[Python] Pandas 사용법 - 두가지의 DataFrame 합치기 (merge, join) — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2020-11-06 오후 5:26

URL: https://continuous-development.tistory.com/136?category=736681

Python

[Python] Pandas 사용법 - 두가지의 DataFra me 합치기 (merge, join)

2020. 10. 16. 17:12 수정 삭제 공개

DataFrame merge

두 개의 데이터 프레임을 하나로 합치는 함수

```
pd.merge(dataframe1,dataframe2)
```

```
In [240]: pd.merge(stu_df,major_df)

Out[240]: 

한번 이름 한년 학관 학점

0 1 업업이 2 산공 4.1

1 2 김완선 4 통계 4.4

2 3 양세형 1 점공 4.3
```

pd.merge를 통해 두 데이터 프레임을 합친다.

기본적으로 merge를 했을 때 같은 칼럼을 기준으로 한다. 이건 자동으로 찾아준다. 여기서는 학번을 기준으로 합쳐져서 겹쳐지는 값들만 가지고나왔다.

만약 컬럼을 지정해주고 싶을 경우에는 **on이라는** 옵션을 사용한다. 아래와 같이 on을 통해 기준이 되는 칼럼을 학번으로 지정해 줄 수 있다. 또한 **how**라는 옵션이 있는데 sql 의 join의 inner, left, right, outer 개념과 같다.

how는 어떻게 합칠 것인가를 나타낸다.

how 옵션

how = inner

-두 컬럼을 기준으로 공통되는 것만 나온다.(기본값이다)

```
In [241]: pd.merge(stu_df,major_df, on='학번',how='inner')
Out[241]:

학번 이름 학년 학과 학점

0 1 선접이 2 산공 4.1

1 2 김완선 4 통계 4.4

2 3 양세형 1 컴공 4.3
```

how = left

- 왼쪽 데이터 프레임을 기준으로 왼쪽에 맞춰 합친다.(왼쪽 칼럼의 기준은 다 나오되 합치는 데 있어서 없는 값은 NaN이 나온다.)



how = right

- 오른쪽 데이터 프레임을 기준으로 왼쪽에 맞춰 합친다.(오른쪽 컬럼의 기준은 다 나오되 합치는 데 있어서 없는 값은 NaN이 나온다.)



right 일 때는 오른쪽을 기준으로 값이나온다.

how = outer

두 데이터 프레임에 대해서 모두 출력한다.



이와 같이 겹치는 칼럼이 없어도 merge 할 수 있다. (그냥 merge를 하면 찾지 못해서 안된다.)

left_on(= 왼쪽에서 지정해 줄 칼럼)과 right_on(= 오른쪽에서 지정해 줄 칼럼)을 통해 지정해줄 칼럼을 설정하고 how를 통해 어떻게 붙일지를 정한다.

여기서 보면 왼쪽 dataframe에서 학번을 기준으로 / 오른쪽 dataframe에서는 과목 코드를 기준으로 합쳐졌다.

위에서 보면 학번과 과목 코드의 값이 똑같은 것 을 볼 수 있다.

right_index=True / left_index =True

해당 데이터 프레임의 인덱스를 기준으로 사용할 때 쓴다.

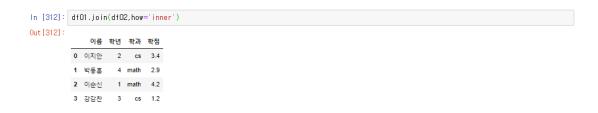


이 같은 데이터 프레임이 있을 때 두 개의 연관관계가 없다. 이럴 때 이것 들을 합치기 위해서



이런 형식으로 된다. 이렇게 되면 왼쪽 데이터 프레임에 있는 인덱스 값과 오른쪽 데이터프레임에 있는 인덱스 값을 기준으로 두 데이터 프레임을 합친다.

join



join으로도 위와 같은 형태의 데이터 프레임을 만들 수 있다.

'Python' 카테고리의 다른 글□

[Python] Pandas 사용법 - 피벗 테이블 생성(pivot,pivot_table)□

[Python] Pandas 사용법 - 그룹화 및 그룹 함수 (groupby, qcut, cut, transfrom ...

[Python] Pandas 사용법 - 두가지의 DataFrame 합치기 (merge, join) 🗆

[Python] Pandas 사용법 - 다양한 인덱스 함수(reset_index,set_index,sort_index) 🗆

[Python] Pandas 사용법 - 인덱싱 접근,데이터 조작, 인덱스조작(loc,iloc)□

[Python] Pandas 사용법 - 다양한 함수 사용(데이터 입출력, 대소문자변환, 공백...

dataframe merge dataframe 합치기 pandas merge

데이터프레임 병합

데이터프레임 합치기



나무늘보스

혼자 끄적끄적하는 블로그 입니다.