

노트북: 첫 번째 노트북

만든 날짜: 2020-11-08 오후 9:25

업데이트: 2020-11-08 오후 9:26

URL: <https://continuous-development.tistory.com/142?category=736681>

---

[Python](#)

# [Python] Pandas 사용법 - 피벗 테이블 생성(pivot,pivot\_table)

2020. 10. 19. 16:47 [수정삭제공개](#)

## -pivot : 피벗 테이블

- 데이터 프레임에서 두 개의 열을 이용하여 행/열 인덱스 reshape 된 테이블을 의미한다.
- 새로운 테이블에서 새로운 기준으로 집계

두 가지 방법으로 피벗테이블을 만든다.

- pivot(index, columns, values) - groupby가 필요
  - pd.pivot\_table(data, values, index, columns, aggfunc='mean(함수)') - groupby 없이 가능
- or
- dataFrame.pivot\_table(values(칼럼), index, columns, aggfunc='mean(함수)')

## # pivot

`pivot(index, columns, values)`

```
In [78]: titanic_df01 = pd.DataFrame(titanic, columns = ['sex', 'pclass'])
titanic_df01.head()
```

Out[78]:

	sex	pclass
0	male	3
1	female	1
2	female	3
3	female	1
4	male	3

```
In [81]: # 성별과 객실등급에 따른 승객 수 집계
titanic_df01 = titanic.groupby(['sex', 'pclass']).size().reset_index(name='cnt')
titanic_df01
```

Out[81]:

	sex	pclass	cnt
0	female	1	94
1	female	2	76
2	female	3	144
3	male	1	122
4	male	2	108
5	male	3	347

pivot을 하기 전에는 그전에 groupby로 필요한 값들을 묶어줘야 한다.

```
In [83]: titanic_df01.pivot('sex', 'pclass', 'cnt')
```

Out[83]:

	pclass	1	2	3
sex				
female		94	76	144
male		122	108	347

이렇게 pivot을 구성하면 index = sex , columns = pclass , value = cnt 인 테이블이 생성된다.

## # pivot table

```
pd.pivot_table(data, values, index, columns, aggfunc='mean(함수)')
```

or

```
dataFrame.pivot_table(values(컬럼), index, columns, aggfunc='mean(함수)')
```

```
In [98]: titanic.pivot_table('cnt', 'sex', 'survived', aggfunc=np.sum)
```

Out[98]:

	survived	0	1
sex			
female		81	233
male		468	109

해당 데이터프레임에 cnt, sex, survived

## # pivot table option

- data : 데이터 프레임
- values : 분석할 열
- index
- columns
- aggfunc : 집계 함수
- fill\_value : Nan 대체 값
- margins : 분석 결과를 오른쪽과 아래에 붙일지 여부
- margins\_name : 마진 열(행)의 이름

```
In [106]: pivot_sample_df.pivot_table('인구', '도시', '연도', margins= True)
```

Out [106]:

연도	2005	2010	2015	All
도시				
부산	3512547.0	3393191.0	3448737.0	3.451492e+06
서울	9762546.0	9631482.0	9904312.0	9.766113e+06
인천	NaN	263203.0	2890451.0	1.576827e+06
All	6637546.5	4429292.0	5414500.0	5.350809e+06

margin을 하면 분석 결과가 나오고 True 값일 때 밑에다가 붙여서 넣는다.

```
In [107]: pivot_sample_df.pivot_table('인구', '도시', '연도', margins= True, margins_name = '분석결과')
```

Out [107]:

연도	2005	2010	2015	분석결과
도시				
부산	3512547.0	3393191.0	3448737.0	3.451492e+06
서울	9762546.0	9631482.0	9904312.0	9.766113e+06
인천	NaN	263203.0	2890451.0	1.576827e+06
분석결과	6637546.5	4429292.0	5414500.0	5.350809e+06

margins\_name 은 칼럼과 열의 이름을 넣는다.

나무늘보스  
혼자 끄적끄적하는 블로그 입니다.





