```
import pandas as pd
```

```
penguins = pd.read_csv("penguins.csv")
```

```
# preview first 5 rows
penguins.head()
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007

```
# preview last 5 rows
penguins.tail()
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
339	340	Chinstrap	Dream	55.8	19.8	207.0	4000.0	male	2009
340	341	Chinstrap	Dream	43.5	18.1	202.0	3400.0	female	2009
341	342	Chinstrap	Dream	49.6	18.2	193.0	3775.0	male	2009
342	343	Chinstrap	Dream	50.8	19.0	210.0	4100.0	male	2009
343	344	Chinstrap	Dream	50.2	18.7	198.0	3775.0	female	2009

```
#shap of dataframe
penguins.shape
```

```
(344, 9)
```

```
penguins.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 9 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   rowid              344 non-null    int64  
 1   species            344 non-null    object 
 2   island             344 non-null    object 
 3   bill_length_mm     342 non-null    float64
 4   bill_depth_mm      342 non-null    float64
 5   flipper_length_mm  342 non-null    float64
 6   body_mass_g        342 non-null    float64
 7   sex                333 non-null    object 
 8   year               344 non-null    int64  
dtypes: float64(4), int64(2), object(3)
memory usage: 24.3+ KB
```

```
# select columns
penguins["species"]
penguins.species
```

```
0        Adelie
1        Adelie
2        Adelie
3        Adelie
4        Adelie
        ...
339    Chinstrap
340    Chinstrap
341    Chinstrap
342    Chinstrap
343    Chinstrap
Name: species, Length: 344, dtype: object
```

```
# select many columns
penguins[ ['species','island','sex']].head(10)
```

	species	island	sex
0	Adelie	Torgersen	male
1	Adelie	Torgersen	female
2	Adelie	Torgersen	female
3	Adelie	Torgersen	NaN
4	Adelie	Torgersen	female
5	Adelie	Torgersen	male
6	Adelie	Torgersen	female
7	Adelie	Torgersen	male
8	Adelie	Torgersen	NaN
9	Adelie	Torgersen	NaN

```
# Integer location based indexing (iloc)
penguins.iloc[ 0 : 5, [1,2,5]]
```

	species	island	flipper_length_mm
0	Adelie	Torgersen	181.0
1	Adelie	Torgersen	186.0
2	Adelie	Torgersen	195.0
3	Adelie	Torgersen	NaN
4	Adelie	Torgersen	193.0

```
# filter rows by a condition
penguins['island'] == 'Torgersen'
```

```
0       True
1       True
2       True
3       True
4       True
       ...
339    False
340    False
341    False
342    False
343    False
Name: island, Length: 344, dtype: bool
```

```
# filter rows by a condition
penguins[penguins['island'] == 'Torgersen']
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007
6	7	Adelie	Torgersen	38.9	17.8	181.0	3625.0	female	2007
7	8	Adelie	Torgersen	39.2	19.6	195.0	4675.0	male	2007
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
9	10	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN	2007
10	11	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN	2007
11	12	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN	2007
12	13	Adelie	Torgersen	41.1	17.6	182.0	3200.0	female	2007
13	14	Adelie	Torgersen	38.6	21.2	191.0	3800.0	male	2007
14	15	Adelie	Torgersen	34.6	21.1	198.0	4400.0	male	2007
15	16	Adelie	Torgersen	36.6	17.8	185.0	3700.0	female	2007
16	17	Adelie	Torgersen	38.7	19.0	195.0	3450.0	female	2007
17	18	Adelie	Torgersen	42.5	20.7	197.0	4500.0	male	2007
18	19	Adelie	Torgersen	34.4	18.4	184.0	3325.0	female	2007
19	20	Adelie	Torgersen	46.0	21.5	194.0	4200.0	male	2007
68	69	Adelie	Torgersen	35.9	16.6	190.0	3050.0	female	2008
69	70	Adelie	Torgersen	41.8	19.4	198.0	4450.0	male	2008
70	71	Adelie	Torgersen	33.5	19.0	190.0	3600.0	female	2008
71	72	Adelie	Torgersen	39.7	18.4	190.0	3900.0	male	2008
72	73	Adelie	Torgersen	39.6	17.2	196.0	3550.0	female	2008
73	74	Adelie	Torgersen	45.8	18.9	197.0	4150.0	male	2008
74	75	Adelie	Torgersen	35.5	17.5	190.0	3700.0	female	2008
75	76	Adelie	Torgersen	42.8	18.5	195.0	4250.0	male	2008
76	77	Adelie	Torgersen	40.9	16.8	191.0	3700.0	female	2008
77	78	Adelie	Torgersen	37.2	19.4	184.0	3900.0	male	2008
78	79	Adelie	Torgersen	36.2	16.1	187.0	3550.0	female	2008
79	80	Adelie	Torgersen	42.1	19.1	195.0	4000.0	male	2008
80	81	Adelie	Torgersen	34.6	17.2	189.0	3200.0	female	2008

81	82	Adelie	Torgersen	42.9	17.6	196.0	4700.0	male	2008

