

In [1]: `import pandas as pd`

In [8]: `import numpy as np`

In [3]: `df=pd.read_csv("C:\\Users\\admn\\Desktop\\Ex-02_DS_Outlier-main\\weight.csv")`

In [4]: `df`

Out[4]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

10000 rows × 3 columns

In [5]: `df.drop("Gender", axis=1,inplace=True)`

In [6]: `df`

Out[6]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

10000 rows × 2 columns

In [9]: `df.boxplot()`



In [18]: `from scipy import stats`

In [16]: `z=np.abs(stats.zscore(df))`

In [17]: `z`

Out[17]:

	Height	Weight
0	1.944061	2.505797
1	0.627537	0.027101
2	2.012443	1.597806
3	1.394060	1.825222
4	0.913421	1.398750
...
9995	0.050660	0.768151
9996	0.181839	0.293631
9997	0.649688	1.026730
9998	0.693125	0.075127
9999	1.149708	1.488507

10000 rows × 2 columns

In [13]: `df`

Out[13]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

10000 rows × 2 columns

In [18]: `df1=df.copy()`

In [19]: `df1=df1[(z<3).all(axis=1)]`

In [28]: `df1.boxplot()`



In [21]: `df1`

Out[21]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

9993 rows × 2 columns

In [22]: `df2=df.copy()`

In [23]: `q1=df2.quantile(0.25)`

In [24]: `q3=df2.quantile(0.75)`

In [25]: `IQR=q3-q1`

In [26]: `IQR`

Out[26]: `Height 5.668641`
`Weight 51.351474`
`dtype: float64`

In [27]: `IQR.Height`

Out[27]: `5.668641245615746`

In [28]: `df2_new=df2[((df2>=q1-1.5*IQR)&(df2<=q3+1.5*IQR)).all(axis=1)]`

In [29]: `df2`

Out[29]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

10000 rows × 2 columns

In []: