remove_outlier_using_Iqr

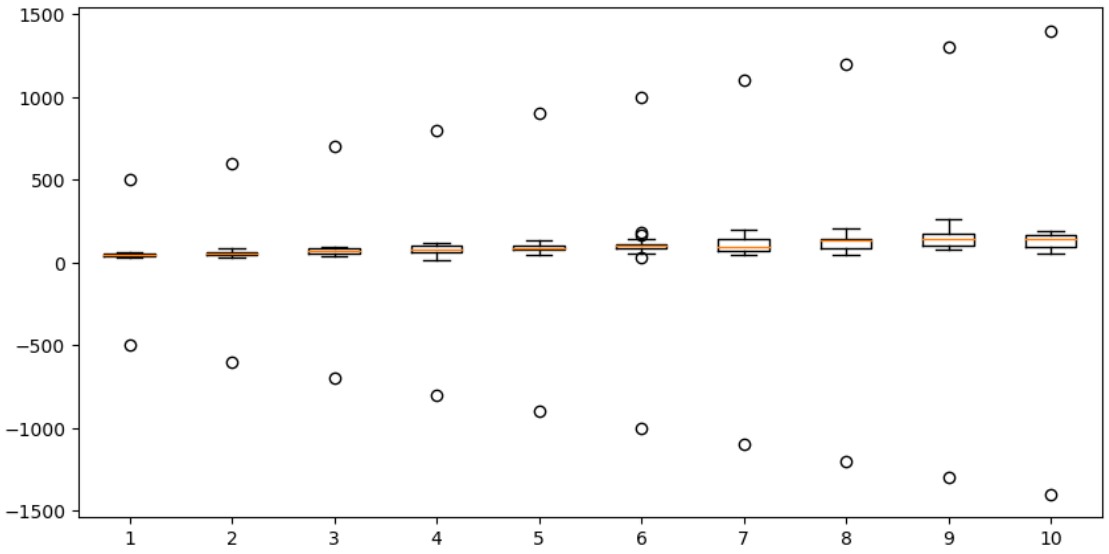
July 10, 2024

[90]:
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("Outiers_iqr.csv")
```

[ ]:
```
```

[ ]:
```
df.describe()
```

[63]:
```
plt.figure(figsize= (10,5))
plt.boxplot(df)
plt.show()
```



[ ]:
```
```

[3]:
```
df.shape
```

[3]:
```
(22, 10)
```

0.0.1 Calculate the IQR for 'Col1' in the DataFrame.

[4]:
```
Q1 = df["Col1"].quantile(0.25)
Q3 = df["Col1"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[4]:
```
13.526885098934336
```

0.0.2 Identify outliers in 'Col1' using the IQR method.

[5]:
```
upper_limit = Q3 +1.5 * IQR
lower_limit = Q1-1.5 * IQR
upper_limit , lower_limit
```

[5]:
```
(75.60131335782432, 21.493772962086982)
```

[6]:
```
df[(df["Col1"]>= 75.60)|(df["Col1"]<= 21.49)]
```

[6]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.3 Remove the outliers from 'Col1' using the IQR method:

[7]:
```
df_col1 = df[(df["Col1"]<=75.60)&(df["Col1"]>=21.49)]
df_col1.head()
```

[7]:

	Col1	Col2	Col3	Col4	Col5	Col6	\
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	

	Col7	Col8	Col9	Col10
0	141.641278	130.235697	81.265916	174.411704
1	73.624502	178.821424	169.354230	92.856334
2	166.111772	47.663254	187.929779	81.100913
3	53.925957	128.308524	88.965884	166.535983
4	133.474284	131.694726	178.168806	127.709547

[8]:
```
import matplotlib.pyplot as plt
# before removal outliers
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col1"])
```

```
plt.title("Before removal outiers")

# after removal outliers
plt.subplot(1,2,2)
plt.boxplot(df_col1["Col1"])
plt.title("after removal outliers")

plt.show()
```

0.0.4 Calculate the IQR for 'Col2' in the DataFrame.

[9]:
```
Q1 = df["Col2"].quantile(0.25)
Q3 = df["Col2"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[9]: 20.00303880789825

[10]:
```
upper_limit = Q3 +1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[10]: (93.0925032901601, 13.08034805856709)

0.0.5 Identify outliers in 'Col2' using the IQR method.

[11]:
```
df[(df["Col2"]>=242.36)|(df["Col2"]<=-31.67)]
```

[11]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

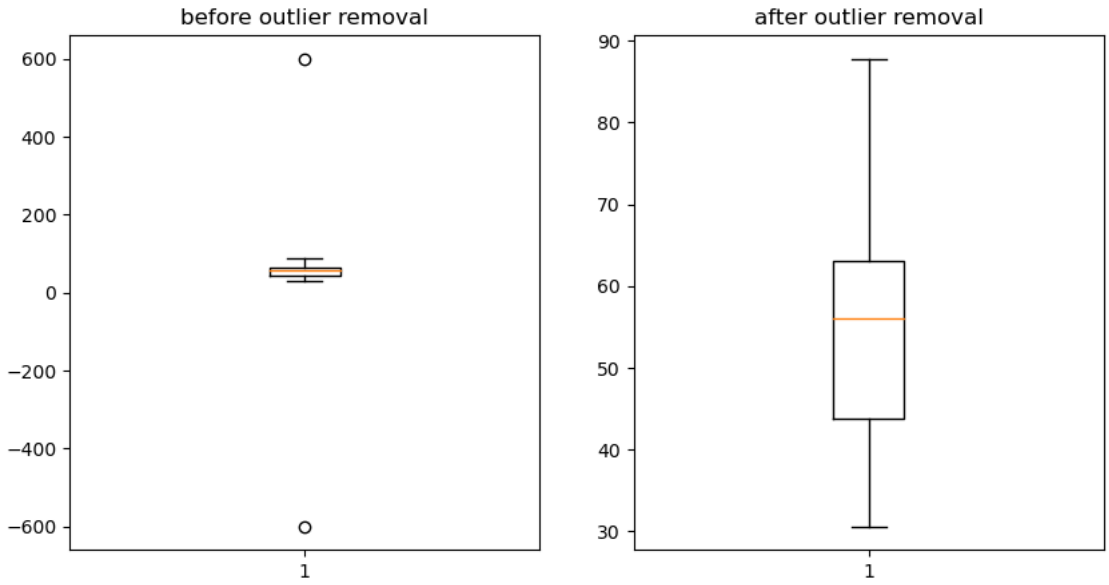
0.0.6 Remove the outliers from 'Col2' using the IQR method.

[12]:
```
df_col2 = df[(df["Col2"]<=242.36)&(df["Col2"]>=-31.67)]
df_col2.head()
```

[12]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	141.641278	130.235697	81.265916	174.411704
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	73.624502	178.821424	169.354230	92.856334
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	166.111772	47.663254	187.929779	81.100913
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	53.925957	128.308524	88.965884	166.535983
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	133.474284	131.694726	178.168806	127.709547

[13]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col2"])
plt.title("before outlier removal")

# after outlier

plt.subplot(1,2,2)
plt.boxplot(df_col2["Col2"])
plt.title("after outlier removal")

plt.show()
```



[14]:
```
df["Col2"].describe()
```

[14]:
```
count     22.000000
mean      50.918521
std      186.408413
min     -600.000000
25%       43.084906
50%       56.118975
75%       63.087945
max      600.000000
Name: Col2, dtype: float64
```

[15]:
```
df_col2["Col2"].describe()
```

[15]:
```
count    20.000000
mean     56.010373
std      14.520581
min      30.604948
25%      43.784526
50%      56.118975
75%      62.997927
max      87.784173
Name: Col2, dtype: float64
```

0.0.7 Calculate the IQR for 'Col3' in the DataFrame.

[16]:
```
Q1 = df["Col3"].quantile(0.25)
Q3 = df["Col3"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[16]:
```
26.59240400352381
```

0.0.8 Identify outliers in 'Col3' using the IQR method.

[17]:
```
upper_limit = Q3+ 1.5* IQR
lower_limit = Q1 - 1.5* IQR
upper_limit , lower_limit
```

[17]:
```
(124.0239861494212, 17.65437013532597)
```

[18]:
```
df[(df["Col3"]>=242.36)|(df["Col3"]<=-31.67)]
```

[18]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.9 Remove the outliers from 'Col3' using the IQR method.

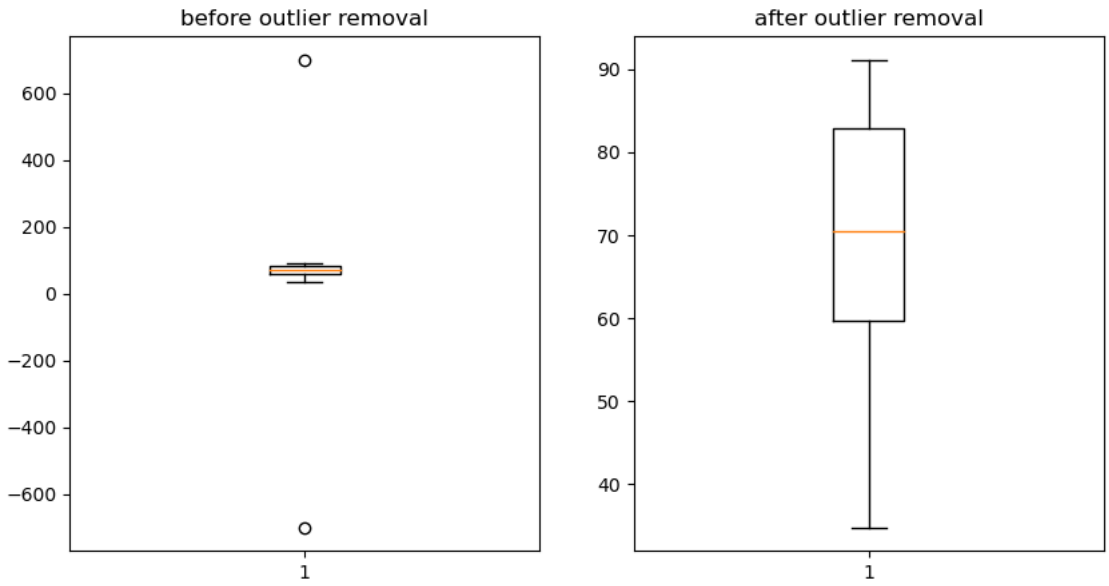
[19]:
```
df_col3 = df[(df["Col3"]<=242.36)&(df["Col3"]>=-31.67)]
df_col3.head()
```

[19]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	141.641278	130.235697	81.265916	174.411704
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	73.624502	178.821424	169.354230	92.856334
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	166.111772	47.663254	187.929779	81.100913
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	53.925957	128.308524	88.965884	166.535983
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	133.474284	131.694726	178.168806	127.709547

[20]:
```
plt.figure(figsize=(10,5))
# before removal outlier
plt.subplot(1,2,1)
plt.boxplot(df["Col3"])
plt.title("before outlier removal")
```

```
# after outlier removal
plt.subplot(1,2,2)
plt.boxplot(df_col3["Col3"])
plt.title("after outlier removal")

plt.show()
```



0.0.10 Calculate the IQR for 'Col4' in the DataFrame.

[21]:
```
Q1 = df["Col4"].quantile(0.25)
Q3 = df["Col4"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[21]: 35.57979273237832

0.0.11 Identify outliers in 'Col4' using the IQR method.

[22]:
```
upper_limit = Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[22]: (153.85864744556667, 11.539476516053377)

[23]:
```
df[(df["Col4"]>=242.36)|(df["Col4"]<=-31.67)]
```

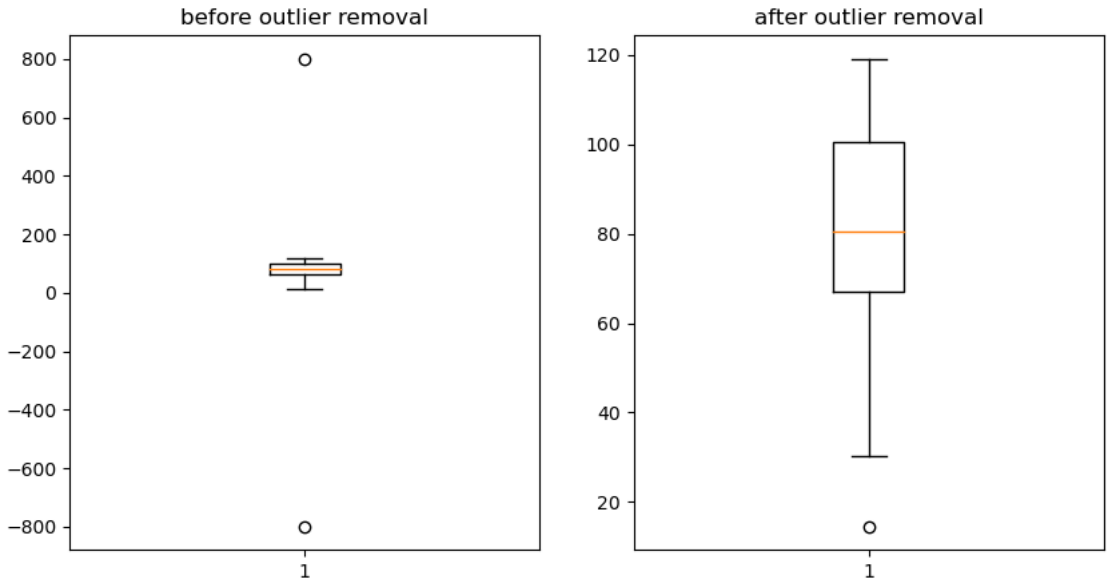
[23]:
```
      Col1   Col2   Col3   Col4   Col5    Col6    Col7    Col8    Col9   Col10
20   500.0  600.0  700.0  800.0  900.0  1000.0  1100.0  1200.0  1300.0  1400.0
21  -500.0 -600.0 -700.0 -800.0 -900.0 -1000.0 -1100.0 -1200.0 -1300.0 -1400.0
```

0.0.12 Remove the outliers from 'Col4' using the IQR method.

[24]:
```
df_col4 = df[(df["Col4"]<=242.36)&(df["Col4"]>=-31.67)]
df_col4.head()
```

[24]:
```
        Col1       Col2       Col3        Col4        Col5       Col6  \
0  54.967142  81.984732  84.769332   68.020644   83.409843  50.462024
1  48.617357  56.613355  73.427366   75.358526  100.713377  85.277414
2  56.476885  61.012923  67.687034   52.341626  134.336821  88.004992
3  65.230299  38.628777  63.977926   50.094834   74.451893  71.920296
4  47.658466  51.834259  40.429560  100.313146   65.745192  94.355000

         Col7        Col8        Col9       Col10
0  141.641278  130.235697   81.265916  174.411704
1   73.624502  178.821424  169.354230   92.856334
2  166.111772   47.663254  187.929779   81.100913
3   53.925957  128.308524   88.965884  166.535983
4  133.474284  131.694726  178.168806  127.709547
```

[25]:
```
plt.figure(figsize=(10,5))
# before removal outlier
plt.subplot(1,2,1)
plt.boxplot(df["Col4"])
plt.title("before outlier removal")

# after outlier removal

plt.subplot(1,2,2)
plt.boxplot(df_col4["Col4"])
plt.title("after outlier removal")

plt.show()
```



0.0.13 Calculate the IQR for 'Col5' in the DataFrame.

[26]:
```
Q1 = df["Col5"].quantile(0.25)
Q3 = df["Col5"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[26]:
```
25.924923919503016
```

0.0.14 Identify outliers in 'Col5' using the IQR method.

[27]:
```
upper_limit = Q3 + 1.5* IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[27]:
```
(139.38805206071646, 35.688356382704384)
```

[28]:
```
df[(df["Col5"]>=242.36)|(df["Col5"]<=-31.67)]
```

[28]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.15 Remove the outliers from 'Col5' using the IQR method.

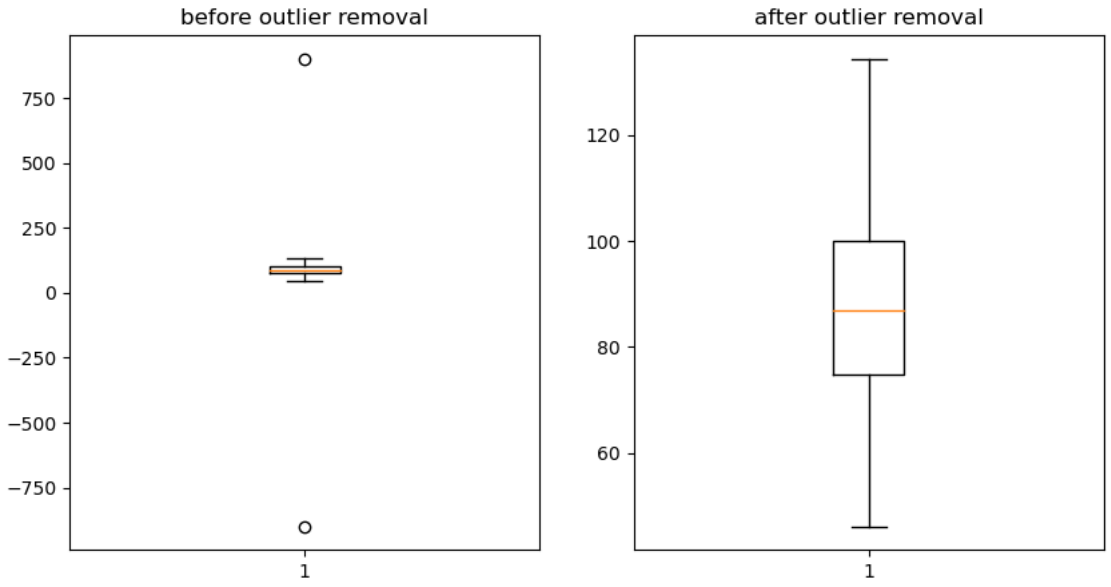
[29]:
```
df_col5 = df[(df["Col5"]<=242.36)&(df["Col5"]>=-31.67)]
df_col5.head()
```

[29]:

	Col1	Col2	Col3	Col4	Col5	Col6	\
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	

	Col7	Col8	Col9	Col10
0	141.641278	130.235697	81.265916	174.411704
1	73.624502	178.821424	169.354230	92.856334
2	166.111772	47.663254	187.929779	81.100913
3	53.925957	128.308524	88.965884	166.535983
4	133.474284	131.694726	178.168806	127.709547

[30]:
```
# before outlier removal
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col5"])
plt.title("before outlier removal")

# after outlier removal
plt.subplot(1,2,2)
plt.boxplot(df_col5["Col5"])
plt.title("after outlier removal")
plt.show()
```



0.0.16 Calculate the IQR for 'Col6' in the DataFrame.

[31]:
```
Q1= df["Col6"].quantile(0.25)
Q3 = df["Col6"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[31]:
```
27.28556597900065
```

0.0.17 Identify outliers in 'Col6' using the IQR method.

[32]:
```
upper_limit = Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[32]:
```
(154.17322320530172, 45.03095928929911)
```

[33]:
```
df[(df["Col6"]>=242.36)|(df["Col6"]<=-31.67)]
```

[33]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.18 Remove the outliers from 'Col6' using the IQR method.

