Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

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

info on dataset

```
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
```

```
df.describe()
```

```
         Category       Weight      Height       Width     Length1  \
count  159.000000   159.000000  159.000000  159.000000  159.000000
mean     3.264151   398.326415    8.970994    4.417486   26.247170
std      1.704249   357.978317    4.286208    1.685804    9.996441
min      1.000000     0.000000    1.728400    1.047600    7.500000
25%      2.000000   120.000000    5.944800    3.385650   19.050000
50%      3.000000   273.000000    7.786000    4.248500   25.200000
75%      4.500000   650.000000   12.365900    5.584500   32.700000
max      7.000000  1650.000000   18.957000    8.142000   59.000000

          Length2     Length3
count  159.000000  159.000000
mean    28.415723   31.227044
std     10.716328   11.610246
min      8.400000    8.800000
25%     21.000000   23.150000
50%     27.300000   29.400000
```

```
75%     35.500000   39.650000
max     63.400000   68.000000
```

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

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

```
df.columns
```

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

defining X and y

```
X = df.drop(['Category', 'Species', 'Weight'],axis=1)
y = df['Weight']
```

```
X.shape, y.shape
```

```
((159, 5), (159,))
```

Splitting the dataset into the Training set and Test set

```
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 = 2)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

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

LINEAR REGRESSION

```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(X_train,y_train)
```

```
LinearRegression()
```

ModelEvaluation

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

```
from sklearn.metrics import mean_absolute_percentage_error
```

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

```
1.3055026838184
```