# pandas_merge (7)

January 12, 2020

#

Pandas Merge Tutorial

### 0.1 Basic Merge Using a Dataframe Column

[1]:
```
import pandas as pd
df1 = pd.DataFrame({
    "city": ["new york","chicago","orlando"],
    "temperature": [21,14,35],
})
df1
```

[1]:

| | city | temperature |
|---|---|---|
| 0 | new york | 21 |
| 1 | chicago | 14 |
| 2 | orlando | 35 |

[2]:
```
df2 = pd.DataFrame({
    "city": ["chicago","new york","orlando"],
    "humidity": [65,68,75],
})
df2
```

[2]:

| | city | humidity |
|---|---|---|
| 0 | chicago | 65 |
| 1 | new york | 68 |
| 2 | orlando | 75 |

[3]:
```
df3 = pd.merge(df1, df2, on="city")
df3
```

[3]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21 | 68 |
| 1 | chicago | 14 | 65 |
| 2 | orlando | 35 | 75 |

### 0.2 Type Of DataBase Joins

<img src="db_joins.jpg" height="800", width="800">

[4]:
```
df1 = pd.DataFrame({
    "city": ["new york","chicago","orlando", "baltimore"],
    "temperature": [21,14,35, 38],
})
df1
```

[4]:

| | city | temperature |
|---|---|---|
| 0 | new york | 21 |
| 1 | chicago | 14 |
| 2 | orlando | 35 |
| 3 | baltimore | 38 |

[5]:
```
df2 = pd.DataFrame({
    "city": ["chicago","new york","san diego"],
    "humidity": [65,68,71],
})
df2
```

[5]:

| | city | humidity |
|---|---|---|
| 0 | chicago | 65 |
| 1 | new york | 68 |
| 2 | san diego | 71 |

[6]:
```
df3=pd.merge(df1,df2,on="city",how="inner")
df3
```

[6]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21 | 68 |
| 1 | chicago | 14 | 65 |

[7]:
```
df3=pd.merge(df1,df2,on="city",how="outer")
df3
```

[7]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21.0 | 68.0 |
| 1 | chicago | 14.0 | 65.0 |
| 2 | orlando | 35.0 | NaN |
| 3 | baltimore | 38.0 | NaN |
| 4 | san diego | NaN | 71.0 |

[8]:
```
df3=pd.merge(df1,df2,on="city",how="left")
df3
```

[8]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21 | 68.0 |
| 1 | chicago | 14 | 65.0 |
| 2 | orlando | 35 | NaN |
| 3 | baltimore | 38 | NaN |

[9]:
```
df3=pd.merge(df1,df2,on="city",how="right")
df3
```

[9]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21.0 | 68 |
| 1 | chicago | 14.0 | 65 |
| 2 | san diego | NaN | 71 |

### 0.3 indicator flag

[10]:
```
df3=pd.merge(df1,df2,on="city",how="outer",indicator=True)
df3
```

[10]:

| | city | temperature | humidity | _merge |
|---|---|---|---|---|
| 0 | new york | 21.0 | 68.0 | both |
| 1 | chicago | 14.0 | 65.0 | both |
| 2 | orlando | 35.0 | NaN | left_only |
| 3 | baltimore | 38.0 | NaN | left_only |
| 4 | san diego | NaN | 71.0 | right_only |

### 0.4 suffixes

[11]:
```
df1 = pd.DataFrame({
    "city": ["new york","chicago","orlando", "baltimore"],
    "temperature": [21,14,35,38],
    "humidity": [65,68,71, 75]
})
df1
```

[11]:

| | city | temperature | humidity |
|---|---|---|---|
| 0 | new york | 21 | 65 |
| 1 | chicago | 14 | 68 |
| 2 | orlando | 35 | 71 |
| 3 | baltimore | 38 | 75 |

[12]:
```
df2 = pd.DataFrame({
    "city": ["chicago","new york","san diego"],
    "temperature": [21,14,35],
    "humidity": [65,68,71]
})
df2
```

[12]:
```
         city  temperature  humidity
0     chicago           21        65
1    new york           14        68
2   san diego           35        71
```

[13]:
```
df3= pd.merge(df1,df2,on="city",how="outer", suffixes=('_first','_second'))
df3
```

[13]:
```
         city  temperature_first  humidity_first  temperature_second  \
0    new york               21.0            65.0                14.0
1     chicago               14.0            68.0                21.0
2     orlando               35.0            71.0                 NaN
3   baltimore               38.0            75.0                 NaN
4   san diego                NaN             NaN                35.0

   humidity_second
0             68.0
1             65.0
2              NaN
3              NaN
4             71.0
```

### 0.5 join

[14]:
```
df1 = pd.DataFrame({
    "city": ["new york","chicago","orlando"],
    "temperature": [21,14,35],
})
df1.set_index('city',inplace=True)
df1
```

[14]:
```
          temperature
city
new york           21
chicago            14
orlando            35
```

[15]:
```
df2 = pd.DataFrame({
    "city": ["chicago","new york","orlando"],
    "humidity": [65,68,75],
})
df2.set_index('city',inplace=True)
df2
```

[15]:
```
          humidity
city
chicago         65
```

```
new york        68
orlando         75
```

[16]:
```
df1.join(df2,lsuffix='_l', rsuffix='_r')
```

[16]:
```
          temperature  humidity
city
new york           21        68
chicago            14        65
orlando            35        75
```