In [26]:

```
import pandas as ps
```

In [27]:

```
employee = ps.read_table("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Tab_separated_values.tsv")
employee.head()

cols = ["Name", "Position"]
employee[cols]
```

Out[27]:

	Name	Position
0	Airi Satou	Accountant
1	Angelica Ramos	Chief Executive Officer (CEO)
2	Ashton Cox	Junior Technical Author
3	Bradley Greer	Software Engineer
4	Brenden Wagner	Software Engineer
5	Brielle Williamson	Integration Specialist
6	Bruno Nash	Software Engineer
7	Caesar Vance	Pre-Sales Support
8	Cara Stevens	Sales Assistant
9	Cedric Kelly	Senior Javascript Developer
10	Charde Marshall	Regional Director
11	Colleen Hurst	Javascript Developer
12	Dai Rios	Personnel Lead
13	Donna Snider	Customer Support
14	Doris Wilder	Sales Assistant
15	Finn Camacho	Support Engineer
16	Fiona Green	Chief Operating Officer (COO)
17	Garrett Winters	Accountant
18	Gavin Cortez	Team Leader
19	Gavin Joyce	Developer
20	Gloria Little	Systems Administrator
21	Haley Kennedy	Senior Marketing Designer
22	Hermione Butler	Regional Director
23	Herrod Chandler	Sales Assistant
24	Hope Fuentes	Secretary
25	Howard Hatfield	Office Manager
26	Jackson Bradshaw	Director
27	Jena Gaines	Office Manager
28	Jenette Caldwell	Development Lead
29	Jennifer Acosta	Junior Javascript Developer
30	Jennifer Chang	Regional Director
31	Jonas Alexander	Developer
32	Lael Greer	Systems Administrator
33	Martena Mccray	Post-Sales support
34	Michael Bruce	Javascript Developer

	Name	Position
35	Michael Silva	Marketing Designer
36	Michelle House	Integration Specialist
37	Olivia Liang	Support Engineer
38	Paul Byrd	Chief Financial Officer (CFO)
39	Prescott Bartlett	Technical Author
40	Quinn Flynn	Support Lead
41	Rhona Davidson	Integration Specialist
42	Sakura Yamamoto	Support Engineer
43	Serge Baldwin	Data Coordinator
44	Shad Decker	Regional Director
45	Shou Itou	Regional Marketing
46	Sonya Frost	Software Engineer
47	Suki Burks	Developer
48	Tatyana Fitzpatrick	Regional Director
49	Thor Walton	Developer
50	Tiger Nixon	System Architect
51	Timothy Mooney	Office Manager
52	Unity Butler	Marketing Designer
53	Vivian Harrell	Financial Controller
54	Yuri Berry	Chief Marketing Officer (CMO)
55	Zenaida Frank	Software Engineer
56	Zorita Serrano	Software Engineer

In [28]:

```
employee.dtypes
```

Out[28]:

```
Name          object
Position      object
Office        object
Age            int64
Start date    object
Salary        object
dtype: object
```

In [29]:

```
import numpy as np
employee = ps.read_table("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Tab_separated_values.tsv")
#employee.select_dtypes(include=[np.number]).dtypes
```

In [11]:

```
employee.describe()
```

Out[11]:

	Age
count	57.000000
mean	42.736842
std	14.877507
min	19.000000
25%	30.000000
50%	42.000000
75%	56.000000
max	66.000000

In [15]:

```
#find out number of columns and rows
employee.shape
```

Out[15]:

```
(57, 6)
```

In [30]:

```
employee["Name Salary"] = employee["Name"]+employee["Salary"]
employee.head()
```

Out[30]:

	Name	Position	Office	Age	Start date	Salary	Name Salary
0	Airi Satou	Accountant	Tokyo	33	2008/11/28	$162,700	Airi Satou$162,700
1	Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09	$1,200,000	Angelica Ramos$1,200,000
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000	Ashton Cox$86,000
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000	Bradley Greer$132,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850	Brenden Wagner$206,850

In [33]:
```
employee.drop(['Age'], axis=1, inplace=True)
employee.head()
```

Out[33]:

	Name	Position	Office	Start date	Salary	Name Salary
0	Airi Satou	Accountant	Tokyo	2008/11/28	$162,700	Airi Satou$162,700
1	Angelica Ramos	Chief Executive Officer (CEO)	London	2009/10/09	$1,200,000	Angelica Ramos$1,200,000
2	Ashton Cox	Junior Technical Author	San Francisco	2009/01/12	$86,000	Ashton Cox$86,000
3	Bradley Greer	Software Engineer	London	2012/10/13	$132,000	Bradley Greer$132,000
4	Brenden Wagner	Software Engineer	San Francisco	2011/06/07	$206,850	Brenden Wagner$206,850

