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

Joining Pandas Objects Together

(1) Concatenating Multiple Data Frames Together

In [2]:
```
stocks_2016 = pd.read_csv('stocks_2016.csv', index_col = 'Symbol')
stocks_2017 = pd.read_csv('stocks_2017.csv', index_col = 'Symbol')
```

In [3]:
```
stocks_2016
```

Out[3]:

Symbol	Shares	Low	High
AAPL	80	95	110
TSLA	50	80	130
WMT	40	55	70

In [4]:
```
stocks_2017
```

Out[4]:

Symbol	Shares	Low	High
AAPL	50	120	140
GE	100	30	40
IBM	87	75	95
SLB	20	55	85
TXN	500	15	23
TSLA	100	100	300

In [5]:
```
stocks_2016.append(stocks_2017)
```

Out[5]:

Symbol	Shares	Low	High
AAPL	80	95	110
TSLA	50	80	130

Symbol	Shares	Low	High
WMT	40	55	70
AAPL	50	120	140
GE	100	30	40
IBM	87	75	95
SLB	20	55	85
TXN	500	15	23
TSLA	100	100	300

In [ ]:
```
from IPython.display import HTML
HTML('<iframe src=https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.con
```

In [6]:
```
s_list = [stocks_2016, stocks_2017]
pd.concat(s_list)
```

Out[6]:

Symbol	Shares	Low	High
AAPL	80	95	110
TSLA	50	80	130
WMT	40	55	70
AAPL	50	120	140
GE	100	30	40
IBM	87	75	95
SLB	20	55	85
TXN	500	15	23
TSLA	100	100	300

In [7]:
```
stocks_1 = pd.concat([stocks_2016, stocks_2017])
stocks_1
```

Out[7]:

Symbol	Shares	Low	High
AAPL	80	95	110
TSLA	50	80	130
WMT	40	55	70
AAPL	50	120	140

Symbol	Shares	Low	High
GE	100	30	40
IBM	87	75	95
SLB	20	55	85
TXN	500	15	23
TSLA	100	100	300

In [8]:
```
stocks_1.loc['AAPL']
```

Out[8]:

Symbol	Shares	Low	High
AAPL	80	95	110
AAPL	50	120	140

In [9]:
```
stocks_2 = pd.concat([stocks_2016, stocks_2017], axis = 'columns')
stocks_2
```

Out[9]:

Symbol	Shares	Low	High	Shares	Low	High
AAPL	80.0	95.0	110.0	50.0	120.0	140.0
TSLA	50.0	80.0	130.0	100.0	100.0	300.0
WMT	40.0	55.0	70.0	NaN	NaN	NaN
GE	NaN	NaN	NaN	100.0	30.0	40.0
IBM	NaN	NaN	NaN	87.0	75.0	95.0
SLB	NaN	NaN	NaN	20.0	55.0	85.0
TXN	NaN	NaN	NaN	500.0	15.0	23.0

In [10]:
```
stocks_2.loc[:, 'Shares']
```

Out[10]:

Symbol	Shares	Shares
AAPL	80.0	50.0
TSLA	50.0	100.0
WMT	40.0	NaN
GE	NaN	100.0

Symbol	Shares	Shares
IBM	NaN	87.0
SLB	NaN	20.0
TXN	NaN	500.0

In [11]:
```
pd.concat([stocks_2016, stocks_2017], keys = ['2016', '2017'])
```

Out[11]:

	Symbol	Shares	Low	High
2016	**AAPL**	80	95	110
	TSLA	50	80	130
	WMT	40	55	70
2017	**AAPL**	50	120	140
	GE	100	30	40
	IBM	87	75	95
	SLB	20	55	85
	TXN	500	15	23
	TSLA	100	100	300

In [12]:
```
pd.concat([stocks_2016, stocks_2017], keys = ['2016', '2017'], names = ['Year'])
```

Out[12]:

Year	Symbol	Shares	Low	High
2016	**AAPL**	80	95	110
	TSLA	50	80	130
	WMT	40	55	70
2017	**AAPL**	50	120	140
	GE	100	30	40
	IBM	87	75	95
	SLB	20	55	85
	TXN	500	15	23
	TSLA	100	100	300

In [13]:
```
pd.concat([stocks_2016, stocks_2017], keys = ['2016', '2017'], names = ['Year', 'Symbol
```

Out[13]:

Year	Symbol_2	Shares	Low	High
2016	**AAPL**	80	95	110
	TSLA	50	80	130
	WMT	40	55	70
2017	**AAPL**	50	120	140
	GE	100	30	40
	IBM	87	75	95
	SLB	20	55	85
	TXN	500	15	23
	TSLA	100	100	300

In [14]:
```
stocks_concat = pd.concat([stocks_2016, stocks_2017], keys = ['2016', '2017'], names =
stocks_concat
```

Out[14]:

Year	Symbol	Shares	Low	High
2016	**AAPL**	80	95	110
	TSLA	50	80	130
	WMT	40	55	70
2017	**AAPL**	50	120	140
	GE	100	30	40
	IBM	87	75	95
	SLB	20	55	85
	TXN	500	15	23
	TSLA	100	100	300

In [15]:
```
stocks_concat.loc['2016']
```

Out[15]:

Symbol	Shares	Low	High
AAPL	80	95	110
TSLA	50	80	130
WMT	40	55	70

In [16]:
```
stocks_concat.loc['2016'].loc['AAPL']
```

Out[16]:
```
Shares     80
Low        95
High      110
Name: AAPL, dtype: int64
```

In [17]:
```
pd.concat([stocks_2016, stocks_2017], join = 'outer', axis = 'columns', keys = ['2016',
```

Out[17]:

Year	2016			2017		
	Shares	Low	High	Shares	Low	High
Symbol						
AAPL	80.0	95.0	110.0	50.0	120.0	140.0
TSLA	50.0	80.0	130.0	100.0	100.0	300.0
WMT	40.0	55.0	70.0	NaN	NaN	NaN
GE	NaN	NaN	NaN	100.0	30.0	40.0
IBM	NaN	NaN	NaN	87.0	75.0	95.0
SLB	NaN	NaN	NaN	20.0	55.0	85.0
TXN	NaN	NaN	NaN	500.0	15.0	23.0

In [18]:
```
pd.concat([stocks_2016, stocks_2017], join = 'inner', axis = 'columns', keys = ['2016',
```

Out[18]:

Year	2016			2017		
	Shares	Low	High	Shares	Low	High
Symbol						
AAPL	80	95	110	50	120	140
TSLA	50	80	130	100	100	300

In [19]:
```
stocks_concat_2 = pd.concat([stocks_2016, stocks_2017], axis = 'columns', keys = ['2016
stocks_concat_2
```

Out[19]:

Year	2016			2017		
	Shares	Low	High	Shares	Low	High
Symbol						
AAPL	80.0	95.0	110.0	50.0	120.0	140.0
TSLA	50.0	80.0	130.0	100.0	100.0	300.0
WMT	40.0	55.0	70.0	NaN	NaN	NaN
GE	NaN	NaN	NaN	100.0	30.0	40.0
IBM	NaN	NaN	NaN	87.0	75.0	95.0
SLB	NaN	NaN	NaN	20.0	55.0	85.0

Year	2016			2017		
	Shares	Low	High	Shares	Low	High
Symbol						
TXN	NaN	NaN	NaN	500.0	15.0	23.0

In [20]:
```
stocks_concat_2.loc['AAPL']
```

Out[20]:
```
Year
2016  Shares     80.0
      Low        95.0
      High      110.0
2017  Shares     50.0
      Low       120.0
      High      140.0
Name: AAPL, dtype: float64
```

In [21]:
```
stocks_concat_2.loc['AAPL', '2016']
```

Out[21]:
```
Shares     80.0
Low        95.0
High      110.0
Name: AAPL, dtype: float64
```

In [22]:
```
stocks_concat_2.loc['AAPL', '2016'].loc['Shares']
```

Out[22]:
```
80.0
```

(2) Merging Multiple Data Frames Together

In [ ]:
```
from IPython.display import HTML
HTML('<iframe src=https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Dat
```

In [23]:
```
faculty = pd.DataFrame({'faculty_n': ['01', '02', '03', '04'],
                        'Name': ['John', 'Jack', 'Jane', 'Grace'],
                        'Gender': ['M', 'M', 'F', 'F']})

course = pd.DataFrame({'class_n': ['DSA_01', 'DSA_02', 'DSA_03','DSA_04', 'DSA_05', 'DS
                       'faculty_n': ['01', '02', '03', '04', '02', '03'],
                       'date': ['MW', 'MWF', 'TT', 'Sat', 'TT', 'MW'],
                       'semester': ['spring', 'fall', 'fall', 'spring', 'spring', 'spri
```

In [24]:
```
faculty
```

Out[24]:

	faculty_n	Name	Gender
0	01	John	M
1	02	Jack	M

	faculty_n	Name	Gender
2	03	Jane	F
3	04	Grace	F

In [25]:
```
course
```

Out[25]:

	class_n	faculty_n	date	semester
0	DSA_01	01	MW	spring
1	DSA_02	02	MWF	fall
2	DSA_03	03	TT	fall
3	DSA_04	04	Sat	spring
4	DSA_05	02	TT	spring
5	DSA_06	03	MW	spring

