هایی برای

داده و تحلیل برای برنامه نویسی در زبان پایتون برای کاربر فراهم میکند

پانداس یک کتابخانه فوق العاده قوی برای تحلیل داده و بصری سازی داده هاست .

کاربران این کتابخانه در سال های 2014 تا 2018 از 5 میلیون به 10 میلیون نفر افزایش پیدا کرد و در

زمان کنونی از آن برای کارهای مربوط به علم داده استفاده میشود.

استفاده از پانداس مزایایی دارد مثل 1. پانداس میتواند داده ها را با استفاده از ساختارهای series و

dataframe به قالبی برای تحلیل داده مناسب تبدیل کند

2. پانداس شامل تعداد زیادی متد برای پالایش مناسب داده ها است

3. پانداس دارای ابزار های متعددی برای عملیات ورودی و خروجی است و میتواند داده ها را از

فرمت های گوناگون شامل ms excel, tsv, csv بخواند

برای نصب ماژول پانداس در قسمت cmd ویندوز عبارت pip install pandas را مینویسیم

تا کتابخانه نصب شود.

ساختار داده ای پانداس از دو نوع series و dataframe برای ذخیره سازی داده ها تشکیل شد

ه

Series: یک series مثل یک آرایه یک بعدی است و میتواند هر نوع داده ای را ذخیره کند. مقادیری که

در series قرار میگیرند قابل تغییر هستند.

اولین عنصر در series اندیس صفر میگیرد یعنی اگر ما هر تعداد عنصر داشتیم آن را منهای یک

میکنیم بعد تعداد اندیس ها را متوجه میشویم .

گرامر فراخوانی کتابخانه پانداس به شکل زیر است:

Import pandas as pd

برای ساخت و نمایش series از متد pd.series() استفاده میشود. گرامر به شکل زیر است:

```
Series1= pd.series([2,4,6,8])
```

```
0    2
```

```
1    4
```

```
2    6
```

```
3    8
```

نتیجه کد به شکل روبه رواست<<<

میبینیم که دو ستون ایجاد شده. ستون اول که با صفر شروع شده ستون اندیس هاست و ستون دوم که با 2 شروع شده مقادیری هستند که ما وارد کردیم

Dataframe: ساختار دیتافریم را میتوان بصورت یک جدول دانست. دیتا فریم ها داده ها را در سطر ها و ستون ها سازماندهی میکنند و از انها ساختاری دو بعدی میسازند. ستون ها میتوانند شامل انواع زیادی از مقادیر باشند و اندازه دیتا فریم هم قابل تغییر است پس کاربر میتواند ان را ویرایش کند. برای ایجاد یک دیتافریم متوانیم از اول شروع کنیم و ارایه های نام پای را به دیتا فریم تبدیل کنیم.

گرامر تعریف یک دیتافریم به شکل زیر میباشد:

```
Pandas.dataframe(data,index,column,dtype,copy)
```

Data: داده ای که از ورودی دریافت و میتواند به شکل لیست یا مقدار ثابت ویا دیکشنری باشد

Index: برای برچسب دهی داده ها استفاده میشود که اگر توسط کاربر مقدار نگیرد بصورت

پیشفرض از عدد صفر تا x-1 ادامه میابد که x تعداد داده های ماست

Columns: برای برچسب دهی به ستون ها استفاده میشود اگر کاربر به ان مقدار ندهد بصورت

خودکار از صفر شروع و تا x-1 که xتعداد داده هاست را میگرد

Dtype: نوع ستون ها را مشخص میکند

Copy: اگر مقدار ان **true** باشد از داده های برنامه نسخه کپی ایجاد میکند اما مقدار پیشفرض ان

همیشه **false** است

کار با سطر و ستون در دیتا فریم

کاربر میتواند روی دیتا فریم ها عملیاتی مانند انتخاب. اضافه کردن. حذف و تغییر نام سطر ها و ستون ها را انجام دهد

انتخاب ستون ها: برای انتخاب یک ستون میتوان از نام آن استفاده کرد
انتخاب سطرها: سطر ها را میتوان با دستور loc,iloc,ix انتخاب کرد و برای انتخاب باید از اندیس
سطر ها استفاده کرد

داده های csv

Csv یک فایل متنی با مقادیری هستند که بوسیله کاما از یکدیگر جدا میشوند این نوع فایل غالباً شناخته شده است و برای استفاده بیشتر از این استفاده میکنیم.
کاربر میتواند با کتابخانه پانداس یک فایل csv را بصورت کامل و یا بخشی از بخواند و استفاده کند
در csv میتوان داده ها را کپی کرد و یا در نرم افزار notepad چسباند و ذخیره کرد

وارد کردن داده های اکسل

پانداس از تابع read_excel استفاده میکند که میتوان با آن داده های اکسل را خواند.
پس از فراخوانی تابع read_excel ، نام فایل به عنوان آرگومان به آن پاس داده
میشود read_excel. برای باز کردن/بارگذاری فایل و سپس، تجزیه داده ها مورد استفاده قرار
میگیرد. همانطور که از مثال پیشین مشهود است، تابع print() به کاربر کمک می کند تا محتوای
دیتافریم را نمایش دهد.

پیش پردازش داده ها

دستکاری یا پیش پردازش داده برای آماده کردن آنها برای مرحله بعدی پروژه انجام میشود
فرایند پیش پردازش ها شامل ادغام کردن... گروه بندی...الحاق کردن هستند

ادغام کردن

کتابخانه پانداس این امکان را برای کاربر فراهم می کند که اشیای دیتافریم را با تابع merge() به
یکدیگر متصل کنند. در ادامه، دو دیتافریم ساخته و روش ادغام کردن آن ها با یکدیگر نمایش داده
شده است.

گروه بندی در پانداس

پانداس برای کشف و سازماندهی داده های جدولی که حجم زیادی دارند استفاده میشود. کاربران برای انجام تحلیل بیشتر به سازمان دهی دیتا فریم ها در زیر گروه های مختلفی نیاز دارد الحاق کردن اضافه کردن یک مجموعه از داده ها به یک مجموعه داده دیگر به وسیله فراخوانی تابع `concat()`

آمار توصیفی

با استفاده از تابع `describe()` ، آمار توصیفی برای ستون های عددی ارائه می شود، اما ستون های حاوی کاراکتر توسط این تابع در نظر گرفته نمی شوند. مرتب سازی عنصرهای دیتا فریم در پانداس کاربر به دو صورت صعودی و نزولی میتواند مرتب سازی کند. پارامترهای مرتب سازی شامل:

By: نام یک یا مجموعه ای از ستون ها برای مرتب کردن داده بر مبنای ستون هاست
Axis: عدد 1 یا `column` برای ستون ها و عدد 0 یا `index` برای سطر ها استفاده میشود
Ascending: یک مقدار بولین است که اگر مقدار آن برابر با `true` باشد صعودی و اگر مقدار آن برابر با `false` باشد مرتب سازی نزولی است
Inplace: یک مقدار بولین است که اگر مقدار آن `true` باشد در انتقال دیتا فریم به تابع تغییراتی اعمال میشود
kind: مقدار رشته ای است که میتواند مقادیر `'quicksort'`، `'mergesort'`، `'heapsort'` را بگیرد مرتب سازی بر اساس الگوریتم آن انجام شود
Na_position: یکی از دو مقدار رشته ای `last` یا `first` را بگیرد و بر اساس آن مشخص کند که مکان مقادیر `nan` یا همان خالی بعد از مرتب سازی در اول جدول یا آخر جدول قرار گیرد. دیتا فریم برگشت داده شده از تابع مرتب سازی همان ابعاد دیتا فریم اولیه را دارد

<https://7learn.com/blog/pandas-library-tutorial>

<https://blog.faradars.org/pandas-from-zero-to-hero>

/