In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

fish data

In [5]:

In [6]:

Out[6]:

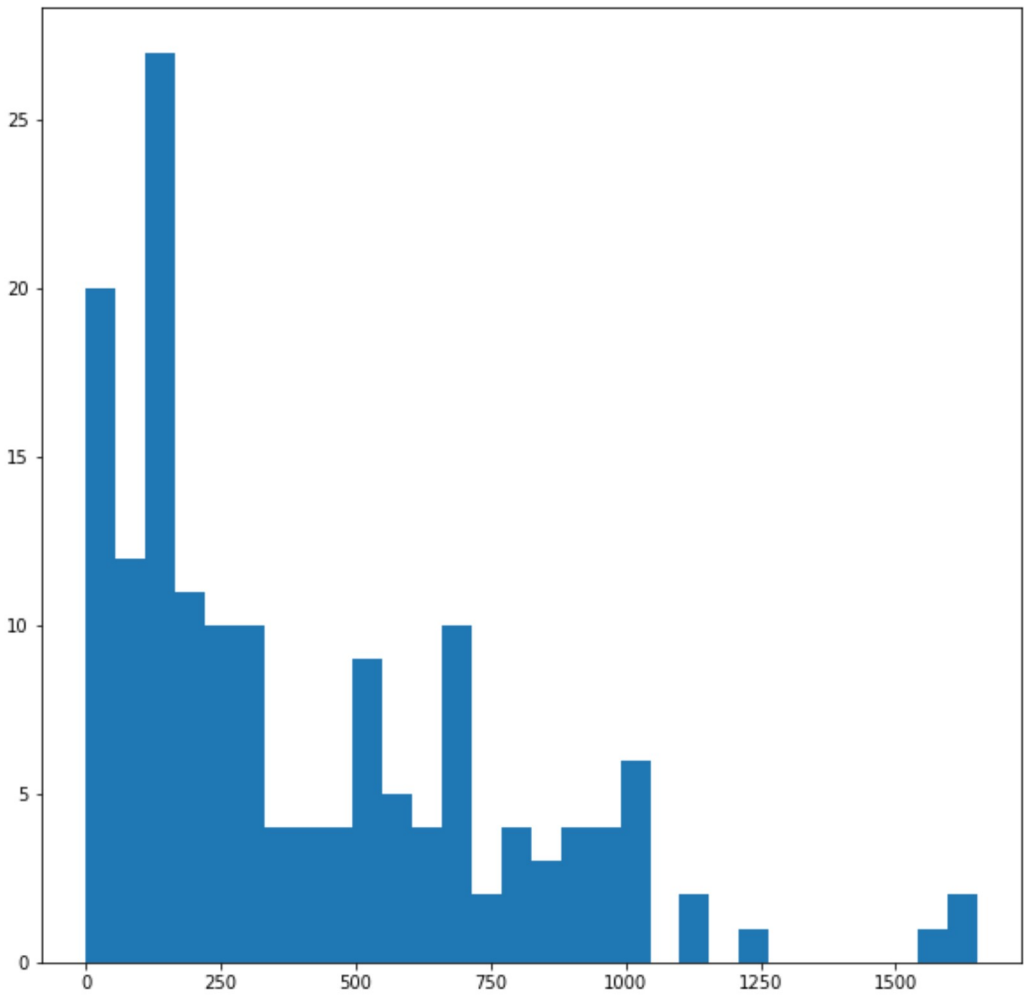
	Species	Weight	Length1	Length2	Length3	Height	Width
0	Bream	242.0	23.2	25.4	30.0	11.5200	4.0200
1	Bream	290.0	24.0	26.3	31.2	12.4800	4.3056
2	Bream	340.0	23.9	26.5	31.1	12.3778	4.6961
3	Bream	363.0	26.3	29.0	33.5	12.7300	4.4555
4	Bream	430.0	26.5	29.0	34.0	12.4440	5.1340

In [7]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159 entries, 0 to 158
Data columns (total 7 columns):
Species    159 non-null object
Weight     159 non-null float64
Length1    159 non-null float64
Length2    159 non-null float64
Length3    159 non-null float64
Height     159 non-null float64
Width      159 non-null float64
dtypes: float64(6), object(1)
memory usage: 8.8+ KB
```

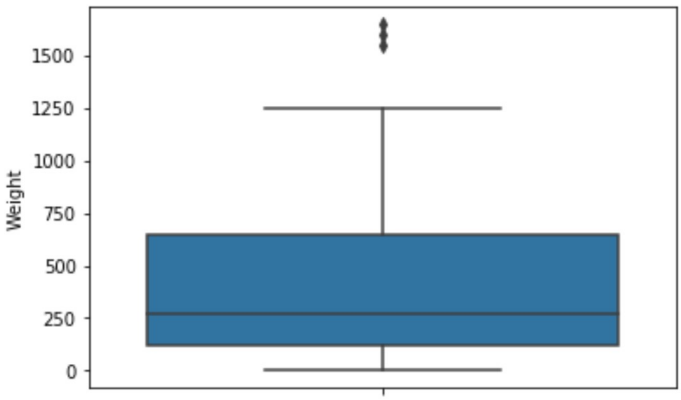
In [8]:
```
#check continuous data for target
plt.figure(figsize=(10,10))
```

Out[8]:
```
(array([20., 12., 27., 11., 10., 10.,  4.,  4.,  4.,  9.,  5.,  4., 10.,
         2.,  4.,  3.,  4.,  4.,  6.,  0.,  2.,  0.,  1.,  0.,  0.,  0.,
         0.,  0.,  1.,  2.]),
 array([   0.,   55.,  110.,  165.,  220.,  275.,  330.,  385.,  440.,
         495.,  550.,  605.,  660.,  715.,  770.,  825.,  880.,  935.,
         990., 1045., 1100., 1155., 1210., 1265., 1320., 1375., 1430.,
        1485., 1540., 1595., 1650.]),
 <a list of 30 Patch objects>)
```



In [9]:

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x2084094f5c8>



In [10]:

Out[10]:
```
count      159.000000
mean       398.326415
std        357.978317
min          0.000000
25%        120.000000
50%        273.000000
75%        650.000000
max       1650.000000
Name: Weight, dtype: float64
```

In [32]:

Out[32]:

	Species	Weight	Length1	Length2	Length3	Height	Width
142	Pike	1600.0	56.0	60.0	64.0	9.600	6.144
143	Pike	1550.0	56.0	60.0	64.0	9.600	6.144
144	Pike	1650.0	59.0	63.4	68.0	10.812	7.480

In [31]:

Out[31]:

	Species	Weight	Length1	Length2	Length3	Height	Width
0	Bream	242.0	23.2	25.4	30.0	11.5200	4.0200
1	Bream	290.0	24.0	26.3	31.2	12.4800	4.3056
2	Bream	340.0	23.9	26.5	31.1	12.3778	4.6961
3	Bream	363.0	26.3	29.0	33.5	12.7300	4.4555
4	Bream	430.0	26.5	29.0	34.0	12.4440	5.1340
...	...	...	...	...	...	...	...
154	Smelt	12.2	11.5	12.2	13.4	2.0904	1.3936
155	Smelt	13.4	11.7	12.4	13.5	2.4300	1.2690
156	Smelt	12.2	12.1	13.0	13.8	2.2770	1.2558
157	Smelt	19.7	13.2	14.3	15.2	2.8728	2.0672
158	Smelt	19.9	13.8	15.0	16.2	2.9322	1.8792

156 rows × 7 columns

In [15]:

```
Out[15]: Perch        56
         Bream        35
         Roach        20
         Pike         17
         Smelt        14
         Parkki       11
         Whitefish     6
         Name: Species, dtype: int64
```

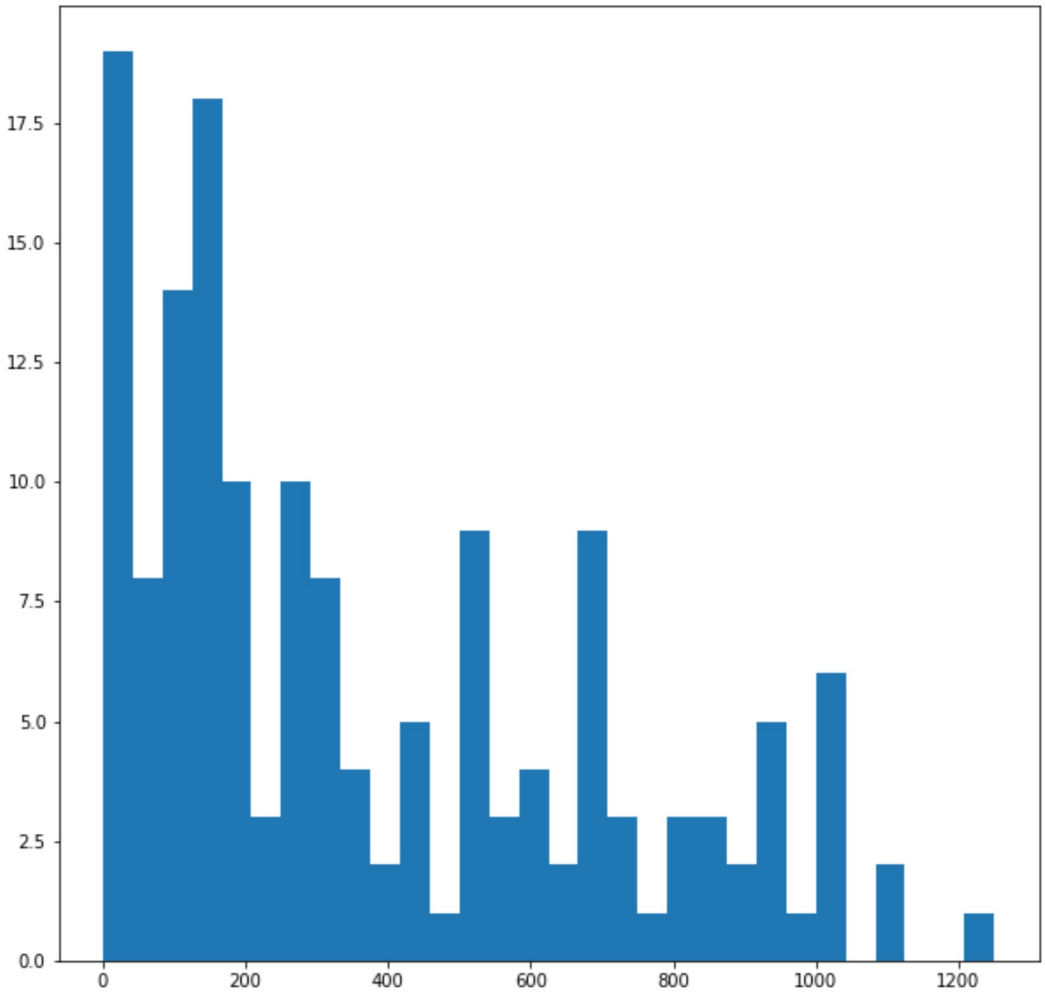
In [33]:

In [35]:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 156 entries, 0 to 158
Data columns (total 7 columns):
Species    156 non-null object
Weight     156 non-null float64
Length1    156 non-null float64
Length2    156 non-null float64
Length3    156 non-null float64
Height     156 non-null float64
Width      156 non-null float64
dtypes: float64(6), object(1)
memory usage: 9.8+ KB
```

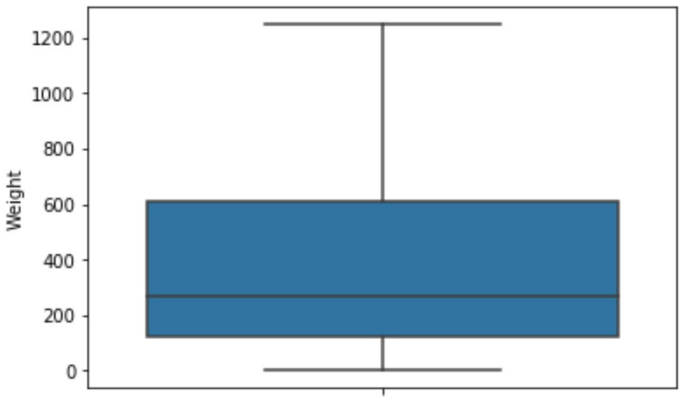
In [40]:
```
#check continuous data for target
plt.figure(figsize=(10,10))
```

Out[40]:
```
(array([19.,  8., 14., 18., 10.,  3., 10.,  8.,  4.,  2.,  5.,  1.,  9.,
         3.,  4.,  2.,  9.,  3.,  1.,  3.,  3.,  2.,  5.,  1.,  6.,  0.,
         2.,  0.,  0.,  1.]),
 array([   0.        ,   41.66666667,   83.33333333,  125.        ,
        166.66666667,  208.33333333,  250.        ,  291.66666667,
        333.33333333,  375.        ,  416.66666667,  458.33333333,
        500.        ,  541.66666667,  583.33333333,  625.        ,
        666.66666667,  708.33333333,  750.        ,  791.66666667,
        833.33333333,  875.        ,  916.66666667,  958.33333333,
       1000.        , 1041.66666667, 1083.33333333, 1125.        ,
       1166.66666667, 1208.33333333, 1250.        ]),
 <a list of 30 Patch objects>)
```



In [36]:

Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x2083f8c63c8>



In [38]:

Out[38]:
```
Perch        56
Bream        35
Roach        20
Smelt        14
Pike         14
Parkki       11
Whitefish     6
Name: Species, dtype: int64
```

In [39]:

In [41]:

In [42]:

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCop
yWarning:
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: http://pandas.pydata.org/pandas-docs/stable/
user_guide/indexing.html#returning-a-view-versus-a-copy (http://pandas.pydata.org/
pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [43]:

Out[43]:
```
2    56
0    35
4    20
5    14
3    14
1    11
6     6
Name: Species, dtype: int64
```

In [44]:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 156 entries, 0 to 158
Data columns (total 7 columns):
Species    156 non-null int32
Weight     156 non-null float64
Length1    156 non-null float64
Length2    156 non-null float64
Length3    156 non-null float64
```

In [45]:

Out[45]:

	Species	Weight	Length1	Length2	Length3	Height	Width
0	0	242.0	23.2	25.4	30.0	11.5200	4.0200
1	0	290.0	24.0	26.3	31.2	12.4800	4.3056
2	0	340.0	23.9	26.5	31.1	12.3778	4.6961
3	0	363.0	26.3	29.0	33.5	12.7300	4.4555
4	0	430.0	26.5	29.0	34.0	12.4440	5.1340

In [46]:

Out[46]:

	Species	Weight	Length1	Length2	Length3	Height	Width
count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000
mean	2.250000	375.217308	25.655769	27.786538	30.571154	8.951128	4.375719
std	1.717556	319.546978	9.119630	9.792651	10.695359	4.324325	1.672188
min	0.000000	0.000000	7.500000	8.400000	8.800000	1.728400	1.047600
25%	1.000000	120.000000	19.000000	21.000000	23.025000	5.931675	3.369600
50%	2.000000	271.000000	25.000000	26.750000	29.250000	7.647800	4.243300
75%	4.000000	612.500000	32.125000	35.000000	39.425000	12.378550	5.424375
max	6.000000	1250.000000	52.000000	56.000000	59.700000	18.957000	8.142000

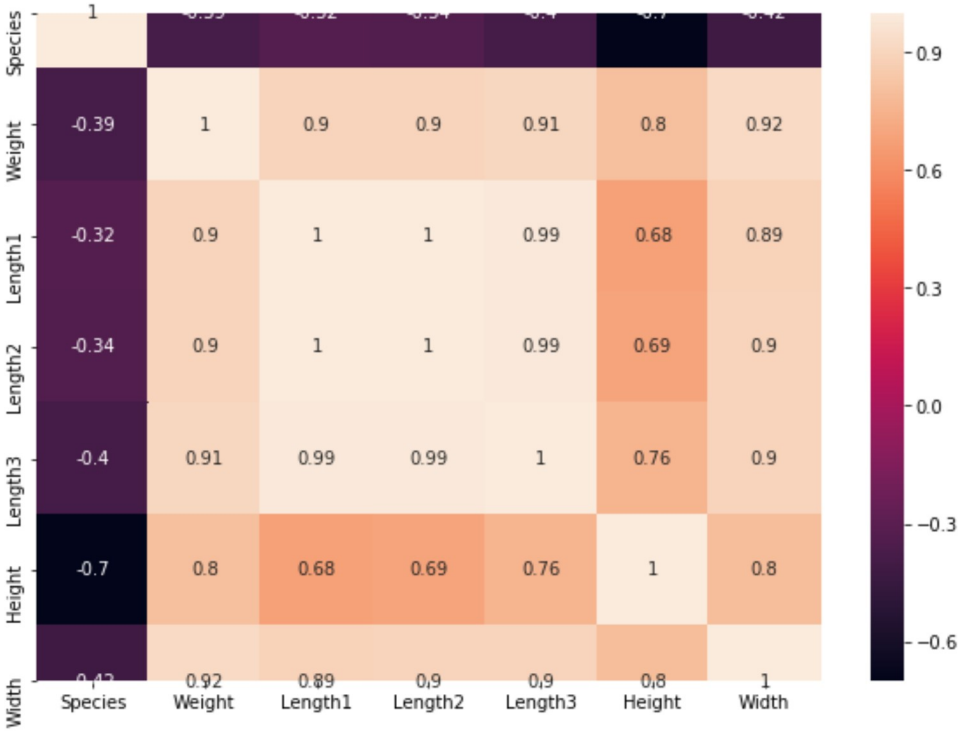
In [47]:

Out[47]:

	Species	Weight	Length1	Length2	Length3	Height	Width
Species	1.000000	-0.386398	-0.324187	-0.342145	-0.395344	-0.701987	-0.416384
Weight	-0.386398	1.000000	0.895743	0.899759	0.907292	0.801871	0.923432
Length1	-0.324187	0.895743	1.000000	0.999418	0.990612	0.676326	0.889166
Length2	-0.342145	0.899759	0.999418	1.000000	0.993049	0.691954	0.895584
Length3	-0.395344	0.907292	0.990612	0.993049	1.000000	0.756019	0.897053
Height	-0.701987	0.801871	0.676326	0.691954	0.756019	1.000000	0.800542
Width	-0.416384	0.923432	0.889166	0.895584	0.897053	0.800542	1.000000

In [48]:
```
plt.figure(figsize=(10,7))
```

Out[48]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x208424ba8c8>
```



In [51]:

Out[51]:
```
Index(['Species', 'Weight', 'Length1', 'Length2', 'Length3', 'Height',
       'Width'],
      dtype='object')
```

In [76]:
```
x=df1[['Species','Length1','Length2','Length3','Height','Width']].values
```

In [77]:

Out[77]:
```
array([[ 0.  , 23.2 , 25.4 , 30.  , 11.52  ,  4.02  ],
       [ 0.  , 24.  , 26.3 , 31.2 , 12.48  ,  4.3056],
       [ 0.  , 23.9 , 26.5 , 31.1 , 12.3778,  4.6961],
       [ 0.  , 26.3 , 29.  , 33.5 , 12.73  ,  4.4555],
       [ 0.  , 26.5 , 29.  , 34.  , 12.444 ,  5.134 ],
       [ 0.  , 26.8 , 29.7 , 34.7 , 13.6024,  4.9274],
       [ 0.  , 26.8 , 29.7 , 34.5 , 14.1795,  5.2785],
       [ 0.  , 27.6 , 30.  , 35.  , 12.67  ,  4.69  ],
       [ 0.  , 27.6 , 30.  , 35.1 , 14.0049,  4.8438],
       [ 0.  , 28.5 , 30.7 , 36.2 , 14.2266,  4.9594],
       [ 0.  , 28.4 , 31.  , 36.2 , 14.2628,  5.1042],
       [ 0.  , 28.7 , 31.  , 36.2 , 14.3714,  4.8146],
       [ 0.  , 29.1 , 31.5 , 36.4 , 13.7592,  4.368 ],
       [ 0.  , 29.5 , 32.  , 37.3 , 13.9129,  5.0728],
       [ 0.  , 29.4 , 32.  , 37.2 , 14.9544,  5.1708],
       [ 0.  , 29.4 , 32.  , 37.2 , 15.438 ,  5.58  ],
       [ 0.  , 30.4 , 33.  , 38.3 , 14.8604,  5.2854],
       [ 0.  , 30.4 , 33.  , 38.5 , 14.938 ,  5.1975],
       [ 0.  , 30.9 , 33.5 , 38.6 , 15.633 ,  5.1338],
```

In [78]:

Out[78]:

```
array([ 242. ,  290. ,  340. ,  363. ,  430. ,  450. ,  500. ,  390. ,
        450. ,  500. ,  475. ,  500. ,  500. ,  340. ,  600. ,  600. ,
        700. ,  700. ,  610. ,  650. ,  575. ,  685. ,  620. ,  680. ,
        700. ,  725. ,  720. ,  714. ,  850. , 1000. ,  920. ,  955. ,
        925. ,  975. ,  950. ,   40. ,   69. ,   78. ,   87. ,  120. ,
          0. ,  110. ,  120. ,  150. ,  145. ,  160. ,  140. ,  160. ,
        169. ,  161. ,  200. ,  180. ,  290. ,  272. ,  390. ,  270. ,
        270. ,  306. ,  540. ,  800. , 1000. ,   55. ,   60. ,   90. ,
        120. ,  150. ,  140. ,  170. ,  145. ,  200. ,  273. ,  300. ,
          5.9,   32. ,   40. ,   51.5,   70. ,  100. ,   78. ,   80. ,
         85. ,   85. ,  110. ,  115. ,  125. ,  130. ,  120. ,  120. ,
        130. ,  135. ,  110. ,  130. ,  150. ,  145. ,  150. ,  170. ,
        225. ,  145. ,  188. ,  180. ,  197. ,  218. ,  300. ,  260. ,
        265. ,  250. ,  250. ,  300. ,  320. ,  514. ,  556. ,  840. ,
        685. ,  700. ,  700. ,  690. ,  900. ,  650. ,  820. ,  850. ,
        900. , 1015. ,  820. , 1100. , 1000. , 1100. , 1000. , 1000. ,
        200. ,  300. ,  300. ,  300. ,  430. ,  345. ,  456. ,  510. ,
        540. ,  500. ,  567. ,  770. ,  950. , 1250. ,    6.7,    7.5,
          7. ,    9.7,    9.8,    8.7,   10. ,    9.9,    9.8,   12.2,
```

In [79]:

In [80]:

In [81]:

In [82]:

In [83]:

Out[83]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

In [84]:

Out[84]: 0.915053474158038

In [85]:

Out[85]: 0.9063763829723153

In [ ]: