DATA 698 Project Proposal 2: Executive Order 88 Energy Analytics & Data Cleansing

Dan Smilowitz

September 25, 2017

Background

On December 28, 2012, New York Governor Andrew Cuomo issued Executive Order 88 requiring that

> By April 1, 2020, all Affected State Entities shall collectively reduce the average EUI [Energy Use Intensity] in State-owned and managed buildings by at least 20% from a baseline of the average EUI of such buildings for State fiscal year 2010/2011

The New York Power Authority was charged with establishing a central management and implementation team to administer the executive order, directed and authorized to (among other responsibilities):

- Take all appropriate measures to ensure that the order's target is met
- Direct Affected State Entities to comply with the requirements of this order
- Provide strategic, technical, and other assistance to each entity to support implementation
- Develop annual milestones for achieving the target
- Develop and implement reporting requirements to document each entity's progress toward the target
- Develop a comprehensive operations and maintenance plan for the State's building portfolio to help achieve no cost and low cost efficiency improvements and ensure that efficiency savings are sustained

The Executive Order (EO88) led to the creation of NYPA's BuildSmart NY program. BuildSmart NY is a key part of Governor Cuomo's Reforming the Energy vision energy strategy and is the premier energy efficiency program referenced in the 2015 New York State Energy Plan.

Problem Description

In order to track state entities' performance against their 2010/2011 baseline, NYPA collects utility bill information for both electricity and all other fuels (natural gas, propane, water, steam, oil, etc.) for all covered facilities and compares usage to the square footage of each facility. This information is reported by each agency and submitted to NYPA in the form of an Excel spreadsheet template.

Presently, agencies are responsible for performing their own data validation — the performance of this task has proved incomplete. Many facilities have shown an increase in reported EUI relative to the baseline year – this has, on investigation, been shown to be due to missing reported data for the baseline year. Additionally, some facilities have been observed to show rather large spikes or dips in their reported data – these are believed to be due to data entry errors and have been excluded from reporting.

In order to meet its reporting requirements for Executive Order 88, NYPA must ensure accurate data is used for the establishment of baselines, and must continue to ensure that reported data is feasible and not the result of data entry errors. With the accuracy of the data ensured, the accuracy of performance reports can be enhanced.

The final portion of this project will be the production of detailed analyses to identify macro trends between and among agencies. Presently, the annual EO88 report produced by the BuildSmart NY team identifies performance at the state, agency, and facility levels, but does not sufficiently highlight trends in the data. The use of advanced analytics will enable the identification of commonalities or differences in performance by facility characteristics to allow for the development of new energy service offerings to help customer meet their required energy reductions.

This problem presents an interesting challenge, as it addresses two key charges faced by NYPA's Commercial Operations department: the Governor's EO88-related charges; and NYPA leadership's efforts to become the nation's first digital utility. BuildSmart NY is an important part of NYPA's energy services strategy, and strongly supports one of its key strategic initiatives through 2020; the execution of this project will improve the accuracy and insights of the analyses and recommendations provided by one of NYPA's flagship customer offerings.

Previous Approaches & Hypothesis

As stated above, EO88 data is reported by state agencies via Excel spreadsheet. This data is presented to NYPA and customer staff through the portal of NYPA's New York Energy Manager (NYEM) platform. Due to the large volume of this data, it has previously been reported only in terms of percent change from baseline year on a facility basis, which can be aggregated by agency or to the entire state portfolio.

The data is hosted on a multi-tenant database owned by a software partner that NYPA initially partnered with in 2013. The vendor has provided extracts of the data, which is organized in a schema that provides a great deal of redundancy and unused database columns; this harms the efficiency of computations based on this data. Further, the data is currently stored in a "wide" format, with each facility having one row for each fiscal year, containing hundreds of variables (multiple units across many fuels), making analysis more cumbersome. Finally, the vendor's platform has been unable to properly handle the quality issues of reported data.

In order to allow for filtering of analysis by facility characteristics and fuels, the data will be converted to more condensed format, following the "tidy data" principles proposed by Hadley Wickham. This transformed data will be stored in a SQL database, as the volume makes in-memory storage of the entire dataset impractical. For missing data, it is expected that values will be imputed reported data for a given facility using other reported values for that facility, in combination with data for facilities of similar size and type, along with weather data. The NYEM platform provides weather data through its portal – this will be used if possible, otherwise another appropriate source of weather data will be identified. A method for flagging possibly-abberant data will also be established, likely based on percentage change from previous months. Those data points flagged as potential errors will be compared to values from the same imputation method for missing data – if there is an intolerable difference between the reported and calculated values, the reported value will be deemed an error and replaced with the imputed value; otherwise the reported data will be deemed feasible and remain.

The cleansed data will be compared not only to the baseline year, but to other facilities reporting. It is expected that useful insights may be provided by isolating performance across and within many building characteristics:

- Location
- Facility type
- Size
- Fuel type(s), including use of renewable energy sources

These findings may be compared to benchmarks from Energy Star's Building Portfolio Manager to determine if state agencies in New York are performing better or worse on aggregate than national standards. Finally, these trends will, where possible, be used to identify possible energy services that NYPA may develop and offer to state agencies to help them meet their EO88 mandates.

It is possible that data may need to be anonymized for academic consumption – if this is necessary, it will be done in a replicable way that will allow reversal so that the results of this project can be utilized within NYPA. The analysis outlined above, including creation and manipulation of a SQL database, will be performed using R. Key findings from this analysis will be contained in a section of the final report for this project. It is possible that a Shiny application may be developed to allow the the BuildSmart NY and Business & Market Development teams to explore performance data.