Mining Energy Efficiency Program Data

Hassan Shaban, ICF International
Haider Khan, ICF International
David Meisegeier, ICF International

ABSTRACT

While much discussion has been devoted to advanced metering data, a lot of insight can also be derived from the vast amount of data that is routinely collected by energy efficiency programs. This data can be very detailed and contains a wealth of information about customers, building and equipment characteristics, measure types, and expected savings. Although most of these datasets do not classify as “big” data, mining them using machine learning and advanced analytics techniques can uncover useful but previously hidden trends. The derived insights can improve program evaluation, customer targeting and customer engagement. In this paper, we provide detailed examples of the insights that can be gleaned by mining data collected by energy efficiency programs. In the first example, we examine a dataset of home improvement program customers using a dimensionality reduction algorithm and explore their energy savings potential. Conclusions from this analysis include that homes smaller than 2,000 sq.ft. or homes built after 1990 have very low energy savings potential. On the other hand, homes built in the 1970s and 1980s and large homes built prior to 1990 have the largest energy savings potential. In the second example, we utilize a machine learning approach to estimate the expected energy savings and recommended efficiency measures for homes without running a complete energy model or requiring an on-site energy audit. This paper demonstrates that, with the usual caveats regarding data type and quality, mining energy efficiency program data has massive potential. Readers will gain a new perspective on leading-edge methods for mining data from existing programs to improve their design and performance.

INTRODUCTION

The major challenge encountered by energy efficiency program implementers is the need to increase energy savings while operating on limited budgets. The total energy savings realized by an energy efficiency program may be improved either by increasing program participation or by increasing the savings achieved per participating customer (York, 2015). Different marketing approaches can be employed to encourage customers to participate in energy efficiency programs, often supplemented with a variety of propensity modeling and customer segmentation techniques in order to make them more effective. Marketing efforts may also be fine-tuned to increase the savings per participating customer by targeting the customers with the largest savings potential.

Energy efficiency programs routinely collect a significant amount of data about participants and participating buildings. Well-established programs have been collecting this data for years, if not decades. The data typically includes geographical information about the participating buildings as well as program-specific pre and post-retrofit information. These multi-dimensional datasets contain a wealth of information about customers, building and equipment characteristics, energy efficiency measure types and expected savings.

Previous studies have exploited this information for a variety of purposes. Fleiter et al. (2012) used audit data collected in small and medium-sized enterprises to explore the barriers to adoption of energy efficiency measures. Hsu (2014) analyzed a multifamily building performance dataset and concluded that building-level data, especially square footage, was more correlated to building energy use than the equipment information typically collected in engineering audits. A very large database of building-level characteristics was recently made available by Mathew et al. (2015), to facilitate the study of the effects of building information on energy use. Kontokosta (2012) and Kontokosta (2015) used very granular building-level data available from New York City's Local Law 84 energy disclosure database, among others, to investigate the significance of different commercial building characteristics on their energy use intensity. Kontokosta (2015) suggested that the wide availability of market-specific energy disclosure data allows the development of more accurate market-specific benchmarking models.

In this paper, we present two examples of how the data from a residential home improvement program may be used for improving program savings by enhancing marketing efforts. The first example shows how the characteristics of buildings with the largest savings potential may be identified using a dimensionality reduction technique, allowing the micro-targeting of these highly desirable customers. The second example is a simple customer engagement tool that uses a machine learning regression algorithm to give users (home-owners) an idea of the energy savings they can expect without requiring an on-site energy audit. Both of these analytical techniques can be added relatively inexpensively to the marketing arsenal of program implementers while reducing implementation costs and improving customer acquisition.

ENERGY EFFICIENCY PROGRAM DATA

Energy efficiency programs collect a significant amount of information about their participants in order to check their eligibility to participate, track pre and post improvement conditions, record installed improvements, estimate savings, calculate incentives etc. The most basic attributes collected about participants are their names and building addresses. Building addresses yields valuable geographic information that could inform the program about the success or failure of its marketing efforts in different locations. The savings and incentives attributed to energy efficiency programs are usually tied to the efficient equipment or service that is being incentivized compared to a baseline (i.e., the existing equipment or current condition in the case of retrofit programs). For example, direct install programs typically require that the auditor document the bulbs and fixtures that were replaced in addition to the efficient bulbs and fixtures that were installed. The annual lighting hours are typically determined based on technical resource manuals or other assumptions. Similarly, HVAC programs collect information about the existing and installed equipment, including type, capacity and efficiency. Programs that offer incentives for HVAC tune-ups, duct sealing and air sealing may require results of AC performance tests, duct leakage tests, and blower door tests respectively. Provided that these tests are conducted using reliable techniques, they may also provide a rich source of information about the effectiveness of these programs. Energy audits required by home improvement programs can provide the most detailed source of information about a home's energy usage. Besides information about the existing equipment, detailed building-level information is collected, such as the building's age, occupancy, and floor area as well as the types of measures that are recommended by energy auditors.

