Ex02-Outlier

You are given bhp.csv which contains property prices in the city of banglore, India. You need to examine price_per_sqft column and do following,

(1) Remove outliers using IQR

(2) After removing outliers in step 1, you get a new dataframe.

(3) use zscore of 3 to remove outliers. This is quite similar to IQR and you will get exact same result

(4) for the data set height_weight.csv find the following

```
(i) Using IQR detect weight outliers and print them

(ii) Using IQR, detect height outliers and print them
```

CODE

bhp.csv

```
import numpy as np
import pandas as pd
from scipy import stats

a = pd.read_csv('bhp.csv')
df = pd.DataFrame(a['price_per_sqft'])
median = df.quantile(0.5)
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
low = Q1 - 1.5 * IQR
high = Q3 + 1.5 * IQR
df1 = df[((df >= Q1 - 1.5 * IQR) & (df <= Q3 + 1.5 * IQR))]

print(df1)


z = np.abs(stats.zscore(df))
df1 = df1[(z < 3)]
print(df1)
```

height_weight.csv

```python
import numpy as np
import pandas as pd
from scipy import stats

a = pd.read_csv('height_weight.csv')
df = pd.DataFrame(a['height'])
print(df)
median = df.quantile(0.5)
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
low = Q1 - 1.5 * IQR
high = Q3 + 1.5 * IQR
df1 = df[((df >= Q1 - 1.5 * IQR) & (df <= Q3 + 1.5 * IQR))]
print(df1)
df2 = pd.DataFrame(a['weight'])
print(df2)
q1 = df2.quantile(0.25)
q3 = df2.quantile(0.75)
IQR = q3 - q1
df2_new = df2[((df2 >= q1 - 1.5 * IQR) & (df2 <= q3 + 1.5 * IQR))]
print(df2_new)
```

OUTPUT

bhp.csv

```
/home/sec/PycharmProjects/pythonProject2/venv/bin/python /home/sec/PycharmProjects/pythonProject2/main.py
       price_per_sqft
0             3699.0
1             4615.0
2             4305.0
3             6245.0
4             4250.0
...              ...
13195         6689.0
13196        11111.0
13197         5258.0
13198        10407.0
13199         3090.0

[13200 rows x 1 columns]
       price_per_sqft
0             3699.0
1             4615.0
2             4305.0
3             6245.0
4             4250.0
...              ...
13195         6689.0
13196        11111.0
13197         5258.0
13198        10407.0
13199         3090.0

[13200 rows x 1 columns]

Process finished with exit code 0
```

height_weight.csv

```
         height
0       73.847017
1       68.781904
2       74.110105
3       71.730978
4       69.881796
...          ...
9995    66.172652
9996    67.067155
9997    63.867992
9998    69.034243
9999    61.944246

[10000 rows x 1 columns]
         height
0       73.847017
1       68.781904
2       74.110105
3       71.730978
4       69.881796
...          ...
9995    66.172652
9996    67.067155
9997    63.867992
9998    69.034243
9999    61.944246

[10000 rows x 1 columns]
         weight
0       241.893563
1       162.310473
2       212.740856
3       220.042470
4       206.349801
...          ...
9995    136.777454
9996    170.867906
9997    128.475319
```

```
...         ...
9995    66.172652
9996    67.067155
9997    63.867992
9998    69.034243
9999    61.944246

[10000 rows x 1 columns]
          weight
0       241.893563
1       162.310473
2       212.740856
3       220.042470
4       206.349801
...            ...
9995    136.777454
9996    170.867906
9997    128.475319
9998    163.852461
9999    113.649103

[10000 rows x 1 columns]
          weight
0       241.893563
1       162.310473
2       212.740856
3       220.042470
4       206.349801
...            ...
9995    136.777454
9996    170.867906
9997    128.475319
9998    163.852461
9999    113.649103

[10000 rows x 1 columns]

Process finished with exit code 0
```