Ex-02_DS_Outlier

› AIM

To read the given data and check the outliers and to remove them

› ALGORITHM

› STEP 1

Read the given Data

› STEP 2

Convert the data into a data frame by importing pandas

› STEP 3

Study the outliers and remove the unnecessary data columns

› STEP 4

Import numpy and from scipy import stats to find the z-score

› STEP 5

Set the range for z-score and find IQR

› STEP 6

Plot the results

› CODE

OUTPUT

```
In [10]: import pandas as pd
         df=pd.read_csv("weight.csv")
         df
```

Out[10]:

	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 [11]: df.drop("Gender",axis=1,inplace=True)
         df
```
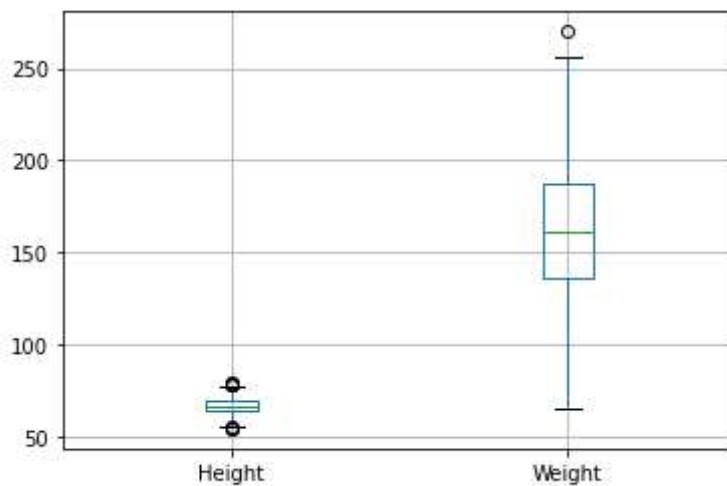
Out[11]:

	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 [19]: df.boxplot()
```

Out[19]: <AxesSubplot:>

```
In [13]: from scipy import stats
         import numpy as np
         z=np.abs(stats.zscore(df))
         cp1=df.copy()
         cp1=cp1[(z<3).all(axis=1)]
         cp1
```
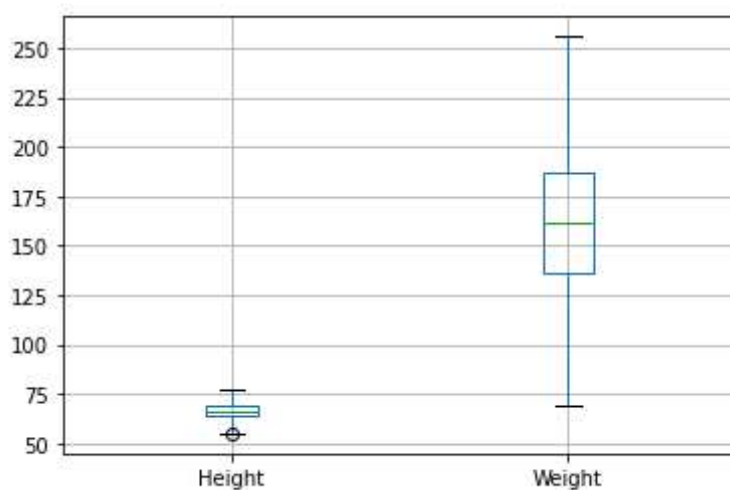
Out[13]:

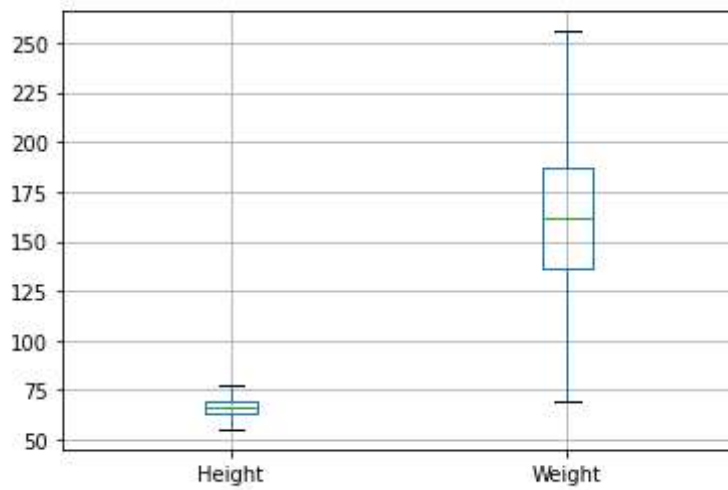|	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...	...	...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

9993 rows × 2 columns

```
In [16]: cp1.boxplot()
```

Out[16]: <AxesSubplot:>

9992 rows × 2 columns

RESULT

Thus the outiers has been successfully removed from the data frame.