

[Python pandas] DataFrame, Series의 행, 열 개수 세기
Python 분석과 프로그래밍/Python 데이터 전처리 2019. 7. 3. 23:49

이번 포스팅에서는 Python pandas의 DataFrame, Series의 행, 열의 개수를 세는 방법(how to count the number of rows and columns of pandas DataFrame and Series)을 소개하겠습니다.

간단한 것들이고, 이미 소개한 것들이긴 한데요, Stackoverflow에 깔끔하게 유형별로 정리한 표가 있어서 옮겨보았습니다.

Series의 행 개수를 셀 때 s.size와 같이 뒤에 ()가 없다는 것 조심해야 합니다.그리고 **> count()**는 **Null 값이 아닌 행(count Non-null rows)만 세며, > size()**는 **Null 값인 행도 모두 포함해서 행(size of all rows)을 센다**는 것도 유념하면 좋겠습니다.

구분	pandas DataFrame (df)	pandas Series (s)
행 개수 세기 (row count)	len(df) df.shape[0] len(df.index)	len(s) s.size len(s.index)
열 개수 세기 (column count)	df.shape[1] len(df.columns)	N/A
Null 값이 아닌 행 개수 세기 (Non-null row count)	df.count()	s.count()
그룹 별 행 개수 세기 (Row count per group)	df.groupby(...).size()	s.groupby(...).size()
그룹 별 Null 값이 아닌 행 개수 세기 (Non-null row count per group)	df.groupby(...).count()	s.groupby(...).count()

간단한 예제를 아래에 소개합니다.
numpy와 pandas 라이브러리 불러오고, DataFrame과 Series 데이터셋 만들어보겠습니다.

```
import numpy as np
import pandas as pd
```

```
df = pd.DataFrame({'grp': ['A', 'A', 'A', 'B', 'B', 'B', 'C', 'C', 'C'],
                   'val': [1, 2, np.nan, 4, np.nan, np.nan, 7, 8, 9]})
```

```
In [02]: df
Out[02]:
grp val
0 A 1.0
1 A 2.0
2 A NaN
3 B 4.0
4 B NaN
5 B NaN
6 C 7.0
7 C 8.0
8 C 9.0
```

```
s = pd.Series([1, 2, np.nan, 4, np.nan, np.nan, 7, 8, 9])
In [03]: s
Out[03]:
0 1.0
1 2.0
2 NaN
3 4.0
4 NaN
5 NaN
6 7.0
7 8.0
8 9.0
dtype: float64
```

구분	DataFrame (df)	Series (s)
행의 개수 세기 (Row count)	In [04]: len(df) Out[04]: 9 In [05]: df.shape[0] Out[05]: 9 In [06]: len(df.index) Out[06]: 9	In [08]: len(s) Out[08]: 9 In [09]: s.size Out[09]: 9 In [10]: len(s.index) Out[10]: 9
열의 개수 세기 (Column count)	In [11]: df.shape[1] Out[11]: 2 In [12]: len(df.columns) Out[12]: 2	N/A
Null 이 아닌 행의 개수 세기 (Non-null row count, ignore NaNs)	In [13]: df.count() Out[13]: grp 9 val 6 dtype: int64 In [14]: df['val'].count() Out[14]: 6	In [15]: s.count() Out[15]: 6
그룹 별 행의 개수 세기 (Row count per group)	In [16]: df.groupby('grp').size() Out[16]: grp A 3 B 3 C 3 dtype: int64	In [17]: s.groupby(df.grp).size() Out[17]: grp A 3 B 3 C 3 dtype: int64
그룹별 Null 이 아닌 행의 개수 세기 (Non-null row count per group)	In [18]: df.groupby('grp').count() Out[18]: val grp A 2 B 1 C 3	In [19]: s.groupby(df.grp).count() Out[19]: grp A 2 B 1 C 3 dtype: int64

많은 도움이 되었기를 바랍니다.

20년 오렌노하우기술로 최상의 웹사이트 품질티보장, 디자인전문기업 웹개발, 하이드리드앱개발.

열기

9

구독하기

1111

카카오톡

트위터

페이스북

좋아요

친구들이 무엇을 좋아하는지 알아보면 기쁩니다

'Python 분석과 프로그래밍 > Python 데이터 전처리' 카테고리의 다른 글		
[Python pandas] DataFrame을 정렬한 후에, 그룹별로 상위 N개 행 선택하기 (sort DataFrame by value and select top N rows by group) (2)	2019.07.13	
[Python pandas] pivot_table() 할 때 DataError: No numeric types to aggregate 에러 대처방법 aggfunc="first" (0)	2019.07.11	
[Python pandas] DataFrame, Series의 행, 열 개수 세기 (0)	2019.07.03	
[Python pandas] DataFrame의 문자열 칼럼을 분할하여 일부분으로 새로운 칼럼 만들기 (0)	2019.07.01	
[Python] 텍스트를 단어 단위로 파싱해서 One-hot encoding 하기 (parsing text and one-hot encoding at word-level) (0)	2019.05.22	
[Python] 텍스트 파일 읽어와서 숫자형 데이터 표준화하기 (reading csv or text file, standardizing or normalizing of numeric data) (0)	2019.05.21	

Posted by R_Friend R_Friend

df.count(), df.group(...).size(), len(), python pandas DataFrame의 그룹별 Null 값이 아닌 행 개수 세기, python pandas DataFrame의 그룹별 행 개수 세기, python pandas DataFrame의 열 개수 세기, python pandas DataFrame의 행 개수 세기, python pandas Series의 그룹별 Null 값이 아닌 행 개수 세기, python pandas Series의 그룹별 행 개수 세기, python pandas Series의 열 개수 세기

트랙백 0개, 댓글 0개가 달렸습니다

http://: 홈페이지 ☐ 비밀번호

댓글 달기

R Friend

http://rfriend.tistory.com

R, Python 분석과 프로그래밍, 통계, Machine Learning, Greenplum, PostgreSQL, Hive, 분석으로 세상보기, 독서일기 R_Friend

➡ 관리자 ✎ 글쓰기

검색

카테고리

- 분류 전체보기 (508)
- R 분석과 프로그래밍 (234)
- Python 분석과 프로그래밍 (184)
- Greenplum and PostgreSQL Da.. (37)
- Hive (11)
- Deep Learning with Tensorf.. (7)
- 분석으로 세상보기 (20)
- 독서일기 (15)

태그목록

HIVE Python Graphical Parameters
R graphics 파이썬 dplyr 패키지 데이
터 변환 PL/R greenplum db 데이터 재
구조화 기계학습 dataframe 판다스
min() Cluster Analysis max() 선형
대수 군집분석 넘파이 low level graphic
function postgresql db 다변량 그래프
낮은 수준의 그래프 함수 Greenplum
Database ggplot2 dplyr package
numpy PostgreSQL pandas

RSS FEED

Total : 4,005,473

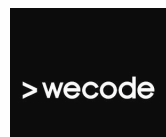
Today : 524 Yesterday : 9,911

>wecode

나의 커리어에 대해
진지하게 고민한다면,
일두 위주의
코딩 부트캠프

10월 개강
14기 모집중

90일, 인생이 변하는 기간



프로그래밍을 전혀
몰라도 괜찮아요, 국내
유일 오프라인 코딩
부트캠프



기획형 홈페이지제작 매스티지

masstige.biz

1:1 전담 상담매니저
홍대디자인

20년 오렌노하우기술로
최상의 웹사이트
퀄리티보장,
디자인전문기업 웹개발,
하이드리드앱개발.

열기