



데이터 요약



판다스로 기본통계하기



Summarize Data_데이터 요약하기

df['w'].value_counts()

Count number of rows with each unique value of variable

len(df)

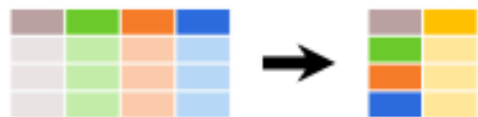
of rows in DataFrame.

df['w'].nunique()

of distinct values in a column.

df.describe()

Basic descriptive statistics for each column (or GroupBy)



pandas provides a large set of **summary functions** that operate on different kinds of pandas objects (DataFrame columns, Series, GroupBy, Expanding and Rolling (see below)) and produce single values for each of the groups. When applied to a DataFrame, the result is returned as a pandas Series for each column. Examples:

sum()

Sum values of each object.

count()

Count non-NA/null values of each object.

median()

Median value of each object.

quantile([0.25, 0.75])

Quantiles of each object.

apply(function)

Apply function to each object.

min()

Minimum value in each object.

max()

Maximum value in each object.

mean()

Mean value of each object.

var()

Variance of each object.

std()

Standard deviation of each object.

판다스로 기본통계하기



Summarize Data_데이터 요약하기

01

value_counts()

- 시리즈 객체의 고유 값(Unique Value) 개수를 세는 데 사용함

02

nunique()

- 명확한 유일 값(Distict Unique)의 개수 확인

03

describe() *(include =/ exclude =)

- 산술(숫자) 데이터를 갖는 열에 대한 주요 기술 통계 정보(평균, 표준편차, 최대값, 최소값, 중간값) 요약해서 출력해줌

판다스로 기본통계하기



Summarize Data_데이터 요약하기

sum()	합계	max()	최대값
count()	개수	var()	분산
median()	중간 값, 중위수	std()	표준편차
mean()	평균	corr()	두 열 간의 상관계수
min()	최소 값		



이론 영상 후 실습 영상 제시



Apply 함수

01 apply()

- 인자로 전달하는 매핑 함수에 시리즈의 모든 원소를 하나씩 입력하고 함수의 리턴 값을 돌려 받음
- 시리즈 원소의 개수 만큼 리턴 값을 받아서 같은 크기의 시리즈 객체로 반환함

02 lambda()

- 사용자 직접 만든 함수



이론 영상 후 실습 영상 제시





결측치 다루기



dropna(), fillna(), isnull(), notnull()



결측치 다루기

Handling Missing Data

df.dropna()

Drop rows with any column having NA/null data.

df.fillna(value)

Replace all NA/null data with value.

01 dropna()

- NA/ null 값이 존재 시, 제거(Drop)
- axis = 0 : 행, axis =1 : 열
- how = 'all', how = 'any'

02 fillna()

- NA/ null 값을 대체

dropna(), fillna(), isnull(), notnull()



결측치 다루기

Handling Missing Data

df.dropna()

Drop rows with any column having NA/null data.

df.fillna(value)

Replace all NA/null data with value.

03

isnull() / notnull()

- `isnull().sum()` : 결측치가 몇 개인지 확인
 - 대량의 데이터 분석 시 유용함

<출처 : <https://dataitgirls2.github.io/10minutes2pandas/>>



dropna(), fillna(), isnull(), notnull()

이론 영상 후 실습 영상 제시

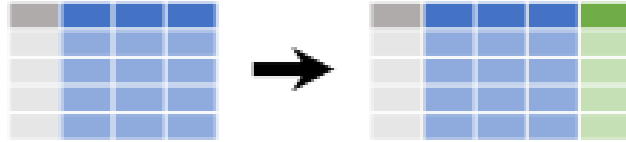




새로운 컬럼 만들기



Assign으로 새로운 컬럼 만들기



```
df.assign(Area=lambda df: df.Length*df.Height)
```

Compute and append one or more new columns.

```
df['Volume'] = df.Length*df.Height*df.Depth
```