```
# filter rows by a condition
penguins[penguins['bill_length_mm'] > 34]
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007
...	...	...	...	...	...	...	...	...	...
339	340	Chinstrap	Dream	55.8	19.8	207.0	4000.0	male	2009
340	341	Chinstrap	Dream	43.5	18.1	202.0	3400.0	female	2009
341	342	Chinstrap	Dream	49.6	18.2	193.0	3775.0	male	2009
342	343	Chinstrap	Dream	50.8	19.0	210.0	4100.0	male	2009
343	344	Chinstrap	Dream	50.2	18.7	198.0	3775.0	female	2009

338 rows × 9 columns

131	132	Adelie	Torgersen	43.1	19.2	197.0	3500.0	male	2009

```
# filter more than one condition
penguins[(penguins['island'] == 'Torgersen') & (penguins['bill_length_mm'] < 35)]
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
14	15	Adelie	Torgersen	34.6	21.1	198.0	4400.0	male	2007
18	19	Adelie	Torgersen	34.4	18.4	184.0	3325.0	female	2007
70	71	Adelie	Torgersen	33.5	19.0	190.0	3600.0	female	2008
80	81	Adelie	Torgersen	34.6	17.2	189.0	3200.0	female	2008

```
# filter more than one condition
penguins[(penguins['island'] == 'Torgersen') | (penguins['bill_length_mm'] < 35)]
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007
6	7	Adelie	Torgersen	38.9	17.8	181.0	3625.0	female	2007
7	8	Adelie	Torgersen	39.2	19.6	195.0	4675.0	male	2007
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
9	10	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN	2007
10	11	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN	2007
11	12	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN	2007
12	13	Adelie	Torgersen	41.1	17.6	182.0	3200.0	female	2007
13	14	Adelie	Torgersen	38.6	21.2	191.0	3800.0	male	2007
14	15	Adelie	Torgersen	34.6	21.1	198.0	4400.0	male	2007
15	16	Adelie	Torgersen	36.6	17.8	185.0	3700.0	female	2007
16	17	Adelie	Torgersen	38.7	19.0	195.0	3450.0	female	2007
17	18	Adelie	Torgersen	42.5	20.7	197.0	4500.0	male	2007
18	19	Adelie	Torgersen	34.4	18.4	184.0	3325.0	female	2007
19	20	Adelie	Torgersen	46.0	21.5	194.0	4200.0	male	2007
54	55	Adelie	Biscoe	34.5	18.1	187.0	2900.0	female	2008
68	69	Adelie	Torgersen	35.9	16.6	190.0	3050.0	female	2008
69	70	Adelie	Torgersen	41.8	19.4	198.0	4450.0	male	2008
70	71	Adelie	Torgersen	33.5	19.0	190.0	3600.0	female	2008
71	72	Adelie	Torgersen	39.7	18.4	190.0	3900.0	male	2008
72	73	Adelie	Torgersen	39.6	17.2	196.0	3550.0	female	2008
73	74	Adelie	Torgersen	45.8	18.9	197.0	4150.0	male	2008
74	75	Adelie	Torgersen	35.5	17.5	190.0	3700.0	female	2008
75	76	Adelie	Torgersen	42.8	18.5	195.0	4250.0	male	2008
76	77	Adelie	Torgersen	40.9	16.8	191.0	3700.0	female	2008
77	78	Adelie	Torgersen	37.2	19.4	184.0	3900.0	male	2008
78	79	Adelie	Torgersen	36.2	16.1	187.0	3550.0	female	2008
79	80	Adelie	Torgersen	42.1	19.1	195.0	4000.0	male	2008

80	81	Adelie	Torgersen	34.6	17.2	189.0	3200.0	female	2008

```
# filter with .query()
penguins.query('island == "Torgersen"') # "island == 'Torgersen'"
```

92	93	Adelie	Dream	34.0	17.1	185.0	3400.0	female	2008
98	99	Adelie	Dream	33.1	16.1	178.0	2900.0	female	2008
116	117	Adelie	Torgersen	38.6	17.0	188.0	2900.0	female	2009
117	118	Adelie	Torgersen	37.3	20.5	199.0	3775.0	male	2009
118	119	Adelie	Torgersen	35.7	17.0	189.0	3350.0	female	2009
119	120	Adelie	Torgersen	41.1	18.6	189.0	3325.0	male	2009
120	121	Adelie	Torgersen	36.2	17.2	187.0	3150.0	female	2009
121	122	Adelie	Torgersen	37.7	19.8	198.0	3500.0	male	2009
122	123	Adelie	Torgersen	40.2	17.0	176.0	3450.0	female	2009
123	124	Adelie	Torgersen	41.4	18.5	202.0	3875.0	male	2009
124	125	Adelie	Torgersen	35.2	15.9	186.0	3050.0	female	2009
125	126	Adelie	Torgersen	40.6	19.0	199.0	4000.0	male	2009
126	127	Adelie	Torgersen	38.8	17.6	191.0	3275.0	female	2009
127	128	Adelie	Torgersen	41.5	18.3	195.0	4300.0	male	2009
128	129	Adelie	Torgersen	39.0	17.1	191.0	3050.0	female	2009
129	130	Adelie	Torgersen	44.1	18.0	210.0	4000.0	male	2009
130	131	Adelie	Torgersen	38.5	17.9	190.0	3325.0	female	2009
131	132	Adelie	Torgersen	43.1	19.2	197.0	3500.0	male	2009
142	143	Adelie	Dream	32.1	15.5	188.0	3050.0	female	2009

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007
6	7	Adelie	Torgersen	38.9	17.8	181.0	3625.0	female	2007
7	8	Adelie	Torgersen	39.2	19.6	195.0	4675.0	male	2007
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
9	10	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN	2007
10	11	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN	2007
11	12	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN	2007
12	13	Adelie	Torgersen	41.1	17.6	182.0	3200.0	female	2007
13	14	Adelie	Torgersen	38.6	21.2	191.0	3800.0	male	2007
14	15	Adelie	Torgersen	34.6	21.1	198.0	4400.0	male	2007
15	16	Adelie	Torgersen	36.6	17.8	185.0	3700.0	female	2007
16	17	Adelie	Torgersen	38.7	19.0	195.0	3450.0	female	2007
17	18	Adelie	Torgersen	42.5	20.7	197.0	4500.0	male	2007
18	19	Adelie	Torgersen	34.4	18.4	184.0	3325.0	female	2007
19	20	Adelie	Torgersen	46.0	21.5	194.0	4200.0	male	2007
68	69	Adelie	Torgersen	35.9	16.6	190.0	3050.0	female	2008
69	70	Adelie	Torgersen	41.8	19.4	198.0	4450.0	male	2008
70	71	Adelie	Torgersen	33.5	19.0	190.0	3600.0	female	2008
71	72	Adelie	Torgersen	39.7	18.4	190.0	3900.0	male	2008
72	73	Adelie	Torgersen	39.6	17.2	196.0	3550.0	female	2008
73	74	Adelie	Torgersen	45.8	18.9	197.0	4150.0	male	2008
74	75	Adelie	Torgersen	35.5	17.5	190.0	3700.0	female	2008
75	76	Adelie	Torgersen	42.8	18.5	195.0	4250.0	male	2008
76	77	Adelie	Torgersen	40.9	16.8	191.0	3700.0	female	2008
77	78	Adelie	Torgersen	37.2	19.4	184.0	3900.0	male	2008
78	79	Adelie	Torgersen	36.2	16.1	187.0	3550.0	female	2008
79	80	Adelie	Torgersen	42.1	19.1	195.0	4000.0	male	2008
80	81	Adelie	Torgersen	34.6	17.2	189.0	3200.0	female	2008

81	82	Adelie	Torgersen	42.9	17.6	196.0	4700.0	male	2008

```
# filter with .query()
penguins.query('island == "Torgersen" | bill_length_mm < 35')
```

117	118	Adelie	Torgersen	37.3	20.5	199.0	3775.0	male	2009
118	119	Adelie	Torgersen	35.7	17.0	189.0	3350.0	female	2009
119	120	Adelie	Torgersen	41.1	18.6	189.0	3325.0	male	2009
120	121	Adelie	Torgersen	36.2	17.2	187.0	3150.0	female	2009
121	122	Adelie	Torgersen	37.7	19.8	198.0	3500.0	male	2009
122	123	Adelie	Torgersen	40.2	17.0	176.0	3450.0	female	2009
123	124	Adelie	Torgersen	41.4	18.5	202.0	3875.0	male	2009
124	125	Adelie	Torgersen	35.2	15.9	186.0	3050.0	female	2009
125	126	Adelie	Torgersen	40.6	19.0	199.0	4000.0	male	2009
126	127	Adelie	Torgersen	38.8	17.6	191.0	3275.0	female	2009
127	128	Adelie	Torgersen	41.5	18.3	195.0	4300.0	male	2009
128	129	Adelie	Torgersen	39.0	17.1	191.0	3050.0	female	2009
129	130	Adelie	Torgersen	44.1	18.0	210.0	4000.0	male	2009
130	131	Adelie	Torgersen	38.5	17.9	190.0	3325.0	female	2009
131	132	Adelie	Torgersen	43.1	19.2	197.0	3500.0	male	2009

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007
6	7	Adelie	Torgersen	38.9	17.8	181.0	3625.0	female	2007
7	8	Adelie	Torgersen	39.2	19.6	195.0	4675.0	male	2007
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
9	10	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN	2007
10	11	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN	2007
11	12	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN	2007
12	13	Adelie	Torgersen	41.1	17.6	182.0	3200.0	female	2007
13	14	Adelie	Torgersen	38.6	21.2	191.0	3800.0	male	2007
14	15	Adelie	Torgersen	34.6	21.1	198.0	4400.0	male	2007
15	16	Adelie	Torgersen	36.6	17.8	185.0	3700.0	female	2007
16	17	Adelie	Torgersen	38.7	19.0	195.0	3450.0	female	2007
17	18	Adelie	Torgersen	42.5	20.7	197.0	4500.0	male	2007
18	19	Adelie	Torgersen	34.4	18.4	184.0	3325.0	female	2007
19	20	Adelie	Torgersen	46.0	21.5	194.0	4200.0	male	2007
54	55	Adelie	Biscoe	34.5	18.1	187.0	2900.0	female	2008
68	69	Adelie	Torgersen	35.9	16.6	190.0	3050.0	female	2008
69	70	Adelie	Torgersen	41.8	19.4	198.0	4450.0	male	2008
70	71	Adelie	Torgersen	33.5	19.0	190.0	3600.0	female	2008
71	72	Adelie	Torgersen	39.7	18.4	190.0	3900.0	male	2008
72	73	Adelie	Torgersen	39.6	17.2	196.0	3550.0	female	2008
73	74	Adelie	Torgersen	45.8	18.9	197.0	4150.0	male	2008
74	75	Adelie	Torgersen	35.5	17.5	190.0	3700.0	female	2008
75	76	Adelie	Torgersen	42.8	18.5	195.0	4250.0	male	2008
76	77	Adelie	Torgersen	40.9	16.8	191.0	3700.0	female	2008
77	78	Adelie	Torgersen	37.2	19.4	184.0	3900.0	male	2008
78	79	Adelie	Torgersen	36.2	16.1	187.0	3550.0	female	2008
79	80	Adelie	Torgersen	42.1	19.1	195.0	4000.0	male	2008

80	81	Adelie	Torgersen	34.6	17.2	189.0	3200.0	female	2008

```
# check missing in each column
penguins.isna().sum()
```

```
rowid                0
species              0
island               0
bill_length_mm       2
bill_depth_mm        2
flipper_length_mm    2
body_mass_g          2
sex                 11
year                 0
dtype: int64
```

121	122	Adelie	Torgersen	37.7	19.8	198.0	3500.0	male	2009

```
# filter missng value in column
penguins[penguins['sex'].isna()]
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
8	9	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN	2007
9	10	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN	2007
10	11	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN	2007
11	12	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN	2007
47	48	Adelie	Dream	37.5	18.9	179.0	2975.0	NaN	2007
178	179	Gentoo	Biscoe	44.5	14.3	216.0	4100.0	NaN	2007
218	219	Gentoo	Biscoe	46.2	14.4	214.0	4650.0	NaN	2008
256	257	Gentoo	Biscoe	47.3	13.8	216.0	4725.0	NaN	2009
268	269	Gentoo	Biscoe	44.5	15.7	217.0	4875.0	NaN	2009
271	272	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN	2009

```
# drop na
clean_penguins = penguins.dropna()

clean_penguins.head()
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
5	6	Adelie	Torgersen	39.3	20.6	190.0	3650.0	male	2007

```
## fill missing values
top5_penguins = penguins.head(5)
```

```
avg_value = top5_penguins['bill_length_mm'].mean()
print(avg_value)
```

```
38.9
```

```
top5_penguins = top5_penguins['bill_length_mm'].fillna(value = avg_value)
top5_penguins
```

```
0    39.1
1    39.5
2    40.3
3    38.9
4    36.7
Name: bill_length_mm, dtype: float64
```

```
# sort bill_length_mm low to high, high to low
penguins.dropna().sort_values('bill_length_mm', ascending = False).head(10)
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
185	186	Gentoo	Biscoe	59.6	17.0	230.0	6050.0	male	2007
293	294	Chinstrap	Dream	58.0	17.8	181.0	3700.0	female	2007
253	254	Gentoo	Biscoe	55.9	17.0	228.0	5600.0	male	2009
339	340	Chinstrap	Dream	55.8	19.8	207.0	4000.0	male	2009
267	268	Gentoo	Biscoe	55.1	16.0	230.0	5850.0	male	2009
215	216	Gentoo	Biscoe	54.3	15.7	231.0	5650.0	male	2008
307	308	Chinstrap	Dream	54.2	20.8	201.0	4300.0	male	2008
315	316	Chinstrap	Dream	53.5	19.9	205.0	4500.0	male	2008
259	260	Gentoo	Biscoe	53.4	15.8	219.0	5500.0	male	2009
305	306	Chinstrap	Dream	52.8	20.0	205.0	4550.0	male	2008

```
# sort mutiple columns
penguins.dropna().sort_values(['island','bill_length_mm'])
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
54	55	Adelie	Biscoe	34.5	18.1	187.0	2900.0	female	2008
52	53	Adelie	Biscoe	35.0	17.9	190.0	3450.0	female	2008
100	101	Adelie	Biscoe	35.0	17.9	192.0	3725.0	female	2009
25	26	Adelie	Biscoe	35.3	18.9	187.0	3800.0	female	2007
66	67	Adelie	Biscoe	35.5	16.2	195.0	3350.0	female	2008
...	...	...	...	...	...	...	...	...	...
81	82	Adelie	Torgersen	42.9	17.6	196.0	4700.0	male	2008
131	132	Adelie	Torgersen	43.1	19.2	197.0	3500.0	male	2009
129	130	Adelie	Torgersen	44.1	18.0	210.0	4000.0	male	2009
73	74	Adelie	Torgersen	45.8	18.9	197.0	4150.0	male	2008
19	20	Adelie	Torgersen	46.0	21.5	194.0	4200.0	male	2007

333 rows × 9 columns

```
# unique values
penguins['species'].unique()
```

```
array(['Adelie', 'Gentoo', 'Chinstrap'], dtype=object)
```

```
# count values
penguins['species'].value_counts()
```

```
Adelie       152
Gentoo       124
Chinstrap     68
Name: species, dtype: int64
```

```
# count more than one columns
result = penguins[['island','species']].value_counts().reset_index()
result.columns = ["island", 'species','count']
result
```

