Import Library

```
import pandas as pd
```

```
import numpy as np
```

Import CSV as DataFrame

```
df = pd.read_csv(r'https://github.com/YBI-Foundation/Dataset/raw/main/Fish.csv')
```

```
df.head()
```

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

Get the information About the data set

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159 entries, 0 to 158
Data columns (total 8 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Category  159 non-null    int64  
 1   Species   159 non-null    object 
 2   Weight    159 non-null    float64
 3   Height    159 non-null    float64
 4   Width     159 non-null    float64
 5   Length1   159 non-null    float64
 6   Length2   159 non-null    float64
 7   Length3   159 non-null    float64
dtypes: float64(6), int64(1), object(1)
memory usage: 10.1+ KB
```

Get the Summary Statistics

```
df.describe()
```

	Category	Weight	Height	Width	Length1	Length2	Leng
count	159.000000	159.000000	159.000000	159.000000	159.000000	159.000000	159.0000
mean	3.264151	398.326415	8.970994	4.417486	26.247170	28.415723	31.227
std	1.704249	357.978317	4.286208	1.685804	9.996441	10.716328	11.610
min	1.000000	0.000000	1.728400	1.047600	7.500000	8.400000	8.800
25%	2.000000	120.000000	5.944800	3.385650	19.050000	21.000000	23.150
50%	3.000000	273.000000	7.786000	4.248500	25.200000	27.300000	29.400
75%	4.500000	650.000000	12.365900	5.584500	32.700000	35.500000	39.650
max	7.000000	1650.000000	18.957000	8.142000	59.000000	63.400000	68.000

```
df.shape
```

```
(159, 8)
```

Get column names

```
df.columns
```

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

Define y and X

```
y = df ['Weight']
```

```
y.shape
```

```
(159,)
```

```
y
```

```
0      242.0
1      290.0
2      340.0
3      363.0
4      430.0
```

```
              ...
       154     12.2
       155     13.4
       156     12.2
       157     19.7
       158     19.9
       Name: Weight, Length: 159, dtype: float64
```

```
X = df [[ 'Width', 'Height', 'Length1',
       'Length2', 'Length3']]
```

```
X.shape
```

```
     (159, 5)
```

```
X
```

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

159 rows × 5 columns

train_test_split

```
from sklearn.model_selection import train_test_split
```

```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size = 0.3,random_state = 2529)
```

```
X_train.shape,X_test.shape,y_train.shape,y_test.shape
```

```
     ((111, 5), (48, 5), (111,), (48,))
```

```
from sklearn.linear_model import LinearRegression
```

```
model = LinearRegression()
```

```
model.fit(X_train,y_train)
```

```
     LinearRegression()
```

```
y_pred = model.predict(X_test)
```

```
y_pred.shape
```

```
     (48,)
```

```
y_pred
```

```
     array([ 485.76826299,  502.24720857,   94.72381964,  876.5711712 ,
             184.0789176 ,  219.30130488,  322.32532246,  376.22325991,
             372.35730485, -182.67537078, -160.60486837,  454.33586185,
             159.59755829,  843.48525226,  587.21680573,  299.53521445,
             597.72950823,  197.14605397,  639.89046741,   91.20067876,
             150.95424753, -103.08320574,  627.19712753,  795.69176861,
             814.68732975, -204.1496511 ,  329.98746856,  715.89288013,
             359.75634357,  792.3243925 ,  532.7036706 ,  552.00832342,
             433.48472727,  687.61750267, -204.76362537,  932.53668294,
             810.74234216,  -80.06217174,  284.36287887,  907.08036021,
             642.5828335 ,  959.33848223,  675.28792291,  718.86305458,
             623.89849226,  376.48346981,  530.83828119,  -86.2357066 ])
```

Get Model Evaluation

```
from sklearn.metrics import mean_squared_error,mean_absolute_error,mean_absolute_percentag
```

```
mean_squared_error(y_test,y_pred)
```

```
     16397.344524411415
```

```
mean_absolute_error(y_test,y_pred)
```

```
     103.02952922678553
```

```
mean_absolute_percentage_error(y_test,y_pred)
```

```
     2.508285347160026
```

```
r2_score(y_test,y_pred)
```

```
0.8349141424416874
```

Get Future predictions

```
df_new = df.sample(1)
```

```
df_new
```

	Category	Species	Weight	Height	Width	Length1	Length2	Length3
59	7	Whitefish	800.0	11.7612	6.5736	33.7	36.4	39.6

```
X_new = df_new [['Height',  'Width','Length1'   ,'Length2'  ,'Length3']]
```

```
X_new.shape
```

```
(1, 5)
```

```
y_pred_new = model.predict(X_new)
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:493: FutureWarning: The featur
Feature names must be in the same order as they were in fit.

  warnings.warn(message, FutureWarning)
```

```
y_pred_new
```

```
array([548.60307604])
```

✓ 0s completed at 5:21 PM ● ×