Project Title : Analytics for Hospitals' Health-Care Data

In [ ]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.api.types import is_numeric_dtype
sns.set()
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
sns.set_style("darkgrid")
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from sklearn.tree import DecisionTreeRegressor

from sklearn import  metrics
%matplotlib inline
```

LOAD THE DATASET

In [2]:
```python
abalone = pd.read_csv('abalone.csv', sep=',')
```

In [3]:
```python
abalone.head()
```

Out[3]:

	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

UNIVARIATE ANALYSIS

In [4]:
```python
rows = 2
cols = 2
i = 0

plt.figure(figsize=(cols * 5, rows * 5))

i += 1
plt.subplot(rows, cols, i)
plt.xticks(range(0, 31, 4))
plt.xlim(0, 30)
_ = sns.distplot(abalone['Rings'], kde=False, bins=range(0, 31, 2))

i += 1
plt.subplot(rows, cols, i)
_ = sns.distplot(abalone['Rings'])

i += 1
plt.subplot(rows, cols, i)
plt.xticks(range(0, 31, 4))
plt.xlim(0, 30)
_ = sns.boxplot(abalone['Rings'])
```

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWar
ning: `distplot` is a deprecated function and will be removed in a future versi
on. Please adapt your code to use either `displot` (a figure-level function wit
h similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
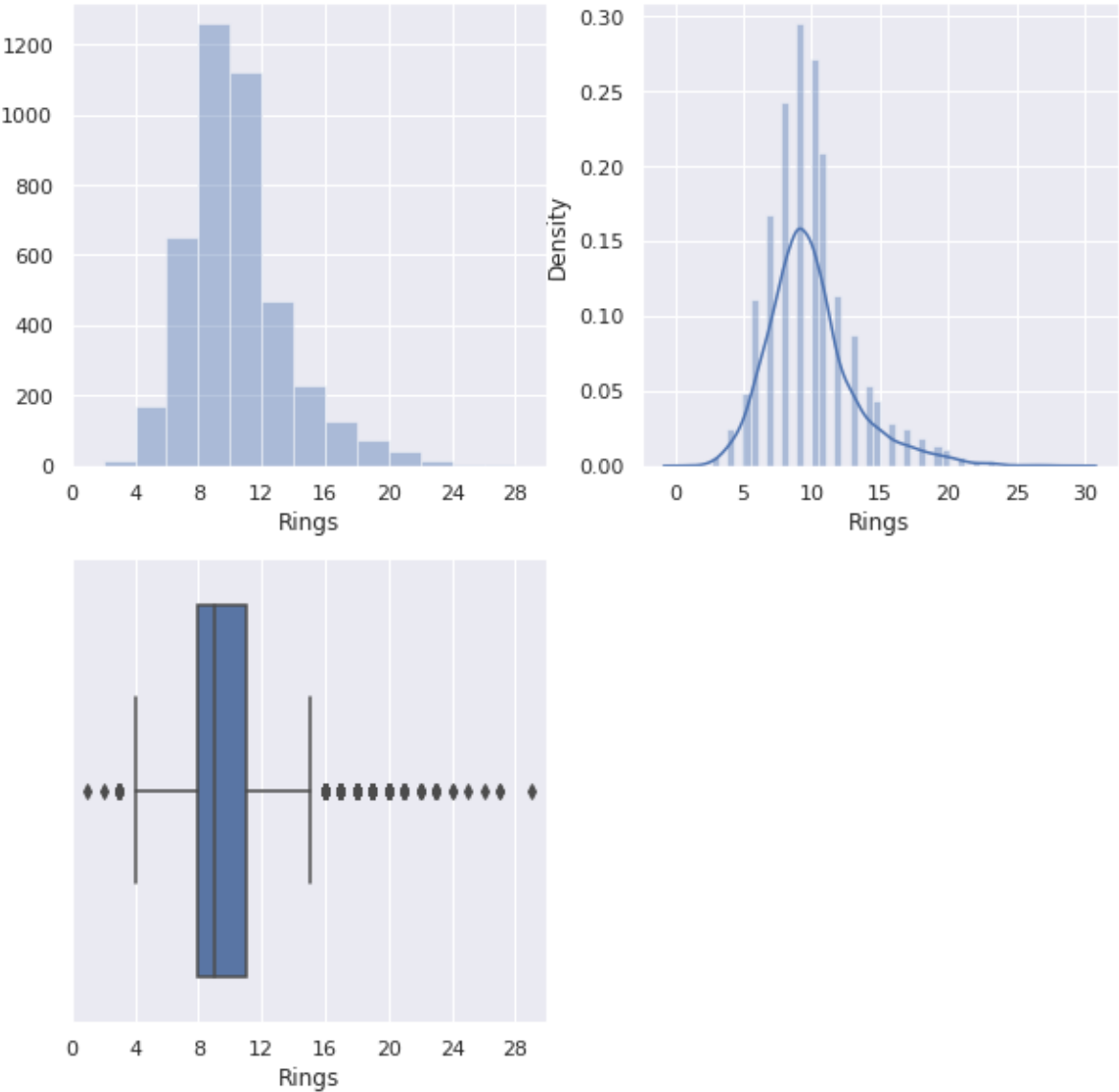
In [5]:
```python
abalone = abalone[abalone['Height'] < 0.4]

plt.figure(figsize=(15, 15))

colors = sns.color_palette()

lines = 3
rows = 3
i = 0

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Length'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Diameter'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Height'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Length'], kde=False, bins=np.arange(0.0, 0.9, 0.05), c

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Diameter'], kde=False, bins=np.arange(0.0, 0.7, 0.05),

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Height'], kde=False, bins=10, color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(abalone['Length'], color=sns.color_palette()[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(abalone['Diameter'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(abalone['Height'], color=colors[i % 3])
```
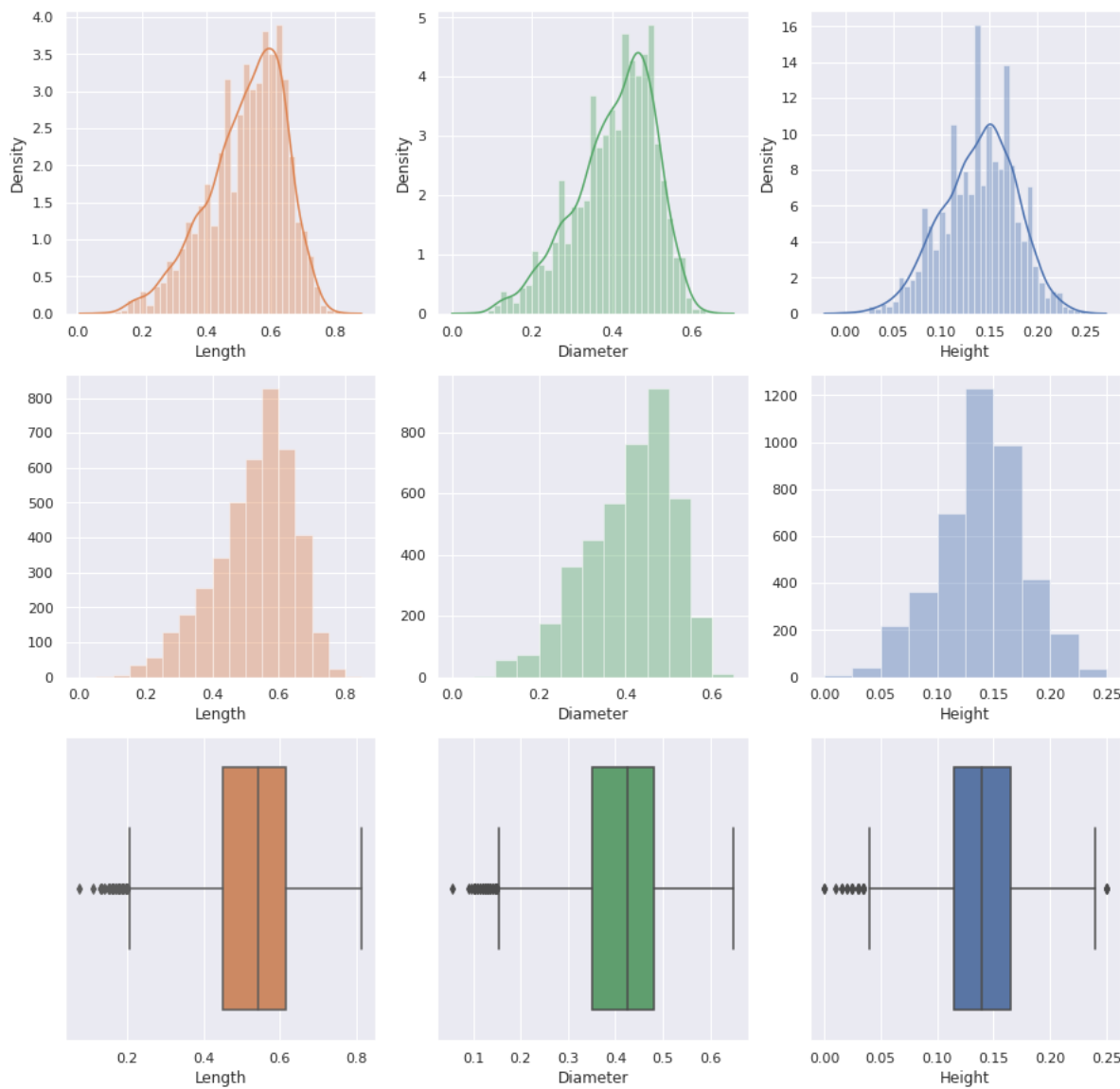
```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
```

```
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

```
In [6]: plt.figure(figsize=(20, 15))

        colors = sns.color_palette()

        rows = 3
        cols = 4
        i = 0

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Whole weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Shucked weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Viscera weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Shell weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Whole weight'], kde=False, bins=14, color=colors[i % cd

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Shucked weight'], kde=False, bins=14, color=colors[i %

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Viscera weight'], kde=False, bins=16, color=colors[i %

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.distplot(abalone['Shell weight'], kde=False, bins=20, color=colors[i % cd

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.boxplot(abalone['Whole weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.boxplot(abalone['Shucked weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.boxplot(abalone['Viscera weight'], color=colors[i % cols])

        i += 1
        plt.subplot(rows, cols, i)
        _ = sns.boxplot(abalone['Shell weight'], color=colors[i % cols])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureW
arning: `distplot` is a deprecated function and will be removed in a future v
ersion. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histog
rams).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
ng: Pass the following variable as a keyword arg: x. From version 0.12, the o
nly valid positional argument will be `data`, and passing other arguments wit
hout an explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
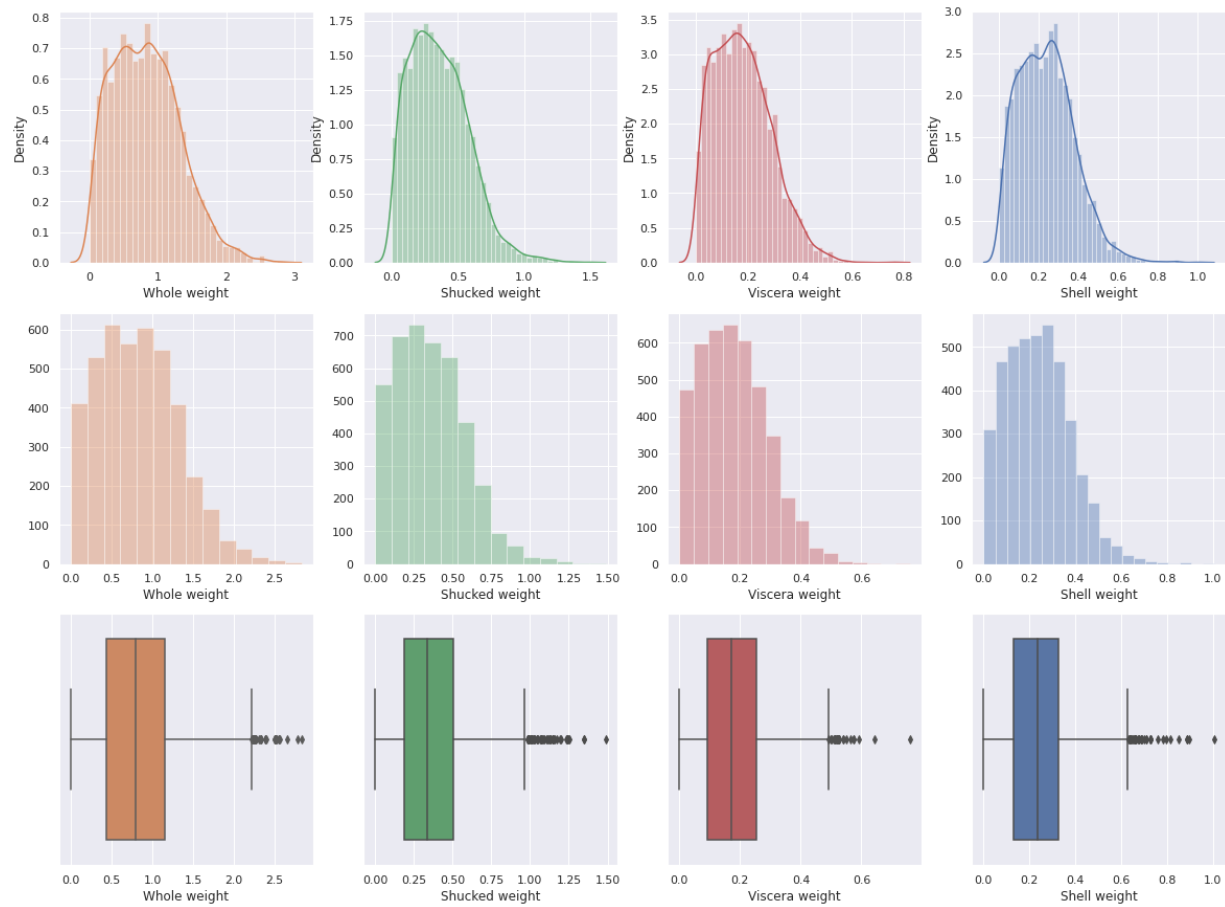
In [7]:
```python
plt.figure(figsize=(5,5))
_ = sns.countplot(abalone.Sex)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: 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.
  FutureWarning
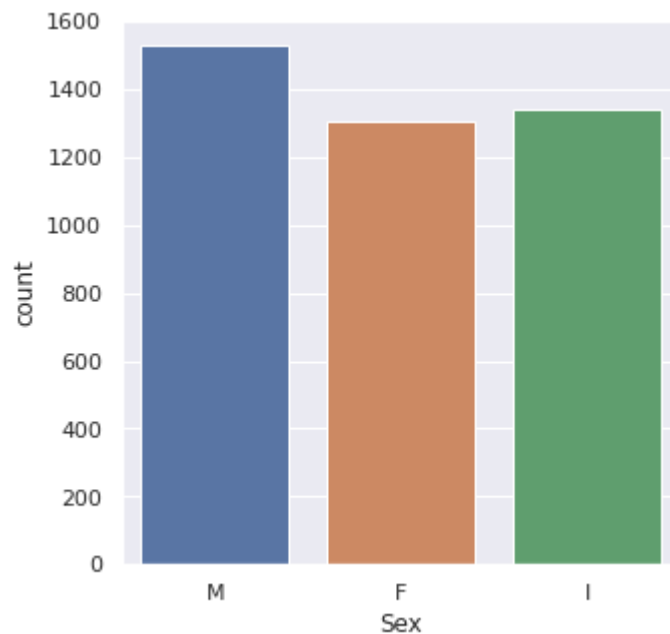
BIVARATE ANALYSIS

In [8]:
```python
plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=abalone, x='Sex', y='Rings')
```
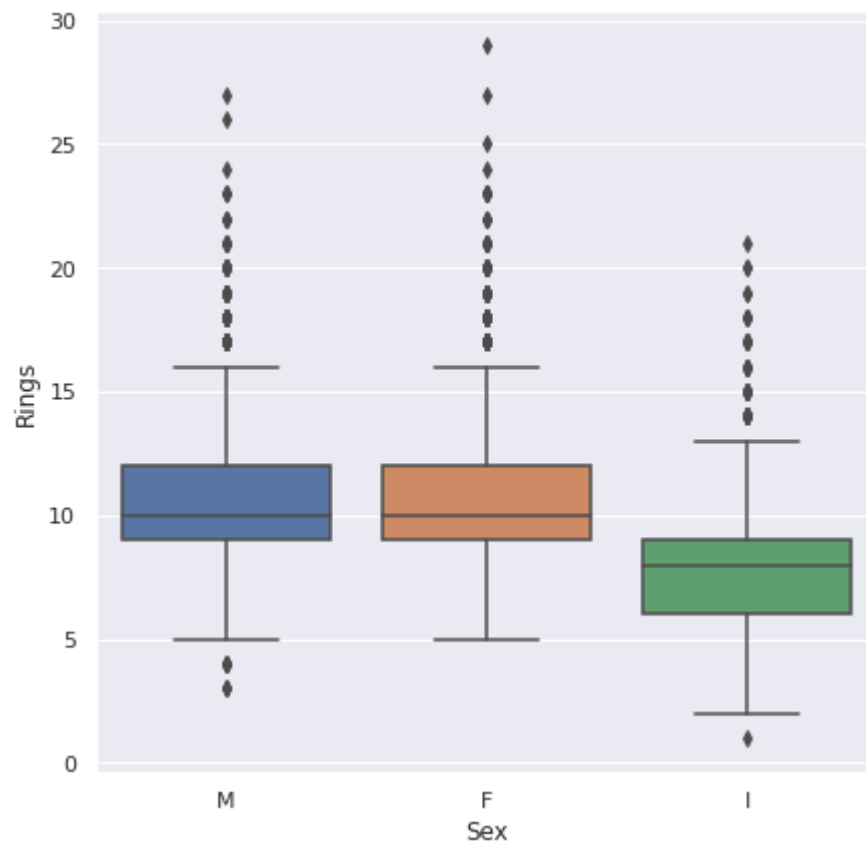
In [9]:
```python
plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=abalone, x='Sex', y='Height')
```
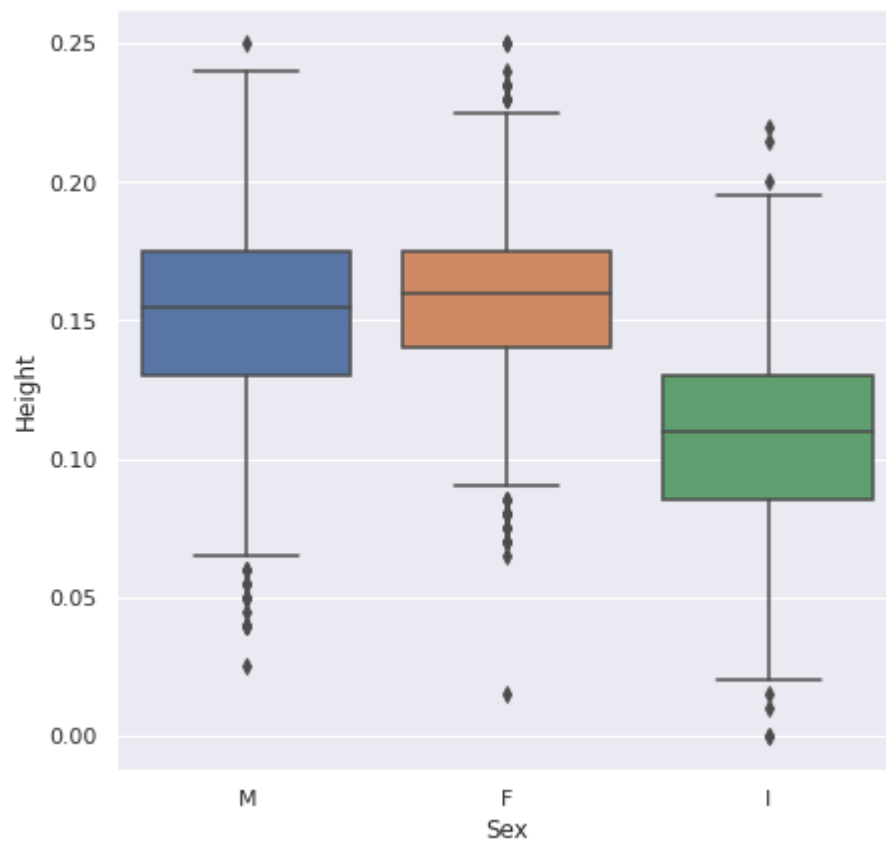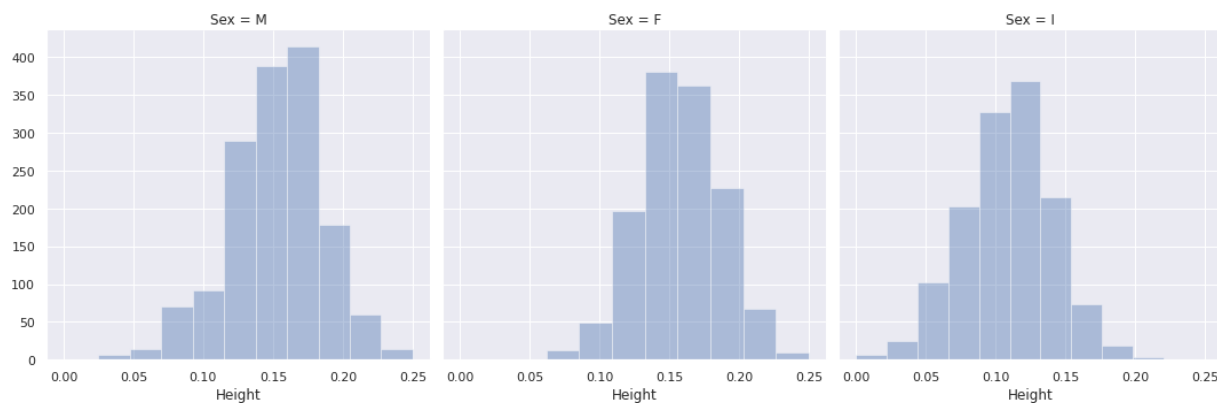
```
In [10]: g = sns.FacetGrid(abalone, col='Sex', margin_titles=True, size=5)
         _ = g.map(sns.distplot, 'Height', kde=False, bins=10)
```

/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: Th
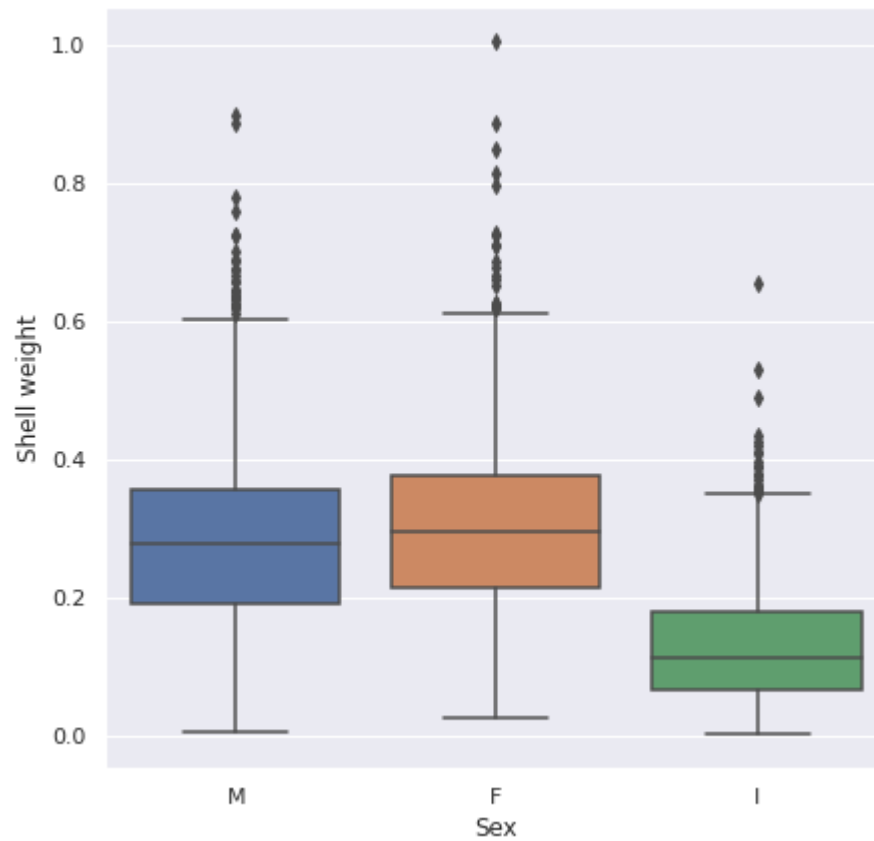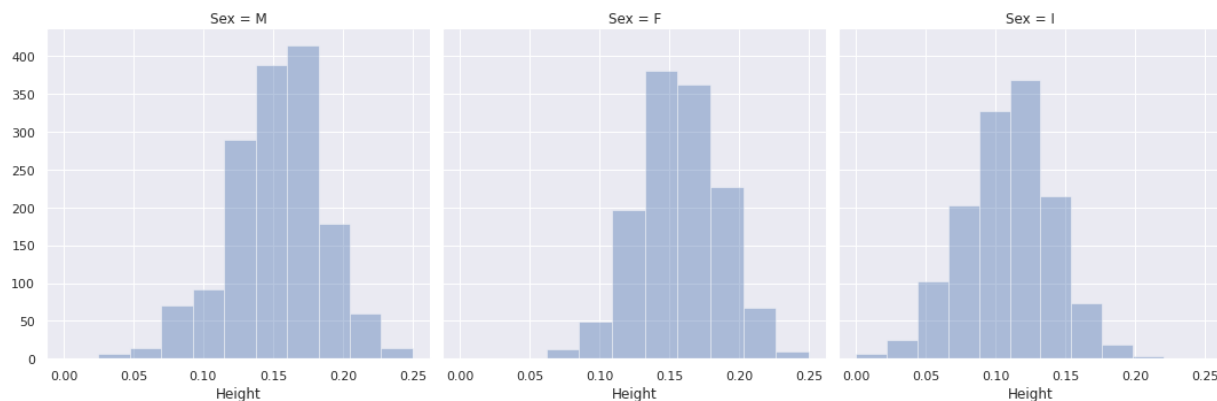e `size` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWar
ning: `distplot` is a deprecated function and will be removed in a future versi
on. Please adapt your code to use either `displot` (a figure-level function wit
h similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

In [11]:
```python
plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=abalone, x='Sex', y='Shell weight')
```

In [12]:
```python
g = sns.FacetGrid(abalone, col='Sex', margin_titles=True, size=5)
_ = g.map(sns.distplot, 'Height', kde=False, bins=10)
```

/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: Th
e `size` parameter has been renamed to `height`; please update your code.
 warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWar
ning: `distplot` is a deprecated function and will be removed in a future versi
on. Please adapt your code to use either `displot` (a figure-level function wit
h similar flexibility) or `histplot` (an axes-level function for histograms).
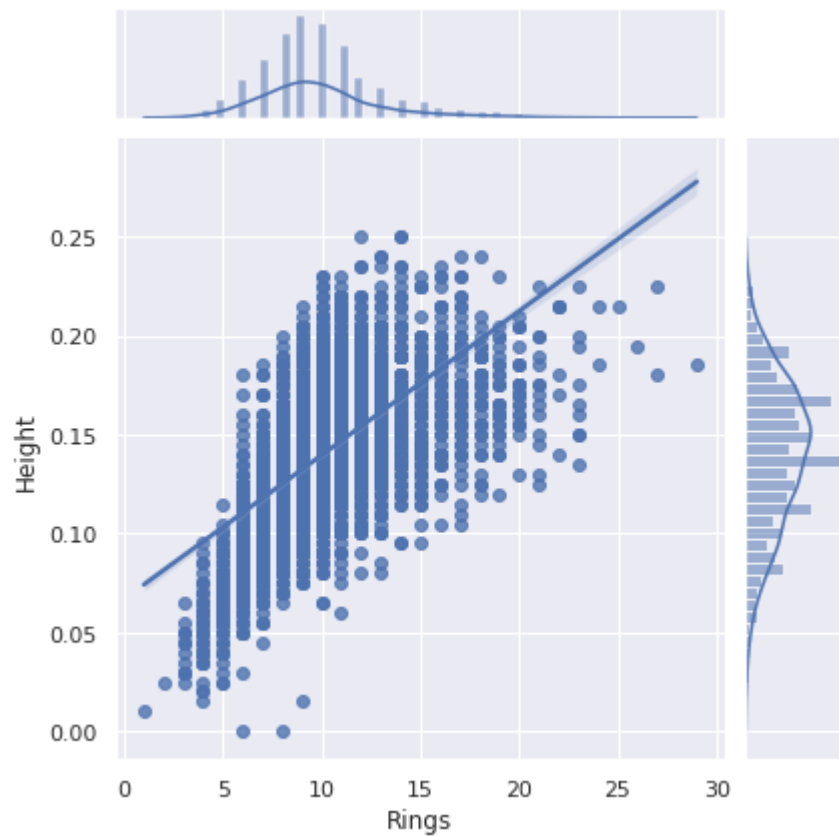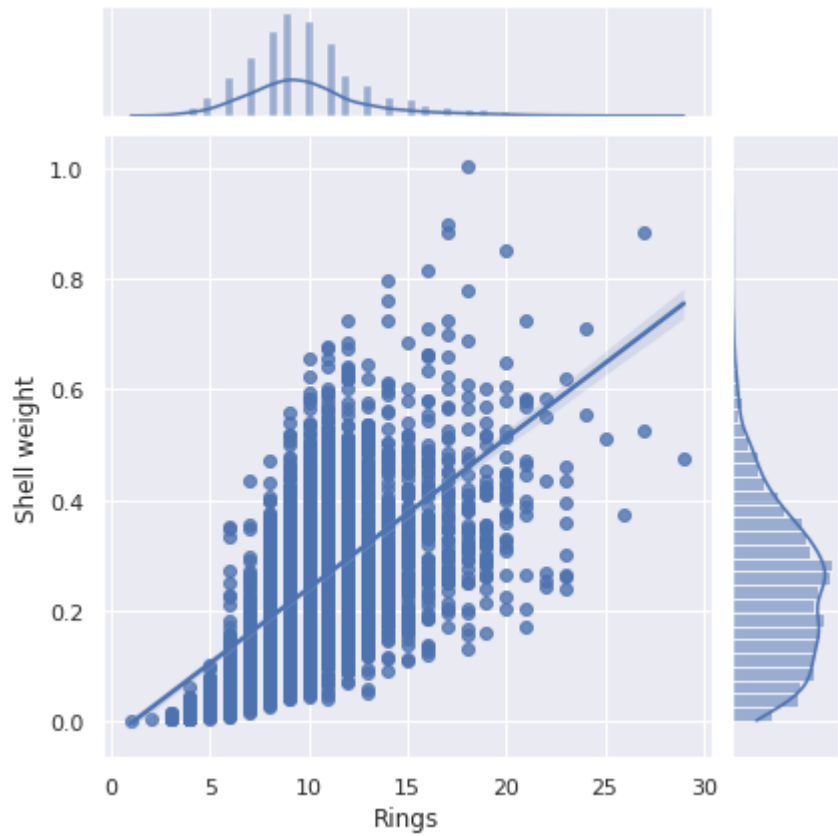 warnings.warn(msg, FutureWarning)

```
In [13]: plt.figure(figsize=(10, 10))
         corr = abalone.corr()
         _ = sns.heatmap(corr, annot=True)
```
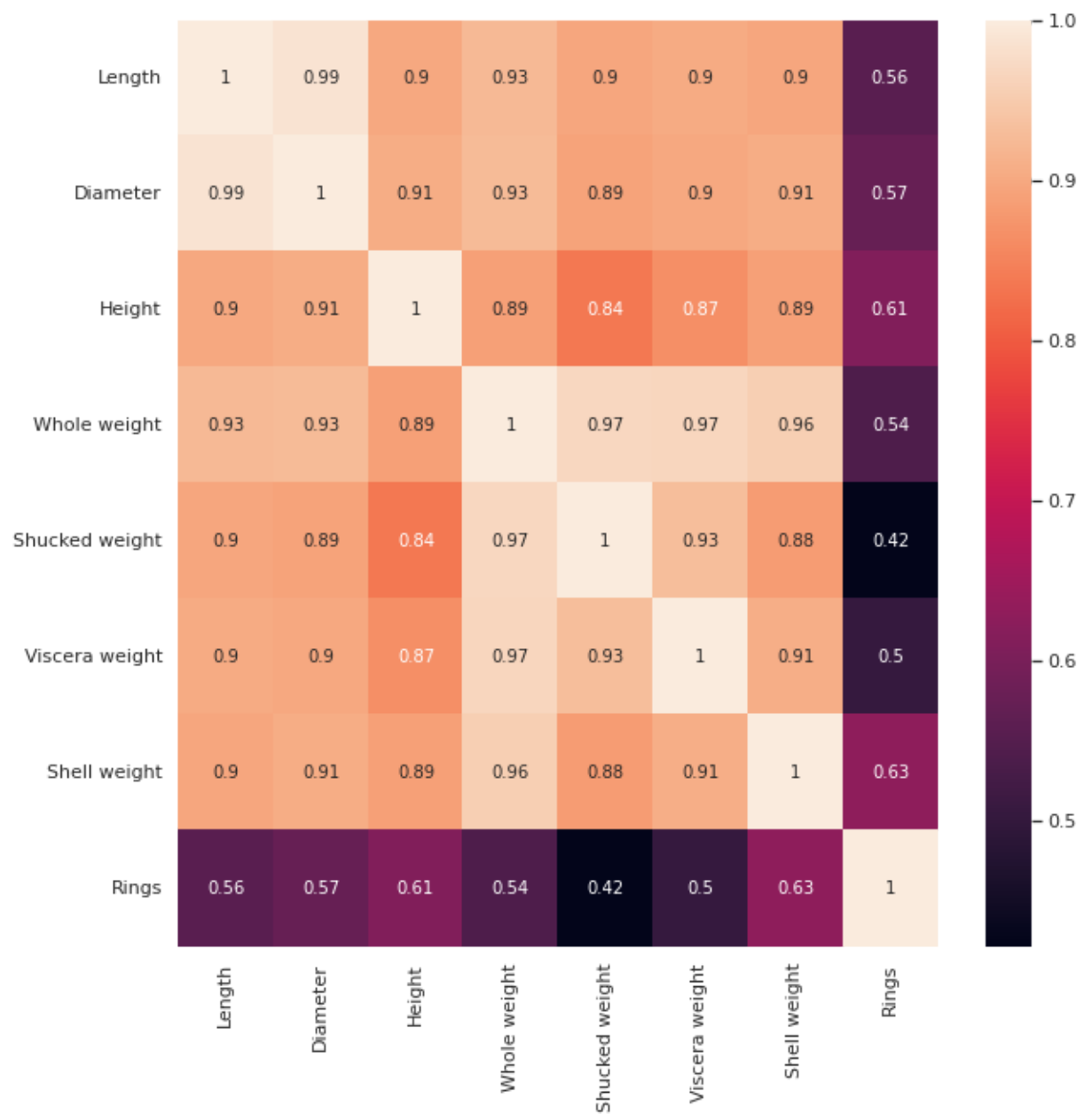
```
In [14]: plt.figure(figsize=(20, 5))

_ = sns.jointplot(data=abalone, x='Rings', y='Height', kind='reg')
_ = sns.jointplot(data=abalone, x='Rings', y='Shell weight', kind='reg')
```

<Figure size 1440x360 with 0 Axes>

DESCRIPTIVE STATISTICS

In [15]: `abalone.describe().T`

Out[15]:

	count	mean	std	min	25%	50%	75%	max
Length	4175.0	0.523965	0.120084	0.0750	0.45000	0.5450	0.61500	0.8150
Diameter	4175.0	0.407856	0.099230	0.0550	0.35000	0.4250	0.48000	0.6500
Height	4175.0	0.139189	0.038489	0.0000	0.11500	0.1400	0.16500	0.2500
Whole weight	4175.0	0.828468	0.490027	0.0020	0.44150	0.7995	1.15300	2.8255
Shucked weight	4175.0	0.359195	0.221713	0.0010	0.18600	0.3360	0.50175	1.4880
Viscera weight	4175.0	0.180536	0.109534	0.0005	0.09325	0.1710	0.25275	0.7600
Shell weight	4175.0	0.238791	0.139162	0.0015	0.13000	0.2340	0.32875	1.0050
Rings	4175.0	9.934132	3.224802	1.0000	8.00000	9.0000	11.00000	29.0000

HANDLING WITH MISSING DATA

In [16]:
```python
df = pd.DataFrame(abalone)
df.isnull()
```

Out[16]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
4172	False	False	False	False	False	False	False	False	False
4173	False	False	False	False	False	False	False	False	False
4174	False	False	False	False	False	False	False	False	False
4175	False	False	False	False	False	False	False	False	False
4176	False	False	False	False	False	False	False	False	False

4175 rows × 9 columns

In [17]:
```python
df.fillna(0)
```

Out[17]:

|	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	0.1500	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7
...	...	...	...	...	...	...	...	...	...
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

4175 rows × 9 columns

OUTLIERS IN EACH ATTRIBUTES

In [18]: `sns.boxplot(df['Length'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

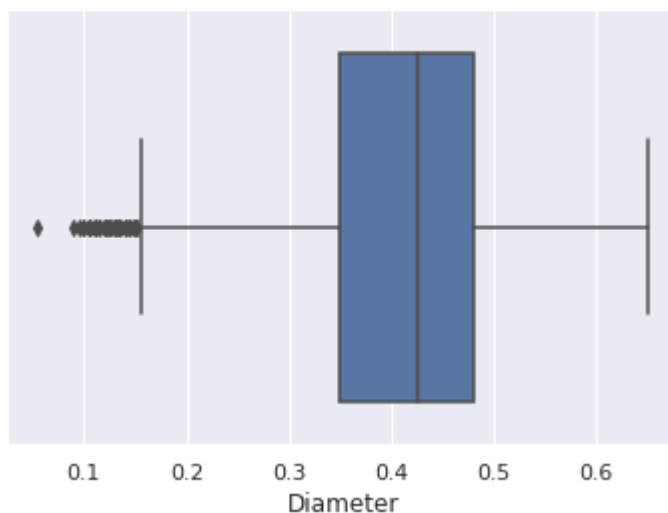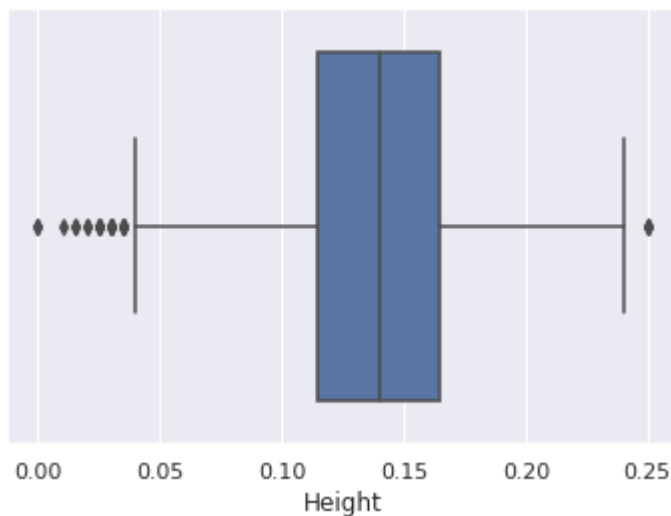Out[18]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc658ae4d0>`
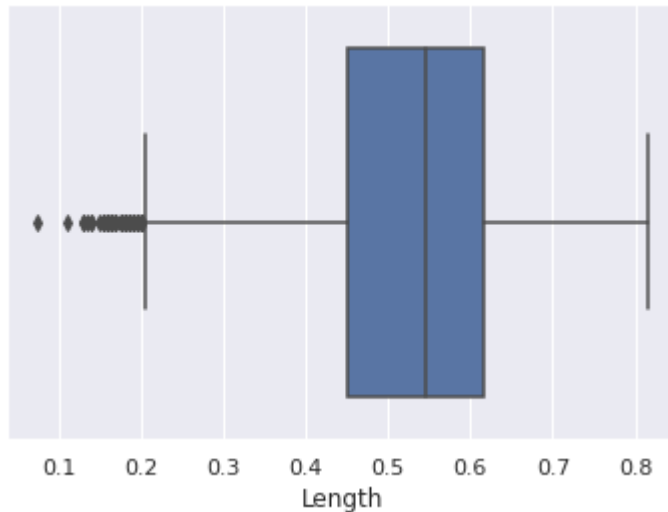
In [19]: `sns.boxplot(df['Diameter'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
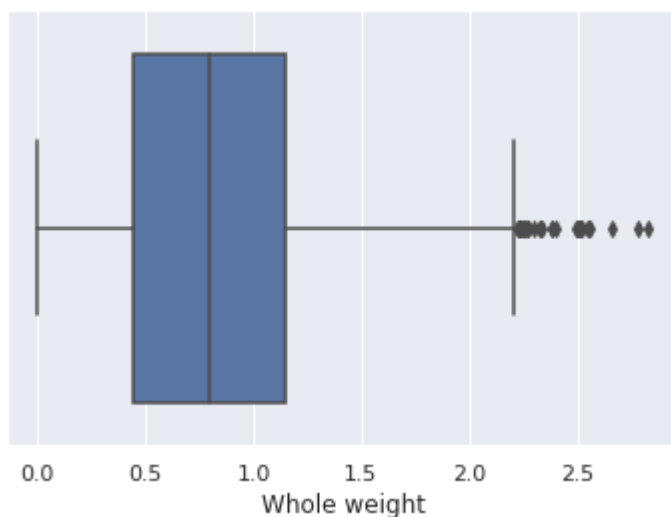
Out[19]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc63c9f190>`

In [20]: `sns.boxplot(df['Height'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

Out[20]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc65679450>`

In [21]: `sns.boxplot(df['Whole weight'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
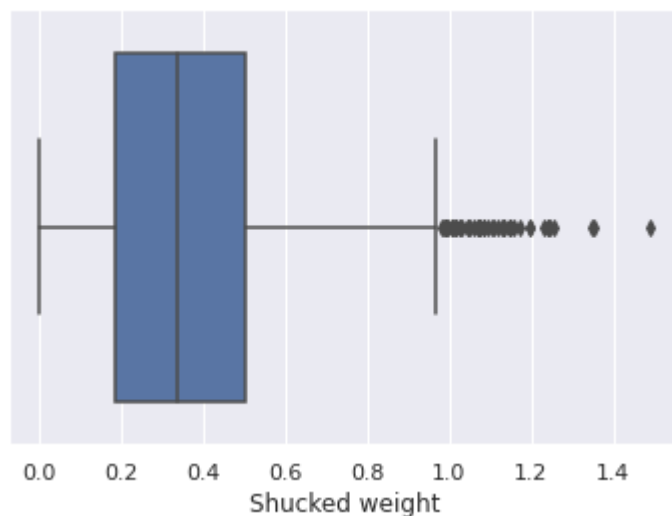
Out[21]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc65661610>`

In [22]: `sns.boxplot(df['Shucked weight'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
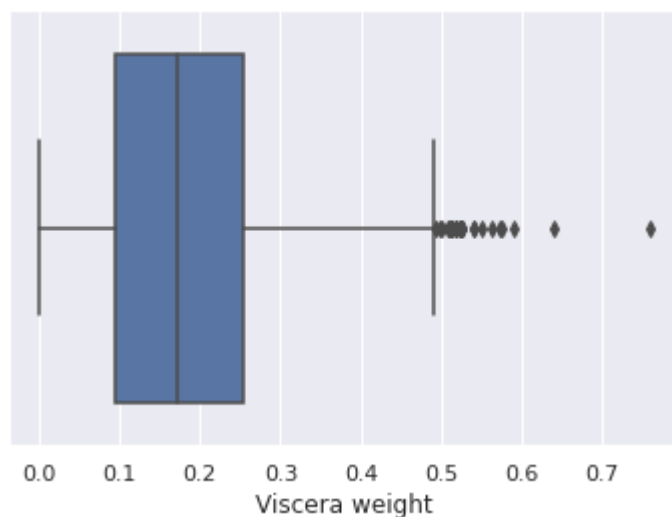
Out[22]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc6563a350>`

In [23]: `sns.boxplot(df['Viscera weight'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
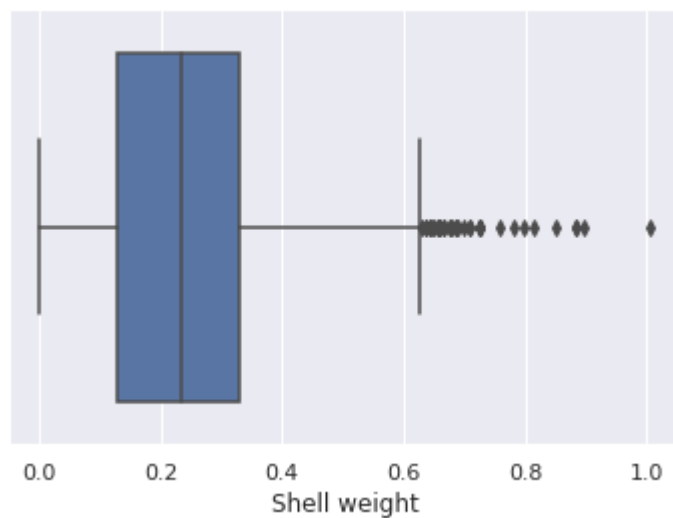
Out[23]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc6565a2d0>`

In [24]: `sns.boxplot(df['Shell weight'],data=df)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
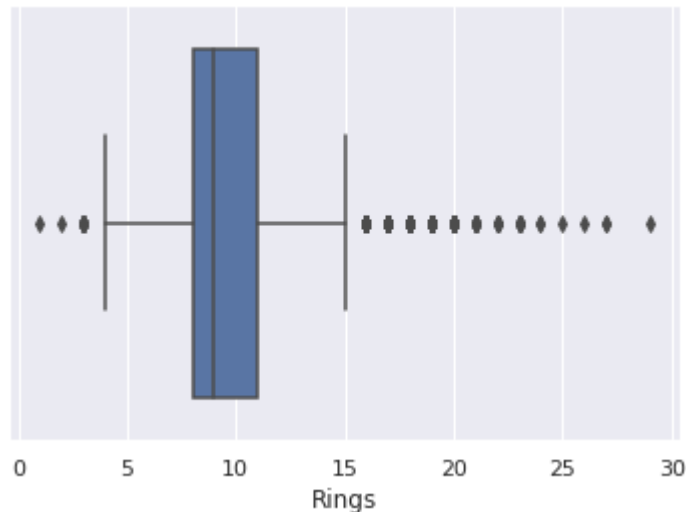
Out[24]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc65a26290>`

In [25]:
```python
sns.boxplot(df['Rings'],data=df)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc661856d0>

In [26]:
```python
Q1 = abalone.quantile(0.25)
Q3 = abalone.quantile(0.75)
IQR = Q3-Q1
print(IQR)
```

```
Length            0.16500
Diameter          0.13000
Height            0.05000
Whole weight      0.71150
Shucked weight    0.31575
Viscera weight    0.15950
Shell weight      0.19875
Rings             3.00000
dtype: float64
```

In [27]:
```python
abalone = abalone[~((abalone < (Q1 - 1.5 * IQR)) |(abalone > (Q3 + 1.5 * IQR))).a
abalone.shape
```
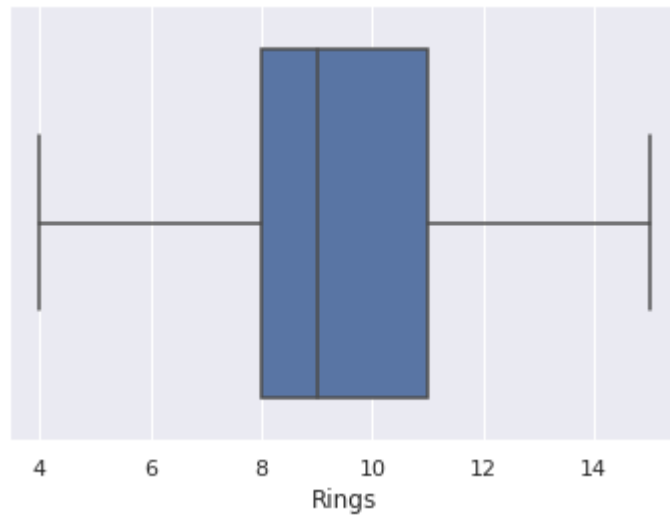
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning:
Automatic reindexing on DataFrame vs Series comparisons is deprecated and will
raise ValueError in a future version. Do `left, right = left.align(right, axis
=1, copy=False)` before e.g. `left == right`
  """Entry point for launching an IPython kernel.

Out[27]: (3781, 9)

```
In [28]: sns.boxplot(abalone['Rings'],data=abalone)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning

Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc6592e290>

In [29]:
```python
sns.boxplot(abalone['Length'],data=abalone)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
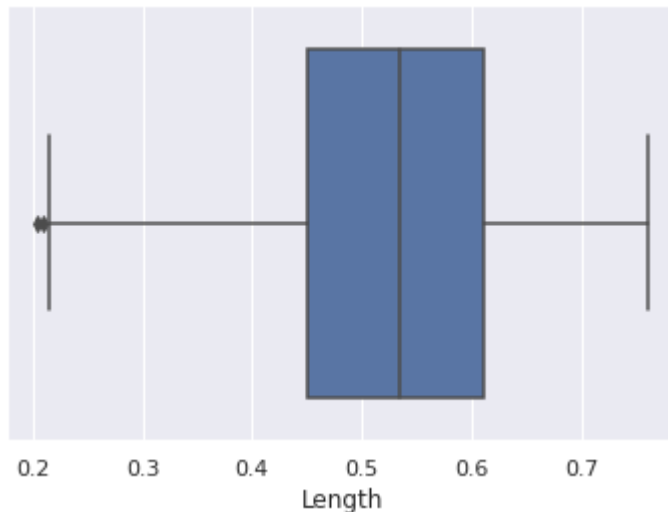 FutureWarning

Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc658a63d0>

In [30]:
```python
sns.boxplot(abalone['Height'],data=abalone)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
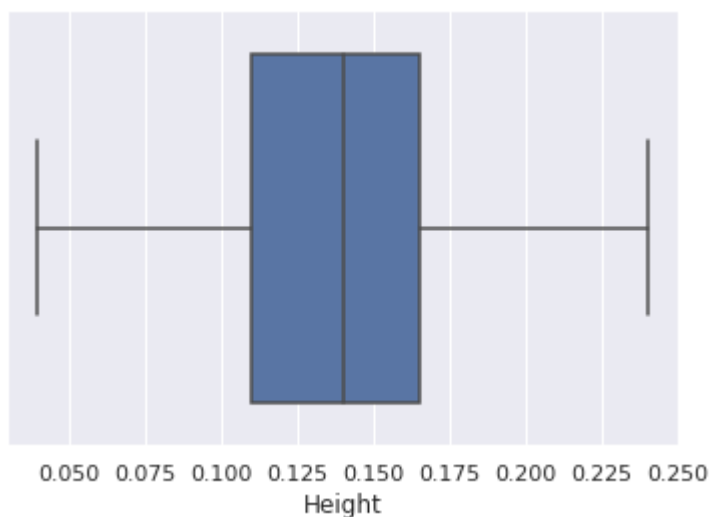 FutureWarning

Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc65975710>

In [31]: `sns.boxplot(abalone['Diameter'],data=abalone)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
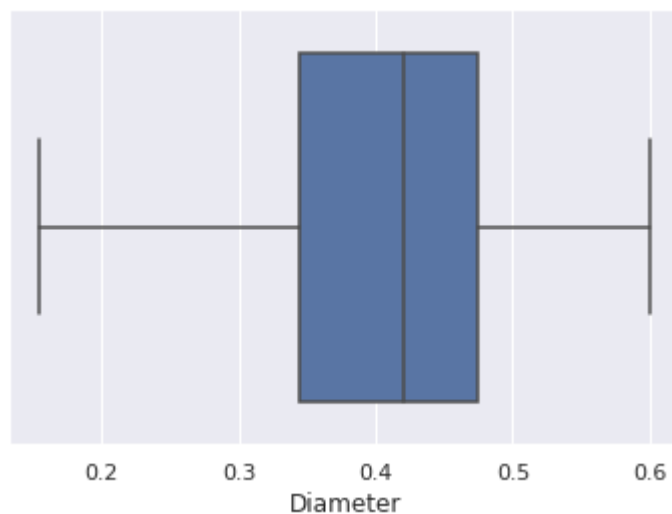
Out[31]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc6377bb10>`

In [32]: `sns.boxplot(abalone['Whole weight'],data=abalone)`

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
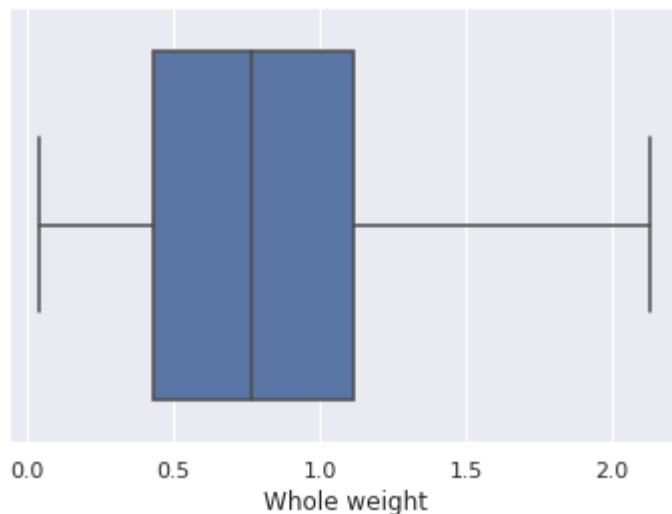n explicit keyword will result in an error or misinterpretation.
  FutureWarning

Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc6555d210>

In [33]: `sns.boxplot(abalone['Shucked weight'],data=abalone)`

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning

Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0x7efc63b167d0>

In [34]: `sns.boxplot(abalone['Shell weight'],data=abalone)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
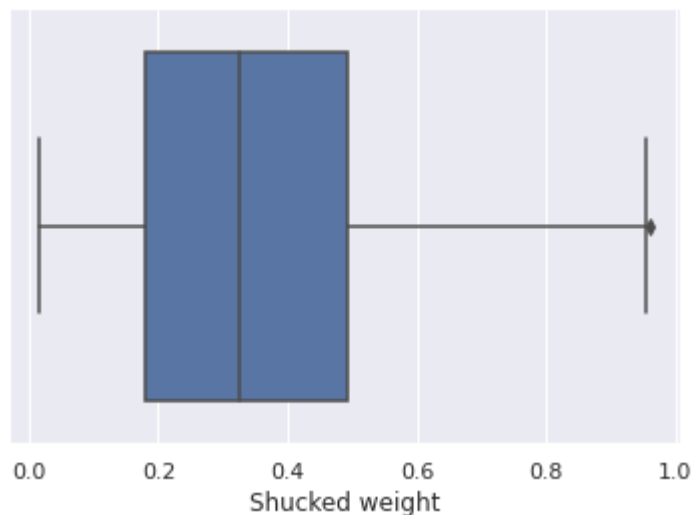
Out[34]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc6568c250>`

In [35]: `sns.boxplot(abalone['Viscera weight'],data=abalone)`

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarnin
g: 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 a
n explicit keyword will result in an error or misinterpretation.
  FutureWarning
```
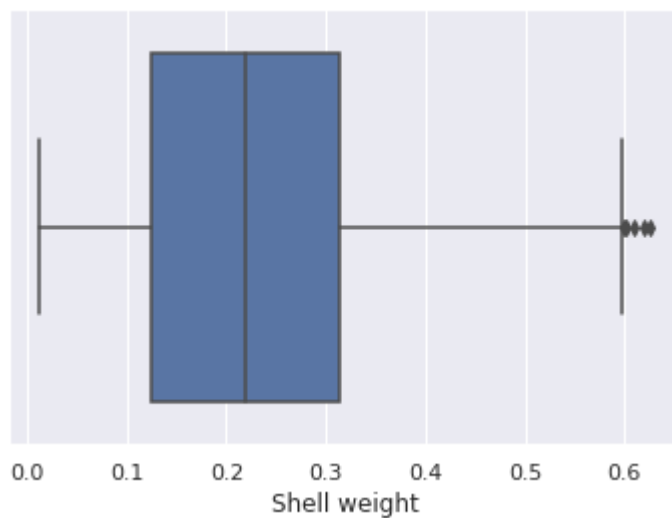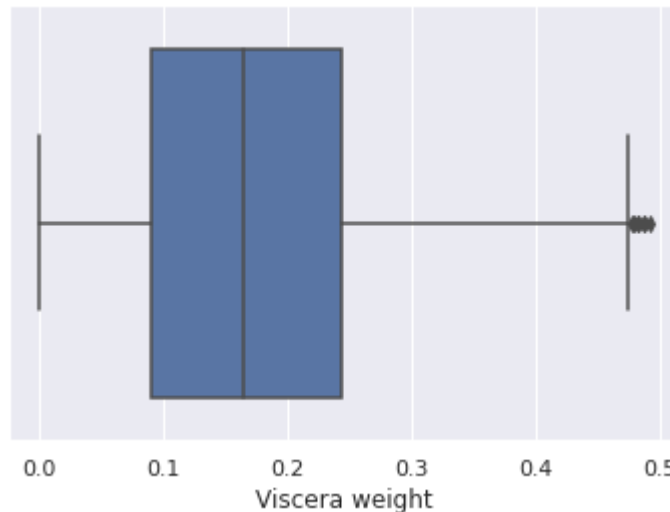
Out[35]: `<matplotlib.axes._subplots.AxesSubplot at 0x7efc63a00a10>`

LABEL ENCODING FOR CATEGORICAL DATA

In [36]: `le=LabelEncoder()`
`abalone['Sex']=le.fit_transform(abalone['Sex'])`

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
```

In [ ]:

In [37]: abalone

Out[37]:

	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	0.1500	15
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7
...	...	...	...	...	...	...	...	...	...
4172	0	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	11
4173	2	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	10
4174	2	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	9
4175	0	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
4176	2	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950	12

3781 rows × 9 columns

Spliting the Data into dependent and Independent Variables

```
In [38]: X = abalone.iloc[:, :-1].values
         y = abalone.iloc[:, -1].values
```

Scaling independent variables

```
In [39]: scaler = StandardScaler()
         scaler.fit(abalone)
```

Out[39]: StandardScaler()

Spliting training and test data

```
In [40]: train_X,val_X,train_y,val_y = train_test_split(X, y, test_size = 0.2, random_stat
```

```
In [41]: print("Shape of Training X :",train_X.shape)
         print("Shape of Validation X :",val_X.shape)

         Shape of Training X : (3024, 8)
         Shape of Validation X : (757, 8)
```

In [42]:
```python
print("Shape of Training y :",train_y.shape)
print("Shape of Validation y :",val_y.shape)
```

```
Shape of Training y : (3024,)
Shape of Validation y : (757,)
```

##LINEAR REGRESSION

In [43]:
```python
lr = LinearRegression()
lr.fit(train_X,train_y)
```

Out[43]: LinearRegression()

In [44]:
```python
%%time
y_pred_val_lr = lr.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_lr)
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_lr))
print("\n")
print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pre
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_lr))
print("\n")
```

```
MAE on Validation set : 1.2719689486359298


MSE on Validation set : 2.7606215450501024


RMSE on Validation set : 1.127816008325795


R2 Score on Validation set : 0.5119499107890585


CPU times: user 5.67 ms, sys: 859 µs, total: 6.53 ms
Wall time: 6.14 ms
```

##SUPPORT VECTOR MACHINE

In [45]:
```python
svm = SVR()
svm.fit(train_X,train_y)
```

Out[45]: SVR()

```
In [46]: %%time
         y_pred_val_svm = svm.predict(val_X)
         print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_svm
         print("\n")
         print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_svm)
         print("\n")
         print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pre
         print("\n")
         print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_svm))
         print("\n")
```

MAE on Validation set : 1.2208952787270895

MSE on Validation set : 2.7012620714060267

RMSE on Validation set : 1.1049413010323623

R2 Score on Validation set : 0.5224440679687887

CPU times: user 146 ms, sys: 0 ns, total: 146 ms
Wall time: 145 ms

##DECISION TREE REGRESSOR

```
In [47]: dc = DecisionTreeRegressor(random_state = 0)
         dc.fit(train_X,train_y)
```

Out[47]: DecisionTreeRegressor(random_state=0)

In [48]:
```python
%%time
y_pred_val_dc = dc.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_dc)
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_dc)
print("\n")
print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pre
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_dc))
print("\n")
```

MAE on Validation set : 1.6393659180977542

MSE on Validation set : 4.88110964332893

RMSE on Validation set : 1.2803772561623212

R2 Score on Validation set : 0.13706896870869845

CPU times: user 10.1 ms, sys: 1.1 ms, total: 11.2 ms
Wall time: 24.9 ms

In [ ]:

##OVERVIEW OF R2 SCORES OF ALL MODELS

In [49]:
```python
print('Logistic Regression R2 Score on Validation set :',metrics.r2_score(val_y,
print('SVR R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_svm)
print('Decision Tree Regressor R2 Score on Validation set :',metrics.r2_score(val
```

Logistic Regression R2 Score on Validation set : 0.5119499107890585
SVR R2 Score on Validation set : 0.5224440679687887
Decision Tree Regressor R2 Score on Validation set : 0.13706896870869845