In [26]:
```
result = pd.merge(faculty, course, on = 'faculty_n')
result
```

Out[26]:

	faculty_n	Name	Gender	class_n	date	semester
0	01	John	M	DSA_01	MW	spring
1	02	Jack	M	DSA_02	MWF	fall
2	02	Jack	M	DSA_05	TT	spring
3	03	Jane	F	DSA_03	TT	fall
4	03	Jane	F	DSA_06	MW	spring
5	04	Grace	F	DSA_04	Sat	spring

In [27]:
```
faculty = pd.DataFrame({'faculty_n': ['01', '02', '03', '04', '05', '06'],
                        'Name': ['John', 'Jack', 'Jane', 'Grace', 'Jeff', 'Sophia'],
                        'Gender': ['M', 'M', 'F', 'F', 'M', 'F']})

course = pd.DataFrame({'class_n': ['DSA_01', 'DSA_02', 'DSA_03','DSA_04', 'DSA_05', 'DS
                         'faculty_n': ['01', '02', '03', '04', '02', '03', np.nan],
                         'date': ['MW', 'MWF', 'TT', 'Sat', 'TT', 'MW', np.nan],
                         'semester': ['spring', 'fall', 'fall', 'spring', 'spring', 'spri
```

In [28]:
```
faculty
```

Out[28]:

	faculty_n	Name	Gender
0	01	John	M
1	02	Jack	M

	faculty_n	Name	Gender
2	03	Jane	F
3	04	Grace	F
4	05	Jeff	M
5	06	Sophia	F

In [29]:
```
course
```

Out[29]:

	class_n	faculty_n	date	semester
0	DSA_01	01	MW	spring
1	DSA_02	02	MWF	fall
2	DSA_03	03	TT	fall
3	DSA_04	04	Sat	spring
4	DSA_05	02	TT	spring
5	DSA_06	03	MW	spring
6	DSA_07	NaN	NaN	NaN

In [30]:
```
result = pd.merge(faculty, course, on = 'faculty_n')
result
```

Out[30]:

	faculty_n	Name	Gender	class_n	date	semester
0	01	John	M	DSA_01	MW	spring
1	02	Jack	M	DSA_02	MWF	fall
2	02	Jack	M	DSA_05	TT	spring
3	03	Jane	F	DSA_03	TT	fall
4	03	Jane	F	DSA_06	MW	spring
5	04	Grace	F	DSA_04	Sat	spring

In [31]:
```
result = pd.merge(faculty, course, how = 'outer', on = 'faculty_n')
result
```

Out[31]:

	faculty_n	Name	Gender	class_n	date	semester
0	01	John	M	DSA_01	MW	spring
1	02	Jack	M	DSA_02	MWF	fall
2	02	Jack	M	DSA_05	TT	spring
3	03	Jane	F	DSA_03	TT	fall
4	03	Jane	F	DSA_06	MW	spring

	faculty_n	Name	Gender	class_n	date	semester
5	04	Grace	F	DSA_04	Sat	spring
6	05	Jeff	M	NaN	NaN	NaN
7	06	Sophia	F	NaN	NaN	NaN
8	NaN	NaN	NaN	DSA_07	NaN	NaN

In [32]:
```
result = pd.merge(faculty, course, how = 'left', on = 'faculty_n')
result
```

Out[32]:

	faculty_n	Name	Gender	class_n	date	semester
0	01	John	M	DSA_01	MW	spring
1	02	Jack	M	DSA_02	MWF	fall
2	02	Jack	M	DSA_05	TT	spring
3	03	Jane	F	DSA_03	TT	fall
4	03	Jane	F	DSA_06	MW	spring
5	04	Grace	F	DSA_04	Sat	spring
6	05	Jeff	M	NaN	NaN	NaN
7	06	Sophia	F	NaN	NaN	NaN

In [33]:
```
result = pd.merge(faculty, course, how = 'right', on = 'faculty_n')
result
```

Out[33]:

	faculty_n	Name	Gender	class_n	date	semester
0	01	John	M	DSA_01	MW	spring
1	02	Jack	M	DSA_02	MWF	fall
2	03	Jane	F	DSA_03	TT	fall
3	04	Grace	F	DSA_04	Sat	spring
4	02	Jack	M	DSA_05	TT	spring
5	03	Jane	F	DSA_06	MW	spring
6	NaN	NaN	NaN	DSA_07	NaN	NaN