# 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.

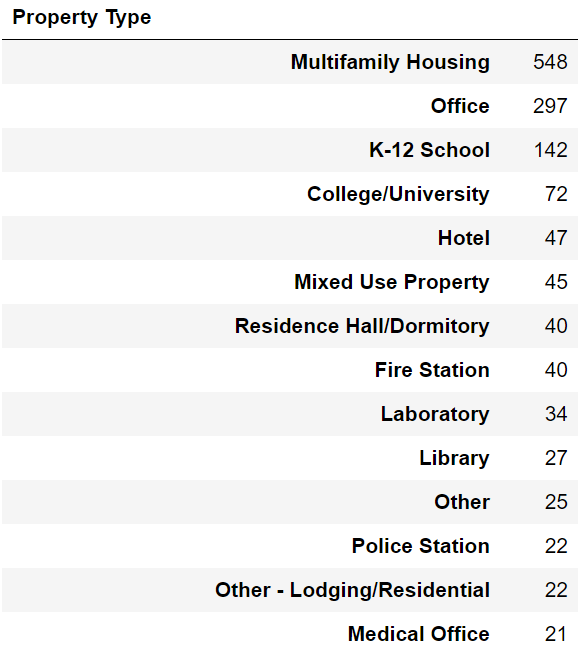
The strategy dealing with missing values or inconsistent entries is simply remove it. The percentage of this process would normally cost less then 1% of all data.

### 3 Data Analysis

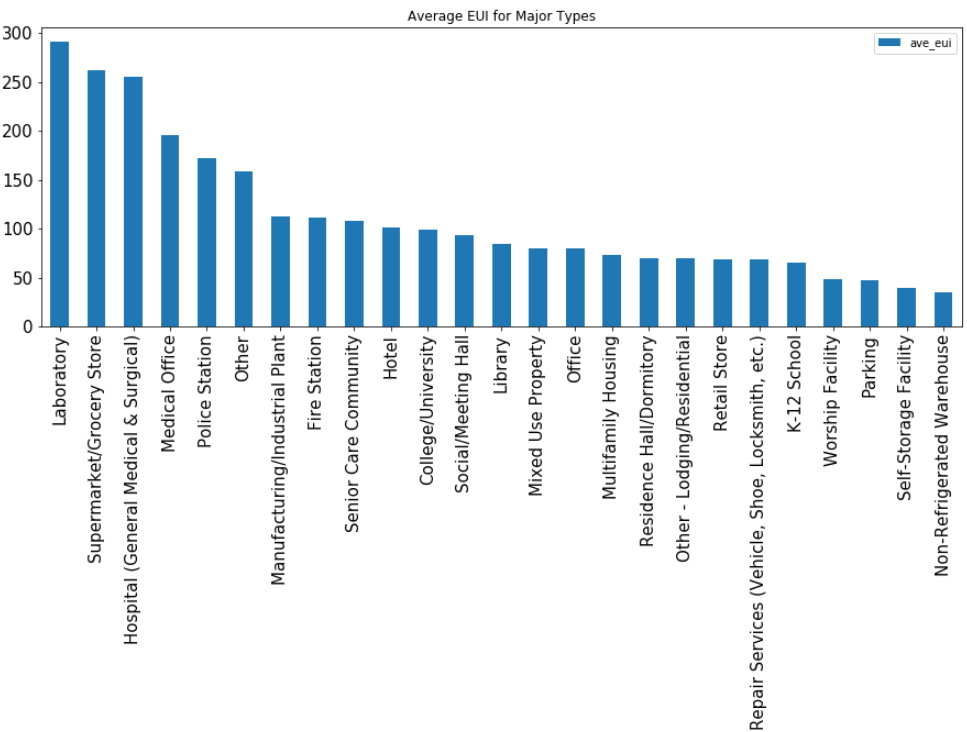
**3.1 Average EUI for each property type building 2017**

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.

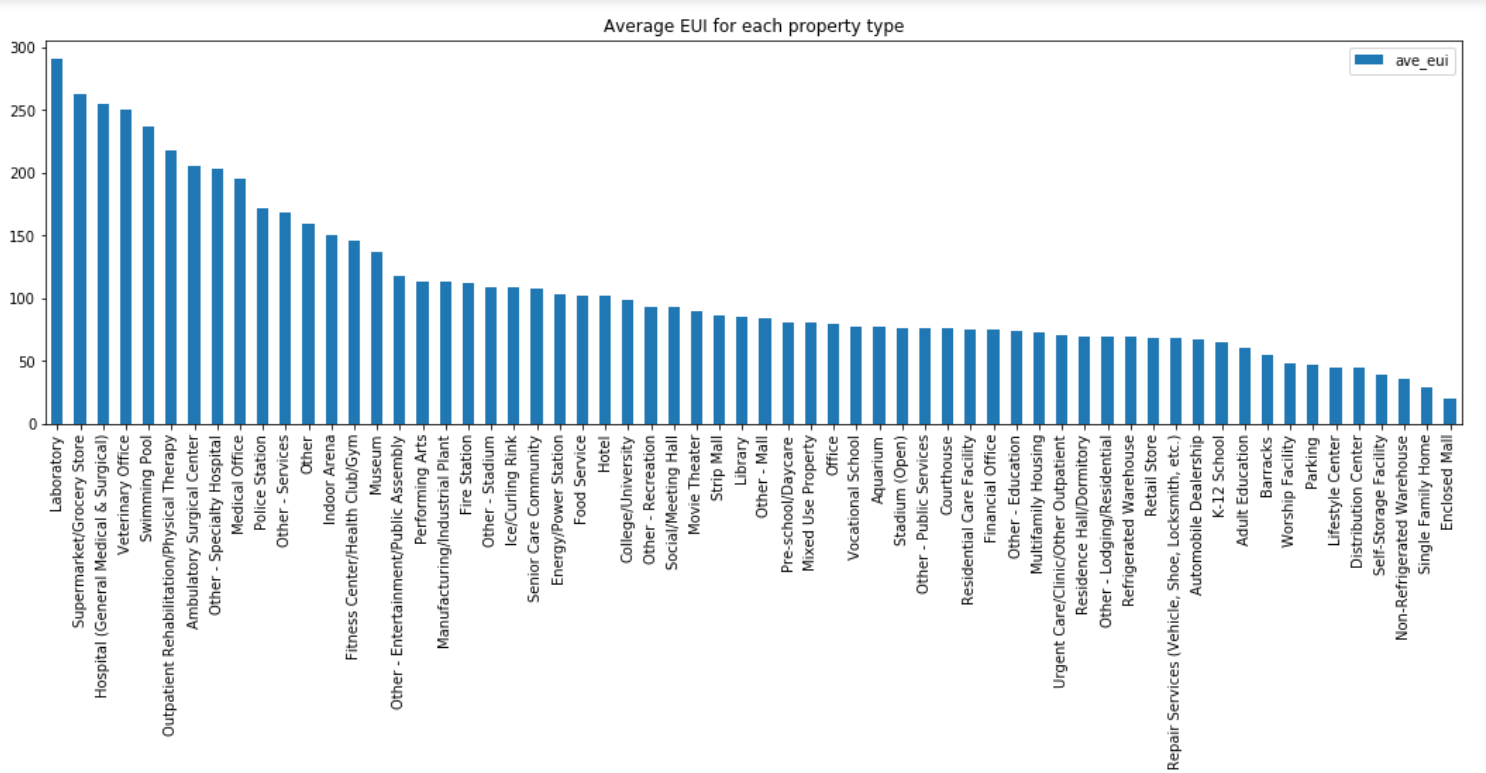
1. Remove the unavailable EUI. The total number of building data reduced from 1800 to 1664. Detect and Remove Outliers
2. Count the top most property types of buildings.



1. Average EUI for major types of building (>10)

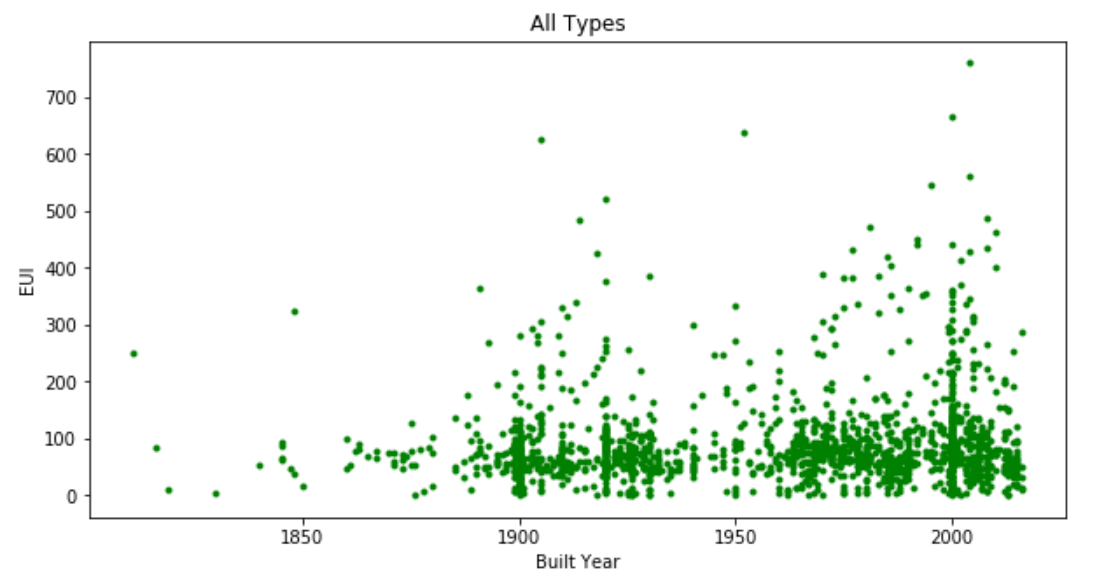


1. Average EUI for all types of building



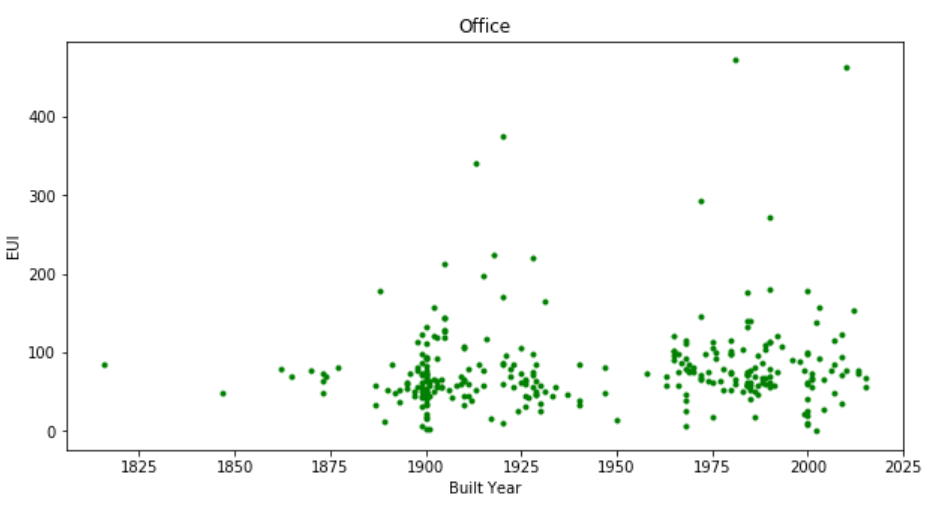
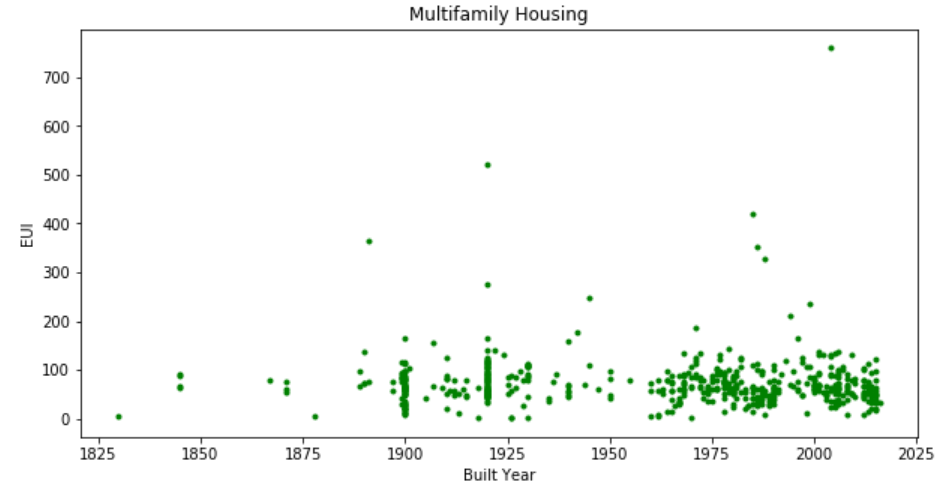
**3.2 How (much) does the building’s built year affect average EUI for each property type of building**

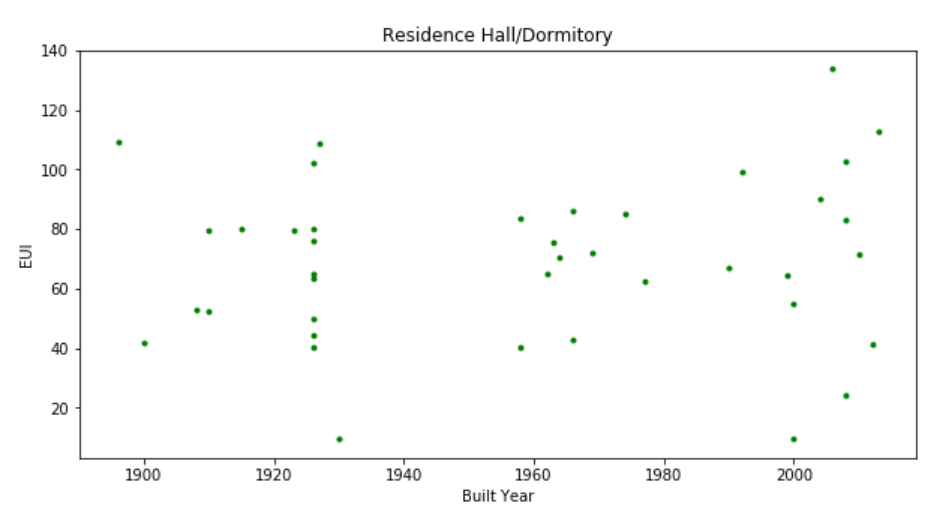
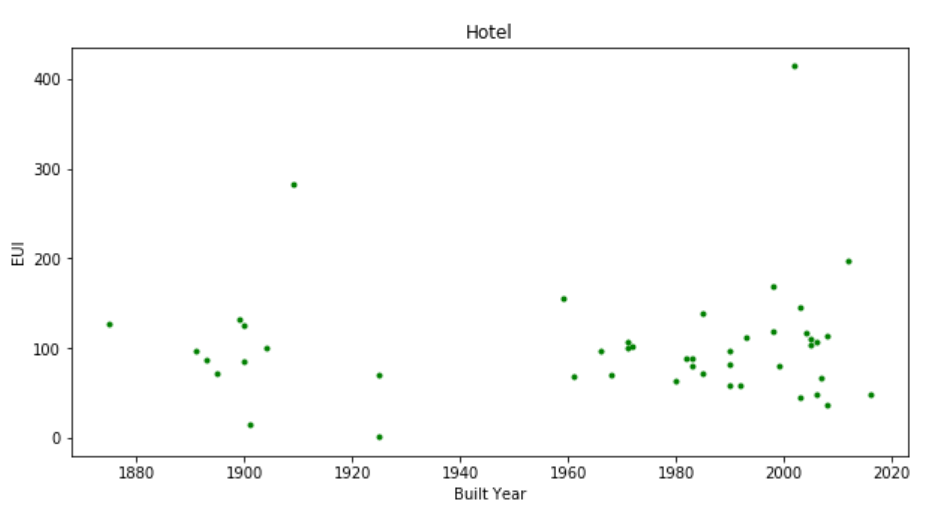
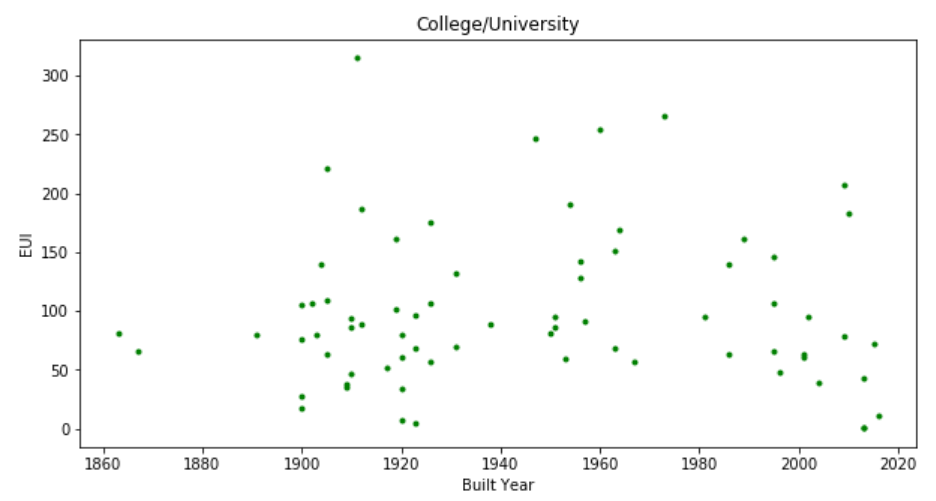
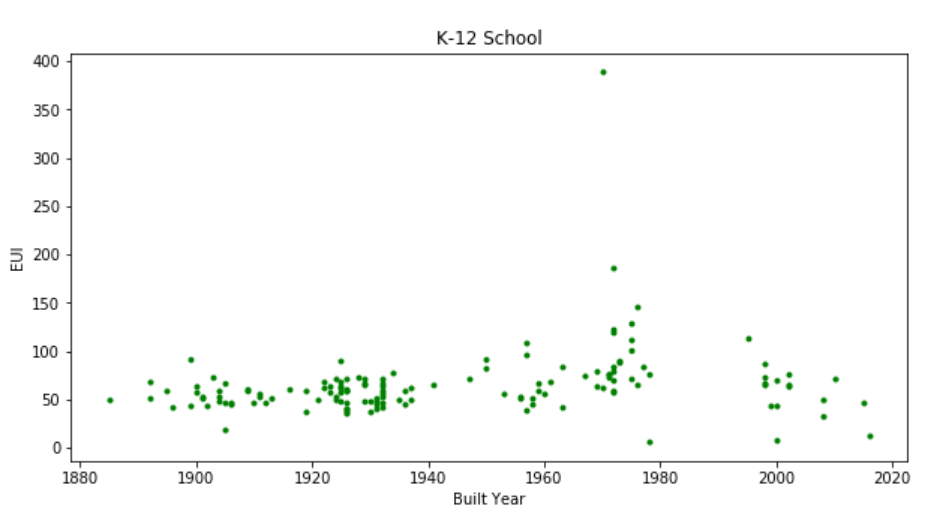
1. How (much) does average EUI vary with the built year of **all** types of buildings

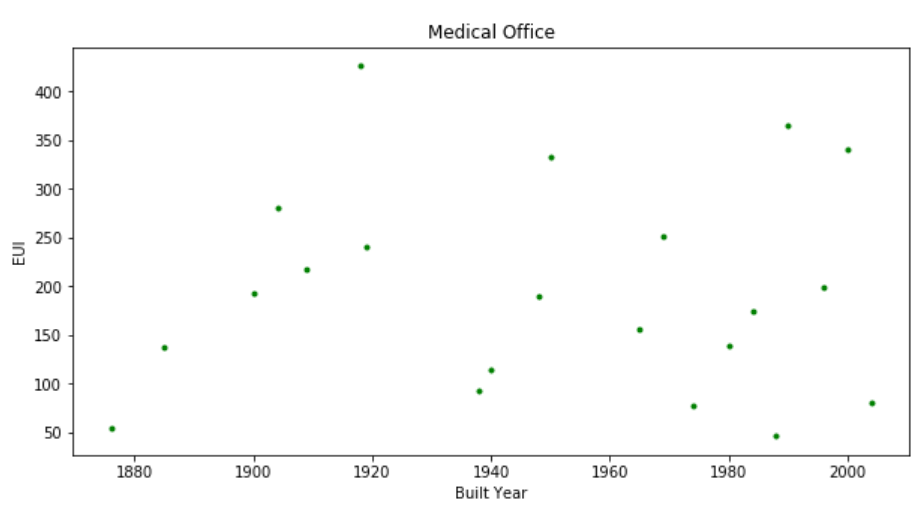
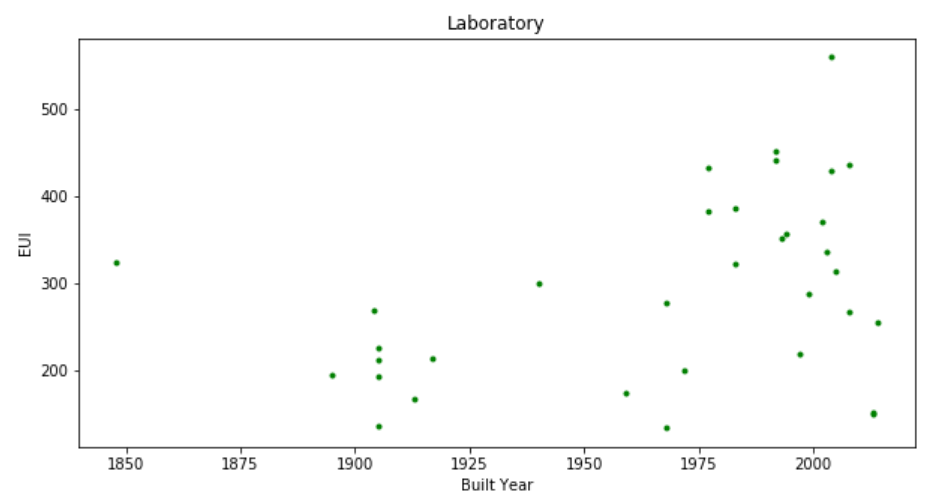


1. How (much) does average EUI vary with the built year of the following types of buildings:

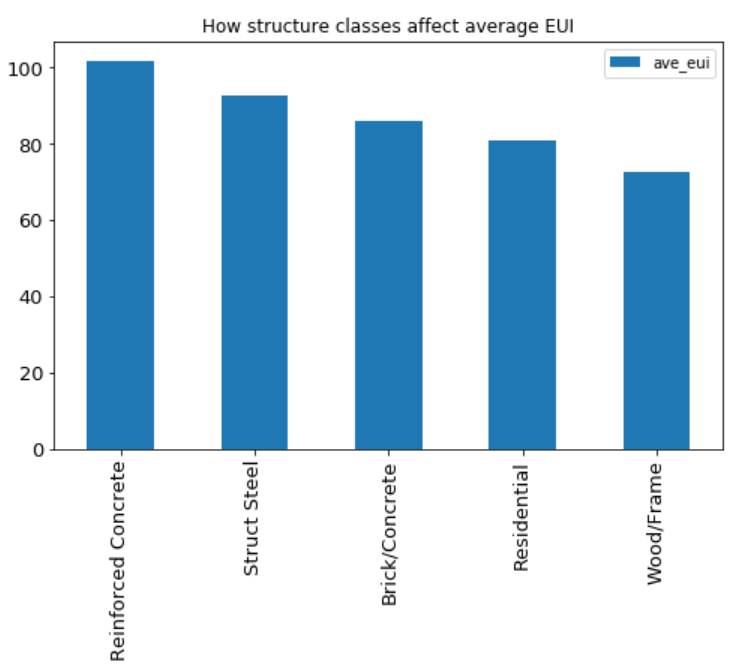
Multifamily Housing; Office; School; College/University; Hotel; Residence Hall/Dormitory; Laboratory; Medical Office





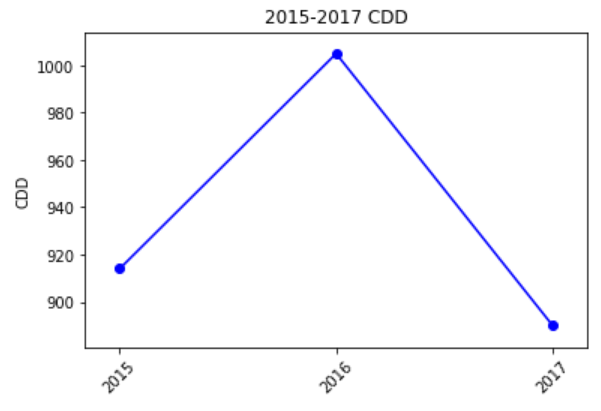


**3.3 How (much) does structure class affect average EUI for each property type of building**

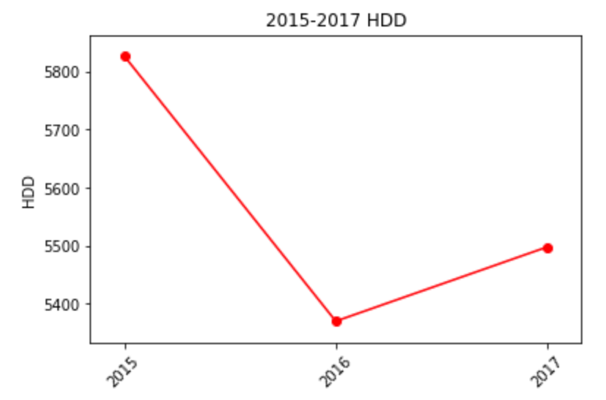


**3.4 How (much) does temperature affect average EUI for each property type of building**

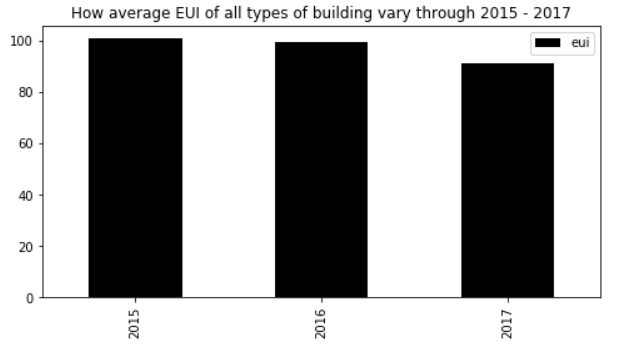
1. Cooling Degree Days (CDD) 2015-2017. Cooling degree days (CDD) is a measurement designed to quantify the demand for energy needed to cooling a building.



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.

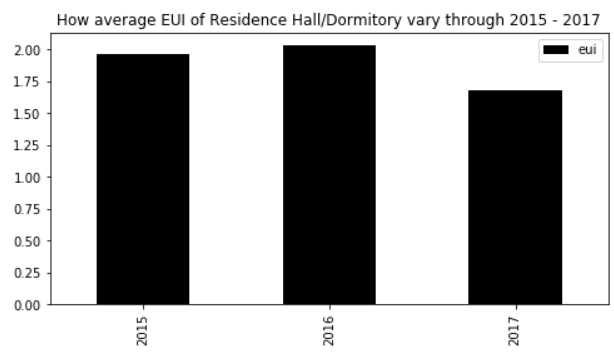
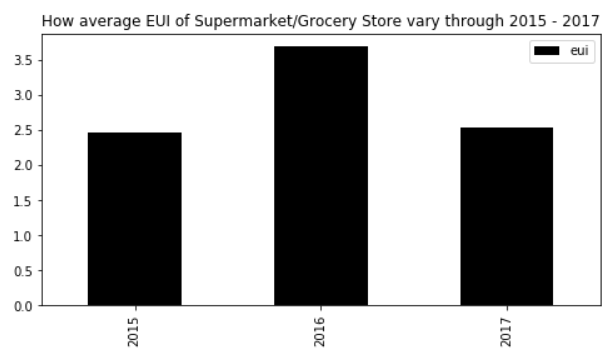
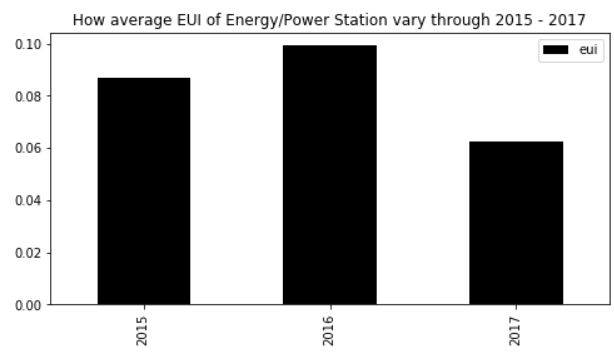
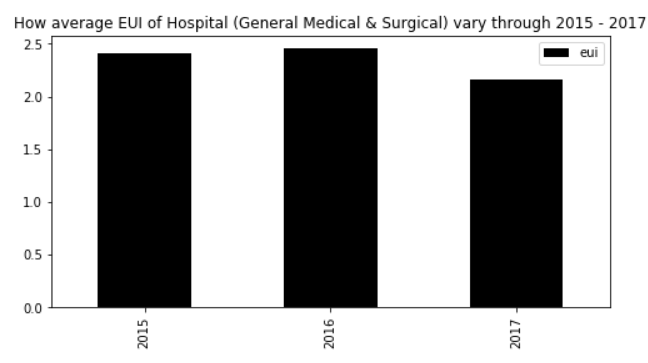
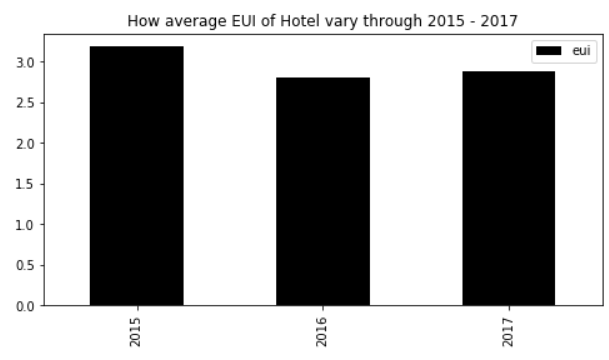
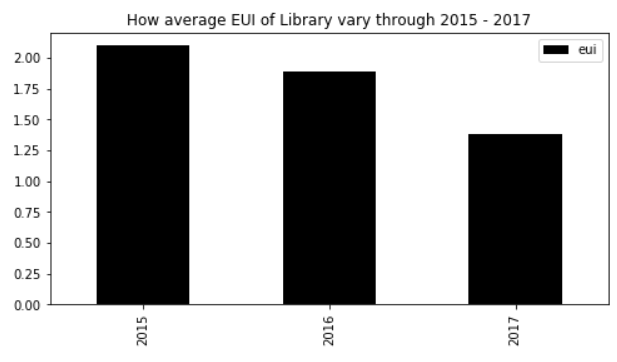
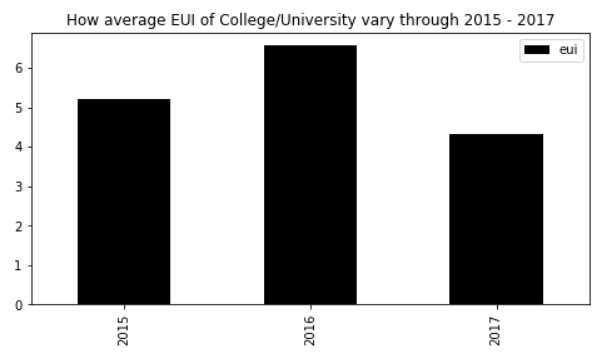
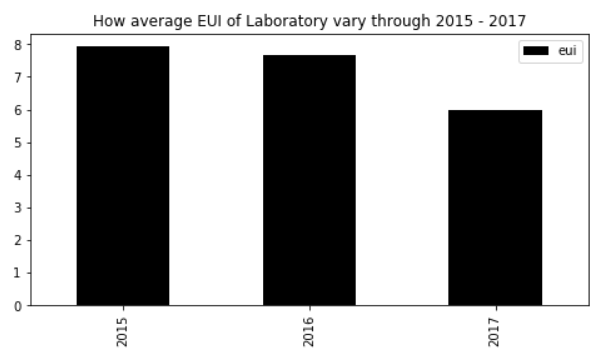


3) How total EUI varies through 2015-2017 for all types of building



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:**

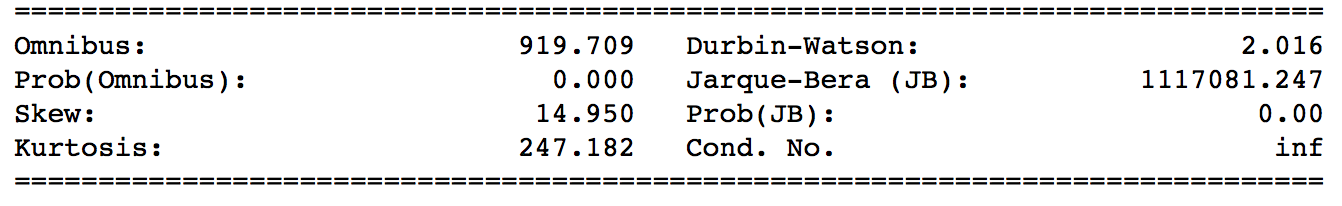
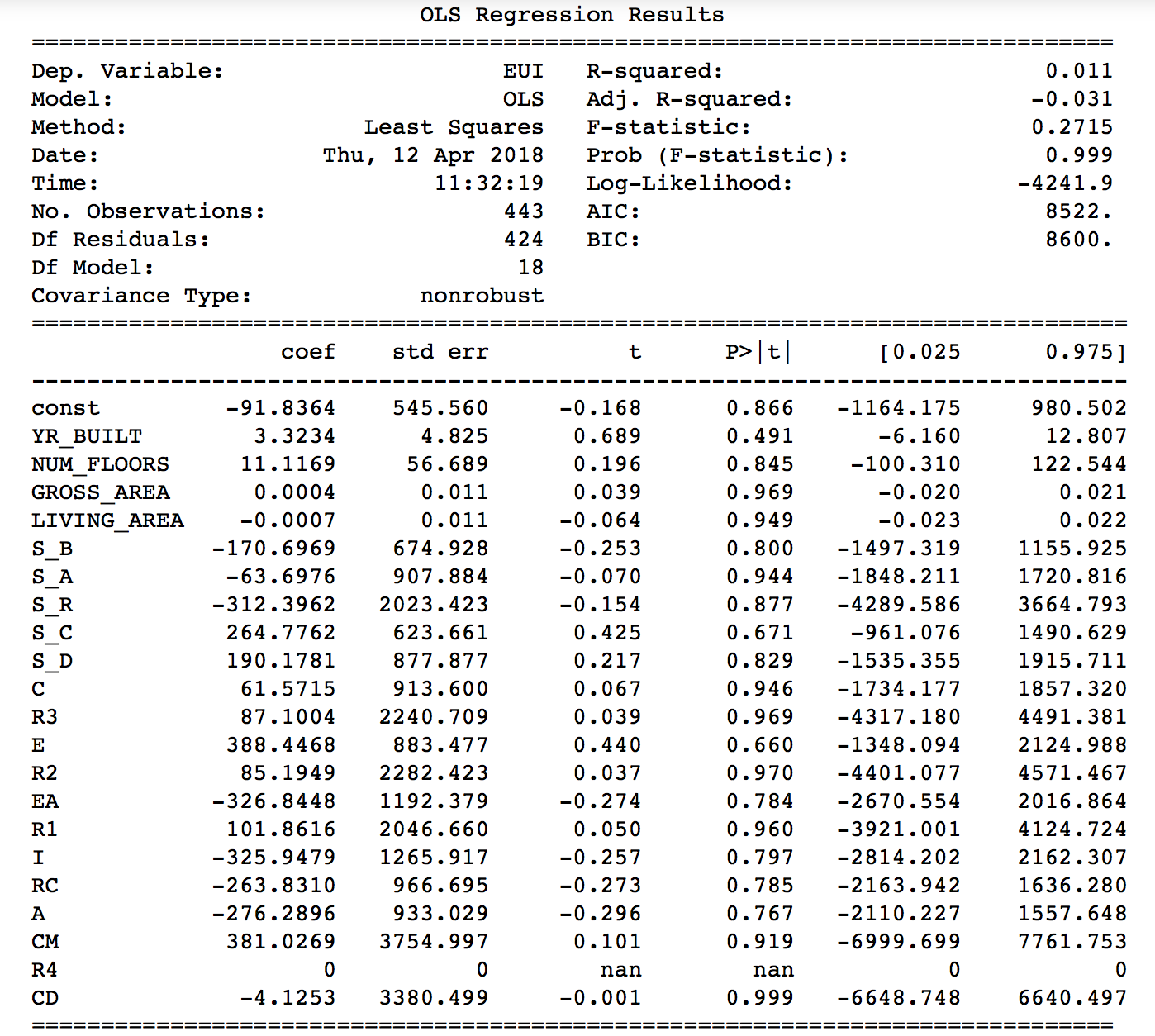
(1) College/University, Energy/Power Station and Supermarket/ Grocery Store are those types of building that are more sensitive, and specifically, proportional to CDD trend.

(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.



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

### 5 Conclusions

Your conclusions should answer your questions. How do your results answer your questions? Did you get the results that you were expecting? Do you feel that some questions need to be addressed again, or perhaps cannot be addressed at all (and why)?

   2. Did you use all of the collected data? Was the amount of data sufficient for confidently answering your questions (yes/no and why)?

   3. What conclusions did you derive from your experimental results? Which algorithm performed better and why do you believe this was the case? Did you attempt any improvements and what were the observations/conclusions after applying these improvements?

### 6 Future Steps

So far we expect from you to be able the majority of the above questions. However, if you have not, please explain in this section what your future steps will be. These can include steps that you have not taken yet, or ideas that you are planning to implement during the following weeks. In general, describe what further analyses and procedures you would like to perform during the following weeks