In [11]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split

df=pd.read_csv("Iris1.csv")
df.head()
```

Out[11]:

	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	NaN	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

In [12]:
```
df.isnull().sum()
```

Out[12]:
```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    8
PetalWidthCm     0
Species          0
dtype: int64
```

In [13]:
```
df['PetalLengthCm'].mean()
```

Out[13]:
```
4.246575342465756
```

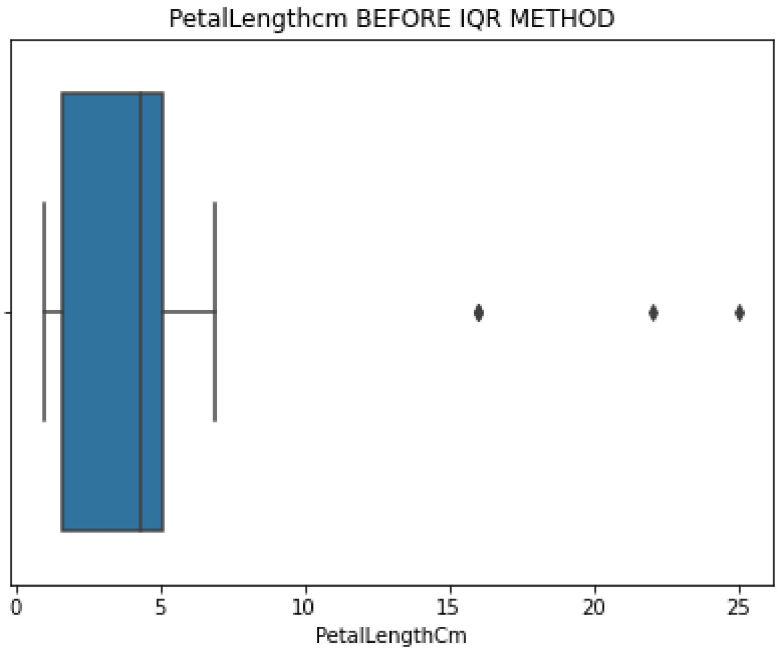
In [14]:
```
df['PetalLengthCm'].fillna(df['PetalLengthCm'].mean(),inplace=True)
```

In [15]:
```
df['PetalLengthCm'].isnull().sum()
```

Out[15]:
```
0
```

In [24]:
```
plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
plt.title('PetalLengthcm BEFORE IQR METHOD')
sns.boxplot(df['PetalLengthCm'])
```

Out[24]:
```
C:\Users\akshu\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only valid po
sitional argument will be `data`, and passing other arguments without an explicit key
word will result in an error or misinterpretation.
  warnings.warn(
<AxesSubplot:title={'center':'PetalLengthcm BEFORE IQR METHOD'}, xlabel='PetalLengthC
m'>
```



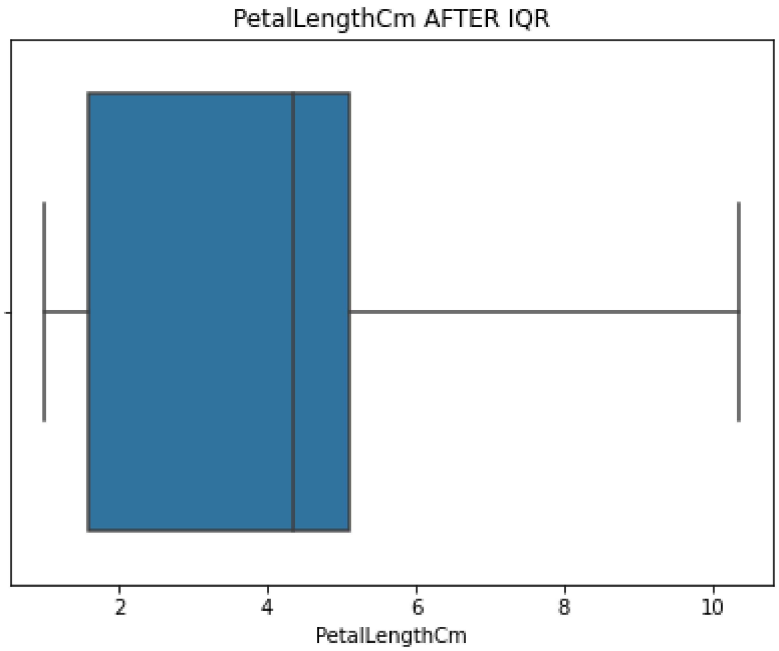
In [17]:
```
q1=df['PetalLengthCm'].quantile(0.25)
q3=df['PetalLengthCm'].quantile(0.75)
iqr=q3-q1
q1,q3,iqr
upper_limit=q3+(1.5*iqr)
lower_limit=q1-(1.5*iqr)
upper_limit,lower_limit
```

Out[17]:
```
(10.349999999999998, -3.649999999999999)
```

In [25]:
```
df.loc[(df['PetalLengthCm']>upper_limit),'PetalLengthCm']=upper_limit
df.loc[(df['PetalLengthCm']<lower_limit),'PetalLengthCm']=lower_limit
plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
plt.title('PetalLengthCm AFTER IQR')
sns.boxplot(df['PetalLengthCm'])
```

