Assignment -3

Python Programming

Assignment Date	5 October 2022
Student Name	R. Vigneshwari
Student Roll Number	820419106070
Maximum Marks	2 Marks

1. Download the dataset

2. Load the dataset into the tool

In [193]:
```
import pandas as pd
data = pd.read_csv("abalone.csv")
data.head()
```

Out[193]:

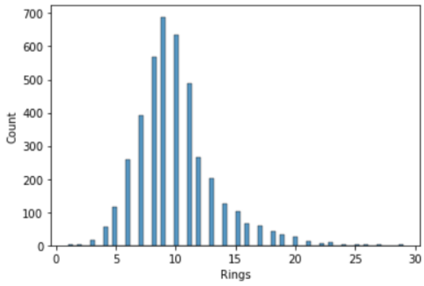
	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

3. Perform Below Visualizations.

· Univariate Analysis

In [194]:
```
import seaborn as sns
sns.histplot(data, x="Rings")
```

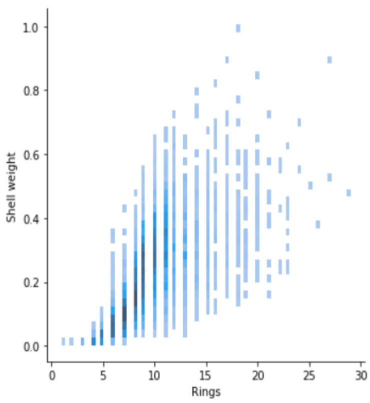
Out[194]: <AxesSubplot:xlabel='Rings', ylabel='Count'>



.Bi-Variate Analysis

In [196]:
```
sns.displot(data, x="Rings", y="Shell weight")
```

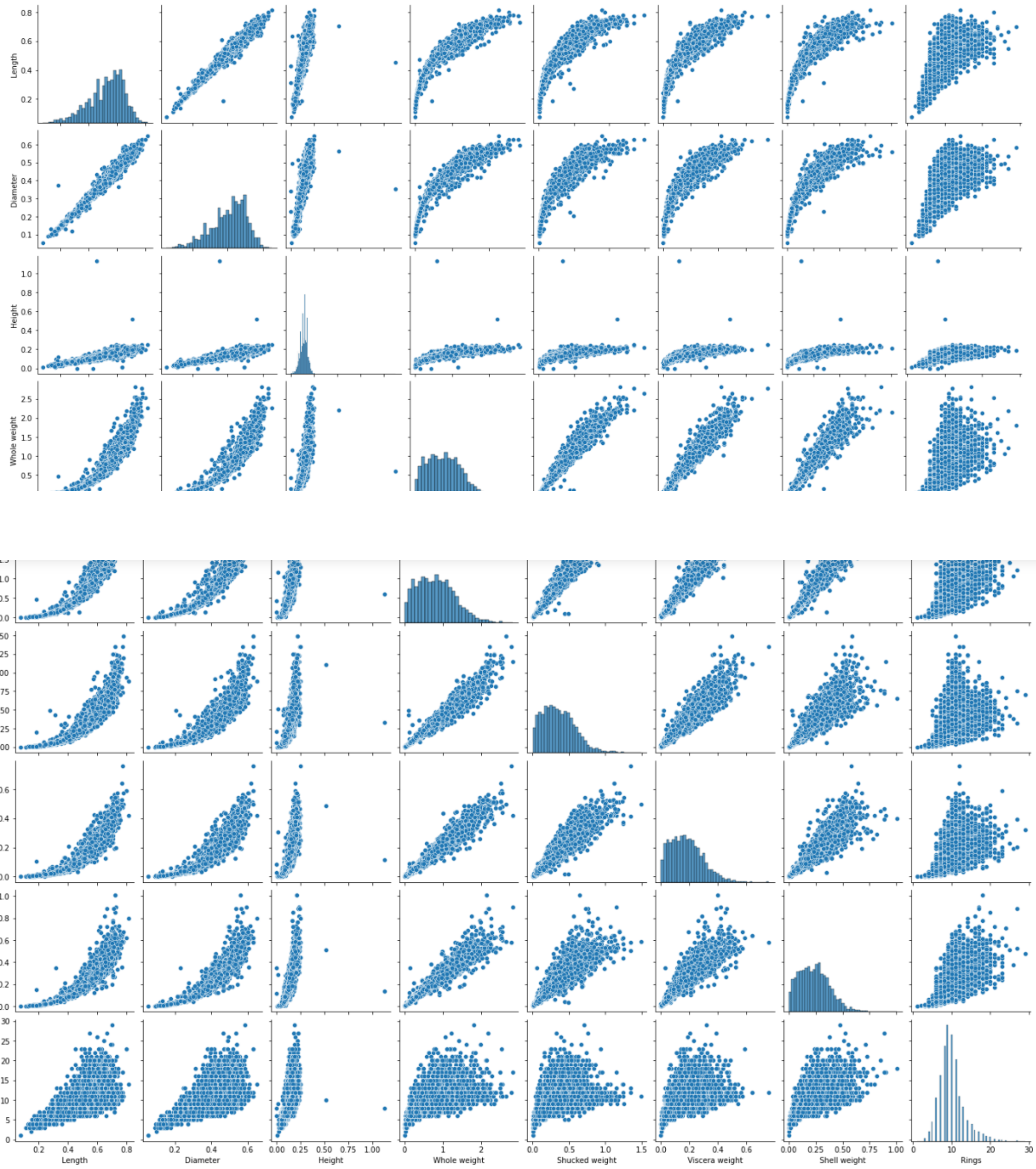
Out[196]: <seaborn.axisgrid.FacetGrid at 0x1dd120f6160>



. Multi-Variate Analysis

In [198]:
```
sns.pairplot(data)
```

Out[198]:
```
<seaborn.axisgrid.PairGrid at 0x1dd13611850>
```



4. Perform descriptive statistics on the dataset.

In [199]:
```
data.mean()
```

```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/531903386.py:1: FutureWarning: Dropping of nuisance columns in DataFrame red
uctions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns bef
ore calling the reduction.
  data.mean()
```

Out[199]:
```
Length            0.523992
Diameter          0.407881
Height            0.139516
Whole weight      0.828742
Shucked weight    0.359367
Viscera weight    0.180594
Shell weight      0.238831
Rings             9.933684
dtype: float64
```

In [200]:
```
data.median()
```

```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/4184645713.py:1: FutureWarning: Dropping of nuisance columns in DataFrame re
ductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns be
fore calling the reduction.
  data.median()
```

Out[200]:
```
Length            0.5450
Diameter          0.4250
Height            0.1400
Whole weight      0.7995
Shucked weight    0.3360
Viscera weight    0.1710
Shell weight      0.2340
Rings             9.0000
dtype: float64
```

In [201]:
```
data.mode()
```
Out[201]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	M	0.550	0.45	0.15	0.2225	0.175	0.1715	0.275	9.0
1	NaN	0.625	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [202]:
```
data.skew()
```
```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/1188251951.py:1: FutureWarning: Dropping of nuisance columns in DataFrame re
ductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns be
fore calling the reduction.
  data.skew()
```
Out[202]:
```
Length           -0.639873
Diameter         -0.609198
Height            3.128817
Whole weight      0.530959
Shucked weight    0.719098
Viscera weight    0.591852
Shell weight      0.620927
Rings             1.114102
dtype: float64
```

In [203]:
```
data.kurt()
```
```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/2907027414.py:1: FutureWarning: Dropping of nuisance columns in DataFrame re
ductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns be
fore calling the reduction.
  data.kurt()
```
Out[203]:
```
Length             0.064621
Diameter          -0.045476
Height            76.025509
Whole weight      -0.023644
Shucked weight     0.595124
Viscera weight     0.084012
Shell weight       0.531926
Rings              2.330687
dtype: float64
```

In [204]:
```
data.std()
```
```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/2723740006.py:1: FutureWarning: Dropping of nuisance columns in DataFrame re
ductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns be
fore calling the reduction.
  data.std()
```
Out[204]:
```
Length            0.120093
Diameter          0.099240
Height            0.041827
Whole weight      0.490389
Shucked weight    0.221963
Viscera weight    0.109614
Shell weight      0.139203
Rings             3.224169
dtype: float64
```

In [205]:
```
data.var()
```
```
C:\Users\welcome\AppData\Local\Temp/ipykernel_2888/445316826.py:1: FutureWarning: Dropping of nuisance columns in DataFrame red
uctions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns bef
ore calling the reduction.
  data.var()
```
Out[205]:
```
Length             0.014422
Diameter           0.009849
Height             0.001750
Whole weight       0.240481
Shucked weight     0.049268
Viscera weight     0.012015
Shell weight       0.019377
Rings             10.395266
dtype: float64
```

5. Check for Missing values and deal with them.

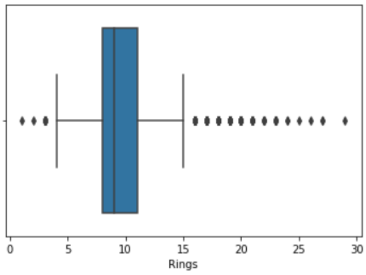
In [206]:
```
data.isna().sum()
```
Out[206]:
```
Sex               0
Length            0
Diameter          0
Height            0
Whole weight      0
Shucked weight    0
Viscera weight    0
Shell weight      0
Rings             0
dtype: int64
```

6. Find the outliers and replace them outliers

In [207]:
```
sns.boxplot(data['Rings'])
```

```
C:\Users\welcome\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[207]: <AxesSubplot:xlabel='Rings'>



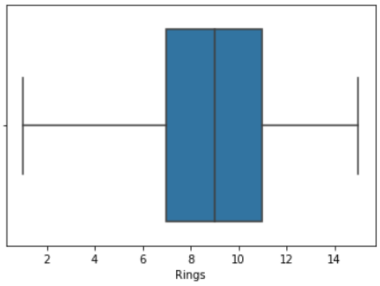
In [208]:
```
import numpy as np

data['Rings']=np.where(data['Rings']>15,5,data['Rings']) #replacing
```

In [209]:
```
sns.boxplot(data['Rings'])
```

```
C:\Users\welcome\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[209]: <AxesSubplot:xlabel='Rings'>



7. Check for Categorical columns and perform encoding.

In [210]:
```
data.tail()
```

Out[210]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
4172	F	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	11
4173	M	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	10
4174	M	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	9
4175	F	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
4176	M	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950	12

In [211]:
```
from sklearn.preprocessing import LabelEncoder
```

In [212]:
```
le=LabelEncoder()
```

In [213]:
```
data['Sex']=le.fit_transform(data['Sex'])
```

In [214]:
```
data.head()
```

Out[214]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

8. Split the data into dependent and independent variables ¶

In [215]:
```
y=data['Rings']
y.head()
```

Out[215]:
```
0    15
1     7
2     9
3    10
4     7
Name: Rings, dtype: int64
```

In [216]:
```
x=data.drop(columns=['Rings'])
x.head()
```

Out[216]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055

9. Scale the independent variables

In [217]:
```
from sklearn.preprocessing import scale
x=scale(x)
x
```

Out[217]:
```
array([[ 1.15198011, -0.57455813, -0.43214879, ..., -0.60768536,
        -0.72621157, -0.63821689],
       [ 1.15198011, -1.44898585, -1.439929  , ..., -1.17090984,
        -1.20522124, -1.21298732],
       [-1.28068972,  0.05003309,  0.12213032, ..., -0.4634999 ,
        -0.35668983, -0.20713907],
       ...,
       [ 1.15198011,  0.6329849 ,  0.67640943, ...,  0.74855917,
         0.97541324,  0.49695471],
       [-1.28068972,  0.84118198,  0.77718745, ...,  0.77334105,
         0.73362741,  0.41073914],
       [ 1.15198011,  1.54905203,  1.48263359, ...,  2.64099341,
         1.78744868,  1.84048058]])
```

10. Split the data into training and testing

In [218]:
```
from sklearn.model_selection import train_test_split
```

In [219]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
```

In [220]:
```
print(x_train.shape,y_train.shape)
```

```
(3341, 8) (3341,)
```

In [221]:
```
print(x_test.shape,y_test.shape)
```

```
(836, 8) (836,)
```

11. Build the Model

In [231]:
```
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
```

12. Train the Model

In [223]:
```
model.fit(x_train,y_train)
```

Out[223]:
```
DecisionTreeRegressor()
```

13. Test the Model

In [224]:
```
pred2=model.predict(x_test)

pred2
```

Out[224]:
```
array([13.,  5.,  5., 10., 14.,  5.,  8.,  8.,  4., 13., 13.,  9., 13.,
        7.,  7., 12., 10., 10., 11.,  9.,  5.,  9.,  7., 10.,  9.,  6.,
        7.,  6.,  9.,  5.,  8., 14.,  6., 12.,  8., 11.,  6.,  3.,  7.,
        6., 10.,  5., 12.,  9., 13., 11.,  8.,  8., 14.,  5.,  6.,  5.,
        9.,  6.,  5.,  9.,  9.,  9., 12., 10., 10.,  5., 12.,  7.,  8.,
       11.,  4.,  5.,  8.,  9., 14.,  9.,  8.,  9.,  9.,  9.,  5., 11.,
        9.,  7., 14., 14., 14.,  8.,  6., 13.,  9.,  9., 11., 11.,  5.,
       13., 10., 13., 11.,  9.,  8.,  9.,  7.,  6.,  5., 10., 11., 11.,
        7.,  9.,  8.,  8., 10.,  8., 10.,  9., 10., 10., 10.,  5.,  7.,
        7.,  9., 15., 11., 10.,  8., 11., 11., 13., 11., 12., 11.,  5.,
       10.,  9.,  9.,  4.,  5.,  9.,  9., 10.,  6., 11.,  8., 11.,  8.,
       11., 10., 13.,  5.,  5.,  8., 11.,  9.,  9., 11., 15.,  7., 11.,
        9.,  5., 10.,  6., 10.,  6.,  8., 13.,  7., 12., 11.,  9., 12.,
        9., 10., 12., 10.,  8.,  5., 10.,  8.,  9., 12.,  7., 14., 11.,
        6.,  8.,  7., 13.,  8., 11.,  9.,  5.,  9., 12., 11., 10., 11.,
        9., 11.,  5.,  9., 12.,  8., 13., 13., 11.,  7.,  9., 10.,  9.,
        7., 10.,  5., 11., 11.,  8.,  5., 13.,  5.,  6., 10.,  5., 11.,
       10., 10., 12., 10.,  7.,  8.,  9., 12., 11.,  4.,  5., 11., 14.,
       10.,  9.,  8.,  9.,  5., 15.,  9., 15.,  9.,  7., 14., 10.,  5.,
       10.,  8.,  6.,  8., 11., 11.,  8., 11., 11., 10.,  5., 11., 10.,
        5.,  7., 12.,  9.,  8.,  7.,  5., 13., 10.,  7.,  8.,  8.,  8.,
       11.,  8., 13., 11., 15.,  9.,  4.,  9.,  8., 10.,  9.,  5.,  5.,
       10.,  6., 11., 13.,  8.,  3., 13.,  5.,  5., 11.,  5.,  5.,  6.,
        9.,  9., 10.,  7., 13.,  9.,  9., 11.,  9.,  5.,  5., 10.,  6.,
        7.,  9., 10.,  9.,  4., 15.,  8.,  7., 13.,  7.,  7., 12., 11.,
        8., 10.,  5.,  8.,  9.,  7., 12.,  8., 10., 11.,  8., 10., 12.,
       10.,  9.,  9.,  9., 15., 11.,  8.,  7.,  9., 10., 11., 12., 11.,
       15.,  9.,  5., 11.,  4.,  6.,  6.,  4.,  8., 10.,  6., 11., 11.,
```

```
        10.,  7., 15., 15.,  7.,  8., 11., 12., 10.,  9., 10.,  7.,  9.,
         7.,  6.,  8., 11.,  8., 14., 11., 13., 14.,  6., 10.,  5.,  6.,
         5.,  9., 11., 11.,  9., 13., 12., 12.,  8., 13.,  5.,  9., 15.,
         8.,  9., 13., 10.,  5., 13.,  6., 12.,  6.,  9.,  8.,  6., 12.,
        11.,  9., 11.,  9.,  7.,  8., 14.,  9.,  7.,  9.,  6.,  8., 10.,
        12.,  8., 11.,  5.,  6.,  8., 14., 11., 10.,  8.,  6.,  9.,  9.,
         8., 10.,  9., 10.,  7., 13., 12.,  7.,  5.,  5.,  8.,  9.,  8.,
        12.,  6., 15.,  9., 10.,  9.,  8.,  9.,  6., 11.,  5.,  7.,  8.,
        13., 12., 13., 11.,  9.,  9., 10.,  5.,  5.,  8.,  9.,  7.,  8.,
         9., 10., 11.,  7., 13., 10., 12.,  7.,  8.,  3., 10.,  9., 11.,
         7.,  7.,  9.,  7.,  7.,  9.,  9., 14.,  5., 12.,  8., 10.,  4.,
        15.,  5.,  8., 15.,  9.,  9., 10.,  8.,  4.,  9., 12.,  5.,  7.,
         8., 15., 10.,  8.,  8., 10., 14., 11.,  6.,  6.,  9.,  5., 13.,
         9., 11., 12.,  9.,  7.,  7., 13., 11., 12.,  8., 11.,  8.,  7.,
         8.,  8.,  9., 10.,  7.,  8.,  8.,  9.,  8.,  6.,  7.,  7.,  5.,
         5., 11.,  9., 12.,  9., 15.,  8.,  9.,  8., 11.,  8.,  7.,  8.,
         9., 10.,  6., 14., 11., 12., 14., 10., 14.,  6., 11.,  8.,  7.,
         6.,  3.,  9.,  8., 11., 11., 12.,  7., 10.,  9.,  8.,  8.,  6.,
        10., 11.,  5.,  5., 11., 14.,  8.,  8., 10.,  6., 11., 11.,  9.,
        15.,  8., 11.,  8.,  5.,  6., 11.,  8.,  8., 12.,  8.,  8.,  5.,
        12.,  6.,  5., 12., 10., 10.,  5.,  7., 12., 14.,  8.,  8.,  5.,
         5.,  8.,  8., 12., 10.,  4., 11.,  6.,  9.,  9.,  9., 12.,  9.,
         6.,  9., 12., 10.,  9.,  9.,  5., 10., 13.,  7., 11., 15.,  9.,
         7.,  9.,  9., 12., 13., 12.,  8.,  8., 13., 11.,  9.,  5., 13.,
        10.,  9.,  7.,  8.,  6.,  7.,  8.,  6., 10.,  5., 10.,  4., 13.,
         8.,  5.,  8.,  9.,  9., 11., 11.,  8., 11., 11.,  8.,  8.,  8.,
         6.,  6., 13.,  8.,  9.,  8., 11.,  5.,  8.,  9.,  9.,  8., 10.,
        13.,  8., 12., 14., 15.,  7.,  8.,  8., 10.,  7., 11.,  5., 12.,
        12., 10., 11.,  8.,  9., 10., 11., 15., 12., 12., 11., 12.,  9.,
        11.,  9.,  8.,  9.,  9., 11.,  8., 14., 11.,  3., 10.,  7., 12.,
        14., 12., 12.,  9., 15., 11.,  8.,  7.,  6.,  9.,  4.,  5.,  5.,
        14., 10.,  5., 13., 11., 11.,  3.,  5.,  5., 13., 10., 10.,  5.,
         9.,  5., 13.,  5., 10.,  8., 11.,  9., 10., 10., 10., 11.,  8.,
         8.,  8.,  8., 11.,  9.,  6.,  8.,  5.,  7.,  8.,  8., 11.,  9.,
        12.,  9.,  8.,  7.])
```

14. Measure the performance using Metrics.

In [225]:
```
from sklearn import metrics
```

In [226]:
```
metrics.confusion_matrix(y_test,pred2)
```

Out[226]:
```
array([[ 1,  2,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0],
       [ 3,  4,  5,  1,  0,  0,  0,  0,  0,  0,  0,  0,  0],
       [ 2,  5, 26,  7,  2,  5,  5,  3, 11,  7,  5,  3,  5],
       [ 0,  1,  5, 16,  9, 11,  3,  0,  2,  1,  0,  0,  0],
       [ 0,  1,  5, 17, 18, 23, 10,  4,  2,  1,  2,  0,  1],
       [ 0,  0,  4,  6, 16, 28, 17, 13,  6,  7,  1,  0,  1],
       [ 0,  0,  6,  2,  9, 28, 30, 22, 23, 12,  7,  2,  1],
       [ 0,  0,  8,  1, 10, 18, 33, 18, 18, 11, 10,  6,  6],
       [ 0,  0,  7,  1,  4,  8, 21, 17, 14,  8,  5,  5,  3],
       [ 0,  0,  7,  0,  2,  4,  5,  8, 13,  5,  5,  1,  1],
       [ 0,  0,  3,  2,  0,  1,  3,  7,  6,  1,  3,  2,  3],
       [ 0,  0,  4,  1,  0,  3,  3,  2,  5,  1,  2,  5,  0],
       [ 0,  0,  3,  0,  0,  0,  4,  1,  5,  3,  4,  1,  0]], dtype=int64)
```

In [227]:
```
print('DT model ACcuracy Score:',metrics.accuracy_score(y_test,pred2))
```

```
DT model ACcuracy Score: 0.20095693779904306
```

In [228]:
```
acc=metrics.accuracy_score(y_test,pred2)
acc
```

Out[228]:
```
0.20095693779904306
```

In [229]:
```
1-acc
```

Out[229]:
```
0.799043062200957
```

In [230]:
```
data.head()
```

Out[230]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7