```
import pandas as pd
```

```
df = pd.read_csv('/content/penguins.csv')
df.head()
```

	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	39.1	18.7	181.0	3750.0	MALE
1	39.5	17.4	186.0	3800.0	FEMALE
2	40.3	18.0	195.0	3250.0	FEMALE
3	NaN	NaN	NaN	NaN	NaN
4	36.7	19.3	193.0	3450.0	FEMALE

```
df.isna().sum()
```

```
culmen_length_mm     2
culmen_depth_mm      2
flipper_length_mm    2
body_mass_g          2
sex                  9
dtype: int64
```

```
df = df.dropna()
df.isna().sum()
```

```
culmen_length_mm     0
culmen_depth_mm      0
flipper_length_mm    0
body_mass_g          0
sex                  0
dtype: int64
```

```
df= df.replace({'MALE': 1, 'FEMALE': 0})
```

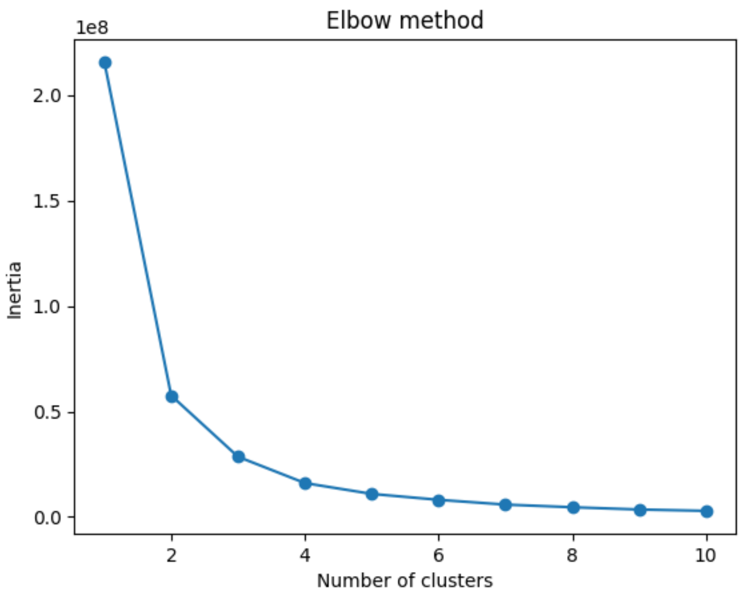
```
df=df.replace({'.': 1})
df.head()
```

	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	39.1	18.7	181.0	3750.0	1
1	39.5	17.4	186.0	3800.0	0
2	40.3	18.0	195.0	3250.0	0
4	36.7	19.3	193.0	3450.0	0
5	39.3	20.6	190.0	3650.0	1

```
x = df['culmen_length_mm']
y = df['body_mass_g']
```

```
#to find the how many number of clusters are required for our data we use elbow method


from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

data = list(zip(x, y))
#print((data))
inertias = []
#print(inertias)

for i in range(1,11):
    kmeans = KMeans(n_clusters=i)
    kmeans.fit(data)
    inertias.append(kmeans.inertia_)
    print(inertias)

plt.plot(range(1,11), inertias, marker='o')
plt.title('Elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
[215716082.37832844]
[215716082.37832844, 57560981.11003442]
[215716082.37832844, 57560981.11003442, 28539957.375238292]
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296]
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522]
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522, 8118592.124499304]
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522, 8118592.124499304, 5856565.79304
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522, 8118592.124499304, 5856565.79304
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522, 8118592.124499304, 5856565.79304
[215716082.37832844, 57560981.11003442, 28539957.375238292, 16063634.734553296, 10921701.34352522, 8118592.124499304, 5856565.79304
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
```



```
# since in the above graph there is a break at 2, so 2 cluster are best for our dataset

kmeans = KMeans(n_clusters=2)
kmeans.fit(data)

plt.scatter(x, y, c=kmeans.labels_)
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
  warnings.warn(
```