	island	species	count
0	Biscoe	Gentoo	124
1	Dream	Chinstrap	68
2	Dream	Adelie	56
3	Torgersen	Adelie	52
4	Biscoe	Adelie	44

```
# summarise dataframe
penguins.describe(include = 'all')
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	yea
count	344.000000	344	344	342.000000	342.000000	342.000000	342.000000	333	344
unique	NaN	3	3	NaN	NaN	NaN	NaN	2	NaN
top	NaN	Adelie	Biscoe	NaN	NaN	NaN	NaN	male	NaN
freq	NaN	152	168	NaN	NaN	NaN	NaN	168	NaN
mean	172.500000	NaN	NaN	43.921930	17.151170	200.915205	4201.754386	NaN	200
std	99.448479	NaN	NaN	5.459584	1.974793	14.061714	801.954536	NaN	0.8
min	1.000000	NaN	NaN	32.100000	13.100000	172.000000	2700.000000	NaN	200
25%	86.750000	NaN	NaN	39.225000	15.600000	190.000000	3550.000000	NaN	200
50%	172.500000	NaN	NaN	44.450000	17.300000	197.000000	4050.000000	NaN	200
75%	258.250000	NaN	NaN	48.500000	18.700000	213.000000	4750.000000	NaN	200
max	344.000000	NaN	NaN	59.600000	21.500000	231.000000	6300.000000	NaN	200

```
# average, mean
penguins['bill_length_mm'].std()
```

```
5.4595837139265315
```

```
# group by + sum/mean
penguins[penguins['species'] == 'Adelie']['bill_length_mm'].mean()

penguins.groupby('species')['bill_length_mm'].sum()
```

```
species
Adelie       5857.5
Chinstrap    3320.7
Gentoo       5843.1
Name: bill_length_mm, dtype: float64
```

```
# group by + sum/mean
# penguins[penguins['species'] == 'Adelie']['bill_length_mm'].mean()

penguins.groupby('species')['bill_length_mm'].agg(['min','mean','median','std','m
```

	min	mean	median	std	max
species					
Adelie	32.1	38.791391	38.80	2.663405	46.0
Chinstrap	40.9	48.833824	49.55	3.339256	58.0
Gentoo	40.9	47.504878	47.30	3.081857	59.6

```
# group by more than one columns
result = penguins.groupby(['island','species'])['bill_length_mm'].agg(['min','mea

result.to_csv('result.csv')
```

```
# if your code is long
penguins.groupby(['island','species'])['bill_length_mm']\
    .agg(['min','mean','max'])\
    .reset_index()
```

	island	species	min	mean	max
0	Biscoe	Adelie	34.5	38.975000	45.6
1	Biscoe	Gentoo	40.9	47.504878	59.6
2	Dream	Adelie	32.1	38.501786	44.1
3	Dream	Chinstrap	40.9	48.833824	58.0
4	Torgersen	Adelie	33.5	38.950980	46.0

```
# map values MALE: m, FEMALE: f
penguins['sex'].head()
```

```
0      male
1    female
2    female
3       NaN
4    female
Name: sex, dtype: object
```

```
# map values MALE:m, FEMALE:f
# penguins['sex'].head()

penguins['sex_new'] = penguins['sex'].map( {'male': 'm', 'female': 'f'} ).head(10

penguins.head()
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year	sex
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007	m
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007	f
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007	f
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007	oth
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007	f

```
# pandas style
penguins['bill_length_mm'].mean()
```

```
43.9219298245614
```

```
# numpy
import numpy as np
np.mean(penguins["bill_length_mm"])
```

```
43.9219298245614
```

```
# other numpy fuction
print(np.sum(penguins['bill_depth_mm']))
print(np.std(penguins['body_mass_g']))
```

```
5865.7
800.781229238452
```

```
# where fuction
score = pd.Series([80,55,62,95,20])
```

```
grade = np.where(score >= 80,"passed","failed")
print(grade)
```

```
['passed' 'failed' 'failed' 'passed' 'failed']
```

```
penguins.head()
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year	sex
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007	m
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007	f
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007	f
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007	oth
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007	f

```
df = penguins.query("species == 'Adelie'")[['species','island','bill_length_mm']]
```

```
df['new_column'] = np.where(df['bill_length_mm']>40, True,False) #boolean
```

```
df.head(10)
```

	species	island	bill_length_mm	new_column
0	Adelie	Torgersen	39.1	False
1	Adelie	Torgersen	39.5	False
2	Adelie	Torgersen	40.3	True
4	Adelie	Torgersen	36.7	False
5	Adelie	Torgersen	39.3	False
6	Adelie	Torgersen	38.9	False
7	Adelie	Torgersen	39.2	False
8	Adelie	Torgersen	34.1	False
9	Adelie	Torgersen	42.0	True
10	Adelie	Torgersen	37.8	False

Merge Dataframes

```
left = {
    'key': [1,2,3,4],
    'name': ['yim','praew','nut','toey'],
    'age': [25,26,25,28]
}

right = {
    'key': [1,2,3,4],
    'city': ['Bangkok','nakorn sawan','suphanburi','phitsanulok'],
    'zip': [1001,2504,2094,9802]
}

df_left = pd.DataFrame(left)
df_right = pd.DataFrame(right)
```

```
df_left
```

	key	name	age
0	1	yim	25
1	2	praew	26
2	3	nut	25
3	4	toey	28

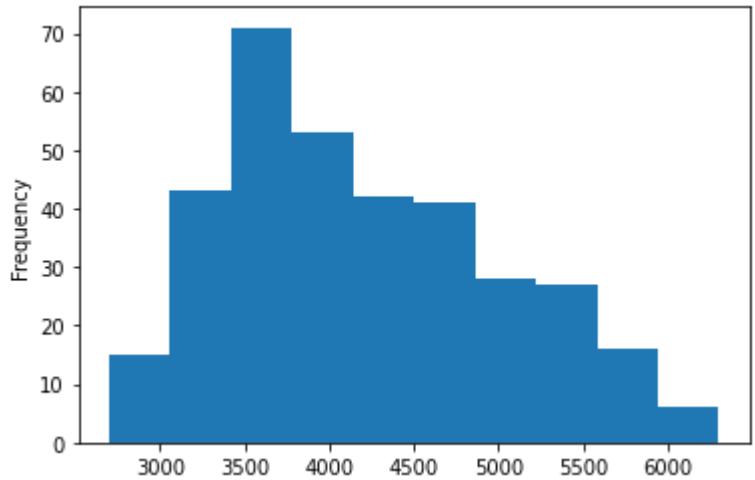
```
import pandas as pd
df_result = pd.merge(df_left,df_right, on='key')
df_result
```

	key	name	age	city	zip
0	1	yim	25	Bangkok	1001
1	2	praew	26	nakorn sawan	2504
2	3	nut	25	suphanburi	2094
3	4	toey	28	phitsanulok	9802

```
# histogram one column
penguins['body_mass_g'].plot(kind='hist')
```

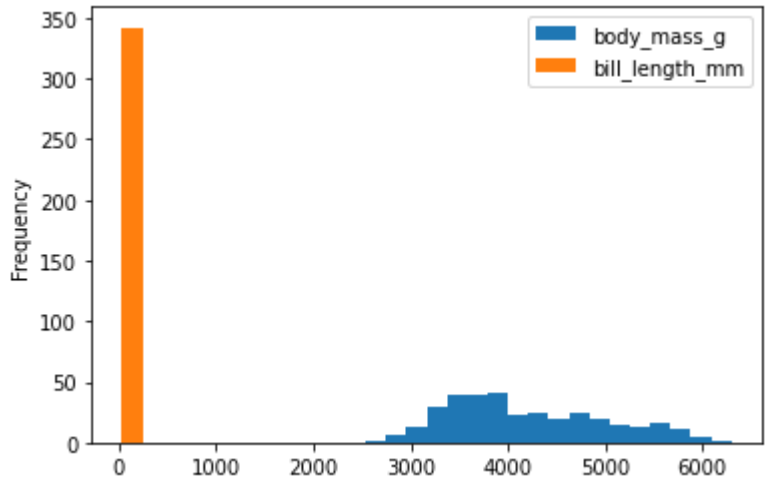
```
<AxesSubplot:ylabel='Frequency'>
```

Download



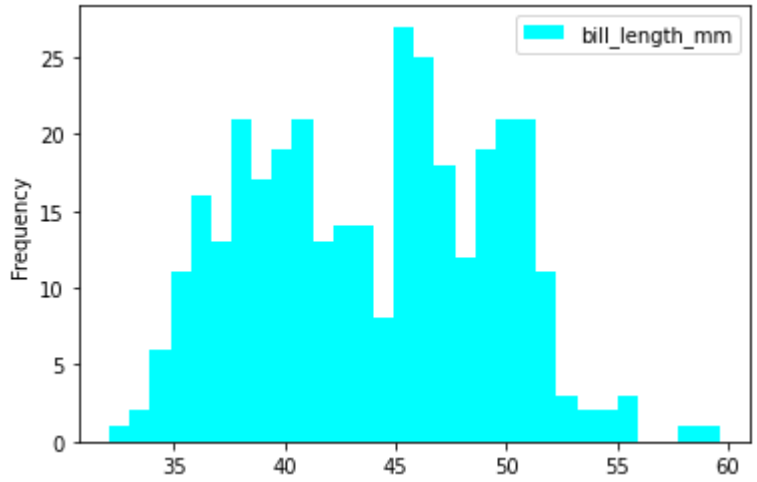
```
# histogram two column
penguins[['body_mass_g','bill_length_mm']].plot(kind='hist',bins=30);
```

Download



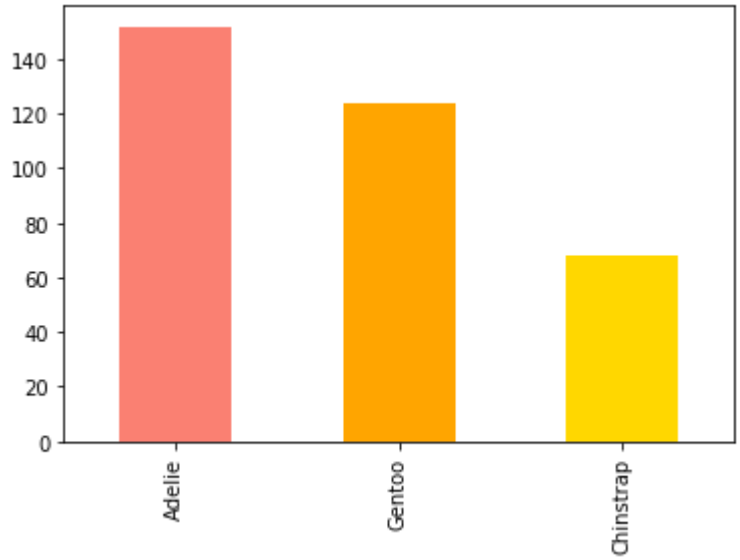

```
penguins[['bill_length_mm']].plot(kind='hist',bins=30, color='cyan');
```

Download



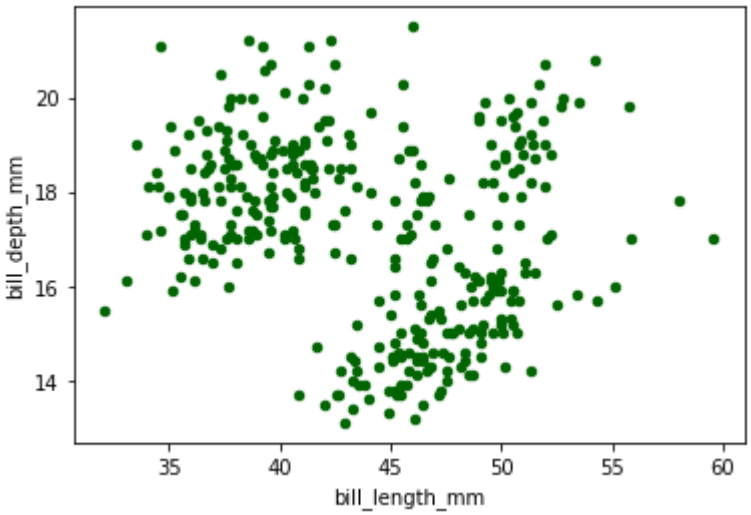

```
# bar plot for species
penguins['species'].value_counts().plot(kind='bar',color=['salmon','orange','gold
```

Download




```
# scatter plot
penguins[['bill_length_mm','bill_depth_mm']]\
    .plot(x='bill_length_mm', y='bill_depth_mm',kind='scatter',color='darkgreen')
```

Download




```
penguins
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
0	1	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male	2007
1	2	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female	2007
2	3	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female	2007
3	4	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	2007
4	5	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female	2007
...	...	...	...	...	...	...	...	...	...
339	340	Chinstrap	Dream	55.8	19.8	207.0	4000.0	male	2009
340	341	Chinstrap	Dream	43.5	18.1	202.0	3400.0	female	2009
341	342	Chinstrap	Dream	49.6	18.2	193.0	3775.0	male	2009
342	343	Chinstrap	Dream	50.8	19.0	210.0	4100.0	male	2009
343	344	Chinstrap	Dream	50.2	18.7	198.0	3775.0	female	2009

344 rows × 10 columns