```
C:\Users\akshu\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only valid po
sitional argument will be `data`, and passing other arguments without an explicit key
word will result in an error or misinterpretation.
  warnings.warn(
```

Out[25]:
```
<AxesSubplot:title={'center':'PetalLengthCm AFTER IQR'}, xlabel='PetalLengthCm'>
```



In [20]:
```
x=df.drop(columns=['SepalWidthCm'],axis=1)
y=df['SepalWidthCm']
x_train,x_test,y_test,y_train=train_test_split(x,y,test_size=0.2,random_state=5)
print(x_test)
```

	Id	SepalLengthCm	PetalLengthCm	PetalWidthCm	Species
148	149	6.2	5.400000	2.3	Iris-virginica
21	22	5.1	1.500000	0.4	Iris-setosa
48	49	5.3	1.500000	0.2	Iris-setosa
39	40	5.1	4.246575	0.2	Iris-setosa
139	140	6.9	5.400000	2.1	Iris-virginica
123	124	6.3	4.900000	1.8	Iris-virginica
1	2	4.9	1.400000	0.2	Iris-setosa
98	99	5.1	3.000000	1.1	Iris-versicolor
42	43	4.4	1.300000	0.2	Iris-setosa
72	73	6.3	4.900000	1.5	Iris-versicolor
40	41	5.0	1.300000	0.3	Iris-setosa
134	135	6.1	5.600000	1.4	Iris-virginica
137	138	6.4	5.500000	1.8	Iris-virginica
114	115	5.8	4.246575	2.4	Iris-virginica
111	112	6.4	5.300000	1.9	Iris-virginica
84	85	5.4	4.500000	1.5	Iris-versicolor
61	62	5.9	4.200000	1.5	Iris-versicolor
63	64	6.1	4.700000	1.4	Iris-versicolor
136	137	6.3	5.600000	2.4	Iris-virginica
93	94	5.0	3.300000	1.0	Iris-versicolor
12	13	4.8	1.400000	0.1	Iris-setosa
150	151	54.0	16.000000	23.0	Iris-virginica
23	24	5.1	1.700000	0.5	Iris-setosa
127	128	6.1	4.900000	1.8	Iris-virginica
74	75	6.4	4.246575	1.3	Iris-versicolor
135	136	7.7	6.100000	2.3	Iris-virginica
25	26	5.0	1.600000	0.2	Iris-setosa
26	27	5.0	1.600000	0.4	Iris-setosa
124	125	6.7	5.700000	2.1	Iris-virginica
79	80	5.7	3.500000	1.0	Iris-versicolor
104	105	6.5	5.800000	2.2	Iris-virginica

In [21]:
```
print(x_train)
```

```
      Id  SepalLengthCm  PetalLengthCm  PetalWidthCm          Species
71    72            6.1            4.0           1.3  Iris-versicolor
69    70            5.6            3.9           1.1  Iris-versicolor
57    58            4.9            3.3           1.0  Iris-versicolor
81    82            5.5            3.7           1.0  Iris-versicolor
10    11            5.4            1.5           0.2      Iris-setosa
..   ...            ...            ...           ...              ...
8      9            4.4            1.4           0.2      Iris-setosa
73    74            6.1            4.7           1.2  Iris-versicolor
144  145            6.7            5.7           2.5   Iris-virginica
118  119            7.7            6.9           2.3   Iris-virginica
99   100            5.7            4.1           1.3  Iris-versicolor

[123 rows x 5 columns]
```

In [22]:
```
print(y_test)
```

```
71     2.8
69     2.5
57     2.4
81     2.4
10     3.7
      ...
8      2.9
73     2.8
144    3.3
118    2.6
99     2.8
Name: SepalWidthCm, Length: 123, dtype: float64
```

In [23]:
```
print(y_train)
```

```
148     3.4
21      3.7
48      3.7
39      3.4
139     3.1
123     2.7
1       3.0
98      2.5
42      3.2
72      2.5
40      3.5
134     2.6
137     3.1
114     2.8
111     2.7
84      3.0
61      3.0
63      2.9
136     3.4
93      2.3
12      3.0
150    15.0
23      3.3
127     3.0
74      2.9
135     3.0
25      3.0
26      3.4
124     3.3
79      2.6
104     3.0
Name: SepalWidthCm, dtype: float64
```

In [ ]: