IMPORTING LIBRARIES:

In [22]:
```
import pandas as pd
```

In [23]:
```
import numpy as np
```

In [24]:
```
import matplotlib.pyplot as mpl
```

In [25]:
```
import seaborn as sns
```

In [26]:
```
from sklearn.linear_model import LinearRegression
```

IMPORTING DATASET:

In [27]:
```
database = pd.read_csv('iris.csv')
```

In [28]:
```
database
```

Out[28]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

DESCRIBING THE DATASET:

In [29]:
```
database.head(15)
```

Out[29]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa

In [30]:
```
database.tail(15)
```

Out[30]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
135	136	7.7	3.0	6.1	2.3	Iris-virginica
136	137	6.3	3.4	5.6	2.4	Iris-virginica
137	138	6.4	3.1	5.5	1.8	Iris-virginica
138	139	6.0	3.0	4.8	1.8	Iris-virginica
139	140	6.9	3.1	5.4	2.1	Iris-virginica
140	141	6.7	3.1	5.6	2.4	Iris-virginica
141	142	6.9	3.1	5.1	2.3	Iris-virginica
142	143	5.8	2.7	5.1	1.9	Iris-virginica
143	144	6.8	3.2	5.9	2.3	Iris-virginica
144	145	6.7	3.3	5.7	2.5	Iris-virginica
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

In [31]:
```
database.shape
```

Out[31]:
```
(150, 6)
```

In [32]:
```
print(database.keys())
```

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

In [33]:
```
database.dtypes
```

Out[33]:
```
Id                int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

In [34]:
```
database.isnull()
```

Out[34]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
145	False	False	False	False	False	False
146	False	False	False	False	False	False
147	False	False	False	False	False	False
148	False	False	False	False	False	False
149	False	False	False	False	False	False

150 rows × 6 columns

In [35]:
```
database.isnull().sum()
```

Out[35]:
```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

In [36]:
```
data = database.groupby('Species')
```

In [37]:
```
data.head()
```

Out[37]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
50	51	7.0	3.2	4.7	1.4	Iris-versicolor
51	52	6.4	3.2	4.5	1.5	Iris-versicolor
52	53	6.9	3.1	4.9	1.5	Iris-versicolor
53	54	5.5	2.3	4.0	1.3	Iris-versicolor
54	55	6.5	2.8	4.6	1.5	Iris-versicolor
100	101	6.3	3.3	6.0	2.5	Iris-virginica
101	102	5.8	2.7	5.1	1.9	Iris-virginica
102	103	7.1	3.0	5.9	2.1	Iris-virginica
103	104	6.3	2.9	5.6	1.8	Iris-virginica
104	105	6.5	3.0	5.8	2.2	Iris-virginica

In [38]:
```
database.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             150 non-null    int64  
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object 
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

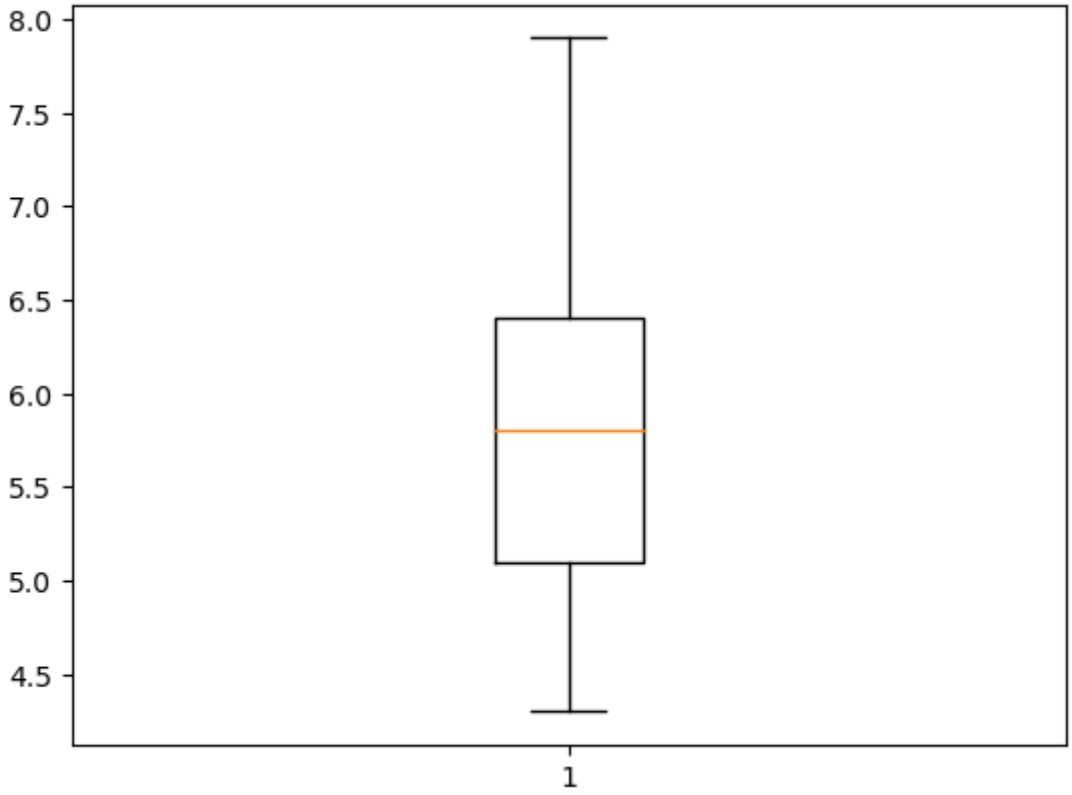
In [39]:
```
database['Species'].unique()
```

Out[39]:
```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

VISUALIZING THE DATASET:

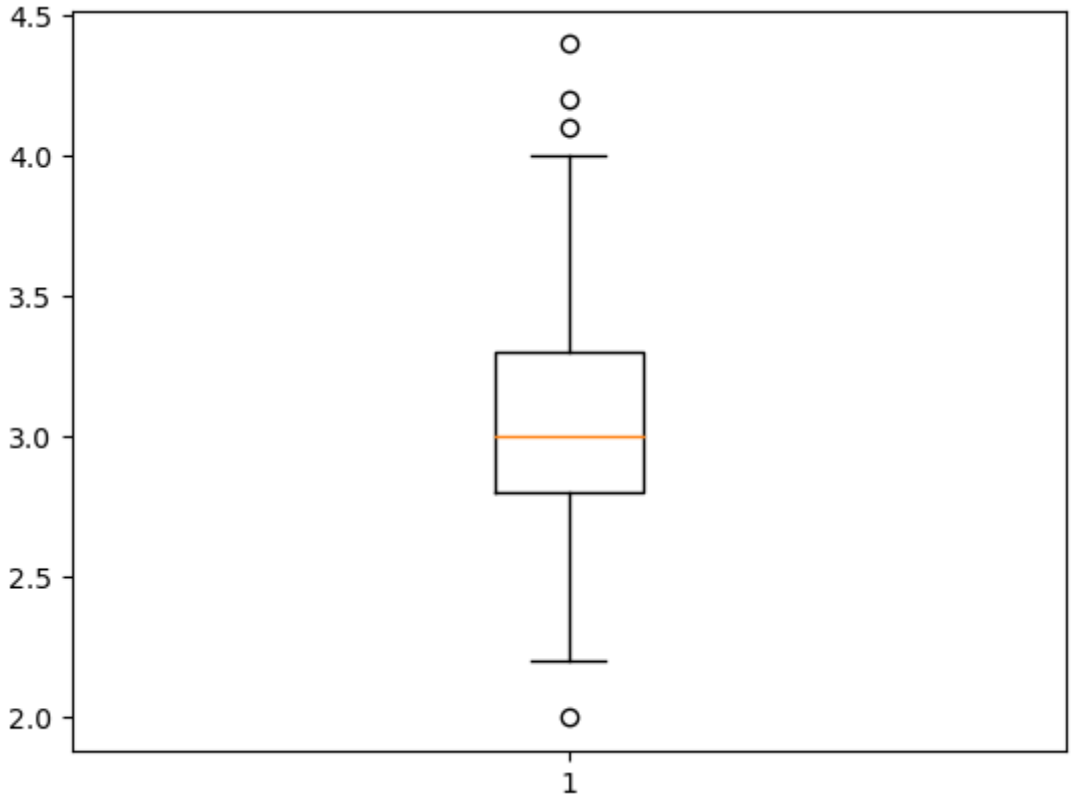
In [40]:
```
mpl.boxplot(database['SepalLengthCm'])
```

Out[40]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x14d321c6320>,
  <matplotlib.lines.Line2D at 0x14d321c65c0>],
 'caps': [<matplotlib.lines.Line2D at 0x14d321c6860>,
  <matplotlib.lines.Line2D at 0x14d321c6b00>],
 'boxes': [<matplotlib.lines.Line2D at 0x14d321c6080>],
 'medians': [<matplotlib.lines.Line2D at 0x14d321c6da0>],
 'fliers': [<matplotlib.lines.Line2D at 0x14d321c7040>],
 'means': []}
```



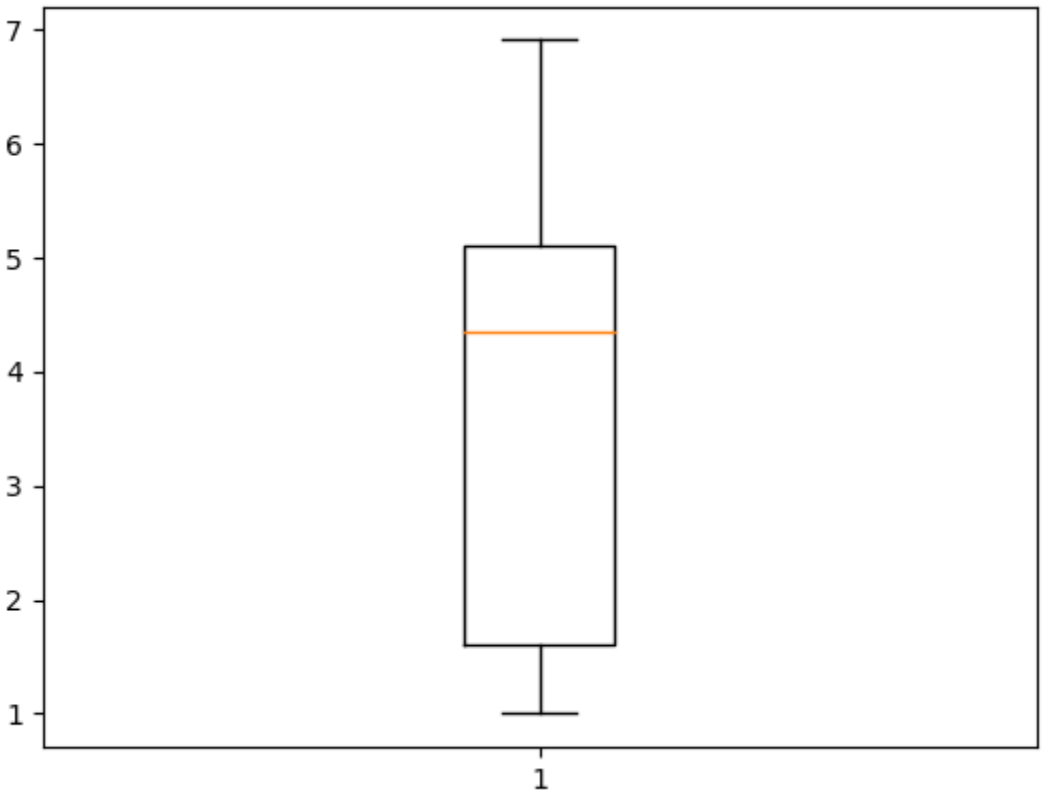
In [41]:
```
mpl.boxplot(database['SepalWidthCm'])
```

Out[41]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x14d32ac9d50>,
  <matplotlib.lines.Line2D at 0x14d32ac9ff0>],
 'caps': [<matplotlib.lines.Line2D at 0x14d32aca290>,
  <matplotlib.lines.Line2D at 0x14d32aca530>],
 'boxes': [<matplotlib.lines.Line2D at 0x14d32ac9ab0>],
 'medians': [<matplotlib.lines.Line2D at 0x14d32aca7d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x14d32acaa70>],
 'means': []}
```



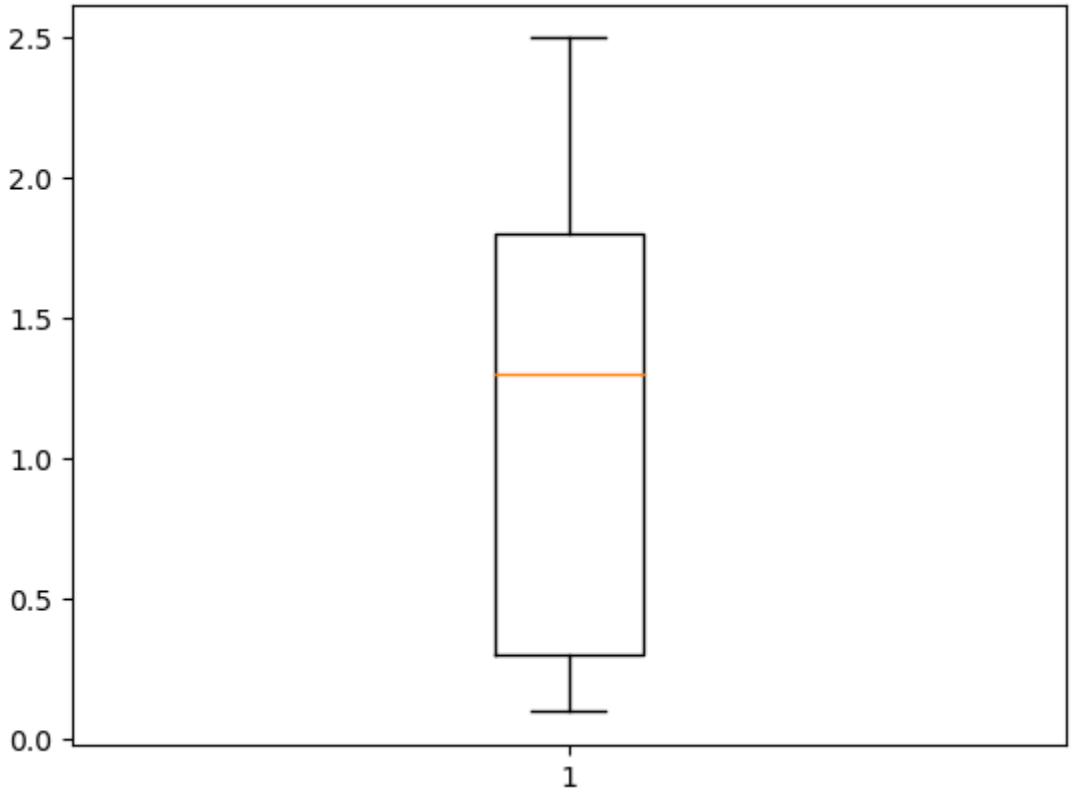
In [42]:
```
mpl.boxplot(database['PetalLengthCm'])
```

Out[42]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x14d32270b50>,
  <matplotlib.lines.Line2D at 0x14d32270df0>],
 'caps': [<matplotlib.lines.Line2D at 0x14d32271090>,
  <matplotlib.lines.Line2D at 0x14d32271330>],
 'boxes': [<matplotlib.lines.Line2D at 0x14d322708b0>],
 'medians': [<matplotlib.lines.Line2D at 0x14d322715d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x14d32271870>],
 'means': []}
```



In [43]:
```
mpl.boxplot(database['PetalWidthCm'])
```

Out[43]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x14d322c3d90>,
  <matplotlib.lines.Line2D at 0x14d322f0070>],
 'caps': [<matplotlib.lines.Line2D at 0x14d322f0310>,
  <matplotlib.lines.Line2D at 0x14d322f05b0>],
 'boxes': [<matplotlib.lines.Line2D at 0x14d322c3af0>],
 'medians': [<matplotlib.lines.Line2D at 0x14d322f0850>],
 'fliers': [<matplotlib.lines.Line2D at 0x14d322f0a60>],
 'means': []}
```



DATA PREPARATION:

In [44]:
```
X = database.iloc[:,0:4]
```

In [45]:
```
Y = database.iloc[:,4]
```

In [46]:
```
X
```

Out[46]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm
0	1	5.1	3.5	1.4
1	2	4.9	3.0	1.4
2	3	4.7	3.2	1.3
3	4	4.6	3.1	1.5
4	5	5.0	3.6	1.4
...	...	...	...	...
145	146	6.7	3.0	5.2
146	147	6.3	2.5	5.0
147	148	6.5	3.0	5.2
148	149	6.2	3.4	5.4
149	150	5.9	3.0	5.1

150 rows × 4 columns

In [47]:
```
Y
```

Out[47]:
```
0      0.2
1      0.2
2      0.2
3      0.2
4      0.2
      ...
145    2.3
146    1.9
147    2.0
148    2.3
149    1.8
Name: PetalWidthCm, Length: 150, dtype: float64
```

TRAINING THE MODEL:

In [48]:
```
from sklearn.model_selection import train_test_split
```

In [49]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size= 0.5, random_s
```

In [50]:
```
from sklearn.linear_model import LinearRegression
```

In [51]:
```
reg=LinearRegression()
```

In [53]:
```
reg.fit(X,Y)
```

Out[53]:
```
▾ LinearRegression
LinearRegression()
```

In [54]:
```
reg.score(X,Y)
```

Out[54]:
```
0.9462594479660653
```

In [56]:
```
reg.intercept_
```

Out[56]:
```
-0.3811353666896995
```

PREDICTIONS:

In [57]:
```
Y_pred=reg.predict(X_test)
```

MODEL EVALUATION:

In [58]:
```
print("Mean Squared Error: %.2f" % np.mean((Y_pred - Y_test) ** 2))
```

```
Mean Squared Error: 0.03
```

In [ ]: