```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression



df = pd.read_csv('FishData.csv')
print(df.head())
```

```
  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
```

```
print(df.shape)
```

```
(159, 7)
```

```
print(df.isnull().sum())
```

```
Species    0
Weight     0
Length1    0
Length2    0
Length3    0
Height     0
Width      0
dtype: int64
```

```
print(df.describe())
```

```
            Weight     Length1     Length2     Length3      Height       Width
count   159.000000  159.000000  159.000000  159.000000  159.000000  159.000000
mean    398.326415   26.247170   28.415723   31.227044    8.970994    4.417486
std     357.978317    9.996441   10.716328   11.610246    4.286208    1.685804
min       0.000000    7.500000    8.400000    8.800000    1.728400    1.047600
25%     120.000000   19.050000   21.000000   23.150000    5.944800    3.385650
50%     273.000000   25.200000   27.300000   29.400000    7.786000    4.248500
75%     650.000000   32.700000   35.500000   39.650000   12.365900    5.584500
max    1650.000000   59.000000   63.400000   68.000000   18.957000    8.142000
```

```
df['Species'].value_counts()
```

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

```
df1 = df.drop(['Length1','Length2','Length3'], axis=1)


df2= df1.drop('Species', axis=1)


print(df2.head())
```

```
   Weight   Height   Width
0   242.0  11.5200  4.0200
1   290.0  12.4800  4.3056
2   340.0  12.3778  4.6961
3   363.0  12.7300  4.4555
4   430.0  12.4440  5.1340
```

```
X ,Y = df2.drop('Weight', axis=1), df2['Weight'].values


X_train, X_test, Y_train, Y_test = train_test_split(X, Y, random_state=42, train_size=0.7)


reg=LinearRegression()
reg.fit(X_train, Y_train)
```

```
LinearRegression()
```

```
print(reg.coef_)
```

```
[  0.5069445  189.80590285]
```

```
r2_score = reg.score(X_test, Y_test)
print(r2_score)
```

```
0.8420412366115981
```

```
reg.predict([[11, 4]])
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/base.py:450: UserWarning: X does not have valid feature names, but LinearRegression was f
  warnings.warn(
array([309.3956591])
```

```
Y_prediction=reg.predict(X_test)
```

```
from sklearn.metrics import r2_score
r2_score(Y_test,Y_prediction)
```

```
0.8420412366115981
```