Intermediate Pandas II

```
import pandas as pd
```

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

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

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex

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

```
44.45
```

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

print(penguins.groupby("species")["bill_length_mm"].mean())

print(penguins.groupby("species")["bill_length_mm"].sum())

print(penguins.groupby("species")["bill_length_mm"].median())
```

```
species
Adelie       38.791391
Chinstrap    48.833824
Gentoo       47.504878
Name: bill_length_mm, dtype: float64

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

species
Adelie       38.80
Chinstrap    49.55
```

```
Gentoo       47.30
Name: bill_length_mm, dtype: float64
```

```
#aggregate function

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

	min	mean	median	std	max

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

print(result)

result.to_csv("result.csv")
```

```
      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
```

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

	island	species	min	mean	max

```
# map values MALE: m, FEMALE: f

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

penguins.head()
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new

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

```
43.9219298245614
```

numpy

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

```
43.9219298245614
```

```
#other function of numpy
print(np.sum(penguins["bill_depth_mm"]))
print(np.std(penguins["body_mass_g"]))
```

```
5865.700000000001
800.7812292384522
```

```
# WHERE
score = pd.Series([80, 55, 62, 95, 70])
```

```
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

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

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

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

	species	island	bill_length_mm	new_column

```
# MERGE Dataframes

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

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

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

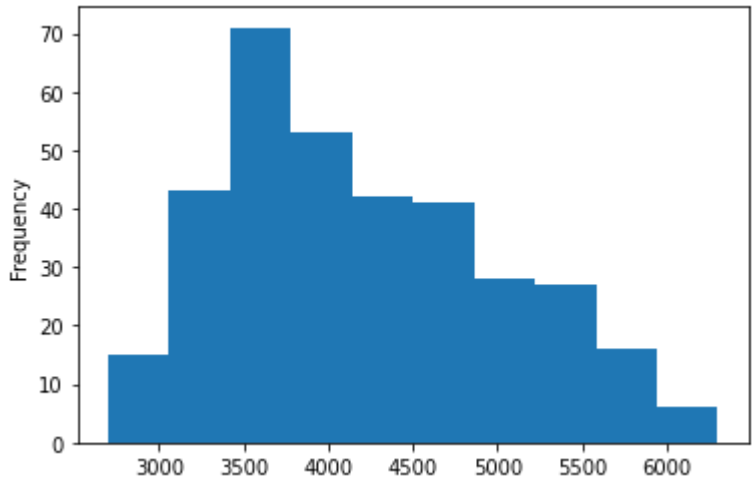
```
import pandas as pd
df_result = pd.merge(df_left, df_right, on = "key")

df_result
```

	key	name	age	city	zip

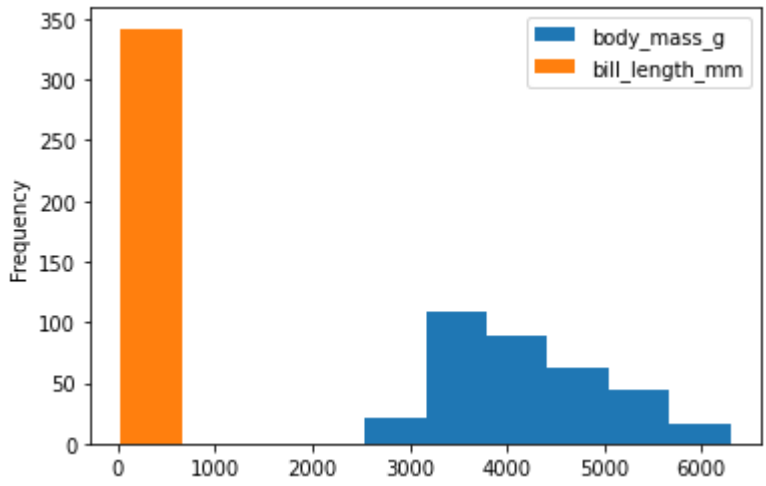
```
#histogram one column
penguins["body_mass_g"].plot(kind = "hist");
```

Download



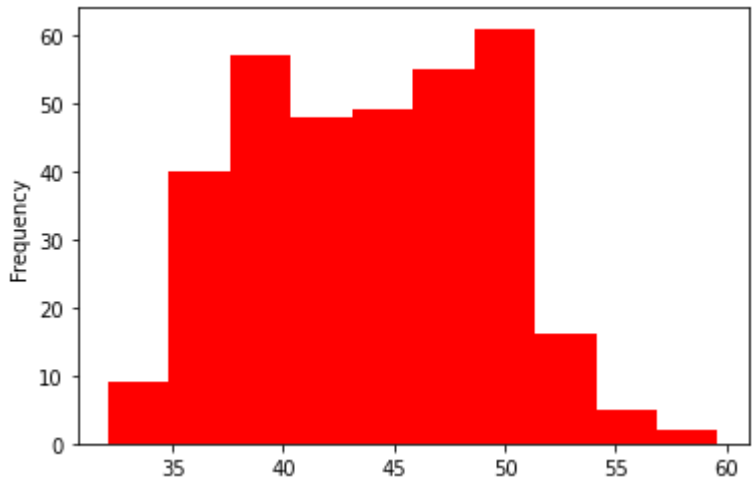

```
# histogram two column
penguins[["body_mass_g", "bill_length_mm"]].plot(kind= "hist");
```

Download



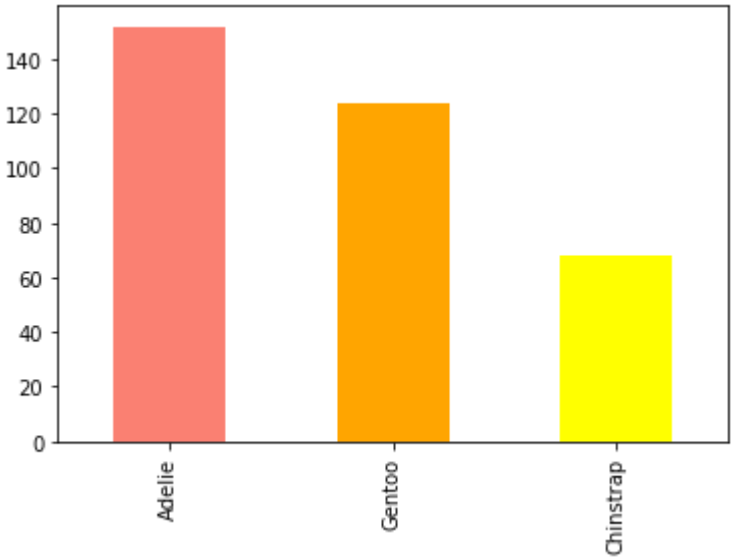

```
penguins["bill_length_mm"].plot(kind= "hist", color="red");
```

Download



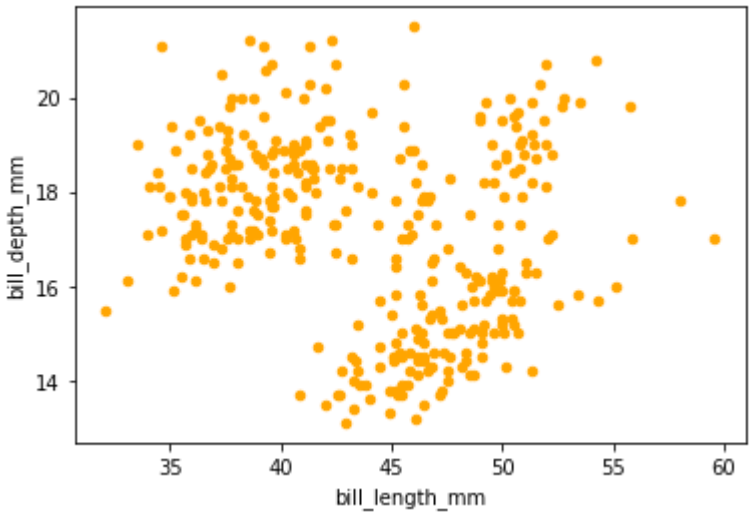

```
#barplot for species
penguins["species"]. value_counts().plot(kind="bar", color = ["salmon", "orange", "
```

Download




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

Download




```
penguins
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	sex_new

```
penguins
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex