ASSIGNMENT-3

In [4]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error,explained_variance_score,
```

In [5]:

```
data = pd.read_csv("C:/Users/MANOHARI/Downloads/abalone.csv")
```

In [6]:

```
data.head()
```

Out[6]:

	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

In [7]:

```
data.describe()
```

Out[7]:

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shel
count	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177
mean	0.523992	0.407881	0.139516	0.828742	0.359367	0.180594	0
std	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0
min	0.075000	0.055000	0.000000	0.002000	0.001000	0.000500	0
25%	0.450000	0.350000	0.115000	0.441500	0.186000	0.093500	0
50%	0.545000	0.425000	0.140000	0.799500	0.336000	0.171000	0
75%	0.615000	0.480000	0.165000	1.153000	0.502000	0.253000	0
max	0.815000	0.650000	1.130000	2.825500	1.488000	0.760000	1

In [8]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4177 entries, 0 to 4176
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Sex             4177 non-null   object 
 1   Length          4177 non-null   float64
 2   Diameter        4177 non-null   float64
 3   Height          4177 non-null   float64
 4   Whole weight    4177 non-null   float64
 5   Shucked weight  4177 non-null   float64
 6   Viscera weight  4177 non-null   float64
 7   Shell weight    4177 non-null   float64
 8   Rings           4177 non-null   int64
dtypes: float64(7), int64(1), object(1)
memory usage: 293.8+ KB
```

In [9]:

```
data['Age']=data['Rings']+1.5
data.drop(['Rings'],axis=1, inplace=True)
data.head()
```

Out[9]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Age
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	16.5
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	8.5
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	10.5
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	11.5
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	8.5

In [10]:

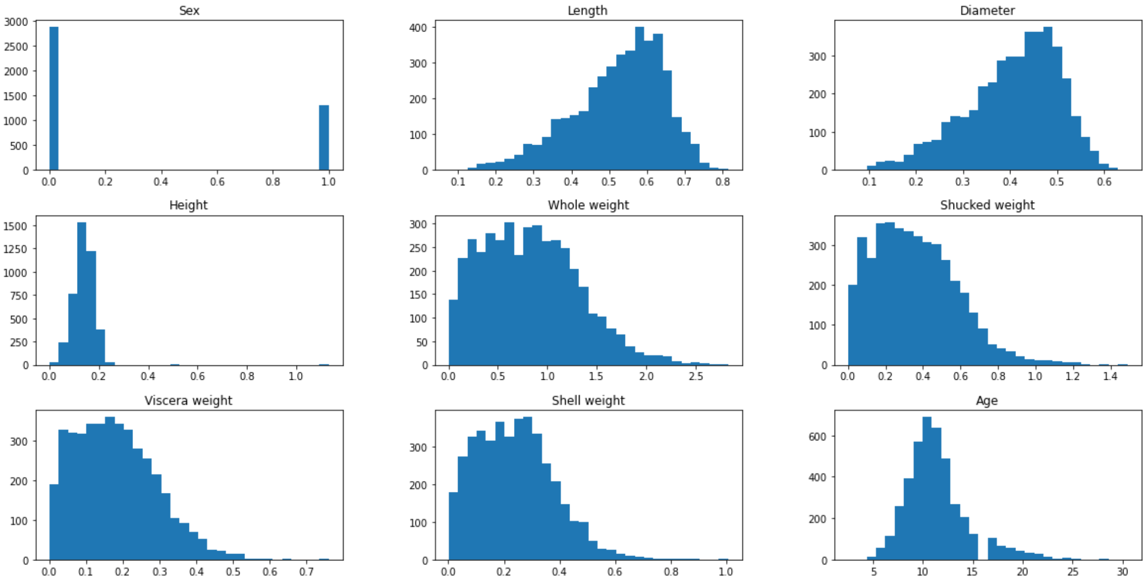
```
enc = OneHotEncoder(handle_unknown='ignore',sparse=False)
data['Sex']=enc.fit_transform(data[['Sex']])
```

In [11]:

```
data.hist(figsize=(20,10), grid=False, layout=(3, 3), bins = 30)
```

Out[11]:

```
array([[<AxesSubplot:title={'center':'Sex'}>,
        <AxesSubplot:title={'center':'Length'}>,
        <AxesSubplot:title={'center':'Diameter'}>],
       [<AxesSubplot:title={'center':'Height'}>,
        <AxesSubplot:title={'center':'Whole weight'}>,
        <AxesSubplot:title={'center':'Shucked weight'}>],
       [<AxesSubplot:title={'center':'Viscera weight'}>,
        <AxesSubplot:title={'center':'Shell weight'}>,
        <AxesSubplot:title={'center':'Age'}>]], dtype=object)
```



UNIVARIATE ANALYSIS

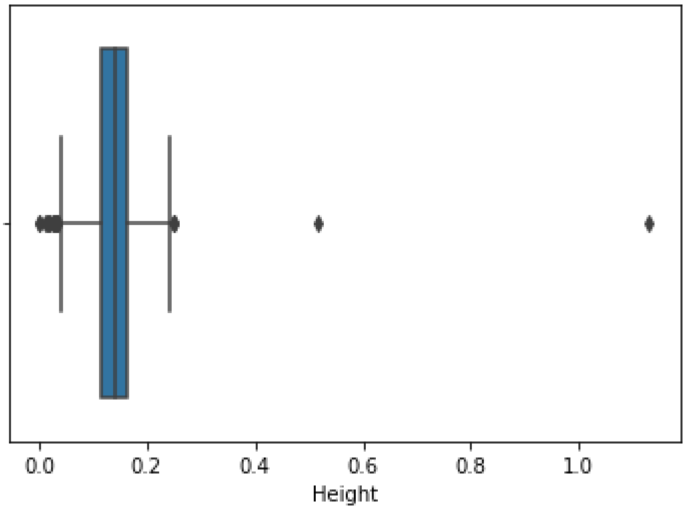
In [12]:

```
sns.boxplot(data['Height'])
```

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

Out[12]:

```
<AxesSubplot:xlabel='Height'>
```



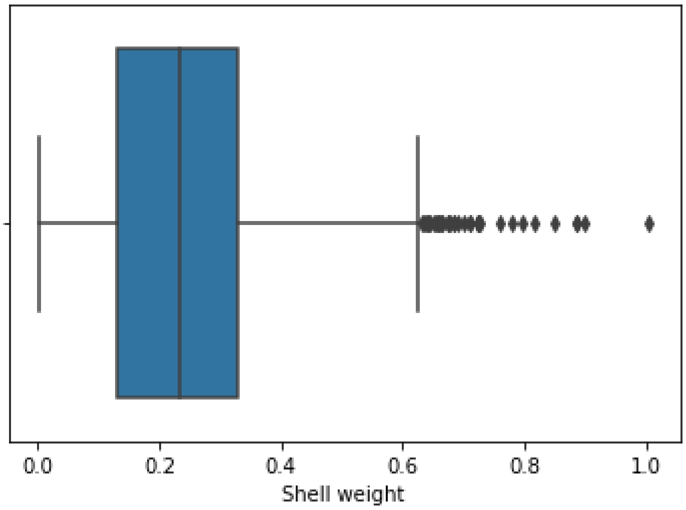
In [13]:

```
sns.boxplot(data['Shell weight'])
```

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

Out[13]:

```
<AxesSubplot:xlabel='Shell weight'>
```



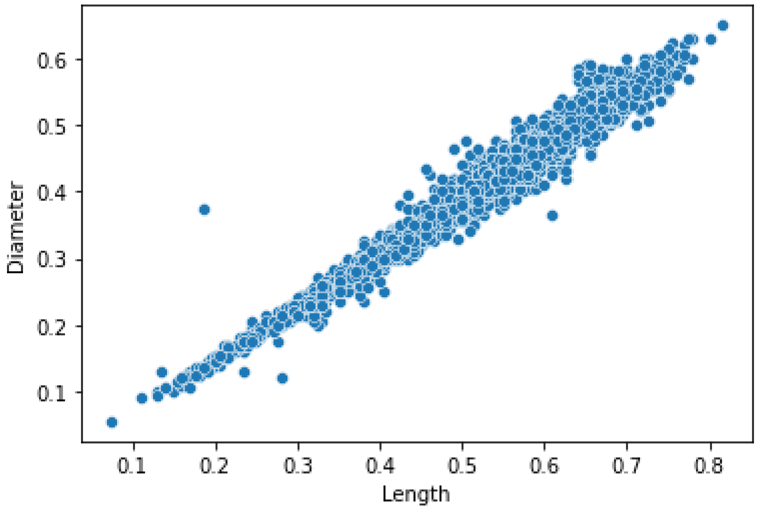
BIVARIATE ANALYSIS

In [14]:

```
sns.scatterplot(x=data['Length'],y=data['Diameter'])
```

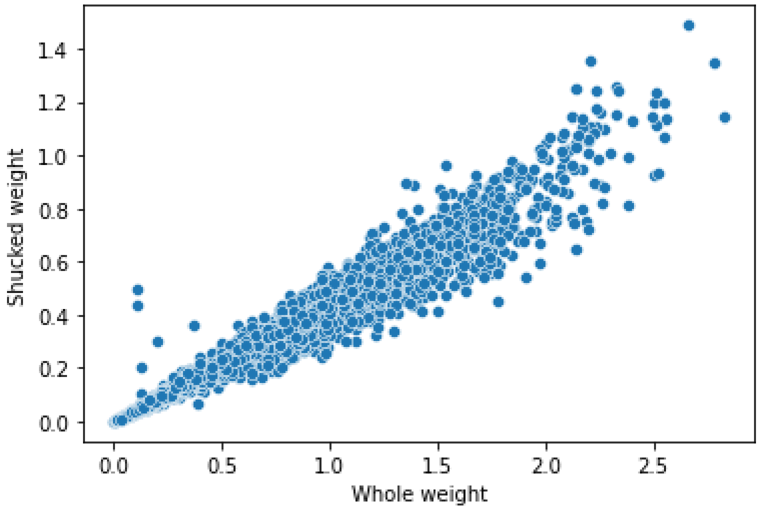
Out[14]:

```
<AxesSubplot:xlabel='Length', ylabel='Diameter'>
```



```
In [15]: sns.scatterplot(x=data['Whole weight'],y=data['Shucked weight'])
```

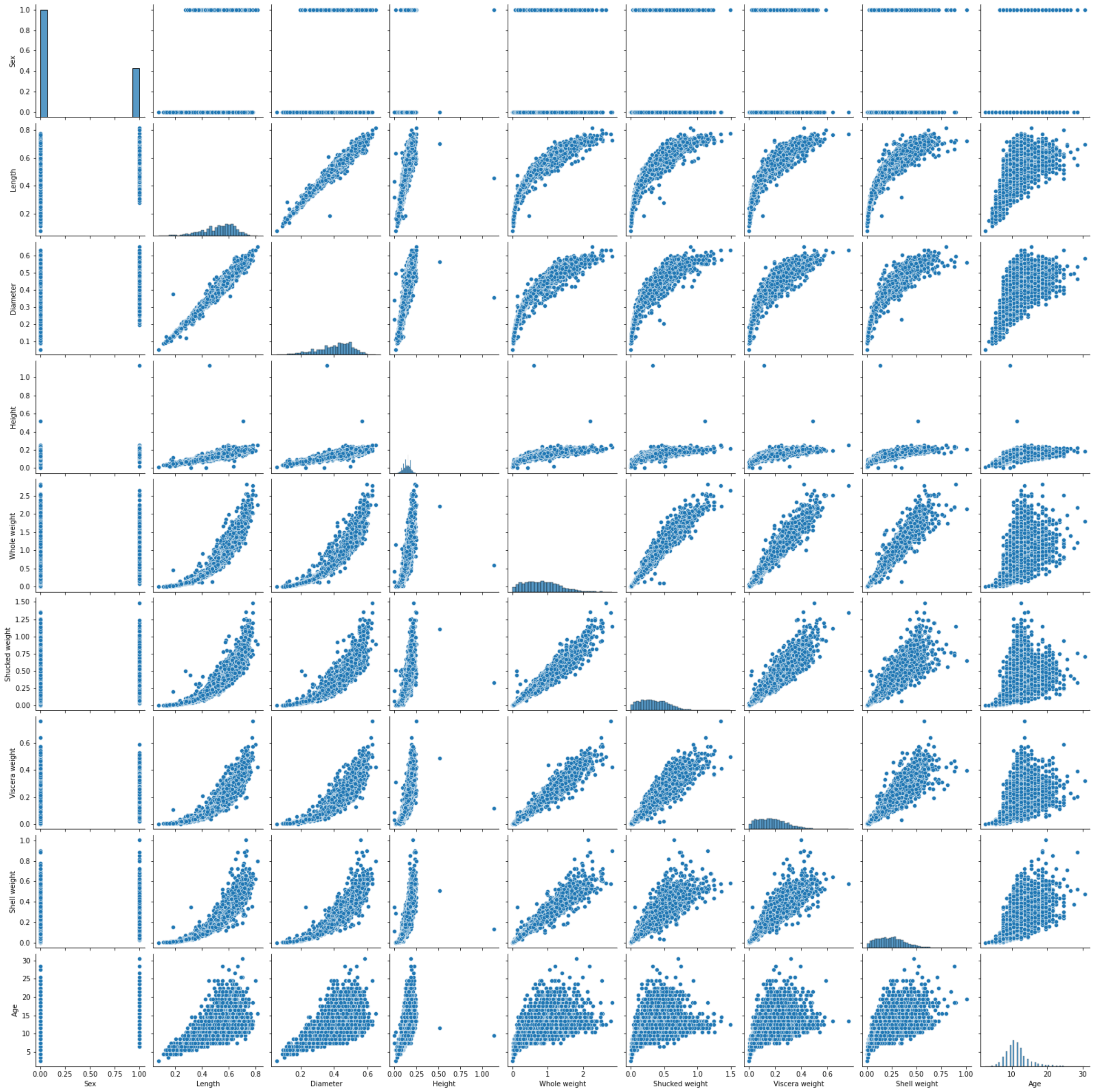
```
Out[15]: <AxesSubplot:xlabel='Whole weight', ylabel='Shucked weight'>
```



MULTIVARIATE ANALYSIS

```
In [16]: sns.pairplot(data)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1b3fcfb9eb0>
```



```
In [18]: X=data.drop(['Age'],axis=1)
         X= StandardScaler().fit_transform(X)
         y=data['Age']
```

TRAIN AND TEST SPLIT

```
In [19]: X_train, X_test, y_train, y_test = train_test_split(X,y,random_state=45,t
```

```
In [20]: model = LinearRegression()
         model.fit(X_train,y_train)

Out[20]: LinearRegression()
```

```
In [21]: model.score(X_test,y_test)

Out[21]: 0.5033622966639644
```

```
In [22]: y_predict=model.predict(X_test)
         mean_absolute_error(y_test,y_predict)

Out[22]: 1.6382175500512406
```

```
In [23]: explained_variance_score(y_test,y_predict)
```

```
Out[23]: 0.503601447711119
```

```
In [24]: r2_score(y_test,y_predict)
```

```
Out[24]: 0.5033622966639644
```