In [35]:
```
employee = ps.read_table("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkil
lDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Ta
b_separated_values.tsv")
employee.head()
cols = ["Name", "Pos", "Ofc", "Age", "S_date", "Sal"]
employee.columns = cols
employee.head()
```

Out[35]:

	Name	Pos	Ofc	Age	S_date	Sal
0	Airi Satou	Accountant	Tokyo	33	2008/11/28	$162,700
1	Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09	$1,200,000
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850

In [36]:

```
employee.sort_values(by="Age", ascending=True)
```

Out[36]:

	Name	Pos	Ofc	Age	S_date	Sal
48	Tatyana Fitzpatrick	Regional Director	London	19	2010/03/17	$385,750
45	Shou Itou	Regional Marketing	Tokyo	20	2011/08/14	$163,000
7	Caesar Vance	Pre-Sales Support	New York	21	2011/12/12	$106,450
32	Lael Greer	Systems Administrator	London	21	2009/02/27	$103,500
18	Gavin Cortez	Team Leader	San Francisco	22	2008/10/26	$235,500
40	Quinn Flynn	Support Lead	Edinburgh	22	2013/03/03	$342,000
9	Cedric Kelly	Senior Javascript Developer	Edinburgh	22	2012/03/29	$433,060
46	Sonya Frost	Software Engineer	Edinburgh	23	2008/12/13	$103,600
14	Doris Wilder	Sales Assistant	Sidney	23	2010/09/20	$85,600
39	Prescott Bartlett	Technical Author	London	27	2011/05/07	$145,000
13	Donna Snider	Customer Support	New York	27	2011/01/25	$112,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850
30	Jennifer Chang	Regional Director	Singapore	28	2010/11/14	$357,650
34	Michael Bruce	Javascript Developer	Singapore	29	2011/06/27	$183,000
27	Jena Gaines	Office Manager	London	30	2008/12/19	$90,560
31	Jonas Alexander	Developer	San Francisco	30	2010/07/14	$86,500
28	Jenette Caldwell	Development Lead	New York	30	2011/09/03	$345,000
0	Airi Satou	Accountant	Tokyo	33	2008/11/28	$162,700
12	Dai Rios	Personnel Lead	Edinburgh	35	2012/09/26	$217,500
10	Charde Marshall	Regional Director	San Francisco	36	2008/10/16	$470,600
36	Michelle House	Integration Specialist	Sidney	37	2011/06/02	$95,400
51	Timothy Mooney	Office Manager	London	37	2008/12/11	$136,200
42	Sakura Yamamoto	Support Engineer	Tokyo	37	2009/08/19	$139,575
6	Bruno Nash	Software Engineer	London	38	2011/05/03	$163,500
11	Colleen Hurst	Javascript Developer	San Francisco	39	2009/09/15	$205,500
54	Yuri Berry	Chief Marketing Officer (CMO)	New York	40	2009/06/25	$675,000
24	Hope Fuentes	Secretary	San Francisco	41	2010/02/12	$109,850
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000
19	Gavin Joyce	Developer	Edinburgh	42	2010/12/22	$92,575
29	Jennifer Acosta	Junior Javascript Developer	Edinburgh	43	2013/02/01	$75,650
21	Haley Kennedy	Senior Marketing Designer	London	43	2012/12/18	$313,500
33	Martena Mccray	Post-Sales support	Edinburgh	46	2011/03/09	$324,050
8	Cara Stevens	Sales Assistant	New York	46	2011/12/06	$145,600
52	Unity Butler	Marketing Designer	San Francisco	47	2009/12/09	$85,675
22	Hermione Butler	Regional Director	London	47	2011/03/21	$356,250