Add single column.

assign()

새로운 컬럼을 데이터프레임에 배정(Assign)하는 함수



Assign으로 새로운 컬럼 만들기

이론 영상 후 실습 영상 제시



Qcut으로 범주화하기

```
pd.qcut(df.col, n, labels=False)
```

Bin column into n buckets.



pandas provides a large set of **vector functions** that operate on all columns of a DataFrame or a single selected column (a pandas Series). These functions produce vectors of values for each of the columns, or a single Series for the individual Series. Examples:

```
max(axis=1)
```

Element-wise max.

```
min(axis=1)
```

Element-wise min.

```
clip(lower=-10, upper=10) abs()
```

Trim values at input thresholds Absolute value.

01 qcut()

- Bin Column을 n개의 Bucket으로 변화시킴
- 숫자형 데이터를 n개의 범주로 구분할 때 사용함

Qcut으로 범주화하기

02 `max(axis = 1)/ min(axis = 1)`

- `axis = 1` : 열(Column)
- 열에서 가장 큰 값/ 작은 값을 가져올 때 사용함

03 `clip()`

- 임계치 값을 지정해서 값을 변화시켜줄 때 사용함

04 `abs()`

- 절대값



Qcut으로 범주화하기

이론 영상 후 실습 영상 제시



과목명	데이터시각화	주차명	03. 데이터 요약, 결측치 다루기, 새로운 컬럼 만들기	페이지번호	05_01_03
-----	--------	-----	---------------------------------	-------	----------

주차명

03. 데이터 요약, 결측치 다루기, 새로운 컬럼 만들기

페이지번호

05_01_03

학습목차

평가하기
- 학습평가

정리하기
- 학습정리

학습평가

학습한 내용을 바탕으로 다음 문제를 풀어 보세요.

번호	문제	정답	해설
1	인자로 전달하는 매핑 함수에 시리즈의 모든 원소를 하나씩 입력하고 함수의 리턴 값을 돌려 받는 함수는? ① assign()함수 ② apply() 함수 ③ clip() 함수 ④ qcut()함수	2	apply 함수에 대한 설명이다.
2	숫자형 데이터를 n개의 범주로 구분할 때 사용하는 함수는? ① assign()함수 ② apply() 함수 ③ clip() 함수 ④ qcut()함수	4	qcut 함수에 대한 설명이다.
3	df.max(axis=0)의 데이터프레임에서 어떤 값을 가져오는 판다스 코드인가? ① x축에서 최대값 가져오기 ② y축에서 최대값 가져오기 ③ 행에서 최대값 가져오기 ④ 열에서 최대값 가져오기	3	axis = 0 : 행, axis =1 : 열

화면설명

[학습평가 페이지]
- 페이지 퀴즈 컴포넌트 사용하여 페이지 개발

학습평가

화면설명

평가하기

- 학습평가

정리하기

- 학습정리

학습한 내용을 바탕으로 다음 문제를 풀어 보세요.

번호	문제	정답	해설
1	인자로 전달하는 매핑 함수에 시리즈의 모든 원소를 하나씩 입력하고 함수의 리턴 값을 돌려 받는 함수는? ① assign()함수 ② apply() 함수 ③ clip() 함수 ④ qcut()함수	2	apply 함수에 대한 설명이다.
2	숫자형 데이터를 n개의 범주로 구분할 때 사용하는 함수는? ① assign()함수 ② apply() 함수 ③ clip() 함수 ④ qcut()함수	4	qcut 함수에 대한 설명이다.
3	df.max(axis=0)의 데이터프레임에서 어떤 값을 가져오는 판다스 코드인가? ① x축에서 최대값 가져오기 ② y축에서 최대값 가져오기 ③ 행에서 최대값 가져오기 ④ 열에서 최대값 가져오기	3	axis = 0 : 행, axis =1 : 열

[학습평가 페이지]

- 페이지 퀴즈 컴포넌트 사용하여 페이지 개발