EDA of Heat load Calculation

```
#Import basic libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')

#Import the dataset
df = pd.read_excel('/Users/ajithsmacbookair/Downloads/Data Science
/Datasets/ENB2012_data.xlsx')
df.head()
```

```
     X1     X2     X3      X4   X5  X6   X7  X8     Y1     Y2
0  0.98  514.5  294.0  110.25  7.0   2  0.0   0  15.55  21.33
1  0.98  514.5  294.0  110.25  7.0   3  0.0   0  15.55  21.33
2  0.98  514.5  294.0  110.25  7.0   4  0.0   0  15.55  21.33
3  0.98  514.5  294.0  110.25  7.0   5  0.0   0  15.55  21.33
4  0.90  563.5  318.5  122.50  7.0   2  0.0   0  20.84  28.28
```

```
#From the attributes of the dataset, we find :
df.columns = ['Relative_Compactness', 'Surface_Area', 'Wall_Area',
'Roof_Area', 'Overall_Height', 'Orientation', 'Glazing_Area',
'Glazing_Area_Distribution', 'Heating_Load','Cooling_Load']

df.head()
```

```
   Relative_Compactness  Surface_Area  Wall_Area  Roof_Area
Overall_Height  \
0                  0.98         514.5      294.0     110.25
7.0
1                  0.98         514.5      294.0     110.25
7.0
2                  0.98         514.5      294.0     110.25
7.0
3                  0.98         514.5      294.0     110.25
7.0
4                  0.90         563.5      318.5     122.50
7.0

   Orientation  Glazing_Area  Glazing_Area_Distribution  Heating_Load
\
0            2           0.0                          0         15.55

1            3           0.0                          0         15.55

2            4           0.0                          0         15.55
```

```
3              5           0.0                          0          15.55

4              2           0.0                          0          20.84
```

```
   Cooling_Load
0         21.33
1         21.33
2         21.33
3         21.33
4         28.28
```

Statistical Analysis

```
df.shape  #Number of rows and columns
```

```
(768, 10)
```

```
df.describe().T  #Tells the distribution of all columns
```

```
                           count        mean        std     min
25%  \
Relative_Compactness       768.0    0.764167   0.105777    0.62
0.6825
Surface_Area               768.0  671.708333  88.086116  514.50
606.3750
Wall_Area                  768.0  318.500000  43.626481  245.00
294.0000
Roof_Area                  768.0  176.604167  45.165950  110.25
140.8750
Overall_Height             768.0    5.250000   1.751140    3.50
3.5000
Orientation                768.0    3.500000   1.118763    2.00
2.7500
Glazing_Area               768.0    0.234375   0.133221    0.00
0.1000
Glazing_Area_Distribution  768.0    2.812500   1.550960    0.00
1.7500
Heating_Load               768.0   22.307195  10.090204    6.01
12.9925
Cooling_Load               768.0   24.587760   9.513306   10.90
15.6200

                             50%       75%     max
Relative_Compactness        0.75    0.8300    0.98
Surface_Area              673.75  741.1250  808.50
Wall_Area                 318.50  343.0000  416.50
Roof_Area                 183.75  220.5000  220.50
Overall_Height              5.25    7.0000    7.00
Orientation                 3.50    4.2500    5.00
Glazing_Area                0.25    0.4000    0.40
```

```
Glazing_Area_Distribution    3.00    4.0000    5.00
Heating_Load                18.95   31.6675   43.10
Cooling_Load                22.08   33.1325   48.03
```

```
df.dtypes  #Tells the types of data in the dataset
```

```
Relative_Compactness         float64
Surface_Area                 float64
Wall_Area                    float64
Roof_Area                    float64
Overall_Height               float64
Orientation                    int64
Glazing_Area                 float64
Glazing_Area_Distribution      int64
Heating_Load                 float64
Cooling_Load                 float64
dtype: object
```

```
df.nunique()  #Tells the number of unique values in each columns
```

```
Relative_Compactness          12
Surface_Area                  12
Wall_Area                      7
Roof_Area                      4
Overall_Height                 2
Orientation                    4
Glazing_Area                   4
Glazing_Area_Distribution      6
Heating_Load                 587
Cooling_Load                 636
dtype: int64
```

```
df.isnull().sum()   #Tells the number of Nan values in each columns
```

```
Relative_Compactness         0
Surface_Area                 0
Wall_Area                    0
Roof_Area                    0
Overall_Height               0
Orientation                  0
Glazing_Area                 0
Glazing_Area_Distribution    0
Heating_Load                 0
Cooling_Load                 0
dtype: int64
```

```
df.duplicated().sum() #tells the sum of duplicate values in the
columns
```

```
0
```

```
df.corr()  #Tells the co-relation between each columns
```

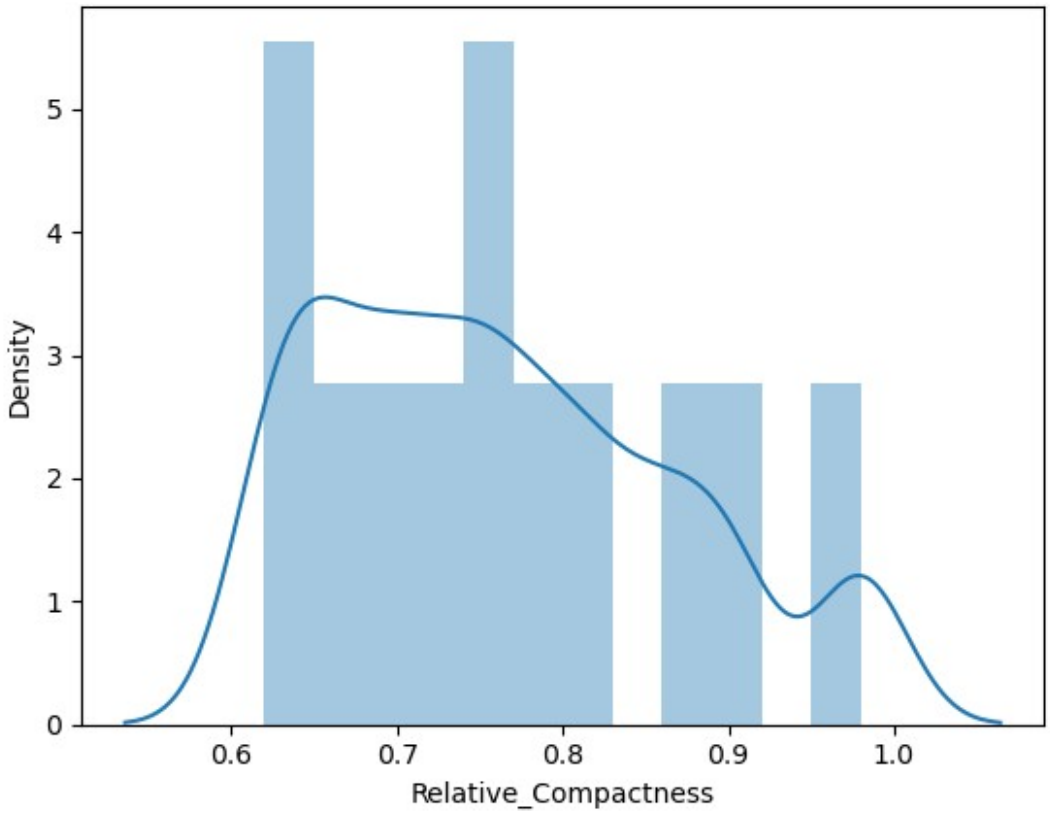
```
                           Relative_Compactness  Surface_Area
Wall_Area  \
Relative_Compactness               1.000000e+00 -9.919015e-01 -
2.037817e-01
Surface_Area                      -9.919015e-01  1.000000e+00
1.955016e-01
Wall_Area                         -2.037817e-01  1.955016e-01
1.000000e+00
Roof_Area                         -8.688234e-01  8.807195e-01 -
2.923165e-01
Overall_Height                     8.277473e-01 -8.581477e-01
2.809757e-01
Orientation                        4.678592e-17 -3.459372e-17 -
2.429499e-17
Glazing_Area                      -2.960552e-15  3.636925e-15 -
8.567455e-17
Glazing_Area_Distribution         -7.107006e-16  2.438409e-15
2.067384e-16
Heating_Load                       6.222719e-01 -6.581199e-01
4.556714e-01
Cooling_Load                       6.343391e-01 -6.729989e-01
4.271170e-01

                              Roof_Area  Overall_Height   Orientation
\
Relative_Compactness      -8.688234e-01    8.277473e-01  4.678592e-17

Surface_Area               8.807195e-01   -8.581477e-01 -3.459372e-17

Wall_Area                 -2.923165e-01    2.809757e-01 -2.429499e-17

Roof_Area                  1.000000e+00   -9.725122e-01 -5.830058e-17

Overall_Height            -9.725122e-01    1.000000e+00  4.492205e-17

Orientation               -5.830058e-17    4.492205e-17  1.000000e+00

Glazing_Area              -1.759011e-15    1.489134e-17 -9.406007e-16

Glazing_Area_Distribution -1.078071e-15   -2.920613e-17 -2.549352e-16

Heating_Load              -8.618281e-01    8.894305e-01 -2.586763e-03

Cooling_Load              -8.625466e-01    8.957852e-01  1.428960e-02


                           Glazing_Area  Glazing_Area_Distribution  \
Relative_Compactness      -2.960552e-15              -7.107006e-16
Surface_Area               3.636925e-15               2.438409e-15
```

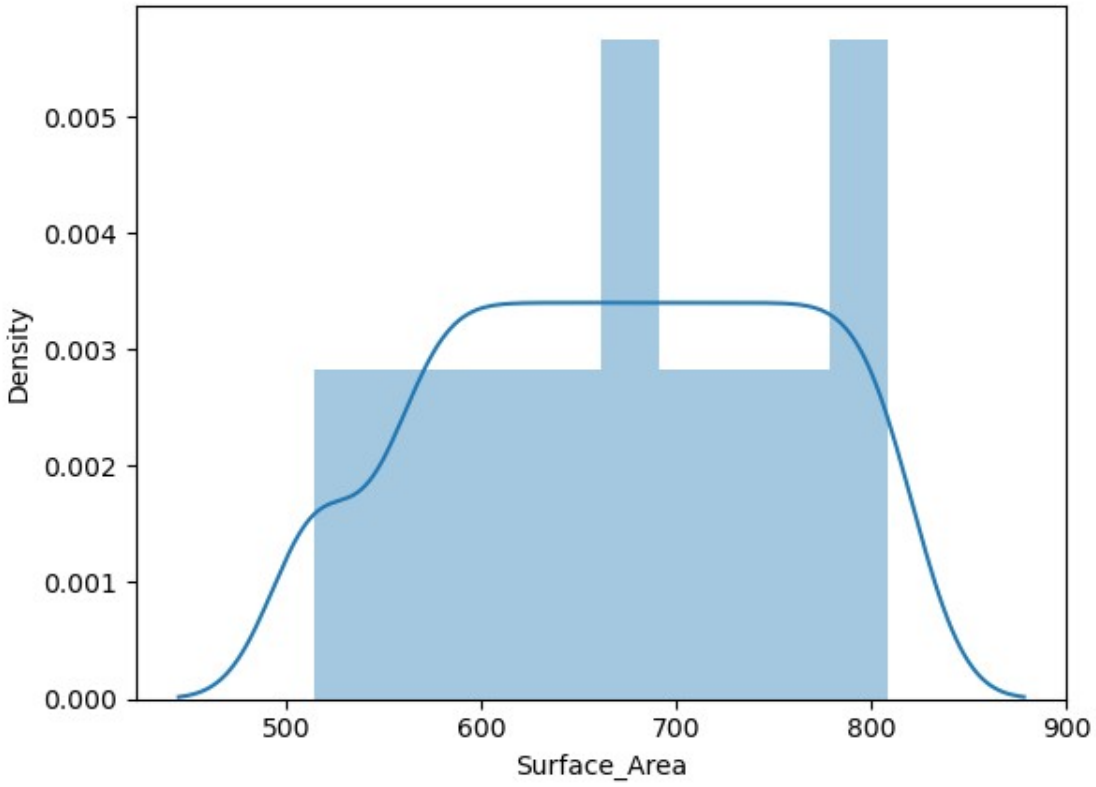
```
Wall_Area                 -8.567455e-17              2.067384e-16
Roof_Area                 -1.759011e-15             -1.078071e-15
Overall_Height             1.489134e-17             -2.920613e-17
Orientation               -9.406007e-16             -2.549352e-16
Glazing_Area               1.000000e+00              2.129642e-01
Glazing_Area_Distribution  2.129642e-01              1.000000e+00
Heating_Load               2.698417e-01              8.736846e-02
Cooling_Load               2.075050e-01              5.052512e-02

                          Heating_Load  Cooling_Load
Relative_Compactness          0.622272     0.634339
Surface_Area                 -0.658120    -0.672999
Wall_Area                     0.455671     0.427117
Roof_Area                    -0.861828    -0.862547
Overall_Height                0.889430     0.895785
Orientation                  -0.002587     0.014290
Glazing_Area                  0.269842     0.207505
Glazing_Area_Distribution     0.087368     0.050525
Heating_Load                  1.000000     0.975862
Cooling_Load                  0.975862     1.000000

df.cov()

                          Relative_Compactness  Surface_Area
Wall_Area  \
Relative_Compactness              1.118887e-02 -9.242069e+00 -
9.403911e-01
Surface_Area                     -9.242069e+00  7.759164e+03
7.512907e+02
Wall_Area                        -9.403911e-01  7.512907e+02
1.903270e+03
Roof_Area                        -4.150839e+00  3.503937e+03 -
5.759896e+02
Overall_Height                    1.533246e-01 -1.323703e+02
2.146545e+01
Orientation                      -1.447488e-19  7.411137e-16
0.000000e+00
Glazing_Area                     -1.085616e-19 -3.520290e-16 -
7.411137e-17
Glazing_Area_Distribution         1.375113e-18  8.522807e-15
0.000000e+00
Heating_Load                      6.641610e-01 -5.849415e+02
2.005866e+02
Cooling_Load                      6.383312e-01 -5.639665e+02
1.772672e+02

                             Roof_Area  Overall_Height   Orientation
\
Relative_Compactness      -4.150839e+00        0.153325 -1.447488e-19
```

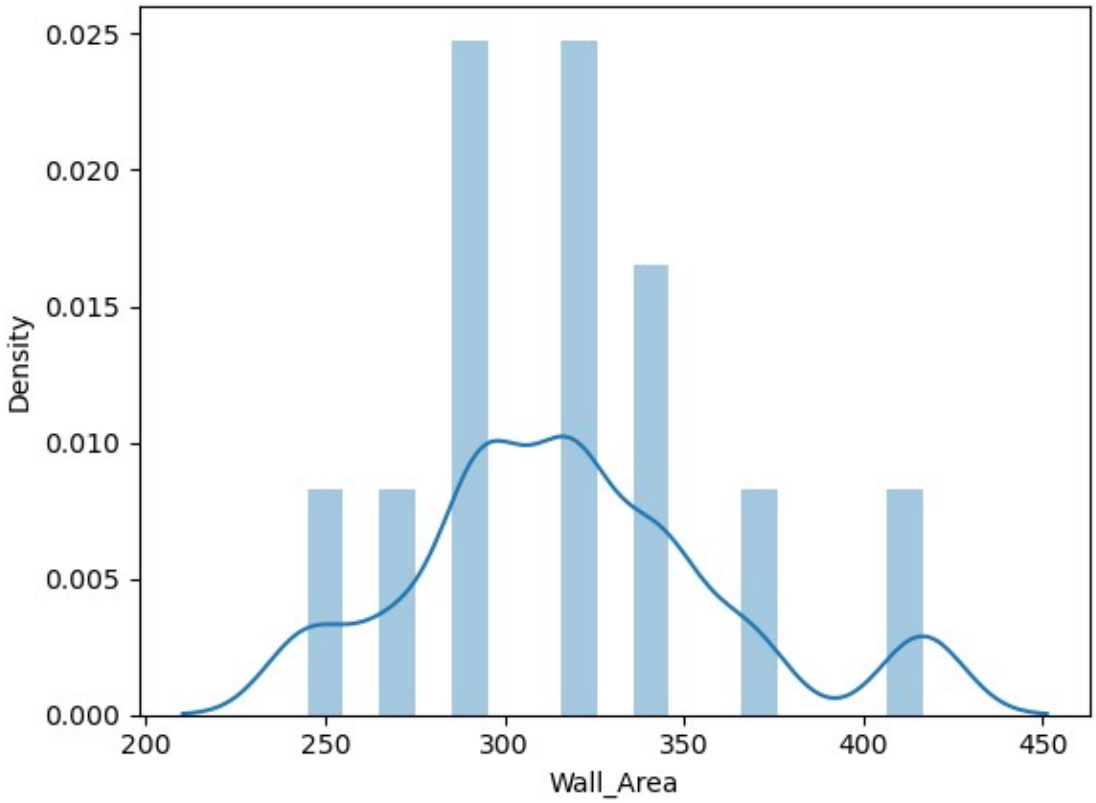
```
Surface_Area                3.503937e+03     -132.370274  7.411137e-16

Wall_Area                  -5.759896e+02       21.465450  0.000000e+00

Roof_Area                   2.039963e+03      -76.917862  0.000000e+00

Overall_Height             -7.691786e+01        3.066493  0.000000e+00

Orientation                 0.000000e+00        0.000000  1.251630e+00

Glazing_Area                3.659249e-16        0.000000  2.026483e-18

Glazing_Area_Distribution  -3.149733e-15        0.000000  0.000000e+00

Heating_Load               -3.927640e+02       15.715671 -2.920078e-02

Cooling_Load               -3.706169e+02       14.923005  1.520860e-01


                            Glazing_Area  Glazing_Area_Distribution  \
Relative_Compactness      -1.085616e-19               1.375113e-18
Surface_Area              -3.520290e-16               8.522807e-15
Wall_Area                 -7.411137e-17               0.000000e+00
Roof_Area                  3.659249e-16              -3.149733e-15
Overall_Height             0.000000e+00               0.000000e+00
Orientation                2.026483e-18               0.000000e+00
Glazing_Area               1.774772e-02               4.400261e-02
Glazing_Area_Distribution  4.400261e-02               2.405476e+00
Heating_Load               3.627273e-01               1.367273e+00
Cooling_Load               2.629852e-01               7.454857e-01

                            Heating_Load  Cooling_Load
Relative_Compactness            0.664161      0.638331
Surface_Area                 -584.941509   -563.966469
Wall_Area                     200.586579    177.267243
Roof_Area                    -392.764044   -370.616856
Overall_Height                 15.715671     14.923005
Orientation                    -0.029201      0.152086
Glazing_Area                    0.362727      0.262985
Glazing_Area_Distribution       1.367273      0.745486
Heating_Load                  101.812216     93.674133
Cooling_Load                   93.674133     90.502983

df.skew() #Tells the skewnws of each column

Relative_Compactness         0.495513
Surface_Area                -0.125131
Wall_Area                    0.533417
Roof_Area                   -0.162764
```

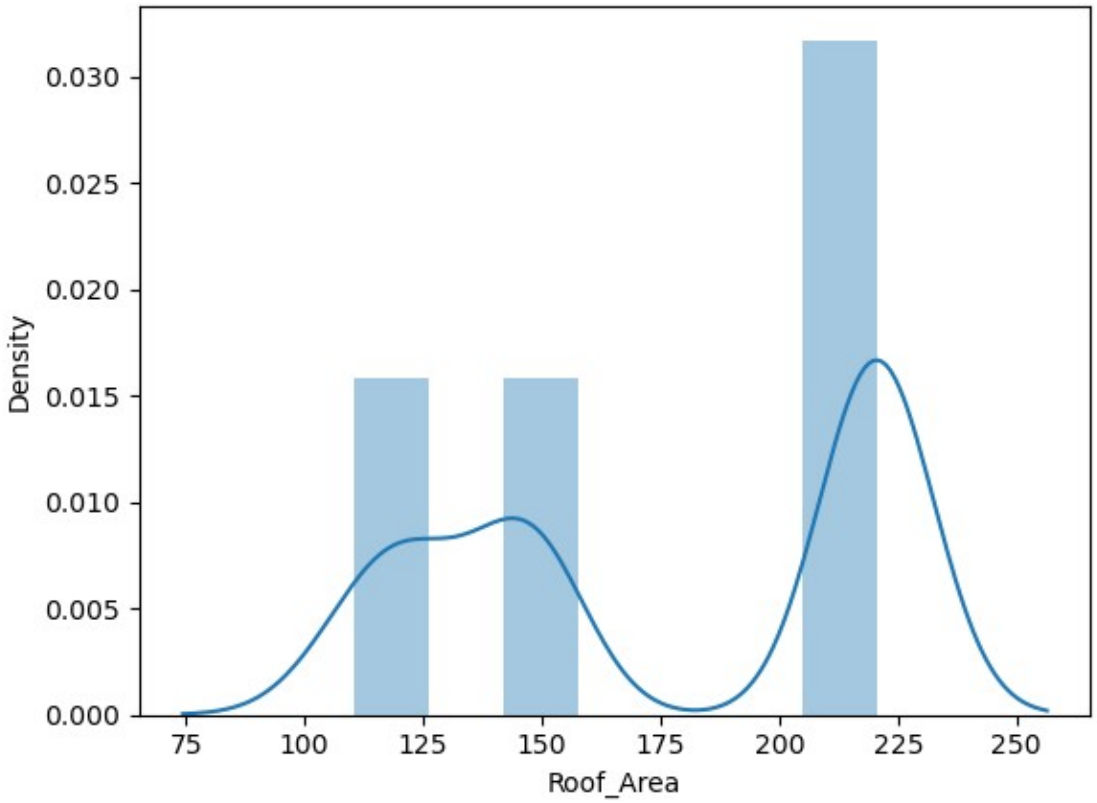
```
Overall_Height               0.000000
Orientation                  0.000000
Glazing_Area                -0.060254
Glazing_Area_Distribution   -0.088689
Heating_Load                 0.360446
Cooling_Load                 0.395992
dtype: float64
```

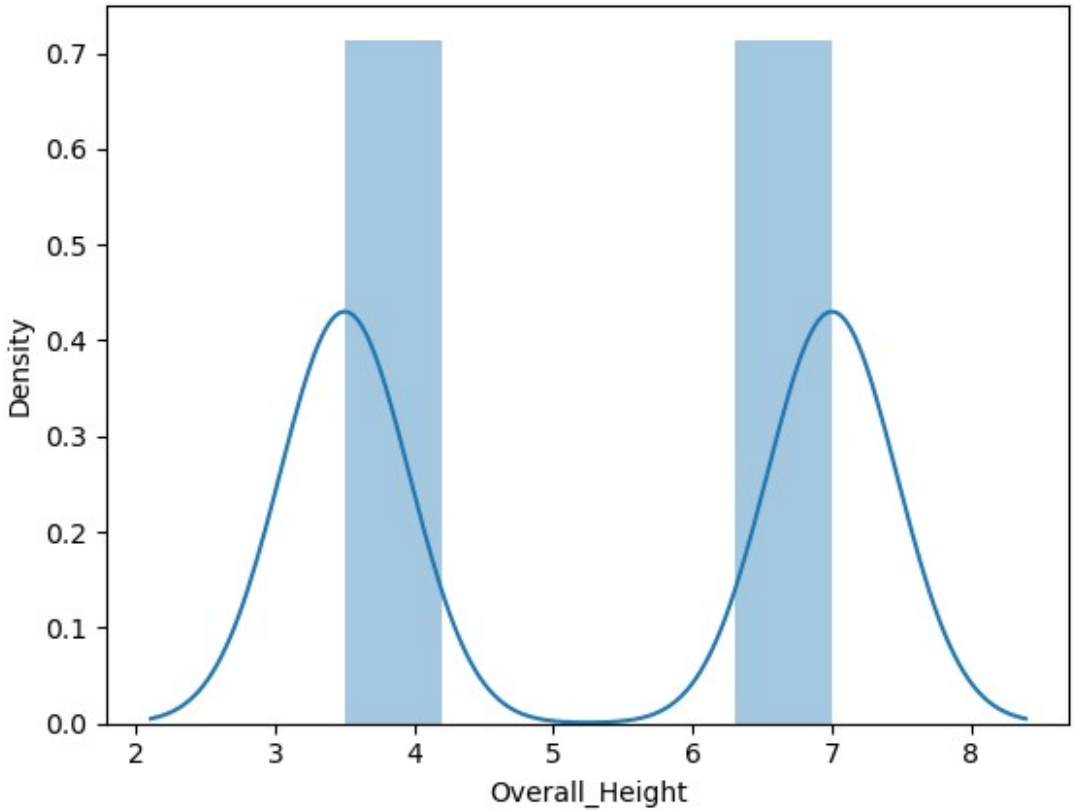
```
for i in df.columns:
    sns.distplot(df[i])     #Shows the distribution of all columns
    plt.show()
```

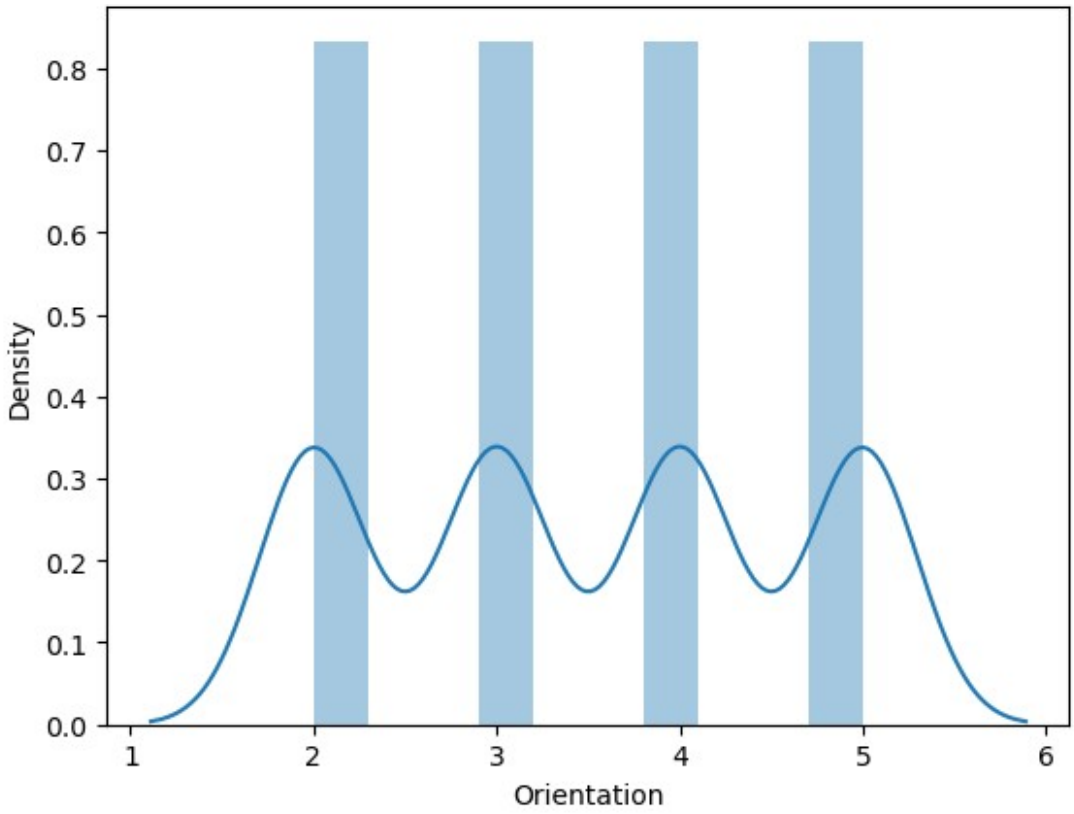


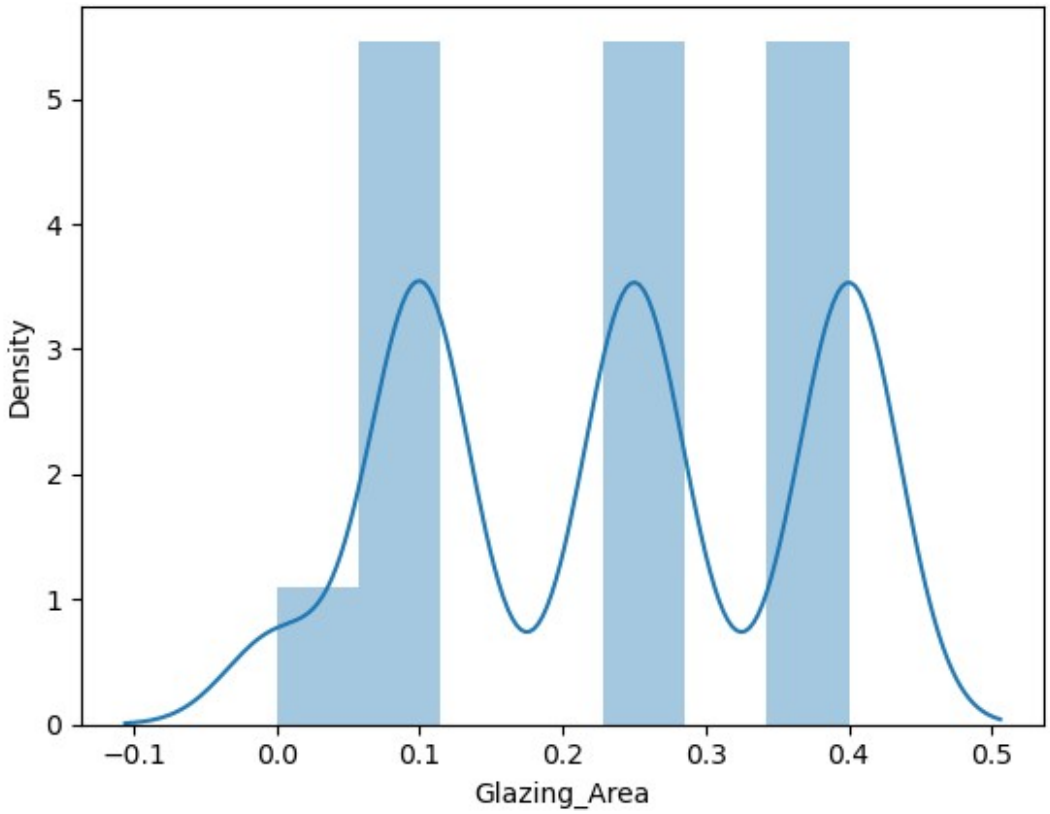

0.005
0.004
0.003
0.002
0.001
0.000
Density
500
600
700
800
900
Surface_Area

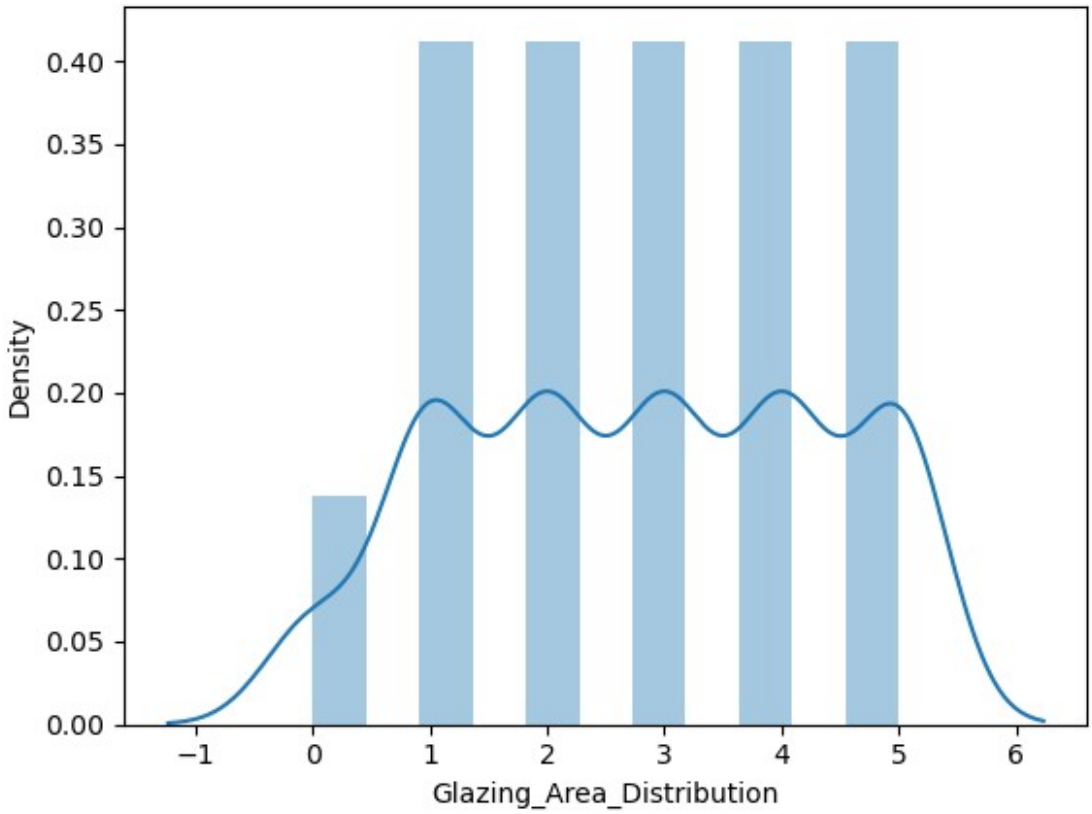

0.025
0.020
0.015
0.010
0.005
0.000
Density
200
250
300
350
400
450
Wall_Area

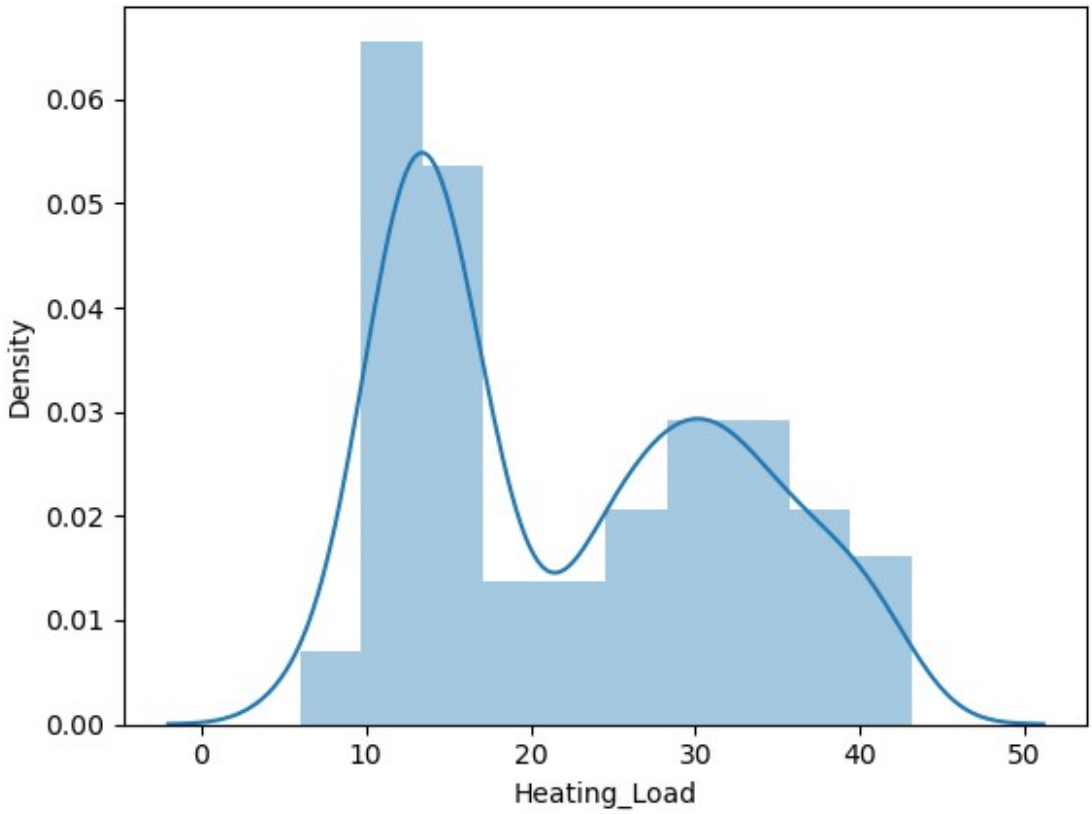

0.030
0.025
0.020
0.015
0.010
0.005
0.000
Density
75
100
125
150
175
200
225
250
Roof_Area

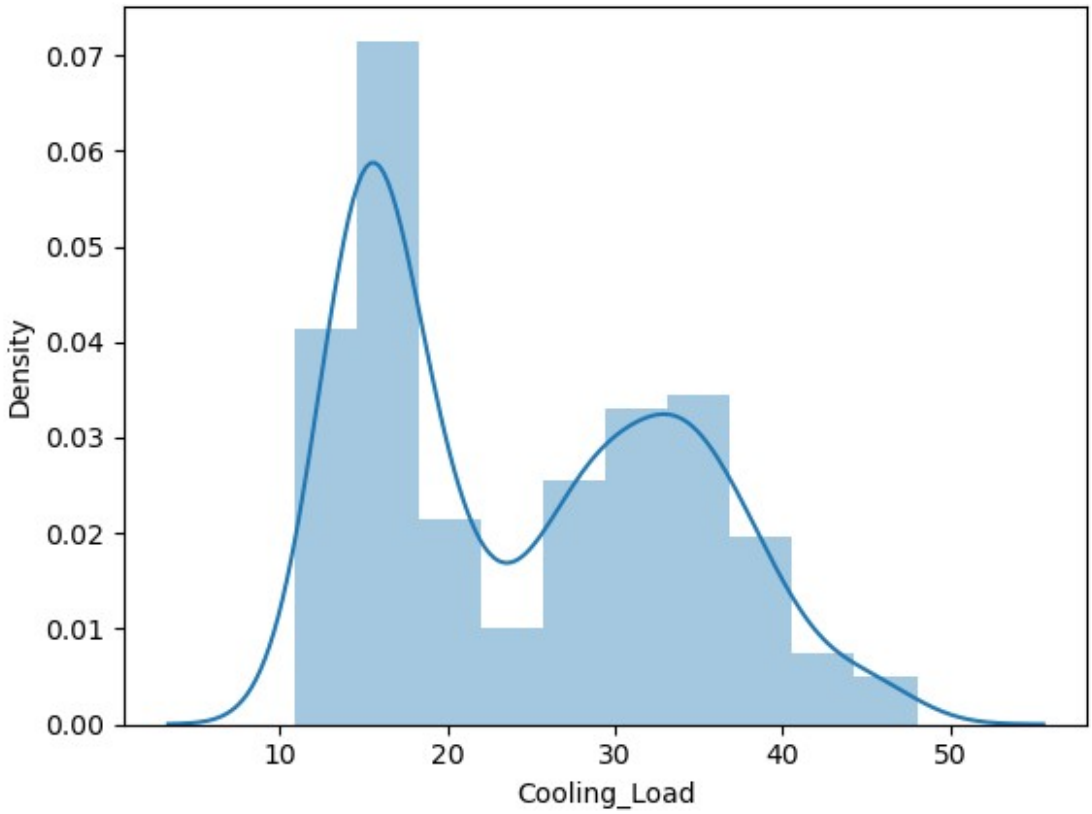

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Density
2
3
4
5
6
7
8
Overall_Height


Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1
2
3
4
5
6
Orientation


Density
0
1
2
3
4
5
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
Glazing_Area


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Density
−1
0
1
2
3
4
5
6
Glazing_Area_Distribution


0.06
0.05
0.04
0.03
0.02
0.01
0.00
Density
0
10
20
30
40
50
Heating_Load




```
df.groupby('Overall_Height').mean()  #groupby function to gain
insights from a categorical column.
```

```
                 Relative_Compactness  Surface_Area  Wall_Area
Roof_Area  \
Overall_Height

3.5                          0.676667    747.250000     306.25
220.500000
7.0                          0.851667    596.166667     330.75
132.708333

                 Orientation  Glazing_Area
Glazing_Area_Distribution  \
Overall_Height

3.5                      3.5      0.234375                      2.8125

7.0                      3.5      0.234375                      2.8125


                 Heating_Load  Cooling_Load
Overall_Height
```

```
3.5                13.338505     16.071432
7.0                31.275885     33.104089

df.groupby('Overall_Height').max()  #groupby function to gain insights
from a categorical column.

                Relative_Compactness  Surface_Area  Wall_Area
Roof_Area  \
Overall_Height

3.5                             0.74         808.5      367.5
220.5
7.0                             0.98         661.5      416.5
147.0

                Orientation  Glazing_Area
Glazing_Area_Distribution  \
Overall_Height

3.5                       5           0.4                          5

7.0                       5           0.4                          5


                Heating_Load  Cooling_Load
Overall_Height
3.5                    19.52         22.73
7.0                    43.10         48.03

df.groupby('Orientation').mean()  #groupby function to gain insights
from a categorical column.

             Relative_Compactness  Surface_Area  Wall_Area   Roof_Area
\
Orientation

2                        0.764167    671.708333      318.5  176.604167

3                        0.764167    671.708333      318.5  176.604167

4                        0.764167    671.708333      318.5  176.604167

5                        0.764167    671.708333      318.5  176.604167


             Overall_Height  Glazing_Area
Glazing_Area_Distribution  \
Orientation

2                      5.25      0.234375                       2.8125
```

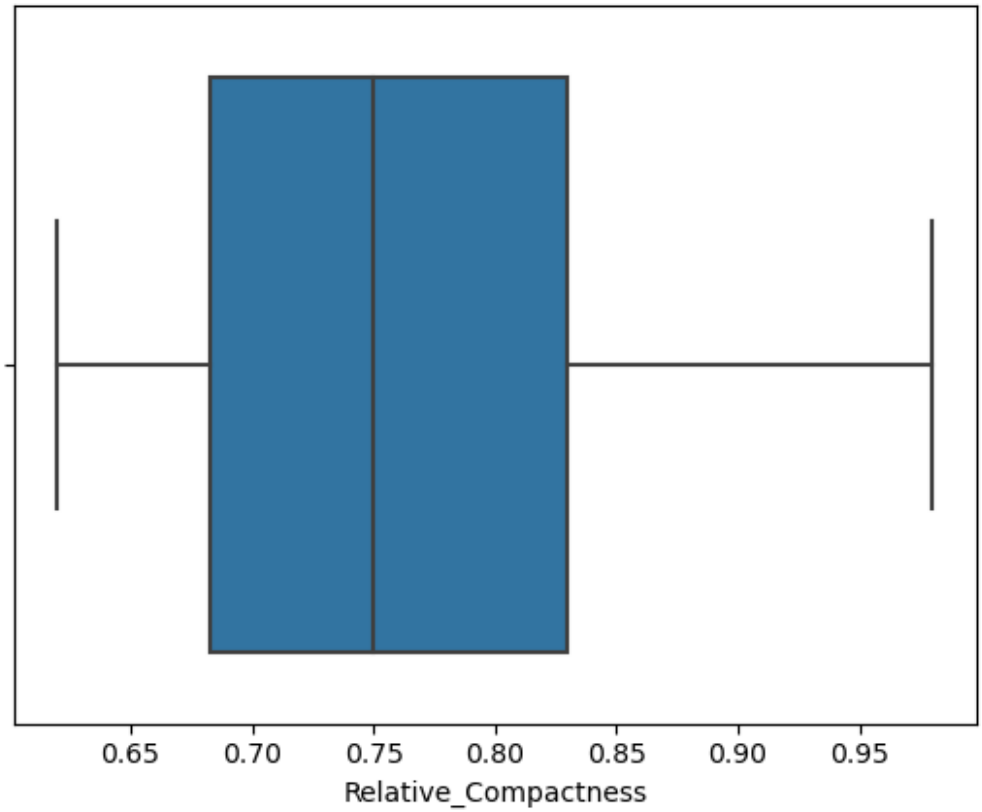
```
3                    5.25      0.234375                    2.8125

4                    5.25      0.234375                    2.8125

5                    5.25      0.234375                    2.8125
```

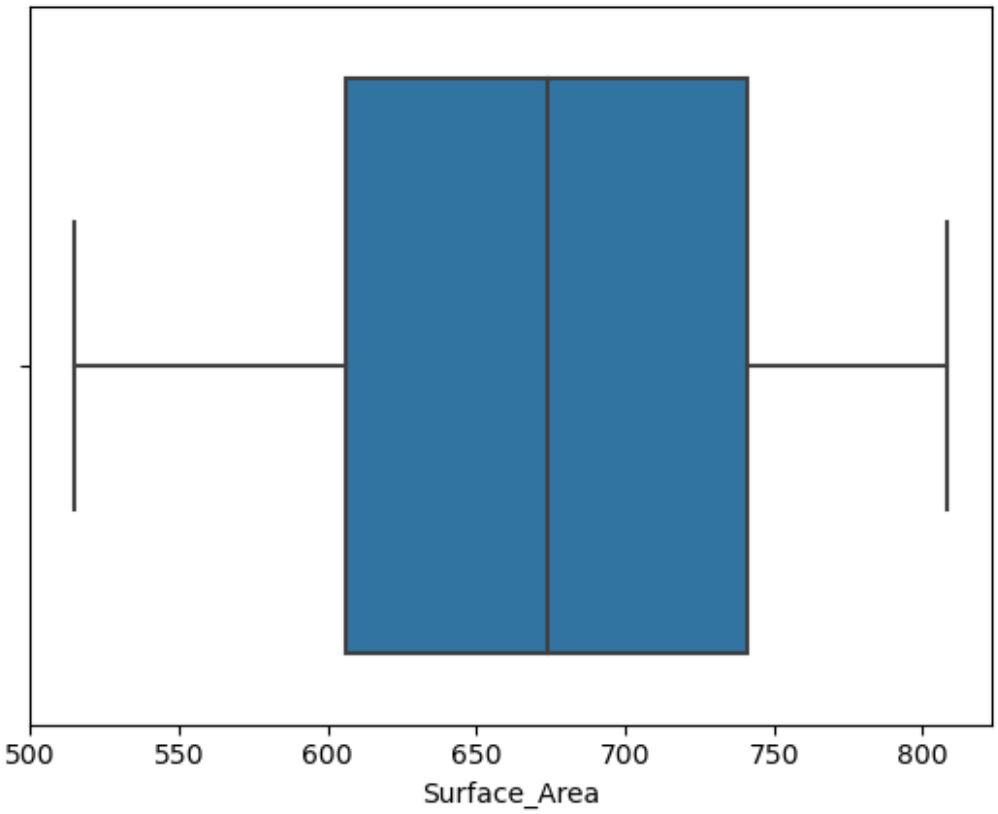
```
             Heating_Load  Cooling_Load
Orientation
2               22.312865     24.604531
3               22.380677     24.312552
4               22.259875     24.480313
5               22.275365     24.953646
```

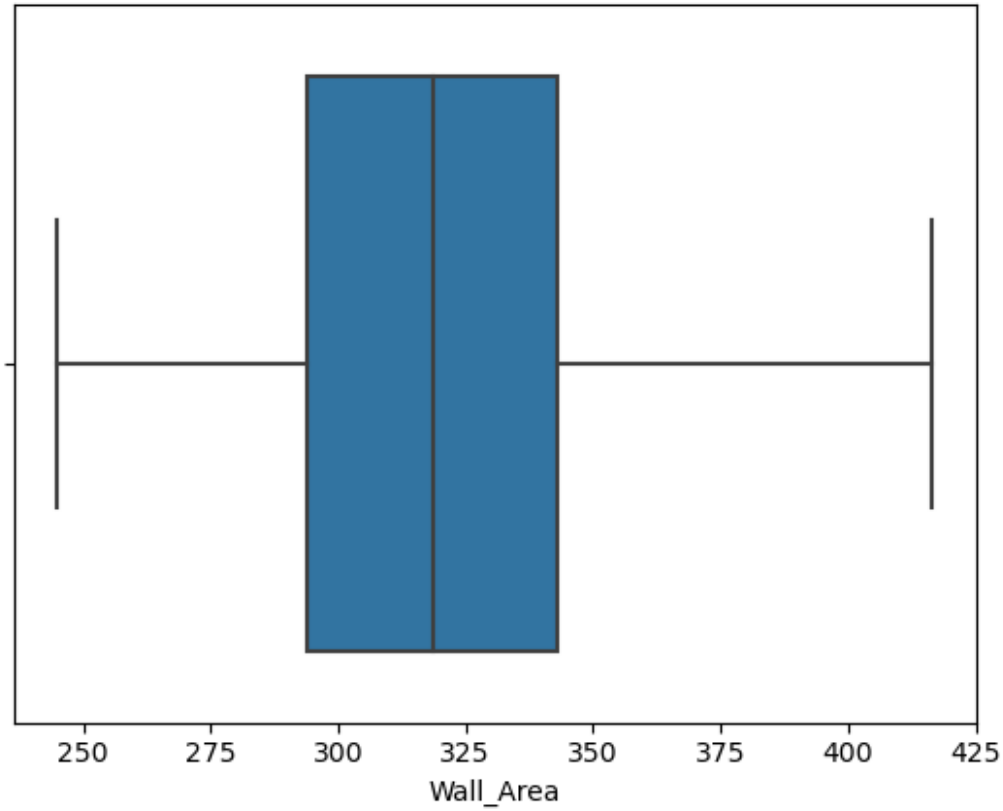
```
from scipy.stats import normaltest
normaltest(df['Overall_Height'])      #Pvalue
```

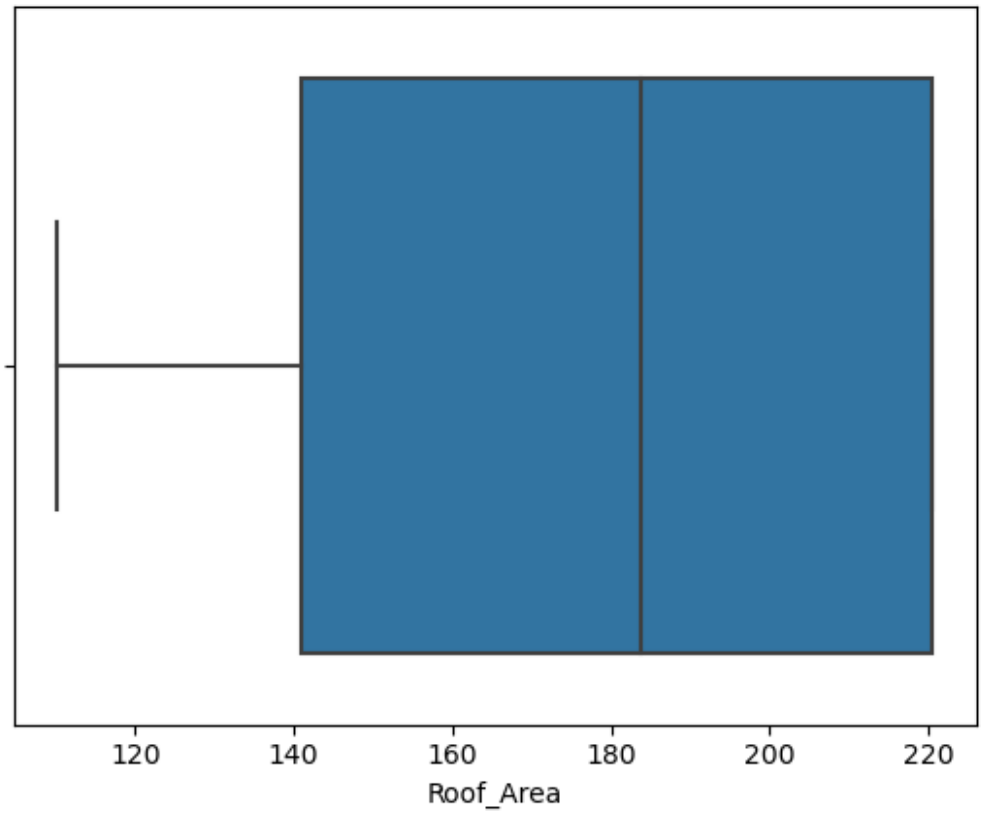
```
NormaltestResult(statistic=2977.1715748341653, pvalue=0.0)
```

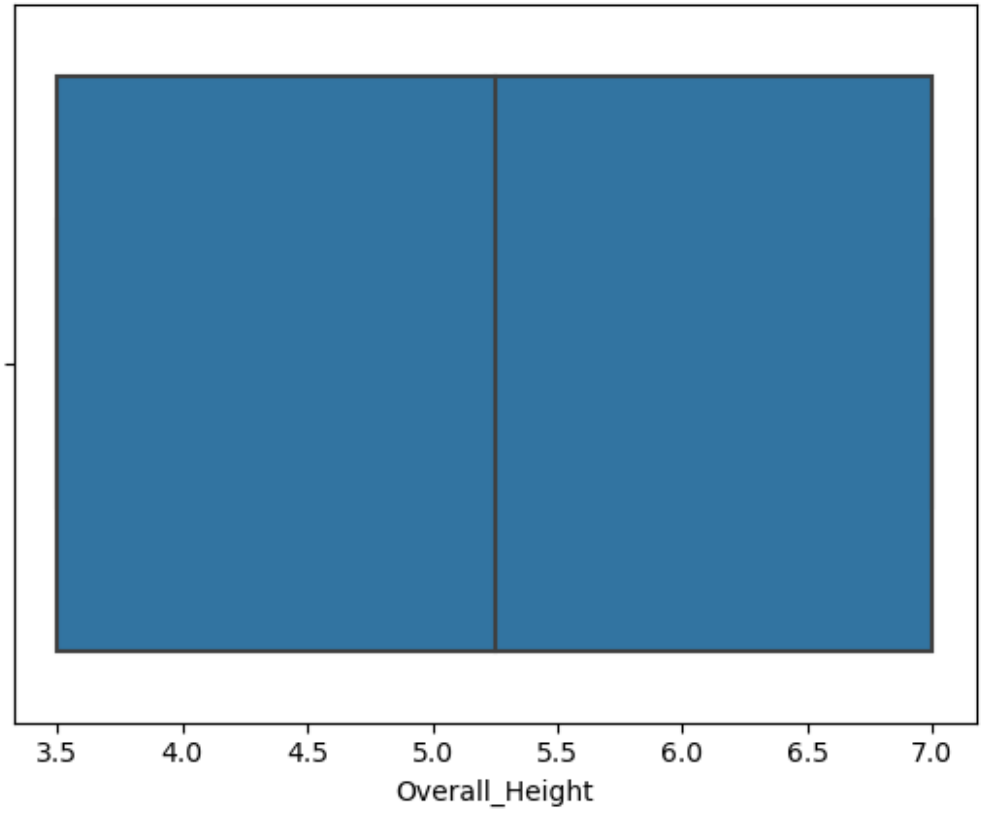
```
for i in df:
    sns.boxplot(df[i])         #Using the boxplot to check ooutliers
    plt.show()
```

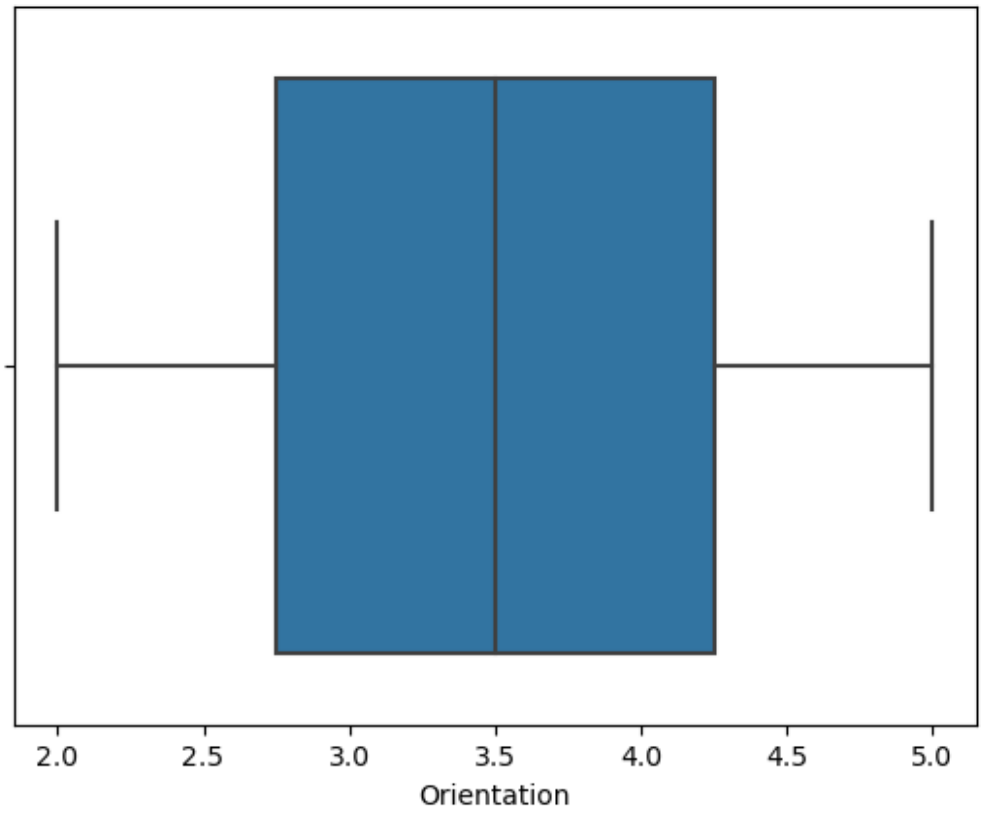


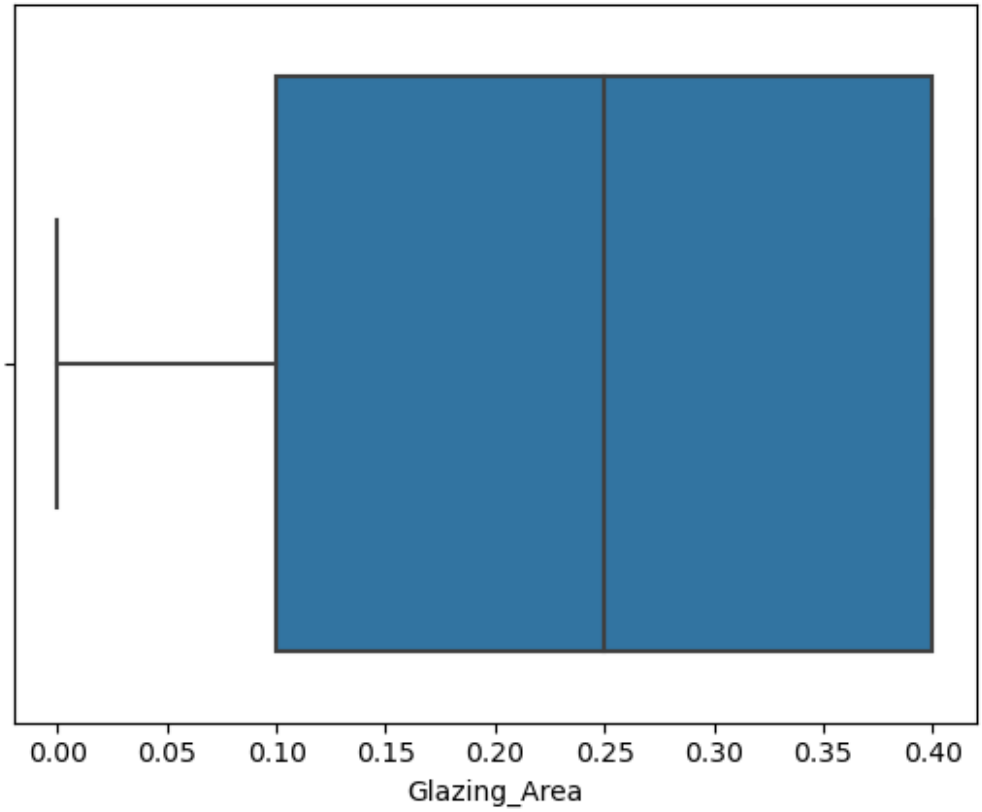

500
550
600
650
700
750
800
Surface_Area

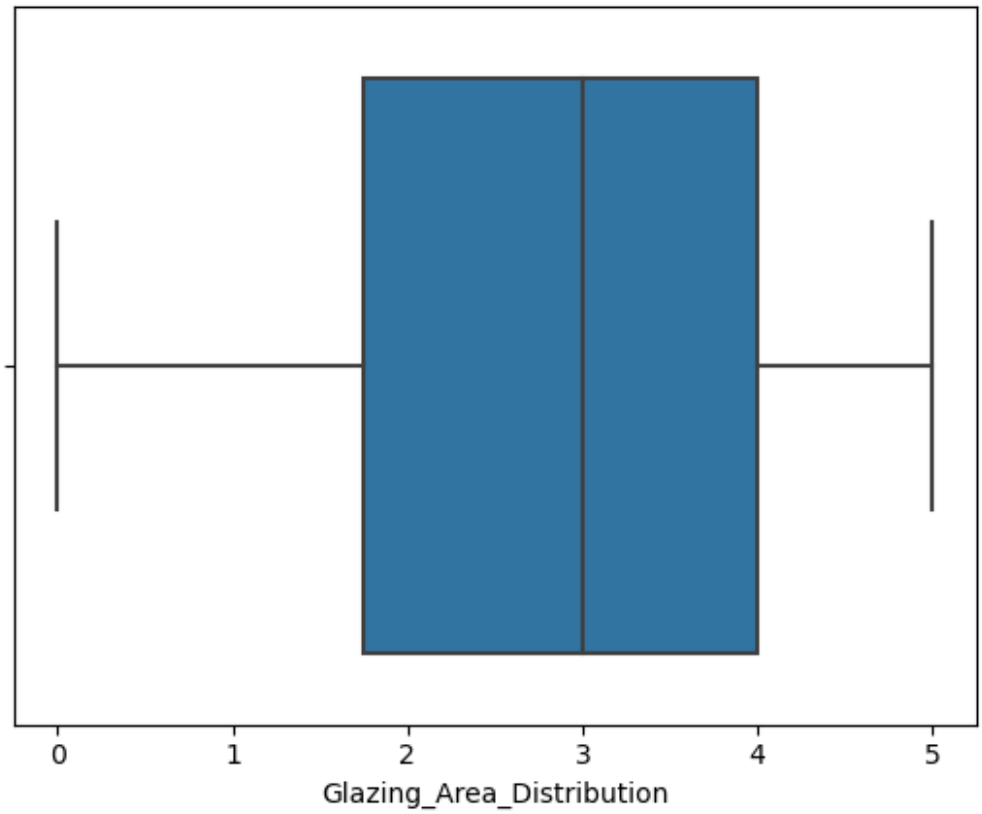

250
275
300
325
350
375
400
425
Wall_Area

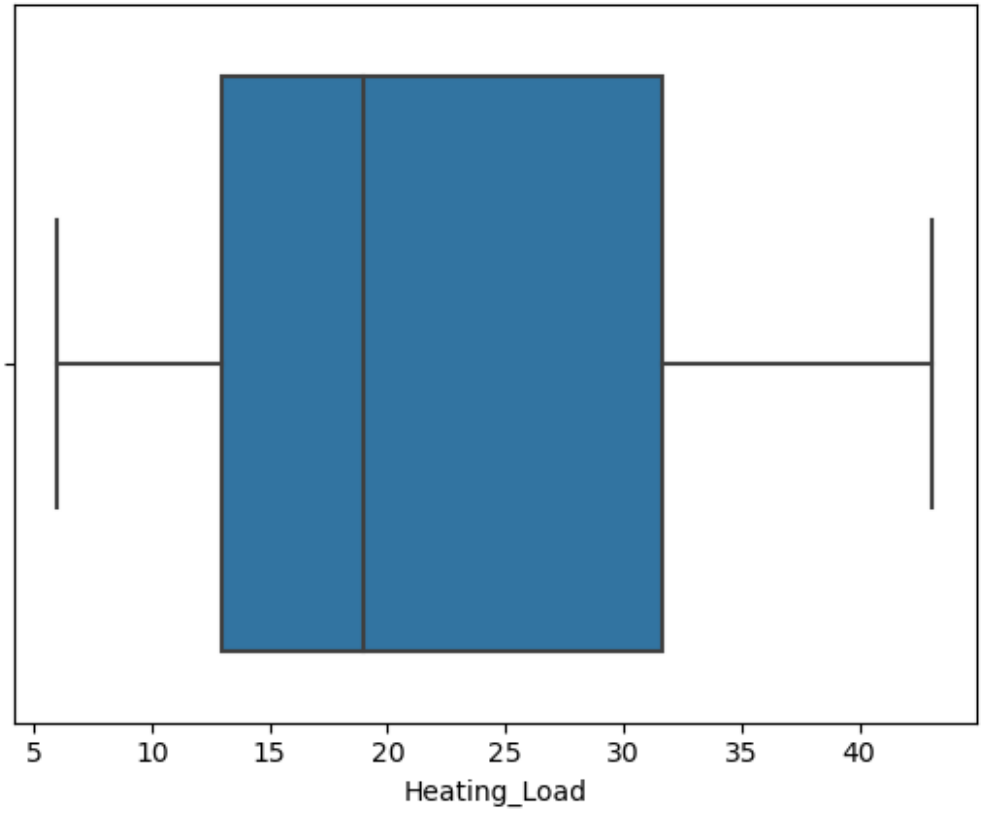

120
140
160
180
200
220
Roof_Area

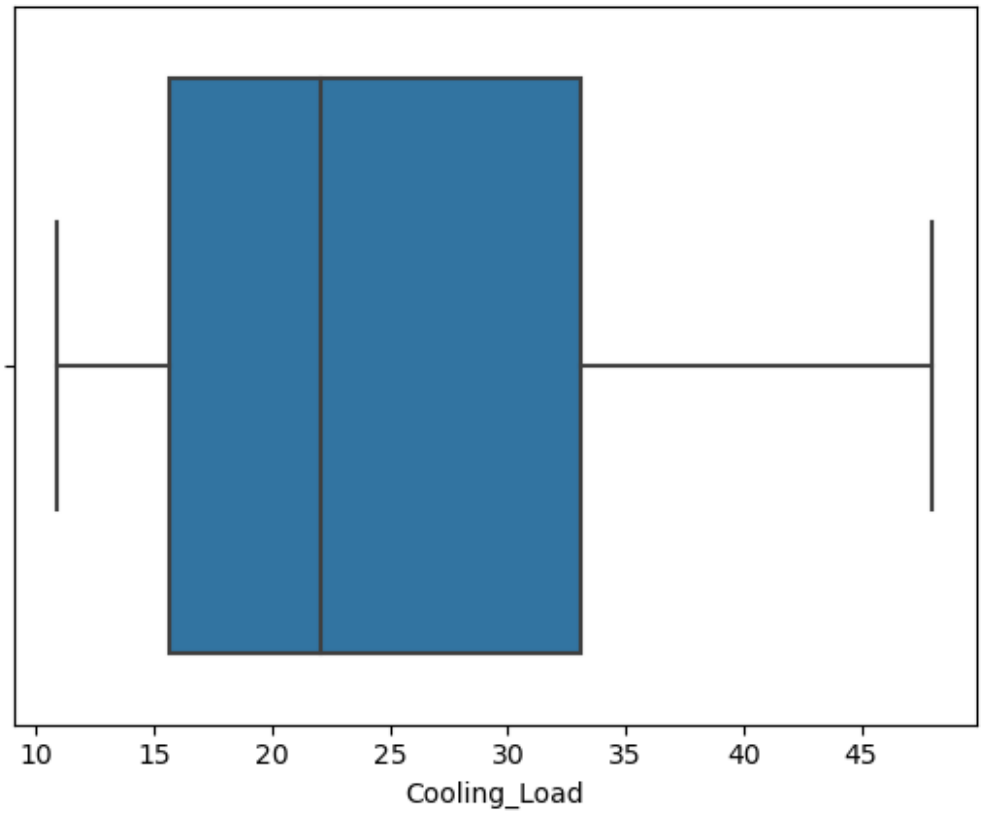

3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
Overall_Height


2.0
2.5
3.0
3.5
4.0
4.5
5.0
Orientation

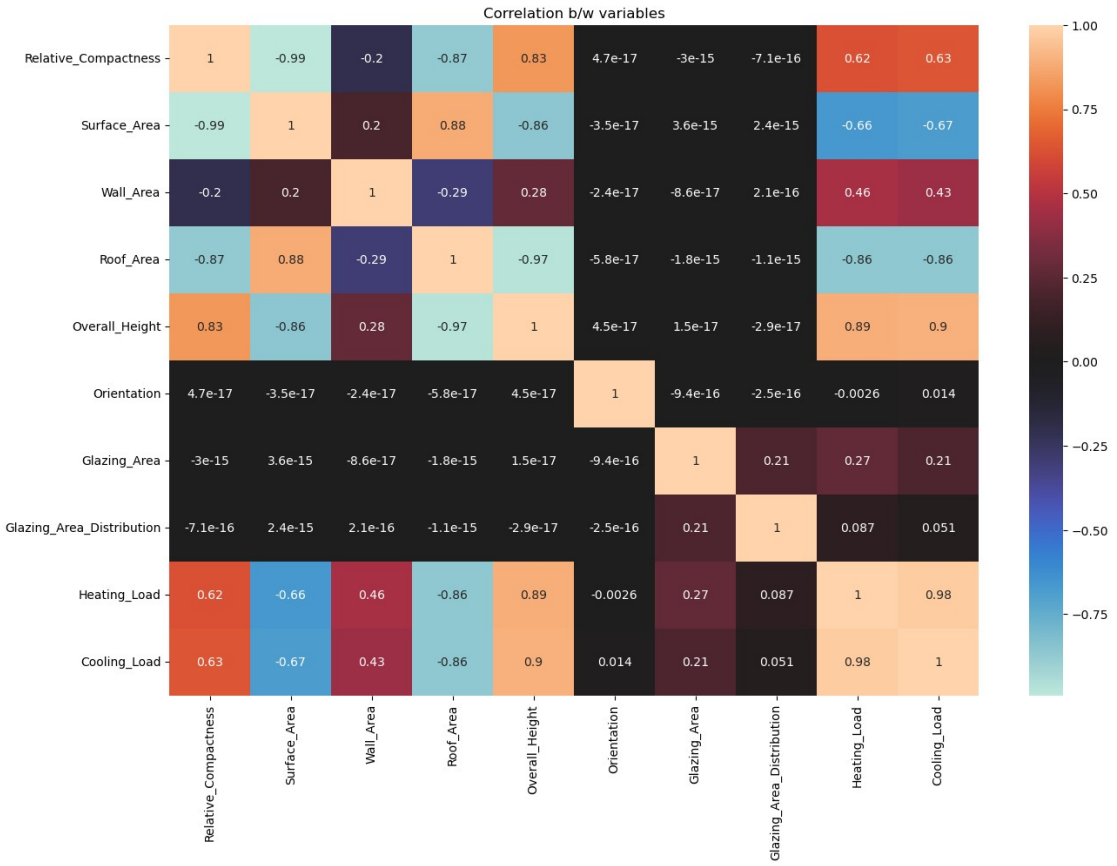

0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Glazing_Area


0
1
2
3
4
5
Glazing_Area_Distribution


5
10
15
20
25
30
35
40
Heating_Load

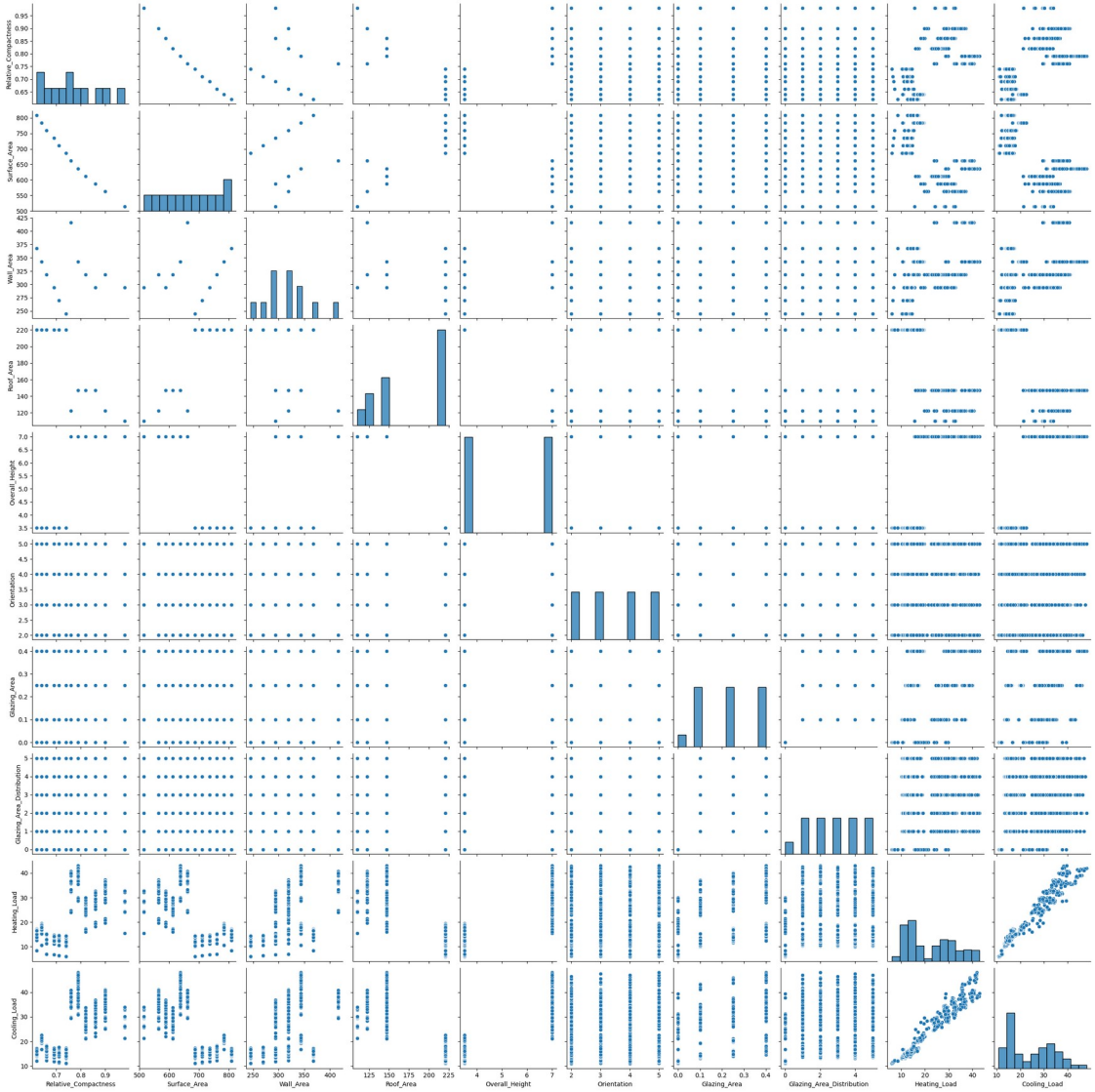



```
# Let's Perform outlier treatment
def outlier_treatment(feature_name):
    q1 = df[feature_name].quantile(0.25)
    q3 = df[feature_name].quantile(0.75)
    IQR = q3-q1
    higher_fence = q3 + 1.5 * IQR
    lower_fence = q1 - 1.5 * IQR
    df.loc[df[feature_name]> higher_fence, feature_name] =
higher_fence
    df.loc[df[feature_name]< lower_fence, feature_name] = lower_fence

for i in df.columns:
    outlier_treatment(i)

sns.heatmap(df.corr(), cmap = 'icefire', annot = True);
fig = plt.gcf()
fig.set_size_inches(15,10)
plt.title("Correlation b/w variables")
plt.show()
```




```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7fa41223a3d0>
```


Relative_Compactness
Surface_Area
Wall_Area
Roof_Area
Overall_Height
Orientation
Glazing_Area
Glazing_Area_Distribution
Heating_Load
Cooling_Load