Main Report

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 much 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. Taking temperature as an example, “how does temperature varies among 2015-2017”, and “how does EUI of each type of building varies among 2015-2017”, etc.

1.3 Main Methods

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

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

3.1 Average EUI for each property type building 2017

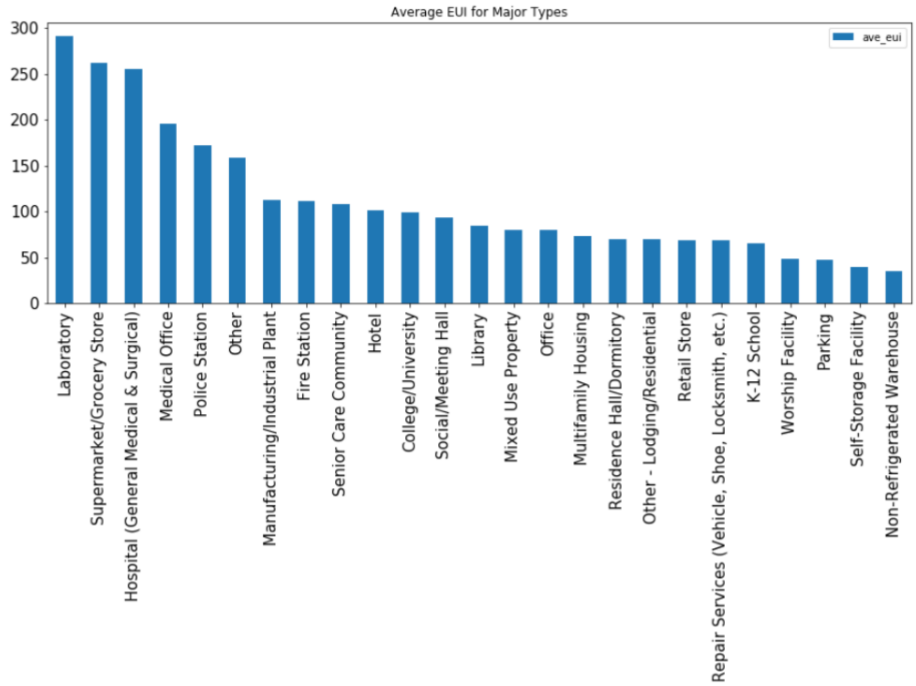
The dataset we used is BERDO_2017. The reasons why we use this instead of the integrated dataset of BERDO and Property Assessment are : 1.BERDO already has all information we need; 2.As discussed before, the integrated dataset has less entries then BERDO.

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

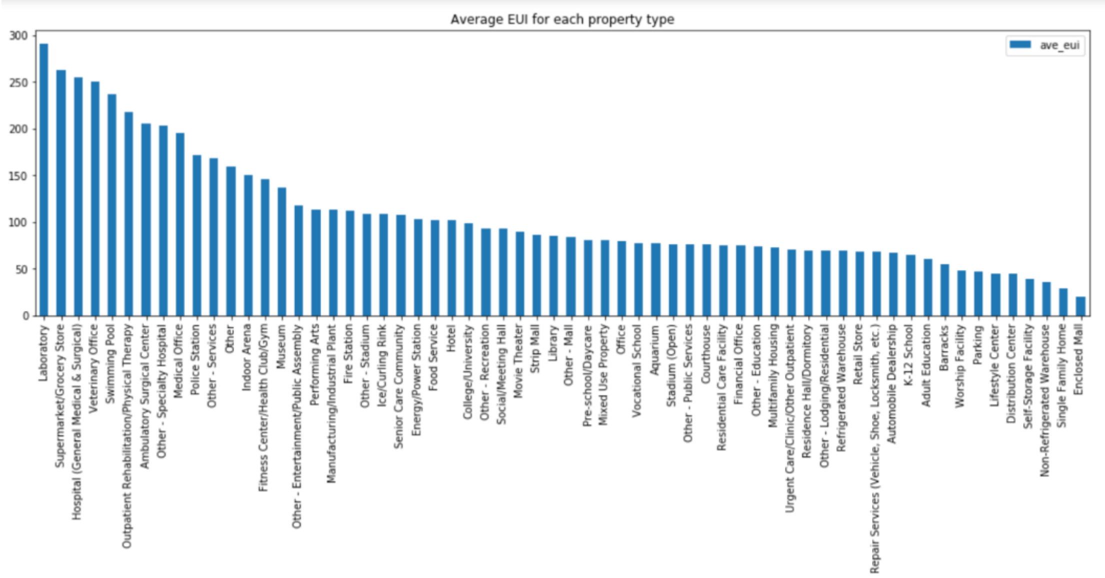
2) Count the top most property types of buildings.

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

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

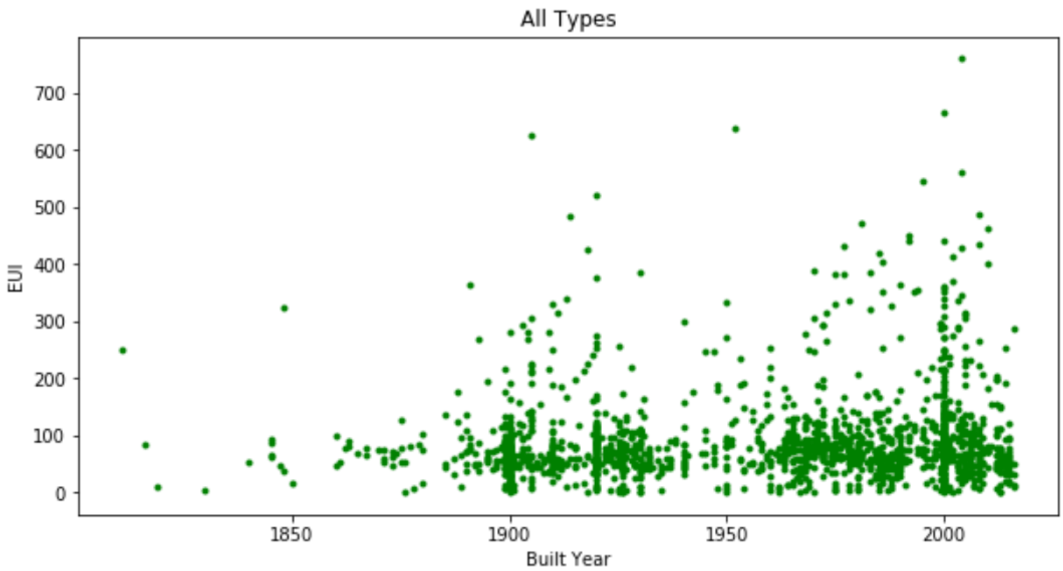


4) Average EUI for all types of building



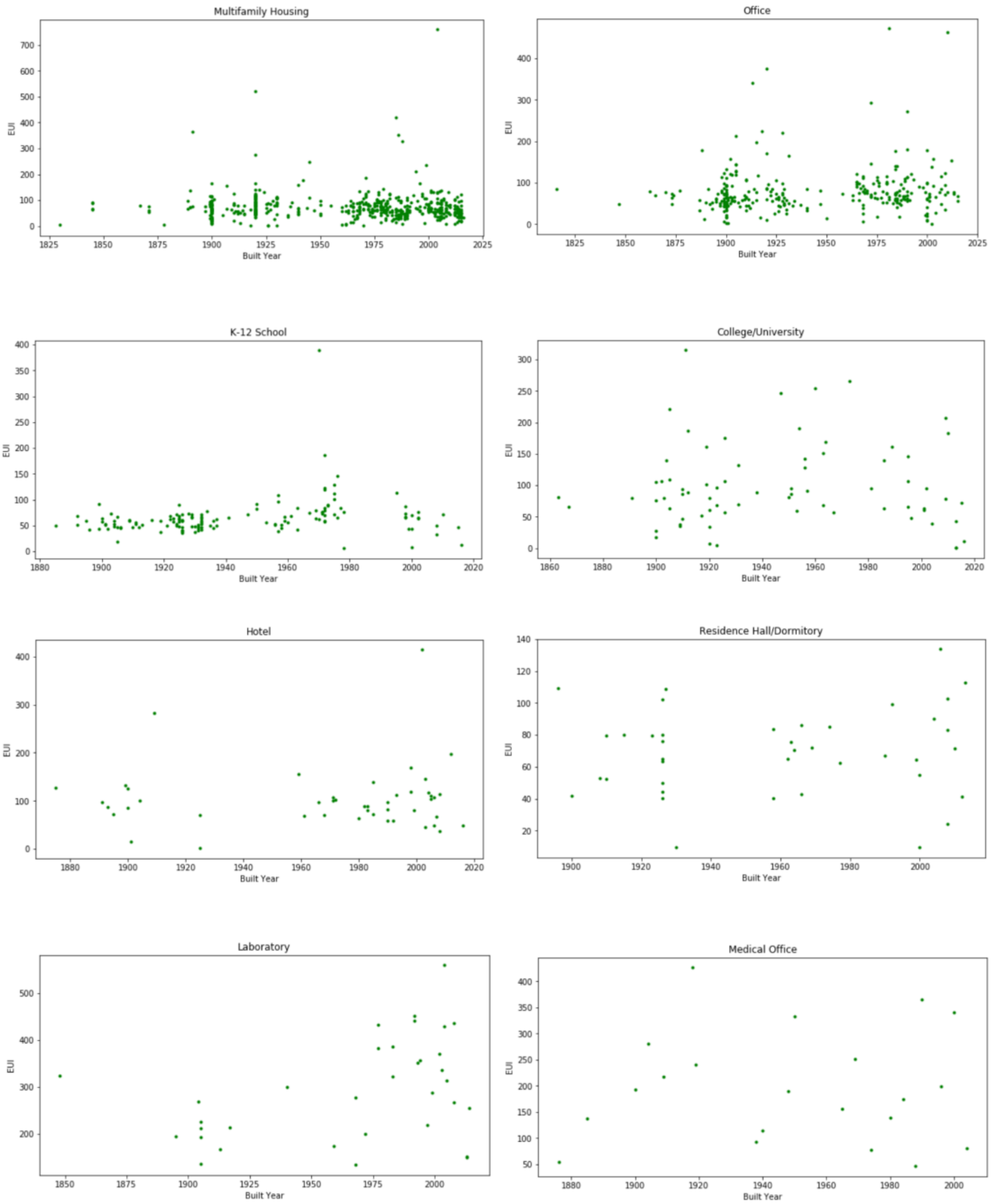
3.2 How (much) does the building's built year affect average EUI for each property type of building?

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

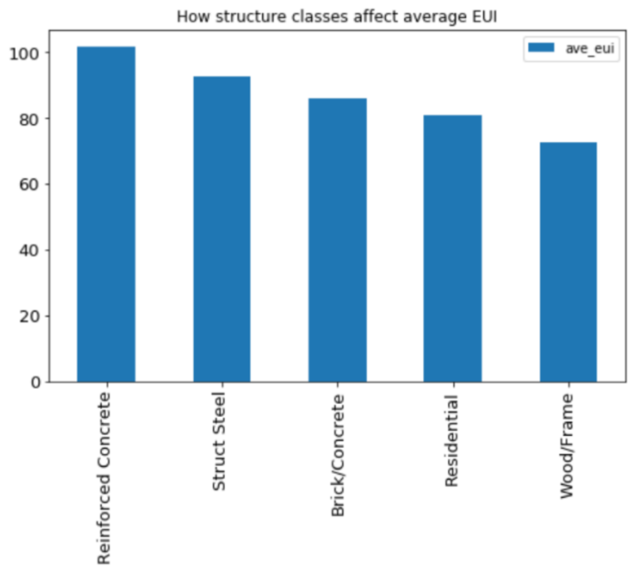


2) How (much) does average EUI vary with the built year of the following types of buildings:

Multifamily Housing; Office; School; College/University; Hotel; Residence Hall/Dormitory; Laboratory; Medical Office?

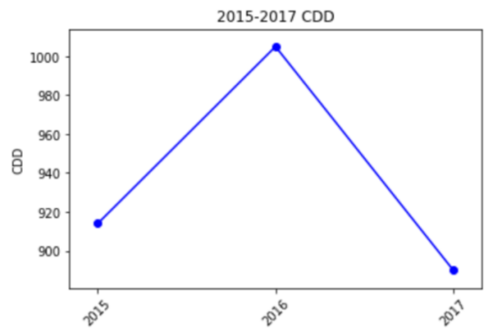


3.3 How (much) does structure class affect average EUI for each property type of building?

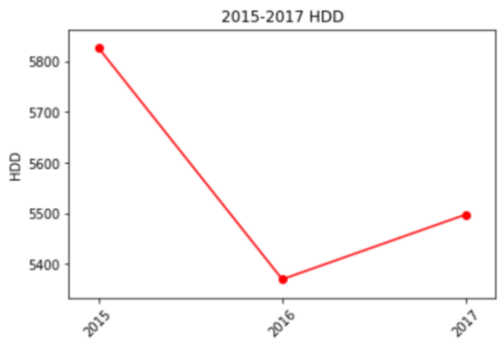


3.4 How (much) does temperature affect average EUI for each property type of building?

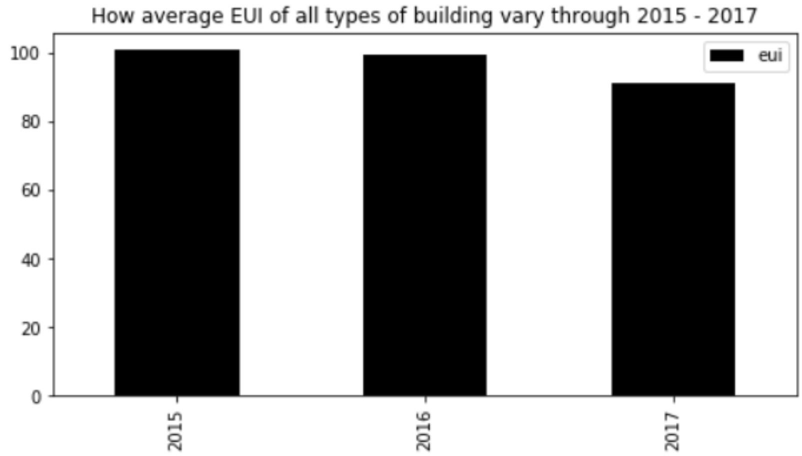
1) Cooling Degree Days (CDD) 2015-2017. Cooling degree days (CDD) is a measurement designed to quantify the demand for energy needed to cooling a building.



2) 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.

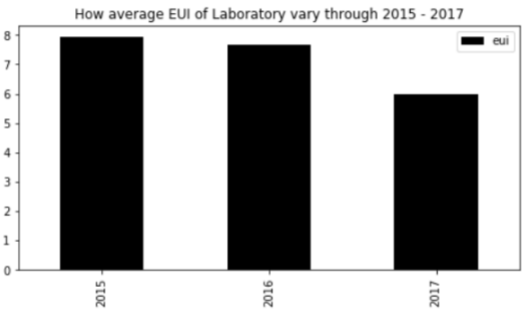


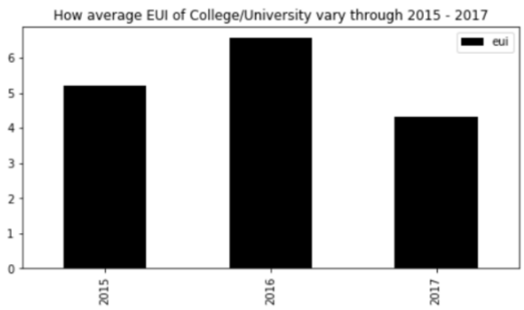
3) How total EUI varies through 2015-2017 for all types of building?

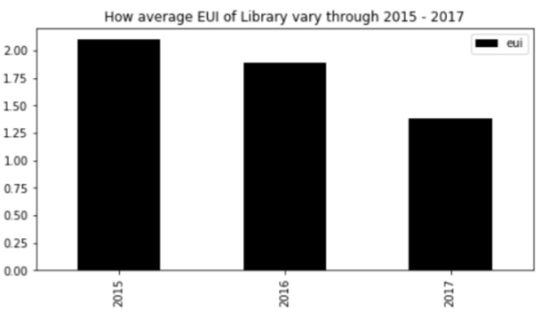


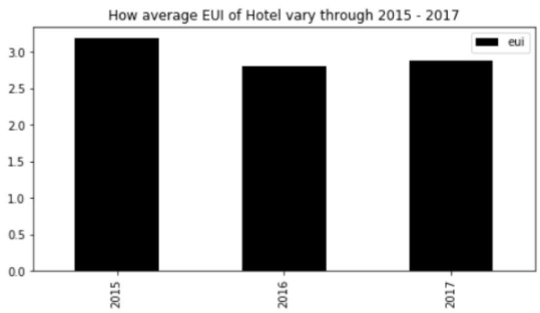
4) How total EUI varies through 2015-2017 for following major types of building:

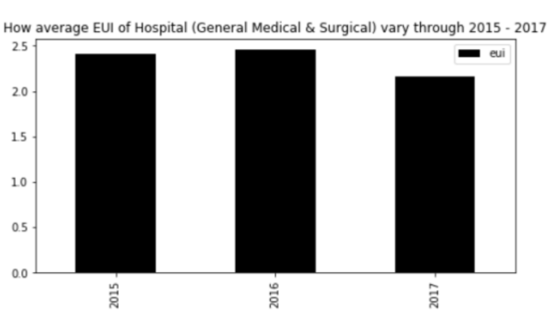
Laboratory; College/University; Library; Hotel; Hospital; Energy/Power Station; Supermarket/Grocery Store; Residence Hall/Dormitory?

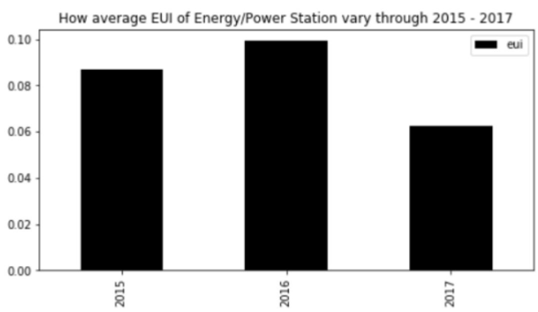


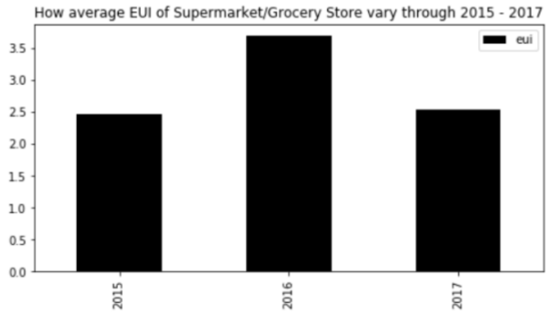


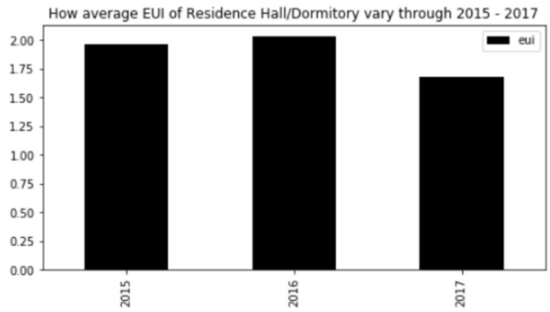












Tentative Conclusion:

(1) College/University, Energy/Power Station and Supermarket/ Grocery Store are those types of building that are more sensitive, and specifically, proportional to CDD trend.

(2) Others, including Laboratory, Library, Hotel, Hospital and Residence Hall/ Dormitory, are less sensitive to CDD or HDD trend.

3.5 Prediction of EUI for a building with given information

The algorithm we used to predict of EUI is linear regression.

The reason why we chose this specific algorithm is, in order to predict EUI with given information of a building, we need a general prediction equation to describe a prediction model. Since there is more than one independent variable, including property type, built year, temperature, gross area, living area, number of floors, structure class, etc., specifically, we need to build a multiple linear regression model. For LU/structure class which are categorical, we use the "indicator vector" representation. And the dependent variable is EUI.

OLS Regression Results

Dep. Variable:	EUI	R-squared:	0.011
Model:	OLS	Adj. R-squared:	-0.031
Method:	Least Squares	F-statistic:	0.2715
Date:	Thu, 12 Apr 2018	Prob (F-statistic):	0.999
Time:	11:32:19	Log-Likelihood:	-4241.9
No. Observations:	443	AIC:	8522.
Df Residuals:	424	BIC:	8600.
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-91.8364	545.560	-0.168	0.866	-1164.175	980.502
YR_BUILT	3.3234	4.825	0.689	0.491	-6.160	12.807
NUM_FLOORS	11.1169	56.689	0.196	0.845	-100.310	122.544
GROSS_AREA	0.0004	0.011	0.039	0.969	-0.020	0.021
LIVING_AREA	-0.0007	0.011	-0.064	0.949	-0.023	0.022
S_B	-170.6969	674.928	-0.253	0.800	-1497.319	1155.925
S_A	-63.6976	907.884	-0.070	0.944	-1848.211	1720.816
S_R	-312.3962	2023.423	-0.154	0.877	-4289.586	3664.793
S_C	264.7762	623.661	0.425	0.671	-961.076	1490.629
S_D	190.1781	877.877	0.217	0.829	-1535.355	1915.711
C	61.5715	913.600	0.067	0.946	-1734.177	1857.320
R3	87.1004	2240.709	0.039	0.969	-4317.180	4491.381
E	388.4468	883.477	0.440	0.660	-1348.094	2124.988
R2	85.1949	2282.423	0.037	0.970	-4401.077	4571.467
EA	-326.8448	1192.379	-0.274	0.784	-2670.554	2016.864
R1	101.8616	2046.660	0.050	0.960	-3921.001	4124.724
I	-325.9479	1265.917	-0.257	0.797	-2814.202	2162.307
RC	-263.8310	966.695	-0.273	0.785	-2163.942	1636.280
A	-276.2896	933.029	-0.296	0.767	-2110.227	1557.648
CM	381.0269	3754.997	0.101	0.919	-6999.699	7761.753
R4	0	0	nan	nan	0	0
CD	-4.1253	3380.499	-0.001	0.999	-6648.748	6640.497