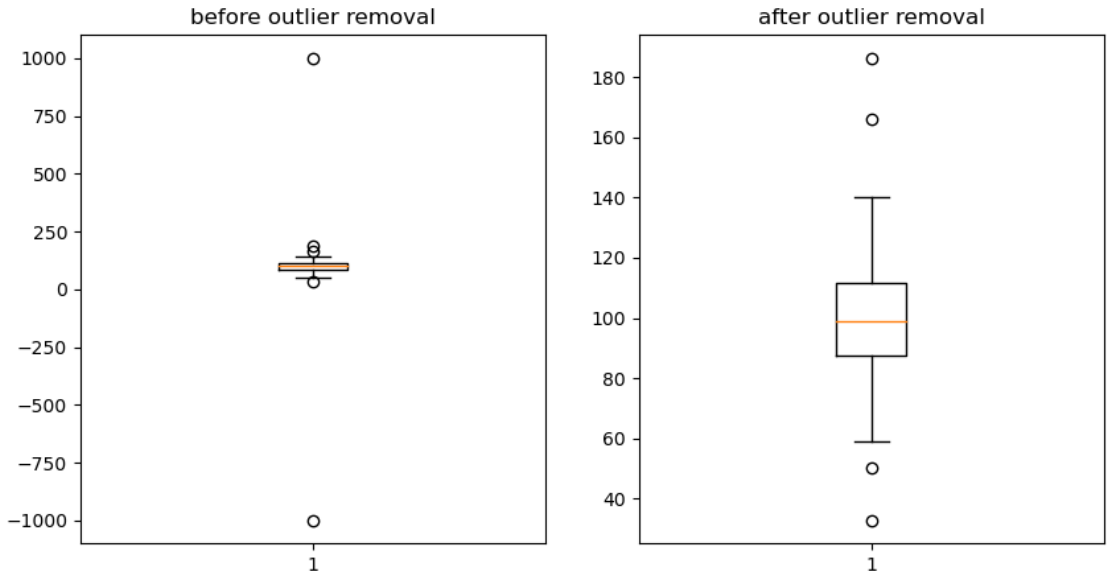
[34]:
```
df_col6 = df[(df["Col6"]<=242.36)&(df["Col6"]>=-31.67)]
df_col6.head()
```

[34]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	141.641278	130.235697	81.265916	174.411704
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	73.624502	178.821424	169.354230	92.856334
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	166.111772	47.663254	187.929779	81.100913
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	53.925957	128.308524	88.965884	166.535983
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	133.474284	131.694726	178.168806	127.709547

[35]:
```
plt.figure(figsize=(10,5))
 # after removal outlier
```

```
plt.subplot(1,2,1)
plt.boxplot(df["Col6"])
plt.title("before outlier removal")

#after removal outlier

plt.subplot(1,2,2)
plt.boxplot(df_col6["Col6"])
plt.title("after outlier removal")

plt.show()
```



0.0.19 Calculate the IQR for 'Col7' in the DataFrame.

[36]:
```
Q1 = df["Col7"].quantile(0.25)
Q3 = df["Col7"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[36]: 68.50986144555279

0.0.20 Identify outliers in 'Col7' using the IQR method.

[37]:
```
upper_limit = Q3 + 1.5* IQR
lower_limit = Q1 - 1.5* IQR
upper_limit , lower_limit
```

[37]:
```
(242.3643215176233, -31.675124264587865)
```

[38]:
```
df[(df["Col7"]>=242.36)|(df["Col7"]<=-31.67)]
```

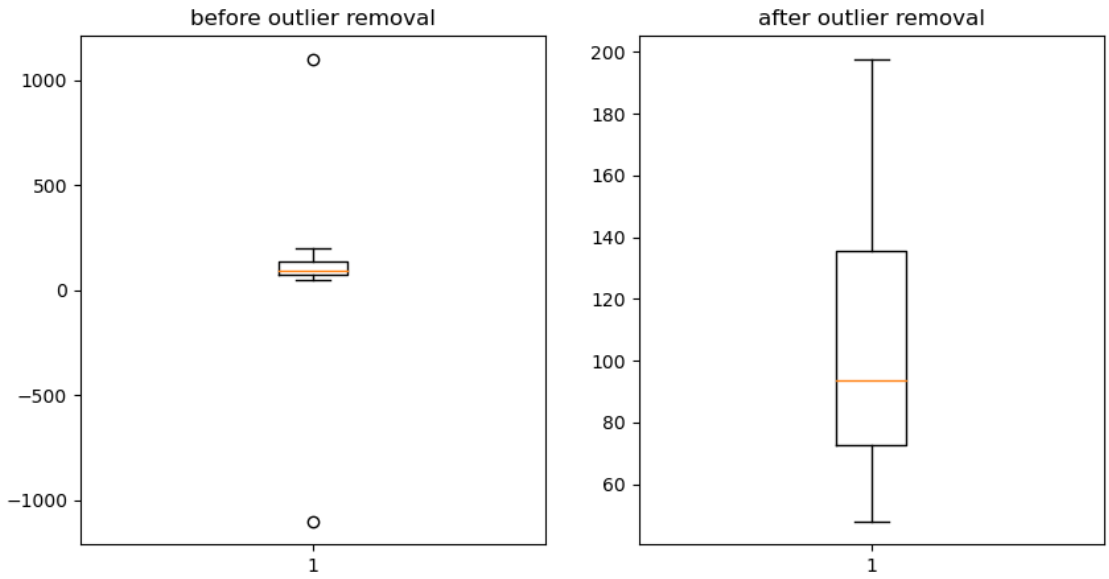
[38]:
```
      Col1   Col2   Col3   Col4   Col5    Col6    Col7    Col8    Col9   Col10
20   500.0  600.0  700.0  800.0  900.0  1000.0  1100.0  1200.0  1300.0  1400.0
21  -500.0 -600.0 -700.0 -800.0 -900.0 -1000.0 -1100.0 -1200.0 -1300.0 -1400.0
```

0.0.21 Remove the outliers from 'Col7' using the IQR method.

[39]:
```
df_col7 = df[(df["Col7"]<=242.36)&(df["Col7"]>=-31.67)]
df_col7.head()
```

[39]:
```
         Col1       Col2       Col3        Col4        Col5       Col6  \
0   54.967142  81.984732  84.769332   68.020644   83.409843  50.462024
1   48.617357  56.613355  73.427366   75.358526  100.713377  85.277414
2   56.476885  61.012923  67.687034   52.341626  134.336821  88.004992
3   65.230299  38.628777  63.977926   50.094834   74.451893  71.920296
4   47.658466  51.834259  40.429560  100.313146   65.745192  94.355000

         Col7        Col8        Col9       Col10
0  141.641278  130.235697   81.265916  174.411704
1   73.624502  178.821424  169.354230   92.856334
2  166.111772   47.663254  187.929779   81.100913
3   53.925957  128.308524   88.965884  166.535983
4  133.474284  131.694726  178.168806  127.709547
```

[40]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col7"])
plt.title("before outlier removal")

# after outlier removal

plt.subplot(1,2,2)
plt.boxplot(df_col7["Col7"])
plt.title("after outlier removal")

plt.show()
```



0.0.22 Calculate the IQR for 'Col8' in the DataFrame.

[41]:
```
Q1 = df["Col8"].quantile(0.25)
Q3 = df["Col8"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[41]:
```
54.705786675975844
```

0.0.23 Identify outliers in 'Col8' using the IQR method.

[42]:
```
upper_limit= Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[42]:
```
(225.00482821379168, 6.181681509888307)
```

[43]:
```
df[(df["Col8"]>=225.00)|(df["Col8"]<=6.18)]
```

[43]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.24 Remove the outliers from 'Col8' using the IQR method.

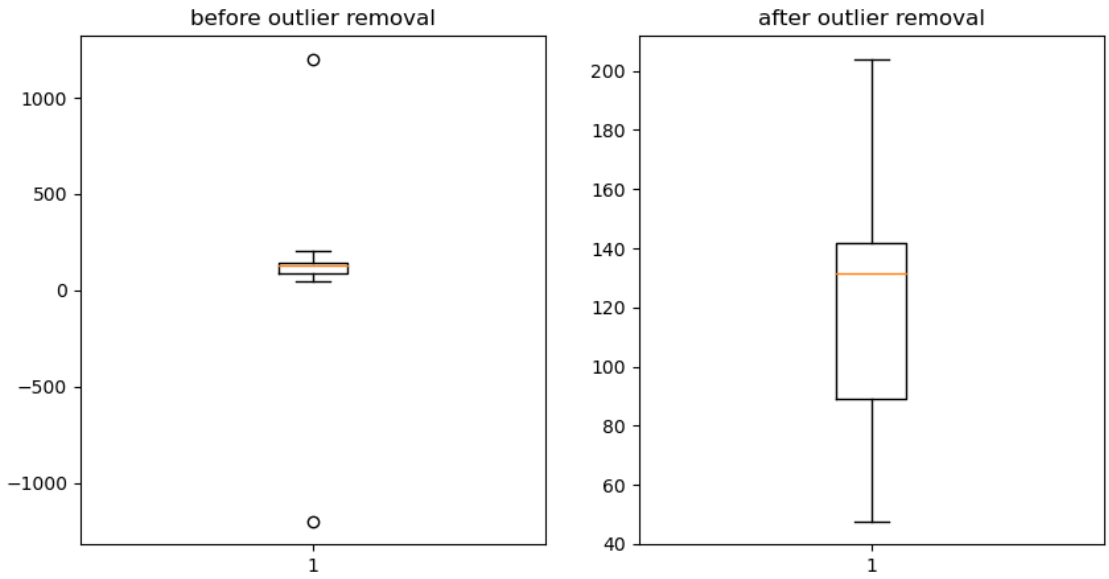
[44]:
```
df_col8 = df[(df["Col8"]<=225.00)&(df["Col8"]>=6.18)]
df_col8.head()
```

[44]:

	Col1	Col2	Col3	Col4	Col5	Col6	\
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	

	Col7	Col8	Col9	Col10
0	141.641278	130.235697	81.265916	174.411704
1	73.624502	178.821424	169.354230	92.856334
2	166.111772	47.663254	187.929779	81.100913
3	53.925957	128.308524	88.965884	166.535983
4	133.474284	131.694726	178.168806	127.709547

[45]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col8"])
plt.title("before outlier removal")

# after outlier removal

plt.subplot(1,2,2)
plt.boxplot(df_col8["Col8"])
plt.title("after outlier removal")

plt.show()
```



0.0.25 Calculate the IQR for 'Col9' in the DataFrame.

[46]:
```
Q1 = df["Col9"].quantile(0.25)
Q3 = df["Col9"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[46]:
```
78.03971203849267
```

0.0.26 Identify outliers in 'Col9' using the IQR method.

[47]:
```
upper_limit = Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[47]:
```
(293.52596351735133, -18.63288463661931)
```

[48]:
```
df[(df["Col9"]>=293.25)|(df["Col9"]<=-18.63)]
```

[48]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.27 Remove the outliers from 'Col9' using the IQR method.

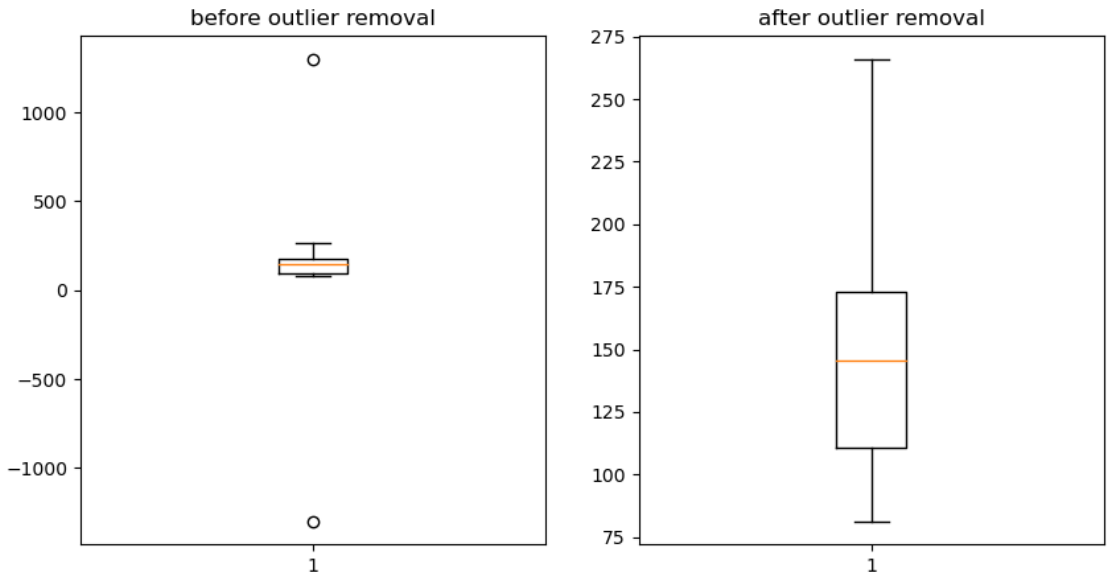
[49]:
```
df_col9 = df[(df["Col9"]<=293.25)&(df["Col9"]>=-18.63)]
df_col9.head()
```

[49]:

	Col1	Col2	Col3	Col4	Col5	Col6	\
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	

	Col7	Col8	Col9	Col10
0	141.641278	130.235697	81.265916	174.411704
1	73.624502	178.821424	169.354230	92.856334
2	166.111772	47.663254	187.929779	81.100913
3	53.925957	128.308524	88.965884	166.535983
4	133.474284	131.694726	178.168806	127.709547

[50]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col9"])
plt.title("before outlier removal")

# after outlier removal

plt.subplot(1,2,2)
plt.boxplot(df_col9["Col9"])
plt.title("after outlier removal")

plt.show()
```



0.0.28 Calculate the IQR for 'Col10' in the DataFrame.

[51]:
```
Q1 = df["Col10"].quantile(0.25)
Q3 = df["Col10"].quantile(0.75)
IQR = Q3 - Q1
IQR
```

[51]:
```
73.06420162895085
```

0.0.29 Identify outliers in 'Col10' using the IQR method.

[52]:
```
upper_limit = Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
upper_limit , lower_limit
```

[52]:
```
(274.4154684283565, -17.8413380874469)
```

[53]:
```
df[(df["Col10"]>=274.41)|(df["Col10"]<=-17.84)]
```

[53]:

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
20	500.0	600.0	700.0	800.0	900.0	1000.0	1100.0	1200.0	1300.0	1400.0
21	-500.0	-600.0	-700.0	-800.0	-900.0	-1000.0	-1100.0	-1200.0	-1300.0	-1400.0

0.0.30 Remove the outliers from 'Col10' using the IQR method.

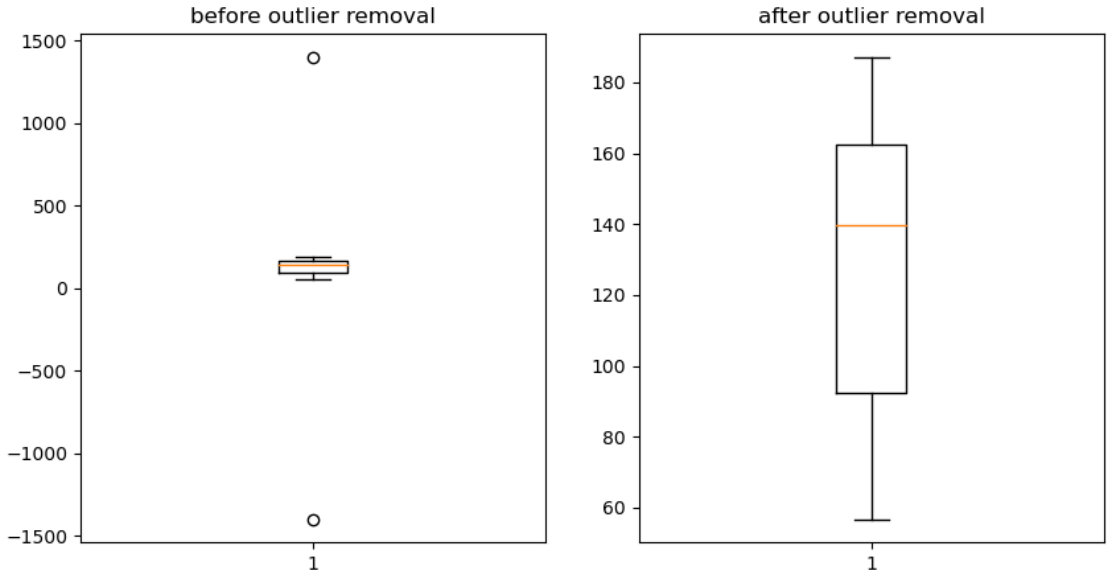
[54]:
```
df_col10 = df[(df["Col10"]<=274.41)&(df["Col10"]>=-17.84)]
df_col10.head()
```

[54]:

	Col1	Col2	Col3	Col4	Col5	Col6	\
0	54.967142	81.984732	84.769332	68.020644	83.409843	50.462024	
1	48.617357	56.613355	73.427366	75.358526	100.713377	85.277414	
2	56.476885	61.012923	67.687034	52.341626	134.336821	88.004992	
3	65.230299	38.628777	63.977926	50.094834	74.451893	71.920296	
4	47.658466	51.834259	40.429560	100.313146	65.745192	94.355000	

	Col7	Col8	Col9	Col10
0	141.641278	130.235697	81.265916	174.411704
1	73.624502	178.821424	169.354230	92.856334
2	166.111772	47.663254	187.929779	81.100913
3	53.925957	128.308524	88.965884	166.535983
4	133.474284	131.694726	178.168806	127.709547

[55]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.boxplot(df["Col10"])
plt.title("before outlier removal")

# after outlier removal

plt.subplot(1,2,2)
plt.boxplot(df_col10["Col10"])
plt.title("after outlier removal")

plt.show()
```


before outlier removal
1500
1000
500
0
−500
−1000
−1500
1
after outlier removal
180
160
140
120
100
80
60
1

[ ]: