Revised Report

Team Member: Mengxi Wang, Rui Chen, Jialiang Shi

Github: https://github.com/markshibu/Mapping-the-Intensity-of-Energy-Use-Across-BU

1 Project Description

1.1 Topic and Goals

The goal of this project is to analyze BU's building energy usage intensity (EUI) and discover relationships between driving factors to help the University achieve its goal of becoming carbon neutral by 2040. Ideally, the analysis would help inform decisions made by sustainability@BU, the Climate Action Plan and the buildings team of Carbon Free Boston.

1.2 Main Questions

The main question we are trying to answer is "How does BU's building energy use intensity vary with property type/year built/ temperature?".

Several specific questions are raised to help answering the major question.:

1. What is the average EUI for each property type building in 2017?
2. How (much) does the building's built-year affect average EUI for each property type of building?
3. How (much) does structure class affect EUI of buildings?
4. How do buildings with different property type vary with temperature?
5. How do some variables affect EUI of multi-family/office building?

1.3 Main Methods

The methods we used includes data scraping, classification, linear regression, probability and statistics.

In order to achieve the results, we have done the following things:

1. Clean the energy consumption data among 2015-2017
2. For building type /year of built /building structure: we use the data from BERDO to analyze.
3. For temperature: we use extra data from datasets like Cooling Degree Days (CDD) 2015-2017, Heating Degree Days (HDD) 2015-2017 and use linear regression to analyze the relationship between temperature and EUI.

1.4 Results

Up to now, we integrated two datasets - Building Energy Reporting and Disclosure Ordinance of Boston(BERDO) 2015-2017, and Property Assessment of Boston 2015-2017; analyzed how (much) property type/year built/temperature affect EUI; made some effort on predicting EUI of a building with given information.

Most of the results shows what we were expecting, although part of them won't give much interesting conclusion.

2 Data Description

2.1 Datasets Source

The datasets we are using are retrieved or downloaded directly from the government official websites.

2.2 Datasets Combination

BERDO has our core attribute of each building - EUI. While, we may also need several other attributes in different datasets to find out how they are related. So, when processing the data, we combined two datasets using the address in BERDO to find the matching building in Property Assessment dataset, and generated a single .csv file which have all attributes we need for each building. However, in this process, the total number of buildings dropped from around 1800 to 746, since most of the buildings don't match with the other one. However this may not cause severe consequences due to the scarcity of data, because we didn't do much of our analysis solely upon the integrated dataset.

2.3 Outliers and Missing Data

When cleaning BERDO 2015-2017, we detected and removed most of outliers based on our observation. We made this tentative conclusion that, "Normally, EUI should never be over 800 kBTU/sf". This can be proved by listing all the buildings of same property type, and if only one or two of the data would be at least 10 times greater then others, then is obviously a contaminated data.

The strategy dealing with missing values or inconsistent entries is simply remove it. The percentage of this process would normally cost less then 1% of all data.

3 Data Analysis & Conclusions

3.1 Average EUI for each property type building in 2017

3.1.1 Data processing

1) Datasets:

- BERDO_2017
- Property Assessment

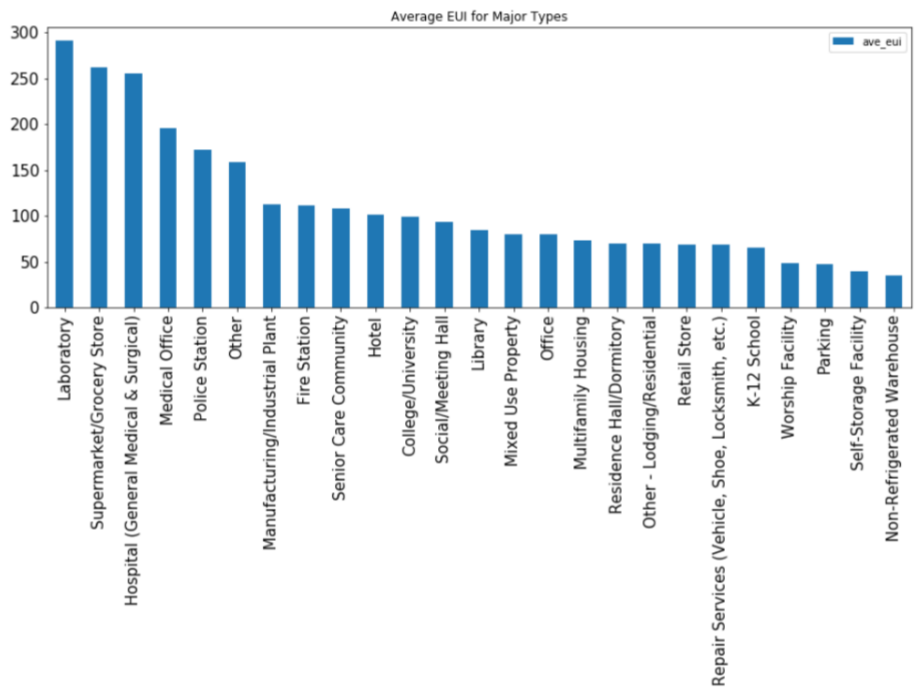
2) Remove the unavailable EUI. The total number of building data reduced from 1800 to 1664. Detect and Remove Outliers

3) Count the top most property types of buildings.

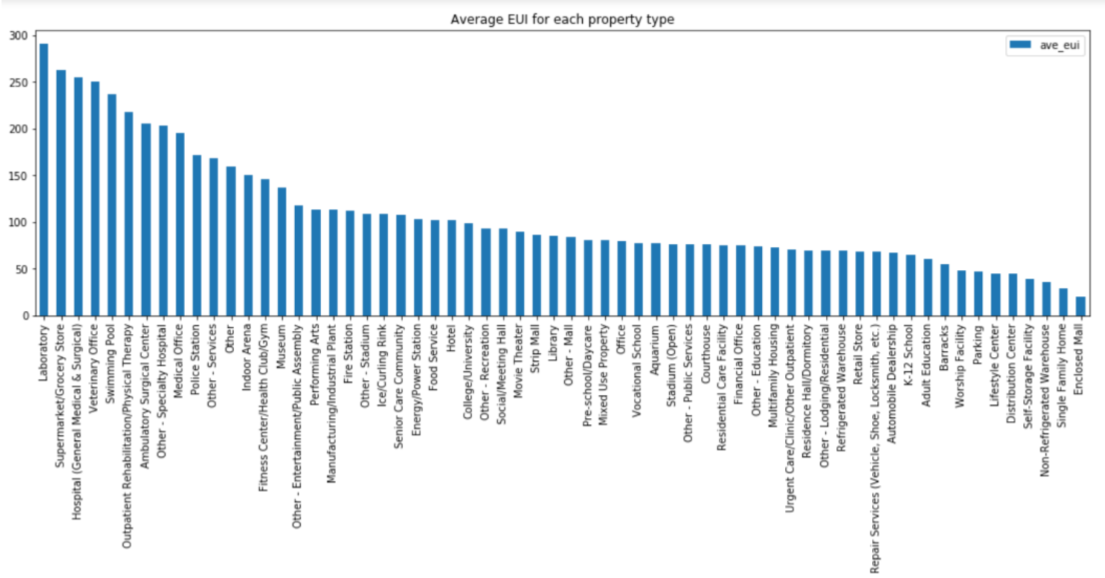
Property Type	
Multifamily Housing	548
Office	297
K-12 School	142
College/University	72
Hotel	47
Mixed Use Property	45
Residence Hall/Dormitory	40
Fire Station	40
Laboratory	34
Library	27
Other	25
Police Station	22
Other - Lodging/Residential	22
Medical Office	21

Img1: property types

4) Average EUI for major types of building (>10 samples)



Img2: Average EUI for major types of building



Img3: Average EUI for all types of building

3.1.2 Data analysis

From plots above, we have following conclusions:

1. Among all major types, laboratory, supermarket/grocery store and hospital (general medical & surgical) are top3 buildings that consume more energy than other buildings;
2. Hypothetical conclusion:

1) For laboratories (top1): as we all know, there are many universities and colleges in Boston, so it's obvious that there are many laboratories in Boston city.

    a. Some laboratories may run all day and night even during holidays, which consume much more energy than other buildings;

    b. Some laboratories may be used to conduct energy-consuming experiences like experiences related to electricity.

2) For supermarket/grocery store (top2): supermarkets or grocery stores are distributed everywhere in Boston city, and they have regular hours of operation.

    a. Supermarkets or grocery stores are usually open every day in daily life, even during holidays, some even open round-the-clock.

    b. Whenever a supermarket is open, it is consuming energy in every aspects, for example lighting, refrigeration, electronic devices and so forth.

3) For hospital (general medical & surgical) (top3): hospitals are expected to consume more energy than other building types, because it should be open round-the-clock even the emergency like power outage around the whole city.

Not to mention the energy consumption from lighting, but the medical devices consume much more energy than other devices.

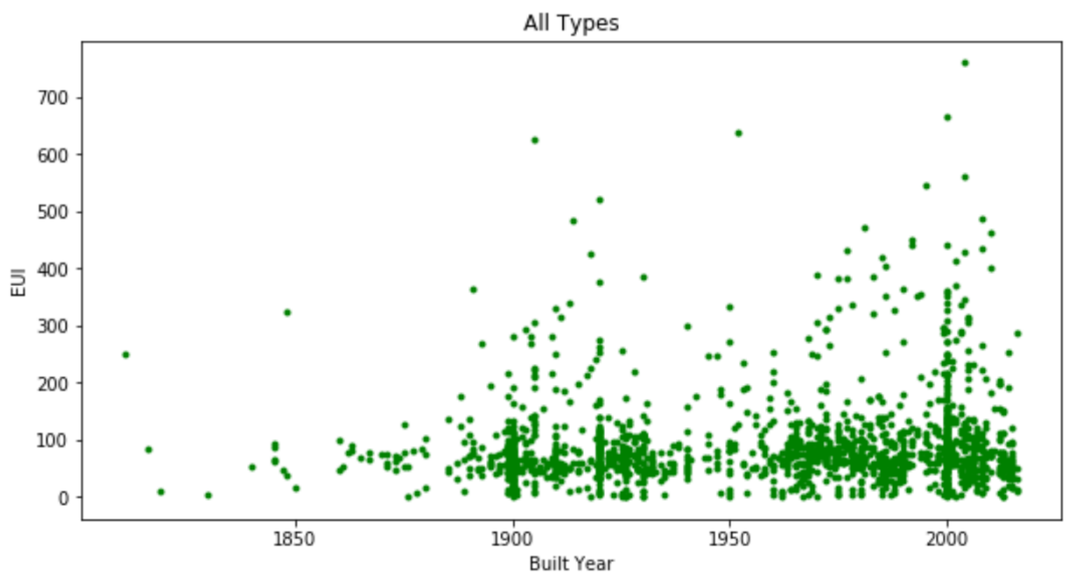
3.2 How (much) does the building's built year affect average EUI for each property type of building?

3.2.1 Data processing

1) Datasets:

   - BERDO_2017
   - Property Assessment

2) How (much) does average EUI vary with the built year of **all** types of buildings?

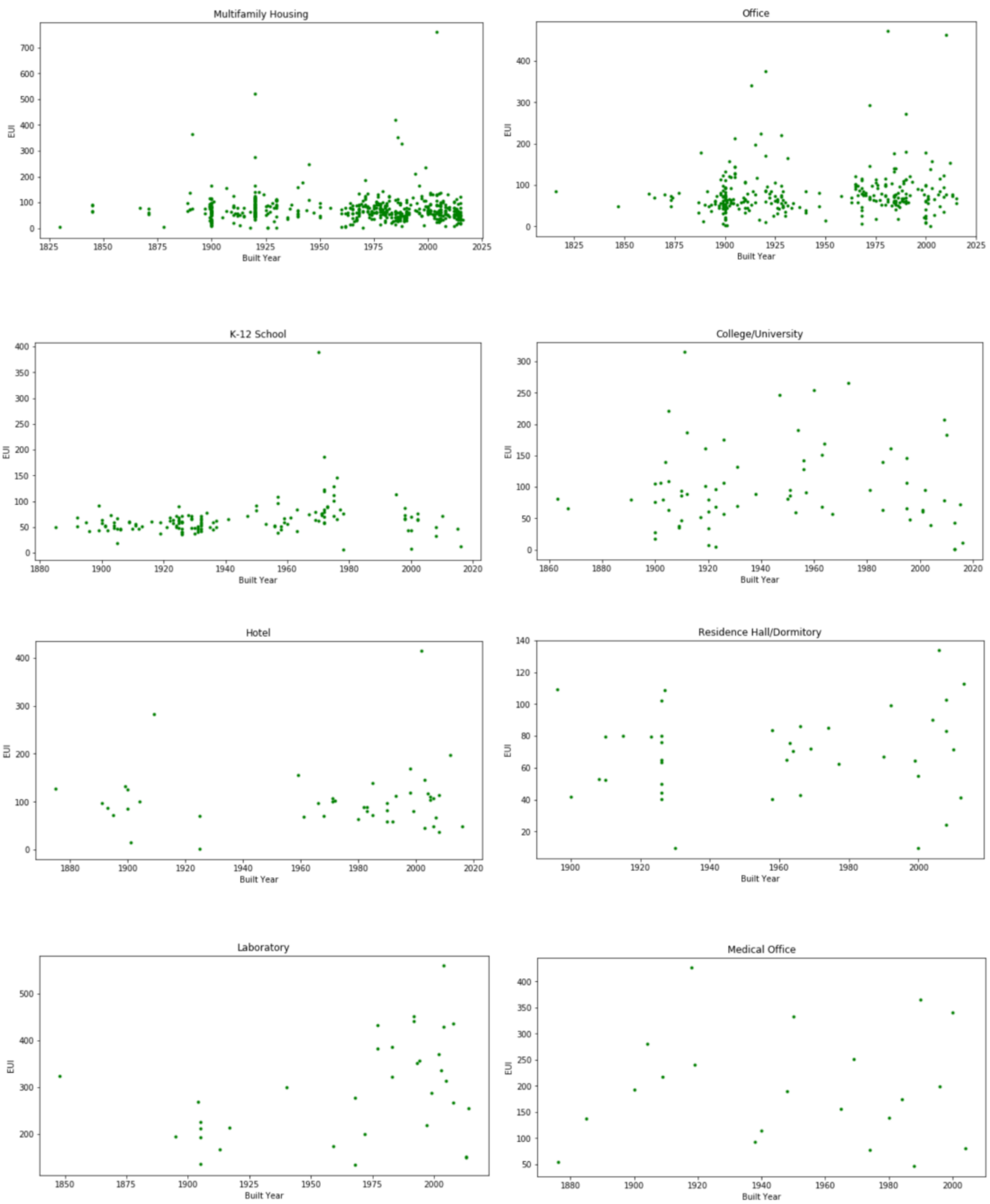


Img4: Average EUI for all buildings of different year of built

We analyzed the average EUI among all types of buildings first, but we noticed that year of built may not be the only reason which effect EUI, according to 3.1, we decided to analyze EUI vary with the year of built in different types buildings.

3) How (much) does average EUI vary with the built year of the following types of buildings:

Types listed: Multifamily Housing; Office; School; College/University; Hotel; Residence Hall/Dormitory; Laboratory; Medical Office



Img5: Average EUI for all buildings of different year of built in different types

3.2.2 Data analysis

From the plots above, we have the following conclusions:

1. In order to analyze this aspect more accurately, we divided the data into different classes according to their types.
2. Building age is negatively correlated with EUI. Therefore, older buildings are found to be more efficient than those built more recently.

Hypothetical conclusion:

1) Newly-built buildings are designed under the consideration of energy-saving, but they may be used in modern ways which consume more energy than we expected.

2) Older buildings are designed without energy-saving consciousness, but are used in outdated and traditional ways which consume less energy.

3.3 How (much) does structure class affect average EUI?

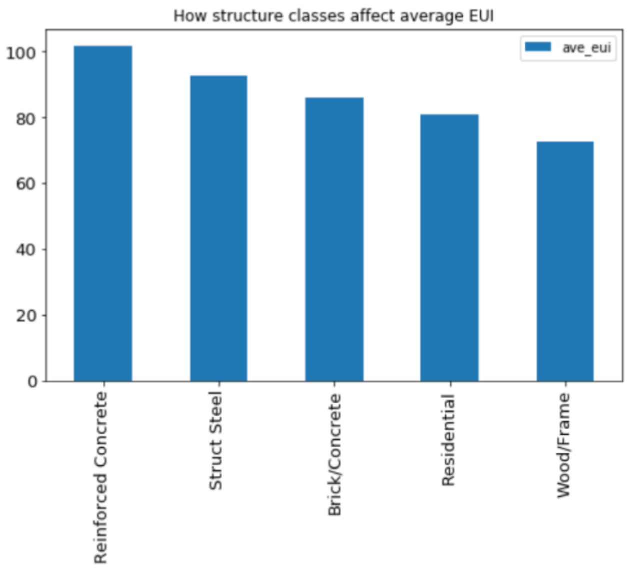
3.3.1 Data processing

1) Datasets:

- BERDO_2017
- Property Assessment

2) How structure affects average EUI for each type of building?

For all 5 types of structure classes, we sorted them in terms of average EUI as follow.



Img6: Average EUI for all buildings of different building structures

3.3.2 Data analysis

From plot above, we have following conclusions:

1. Reinforce concrete is synthetic material which has poor performance in heat preservation and thermal insulation aspects, so reinforce concrete buildings consumes more energy.

2. The average EUI of wood/frame buildings are lowest among others.

   Wood/frame is natural material which has better performance in heat preservation and thermal insulation aspects, so wood/frame buildings consumes less energy.

3.4 How does each property type of buildings vary with temperature?

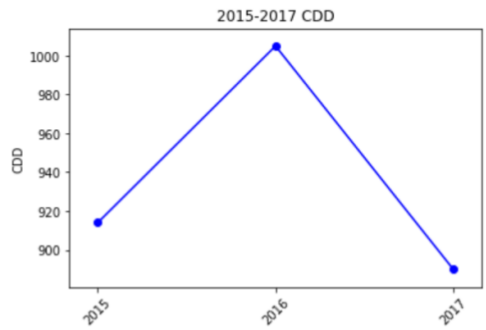
By doing this, we substitute the term temperature with CDD/HDD, which is the kind of property that we really care about – energy consuming needs.

3.4.1 Data processing

1) Datasets:

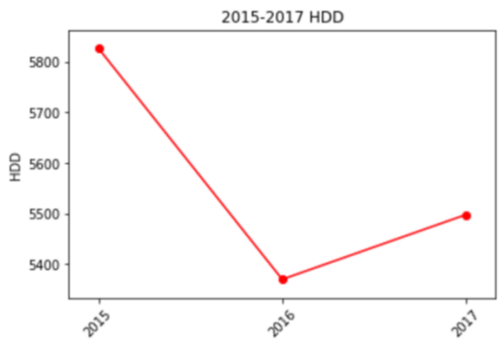
- BERDO_2017
- BERDO_2016
- BERDO_2015
- Cooling Degree Days (CDD) 2015-2017
- Heating Degree Days (HDD) 2015-2017

2) Cooling Degree Days (CDD) 2015-2017. Cooling degree days (CDD) is a measurement designed to quantify the demand for energy needed to cooling a building.



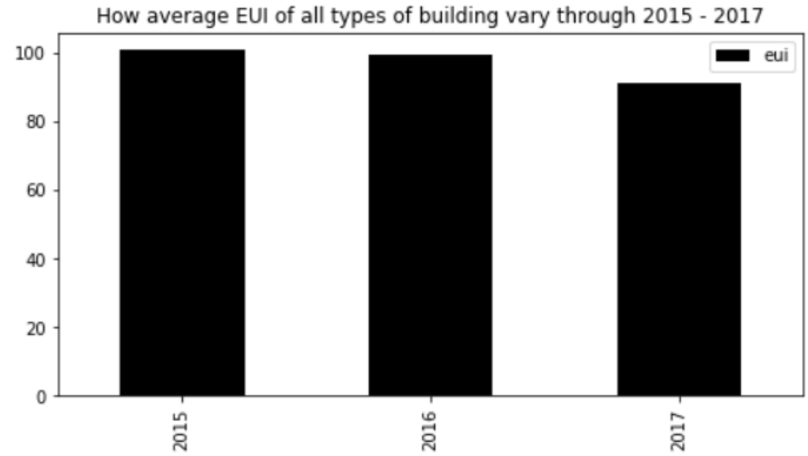
Img7: 2015-2017 CDD

3) Heating Degree Days (HDD) 2015-2017. Heating degree days (HDD) is a measurement designed to quantify the demand for energy needed to heat a building.



Img8: 2015-2017 HDD

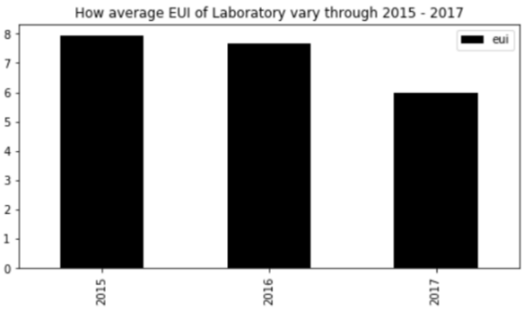
3) How total EUI varies through 2015-2017 for all types of building?

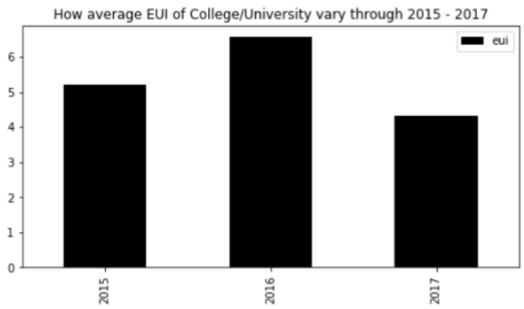


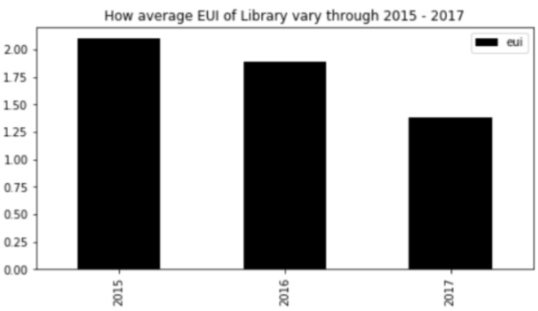
Img9: 2015-2017 EUI for all types of buildings

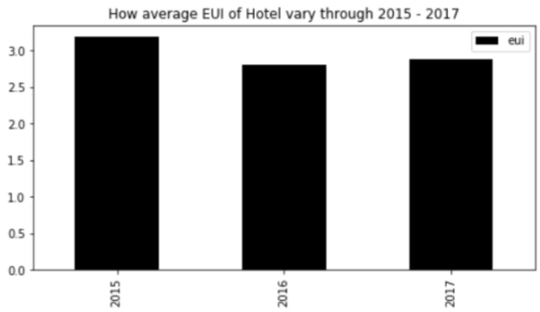
4) How total EUI varies through 2015-2017 for following major types of building:

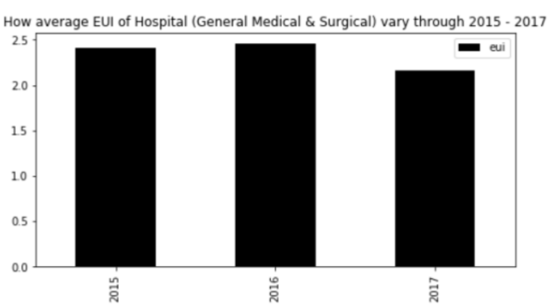
Laboratory; College/University; Library; Hotel; Hospital; Energy/Power Station; Supermarket/Grocery Store; Residence Hall/Dormitory?

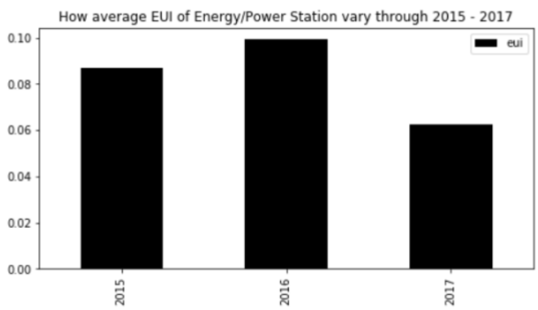


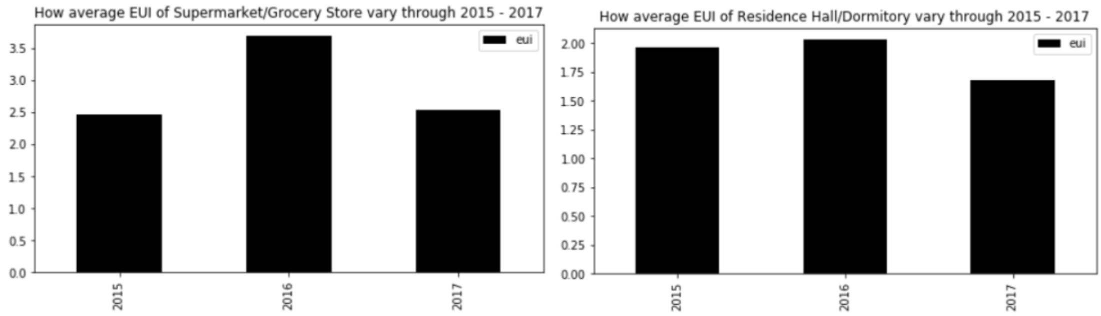












Img10: 2015-2017 EUI for different types of buildings

3.4.2 Data analysis:

From the plots above, we have following conclusions:

1. College/University, Hotel, Hospital, Energy/Power Station, Supermarket/ Grocery Store and Residence Hall/ Dormitory are those types of building that are more sensitive, and specifically, proportional to CDD or HDD trend.
2. Others, including Laboratory and Library are less sensitive to CDD or HDD trend.
3. For those buildings which are more sensitive to CDD or HDD trend, we analyzed the reasons as following:
   a. Building types like university, hotel, hospital, supermarket are public facilities which means that they are open to society. It makes sense to us that the EUI is proportional to temperature change, for example, the temperature is lower than usual, it consume more heating energy for people who work or stay there need to warm themselves.
   b. Building types like residence hall or dormitory, it is obvious that they are residential hall which means that people live there need warm themselves using heating energy or cool themselves using cooling energy which makes the EUI proportional to temperature change.
4. For those buildings which are less sensitive to CDD or HDD trend, we analyzed the reasons as following:
   a. Building types like laboratory and library are non-residential hall, which means that people usually don't spend all day there.
   b. For laboratories, although they need energy to maintain their daily usage, yet the main reason which causes the change of EUI is not just temperature change, for example, the experiments scientists or students conduct there. Also, laboratories may not open every day.

3.5 Which variables significantly affect the EUI of multi-family/office building?

Regression Results, Source EUI as Dependent Variables, Multi-Family Buildings

OLS Regression Results

Dep. Variable:	EUI	R-squared:	0.023
Model:	OLS	Adj. R-squared:	-0.054
Method:	Least Squares	F-statistic:	0.3001
Date:	Wed, 02 May 2018	Prob (F-statistic):	0.952
Time:	14:25:23	Log-Likelihood:	-686.03
No. Observations:	97	AIC:	1388.
Df Residuals:	89	BIC:	1409.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	148.1130	93.608	1.582	0.117	-37.885	334.111
YR_BUILT	-1.0279	0.844	-1.218	0.227	-2.706	0.650
NUM_FLOORS	-1.0927	9.038	-0.121	0.904	-19.050	16.865
GROSS_AREA	-0.0002	0.001	-0.314	0.754	-0.002	0.001
S_B	28.7732	88.530	0.325	0.746	-147.134	204.680
S_A	44.9877	177.067	0.254	0.800	-306.841	396.817
S_R	51.2359	105.782	0.484	0.629	-158.951	261.423
S_C	62.3370	62.814	0.992	0.324	-62.474	187.148
S_D	-39.2209	130.835	-0.300	0.765	-299.187	220.745

Omnibus:	200.817	Durbin-Watson:	2.112
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28924.145
Skew:	8.966	Prob(JB):	0.00
Kurtosis:	85.674	Cond. No.	3.18e+21

Regression Results, log Source EUI as Dependent Variables, Multi-Family Buildings

OLS Regression Results

Dep. Variable:	y	R-squared:	0.046
Model:	OLS	Adj. R-squared:	-0.029
Method:	Least Squares	F-statistic:	0.6194
Date:	Wed, 02 May 2018	Prob (F-statistic):	0.739
Time:	14:44:25	Log-Likelihood:	-131.24
No. Observations:	97	AIC:	278.5
Df Residuals:	89	BIC:	299.1
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.5994	0.307	11.719	0.000	2.989	4.210
YR_BUILT	-0.0019	0.003	-0.687	0.494	-0.007	0.004
NUM_FLOORS	0.0260	0.030	0.876	0.383	-0.033	0.085
GROSS_AREA	-2.601e-06	2.22e-06	-1.172	0.244	-7.01e-06	1.81e-06
S_B	0.8099	0.290	2.788	0.006	0.233	1.387
S_A	1.0380	0.581	1.787	0.077	-0.116	2.192
S_R	0.9778	0.347	2.817	0.006	0.288	1.668
S_C	0.5571	0.206	2.703	0.008	0.148	0.967
S_D	0.2167	0.429	0.505	0.615	-0.636	1.070

Omnibus:	32.958	Durbin-Watson:	2.299
Prob(Omnibus):	0.000	Jarque-Bera (JB):	165.487
Skew:	-0.892	Prob(JB):	1.16e-36
Kurtosis:	9.145	Cond. No.	3.18e+21

Regression Results, Source EUI as Dependent Variables, Office Buildings

OLS Regression Results

Dep. Variable:	EUI	R-squared:	0.178
Model:	OLS	Adj. R-squared:	0.074
Method:	Least Squares	F-statistic:	1.704
Date:	Wed, 02 May 2018	Prob (F-statistic):	0.127
Time:	14:57:44	Log-Likelihood:	-305.44
No. Observations:	63	AIC:	626.9
Df Residuals:	55	BIC:	644.0
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	59.7705	10.408	5.743	0.000	38.912	80.629
YR_BUILT	-0.0569	0.133	-0.428	0.670	-0.323	0.210
NUM_FLOORS	-0.4187	0.826	-0.507	0.614	-2.073	1.236
GROSS_AREA	3.005e-05	2.05e-05	1.463	0.149	-1.11e-05	7.12e-05
S_B	34.4154	7.699	4.470	0.000	18.987	49.844
S_A	4.7279	13.457	0.351	0.727	-22.240	31.696
S_R	28.7684	14.928	1.927	0.059	-1.148	58.684
S_C	16.8879	9.987	1.691	0.096	-3.126	36.902
S_D	-25.0290	19.910	-1.257	0.214	-64.929	14.871

Omnibus:	19.505	Durbin-Watson:	1.872
Prob(Omnibus):	0.000	Jarque-Bera (JB):	25.882
Skew:	1.244	Prob(JB):	2.40e-06
Kurtosis:	4.916	Cond. No.	1.03e+22

Regression Results, log Source EUI as Dependent Variables, Office Buildings

OLS Regression Results

Dep. Variable:	y	R-squared:	0.276
Model:	OLS	Adj. R-squared:	0.184
Method:	Least Squares	F-statistic:	3.002
Date:	Wed, 02 May 2018	Prob (F-statistic):	0.00963
Time:	14:59:38	Log-Likelihood:	-24.728
No. Observations:	63	AIC:	65.46
Df Residuals:	55	BIC:	82.60
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.4010	0.121	28.139	0.000	3.159	3.643
YR_BUILT	-0.0003	0.002	-0.192	0.848	-0.003	0.003
NUM_FLOORS	-0.0032	0.010	-0.335	0.739	-0.022	0.016
GROSS_AREA	4.123e-07	2.39e-07	1.728	0.090	-6.59e-08	8.9e-07
S_B	1.0171	0.089	11.377	0.000	0.838	1.196
S_A	0.6350	0.156	4.063	0.000	0.322	0.948
S_R	0.9312	0.173	5.372	0.000	0.584	1.279
S_C	0.8099	0.116	6.984	0.000	0.578	1.042
S_D	0.0078	0.231	0.034	0.973	-0.456	0.471

Omnibus:	0.674	Durbin-Watson:	1.735
Prob(Omnibus):	0.714	Jarque-Bera (JB):	0.535
Skew:	0.223	Prob(JB):	0.765
Kurtosis:	2.925	Cond. No.	1.03e+22

3.5.1 Data Analysis

1. Older buildings are shown to be correlated with lower EUIs for both multi-family and office buildings. This finding is consistent with the results before and reinforces the link between older buildings and energy efficiency.
2. There is a positive correlation between EUI and gross area for office buildings. Larger office buildings, therefore, are shown to have higher EUIs
3. There is a negative correlation between EUI and gross area for multi-family buildings. Contrary to the findings for office buildings, larger multi-family buildings are found to be more efficient.
4. The regression results show that the coefficients for number of floor of office buildings are negative, which means that offices with more floors are more efficient.

4 Future Steps

1. Analyze the main factors of EUI for other types of buildings like university, hotel and laboratory

2. Develop a predictive model to create an energy performance benchmark or to estimate energy consumption for buildings. Method : Robust multiple regression techniques

3. Try to solve the question that which structure or material should be used to construct each type of buildings to make them more efficient.

Method :

a. Collect data in terms of the structure or material for each type of building as much as possible.

b. Try to compare the average of EUI with different structures or materials (controlling the other variables)

4. Analyze why older buildings are more efficient and give the details and factors which make the older buildings efficient compared to the buildings recently.

5. Analyze how other variables (Occupancy/Energy Source) affect the EUI of Buildings.