

CUNY MSDS Capstone Project  
Draft Proposal

---

**COMMERCIAL BUILDING ENERGY  
CONSUMPTION**

**ANALYSIS AND PREDICTION**

---

February 11, 2019

John Grando  
[john.grando@spsmail.cuny.edu](mailto:john.grando@spsmail.cuny.edu)

---

## Background

Commercial Building Energy Consumption accounts for approximately 25%<sup>1</sup> of the United States energy production profile. Many economical and sociological factors are pushing owners of these buildings to reduce energy consumption and optimize performance. However, it is difficult to say whether a building is operating efficiently or not. Comparing simple summary statistics, such as energy use per square foot (Energy Use Intensity – EUI) is not as simple as it seems because there are a multitude of factors that affect a building’s energy consumption profile. It is very difficult to make similar comparisons and, even then, it is difficult to determine whether one building is performing consistent with other standard practice buildings. In order to determine a level of energy consumption optimization for a randomly selected building, analysis must be performed.

Commercially, ENERGY STAR<sup>2</sup> has implemented a benchmarking algorithm that scores buildings on a scale from 1 – 100 using market-available data. While it is unclear the number of sources used, one is definitely the Commercial Buildings Energy Consumption Survey (CBECS) (). The output of this benchmarking algorithm is a unit-less score, as well as a reference ‘baseline’ building; however, the methodology is not released and it is unclear what factors are important to influence the energy consumption of the building. These barriers make it difficult to provide custom comparisons and nearly impossible to make batch predictions from a set of buildings.

### Source - Commercial Buildings Energy Consumption Survey (CBECS)

CBECS is a national sample survey that collects information on the stock of U.S. commercial buildings, including their energy-related building characteristics and energy usage data (consumption and expenditures). Commercial buildings include all buildings in which at least half of the floorspace is used for a purpose that is not residential, industrial, or agricultural. By this definition, CBECS includes building types that might not traditionally be considered commercial, such as schools, hospitals, correctional institutions, and buildings used for religious worship, in addition to traditional commercial buildings such as stores, restaurants, warehouses, and office buildings<sup>3</sup>.

## Preliminary Analysis

The survey results from the publicly available data set are available in csv format on the EIA website, labeled as a microdata file.<sup>4</sup> This file consists of 6,720 records that represent an estimated 5.6 million total buildings in the United States, using a complex sampling design that is explained in detail in the user’s guide. This data set consists of 1,119 variables, which includes various major fuel consumption values, such as electricity, natural gas, etc. From initial inspection on these proposed response variables,

---

<sup>1</sup>EIA - [https://www.eia.gov/energyexplained/index.php?page=us\\_energy\\_commercial](https://www.eia.gov/energyexplained/index.php?page=us_energy_commercial)

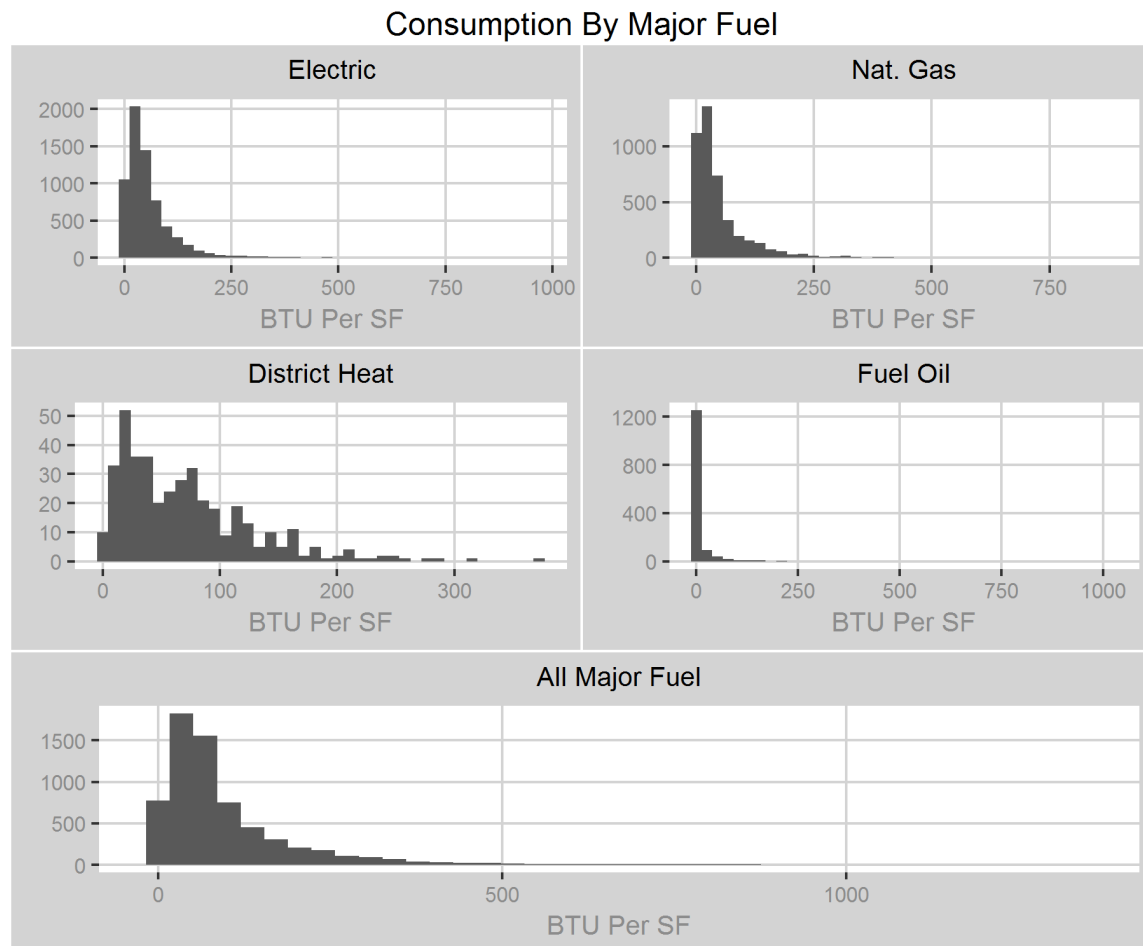
<sup>2</sup>ENERGY STAR - <https://www.energystar.gov/>

<sup>3</sup>[href\(https://www.eia.gov/consumption/commercial/about.php\)](https://www.eia.gov/consumption/commercial/about.php)EIA - <https://www.eia.gov/consumption/commercial/about.php>

<sup>4</sup>microdata - <https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata>

---

after normalizing per square foot, it appears that they have a unimodal right skew, and can possibly be modeled.



## Objective

The goal of this project is two-fold. First, it is important to explore the data set and determine the influencing factors for each planned response variable. Second, a predictive model will be constructed for each major fuel consumption type (Electricity, Natural Gas, Fuel Oil, and District Heat, etc.) as well as an additional model for the overall major fuel consumption, which is the sum of the fuel consumption types noted above. Efforts will be made to explain the model inputs, but the ultimate goal is the performance of predictions. However, the most parsimonious set of predictors will be used amongst the candidate models, in order to maximize our explanatory ability.

## Deliverable

A summary will be provided which will try to explain the domain as well as any major indicators of energy use. Additionally an interface will be provided that can take inputs and provide predictive outputs on anticipated energy consumption.