Intermediate Pandas

Penguins CSV

```python
import pandas as pd
```

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

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	FEMALE
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	MALE
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	FEMALE
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	MALE

```python
# shape of dataframe
penguins.shape
```

```
(344, 7)
```

```python
# information of dataframe
penguins.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 7 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   species            344 non-null    object
 1   island             344 non-null    object
 2   bill_length_mm     342 non-null    float64
 3   bill_depth_mm      342 non-null    float64
 4   flipper_length_mm  342 non-null    float64
 5   body_mass_g        342 non-null    float64
 6   sex                333 non-null    object
dtypes: float64(4), object(3)
memory usage: 18.9+ KB
```

Select columns

```python
# select columns
penguins[ ['species', 'island', 'sex'] ].tail(8)
```

	species	island	sex
336	Gentoo	Biscoe	NaN
337	Gentoo	Biscoe	MALE
338	Gentoo	Biscoe	FEMALE
339	Gentoo	Biscoe	NaN
340	Gentoo	Biscoe	FEMALE
341	Gentoo	Biscoe	MALE
342	Gentoo	Biscoe	FEMALE
343	Gentoo	Biscoe	MALE

Filter Rows

```python
# integer location based indexing (iloc)
mini_penguins = penguins.iloc[ 0:5, 0:3 ]
mini_penguins
```

	species	island	bill_length_mm
0	Adelie	Torgersen	39.1
1	Adelie	Torgersen	39.5
2	Adelie	Torgersen	40.3
3	Adelie	Torgersen	NaN
4	Adelie	Torgersen	36.7

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	MALE
6	Adelie	Torgersen	38.9	17.8	181.0	3625.0	FEMALE
7	Adelie	Torgersen	39.2	19.6	195.0	4675.0	MALE
8	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN
9	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN
10	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN
11	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN
12	Adelie	Torgersen	41.1	17.6	182.0	3200.0	FEMALE
13	Adelie	Torgersen	38.6	21.2	191.0	3800.0	MALE
14	Adelie	Torgersen	34.6	21.1	198.0	4400.0	MALE
15	Adelie	Torgersen	36.6	17.8	185.0	3700.0	FEMALE
16	Adelie	Torgersen	38.7	19.0	195.0	3450.0	FEMALE
17	Adelie	Torgersen	42.5	20.7	197.0	4500.0	MALE
18	Adelie	Torgersen	34.4	18.4	184.0	3325.0	FEMALE
19	Adelie	Torgersen	46.0	21.5	194.0	4200.0	MALE
68	Adelie	Torgersen	35.9	16.6	190.0	3050.0	FEMALE
69	Adelie	Torgersen	41.8	19.4	198.0	4450.0	MALE
70	Adelie	Torgersen	33.5	19.0	190.0	3600.0	FEMALE
71	Adelie	Torgersen	39.7	18.4	190.0	3900.0	MALE
72	Adelie	Torgersen	39.6	17.2	196.0	3550.0	FEMALE
73	Adelie	Torgersen	45.8	18.9	197.0	4150.0	MALE
74	Adelie	Torgersen	35.5	17.5	190.0	3700.0	FEMALE
75	Adelie	Torgersen	42.8	18.5	195.0	4250.0	MALE
76	Adelie	Torgersen	40.9	16.8	191.0	3700.0	FEMALE
77	Adelie	Torgersen	37.2	19.4	184.0	3900.0	MALE
78	Adelie	Torgersen	36.2	16.1	187.0	3550.0	FEMALE
79	Adelie	Torgersen	42.1	19.1	195.0	4000.0	MALE
80	Adelie	Torgersen	34.6	17.2	189.0	3200.0	FEMALE
81	Adelie	Torgersen	42.9	17.6	196.0	4700.0	MALE
82	Adelie	Torgersen	36.7	18.8	187.0	3800.0	FEMALE
83	Adelie	Torgersen	35.1	19.4	193.0	4200.0	MALE
116	Adelie	Torgersen	38.6	17.0	188.0	2900.0	FEMALE

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

```python
penguins[ penguins['bill_length_mm'] > 34 ]
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	MALE
...	...	...	...	...	...	...	...
338	Gentoo	Biscoe	47.2	13.7	214.0	4925.0	FEMALE
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	FEMALE
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	MALE
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	FEMALE
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	MALE

338 rows × 7 columns

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

```python
filtered_penguins
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex

0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	MALE
6	Adelie	Torgersen	38.9	17.8	181.0	3625.0	FEMALE
7	Adelie	Torgersen	39.2	19.6	195.0	4675.0	MALE
8	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN
9	Adelie	Torgersen	42.0	20.2	190.0	4250.0	NaN
10	Adelie	Torgersen	37.8	17.1	186.0	3300.0	NaN
11	Adelie	Torgersen	37.8	17.3	180.0	3700.0	NaN
12	Adelie	Torgersen	41.1	17.6	182.0	3200.0	FEMALE
13	Adelie	Torgersen	38.6	21.2	191.0	3800.0	MALE
14	Adelie	Torgersen	34.6	21.1	198.0	4400.0	MALE
15	Adelie	Torgersen	36.6	17.8	185.0	3700.0	FEMALE
16	Adelie	Torgersen	38.7	19.0	195.0	3450.0	FEMALE
17	Adelie	Torgersen	42.5	20.7	197.0	4500.0	MALE
18	Adelie	Torgersen	34.4	18.4	184.0	3325.0	FEMALE
19	Adelie	Torgersen	46.0	21.5	194.0	4200.0	MALE
54	Adelie	Biscoe	34.5	18.1	187.0	2900.0	FEMALE
68	Adelie	Torgersen	35.9	16.6	190.0	3050.0	FEMALE
69	Adelie	Torgersen	41.8	19.4	198.0	4450.0	MALE
70	Adelie	Torgersen	33.5	19.0	190.0	3600.0	FEMALE
71	Adelie	Torgersen	39.7	18.4	190.0	3900.0	MALE
72	Adelie	Torgersen	39.6	17.2	196.0	3550.0	FEMALE
73	Adelie	Torgersen	45.8	18.9	197.0	4150.0	MALE
74	Adelie	Torgersen	35.5	17.5	190.0	3700.0	FEMALE
75	Adelie	Torgersen	42.8	18.5	195.0	4250.0	MALE
76	Adelie	Torgersen	40.9	16.8	191.0	3700.0	FEMALE
77	Adelie	Torgersen	37.2	19.4	184.0	3900.0	MALE
78	Adelie	Torgersen	36.2	16.1	187.0	3550.0	FEMALE
79	Adelie	Torgersen	42.1	19.1	195.0	4000.0	MALE
80	Adelie	Torgersen	34.6	17.2	189.0	3200.0	FEMALE
81	Adelie	Torgersen	42.9	17.6	196.0	4700.0	MALE
82	Adelie	Torgersen	36.7	18.8	187.0	3800.0	FEMALE
83	Adelie	Torgersen	35.1	19.4	193.0	4200.0	MALE
92	Adelie	Dream	34.0	17.1	185.0	3400.0	FEMALE
98	Adelie	Dream	33.1	16.1	178.0	2900.0	FEMALE
116	Adelie	Torgersen	38.6	17.0	188.0	2900.0	FEMALE
117	Adelie	Torgersen	37.3	20.5	199.0	3775.0	MALE

118	Adelie	Torgersen	35.7	17.0	189.0	3350.0	FEMALE
119	Adelie	Torgersen	41.1	18.6	189.0	3325.0	MALE
120	Adelie	Torgersen	36.2	17.2	187.0	3150.0	FEMALE
121	Adelie	Torgersen	37.7	19.8	198.0	3500.0	MALE
122	Adelie	Torgersen	40.2	17.0	176.0	3450.0	FEMALE
123	Adelie	Torgersen	41.4	18.5	202.0	3875.0	MALE
124	Adelie	Torgersen	35.2	15.9	186.0	3050.0	FEMALE
125	Adelie	Torgersen	40.6	19.0	199.0	4000.0	MALE
126	Adelie	Torgersen	38.8	17.6	191.0	3275.0	FEMALE
127	Adelie	Torgersen	41.5	18.3	195.0	4300.0	MALE
128	Adelie	Torgersen	39.0	17.1	191.0	3050.0	FEMALE
129	Adelie	Torgersen	44.1	18.0	210.0	4000.0	MALE
130	Adelie	Torgersen	38.5	17.9	190.0	3325.0	FEMALE
131	Adelie	Torgersen	43.1	19.2	197.0	3500.0	MALE
142	Adelie	Dream	32.1	15.5	188.0	3050.0	FEMALE

Query

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
8	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN
14	Adelie	Torgersen	34.6	21.1	198.0	4400.0	MALE
18	Adelie	Torgersen	34.4	18.4	184.0	3325.0	FEMALE
70	Adelie	Torgersen	33.5	19.0	190.0	3600.0	FEMALE
80	Adelie	Torgersen	34.6	17.2	189.0	3200.0	FEMALE

Missing Values NAN

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

```
# filter missing values in column sex
penguins[penguins['bill_length_mm'].isna()]
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN

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

clean_penguins.head(10)
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	MALE
6	Adelie	Torgersen	38.9	17.8	181.0	3625.0	FEMALE
7	Adelie	Torgersen	39.2	19.6	195.0	4675.0	MALE
12	Adelie	Torgersen	41.1	17.6	182.0	3200.0	FEMALE
13	Adelie	Torgersen	38.6	21.2	191.0	3800.0	MALE
14	Adelie	Torgersen	34.6	21.1	198.0	4400.0	MALE

Fill Missing Values

```
# 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
```

Sort DataFrame and Value Count

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
253	Gentoo	Biscoe	59.6	17.0	230.0	6050.0	MALE
169	Chinstrap	Dream	58.0	17.8	181.0	3700.0	FEMALE
321	Gentoo	Biscoe	55.9	17.0	228.0	5600.0	MALE
215	Chinstrap	Dream	55.8	19.8	207.0	4000.0	MALE
335	Gentoo	Biscoe	55.1	16.0	230.0	5850.0	MALE
283	Gentoo	Biscoe	54.3	15.7	231.0	5650.0	MALE
183	Chinstrap	Dream	54.2	20.8	201.0	4300.0	MALE
191	Chinstrap	Dream	53.5	19.9	205.0	4500.0	MALE
327	Gentoo	Biscoe	53.4	15.8	219.0	5500.0	MALE
181	Chinstrap	Dream	52.8	20.0	205.0	4550.0	MALE

```python
# sort multiple columns
penguins.dropna().sort_values(['island', 'bill_length_mm'], ascending=[True, False])
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
253	Gentoo	Biscoe	59.6	17.0	230.0	6050.0	MALE
321	Gentoo	Biscoe	55.9	17.0	228.0	5600.0	MALE
335	Gentoo	Biscoe	55.1	16.0	230.0	5850.0	MALE
283	Gentoo	Biscoe	54.3	15.7	231.0	5650.0	MALE
327	Gentoo	Biscoe	53.4	15.8	219.0	5500.0	MALE
...	...	...	...	...	...	...	...
83	Adelie	Torgersen	35.1	19.4	193.0	4200.0	MALE
14	Adelie	Torgersen	34.6	21.1	198.0	4400.0	MALE
80	Adelie	Torgersen	34.6	17.2	189.0	3200.0	FEMALE
18	Adelie	Torgersen	34.4	18.4	184.0	3325.0	FEMALE
70	Adelie	Torgersen	33.5	19.0	190.0	3600.0	FEMALE

333 rows × 7 columns

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

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

```
# 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

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

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

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

44.45

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

```python
# group by + agg
penguins.groupby('species')['bill_length_mm'].agg(['min', 'mean', 'median', 'std', 'max'])
```

	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

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

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

```python
# 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

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

penguins['sex_new'] = penguins['sex'].map( {'MALE': 'm', 'FEMALE': 'f'} ).fillna('other')

penguins.head()
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE	m
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE	f
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE	f
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	other
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE	f

What is numpy?

Numpy stands for `Numerical Python`, a very popular library for data science and computation.

```python
penguins.head()
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE	m
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE	f
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE	f
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	other
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE	f

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

```
43.92192982456142
```

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

```
43.92192982456142
```

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

```
5865.700000000001
800.7812292384519
```

Numpy Where

```python
import numpy as np
```

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

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

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

```
penguins.head()
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE	m
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE	f
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE	f
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	other
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE	f

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

```python
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

```python
left = {
    'key': [1, 2, 3, 4],
    'name': ['toy', 'joe', 'jane', 'anna'],
    'age': [25, 28, 30, 22]
}

right = {
    'key': [1, 2, 3, 4],
    'city': ['Bangkok', 'London', 'Seoul', 'Tokyo'],
    'zip': [1001, 2504, 2094, 9802]
}

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

df_left

	key	name	age
0	1	toy	25
1	2	joe	28
2	3	jane	30
3	4	anna	22

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

df_result

	key	name	age	city	zip
0	1	toy	25	Bangkok	1001
1	2	joe	28	London	2504
2	3	jane	30	Seoul	2094
3	4	anna	22	Tokyo	9802

Pandas Plots

- histogram
- histogram 2 columns
- bar plot
- scatter plot

```python
# histogram one column
penguins['body_mass_g'].plot(kind='hist');
```
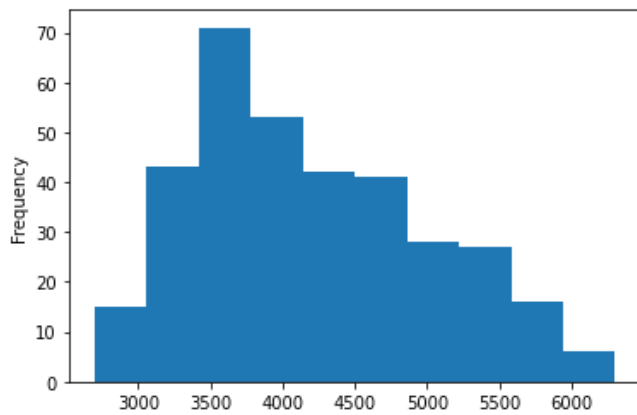
```
# histogram two columns
penguins[['body_mass_g', 'bill_length_mm']].plot(kind='hist', bins=30)
```
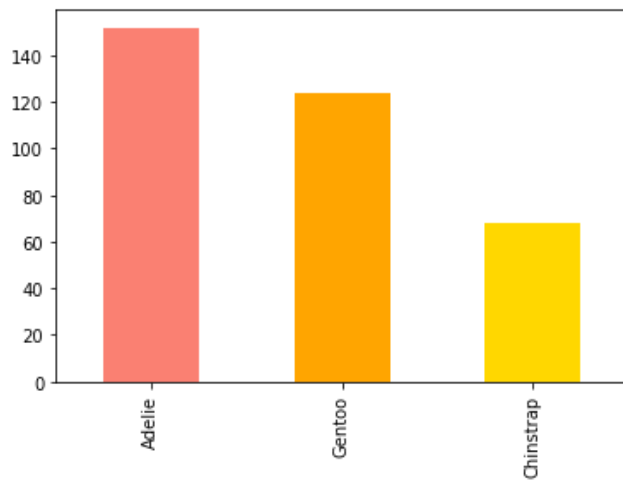
```
<AxesSubplot:ylabel='Frequency'>
```
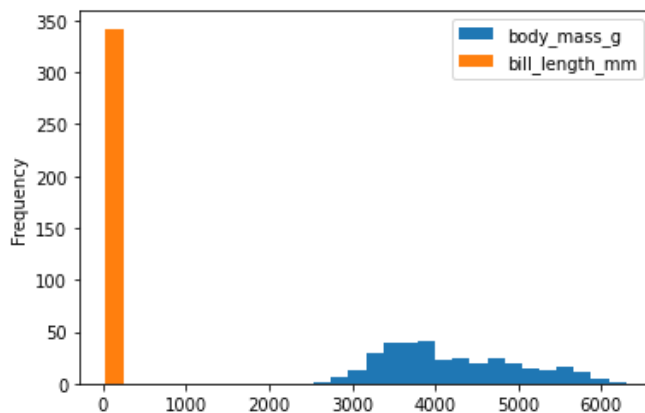


```
penguins['bill_length_mm'].plot(kind='hist', bins=30, color="orange");
```
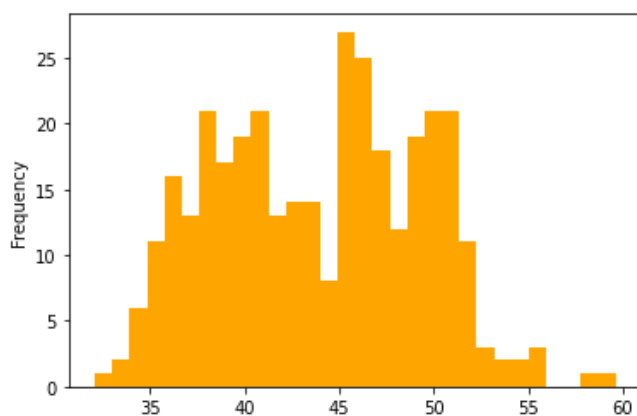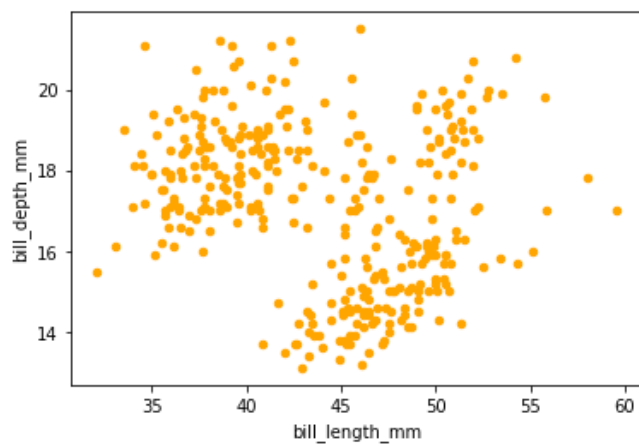


```
# bar plot for species
penguins['species'].value_counts().plot(kind='bar', color=['salmon','orange', 'gold'])
```

```
<AxesSubplot:>
```

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

```
<AxesSubplot:xlabel='bill_length_mm', ylabel='bill_depth_mm'>
```



```
penguins
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE	m
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE	f
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE	f
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN	other
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE	f
...	...	...	...	...	...	...	...	...
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN	other
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	FEMALE	f
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	MALE	m
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	FEMALE	f
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	MALE	m

344 rows × 8 columns