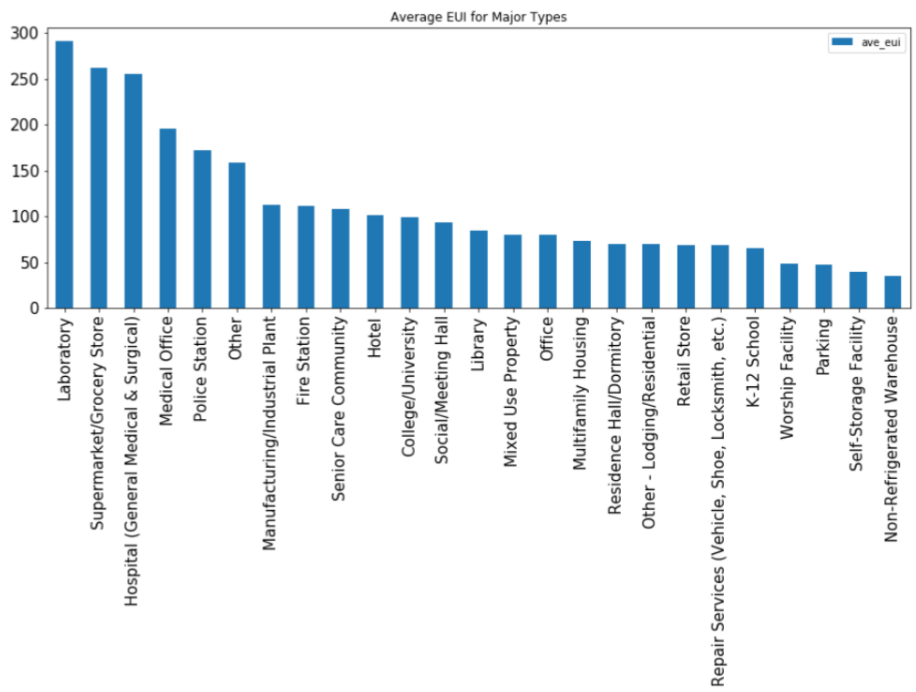
Omnibus:	919.709	Durbin-Watson:	2.016
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1117081.247
Skew:	14.950	Prob(JB):	0.00
Kurtosis:	247.182	Cond. No.	inf

We notice the R-squared is too low (0.011), so these variables are not enough to predict the EUI.

5 Conclusions

5.1 What is average EUI for each type of building? Which contributes the most?

Among major types of buildings, following types has the greatest Average Energy Usage Intensity:

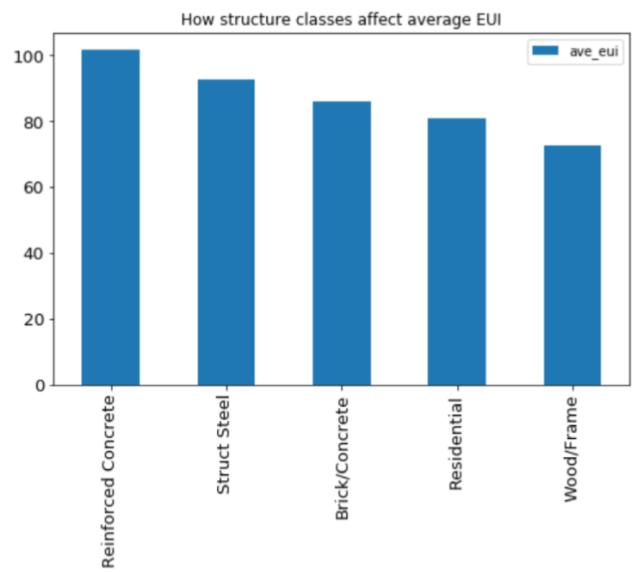


5.2 How (much) does the building's built year affect average EUI for each property type of building?

After analyzing how (much) does the building' built year affect average EUI for all types, or each specific type of building, we come to a conclusion that the built year do not affect EUI as much as we expected.

5.3 How (much) does structure class affect average EUI for each property type of building?

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



5.4 How (much) does temperature affect average EUI for each property type of building?

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.

Tentative conclusions are as follow:

(1) College/University, Energy/Power Station and Supermarket/ Grocery Store are those types of building that are more sensitive, and specifically, proportional to CDD trend.

(2) Others, including Laboratory, Library, Hotel, Hospital and Residence Hall/ Dormitory, are less sensitive to CDD or HDD trend.

5.5 Prediction of EUI for a building with given information.

Based on the results of multiple linear regression analysis, R square is not high enough for us to safely predict EUI for a building with given information.

6 Future Steps

- Analyze the reason why each element has different effect on the EUI of a building specifically.
- Try to search for more information and analyze other elements to predict the EUI of a building.
- Give a final conclusion about how these elements effect the EUI of a building.