In order to demonstrate the effectiveness that this level of data could provide in program marketing efforts, we used a dataset of audit data for over 9,000 homes in Maryland that participated in a home improvement program run by ICF International. More than 250 attributes were collected for each home by professional energy auditors, including building envelope information, equipment condition and specifications, blower door test results, billing history, recommended measures, and energy savings attributed to these measures using engineering calculations. For this study, a subset of these attributes was selected for further analysis. The selected attributes were as follows:

- Postal code
- Year constructed
- Conditioned floor area (sq. ft.)
- Number of occupants
- Number of bedrooms
- Number of stories above grade
- Orientation of the front of the building
- Main HVAC system type
- Main domestic water heating fuel
- Building type (single family/townhouse, end unit/interior unit townhouse)
- Total annual energy savings resulting from energy efficiency measures

The final attribute in the list above is provided as an output from our analysis while the other attributes are inputs. These attributes were selected because they could be easily provided by typical home occupants, which was one of the goals of this study. Additionally, this was a required and well-validated field in the dataset. A separate table was also created containing the individual energy efficiency measures recommended by auditors for each home along with their associated energy savings.

EXPLORING SAVINGS POTENTIAL USING DIMENSIONALITY REDUCTION

Exploratory analysis and visualization of a dataset can reveal valuable qualitative insights about the trends and correlations between its different attributes. However, datasets that contain a large number of attributes can be difficult to visualize effectively, since regular scatter plots or bar graphs are limited to two or three dimensions, i.e., only two or three attributes can be visualized simultaneously at any one time. One solution to this problem is to use a mathematical technique known as dimensionality reduction. This technique allows datasets comprising tens, hundreds or even thousands of attributes to be displayed on a two or three dimensional plot. Data exploration for the entire dataset may then be performed intuitively, allowing the analyst to uncover trends and correlations as well as dispel myths about expected trends and correlations.

To demonstrate the effectiveness of dimensionality reduction for analyzing large data sets, we used nine of the attributes from the main dataset shown in the previous section (postal code was excluded as it had no significant effect on savings potential). The primary goal was to explore the correlation of the different attributes with the annual energy savings assessed by an energy auditor in order to identify the general building characteristics that may imply large energy savings potential. A dimensionality reduction algorithm was applied to this dataset which reduced the entire dataset to a two-dimensional “map” as shown in Figure 1. Each point on this map represents a single home taking into account all of the attributes in the dataset. Thus, homes that have similar attributes tend to be located closer to each other compared to homes that have very different attributes. It was then possible to apply different coloring to the points based on different attributes, analogous to the application of layers to a geographical map. The axes of this map are mathematical co-ordinates and calculated by the dimensionality reduction algorithm from the input attributes. In effect, this map clusters the different homes using the combination of attributes to determine their location on the map.

We started coloring the data points by potential annual savings to identify the map regions with low and high savings potential (Figure 1). It is clear from the clustering of similar colors in Figure 1 that the data points are stratified by annual energy savings potential, with homes that were assessed the lowest savings lying towards the bottom right of the map and homes with the largest savings lying near the top left. We then overlaid different attributes from this dataset to analyze what their savings potential would be. For example, Figure 2 shows the same two-dimensional map with the data points colored by conditioned area in square feet. The direction of increasing annual savings potential (for energy efficiency measures identified by an auditor) is also indicated in the figure for reference. The smallest homes, with conditioned areas in the range 1,000 – 2,000 sq.ft., are clustered near the bottom left of the map indicating that these are typically assessed less than $200 in annual energy savings. This can be explained by the fact that HVAC loads tend to drive a large portion of energy efficiency savings but smaller homes have smaller loads, and hence smaller savings potential.

Homes are relatively uniformly distributed across the different savings ranges, which indicates that in general, significant energy savings may be realized in homes larger than 2,000 sq.ft. but other factors may affect this potential in individual homes. Some homes smaller than 2000 sq.ft are assessed energy efficiency measures with as much savings potential as some of te largest homes in the dataset.

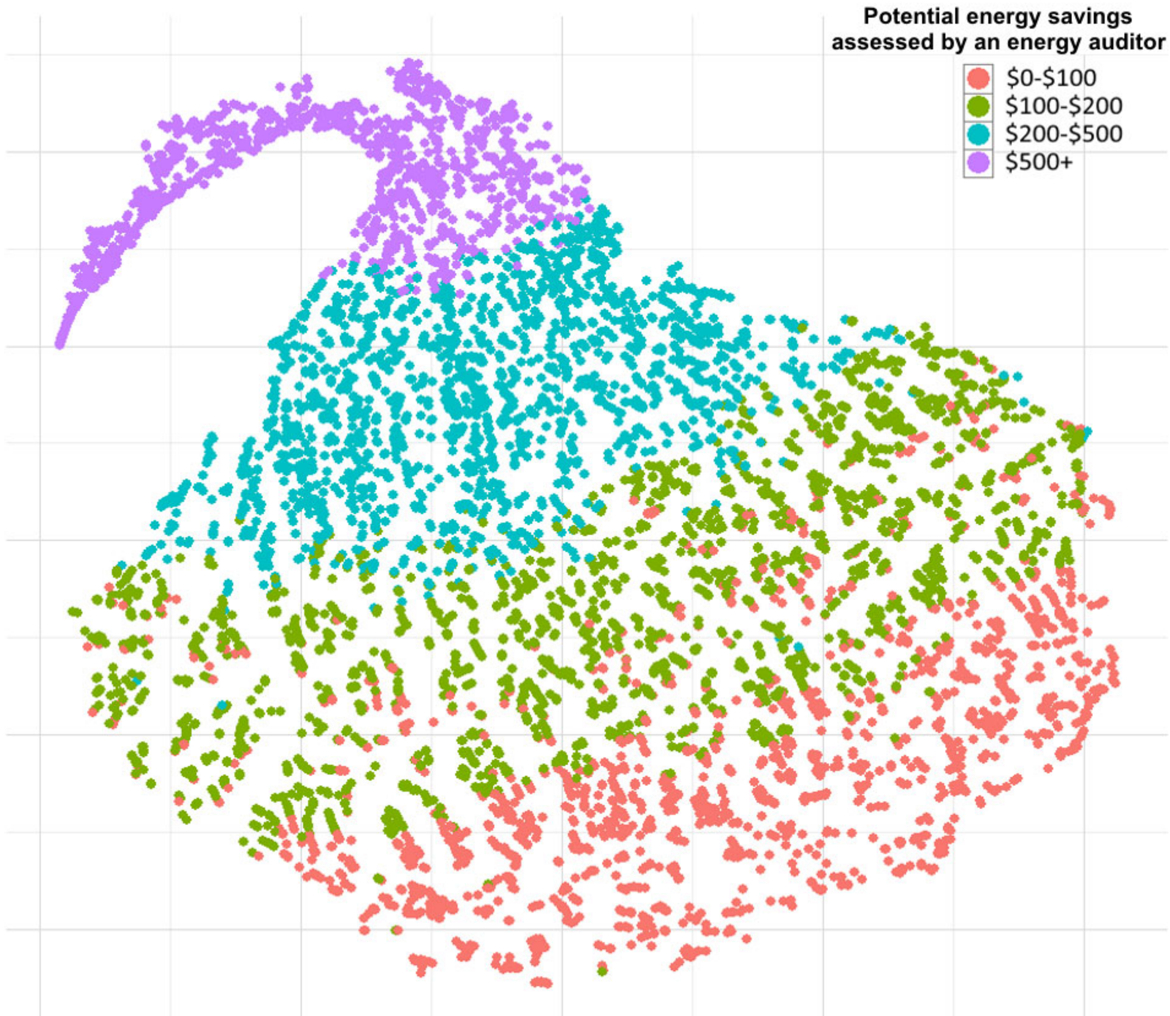

Figure 1. Two-dimensional map showing the annual energy savings ($) assessed by an auditor for homes participating in a home improvement program.

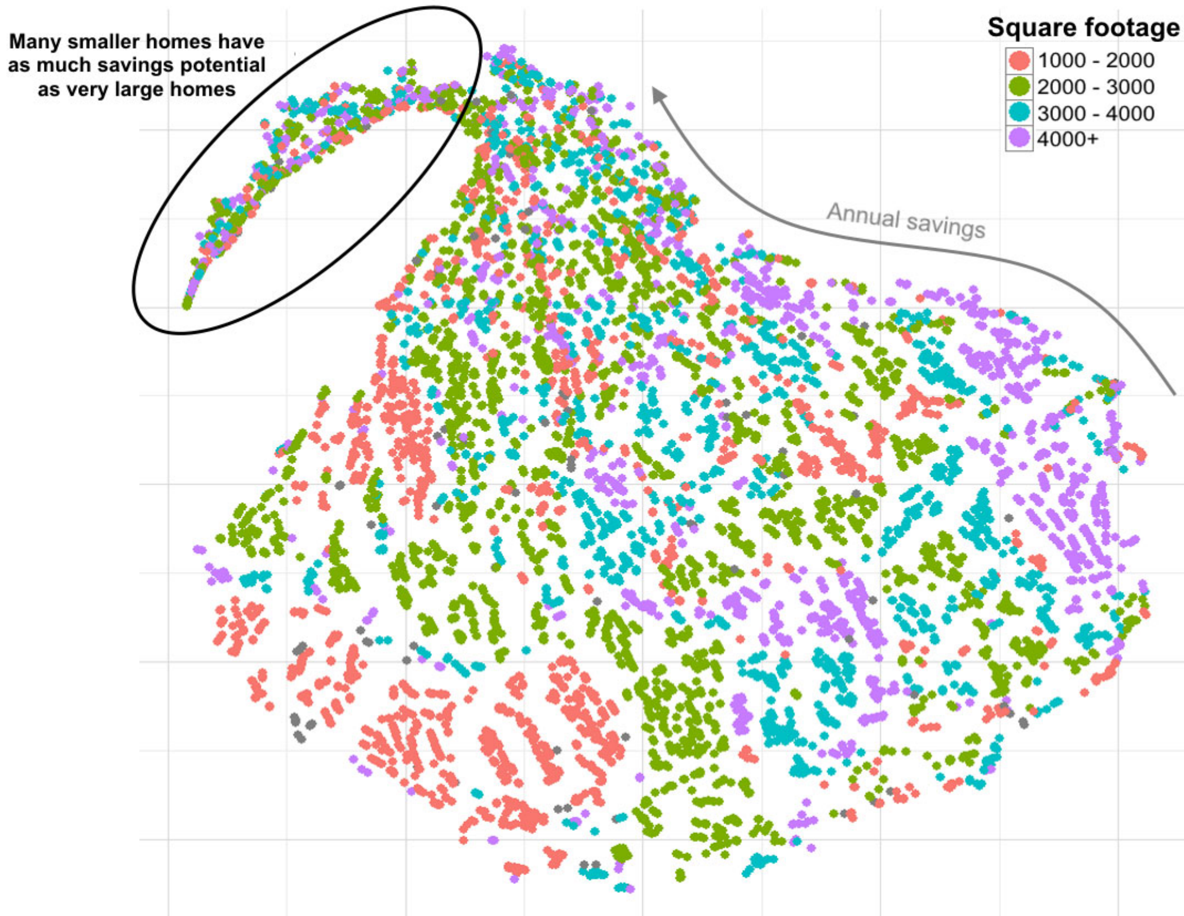

Figure 2. Two-dimensional map showing conditioned area of homes participating in a home improvement program.

The next layer we explored was the age of the home, as indicated by the year it was built. The years were divided into four time periods, as shown in the legend of Figure 3. The most interesting observation is that homes with the largest energy savings potential are those built in the 70s and 80s followed by those built prior to 1950. This may imply that the homes built in this region during these periods were of inferior quality from an energy efficiency perspective (e.g., lower insulation levels, leakier envelopes, etc.) than both older and newer homes. Figure 3 also shows that newer homes are typically assessed very low savings during audits as they are located towards the bottom right of the map. These homes happen to coincide with the larger homes from the Figure 2, which leads us to conclude that newer homes do not have much energy saving potential even if they have very large conditioned areas.

Finally, we explored the impact of the number of occupants on the assessed energy savings potential. Figure 4 shows the homes colored by the number of occupants in the household. It appears that occupancy may not have a very strong effect on energy savings potential and some homes with one or two occupants are assessed with as much savings as those having 5 – 10 occupants. The homes that have a large kWh/occupant would most likely benefit from behavioral measures in addition to other standard energy efficiency measures.

We have limited this analysis to four attributes for the purposes of demonstrating how program implementers could apply dimensionality reduction to large data sets to better guide their marketing efforts. However, other attributes may also be assessed depending on the level of detail required for a given analysis. Conclusions from this analysis include that homes smaller than 2,000 sq.ft. or homes built after 1990 have very low energy savings potential. On the other hand, homes built in the 70s and 80s and large homes built prior to 1990 have the largest energy savings potential. In addition, homes with one or two occupants should not be ignored by marketing efforts since they may have as much energy savings potential as homes with three times as many occupants.

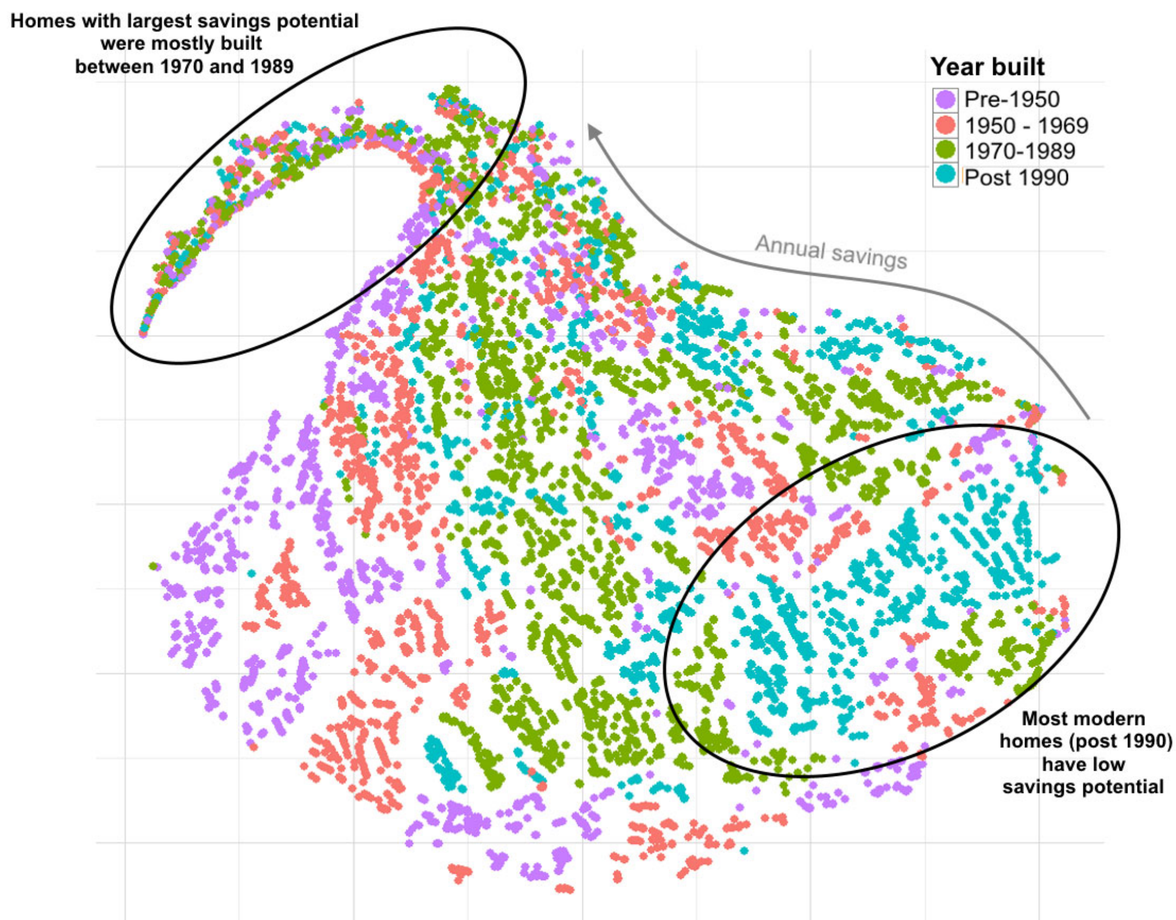

Figure 3. Two-dimensional map showing age of homes participating in a home improvement program.

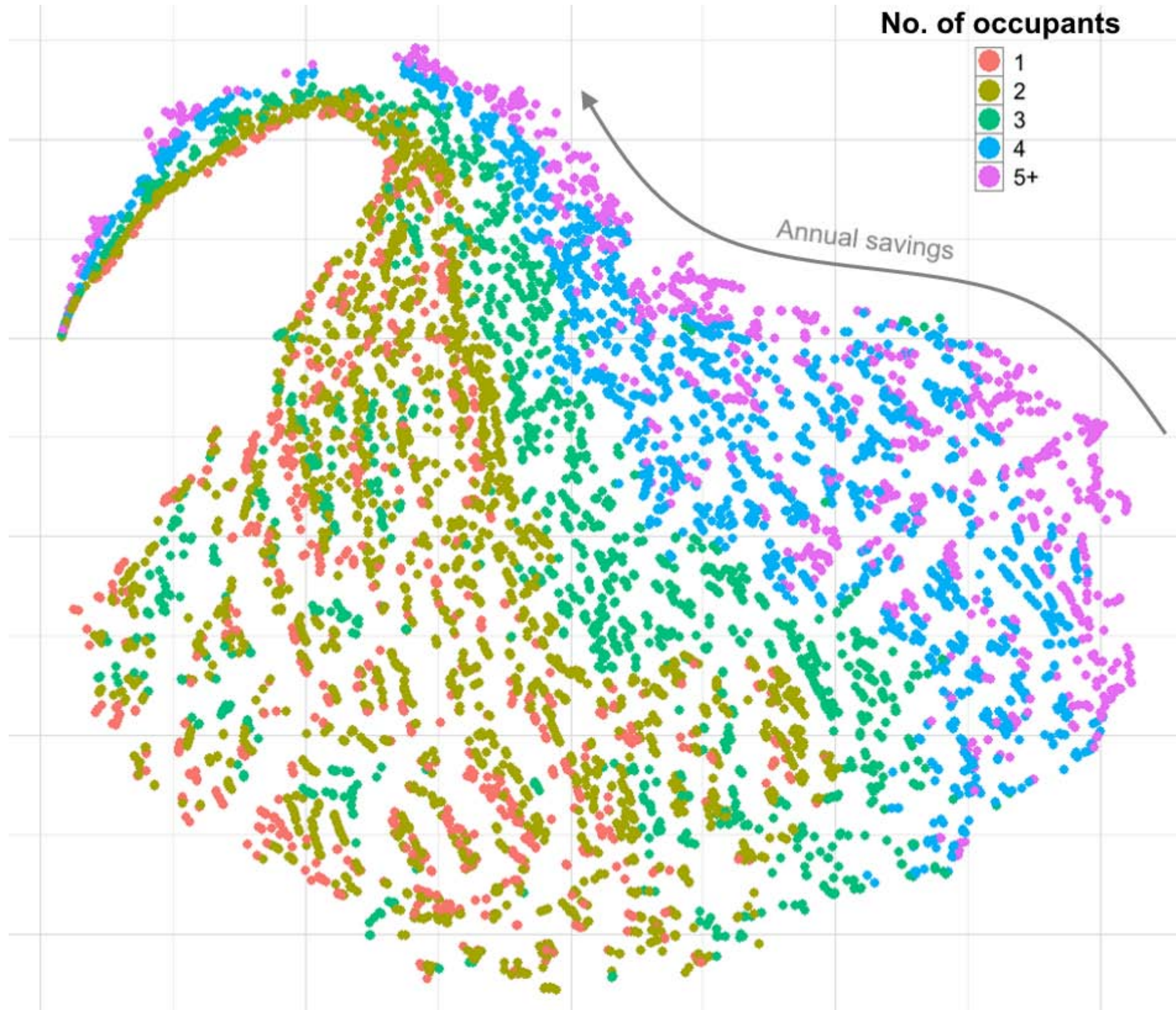

Figure 4. Two-dimensional map showing number of occupants in homes participating in a home improvement program.

PREDICTING SAVINGS USING MACHINE LEARNING

Participants in home improvement programs that have a comprehensive energy audit component are usually required to contribute to the cost of the audit as well as provide their time for being home during the audit. These requirements may turn potential customers away if they are not aware of the benefits of participating in such a program. One way to deal with this issue is to provide a quick estimate of the potential financial benefits that could be identified through a comprehensive home energy audit, if the identified energy efficiency measures were installed. For example, one recent study found that potential energy cost savings are one of the main determinants for participation in retrofit programs (Gamtessa, 2013). In the second part of this paper, we demonstrate how machine learning can be applied to the dataset described earlier to build a model that estimates the energy efficiency potential of a home from very simple user inputs. We designed a customer engagement tool around this model called Beacon Predictive Analytics™ (ICF International, 2015) and packaged it as a web application aimed at residential customers who are considering participating in a home improvement program. A screen shot of this tool is presented in Figure 5.

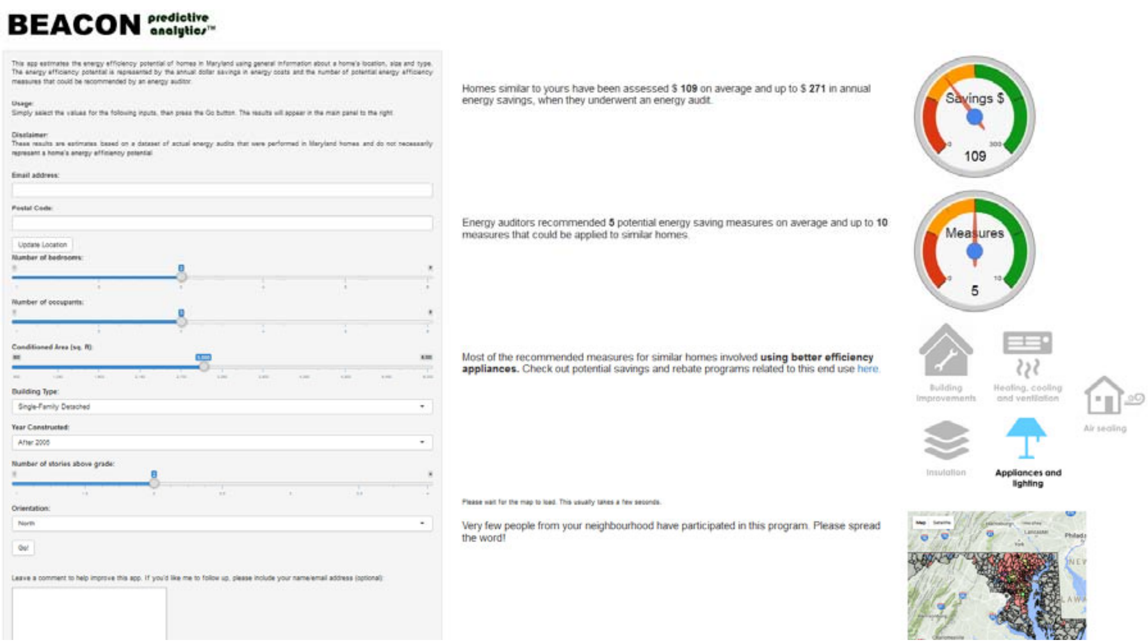

Figure 5. Screenshot of Beacon Predictive Analytics™ web application for estimation of energy savings assessed by an energy auditor.

Users of this tool provide a set of inputs through simple graphical user interface (GUI) widgets such as sliders and drop-down lists. More specifically, the information collected includes the user's email address and postal code, along with some basic information about their home (number of stories, bedrooms and occupants, conditioned area, building type, year constructed and orientation). The tool then uses a machine learning algorithm calibrated with historical participant data to calculate the following three parameters, which are displayed to the user graphically and as text:

- The average and upper limit of energy savings that were assessed from similar homes participating in the program (Figure 6a).
- The average and upper limit of the number of energy efficiency measures recommended by auditors for similar homes.
- The most common end use for which the recommended energy efficiency measures apply, along with a web link to the website of any relevant energy efficiency programs run in their jurisdiction (Figure 6b).

Additionally, a map of the jurisdiction shows the customer what percentage of homes in their zip code have already participated in the program. A custom message is displayed next to the map describing the participation level in the user's zip code compared to other zip codes in the utility's service territory.

