

```
In [86]: ▶ import numpy as np
import pandas as pd
```

## Merging Datasets

```
In [87]: ▶ df1 = pd.DataFrame({'Name': ['ali', 'taha', 'sara'], 'C++': [18, 12, 17]})
df1
```

Out[87]:

	Name	C++
0	ali	18
1	taha	12
2	sara	17

```
In [88]: ▶ df2 = pd.DataFrame({'Name': ['ali', 'taha', 'omid'], 'Python': [13, 14, 16]})
df2
```

Out[88]:

	Name	Python
0	ali	13
1	taha	14
2	omid	16

```
In [89]: ▶ dfm = df1.merge(df2, on='Name')
dfm
```

Out[89]:

	Name	C++	Python
0	ali	18	13
1	taha	12	14

```
In [90]: ▶ df1.merge(df2, how='inner', on='Name')
```

Out[90]:

	Name	C++	Python
0	ali	18	13
1	taha	12	14

```
In [91]: df1.merge(df2, how='outer', on='Name')
```

Out[91]:

	Name	C++	Python
0	ali	18.0	13.0
1	taha	12.0	14.0
2	sara	17.0	NaN
3	omid	NaN	16.0

```
In [92]: df1.merge(df2, how='left', on='Name')
```

Out[92]:

	Name	C++	Python
0	ali	18	13.0
1	taha	12	14.0
2	sara	17	NaN

```
In [93]: df1.merge(df2, how='right', on='Name')
```

Out[93]:

	Name	C++	Python
0	ali	18.0	13
1	taha	12.0	14
2	omid	NaN	16

```
In [94]: #
```

```
In [95]: df1 = pd.DataFrame({'X': ['ali', 'sara', 'taha', 'ali'], 'V': [1, 2, 3, 5]})  
df1
```

Out[95]:

	X	V
0	ali	1
1	sara	2
2	taha	3
3	ali	5

```
In [96]: ▶ df2 = pd.DataFrame({'Y': ['ali', 'sara', 'taha', 'ali'], 'V': [5, 6, 7, 8]})  
df2
```

Out[96]:

	Y	V
0	ali	5
1	sara	6
2	taha	7
3	ali	8

```
In [97]: ▶ df1.merge(df2, left_on='X', right_on='Y')
```

Out[97]:

	X	V_x	Y	V_y
0	ali	1	ali	5
1	ali	1	ali	8
2	ali	5	ali	5
3	ali	5	ali	8
4	sara	2	sara	6
5	taha	3	taha	7

```
In [98]: ▶ df1.merge(df2, left_on='X', right_on='Y', suffixes=('_df1', '_df2'))
```

Out[98]:

	X	V_df1	Y	V_df2
0	ali	1	ali	5
1	ali	1	ali	8
2	ali	5	ali	5
3	ali	5	ali	8
4	sara	2	sara	6
5	taha	3	taha	7

```
In [99]: ▶ #
```

```
In [100]: ▶ df1 = pd.DataFrame({'X': ['b', 'b', 'a', 'c', 'a', 'a', 'b'], 'Y': range(7)})  
df1
```

Out[100]:

	X	Y
0	b	0
1	b	1
2	a	2
3	c	3
4	a	4
5	a	5
6	b	6

```
In [101]: ▶ df2 = pd.DataFrame({'X': ['a', 'b', 'd'], 'Z': range(3)})  
df2
```

Out[101]:

	X	Z
0	a	0
1	b	1
2	d	2

```
In [102]: ▶ pd.merge(df1, df2)
```

Out[102]:

	X	Y	Z
0	b	0	1
1	b	1	1
2	b	6	1
3	a	2	0
4	a	4	0
5	a	5	0

```
In [103]: ▶ pd.merge(df1, df2, on='X')
```

Out[103]:

	X	Y	Z
0	b	0	1
1	b	1	1
2	b	6	1
3	a	2	0
4	a	4	0
5	a	5	0

```
In [104]: ▶ pd.merge(df1, df2, how='outer')
```

Out[104]:

	X	Y	Z
0	b	0.0	1.0
1	b	1.0	1.0
2	b	6.0	1.0
3	a	2.0	0.0
4	a	4.0	0.0
5	a	5.0	0.0
6	c	3.0	NaN
7	d	NaN	2.0

```
In [105]: ▶ #####
```

```
In [106]: ▶ frame1 = pd.DataFrame({'A': ['b', 'b', 'a', 'c', 'a', 'a', 'b'], 'B': range(7)})  
frame1
```

Out[106]:

	A	B
0	b	0
1	b	1
2	a	2
3	c	3
4	a	4
5	a	5
6	b	6

```
In [107]: ▶ frame2 = pd.DataFrame({'C': ['a', 'b', 'd'], 'D': range(3)})  
frame2
```

Out[107]:

	C	D
0	a	0
1	b	1
2	d	2

```
In [108]: ▶ pd.merge(frame1, frame2, left_on='A', right_on='C')
```

Out[108]:

	A	B	C	D
0	b	0	b	1
1	b	1	b	1
2	b	6	b	1
3	a	2	a	0
4	a	4	a	0
5	a	5	a	0

```
In [109]: ▶ #####
```

```
In [110]: ▶ df1 = pd.DataFrame({'X': ['b', 'b', 'a', 'c', 'a', 'b'], 'Y': range(6)})  
df1
```

Out[110]:

	X	Y
0	b	0
1	b	1
2	a	2
3	c	3
4	a	4
5	b	5

```
In [111]: ▶ df2 = pd.DataFrame({'X': ['a', 'b', 'a', 'b', 'd'], 'Z': range(5)})  
df2
```

Out[111]:

	X	Z
0	a	0
1	b	1
2	a	2
3	b	3
4	d	4

```
In [112]: ▶ pd.merge(df1, df2)
```

Out[112]:

	X	Y	Z
0	b	0	1
1	b	0	3
2	b	1	1
3	b	1	3
4	b	5	1
5	b	5	3
6	a	2	0
7	a	2	2
8	a	4	0
9	a	4	2

```
In [113]: ▶ pd.merge(df1, df2, how='left')
```

```
Out[113]:
```

	X	Y	Z
0	b	0	1.0
1	b	0	3.0
2	b	1	1.0
3	b	1	3.0
4	a	2	0.0
5	a	2	2.0
6	c	3	NaN
7	a	4	0.0
8	a	4	2.0
9	b	5	1.0
10	b	5	3.0

```
In [114]: ▶ pd.merge(df1, df2, how='right')
```

```
Out[114]:
```

	X	Y	Z
0	a	2.0	0
1	a	4.0	0
2	b	0.0	1
3	b	1.0	1
4	b	5.0	1
5	a	2.0	2
6	a	4.0	2
7	b	0.0	3
8	b	1.0	3
9	b	5.0	3
10	d	NaN	4

```
In [115]: ▶ #####
```



```
In [116]: ▶ frame1 = pd.DataFrame({'k1': ['foo', 'foo', 'bar'], 'k2': ['one', 'two', 'one']})
frame1
```

Out[116]:

	k1	k2	Z
0	foo	one	1
1	foo	two	2
2	bar	one	3

```
In [117]: ▶ frame2 = pd.DataFrame({'k1': ['foo', 'foo', 'bar', 'bar'],
                                   'k2': ['one', 'one', 'one', 'two'],
                                   'W': [4, 5, 6, 7]})
frame2
```

Out[117]:

	k1	k2	W
0	foo	one	4
1	foo	one	5
2	bar	one	6
3	bar	two	7

```
In [118]: ▶ pd.merge(frame1, frame2, on=['k1', 'k2'], how='outer')
```

Out[118]:

	k1	k2	Z	W
0	foo	one	1.0	4.0
1	foo	one	1.0	5.0
2	foo	two	2.0	NaN
3	bar	one	3.0	6.0
4	bar	two	NaN	7.0

```
In [119]: ▶ pd.merge(frame1, frame2, on='k1', suffixes=('_df1', '_df2'))
```

Out[119]:

	k1	k2_df1	Z	k2_df2	W
0	foo	one	1	one	4
1	foo	one	1	one	5
2	foo	two	2	one	4
3	foo	two	2	one	5
4	bar	one	3	one	6
5	bar	one	3	two	7

## Merging on Index

```
In [120]: ▶ df1 = pd.DataFrame({'X': ['a', 'b', 'a', 'a', 'b', 'c'], 'Y': range(6)})  
df1
```

Out[120]:

	X	Y
0	a	0
1	b	1
2	a	2
3	a	3
4	b	4
5	c	5

```
In [121]: ▶ df2 = pd.DataFrame({'Z': [18, 15]}, index=['a', 'b'])  
df2
```

Out[121]:

	Z
a	18
b	15

```
In [122]: ▶ pd.merge(df1, df2, left_on='X', right_index=True)
```

Out[122]:

	X	Y	Z
0	a	0	18
2	a	2	18
3	a	3	18
1	b	1	15
4	b	4	15

```
In [123]: ▶ pd.merge(df1, df2, left_on='X', right_index=True, how='outer')
```

Out[123]:

	X	Y	Z
0	a	0	18.0
2	a	2	18.0
3	a	3	18.0
1	b	1	15.0
4	b	4	15.0
5	c	5	NaN

In [124]: `####`

In [125]: `d = {'X': ['H', 'H', 'H', 'N', 'N'], 'Y': [1397, 1398, 1399, 1398, 1399], 'Z':  
frame1 = pd.DataFrame(d)  
frame1`

Out[125]:

	X	Y	Z
0	H	1397	0.0
1	H	1398	1.0
2	H	1399	2.0
3	N	1398	3.0
4	N	1399	4.0

In [126]: `i = [['N', 'N', 'H', 'H', 'H', 'H'], [1398, 1397, 1397, 1397, 1398, 1399]]  
frame2 = pd.DataFrame(np.arange(12).reshape((6, 2)), index=i, columns=['A', '  
frame2`

Out[126]:

		A	B
N	1398	0	1
	1397	2	3
H	1397	4	5
	1397	6	7
	1398	8	9
	1399	10	11

In [127]: `pd.merge(frame1, frame2, left_on=['X', 'Y'], right_index=True)`

Out[127]:

	X	Y	Z	A	B
0	H	1397	0.0	4	5
0	H	1397	0.0	6	7
1	H	1398	1.0	8	9
2	H	1399	2.0	10	11
3	N	1398	3.0	0	1

```
In [128]: ▶ pd.merge(frame1, frame2, left_on=['X', 'Y'], right_index=True, how='outer')
```

Out[128]:

	X	Y	Z	A	B
0	H	1397	0.0	4.0	5.0
0	H	1397	0.0	6.0	7.0
1	H	1398	1.0	8.0	9.0
2	H	1399	2.0	10.0	11.0
3	N	1398	3.0	0.0	1.0
4	N	1399	4.0	NaN	NaN
4	N	1397	NaN	2.0	3.0

## join

```
In [129]: ▶ df1 = pd.DataFrame([[19, 12], [13, 18], [5, 16]], index=['ali', 'sara', 'taha'], df1
```

Out[129]:

	C++	Python
ali	19	12
sara	13	18
taha	5	16

```
In [130]: ▶ df2 = pd.DataFrame([[17, 12], [19, 20], [11, 6], [13, 18]], index=['farid', 'sara', 'mahsa', 'taha'], df2
```

Out[130]:

	Java	PHP
farid	17	12
sara	19	20
mahsa	11	6
taha	13	18

```
In [131]: ▶ #df1.merge(df2)
```

In [132]: `df1.merge(df2, left_index=True, right_index=True)`

Out[132]:

	C++	Python	Java	PHP
sara	13	18	19	20
taha	5	16	13	18

In [133]: `df1.join(df2)`

Out[133]:

	C++	Python	Java	PHP
ali	19	12	NaN	NaN
sara	13	18	19.0	20.0
taha	5	16	13.0	18.0

In [134]: `df2.join(df1)`

Out[134]:

	Java	PHP	C++	Python
farid	17	12	NaN	NaN
sara	19	20	13.0	18.0
mahsa	11	6	NaN	NaN
taha	13	18	5.0	16.0

In [135]: `df2.join(df1, how='outer')`

Out[135]:

	Java	PHP	C++	Python
ali	NaN	NaN	19.0	12.0
farid	17.0	12.0	NaN	NaN
mahsa	11.0	6.0	NaN	NaN
sara	19.0	20.0	13.0	18.0
taha	13.0	18.0	5.0	16.0

```
In [136]: ▶ df3 = pd.DataFrame([[15, 6], [17, 18], [19, 20], [3, 9]],
                                index=['ali', 'sara', 'taha', 'farshid'],
                                columns=['Pascal', 'C#'])
df3
```

Out[136]:

	Pascal	C#
ali	15	6
sara	17	18
taha	19	20
farshid	3	9

```
In [137]: ▶ x = df2.join(df3, how='outer')
df1.join(x)
```

Out[137]:

	C++	Python	Java	PHP	Pascal	C#
ali	19	12	NaN	NaN	15.0	6.0
sara	13	18	19.0	20.0	17.0	18.0
taha	5	16	13.0	18.0	19.0	20.0

```
In [138]: ▶ df1.join([df2, df3])
```

Out[138]:

	C++	Python	Java	PHP	Pascal	C#
ali	19.0	12.0	NaN	NaN	15.0	6.0
sara	13.0	18.0	19.0	20.0	17.0	18.0
taha	5.0	16.0	13.0	18.0	19.0	20.0

## concat

```
In [139]: ▶ df1
```

Out[139]:

	C++	Python
ali	19	12
sara	13	18
taha	5	16

In [140]: `df2`

Out[140]:

	Java	PHP
<b>farid</b>	17	12
<b>sara</b>	19	20
<b>mahsa</b>	11	6
<b>taha</b>	13	18

In [141]: `pd.concat([df1, df2])`

Out[141]:

	C++	Python	Java	PHP
<b>ali</b>	19.0	12.0	NaN	NaN
<b>sara</b>	13.0	18.0	NaN	NaN
<b>taha</b>	5.0	16.0	NaN	NaN
<b>farid</b>	NaN	NaN	17.0	12.0
<b>sara</b>	NaN	NaN	19.0	20.0
<b>mahsa</b>	NaN	NaN	11.0	6.0
<b>taha</b>	NaN	NaN	13.0	18.0

In [142]: `pd.concat([df1, df2], axis=1)`

Out[142]:

	C++	Python	Java	PHP
<b>ali</b>	19.0	12.0	NaN	NaN
<b>sara</b>	13.0	18.0	19.0	20.0
<b>taha</b>	5.0	16.0	13.0	18.0
<b>farid</b>	NaN	NaN	17.0	12.0
<b>mahsa</b>	NaN	NaN	11.0	6.0

In [143]: `#`

In [144]: `s1 = pd.Series([12, 18], index=['ali', 'sara'])`  
`s1`

Out[144]: `ali 12`  
`sara 18`  
`dtype: int64`

```
In [145]: ▶ s2 = pd.Series([13, 20, 14], index=['taha', 'mahsa', 'sara'])
s2
```

```
Out[145]: taha    13
mahsa    20
sara     14
dtype: int64
```

```
In [146]: ▶ pd.concat([s1, s2])
```

```
Out[146]: ali      12
sara      18
taha      13
mahsa     20
sara      14
dtype: int64
```

```
In [147]: ▶ pd.concat([s1, s2], axis=1)
```

```
Out[147]:
```

	0	1
<b>ali</b>	12.0	NaN
<b>sara</b>	18.0	14.0
<b>taha</b>	NaN	13.0
<b>mahsa</b>	NaN	20.0

```
In [148]: ▶ pd.concat([s1, s2], axis=1, keys=['c++', 'python'])
```

```
Out[148]:
```

	c++	python
<b>ali</b>	12.0	NaN
<b>sara</b>	18.0	14.0
<b>taha</b>	NaN	13.0
<b>mahsa</b>	NaN	20.0

```
In [149]: ▶ pd.concat([s1, s2], keys=['c++', 'python'])
```

```
Out[149]: c++    ali    12
           sara    18
python  taha    13
           mahsa   20
           sara    14
dtype: int64
```



```
In [150]: ▶ r = pd.concat([s1, s2], keys=['c++', 'python'])
r
```

```
Out[150]: c++      ali      12
           sara      18
python   taha      13
           mahsa     20
           sara      14
dtype: int64
```

```
In [151]: ▶ r.unstack()
```

```
Out[151]:
```

	ali	mahsa	sara	taha
c++	12.0	NaN	18.0	NaN
python	NaN	20.0	14.0	13.0

## combine\_first

```
In [152]: ▶ df1 = pd.DataFrame({'C++': [None, 12], 'Python': [None, 14]}, index=['Ali', 'Taha'])
df1
```

```
Out[152]:
```

	C++	Python
Ali	NaN	NaN
Taha	12.0	14.0

```
In [153]: ▶ df2 = pd.DataFrame({'C++': [None, 15], 'Python': [13, None]}, index=['Ali', 'Taha'])
df2
```

```
Out[153]:
```

	C++	Python
Ali	NaN	13.0
Taha	15.0	NaN

```
In [154]: ▶ df1.combine_first(df2)
```

```
Out[154]:
```

	C++	Python
Ali	NaN	13.0
Taha	12.0	14.0

In [155]: `df2.combine_first(df1)`

Out[155]:

	C++	Python
<b>Ali</b>	NaN	13.0
<b>Taha</b>	15.0	14.0

In [156]: `#`

In [157]: `df1 = pd.DataFrame({'C++': [None, 12], 'Python': [14, None]}, index=['Ali', 'Taha'])`

Out[157]:

	C++	Python
<b>Ali</b>	NaN	14.0
<b>Taha</b>	12.0	NaN

In [158]: `df2 = pd.DataFrame({'Python': [13, 20, None], 'Java': [None, 11, 17]}, index=['Ali', 'Taha', 'Mahsa'])`

Out[158]:

	Python	Java
<b>Ali</b>	13.0	NaN
<b>Taha</b>	20.0	11.0
<b>Mahsa</b>	NaN	17.0

In [159]: `df1.combine_first(df2)`

Out[159]:

	C++	Java	Python
<b>Ali</b>	NaN	NaN	14.0
<b>Mahsa</b>	NaN	17.0	NaN
<b>Taha</b>	12.0	11.0	20.0

In [160]: `df2.combine_first(df1)`

Out[160]:

	C++	Java	Python
<b>Ali</b>	NaN	NaN	13.0
<b>Mahsa</b>	NaN	17.0	NaN
<b>Taha</b>	12.0	11.0	20.0

## pivot

```
In [161]: ▶ df = pd.DataFrame({
    "A": ['one', 'one', 'one', 'two', 'two', 'two'],
    "B": [1, 1, 2, 1, 1, 2],
    "C": [1, 2, 1, 2, 1, 2],
    "V": [0, 1, 2, 3, 4, 5]})
df
```

Out[161]:

	A	B	C	V
0	one	1	1	0
1	one	1	2	1
2	one	2	1	2
3	two	1	2	3
4	two	1	1	4
5	two	2	2	5

```
In [162]: ▶ df.pivot(index="A", columns=["B", "C"], values="V")
```

Out[162]:

	B		C	
A	1	2	1	2
one	0.0	1.0	2.0	NaN
two	4.0	3.0	NaN	5.0

```
In [163]: ▶ df.pivot(index=["A", "B"], columns=["C"], values="V")
```

Out[163]:

	C	
A B	1	2
one 1	0.0	1.0
2	2.0	NaN
two 1	4.0	3.0
2	NaN	5.0

## melt

```
In [164]: ▶ mydict = {'Name': {0: 'Ali', 1: 'Sara', 2: 'Mahsa'}, 'C++': {0: 12, 1: 13, 2: 20}, 'Python': {0: 14, 1: 16, 2: 8}}
df = pd.DataFrame(mydict)
df
```

Out[164]:

	Name	C++	Python
0	Ali	12	14
1	Sara	13	16
2	Mahsa	20	8

```
In [165]: ▶ df.melt(['Name']) # df.melt(id_vars=['Name'])
```

Out[165]:

	Name	variable	value
0	Ali	C++	12
1	Sara	C++	13
2	Mahsa	C++	20
3	Ali	Python	14
4	Sara	Python	16
5	Mahsa	Python	8

```
In [166]: ▶ m = pd.melt(df, ['Name'])
r = m.pivot('Name', 'variable', 'value')
r
```

Out[166]:

	variable	C++	Python
	Name		
	Ali	12	14
	Mahsa	20	8
	Sara	13	16

```
In [167]: ▶ r.reset_index()
```

Out[167]:

	variable	Name	C++	Python
0	Ali	12	14	
1	Mahsa	20	8	
2	Sara	13	16	

```
In [168]: df
```

```
Out[168]:
```

	Name	C++	Python
0	Ali	12	14
1	Sara	13	16
2	Mahsa	20	8

```
In [169]: df.melt(['Name'], value_vars=['C++'], var_name='Dars', value_name='Score')
```

```
Out[169]:
```

	Name	Dars	Score
0	Ali	C++	12
1	Sara	C++	13
2	Mahsa	C++	20

```
In [170]: df.melt(['Name'], value_vars=['C++'])
```

```
Out[170]:
```

	Name	variable	value
0	Ali	C++	12
1	Sara	C++	13
2	Mahsa	C++	20

دانشگاه شهید مدنی آذربایجان  
برنامه نویسی پیشرفته با پایتون  
امین گلزاری اسکویی  
۱۴۰۰-۱۴۰۱

[Codes and Projects \(click here\)](https://github.com/Amin-Golzari-Oskouei/Python-Programming-Course-Advanced-2021) (<https://github.com/Amin-Golzari-Oskouei/Python-Programming-Course-Advanced-2021>) [slides and videos \(click here\)](https://drive.google.com/drive/folders/1Dx3v7fD1QBWL-MNP2hd7ilxaRbeALkKA) (<https://drive.google.com/drive/folders/1Dx3v7fD1QBWL-MNP2hd7ilxaRbeALkKA>)