



## Introduction

In the era of smart homes, the ability to predict energy consumption not only save money for users but also help in generating money for the user by giving excess energy back to Grid (in case of solar panels usage). In this case regression analysis will be used to predict Appliance energy usage based on data collected from various sensors. The energy prediction will come under supervised machine learning task aiming to Appliance energy consumption for a house based on factors like temperature, humidity & pressure. Many techniques, Gradient descent algorithm, and linear regression (in built function) have been applied to credit predict the energy consumption.



## **Dataset**

The dataset (Appliances Energy Prediction)

The data set is at 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. The dataset has 19735 rows and 29 variables include lights, date, Temperature and Humidity in various places in the house, pressure etc., The number of missing values and null values is zero. Number of weekdays is 14263 and weekend (Saturday and Sunday) is 5472.

## **Observations**



Al

Temperature columns - Temperature inside the house varies between 14.89 Deg & 29.85 Deg, temperature outside (T6) varies between -6.06 Deg to 28.29 Deg. The reason for this variation is sensors are kept outside the house

Humidiy columns - Humidity inside house varies is between 20.60% to 63.36% with exception of RH\_5 (Bathroom) and RH\_6 (Outside house) which varies between 29.82% to 96.32% and 1% to 99.9% respectively.

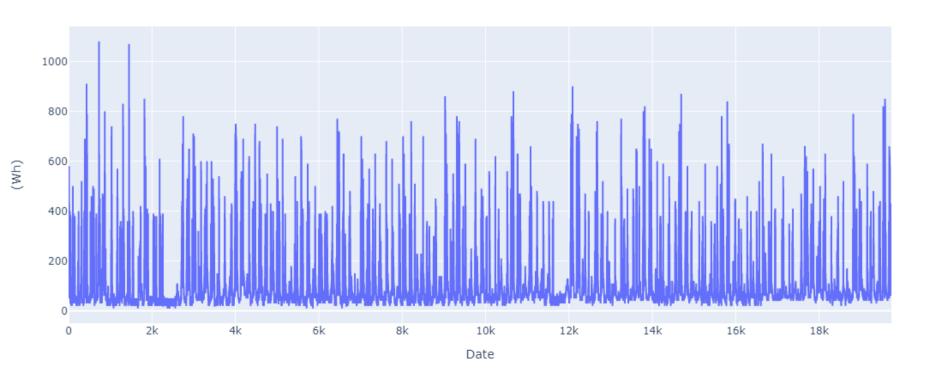
Appliances - 75% of Appliance consumption is less than 100 Wh . With the maximum consumption of 1080 Wh , there will be outliers in this column and there are small number of cases where consumption is very high

Lights column - Intially I believed lights column will be able to give useful information. With 11438 0 (zero) enteries in 14801 rows, this column will not add any value to the model. I believed light consumption along with humidity level in a room will give idea about human presence in the room and hence its impact on Appliance consumption. Hence for now, I will dropping this column

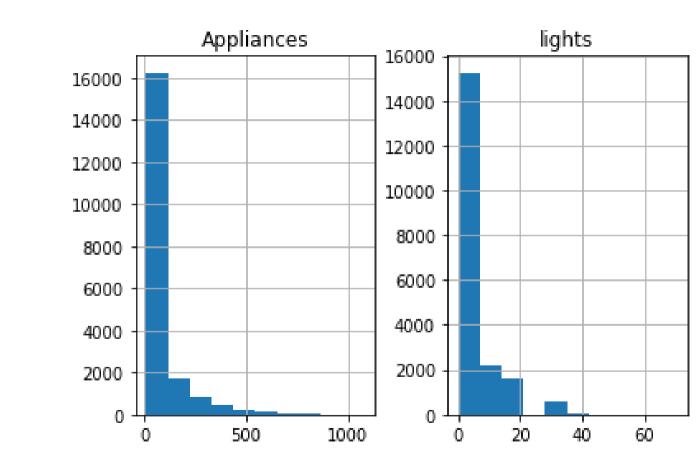


# **Energy Consumption:**

Appliance energy consumption measurement



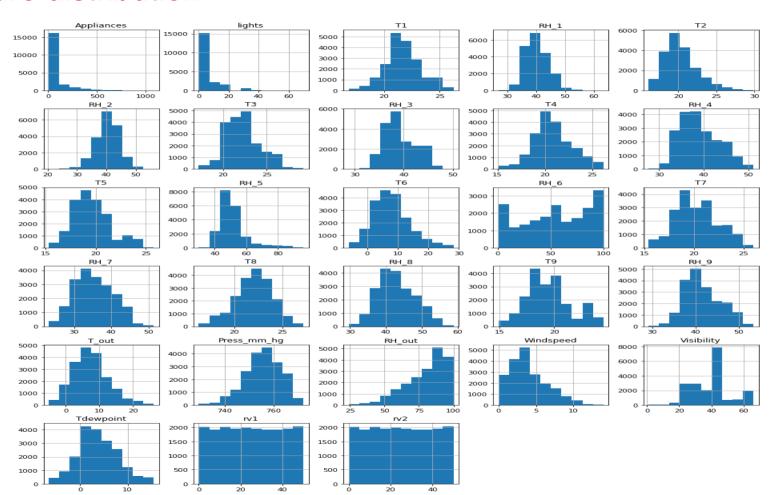


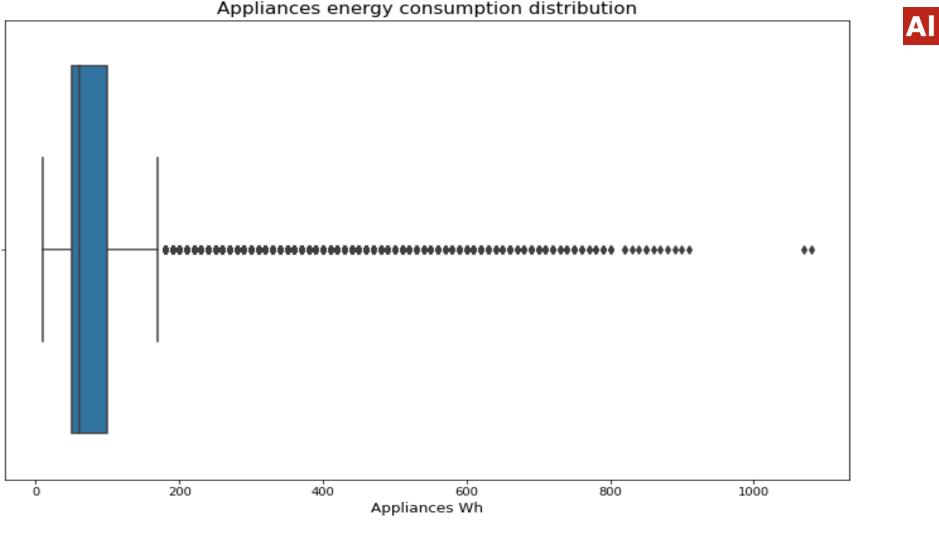


Appliance = lights

### Feature distribution







#### correlation of all the columns in the dataframe 2 0 0550 0860 12 0 060 0850 0360 040 0170 020 0070 120 0830 0260 0560 040 0940 010 0510 0950 0350 150 0807000 00010 0110 01 Appliances 0.0240.1D.005605D.0970.130.0089D110.0790.140.0790.15 0.140.0390.07D 0130.160 0088.0740 01D.0690.06 0.020.036000350003 T1 -0.055,024 1 0.16 0.84 0025,890 0290.880 0980.890 0150.65 0.62 0.84 0.14 0.83 00640,840 0720.68 0.15 0.350 0880 0760.570 0062006 0.8 0.25 0.84 0.11 0.88 0.21 0.3 0.32 0.250.021 0.03 0.74 0.12 0.76 0.34 0.29 0.27 0.2 0.02 0.64 0.007000 0.23 0.72 0.03 0.8 0.58 0.66 ( 23 0.580.069 0.68 0.16 0.79 0.13 0.510.0520.07 0.580.010.01 T2 -0.120.00560.84 0.27 1 0.17 0.74 0.12 0.76 0 1 0.14 0.68 0.047 0.72 0.11 0.25 0.009 0.39 0.05 0.69 0.04 0.68 0.05 0.68 0.03 40.26 0.58 0.069 0.05 0.05 0.068 0.06 - 0.8 T3 -0.0850.0970.89 0.25 0.74 0.14 1 0.0110.85 0.12 0.890.0660.69 0.65 0.85 0.17 0.8 0.044 0.9 0.13 0.7 0.19 0.28 0.1 0.1 0.65 0.052005 RH 3 -0.0360.130.0290.84 0.12 0.680.011 1 0.14 0.9 0.05 0.380.0770.51 0.25 0.83 0.28 0.83 0.2 0.83 0.12 0.23 0.36 0.260.017.0.40.00048004 T4 -0.040.0089.88 0.11 0.760.0470.85 0.14 1 0.0490.870.0760.65 0.7 0.880.044 0.8 0.0950.890.0260.660.0750.39 0.19 0.1 0.520.0018001 RH 4 -0.0170.110.0980.88 0.23 0.72 0.12 0.9 0.049 1 0.0920.35 0.26 0.39 0.13 0.89 0.17 0.850.0450.86 0.29 0.25 0.34 0.30 002 0.62 0018001 T5 -0.020.0790.89 0.21 0.72 0.11 0.89 0.05 0.870.092 1 0.0330.63 0.87 0.15 0.820.0160.910.072.0.65 0.17 0.27 0.150.0840.550 0.055005 - 0.6 RH 5 -0.0070.140.015 0.3 0.03 0.250.066 0.380.0760.350.033 1 0.0780.26 0.14 0.330.0870.36 0.14 0.270.0530.12 0.190.0820.0130.0780.010.01 0.80,00970.690,0770.65,0.26,0.630,078,1,0.67,0.62,0.26,0.480,0740.67,0.18,0.97,0.14,0.57,0.170,0810,760,0150,015 17 -0.0260.14 0.840.0210.660.0510.85 0.25 0.88 0.13 0.87 0.14 0.62 0.75 1 0.0340.88 0.21 0.940.078 0.630.098 0.41 0.19 0.11 0.47 0.039003 RH 7 -0.0560.0350.14 0.8 0.23 0.69 0.17 0.830.0440.89 0.15 0.33 0.26 0.360 0.34 1 0.12 0.880.0280.86 0.29 0.27 0.38 0.210.007 0.640 0018001 T8 -0.040.0710.83 0.03 0.580.041 0.8 0.28 0.8 0.17 0.820.0870.48 0.67 0.88 0.12 1 0.21 0.87 0.16 0.5 0.16 0.3 0.22 0.06 0.350.032 0. - 0.4 RH 8 -0.0940.018.006-0.740.0690.680.0440.830.0950.850.0160.360.0740.49 0.21 0.88 0.49 0.2 0.046 0.50.0045004 -0.01 0.16 0.84 0.12 0.680.055 0.9 0.2 0.890.0450.91 0.14 0.67 0.74 0.940.0280.87 0.11 1 0.008 0.67 0.16 0.32 0.18 0.1 0.580.00 II2001 RH 9 -0.050.0088.0720.76 0.16 0.68 0.13 0.830.0260.860.0720.27 0.18 0.390.0780.86 0.16 0.860.0087 1 0.22 0.18 0.36 0.240.008 0.54 0.0030.00 T out -0.0990.0740.68 0.34 79<mark>0.034 0.7 | 0.12 0.66 0.29 0.65 0.0530.97</mark> 0.64 0.63 0.29 | 0.5 | 0.12 0.67 0.22 Press mm hg -0.0350.0110.15 0.29 0.13 0.26 0.19 0.230.0750.25 0.17 0.12 0.140.066.0980.27 0.16 0.23 0.16 0.18 0.14 RH out -0.150,069 0.35 0.27 0.51 0.58 0.28 0.36 0.39 0.34 0.27 0.19 0.57 0.72 0.41 0.38 0.3 0.49 0.32 0.36 0.570 0.92 - 0.2 Windspeed 0.0870.060.088.0.2.0.0520.069.0.1.0.26.0.19.0.3.0.150.0820.170.0980.19.0.21.0.22.0.2.0.18.0.24.0.19.0.24.0.18 Visibility -00028 020 076 0210 07 00540 1 0.017 0.10 0026 0840 0130 0810 11 0.11 0.0720 060 046 0.10 0080 0770 040 088 0075 1 0.042 0059005 Tdewpoint -0.0150.036 0.57 0.64 0.58 0.5 0.65 0.41 0.52 0.62 0.590.078 0.76 0.26 0.47 0.64 0.39 0.5 0.58 0.54

Windspeed

Visibility Tdewpoint

Appliances lights



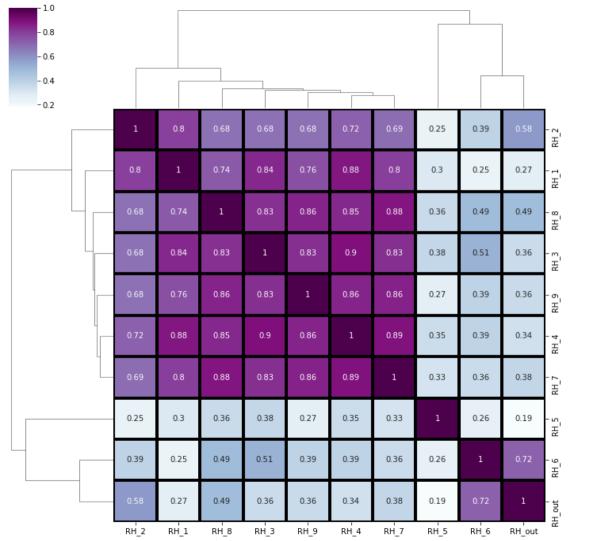
# Observations based on correlation plot

All the temperature variables from T1-T9 and T\_out have positive correlation with the target Appliances . For the indoortemperatures, the correlations are high as expected, since the ventilation is driven by the HRV unit and minimizes air tempera-ture differences between rooms. Four columns have a high degree of correlation with T9 - T3,T5,T7,T8 also T6 & T\_Out has high correlation (both temperatures from outside) . Hence T6 & T9 can be removed from training set as information provided by them can be provided by other fields.

Weather attributes - Visibility, Tdewpoint, Press\_mm\_hg have low correlation values

Humidity - There are no significantly high correlation cases (> 0.9) for humidity sensors.

Random variables have no role to play



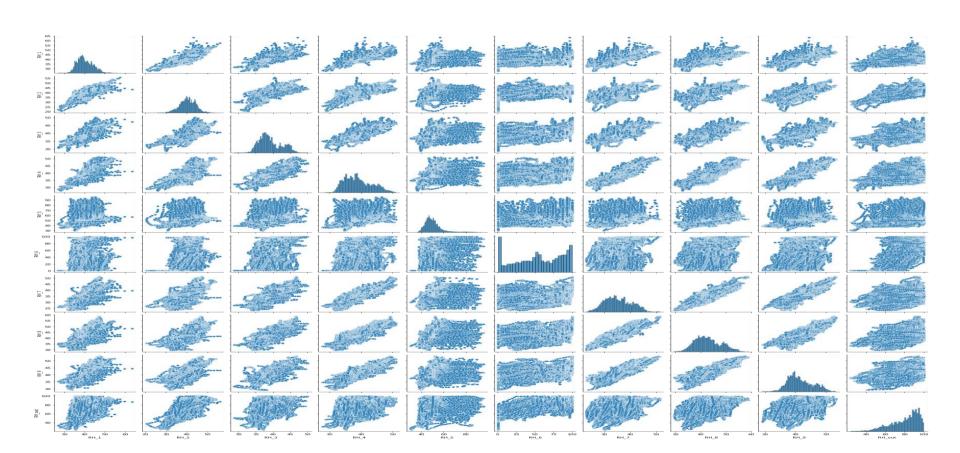


# Correlation matrix for humdity indoor and outdoor

RH\_1, Humidity in kitchen area, in % RH\_2, Humidity in living room area, in % RH\_3, Humidity in laundry room area, in % RH\_4, Humidity in office room, in % RH\_5, Humidity in bathroom, in % RH\_6, Humidity outside the building (north side), in % RH\_7, Humidity in ironing room, in % RH\_8, Humidity in teenager room 2, in % RH\_9, Humidity in parents room, in % RH\_out, Humidity outside (from Chievres weather station), in %

# Pariplot for humidity







## Conclusion:

This analysis shows that how much impact weather conditions have in our day-today energy consumption in home. Whenever the Atomospheric pressure is low or high, there's a evident change in the way enegry is consumed

