

Module 02

Pandas: Merging, Joining, and Concatenating

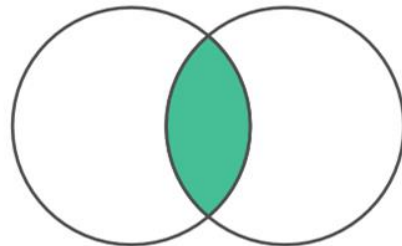
Data Science Developer

Outline

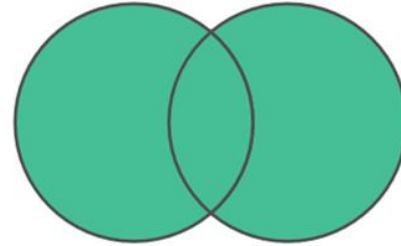
- Merging and joining Dataframe
 - Inner Join
 - Outer Join
 - Left Join
 - Right Join
- Concatenating Dataframe

Merging and Joining

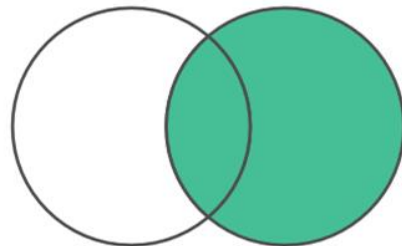
Joining Methods



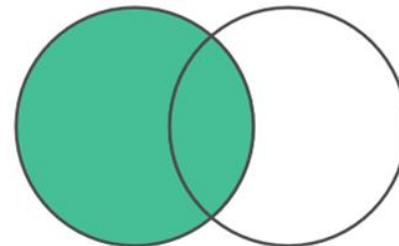
INNER JOIN



FULL OUTER JOIN



RIGHT OUTER JOIN



LEFT OUTER JOIN

The **merge** function allows you to merge DataFrames together using a similar logic as merging SQL Tables together.

Dataframe to join

In [12]: left

Out[12]:

	key	A	B
0	K0	A0	B0
1	K1	A1	B1
2	K2	A2	B2
3	K3	A3	B3

In [13]: right

Out[13]:

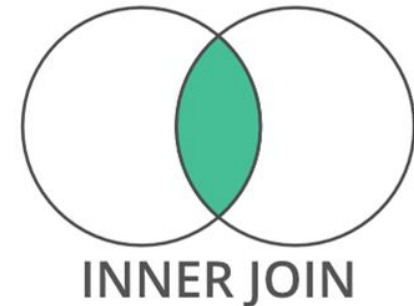
	key	C	D
0	K0	C0	D0
1	K1	C1	D1
2	K3	C2	D2
3	K4	C3	D3

Inner Join and Outer Join

```
In [14]: pd.merge(left,right,on='key')
```

Out[14]:

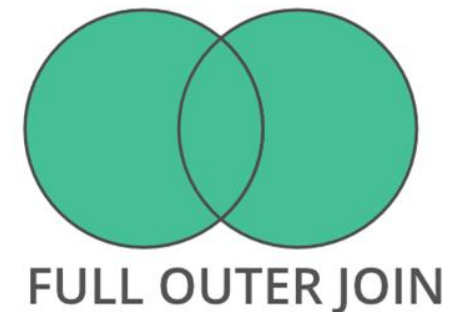
	key	A	B	C	D
0	K0	A0	B0	C0	D0
1	K1	A1	B1	C1	D1
2	K3	A3	B3	C2	D2



```
In [15]: pd.merge(left,right,how='outer', on='key')
```

Out[15]:

	key	A	B	C	D
0	K0	A0	B0	C0	D0
1	K1	A1	B1	C1	D1
2	K2	A2	B2	NaN	NaN
3	K3	A3	B3	C2	D2
4	K4	NaN	NaN	C3	D3

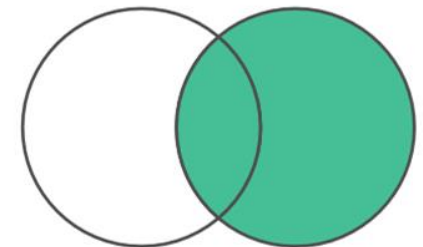


left Join and Right Join

```
In [16]: pd.merge(left, right, how= 'right', on= 'key')
```

Out[16]:

	key	A	B	C	D
0	K0	A0	B0	C0	D0
1	K1	A1	B1	C1	D1
2	K3	A3	B3	C2	D2
3	K4	NaN	NaN	C3	D3

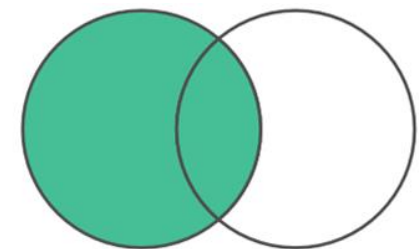


RIGHT OUTER JOIN

```
In [17]: pd.merge(left, right, how= 'left', on= 'key')
```

Out[17]:

	key	A	B	C	D
0	K0	A0	B0	C0	D0
1	K1	A1	B1	C1	D1
2	K2	A2	B2	NaN	NaN
3	K3	A3	B3	C2	D2



LEFT OUTER JOIN

Joining by index

In [19]: left

Out[19]:

	A	B
K0	A0	B0
K1	A1	B1
K2	A2	B2
K3	A3	B3

In [20]: right

Out[20]:

	C	D
K0	C0	D0
K1	C1	D1
K3	C2	D2
K4	C3	D3

Joining is a convenient method for combining the columns of two potentially differently-indexed DataFrames into a single result DataFrame.

In [21]: left.join(right)

Out[21]:

	A	B	C	D
K0	A0	B0	C0	D0
K1	A1	B1	C1	D1
K2	A2	B2	NaN	NaN
K3	A3	B3	C2	D2

In [22]: left.join(right, how='outer')

Out[22]:

	A	B	C	D
K0	A0	B0	C0	D0
K1	A1	B1	C1	D1
K2	A2	B2	NaN	NaN
K3	A3	B3	C2	D2
K4	NaN	NaN	C3	D3

Concatenating

The DataFrames

In [5]:

df1

Out[5]:

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3

In [6]:

df2

Out[6]:

	A	B	C	D
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7

In [7]:

df3

Out[7]:

	A	B	C	D
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11

Concatenation

In [8]: `pd.concat([df1,df2,df3])`

Out[8]:

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11

Concatenation

```
In [46]: pd.concat([df1,df2,df3],axis=1)
```

```
Out[46]:
```

	A	B	C	A	B	C	A	B	C
0	A0	B0	C0	A4	B4	C4	A8	B8	C8
1	A1	B1	C1	A5	B5	C5	A9	B9	C9
2	A2	B2	C2	A6	B6	C6	A10	B10	C10
3	A3	B3	C3	A7	B7	C7	A11	B11	C11

Reference

- Merge, join, concatenate and compare.
https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html