	Name	Pos	Ofc	Age	S_date	Sal
1	Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09	$1,200,000
15	Finn Camacho	Support Engineer	San Francisco	47	2009/07/07	$87,500
16	Fiona Green	Chief Operating Officer (COO)	San Francisco	48	2010/03/11	$850,000
44	Shad Decker	Regional Director	Edinburgh	51	2008/11/13	$183,000
25	Howard Hatfield	Office Manager	San Francisco	51	2008/12/16	$164,500
47	Suki Burks	Developer	London	53	2009/10/22	$114,500
41	Rhona Davidson	Integration Specialist	Tokyo	55	2010/10/14	$327,900
56	Zorita Serrano	Software Engineer	San Francisco	56	2012/06/01	$115,000
23	Herrod Chandler	Sales Assistant	San Francisco	59	2012/08/06	$137,500
20	Gloria Little	Systems Administrator	New York	59	2009/04/10	$237,500
49	Thor Walton	Developer	New York	61	2013/08/11	$98,540
50	Tiger Nixon	System Architect	Edinburgh	61	2011/04/25	$320,800
5	Brielle Williamson	Integration Specialist	New York	61	2012/12/02	$372,000
53	Vivian Harrell	Financial Controller	San Francisco	62	2009/02/14	$452,500
17	Garrett Winters	Accountant	Tokyo	63	2011/07/25	$170,750
55	Zenaida Frank	Software Engineer	New York	63	2010/01/04	$125,250
43	Serge Baldwin	Data Coordinator	Singapore	64	2012/04/09	$138,575
38	Paul Byrd	Chief Financial Officer (CFO)	New York	64	2010/06/09	$725,000
37	Olivia Liang	Support Engineer	Singapore	64	2011/02/03	$234,500
26	Jackson Bradshaw	Director	New York	65	2008/09/26	$645,750
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000
35	Michael Silva	Marketing Designer	London	66	2012/11/27	$198,500

In [39]:
```
employee["Name"].sort_values().head()
```

Out[39]:
```
0          Airi Satou
1      Angelica Ramos
2          Ashton Cox
3       Bradley Greer
4      Brenden Wagner
Name: Name, dtype: object
```

In [42]:
```
b
```

Out[42]:

	Name	Pos	Ofc	Age	S_date	Sal
48	Tatyana Fitzpatrick	Regional Director	London	19	2010/03/17	$385,750
45	Shou Itou	Regional Marketing	Tokyo	20	2011/08/14	$163,000
7	Caesar Vance	Pre-Sales Support	New York	21	2011/12/12	$106,450
32	Lael Greer	Systems Administrator	London	21	2009/02/27	$103,500
9	Cedric Kelly	Senior Javascript Developer	Edinburgh	22	2012/03/29	$433,060
18	Gavin Cortez	Team Leader	San Francisco	22	2008/10/26	$235,500
40	Quinn Flynn	Support Lead	Edinburgh	22	2013/03/03	$342,000
14	Doris Wilder	Sales Assistant	Sidney	23	2010/09/20	$85,600
46	Sonya Frost	Software Engineer	Edinburgh	23	2008/12/13	$103,600
13	Donna Snider	Customer Support	New York	27	2011/01/25	$112,000
39	Prescott Bartlett	Technical Author	London	27	2011/05/07	$145,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850
30	Jennifer Chang	Regional Director	Singapore	28	2010/11/14	$357,650
34	Michael Bruce	Javascript Developer	Singapore	29	2011/06/27	$183,000

In [43]:
```
employee[(employee.Age < 30 ) & (employee.Ofc == "London")]
```

Out[43]:

	Name	Pos	Ofc	Age	S_date	Sal
32	Lael Greer	Systems Administrator	London	21	2009/02/27	$103,500
39	Prescott Bartlett	Technical Author	London	27	2011/05/07	$145,000
48	Tatyana Fitzpatrick	Regional Director	London	19	2010/03/17	$385,750

In [44]:
```
#filter columns
cols = ["Name", "Pos"]
employee[(employee.Age<40) & (employee.Ofc=="London")][cols]
```

Out[44]:

	Name	Pos
6	Bruno Nash	Software Engineer
27	Jena Gaines	Office Manager
32	Lael Greer	Systems Administrator
39	Prescott Bartlett	Technical Author
48	Tatyana Fitzpatrick	Regional Director
51	Timothy Mooney	Office Manager

In [46]:

```
#chapter 6 - •  Filtering
employee[(employee.Name!="Airi Satou") & (employee.Name!="Angelica Ramos") & (
employee.Name!="Bradley Greer")]
```

Out[46]:

	Name	Pos	Ofc	Age	S_date	Sal
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850
5	Brielle Williamson	Integration Specialist	New York	61	2012/12/02	$372,000
6	Bruno Nash	Software Engineer	London	38	2011/05/03	$163,500
7	Caesar Vance	Pre-Sales Support	New York	21	2011/12/12	$106,450
8	Cara Stevens	Sales Assistant	New York	46	2011/12/06	$145,600
9	Cedric Kelly	Senior Javascript Developer	Edinburgh	22	2012/03/29	$433,060
10	Charde Marshall	Regional Director	San Francisco	36	2008/10/16	$470,600
11	Colleen Hurst	Javascript Developer	San Francisco	39	2009/09/15	$205,500
12	Dai Rios	Personnel Lead	Edinburgh	35	2012/09/26	$217,500
13	Donna Snider	Customer Support	New York	27	2011/01/25	$112,000
14	Doris Wilder	Sales Assistant	Sidney	23	2010/09/20	$85,600
15	Finn Camacho	Support Engineer	San Francisco	47	2009/07/07	$87,500
16	Fiona Green	Chief Operating Officer (COO)	San Francisco	48	2010/03/11	$850,000
17	Garrett Winters	Accountant	Tokyo	63	2011/07/25	$170,750
18	Gavin Cortez	Team Leader	San Francisco	22	2008/10/26	$235,500
19	Gavin Joyce	Developer	Edinburgh	42	2010/12/22	$92,575
20	Gloria Little	Systems Administrator	New York	59	2009/04/10	$237,500
21	Haley Kennedy	Senior Marketing Designer	London	43	2012/12/18	$313,500
22	Hermione Butler	Regional Director	London	47	2011/03/21	$356,250
23	Herrod Chandler	Sales Assistant	San Francisco	59	2012/08/06	$137,500
24	Hope Fuentes	Secretary	San Francisco	41	2010/02/12	$109,850
25	Howard Hatfield	Office Manager	San Francisco	51	2008/12/16	$164,500
26	Jackson Bradshaw	Director	New York	65	2008/09/26	$645,750
27	Jena Gaines	Office Manager	London	30	2008/12/19	$90,560
28	Jenette Caldwell	Development Lead	New York	30	2011/09/03	$345,000
29	Jennifer Acosta	Junior Javascript Developer	Edinburgh	43	2013/02/01	$75,650
30	Jennifer Chang	Regional Director	Singapore	28	2010/11/14	$357,650
31	Jonas Alexander	Developer	San Francisco	30	2010/07/14	$86,500
32	Lael Greer	Systems Administrator	London	21	2009/02/27	$103,500
33	Martena Mccray	Post-Sales support	Edinburgh	46	2011/03/09	$324,050
34	Michael Bruce	Javascript Developer	Singapore	29	2011/06/27	$183,000
35	Michael Silva	Marketing Designer	London	66	2012/11/27	$198,500
36	Michelle House	Integration Specialist	Sidney	37	2011/06/02	$95,400
37	Olivia Liang	Support Engineer	Singapore	64	2011/02/03	$234,500

	Name	Pos	Ofc	Age	S_date	Sal
38	Paul Byrd	Chief Financial Officer (CFO)	New York	64	2010/06/09	$725,000
39	Prescott Bartlett	Technical Author	London	27	2011/05/07	$145,000
40	Quinn Flynn	Support Lead	Edinburgh	22	2013/03/03	$342,000
41	Rhona Davidson	Integration Specialist	Tokyo	55	2010/10/14	$327,900
42	Sakura Yamamoto	Support Engineer	Tokyo	37	2009/08/19	$139,575
43	Serge Baldwin	Data Coordinator	Singapore	64	2012/04/09	$138,575
44	Shad Decker	Regional Director	Edinburgh	51	2008/11/13	$183,000
45	Shou Itou	Regional Marketing	Tokyo	20	2011/08/14	$163,000
46	Sonya Frost	Software Engineer	Edinburgh	23	2008/12/13	$103,600
47	Suki Burks	Developer	London	53	2009/10/22	$114,500
48	Tatyana Fitzpatrick	Regional Director	London	19	2010/03/17	$385,750
49	Thor Walton	Developer	New York	61	2013/08/11	$98,540
50	Tiger Nixon	System Architect	Edinburgh	61	2011/04/25	$320,800
51	Timothy Mooney	Office Manager	London	37	2008/12/11	$136,200
52	Unity Butler	Marketing Designer	San Francisco	47	2009/12/09	$85,675
53	Vivian Harrell	Financial Controller	San Francisco	62	2009/02/14	$452,500
54	Yuri Berry	Chief Marketing Officer (CMO)	New York	40	2009/06/25	$675,000
55	Zenaida Frank	Software Engineer	New York	63	2010/01/04	$125,250
56	Zorita Serrano	Software Engineer	San Francisco	56	2012/06/01	$115,000

In [47]:

```
#chapter 6 - mean()
employee.mean()
```

Out[47]:

```
Age    42.736842
dtype: float64
```

In [49]:

```
#Chapter 6 - •  String manipulation techniques (upper, lower, replace)
employee.Name.str.upper().head()
```

Out[49]:

```
0          AIRI SATOU
1      ANGELICA RAMOS
2          ASHTON COX
3       BRADLEY GREER
4      BRENDEN WAGNER
Name: Name, dtype: object
```

In [52]:

```
employee.Pos.str.contains("Software").head()
```

Out[52]:

```
0    False
1    False
2    False
3     True
4     True
Name: Pos, dtype: bool
```

In [57]:
```
employee[employee.Pos.str.contains("Software")].head()
```

Out[57]:

	Name	Pos	Ofc	Age	S_date	Sal
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850
6	Bruno Nash	Software Engineer	London	38	2011/05/03	$163,500
46	Sonya Frost	Software Engineer	Edinburgh	23	2008/12/13	$103,600
55	Zenaida Frank	Software Engineer	New York	63	2010/01/04	$125,250

In [55]:
```
#Chapter 6- o   String replace
employee.Pos.str.replace("Engineer", "Developer").head()
```

Out[55]:
```
0                    Accountant
1     Chief Executive Officer (CEO)
2           Junior Technical Author
3                Software Developer
4                Software Developer
Name: Pos, dtype: object
```

In [59]:
```
#chapter 6 - Aggregations
employee.Age.min()
employee.Age.max()
```

Out[59]:
```
19
```

In [61]:
```
#chapter 6 - •  Groupby clauses (like you use in SQL Aggregations)
employee.groupby("Pos").Age.min().head()
```

Out[61]:
```
Pos
Accountant                       33
Chief Executive Officer (CEO)    47
Chief Financial Officer (CFO)    64
Chief Marketing Officer (CMO)    40
Chief Operating Officer (COO)    48
Name: Age, dtype: int64
```

In [64]:
```
#chapter 6 - •  Groupby clauses (like you use in SQL Aggregations)
employee.groupby("Pos").Age.agg(['count', 'min', 'max']).head()
```

Out[64]:

	count	min	max
Pos			
Accountant	2	33	63
Chief Executive Officer (CEO)	1	47	47
Chief Financial Officer (CFO)	1	64	64
Chief Marketing Officer (CMO)	1	40	40
Chief Operating Officer (COO)	1	48	48

In [68]:

```
#chapter homework - calculate the mean rowwise and columnwise
import numpy as np
iris = ps.read_table("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDev
elopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Iris.c
sv", sep=",")
```

In [66]:

```
iris.head()
```

Out[66]:

	sepal length,sepal width,petal length,petal width,iris,,,,
0	5.1,3.5,1.4,0.2,Iris-setosa,,,,
1	4.9,3,1.4,0.2,Iris-setosa,,,,
2	4.7,3.2,1.3,0.2,Iris-setosa,,,,
3	4.6,3.1,1.5,0.2,Iris-setosa,,,,
4	5,3.6,1.4,0.2,Iris-setosa,,,,

In [69]:

```
iris.mean()
```

Out[69]:

```
sepal length                            (5.8433333333333355+0j)
sepal width                              (3.054000000000001+0j)
petal length                            (3.7586666666666697+0j)
petal width                             (1.1986666666666674+0j)
iris            (8.30747206291045e-309+9.197469912385895e-309j)
Unnamed: 5                                                  NaN
Unnamed: 6                                                  NaN
Unnamed: 7                                                  NaN
Unnamed: 8                                                  NaN
dtype: complex128
```

In [72]:

```
#add a new column to store mean that's calculated by rows
iris["mean"]=iris.mean(axis=1)
iris.head()
```

Out[72]:

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	mean
0	5.1	3.5	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
1	4.9	3.0	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.375
2	4.7	3.2	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
3	4.6	3.1	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
4	5.0	3.6	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550

In [73]:

```
#calculate mean by columns
iris.mean(axis=0)
```

Out[73]:

```
sepal length                             (5.8433333333333355+0j)
sepal width                               (3.054000000000001+0j)
petal length                             (3.7586666666666697+0j)
petal width                              (1.1986666666666674+0j)
iris            (2.976469694290114e-88+2.2373972069411116e-145j)
Unnamed: 5                                                   NaN
Unnamed: 6                                                   NaN
Unnamed: 7                                                   NaN
Unnamed: 8                                                   NaN
mean                                      (3.4636666666666666+0j)
dtype: complex128
```

In [75]:

```
#chapter 7 - loc : select rows and columns, like table slicing
iris.loc[0, :]
```

Out[75]:

```
sepal length            5.1
sepal width             3.5
petal length            1.4
petal width             0.2
iris            Iris-setosa
Unnamed: 5              NaN
Unnamed: 6              NaN
Unnamed: 7              NaN
Unnamed: 8              NaN
mean                   2.55
Name: 0, dtype: object
```

In [76]:

```
iris.loc[0:3, :]
```

Out[76]:

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	mean
0	5.1	3.5	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
1	4.9	3.0	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.375
2	4.7	3.2	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
3	4.6	3.1	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350

In [81]:

```
iris.loc[0:3, ["sepal length", "petal width", "iris"]]
```

Out[81]:

	sepal length	petal width	iris
0	5.1	0.2	Iris-setosa
1	4.9	0.2	Iris-setosa
2	4.7	0.2	Iris-setosa
3	4.6	0.2	Iris-setosa

In [83]:

```
iris.loc[iris.iris=="Iris-setosa"]
```

Out[83]:

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	mean
0	5.1	3.5	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
1	4.9	3.0	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.375
2	4.7	3.2	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
3	4.6	3.1	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
4	5.0	3.6	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
5	5.4	3.9	1.7	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.850
6	4.6	3.4	1.4	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.425
7	5.0	3.4	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.525
8	4.4	2.9	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.225
9	4.9	3.1	1.5	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.400
10	5.4	3.7	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.700
11	4.8	3.4	1.6	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.500
12	4.8	3.0	1.4	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.325
13	4.3	3.0	1.1	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.125
14	5.8	4.0	1.2	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.800
15	5.7	4.4	1.5	0.4	Iris-setosa	NaN	NaN	NaN	NaN	3.000
16	5.4	3.9	1.3	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.750
17	5.1	3.5	1.4	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.575
18	5.7	3.8	1.7	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.875
19	5.1	3.8	1.5	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.675
20	5.4	3.4	1.7	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.675
21	5.1	3.7	1.5	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.675
22	4.6	3.6	1.0	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	mean
23	5.1	3.3	1.7	0.5	Iris-setosa	NaN	NaN	NaN	NaN	2.650
24	4.8	3.4	1.9	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.575
25	5.0	3.0	1.6	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.450
26	5.0	3.4	1.6	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.600
27	5.2	3.5	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.600
28	5.2	3.4	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
29	4.7	3.2	1.6	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.425
30	4.8	3.1	1.6	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.425
31	5.4	3.4	1.5	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.675
32	5.2	4.1	1.5	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.725
33	5.5	4.2	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.825
34	4.9	3.1	1.5	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.400
35	5.0	3.2	1.2	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.400
36	5.5	3.5	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.625
37	4.9	3.1	1.5	0.1	Iris-setosa	NaN	NaN	NaN	NaN	2.400
38	4.4	3.0	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.225
39	5.1	3.4	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.550
40	5.0	3.5	1.3	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.525
41	4.5	2.3	1.3	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.100
42	4.4	3.2	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.275
43	5.0	3.5	1.6	0.6	Iris-setosa	NaN	NaN	NaN	NaN	2.675
44	5.1	3.8	1.9	0.4	Iris-setosa	NaN	NaN	NaN	NaN	2.800
45	4.8	3.0	1.4	0.3	Iris-setosa	NaN	NaN	NaN	NaN	2.375

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	mean
46	5.1	3.8	1.6	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.675
47	4.6	3.2	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.350
48	5.3	3.7	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.675
49	5.0	3.3	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN	2.475

In [87]:

```
employee = ps.read_table("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Tab_separated_values_missingvalues.tsv")
employee.head()
employee.shape
```

Out[87]:

```
(57, 6)
```

In [89]:

```
#drop rows with any empty value
employee.dropna(how="any").shape
```

Out[89]:

```
(54, 6)
```

In [91]:

```
employee.fillna(value=0).head()
```

Out[91]:

	Name	Position	Office	Age	Start date	Salary
0	Airi Satou	Accountant	Tokyo	33	2008/11/28	0
1	Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09	0
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	0

In [93]:

```
#chapter 8 - work with plot
iris = ps.read_csv("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/Iris.csv", sep=",")
iris.head()
iris.shape
```

Out[93]:

```
(150, 9)
```

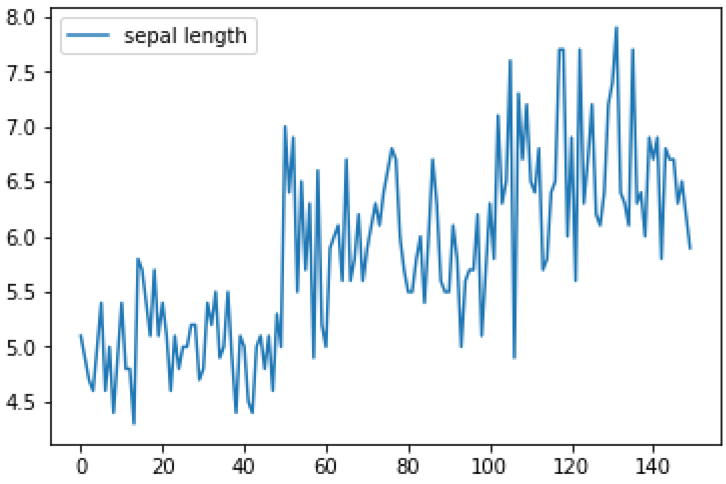
In [97]:
```
iris.head()
```

Out[97]:

	sepal length	sepal width	petal length	petal width	iris	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8
0	5.1	3.5	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN
1	4.9	3.0	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN
2	4.7	3.2	1.3	0.2	Iris-setosa	NaN	NaN	NaN	NaN
3	4.6	3.1	1.5	0.2	Iris-setosa	NaN	NaN	NaN	NaN
4	5.0	3.6	1.4	0.2	Iris-setosa	NaN	NaN	NaN	NaN

In [100]:
```
cols=["sepal length", "iris"]
iris[cols].plot()
```

Out[100]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x164ce06e940>
```



In [103]:
```
#join tables
dataframe1 = ps.DataFrame({"employee":["ABC", "XYZ", "MNO"], "age":[22, 35, 43
]})
dataframe1
```

Out[103]:

	employee	age
0	ABC	22
1	XYZ	35
2	MNO	43

In [105]:

```
dataframe2 = ps.DataFrame({"employee":["ABC", "XYZ", "EFG"], "salary":[10000,
30000, 25000]})
dataframe2
```

Out[105]:

	employee	salary
0	ABC	10000
1	XYZ	30000
2	EFG	25000

In [107]:

```
#inner join
dataframe3 = ps.merge(dataframe1, dataframe2, on="employee")
dataframe3
```

Out[107]:

	employee	age	salary
0	ABC	22	10000
1	XYZ	35	30000

In [108]:

```
#outer join
dataframe4 = ps.merge(dataframe1, dataframe2, how="outer")
dataframe4
```

Out[108]:

	employee	age	salary
0	ABC	22.0	10000.0
1	XYZ	35.0	30000.0
2	MNO	43.0	NaN
3	EFG	NaN	25000.0

In [109]:

```
dataframe3 = ps.merge(dataframe1, dataframe2, how="left")
dataframe3
```

Out[109]:

	employee	age	salary
0	ABC	22	10000.0
1	XYZ	35	30000.0
2	MNO	43	NaN

In [110]:

```
dataframe3 = ps.merge(dataframe1, dataframe2, how="right")
dataframe3
```

Out[110]:

	employee	age	salary
0	ABC	22.0	10000
1	XYZ	35.0	30000
2	EFG	NaN	25000

In [114]:

```
#chapter 9 - pivot data
webhits = ps.read_csv("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/webhits.csv")
webhits
```

Out[114]:

	Date	Page_Name	Hits
0	1/1/2019	ABC	33
1	1/2/2019	EFG	22
2	1/3/2019	XYZ	44
3	1/1/2019	MNO	55
4	1/2/2019	ABC	21
5	1/3/2019	EFG	30
6	1/1/2019	XYZ	40
7	1/2/2019	MNO	51
8	1/3/2019	ABC	34
9	1/1/2019	EFG	67
10	1/2/2019	XYZ	25
11	1/3/2019	MNO	43

In [115]:

```
#pivode data
webhits.pivot(index="Page_Name", columns="Date")
```

Out[115]:

	Hits		
Date	**1/1/2019**	**1/2/2019**	**1/3/2019**
Page_Name			
ABC	33	21	34
EFG	67	22	30
MNO	55	51	43
XYZ	40	25	44

In [116]:

```
##pivot table
webhits.pivot_table(index="Page_Name", aggfunc="sum")
```

Out[116]:

```
           Hits
Page_Name
      ABC    88
      EFG   119
      MNO   149
      XYZ   109
```

In [117]:

```
webhits.pivot_table(index="Page_Name", aggfunc="mean")
```

Out[117]:

```
           Hits
Page_Name
      ABC  29.333333
      EFG  39.666667
      MNO  49.666667
      XYZ  36.333333
```

In [120]:

```
webhits.pivot_table(index="Page_Name", aggfunc="count")
```

Out[120]:

```
           Date  Hits
Page_Name
      ABC     3     3
      EFG     3     3
      MNO     3     3
      XYZ     3     3
```

In [121]:
```
webhits
```

Out[121]:

	Date	Page_Name	Hits
0	1/1/2019	ABC	33
1	1/2/2019	EFG	22
2	1/3/2019	XYZ	44
3	1/1/2019	MNO	55
4	1/2/2019	ABC	21
5	1/3/2019	EFG	30
6	1/1/2019	XYZ	40
7	1/2/2019	MNO	51
8	1/3/2019	ABC	34
9	1/1/2019	EFG	67
10	1/2/2019	XYZ	25
11	1/3/2019	MNO	43

In [122]:
```
webhits.shift(1)
```

Out[122]:

	Date	Page_Name	Hits
0	NaN	NaN	NaN
1	1/1/2019	ABC	33.0
2	1/2/2019	EFG	22.0
3	1/3/2019	XYZ	44.0
4	1/1/2019	MNO	55.0
5	1/2/2019	ABC	21.0
6	1/3/2019	EFG	30.0
7	1/1/2019	XYZ	40.0
8	1/2/2019	MNO	51.0
9	1/3/2019	ABC	34.0
10	1/1/2019	EFG	67.0
11	1/2/2019	XYZ	25.0

In [123]:
```
webhits.shift(-1)
```

Out[123]:

	Date	Page_Name	Hits
0	1/2/2019	EFG	22.0
1	1/3/2019	XYZ	44.0
2	1/1/2019	MNO	55.0
3	1/2/2019	ABC	21.0
4	1/3/2019	EFG	30.0
5	1/1/2019	XYZ	40.0
6	1/2/2019	MNO	51.0
7	1/3/2019	ABC	34.0
8	1/1/2019	EFG	67.0
9	1/2/2019	XYZ	25.0
10	1/3/2019	MNO	43.0
11	NaN	NaN	NaN

In [125]:
```
webhits.shift(-1).to_csv("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/webhits_shiftup.csv")
```

In [126]:
```
webhits.shift(2).to_csv("C:/MyDocuments/ProfessionalDevelopment/TechnicalSkillDevelopment/DataScience/UdemyCourses/LearnDataAnalysisWithPandas/Resources/webhits_shift_somecolumns.csv", columns=["Date", "Page_Name"])
```

In [129]:

```
#use shift to calculate delta
webhits["Prev_hits"]=webhits["Hits"].shift(1)
webhits
```

Out[129]:

	Date	Page_Name	Hits	prev_hits	Prev_hits
0	1/1/2019	ABC	33	NaN	NaN
1	1/2/2019	EFG	22	33.0	33.0
2	1/3/2019	XYZ	44	22.0	22.0
3	1/1/2019	MNO	55	44.0	44.0
4	1/2/2019	ABC	21	55.0	55.0
5	1/3/2019	EFG	30	21.0	21.0
6	1/1/2019	XYZ	40	30.0	30.0
7	1/2/2019	MNO	51	40.0	40.0
8	1/3/2019	ABC	34	51.0	51.0
9	1/1/2019	EFG	67	34.0	34.0
10	1/2/2019	XYZ	25	67.0	67.0
11	1/3/2019	MNO	43	25.0	25.0

In [130]:

```
webhits["Hits_delta"] = webhits["Hits"]-webhits["Prev_hits"]
webhits
```

Out[130]:

	Date	Page_Name	Hits	prev_hits	Prev_hits	Hits_delta
0	1/1/2019	ABC	33	NaN	NaN	NaN
1	1/2/2019	EFG	22	33.0	33.0	-11.0
2	1/3/2019	XYZ	44	22.0	22.0	22.0
3	1/1/2019	MNO	55	44.0	44.0	11.0
4	1/2/2019	ABC	21	55.0	55.0	-34.0
5	1/3/2019	EFG	30	21.0	21.0	9.0
6	1/1/2019	XYZ	40	30.0	30.0	10.0
7	1/2/2019	MNO	51	40.0	40.0	11.0
8	1/3/2019	ABC	34	51.0	51.0	-17.0
9	1/1/2019	EFG	67	34.0	34.0	33.0
10	1/2/2019	XYZ	25	67.0	67.0	-42.0
11	1/3/2019	MNO	43	25.0	25.0	18.0