CUNY MSDS Capstone Project
Proposal

COMMERCIAL BUILDING ENERGY CONSUMPTION

ANALYSIS AND PREDICTION

February 15, 2019

John Grando
john.grando@spsmail.cuny.edu

Background

Commercial Building Energy Consumption accounts for approximately 25% [1] 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 summary statistics, such as energy use per square foot, is not as simple as it seems because there are a multitude of factors that affect a building's energy consumption profile. The complexity of making similar comparisons creates a situation where it is difficult to determine whether a building is performing consistent with, or better than, other standard practice buildings.

Commercially, ENERGY STAR [2] 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[3].

Preliminary Analysis

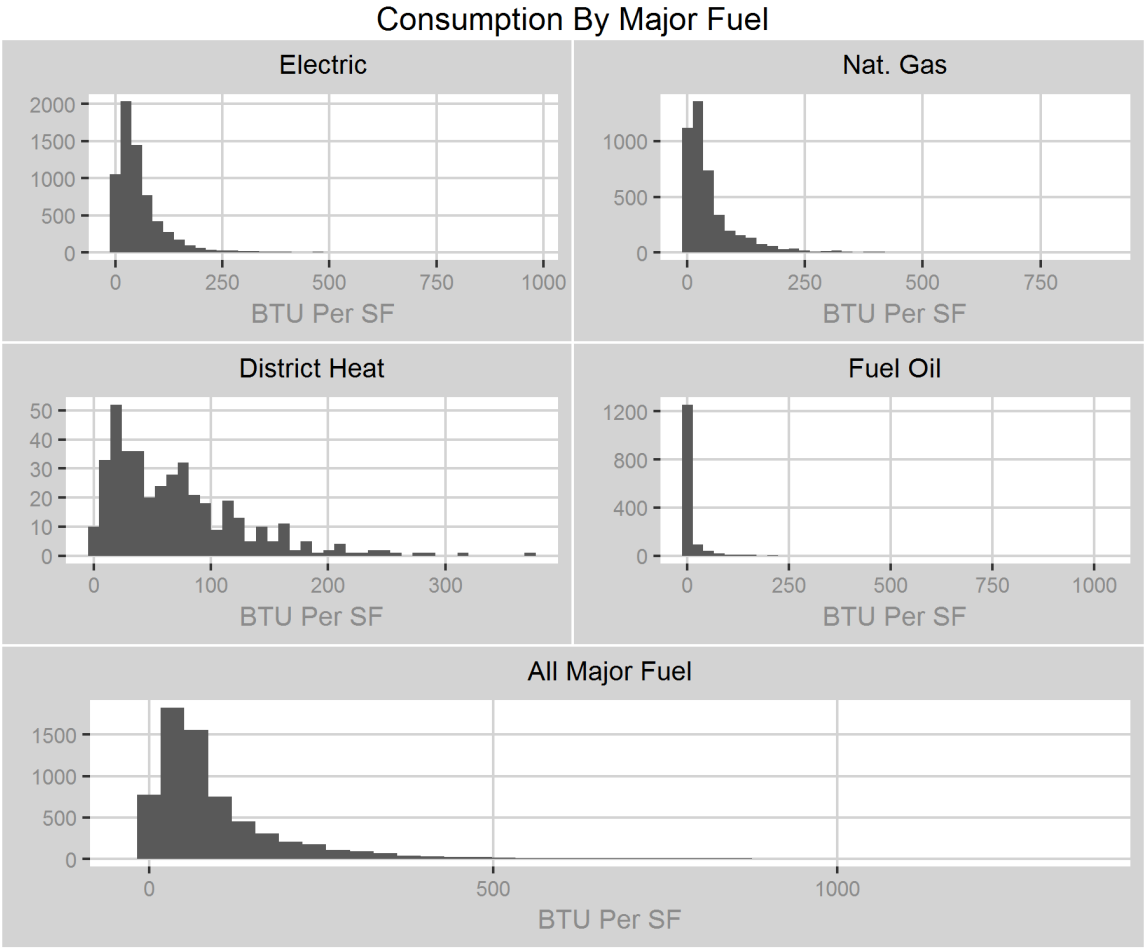
The CBECS survey results from the are available in csv format on the EIA website, labeled as a microdata file. [4] This file consists of 6,720 records that represent an estimated 5.6 million total buildings in the United States, using a complex sampling design that is explained in the accompanied user's guide. This data set consists of 1,119 variables, which include various major fuel consumption values, such as electricity, natural gas, etc. From initial inspection on these proposed response variables, after normalizing per square foot, it appears that they have a unimodal right skew. All fuel sources appear to be operating on relatively the same scale with a few high usage cases that may or may not be outliers.

[1] EIA - https://www.eia.gov/energyexplained/index.php?page=us_energy_commercial

[2] ENERGY STAR - https://www.energystar.gov/

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

[4] microdata - https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata



Objective

The goal of this study is to provide accurate predictions of energy consumption while requiring as few inputs as possible. In order to do this, it is important to explore the data set and determine the influencing factors for each planned response variable. Once variables are selected, a predictive model can be constructed for each major fuel consumption type (Electricity, Natural Gas, Fuel Oil, and District Heat, etc.). Additionally, an attempt will be made to predict total energy consumption using a single model. Given the complexity of the problem, it is unlikely that the final version of any type of regression model would be easily explainable; therefore, the focus will be on the quantity and attainability of the inputs, and accuracy of predictions.

Deliverable

A summary will be provided which will evaluate and justify a subset of selected variables for each fuel source. Additionally, a set of models will be trained in order to provide predictions on major fuel consumption for a proposed building.