

```
from google.colab import files
uploaded = files.upload()
```

abalone.csv

- **abalone.csv**(text/csv) - 191962 bytes, last modified: 11/7/2022 - 100% done
Saving abalone.csv to abalone.csv

```
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
```

```
abalone = pd.read_csv('abalone.csv', sep=',')
```

```
abalone.head()
```

	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

```
rows = 2
cols = 2
i = 0
```

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

```
i += 1
```

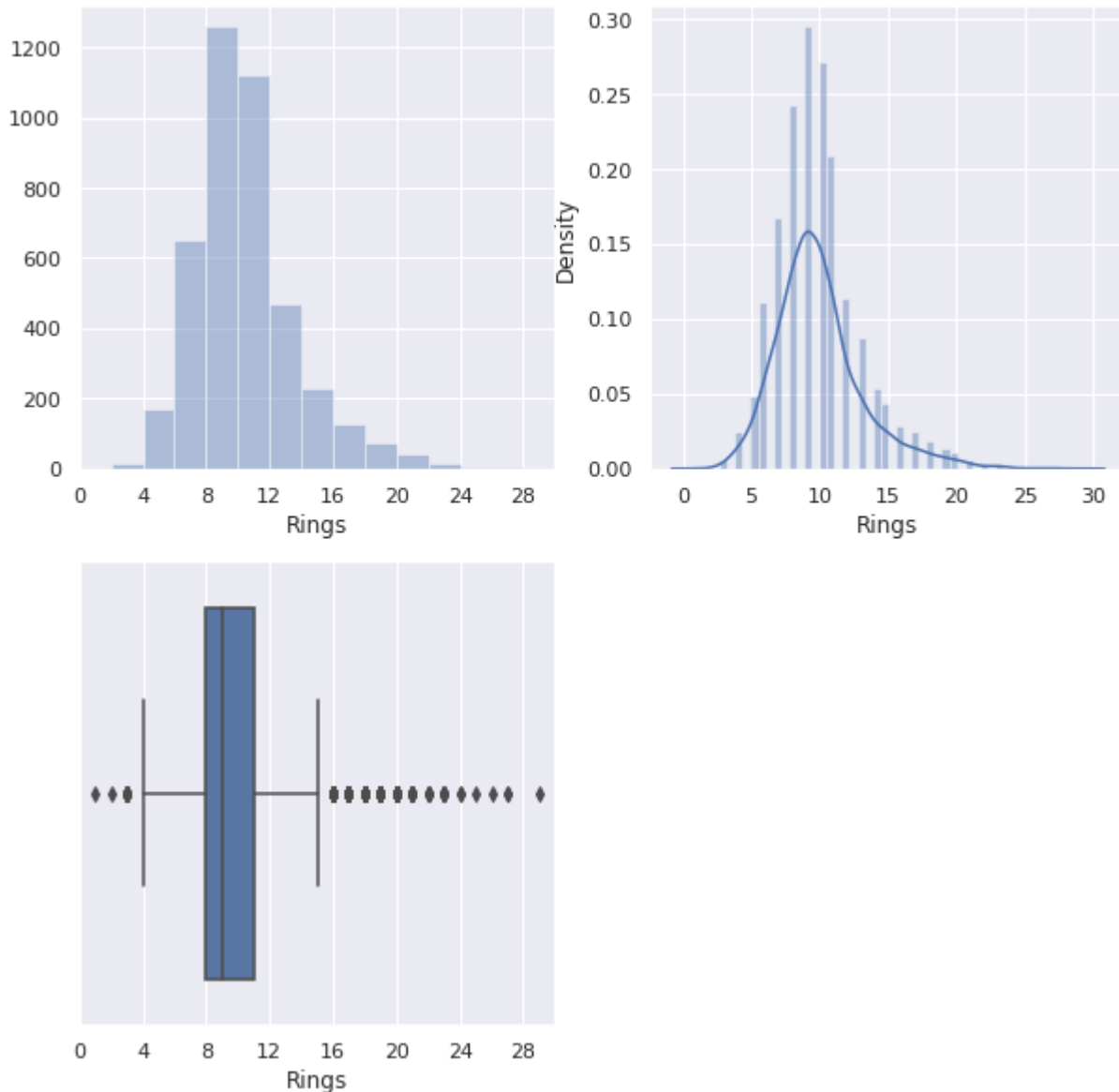
```
plt.subplot(rows, cols, i)
plt.xticks(range(0, 31, 4))
plt.xlim(0, 30)
_ = sns.distplot(abalone['Rings'], kde=False, bins=range(0, 31, 2))
```

```
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: FutureWarning: `d`
warnings.warn(msg, FutureWarning)

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



```
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), color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(abalone['Diameter'], kde=False, bins=np.arange(0.0, 0.7, 0.05), color=colors[i % 3])

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])

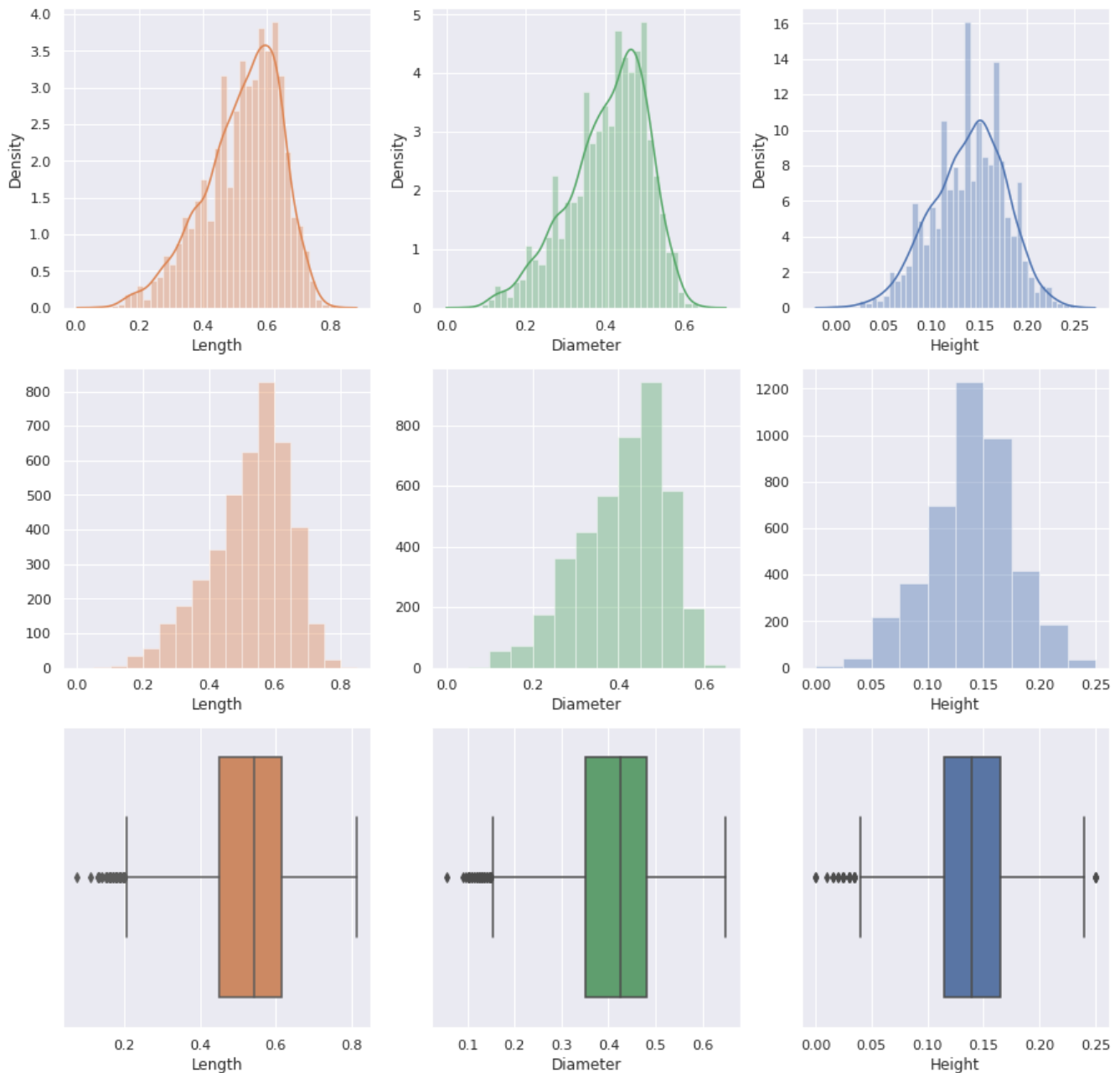
i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(abalone['Diameter'], color=colors[i % 3])

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: FutureWarning: `d
warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `d
warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `d
warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `d
warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning

```



```
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 % cols])

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

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

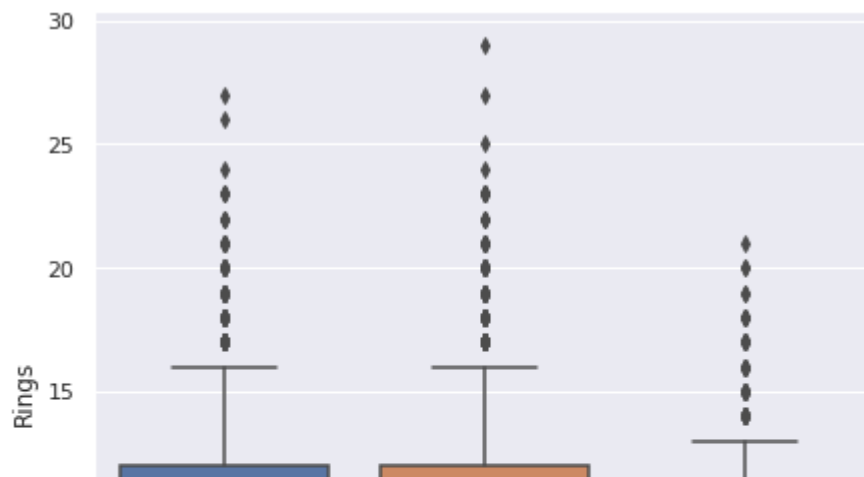
i += 1
plt.subplot(rows, cols, i)
_ = sns.distplot(abalone['Shell weight'], kde=False, bins=20, color=colors[i % cols])

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

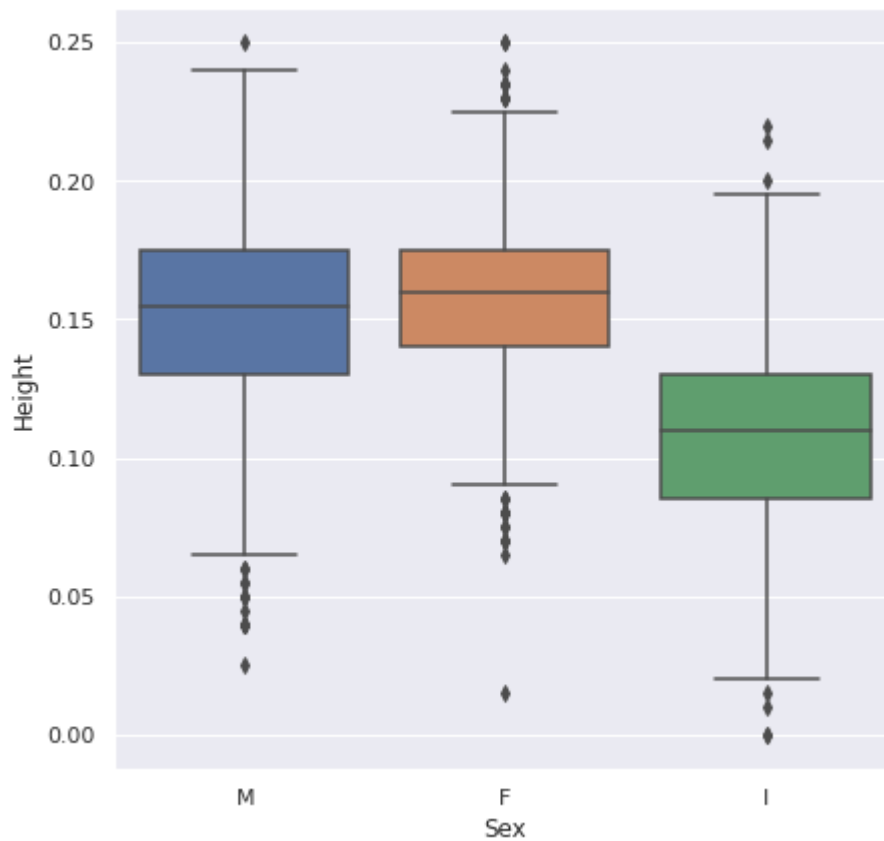
i += 1
plt.subplot(rows, cols, i)
_ = sns.boxplot(abalone['Shucked weight'], color=colors[i % cols])
```

```
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])
```

```
plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=abalone, x='Sex', y='Height')
```



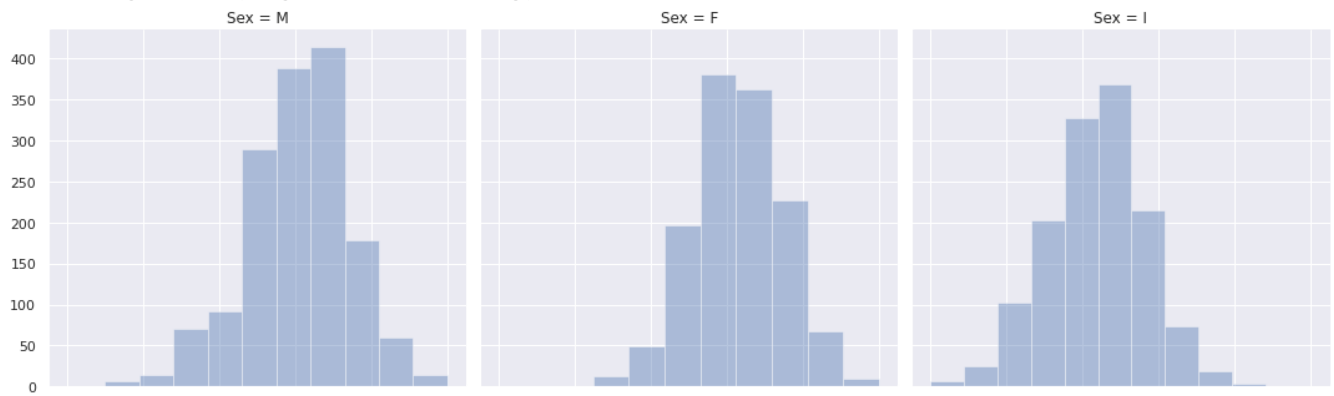
```
g = sns.FacetGrid(abalone, col='Sex', margin_titles=True, size=5)
_ = g.map(sns.distplot, 'Height', kde=False, bins=10)
```



```

/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `d
warnings.warn(msg, FutureWarning)

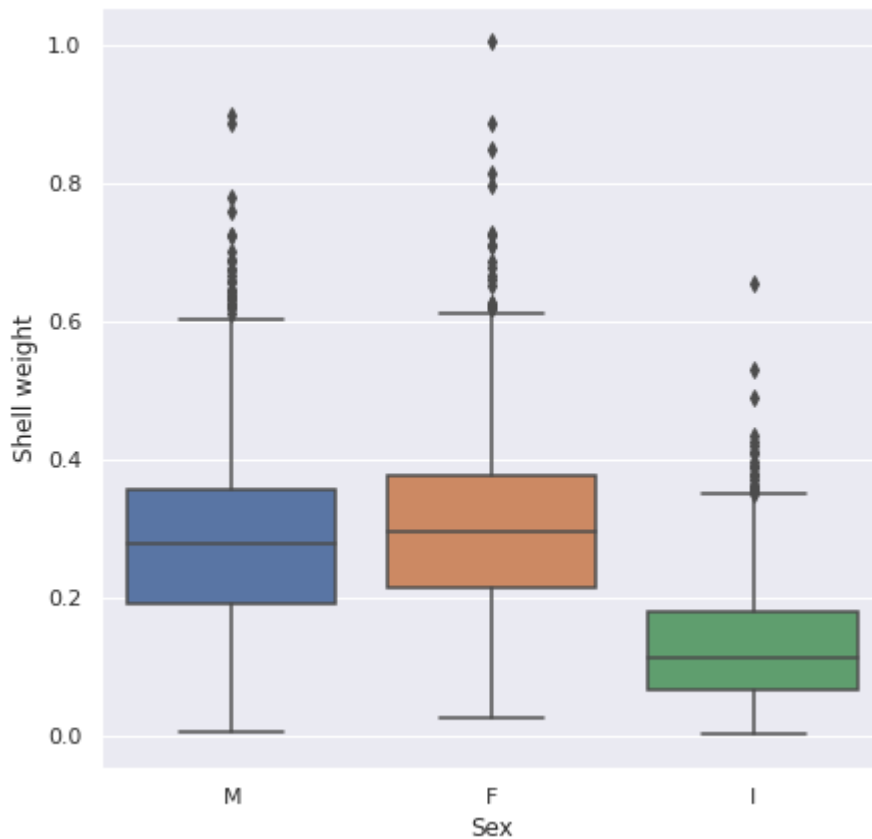
```



```

plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=abalone, x='Sex', y='Shell weight')

```

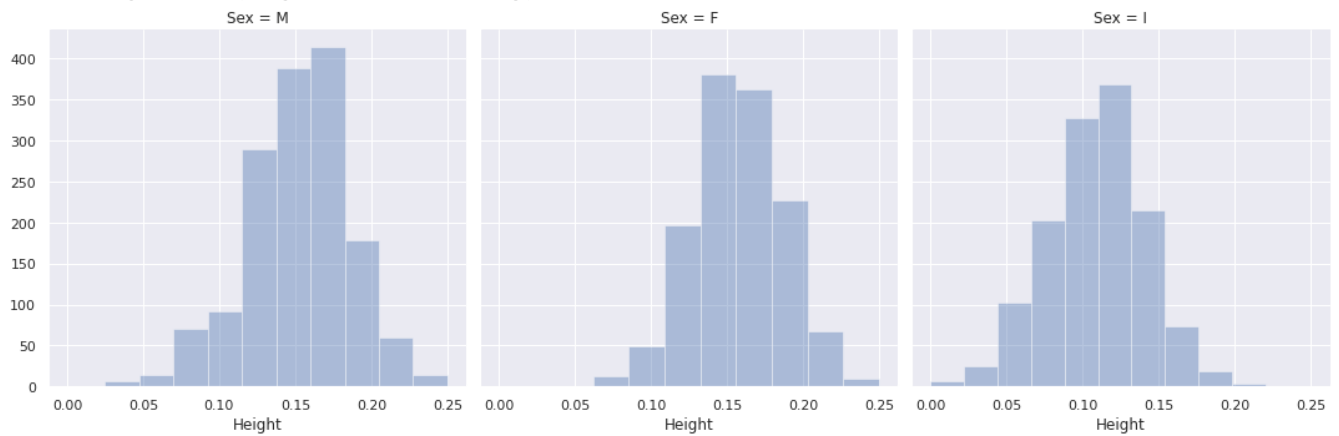


```

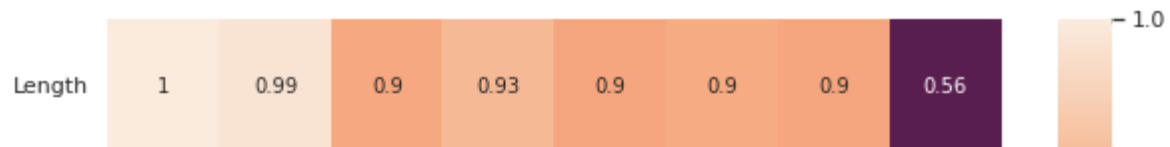
g = sns.FacetGrid(abalone, col='Sex', margin_titles=True, size=5)
_ = g.map(sns.distplot, 'Height', kde=False, bins=10)

```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`  
warnings.warn(msg, UserWarning)  
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `d`  
warnings.warn(msg, FutureWarning)
```



```
plt.figure(figsize=(10, 10))  
corr = abalone.corr()  
_ = sns.heatmap(corr, annot=True)
```

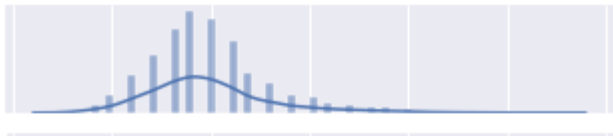


```
plt.figure(figsize=(20, 5))
```

```
_ = 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>



```
abalone.describe().T
```

	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

```
df = pd.DataFrame(abalone)
df.isnull()
```

	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

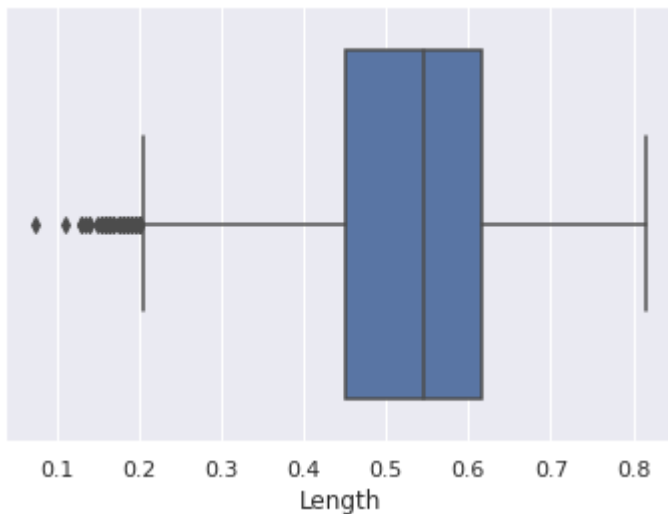
```
df.fillna(0)
```

	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

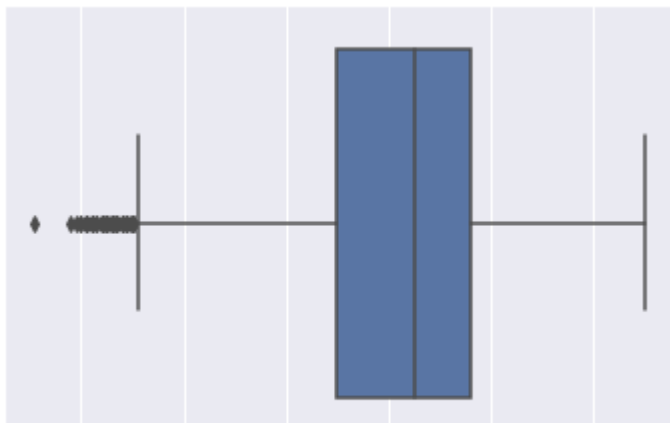
```
sns.boxplot(df['Length'],data=df)
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f5727658a90>



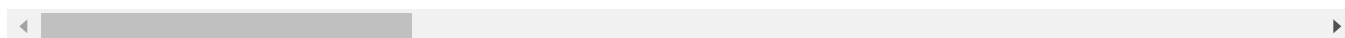
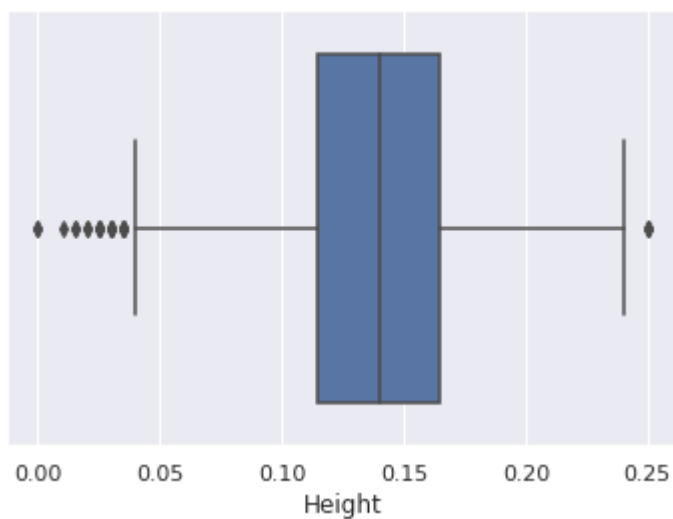
```
sns.boxplot(df['Diameter'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f572766a210>
```



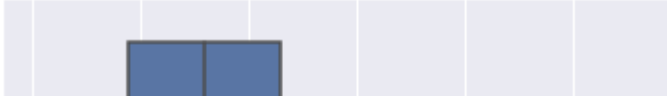
```
sns.boxplot(df['Height'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57294bc290>
```



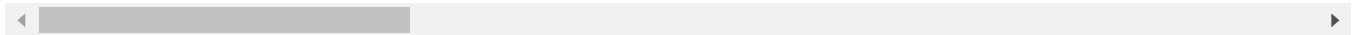
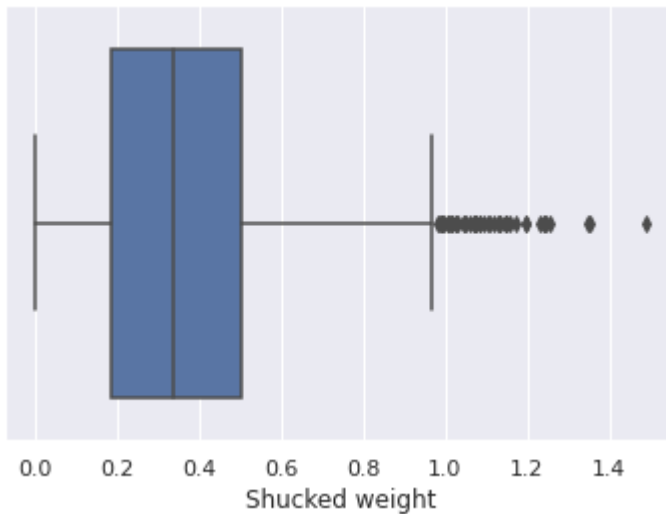
```
sns.boxplot(df['Whole weight'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57274b0650>
```



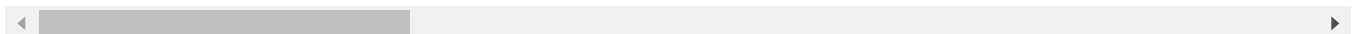
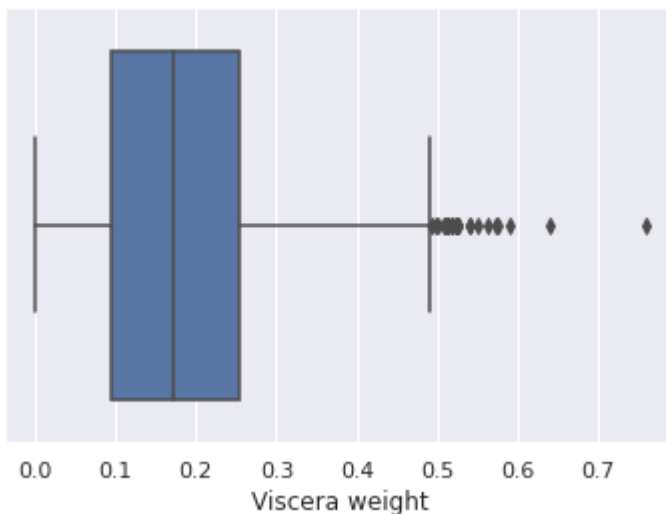
```
sns.boxplot(df['Shucked weight'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f5729378d50>
```



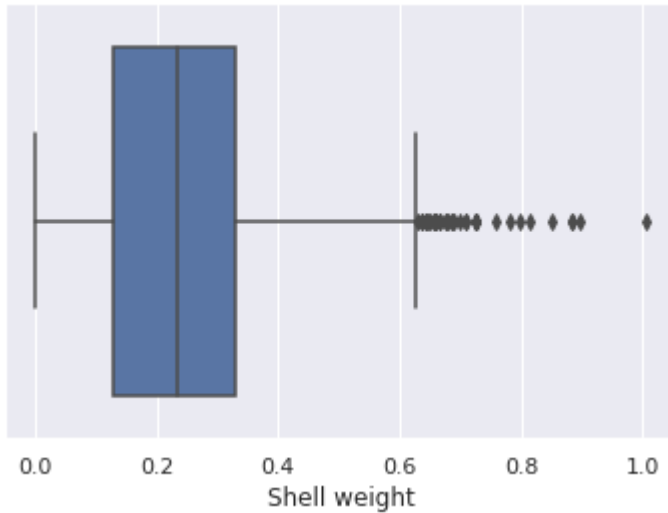
```
sns.boxplot(df['Viscera weight'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57294bc990>
```



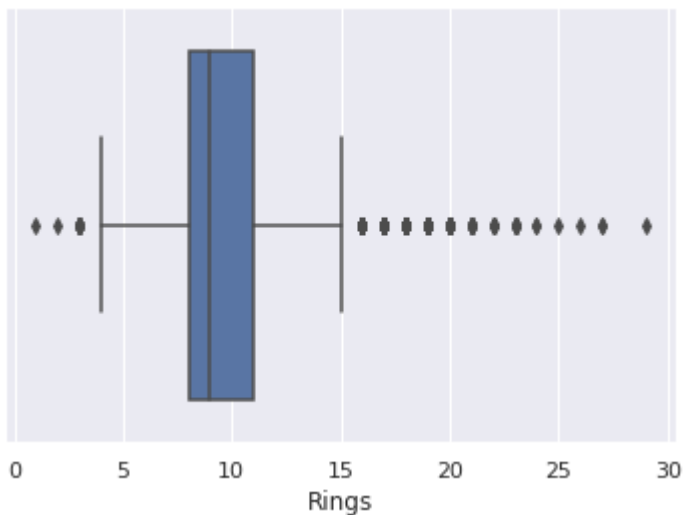
```
sns.boxplot(df['Shell weight'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f572956a610>
```



```
sns.boxplot(df['Rings'],data=df)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57273f63d0>
```



```
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
```



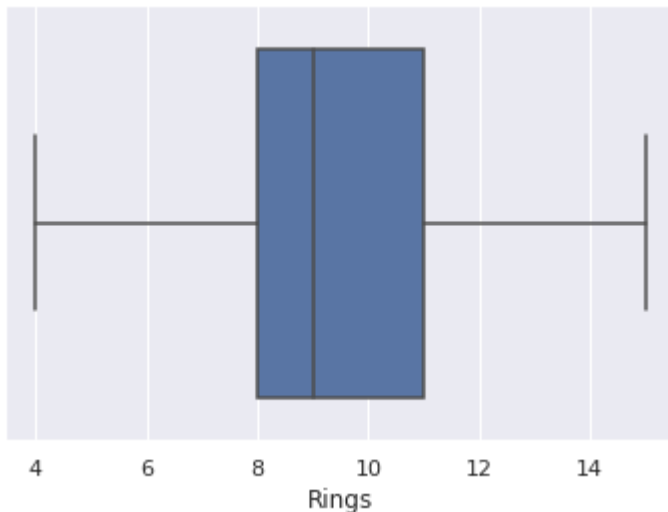
```
Shucked weight    0.31575
Viscera weight    0.15950
Shell weight      0.19875
Rings             3.00000
dtype: float64
```

```
abalone = abalone[~((abalone < (Q1 - 1.5 * IQR)) |(abalone > (Q3 + 1.5 * IQR))).any(axis=1)]
abalone.shape
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Automati
    """Entry point for launching an IPython kernel.
(3781, 9)
```

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

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
    FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57290b7210>
```



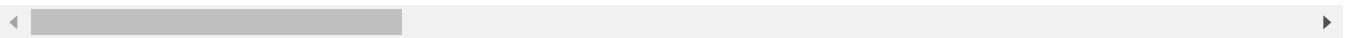
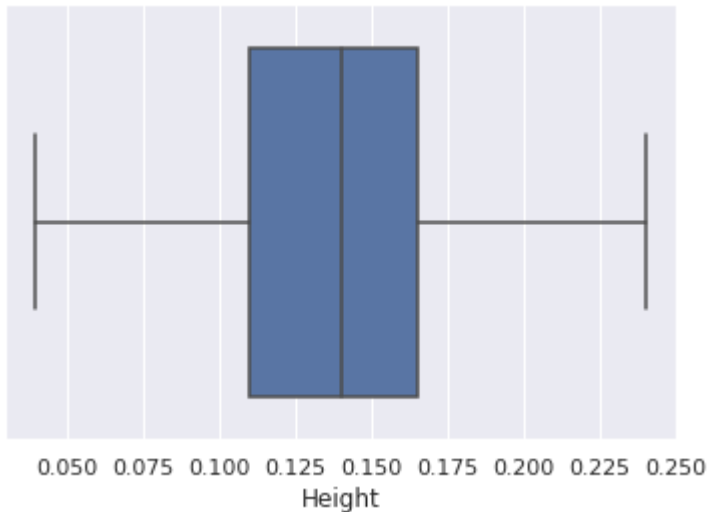
```
sns.boxplot(abalone['Length'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f5729a98710>
```



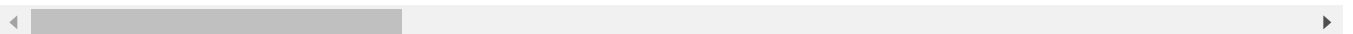
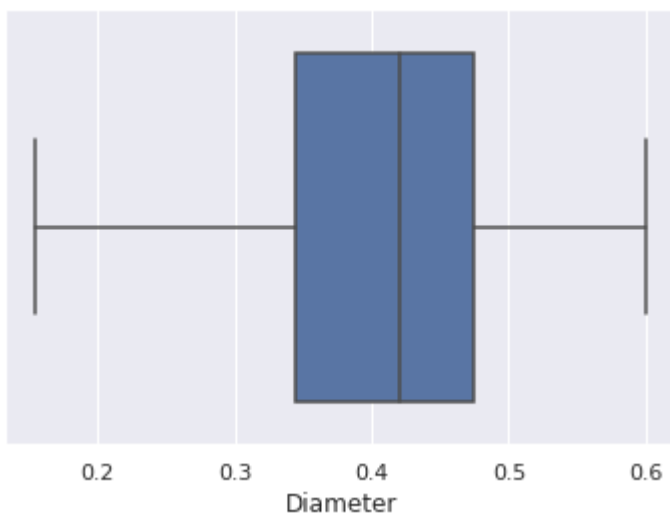
```
sns.boxplot(abalone['Height'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57295c24d0>
```



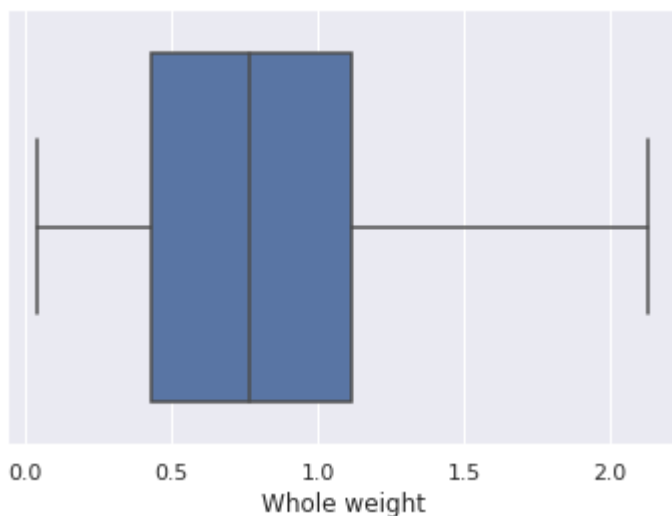
```
sns.boxplot(abalone['Diameter'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57290d8410>
```



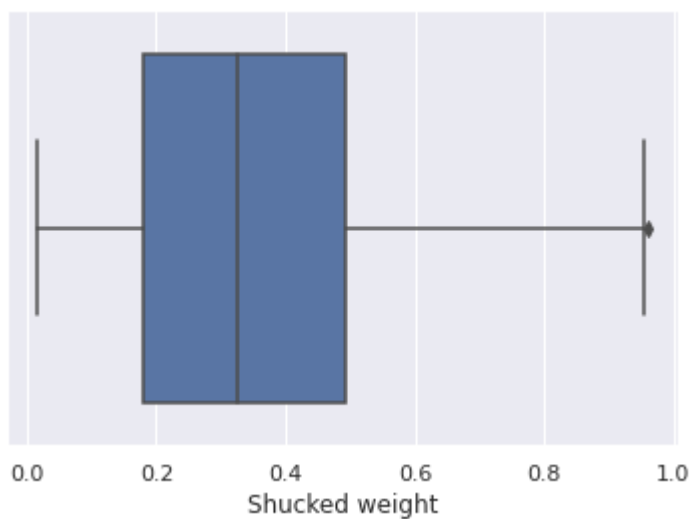
```
sns.boxplot(abalone['Whole weight'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57291d5c10>
```



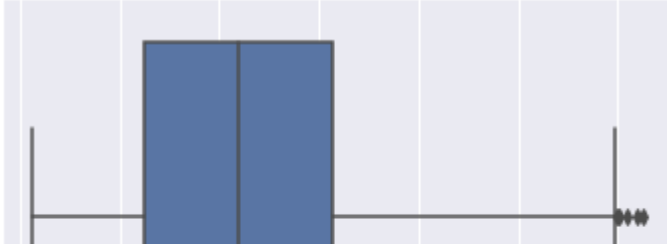
```
sns.boxplot(abalone['Shucked weight'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57291e7910>
```



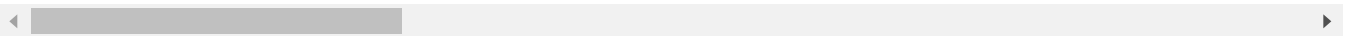
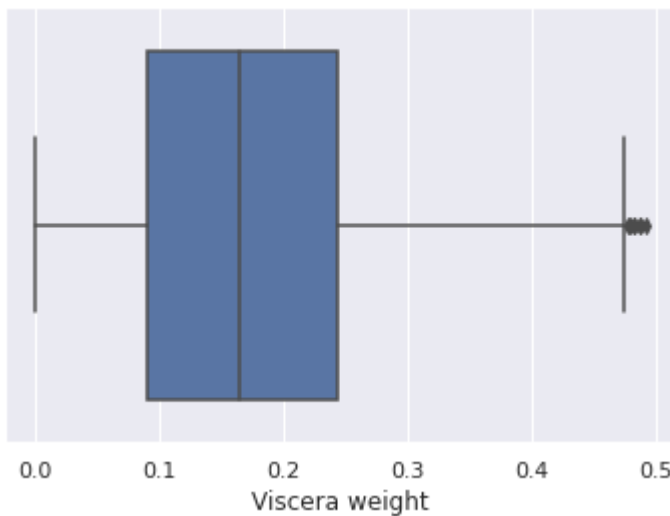
```
sns.boxplot(abalone['Shell weight'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f572767ba50>
```



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

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f57276cdfd0>
```



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

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
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/stable/user>



abalone

	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

Splitting the Data into dependent and Independent Variables

4175	0	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
-------------	---	-------	-------	-------	--------	--------	--------	--------	----

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

Scaling independent variables

```
scaler = StandardScaler()
scaler.fit(abalone)
```

```
StandardScaler()
```

Splitting training and test data

```
train_X, val_X, train_y, val_y = train_test_split(X, y, test_size = 0.2, random_state = 0)
```

```
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)
```

```
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

```

lr = LinearRegression()
lr.fit(train_X,train_y)

LinearRegression()

%%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_pred_val_lr)))
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 11.2 ms, sys: 56 µs, total: 11.2 ms
Wall time: 29.2 ms

```

SUPPORT VECTOR MACHINE

```

svm = SVR()
svm.fit(train_X,train_y)

SVR()

%%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_pred_val_svm)))
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 170 ms, sys: 0 ns, total: 170 ms

Wall time: 190 ms

DECISION TREE REGRESSOR

```
dc = DecisionTreeRegressor(random_state = 0)
dc.fit(train_X,train_y)
```

```
DecisionTreeRegressor(random_state=0)
```

```
%%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_pred_val_dc)))
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 10.4 ms, sys: 0 ns, total: 10.4 ms

Wall time: 13.4 ms

OVERVIEW OF R2 SCORES OF ALL MODELS

```
print('Logistic Regression R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_l  
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_y, y_pred_v
```

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

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 3:33 PM