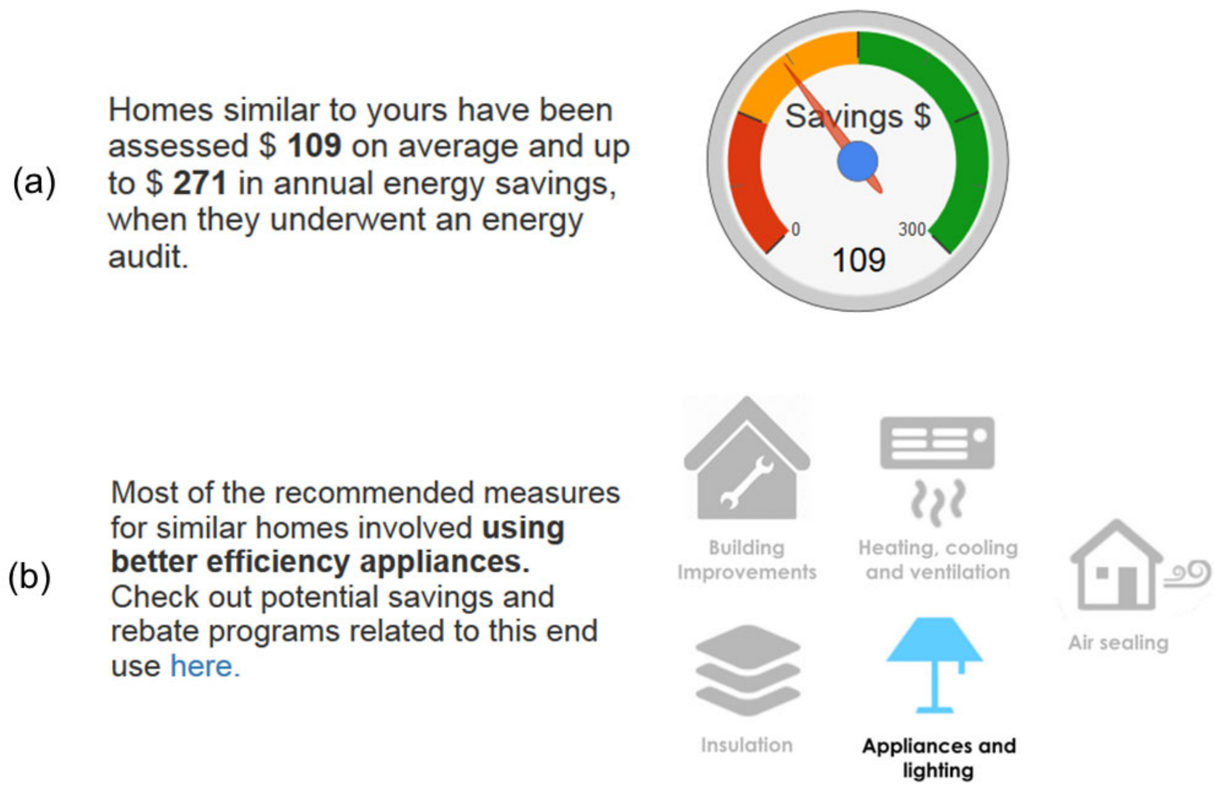

Figure 6. Detail of results displayed by Beacon Predictive Analytics™ (a) Annual energy savings assessed by energy auditors, (b) End use of most recommended measure.

Beacon Predictive Analytics™ gives the non-technical user a quick, quantitative estimate of the financial benefits of participating in home improvement programs as well as qualitative information that could entice them to participate in energy efficiency programs run by their utility. Its simple interface and minimal input requirements make it an ideal customer engagement tool that can give a boost to the marketing and customer acquisition efforts of energy efficiency programs.

CONCLUSIONS

We have presented example applications of advanced analytics techniques for extracting value from the large amount of data that is routinely collected by energy efficiency programs. This data contains a wealth of information about customers, building and equipment characteristics, measure types, and expected savings. The first application demonstrated how multi-dimensional datasets can be easily visualized and explored using dimensionality reduction. The second application was a customer engagement tool that allowed non-expert users to get a quick estimate of the energy savings they could expect to be assessed by an energy auditor. This paper has demonstrated that mining energy efficiency program data has massive potential and can enhance the marketing power of energy efficiency programs.

As part of our ongoing work, we are also applying similar techniques to larger datasets. We are also exploring the possibility of integrating publicly available data sources as well as utility billing history and smart meter data to derive further insights from energy efficiency program datasets, with the ultimate goal of optimizing program designs and increasing savings.

REFERENCES

Gamtessa, S. 2013. "An explanation of residential energy-efficiency retrofit behavior in Canada." *Energy and Buildings* 57: 155–164.

Fleiter, T., Schleich, J., Ravivanpong, P. 2012. "Adoption of energy-efficiency measures in SMEs—An empirical analysis based on energy audit data from Germany." *Energy Policy* 51: 863-875.

Hsu, D. 2014. "How much disclosure of building energy performance is necessary?" *Energy Policy* 64: 263-272.

ICF International. 2015. Beacon Predictive Analytics™ : A customer engagement tool for estimating energy savings potential. Internal Report.

Mathew, P., Dunn, L., Sohn, M., Mercado, A., Custudio, C., Walter, T. 2015. "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use." *Applied Energy* 140: 85-93.

Kontokosta, C.E. 2012. "Predicting building energy efficiency using New York City benchmarking data." 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, USA, August 12 – August 17, 2012.

Kontokosta, C.E. 2015. "A market-specific methodology for a commercial building energy index." *The Journal of Real Estate Finance and Economics* 51: 288-316.

York, D., M. Neubauer, S. Nowak, M. Molina. 2015. *Expanding the Energy Efficiency Pie: Serving More Customers, Saving More Energy Through High Program Participation*. Washington, D.C.: American Council for an Energy-Efficient Economy. http://aceee.org/research-report/u1501.