Working with Tables II: Merging

In [1]:
```
import pandas as pd
```

Example table

In [2]:
```
t1 = pd.DataFrame({'A':[1,2,3],'B':[4,5,6], 'X':['a','b','c']})
t1
```

Out[2]:

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

In [3]:
```
t2 = pd.DataFrame({'C':[11,22,33],'D':[44,55,66], 'X':['b','c','d']})
t2
```

Out[3]:

	C	D	X
0	11	44	b
1	22	55	c
2	33	66	d

Merge vertically:

In [4]:
```
t1.append(t2)
```

Out[4]:

	A	B	X	C	D
0	1.0	4.0	a	NaN	NaN
1	2.0	5.0	b	NaN	NaN
2	3.0	6.0	c	NaN	NaN
0	NaN	NaN	b	11.0	44.0
1	NaN	NaN	c	22.0	55.0
2	NaN	NaN	d	33.0	66.0

In [5]:
```
t1.append(t2, ignore_index=True)  # create new index
```

Out[5]:

	A	B	X	C	D
0	1.0	4.0	a	NaN	NaN
1	2.0	5.0	b	NaN	NaN

	A	B	X	C	D
2	3.0	6.0	c	NaN	NaN
3	NaN	NaN	b	11.0	44.0
4	NaN	NaN	c	22.0	55.0
5	NaN	NaN	d	33.0	66.0

Or with concat:

In [6]:
```
pd.concat([t1,t2], ignore_index=True)
```

Out[6]:

	A	B	X	C	D
0	1.0	4.0	a	NaN	NaN
1	2.0	5.0	b	NaN	NaN
2	3.0	6.0	c	NaN	NaN
3	NaN	NaN	b	11.0	44.0
4	NaN	NaN	c	22.0	55.0
5	NaN	NaN	d	33.0	66.0

Merge horizontally:

In [7]:
```
t1.merge(t2, on='X')  # Merge on X
```

Out[7]:

	A	B	X	C	D
0	2	5	b	11	44
1	3	6	c	22	55

In [8]:
```
t1.merge(t2, on='X', how='inner')  # 'inner is default'
```

Out[8]:

	A	B	X	C	D
0	2	5	b	11	44
1	3	6	c	22	55

Keep all rows:

In [9]:
```
t1.merge(t2, on='X', how='outer')
```

Out[9]:

	A	B	X	C	D
0	1.0	4.0	a	NaN	NaN
1	2.0	5.0	b	11.0	44.0
2	3.0	6.0	c	22.0	55.0

	A	B	X	C	D
3	NaN	NaN	d	33.0	66.0

Keep rows of left table:

In [10]:
```
t1.merge(t2, on='X', how='left')   # keep rows of left table
```

Out[10]:

	A	B	X	C	D
0	1	4	a	NaN	NaN
1	2	5	b	11.0	44.0
2	3	6	c	22.0	55.0

Keep all rows of the right table:

In [11]:
```
t1.merge(t2, on='X', how='right')   # keep rows of left table
```

Out[11]:

	A	B	X	C	D
0	2.0	5.0	b	11	44
1	3.0	6.0	c	22	55
2	NaN	NaN	d	33	66

Create another table:

In [12]:
```
t3 = pd.DataFrame({'C':[111,222,333],'D':[444,555,666], 'Y':['c','d','e']})
t3
```

Out[12]:

	C	D	Y
0	111	444	c
1	222	555	d
2	333	666	e

Merge t1 and t3, keep all rows:

In [13]:
```
t1.merge(t3, left_on='X', right_on='Y', how='outer')
```

Out[13]:

	A	B	X	C	D	Y
0	1.0	4.0	a	NaN	NaN	NaN
1	2.0	5.0	b	NaN	NaN	NaN
2	3.0	6.0	c	111.0	444.0	c
3	NaN	NaN	NaN	222.0	555.0	d
4	NaN	NaN	NaN	333.0	666.0	e

Create 2 more tables:

In [14]:
```
t4 = pd.DataFrame({'A':[1,2,3],    'B':[4,5,6]},     index=['a','b','c'])
t5 = pd.DataFrame({'C':[11,22,33],'D':[44,55,66]}, index=['b','c','d'])
```

In [15]:
```
t4
```

Out[15]:

	A	B
a	1	4
b	2	5
c	3	6

In [16]:
```
t5
```

Out[16]:

	C	D
b	11	44
c	22	55
d	33	66

Merge t4 and t5:

In [17]:
```
t4.join(t5)   # -> merge by index, how = 'left' (default)
```

Out[17]:

	A	B	C	D
a	1	4	NaN	NaN
b	2	5	11.0	44.0
c	3	6	22.0	55.0

In [19]:
```
t4.join(t5, how='left')     # default
t4.join(t5, how='right')
t4.join(t5, how='inner')
t4.join(t5, how='outer')
```

Out[19]:

	A	B	C	D
a	1.0	4.0	NaN	NaN
b	2.0	5.0	11.0	44.0
c	3.0	6.0	22.0	55.0
d	NaN	NaN	33.0	66.0